UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Estimating Ancestry Among South African Ethnic Groups

by

Okuhle Sapo

11061716

Submitted in partial fulfilment of the requirement of the degree

MSc Anatomy (Physical Anthropology)

Contact Details: o.sapo93@gmail.com

Supervisor: Prof. EN L'Abbé

Co-Supervisor: Prof. KE Stull

University of Pretoria

Faculty of Health Sciences

Department of Anatomy

2021

DECLARATION OF ORIGINALITY

UNIVERSITY OF PRETORIA

The Department of Anatomy places great emphasis upon integrity and ethical conduct in the preparation of all written work submitted for academic evaluation.

While academic staff teaches you about referencing techniques and how to avoid plagiarism, you too have a responsibility in this regard. If you are at any stage uncertain as to what is required, you should speak to your lecturer before any written work is submitted.

You are guilty of plagiarism if you copy something from another author's work (e.g., a book, an article or a website) without acknowledging the source and pass it off as your own. In effect, you are stealing something that belongs to someone else. This is not only the case when you copy work word-for-word (verbatim), but also when you submit someone else's work in a slightly altered form (paraphrase) or use a line of argument without acknowledging it. You are not allowed to use work previously produced by another student. You are also not allowed to let anybody copy your work with the intention of passing it off as his/her work.

Students who commit plagiarism will not be given any credit for plagiarised work. The matter may also be referred to the Disciplinary Committee (Students) for a ruling. Plagiarism is regarded as a serious contravention of the University's rules and can lead to expulsion from the University.

The declaration which follows must accompany all written work submitted while you are a student of the Department of Anatomy. No written work will be accepted unless the declaration has been completed and attached.

Full names of student: .....Okuhle Sapo.....

Student number: …….....11061716.....

Topic of work: .....Estimating the Ancestry of South African Ethnic Groups.....
.....
.....

Declaration

1.I understand what plagiarism is and am aware of the University's policy in this regard.

2.I declare that this …Masters Dissertation…… (e.g., essay, report, project, assignment, dissertation, thesis, etc.) is my own original work. Where other people's work has been used (either from a printed source, The Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.

3. I have not used work previously produced by another student or any other person to hand in as my own.

4. I have not allowed, and will not allow, anyone, to copy my work with the intention of passing it off as his or her own work.

SIGNATURE .....  .....

List of Table and Figures

Table of Contents

List of Abbreviations

Abbreviations	Explanation
ANOVA	Analysis of variance
DFA	Discriminant Function Analysis
FARC	Forensic Anthropology Research Centre
FPS	Forensic Pathology Service
SA	South African
SAPS	South African Police Services
TEM	Technical error of measurement
Tukey's HSD	Tukey's Honest Significant Difference test
VIC	Victim Identification Centre

Acknowledgements

I would like to thank my supervisors, Prof. Ericka L'Abbé and Prof. Kyra Stull, that have guided me throughout this journey of doing this research project. The completion of this project would not have been possible without it. I would like to thank the National Research Foundation who funded this project and without the initial financial assistance, I would not have pursued this research project. I would like to thank Gabi Krüger who was always able to assist whenever I encountered a challenge throughout this journey. I would like to thank Meg-Kyla Erasmus who assisted me with the designs of the maps that I used in this project. I would like to thank both Brendon Billings and Nicholas Bacci who aided me with skulls during my data collecting at the Raymond A . Dart Collection.

Finally, I would like to thank my family that have been with me throughout providing the moral support and encouragement to complete this project. Bendingeke ndikwazi ukuyenza lento ngapandle kwenxaso yenu nonke. Mama no Tata imithandazo yenu yemihlanezolo ifezekile ngoku jengoba ndizolanda isidanga sami sesithathu. Ndiswelwe ngama gama ombulelo. Enkosi kakhulu Jola no Majali.

Abstract

The objective of this research project is to assess craniometric differences among the socially defined black South African groups, namely Zulu, Sotho, Pedi, Venda, Tshwane, Tsonga, Swazi, Xhosa, and Ndebele. Current ancestry estimation methods pool these groups into the broad category of black South Africans. This general description of the population as 'black South Africans' may be problematic as various ethnic groups comprise this broad classification. The refinement of self-identification based on ethnicity may improve biological profiles, possibly improving the identification of missing persons from their skeletal remains.

A total of 365 male, adult crania of black South Africans were selected from the Pretoria Bone Collection, the University of Pretoria, and the Raymond A. Dart Collection, at the University of Witwatersrand.

Eighty-five standard cranial landmarks were collected using the 3Skull programme and a Microscribe G2 digitizer (Ousley, 2004). The technical error of measurement (TEM) displayed great intra- and inter-observer agreement. An analysis of variance (ANOVA) was conducted, and twenty-three measurements were found to be statistically significant. A Tukey's Honest Significant Difference (HSD) post hoc test was conducted using the statistically significant cranial measurements and demonstrated that the midface and occipital bones had the most intergroup differences, while the lambda-subtense fraction (OCF) had the most post-hoc pairwise comparisons. The most prevalent inter-group difference was observed between the Swazi and Sotho group. Discriminant Function Analysis (DFA) assessed relationships between size among the groups. A stepwise selection was used to obtain the variables that were best at separating the different groups in the different DFA models. The groups were tested individually, based on geographical location and historical linguistic lineage clusters. The skull was subdivided into the cranial vault, cranial base and splanchnocranium. The various DFA models had overall model classification accuracies that were greater than chance, but their percentages were not high enough to be used for classification purposes in a forensic setting. Clustering the different groups based on their geographical location and historical linguistic lineages resulted in higher overall DFA model classification accuracies than when the groups were assessed as separate groups. The use of historical linguistic lineages may possibly be an alternative manner to refine the black South African classification.

Chapter 1 : Introduction

Public demands are high for forensic scientists to obtain a positive identification of unknown remains found in the South African veldt. With large numbers of migrants in major cities such as Pretoria and Johannesburg, this is a particularly difficult task (L'Abbé and Steyn, 2012, Nienaber, 2015). During 2011-2016, approximately 1 216 258 migrants flowed into the Gauteng province while the Eastern Cape and Limpopo provinces experienced an outflow of 247 437 and 305 030 migrants, respectively (Statistics South Africa, 2016b). Furthermore, 383 345 of the 1 216 258 migrants entering the Gauteng province are from outside South Africa (Statistics South Africa, 2016b). The continual flow of migrants into Gauteng complicates our evaluation of the skeletal remains of the deceased, decreases our likelihood of obtaining identification from the unknown, and contributes to our ever-increasing number of missing people (L'Abbé and Steyn, 2012, Steyn *et al.*, 1997).

South African forensic pathologists frequently ask forensic anthropologists whether a set of unidentified remains can be assigned to a particular ethnic group. In other words, can a forensic anthropologist provide a probability that an unknown person was Pedi, Sotho, Xhosa, Zulu, or whether they can exclude the person from being South African altogether?

If such information is available, even exclusion information, it can assist in narrowing down a list of missing persons and may even improve the rate of positive identification in the country. Yet to reach these goals, research into evaluating differences among ethnically defined South African groups is required. Previous research (De Villiers 1970; Franklin *et al*. 2007) conducted on black South African ethnic groups have produced contradictory information as to the degree of human variation present between and among ethnics groups. The former states that the groups are so similar to one another that it is not possible to separate them sufficiently while the latter visualized observable differences in these groups. Furthermore, the applied statistical analyses in these studies cannot be extrapolated to an unknown person. These studies sought to assess cranial variation among these groups without any inference as to how such information could be useful for forensic purposes. While this research study also addressing human variation among black South African groups, the applied statistical methods that are used will help us to discern whether the available variation among the various groups is useful for establishing a probable ethnicity of an unknown person.

So, what is forensic anthropology? And why are forensic anthropologists focused on using human variation to explain biological parameters (ancestry, sex, age and stature) from the deceased? Forensic anthropology is defined as a scientific discipline that seeks to extrapolate the history, death, and post-life history of a particular individual as depicted by that individual's skeletal remains and the context in which the individual's remains are found (Dirkmaat *et al.*, 2008). Forensic anthropology often focuses on estimating the identification of an unknown person from their skeletal remains but also includes an analysis of taphonomic changes and bone trauma (Symes *et al.*, 2012, Krüger *et al.*, 2018, Spradley *et al.*, 2008, Christensen and Crowder, 2009, Konigsberg *et al.*, 2009, Ousley *et al.*, 2009, L'Abbé and Steyn, 2012, Dirkmaat *et al.*, 2008).

Forensic anthropology is a multidisciplinary field that applies the methods of biological anthropology in a forensic setting (Cattaneo, 2007). A forensic anthropologist typically forms part of a team that assists in constructing the circumstances of a crime scene and an individual death (Cattaneo and Baccino, 2002). In the USA, forensic anthropology is recognised as a forensic science discipline, while in Europe and South Africa, the discipline has yet to acquire an equivalent status (L'Abbé and Steyn, 2012, Cattaneo and Baccino, 2002, Steyn *et al.*, 1997). The reasons for this could be a lack of governmental funding, poor communication amongst academia, forensic medicine, government and industry (NGOs) regarding missing persons in South Africa – aside from those who were disappeared from recent conflicts. Also, there is no nationally recognised organisation where South African forensic anthropologists are professionally accredited which would enhance their prominence in the medicolegal community.

Within a forensic team structure, anthropologists are usually required to create a biological profile from an unknown individual's skeletal remains which contain estimations about sex, age, ancestry and stature (Dirkmaat *et al*., 2008, Spradley *et al*., 2008, Christensen and Crowder, 2009, Konigsberg *et al*., 2009, Ousley *et al*., 2009, L'Abbé *et al*., 2011). The methods for establishing a biological profile, and hence a presumptive identification, from unknown remains, have been used for single cases, war crime victims and in mass disasters (Cattaneo, 2007). For example, the recovery and identification of war victims in Kosovo, Rwanda and the Former Yugoslavia as well as the World Trade Centre disaster (Cattaneo, 2007). However, the standard methodology for establishing a biological profile is temporally and geographically specific and thus needs continual improvement and revisions to maintain high standards, quality assurance and alignment with the goal of victim identification.

Recent scientific advancements have improved the field of forensic anthropology and include advanced statistical analyses for establishing the biological profile; skeletal trauma analysis; and forensic taphonomy (Dirkmaat *et al.*, 2008). These advancements were spurred on from recent challenges to the discipline concerning the application of repeatable, reliable, testable, peer-reviewed, and scientifically valid methods during expert testimony, also known as the Daubert criteria (Christensen and Crowder, 2009, Dirkmaat *et al.*, 2008). These advancements have also broadened the paradigm of thinking regarding the type of information that can be elicited from skeletal remains, especially ancestry (Dirkmaat *et al.*, 2008, Christensen and Crowder, 2009).

At the turn of the century in North America, forensic anthropologists began to use advanced statistical analyses to calculate sex, ancestry and stature from unidentified skeletal remains (Dirkmaat *et al.*, 2008). As all variables of the biological profile are population-specific, which implies that methods can only be accurately used on the populations from which they are derived, researchers began to collect large, representative databases of population groups (Urbanová *et al.*, 2014). The establishing of large, representative modern population databases aided in alleviating problems with outdated reference samples and previously poor scientific studies with little validity or reliability (Dirkmaat *et al.*, 2008).

As technology improved, computational advancements enabled researchers to develop software programs that incorporated multivariate statistical analyses as a means to analyse the population databases. (Urbanová *et al.*, 2014). One software programme was FORDISC which was created under the auspices of the University of Tennessee by Stephen Ousley and Richard Jantz (Elliott, 2008). Three versions of FORDISC software have been released since the first version was developed in 1992 (Elliott, 2008). FORDISC is widely used by forensic anthropologists to ascertain the biological profiles of an unknown person, using population-specific databases, in medicolegal investigations (L'Abbe *et al.*, 2013, Dirkmaat *et al.*, 2008). FORDISC conducts multivariate statistical analyses, primarily discriminant function analysis, of unknown remains based on the several population groups found in the Forensic Anthropology Data Bank (Urbanová *et al.*, 2014). In South Africa, researchers at the Forensic Anthropology Research Centre (FARC) have developed custom crania, dental and postcrania databases for South Africans which are used in FORDISC 3.1.(Shakoane, 2020, Krüger, 2015, Stull *et al.*, 2014, Liebenberg *et al.*, 2015).

When an unknown is compared to one or more of these databases, the individual is classified into a group with a probability or the percentage of chance that the person belongs to that group. Each probability also has a suite of typicalities, which means that if that person belongs to this group, then these values represent how typical that person is of that group to which they may belong. The unknown is classified into the group with the highest probabilities and significant typicalities. Because of human variation within and between groups, an unknown person can only be ascertained as to the probability of belonging to a group which is a statistical suggestion of group membership, not an exact determination. Furthermore, investigation of the unknown case is required to find possible relatives for DNA comparisons.

My research question derives from the frequent inquires of forensic pathologists in South Africa: can forensic anthropologists reliably estimate the presumed ethnicity of an unknown person? In this research project, cranial variation amongst males from ten socially defined South African ethnic groups, namely the Sotho, Pedi, Xhosa, Zulu, Tswana, Tsonga, Swati, Ndebele, and Venda, is investigated using traditional linear measurements and non-standard inter-landmark distances. Various statistical analyses, namely ANOVA, Tukey's HSD are used to describe, compare and evaluate statistically significant relationships among the ten groups. Discriminant function analyses (DFA) are performed on various cranial permutations to decide whether group separation at the level of ethnicity is possible.

Chapter 2 : Literature Review

2.1 Legislation for Unidentified Human Remains in South Africa

The problem of unidentified remains in the Gauteng province dates as far back as 1995 when 2 008 unidentified persons were buried in paupers graves within the province (Steyn *et al.*, 1997). The problem of identification of these remains was further exacerbated by the lack of medical and dental records that could be used to positively identify them (L'Abbé and Steyn, 2012, Nienaber, 2015, Steyn *et al.*, 1997). High rates of interpersonal violence and also of the inadvertently discovered clandestine graves in construction sites in South Africa further highlight the need for the development of forensic anthropology methods for South African groups (L'Abbé and Steyn, 2012).

In most cases of inadvertently discovered graves, it is required to further differentiate between remains that are of forensic and archaeological nature, due to the legal implications of each type (Nienaber, 2015). The National Heritage Resource Act (Act 25 of 1999) in Chapter 2, Part 2: General Protections, section 36 (6a,b) states if any individual encounters an unknown grave during any course of action they must immediately stop that activity (Arts and Culture Department, 1999). The individual must report to the relevant heritage resources authority who must conduct a coordinated investigation with the South African Police Services (SAPS).

If such a grave is covered by the act, then the respective procedures are followed per the Act (Nienaber, 2015, Arts and Culture Department, 1999). The types of graves that are covered by the Act are namely, ancestral graves, royal graves and graves of traditional leaders, graves of victims of conflict, graves of individuals designated by the Minister by notice in the Gazette and lastly other human remains which are not covered by in terms of the Human Tissue Act,1983 (Act No. 65 of 1983)(Arts and Culture Department, 1999, Bernitz *et al.*, 2015). The types of human remains that typically fall within the ambit of the act are those remains that are older than 60 years, which are usually not regarded as being forensic in nature and human archaeological remains that are older than 100 years (Nienaber, 2015).

The inadvertent discovery and improper excavating by SAPS of human remains from heritage sites also cause havoc with regard to the list of unidentified persons within the province. This also creates a backlog for forensic anthropologists who often encounter these remains.

Skeletal remains that are forensic fall under the ambit of the National Health Act (Act 61 of 2003) which stipulates that investigations related to unnatural deaths be under the Regulations Regarding the Rendering of Forensic Pathology Service (R363 of 2007) and the Criminal Procedures Act (Act 51 of 1977) (Nienaber, 2015, Bernitz *et al.*, 2015). The Criminal Procedures Act outlines the procedures that must be followed when obtaining evidence from a crime scene and how criminal investigations must be conducted concerning forensic crime scenes (Nienaber, 2015, Bernitz *et al.*, 2015). The National Health Act, 2003 (Act No. 61 of 2003) stipulates that provincial Forensic Pathological Services are responsible for "the handling and transporting of a body in a designated vehicle to a designated facility" (Department of Health, 2007). The act defines a body as being "a dead human body or the remains thereof. A designated vehicle as being a "public mortuary vehicle, specially adapted in terms of applicable specifications to transport bodies"(Department of Health, 2007). A designated facility is defined by the act as "a medico-legal mortuary specially designed for purposes of storing bodies and where applicable, to perform post-mortem examinations and autopsies"(Department of Health, 2007).

Human remains can only be removed from a crime scene by the Forensic Pathology Service (FPS) once the South African Police Service has provided "written authority" for the removal of the body. The Act further stipulates that a forensic pathologist or forensic medical officer can consult "other qualified professionals and request such professionals to participate in the post-mortem examination and contribute to the further examination of such a body" (Department of Health, 2007). The expertise of a forensic anthropologist would be to aid a forensic pathologist when skeletonised or decomposed remains are sent to the medico-legal laboratories for the examination of the possible cause of death or biological profile of the remains.

Forensic pathologists, as well as forensic anthropologists via the SAPS (Victim Identification Centre), have referred numerous forensic cases to the FARC in the Department of Anatomy at the University of Pretoria. The cases are often decomposed or skeletonized with expertise being required to construct a biological profile, of which ancestry is of particular importance (L'Abbé and Steyn, 2012).

2.2 Forensic Anthropology Research Centre

In 2008, the Department of Anatomy at the University of Pretoria established the FARC(Bernitz *et al.*, 2015). As early as 1993, anthropologists within the Department had been receiving skeletonized remains from law enforcement to establish a presumptive identification (Steyn *et al.*, 1997, L'Abbé and Steyn, 2012, Bernitz *et al.*, 2015). During the 1993 to 1995 period, only 32 cases were received. Of these cases, traditional methods were used with the majority being middle-aged males of probable African origins found in the veldt (Steyn *et al.*, 1997), with various traumatic injuries from suspected ritual killings, sharp force injury and gunshot wounds. None of these remains were identified.

From 2002 to 2012, approximately 555 cases were received, ranging from 31 to 89 cases per year (Bernitz *et al.*, 2015, Steyn *et al.*, 1997, L'Abbé and Steyn, 2012). Similar demographics are seen with the majority constituting middle-aged, black South African males. (L'Abbé and Steyn, 2012, Bernitz *et al.*, 2015). Again, a variety of traumatic injuries were present along with various taphonomic influences of the bones associated with being exposed in the open veldt, such as carnivore damage and burning (L'Abbé and Steyn, 2012). Only one or two individuals were positively identified. In 2009, research was undertaken at FARC to examine ancestry among modern South Africans to create a more accurate biological profile. Research started with an examination of black and white South Africans and later included coloured South Africans. The concentrated focus on ancestry estimation was to establish population-specific methods equivalent to international forensic anthropology standards and forensic applicability (Bernitz *et al.*, 2015). This type of research sought to comprehend further the dynamics of genetic admixture that resulted in the present South African groups (Bernitz *et al.*, 2015). Hence the research focuses on South African black ethnic groups in this current research project.

2.3 South African Populations

South Africa contains approximately 55 908 900 people who are comprised of four major socially defined groups, namely black South Africans (80.7%), with the remaining 19.3% being from white, coloureds and, Asian/Indians South African (Statistics South Africa, 2016b). South African black people can be further divided into their different cultural/ethnic groups namely, Xhosa, Zulu, Swati, Venda, Tsonga, Ndebele, Sotho, Pedi and Tswana. The history of black South Africans has been extensively studied and has generated great debate among scholars.

2.3.1 **Bantu-speaking people history**

Wilhelm Heinrich Immanuel Bleek is credited with coining the word "Bantu" in 1858 to describe the present-day group of languages that are referred to as Bantu-speaking languages (Herbert and Huffman, 1993, Maake, 1991). The term "Bantu," which means people, was used to classify the different Southern African populations into a single race, culture, and language (Hall, 1987, Liebenberg *et al.*, 2015). This classification was justified by the belief at that time that the similarity among languages stemmed from a shared ancestral language (Hall, 1987). The origin of the early Bantu speakers is believed to be in Western Africa most probably in Cameroon or along Cameroon's border with Nigeria (Vansina, 1980, Huffman, 2007, Hall, 1987, Franklin *et al.*, 2007). An estimated 300-800 Bantu languages are spoken throughout the African continent that has diversified from the originally spoken Bantu languages (Herbert and Huffman, 1993, Huffman, 2007, Huffman, 1989).

The migration of the early Bantu speakers represents one of the large-scale movements that occurred before colonisation in Africa (Rexová *et al.*, 2006). The most contested topic among Bantu history scholars is the migration model that enabled them to spread throughout the African continent (Hall, 1987, Vansina, 1980, Herbert and Huffman, 1993, Huffman, 2007, Huffman, 1970, Franklin *et al.*, 2007, Pakendorf *et al.*, 2011). The debates among scholars reside in the reasons that underpinned the need for migrating, the travelled routes while migrating, the areas in which the migrants settled and whether to use linguistic or archaeological data to explain the migration (Franklin *et al.*, 2007, Herbert and Huffman, 1993, Hall, 1987, Huffman, 2007, Vansina, 1980). The early Bantu speakers that resided in Western Africa are believed to have migrated into two subsequent groups (Hall, 1987). The first migrating group moved along the North-Eastern parameters of the equatorial forest and later moved to the inter-lacustrine region associated with both the Urewe pottery and later with Early Iron Age ceramic traditions (Hall, 1987, Herbert and Huffman, 1993, Huffman, 2007, Huffman, 1970, Rexová *et al.*, 2006). The second migrating group traversed the equatorial forest into Angola and Northern Namibia and expanded eastwardly in the first millennium. This group is associated with Late Iron Age pottery (Hall, 1987, Herbert and Huffman, 1993, Rexová *et al.*, 2006). The movements relating to the cultural evolution associated with migration patterns of the Bantu speakers about their pottery style, as illustrated in Figure 1.

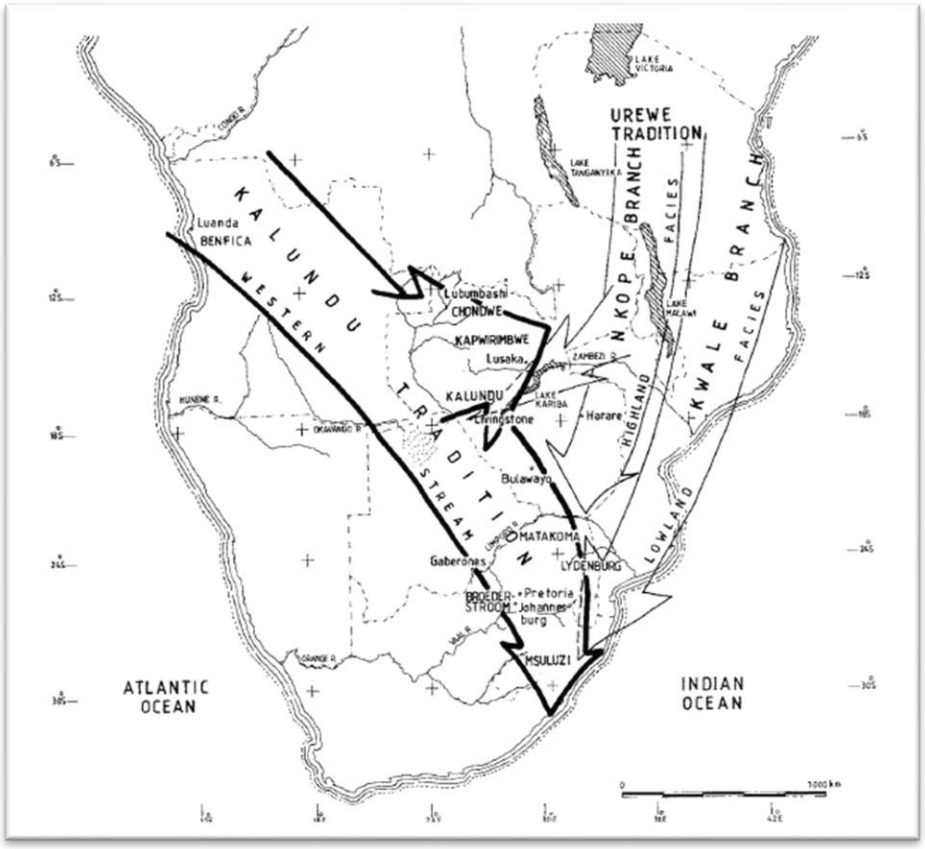


Figure 1. The proposed movements patterns of the Urewe and Kalundu pottery traditions (Taken from Huffman 1989:181).

The archaeological evidence relating to the migration suggests that the Chifumbaze Complex, a commonly observed Early Iron Age ceramic style throughout Eastern and Southern Africa, is divided into two streams (Herbert and Huffman, 1993, Huffman, 2007, Huffman, 1970). The two streams are termed the western and eastern streams and are associated with two different ceramic pottery styles (Huffman, 2007). These streams are also termed as the Kalundu and Urewe traditions with the Urewe tradition being further divided into the Kwale and Nkope branches (Huffman, 2007). The respective streams are associated with both Eastern and Western Bantu speakers who display ideological differences in both their material cultures and life perspectives (Vansina, 1980, Huffman, 2007, Huffman, 2010, Herbert and Huffman, 1993).

The Western Bantu had a matrilineal ideology meaning they believed that the conduit in which people were created was through their mother's bloodline (Huffman, 1970, Huffman, 2007, Herbert and Huffman, 1993, Vansina, 1980). Women in Western Bantu villages played an integral role in the social structure of these societies (Huffman, 1970, Huffman, 2007, Herbert

and Huffman, 1993, Vansina, 1980). Female huts formed the central domain of these villages while male huts were on the outer parameters of the villages (Huffman, 2007). The matrilineal ideology did not ascribe to both the male hereditary leadership practice and the lobola practice (Huffman, 2007, Vansina, 1980). Their marriages were arranged by service to the future in-laws (Huffman, 2007, Huffman, 1970, Vansina, 1980). The Eastern Bantu had a patrilineal worldview about procreation, used the lobola practice as part of their marriage custom, believed in male hereditary leadership and acknowledged the positive role that their male ancestors played in their everyday life (Vansina, 1980, Huffman, 2007, Huffman, 2010, Herbert and Huffman, 1993). The social structure of these communities dictated the activities that both males and females partook in (Huffman, 2007). Pastoralist activities were done by males while females were involved in agricultural activities (Huffman, 2007).

2.3.2 **Early Bantu farmers and the formation of ethnic groups in Southern Africa**

The first Bantu farmers entered the Mapungubwe region of Southern African between 350 and 450 A.D (Huffman, 2007). This region encompasses the valley system near the Shashe-Limpopo confluence and other surrounding landscapes that form part of South Africa, Botswana and Zimbabwe (Huffman, 2009). Around 1000 A.D, drastic socio-political changes occurred in this area and a few elite leaders managed to centralised the channels for trade and cultivation of land for crops (Delius *et al.*, 2014). Two different ceramic pottery styles have been encountered namely, Gokomere and Happy Rest pottery (Huffman, 2007), during this period. These pottery types were found in specific geographic regions relative to the Shashe-Limpopo confluence (Huffman, 2007). The Gokomere pottery was found north of the Limpopo River, while Happy Rest pottery has associated with the region south of the river (Huffman, 2007).

The different climatic conditions that early Bantu farmers encountered influenced the type of food production they adopted within Southern Africa (Hall, 1987). The arid conditions of Western and South-Western regions of Southern Africa were suitable for livestock herding and hunter-gather lifestyle (Hall, 1987), whereas the high annual rainfall of the Central, Eastern and South-Eastern regions of Southern Africa were amendable to crop cultivation and mixed farming economies (Hall, 1987). The settlements that were established by these early farmers were built to access fertile soils along rivers and grazing land (Delius *et al.*, 2014). These farmers introduced various cereal crops into the Southern African region such as sorghum, millet, beans, gourds, and melons (Delius *et al.*, 2014). Domesticated livestock brought into the region included: cattle, goats, sheep, and chickens (Delius *et al.*, 2014), while technological advancements included the

smelting of iron which was used for manufacturing weaponry and agricultural equipment such as hoes (Delius *et al.*, 2014).

While settling into the region, the early Bantu farmers, encountered the Khoi and San who had previously been the only inhabitants of Southern Africa (Franklin *et al.*, 2007). The interactions among these different groups resulted in an exchange of both material goods and culture (Delius *et al.*, 2014). Bantu farmers integrated their own beliefs, knowledge, and language with those of the Khoi and San (Delius *et al.*, 2014). A typical example is the Xhosa people who were in close contact with the Khoi and incorporated parts of the Khoi cultures with their own culture (Franklin *et al.*, 2007). The Xhosa language consists of clicks that are similar to those used in the Khoi and San languages. The population growth that early Bantu settlements experienced due to the adoption of subsistence farming practices resulted in a splitting of the main group of farmers and the migration of these people into unpopulated areas within Southern Africa (Franklin *et al.*, 2007).

Small migrating groups consisted of individuals that were similarly based on language, culture and history (Franklin *et al.*, 2007). This resulted in the formation of the current structure of the Bantu-speaking groups within South African that consist of the Pedi, Sotho, Venda, Swazi, Xhosa, Ndebele, Tsonga and Zulu. The cultural, linguistically and social dynamics of these groups have been the focus of much research within the last 100 years. However, despite their similar origins, are these groups heterogeneous or homogenous? Would we be able to use craniometric variation to distinguish the possible ethnic origin of an unknown person? And can this information be useful in establishing a presumptive identification?

2.4 Ancestry Research in South Africa

South African research studies concentrate on inter-and intra-population variation as a means to create population-specific standards relating to sex, stature, ancestry, and age (L'Abbé and Steyn, 2012). Research on ancestry estimation has evaluated both cranial (Stull *et al.*, 2014) and postcranial data (Liebenberg *et al.*, 2015) along with morphoscopic traits (L'Abbé *et al*. 2011).

The suitability of North American groups, South African groups and African groups from Howell's database using the software program FORDISC 3.0 were evaluated for the estimation of ancestry among 187 black and white South Africans (L'Abbe *et al.*, 2013). As expected, the use of population-specific South African databases in FORDISC 3.0 was pivotal in obtaining the

correct ancestry classification (L'Abbe *et al.*, 2013). This led to the development of custom databases on the craniometrics of white and black South Africans (L'Abbe *et al.*, 2013).

Craniometric and geometric morphometric data were analysed from the crania of three South African population groups, namely South African coloureds, whites and blacks (Stull *et al.*, 2014). The statistical analyses used were general Procrustes analyses, principal component analysis and linear discriminant analysis (Stull *et al.*, 2014). The shape differences from the three South African groups were better extrapolated using geometric morphometrics than traditional craniometrics, with high cross-validation accuracies (89%) (Stull *et al.*, 2014). The study demonstrated also that it was possible to accurately estimate the ancestry of genetically diverse population groups contrary to what has been previously expressed in other studies (Stull *et al.*, 2014). This assertion is affirmed by the high cross-validation percentage obtained using linear discriminant analysis (Stull *et al.*, 2014). The study further showed that geometric morphometrics was better at estimating the ancestry of unknown South African than traditional craniometrics and principal component analysis (Stull *et al.*, 2014).

In many unidentified cases, the cranial remains are not recovered or are badly damaged from either perimortem or post-mortem trauma. In response to this problem, Liebenberg *et al.* (2015) evaluated postcranial data instead of craniometric data to assess the question of estimating ancestry among South African groups. The study used 39 standard measurements from 11 postcranial bones of 360 black, white and coloured South African groups (Liebenberg *et al.*, 2015). The statistical analyses included the analysis of variance (ANOVA) and Tukey's honestly significant difference test. The study used linear and flexible discriminant analyses with bone models and other various multivariate subsets to ascertain which method and model obtained the best cross-validation classification among these South African groups (Liebenberg *et al.*, 2015). The cross-validation percentages from flexible discriminant analysis bone models were slightly better than the linear discriminant analysis bone models obtaining ranges between 41-66% and 41-63%, respectively (Liebenberg *et al.*, 2015). The cross-validation percentages from bone models were lower than those that were obtained from multivariate subsets when using both linear and flexible discriminant analyses and obtained accuracies ranging between 63-85% and 62-87%, respectively (Liebenberg *et al.*, 2015).

The study concluded that the high accuracies obtained from the multivariate subsets illustrated that it was possible to estimate ancestry from postcraniometric data (Liebenberg *et al.*, 2015). This work has advanced our understanding of evaluating ancestry from postcranial remains as

previous studies had low sample sizes and inadequate statistical analyses. The work allows forensic analysts to estimate ancestry when the cranial remains are absent.

Recent research has shown great potential in addressing variation among South African groups, and their application in forensic analyses. However, prior research on black South Africans has also shown some discrepancies in population variation.

Several studies have addressed cranial variation among indigenous Southern African groups with the use of various statistical analyses (Franklin *et al.*, 2007, De Villiers, 1968). In a craniometric and morphological assessment of 745 black South African crania, De Villiers evaluated four main groups: Natal Nguni, Cape Nguni, Sotho, and Tsonga and with the use Student's T-test and Penrose distances statistics, she concluded that a great deal of overlap existed among the groups such that it was "virtually impossible to separate" them (De Villiers, 1968:175). For future research, De Villiers (1968) suggested that an approach needs to combine metric, genetic and epigenetic analyses to resolve the question of craniometric variability among black South Africans (De Villiers, 1968).

With the use of traditional craniometrics and geometric morphometrics, Franklin and colleagues (2007) demonstrated that despite considerable overlap among groups, cranial differences did exist among Bantu-speaking populations (Franklin *et al.*, 2007). In their study, 96 cranial landmarks were used in both traditional craniometrics (De Villiers, 1968, Howells, 1973, Bass, 1971) and geometric morphometrics (Milne and O'Higgins, 2002, Viðarsdóttir *et al.*, 2002) studies. Ten Bantu-speaking groups and two Khoisan groups were selected for analysis. The ten Bantu-speaking groups were namely, Southern Sotho, Xhosa, Zulu, Venda, Shangaan (whose correct classification is Tsonga), Malawi, Swazi, Tswana, Ndebele and finally Kalanga. The statistical analyses that were used to assess intra- and inter-morphological variability between the Southern African populations were principal component analyses and discriminant analyses.

The findings from Franklin and colleagues (2007) showed that the Bantu-speaking groups and the Khoisan groups are morphologically different from one another. The crania of Bantu-speaking groups were morphologically determined as being larger than the Khoisan crania. The study further determined that the most morphologically similar Bantu-speaking groups were the Xhosa, Zulu and Southern Sotho. Swati crania were found to be more morphologically divergent to Zulu crania than it was to other Bantu-speaking groups. This finding contradicts the assertion made by de Villiers (1968) that the Swazi and Zulu can be grouped as the Natal

Nguni due to cultural, historical and linguistic evidence. The findings from the Franklin and colleagues (2007) study also contradicts the main finding of de Villiers (1968) study that there is considerable overlap among the Bantu-speaking groups that makes it impossible to separate the groups morphologically. Though the findings from Franklin and colleagues (2007) adequately address morphological differences among Bantu-speaking people, they were inept in addressing ancestry estimation of blacks South African to the satisfactory standards of the forensic anthropology discipline. The study also did not seek to assess these individuals to estimate ancestry but rather to look for variation among them. This study will evaluate the viability of using the craniometric data from black South African ethnic groups to estimate ancestry.

In addition, this study will be using non-standard inter-landmark distances which have shown to obtain better classification accuracies than traditional inter-landmark distances (Katherine Spradley and Jantz, 2016). Unlike the traditional inter-landmark distance, non-standard inter-landmark distances are not subjected to the limitations that are underpinned using sliding and spreading callipers when measuring shape (Katherine Spradley and Jantz, 2016). These instruments usually oversimplify the form of complex shapes into the domain of only breadths, lengths and heights (Katherine Spradley and Jantz, 2016). While non-standard inter-landmarks are typically oblique to the planes that these callipers use to quantify shape thus provide a better perception of object form(Katherine Spradley and Jantz, 2016).

2.5 Research Objectives

The purpose of this study is to investigate cranial variation among males from ten socially defined South African ethnic groups, namely the Sotho, Pedi, Xhosa, Zulu, Tswana, Tsonga, Swati, Ndebele, and Venda using traditional linear measurements, and non-standard inter-landmark distances.

- To capture 85 external cranial landmarks from male skulls housed in the Dart Collection and Pretoria Bone Collection from ten South African ethnic groups using a G2 Microscribe.

- To capture cranial landmark data that will be analysed using technical measurement of error (TEM), ANOVA, Tukey's HSD, using correlation plots and discriminant function analysis.

- To assess with which region (cranial base, cranial vault and splanchnocranium) in the skull, is best at separating the different groups

- To assess whether clustering the different groups based on their geographical location namely, Zulu-Xhosa, Vend-Tsonga, Sotho-Tswana and Swazi- Ndebele, provides better overall DFA model accuracies than when groups are assessed separately.

- To assess whether clustering the different groups based on their historical linguistic lineages namely, Nguni, Sotho-Tswana and Venda-Tsonga Zulu-Xhosa and Vend-Tsonga, provides better overall DFA model accuracies than when groups are assessed separately.

Chapter 3 : Materials and Methods

3.1 Materials

A total of 365 male, adult crania of black South Africans with known sex and ethnicity were obtained from the Pretoria Bone collection and Raymond A. Dart collections at the University of Pretoria and the University of Witwatersrand, respectively (Table 1). The Pedi group was excluded from the study due to a small sample size thus 357 individuals were used in the study. Therefore, eight groups were analysed instead of nine groups that were initially intended for this study. Only males were used because too few females exist within the skeletal collections for accurate statistical analyses (L'Abbe *et al.* 2005; Dayal *et al.* 2009). Furthermore, the most common unidentified persons in South Africa are males between 20 and 60 years of age.

Table 1. Sample size and digitised individuals from each group.

Cultural Group	Total number of males per group	Number of individuals digitised
Tswana	63	48
Venda	50	37
Xhosa	157	49
Zulu	329	51
Ndebele	41	36
Swazi	69	37
Tsonga	105	49
Pedi	18	8
Sotho	290	50
Total number	1122	365

3.1.1 **Pretoria Bone Collection**

During the establishment of a Medical School in the University of Pretoria in 1942, the Anatomy Department began to collect and retain skeletal material for teaching (L'Abbé *et al.*, 2005). In 1943, the skeletal collection was used to facilitate the learning of undergraduates students and staff members who were involved in medicine orientated disciplines (L'Abbé *et al.*, 2005). The collection consists of individuals at varying degree of skeletal completeness from fully complete individuals to incomplete for either or both crania and postcrania (L'Abbé *et al.*, 2005). This research collection is mainly comprised of black and white South Africans from young to older age adults, with a skewed distribution of males over females. In the 20th century, black South Africans males migrated from rural regions of South Africa for better work opportunities in the mines and other labour-intensive occupations in the larger cities, such as Pretoria and Johannesburg (Malan, 1985, Clark *et al.*, 2007, Ngwane, 2003). Currently, the Pretoria Bone Collection has 1693 individuals that are housed in different accessioned boxes. The Pretoria Bone Collection consists of skeletal material that is either donated or unclaimed (L'Abbé *et al.*, 2005). The annual intake of donated and unclaimed bodies in the Department of Anatomy is between 50-100 individuals (L'Abbé *et al.*, 2005). The Department of Anatomy receives unclaimed bodies from all the local hospitals that are based within the Tshwane Metropolitan Region (L'Abbé *et al.*, 2005). All black South Africans are unclaimed, we know who they are, but no one claimed them (or are looking for them).

Both donated and unclaimed bodies within the collection are often from persons of lower socioeconomic status. This may influence the interpretations that can be made about the general health of the South African population when using this collection. Individuals that are from a lower socioeconomic status may present more with certain skeletal manifestations such as spondylolysis and extensive osteophyte formation, which are associated with strenuous manual labour (Van der Merwe *et al.*, 2010), and well as dental pathologies such as calculus formation which relates to poor oral hygiene (Waldron, 2008)

3.1.2 **Raymond A. Dart Collection**

This collection is situated at the School of Anatomical Sciences at the University of Witwatersrand (Dayal *et al.*, 2009). It is a collection that was established in the early 1920s and houses over 2500 modern human skeletons (Dayal *et al.*, 2009, L'Abbé and Steyn, 2012). The collection consists of 2605 cadaver-derived skeletons where 72% is South African (SA)

African, and the remaining 18% is made up of SA Whites, Coloured, and Indian, Other SA, Other African, Other (Worldwide) and Unknown (Dayal *et al.*, 2009). The collection has a highly male-skewed composition with 1840 males (74%) and 756 female skeletons (26%) (Dayal *et al.*, 2009). Similar to the Pretoria Bone Collection this collection has a highly male-skewed sample size due to the migrant labour system that is present in the country. This increased the likelihood that the male individual would form a large percentage of the unclaimed bodies that were used to comprise these collections.

A limitation that would be beyond our control is whether the ethnicities recorded from these collections were correct. In the periods in which these collections were assembled the proper documentation of black individuals were not seen as a priority. There may have been mistakes done on the recording of an individual's ethnicity, as the people receiving the bodies would have just put them in a general black category, and possible guessed on ethnicity.

3.1.3 **Grouping of the different groups based on geographical location and historical linguistic lineages**

The different groups were clustered in terms of where certain languages are predominantly spoken in the nine provinces of South Africa. The languages that were mostly spoken in geographically adjacent provinces to one another or one province were made to be a cluster. According to the Community Survey 2016 conducted by Statistics South Africa, it found that IsiXhosa is spoken in 82.7% and 31.1% of households in the Eastern and Western Cape, respectively. IsiZulu is spoken in 82.5% of the households in KwaZulu Natal. Setswana is spoken in 71.5% and 33.4% in households in North West and Northern Cape, respectively. Sesotho is spoken in 71.9% of the households in Free State. Tshivenda and Xitsonga are spoken in 17.1% and 16.6% of the household in Limpopo. SiSwati and IsiNdebele are spoken in 29.1% and 10.1% of the household in Mpumalanga, respectively (Statistics South Africa, 2016a). This is formed the basis for the following groupings namely Venda-Tsonga, Swati-Ndebele, Zulu-Xhosa and Sotho-Tswana (Figure 2).

The groups were also clustered based on their historical linguistics lineages that are based on the phonetic similarities of the languages (Zeller, 2004, Maake, 1991). Three of these clusters were formed namely Nguni, Venda-Tsonga and Sotho-Tswana. The Nguni cluster consists of IsiXhosa, IsiZulu, SiSwati and IsiNdebele (Figure 3)(Maake, 1991, Zeller, 2004).These

clusters are not only based on historical linguistics lineages but also as seen in Figure 3 they can be also associated with geographical locations within South Africa.

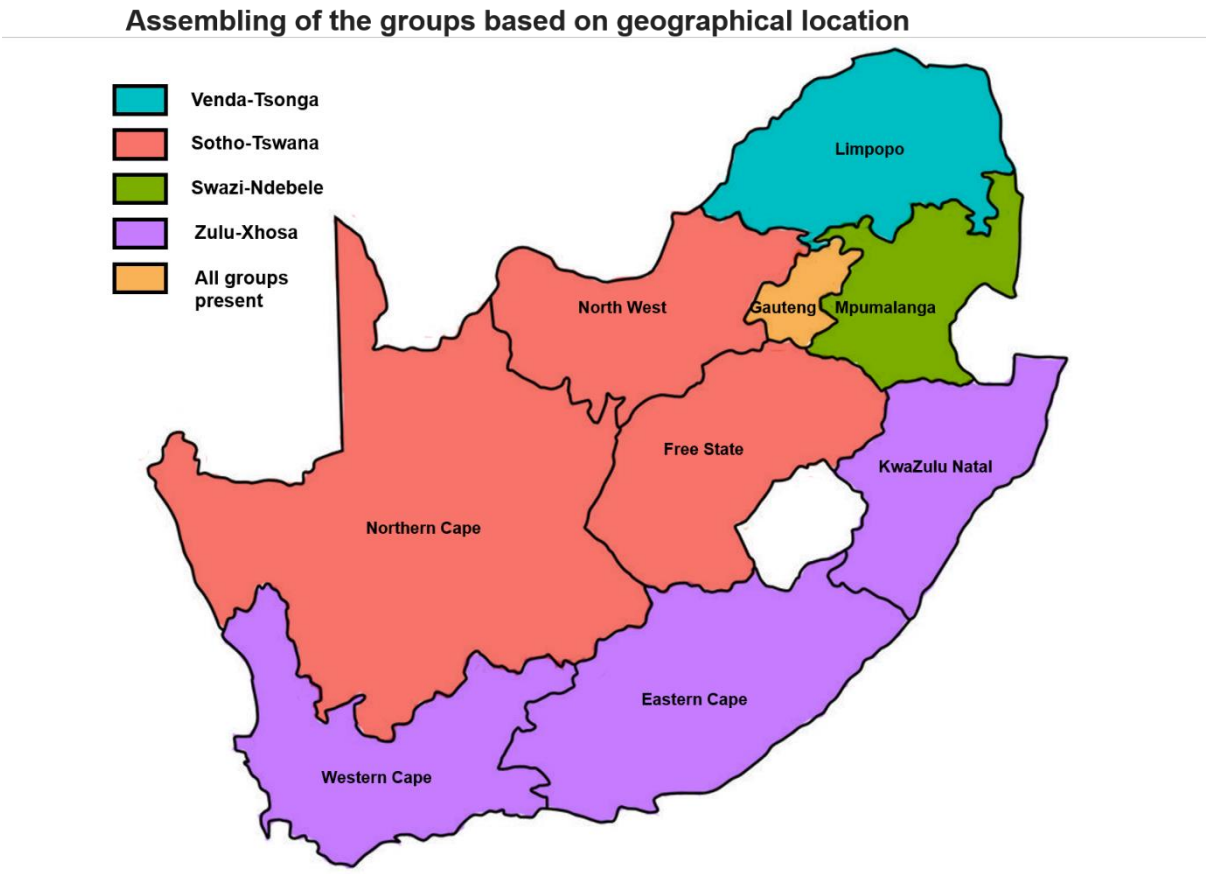


Figure 2. The assembling of the groups based on geographical location.

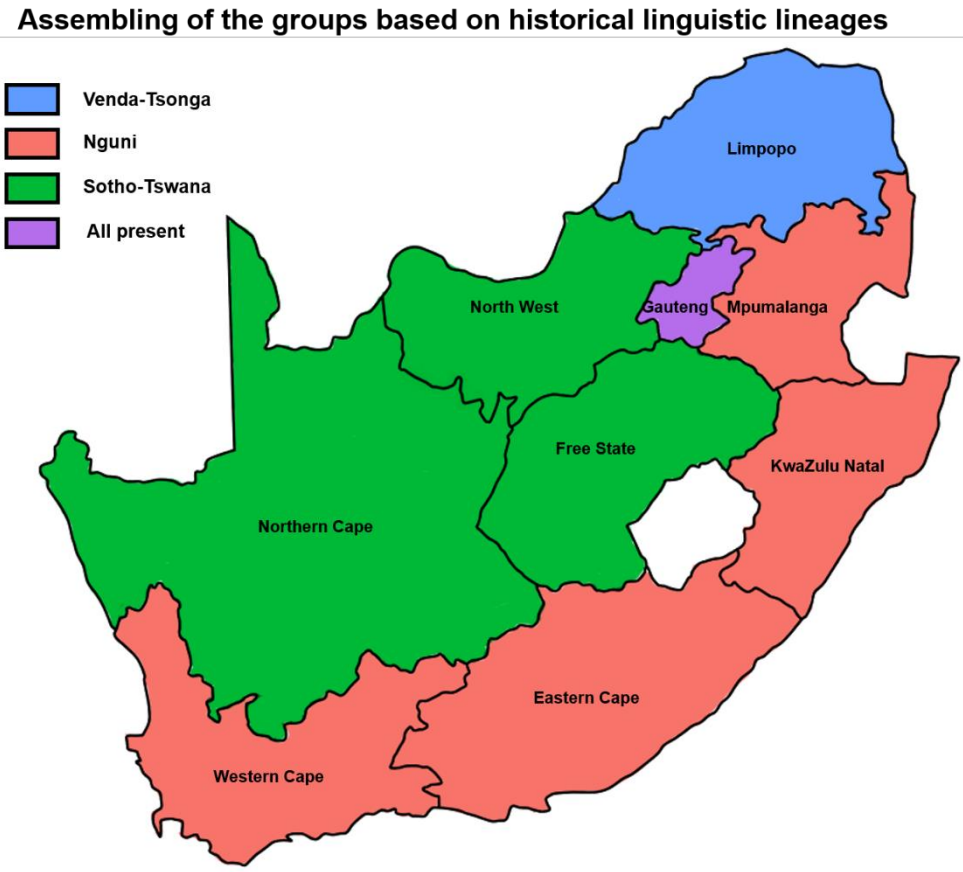


Figure 3. The assembling of the groups based on historical linguistic lineages.

3.2 Methods

Eighty-five standard cranial landmarks were digitised using a Microscribe G2 digitizer with an accuracy of 0.2286 mm (Ousley and McKeown, 2001, Howells, 1973). 3Skull 2.0.176 software was used to aid in the data collection process. Data Collection Procedure

Prior to being digitising the various cranial landmarks, the type 1 and 2 landmarks were marked with a pencil. Cranial landmarks are differentiated into three types, namely Type 1-3. Type 1 landmarks are associated with distinct anatomical locations i.e. where sutures meet such as bregma (Humphries *et al.*, 2015). Type 2 landmarks are associated with points of maximum curvature such as prosthion (Humphries *et al.*, 2015). While type 3 landmarks are located at extremity points where the distance between the points is measured i.e. maximum cranial breadth from euryon to euryon (Humphries *et al.*, 2015). To effortlessly access all cranial landmarks three pillars of clay were used to steady the skull. These pillars of clay are arranged in the shape of a tripod. With the apex of the tripod placed on the dental arcade and the two posterior parallel pillars are placed on the occipital bone. The posterior pillars were placed medial to the mastoid processes so that they did not cover the most inferior point of the mastoid.

One important consideration that was made was to check how close these pillars were to the digitizer. This was done by rotating the arm of the digitizer around the skull and ascertaining whether all landmarks can be digitised including the inferior cranial landmarks.

The accuracy of the X, Y, Z distances were checked before digitising. The 3Skull program software has an option box to test this accuracy. Once this option has been selected, a ruler was used to check the accuracy. The ruler is placed on the flat on a firm surface and orientated horizontally and vertical to check for the accuracy of the X and Y distances, respectively. The accuracy of the Z distances can be checked by slightly slanting the ruler along a vertical surface. If these distances are not accurate usually placing the digitizer in the home position (original position) recalibrates the digitizer.

Figures 2 to 5 illustrates all the standard cranial landmarks and their definitions. For example, landmark 53 refers to the glabella which is the most anteriorly projecting a point in the mid-sagittal plane at the lower margin of the frontal bone and lies above the nasal root and between the superciliary arches. For complete, standard definitions of these landmarks, see Appendix (pg. 82 to 91). Landmarks 78 to 102 refer to the mandible were excluded because of broken, lost and cut mandibles in the two skeletal collections. The actual capturing of the landmark is as follows: the respective landmark is identified on the skull and the tip of the arm of the digitiser is placed on the landmark. A clicker attached to the digitiser is pressed to obtain the X, Y and Z coordinates of the identified landmark. This is the process followed to capture all the external cranial landmarks. This information is then stored in the 3Skull software for each skull and will be exported from the software for analysis.

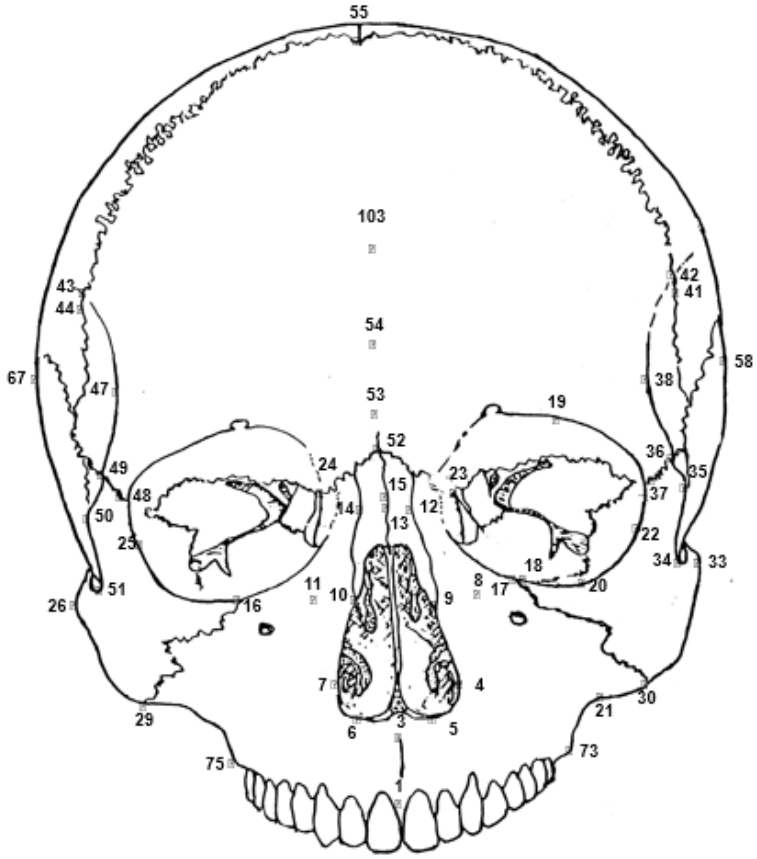


Figure 4. The anterior cranial landmarks that were digitised (Taken from Ousley and McKeown, 2001:178).

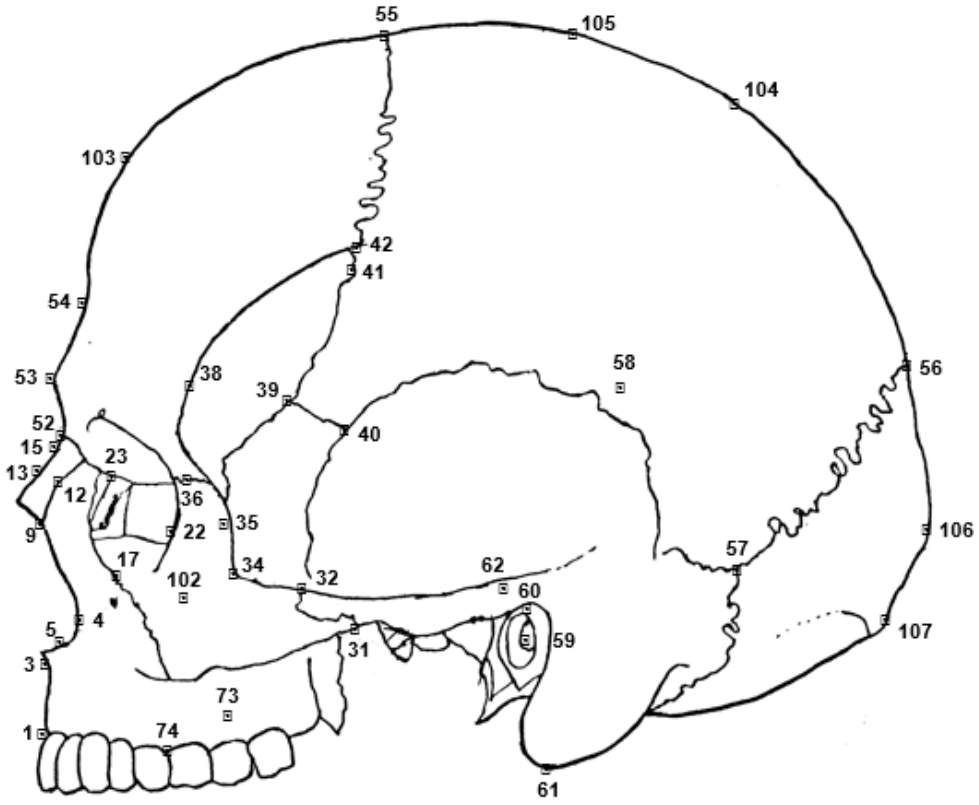


Figure 5. The lateral cranial landmarks that were digitised (Ousley and McKeown, 2001).

Index	Landmark	Abbrev.	Associated Measurement	Location
1	prosthion-estimated	proHEST	BPL, NPH	MS
2	prosthion-ACTUAL	proH		MS
3	Subspinale	ssp	SSR, SSS	MS
4	alare L	alarl	NLB	L
5	most inferior nasal border L	nlhil	NLH	L
6	most inferior nasal border R	nlhir	NLH	R
7	alare R	alarr	NLB	R
8	alpha L	alphal		L
9	nasale inferius L	nasil		L
10	nasale inferius R	nasir		R
11	alpha R	alphar		R
12	nasomaxillary suture pinch L	wnbl	WNB	L
13	nasal bone elevation	sispt	SIS, SIA	MS
14	nasomaxillary suture pinch R	wnbr	WNB	R
15	deepest point on nasal bone profile	ndspt	NDS, NDA	MS
16	zygoorbitale R	zygoor	MOW, IML, XML	R
17	zygoorbitale L	zygool	MOW, IML, XML	L
18	lower orbital border L/R	obhi	OBH (inferior point)	L/R
19	upper orbital border L/R	obhs	OBH (superior point)	L/R
20	cheek height sup point L/R	wmhs	WMH	L/R
21	cheek height inf point L/R	wmhi	WMH	L/R
22	ectoconchion L	ectl	OBB, EKB	L
23	dacryon L	dacl	OBB, DKB	L
24	dacryon R	dacr	DKB	R
25	ectoconchion R	ectr	EKB	R
26	zygion R	zygr	ZYB	R
27	zygotemporale inferior R	zytinr	IML, XML	R
28	zygotemporale superior R	zytspr		R
29	zygomaxilare R	zygomr	ZMB	R
30	zygomaxilare L	zygoml	ZMB, IML	L
31	zygotemporale inferior L	zytinl	IML, XML	L
32	zygotemporale superior L	zytspl		L
33	zygion L	zygl	ZYB	L
34	jugale L	jugl	JUB	L
35	marginal process lateral L	mpll		L
36	frontomalare temporale L	fmtl	Upper facial breadth	L
37	frontomalare anterior L	fmal	FMB, NAS	L
38	frontotemporale L	wfbl	WFB	L
39	Sphenion L	sphl		L
40	Krotaphion L	krol		L
41	Maximum frontal point L	xfbl	XFB	L
42	stephanion L	stpl	STB, STS	L
43	stephanion R	stpr	STB, STS	R
44	Maximum frontal point R	xfbr	XFB	R
45	krotaphion R	kror		R
46	sphenion R	sphr		R
47	frontotemporale R	wfbr	WFB	R
48	frontomalare anterior R	fmar	FMB, NAS	R
49	frontomalare temporale R	fmtr	UFBR	R
50	marginal process lateral R	mplr		R
51	jugale R	jugr	JUB	R
52	nasion	nas	NOL, NLH, NAS, etc.	MS
53	glabella	glb	GOL	MS
54	supraglabellare	spglb	GLS	MS
55	bregma	brg	FRC, PAC, BBH, etc.	MS
56	lambda	lam	PAC, OCC	MS
57	asterion L	astl	ASB	MS
58	eurion L	eurl	XCB	R
59	radiometer point L	radptl	radii (NAR, BRR, etc.)	L
60	porion L	porl	MDH	L
61	mastoideale L	mastl	MDH	L
62	radiculare L	aubl	AUB	L
63	radiculare R	aubr	AUB	R

Figure 6. The respective cranial landmark names that are enumerated in the lateral and anterior skull views, excluding landmarks 78-102 (Taken from Ousley and Mckeown,2001:179)

continuation of Figure 6

Index	Landmark	Abbrev.	Associated Measurement	Location
64	radiometer point R	radptr	radii (NAR, BRR, etc.)	R
65	porion R	porr	MDH	R
66	mastoideale R	mastr	MDH	R
67	eurion R	eurr	XCB	MS
68	asterion R	astr	ASB	L
69	opisthion	ops	FOL	MS
70	basion	bas	BBH, BNL, etc.	MS
71	FOB point R	fobr	FOB	R
72	FOB point L	fobl	FOB	L
73	ectomolare L	ecml	MAB	L
74	M1 anterior point L	avrpt	AVR	L
75	ectomolare R	ecmr	MAB	R
76	hormion	hor		MS
77	alveolon (rubber band)	alv	MAL	MS
78	4 corners	4corn		MS
79	pogonion (vert. projection)	malapt	XRL, MAN	MS
80	gnathion	gniipt	GNI	MS
81	infradentale	gnispt	GNI	MS
82	HMF inf pt	hmfipt	HMF	L
83	HMF sup pt	hmfspt	HMF	L
84	TMF buccal pt	tmfbpt	TMF	L
85	TMF lingual pt	tmflpt	TMF	L
86	gonion L	gonl	GOG	L
87	Left angle base	manlipt	MAN	L
88	Coronion L	coronl		L
89	condylion laterale L	latcndl	bicondylar breadth	L
90	Superior condyle L	supcndlS		L
91	L sup condyle post	supcndlP	CDL, MAN	L
92	condylion mediale L	medcndl		L
93	condylion mediale R	medcndr		R
94	Right superior condyle post.	supcndrP	CDL, MAN	R
95	Superior condyle R	supcndrS		R
96	condylion laterale R	latcndr	bicondylar breadth	R
97	coronion R	coronr		R
98	Right angle base	manripr	MAN	R
99	gonion R	gonr	GOG	R
100	WRB posterior pt	wrbppt	WRB	L/R
101	WRB anterior pt	wrbapt	WRB	L/R
102	maximum malar projection point L/R	mlspt	MLS	L/R
103	metopion	met	FRF, FRS	MS
104	parietal subtense point	paspt	PAF, PAS	MS
105	vertex radius pt	vrrpt	VRR	MS
106	opisthocranion (GOL)	opg	GOL	MS
107	occipital subtense point	ocspt	OCF, OCS	MS

3.3 Statistical Analysis

3.3.1 **Inter and Intraobserver error**

Technical error of measurements, or TEM, is a statistical analysis used to examine the precision of repeated measurements by the same observers or between observers (Stomfai *et al.*, 2011, Harris and Smith, 2009, Liebenberg *et al.*, 2015). The mean that is obtained from repeated measurements of a specimen provides a good estimate of the actual size of the measured specimen (Harris and Smith, 2009). All statistical analyses inherently contain an error in them and reducing the error of repeated measurements increases the between-to-within ratio of variances (Harris and Smith, 2009). The likelihood of obtaining any pre-existing statistical

difference from measured specimens increases by reducing the repeatability error (Harris and Smith, 2009). The equation that is used to measure the absolute technical error of measurement is TEM = $\sqrt{(\sum D^2)}$/ 2N, where D is the difference between measurements and N is the number of the measured subjects (Stomfai *et al.*, 2011). The following equation, R%= 1– (total TEM^2 / SD^2) measures the percentage of the coefficient of reliability which estimates the proportion of between-subject variance of repeated measurements that are free from measurement error (Stomfai *et al.*, 2011). The calculation of the relative technical error of measurement is calculated by using the following equation, % TEM = (TEM/mean) x 100, where the mean is the average of the repeated measurements (Stomfai *et al.*, 2011). The technical error of measurements can be graphically visualised using a Bland and Altman plot (Liebenberg *et al.*, 2015).

3.3.2 **Correlation among landmark variables**

The correlation coefficient is described as the degree or the extent to which two variables are associated with one another(Mukaka, 2012, Asuero *et al.*, 2006, Onwuegbuzie and Daniel, 1999). This correlation coefficient ranges between -1 and + 1, with the sign indicating whether the linear relationship among the continuous variables is positive or negative(Mukaka, 2012, Asuero *et al.*, 2006, Onwuegbuzie and Daniel, 1999). The strength of association between the two variables is at its strongest the closer it is to either -1 or 1 (Onwuegbuzie and Daniel, 1999). The correlation relationships among variables were investigated to ascertain whether the digitised cranial landmarks display collinearity with one another. The correlation relationships between two variables that were being assessed can range from no correlation to a strong correlation. The inter-variable correlations that display no correlation ($r < \pm 0.1$)(Mukaka, 2012, Asuero *et al.*, 2006). The inter-variable correlations that were regarded as having a weak correlation (r = ±0.1-0.5) (Mukaka, 2012, Asuero *et al.*, 2006). The inter-variable correlations that were deemed as being moderate (r = ± > 0.5- < 0.8) (Mukaka, 2012, Asuero *et al.*, 2006). There are inter-variable correlations that were regarded as being strong (r = ± > 0.8-1) (Mukaka, 2012, Asuero *et al.*, 2006). Inter-variable correlations of 0.9 and above variable indicate a collinear relationship between two variables.

The digitised landmarks were used to obtain Procrustes coordinates in MorphoJ and non-standard landmarks. These data sets were imported into FORDISC using the custom import function to build multivariate models based on Mahalanobis distances. These data sets were subjected to the discriminant functional analysis used in FORDISC. The variables that optimise differences among the various groups were elucidated using stepwise selection. Stepwise

selection also minimises the overfitting of data and multicollinearity (Katherine Spradley and Jantz, 2016). Once this has been done interpretations are made on the distances matrices and classification accuracies obtained.

All the individual specimens that were used in this study had a unique identity profile that is only specific to the said specimen. There was no duplicity of identities with regard to the specimens. The total sample size of the individuals that were used in this study is 357. The Ndebele group had the lowest number of individuals at 36 individuals. Five out of the other seven remaining groups had a sample size that were above 40 individuals, with only Swazi and Venda groups had less than 40 individuals. These two groups both had 37 individuals. All the DFA models that have been conducted in this study consisted of ten variables after a stepwise selection had been conducted. Only three groups out of the eight groups do not meet the requirement that there should be at least four times as many samples as there are variables. The sample sizes of these three groups far exceed the minimum sample size of twelve individuals since we are using 10 variables in this study for the DFA models. Before any analysis been conducted for discriminant function analysis, boxplots were used to identify outliers, and some were removed in the process.

3.3.3 **ANOVA**

An analysis of variance (ANOVA) is a statistical method that is used to compare the means of more than two groups (Kim, 2014). This method uses the squared differences between groups and within groups to determine the variance among groups (St and Wold, 1989). This statistical method makes three assumptions to its null hypothesis, namely that the data that is being assessed follow a normal distribution, there is equal residual variance (homoscedasticity) and that there is a transformation of data (St and Wold, 1989). The ANOVA statistical analysis was performed to ascertain whether there are any meaningful differences among the cranial variables that will be used for discriminant function analysis (Kim, 2014, St and Wold, 1989, McHugh, 2011).

3.3.4 **Tukey's HSD Test**

The Tukey's honestly significant differences (HSD) is performed in addition to an ANOVA as a post hoc test since ANOVA analysis does not provide any information with regard to intergroup differences (McHugh, 2011, Smith, 1971). Tukey's HSD test determines statistically significant pairwise differences utilizing a q-value which is a statistic that is where the smallest mean is subtracted from the largest mean among the groups that are being

compared(Smith, 1971, McHugh, 2011). The difference among those two group means is divided by the overall group standard error of the mean (Smith, 1971, McHugh, 2011). A Turkey's honestly significant difference (HSD) test would be further be conducted on the statistically significant variables to ascertain which pairwise comparisons were found in each of these variables (Abdi and Williams, 2010, Smith, 1971). This will provide further information on which intergroup differences are prevalent the most and which groups differ the most from one another.

3.3.5 **Discriminant Function Analysis**

Discriminant Function Analysis (DFA) is a statistical tool used to classify unknown individuals into one or two groups using measurements (Afifi *et al.*, 2011, Ousley and Jantz, 2005, Poulsen and French, 2008). The reference groups used to classify unknown individuals contain demographical information of each individual associated with these reference groups (Ousley and Jantz, 2005). Discriminant function analysis can also be used to ascertain which variables were used to classify an unknown individual to a particular reference group (Afifi *et al.*, 2011). This statistical tool follows a two-step process. First, where the significance of a set of discriminant function equations are ascertained, and then the data is used to classify an unknown individual to a particular reference group (Poulsen and French, 2008). An unknown individual is classified into particular reference groups based on how close the unknown's discriminant score is to the mean score of a certain reference group (Ousley and Jantz, 2005).

The most widely used type of discriminant function analysis is a linear discriminant function which maximises inter-group differences by converting measurements into discriminant function scores (Ousley and Jantz, 2005). The assumptions made when using this type of statistical analysis are: the sample sizes are large and representative, the populations follow a multivariate normal distribution, the amount of variance within the various reference groups is homogenous, there are no outliers and the independent variables used to have low or no multicollinearity (Ousley and Jantz, 2005, Poulsen and French, 2008).

The criterion that is mostly used to determine the effectiveness of the discriminant function model is its classification accuracy percentages (Ousley and Jantz, 2005). Leave-one-out-cross-validation is regarded as one of the best estimators for model classification accuracy (Ousley and Jantz, 2005). Briefly, leave-one-out cross-validation operates in the following manner that an individual in a reference group is removed from the groups and the parameters of the model are re-evaluated using the remaining individuals in the different groups to classify the removed

individual (Ousley and Jantz, 2005). This procedure is done to all individuals in the different reference groups, once this has been completed, overall cross-validation is obtained for the discriminant function analysis model (Ousley and Jantz, 2005). Non-standard inter-landmark distances obtained from the 3Skull software were used to conduct this analysis. Inter-landmark distances are the measured distances between landmarks in two- or three-dimensions (Katherine Spradley and Jantz, 2016).

3.4 Ethical Considerations

Ethics approval for this study was obtained in 2017 (Ethics Reference No: 433/2017). The human remains that were used in this study were handled in accordance with the National Health Act, 2003 (Act No. 61 of 2003) and the Human Tissue Act (Act 65 of 1983).

Chapter 4 : Results

4.1 Technical error of measurement

Intra- and inter-observer error rates in cranial measurements as measured by the absolute technical error of measurement (%TEM) are presented in Table 2. %TEM was used as it removes the effect of size and the results are given in percentages. All error rates were below the standard cut-off of 7.5% for all cranial measurements, except the NDS (naso-dacryal subtense), GLS (glabella projection) and OCF (lambda-subtense fraction). All the non-repeatable cranial variables were non-standard landmarks and were removed from further analyses. All standard cranial landmarks were within the acceptable levels of %TEM. The intra-observer error rate ranged from 0 to10.908 (lowest to highest) and the inter-observer error rate ranged from 0.273 to 20.328 (lowest to highest). Repeatability was generally better with intra-observer tests than with inter-observer.

Most of the measurements in the intra-observer Bland and Altman plot are clustered within a 1 mm difference and have fewer measurements than exceed a difference of 2 mm when compared to the inter-observer agreements (Figure 7). The inter-observer Bland and Altman plot has a similar clustering of measurements as the intra-observer plot with most measurements falling within a 2 mm difference. However, the inter-observer plot has more measurements (n=13) than the intra-observer plot (n=9) that exceed a difference of 2 mm but which still lie within a two standard deviation mean difference (Figure 8).

Table 2. TEM for both inter- and intra-observer for cranial landmarks rounded off to three decimal points. Bold indicates variables that were not repeatable (relative %TEM > 7.5).

Standard Inter-landmark Distances				
Measurement	**Inter-observer**		**Intra-observer**	
	TEM	**%TEM**	**TEM**	**%TEM**
GOL	0.775	0.421	0.500	0.270
BNL	0.707	0.705	0.671	0.667
BBH	0.387	0.293	0.387	0.293
XCB	2.260	1.719	2.086	1.585
XFB	1.830	1.610	0.975	0.855
WFB	1.533	1.594	0.632	0.666
ZYB	0.500	0.389	0.316	0.250
AUB	0.316	0.273	0.224	0.192
ASB	1.224	1.146	0.447	0.411
BPL	0.707	0.701	0.289	0.293
NPH	0.866	1.322	0.577	0.916
NLH	0.707	1.464	0.775	1.600
NLB	0.387	1.464	0.387	1.498
MAL	1.199	2.108	1.436	2.605
MDH	1.746	6.362	0.592	2.040
OBH	0.548	1.606	0.447	1.253
OBB	0.671	1.713	0.548	1.392
DKB	0.775	3.162	0.447	1.966
FRC	0.388	0.344	0.316	0.282
PAC	0.837	0.750	1.360	1.212
EKB	0.922	0.928	0.592	0.602
OCC	0.949	0.981	1.224	1.240
FOL	0.742	1.991	0.806	2.127
FOB	1.360	4.572	1.049	3.394
UFBR	1.581	1.509	0.632	0.613
Non-Standard Inter-landmark Distances				
Measurement	**Inter-observer**		Intra-observer	
	TEM	**%TEM**	**TEM**	%TEM
NOL	0.548	0.301	0.387	0.211
JUB	0.387	0.340	0.447	0.403
NDS	1.500	**13.100**	0.408	3.733
WNB	0.294	3.030	0.203	2.275
SIS	NA	NA	0.129	3.285
ZMB	1.517	1.563	0.592	0.635
SSS	0.775	3.383	0.806	3.416
FMB	0.548	0.548	0.387	0.392
NAS	0.316	1.807	0.000	0.000
DKS	0.922	7.715	0.806	6.950
IML	1.025	2.842	0.632	1.815
XML	0.806	1.514	0.387	0.754

MLS	NA	NA	0.624	5.065
WMH	0.949	4.352	1.204	5.874
GLS	0.447	**20.328**	0.224	**10.908**
STB	2.225	2.014	2.617	2.385
FRS	0.387	1.391	0.447	1.641
FRF	2.049	4.124	2.617	5.082
PAS	0.500	2.012	0.806	3.623
PAF	1.746	2.938	3.428	5.956
OCS	0.500	1.739	0.922	3.073
OCF	2.012	4.162	4.266	**8.448**
NAR	1.360	1.452	0.316	0.334
SSR	0.837	0.875	0.707	0.738
PRR	1.118	1.073	0.408	0.401
DKR	1.000	1.215	0.742	0.891
ZOR	0.742	0.912	0.316	0.391
FMR	1.183	1.543	0.894	1.138
EKR	1.414	1.964	0.894	1.223
ZMR	1.204	1.644	0.592	0.814
BRR	0.671	0.577	0.548	0.469
VRR	0.632	0.517	0.632	0.524
LAR	0.632	0.587	0.592	0.543
OSR	1.225	2.937	0.866	2.082
BAR	0.922	5.854	0.866	5.550
MOW	1.549	2.648	0.447	0.768
NAA	0.612	0.855	0.463	0.727
PRA	0.661	0.937	0.707	0.982
BAA	0.433	1.140	0.289	0.772
NBA	0.632	0.828	0.447	0.584
BBA	0.387	0.691	0.387	0.696
BRA	0.447	0.938	0.447	0.936
SSA	1.533	1.185	1.432	1.135
NFA	0.671	0.474	0.000	0.000
DKA	2.950	2.043	2.280	1.566
NDA	5.514	5.841	2.147	2.298
SIA	NA	NA	2.603	2.518
FRA	0.894	0.705	0.671	0.525
PAA	0.975	0.739	1.140	0.834
OCA	1.245	1.052	2.049	1.748
RFA	0.775	1.212	0.632	1.002
RPA	0.775	1.302	0.866	1.457
ROA	1.140	1.798	1.775	2.743
BSA	3.373	2.035	2.864	1.733
SBA	0.837	0.799	0.592	0.559
SLA	0.447	0.517	0.500	0.582
TBA	1.844	1.233	1.581	1.053

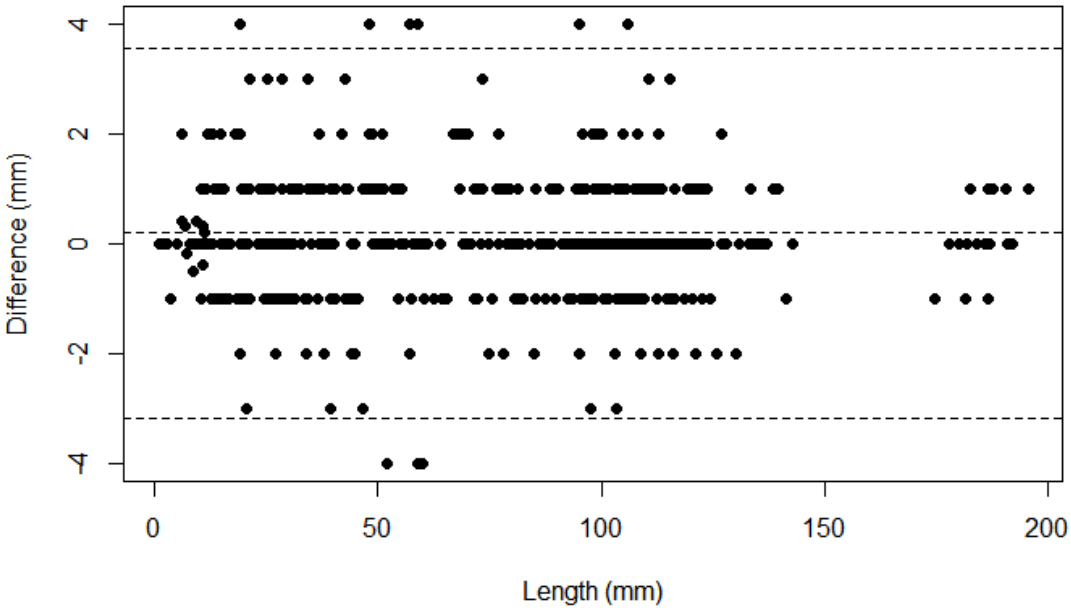


Figure 7. Bland and Altman plot showing the intraobserver error, from ten randomly selected individuals, all but nine of the measurements fall within the upper and lower agreement levels indicated by the dashed line**s.**

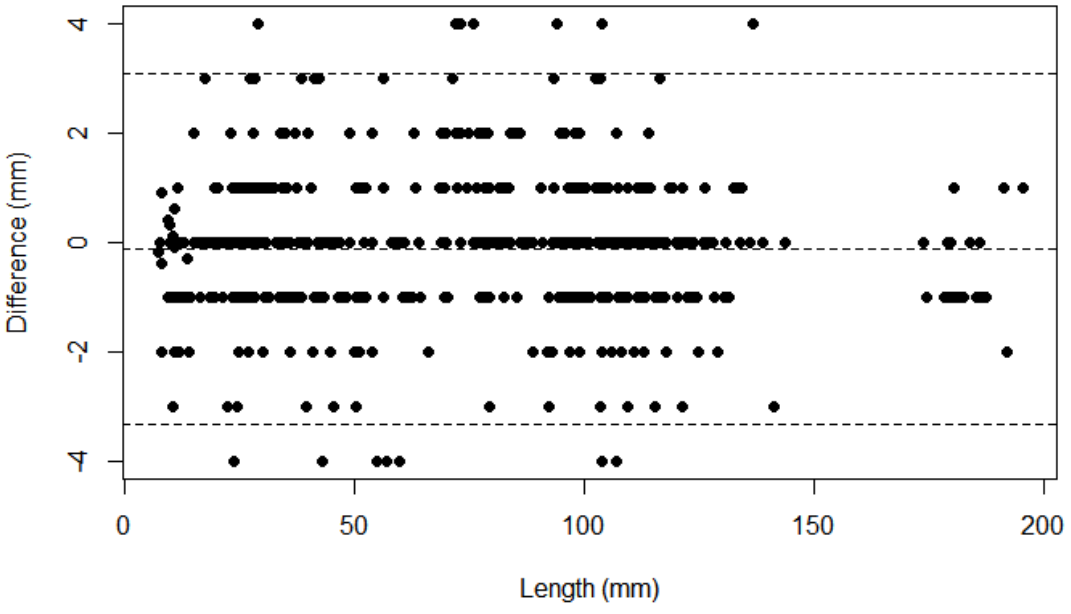


Figure 8. Bland and Altman plot showing the inter-observer error, from ten randomly selected individuals, most of the measurements fall within the upper and lower agreement levels indicated by the dashed lines except for 14 measurements.

4.2 Descriptive Statistics

4.2.1 **Standard Inter-landmark Distances**

Minimum frontal breadth (WFB) was the only standard inter-landmark distance taken on all individuals (n = 357) and maximum alveolar length (MAL) had the smallest sample size (n = 149). The variable that had the greatest deviation from the mean (SD) was the cranial length (GOL) with a standard deviation of 6.034 while orbital breath (OBB) had the least amount of deviation from the mean with a standard deviation of 1.774 (see Table 3).

4.2.2 **Non-standard Inter-landmark Distances**

Nasal subtense (NAS) and the WHM had the largest sample size (n = 355) while simotic subtense (SIS) had the least number of individuals (n = 260). The non-standard inter-landmark distance with the greatest standard deviation was the nasio-occipital length (NOL) (5.959) and the non-standard inter-landmark distance smallest standard deviation was SIS (1.031) (see Table 3).

Table 3. Descriptive statistics of the standard and non-standard landmark distances, n = no of individuals consisting of that particular measurement.

Standard Inter-landmark Distances			
Measurement	**N**	**Mean**	**SD**
GOL	353	187.6	6.034
BNL	349	101.8	4.011
BBH	352	132.9	5.237
XCB	353	128.1	5.472
XFB	353	114.1	5.086
WFB	357	97.34	4.622
ZYB	311	128.3	5.053
AUB	352	117.5	4.569
ASB	351	107.0	4.659
BPL	151	102.4	5.017
NPH	153	68.4	4.092
NLH	355	48.06	2.779
NLB	356	28.25	2.249
MAL	149	57.32	3.311
MDH	351	31.29	3.208
OBH	354	33.99	2.001
OBB	356	40.06	1.774
DKB	356	25.29	2.659

EKB	353	100.9	3.942
UFBR	353	105.8	4.224
Non-Standard Interlandmark Distances			
Measurement	**N**	**Mean**	**SD**
NOL	353	185.2	5.959
JUB	345	115.2	4.711
NDS	264	9.561	1.593
WNB	351	9.119	2.126
SIS	260	2.765	1.031
ZMB	354	94.97	5.250
SSS	349	24.87	3.091
FMB	354	101.7	3.989
NAS	355	18.02	2.526
IML	352	37.6	3.452
DKS	354	12.66	2.324
MLS	319	12.76	1.779
WHM	355	19.26	2.402
STB	351	110	5.698
ZOR	350	82.03	3.662
VRR	350	121.5	4.612

4.3 Correlation plots

The linear relationship among standard and non-standard cranial variables was evaluated using Pearson's product-moment correlations with Holms adjustments (r-squared) and visualised using a corrplot (Figure 9). The correlations among variables were investigated to explore variable relationships. Inter-variable correlations of 0.9 and above indicate collinearity. A total of 195 correlations were observed that ranged from weak to strong relationships. The results displayed minimal collinearity among the variables. Forty-four (23%) inter-variable correlations displayed no correlation ($r < \pm 0.1$), whereas 114 (58%) inter-variable correlations were weak ($r = \pm 0.1\text{-}0.5$). A total of 30 (15%) inter-variable correlations were deemed to be moderate ($r = \pm > 0.5\text{-} < 0.8$), while seven (4%) inter-variable correlations were strong ($r = \pm > 0.8\text{-}1$). Three variable relationships displayed multi-collinearity, namely NOL-GOL, UFBR-FMB and FMB-EKB. The NOL-GOL correlation coefficient was the highest among the three collinear relationships at a value of 0.98 (Table 43).

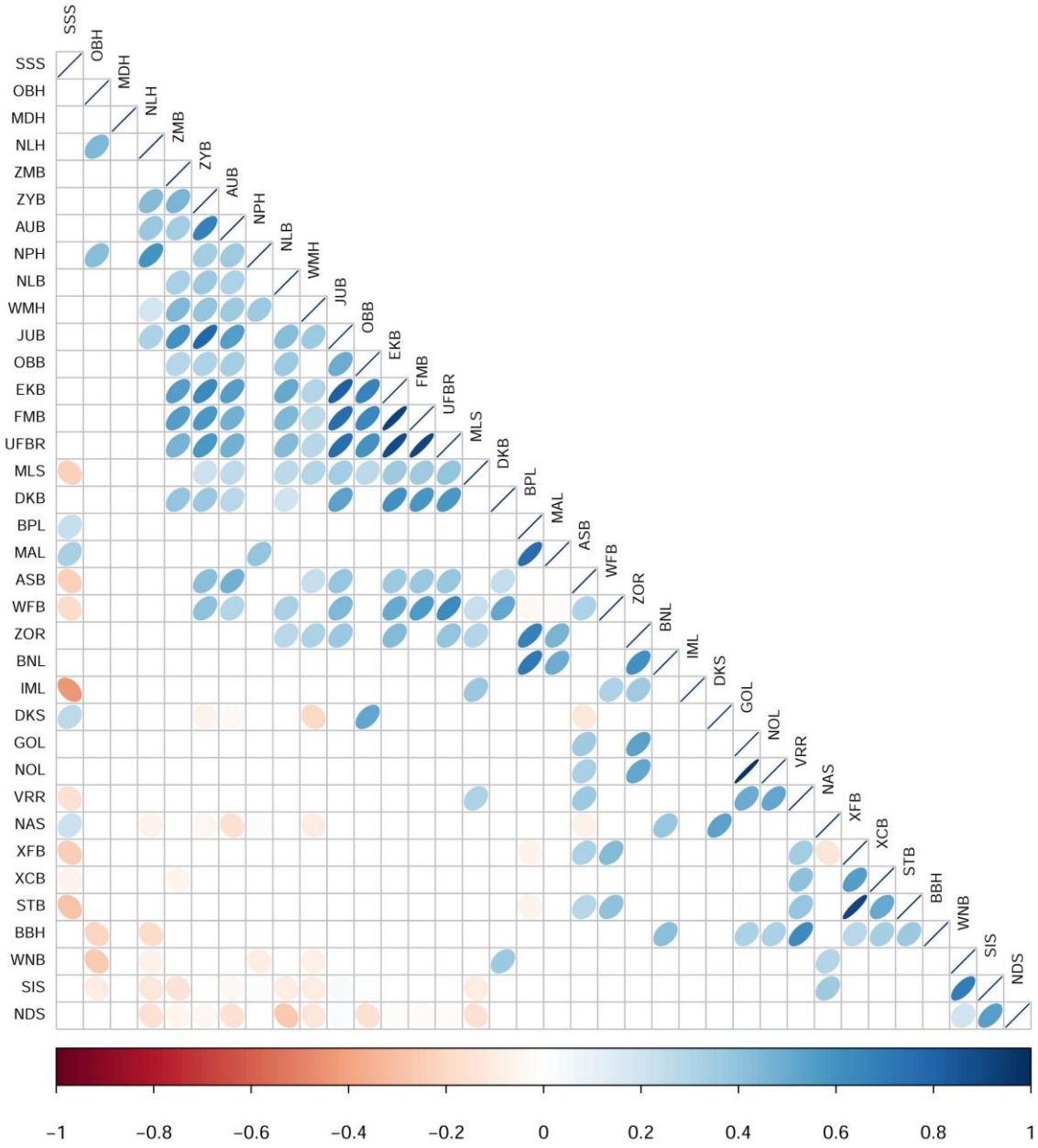


Figure 9. The correlation plot depicting the inter-variable correlation of the cranial measurements. The ellipses that indicate a high correlation between variables are smaller and darker in colour while ellipses that indicate a low correlation between variables are larger and lighter (see Appendix for correlation values).

4.4 ANOVA and Tukey's HSD Test

4.4.1 **ANOVA**

An analysis of variance (ANOVA) was used to ascertain statistically significant differences ($p > 0.05$) among standard cranial landmarks and non-standard interlandmark distances concerning all 8 groups. Five statistically significant standard inter-landmark variables include maximum cranial breadth (XCB), orbital breadth (OBB), interorbital breadth (DKB), occipital chord (OCC) and foramen magnum breadth (FOB). Seventeen non-standard inter-landmark distances were found to be statistically significant and include naso-dacryal subtense (NDS), simotic chord (WNB), simotic subtense (SIS), nasio-frontal subtense (NAS), dacryon subtense (DKS), cheek height (WHM), bistephanic breadth (STB), lambda-opisthion subtense (OCS), lambda-subtense fraction (OCF), (LAR), basion radius (BAR), nasion angle (NAA), nasio-frontal angle (NFA), naso-dacryal angle (NDA), occipital chord (OCC),radio-frontal angle (RFA), radio-parietal angle (RPA) and sub-lambda angle (SLA). Most of the statistically significant standard and inter-landmark distances are associated with mid-facial variables, such as the size of the eyes, shape/size of the nasal complex, as well as length and breadth of the cranium (Table 4).

Table 4. Analysis of variance of standard and non-standard cranial landmarks. Bold indicates statistically significant variables.

Standard Inter-landmark Distances		
Measurement	**F value**	**Pr (> F)**
GOL	1.9503	0.06112
BNL	0.7269	0.647
BBH	1.8152	0.08337
XCB	3.3867	**< 0.01**
XFB	3.0058	0.1958
WFB	1.6958	0.1089
ZYB	0.575	0.7762
AUB	1.8781	0.0722
ASB	1.6449	0.1219
BPL	0.7002	0.6718
NPH	0.5753	0.7751
NLH	1.5125	0.1618
NLB	0.6683	0.699
MAL	1.098	0.3678
MDH	1.6852	0.116
OBH	1.3451	0.2281

OBB	3.5273	**< 0.01**
DKB	2.6185	**<0.05**
OCC	3.7526	**<0.001**
FRC	1.0604	0.3888
FOL	0.3326	0.9389
FOB	2.5496	<**0.05**
EKB	0.2527	0.9711
PAC	1.4796	0.1734
UFBR	0.4353	0.8798
Non-Standard Inter-landmark Distances		
Measurement	**F value**	**Pr (> F)**
NOL	1.7199	0.1032
JUB	0.6472	0.7167
NDS	4.9713	**<0.001**
WNB	2.2659	**<0.05**
SIS	2.2085	**<0.05**
ZMB	1.0247	0.4134
SSS	1.3637	0.2198
FMB	0.2982	0.9543
NAS	3.1149	**< 0.01**
DKS	2.1824	**< 0.05**
IML	1.1262	0.3459
XML	0.5586	0.7893
MLS	1.2886	0.2553
WHM	3.3687	**< 0.01**
STB	2.7236	**<0.001**
FRS	1.5928	0.1364
FRF	0.9902	0.4381
PAS	1.6614	0.1175
PAF	0.5987	0.757
OCS	2.9137	**<0.01**
OCF	6.2565	**<0.001**
NAR	0.8505	0.5462
SSR	0.7888	0.597
PRR	0.8155	0.5758
DKR	1.1881	0.3088
ZOR	0.8965	0.5093
FMR	0.8661	0.5335
EKR	1.1984	0.3029
ZMR	1.051	0.3951
BRR	1.0938	0.3666
VRR	0.7178	0.657
LAR	2.1135	**<0.05**
OSR	1.4086	0.2007
BAR	2.1835	**<0.05**
MOW	1.1514	0.3304
NAA	2.2493	<**0.05**
PRA	1.5366	0.1595
BAA	1.2597	0.2747
NBA	1.1717	0.3183
BBA	1.5138	0.1614

BRA	1.5786	0.1406
SSA	1.3104	0.2443
NFA	2.8725	<**0.01**
DKA	1.9392	0.0628
NDA	3.3913	<**0.01**
SIA	1.4394	0.19
FRA	1.8857	0.0711
PAA	1.7889	0.0885
OCA	3.0229	<**0.01**
RFA	2.5438	< **0.05**
RPA	3.0254	<**0.01**
ROA	1.553	0.1485
BSA	0.6967	0.6747
SBA	1.1413	0.3367
SLA	2.982	<**0.01**
TBA	1.7548	0.0956

4.4.2 **Tukey's HSD Test**

To discover specific differences among the 8 cultural groups, a Tukey's Honest Significant Difference (HSD) post hoc test was conducted on the five standard and eighteen non-standard inter-landmark distances found to be statistically significant in the ANOVA.

The lambda-subtense fraction (OCF) had the most post-hoc pairwise comparisons, and are compiled with Sotho-Ndebele, Swazi-Sotho, Tsonga-Sotho, Tswana-Sotho, Venda-Sotho, Xhosa-Sotho. The most prominent inter-group differences for six variables were from the mid-face (orbit and nose) and cranial vault, particularly the occipital bone, namely orbital breadth (OBB), nasio-frontal subtense (NAS), lambda-subtense fraction (OCF), lambda-opisthion subtense (OCS), nasio-frontal angle (NFA), and occipital angle (OCA) was observed between Swazi and Sotho groups. The naso-dacryal subtense (NDS) had the second most amount of post-hoc pairwise comparisons (five) and it consists of all the OCF, except for the Tswana-Sotho post-hoc pairwise comparison. The OBB variable had four post-hoc pairwise comparisons and like the NDS variable, it had similar post-hoc pairwise comparisons to the OCF variable except for Venda-Sotho and Tsonga-Sotho. The maximum cranial breadth (XCB), nasio-frontal subtense (NAS), cheek height (WMH) and naso-dacryal angle (NDA) variables had three statistically significant post-hoc pairwise comparisons. Seven variables had two post-hoc-pairwise comparisons, namely bistephanic breadth (STB), occipital chord (OCC), basion radius (BAR), nasio-frontal angle (NFA), occipital angle (OCA), radio-parietal angle (RPA) and sub-lambda angle (SLA). The interorbital breadth (DKB), simotic Chord (WNB),

lambda-opisthion subtense (OCS),radio-frontal angle (RFA) had only one post-hoc pairwise comparison. The simotic subtense (SIS) and nasion angle (NAA) did not have any statistically significant inter-group differences (Table 5). Variables located in the splanchnocranium constitute the majority of the interlandmark distances that were found to have statistically significant intergroup differences (Table 5).

Several other post-hoc pairwise comparisons did not have a statistically significant inter-group difference among them. These post-hoc pairwise comparisons were the following: Swazi-Ndebele, Tsonga-Ndebele, Tswana-Ndebele, Venda-Ndebele, Xhosa-Ndebele, Zulu-Ndebele, Tswana-Swazi, Venda-Swazi, Zulu-Swazi, Xhosa-Tswana, Zulu-Tswana and Xhosa-Venda (Table 5).

Table 5. Tukey's HSD Test of the statistically significant variable and the statistically significant group differences observed in each variable. (See figures 10 -20)

Measurement	**Significant Group differences**
XCB	Tsonga-Sotho, Tsonga-Swazi, Xhosa-Tsonga
OBB	Sotho-Ndebele, Tswana-Sotho, Xhosa-Sotho, Swazi-Sotho
DKB	Xhosa-Sotho
NDS	Sotho-Ndebele, Tsonga-Sotho, Venda-Sotho, Xhosa-Sotho, Zulu-Sotho
WNB	Sotho-Ndebele
SIS	None
NAS	Swazi-Sotho, Tswana-Sotho, Tsonga-Swazi
WMH	Venda-Tsonga, Venda-Tswana, Zulu-Venda
STB	Sotho-Ndebele, Swazi-Sotho
OCC	Tswana-Tsonga, Zulu-Tsonga,
OCS	Swazi-Sotho
OCF	Sotho-Ndebele, Swazi-Sotho, Tsonga-Sotho, Tswana-Sotho, Venda-Sotho, Xhosa-Sotho
FOB	Zulu-Xhosa
LAR	Tsonga-Sotho
BAR	Xhosa-Swazi, Zulu-Xhosa

NAA	None
NFA	Swazi-Sotho, Tsonga-Swazi
NDA	Sotho-Ndebele, Tsonga-Sotho, Zulu-Sotho
OCA	Swazi-Sotho, Venda-Sotho
RFA	Tswana-Tsonga
RPA	Zulu-Venda, Zulu-Tsonga
SLA	Zulu-Venda, Zulu-Sotho

Table 6. The different post-hoc pairwise comparisons and the statistically significant variables after a Tukey's HSD Test was conducted. "X" indicates that the variable was found to be statistically significant in that post-hoc comparison.

	Measurement abbreviation																					
	XCB	OBB	DKB	NDS	WNB	SIS	NAS	WMH	STB	OCC	OCS	OCF	FOB	LAR	BAR	NAA	NFA	NDA	OCA	RFA	RPA	SLA
Sot-Nde		X		X	X				X			X						X				
Swa-Nde																						
Tso-Nde																						
Tsw-Nde																						
Ven-Nde																						
Xho-Nde																						
Zul-Nde																						
Ven-Sot				X								X							X			
Swa-Sot		X					X		X		X	X					X		X			
Tso-Sot	X			X								X		X				X				
Tsw-Sot		X					X					X										
Zul-Sot				X														X				X
Xho-Sot		X	X	X								X										
Tso-Swa	X						X										X					
Tsw-Swa																						
Ven-Swa																						
Xho-Swa															X							
Zul-Swa																						
Tsw-Tso										X										X		
Ven-Tso								X														
Xho-Tso	X																					
Zul-Tso										X											X	
Ven-Tsw								X														
Xho-Tsw																						
Zul-Tsw																						
Xho-Ven																						
Zul-Ven								X													X	X
Zul-Xho													X		X							

4.4.2.1 **Boxplots of the statistically significant variables with three or more post-hoc pairwise comparisons (Figure 10-14).**

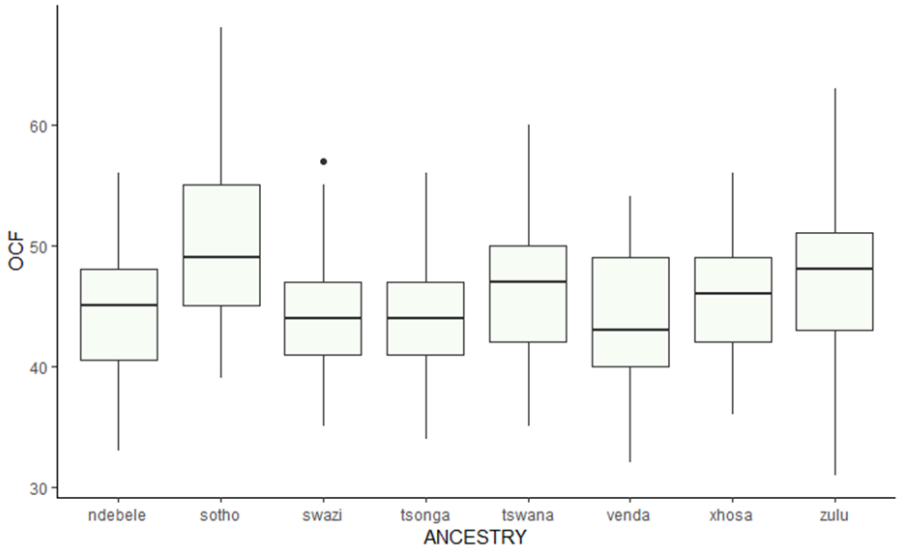


Figure 10. The boxplot of the lambda-subtense fraction (OCF) variable.

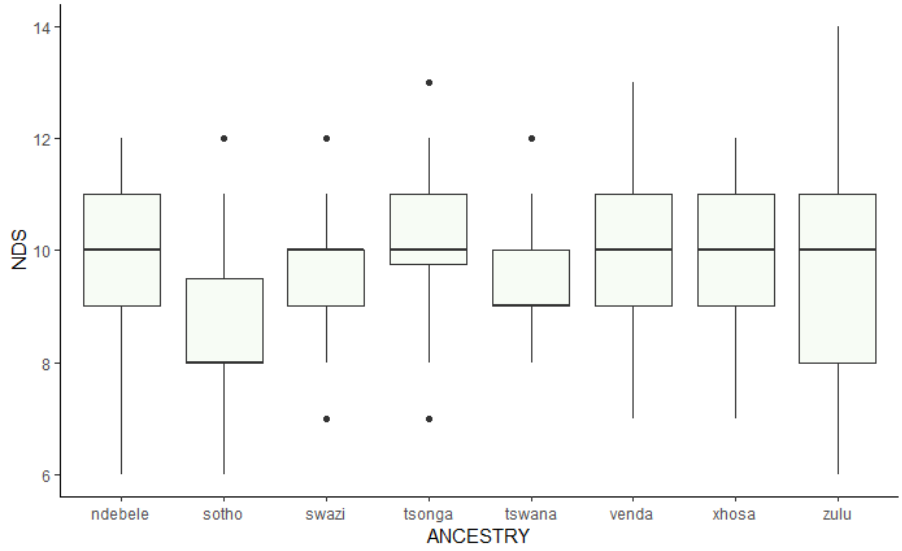


Figure 11. The boxplot of the naso-dacryal subtense (NDS) variable.

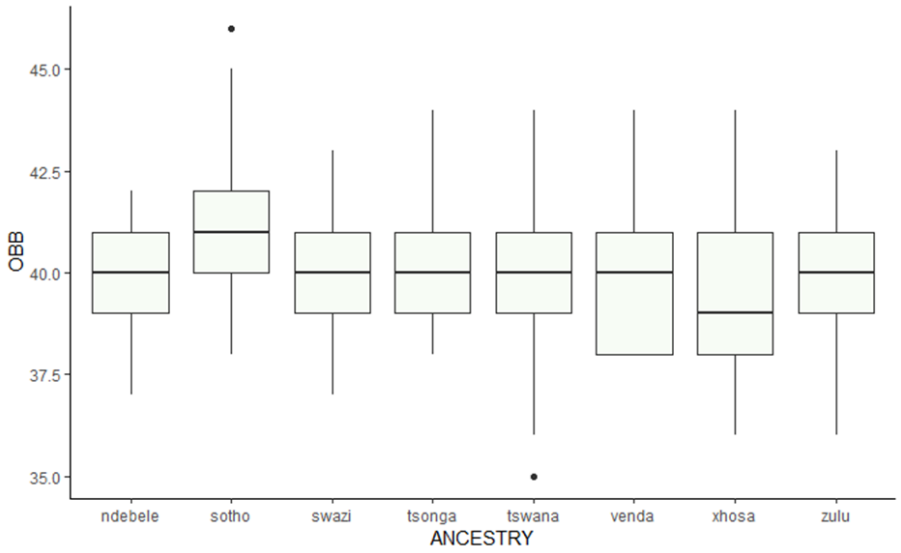


Figure 12. The boxplot of the orbital breadth (OBB) variable.

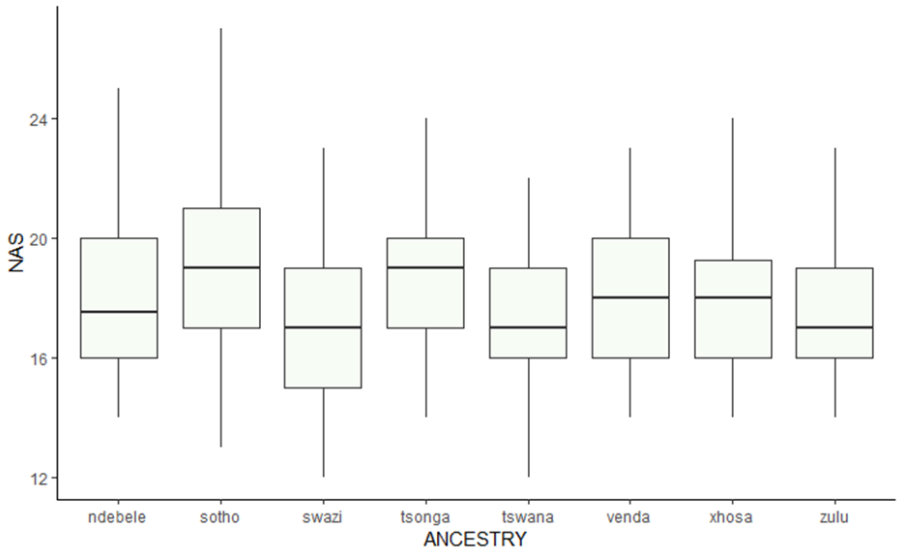


Figure 13. The boxplot of the nasio-frontal subtense (NAS) variable .

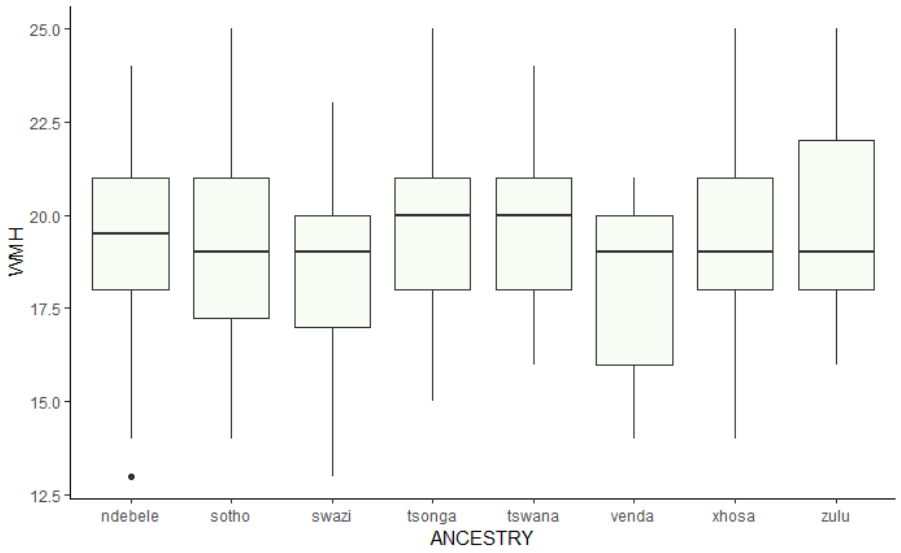


Figure 14. The boxplot of the cheek height (WMH) variable.

4.4.2.2 **The Tukey's plots displaying the significant intergroup from variables that have three or more inter groups differences. Five of the six variables are found in the splanchnocranium. Statistically significant inter-groups differences are displayed in red (Figure 15-20).**

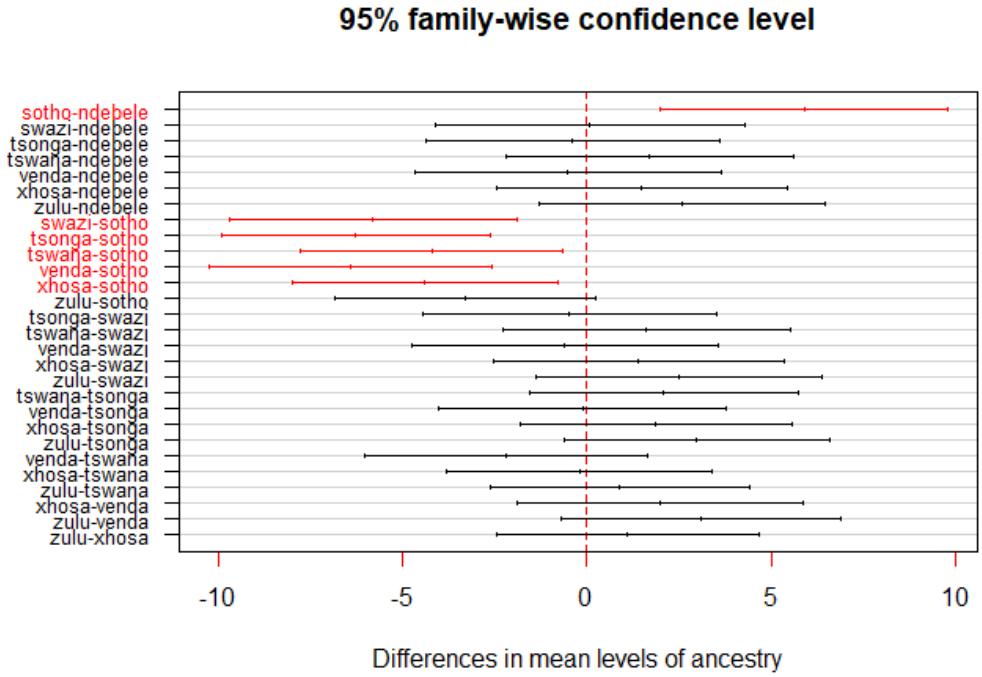


Figure 15. The Tukey's plot of the lambda-subtense fraction (OCF) variable. The statistically significant post-hoc pairwise comparisons are indicated in red.

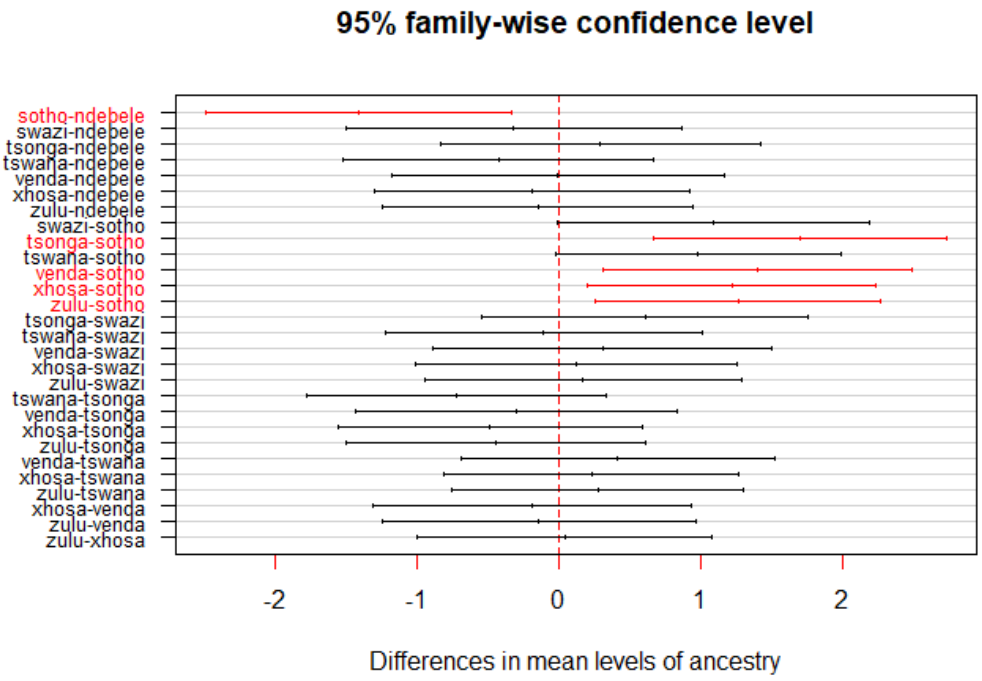


Figure 16. The Tukey's plot of the naso-dacryal subtense (NDS) variable. The statistically significant post-hoc pairwise comparisons are indicated in red.

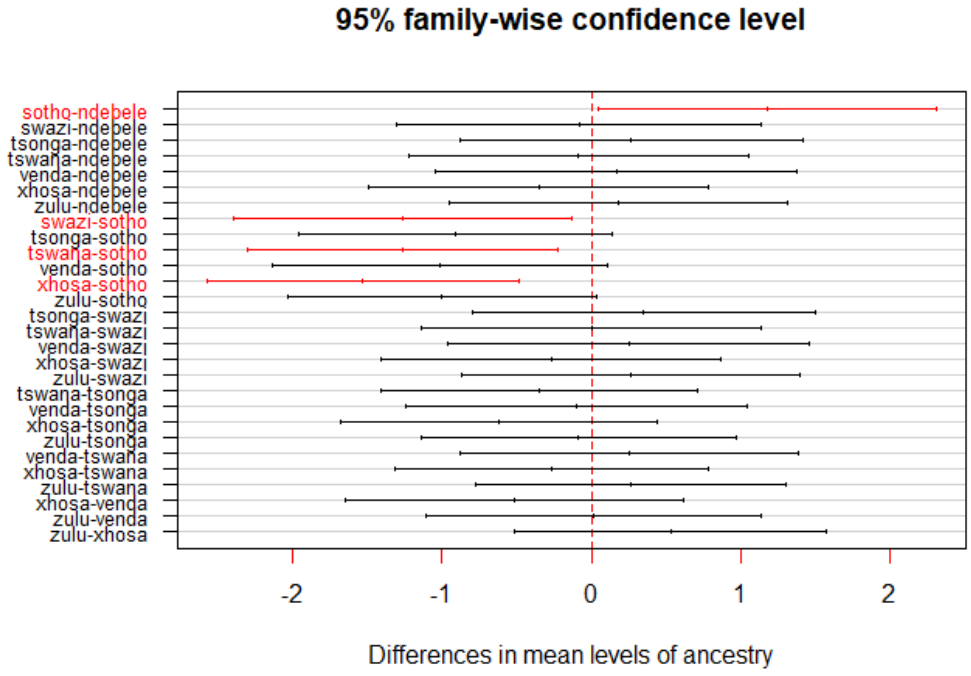


Figure 17. The Tukey's plot of the orbital breadth (OBB) variable. The statistically significant post-hoc pairwise comparisons are indicated in red.

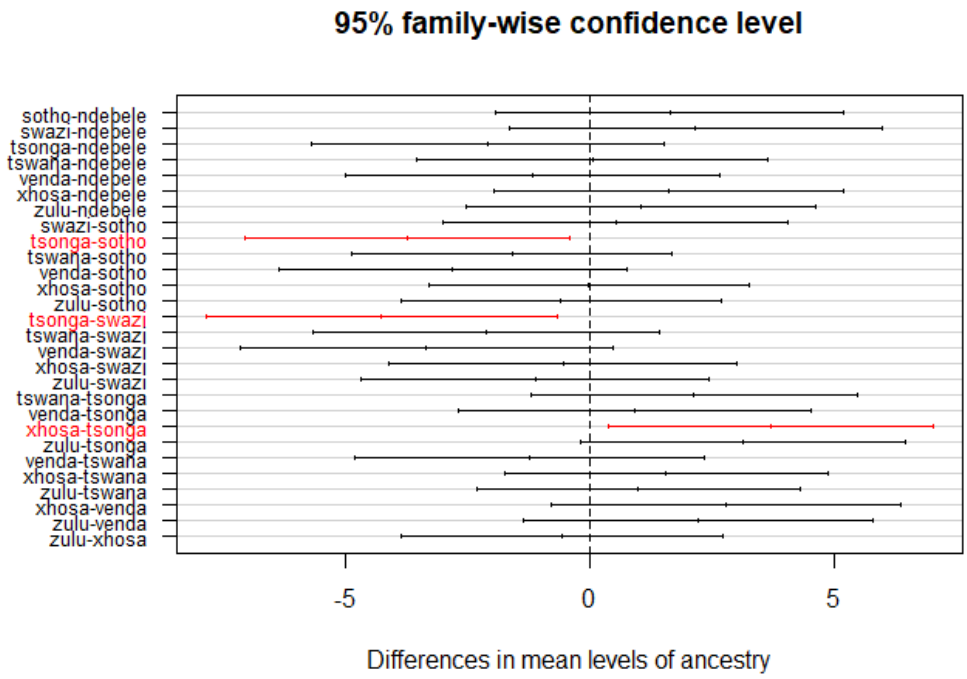


Figure 18. The Tukey's plot of the maximum cranial breadth (XCB) variable. The statistically significant post-hoc pairwise comparisons are indicated in red.

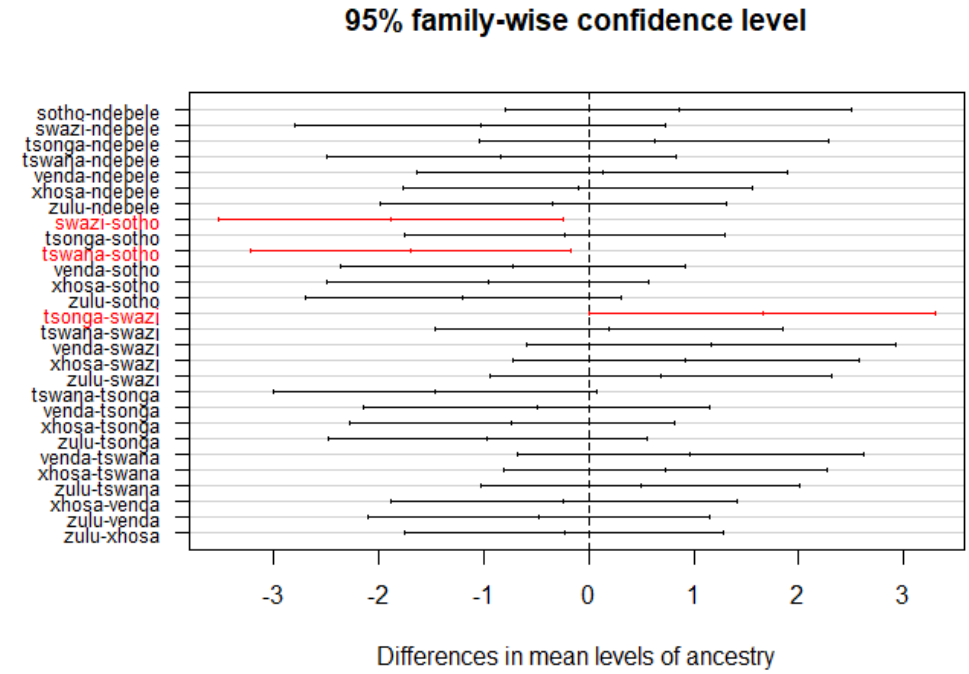


Figure 19. The Tukey's plot of the nasio-frontal subtense(NAS) variable. The statistically significant post-hoc pairwise comparisons are indicated in red.

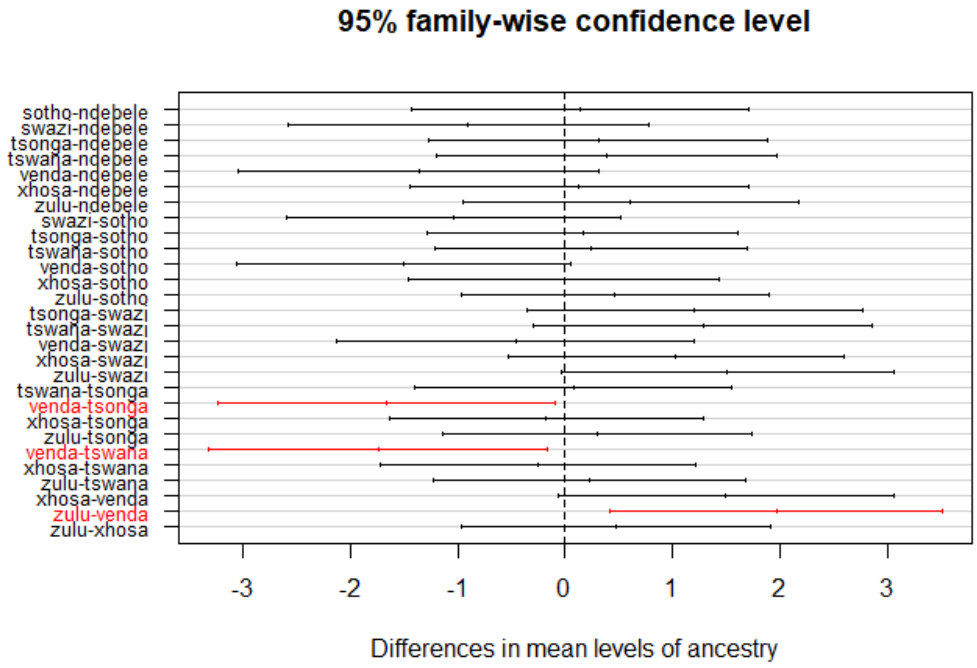


Figure 20. The Tukey's plot of the cheek height (WMH) variable. The statistically significant post-hoc pairwise comparisons are indicated in red.

4.5 Discriminant Function Analysis (DFA)

Standard landmarks and non-standard inter-landmark distances were subdivided into three anatomical regions of the skull, namely the cranial base, cranial vault (neurocranium) and splanchnocranium (mid-face). The three cranial subdivisions were run on all variables, for all groups, using a stepwise selection, as a means to gauge where the size variation among the variables can be used to distinguish among the groups.

Two cultural group subdivisions were created based on geography and linguistics. The first subdivision divides the 8 groups based on geographically adjacent languages and current South African provinces (see Figure 2). A stepwise selection was then conducted on all variables, for all cranial subdivisions, as a means to gauge the effect of geography on cranial variation among black South Africans. The second subdivision divides the 8 cultural groups based on the historical linguistic similarity of all black South Africans (see Figure 3). A stepwise selection was then conducted on all variables, for all cranial subdivisions, as a means to gauge the effect of historical linguistics on cranial variation. Overall, the geographic and linguistic groupings are used to explore various discriminant function analysis models that may better illustrate cranial variation among the 8 groups.

4.5.1 **All groups and All landmarks**

4.5.1.1 **Cranium: Stepwise selection**

This test used a stepwise selection of all the cranial landmarks when analysing all groups individually.

The variables that were removed using stepwise selection were the following.

NAS, IML, RFA, ZMB, NFA, WMH, TBA, MOW, FRF, SSA, ZOR, SBA, PAF, OCA, MLS, BRA, DKR, BSA, WNB, BRR, STB, OCS, XML, DKA, JUB, SSR, SSS, FMB, NBA, FOB, AUB, FRC, BBH, FOL, DKB, OCC, ZYB, ASB, GOL, PAC, BNL, NLH

The variables used in the linear discriminant function.

NOL, XCB, XFB, WFB, NLB, MDH, OBB, EKB, DKS, FRS, NAR, FMR, EKR, VRR, LAR, OSR, BBA, PAA, ROA, SLA

4.5.1.2 **The percentage distribution of the different linear discriminants.**

All of the cranial landmarks were used, and a stepwise selection was performed. The first linear discriminant accounted for 38.58% of the variation observed in this LDA model (Table 7).

Table 7. The proportion of trace of the different linear discriminants

LD1	LD2	LD3	LD4	LD5	LD6	LD7
0.3858	0.2477	0.1604	0.0788	0.0740	0.0362	0.0170

4.5.1.3 **DFA plot of the cranial landmarks after stepwise selection**

The DFA plot displays considerable overlap among the various groups particularly the Ndebele, Tswana, Zulu, Xhosa, Venda, and Tsonga. The Swazi group has some overlap with Venda and Zulu groups. The Swazi are slightly separated from the rest of the other groups. In contrast, the Sotho group is completely separated from the rest of the other group and does not have any overlap with other groups (Figure 21).

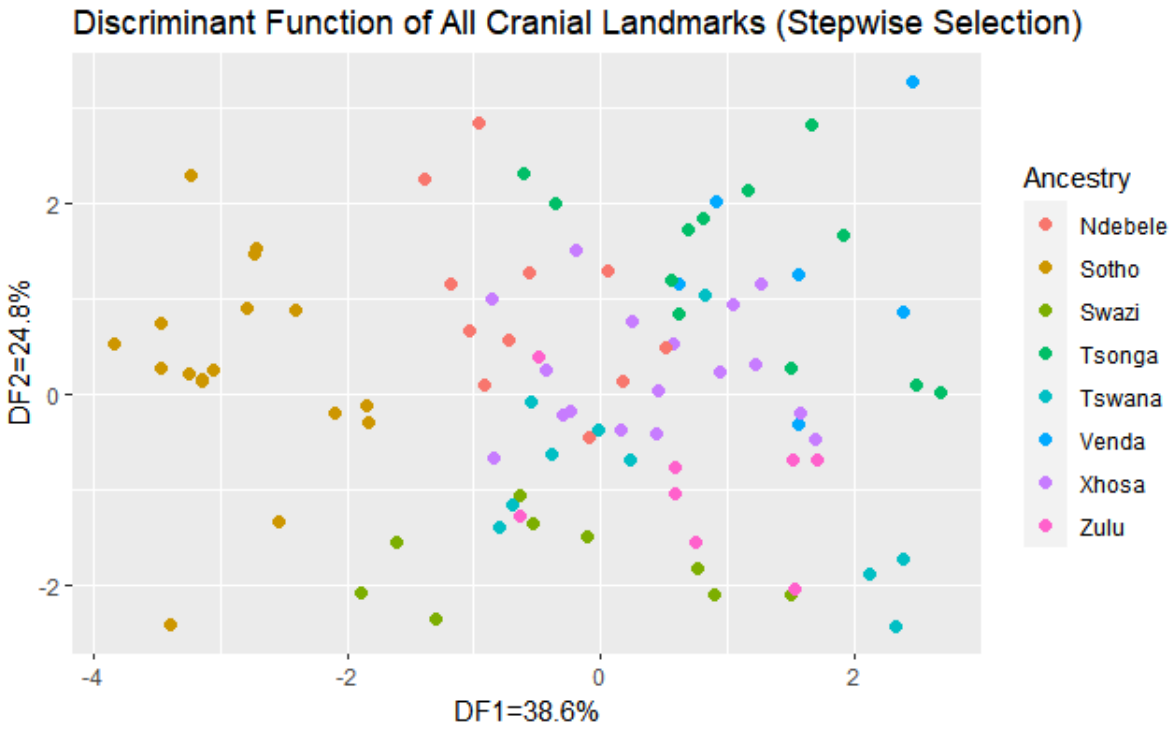


Figure 21. The DF1 vs DF2 plot of all cranial landmarks after stepwise selection

4.5.1.4 **Coefficients of linear discriminant of all variables after stepwise selection**

OBB (orbital breadth) was the variable that accounted for most of the variation observed in the first and fourth linear discriminants. Sub-lambda angle (SLA), dacryon subtense (DKS), vertex radius (VRR), ectoconchion radius (EKR) and frontomalare radius (FMR) were the variables that accounted for most of the variation that was observed in the second, third, fifth, sixth and seventh linear discriminants, respectively. Five out of the six variables mentioned in the previous sentences are non-standard inter-landmark distances. The landmarks that accounted for most of the variation observed in the model are located in the midface of the skull except for the vertex radius that is located on the cranial vault (Table 8).

Table 8. The coefficients of linear discriminant of all variables after stepwise selection

Variable	LD1	LD2	LD3	LD4	LD5	LD6	LD7
NOL	0.354444925	0.04101037	0.11286679	-0.125532996	-0.093847807	0.044223277	0.03091157
XCB	-0.202608249	-0.09310059	-0.05630688	0.058888937	-0.061740610	-0.027172368	-0.01747629
XFB	0.229733952	-0.06149300	-0.02452845	-0.135218785	0.029709429	0.005986434	0.06386363
WFB	-0.190022526	0.14939962	-0.08096695	0.013739217	0.021808895	-0.031221580	0.03876647
NLB	-0.049667605	-0.22626539	-0.04002809	-0.004985617	-0.109926233	0.058473133	0.09091558
MDH	0.124057020	-0.14178075	-0.02994132	0.065003988	0.022400778	-0.108061057	0.09951167
OBB	**-0.665790990**	0.25080008	-0.54943532	**-0.679947809**	0.093474776	0.402526180	0.07542859
EKB	0.238101897	0.03996305	0.23321554	0.288134295	-0.124190314	-0.061689616	-0.12139811
DKS	0.217474241	-0.33832278	**0.58479744**	0.179450672	-0.176394986	0.028339827	0.12501565
FRS	-0.288106938	0.09968140	0.15177204	-0.003976085	0.048580078	-0.122740942	-0.10201693
NAR	-0.211556683	0.10684440	-0.33734942	-0.077640335	0.036151883	-0.224574860	-0.04212105
FMR	0.047724707	0.27231714	0.03589315	0.063575764	0.067368841	-0.208343540	-0.38171546
EKR	-0.108275876	-0.25126061	0.30953498	0.151620629	0.003044612	0.423450642	**0.38507950**
VRR	0.086823384	0.29735997	-0.27244840	0.046064085	0.162512281	**0.528191572**	-0.11717517
LAR	-0.227979487	-0.42300539	0.20138649	0.028402554	**-0.209351709**	-0.209351709	0.15613464
OSR	-0.340913723	0.11003450	-0.33702011	0.087618188	0.087917984	0.146890364	-0.27864150
BBA	-0.180373593	-0.09726274	-0.28281490	-0.015473446	-0.003374609	0.105876190	-0.10051915
PAA	-0.008681463	0.19421066	-0.18607715	-0.048518358	0.035997522	0.305935735	-0.15441411
ROA	-0.039083744	0.01785682	-0.23013815	0.066532709	-0.044791788	0.073623783	-0.12960222
SLA	0.326236886	**-0.45247744**	0.42674958	-0.186136717	-0.063322846	-0.258031510	0.13566291

4.5.1.5 **Classification matrix of the different ancestral groups.**

The kappa value for this model is 30.68 %. The low kappa value indicates that the DFA model does not effectively separate the groups. The overall model classification accuracy is 40%. All groups have a correct classification percentage greater than chance (12.5%). The Sotho group had the highest correct classification percentage at 73.33% with the Tswana and Venda having the lowest correct classification percentage at 20% (Table 9).

Table 9. Classification matrix of the individual ancestral groups

	Number	Ndebele	Sotho	Swazi	Tsonga	Tswana	Venda	Xhosa	Zulu	Correct Classification Percentage (%)
Ndebele	13	5	0	3	3	1	1	0	0	38.46
Sotho	15	3	11	1	0	0	0	0	0	73.33
Swazi	8	1	1	3	0	0	0	0	3	37.5
Tsonga	16	0	1	0	7	2	2	4	0	43.75
Tswana	10	0	0	4	1	2	1	2	0	20
Venda	5	0	0	0	2	0	1	0	2	20
Xhosa	16	3	0	1	2	2	1	5	2	31.25
Zulu	7	0	1	2	0	0	0	2	2	28.57
Total	100	12	14	14	15	7	6	13	6	
Overall model classification accuracy										40

4.5.1.6 **Cranial base: All Groups**

4.5.1.7 **All groups: Coefficients of linear discriminant**

The cranial base refers to the posterior surface of the skull and consists of five standard landmarks, namely: FOB, AUB, ASB, MDH and FOL. No stepwise selection was performed for these variables. The first linear discriminant accounted for 45.64% of the variation that was observed from the DFA model for all groups. The proportion of trace is the percentage of separation achieved by each discriminant function (Table 10). It indicates the amount of the observed variation, as a percentage, of the DFA model that can be attributed to each linear discriminant (Table 10).

Table 10. The proportion of trace of various linear discriminants using cranial base variables

LD1	LD2	LD3	LD4	LD5
0.4564	0.3183	0.1711	0.0443	0.0099

4.5.1.8 **DFA plot of the cranial base using all groups**

The Ndebele group overlaps with four groups namely with the Xhosa, Sotho, Tsonga and Swazi groups. The Venda and Xhosa are most similar to each other for cranial base variables (Figure 22).

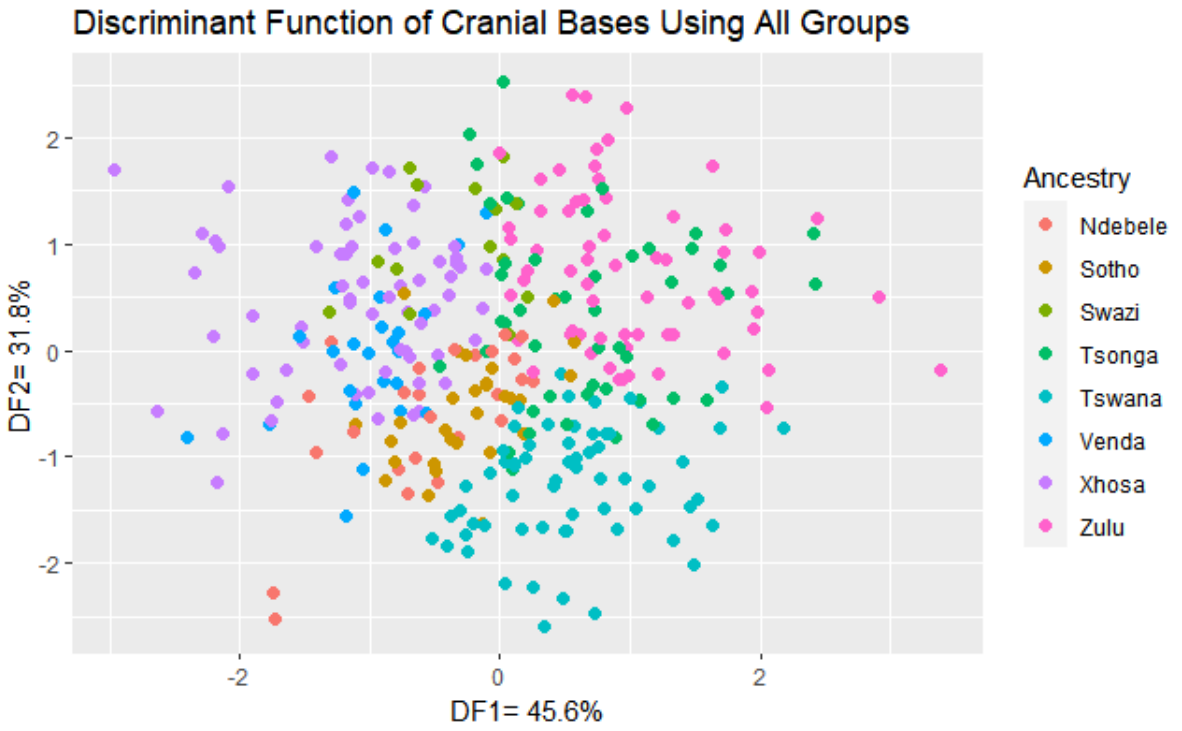


Figure 22. DF1 vs DF2 plot of all groups using cranial bases variables.

4.5.1.9 **The coefficients of linear discriminants for cranial bases variables using all groups**

The bold numbers indicate the variable that accounts for the most variation observed in that specific linear discriminant. FOB (foramen magnum breadth) was the variable that accounts for most of the observed variation in the first linear discriminant. While MDH (mastoid height) accounts for most of the observed variation in the second and third linear discriminants. The fourth and fifth linear discriminants have AUB (auricular breadth) and FOL (foramen magnum length) as the variables that account for the variation observed in those linear discriminants, respectively (Table 11).

Table 11. Coefficients of linear discriminants for cranial bases variables using all groups

Variable	LD1	LD2	LD3	LD4	LD5
FOB	**0.45894649**	0.18733968	0.01381705	0.02589328	-0.175886020
AUB	0.02617507	0.01224452	0.13506718	**-0.18550385**	0.092816731
ASB	0.06031380	-0.18184564	-0.14590372	0.02516391	-0.004573749
MDH	-0.03618249	**0.21548015**	**-0.24329321**	-0.06334367	0.065389300
FOL	-0.24999806	0.02590538	-0.05248961	-0.17614247	**-0.279337368**

Table 12. The classification matrix of the various groups using cranial base variables

	Number	Ndebele	Sotho	Swazi	Tsonga	Tswana	Venda	Xhosa	Zulu	Correct Classification Percentage (%)
Ndebele	36	1	2	3	5	9	2	10	4	2.78
Sotho	47	3	4	0	7	10	5	11	7	8.51
Swazi	35	6	1	2	4	5	4	6	7	5.71
Tsonga	47	4	5	1	9	8	2	7	11	19.15
Tswana	45	4	5	3	5	15	3	4	6	33.33
Venda	31	4	7	1	3	2	2	7	5	6.45
Xhosa	45	3	5	3	5	5	7	10	7	22.22
Zulu	49	2	2	2	7	8	2	7	19	38.78
Total	335	27	31	15	45	62	27	62	66	
Overall model classification accuracy										18.51

4.5.1.10 **Cranial base: Classification matrix of the different groups**

The kappa value relating to the model classification accuracy is 6.2%. The overall model classification accuracy was 18.51% and the very low kappa value indicates that the model performed poorly for group separation using the cranial base variables. Four out of the eight groups had correct classification percentages less than chance (12.5%), namely the Ndebele,

Sotho, Swazi and Venda. The Venda had the lowest correct classification at 2.78% and the Zulu group having the highest correct classification at 38.78% (Table 12).

4.5.1.11 **Splanchnocranium: All Groups**

The Splanchnocranium stepwise selection of all the splanchnocranium landmarks when analysing all groups individually.

The variables that were removed using stepwise selection were the following.

NLB, IML, ZOR, SIS, OBH, ZMR, OBB, FMB, ZMB, NFA, XML, MOW, ZYB, SSA, UFBR, WFB, WNB, DKR, DKB, DKS, EKR, NAR, NLH

The variables used in the linear discriminant function for the facial skeleton included:

XFB, EKB, JUB, NDS, SSS, NAS, MLS, WMH, SSR, FMR

4.5.1.12 **The percentage distribution of the different linear discriminants for the splanchnocranium using all groups.**

After the stepwise selection was conducted on the splanchnocranium landmarks, the first linear discriminant accounted for 60.03% of the variation that was seen in this specific LDA model (Table 13).

Table 13. The proportion of trace of the different linear discriminants

LD1	LD2	LD3	LD4	LD5	LD6	LD7
0.6003	0.1413	0.1270	0.0633	0.0414	0.0177	0.0090

4.5.1.13 **DFA plot of the splanchnocranium stepwise selection using all groups**

The majority of the groups are overlapping with one another except for the Swazi and Sotho. These two groups do overlap with some of the groups but are noticeable on the periphery of all the groups (Figure 23).

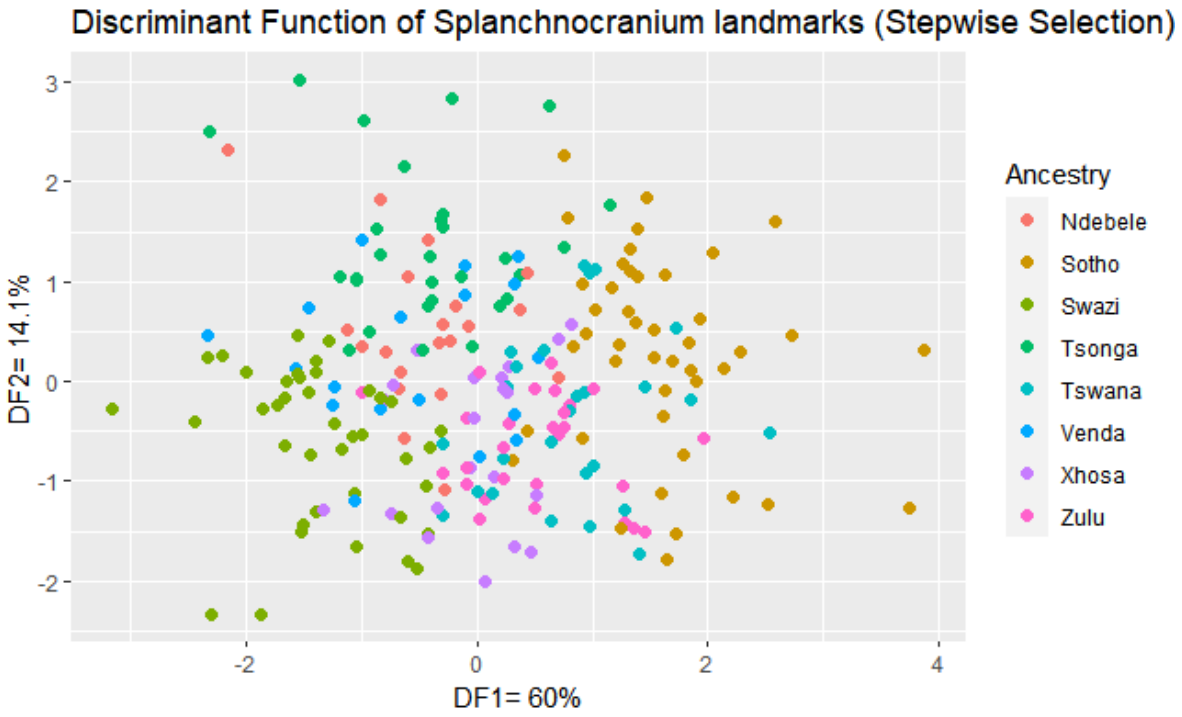


Figure 23. DF1 vs DF2 of the Splanchnocranium variables after a stepwise selection was conducted.

4.5.1.14 **Splanchnocranium Stepwise: Coefficients of linear discriminant using all groups**

The variable that accounted for most of the variation observed in the first and fourth linear discriminant is NDS (Naso-Dacryal Subtense). The MLS (Malar Subtense) variable contributed the most towards the variation that was observed in the second and fifth linear discriminant. While the WMH (Cheek Height) and SSS (Zygomaxillary Subtense) accounted for most of the variation observed in the third and sixth linear discriminant, respectively. All of these variables that accounted for most of the variation observed in the different linear discriminants are nonstandard inter-landmark distances and indicate variation on the mid-facial, particularly the maxilla (Table 14).

Table 14. The coefficients of the linear discriminants of splanchnocranium variables

Variable	LD1	LD2	LD3	LD4	LD5	LD6	LD7
XFB	-0.08636685	-0.07324867	-0.01770045	0.093277262	0.02286452	-0.007869982	0.130106718
EKB	0.15656267	0.02520625	0.03063880	0.103228136	-0.13224951	0.009290910	0.007887866
JUB	-0.16646487	0.05479749	-0.03320793	0.007175691	0.01872362	0.018739079	-0.173816011
NDS	**-0.29856053**	0.11722222	-0.26328046	**-0.311127779**	-0.05124794	0.200027957	0.157230522
SSS	0.22362112	-0.07103500	-0.02060345	0.048231552	-0.02191336	**0.339495908**	0.024860818
NAS	0.11077523	0.14713835	0.19663984	0.096718933	0.22301981	-0.142174359	**0.196466960**
MLS	0.15704664	**-0.41871156**	0.19850065	-0.207899424	**-0.25852417**	-0.035234205	0.118813774
WMH	0.28112338	0.10762088	**-0.34499564**	0.039181747	-0.09684154	-0.092070793	0.137402825
SSR	-0.17641171	0.07314931	0.06902996	-0.137888887	-0.14554149	-0.104237803	0.046692654
FMR	0.18544424	-0.07104845	-0.04130791	-0.059364905	0.23458745	-0.014880691	-0.076041864

4.5.1.15 **Splanchnocranium Stepwise: Classification matrix of the different groups.**

The kappa value that is associated with the overall model classification accuracy is 9.84%. This low kappa value indicates that the model performed poorly in separating the different groups. The overall model classification accuracy is 21.33% when splanchnocranium variables were used. Only the Ndebele and Tswana had correct classification percentages, 8.00 and 6.06 respectively, that were less than chance (12.55%). The Sotho and Swazi had significantly higher correct classification, 44.12% and 41.67% respectively (Table 15).

Table 15. Classification matrix of the splanchnocranium stepwise for the various groups

	Number	Ndebele	Sotho	Swazi	Tsonga	Tswana	Venda	Xhosa	Zulu	Correct Classification Percentage (%)
Ndebele	25	2	1	6	6	3	1	2	4	8
Sotho	33	1	15	1	3	7	3	1	3	44.12
Swazi	24	1	2	10	3	4	2	1	1	41.67
Tsonga	31	6	5	4	7	2	2	2	3	22.58
Tswana	33	7	8	5	1	2	0	4	6	6.06
Venda	17	2	2	5	3	1	2	2	2	10.53
Xhosa	28	2	4	7	1	1	4	4	5	14.29
Zulu	31	3	6	4	1	3	2	6	6	19.35
Total	222	17	43	42	22	23	16	22	30	
Overall model classification accuracy										21.33

4.5.1.16 **Cranial vault: Stepwise selection**

Stepwise selection of all the cranial landmarks when analysing all groups individually.

The variables that were removed using stepwise selection were the following.

NBA, PAF, SLA, RFA, FRF, NOL, OCS, PAA, OCF, OCA, GOL, FRC, BBH, OCC, PAC

The variables used in the linear discriminant function.

BNL, XCB, STB, FRS, PAS, BBA, RPA, ROA, SBA, TBA

4.5.1.17 **Cranial vault stepwise: The percentage of the different linear discriminants when using all groups**

The first linear discriminant accounted for 38.10% of the variation that was observed in the DFA model when stepwise selection was used for cranial vault variables (Table 16).

Table 16. The proportion of trace of linear discriminants of the cranial vault stepwise

LD1	LD2	LD3	LD4	LD5	LD6	LD7
0.3810	0.2290	0.2152	0.0778	0.0642	0.0238	0.0089

4.5.1.18 **DFA plot of the cranial vault landmarks after a stepwise selection**

Most of the groups are overlapping with one another with regard to the cranial vault variables. The Ndebele group appears to have considerable overlap with most groups when it relates to cranial vault variables. The Sotho, Zulu and Tsonga groups are on the periphery of the cluster relating to other groups (Figure 24).

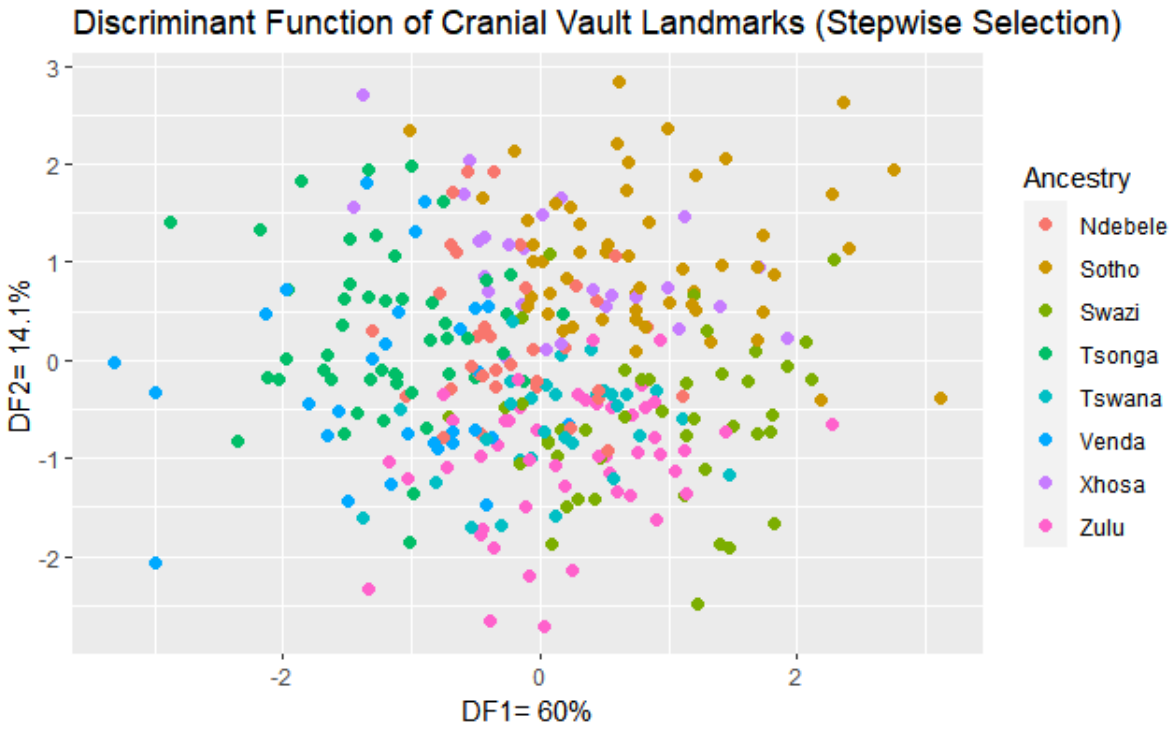


Figure 24. DF1 vs DF2 plot of cranial vault landmarks using all groups

4.5.1.19 **Cranial vault stepwise: Coefficients of linear discriminant when using all groups**

The FRS (Nasion-Bregma Subtense) variable accounted for most of the variation that was observed in the second, fourth and seventh linear discriminants. The XCB (Maximum cranial breadth), PAS (Bregma-Lambda Subtense), BBA (Basion Angle) and RPA (Radio-Parietal Angle) variables accounted for most of the variation that is observed in the first, fifth, sixth and third linear discriminants, respectively. Most of the observed variation is from non-standard inter-landmark distances (Table 17).

Table 17. The coefficients of linear discriminants of cranial vault variables

Variable	LD1	LD2	LD3	LD4	LD5	LD6	LD7
BNL	-0.08136644	0.008810194	0.094270816	-1.045515e-01	-0.01843308	0.149962630	-0.088419833
XCB	**0.15026090**	0.033770936	0.044664847	5.594366e-02	-0.00662800	-0.088614274	-0.096544975
STB	-0.03340186	-0.117562562	-0.159348454	-9.167636e-05	-0.05992630	0.024662426	-0.007404713
FRS	0.13350497	**0.271065257**	-0.032553947	**-1.827870e-01**	0.21462955	-0.053910432	**0.161307447**
PAS	0.01869481	0.176950447	-0.001895497	-1.113179e-01	**-0.32899834**	0.041406189	0.108438371
BBA	0.03845655	-0.173851075	-0.101090753	1.286629e-01	-0.05914359	**0.323005018**	-0.028247609
RPA	-0.13737884	0.023513968	**-0.162677035**	8.420592e-02	0.14330550	-0.123257502	-0.040069670
ROA	-0.05462126	-0.050161398	0.038581862	8.476323e-02	-0.05394080	-0.017529861	0.125077214
SBA	0.11222963	0.153810150	-0.139344926	-4.592588e-03	-0.09009838	-0.022054224	0.123392214
TBA	-0.08420163	0.118890363	0.044400269	9.310093e-02	-0.01782820	0.003339628	0.000866436

4.5.1.20 **Cranial vault stepwise: Classification matrix of the different ancestral groups**

The overall model classification accuracy is 19.94% after using stepwise selection. The kappa value that is associated with the overall model classification accuracy is 8.44%. The low kappa value alludes to the fact that the model performed poorly in separating the groups when cranial vault variables were used. Four groups had a correct percentage that was less than chance, namely Xhosa, Venda, Tswana and Ndebele, Similar to the previous section the Sotho and Swazi groups obtained higher correct classification, 35.55% and 40.63% respectively than other groups (Table 18).

Table 18. The classification matrix of different groups using cranial vault variables

	Number	Ndebele	Sotho	Swazi	Tsonga	Tswana	Venda	Xhosa	Zulu	Correct Classification Percentage (%)
Ndebele	34	1	4	5	7	3	4	6	4	2.94
Sotho	45	2	16	7	4	1	4	6	5	35.55
Swazi	32	4	0	13	3	6	0	1	5	40.63
Tsonga	41	5	5	2	10	3	10	4	2	24.39
Tswana	44	6	6	7	5	5	5	0	9	11.63
Venda	31	5	3	3	6	4	3	2	5	9.68
Xhosa	45	7	10	3	8	4	3	4	6	8.89
Zulu	34	1	6	9	7	4	3	4	11	24.44
Total	304	31	50	49	50	30	32	27	47	
Overall model classification accuracy										19.94

4.6 Geographical Location

The general overlapping of these groups may mask differences among the groups which can be further examined with clustering of groups based on geography. The different geographically adjacent groups were clustered in the following combinations Sotho-Tswana(sot_tsw), Zulu-Xhosa (zul_xho), Swazi-Ndebele (swa_nde) and Venda-Tsonga (ven_tso) (see Figure 2). Historical linguistic groupings include nine official languages predominately associated with black South Africans. These groups were chosen as they are still reflective of modern-day clusters in the predominantly spoken languages by black South Africans from geographically adjacent provinces of South Africa (see Figure 2).

4.6.1 **Geography: All cranial variables and all groups**

Stepwise selection of all the cranial landmarks for the different variables based on geography.

The variables that were removed by means of stepwise selection were the following.

RFA, SLA, SSS, WNB, OCS, NOL, ZMR, FMR, BRR, OCA, PAF, MOW, OSR, DKA, JUB, EKR, VRR, DKS, NFA, WFB, WMH, NAS, SIS, STB, XML, DKR, BRA, FRF, FRS, LAR, FRA, SSR, PAS, ZOR, NAR, AUB, BBH, XFB, NLB, EKB, ZYB, OCC, FRC, OBH, NLH, PAC, XCB, GOL, ASB, FOB, OBB

The variables used in the linear discriminant function.

BNL, MDH, DKB, NDS, ZMB, FMB, IML, MLS, OCF, FOL, BAR, UFBR, NBA, BBA, SSA, PAA, RPA, ROA, SBA, TBA

4.6.1.1 **The percentage distribution of the different linear discriminants**

The first linear discriminant for the stepwise selection of geographically adjacent group accounted for 52.48% (Table 19).

Table 19. Proportion of trace for the various linear discriminant

LD1	LD2	LD3
0.5248	0.2844	0.1909

4.6.1.2 **DFA plot of using cranial landmarks after stepwise selection based on geography**

There is a complete overlap between Swazi-Ndebele and Zulu-Xhosa cluster groups. The Sotho-Tswana cluster has some overlap with the Zulu-Xhosa cluster and does not overlap with other clusters. The Venda-Tsonga appears to be a distinct cluster on its own, but it is still near the other clusters (Figure 25).

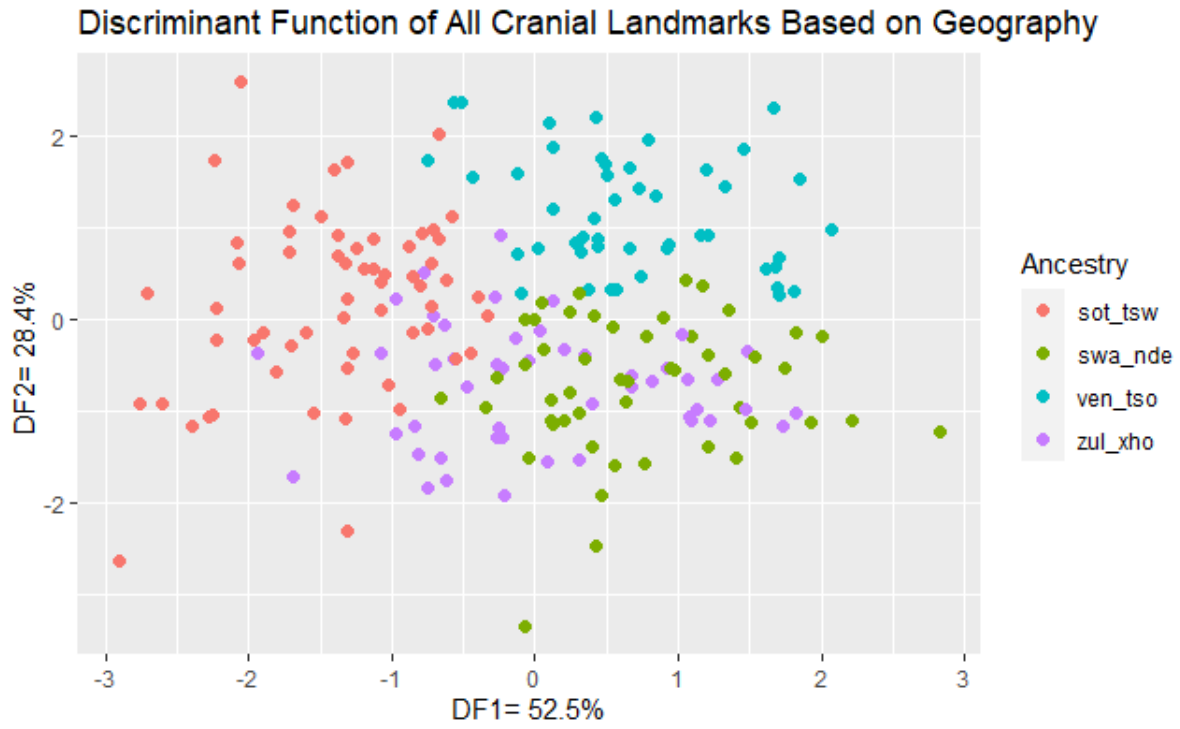


Figure 25. DF1 vs DF2 plot of all cranial landmarks based on geography after a stepwise selection was conducted.

4.6.1.3 **Coefficients of linear discriminants of the different variables based on geography after stepwise selection.**

The variable that accounts for most of the variation in the first linear discriminant is NDS (Naso-Dacryal Subtense), while MDH (Mastoid height) and BAR (Basion Radius) accounts for most of the variation in the second and third linear discriminants, respectively. Two of the three variables which account for most of the observed variation are nonstandard inter-landmark distances (Table 20).

Table 20. Coefficients of linear discriminants of the different variables based on geography

Variable	LD1	LD2	LD3
BNL	-0.06793956	0.067532950	0.134953840
MDH	0.03978039	-0.241807341	0.099240846
DKB	0.07255623	-0.126415555	-0.008637706
NDS	**0.43624969**	-0.003567702	0.137746874
ZMB	-0.02932187	0.097857918	-0.113507989
FMB	-0.14218128	**0.236123283**	0.222844185
IML	0.09835276	0.152876863	-0.046840842
MLS	-0.13622725	-0.172652781	0.191702325
OCF	-0.08562750	0.014054814	0.004763241

FOL	0.09499646	-0.191029075	0.017445218
BAR	-0.02988235	0.106997736	**0.400751962**
UFBR	0.08714335	-0.201282517	-0.341373425
NBA	-0.02603677	-0.161972592	0.104246807
BBA	-0.12259996	-0.185667581	-0.048086419
SSA	0.04302053	-0.041239974	0.007194667
PAA	-0.07610783	0.036838878	-0.091180422
RPA	0.03823640	0.010317037	-0.064501699
ROA	0.04234117	0.04234117	0.037637819
SBA	0.01238772	-0.170912033	0.053792220
TBA	-0.03695991	0.101221485	0.303667015

4.6.1.4 **Classification matrix of the various based to their geographical location**

All group clusters of the geographically adjacent language groups obtained correct classification accuracies that were greater than chance (25%). The Venda and Tsonga group obtained the highest correct classification (60%).The overall model classification for this model is 45.63%. (Table 21).

Table 21.Classification matrix of the various based to their geographical location

	Number	sot_tsw	swa_nde	ven_tso	zul_xho	Correct Classification Percentage (%)
sot_tsw	58	31	5	9	13	53.45
swa_nde	46	7	17	11	11	36.96
ven_tso	45	3	9	27	6	60.00
zul_xho	57	14	15	9	19	33.33
Total	206	55	46	56	49	
Overall model classification accuracy						45.63

4.6.1.5 **The percentage of distribution of the linear discriminants of the cranial base based on geography.**

The cranial base landmarks consist of five variables, namely: FOB, AUB, ASB, MDH and FOL. The first discriminant accounted for 53.9% of the variation that was observed from the DFA model of the geographically adjacent groups when cranial base landmarks were used (Table 22).

Table 22. Cranial base: The percentage distribution of the different linear discriminants.

Proportion of trace		
LD1	LD2	LD3
0.5390	0.3233	0.1376

4.6.1.6 **Cranial base separated by geography: Coefficients of linear discriminant**

Mastoid Height (MDH), Bi-Asterionic Breadth (AUB) and Foramen Magnum Breadth (FOB) were the variables that contributed to the amount of variation that was observed in the LD1, LD2 and LD3 respectively (Table 23).

Table 23. Cranial base separated by geography: Coefficients of linear discriminant

Variable	LD1	LD2	LD3
FOB	-0.07069654	0.14088505	**-0.300444211**
AUB	0.08653863	0.09489509	-0.142564430
ASB	0.11099694	**-0.19190388**	-0.005885765
MDH	**-0.29391879**	-0.13852833	-0.087475410
FOL	-0.06248056	-0.11773064	0.101508894

4.6.1.7 **Cranial base: Classification matrix of the different groups based on geography**

Three out of the four group clusters of the geographically adjacent language groups obtained correct classification accuracies that were greater than chance (25%). The Swazi and Ndebele group cluster did not obtain a correct classification that was greater than chance with regard to cranial base landmarks. The Kappa value relating to the overall model classification accuracy of this model is 3.76% (Table 24).

Table 24. Cranial base: Classification matrix of the different groups based on geography

Classification matrix						
	Number	sot_tsw	swa_nde	ven_tso	zul_xho	Correct Classification Percentage (%)
sot_tsw	92	34	13	21	24	36.96
swa_nde	71	19	12	16	24	16.90
ven_tso	78	22	17	18	21	23.08
zul_xho	94	24	20	19	31	32.98
Total	335	99	62	74	100	
Overall model classification accuracy						28.36

4.6.1.8 **Cranial vault: The percentage distribution of the different linear discriminants.**

The first linear discriminant accounted for 48.55% of the variation that was observed in the DFA model for geographical location when cranial vault variables were used (Table 25).

Table 25. Cranial vault: The percentage distribution of the different linear discriminants.

Proportion of trace		
LD1	LD2	LD3
0.4855	0.3711	0.1434

4.6.1.9 **Cranial vault: Coefficients of linear discriminant**

The variables that contributed to most of the variation that was observed in the model for the three linear discriminants were BBA (Basion Angle) for the first linear discriminant, while for the second and third linear discriminant is PAS (Bregma-Lambda Subtense). The variables that accounted for most of the variation that is observed in the different linear discriminants (Table 26).

Table 26. Cranial vault: Coefficients of linear discriminant

Variable	LD1	LD2	LD3
GOL	0.067567467	-0.1126500795	0.269498508
BNL	0.242760954	-0.1881573180	-0.232053844
BBH	0.199025131	-0.4378799518	-0.158610427
XCB	-0.091435139	-0.0934950310	-0.0934950310
OCC	-0.375482885	-0.0279791499	-0.054419776
FRC	-0.679887667	0.8164511963	0.095416061
PAC	0.307982826	-0.2694337714	0.647426016
NOL	0.014332106	0.0574751148	-0.269597097
STB	0.139114628	-0.0319690439	0.054954417
FRF	-0.037214347	-0.0000933682	-0.003043655
FRS	-0.004295279	-0.0273172244	-0.299686158
PAS	-0.333896744	**1.4547677242**	**-0.843546789**
PAF	-0.050290001	0.0841444510	-0.012278349
OCS	0.392896843	-0.1894125238	-0.156452109
OCF	-0.065282871	0.0063011410	-0.026641671
NBA	0.362215822	-0.2978038226	-0.471220134
BBA	**1.265828035**	-1.1781103686	-0.029671989
PAA	-0.264210771	0.8680344448	-0.561101057
OCA	0.251755931	-0.1459262005	-0.055487134
RFA	0.114402533	-0.5273718920	-0.336162778
RPA	-0.365655542	-0.1331017104	-0.838678472
ROA	0.170975630	0.0642286099	0.148962556
SBA	0.046920978	-0.1565216768	-0.072902938
SLA	-0.012993650	-0.1082393074	0.075256066
TBA	-0.043952244	-0.0780352841	-0.225679999

4.6.1.10 **Cranial vault: Classification matrix of the different groups based on geography**

Three of the four groups had correct classification accuracies that were greater than chance. The Zulu and Xhosa group cluster was the only group cluster that had a correct classification accuracy that was less than chance (25%) with regard to cranial vault landmarks. The Kappa value that is associated with the overall model classification accuracy is 12.22%. The current DFA model was not effective in differentiating the various groups adequately about the geographical location as indicated by the low kappa value (12.22%) (Table 27).

Table 27. Cranial vault: Classification matrix of the different groups based on geography

Classification matrix						
	Number	sot_tsw	swa_nde	ven_tso	zul_xho	Correct Classification Percentage (%)
sot_tsw	72	28	11	14	19	38.89
swa_nde	56	6	23	16	11	41.07
ven_tso	59	13	10	22	14	37.29
zul_xho	83	22	20	22	19	22.89
Total	270	69	64	74	63	
Overall model classification accuracy						34.07

4.6.1.11 **Splanchnocranium: The percentage distribution of the different linear discriminants**

The first linear discriminant accounted for 53.36% variation that was observed in the model (Table 28).

Table 28. Splanchnocranium: The percentage distribution of the different linear discriminants.

Proportion of trace		
LD1	LD2	LD3
0.5336	0.3225	0.1440

4.6.1.12 **Splanchnocranium: Classification matrix of the different groups based on geography**

Most of the group clusters had correct classification accuracy that were greater than chance. The Swazi and Ndebele group did have a correct classification accuracy greater than chance (25%). The Kappa value associated with the overall model classification accuracy is 14.17%. The low kappa value that was obtained by this model indicates that the models were unable to distinguish individuals from the various adequately (Table 29).

Table 29. Splanchnocranium: Classification matrix of the different groups based on geography

Classification matrix						
	Number	sot_tsw	swa_nde	ven_tso	zul_xho	Correct Classification Percentage (%)
sot_tsw	57	25	4	16	12	43.86
swa_nde	43	7	10	11	15	23.26
ven_tso	41	8	12	17	4	41.46
zul_xho	49	13	14	6	16	32.65
	190	53	40	50	47	
Overall model classification accuracy						35.79

4.7 Linguistics separation of groups

Apart from grouping the different groups based on their geographical location. The groups were clustered based on their historical linguistic lineages. These groups were clustered into three main groups namely, Nguni, Sotho-Tswana and Venda-Tsonga (See figure 3). The Nguni clustered comprised of the Zulu, Xhosa, Swati and Ndebele groups.

4.7.1 **Stepwise linguistics**

Variables that were removed in the stepwise selection:

NLH, OSR, WNB, BNL, ZOR, SSS, SSR, JUB, OCF, NOL, BAR, FMR. ASB, IML. EKR, ZYB, DKA, NAR, NAS, OCC, BRA, WFB, ZMB, STB, NBA, OCA, NFA, VRR, BBH, AUB, RPA, FMB, EKB, RFA, FRC, NLB, SLA, FOL, ROA, DKS, WMH, PAF, SSA, SBA, OCS, MLS, OBB, BBA, DKR, FRF, FRS, FOB, BRR, LAR, MDH, TBA, PAA, ZMR

Variables in the model are the following:

GOL, XCB, OBH, DKB, XML, PAC, PAS, MOW, UFBR, FRA

4.7.1.1 **Stepwise linguistics: The percentage distribution of the different linear discriminants using all variables**

The first discriminant function is accounted for 73.76% of the variation that was observed in the model (Table 30).

Table 30. Stepwise linguistics: The percentage distribution of the different linear discriminants

Proportion of trace	
LD1	LD2
0.7376	0.2624

4.7.1.2 **Stepwise linguistics: Coefficients of linear discriminant using all variables**

The variables that accounted for the most variation observed in the first and second linear discriminants were orbital height (OBH) and interorbital breadth (DKB) respectively (Table 31).

Table 31. Stepwise linguistics: Coefficients of linear discriminant

Variable	LD1	LD2
GOL	-0.07869327	0.09090915
XCB	-0.10415348	0.06013702
OBH	**0.17357421**	0.21259423
DKB	0.10388119	**0.36787812**
XML	0.11992018	-0.06917079
PAC	0.09111398	-0.12104866
PAS	0.03753227	0.23808921
MOW	0.09807527	-0.05743654
UFBR	-0.13700450	-0.16624665
FRA	0.02660505	0.04095824

4.7.1.3 **DFA plot of all landmarks after stepwise selection based on linguistics**

The various linguistic groups have minimal to no overlap among them even though they congregate closely with one another with regard to all cranial landmarks (Figure 26).

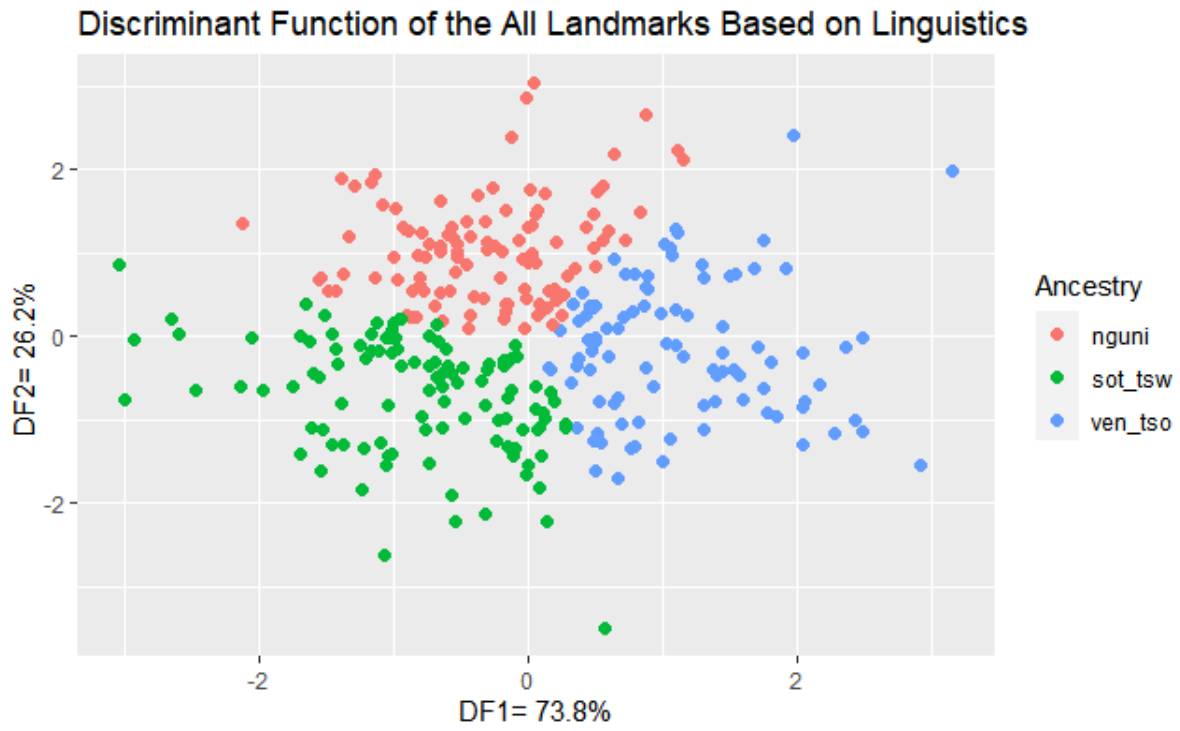


Figure 26. DFA plot of all landmarks after stepwise selection based on linguistics was conducted

4.7.1.4 **Stepwise linguistics: Classification matrix of the different groups using all variables**

All three linguistic clusters had a correct classification percentage that was greater than chance (33%). The overall model classification of this DFA model is 45.19%. The Kappa value for this DFA model is 16.94% thus indicating that the model was unable to effectively distinguish the individuals from the different groups (Table 32).

Table 32. Stepwise linguistics: Classification matrix of the different groups based on linguistics

Classification matrix					
	Number	nguni	sot_tsw	ven_tso	Correct Classification Percentage (%)
Nguni	147	64	47	36	43.54
sot_tsw	93	27	43	23	46.24
ven_tso	72	21	17	34	47.22
Total	312	112	107	93	
Overall model classification accuracy					45.19

4.7.1.5 **Cranial base linguistics: The percentage distribution of the different linear discriminant**

The first linear discriminant accounted for the 60.7% of the variation that was observed in the model (Table 33).

Table 33. Cranial base linguistics: The percentage distribution of the different linear discriminants

Proportion of trace	
LD1	LD2
0.607	0.393

4.7.1.6 **The coefficients of the linear discriminants of the cranial base variables based on linguistics**

The mastoid height was the variable that accounted for most of the variation that was observed in both the first and second linear discriminant (Table 34).

Table 34. Cranial base linguistics: Coefficients of linear discriminant

Variable	LD1	LD2
FOB	-0.09077769	0.1102331
AUB	0.05269806	0.1288275
ASB	0.18252712	-0.1261764
MDH	**-0.19483451**	**-0.2468160**
FOL	0.01679822	-0.1374992

4.7.1.7 **The DFA plot of the cranial base variables based on linguistics**

The different linguistic groups are congregated closely with one another with minimal to no overlap among the three different with regard to the cranial base landmarks. A similar congregation pattern is observed with all cranial landmarks but has a different pattern in how the groups are arranged (Figure 27).

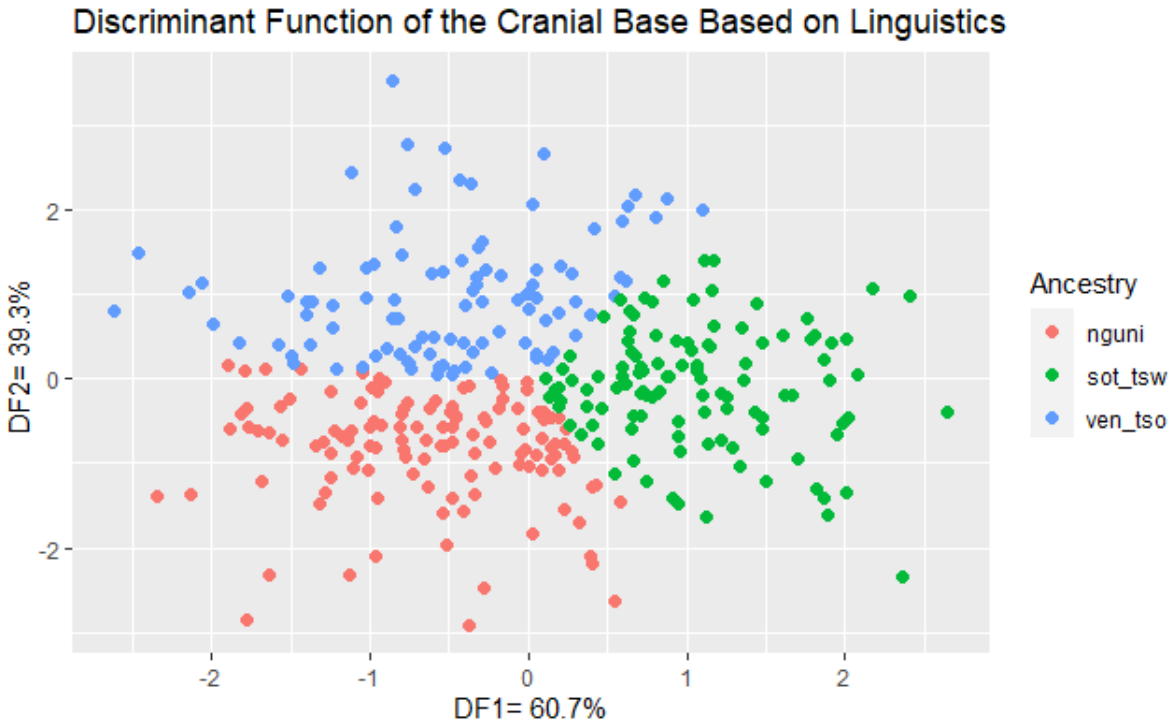


Figure 27. The DFA plot of cranial base variables based on linguistics

4.7.1.8 **The classification matrix of the cranial base landmarks based on linguistics**

The overall model classification accuracy of this model is 37.31% with only the Venda- Tsonga cluster obtaining a correct classification percentage less than chance (33%). The model had a kappa value is 4.81%. The low kappa value indicates that the model was unable to separate the individuals from the various groups adequately (Table 35).

Table 35. Cranial base linguistics: Classification matrix of the cranial base landmarks based on linguistics

Classification matrix					
	Number	nguni	sot_tsw	ven_tso	Correct Classification Percentage (%)
nguni	165	64	51	50	38.78
sot_tsw	92	30	39	23	42.39
ven_tso	78	27	29	22	28.21
Total	335	121	119	95	
Overall model classification accuracy					37.31

4.7.1.9 **Splanchnocranium linguistics**

Variables that were removed in stepwise selection:

SSA, WNB, OBH, XFB, DKB, ZYB, OBB, SIS, FMR, ZOR, MOW, WFB, NAS, NLB, NFA, SSS, EKB, XML, UFBR, SSR DKS, DKR, EKR

Variables that are part of the DFA model:

NLH, JUB, NDS, ZMB, FMB, IML, MLS, WMH, NAR, ZMR

4.7.1.10 **Splanchnocranium linguistics: The percentage distribution of the different linear discriminants**

The first linear discriminant function accounted for 78.92% of the variation that was observed in the model (Table 36).

Table 36. Splanchnocranium linguistics: The percentage distribution of the different linear discriminants

Proportion of trace	
LD1	LD2
0.7892	0.2108

4.7.1.11 **Splanchnocranium linguistics: Coefficients of linear discriminant**

NDS (naso-dacryal subtense) was the variable that accounted for most of the variation that was observed on the first linear discriminant while the MLS (malar subtense) accounted for most of the variation observed in the second linear discriminant (Table 37).

Table 37. Splanchnocranium linguistics: Coefficients of linear discriminant

Variable	LD1	LD2
NLH	-0.10864649	-0.277000493
JUB	-0.09345632	0.001056669
NDS	**-0.44499445**	-0.151613368
ZMB	0.02093789	0.127977067
FMB	0.07730112	0.008222465
IML	-0.09505692	0.245442434
MLS	0.19945301	**-0.355294157**
WMH	0.27918439	0.009828096
NAR	0.11915072	0.132668366
ZMR	-0.13574624	-0.126625707

4.7.1.12 **The DFA plot of splanchnocranium landmarks based on linguistics**

The different groups congregate in a slightly similar orientation as in the DFA plot of cranial base variables; however, the individual points are sparsely located than the previous DFA plot. The individual groups minimally or do not overlap with each other as in the previous two DFA plots (Figure 28).

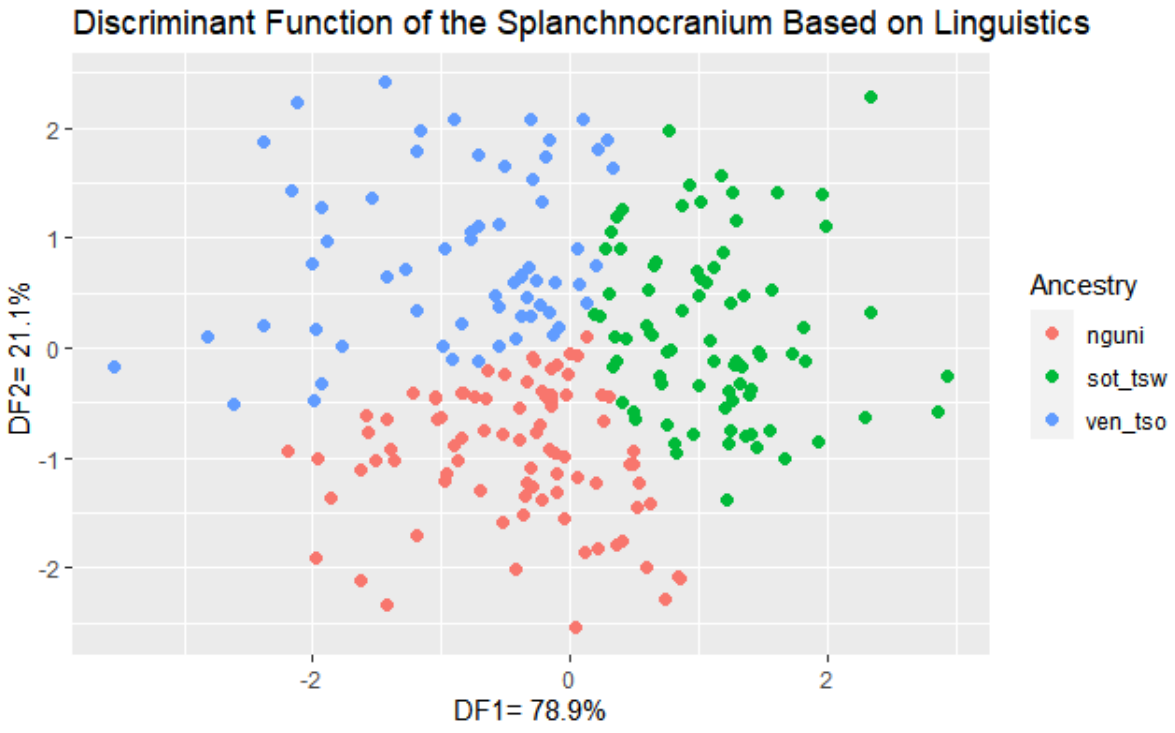


Figure 28. DFA plot of splanchnocranium landmarks based on linguistics

4.7.1.13 **Splanchnocranium linguistics: Classification matrix of the splanchnocranium landmarks based on linguistics**

The overall correct classification of this model is 49.78%. All of the groups obtained a correct classification percentage that is greater than chance (33%) with the Sotho-Tswana cluster obtaining the higher correct classification percentage (59.09%). The Kappa value of this model is 23.09% which indicates that the model was unable to separate groups adequately (Table 38).

Table 38. Splanchnocranium linguistics: Classification matrix of the splanchnocranium landmarks based on linguistics

Classification matrix					
	Number	nguni	sot_tsw	ven_tso	Correct Classification Percentage (%)
nguni	112	54	29	29	48.21
sot_tsw	66	15	39	12	59.09
ven_tso	53	20	11	22	41.51
Total	231	89	79	63	
Overall model classification accuracy					49.78

4.7.1.14 **Cranial vault linguistics**

Variables that were removed in stepwise selection:

FRC, BNL, OCS, RFA, PAF, SLA, SBA, PAS, NBA, OCA, BBA, FRF, XCB, ROA, BBH

Variables that are part of the DFA model:

GOL, OCC, PAC, NOL, STB, FRS, OCF, PAA, RPA, TBA

4.7.1.15 **Cranial vault linguistics: The percentage distribution of the different linear discriminants**

The first linear discriminant function accounted for 73.44% of the variation that was observed in the model (Table 39).

Table 39. Cranial vault linguistics: The percentage distribution of the different linear discriminants

Proportion of trace	
LD1	LD2
0.7344	0.2656

4.7.1.16 **Splanchnocranium linguistics: Coefficients of linear discriminant**

FRS (Nasion-Bregma Subtense) was the variable that accounted for most of the variation that was observed on the first linear discriminant while the GOL (Glabello-Occipital Length) accounted for most of the variation observed in the second linear discriminant (Table 40).

Table 40. Splanchnocranium linguistics: Coefficients of linear discriminant

Variable	LD1	LD2
GOL	0.066520871	**0.3365763151**
OCC	-0.049694541	0.0710843746
PAC	0.009270095	-0.1284302959
NOL	-0.072674400	-0.2580328730
STB	0.075935828	0.1151716458
FRS	**-0.165151733**	-0.0294064619
OCF	-0.107919516	-0.0443693943
PAA	-0.051617899	-0.0848355294
RPA	0.026235448	0.0006144923
TBA	-0.007358217	-0.0041805772

4.7.1.17 **DFA plot of cranial vault landmarks based on linguistic**

The groups congregate in a different orientation from the previous DFA plots relating to the linguistics but the similar to those plots the groups do not overlap with one another (Figure 29).

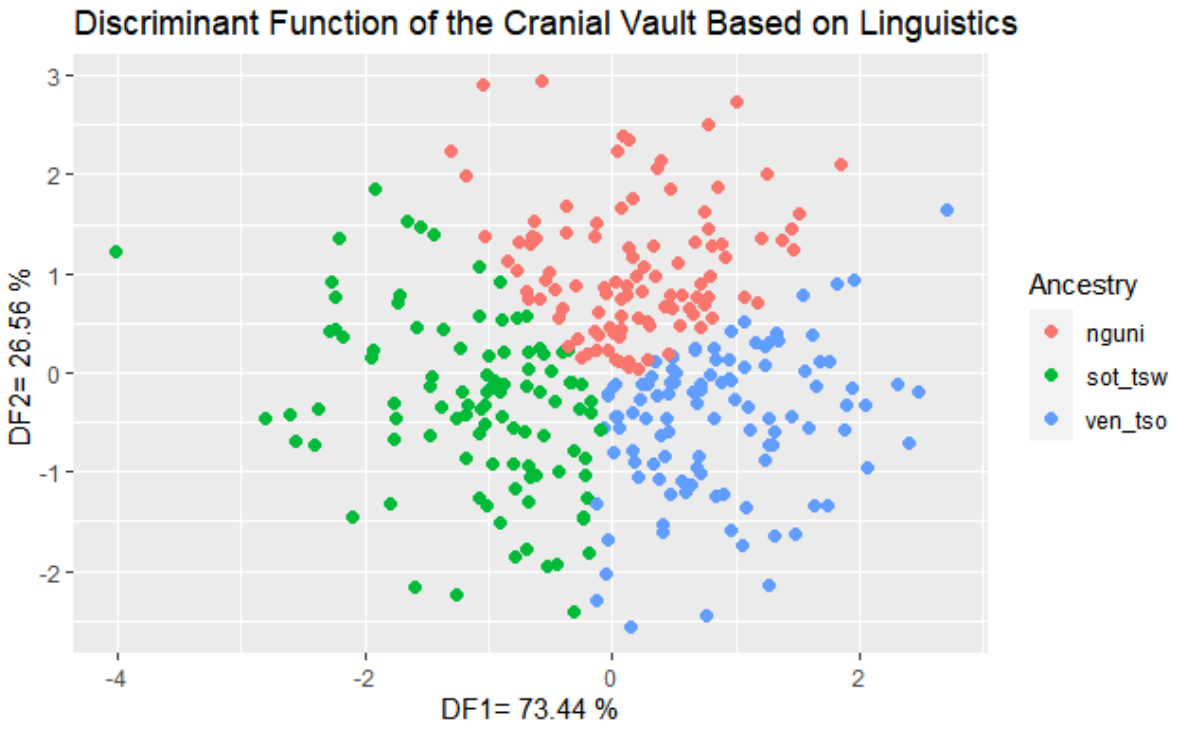


Figure 29. DFA plot of cranial vault landmarks based on linguistic

4.7.1.18 **Cranial vault linguistics: Classification matrix of the cranial vault landmarks based on linguistics**

The overall correct classification of this model is 42.90%. All of the groups obtained a correct classification percentage that is greater than chance (33%) with the Sotho-Tswana cluster obtaining the higher correct classification percentage (48.28%). The Kappa value of this model is 14.27% which indicates that the model was unable to separate groups adequately (Table 41).

Table 41. Cranial vault linguistics: Classification matrix of the cranial vault landmarks based on linguistics

Classification matrix					
	Number	nguni	sot_tsw	ven_tso	Correct Classification Percentage (%)
nguni	150	56	47	47	37.33
sot_tsw	87	25	42	20	48.28
ven_tso	73	23	15	35	47.95
Total	310	104	104	102	
Overall model classification accuracy					42.90

Chapter 5 : Discussion

The objective of this study was to assess cranial variation among black South African males using traditional linear measurements and non-standard inter-landmark distances. This was to ascertain whether the black South African category in FORDISC 3.1 database could be refined for classification purposes. Considerable overlap exists among all eight groups when they were assessed as separate groups when using all landmarks through stepwise selection and using the cranial base and cranial vault variables. A similar pattern was observed when the groups were grouped based on geographical location. However, when they were grouped based on historical linguistics classifications, they did not display similar overlapping patterns, as they did when they were assessed individually or based on geographical location. Overall, the model classification accuracies were low. The model accuracies of these various DFA models were greater than chance but were inadequate for them to be considered for classification purposes within a forensic context.

The use of non-standard inter-landmark distances was beneficial in further elucidating cranial variation among black male South Africans. Spradley and Jantz (2016) noted similar results when using non-standard inter-landmark distances to assess cranial variation of diverse American black, American white and Hispanic populations. The study found that the greatest differences in non-standard interlandmark distances were associated with the mid-face and cranial vault for both males and females.

Among black South Africans, most variation was found in the mid-face and the cranial vault, such as occipital fraction (OCF), naso-dacryal subtense (NDS), nasio-frontal subtense (NAS), cheek height (WMH), maximum cranial breadth (XCB) and naso-dacryal angle (NDA) when compared within and between the groups, and with the DFA models. The current study has shown that DFA models that assess groups individually do not have high overall model classification accuracies as well as kappa values. Classification models looking exclusively at variables of the splanchnocranium obtained the highest accuracies irrespective of geographic or linguistic separations. The clustering of the groups into their historical linguistic lineage obtained the highest overall model classification accuracy of 49.78%. This classification accuracy is higher in comparison to when all groups were used (21.33%) and when the groups were clustered based on their geographical location (35.79%).

The overlap that was observed among the groups is similar to what was observed in the De Villiers (1968) study wherein the various DFA scatterplots, the groups clustered and overlapped each other. However, the further assertions of the De Villiers (1968) study of them being "virtually impossible to separate" cannot be fully supported by this study. Including the rationale of clustering the Swazi and Zulu as a single group due to their similarities that De Villiers (1968) highlighted. As the Zulu and Swati groups did not significantly misclassify as the other when compared to other groups in the various classification matrices. In the current study, no justifiable reason exists to group Swazi and Zulu individuals into a single, homogeneous group as was stated by De Villiers (1968). When all the landmarks and the individual groups were assessed separately, the Sotho and Swazi groups were divergent from the other groups. The most seen post-hoc pairwise comparison was between Swazi and Sotho individuals. The Sotho group was completely separated from the other groups (See Figure 21). These two groups not only were divergent from the other groups but had higher correct classification percentages when the splanchnocranium and cranial vault were assessed. The findings of this study demonstrate that South African groups, which are commonly classified as black South Africans, are not a homogeneous population, which was also noted in research by Franklin and colleagues (2007).

All black South African cultural groups, while heterogenous, do display overlap even when separated into geographic and linguistic clusters. The overlap among ethnic groups is likely based on historical circumstances. During the colonial and *apartheid* eras, intermarriages of the different cultural groups were never prohibited through laws as it was between white and black South Africans (e.g., Stull *et al*. 2014). The migrant labour system that was enforced on black, male South Africans ensured, through both circumstance and necessity, frequent intermarriage among groups in the 20th century. Similarly, economic migrants, most of whom are black, male South Africans, in larger cities such as Johannesburg and Pretoria continue to maintain intermarriage across cultural/linguistics groups in South Africa (South African census 2016). Franklin and colleagues (2007) also concluded that the fluid movement of people throughout the country during *apartheid*, and likely previously, increased intermarriage among cultural/linguistics groups and increased group homogeneity, despite the evidence being available on group variation. Gene flow among the early Bantu groups that occurred can also explain the overlap that we see in the current study from the various DFA models. Pakendorf and colleagues (2011) highlight that the genetic variation observed in the mitochondrial DNA and the Y chromosome in Bantu speakers is indicative that intermarriage among the groups

was a common practice. Furthermore that the language differences did not inhibit the likelihood of the different Bantu groups from intermarrying due to multilingualism being common within the continent (Pakendorf *et al.*, 2011). The early Bantu speakers who migrated from Western African never had prolonged periods of isolation from one another. Pakendorf and colleagues (2011) state the studies that assessed the mitochondrial DNA diversity of African populations indicate that males chose wives from a different cultural group due to that cultural group residing near them. This study found that the genetic diversity observed in mitochondrial DNA was influenced more by geographical location instead of the language that the female spoke (Pakendorf *et al.*, 2011). The patriarchal social structure of the African population necessitates that newlywed women leave their place of birth and reside with their husband's family (Pakendorf *et al.*, 2011). The possibility of gene flow through intermarriages was likely as these groups were residing close to each other and also shared the same ideological paradigm with worldwide outlook i.e *lobola* (bride price) and patriarchal hierarchy as well as similar material culture that is well documented in archaeological literature(Huffman, 2007, Pakendorf *et al.*, 2011).

The pooling of the groups of the historical linguistic lineages resulted in a clear separation of the different three clusters. The result was unexpected considering the amount of overlap that was seen when various models were used. Not only was there clear distinct three clusters, but it also produced the highest overall model accuracy (49.78%) when the splanchnocranium variables were observed. The pooling of the groups in the historical linguistic lineages produced enough differences among the clusters adequately that resulted in them irrespective of which type of cranial variables were used. This is a novel finding regarding ancestry estimation studies relating to South African black groups. No study has ever assessed these groups using the historical lineages for classification purposes. This may alternatively be a means to separate the groups rather than looking at the different groups individually. Most studies that have been conducted relating to the black South African groups have typically assessed these groups individually.

One of the objectives of this study was to see whether the black South African category in the classification software FORDISC 3.1 could be further refined. The clustering of these groups based on historical linguistic lineage may possibly be an alternative way to classify unknown individuals instead of the broad category of black South African. This may provide a better possibility of narrowing down the options that are associated with the classification of an unidentified individual as a black South African in this country. An ancestry estimation with a

historical linguistic lineage estimation, associated with specific languages, may address the query of whether we can state that when we classify unknown remains as being most likely those of a black South African person who may have likely spoken a certain language. This query may seem trivial given that there has not been any literature that associates an ancestry estimation to a language that the deceased individual may have spoken. In the viewpoint of the law enforcement official that lessens the number of options in the broad classification of black South African in this country where there are nine cultural groups encompassed by that estimation. Given that certain languages are spoken predominately in specific provinces in the country so these historical linguistic lineages estimations may provide a possible geographical location where an unknown person may have likely lived in the country. A good example of this is the clustering of the Nguni languages which are mostly spoken in the eastern, western and southern parts of the country. The prevalence of certain historical languages to specific regions of the country occurred not through chance but rather as a design of the racist *apartheid* regime. The remnants of the Group Areas Act of 1950 which dictated where certain area based on their skin colour and language that they spoke are still fully intact in the modern-day South African society.

The historical linguistic lineage data can also be useful for exclusionary purposes for individuals that produce atypical results when run through classification software such as FORDISC 3.1. The data may provide some insight into whether such atypical individuals are such because they are not likely to be part of historical linguistic lineages of South Africa or just individuals are outliers but are part still of the South African groups. This sort of information may be useful for a province such as Gauteng that has numerous unclaimed bodies that are lying in state morgues as highlighted by studies such as Steyn and colleagues (1997), L'Abbé and Steyn (2012) and Nienaber (2015). The movement of economic migrants from within the country and from neighbouring countries to the Gauteng province, for better employment opportunities, has compounded this ever-growing issue.

Chapter 6 : Conclusion

Considerable overlap has been observed among when South African groups were assessed individually for ancestry estimation. The overall model accuracies of all the DFA models were greater than chance but were too low to be reliably used for ancestry estimation purposes. The splanchnocranium cranial variables produced the DFA models with the highest overall model accuracies. Intergroup differences were predominately found in nonstandard interlandmark distances which were mostly located in the splanchnocranium. The pooling of the groups into their historical linguistic lineages was able to ensure to separate the different groups into separate clusters. The findings of this study dispute the finding that South African black groups are homogenous as De Villiers (1968) suggested. There are differences among these groups but not quantifiable enough to be separate the groups for forensic purposes. However, the use of historical linguistic lineages may potentially be an alternative way to refine the black South African category in the South African database in FORDISC 3.1.

This study had various limitations in the manner it was conducted. The specimens that were used in the study were those of South African Black male individuals. The outcomes of this study cannot apply to South African Black female individuals with regard to cranial variation. The study consisted of 8 groups that were investigated instead of the 9 groups that were initially envisaged. Thus, the outcomes are not necessarily the entire reflection of the amount of cranial variation among the crania of South African Black males, as the Pedi males did not form part of the study due to the low sample size. This study only made use of Discriminant Function Analysis to assess the cranial among South African black males which only looks at size differences. In future, the use of other statistical analysis such as geometric morphometrics can provide greater insight into the crania of South African black males.

Chapter 7 : References

Abdi & Williams 2010. Tukey's honestly significant difference (HSD) test. *Encyclopedia of research design,* 3**,** 1-5.

Afifi, May & Clark 2011. *Practical multivariate analysis*, CRC Press.

Arts and Culture Department 1999. Republic of South Africa Government Gazette No. 19974 (No. 25 of 1999 : National Heritage Resources Act, 1999). *In:* CULTURE, A. A. (ed.). Cape Town, South Africa.

Asuero, Sayago & Gonzalez 2006. The correlation coefficient: An overview. *Critical reviews in analytical chemistry,* 36**,** 41-59.

Bass 1971. Human osteology: a laboratory and field manual of the human skeleton.

Bernitz, Kenyhercz, Kloppers, L'Abbé, Labuschagnes & Olckers 2015. The history and current status of forensic science in South Africa. *The Global Practice of Forensic Science.* John Wiley & Sons, Chichester.

Cattaneo 2007. Forensic anthropology: developments of a classical discipline in the new millennium. *Forensic Science International* 165**,** 185-93.

Cattaneo & Baccino 2002. A call for forensic anthropology in Europe. *International Journal of Legal Medicine,* 116**,** N1-N2.

Christensen & Crowder 2009. Evidentiary standards for forensic anthropology. *Journal of Forensic Science,* 54**,** 1211-6.

Clark, Collinson, Kahn, Drullinger & Tollman 2007. Returning home to die: circular labour migration and mortality in South Africa 1. *Scandinavian Journal of Public Health,* 35**,** 35-44.

Dayal, Kegley, Štrkalj, Bidmos & Kuykendall 2009. The history and composition of the Raymond A. Dart Collection of human skeletons at the University of the Witwatersrand, Johannesburg, South Africa. *American Journal of Physical Anthropology,* 140**,** 324-335.

De Villiers 1968. *The skull of the South African negro,* Witwatersrand University Press.

Delius, Schoeman & Maggs 2014. *Forgotten World: The Stone Walled Settlements of the Mpumalanga Escarpment*, Wits University Press.

Department of Health 2007. Government Gazette No.30075, National Health Act, 2003 (Act No. of 2003) Regulations Regarding the Rendering of Forensic Pathology Service (No. R. 636). *In:* HEALTH (ed.).

Dirkmaat, Cabo, Ousley & Symes 2008. New perspectives in forensic anthropology. *American Journal of Physical Anthropology,* 137**,** 33-52.

Elliott. 2008. *FORDISC and the determination of ancestry from craniometric data.* Dept. of Archaeology-Simon Fraser University.

Franklin, Freedman, Milne & Oxnard 2007. Geometric morphometric study of population variation in indigenous southern African crania. *American Journal of Human Biology,* 19**,** 20-33.

Hall 1987. *The Changing Past: Farmers, Kings and Traders in Southern Africa, 200-1860,* D. Philip.

Harris & Smith 2009. Accounting for measurement error: a critical but often overlooked process. *Archives of oral biology,* 54**,** S107-S117.

Herbert & Huffman 1993. A new perspective on Bantu expansion and classification: linguistic and archaeological evidence fifty years after Doke. *African Studies,* 52**,** 53-76.

Howells 1973. Cranial variation in man. A study by multivariate analysis of patterns of difference. Among recent human populations. *Papers of the Peabody Museum of Archaeology and Ethnology***,** 1-259.

Howells 1989. Skull shapes and the map: craniometric analyses in the dispersion of modern Homo. *Papers of the Peabody museum of Archaeology and Ethnology,* 79.

Huffman 1970. The Early Iron Age and the spread of the Bantu. *The South African Archaeological Bulletin,* 25**,** 3-21.

Huffman 1989. Ceramics, settlements and late Iron Age migrations. *African archaeological review,* 7**,** 155-182.

Huffman 2007. *Handbook to the Iron Age: The Archaeology of Pre-colonial Farming Societies in Southern Africa,* University of KwaZulu-Natal Press.

Huffman 2009. Mapungubwe and Great Zimbabwe: The origin and spread of social complexity in southern Africa. *Journal of Anthropological Archaeology,* 28**,** 37-54.

Huffman 2010. Debating the central cattle pattern: a reply to Badenhorst. *The South African Archaeological Bulletin***,** 164-174.

Humphries, Maxwell, Ross & Ubelaker 2015. A Geometric Morphometric Study of Regional Craniofacial Variation in Mexico. *International Journal of Osteoarchaeology,* 25**,** 795-804.

Katherine Spradley & Jantz 2016. Ancestry estimation in forensic anthropology: geometric morphometric versus standard and nonstandard interlandmark distances. *Journal of Forensic Sciences,* 61**,** 892-897.

Kim 2014. Analysis of variance (ANOVA) comparing means of more than two groups. *Restorative dentistry & endodontics,* 39**,** 74.

Konigsberg, Algee-Hewitt & Steadman 2009. Estimation and evidence in forensic anthropology: sex and race. *Am J Phys Anthropol,* 139**,** 77-90.

Krüger. 2015. *Comparison of sexually dimorphic patterns in the postcrania of South Africans and North Americans.* University of Pretoria.

Krüger, Liebenberg, Myburgh, Meyer, Oettlé, Botha, Brits, Kenyhercz, Stull & Sutherland 2018. Forensic anthropology and the biological profile in South Africa. *New Perspectives in Forensic Human Skeletal Identification.* Elsevier.

L'Abbe, Kenyhercz, Stull, Keough & Nawrocki 2013. Application of FORDISC 3.0 to explore differences among crania of North American and South African blacks and whites. *Journal of Forensic Sciences,* 58**,** 1579-83.

L'Abbé & Steyn 2012. The establishment and advancement of forensic anthropology in South Africa. *A Companion to Forensic Anthropology***,** 626-638.

L'Abbé, Loots & Meiring 2005. The Pretoria bone collection: a modern South African skeletal sample. *HOMO-Journal of Comparative Human Biology,* 56**,** 197-205.

Liebenberg, L'Abbe & Stull 2015. Population differences in the postcrania of modern South Africans and the implications for ancestry estimation. *Forensic Science International* 257**,** 522-9.

Maake 1991. Language and politics in South Africa with reference to the dominance of the Nguni languages. *English Studies in Africa,* 34**,** 55-64.

Malan 1985. Migrant labour in southern Africa. *Africa Insight,* 15**,** 103-114.

McHugh 2011. Multiple comparison analysis testing in ANOVA. *Biochemia medica,* 21**,** 203-209.

Milne & O'Higgins 2002. Inter-specific variation in Macropus crania: form, function and phylogeny. *Journal of Zoology,* 256**,** 523-535.

Mukaka 2012. A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal,* 24**,** 69-71.

Ngwane 2003. 'Christmas time'and the struggles for the household in the countryside: rethinking the cultural geography of migrant labour in South Africa. *Journal of Southern African Studies,* 29**,** 681-699.

Nienaber 2015. The archaeological investigation of crime scenes and humanitarian cases that involve graves and human remains in South Africa. *Forensic Archaeology: A Global Perspective.*

Onwuegbuzie & Daniel 1999. Uses and misuses of the correlation coefficient.

Ousley & Jantz 2005. FORDISC 3.1. *The University of Tennessee, Knoxville***,** 36-38.

Ousley, Jantz & Freid 2009. Understanding race and human variation: why forensic anthropologists are good at identifying race. *Am J Phys Anthropol,* 139**,** 68-76.

Ousley & McKeown 2001. Three dimensional digitizing of human skulls as an archival procedure. *BAR International Series,* 934**,** 173-186.

Pakendorf, de Filippo & Bostoen 2011. Molecular perspectives on the Bantu expansion: a synthesis. *Language Dynamics and Change,* 1**,** 50-88.

Poulsen & French 2008. Discriminant function analysis. *Retrieved from.*

Rexová, Bastin & Frynta 2006. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. *Naturwissenschaften,* 93**,** 189-194.

Shakoane. 2020. *Evaluating sexual dimorphism among South African groups using the dentition* Masters, University of Pretoria.

Smith 1971. The effect of unequal group size on Tukey's HSD procedure. *Psychometrika,* 36**,** 31-34.

Spradley, Jantz, Robinson & Peccerelli 2008. Demographic change and forensic identification: problems in metric identification of Hispanic skeletons. *Journal of Forensic Science,* 53**,** 21-8.

St & Wold 1989. Analysis of variance (ANOVA). *Chemometrics and intelligent laboratory systems,* 6**,** 259-272.

Statistics South Africa 2016a. Community Survey 2016, Statistical release P0301 *In:* PRESIDENCY, D. I. T. (ed.). Pretoria

Statistics South Africa 2016b. Mid-year population estimates 2016. *In:* PRESIDENCY (ed.). Pretoria,South Africa.

Steyn, Meiring & Nienaber 1997. Forensic anthropology in South Africa: a profile of cases from 1993 to 1995 at the Department of Anatomy, University of Pretoria. *South African Journal of Ethnology,* 20**,** 23-26.

Stomfai, Ahrens, Bammann, Kovacs, Mårild, Michels, Moreno, Pohlabeln, Siani & Tornaritis 2011. Intra-and inter-observer reliability in anthropometric measurements in children. *International Journal of Obesity,* 35**,** S45-S51.

Stull, Kenyhercz & L'Abbe 2014. Ancestry estimation in South Africa using craniometrics and geometric morphometrics. *Forensic Science International* 245C**,** 206 e1-206 e7.

Symes, Ericka, L'Abbé, Wolff & Dirkmaat 2012. Bone in medicolegal investigations. *A companion to forensic anthropology,* 10**,** 340.

Urbanová, Ross, Jurda & Nogueira 2014. Testing the reliability of software tools in sex and ancestry estimation in a multi-ancestral Brazilian sample. *Legal Medicine,* 16**,** 264-273.

Van der Merwe, Steyn & L'Abbé 2010. Trauma and amputations in 19th century miners from Kimberley, South Africa. *International Journal of Osteoarchaeology,* 20**,** 291-306.

Vansina 1980. Bantu in the crystal ball, II. *History in Africa,* 7**,** 293-325.

Viðarsdóttir, O'Higgins & Stringer 2002. A geometric morphometric study of regional differences in the ontogeny of the modern human facial skeleton. *Journal of Anatomy,* 201**,** 211-229.

Waldron 2008. *Palaeopathology,* Cambridge University Press.

Zeller 2004. Relative clause formation in the Bantu languages of South Africa. *Southern African Linguistics and Applied Language Studies,* 22**,** 75-93.

Chapter 8 : Appendix

8.1 Landmark Short Names

The measurements used in the study according to (Howells,1973: pg. 159-187),(Howells, 1989) a (Langley *et al*., 2016: pg. 61-70) (Table 42).

Table 42. The short names of the different cranial variables.

SHORT NAMES IN ALPHABETICAL ORDER	
ASB = Biasterionic Breadth	FRF = Nasion-Subtense Fraction
AUB = Biauricular Breadth	FRS = Nasion-Bregma Subtense
AVR = MI Alveolus Radius	GLS = Glabella Projection
BAA = Basion Angle, na-pr	GOL = Glabello-Occipital Length
BBA = Basion Angle, na-br	IML = Malar Length Inferior
BBH = Basion-Bregma Height	JUB = Bijugal Breadth
BNL = Basion-Nasion Length	LAR = Lambda Radius
BPL = Bas/ion-Prosthion Length	MAB = Palate Breadth
BRR= Bregma Radius	MDB = Mastoid Width
BRA=Bregma Angle	MDH = Mastoid Height
BSA= Basal Angle	MLS = Malar Subtense
DKA = Dacryal Angle	MOW= Midorbital Width
DKB = Interorbital Breadth	NAA = Nasion Angle, ba-pr
DKR = Dacryon Radius	NAR = Nasion Radius

DKS = Dacryon Subtense	NAS = Nasio-Frontal Subtense
EKB = Biorbital Breadth	NBA = Nasion Angle, ba-br
EKR = Ectoconchion Radius	NDA = Naso-Dacryal Angle
FMB = Bifrontal Breadth	NDS = Naso-Dacryal Subtense
FMR = Frontomalare Radius	NFA = Nasio-Frontal Angle
FOL = Foramen Magnum Length	NLB = Nasal Breadth
FRA = Frontal Angle	NLH = Nasal Height
FRC = Nasion-Bregma Chord	NOL = Nasio-Occipital Length
SHORT NAMES IN ALPHABETICAL ORDER	
NPH = Nasion-Prosthion Height	SIS = Simotic Subtense
OBB = Orbit Breadth Left	SLA= Sub-Lambda Angle
OBH = Orbit Height Left	SOS = Supraorbital Projection
OCA = Occipital Angle	SSA = Zygomaxillare Angle
OCC = Lambda-Opisthion Chord	SSR = Subspinale Radius
OCF = Lambda-Subtense Fraction	SSS = Zygomaxillary Subtense
OCS = Lambda-Opisthion Subtense	STB = Bistephanic Breadth
OSR = Opisthion Radius	TBA= Trans-Basal Angle
PAA = Parietal Angle	VRR = Vertex Radius

PAC = Bregma-Lambda Chord	WCB = Minimum Cranial Breadth
PAF = Bregma-Subtense Fraction	WMH = Cheek Height
PAS = Bregma-Lambda Subtense	WNB = Simotic Chord
PRA = Prosthion Angle, na-ba	XCB = Maximum Cranial Breadth
PRR = Prosthion Radius	XFB = Maximum Frontal Breadth
RFA= Radio-Frontal Angle	XML = Malar Length Maximum
ROA= Radio-Occipital Angle	ZMB = Bimaxillary Breadth
RPA= Radio-Parietal Angle	ZMR = Zygomaxillare Radius
SBA=Sub-Bregma Angle	ZOR = Zygoorbitale Radius
SIA = Simotic Angle	ZYB = Bizygomatic Breadth

Landmark Definitions

Alveolon aIv (Landmark 77)

The point where the mid-sagittal plane of the palate is intersected by a line connecting the posterior borders of the alveolar crests. To assist in locating the point, place a rubber band or wire against the posterior margins of the alveolar processes of the maxilla

Alare alar (Landmark 4 & 7)

The most lateral points on the nasal aperture.

Asterion as (Landmark 57 &68)

The common meeting point of the temporal, parietal, and occipital bones, on either side.

If the meeting point is occupied by a wormian bone (Os astericum), extend the lambdoid suture onto its surface, and then extend the other two sutures (temporo-parietal, temporo-occipital) to the first line, finding asterion as the point midway between the intersections if these do not coincide (BM). Use only the part of the last two sutures (ca. 1 cm) which is nearest the point, in finding these directions. If the lambdoid (or other) suture is complex or composed of wormian bones, trace a pencil line along the center of the area covered by the complexity, as well as can be done, to find the main axis of the suture. If the sutures gape at this point, leaving an open space, find asterion on the edge of the occipital bone.

Basion ba (Landmark 70)

On the anterior border of the foramen magnum, in the midline, at the position pointed to by the apex of the triangular surface at the base of either condyle, i.e., the average position from the crests bordering this area. Mark carefully with a pencil.

In the most usual specimen, the border of the foramen will have a thickness of 1-2 mm and a rounded edge. The position chosen will be about halfway between the inner border directly facing the posterior border (opisthion) and the lowermost point on the border, i.e., between the points usually designated endobasion and hypobasion respectively. As a practical matter, the point will almost always be the endpoint of the basion-nasion length if the calliper is applied here to find a maximum. It will correspond with endobasion only if there is a thin, sharp border to the foramen. The variation in structure here is considerable: in the thickness of the border, and the presence of a small tubercle or a larger articular surface. In the case of the former, displace basion to one side or the other; in the case of an articular surface, place the point on this, trying to estimate the position from directions above. To estimate basion in a damaged skull, use a transverse line connecting the posterior limits of the bases of the spinous processes on either side. The elevation of basion in such a case can, however, be only guessed at.

Bregma br (Landmark 55)

The posterior border of the frontal bone in the median plane.

Normally this is the meeting point of the coronal and sagittal sutures. The latter may diverge from the midline here, however, and should not then be followed. (Metopic sutures should be disregarded.) More commonly, the coronal suture may project slightly backwards in a small point here, from the general smooth curve of the suture on either side; or the two halves of the

suture may meet in a short antero-posterior line, i.e., one-half may lie forward of the other where they reach the midline. In these and similar cases, the general course of the suture should be lightly drawn with a pencil, and the bregma established on this. The point should mark the limits of the frontal and parietal segments of the vault generally, not minor sutural variations. If the coronal suture is nearly obliterated, its course must be established from any remaining traces; if it is completely obliterated, there is nothing to do but estimate the position of bregma. The sutures may meet with rounded external edges, resulting in a cleft or depression at their junction. Bregma is then to be established "in the air" (BM), i.e., in its correct position but at the level of the general surface of the bone and not, by sinking the calliper point into such a fissure in measuring, below this surface. Some device, such as displacing bregma slightly to one side in the same transverse plane, may be followed. In any other questionable case where such a choice may be necessary, bregma is considered to be on the frontal bone.

Dacryon dk (Landmark 23 & 24)

The apex of the lacrimal fossa, as it impinges on the frontal bone. Mark with a pencil point on both sides.

In the ideal well-preserved specimen, the groove will be clearly defined and sharply apexed, the apex corresponding well with the inner wall of the orbit as viewed by the sighting. The groove will be bisected by the lacrimo-maxillary suture, which will meet the fronto-lacrimal and fronto-maxillary sutures (the frontal bone) at the groove's apex. The inner border of the orbit, curving down from above, will form a slight promontory overhanging the apex of the groove and just lateral to it. The point determined should be on the frontal bone. There is much variation from the above pattern: the fossa may be shallow with a broad or ill-defined apex; the suture may be obliterated; the lacrimal bone itself may be absent anatomically or lost postmortem. Approximate the point defined above, i.e., the apex, by using, in order of priority:

(a) the lacrimal fossa observed from directly above, a view that makes it easy to determine its course and the proper point of its apex.

(b) the promontory on the frontal bone-the best guide when the lacrimal bone is broken out entirely.

(c) the posterior border of the fossa - the point never lies posterior or lateral to it but may approach it.

(d) the lacrimo-maxillary suture, when the structures are whole, but the form of the fossa is shallow and undefined.

(e) there is often a small foramen just at the apex of the fossa, which may be used as a guide, though it is apt to lie slightly mesial to the apex proper.

Ectoconchion ek (Landmark 25)

The intersection of the most anterior surface of the lateral border of the orbit and a line bisecting the orbit along its long axis. Mark both sides with a pencil.

Hold the flat of a pencil lead so that this surface is perpendicular to the median plane of the skull, i.e., tangent to the most anterior curvature of the orbital margin, and use it to draw a line along this crest. Turn the skull to be able to sight along the long axis of the orbit, however oblique, and to bisect the orbit visually, with the pencil as a sighting guide. Make a tick with the pencil where this axis appears to intersect the line already made.

Frontomalare anterior fm: a (Landmark 37 & 48)

The most anterior point on the frontomalar suture. It may be found with the side of a pencil lead held in the transverse plane.

This is neither the orbital nor the temporal point of the suture (Martin's frontomalare orbitale and temporale respectively), nor is it the line of the orbital border. It is strictly the most anteriorly projecting point and is used for measurements relating to such projection. The suture may be and usually is, quite irregular in its course here. No corrective for this is offered: the point is placed on the suture wherever it may

Frontomalare fmt (Landmark 36 & 49)

The most laterally positioned point on the fronto-malar suture.

Frontotemporale ft (Landmark 38 & 47)

A point located generally forward and inward on the superior temporal line directly above the zygomatic process of the frontal bone. The right and left frontotemporale form the endpoints of the minimum frontal breadth measurement.

Glabella g (Landmark 53)

The most anteriorly projecting point in the mid-sagittal plane at the lower margin of the frontal bone, which lies above the nasal root and between the superciliary arches. The point of glabella is depressed between the confining bony ridges and is often delineated superiorly by a shallow gutter or a transversely running indentation on the surface of the frontal bone (Langely *et al*.,2016: 63).

Hormion hor (Landmark 76)

The intersection of the sphenoid bone and the posterior border of the vomer.

Jugale jug (Landmark 34 & 51)

The deepest point in curvature between the frontal and temporal process of the zygomatic bone.

Krotaphion Kr (Landmark 40 & 45)

The intersection of the squamosal and sphenoidal suture.

Lambda la (Landmark 56)

The apex of the occipital bone at its junction with the parietals, in the midline.

This is normally the meeting of the sagittal and lambdoid sutures, but must be placed in the midline. The ruling principle, as in the case of bregma, is to divide the parietal and occipital segments of the sagittal section of the skull (BM). There is often an intercalary or apical bone at the site, in which case lambda is to be found by extending the general curving course of each half of the lambdoid suture to their intersections with the midline; if these extensions do not meet the midline in a single point, lambda is halfway between such intersections (BM). In occasional cases, the apex of the occipital makes an obvious forward excursion along the midline, away from the general course of the suture, and probably resulting from an apical bone joined to the occipital. The same procedure as immediately above is followed. The lambdoid

suture itself may be very complex or composed largely of wormian bones. Trace a pencil line along the center of such an area on each side, to find lambda as above.

Marginal process lateral mpl (Landmark 35 & 50)

The most posteriorly projecting point on the frontal process of the zygomatic bone.

Mastoidale ms (landmark 61 & 66)

The most inferior point on the tip of the mastoid process.

Nasale inferius (Landmark 9 & 10)

The endpoint of the nasomaxillary suture.

Nasale superius

The intersection of the nasomaxillary suture and the frontal bone.

Nasion na (Landmark 52)

The intersection of the fronto-nasal suture and the median plane. Mark with a pencil.

This does not refer to the internasal suture in any way. If there is irregularity near the midline, rectify the general curve of the fronto-nasal suture with a pencil to find the correct level for nasion. Except for this last, a general rule is to consider nasion as on the frontal bone (BM, V). I.e., if the fronto-nasal suture forms a cleft or gap, locate nasion on the midline just at the angle between the facial and sutural surfaces of the frontal bone itself. Equivalent definitions: BM, V, M.

Nasomaxillary suture pinch wnb (Landmarks 12 & 14)

The most pinched in areas of the nasomaxillary sutures.

Opisthocranion op (Landmark 106)

The most distant point posteriorly from glabella on the occipital bone, located in the mid-sagittal plane. Opisthocranion almost always falls on the superior squama of the occipital bone, and only occasionally on the external occipital protuberance. Opisthocranion is established by obtaining the measurement of maximum cranial length.

Opisthion Os (Landmark 69)

The inferior edge of the posterior border of the foramen magnum in the midline.

The posterior border is virtually always either a sharp edge or one which makes a clear angle between the external surface and the actual border of the foramen. Equivalent definitions: BM, V (?)

Prosthion Pr (Landmark 1)

The most anteriorly prominent point, in the midline, on the alveolar border, above the septum between the central incisors. Mark with a pencil.

This most anterior point (with the skull erect in a position corresponding to the eye-ear plane) is generally easily seen, but if the bone descends smoothly into the interincisor septum, while also continuing to slope forward, find the point in the general line of the border elsewhere along with the gum. There is apt to be thickening or reinforcement of the border slightly below the arcs of exposure of the tooth roots along the row, especially if there has been some slight (not marked) retraction of the edges of the alveoli - this is the level sought for prosthion. Because of frequent damage, and of ante-mortem loss or avulsion of the incisors, the location of prosthion for actual measurement may be difficult, resulting in the need for approximations or estimates of its original position. Sighting along the border on either side of the central region will help. The point is often slightly recessed between the incisors themselves. If the recession is exceptionally deep, the point should correspond to the center of a gentler concavity, i.e., about that which would meet the rounded tip of the spreading calliper, so that this may be used directly in measuring basion-prosthion length.

Radiculare ra (Landmark 62 & 63)

The point on the lateral aspect of the root of the zygomatic process at the deepest incurvature.

Staurion

The intersection of the incisive suture and the palatine sutures.

Stephanion st (Landmark 42 & 43)

The intersection of the coronal suture and the limit of the temporal muscle (the inferior temporal line). Mark with a pencil on both sides.

The temporal line is generally divided before reaching the suture; if not, the upper limit is used. The lower line may follow a mutual course with the suture for a short distance, in which case the posterior end of this course is used. In general, it may help to imagine the point as the end of the inferior temporal line on a detached frontal bone. In some skulls surface erosion or poor definition make finding the point difficult. A better definition on one side may help on the other. Also, the point usually coincides with the lower end of the pan complicata of the suture, if this is discernible.

Subspinale ss (Landmark 3)

The deepest point is seen in the profile below the anterior nasal spine.

Practically, this relates to the deepest points found in measuring zygomaxillary subtense and subspinale radius; and one must bear in mind that the point is on the crest, or profile, not in the fissure of the intermaxillary suture. If the nasal spine is small or eroded, the Point is difficult to locate. It should not be placed internal to the outermost limit of the lower border of the aperture.

Zygomaxillare anterior zm: a (Landmark 29 & 30)

The intersection of the zygomaxillary suture and the limit of the attachment of the masseter muscle, on the facial surface. Mark with a pencil on both sides.

In very rare cases, and association with an os japonicum, the suture runs more posteriorly, and over a centimetre lateral to the anterior end of the masseter attachment. It then is apt to be located beyond the angle of the malar; in such cases, it is recommended that the point be placed on the facial surface, on the masseter limit, not more than 6-8 mm from the anterior end of the masseter area. If obliteration makes the facial part of the suture difficult to follow, an inspection of its course, if present, on the internal surface of the arch may help.

Zygion zy (Landmark 26 & 33)

The most laterally positioned point on the zygomatic arches. The landmark is used to obtain bizygomatic breadth.

Zygoorbitale zo (Landmark 16 & 17)

The intersection of the orbital margin and the zygomaxillary suture. Mark with a pencil.

Since the orbital border is usually softly rounded here, the point should be found midway between the facial and orbital surfaces. A small process of the malar may extend several millimetres mesially from the rest of the bone just here, pushing the suture and point well inward along the orbital margin, and increasing the measurement of malar length by the length of this sliver. As a convention, the point is never placed mesial to the plane of the medial border of the infraorbital foramen.

8.2 Boxplots

The boxplots of the individual cranial variables (Figure 30-106).

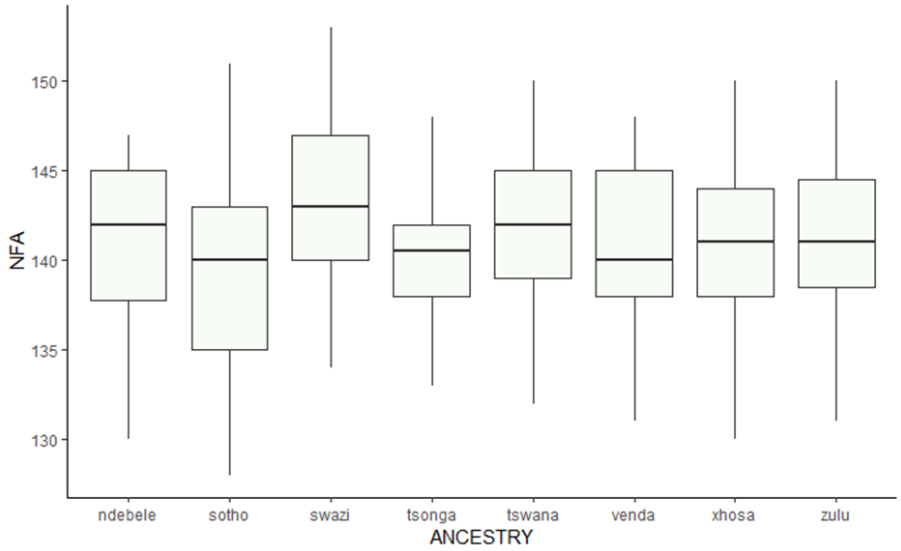


Figure 30. The boxplot of the nasio-frontal angle (NFA) variable.

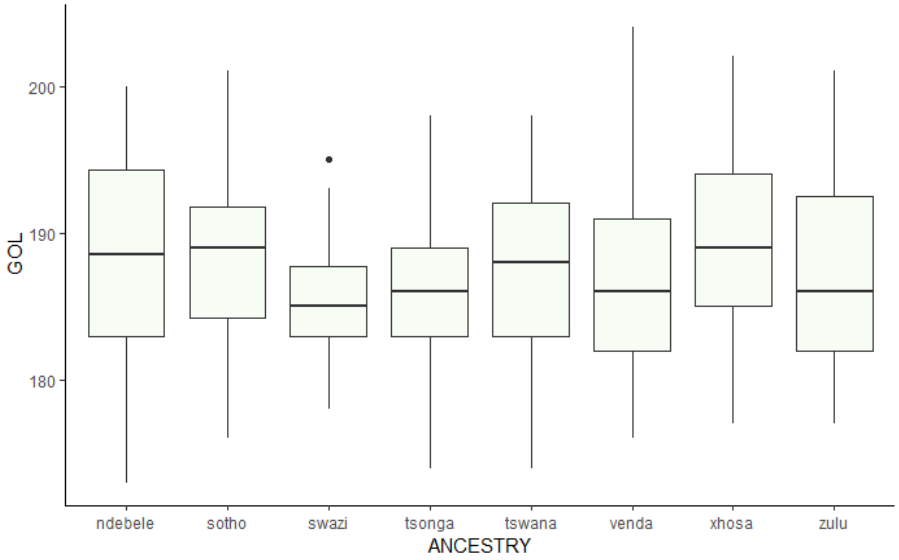


Figure 31. The boxplots of the glabello-occipital length (GOL) variable.

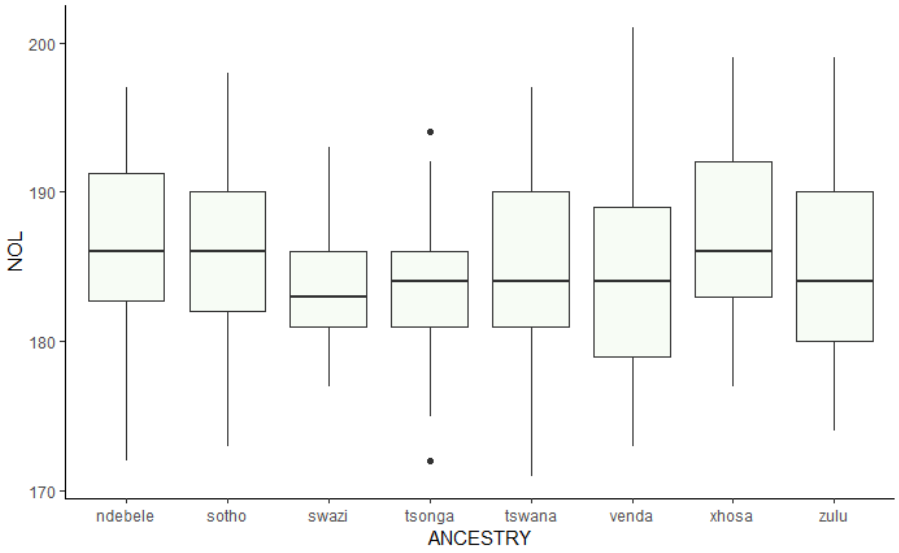


Figure 32. The boxplot of the nasio-occipital length (NOL) variable.

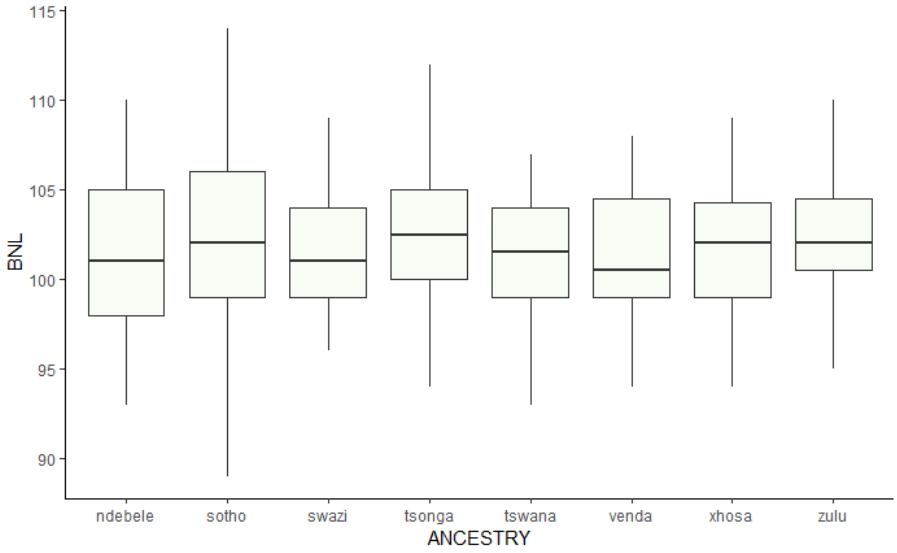


Figure 33. The boxplot of the basion-nasion length (BNL) variable.

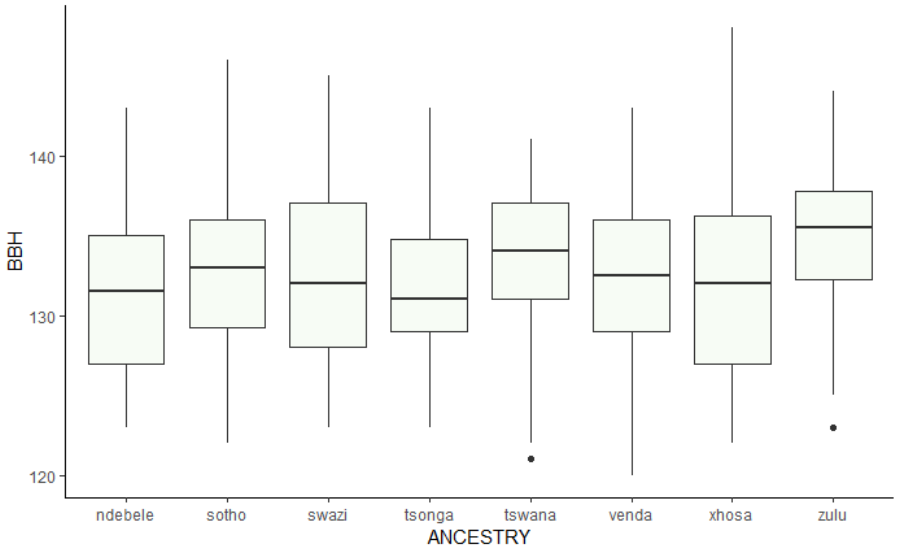
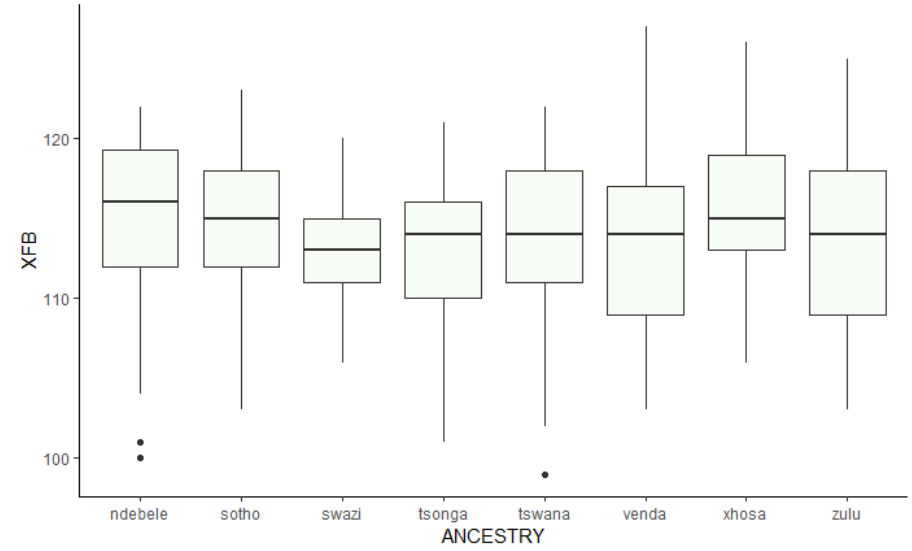


Figure 34. The boxplot of the basion-bregma height (BBH) variable.



.

Figure 35. The boxplot of the maximum frontal breadth (XFB) variable.

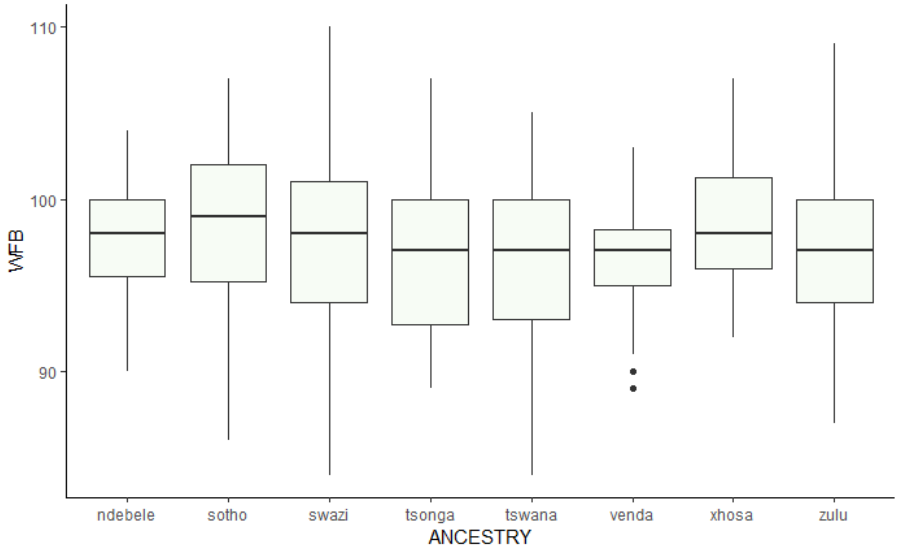


Figure 36. The boxplot of the minimum frontal breadth (WFB) variable.

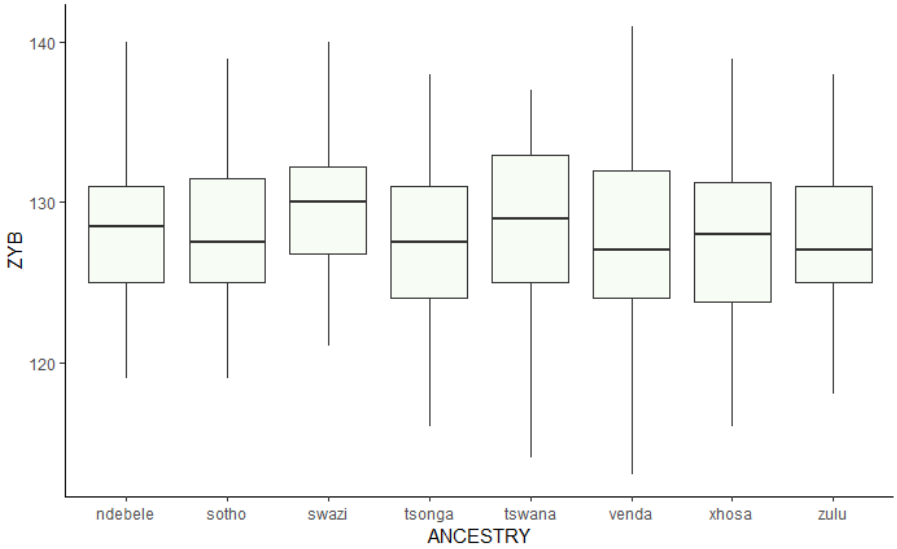


Figure 37. The boxplot of the zygomatic breadth (ZYB) variable.

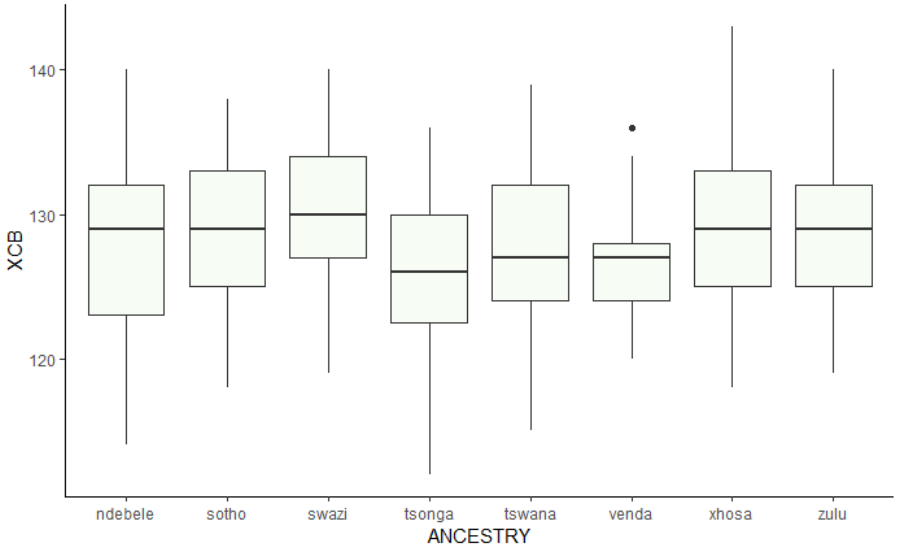


Figure 38. The boxplot of the maximum cranial breadth (XCB) variable.

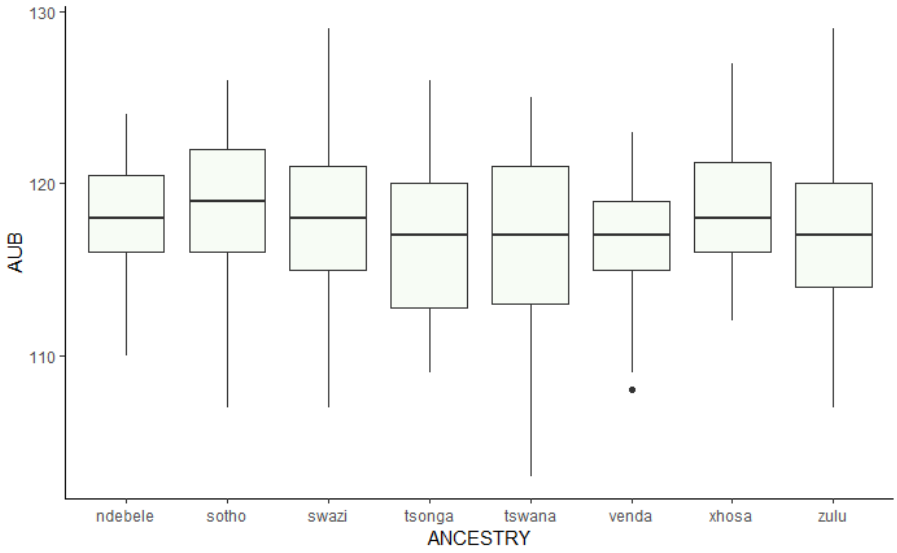


Figure 39. The boxplot of the biauricular breadth (AUB) variable.

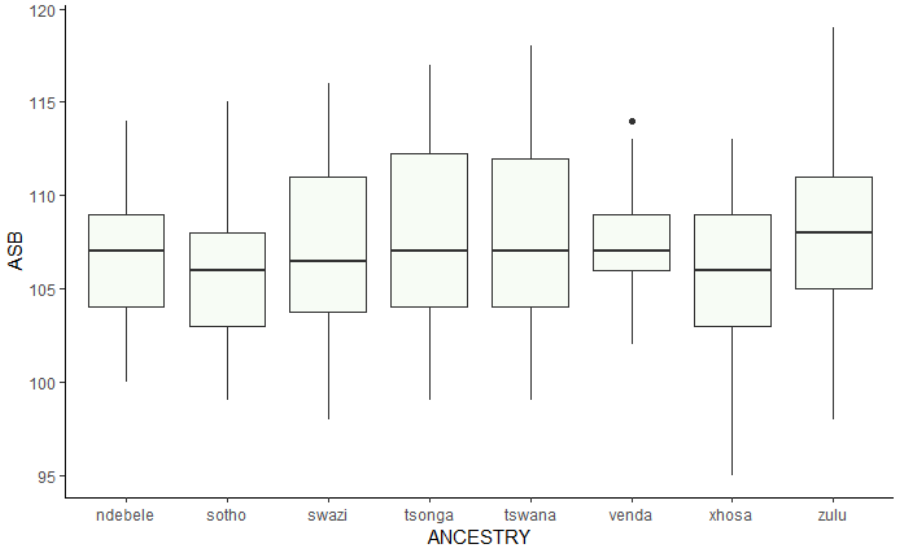


Figure 40. The boxplot of the biasterionic breadth (ASB) variable.

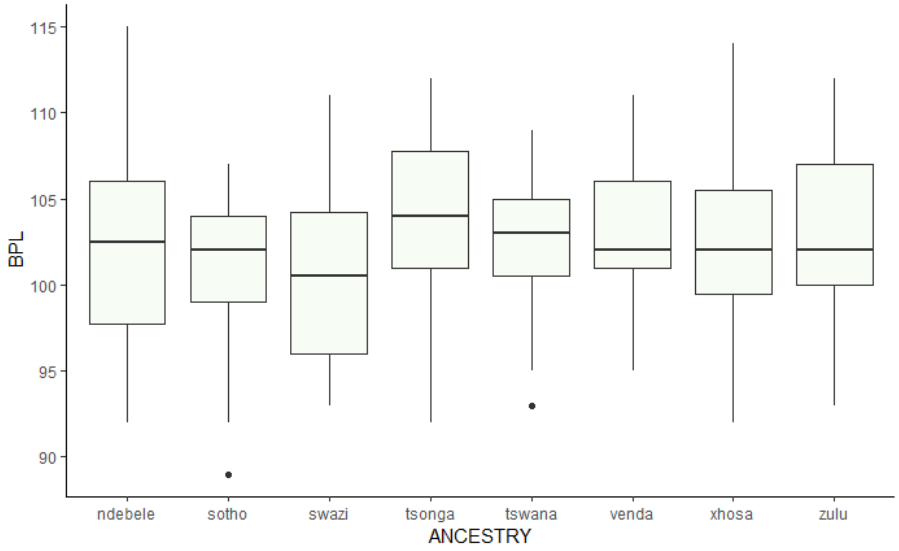


Figure 41. The boxplot of basion-prosthion length (BPL) variable.

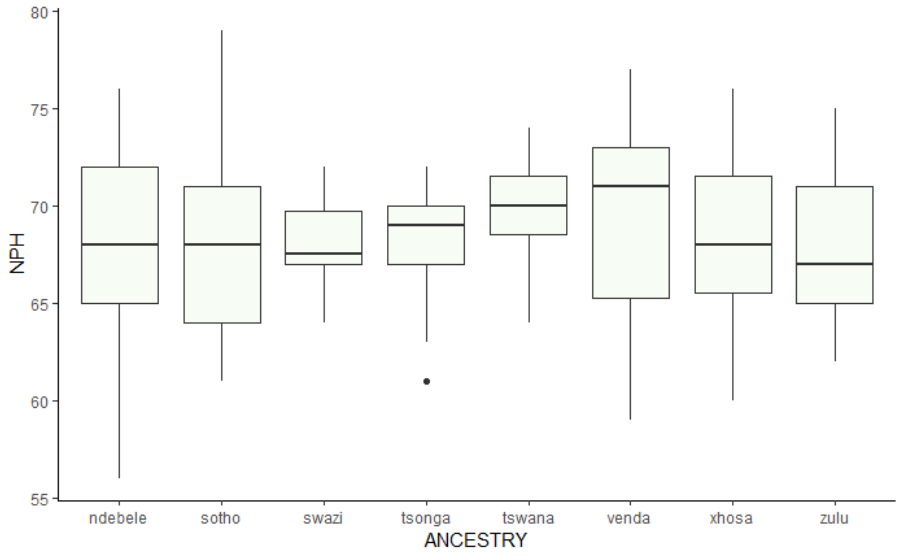


Figure 42. The boxplot of nasion- prosthion height (NPH) variable.

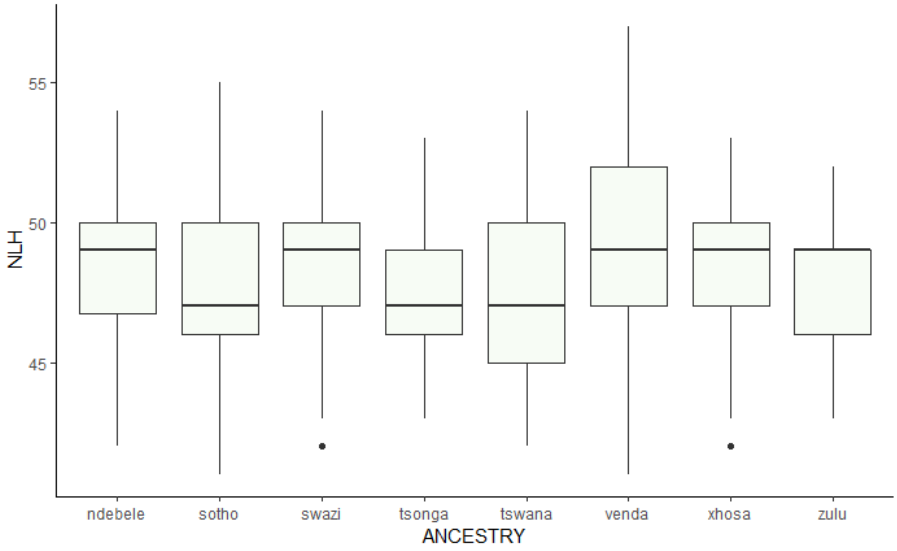


Figure 43. The boxplot of the nasal height (NLH) variable.

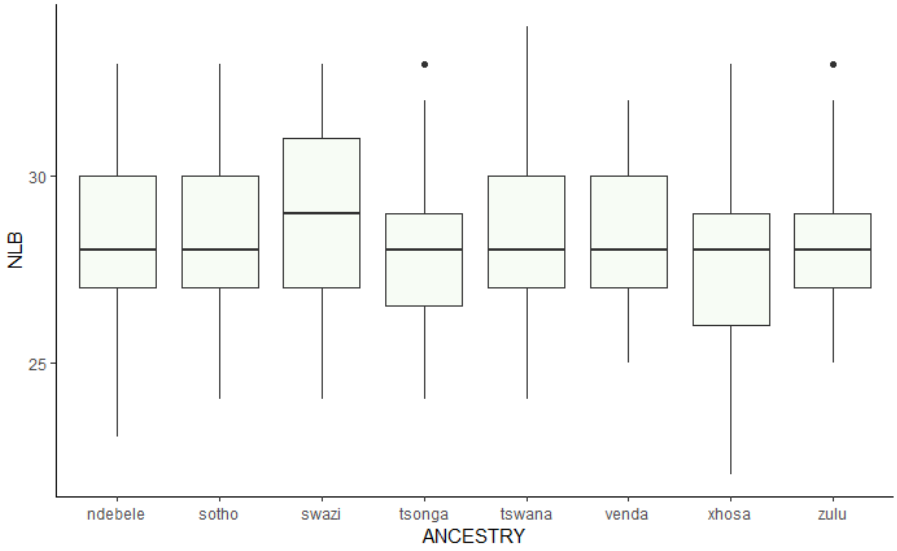


Figure 44. The boxplot of the nasal breadth (NLB) variable.

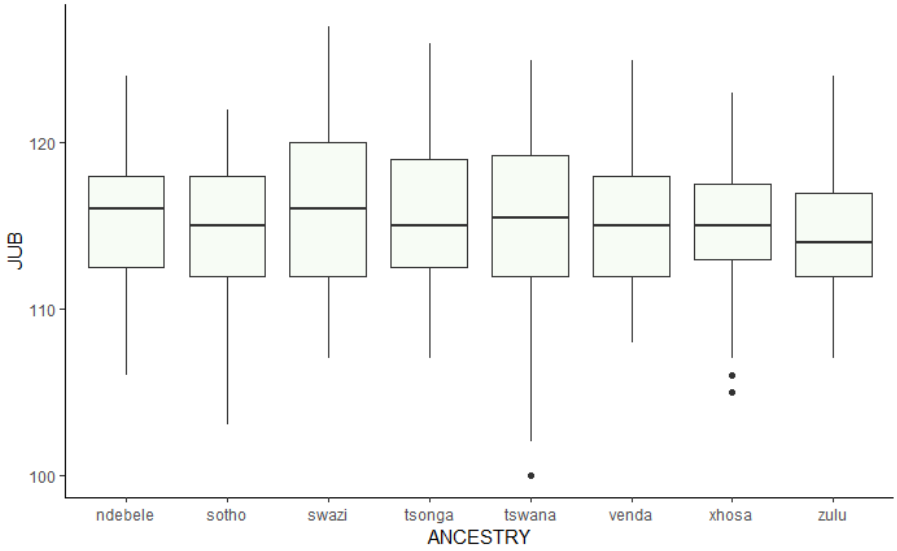


Figure 45. The boxplot of the bijugal breadth (JUB) variable.

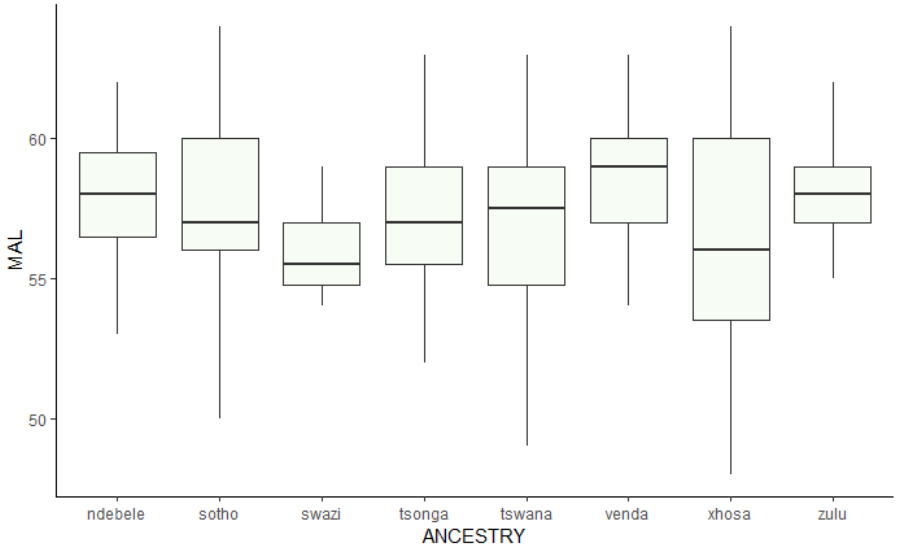


Figure 46. The boxplot of the maximum alveolar length (MAL) variable.

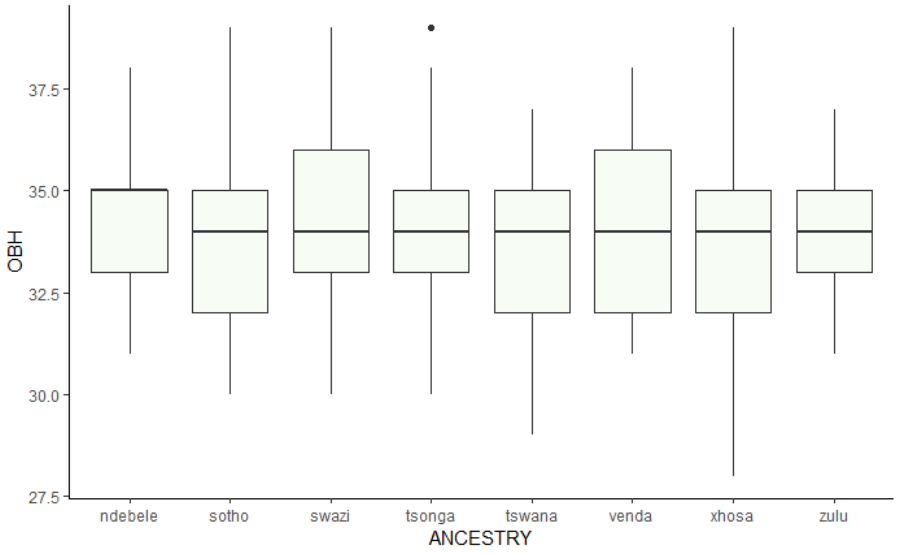


Figure 47. The boxplot of the orbital height (OBH) variable.

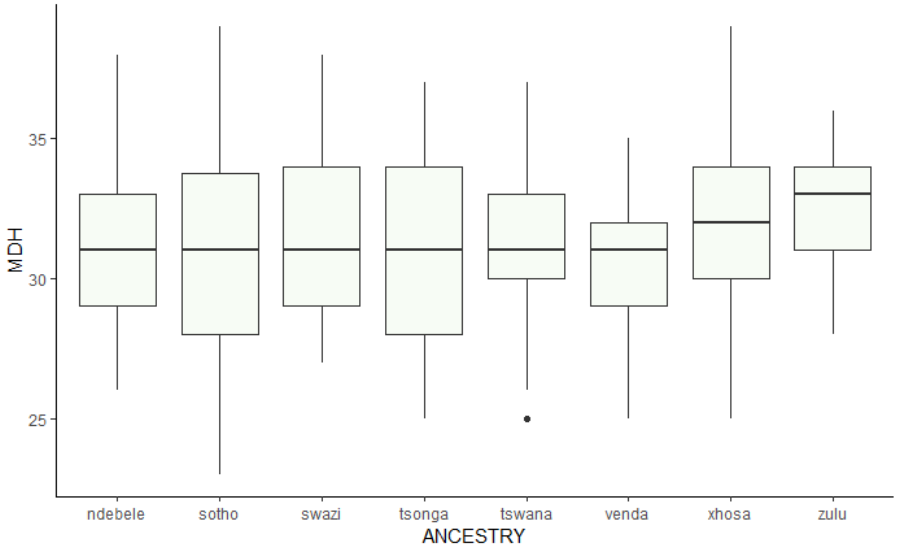


Figure 48. The boxplot of the mastoid height (MDH) variable.

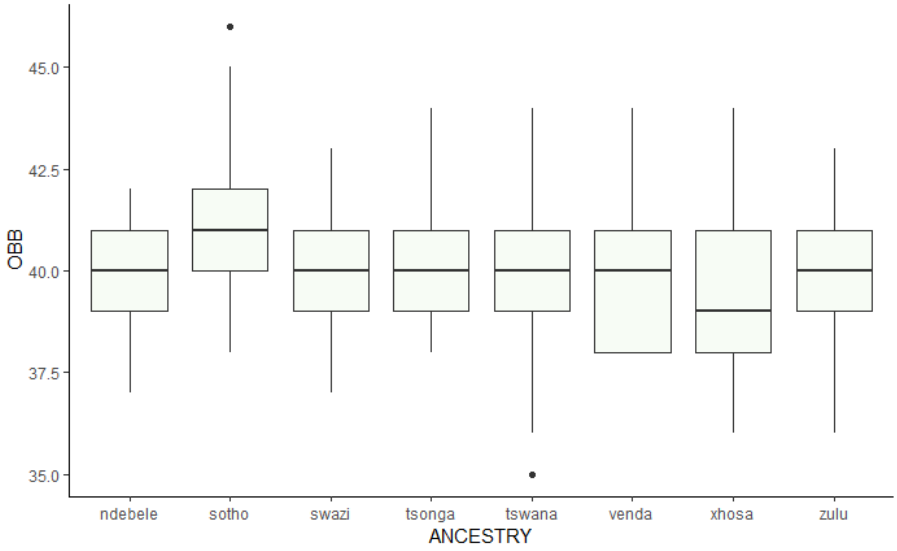


Figure 49. The boxplot of the orbital breadth (OBB) variable.

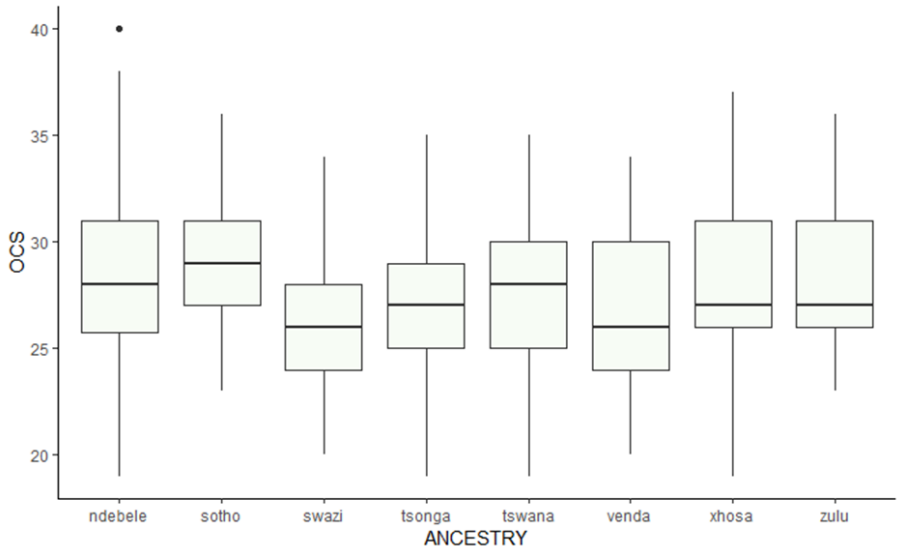


Figure 50. The boxplot of the lambda- opisthion subtense (OCS) variable.

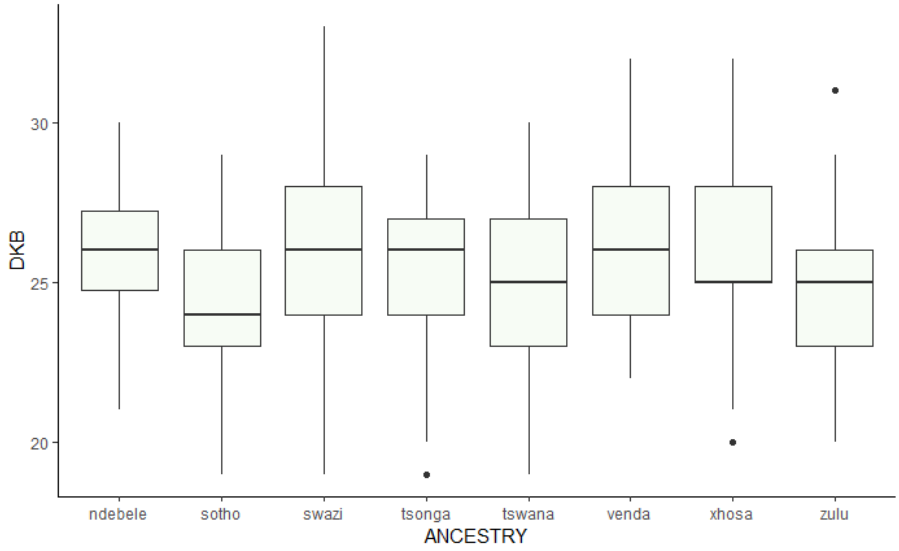


Figure 51. The boxplot of the interorbital breadth (DKB) variable.

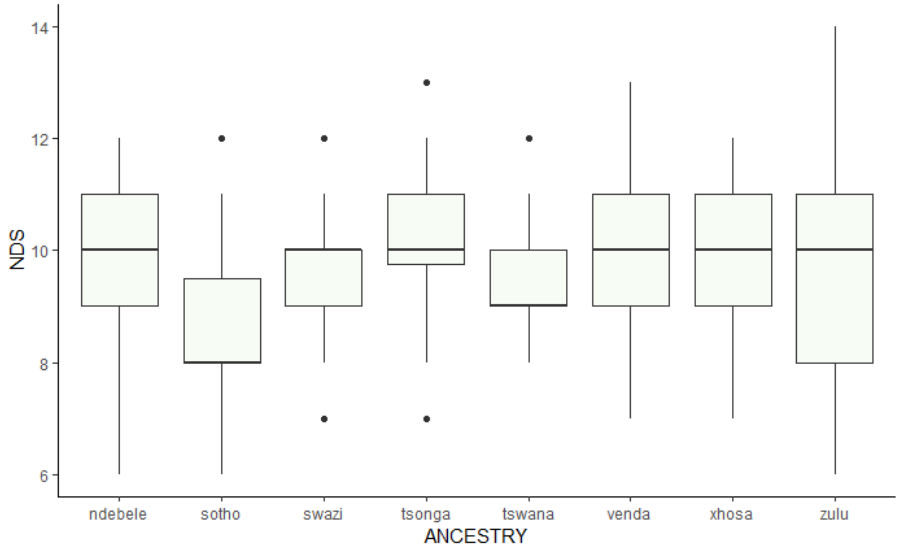


Figure 52. The boxplot of the naso-dacryal subtense (NDS) variable.

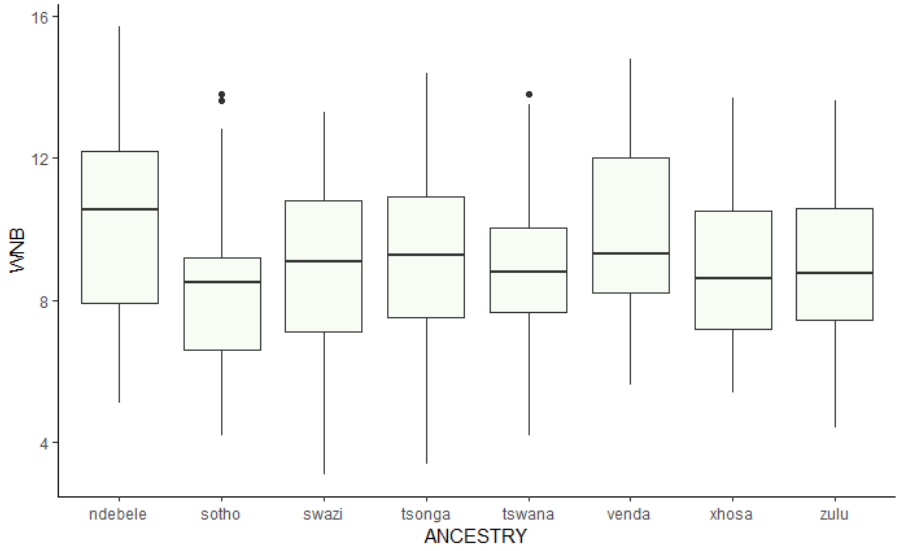


Figure 53. The boxplot of the simotic chord (WNB) variable.

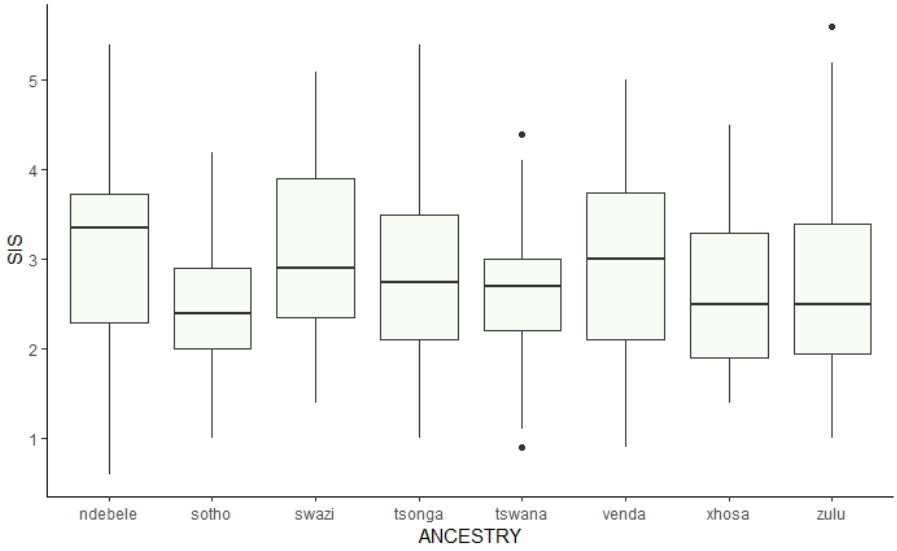


Figure 54. The boxplot of the simotic subtense (SIS) variable.

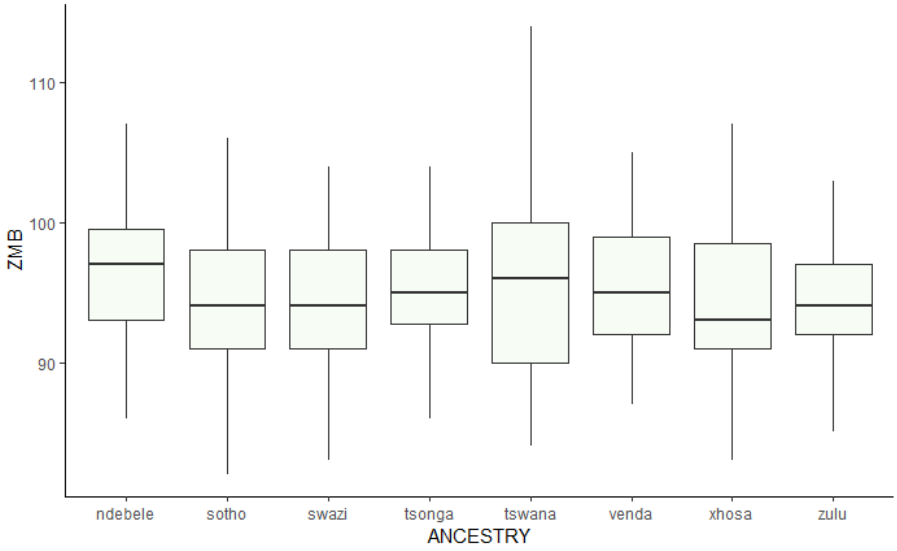


Figure 55. The boxplot of the bimaxillary breadth (ZMB) variable.

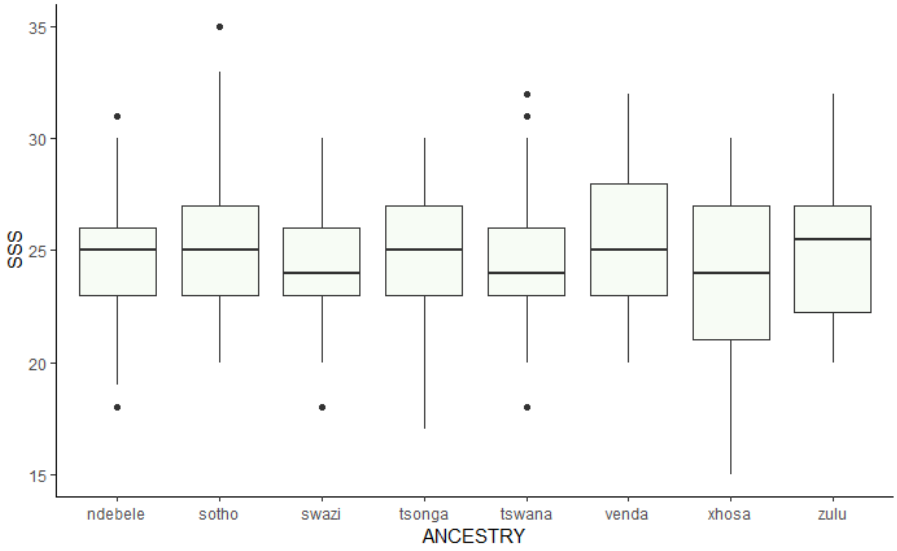


Figure 56. The boxplot of the zygomaxillary subtense (SSS) variable.

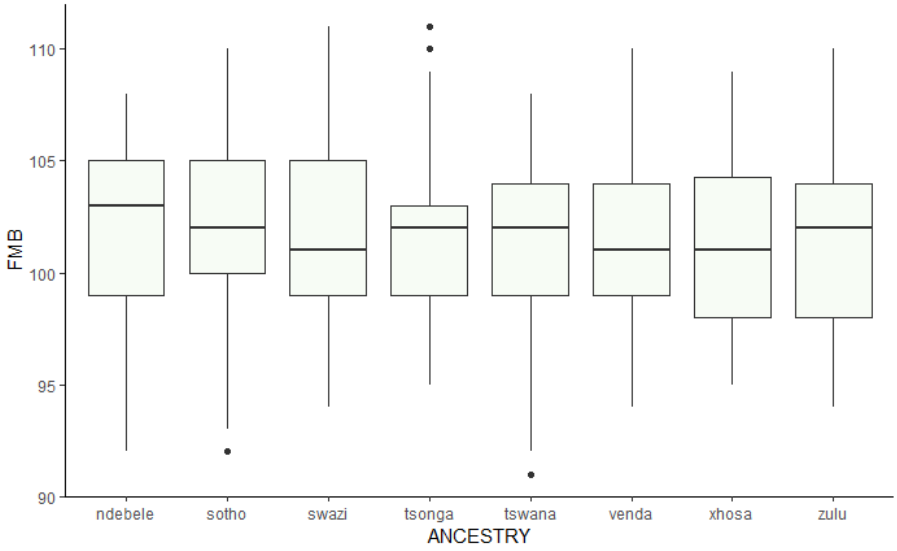


Figure 57. The boxplot of the bifrontal breadth (FMB) variable.

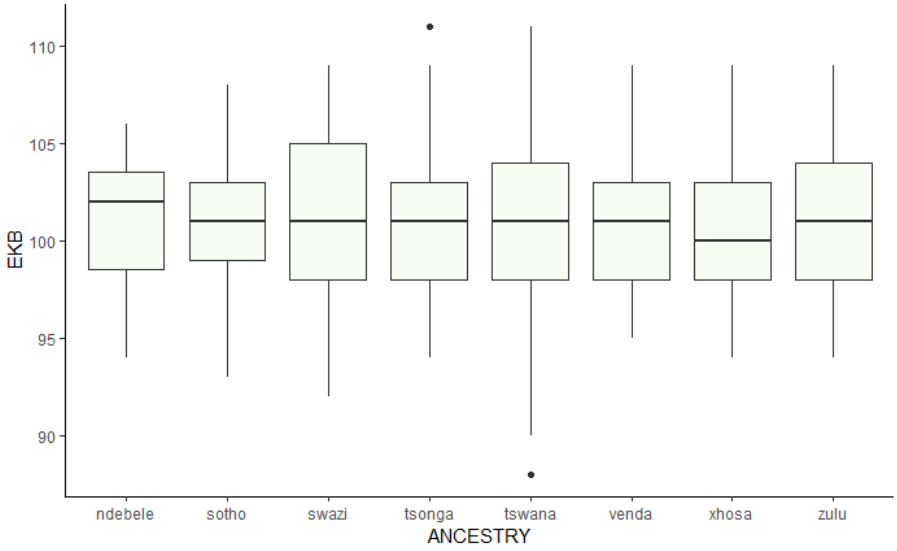


Figure 58. The boxplot of the biorbital breadth (EKB) variable.

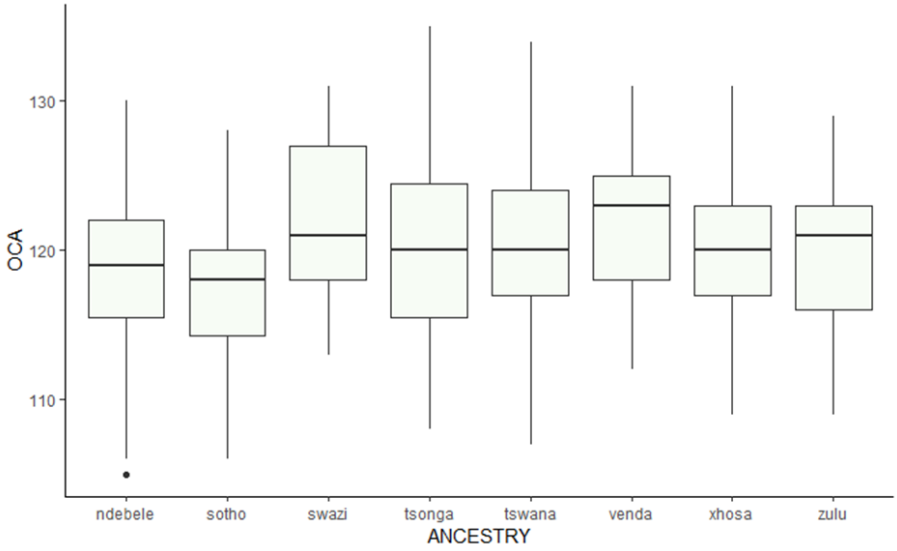


Figure 59. The boxplot of the occipital angle (OCA) variable.

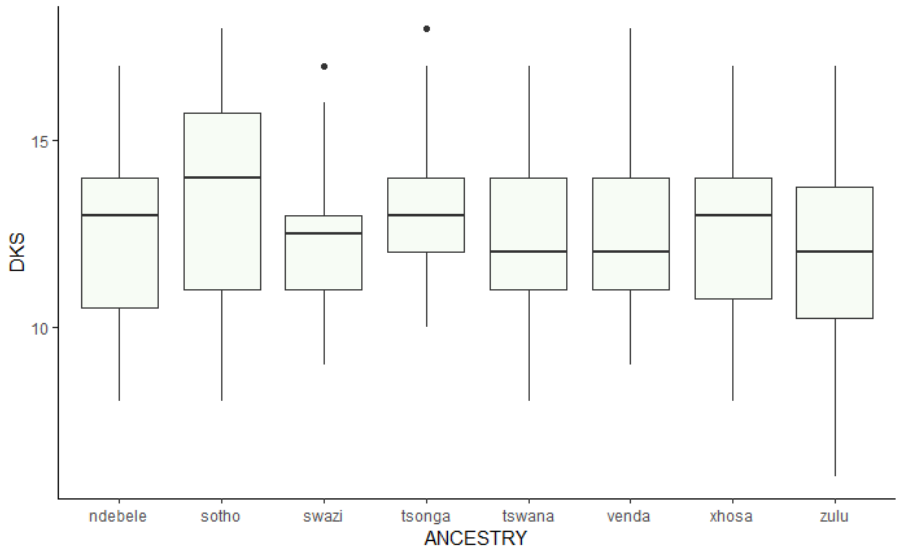


Figure 60. The boxplot of the dacryon subtense (DKS) variable.

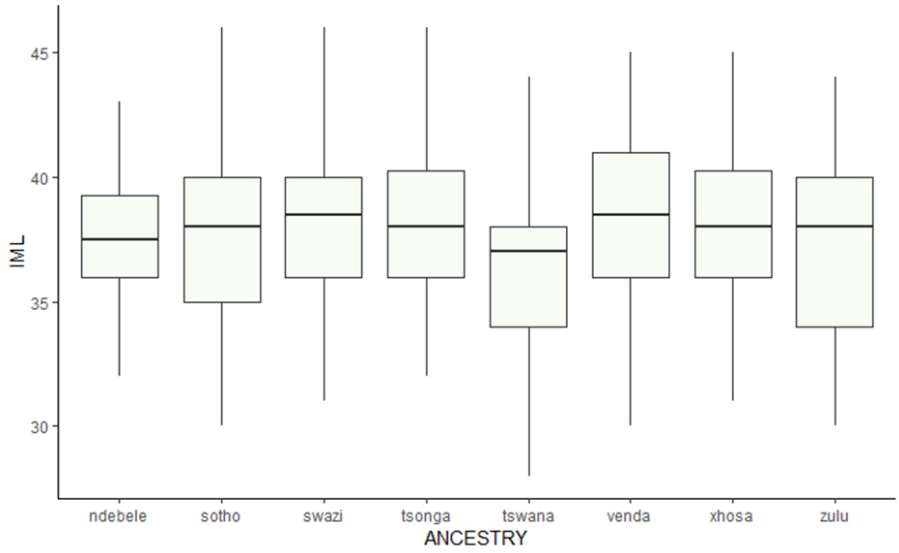


Figure 61. The boxplot of the malar length inferior (IML) variable.

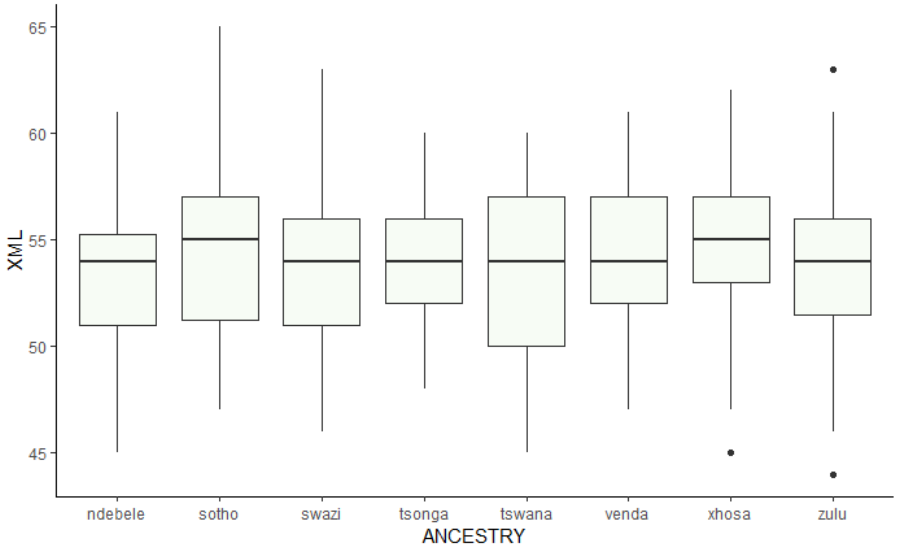


Figure 62. The boxplot of the malar length maximum (XML) variable.

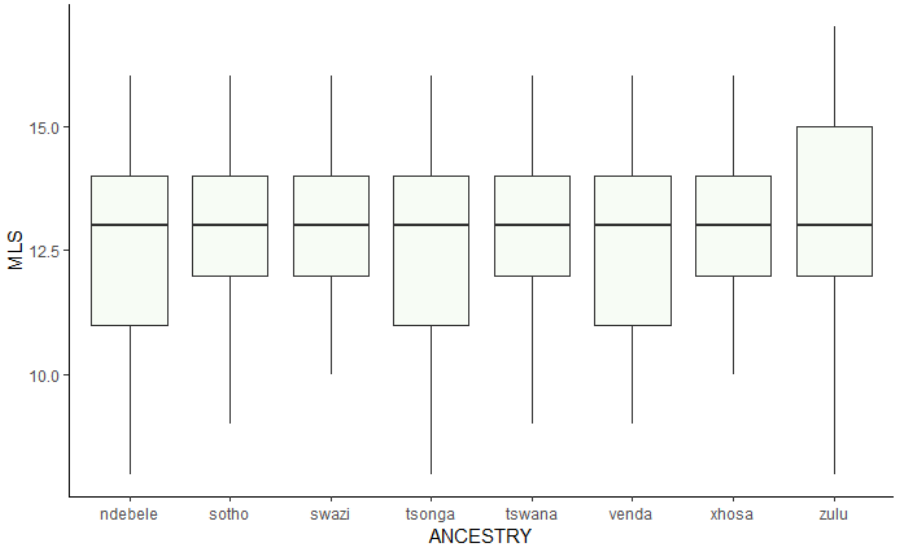


Figure 63. The boxplot of the malar subtense (MLS) variable.

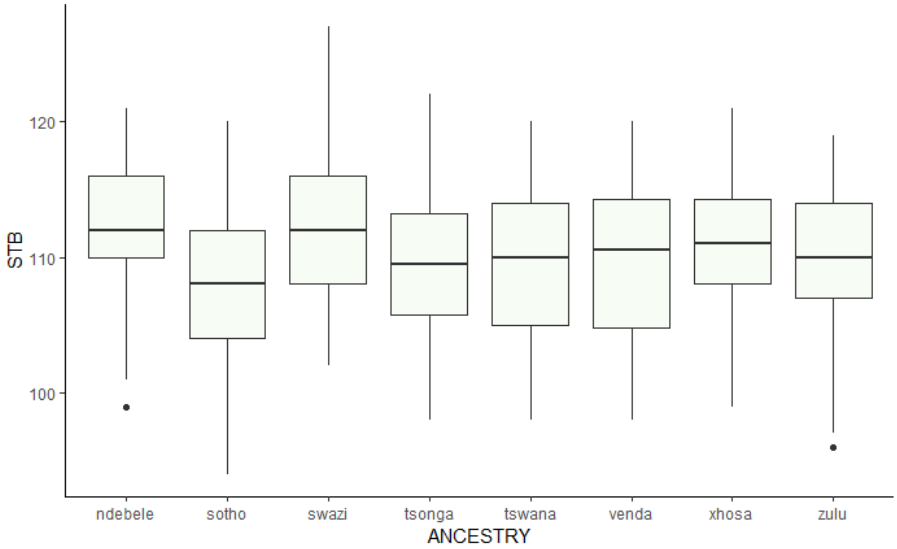


Figure 64. The boxplot of the bistephanic breadth (STB) variable.

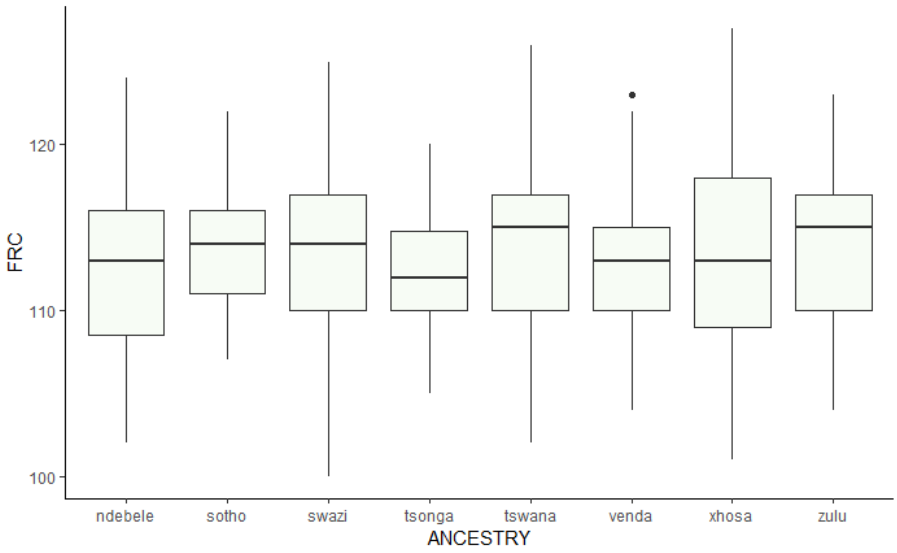


Figure 65. The boxplot of the frontal chord (FRC) variable.

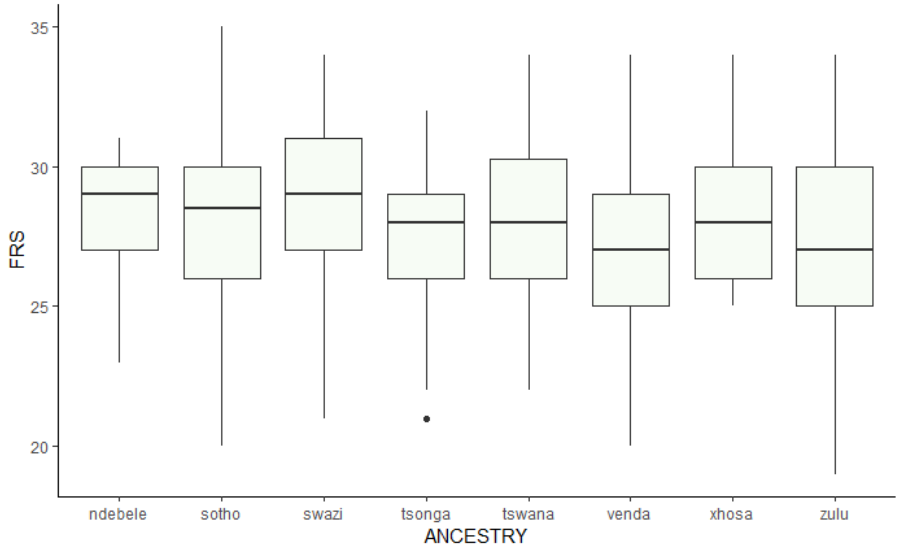


Figure 66. The boxplot of the nasion-bregma subtense (FRS) variable.

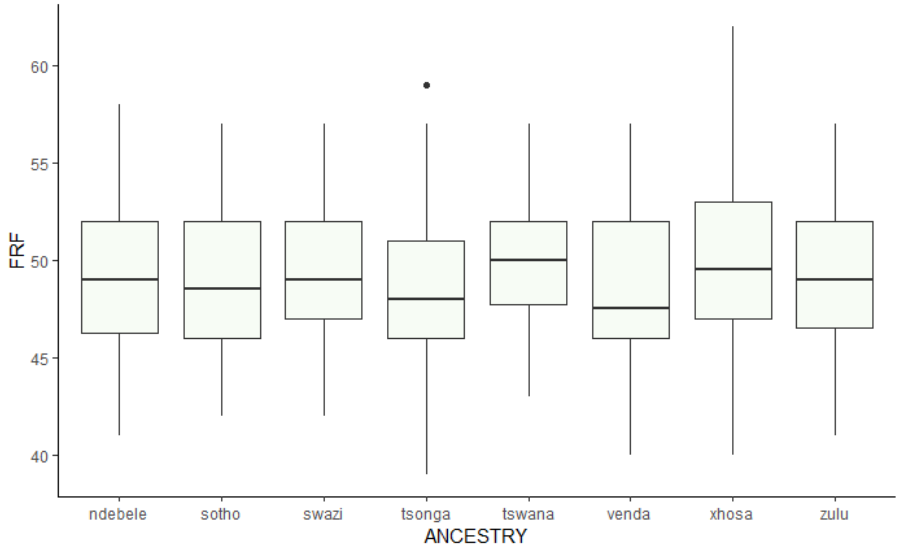


Figure 67. The boxplot of the nasion-subtense fraction (FRF) variable.

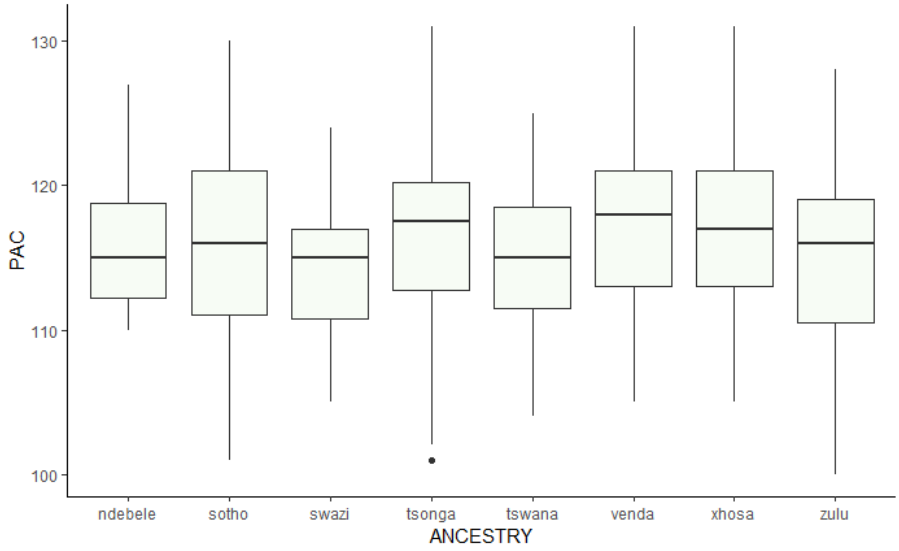


Figure 68. The boxplot of the bregma- lambda chord (PAC) variable.

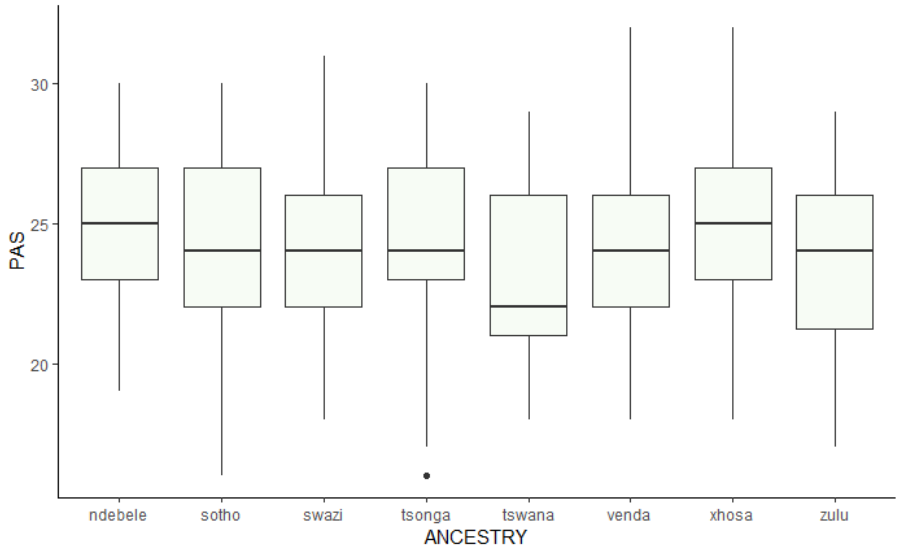


Figure 69. The boxplot of bregma-lambda subtense (PAS) variable.

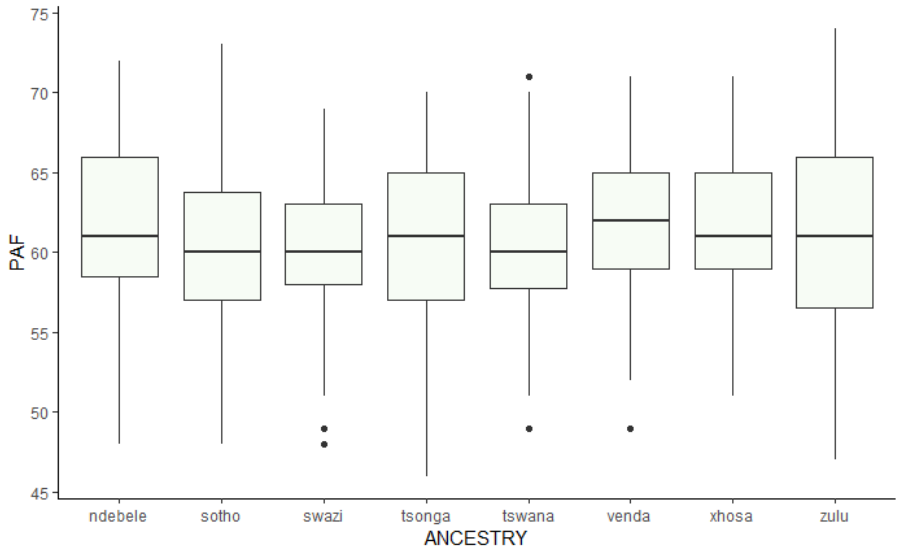


Figure 70. The boxplot of the bregma subtense fraction (PAF) variable.

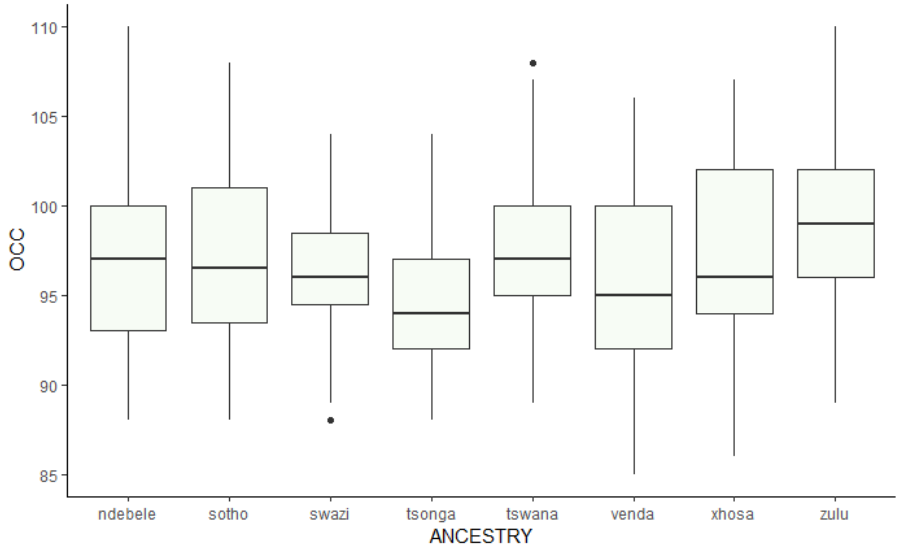


Figure 71. The boxplot of the lambda-opisthion chord (OCC) variable.

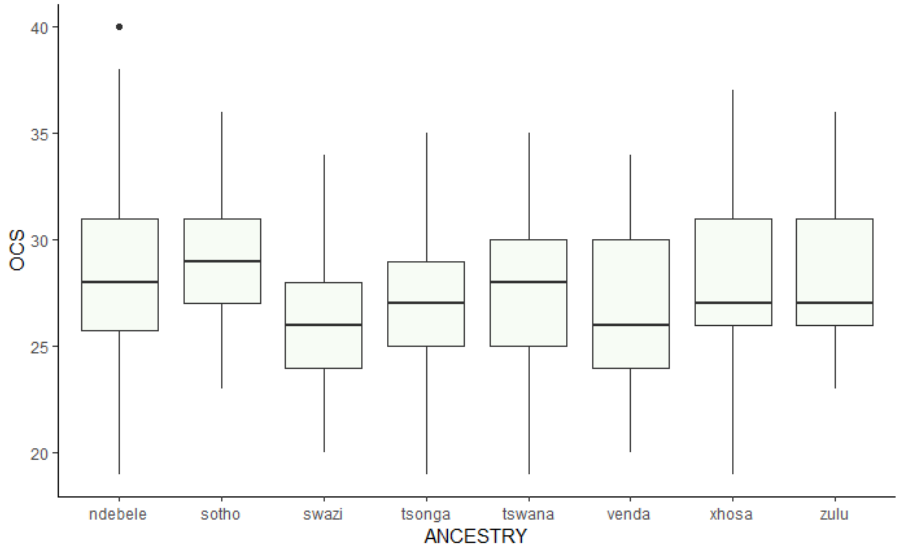


Figure 72. The boxplot of the lambda-opisthion subtense (OCS) variable.

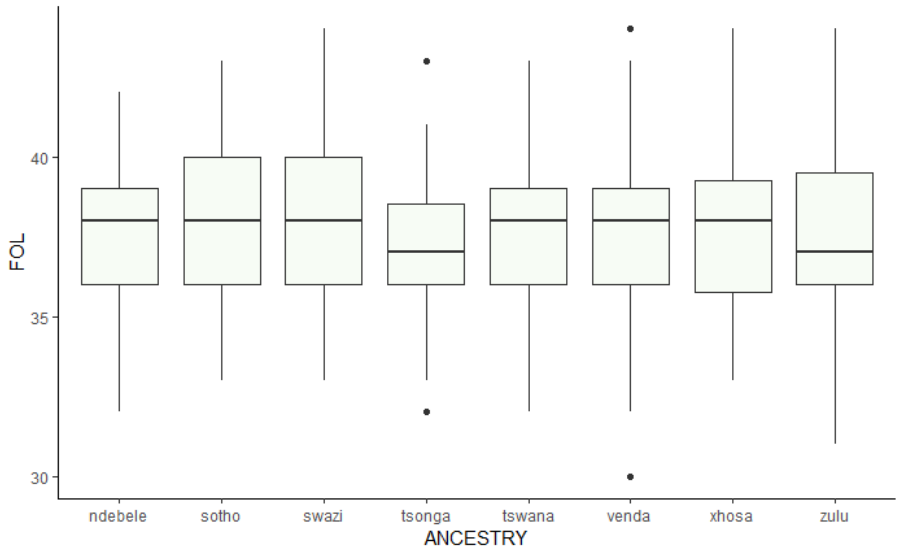


Figure 73. The boxplot of the foramen magnum length (FOL) variable.

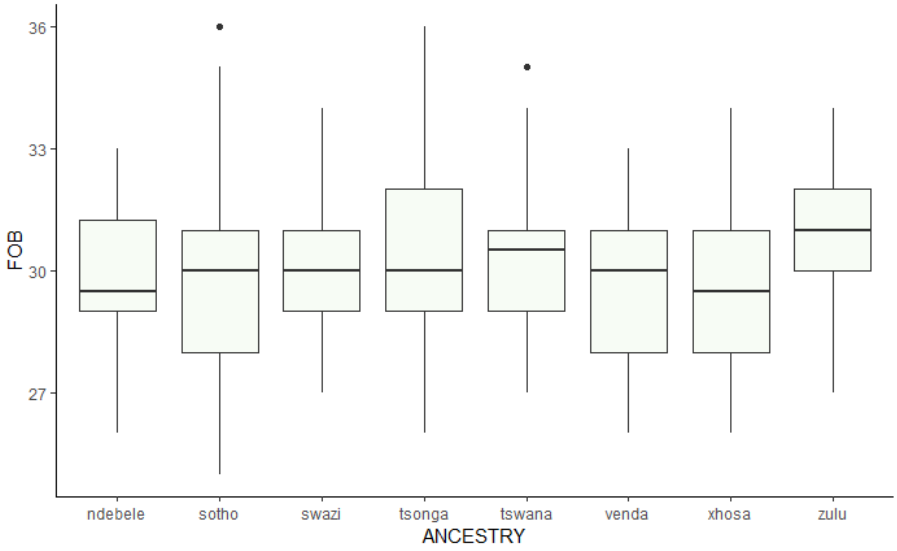


Figure 74. The boxplot of the foramen magnum breadth (FOB) variable.

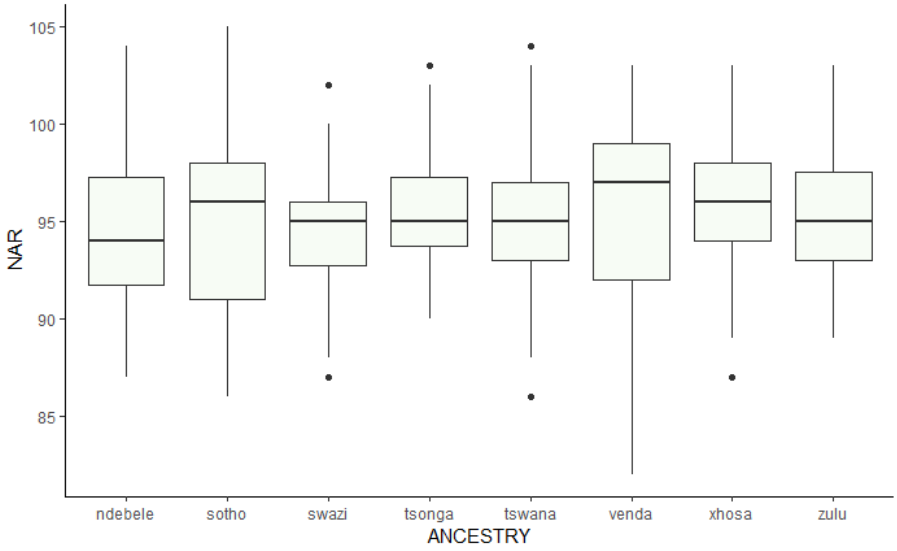


Figure 75. The boxplot of the nasion radius (NAR) variable.

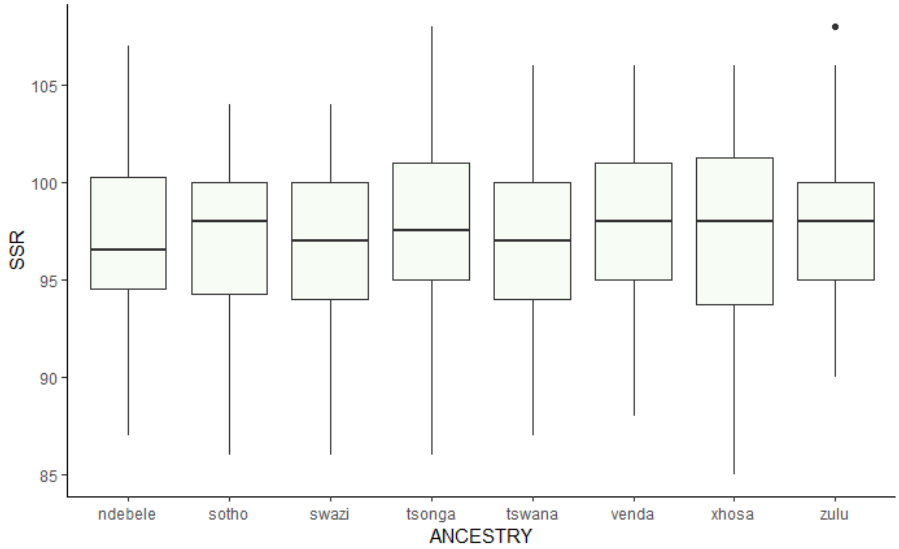


Figure 76. The boxplot of the subspinale radius (SSR) variable.

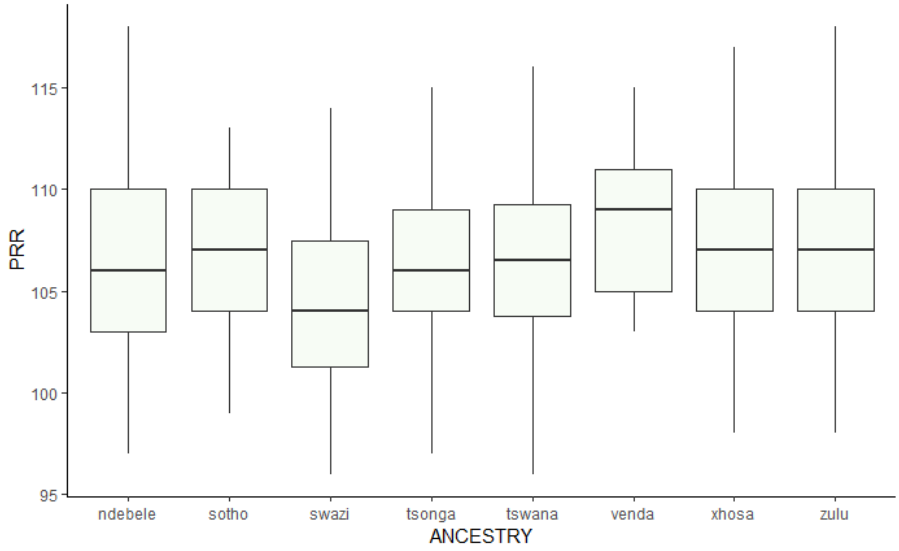


Figure 77. The boxplot of the prosthion radius (PRR) variable.

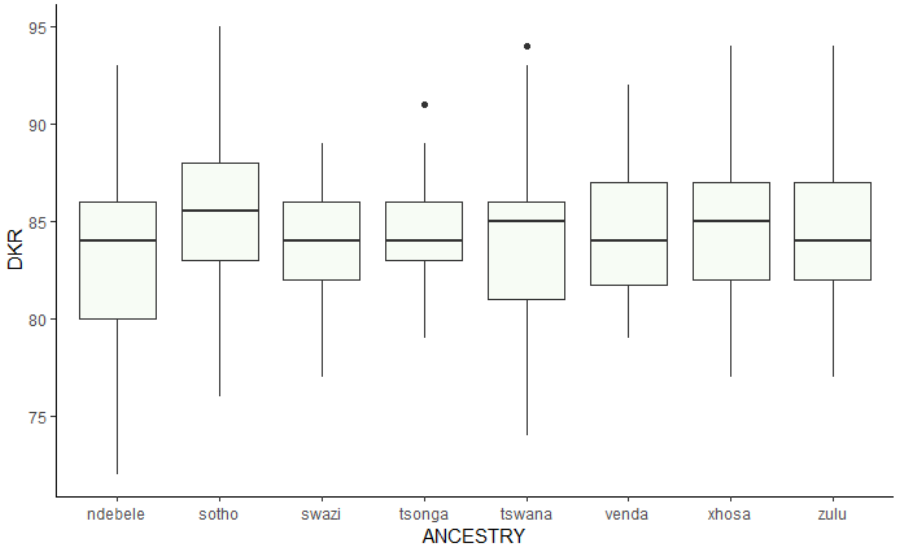


Figure 78. The boxplot of the dacryon radius (DKR) variable.

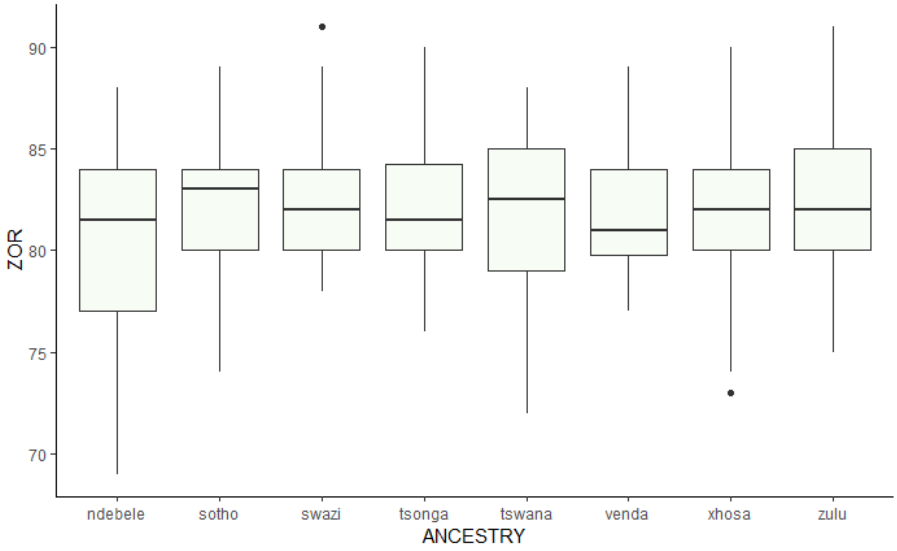


Figure 79. The boxplot of the zygoorbitale radius (ZOR) variable.

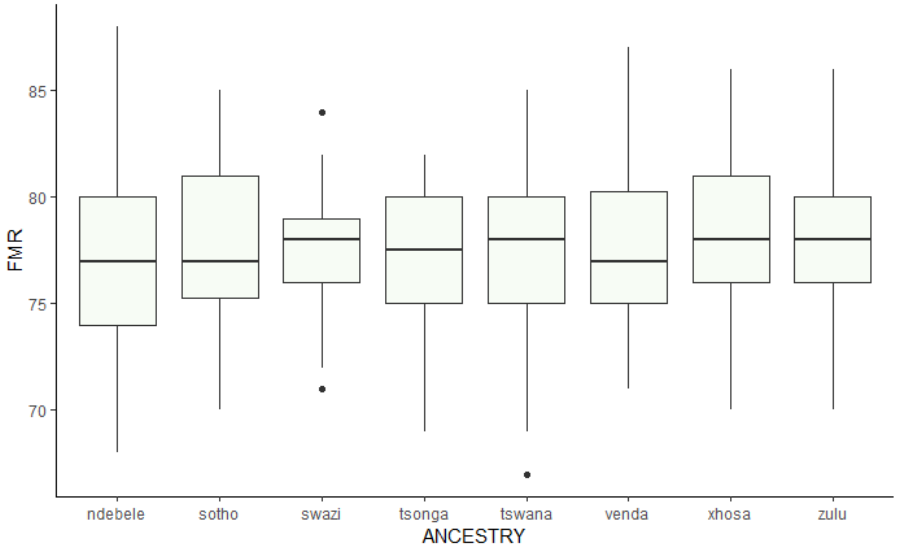


Figure 80. The boxplot of the frontomalare radius (FMR) variable.

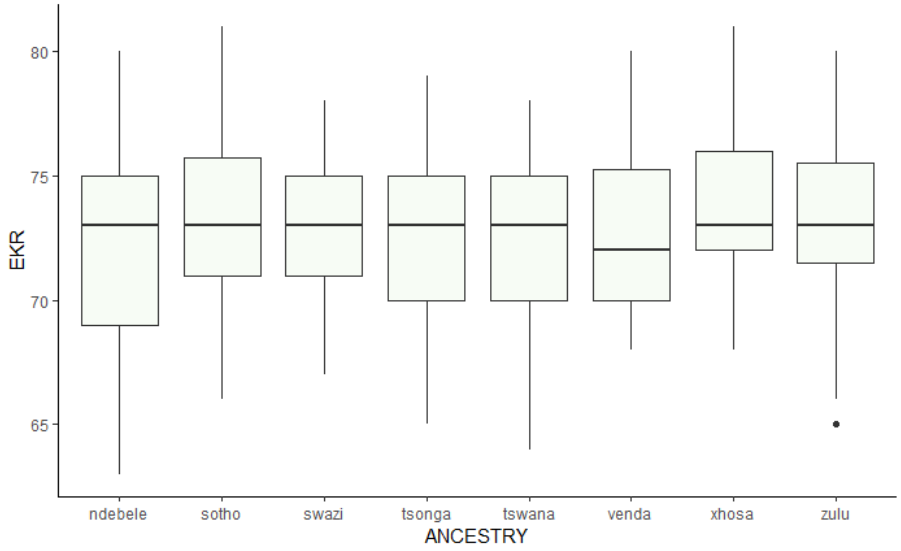


Figure 81. The boxplot of the ectoconchion radius (EKR) variable.

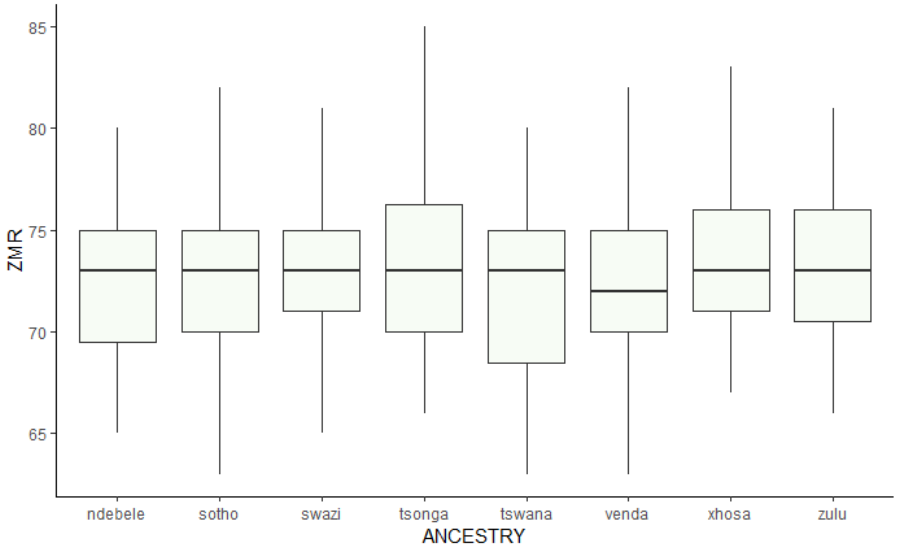


Figure 82. The boxplot of the zygomaxillare radius (ZMR) variable.

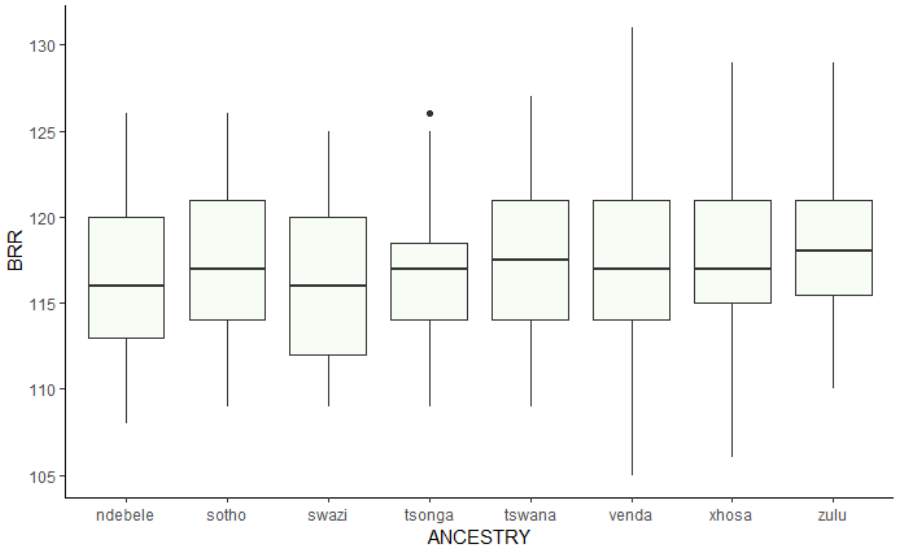


Figure 83. The boxplot of the bregma radius (BRR) variable.

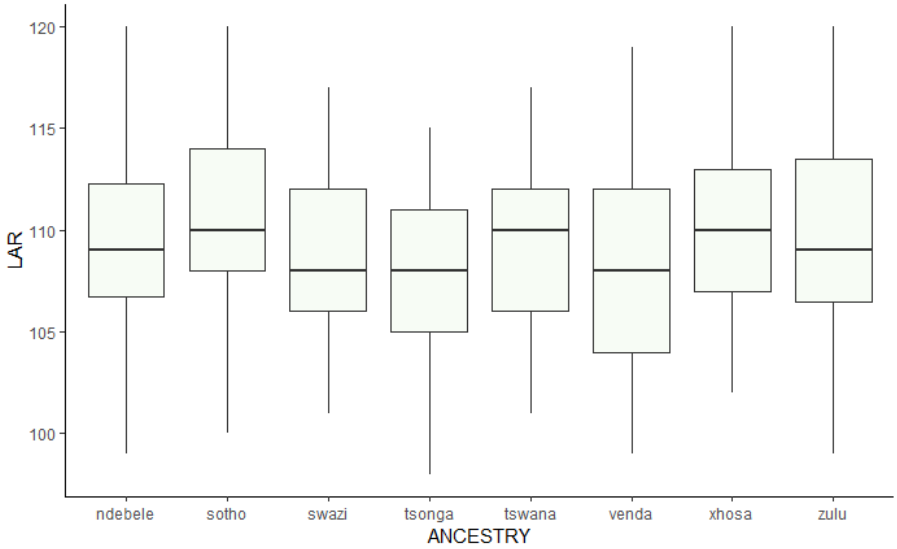


Figure 84. The boxplot of the lambda radius (LAR) variable.

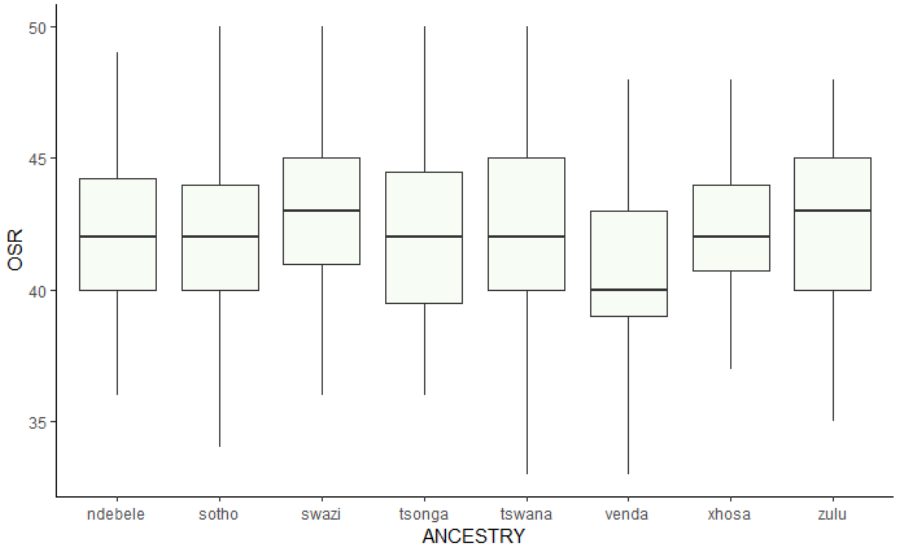


Figure 85. The boxplot of the opisthion radius (OSR) variable.

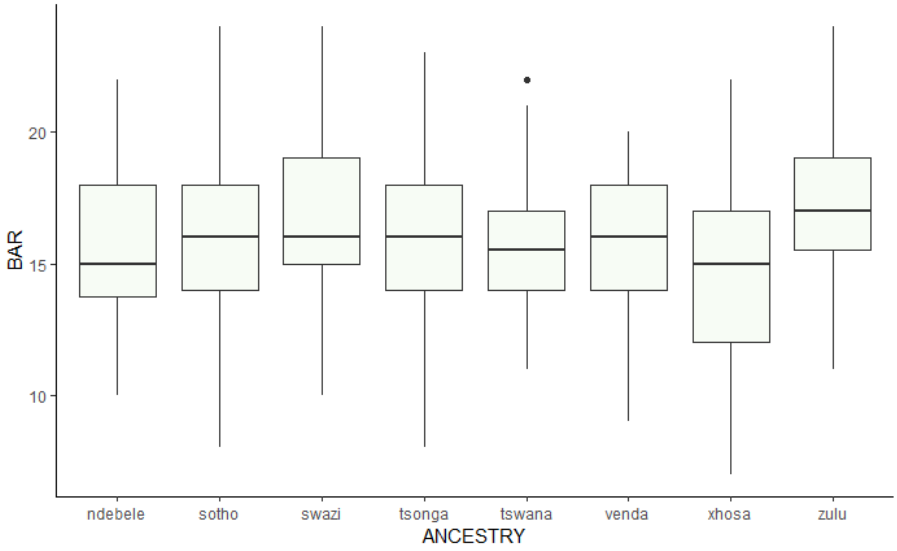


Figure 86. The boxplot of the basion radius (BAR) variable.

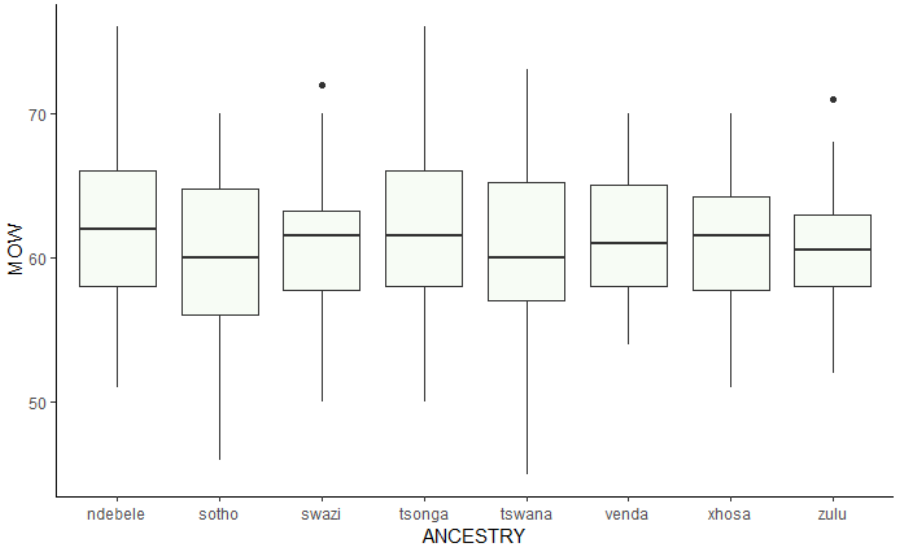


Figure 87. The boxplot of the midorbital width (MOW) variable.

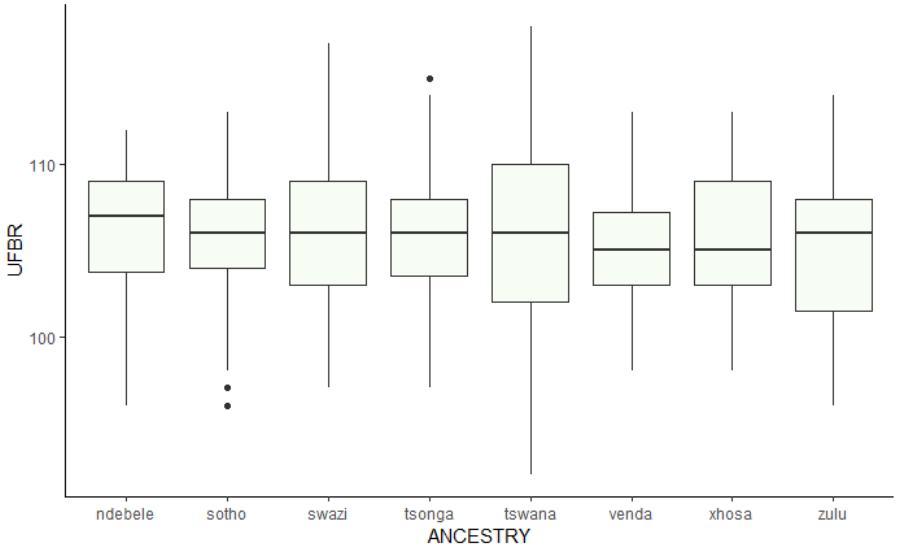


Figure 88. The boxplot of the upper facial breadth (UFBR) variable.

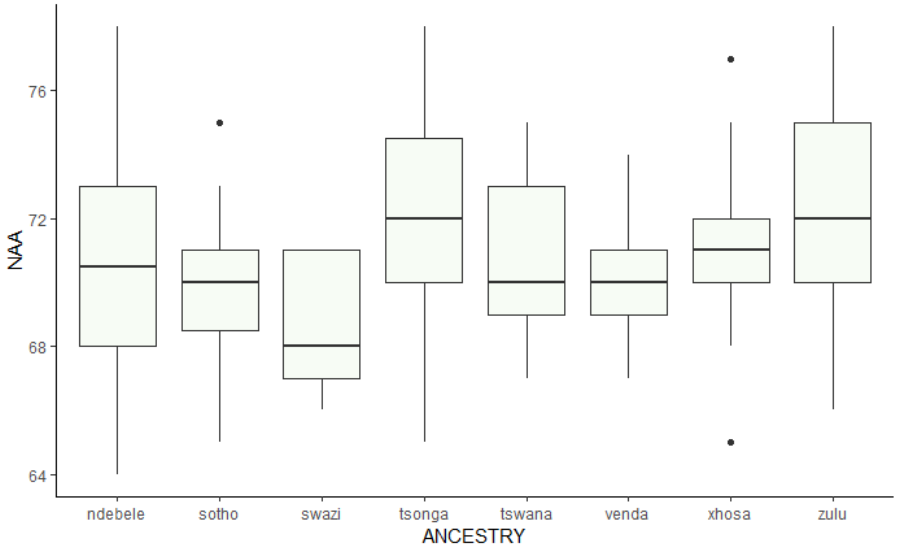


Figure 89. The boxplot of the nasion angle,ba-pr (NAA) variable.

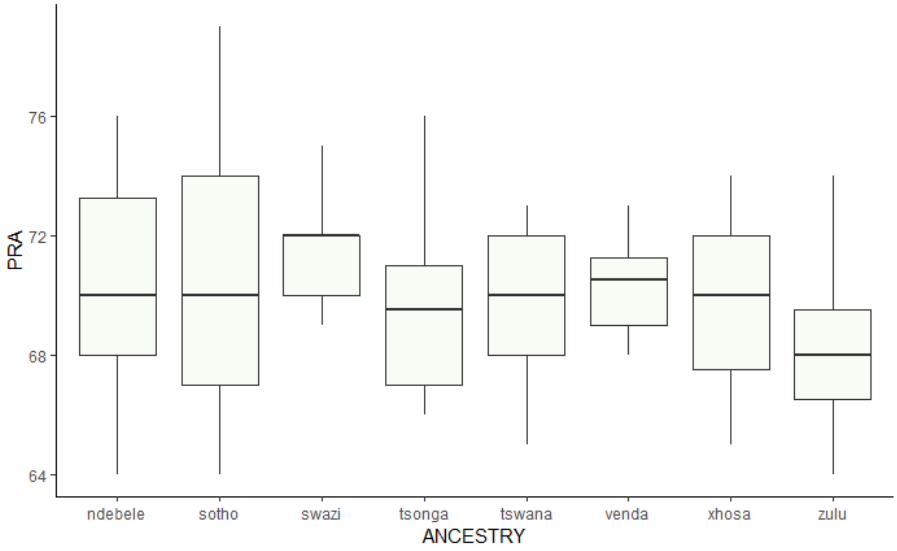


Figure 90. The boxplot of the prosthion angle (PRA) variable.

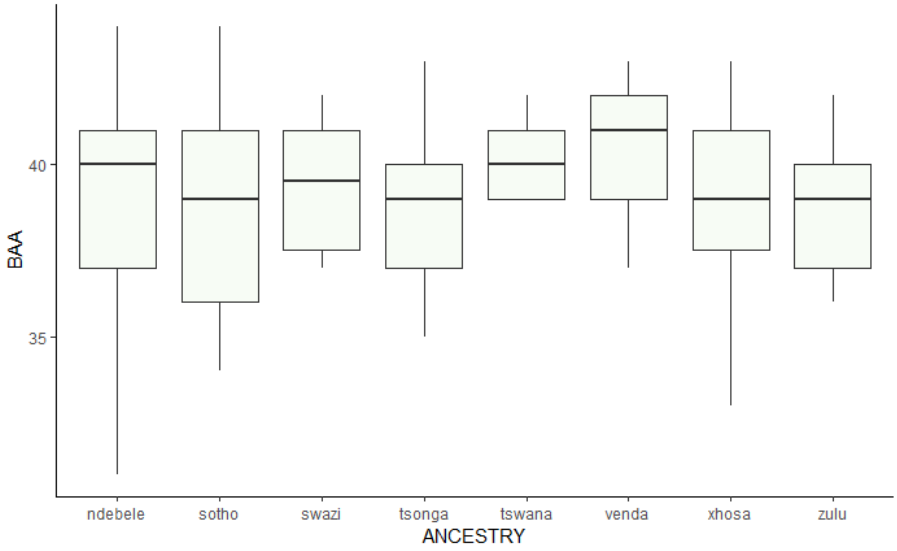


Figure 91. The boxplot of the basion angle,na- pr (BAA) variable.

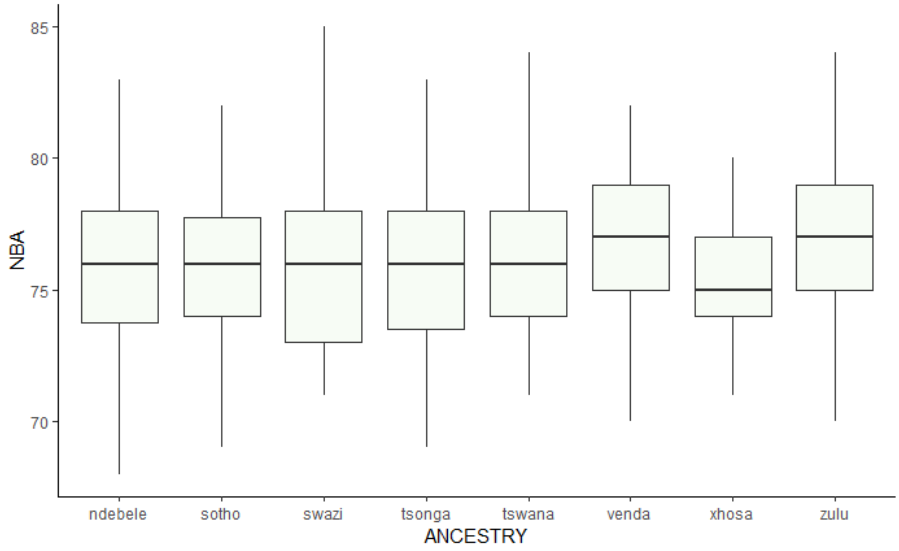


Figure 92. The boxplot of the nasion angle,ba-br (NBA) variable.

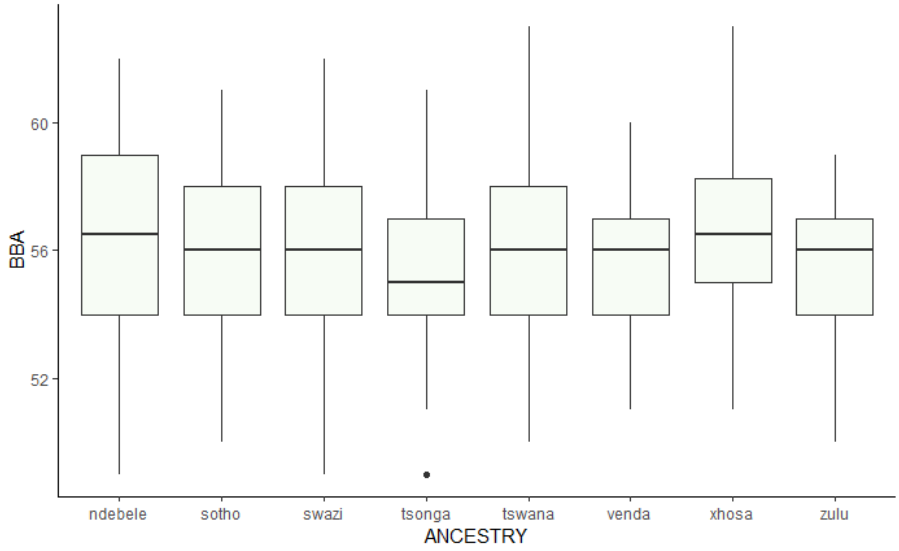


Figure 93. The boxplot of the basion angle,na-br (BBA) variable.

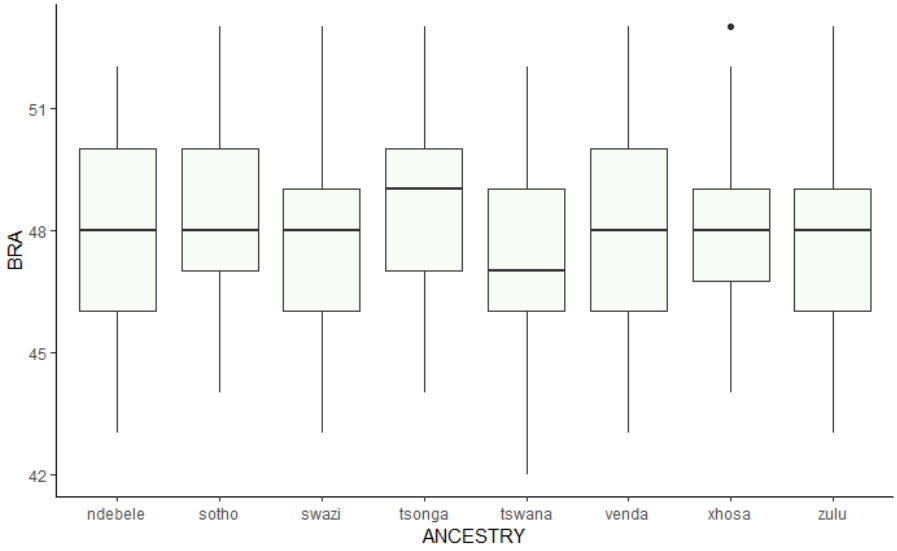


Figure 94. The boxplot of the bregma angle (BRA) variable.

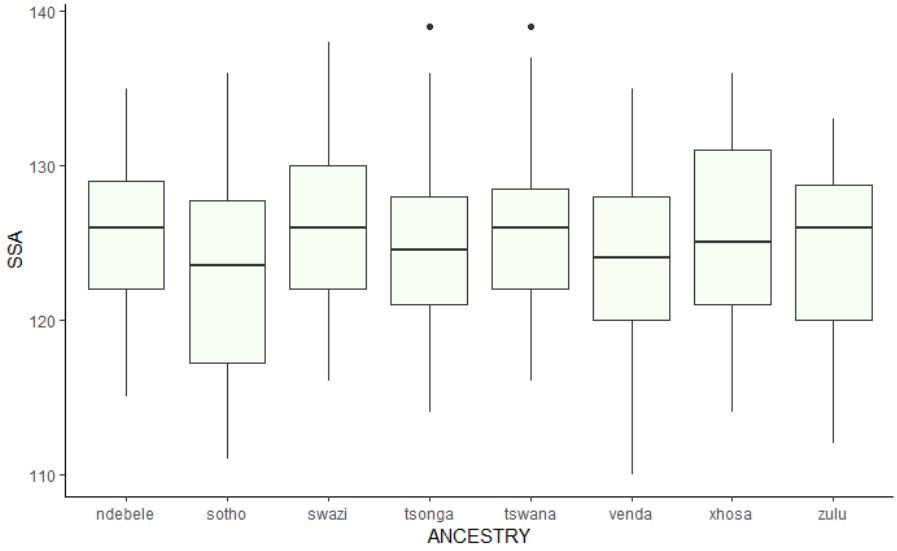


Figure 95. The boxplot of the zygomaxillare angle (SSA) variable.

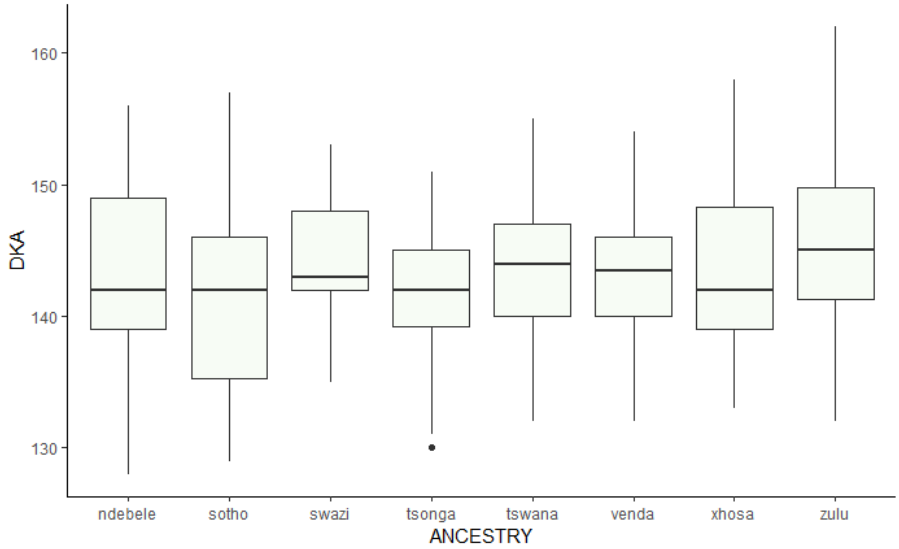


Figure 96. The boxplot of the dacryal angle (DKA) variable.

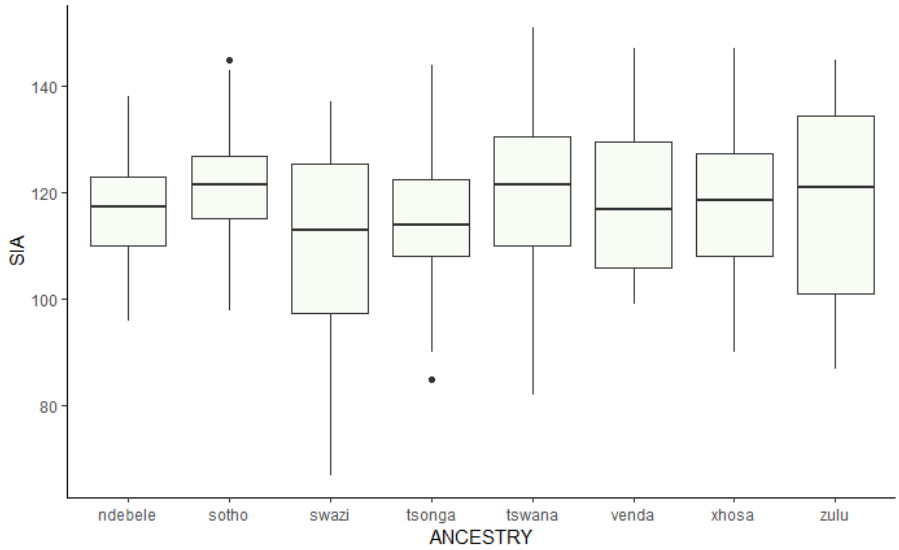


Figure 97. The boxplot of the simotic angle (SIA) variable.

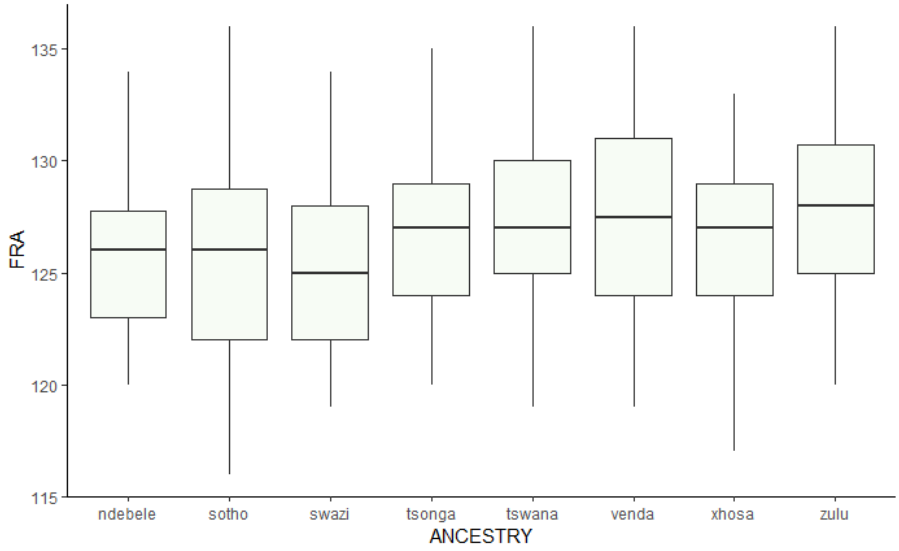


Figure 98. The boxplot of the frontal angle (FRA) variable.

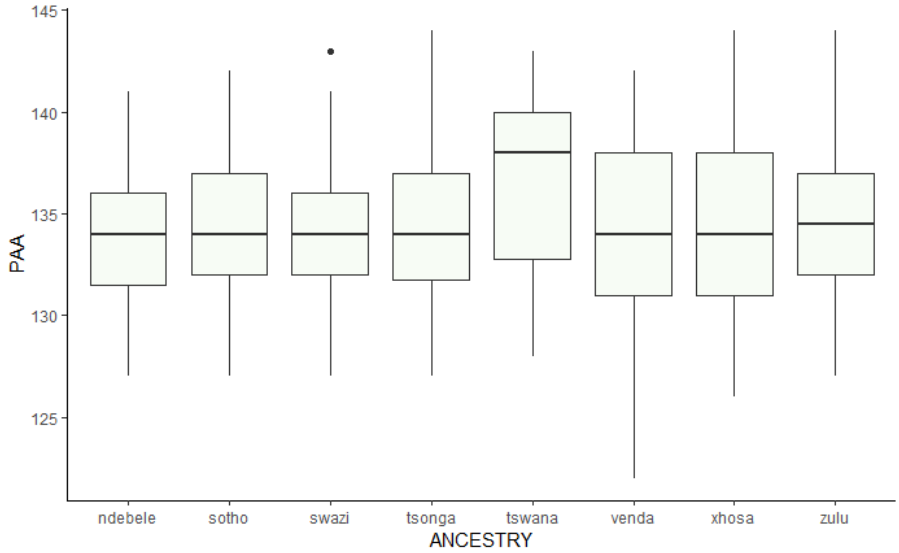


Figure 99. The boxplot of the parietal angle (PAA) variable.

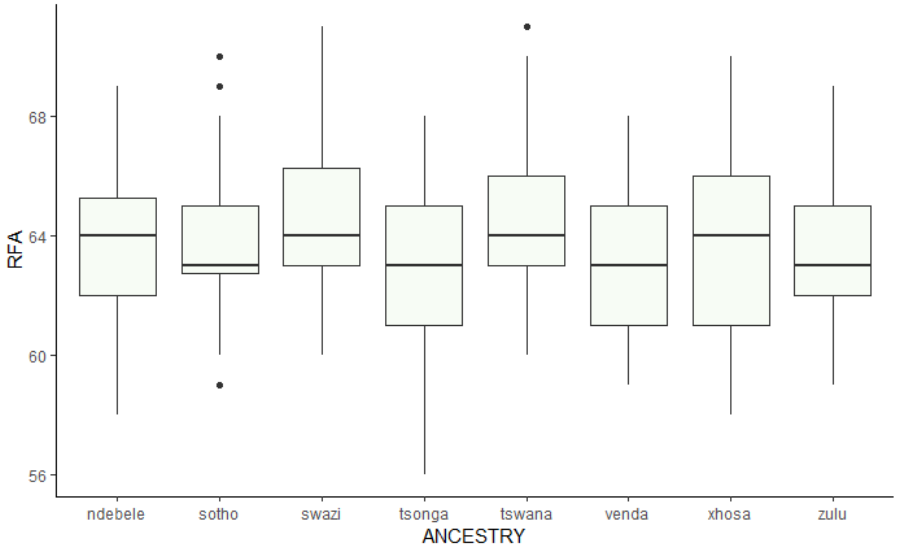


Figure 100. The boxplot of the radio-frontal (RFA) variable.

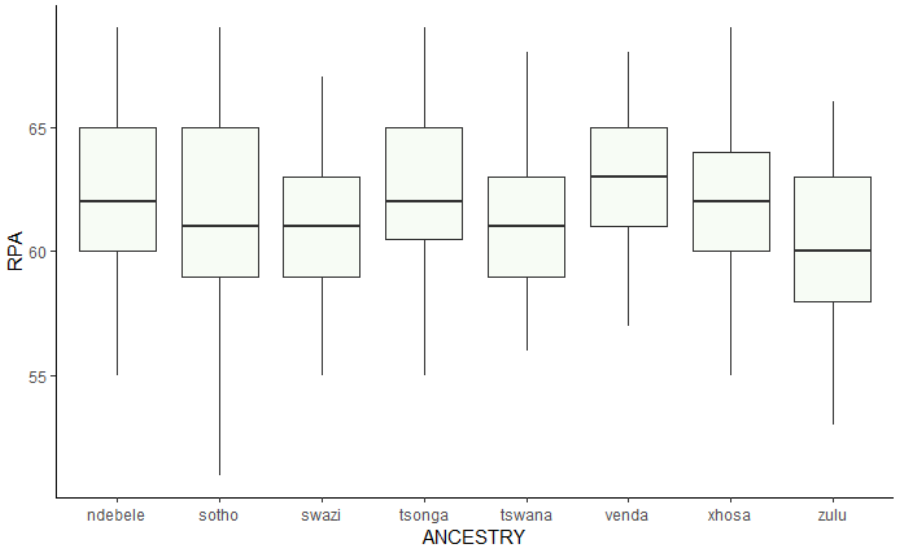


Figure 101. The boxplot of the radio-parietal angle (RPA) variable.

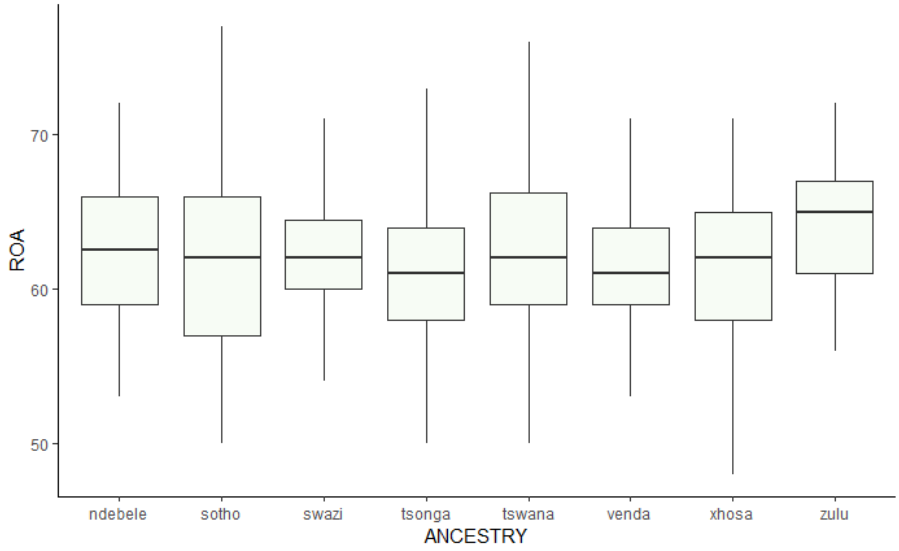


Figure 102. The boxplot of the radio-occipital angle (ROA) variable.

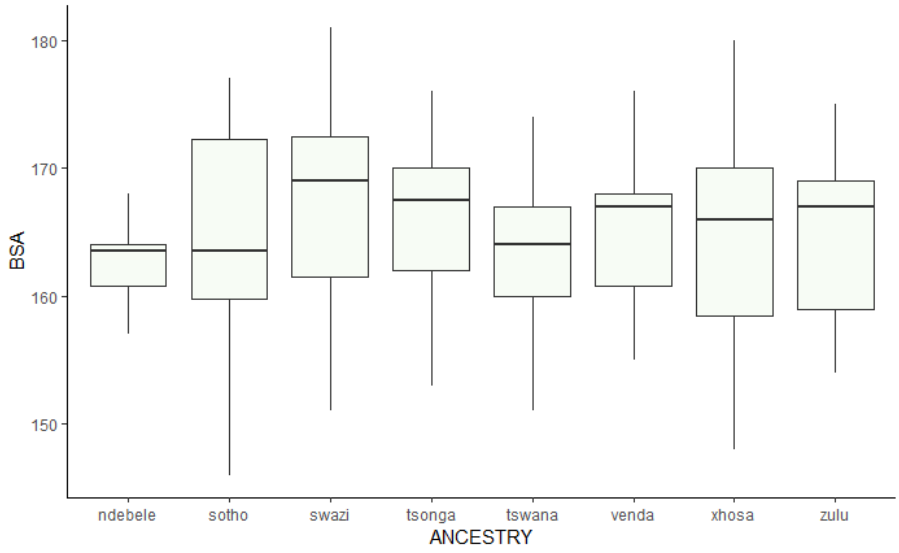


Figure 103. The boxplot of the basal angle (BSA) variable.

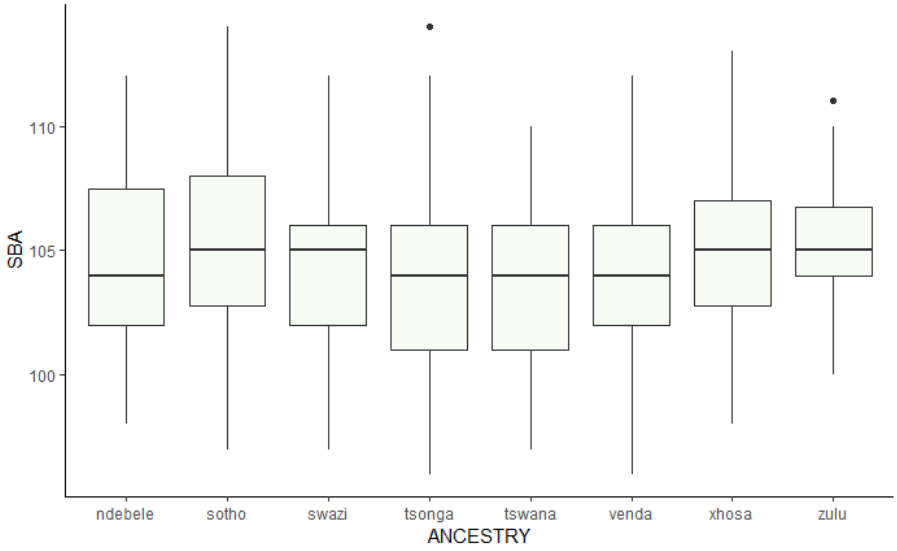


Figure 104. The boxplot of the sub-bregma (SBA) variable.

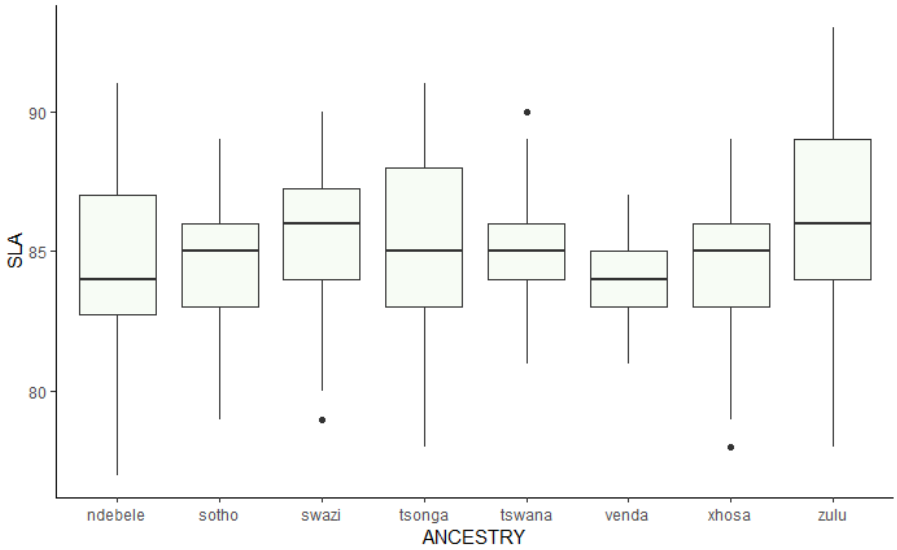


Figure 105. The boxplot of the sub-lambda angle (SLA) variable.

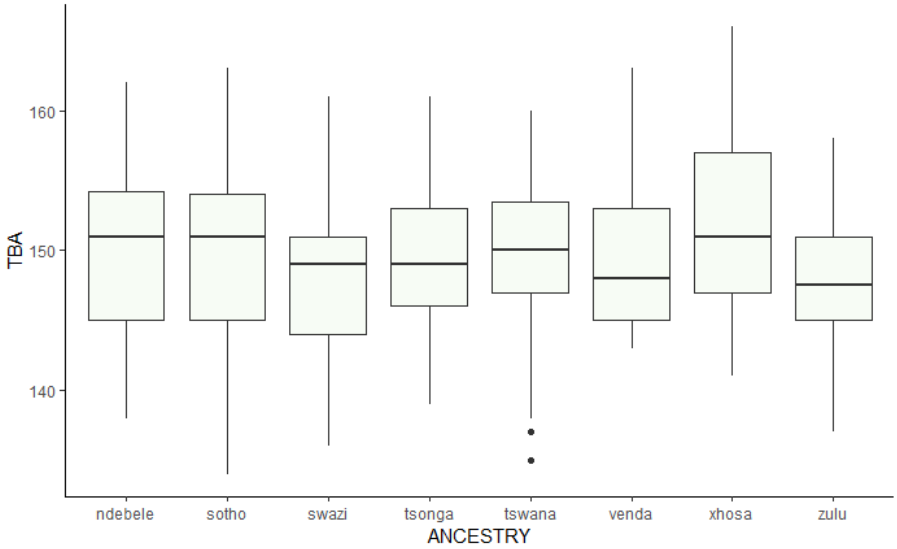


Figure 106. The boxplot of the trans-basal angle (TBA) variable.

8.3 Correlation Table

The correlation values of the different inter-variable relationships. Inter-variable relationships that reflect multicollinearity are bolded (Table 43).

Table 43. Correlation Table of the inter-variable relationships

Variables	Correlation Coeff.	Variables	Correlation Coeff.
NLH-OBH	0.412	JUB-NLH	0.220
ZYB-NLH	0.276	JUB-ZMB	0.614
ZYB-ZMB	0.470	JUB-ZYB	0.803
AUB-NLH	0.259	JUB-AUB	0.518
AUB-ZMB	0.394	JUB-NLB	0.433
AUB-ZYB	0.691	JUB-WMH	0.281
NPH-OBH	0.320	OBB-ZMB	0.242
NPH-NLH	0.641	OBB-ZYB	0.397
NPH-ZYB	0.154	OBB-AUB	0.239
NPH-AUB	0.289	OBB-NLB	0.274
NLB-ZMB	0.348	OBB-JUB	0.480
NLB-ZYB	0.335	EKB-ZMB	0.538
NLB-AUB	0.239	EKB-ZYB	0.637
WMH-NLH	0.243	EKB-AUB	0.440
WMH-ZMB	0.212	EKB-NLB	0.469
WMH-ZYB	0.219	EKB-WMH	0.175
WMH-AUB	0.206	EKB-JUB	0.809
WMH-NPH	0.298	EKB-OBB	0.641
FMB-ZMB	0.501	UFBR-ZMB	0.470
FMB-ZYB	0.629	UFBR- ZYB	0.632
FMB-AUB	0.435	UFBR- AUB	0.430
FMB-NLB	0.441	UFBR-NLH	0.176
FMB-WMH	0.211	UFBR-WMH	0.262
FMB-JUB	0.775	UFBR-JUB	0.766

FMB-OBB	0.591	UFBR-OBB	0.563
FMB-EKB	**0.928**	UFBR-EKB	0.870
MLS-SSS	-0.126	UFBR-FMB	**0.923**
MLS-ZMB	0.225	MLS-OBB	0.165
MLS-ZYB	0.147	MLS-EKB	0.280
MLS-NLB	0.174	MLS-FMB	0.236
MLS-WMH	0.225	MLS-UFBR	0.280
MLS-JUB	0.288	DBK-ZMB	0.427
DBK-JUB	0.545	DBK-ZYB	0.415
DBK-EKB	0.641	DBK-AUB	0.291
DBK-FMB	0.632	DBK-NLB	0.335
DBK-UFBR	0.598	BPL-SSS	0.326
MAL-SSS	0.398	ASB-SSS	-0.092
MAL-NPH	0.247	ASB-ZYB	0.276
MAL-BPL	0.670	ASB-AUB	0.411
ASB-FMB	0.254	ASB-WMH	0.157
ASB-UFBR	0.264	ASB-JUB	0.234
ASB-DKB	0.083	ASB-EKB	0.224
WFB-BPL	0.098	WFB-MAL	0.081
WFB-SSS	-0.031	WFB-ZYB	0.430
WFB-AUB	0.394	WFB-NLB	0.238
WFB-JUB	0.483	WFB-EKB	0.592
WFB-FMB	0.619	WFB-UFBR	0.679
WFB-MLS	0.189	WFB-DKB	0.482
WFB-ASB	0.246	ZOR-NLB	0.204
ZOR-WMH	0.385	ZOR-JUB	0.388
ZOR-EKB	0.369	ZOR-UFBR	0.388
ZOR-MLS	0.285	ZOR-BPL	0.643
ZOR-MAL	0.414	BNL-BPL	0.695
BNL-MAL	0.448	BNL-ZOR	0.609
IML-SSS	-0.189	IML-MLS	0.337
IML-WFB	0.221	IML-ZOR	0.369
DKS-SSS	0.272	DKS-ZYB	0.024
DKS-AUB	-0.014	DKS-WMH	-0.068
DKS-OBB	0.506	DKS-ASB	-0.089
GOL-ASB	0.398	GOL-ZOR	0.470
NOL-ASB	0.375	NOL-ZOR	0.475
NOL-GOL	**0.980**	VRR-SSS	-0.029
VRR-MLS	0.261	VRR-ASB	0.290
VRR-GOL	0.469	VRR-NOL	0.476
NAS-SSS	0.305	NAS-NLH	0.092
NAS-ZYB	0.084	NAS-AUB	0.056
NAS-WMH	0.083	NAS-ASB	-0.001
NAS-BNL	0.431	NAS-DKS	0.555
XFB-SSS	-0.129	XFB-BPL	-0.073
XFB-ASB	0.245	XFB-WFB	0.453
XFB-VRR	0.413	XFB-NAS	-0.143
XCB-SSS	-0.047	XCB-ZMB	0.025
XCB-VRR	0.380	XCB-XFB	0.554
STB-SSS	-0.114	STB-BPL	-0.102
STB-ASB	0.141	STB-WFB	0.418
STB-VRR	0.425	STB-XFB	0.891
STB-XCB	0.464	BBH-OBH	-0.084

BBH-NLH	0.011	BBH-BNL	0.399
BBH-GOL	0.251	BBH-NOL	0.254
BBH-VRR	0.654	BBH-XFB	0.239
BBH-XCB	0.283	BBH-STB	0.284
WNB-OBH	-0.088	WNB-NLH	0.053
WNB-NPH	-0.034	WNB-WMH	0.000
WNB-DKB	0.389	WNB-NAS	0.323
SIS-OBH	-0.029	SIS-NLH	0.040
SIS-ZMB	0.039	SIS-AUB	0.057
SIS-NLB	0.075	SIS-WMH	0.087
SIS-MLS	-0.067	SIS-NAS	0.206
SIS-WNB	0.619	SIS-JUB	0.110
SIS-NPH	0.043	NDS-NLH	-0.031
NDS-ZMB	0.091	NDS-ZYB	0.076
NDS-AUB	0.100	NDS-NLB	0.003
NDS-WMH	-0.022	NDS-JUB	0.094
NDS-OBB	-0.075	NDS-EKB	0.160
NDS-FMB	0.149	NDS-UFBR	0.149
NDS-MLS	-0.015	NDS-WNB	0.169
NDS-SIS	0.497		

8.4 Tukey's Plots

The Tukey's plot of the statistically significant variables with the statistically significant post-hoc pairwise comparisons are indicated in red (Figure 107-122).

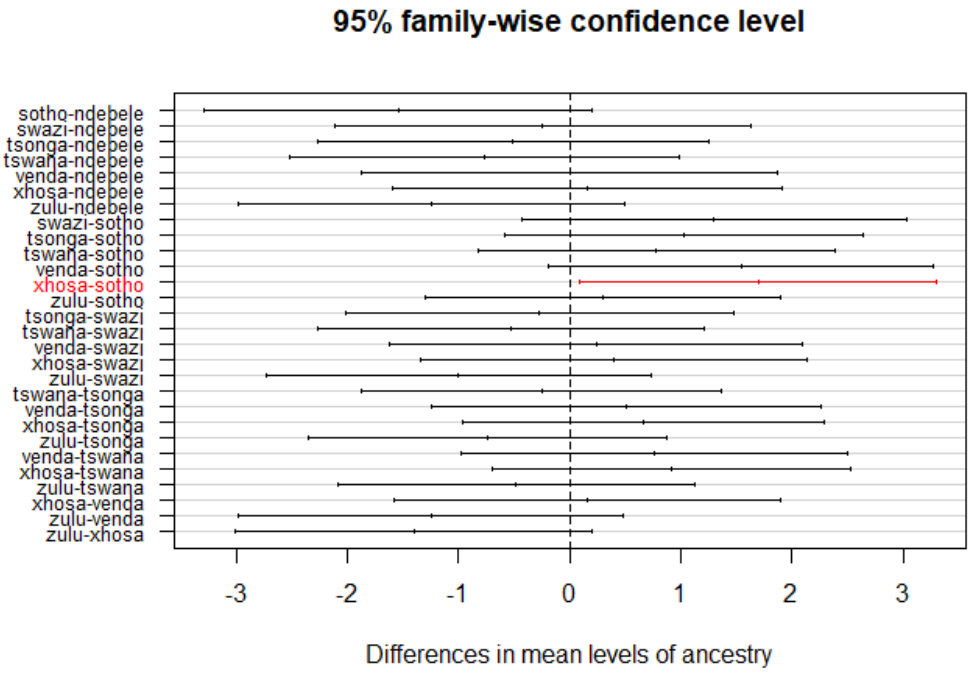


Figure 107. The Tukey's plot of the interorbital breadth (DKB) variable.

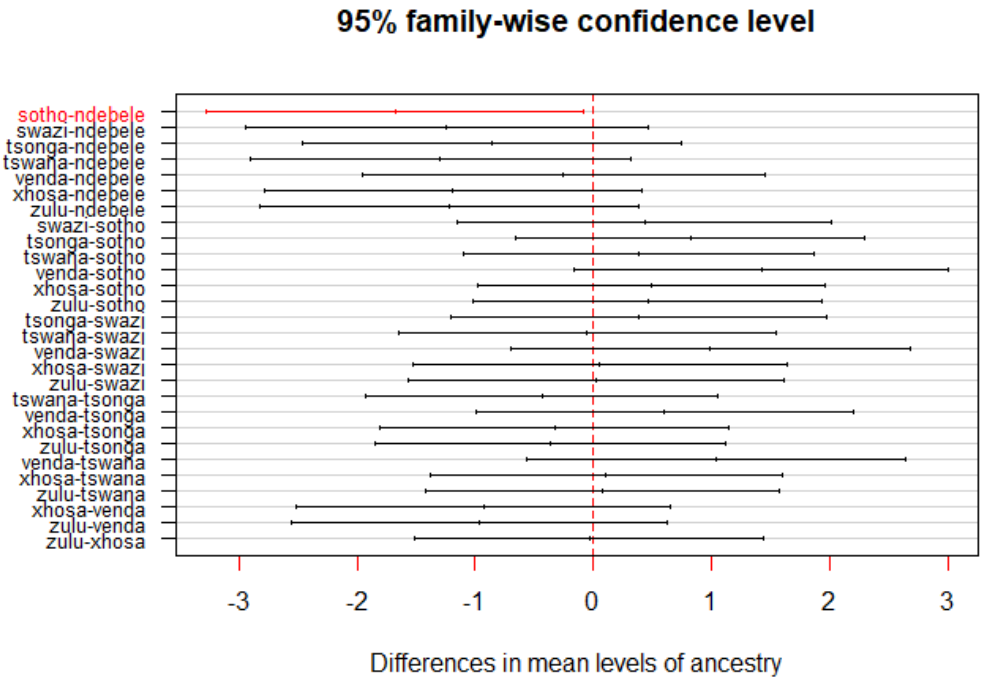


Figure 108. The Tukey's plot of the simotic chord (WNB) variable.

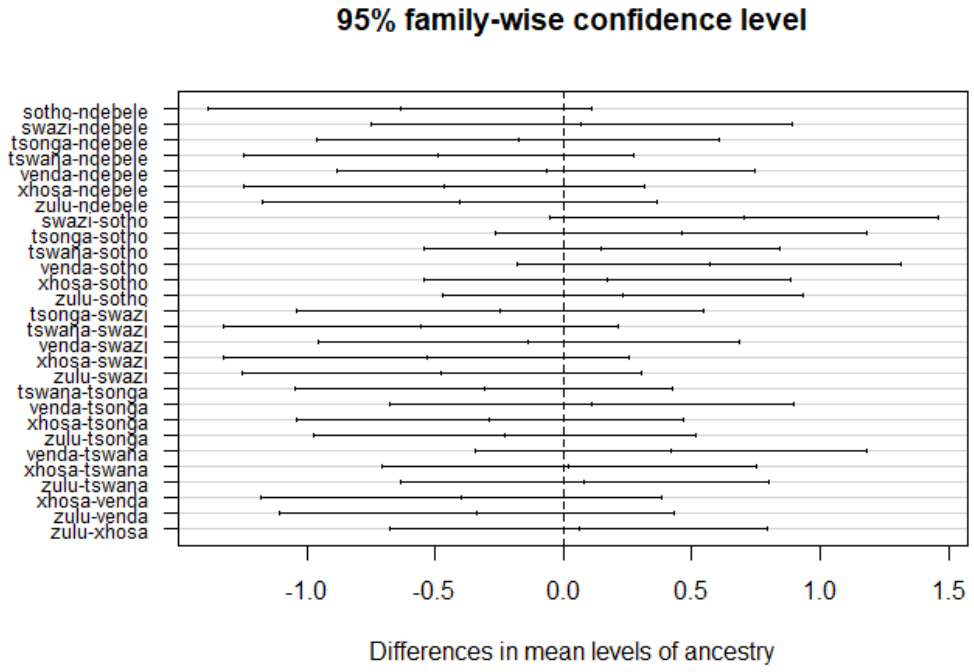


Figure 109. The Tukey's plot of the simotic subtense (SIS) variable.

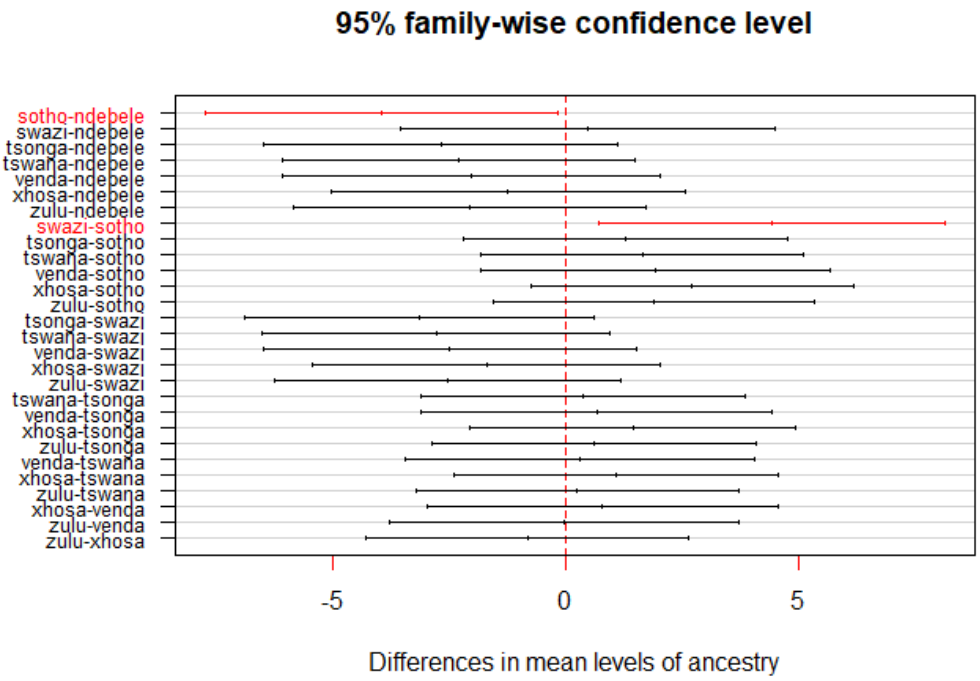


Figure 110. The Tukey's plot of the bistephanic breadth (STB) variable.

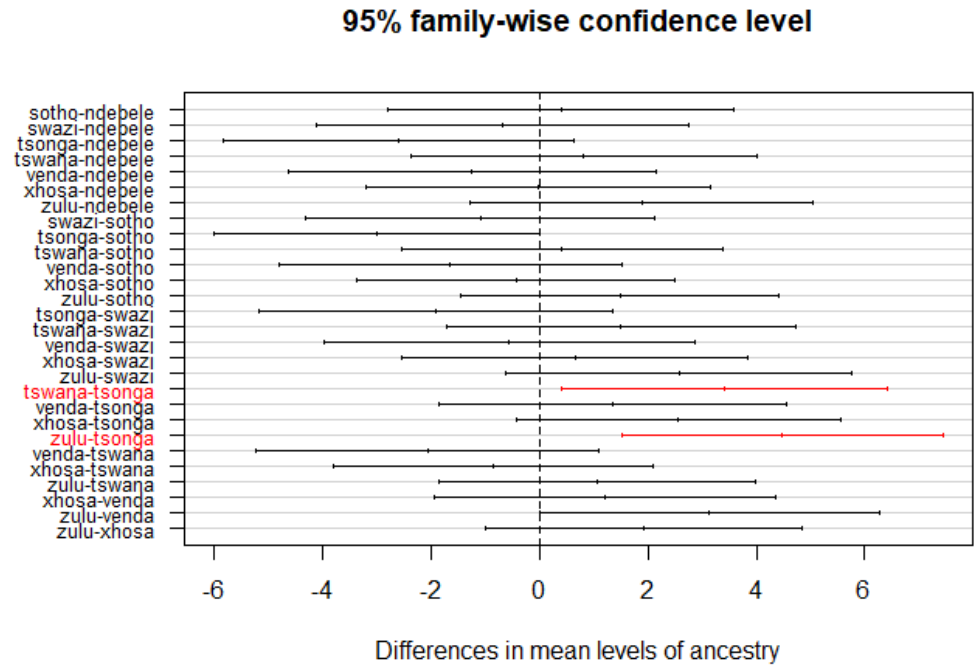


Figure 111. The Tukey's plot of the lambda-opisthion chord (OCC) variable.

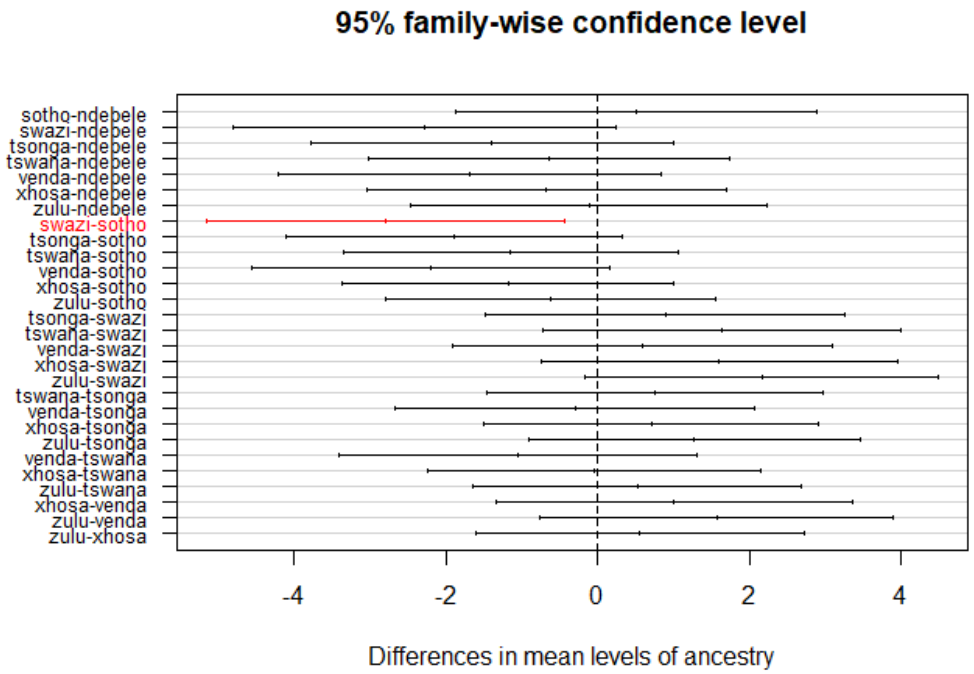


Figure 112. The Tukey's plot of the lambda- opisthion subtense (OCS) variable.

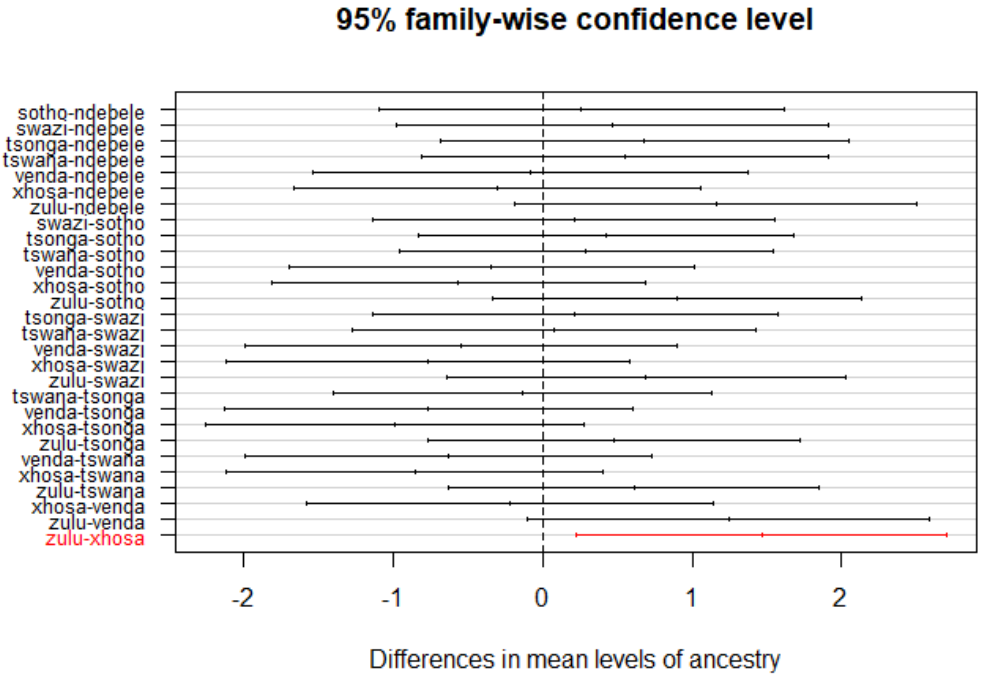


Figure 113. The Tukey's plot of the foramen magnum breadth (FOB) variable.

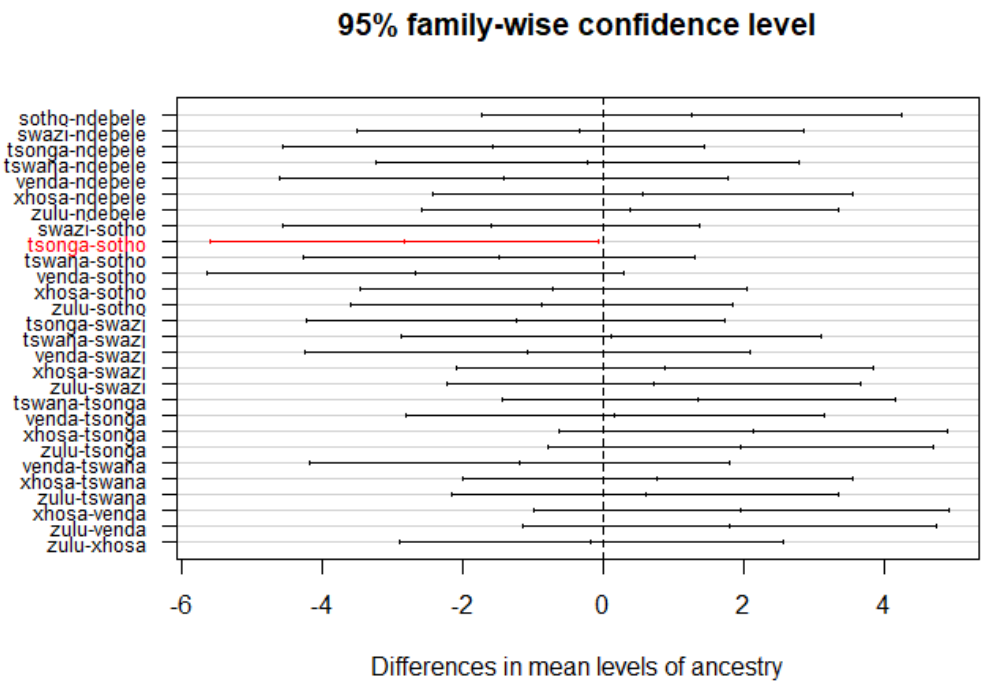


Figure 114. The Tukey's plot of the lambda radius (LAR) variable.

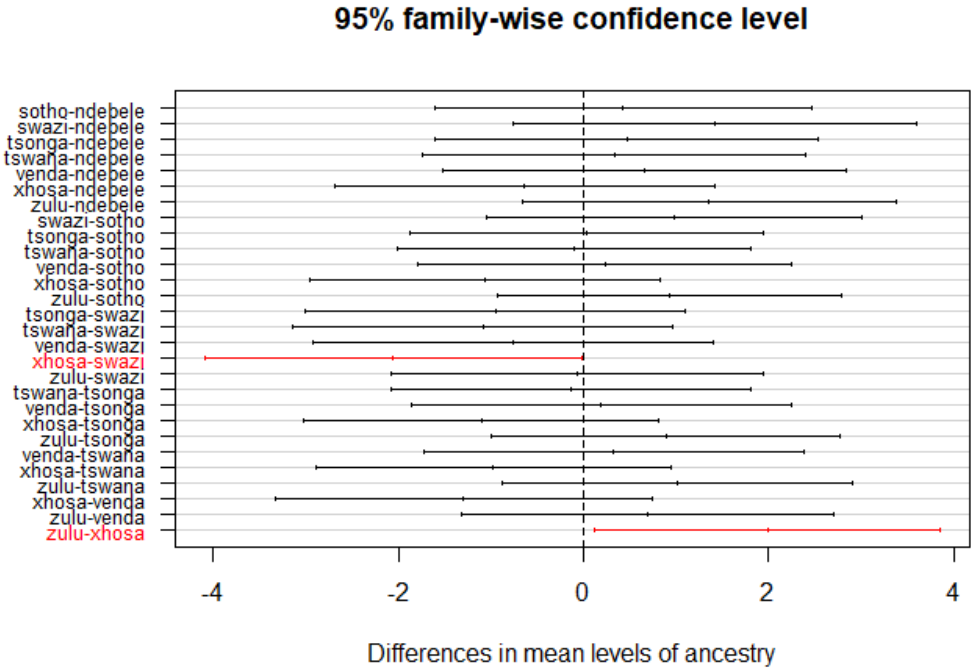


Figure 115. The Tukey's plot of the basion radius (BAR) variable.

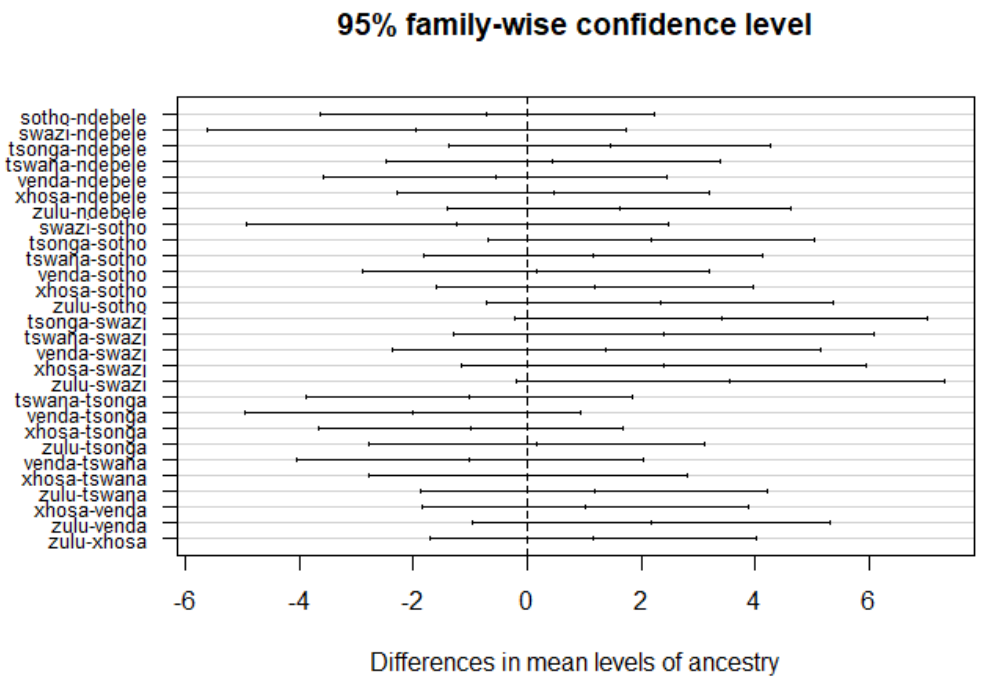


Figure 116. The Tukey's plot of the nasion angle (NAA) variable.

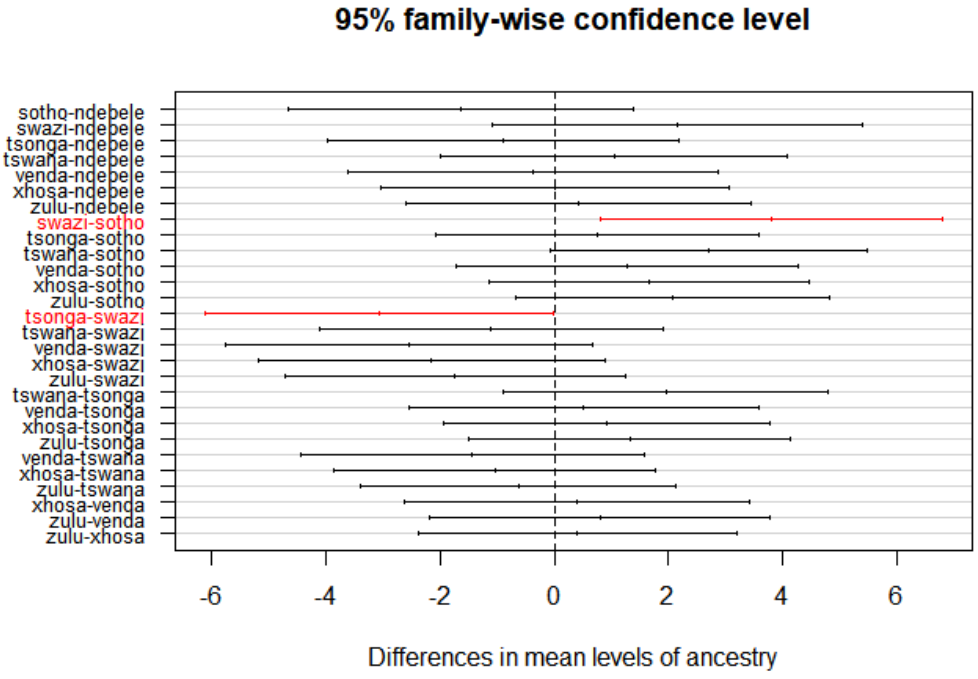


Figure 117. The Tukey's plot of the nasio-frontal angle (NFA) variable.

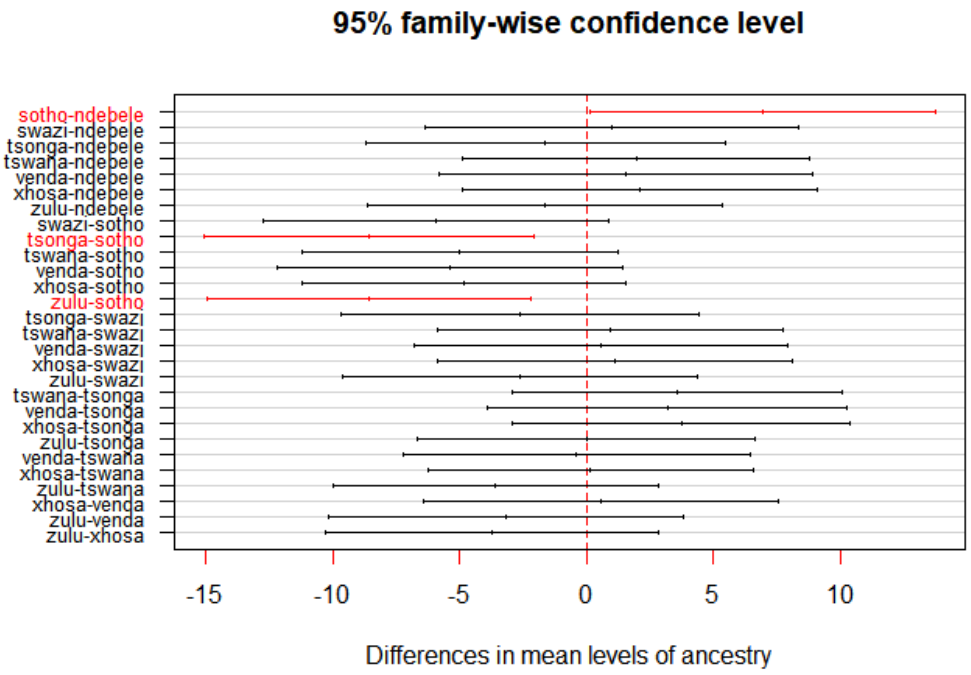


Figure 118. The Tukey's plot of the naso- dacryal angle (NDA) variable.

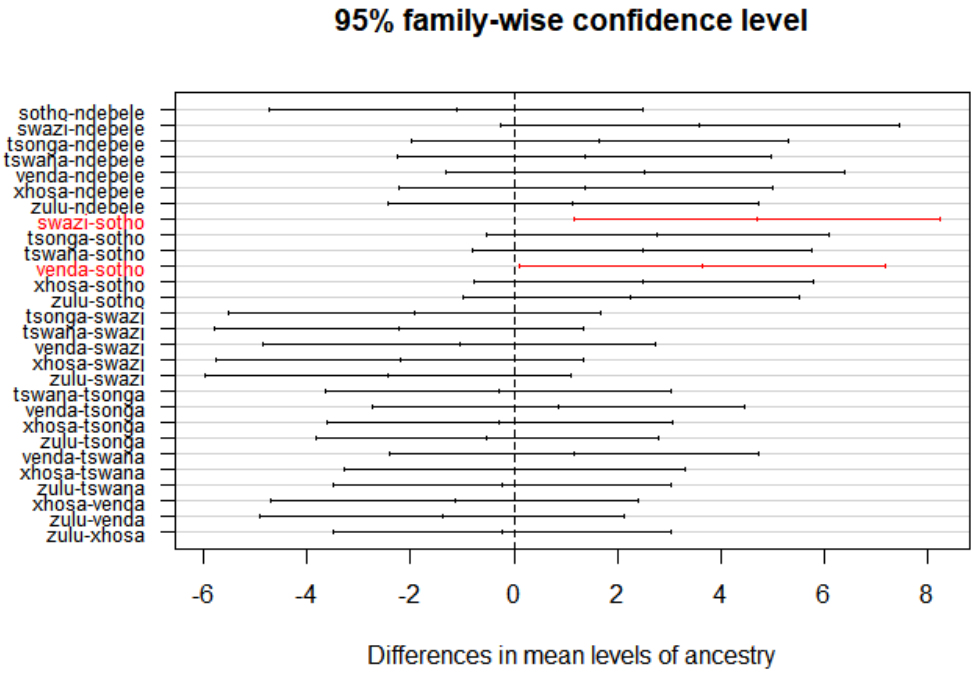


Figure 119. The Tukey's plot of the occipital angle (OCA) variable.

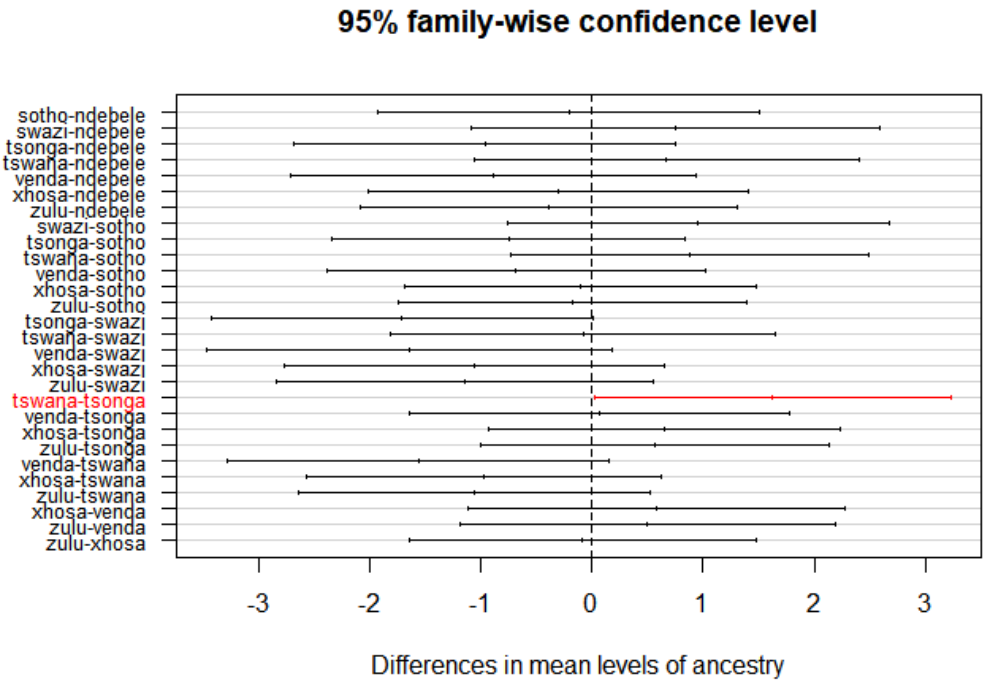


Figure 120. The Tukey's plot of the radio-frontal angle (RFA) variable.

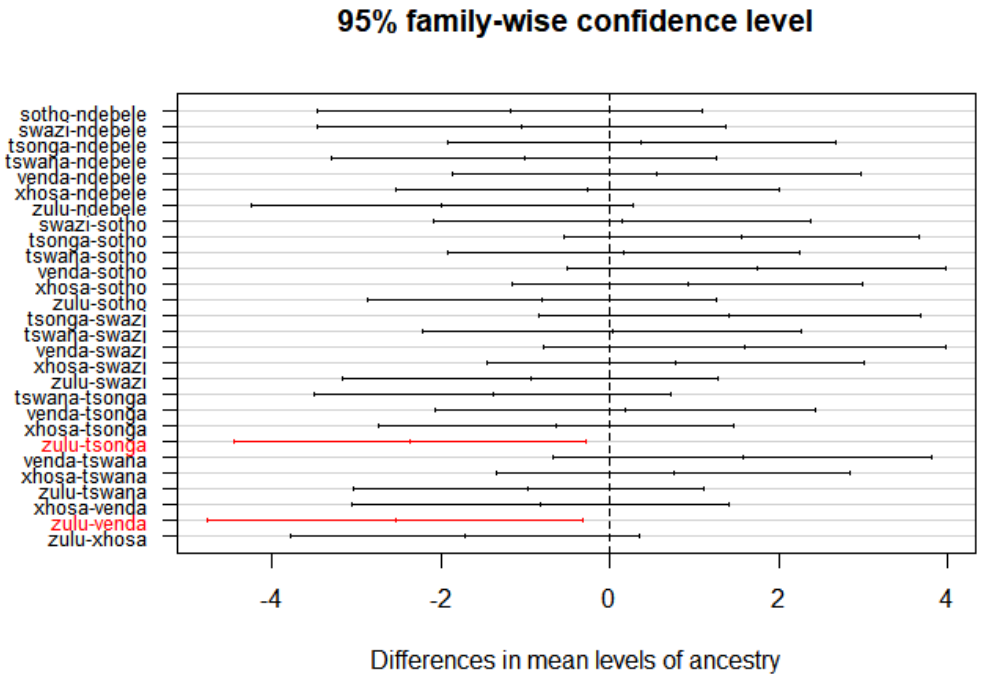


Figure 121. The Tukey's plot of the radio-parietal angle (RPA) variable.

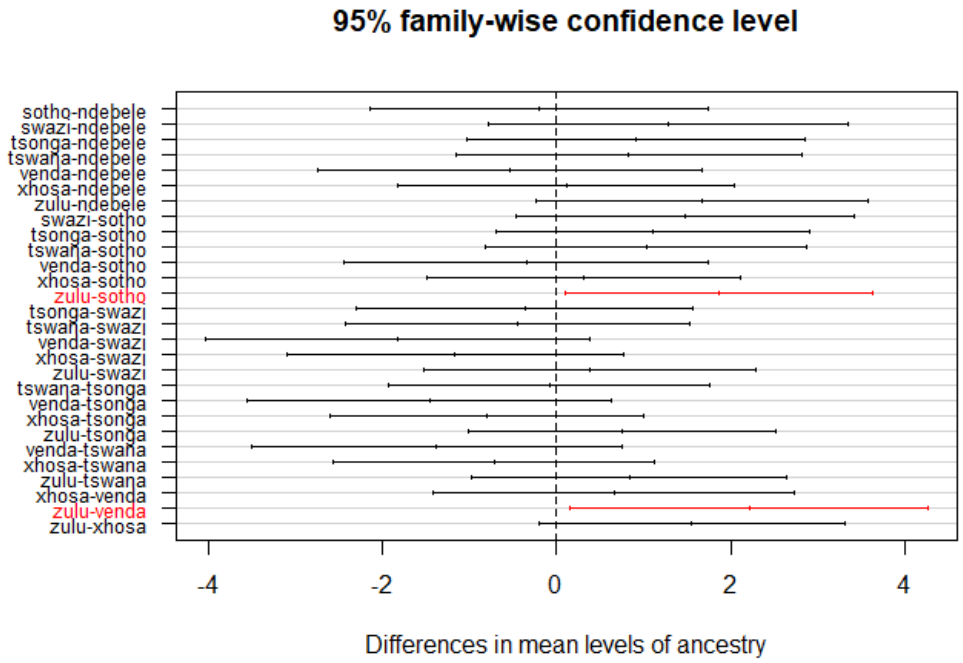


Figure 122. The Tukey's plot of the sub-lambda angle (SLA) variable.