Spring 5-2022

TEACHING OLD CALIPERS NEW TRICKS: USING CRANIOMETRICS FOR ANCESTRY ADMIXTURE ESTIMATION VIA FUZZY MATH

Kristi Carnahan

TEACHING OLD CALIPERS NEW TRICKS: USING CRANIOMETRICS FOR ANCESTRY ADMIXTURE ESTIMATION VIA FUZZY MATH

by

Kristi Carnahan

A Thesis
Submitted to the Graduate School,
the College of Arts and Sciences
and the School of Social Science and Global Studies
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts

Approved by:

Dr. Marie Danforth, Committee Chair
Dr. Bridget Hayden
Dr. Sharon Young

May 2022

Published by the Graduate School

ABSTRACT

Cranial measurements have been a cornerstone of physical anthropology since its formation as a discipline in the early 1900s. However, most other ancestry determination methods come with a significant epistemological issue: they differentiate individuals into discrete categories without accounting for the issue of admixture. Advances in data mining and analysis techniques can now be used to help resolve this issue through soft computing, also known as “fuzzy math”. This type of advanced computational math requires specialized knowledge in computer programming, statistics, and data analysis techniques unless one is using computer programs specially designed to run these analyses.

This project compiled a database from multiple open-source craniometrics data and utilized prepared packages within the R statistical environment to find a valid soft computing method for fuzzy ancestry determination that does not require extensive knowledge in computer programming or data mining. Exploration of database demographics notes an excess of White-identified individuals, and when tested, this demographic skew impacts the ability of the given package to return valid results. The package chosen was valid using the compiled database. Exploration of causes for the invalid results, including a significant White skew in the underlying database due to accessibility of metric databases, overfitting, and the inherent issues of admixture on craniometric research, are explored, and future directions discussed.

ACKNOWLEDGMENTS

My most sincere thanks to Dr. Marie Danforth, for never giving up hope that I could finish this project even as I worked on the frontlines of the COVID-19 pandemic. Your confidence in me has been invaluable in my journey through this program, and because of your encouragement I have exceeded even my own expectations for my journey. Plus, I know so much more about teeth than I ever wanted to, after cataloguing the dentition of the entire Tipu collection. Thank you is not strong enough.

To Dr. Bridget Hayden, who never let me slack away from my potential, even when it frustrated me, and who encouraged me endlessly to continue to push the boundaries of what I knew, what I understood, and how it impacted my life and the work I am interested in pursuing.

To the teams that have created and continue to maintain open-source databases such as UTK's Forensic Databank and NCSU's online database of craniometrics—thank you for making this and much more work possible.

DEDICATION

To my family, thank you for loving me and encouraging me to follow my dreams, even when you don't understand a word that I'm saying. Mom isn't around for this piece, but I know she's proud that I finally finished this. She's going to have to wait a while for the PhD, though.

Nichole, you've kept me sane and loved me through it all, even when you didn't have to do so. I would have to write another thesis to give you the praise you're due. Suffice it to say: You're stuck with me. For Good.

Rebecca, Mel, Nicola, and Katie—Your countless hours spent supporting and sprinting me through the research, writing, and revision process will never be forgotten, even if I didn't have the words to tell you how much your support meant at the time. Without you, this thesis would never have been written—very literally.

To Balt, Brittany, and Jasleen—thank you for your support and encouragement through these final, painstaking pieces. Y'all came into my life in a wild time, and I'm forever grateful for your love and laughter in my world.

To my ER, COVID ops, and many other healthcare friends around the US—thank you for the laughter and encouragement, for trying to explain my thesis to unsuspecting interns and attendings alike (with questionable and hilarious results) when I wasn't there, and for never giving up even when it feels like the world is against us. We can't fix stupid, but we can laugh our way through it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I - Introduction

As far back as Plato and Dionysus, humans have been interested in identifying what separates them from other people, but the modern concept of race as a form of human variation began in earnest with European expansion into global commerce and colonization in the sixteenth century (Molnar 2006). With the publication and general acceptance of Darwin's theories of evolution, the idea of racial differences as static and immutable lost significant credibility. Within 60 years of Darwin's publication of *On the Origin of Species*, physical anthropology would be professionalized with a near-exclusive focus on identifying and classifying racial typologies.

Cranial measurements have been a foundational metric in physical anthropology's typological methodology since its formation as a discipline in the early 1900s. They quickly became a cornerstone in the metric determination of ancestry (formerly "race"), using methods such as the cephalic index (Armstrong-Fumero 2014). These methods attempt to identify and classify individuals based on craniofacial variation, which have been shown to change with both short- and long-term in the form of plasticity and secular change (Ousley, Jantz and Freid 2009).

Plasticity is a temporary change whereupon bone adjusts to a variety of stressors placed upon it (Baab, et al. 2010, Evteev, et al. 2014). These stressors are caused by external factors, such as climate differences in humidity and temperature or cultural factors like food preparation practices and other subsistence methods, and intrinsic factors, such as sexual dimorphism and genetic inheritance (Ross, Ubelaker and Kimmerle 2011). When looking at correlating an individual's variations into larger ancestry groups, these ancestry determination methods—such as the cephalic index—rely

upon secular change, the process by which once reversible traits become encoded genetically and passed down to future generations (Jantz and Jantz 2000).

The cephalic index, as well as the majority of ancestry determination methods that would follow it, came with a significant epistemological issue: it differentiated individuals into discrete categories without accounting for the issue of admixture more than in noting it as a complicating factor in determining an appropriate grouping for 20% or more of their samples (Hefner 2009). Giles and Elliot (1962) argue that the unknown amounts of admixture in non-White samples impacted their overall accuracy—ranging from 82% to 88%. Brues (1990) notes that the Howells's (1973) multivariate analysis has a higher accuracy rate than Giles and Elliott's bcause of the racial ambiguity of the source collections that Howell was able to circumvent with his sample methodology. All of these methods assigned the individual into a single, or "hard", classification, despite the notation of their admixture and without attempting to explore methods that allowed for multi-categorical classifications, until more recent work.

Admixture studies began initially in genetic studies but were largely avoided by craniometric studies until the late 2010s. The first significant publication of methodology whereby an anthropologist was able to successfully validate an admixture method was in 2016, with Dr. Brigette Algee-Hewitt's publication of "Population Inference from Contemporary American Craniometrics" in the *American Journal of Physical Anthropology* (*AJPA*). She utilized soft computing, a form of data mining and analysis which allow for some measure of "imprecision, uncertainty, and partial truth" (Maimon and Rokach 2008b, 1), to allow for a larger amount of information given from the craniometric analysis. Instead of an individual being identified as "White" or "Black",

their ancestry is more precisely expressed as overlapping classes—for example, as 50% Black, 30% White, and 20% Asian. However, this method required complex knowledge of computer programming, data mining, and statistical analysis to be validated.

Knowing it is possible to get such information in a fully validated method was a significant step in bringing new, advanced statistical methods into bioanthropology. Yet the statistical and computational complexity of the methodology meant there was minimal application within bioanthropology as a field. Therefore, the goal of this research was to see if there was an equally valid but more accessible method that biological anthropologists could utilize to get a similar result. The author compiled an initial, unrefined database from two open source craniometric databases—the Forensic Data Bank maintained by the Forensic Anthropology Center at the University of Tennessee at Knoxville (UTK) and the online database of craniometrics from Latin American nations maintained by the Forensic Analysis Lab at North Carolina State University (NCSU). The databases were joined using Microsoft Access, then refined using measurements proven by Jantz and Jantz (2000) to demonstrate the most craniofacial variations in shape and size. Given the size and diversity of the databases, they offered the opportunity to create a statistically significant database for computational testing. If successful, this will offer an easily accessible program that is statistically robust and requires less coding knowledge to obtain valid information on admixture and ancestry of human crania, with subsequent applications across multiple biological anthropology fields.

CHAPTER II – Literature Review

It only takes a quick glance around in a crowd to see the spectrum of human appearance, yet seeing variation does not easily translate into grasping its creation. People have always known that appearance was, in some way, inherited from our parents but little was understood beyond this. American biologists and anthropologists of the late 19th and early 20th centuries assumed morphological stasis—no change in the facial shape or structure—until Franz Boas used craniometric studies to demonstrate the inherent plasticity of the human skull (Cartmill 1998). Since Boas' initial publication, plasticity studies have compiled an extensive list of factors contributing to the final shape of the human body, most particularly the human skull (Hulse 1981).

Historical Approaches to Human Variation

Pre-Darwinian Approaches

Molnar (2006, 3) places the curiosity regarding human variation that would lead to the modern conception of race as beginning in earnest in the late sixteenth century with European expansion into global exploration, trade and commerce, and colonization. Andrea Vesalius's *On the Fabric of the Human Body*, which appeared in 1543, was an intentionally provocative work that illustrated human anatomy and variation in-depth through the use of detailed and accurate renderings of the body and is often posited as the first published work on modern morphological variation. Vesalius's publication would be the first in what would become a major point of scientific and cultural curiosity that continued throughout the sixteenth and into the seventeenth centuries (Marks 2011).

Scientific work done throughout the sixteenth century on human variation such as Vesalius, comparative anatomy such as Edward Tyson's 1699 *Orang-Outang sive Homo*

Sylvestris, and discrediting entrenched biblically based beliefs such as Isaac de la Peyrère's 1655 *Pre-Adamites*, converged in Carl Linnaeus's taxonomic work *Systema Naturae*, published in 1735 (Marks 2011). The taxonomies within his work were not bereft of all religious doctrine, as their construction was in line with the commonly held view that species had been fixed in appearance and number since their creation; it also was not free from the misunderstandings created by purely visual discernment, such as the creation of four 'subspecies' of *Homo sapiens*—American, European, Asiatic, and Negro—based upon the visual differences in the cranial morphology of people discovered since Europe began its global explorations in the fifteenth century (Molnar 2006).

The premise underlying Linnaeus's taxonomy of human variation as different subspecies was not frequently questioned at this time. Contemporaneous scientists critiqued the criteria upon which these categories were differentiated, but never critiqued the idea of multiple human 'subspecies' in and of itself. In 1779, Johann Blumenbach, often cited as 'the father of physical anthropology,' expanded this into five discrete categories—Caucasoid, Mongoloid, American Indian, Ethiopian, Malay—also based on cranial features, despite his observation of the overlapping nature of physical traits between these groups; Cuvier decreased the categories to three—Caucasoid, Mongoloid, Negroid—in 1817. Some scientists, such as Blumenbach and James Cowles Prichard, did question the arbitrary nature of these racial divisions; Blumenbach noted in his work that there was more variation between multiple individuals identified as African than there were between individuals labeled separately as African and European. Prichard eventually determined and published the belief that environmental influences were

responsible for much of human variation (Molnar 2006, Marks 2011). The debates between scientists regarding the extent and causal factors for human variation continue today, but the shape of these arguments was changed by the next major theoretical turn in science: Darwinian evolution.

Darwinian Evolution Emerges

On November 24, 1859, Charles Darwin published *On the Origin of Species by Means of Natural Selection*. In this publication, he made the argument that natural selection—the inheritable evolution of species in response to external stressors—is a process of selective advantage wherein traits that allow organisms to successfully reproduce persist while those that hinder reproduction do not (Marks 2011). Since Darwin's original publication, other methods besides natural selection have been theorized for macro- and micro-level evolution, such as niche construction, mutation, gene flow, and genetic drift. However, Darwin's proposition of evolution had wide-reaching, lasting consequences on how human history has developed; no longer could the idea of a fixed number and static appearance of species be supported. Instead, the prevailing belief was that we have been slowly changing over many years; this complicated the idea of racial differences, which by this time were well entrenched into society and the basis for ongoing structural violence such as slavery, indentured servitude, colonialism/imperialism, and more. Within 60 years of Darwin's revelation, the field of physical anthropology would be professionalized by Aleš Hrdlička, focusing almost exclusively on classificatory and descriptive racial typologies. As a field known for its research focus on racial differences, much of the world looked to anthropology for cues on how to differentiate between races; discord amongst early physical

anthropologists on the cause and meaning of race demonstrated the complexity of the subject, which will be considered in the next section.

Early Physical Anthropology

Early physical anthropology relied heavily on measurements of the cranium (craniometrics) and observable but unmeasured morphological variations (non-metric traits) for racial classification (Armstrong-Fumero 2014, Hefner 2009); contemporary anthropologists continue to use both methods in current research. The underlying assumption of classification based on non-metric traits is the capacity for visual discernment, or the idea that simple observation allows sufficient evidence to successfully interpret complex phenomena. Fernando Armstrong-Fumero (2014) notes the persistence of this idea—that "seeing is believing"—not only in a variety of practices within physical anthropology, like ancestry determination, but more importantly in the public perception and understanding of race and ancestry. He suggests the entwining of visual discernment and scientific (more specifically, statistical) methods began as early as the works of Samuel Morton, one of the earliest physical anthropologists, in the first half of the 19th century (Armstrong-Fumero 2014).[1]

Morton regularly used the cephalic index, one of the earliest forms of craniometrics, calculated as the ratio of cranial breath to cranial length. This proportion

[1] This is not to say that early 19th century scientists studying human variation disregarded measurements completely. Mathematician Francis Galton, founder of the Galton Laboratory for National Eugenics, focused was the development of biometrics to mathematically identify 'normal' human variation for a variety of physical and social traits, like body size and social achievement. Many at this time believed these traits were specific to different 'racial stock', which while problematic, their methods greatly improved the fields of mathematics and biometrics (Molnar 2006: 14).

was touted as an immutable physical trait able to assign individuals into discrete, essentialized, biologically determinant racial categories (Caspari 2003, Ta'ala 2015). Morton used rudimentary statistical analyses on this index and other craniometrics to justify the prevailing, but undeniably racist, views of the time, such as race being a biological truth that impacts moral and intellectual capacities. However, Morton admitted—though he did not see it as problematic—that his method often favored his opinion of race based on visual examination over any interpretations supported by his chosen statistical methods. "In several cases, Morton found himself sidelining the statistical material, or finding ways to explain why it diverged from conclusions that he derived through other means, and in which he seemed to place more faith" (Armstrong-Fumero 2014, 6). Modern standards would consider Morton's work to be superficially scientific at best, but in the 1800s, it effectively connected the idea that visual discernment can support and be supported by scientific inquiry (Armstrong-Fumero 2014). While Samuel Morton's investigative or interpretative methods did not stand the test of time, his work in measurement, incorporation of statistical methods, and reliance on visual discernment continue to impact biological anthropology research, for better or for worse; all three of these methods are still utilized across the field of anthropology as a whole, not just racial or ancestral classification within biological anthropology.

Morton's methods, and the cultural construction of race supporting them, continued with little dissent until the early twentieth century. Ales Hrdlička (1919, 22) wrote that the largest scientific goal of physical anthropology moving forward was "the gradual completion…of the study of the normal white man living under ordinary conditions. … Such knowledge of the white race is eventually indispensable for

anthropological comparisons." He goes on to discuss the necessity of a thorough study of more 'primitive' human races, stating there is "…not a single instance [that] we can say that we possess even a fairly complete record of any of the colored peoples. …It may not be of special benefit to the more primitive groups themselves, but we must have it not alone for descriptive and statistical purposes, but for a proper understanding of the fundamental problems of our own race and of humanity in general" (1919, 23). The view espoused here by Hrdlička became commonplace in the early 20th century, holding such a prominent place in physical anthropology that the field became known almost exclusively as 'the study of race' (Caspari 2003, 2009).

Along with Hrdlička, prominent twentieth century physical anthropologist Earnest Hooton followed closely in Morton's footsteps in studies on the capacity to visually discern human variation. Hooton oversaw the doctoral studies of physical anthropologists at Harvard, one of the earliest of such programs in the United States. In total, he trained 28 students who went on to accept positions across the United States and shape the field as we know it today (Caspari 2009). Methodologically, Hooton is most well-known for creating a standardized suite of non-metric traits he claimed were useful in classification, though not strictly for race or ancestry determination (Hefner 2009). His first paper published in the *American Journal of Physical Anthropology* (*AJPA*) suggested that Icelandic "Eskimo" populations could be distinguished from other populations through four non-metric craniofacial traits; however, he did not position these as strictly racial traits, but traits that were functionally developed as an adaptation to the frigid sub-Arctic temperatures (Caspari 2009).

Beyond their methodological approaches supporting visual discernment, Hrdlička and Hooton shared other theoretical foundations that led them be the only physical anthropologists to ever serve on the Committee on Anthropology of the National Research Council (NRC). The *AJPA*, first published in 1918, was established to be a way to further the main focus of the committee: eugenic-based racial anthropology aligning with contemporaneous European physical anthropology. However, the appointment of only two physical anthropologists to the Committee on Anthropology (all the rest were pro-eugenics scientists from other scientific disciplines) created noticeable tension among much of the fledgling anthropology profession at the time; this tension was heightened by the fact that both anthropologists were pro-eugenics, leaving the committee with no member to temper its ideas with an alternative point of view. Many felt Franz Boas would have been an acceptable choice to serve on the committee as a non-eugenicist member (Caspari 2009).

Franz Boas, a contemporary of Hrdlička and Hooton, was staunchly against eugenics and questioned anthropology's understanding and use of race. While not the first to contest the idea of race as a biological reality, he was the first American-based anthropologist to put forth a report utilizing empirically collected and mathematically analyzed data (to the degree allowable at the time) to refute the idea of morphological stasis and a racial essence, disrupting the base assumption permeating anthropology at the time that race was a fixed biological reality (Caspari 2003; Gravlee, Bernard and Leonard 2003). Analyzing multiple cranial measurements to examine the difference between European-born and US-born children of immigrants, Boas noted small but distinctive differences between these two groups. While the differences seem minute—less than a

centimeter—when individual measurements are considered, the suite of measurements shows significant changes in the overall form (both shape and size) of the cranium (Cole III 1996). Boas posited the changes seen were not based on racial changes but influenced by pre- and post-natal environments (Gravlee, Bernard and Leonard 2003).

Despite Boas' research, the belief in and attempts to scientifically determine biologically discrete racial categories stubbornly persisted within anthropology. In fact, some of his students continued searching for these categories despite their teacher's findings (Anderson 2012). Boas' students were not alone; over the remainder of the nineteenth and into the twentieth centuries, measurements and mathematical formulas became increasingly complex in the search for how they could accurately determine race (Selcer 2012). However, these early studies laid the groundwork for the refutation of biological race by anthropologists, a contentious debate in physical anthropology leading up to Washburn's "new" physical anthropology (Washburn 1951) and Carlton Coon's 1962 publication *The Origins of Human Race.*

The "New" Physical Anthropology

The rise of Nazi Germany, the Holocaust and other atrocities during World War II demonstrated the undeniable racial biases of contemporaneous science. Physical anthropology, realizing its role in the eugenics movement so adamantly pursued by Nazi Germany, began to move away from these classificatory and descriptive research goals (Fuentes 2010). The shift away from typology began in earnest with Ashley Montagu's *Man's Most Dangerous Myth: The Fallacy of Race* in 1942. However, this was not Montagu's first attempt to move physical anthropology away from typological race studies and towards a broader, more (truly) anthropological base. In 1940, he published

the article entitled “A Cursory Examination of the Relations between Physical and Social Anthropology”, wherein he laid out a case for the integration of sociocultural factors into the biological exploration of humanity. In this publication, he notes:

> …not only have physical anthropologists a great deal to learn from the findings of social anthropologists, but we shall also see that unless they make certain of these findings part of their methodological procedures, much of their labor is likely to prove abortive. … The physical factors involved in social development, and the social factors involved in physical development are relationships of obvious importance which up to the present time have been virtually completely neglected by the anthropologist. (Ashely-Montagu 1940, 42 - 43, 61).

Within two years, he would publish his seminal work on the racial fallacy and call for a move away from the understanding of race as a biological concept and towards one reflecting ethnicity, with the understanding that ethnicity describes the sociocultural connection between people beyond their physical similarities. Montagu’s push away from racial science and towards a more culturally centered view of ethnicity received significant, though not unexpected, objections across the field. According to Littlefield and colleagues (1982), he was the sole champion of that fight until the early 1960s.

While Montagu may have been alone in the fight to retire the use of race and rejoin two subfields that had been torn asunder in anthropology, there were others also working to move physical anthropology beyond the study of typological race. In this venture, Montagu was joined by Sherwood Washburn, most notably, and other anthropologists and scientists who would soon become proponents of the “modern synthesis” of evolutionary biology and physical anthropology, such as Theodosius Dobzhansky. In 1950, Washburn and Dobzhansky convened the Cold Spring Harbor Symposium to explore how the disjointed fields of evolutionary biology and physical anthropology could be united once again. At the next annual meeting of the American Association of Physical Anthropologists in 1951, Washburn called for a ‘new’ physical

anthropology which moved towards populational and evolutionary-level studies of humanity instead of the individualistic, typological and classificatory studies that were currently the focus. He went on to write multiple publications (e.g., Washburn 1951) and organize other conferences over the next few years to further develop this idea and its implementation steps (Ellison 2018). However important this paradigmatic shift was for expanding the breadth and goals of physical anthropology research, it has not fully eliminated race as a focus of research—especially in forensic studies. Instead, it has altered how biological anthropology conceptualizes and studies race and utilizes early methodological work (Ta'ala 2015).

Accounting for Admixture

Early attempts. Prior to studies at the population level, admixture—the phenotypic and genotypic results of interbreeding between individuals from previously geographically isolated populations—was seen more as a nuisance preventing the data from showing 'pure' racial typologies instead of an area of potential research. Early researchers such as Giles and Elliot (1962) and Howells (1973) noted that the presence of population admixture obscured their results, but their focus on discrete racial categorization produced a fixation on the assignment of "hard," or single, classifications. Even today, biological anthropologists discuss admixture—especially within forensic anthropology—as a reason for misidentification or inconclusive results. This can be seen in discussions regarding the different accuracy rates of two seminal ancestry classification works: Giles and Elliot's discriminant function analysis based on three U.S.-based sample collections (1962) and Howells's multivariate analysis based on 18 different populations across the globe (1973).

Giles and Elliot collected eight measurements from individuals in the Terry collection (curated at that time in the Department of Anatomy at Washington University's School of Medicine in St. Louis, Missouri, now located at the Smithsonian Institute in Washington, D.C.), the Todd collection (curated at that time in the Department of Anatomy at Western Reserve University's School of Medicine in Cleveland, Ohio, now located at the Cleveland Museum of Natural History), and the Indian Knoll collection (still curated at the University of Kentucky) to formulate a series of equations to differentiate between White, Black, and Native American individuals. Their initial accuracy rates ranged between 82% and 88% overall (on both model and test case individuals), but a wider variation in accuracy of just the test cases, ranging from 76.9% accuracy in identifying Native American males to 100% accuracy in identifying White females. Giles and Elliot (1962) suggested that the unknown admixture amount in the non-White samples likely impacted accuracy. When discussing the collections used for their sample database, they noted a large chronological gap among the three collections (the Terry and Todd collections contain mostly 20th century White and Black individuals, while the Indian Knoll collection is a prehistoric Native American collection) but did not appear to factor in the impact secular change over the span of this chronological difference would have on the measurements upon which they were basing their linear regressions. Additionally, they assumed that "any person showing any phenotypic evidence of Negroid admixture was considered a 'Negro'" (148), which is a questionable practice and could have a significant impact on their identifications. Despite these limitations, they still created linear function equations to separate among White, Black,

and Native American skulls. The potential effects of secular change and obscure admixture could not outweigh the desire for discrete categorization.

Between 1965 and 1980, Howells personally collected approximately 30 measurements from over 2,500 crania in 18 populations and used these measurements to develop multivariate methods for ancestry classification (Howells 1973). In reporting the accuracy of his methods in his initial publication, *Cranial Variation in Man*, Howells notes that his multivariate methods accurately classified 92% of the crania used to develop the methods and an undisclosed "meager" number of crania not used in the development of his methods (Howells 1973, v). (The author also checked his results and different sections and could not find much more clarity on accuracy rates of the test cases alone. The test bank has 524 cases total, but all accuracy rates discuss more than 524 cases total as the basis of their accuracy rates). Brues (1990) explains the higher accuracy rate in Howells's (1973) multivariate analysis compared to those of Giles and Elliot's (1962) as resulting from the "racial ambiguity of the Todd collection 'Negroes,' who must have included many with appreciable White admixture" (6). Hefner (2009) notes that "[w]hen ambiguous or discordant trait values are encountered, admixture or individual idiosyncrasy is invoked…" (986) and suggests that anthropologists who fall back onto admixture in these cases simply do not know enough about the variation of traits in the individual target populations—again, utilizing hard classification of single ancestries (994).

More recent attempts. Anthropologists within and outside of the biological subfield criticize the ongoing practice of single, hard ancestral classifications as a reification of the biological race concept (Armstrong-Fumero 2014). As an alternative to

hard classification, the study of admixture provides several research opportunities, such as population structure, human migration, or the racialization process. Only very recently have anthropologists developed approaches using craniometrics to determine admixture proportions with the goal to further studies of cultural phenomena. Thus far only one researcher, Dr. Bridget Algee-Hewitt, has detailed a novel ancestry estimation method that accounts for admixture. In her 2011 dissertation, she used craniometric data in a finite mixture analysis to determine admixture proportions based on the statistical probability of inclusion into a group and has since utilized the method to further examine biogeographic population structure across space and time (Algee-Hewitt 2016, 2017a, 2017b).

Admixture has the potential to partially address the longstanding critique that ancestry estimation methods reify the biological race concept. A main tenet of racial typologies was the previous existence of pristine or pure races with traits that were once unique and well-delineated from one another but became muddied by European colonization and globalization (Cartmill 1998, 653). However, current research disproves this tenet. Research has consistently demonstrated that both phenotypic and genetic trait distributions geographically overlap in what is commonly referred to as clines (Caspari 2003). Genetic evidence supports continued migration and subsequent gene flow between human populations, despite geographic distance (Cartmill 1998). This means that no population of *Homo sapiens* has ever been fully distinct from one another. Instead, genetic adaptations pass between populations by gene flow resulting from persistent human interbreeding despite variable amounts of geographic isolation over time. This is supported by the work of Algee-Hewitt (2017a) which demonstrates variances in

admixture proportions between four U.S. regions (Midwest, Northeast, South, and West) which align with unique, known population migration histories for these areas. Utilizing this method on a larger scale temporally and spatially may provide more tangible evidence that "pristine" races have never existed than current methods have thus far.

Sources of Craniofacial Variation

Plasticity and Secular Change

Bone is a living tissue that is both solid and plastic; its hardened calcified form adjusts to an almost infinite number of factors and demands placed upon it. Early physical anthropologists attempted to find a static physical 'essence' that could accurately categorize people into discrete racial categories. However, this proved to be impossible because of the gradual rate of change among populations for any given trait and the adaptive process of human plasticity. Plasticity, a temporary change which allows adjustment to a variety of demands placed upon the bones, was first demonstrated in American anthropology by Franz Boas' studies of immigrant families and cranial change (Gravlee, Bernard and Leonard 2003); plasticity studies have since expanded to include the entire skeletal system, not just the cranium (Jantz and Jantz 2000). Factors, such as humidity or temperature, create unique suites of morphological traits that are similar across regions which face similar environmental conditions; long-term differences in cultural strategies, such as subsistence practices, help humans adapt to their environment and create even more diversity within a given region (Baab, et al. 2010; Evteev, et al. 2014; Maddux, et al. 2017; Menéndez, et al. 2014; Paschetta, et al. 2010).

Plasticity persisting over long periods of time is known as secular change. As time progresses, small phenotypic changes may become encoded in the epigenetic, and then

the genetic, code, and passed onto future generations as heritable, permanent traits. The unique combinations of physical traits created by regional secular change and plasticity across the globe allow anthropologists to identify where ancestral groups settled and adapted over many thousands of years, the process of ancestry estimation (DiGangi and Hefner 2013, Menéndez, et al. 2014, Ross, Ubelaker and Kimmerle 2011). Anthropologists utilize this relationship between phenotype and genotype to use craniofacial morphology as a genetic proxy in studies in which genetic examination is inaccessible for any number of reasons, as well as in a number of other studies in human evolution (von Cramon-Taubadel and Lycett 2008, Roseman 2016), dietary reconstruction (Paschetta, et al. 2010, Perez, et al. 2011), population histories (especially migrations) (von Cramon-Taubadel 2011, Hughes, et al. 2013), and biodistance analysis (Stojanowski and Schillaci 2006, Wijsman and Neves 1986).

Intrinsic and Extrinsic Factors in Cranial Morphology

Since Boas' initial publication, plasticity studies have demonstrated an impressive number of factors contributing to the final shape and size of the human body. Multiple extrinsic and intrinsic factors influence cranial morphology—including genetic inheritance, sexual dimorphism, climate, nutrition and health status, and dietary intake—creating the craniofacial variation which grounds ancestry estimation and other variation-based research (Ross, Ubelaker and Kimmerle 2011, Menéndez, et al. 2014). Extrinsic factors are those occurring outside the body, many of which necessitate a temporary adaptive response from the body. Anthropologists can group these factors based on whether they have a direct impact, such as environmental or biomechanical forces, or an indirect impact, such as cultural factors.

The environment is an omnipresent factor in human life, to which the body adapts through the process of acclimatization. Changes brought about by acclimatization can be reversed, but if an individual is continually exposed, and therefore adapting, to a specific environment throughout childhood and into puberty, the acclimatization adaptation largely become permanent after puberty (Frisancho 2010). Research demonstrates a correlation between mid-facial variation and environmental factors like temperature and humidity/aridity. Evteev and colleagues (2014) found a significant association between the nasal and maxillary shapes and the climate (either cold and dry North Asia or more temperate Eastern Asian). The relationship between geographically patterned climate and nasal form was supported by further research by Maddux et al. (2017), who showed that the bony nasal aperture and internal nasal fossa demonstrate changes related to ecogeographic variation, but the soft tissues of the nose do not.

Direct impact to the cranium from biomechanical forces is generally restricted to areas of muscle movement and other load bearing forces of mastication. Paschetta and colleagues (2010) examined how masticatory loading changed craniofacial shape in the Ohio Valley by comparing the craniofacial shape of three prehistoric populations in the middle and upper Ohio Valley. Each population was from a different time period, and each had archaeological evidence of dietary changes as a result of subsistence and technological changes. Their results indicate that different levels of masticatory loading, approximated by changes in diet preparation and intake, alter cranial morphology at several points: the temporal fossa at the attachment sites of masticatory muscles, the general shape of the neurocranium, the zygomatic arch, and the palate (Paschetta, et al. 2010).

Intrinsic factors are internal physiological processes, some of which are under genetic control. One of the most commonly discussed intrinsic factors impacting craniofacial morphology is sexual dimorphism. Sexual dimorphism impacts the size, shape, and developmental timeline of the skull, with males having a larger size, more robust or pronounced cranial shapes, and development of features at a later age than their female counterparts. These attributes are generally regarded as under genetic control, due to the sex chromosomes and other genetic factors signaling the release of hormones throughout the lifetime; however, extrinsic factors such as dietary intake have been demonstrated to have a significant impact on the expression of sexually dimorphic traits, such as the impact of an individual's nutritional status on their stature or the onset of menarche (Moore 2013). In recent literature (such as the sources discussed next), few intrinsic factors (like sexual dimorphism or age/ontogeny) are discussed in isolation because most depend upon or respond to extrinsic factors, making their independent impact difficult to tease apart. Factors like dietary intake and population disease loads have multifaceted impacts on the skeleton by their impact on both physiological functions and biomechanical forces.

Cultural factors also have a profound, though more indirect, impact on craniofacial morphology. For example, Bigoni et al. (2013) explore the morphological impact of different socioeconomic statuses in a medieval Czech Republic population using geomorphometric shape analysis of skull asymmetry. Their results suggest that cranial morphology is impacted by socioeconomic classes due to differential experiences of developmental stress and the differences in types of and access to food resources between classes (Bigoni, et al. 2013). In another study exploring the morphological

impact of cultural differences, Weisensee and Jantz (2016) examined how the epidemiological transition—improved public health and medical initiatives resulting in decreased infant and child mortality and population disease load, and increased overall lifespan—changed cranial morphology of one United States and one Portuguese population over a time span of 150 years, and found that both populations experienced significant secular change but in different areas of the skull; that is to say while morphological change was present, the specific changes were population dependent and not uniform as many expected them to be. These public health initiatives, aimed at changing cultural norms around hygiene and illness, impacted intrinsic and extrinsic factors—dietary intake, nutritional status, and disease load of the populations—resulting in measurable differences in facial size and the size and shape of the cranial base and posterior and lateral cranial fossa.

Most research exploring contributing factors of human craniofacial variability has been conducted within the last 100 years, give or take a decade. Sparked by empirical evidence gathered by Franz Boas, the father of American anthropology, the study of plasticity and morphological change continued in earnest after the first World War (Jantz and Jantz 2000). Human variation has been around as long as humans themselves, but the modern field of physical anthropology was founded upon and continues to influence the research and understanding of human variation.

Ancestry Estimation Methods

Ancestry estimations in bioarchaeology and forensic anthropology rely upon a correlation between biogeographic ancestry/origins and socially constructed racial categories (Ousley, Jantz and Freid 2009). Within bioarchaeology, this association allows

for a variety of studies, such as the process of identity formation (Knudson and Stojanowski 2009), racialization (Geller and Stojanowski 2016), mortuary practices of a given time and space (Rakita, et al. 2005), and other parts of the human experience for which race has had a biological or social impact. Within forensic anthropology, ancestry estimation methods apply what is known about ancestry from population level variation studies to help positively identify a decedent, as part of the biological profile. There are three methods used alone or in combination with one another in modern anthropological and forensic work to estimate ancestry: DNA analysis, non-metric traits, and craniometrics. No method is perfect; each method has its strengths and its drawbacks.

DNA Analysis

Significant theoretical and methodological advances in ancestry estimation have come from the use of DNA (Algee-Hewitt 2016). Genetics researchers utilizes robust and revolutionary methods of estimating admixture proportions ('soft' labeling) over single-category ancestry categorization ('hard' labeling) (Algee-Hewitt 2016; White, Black and Folkens 2012). Genotype determination starts with a "read" of the gene being mapped, where the allele is scanned and determined, or "called", as either matching the reference allele or an alternative allele; this read is often done at least two times (often significantly more) on human genetic material because of the diploid nature of human cells. The proportion of reference to alternative allele reads determines the likelihood of a site being homozygous or heterozygous. Schraiber and Akey (2015) note a significant problem with low-depth scanning is that some alleles are only partially sampled by the reads (i.e., not read with every scan, but only on some of the scans) or not sampled at all by any reads. They note that "[a]ccurate calling of heterozygous sites requires high-coverage data to

mitigate the effects of sequencing errors and the stochasticity inherent in sampling each allele" (729).

The process of determining ancestry and admixture from raw genetic information is complex and riddled with choices that may unintentionally impact later analyses, especially when looking to use this information for population history research (Schraiber and Akey 2015). The issues with the most potential for impacting current ancestry determination methods include cost and sequencing depth (high or low). Cost of genetic sequencing continually decreases, but the current cheaper alternatives are low-depth sequencing in which the chromosomes are sampled with replacement instead of directly testing the full genome; high quality, high-depth whole genome sequencing is still expensive, especially when more than one individual is involved (Skotte, Korneliussen and Albrechtsen 2013). Low-depth sequencing data retains most genomic information, making it an acceptable option for large-scale needs with proper methodological understanding and mitigation of the known issues, such as the model provided by Skotte, Korneiliussen, and Albrechtsen (2013).

The issues in genetic testing of modern populations, such as with forensic cases, are compounded by multiple other factors when dealing with historic and prehistoric remains. Genetic analyses of historic or prehistoric remains come with definite drawbacks, including but not limited to sample size, high risk of contamination and taphonomy-related sample degradation, and destructiveness inherent in sample acquisition (White, Black and Folkens 2012). After death, the human body, including its DNA, begins to internally decompose and, depending on the circumstances of death and cultural burial rituals, may experience a variety of taphonomic degradation forces such as

water, animals, or insects. When examining human remains dating back thousands of years, very little of the remains have viable genetic material, causing most ancient DNA (aDNA) studies to work with small sample sizes of highly degraded material. Nieves-Colón et al. (2018), when seeking to compare the efficacy of two different DNA extraction methods for individuals excavated from tropical or semi-tropical sites, had only twelve individuals with skeletal materials (teeth and the petrous portion of the temporal bone) adequate for DNA sequencing from the three sites (Tanzania, Mexico, and Puerto Rico) which fit the climatic restrictions of their study. Researchers at the Arizona State University Ancient DNA Laboratory extracted less than one hundred base pairs from each element provided for sampling, which is a typical finding for aDNA studies. Additionally, because historic and prehistoric samples are often handled by more than one individual prior to DNA extraction (during excavation, transport, cleaning, etc.), care must be taken to properly decontaminate the test specimen as well as prevent recontamination by researchers during the sampling and testing processes.

Genetic testing of prehistoric and historic populations necessitates removal of part of the body for the testing, even on mummified remains. Because of the destructive nature of DNA analysis, descendants of some Indigenous groups limit or deny this testing (Mayes 2010); destructive testing of this nature without consultation or permission of descendant groups, despite the protections called for in the Native American Grave Protection and Repatriation Act of 1990, has caused increased tension between anthropologists and Indigenous tribes (see Balter 2017, Eveleth 2015). In order to continue our work without causing harm to living or past populations, anthropologists need robust, accurate, non-destructive methods for estimating ancestry and admixture.

Non-metric Traits

One of two non-destructive methods available to anthropologists for ancestry and admixture estimation is the use of non-metric traits. Non-metric traits, also known as discrete traits, are the variable shape expressions of bones and teeth unrelated to pathology. According to Hefner (2009), there are three methods of identification and interpretation used by current methods on a regular basis: 1) description of the bone's shape or readily observable feature (e.g., cranial suture pattern); 2) dichotomous designation (i.e., the presence or absence of a feature); or 3) categorization of a feature along a pre-determined nominal or ordinal scale (e.g., the degree of concavity seen on the nasal profile). Thus far, anatomical sites and the overall methodology have not changed significantly since Hooton's work, but new statistical models and technological advances show potential for altering both the identification and interpretation of these traits.

The strength of using this type of data lies in its ability to be utilized in any laboratory, because it does not require specialized equipment (though Hefner (2009) suggests an inexpensive contour gauge be used for better visualization of the nasal contour) and its ability to be applied to incomplete remains (White, Black and Folkens 2012). However, the drawbacks of this method are substantial, including but not limited to subjectivity in scoring, variable rates of interobserver error, and minimal use of sound statistical analysis methods (Hefner 2009, Klales and Kenyhercz 2015). Hefner (2009, 986) eloquently notes non-metric analysis as is generally used today is as much art as it is science, "an art that is intuitive, untestable, unempirical, and consequently unscientific."

Craniometrics

Craniometrics, the distance between two points on the cranium measured using calipers or (more recently) laser scanning technology and computer graphics software, circumvent some of the subjectivity in application of non-metric methods. Especially in early craniometric studies, the use of non-standardized points for measurements inhibited comparative studies using craniometric data. A century ago, German anthropologist Rudolf Martin (1928) gained scientific prominence throughout Germany and northern and western Europe for his standardization of anthropological measuring techniques and methods, including craniometric landmarks, in *Lehrbuch der Anthropology*. American anthropology did not directly adopt his methods, but many of the current point standards—including those of Howells (1973) as well as Buikstra and Ubelaker (1994)—were influenced by his point definitions (Morris-Reich 2013).

One of the seminal craniometric works for American anthropology was W. W. Howells' 1973 publication *Cranial Variation in Man*. This work had two major consequences, one expected and one unexpected. One of Howell's intentions, which he achieved, was to demonstrate the ability to apply mathematical methods of classification to ancestry estimation techniques. There is a noted shift in ancestry determination publications towards craniometrics and mathematical methods of classification, and away from non-metric modes. The second, and unintentional per Dr. Howells, consequence was the creation of a series of well-defined and illustrated landmark sites, measurements, indices, and angles for craniometric data collection which have become the cornerstone of craniometrics (Howells 1996).

Ancestral Admixture Estimation

Admixture estimation provide a means to address the longstanding typological critique of physical anthropology methods and contribute to populational and evolutionary-level studies. A significant criticism of physical anthropology in the early twentieth century was the reliance on typological or classificatory models, where an entire population is described in terms of traits held by a limited number of individuals (DiGangi and Moore 2013). A better understanding of, and ability to detect, admixture diminishes this by showing the presence and broad variation of traits across global populations. Typological models were based on the idea that traits that were particularly advantageous in a specific climate will increase in frequency, based on the process of natural selection. While this fact does not preclude their appearance in other geographic areas, it was believed that if they do not confer any particularly strong advantage in that climate, the trait will have a generally lower expression frequency. Admixture analysis allows us to examine the expression frequency of different traits, which in turn can be used to explore how populations have interacted with one another as seen with increased gene flow between populations increasing the presence of traits between the populations and/or evolved over time due to other circumstances, such as climate change causing a decreased need for cold-adapted traits.

In 1953, when Francis Watson and James Crick described the double helical structure of the DNA, scientific understanding and available technology moved genetic research into the world of molecular genetics (Gayon 2016), and physical anthropology was along for the ride. Investigations regarding racial admixture initially attempted to determine the correlation between morphological traits and perceived genetic admixture;

after amassing sufficient evidence to comfortably posit the phenotype is, in fact, generally indicative of the genotype, studies linking morphology to genetics all but disappeared. In the late twentieth century and into the twenty-first century, genetic admixture reigned supreme when looking for evidence of population variation and admixture (e.g., Chakraborty 1975, Relethford and Lees 1982, and Parra, et al. 2001).

Algee-Hewitt (2016) notes a shift in human genetics research towards a population level approach to genetic analysis, which has resulted in greater understanding of the scope and consequences of admixture events, as seen in the population structure, ancestry proportions, and degrees of admixture among groups. Studies broaching these topics utilize both short and extended time frames to understand the genetic and phenotypic alterations and infer social changes associated with these admixture events. These correlative studies provide the groundwork for the study of biological distance, or biodistance, through the correlation of genotype and phenotype, and have created new analytical methodologies, such as unsupervised cluster modeling, for better visualization and understanding of hybrid populations (Algee-Hewitt 2016).

Seeking characteristics which would differentiate races, intensive study on the impacts of racial admixture on cranial morphology was largely ignored. Despite the theoretical advances created by genetic admixture studies, Algee-Hewitt (2016) notes the near tunnel vision of craniometric studies on ancestries confined to a single large population (White, Black, Asian, etc.) as determined by supervised classificatory methods or cranial diversity studies requiring complex methodology. This tunnel vision had been in place since the earliest craniometric pursuits. Hrdlička (1919, 24) remarks on the importance of study of ‘primitive’ people because it was believed they are “less

mixed, less abnormal, less pathological, perhaps less aberrant than those of more civilized communities...” and called for “investigations into the physical, physiological, and intellectual effects of racial mixtures on progeny”, stating admixture as a serious concern for “many nations, particularly the American.” However, he notes that these are prospects for future investigations, not studies being actively engaged in the early 1900s.

Algee-Hewitt (2016) suggests the largest problem with traditional cranial metric and non-metric methods is their inability to address proportions of ancestral admixture, and instead proposes using statistical analysis methodology currently employed in genetic admixture testing. Since the practice of craniometrics as proxies for genetic markers has been well established using evolutionary models (see Strauss and Hubbe 2010, Hughes et al. 2013), she chose an unsupervised model of finite mixture analysis along with a traditional three contributor model to explore the underlying population structure and generate admixture proportions of self-declared American Black, White, Hispanic, Native American, and Asian individuals from the Forensic Anthropology Data Bank, a collection of measurements for over 3,400 individuals compiled from individuals in large skeletal collections (such as Terry collection curated at the Smithsonian Institution) measured by the database’s creators (Richard Jantz and Stephen Ousley) and cases supplied by over 100 forensic anthropologists across the United States (Forensic Anthropology Data Bank | Forensic Anthropology Center n.d.). This study is the only one found which uses *only* craniometrics and provides ancestry admixture *proportion* estimation.

Accuracy Rates

Published accuracy rates within anthropology vary widely, and as such must be closely examined for true accuracy versus methodological choices which artificially inflate them. Using 99 identifications from recent Federal Bureau of Investigation (FBI) forensic cases that were positively identified using forensic anthropology ancestry estimation techniques, Thomas, Parks and Richard (2017) determined a correlation rate, which they term an accuracy rate, of 90.9% between the forensic anthropologist's estimated ancestry (determined by unspecific methods) and the self-identified social race of the identified decedant (determined by driver's license or other forms of identification after forensic identification). However, this study shows accuracy rates higher than normally seen in the literature. Liebenberg et al. (2015) found an accuracy rate of 40 – 79% with five cranial indices (cranial index, upper facial index, orbital index, nasal index, and gnathic index) and 83 – 84% with linear discriminant analyses for a population of modern South Africans. Ousley et al. (2009) attempted ancestry determination using Howells' global craniometric data to assess the ability of multivariate methods to classify individuals into ancestry groups consistent with the ancestry assigned by Howells and, further, to attempt classification by region or continent of this ancestry. They found with only 10 measurements, multivariate methods identified 70% of individuals into the correct continent/region and 50% of individual ancestry was accurate; with 24 stepwise-selected variables, these methods identified 89% into the correct continent/region but only 75% of individual ancestry assignments were accurate. Hefner and Ousley (2014) utilized morphoscopic methods to determine ancestry using 10

different classification methods, with classification accuracy ranging form 66.4% (logistic regression) to 87.8% (neural network).

Thomas and colleagues' high accuracy rate likely results from three methodological choices: the inclusion of only positively identified individuals that utilized craniometric analysis and had a determined ancestry (as opposed to a finding of "undetermined" ancestry), the collapsing of multiple Asian-based categories (Asian, Hispanic, and Native American) into one group, and the choice to consider the correct assignment of ancestry as equal whether one or multiple ancestries were estimated regardless of method used. When broken down further, they note that accuracy rates increased when *one or more* ancestries were estimated (90.9%, as previously mentioned) over 'hard' or single ancestry estimations (88.3%). Algee-Hewitt (2016) reports 71 – 75% mean matching accuracy using the finite mixture analysis, when comparing the 'true' identifier (the ancestry recorded in the FDB or Howells records) to the largest cluster membership proportion. The higher accuracy occurrs when she uses only three clusters, effectively collapsing the groups into White, Black, and Indigenous. As she determines accuracy using only the largest cluster membership, her accuracy rates may also demonstrate artificial inflation, as seen in Thomas and colleagues work, if she considered the top two cluster memberships.

Confirmation and Cognitive Biases

The methodological issues with Thomas, Parks, and Richard (2017) demonstrate multiple layers of confirmation bias. Kerstholt et al. (2010) define confirmation bias as "the tendency to selectively gather and process information such that it fits existing beliefs" (138); Thomas and colleagues leave much to be desired in determining and

discussing accuracy rates for forensic anthropological work. Choosing only cases with positively identified remains with ancestry estimation from skeletal remains only, they did not include any cases with an indeterminate ancestry estimation, citing "this study aims to examine the accuracy of ancestry esitmation when an estimation is offered" (2). Their methodological choice implies that because definitive ancestry could not be determined that no estimate was offered, when the opposite is true; ancestry estimation was attempted and could not be accurately determined, which is important to factor in when determining accuracy rates of current methods. While these cases would, undoubtedly, decrease the accuracy rate offered in the study, they also give a truer representation of the accuracy of ancestry estimation methods. By removing cases that would unquestionably decrease the accuracy rate, Thomas and colleagues carefully selected information with the potential to prove a pre-existing belief or idea about the accuracy of ancestry estimation methods, though this may not have been a conscious process.

The other significant issues with the methodology employed by Thomas, Parks, and Richard (2017) are their choice of ancestral categories and their determination of "correct" classification when more than one ancestry was suggested. This study utilized only three ancestral categories: White, Black, and Asian. They consciously collapsed Native Americans, Asian, and Hispanic populations into a single identificaiton category of Asian, despite the ongoing concern within bioarchaeology and forensic anthropology that the classifications of Asian and Hispanic are too broad to be operationally or contextually useful (see Spradley 2014, Tallman and Winburn 2015). This is, effectively, a single category for three-quarters or more of the global population, the epistemological

equivalent of hitting the broadside of a barn! This, again, effectively increases the accuracy rates they will determine, because a determination of Hispanic for an individual who self-identified as Asian will not be flagged as an inaccurate estimation; the probability of being wrong dropped from 80% (four out of five possible classification choices) to 60% (three out of five possible classification choices), with one of those classifications encompassing over 75% of the world population.

Their determination of accuracy rate is further skewed when looking more closely at those instances where the forensic anthropologist suggested more than one possible ancestry. If the anthropologists suggested more than one possible ancestry, Thomas and colleagues considered the ancestry estimation to be correct if the identified ancestry matched either of the estimations; combined with the use of only three ancestry categories, this futher skews the actual accuracy of the methods. This, again, demonstrates their confirmation bias in their choosing an accuracy determination method that artificially inflates their accuracy rates. Thomas, Parks, and Richard (2017, 2) defend this choice stating "[a]lthough this may inflate the overall accuracy rate…it was considered correct because it did not falsely limit the pool of possible missing person matches". However, given the breadth of populations included in their groups, more than one ancestry does not really assist the search either. Their methods created the illusion of improved accuracy, instead of giving a clear representation of the accuracy of current ancestry estimation methods used by forensic anthropologists.

These issues demonstrate not only confirmation bias but also cognitive bias, or the influence of human cognitive processes on the decision-making process, in studies of ancestry determination. Cognitive bias, like confirmation bias, is receiving increased

study and scrutiny in forensic sciences within the last decade. Multiple studies have demonstrated that *a priori* information greatly impact the results across forensic science work (see Canter 2013; Fraser-Mackenzie, Dror, and Wertheim 2013; Nakhaeizadeh, Dror, and Morgan 2014; Nakhaeizadeh, Morgan, et al. 2018). For forensic anthropology specifically, Nakhaeizadeh, Mortan, et al. (2014, 2018) studied the impact of exposure to contextual information on the sex estimation, ancestry estimation using non-metric methods, and/or age estimations; both studies found that exposure to potentially skewing information significantly impacted the anthropologist's results. the author found no studies on the potential for confirmation or cognitive bias in metric work, though this is a fruitful future direction.

Confirmation and cognitive biases are issues receiving increased scrutiny within the forensic sciences (among other scientific fields); these studies have become so prevalent and important that a new subfield—cognitive forensics—has emerged as a focus within the forensic sciences (Nakhaeizadeh, Morgan, et al. 2018). Further research is needed on the extent of cognitive and confirmation bias in other anthropological methods, however, because the methods are only as good as the researcher. When applied properly, data mining and analysis techniques like unsupervised methods (discussed later) are well-suited to reduce cognitive and confirmation bias in data analysis.

Data Mining and Analysis Techniques

Soft Computing

The statistical estimation of ancestral admixture utilizes a relatively new form of data mining and analysis called 'soft computing'. According to Maimon and Rokach (2008a), data mining "tries to solve the crisis of information overload by exploring large

and complex bodies of data in order to discover useful patterns" (vii). In particular, methods within soft computing "exploit a tolerance for imprecision, uncertainty, and partial truth to achieve tractability, robustness, and low-cost solutions" (Maimon and Rokach 2008b, 1). The tolerance for overlapping mixed results separates soft computing from 'hard' computing. In the context of ancestry estimation, you can see the difference between 'hard' and 'soft' in the difference in information given when identifying an individual as "White" or "Black"—single, discrete "hard" classifications—and identifying them as expressing traits that are 50% Black, 30% White, and 20% Asian or other ancestry—'soft' classification that tolerates the limitations of skeletal variation and the clinal distribution of traits associated with human variation. The hard classification, including most of the current and traditional methods, carries the connotation or implicit suggestion that the individual's self-identification aligns with the predominant ancestral geographic profile. Increased utilization of soft classification methods seeks to offset these ideas through demonstrating the lack of "pure" or single-origin ancestries. Further research, such as has been started by Algee-Hewitt (2017a, 2017b), demonstrates the differing population histories associated with different social race categorization and therefore the multiplicity of ways that one may present physically compared to their personal racial identification.

Posterior Probabilities

Soft computing for admixture estimation relies upon the abilities of these methods to supply posterior probabilities. Posterior probabilities are specific to Bayesian statistical analysis methods, combining information from prior probabilities and likelihood functions. Prior probabilities are the probability of an event occurring determined before,

or prior to, data collection, and is determined without factoring in any conditions on the realistic outcomes (i.e., probability of any and every theoretically possible outcome). Likelihood estimates are the probability of an event occurring based on current knowledge (i.e., after data collection) without considering outcomes that did not occur during the data collection. Posterior probabilities combine the information from these two theories, weighting the infinite probabilities of an event occurring (or in this case, membership in a cluster) based on the likelihood of such an event based on current knowledge; in shorthand, posterior probabilities are defined as prior probability multiplied by likelihood (Lee 2012). During the analysis, these probabilities are calculated based on the relative distance from the data point to the center, or centroid, of the various clusters (Moore 2013). Utilizing posterior probabilities in combination with fuzzy methods allows for the populational and evolutionary research discussed previously. No trait is assumed to be group-limited, but groups impacted by similar biogeographical influences can be connected by the strength and likelihood of trait expression in any given cluster.

Clustering Methods

Clustering analysis of continuous variables, like craniometrics, is based on the principles of distance and dissimilarity—the traits of those 'inside' the clusters are more similar to, and therefore of a 'closer' distance when graphed—to others in the cluster than to those 'outside' the cluster. Clustering methods generally take one of three approaches to cluster identification: partitioned, hierarchical, and fuzzy. Both hierarchical and partitioned methods are considered 'hard' computing in their traditional forms, while fuzzy methods are, by definition, soft computing methods (Kubat 2017).

Partitioned and hierarchical methods are similar in outcome but different in approach. Both hierarchical and partitioning methods begin with a predetermined number of clusters as input by the researcher, into which they divide individual data points. Partitioning methods create clusters by minimizing the variation within each cluster, with a goal of the most homogenized clustering achievable given the level of heterogeneity inherent in most datasets. A strength inherent in partitioning methods is the ability for affiliation changes throughout the clustering process; the partitioning algorithm moves individual data points as it gains more data to minimize the differences within the cluster group (Sarstedt and Mooi 2014). One of the most well-known and commonly used cluster methods, k-means clustering, is a partitioning method. For data analysis purposes, *k* stands for the number of clusters, so the k-means clustering occurs through analysis of and relationship of data points to the mean—center or centroid—of the cluster. K-means clustering methods are commonly used in a variety of analyses due to the simplicity of its methodology (Kubat 2017). Everitt and colleagues (2011) note that the field of archaeology uses a variety of cluster analyses, including k-means clustering, to uncover patterns of artifact distribution over time and space.

Hierarchical clustering is named after the structure created through the analysis, which resembles a hierarchy structure map. Unlike partitioned clustering, once a data point is assigned to a cluster during the hierarchical clustering process, it is not moved. As the process continues, the initial clusters are refined to smaller clusters until the best clustering fit has been achieved or the number of clusters indicated by the researcher has been met. Refinement occurs through one of two methods:

1) ‘bottom up’ or agglomerative approaches that build clusters by examining the relationship between individual data points and building clusters ‘up’ as they examine the relationship of each new data point to previously established ones,
2) ‘top down’ or divisive approaches that build clusters by initially considering all data points as one cluster and building the clusters ‘down’ by examining the larger cluster for smaller, concentrated clusters.

Hierarchical methods can be represented by a dendrogram, such as seen in taxonomic work (Everitt, et al. 2011, Sarstedt and Mooi 2014). Depending upon the purpose of the analysis, this approach can be useful. However, the rigidity of initial placement (once it is assigned, it is not moved from that place) makes hierarchical methods less than ideal for admixture analysis, compared to other methods that allow for reassignment to better fitting cluster as the algorithm gains more information.

In the statistical and computational analysis jargon, methods that result in non-discrete or overlapping clusters are ‘fuzzy’ clustering methods. Where partitioned and hierarchical methods give cluster assignments as either ‘in’ (noted as a 1) or ‘out’ (noted as a 0) of a single cluster (known as ‘crisp’ methods), fuzzy clustering indicates the strength or probability of membership in *some or all* clusters. Fuzzy clustering relies upon fuzzy logic, described by Everitt and colleagues (2011) as “an extension of Boolean logic in which the concepts of true and false are replaced by that of partial truth” (244). Obviously, not all clustering methods are fuzzy or can be modified to be fuzzy, but the number of fuzzy methods continually increases. Some initial fuzzy methods, such as the fuzzy k-means cluster and the fuzzy k-nearest neighbor, derived directly from traditional

hard computing methods. (Nearest neighbor is one of many different agglomerative hierarchical methods available.)

Classification Methods Used in Anthropology

Anthropologists have used some variation of hard and/or soft computing techniques for many years. Some of the earliest, and still more widely used, linear modeling methods are hard computing methods, such as discriminant function analyses. However, using craniometrics (specifically) in these methods is problematic because they invalidate multiple of the base assumptions of the methods (i.e., independence of the variables), therefore causing questionable validity of the results. Advances in computational science have developed a variety of nonlinear modeling techniques, like neural networks, capable of powerful analyses without the limitations imposed by linear modeling methods.

Discriminant function analysis. Discriminant function analyses (DFAs) are a form of linear modeling used to isolate or identify at least two variables that identify or predict membership in two or more groups. This analysis class includes two-group discriminant function, stepwise discriminant analyses, and canonical analyses, and requires base assumptions of normal distribution of data, homogeneity of variance and covariance, and independence of variables. Two-group discriminant functions are analogous to multiple regression and, similarly, can be expressed in a linear equation with a regression coefficient, correlating to the possibility of group membership, for each variable. When attempting to understand the impact of variables on group membership, researchers can utilize stepwise discriminant function analyses to create different models of the group by including or excluding variables to see which variables have the largest impact on group

membership. Finally, the most complex version of discriminant function analyses is the canonical analysis. This is used to differentiate between multiple groups by determining a series of 'functions', where the first function is a suite of variables that provides the most differentiation between groups, the second function provides the next most amount of differentiation, etc. Each function will be independent of one another, in that their discriminatory capabilities will not overlap with that of any other function (Statsoft, Inc. 2013).

Because of their differential capabilities, DFAs have been extensively used in bioanthropological studies since the discipline as a whole has moved—in theory, even if not fully in practice—towards population level studies. Theoretically, this makes sense, as DFAs can analyze multiple variables and help delineate otherwise obscure differences between two populations. While not the first to use it, Giles and Elliot (1962) utilized this method in what has become a seminal work in racial classification using craniometrics (discussed earlier). Dividing the sample by known or estimated sex, they used eight cranial measurements to create two series of formulas (one for males and one for females) by which researchers could delineate—in hard classifications—between White, Negro, and American Indian individuals. Giles and Elliot also used five cranial measurements to create a discriminant function formula by which researchers could determine the sex of an individual (Giles and Elliot 1963). While these two publications are not the only ones to utilize discriminant function analysis, they are some of the earliest to do so and are still widely recognized within the field for their groundbreaking nature.

DFAs are created on the basis of larger-scale collections or populations, common uses of discriminant functions within the forensic branch of bioanthropology tend to remain individualized—to classify an individual whose status is otherwise unknown or questionable, such as seen in the computer program FORDISC, which uses DFA to determine ancestry and sex of individuals (Jantz and Ousley 2005). FORDISC utilizes a customizable database comprised of measurements from Howells' global database and the FDB as a basis for its analysis; users can choose which populations to include in their analysis, and proper selection of incorporated populations has been shown to have a significant impact on the results. Recent criticism of the program has focused on accuracy rates, even with appropriate population base selection, (Ubelaker, Ross and Graver 2002, Elliott and Collard 2009) and the limitations of the current database for the global population, especially those of significant admixture (L'Abbé, et al. 2013, Urbanová, et al. 2014, Dudzik and Jantz 2016).

Additionally, variables used in the analysis should be carefully selected to ensure their independence from one another to meet the underlying assumptions of DFAs. This method relies on assumptions that cannot always be met with real-life datasets. Specifically, DFA assumes that variables are not correlated with one another (non-collinearity), that each independent variable is normally distributed (multivariate normality), and that variation between the group variables equals the variation between the prediction variables (homoscedasticity) (Statsoft, Inc. 2013). Cluster analyses require none of these assumptions because the main goal of cluster analyses is to *discover* or *uncover* groups within the data (Everitt, et al. 2011). The potential redundancy of variables, as seen in craniometrics, would undermine the overall validity of the results

(Statsoft, Inc. 2013). Additionally, the independence of craniometric variables is questionable because of the overlap in measurements (e.g., biorbital breadth includes interorbital breadth and the majority of the orbital breadth measurements) and the interrelationship between measurements (i.e., when one measurement—like the nasal height—changes, it will often necessitate a change in other measurements—like the upper facial height).

Without assurance of these assumptions always being met, the validity of the testing is compromised. However, advances in computational methods have created a variety of powerful, nonlinear modeling techniques, such as neural networks, which do not have the same base assumptions of variable independence or normal distribution of variables. These methods require varying levels of technical and theoretical understanding, but their potential for revealing previously obscure differences or information about populations is thus far under-explored within anthropology.

Neural networks. Neural networks are a class of nonlinear modeling techniques that can identify patterns or relationships between variables in datasets and model complex functions, without many of the validity issues that plague linear modeling methods. Computational—*artificial*—neural networks (ANNs) are based on the human brain—the *biological* neural network—and have been used for a variety of data mining tasks, including classification, clustering, and predictions. The increased use of neural networks is related to two important features: their power and user accessibility (Statsoft, Inc. 2013, Zhang 2008).

The power of neural networks comes from four central characteristics of this class: 1) adaptive, data-driven learning, 2) processing of complex

relationships/functions, 3) non-linear modeling, 4) ability to process large amounts of imprecise data, including non-metric data such as the non-metric traits used in biological anthropology. ANNs, like biological neural networks, learn from experience, meaning they require training algorithms, where they use representative data to discover and understand the structure of the data it will be processing. The output of neural networks improves through the innate learning of the algorithms as more data is received and processed, which contributes significantly to the user-friendliness of the method, as it does not necessarily require any reprogramming or adjustment of the algorithm if the user has completed the appropriate preparatory work and chosen the correct neural network for the analysis (Zhang 2008, Statsoft, Inc. 2013, Hefner and Ousley 2014).

Despite the power of ANN to work with metric data, all instances of neural network usage in biological anthropology thus far have used non-metric variables. Hefner and Ousley (2014) were the first to apply ANN to biological anthropology work with skeletal work, in their exploration of eleven statistical methods, ranging from logistic regression to neural networks, for their accuracy in determining ancestry based on six cranial morphoscopic traits. They found that ANNs had the highest rate of correct classification of the methods used, with multiple other nonlinear, machine learning methods (random forest modeling, support vector machine) having similar but slightly lower accuracy rates. Comparably, Cavalli, Lusnig, and Trentin (2017) used the shape of the calvarium as determined from computed tomography (CT) scans and a pattern recognition ANN to determine sex on a total of 1,700 individuals, achieving an accuracy rate of approximately 81% using multiple classifiers. However, this study has a grave limitation in that it was only completed on healthy, adult Caucasian individuals, so its

accuracy and validity on any other ancestry or on those with potential health issues impacting the shape or size of the cranium is questionable at this time. Finally, Trentin, Lusnig, and Cavalli (2018), developed a new type of neural network, called a Parzen neural network and used it to determine sex on 1,400 healthy, adult Caucasian individuals based on cranial CT morphology. Again, their work suffers from sampling limitations, but with an accuracy rate of 81%, also shows the power of ANNs, especially for non-metric traits, as are commonly used in biological anthropology.

Bioanthropology as a whole has, at least theoretically, attempted to correct and move beyond its early beginnings as a predominantly racial science, which included a significant reliance upon visual discernment over mathematical or statistical evidence. Biometric data as collected from human skeletal remains has come with its own set of challenges when attempting statistical analysis, such as issues of discrete classification or finding a reasonable number of measurements that neither overfit the data nor violate the basic assumptions of the statistical model being utilized. The more recent trend towards soft, or fuzzy, computing is one attempt to help address these technical issues, as well as relevant cultural issues, such as the critique of hard (single) classification as a reification of the concept of biological race. However, knowing these methods can help address these issues does very little if the methods themselves are not approachable by more than a small number of bioanthropologists. The research developed for this thesis, as explained in the next chapter, is an exploration of the possibility ways in which valid methods could be made approachable to a wider selection of researchers.

CHAPTER III - Materials and Methods

The materials and methods described in the following sections were developed to test multiple hypotheses on the capability of the FANNY package in the R environment to accurately determine ancestry using craniometric measurements. If found to be adequately accurate, this package offers the ability to determine ancestry admixture rates as well as "hard" ancestry determination. The main goal of this research is test if FANNY, part of the open-source package 'cluster' in the R environment, is a valid, reliable, and stable method for accurately estimating ancestry of a large sample of diverse crania compiled from freely available craniometric databases. Accuracy rates, used for determining validity, reliability, and stability, will be established by comparing the ancestry determined or self-identified in their respective source database (Forensic Databank (FDB) at the University of Tennessee-Knoxville or NCSU's online database of craniometrics from Latin American nations) with the ancestry determined to be the largest proportion of the individual's; my goal is that the accuracy rates will meet or exceed 70%, putting them equal with other craniometric methods. If FANNY is found to be valid, reliable, and stable, it would create easier access for many bioanthropologists to explore the use of fuzzy methods in their own work.

Materials

Craniometric Databases.

A persistent issue in forensic anthropology research is procuring an adequately-sized sample representative of the time period under question. This research combined two existing craniometric databases, the Forensic Anthropology Databank and North Carolina State University's database, to create the needed sample.

Forensic Anthropology Databank. Drs. Douglas Ousley and Richard Jantz created the Forensic Anthropology Data Bank (FDB) in 1986 with the assistance of a National Institute of Justice grant to act as a repository for demographic and skeletal data from three sources of modern remains: identified and unidentified forensic cases contributed by forensic anthropologists nationwide, including over 400 cases analyzed by J. Lawrence Angel, and the donated remains curated by the Forensic Anthropology Center at the University of Tennessee at Knoxville (UTK). The FDB currently contains over 4,000 individuals and regularly adds others submitted by forensic anthropologists across the United States. Most individuals have confirmed basic demographic data (sex, race, birth year, age at death), but some individuals have only demographics determined from skeletal analysis, no confirmed data (Forensic Anthropology Data Bank | Forensic Anthropology Center n.d.). Measurements used in the FDB primarily follow the definitions set by W.W. Howells (1973), and all others can be found in the informational package that comes with the FORDISC program (personal communication with Dr. Jantz, 2017). Through email correspondence, Dr. Jantz has graciously supplied craniometric data from the entire FDB, containing 2519 individuals total, 2481 of whom have between 1 and 61 measurements. Sampling procedure will be discussed below.

NCSU database. North Caroline State University's (NCSU) Forensic Analysis Lab, under the direction of Dr. Ann H. Ross, maintains an online database of craniometrics from a variety of Latin American nations, found at https://sites.google.com/a/ncsu.edu/craniometrics-database/database. This database was initially created to investigate admixture in "Hispanic" populations. Samples include individuals from pre-contact societies through modern forensic cases. Because of the

extensive time period of this database, the integrated search function was used to obtain only forensic cases to minimize the chance of secular change issues; the search returned a total of 20 cases from Panama and Peru. These cases were exported to an Excel document and subjected to the sampling procedures discussed below (NCSU Forensic Analysis Lab, n.d.). Because of the limited scope of the database (i.e., only cases from Latin America), there is no designated race on these individuals when exported, so all were assigned the label of "Hispanic" in the database to be used for later accuracy assessment. The nationalities of the individualities were placed under ethnicity.

Software Programs

Readily available software resources make combining, refining, and testing large datasets easier and faster than ever before. This research utilizes two software programs to complete its research. While Microsoft Office software is not free, anyone can have access to it for a fee, or a similar database program capable of the same work, such as LibreOffice, for free (Smith 2021); Microsoft Access, the database program of Microsoft Office, is used to compile databases and refine them to create the main sample database (Microsoft Office Support 2018). It used the open-source computer software known as the R environment to create randomly generate test and split datasets (detailed below) and complete the necessary statistical tests (The R Foundation n.d.).

Microsoft Access. Microsoft Access is the database management software included in the suite of Microsoft Office products. Utilizing a database, as opposed to a spreadsheet, helps prevent inconsistent or redundant data entry and expands the functionality of the data to include multiple types of search queries, forms, reports, and more (Microsoft Office Support 2018). For this research, the FDB and NCSU databases

were uploaded as separate tables within a single database, and append queries were utilized to create the main sample database (discussed later).

R Statistical Environment. Though fuzzy testing is becoming increasingly common, there are very few commercial sources that support this testing without buying additional upgrades; even with the upgrades, one must still know how to compile and execute some form of computer coding to use the software and its upgrades to complete the needed testing. Given this, the most accessible software for this testing is the free, open-source computer environment of R, as it requires a similar level of computer coding knowledge but is free and open to the public (The R Foundation n.d.).

There are two components of the R environment: the programming language and the software suite. The R Foundation, which created and maintains the software, describes the language as "well-developed, simple and effective" (The R Foundation n.d.). For those who are not familiar with the language, there are multiple free, online tutorials, such as DataCamp (https://www.datacamp.com/courses/free-introduction-to-r), that teach the programming language. The software suite allows for statistical analysis and graphic representation of large volume datasets through individual programming of programs, an extensive collection of 'packages' created and published for R by other users, or some mixture of the two. This research utilizes 'packages' found in R that were programmed and released for general use by individuals throughout the world. These coding packages are available for search and download through the Comprehensive R Archive Network (CRAN), a series of globally mirrored servers maintained by the R Project specifically for package storage and distribution.

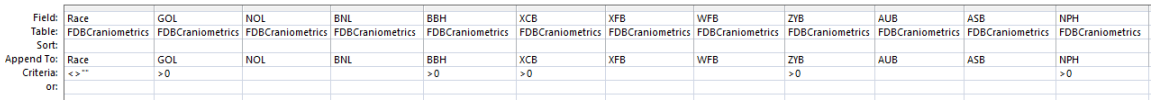
Utilizing k-means clustering allows variability in number of clusters. Increasing the number of clusters could reveal underlying biogeographic patterning to the craniometric data that is lost due to the way in which current identification nomenclature groups large geographic areas into a singular identity, such as Asian and Hispanic populations. This research tests the algorithm using between three and eight designated clusters, assessing for clustering composition and reliability of each cluster designation, with the goal of assessing the potential of these algorithms to differentiate smaller clusters aligned with the regional biogeographical patterning instead of larger continental patterns.

Sampling and Database Demographics

Sampling. Of the 2,481 FDB individuals with measurements, 2,451 individuals in total have an associated social race. The author combined those individuals from the FDB with those from the NCSU database sample. To ensure the data provided for the analysis represents both general skull shape and mid-facial differences, the combined sample was then further refined by craniofacial markers. In assessing for craniofacial changes between the mid-nineteenth and twentieth centuries, Jantz and Jantz (2000) used five craniofacial markers to look for changes in the shape and size of the cranial value and anterior face: glabello-occipital length (GOL), basion-bregma height (BBH), maximum cranial breadth (XCB), bizygomatic breadth (ZYB), and nasion-prosthion height (NPH). While these areas are subject to secular change (morphological changes over time), they are also the minimum number of areas which ensure adequate representation of craniofacial shape and size variation across this sample. All available measurements on included individuals were used in the analysis, not just the five measurements used to

refine the sample. This was intended to offset the secular change presented by these measurements by having measurements across the spectrum of susceptibility, including those with very little demonstrable change over time.

Both data sets (FDB and NCSU) were uploaded to Microsoft Access and append queries used to create the sample based on the presence of a social race and the five craniofacial markers (see Figure 1). From an initial pool of 2,539 individuals, this creates



Field:	Race	GOL	NOL	BNL	BBH	XCB	XFB	WFB	ZYB	AUB	ASB	NPH
Table:	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics	FDBCraniometrics
Sort:												
Append To:	Race	GOL	NOL	BNL	BBH	XCB	XFB	WFB	ZYB	AUB	ASB	NPH
Criteria:	<>""	>0			>0	>0			>0			>0
or:												

Figure 1 Partial Microsoft Access Append Query for FDB Demonstrating Use of Refinement Criteria

a final sample of 1,924 individuals. The social race of all individuals was then standardized to reflect five categorizations based on the FDB system: White (W), Black (B), Native American (NA), Asian (EA), and Hispanic (H). The individuals who had an already admixed designation retained their original designation. The random number generator contained in the R environment, coded as the command "sample" (see Figure 2), was used to generate a testbank containing 500 randomly chosen individuals; the rest of the sample (1,424 individuals) was designated as the main sample (uploaded into R as "S1"). Two databases were created for the main sample, one containing all variables as described above and one with only the five variables used to cull the original database into the sample database (GOL, BBH, XCB, ZYB, and NPH). The variable limited sample was used to test the hypothesis that too many variables can decrease the accuracy (discussed later).

The main sample was then split into multiple separate, smaller databases, again using the R random number generator. First, two databases were created for stability

testing by splitting the main sample into two equal sized secondary testbanks (712 individuals each; see Figure 2). The third and fourth databases were created to determine

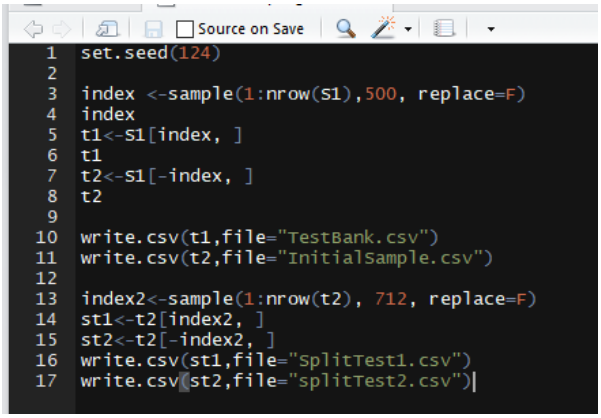


Figure 2 R Coding to Create and Export Test and Split Databases from Main Sample (S1)

if the disproportionately high number of White individuals in the main sample would impact the accuracy of the clustering method. The third database totaled 60 individuals—all available EA and admixed individuals (n=20) plus ten individuals from each of the four remaining ancestry categories (W, B, H, NA, n=40). The fourth database followed the same pattern as the third, but with a larger number of individuals; it included all available EA, NA, and admixed individuals (n=37) plus 50 randomly selected individuals from the remaining ancestry categories (W, B, H, n=150) for a total of 187 individuals.

Database demographics. Understanding the demographic representation of the data is important for result interpretation. The full sample database contained 1,924 individuals with birth years ranging from 1892 to 1990 and ages (both estimated and known) ranging from 16 years old to 101 years old. Each individual had a known or estimated sex, designated as male (M), female (F), or unknown (U). Table 1 (below) demonstrates the demographic makeup of the full sample by age and sex.

Table 1 Full Sample Demographics by Race and Sex

Designation	**Male**	**Female**	**Unknown**	**Total**
Black (B)	250	162		412
East Asian (EA)	6	7		13
Hispanic (H)	180	44	4	228
Native American (NA)	19	7		26
White (W)	768	463	1	1,232
B/NA	2			2
W/B	2	1		3
W/EA	1	1		2
W/H	1	4		5
W/NA	1			1
Total	1,230	689	5	**1924**

The demographics show a disproportionate amount of White, male, and White male individuals within the sample; Whites and males make up the majority of their respective samples by size (approximately 64%), and there are almost twice as many White males (768 or 40% of the total sample) as any other category. The reality of donation-based collections, such as the FDB, is that they reflect the demographics of the society from which they are derived. Despite contributions of forensic anthropologists around the US, the primary source of participating individuals is Tennessee and surrounding areas; the latest census bureau statistics estimated Tennessee to be 79% White, slightly higher than their representation in the current sample (United States Census Bureau 2016).

Tests of stability and reliability (discussed later) are also impacted by the demographics of their respective databanks. The demographics of each sample set are detailed in Tables 2 through 5.

Table 2 Test Bank Demographics by Race and Sex

Designation	**Male**	**Female**	**Unknown**	**Total**
B	60	41		101
EA	2	2		4
H	60	13	2	75
NA	6	3		9
W	197	112		309
W/EA	1			1
W/H		1		1
Total	326	172	2	**500**

Table 3 Initial Testing Sample Demographics by Race and Sex

Designation	**Male**	**Female**	**Unknown**	**Total**
B	190	121		311
EA	4	5		9
H	120	31	2	153
NA	13	4		17
W	571	351	1	923
B/NA	2			2
W/B	2	1		3
W/EA		1		1
W/H	1	3		4
W/NA	1			1
Total	904	517	3	**1424**

Table 4 Split Table #1 Demographics by Race and Sex

Designation	**Male**	**Female**	**Unknown**	**Total**
B	101	58		159
EA	2	3		5
H	60	16	2	78
NA	7	3		10
W	278	174	1	453
B/NA	2			2
W/B	2	1		3
W/H		2		2
Total	452	257	3	**712**

Table 5 Split Table #2 Demographics by Race and Sex

Designation	**Male**	**Female**	**Unknown**	**Total**
B	89	63		152
EA	2	2		4
H	60	15		75
NA	6	1		7
W	293	177		470
W/EA		1		1
W/H	1	1		2
W/NA	1			1
Total	452	260	0	**712**

There is a risk with random sampling of skewed representation in any given sample, but each of the data sets demonstrated the same approximate proportions as the larger, full sample (i.e., approximately twice as many males than females, Whites than other social race, etc.). Only the second split table (Figure 7) lacked any individuals of unknown sex. The test bank had only two individuals with originally designated admixture, which was half as many as the other datasets in this research.

Variable determination and missing values. A common concern in forensic and bioarchaeological skeletal analysis is an inability to obtain complete measurements on remains because of issues like taphonomy, antemortem and/or postmortem treatment of the body, and issues with full skeletal recovery and preservation. As a result, it may not be feasible to obtain many cranial measurements. Missing values create an issue for cluster analysis, as the algorithm determines clusters based on comparison of all variables. Missing values must be addressed to avoid—or at least be aware of—the potential for skewing the results. There are multiple ways to deal with the issue of missing values, with varying impacts upon the resulting analysis.

The two simplest ways of dealing with missing values are 1) to replace the empty cell with a “0” or 2) to replace the empty cell with the mean value of that variable. The most significant drawback of the first method is that potential for a heavy skew of the centroid (which is determined by the average of all individuals in the cluster), depending on how many individuals have missing values for a given variable. Therefore, this research utilized the second option, determining the mean measurement (i.e., not including the individuals with missing variables) for each variable using the AVERAGEIF function in Microsoft Excel and using it to replace any missing values. This helped prevent skewing the data towards lower centroids due to the presence of zeros in the absence of any given measurement. However, having too many individuals with the average inserted can also skew the data, though not as drastically as having zeroes. To minimize this, each variable was examined to determine how many individuals have missing values, and those for which more than 10% of individuals (from the sample as a whole [n=1924]) with a missing value were eliminated from the variable pool. For all remaining variables, the mean was inserted into any empty cells for that variable, as discussed above. Following these parameters, this research utilizes a total of 22 variables that were included in both the FDB and NCSU databases. The variable names and abbreviations are detailed in Table 6 and were measured in accordance with W.W. Howells (1973) published instructions.

Table 6 Variables Included, Measured According to Howells (1973) Instructions

Abbreviation	Name	Abbreviation	Name
BNL	Cranial Base Length	MDH	Mastoid Height
BBH	Basion-Bregma Height	OBH	Orbital Height
NLH	Nasal Height	OBB	Orbital Breadth
NLB	Nasal Breadth	DKB	Interorbital Breadth
ZYB	Bizygomatic Breadth	EKB	Biorbital Breadth
AUB	Biauricular Breadth	FRC	Frontal Chord
BPL	Basion-Prosthion Length	PAC	Parietal Chord
NPH	Nasion-Prosthion Height	OCC	Occipital Chord
GOL	Maximum Cranial Length	MAL	Maxillo-Alveolar Length
XCB	Maximum Cranial Breadth	FOL	Foramen Magnum Length
WFB	Minimum Frontal Breadth	FOB	Foramen Magnum Breadth

Methods

Clustering Method and Package Selection

Posterior probabilities obtained from any fuzzy method will be, in part, dependent upon the clustering method chosen. Multiple pre-existing R packages exist for most clustering methods, so each must be examined to ensure use of the most appropriate package. In the case of fuzzy k-means clustering, the method of choice for this research (discussed next), there are five pre-existing R packages (ppclust, fclust, FuzzyR, RCmdr.FuzzyPlugin, and FANNY) available for data processing.

Clustering method. Ancestry estimation focuses on biogeographic skeletal features, expecting some features to cluster together and overlap based on the limited number of biological responses to geographic environments. As such, the most appropriate soft computing methods for admixture estimation should also focus on fuzzy clustering methods with the ability to identify posterior probabilities for cluster membership. With the ongoing advancements in soft computing, there are now multiple

‘fuzzy’ options for k-means clustering. Theoretical simplicity, flexible affiliation of the traditional method, and the ability to modify the traditional algorithm to allow cluster overlap, or ‘fuzzify’ the cluster boundaries made k-means clustering an excellent option for this project’s goal of extending the approachability of admixture estimation.

Package selection. Once the appropriate clustering method was determined, published R packages were examined for applicability and ease of use. The CRAN site maintains a list of available packages written and published by individuals or groups in several fields, from horticulture to psychology. A search of that site provided five packages likely to fit the needs of this research: ppclust, FuzzyR, fclust, R Commander with the fuzzy numbers extension, and FANNY.

The package ppclust (Cebeci, et al. 2018) provides a range of probabilistic and possibilistic cluster analysis algorithms, including fuzzy and hard c-means clustering, fuzzy possibilistic product partition c-means clustering, and modified fuzzy possibilistic c-means clustering. Because of the variety of algorithms provided, ppclust is an excellent package for those well-versed in the differences between the different clustering algorithms, as the differences are nuanced but important depending on the type and amount of data being tested; this same variety, however, makes it a difficult package to be used by beginners to both cluster analysis and R programming.

The second package explored, FuzzyR, ultimately suffered from many of the same setbacks as ppclust—namely, an abundance of algorithmic options with the potential to overwhelm novices to the field. However, FuzzyR has one addition which works in its favor despite the number of algorithms present: a graphical user interface (GUI). The software program of R is a command line program, meaning that each line is

its own command enacted by the program, like Figure 2. GUIs utilize graphics, such as windows, buttons, or boxes, to provide a visual interface which controls the program; these have become the norm in software programs, such as the Windows or Mac software GUIs (see Figure 3). GUIs lessen the tension of computer use and make novice users feel more comfortable (Levy 2018).

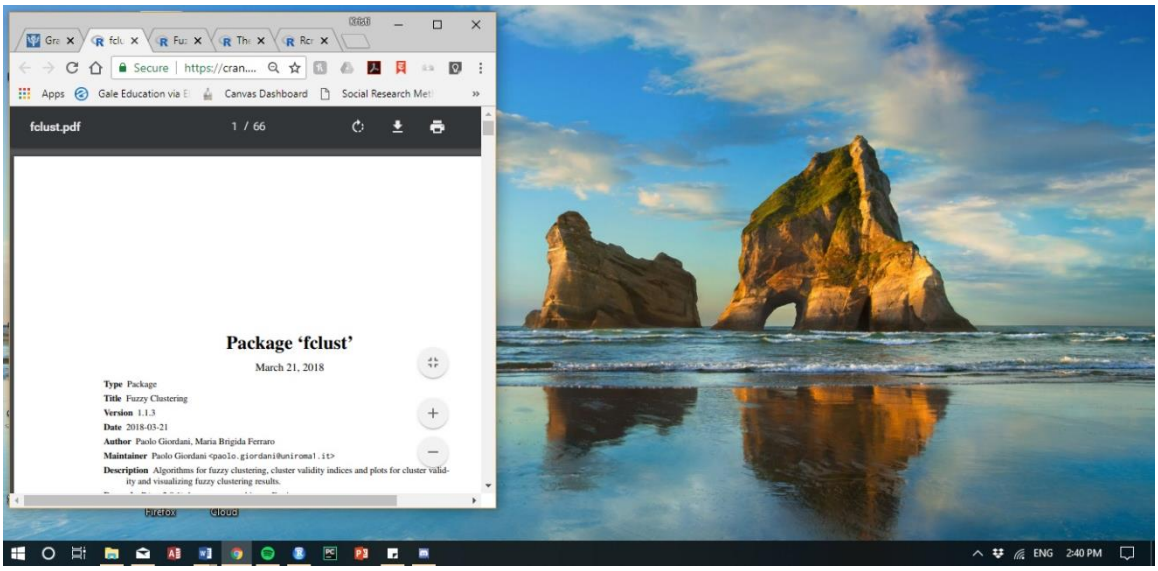


Figure 3 Windows with the GUI of Google Chrome

The third available package, fclust, focuses on fuzzy k-means cluster techniques. It includes a basic fuzzy k¬-means algorithm (FKM) along with 15 variations of the algorithm designed for specific purposes. This package is a line-command only (Giordani and Brigida Ferraro 2018). The ability to choose only the fuzzy k-means algorithm made this package more appealing than the previous two mentioned, but the variety of other algorithms and the lack of GUI made it less appealing than the next package option, RcmdrPlugin.FuzzyClust.

The fourth package was explored for this research based on two criteria: GUI interface and limited choice of algorithms. This package requires the use of an interface

package called, Rcmdr, to function. R Commander is a GUI for R that is regularly updated, as well as modifiable for specialized tasks through the use of plug-in packages, such as the fuzzy clustering plug-in used in this research (Fox and Bouchet-Valat 2017). Once both packages are installed and the R Commander GUI initiated (see Figure 4), the fuzzy cluster plugin is engaged, giving the GUI the ability to complete the two fuzzy cluster algorithms included in the program: fuzzy k-means and Gustafson Kessel.

```
1  install.packages(Rcmdr)
2  install.packages(RcmdrPlugin.FuzzyClust)
3  library(Rcmdr)
4
5
```

Figure 4 Code to Install Rcmdr and RcmdrPlugin.FuzzyClust Packages, and Start the R commander GUI

The intuitive, easy-to-use design of R Commander, as well as the limited options of the fuzzy cluster plug-in, make it an excellent choice for novices (see Figure 5).

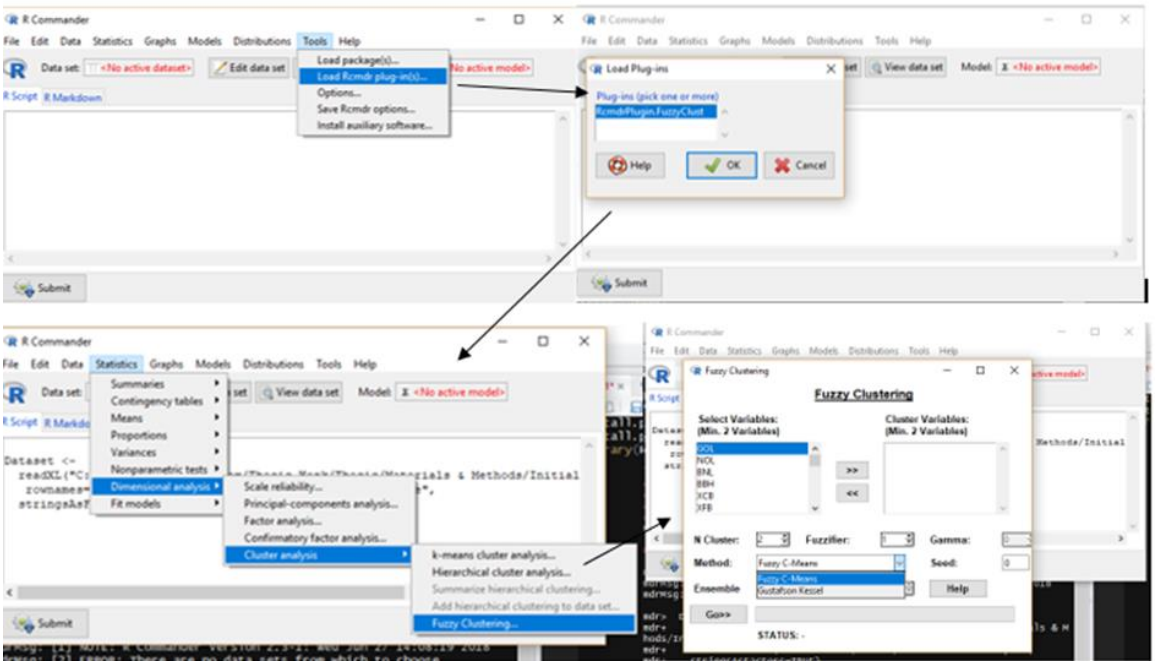


Figure 5 Initiating the Fuzzy Cluster Plugin with the R Commander GUI

The R Commander fuzzy cluster allows for customization of cluster numbers and fuzzifier factor, as well as individual determination of variables used to create the clusters. For this research, all available craniometrics were used as variables. The data was tested starting with the cluster number at three and increasing to eight; the fuzzifier factor, which determines the amount of allowable group overlap, was kept stable at two for all tests.

Once the analysis is completed, a results screen is generated with a visual scatterplot based on the first two principal components analyses, the table of cluster membership probabilities, cluster centroid information, and statistical tests of validation and group differences (multivariate analysis of variance or MANOVA) (see Figure 6).

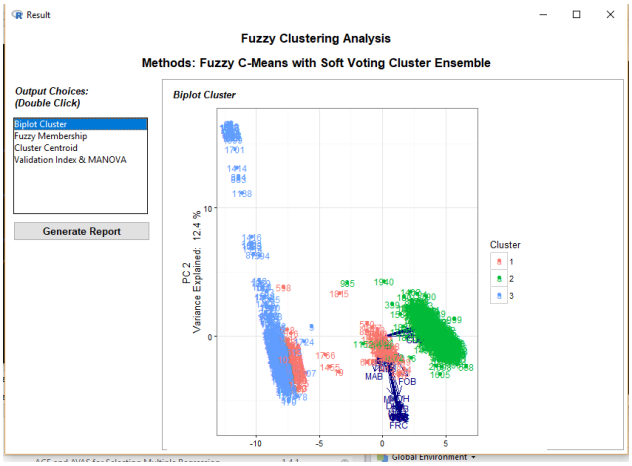


Figure 6 Fuzzy Cluster Results GUI with Report Generator

The program automatically saves generated reports as text documents; the charts are later transferred to a spreadsheet for further analysis.

The Rcmdr.FuzzyPlugin was the most visually user-friendly option because of the GUI used and provided a wealth of information about the data, but the program overall was less than ideal because of the manner of data output and the difficulty in adequately and accurately identifying centroids and any surrounding individuals. Centroids of the

individual clusters were given as the average numbers for each measurement, requiring the author to separate out all the individuals in a cluster then try to find individuals with similar numbers to average to determine ancestry identification for each cluster. This could end up with anywhere from three to six individuals, none with more similar numbers than any other, making it difficult to determine which one(s) would be closest to the true centroid. Easy identification of the centroids was essential to cluster assignment and therefore tests of validity, stability, and reliability, so this method was given high potential for use in this research, but the author wanted to explore the last previously formulated coding collection, FANNY, as a possible better fit.

Within the ‘cluster’ package is a class, or subset of programming, specific to fuzzy cluster analyses, known as FANNY. An immediate downfall of FANNY is that it is a line command program and not a GUI. However, the coding necessary to obtain results is less intensive and has fewer options (overall) than previous line command package options. Additionally, exporting results from FANNY can be individualized to only the information needed by the researcher, instead of having to save everything, as was done in the Rcmdr.FuzzyPlugin exports. Finally, the programming can be utilized by the ‘fviz_cluster’ class within the ‘factoextra’ package for visualization purposes; this is an improvement over the data visualization from Rcmdr.FuzzyPlugin because fviz_cluster allows individualization of the data plots, including the inclusion of identified cluster centers instead of having to attempt to find individuals closest to the centroid measurements as was done by Rcmdr.FuzzyPlugin. For these reasons, the author felt the FANNY class of the cluster package was the best option for testing purposes for this research project.

FANNY allowed the individualization of several details pertinent to fuzzy analyses, including the number of desired clusters (where *k* must be greater than 0 but no more than half the number of observations included in the analysis), the ability to use dissimilarity matrices or standardized data instead of observed variable matrices, different calculating metric (Euclidean, Manhattan, or squared Euclidean), specifying the maximum number of testing iterations (default set to 500), and more. This research tested the capabilities of this program to separate the data into between three and seven clusters. Because the data used was actual observational data, it was not considered standardized or a dissimilarity matrix, so the author included information to reflect this in the coding (stand = FALSE). According to the package reference manual, the squared Euclidean metric "is equivalent (but somewhat slower) to computing so called 'fuzzy c-means'" (Maechler, et al. 2019, 38), so the author specified this metric (metric="SqEuclidean") as seen in Figure 7.

```
13
14  library(cluster)
15  ISK3A<-fanny(InitialSample, 3, stand = FALSE, metric="SqEuclidean")
16  ISK3A
17
```

Figure 7 Coding for FANNY Class Testing of the Initial Sample into 3 Clusters

Results from testing was then exported using the 'xlsx' package. This package has multiple functionalities, including use in reading and importing, creating and exporting to, and formatting existing Microsoft Excel files. It also allows the user to import multiple different datasets into the same Excel file using the "append" function (append=TRUE), which creates a new sheet within the file and writes the data there. Using this package, the author was able to export all FANNY results into a single file for each sample and number of clusters tested (see Figure 8).

```
library(xlsx)
write.xlsx(ISK3A$data, "ISK3.xlsx", sheet="data")
write.xlsx(ISK3A$membership, "ISK3.xlsx",sheet="membership", append=TRUE)
write.xlsx(ISK3A$coeff, "ISK3.xlsx", sheet="coeff",append=TRUE)
write.xlsx(ISK3A$memb.exp,"ISK3.xlsx", sheet="membexp", append=TRUE)
write.xlsx(ISK3A$clustering,"ISK3.xlsx", sheet="clustering", append=TRUE)
write.xlsx(ISK3A$k.crisp,"ISK3.xlsx", sheet="hardcat", append=TRUE)
write.xlsx(ISK3A$objective, "ISK3.xlsx", sheet="objective", append=TRUE)
write.xlsx(ISK3A$convergence, "ISK3.xlsx", sheet="convergence", append=TRUE)
write.xlsx(ISK3A$silinfo$widths, "ISK3.xlsx", sheet="silinfo",append=TRUE)
write.xlsx(ISK3A$silinfo$clus.avg.widths,"ISK3.xlsx", sheet ="clusavgwid",append=TRUE)
write.xlsx(ISK3A$silinfo$avg.width, "ISK3.xlsx",sheet="avgwid",append=TRUE)
```

Figure 8 Coding for Data Export using 'xlsx' Package

Finally, results were visualized using the ‘fviz_cluster’ class of the ‘factoextra’ package. This package allowed significant personalization of each graph to make the data more comprehensible (see Figure 9).

```
library(factoextra)
fviz_cluster(ST23, show.clust.cent=TRUE, repel=TRUE, pointsize=0.5,labelsize=2)
fviz_silhouette(ST23)
```

Figure 9 Coding for Data Visualization using the 'factoextra' Package

This included plotting of the cluster center (show.clust.cent=TRUE), which made identification of individuals closest to the centroid more manageable than previous programs explored. It also allowed the author to use different sizes for the individual points within the plot as well as their labels, to make them more easily visualized (pointsize=0.5, labelsize=2). Finally, the program has a “repel” function (repel=TRUE) which, when activated, keeps the program from “overplotting” or putting in every text label for each point. This, again, makes understanding key aspects of the data easier for the user. The cluster plots for the five variable testing (discussed later) can be found in Appendix A.

Validity, Stability, and Reliability Testing

While traditional and fuzzy k-means clustering are well-documented and researched statistical methods, each application of an analytical method to new data must

address three major concerns—validity, stability, and reliability—before researchers can approach interpretation.

Validity. Concurrent validity is the ability of a method to differentiate between the groups for which it was developed to discern (Trochim 2006). Because there is no way to trace full ancestry for any person or group, especially considering rapidly increasing globalization, the groups (classes) available to validate the methodology are social race groups. One's self-identification with a given racial group does not equate to any single geographic ancestry population, but research has revealed patterns within the U.S. regarding biogeographic ancestry, population histories, and racial self-identification (Ousley, Jantz and Freid 2009, Algee-Hewitt 2017a). While not a perfect measure and seemingly counter-intuitive, as one attraction of admixture estimation is the potential to move away from hard classification, it was the only categorization available for cluster validation.

Fuzzy k-means clustering is a form of 'unsupervised' clustering, meaning that it is completed without class identifiers, to ensure the relationships are based off the variable data and not influenced by class data. Utilizing the class of the cluster centroid, each data point was identified by that label (as, for example, predominantly White or Black ancestry) (Everitt, et al. 2011). For points of overlap, labels were assigned according to the highest posterior probability. These were then compared to the recorded class labels to determine the accuracy, or validity, rate. Finally, the relationship of this accuracy rate to other published accuracy rates—in this case, Algee-Hewitt's (2016) finite mixture analysis of 71 – 75% mean matching accuracy—was explored as a means of ensuring a minimum level of accuracy (Kubat 2017).

Reliability. Reliability is the ability of the method to produce consistent results over time. As such, reliability is tested by "critically revisiting and replicating the clustering results at a later point in time", preferably with a newly collected dataset (Sarstedt and Mooi 2014, 260). Large datasets of newly collected craniometrics are difficult to obtain, leaving three options to test reliability: 1) running the entire data set at a later date, 2) using random sampling of the entire data set to create a test bank, or 3) using Howells' world-wide craniometrics data set as an alternative data set, though not newly collected. The author tested the initial reliability of the algorithm, regardless of its validity, by running the initial sample through the program on different dates using the same code and comparing the number of individuals in each cluster to ensure the underlying algorithm was reliable (regardless of validity). The algorithm was found to be reliable to itself using the initial sample dataset, allowing the author to focus on ensuring it would then be reliable in how it was categorizing individuals using a different method.

Given how temporally expansive Howell's dataset is (extending from prehistoric groups to the 1970s), there are issues with the potential for craniofacial changes over time impacting the accuracy of the results. Because of these issues, this research used a random subsampling of 500 individuals to use as a test databank, created using the imbedded random number generator (RNG) in R (see Figure 10).

```
library(readxl)
s1<-read_excel ("C:/Users/Kristi/Dropbox/Thesis Work/Thesis/Materials & Methods/Sample.xlsx")
view (s1)


set.seed(124)

index <-sample(1:nrow(s1),500, replace=F)
index
t1<-s1[index, ]
t2<-s1[-index, ]

write.csv(t1,file="TestBank.csv")
write.csv(t2,file="InitialSample.csv")

index2<-sample(1:nrow(t2), 712, replace=F)
st1<-t2[index2, ]
st2<-t2[-index2, ]
write.csv(st1,file="SplitTest1.csv")
write.csv(st2,file="splitTest2.csv")
```

Figure 10 Use of Imbedded Random Number Generator for Subsample Creation

To get a sampling that is replicable in the future, the author set a "seed", or starting number, for use by the RNG algorithm. This allowed any future testing to start at the same point in the data and pull the sample according to the same algorithm, improving the replicability of future testing. The author then created an "index" command, which tells the program to use the RNG to sample the entire initial sample (labeled as "s1" for this program) for a total of 500 individuals, and to not replace drawn numbers back into the pool of potential sample subjects to prevent individual duplication within the new subsample. This test bank was labeled as "t1" and subsequently downloaded into a file named "TestBank.csv". The remaining sample was compiled into a sample bank labeled "t2" using the negate function of the index, which told the program to use all individuals not included in the original index. This bank (t2) was then downloaded into a file named "InitialSample.csv". This file was used for initial testing and was split again for testing of methodological stability, discussed in the next section.

Stability. Stability is the ability of the method to consistently identify and cluster the same (or similar) individuals. There are multiple ways to test stability of a method, but the two most common methods are the inter-methodology comparison and split

datasets. For inter-methodology comparison, the researcher runs the given dataset through a different clustering method and compares cluster affiliations. Significant change in cluster affiliations (more than 20% difference is the widely accepted standard) should trigger a reassessment of the clustering methods, variables used, and program set-up (Kubat 2017). The second method to ascertain stability is the 'split dataset' method, in which the dataset under question is split into two separate subsets and run through the program. Cluster centroids are then compared using traditional comparative methods (independent samples t-test and/or analysis of variance [ANOVA]) for significant differences. This research used the latter procedure—splitting the initial dataset used—to explore the stability of fuzzy k-means clustering methods on datasets of different sizes (Sarstedt and Mooi 2014).

Summary

Testing for the validity, stability, and reliability of statistical methods required a sample of adequate number and, hopefully, physical diversity, while controlling for issues such as the significant cranial plasticity that has been demonstrated over hundreds of years. For this reason, the sample used for this research was restricted to more modern samples; this allowed for control over plasticity over time while allowing for an excellent sample size. The sample for this research was compiled from multiple open-source compilations, including UTK's Forensic Data Bank and NCSU's online craniometric database of Latin American individuals. The sample was the parsed using Microsoft Access from the original 2,539 individuals to the sample of 1,924 used for testing by eliminating individuals that did not have five measurements (TKB) that have been shown to have the most craniofacial variability. Further random sampling into testbanks and

split databases utilized the imbedded random number generator found within the R environment.

The author explored five different packages within the R environment for their ability to perform fuzzy k-means clustering tests, their ease of use, and the effort involved in determining their validity, stability, and reliability from their results and export methods. From the five packages (ppclust, fclust, FuzzyR, Rcmdr.FuzzyPlugin, and FANNY), this research utilized FANNY in conjunction with the 'xlsx' and 'factoextra' packages for testing, data export, and results visualization. While FANNY is a line command programming option, it is easy to use, with clear instructions for which functions and metrics to use for the data type (observed variables) and statistical method (fuzzy k-means) used in this research. Data was easily exported using the 'xlsx' package into single Microsoft Excel files for each sample and cluster number. Data was then visualized using the 'fviz_cluster' class of the 'factoextra' package, which allowed for individualization of each graph to fit its needs.

CHAPTER IV – Results and Discussion

Admixed ancestry determination as calculated using fuzzy math could have applications across biological anthropology, from understanding historical population movements to forensic identification of individuals. Currently there are multiple pre-coded, readily available statistical packages that are readily accessible for researchers without a strong background in coding and statistical computations. However, they still need to be tested for their ability to accurately differentiate between groups using criteria applicable to bioanthropological studies, such as the craniometrics used in this research. As such, the goal of this research was to determine the accuracy of FANNY, a readily available package within the R environment, in clustering groups into ancestral groups using a series of cranial measurements. An extensive databank of craniometrics was sourced from multiple databases stored either online for open use (such as the North Carolina State University Forensic Analysis Lab) or available via email, such as the FDB. Testing databases were created and clustered using the FANNY and cluster charting packages. FANNY does not allow blank variable fields, so means were determined based on all available data for the variable and input for all missing variables. Centroids were identified by determining the ancestry as indicated in the initial database for at least three individuals closest to the centroid and used to label clusters as based on majority rule.

After the initial testing, it became evident that other variables impacting the accuracy of the results needed to be explored, namely the high number of variables initially used (22) and the disproportionate number of White individuals in the sample (924 out of 1,424). Consequently, further databases were created with fewer variables (5) and more proportionate ancestry representations and subjected to the same testing

procedures with variable results on the accuracy rates. The implications of these findings are discussed in more detail below.

Results

To explore the effectiveness of pre-coded, fuzzy math statistical programs in assessing admixture, a broad base sampling of 2,539 was compiled from UTK's FDB and NCSU's open-access database online. It was then parsed down from 2,539 to 1,924 individuals using five measurements previously determined by Jantz and Jantz (2000) to show the most craniofacial variation with minimal impact without being significantly impacted by cranial plasticity: glabello-occipital length (GOL), basion-bregma height (BBH), maximum cranial breadth (XCB), bizygomatic breadth (ZYB), and nasion-prosthion height (NPH). These measurements were the only ones mandated in the creation of the databank, meaning any of the other 22 variables could be missing for the remaining sample. However, the package used for testing, FANNY, does not allow blank variable fields. In the first attempt to overcome this, an average of all available data for each variable was compiled then input in all blank fields. This choice likely influenced grouping choices made by the program, contributing to the high number of clusters designated as White.

The FANNY package was used to cluster the initial sample of 1,424 individuals into three, four, five, six, and seven different groups (see Table 7 for total number of individuals for each cluster by test). The k-6 testing, which limits the total results groups to six, returned hard clustering of only five groups, despite giving posterior probabilities of cluster membership for six clusters for each individual; no individual or group in this test scored a sufficiently high likelihood of membership in the sixth group (i.e., none

were sufficiently different enough from membership in the other five groups) to be 'hard clustered' into the sixth group. All other tests returned the designated number of groups, but it should be noted that the final cluster (7) in the k-7 test contains only one individual—number 181. Apart from the k-3 test, cluster 1 was the largest cluster; the cluster size for clusters 1 and 2 were approximately the same in the k-3 test.

Table 7 Total Number of Individuals in Each Cluster, by k-test, for the Initial Sample

Test/Cluster	1	2	3	4	5	6	7
K-3	532	574	318	—	—	—	—
K-4	658	427	333	6	—	—	—
K-5	692	253	123	354	2	—	—
K-6	527	410	318	164	5	—	—
K-7	700	289	30	320	75	9	1

This method demonstrated a large skew towards one cluster grouping over all the others; in every test except k-3, the first group included the largest number of individuals. While not unexpected—there was a higher probability that the program would get a White individual over any other individual due to the high skew—it is worth mentioning due to the presence of the skew. The program noted not only that many individuals belonged together, but also likely began grouping them from the start of the run. FANNY also demonstrated this recognition across almost all initial sample testing—the largest group was consistently the first one, and, as demonstrated later, identified as White per centroid. Additionally, in the k-3 test, the first cluster was only marginally smaller than the second and largest cluster. As will be discussed later, the first cluster was also consistently designated as White. As such, the test did, in some respect, recognize the skew towards White individuals within the sample and grouped them accordingly.

Identifying the overall cluster label for ancestry was complicated by the overwhelming presence of White identified individuals in the sample. For all five tests, cluster 1 was overwhelmingly White, with over 70% of the cluster identified as such. Additionally, and unsurprisingly, the centroids also designated cluster 1 as White across all tests on the initial sample. Twenty-one of the total 24 clusters were at least 50% or more White; the three exceptions were cluster 4 in the k-4 test (n=6) and clusters 3 and 7 in the k-7 test (n=30 and 1, respectively).

Table 8 Breakdown of Races Included in Each Cluster for the k-7 Test--Initial Sample

Race	**1**	**2**	**3**	**4**	**5**	**6**	**7**
B	132 (18.9%)	63 (21.8%)	12 (40.0%)	77 (24.1%)	25 (33.3%)	2 (22.2%)	0.0%
B/NA	2 (0.3%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
EA	1 (0.1%)	3 (1.0%)	0.0%	5 (1.6%)	0.0%	0.0%	0.0%
H	54 (7.7%)	45 (15.6%)	5 (16.7%)	36 (11.3%)	10 (13.3%)	3 (33.3%)	0.0%
NA	6 (0.9%)	6 (2.1%)	1 (3.3%)	4 (1.3%)	0.0%	0.0%	0.0%
W	502 (71.7%)	171 (59.2%)	12 (40.0%)	193 (60.3%)	40 (53.3%)	4 (44.4%)	1 (100.0%)
W/B	2 (0.3%)	0.0%	0.0%	1 (0.3%)	0.0%	0.0%	0.0%
W/EA	0.0%	0.0%	0.0%	1 (0.3%)	0.0%	0.0%	0.0%
W/H	0.0%	1 (0.3%)	0.0%	3 (0.9%)	0.0%	0.0%	0.0%
W/NA	1 (0.1%)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Only two of the 24 clusters identified across the five tests had individuals easily designated as centroids: clusters 5 and 7, both of the k-7 test. Of these clusters, Cluster 5 was the only one with more than one individual in the cluster; the k-7 test singled out individual 181 as its own cluster (7), making it both the centroid and the entire cluster. For the remaining clusters, up to three of the individuals closest to the centroid were

recorded. The majority ancestry was used when possible; otherwise, the clusters were designated as "admixed" if consistent ancestries were not noted.

Table 9 Race Designation by Cluster--Initial Sample, 22 Variables

Test/Cluster	1	2	3	4	5	6	7
K-3	White	White	Admixed	—	—	—	—
K-4	White	White	White	White	—	—	—
K-5	White	White	Black	White	White	—	—
K-6	White	White	White	White	White	—	—
K-7	White	Hispanic	Admixed	White	White	Hispanic	White

The variation and spread of White individuals in each test technically invalidated this method, without needing to test on smaller samples or for stability and reliability. In all tests except k-7, the majority of clusters were designated as White by centroids, leaving over 30% of the sample misidentified. Of the 24 identified clusters, only four of them had a non-White designation—cluster 3 in the k-3, k-5, and k-7 tests, and cluster 2 in the k-7 test. This raised three issues that may impact validity that needed to be explored: the effect of using database-wide means as substitutes for missing variables, overfitting (i.e., too many test variables), and representative sample proportions.

Issues Impacting Testing Validity

Insisting everyone included in the sample had every measurement (over the allotted five) would have decreased the sample, potentially to levels too small to be statistically significant. A different way to deal with the blank datapoints would have been to substitute mean values based on designated ancestry associations (when possible), either 1) based on values averages from the data provided within the database or 2) based on averages by ancestry for all the measurements included in the 22-variable database published in other studies using the FDB published averages. Addressing the latter, no such publications could be easily accessed.

Addressing the former, there were two issues. First, there were no group averages possible for some of the measurements, because the collective group had no measurements for those variables. Additionally, this issue had already been corrected, to some degree, with the adjusted database created with only five variables to address the possibility of overfitting (discussed later). Finally, there would have been a question of how to best determine the mean for individuals with designated admixture (e.g., East Asian/White or Native American/White), but whose admixture groups did not have enough individuals to make an average measurement calculate a reliable mean. One possible way to create idealized means for these purposes would have been to determine averages for each of the two separate groups involved in the admixture (e.g., East Asian and White), but it was unclear at the time how accurate this method might have been in reflecting the actual averages of those admixed communities. Given that the testing on the initial sample never proved adequately valid, further testing on reliability and stability would have been unnecessary. However, the patterns seen in testing of the initial sample and subsequent subsegments of the initial sample revealed issues with the database formation and sample selection, the most significant of which is the disproportionate number of White individuals within readily available research databases such as the FDB. These issues are important to understand not only for the use of data mining and soft computing techniques in bioanthropology, but to any research done using freely available databases such as the FDB going forward.

The second concern addressed was the question of variable impact on validity, or the idea that the sample uses too many variables (the 22 different measurements) which may unnecessarily obscure the test data. This issue, known as overfitting, has been

documented in previous craniometric work with by Hefner and Ousley (2014) and Monsalve and Hefner (2016). Determining the precise number of variables to use to minimize overfitting while maximizing the variability within the available data has been an ongoing dilemma for research such as this anthropological research. Given the potential skew issues arising from the use of database-wide averages to fill missing measurements, the only option to avoid this issue while checking for overfitting was to use only the initial five variables employed to parse the full database into a testable database. Therefore, to address this concern, a variable-limited version of the initial sample of 1,424 individuals was created using five measurements determined by Jantz and Jantz (2000) to show the most craniofacial variation without being significantly impacted by cranial plasticity: glabello-occipital length (GOL), basion-bregma height (BBH), maximum cranial breadth (XCB), bizygomatic breadth (ZYB), and nasion-prosthion height (NPH). The new variable-limited initial sample set was subjected to the same series of FANNY tests as the original sample—attempting between three and seven clusters (see Table 4 for total number of individuals for each cluster by test and Appendix A for representative cluster plots).

Table 10 Total Number of Individuals Per Cluster by Test Using the Five Variable Initial Sample

Test/Cluster	1	2	3	4	5	6	7
K-3	486	443	495	—	—	—	—
K-4	331	384	339	370	—	—	—
K-5	249	319	293	277	286	—	—
K-6	241	276	247	207	209	244	—
K-7	147	260	199	182	233	188	215

Limiting the variables, interestingly, resulted in more evenly distributed groups, despite the known numerical skew of White individuals (see table 4). The ability of the

variable limiting to create more evenly distributed groups added support for the investigation of this method using a more proportionate sample. The more proportionate groupings meant that no group was comprised of less than one hundred individuals, but White individuals still dominated since every cluster in every test was at least 50% White. Table 5 shows the count and percentage breakdown of the clusters in the k-7 test of the five-variable initial sample, demonstrating the continued impact of the high number of White individuals on this test.

Table 11 Breakdown of Races by Cluster for the k-7 Test--Five Variable Initial Sample

Race	**1**	**2**	**3**	**4**	**5**	**6**	**7**
B	19 (12.93%)	30 (11.54%)	43 (21.61%)	34 (18.68%)	79 (33.91%)	40 (21.28%)	60 (30.70%)
B/NA	0.00%	1 (0.38%)	0.00%	1 (0.55%)	0.00%	0.00%	0.00%
EA	0.00%	0.00%	2 (1.01%)	1 (0.55%)	0.00%	4 (2.13%)	2 (0.93%)
H	6 (4.08%)	12 (4.62%)	20 (10.05%)	35 (19.23%)	15 (6.44%)	39 (20.74%)	26 (12.09%)
NA	1 (0.68%)	1 (0.38%)	1 (0.50%)	5 (2.75%)	2 (0.86%)	4 (2.13%)	3 (1.40%)
W	120 (81.63%)	216 (83.08%)	132 (66.33%)	104 (57.14%)	137 (58.80%)	99 (52.66%)	115 (53.49%)
W/B	1 (0.68%)	0.00%	1 (0.50%)	1 (0.55%)	0.00%	0.00%	0.00%
W/EA	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1 (0.47%)
W/H	0.00%	0.00%	0.00%	0.00%	0.00%	2 (1.06%)	2 (0.93%)
W/NA	0.00%	0.00%	0.00%	1 (0.55%)	0.00%	0.00%	0.00%

Using the same method of identifying at least three individuals closest to the centroid and basing cluster designation on the majority rule, we saw more clusters designated as non-White (seven out of 25 clusters as compared to the five out of 24 clusters of the initial method). However, this method did not solve the immediate validity issue, as the spread of White individuals across all groups continued to skew centroids towards White and impact the accuracy rate of the method (see Table 6). Even with an

increased number of groups designated as non-White, over 30% of the sample would be incorrectly identified between misidentified White individuals in the non-White clusters and misidentified non-White individuals in the White clusters.

Table 12 Race Designation for Each Cluster using Centroids--Five Variable Initial Sample

Test/Cluster	1	2	3	4	5	6	7
K-3	White	Black	Admixed	—	—	—	—
K-4	White	White	White	White	—	—	—
K-5	Admixed	White	White	White	White	—	—
K-6	Black	White	White	White	White	White	—
K-7	White	Black	Black	Black	White	Admixed	White

Decreasing the variables created two interesting results across the cluster testing. The first and most noticeable result was that across all cluster tests, the number of individuals included in each cluster was more evenly distributed than with the larger variable sampling. Interpreting this change was difficult; the biggest concern was that decreasing the variables decreased the tests sensitivity to the skew of the White population within the sample. However, this could also have been a result of eliminating the potential skew issues inherent in taking the average of the measurements from a database whose composition already skewed heavily to White individuals.

The second result noted from the limited variable testing was the slight increase in clusters identified as non-White. In the initial 22-variable testing, only 5 out of 24 clusters were identified as non-White, while the limited, 5-variable testing designated 8 clusters out of 25 clusters as non-White. For both databases, the k-4 cluster test produced only White centroids. However, the limited variable testing designated at least one non-White centroid for every other test, with the k-3 test having 2 out of the 3 clusters (up an

increase from 1 out of 3) as non-White and k-7 designating having 4 non-White clusters out of 7 (an increase from 3 out of 7). When combined with the more even distribution of individuals among groups in this restricted variable testing, the higher number of non-White clusters in the k¬-3 and k-7 tests pushed the number of misidentified individuals from non-White to White; the two clusters of non-White individuals in the k-3 testing included a total of 938 individuals, while the four clusters in the k-7 testing included 829 individuals. For each of these tests, at least 50% of the White individuals in the data set were identified as non-White instead of as White, effectively mirroring the misidentification issue seen with the 22-variable testing results. This meant that, while the increased number of non-White-identified groups was initially encouraging to see, its importance was negated somewhat for the k-3 and k-7 test groups by the even distribution of individuals in the group, which decreased accuracy and therefore the validity of the test.

The skew created by the disproportionate representation in the sample impacted the validity in both previous tests, using 22 variables and 5 variables, respectively. This suggested that a possible solution to the issue of validity could be the creation of a more proportionate testing sample. However, the small number of individuals in the Native American and Hispanic groups within this compiled database prevented the creation of a sample base that was both proportional across all ancestry groups and of an adequate sample size (preferably, greater than 100 individuals). This issue was a significant drawback throughout the research, and at this point became a deterrent unable to be overcome by this research project.

CHAPTER V – Conclusions and Future Directions

The focus of this research was assessing the validity of the FANNY package in providing reasonably accurate (70% or more) ancestry determinations. A craniometric database was created from the Forensic Databank developed and maintained by the University of Tennessee-Knoxville and the online database of craniometrics from various Latin American nations maintained by North Carolina State University. The combined database of over 2,000 individuals was refined to ensure every individual had a baseline five craniometrics (GOL, BBH, XCB, ZYB, NPH) that were previously shown as useful for identifying craniofacial variation without being greatly affected by more temporary cranial plasticity (Jantz and Jantz 2000). A random number generator was used to create separate sample and test banks for validity, stability, and reliability testing.

The sample databank was tested using FANNY, a previously validated package for fuzzy *k*-means clustering via the R statistical program. Multiple iterations of the test were completed, looking at the package's ability to cluster into anywhere from three to seven groups, as well as using between 5 and 22 variables to cluster the sample bank. The results of this research demonstrated that, while there is underlying promise to the program, there are multiple issues the program could not overcome, and therefore none of the tests passed the threshold for validity.

When looking from the systemic view, an important issue underlying the validity of the study lies with the significant White skew in the underlying database used for this research. The database was approximately 64% White (932 White individuals out of the 1,424 total). There are multiple possible reasons for the disproportionately high number of White individuals compared to non-White individuals, including the limited number of

readily available/freely accessible metric databases, issues inherent to the post-mortem body donation process, and issues with determination of ancestry on unknown individuals within these databases.

The Forensic Databank is a valuable research resource, widely known in the bioarchaeology and forensic anthropology fields as an excellent source for research requiring metric variables. It is easy to obtain information from the databank, involving only an email detailing the desired data sent to the curators, who in turn prepare an appropriate spreadsheet. North Carolina State University's metric skeletal database is even easier to access than the FDB, as it can be easily downloaded from their website in full. However, these are the only anthropological databases with relevant forensic cases readily available in the US. There are multiple collections of human remains appropriate for forensic anthropology work, especially outside of the United States, but they are not as well-known and require significant internet searching or notation from other publications to find. Once located, they require an application to access, travel to the curation site, and researcher measurement of each cranium. Development of accurate and readily accessible databanks on a global scale would go far to in increasing the POC representation in ongoing research.

A significant portion of the FDB is comprised of individuals whose remains were donated, whether by their stated wish and pre-planning or by the choice of their families after their death. However, they arguably form a highly selective sample. Many individuals may choose not to donate their remains based on an underlying issue not addressed by any of these methods of decedent donation, such as mistrust of the biomedical and medicolegal system. However, peer-reviewed data on the reasons behind

lack of body donation or participation in post-mortem research opportunities in POC communities is nearly non-existent, and this presents its own set of challenges considering the very real impact that it imparts upon biomedical and medicolegal studies. However, there is a small subset of articles which notate the low number of post-mortem donations among minority populations (Goldberg, et al. 2020).

This issue, however, remains a double-edged sword. While the academic and medicolegal research would greatly benefit from inclusion of minority communities specifically to address issues related to them, it should not be done with the purpose of increasing or equalizing numbers in research studies alone. Researchers should engage communities of color to help answer questions they may have about their family and communities that can be answered by our postmortem research, or existing research that can be furthered through postmortem research. Bioanthropologists can also work with rapidly expanding data sets obtained using geomorphometrics, including increasing research on the correlation between geomorphometric measurements and those of dry bones. This would enable, such that we can use of data readily available through CAT scans or other noninvasive techniques taken on the living to further refine and expand our datasets to include more members of minority communities to better serve them in the contexts that we are able.

There is an abundance of future directions that this research study brought to light, likely more than the author can begin to enumerate. However, there are two significant fields of improvement to which the author believes this research should bring attention: variable management and data availability.

Databanks, such as FDB, give an impressive number of measurements per individual whenever available, and the increasing use of three-dimensional measuring techniques and technology, such as geomorphometrics, are soon to make the number of variables practically infinite. Yet there is very little research on which variables or combinations of variables show the most distinctions. As our data sources and technologies change, these should be an ongoing and upfront focus. It has been shown that too many variables are as problematic as too few because of the issue of overfitting. Admixture through globalization, along with other physiological reactions to geographic and climate changes (among other things), will continue to impact the shapes of human bodies and create an ongoing need for routine research on what measurements are showing the highest variability between groups or individuals at any given time. As we look more into newer, technologically based research models such as shape analysis using geomorphometrics, researchers should regularly check for secular changes that impact findings and determine what measurements truly give the greatest variability across the body.

Making databanks readily available and easily accessible should also be a top priority, but neither of these does any good if there is significant skew within the databanks themselves. Post-mortem researchers most often rely on an individual's or family's willingness to donate remains to generate our data, yet the communities we need the most are also the most vulnerable among us. Research into non-invasive methods, such as measurements from computed tomography scans, that can be used accurately for post-mortem analysis contexts would be a high value option. However, the best option would be to work with those most impacted by the subject matter in which we study

(such as discussed in Franklin et al. 2020 and Flewellen et al. 2021) would be a significant step in rebuilding the trust between minority communities and biomedical and bioarchaeological researchers.

Ancestry determination within the field of anthropology is rife with issues. Most widely used ancestry determination methods in anthropology currently place an individual within a single classification, despite the complicating issues spanning historical contexts of population movement, colonization and globalization, rapidly changing climates, subsistence patterns, and the complex interplay between social race and ancestral biogeographic traits. Advanced statistical and computational methods, if they can be validated as stable and reliable for use with the data that anthropological explorations can provide (such as craniometrics), offer an opportunity for a more nuanced exploration of human variation. A more nuanced understanding of human variation can be used as educational tools for the general public, for helping obtain more geographic understanding of biogeographic variation to help more accurately identify forensic cases, and can help to deepen the biocultural understanding of population movement and interactions across space and time.

Methods that can help shed light on any of these complexities are not regularly used by anthropologists due to significant barriers. Lack of knowledge of advances in data mining and computational statistics, lack of knowledge of complex computer coding needed to run the tests, and cost or operational barriers to using user-friendly software made to help the average user are a few of the issues facing anthropologists who could most benefit from the advanced methods. However, we must not put all our faith into more advanced methodology, because despite advances in mathematics, data mining, and

computational statistics, the complexity of human variation may not be something that can be encoded and identified by measurements and machines.

APPENDIX Cluster Plots for Initial Sample Tests K-3 through K-7

Figure A1. Cluster Plot for Initial Sample K-3 Testing with 5 Variables

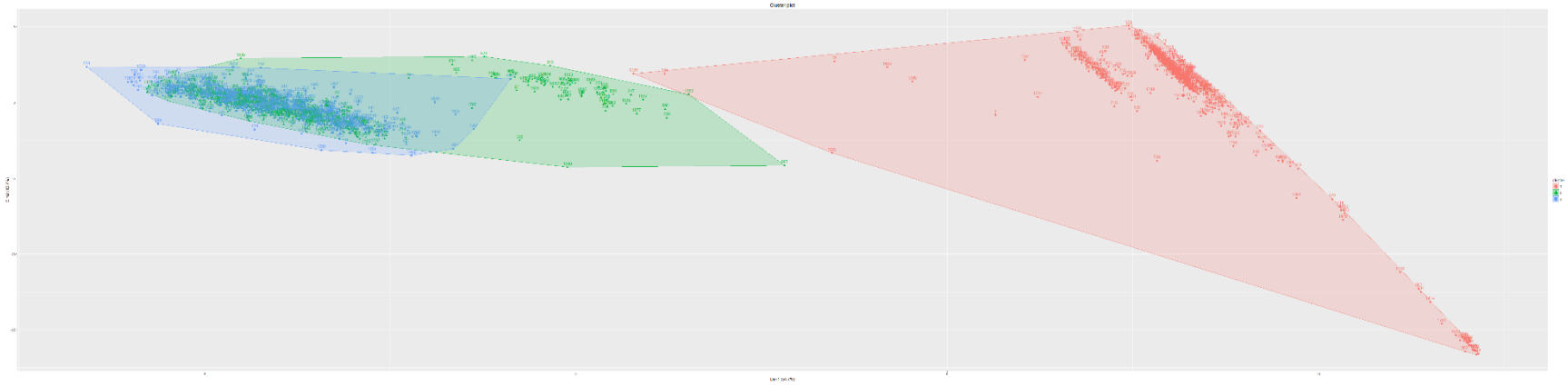


Figure A2. Cluster Plot for Initial Sample K-4 Testing with 5 Variables

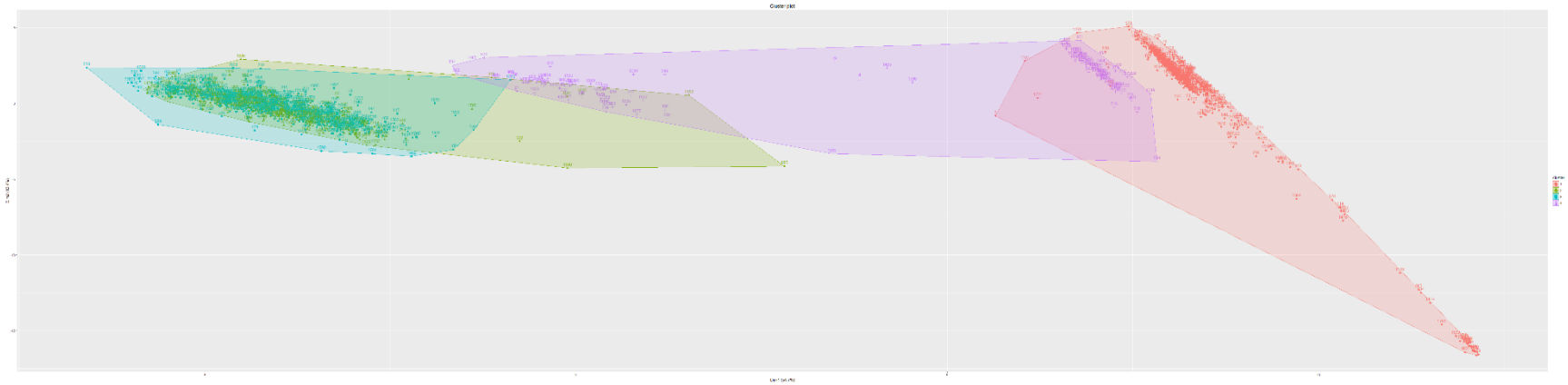


Figure A3. Cluster Plot for Initial Sample K-5 Testing with 5 Variables

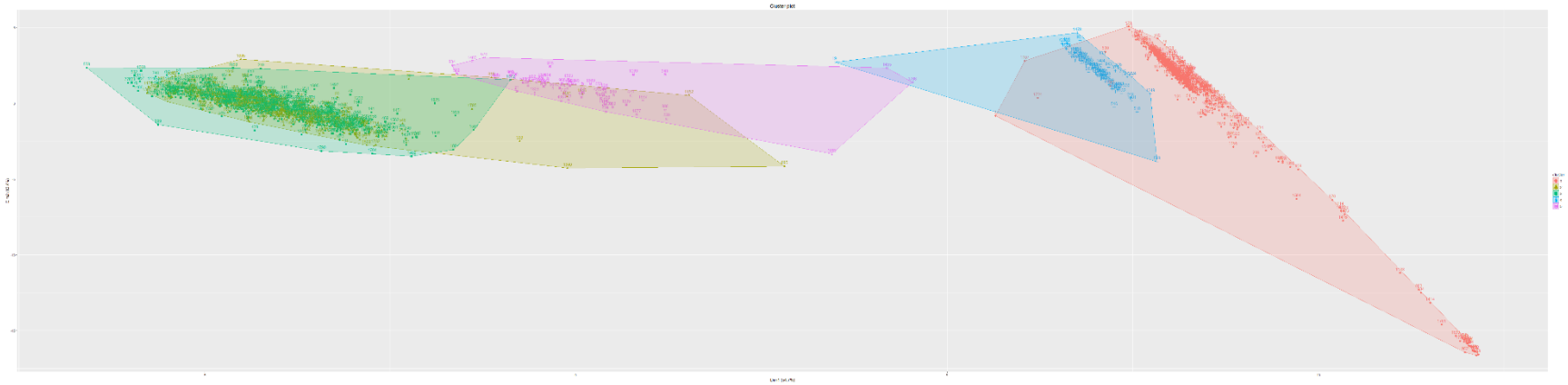


Figure A4. Cluster Plot for Initial Sample K-6 Testing with 5 Variables

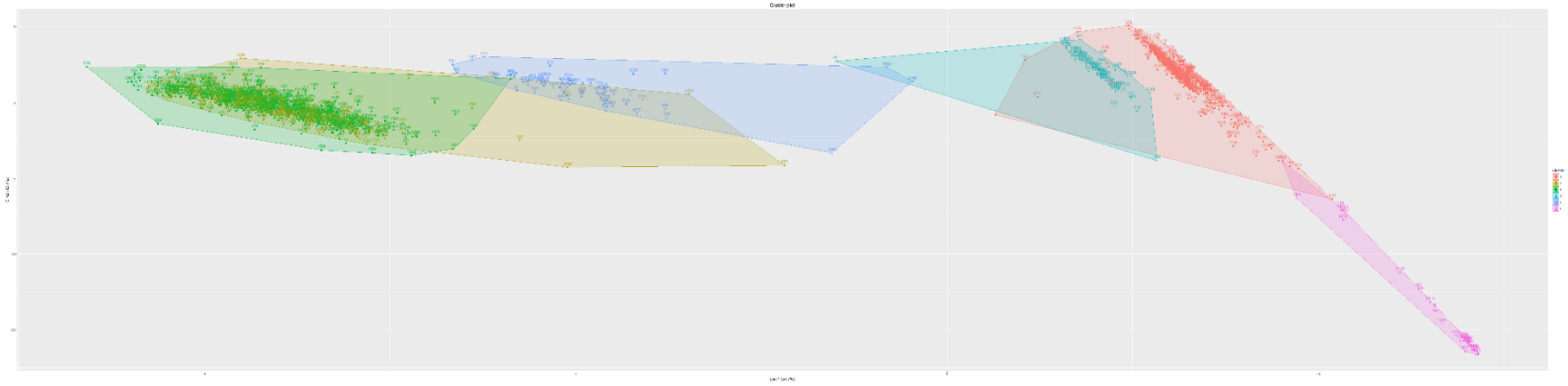
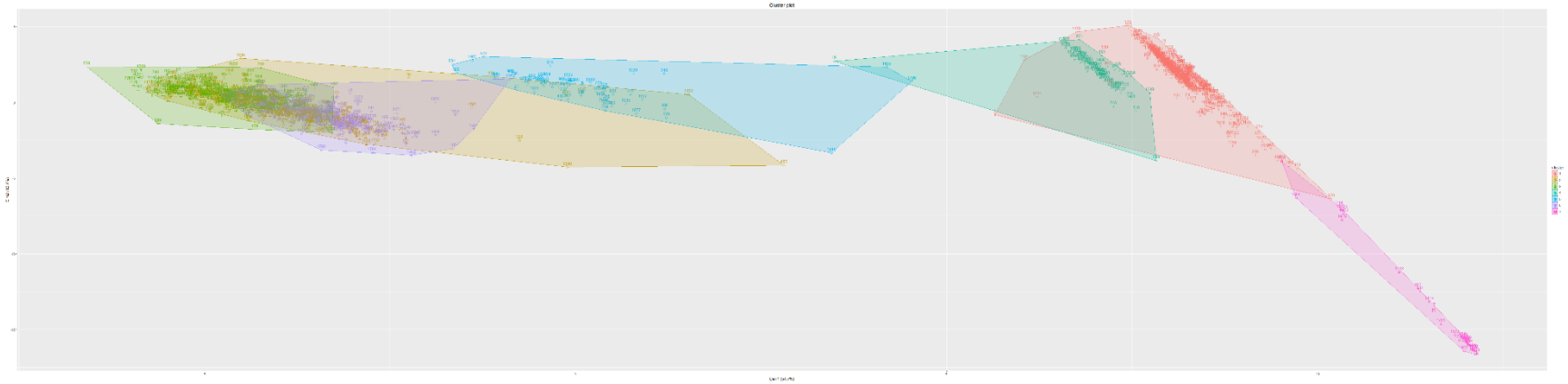


Figure A5. *Cluster Plot for Initial Sample K-7 Testing with 5 Variables*



Works Cited

Algee-Hewitt, Bridget F. B. 2016. "Population Inference from Contemporary American Craniometrics." *American Journal of Physical Anthropology* 160: 604-624.

Algee-Hewitt, Bridget F. B. 2017a. "Geographic Substructure in Craniometric Estimates of Admixture for Contemporary American Populaitons." *American Journal of Physical Anthropology* 164 (2): 1-21.

Algee-Hewitt, Bridget F. B. 2016. "Population Inference from Contemporary American Craniometrics." *American Journal of Physical Anthropology* 160: 604-624.

Algee-Hewitt, Bridget F. B. 2017b. "Temporal Trends in Craniometric Estimates of Admixture for a Modern American Sample." *American Journal of Physical Anthropology* 163 (4): 1-12.

Anderson, Warwick. 2012. "Racial Hybridity, Physical Anthropology, and Human Biology in the Colonial Laboratories of the United States." *Current Anthropology* 53 (S5): S95-S107.

Armstrong-Fumero, Fernando. 2014. "'Even the Most Careless Observer': Race and Visual Discernment in Physical Anthropology from Samuel Morton to Kennewick Man." *American Studies* 53 (2): 5-29.

Ashely-Montagu, M. F. 1940. "A Cursory Examination of the Relations between Physical and Social Anthropology." *American Journal of Physical Anthropology* 26 (1): 41 - 61.

Baab, Karen L., Sarah E. Freidline, Steven L. Wang, and Timothy Hanson. 2010. "Relationship of Cranial Robusticity to Cranial Form, Geography, and Climate in Homo sapiens." *American Journal of Physical Anthropology* 141: 97 - 115.

Balter, Michael. 2017. "The Ethical Battle over Ancient DNA." *SAPIENS.* March 30. Accessed February 27, 2018. https://www.sapiens.org/archaeology/chaco-canyon-nagpra/.

Bigoni, Lucie, Václav Krajícek, Vladimír Sládek, Petr Velemínský, and Jana Velemínská. 2013. "Skull Shape Asymmetry and the Socioeconomic Structure of an Early Medieval Central European Society." *American Journal of Physical Anthropology* 150: 349 - 364.

Brues, Alice M. 1990. "The Once and Future Diagnosis of Race." In *Skeletal Attribution of Race*, edited by George W. Gill and Stanley Rhine, 1-7. Albuquerque, New Mexico: Maxwell Museum of Anthropology.

Buikstra, Jane E., and Douglas H. Ubelaker. 1994. *Standards for Data Collection from Human Skeletal Remains.* Fayetteville, Arkansas: Arkansas Archaeological Survey.

Canter, David, Laura Hammond, and Youngs Donna. 2013. "Cognitive Bias in Line-Up Identifications: The Impact of Administrator Knowledge." *Science and Justice* 53: 83 - 88.

Cartmill, Matt. 1998. "The Status of the Race Concept in Physical Anthropology." *American Anthropologist* 100 (3): 651-660.

Caspari, Rachel. 2009. "1918: Three Perspectives on Race and Human Variation." *American Journal of Physical Anthropology* 139: 5 - 15.

Caspari, Rachel. 2003. "From Types to Populations: A Century of Race, Physical Anthropology, and the American Anthropological Association." *American Anthropologist* 105 (1): 65-76.

Cavalli, Fabio, Luca Lusnig, and Edmondo Trentin. 2017. "Use of Pattern Recognition and Neural Networks for Non-Metric Sex Diagnosis from Lateral Shape of Calvarium: An Innovative Model for Computer-Aided Diagnosis in Forensic and Physical Anthropology." *International Journal of Legal Medicine* 131: 823 - 833.

Cebeci, Zeynel, Figen Yildiz, Alper Tuna Kavlak, Cagatay Cebeci, and Hasan Onder. 2018. "ppclust.pdf." *cran.r-project.org.* January 4. Accessed June 21, 2018. https://cran.r-project.org/web/packages/ppclust/ppclust.pdf.

Chakraborty, Ranajit. 1975. "Estimation of Race Admixture: A New Method." *American Journal of Physical Anthropology* 42: 507 - 512.

Cole III, Theodore M. 1996. "Historical Note: Early Anthropological Contributions to Geometric Morphometrics." *American Journal of Physical Anthropology* 101: 291 - 296.

DiGangi, Elizabeth A., and Joseph T. Hefner. 2013. "Ancestry Estimation." In *Research Methods in Human Skeletal Biology*, edited by Elizabeth A. DiGangi and Megan K. Moore, 117 - 149. Waltham, MA: Elsevier.

DiGangi, Elizabeth A., and Megan K. Moore. 2013. "Introduction to Skeletal Biology." In *Research Methods in Human Skeletal Biology*, edited by Elizabeth A. DiGangi and Megan K. Moore, 3 - 28. Waltham, MA: Elsevier.

Dudzik, Beatrix, and Richard L. Jantz. 2016. "Misclassifications of Hispanics Using Fordisc 3.1: Comparing Cranial Morphology in Asian and Hispanic Populations." *Journal of Forensic Sciences* 61 (5): 1311 - 1318.

Elliott, Marina, and Mark Collard. 2009. "FORDISC and the Determination of Ancestry from Cranial Measurements." *Biology Letters* 5: 849 - 852. doi:10.1098/rsbl.2009.0462.

Ellison, Peter T. 2018. "The Evolution of Physical Anthropology." *American Journal of Physical Anthropology* 165: 615 - 625.

Eveleth, Rose. 2015. "Genetic Testing and Tribal Identity." *The Atlantic.* January 26. Accessed February 27, 2018. https://www.theatlantic.com/technology/archive/2015/01/the-cultural-limitations-of-genetic-testing/384740/.

Everitt, Brian S., Sabine Landau, Morven Leese, and Daniel Stahl. 2011. *Cluster Analysis.* 5th ed. West Sussex, UK: John Wiley & Sons.

Evteev, Andrej, Andrea L. Cardini, Irina Morozova, and Paul O'Higgins. 2014. "Extreme Climate, Rather Than Population History, Explains Mid-Facial Morphology of Northern Asians." *American Journal of Physical Anthropology* 153: 449 - 462.

Flewellen, Ayana Omilade, Justin Dunnavant, Alicia Odewale, Alexandra Jones, Tsione, Crossland, Zoë Wolde-Michael, and Maria Franklin. 2021. "The Future of

Archaeology is Antiracist: Archaeology in the Time of Black Lives Matter." *American Antiquity* (Cambridge University Press) 86 (2): 1 - 20.

Fox, John, and Milan Bouchet-Valat. 2017. "Getting Started with the R Commander." January 24. Accessed June 27, 2018. https://socialsciences.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf.

Franklin, Maria, Justin P. Dunnavant, Ayana Omilade Flewellen, and Alicia Odewale. 2020. "The Future is Now: Archaeology and the Eradication of Anti-Blackness." *International Journal of Historical Archaeology* 24: 753 - 766. Accessed December 26, 2020. https://link.springer.com/article/10.1007/s10761-020-00577-1.

Fraser-Mackenzie, Peter A.F., Itiel E. Dror, and Kasey Wertheim. 2013. "Cognitive and Contextual Influences in Determination of Latent Fingerprint Suitability for Identification Judgments." *Science and Justice* 53: 144 - 153.

Frisancho, A. Roberto. 2010. "The Study of Human Adaptation." In *Human Evolutionary Biology*, edited by Michael P. Muehlenbein, 17-28. Cambridge: Cambridge University Press.

Fuentes, Agustin. 2010. "The New Biological Anthropology: Bringing Washburn's New Physical Anthropology into 2010 and Beyond—The 2008 AAPA Luncheon Lecture." *Yearbook of Physical Anthropology* 53: 2-12.

Gayon, Jean. 2016. "De Mendel a L'epigenetique: Histoire de la Genetique." *Comptes Rendus Biologies* 339: 225-230.

Geller, Pamela L., and Christopher M. Stojanowski. 2016. "The Vanishing Black Indian: Revisiting Craniometry and Historic Collections." *American Journal of Physical Anthropology* 1-18.

Giles, Eugene, and Orville Elliot. 1962. "Race Identification from Cranial Measurements." *Journal of Forensic Sciences* 7 (2): 147-157.

Giles, Eugene, and Orville Elliot. 1963. "Sex Determination by Discriminant Function Analysis of Crania." *American Journal of Physical Anthropology* 21 (1): 53 - 68.

Giordani, Paolo, and Maria Brigida Ferraro. 2018. "Package 'fclust'." March 21. Accessed June 27, 2018. https://cran.r-project.org/web/packages/fclust/fclust.pdf.

Goldberg, David, Brianna Doby, Laura Siminoff, Malay Shah, and Raymond Lynch. 2020. "Rejecting Bias: The Case Against Race Adjustment for OPO Performance in Communities of Color." *American Journal of Transplantation* 20: 2337-2342. doi:https://doi.org/10.1111/ajt.15865.

Gravlee, Clarence C., H. Russell Bernard, and William R. Leonard. 2003. "Boas's 'Changes in Bodily Form': The Immigrant Study, Cranial Plasticity, and Boas's Physical Anthropology." *American Anthropologist* 105 (2): 326-332.

Hefner, Joseph T. 2009. "Cranial Nonmetric Variation and Estimating Ancestry." *Journal of Forensic Sciences* 54 (5): 985-995.

Hefner, Joseph T., and Stephen D. Ousley. 2014. "Statistical Classification Methods for Estimating Ancestry Using Morphoscopic Traits." *Journal of Forensic Sciences* 59 (4): 883-890.

Howells, W. W. 1973. *Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference Among Recent Human Populations.* Vol. 67, in *Papers of the Peabody Museum of Archaeology and Ethnology*, 259. Cambridge, MA: Peabody Museum.

Howells, W. W. 1996. "Howells' Craniometric Data on the Internet." *American Journal of Physical Anthropology* 101: 441-442.

Hrdlička, Aleš. 1919. *Physical Anthropology: Its Scope and Aims.* Philadelphia: The Wistar Institute of Anatomy and Biology.

Hughes, Cris E., Meredith L. Tise, Lindsay H. Trammell, and Bruce E. Anderson. 2013. "Cranial Morphological Variation Among Contemporary Mexicans: Regional Trends, Ancestral Affinities, and Genetic Comparisons." *American Journal of Physical Anthropology* 151: 506-517.

Hulse, F. S. 1981. "Habits, Habitats, and Heredity: A Brief History of Studies in Human Plasticity." *AMerican Journal of Physical Anthropology* 56: 495-501.

Jantz, R. L., and Lee Meadows Jantz. 2000. "Secular Change in Craniofacial Morphology." *American Journal of Human Biology* 12: 327 - 338.

Jantz, R. L., and S. D. Ousley. 2005. *FORDISC 3.1 Personal Computer Forensic Discriminant Functions.* University of Tennessee-Knoxville. Accessed February 17, 2020. https://fac.utk.edu/fordisc-3-1-personal-computer-forensic-discriminant-functions/.

Kerstholt, Jose, Aletta Eikelboom, Tjisse Dijkman, Reinoud Stoel, Rob Hermsen, and Bert van Leuven. 2010. "Does Suggestive Information Cause a Confirmation Bias in Bullet Comparisons." *Forensic Science International* 198: 138 - 142.

Klales, Alexandra R., and Michael W. Kenyhercz. 2015. "Morphological Assessment of Ancestry Using Cranial Macromorphoscopics." *Journal of Forensic Sciences* 60 (1): 13-20.

Knudson, Kelly J., and Christopher M. Stojanowski. 2009. *Bioarchaeology and Identity in the Americas.* Gainesville, FL: University Press of Florida.

Kubat, Miroslav. 2017. *An Introduction to Machine Learning.* 2nd ed. New York: Springer.

L'Abbé, Ericka N., Michael Kenyhercz, Kyra E. Stull, Natalie Keough, and Stephen 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* 68 (6): 1579 - 1583.

Lee, Peter M. 2012. *Bayesian Statistics: An Introduction.* 4th ed. West Sussex, UK: John Wiley and Sons.

Levy, Steven. 2018. "Graphical User Interface." *Encyclopaedia Britannica.* March 29. Accessed June 21, 2018. https://www.britannica.com/technology/graphical-user-interface.

Liebenberg, Leandi, Kyra E. Stull, Ericka N. L'Abbe, and Deona Botha. 2015. "Evaluating the Accuracy of Cranial Indices in Ancestry Estimation Among South African Groups." *Journal of Forensic Sciences* 60 (5): 1277-1282.

Littlefield, Alice, Leonard Lieberman, and Larry T. Reynolds. 1982. "Redefining Race: The Potential Demise of a Concept in Physical Anthropology." *Current Anthropology* 23 (6): 105-116.

Maddux, Scott D., Lauren N. Butaric, Todd R. Yokley, and Robert G. Franciscus. 2017. "Ecogeographic Variation across Morphofunctional Units of the Human Nose." *American Journal of Physical Anthropology* 162: 103 - 119.

Maechler, Martin, Anja Struyf, Mia Hubert, Kurt Hornik, Matthias Studer, Pierre Roudier, Juan Gonzalez, Kamil Kozlowski, Erich Schubert, and Keefe Murphy. 2019. *Package 'cluster'.* June 19. Accessed October 19, 2019. https://cran.r-project.org/web/packages/cluster/cluster.pdf.

Maimon, Oded, and Lior Rokach. 2008b. "Introduction to Soft Computing for Knowledge Discovery and Data Mining." In *Soft Computing for Knowledge Discovery and Data Mining*, edited by Oded Maimon and Lior Rokach, 1 - 13. New York: Springer.

Maimon, Oded, and Lior Rokach. 2008a. "Preface." In *Soft Computing for Knowledge Discovery and Data Mining*, edited by Oded Maimon and Lior Rokach, vii - viii. New York: Springer.

Marks, Jonathan. 2011. *The Alternative Introduction to Biological Anthropology.* New York: Oxford University Press.

Martin, Rudolph. 1928. *Lehrbuch der Anthropology.* Gustav Fischer Verlag.

Mayes, Arion T. 2010. "These Bones are Read: The Science and Politics of Ancient Native America." *American Indian Quarterly* 34 (2): 131-156.

Menéndez, Lumila, Valeria Bernal, Paula Novellino, and S. Ivan Perez. 2014. "Effect of Bite Force and Diet Composition on Craniofacial Diversification of Southern South American Human Populations." *American Journal of Physical Anthropology* 155: 114 - 127.

Microsoft Office Support. 2018. *Database basics - Access.* Accessed June 14, 2018. https://support.office.com/en-GB/article/database-basics-a849ac16-07c7-4a31-9948-3c8c94a7c204.

Molnar, Stephen. 2006. *Human Variation: Races, Types, and Ethnic Groups.* 6th ed. New York: Routledge.

Monsalve, Timisay, and Joseph T. Hefner. 2016. "Macromorphoscopic Trait Expression in a Cranial Sample from Medellin, Colombia." *Forensic Science International* 266: 574.e1-574.e8.

Montagu, M. F. Ashley. 1942. *Man's Most Dangerous Myth: The Fallacy of Race.* New York: Columbia University Press.

Moore, Megan K. 2013. "Sex Estimation and Assessment." In *Research Methods in Human Skeletal Biology*, edited by Elizabeth A. DiGangi and Megan K. Moore, 91-116. Waltham, MA: Academic Press.

Morris-Reich, Amos. 2013. "Anthropology, Standardization and Measurement: Rudolf Martin and Anthropometric Photography." *British Journal for the History of Science* 46 (3): 487-516.

Nakhaeizadeh, Sherry, Itiel E. Dror, and Ruth M. Morgan. 2014. "Cognitive Bias in Forensic Anthropology: Visual Assessment of Skeletal Remains is Susceptible to Confirmation Bias." *Science and Justice* 54: 208 - 214.

Nakhaeizadeh, Sherry, Ruth M. Morgan, Carolyn Rando, and Itiel E. Dror. 2018. "Cascading Bias of Initial Exposure to Information at the Crime Scene to the Subsequent Evaluation of Skeletal Remains." *Journal of Forensic Sciences* 63 (2): 403 - 411.

Nieves-Colón, Maria A., Andrew T. Ozga, William J. Pestle, Andrea Cucina, Vera Tiesler, Travis W. Stanton, and Anne C. Stone. 2018. "Comparison of Two Ancient DNA Extraction Protocols for Skeletal Remains from Tropical Environments." *American Journal of Physical Anthropology* 166 (4): 1 - 13.

North Carolina State University Forensic Analysis Lab. n.d. *Craniometrics Database.* Accessed April 2018, 17. https://sites.google.com/a/ncsu.edu/craniometrics-database/home.

Ousley, Stephen, Richard Jantz, and Donna Freid. 2009. "Understanding Race and Human Variation: Why Forensic Anthropologists are Good at Identifying Race." *American Journal of Physical Anthropology* 139: 68-76.

Paschetta, Carolina, Soledad de Azevedo, Lucía Castillo, Neus Martínez-Abadías, Miquel Hernández, Daniel E. Lieberman, and Rolando González-José. 2010. "The Influence of Masticatory Loading on Craniofacial Morphology: A Test Case Across Technological Transitions in the Ohio Valley." *American Journal of Physical Anthropology* 141: 297-314.

Perez, S. Ivan, Verónica Lema, José Alexandre Felizola Diniz-Filho, Valeria Bernal, Paula N. Gonzalez, Diego Gobbo, and Héctor M. Pucciarelli. 2011. "The Role of Diet and Temperature in Shaping Cranial Diversification of South American Human Populations: An Approach Based on Spatial Regression and Divergence Rate Tests." *Journal of Biogeography* 38: 148 - 163.

Rakita, Gordon F. M., Jane E. Buikstra, Lane A. Beck, and Sloan R. Williams, . 2005. *Interacting with the Dead: Perspectives on Mortuary Archaeology for the New Millennium.* Gainesville, FL: University Press of Florida.

Relethford, John H., and Francis C. Lees. 1982. "The Use of Quantitative Traits in the Study of Human Population Structure." *Yearbook of Physical Anthropology* 25: 113 - 132.

Roseman, Charles C. 2016. "Random Genetic Drift, Natural Selection, and Noise in Human Cranial Evolution." *American Journal of Physical Anthropology* 160: 582 - 592.

Ross, Ann H., Douglas H. Ubelaker, and Erin H. Kimmerle. 2011. "Implications of Dimorphism, Populationa Variation, and Secular Change in Estimating Population Affinity in the Iberian Peninsula." *Forensic Science International* 206: 214.e1 - 214.e5.

Sarstedt, Marko, and Erik Mooi. 2014. *A Concise Guide to Market Research.* 2nd ed. New York: Springer.

Schraiber, Joshua G., and Joshua M. Akey. 2015. "Methods and Models for Unravelling Human Evolutionary History." *Nature Reviews Genetics* 16: 727 - 740.

Selcer, Perrin. 2012. "Beyond the Cephalic Index: Negotiating Politics to Produce UNESCO's Scientific Statements on Race." *Current Anthropology* 53 (S5): S173-S184.

Skotte, Line, Thorfinn Sand Korneliussen, and Anders Albrechtsen. 2013. "Estimating Individual Admixture Proportions from Next Generation Sequencing Data." *Genetics* 195: 693 - 702.

Smith, Matthew S. 2021. "The Best Microsoft Office Alternatives of 2021." *Business Insider.* August 26. Accessed February 2, 2022. https://www.businessinsider.com/guides/tech/best-microsoft-office-alternatives.

Spradley, M. Katherine. 2014. "Toward Estimating Geographic Origin of Migrant Remains Along the United States-Mexico Border." *Annals of Anthropological Practice* 38 (1): 101 - 110.

Statsoft, Inc. 2013. *Electronic Statistics Textbook.* Accessed January 4, 2018. http://www.statsoft.com/textbook/.

Stojanowski, Christopher M., and Michael A. Schillaci. 2006. "Phenotypic Approaches for Understanding Patterns of Intracemetery Biological Variation." *Yearbook of Physical Anthropology* 49: 49 - 88.

Strauss, André, and Mark Hubbe. 2010. "Craniometric Similarities Within and Between Human Populations in Comparison with Netural Genetic Data." *Human Biology* 82: 315 - 330.

Ta'ala, Sabrina C. 2015. "A Brief History of the Race Concept in Physical Anthropology." In *Biological Affinity in Forensic Identification of Human*

Skeletal Remains: Beyond Black and White, edited by Gregory E. Berg and Sabrina C. Ta'ala, 1-15. Boca Raton, FL: CRC Press.

Tallman, Sean D., and Allysha P. Winburn. 2015. "Forensic Applicability of Femur Subtrochanteric Shape to Ancestry Assessment in Thai and White American Males." *Journal of Forensic Sciences* 60 (5): 1283 - 1289.

The R Foundation. n.d. "R: What is R?" *The R Project for Statistical Computing.* Accessed May 1, 2018. https://www.r-project.org/about.html.

Thomas, Richard M., Connie L. Parks, and Adam H. Richard. 2017. "Accuracy Rates of Ancestry Estimation by Forensic Anthropologists Using Identified Forensic Cases." *Journal of Forensic Sciences* 1-4.

Trentin, Edmondo, Luca Lusnig, and Fabio Cavalli. 2018. "Parzen Neural Networks: Fundamentals, Properties, and an Application to Forensic Anthropology." *Neural Networks* 97: 137 - 151.

Trochim, William M. 2006. *The Research Methods Knowledge Base.* Cengage Publishing. October 20. Accessed August 29, 2017. https://www.socialresearchmethods.net/kb/.

Ubelaker, Douglas H., Ann H. Ross, and Sally M. Graver. 2002. "Application of Forensic Discriminant Functions to a Spanish Cranial Sample." *Forensic Science Communciations* 4 (3). Accessed February 17, 2020. https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2002/ubelaker1.htm.

United States Census Bureau. 2016. *U. S. Census Bureau QuickFacts: Tennessee.* Accessed June 14, 2018. https://www.census.gov/quickfacts/fact/table/TN/PST045217.

University of Tennessee-Knoxville. n.d. *Forensic Anthropology Data Bank | Forensic Anthropology Center.* Accessed August 30, 2017. https://fac.utk.edu/background/.

Urbanová, Petra, Ann H. Ross, Mikolás Jurda, and Maria-Ines Nogueira. 2014. "Testing the Reliability of Software Tools in Sex and Ancestry Estimation in a Multi-Ancestral Brazilian Sample." *Legal Medicine* 16: 264 - 273.

von Cramon-Taubadel, Noreen. 2011. "The Relative Efficacy of Functional and Developmental Cranial Modules for Reconstructing Global Human Population History." *American Journal of Physical Anthropology* 146: 83 - 93.

von Cramon-Taubadel, Noreen, and Stephen J. Lycett. 2008. "Brief Communication: Human Cranial Variation Fits Iterative Founder Effect Model with African Origin." *American Journal of Physical Anthropology* 136: 108-113.

Wagner, K. 1932. "The Variability of Hybrid Populations." *American Journal of Physical Anthropology* 16 (3): 283-307.

Washburn, Sherwood Larned. 1951. "Section of Anthropology: the New Physical Anthropology." *Transactions of the New York Academy of Sciences* 13 (7) Series II: 298 - 304.

Weisensee, Katherine E., and Richard L. Jantz. 2016. "An Examination of the Differential Effects of the Modern Epidemiological Transition on Cranial Morphology in the United States and Portugal." *Human Biology* 88 (1): 30 - 37.

White, Tim D., Michael T. Black, and Pieter A. Folkens. 2012. *Human Osteology.* 3rd ed. Burlington, MA: Academic Press.

Wijsman, Ellen M., and Walter A. Neves. 1986. "The Use of Nonmetric Variation in Estimating Human Population Admixture: A Test Case with Brazilian Blacks, Whites, and Mulattos." *American Journal of Physical Anthropology* 395-405.

Zhang, G. Peter. 2008. "Neural Networks for Data Mining." In *Soft Computing for Knowledge Discovery and Data Mining*, edited by Oded Maimon and Lior Rokach, 17 - 44. New York: Springer.