ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb569

Popularizing Multi-Dimensional Data Bucketing for the Social Sciences Through Inductive Research: Another Endeavour for Improved Social Sciences Research

Sujay Rao Mandavilli

Publication Date: 2026/02/14

Abstract: The objective of this paper is to emphatically proclaim that knowledge is not infallible, and that all paradigms and models in science are subject to constant revision and modification, subject of course to rock solid data and reliable evidence. The main vehicle upon which this paper rides is of course the inductive approach to research which me must support and fight for tooth and nail, subject of course, to cost and time concerns and considerations. We begin this paper by furnishing quotes of eminent scientists and thinkers in support of our stance, and then review the concept of data analysis, along with the different types of data. The core concepts and postulates of our paper are then presented along with what we call the "primary axis", "secondary axis", and pattern identification. Multi-dimensional data bucketing is almost always required, and these need to be vetted and validated against real-world data. The latter needs to be a continuous process, and theories and frameworks need to be revalidated constantly and continuously. This paradigm evolved in the context of geographical analysis such as the claimed out of Africa dispersal of humans, and the origin of language, but this can be used for temporal analysis and other forms of analysis as well. Therefore, some other examples are also provided. We do therefore, hope, expect and anticipate that this paper will become an important one in the twenty-first century philosophy of science.

How to Cite: Sujay Rao Mandavilli (2026) Popularizing Multi-Dimensional Data Bucketing for the Social Sciences Through Inductive Research: Another Endeavour for Improved Social Sciences Research. *International Journal of Innovative Science and Research Technology*, 11(2), 607-613. https://doi.org/10.38124/ijisrt/26feb569

I. INTRODUCTION

The objective of this paper is to emphatically proclaim that knowledge is not infallible, and that all paradigms and models in science are subject to constant revision and modification, subject of course to rock solid data and reliable evidence. The main vehicle upon which this paper rides is of course the inductive approach to research which me must support and fight for tooth and nail, subject of course, to cost and time concerns and considerations. We begin this paper by furnishing quotes of eminent scientists and thinkers in support of our stance, and then review the concept of data analysis, along with the different types of data. The core concepts and postulates of our paper are then presented along with what we call the "primary axis", "secondary axis", and pattern identification. Multi-dimensional data bucketing is almost always required, and these need to be vetted and validated against real-world data. The latter needs to be a continuous process, and theories and frameworks need to be revalidated constantly and continuously. This paradigm evolved in the context of geographical analysis such as the claimed out of Africa dispersal of humans, and the origin of language, but this can be used for temporal analysis and other forms of analysis as well. Therefore, some other examples are also provided. We do therefore, hope, expect and anticipate that this paper will become an important one in the twenty-first century philosophy of science. [1]

John von Neumann needs no introduction to most people familiar with science. John von Neumann was an eminent Hungarian-American mathematician and physicist. He was known to have integrated pure and applied sciences, having made praiseworthy and noteworthy contributions to many fields, including mathematics, physics, and statistics. He impacted the field of economics too with his pioneering

[1] Trochim, W.M.K, (2006). Research Methods Knowledge Base.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb569

contributions to game theory; John von Neumann viewed science as being based on rigour and discipline rather than a search for ultimate explanations which could sometimes prove to be elusive. Therefore, and in one short sentence, conclusions can never be final, and may be subject to revision. That is why Carl Sagan noted and observed, science is "a way of thinking much more than it is a body of knowledge". Generalizations might quite often turn out to be false as noted by both Mark Twain and Alexandre Dumas. There is yet another interesting quote and saying, "In seeking absolute truth we aim at the unattainable and must be content only with broken portions". This saying is attributed to William Osler, the famous Canadian physician, who advocated for intellectual humility over absolute certainty. Critical thinking is also extremely important, and its importance was realized by Albert Einstein and others. As Einstein indeed stated, commenting on Thomas Alva Edison's opinion that a college education is useless, "The value of an education is not the learning of many facts", he said, "but the training of the mind to think about the many things that cannot be learnt from textbooks". However, scientific method and the philosophy of science have both evolved since then. Knowledge must be based on facts as far as possible, and while hypothesis and theories must be constructed, they must always be subject to revalidation and reassessment.

As Corey Christen said in "The price of illiteracy", "When a society loses the ability to differentiate knowledge from opinion, it loses everything, not just truth, but the capacity to act as well." People must be taught how to think, not what to think, and this idea was first championed by anthropologist Margaret Mead and others. As the noted management guru once observed, knowledge has to be improved, challenged and increased constantly, otherwise, it vanishes. Again, we need no hasty generalization, we need rock solid data driven models. Hasty generalization in plain and simple English means jumping to conclusions. A hasty generalization fallacy is form of a logical fallacy where a broad conclusion is drawn from a tiny or unrepresentative sample, with ample scope for personal bias and judgmental conclusions. Scientific models must also be constructed wherever possible; Scientific models are simplified yet accurate representations of complex physical objects or conceptual systems such as models of the solar system, the milky way galaxy, or models of the atom such as Niels Bohr's representation. We also then have the plate tectonics model, the tree of life, the food web model, and the water cycle model. All these must not necessarily overly abstract but must be driven by real-world data as far as possible. Anticipated and expected future data must also be seamlessly fitted into existing models, but admissions of error made wherever necessary. Models can then be revised by the same researcher, or by other researchers. While having two or more schools of thought can be beneficial, there must not be too much of rampant careerism. [2] [3] [4] [5] [6] [7]

- *What is Data Analysis?*

Data analysis is extremely common in today's world; it is so common that it is inspiring and drawing legions of data scientists into the field- data science has even become an extremely common and popular profession among today's youth. A data scientist is a typically twenty-first century multidisciplinary and trans disciplinary professional who combines expertise drawn or culled from multiple fields judiciously and effectively to offer new insights into data. Expertise could include expertise in multiple fields and professions such as computer science, statistics, and scientific analysis techniques to analyze, construct, visualize, and interpret highly complex and insightful large-scale data models. Often, data inspection, cleansing, transformation, and modeling may also be required. Such data can be used by organizations, businesses, governments, scientists, computer departments, and individuals alike, to aid in their decision making processes. Other secondary techniques such as data mining can also be used; the latter is extremely important from our perspectives and can be used to discover hidden or underlying patterns or correlations. Data analysis is sometimes divided into descriptive statistics, exploratory data analysis which is used to draw insights and correlations, and confirmatory data analysis which used to validate or contradict preliminary hypothesis. Predictive analytics is another important technique that focuses on the application of statistical models for predictive forecasting or behavior modeling. Data modeling is used to create and model schemas on how data in constructed, stored and analyzed. Data bucketing and data categorization or data classification are used to categorize or classify date into buckets or classification. We also have univariate, bivariate and multivariate analysis, and such a classification would indeed be dependent on the number of variables involved; this concept was carried forward into one way and two way ANOVA analysis, where the term factor is used, and these were

[2] Adami, Christoph (2024). *The Evolution of Biological Information: How Evolution Creates Complexity, from Viruses to Brains*. Princeton: Princeton University Press. pp. 189–190.
[3] Wehrung, Friedrich (2006). "Von Neumann coordinatization is not first-order". *Journal of Mathematical Logic*. **6** (1): 1–24
[4] v. Neumann, J. (1929). "Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen". *Mathematische Zeitschrift* (in German). **30** (1)
[5] von Neumann, John; Bochner, Salomon (1935). "Almost Periodic Functions in Groups, II". *Transactions of the American Mathematical Society*. **37** (1): 21–50
[6] Callaham, Michael; Wears, Robert; Weber, Ellen L. (2002). "Journal Prestige, Publication Bias, and Other Characteristics Associated With Citation of Published Studies in Peer-Reviewed Journals". *JAMA*. **287** (21): 2847–50
[7] Eisner, E. W. (1981). "On the Differences between Scientific and Artistic Approaches to Qualitative Research". *Educational Researcher*. **10** (4): 5–9

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb569

techniques that were developed by RA Fischer and George W. Snedecor in agricultural contexts. [8] [9] [10]

➢ *What is a Dimension?*

A dimension may be defined as the measurement of the size or the extent in any one particular direction (most common examples being length, width, height, or time, and in simple terms), represents the number of coordinates that are required to locate a point in space. For example, one dimension is length, two dimensions in a plane with both length and width, while a three dimensional object may be either a cube or a cuboid, for example. These concepts are often widely used in science, engineering and physics. The term "dimension" may have other meanings as well. For example, it may refer to the scope or the magnitude of some entity, a common example being the "many dimensions" of a problem. Another connotation or implication of the term dimension is the process or act of measuring a single or multiple aspect of a problem, trait, or any other attribute. For example, a unidimensional attitude scale, personality scale, behavioural scale, attribute scale, or other scale would contain items related only to the respective concept of interest. Therefore, while unidimensional models always measure, analyze, represent or depict data using a single dimension, factor, dimension, or a variable, (thereby providing a simple, and a linear overview in terms of say weight or temperature, multiple personality attributes, or proficiency or competence in more than one subject. multidimensional approaches and multidimensional research models are far more complex, and may attempt to integrate multiple, distinct, though interrelated dimensions, thereby furnishing a much more comprehensive, and a multilayered understanding of issues. Multidimensional models therefore not only provide much higher levels of accuracy but also richer an deeper insights into complex behaviors or systems. This is what we recommend in this paper for various fields of the social sciences, and our entire paper is built on the aforesaid principle. [11] [12] [13]

➢ *Inductive Approach*

Inductive research is a widely used approach and technique in various branches and fields of the social sciences and humanities. Instead of basing research with an already existing theory, inductive approaches to research are based on a large number of observations of real world data (systematically gathered from large, though relevant and interrelated scenarios) and uses such data to systematically and meticulously build and construct more complex concepts, patterns, and eventually testable hypotheses and theories. The process of testing theories and hypotheses itself may become much simpler, given the fact that data is always used, and forms the basis of their construction. However, theories and hypotheses may be modified in due course as and when more data presents itself, or becomes available; inductive approaches can also be used in conjunction with other approaches, and a research study may typically combine several approaches and techniques for greater efficacy and flexibility. On the other hand, deductive approaches to research represent a top down approach to research, and theories and hypothesis are often formulated first, and then tested as more and more data becomes available. Deductive approaches are associated with top down reasoning, and the hypothetico deductive approach and the deductive nomological approach and technique are often used. While inductive research processes are often laborious and time-consuming, we recommend them as far as practically possible. As a matter of fact, our entire method- the approach espoused in this paper – in naturally built upon inductive research models and techniques. [14] [15]

We will also now briefly review the different types of data sources that are used in social sciences research, and research in the humanities. There represent mostly qualitative data, though quantification techniques can also be used. This is because our work deals primarily with the social sciences and the humanities, though these approaches and techniques can be used in other fields of research and application as well. We may make use of ethnographic data as well – these are often collected over protracted periods of time, say 15 to 18 months in a native setting, through what is most commonly known as the participant observation method. This method came of age in the 1920's, through the efforts of Franz Boas and Bronislaw Malinowski who were famous ethnographers. Data from other social science research techniques may also be used; for example, we may have interviews of various types, structured, semi-structured or unstructured, questionnaires of various types such as open ended or closed ended, surveys, focused group

[8] Tabachnick, B.G.; Fidell, L.S. (2007). *Using Multivariate Statistics, 5th Edition.* Boston: Pearson Education, Inc. / Allyn and Bacon

[9] Lewis-Beck, Michael S. (1995). *Data Analysis: an Introduction*, Sage Publications Inc

[10] Juran, Joseph M.; Godfrey, A. Blanton (1999). *Juran's Quality Handbook, 5th Edition.* New York

[11] Cohen, N.; Arieli, T. (2011). "Field research in conflict environments: Methodological challenges and snowball sampling". *Journal of Peace Research*. **48** (4): 423–436

[12] Creswell, John W. (2008). *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson

[13] Kara, Helen (2012). *Research and Evaluation for Busy Practitioners: A Time-Saving Guide*. Bristol: The Policy Press

[14] Reverby, Susan M. (1 April 2012). "Zachary M. Schrag. Ethical Imperialism: Institutional Review Boards and the Social Sciences, 1965–2009. Baltimore: Johns Hopkins University Press. 2010. Pp. xii, 245. $45.00". *The American Historical Review*. **117** (2): 484–485

[15] Rocco, T.S., Hatcher, T., & Creswell, J.W. (2011). The handbook of scholarly writing and publishing. San Francisco, CA: John Wiley & Sons. 2011

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb569

discussions, etc. Quantification techniques may also be used such as Thurstone scale and Likert scale. Quantification techniques are also an important part and parcel of this paper, and these must be widely used and applied in various fields of the social sciences. [16] [17]

➢ *Steps and the Process Involved*

The following are the various steps and processes involved in relation to the aforesaid technique. The first step would involve identifying attributes. An attribute in this connection and context would mean a quality or a feature which is an innate characteristic or an inherent part of something usually or typically a phenomenon or a research study. The next step would involve identifying parameters. But what is a parameters? A parameter is a numerical or any other measurable factor that primarily defines a system or defines the conditions of its operation. We must also identify threshold values. What is a threshold value? A threshold value is the magnitude or intensity that must be exceeded for a specific action, reaction, result, phenomenon, occurrence, or condition to manifest itself or be demonstrated. We must also then identify ranges of data. The range represents the variety of values in data sets, or much more simply, the difference between maximum values and minimum values in a data set which, in the case of quantitative date, be derived by means of a formula. [18] [19] [20]

➢ *Types of Data*

For every field of inquiry in the sciences we need data and evidence. This applies not just for the social sciences, but also for all other fields of the sciences. For example, if we were to probe and investigate Sir Roger Penrose's theory that multiple universes existed before the present one, we need data and evidence. If we are to probe and investigate the multiverse theory developed by Hugh Everett in 1957, we need data and evidence. Let us now attempt to differentiate between data and evidence. Data generally refers to raw, unprocessed, or unanalyzed facts and or observations, either quantitative or qualitative. Evidence on the other hand, refer to data that has been scrutinized, interpreted, analyzed, or advanced in support or refutation of a theory, hypothesis, or set of observations. Evidence is therefore, data in action. Let us now attempt to analyze the different types of data below:

- Quantitative data: Quantitative data in sum, refers to information that can be measured, counted, and expressed numerically, and such data is widely used in statistical analysis. For example, heights of students, weight of students, etc.
- Qualitative data: Qualitative data connotes non-numerical, descriptive information that is used to understand concepts, emotions, and behaviors. Qualitative data is most commonly obtained through interviews, focus group discussions, questionnaires or subject-centric survey responses. Qualitative data provides context-rich, in-depth and subjective insights into social and cultural phenomena. These data may or may not be later quantified depending on the context or requirement.
- Statistical data: Statistical data refers to the collection of numerical or statistically analyzable facts gathered through experiments, or observations (sometimes social science research techniques) to analyze, describe, and interpret real-world observations or phenomena. Statistical data is extensively used in fields as diverse as economics, product engineering and analysis, and demographics—it is pivotal and central for identifying trends, making data-driven predictions, and arriving at evidence-based decisions.
- Non-statistical data: Non-statistical data refers to information that does not depend on numerical, statistical or probability-based analysis to draw conclusions, and instead relies on largely subjective data sets for analysis. Examples could include boolean data, categorical answers, or non-quantifiable feedback.
- Ordinal data: Ordinal data refers to categorical, statistical data where data is ranked systematically in a order. Example, ranks of student heights or test scores. Complex downstream mathematical or statistical analysis may not be possible with this type of data.
- Cardinal data: Cardinal data refers to ordinary cardinal numbers that are widely used for counting and measuring quantity in daily life. Decimal numbers are also often used in such data. Therefore, magnitude or size in measured, not position or order.
- Discrete data: "What exactly is discrete data?", we may ask. Data that can only take certain values within a given range of values is referred to as discrete data or discrete values. This data can most commonly be counted, and takes on the form of whole numbers or integers, most often though not always.
- Continuous data: Continuous data refers to quantitative information that can take any value within a given range,

[16] Thurstone, L. L. (1927b) The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400

[17] Krus, D.J., & Kennedy, P.H. (1977) Normal scaling of dominance matrices: The domain-referenced model. *Educational and Psychological Measurement*, 37, 189-193

[18] Anderson, D.R.; Sweeney, D.J.; Williams, T.A. (1994) *Introduction to Statistics: Concepts and Applications*, pp. 5–9. West Group

[19] Higham, Nicholas J. (1998). "Aids and Resources for Writing and Research". *Handbook of Writing for the Mathematical Sciences*. Society for Industrial and Applied Mathematics. p. 214

[20] *Sheynin, Oscar (2010). "Statistics, History of". In Lovric, Miodrag (ed.). International Encyclopedia of Statistical Science. Springer. pp. 1493–1504*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb569

including infinitesmally small values within a given rage, characterized by infinite divisibility. This data type often allows for a great deal of precision or accuracy to be obtained.

- Interval data: Interval data refers to a type of quantitative, numerical measurement in which variables are ordered and separated by equal, consistent distances. Data concentration can be measured for different intervals conveniently, or absence of data concentration.
- Grouped data: Grouped data refers to a method of organizing raw data also known as ungrouped, unbinned or unsegregated data into convenient or easily manageable categories, classes, or intervals in such a way that large datasets can be made much more manageable and easy to analyze. Data grouping is often found in frequency distribution analysis such as normal distribution, histograms, and frequency polygons. This grouping is further used to derive mean, median and mode including geometric mean, harmonic mean, quartiles, deciles, quintiles, percentiles, etc.
- Ungrouped data: Ungrouped data, which is also often referred to as raw data in common, colloquial parlance or everyday usage, refers to data collected in its original and unsorted or non-aggregated (i.e. ungrouped) form without being organized into categories, classes or frequency distributions.
- Categorical data: Categorical data represents classes or categories of information that are neatly or conveniently subdivided into groups or labels rather than numerical measurements alone. This is akin to data bucketing, and this forms the foundational and fundamental premise or foundation of this paper. The term "data bucketing" is already widely used in statistics, and this refers to grouping and assortment of data into some convenient basis.
- Multi-dimensional data: We may also then have multi-dimensional data or data that is sorted or aggregated based on multiple dimensions or criteria. This is of course the foundational premise of this paper. There may be independent categories or dimensions, and sub categories or sub dimensions. We had also defined the concept of primary axis or the primary basis of measurement in this paper.

➢ *Quantification Techniques*

The term "quantification" which first originated in the field of positivism, is a widely used term in social sciences research and it refers to the process of measuring, counting, converting (also objectivizing and standardizing) abstract and seemingly non-quantifiable social phenomena into operable and statistically usable numerical data that may be further used to identify patterns, test hypotheses, and analyze relationships using statistics. Key aspects of quantification in social research include measurement of data, binning of data, discarding of unwanted data, statistical analysis and the like. Original, non-numerical data may be derived from disciplines and subdisciplines as far apart and widely removed such as phenomenology, structuralism, and emic and etic studies. [21]

➢ *Data Patterns*

"What are patterns?", we may ask. "What are data patterns?", we may ask. A pattern is nothing but a predictable, repeated occurrence with observable and analyzable attributes. These are easily observable in elements like shapes, colors, textures, surfaces, or in a more mathematical and statistical sense, numbers or more complex statistical data. In the realm of social sciences, patterns may be observable in behavior as well. Patterns may also be used to group and bucket raw data into easily manageable and analyzable classes, which can help understand or derive rules, or even make predictions in many cases. Patterns in data analysis help identify and measure recurring themes, trends, or relationships within datasets, including, from our perspective, social and cultural data.

➢ *What then is Our Approach?*

Our approach therefore, implies sorting data on the basis of the primary axis, and then sub-sorting it based on some other dimension. We have already explained the idea of a primary axis before, and simply put it refers to an categorization of data that is rigid, and does not change during the course of the analysis or research study. Examples of primary axis include time in case of temporal data (given the fact that time frames most often do not change), space coordinates in cases of spatial and geographical data, and social structures as in the case of hierarchical analysis, Therefore, the following are the primary uses of our approach and technique.

➢ *Use in Spatial and Geographical Data*

Spatial data, which is also sometimes referred to as geospatial data, refers to information that identifies the geographic location, attributes and characteristics of natural or manmade features or objects on the earth's surface. Such data is increasingly being stored and managed in GIS systems or "Geographic information systems" for further downstream analysis, and decision-making. Geographical data on the other hand refers to data that is organized or sorted based on some geographical attribute, for example, county, state, or region.

➢ *Use in Temporal Data*

Temporal data on the other hand, refers to information that is sorted and indexed by time, also seeking to capture and isolate how data sets, (events, observed phenomena and occurrences included) or measurements change over periods of time that are factored into an analysis. often, scales such as years, months and days are used, and start and end dates representing the period of the study are used too.

[21] Adam, John A. *Mathematics in Nature: Modeling Patterns in the Natural World*. Princeton, 2006

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb569

➢ *Use in Other Forms of Data*

This kind of an analysis, lends itself quite brilliantly to other forms of data such as hierarchical data, though this research technique indeed originated in geographical contexts – study of the purported or claimed Out of Africa analysis, and various now obsolete and defunct theories dealing with the origin of language. We may recall here, and at this juncture, that we had proposed the "epochal polygenesis" approach instead. For example, we may study how various attributes, characteristics, or behavioural patterns change with income levels, or class hierarchy – therefore, income levels and the class hierarchy represent the primary axis here. Of course, the primary axis doesn't change throughout the course of the study, and income levels and class hierarchy do not change during the entire course of the study. Observed versus expected values must be constantly compared with each other, and models must be continuously and constantly redeveloped and recast. Therefore, expected values may be determined mathematically or non-mathematically, and this must be done on a constant basis. A comparison of expected and actual values may also be performed judgmentally, and not necessarily mathematically or statistically, as in the case of a Chi-square test.

The entire objective of this exercise is to drive home the fact that knowledge is most often provisional and not final. Many researchers conveniently ignore this fact as observed in case of the Out of Africa theory, or archaic hypotheses, on the origin of languages. Comparison with actual values must be performed constantly, or if not constantly, as frequently as possible. Expected values for time series and non time series data must be obtained and patterns identified including data based on age, gender, income level, nationality, religion, etc. We must also avoid walking into the hasty generalization trap, and must make assessments and reassessments constantly. We may also make predictions through data extrapolations, and may resort to altering the scales or the granularity of the scales. Let us now present some interesting examples below. In many cases of study, both a primary and a secondary axis of study may be used, rarely a tertiary axis, and many other dimensions made subservient to them. For example, we may want to study types of observations made against an urban rural continuum (quantified based on the population size of the habitat or settlement), and further based on countries or regions. The latter would constitute a secondary axis. Of course, everything would be dependant based on the nature of study involved. [22]

➢ *Out of Africa Theory Versus Multiregional Hypothesis*

The "Out of Africa" theory which is also sometimes referred to as the recent African origin of anatomically modern humans proposes that Homo Sapiens evolved solely in Africa-different parts of Africa proposed by different scholars and dispersed to other parts of the world, though the reasons for such dispersal are not clearly explained (for us they would be somewhat, if not highly unreasonable). As a matter of fact, dispersal routes in many cases are assumed based on existing fossil evidence in different parts of the world, which is somewhat fallacious, if not highly erroneous. The date of dispersal is also notoriously uncertain, and varies from scholar to scholar. The upper date limit is generally 200,000 years and the lower date limit is generally 60,000 years. Key variants of the Out of Africa theory therefore include the recent African origin hypothesis, assimilation model or partial replacement model (where it is postulated that Homo Sapiens interbred with other archaic hominins; the origin of the other hominins is however nether clearly nor ambiguously explained. Some models call for multiple dispersals in waves; while others claim that it was homo erectus that emerged from Africa much, much earlier, say two million years ago – therefore, the issue is clearly unsettled due to the absence of sufficient data and evidence. Out of Africa theories differ considerably from non origin theories and multiregional models which propose that Homo sapiens evolved simultaneously and in parallel in different continents or regions of the world, from other archaic populations with differing levels of gene flow across regions. [23] [24]

However, it is our firm and equivocal belief, (it has been for several decades now), that the dust has not settled, and meaningful conclusions cannot be drawn due to paucity and dearth of data. Combinations of different theories are also possible, and this would not constitute a mix up in the traditional sense of the term; The characteristics and attributes of human fossils such as cranial capacity and the size of femur bones must be studied from diverse contexts and situations, and in such as case, the primary axis would be geographical location, and the secondary axis would be time; everything else would be subservient to these, including the size of the bones. However, the latter must be must critically in data analysis. This is the way to go, and data-driven models must be refined constantly or continuously, or left as they are. There is also a wide range of anatomical diversity within Africa, and this may work against the out of Africa theory, rather than in favour of it. Which variant migrated outwards? Again, we have no clue here. Primates also comprise both platyrrhine and catarrhine, or new world and old world monkeys respectively. There is too little data to construct robust evolutionary models, and most fossils are found by accident not conscious digging. However, model, model, model, we must and improve them as and when new data appears.

[22] Campbell, Ian (2007-08-30). "Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations". *Statistics in Medicine*. **26** (19): 3661–3675

[23] Robinson D, Ash PM (2010). *The Emergence of Humans: An Exploration of the Evolutionary Timeline*. New York

[24] Shea, John J. (2003). "Neandertals, competition, and the origin of modern human behavior in the Levant". *Evolutionary Anthropology: Issues, News, and Reviews*. **12** (4): 173–187

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb569

For example, we have Sahelanthropus tchadensis from the Torros-Menalla area of Chad in Africa, The brain size of these fossils are a puny 320-380 CC; these are six to seven million years old, and further details of the fossils must be obtained such as postulated age and gender. We then have Orrorin tugenensis from Kenya which are between 6.2 and five million years old. Their cranial capacity is unknown, and in 1992, the Japanese anthropologist Gen Suwa discovered what are today known as Ardipithecus ramidus fossils in Ethiopia. Giday Woldegabriel and others also made several interesting and noteworthy finds. The list goes on and on. Fossils have been found as far apart as in China, Siberia, the middle east, and Indonesia. We also need interdisciplinary models, and competent interdisciplinary synthesizers. We have a long way to go to make science reliable and robust. The journey and the voyage have just begun. There could be other via media proposals too; for example, evolution may have begun earlier in Africa, and fossils may have been better preserved in Africa. There could be many different interpretations. Theories on origin and spread of languages must also be developed accordingly and in unison, and the characteristics and attributes of language observed carefully; Isoglosses must be mapped as well. Theories on the origin of language must also ride piggy back on human dispersal models as far as possible, and dependant on the time frame.

We are currently too obsessed with rampant theorization and hypothesis building; there is too much careerism, and too much of a publish or perish culture; cross-cultural collaboration is sorely lacking; interdisciplinary research is relatively low; barely and scarcely enough modeling is down; to reduce vested interests that impede the healthy growth of science; we need multiple centres and pillars or power in science; non just US centric science; however, ideologies must be booted out, and so must other egregious approaches; we may use this technique to perform other forms and other kinds and types of analysis; for example, we could measure increases in wealth versus increase in obesity. We could use this technique to make preliminary observations, and then further additional observations. We may map increases in literacy versus years – various attributes can be used here, and cause and effect analysis be performed as well in some cases, along with a root cause analysis. likewise, changes in customer tastes and preferences can be analyzed, changes in market conditions can be analyzed, changes in economic conditions can be analyzed, changes in obesity and morbidity rates can be analyzed, and of course many, many more. Always, preliminary observations can be made along with additional observations, and several rounds of iterations are always required in research as more and more data is obtained, not assertions of fait accompli. The rather could be extremely dangerous; As Nei deGrasse Tyson once put it, “Science works on the frontier between knowledge and ignorance. Scientists must not be afraid to admit that they do not know. There is no shame in that. The only shame it to pretend that they have all the answers.”

II. CONCLUSION

The objective of this paper was to emphatically proclaim that knowledge is not infallible, and that all paradigms and models in science are subject to constant revision and modification, subject of course to rock solid data and reliable evidence. The main vehicle upon which this paper rides is of course the inductive approach to research which me must support and fight for tooth and nail, subject of course, to cost and time concerns and considerations. We begin this paper by furnishing quotes of eminent scientists and thinkers in support of our stance, and then review the concept of data analysis, along with the different types of data. The core concepts and postulates of our paper are then presented along with what we call the “primary axis”, “secondary axis”, and pattern identification. Multi-dimensional data bucketing is almost always required, and these need to be vetted and validated against real-world data. The latter needs to be a continuous process, and theories and frameworks need to be revalidated constantly and continuously. This paradigm evolved in the context of geographical analysis such as the claimed out of Africa dispersal of humans, and the origin of language, but this can be used for temporal analysis and other forms of analysis as well. Therefore, some other examples are also provided. We do therefore, hope, expect and anticipate that this paper will become an important one in the twenty-first century philosophy of science.