Intersect, Vol 10, No 1 (2016)

Ambiguous Complexity: Visualizations of Big Data in Microbiome Science

Courtney Nguyen
University of California Los Angeles

Abstract

The purpose of this research is to identify the various contexts in which visualizations create social inequalities in the realm of big data science. In analyzing the visuals from two microbiome case studies, it becomes clear that the interplay of the visualizations conflates with the complexities of the datasets from which they were derived. Continuing visualization studies through interdisciplinary lenses will further understanding of the inequalities in science and society.

Introduction

Gathering large datasets often introduces more questions than answers due to the incredible influx of multifunctional raw data. “Far better an approximate answer to the right question, which is often vague,” stated statistician John Tukey, “than an exact answer to the wrong question, which can always be made precise.” Is today’s science characterized by inaccurate simplicity or ambiguous complexity? Because of the dense, vague nature of big data, creating graphic representations of this data proves equally difficult.

Today, visualizations of big data create both understanding as well as confusion among viewers. Not all interpretations are created equal and “not all numbers are neutral” (Boyd & Crawford, 2012). Translating big data collections into visualizations, interactive graphics, and images demands caution and a critical eye. Considerations of the audience and how this data is treated and portrayed may reveal deeper understandings about the surrounding context of big data science—its social, political, historical contexts—than the surface-level scientific interpretation can provide.

In this paper, I will examine the types of visual representations of big data in two case studies related to the microbiome and analyze the modes in which these visuals present the data. The microbiome, literally a microscopic community from *micro-* and *-biome*, comprises a vast catalog of microbes and their genes (Ursell *et al.*, 2012). Understanding the ambiguity as well as complexity of microbiome science through the tool of visualizations is essential to unveiling how big data can create inequalities. There are many themes of this paper, including, but not limited to, flaws in big data experimentation, complexities of racial profiling, quantitative representation of qualitative factors, relatability in science, and aggressive advancement of technology.

Microbiome science, in particular, embodies the ambitions and dilemmas of mass sequencing and analysis in the age of big data science. Interchangeably known as metabonomics and metabolomics, microbiome science measures a diverse multitude of biological molecules and is regarded as the “most advanced downstream product” used to holistically study molecules (Horgan & Kenny, 2011). Deconstructing how to understand visualizations of microbiome science can provide a good model into addressing unique challenges of big data science overall (Jansson & Baker, 2016).

Analysis begins with visualization studies. Understanding visualizations, first and foremost, allows for the framework of this paper to form. How, why, and when visualizations are created help gather insights about how the audience must come to view them. I intend to study two microbiome cases through a social science lens in order to demonstrate how science is not a standalone complex, but rather a process that receives influences from its own cultural and historical time frame. The first case study analyzes a live conference presentation on the microbiome. The second case study revolves around a controversial project that takes place in an urban environment. Both studies not only reveal similarities and differences in the interpretations of their visualizations, but they also show how inequalities of science proxy as social inequalities.

Visualization Process

The investigation of visualizations in biomedical research initially requires a glance into the collection of works revolving around imaging practices and data analysis products. As end-products of big data, visualizations utilize specific tools useful in forming specific interpretations and meanings. Many scholars have looked into this understanding of the visualization's performative value as enabling of a particular meaning, or meanings, over others, which is especially important in this paper as these help reinforce a certain message (Burri & Dumit, 2008; Vertesi, 2014; Coopmans, 2010).

Scholars in the field of visualization studies also lead the research on viewing specific images that permit the audience to reach certain understandings through the power and persuasiveness of images (Burri & Dumit, 2008). This ability derives, in part, from the idea that science and technology have been seen to provide the highest possible level of objectivity—a tool for knowing truth—but only due to their respective societal and historical context. In contrast to today's value in mechanical objectivity—societally given value in knowledge obtained from technological advancements—people once preferred truth-to-nature objectivity. This paradigm shift not only parallels the change in how people understand images over time but also how scientific images rely on cultural shifts or preferences to create representations. The anthropological progress of science changes in tandem with its societal values. Volumes of work completed in this field reveal a long-existing desire to associate truth with seeing the underlying meanings in visualizations (Burri & Dumit, 2008; Beaulieu, 2001; Daston & Gallison, 2007).

The process of viewing visualizations depends on the social, cultural, and historical contexts in which the visual is generated. Awareness of this provides experts in various disciplines greater ease in reproducing their own interpretations for their audiences (Vertesi, 2014; Coopmans, 2010). It is the "intersection of scientific imagery with popular narratives and culture" that molds the conversations of scientific knowledge throughout non-academic society (Burri & Dumit, 2008). Scholars have studied how the collective expertise and collaboration allow for sharing of numerous ideas in visual and interactive instruments, algorithms, and concepts (Burri & Dumit, 2008; Stevens, 2015). This theme reappears through the case studies in the paper as the foundation for how these images are then used in particular ways.

The flexibility of visualizations to contain multiple interpretations creates problematic opportunities to develop different understandings of subjects and objects in the arena of knowledge production. Scholars note the disparity between what exists in the contents of the data and what can be done with the data (Burri & Dumit, 2008). Decisions made in image post-processing stages do not depend on technical and professional standards alone, but again on cultural and aesthetic conventions or individual expertise (Burri, 2001; Frow, 2012). The ambiguous nature of images affects the process in which society comes to view or understand these big data products.

Additionally, by defining big data images as having performative value, attending to visualizations as interactive also requires attention to the researchers' engagement with computers and other instruments (Burri & Dumit, 2008). More research must be done on the deployment of visuals and its relationship with technological advancement. Hype in visual persuasiveness (using visualizations as truth) remains a crucial part of contemporary science authority. Because of this, contemporary visuals created from scientific imaging technology garner great amounts of approval and promotion regardless of the data's underlying complexities and opaqueness (Burri and Dumit; 2008; Biteen *et al.*, 2015). This idea pans out not only in the case studies discussed below, but also in the rhetoric of modern day scientists.

Case Study 1: The Spectacular Microbiome

In early February 2016, I attended a few seminars in UCLA's annual International Symposium on Nanotechnology. Speakers from around the globe presented their research findings to fellow scientists and future sponsors alike. Most of the audience members were middle-aged and graying, mainly business-attired professionals. A packet that I picked up at the beginning of the conference doubled both as a research paper about nanotechnology research and as a detailed conference brochure. It may very well note the underlying message of this grand three-day conference. According to this paper, a main point of this event lies in attracting generous private investors and sponsors for future research projects that combine nanoscience technology with other fields of science. This conference did not focus on one discipline alone, but rather a combination of nanotechnology and the life sciences, physical sciences and medicine.

A talk by Dr. Pieter Dorrestein titled "3-D Cartography of Microbiome Chemistry" stood out in the midst of the microbiome series of the conference. Prefaced by a prominent UCLA microbiology professor, Dr. Dorrestein spoke about visualizing microbial diversity across the body in "more efficient ways." By "efficient," he meant using newer, more advanced chemistry technology. He spoke briefly about utilizing mass spectrometry and sequencing to calculate the number of microbes in his sample but did not go in depth about how he chose and collected his samples on the human body. Before even discussing his visualizations, he announced his latest research projects and visions of developing new algorithms and software for promising, to-be-uncovered spectrometry findings.

As he began to explain his 3-D visualizations, the most striking aspect of his work in the presentation was the flashing neon colors on the rotating human models. Red, blue, green, and yellow dots clumped together in some body parts like the hands and mouth. Other body parts like the forehead and thigh showed very little color. Dr. Dorrestein called this the ultimate 'spatial awareness' as these images work to show varying concentrations of microbes across the body. His next few slides showed his research in an article. This one in particular had been on the main page of CNN for four weeks and even featured in a children's book. He asserted that his visualizations highlight connections and relevancies about science research and human bodies that anyone can understand regardless of the complexities of studying microbiome science.

Ending his speech, he reiterated the need for advancement in technologies to continue relevant research similar to what he had shown. The audience clapped, and no one raised any post-presentation comments or questions. Every other research presentation before Dr. Dorrestein's had received multiple questions about clarification or future projects, yet the silence broke that tradition. The significant lack of questions at the presentation signals one of two things: all participants of the presentation understood the presentation or the blinking lights and rotating human models on the screen acted as a spectacular distraction from a complicated entanglement.

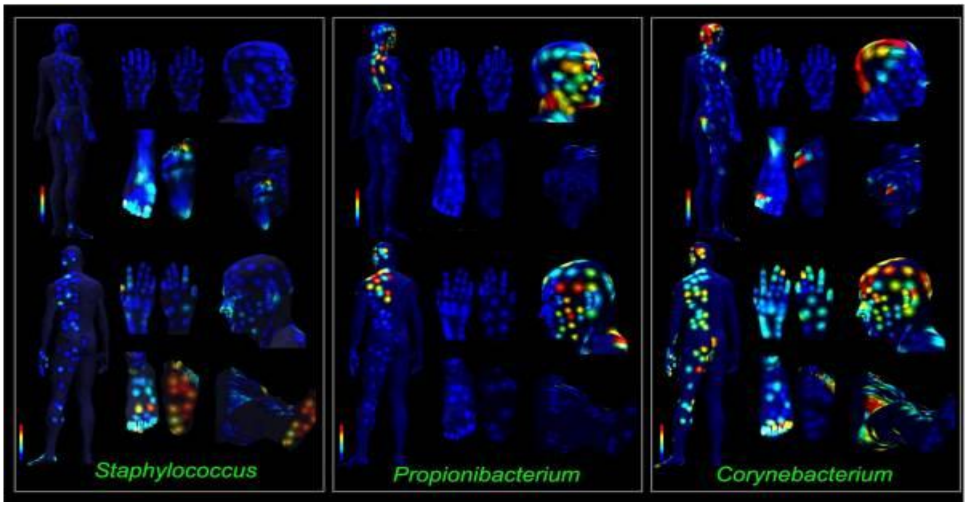

FIGURE 1. The topographical map of representative distributions of bacteria from the genera Staphylococcus, Propionibacterium, and Corynebacterium, based on their relative amounts at each body location, for the female and male individuals (image from Bouslimani *et al.*, 2015).

The research paper this presentation derives from hardly contains the same impressive features as shown at the conference. One of the most interesting parts of the presentation, the dynamic visualizations of proportions of microbes on the human body, was not present in the paper. Instead, the paper used two-dimensional static images that relayed only segments of what the media visualization showed (Figure 1). Notably, the researchers themselves noted that the whole body was not sampled. Specific pre-determined areas such as the hands, face, and feet were the main targeted areas of investigation (Bouslimani *et al.,* 2015). The actual lack of comprehensive sampling conflicts with the full-body human representations of the human models in the presentation and the paper as those images imply that the entire body surface was sampled. Selective presentation of data during the viewing experience promotes to the audience a specific interpretation of the data from the experience of viewing the research's information. This most notably reveals the ability of the visualization to manipulate interpretations beyond its superficial face value and questions the genuineness of science's methods and constitution (Coopmans, 2010). Understanding the relationship between appearance of the tactful presentation and reality of the complex, messy dataset unveils a much-needed investigation into the research.

In the sciences, neglecting to study a specific body part implies not only a hierarchy of body part importance but also a variety of other disparities. The selection of specific body parts to sample presents a preexisting bias in the realm of microbiome science. If future skin microbiome studies of other body parts are conducted, this research (containing no data on non-hand, face, or feet areas) may impede newer research due to bias and lack of data. Moreover, the reason why these researchers decided to collect data from hands, feet, and face only may originate from a non-scientific understanding of how humans acquire disease and illness.

Cultural understandings of how disease is acquired may manifest in the form of hesitance to shake the hand of someone who just sneezed or can explain why some people dislike wearing shoes in their bed. The idea of seeing "dirt" or "sickness" physically move from object to hand, human to human, unnerves many. The same theory applies to the stigmatization of diseases in the 1980s—it was believed that any type of contact, sexual or non-sexual, to anyone with HIV would instantly transfer the disease to oneself (Francoeur & Noonan, 2004). Visualizations that pinpoint the identities of foreign and unknown bacteria on the body embolden inaccurate interpretations of a tight relationship between bacteria and disease (Figure 2). Unfortunately, this concept of the human body as a tightly-bound, individual unit does not follow current understandings of health in the human body. Health does not rely on the human body avoiding foreign objects, but relies on the intertwined relationship between the body and the environment (Allen, 2014). While this research seeks to show how unknown microbes may non-toxically coexist on the skin, the connection it makes between folk understanding of disease acquisition and current data still enforces the prior.

In fact, important consideration must be taken in pointing out the persuasive performance of these visualizations at the presentation. While the research papers present a static representation of a supposedly live and breathing subject, the dynamic images with blinking lights act just like the moving microbes on the body. The multimedia essence of the image allows for it to represent a dimension beyond spatial ones: time. The inclusion of time and movement mimic the natural world in which we live and perceive. Therefore, the movie-like images in the presentation allowed the audience to automatically perceive the movement of microbes. In a sense, these visualizations garner much power, mainly due to their close familiarity to human perceptions of nature (Burri & Dumit, 2008). Shown to a live audience, a two-dimensional graph that plots time performs differently from a moving video showing time passing. This push for replicating time-sensitive results found in nature is a repeating theme in Dorrestein's work, and he frequently cites the importance of advancing these objective technologies (Hsu, 2013; Hsu, 2015; Bouslimani, 2014). This drive for newer, faster, and more precise technology also comments on the scientific trend of automatization.

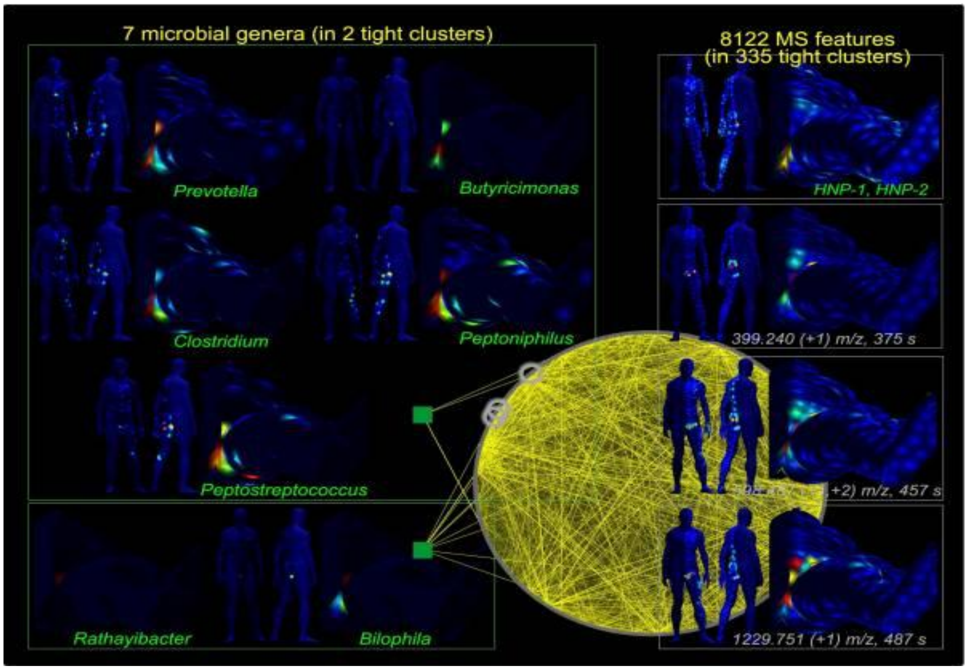

FIGURE 2. Molecular and bacterial communities found to be co-localized in an individual (image from Bouslimani *et al.*, 2015).

Modern machines and technologies, such as those used in Dorrestein's research, strive towards capturing realism in complex datasets. More frequently, interpretations from these technologies are seen as objective due to the diminished presence of human interference or subjectivity. Thus, viewers place greater authority in such visualizations deriving from a mechanical or technological production, which may allow scientists to place value in the same modes of understanding. Visualizations assume the traits of objectivity and hide flaws in current science experimentation through the veil of simplicity and suggestion.

Case Study 2: The Traveling Microbiome

The old saying "using a subway handrail is like shaking the hands of 100 strangers" has never been so understated in a city of now over eight million people. In New York City, daily exchanges involving contact between strangers on the streets, buildings, parks, and bodegas occur with similar frequencies in the underground world of subway tunnels. With 659 miles of subway track, 468 subway stations, and 24 various subway lines, the Metropolitan Transport Authority boasts a daily weekday ridership of 7,660,605 passengers (MTA, 2015). The subway is filled with people taking their first ride a well as people taking their daily commute home. Down the street or across the bridge to another borough, the subway system transports people both near and far, reaching destinations as far from Manhattan as Queens and Coney Island. As an unintended result, this diverse, cheap, and convenient system also disperses a plethora of bacteria and other microorganisms with the same vigor it carries its human passengers from station to station.

A recent project of Dr. Christopher Mason at Weill-Cornell Medical College involving the study of New York City's subway microbiome has been the focus of harsh criticism and unexpected controversy. In 2015, Mason launched his highly interactive site, PathoMap, based off of his data from collecting and identifying microbes on the surfaces of New York City. Expert scientists from the fields of virology and genetics, confounded by the results, attacked Mason's work, and a heated debate regarding the validity of the data and conclusions generated in Mason's project ensued. The discourse between these scientists follows the emergence of microbiome awareness in daily life conversations; more people outside of academia are familiar with the concept of the microbiome. The data shown via visualizations both obscures and reveals to create certain meanings, namely interpretations that vary from expert to expert. Thus, these controversies among experts segue into recurring issues of studying microbiome science in populations.

Due to the large-scale nature of this project, the research team called for crowdsourcing, or the inclusion of people outside of the research laboratory, to help with the project. Researchers and interns, along with student volunteers from local colleges and public citizens of New York City, collected samples from the entire New York City MTA subway system in triplicate—two swabs from the station and one swab from within the actual train (PathoMap Research Team, 2013; Afshinnekoo, 2015). Objects with a presumed high level of human-surface interaction, such as turnstiles and kiosks, were prioritized over other objects depending on who collected the samples (Afshinnekoo, 2015). For the scientists, the preference of sampling one type of surface would allow them “the greatest comparability options” among the different data sets gathered. On the contrary, the individual preferences among the volunteers may have caused discrepancies in the comparability of the samples they collected as the surfaces they choose reflected their own interests. With the volunteer’s free-for-all mindset and few limitations on what they could collect, scientists analyzing the data neglected to ask more questions of the identity of collectors as a variable before beginning to compare samples.

In the visualization result, as seen in the research graphics (Figure 3), the high level of microbial diversity of a subway station corresponds to a supposed large amount of population diversity. Elusively, the actual dataset hardly explains the sampling process. According to their interactive PathoMap, each point on the map represents one collection swab (Figure 4). A closer look into this graphic reveals uncertainties regarding the quality and accuracy of each subway’s sample. The sampling goal was to have at least two swabs at each station. Procedurally, each point on the map representing a specific swabbed location should cluster in a group of two or more points. Representation, as in the location and type of sample, plays a role in creating disruptions between the data and its product. The map illustrates how areas farther away from Manhattan have fewer than the required number of samples (Figure 4). As evidenced by the visualization, the study shows that more volunteers sampled subway stations in Manhattan than in less populated boroughs like Staten Island or Bronx. Even with all locations sampled, the quantity of samples varies location to location, and the amount of available data gets called into question.

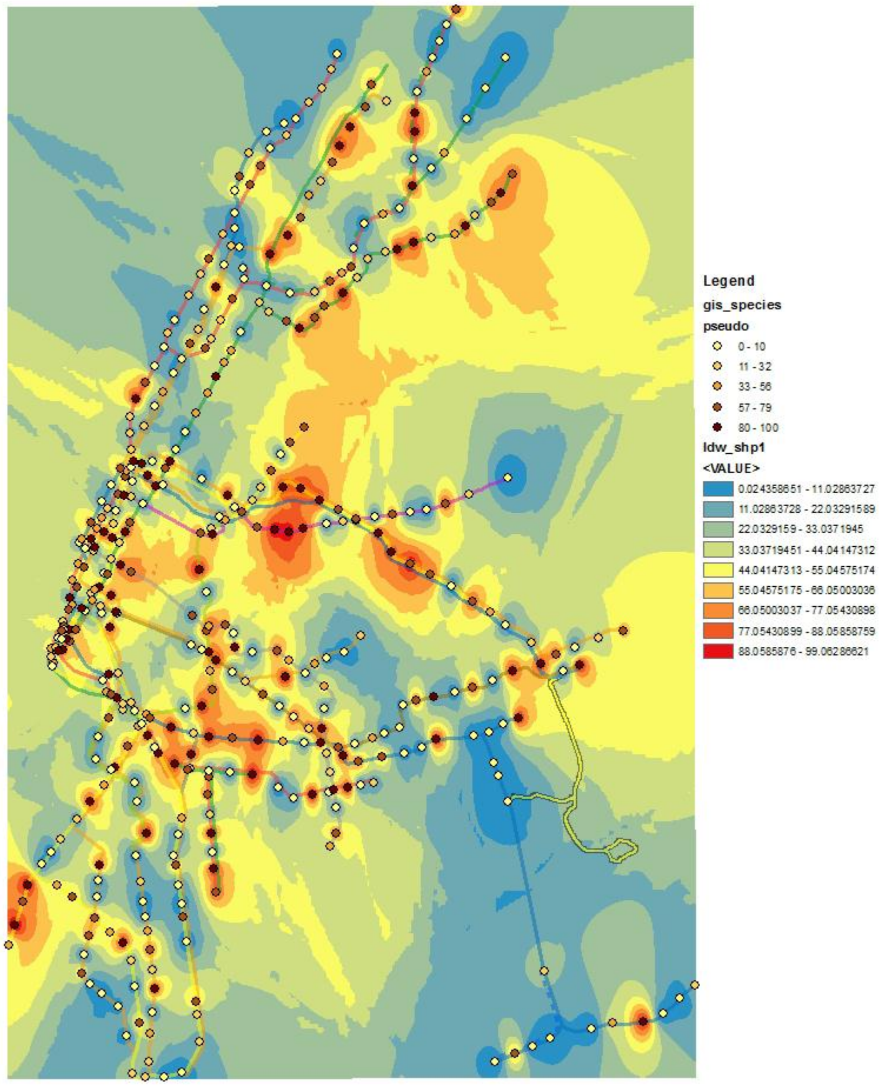

FIGURE 3. Heatmap of the Pseudomonas genus, the most abundant genus found across the city (image from Afshinnekoo, 2015).

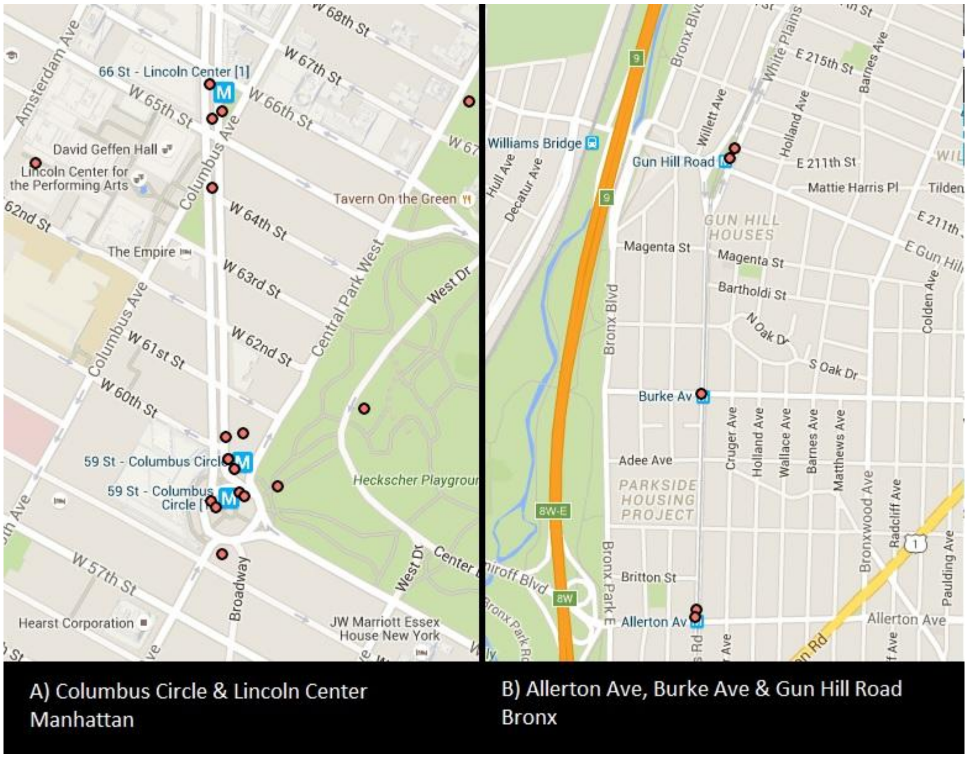

FIGURE 4. Comparison of amount of samples taken in Manhattan subway stations and samples taken in Bronx subway stations. Source data from pathomap.giscloud.com.

Due to the lack of clarity in the sampling methods of Mason's project, other aspects in the production and analysis of the data can also be shown to be overly complex. Obscurities in the experimentation and data processing directly affect the other results and graphics of this research. The most prominent and influence-heavy visualization in Mason's data uses the collected DNA microbial sequences at each station location to link calculated ancestry with census data. Their goal was to show how microbiome science has the capability to identify population demographics (Figure 5). Because of growing ease and access to genetic testing, scientists have been increasingly interested in using genomics to provide scientific evidence for racial categorization (White & Duster, 2011).

Mason's project is not exempt from trying to tie race into biology. The image contains three columns shown side-by-side: the genetic testing results, the 2010 census data, and a column showing an estimated statistical difference between the first two columns' racial percentages. While most of the estimated ancestry shows high correlation to the census data, the first two columns also reveal jarring differences. The third column, appropriately titled "deviation," represents the statistical differences between the first two columns.

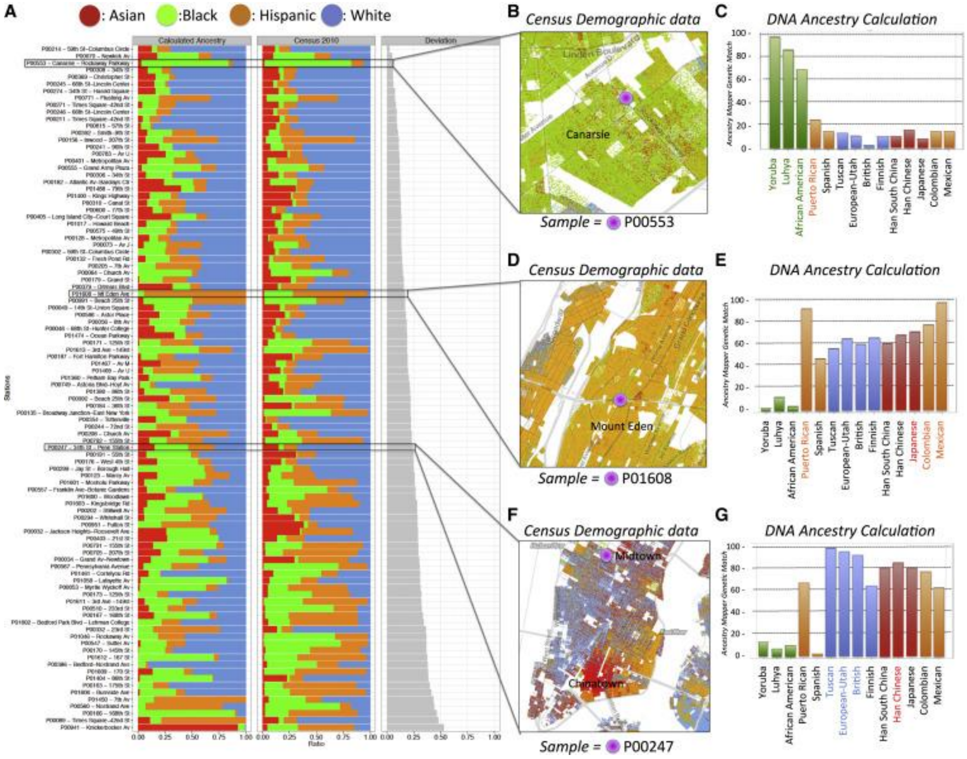

FIGURE 5. Human Ancestry Predictions from Subway Metagenomic Data Mirror Census Data Using ancestry (Afshinnekoo *et al.*, (2015).

Unfortunately, no portion of Dr. Mason's paper explains deviation or the factors that lead to it. The other images in this visualization explore the calculated ancestry and census data correspondence, ignoring the details about how "deviations" in the data arise. Close examination of the race data shows that high amounts of deviation (over ~25%) come from stations where the estimated proportion of white inhabitants are higher than what the census data shows (Figure 5). No exact proportions exist in the data or visualization, so analysis of this graphic requires approximations and individual interpretation of the values.

Noticeably, the calculated ancestry column tends to underestimate the proportion of Hispanics in the majority of stations; however, it is difficult to ascertain if the black population also faces this issue. The top portion of the calculated ancestry chart seems to overestimate the proportion of blacks yet towards the bottom of the chart, with a higher percentage of deviation, black populations are underestimated. In the areas with higher deviations, the calculated proportions of whites are overestimated (Figure 5). The deviation column utilizes a numerical value that cannot adequately address potentially more descriptive values—social, cultural, ethnic, or political issues—occurring in these locations. The preference for quantitative results over qualitative effects shadows the demand for a more holistic understanding of the resulting science and society difference.

Understanding how this project's sampling works roots out its issues of using insufficient data to paint a picture, or in this case, generate a microbiome map. Not enough data has been released to fully understand PathoMap visualizations, yet even with the lack of labels and details, the images present a persuasive message. In Mason's chosen census demographic data in the graphic, the subway systems near Canarsie and Mount Eden incorporate highly racially homozygous populations and have lower levels of deviation; however, these locations lack as extensive a swab sampling collection as stations in the heart of Manhattan (Figure 5, B-G). The snapshot of subway DNA data at a remarkably more diverse and heterozygous populated station, Penn Station, shows higher levels of deviation. The quality and quantity of samples taken at Penn Station outnumber those gathered in neighboring boroughs.

Once again, these observations touch upon the topic of sampling, but more specifically the perspective and interests of volunteers collecting the samples. As a project relying on the effort of people outside the laboratory team, PathoMap needs further analysis into the identity of their volunteers. Sampling inequalities show through the varying quantity of submissions and how those outside the laboratory realm most likely do not operate with the same scientific mindset as those inside. How these people understand the PathoMap project and its technology may factor into which subway stations, if any at all, they swab and what they swab.

Crowdsourcing research experiments create a boondoggle for scientists analyzing the collections who then have to decide whether or not to overlook disparities, such as sampling inequalities, in the research. However, aside from inexplicit sampling methods, there exists another disruption in comparing data from genetic ancestry tests and data from the census. Utilizing race as a category in any health-related science prompts a discussion on racial profiling in medicine and health. Ultimately, as PathoMap seeks to prevent permanent disease outbreaks, gain knowledge of bioterrorism, and manage health in the context of large populations, the data from this research is not immune to conversations of race in health (Afshinnekoo *et al.*, 2015). In fact, "emerging genetic knowledge has the potential to transform contemporary notions of social coherence and group identity" (Brodwin, 2002). In Mason's research, the comparison between expected genetic results and actual census data shows outside factors that work against using science to define identity. Biology and genetic testing may only somewhat explain the identity of populations. With carefully crafted graphs and maps, the researchers struggled with defining social constructs, like race, that appear in science. The deviations that occurred between calculated ancestry and census data does not solely stem from problematic sampling, but also from trying to measure qualitative ideas with quantitative data. Endeavoring to measure identity, as PathoMap attempted, demonstrates the difficulty of measuring race, which may have societal origins that may not be accounted for in microbiome sciences.

Conclusion

With complex datasets found in the realm of microbiome science, it is easy to automatically accept convenient interpretations of visualizations. These extensively dense datasets form a multiplicity of interpretations outside a strictly molecular understanding. From attending a presentation on the microbiome to studying controversial cases like the New York City subway PathoMap, it is clear that interpreting big data visualizations, in all situations involving scientific research, warrants caution.

Throughout this paper, scientific controversies and disjunctions appear in visualizations to show different types of inequalities in the realm of big data. Sparse sampling and lack of clarity in experimentation prove problematic in creating fluid and encompassing associations. Utilizing race as a category becomes an issue as racial profiling in health and medicine requires extensive definitions of race in order to create a cohesive argument. Like most genomic sciences, challenges arise when qualitative factors such as race are subject to quantitative representation. The capacity of visualizations to be persuasive comes from their relatability and ambiguity. Lastly, aggressive pushes for technological advancement reveal cultural and social shifts towards automatization. The conclusion from the interplay of the interactive maps and research visualizations illustrates the idea that images produced from complex datasets contain innate difficulties. These complexities dwell in the overwhelming nature of current microbiome science.

Considering these limitations of big data visualizations, further questions must be asked about the outcome of big data. An important concern of microbiome science revolves around continued funding. Considerations of non-scientific entities come into play when applying for grants and funding sources. Currently, large corporations and government institutions in healthcare fields lead the search for solutions driven by big data science with private interests following closely behind (Bouslimani, 2014; Biteen, 2015). How can the extraneous agents of science, such as funding, ethnics, and media, affect the way visualizations are produced and deployed? Should these research agendas proceed to play a major part in big data sciences, and how can visualizations avoid the influence of these politics?

Continuing visualization studies in the realm of big data science is vital to better understanding the inequalities present in both science and society. Progress in the field of microbiome visualizations means greater transparency in how the data is used. Small inclusions to the big data visualizations (such as explicitly stressing experimental and analytic limitations up front) enhance clarity and are vital to the improvement of the field. Larger efforts to optimize and standardize the processes and protocol of visualizing big data requires the combined effort of scientists in all fields, from social scientists to medical scientists.

References

Afshinnekoo, E. *et al*. (2015). Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. *Cell Systems, 1*(1), 72-87.

Allen, M. (2014). *Misconceptions in Primary Science*. McGraw Hill: NY.

Alois, J. D. (2015). PathoMap Finds DNA Fragments of Anthrax & Bubonic Plague in NYC Subway. Retrieved from http://www.crowdfundinsider.com/2015/02/62293-pathomap-finds-dna-fragments-of-anthrax-bubonic-plague-in-nyc-subway/

Beaulieu, A. (2001). Voxels in the brain: Neuroscience, informatics and changing notions of objectivity. *Social Studies of Science, 31*(5), 635-680.

Biteen, J. S., Blainey, P. C., Cardon, Z. G., Chun, M., Church G. M., Dorrestein, P. C. … Young, T. D. (2015). Tools for the microbiome: Nano and beyond. *ACS Nano, 10*, 6-37.

Bouslimani, A., *et al.* (2014). Mass spectrometry of Natural Products: Current, Emerging and Future technologies. *Nat Prod Rep, 31*(6); 718-729.

Bouslimani, A., *et al.* (2015). Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci USA, 112*(17); E2120-9.

Boyd, D. & Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Brodwin, P. (2002). Genetics, Identity, and the Anthropology of Essentialism. *Anthropological Quarterly, 75*(2), 323-330.

Burri, R. V. & Dumit, J. (2008). Social studies of scientific imaging and visualizations. In E. J. Hackett, O. Amerstdamska, M. Lynch, & J. Wajcman (Eds.), *The Handbook of Science and Technology Studies* (pp. 297-317). Cambridge, MA: MIT.

Coopmans, C. (2010). 'Face value': New medical imaging software in commercial view. *Social Studies of Science, XX*(X), 1-22.

Daston, L. J. & Gallison, P. (2007). Epistemologies of the Eye. *Objectivity* (pp. 17-53). Brooklyn, NY: Zone.

Francoeur, R. T. & Noonam, R. J. (2004) *The Continuum Complete International Encyclopedia of Sexuality.* Continuum International: NY.

Frow, E. K. (2012). Drawing a line: Setting guidelines for digital image processing in scientific journal articles. *Social Studies of Science, 42*(3), 369-392.

Jansson, J. K. & Baker, E. S. (2016). A multi-omic future for microbiome studies. *Nature, 1*(5):1-3.

Horgan, R. P. & Kenny, L. C. (2011). SAC review 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet Gynecol, 13*, 189-95.

Metropolitan Transportation Authority. (Feb 2015). *Public Transportation for the New York Region.* Retrieved from web.mta.info/mta/network.htm

PathoMap Research Team (2013). Prezi; https://prezi.com/wr203owl13io/pathomap-kickstarter-video/

Stevens, H. (2015). Networks: representations and tools in postgenomics. In S. S. Richardson and Hallam Stevens (Eds.), *Postgenomics: Perspectives on Biology after the Genome* (pp. 103-125). Westchester Publishing.

Ursell, L. K., Metcalf, J. L., Parfrey, L. W., & Knight, R. (2012). Defining the Human Microbiome. *Nutr Rev, 70* (Suppl 1): S38-44.

Vertesi, J. (2014). Drawing as: distinctions and disambiguation in digital images of Mars. In C. Coopmans, J. Vertesi, M. Lynch, & S. Woolgar (Eds.), *Representation in Scientific Practice Revisited* (pp. 15-36).

White, P. C. & Duster, T. (2011). Do Health and Forensic DNA Databases Increase Racial Disparities? *PLoS Med, 8*(10).