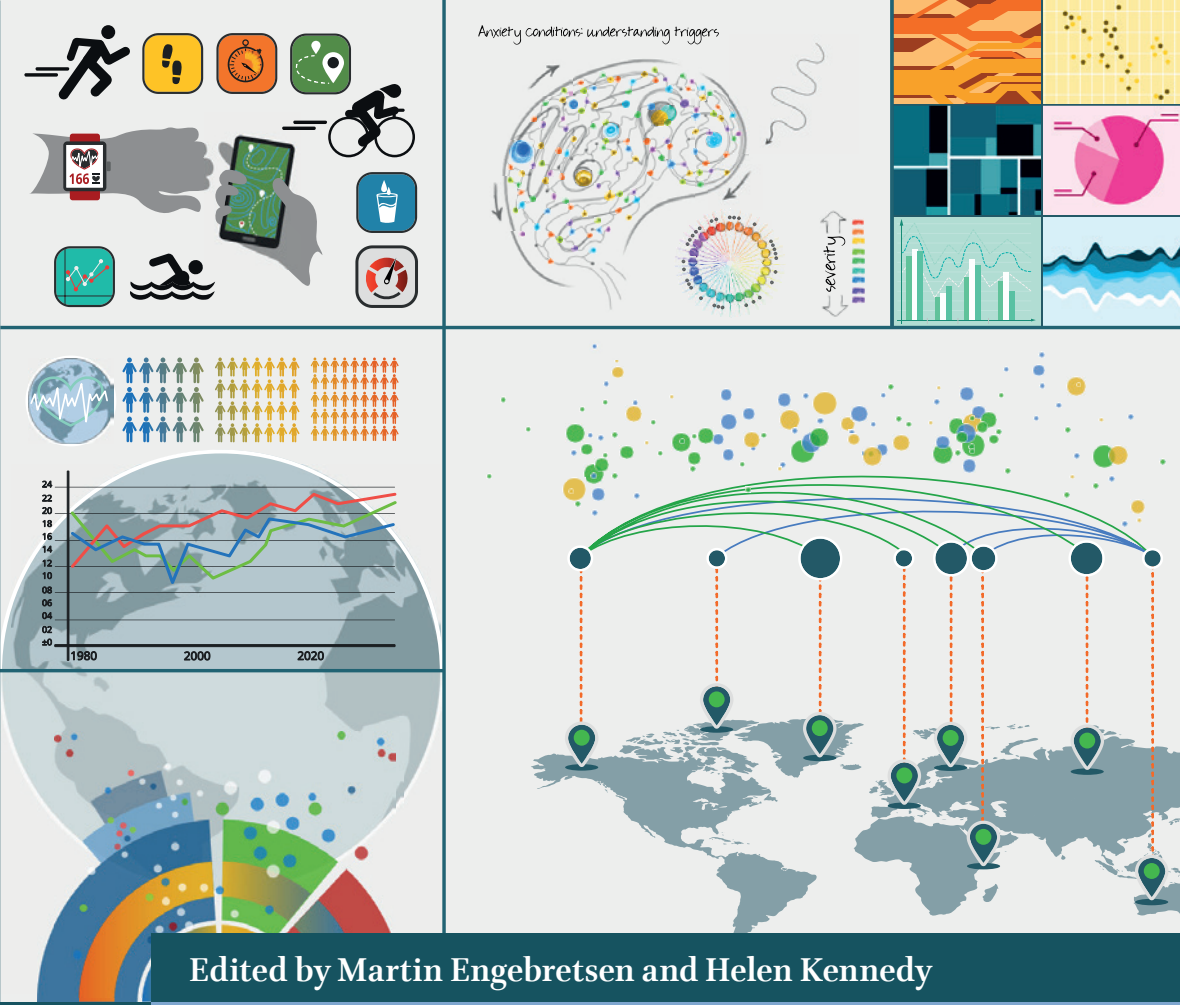

Edited by Martin Engebretsen and Helen Kennedy

Data Visualization in Society

Amsterdam University Press

22. Visualizing diversity: Data deficiencies and semiotic strategies

John P. Wihbey, Sarah J. Jackson, Pedro M. Cruz, and Brooke Foucault Welles

Abstract

This chapter explores the complicated dynamics that are inherent to the practice of data visualization involving issues of race and identity. We focus on data from the US Census and the profound questions that are raised as visual forms purport to represent groups. After reviewing historical context and related limitations and controversies, we present a project that explores a novel approach to visualizing US immigration patterns, an approach that relies on visual metaphors and algorithmic construction of visualization patterns based on massive sampling of Census microdata. The chapter suggests that the use of innovative expressive techniques to convey insights through poetic, and thus less literal, and limiting, forms is a way of grappling with underlying deficiencies in administrative population data.

Keywords: Data visualization; Immigration; Race; Diversity; Computational design; Data art

Introduction

The vocabulary of diversity, pluralism, multiculturalism, and the proverbial 'melting pot' are often invoked in contemporary discourse to characterize the complex, highly fraught, and extraordinarily multilayered history of immigration, race, and cultural identity in the United States. Ideas about American identity—who people truly 'are' at some essential level, where they come from, how they choose to be identified, and how majority cultures may identify them—continue to evolve over time. Through this discursive

Engebretsen, M. and H. Kennedy (eds.), *Data Visualization in Society*. Amsterdam: Amsterdam University Press, 2020
DOI 10.5117/9789463722902_CH22

space, groups work to access forms of cultural and political recognition and resources, all the while potentially excluding and/or including others as the boundaries of identity are asserted, renegotiated, and contested.

These dynamics echo throughout American history, and manifest in a series of vexing questions: Who is to be counted as a citizen, with full associated rights? Who is included or excluded from a wide variety of identifying categories, such as 'Indian/Native American', 'Asian', 'Latino', or white? What is 'blackness' or 'whiteness'? Who should use hyphenated identities based on unique descent and ancestry, and why? How are multiracial persons, a growing portion of the population, to be identified? While cultural debate has been, and likely always will be, sprawling and unsettled around such questions, the formal locus of this debate is the decennial US Census, mandated by the Constitution to count persons.

Media representations of many kinds—novels, films, songs, paintings, journalism—have been used to explore the changing nature of the country, bringing to light, for example, how enslaved and indigenous peoples, and their descendants, have struggled to gain equality and how waves of immigrants have entered the country and challenged dominant power structures maintained by white Protestants of European descent. Such media representations have played a vital role in reconceptualizing notions of what it is to be 'American' and in surfacing important experiences that may have otherwise been culturally marginal.

As a relatively newer form of media particular to the digital era, interactive data visualization provides novel affordances that open up new possibilities for exploring evolving notions of human identity. In this chapter, we present an example from our own work which attempts to push the boundaries of discourse about labelling and identity. This unique project leverages administrative data from the US Census to tell the sweeping story of immigration history and cultural identity in America. The project, which draws upon Census recordings of persons' countries of origin primarily over the period 1830 to 2015, deploys visual metaphor and computational techniques to expand the expressive meanings and possibilities around themes of diversity. We see the project as a particular form of discourse that both grapples with the challenges of reductionism and inclusivity/exclusivity, and that semiotically projects complex ideational and compositional meanings that speak deeply to a general theme of cultural diversity. Because deficiencies in data are a problem for all sorts of reasons, including visualization challenges, we sought to address these by experimenting with a form of visualization that works with limited/deficient data.

To contextualize the experimental case study we produce, we first situate this visualization work in the intellectual history relating to the underlying census data and the limitations embedded in it. We explore how any picture of 'diversity' based on these administrative data necessarily, and tragically, excludes certain types of persons, with African Americans and Native Americans being two categories of persons whose origins in this country do not fit into narratives about diversity through immigration.

With these caveats in mind, our project nevertheless focuses on the census 'country of origin' information, using statistical estimates, to render a picture of American diversity that evolves, grows, and complicates understanding over time. There are, of course, many ways of portraying diversity, and immigration is one of those: it is a subset of diversity. As will be explained, given how unrepresentative race data in the census are, we focused on immigration specifically, as extracting the immigration data is a much more accurate task that provides a reliable basis for visualization. We chose to deploy the visual metaphor of tree rings to evoke the complexity and interdependence of a biological ecosystem. Historical immigration patterns are shown as a set of tree rings, which are encoded by processing millions of samples of US Census microdata, from a pool of nearly 2 billion individual records. As time advances, the tree grows, forming rings of immigration. Each ring corresponds to a decade. Cells are deposited in layers, with each cell corresponding to 100 immigrants.

Our efforts focused on a central research question: *Given the known constraints, what would a dynamic picture of US diversity, as a function of immigration, look like?* Further, how might artistic, design, and poetic strategies work to enhance knowledge and interest in the diversity of the country, signifying truths and conveying important insights that may transcend the limitations of underlying, literal data? Interventions around such a question, of course, bear crucially on urgent political questions and current discourses about cultural diversity and public policy proposals, and we take up these questions with this background in mind.

Visualizing migration and identity: A brief US history

During the second half of the nineteenth century, visualizations of Census-based numbers first began appearing in government statistical abstracts; some of these figures began examining the distribution of different ethnic groups throughout the country as a function of immigration (US Census Bureau (n.d.). Statistical abstracts of the United States). Immigration as a

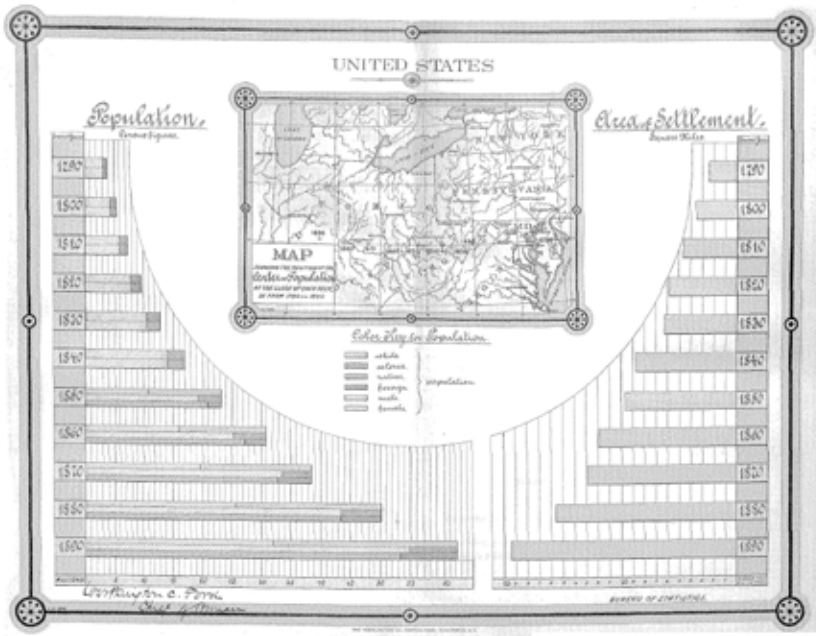

Figure 22.1. Map/chart included in 1896 US Census documents, showing growth of racial and demographic groups and territorial expansion. From US Census Bureau (1896). Statistical abstract of the United States 1897—Part 2. (https://www.census.gov/library/publications/1898/compendia/statab/20ed.html). Public domain.

phenomenon also became known through non-quantitative representations such as drawings, posters, paintings, and other hand-drawn and printed media forms. The question of place of birth was added in 1850, following the beginning of a dramatic increase in immigration (Gibson & Jung, 2006).

Of course, migration and the movement of peoples are network-driven processes, lending themselves readily to visualization (Portes & Rumbaut, 1990). Immigration records are limited, though, and certain ethnic groups can only be traced back so far; thus most representations are constrained by the available data (Daniels, 1989). Full-scale histories that attempt to recover the nuances of European and American migration, for example, have rarely been attempted (Nugent, 1995). In any case, the United States began keeping records of persons entering the country at ports in 1820 and, although prone to inaccuracy, this gave way to an idea of change in population volume due to external flows and eventually to visualizations of these numbers (Handlin, 1959).

The history of data visualization relating to US immigration is not well documented, and to our knowledge, there is no extant comprehensive history. We performed an environmental scan of the literature/relevant materials

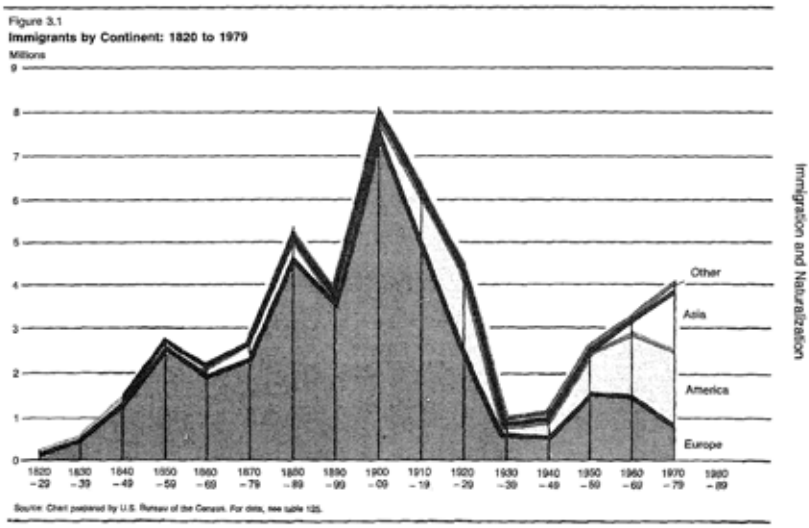

Figure 22.2. Chart included in 1983 US Census Bureau documents, showing the relative contribution of various continents to immigration totals in the United States. From US Census Bureau (1983). Statistical abstract of the United States: 1984—Section 1 Population. (https://www.census.gov/library/publications/1983/compendia/statab/104ed.html). Public domain.

and found a number of visualizations in the US Library of Congress's virtual trove of historical documents; US Census Bureau materials, especially the statistical abstracts; statistical atlases of censuses; books about immigration; and on digital news sites and data blogs. While far from a comprehensive search, we examined numerous visualizations spanning from 1828 to 2018. Representative examples included a 1984 map (see Figure 22.1) from the US Census that depicts immigrants by origin from 1820 to 1979 (US Census Bureau, 1983), as well as data visualizations from contemporary media outlets such as Vox that show 200 years' worth of data trends (Chang, 2017).

Traditionally, most visualizations of immigration to the United States have involved some sort of map, including land plot, county, density, and flowchart maps. An early example printed by the Census Bureau (1896) illustrates how the intersection of identity and geography were being represented and imagined in the nineteenth century, with categories of 'white', 'coloured', 'native', and 'foreign' delineated (see Figure 22.2).

As will be discussed, certain classes of people are wholly excluded from any such maps. Glaringly, the precise African countries of origin of slaves and their ancestors are not included in this historical narrative, nor are the indigenous nations from which Native Americans came, even as they became US citizens through subjugation. That said, it is a point that bears further

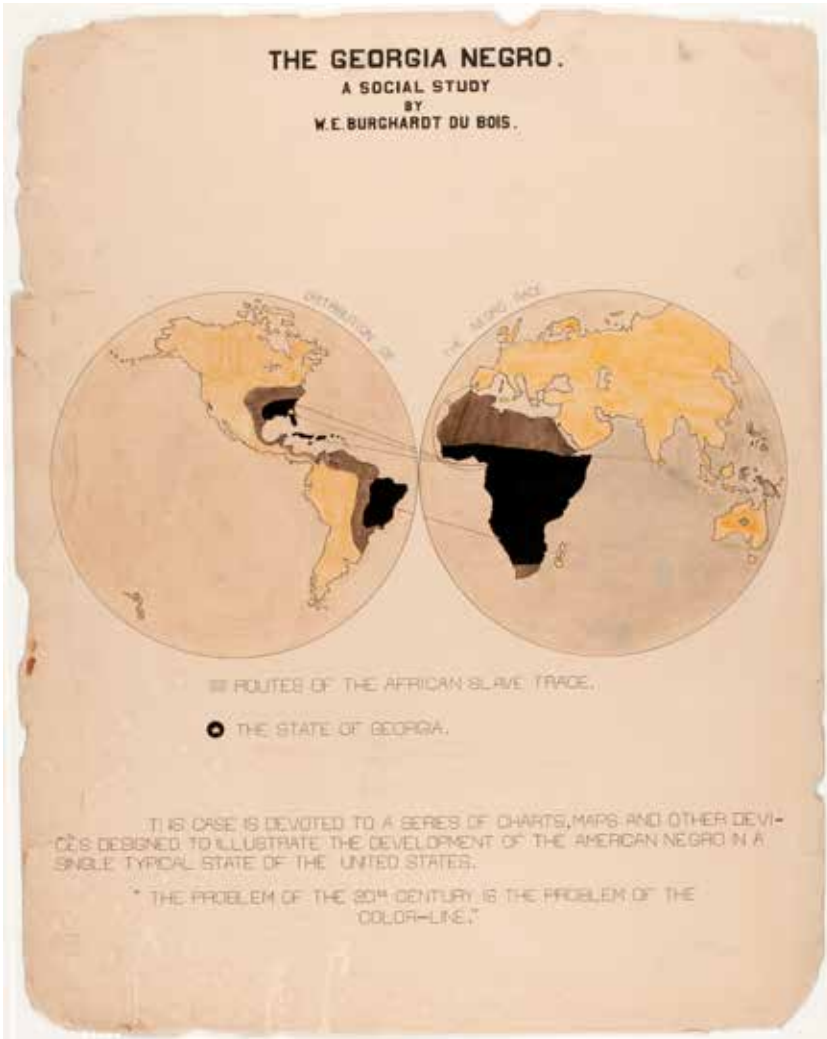

Figure 22.3. Historical rendering by sociologist W.E.B. Du Bois of trajectory of African slave trade to the Americas. From Du Bois, W.E.B. (1900). *The Georgia Negro: A social study.* [Map] Library of Congress Prints and Photographs Division, Washington, D.C. Public domain.

research that there are early examples of both African-American scholars and folk artists and Native Americans tracing their own history through visualizations. These would include, for example, a 1900 map created by the pioneering social scientist W. E. B. Du Bois about the trajectory of the slave trade from regions of the African continent to the Americas (see Figure 22.3).

For Native Americans, there are examples of data-related artefacts such as 'winter count' calendar-pictorial cloths and skins (Lakota, 1902), as well

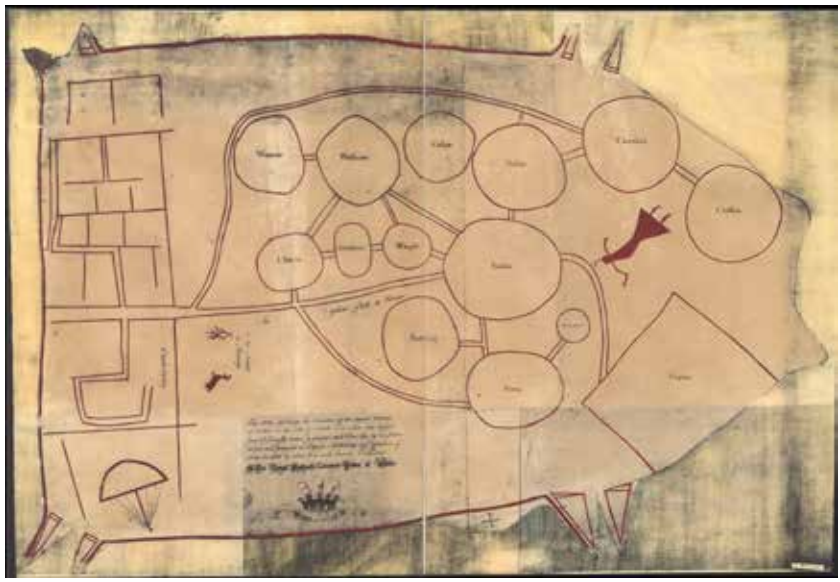

Figure 22.4. Native American map rendering on deerskin of tribal information and location. From Nicholson, F. (1724/1900). *Map of the several nations of Indians to the northwest of South Carolina*. [S.l.: s.n.] [Map]. Retrieved from the Library of Congress. (https://www.loc.gov/item/2005625337/). Library of Congress, Geography and Map Division. Public domain.

as examples of local and regional maps with population features done by indigenous persons (see Figure 22.4).

Exclusion in Census data

The US Census is intimately tied not only to questions of identity but to power and inequality, and the racial categories encoding many generations of non-white persons remain highly problematic as data sources. The categories included in the first Census in 1790 speak to this: 'free white' men and women were counted with specifics about their age, dates of birth or death, profession, and familial role. Next to these categories appear only two others: 'non-taxed Indian' and 'Slave'. Together these categories reflect the power and entrenched racial ideologies central to the nation's inception: white men and women mattered and were counted in detail because their numbers (and economic and social capital) mattered to questions of political representation and policy; Native Americans—who include over 500 tribes with distinct cultural practices and languages—are lumped into the category of 'non-taxed Indian', a term that reflects their non-citizen

status and non-inclusion in apportionment counts in determining political representation; and finally, 'slaves', a term that, to those in government, stood in for people of African descent in the Americas so clearly that they did not bother to specify what we now think of as a racial category because 'slave' and Black were presumed to be synonymous (Zinn, 2015).

Rendered invisible for decades following the first census were Asian Americans and Latinxs—groups that have long been a part of the American fabric but whose early numbers were considered too low to matter enough to count, who were not yet 'raced' in the American imagination, or who primarily resided in parts of the country yet to become politically consequential. For example, in the 1860 census, the brand new state of California included 'Chinese' as a category in the census—a reflection of the presence of Chinese labourers in the West—but this category was not included in any other state (Hart, 2009).

Likewise, much of the Southwest between Census years 1790 and 1860 was either not yet a part of the United States (rather controlled variously by Spanish or Mexican governments) or relatively new territories and states without much population or political representation. Thus, the need to count those who would now be considered Latinx—and in fact even a federally recognized racial or ethnic category to describe their various origins—simply did not exist. It was well over 100 years after the 1860 census, in 1970, before the federal government would make the first attempt to count 'Hispanics' as an ethnicity (Cohn, 2010; US Census Bureau (n.d.). Measuring Race and Ethnicity Across the Decades: 1790-2010).

The evolution of census categories is a clear example of how racial categories—while socially constructed and ever changing in response to sociocultural context—are central to questions of inequality and belonging in the United States. As various groups have sought to maintain and gain power throughout American history, and as socio-political contexts have been shifted by war, labour demands, economic upheaval, migration, and activism, 'race' as an identity category worth counting has shifted, as well. Over time, the US Census Bureau, entrenched in the original exclusionary ways of thinking about identity visible in the 1790 census, has responded, sometimes slowly and under pressure and sometimes rapidly when groups are deemed a threat, to these shifts. Generally, changes in census categories are spurred by new understandings of who in America should be counted—who matters. Importantly, however, mattering and being counted are not always a positive thing: take, for example, the case of 'slaves' who mattered to their owners for the purposes of economic gain and political representation; or the Chinese whose counting lead to the passage of the draconian 1882 Chinese Exclusion Act.

The creative data visualization project we discuss below begins in the nineteenth century; with its focus on immigration, the project's substantial visual forms really begin to take shape with the 1870 census—the first after the Civil War. As important context, it is worth noting that this census reflects how deeply important questions of white racial purity became to those in power in the context of reconstruction and the attempted social gains of African Americans. The category of 'slave' as a stand-in for African American is removed. In its place new categories arise alongside 'White' and 'Indian': 'Black', 'mulatto', 'quadroon', and 'octroon'. These categories reflect the racial anxieties of whites in power during reconstruction who embraced racial pseudo-science based on mythologies of 'Black blood' to justify their fear that increased gains by African Americans would lead to mixed-race children who would sully the purity of the 'white race' and throw the existing racial order into chaos. Suffice it to say, census workers carrying out counts in the nineteenth century—and well into the twentieth—were given detailed instructions that would both offend and appal relative to contemporary standards, about how to assess and record racial distinctions and determine the cultural identity of many different kinds of people.

It was in 1890 that Asian Americans began to be counted in the national census—largely as a result of increased immigration of Japanese and Chinese men who worked first as labourers in agriculture and railroads and whose increasing numbers were perceived as an economic and cultural threat. This is reflected in the fact that *only* the categories 'Chinese' and 'Japanese' are added to the census at the time despite the lesser presence of Korean, Filipino, and other Asian labourers in the same industries (Takaki, 2012). Over time, the counting of Asian Americans by the census became more inclusive, sometimes in response to perceived threats, and other times as a result of political activism and lobbying by Asian American groups who sought to challenge the perception of Asianness as an always-unamerican-Other category. Among the most shameful examples of how questions of power and oppression are tied to the census is that the United States government used records from the 1940 census to find and intern Japanese families during World War II—illustrating that a demographic survey in the context of xenophobic ideology is anything but a simple count (Aranti, 2018).

Between 1900 and 1940 the mulatto, quadroon, and octroon categories were dropped from the census as racist blood quantum science was debunked and it became clear that the mere existence of mixed-race African Americans would not, in the context of entrenched American anti-blackness, dissolve the conditions of the black/white racial binary. During this period also, the 'race' category of 'Mexican' came and went from the 1930 census, and the

categories 'Hindu' and 'Korean' were added in response to increasingly visible populations of people with South Asian and Korean origin. To be clear, none of these categories are a race—Mexican and Korean are nationalities within the Latinx and Asian ethnic and racial groups—further examples of how the federal government itself has contributed to mischaracterizations and misunderstandings about race, national origin, and ethnicity.

Likewise, the appearance of 'Indian' and 'Hindu' on the 1940 census as race categories is almost amusing in retrospect given that neither is an accurate term for the people they are supposed to describe, and if used as they were then now would cause great confusion. In 1940 'Indian' was still inaccurately being used to describe Native Americans and 'Hindu' still being used to describe immigrants from India, Pakistan, and Bangladesh who, notably, were not all religiously Hindu but include Christians, Muslims, Buddhists, Hindus, and other religions. This conflation of race and nationality with a religion—the most visible to outsiders in India—again shows how imperfect census categories can be, especially as defined by those with racial and political power who often misunderstand enormously large and diverse ethnic and racial groups. It was not until 1950 that the Census Bureau changed 'Indian' to 'American Indian' in the census and 1980 until both distinctions among Native American groups and South Asian groups began to be disaggregated.

The census categories, of course, cannot tell us specifics regarding experiences of racialization in the United States, as questions of identity weigh heavily and uniquely on communities because of other forms of de jure and de facto policy and tradition. Even after the 2000 census allowed responders to acknowledge the very American experience of being descended from multiple groups by checking more than one box, some worried this was a blow to the power of collective identity politics even as others felt seen for the first time. Among Native Americans, for example, the possibility of checking more than one box falls within weighty debates and policies about blood quantum and tribal membership laws, 'real Indians', and federal recognition of tribal status. Two Americans who check 'American Indian' in the census may, for example, have radically different understandings of the political, social, and cultural weight of that identity depending on their phenotypical experience, the families and communities to which they formally and informally belong, and federal tribal recognition policy (Jarvis, 2017; Schmidt, 2011). Likewise, because 'Hispanic or Latino' designates *ethnicity* as opposed to race and can apply to anyone from Latin America and other countries colonized by Spain, Portugal, and France, a black-skinned Haitian American, white-skinned Chilean American, indigenous Mexican American, and Asian Filipino American might all check the category (and

one or more others), but understand these categories—and themselves—in radically different ways (Amaro & Zambrana, 2000).

As the United States approaches the 2020 census, new debates and concerns about the visibility and counting of identity have arisen. In particular, the Trump Administration has introduced a question asking respondents to specify if they are, or are not, American citizens which has raised concerns among human rights and immigrant rights groups. These groups fear that at the least the citizenship question might dissuade people from responding, leading to inaccurate counts in particular of immigrants of colour who seek to have an increased voice in American politics, and at worst might be used, as has been the case in the past, to target immigrant communities.

Visualizing Immigration and Identity

To address the exclusions discussed above and attempt to produce a visualization of available data, we endeavoured to create a project about US immigration that would explore novel expressive forms. We chose to sample from nearly 2 billion instances of microdata in order to get the finest granularity in terms of location of origin that we could, per state, displayed in decennial increments, and dating to as far back as 1790 when available. Census summary tables frequently lack all of this information, necessitating a sampling method. Furthermore, using the finest granularity is the most accurate way of extracting immigration counts and accounting for subtle differences in place of origin. In the face of inherent data problems, we explore new visualization forms, specifically tailored to the dataset and its context. Knowing the profoundly problematic nature of racial designations, we chose instead to focus on country of origin reports in the census data in order to gesture broadly at the diversity of the country and show its layers of complexity.

As mentioned, our case study explores historical immigration patterns (1830-2015), which are shown as a set of tree rings, drawing on millions of samples of US Census microdata, from a pool of nearly 2 billion individual records. As time advances, the tree grows, forming rings of immigration. Each ring corresponds to a decade. Cells are deposited in layers, and each cell corresponds to 100 immigrants.

The underlying dataset consists of samples of questionnaires from the US Census made available through IPUMS, a repository for statistical agencies that is maintained by the University of Minnesota (Ruggles et al., 2017). We queried the US state of residence, age, and place of origin of each person since 1790. (It should be noted that a large amount of territories were only

incorporated as states after 1790, meaning that data for these states were only available after a certain year.) The places of origin originally had 571 denominations. Using these data, we calculated estimates for the number of native-born persons and the number of immigrants who arrived in each decade. After reviewing the data, these places of origin were grouped into seven cultural-geographical groups: Canada, Europe, Latin America, Asia, Oceania, Africa, and the Middle East. Colours were assigned accordingly, creating a swirling spiral of various hues.

The precise evolution of the result is detailed further below, but first we present here the general pattern:

We employ visual metaphor for a variety of reasons. First, metaphor is useful to suggest other ways of thinking about the data, generating meanings that a bar chart, for example, cannot. Metaphor is also useful to embed meaning by the authors—in this case, and among others, inclusiveness. Metaphors can be used to convey figurative meanings that are recognizable and familiar, contributing to memorability (Cox, 2006); and figurative approaches allow for expressiveness and uniqueness, which contribute to stickiness (Borkin et al., 2013).

Lakoff and Johnson (1980) pioneered the view of metaphors from a cognitive perspective, framing a theory on 'conceptual metaphors', which map structural properties between a source domain and a target domain and represent a cross-domain mapping process. With this, one can understand one domain in terms of another. The metaphorical expression to convey such processes is just a linguistic expression, a surface realization of such cross-domain mapping (Lakoff, 1993; Lakoff & Johnsen, 1980; Chandler, 2017).

The meaning of the visualization is intertwined with the aesthetic qualities of the artefact, as it attempts to connote notions of wonder and to play with ideas of transformation, recurring growth, and evolution. (The study of tree rings is called 'dendrochronology', a term we have used and playfully co-opted in exhibiting the data to audiences; likewise, a studio exhibit at our home institution of Northeastern University that shows prints of the tree rings was entitled 'Naturalizing Immigration'.) The rather dry, clinical, and exacting qualities of traditional data visualization forms (Tufte, 1983)—bar charts, pie charts, line graphs—are eschewed in favour of a ludic, curiosity-evoking, and, we hope, more sublime figurative style that attempts to match thematically the country's diversity itself, while avoiding claims of finality and starkly direct quantitative comparisons among groups, whose essential nature are highly problematic (Cruz, 2015). We attempt to solve the very real problems of data integrity by, in effect, moving to a poetic and expressive level.

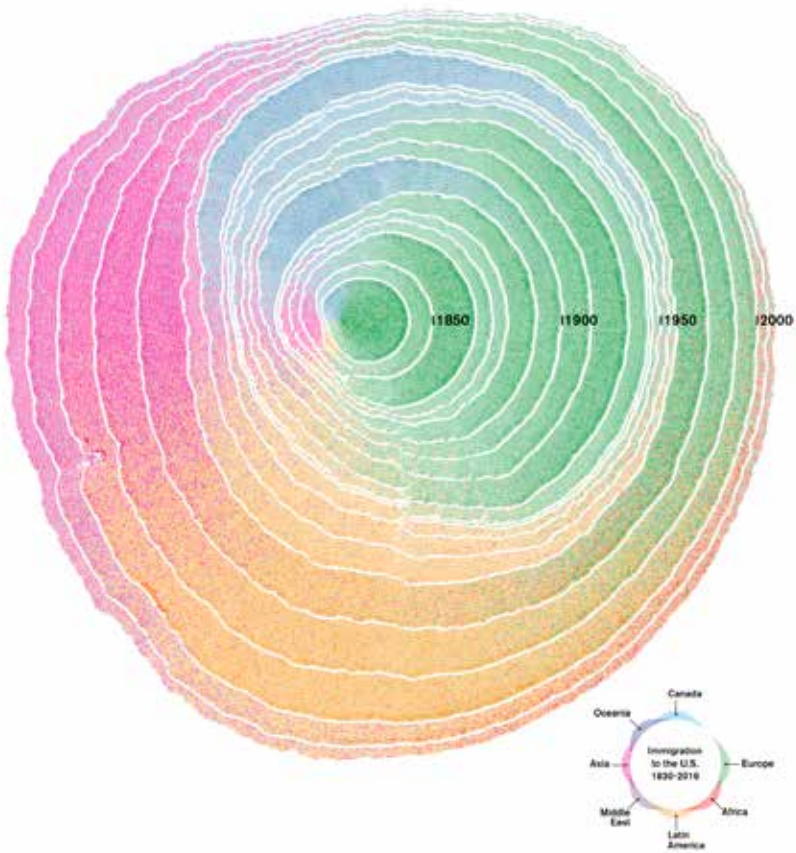

Figure 22.5. Visualization of US immigration as metaphorical rings in a growing tree trunk, with each dash representing 100 immigrants and each ring representing one decade. The image is based on Census data relating to persons' origin at birth, 1830-2015.

The video version of the visualization, some six minutes long, produces perhaps an even more powerful effect than the still images, as it allows the viewer to experience the full growth of the tree rings. That video can be found at: https://vimeo.com/276140430.

The video shows the simulation of the system: as data are injected into the visualization, new cells spawn that represent incoming immigrants in a given period in time. The specific places of origin of immigration for a certain decade are displayed as a list on the left side of the canvas, sorted by descending number of immigrants. As times passes, the tree registers every immigrant who arrived according to the dataset. One can observe the tree's state at six points in time in (Figure 22.5): 1880, 1910, 1940, 1970, 2000, and 2015.

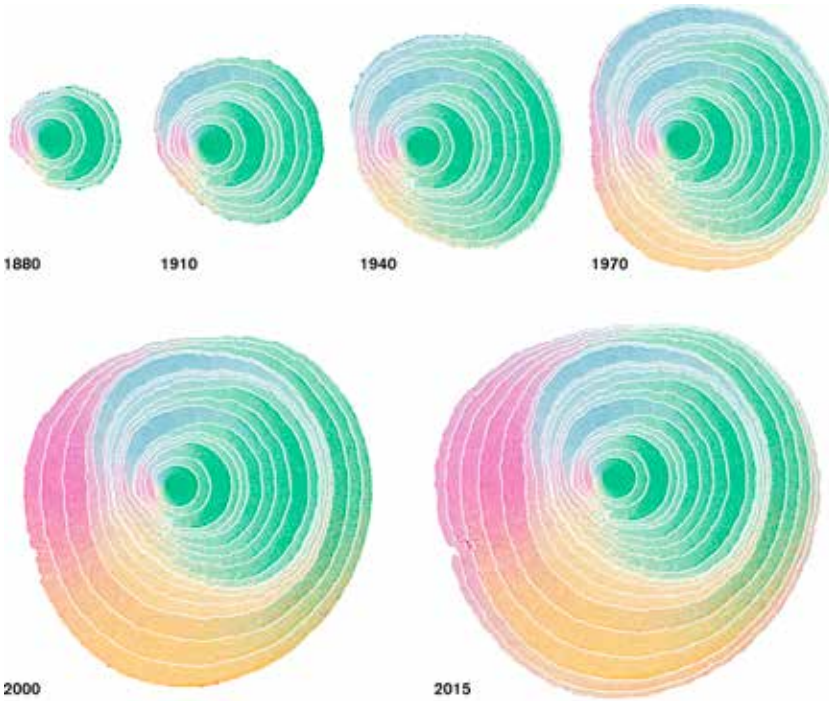

Figure 22.6. Evolution over time of visual simulation of US immigration as metaphorical rings in a growing tree trunk.

The biological metaphor that inspires this visualization was chosen for its connotations and the form of discourse it evokes and produces. Trees in their natural setting have annual growth rings that reflect varying environmental conditions; the rings' forms are neither perfect circles nor ellipses. Our algorithm is inspired by this variation and accordingly deposits immigrant cells in specific directions, depending on the geographic origin of the immigrant. Rings that are more skewed toward the country's East, for example, show more immigration from Europe, while rings skewed South show more immigration from Latin America. With this, it is possible to observe the quantity of immigration through the thickness of the rings. As mentioned, the colour of the cells corresponds to specific cultural-geographical regions, which the key and labels indicate.

Like countries, trees can be hundreds, even thousands, of years old. The cells grow slowly, and their pattern of growth influences the shape of the tree's trunk. They are all part and parcel of the organism's growth. This idea lends itself to the representation of history, as it shows a sequence of events that have left a mark and shaped the present. Just as cells leave an informational mark in the tree, so too can incoming immigrants be seen as natural contributors. Our visualization suggests that these marks of the past

are immutable and cannot be erased, regardless of how one reads them—or how one might prefer to shape the marks of the future.

Our data story uses an algorithmic 'physics engine' to simulate how cells interact with each other while creating a visualization with emergent patterns as one watches. This means that as new cells grow in the system, they are simulated as physical bodies that push and compress nearby bodies. As new cells grow, the physical constraints are solved in a certain number of steps, enabling the system to reach to near equilibrium states before advancing to the next data injection (Jakobsen, 2001; Press, Teukolsky, Vetterling, & Flannery, 2007). This causes a cascade of actions-reactions that result in the visual organization of our data, by simulating natural phenomena and obtaining a visual resemblance with organic forms (Cruz, 2017).

The cells and rings in a tree are nature's own way of organizing information. The United States is, of course, currently organized into fifty distinct states. Each state has grown at different rates and with varying immigration profiles. Some will be larger, some will be smaller, some will have complex shapes that represent waves of immigrants, and others will be perfectly circular due to the absence of immigration. Each state has its own signature and can be characterized individually. The country can therefore also be envisioned as a forest of trees, providing additional layers of complexity that tell the evolving story of American diversity. The visualization of such trees are cross-sections of their trunks that reveal the tree rings inside. In fact, when one looks at a set of tree rings, what is presented is a sample of all of the tree's cells. This dynamic can be observed in our project's context as well, in the sense that the visualized data are a sample of the universe of study.

US immigrants come from multiple geographical directions, so it makes sense that a tree can grow more in the direction where immigration is coming from. In order to do this, the seven cultural-geographical groups were attributed to specific directions (e.g. Canada → North, Europe → East, Latin America → South, and so on). With these directions, a Gaussian (normal) distribution can be created for each immigration group, with the average centred on the corresponding direction. This results in each state having its own form derived from data. Rings that are more skewed toward East, for example, show more immigration from Europe, while rings skewed South show more immigration from Latin America. Fifty sets of tree rings were simulated to show different profiles of growth and immigration for each US state (see Figure 22.7 for individual examples). An algorithm was devised in order to attain a resemblance with tree rings while carrying the semantic context that has been described. This algorithm was the result of an iterative design process that is described elsewhere (Cruz et al., 2018).

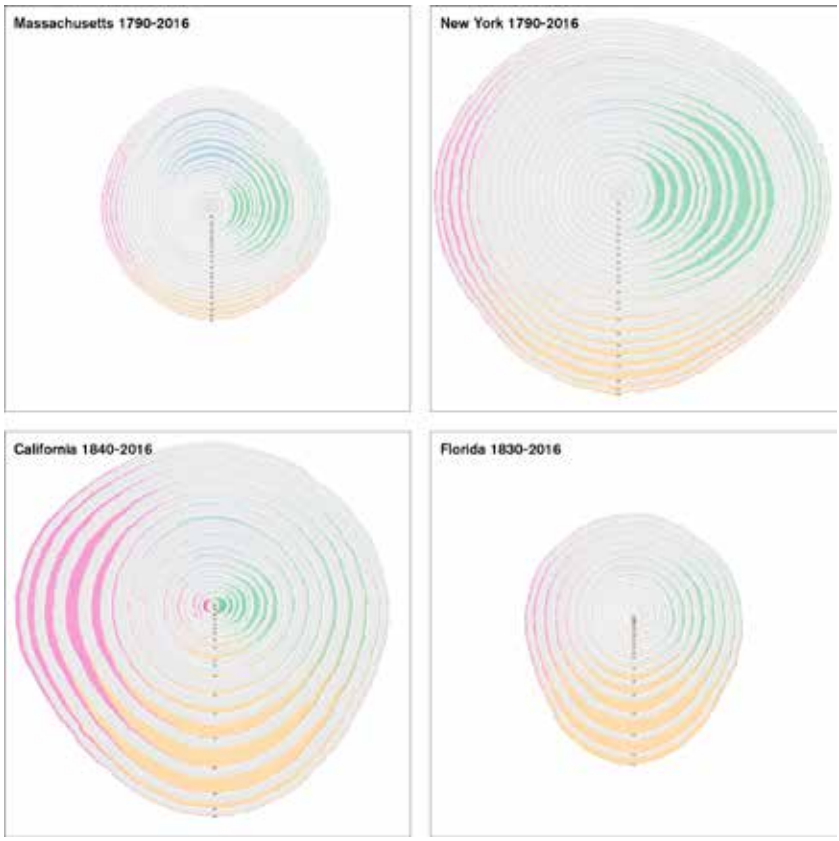

Figure 22.7. Examples of visual patterns in specific US states relating to immigration as metaphorical rings in a growing tree trunk. White cells represent native-born persons, while coloured cells represent immigrants.

Transcending literal representation

No visual form could, of course, capture the diversity of the US population, which in a very real sense mirrors the diversity of the globe itself. In this way, the project case study we present is inherently an artificial exercise. Yet, by moving from traditional conventions of precise correspondence to visual metaphor, we believe data visualization might indeed *more accurately* capture this sprawling and endlessly nuanced historical pattern and phenomenon, as compared with less artistic data forms. We use design elements to amplify and communicate messages of interrelatedness, cultural accumulation and accretion, and complex evolution; this might be contrasted with more literal visual expressions of diversity, where precise racial and origins categories might be directly compared in a more clinical and purely

scientific way. Our visualization aims to embed more meaning, and to produce both more emotion and curiosity in the viewer, than would a more standard, minimal, and sterile depiction of a dataset. The use of a physics engine algorithm that introduces indeterminacy and produces emergent phenomena—rendering simulations novel and unpredictable to some degree—is also faithful to the phenomenon of immigration itself, which is a function of a non-linear, complex global system of push and pull factors.

Through the visual metaphor employed here, and the figurative representations, interesting and important historical patterns can be discerned. Through the visualization, the viewer may note that the origins of US immigrant populations transform from era to era. In the 1840s and 1880s, European immigrants came mainly from northern and western Europe, whereas the famous influx of the early 1900s, symbolized by Ellis Island's gateway, emanated mostly from southern and eastern Europe. Immigration from Asia rose between 1970 and 2000, while large-scale immigration from Latin America began in 1950 and lasted for half a century. Immigration from Africa only becomes visible in the 21st century.

As discussed, no data picture of diversity in the United States can fully account for the lack of data for certain marginalized groups. In addition, the categories in each successive decade after the country's founding may have gotten more inclusive and precise over time, but they still often suffer from certain historical biases. Undercounting was inevitable, and still is, with regard to newly arrived, and thus highly transient and vulnerable, populations.

There can be no neutral rendering in this domain, no purely objective point of view, and thus no representational act that avoids questions of exclusivity. Knowing this, we choose through the creation of this case study to accept the burden of fallibility in the service of trying to convey a higher set of insights. Imperfect and tragically flawed in its origins as well as its current history, the country is becoming more diverse ethnically and racially each year now, and the percentage of the population made up of foreign born persons is approaching historic levels (Zong & Batalova, 2017), even as a policy backlash and anti-immigration sentiment continue to simmer.

The counting of populations and the rendering of pictures based on those data are an inherently political act. Controversies continue to grow over how resources will be used in future census counts; there remain grave concerns that persons of colour and marginalized groups, in particular, will not be sufficiently represented in official statistics (Chevat & Lowenthal, 2015). Data visualization that is rigorously rooted from a computational and statistical perspective, while at the same time innovative in generating

ideational and compositional meanings, can help transcend limitations of administrative data and produce new discourses about diversity and its importance in society.

References

Amaro, H., & Zambrana, R. E. (2000). Criollo, mestizo, mulato, LatiNegro, indígena, white, or black? The US Hispanic/Latino population and multiple responses in the 2000 census. *American Journal of Public Health*, *90*(11), 1724-1727.

Aranti, L. (2018, April 6). Secret use of Census info helped send Japanese Americans to internment camps in WWII. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/retropolis/wp/2018/04/03/secret-use-of-census-info-helped-send-japanese-americans-to-internment-camps-in-wwii/

Armstrong Dunbar, E. (2008). *A fragile freedom: African American women and emancipation in the antebellum city.* New Haven: Yale University Press.

Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2306-2315. http://doi.org/10.1109/TVCG.2013.234

Chandler, D. (2017). *Semiotics: The basics*. New York: Taylor & Francis.

Chang, A. (2017, August 3). *Watch how immigration in America has changed in the last 200 years*. Vox. Retrieved from https://www.vox.com/2016/1/4/10709366/immigration-america-200-years

Chevat, B., & Lowenthal, T. A. (2015, March 3). Meeting the challenges of the 2020 Census. *Center for American Progress*. Retrieved from https://www.americanprogress.org/issues/race/reports/2015/03/03/106224/meeting-the-challenges-of-the-2020-census/

Cohn, D. (2010, March 3). Census history: Counting Hispanics. *Pew Research Center*. Retrieved from http://www.pewsocialtrends.org/2010/03/03/census-history-counting-hispanics-2/

Cox, D. (2006). Metaphoric mappings: The art of visualization. In: P. A. Fishwick (Ed.), *Aesthetic Computing*. (pp. 89-114). Cambridge: MIT Press.

Cruz, P. (2015). Wrongfully right: Applications of semantic figurative metaphors in information visualization. *Proceedings of IEEE VIS 2015 Arts Program, VISAP'15: Data Improvisations*, Chicago, Illinois, 14-21.

Cruz, P. (2017). Adapted Dorling cartogram on wage inequality in Portugal. *Proceedings of the IEEE VIS Arts Program* (*VISAP*), Phoenix, Arizona.

Cruz, P.M., Wihbey, J., Ghael, A., Costa, S., Chao, R., & Shibuya, F. (2018). Process of simulating tree rings for immigration in the US. *IEEE VIS Arts Program Annotated Projects*.

Daniels, R. (1989). *History of Indian immigration to the United States: An interpretive essay*. New York: Asia Society.

Du Bois, W. E. B. (1900). *The Georgia Negro: A social study* [Map]. Library of Congress Prints and Photographs Division Washington, D.C.

Gibson, C., & Jung, K. (2006). *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000—U.S. Census Bureau Working Paper No. 81*. US Census Bureau. Retrieved October 11, 2018 from https://www.census.gov/population/www/documentation/twps0081/twps0081.html

Handlin, O. (1959). *Immigration as a factor in American history*. Englewood Cliffs, NJ: Prentice-Hall.

Harding, S. (1993). *The 'racial' economy of science: Toward a democratic future*. Bloomington: Indiana University Press.

Hart, P. (2009, February 5). Census Bureau tracks changing face of US University Times. Retrieved from https://www.utimes.pitt.edu/archives/?p=8642

Horton, L. E. (1985). Community organization and social activism: Black Boston and the antislavery movement. *Sociological Inquiry, 55*(2), 182-199. http://doi.org/10.1111/j.1475-682x.1985.tb00858.x

Ignatiev, N. (2012). *How the Irish became white*. New York: Routledge.

IPUMS USA. (n.d.). 1890 Census: Instructions to enumerators. Retrieved June 1, 2018 from https://usa.ipums.org/usa/voliii/inst1890.shtml

Jakobsen, T. (2001). Advanced character physics. In: Proceedings of Game Developers Conference, San Francisco, pp. 383-401.

Jarvis, B. (2017, January 18). Who decides who counts as Native American? *The New York Times Magazine*. Retrieved from https://www.nytimes.com/2017/01/18/magazine/who-decides-who-counts-as-native-american.html

Kertzer, D. I., & Arel, D. (2002). *Census and identity: The politics of race, ethnicity, and language in national censuses*. Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: The University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In: A. Ortony (ed.), *Metaphor and thought*. (pp. 202-251). Cambridge: Cambridge University Press.

Lakota, H. (1902). *Winter Count on cloth by Long Soldier* [Painting]. Smithsonian, National Museum of the American Indian. Retrieved October 1, 2018 from https://americanindian.si.edu/exhibitions/horsenation/wintercount.html

Portes, A., & Rumbaut, R. G. (1990). *Immigrant America: A portrait*. Berkeley and Los Angeles, CA: University of California Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes* (3rd ed.). Cambridge: Cambridge University Press.

Nicholson, F. (1724/1900). *Map of the several nations of Indians to the northwest of South Carolina* [Map]. Library of Congress. Retrieved from http://hdl.loc.gov/loc.gmd/g3860.ct000734

Nugent, W. T. K. (1995). *Crossings: The great transatlantic migrations 1870-1914*. Bloomington: Indiana University Press.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., & Sobek, M. (2017). *Integrated public use microdata series: Version 7.0* [Data Set]. Minneapolis: University of Minnesota. https://doi.org/10.18128/D010.V7.0.

Schmidt, R. W. (2011). American Indian identity and blood quantum in the 21st century: A critical review. *Journal of Anthropology,* 2011. http://dx.doi.org/10.1155/2011/549521

Steinmetz, K. (2018, March 27). The debate over a new citizenship question isn't the first Census fight. Here's why the count is controversial. *Time*. Retrieved from http://time.com/5217151/census-questions-citizenship-controversy/

Takaki, R. T. (2012). *Strangers from a different shore a history of Asian Americans* (Updated and Revised). Boston, New York, & London: Little, Brown and Company.

Truesdell, L. E. (1930). 1930 Census: Volume 6. [Supplement] Special report on foreign-born white families by country of birth of head with an appendix giving statistics for Mexican, Indian, Chinese, and Japanese families. US Census Bureau. Retrieved from https://catalog.hathitrust.org/Record/001310758

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

US Census Bureau. (n.d.). *Measuring Race and Ethnicity Across the Decades: 1790-2010*. Retrieved June 15, 2018 from https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html

US Census Bureau. (n.d.). *Statistical abstracts of the United States*. Retrieved August 15, 2018 from https://www.census.gov/library/publications/time-series/statistical_abstracts.html

US Census Bureau. (1896). *Statistical abstract of the United States 1897—Part 2*. Retrieved August 15, 2018 from https://www.census.gov/library/publications/1898/compendia/statab/20ed.html

US Census Bureau. (1983). *Statistical abstract of the United States: 1984—Section 1 Population*. Retrieved August 15, 2018 from https://www.census.gov/library/publications/1983/compendia/statab/104ed.html

Zinn, H. (2015). *A people's history of the United States: 1492-Present* (3rd ed.). London: Routledge.

Zong, J., & Batalova, J. (2017, March 8). Frequently requested statistics on immigrants and immigration in the United States. *Migration Policy Institute*. Retrieved from https://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states-6

About the authors

John Wihbey is an Assistant Professor of journalism and media innovation at Northeastern University in Boston, where he heads the graduate programmes in the School of Journalism. He is a faculty director of the Co-Laboratory for Data Impact, based at Northeastern's College of Arts, Media and Design.

Pedro Cruz is an Assistant Professor in the Department of Art + Design at Northeastern University. He is a faculty director of the Co-Laboratory for Data Impact.

Sarah J. Jackson is an Associate Professor of Communication Studies at Northeastern University, where she is also a faculty affiliate of the Department of Cultures, Societies and Global Studies.

Brooke Foucault Welles is an Associate Professor of Communication Studies at Northeastern University and a core faculty member at the Network Science Institute.