CHANCE

ISSN: 0933-2480 (Print) 1867-2280 (Online) Journal homepage: http://www.tandfonline.com/loi/ucha20

Interview with a Centennial Chart

Heike Hofmann

To cite this article: Heike Hofmann (2007) Interview with a Centennial Chart, CHANCE, 20:2, 26-35, DOI: 10.1080/09332480.2007.10722843

To link to this article: http://dx.doi.org/10.1080/09332480.2007.10722843

Interview with a

Heike Hofmann

Informative and aesthetically pleasing graphics have a long history. Here is an example from the 1870 census.

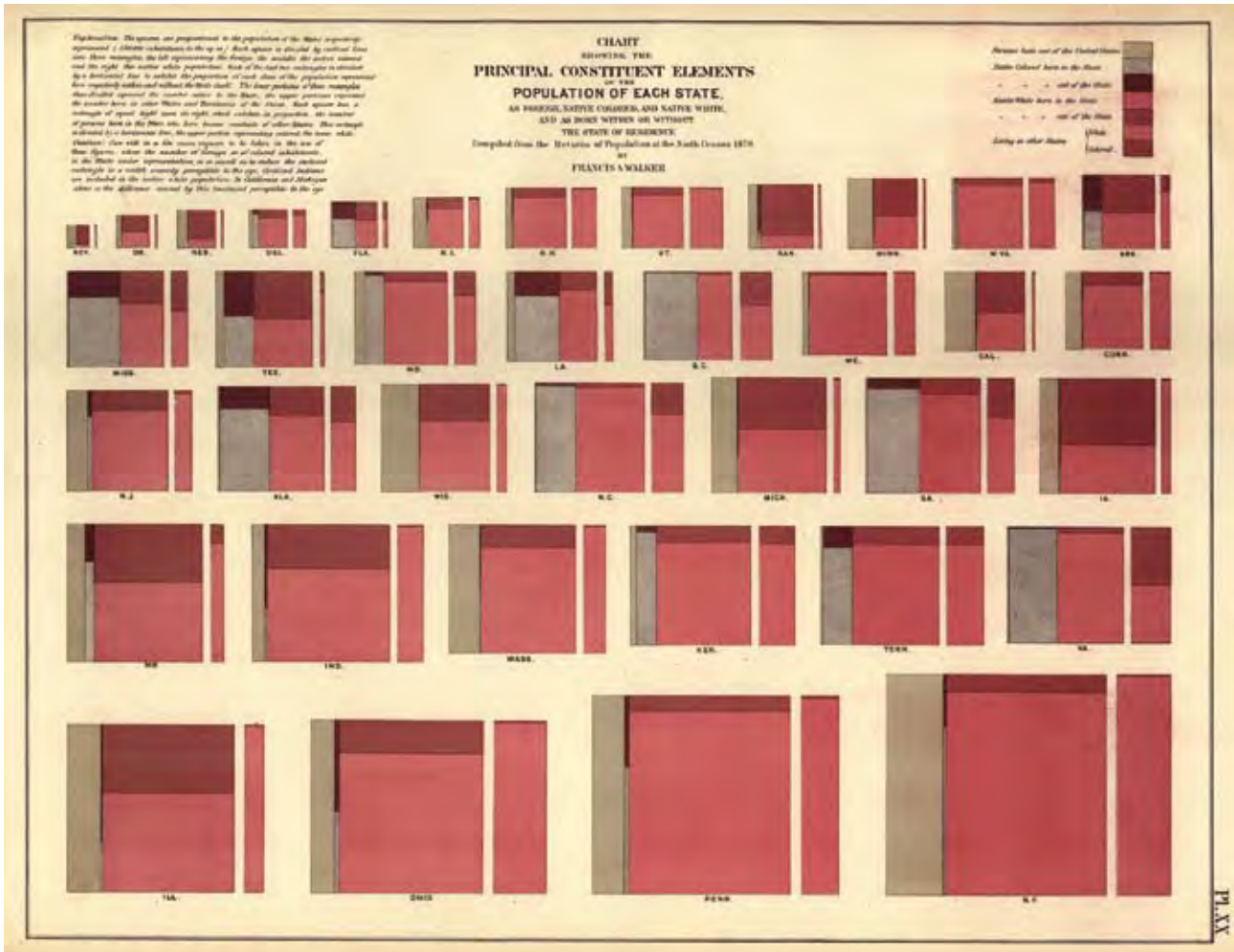


Figure 1. "Beautiful Mosaic," by Francis A. Walker and published in *Statistical Atlas of the Ninth U.S. Census of 1870*, shows total population in each state classified according to nationality, race, and place of birth.

Centennial Chart

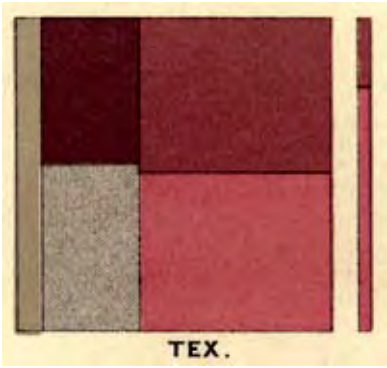


Figure 2. A close-up of the mosaic of Texas. The underlying numbers are shown in Table 1.

Table 1— Numbers for Texas Based on a 1% Sample of the 1870 Census

# Foreign-Born			616
# Native	Black	Born in-state	1071
		Born out-of-state	1301
	White	Born in-state	3190
		Born out-of-state	2581
# Native	Black	Living out-of-state	86
# Native	White	Living out-of-state	184

A graphic is worth a thousand words. Alas, past experience with mosaic plots has shown it might need a couple hundred words of description before the chart starts talking on its own. A rough first count of words shows that Francis A. Walker needed approximately 136 words (see Figure 3) for his beautiful example of a mosaic plot published in the *Statistical Atlas* of 1874. A scaled-down copy of the original chart is displayed in Figure 1.

Mosaic plots have been known in the statistical community since the 1980s, when John Hartigan and Beat Kleiner introduced them as a way of visualizing multiway contingency tables. Earlier displays that we would classify as mosaic plots today do exist, however. Here, we are discussing a chart from the *Statistical Atlas* of the 9th U.S. Census, published in 1874.

The chart consists of a series of mosaic plots, one for each U.S. state existent in 1870. The overall size of each mosaic represents the total population of the state. The states are ordered according to size in a "snake-like" design (i.e., the 12 smallest states are in the top row, smallest to largest from left to right). The next-smallest state, Connecticut, is drawn on the very right in the second row just below the largest state of the top row. The next-largest states are then drawn from right to left in the second row, and, again, from left to right in the third row, etc. Note that, this way, the rows have different heights. Walker was even able to move rows two and three into each other by aligning all mosaic plots of row two along their top lines and mosaics of row three along their bottom lines, which helps the layout to make parsimonious use of space.

Each mosaic plot is color-coded and contains the following information, from left to right (see Figure 2 for a close-up of one of the states):

a) Number of "foreign born" people living in the state (brown bar)

b) Number of "native colored" people, classified according to whether an individual was born in the state (grey area) or outside the state (dark red area on top). Walker, here, uses "colored" to refer to black and mulatto people in the same way the word is used throughout the first 12 U.S. census surveys.

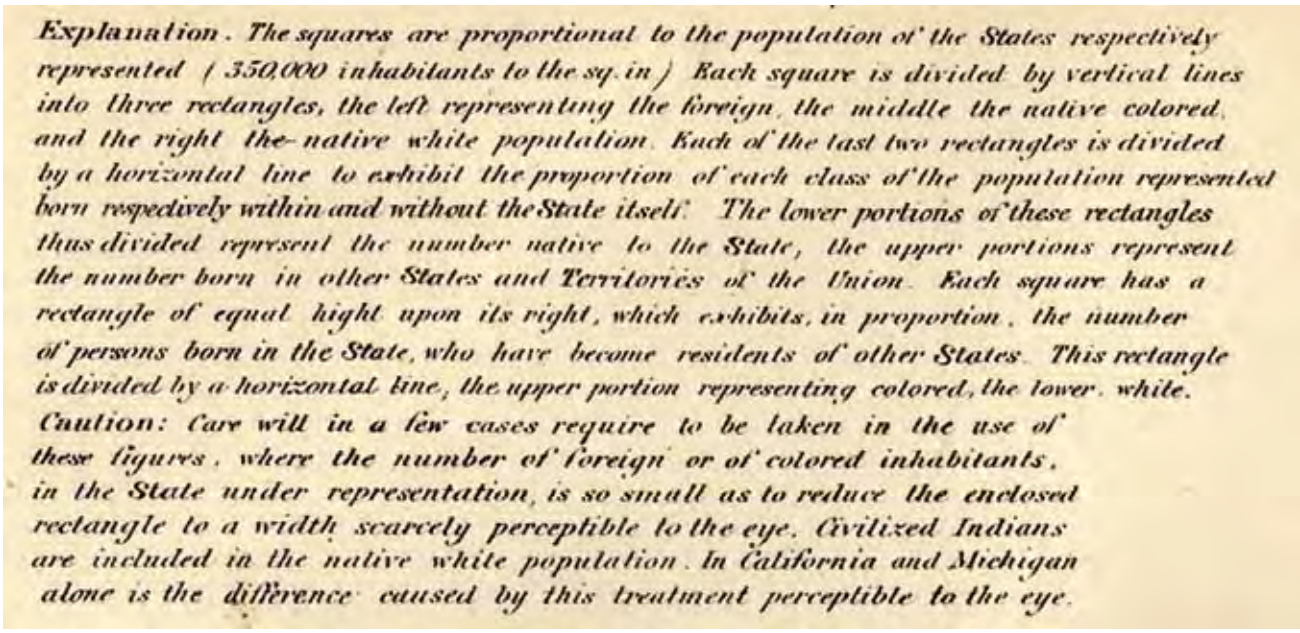

Explanation. The squares are proportional to the population of the States respectively represented (350,000 inhabitants to the sq. in.) Each square is divided by vertical lines into three rectangles, the left representing the foreign, the middle the native colored, and the right the native white population. Each of the last two rectangles is divided by a horizontal line to exhibit the proportion of each class of the population represented born respectively within and without the State itself. The lower portions of these rectangles thus divided represent the number native to the State, the upper portions represent the number born in other States and Territories of the Union. Each square has a rectangle of equal hight upon its right, which exhibits, in proportion, the number of persons born in the State, who have become residents of other States. This rectangle is divided by a horizontal line, the upper portion representing colored, the lower, white. Caution: Care will in a few cases require to be taken in the use of these figures, where the number of foreign or of colored inhabitants, in the State under representation, is so small as to reduce the enclosed rectangle to a width scarcely perceptible to the eye. Civilized Indians are included in the native white population. In California and Michigan alone is the difference caused by this treatment perceptible to the eye.

Figure 3. Francis A. Walker's original description of "Beautiful Mosaic."

c) Number of "native white" people, also classified according to being born in the state (pink) or outside (dark pink). "Native" is used to refer to people born in a U.S. state or territory.

d) Number of people born in the state, but living outside it, with white people represented in the pink area and colored people represented by the dark pink area on top.

Adding the values in (a) to (c) gives the total population in each state. The mosaic, itself, is drawn by starting from a square representing this total population and cutting it into three vertical slices according to the numbers of foreigners, blacks, and whites. Figure 3 shows Walker's original description of the chart. Using a square to represent the size of the state's population is a crucial step in making this chart work: States' populations range from only 42,491 in Nevada to more than 4 million in New York. Walker implicitly applies a square root transformation to population size to make heights of mosaic plots easier to compare (heights, rather than areas, as ocular estimates of area comparisons are notoriously wrong).

The bin on the right of each state's square corresponds to the white and black populations born, but not living, in the state in 1870. This has both pros and cons. A con is that the color-coding is a bit misleading: The colors of this bin are repeating the colors of the bin corresponding to the native white population. In this regard, it might make more sense to place this rectangle underneath the state's square with the part corresponding to the white population on the right and the part corresponding to the black population on the left. This would lead to a less efficient use of space, however. One advantage of placing the rectangle next to the state's square is that the height of the bin is naturally adjusted. By using the same height, we can interpret the width of the bin with respect to the overall width of the square, or, possibly easier, interpret the area of the rectangle as a proportion of the area of the square. This lets the width represent the rate of immigrants in proportion to the total population of the state.

What Does It Say?

Each individual chart already contains a lot of information, but the true beauty of the chart comes from the overview it provides.

Either Foreign or Black. By searching for grey and brown bars in Figure 1, we see that large bars in these colors do not occur together in a single mosaic plot, leaving us with the conclusion that states have either a large foreign-born population or a large black population. This can be seen in more detail in Figure 4. The maps of Figure 5 show choropleth maps; states are shaded between light yellow and dark green according to their percentage of black population (top map) and foreign-born population (bottom map). Both maps show strong spatial patterns. Not surprisingly, the former slave states (slavery was abolished in 1865) of the southeast show the highest percentages of black population.

Related to the question of the abolishment of slavery in 1865, we might be asking whether race has an influence on the choice of state when moving into a state. From the chart in Figure 1, we can check the difference in heights of the dark-colored areas of bins two and three. Whenever the black population is small, there seems to be a higher rate of blacks moving into the state. This finding remains inconclusive, though.

In-State or Out-of-State? Figure 6 shows a scatterplot of the percentage of white people from out-of-state versus the

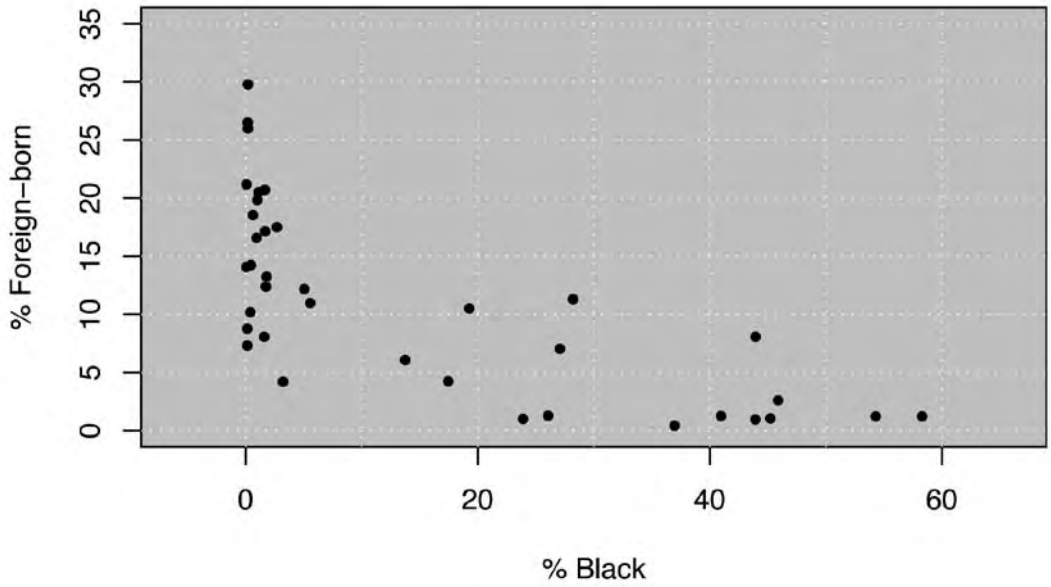


Figure 4. Scatterplot of the rates of the foreign-born people versus the black population in each state. States with a higher rate of foreign-born people tend to have a smaller rate of black population.

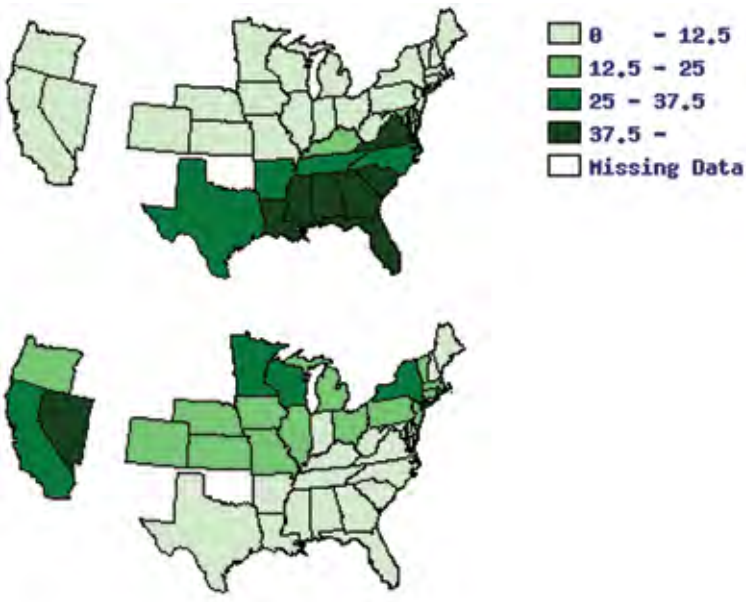


Figure 5. Choropleth maps of the rates of foreign-born people (bottom) and the black population (top) by state. States with high percentages of black population are in the southeast, while states with a high immigration rate are in the north and west.

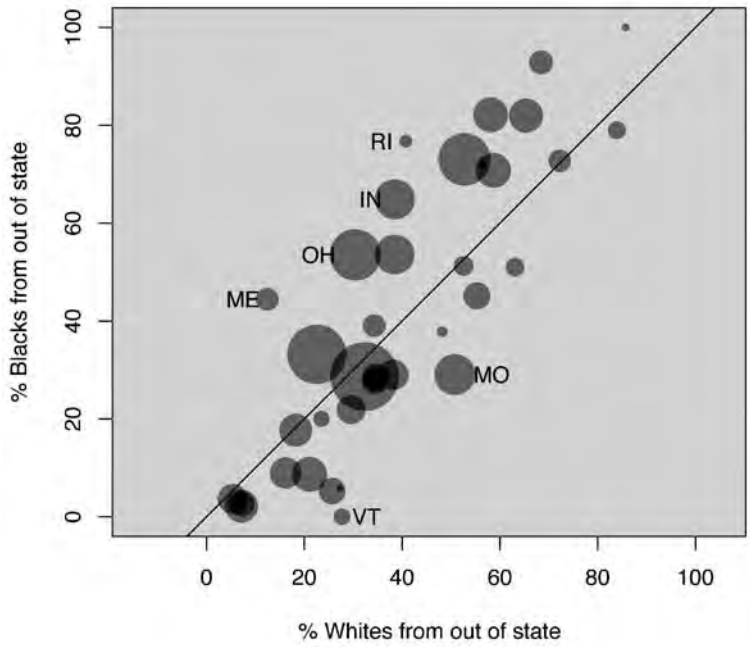


Figure 6. Scatterplot of the percentages of white and black people from outside the state. The size of a point reflects the overall population of a state. Percentages show a linear relationship.

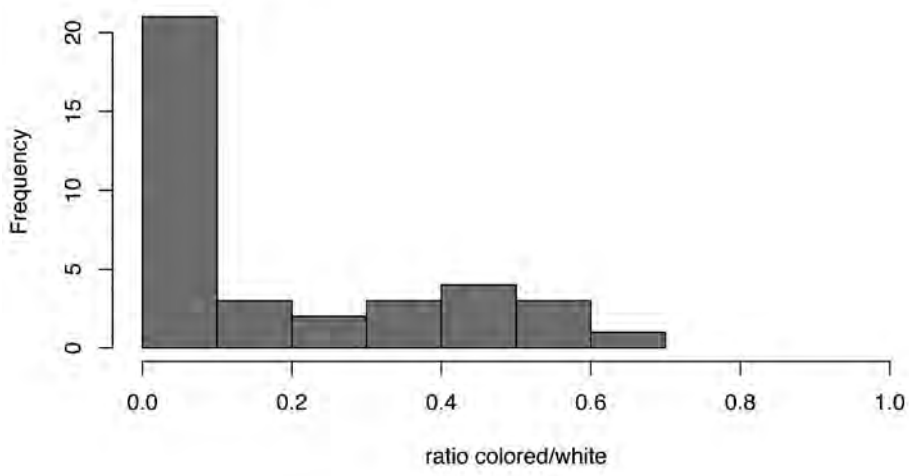


Figure 7. Histogram of the ration of black versus white people living outside the state in which they were born.

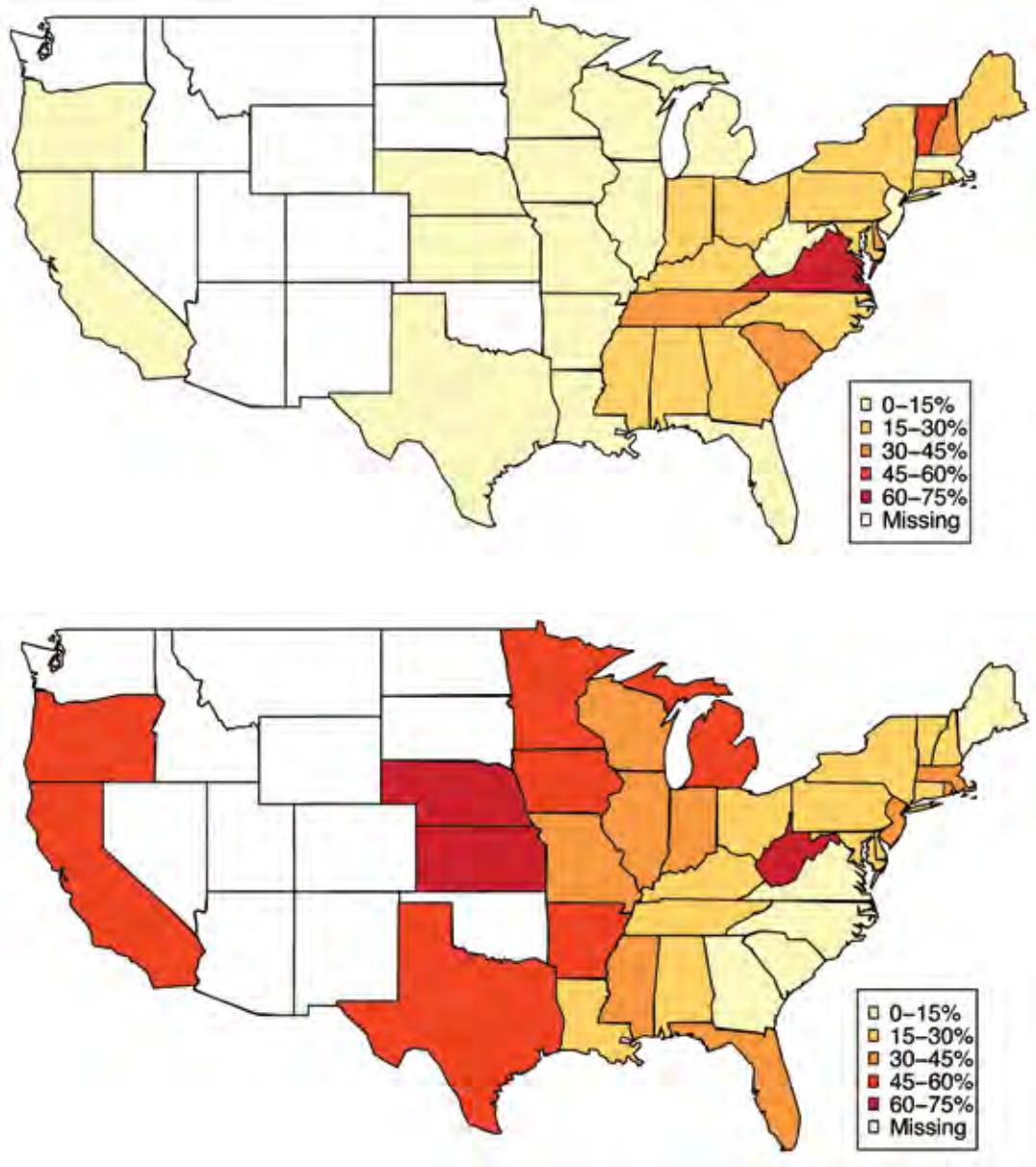


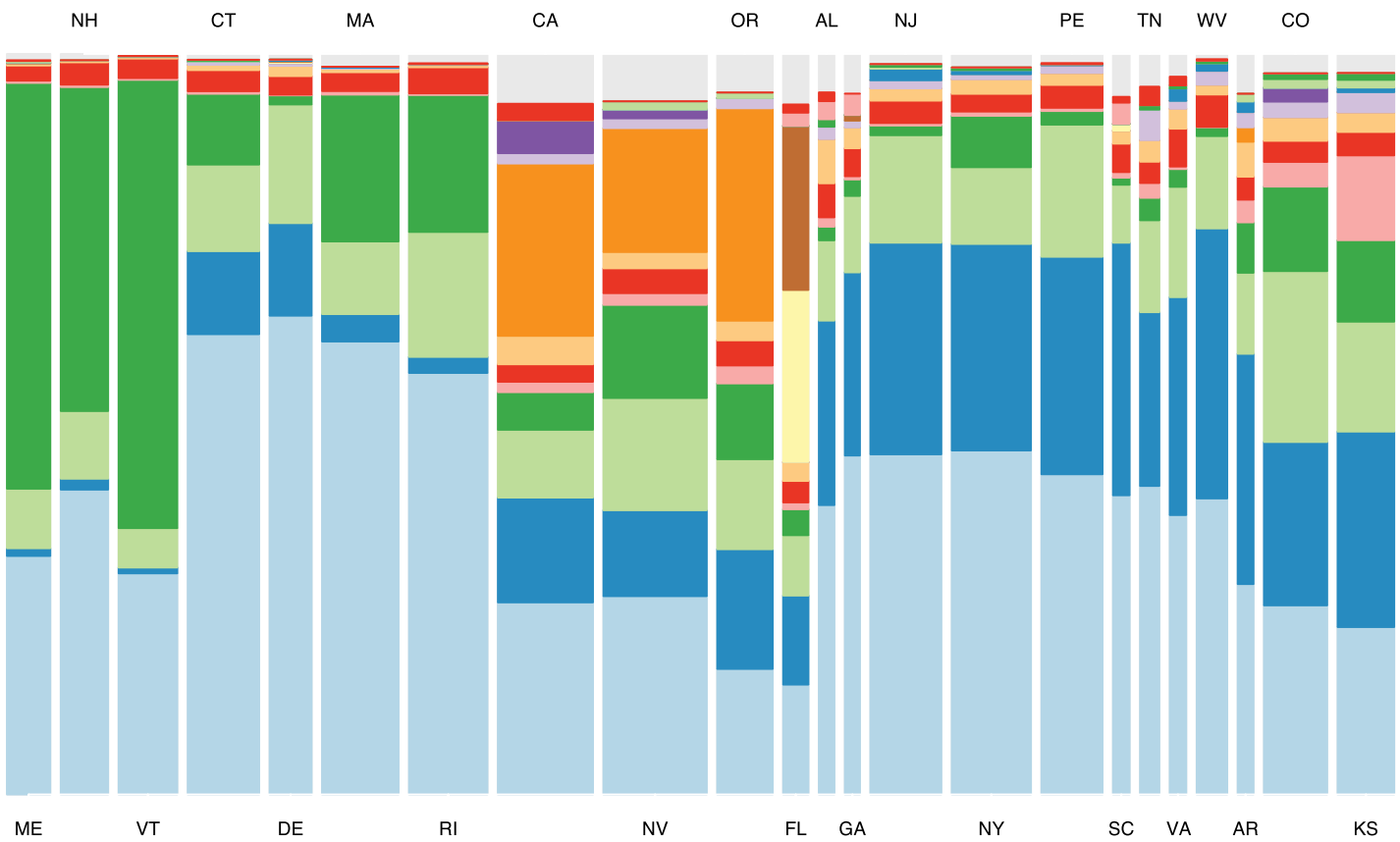
Figure 8. Choropleth maps of the United States in 1870. Each state is colored according to the rate of people born in the state but living outside (top) and the rate of people being born outside the state in which they now live (bottom).

percentage of black people from out-of-state for each state. The size of a dot corresponds to a state's overall population. The relationship between these percentages is linear, with a slope close to one, suggesting that out-of-state rates do not depend on race.

Moving on. Comparing the ratio of black to white people who live outside the state in which they were born (rightmost bin in each of the mosaics), we see that the numbers for black people are lower than the numbers for white people. This suggests the black population moves less in all states. A summary is given in the histogram in Figure 7.

States Going out of Fashion? The overall numbers of people living outside suggest the "older" states on the East Coast have higher rates of people who are born in the state, but no longer live there, than the younger states farther to the west (i.e., the eastern states are "producing" population for the western states). This is mirrored in the higher rates of people living in a western state in 1870 who were not born there. A

Figure 9. A modern mosaic plot of state by place of origin. The width of each bin corresponds to the size of the state's foreign-born population. Bins are ordered according to a combination of results from clustering and a principal component analysis to place states with similar makeups of foreign-born populations close to each other. The legend shows the marginal distribution of immigrants by birthplace.



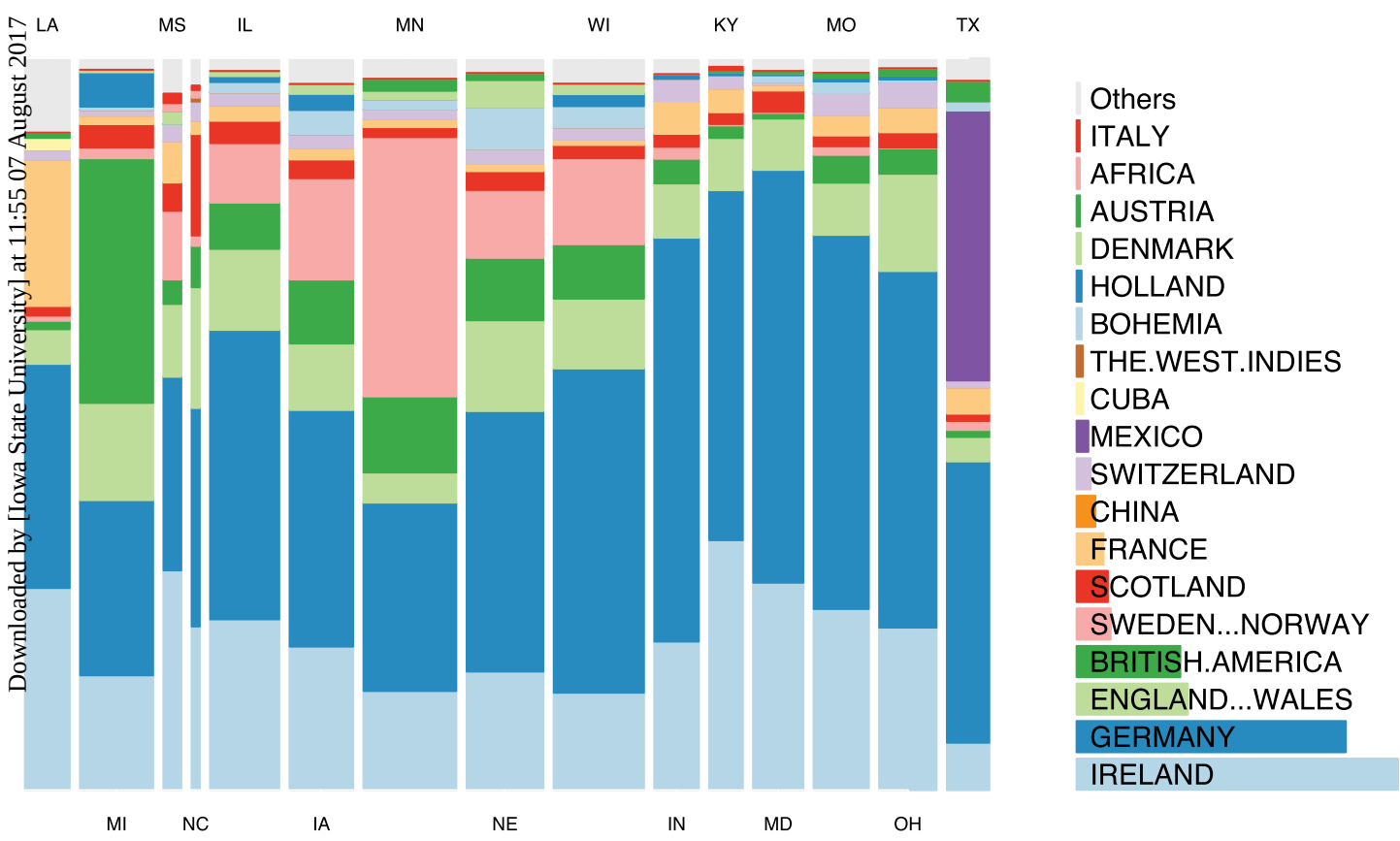


Figure 10. A map of the United States of 1870. Colors give the rates of the Chinese population among foreign-born people in each state. Note that the color scale is not continuous, but focuses on the lower and upper ends of the data scale.

summary is shown in Figure 8. The upper map shows rates of "moving out" of a state; the lower map shows rates of "moving into" a state. No doubt about it, the possibility of free land was a strong economic incentive to move into the brand new states. The curious outlier of West Virginia—a state in the eastern part of the United States with high "moving in" and low "moving out" rates—turns out to be more of a spatial outlier in terms of its history: Even though West Virginia is in the east, it only became a state in 1863 by splitting off from Virginia, which explains both the 'loss' of population in Virginia and the 'gain' in West Virginia.

What Does It Not Talk About?

Like any grumpy old man, there are things the chart does not want to talk about. This might or might not have a political agenda; we will leave that judgment to the reader.

The overall layout of the graphic ensures the available space is used as efficiently as possible. The information coded by this layout is limited, however. The lack of spatial information is particularly obvious. While the Statistical Atlas contains a lot of maps, none of them are choropleth maps on a state level. In that way, state-level summaries are kept separately from spatial information.

On a different note, it is curious how Walker decided to not include information he clearly had access to and must have evaluated in the process of drawing the chart. As Walker explains, the Indian population has been ignored by accumulating "civilized" Indians to the white population.

By "civilized" Indians, Walker most likely meant members of one of the five Native-American nations—Cherokee, Chickasaw, Choctaw, Creek, and Seminole—considered civilized by white society because they had adopted many of the colonists' customs and generally had good relations with their neighbors.

The considerable Chinese population is completely disregarded by him in both word and picture. This information would have been easy enough to include in the existing chart by cutting the bin corresponding to foreign-born population according to places of origin. Of the foreign-born population, Chinese make up almost a third of the population in California, more than half in Idaho (which did not become a state until 1890), and more than 20% in both Nevada and Oregon. On a map, these states can be identified easily as the western part of the United States (see Figure 10). The modern mosaic plot in Figure 9 gives an overview of (self-reported) birthplaces of foreign-born inhabitants in each state. The size of each bin corresponds to the number of foreign-born people compared to the state's total population. Places of origin are sorted according to their overall contribution of people to the United States of 1870. For practical purposes, only the 18 places with the highest immigration rates to the United States are considered. The remaining nine are summarized in the category "others." The highest overall percentage of immigrants from these birthplaces is less than 0.1%. Percentages below 1% are almost impossible to detect in the chart.

The overall largest number of people originates from Ireland, followed by Germany and the UK.

The order in which states are drawn is given by a combination of the results from a cluster analysis and a principal component analysis. First, states are clustered according to the makeup of their foreign population. This yields nine obvious clusters of states. These clusters are then sorted according to their average value on the first principal component of the same data, which basically sorts clusters according to the percentages of Irish and German immigrants. As can be seen in Figure 9, the rates of Irish immigrants tend to decrease from left to right, while rates of German immigrants tend to increase.

The bars exhibit color patterns according to the composition of immigrants in a state. Some compositions are more frequent than others (e.g., New Hampshire, Maine, and Vermont show a similar immigrant makeup, with large numbers of Irish and British Americans and only a few Germans and English). The bars for Nevada, California, and Colorado also show similar color patterns, predominantly, as discussed, because of their large Chinese population.

Florida and Texas show up with unique compositions of their foreign-born population. Those in the second-largest group of foreign-born people in Texas are Mexican. With 36.9% of all immigrants in Texas being Mexican, this is by

Table 2— Clusters of States with Similar Makeups in Birthplaces of Their Foreign-Born Population

Cluster	States	Distinguishing Feature
1	ME, NH, VT	High rate of British Americans
2	CT, DE, MA, RI	High rate of Irish
3	CA, NV, OR	High rate of Chinese
4	FL	Immigrants from Cuba and the West Indies
5	AL, GA, NJ, NY, PA, SC, TN, VA, WV	High rates of Germans and Irish
6	AR, CO, KS, LA, MI, MS, NC	"Left over" group—large diversity between group members: NC has a large Scottish population; LA has a large French population; MI has a high rate of Dutch immigrants
7	IL, IA, MN, NE, WI	High rates of Norwegians, Swedes, and Bohemians
8	IN, KY, MD, MO, OH	Highest rates of Germans, high rates of French and Swiss
9	TX	High rate of Mexicans

far the highest rate of Mexicans in the United States in 1870. The next-highest rates are 4.4% in California and 1.9% in Colorado. This is a direct result of the Mexican-American war, which ended in 1848 with the annexation of the former Mexican states of Texas, New Mexico, and parts of Colorado and California. Florida is the only state to have significant numbers of immigrants from Cuba and the West Indies, which make up almost half (45.3%).

Swedes and Norwegians have highest rates of immigration in what are now the midwestern states, among them Minnesota, Iowa, Wisconsin, and Illinois. The rate of Scottish immigrants is at a fairly stable rate of 3%, except for a soaring 13.9% in North Carolina.

The Art of Graphics

Most of these results might not come as a big surprise to anybody, but it is good to see them emerge from glancing at a single chart that is, admittedly, a bit on the talkative side. The art of great graphics does not stop with Charles Joseph Minard's "Napoleon's March to Moscow." Beautiful charts do exist in other places, but we need to take time to read them carefully. Only this will lead to fulfilling the U.S. Census and this chart's purpose, which is to keep history alive by talking about it. ◘

Further Reading

Walker, F.A. (1874). *Statistical Atlas of the United States Based on the Results of the Ninth Census 1870.* United States: Census Office.

Friendly, M. (2002). "A Brief History of the Mosaic Display." *Journal of Computational and Graphical Statistics*, 11, 89–107.

Minnesota Population Center, University of Minnesota, *www.ipums.org/usa/index.html.*

Historical Census Browser, Geospatial and Statistical Data Center, University of Virginia, *http://fisher.lib.virginia.edu/collections/stats/histcensus.*