ARTICLE

Scaling the Interactive Dot Map

Kyle E. Walker
Department of Geography / Texas Christian University / Fort Worth / TX / USA

ABSTRACT

Dot maps are effective for cartographic visualization of categorical data. Recent advances in Web mapping technology have facilitated the development of interactive dot maps, in which users can pan and zoom to view data distributions for different areas. This interactivity, however, introduces multiple cartographic challenges, as design decisions that are appropriate at large scales can lead to clutter and illegibility at small scales. This article considers these challenges in the context of an applied example – an interactive dot map of educational attainment in the United States. It covers the methodology of the map's creation as well as how it addresses the cartographic challenges of interactive dot mapping.

Keywords: interface design,demographics,interactive maps,visualization,dot maps

RÉSUMÉ

Les cartes de points sont un instrument efficace de visualisation cartographique de données catégorielles. Les progrès récents de la technologie cartographique Web ont facilité l'élaboration de cartes de points interactives permettant aux utilisateurs de faire un panoramique et un zoom afin de voir les distributions de données pour différents secteurs. Cette interactivité soulève toutefois maintes difficultés cartographiques, car les décisions conceptuelles qui sont appropriées aux grandes échelles peuvent mener à la confusion et à l'illisibilité à petite échelle. L'auteur se penche sur ces difficultés dans le contexte d'un exemple de cartographie appliquée : celui d'une carte de points interactive du niveau de scolarité aux États-Unis. Il s'intéresse à la méthodologie de création de la carte ainsi qu'à la façon dont elle résout les difficultés cartographiques de l'élaboration de cartes de points interactives.

Mots clés : cartes de points,cartes interactives,conception d'interface,démographie,visualisation

Introduction

Cartographers have long used *dot maps* as a technique for visualizing the geographic distribution of either univariate or multivariate count data. To construct a dot map, the cartographer must make a series of decisions to facilitate the map's interpretability: selection of dot size and colour; selecting the dot value, which refers to the number of units represented by each dot; and choosing an algorithm to place dots on the map to best represent the geographical distribution of the data (Kimerling 2009). In turn, cartographers have developed a series of methods and algorithms to provide optimal solutions to these decisions (Kimerling 2009; Hey and Bill 2014; Chua and Vande Moere 2017).

In recent years, advances in Web mapping technology have assisted the development of interactive dot maps that allow panning and zooming functionality. For the United States, developers often design these maps to cover the entire country, addressing topics such as race and ethnicity (Cable 2013b); immigration (Walker 2015); electoral politics (Cable 2013a; Field 2013); employment (Manduca 2015); and housing (Schwenke 2015). While such national dot-density maps are richly informative and of interest to a wide audience, they present a new set of visualization challenges. Whereas a static dot-density map must be designed to maximize comprehension at a specific view and zoom level, the Web cartographer must design a styling specification that is suitable for millions of potential views, ranging from sparsely populated rural areas to high-density cities.

This article covers the creation of such an interactive dot-density map: Educational Attainment in America, viewable at http://personal.tcu.edu/kylewalker/maps/education. The map visualizes educational attainment for the United States population aged 25 and up using data from the 2011–2015 American Community Survey. The sections in this article address a series of methodological and interface design choices intended to improve the legibility of the interactive dot map, with the goal of making the map comprehensible to a wide range of potential map viewers. Section 2 briefly reviews the literature on dot mapping and introduces some examples of popular interactive dot maps from recent years. Section 3 addresses the project's data and methodology, with a focus on the use of dasymetric dot placement to improve the map's comprehensibility. In Section 4, a series of interactive methods used to improve the legibility of interactive dot maps are considered. These include zoom-dependent data and styling; interactive data filters; linked charts to summarize the visible data on the map; and responsive map display depending on the size of the viewer's screen. Section 5 then covers some of the challenges presented by interactive dot map development, especially when designing for a wide public audience that may include both subject experts and inexperienced map readers.

Interactive Dot Mapping

Dot maps have a long history within cartography, with example uses spanning the past 100 years (de Geer 1922; Roth 2010). Commonly, the choice of dot size and dots-to-data ratio is a subjective decision made by the cartographer based on a visual assessment of the legibility of the map (Kimerling 2009; Roth 2010). However, cartographers have also established a set of guidelines for optimizing the placement of dots on dot maps. A common approach is the Mackay nomograph, which is a tool to help determine optimal dot size and dot density given the area and scale of a cartographer's map (Mackay 1949). Other more recent approaches have sought to improve upon the nomograph (de Berg and others 2004; Kimerling 2009; Chua and Vande Moere 2017) or provide automated solutions for dot mapping (Hey 2012; Hey and Bill 2014).

Commonly, cartographers design dot maps as *pointillist* maps, in which the colour of dots is mapped to a set of categorical values. In turn, pointillist dot mapping allows cartographers not only to visualize the distribution of a geographic phenomenon but also to consider how this distribution varies among subcategories. Further, as Jenks (1953, 5) explains, "As the distributions of the phenomena being mapped change, the balance between the colored dots will change and new distinctive colors will result." Pointillist mapping can thus visualize areas of categorical homogeneity but also areas of diversity as the colours for different categories blend together. As Rankin (2010) discusses, pointillist dot mapping offers some sociological improvements on choropleth maps of social data, as it is better suited to showing the internal heterogeneity of areal units. Also, pointillist mapping can illustrate smoother demographic transitions between neighbourhoods rather than abrupt ones at enumeration unit boundaries.

In recent years, Web cartographers have combined modern Web mapping technologies with dot mapping techniques to visualize datasets at a large scale. Whereas the aforementioned techniques are useful for optimizing the display of static dot-density maps, which are generally confined to a particular map zoom level and extent, giving a snapshot of a geographic phenomenon, interactive mapping tools allow users to pan and zoom, and can represent much larger volumes of data. Interactive dot-density maps present additional challenges, as the scale and extent of the map are not fixed, but rather can cover millions of potential views. Further, such interactive maps should incorporate not only cartographic principles but also principles from user experience and interaction design (Roth 2012; MacEachren 2013). Roth (2012, 378) describes cartographic interaction design as "the careful decision making leading to a successful experience with an interactive map." This involves anticipating, as well as possible, the ways in which the intended audience will use the map and outlining what map viewers should take away from the map (Muehlenhaus 2014). This is especially important for Web cartographers designing maps for public consumption, as many map viewers may be using mobile devices that offer a different environment for interaction than desktop computers (Muehlenhaus 2014; Roth 2015).

Arguably the most popular interactive dot map implementation is the University of Virginia's Racial Dot Map (Cable 2013b). Developed by Dustin Cable, this map uses data from the 2010 US Census to visualize the geographic distribution of the US population using five colour-coded categories: White, Black, Asian, Hispanic, and Other. On the map, each dot represents a single Census respondent at the Census block level, and dots are randomly placed within Census blocks. Cable uses Python to iterate through Census blocks in the United States and generate randomly placed dots for each defined racial/ethnic category, and then writes the result to a database. The workflow then uses Processing to read the *XY* coordinates from the dataset and generate PNG map tiles. The Racial Dot Map includes approximately 1.2 million PNG tiles, which are displayed with the Google Maps API.

The Racial Dot Map and its predecessors inspired other dot maps of public data that employ the same tile-generation methodology. The Brazil Racial Dot Map (Pata 2015) replicates Cable's methodology using data from the 2010 Brazilian Census. Manduca's (2015) map *Where Are the Jobs?* similarly applies this methodology, but to a map of the locations of jobs colour-coded by major category (Manufacturing, Professional Services, Healthcare/Education/Government, and Retail/Hospitality). Data for this map come from the US Census Bureau's Longitudinal Employer–Household Dynamics dataset. Cable (2013a) also applies this methodology to a dot map of the 2008 US Presidential election results, revealing precinct-level patterns of voting for the Democratic and Republican candidates.

Not all popular examples of interactive dot mapping employ this methodology with its one-to-one dot-to-data relationships. Walker (2015) maps the US immigrant population from the 2009–2013 American Community Survey, with each dot representing approximately 20 immigrants, and dots colour-coded by major region of origin. Schwenke's (2015) dot map of renter-occupied versus owner-occupied housing units uses dot-density where each dot represents 25 housing units. Notably, both of these maps use Mapbox libraries to construct their visualizations, as opposed to Processing and the Google Maps API. Field's (2013) dot-density map of the 2012 US presidential election results employs ArcGIS Online technology to render the interactive map and extends the interactive dot-density mapping methodology by employing zoom-dependent data sources. The dot-to-data ratios vary depending on the zoom level of the map, allowing the map, according to Field, "flexibility to be able to show overall patterns at a national and regional scale but which adds detail at each successively larger scale." Cartographer and developer Eric Fischer has incorporated this concept of scalability into his *tippecanoe* software (Fischer 2017), which algorithmically can add or remove dots (or other geographic features) depending on the map's zoom level.

Data and Methodology: Educational Attainment in America

Data for the Educational Attainment in America map come from the National Historical Geographic Information System, or NHGIS (Manson and others 2017). Datasets include 2015 Census tract shapefiles, which are derived from the US Census Bureau's TIGER/Line shapefiles and processed by NHGIS, and 2011–15 American Community Survey Table B15003, Educational Attainment For the Population 25 Years and Over. Census tracts are preferred to block groups in this project due to their smaller margins of error.

The decision to use Census tracts as sources from which to draw aggregated data has consequences, however, as some Census tracts are large in area, especially in rural areas, and others include both land and water areas within their boundaries. Placing dots randomly within those Census tracts has the potential to create a misleading representation of population distribution. To resolve this, the map incorporates a *dasymetric mapping* strategy (Eicher and Brewer 2001; Mennis 2003) in which ancillary information that describes the underlying population surface is used to improve the placement of dots. The dasymetric dot-density mapping technique is used by Field (2013), which maps county-level presidential election results from 2012 in relationship to land-cover data from the National Land Cover Database.

In this project, Census blocks are used as the ancillary information for dasymetric mapping. Census blocks are the smallest areal unit in which the US Census Bureau aggregates data, and are often analogous to a city block, yet may also cover nonpopulated areas such as parks, roads, or lakes. For the 2010 US Census, the US Census bureau defined 11,078,297 blocks in the 50 US states plus the District of Columbia. Of these 11 million blocks, however, 4,871,270 had a recorded population of 0 for 2010, meaning that these areas within Census tracts would be unsuitable for the placement of dots. To resolve this, data on block geography and population were obtained from NHGIS and used to isolate the 4.87 million blocks with no population in 2010. Using ArcGIS Pro, these block areas were erased from the NHGIS Census tract shapefile, creating Census tracts that in many cases were collections of polygons that only represented the populated areas within those tracts.

The dasymetric Census tract layer was then merged with data from ACS Table B15003, which is aggregated into five categories: less than high school; high school or equivalent; some college (two-year degree, or attended college but did not finish); bachelor's degree; and graduate degree (encompassing master's, professional, and doctoral degrees). Dots for the map were generated using the *pandas* library and the *arcpy* site package in Python to generate random points within each dasymetric Census tract polygon in the United States corresponding to quantities for each of the five categories. These datasets were then combined into a master CSV file with three columns: "X" (for longitude), "Y" (for latitude), and "level" for educational level represented by the dot. This process was completed for five levels of aggregation: 1 dot corresponding to an average of 25, 50, 100, 200, or 500 people aged 25 and over. In total, this process generated 16,267,458 dots across the five levels of aggregation. Finally, the ordering of the output rows was randomized to improve visual presentation on the eventual map.

The generated CSV files were uploaded to Mapbox Studio for conversion into vector tiles to be used in Web mapping. The code corresponding to the above process is viewable at https://github.com/walkerke/education_map. Map design was completed with Mapbox GLJS, a JavaScript library that enables the creation of WebGL maps that consume Mapbox data and services. The end result is an interactive dot map that enables users to pan and zoom around the United States, comparing trends both within and among regions. A variant of the Mapbox Dark base map is used for the background tiles; ColorBrewer Set1 is used for the colour scheme, although the hex colour #984ea3 (purple) is omitted, as it was difficult to view against the dark base map. The five Set1 colours, while categorical, were then arranged to correspond

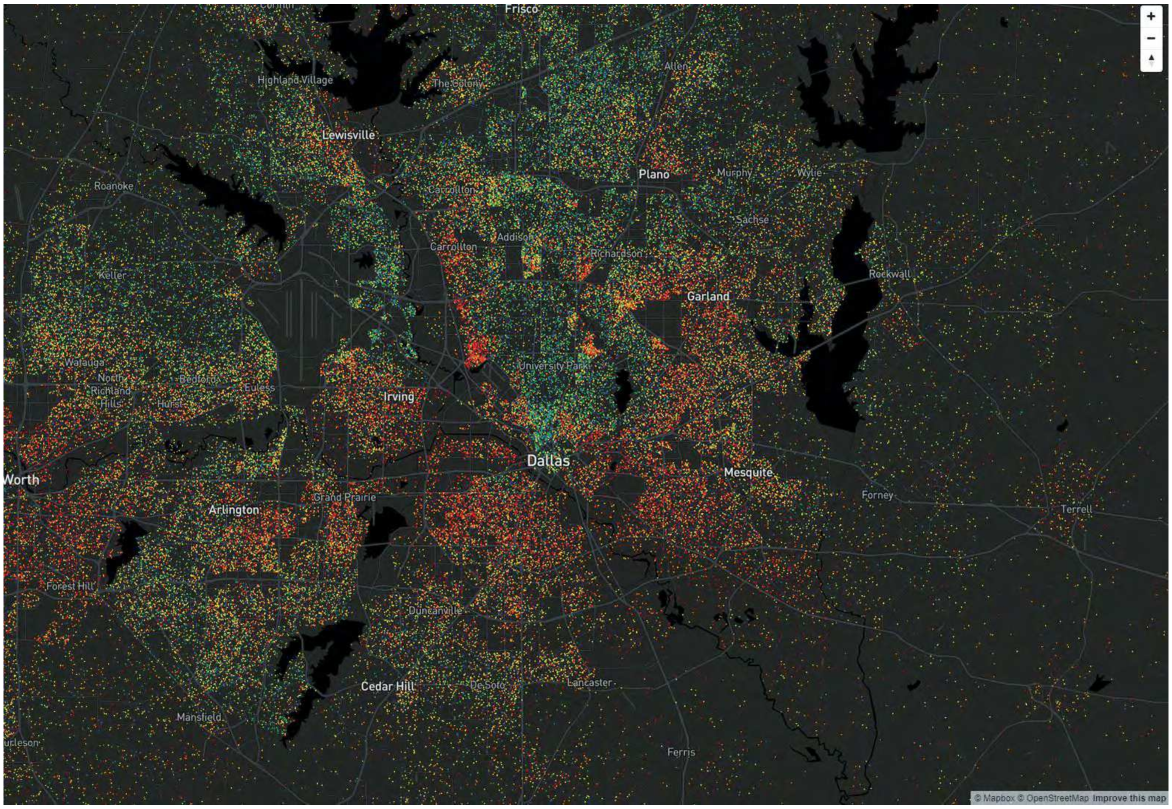

Figure 1. Dasymetric dot map representation of educational attainment for the Dallas, Texas area
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

with the semi-ordinal nature of the categories. Red dots (#e41a1c) represent less than high school; orange (#ff7f00), high school; yellow (#ffff33), some college; green (#4daf4a), bachelor's degree; and blue (#377eb8), graduate degree.

Figure 1, which displays a slice of the map for the area around Dallas, TX, illustrates the utility of the dasymetric dot-mapping technique. There are areas around Dallas that are known to have no population to map viewers – such as the Dallas–Fort Worth and Dallas Love Field Airports, major lakes, and industrial areas. While these areas are located within Census tracts that in many cases have a recorded population, the dasymetric technique used in this map avoids plotting dots in those areas and thus portrays a more realistic population surface.

As evident in Figure 1, dasymetric techniques can help improve the legibility of dot-density maps, especially those that are intended for a public audience that is not necessarily familiar with the ways in which dot maps place dots randomly within enumeration units. However, this technique is not unique to the *interactive* dot map, as it can be employed for static dot mapping as well. The following section outlines methods in interactive cartography that can enhance users' comprehension of dot-density maps. These methods attempt to improve the map's legibility given the ability of users to pan, zoom, and view multiple different maps within a given viewing session. Such map features include zoom-dependent data and styling; interactive data filtering; and linked interactive summary charts.

Methods for Improving the Legibility of Interactive Dot Maps

ZOOM-DEPENDENT DATA AND STYLING

A principal challenge of interactive dot mapping for the entirety of the United States is defining cartographic rules that make the map performant at both small and large scales, and in high-density cities as well as in low-density rural areas. Typically, a dot-density map designer will make decisions about dot size and aggregation level for a specific region, which is where utilities such as the Mackay nomograph are quite useful. However, for a map that is intended for public consumption *and* covers areas at very high and very low densities, consistent dot definitions are important for the viewers.

As mentioned earlier in the article, dots in the map correspond to levels of aggregation from 1:25 up to 1:500. A principal reason for this is to avoid over- or underplotting, depending on the map's zoom level. The map's minimum zoom level is set to 4, which will show the entirety of the

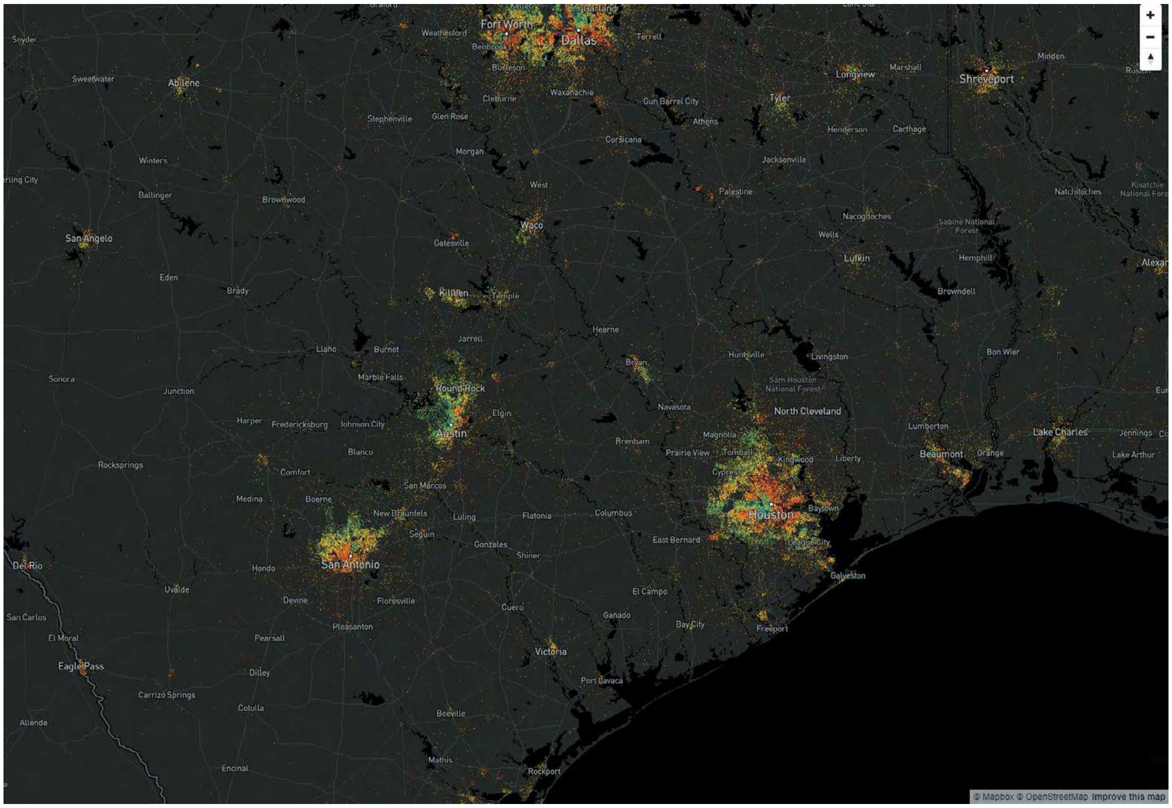

Figure 2. View of east central Texas at zoom level 7 and a 1:100 dots-to-data ratio
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

continental United States; the maximum level is 13.5, which will show a collection of neighbourhoods within a city. In turn, the map will either show or hide the five dot datasets depending on the zoom level of the map. For zooms less than 5, each dot represents approximately 500 people aged 25 and up from the ACS; for zooms 5 to 7, each dot represents 200; for zooms 7 to 9, the ratio is 1:100; for zooms 9 to 11, 1:50; and for zooms greater than 11, the ratio is 1:25. Rather than using an algorithm to select the appropriate representation of dots per zoom level, whole numbers are used to facilitate comprehension by the target non-expert audience for the map.

Dots are progressively added or removed from the map as the user zooms in and out to explore patterns around the United States. In concert with this zoom-dependent data, the map uses zoom-dependent styling available in Mapbox GL JS to make dots appear at smaller sizes when the map is viewed at a smaller zoom level, and then progressively increase in size as the user zooms in to facilitate viewing. Even at small zoom levels with a smaller number of visible dots, dots will cluster together, giving users a general sense of patterns both around the United States and in different parts of metropolitan areas. Figure 2 is an example for east central Texas at zoom level 7, where dots represent a ratio of 1:100.

In the figure, general zones of educational attainment are visible for the three major metropolitan areas on the map: Houston, Austin, and San Antonio. In Austin, for example, areas on the western side of the metropolitan area tend to have greater educational attainment, represented by the bluer hues of the dots; areas on the eastern side of the metropolitan area are represented with more red and orange hues. The general patterns characteristic of metropolitan area are still visible with fewer dots, allowing the user to compare patterns in the same view with other metropolitan areas in Texas, such as Houston and San Antonio. Further, the pointillist technique employed by the map illustrates, even at this zoom level, the ways in which populations with different educational attainment blend into one another geographically in the metropolitan areas.

Information about the dots-to-data ratio and how this varies as the user zooms in and out is available in the map's sidebar as dynamic text. The minimum zoom level, the maximum zoom level, and an associated text ID are available to the map as a JavaScript array. The map's script includes logic to fetch the zoom level of the map at each zoom by the user, and then shows or hides layers according to their associated zoom levels and modifies the text in the sidebar accordingly. The array appears as follows:

```
var layerids = [
{id: 'education-25', minzoom: 11, maxzoom: 14, textid: 'layer25'},
{id: 'education-50', minzoom: 9, maxzoom: 10.99999, textid: 'layer50'},
{id: 'education-100', minzoom: 7, maxzoom: 8.99999, textid: 'layer100'},
{id: 'education-200', minzoom: 5, maxzoom: 6.99999, textid: 'layer200'},
{id: 'education-500', minzoom: 3, maxzoom: 4.99999, textid: 'layer500'}
];
```

Associating the five layer IDs with minimum zooms, maximum zooms, and a text ID string in this fashion allows additional logic to build zoom-dependent styling into the Mapbox map. The first piece of code below illustrates how the map first iterates through the array to establish zoom ranges for each layer, which is identified by its id property. The second component of the code responds to user zoom events by checking the current zoom level of the map and then modifying the HTML element that displays the dot information to the user, which is identified in the map's HTML by the textid property:

```
// Set the layer zoom range

layerids.map(function(id) {
  map.setLayerZoomRange(id.id, id.minzoom, id.maxzoom);
});

// Modify dot info in sidebar

layerids.map(function(x) {
  map.on('zoom', function() {

    var d = document.getElementById(x.textid);
    var zoom = map.getZoom();

    if (zoom >= x.minzoom && zoom <= x.maxzoom ) {
      d.style.display = 'block';
    } else {
      d.style.display = 'none';
    }
  });
});
```

In turn, map viewers will always have information about the dots-to-data ratio as they dynamically zoom in and out, helping facilitate comprehension of the map's content.

INTERACTIVE DATA FILTERS

Zoom-dependent data and styling are useful in preserving the legibility of a map while maintaining comprehensibility as the user zooms in and out. One issue that arises, however, relates to the widely varying population densities of communities across the United States. Given that the map covers the entirety of the United States, styling is preserved within zoom levels – which means that dot sizes and densities at a given zoom level will be the same in Manhattan as in less dense cities and rural areas. A styling optimized for very high-density areas would make lower-density areas illegible; however, the consequence of this is the overplotting of dots in these highly dense areas, as evidenced in Figure 3.

The mixing of colours on a multivariate dot map represents one aspect of the pointillist cartographic technique, as the blending of colours in areas of educational heterogeneity is distinguished from the chromatic homogeneity in areas where one or two categories predominate. However, as Figure 3 also illustrates, dots in Manhattan can occlude the visibility of all dots in the area, given the very high dot density. This issue is addressed in two ways. First, before the longitude/latitude

  doi:10.3138/cart.53.3.2017-0021

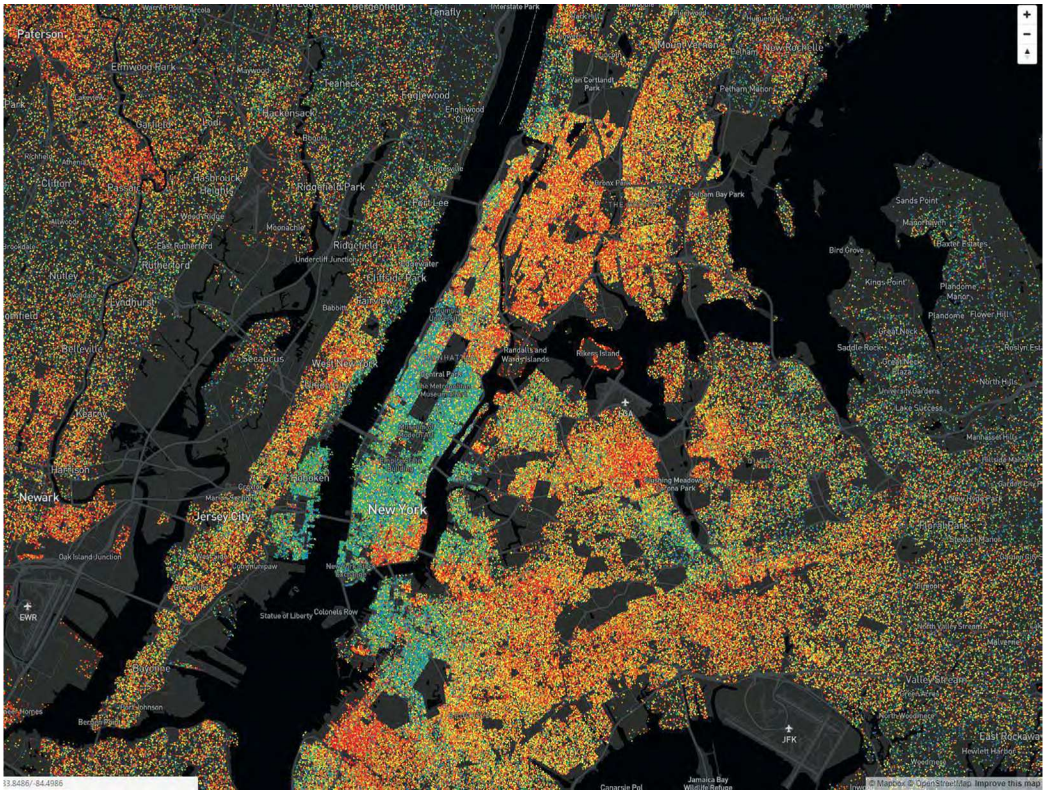

Figure 3. New York City, with all dot categories displayed
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

CSV files were uploaded to Mapbox Studio, the ordering of the rows in the dataset was randomized to prevent categories from appearing "layered" on top of one another. This allows for a general visual representation of the overall dot breakdown in highly dense areas like Manhattan.

Despite this, discerning patterns for individual groups can be difficult in very dense areas. In response to this, the map includes *interactive data filters*. The Mapbox GL JavaScript library includes functionality to filter layers interactively based on categories within a dataset with an expression. For this map, the original filter (when all categories are visible) is as follows:

```
var filter = ["in", "level", "less_than_hs", "high_school",
              "some_college", "bachelors", "graduate"];
```

This filter expression is a JavaScript array that specifies the categories **in** the "level" attribute of the dataset that should be visible on map. By default, the map will show all categories.

To filter categories on the map, users can click on the corresponding categories in the legend. If a category is visible and a user of the map clicks its entry in the legend, this will fire an event that will remove the corresponding entry from the filter expression and in turn hide the category on the map. Conversely, if the category is already hidden, the map will fire an event to add the category to the filter expression. The interactive filter in turn allows for the map viewer to customize the viewing experience and improve upon some of the legibility issues inherent in dot-density mapping. For example, if a map user wants to view the distribution of graduate degree holders exclusively in the New York City area, all other categories can be turned off in the legend to isolate dots representing graduate degree holders on the map. An example of this is illustrated in Figure 4.

As shown in Figure 4, graduate degree holders in central New York City are most densely clustered on the Upper East and Upper West Sides of Manhattan; far fewer graduate degree holders are found in the Bronx and western Queens. In Hudson County, NJ across the Hudson River, graduate degree holders tend to be more concentrated in and around Hoboken than in other parts of the area. While these patterns are discernible when all educational

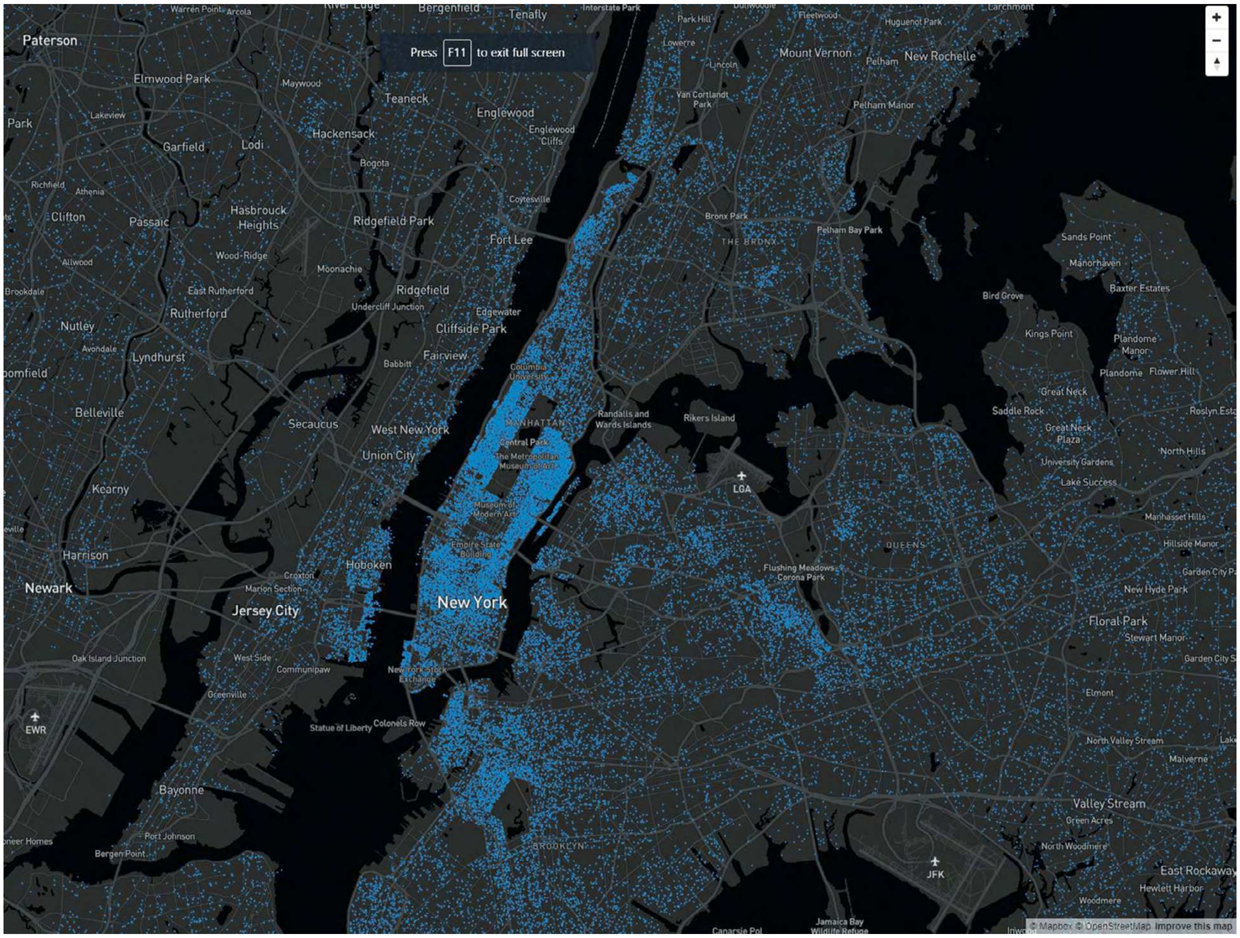

Figure 4. New York City, with only dots representing graduate degree holders displayed
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

attainment categories are visible on the map, they are far easier to interpret when graduate degree holders are isolated using the filters.

This functionality also illustrates the utility of using vector tiles as opposed to the traditional raster tiles for interactive mapping. When users upload data to Mapbox Studio, as I did for this project, Mapbox will convert their data to vector tiles, which in turn can be displayed in a Web map using the Mapbox GL JavaScript library. As opposed to raster-tile-based Web maps, which show tessellations of PNG images depending on the zoom level and extent of the user's screen, vector tile Web maps can expose the data to the user, allowing in-browser manipulation such as the interactive filters discussed above. By exposing the underlying data to the Web client in this way, vector tile Web maps allow the developer to incorporate additional functionality as well as to summarize and explore the visible data on the screen. An example of this, a summary chart that is linked to the visible dot-density map, is discussed in the next subsection.

LINKED SUMMARY CHARTS

Even with the implementation of functionality to improve the legibility of interactive dot-density maps, it can be difficult for users to summarize the data that appear on the screen. Dot maps are particularly well suited for demonstrating where proportionally "more" or "less" of a given population is located relative to other places; however, some quantitative assessment of this comparison can also assist map readers with interpretation of the visible patterns. To aid in this, Educational Attainment in America includes an option for map users to generate a bar chart that visualizes the percentage breakdown of visible dots on the screen, shown in Figures 5 and 6.

Figures 5 and 6 compare the north and south sides of the city of Chicago with charts that have been generated using this method. As illustrated in the figure, the north side of Chicago has proportionally more residents age 25 and up with bachelor's and graduate degrees, whereas the south side of Chicago has proportionally more residents without a higher education. While this information is

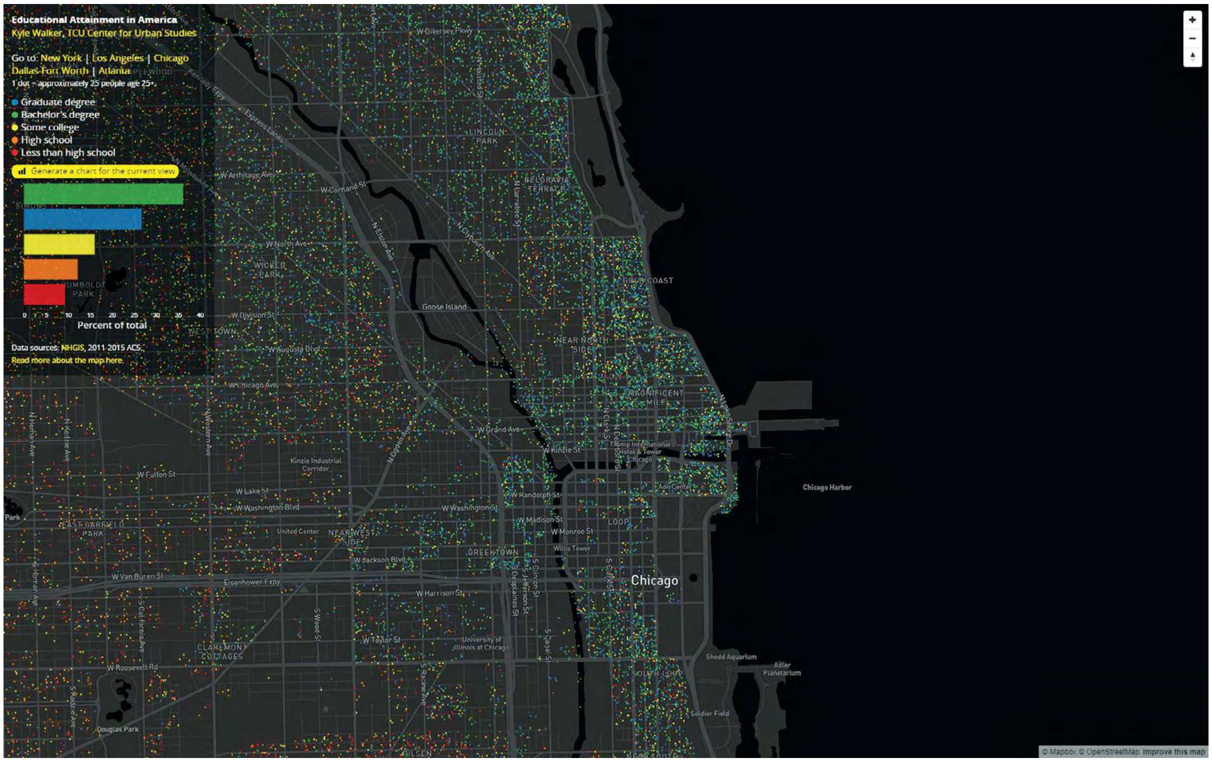

Figure 5. The north side of Chicago, IL with a linked summary chart displayed
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

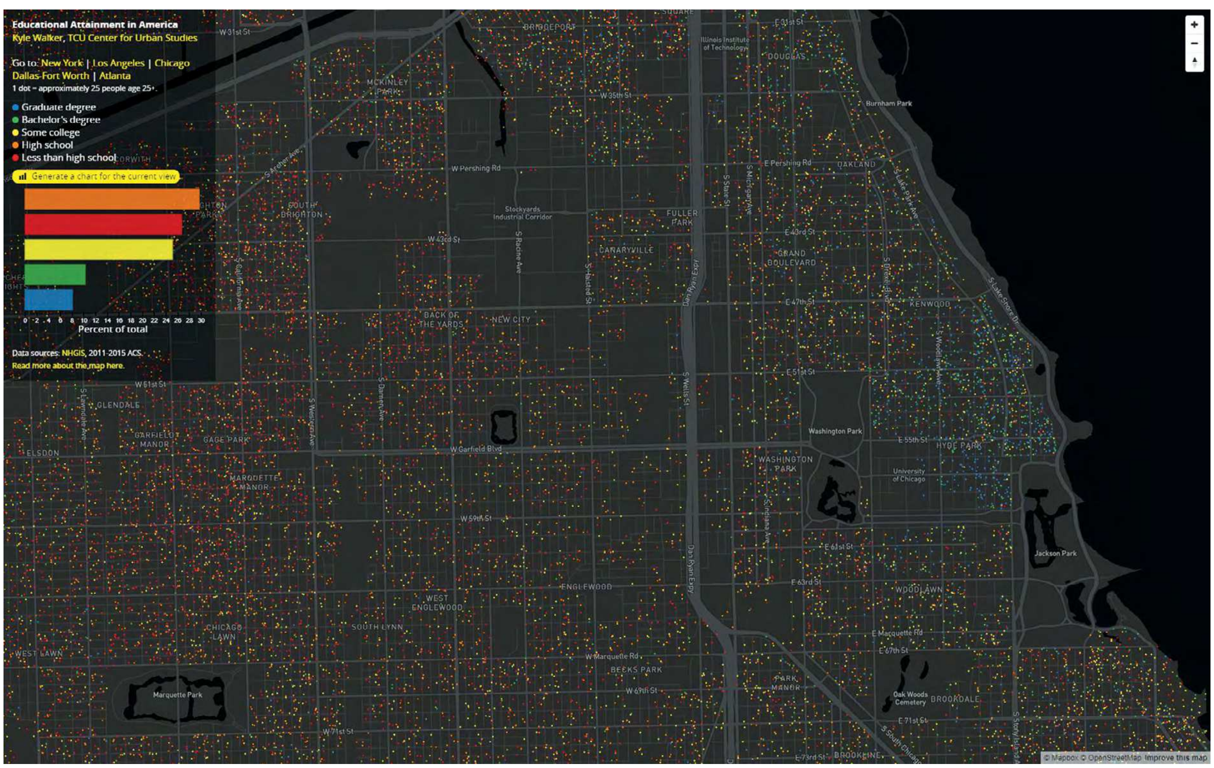

Figure 6. The south side of Chicago, IL with a linked summary chart displayed
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

discernible by viewing the dots on the map as the user pans around Chicago, the chart tool enables the user to make clearer comparisons between these different areas of the city.

Users can create the summary chart by clicking a button on the map's sidebar labeled "Generate a chart for the current view." When clicking the button, the user fires an event that invokes a defined function, get_percentages(), that calculates the percentage of the total visible dots on the screen constituted by each of the five groups and returns the result as a JavaScript array. This is made possible by the queryRenderedFeatures() function in the Mapbox GL JS library, which developers can build into their maps to allow users to retrieve map data from the loaded vector tiles. The JavaScript array is then passed to the chart-generating code, which uses the Dimple.js wrapper around the D3.js visualization library (Kiernander 2017; Bostock, Ogievetsky, and Heer 2011). If the users modifies the map view, clicking the button again will regenerate the chart. While Mapbox GL JS includes functionality to regenerate the percentage data with every move of the map, this leads to significant performance problems; thus, the button-click event approach is adopted here. The Dimple.js library allows an animation as the chart regenerates as well as hover tooltips for users to access the underlying percentage values.

For the chart, percentages are preferred to raw counts for multiple reasons. First, ACS data are not intended for measuring counts, but rather for comparing categories, given that they are based on a sample subject to a margin of error. Percentages facilitate these types of comparisons, whereas counts could be more misleading. Further, not every dot represents the same quantity – for example, in Figure 6, some dots will represent more than 25 people and some will represent fewer – meaning that a percentage breakdown is more appropriate for making the general comparisons that are available on the map.

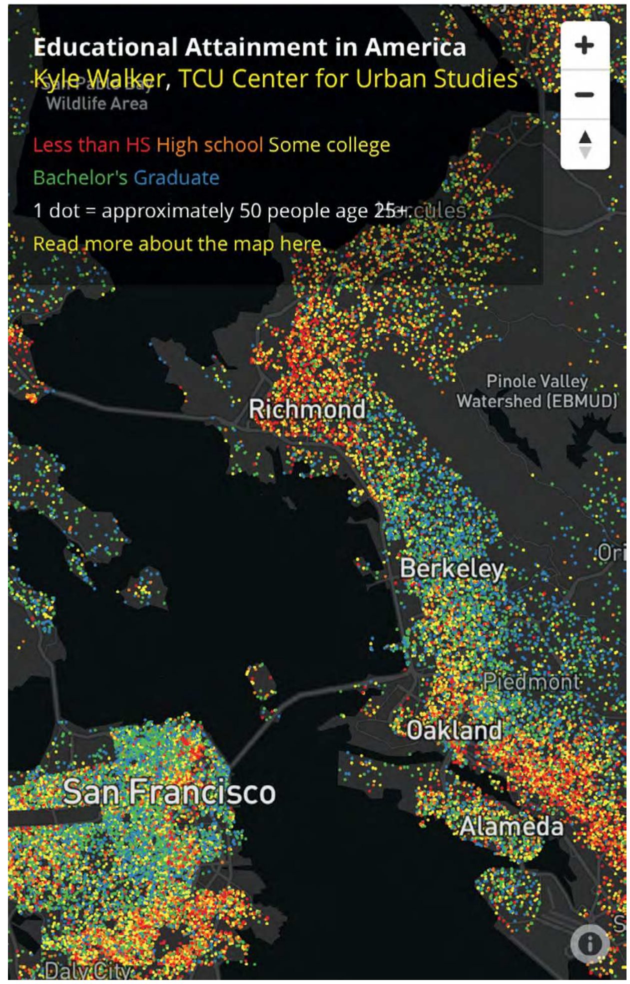

Figure 7. Mobile display of the map, as seen in the Chrome browser on an iPhone
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

Challenges in Interactive Dot-Density Mapping

As discussed in the previous section, the use of Web technologies to build interactive dot maps offer cartographers additional options to improve the viewing experience for their map readers. However, the use of such mapping technologies also introduces additional challenges for the cartographer. One very prominent challenge is the fact that the map will be viewed on a variety of devices – from widescreen monitor to mobile phone, and everything in between – with screen sizes of varying dimensions. Thus, to reach the greatest audience successfully, the map needs to be performant across these devices and screen sizes.

When *Educational Attainment in America* is viewed on a monitor or laptop or most tablets, functionality such as the interactive legend and linked interactive chart appears in the map's sidebar. On a phone, however, a sidebar with all of these features would obscure the entire screen. In response to this, the map displays a more limited sidebar for viewing on mobile devices that functions as an informational header for the map. In many ways, this involves a design of two map layouts: one that appears when the user's screen width is more than 650 pixels wide, and another for when it is less than 650 pixels wide. This is accomplished by defining the CSS classes mobile-hide and mobile-show in the map's stylesheet, which re then applied to the HTML elements that I wish to hide or show on different devices. An example of the map as viewed in a Chrome browser on an iPhone is shown in Figure 7.

In the mobile map, some of the information in the sidebar is preserved. Title and author information and links still appear, as do the dynamic text that informs map

  doi:10.3138/cart.53.3.2017-0021

readers of the dots-to-data ratio and a link to the project GitHub page. The legend is replaced with text for each educational attainment category in the corresponding dot colour on the map, and the interactive chart is omitted entirely.

These mobile-specific modifications are important, given the way that users have accessed the map. Since its publication in March 2017, 54.1 percent of visitors have accessed the map from desktop computers, 39.6 percent visited from cell phones, and the remaining 6.3 percent viewed the map on tablets. Of all of the mobile views (phone + tablet), over 40 percent of visitors viewed the map on iPhones, which are represented by the image in Figure 7. In turn – while a minority of users viewed the mobile-specific map – they nonetheless constitute a significant proportion of map viewers, which points to the importance of mobile-friendly map design.

Another issue involves colour selection on the map. The map adopts a semi-linear approach to visualizing educational attainment by selecting a categorical colour palette arranged to represent similar levels of educational attainment with similar hues. This meant that the brightest colour used on the map – #ffff33, for Some College – is the middle category. If the classes were interpreted as categorical this would not be a major issue, but it is generally inadvisable for diverging palettes. Additionally, categorical palettes are difficult for colourblind map readers to interpret. Thus, a colourblind-safe alternative version of the map is linked to on the project GitHub page. This map is identical in content to the original Educational Attainment in America Map, but it uses the diverging ColorBrewer palette RdYlBu instead of the categorical Set1 palette. An example of such a map is found in Figure 8.

While this map has some disadvantages relative to the original map – namely the difficulty in distinguishing neighboring categories – the map is much more legible for colour blind viewers, allowing greater accessibility.

The target audience of the map also raises questions about its interpretability. A cartographically literate audience would have little difficulty interpreting a dot-density map, given that it is likely familiar with the underlying data-generating process. However, interactive maps such as this have the potential to reach a very broad

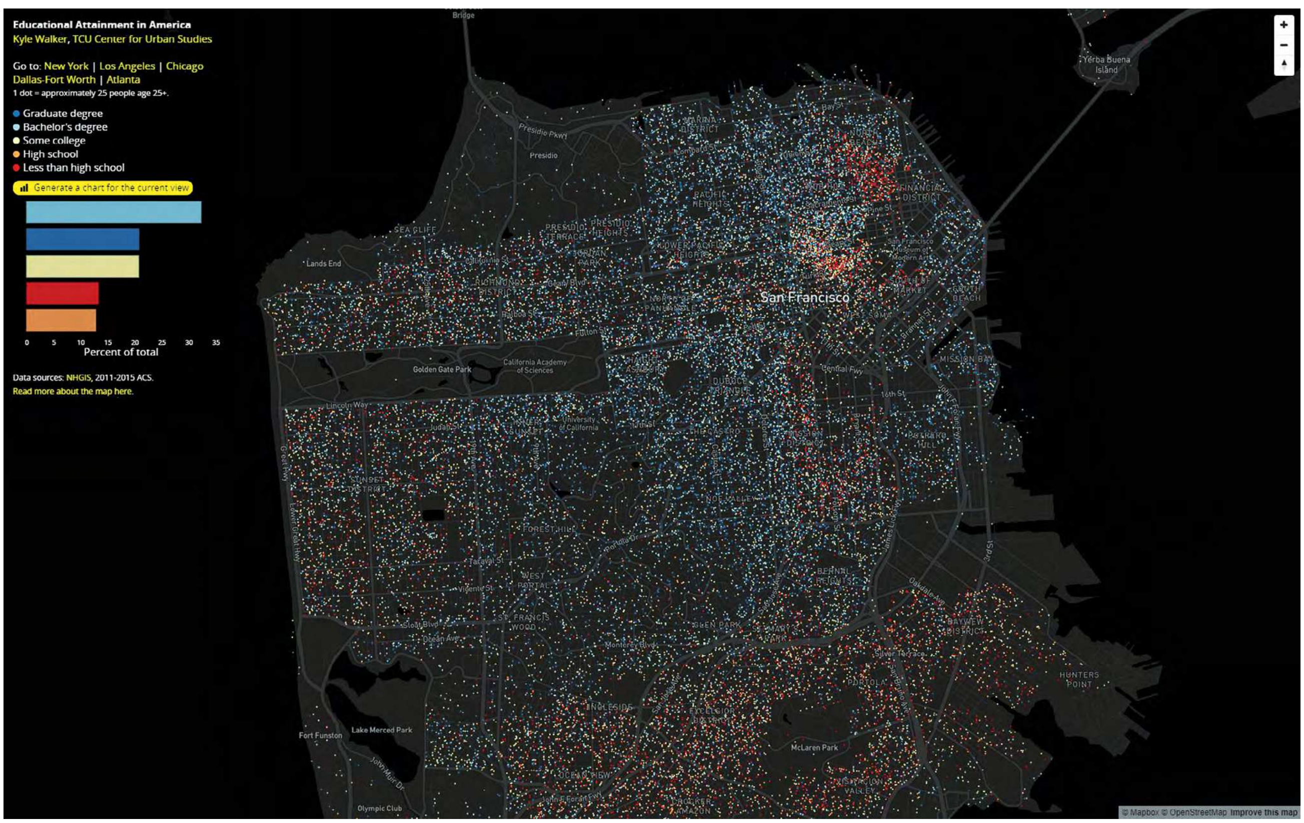

Figure 8. The *Educational Attainment in America map* with a colourblind-safe palette
Source: Walker (2017). © Mapbox © OpenStreetMap contributors. OpenStreetMap data is available under the Open Database License.

audience by attracting media attention. After the map's publication, it received coverage in outlets including *CityLab* from *The Atlantic* (Misra 2017), *Business Insider* (Weller 2017), *The Boston Globe* (Salinas 2017) and *D Magazine* (Goodman 2017), which were republished on sites such as *SFGate* and *The Houston Chronicle*. Since its publication in mid-March 2017, the map has been visited over 163,000 times by over 140,000 unique users as of July 2018.

This makes the map subject to feedback from both technical experts and the general population, both directly and through social media. Notably, this includes multiple e-mails either from users who have been unable to find themselves on the map and asked me to add them to it, or from users who are upset that their neighbourhoods are missing dots that are representative of their own educational achievements. As discussed in this article, there are multiple factors that introduce uncertainty into the visualization. As the data are from the five-year 2011–15 American Community Survey, they represent a sample of the US population subject to a margin of error *and* an average over a five-year period that is between two and six years prior to the making of the map. Additionally, dots represent aggregations of ACS data – between 25 and 500 depending on the zoom level – that are randomly placed within Census tracts.

These concepts, however, may be unfamiliar to lay readers, who in turn may interpret individual dots on the map as individual people, leading to misinterpretation of the map's content. In part, this has to do with the ability to zoom on these dot maps – leading to an adjustment of the maximum zoom level from 14.5 to 13.5 to try to minimize these sorts of misinterpretations. This points to worthy avenues for future research on both non-expert interpretation of dot-density maps and public responses to "viral" map content, given the popularity of map-based content both in data journalism and through social media.

Conclusion

Pointillist interactive dot mapping has proven a popular method for visualizing large amounts of social data on the Web in recent years, as it is well suited to allow map viewers to explore patterns in categorical data and observe areas where these categories blend together. This article outlines a series of analytical and visual methods for creating interactive and legible dot maps. These methods include dasymetric dot mapping to reduce the improper placement of dots; the use of zoom-dependent data and styling; interactive filters to reduce overplotting and allow a customizable user experience; and linked charts to summarize data on the map, which can be difficult for users to do on their own. While these techniques, made possible by interactive mapping technologies, can improve the user experience for viewers of dot maps, the use of such technologies also introduces some challenges for cartographers. These include the necessity of building the map to be legible on mobile devices as well as desktop computers and accounting for the diversity of the audience in regard to colour selection and interpretability of the data.

A major reason for the widespread popularity of interactive dot maps such as *Educational Attainment in America* is their appeal to a broad audience. The use of modern Web mapping technology such as Mapbox Studio allows users to customize the map viewing experience for any corner of the United States and potentially relates these maps to their own experiences. *Educational Attainment in America* attempts to leverage features incorporated into these modern Web technologies to improve the user experience of interactive dot maps. However, use of such technologies makes possible additional contributions of the map beyond the cartographic product itself. As all of the map's source code is available via GitHub at https://github.com/walkerke/education_map, cartographers can use the examples discussed in this article in their own projects.

Two notable examples include Passos (2017), which adapts this methodology to map educational attainment in Sao Paolo and Rio de Janeiro, Brazil, and the adaptation of the *Educational Attainment in America* map by the Texas Higher Education Coordinating Board (2017) as part of their efforts to improve educational achievement in the state. In the Texas map, the cartographers adapted the aforementioned methods to create a tailored solution that corresponded specifically to their organizational goals, focusing on the population aged 25–34. Passos, on the other hand, adapts the methods to a different dataset – the Brazilian Census. This open approach to the map's source code is one aspect of the Open Geographic Information Science framework proposed by Singleton, Spielman, and Brunsdon (2016). By using open data and making the source code publicly available, the framework outlined in this article not only can scale a map to hundreds of thousands of public viewers, but also can scale it to other developers who can make their own public-facing contributions.

Author Information

Kyle E. Walker is an associate professor of geography and the Director of the Center for Urban Studies at Texas Christian University. His current research focuses on the spatial demography of US metropolitan areas, demographic data visualization, and tools for open data science. E-mail: kyle.walker@tcu.edu.

  doi:10.3138/cart.53.3.2017-0021

References

Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer. 2011. "D3 Data-Driven Documents." *IEEE Transactions on Visualization and Computer Graphics* 17(12): 2301–9. https://doi.org/10.1109/TVCG.2011.185.

Cable, Dustin. 2013a. "A Map That Shows What's Wrong with Washington." StatChat from the Demographics Research Group at the University of Virginia. Available at http://statchatva.org/2013/10/28/a-map-that-shows-whats-wrong-with-washington.

Cable, Dustin. 2013b. "The Racial Dot Map." Weldon Cooper Center for Public Service, University of Virginia. Available at https://demographics.virginia.edu/DotMap.

Chua, Alvin, and Andrew Vande Moere. 2017. "BinSq: Visualizing Geographic Dot Density Patterns with Gridded Maps." *Cartography and Geographic Information Science* 44(5): 390–409. https://doi.org/10.1080/15230406.2016.1174623.

de Berg, Mark, Prosenjit Bose, Otfried Cheong, and Pat Morin. 2004. "On Simplifying Dot Maps." *Computational Geometry* 27(1): 43–62. https://doi.org/10.1016/j.comgeo.2003.07.005.

de Geer, Sten. 1922. "A Map of the Distribution of Population in Sweden: Method of Preparation and General Results." *Geographical Review* 12(1): 72–83. https://doi.org/10.2307/208657.

Eicher, Cory L., and Cynthia A. Brewer. 2001. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation." *Cartography and Geographic Information Science* 28(2): 125–38. https://doi.org/10.1559/152304001782173727.

Field, Kenneth. 2013. "Red, Blue and Purple: Mapping the 2012 US Presidential Election." ArcGIS Blog. Available at https://blogs.esri.com/esri/arcgis/2013/04/29/red-blue-and-purple-mapping-the-2012-us-presidential-election.

Fischer, Eric. 2017. "Tippecanoe: Build Vector Tilesets from Large Collections of GeoJSON Features." Available at https://github.com/mapbox/tippecanoe.

Goodman, Matt. 2017. "Use This Interactive Map To Dissect Dallas by Educational Achievement." *D Magazine*, March. Available at https://www.dmagazine.com/frontburner/2017/03/use-this-interactive-map-to-dissect-dallas-by-educational-achievement.

Hey, Annette. 2012. "Automated Dot Mapping: How to Dot the Dot Map." *Cartography and Geographic Information Science* 39(1): 17–29. https://doi.org/10.1559/1523040639117.

Hey, Annette, and Ralf Bill. 2014. "Placing Dots in Dot Maps." *International Journal of Geographical Information Science* 28(12): 37–41. https://doi.org/10.1080/13658816.2014.928822.

Jenks, George F. 1953. "'Pointillism' as a Cartographic Technique." *Professional Geographer* 5(5): 4–6. https://doi.org/10.1111/j.0033-0124.1953.055_4.x.

Kiernander, John. 2017. "Dimple: An Object-Oriented API for Business Analytics Powered by D3." Available at http://dimplejs.org.

Kimerling, A. Jon. 2009. "Dotting the Dot Map, Revisited." *Cartography and Geographic Information Science* 36(2): 165–82. https://doi.org/10.1559/152304009788188754.

MacEachren, Alan M. 2013. "Cartography as an Academic Field: A Lost Opportunity or a New Beginning?" *Cartographic Journal* 50(2): 166–170. https://doi.org/10.1179/0008704113Z.00000000083.

Mackay, J.R. 1949. "Dotting the Dot Map: An Analysis of Dot Size, Number, and Visual Tone Density." *Surveying and Mapping* 9(1): 3–10.

Manduca, Robert. 2015. "Where Are the Jobs?" Available at http://www.robertmanduca.com/projects/jobs.html.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2017. *IPUMS National Historical Geographic Information System: Version 12.0* [Database]. Minneapolis: University of Minnesota. https://doi.org/10.18128/D050.V12.0.

Mennis, Jeremy. 2003. "Generating Surface Models of Population Using Dasymetric Mapping." *Professional Geographer* 55(1): 31–42. https://doi.org/10.1111/0033-0124.10042.

Misra, Tanvi. 2017. "Mapping the Achievement Gap." *CityLab*, March. Available at https://www.citylab.com/equity/2017/03/a-us-dot-map-of-educational-achievement-segregation/519063.

Muehlenhaus, Ian. 2014. *Web Cartography: Map Design for Interactive and Mobile Devices*. Boca Raton, FL: CRC Press.

Passos, Carolina. 2017. "Educational Landscape of the Adult Population in Metropolitan Regions." Mapping-Lab. Available at https://www.mappinglab.me/education-level.

Pata. 2015. "Brazil Racial Dot Map." Available at http://patadata.org/maparacial/en.html.

Rankin, William. 2010. "Cartography and the Reality of Boundaries." *Perspecta* 42: 42–45.

Roth, Robert E.2010. "Dot Density Maps." *Encyclopedia of Geography*, vol. 2, ed. Barney Warf, 787–90. London: Sage. https://doi.org/10.4135/9781412939591.n303.

Roth, Robert E. 2012. "Cartographic Interaction Primitives: Framework and Synthesis." *Cartographic Journal* 49(4): 376–95. https://doi.org/10.1179/1743277412Y.0000000019.

Roth, Robert E. 2015. "Interactivity and Cartography: A Contemporary Perspective on User Interface and User Experience Design from Geospatial Professionals." *Cartographica* 50: 94–115. https://doi.org/10.3138/cart.50.2.2427.

Salinas, Sara. 2017. "Where Are Boston's Most Educated Neighborhoods?" *The Boston Globe*, 16 August. Available at http://www.bostonglobe.com/metro/2017/08/16/where-are-boston-most-educated-neighborhoods/mLs1V99GdwITcw41cyteiO/story.html.

Schwenke, Ken. 2015. "Where the Renters Are." Available at http://schwanksta.com/maps/where-the-renters-are-dots#4/38.73/-97.53.

Singleton, Alex David, Seth Spielman, and Chris Brunsdon. 2016. "Establishing a Framework for Open Geographic Information Science." *International Journal of Geographical Information Science* 30(8): 1–15. https://doi.org/10.1080/13658816.2015.1137579.

Texas Higher Education Coordinating Board. 2017. "60×30TX: Texas's Educated Population Ages 25–34." Available at https://www1.thecb.state.tx.us/map/attainment/.

Walker, Kyle. 2015. "Mapping Immigrant America." Available at http://personal.tcu.edu/kylewalker/immigrant-america.

Walker, Kyle. 2017. "Educational Attainment in America." Available at http://personal.tcu.edu/kylewalker/maps/education.

Weller, Chris. 2017. "Mesmerizing Maps Show Where the Most Educated Americans Live." *Business Insider*, August 15. Available at http://www.businessinsider.com/where-most-educated-americans-live-2017-8.