

Nov 25, 2024

Data Visualization

Week 8. Visualizing geospatial data

Reminder

Time series have two components:

- observation
- regarding time stamp

Introduction

Data sets typically consist of numeric and categorical variables.

- To explain the impact of location on observations, data sets utilize coordinate information (such as neighborhood, city, region, country, etc.).
- These types of variables are referred to as spatial variables.
- Spatial variables are often visualized on maps.
- Choropleth and cartogram are among the most well-known methods of map visualization.

Choropleth

There may be a need to show how certain quantities vary across locations. This can be done by individually coloring the regions on a map based on the data dimension to be visualized. Such maps are called choropleth maps.

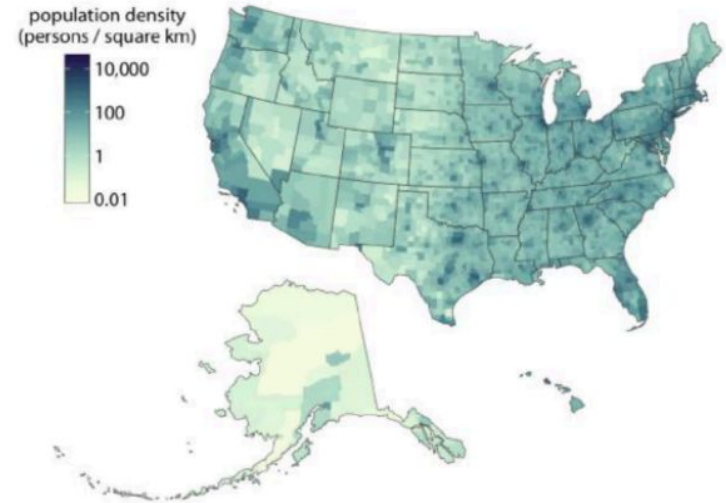


Figure 15-11. Population density in every US county, shown as a choropleth map. Population density is reported as persons per square kilometer. Data source: 2015 Five-Year American Community Survey.

Things to Consider in Choropleths

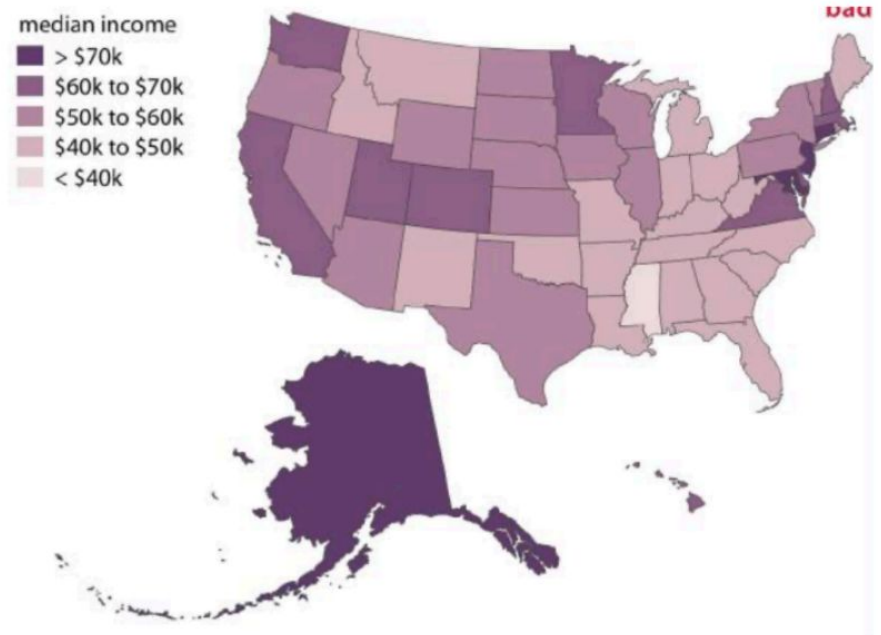
The color issue

- When the range of the variable used for coloring is very wide, it may become difficult to distinguish color transitions and the values corresponding to specific shades.
- In such cases, the variable used for coloring can be divided into specific intervals, making it easier to differentiate the corresponding colors.

Things to Consider in Choropleths

The size issues

- The sizes of regions displayed on maps can lead to misleading interpretations.
- For example, although Alaska occupies the largest area on the map, its population is significantly lower compared to other states.
- The high income levels observed in Alaska could lead to an overly positive and inaccurate interpretation of the overall economic situation in the United States.
- In such cases, the use of choropleth maps should be approached with caution.



Cartogram

Not every map-like visualization needs to be geographically accurate to be useful. A cartogram is a type of map in which the sizes of regions are adjusted based on a specific variable.

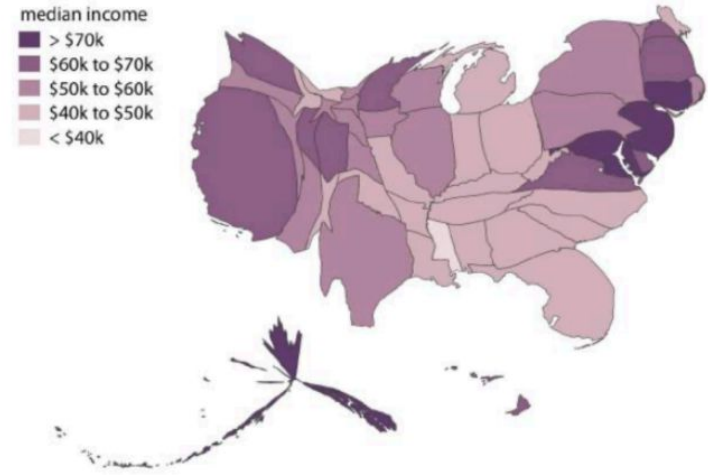


Figure 15-15. Median income in every US state, shown as a cartogram. The shapes of individual states have been modified such that their area is proportional to their number of inhabitants. Data source: 2015 Five-Year American Community Survey.

Cartogram

As an alternative to a cartogram with distorted shapes, a much simpler cartographic heatmap can be used, where each region is represented by a colored square.



Figure 15-16. Median income in every US state, shown as a cartogram heatmap. Each state is represented by an equally sized square, and the squares are arranged according to the approximate position of each state relative to the other states. This representation gives the same visual weight to each state. Data source: 2015 Five-Year American Community Survey.

Cartogram

More complex cartograms can also be created by placing individual graphs at the location of each state.

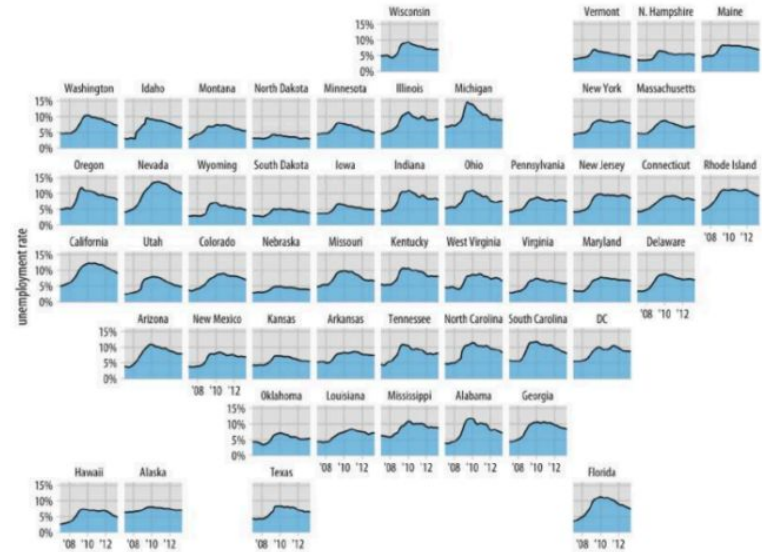


Figure 15-17. Unemployment rate leading up to and following the 2008 financial crisis, by state. Each panel shows the unemployment rate for one state, including the District of Columbia (DC), from January 2007 through May 2013. Vertical grid lines mark January of 2008, 2010, and 2012. States that are geographically close tend to show similar trends in the unemployment rate. Data source: US Bureau of Labor Statistics.

Reference

The notes and plots in the presentation are compiled from Claus O. Wilke's book, *Fundamentals of Data Visualization*.

