

Solutions To The Crowded Scatterplot Problem

A scatterplot visualizes the relationship between two numeric variables. One variable provides the x-axis coordinate for a shape (typically a dot) and the other variable provides the y-axis coordinate. However, scatterplots are ineffective if many shapes have the same coordinates because most the shapes cannot be seen beneath the shape on top. This poster discusses four solutions to the crowded scatterplot problem, using a population map of the United States as an example.

The US population is a crowded scatterplot in that a large segment of the population lives in high population density urban areas that are too small to be visible on a typical map. For example, New York City has over amillion more residents than Idaho, Montana, Wyoming, North Dakota, South Dakota, and Nebraska combined. However, these states occupy 1,200 times more surface area on a US map. If you place a dot on a US map for every resident, New Yorkers would be much less visible than the residents of the aforementioned states. Even within those states, the same challenges occurs. For example, the city of Cheyenne holds nearly 10% of the population of Wyoming. However, the city occupies only 0.02% of the surface area of Wyoming, making these residents less visible than other Wyomingites.

This R code underlying this project is on GitHub at: https://github.com/sjoshuam/crowded_scatterplot