Homework 1b: Construction

Scott Thatcher

10/15/2021

1 Introduction

We've now reviewed the mechanics of ggplot and the grammar of graphics. The purpose of this assignment is to give you more practice using those tools, and to help you see connections between the elements of design and the choices you make in graphing data.

You'll be looking at the Missouri health data set `2016_MO_CLS_PUF_w_DEMOGRAPHICS.xls` that can be found with this assignment on Blackboard. This data set contains county-level data on a variety of health-related variables, as well as demographic data for the counties. Information about each variable can be found on a separate tab on the spreadsheet.

Make sure to download it and put it where you can load it in as part of this assignment. You'll need to use something like the following code:

```
library(tidyverse)
library(readxl)
mohealth <- read_excel("2016_MO_CLS_PUF_w_DEMOGRAPHICS.xls")
```

2 Assignment

1. After reading the data documentation, choose two main variables whose relationships you'd like to explore. Then, choose a third variable that might be of interest, and that could conceivably be related to the first two variables.
2. Think about multiple ways you might display the relationships between these three variables. With three variables, you'll have to pay attention to mapping at least one variable to an aesthetic different from the x and y axes. It might help to make some "by-hand" sketches at this point.
3. Now, make at least two different graphs that display the inter-relatedness of these three variables in very different ways. To create your different graphs, you might
    - choose different aesthetics for the third variable,
    - choose to switch around which aesthetics are used for all three variables,
    - experiment with treating one variable as discrete, rather than continuous (above or below a threshold value, or creating bins for a numeric variable, for example), or
    - use `scale` or `theme` settings to vary the way that a particular aesthetic is realized on the page.

While all graphs should strive to be accurate representations of the underlying data, don't worry if one graph seems to do a much better job than another—the goal here is to explore, learn, and be able to comment on what makes the good one good (if there is a good one).

3 Submitting Your Work

- Once you've made your two graphs, save them as PNG files. Then start a discussion thread in the **Module 1: Construction** forum, and upload your images to that thread.
- In your forum post, add a verbal description of the choices you made in each graph. Make sure to tie it in with the vocabulary on elements of design that we saw in the first lecture. Explain which graph feels the most "successful" to you in communicating about the data? Briefly describe any relationships that are present in the data, according to your graphs. Make sure to refer to the specific real-world meaning of your variables.
- Finally, comment on others' posts, similar to the previous assignment.

4 Suggestions and Hints

- It can help to sketch out ideas on paper first. If you've already got a picture in your head, it's easier to google "how to do X in ggplot?"
- We'll talk more about color later, but for quick access to a variety of color choices, you might use the `scale_color_brewer` and/or `scale_fill_brewer` commands, and play around with the palettes listed in the help (`?scale_color_brewer`)
- As mentioned previously, not all of your choices may be good ones, but the purpose of this assignment is exploration. We'll talk more about the "rules" in Module 2.
- Make sure that your responses to the prompts above do involve using combinations of the ggplot commands we've seen in the lecture. I'd like you to play around with `scale` and `theme` commands (perhaps more in this assignment than in some future assignments!)

5 Assignment Grading

Points will be assigned as follows:

Points	Description
10 pts	First graph.
10 pts	Second graph.
15 pts	Written description/explanation.
5 pts	Responses/critiques of others' posts.