Problem Set 2 - PPOL 6802 Spring 2024

Alex Lundry

2024-02-02

Problem Set #2

To complete this problem set, you will need to download the necessary data from Github. You can do that by running the code below:

```r
library(tidyverse)

d1 <- read_csv("https://raw.githubusercontent.com/alexlundry/ppol6802_spring24/main/problem_sets/dataset

d2 <- read_csv("https://raw.githubusercontent.com/cphalpert/census-regions/master/us%20census%20bureau%

d3 <- left_join(d1, d2, by = "State")
```

Question 1 - Bar Chart

Using ggplot2, make a bar chart of the number of counties in each census Division.

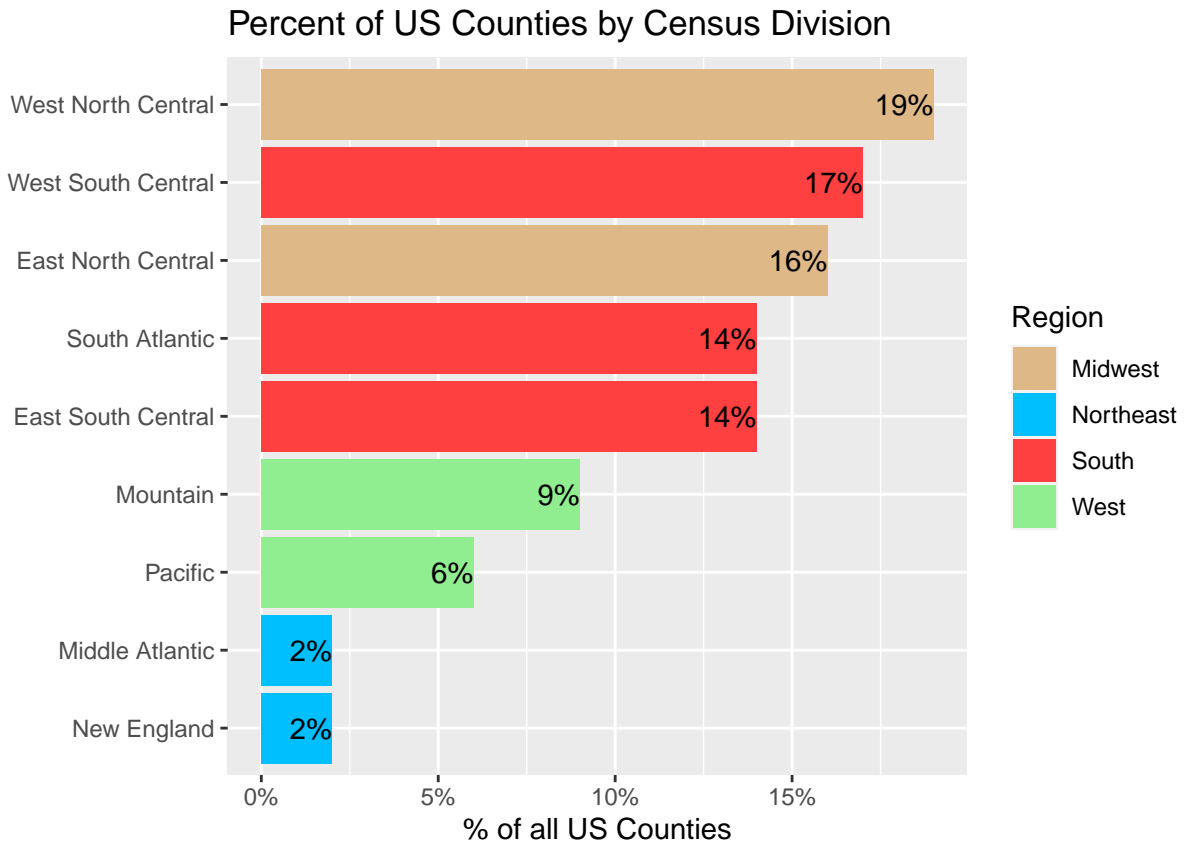**Bonus (not required)**: Sort the bars in order (+1 point)

Question 2 - Histogram

Using ggplot2, create a histogram of `pct_bachelors` that is split into 20 bins.

Question 3 - Bar Chart Reproduction

Using ggplot2, reproduce the bar chart below. To help you do this, here are a few key notes:

- You must first count the counties (rows) within each `Division`
- Then the percentage is calculated; it is rounded to 2 decimal places
- These are horizontal bars.
- The bars are ordered descending by their percent.
- The bars are colored by `Region`
- The bars are labeled with their percent, and the position of that label is on the inside end of the bar
- I am manually setting colors. They are "deepskyblue", "burlywood", "brown1", and "lightgreen".
- The axis labels are formatted to be in %
- There is a title and labels.

Percent of US Counties by Census Division

Question 4 - Research Question

Oh no! Your try-hard problematic Gen X boss that wants to be a Gen Z just texted you:

> "OMG - I don't know how Trump lost in 2020 when he won so many more counties than Biden did!! I told my boss I would blow her mind with a pie chart I made about this. Can you put one together for my meeting with her in 30 minutes? And we need a chart that EATS: title, value labels and appropriate colors. Thanks king!"

What do you send your boss?

BUTTERUP YOUR BOSS BONUS (not required): Draft an email to your boss giving them the chart, but with a warning about why this isn't the most honest way to show the 2020 results. (up to 2 points)