**Problem Set 2**

PSCI 107

24 March 2021

21249856

**Question 1:**

**a.** Using a for loop, I calculated the average temperature within each county for 1980 to 2019. I then visualized these average temperatures using a bar plot, categorizing counties by region:

Chart, box and whisker chart

Description automatically generated

According to this plot, it seems that counties in the South had the highest median average temperature from 1980 to 2019, at around 60 degrees (F). The other three regions had very similar median average temperatures, at around 50 degrees. In the West, there are a number of outlier counties with very high average temperatures. This makes sense, as counties in California, for instance, are probably much warmer than counties in Washington state or Montana.

**b.** To find the highest and lowest average temperatures in each state, I used a for loop to find the maximum and minimum, then saved the results in two new variables, respectively. Then, to find the states with the most and least consistent average temperatures, I created a new variable which recorded the difference (i.e. the range) between the maximum and minimum in each state.

The five states with the most consistent average temperature were: DC, Delaware, Rhode Island, Connecticut, and Vermont. (DC has the most consistent average temperature out of necessity, because it has only one county).

The five states with the least consistent average temperature were: Colorado, Texas, Nevada, Arizona, and California.

Here is a bar graph of each state, ordered from most consistent to least consistent (i.e. the lowest average temperature range to the highest):

Chart, histogram

Description automatically generated

Notably, bigger states are less likely to have a consistent average temperature, and smaller states are more likely to have a consistent average temperature.

**c.** I used the *paste()* command to create 9 unique subregions: “Midwest.ENC”, “Midwest.WNC”, “Northeast.MA”, “Northeast.NE”, “South.ESC”, “South.SA”, “South.WSC”, “West.M”, and “West.P”.

**d.** Using a double for loop, I found the correlations between temperature and population for each subregion for every year. I excluded all missing values from the correlation calculations with “pairwise.complete”. With the resulting correlations, I populated a matrix, *temp.pop.cor*.

Bonus:

To visualize these correlations between temperature and population over time, I created a line plot using *matplot()*, with each line representing a subregion:

Chart, line chart

Description automatically generated

According to this plot, a higher temperature was associated with a higher population for all subregions. The only subregions with moderately strong associations seem to be in the Northeast, at a correlation of around 0.5 from 1980 to 2019. Perhaps the Northeast, with generally colder temperatures, has a higher likelihood of bigger populations in warmer counties. The relationship between temperature and population was fairly consistent over time for all subregions, though the “South.SA” subregion had a slight noticeable increase in the strength of correlation, from 0.13 in 1980 to 0.25 in 2019.

**Question 2:**

I found an article in FiveThirtyEight by Nathaniel Rakich, titled: [The Strongest House Candidates In 2020 Were (Mostly) Moderate](https://fivethirtyeight.com/features/the-strongest-house-candidates-in-2020-were-mostly-moderate/)

**a.** Many of the House candidates with the strongest performances in 2020 were Moderates.

**b.** Outline of the article’s 12 paragraphs:

**1.** Introduces the argument and key concepts: Moderates in Congress are becoming rarer in today’s political polarization, but they were the strongest House candidates in 2020.

**2.** Introduces the data and methods used to support the argument: The author calculated candidates’ strength by comparing them with their district’s presidential vote share.

**3**. Further emphasizes the argument and summarizes findings: Though most House candidates performed according to presidential partisanship, many of the strongest exceptions were moderate candidates.

**4.** Introduces a visual representation of the data: The visual displays top 10 strongest Democratic candidates in 2020, measured by difference in margins between presidential and House candidates.

**5.** Explains one of the data points: Though Democratic Rep. Collin Peterson did not win his seat in 2020, he was still the strongest Democratic House candidate.

**6.** Explains some of the other data points: Reps. Henry Cuellar, Ed Case, Jared Golden, and former Rep. Anthony Brindisi are all Democrat moderates with strong performances in 2020.

**7.** Introduces another visual representation of the data: The visual displays top 10 strongest Republican candidates in 2020, measured by difference in margins between presidential and House candidates.

**8.** Explains some of the data points: Reps. John Katko, Brian Fitzpatrick, Dan Newhouse, and Adam Kinzinger are all Republican moderates with strong 2020 performances.

**9.** Explains some unique data points: The Republican moderates in Minnesota 5th and Maryland 7th districts, with Democratic incumbents, had unusually strong performances.

**10.** Explores a possible explanation for the unique data: The apparent strength in performance of Republican moderates in Minnesota 5th and Maryland 7th may be due to their opponents’ weaknesses.

**11.** Further explores a possible explanation for the unique data: Rep. Ilhan Omar may have lost because of factors like national unfavorability or scandal, not because of Rep. Lacy Johnson’s strength as a moderate candidate.

**12.** Introduces caveats but then reasserts argument: Though factors like opponent quality can determine candidate quality, and some radical Reps. were strong candidates, the general rule is that moderate candidates had the strongest performances in 2020.

**c.** I like that this article was short and to the point, while also supporting the author’s main argument sufficiently. I also liked how the article was organized in a repetitive format, where the first few paragraphs discussed the strongest Democrat House candidates, and the next few paragraphs discussed the strongest Republican House candidates. I would probably break the paragraphs into smaller, more digestible chunks; for instance, paragraphs **8** and **9** were large blocks of text that could be split into two or three separate paragraphs.