

DSO 545: Statistical Computing and Data Visualization

Homework 05

Due Date: Friday November 8, 2019 (EOD at 11:59pm)

Instructions

- Use Python to answer all questions (Feel free to use any IDE or Jupyter notebook)
 - Submit your Python file to blackboard with comments when needed
 - USC won't tolerate any kind of cheating
 - Good luck
-

Which Democrats Are Leading the 2020 Presidential Race?

Each week, the New York Times tracks how the Democratic presidential candidates are doing and who is breaking out of the pack in the historic race for the 2020 nomination.

You can view the latest polls and analysis here:

<https://www.nytimes.com/interactive/2020/us/elections/democratic-polls.html>

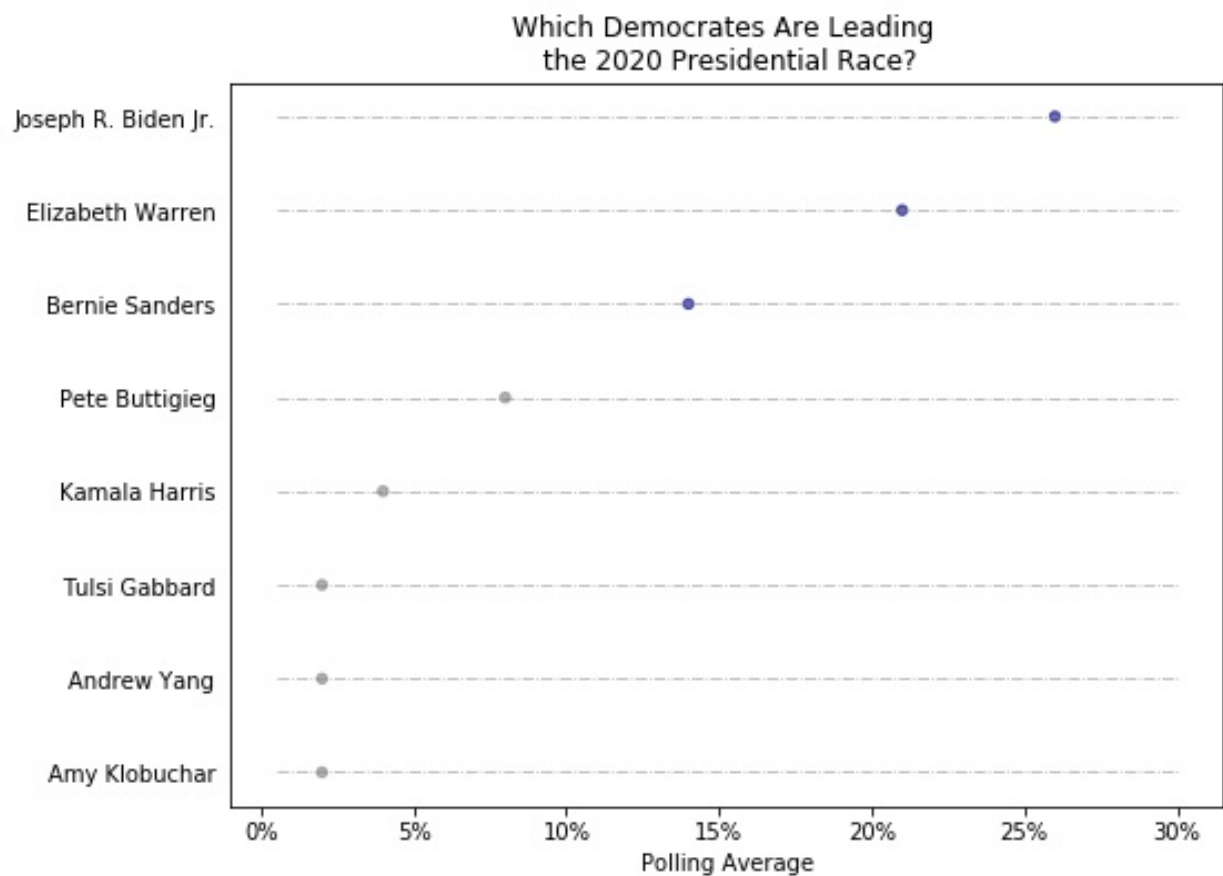
In this homework, you are expected to scrape the latest poll results so we can understand better who is leading the race to be the Democratic party candidate for the next presidential elections.

1. (4 points) Scrape the "Current State of the Race" table (first table) from the NYT webpage.
2. (2 points) Clean the scrapped data to make sure that all the variables ("National Polling Average", "Individual Contributions", "Weekly News Coverage") are numeric.
3. (2 points) Create a dot plot that shows the top candidates who have more than 2% support. (Please note that your graph might not look exactly the same in case NYT updated the polls before next week)

Plot Specification:

- The top three candidates have blue color dots

- The remaining candidates have grey color dots
- The horizontal lines are "dash dotted"



4. (2 points) Create a tree map as follows. The percentage for each candidate is out of the total money received by the candidates who has at least 2% polling average.

Plot Specification:

- The percentages are rounded using the python "round()" function
- The color scheme is 'Spectral'

Who is Leading the Money Race?

