

Final Exam

DATA 101

Due October 25th at 11:59pm

This exam consists of 16 questions. You must hand in an .rmd file as well as a knitted pdf or word document. The rmd of this exact file is available, and you are welcome to use that as a template for your answers.

You are not allowed to share code with classmates. You may ask clarifying questions to TAs/me. If you are stuck on something and can't continue to the next part of the assignment, you can ask a TA or me to give you the code to continue, but do expect to lose a couple of points. Make sure to take the time to add titles, labels, etc to make your graphs look professional.

For the final exam we are going to be working with elections returns from US Senate elections.

Comprehensive results for Senate races have been compiled by the MIT Election Data and Science lab. We're going to do some cleaning to get this into a format that we can use to analyze using a map. The changes we make in the data are going to be cumulative, so you should assume that changes you make to the data in one question (filtering, selecting variables etc) apply to all subsequent questions. For example, in Q6 you will remove Independent candidates from the data. All subsequent questions make use of this filtered data with no Independents.

1. **First, import this data using [this link](#). Download the spreadsheet as a csv and load the election results into R.**
2. **Our first step to clean this data is removing non-substantive columns. Keep only the variables year, state, state_po, stage, candidate, party_detailed, candidatevotes, totalvotes.**
3. **Next, we're going to remove rows with some incomplete data. Remove any rows that have missing data in the "candidate" or "party" columns.**
4. **Next, create a new variable, "party3" which recodes the "party" column into "D" for Democrats, "R" for Republicans, and "I" for all other parties. (Hint: you may want to first create this column so that all rows equal "I". Then use the ifelse() function to recode to "R" if the row represents a Republican and otherwise stays equal to its current value.)**
5. **How many Democrats are in this dataset? How many Republicans? How many Independents?**
6. **Now let's look at the 2-party vote in these data. First, remove the independent candidates from the data. Next, remove all the rows where "stage" is not equal to "gen". This ensures that we only get results from the general election.**
7. **How many races were contested between more than two candidates? Which state had the most of these races?**
8. **For Democratic and Republican candidates create a figure that displays year on the x-axis and each candidate's percent of the vote on the y-axis. Be sure to color code each candidate by their respective party. Add two lines – one for each party – that represents the trend in that parties' support overtime.**

9. Let's take a look at the races from 2012. Filter your dataset so that it only contains the results for 2012, and only the columns year, state, party3, and the candidate percent you calculated in the previous question. Reshape this data so that there is only one row per state, and two columns that represent the percent of the vote won by the Republican candidate and the percent of the vote won by the Democratic candidate. Note that you will not have 50 rows because not all states have a Senate election in an election year.
10. Create a variable "demwin" that records if the Democrat received a higher vote share than the Republican in each race in 2012.
11. Create a variable "demdiff" that records the difference between the Democratic and Republican share of the vote in each race in 2012.
12. Next, we're going to do some analysis to map this data. Load in the state-level mapping data that we've worked with from the package mapdata
13. Join the 2012 Senate election data to this mapping data. Be cautious about the format of the state names!
14. Create a map that shows the winner of each Senate contest in 2012, with Democrats in blue and Republicans in red. If there was no Senate contest in a state (or if a party other than Democrats or Republicans won the seat), leave the state blank.
15. Create a map that shades each state by the Democratic vote difference you created above. Again, If there was no Senate contest in a state (or if a party other than Democrats or Republicans won the seat), leave the state blank.