Problem Set 6: Presidential Election Results, Newspaper Coverage

Due 5/7

This problem set consists of two questions that are unrelated to each other. Submit a short writeup of your answers as well as your R code (in a separate file) on Blackboard.

Question 1

First, we'll create maps of the results of three Presidential elections at the state level. The data file is available in CSV format as `elections.csv`. Each row of the data set represents the distribution of votes in that year's presidential election from each state in the United States. The table below presents the names and descriptions of variables in this data set.

Name	Description
`state`	Full name of 48 states (excluding Alaska, Hawaii, and the District of Columbia)
`year`	Election year
`rep`	Popular votes for the Republican candidate
`dem`	Popular votes for the Democratic candidate
`other`	Popular votes for other candidates

Begin with the 1960 election. Visualize the state-level outcome by coloring states based on the two-party vote share. The color should range from pure blue (100% Democratic) to pure red (100% Republican) using the RGB color scheme. Use the `state` database in the `maps` package (refer to the code from Class 21).

Then, also plot the state-level outcomes for the 1984 and 2012 elections. Briefly comment on how the geography of election outcomes has changed over time (which regions/states became more/less Democratic/Republican).

Question 2

Second, we'll analyze data from newspapers across the country to see what topics they cover and how those topics are related to their ideological bias. The data come from 2005. You will need to load the R packages `wordcloud` and `slam`.

You will use two data sources for this analysis. The first, `dtm`, is a document term matrix with one row per newspaper, containing 1000 phrases – stemmed and processed – that do the best job of identifying the speaker as a Republican or a Democrat. For example, "living in poverty" is a phrase most frequently spoken by Democrats, while "global war on terror" is a phrase most frequently spoken by Republicans; a phrase like "exchange rate" would not be included in this dataset, as it is used often by members of both parties and is thus a poor indicator of ideology. You can load this matrix using the `load` command.

The second object, `papers.csv`, contains some data on the newspapers on which `dtm` is based. The row names in `dtm` correspond to the `newsid` variable in `papers`. The variables are:

Name	Description
`newsid`	The newspaper ID
`paper`	The newspaper name
`city`	The city in which the newspaper is based
`state`	The state in which the newspaper is based

We will explore the content of some newspapers. First, make a word cloud of the top words (at most 20) of your hometown newspaper, or the newspaper that's closest to where you grew up (refer to the code from Class 22 for help).

Then, look at one newspaper that is considered to be very liberal (The Philadelphia Daily News, Philadelphia, PA) and one that is considered to be very conservative (The Daily Sentinel, Grand Junction, CO). Create a word cloud for each (again at most 20 words). How does their language differ? Do they have anything in common?

Finally, make a word cloud for the top words in all newspapers combined (hint: use `colSums`). What were the biggest topics in the news overall, and how much is this reflected in the top topics in The Philadelphia Daily News and The Daily Sentinel?