Introduction:

* What is your project? You should summarize what you wrote in your twitter identification.
  + Everyone wants to know who the next president is going to be in the United States. What better way than to actually retrieve data from online sources or directly from someone’s social media account. Although we cannot extend this to the population, we can draw conclusions that give us better insight into what our future holds.
  + 1. We are going to mine twitter and potentially other sources for a data in the form of a word text. Utilizing text analysis and other factors that will be outlined in our report, we will attempt to predict the outcome of the 2020 election.
  + 2. We will further attempt to display the results of our analysis in a (hopefully) color coded map of the demography of of sentiment that we hope to glean.
  + 3. Indeed we intend to. However, until we have sufficient programming and analysis, this will be performed at a later date.
  + 4. Without question we will perform a classification and perhaps a random forest analysis of the data that we mine in order to reach our conclusions.
* Why is this interesting and how does it relate to society
  + What is more important than the Presidential Election? This determines the individual who has our future in their hands. Which as we all know, affects almost all aspects of the average American day to day life.
  + There are tons of statisticians and other people who are trying their best to come up with methods for predicting the president of the United States.

Materials:

* Describe your data and how you obtained it. (Was it from a hash tag or account ....)
  + We scraped tweets using #biden & #trump (there are more below in the “other” section) from a very popular social media platform named twitter. Twitter was founded in the beginning of 2006.
  + We then imported this into MySQL, to be able to write R code for statistical analysis. We also imported the data into Tableau in order to create visualizations.
  + Electoral votes per state by political party.
* What population does your data represent and can it be extended to the entire population of the USA? Explain why or why not?
  + No this research/conclusions cannot be extended to the entire population because we are retrieving data from a social media platform in which not everyone has. Even if the an individual has twitter, they may not be into politics and writing political hashtags, which our data is based on.
    - 44% of 18–24 year olds use Twitter. 31% of 25–30 year olds use Twitter. 26% of 30–49 year olds use Twitter. 17% of 50–64 year olds use Twitter.

Methods

* What statistical methods are you using. Explain in some detail and very briefly
  + Sentiment Analysis
  + Wordcloud
  + Data:
    - Glm
    - Lda
    - Classification: Random Forest
* What are your findings
* Present the results of your analysis.

Discussion and conclusion

* interpret your findings (this should be detailed). Is there a word that differentiates your outcome that you are model? What does that word represent?
* How sure can you be of your findings?
* Describe either the ideas/words or locations that appear to mean regarding the output variable.
* What does this analysis say about society and your demographic?
* What is your conclusion?

Other:

Other hastags:

* #trump
* #republican
* #donaldtrump
* #maga
* #teamtrump
* #trump2020
* #biden
* #democrat
* #joebiden
* #teambiden
* #biden2020