For project 2 we propose to focus on creating descriptive statistics and visualizations from Political TV Ad Data and additional political demographic datasets. We hope to detail what ads are being aired by region, time, party and current election results. Parameters we would like the unit of analysis to be state and only presidential primary election data, excluding senate and house.

Owen would like to focus on the relationship between television advertisements and receiving votes. This will be done by examining the correlation between number of adds run in a state by groups supporting a particular candidate and the percentage of votes that candidate received. This will be further broken down by the various types of ads run to reach a more detailed level of analysis. In the dataset, an ad will count as supporting a candidate regardless of whether it is run by the campaign itself or by an outside organization. Once the overall correlation between ads and votes has been calculated, the ads dataset will be broken down by the following categories: message, program type, time of day, sponsor type, and state.

While Elizabeth would like to focus on creating visualizations to detail how data from the Political TV Ad data and County Facts data clusters over a map of the United States. The visualizations should support all analyses completed by Owen and will additionally focus on density of each cluster and the states and/or overlapping states. Elizabeth also hopes to provide visualizations showing bucketed groups of time that ads will be displayed taking into consideration time zones and day of the week and corresponding descriptive statistics, such as min, max, sd,… for each visualization. For any additional measures made a data glossary will be provided and a data dictionary will be made for the final datasets used.

|  |
| --- |
| Who are the people that put up the advertisement and do they support or oppose that person? |
|  |

Links to additional datasets and packages with small description of dataset and how it will be used in the analysis:

1. Geopy
   * 1. Geopy has a library with coordinates for counties and states within the US and additional datasets for coordinates within the world
     2. <https://github.com/geopy/geopy>
2. 2016-US-Election
   * 1. <https://www.kaggle.com/benhamner/2016-us-election>
3. Bokeh
   * 1. A visualization library
     2. <http://bokeh.pydata.org/en/latest/>
4. Matplotlib
5. D3
   * 1. D3py
     2. Elizabeth hopes to use the d3py module to generate xml pages, they will be added to the report and the link will be provided

Please visit our repo for further information and to review each dataset, developing data dictionaries and glossaries.

Link: <https://github.com/CHABOBO/project_ad_el_archive>