**­­­**

**Statistical Methods for Data Science**

Spring 2017

Mini Project 2

Contributing members:

**Akhilesh Kumar Kagalvadi Chinnaswamy (axk167131)**

**Vidya Sri Mani (vxm163230)**

**Contribution of Group Members:**

We worked on the problems together trying to find different options to find the results and worked together to compile the report.

Exercise 1 (10 points): (a) Use R to make a map of state-by-state percentage vote share of President Trump in the 2016 US presidential election.

The data are stored in a CSV file on eLearning.

Requirement

Be sure to include Alaska and Hawaii in your map. Get the color scheme in your map to look like the one shown at http://projects.fivethirtyeight.com/2016-election-forecast/?ex\_cid=rrpromo (for an unrelated dataset). It uses shades of red for Donald Trump and shades of blue for Hillary Clinton. (b) What does the map show? Justify your conclusions.

The map plots the difference or margin of Trump % votes minus the Clinton % votes. The range of this attribute is around [-86.8 46.3 ], The graph is plot with more blue color to Hillary and more orange color to Trump. The range of -90 to 90 is mapped as [-90 to 0] are blue with Clinton as. White [0] and Orange for Trump with margin [0 to 90].

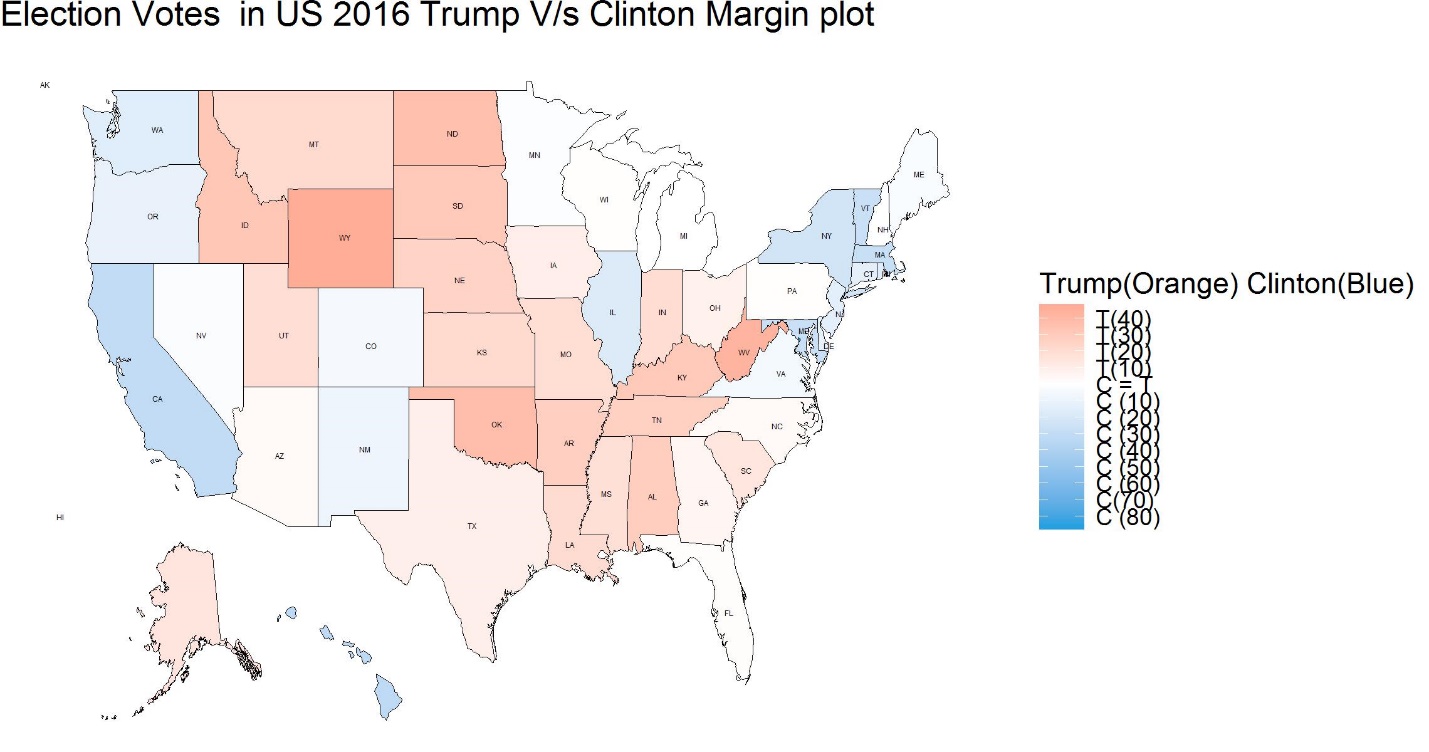
**The following are the observations:**

1. The Hillary had high margins in certain states:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Standard Deviation | Range |
| Margin | 3.66% | 23.82% | [-86.8,46.3] |
| Hillary percentage | 44.72% | 12.17% | [21.9,90.9] |
| Trump Percentage | 48.38% | 11.96% | [4.1,68.6] |

* 1. The range of margin are -86.8 %, 46.3 % which means Hillary had higher margin of leads than Trump.
  2. The mean of margin is 3.66%, which means on an average the Donald Trump had 4% lead over Hillary in all states.
  3. Standard deviation is 23.82% , Which means the change in leads in states vary about 23 % from the mean.
  4. Trump has of 48% average votes over Clinton 44.72% votes.
  5. Standard deviations of both Trump and Clinton votes are equal.
  6. Min vote that Clinton has in any state is 22% over Trump’s 4.1 % and Maximum votes in
  7. any state Clinton has 90.9% over Trump’s 68.6%.

**Observation from the Graph:**



The plot uses two colors to show Votes for Trump and Clinton. While Trump is represented by Orange Clinton is represented by Blue. A deeper shade of color shows a greater margin or votes for the candidate.

For Instance,

* CA state is represented with Deep blue. This shows their high support towards Clinton. From the data CA has around 86% lead.
* Similarly states like WY, OK, WV, ND have higher leads for Trump.
* Mildly Colored states represent not so strong support towards the candidates. However, if the number of votes for Clinton exceed that of Trumps, blue is used, otherwise red.
* States like OR, CO, NM, MN, NJ, CTRI have medium leads to Hillary.
* Similarly states like MT, UT, Alaska, KS, IA, MO, AR, AL, TN, SC have good medium leads to Trump.
* The states like MI, PA, NC, FL, NH, AZ have a little neutral white color showing a close competition in this states between the candidates.
* All the mid regions are covered by Orange indicate huge trump and all the states in the Corner are blue supporting Hillary.
* Clinton has huge Margin wins ranging till 80, whereas Trump has ranges of lead with max of 40’s.
* Few states have white colors, indicate the close competition between both candidates was close. States like Florida etc.

**To add the map of Hawaii and Alaska:**

The maps don’t have Hawaii and Alaska data as they are far of in the map. To bring the map below the us, we need to scale the map reduce its latitude and add it to the data. Recently there is a library created called fiftystater has fifty\_states, which provides the map directly containing all the states and Alaska and Hawaii added at the bottom. However the state.center don’t work as they need to be rescaled, this is left as additional exercise in this project.

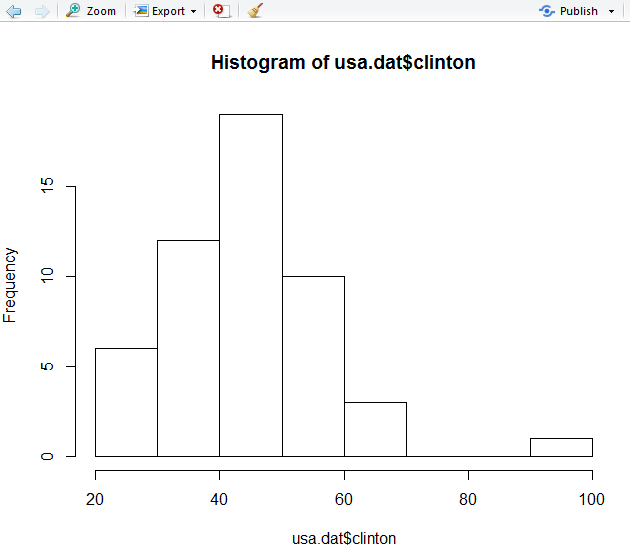
**Additional Analysis:**

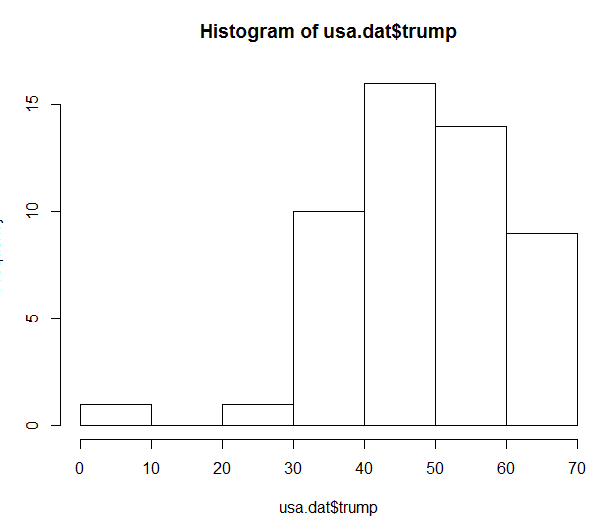
Histogram plot of Clinton, Trump and margin.

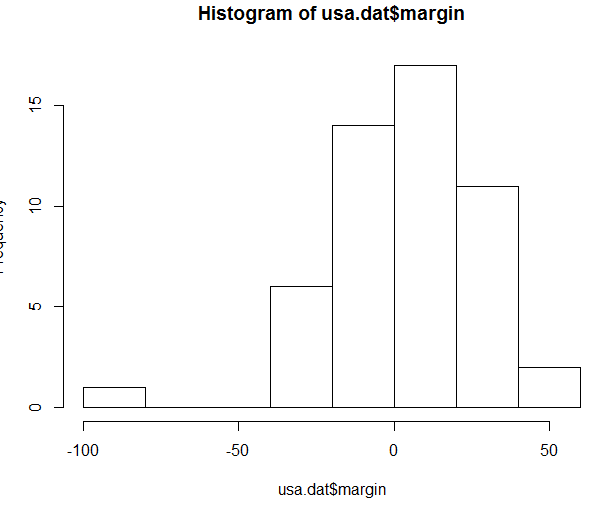
Reveals high frequency votes of 40% and conc votes from 20 -60 and little peaks at 90 for Clinton.

Opp to Clinton, Trump reveals high frequency from 40 to 70 , with others influencing a little change.

Margin has a high frequency near 0, indicating the closer fight,however it is little right sckewed, indicate that



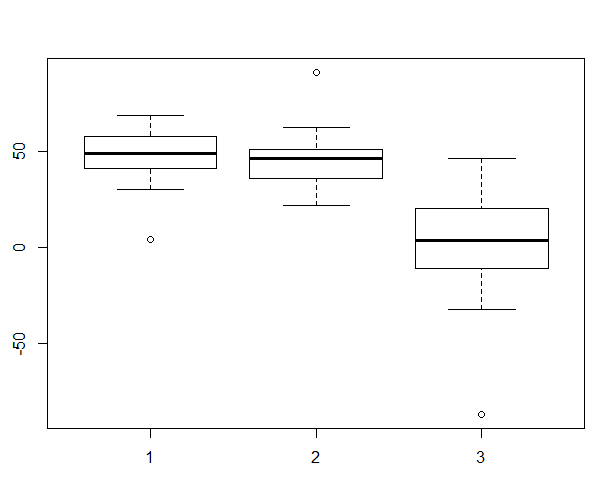




­­

Boxplot

boxplot(usa.dat$trump,usa.dat$clinton,usa.dat$margin)



1.Trump 2. Clinton 3. Margin.  
Trump mean is high over Clinton Mean.

> IQR(usa.dat$trump)

[1] 16.25

> IQR(usa.dat$clinton)

[1] 15.05

> IQR(usa.dat$margin)

[1] 31.35

