

Charlotte Area JMP Users Group Data Challenge Night

Getting started with 2016 primary election data

Source: <https://www.kaggle.com/react/benhamner/2016-us-election>

1. Go to the URL shown above and click the “Download Data” link for the 16.46MB data set (it is found in the 2016 header on the page).
What’s Important in the zip file:
 - a. Primary_results.csv
 - b. County_facts.csv
 - c. The dictionary is helpful, but you don’t need the county shapes. Those are already built into JMP
2. Select one or more of the following questions and create a JMP analysis that answers the question(s). They are roughly grouped by level of difficulty, but feel free to mix and match levels. You can also get creative and look for potential nuggets of insight in the data that are totally unrelated to any of the questions below!
3. Create a page in PowerPoint that “tells the story”, then save the script to your worksheet and to a script file (see instructions for saving scripts and other helpful tips at the end of this document).
4. Send your results to Cheryl at cjohnson6@bankofamerica.com by end of day Friday, March 18th for compilation.

Level 1 Analysis

- Explore county-to county differences in voting behavior at either the party or candidate level.
- Explore county-to-county or state-to-state differences in the demographic data in the County Facts data set.

Level 2 Analysis

- Examine correlations between the various demographic factors overall or on a state-by state basis.
- Examine outlier behavior within the demographic table. Are counties outliers for the same or different demographic reasons?
- Which counties show similar demographics overall? Can you cluster the data into dominant archetypes (e.g., urban white collar vs. rural blue collar)?

Level 3 Analysis

- Is there a demographic that favors one of the candidates?
- Are there any data anomalies with nearly identical demographics and different party affiliations?
- Can any of the county demographic factors be used to predict a dominant candidate or party?
- Are there other data (either at the county or state level) that are predictive? What would you add from the census data?

Hints and Tips

- For navigation tips, steps will be separated by the greater than sign (>) – for example Analyze > Distribution or Graph > Graph Builder
- Some navigation tips will start with Hotspot which indicates you click on the red downward point triangle in the title bar for the platform – for example Hotspot > Script > Local Data Filter
- To save a script once you have the output looking the way you want, save the script to a:
 - Data Table (Hotspot > Script > Save Script to Data Table) then save your table
 - Script Window ((Hotspot > Script > Save Script to Script Window) then save your script file
- Use the field labeled “fips” as the key to join the tables together (Tables > Join)
- Fips can also be dragged into the Map Shape box in Graph Builder to load the county shapes (Graph > Graph Builder)
- Since the primary data is for only four states, you may want to employ the local data filter to focus in on one state at a time (Hotspot > Script > Local Data Filter)
 - Select the variable (or variables) on which you want to filter
 - Highlight one or more levels and watch the data change dynamically
 - Hold down the <Ctrl> key to highlight more than one factor at a time
- Drag any other variable of interest into the color role and data will be displayed as a heat map
- Use the column switcher to rapidly switch between the demographic factors in the County Facts data (Hotspot > Script > Column Switcher)
 - Choose the column to switch
 - Select all the demographic factors when asked for set of factors
 - Click the play (triangle) button on the column switcher
- Use Scatterplot Matrix to compare all pairs of factors against each other (Graph > ScatterPlot Matrix)
- To group counties by similar demographics consider using cluster analysis Analyze>Multivariate>Cluster
 - County Name goes into the Label role
 - All the demographics go into the Y role
 - Drag the Cluster Number Handle (i.e, diamond shape) right to left to increase the number of distinct clusters
 - Save Cluster numbers (Hotspot> Save Clusters)
 - Compare statistics for different factors by cluster