

Assignment 2: Exploratory Data Analysis with Tableau

In this assignment you will be selecting one of the provided datasets and exploring it with Tableau Desktop. Begin by inspecting the available data without visualizing the data, and write down three hypotheses. Next, investigate each of your hypotheses by visualizing relevant variables (including derived variables, if that helps) in multiple ways. Look for correlations, clusters, outliers, or any other patterns. See if you can find evidence for or against each hypothesis. As you explore, take screen captures of plots that show the development of your analysis. Try to find something unexpected in the data. For **only one** of the hypotheses, describe your exploration process using the template below, noting changes and refinements you made to the visualizations as you went along, as well as what worked or didn't work during your exploration process. For the remaining two hypotheses, simply provide screen captures of the first and last visualizations you made.

Note: this assignment will be graded primarily on how thoroughly you explored the data set with various visualizations. Your first and last visualizations should be quite different. In the template, you should show a variety of plot types as well as refinement within individual plots. Use of Tableau is required. Your writing should also be clear.

Assignment due date: 12 noon PST, day of week 9 live session

[Assignment 2 Template File](#)

Data Sets

[County data from the 2000 Presidential Election in Florida](#) (also see the [text file with column details](#))

For each of the 67 Florida counties, the data include the type of voting machine used, the number of columns in the presidential ballot, the undervote, the overvote, and the official certified votes for each of the twelve presidential candidates. Of particular interest are the Buchanan vote in Palm Beach county, and the overvote as a function of voting machine type and number of columns. Data is from the CMU Statistical Data Repository. (Note: be sure you save the text file properly. Some browsers may add extra characters.) [Background on the Florida butterfly ballot](#)

[Online Dating and Relationships](#)

From the Pew Research Center's Internet and American Life Spring Tracking Survey, April 17-May 19, 2013. N=2,252 adults ages 18+. Interviews were conducted in English and Spanish and on landline and cell phones. This is a national survey of dating and relationships in the digital era, the first dedicated study of this subject by the Pew Research Center's Internet Project since 2005. A list of the [Survey questions](#) is available for download as a PDF.

[NBA Player Statistics](#)

Todor Tzolov (MIMS '15) scraped thousands of web pages to assemble this data set about current NBA players. It includes playing statistics as well as draft and contract information as of 68 games into the 2015 season.

[On-Time Performance of Airlines](#)

This data from September 2001 (yes, it includes 9/11) has been extracted from the On-Time Performance data table of the "On-Time" database from the TranStats data library. See the [readme file](#) for detailed metadata.