**Lab 6**

**Advanced Chart Types**

In this lab, you will learn how to:

* Understand the difference between storing data in **wide** and **narrow** format.
* Import additional datasets to Tableau
* Create treemaps, line plots, word clouds, and bubble charts.

**Video**

<http://www.tableausoftware.com/learn/tutorials/on-demand/treemaps-word-clouds-and-bubble-charts-chart-type-8>

**Wide and Narrow Data**

The two main ways to store data in a table are **wide** and **narrow**. The differences are illustrated with two excel files we will be using in this lab.

USstates.xslx – This is an example of **wide** data. The Excel table records demographic data for each US state. Data stored includes state population, GDP from 2010-2013, and geographic region. Each row of the table corresponds to a specific state and each column records a certain attribute of that state.

|  |  |  |
| --- | --- | --- |
| State | GDP 2010 | GDP 2013 |
| Alaska | 49 billion | 51 billion |
| California | 1.9 trillion | 2 trillion |

StateGDP.xlsx – This is an example of **narrow** data. The Excel table records GDP data for each state from 2010-2013. Here, each row represents a yearly measurement. We now have a ‘Year’ column, which records which year that line measures.

|  |  |  |
| --- | --- | --- |
| State | Year | GDPGDP |
| Alaska | 2010 | 49 billion |
| Alaska | 2013 | 51 billion |
| California | 2010 | 1.9 trillion |
| California | 2013 | 2 trillion |

If you want to make a line plot (for example, show how GDP of different states changes over time), you want to use **narrow** data. Keep this in mind when gathering data for your projects. When you have the choice, you typically want to use the **narrow** format. It is possible to use various software packages (such as R) to convert between formats. For details, see your friendly neighborhood TA-s.

For more information about wide and narrow data: <https://en.wikipedia.org/wiki/Wide_and_narrow_data>

**Practice**

* Connect to the US state dataset in Tableau
* Create and put the following graphs in a dashboard called ‘Practice 1’:
  + **Create a treemap of electoral votes in 2010 by state.**
    - Color the treemap by geographical region.
  + **Create another treemap of electoral votes in 2010 by state.**
    - Color the treemap by state population in 2010.
* Create and place the following graph in a dashboard called ‘Practice 2’:
  + Load the excel file StateGDP.xlsx (Data > Connect to Data). This dataset contains GDP data for each US state from 2010-2013.
  + **Create a line plot showing evolution of GDP over time by region.**
* Create and place the following graphs in a dashboard called ‘Practice 3’:
  + Load the text file poe.txt (Data > Connect to Data). This contains the vocabulary used in two of Edgar Allan Poe’s novels, *The Angel of the Odd* and *The Spectacles.*
  + **Create a word cloud with the words from both novels combined** (hint: make a treemap first. This takes a little computing time.)**.**
  + **Create a bubble chart displaying the words from both novels, colored by novel** (drag ‘Novel’ into Rows, or drag ‘Novel’ into Color)**.**

**Assignment**

**Due Date: Feburary 23, 2018, 6:00 pm. Value: 10 points**

* Connect to the USstates.xlsx in Tableau
* Create and place the following graph in a dashboard called ‘Exercise 1’:
  + **Create a treemap of GDP in 2010 by state.**
    - Color the treemap by geographical region.

(2.5 pts)

* Create and place the following three graphs in a dashboard called ‘Exercise 2’:
  + **Create a treemap of GDP in 2010 by state.**
    - Color the treemap by state population.
  + **Create a treemap of population in 2010 by state.**
    - Color the treemap by GDP in 2010.
  + **Create a scatterplot of GDP in 2010 vs state population in 2010.**
  + Which of the two treemaps do you prefer? Would your answer change based on the goal of the analysis? Answer in an annotation on the dashboard.

(2.5 pts)

* Create and place the following graph in a new dashboard called ‘Exercise 3’:
  + Load the excel file StateGDP.xlsx.
  + **Create a line plot showing evolution of GDP over time by region.**
    - Add a reference line showing the average GDP.

(2.5 pts)

* Create and place the following three plots in a new dashboard called ‘Exercise 4’:
  + Load the text file poe.txt. This file has the words from two novels by Edgar Allen Poe.
  + **Create a single word cloud with the words from both novels combined.**
  + **Create a single bubble chart with the words from both novels combined.**
  + **Create a bubble chart displaying the words from both novels, colored by novel (**there are several ways to do this).
  + Which of the plots displays the information best? Discuss advantages and shortcomings of these plots compared with using barplots. Answer in an annotation on the dashboard. (2.5 pts)