Hands-On Lab

Visualization Styles

Lab version: 1.1.0

Last updated: 5/22/2017

Contents

[Overview 3](#_Toc423254481)

[Exercise 1: Excel Charting 3](#_Toc423254482)

[Exercise 2: Using Excel Helpers + HTML Helpers 7](#_Toc423254483)

[Exercise 3: Import Data to Google Maps 11](#_Toc423254484)

Overview

Using some predefined data sets, we’ll explore a few different ways to display data. Note: We will not be driving toward a final visualization here. Instead, we will be exploring the different kinds of charting and visualization that can be used to understand the data in different ways.

We will start using Excel to explore state jobs data. We will then learn how to use the Excel visualization helper file provided as a part of this class. Then we will look at mapping data points using latitude and longitude data on Google Maps. Finally, we’ll cover a simple d3.js sample. This last one is probably only appropriate for people with some javascript or programming experience.

There are more exercises here than can be completed in an hour. If you feel the need to skip ahead to one that interests you more, please feel free to do so.

**Lab Structure**

This lab includes exercises with the following tasks:

* + Exploring chart types in Excel
  + Explore Excel Size and Color Helpers, render data with HTML helper
  + Using Google Maps to map latitude/longitude data points

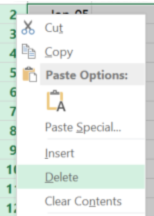
Exercise 1: Excel Charting

In this exercise we will look at a number of ways to chart the state jobs data and select one that most appropriately fits the data we have.

* 1. Open the file “Labs/Lab 3 Assets/Lab 3 – begin.xlsx” (or download it here <https://github.com/matthiasxc/DataVisClass/tree/master/Labs/Lab%203%20Assets> … when you open it up, click “View Raw” to download the file)
  2. In this file we have something of a continuation from our last lab. Instead of just the state data from California and Texas, we have all the state jobs data from all 50 states (and Washington DC). Additionally, we have 4 sheets “Jobs by State”, “Monthly Change”, “Change Since 08”, “% Change Since 08”. Open the “% Change Since 08” sheet.



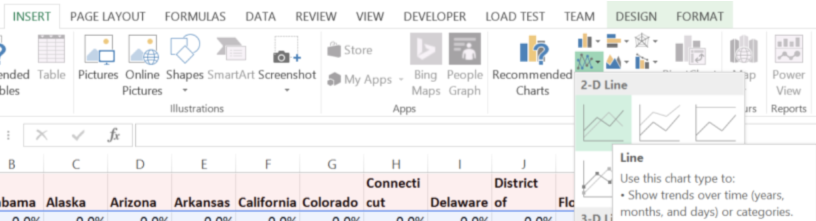
* 1. Click and drag on the rows from 2-37. Right click on them and select “Delete”



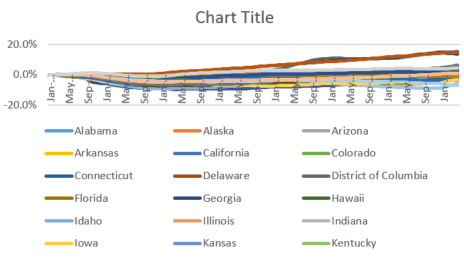
* 1. Select ALL THE DATA from AZ89 to A1.



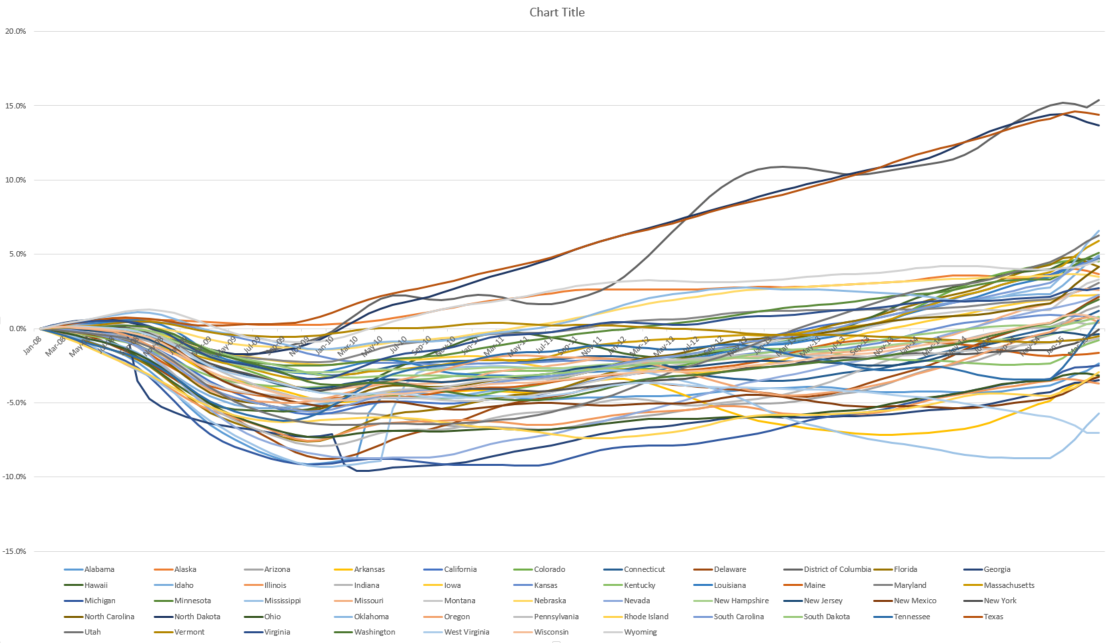
* 1. Click on the “Insert” tab (Mac: “Charts” tab) and click on the line icon. Select 2-D Line.



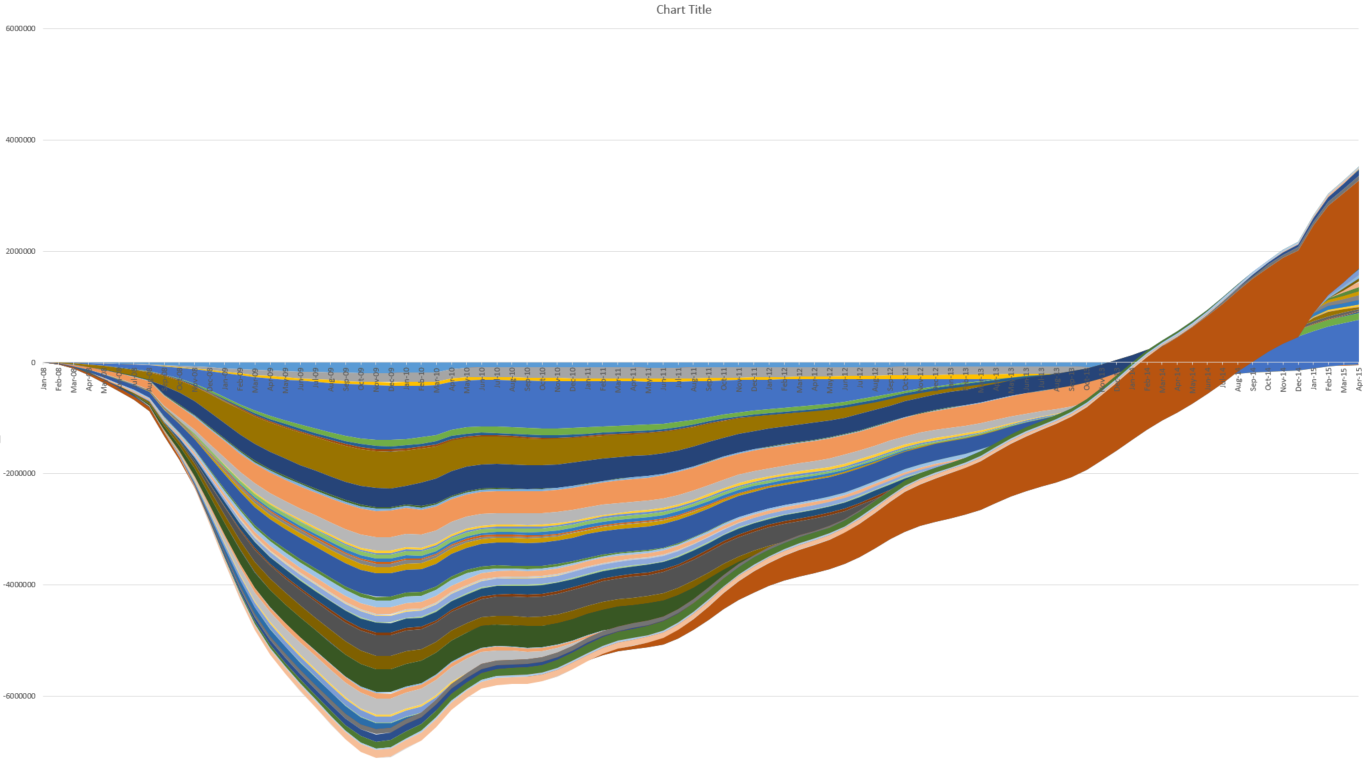
* 1. You should end up with a graph that looks like this.



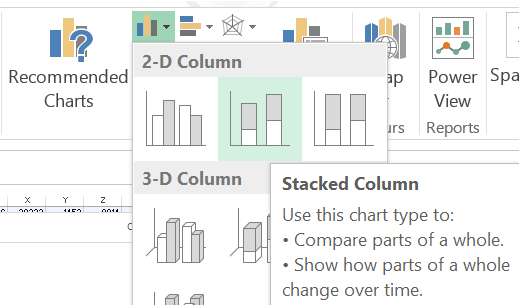
* 1. Enlarge the graph so we can see a little more detail to the data.

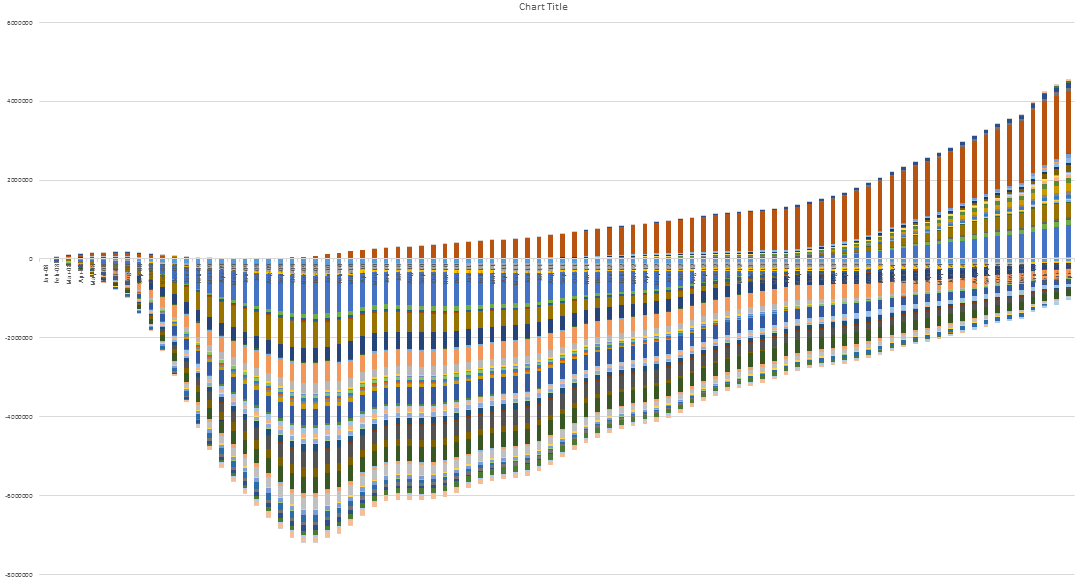


* 1. Now we can see that three states have had enormous growth since the recession: Texas, North Dakota, and Washington DC. (You should be able to see detail on the graph by hovering over the lines.) No other state cracks 7% growth and many states are still struggling to recover from the 2008-2009 recession. But what is the best way to visualize this?
  2. Knowing by intuition (because we’re just so super smart) that DC and North Dakota have very small populations and Texas is huge, let’s try to visualize raw job growth rather than growth as a percentage. Click on the “Change Since 08” sheet and delete rows from 2-37. Select all the data and go to the “Insert” tab and select the “Stacked Area” chart.

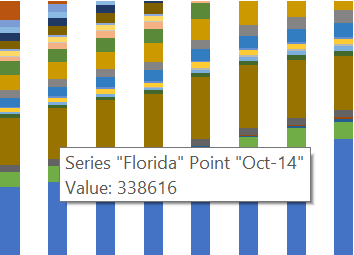


* 1. We can see that something is strange here. The large brown line is Texas, but it seems to be negative for a large portion of this graph. This is because the nature of the stacked area chart requires that the areas maintain their position through the duration of the chart, so it hides the lines that cause the least change in the data. This hides Texas during the low points of the recession and emphasizes it at the high point of the recovery.
  2. Instead, let’s pick a visualization type that separates out the positive states from the negative ones. Select all the data again and go to the “Insert” tab and select “Stacked Column”.



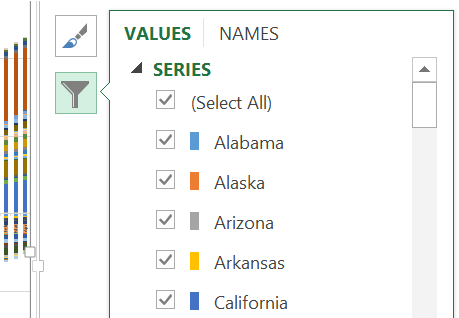


* 1. Even though the columns are separated by white lines (which we could fill in using our graphics application) we get a much better sense of which states saw growth and decline and when. We can zoom into this chart and hover over any point at the chart to see which state it represents and the underlying value.



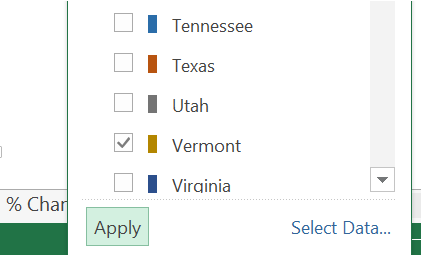
* 1. It’s looking much better. Let’s use the in-chart filtering to explore and limit some of the data. **PC:** Select the chart. Hovering over it, you should see a filter icon appear to the right of the chart. From here we can check just the data we want to see.

Mac: Unfortunately, if you’re using Office 2011, the filter feature isn’t working. We’ll be hiding columns within our initial data set in order to deliver the same effect.



* 1. Let’s look at just the northeastern states. Select Connecticut, Delaware, Maine, Massachusetts, New Hampshire, New York, Pennsylvania, Rhode Island, and Vermont. Press the “Apply” button at the bottom.

Mac: Select the columns for the states that \*aren’t\* listed above. Right click on them and select “Hide”



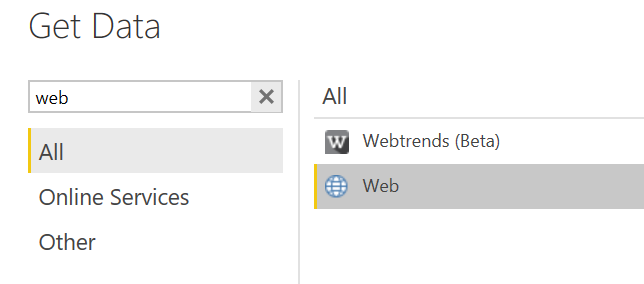
* 1. Now we can see that, regionally, the northeast is still recovering from the recession with Massachusetts really leading the way in job growth in the area.
  2. Nationally, however, it’s still very hard to make out individual states. It might make more sense to view this data on a map, which is what we will do in the next exercise.
  3. If you want to keep working with this data set, here are a things to try:
     1. a) Find the states that have done the best in the recovery. Find the 5 that have created the most jobs (Change Since 08) and the 5 with the highest recovery rate (% Change Since 08).

b) Combine this data with the 2016 election result data (found in the visualization helpers file). You will need to copy values and transpose data in order to do this. Create a scatter plot of the data and add a trendline to it. See if there is a correlation between the 2016 election results and

Exercise 2: Using PowerBI to Create a Choropleth Map

In this exercise we will take our state-by-state jobs data and render it out into a map using the Excel “Visualization Helpers” file and a pre-built HTML page.

1. If you haven’t yet, install Power BI and open it.
2. On the loading screen, click on “Get Data”. This will take you to a screen for selecting one of many kinds of data sources. Type “Web” into the search box and select “Web” from the results. Click “Connect”



1. Power BI is great for importing data from Wikipedia, which is enormously helpful for many kinds of data. Here, we’re going to import the data from the recent presidential election. Do a search for the “2016 Presidential Election” and copy and paste the Wikipedia URL into the Power BI URL box.

For reference, it should be

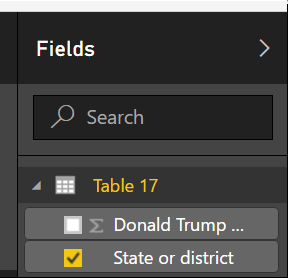
<https://en.wikipedia.org/wiki/United_States_presidential_election,_2016>

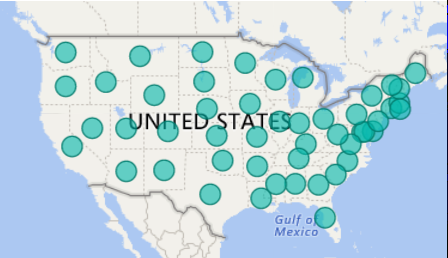
1. From the data navigator, select “Table 17” and press “Edit”. This will take you to a query editor where we will parse the data into something a little more manage-able. Apply the following changes to the data:
   1. Click on the arrow in the second column and make sure only WTA is selected

Electoral 
Sort Ascending 
Sort Descending 
Clear Sort 
Clear Filter 
Remove Empty 
Text Filters 
Sea rch 
(Select All) 
WTA 

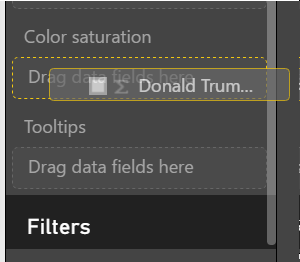
* 1. Drag the “% Donald Trump” column to be next to the “State or district” column. Then select both columns (select one column & then select the other while pressing the “Shift” key) and right-click on the columns. Select “Remove Other Columns” from the menu
  2. Right-click on value 20 that says “Maine (at-large)” and select “Replace Values”. Replace “Maine (at-large)” with “Maine”. Do the same for Nebraska.

1. Click “Close & Apply” to send the data from the Query editor to the main Power BI application.
2. In the “Fields” space, check “State or District” and you should see a map pop up on the left.

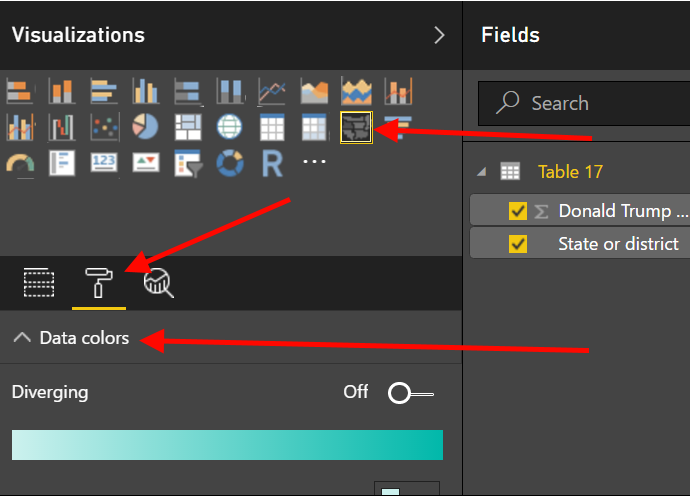




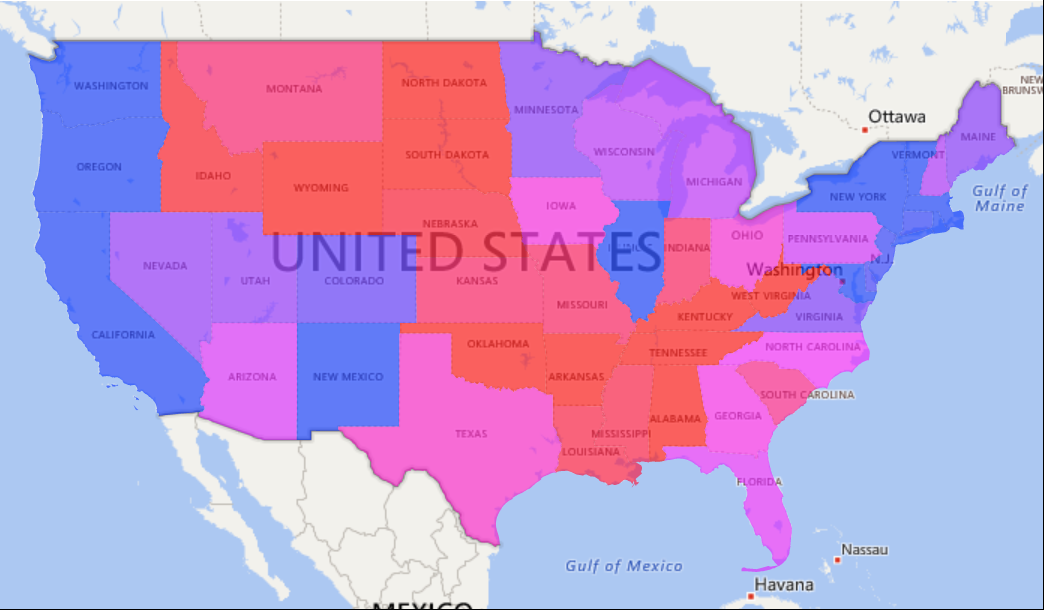
1. Drag the “Donald Trump” column into the “Color Saturation” field.



1. Select the “Filled Map” visualization option, then click on the “Format” icon and open the “Data Colors” option.



1. Change the data colors to a Red-Purple-Blue (I used the “Custom Colors” option) and the Minimum value to “.4” and the Maximum value to “.6” and you should get something like this

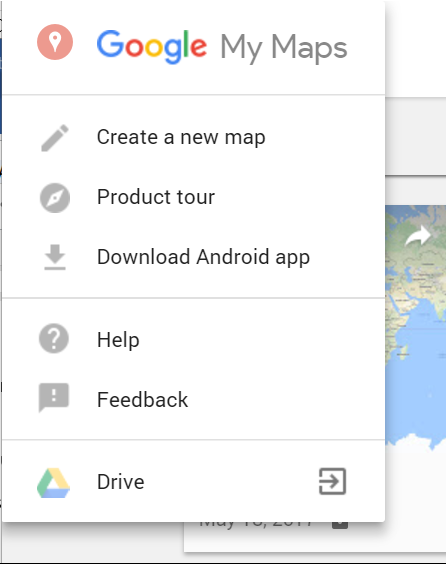


1. Play around with your visualization! Try changing the colors or changing the min-mid-max values.
2. (Optional) Hover over the states to see the detailed information for that state and notice that the information is in decimal form, not percentage form. If you have time right click on the Table 17 data and edit the query. In the “Add Column” tab, select “Custom Column” and see if you can figure out the formula for turning the percentage values of (.6208) into more meaningful values to the average user (62.08).

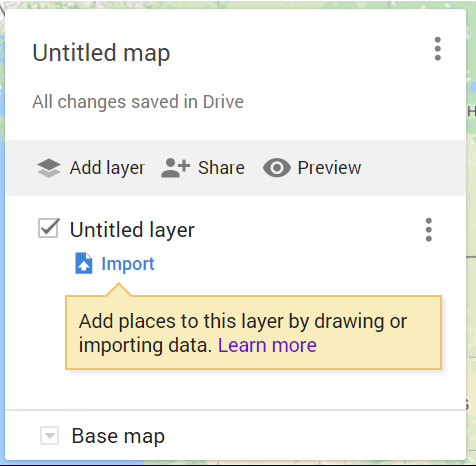
Exercise 3: Import Data to Google Maps

In this exercise we will take a data set with latitude/longitude coordinates and map it onto Google Maps. We will end up with the following visual.

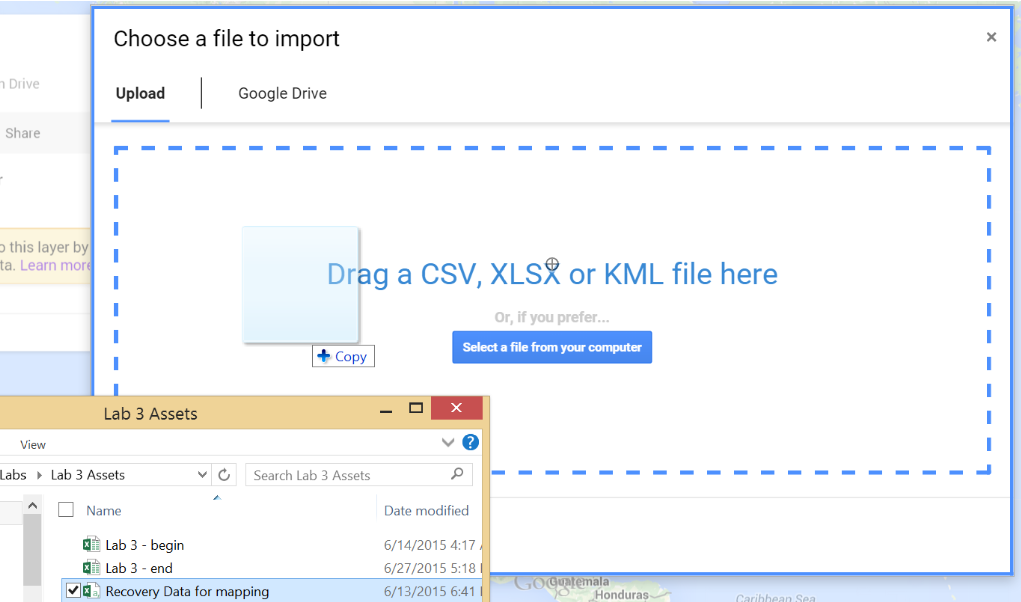
1. In the assets for this lab (found in Labs/Lab 3 Assets/) find and open “Recovery Data For Mapping.csv”
2. This is a small list of the 1,000 largest projects paid for by the ARRA (American Recovery and Reinvestment Act of 2009), popularly known as “The Stimulus”. The original source data (which can be found in the Data folder) is enormous, containing almost 100 fields per row and 615,000 rows of projects accounting for somewhere between 245 billion and 337 billion worth of federal extra-budgetary spending. If you’re feeling brave (and you should by now), feel free to open up that larger file and explore that data.
3. What we care about at the moment is the columns AB and AC. These contain the latitude and longitude of where these awards were spent. We’re going to map this data.
4. In a browser, go to <https://www.google.com/maps/d/> If you don’t have a Google account, sign up for one.
5. Click on the hamburger button in the top left corner and select “Create New Map”



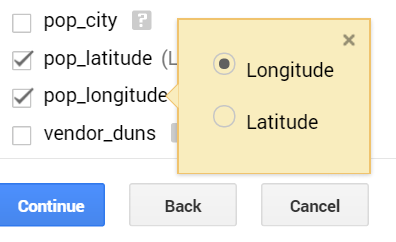
1. Under “Untitled Layer” click “Import”



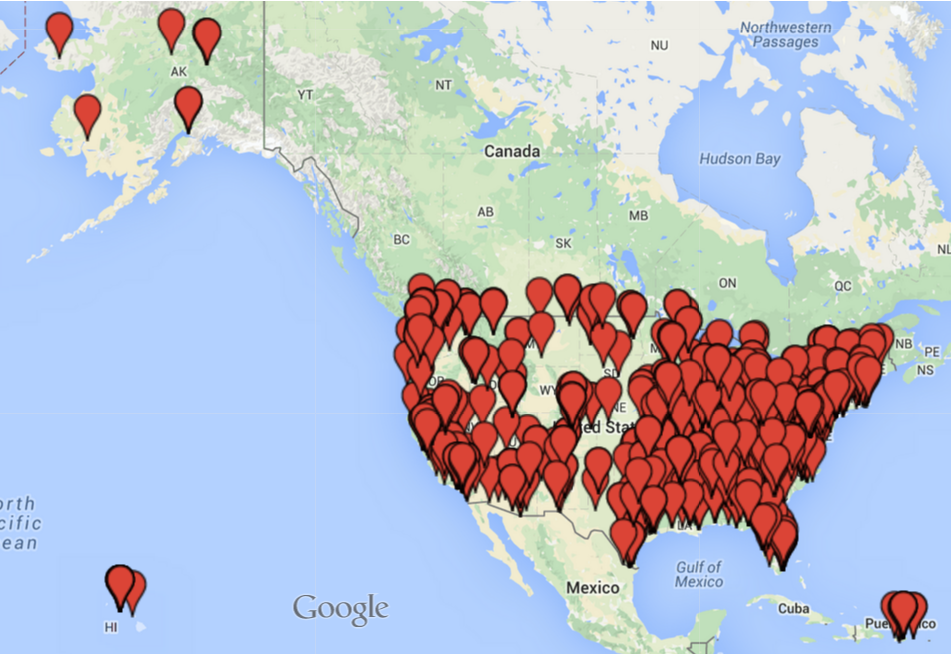
1. This will open up a new tab or window and you’ll see an option to import your data. Click on Import. In what is probably one of the easiest visuals in this training, you merely need to drag the file from the folder you opened it in to the import box on the page.

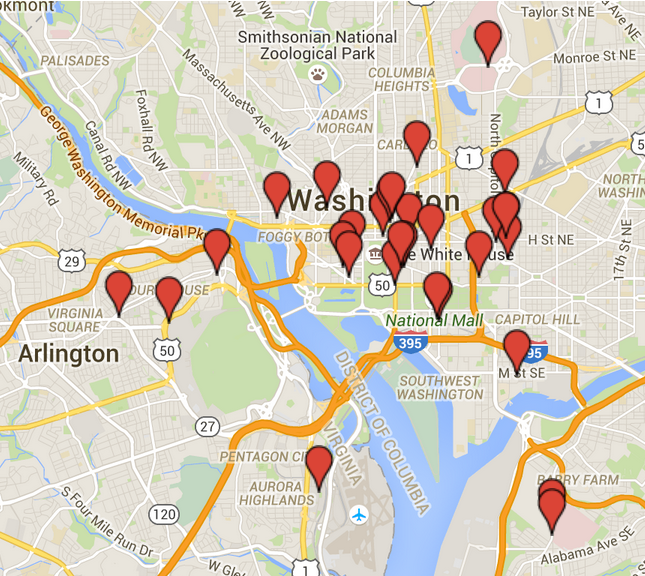


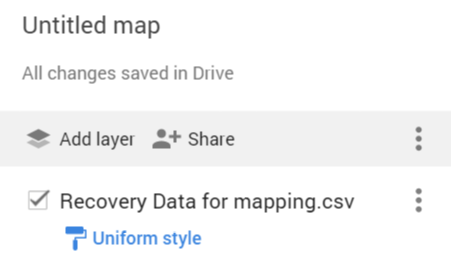
1. Now we need to select the data that will control where our markers are being placed. Scroll to the bottom of the list box and select “pop\_latitude” for the Latitude and “pop\_longitude” for the Longitude. Click “Continue”. Stop me if I’m going too fast.



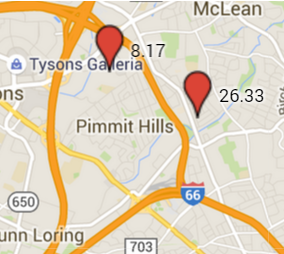
1. Now we need to choose a column to label our markers. Pick from either “award\_amount”, “number\_of\_jobs” or “Cost Per Job”. Google Maps will take care of the rest, rendering it as markers on the map.



1. You will probably find that Google Maps couldn’t render 1 row. If you click through to that data, the reason is that there was no appropriate lat/lon for that row.
2. Zoom in to an area you’re familiar with and examine some of the markers.
3. For one last thing, let’s go to the left hand side bar and click on the “Uniform style” button.



1. In the “Set labels” box, select “number\_of\_jobs”. Now all our labels have a number attached to them. This makes the visualization very messy where zoomed out, but much easier to browse and much more informative when zoomed in.



1. Note that when you click on the marker, you get a number in the label that, while accurate, isn’t very readable. You also get a lot of information in the pop-up that isn’t particularly useful. For further work on your own, try the following exercises:
   1. Delete columns in the “Recovery Data for mapping.csv” file so that you have only the data you want.
   2. Rename the columns to something that is a little more “human readable”
   3. Create a column with the “award\_amount” numbers formatted into a more readable number. Hint: This will actually involve creating 2 columns, using the formula

=TEXT(M2, "#,###,###,###")

And using the “paste values” functionality we’ve practiced.