Hands-On Lab

Visualization Styles

Lab version: 1.0.0

Last updated: 6/28/2015

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 exersices 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

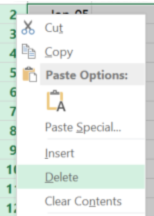
Exercise 1: Excel Charting

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> )
  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.



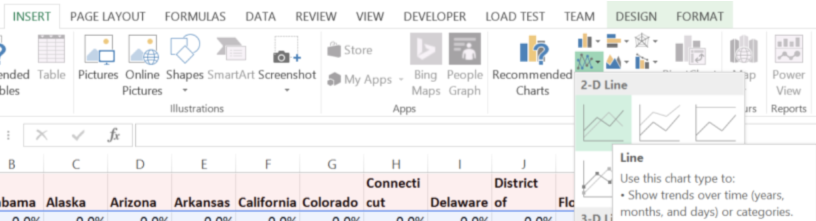
* 1. Click and drag on the rows from 2-37. Right click on them and select “Delete”



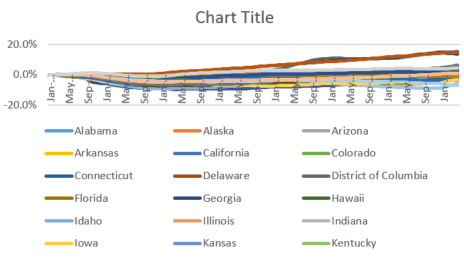
* 1. Select ALL THE DATA from AZ89 to A1.



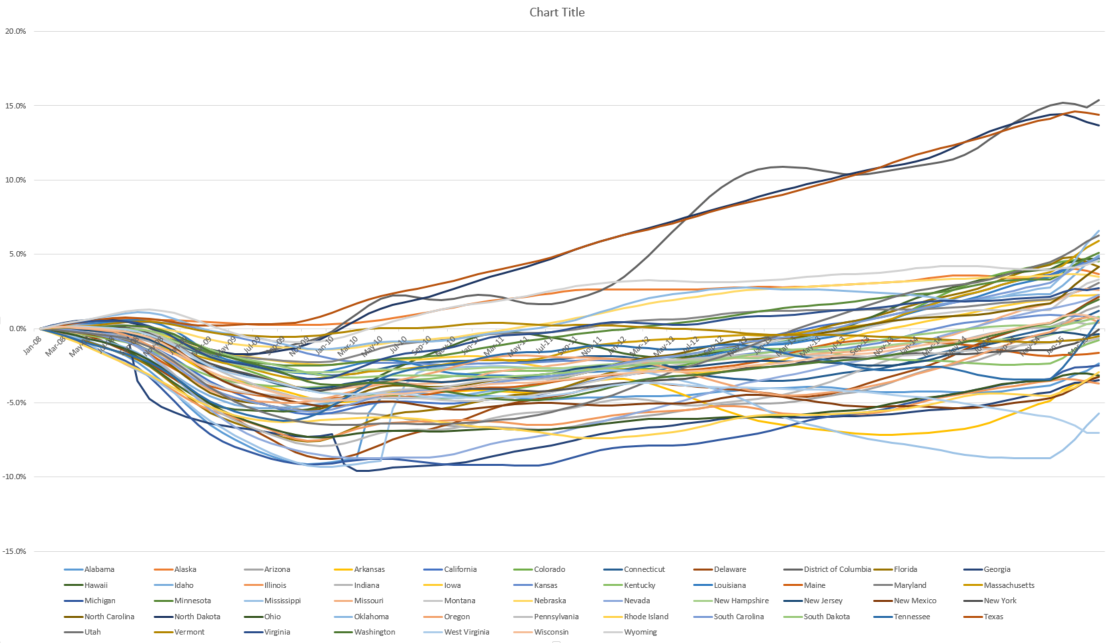
* 1. Click on the “Insert” tab and click on the line icon. Select 2-D Line.



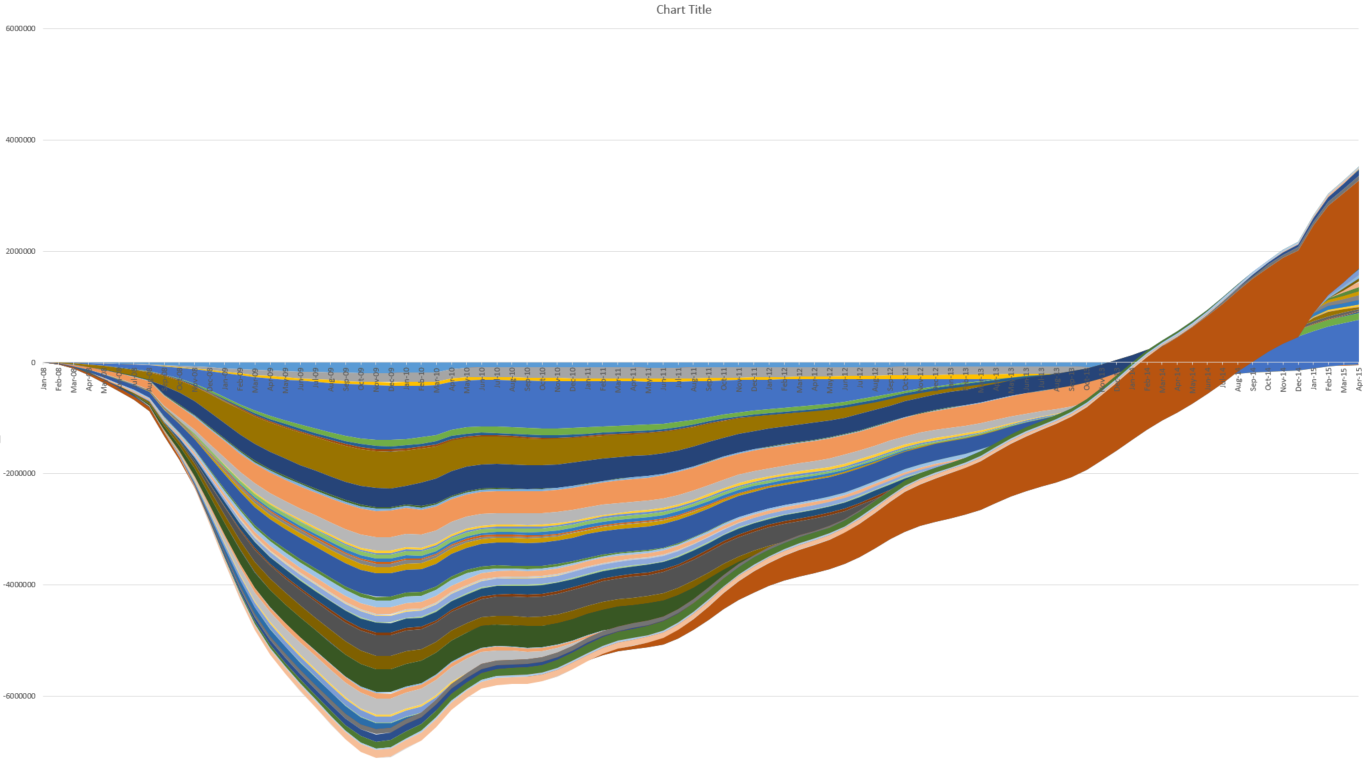
* 1. You should end up with a graph that looks like this.



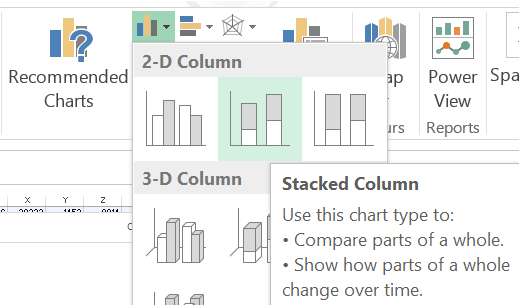
* 1. Enlarge the graph so we can see a little more detail to the data.

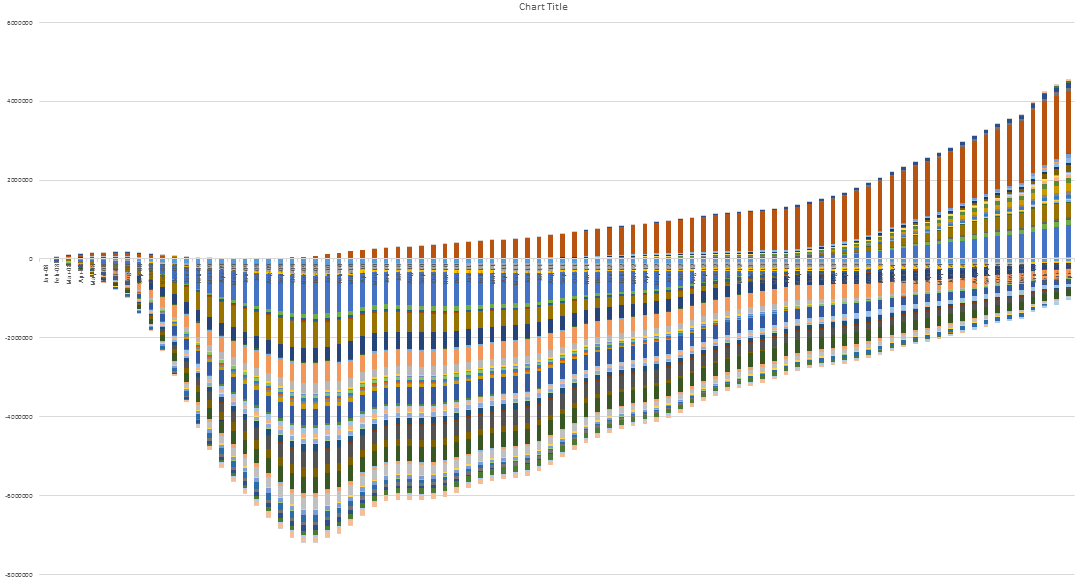


* 1. Now we can see that three states have had enormous growth since the recession: Texas, North Dakota, and Washington DC. No other state cracks 7% growth and many states are still struggling to recover from the 2008-2009 recession. But what is the best way to visualize this?
  2. Knowing by intuition (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.

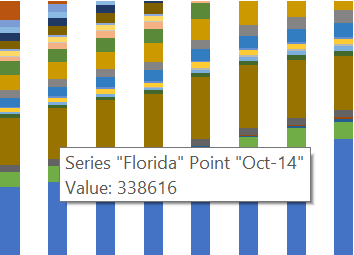


* 1. We can see that something is strange here. The large brown line is Texas, but it seems to be negative for a large portion of this graph. This is because the nature of the stacked area chart requires that the areas maintain their position through the duration of the chart, so it hides the lines that 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”.

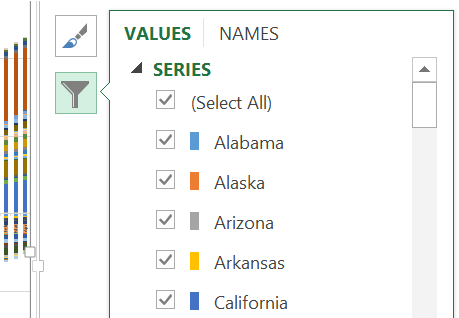




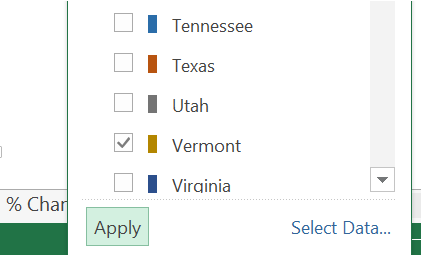
* 1. Even though the columns are separated by white lines (which we could fill in using our graphics application) we get a much better sense of which states saw growth and decline and when. We can zoom into this chart and hover over any point at the chart to see which state it represents and the underlying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



* 1. It’s looking much better. Let’s use the in-chart filtering to explore and limit some of the data. 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.



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



* 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 2008 election result data. 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 2008 election results and

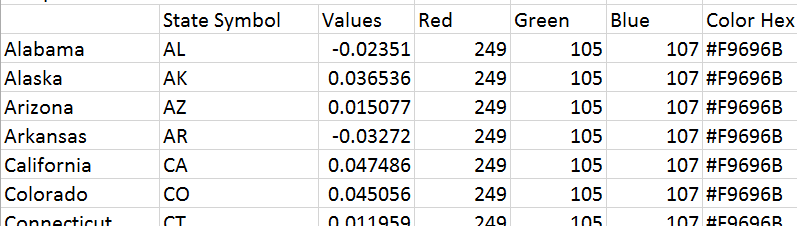
Exercise 2: Using Excel Helpers + HTML Helpers

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.

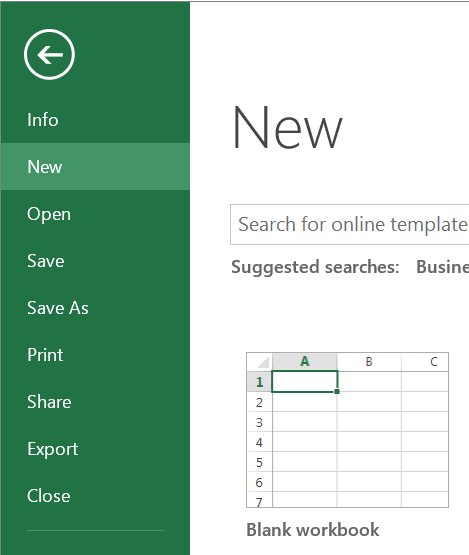
1. Open the folder that you received as part of this class. It should have a “Visualization Helpers” Excel file in it. Open that file.
2. This file has 4 sheets with the calculations necessary to provide the appropriate visualization values for the following scenarios:
   1. Size Calculator – 1 dimensional value-to-size calculation (bar charts)
   2. Area (circles) – 2 dimensional value-to-diameter calculation
   3. Color (2 colors) – input minimum & maximum values & RGB colors and output appropriate RGB and hex color values for a given data value
   4. Color (3 colors) – input minimum, middle, maximum values & colors and output appropriate RGB and hex color values for a given data value
3. Select the “Color (3 colors)” sheet. The preloaded data we’re converting is the percentage of the vote Barack Obama received in any given state in 2008. Therefore, we’re using the following values:
   1. Minimum Value – 40%, Minimum Color – Red - (249, 105, 107)
   2. Middle Value – 50%, Middle Color – Yellow - (255, 235, 132)
   3. Maximum Value – 60%, Maximum Color – Green - (9, 190, 123)
4. Open up the Lab 3 Excel file and open the “% Change Since 08” sheet. Select the values B89 – AZ89 and copy them (ctrl-c or command-c for Mac). Paste them into the row beneath as values by right-clicking on B90 and selecting the “Paste Values” option.

We’re doing this to get just the values because if we transpose the data with the references attached, it will change substantially.

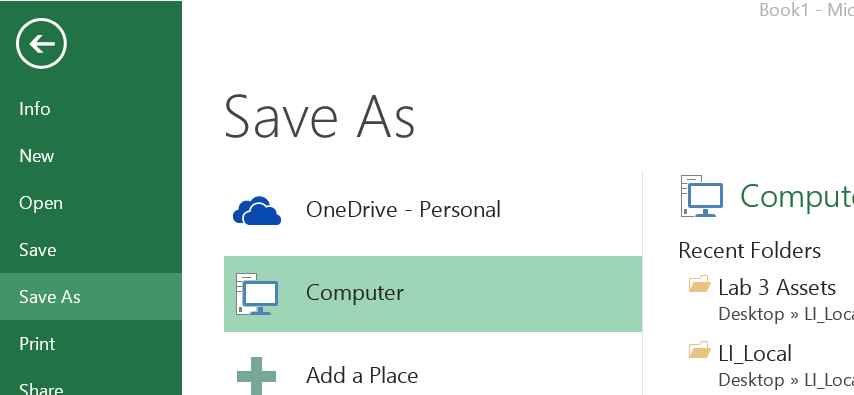
1. Select and copy the values from B90 – AZ90. Go back to the “Visualization Helpers” file. Right click on C11 and select the transpose option from the menu. It should look something like this:



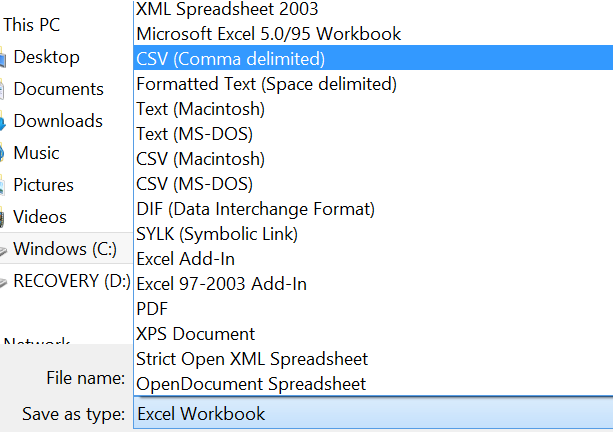
1. You’ll notice all our colors are the same. This is because, while 40% is quite low in a presidential election, it’s quite high when talking about job growth. We need to change the minimum, middle, and maximum values to be right for our data set. Change them to the following values:
   1. B5 = -5%
   2. B6 = 0%
   3. B7 = 5%
2. Now we have a more interesting variation in our color values. We could take these color values and transpose each one by hand into a Photoshop (or equivalent) visualization or into an HTML page. In many cases, that would be the best use of this visualization helper. The data does not \*have\* to be state-by-state data to use this… it can be any values.
3. Create a new Excel file by clicking on “File”. Then click on “New” and select “Blank Workbook”



1. Switch back to the “Visualization Helper” file, make sure you’ve selected the “Color (3 colors)” sheet and copy all the data in this sheet, from A1 to G61. Go back to your new worksheet, select A1 and paste the data in. Click the “File” tab and select “Save As”. Select “Computer” or a recent folder… anything to bring up the “Save As” dialog.



1. Navigate to the root folder of the class files and then go to “Code -> data”. In the “Save as type” drop-down box, select “CSV (Comma delimited)”. Name the file “Job Recovery By State” and save it.

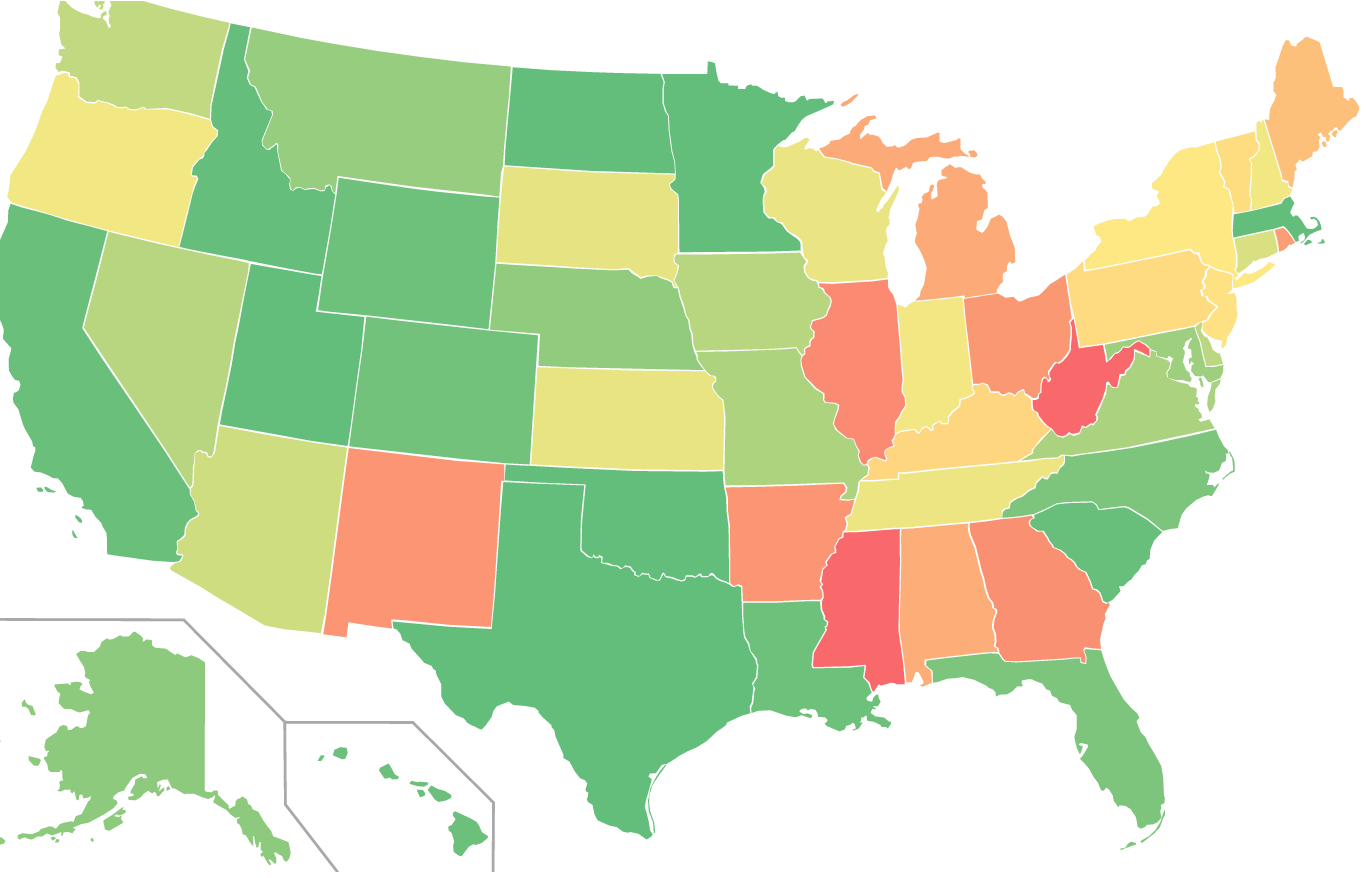


1. This will strip out all formatting and formulas from our spreadsheet and save it as a set of values in which columns are separated by commas and each new line represents a new row. So, for example, we could open our CSV file in a text editing program as see:

*,Red,Green,Blue  
Minimum Color,249,105,107*

…and so on. The CSV format is very common, can be opened with a wide variety of programs and is programmatically easy to parse.

1. In the file system, go to where the class files are, open the Code folder and open the “usMap2ColorExample.html” file in a browser. If you get a warning about blocked content, allow it.
2. You should get the following visual, which is a choropleth map of our state-by-state jobs recovery data.



1. With a little bit of work, you could repeat this process for every year since 2008 and have a simple 7 frame animation showing the Great Recession and subsequent recovery. While this wouldn’t be interactive, it’s a plausible solution for an impressive visual.

This can be used with any state-by-state programmatic data. Please refer back to this tutorial or to the class Google Group if you need help. The visual was created with the following tools

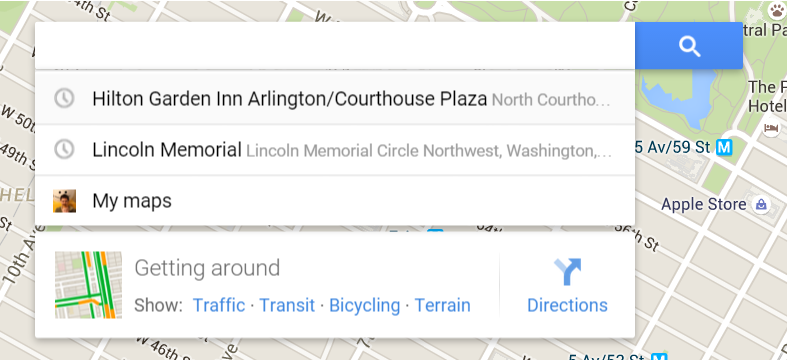
* WikiCommons SVG map of the US
* jquery-2.1.4.js
* jquery.csv-0.71.js
* colorpicker.js

I used Visual Studio Code, a powerful (and free) version of Microsoft’s Visual Studio IDE that is available on every platform.

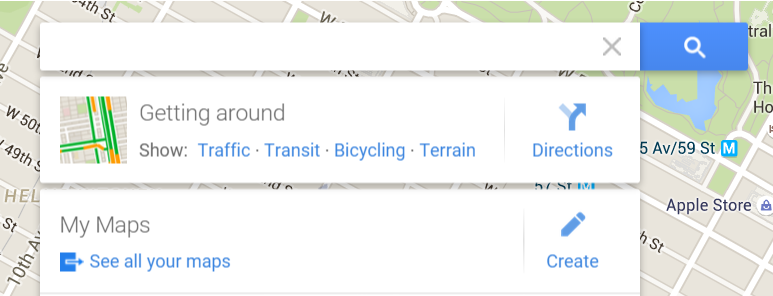
Exercise 3: Import Data to Google Maps

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.

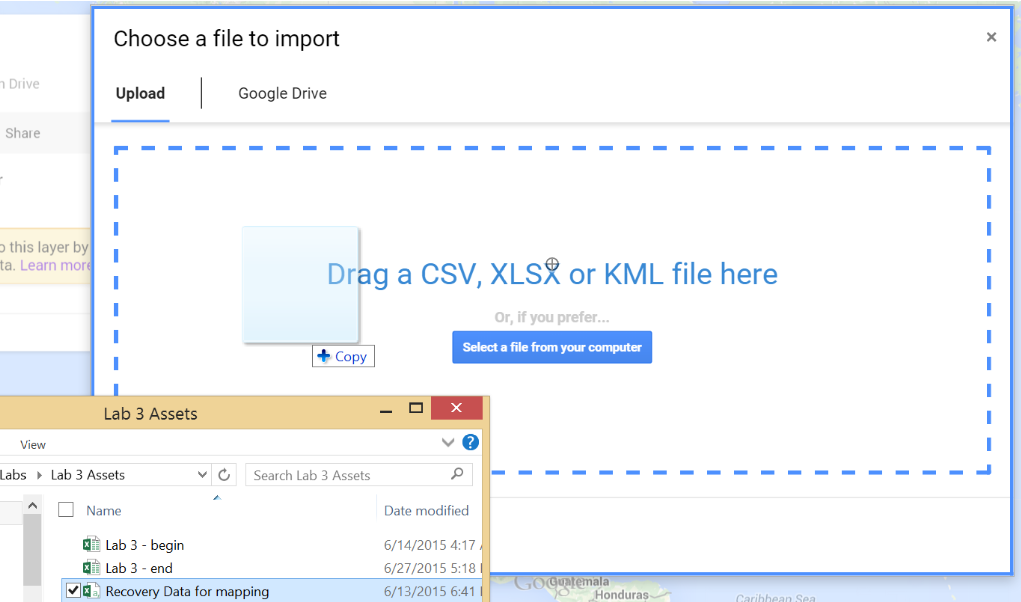
1. In the assets for this lab (found in Labs/Lab 3 Assets/) find and open “Recovery Data For Mapping.csv”
2. This is a small list of the 1,000 largest projects paid for by the ARRA (American Recovery and Reinvestment Act of 2009), popularly known as “The Stimulus”. The original source data (which can 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 <http://maps.google.com> If you don’t have a Google account, sign up for one.
5. Click into the main search box in the upper left of the page and you should see “My Maps” appear.



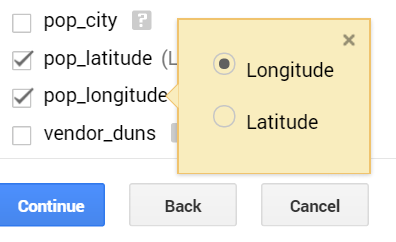
1. Click on it and then click on the “Create” button.



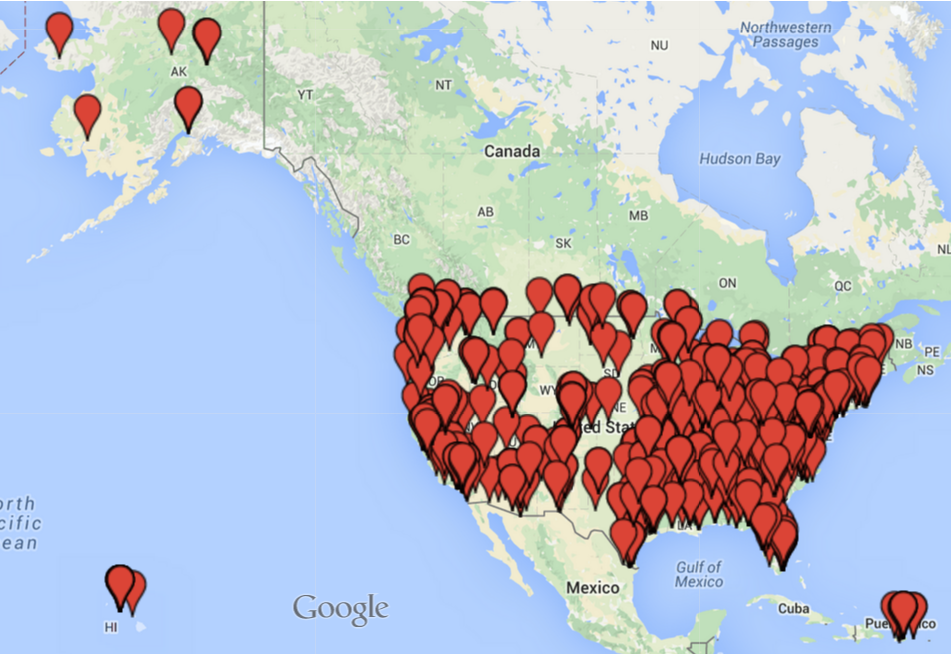
1. This will open up a new tab or window and you’ll see an option to import your data. Click on Import. In what is probably one of the easiest visuals in this training, you merely need to drag the file from the folder you opened it in to the import box on the page.

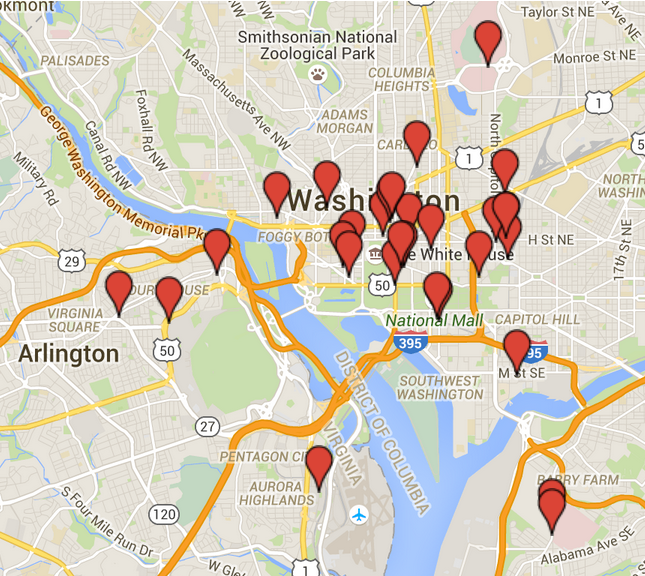


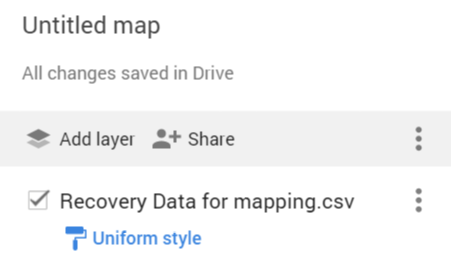
1. Now we need to select the data that will control where our markers are being placed. Scroll to the bottom of the list box and select “pop\_latitude” for the Latitude and “pop\_longitude” for the Longitude. Click “Continue”. Stop me if I’m going too fast.



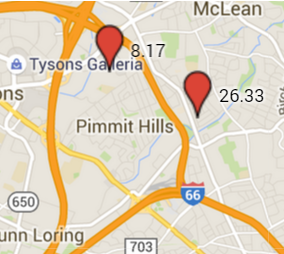
1. Now we need to choose a column to label our markers. Pick from either “award\_amount”, “number\_of\_jobs” or “Cost Per Job”. Google Maps will take care of the rest, rendering it as markers on the map.



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.