Getting Started with Data Visualization I: Tools for Research
Activity Guide

Quick links in this document:

Folder of materials: go.ncsu.edu/startviz

Exercises:

Guiding research questions:
- What patterns in popular baby naming are evident over time in the United States?
- Are there geographical patterns?
- How does the number of occurrences of the most popular names in the past compare to recent decades?

Tableau Public

https://public.tableau.com/s/

- Useful for creating interactive visualizations
- Requires downloading an application and setting up an online account.
- Tableau Public requires saving your visualization to the web
- **For students:** you can get a free download of Tableau Desktop, a fully-featured version of Tableau that allows you to save to your computer. Get it here: http://www.tableau.com/academic/students
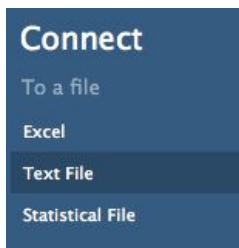
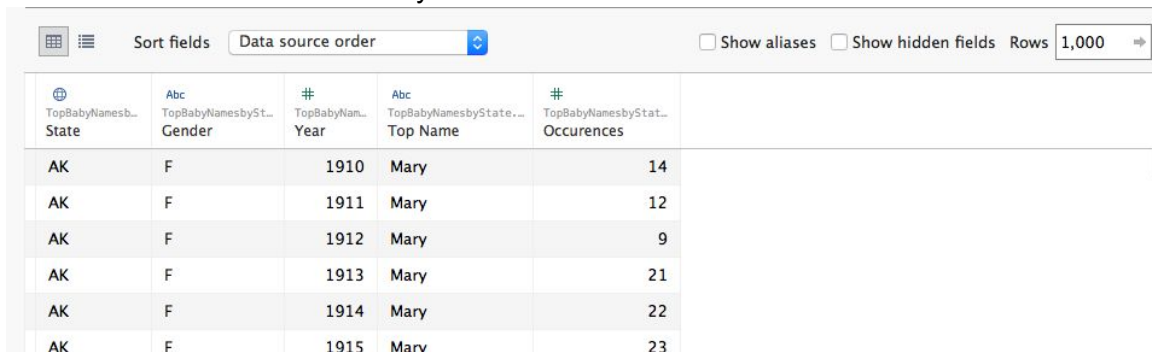Demos of what we're going to make:
Dashboard 1
Dashboard 2 (optional)

Importing Data

1. Download the data: **TopBabyNames.csv** file from **go.ncsu.edu/startviz**

2. Open the .csv file to examine its contents. Then close the file.

3. Open Tableau Public on your computer

4. Under Connect, select "Text File"

Connect
To a file
Excel
Text File
Statistical File

4. Select "TopBabyNames" file from your downloads folder. Click Open

5. Examine the data for accuracy.

| State	Gender	Year	Top Name	Occurences
AK	F	1910	Mary	14
AK	F	1911	Mary	12
AK	F	1912	Mary	9
AK	F	1913	Mary	21
AK	F	1914	Mary	22
AK	F	1915	Mary	23

6. When satisfied that the data is correct, click on "Sheet 1" (orange button at the bottom left)

Exercise 1: Line chart with filter

This chart will show the number of occurrences of the name over time in the data set and will allow the user to filter the chart based on name.

1. Drag Year to columns
2. Drag Occurrences to rows
3. Drag Top Name to the chart and drop it on the white space anywhere in the chart
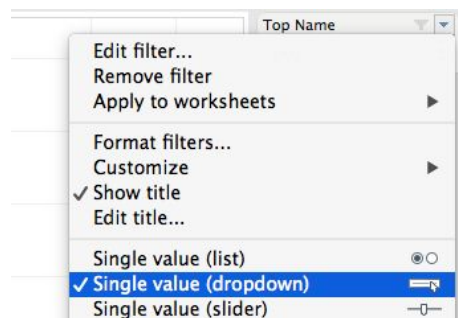
You'll get this jumble of lines:
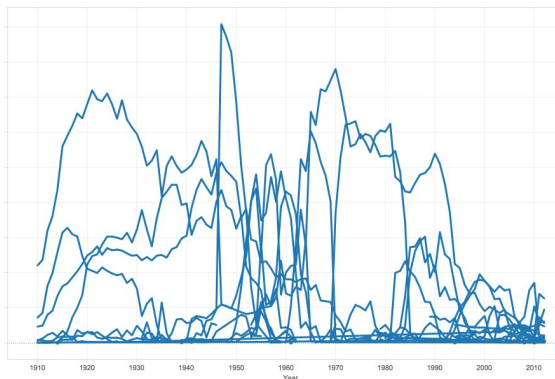
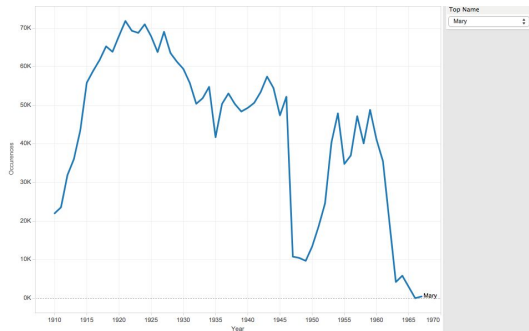4. Drag Top Name to Filters. Click on the Top Name dropdown arrow. Select "Show filter":
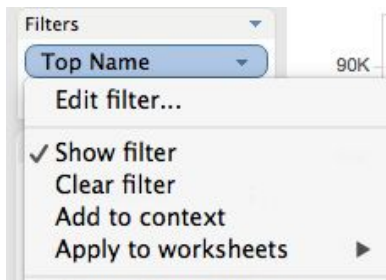
5. A filter should appear on the right side of the workspace. To edit this filter, click on the dropdown arrow at the top of the filter

6. Select Single Value Dropdown

7. Select a name from the dropdown list. The chart should filter out all names but that one.

Exercise 2: Map

This choropleth map will reveal how name popularity varies by state.
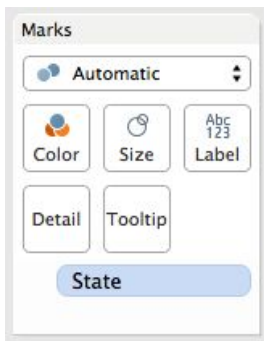
1. Create a new sheet by clicking on the new sheet icon at the bottom of the workspace:

2. Drag Longitude generated to columns
3. Drag Latitude generated to rows
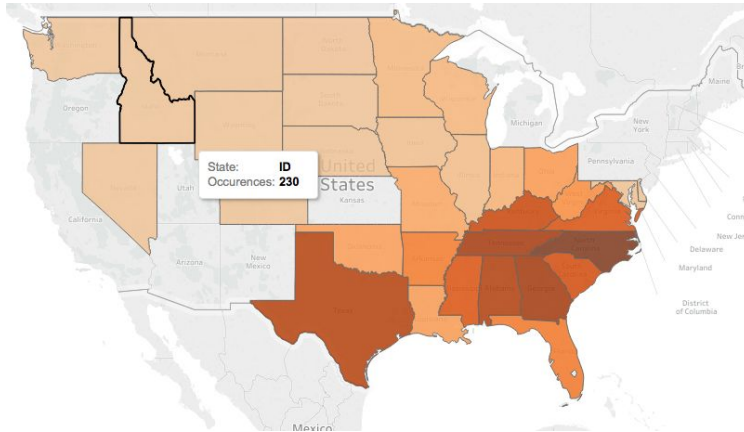4. Drag State to the Marks menu and drop in white space under the buttons:

5. Drag Occurrences onto the Color button

6. Return to Sheet 2, in the filter window, click on the Top Name dropdown. Click Apply to

worksheets… Selected worksheets. Select Sheet 2.

7. Your map should now be filtered based on the filter in the previous sheet.

8. To change map color scale:
> Click on the color button in the Marks menu. Select Edit Colors...
> Click the dropdown to change the color palette to a different color range.

9. Saving individual sheets as images or PDFs
Now you can save your Tableau workbook to the web. Make sure you are logged into your online Tableau account. The workbook will save to your gallery. In the gallery, you can download PDFs or images of individual sheets. An easy alternative is to take a screenshot of a sheet.

Exercise 3: Dashboard with map & line chart

Now, let's bring the two charts we've created together in a dashboard so that viewers can see both sheets in the same place and get more from the data.

1. Create a new dashboard by clicking on the ⊞ icon

2. Set dashboard size to Automatic using the dropdown menu on the left side of the screen



3. Under Objects, Select "Floating"

4. Drag the line chart (Sheet 1) to the Dashboard and place in lower left corner.

5. Under Objects, Select "Tiled"

6. Drag the map (Sheet 2) to the dashboard. The map should take up the entire space.

7. Select the dropdown arrow in the legend (Occurrences") Select "Floating" to make that easier to position.
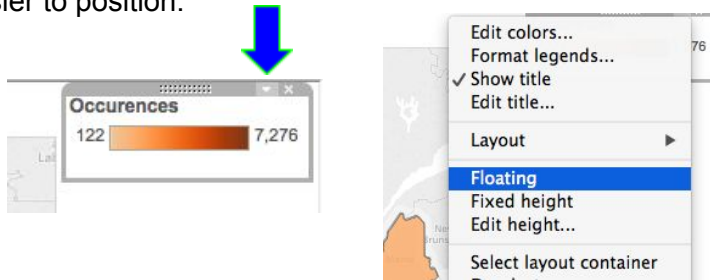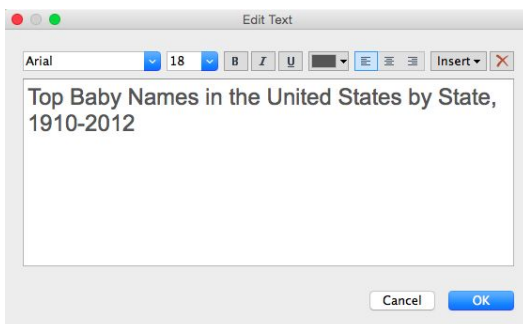

8. Move the legend to the bottom right corner.

9. Hide the map title by double clicking on the title and selecting "Hide title"
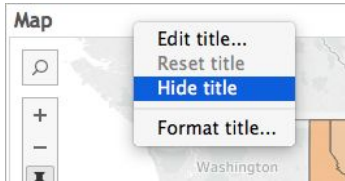
10. Drag Text ( )and drop it above the map (make sure that "Tiled" is selected under New objects")

11. Edit the text box with the new title.

Your dashboard should now have two sheets that are linked to one filter.

Top Baby Names in the United States by State, 1910-2012
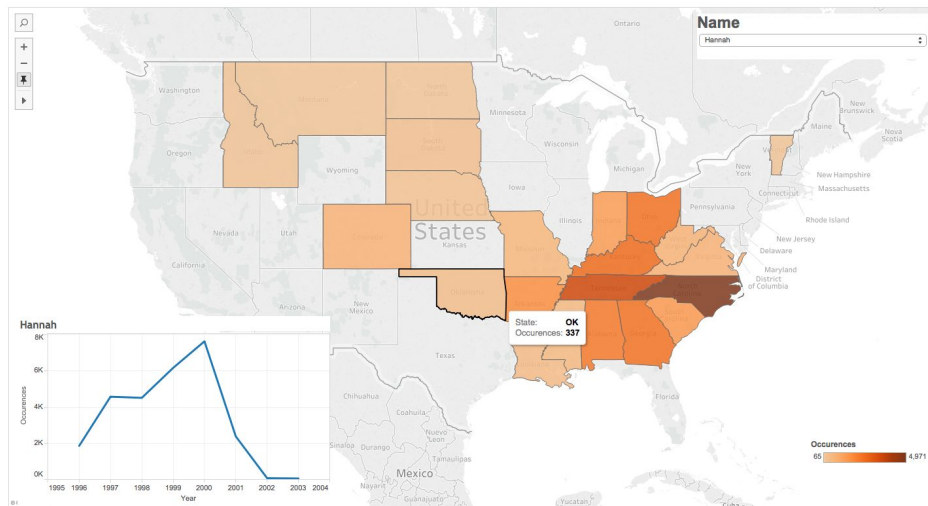
To make your line chart title display the name selected in the dropdown menu,

1. Click on the title
2. Select Insert dropdown button. Select "Top Name" from choices. This will make the top name the title of the chart and will dynamically update based on filtering choices.

Exercise 4: Bar chart challenge: What are the most popular names in the dataset?

Start a new sheet.

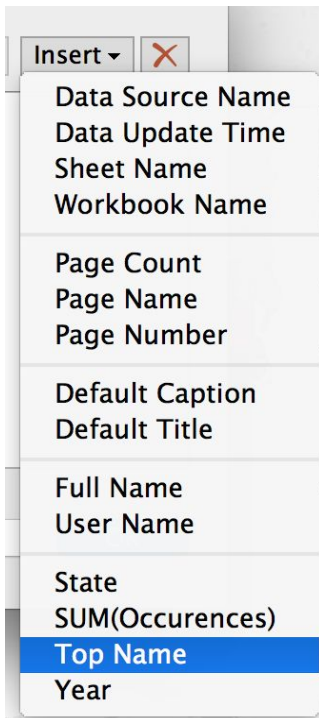Try making a bar chart that shows the number of **occurences** for each **name** in the dataset. Order the bars by size from longest to shortest.

1. To get started: drag **Top Name** to rows

Try to get your chart to look like this example:
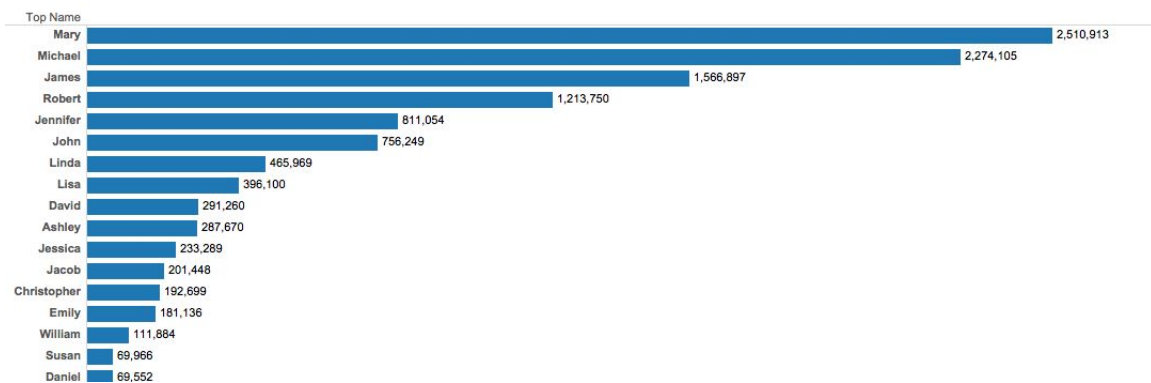
If you make a mistake, you can always go back to previous steps!
Hit CTL+Z on a PC and command +Z on a Mac.

2. Once you have created the chart above, add Gender and Year as filters

Solution to Exercise 4:
Drag & drop:

   Top Names to Rows
   Occurrences to columns
   Occurrences to Label

Click   ⊞↓   to order the bars

Optional Exercises in Tableau

Area graph

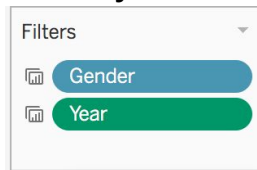An area graph will reveal changes in naming over time in a flow-like effect. The power of this visualization will be in allowing the user to change the range of years and select genders to better see patterns.

1. Drag Year to columns
2. Drag Occurrences to Rows
3. In the Marks menu, change the dropdown to Area

Marks

 ⌃ Area ▼

4. Drag Top Name to the Color button
5. Select "Add All Members"
6. Drag Year to filter. Click on this and select "Show filter"
7. Drag Gender to filter. Click on this and select "Show filter"
8. Select "All" in the Gender filter
9. Drag Top Name to the Label button.

The filters will make analysis easier so there isn't too much data on the screen.
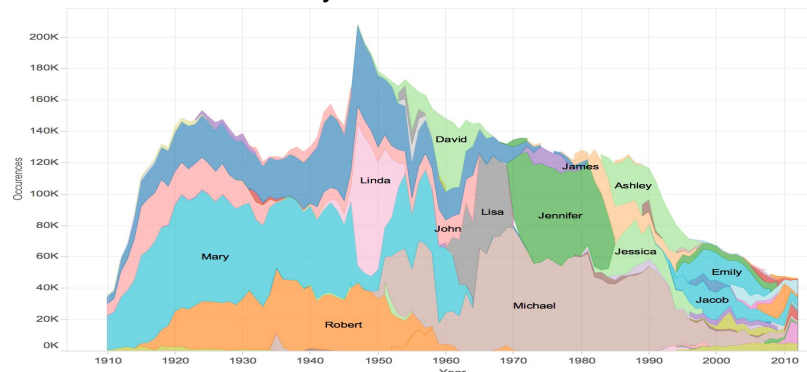
Dashboard 2 Challenge

Try making a dashboard that combines the bar and flow charts where they both update when the user selects filter options. See this Demo as an example.

RAW

http://raw.densitydesign.org/

Raw is an interesting tool for exploring possibilities with data visualization that are more complex than basic charts.

Get Started

Download the data: **TopBabyNames.csv** file from **go.ncsu.edu/startviz**

Open the .csv file

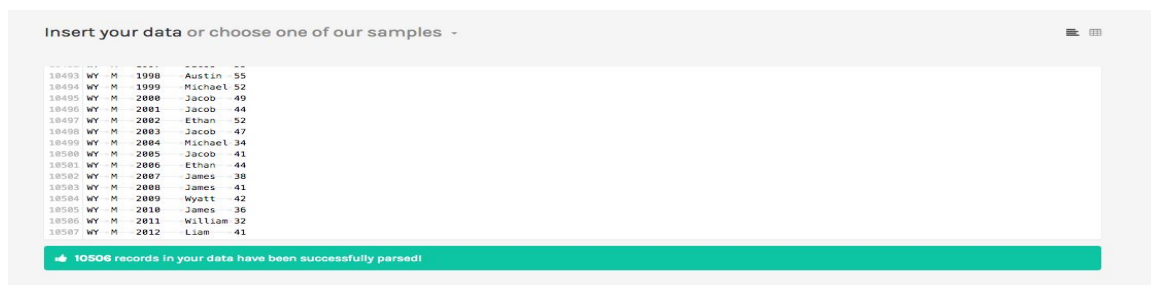**Copy all of your data from the .csv file.**
      -on Mac: Select All: (command A), Copy (command C)
      -on PC: Select All: (control A), Copy (control C)

Open a web browser to the following website: http://raw.densitydesign.org/

Click the Use It Now! Button
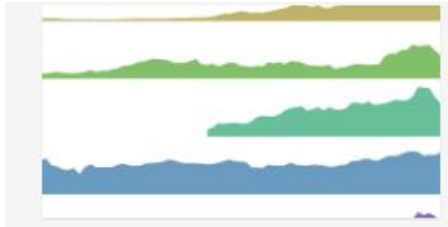
Paste it into the Raw data window
      -on Mac: (command V)
      -on PC: (control V)

Exercise 1: Small multiples chart using baby names dataset

We will make a chart to show changes in name popularity over time organized by individual names. This is an alternative to a traditional line or area chart.

1. Choose the Small Multiples (Area) chart, the 5th visualization in the first column

2. Scroll down. You should now see a list of green buttons (dimensions) and some gray windows.

3. Drag the dimension marked Top Name to Group, Year to Date and Occurrences to Size
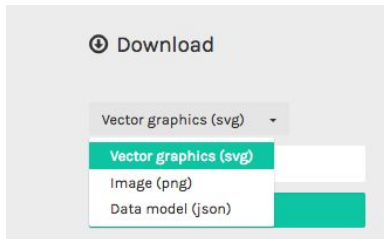

4. Change chart height to 1800 by typing it in the window:

This will create a small multiples chart showing patterns over time for each name.
You cannot filter out names using Raw. To do that, you have to modify your dataset.

5. To save the chart, download as a .png (image) or .svg (vector) file.

Definitions of file formats (from Wikipedia):
SVG Scalable Vector Graphics (SVG) is an XML-based vector image format for two-dimensional graphics with support for interactivity and animation.

Portable Network Graphics (PNG) is a raster graphics file format that supports lossless data compression. Popular on the internet.

JSON, or JavaScript Object Notation, is an open standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs. It is used primarily to transmit data between a server and web application.
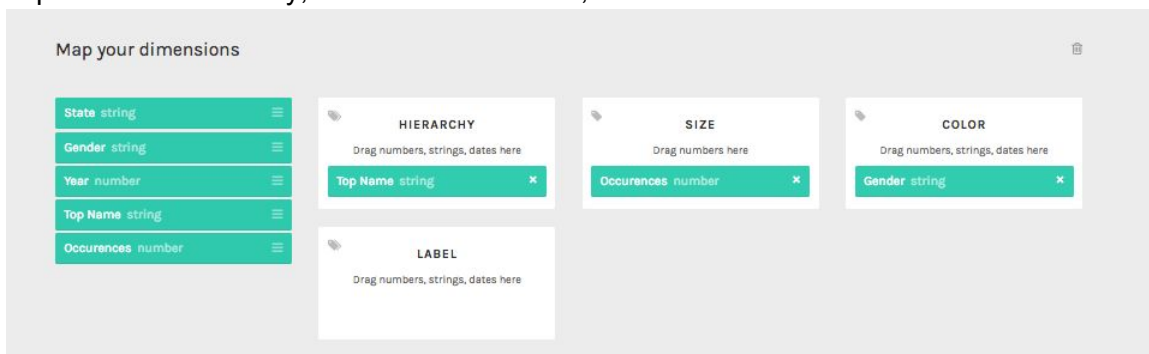
Exercise 2: Circle packing chart

Our next chart will be to create a bubble chart showing the most popular names in the entire dataset.

1. Choose the Circle Packing chart, top row, 3rd on the right.

2. Map the dimensions according to the image:
Top name to Hierarchy, Occurrences to Size, Gender to Color

3. Set Diameter to 1000 and Padding to 35:
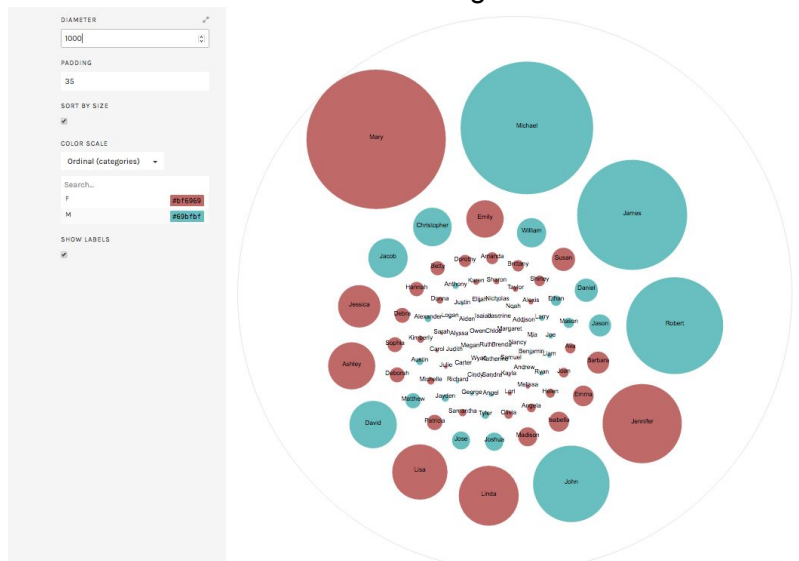
Tip: Adjust diameter and height to avoid overcrowding.

What happens if you go back and drag the Gender dimension above Top Name?
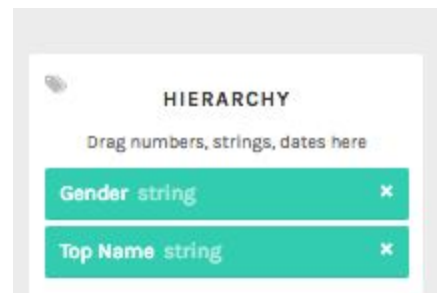
Optional Exercises in Raw

Scatterplot

Choose "Movies" from the datasets dropdown menu:
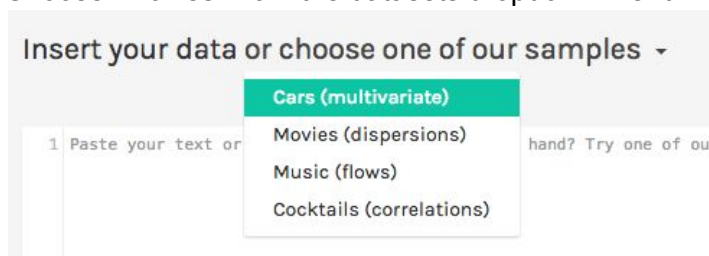
Once the data has successfully loaded, you can choose a chart option.

We are going to start with a simple chart: a scatterplot, to **compare the cost of the film with how much it earned at the box office**.

Click on the scatterplot image:

Now we have to map our dimensions to the scatterplot:

Drag and drop the following dimensions to the boxes indicated:
        Production budget to X-AXIS
        Total Domestic Box Office to Y-AXIS

Scroll down
Set max radius to 10
Check the box Set Origin to 0,0
From here, you can download the SVG graphic.

Cluster Dendrogram

Choose the "Movies" from the sample datasets dropdown menu:

We can create a tree-type visualization using this data and show relationship between genre and movie title.

Select the **Cluster Dendrogram** visualization:

Drag **Genre and Movie** into the Dimensions window

Your visualization should appear below:

Parallel Coordinates to examine multiple variables

A **parallel coordinates chart** allows us to look for correlations between numeric variables. Each line is one row of the data set.

To read a parallel coordinates chart, refer to this diagram. Lines with no correlation crisscross or appear jumbled with no obvious pattern. Lines with correlation either create an X pattern (inverse correlation), or a parallel pattern (positive correlation).

Parallel Coordinates using the Cars sample data set.

Select Cars sample data set from Raw dropdown menu.

Select Parallel Coordinates option from the list of choices

Drag variables you want to compare (must be numeric or dates) into the dimensions window.

Select **Economy, Cylinders, and Power.** Stack in the dimensions window. Top to bottom will be in left to right order in the chart.

This is the visualization you should get, showing an inverse relationship between mpg and cylinders, and positive relationship between cylinders and power:
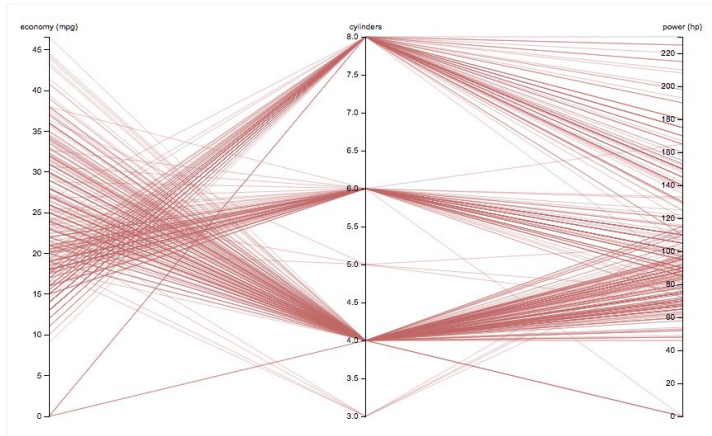
Other free tools for quick data visualization

Plotly - plot.ly
Tool for making simple, interactive charts pretty easily.

Carto - www.carto.com
This is a powerful tool for making interactive web maps. CartoDB can geocode the data based on zip, IP address, lat/long and country name. It's easy for a beginner to get started with and also has features for people familiar with HTML, CSS and Javascript. Works with tabular data and shapefiles. We have a workshop on this!

Voyant - voyant-tools.org
This is a good tool for reading and analyzing digital texts that is popular in digital humanities. It's easy to explore a data set quickly, for example, paper abstracts.