

# One Dataset in 5 Visuals with Tableau

Data Visualization and Design | CUNY Graduate Center | Summer 2019

*This tutorial is adapted from one written by Erin Waldron of Data Dozen*

## Goals

- Make clear and well reasoned design decisions
- Decide when to use interaction to enhance your project
- Control the look and feel of your dashboard

## Data

This Tableau workbook that uses this dataset

Dataset

## Premise

We have a few preliminary questions:

1. What is the distribution of first letters in US Baby names? Are some letters more popular than others?
2. Are there any outliers: Male or female names that have changed or are assigned to different genders more than others?
3. What is the popularity of my name over time?
4. Are there any other interesting trends in baby names that emerge from these exploratory visualizationg?

We will use a few visualizations to complete this

2. A treemap to show the distribution of names
3. A line chart and search bar to search for your own name
4. A scatterplot to show the position of a name between male and female
5. Stacked bar charts to show the % of a name for male vs female
6. A wordcloud to illustrate the distribution of first letters

## Visualizations

1. Treemap of names
2. Search bar for name & line chart
3. Scatterplot of percent frequency male vs. female with a trend line

4. Stacked bar charts to show the ratio of a name between male and female owners
5. Wordcloud of first letters

You might be wondering why we're using all of these plots since length is so much easier to read than area. Well, even though a bar chart does communicate the most efficiently, the world would be a very boring place if we only cared about efficiency. Sometimes, sacrificing readability for engagement, design, or beauty is a desirable tradeoff. Secondly, there is some research to suggest that when information is just slightly more challenging to read, it's more engaging and we learn more from it. So while a bar chart is great, it's not always the most elegant, engaging, or interesting. Sometimes you'll want to use a treemap just because it communicates the overall picture better.

### Treemap (aka Area Chart) of Names

Treemaps are usually made of a series of relatively sized rectangles that are fit together. They are good for representing things that are parts of a whole (such as the portion of the whole a particular name is).

They have their own internal axis: the cartesian plane doesn't mean anything and isn't translated. Therefore, all of the variables you will use are pulled into the marks cards, not the Rows and Columns. We want to make a Treemap of just Male names. You could do the same with just Female or All Names.

The only variables we are interested in are: the SUM of Female Names BROKEN DOWN BY Name

1. Drag the 'CALC: Female Names' Pill to the Color Mark Card
2. Drag the 'CALC: Female Names' Pill (again) to the Size Mark Card
3. Drag the 'Names' Pill to the Detail Mark Card
4. Drag the 'CALC: Female Names' Pill (again) to the Tooltips Mark Card

Customize your Tooltips & Clean up your treemap. I'll only do this once in this tutorial, though you need to do it every time.

1. Click on ToolTip
2. I used this format: `Since 1900, there have been <SUM(CALC: Female)> <Name>'s`
3. Think about what you want to prioritize, is it the name or the number of times it was used? Change your font, font size, color (if you wish) accordingly.
4. Change your title. I'll call it Female Names since 1900
5. Change the legend title (click on the dropdown, and select 'Edit Title')
6. Change the color to something more visually appealing. Click on the Color Marks Card and pick a single color gradient.
7. Remove the null values by clicking on the help box in the lower right corner.

Rename your Sheet ‘Treemap’

For a fun example, duplicate this sheet and filter by name. You might try Harry Potter names: Hermione, Draco, Sirius, or musicians: Drake, Beyonce, Cher, etc.

## Scatterplot

Let’s move on to make a scatterplot of the relationship between the percent each name is used by gender. This is going to be a diagonal line since each name will be some portion of the whole. We’ll also size the dots based on how common the name is.

1. Drag the ‘CALC: Male Percent’ to the Columns
2. Drag the ‘CALC: Female Percent’ to the Rows
3. Drag ‘Occurrences’ to the Size Marks Card
4. Drag the ‘Name’ to the Detail Marks Card
5. Increase the size of the Name

We see that the names at either end (the ones that are more dominantly male or female) are far more common than the gender-agnostic names along the center.

5. Drag the ‘CALC: Female Percent’ to the Color marks card. This gives us a nice gradient (since it’s a continuous numeric variable). However, we want 2 colors. Click on the Color Marks card and select one of the color schemes that have opposite ends (I like Sunrise and Temperature). Now we have a nice visualization showing the relative dominance of names throughout time.
6. Remove the axis labels by right-clicking and going to ‘Tick Marks’ and select ‘None’
7. Rename the axis labels by double clicking and Changing the label to Male and Female
8. Rename your chart

Rename your sheet ‘Scatterplot’

## Wordcloud

Wordclouds are visualizations that size the words according to their relative prominence. They are awful ways to show data: not only are they scaled based on area, but it’s impossible for me to visually understand the difference between how prevalent “Penelope” is versus “Da” in a word cloud. Don’t use them to communicate your primary point (if you need your reader to understand the prevalence of a word versus the others, give a tooltip). However, wordclouds are great at offering a quick snapshot and oftentimes have a cool effect.

We’re going to make one of the first letters (a little bit easier to read)

1. Drag 'CALC: First Letter' to Text on the Marks card.
2. Drag 'CALC: First Letter' to Size on the Marks card.
3. Right-click 'CALC: First Letter' on the Size card and select Measure > Count.
4. You might have to change the Mark type from Automatic to Text.

We want to color this based on the dominance of each letter in Female names. We'll need to make a new calculated field of Female first letters.

1. Click on the dropdown box in Dimensions >> Create Calculated Field
2. Call this field CALC: Female First Letter
3. Enter this formula: `IF [Sex] = "F" THEN LEFT([Name], 1) END`
4. Drag 'CALC: Female First Letter' to Color on the Marks card.
5. Since this is a Dimension you may need to tell it to 'COUNT'. Click on the drop down arrow and select Measure > Count
6. Click on the color to change the color. I like Sunrise-Sunset
7. Rename your chart.

Rename your sheet.

### Stacked Bar Chart

We'll illustrate this with a group of names: Antonios, Kelly, Kiana, Maria, Michelle, Devon

1. Drag 'Name' to the Filter Card and Select 3-7 names.
2. Drag the 'Name' Pill to the Columns
3. Drag the 'Occurrences' Pill to the Rows
4. You see that these are raw counts
5. First we'll drag Sex to the color card to break up the columns by Sex

Now we'll turn our bars into percentages (since Maria and Michelle are disproportionately more common in the US than all of the other names).

3. Right-click SUM(Occurrences) on the Rows shelf, and then click Add Table Calculation.
4. In the Calculation Type drop-down menu select Percent of Total.
5. Compute Using from select Table (Down)

Change the title, axis names, colors if you wish.

### Line Chart with a search bar

Time is generally thought of as a continuous variable. Though we divide our world into hours/days/months/etc., you can divide time into an infinitely small set of pieces and therefore it's continuous (something like number of people is discrete and not continuous: you can have half a second, you can't have half

a person). Discrete variables are well represented by lines. So we are going to plot the popularity of a name over time. There are thousands of names in this dataset. We have a few options:

1. Take the most common 3-7 names
2. Take some outlier names
3. Choose based on another set of criteria (famous people)
4. Let the user choose.

We're going to let our user choose. This is called an 'Action'

First set up your Line Graph so that the occurrences are broken down by Sex. We could have just used our calculated fields, but quite often, when you are doing something simple (i.e., visualizing a variable broken down by another variable such as names broken down by gender, you'll want to specify that with a marks card rather than 2 variables describing the same thing. The real reason we made the Male & Female variables was for illustration and to calculate the percentages)

1. Drag 'Year' to the Columns
2. Drag 'Occurrences' to the Rows
3. Drag 'Sex' to the 'Color' Marks Card

This tells us how many male and female children were registered with SSN every year. We want to get more specific.

1. Drag 'Name' to the Filter Card
2. Select only one name that you are interested in. I'm going to choose Devon because it's interesting: it has changed. And because it's my wife's name.
3. This will be the default name for your visualization. We want to turn this into an action when we make the dashboard.

Rename your sheet 'Line'

In the next session, we will put all of these together into a Dashboard.