Tableau Charts

1. **Scatter Chart**

This scatter chart is showing the relationship between GDP per capita and life expectancy in 2016 by country. I used data from four different datasets, including the population, gdp/capita, and life expectancy data sets from the World Bank and a countries dataset that included region.

I connected all four datasets in Tableau with inner joins to the original gdp/capita dataset to create the final combined dataset. After moving both from measures to dimensions, I assigned Life expectancy to the rows (y-axis) and GDP/capita to columns (x-axis). Region is assigned to “Color” differentiate each country’s region and population is assigned to “Size” to make dot size proportional to each country’s total population.

<https://data.worldbank.org/>

<http://www.countries-list.info/Download-List>

1. **Multi-Line Chart**

This chart is a multi-line time series chart to show average daily flight delays for each airline from January – March, 2015. The data 1,048,575 rows and 9 features, adding a 10th feature of airline name from a separate sheet with an inner join on Iata code.

I first added a top filter by field to reduce the dataset to only the airlines with the top 10 total departure time. I moved departure delay from measures to dimensions and it to the rows with the average measure and date to the columns with day selected. Airline name is added to “Color” to make it a multi-line chart and airline name and total number of records are added to “Detail” to provide additional information.

<https://www.kaggle.com/usdot/flight-delays>

1. **Stacked Bar/Line Chart**

This chart is a dual-axis graph showing total US foreign aid from 1946 – 2016, by region as a bar chart and by assistance type as a multi-line chart, and shows how US aid has changed over the years. The data is from USAID and has 821,744 rows and 49 columns showing each foreign aid project since 1946.

I moved fiscal year from measures to dimensions and assigned it to the columns and the total aid in current USD to the rows and selected bar chart. I added a range filter to remove the year 2017 as it was an incomplete year. I assigned region to “Color” to create a stacked bar chart based on region. I then added total aid in current USD again to the right axis and made this chart a line chart to create the dual axis chart. This second chart is a multi-line time series graph with custom colors that shows US aid split by assistance type (military and economic).

I also made two extra charts out of curiosity. The first chart is a treemap showing the distribution of US aid by country and the second is a multi-axis line graph showing military and economic aid by region.

<https://www.usaid.gov/results-and-data/data-resources>

1. **Symbol Map**

This chart is a symbol map showing all earthquakes with a magnitude of 5.5 or more from 1965 – 2016 and gives a clear trend of where earthquakes happen and general magnitude distribution across the globe. The data is from the US Geological Association and contains 23,412 rows and 21 columns.

I started by assigning Latitude to the rows and Longitude to the columns. After selecting a symbol map, I realized there were too many data points to see the individual earthquakes so set an “at least” filter to only map earthquakes with a magnitude of 7 or more. I then assigned magnitude to size to easily visualize the severity of the earthquakes as bigger dots meant bigger earthquakes. However, the dot sizes were not easily differentiable between the magnitudes, so I created a calculated field called “Calculation1”, which is “10^magnitude”. Assigning this new calculation to “Size” gave larger earthquakes a much larger bubble to show where the most severe earthquakes occur and more accurately depicted the actual increase in severity between magnitudes. I finished by tweaking map and dot aesthetic properties to arrive at the final chart.

<https://www.kaggle.com/usgs/earthquake-database>

1. **Filled Map**

This chart shows the mortality rate (deaths per 100,000) from cardiovascular disease in 2014 in each county of the US. The county-based data is from Institute for Health Metrics and Evaluation and has 3194 rows and 14 columns.

I selected the cardiovascular dataset from the other disease data sheets and removed the extra “(male mortality, female mortality)” from total mortality in 2014 to use that specific year with an integer value rather than string value. I assigned generated Latitude to rows, generated Longitude to columns, Mortality Rate 2014 to the “Color”, and State and County to “Detail” to generate county level coloring. I chose a “temperature diverging” color palette to provide a greater range of colors and more clearly demonstrate which counties had the highest mortality rates, as highlighted in red.

<https://www.kaggle.com/IHME/us-countylevel-mortality>

D3.js Charts

1. **Bar Chart**

This chart is a basic bar chart that plots the median salary of 50 different color majors of the course of their careers. The data is sorted from highest median salary to lowest and each data element is bound to a rectangle svg of with a height corresponding to the median salary. Median salary is assigned to the y-axis and college major is assigned to the x-axis.

<https://www.kaggle.com/wsj/college-salaries>

1. **Scatter Chart**

This scatter plot uses the famous Iris dataset, which contains 4 feature measurements, the species feature, and 150 data elements. Petal width is bound to the y-value, petal height is bound to the x-value, and species to bound to the dot color. Both petal length and petal width are then bound to a dot svg and scaled linearly onto the axes. A legend and mouse over event are created to give information to the chart.

<https://www.kaggle.com/uciml/iris>

1. **Pie Chart**

For the pie chart I used the same airline data as I used for the multi-line chart in Tableau. I used a pivot table in Excel to get the data into a form of having only the columns “airline”, “delay”, and “airlineFullName”, and each row is the total delay time in minutes for each airline over the course of the data set. I created an arc svg variable and defined the radius of the pie chart and labels. I then bound the delay data to an arc and created the pie chart with the airline labels centered on each arc.

1. **Multi-Line Chart**

For this chart, I used the same GDP data as from the scatter chart in Tableau. Using Excel, I sorted by 2016 GDP and selected the top 14 countries. I then used a Tableau ribbon in Excel to reshape the data table from wide to long where each country and year combination are a single row, resulting in three columns and 639 data points. I created the line svg and bound the year (after having formatted it to %Y) and gdp values to the x and y-values, respectively. Country is nested and bound to the color palette and used to map a line svg for each unique country.

1. **Stacked Bar Chart**

For the stacked bar chart, I used the same US Foreign Aid dataset as used for the stacked bar chart in Tableau. Using an Excel pivot table, I summed all foreign aid based on economic or military assistance type for each year. I defined the stacking variables, mapped the assistance type on each bar using layout.stack() to the layout variable, mapped the layout variable to a rect svg to create the stacked bar chart, and used the default colors to do so.