

Data Visualization

Summary: Why should we use data visualizations at all? Humans are much better at interpreting relationships in data when we can see them directly, instead of written out as a bunch of numbers in a table. Consider this set of data, called [Anscombe's Quartet](#)

STEP 1: Extract and Clean

Before you get to creating visualizations, you'll need to understand the different types of data you'll be working with. There are two main types, quantitative and qualitative. Quantitative data is continuous or discrete while qualitative is categorical or ordinal. You may need to change data types and handle missing or messed up data prior to visualization.

Useful Alteryx tool: Select

STEP 2: Explore Data

Next you should get familiar your data. You can use bar graphs for categorical variables to see how the data is distributed between the categories. For continuous variables, you can use histograms to see how they are distributed. Lastly you can use scatter plots to look at relationships between variables. This will help you identify variables that are correlated or have other interesting relationships.

Useful Alteryx tool: Field Summary, Histogram, Scatter Plot

STEP 3: Define Objective

Create a series of questions you want to answer with your data. You may need to go back at this point to aggregate your data to get it at the right level for your visualization. Next you need to determine the best method of presenting insights based on the analysis. There are different recommended visualizations based on if you are showing the relationship between variables, doing a comparison, showing the composition or the distribution. Read more here [Chart Suggestions](#)

STEP 4: Visualize Data

For quantitative data, especially when comparing two variables, dots are preferred. Lines are really good for connecting related data and guiding the eye, such as data over time. When you are comparing groups, bar graphs are excellent. Color is often used to encode a third dimension since it is really difficult to interpret three dimensional data on a two dimensional plane (like your computer screen). This can either be with continuous data like in heatmaps or discrete data when splitting data into groups. Geospatial data (latitude, longitude) is readily viewed on a map.

Useful Alteryx tool: Charting, Export to .tde file to use in Tableau

STEP 5: Share Data

Make sure the amount of information shared is not overwhelming. There are some concerns you should keep in mind when sharing your visualizations. Color blindness, the most typical is red/green so when selecting colors be mindful of colors that would be difficult to differentiate for them. When designing your visualizations, you need to take care to remove anything that distracts from the data. The lie factor compares the size of the effect in the graphic with the size of the effect in the data. Bar charts aren't the only common plot that can be abused. Often you'll see line charts with lines connecting points at irregular intervals. Read more here [Chart Junk](#)

Useful Alteryx tool: Charting, Export to .tde file to use in Tableau