Data Science

Understanding the Basic of Plotting

Tableau Defaults

- Discrete fields (blue colored) - create table headers
- Continuous fields (green colored) - create chart axis

Let's create First Plot

Types of Charts

Show Me Toolbar

- When you have some specific visualization in your mind and you are not sure how to create it
- When you know which fields you want to analyze, but you don't know which type of chart will be suitable for that field.

Types of Charts

- Maps -
    - Adding a geographic field produces a map view with latitude and longitude as axes and the geographic field on the Level of Detail shelf.
    - Subsequent dimensions add rows to the view while subsequent measures further refine the map by adding size and color encoding.
- Line
    - Adding a date dimension and then adding a measure produces a line view.
    - All subsequent clicks result in refinement of the line view.

Types of Charts

- Text Table
  - Adding a dimension first produces a text table (or cross-tab).
  - All subsequent clicks on fields result in refinement of the text table.
- Bars -
  - Adding a measure first and then a dimension produces a bar view.
  - All subsequent clicks result in refinement of the bar view, unless a date dimension is added, at which time the view is changed to a line.
- Scatter
  - Adding a measure and then another measure produces a scatter view.
  - Subsequent dimensions result in refinement to the scatter view. Subsequent measures will create a scatter matrix.

Adding Customizations

Customizations

- Color
- Size
- Label
- Detail
- Tooltip

Line Chart

Line Chart

- Single Line Chart
- Dual Axis Line Chart

Area Chart

Area Chart

- Commonly used to showcase data that depicts a time-series relationship.
- But unlike line charts, it can visually represent volume also.
- Area charts are primarily used when the magnitude of the trend is to be communicated rather than individual data values.
- To showcase this magnitude, the area between the line segments and the axes is highlighted by filling it with color.

Use of Area Chart

- Suitable to
    - Get a sense of volume to your data
    - See part-to-whole relationships between groups
    - Analyze trend of magnitude of a quantitative data
    - Compare trend/proportion of each category
- Not suitable when -
    - There are too many categories, as this can be difficult to read.
    - There are a lot of overlapping values in your data series.

Bar Chart

Bar Chart

- Bar charts provide a visual presentation of data across different categories.
- One axis of the chart shows the specific categories being compared and the other axis represents a numerical value scale.

Bar Chart

- When to Use It
  - You want to compare two or more values in the same category
  - You want to compare parts of a whole
  - You don't have too many groups (less than 10 works best)
  - You want to understand how multiple similar data sets relate to each other
- Don't use a bar chart for the following reasons:
  - There are large number of categories available, i.e more than 10
  - You want to visualize continuous data

Bar Chart

- Variations of Bar Chart
  - Basic Bar Chart
  - Side by side bar chart
  - Stacked bar chart

Histogram

Histogram

- A histogram is a data visualization that shows the distribution of data over a continuous interval or certain time period.
- It's basically a combination of a vertical bar chart and a line chart.
- The continuous variable shown on the X-axis is broken into discrete intervals and the number of data you have in that interval determines the height of the bar.
- It gives an estimate as to where values are concentrated, what the extremes are and whether there are any gaps or unusual values throughout your data set.
- They are also useful for giving a rough view of the probability distribution.

Bar chart vs Histogram

Bar Chart -

- Useful for comparing distinct values of data.
- Bar height represents the actual value of items.
- Used to plot categorical variables (the qualitative data on the x-axis).
- The x-axis in a bar chart represents a discrete variable. Each item on the axis is independent of the other item.

Histogram -

- Useful for comparing distribution of continuous data.
- Bar height represents the frequency of items that fall in each bin.
- Used to plot numerical variables (the quantitative data on the x-axis).
- The x-axis in a histogram represents a continuous variable that has been grouped into multiple bins.