

CS500-Data Science Tools and Technique

Data Visualization

Bahar Ali
PhD Scholar,
National University Of Computer and Emerging Sciences,
Peshawar.

Charts and Graphs

- A graphical technique for representing a data set
- A chance to make a convincing argument through visualizing data
- If we work with chart types that are not the right fit for the data, we won't convince anybody

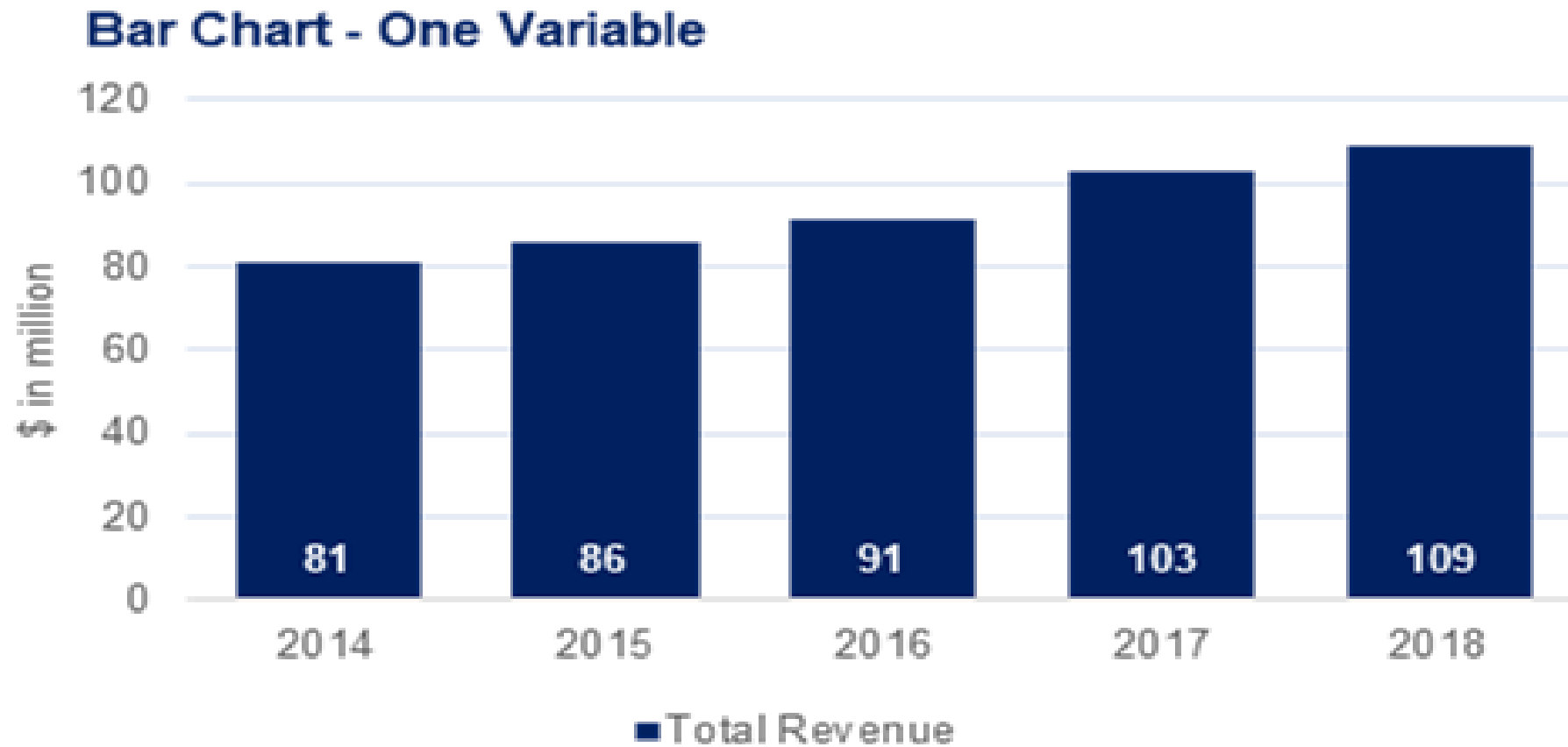
Charts and Graphs

- What are the different chart types and how they are suitable for specific type of data, lets discuss some of them

Bar Graph

- The most frequently used chart types.
- As the name suggests a bar chart is composed of a series of bars illustrating a variable's development.
- Presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent.
- People are generally familiar with them and can understand them easily.

Bar Graph



Bar Graph

- **When to use bar charts**

- When we want to track the development of one or two variables over time. For example, to show how a company's total revenues in one year period.
- Good when your data is in categories (Male, Female counts)

- **When to avoid using bar charts**

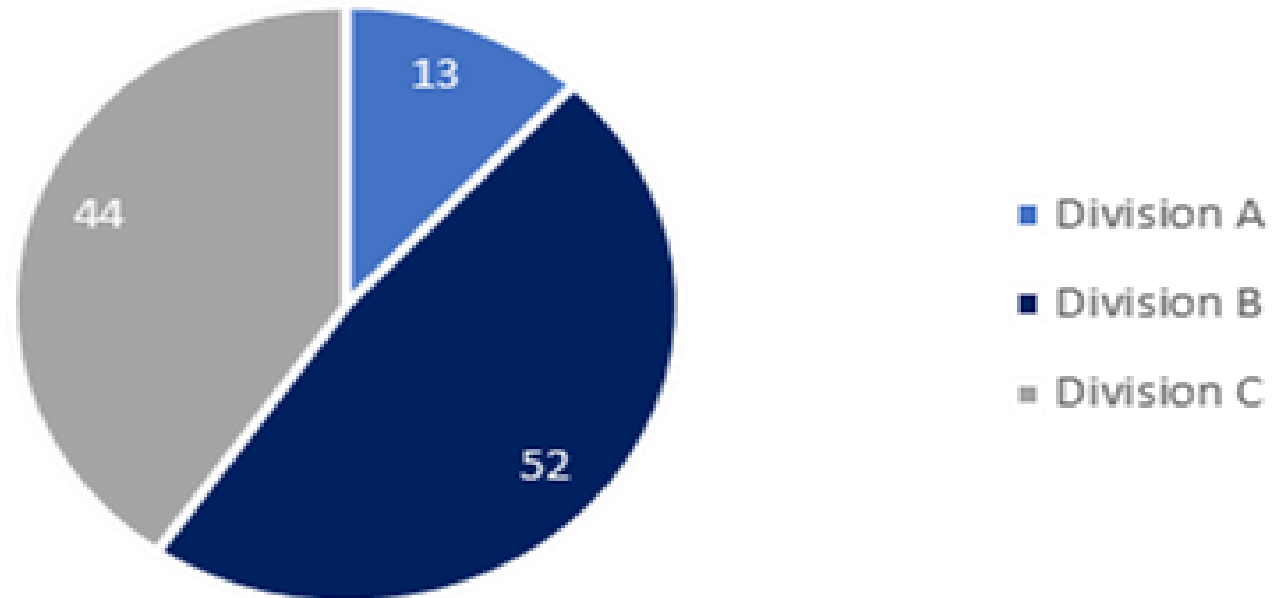
- Bar charts can be confusing, especially if one uses them to compare several variables.
- When you have continuous data (such as a person's height), in such cases using a Histogram or line graph are a better options.

Pie Chart

- A circular statistical graphic which is divided into slices to illustrate numerical proportion.
- In a pie chart, the arc length of each slice, is proportional to the quantity it represents
- The larger a slice is the bigger portion of the total quantity it represents.
- The smaller a slice is the smaller portion of the total quantity it represents.

Pie Chart

Pie Chart



Pie Chart

- **When to use bar charts**

- Best suited to depict sections of a whole.
- If a company operates three separate divisions, at year-end its top management would be interested in seeing what portion of total revenue each division accounted for, then use pie chart

- **When to avoid pie charts**

- In situations when we would like to show how one or more variables develop over time.
- When the sum of the proportions does not make 100%

Doughnut Chart

- A doughnut (or donut) chart is a pie chart with a blank circular area in the center.
- Shows the same information as a pie chart but from a different perspective
- The chart is divided into parts that show the percentage each value contributes to a total.

Doughnut Chart

Doughnut Chart



■ Division A ■ Division B ■ Division C

Doughnut Chart

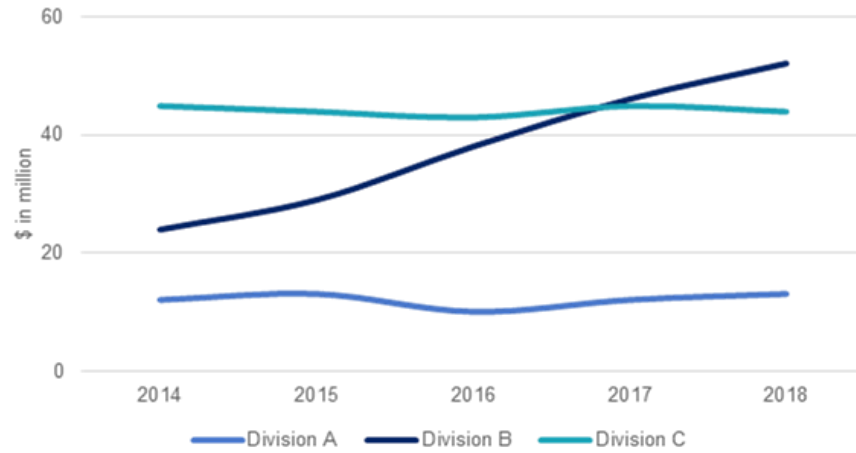
- **When to use bar charts**
- The only important difference is that doughnut charts allow us to indicate the total amount by adding a text box in the middle.
- **When to avoid using doughnut charts**
- In situations when we would like to show how one or more variables develop over time.
- When the sum of the proportions does not make 100%

Line Graph

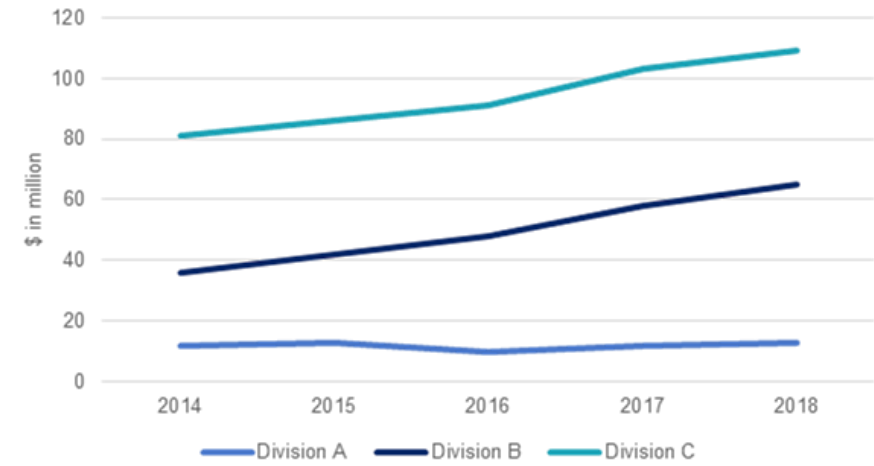
- A line graph, also known as a line chart, is a type of chart used to visualize the value of something over time.
- Consists of a horizontal x-axis and a vertical y-axis.
- Showing how single, or multiple variables develop over time.
- For example, a finance department may plot the change in the amount of cash the company has on hand over time.

Line Graph

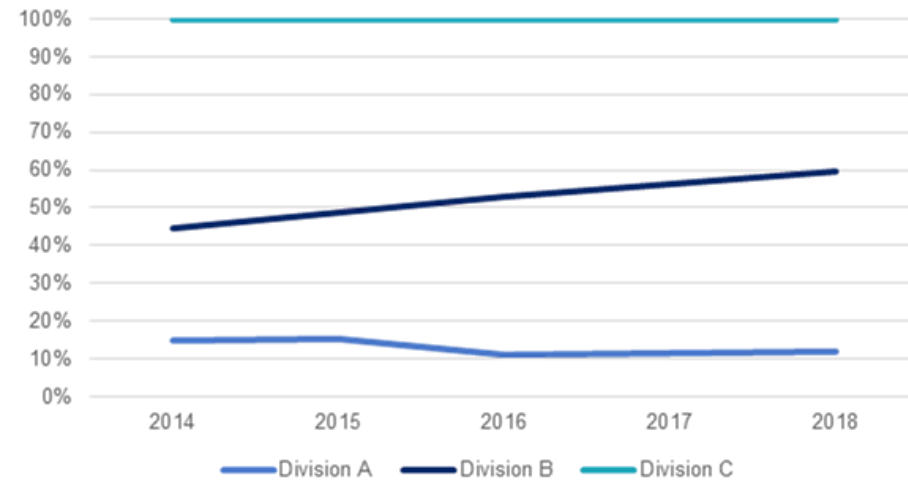
Line Chart



Stacked Line



100% Stacked Line



Line Graph

- **When to use line charts**

- Used to track changes over short and long periods of time.
- To compare changes over the same period of time for more than one group.
- Allows us to track the development of several variables at the same time.
- Useful to show trends and behavior in a data.
- To display several dependent variables against one independent variable

- **When to avoid line charts**

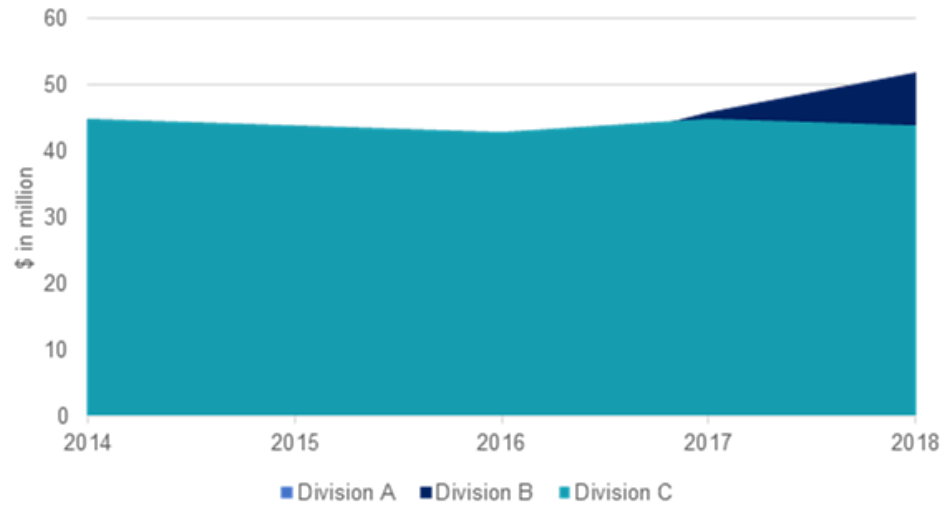
- When you want to show how the individual parts of a whole change over time. Instead use a stacked line chart (where line values accumulate) or a 100% stacked line chart (where lines accumulate to 100%), but a stacked *area* chart would look better.

Area Chart

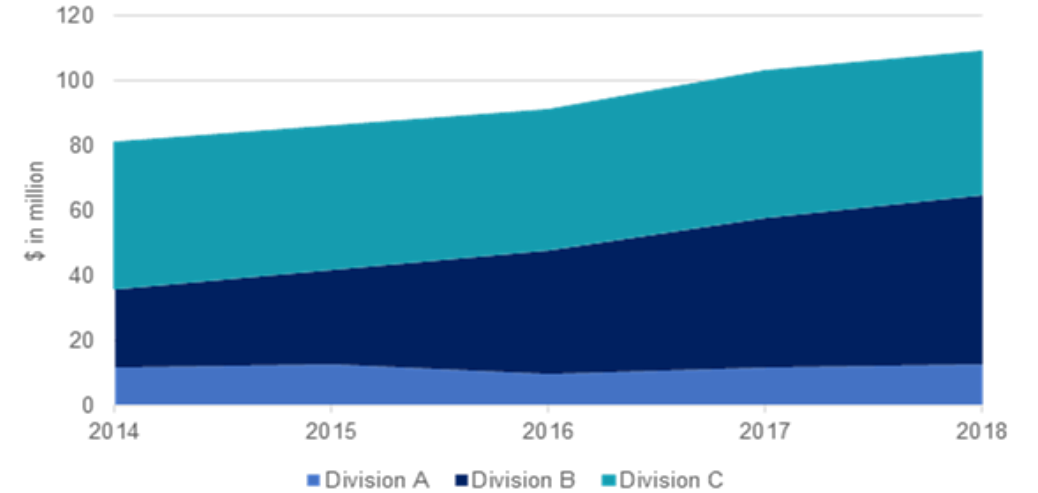
- Represents the change in a one or more quantities over time.
- It's made by plotting a series of data points over time, connecting those data points with line segments, and then filling in the area between the line and the x-axis with color or shading.
- Colored regions (areas) show us the development of each variable over time.
- There are three types of area charts: regular area chart, stacked area chart, and 100% stacked area chart.

Area Chart

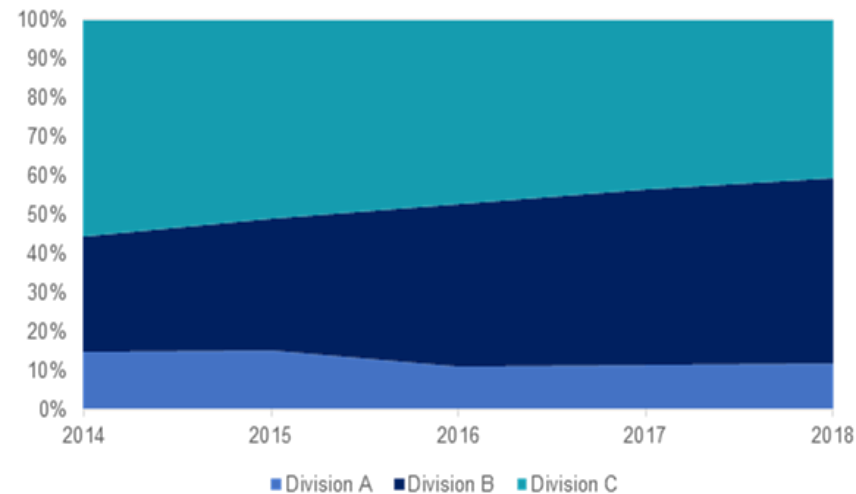
Area chart



Stacked Area Chart



100% Stacked Area Chart



Area Chart

- **When to use an area chart**

- To show how the parts of a whole change over time, For example, if the company has three revenue generating divisions, it is very likely that management would like to see the development of each of these divisions.
- This is a great way to draw attention to the total value and still emphasize an important trend

- **When to avoid area charts**

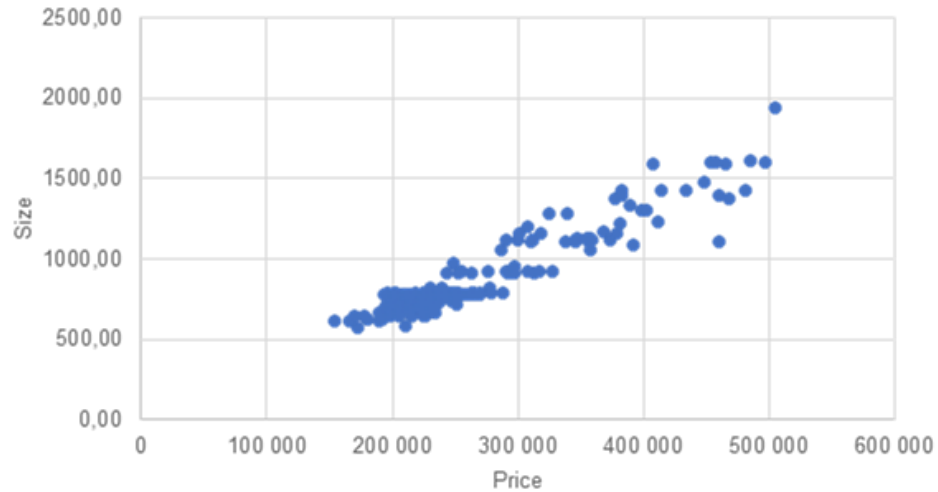
- Not suitable for representing parts of a whole over a single period.
- Area charts can confuse things when showing parts of a whole, then It is recommended to use stacked area chart, and 100% stacked area chart

Scatter Plot

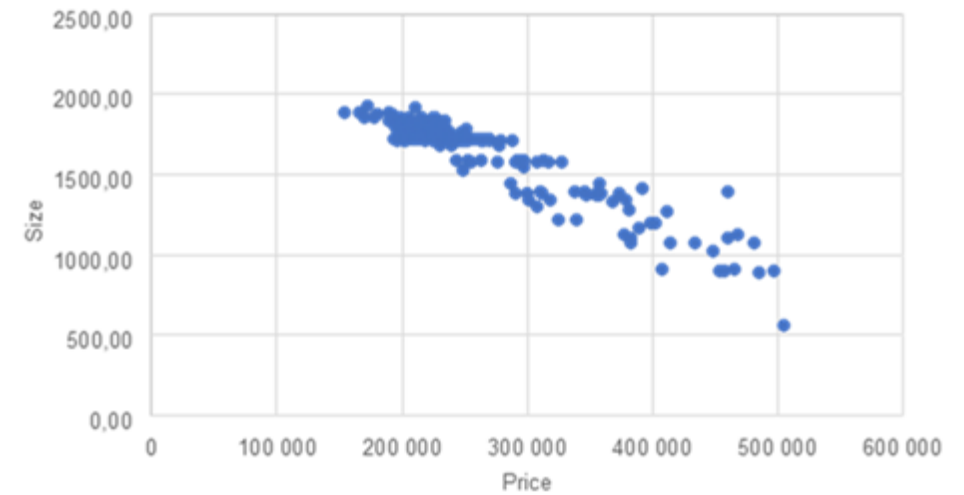
- A type of chart that is often used in the fields of statistics and data science.
- Used to display values for typically two variables for a set of data
- It consists of multiple data points plotted across two axes.
- Scatter plots chart types are excellent for finding correlations

Scatter Plot

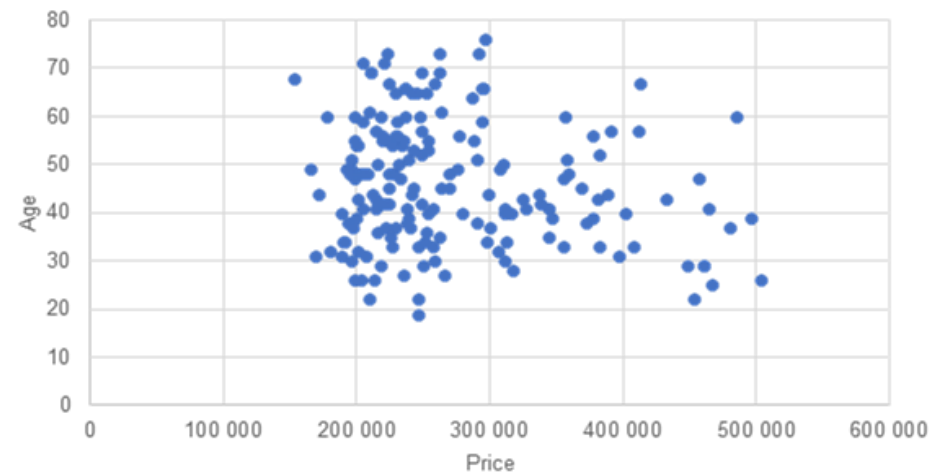
Scatter Plot - Positive Relationship



Scatter Plot - Negative Relationship



Scatter Plot - No Relationship



Scatter Plot

- **When to use scatter plots**

- A great indicator that allows us to see whether there is a pattern to be found between two variables.
- Use it when data is numerical or numbers

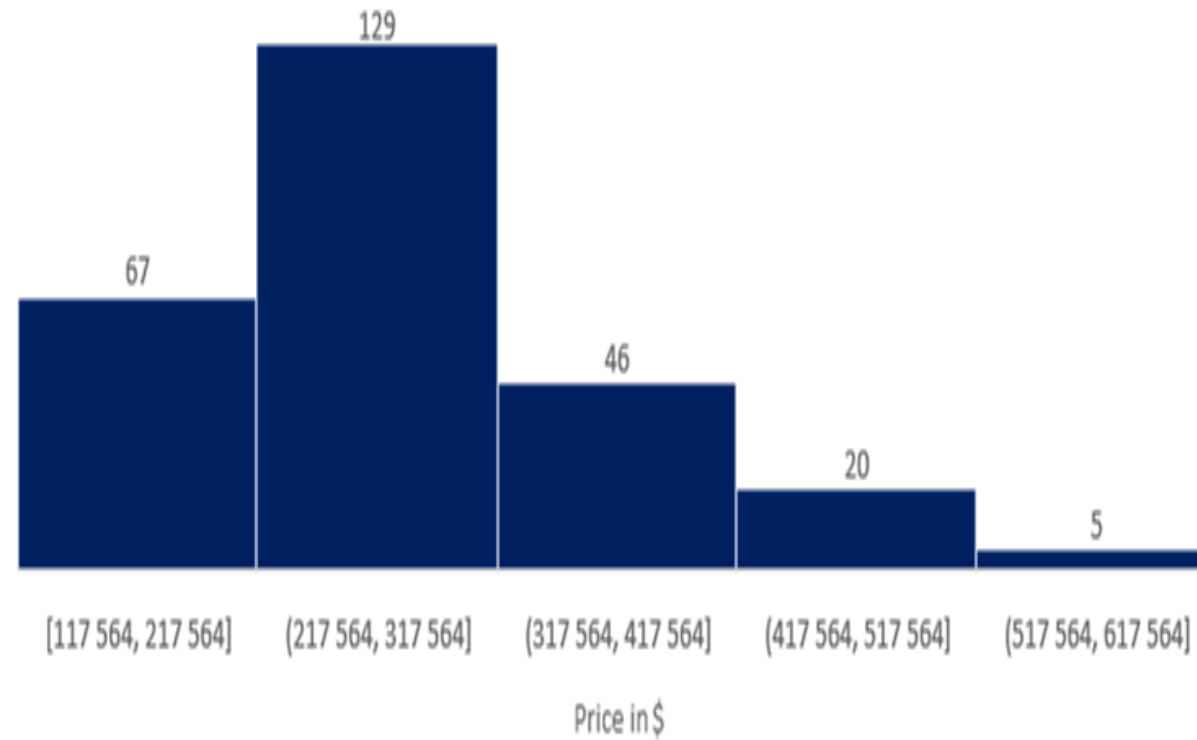
- **When to avoid scatter plots**

- When data is single dimensional
- A scatter plot requires at least two dimensions for our data.
- Not used with categorical type of data

Histogram Chart

- Histogram is a series of bins showing us the frequency of observations of a given variable.
- To show people house cost, a histogram can be used

Histogram Chart



Histogram Chart

- **When to use histograms**
- Histograms are great when we would like to show the distribution of the data we are working with
- This allows us to group continuous data into bins and hence, provide a useful representation of where observations are concentrated
- **When to avoid histograms**
- When the data you are working with contains multiple categories or variables