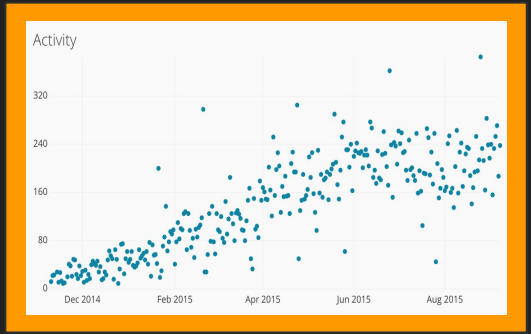
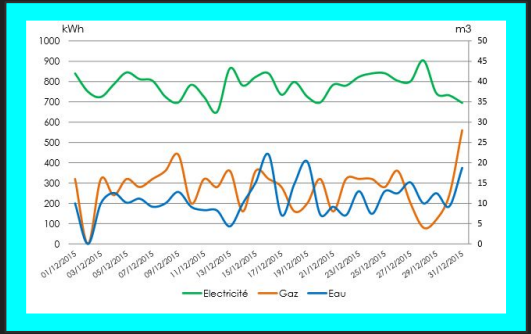




Statistical Charts

A Guide.

SoftStack Factory 2018

We Will divide Statistical Charts by The kind of data that they are best suited to display.

Categorical Vs. Quantitative

Categorical

- Bar Chart
- Pie Chart

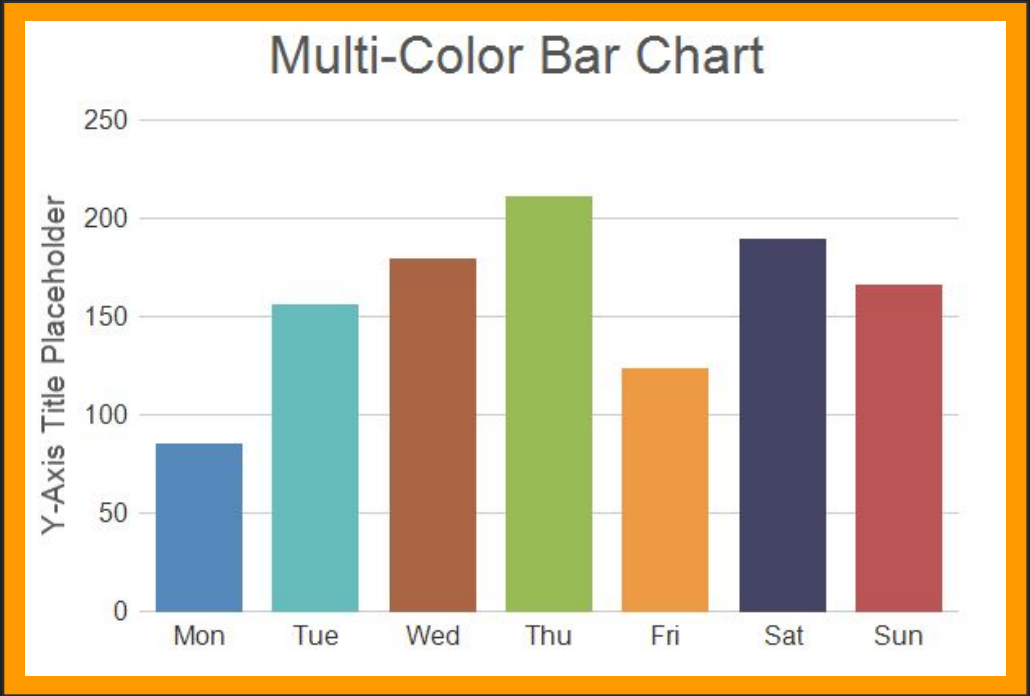
Quantitative

- Histogram
- Box Plot
- Line Chart
- Scatter Plot

Categorical Data

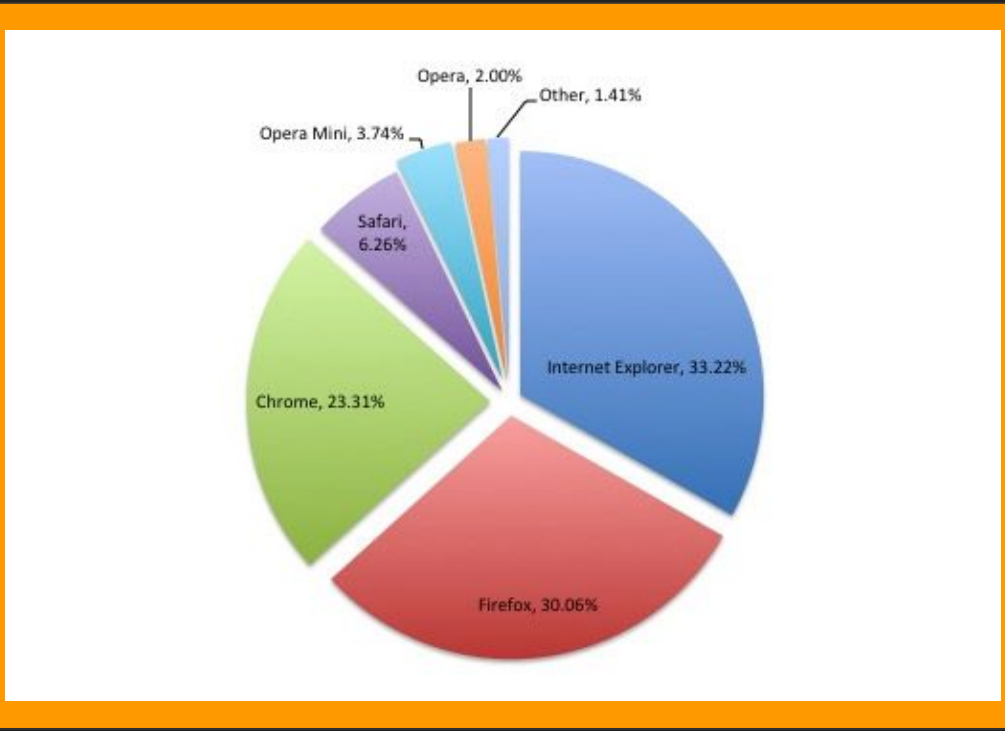
Sample	Size	Color	Shape	Label
1	Large	Red	Round	Y
2	Large	Blue	Round	N
3	Small	Blue	Square	N
4	Small	Red	Square	Y

Bar Chart



- Each Bar Represents a category.
- Magnitude of Bar can represent a relative or absolute value.
- Generally used to show the frequency of a list of categories.
- Example:
  - # of laps run x Day of Week

Pie Chart

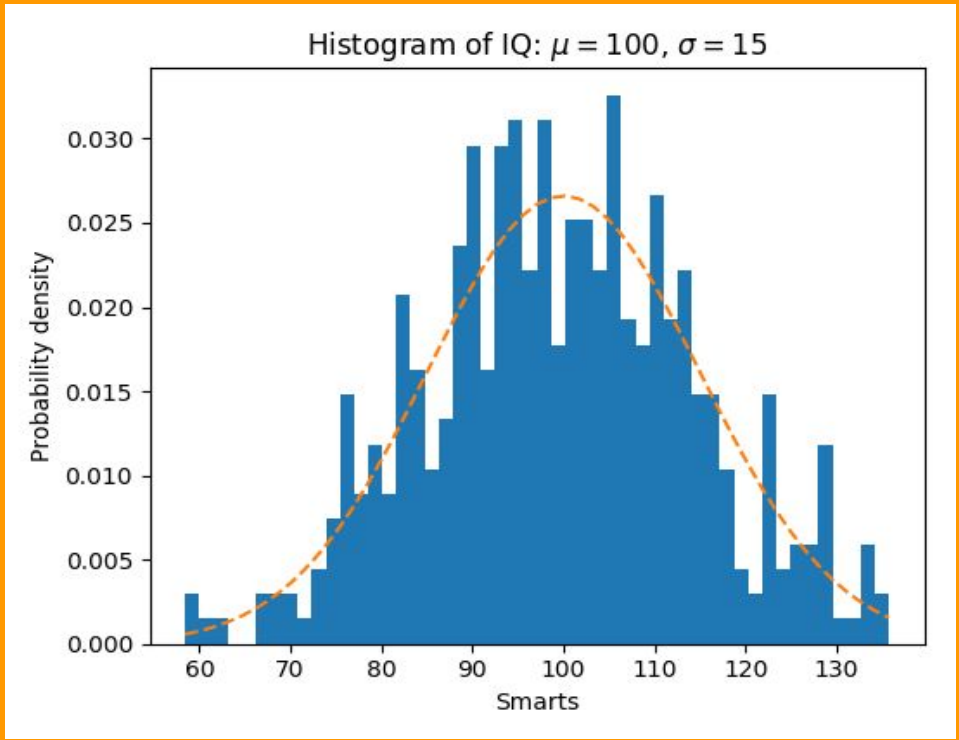


- Each slice Represents a category.
- Generally, each slice is used to show the relative frequency of each category when compared to a whole.
- Example:
  - Market share x browser

Quantitative Data

Sample	pedal_size	num_pedals	stem_size	Label
1	3.2	5	8	Y
2	3.9	5	7	N
3	1.1	4	5	N
4	1.9	3	8	Y

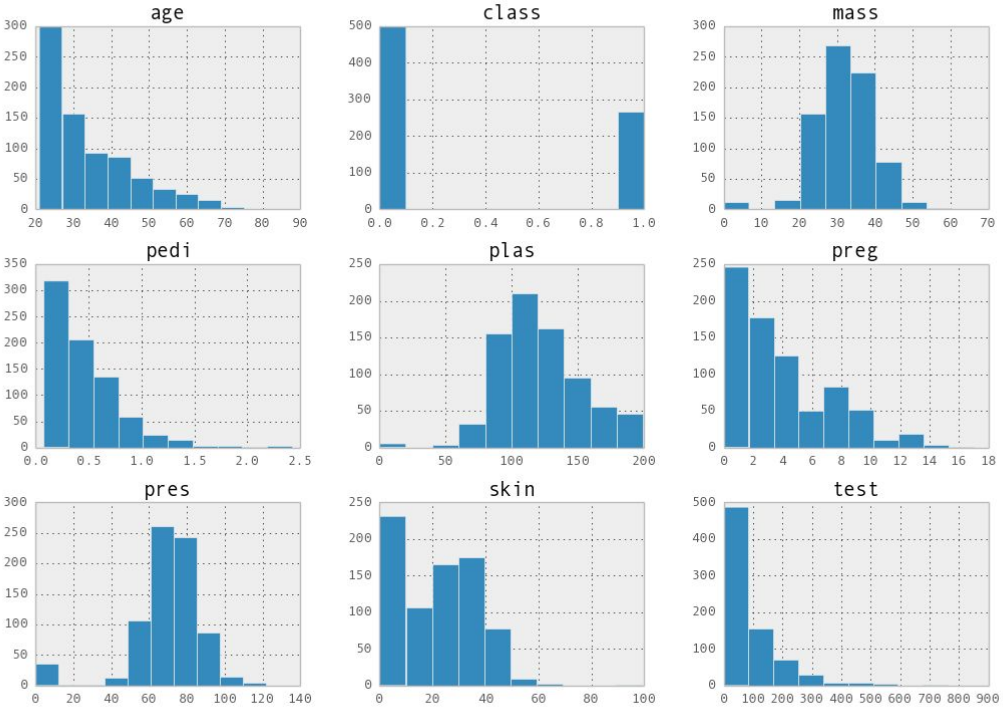
Histogram



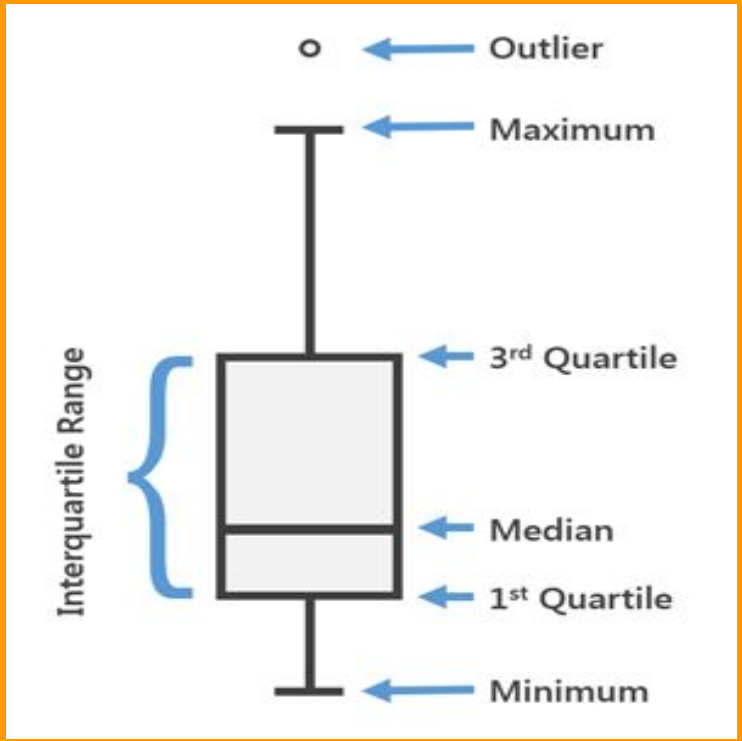
- Summary plot for a SINGLE variable.
- Great for visualizing distribution:
  - Center and Variability
  - Skewness and Modality
  - Outliers or Strange Patterns.
- Powerful when single feature histograms are all viewed side by side.

Ideal Use Case:

Multiple Histograms Viewed Side By Side.



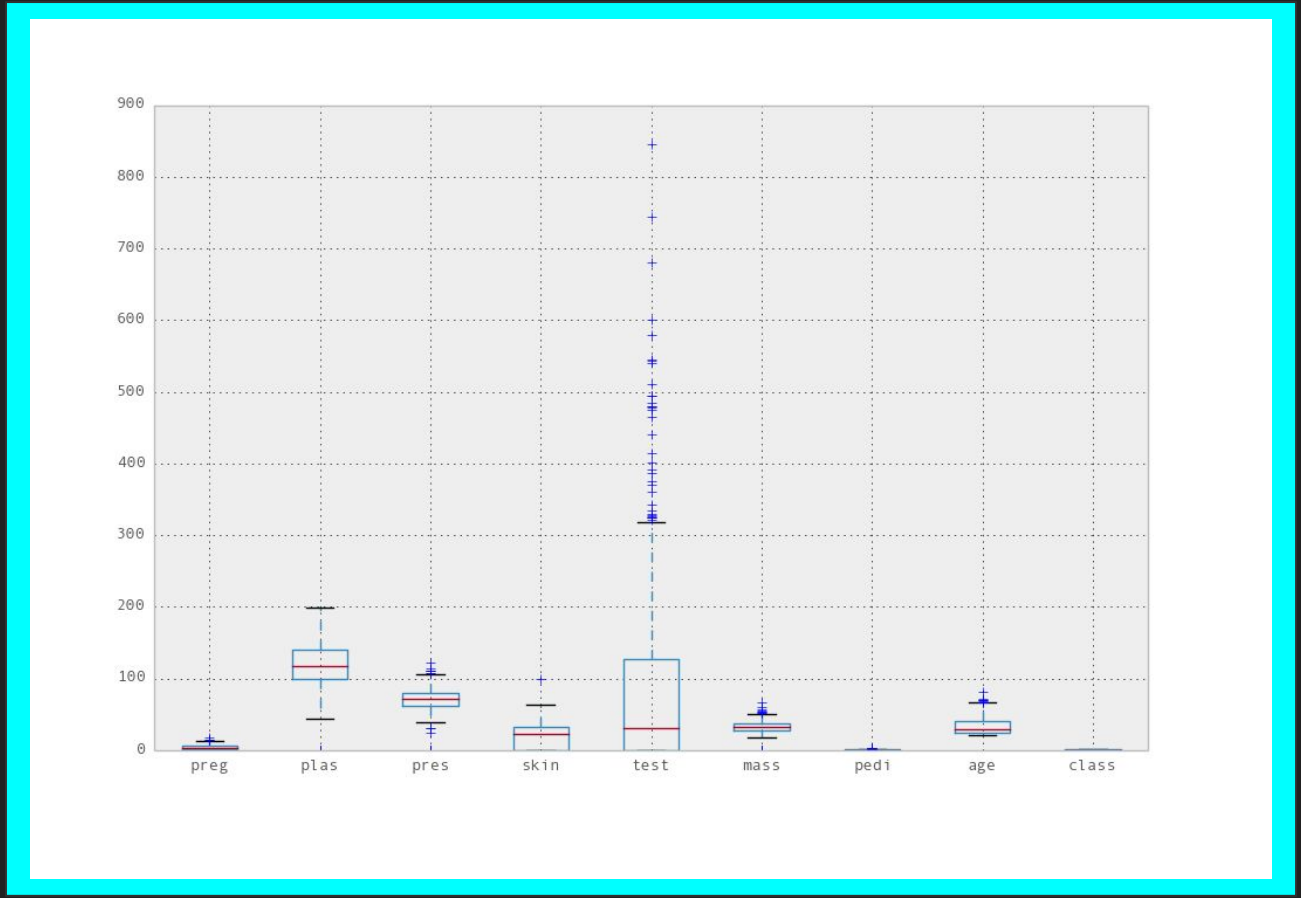
Box Plot



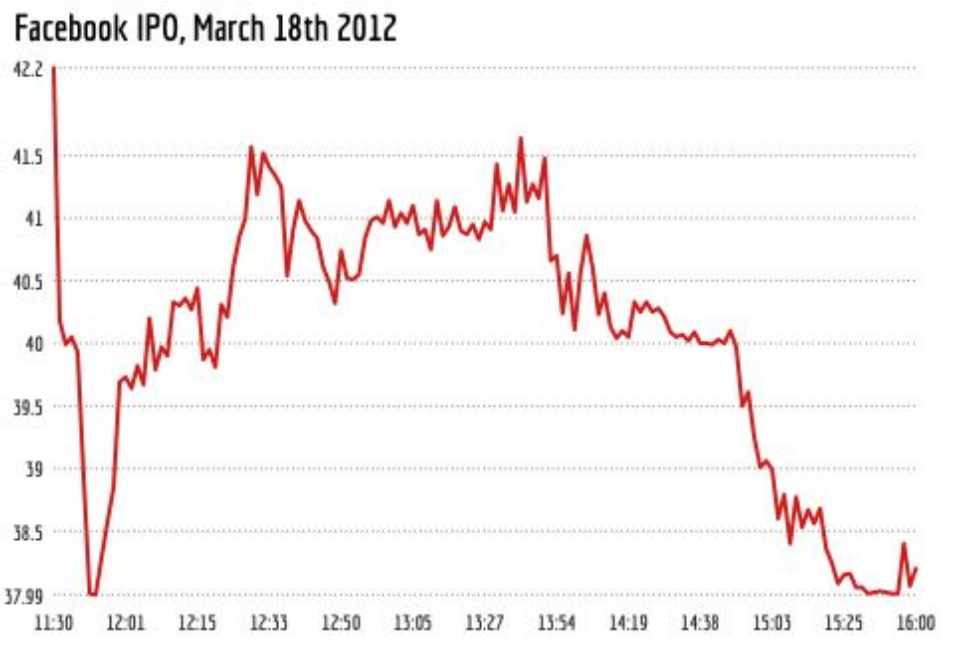
- Shows 5-number Summary for a SINGLE variable.
- Hard to view distribution at large.
- Great when comparing the scale of different features.

Ideal Use Case:

Multiple Box-Plots Viewed Side By Side.

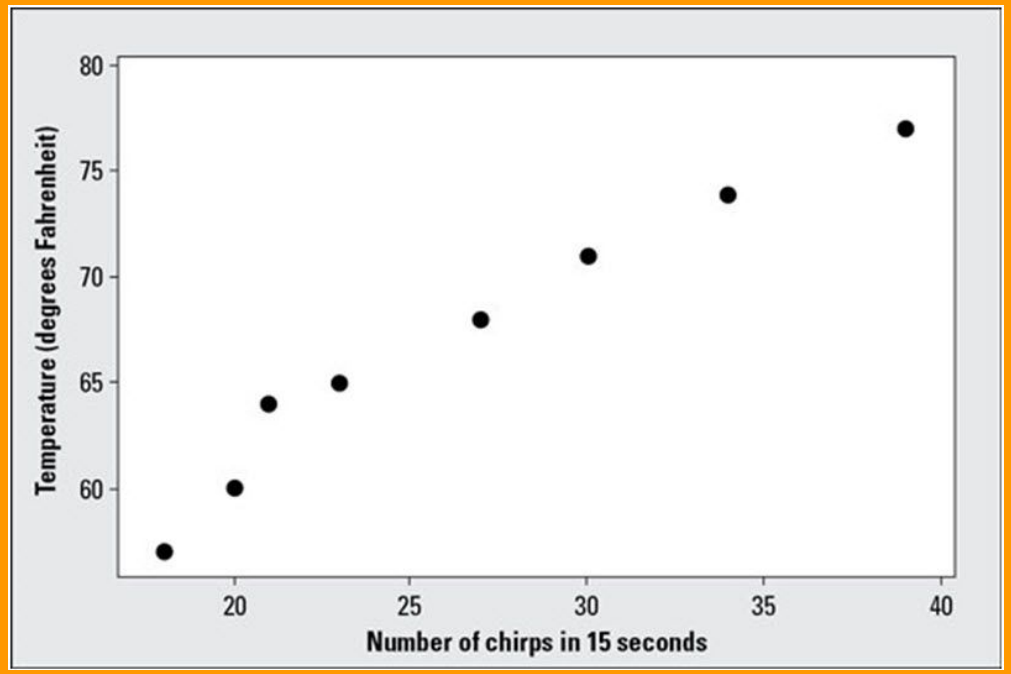


Line Plot



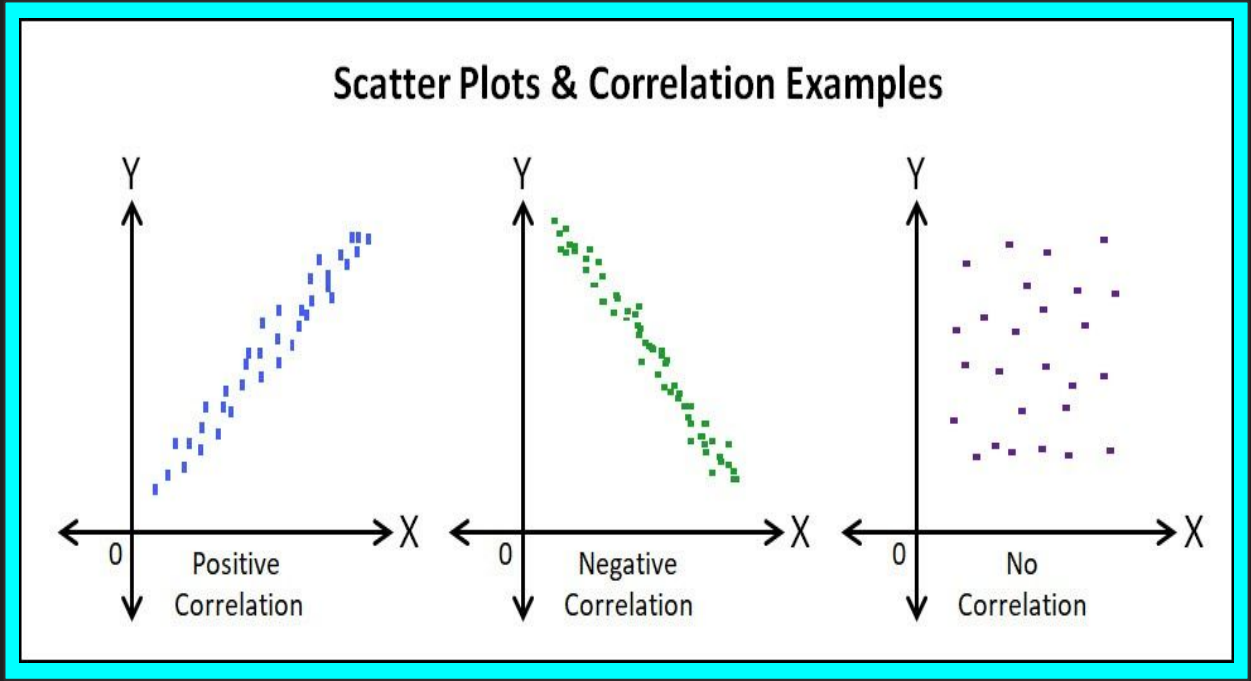
- Generally used to describe the relationship between a SINGLE variable and time.
- Great for finding trends in the data.
- Example:
  - Facebook Stock Prices x Hour of Day

Scatter Plot



- Shows relationship between TWO variables.
- Generally used to discern a correlation between 2 variables.
- Example:
  - Temp vs. cricket chirps in seconds.

Scatter Plot Considerations



Linear: Can I draw a line through the data points?

Strength: How well does my line encompass points on graph?

Direction: Does the relationship have a positive or negative slope?

Outliers: Are there any outliers that may affect the relationship?