**Scatterplots and Correlation**
Please also [read this document](https://chartio.com/learn/charts/what-is-a-scatter-plot/) from Chartio

Scatterplots are used to see if two variables are related to each other. Each dot on the scatterplot is an individual data point such as a person, store, or product. For example the height and weight of 100 children can be displayed as dots on a chart. You could for example plot the X value as the child’s height, and the Y-value as their weight. Together the height and weight values can be used to plot a dot on a chart. If you plot 100 differet data points (ie 100 children) then the data can tell a story.



You may even find that height and weight are related and even correlated, but the correlation is not perfect. While children put on weight as they get taller, sometimes they get taller and remain the same weight, or even lose some weight. Correlation is a measure of how closely two variables co-vary. It is common to draw a regression line to plot the relationship between the data variables.

Scatterlplots are useful to see if two measures are correlated. Each dot represents one row of data in your dataset or perhaps a summarized value (such as zip code or store number). For example weather temperature and ice cream sales. It makes sense that consumers eat more refreshing ice cream during the hotter months, but is this relationship true for all regions? Is this relationship getting stronger over the years, or weaker?

In the scattergram above you can see that there is not much of a relationship between the X and Y variables for the red sub-sample, and a stronger correlation for the green and bluw sub-samples. While looking at how closely two numbers are correlated is interesting and identifying diofferences in sub-samples is also insightful, we much remember that the covariance may be due to chance, as correlation does not mean causation.

If you consider the x-axis a predictor variable then the Y-axis is the outcome variable so you can see the ‘influence’ of X on Y. This is over-reading into the data however, as the data coviance can be due to chance and not repeatable.

In the picture on the right the top-left scatterplot suggests the data are tight and ‘predictable.’ Higher levels of the X-axis variable are associated with higher levels of the Y-axis in most cases, and you could even wager a prediction of what the Y-value could be (ie weight) if given an X-value (height).

The top right scattergram shows the opposite phenomenon, most of the time an increase in the X-value is associated with lower levels of the Y-value. The points are not closely packed together on a line though so the correlation (and any SWAG prediction) is weak and not a good bet.

The scatterplot on the bottom left-hand side does not form into any line going in any direction so the data can be interpreted as having no correlation. The data does not co-vary in any predictable or repeatable way.

In the bottom-right scatterplot, the relationship is called curvilinear. An increase in X is associated with an increase in Y but only yup to a point. For example people eating more potato chips can aloso report being more happy, but only up to a point then the trend reverses. Eating more potato chips makes you feel worse when you start to eat too much. Many relationships are curvilinear and therefore fascinating. While there is often a point of diminshing returns (up then flat-line) here clearly the relationship reverses course.

**Bubble Charts are Scattergrams that have a Third Variable Plotted**[This website from chartio](https://chartio.com/learn/charts/bubble-chart-complete-guide/) also helpful



The size of the dot can be varied to add more information to the chart. Alternatively you can change the darkness of the dot to indicate a higher number on a third value.

PowerBI and Tableau have excellent scatterplot facility and even motion option to look at the relationship over time. The bouncing bubbles that trend up and to the right indicate a positive relationship over time.

References

<https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/regression-analysis/scatter-plot-chart/>

<https://www.mathsisfun.com/data/scatter-xy-plots.html>