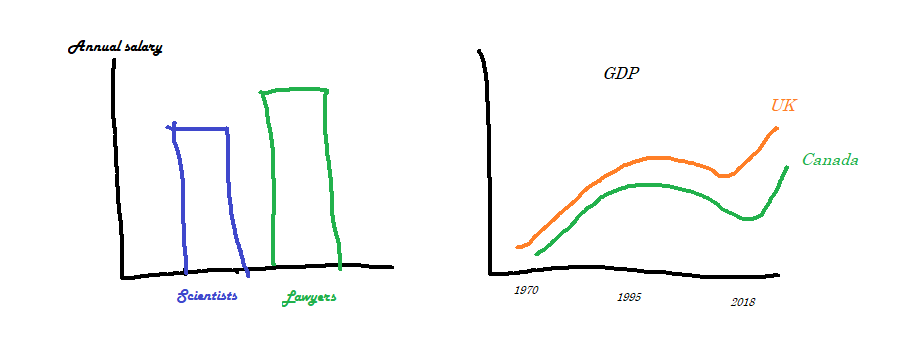
There are 6 guiding principles that can steer data visualization:

1. Comparison
2. Composition
3. Distribution
4. Deviation
5. Relationship
6. Trend

# **1. Comparison**

A comparison tries to set one set of variables apart from another, and display how those variables interact, like the number of visitors to five competing websites in a certain period of time.This kind of visual material compares multiple variables in datasets or multiple categories within a single variable.



These 4 chart types can best achieve this:

****Column chart****Column charts are the standard for showing chronological data, such as growth over specific periods of time, and for comparing data across categories. Grouping a small number of sub-sets (max. 6) gives additional information about the distribution of data over smaller units, i.e. quarters or regional offices.

****Bar chart****

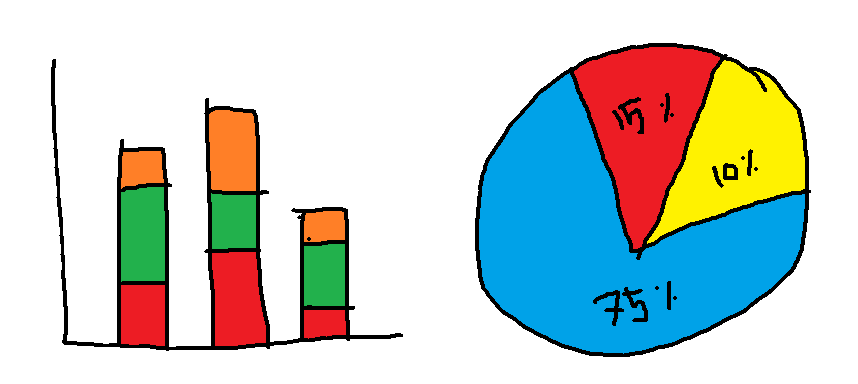
This chart type is typically suitable for comparing the size of data (horizontal axis) for a medium to large data set (vertical axis). Keep data in an order that makes sense. Either list by value or, if that’s not the strength, choose a logic for the labels that makes sense, like listing them alphabetically. Added benefit of this chart type is that it accomodates for long data labels.

****Line chart****The purpose of a line chart is to show trends, accelerations (or decelerations) and volatility. They display relationships in how data changes over a period of time. Choose a line chart also when the number of data points is very high and a column or bar chart would look too cluttered.

****Two-axis chart****This chart type combines two chart types in a single chart: a column chart series and a line graph series. There are two Y-axes: each axis has its own unit and magnitude, and each data series conforms to one of these axis. This chart type is ideal if you have two (or more) variables that you want to show in the same period of time, like temperature and rainfall, and how they correlate over time.

# ****Composition****

Have you heard about stacked bar charts? But I’m sure you know what a pie chart is.The purpose of these charts is to show the composition of one or more variables in absolute numbers and in normalized forms (e.g. percentage).Composition charts are some of the old school visualization techniques that nowadays have limited use cases (do you really need a pie chart to show a composition of yellow 10% and red 15%?). Nevertheless, sometimes they can present information in a visually aesthetic and familiar, vintage fashion.



These 4 chart types can best achieve this:

****Pie or donut chart****This chart type is typically suitable when you need to visualize a part to whole relationship. Pie charts work best if you have no more than 6 categories and when the categories have clearly distinguishable values. The donut chart merely adds a hole to the middle of the pie chart.

****Stacked column chart****This is a variation of the column chart where each column is broken into smaller units (i.e. quarters or regional offices) showing the development of the whole as well as that of the smaller units within the data set. Use a 100% stacked column chart when you want to focus on the differences between units.

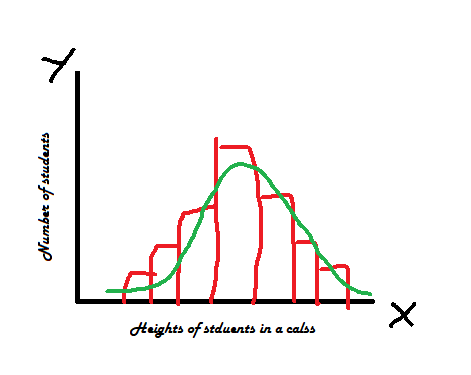
****Stacked area chart****This is a variation of the stacked column chart showing the development of the whole as well as that of the smaller units within the data set in a more continuous flow. Use a 100% stacked area chart when you only want to show relative differences.

****Waterfall chart****This chart type is used to visualize changes in performance. A typical use is to show how an initial value is increased and decreased by a series of intermediate values, leading to a final value.

# ****3.Distribution****

An important concept in statistics and data science is distribution. Distribution generally refers to the probability of occurrence of an outcome. In a distribution of 100 coin flips how many will get heads and how many tails? Frequency distributions like this are presented in histograms or curves.

Below is a representation of studnets’ heights distribution in a swimming class. The x-axis shows different height categories and y-axis has the number of students in each category.



These 4 chart types can best achieve this:

****Column chart****Column charts are the standard for showing chronological data, such as growth over specific periods of time, and for comparing data across categories. Grouping a small number of sub-sets (max. 6) gives additional information about the distribution of data over smaller units, i.e. quarters or regional offices.

****Scatter plot****This chart type works best to find and show the correlations in a fairly large data set. The data sets need to be in pairs with a dependent variable and an independent variable. The dependent (the one the other relies on) becomes the y axis and the independent, the x. Adding a trend line will help show the correlation and how statistically significant it is.

****Line chart****The purpose of a line chart is to show accelerations (or decelerations) and volatility. They display relationships in how data changes over a period of time. Choose a line chart also when the number of data points is very high and a column or bar chart would look too cluttered.

****Two-axis chart****This chart type combines two chart types in a single chart: a column chart series and a line graph series. There are two Y-axes: each axis has its own unit and magnitude, and each data series conforms to one of these axis. This chart type is ideal if you have two (or more) variables that you want to show in the same period of time, like temperature and rainfall, and how they change over time.

# **4. Deviation**

You want to show which values deviate from the norm. This is a variation of one comparison, but aims to highlight the outliers or problem areas that need special attention.

These 4 chart types can best achieve this:

****Column chart****Column charts are the standard for showing chronological data, such as growth over specific periods of time, and for comparing data across categories. Deviations will be clearly visible if one or more columns are much shorter or longer than the others.

****Bar chart****This chart type is typically suitable for comparing the size of data (horizontal axis) for a medium to large data set (vertical axis). Keep data in an order that makes sense. Deviations will be clearly visible if one or more bars are much shorter or longer than the others.

****Line chart****Line charts are more suitable than column charts or bar charts for showing deviations when the differences are fairly small or if there are a lot of data points.

****Area chart****  
This is a visual alternative for the line chart, showing not only the development of data on a graph over time but also the volume of that data.

# **5. **Relationship****

Trees grow taller as they get older in the early years. That’s a relationship between two variables — height and age.

*height = f(age)*

In another example, the price of a house depends on the number of beds, number of bathrooms, location, square footage etc. This is a relationship between one dependent and many explanatory variables.

*price = f(beds, baths, location, area)*If you look at a data set just as numbers, there is no way to identify these relationships. But in fact, you can, without going into complex statistical analysis, with the help of a good visualization.

These 3 chart types can best achieve this:

****Scatter plot****This chart type works best to find and show the correlations between two variables. The data sets need to be in pairs with a dependent variable and an independent variable. The dependent (the one the other relies on) becomes the y axis and the independent, the x. Adding a trend line will help show the correlation and how statistically significant it is.

****Bubble chart****This chart type is a variation on the scatter plot, adding a third (dependent) variable which indicates volume or size.

****Two-axis chart****This chart type combines two chart types in a single chart: a column chart series and a line graph series. There are two Y-axis: each axis has its own unit and magnitude, and each data series conforms to one of these axis.

# **6. Trend**

You want to understand the trend over time of some data variables. You will need to plot the time line on the X-axis.

These 4 chart types can best achieve this:

****Line chart****The line chart displays relationships and the resulting derivative trend of how data changes over a period of time. Choose a line chart also when the number of data points is very high and a column or bar chart would look too cluttered.

****Column chart****Column charts are the standard for showing chronological data, such as growth over specific periods of time, and for comparing data across categories. Grouping a small number of sub-sets (max. 6) gives additional information about the distribution of data over smaller units, i.e. quarters or regional offices.

****Scatter plot****This chart type works best to find and show the correlations between two variables. The data sets need to be in pairs with a dependent variable and an independent variable, in this case time. The dependent (the one the other relies on) becomes the y axis and time becomes the x. Adding a linearly plotted average will show the trend.