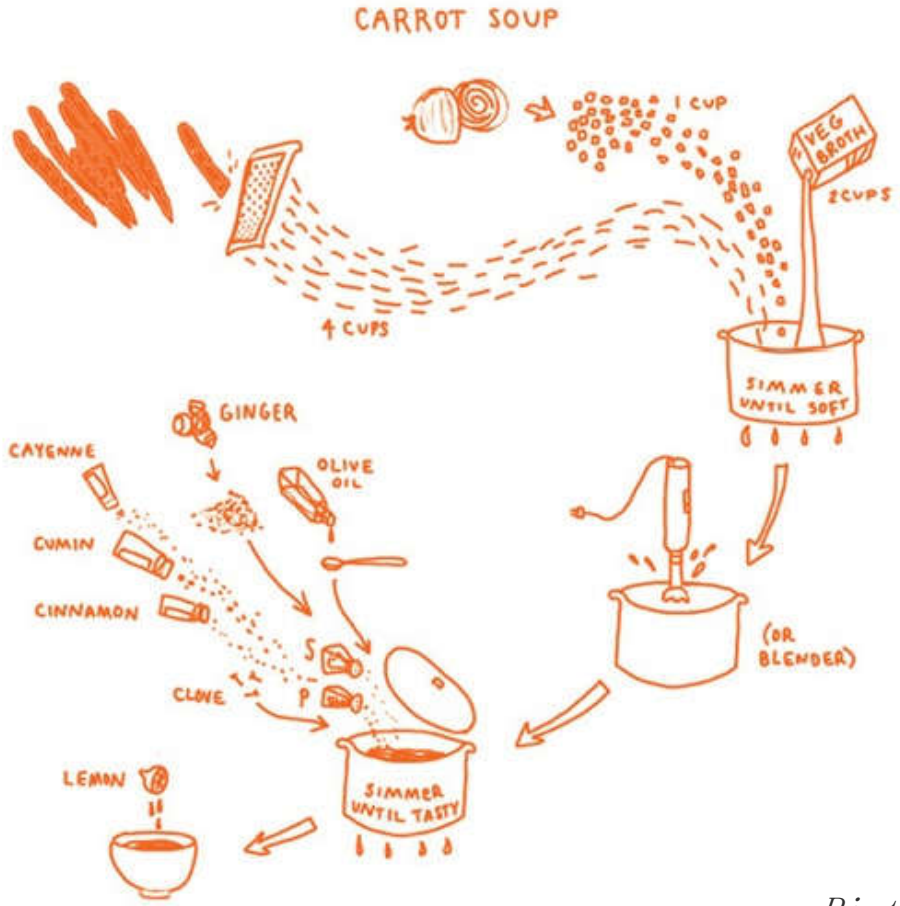
Data Exploration & Visualization

Visualization - Understanding

- One of the most rewarding and useful things you can do to understand your data is to visualize it in a pictorial format. Visualizing your data allows you to interact with it, analyze it in a straightforward way, and identify new patterns, making your would-be complex data more accessible and understandable. The way our brains processes visuals like shapes, colors, and lengths makes looking at charts and graphs more intuitive for us than poring over spreadsheets. Author and illustrator Felicita Sala demonstrates this masterfully:

Visualization Example



Picture Cook: See, Make, Eat by [Katie Shelly](#) (Ulysses

What does it tell you?

- What do you think that she's trying to communicate to you here?
- So, to me rather than a traditional recipe, these visualizations they communicate a couple things, right?
- A couple key things, the **ingredients** that are needed.
- The **amounts of ingredients** that are needed, and
- Even the **order** that they are to be introduced into the recipe, all of this in one fell swoop.

Why is Data Visualization Important?

- Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets.
- Data visualization uses visual data to communicate information in a manner that is universal, fast, and effective. This practice can help companies identify which areas need to be improved, which factors affect customer satisfaction and dissatisfaction, and what to do with specific products (where should they go and who should they be sold to). Visualized data gives stakeholders, business owners, and decision-makers a better prediction of sales volumes and future growth.

Why is Data Visualization Important?

- **Correlations in Relationships:** Without data visualization, it is challenging to identify the correlations between the relationship of independent variables. By making sense of those independent variables, we can make better business decisions.
- **Trends Over Time:** While this seems like an obvious use of data visualization, it is also one of the most valuable applications. It's impossible to make predictions without having the necessary information from the past and present. Trends over time tell us where we were and where we can potentially go.
- **Frequency:** Closely related to trends over time is frequency. By examining the rate, or how often, customers purchase and when they buy gives us a better feel for how potential new customers might act and react to different marketing and customer acquisition strategies.
- **Examining the Market:** Data visualization takes the information from different markets to give you insights into which audiences to focus your attention on and which ones to stay away from. We get a clearer picture of the opportunities within those markets by displaying this data on various charts and graphs.
- **Risk and Reward:** Looking at value and risk metrics requires expertise because, without data visualization, we must interpret complicated spreadsheets and numbers. Once information is visualized, we can then pinpoint areas that may or may not require action.
- **Reacting to the Market:** The ability to obtain information quickly and easily with data displayed clearly on a functional dashboard allows businesses to act and respond to findings swiftly and helps to avoid making

Classification of Charts

Comparison

- Comparison charts are used to compare one or more datasets. They can compare items or show differences over time.

Relationship

- Relationship charts are used to show a connection or correlation between two or more variables.

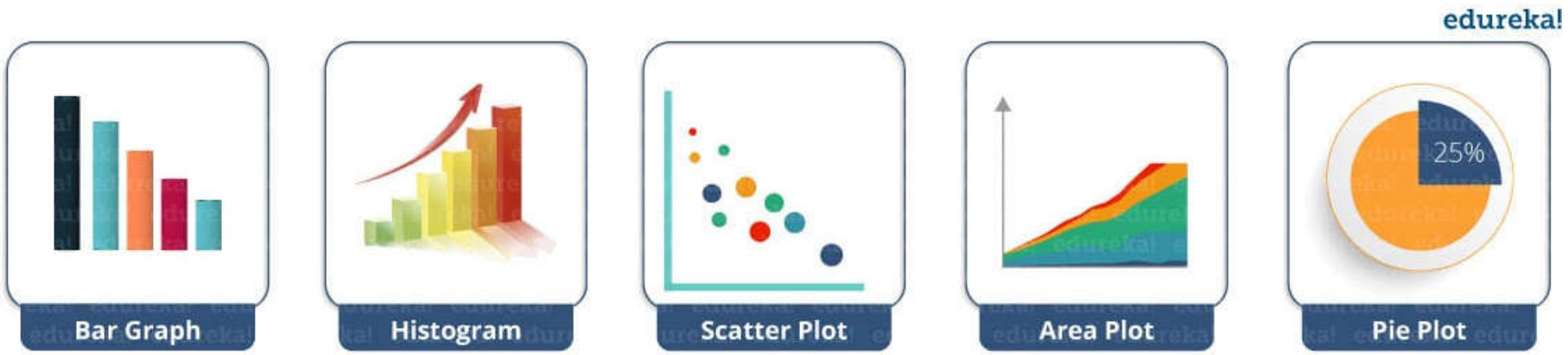
Composition

- Composition charts are used to display parts of a whole and change over time.

Distribution

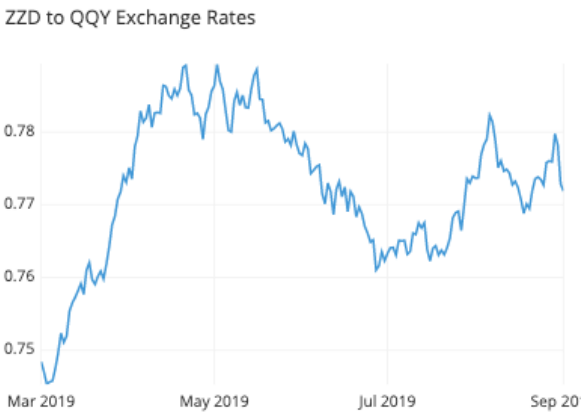
- Distribution charts are used to show how variables are distributed over time, helping identify outliers and trends.

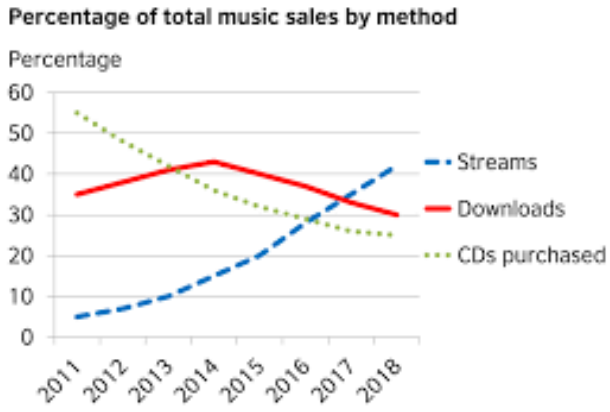
Types of Charts



Line Chart

- A line chart is a chart that shows the data as a series of points connected by lines. They are often used to show trends over time, such as changes in revenue or website traffic etc.
- Multiple lines can also be plotted in a single line chart to compare the trend between series. A common use case for this is to observe the breakdown of the data across different subgroups.





Best practices for creating line charts

- Clearly label your axes - Make sure the viewer knows what they are evaluating.
- Remove distracting chart elements - Grids, varying colors, and bulky legends can distract the viewer from quickly seeing the overall trend.
- Zoom in on the y-axis if your data set starts above zero - In certain cases, changing the scale of the y axis makes it easier for.
- Avoid comparing more than 5-7 lines - You don't want your chart to become cluttered or hard to read. Visualize the data you need to tell your story, nothing more.

Bar Chart Vs Histogram

- Discrete Data
- Frequency/Occurrences
- Order of Bars could be changed

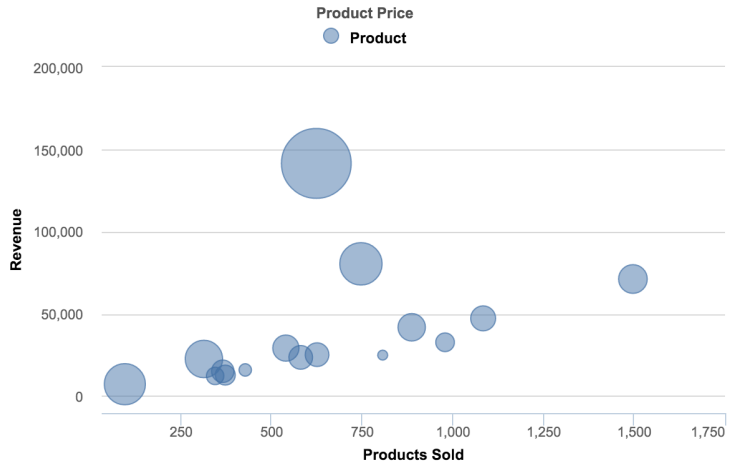
- Continuous Data
- Values grouped in Bins
- Order of Bars not changeable

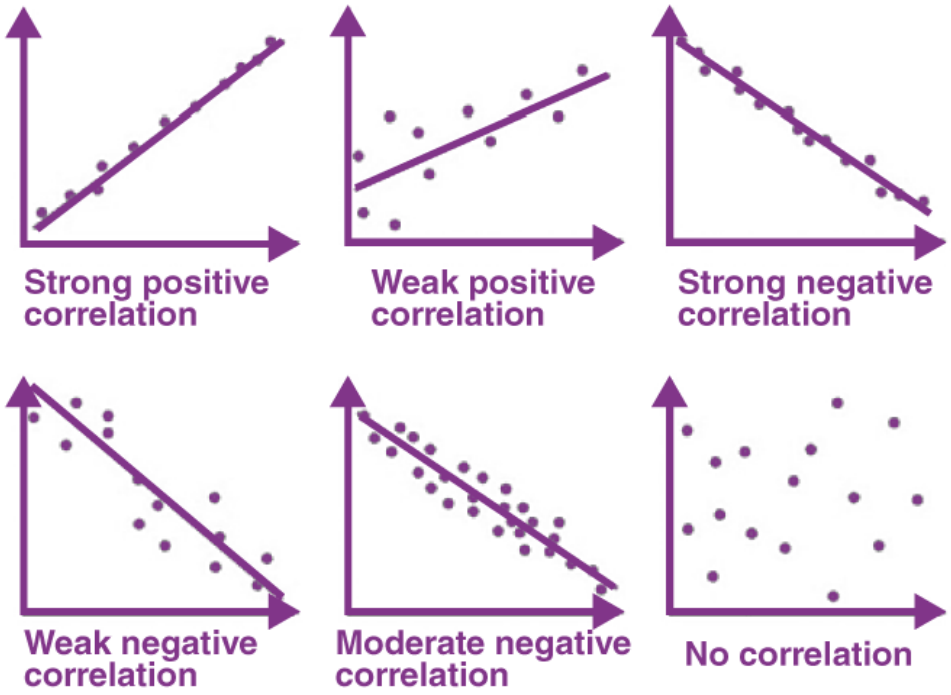
Best practices for creating bar and column charts

- Start the y-axis at zero - Our eyes are sensitive to the area of bars on a chart. If those bars are truncated, the viewer might draw the wrong conclusions.
- Label the axes - Labeling the axes gives your viewer context.
- Put value labels on bars - This helps to preserve the clean lines of the bar lengths.
- Avoid using too many colors "rainbow effect". Using a single color, or varying shades of the same color, is a much better practice. You can highlight one bar in particular if that is the message you want to get across.

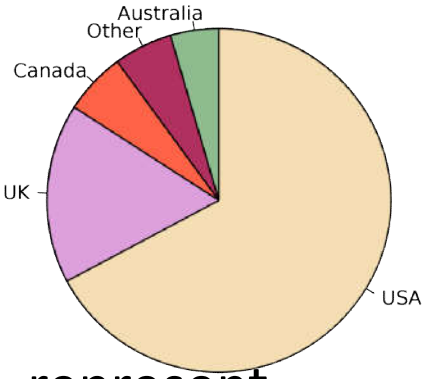
Scatter plot

- A scatter plot is a chart that uses dots to represent individual data points.
- They are often used to show the relationship between two variables, such as the correlation.
- Scatter plots are particularly useful for identifying patterns or trends in the data
- Scatter plots are also useful for identifying outliers or un usual data points.





Pie Plot



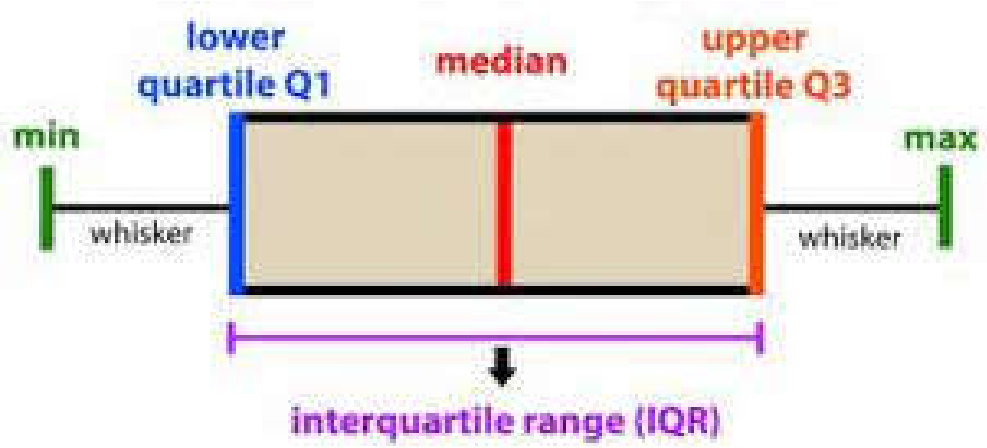
- A pie chart is a circular chart that is divided into slices to represent data as a percentage of the whole. Pie charts are often used to show the relative sizes of different categories or components within a data set.
- However, it is worth noting that pie charts are sometimes criticized for being less effective than other types of charts, such as bar charts or stacked bar charts, in certain situations. This is because it can be difficult to accurately compare the size of different slices of the pie, particularly if there are many slices or if the sizes of the slices are very similar.

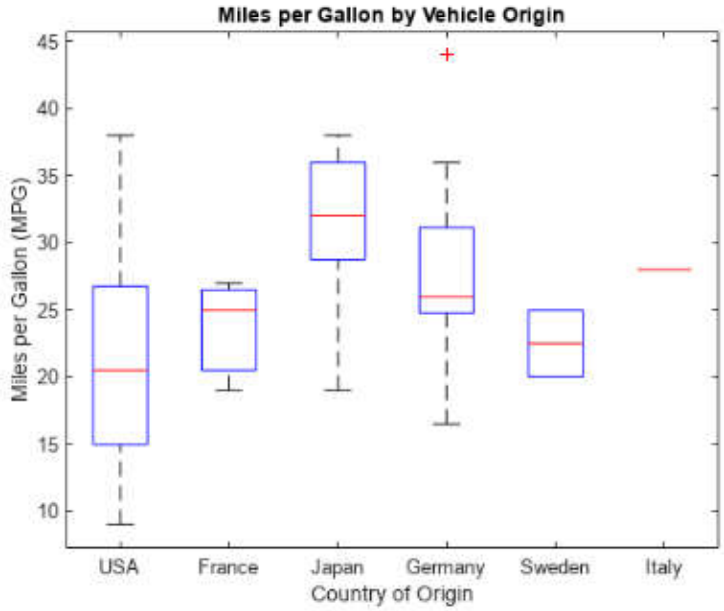
Best practices for creating pie charts

- Make sure your segments add up to 100 - Sounds obvious, but this is a common mistake.
- Keep it clean and consistent. Compare just a few categories to get your point across. If the pie slices have roughly the same size, consider to use a bar or column chart instead.
- Avoid using 3-D imagery or tilt your pie chart - This often makes your data impossible to read, because your viewer is trying to quickly compare angles

Box Plot

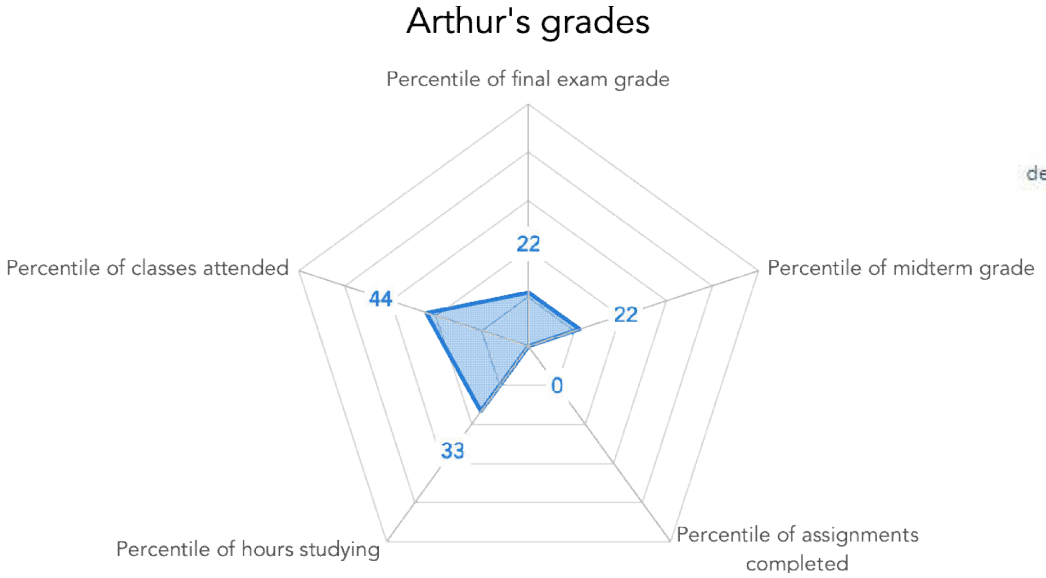
- A box plot, also known as a box-and-whisker plot, is a graphical representation of a dataset that provides a summary of the distribution of the data.
- Box plots are commonly used to Identify outliers, compare the distributions of different data sets, Visualize the shape of the distribution and Assess the spread of the data.

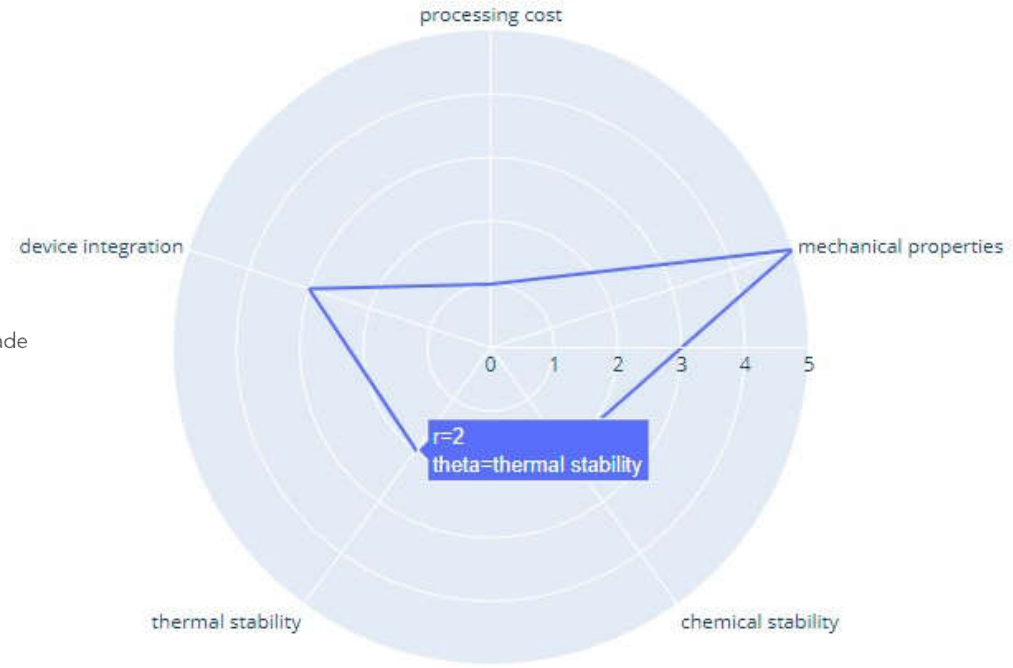




Radar Chart

- It is a graphic displaying data that consists of many independent variables. It is shown as a two-dimensional chart of three or more quantitative variables. These variables are shown on axes starting from the same point.





Word clouds

- Word clouds (also known as tag clouds) are a type of weighted list. Word clouds display text in varying font sizes, weight, or colors to show frequencies or categories. They can be arranged alphabetically or at random. They help people identify trends and patterns that might have been difficult to see otherwise.

Best practices for creating a Word Cloud

- Provide context - Word clouds are visually eye-catching and provide information about frequency but they often don't give the viewer any context.
- Use word clouds to show frequency - Avoid using them to display complex topics like the budget or healthcare crisis.
- Watch your word length - Longer words take up more space and can be misleading.
- Word clouds are great for filtering and analyzing data.
- Avoid making your words too similar in size or color.