Types of Data Visualization

Tushar B. Kute,
http://tusharkute.com

Types of Graphs

- Column Chart
- Bar Graph
- Stacked Bar Graph
- Area Chart
- Dual Axis Chart
- Line Graph
- Pie Chart
- Waterfall Chart
- Scatter Plot Chart
- Histogram
- Funnel Chart
- Heat Map

Bar Graph

- Bar charts are among 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.
- Given that bar charts are such a common chart type, people are generally familiar with them and can understand them easily.
- Examples like this one are straightforward to read.

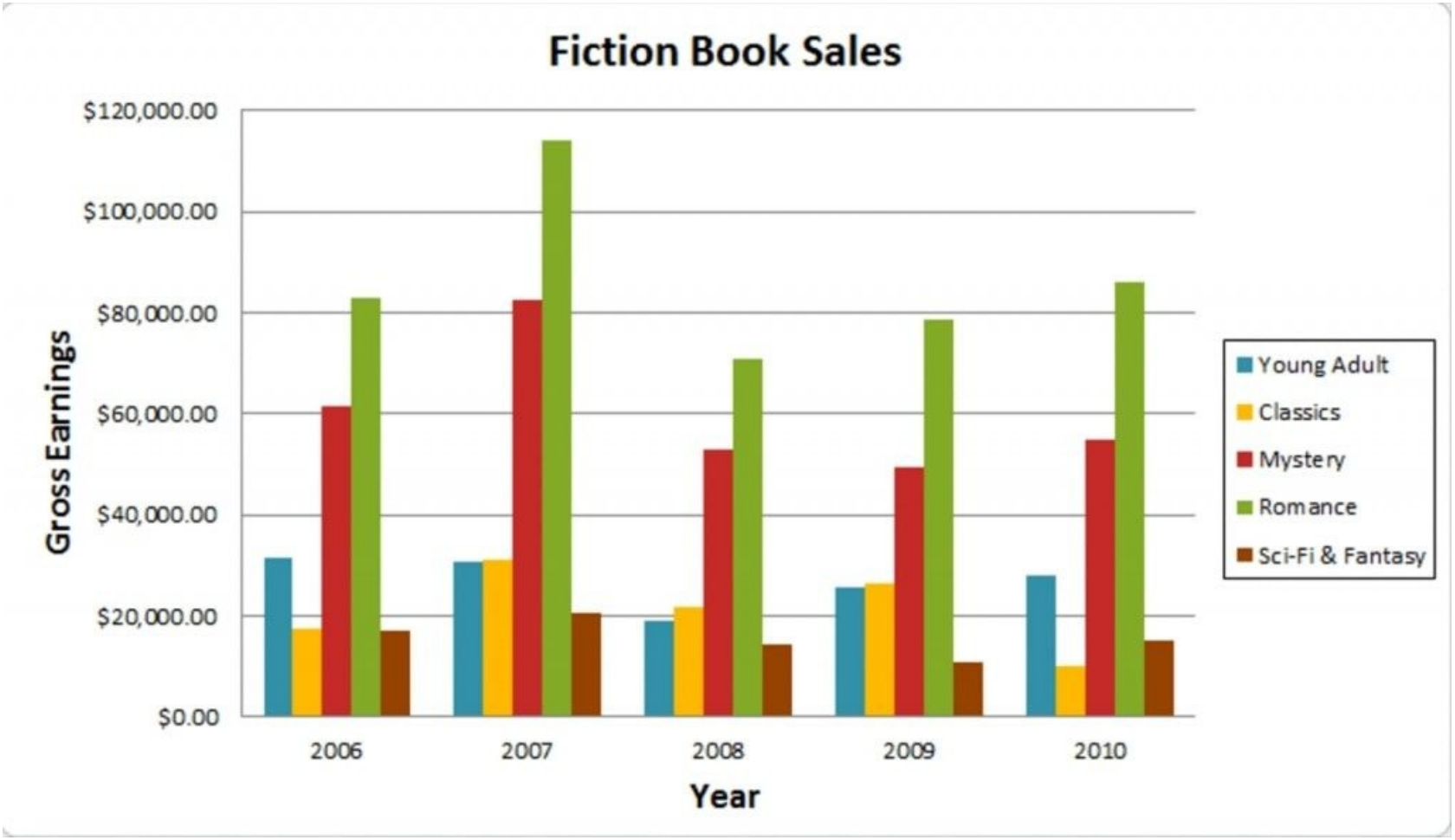
Example:


Bar Chart - One Variable
120
100
80
60
40
20
0
$ in million
81
86
91
103
109
2014
2015
2016
2017
2018
Total Revenue

Bar Graph

- However, please be aware that bar charts can be confusing, too.
- Especially if one uses them to compare several variables. I personally believe that a comparison of more than two variables with a clustered bar chart becomes too cluttered.
- Here is an example of a clustered bar chart that is not exactly crystal clear:

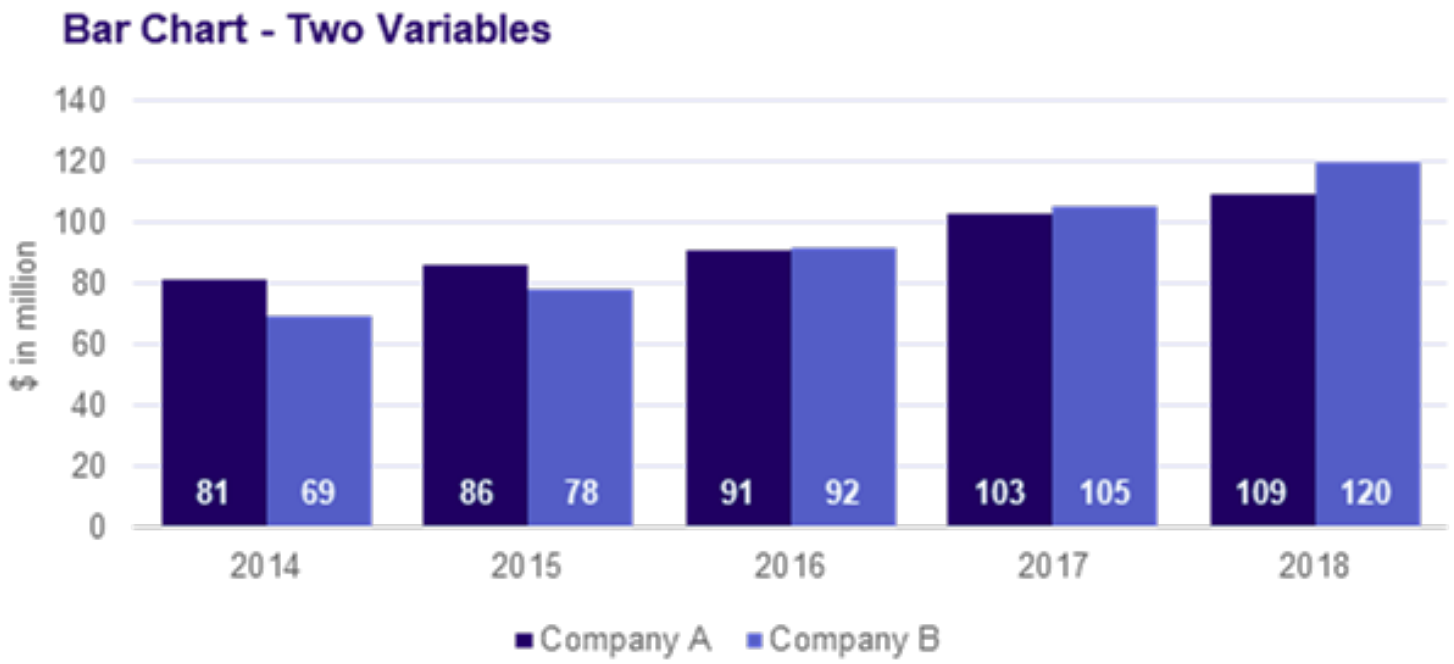
Bar Graph


Fiction Book Sales
$120,000.00
$100,000.00
$80,000.00
$60,000.00
$40,000.00
$20,000.00
$0.00
Gross Earnings
2006
2007
2008
2009
2010
Year
Young Adult
Classics
Mystery
Romance
Sci-Fi & Fantasy

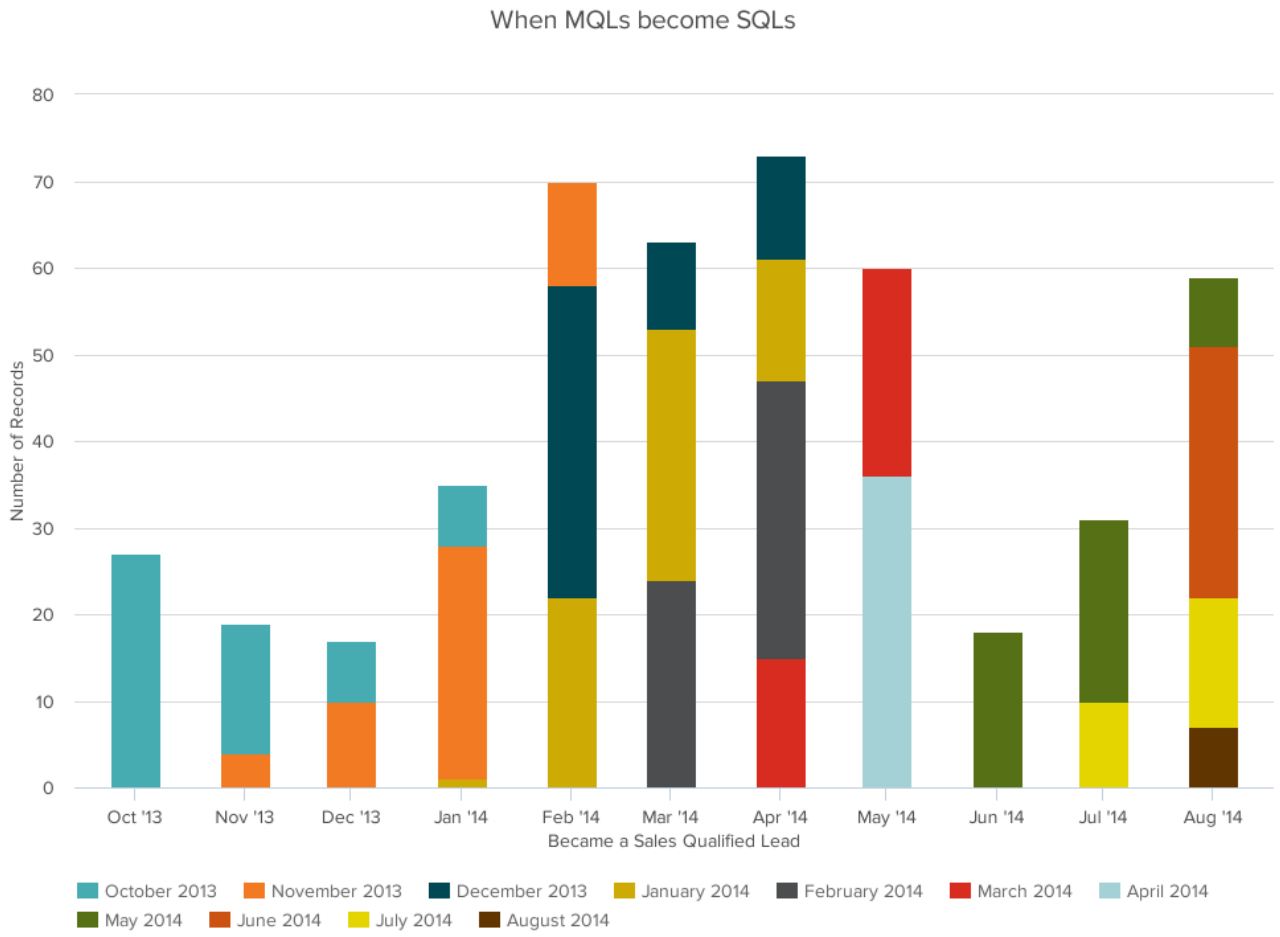
Bar Graph – When to use ?

- Bar charts are nice but limited. We have to consider the type of data we want to visualize and the number of variables that will be added to the chart.
- Bar charts are great when we want to track the development of one or two variables over time.
- For example, one of the most frequent applications of bar charts in corporate presentations is to show how a company's total revenues have developed during a given period.

Bar Chart – two variables



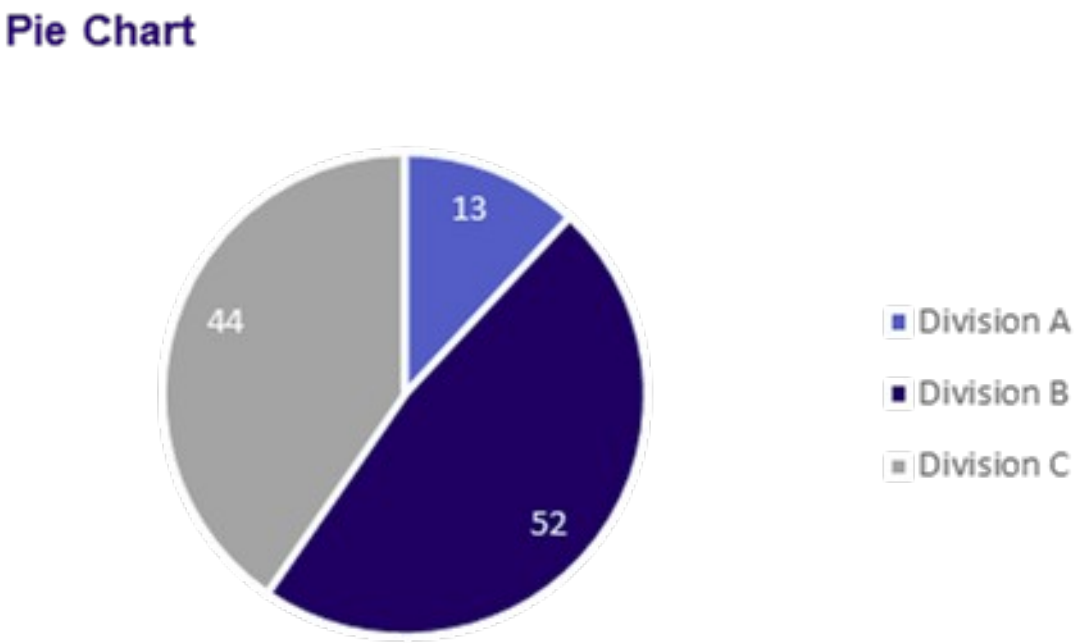
Stacked Bar Chart



Pie Chart

- A pie chart is a circular graph divided into slices. The larger a slice is the bigger portion of the total quantity it represents.
- When to use a pie chart?
  - So, pie charts are best suited to depict sections of a whole.
- What does that mean?
  - 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.

Example:



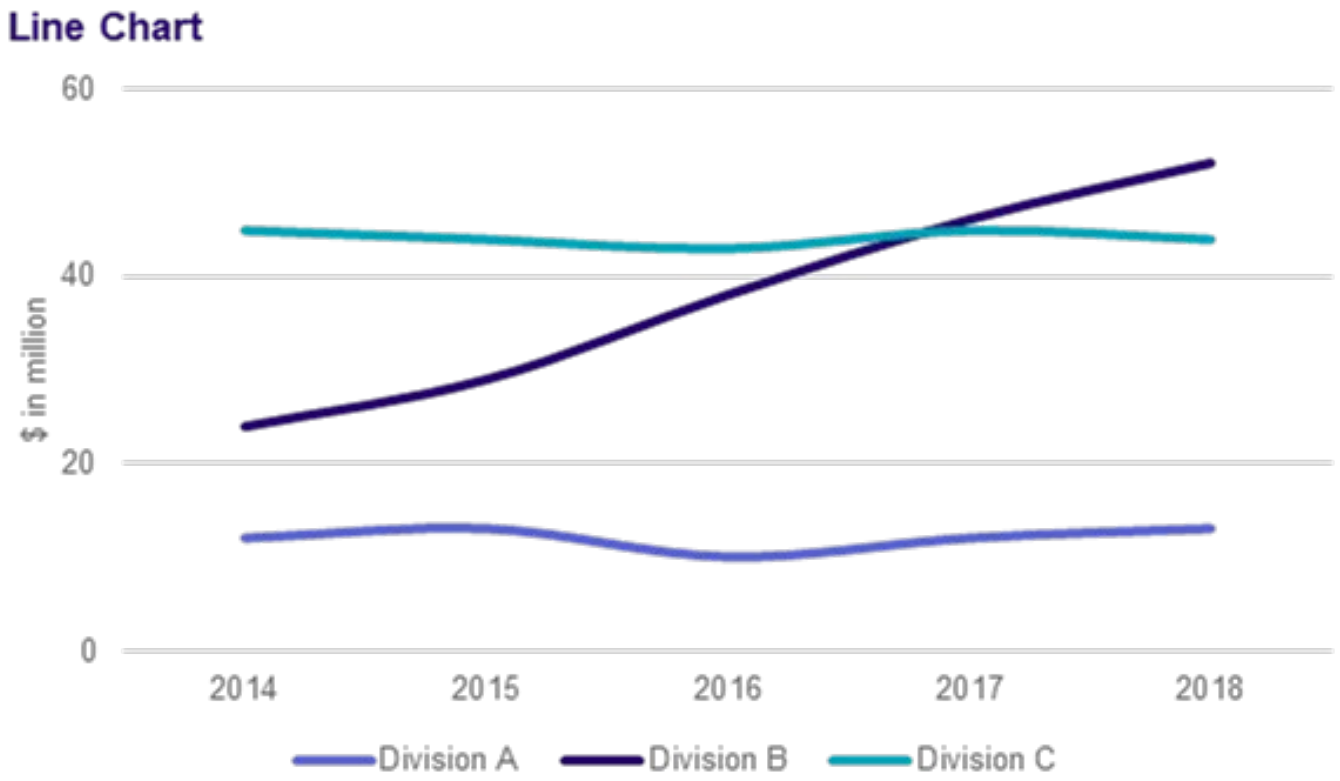
Pie Chart – When to avoid?

- Obviously, we can't use a pie chart in situations when we would like to show how one or more variables develop over time.
- Pie charts are a definite no-go in these cases. Moreover, as mentioned earlier, a pie chart would be misleading if we don't consider all values.
- In the context of our example from earlier, we shouldn't create a pie chart that includes revenue of only two of the firm's three divisions.

Line Chart

- A line chart is, as one can imagine, a line or multiple lines showing how single, or multiple variables develop over time. It is a great tool because we can easily highlight the magnitude of change of one or more variables over a period.
- When to use line charts
  - Remember the awkward ‘Fiction book sales’ chart we saw earlier? Well, a simple line chart would have been much better in that case.
  - A line chart allows us to track the development of several variables at the same time. It is very easy to understand, and the reader doesn’t feel overwhelmed.

Example:


Line Chart
60
40
20
0
$ in million
2014
2015
2016
2017
2018
Division A
Division B
Division C

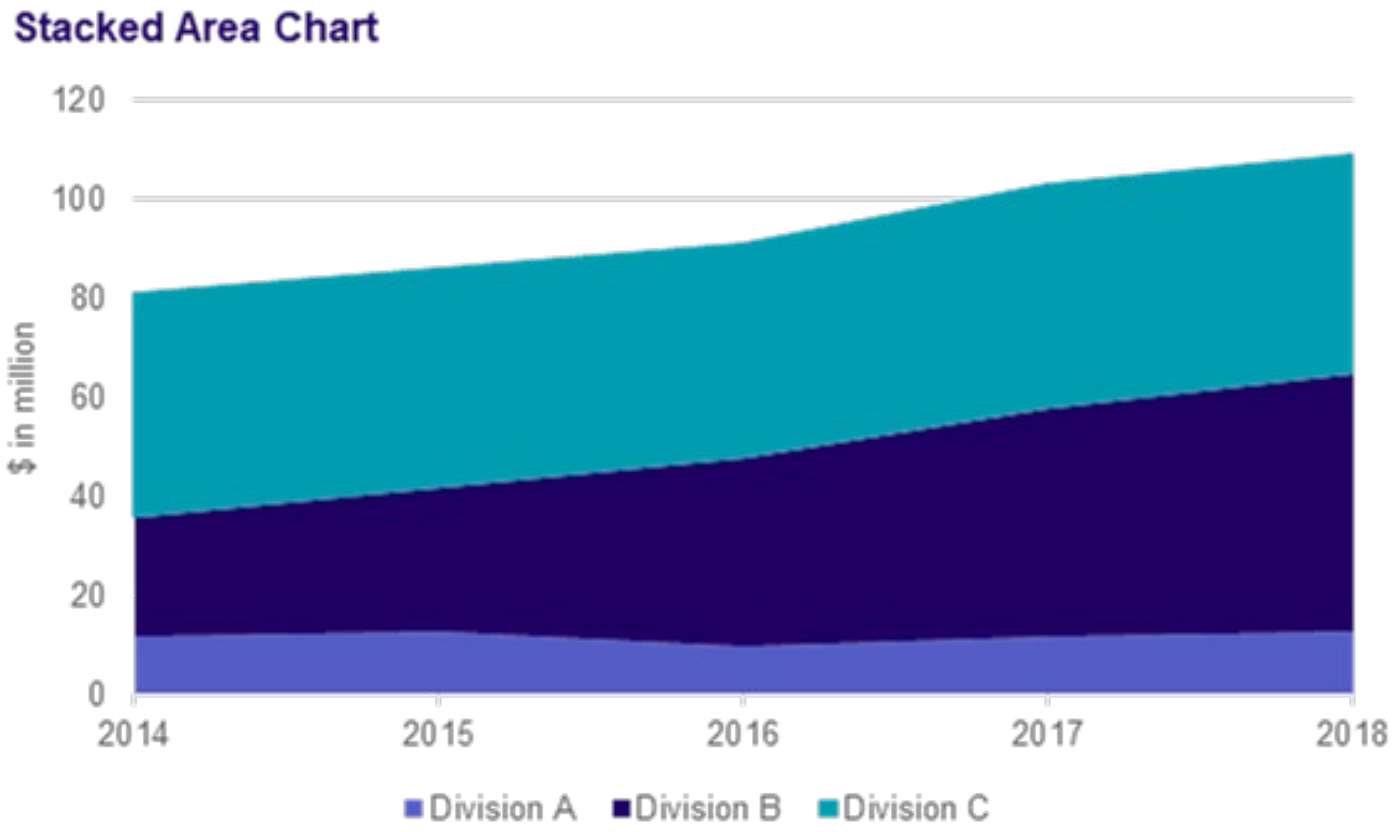
Area Chart

- Area charts are very similar to line charts.
- The idea of an area chart is based on the line chart. Coloured 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.

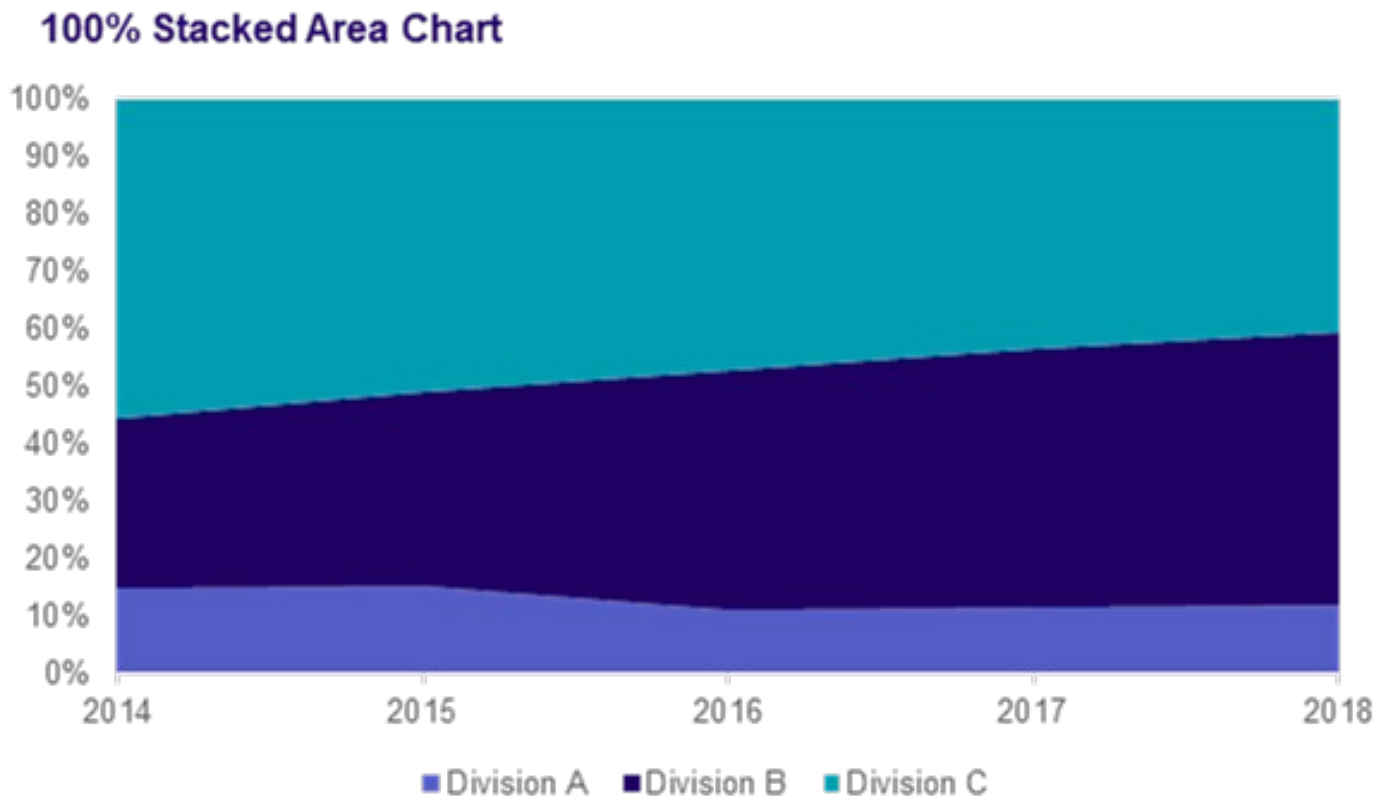
When to use Area Chart?

- Whenever we want to show how the parts of a whole change over time, we should consider an area chart. So, 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 – say, revenues from one division have been growing rapidly while the other two have kept the same level. A stacked area chart is perfect in this case.
- However, if we are interested in the portion of revenue generated by each division and not that much of the total amount of revenues, we can simply use a 100% stacked area chart. This will show each division's percentage contribution over time.

Example:


Stacked Area Chart
120
100
80
60
40
20
0
$ in million
2014
2015
2016
2017
2018
Division A
Division B
Division C

Example:


100% Stacked Area Chart
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
2014
2015
2016
2017
2018
Division A
Division B
Division C

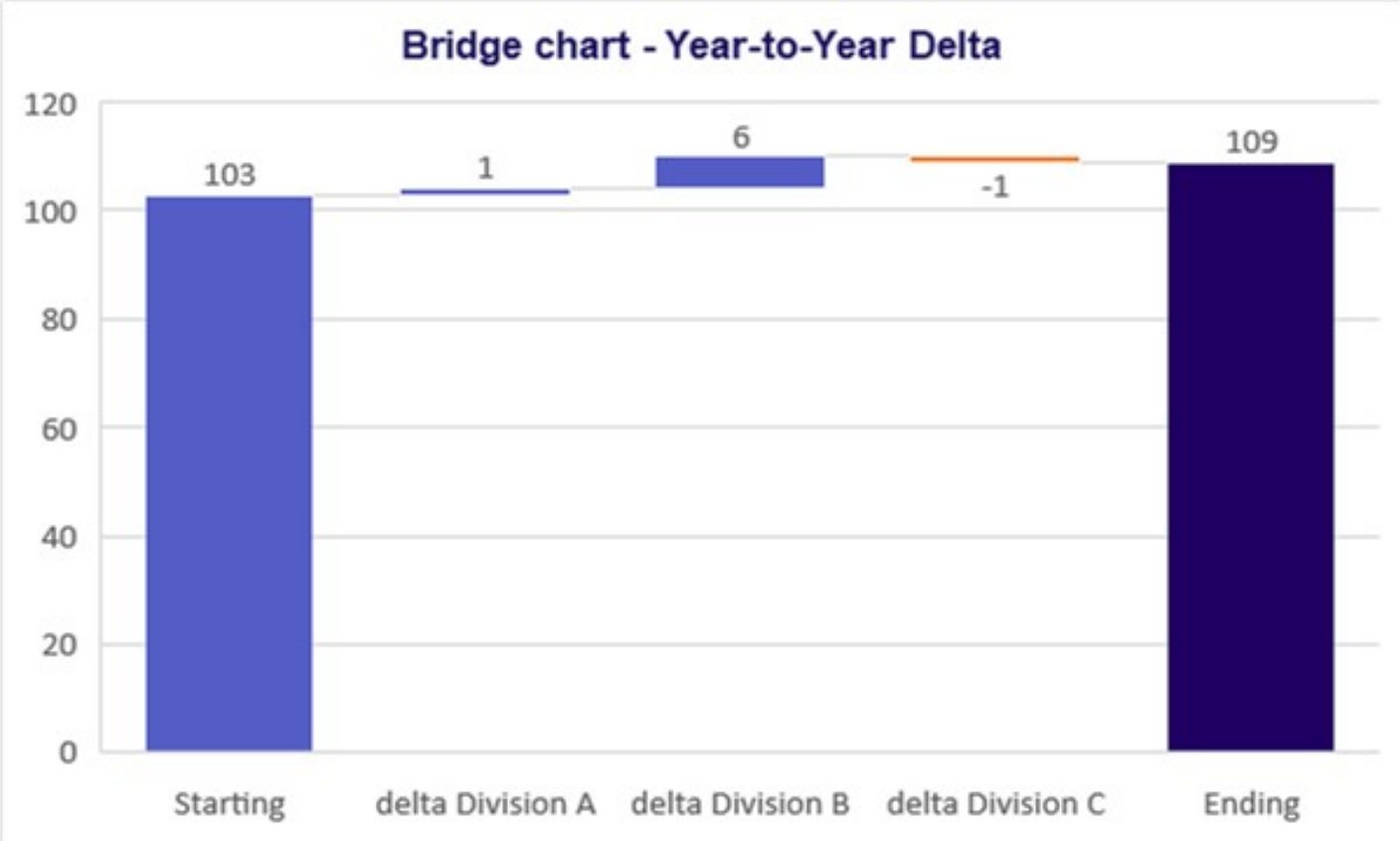
Area Chart – When to avoid?

- Obviously, similarly to line charts, area charts are not suitable for representing parts of a whole over a single period.
- In our example, we can't use an area chart to show the proportion of revenues each division generated in say, 2018 alone. So that's a situation where we can't use an area chart.
- In general, I would stay away from the classical area chart too. It can be very confusing and even Microsoft themselves recommend avoiding it and to consider using a simple line chart.

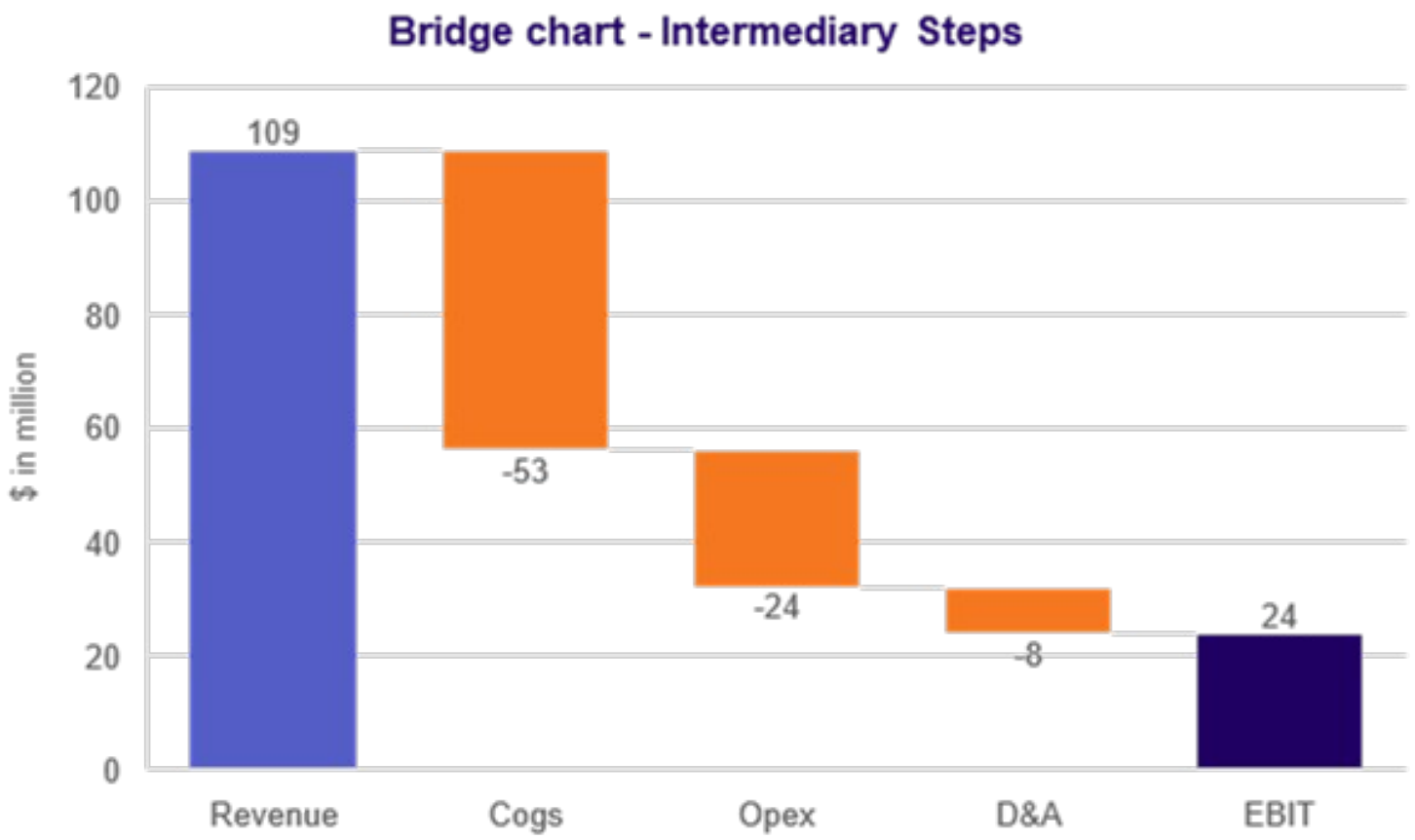
Waterfall Chart

- Waterfall, also known as bridge charts, take their origins from consulting.
- Several decades ago top tier “24/7 at your service” consultants at McKinsey popularized this type of visualization among their clients. And ever since, the popularity of bridge charts has continued to rise.
- Bridge charts are made of bars showing the cumulative effect of a series of positive and negative values impacting a starting and an ending value.

Example:


Bridge chart - Year-to-Year Delta
120
100
80
60
40
20
0
103
1
6
-1
109
Starting
delta Division A
delta Division B
delta Division C
Ending

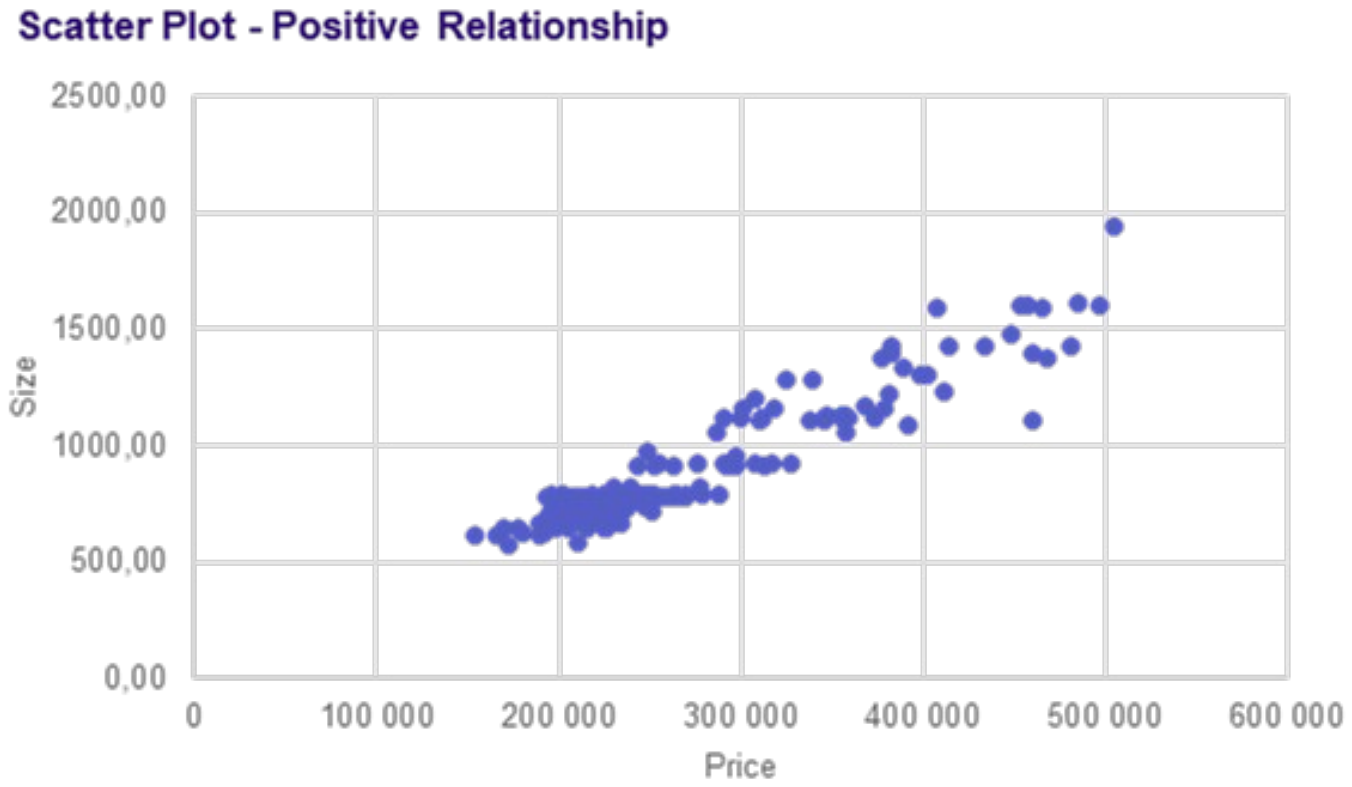
Example:


Bridge chart - Intermediary Steps
$ in million
120
100
80
60
40
20
0
109
-53
-24
-8
24
Revenue
Cogs
Opex
D&A
EBIT

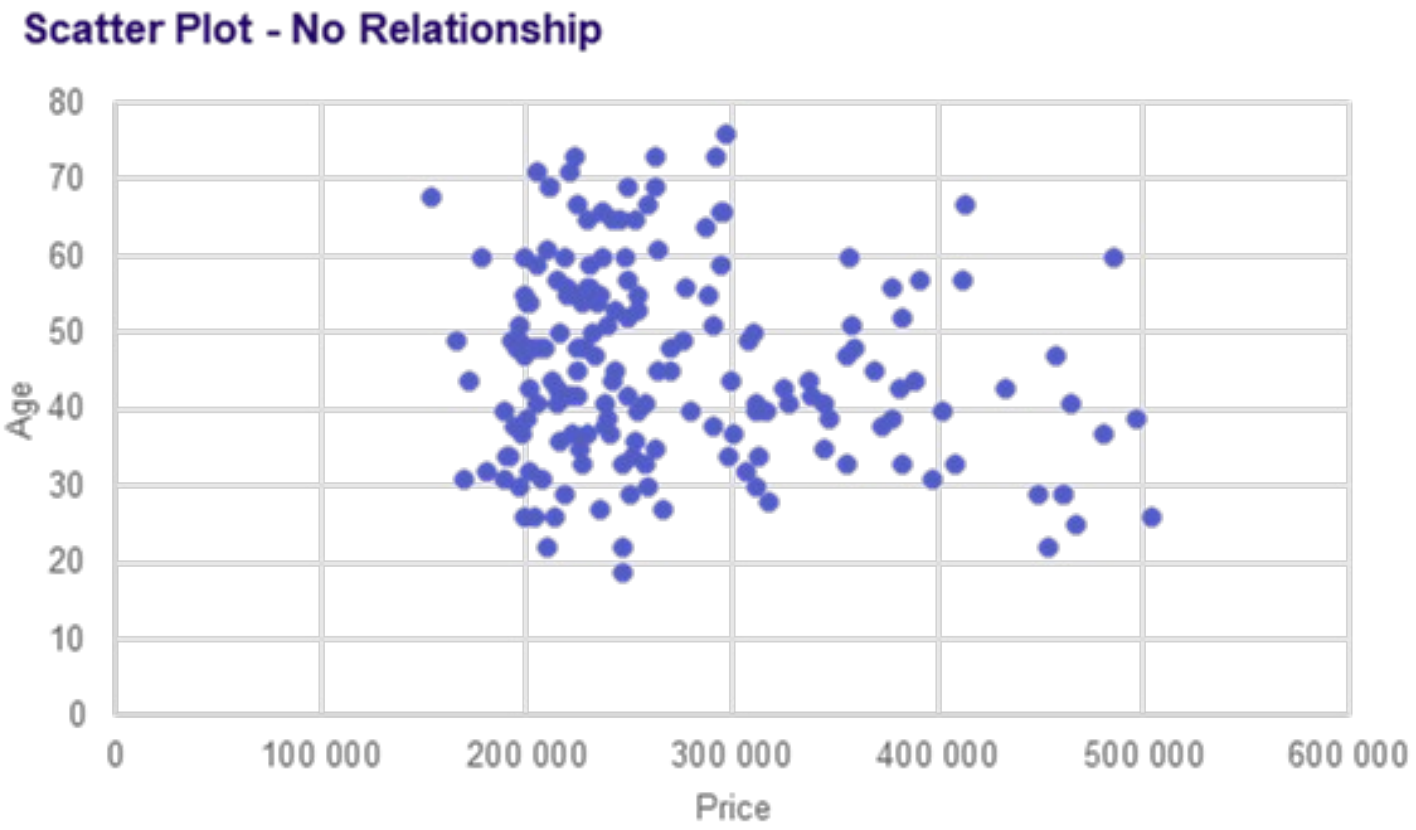
Scatter Plot

- A scatter plot is a type of chart that is often used in the fields of statistics and data science. It consists of multiple data points plotted across two axes.
- Each variable depicted in a scatter plot would have multiple observations. If a scatter plot includes more than two variables, then we would use different colours to signify that.
- When to use scatter plots
  - A scatter plot chart is a great indicator that allows us to see whether there is a pattern to be found between two variables.

Example:


Scatter Plot - Positive Relationship
2500,00
2000,00
1500,00
1000,00
500,00
0,00
Size
0
100 000
200 000
300 000
400 000
500 000
600 000
Price

Example:


Scatter Plot - No Relationship
Age
Price
0
10
20
30
40
50
60
70
80
0
100 000
200 000
300 000
400 000
500 000
600 000

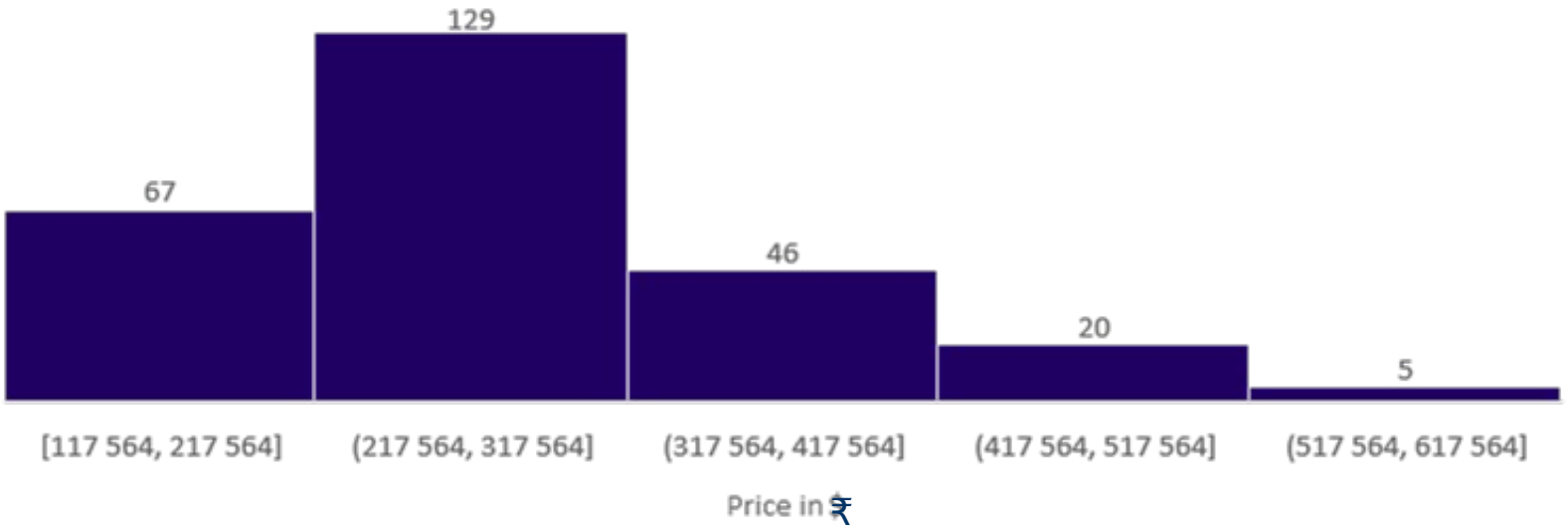
Scatter Plot – When to avoid?

- We can’t use scatter plots when we don’t have bi-dimensional data.
- In our example, we need information about both house prices and house size to create a scatter plot. A scatter plot requires at least two dimensions for our data.
- In addition, scatter plots are not suitable if we are interested in observing time patterns.
- Finally, a scatter plot is used with numerical data, or numbers. If we have categories such as 3 divisions, 5 products, and so on, a scatter plot would not reveal much.

Histogram

- A series of bins showing us the frequency of observations of a given variable. The definition of histogram charts is short and easy.
- Here's an example.
  - An interviewer asked 267 people how much their house cost. Then a histogram was used to portray the interviewer's findings. Some prices were in the range between ₹117-217k, many more in the range ₹217-₹317k, and the rest of the houses were classified in more expensive bins.

Example:

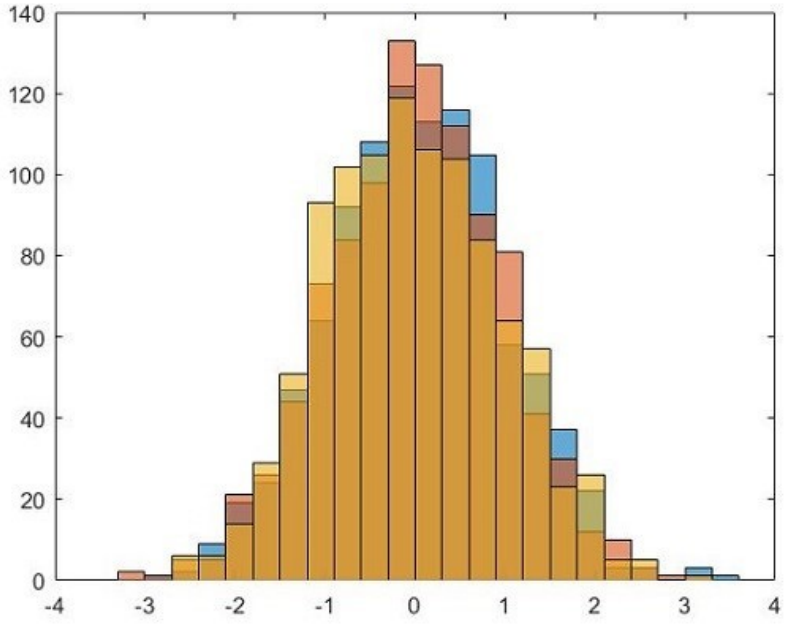

129
67
46
20
5
[117 564, 217 564]
(217 564, 317 564]
(317 564, 417 564]
(417 564, 517 564]
(517 564, 617 564]
Price in ₹

Histogram – When to use?

- 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.

Histogram – When to avoid?

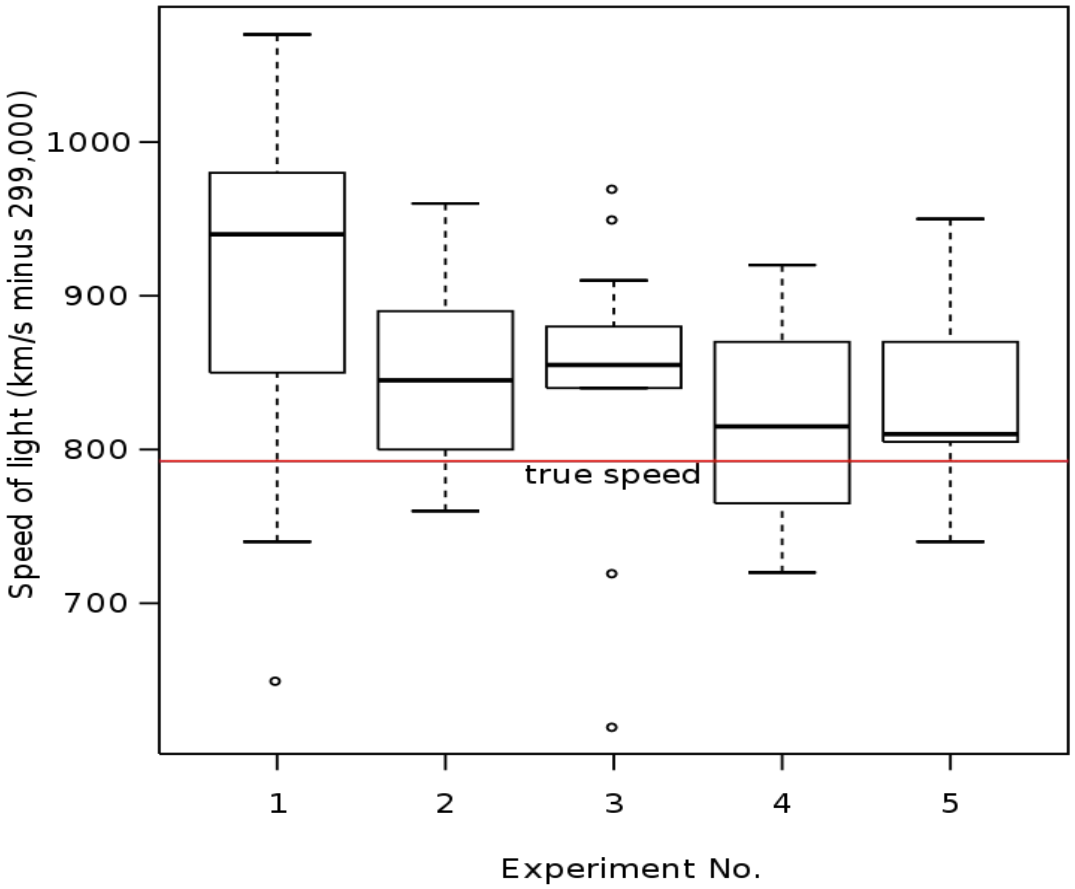
- Be careful when the data you are working with contains multiple categories or variables. Multi-column histograms are among the chart types to be avoided when they look like this.



Box Plot

- A box plot or boxplot is a method for graphically depicting groups of numerical data through their quartiles.
- Box plots may also have lines extending from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram.
- Outliers may be plotted as individual points.

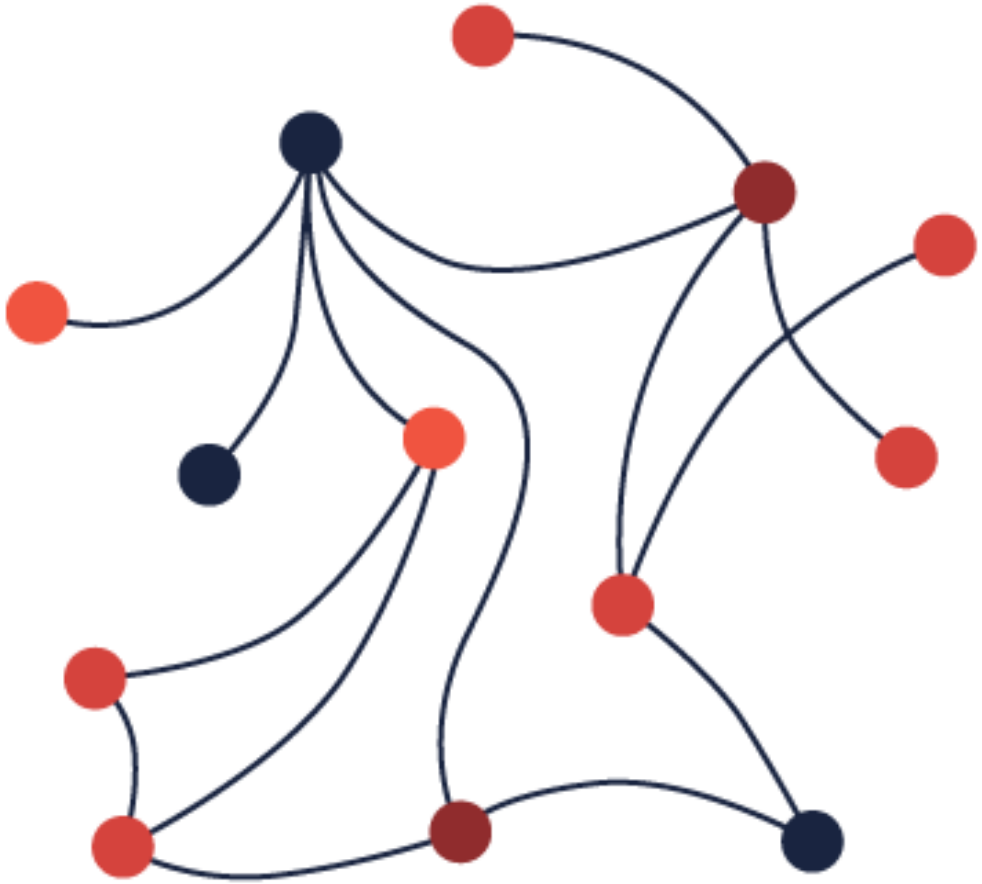
Box Plot


Speed of light (km/s minus 299,000)
1000
900
800
700
true speed
1
2
3
4
5
Experiment No.

Network Visualization

- Network Visualisation (also called Network Graph) is often used to visualise complex relationships between a huge amount of elements.
- A network visualisation displays undirected and directed graph structures. This type of visualization illuminates relationships between entities.
- Entities are displayed as round nodes and lines show the relationships between them.
- The vivid display of network nodes can highlight non-trivial data discrepancies that may be otherwise be overlooked.

Network Visualization

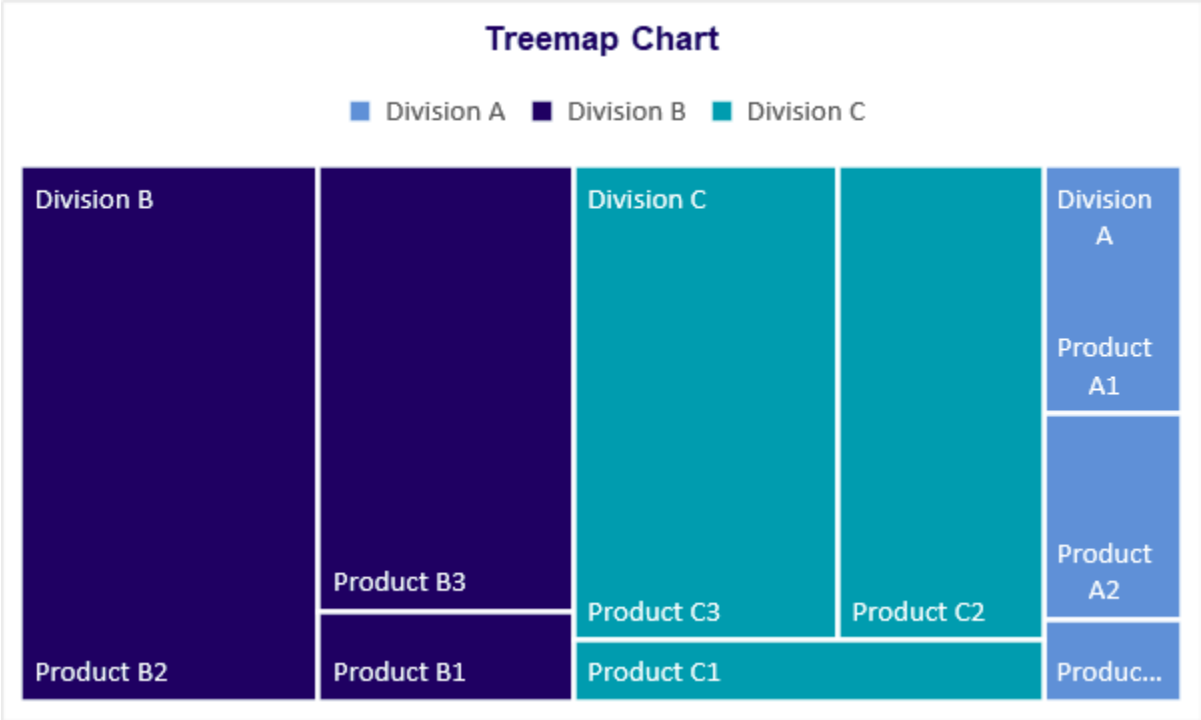


Hierarchy Data Visualization

- Treemaps are a data-visualization technique for large, hierarchical data sets. They capture two types of information in the data: (1) the value of individual data points; (2) the structure of the hierarchy.
- Definition: Treemaps are visualizations for hierarchical data. They are made of a series of nested rectangles of sizes proportional to the corresponding data value.
- A large rectangle represents a branch of a data tree, and it is subdivided into smaller rectangles that represent the size of each node within that branch.

Treemap Chart

- There are some chart types that are effective but often neglected. Treemap charts are a good example. Here is what one looks like.



Treemap – when to use?

- The company we have been looking at so far has three divisions. And each of them has its own products.
- This is the perfect way to provide information about the weight divisions have with respect to the firm's total revenue.
- At the same time, it shows how much each product contributes to the revenue of its division.

Reports

- Data reporting is the process of collecting and formatting raw data and translating it into a digestible format to assess the ongoing performance of your organization.
- Your data reports can answer basic questions about the state of your business.
- They can show you the status of certain information in an Excel file or a simple data visualization tool.
- Static data reports usually use the same format over a period of time and pull from one source of data.

Reports

Reports

- A data report is nothing more than a recorded list of facts and figures. Take the population census, for example.
- This is a technical document that transmits basic information on how many and what kind of people live in a certain country.
- It can be displayed in the text, or in a visual format, such as a graph or chart. But it is static information that can be used to assess current conditions.

Why data reporting is important?

- Data provides a path that measures progress in every area of our lives. It informs our professional decisions as well as our day-to-day matters.
- A data report will tell us where to spend the most time and resources, and what needs more organization or attention.
- Accurate data reporting plays an important role in every industry. The use of business intelligence in healthcare can help physicians save lives by providing more effective and efficient patient care.
- In education, data reports can be used to analyze how attendance records relate to seasonal weather patterns, or how acceptance rates intersect neighborhood areas.

Data reporting skills

- The most effective business analysts master certain skills. An excellent business analyst must be able to prioritize the most relevant information.
- They must be extremely thorough and detail-oriented; there's no room for error in data reports. Another useful skill is the ability to process and collate large amounts of information. And finally, being able to arrange the data and display it in an easy-to-read format is key for all data reporters.
- Excellence in data reporting doesn't mean you have to immerse yourself in code or be an expert at analytics. Other important skills include being able to extract essential information from the data, keeping it simple, and avoiding data hoarding.

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com