

Introduction to Data Visualisation

1.1 Introduction to data visualisation and communicating value

Introduction

Data is omnipresent and growing larger at an astronomical pace each day. This massive collection of information has changed every interface and interaction happening globally.

This week, you will learn basic visualisation principles, explore the methods to choose the right kind of presentation for your purposes, learn how to provide collaborating features and make use of data as a resource.

Activity: Perceiving the natural order in visualisation

Time: 15 minutes

Purpose: To reflect on the role of data visualisation in helping business travellers and travel managers.

Task: Watch the 2-minute 50-second video and answer the question below.

Feedback: To help further the conversation, respond to the posts of at least 2 of your peers. You can ask them questions for clarification on their approach or provide them with feedback. Your facilitator will moderate this discussion.

<https://youtu.be/0J0vduSYSR4> Source: (*Beyond business Travel*, 2021)

Question: How do you think data visualisation is benefitting business travellers or travel managers to manage better?

1.1.1 Telling stories with data

Data visualisation - Telling stories with data

Let us now delve deeper into the world of data visualisation.

Data visualisation is one of the stages in data science methodology that renders information into a visual context, such as information graphics, information visualisation and statistical graphics. It provides a quick and effective way to communicate information universally using visual information. This makes it easier to identify patterns, trends and outliers in large data sets and pull insights from them.



Source: The narrative structure of storytelling with data (Cloudinary, n.d.)

Telling stories with data too has both benefits and common concerns, regardless of the setting and audience type.

Benefits	Common Concerns
<ul style="list-style-type: none"> • The ability to absorb information quickly, improve insights and make quicker decisions. • Enhanced understanding of the possible steps to improve the organisation functions. • An improved ability to preserve the audience's interest with curated information they can assimilate. • An easy dispersal of information with enhanced prospects to share insights with everyone involved. • An increased ability to work on the key findings and implement changes with greater speed and fewer mistakes. 	<ul style="list-style-type: none"> • What is the dataset? Who created the dataset and why? • What is the dataset trying to say, and how can we let it say this? • What are we expecting the others to see from this dataset, and how can we convince them of this? How much work must we do to persuade them? • Who is affected by the processes and outcomes related to this dataset?

There are five core components to the workflow needed to tell stories with data. They are:

- Plan and sketch an endpoint.
- Simulate and consider some valid data.
- Acquire and prepare the real data.
- Explore and understand that dataset.
- Share what we did and what we found.

Watch this 4-minute 46-second video to explore how to tell stories with and within data:

https://youtu.be/r5_34YnCmMY Source: (*Harvard Business Review, 2019*)

1.1.2 Interpret a data visualisation

Activity: A day in the life of Americans

Time: 20 minutes

Purpose: To interpret a visualisation and assess the strengths and weaknesses.

Task: Visit the URL below and read the article. Once done, answer the given questions.

Instructions:

Step 1: Review the article below. It contains animated data visualisation demonstrating people's time spent on daily activities throughout the day. The plot is simple and easy to interpret but also includes a good number of variables, including time, activity type, number of people doing each activity, and the order in which activities are done.

Yau, N (2015). *A Day in the Life of Americans*. This is how America runs. Flowing Data. <https://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/>

Step 2: Based on your understanding of the article, answer the following questions.

Questions

1. When is the major shift in activities during the day, and to which?
2. What are the most common activities are the persons most involved in?
3. What is effective about this visualisation? Do you feel it lacked in providing the correct insights? Explain briefly.

1.2 Analytical reasoning and critical thinking

Introduction

Critical thinking is the ability to identify and extract key information from data to create a workable solution for the identified problem. It critically evaluates the identified problems, verifies the cause and resolves them by developing solutions.

On the other hand, analytical reasoning focuses on assessing your ability to consider a group of facts and rules and, given those facts and rules, determine what could or must be true. It's the ability to look at information, be it qualitative or quantitative, and discern patterns within the information.

Critical thinking

Critical thinking encompasses six vital skills: problem solving, analysis, creative thinking, interpretation, evaluation and reasoning.

Critical thinking and analytical reasoning process

Critical thinking and analytical reasoning process

What is the critical thinking and analytical reasoning process?

Data/Information visualisation is an outline that supports this analytical reasoning and critical thinking process. This framework has three views - a data view, a knowledge view and a navigation view.

- Navigation View: Gives an overview of the exploration process by collating the data view and knowledge view in the form of captured visualisation.
- Knowledge View: Helps the analyst record the analysis in findings and records.
- Data View: Offers collaborating information visualisation tools.

Analytical reasoning can break down data into patterns to discern information. For example, should a business be worried if a decrease in profit is part of a seasonal pattern of ups and downs or part of a greater downward trend? An individual gains insight into the information by learning to identify these patterns in both numbers and written arguments. In contrast, critical thinking is a logical skill, the ability to analyse and interpret information to assess its framework and validity. This can include multiple specific skills: asking questions about information, comparing material to general information and drawing connections between information and its context for analysis.

How does critical thinking differ from thinking?

	Thinking	Critical Thinking
Focus	The focus is on information: data, facts, examples on ideas: opinions, positions	The focus is on ideas: assumptions, biases, flaws in reasoning, point of view, context, implications
Activity	Organising and making connections between pieces of information or ideas, sometimes making basic inferences	Deeply and broadly questioning and testing the ways in which an idea is formed as well as how you have been interpreting and examining the idea. Thinking about your own while you are thinking about the thinking of others
Goal	To form an opinion about what you are thinking about	To apply criteria in forming a conclusion or evaluation about what you have been thinking about and how you have been thinking about it

Refer to the following readings to further your understanding.

- Dr. Wallis, C. (2020). *Introduction: What is Critical Thinking. Understanding, Learning, and Teaching Critical Thinking*. California State University Long Beach. <https://home.csulb.edu/~cwallis/170/text/What%20is%20Critical%20Thinking.pdf>
- Heard J., Scoular, C., Duckworth, D., Ramalingam, D., & Teo, I. (2020). *Critical thinking: Definition and structure*. Australian Council for Educational Research. https://research.acer.edu.au/ar_misc/38

Reflecting on the critical thinking process

Understanding critical thinking process

Time: 10 minutes

Purpose: To understand the critical thinking process.

Task: As critical thinking is a highly complex operation, the following examples are mere sketches of what is involved. Reflect and write your reflections in the box below.

1. Management

Question	Activity
What are the qualities of an effective Manager?	Analysis: breaking down the role of the Manager into tasks and inferring the qualities needed to complete each task effectively.

2. Political Science

Question	Activity
What are some of the barriers to instituting democracy in the nations of the developing world?	Synthesis: making connections by establishing the similarities and differences between a number of developing world countries in terms of problems preventing democratic systems.

3. Biology

Question	Activity
Is this study on the higher disease rate of farmed Chinook salmon credible?	Evaluation: judging the scope, controls and methodology of the study to determine if the scientific method was followed accurately.

1.3 Understanding your audience

Introduction

Any data visualization fails the purpose if not designed to communicate with a specific target audience. It should match with the target audience's expertise and help the viewers see and process data quickly and thoroughly.

Activity: Dashboards and audience

Time: 15 minutes

Purpose: To identify the target audience based on data visualisation.

Task: Review the data visualisation charts and select the appropriate target audience for each chart.



How to impress your audience with your data visuals

Win over your audience with your data visuals

To know your target audience, you must consider how familiar they are with the concepts being represented by the data and if they possess the capacity to interpret the charts and graphs represented easily. Here are some of the things you must consider:

Audience	Who is the audience, and what are the needs of the audience?
Goal	Is the goal to evoke a response, action or generate interest?
Setting	What is the environment in which the data is being shared- professional business or informal reflection?
Mode	How will the visualisations be presented? If you are publishing to Tableau Server, Tableau Online, or Tableau Public, do you need to create different dashboard versions?
Mood	Does the data tell a story that should invoke a certain response from your audience? You'll need to be sensitive towards your audience and the colours for the specific context
Consistency	Use the same fonts, colours, shapes, line thickness, and row-banding throughout all visualisations

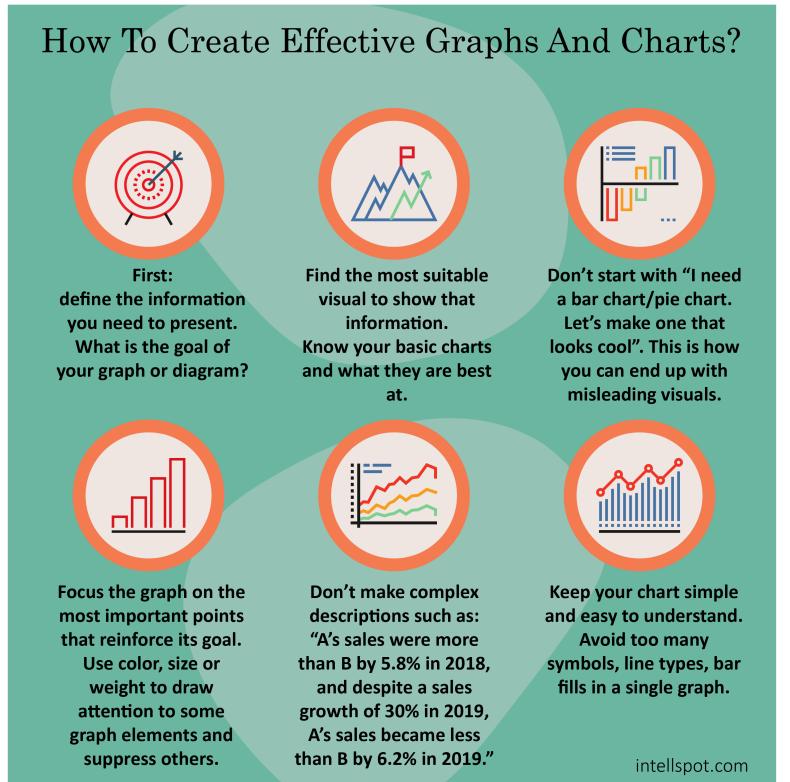
Refer to the following readings to further your understanding:

- Wolff, R. (2021). *Beautiful Data Visualization: 6 Examples of Great Datavis*. MonkeyLearn. <https://monkeylearn.com/blog/data-visualization-examples/>
- Mansurova, M. (2024, February 6). Data Visualisation 101: Playbook for Attention-Grabbing Visuals. Medium. <https://towardsdatascience.com/data-visualisation-101-playbook-for-attention-grabbing-visuals-c5b9faa7a950>
- Klassen, N. (2021). *Tailoring Visualizations to Suit Your Audience*. Air. <https://www.airweb.org/article/2021/08/19/tailoring-visualizations-to-suit-your-audience>

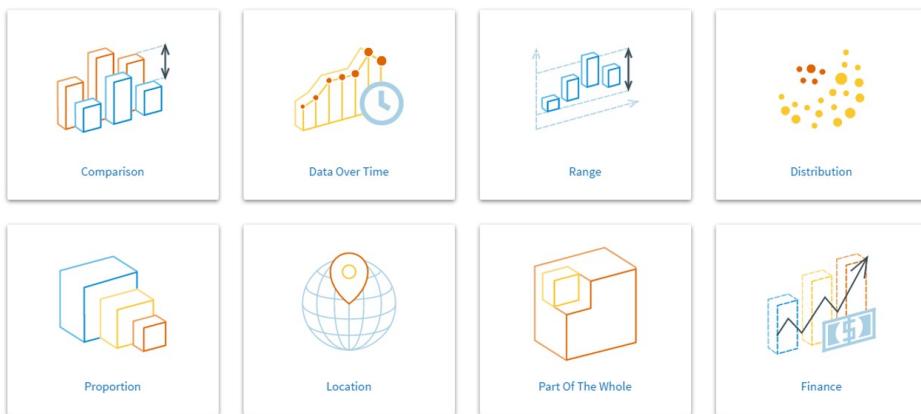
Choosing the right graphs

Determining the types of graphs to use and why may slow down any reporting and analysis. That's because picking the wrong visualisation could confuse the viewer or lead to mistaken data interpretation. You will explore this topic in more detail in the upcoming weeks. To create charts that illuminate and provide the right portrait for analysis, you must first understand the reasons why you might need a chart.

How To Create Effective Graphs And Charts?



It is critical when designing visualisations to consider the course by which the information gets transferred to the viewer. Choosing the correct chart for data analysis and reporting is a tough task. There are so many data visualisation options that aligning the appropriate option for every situation is complicated. Further, applying data to the wrong type of chart will likely relay wrong insights and, subsequently, wrong decisions.



1.3.1 Decoding visualisation as an audience

Time: 15 minutes

Purpose: To understand how to decode visualisation as an audience

Task: Review the visualisation below and answer the question.

Feedback: To help further the conversation, respond to the posts of at least 2 of your peers. You can ask them questions for clarification on their approach or provide them with feedback. Your facilitator will moderate this discussion.

As we have already learnt, you need to define your audience first: who are you creating your data visualisation for? What kind of people or professionals are they, what do they do, and what are they looking for? For example, your CEO's questions and what they need to know are drastically different to what your teams need to know, which is different to what a stakeholder wants to know.

Question

- Suppose your audience is composed of investors or the CFOs of your company evaluating the performance of online marketing efforts. What could be some of the metrics they could be interested in? Look at the following visualisation and answer below.



1.4 Design, simplify and focus

Introduction

Designing visualisation is an iterative process in which understanding the user's perspective, define the problems, identify workable solutions and developing strategies with a predefined approach.

It is a way of thinking and working as well as a collection of hands-on methods.

Activity: Data visualisation and misrepresentation

Visit the following URL to read about 15 misleading data visualisation examples and reflect on how data visualisation can be used to bend the truth and misrepresent trends. Support your reflection with a few latest examples.

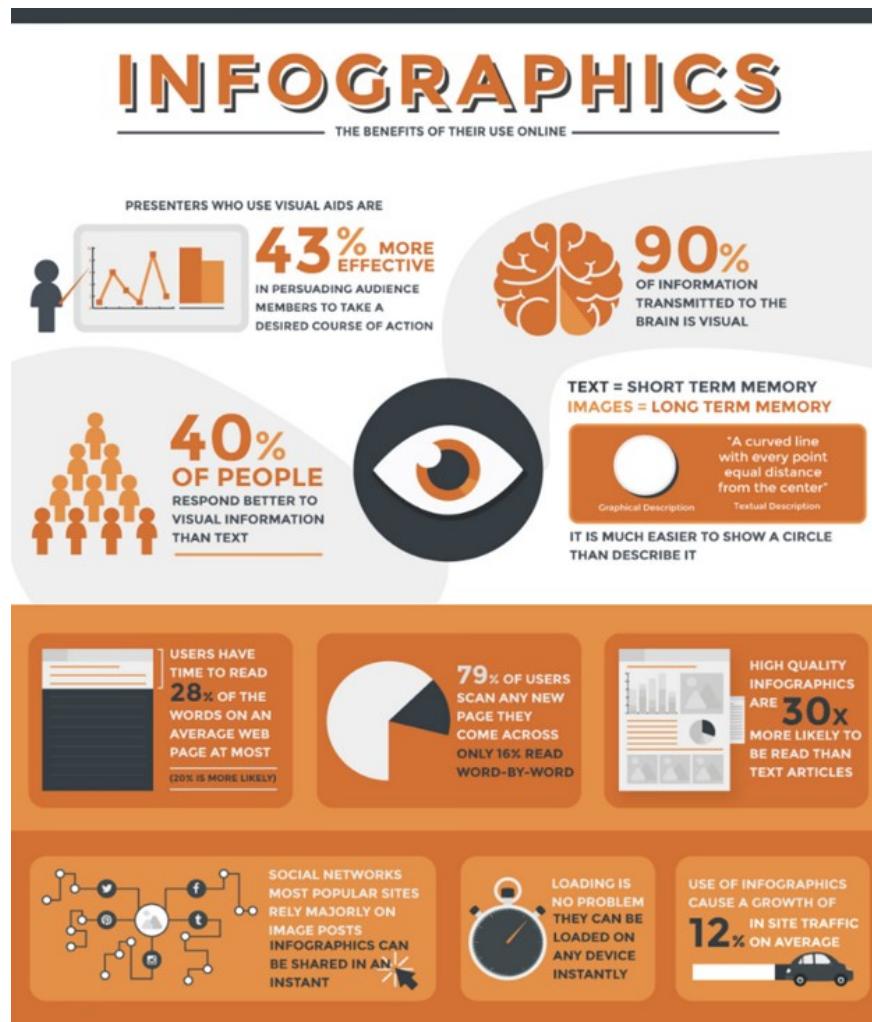
Clayton, T. (2021). *15 Misleading Data Visualization Examples*. Rigorous Themes. <https://rigorousthemes.com/blog/misleading-data-visualization-examples/>

Design thinking in data visualisation

Utilise concepts of design in data visualisation

The purpose of designing any visualisation is that the user understands the data better and the visualisations are intuitive and well-organised.

Look at the sample infographic below:



When designing, make sure you are shooting for clarity, and your data is set out in a logical format. You must therefore take cognisance of the following:

- Placed alphabetically, by value or another criterion, depending on the data.
- Take into consideration the hierarchy of data, place different elements in certain places to attract more attention, and ensure that you have white space in your design.
- The data-ink ratio is well balanced

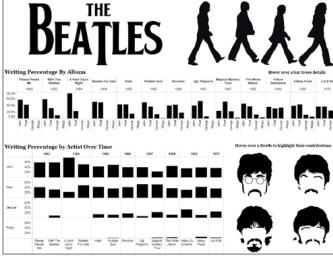
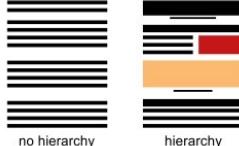
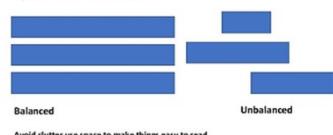
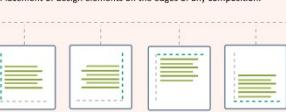
Refer to the following readings to further your understanding.

- Vinzon, E. (2021). 14 Peaks. Tableau Public. https://public.tableau.com/app/profile/ervin.vinzon/viz/Eight-Thousander_16386540239820/dash

- How to think visually using visual analogies. (n.d.). 1000+ Infographics, Posters, Flyers & More | Venngage Gallery. <https://venngage.com/gallery/post/how-to-think-visually-using-visual-analogies/>

Concepts of design

You will now explore a few elements of design to keep in mind while designing a visualisation.

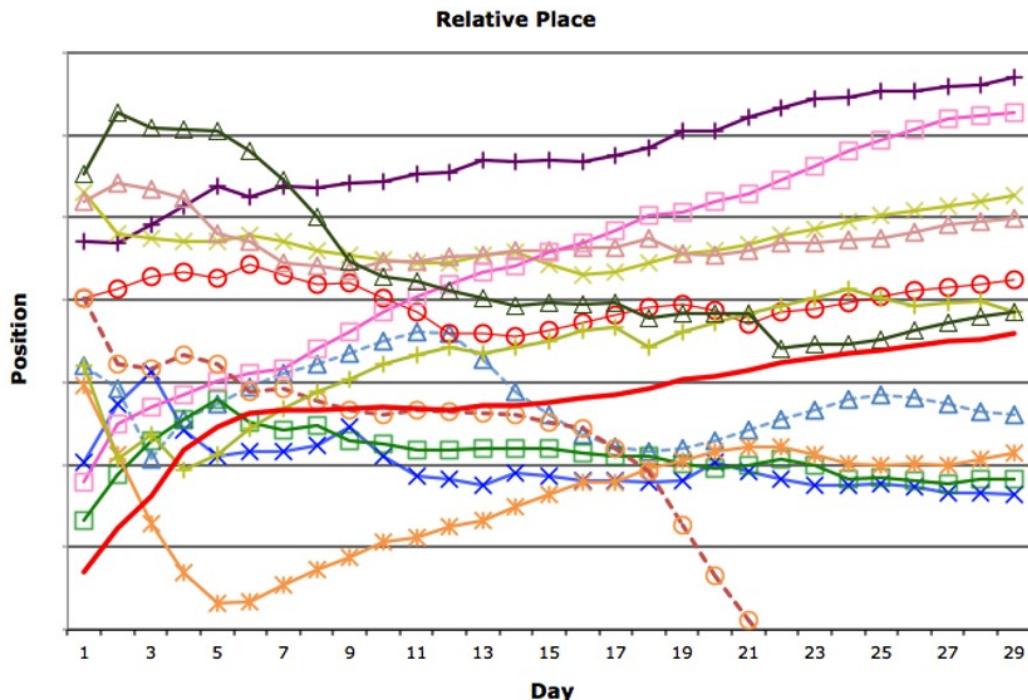
Unity	<p>Unity is the application of methods that ensure that elements in the design appear to 'go together' - (colour, font, & shape consistency). Click on the teal circles to reveal the aspects to ensure unity in a visual design.</p>	 <p>a. Consistent colour: Black b. Consistent font c. Simplify images</p>
Hierarchy	<p>Hierarchy is the application of design methods to indicate the importance and 'flow' within the visual (size, placement).</p>	
Colour	<p>The use of colour provides contrast for data points in opposition and brings attention to relevant elements within the visual.</p>	
Balance & Alignment	<p>Balance and Alignment create harmonious visuals that do not distract from the message being communicated.</p>	<p>Space and Balance</p>  <p>Alignment Design Principle</p> <p>Multiple objects are said to be aligned when they are placed such that their left or right edges, or center-lines line up on a common position.</p> <p>1 Edge Alignment</p> <p>Placement of design elements on the edges of any composition.</p>  <p>2 Center Alignment</p> <p>Placement of design elements along the imaginary central axis of any composition.</p> 

Watch this 3-minute 44-second video to explore Tufte's design principles.<https://youtu.be/mvBJ7i8sEpk> Source: ([ELFT QI](#), 2021)

Recognise and eliminate the clutter clouding your information

As you start designing the visualisation, continually refer to your original objective. You must question whether the elements you add to your visualisation bring you closer to that objective. Eliminate everything and anything that doesn't add value or detracts from the viewers' attention.

The following chart is a good example of how more detail can lead to confusion rather than clarity. There is too much for the eye to focus on and too many threads to follow:



Source: Cluttered chart (Srichinmoyraces, 2012)

The target audience needs to understand the sophisticated content thoroughly. The visualisation of any data should grab an audience's attention and navigate it to the complex and important concepts of the visualisation.

Here are a few important things to keep in mind:

- The process should build context and prioritise data exploration and comparison.
- Select visual and interactive affordances that support the discoverability of core features.
- Motion and interaction should support analytical reasoning and user comprehension by revealing context, insights, associations, and causality.
- Every action, colour, and visual element should support data insights and understanding.

Refer to the following readings to further your understanding.

- Rachel. (2020). *Presenting data visualization to engage your audience*. UX Collective. <https://uxdesign.cc/presenting-data-visualization-to-engage-your-audience-815eb6a43a62>

- Few, S. (n.d.). 35. *Data Visualization for Human Perception*. Interaction Design Foundation. <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception>
- Ramseger, F., Walker, J. *Design tips for data visualizations that capture audience attention*. Tableau. https://www.tableau.com/sites/default/files/webinars/slide_deck_0.pdf

1.4.1 Utilising tableau desktop

How to create a word cloud with tableau desktop for students of art and cinema

Time: 15 minutes

Purpose: To learn how to create a Word Cloud using Tableau for a student audience from the Arts and Cinema discipline.

Task: Utilise this [Movie Review Dataset](#) available at Kaggle and envision word cloud on the content information.

Instruction

1. Open Tableau Desktop and connect to a data source by downloading the data file from [Movie Review Dataset](#) available at Kaggle.
2. Drag the desired dimension to **Text** on the **Marks** card.
3. Drag the same dimension to **Size** on the **Marks** card.
4. Right-click on the dimension on the **Size** card and select **Measure > Count**.
5. If necessary, change the Mark type from **Automatic** to **Text**.
6. To add colour, drag the same dimension to **Colour** on the **Marks** card.
7. Share your work in the discussion forum.

You may refer to Tableau's knowledge base for more help.

1.5 Exploratory analysis

Introduction

Exploratory data analysis (EDA) is the task of analysing data using simple plotting tools from statistics. This approach analyses data sets to summarise their main features. When you are trying to build a visual learning model, you need to ensure that your data makes sense.

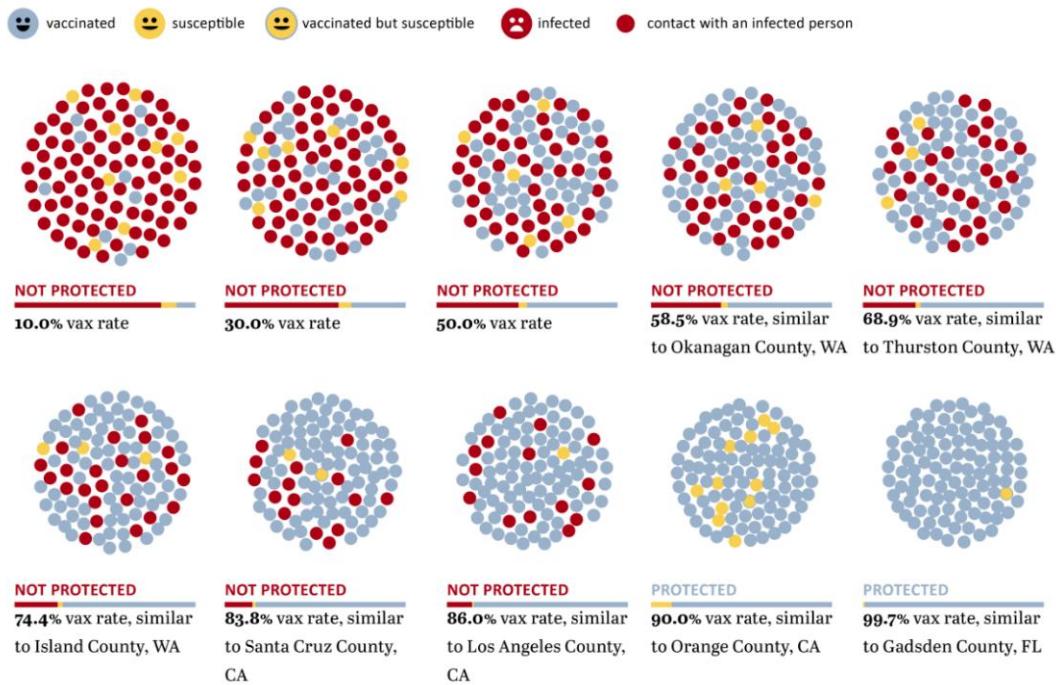
Activity: Simulation of vaccination effectiveness

Time: 15 minutes

Purpose: To analyse vaccination data using a visualisation

Task: Review the visualisation below and decipher if vaccinating against measles has been effective or not if delivered fully in 2 doses. Discuss your analysis supported by statistical information.

Feedback: To help further the conversation, respond to the posts of at least 2 of your peers. You can ask them questions for clarification on their approach or provide them with feedback. Your facilitator will moderate this discussion.



1.5.1 Tools and methods for performing EDA

Tools and methods

There are a lot of tools where one can perform EDA like:

- programming languages used are mainly R and Python

- tools like Tableau, IBM Cognos etc are used which are often known as Business intelligence tools (BI).

You may undertake Exploratory Data Visualization to help you understand the data and look for any patterns. This typically involved plotting different graphs and charts to visualize the data and to help you see any trends or anomalies.

Some methods and plots are distinguished as: -

- Univariate analysis
- Bivariate analysis
- Multivariate analysis

Some of the common graphs used while performing EDA:

- Scatter plot
- Pair plots
- Histogram
- Box plots
- Contour plots
- Violin plots

Exploratory analytics tries to answer the questions of how and why something happened. In this process, once the data is cleaned, it is important to decipher the data by looking at the data types of the variables and if they define the problem well. The second step is to look at the distribution of variables of interest and assess how many categorical and numerical variables are there. Using statistical methods like mean, median, quartiles to calculate also help visualise the data well. Work on the use-case to calculate if the variables are correlated. Then you can choose your favourite visualisation tool to present the data better.

Refer to the following readings to further your understanding.

- IBM Cloud Education. (2020). *Exploratory Data Analysis*.
IBM. <https://www.ibm.com/cloud/learn/exploratory-data-analysis#toc-explorator-kFQ--qH6>
- Restori, M. (n.d.). *What is Exploratory Data Analysis*.
Chartio. <https://chartio.com/learn/data-analytics/what-is-exploratory-data-analysis/>
- SolveXia. (2019). *19 Big data problems you need to solve*.
SolveXia. <https://www.solvexia.com/blog/15-big-data-problems-you-need-to-solve>

How to find data sources

The first step in visualizing data is to load it into your application.

Now two questions arise:

- Do you own the data in a recorded definable, digitizable format?
- How does one quantitatively acquire data?

Thus, the procurement stage covers several tasks that are sometimes complex:

- Unless you are generating your own data, you must find a good source for the data you want.

- If you don't own the data, you must make sure you have the right to use it.
- You must download the data, which may present difficulties if the volume is large, especially if it's dynamic. You may have to go through loops to extract the data from a web page or another source.

Tableau can connect with various data sources such as text, excel file, and databases to big data queries. This section will look at the basics and advance feature of data connectivity with different sources. Here we will also look at Join types, Data Blending, connection with cubes, custom SQL and Google Analytics.

Refer to the following readings to further your understanding.

- everviz. (2021). *How to find data sources for data visualization*.
everviz. <https://www.everviz.com/how-to-find-data-sources-for-data-visualization/>
- Hughes, E. (2020). *Refreshing LIVE vs EXTRACT Data Sources in Tableau*. The Data School. <https://www.thedataschool.co.uk/erica-hughes/refreshing-live-vs-extract-data-sources-in-tableau>

How to select variables

Variable selection means choosing among many variables to include in a model to determine a set of variables that will provide the best fit for the model to facilitate accurate predictions. Variable selection is one of the most difficult aspects of model building, and one must carefully select appropriate variables to avoid including noise variables in the final model.

Determining such an appropriate subset of explanatory variables to be used in any data analysis is called the problem of variable selection. While choosing a subset of explanatory variables, there are two possible options:

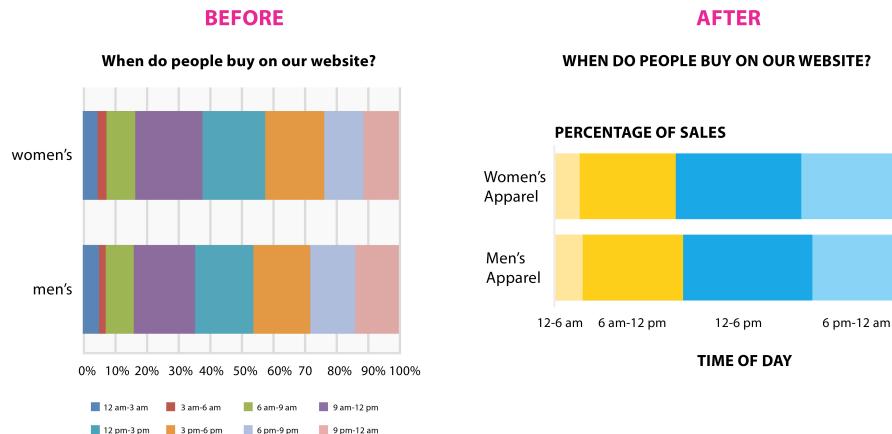
- The analyst may include as many explanatory variables as possible to make the model as accurate as possible.
- To make the model as simple as possible, the analyst may include only a lesser number of explanatory variables.

Watch this 7-minute 7-second video to learn more about model building and variable selection.
<https://youtu.be/Bq-estPj7fo> Source: (*MarinStatsLectures-R Programming & Statistics*, 2021)

Case study examples on common data visualisation problem

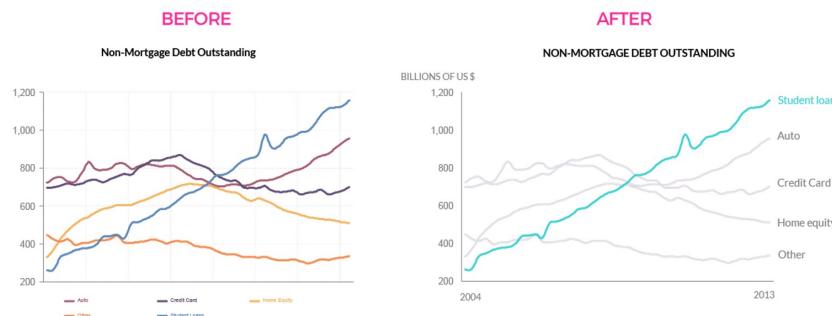
You will now explore a few examples of common data visualisation problems.

- Case 1: **Highlighting your key point**
Ask yourself, "Can this be summarised?" If certain details are not essential, consider summarising certain elements by grouping them into broader categories.



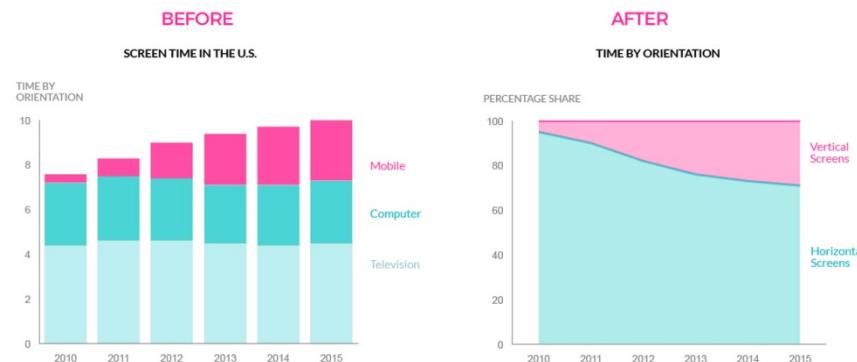
- Case 2: Using the right colour and sizing to emphasise key insights

Use different design techniques to make the important chart elements stand out, such as making the size bigger, making certain lines thicker, applying colour, adding data explainers or labels, among others.



- Case 3: Using minimum colour

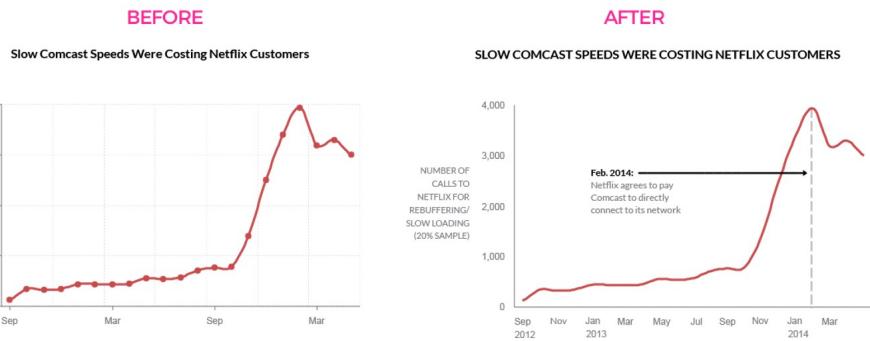
In this example, we see how using only two colours (and summarising information into two groups) makes it easier to look at the key points of the chart. Use only two to three colours to highlight the most important points. Use grey for the rest of the information.



- Case 4: Declutter your charts

Chart gridlines and borders should be minimally used. In this example, the chart gridlines are not necessary to understand the key point of this chart - that slow Comcast speeds caused a spike in customer calls in

2014.



1.5.2 How to pick a chart

Activity: Match the following

Time: 15 minutes

Purpose: To judge how to pick the right chart.

Task: Drag and drop the correct chart type against each visualisation requirement.

Chart	Function
Line	These charts are perfect for comparing one or many value sets, and they can easily show the low and high values in the data sets.
Dual Axis Line	These charts help us know more information about how a dataset performed during a specific time period.
Pie	This type of chart shows how individual parts make up the whole of something.
Bubble	Relationship charts are suited to showing how one variable relates to one or numerous different variables.
Scatter	Distribution charts help you to understand outliers, the normal tendency, and the range of information in your values.

Activity: Kaggle - Predict categories from the San Francisco crime database

Time: 25 minutes

Purpose: To dive into data and generate inferences using different visualisation methods

Task: Explore a dataset using Tableau and answer the given questions.

Feedback: To help further the conversation, respond to the posts of at least 2 of your peers. You can ask them questions for clarification on their approach or provide them with feedback. Your facilitator will moderate this discussion.

[Kaggle](#) has a data science competition to predict the category of crime in San Francisco based on 12 years (From 1934 to 1963) of crime reports across all of San Francisco's neighbourhoods (time, location and other features are given). Explore the [dataset](#) visually using Tableau (to download the data set, you need to register on Kaggle) and find hidden trends like:

Questions:

1. Are there specific clusters with higher crime rates?
2. Is there a yearly/monthly/daily/hourly trend?
3. Is crime distribution even across all geographical areas or different?