NPA Data Science

7: Data Visualisation and Storytelling

Topic Outcomes

- Explain data visualisations and data storytelling
- Explain techniques for data visualisation, data dashboards and data storytelling

Data Visualisation and Storytelling

The process of presenting data in visual form is known as data visualisation. Creating visualisations really helps make things clearer and easier to understand, especially with larger, high dimensional datasets. Visualisations are increasingly seen as powerful tools to engage users with unfamiliar and complex subject matter (Grainger et al. 2016).

Have you heard the phrase, “a picture is worth a thousand words”? This is why data visualisation is so important. A good visualisation tells a story, removing the noise from data and highlighting the useful information. Our eyes are drawn to colours and patterns. We can quickly identify red from blue, square from circle. Our culture is visual, including everything from art and advertisements to TV and films.

Data visualisation is another form of visual art that grabs our interest and keeps our eyes on the message. When we see a chart, we quickly see trends and outliers. If we can see something, we internalise it quickly. It’s storytelling with a purpose. If you’ve ever stared at a massive spreadsheet of data and couldn’t see a trend, you know how much more effective a visualisation can be.

Storytelling with data differs from data visualisation because it requires communicators to offer a larger, holistic, view of their message. Throughout history, storytelling has been an effective way of conveying information and knowledge (Tong et al. 2018). You must focus first on your audience and structure a larger message before any visuals are rendered. You must identify from the start:

- What do I want my audience to know or do with the data I am presenting?
- How will I structure a narrative that leads to desired action?
- How is my data helping drive a decision?

A good data storyteller should know their audience and adjust accordingly. Understanding what each audience segment wants in terms of data complexity and narrative delivery can help you pitch your stories more effectively and efficiently (Feigenbaum and Alamalhodaei,

2020). There is no understating how important it is for all presented data to have a purpose. Every piece of data you include should further this purpose – or it should be left out. The key things to think about for storytelling include understanding the context, choosing an effective visual, eliminating clutter, drawing attention to where you want it, thinking like a designer, and telling the story in a simple way.

Some good examples of storytelling can be found here: https://www.tableau.com/en-gb/learn/articles/best-beautiful-data-visualization-examples

Hans Rosling, famously tells the story of global population growth through data visualisation, which can be found on YouTube, alongside his videos on child mortality rates, the rise of Asia, insights on HIV and poverty. Watching these examples, you can see how data is brought to life to tell stories.

Common dimensions of storytelling with data visualisation according to Tong et al. (2018) are:

Authoring tools - who is creating the story and narrative

User-engagement - related to the audience and why we use storytelling

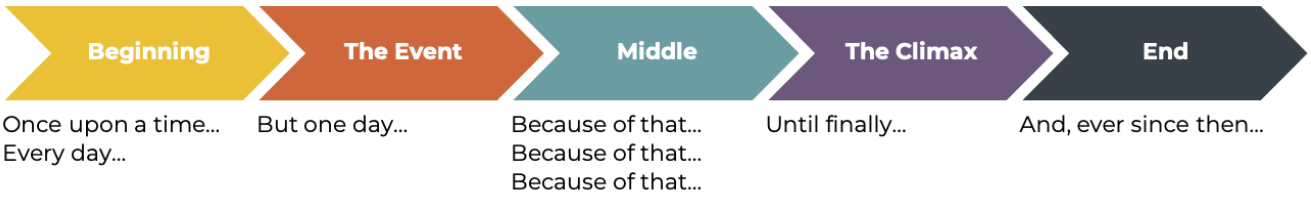
Narratives - how an author tells a story - including the events and characters

Transitions: how the authors tell a story - seamlessly blending events within a story to make it flow and strengthen its coherence

Memorability: why authors present data as a story - a good visualisation technique draws the viewer's attention and increases memorability

Interpretation: data interpretation is the process of critiquing and determining the significance of the data

All stories have a structure and common way to think about the story is shown in the figure below.



For data stories, the beginning should include a problem and context. The event is the analysis undertaken, and the assumptions and data used. The middle will look at relationships between variables and the interpretations of those relationships, and the climax is the implications of the results. The end includes the recommendations and next steps.

It is useful to be aware that the skill and practice of data visualisation and storytelling requires more than data science skills and is useful in many areas of work. There is a

wealth of information out there on successful approaches from across many areas of society and from other disciplines. The Data Storytelling Workbook introduces a variety of practices and techniques that contemporary data storytellers use (Feigenbaum and Alamalhodaei, 2020). It draws from a range of disciplines not just data science to explore different concepts for data visualisation and storytelling. For example, making an analogy with photography, they highlight that the more you zoom in on a data point, the more complexity you will see. Whilst the farther you zoom out, the more context you will see. Hence, when presenting data, the goal is to zoom in enough to show complexity whilst zooming out enough to give the audience the context needed to understand the meaning and significance. This is especially important if the point of your data visualisation and story is to get the audience to feel a certain way or to take a particular action. In this sense, you are trying to reach your audience emotionally, to provoke them to think, feel and do.

Data visualisation tools and examples

Data visualisation tools make the job easier. Visualisation methods are utilised to create tables and diagrams to understand data. Doing this with big data is more difficult than with traditional small data because of the complexity of the four V's. Technically, that simple pie chart you can generate with one-click using Microsoft Excel is a data visualisation. But, as technology has suddenly begun evolving in leaps and bounds over the traditional databases and spreadsheets to which we're accustomed, new kinds of data visualisations have become possible using a host of new tools and tech. Matplotlib is a popular Python library that can be used to create your Data Visualisations quite easily. Other free tools include Tableau Public, Tableau Gallery, Microsoft Power BI, Datawrapper, Google Data Studio, Openheatmap, Leaflet, Chartbuilder, OpenRefine, Information is Beautiful. There are many more tools available for all kinds of data visualisations.

Common general types of data visualisation are Charts, Tables, Graphs, Maps, Infographics and Dashboards. Not all chart types are suitable for all types of data, or the questions being addressed. There are six main types that can address different questions.

Comparison: comparing and sorting data points (e.g bar chart)

Composition: part-to-whole comparisons (e.g. pie chart)

Distribution: Comparison of data points along an axis (e.g histogram)

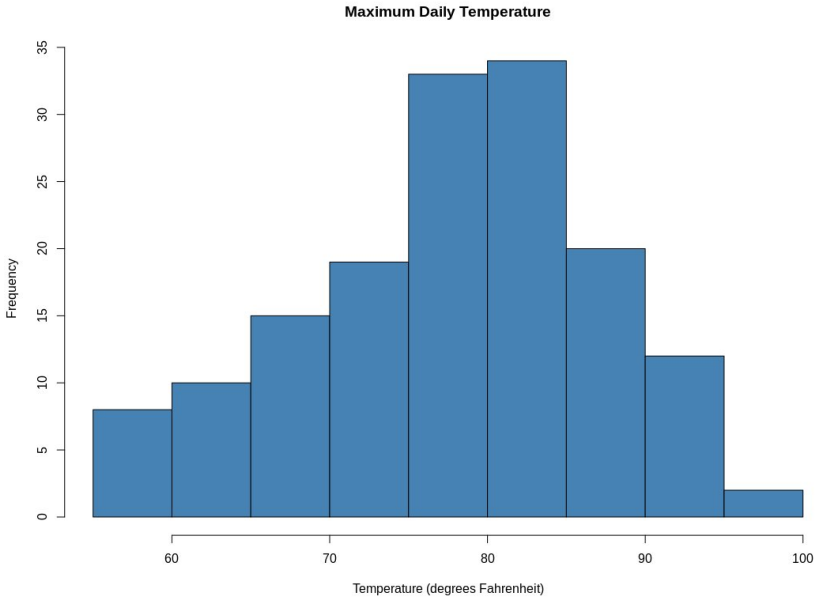
Relationship: Patterns between two or more variables (e.g. line graph)

Evolution: patterns over time (e.g. scatterplot)

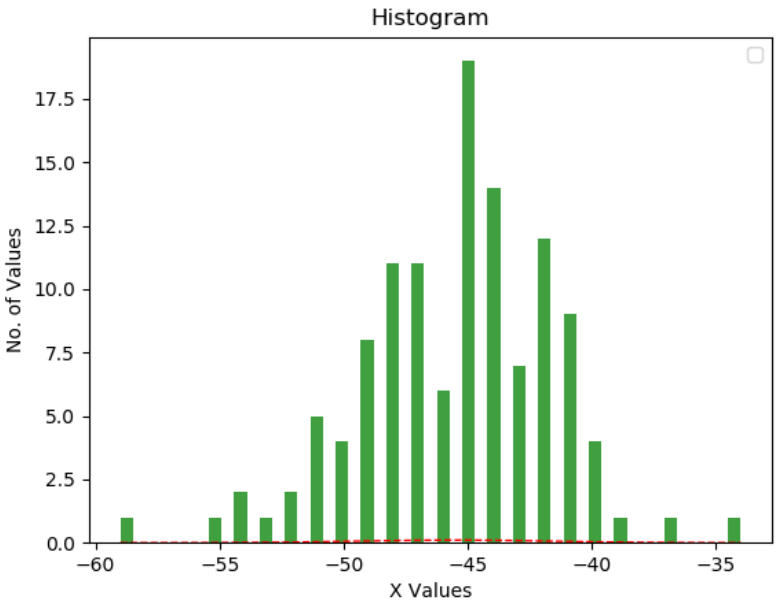
Profiling: pattern comparison (e.g. grouped bar chart)

Here are some examples of different data visualisations that can be created.

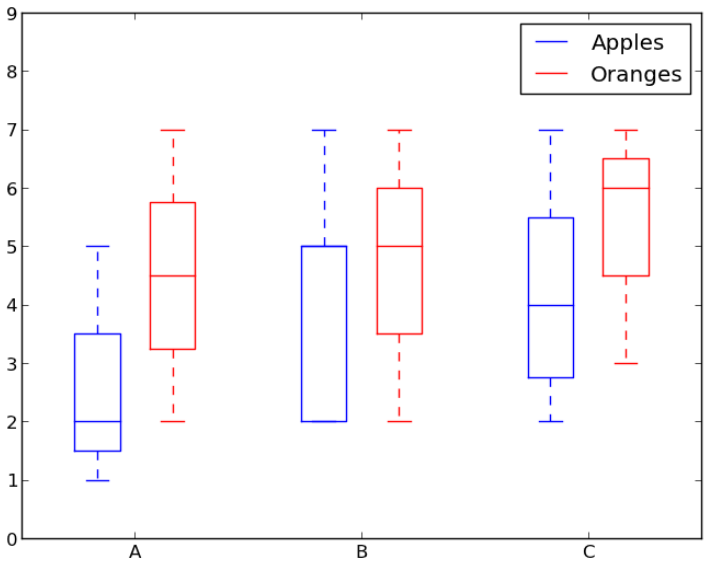
Bar plots are most effective when you are trying to visualise categorical data that has few (probably < 10) categories. If we have too many categories then the bars will be very cluttered in the figure and hard to understand. They're nice for categorical data because you can easily see the difference between the categories based on the size of the bar (i.e magnitude); categories are also easily divided and colour coded too.



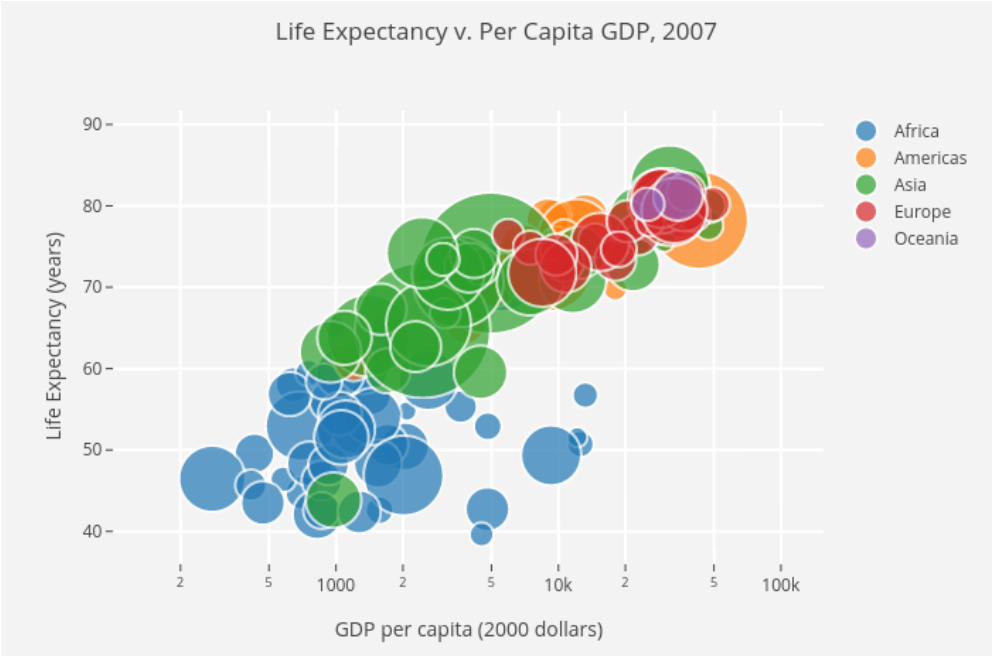
A histogram is a very common plot. It plots the frequencies that data appears within certain ranges.



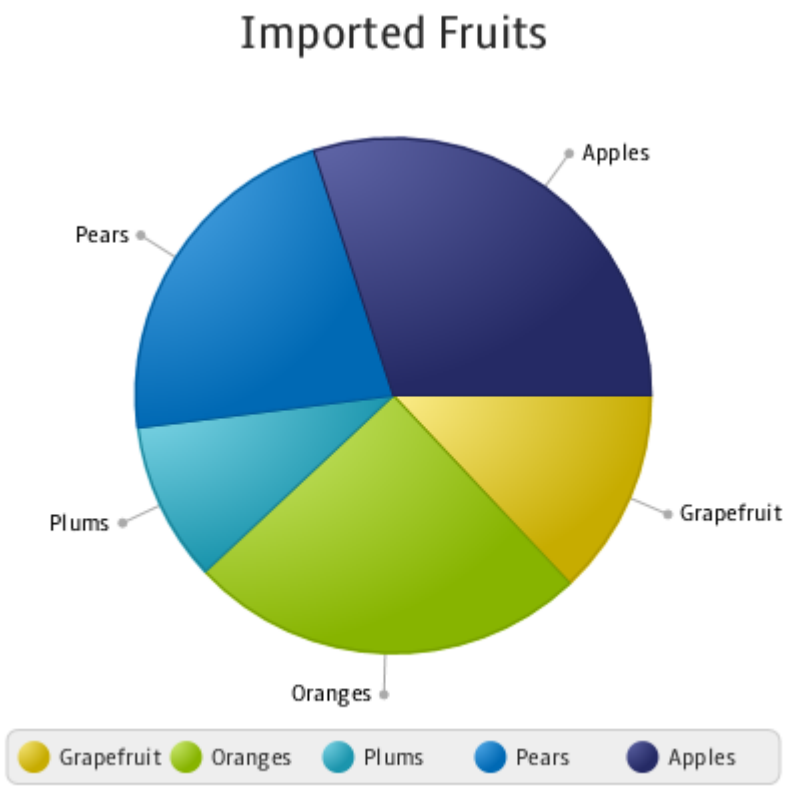
A box plot provides a graphical view of the median, quartiles, maximum, and minimum of a data set. But how would we improve this? By adding a title and axis label.



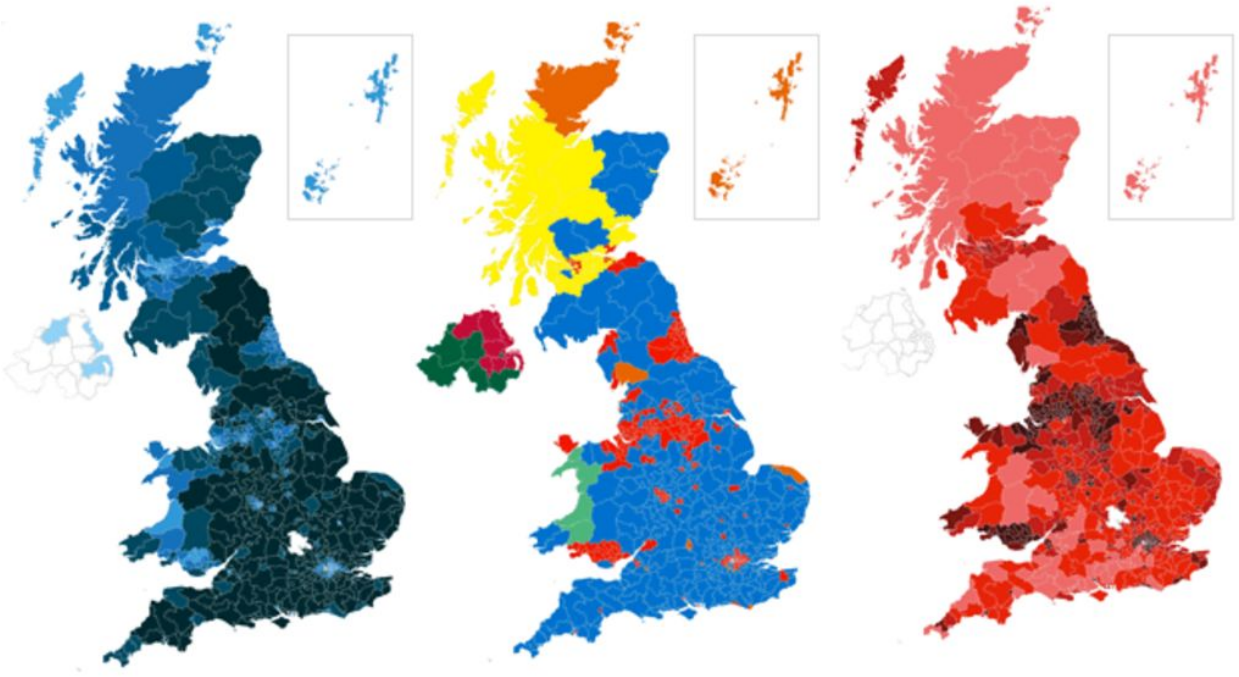
A scatter plot is very useful for showing the relationship between two variables.



A pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion.



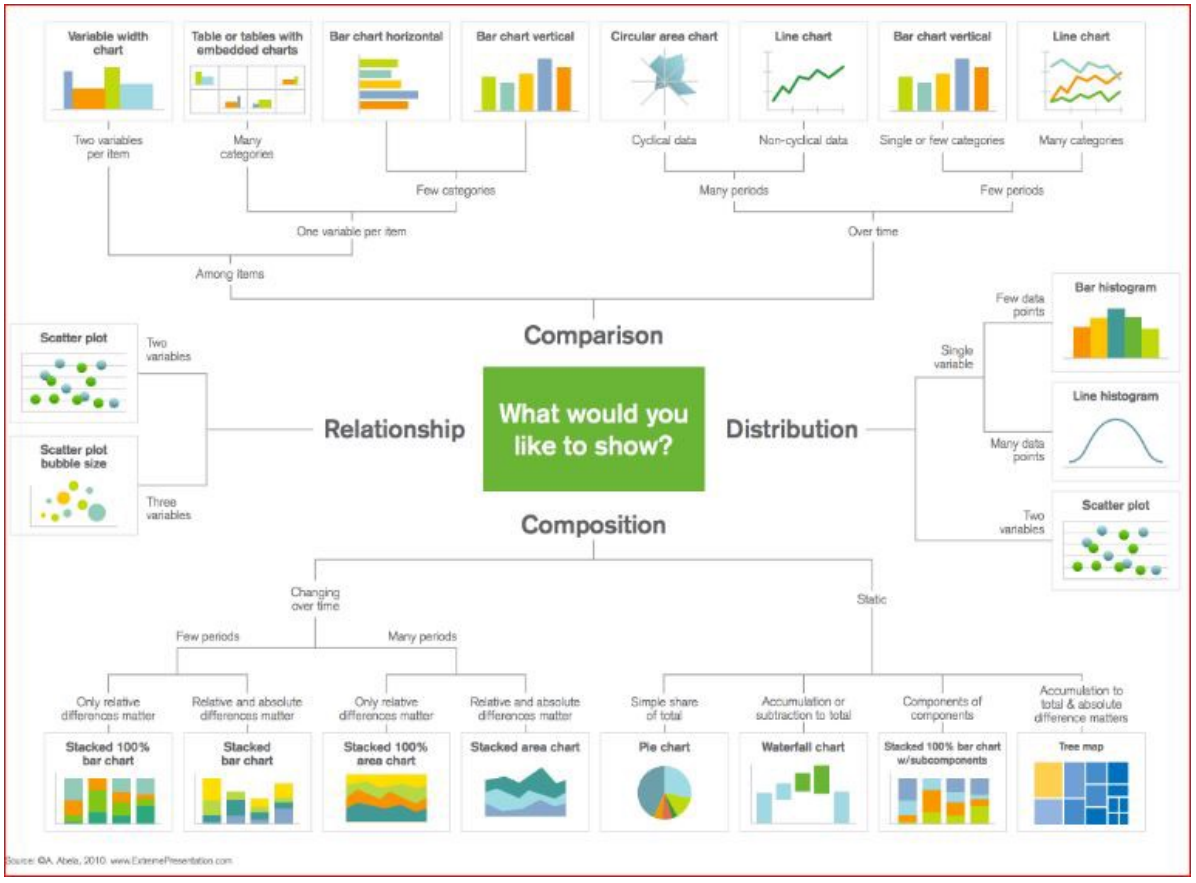
A heatmap can show the geographical spread of data such as polling results.



Word clouds (also known as text clouds or tag clouds) work in a simple way: the more a specific word appears in a source of textual data (such as a speech, blog post, or database), the bigger and bolder it appears in the word cloud. A word cloud is a collection,

or cluster, of words depicted in different sizes. The bigger and bolder the word appears, the more often it's mentioned within a given text and the more important it is.

But how would you know what kind of visualisation to use? The figure below is a handy snapshot of how you might go about making the decision. So if you want to compare variables of data, you would use a bar chart or a line chart, and if you want to assess a relationship, you use a scatter plot. Whereas if you want to measure distribution, you can use a histogram or a scatter plot and for composition of data, options include a pie chart, tree map, stacked bar chart or stacked area chart for example.



Some further information on the relationship between what data you want to visualise and the best way to visualise it, with recommendations on appropriate tools can be found on the Data Visualisation Catalogue: https://datavizcatalogue.com/.

It is also important to be aware of the best practices regarding the use of colour in data visualisations. The use of colour is usually related to the purpose such as:

- **sequential:** a single colour scale from light to dark e.g. a heat map
- **diverging:** two colours with a neutral midpoint
- **categorical:** different colour per category
- **highlight:** a highlight colour against a neutral colour to bring attention to a specific point
- **altering:** use of a colour code to show whether on target or not

Usually one would aim to use no more than three colours on a single graph as a way to allow the brain to spot patterns quicker. Additionally, the issue of colour-blindness is important. Normally this relates to red-green colour blindness but there is also a blue-yellow variant. Finally, there are design considerations to bear in mind, including the use of titles and subtitles, labels and text, legends, captions, axis arrangement and scale, units, range and sorting order. As a minimum all graphs should have a title and be labelled sufficiently for the audience to understand without needing further information. Now that you are aware of this, look again at the graphs above, which ones would meet the standards expected and which are missing information? Are they suitable for colour-blindness?

Dashboards

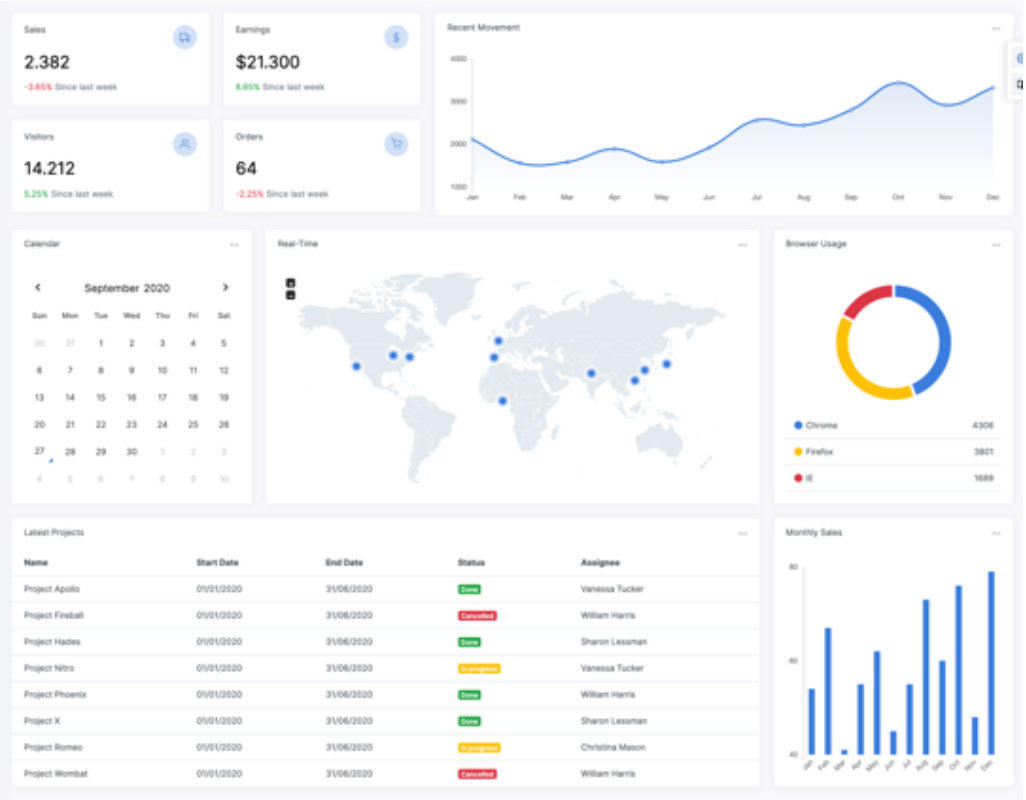
A dashboard is a visual display of the most important information needed to achieve one or more objectives, which has been consolidated on a single computer screen so that it can be monitored at a glance - see the figure below for an example. A data dashboard is named after a car dashboard and should give the user key information within a single eye span.

Dashboards can be strategic, operational or analytical.

A strategic dashboard might be used by a CEO or an organisation, and contain key metrics monitored by the business and show overall progress.

An operational dashboard is generally focused on a specific process within a business. This may be real time information. For example a call centre might display call waiting times.

An analytical dashboard is usually interactive with a filtering capability to dig into the data without the need for an analyst. They enable domain experts to understand why something is happening.



Dashboard software is a data visualisation tool that gathers and displays business data in interactive and customisable visualisations that enable users to monitor a business's health, analyze processes, track KPIs, and discover actionable insights.

Some examples of open sources dashboard software and tools include:

Freeboard http://freeboard.io/

Mozaik http://mozaik.rocks/

Dashbuilder http://dashbuilder.org/

Grafana http://grafana.org/

Stashbord http://www.stashboard.org/

Google Data Studio: https://analytics.google.com/analytics/academy/course/10 (including a course explaining how to use it)

Metabase: https://www.metabase.com/

Here you can see a live demo of the Grafana Dashboard. Consider what kinds of visualisations are being used.

References

Feigenbaum, A. and Alamalhodaei, A. 2020. *The Data Storytelling Workbook.* London: Routledge.

Grainger, S., Mao, F. and Buytaert, W., 2016. Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling & Software*, 85, pp.299-318

Lee, A. 2018. Storytelling with Data: A Data Visualization Guide for Business Professionals. Towards Data Science. Available: https://towardsdatascience.com/storytelling-with-data-a-data-visualization-guide-for-business-professionals-97d50512b407

Martin, N. 2018. Data Visualization: How To Tell A Story With Data. Forbes. Available: https://www.forbes.com/sites/nicolemartin1/2018/11/01/data-visualization-how-to-tell-a-story-with-data

Ostrowski, J. 2019. Data Storytelling: Best Practices and Visualization Tips To Create Remarkable Reports. Ladder. Available: https://blog.ladder.io/data-storytelling/

Rist, O. 2018. 10 Free Data Visualization Tools. PC. Available: https://uk.pcmag.com/business/84085/10-free-data-visualization-tools

Seif, G. 2018. 5 Quick and Easy Data Visualisations in Python With Code. Towards Data Science. Available: https://towardsdatascience.com/5-quick-and-easy-data-visualizations-in-python-with-code-a2284bae952f

Tong, C., Roberts, R., Borgo, R., Walton, S., Laramee, R.S., Wegba, K., Lu, A., Wang, Y., Qu, H., Luo, Q. and Ma, X., 2018. Storytelling and visualization: An extended survey. *Information*, 9(3), p.65.