1495 UNIVERSITY OF ABERDEEN | CELEBRATING 525 YEARS 1495 – 2020

ABERDEEN 2040

Data Visualisation

Data Mining & Visualisation

Lecture 6

2025

Today…

- Intro to data visualisation
- Common types of visualisations
- Making effective visualisations

Descriptive Statistics vs Visualisations

So far, we have discussed several methods for extracting descriptive statistics from a dataset.

These are incredibly useful for starting to summarise, analyse, and understand important aspects of our data.

Descriptive Statistics vs Visualisations

However, what we gain from these approaches can often be a partial understanding, obscuring other important aspects and trends within the data.

Visualising data can be a fast, intuitive, and effective way for us (and others) to gain a broader perspective.

Descriptive Statistics vs Visualisations

Data visualisations do not replace or supersede descriptive statistics, or vice versa.

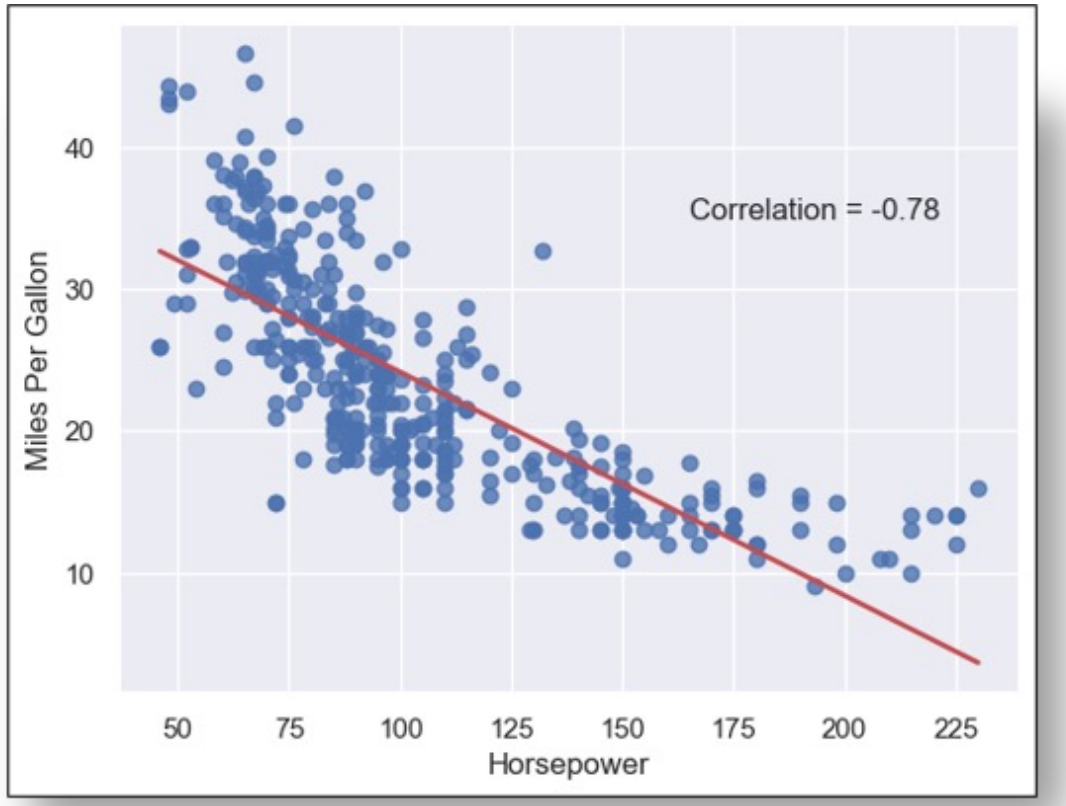
In reality, both methods are frequently used in conjunction, and are important aspects of a data scientist's toolkit.

Intro to Data Visualisation

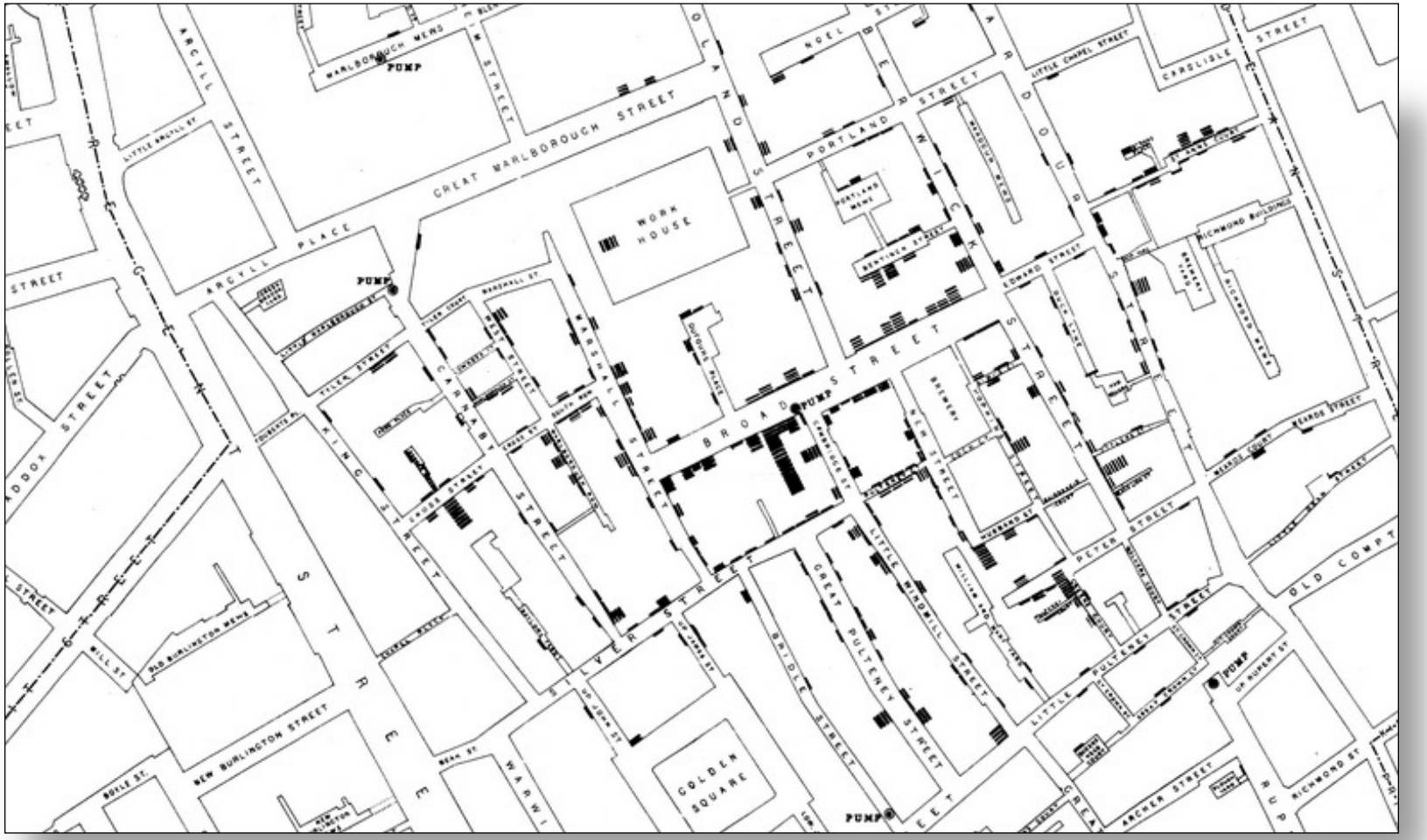
ABERDEEN 2040

What Do We Mean by Data Visualisation

As a noun, a data visualisation is a visual representation of information.

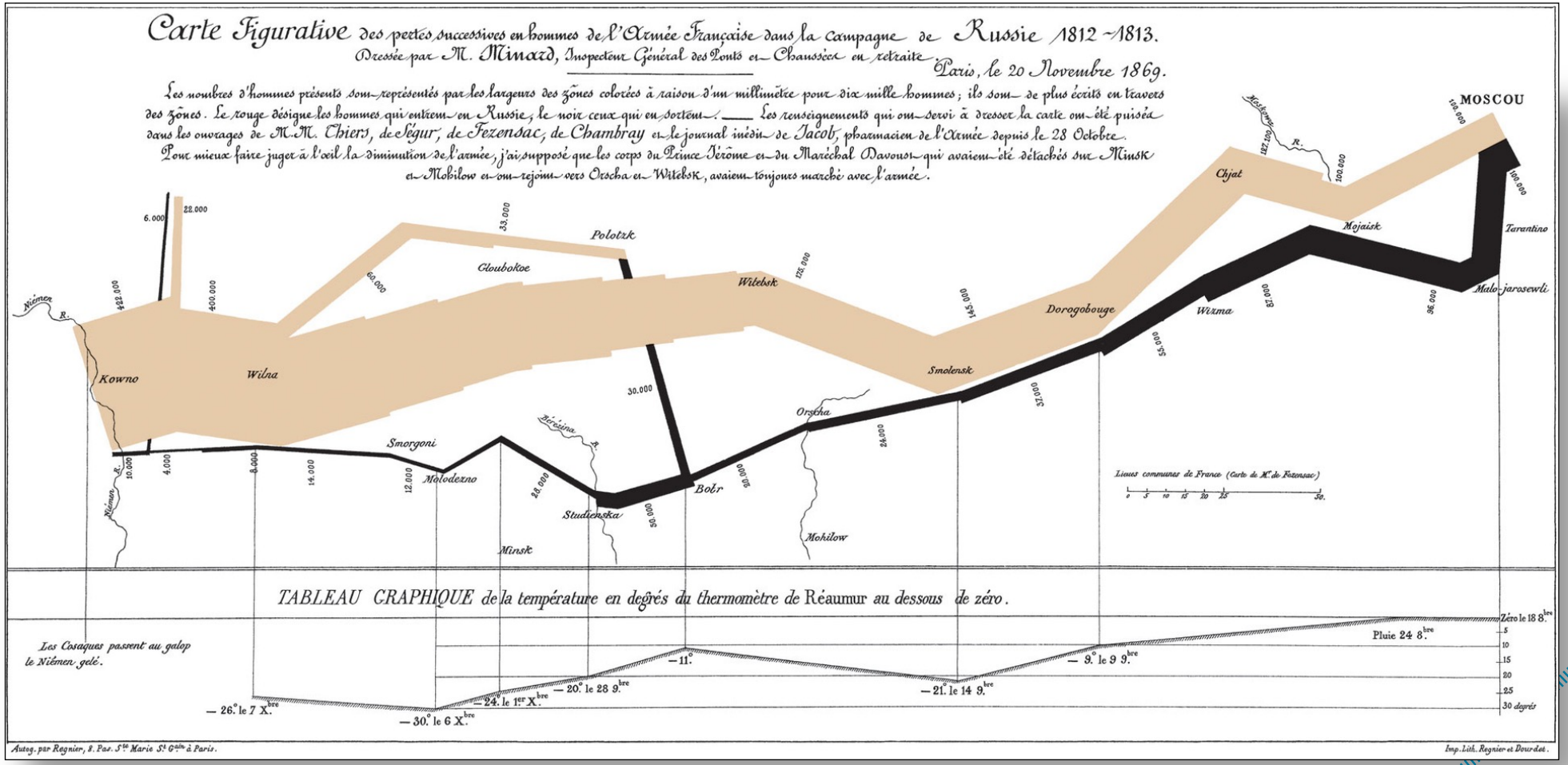


Famous Data Visualisations



Dot map of cholera deaths during the 1854 London cholera outbreak
John Snow (1860)

Famous Data Visualisations



Movements of the Napoleonic Army during the Russian Campaign of 1812-1813
Charles Joseph Minard (1869)

Famous Data Visualisations

London Underground map
Harry Beck (1933)

What Do We Mean by Data Visualisation

More broadly, Data Visualisation (DataVis) is the *process* of designing visual representations, in a form that is easy-to-understand.

In other words, we *show* the data, and let that visualisation inform the decisions being made.

Communicating With Your Audience

We say 'easy-to-understand' -- but for whom?

People will have a broad range of experiences and skills.

When creating effective visualisations, it is important to design for a broad range of people and skillsets.

Who your 'audience’ is will be highly contextual, but we'll talk about some of these considerations soon!

Communicating With Your Audience

In other words, DataVis can help explain complex patterns, trends, and observations to a wide range of people.

Show *them* the data, and help guide *their* decisions.

In this way, it is a very powerful tool for data scientists!

Common Types of Data Visualisations

Common Types of Data Visualisations

Recall our lecture about data types:

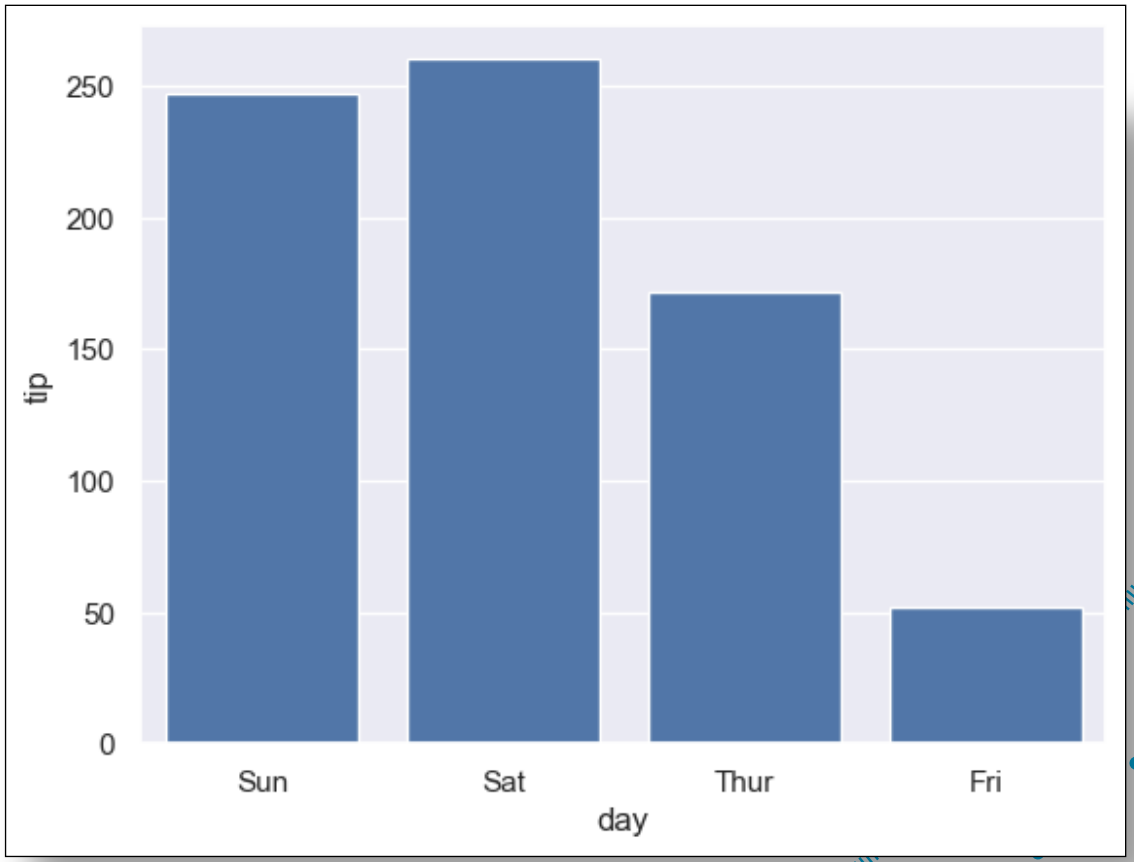
- Categorical vs Quantitative
- Discrete vs Continuous

The types of data that we are working with will also inform which types of visualisations are most appropriate.

Bar Charts

Bar charts allow us to explore the relationship between a categorical variable and a quantitative variable.

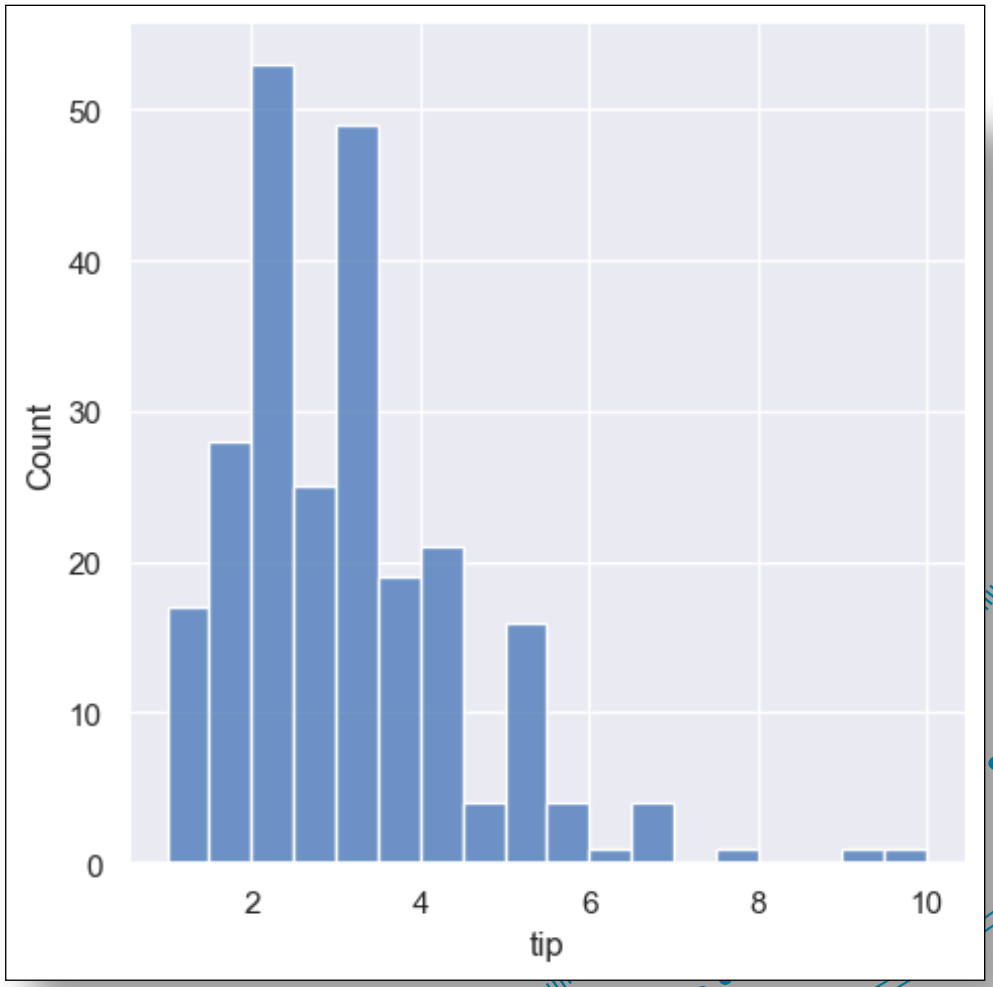
Example: Does the day of the week affect the amount of tips received?



Histogram

Histograms are similar to bar charts, but allow us to view the distribution of categorical or quantitative variables.

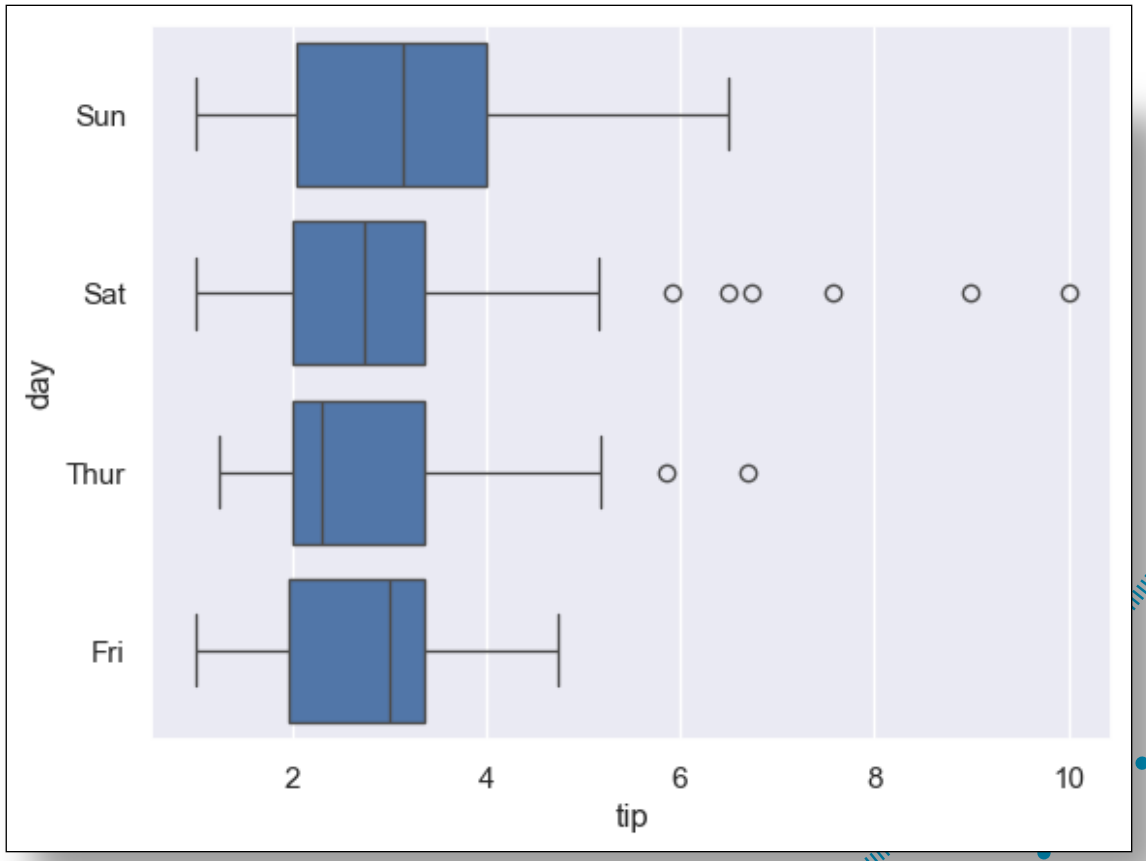
Example: What are the distribution of tips?



Boxplots

Boxplots are used show the distribution of quantitative variables across different categorical variables.

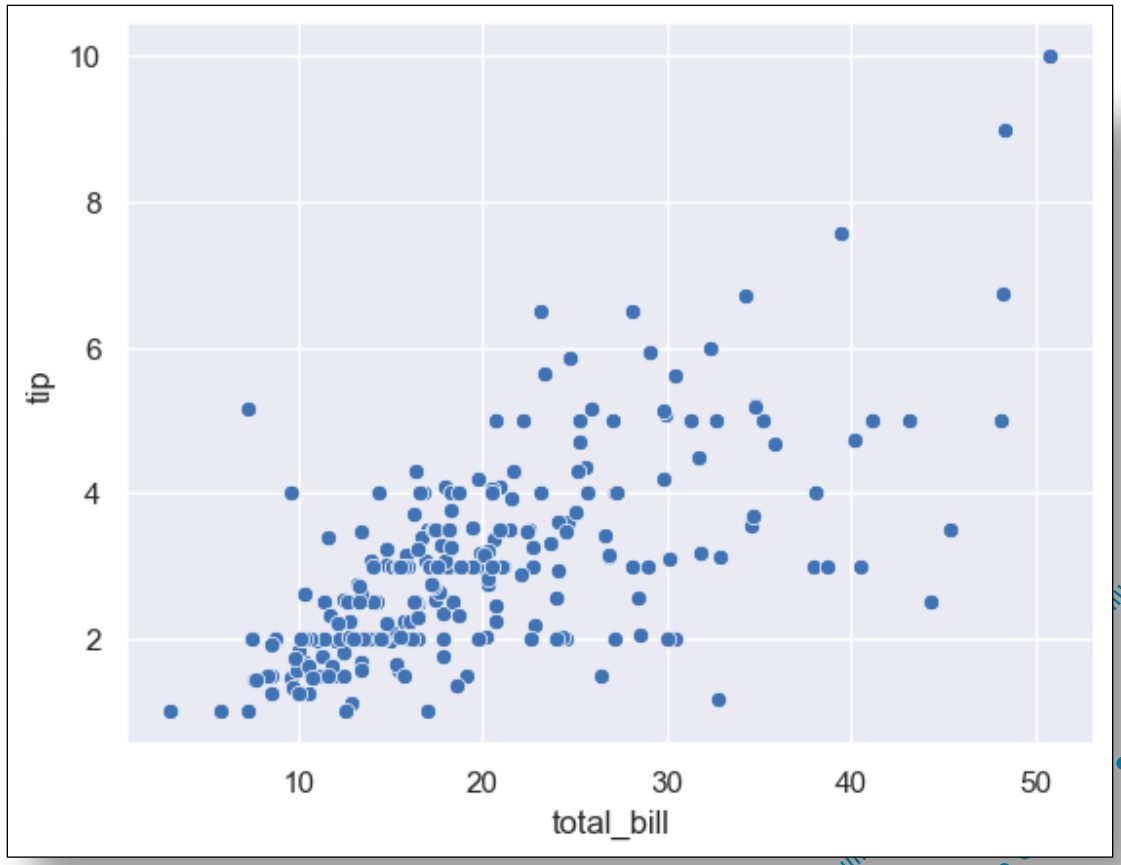
Example: What are the distribution of tips for each day?



Scatter Plot

Scatter plots are used to explore whether a relationship might exist between two quantitative variables.

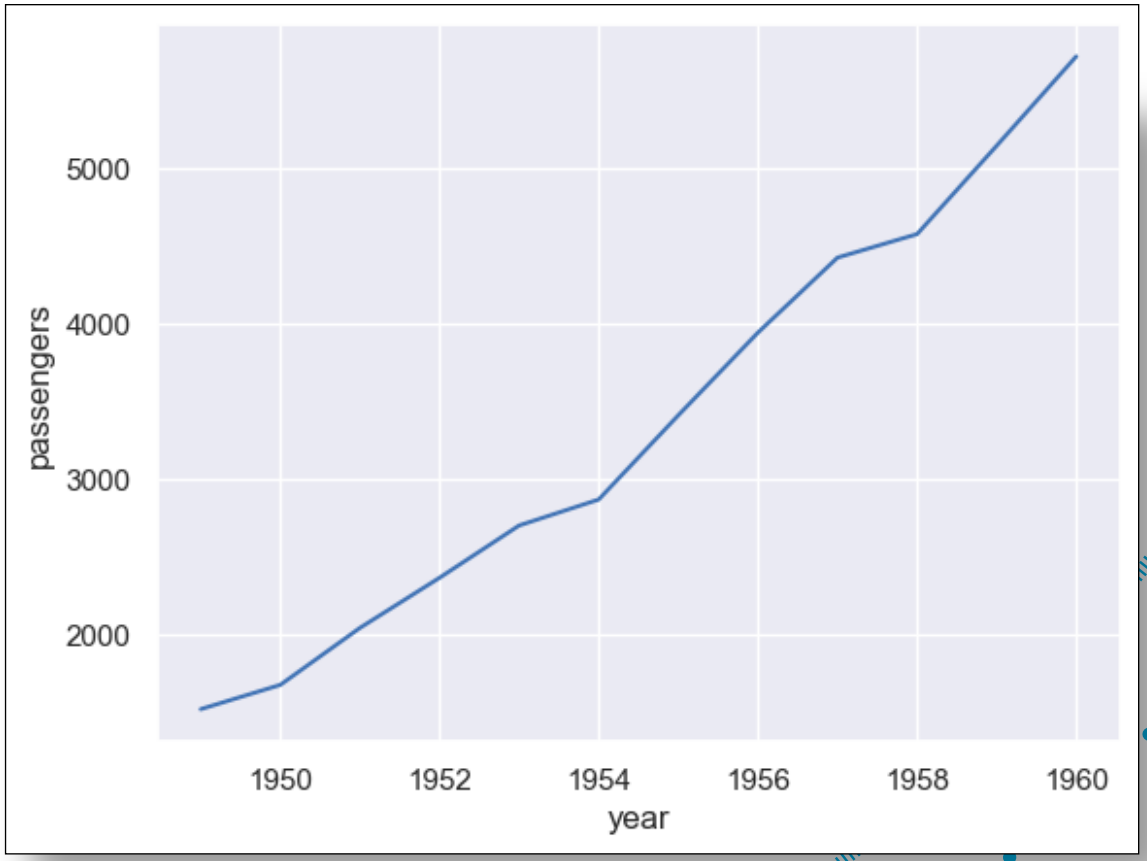
Example: How did the number of flight passengers change over time?



Line Graph

Line Graphs are used show the relation between two quantitative variables.

Example: How did the # of flight passengers change over time?



Other Popular Types of Visualisations

Various other forms of visualisations exist.

Some of the more popular examples include:

- Radar plots
- Violin plots
- Bubble plots
- Lattice plots

Customising Visualisations

Importantly, default Python visualisations will rarely be as clear and as understandable as they could be.

Libraries will often have various options to customise the visualisations, and it will be down to you to adjust them.

Remember: The aim is to make visualisations that speak for themselves!

The Grammar of Graphics

ABERDEEN 2040

The Grammar of Graphics

Introduced by statistician Leland Wilkinson, the Grammar of Graphics is a framework for describing data visualisations.

It gives us a structured and modular approach for designing and understanding the core components of visualisations.

It has inspired various data visualisation tools and libraries widely in use today (including Seaborn, Tableau, ggplot2, etc.).

The Grammar of Graphics

It breaks down a visualisation into several layered core components, including:

- Data
- Aesthetics
- Geometries
- Scales
- Facets
- Coordinate systems
- Statistical transformations

The Grammar of Graphics – Data

The foundation of our entire visualisation is the data.

As such, our starting point is always to identify the raw data (or dataset) that we want to visualise.

From there, we can start to adapt and iterate upon our visualisation.

The Grammar of Graphics – Aesthetics

Next, what are the aesthetic properties of our visualisation?

What variables should map to visual elements, such as:

- Our x axis
- Our y axis
- Colour
- Size
- Shape
- Etc.

The Grammar of Graphics – Geometries

Geometries specify the shape or form that your data takes in the plot. It determines type of plot that is created.

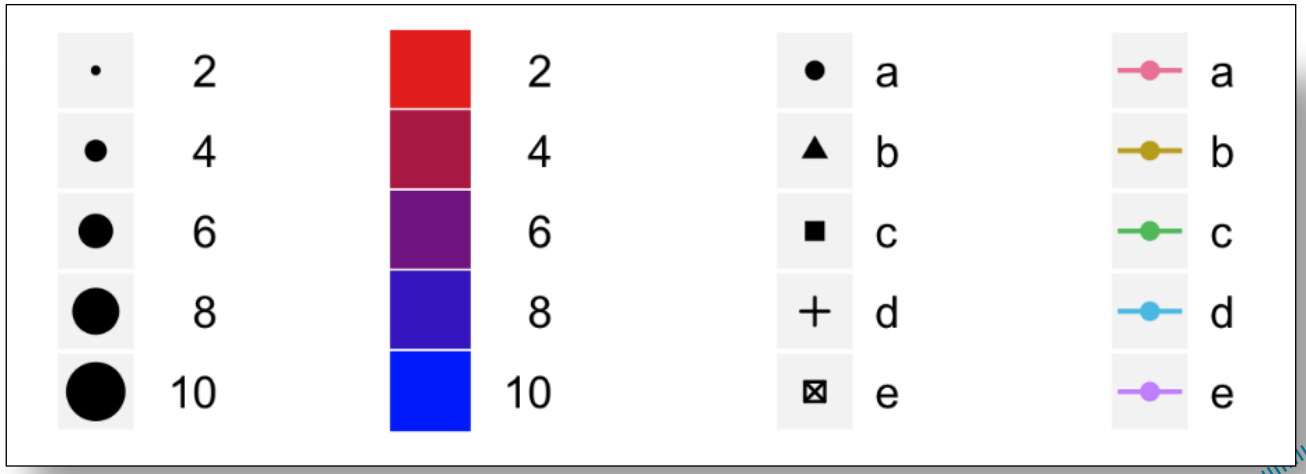
Some common geometric objects (or 'geoms') include:

- Points (for scatter plots)
- Lines (for line graphs)
- Bars (for bar charts and histograms)
- Boxes (for box plots)

The Grammar of Graphics – Scales

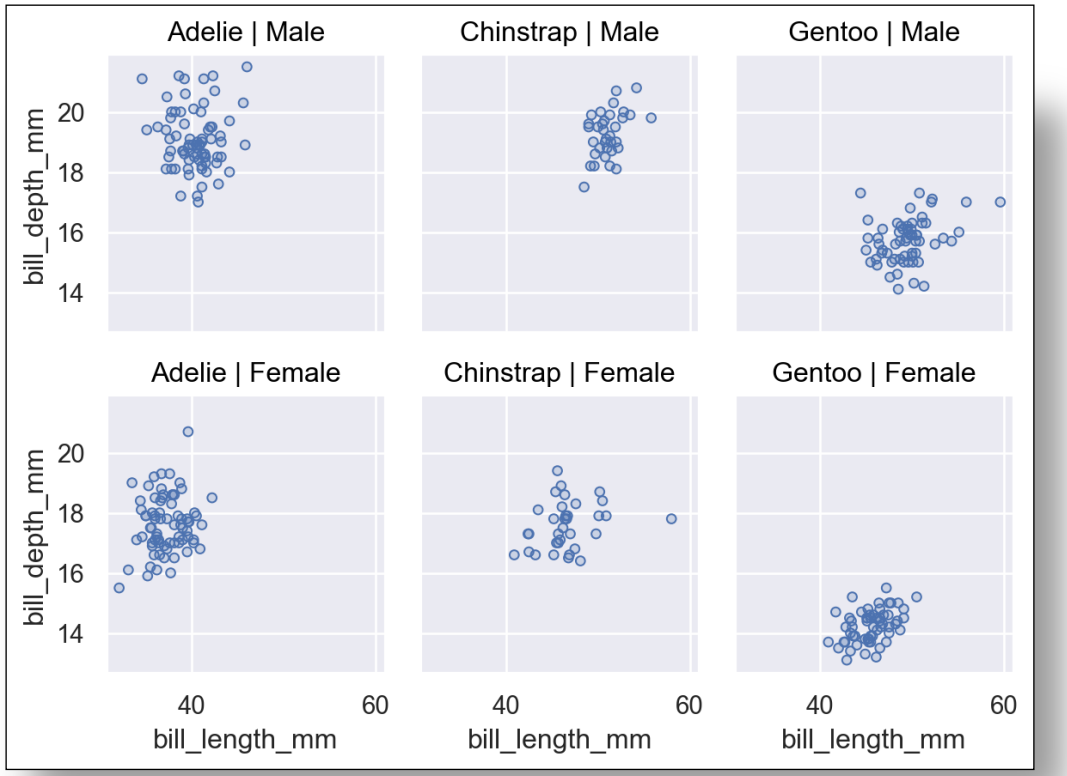
Scales control the mapping from *the values* of our data to the aesthetic attributes of the visualisation.

How should the values of the data map (or scale) to aspects such as individual colours, opacity values, shapes, etc.



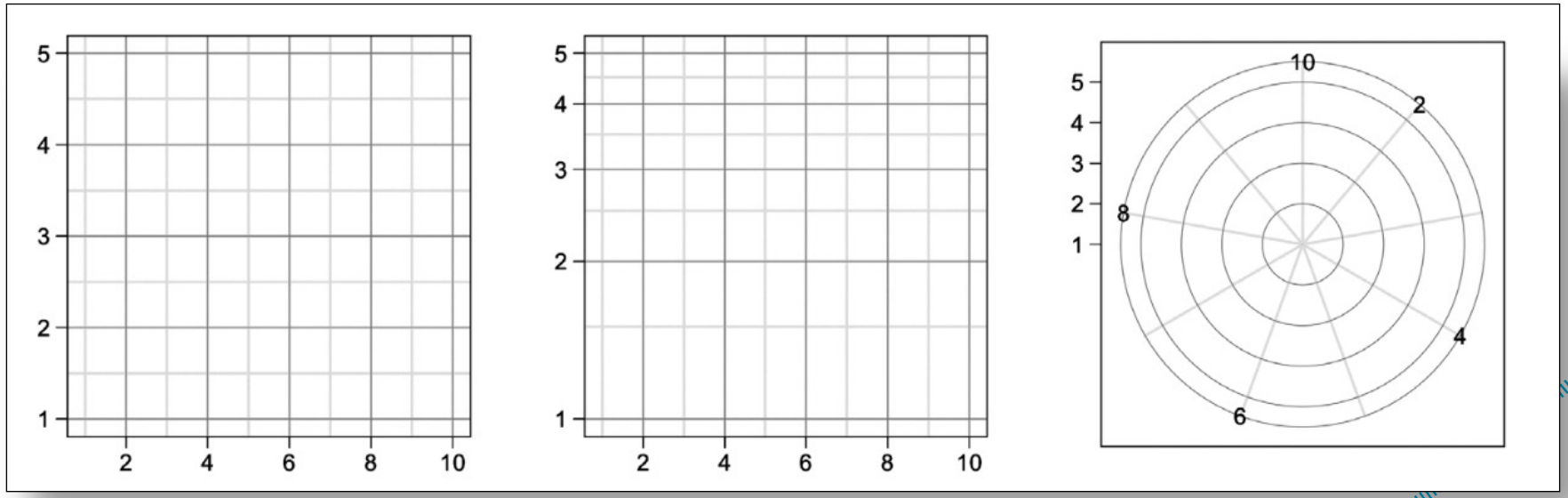
The Grammar of Graphics – Facets

Facets can be used to generate multiple subplots, based on the values of one or more categorical variables. E.g.:



The Grammar of Graphics – Coordinates

Coordinate systems control how data points are laid out on the plot. Some of the more common examples include cartesian, log, and polar coordinate systems.



The Grammar of Graphics – Transformation

Statistical transformations transform the data, often by summarising them in some manner.

This might include smoothing the data, dividing the data into buckets, adding small amounts of random noise, etc.

The Grammar of Graphics

In all, the Grammar of Graphics provides us with a common language for building the ‘layers’ of data visualisations.

By doing so, we can build up highly specific, complex, and tailored visualisations based on what we need.

Customising Visualisations

For your own visualisations, you'll need to think about:

- What variables are called in the axes
- Whether the axis limits are appropriate
- Whether the graph makes sense (data dredging)
- Whether the graph is clear and accessible
- Whether the graph has an accurate title

And of course: Whether you are using the right kind of visualisation in the first place!

Risks of Ineffective Visualisations

However, just like we saw in the previous lecture:

Ineffective visualisations can mislead people, and cause them to draw incorrect conclusions.

As such, effective visualisations are hugely important when it comes to communicating data.