# Intro Visualisation

CAMBRIDGE SPARK

# Why visualise data?

Tables and numbers can be hard to make sense of, a good visualisation can quickly allow to spot trends, different aspects and relations in our data.

We can use many different types of plots to visualise different information from our data set, such as:

- **box plots** to compare relative scales of features
- **scatter plots** for relations between features
- **heatmaps** visualise third entity value on 2D plot

CAMBRIDGE SPARK

# Visualising data
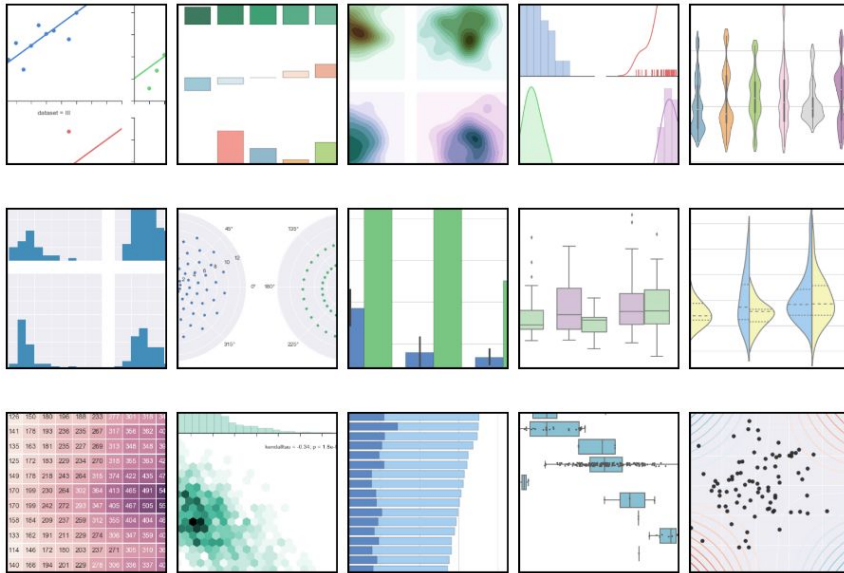


Some important questions to consider:

- What data do I want to plot?

- What type of plot is suitable?

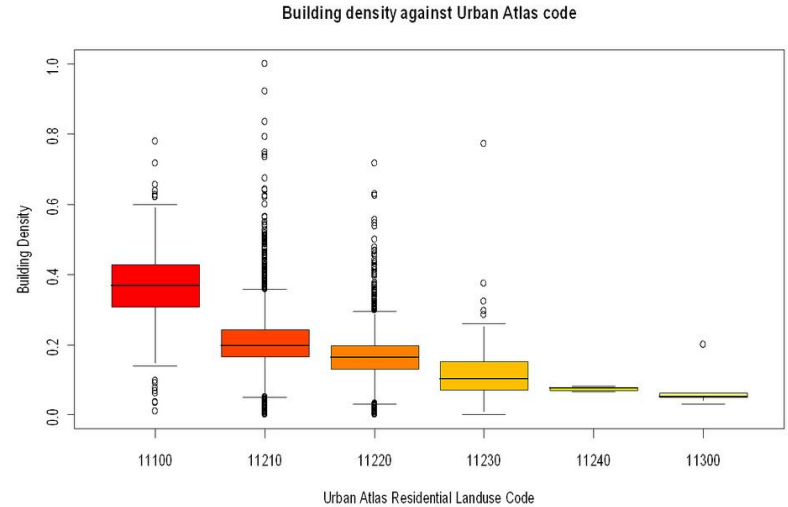- How to convey a message with a plot?

CAMBRIDGE SPARK

# Scatter plots

- Easily spot correlation between two variables

- Each dots X and Y values represented by the variables

- Size and shape of dots can represent additional variables

- A pair plot compares all pairs with each other



CAMBRIDGE SPARK
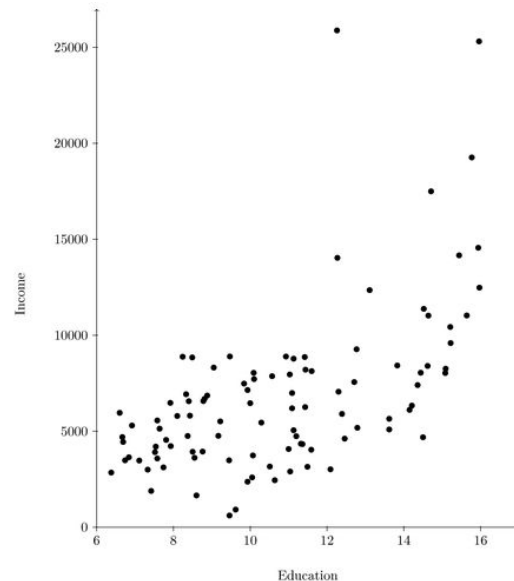
# Box plots

- Show the shape of distribution

- Box spans the interquartile range

- Median shown as vertical line in the box

- Ideal for visualising variable scales



Building density against Urban Atlas code

Building Density

Urban Atlas Residential Landuse Code

CAMBRIDGE SPARK

# Heat maps

- Individual values of a matrix represented as colours

- Colours change proportionately to values in the matrix

- Ideal for representing correlation matrix between variables of our dataset
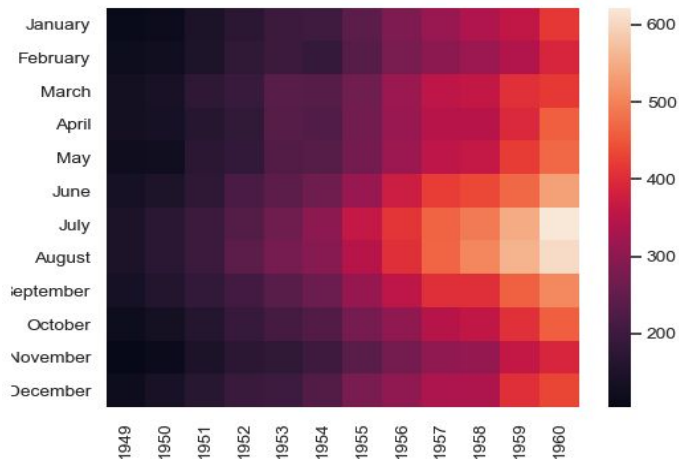


Flights/month/year

# What plot to use?

- Show daily activity on a given social media

- Determine distribution of heights of students in a class

- Checks a car's gas mileage based on horsepower

CAMBRIDGE SPARK

# In Python

- **Matplotlib**: the main library on which others are built. More verbose but allows more flexibility

- **Seaborn**: Built on top of matplotlib, provides fancy plots out-of-the box and works well with Pandas

- "**Pandas plots**": DataFrames have a "plot" method that allows to quickly generate simple plots

- **Bokeh**, **plotly**, etc…

CAMBRIDGE SPARK

Hands-on session

*visualisation.ipynb*