Session 06: Visualisation

wifi: GA-Guest, yellowpencil

GENERAL ASSEMBLY

```
cd ~/Documents/ga-ldn-ds37
git commit -am "your commit message here"
git pull
```

Today's session plan

1800-1845	Standup & Session 05 recap
1845-1900	What makes a good visualisation?
1900-1920	Break
1920-1930	Visualising distributions
1930-1945	API requests with Python, intro to Pandas
2000-2100	Project ideas & finding datasets
Homework: Finalising project ideas	

At the end of the session, you will be able to ...

Understand why visualisation is important

Choose an appropriate visualisation to communicate a message

Use Matplotlib and seaborn to produce visualisations

Data Science Part Time

Review

Computers Out:

Exploring four datasets

1. Open up the notebook ds37-06-01.ipynb
2. Read in the four datasets as instructed
3. Use `describe` to generate high level summary statistics for the datasets
4. What can you conclude about the four datasets?

Data Science Part Time

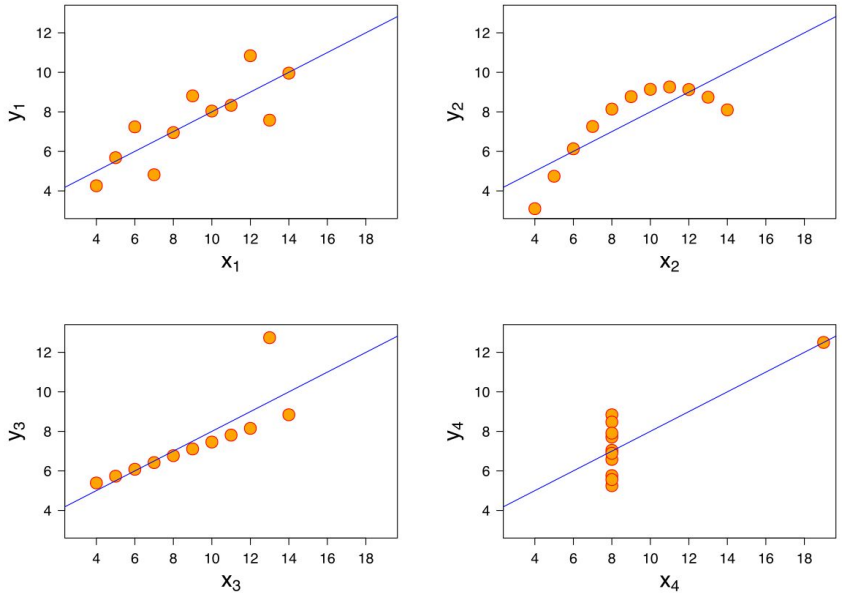
Why is visualisation important?

Anscombe's quartet

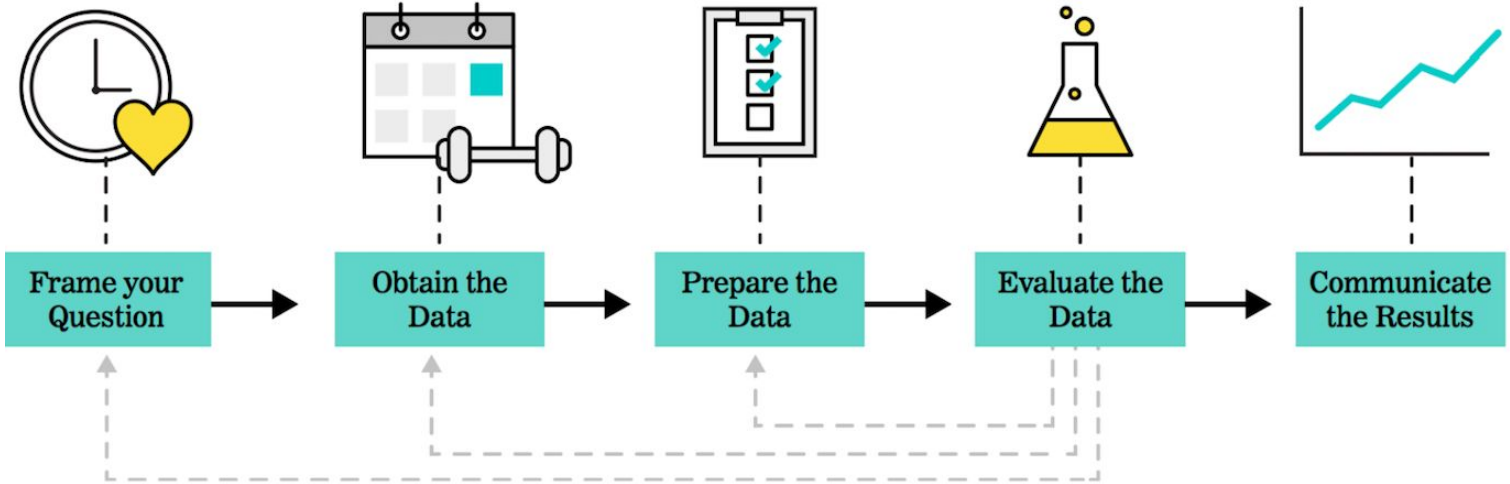
The datasets in the last exercise form **Anscombe's quartet.**

This is a set of four datasets with identical summary statistics but showing very different relationships when visualised.

They demonstrate that statistics can give an incomplete or misleading impression of data, and that visualisation is an important sense check at **all** stages of the data science workflow.



Where are we in the data science workflow?



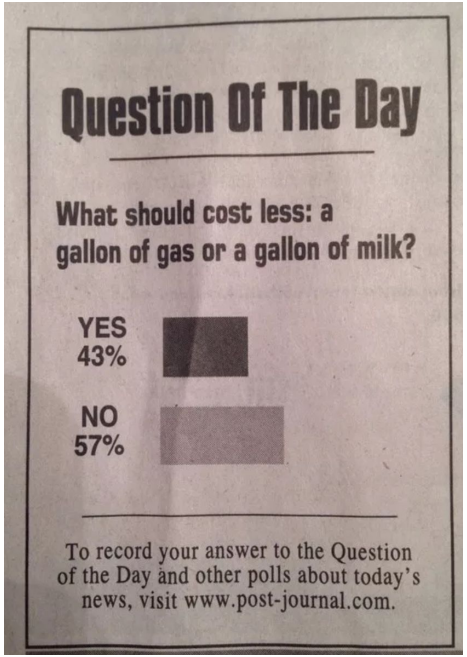
Data Science Part Time

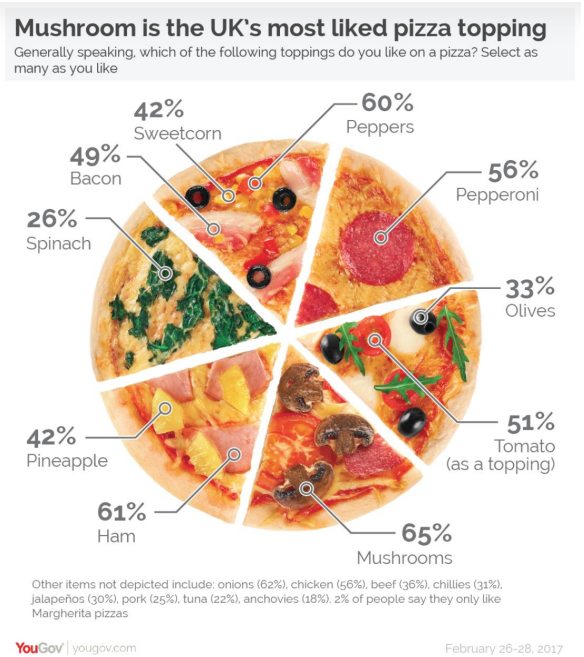
What makes a good visualisation?

Real Cases:
The Good, The Bad and The Ugly

3 minutes

i.imgur.com / Via reddit.com





Group Exercise:

Good and bad visualisations

Visit https://www.reddit.com/r/dataisbeautiful/ and find one example of a **good** data visualisation, and one example of a **poor** data visualisation.

Be ready to discuss your choices and reasons with the class.

Fundamentals of data visualisation

Using data to communicate is a balancing act between detail and ease of use. Which visualisation to choose depends on the message to be communicated. We can refer to a **visual vocabulary** like the one produced by the Financial Times to help us decide which category our visualisation falls into:

Correlation To show relationships between variables

Ranking To show an item's position in an ordered list

Distribution To show the spread of values in a single variable

Change over time Showing changing trends

Magnitude Size comparisons

Part-to-whole Show how a single quantity is broken down into components

Deviation To emphasise variations from a reference point or target

Spatial For representing geographical data

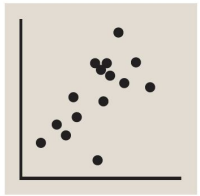
Flow Shows movement between two or more conditions

Group Exercise:

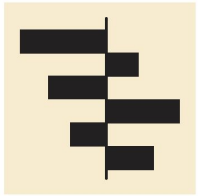
Matching visualisations

Which of the following visualisations would you use to show: flow, ranking, magnitude, distribution, part-to-whole, deviation, change over time, correlation, spatial patterns? (source FT visual vocabulary)

Scatterplot

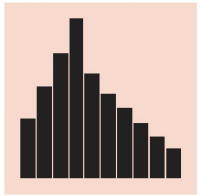


Diverging bar

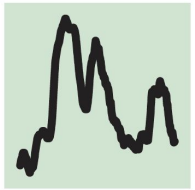


Ordered bar

Histogram

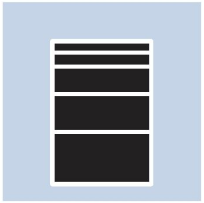


Line



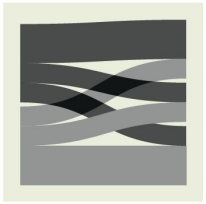
Column

Stacked column/



Basic choropleth

Sankey



Group Exercise:

Choosing appropriate visualisations

Refer to the FT visual vocabulary (in your current folder) to decide on the most appropriate way to visualise:

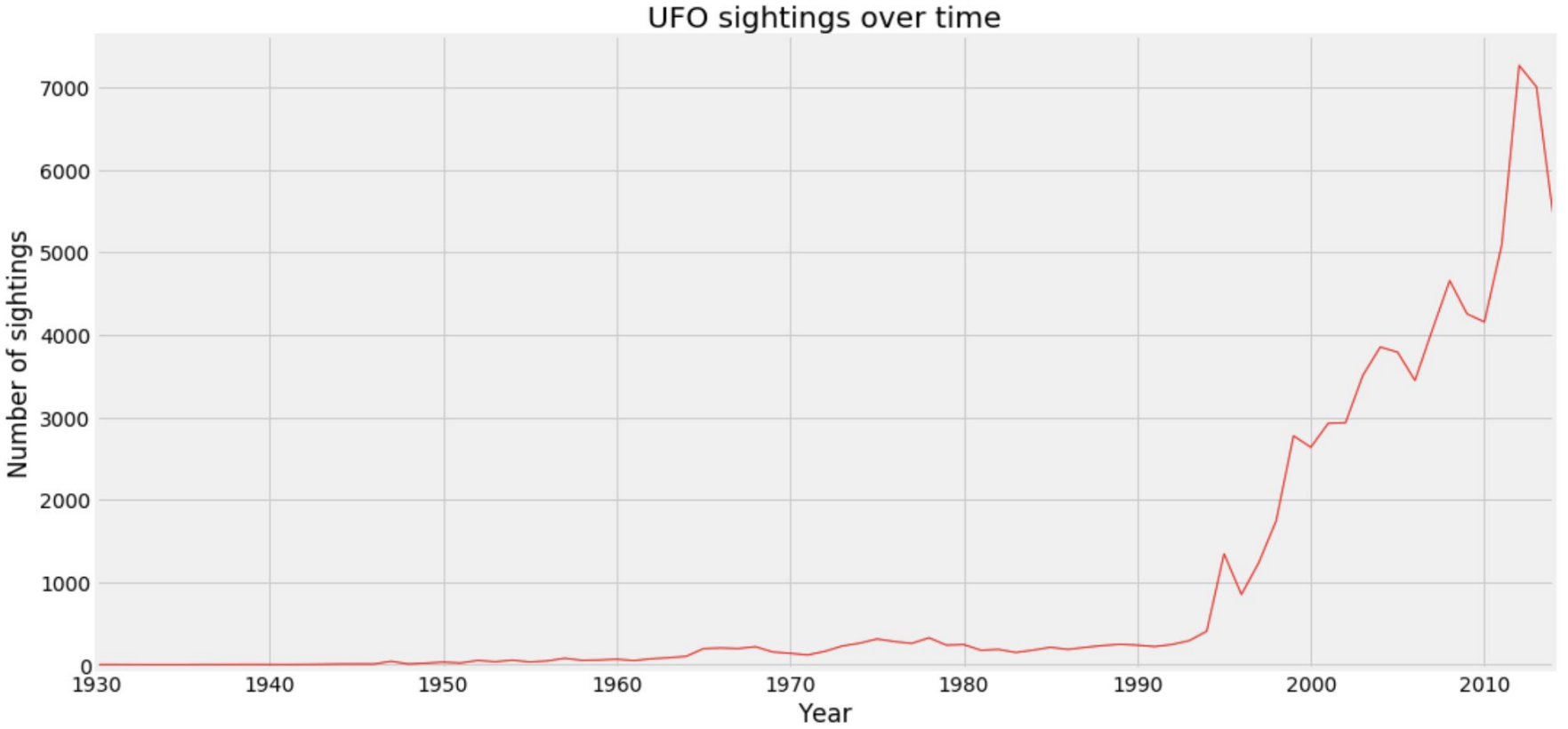
1. The size of the gender pay gap for different industries
2. US presidential election results
3. The results of a vote in Parliament
4. Changes in Google's share price

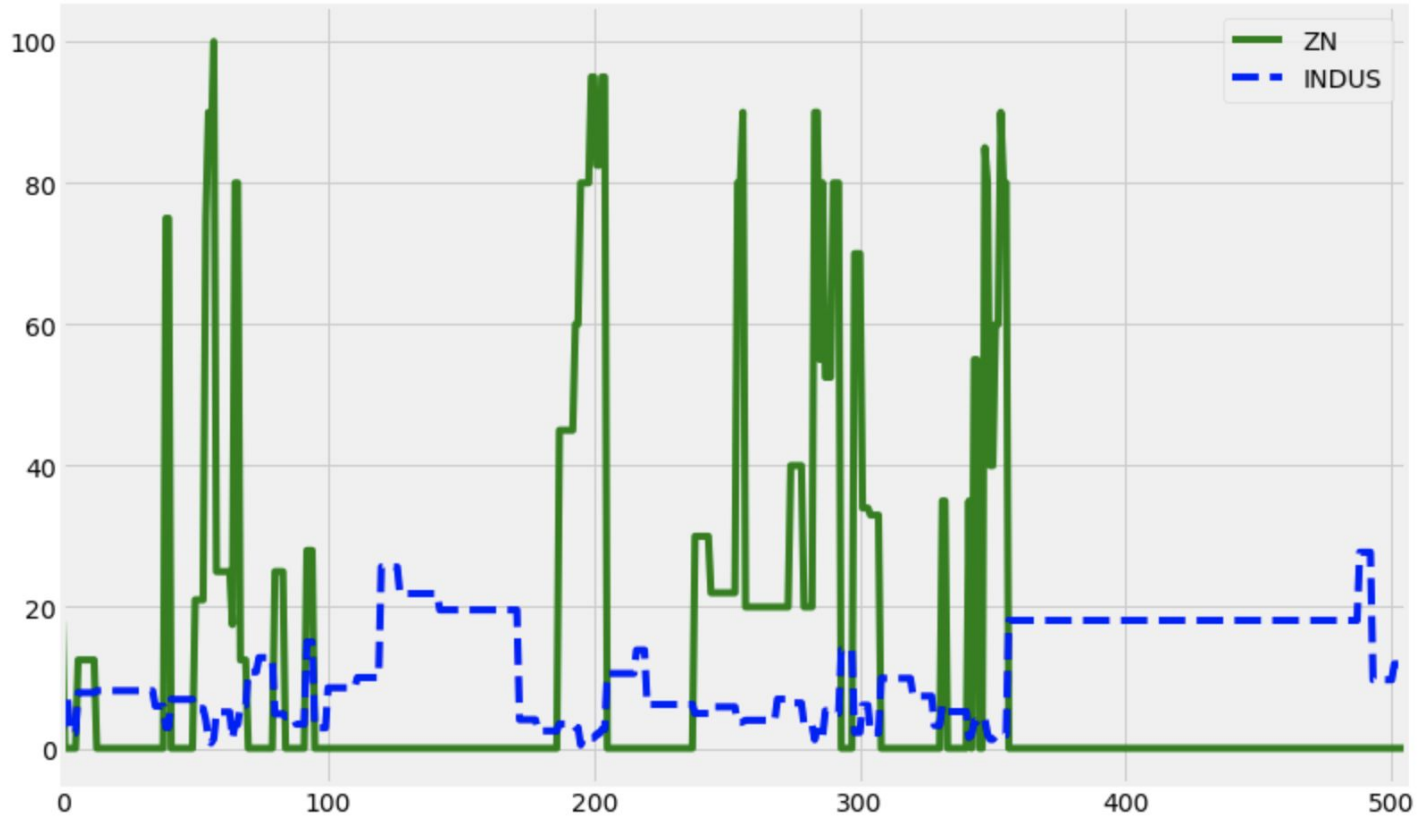
Computers Out:

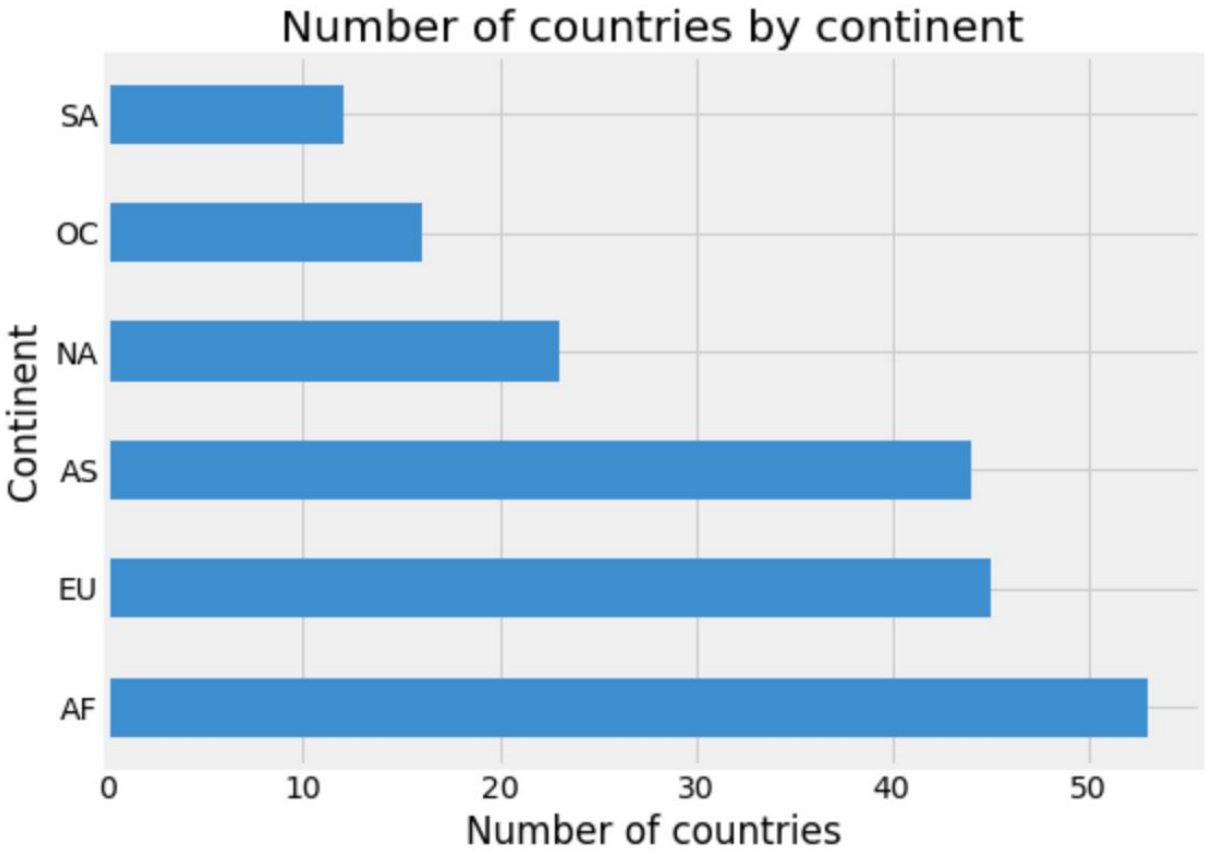
Let's start plotting

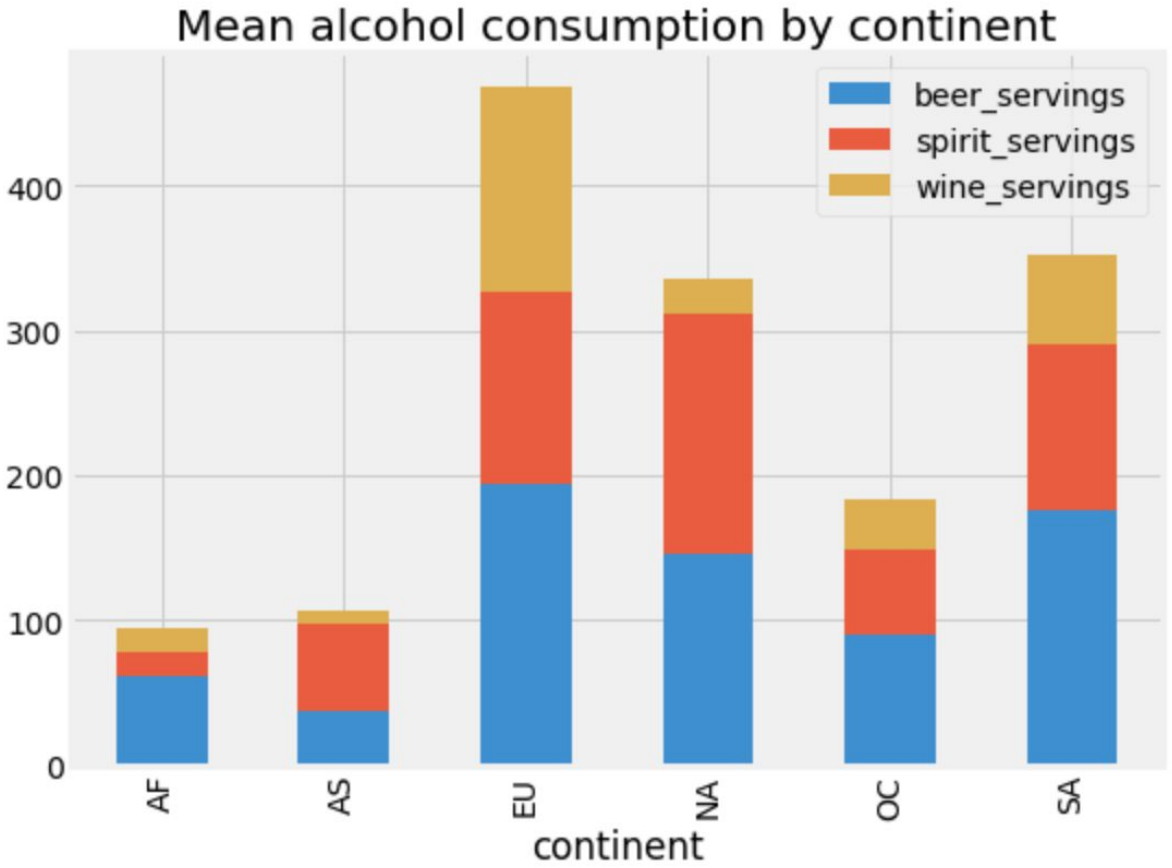
Open the notebook ds37-06-02.ipynb

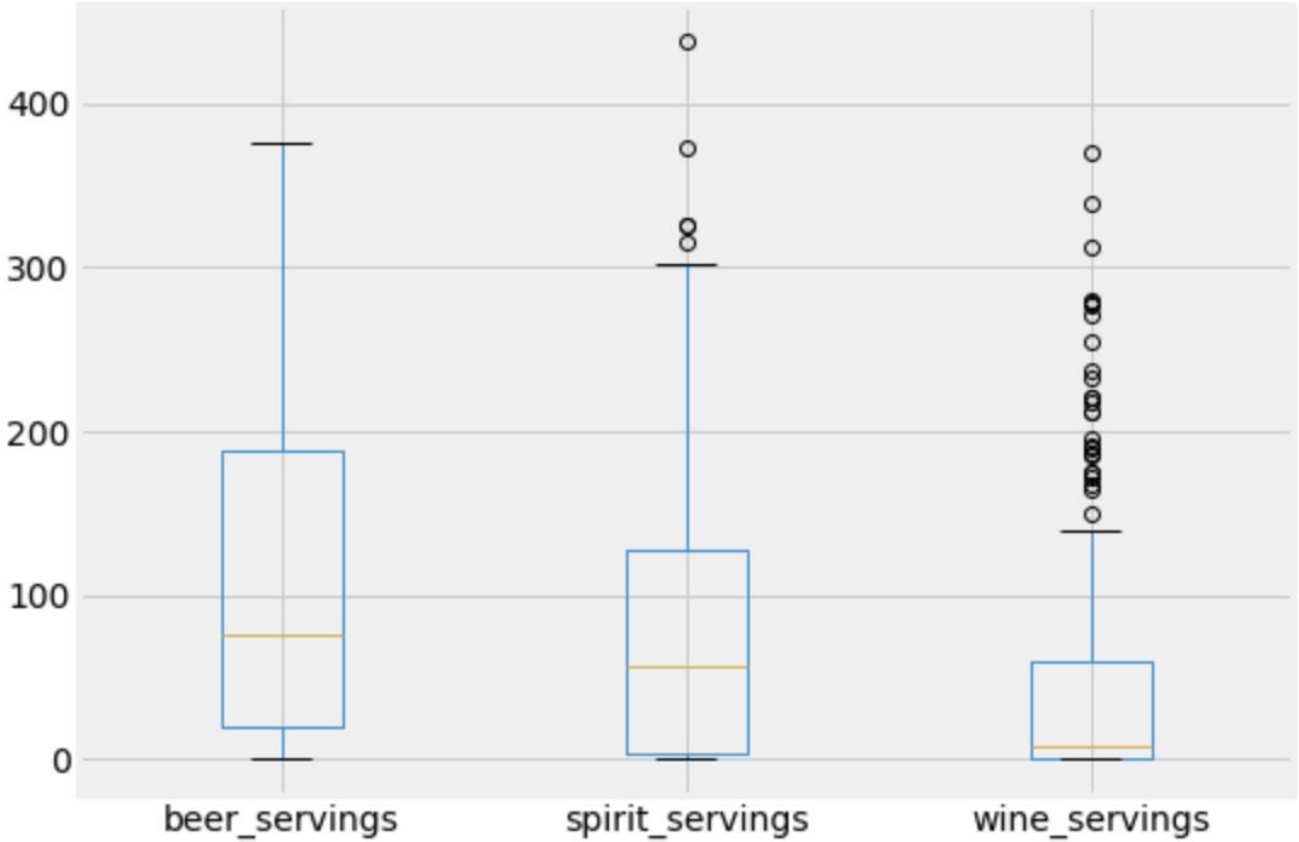
Let's start producing data visualisations using two new libraries; **matplotlib** and **seaborn**

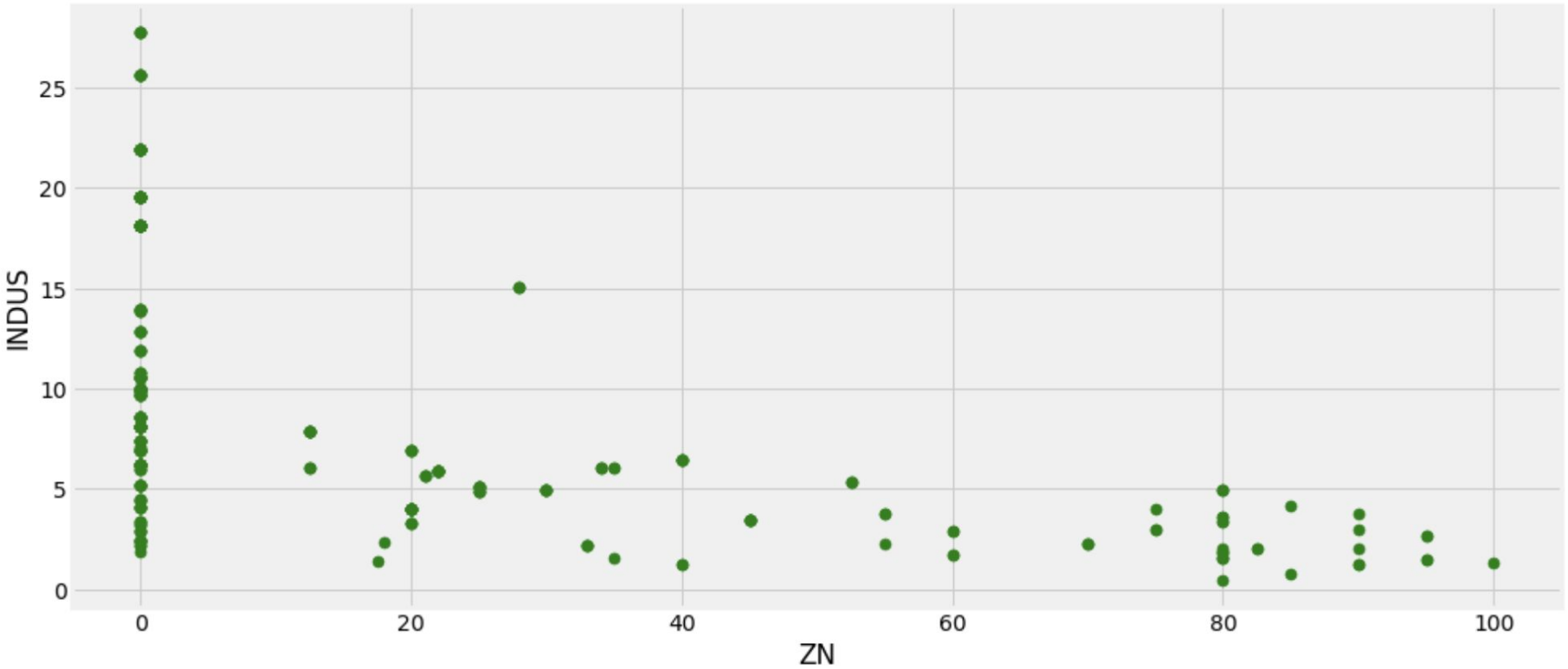











INDUS
25
20
15
10
5
0
0
20
40
60
80
100
ZN

Intro to Python

Let's Review

At the end of the session, you will be able to ...

Understand why visualisation is important

Choose an appropriate visualisation to communicate a message

Use Matplotlib and seaborn to produce visualisations

Feedback

Take a minute to fill out our end of week survey:

http://bit.ly/ds37weekly

Coming up next week...

- Statistics and maths refresher
- Experiments and hypothesis testing


GA