LEARNING LAB, Part 1

Learning Lab, Part 1:

✓The Data Science Process

✓Asking Questions of Data

✓Collecting and Sourcing Data

The Data Science Process

Asking Questions of the Data

Collection

Exploration

Modelling

Communicate the Results

Learning Lab
Part 1

- - - - - - - - -

Part 2

Asking Questions of Data

Questions to consider:

> What is the goal / problem?

> Is the problem quantifiable?

> Domain knowledge

Asking Questions of Data

Who?

Target audience

Why?

Desired effect

What?

Intended meaning

Asking Questions of Data

DEMO

Road Traffic Casualties in Leeds

What questions would you like to answer? Please come up with two questions of your own. Also consider Who, What and Why questions

Asking Questions of Data

Defining the Goal

Practical Example, Learning Lab Part 1:

Where do cycling crashes cluster in Leeds? What factors are significant in causing road casualties?

Collecting Data

Practical Example:

Collecting Data

> What data do we have available?

> What format is it in?

Is it… .xlsx .csv database

> Is the data good quality?

Are there missing values, empty columns, outliers, strange characters?

Collecting Data

DEMO

Collecting Data

Let’s look at the data we have available on Github and Leeds Data Mill.

Collecting Data

<u>On Data Quality</u>

Common problems:

> missing (NA) values

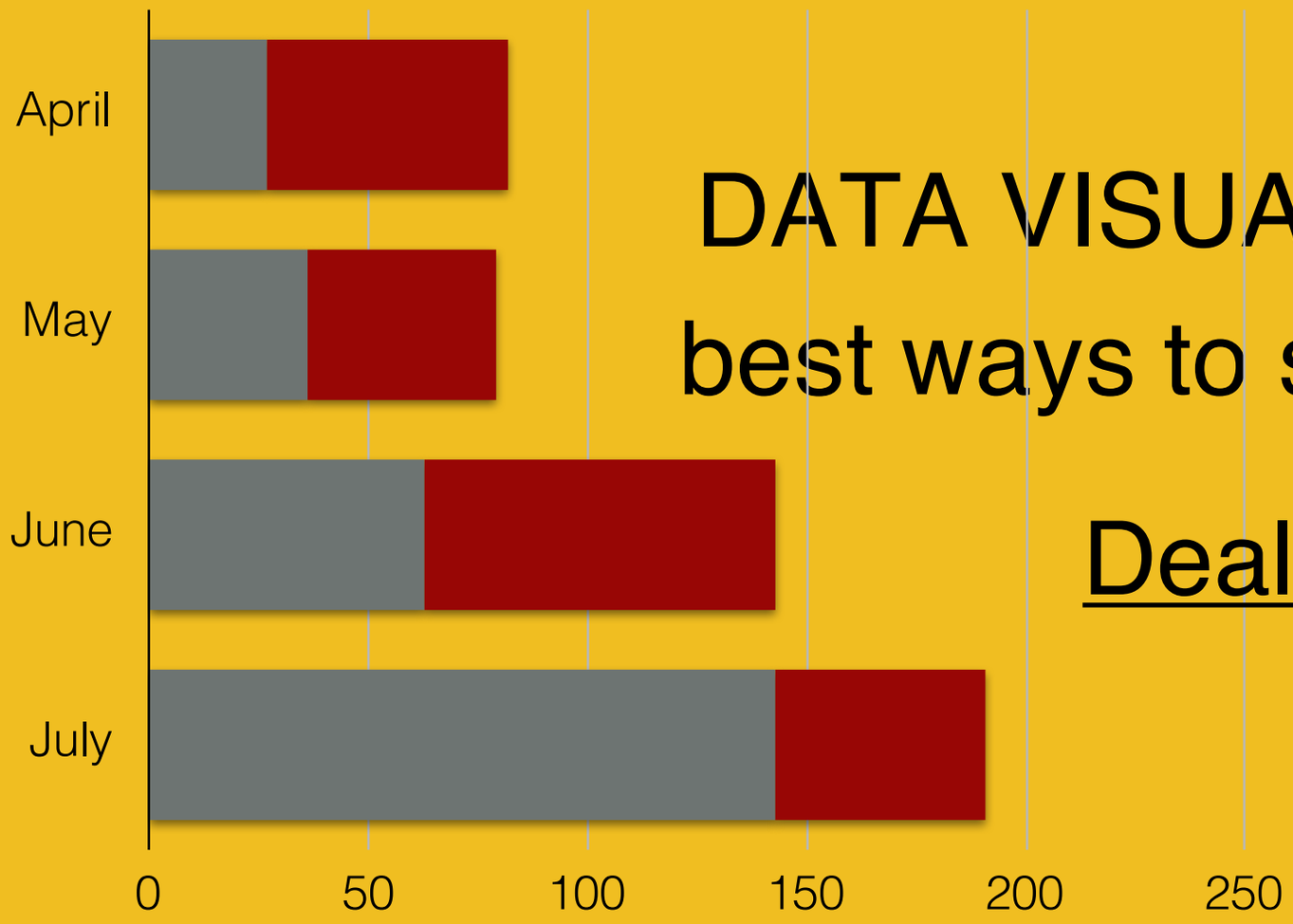
> empty columns

> outliers

> data entry mistakes

> unnecessary characters &@?l+$£

End of Learning Lab, Part 1



<u>How to spot problems</u>

DATA VISUALISATION is one of the best ways to spot problems with data

<u>Dealing with Data Problems</u>

DATA CLEANING

Next…. LEARNING LAB, PART 2