Welcome to CM4125 - Data Viz!

- Dr Carlos Moreno-García
- Senior Lecturer & Research Degrees Co-coordinator
- School of Computing, Engineering and Technology
- email: c.moreno-garcia@rgu.ac.uk

Module aims and objectives

Module aims and objectives

1. Critically appraise a variety of data visualisation techniques in terms of psychology, design, effectiveness and appeal to a wider audience.

Module aims and objectives

1. Critically appraise a variety of data visualisation techniques in terms of psychology, design, effectiveness and appeal to a wider audience.

2. Interpret data visualisations and explain what conclusions can be drawn from them in terms of data analysis.

Module aims and objectives

1. Critically appraise a variety of data visualisation techniques in terms of psychology, design, effectiveness and appeal to a wider audience.

2. Interpret data visualisations and explain what conclusions can be drawn from them in terms of data analysis.

3. Understand the challenges of visualising large datasets.

Module aims and objectives

1. Critically appraise a variety of data visualisation techniques in terms of psychology, design, effectiveness and appeal to a wider audience.
2. Interpret data visualisations and explain what conclusions can be drawn from them in terms of data analysis.
3. Understand the challenges of visualising large datasets.
4. Translate numerical and categorical data into coherent pictorial representations.

Module aims and objectives

1. Critically appraise a variety of data visualisation techniques in terms of psychology, design, effectiveness and appeal to a wider audience.
2. Interpret data visualisations and explain what conclusions can be drawn from them in terms of data analysis.
3. Understand the challenges of visualising large datasets.
4. Translate numerical and categorical data into coherent pictorial representations.
5. Create novel and interactive data visualisations which lucidly exhibit particular dataset features using publicly available data.

How the module will work

What this module **is** about?

What this module **is** about?

- How to get proper data from various sources

What this module **is** about?

- How to get proper data from various sources
- How to manipulate that data to have only what we are interested on

What this module **is** about?

- How to get proper data from various sources
- How to manipulate that data to have only what we are interested on
- How to put that data into meaningful, interesting and interactive pictorial representations

DATA

SORTED

ARRANGED

PRESENTED
VISUALLY

EXPLAINED
WITH A STORY

What this module **is not** about?

What this module **is not** about?

- Computer vision

What this module **is not** about?

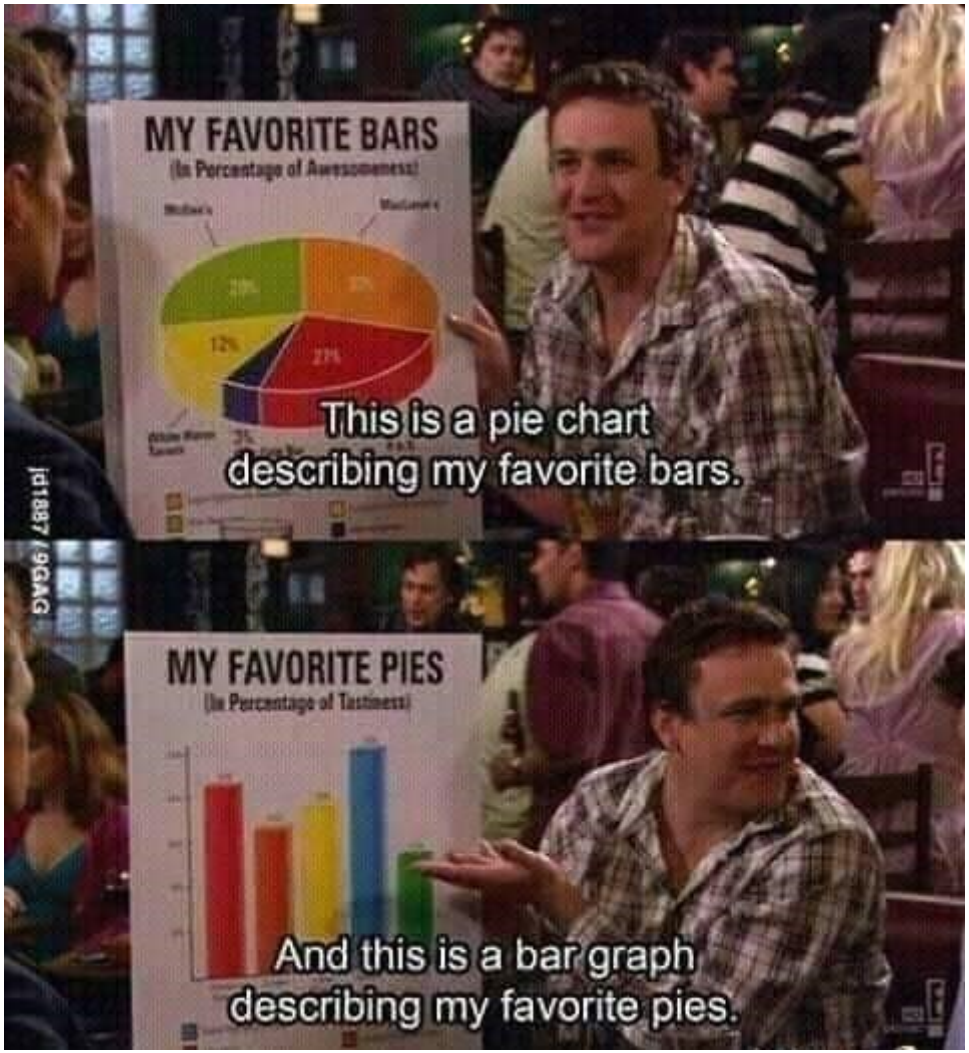
- Computer vision
- Predicting or learning from data (i.e. machine learning)

What this module **is not** about?

- Computer vision
- Predicting or learning from data (i.e. machine learning)
- Flashy images and colours (although this sometimes helps!)

What this module **is not** about?

- Computer vision
- Predicting or learning from data (i.e. machine learning)
- Flashy images and colours (although this sometimes helps!)
- Only bars and pies!


MY FAVORITE BARS
(In Percentage of Awesomeness)
This is a pie chart
describing my favorite bars.
MY FAVORITE PIES
(In Percentage of Tastiness)
And this is a bar graph
describing my favorite pies.
id18877 / 9GAG

Online teaching

Online teaching

- Lectures will run for **one hour**, *usually* every Monday at 10 am

Online teaching

- Lectures will run for **one hour**, *usually* every Monday at 10 am
- Labs will run for **two hours**, *usually* every Monday at 12 pm

Assessment

Assessment

- You will have one coursework composed of two deliverables:
    - A slide deck
    - A visualisation

Assessment

- You will have one coursework composed of two deliverables:
    - A slide deck
    - A visualisation
- Submission date: Thursday, 28th November, 2024, 13:00

Assessment

- You will have one coursework composed of two deliverables:
    - A slide deck
    - A visualisation
- Submission date: Thursday, 28th November, 2024, 13:00
- Next week, we will have a coursework clarification session during the first hour of the lab

Resources

Resources

- In the Moodle accordion, you will see a *Resources* tab.
    - There you will find a link to a Trello board where everyone can participate and add material
        - To become a board member and aadd sources, click here
    - We will also have a DataCamp group for you to do related courses
        - To join the group, click here (use your rgu email to create the account).

Support/Assessments

Support/Assessments

- Mondays 15:00-16:00
- Thursdays 10:00-11:00

Introduction to Data Viz

MENTI POLLS

Go to www.menti.com and enter the code shown in the screen

For those of you who chose Excel...

For those of you who chose Excel...

- Don't get me wrong, Excel is great for initial data exploration

For those of you who chose Excel...

- Don't get me wrong, Excel is great for initial data exploration
- However, it is **not** a proper tool for data visualisation!

- It has finite rows/cols, so by definition it is not suitable for big data!

- It has finite rows/cols, so by definition it is not suitable for big data!

- Excel tends to do weird things with data types

- Excel tends to do weird things with data types

- Excel assumes formats when typing, and autocorrections creates chaos

- Excel assumes formats when typing, and autocorrections creates chaos

In [1]:
```
from IPython.display import HTML
HTML('<iframe width="560" height="315" src="https://www.youtube.com/embed/yb2z
```

```
C:\ProgramData\anaconda3\envs\notebook\Lib\site-packages\IPython\core
\display.py:431: UserWarning: Consider using IPython.display.IFrame in
stead
  warnings.warn("Consider using IPython.display.IFrame instead")
```

Out[1]:

- Spreadsheets are very interactive, but error prone (costing millions of hours and money)

- Spreadsheets are very interactive, but error prone (costing millions of hours and money)
- A researcher was able to get 15k spreadsheets from Enron, and found that 42% didn't have a single formula/calculation!

- Spreadsheets are very interactive, but error prone (costing millions of hours and money)
- A researcher was able to get 15k spreadsheets from Enron, and found that 42% didn't have a single formula/calculation!
- From the remaining 9k, 24% contained "obvious" errors!

- Spreadsheets are very interactive, but error prone (costing millions of hours and money)
- A researcher was able to get 15k spreadsheets from Enron, and found that 42% didn't have a single formula/calculation!
- From the remaining 9k, 24% contained "obvious" errors!
- There's a thing called the European Spreadsheet Risk Interest Group (EuSpRiG), where they have Excel horror stories!

My top 4

Python

Python

- Very high level programming language

Python

- Very high level programming language
- Provides complementary tools/packages such as `matplotlib`, `dash plotly` and `Streamlit.io` that help you visualise data better

Python

- Very high level programming language
- Provides complementary tools/packages such as `matplotlib`, `dash plotly` and `Streamlit.io` that help you visualise data better
- Best ranked programming language by the IEEE

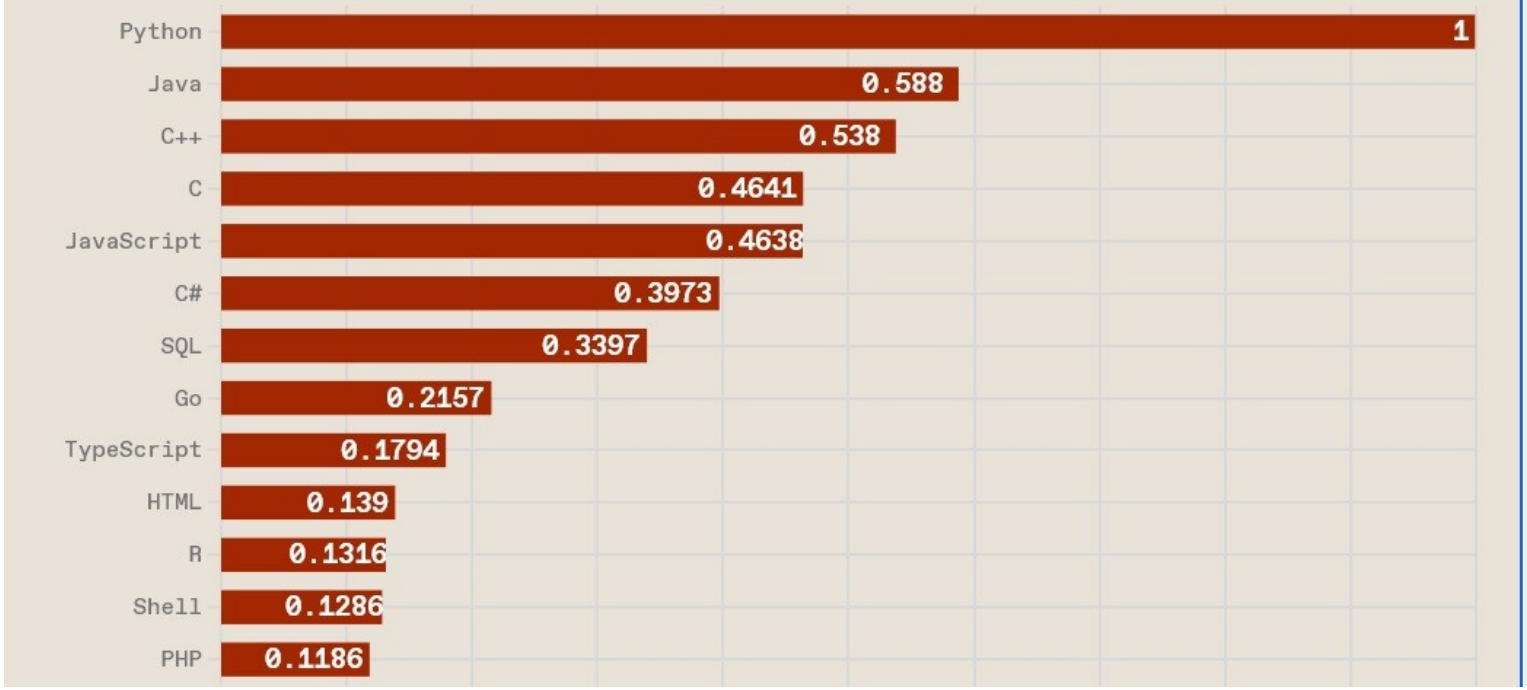
Python

- Very high level programming language
- Provides complementary tools/packages such as `matplotlib`, `dash plotly` and `Streamlit.io` that help you visualise data better
- Best ranked programming language by the IEEE

Top Programming Languages 2023

Click a button to see a differently weighted ranking

Spectrum | Jobs | Trending



- In fact, this slideshow was done using it!

- In fact, this slideshow was done using it!

In [2]:
```
print("Hello")
```

```
Hello
```

R

R

- Even higher level programming language

R

- Even higher level programming language
- Widely used in statistics

R

- Even higher level programming language
- Widely used in statistics
- Also contains numerous packages (`lattice` , `ggplot` , etc.) to do data vis

- Advantage: R gets you better salaries than Python

- Advantage: R gets you better salaries than Python
- Disadvantage(?): Exists within it's own bubble

Tableau

Tableau

- A commercial tool for business analytics to create dashboards

Tableau

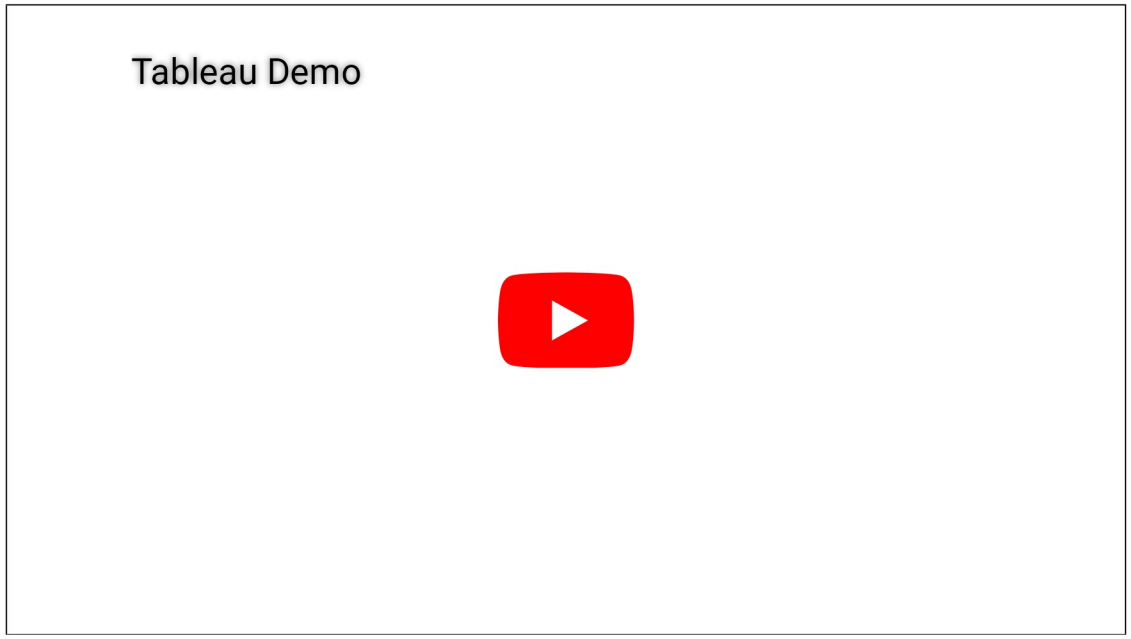
- A commercial tool for business analytics to create dashboards
- Easy to use

Tableau

- A commercial tool for business analytics to create dashboards
- Easy to use
- Connects to data from different sources and can import Python/R code

In [3]:
```
from IPython.display import HTML
HTML('<iframe width="560" height="315" src="https://www.youtube.com/embed/lR5A
```

Out[3]:



Power BI

Power BI

- Microsoft's response to Tableau!

Power BI

- Microsoft's response to Tableau!
- Easi(er) to use

Power BI

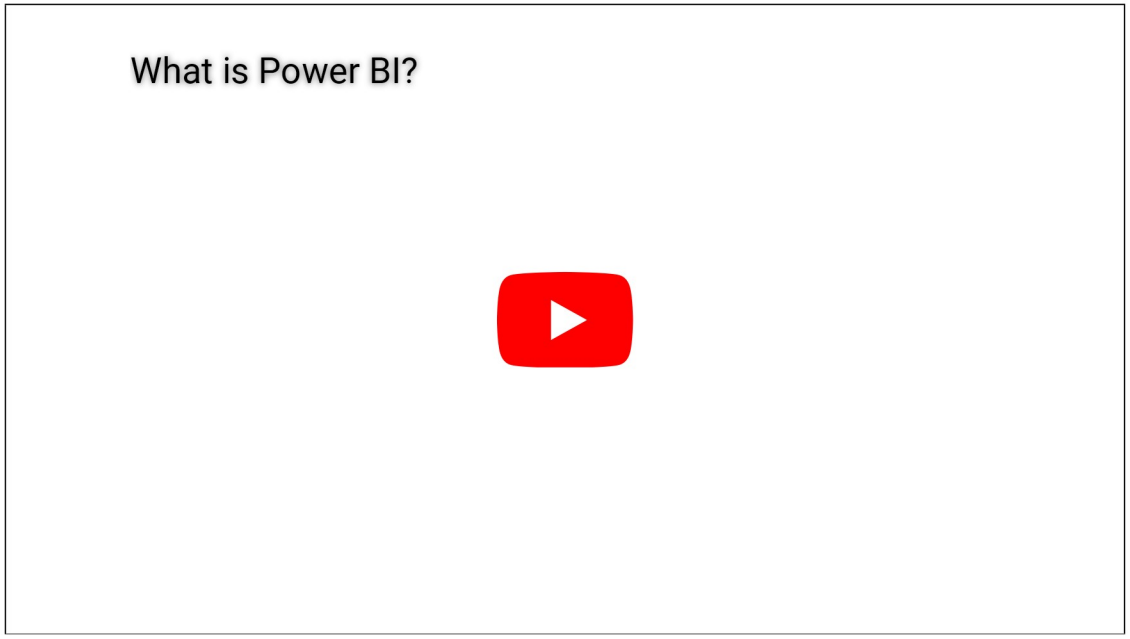
- Microsoft's response to Tableau!
- Easi(er) to use
- Connects to existing projects in `PowerApps`, `Excel` or `Azure`

Power BI

- Microsoft's response to Tableau!
- Easi(er) to use
- Connects to existing projects in `PowerApps`, `Excel` or `Azure`

In [4]:
```
from IPython.display import HTML
HTML('<iframe width="560" height="315" src="https://www.youtube.com/embed/yKTS
```

Out[4]:



How to get into data vis?

1) Get familiar with data repositories

1) Get familiar with data repositories

- Data is literally everywhere!

1) Get familiar with data repositories

- Data is literally everywhere!
- Sites such as Kaggle offer challenges every day!

1) Get familiar with data repositories

- Data is literally everywhere!
- Sites such as Kaggle offer challenges every day!
- If you are familiar with Python/R, you will be aware that they also have data within

2) Watch other people have a go at it!

2) Watch other people have a go at it!

- Sites such as fivethirtyeight.com have lots of examples

2) Watch other people have a go at it!

- Sites such as fivethirtyeight.com have lots of examples
- There are also social media groups
    - Information is beautiful
    - I f*cking love maps
        - FB
        - TW
        - IG

3) Try to think how you will use this in your other modules!

Final Recommendations

Final Recommendations

- If you plan to use Python/R, I **strongly** recommend you installing ANACONDA
    - This way at least you will have two of the discussed tools.

Final Recommendations

- If you plan to use Python/R, I **strongly** recommend you installing ANACONDA
    - This way at least you will have two of the discussed tools.
- As students, you get Tableau and Power BI for free, don't hesitate to give them a try as well!

Lab