Python: Data Analysis and Data Visualization

David Pinezich

Learning Targets

- Import data and pre process
    - formats, parse data, use suited data structures
- Aggregate
    - Perform basic analysis
    - descriptive statistics
    - text analysis
- Visualization types
    - Tables, x-y Plot, Normal distribution, Pie chart, Spider diagram, Word cluster, Histogram, 3D Plot

Schedule Today

- Recap & Questions   ~ 15min

- Intro population.io, JSON & HTTP-Requests   ~ 30min

- Break   ~ 15min

- Exercise Population (individual breaks)   ~ 75min

JSON

- Basics
  - Hierarchical Data
  - Structures
  - XML-Like

```json
{
  "firstName": "John",
  "lastName": "Smith",
  },
  "phoneNumbers": [
    {
      "type": "home",
      "number": "212 555-1234"
    },
    {
      "type": "office",
      "number": "646 555-4567"
    }
  ],
  "children": [],
  "spouse": null
}
```

HTTP Requests

- If a simple website is called, the Browser fires a HTTP Request. The HTTP response contains HTML and CSS. The Browser turns this data into the representation you are looking at.

- By HTTP Requests also non-graphical data can be transmitted.
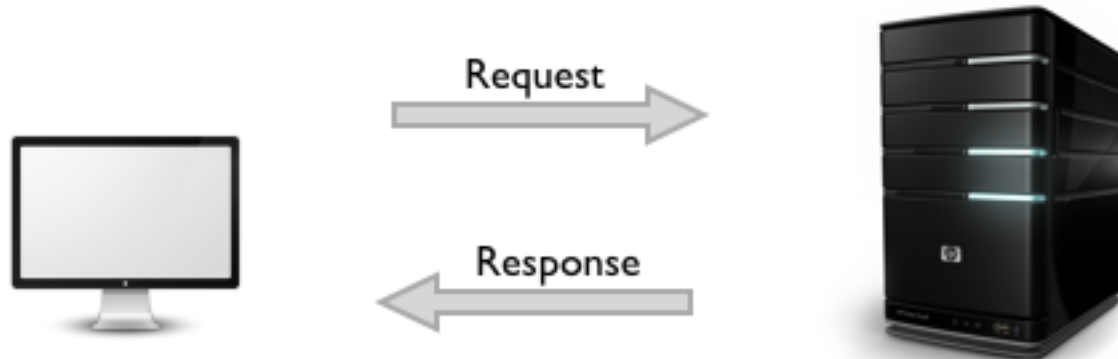
HTTP Requests

Python examples – Showtime!

```python
import requests

r = requests.get('https://api.github.com/events')
print(r.text)
```

```python
import requests

countries = requests.get("http://api.population.io:80/1.0/countries")
countries_list = countries.json()["countries"]
```

Population.io

- http://population.io

- http://api.population.io => *http://d6wn6bmjj722w.population.io*

Exercise 1

- Done together step by step
    - Learn how to structure a program
    - Learn how to use pickle
    - Learn how to extract data from population io

Exercise 2-6

- Read the json strings of the python files

    - 2: Find the top 10 countries with the largest population

    - 3: Find the top 10 countries with the largest relative growth (from today to tomorrow)

    - 4: Find the top 10 countries with the longest life expectancy (Average between woman/men) for people born on "2015-06-30"

    - 5: Extract the correlation (see next slide) between relative growth and the life expectancy

    - 6: Calculate further correlations & have fun with the data

Correlation

- Source: https://www.mathsisfun.com/data/images/correlation-calc1.gif

2 Subtract Mean

3 Calculate ab, a² and b²

Temp °C	Sales	"a"	"b"	a×b	a²	b²
14.2	$215	-4.5	-$187	842	20.3	34,969
16.4	$325	-2.3	-$77	177	5.3	5,929
11.9	$185	-6.8	-$217	1,476	46.2	47,089
15.2	$332	-3.5	-$70	245	12.3	4,900
18.5	$406	-0.2	$4	-1	0.0	16
22.1	$522	3.4	$120	408	11.6	14,400
19.4	$412	0.7	$10	7	0.5	100
25.1	$614	6.4	$212	1,357	41.0	44,944
23.4	$544	4.7	$142	667	22.1	20,164
18.1	$421	-0.6	$19	-11	0.4	361
22.6	$445	3.9	$43	168	15.2	1,849
17.2	$408	-1.5	$6	-9	2.3	36
18.7	**$402**			**5,325**	**177.0**	**174,757**

1 Calculate Means

4 Sum Up

- 1 is a perfect positive correlation
- 0 is no correlation (the values don't seem linked at all)
- -1 is a perfect negative correlation

5 $\dfrac{5,325}{\sqrt{177.0 \times 174,757}}$ = 0.9575