Session 04: APIs

wifi: GA-Guest, yellowpencil

GENERAL ASSEMBLY

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

Today's session plan

1800-1820	Standup & Review
1820-1845	What's an API?
1845-1900	Exploring APIs with your browser
1900-1920	Break
1920-2000	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 ...

Use the requests library to make API requests

Understand the structure of JSON data

Read API documentation to construct complex requests with many parameters

Convert nested JSON into a pandas dataframe

Data Science Part Time

Review

Volfefe Index

Computers Out:
Git in practise

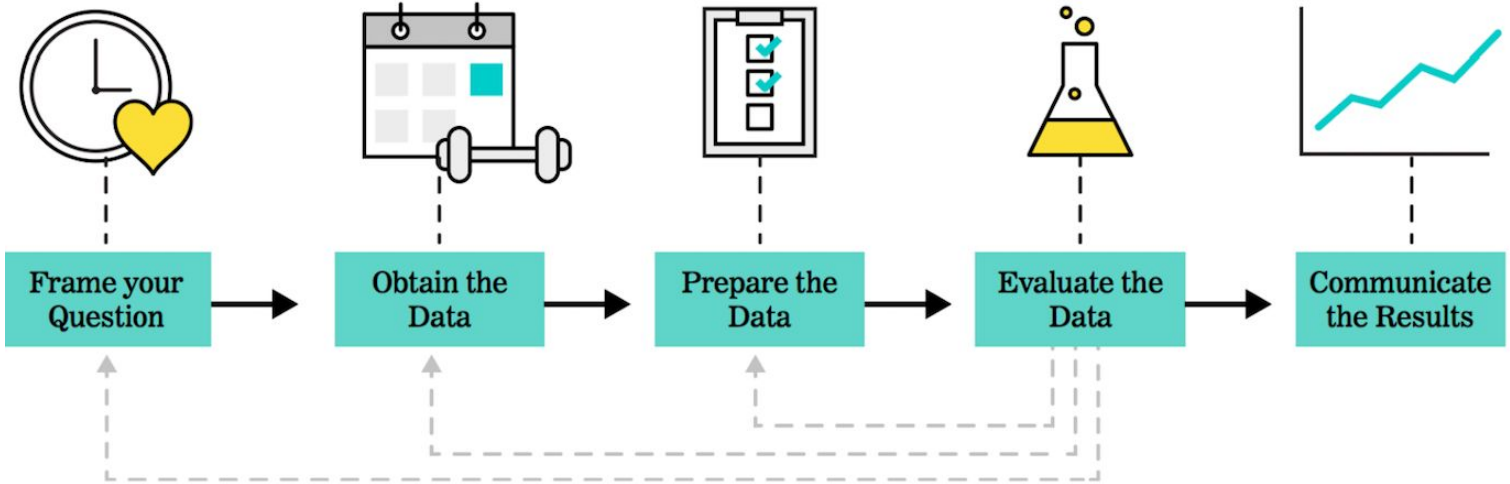
Perform the following tasks using bash and git commands only*:

1. Create a new directory inside your Documents directory called ‘ebooks’ and change to that directory
2. Download the file at this URL as ebook.txt: http://www.gutenberg.org/cache/epub/1497/pg1497.txt
3. Use `head` to inspect the file and figure out what’s in it
4. Use `wc` to figure out how many lines and words the file has
5. Initialise your directory as a git repository
6. Add ‘ebook.txt’ to the staging area and make a commit, with a sensible commit message
7. Use `open ebook.txt` to open up the file in your default text editor
8. Give your laptop to the person sitting next to you. Ask your partner to **secretly** change one word of their choice somewhere in the file, and replace it with another or phrase of their choice. Then ask your partner to save and close the file.
9. Make another commit
10. Use `git log` and `git diff` to find out exactly what your partner changed in your file

Data Science Part Time

What's an API?

Where are we in the data science workflow?



What's an API?

An API is a safe, legal way of accessing large amounts of structured data from the web.

This data could be owned by private companies, governments, or public organisations.

It could be free to access or available on a paid-for basis.

APIs are a widely used data source in data science, and often give us cleaner and better structured data than spreadsheets or other kinds of databases.

Group Exercise:

Understanding the Web

5 minutes

Let's think about the steps needed to access a website in everyday life. When you type **http://www.bbc.co.uk** into your browser's address bar and press enter:

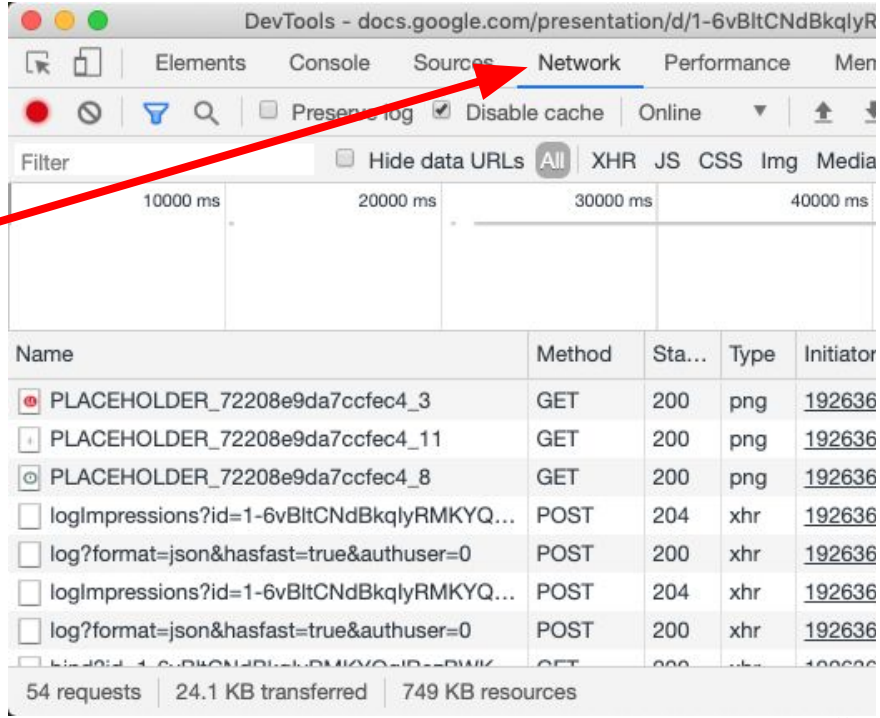
1. The browser sends a request for information to _____'s servers. This is called a 'GET' request.
2. A server is _____.
3. The server responds to the browser's request by sending back a combination of _____, _____ and _____ files.
4. The browser then translates these files and displays the full website to the user.

Computers Out:

Understanding the Web

Let's observe the GET request that your browser makes when it visits a website!

1. Open Chrome.
2. Press Ctrl-Shift-I (on Windows) to open the Developer Tools.
3. Click on the Network tab.
   (The window to the right will be empty.)
4. Refresh your browser window.
   (The window will look similar to the right.)
5. Your instructor will walk you through how to observe the requests



Understanding the web

When you're accessing a website, you are:

Sending a request to a server and getting back HTML, CSS and JavaScript files

When you're using an API, you are:

Sending a request to a server and getting back structured data

This is called **calling an API** or making an **API request**.

Computers Out:

Calling an API: Astronauts

5 minutes

Many APIs are easy to call. Just visit the URL in your browser!

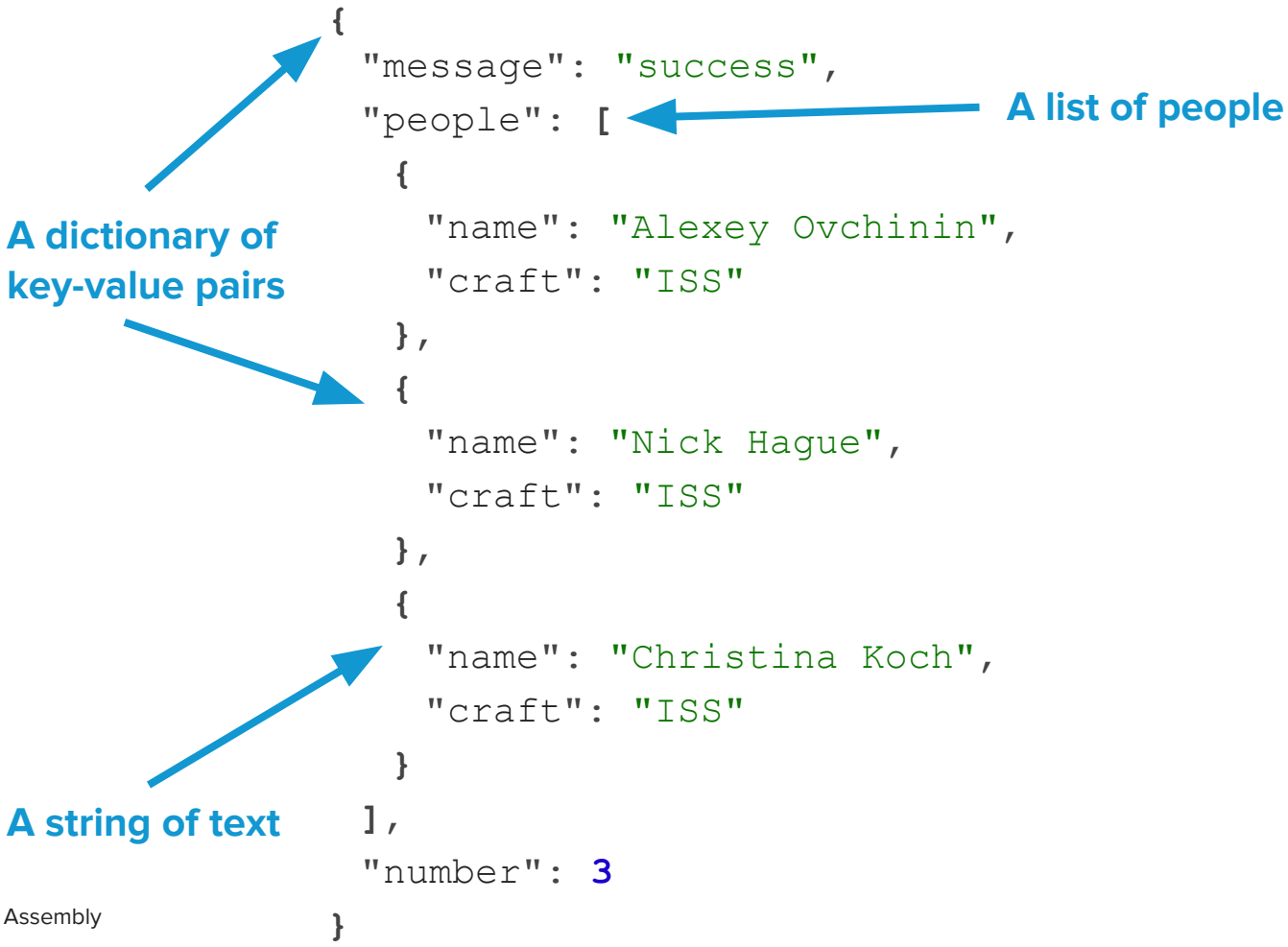
1. Let's view which astronauts are aboard the ISS (International Space Station).

2. In your browser, visit:

   `http://api.open-notify.org/astros.json`

Here is a hint this may return structured data in the JSON text format!

JSON



JSON

This data is returned in JSON (JavaScript Object Notation) format.

It's structured data, and it is fairly readable.

In fact, there is a very strict format that JSON must follow- which looks very similar to the dictionaries we're already familiar with!

In Python, JSON is treated in almost exactly the same way as a dictionary.

Computers Out:

Calling an API with Python

It’s much faster to grab data from APIs using Python.

We do this using a library called **requests.**

Let’s try this out in our Jupyter notebook.

Open ds37-04-01.ipynb to get started!

Group Exercise:

Which organisations have APIs?

Research the following APIs, and find out what sorts of data or services they provide. Are they free to use? Why would the organisation be motivated to make their data available through APIs?

1. Spotify
2. Transport for London
3. Google Maps
4. Google Cloud Vision **(try this one out in your web browser)**
5. Alpha Vantage

Is there an API for that?

There’s no comprehensive directory of every API in the world.

The process of checking if there’s an API to meet your needs will involve a bit of clever Googling.

API keys

Some APIs will require you to sign up for an API **key**, which is a bit like a password.

This helps keep track of how many requests a single person is making, and prevents malicious use of the API.

Complex API requests

We can request very specific datasets from an API, by specifying **parameters** (you can think of these as **options** or **filters**).

Every API will have **documentation** which shows us how to construct an API request (i.e. a URL) that will retrieve exactly the data we want.

Computers Out:

Calling an API: Share prices

Now let's call an API that has parameters, including an api key, shown here as the **`apikey`** parameter.

1. Let's view Microsoft's share prices over the past 24hrs
2. Use the Alpha Vantage documentation to understand what 'TIME_SERIES_INTRADAY', 'MSFT', and '5min' means
3. In your browser, visit:

https://www.alphavantage.co/query?function=TIME_SERIES_INTRADAY&symbol=MSFT&interval=5min&apikey=demo

GET parameters
function=TIME_SERIES_INTRADAY
symbol=MSFT...

4. Use Jupyter Notebook to perform this same request programmatically

Note: This is just a sample API call. You would have to sign up with them to access other companies' share prices

Computers Out:

Calling an API: Police

Now let's try another API.

1. In your browser, visit: https://data.police.uk/docs/

2. What information do we think this API call will return? Think about the different parts of the request
   https://data.police.uk/api/crimes-street/all-crime?lat=52.629729&lng=-1.131592&date=2017-01

3. Use the API documentation to figure out how to construct an API request (i.e. a URL) that will give you data about stop and searches in London in May 2019 and view the results in your browser.
   *hint: use Google maps to get latitude/longitudes

4. Inspect the result in your browser. What information are you getting back?

5. Perform the same request in Python.

6. Use a for loop to build one big JSON object that contains stop and search data for London for the whole of 2019

Intro to Python

Let's Review

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

Use the requests library to make API requests

Understand the structure of JSON data

Read API documentation to construct complex requests with many parameters

Convert nested JSON into a pandas dataframe

Feedback

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

http://bit.ly/ds37weekly

Coming up next week...

- Exploratory data analysis with Pandas
- Data visualisation with matplotlib and seaborn

Homework

Finalise the topic of your final project, and find data sources (bearing in mind the good data checklist from session 1) that can help you investigate your question.

Some suggested sources:

https://archive.ics.uci.edu/ml/index.php
https://www.kaggle.com/
https://data.gov.uk/
https://www.ons.gov.uk/


GA