Social Data Aquisition with APIs

Layla Bouzoubaa

4/14/23

Hello 👋

👩‍🎓 Doctoral student, IS

💊 NLP, social computing, drugs

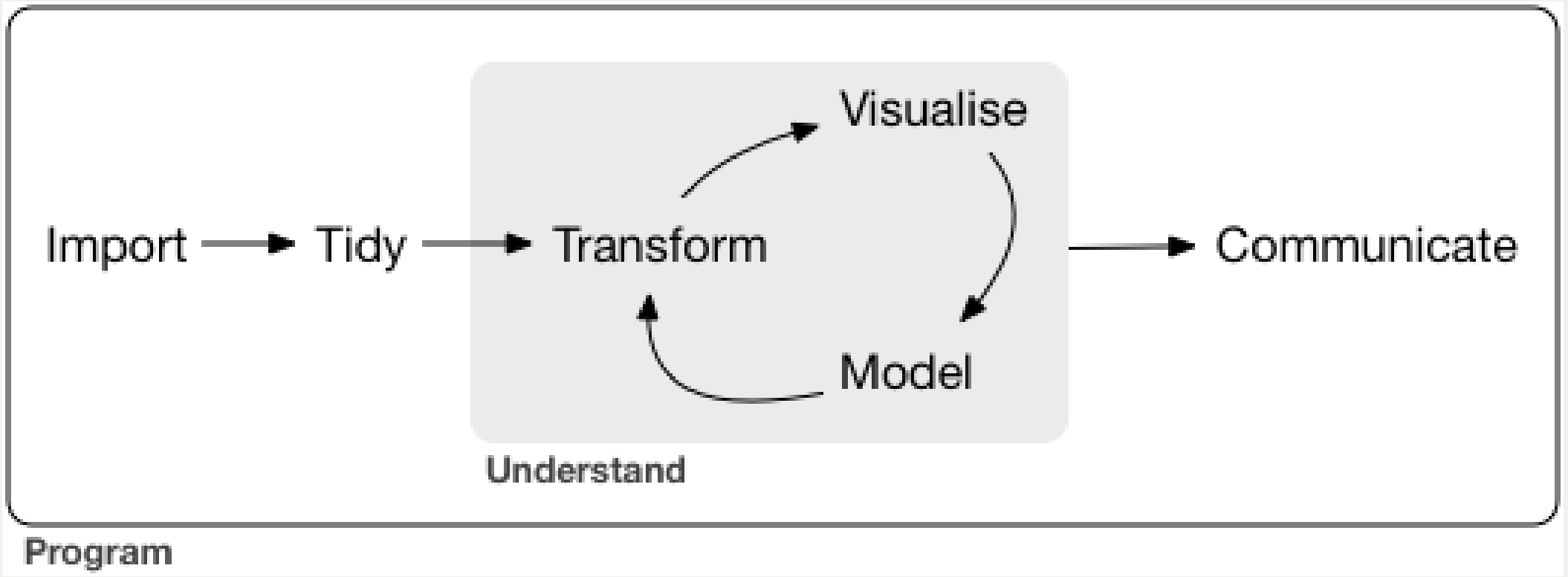
⏱️ Passionate about 🐱 👫 🍷 🍣 🦪 🍵

Outline

- Introduce API
- Setup Reddit API
- Reddit API with examples
- Play time

DATA
LOTS AND LOTS OF DATA
makeameme.org

Typical Data Science Pipeline



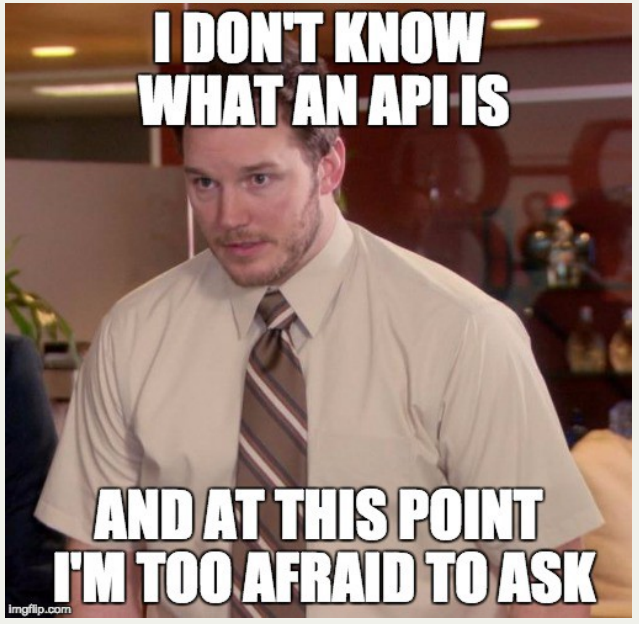
Reddit

- The “front page of the internet”
- Reddit has an API that allows developers to access and extract data from the platform.

What is an API?

Application **P**rogram **I**nterface

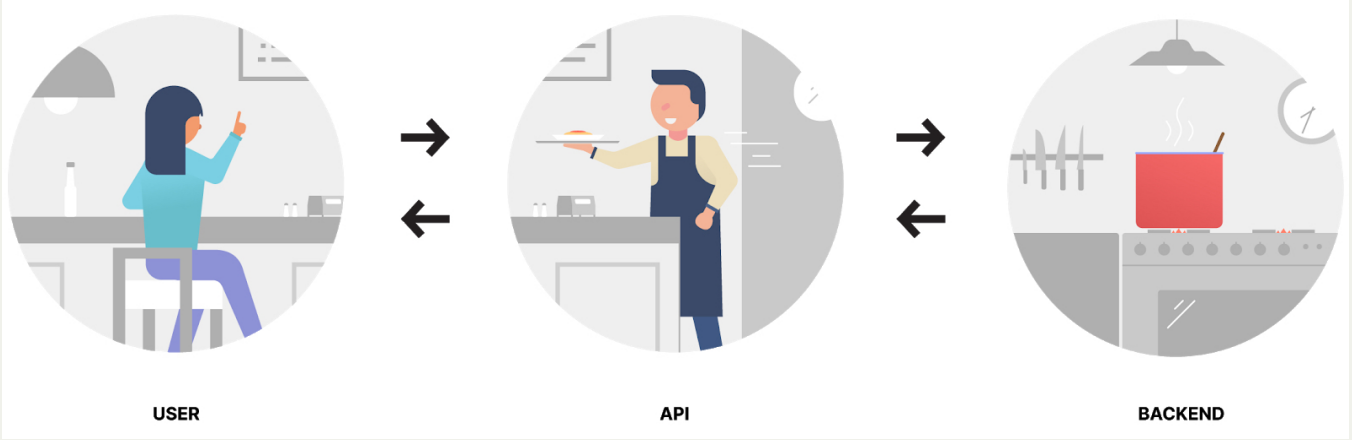
- Allows a user to programatically pull data from a source given that source provides one
  - *ex. The NYT, Twitter, Facebook, Google, US Census*



How Does it Work?

Not Only for Getting Data

- Build APIs for your ML models so others can use them
- Write (post) to a database
  - and update (put) and delete (delete)
- Functions
  - more control over what is exposed to end user



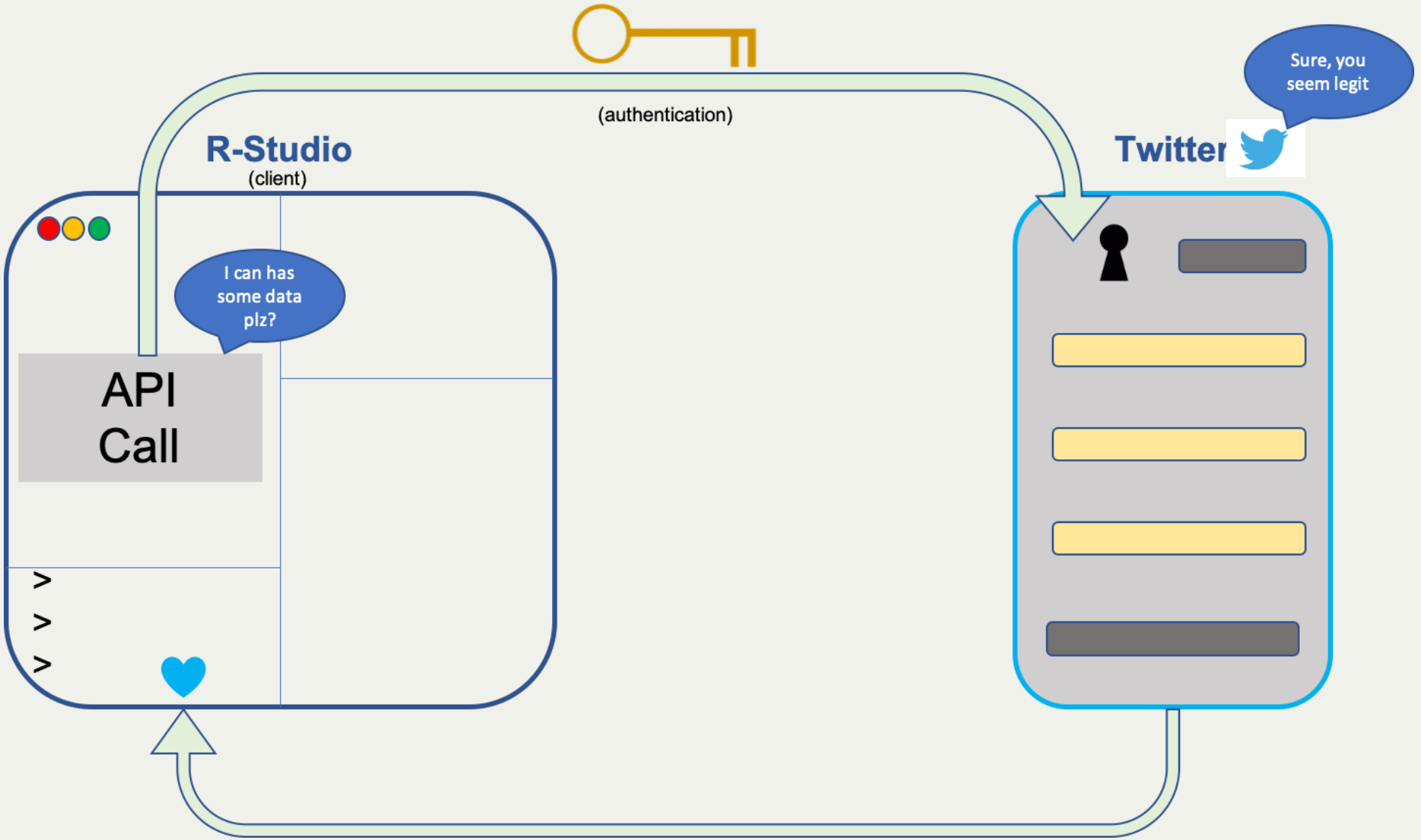
The API

- **G**ET: used to retrieve data from the server
- **P**OST: send data to the server to create a new resource
- **P**UT: update an existing resource on the server
- **D**ELETE: delete an existing resource on the server

There is this App…

It’s that easy?

Keeper of the Keys



Understanding Reddit API

- Reddit API is a RESTful API that uses HTTP requests to communicate with the server.
- API endpoints are used to perform specific tasks such as getting information about a subreddit, submitting posts or comments, and searching for content.
- To access the Reddit API, we need to authenticate with our API key.

The Problem

Which Friends characters are discussed more?

The One with the Steps

Source: reddit.com/r/friends_tv_show

1. Obtain Reddit API Key
2. Install required libraries
    1. PRAW
    2. requests
3. Authenticate with Reddit
4. Do data science

```
pip install praw
pip install requests
```

The One with the Notebook

See the notebook here to follow along

Resources

- What is an API?
- Reddit API Documentation