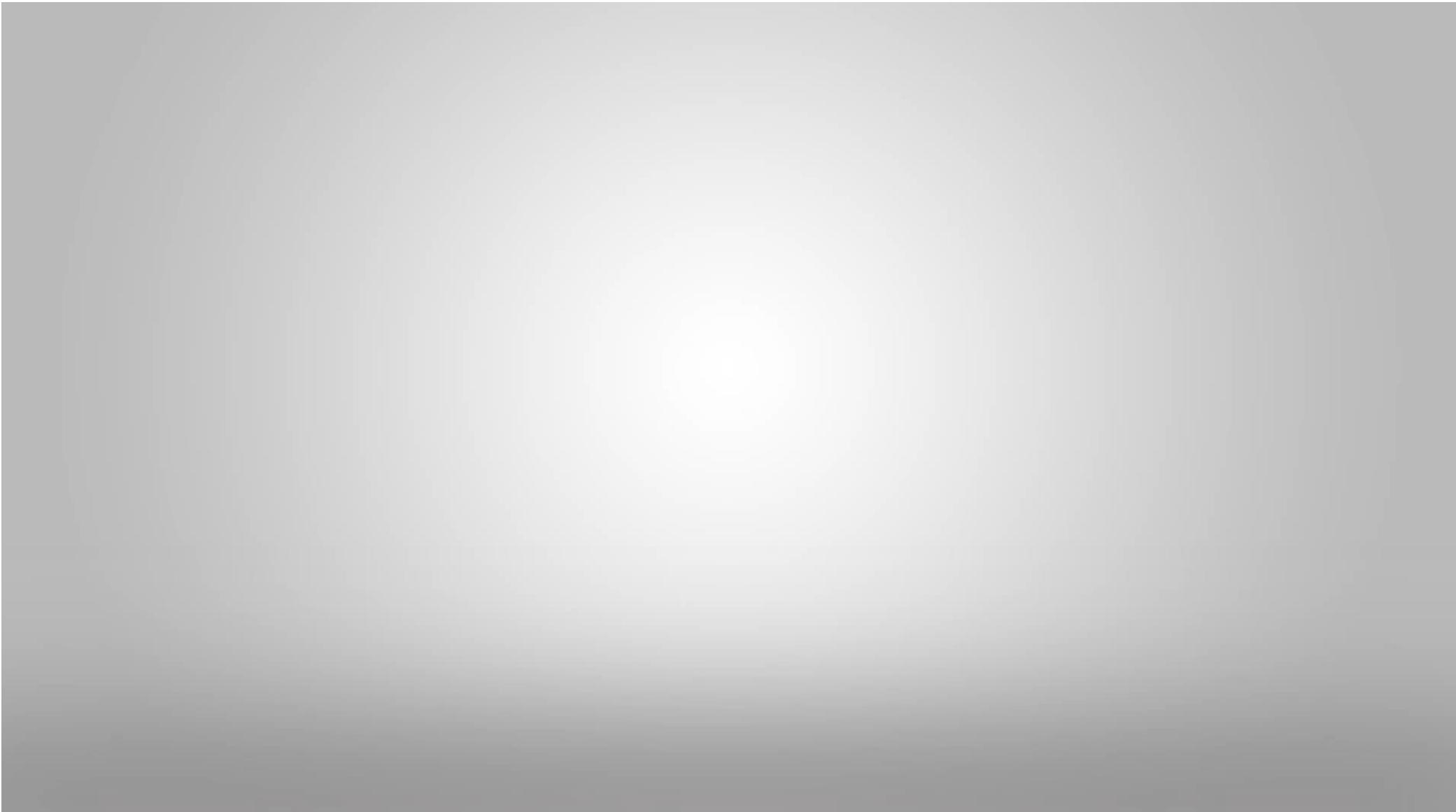


# Intro to Data Collection

APIs



Do no distribute

“Look after  
truth and  
goodness, and  
beauty will look  
after herself”

**The data is the 1<sup>st</sup> priority of  
effective data visualizations**





# Data: facts and statistics collected

Oxford

# Examples of Data Sources

---



## Miranda W.

3 reviews



2 months ago

**Verified customer**

I recently celebrated my birthday here and it was an all-around great experience! The staff treated us very nicely, and they even gave us a complimentary champagne toast. The space was clean and organized, and my guests and I felt very at home. I would definitely recommend this place, and I'll be coming back.

## Reviews

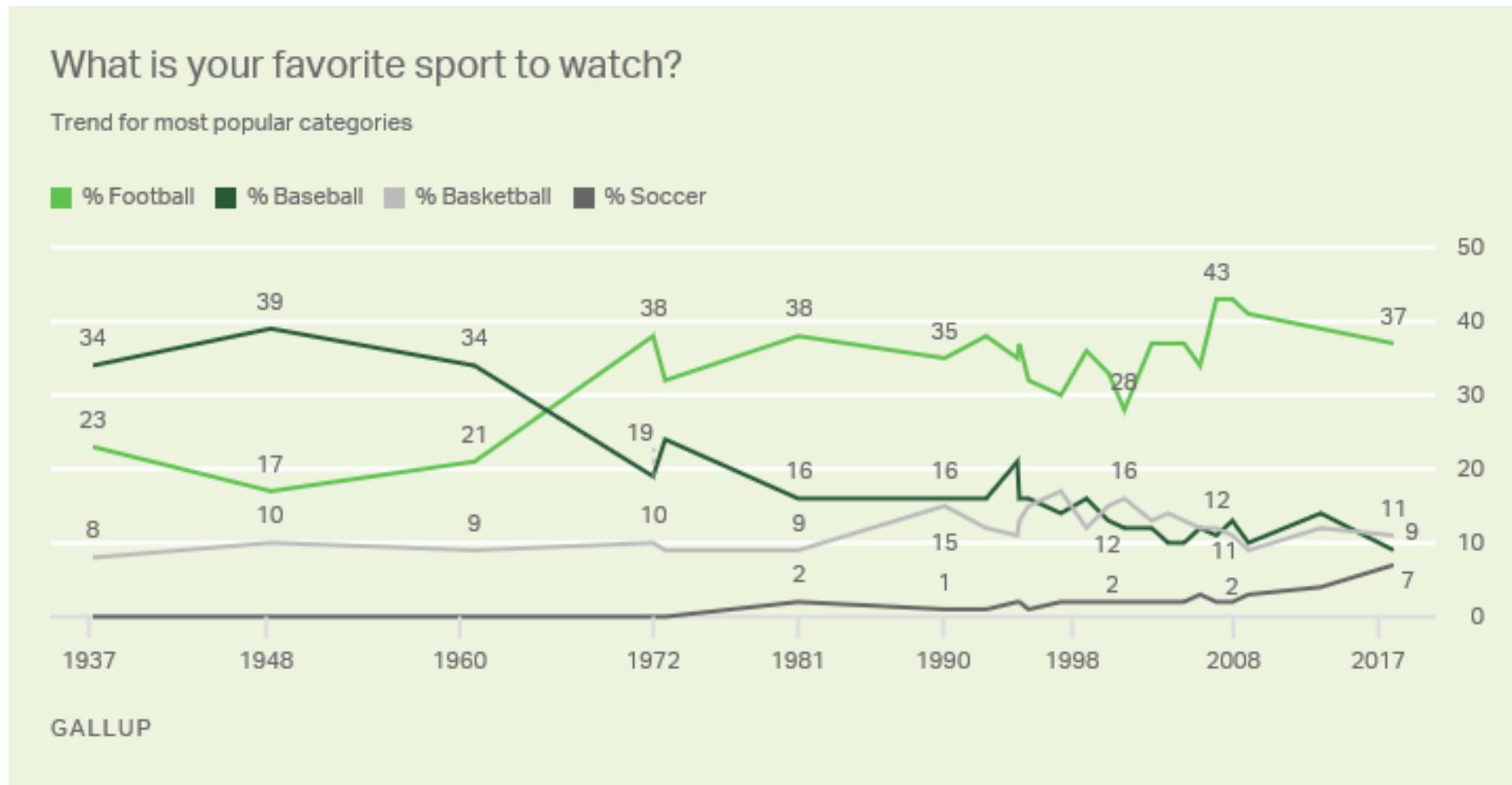


Behavior

Compare SessionCam Price Plans?		
Free Plan	Small/Medium	Enterprise
\$0.00 forever	\$30.00/mth.	Contact Sales Team
500 pages per month	10,000 pages per month \$3 for each extra 1,000 pages	Unlimited pages Full Account Management
90 Days storage Full Dashboard Full Session Recording Full Session replay Full Heatmap Suite Full Conversion Suite	90 Days storage Full Dashboard Full Session Recording Full Session replay Full Heatmap Suite Full Conversion Suite	90+ Days storage Full Dashboard Full Session Recording Full Session replay Full Heatmap Suite Full Conversion Suite
<a href="#">Proceed to Free Plan</a>	<a href="#">Proceed to \$30 Plan</a>	<a href="#">Contact SessionCam</a>

Do no distribute

# Polls



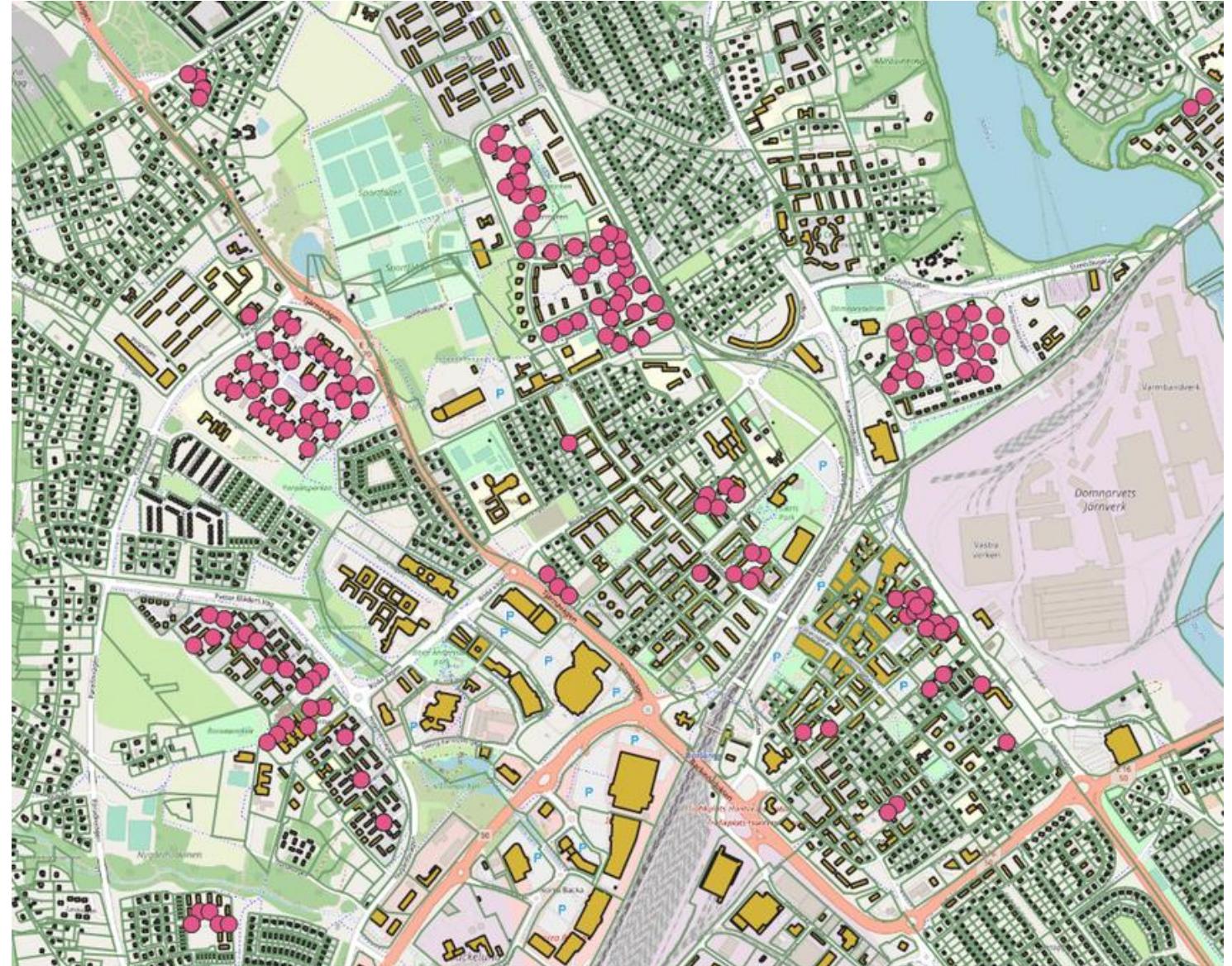
# Measures



© The University of Waikato Te Whare Wānanga o Waikato | [www.sciencelearn.org.nz](http://www.sciencelearn.org.nz)



# Geodata



Do not distribute

# Market Data

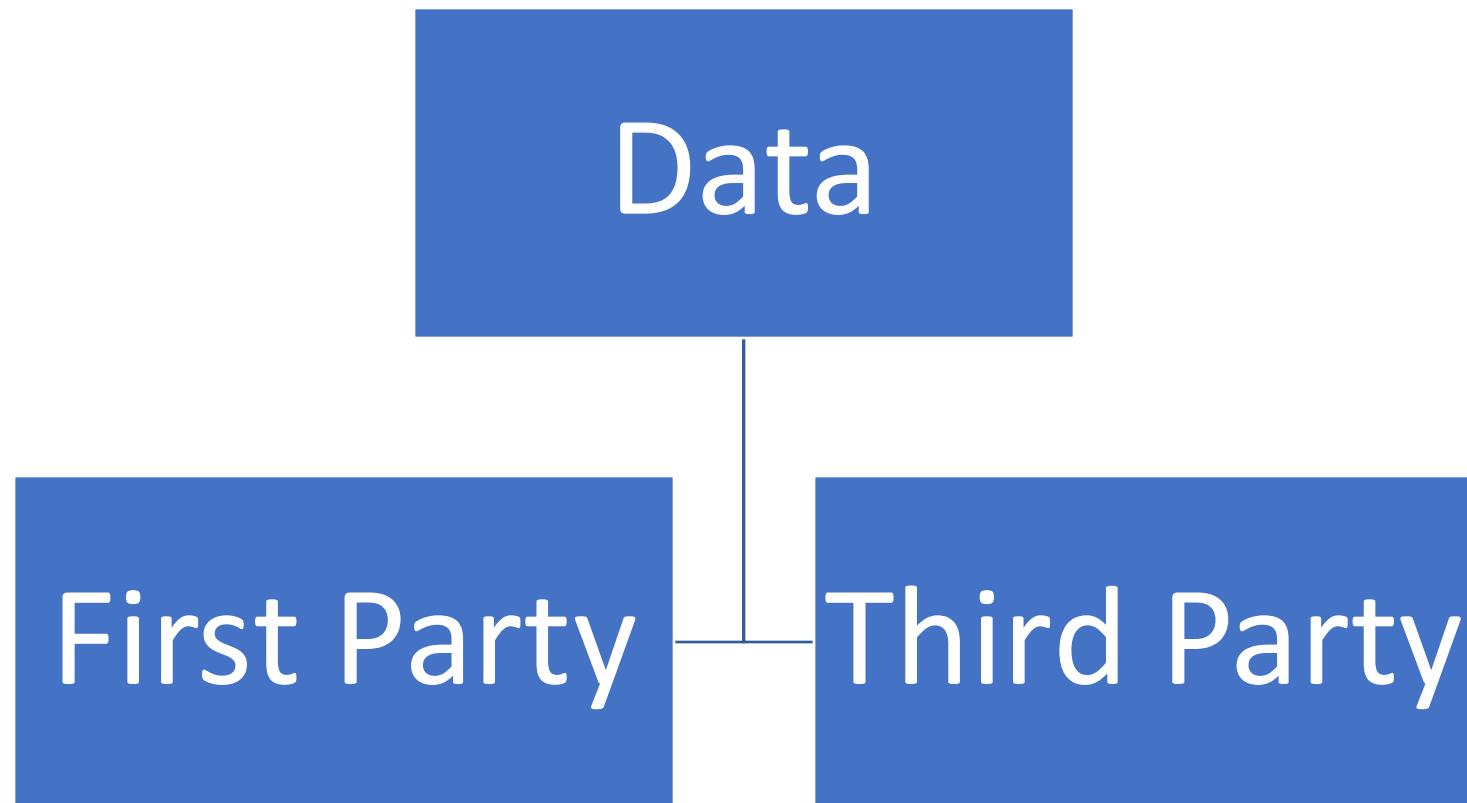
Do not distribute



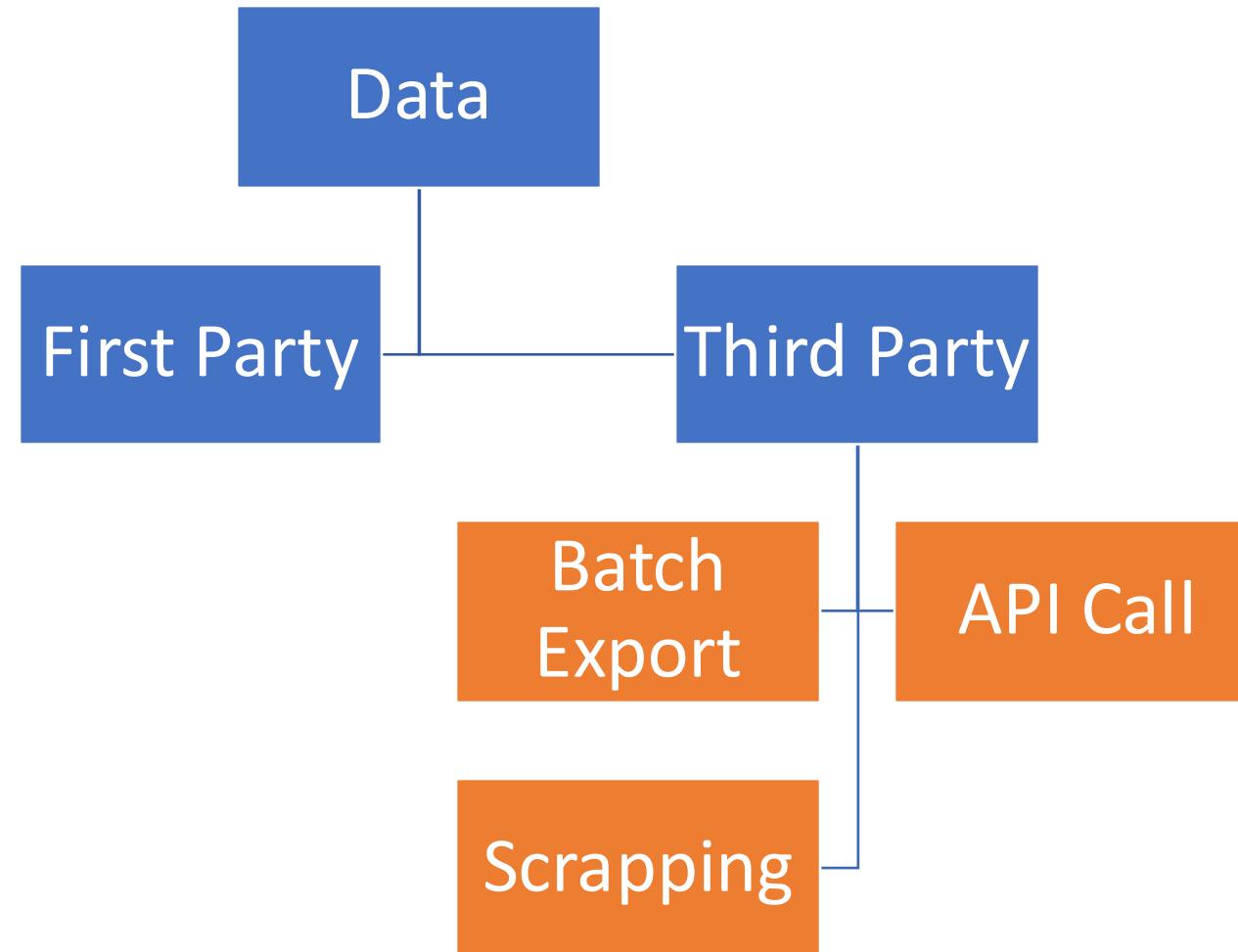
# Transactional Data

Do no distribute

# Primary Ways to Get Data



# Primary Ways to Get Data

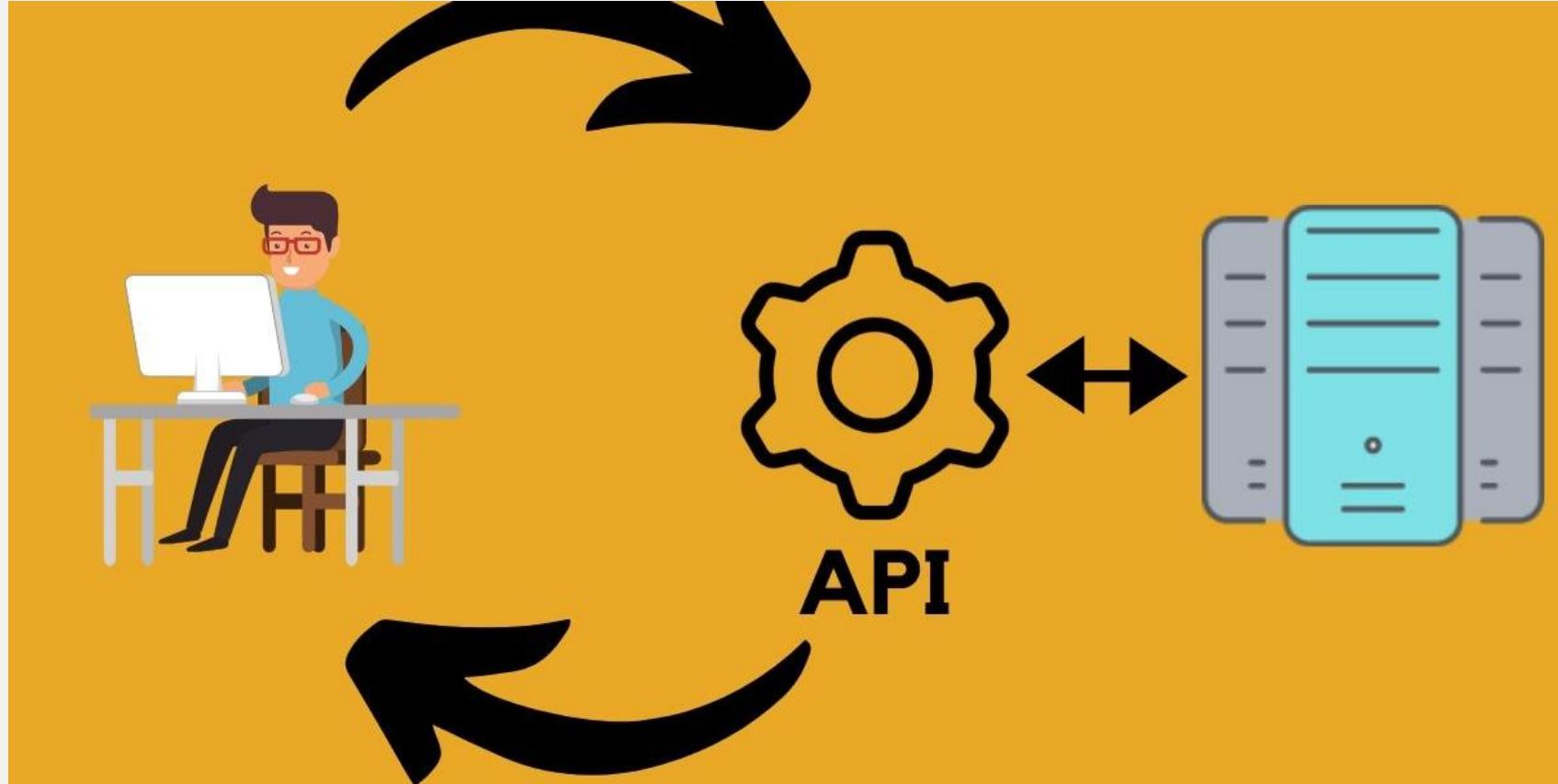


# Batch: Some free services

- <https://www.kaggle.com/datasets>
- <https://datasetsearch.research.google.com/>
- <https://data.gov/>
- <https://datahub.io/collections>
- <https://www.earthdata.nasa.gov/>
- <https://opendata.cern.ch/>
- <https://apps.who.int/gho/data/node.home>

# Web scrapping

Next time



API Calls



# What is an API?

---

- An application programming interface for the web
  - Application programming: software
  - Interface: way to communicate
  - Web: online



# Take 5 Minutes: Google Public APIs

- <https://github.com/public-apis/public-apis>
- <https://rapidapi.com/collection/list-of-free-apis>
- <https://publicapis.dev/>
- <https://free-apis.github.io/#/>
- <https://medium.com/codex/15-fun-and-interesting-apis-to-use-for-your-next-coding-project-in-2022-86a4ff3a2742>

\*\*\*Most popular websites have an API (Facebook, Twitter, Instagram, LinkedIn, etc.)\*\*\*



# Example: Public Holidays

- Call in Web browser
  - <https://date.nager.at/api/v3/publicholidays/2025/US> (you can change the year, or the country. Try DE)
  - <https://api.agify.io/?name=stefan>
- Call in terminal on Mac:
  - curl <https://date.nager.at/api/v3/publicholidays/2025/AT>
  - curl <https://api.agify.io/?name=stefan>
- Call in R (coming soon)

# Wait, what is coming back from the API calls?

- Can be a variety of formats:
  - Most often: JSON
  - Others: XML, plain text



```
[ { "date": "2024-01-01", "localName": "New Year's Day", "name": "New Year's Day", "countryCode": "US", "counties": null }, { "date": "2024-03-29", "localName": "Good Friday", "name": "Good Friday", "countryCode": "US", "counties": [ "US-CT", "US-DE", "US-HI" ] } ]
```

The JSON code is annotated with red labels and brackets:

- A red bracket on the left side groups the entire array as an **Object**.
- A red bracket on the right side groups the two objects within the array as **Key-value pairs**.
- A red bracket below the first object groups its properties as an **Object**.
- A red bracket below the second object groups its properties as an **Object**.
- A red bracket below the "counties" property of the first object groups the three strings as an **Array**.

A complex, abstract digital visualization on the left side of the slide. It features a grid of blue cubes, some of which are highlighted in a bright cyan color. These cubes are interconnected by a network of glowing blue lines and points, forming a three-dimensional lattice structure. The surface of the cubes is covered in a dense pattern of binary digits (0s and 1s). Several small, glowing red and green spheres are scattered throughout the scene, emitting thin blue light trails as they move. The overall effect is one of a high-speed, data-rich digital environment.

# API Protocols

- REST API (representational state transfer): a flexible architecture that allows applications to exchange data in multiple formats. This is the most popular web API architecture
- SOAP API (simple object access protocol): a highly structured architecture that uses XML data format
- RPC (remote procedural call): invoke executable actions or processes remotely

# Get versus Post API Calls

- GET: Pass needed information in the URL:

<https://date.nager.at/api/v2/publicholidays/2024/DE>  
<https://api.agify.io/?name=stefan>

- POST: Pass needed information in the body of the request.

New Request Send Cancel

POST [https://www.handelsregister.de/rp\\_web/search.do](https://www.handelsregister.de/rp_web/search.do)

Request Headers:

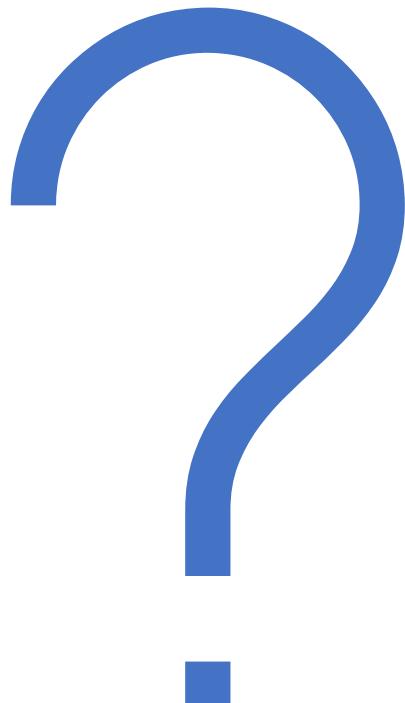
```
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-GB,en;q=0.5
Accept-Encoding: gzip, deflate, br
Referer: https://www.handelsregister.de/rp_web/mask.do?Typ=n
Content-Type: application/x-www-form-urlencoded
Content-Length: 133
Cookie: language=en; JSESSIONID=0D14F2CA4002BE97A282BCCC1AB7080C-n1.tc032n02
Connection: keep-alive
Upgrade-Insecure-Requests: 1
```

Request Body:

```
suchTyp=n&registerArt=&registerNummer=&registergericht=R3101&schlagwoerter=&schlagwortOptionen=2&ergebnisseProSeite=100
```

Do not distribute

# What do you need to make an API call?



- Endpoint: the URL of what you are going to call)
- Key:
  - A String that says you are authorized to make the call
  - Can be used for billing
  - Can be used to restrict the number of calls you make
  - Key might be sent in the URL (get), header, or the body depending on the API

# Let's get a key and make a call

- <https://rapidapi.com/apidojo/api/yahoo-finance1/>
- Create a free account, and subscribe to API (Free Basic Account, 500 calls per month)
- Let's take a few minutes to look around at the options

# Goal: Pull Data Into R to Analyze and Visualize

---



# Studio<sup>®</sup>

# Library Httr2

---

<https://httr2.r-lib.org/>

# In Class Demo

---

<https://www.dropbox.com/scl/fi/b7d4mk2efngz63zc37hlf/InClassAPISample.R?rlkey=9pigsgwhv5w2dpdan4hq1hxyq&dl=0>

# Let's Practice

- Football Stats
  - <https://www.api-football.com/documentation-v3#section/Introduction>
  - A solution for the leagues endpoint:  
<https://www.dropbox.com/scl/fi/8eydqx9vrmrwh1tosbr0d/InClassAPISampleSports.R?rlkey=1wfaykdhqlqvq56uarxfizmnf2&dl=0>
- Currency exchange (a bit harder):
  - Create a tibble that lists the conversion rate between EUR and each supported currency
  - <https://rapidapi.com/fyhao/api/currency-exchange>
  - A solution:  
<https://www.dropbox.com/scl/fi/b0lk8f04tc3hm18whvvh9/InClassAPISampleCountries.R?rlkey=jhhucvkx9zql8szr7difq8s7l&dl=0>

# Look at Assignment

# Potential Final Exam Questions

- Which of the following are examples of data sources?
- What are the two primary ways to get data?
- Third party data typically comes in which of the following formats?
- What is an API call?
- What is the difference between GET and POST API calls?
- List what information you typically need to make an API call?
- What R library can you use to make API calls?
- Some high-level questions about R syntax to make API calls (what certain lines of code do)