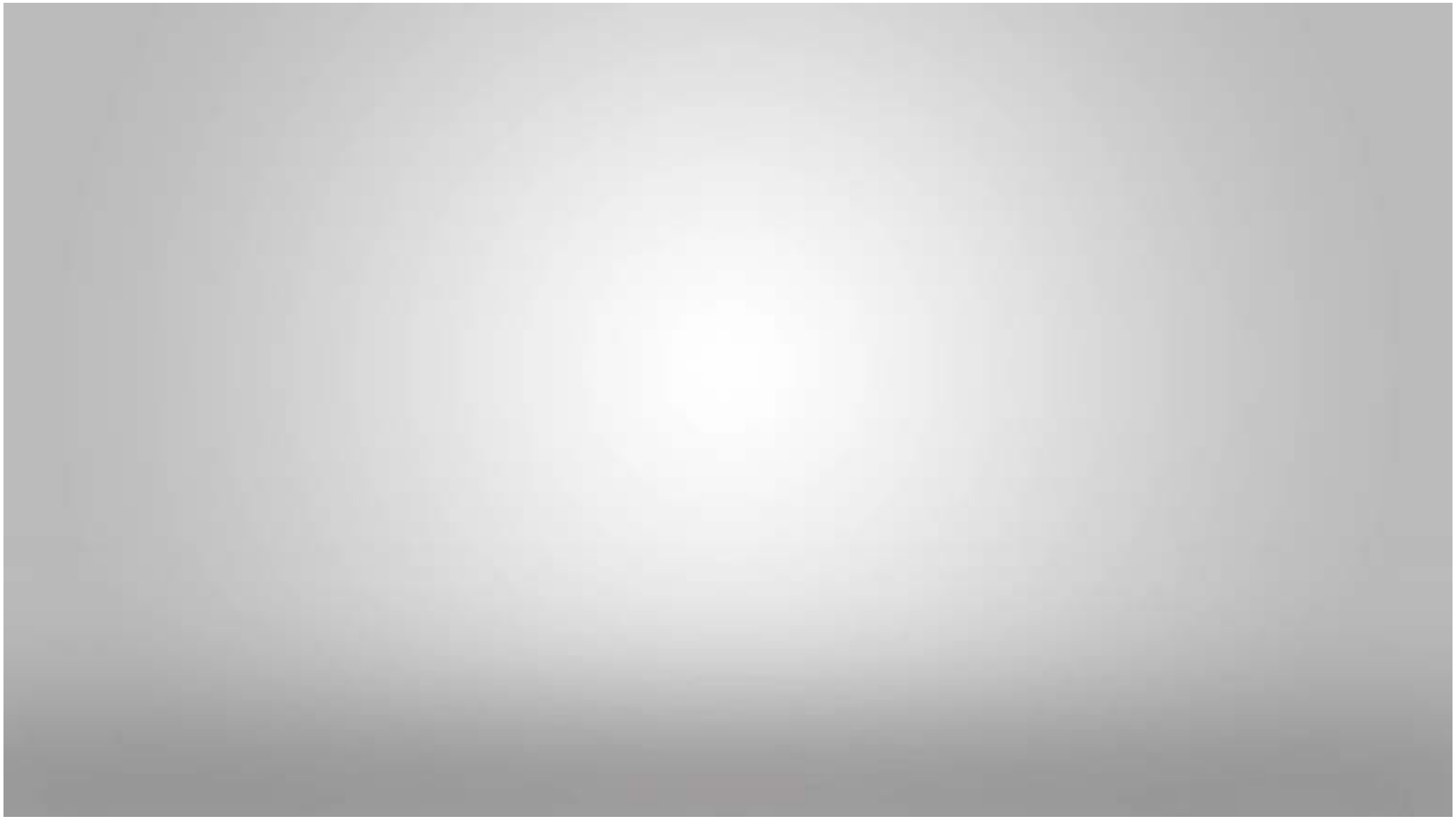
Intro to Data Collection

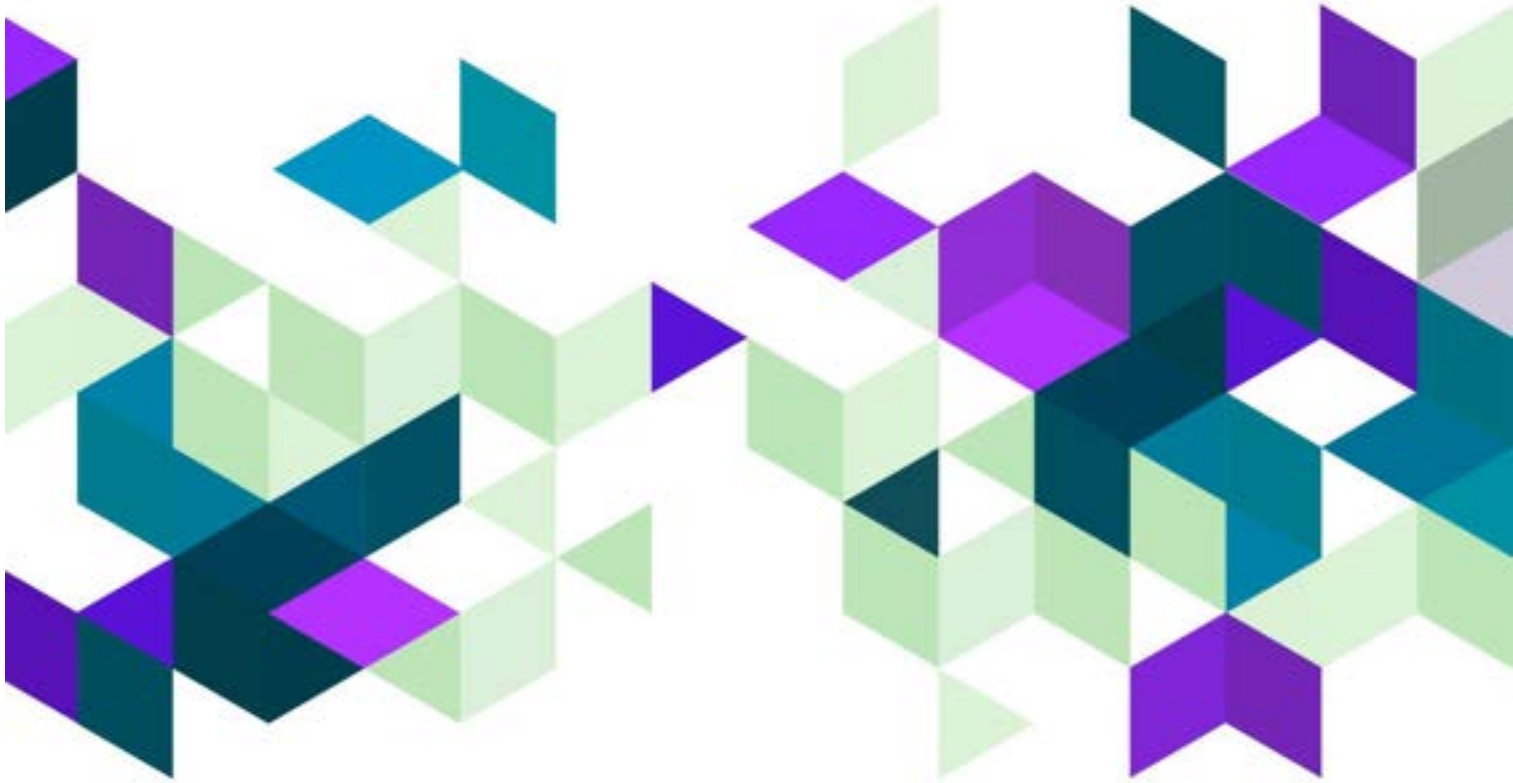
APIs

Do no distribute



Do no distribute

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



The data is the 1st priority of effective data visualizations

Do no distribute

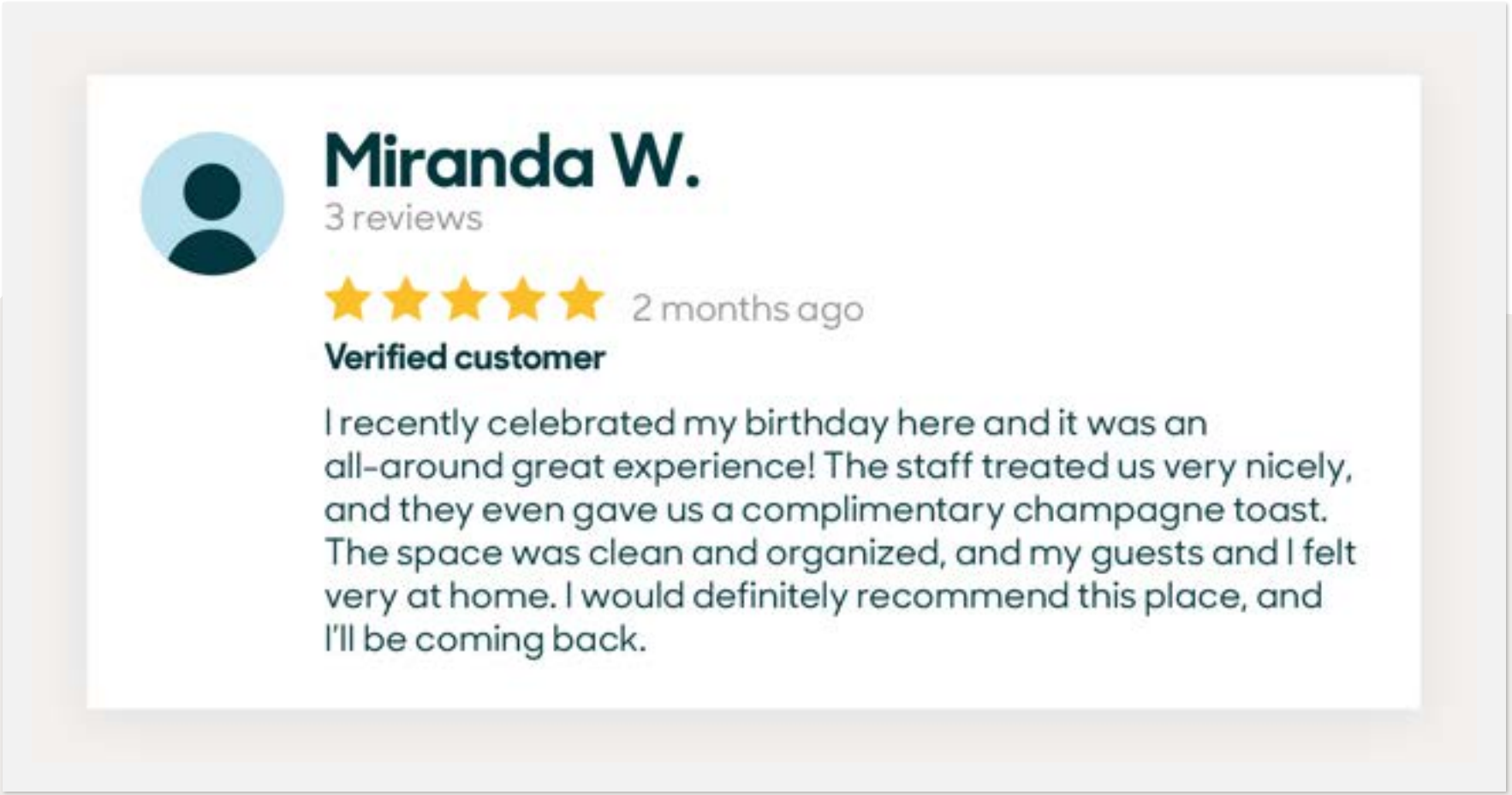
Data: facts and statistics **collected**

Oxford

Do no distribute

Examples of Data Sources

Do no distribute



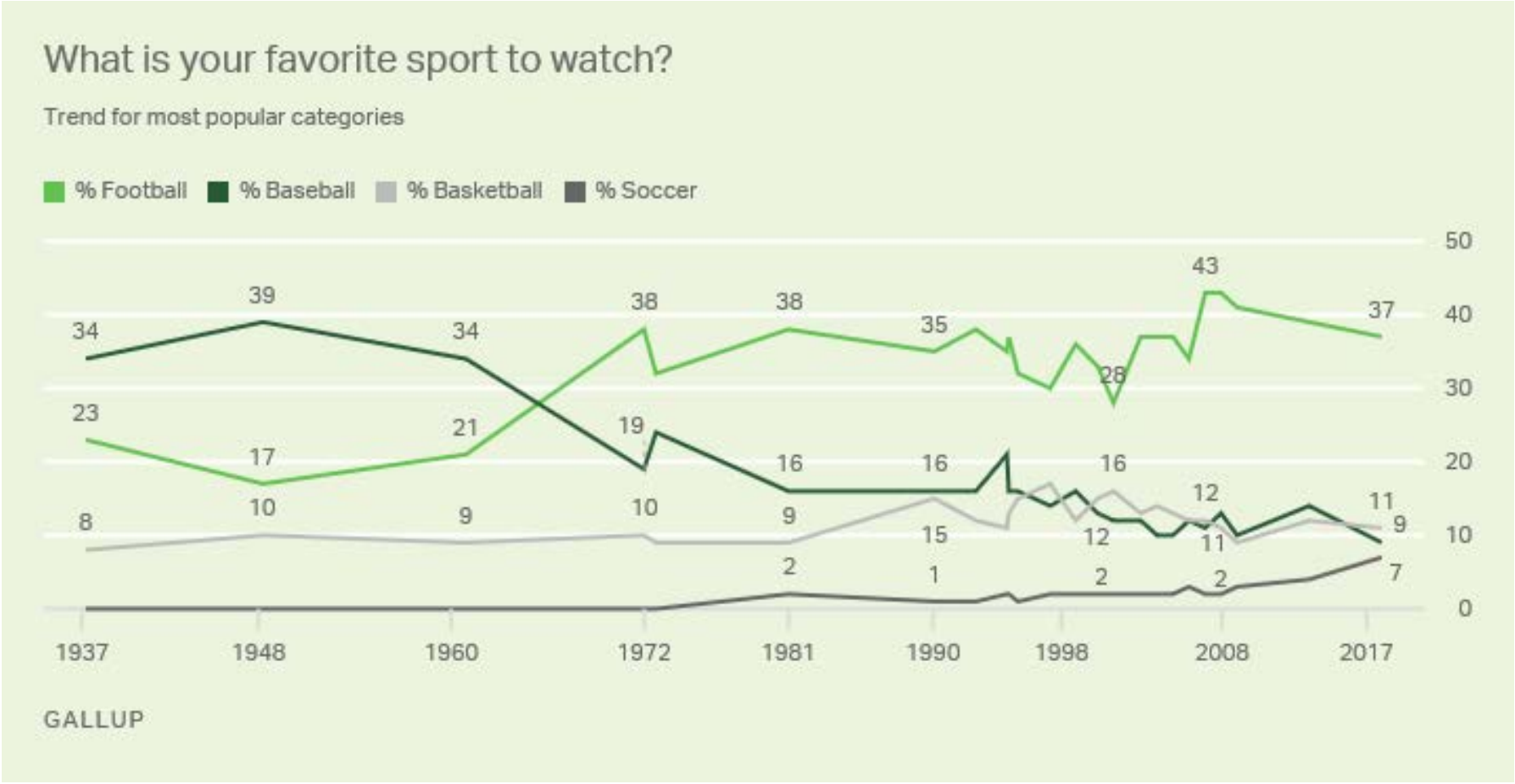
Reviews

Do no distribute

Behavior

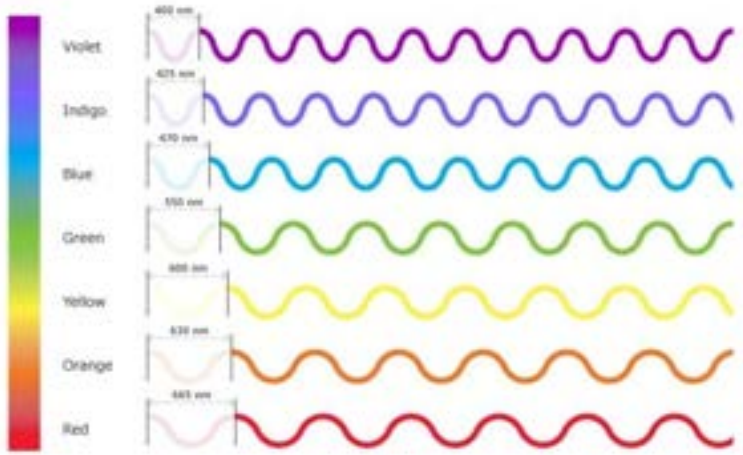
Do no distribute

Polls



Do no distribute

Measures



© The University of Waikato Te Whare Wānanga o Waikato | www.sciencelearn.org.nz

Do no distribute

Geodata

Do no distribute

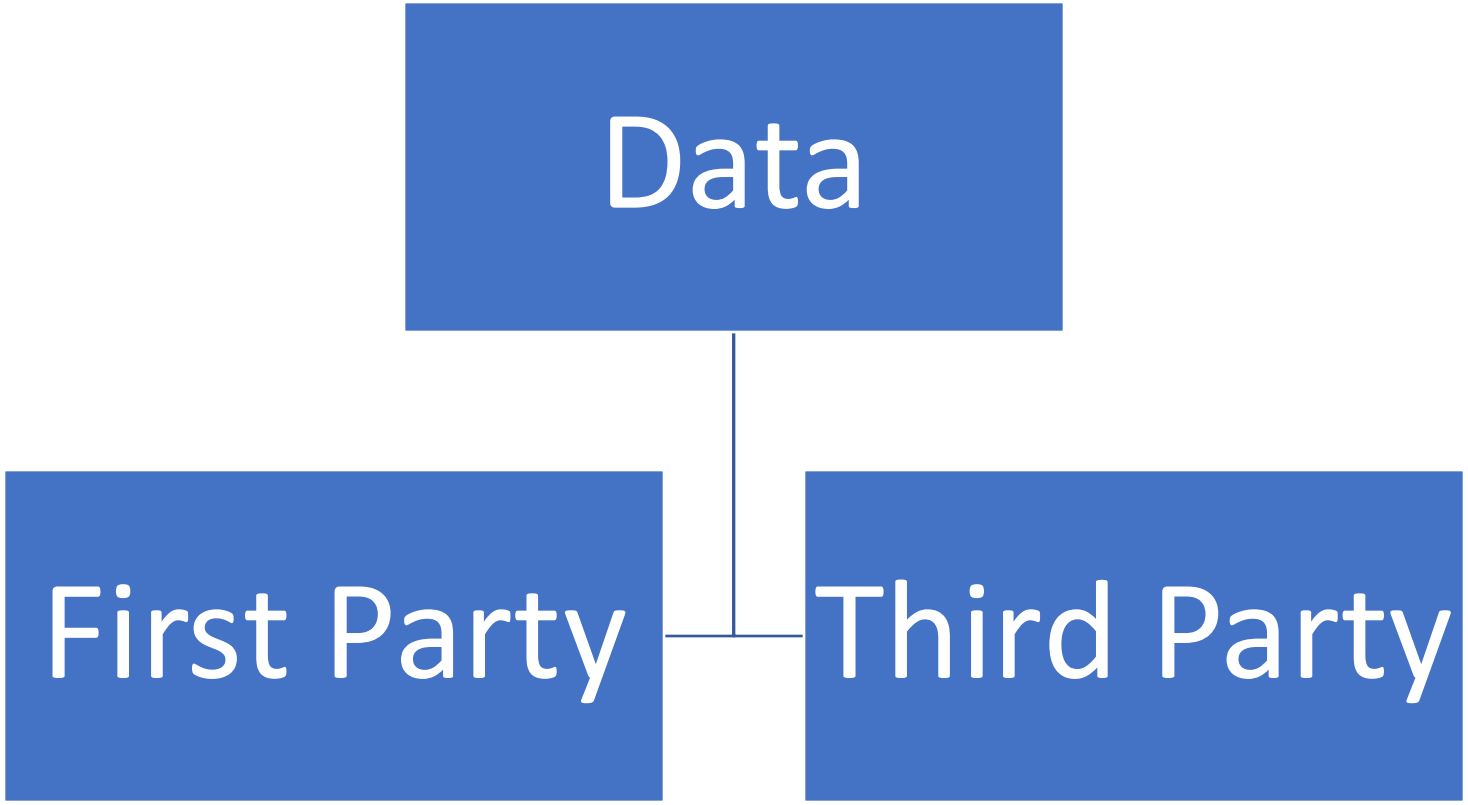
Market Data

Do no distribute

Transactional Data

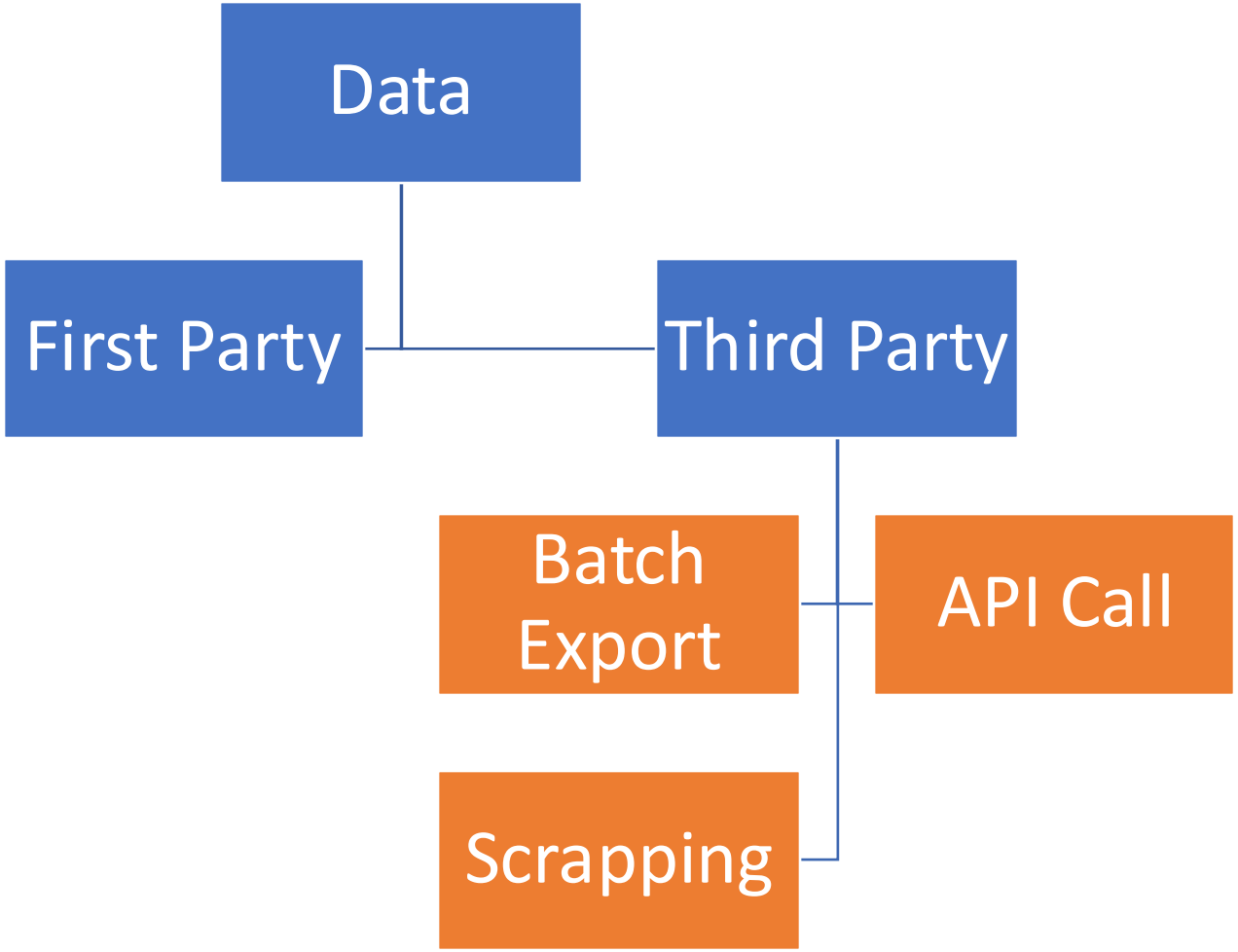
Do no distribute

Primary Ways to Get Data



Do no distribute

Primary Ways to Get Data



Do no distribute

Batch: Some free services

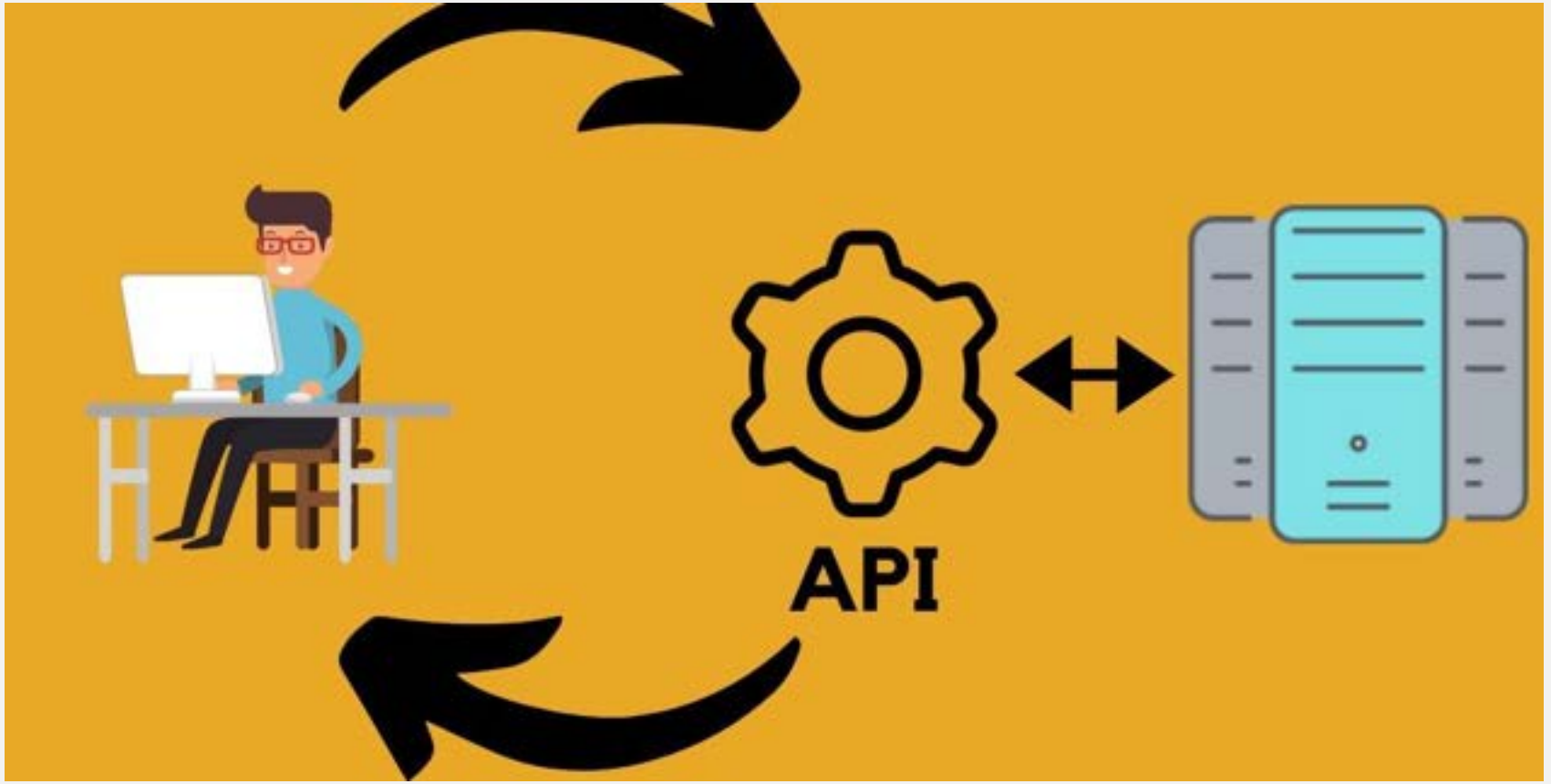
- https://www.kaggle.com/datasets
- https://datasetsearch.research.google.com/
- https://data.gov/
- https://datahub.io/collections
- https://www.earthdata.nasa.gov/
- https://opendata.cern.ch/
- https://apps.who.int/gho/data/node.home

Do no distribute

Web scrapping

Next time

Do no distribute



API Calls

Do no distribute

What is an API?

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

Do no distribute

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.)

Do no distribute

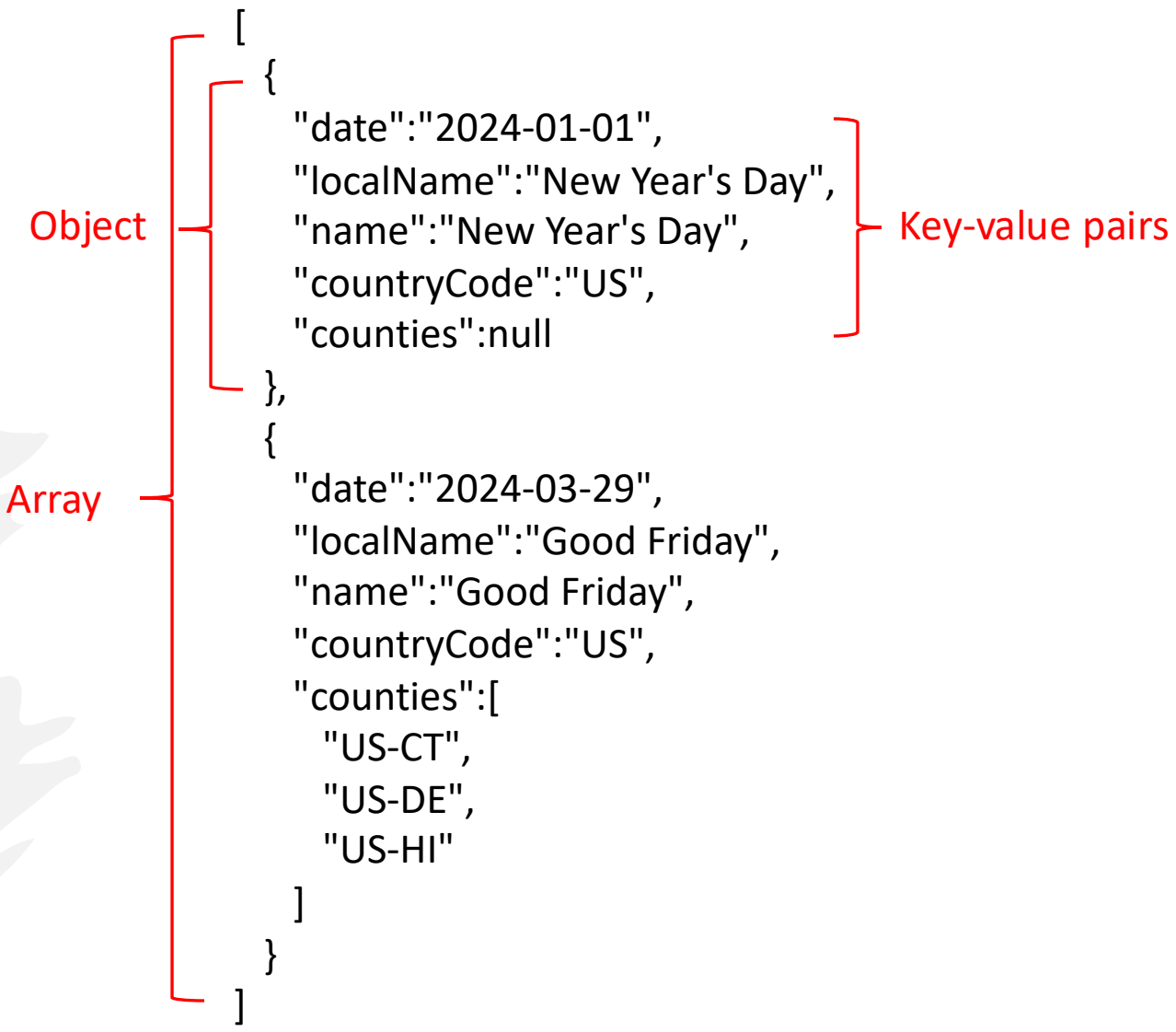
Example: Public Holidays

- 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)

Do no distribute

Wait, what is coming back from the API calls?

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



Do no distribute

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

Do no distribute

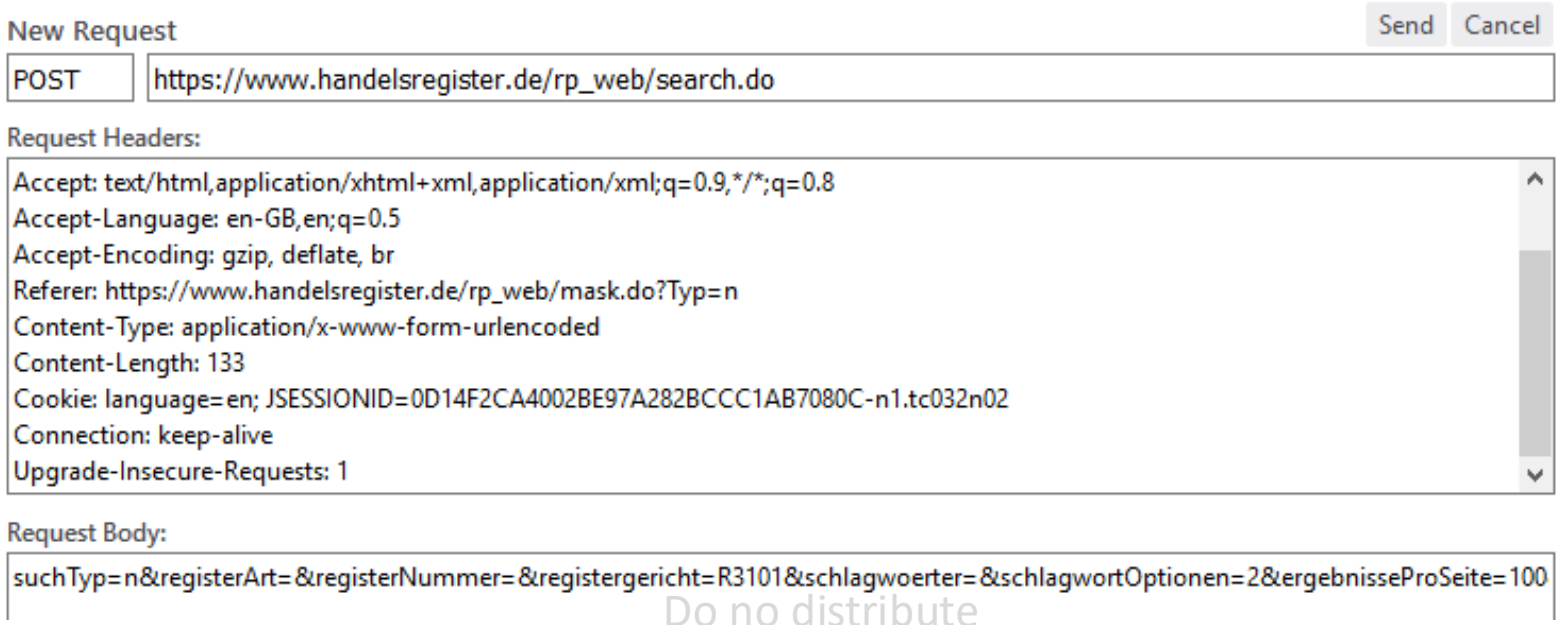
Get versus Post API Calls

- 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.



Do no 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

Do no distribute

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

Do no distribute

Goal: Pull Data Into R to Analyze and Visualize

Do no distribute

Library Httr2

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

Do no distribute

In Class Demo

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

Do no distribute

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/InClassAPISample_Sports.R?rlkey=1wfaykdhlqvq56uarxfizmnf2&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/InClassAPISample_Countries.R?rlkey=jhhucvkx9zql8szr7difq8s7l&dl=0

Do no distribute

Look at Assignment

Do no distribute

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)

Do no distribute