Lecture 2

# DD 324:
# Data Visualisation

## Why ———— visualise?

14 Jan 2025 · Gyan Lakhwani · gyanlakhwani@gmail.com · Department of Design, DTU

DD 324

# Course Objective

Learn to use data as a design material to exhibit, explore, explain, experience and enable.

14 Jan 2025 · Gyan Lakhwani · gyanlakhwani@gmail.com · Department of Design, DTU

# What can we do with data?

## Exhibit

Show raw data

List, Table, Infographic

## Explain

Answer Questions

Data journalism, report

## Explore

Finding what to ask

Dashboards, simulations

## Experience

Finding meaning in data

Data art piece, New media installation

## Enable

Building tools to visualise specific use cases

Software for data viz

14 Jan 2025 · Gyan Lakhwani · gyanlakhwani@gmail.com · Department of Design, DTU

# What can we do with data?

**Exhibit**

Show raw data

List, Table, Infographic

**Explain**

Answer Questions

Data journalism, report

**Explore**

Finding what to ask

Dashboards, simulations

**Experience**

Finding meaning in data

Data art piece, New media installation

**Enable**

Building tools to visualise specific use cases

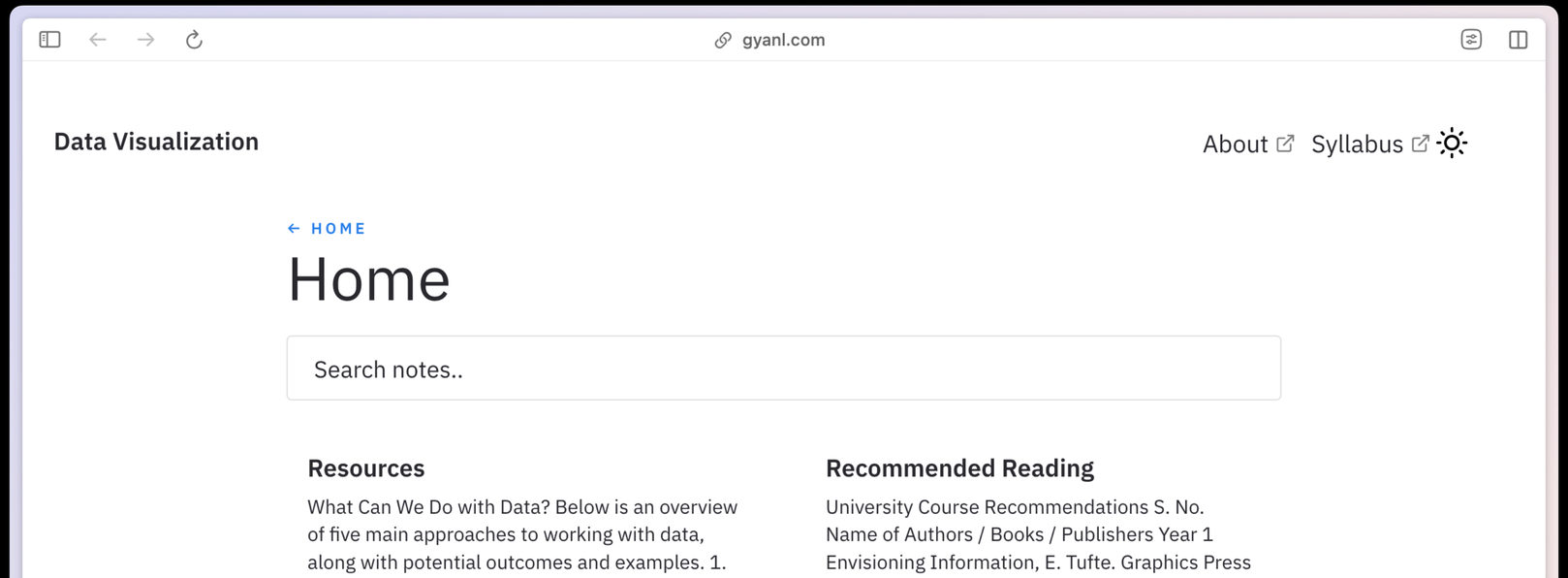
Software for data viz

14 Jan 2025 · Gyan Lakhwani · gyanlakhwani@gmail.com · Department of Design, DTU

## Course Website

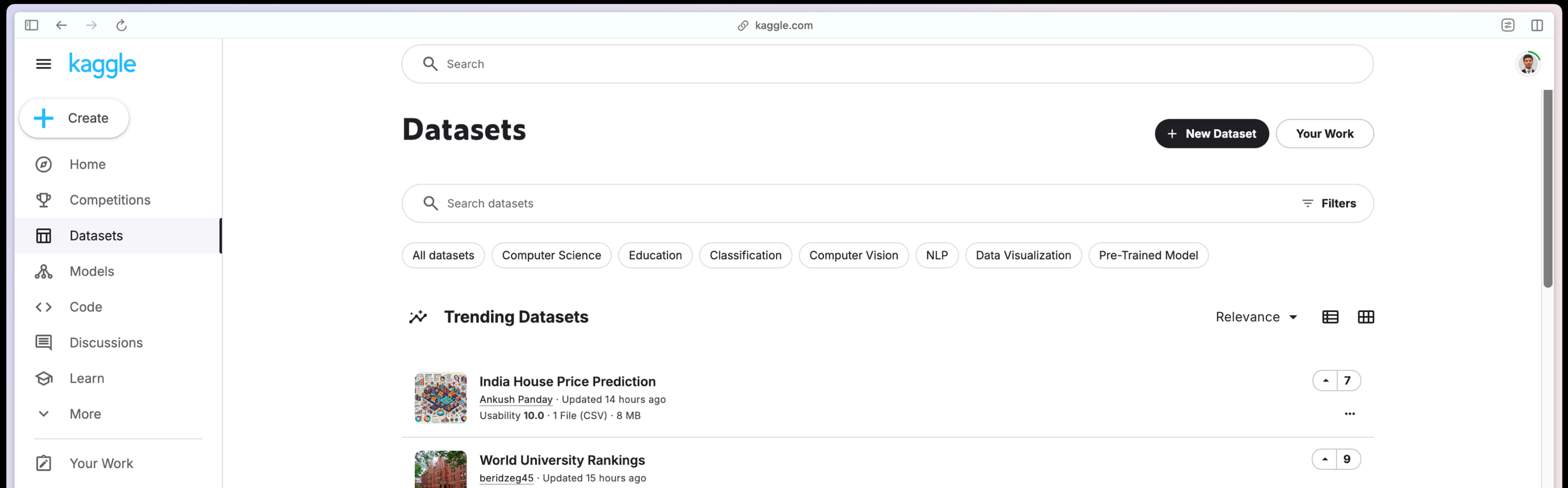
# gyanl.com/dv

(Work in progress)



While we wait...

# kaggle.com/datasets



# Data —— Information

Data —— Information

# Card PINs

3.5 million PINs that were hacked from a bank.

Data —— Information

# Card PINs

3.5 million PINs that were hacked from a bank.

I don’t recommend that you do this.

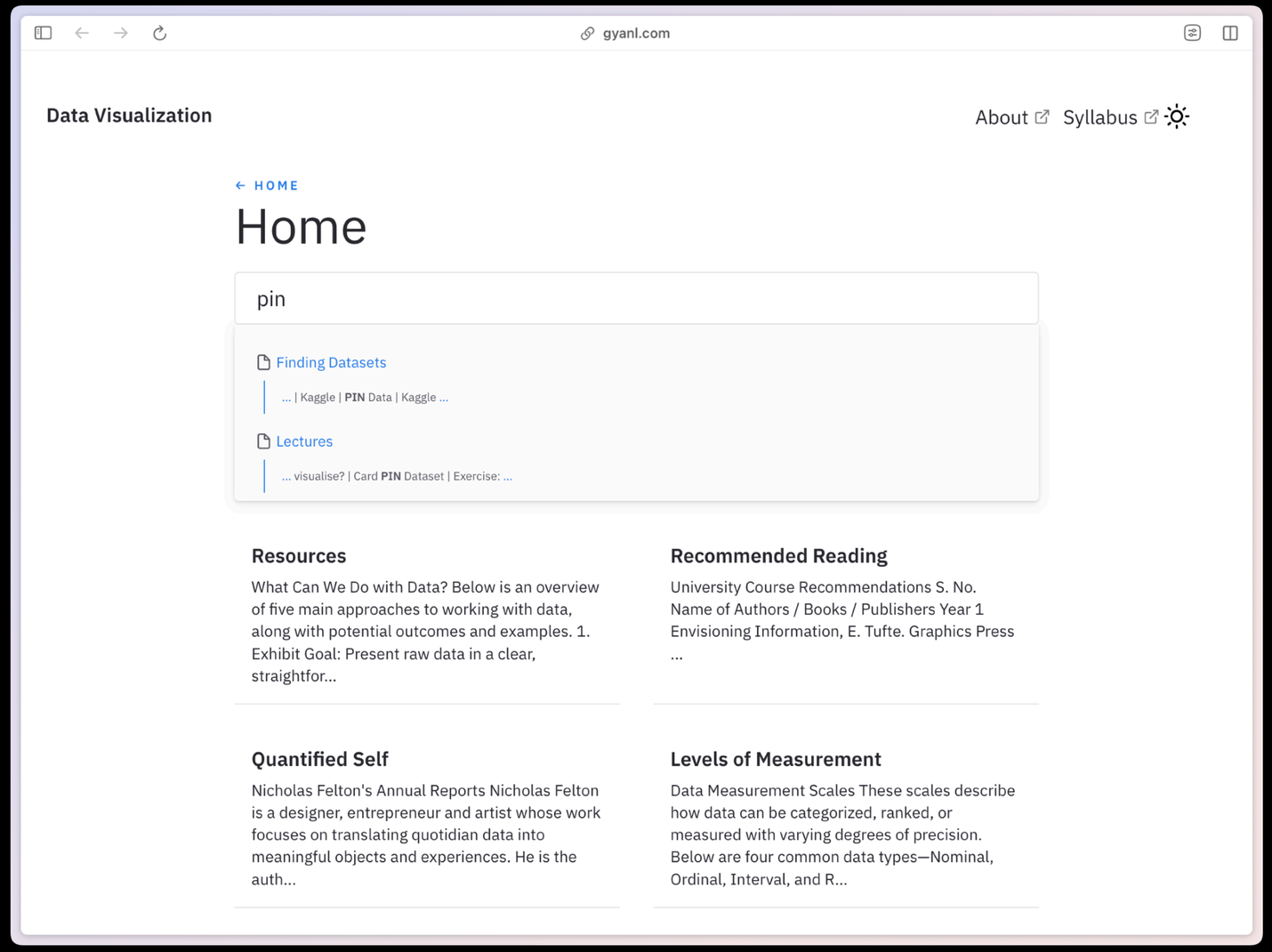
Data —— Information

# Card PINs

3.5 million PINs that were hacked from a bank.

I don’t recommend that you do this.

Here is a smaller .7 million synthetic dataset that is similar to the leaked PINs



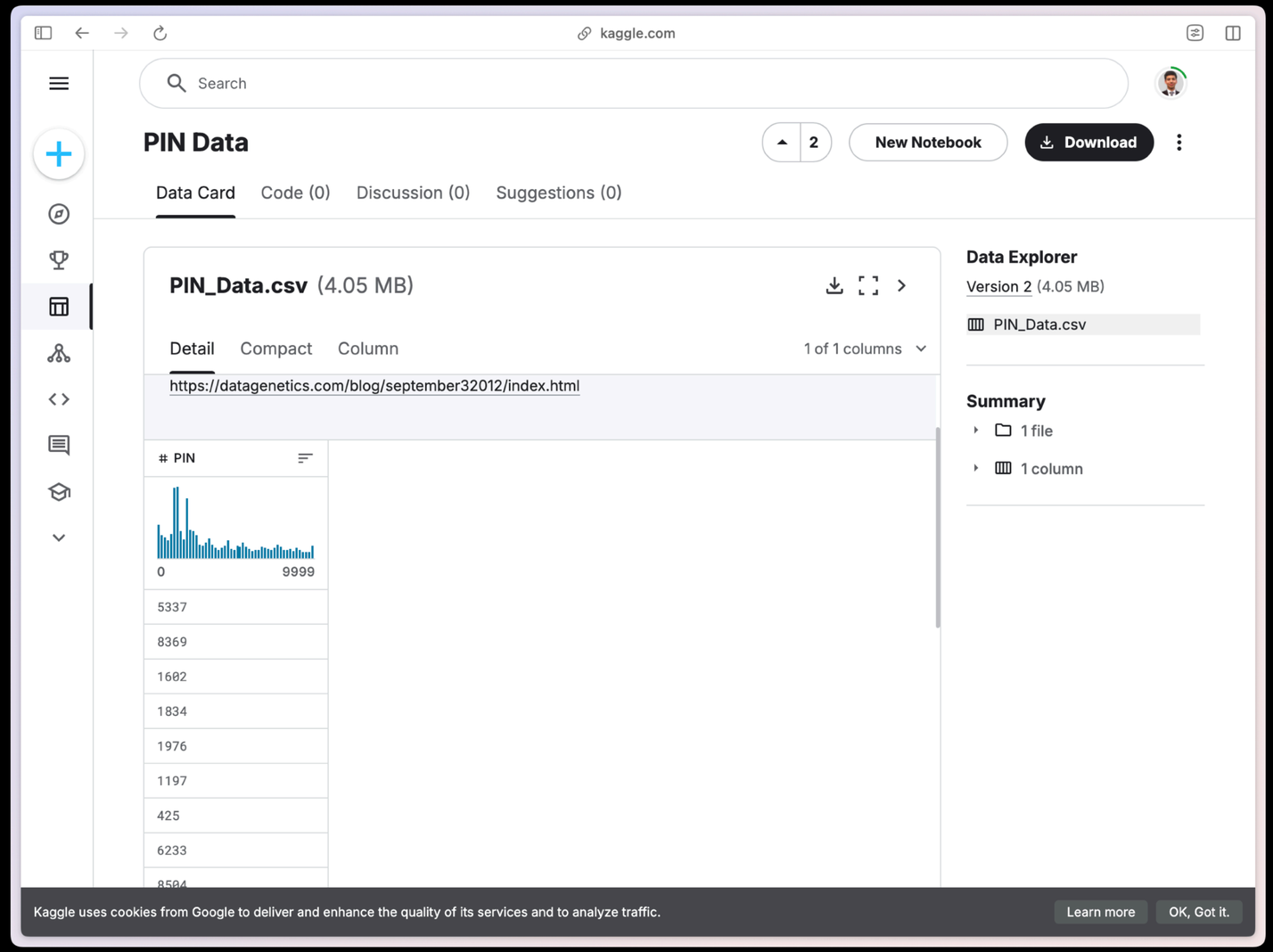
Data —— Information

# Card PINs

3.5 million PINs that were hacked from a bank.

I don't recommend that you do this.

Here is a smaller .7 million synthetic dataset that is similar to the leaked PINs



5337, 8369, 1602, 1834, 1976, 1197, 425, 6233,
8504, 9305, 1497, 1222, 2118, 8650, 4001, 6555,
1760, 8266, 1888, 9703, 1120, 6942, 622, 704,
8976, 9229, 9191, 9055, 2606, 5263, 1236, 103,
287, 5768, 2421, 7942, 7412, 829, 9232, 9790,
8440, 27, 102, 909, 208, 54, 7543, 7601,
3949, 3315, 2215, 3455, 1988, 16, 3207, 2107,
4308, 502, 3333, 9406, 4646, 7774, 3838, 2441,
8073, 6169, 4202, 2751, 1123, 1963, 563, 4242,
. . . and so on

5337, 8369, 1602, 1834, 1976, 1197, 425, 6233,
8504, 9305, 1497, 1222, 2118, 8650, 4001, 6555,
1760, 8266, 1888, 9703, 1120, 6942, 622, 704,
8976, 9229, 9191, 9055, 2606, 5263, 1236, 103,
287, 5768, 2421, 7942, 7412, 829, 9232, 9790,
8440, 27, 102, 909, 208, 54, 7543, 7601,
3949, 3315, 2215, 3455, 1988, 16, 3207, 2107,
4308, 502, 3333, 9406, 4646, 7774, 3838, 2441,
8073, 6169, 4202, 2751, 1123, 1963, 563, 4242,
. . . and so on

5337, 8369, 1602, 1834, 1976, 1197, 0425, 6233,
8504, 9305, 1497, 1222, 2118, 8650, 4001, 6555,
1760, 8266, 1888, 9703, 1120, 6942, 0622, 0704,
8976, 9229, 9191, 9055, 2606, 5263, 1236, 0103,
0287, 5768, 2421, 7942, 7412, 0829, 9232, 9790,
8440, 0027, 0102, 0909, 0208, 0054, 7543, 7601,
3949, 3315, 2215, 3455, 1988, 0016, 3207, 2107,
4308, 0502, 3333, 9406, 4646, 7774, 3838, 2441,
8073, 6169, 4202, 2751, 1123, 1963, 0563, 4242,
. . . and so on

# Data is not always available exactly how you want it.

Sometimes you will need to clean up your data to make it work for your needs.

# Do you notice any trends?

5337, 8369, 1602, 1834, 1976, 1197, 0425, 6233,
8504, 9305, 1497, 1222, 2118, 8650, 4001, 6555,
1760, 8266, 1888, 9703, 1120, 6942, 0622, 0704,
8976, 9229, 9191, 9055, 2606, 5263, 1236, 0103,
0287, 5768, 2421, 7942, 7412, 0829, 9232, 9790,
8440, 0027, 0102, 0909, 0208, 0054, 7543, 7601,
3949, 3315, 2215, 3455, 1988, 0016, 3207, 2107,
4308, 0502, 3333, 9406, 4646, 7774, 3838, 2441,
8073, 6169, 4202, 2751, 1123, 1963, 0563, 4242,

Data ——— Information

# What could we look for?

**Unique PINs**

Which all PINs are people using?

**Average (mean)**

Find the arithmetic mean of the numbers

Data —— Information

# What could we look for?

**Unique PINs**
Which all PINs are people using?

0000, 0001, 0002, 0003…

**Average (mean)**
Find the arithmetic mean of the numbers

Data —— Information

# What could we look for?

**Unique PINs**
Which all PINs are people using?

0000, 0001, 0002, 0003…

**Average (mean)**
Find the arithmetic mean of the numbers

3954

Data —— Information

# What could we look for?

**Unique PINs**
Which all PINs are people using?

**Average (mean)**
Find the arithmetic mean of the numbers

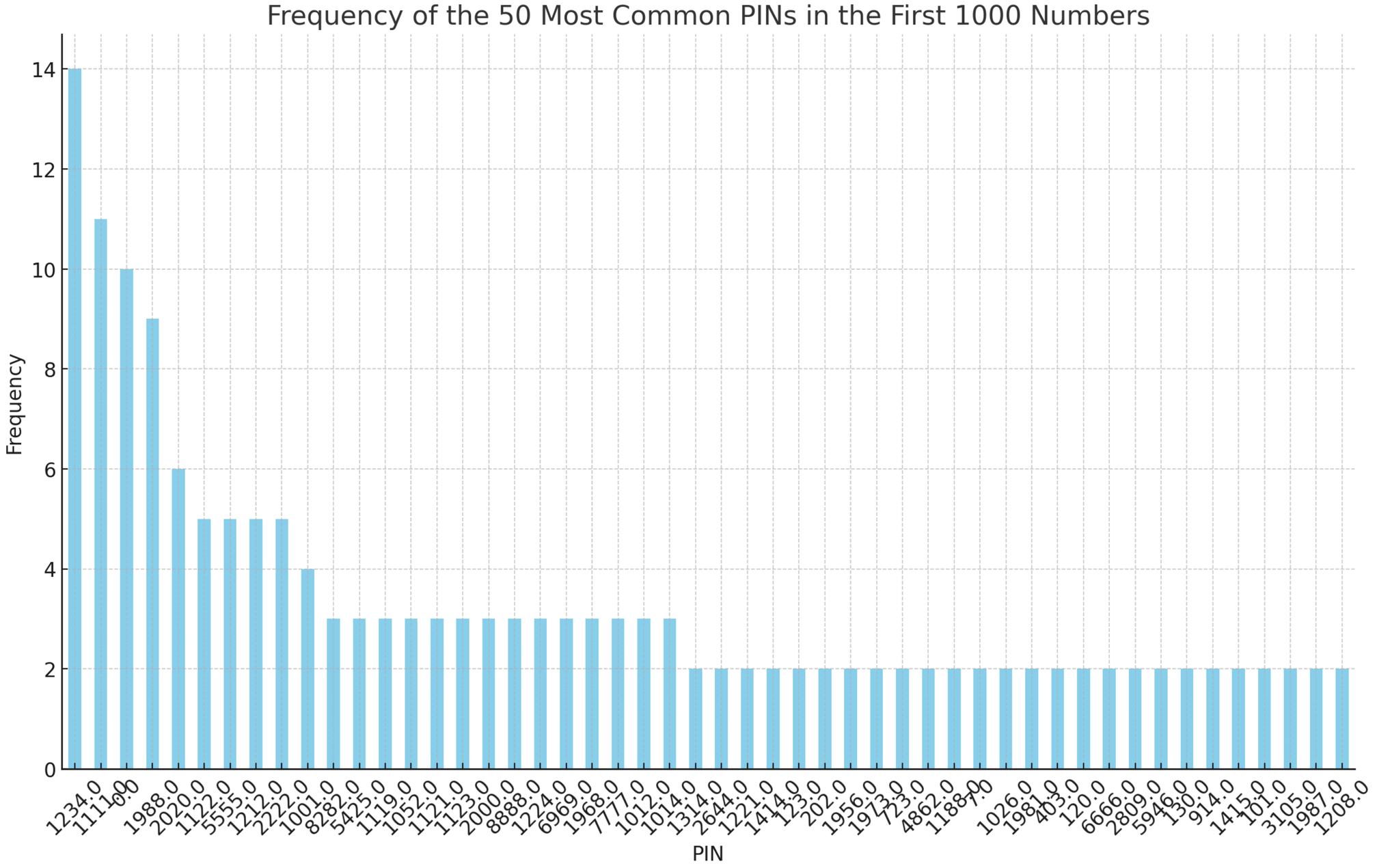
**Frequency**
Count every time each PIN shows up

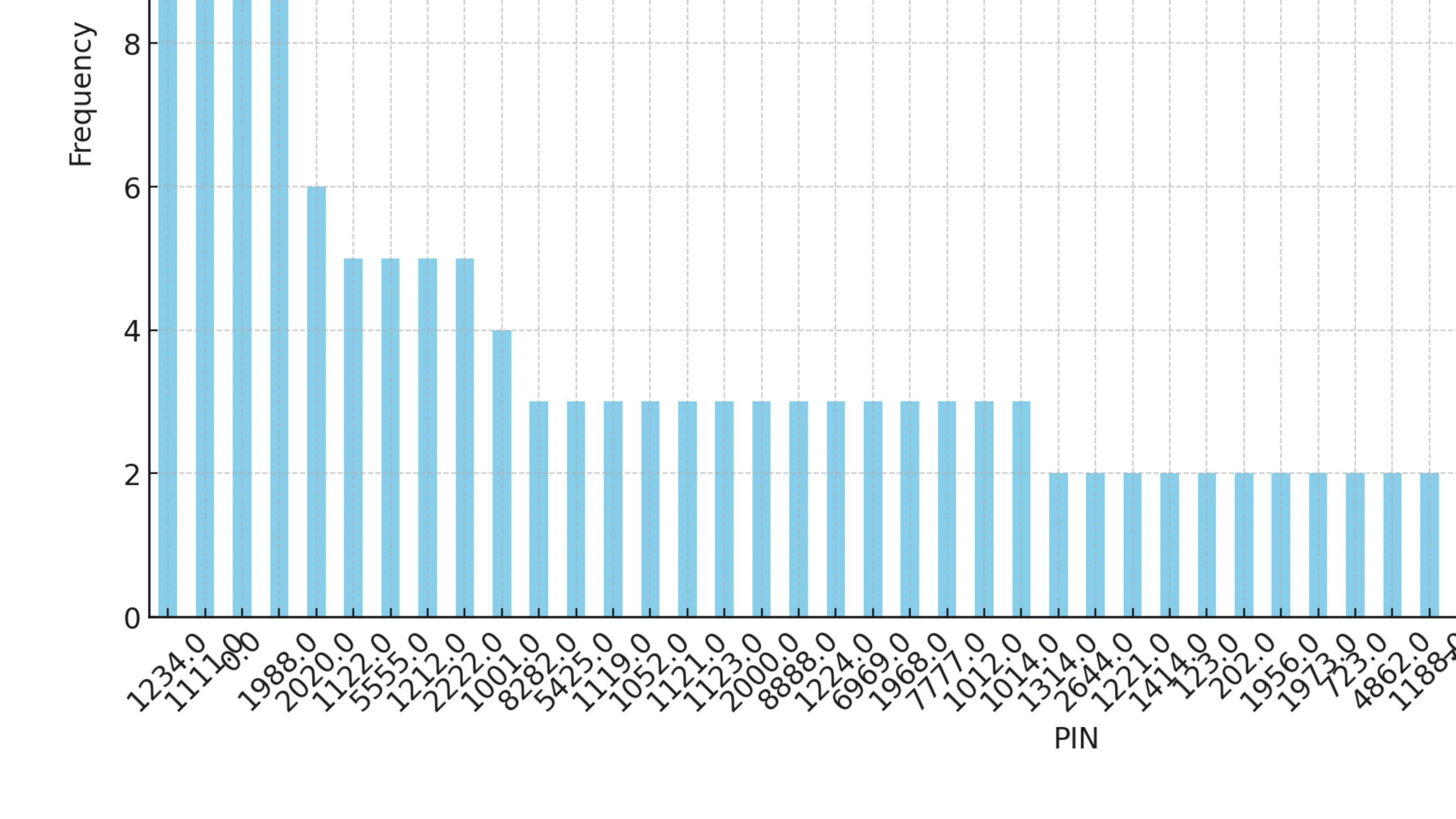
**Number frequency**
How many times 1,2,3,4... show up

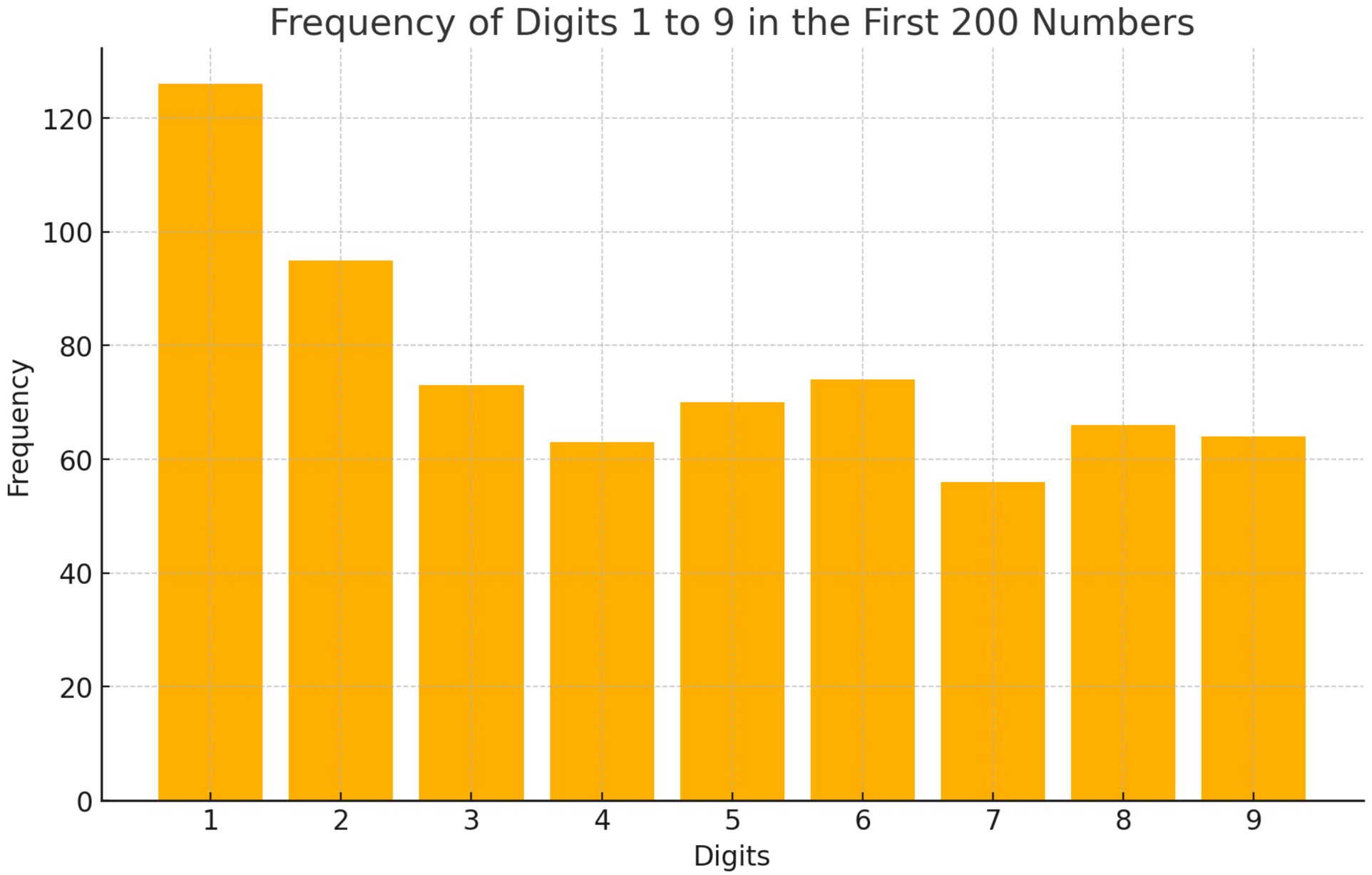
**Most common first/last digit?**
Is 0xxx or 1xxx or xxx0 very common?

**Sequences**
How many people use 1234, 9876, etc?

**Repeated Digits**
How many people use 1111, 2424, etc?


Frequency of the 50 Most Common PINs in the First 1000 Numbers
Frequency
PIN
0
2
4
6
8
10
12
14
1234.0
1111.0
0.0
1988.0
2020.0
1122.0
5555.0
1212.0
2222.0
1001.0
8282.0
5425.0
1119.0
1052.0
1121.0
1123.0
2000.0
8888.0
1224.0
6969.0
1968.0
7777.0
1012.0
1014.0
1314.0
2644.0
1221.0
1414.0
123.0
202.0
1956.0
1973.0
723.0
4862.0
1188.0
7.0
1026.0
1981.0
403.0
120.0
6666.0
2809.0
5946.0
130.0
914.0
1415.0
101.0
3105.0
1987.0
1208.0


Frequency
0
2
4
6
8
1234.0
1111.0
0.0
1988.0
2020.0
1122.0
5555.0
1212.0
2222.0
1001.0
8282.0
5425.0
1119.0
1052.0
1121.0
1123.0
2000.0
8888.0
1224.0
6969.0
1968.0
7777.0
1012.0
1014.0
1314.0
2644.0
1221.0
1414.0
123.0
202.0
1956.0
1973.0
723.0
4862.0
PIN


Frequency of Digits 1 to 9 in the First 200 Numbers
Frequency
0
20
40
60
80
100
120
1
2
3
4
5
6
7
8
9
Digits

Data —— Information

# Card PINs

3.5 million PINs that were hacked from a bank.

| | PIN | Freq |
|---|---|---|
| #1 | 1234 | 10.713% |
| #2 | 1111 | 6.016% |
| #3 | 0 | 1.881% |
| #4 | 1212 | 1.197% |
| #5 | 7777 | 0.745% |
| #6 | 1004 | 0.616% |
| #7 | 2000 | 0.613% |
| #8 | 4444 | 0.526% |
| #9 | 2222 | 0.516% |
| #10 | 6969 | 0.512% |

Data —— Information

# Card PINs

3.5 million PINs that were hacked from a bank.

| | PIN | Freq |
|---|---|---|
| #1 | 1234 | 10.713% |
| #2 | 1111 | 6.016% |
| #3 | 0 | 1.881% |
| #4 | 1212 | 1.197% |
| #5 | 7777 | 0.745% |
| #6 | 1004 | 0.616% |
| #7 | 2000 | 0.613% |
| #8 | 4444 | 0.526% |
| #9 | 2222 | 0.516% |
| #10 | 6969 | 0.512% |

Total: **23.34%**

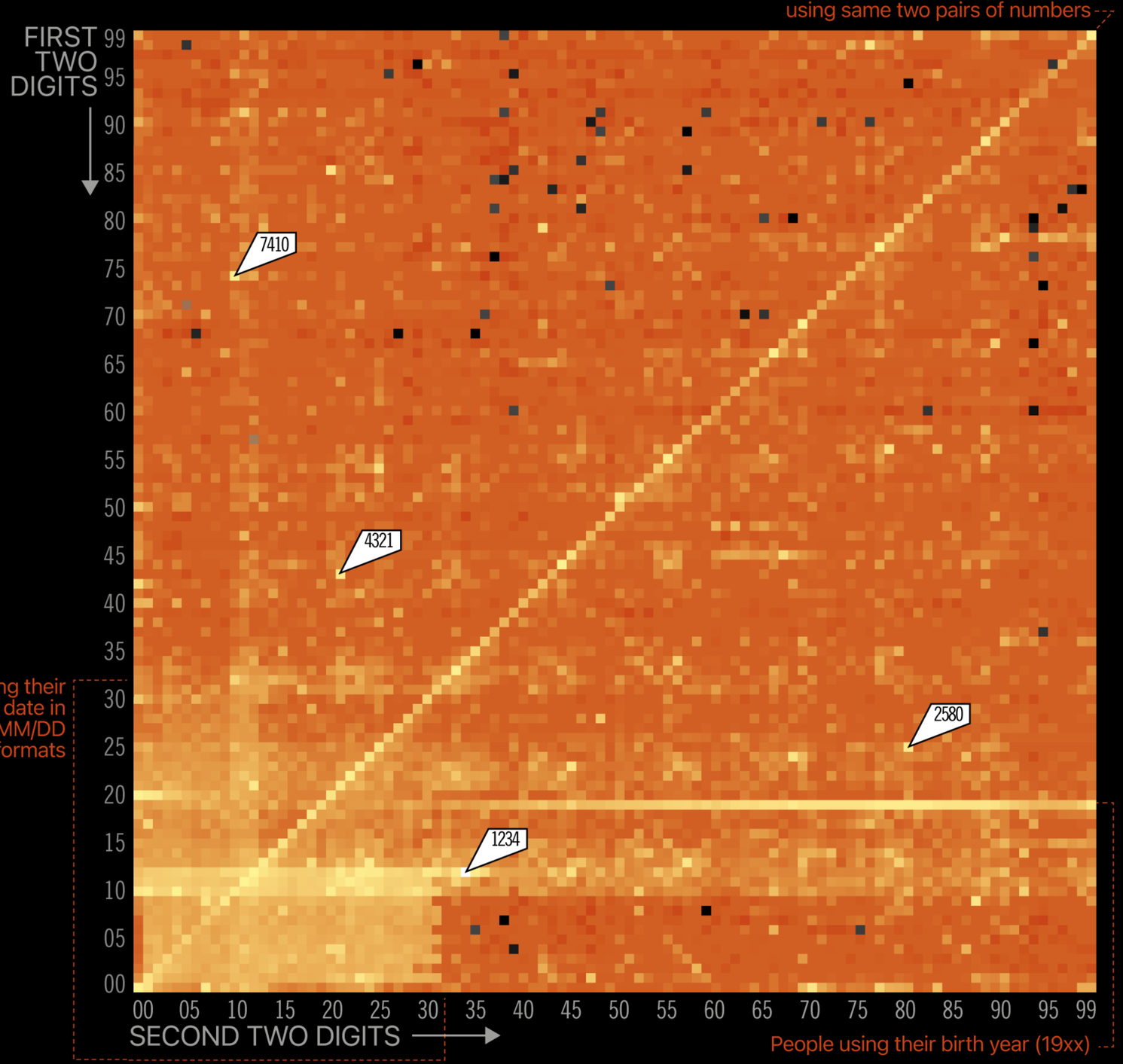
# Visualisation

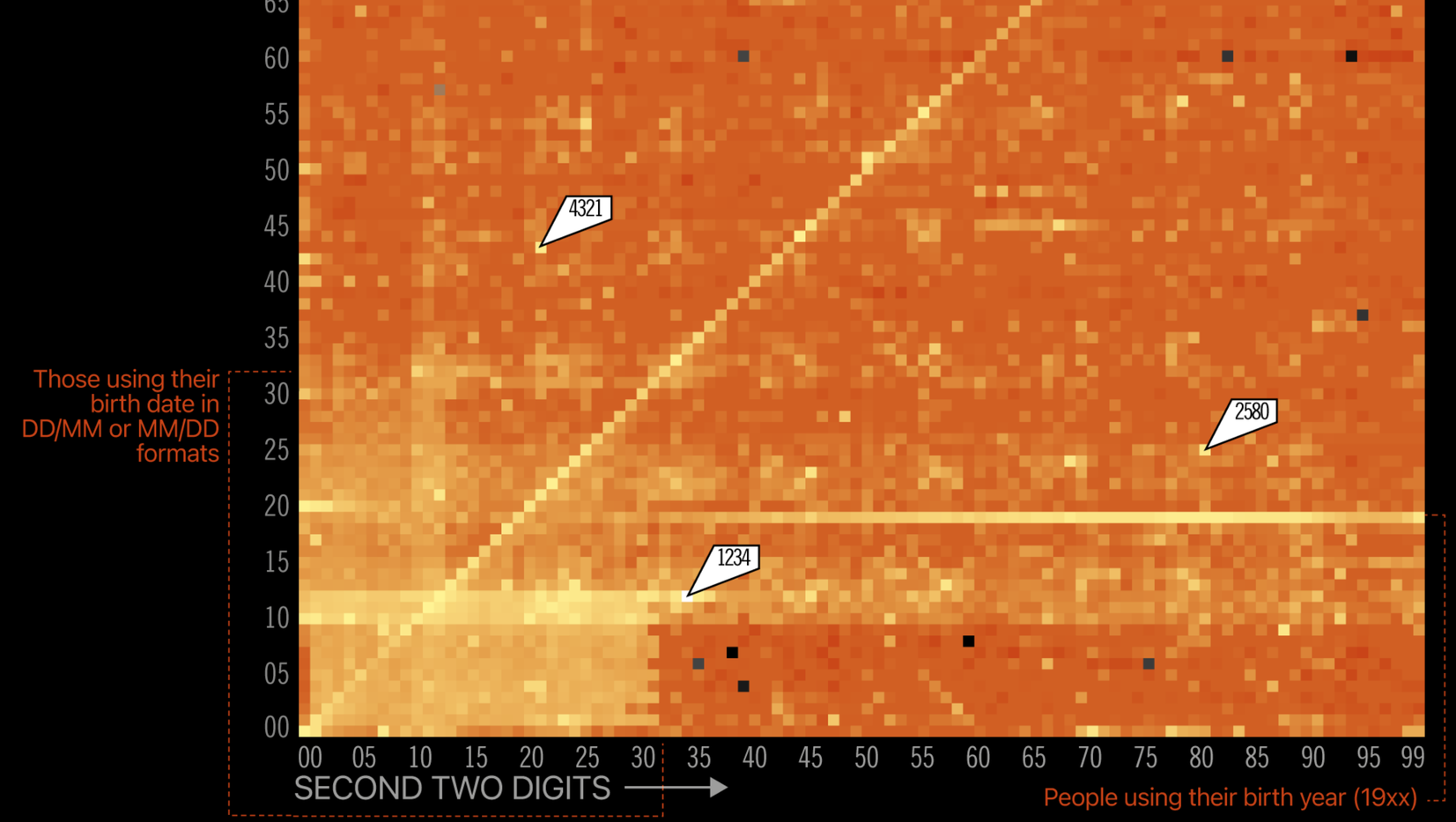
informationisbeautiful.net

# Pin Point

## The most common 4-digit PIN numbers

most common — least






## most common

1234 0000 7777 2000 2222 9999 5555 1122 8888 2001
1111 1212 1004 4444 6969 3333 6666 1313 4321 1010

27% of all PIN numbers

## least common

8557 8438 9539 7063 6827 0859 6793 0738 6835 8093
9047 0439 8196 6093 7394 9480 8398 7637 9629 8068

sources: 3.4 million PIN codes from several data breaches

Data ——— Information

# Did we learn something?

Data —— Information

# Did we learn something?

Data —— Information

# Did we learn something?

It’s really bad to use 1234, 1111 or 0000 as your PIN.

Data —— Information

# Did we learn something?

It’s really bad to use 1234, 1111 or 0000 as your PIN.

You can make sense of a large amount of data *visually* a lot faster than going over it as text.

Data —— Information

# Did we learn something?

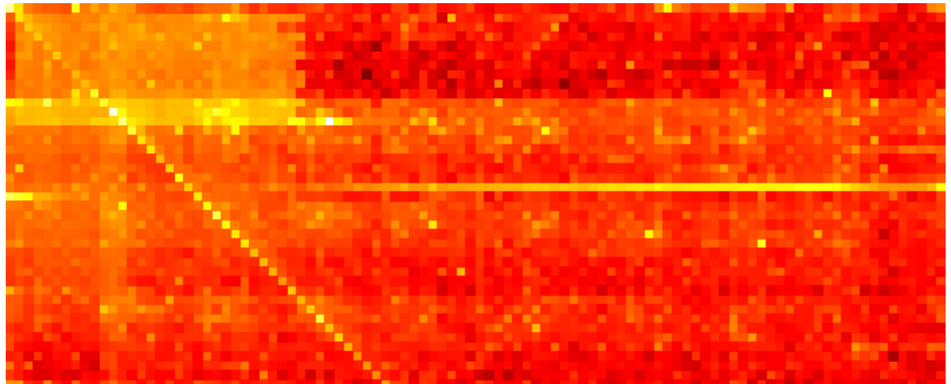
It’s really bad to use 1234, 1111 or 0000 as your PIN.

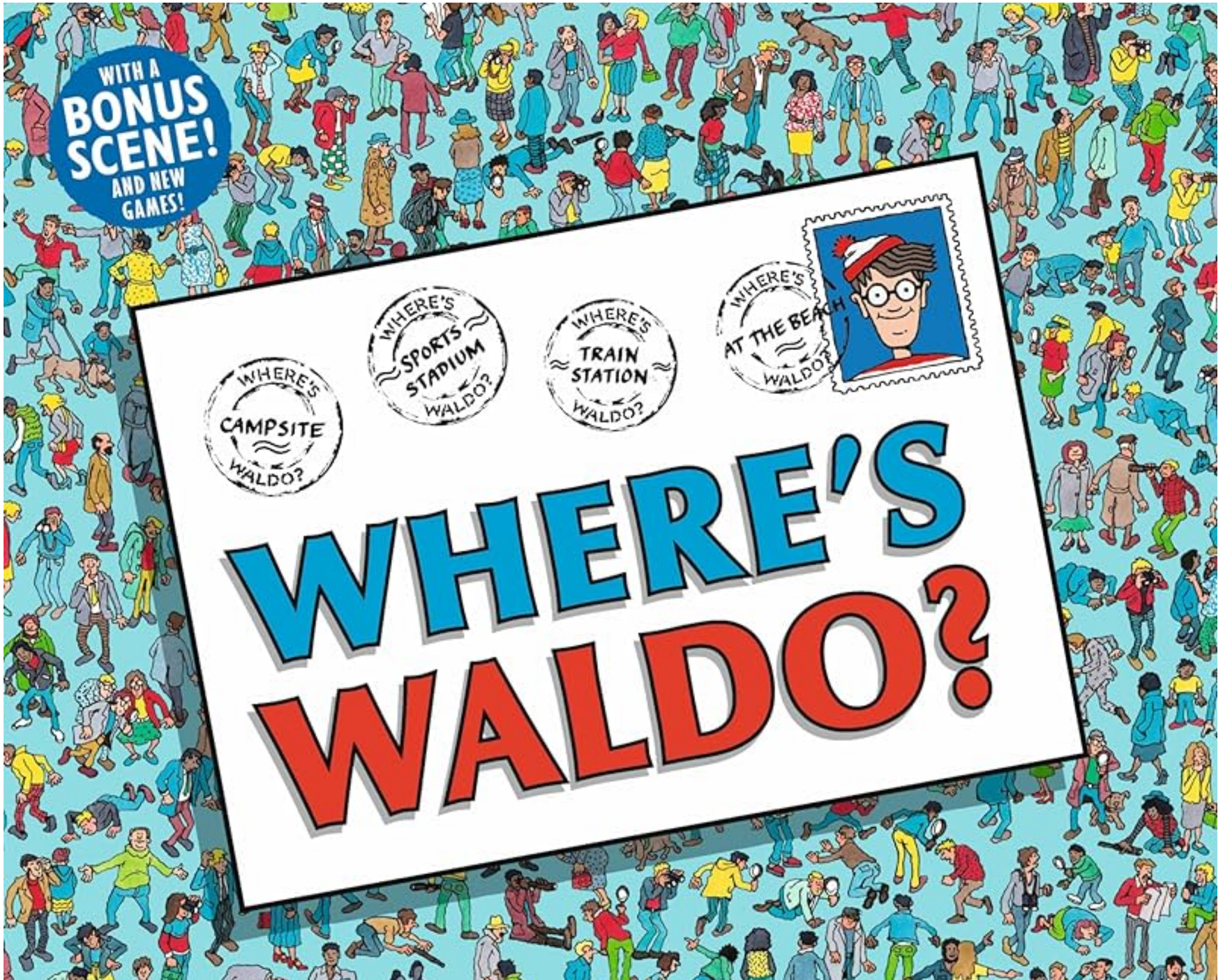
You can make sense of a large amount of data *visually* a lot faster than going over it as text.

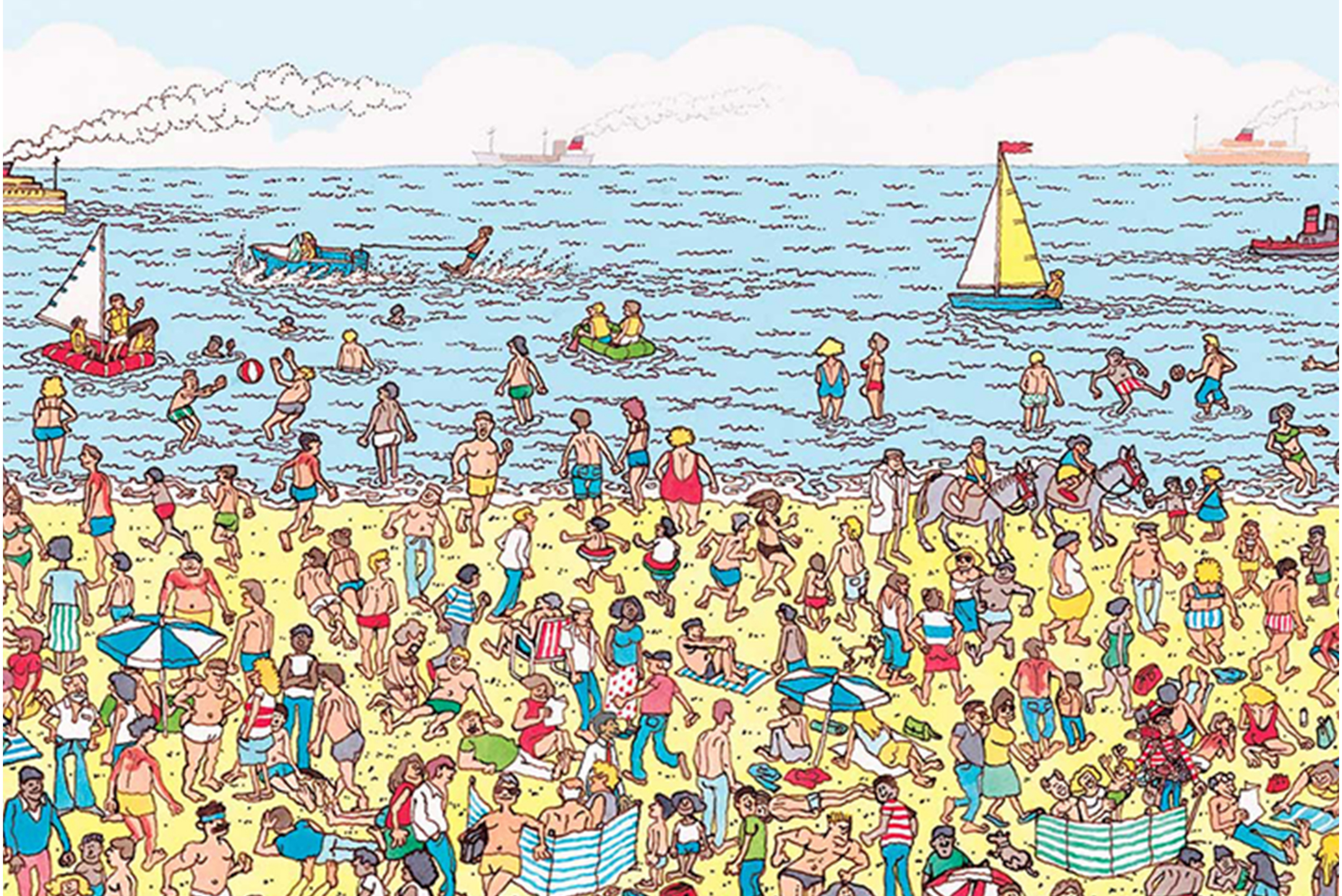
## Text

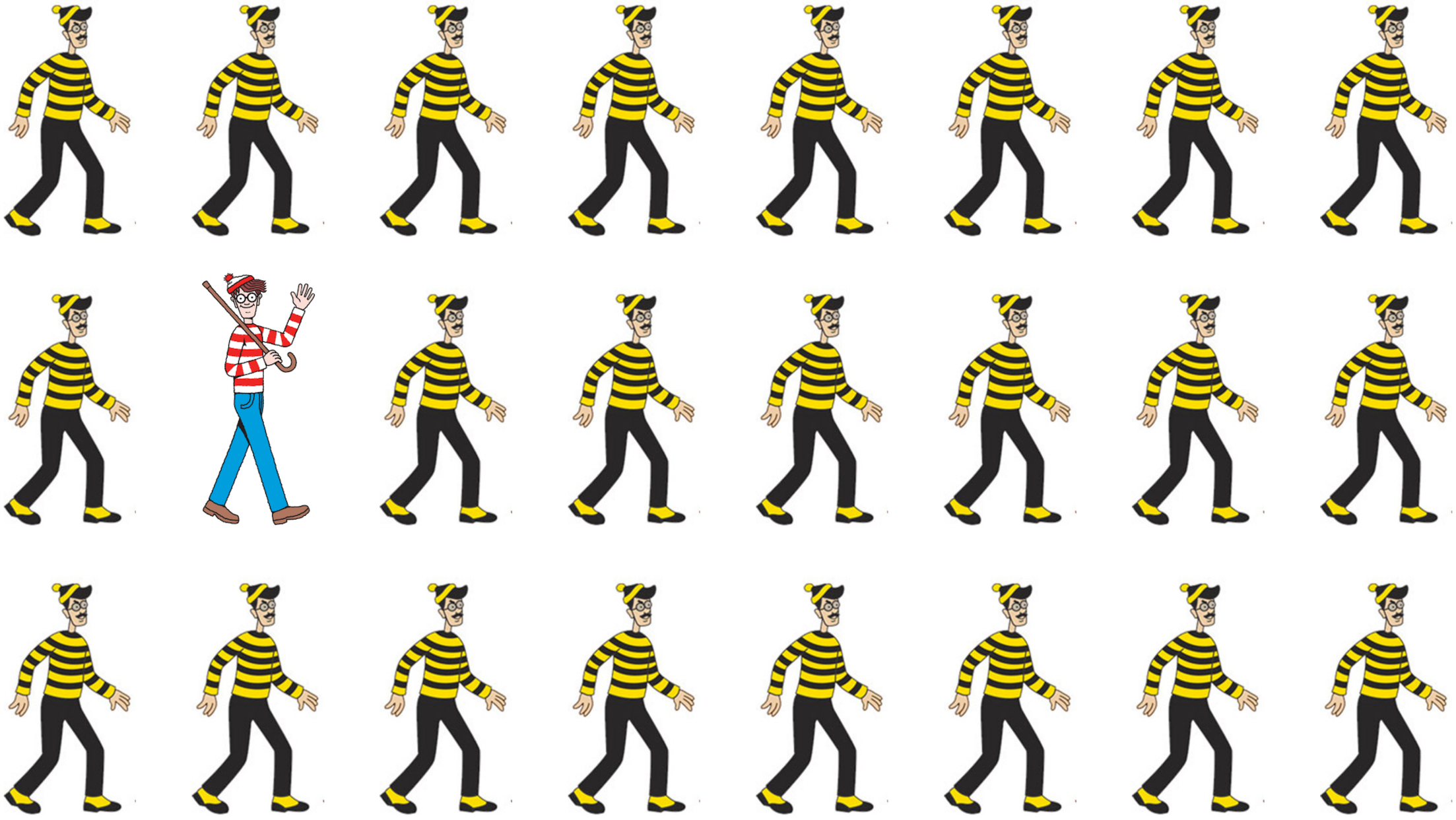
5337, 8369, 1602, 1834, 1976, 1197, 0425, 6233,
8504, 9305, 1497, 1222, 2118, 8650, 4001, 6555,
1760, 8266, 1888, 9703, 1120, 6942, 0622, 0704,
8976, 9229, 9191, 9055, 2606, 5263, 1236, 0103,
0287, 5768, 2421, 7942, 7412, 0829, 9232, 9790,
8440, 0027, 0102, 0909, 0208, 0054, 7543, 7601,
3949, 3315, 2215, 3455, 1988, 0016, 3207, 2107,
4308, 0502, 3333, 9406, 4646, 7774, 3838, 2441,
8073, 6169, 4202, 2751, 1123, 1963, 0563, 4242,

## Visualisation




WITH A
BONUS
SCENE!
AND NEW
GAMES!
WHERE'S
CAMPSITE
WALDO?
WHERE'S
SPORTS
STADIUM
WALDO?
WHERE'S
TRAIN
STATION
WALDO?
WHERE'S
AT THE BEACH
WALDO?
WHERE'S
WALDO?





Data ——— Information

# Did we learn something?

It’s really bad to use 1234, 1111 or 0000 as your PIN.

You can make sense of a large amount of data *visually* a lot faster than going over it as text.

(As long as you use an encoding that makes sense and keep human graphical perception in mind)

Data ——— Information

# Did we learn something?

It’s really bad to use 1234, 1111 or 0000 as your PIN.

You can make sense of a large amount of data *visually* a lot faster than going over it as text.

(As long as you use an encoding that makes sense and keep human graphical perception in mind)

Data ——— Information

# Did we learn something?

Visualising data can help generate an **Insight.**

Data ——— Information

# Did we learn something?

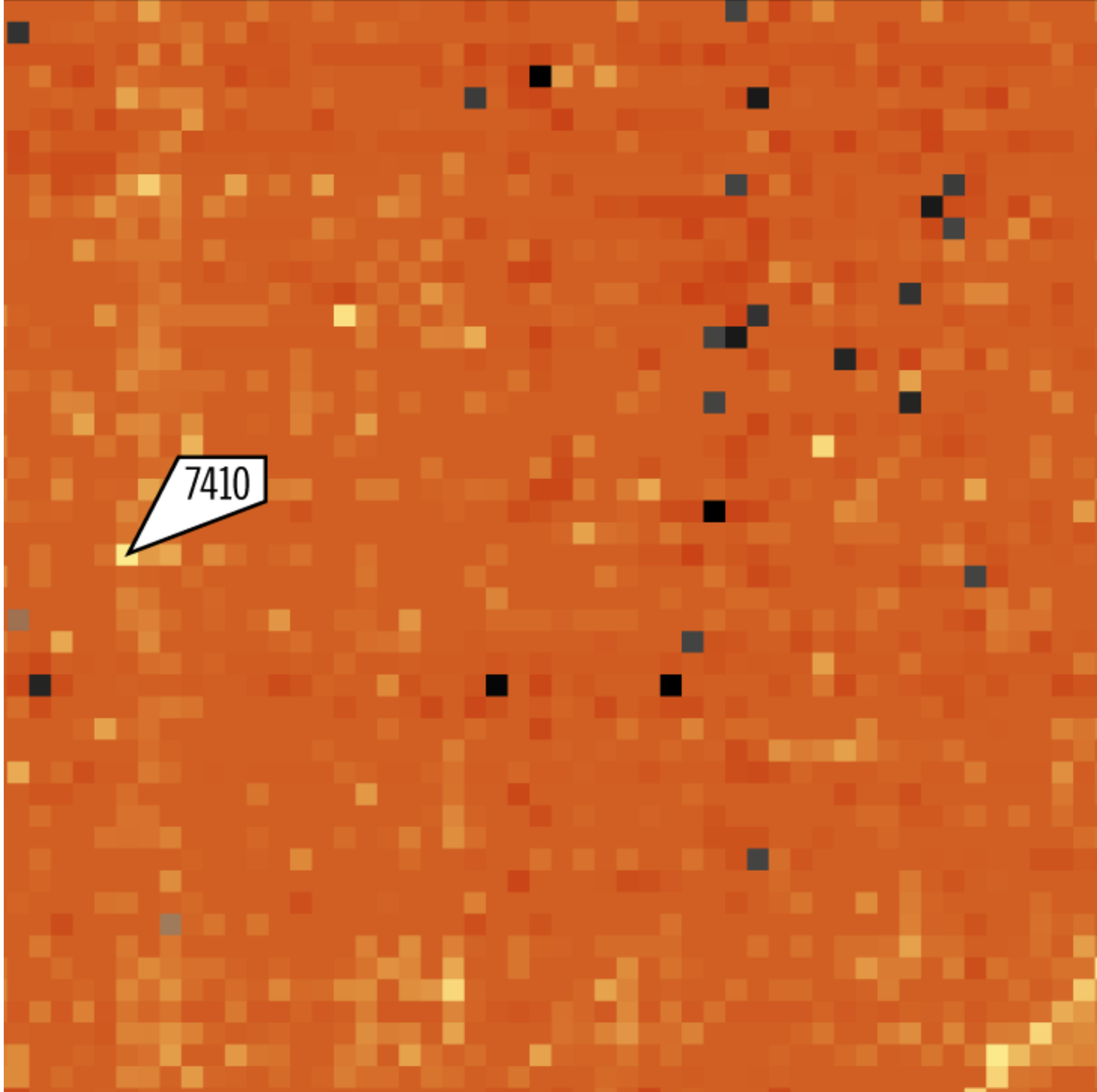
Visualising data can help generate an **Insight.**

an understanding of cause and effect based on the identification of relationships and behaviors within a model, system, context, or scenario

Data —— Information

# Insight

An understanding of cause and effect based on the identification of relationships and behaviors within a model, system, context, or scenario



Data —— Information

# Insight

An understanding of cause and effect based on the identification of relationships and behaviors within a model, system, context, or scenario

Data —— Information

# Insight

An understanding of cause and effect based on the identification of relationships and behaviors within a model, system, context, or scenario

Lecture 2

# DD 324:
# Data Visualisation

## Looking at data

14 Jan 2025 · Gyan Lakhwani · gyanlakhwani@gmail.com · Department of Design, DTU

What are

# Data Models?

a conceptual framework that defines how data is structured

## What are

# Data Models?

### Relational

Rows and columns in a table

### Hierarchical

Parent-child relationships

### Network

Complex connections

### Document

Semi-structured data

### Graph

Emphasizes relationships among data points

## What are

# Data Models?

### Relational

Rows and columns in a table

### Hierarchical

Parent-child relationships

### Network

Complex connections

### Document

Semi-structured data

### Graph

Emphasizes relationships among data points

These are more relevant for data science/database management.

## What are

# Data Models?

**Relational**

Rows and columns in a table

**Hierarchical**

Parent-child relationships

**Network**

Complex connections

**Document**

Semi-structured data

**Graph**

Emphasizes relationships among data points

We might talk about these later. For now, just be aware they exist.

Data Models

# Types of Data

Structured ——— Unstructured

# Structured

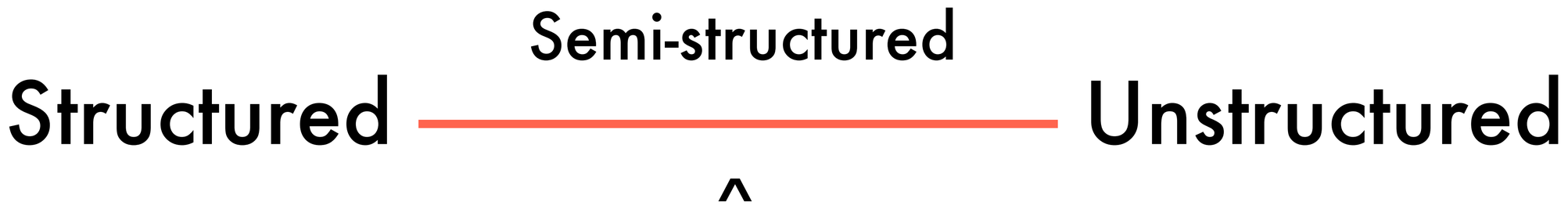
- Follows a **predefined format** and can be put in **rows and columns**.
- **Measurements of some sort**
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- **For eg.** course attendance sheet, sales data for a company, feedback popup with star rating
- **Visualisation is straightforward**

# Unstructured

- Data that **doesn't fit neatly** into a tabular format.
- **Structure is not defined**
  - (Long) Text
  - Images
  - Audio
  - Video
- **For eg.** a moodboard for a project, social media posts, feedback form with a textbox
- **Could need NLP or creativity**

Data Models

# Types of Data



## Structured or Unstructured?

# 10 emails

Structured or Unstructured?

# 10 emails and a rating for spam/ham

Structured or Unstructured?

# Star rating popup

Structured or Unstructured?

# Star rating popup with a textbox for explaining why

Structured or Unstructured?

# Names of 10 songs

Structured or Unstructured?

# Contents of 10 emails

Structured or Unstructured?

# Contents of 10 emails, from, to and date

# Structured or Unstructured?

# Structured or Unstructured?

27 March 2020
1:07 PM

31 March 2020
3:27 PM

2 April 2020
4:58 PM

6 April 2020
6:07 PM

# What structured information could you extract from these?

27 March 2020
1:07 PM

31 March 2020
3:27 PM

2 April 2020
4:58 PM

6 April 2020
6:07 PM

Data Models

# Types of Data

Structured ——— Unstructured

Data Models

# Types of Data

Structured ——— Unstructured

Data Models

# Types of Data

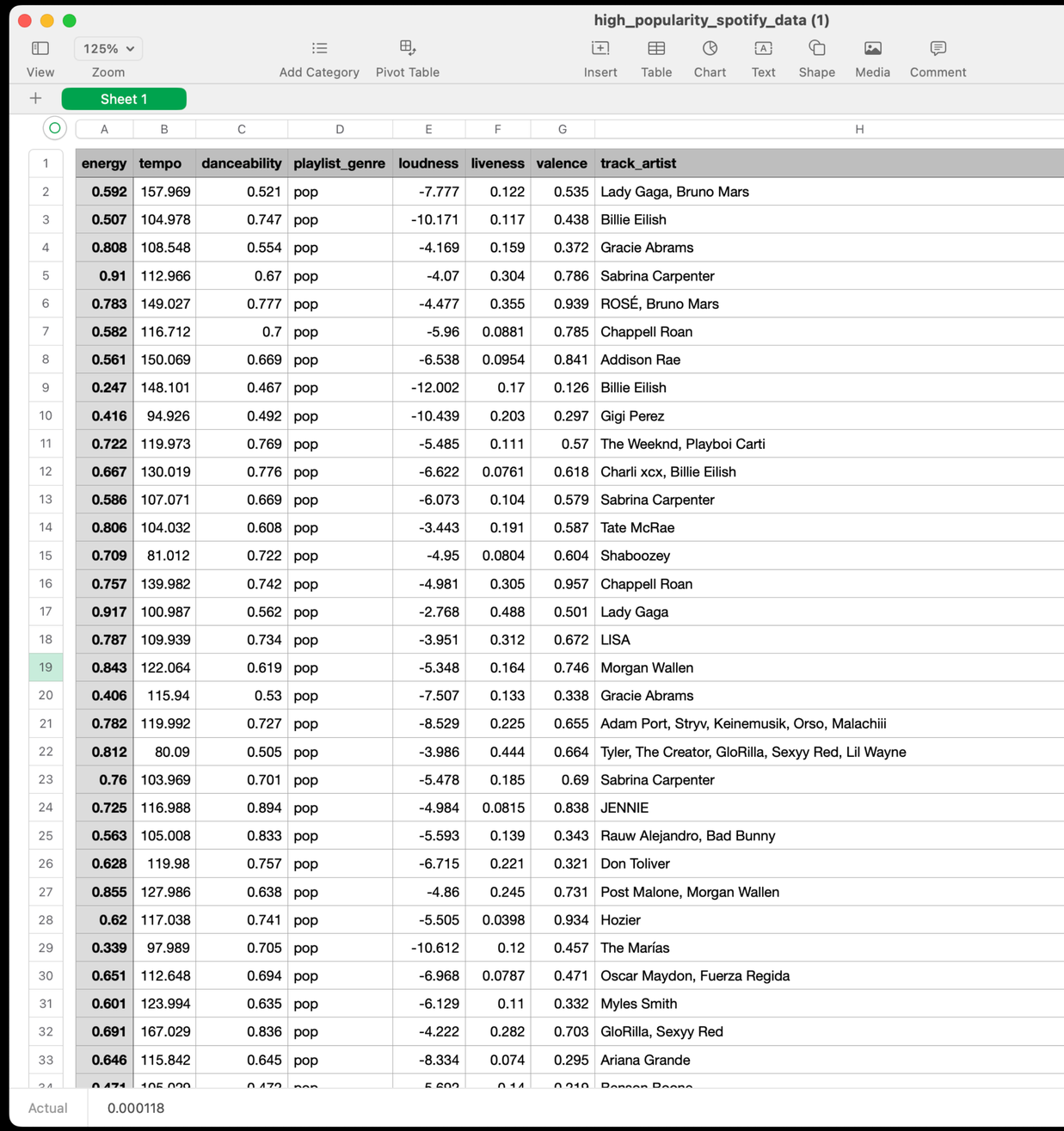
Structured ——— Unstructured

We will focus on structured data today.

Data —— Information

# Lets try and look at some datasets.

Songs from Spotify ▸


high_popularity_spotify_data (1)

125% View Zoom Add Category Pivot Table Insert Table Chart Text Shape Media Comment

Sheet 1

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | energy | tempo | danceability | playlist_genre | loudness | liveness | valence | track_artist |
| 2 | 0.592 | 157.969 | 0.521 | pop | -7.777 | 0.122 | 0.535 | Lady Gaga, Bruno Mars |
| 3 | 0.507 | 104.978 | 0.747 | pop | -10.171 | 0.117 | 0.438 | Billie Eilish |
| 4 | 0.808 | 108.548 | 0.554 | pop | -4.169 | 0.159 | 0.372 | Gracie Abrams |
| 5 | 0.91 | 112.966 | 0.67 | pop | -4.07 | 0.304 | 0.786 | Sabrina Carpenter |
| 6 | 0.783 | 149.027 | 0.777 | pop | -4.477 | 0.355 | 0.939 | ROSÉ, Bruno Mars |
| 7 | 0.582 | 116.712 | 0.7 | pop | -5.96 | 0.0881 | 0.785 | Chappell Roan |
| 8 | 0.561 | 150.069 | 0.669 | pop | -6.538 | 0.0954 | 0.841 | Addison Rae |
| 9 | 0.247 | 148.101 | 0.467 | pop | -12.002 | 0.17 | 0.126 | Billie Eilish |
| 10 | 0.416 | 94.926 | 0.492 | pop | -10.439 | 0.203 | 0.297 | Gigi Perez |
| 11 | 0.722 | 119.973 | 0.769 | pop | -5.485 | 0.111 | 0.57 | The Weeknd, Playboi Carti |
| 12 | 0.667 | 130.019 | 0.776 | pop | -6.622 | 0.0761 | 0.618 | Charli xcx, Billie Eilish |
| 13 | 0.586 | 107.071 | 0.669 | pop | -6.073 | 0.104 | 0.579 | Sabrina Carpenter |
| 14 | 0.806 | 104.032 | 0.608 | pop | -3.443 | 0.191 | 0.587 | Tate McRae |
| 15 | 0.709 | 81.012 | 0.722 | pop | -4.95 | 0.0804 | 0.604 | Shaboozey |
| 16 | 0.757 | 139.982 | 0.742 | pop | -4.981 | 0.305 | 0.957 | Chappell Roan |
| 17 | 0.917 | 100.987 | 0.562 | pop | -2.768 | 0.488 | 0.501 | Lady Gaga |
| 18 | 0.787 | 109.939 | 0.734 | pop | -3.951 | 0.312 | 0.672 | LISA |
| 19 | 0.843 | 122.064 | 0.619 | pop | -5.348 | 0.164 | 0.746 | Morgan Wallen |
| 20 | 0.406 | 115.94 | 0.53 | pop | -7.507 | 0.133 | 0.338 | Gracie Abrams |
| 21 | 0.782 | 119.992 | 0.727 | pop | -8.529 | 0.225 | 0.655 | Adam Port, Stryv, Keinemusik, Orso, Malachiii |
| 22 | 0.812 | 80.09 | 0.505 | pop | -3.986 | 0.444 | 0.664 | Tyler, The Creator, GloRilla, Sexyy Red, Lil Wayne |
| 23 | 0.76 | 103.969 | 0.701 | pop | -5.478 | 0.185 | 0.69 | Sabrina Carpenter |
| 24 | 0.725 | 116.988 | 0.894 | pop | -4.984 | 0.0815 | 0.838 | JENNIE |
| 25 | 0.563 | 105.008 | 0.833 | pop | -5.593 | 0.139 | 0.343 | Rauw Alejandro, Bad Bunny |
| 26 | 0.628 | 119.98 | 0.757 | pop | -6.715 | 0.221 | 0.321 | Don Toliver |
| 27 | 0.855 | 127.986 | 0.638 | pop | -4.86 | 0.245 | 0.731 | Post Malone, Morgan Wallen |
| 28 | 0.62 | 117.038 | 0.741 | pop | -5.505 | 0.0398 | 0.934 | Hozier |
| 29 | 0.339 | 97.989 | 0.705 | pop | -10.612 | 0.12 | 0.457 | The Marías |
| 30 | 0.651 | 112.648 | 0.694 | pop | -6.968 | 0.0787 | 0.471 | Oscar Maydon, Fuerza Regida |
| 31 | 0.601 | 123.994 | 0.635 | pop | -6.129 | 0.11 | 0.332 | Myles Smith |
| 32 | 0.691 | 167.029 | 0.836 | pop | -4.222 | 0.282 | 0.703 | GloRilla, Sexyy Red |
| 33 | 0.646 | 115.842 | 0.645 | pop | -8.334 | 0.074 | 0.295 | Ariana Grande |

Actual 0.000118

## Spotify Dataset

# A dataset can have many **features.**

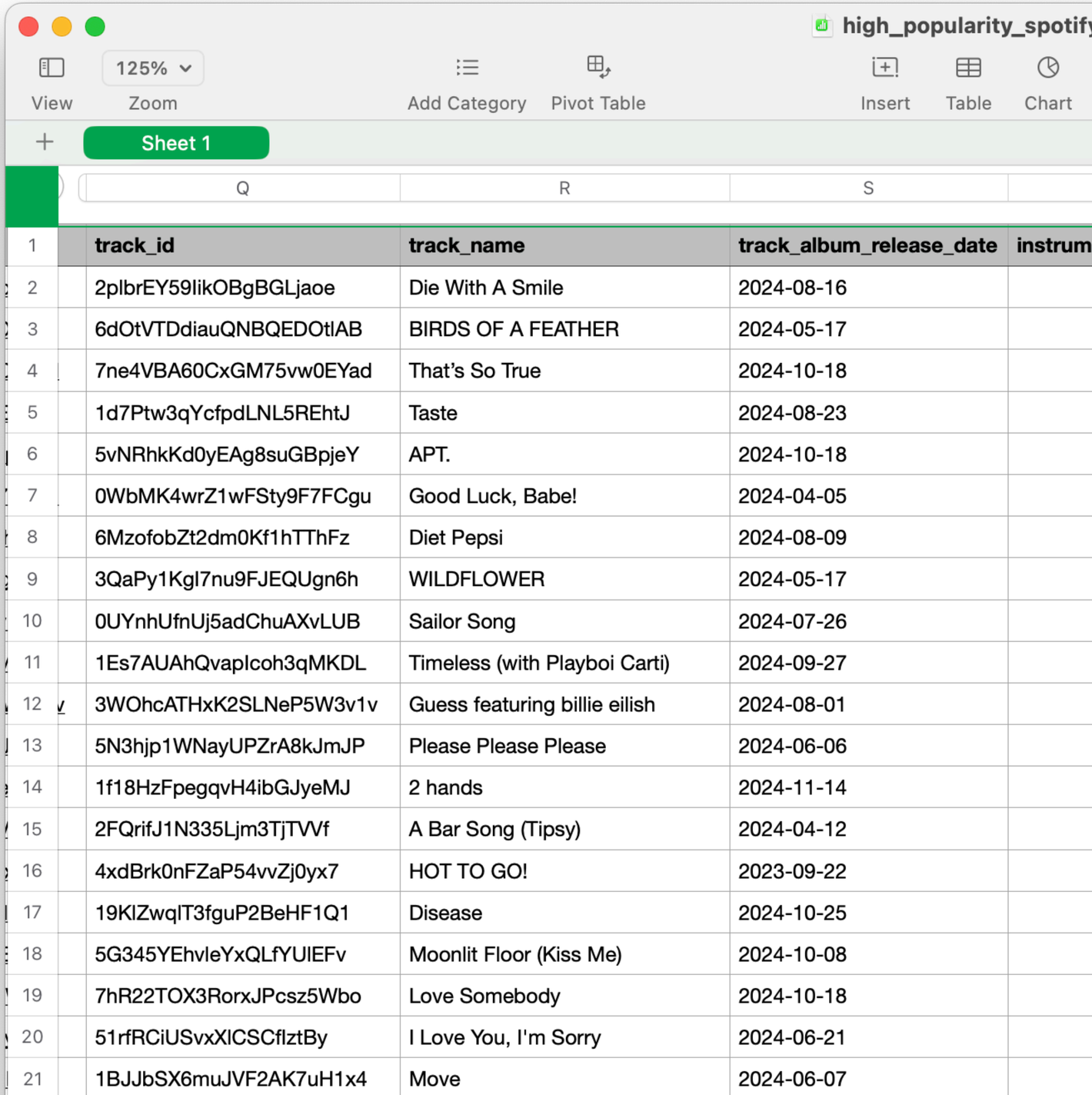
Here are some of the features this dataset has: song name, artist, genre, popularity, energy, tempo, danceability...

| Feature | Description |
|---|---|
| **Energy** | A measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. |
| **Tempo** | The speed of a track, measured in beats per minute (BPM). |
| **Danceability** | A score describing how suitable a track is for dancing based on tempo, rhythm stability, beat strength and overall regularity. |
| **Loudness** | The overall loudness of a track in decibels (dB). Higher values indicate louder tracks overall. |
| **Liveness** | The likelihood of a track being performed live. Higher values suggest more audience presence. |
| **Valence** | The overall musical positiveness(emotion) of a track. High valence sounds happy; low valence sounds sad or angry. |
| **Speechiness** | Measures the presence of spoken words. |
| **Instrumentalness** | The likelihood a track contains no vocals. Values closer to 1.0 suggest solely instrumental tracks. |
| **Mode** | Indicates the modality of the track. |

Features

# Usually the title of each column in a table is the feature name.

At least one feature should have unique values.

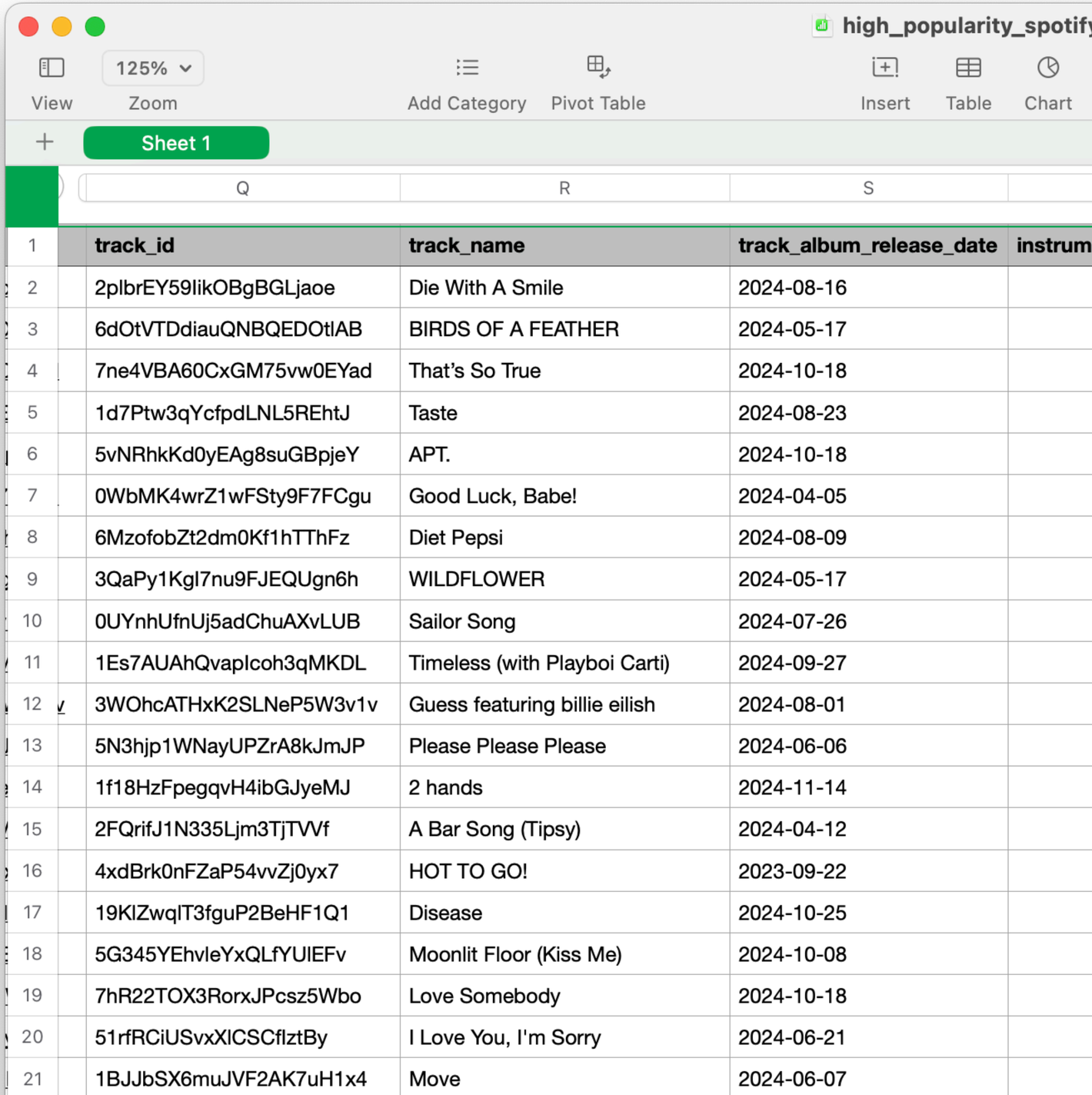



| | Q | R | S |
|---|---|---|---|
| 1 | track_id | track_name | track_album_release_date |
| 2 | 2plbrEY59IikOBgBGLjaoe | Die With A Smile | 2024-08-16 |
| 3 | 6dOtVTDdiauQNBQEDOtlAB | BIRDS OF A FEATHER | 2024-05-17 |
| 4 | 7ne4VBA60CxGM75vw0EYad | That's So True | 2024-10-18 |
| 5 | 1d7Ptw3qYcfpdLNL5REhtJ | Taste | 2024-08-23 |
| 6 | 5vNRhkKd0yEAg8suGBpjeY | APT. | 2024-10-18 |
| 7 | 0WbMK4wrZ1wFSty9F7FCgu | Good Luck, Babe! | 2024-04-05 |
| 8 | 6MzofobZt2dm0Kf1hTThFz | Diet Pepsi | 2024-08-09 |
| 9 | 3QaPy1Kgl7nu9FJEQUgn6h | WILDFLOWER | 2024-05-17 |
| 10 | 0UYnhUfnUj5adChuAXvLUB | Sailor Song | 2024-07-26 |
| 11 | 1Es7AUAhQvapIcoh3qMKDL | Timeless (with Playboi Carti) | 2024-09-27 |
| 12 | 3WOhcATHxK2SLNeP5W3v1v | Guess featuring billie eilish | 2024-08-01 |
| 13 | 5N3hjp1WNayUPZrA8kJmJP | Please Please Please | 2024-06-06 |
| 14 | 1f18HzFpegqvH4ibGJyeMJ | 2 hands | 2024-11-14 |
| 15 | 2FQrifJ1N335Ljm3TjTVVf | A Bar Song (Tipsy) | 2024-04-12 |
| 16 | 4xdBrk0nFZaP54vvZj0yx7 | HOT TO GO! | 2023-09-22 |
| 17 | 19KlZwqlT3fguP2BeHF1Q1 | Disease | 2024-10-25 |
| 18 | 5G345YEhvleYxQLfYUlEFv | Moonlit Floor (Kiss Me) | 2024-10-08 |
| 19 | 7hR22TOX3RorxJPcsz5Wbo | Love Somebody | 2024-10-18 |
| 20 | 51rfRCiUSvxXlCSCflztBy | I Love You, I'm Sorry | 2024-06-21 |
| 21 | 1BJJbSX6muJVF2AK7uH1x4 | Move | 2024-06-07 |

Features

# Usually the title of each column in a table is the feature name.

# At least one feature should have unique values... why?




| | Q | R | S |
|---|---|---|---|
| 1 | track_id | track_name | track_album_release_date |
| 2 | 2plbrEY59IikOBgBGLjaoe | Die With A Smile | 2024-08-16 |
| 3 | 6dOtVTDdiauQNBQEDOtlAB | BIRDS OF A FEATHER | 2024-05-17 |
| 4 | 7ne4VBA60CxGM75vw0EYad | That's So True | 2024-10-18 |
| 5 | 1d7Ptw3qYcfpdLNL5REhtJ | Taste | 2024-08-23 |
| 6 | 5vNRhkKd0yEAg8suGBpjeY | APT. | 2024-10-18 |
| 7 | 0WbMK4wrZ1wFSty9F7FCgu | Good Luck, Babe! | 2024-04-05 |
| 8 | 6MzofobZt2dm0Kf1hTThFz | Diet Pepsi | 2024-08-09 |
| 9 | 3QaPy1Kgl7nu9FJEQUgn6h | WILDFLOWER | 2024-05-17 |
| 10 | 0UYnhUfnUj5adChuAXvLUB | Sailor Song | 2024-07-26 |
| 11 | 1Es7AUAhQvapIcoh3qMKDL | Timeless (with Playboi Carti) | 2024-09-27 |
| 12 | 3WOhcATHxK2SLNeP5W3v1v | Guess featuring billie eilish | 2024-08-01 |
| 13 | 5N3hjp1WNayUPZrA8kJmJP | Please Please Please | 2024-06-06 |
| 14 | 1f18HzFpegqvH4ibGJyeMJ | 2 hands | 2024-11-14 |
| 15 | 2FQrifJ1N335Ljm3TjTVVf | A Bar Song (Tipsy) | 2024-04-12 |
| 16 | 4xdBrk0nFZaP54vvZj0yx7 | HOT TO GO! | 2023-09-22 |
| 17 | 19KlZwqlT3fguP2BeHF1Q1 | Disease | 2024-10-25 |
| 18 | 5G345YEhvleYxQLfYUlEFv | Moonlit Floor (Kiss Me) | 2024-10-08 |
| 19 | 7hR22TOX3RorxJPcsz5Wbo | Love Somebody | 2024-10-18 |
| 20 | 51rfRCiUSvxXlCSCflztBy | I Love You, I'm Sorry | 2024-06-21 |
| 21 | 1BJJbSX6muJVF2AK7uH1x4 | Move | 2024-06-07 |

| | | | | |
|---|---|---|---|---|
| Lover<br>Diljit Dosanjh | 14,35,53,214 | ✓ | 3:10 |

| | | | | |
|---|---|---|---|---|
| Lover<br>Diljit Dosanjh | 14,35,53,214 | ✓ | 3:10 |
| Lover<br>Taylor Swift | 1,65,05,10,116 | | 3:41 |

Spotify Dataset

# What are some Features that might impact each other?

| Feature | Description |
| --- | --- |
| **Energy** | A measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. |
| **Tempo** | The speed of a track, measured in beats per minute (BPM). |
| **Danceability** | A score describing how suitable a track is for dancing based on tempo, rhythm stability, beat strength and overall regularity. |
| **Loudness** | The overall loudness of a track in decibels (dB). Higher values indicate louder tracks overall. |
| **Liveness** | The likelihood of a track being performed live. Higher values suggest more audience presence. |
| **Valence** | The overall musical positiveness(emotion) of a track. High valence sounds happy; low valence sounds sad or angry. |
| **Speechiness** | Measures the presence of spoken words. |
| **Instrumentalness** | The likelihood a track contains no vocals. Values closer to 1.0 suggest solely instrumental tracks. |
| **Mode** | Indicates the modality of the track. |

## Spotify Dataset

# What are some Features that might impact each other?

## Hypothesis

# High energy tracks are Louder.

| Feature | Description |
|---|---|
| **Energy** | A measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. |
| **Tempo** | The speed of a track, measured in beats per minute (BPM). |
| **Danceability** | A score describing how suitable a track is for dancing based on tempo, rhythm stability, beat strength and overall regularity. |
| **Loudness** | The overall loudness of a track in decibels (dB). Higher values indicate louder tracks overall. |
| **Liveness** | The likelihood of a track being performed live. Higher values suggest more audience presence. |
| **Valence** | The overall musical positiveness(emotion) of a track. High valence sounds happy; low valence sounds sad or angry. |
| **Speechiness** | Measures the presence of spoken words. |
| **Instrumentalness** | The likelihood a track contains no vocals. Values closer to 1.0 suggest solely instrumental tracks. |
| **Mode** | Indicates the modality of the track. |

Data ——— Information

# You can use data to prove or disprove a hypothesis.

Data ——— Information

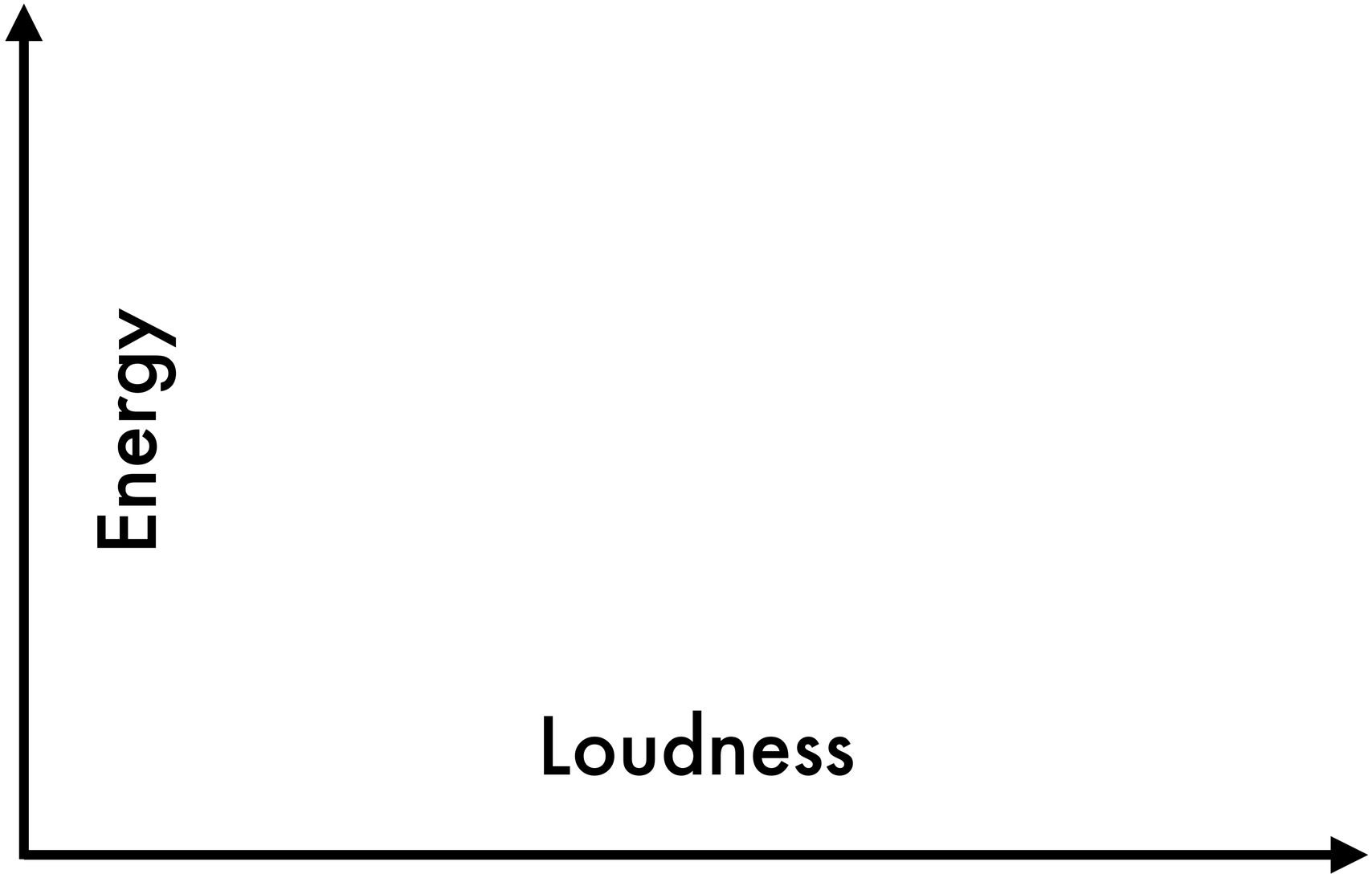
# You can use data visualisation to prove or disprove a hypothesis.

**Hypothesis**

High energy tracks are Louder.

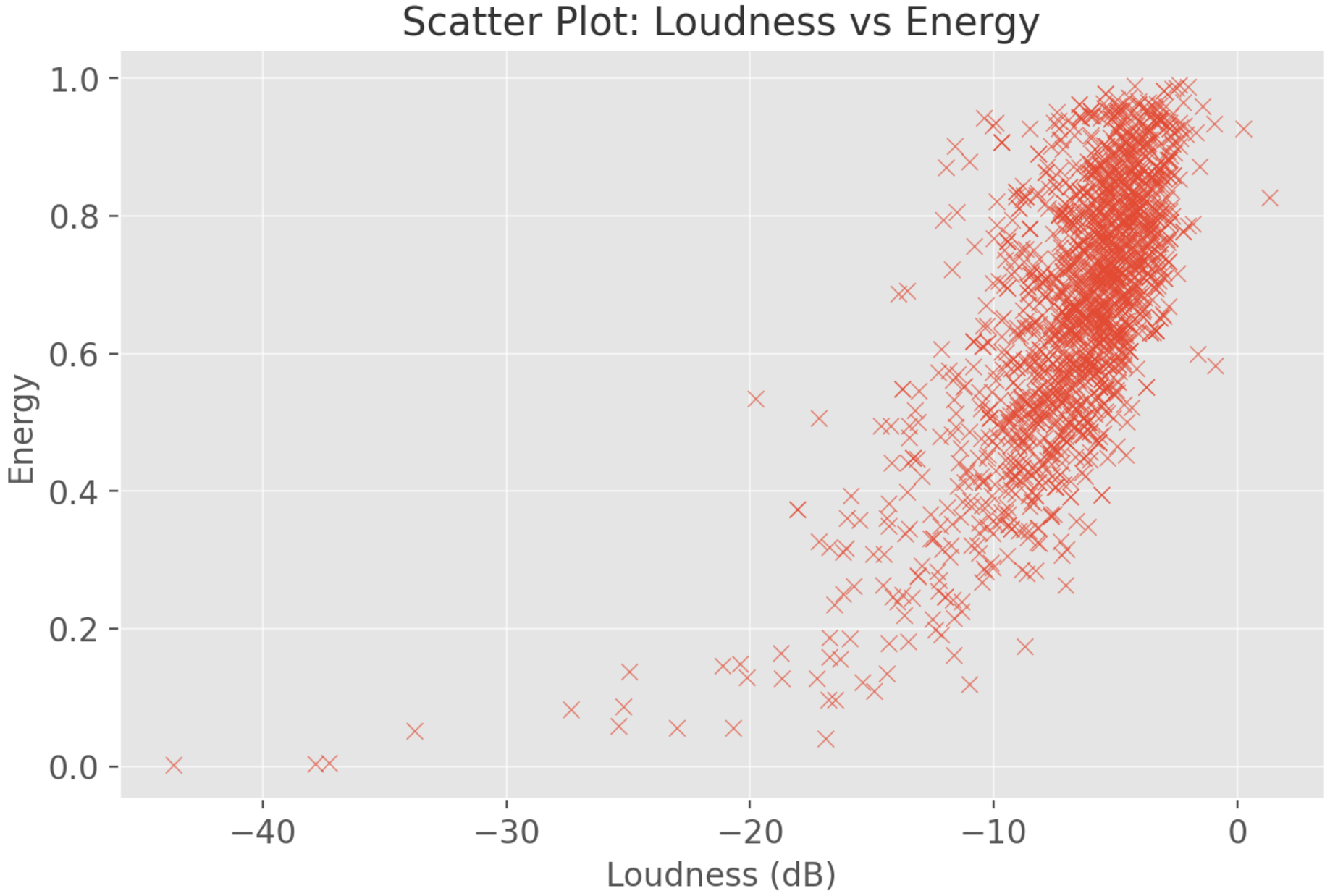
**Hypothesis**

High energy tracks are Louder.



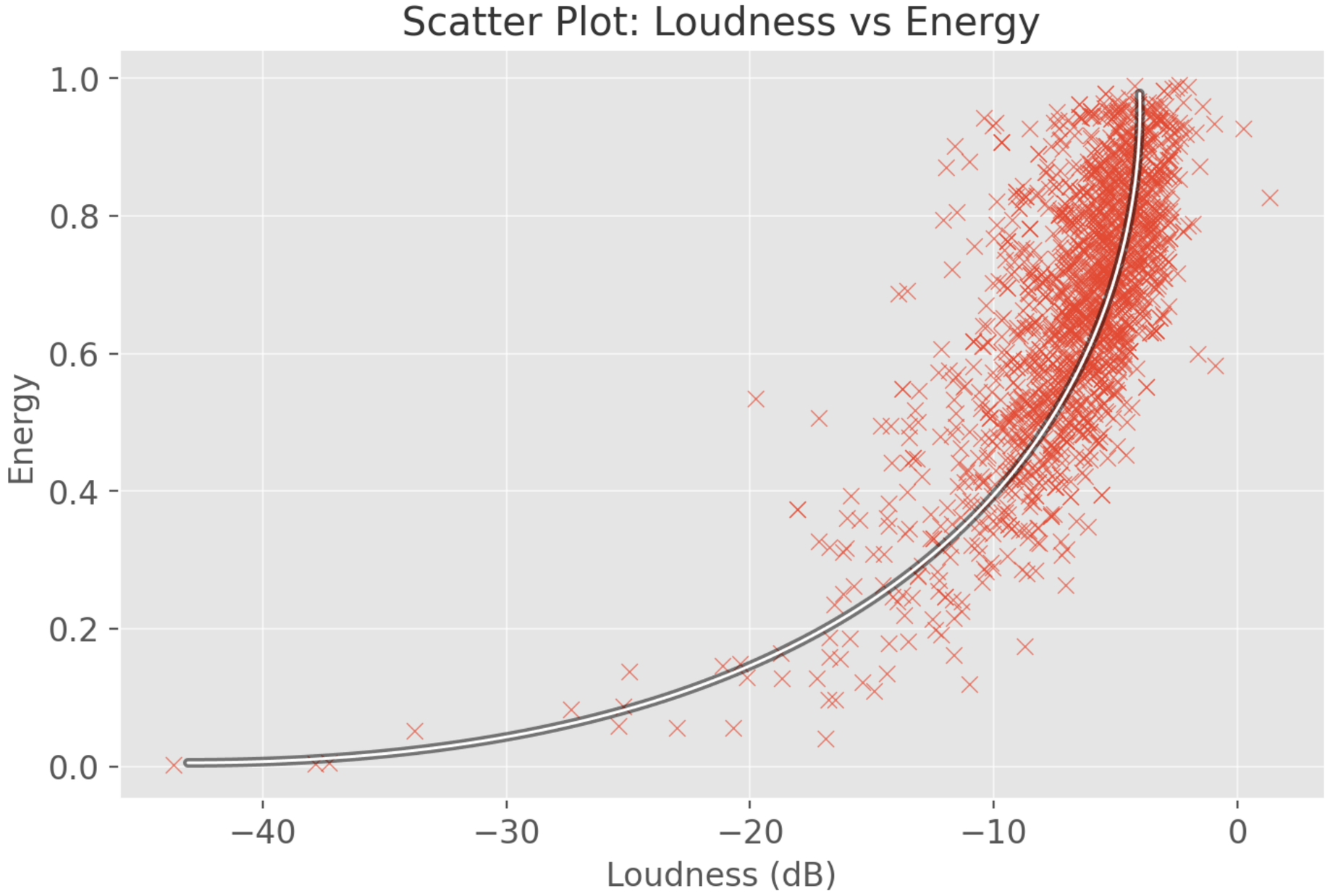
**Hypothesis**

# High energy tracks are Louder.



**Hypothesis**

# High energy tracks are Louder.



Exercise

# Explore a Dataset

## ✅ Dataset with 4+ features

| Date<br>The date when the stock data was recorded. Represents each trading day. | # Adj_Close<br>The adjusted closing price accounting for corporate actions like dividends. | # Close<br>The final price at which the stock was traded on that day. | # High<br>The highest price that Adobe's stock reached on a given day. | # Low<br>The lowest price Adobe's stock reached during a trading day. |
|---|---|---|---|---|
| <br>1986-08-13 2024-12-27 | <br>0.19 688 | <br>0.21 688 | <br>0.22 700 | <br>0.21 679 |
| 1986-08-13 | 0.19805713 | 0.210938007 | 0.21875 | 0.210938007 |
| 1986-08-14 | 0.209059551 | 0.222655997 | 0.230469003 | 0.222655997 |
| 1986-08-15 | 0.205392078 | 0.21875 | 0.222655997 | 0.21875 |

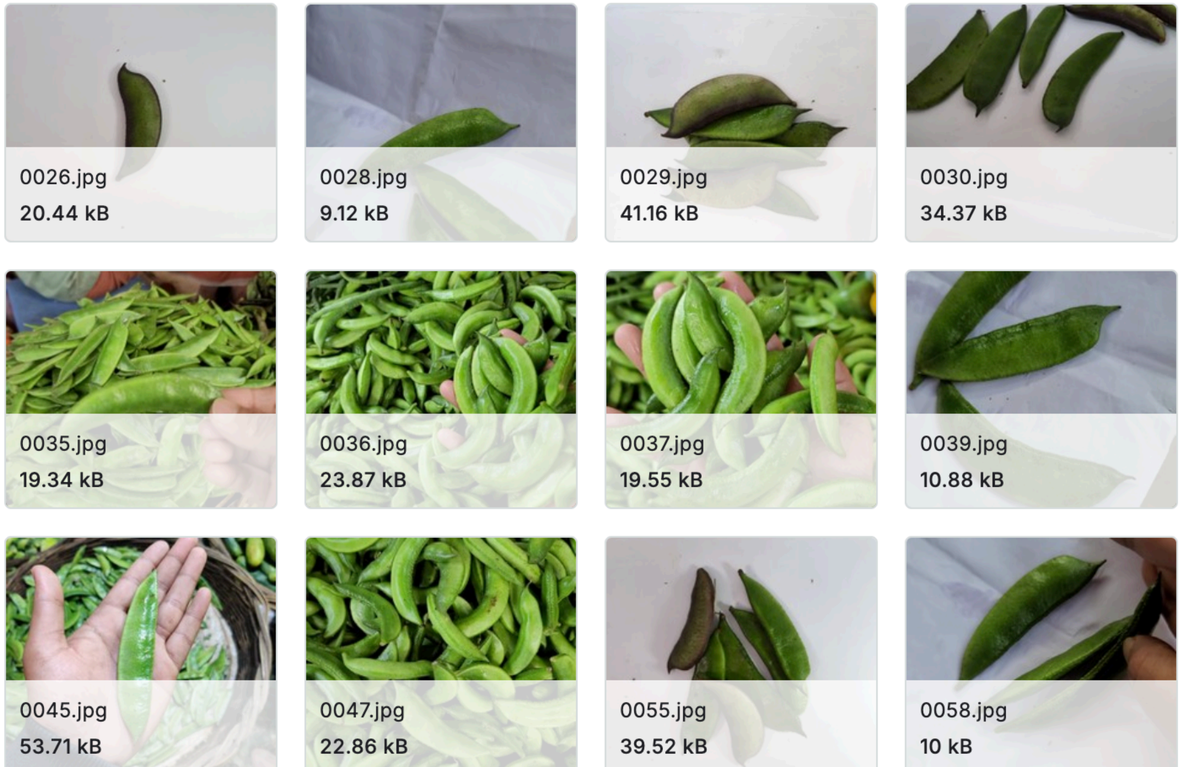
Exercise

# Explore a Dataset

## ❌ Unstructured Text

| Category<br>Label | Message<br>Text |
| --- | --- |
| ham 87%<br>spam 13%<br>Other (1) 0% | 5158<br>unique values |
| | work a... |
| spam | Your free ringtone is waiting to be collected. Simply text the password "MIX" to 85069 to verify. Ge... |
| ham | Watching telugu movie..wat abt u? |
| ham | i see. When we finish we have loads |

## ❌ Images



Exercise

# Explore a Dataset

- Find a dataset in an area you are interested in. Sports, politics, health, music, etc.
- Spend some time looking at it in a tabular format. Try some spreadsheet operations.
- Come up with 3 statements about it that you think might be true.

14 Jan 2025 · Gyan Lakhwani · gyanlakhwani@gmail.com · Department of Design, DTU

Exercise

# Explore a Dataset

- Find a dataset in a
  Spe
  operations.
  p with 3 statements about it that you think might be true.

WITH AI!!!!

14 Jan 2025 · Gyan Lakhwani · gyanlakhwani@gmail.com · Department of Design, DTU