Information Visualization

Daniel Gonçalves, Sandra Gama

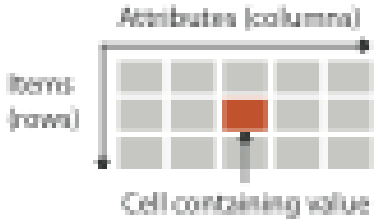
02

Data Abstraction

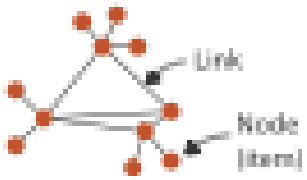
Dataset Types

Dataset Types

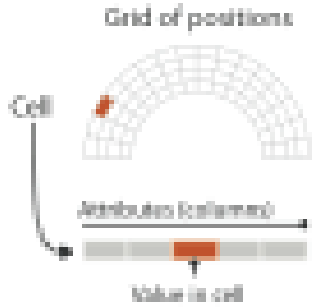
Tables



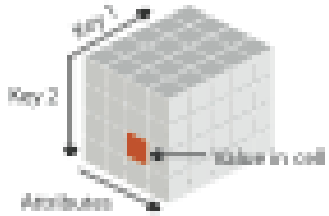
Networks



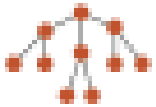
Fields (Continuous)



Multidimensional Table

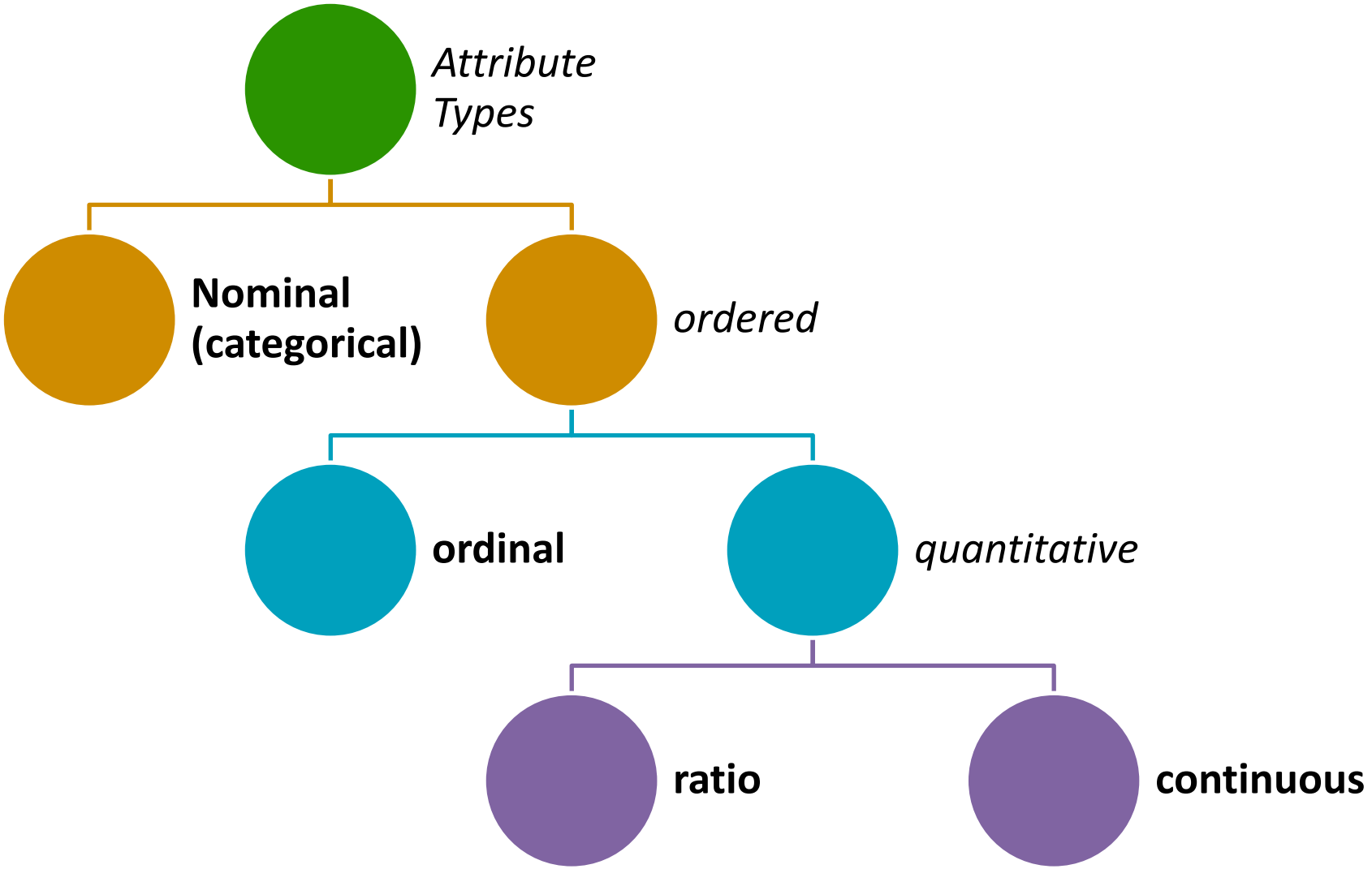


Trees



Geometry (Spatial)

Attribute Types

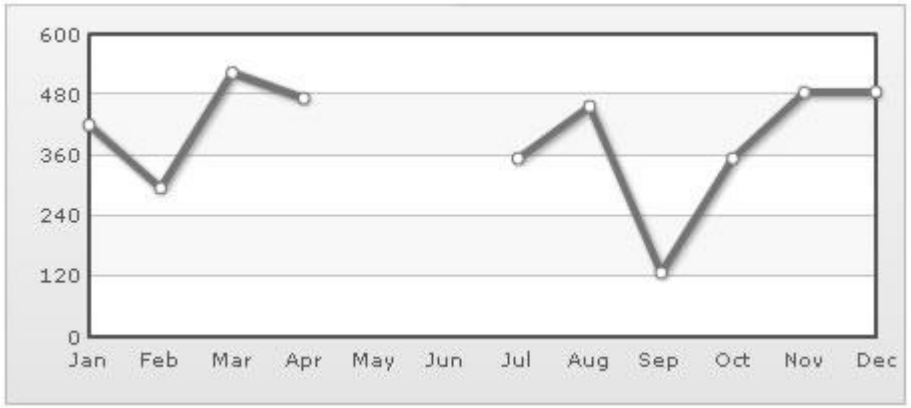


Game PIN

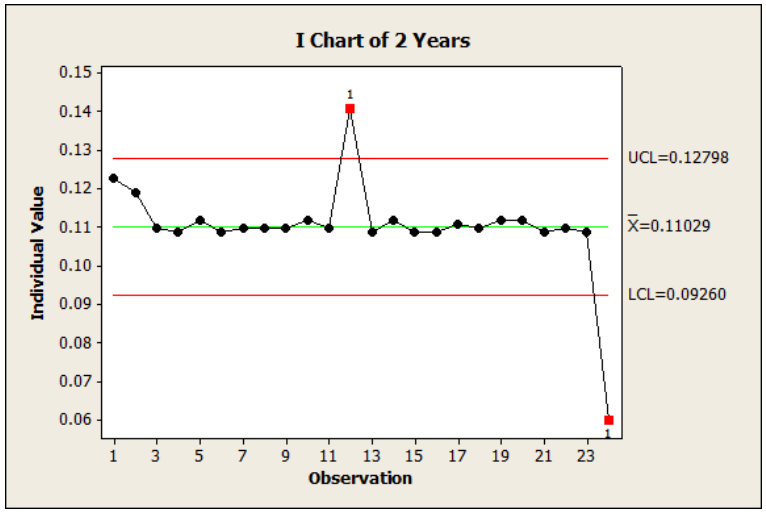
Enter

Data Quality Problems

Easier to spot:



missing values



outliers

Derived Attributes

Cost (2021) = Cost (1921) * inflation

Hamburgers per capita = hamburgers / people

KEEP CALM

TELL ME WHAT TO DO

Tell me what to do!

Derived attributes for this dataset!

Dataset: macroeconomics

Attributes:

- **Imports (in M€)**
- **Exports (in M€)**

Tell me what to do!

Derived attributes for this dataset!

Dataset: weigth of a set of people over time, since Jan 3, 2022

Attributes:

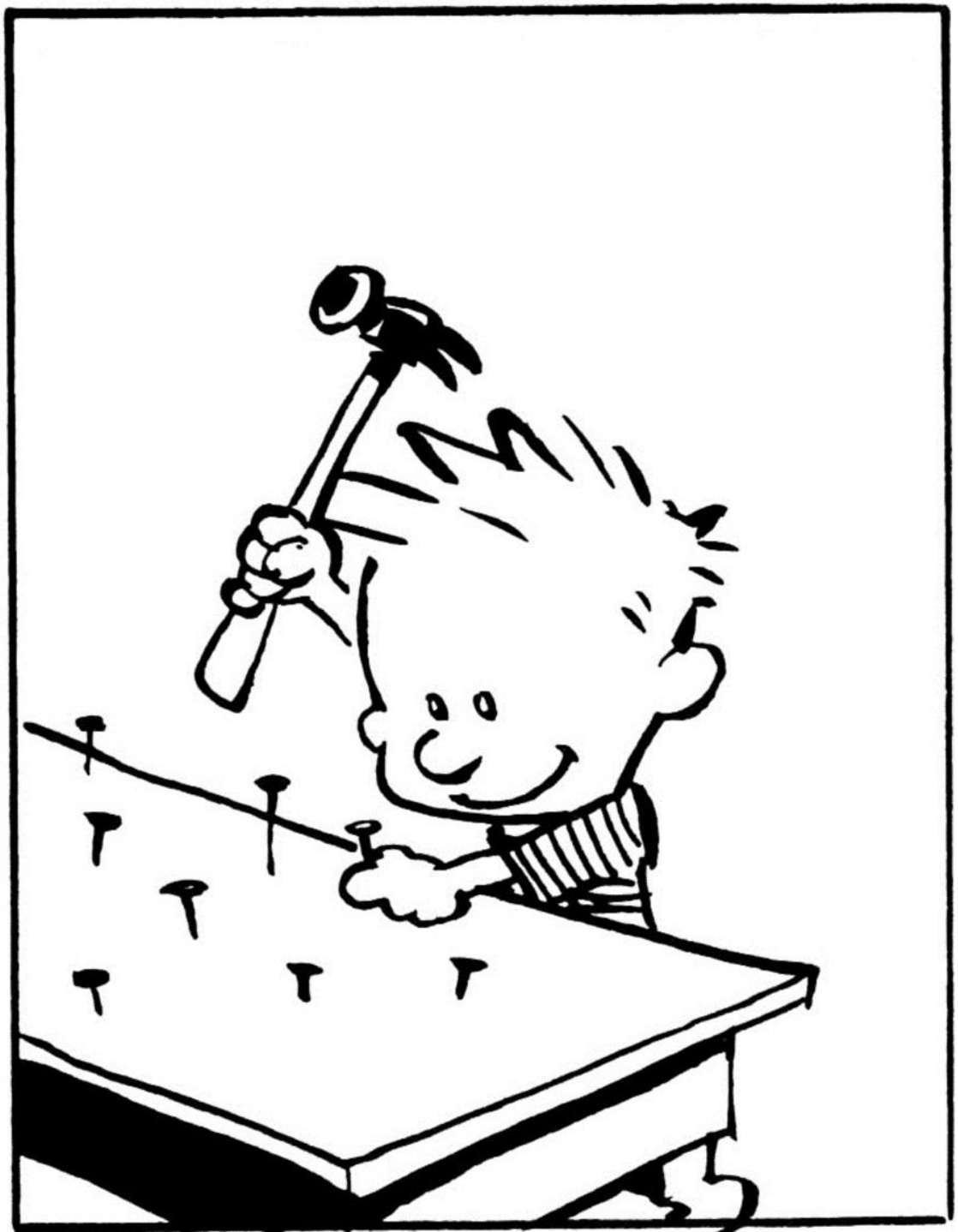
- **personID**
- **weight (in Kg)**
- **week (the date of the Monday of each week)**

KEEP CALM

TELL ME WHAT TO DO


THERE WILL
NOW BE A
BRIEF
INTERMISSION

10:00




THERE WILL
NOW BE A
BRIEF
INTERMISSION

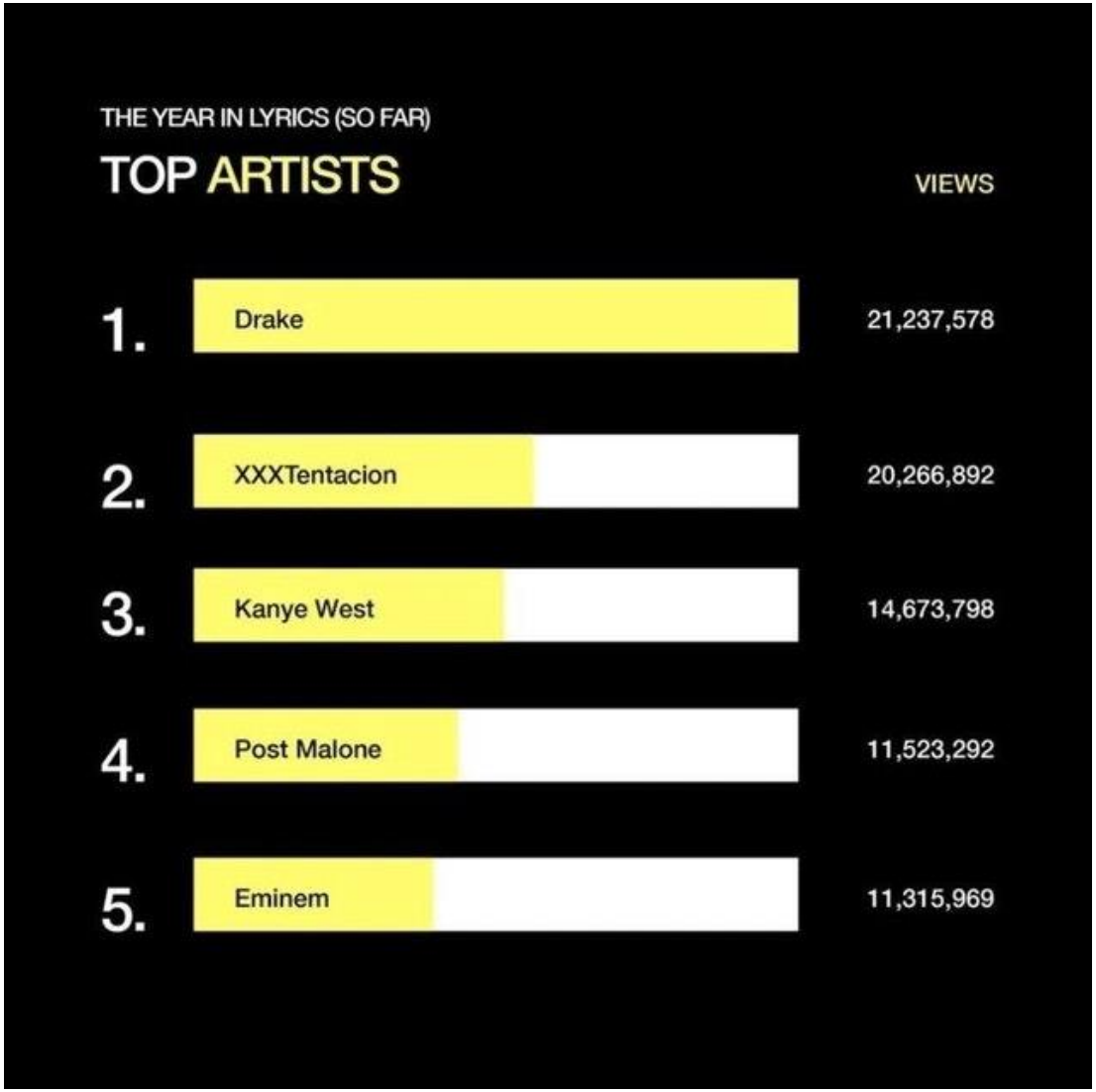

Direct
Poll

Get Ready…

http://etc.ch/dCHw

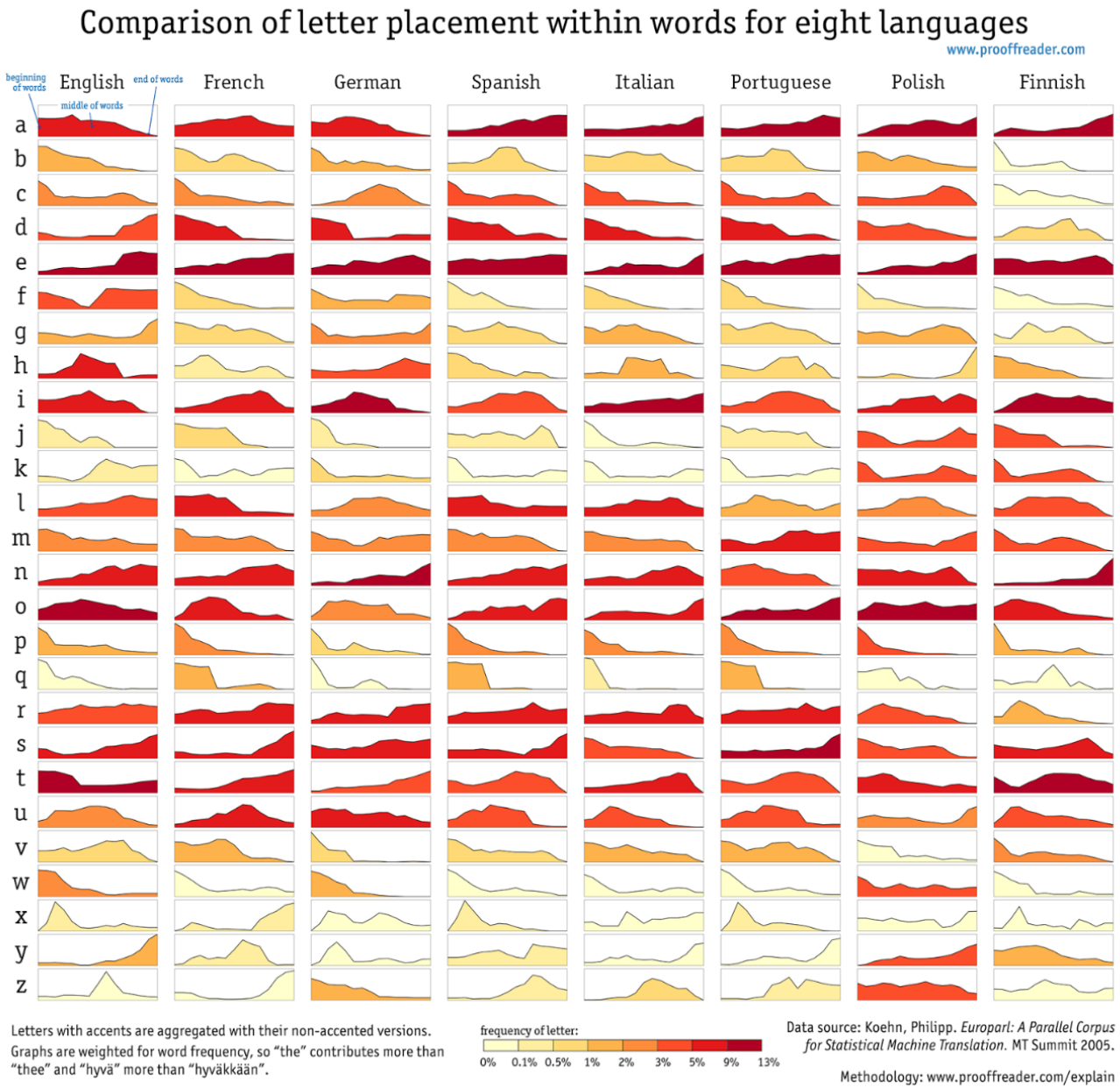

GOOD or EVIL?

Exibit A


THE YEAR IN LYRICS (SO FAR)
TOP ARTISTS
VIEWS
1.
Drake
21,237,578
2.
XXXTentacion
20,266,892
3.
Kanye West
14,673,798
4.
Post Malone
11,523,292
5.
Eminem
11,315,969

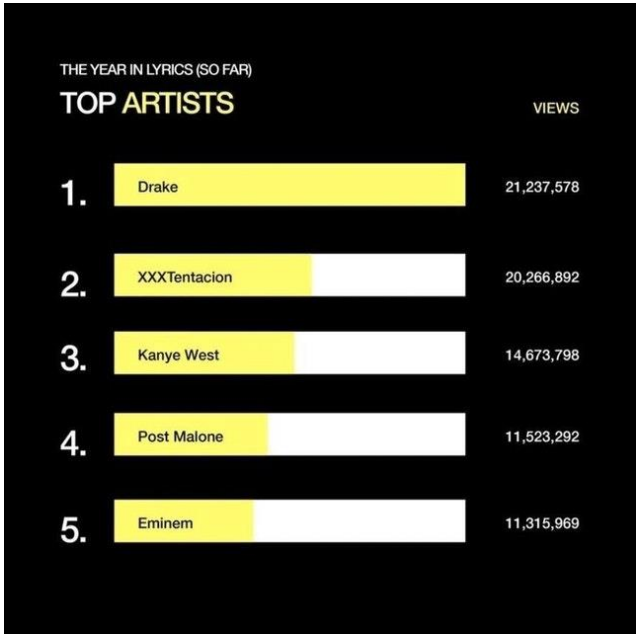
Exibit B

http://www.prooffreader.com/2014/07/comparison-of-letter-positions-in-eight.html

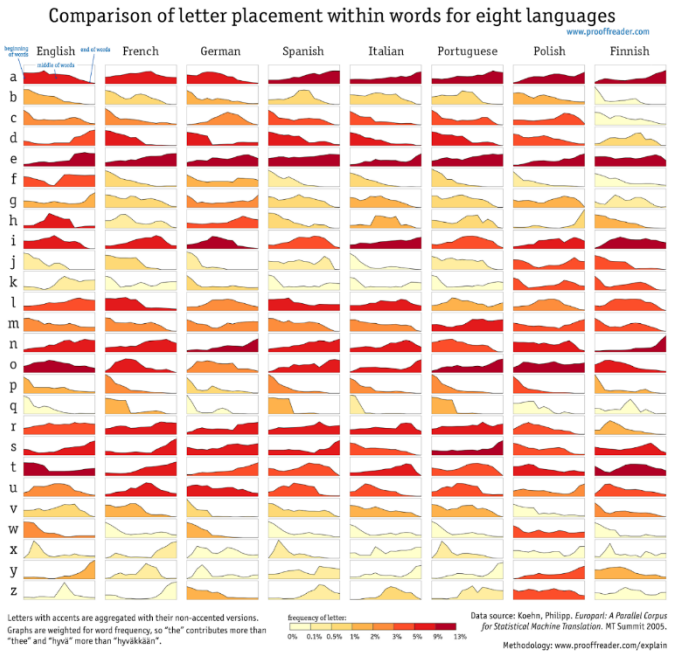


All Together Now!

A

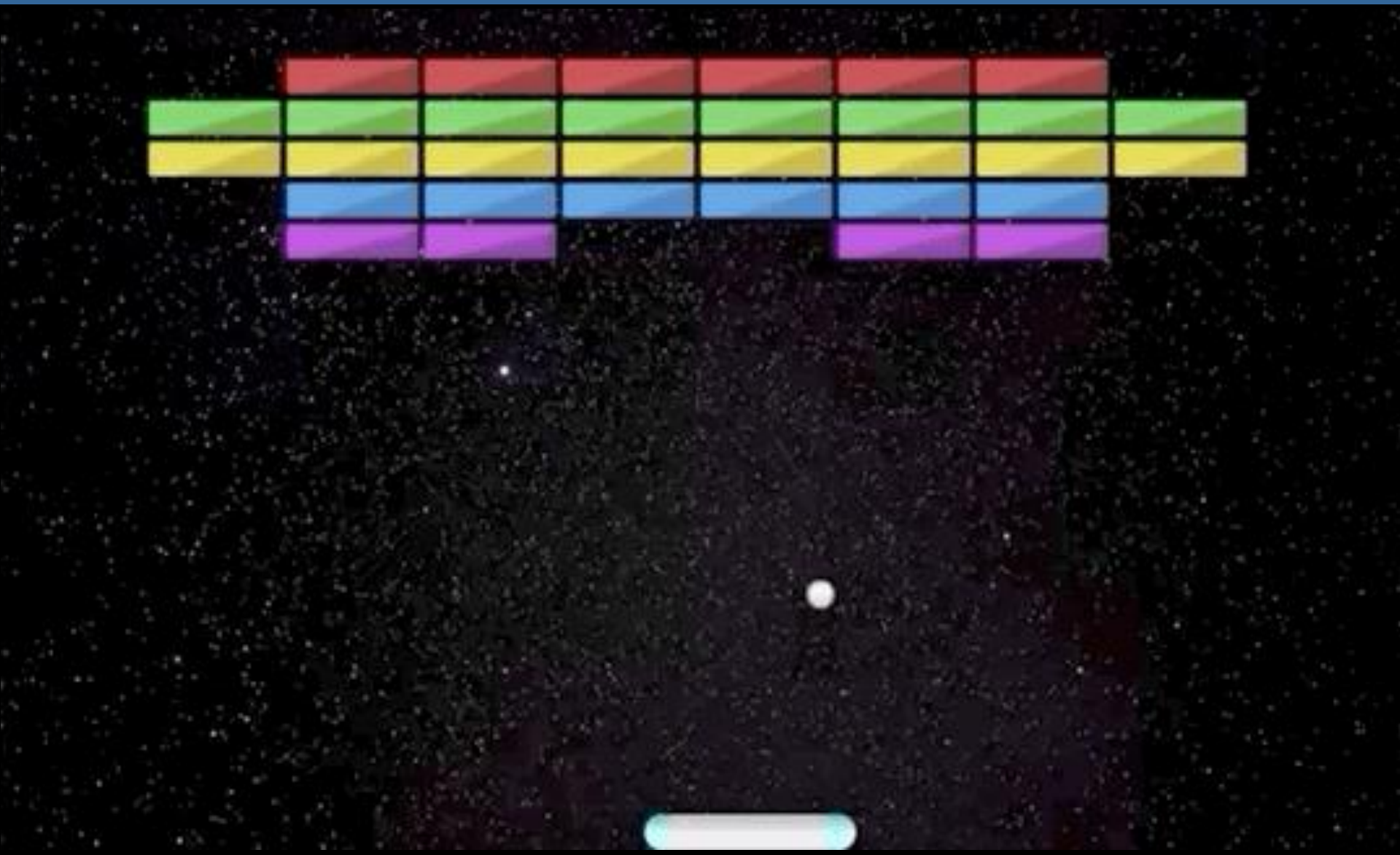


B



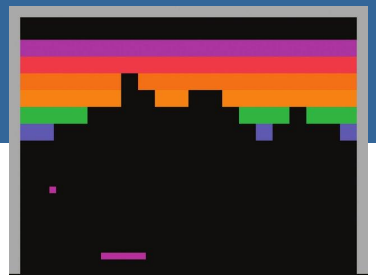
Direct Poll

Breakout!



Breakout!

Example: movie dataset



Download this Dataset (link in the form)

Tell us:

How would you characterize the dataset?

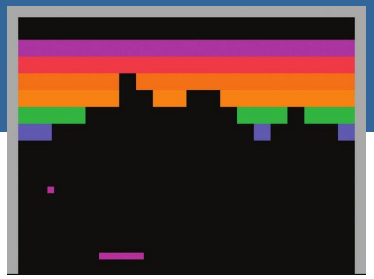
Are there data quality problems? How would you fix them?

What are the different attribute types?

What derived attribute do you think makes sense here?

Three questions you could answer with this dataset

Time to Break Out



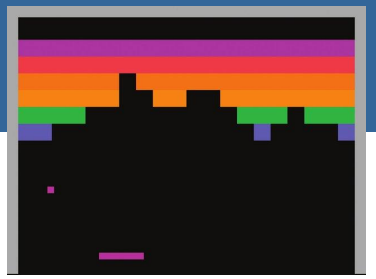
Discuss among yourselves

Fill in the questionnaire

https://bit.ly/3QiVJok

Groups reporting in 30min

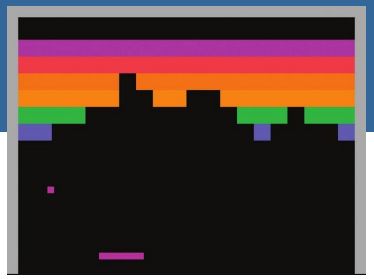
Be quick!



30:00

https://bit.ly/3QiVJok

Data Checklist!



For each dataset:

static or dynamic?

Type: table, network, field, geo?

What are the items, attributes, relationships?

time-based?

For each attribute:

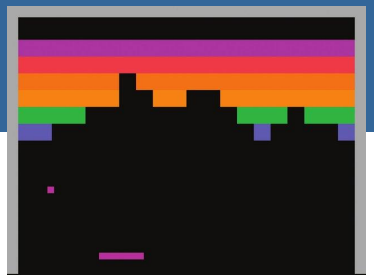
Meaning (semantics)

attribute type (nominal, ordinal, continuous, ratio)

sequential, diverging or cyclic?

hierarchic? (what levels of the hierarchy are relevant?)

Example: movie dataset



Dataset…

Static (JSON file)

Type: Table

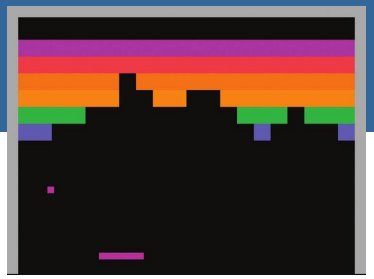
Items: movies

Relationship: Movies that belong to the same collection

Attributes: see next

Time-Based: yes

Example: movie dataset



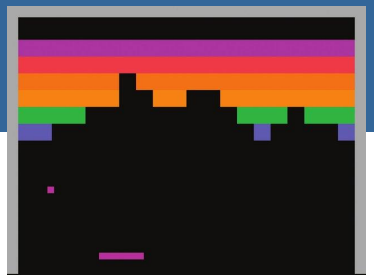
collection

Meaning: whether the movie belongs to a collection, a franchise

Type: nominal

Hierarchic: no

Example: movie dataset



budget

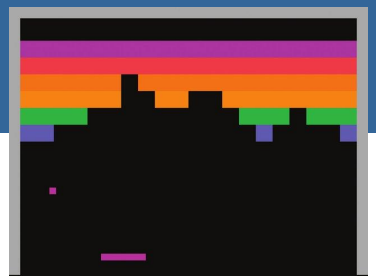
Meaning: how much it cost to produce

Type: ratio

Sequential

Hierarchic: no

Example: movie dataset



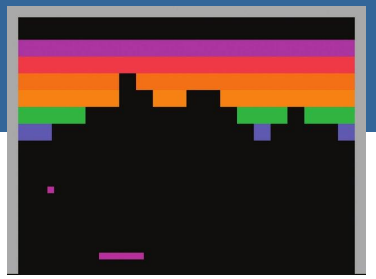
year

Meaning: the year the movie was made in

Type: quantitative

Hierarchic: Yes

Example: movie dataset



Genre

Language

Title

production_companies

release_date

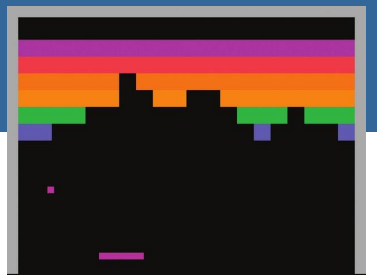
revenue

Runtime

rating

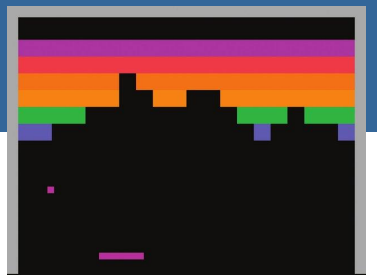
Meaning
Type
Sequential: ?
Hierarchic: ?

Data Quality



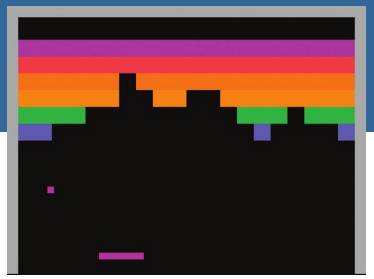
Missing Values?

Outliers?



Derived attribute??

Example: movie dataset



Budget adjusted for inflation!

Meaning: cost in USD at current prices (inflation adj.)

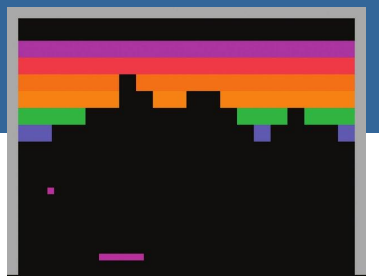
Type: ratio

Sequential

Hierarchic: no

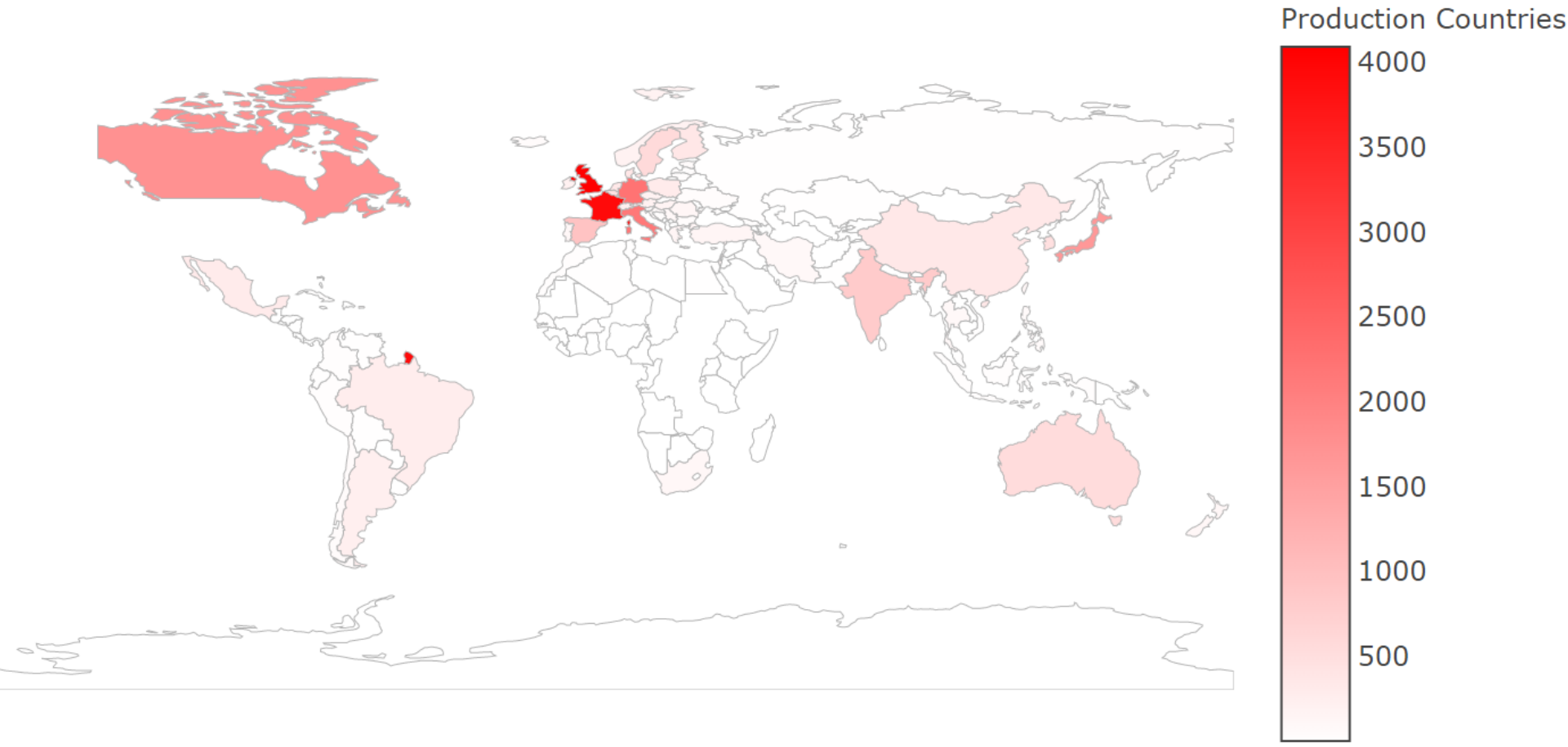
Other: profit, year

Questions we could answer…



Time to present yours!

The US rules movie production. Who's next?

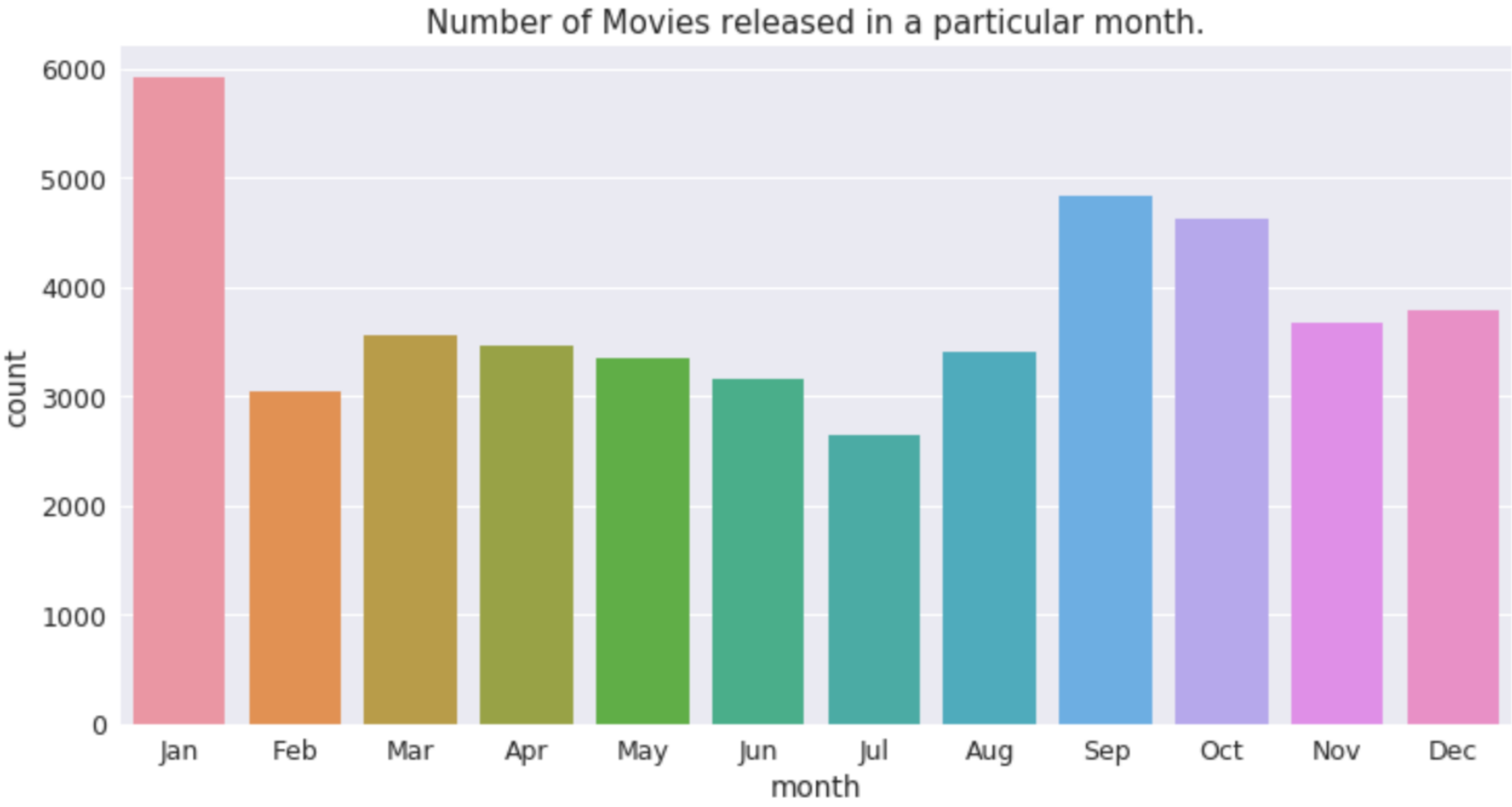


What are the most successful franchises?

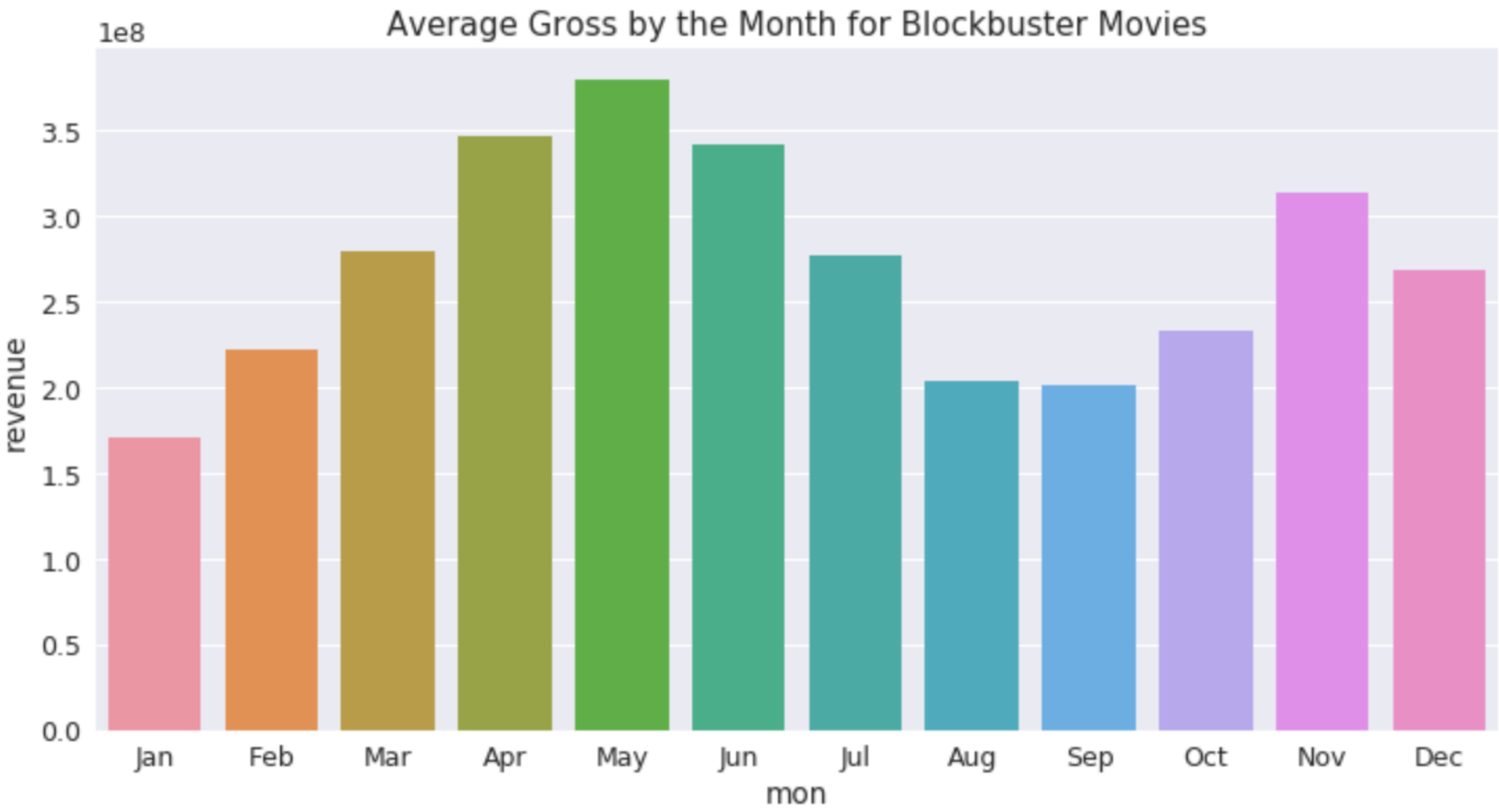
	belongs_to_collection	count	mean	sum
646	James Bond Collection	26	2.733450e+08	7.106970e+09
473	Friday the 13th Collection	12	3.874155e+07	4.648985e+08
976	Pokémon Collection	11	6.348189e+07	6.983008e+08
552	Harry Potter Collection	8	9.634209e+08	7.707367e+09
540	Halloween Collection	8	3.089601e+07	2.471681e+08
29	A Nightmare on Elm Street Collection	8	4.544894e+07	3.635916e+08
1317	The Fast and the Furious Collection	8	6.406373e+08	5.125099e+09
1432	The Pink Panther (Original) Collection	8		
1160	Star Wars Collection	8		
977	Police Academy Collection	7		

	belongs_to_collection	count	mean	sum
552	Harry Potter Collection	8	9.634209e+08	7.707367e+09
1160	Star Wars Collection	8	9.293118e+08	7.434495e+09
646	James Bond Collection	26	2.733450e+08	7.106970e+09
1317	The Fast and the Furious Collection	8	6.406373e+08	5.125099e+09
968	Pirates of the Caribbean Collection	5	9.043154e+08	4.521577e+09
1550	Transformers Collection	5	8.732202e+08	4.366101e+09
325	Despicable Me Collection	4	9.227676e+08	3.691070e+09
1491	The Twilight Collection	5	6.684215e+08	3.342107e+09
610	Ice Age Collection	5	6.433417e+08	3.216709e+09
666	Jurassic Park Collection	4	7.578710e+08	3.031484e+09

Summer blockbusters?



Summer blockbusters?



Breakout Over!

QUESTIONS?