

HCI Assignment 2025-2026

Prof. Fabio Ciravegna
Dipartimento di Informatica
Università di Torino
fabio.ciravegna@unito.it



Anime!!!



Assignment Objective

- Aim is to analyse and visualise a large set of data about anime
 - Using
 - Python
 - Pandas
 - Seaborn
 - GeoPandas
 - Folium
 - etc.

Data: over 130 million data points

- **Anime:**

- 28,955 anime details

- **Characters:**

- 209,963 characters details
- 236,816 characters and roles
- 37,080 nicknames of characters

- **Actors:**

- 20,465 alt names for people (e.g. actors)
- 458,091 roles of people in anime (e.g. singing the theme song)
- 76,699 people details

- **Users:**

- 337,155 user profiles
- 124,298,357 ratings given by users
- 105,249 user recommendations
- 28,955 user stats (ratings, number of completed series, etc.)
- 4,178,747 likes of users

characters_df

✓ [6] 12ms

209,963 rows x 7 cols

character_id	url	name	name_kanji	image	favorites	...
None Count	2	None Count	2	None Count	55480	None Count
	209507		162788		129167	
Unique values		Unique values		Unique values		Unique values
0	280386.0	https://myan...	Envi Mel Cha...	エンヴィ・メル...	https://cdn...	0.0
1	280354.0	https://myan...	Eleven	イレブン	https://cdn...	0.0
2	280353.0	https://myan...	Studio	スタジオ	https://cdn...	0.0

details_df

✓ [7] 44ms

28,955 rows x 29 cols

mal_id	title	title_ja	url	image...	type	...
None Count	0	None Count	0	None Count	0	None Count
	28953		27773		28955	
Unique values		Unique values		Unique values		Unique values
119						
TV	29.1%	Finish				
Movie	17%	Not yet				
Other	54%	Current				

person_details_df

✓ [21] 16ms

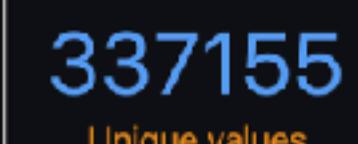
76,699 rows x 10 cols

	person_m...	url	website_...	image_...	name	given_na...	fa...
	None Count 0	None Count 0	None Count 59264	None Count 0	None Count 2	None Count 30253	None C
	 76594 Unique values		nan 77.3% https://stra... 0% Other 22.7%	https://cdn... 62.9% https://cdn... 0% Other 37.1%	 74509 Unique values	nan 39.4% 大輔 0.1% Other 60.4%	nan 佐藤 関 智一 智和 杉田
0		1 https://myan...	NaN	https://cdn...	Tomokazu Seki		関
1		2 https://myan...	https://agrs...	https://cdn...	Tomokazu Sug...		杉田

profiles_df

✓ [29] 653ms

337,155 rows x 10 cols

	username	gender	birthday	location	joined	watching	co...
	None Count 1	None Count 170876	None Count 215826	None Count 0	None Count 1676	None Count 1678	None C
	 337155 Unique values	nan 50.7% Male 35.8% Other 13.5%	nan 64% 2003 0.1% Other 93.9%	Japan 29.2% United Sta... 19.4% Other 51.4%	nan 0.5% Sep 30, 20... 0.2% Other 99.3%	0 11.3% 1 8.6% Other 80.1%	0 0 0
0							
1							

stats_df

✓ [17] 35ms

28,955 rows x 27 cols

None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0	None Count	0
59356	7	146	4	20	20	56036	21	770	8	29	113	2928	451	14953	302	349	6472	7040	781	99500	150	572	9710	2910	2910
7040	781	99500	150	572	9710	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910
7040	781	99500	150	572	9710	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910
7040	781	99500	150	572	9710	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910	2910

Assignment HCI for 2025-2026

1 Introduction

This assignment is primarily concerned with applying the ideas that are being presented in the module on methods for accessing the Web and making sense of its content. In providing a solution, you are required to use the methods and techniques taught in the module.

You are given a set of data about animes over the years.

The data is provided as a set of CSV files.

You are requested to create a system able to support analysis and visualisation by fans and journalists (experts).

2 Requirements

Functionalities:

The goal of the work is to provide the possibility for fans and journalists to access data about anime. You must design a (set of) Jupiter Notebook(s) that clean, analyse and visualise the data contained.

3 The Data Provided

You are provided with the following data files:

- Characters.csv: 209,963 characters details
- Character_anime_works.csv: it contains 236,816 characters and roles in the database
- Character_nicknames.csv: 37,080 nicknames of characters

Note!

- There is a considerable requirement on providing a professional documentation for your project
 - so make sure to take the time to document the project professionally
 - that means your Jupyter Notebook(s) must tell a story in appropriate markdown cells

Telling a Story with Data

- Data Alone Doesn't Speak
 - A dataset with millions of records is just numbers.
 - Without interpretation, it overwhelms rather than informs.
 - A story gives shape and meaning.
- Stories Connect with Humans
 - Humans remember stories better than facts.
 - A well-told data story makes insights stick.
 - Example: “Film lengths doubled in 50 years” is more memorable than a raw histogram
- A Story = A Message
 - The reader should be able to summarise your analysis in one sentence.
 - Without that, you've only shown charts — not insights.

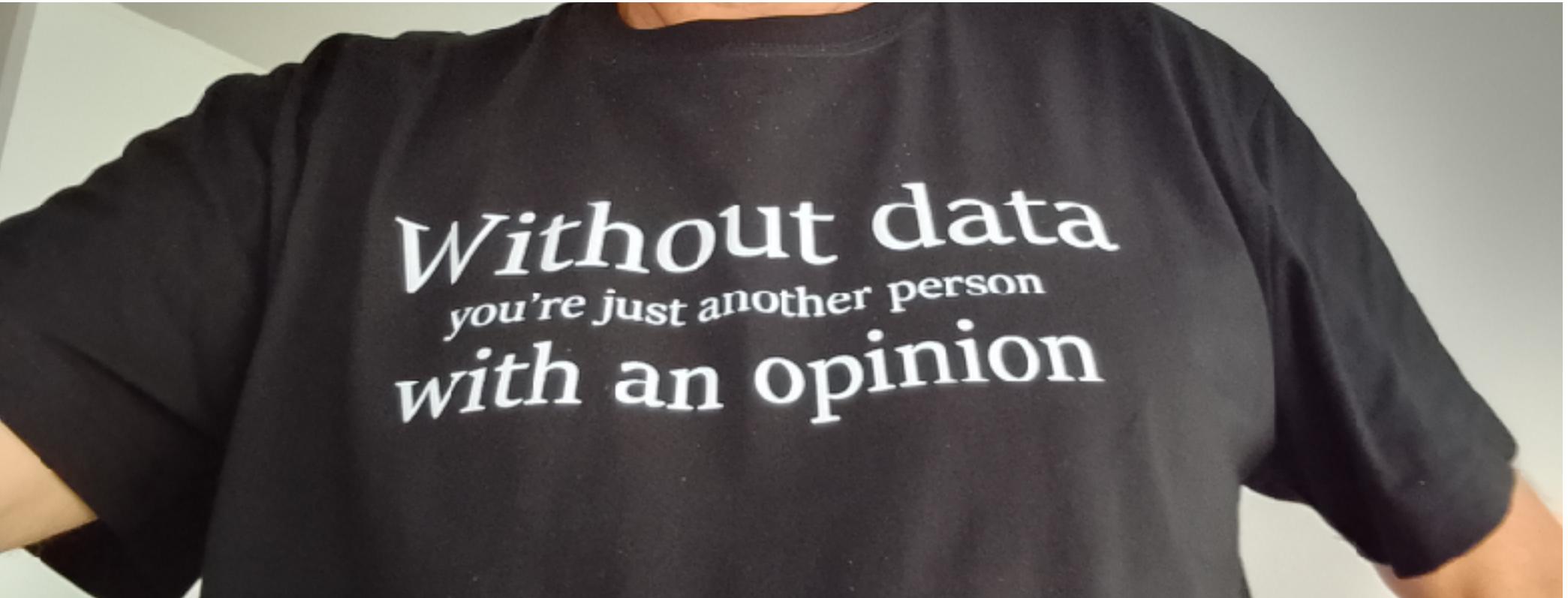
How to Create a Data Story

- Step 1 – Explore the Dataset
 - Look at distributions, outliers, trends.
 - Ask: What stands out? What surprises me?
- Step 2 – Identify a Message
 - Example: “Genres popularity changes over time and so does their characteristics. Comedies for example are shorter than dramas, and this gap is increasing.”
 - **Make sure it uses the large-scale dataset — not just a small slice.**
- Step 3 – Support with Visuals
 - Use charts that highlight your message clearly.
 - Annotate, title, and highlight so the audience “gets it” immediately.
- Step 4 – Provide Context
 - Compare across time, regions, categories.
 - Give the reader a frame of reference.
- Step 5 – Refine for Clarity
 - Keep it simple: each chart should answer one question.
 - Your story should flow logically from introduction → evidence → conclusion.

ALWAYS support ALL your claims
with data



UNIVERSITÀ
DI TORINO



What makes a good data story?



UNIVERSITÀ
DI TORINO

- Big Picture, Not Small Slice
 - Use the whole dataset (millions of records) to give context.
 - Avoid narrowing to a tiny, unrepresentative subset (e.g. only Oscar winners).
 - Example: “Interest in genres has changed over 80 years” → story at scale.
- A Clear Message
 - The story must be summarisable in one sentence.
 - Example: “Independent films are shorter but have grown more numerous after 2000.”
 - If your audience needs a paragraph to explain, the story is not sharp enough.

What Makes a Good Data Story? (2)



- Patterns, Trends, and Outliers
 - A good story reveals:
 - Patterns (e.g. rise of superhero movies since 2010).
 - Trends (film budgets increasing faster than box office revenue).
 - Outliers (one ultra-long experimental film, record-breaking blockbuster).
 - Context and Comparisons
 - Place data points in historical or category context.
 - Compare groups: by year, by country, by genre, by studio.
 - Example: “Italian cinema peaked in the 1960s, but Korean cinema exploded after 2000.”

What Makes a Good Data Story? (3)



UNIVERSITÀ
DI TORINO

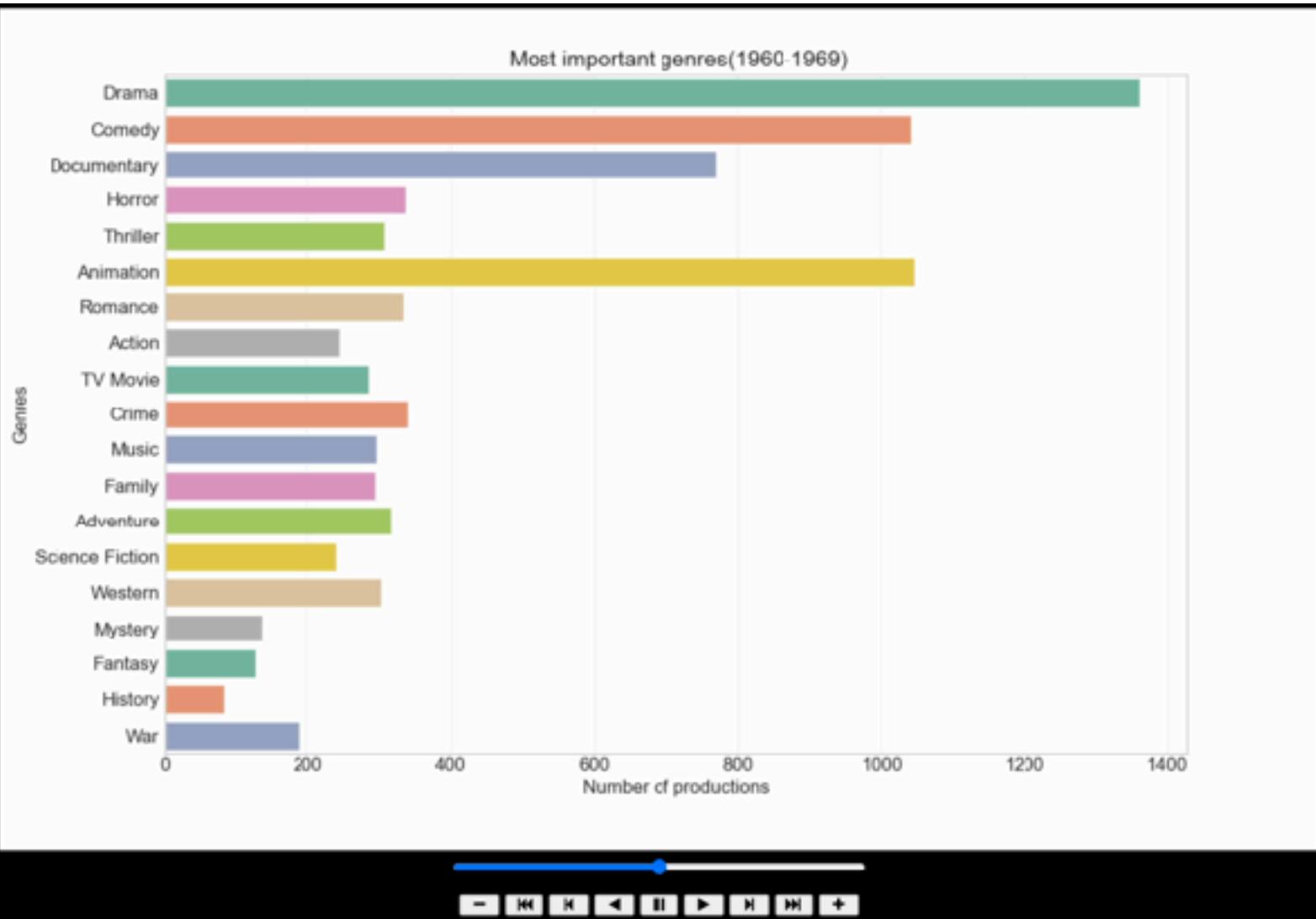
- Audience Focus
 - The story is not about data exploration for its own sake.
 - It must answer:
 - “What insight does this dataset give about the world?”
 - Example:
 - Instead of “distribution of runtimes,”
 - say: “Audiences are watching longer films today than 30 years ago.”
- Visualisation as Narrative
 - Start broad → zoom in → highlight insight.
 - Avoid charts that require the audience to work hard.
 - Use design choices (titles, highlights, annotations) to make the story clear.

What makes a good visualisation?



- Clarity
 - The message should be grasped immediately, without deep reflection.
 - A good visualisation = a headline, not a puzzle.
 - The audience should not need to calculate, mentally align scales, or decipher colour codes.
 - Remove distractions: avoid unnecessary 3D, exotic chart types, or excessive annotations.
 - Rule of thumb: If someone has to “study” your chart, it’s not clear enough.

Good Vs Not Good

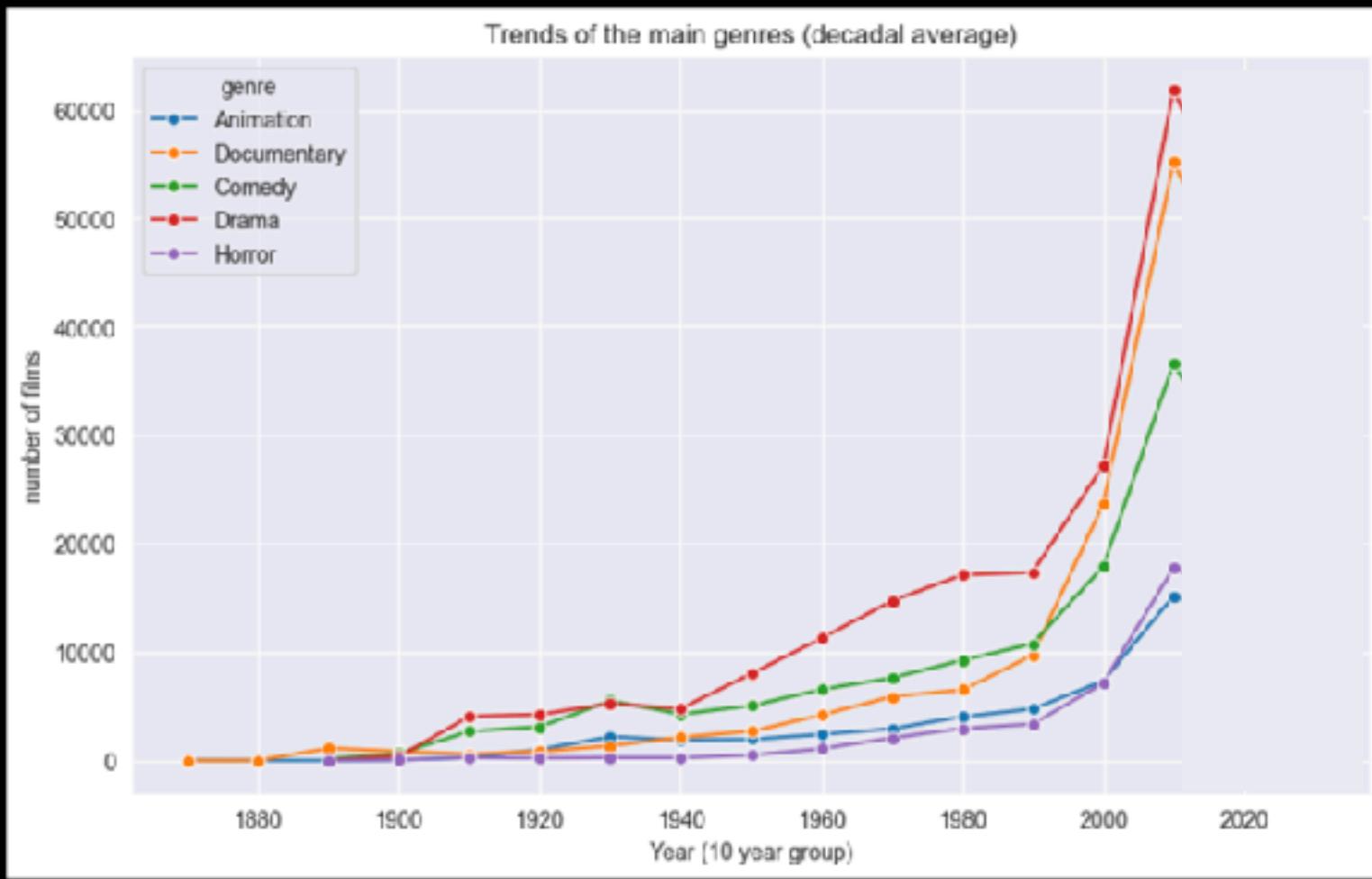


it requires a lot of thinking to get the message and to reason across the different time windows

The intended use was to understand that documentary becomes more popular than Comedy over time - it nearly reaches Drama in the end

Good

Top genres: Drama, Documentary, Comedy, Animation, Horror



the same message is very clear: Documentary overtook Comedy in 1990-2000



UNIVERSITÀ
DI TORINO

However here is a problem:
why are the number of films
decreasing in 2020-2030?

The decade is incomplete

- the data was until mid 2024
- so the last decade is not representative
- Solution?
 - put a note
 - remove the last decade
 - use the projection of the expected value at end of 2030

Even Better

Average Rotten Tomatoes Score by Genre Over Time

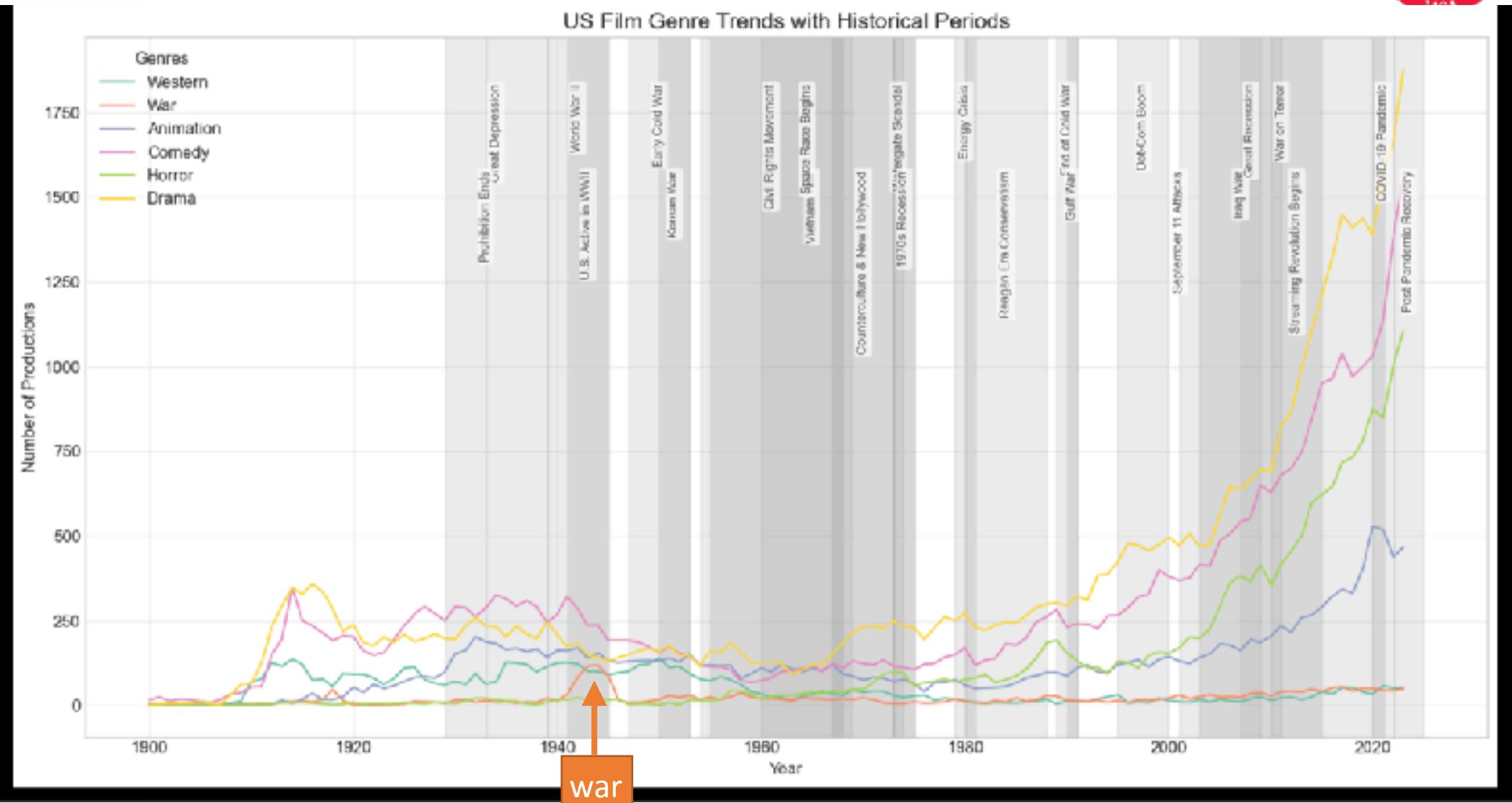


Note: this is not the same graph - the data is different but I hope you get the gist

Better

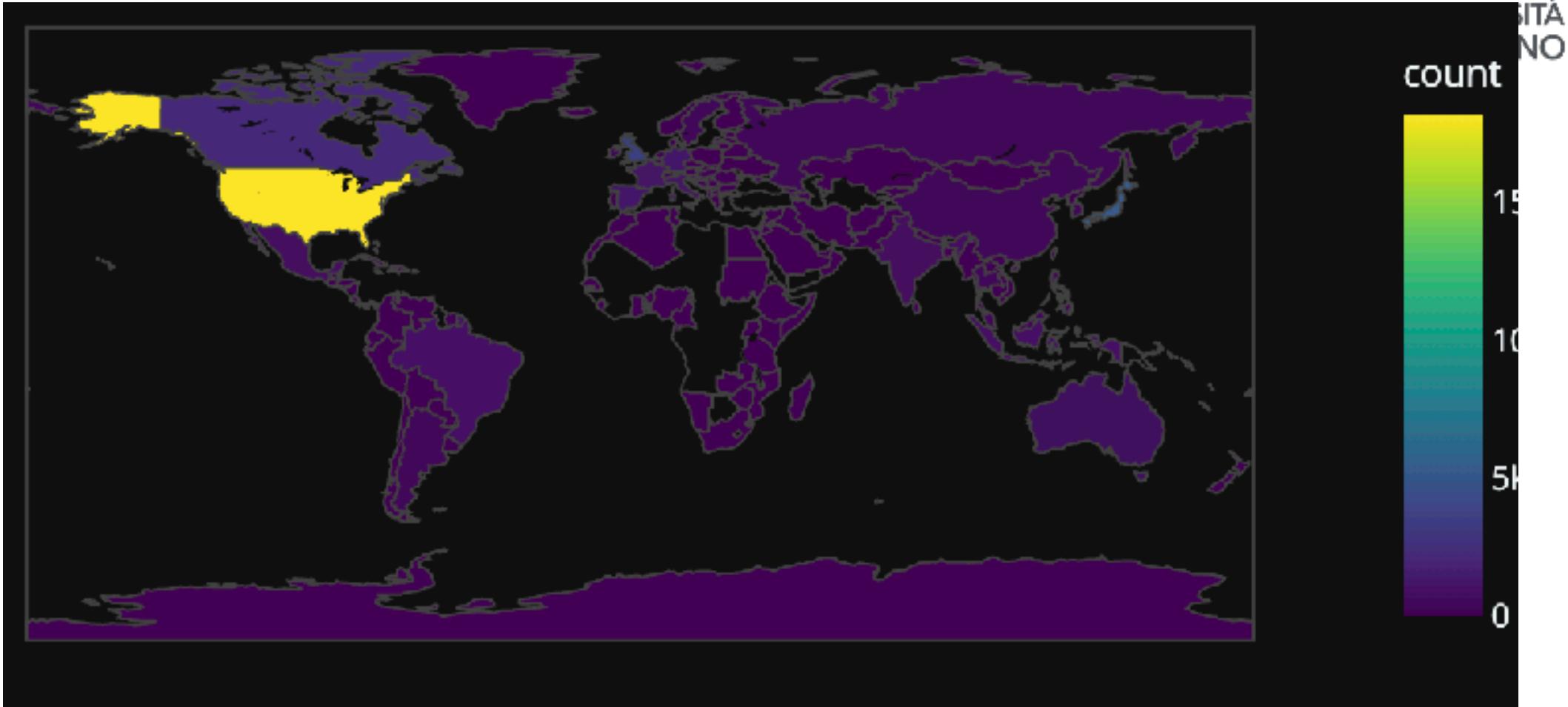


US Film Genre Trends with Historical Periods



the addition of the most important event (grey areas) allows to reason e.g. what happened to specific genres such as war

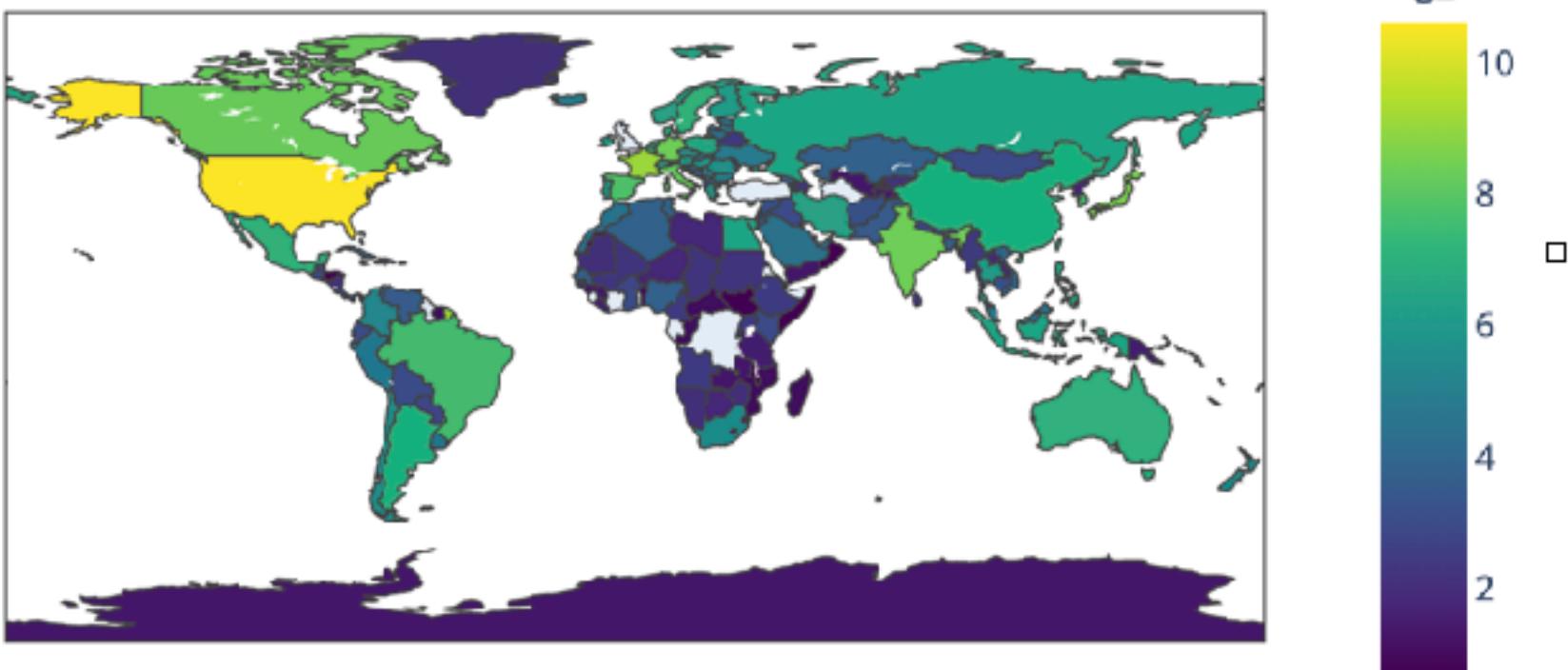
Bad



Number of films produced by Country

- bad because you only see the USA - they are an outlier so they make the rest of the map pointless

Number of Movies Produced by Country (Log Scale)



Using a log scale reduces the peak of the outlier and hence

- the rest of the map is meaningful
 - the predominance of the USA is maintained

Good Viz (ctd)

- Accuracy
 - Always represent data truthfully.
 - Avoid distorted scales (e.g. truncated y-axes that exaggerate differences).
 - Keep proportions correct (a pie slice of 30% must look bigger than 10%).
 - Don't mislead with visual tricks (e.g. 3D depth making values look larger).
- Audience Focus
 - Design for the reader, not for yourself.
 - Ask: What does my audience need to know?
 - Provide the right level of detail:
 - Executives → big picture, one key message.
 - Analysts → detailed breakdown, ability to drill down.
 - Use annotations, highlights, or colour emphasis to guide attention.

Questions?

