The Impact of Datascapes (on Metadata Analytics)

Illustrated by the Virtual Census of Characters in the Japanese Visual Media Graph

Martin Roth and Zoltan Kacsuk

Deutsche Forschungsgemeinschaft
German Research Foundation

第四回EAJS日本会議 // The 4th EAJS Japan Conference
東北大学 // Tohoku University
20-21 September 2025

Outline of the presentation

HOCHSCHULE DER MEDIEN

1. Background: Metadata analytics and the data we rely on

2. Datascapes: What they are and why they matter

3. Application: The datascapes lens and the virtual census of anime characters

HOCHSCHULE
DER MEDIEN

Background:
Metadata analytics
and the data we rely on

Metadata analytics

HOCHSCHULE DER MEDIEN

“The aim of metadata analytics is to generate new insights into phenomena and processes in the domain covered by specific metadata sets by subjecting these sets to quantitative methods of analysis at various levels of granularity, exploiting their qualities as metadata, namely "structuredness," scale, and interconnectedness.”

(Roth and Kacsuk 2025)

Roots of metadata analytics

HOCHSCHULE DER MEDIEN

- **Bibliometrics** and **Scientometrics**
  - Analysis of (academic) publications based on citation metadata and bibliographic information
- **Distant Reading** (Franco Moretti)
  - textual analysis of large text corpuses of literary works, as opposed to analysis of canonical works
- **Cultural Analytics** (Lev Manovich)
  - computatoinal analysis of large quantities of (often visual) data and visualisation of the results of such analysis, including computational feature extraction

Shared problem

HOCHSCHULE DER MEDIEN

Rooted in distant reading, both cultural analytics and metadata analytics share three foundational assumptions / starting points:

1. The universe of cultural products is far larger than the scope any individual researcher or research team can cover manually / qualitatively
2. **We are used to study canonical works**
3. by considering a broader scope of works quantitatively and with new approaches, we may find new patterns and understand culture better

We rely on community generated data

HOCHSCHULE DER MEDIEN

- Databases by enthusiast communities are often the **most up-to-date online resource for checking information**
- Japanese Visual Media Graph project aim:
  Make these databases available for **large-scale quantitative research (=metadata analytics), in collaboration with the communities**
  - Funding: German research foundation, 6 years duration

Source databases

HOCHSCHULE DER MEDIEN

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture
- **The Visual Novel Database** (VNDB): Focused on visual novel games only
- **Anime Characters Database** (ACDB): Focus on one aspect of the domain

Other databases:

- **Wikidata:** Not focused on Japanese visual media
- **Media-Arts Database:** Collects information on manga, animation, games and media art from institutions, creators and publishers in Japan

Source databases for the JVMG project

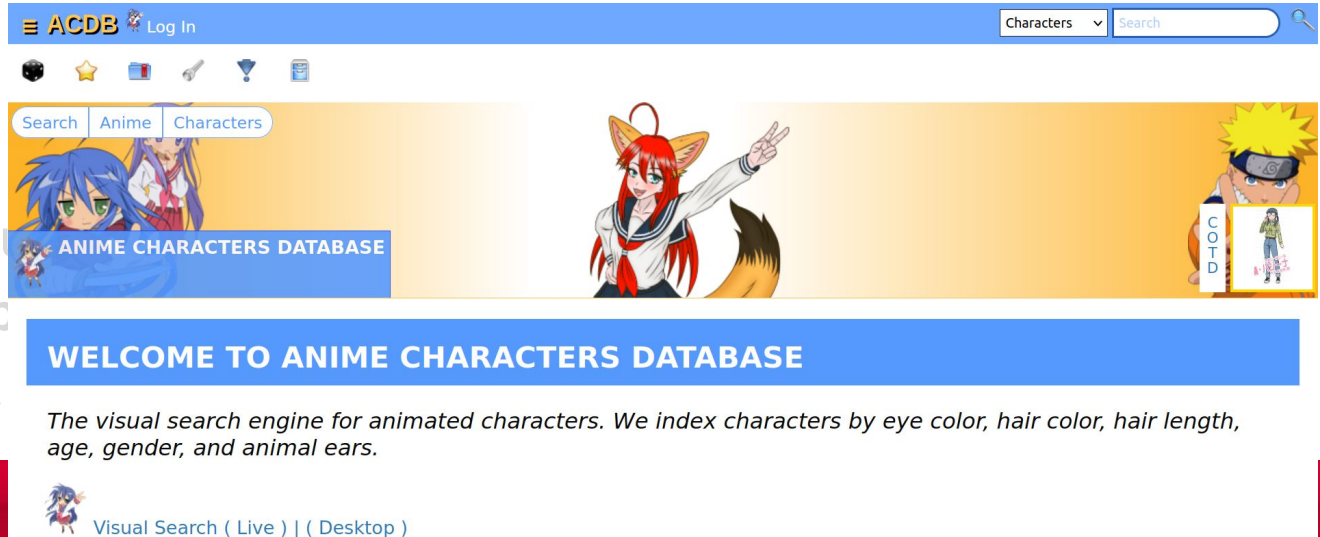
HOCHSCHULE DER MEDIEN

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture
- **The Visual Novel Database** (VNDB): Focused on visual novel games only
- **Anime Characters Database** (ACDB): Focus on one aspect of the domain

Other databases:

- **Wikidata:** Not focu
- **Media-Arts Datab** games and media



Scale

HOCHSCHULE DER MEDIEN

Enthusiast community	Works and media				Company	Characters	Work properties	Character properties	Involved people
ACDB	Work					Character	Work Tag	Character Tag	People
	10.207					107.369	1.088	4.051	5.557
AnimeClick	Animation Work	Comic Work				Character			Staff
	9.491	11.762				102.143			39.604
VNDB			Visual Novel	Release	Producer	Character	Tag	Trait	Staff
			28.190	71.349	10.394	90.077	2.585	2.777	21.164

Scale

Database	Works and media					Characters	
Wikidata	Anime titles	Manga series	Video game	Light novel & LN series		Anime character	Manga character
	4.467	13.871	47.192	867		3.788	2.990
Media-Arts Database	Anime titles	Anime items	Game items	Manga book series	Manga magazine issues		
	12.085	~135.000	~61.000	133.779	170.670		

HOCHSCHULE
DER MEDIEN

Datascapes:
What they are and why they matter

Databases/ data resources are datascapes

HOCHSCHULE DER MEDIEN

- Each data collection / database is always first and foremost a function of the interests of the communities working on them, in terms of granularity, focus, coverage and data completeness
- Each data collection / database is thereby a particular representation of a domain of interest (anime, characters, etc.) and a window into a particular view of the domains of interest by the community that curates it
- **Datascapes describe domains (and as such, may shape their user's perspective on a domain) and reflect their curating community (and as such, may speak to the relevances and practices in the community)**

Datascapes are a tool for critical inquiry

HOCHSCHULE DER MEDIEN

- Against the backdrop of critical data studies, adopting the datascapes perspective forces us to reconsider what metadata describe in a more nuanced way:
  - in what ways and to what degree do datascapes describe empirical reality?
  - in what ways do they describe their knowledge communities?
- When working with data, it is quite easy to fall into the trap of assuming objectivity due to having erased the trajectory or history of a particular piece of information
- **datascapes reiterate the importance of provenance tracking and the fact that there is no “uncooked data”, highlighting the fact that particular data representations of a domain may influence its perception**

Datascapes: what to look for?

HOCHSCHULE DER MEDIEN

- Categories: what categories are used to describe the domain?
- Measurement scales: what mode is used to describe any relevant aspect?
- Granularity: what resolution is applied to describing the domain / any of its relevant aspects?
- Level of precision: how accurate are the descriptions vis-a-vis their objects
- Focus: What parts of the domain are focused on, what parts are left out?
    - Partially missing populations
    - Completely missing populations

HOCHSCHULE
DER MEDIEN

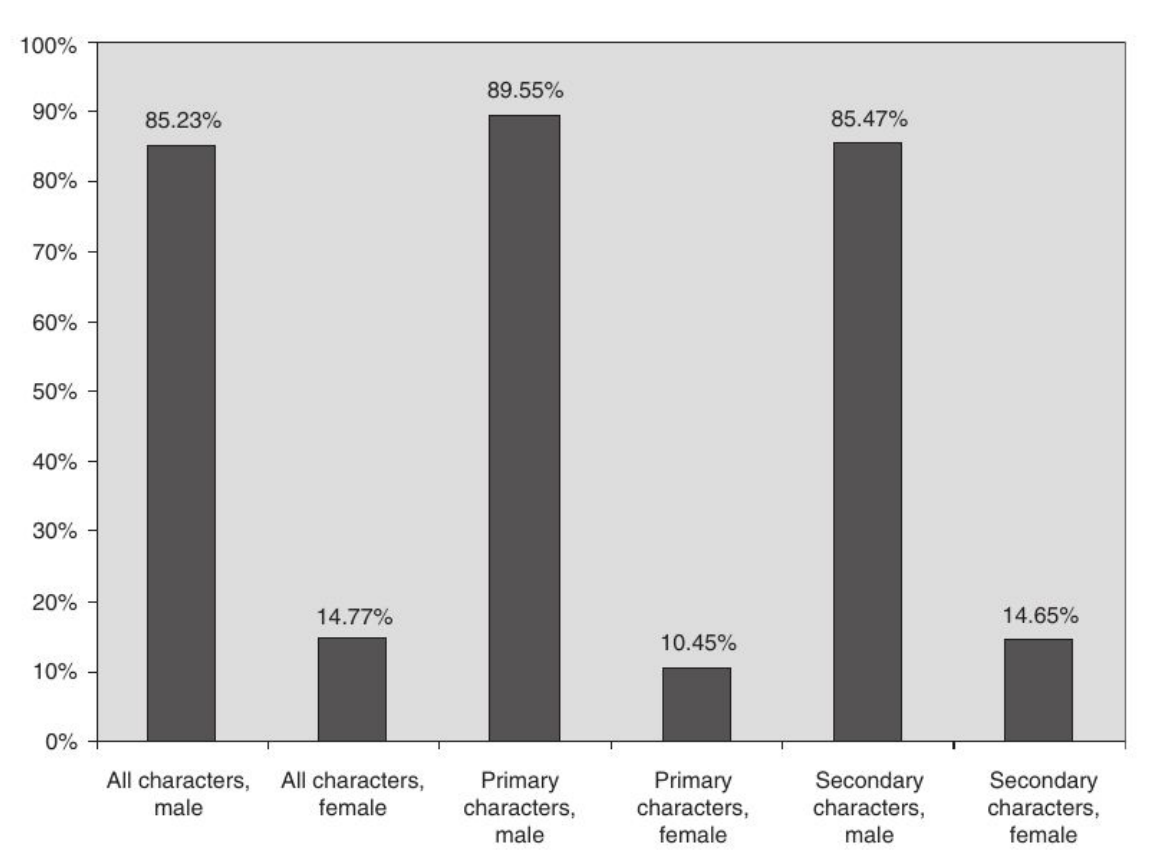
Application:
The datascapes lens and the virtual census of anime characters

Background: The Virtual Census of Videogame Characters (Williams et al. 2009)

HOCHSCHULE DER MEDIEN

- “Measuring the imbalances that exist on the screen can tell us what imbalances exist in social identity formation, social power and policy formation in daily life.” (Williams et al. 2009, 819)
- “The results show a systematic over-representation of males, white and adults and a systematic under-representation of females, Hispanics, Native Americans, children and the elderly” (Ibid., 815)

Williams, Dmitri, et al. "The virtual census: Representations of gender, race and age in video games." New media & society 11.5 (2009): 815-834.



Source: Williams et al. 2009, 825.

. 2009)

the screen can tell
y formation, social
(Williams et al.

resentation of males,
-representation of
hildren and the

Williams, Dmitri, et al. "The virtual census: Representations of gender, race and age in video games." New media & society 11.5 (2009): 815-834.

Virtual census of ACDB characters

HOCHSCHULE DER MEDIEN

- Anime Characters Database
- Over 100.000 characters
- Most characters have age and gender

Virtual census of ACDB characters

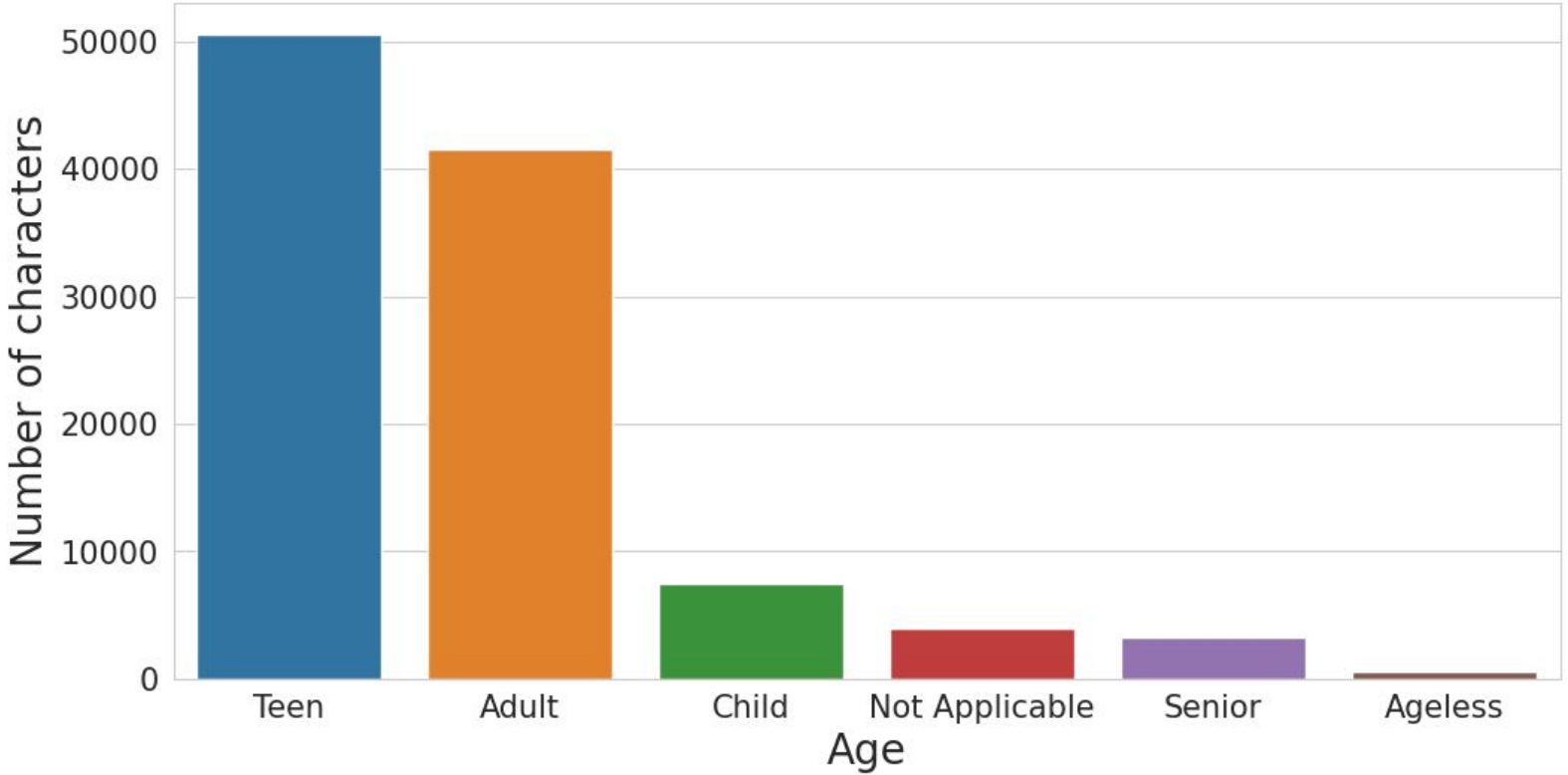
HOCHSCHULE DER MEDIEN

- Anime Characters Database
- Over 100.000 characters
- Most characters have age and gender

Emphasis will be on bringing to the fore the datascapeness of the dataset as opposed to the results of the analysis

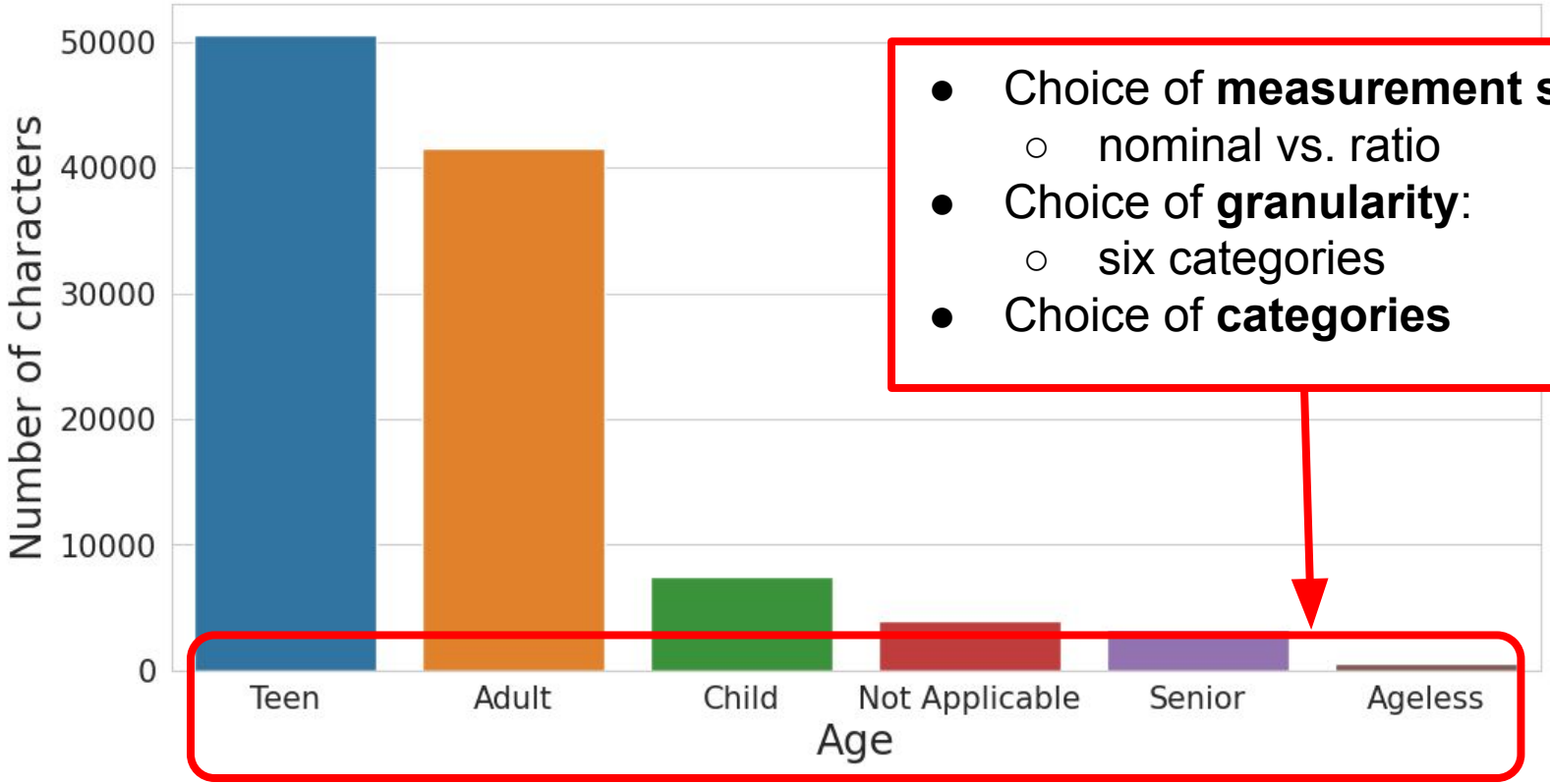
Distribution of all* ACDB characters by age

HOCHSCHULE DER MEDIEN



* Excluding characters with no publication year.

Distribution of all* ACDB characters by age

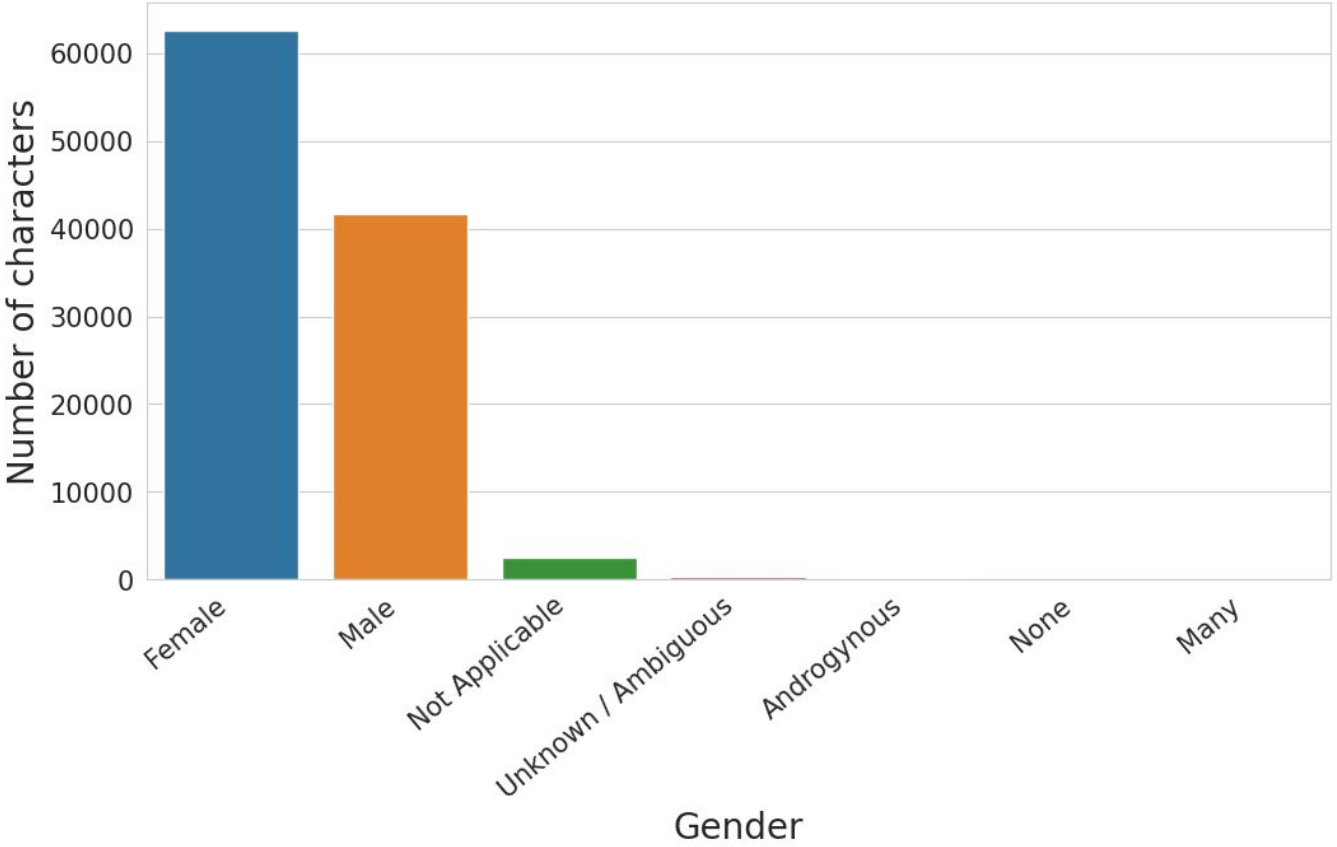


- Choice of **measurement scale:**
  - nominal vs. ratio
- Choice of **granularity**:
  - six categories
- Choice of **categories**

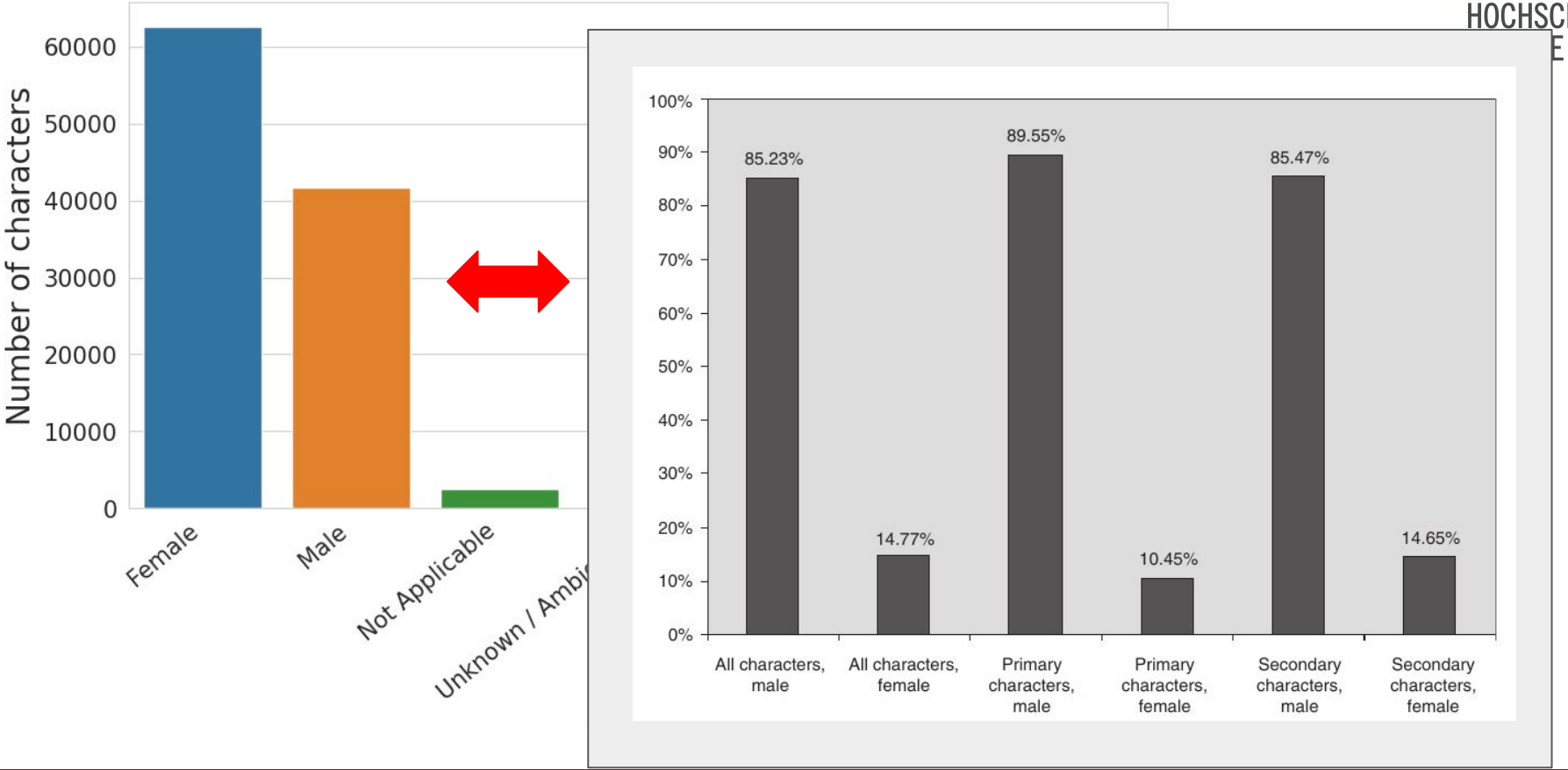
* Excluding characters with no publication year.

Distribution of all ACDB characters by gender

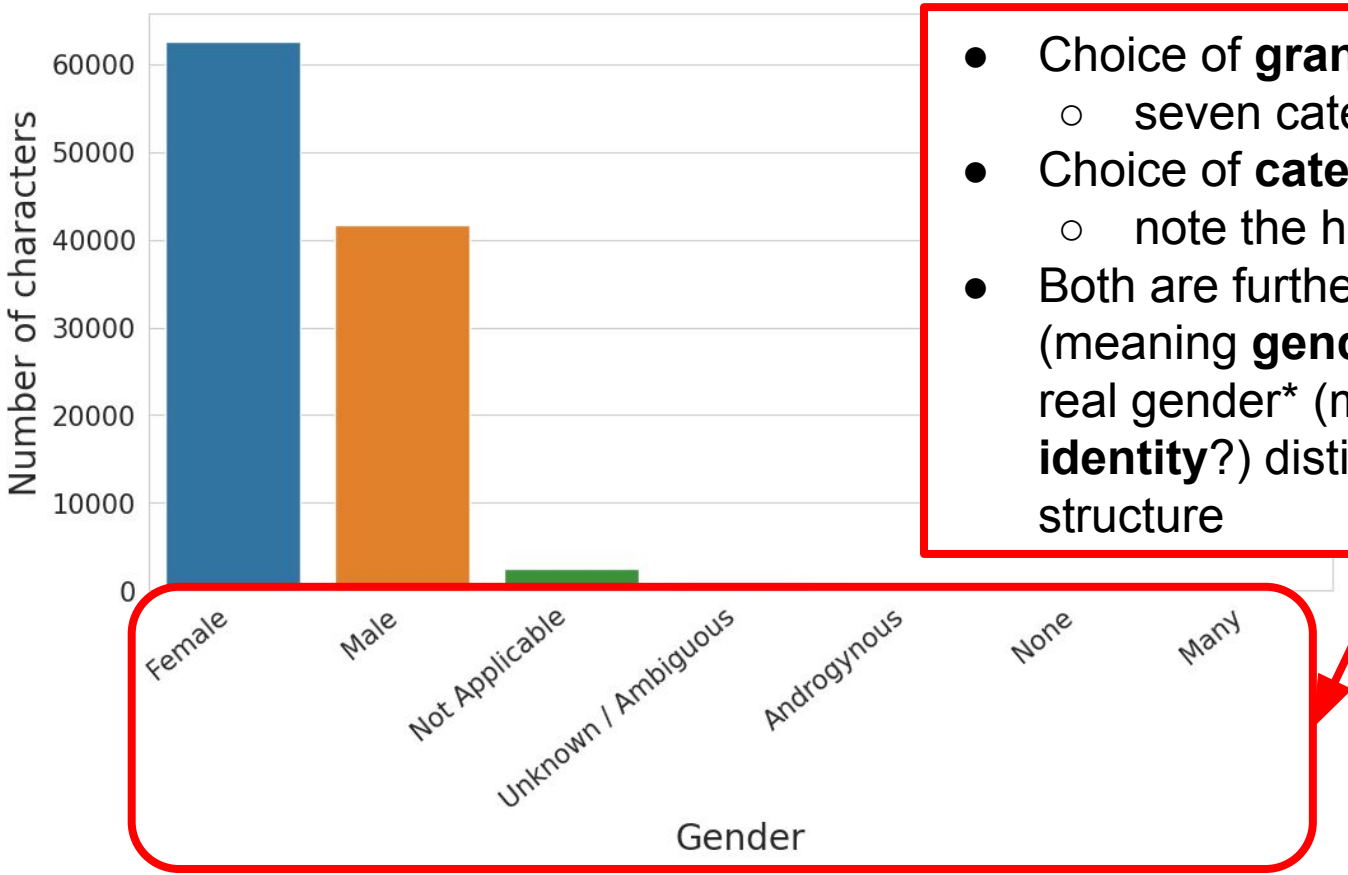
HOCHSCHULE DER MEDIEN



Distribution of all ACDB characters by gender



Distribution of all ACDB characters by gender



- Choice of **granularity**:
  - seven categories
- Choice of **categories**
  - note the handling of **non-binary**
- Both are further nuanced by gender* (meaning **gender expression**?) and real gender* (meaning **gender identity**?) distinction in the data structure

Zoom In ☰ Links

Rimuru Tempest

Comment Groups Votes

Fav 50 | Up 20 | Down 1 | Love 22 | Hate 0

Profile

Anime character Rimuru Tempest is a character from That Time I Got Reincarnated as a Slime.

They have been indexed as androgynous teen with yellow eyes and blue hair that is to waist length.

Trait	Appears as	Official
Gender	Androgynous	**(Male)**
Eye Color	Yellow	
Hair Color	Blue	
Hair Length	To Waist	
Apparent Age	Teen	
Animal Ears	No	

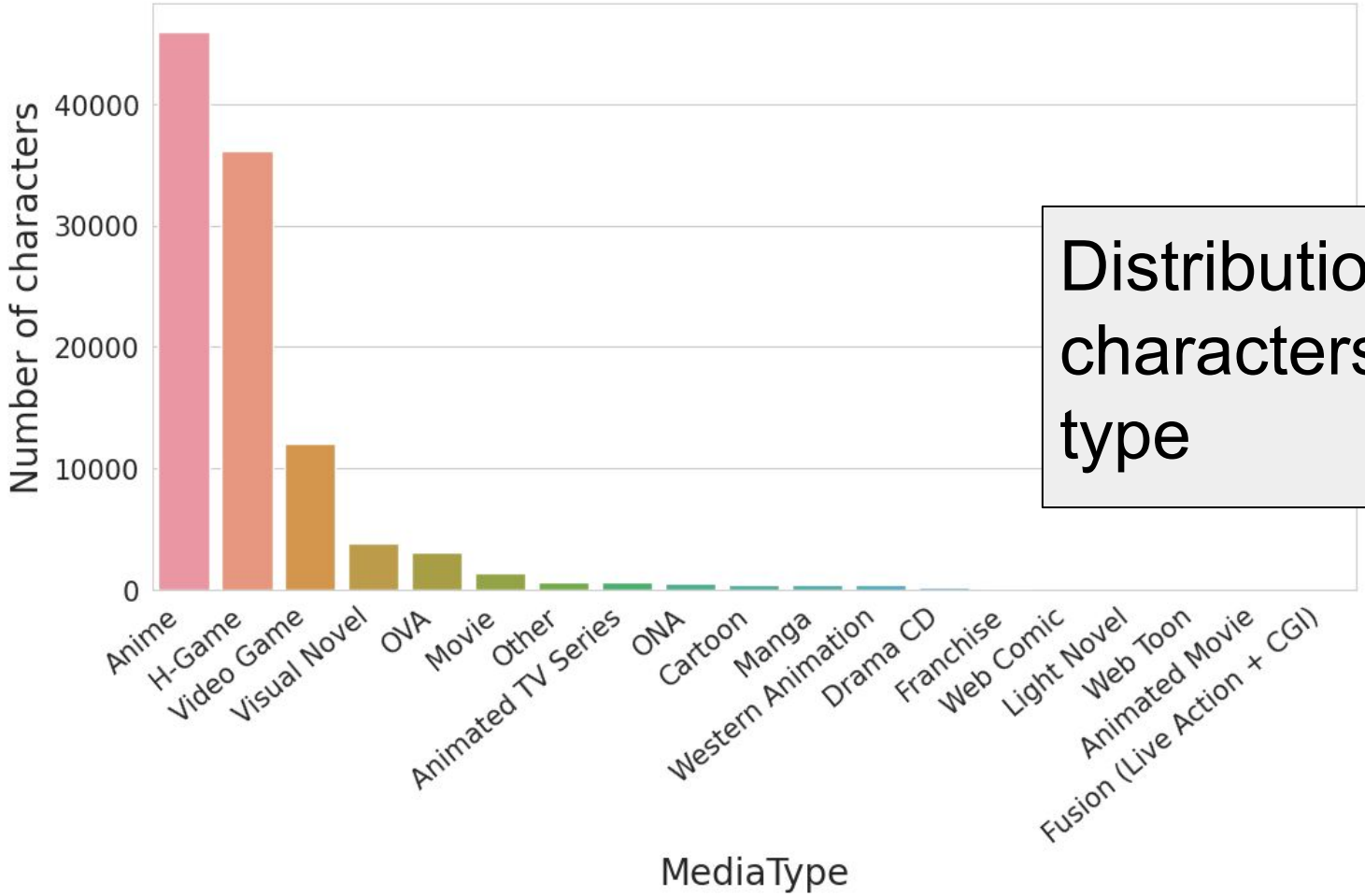
gender expression

gender identity

Relations | Relations Graph

Add a Relation

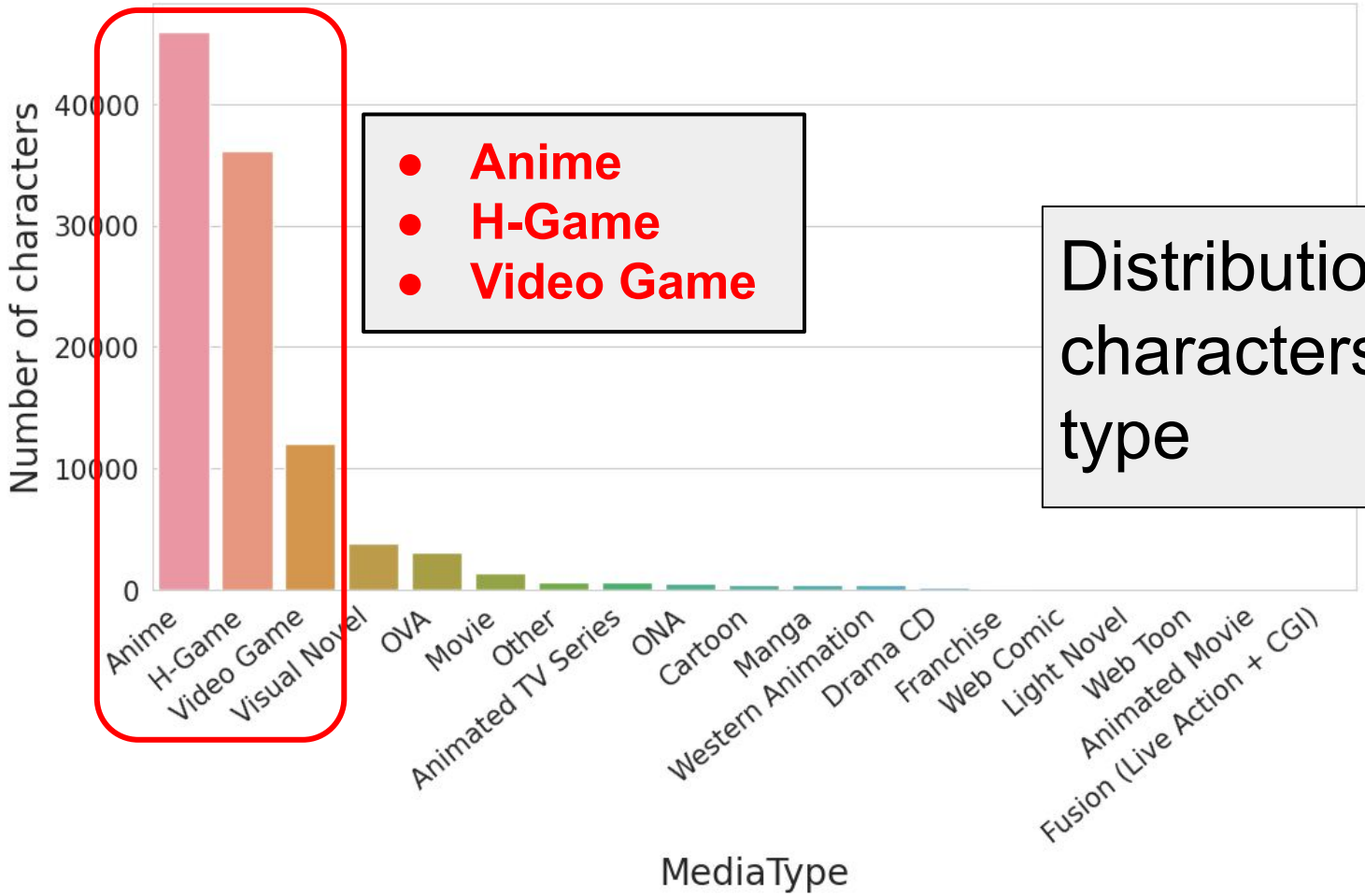
Source: https://www.animecharactersdatabase.com/characters.php?id=91525



Distribution of ACDB characters* by media type

HOCHSCHULE DER MEDIEN

* Removed characters with publication year 2021 and/or with missing media type.

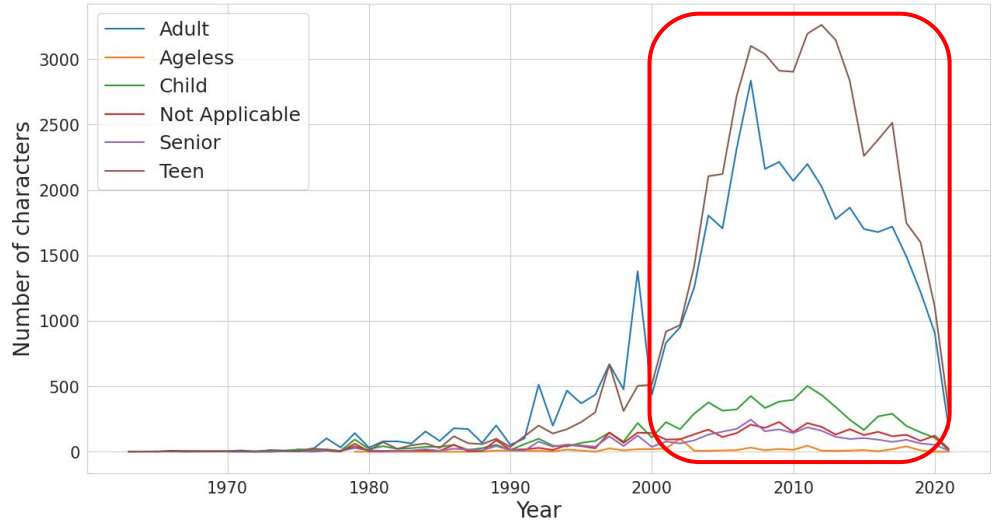


- **Anime**
- **H-Game**
- **Video Game**

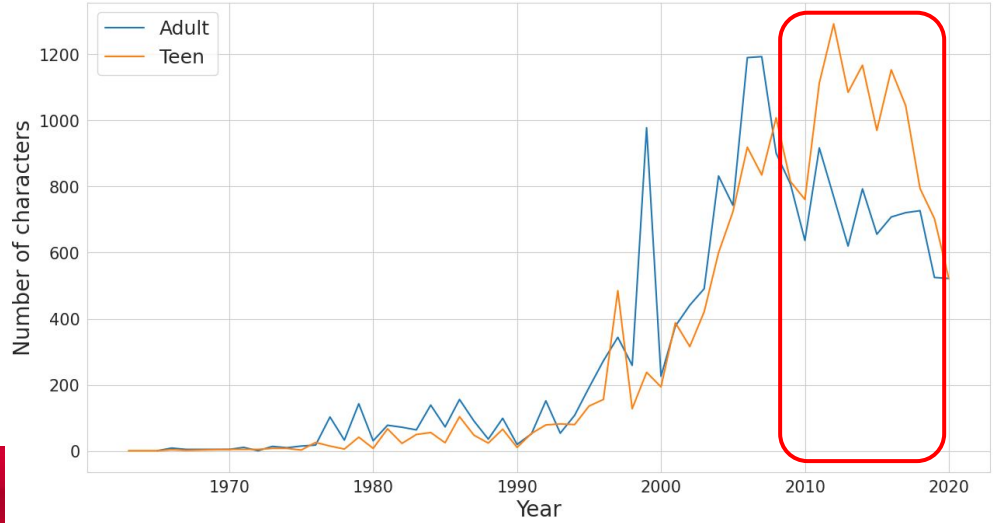
Distribution of ACDB characters* by media type

HOCHSCHULE DER MEDIEN

* Removed characters with publication year 2021 and/or with missing media type.

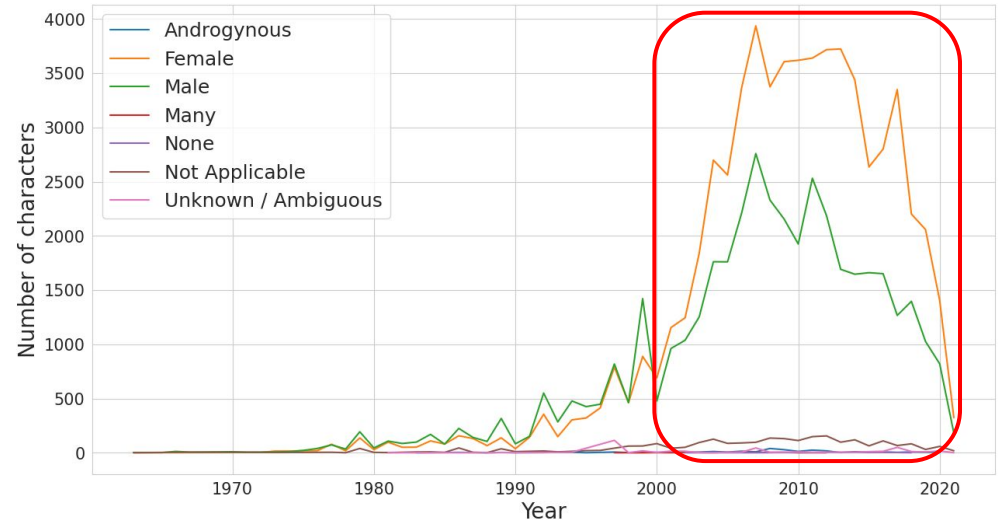


All ACDB characters:
age by year of publication

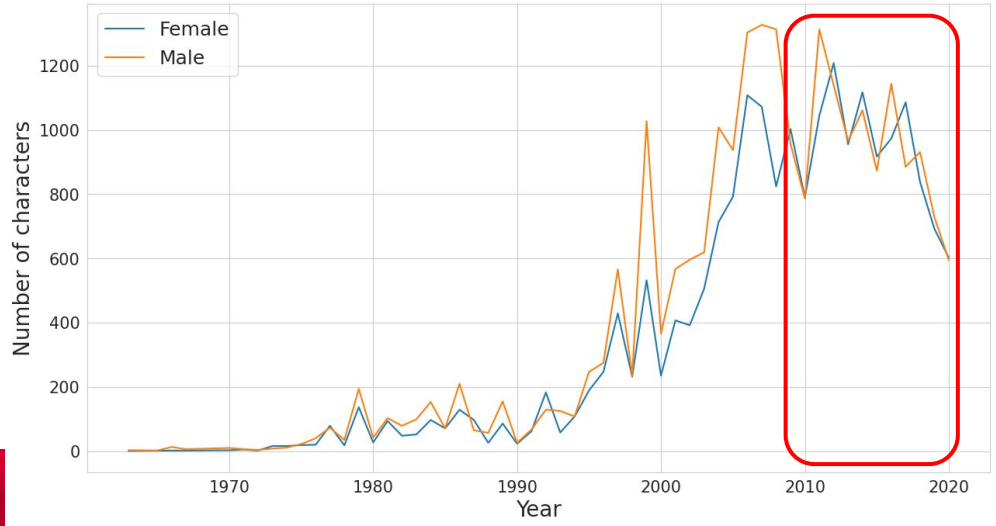


Anime characters only:
age by year of publication

HOCHSCHULE
DER MEDIEN



All ACDB characters:
gender by year of publication



Anime characters only:
gender by year of publication

Concluding points

HOCHSCHULE DER MEDIEN

- How much of what we see is a representation of the space of anime characters and how much of it is just the lens of the datascape itself?
  - **It’s both**, in each subdomain in its own way
  - Both are continuously negotiated
- Datascapes, thereby, turn into representations of cultural domains (in the example case: anime characters)

Thank you for your attention!

HOCHSCHULE DER MEDIEN

Get in touch with us:

Email: roth1003@fc.ritsumei.ac.jp,

kacsuk@hdm-stuttgart.de

Website: **https://mediagraph.link/**