Introducing the Japanese Visual Media Graph

Zoltan Kacsuk

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

Outline of the presentation

HOCHSCHULE DER MEDIEN

1. **Key characteristics** and **data sources** of the JVMG project
2. Measuring **data quality**
3. **Data integration** and **matching**
4. **Legal harmonization** of the data
5. The JVMG **web frontend**
6. Future directions
7. The **Tiny Use Case** workflow methodology
8. An example TUC: A **virtual census** of anime characters

Introducing the JVMG project

HOCHSCHULE DER MEDIEN

- Databases by enthusiast communities are often the **most up-to-date online resource for checking information**

- Project aim: Make these databases available for **large-scale quantitative research, in collaboration with the communities**
    - Funding: German Research Foundation, 6 years duration

The current JVMG project team

- Project partners
  - **Magnus Pfeffer**, Hochschule der Medien Stuttgart
  - **Martin Roth**, Hochschule der Medien Stuttgart / Ritsumeikan University Kyoto
  - Martin Hennig, IZEW Tübingen
  - Olaf Brandt, Universitätsbibliothek Tübingen

- Project employees
  - Zoltan Kacsuk
  - Tobias Malmsheimer
  - Christopher Zysik
  - Saskia Dreßler

Key characteristics of the JVMG project

HOCHSCHULE DER MEDIEN

COLLABORATION WITH COMMUNITIES

We work with diverse fan and enthusiast communities to make their data available to researchers. We respect their wishes and conditions for the use of the data.

SUITABILITY FOR RESEARCH

The needs of researchers drive all aspects of the development of the graph database, from the choice of data sources and the data model to specific representation details.

OPEN DEVELOPMENT

We document the process of data integration and the associated research in an open manner. You will find regular updates on the project blog.

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

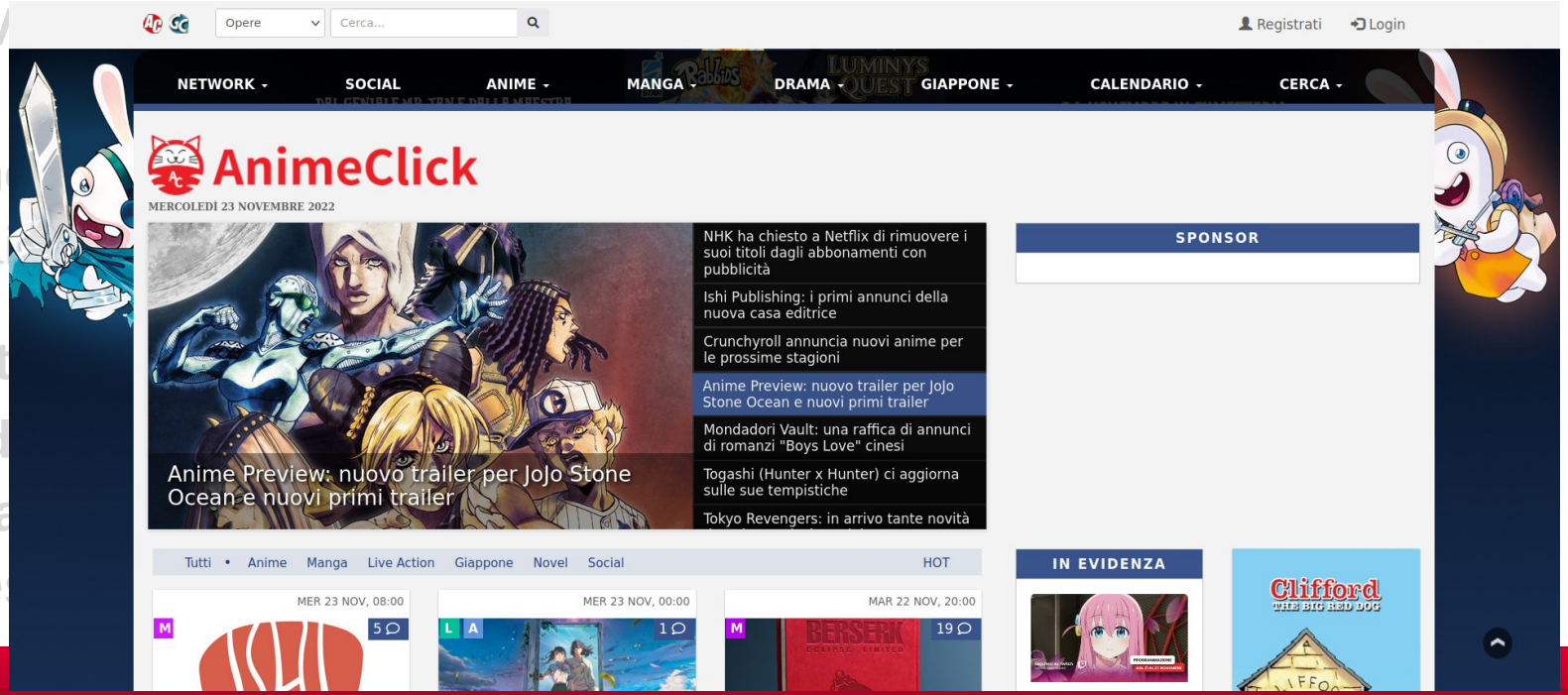
Source databases

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture
- The V
  only
- Anim
  doma

Other dat

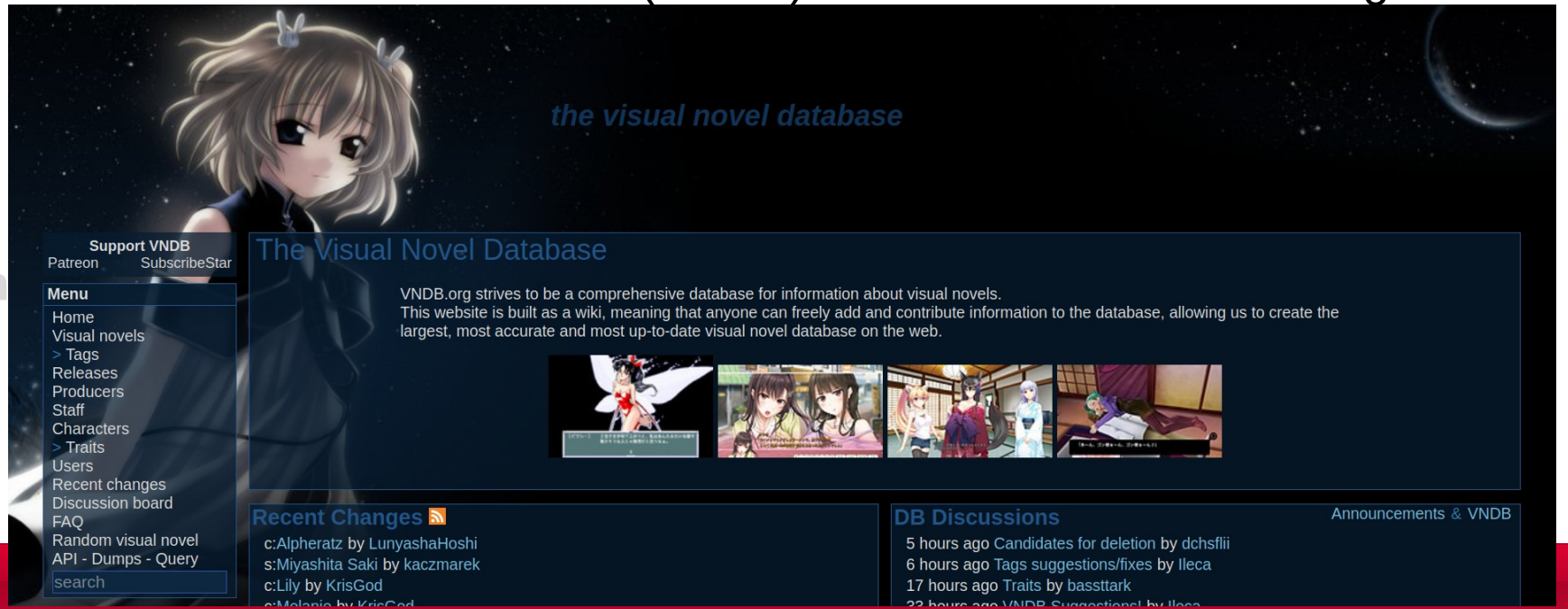
- Wikid
- Media
  game



Source databases

Fan/enthusiast community databases:

- **AnimeClick:** Wide interest in Japanese visual media and culture
- **The Visual Novel Database** (VNDB): Focused on visual novel games



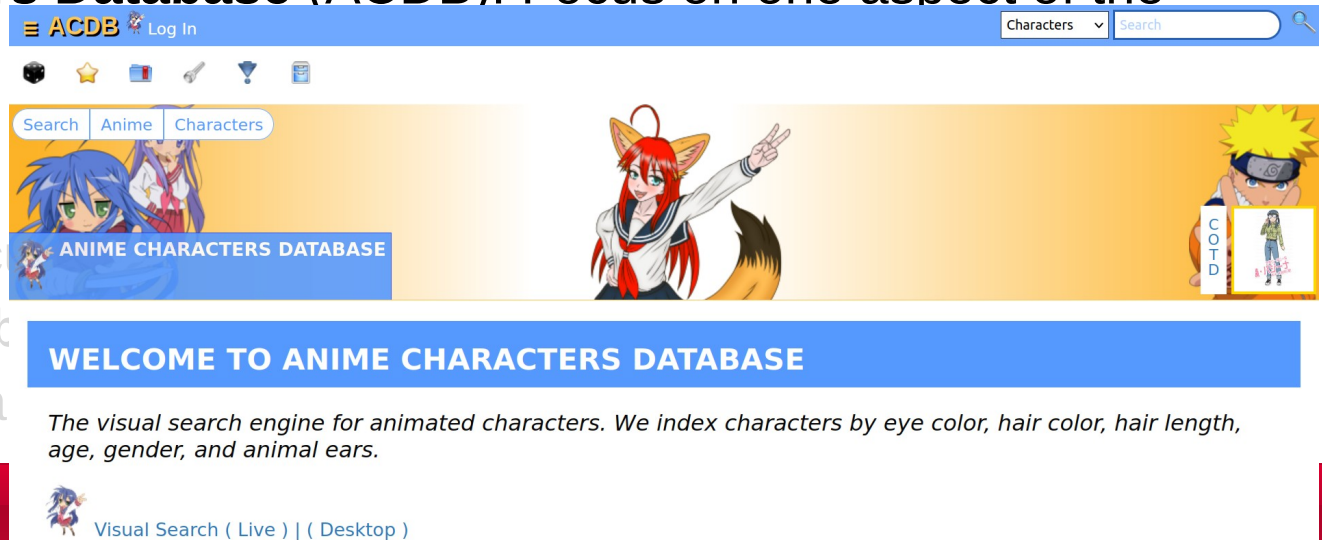
Source databases

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 foc
- **Media-Arts Datab** games and media



HOCHSCHULE DER MEDIEN

nd culture

al novel games

aspect of the

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

- **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

Entity and concept numbers

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

Entity and concept numbers

HOCHSCHULE DER MEDIEN

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		

Measuring data quality

Assessing data accuracy

HOCHSCHULE DER MEDIEN

- **Random sample** of anime (or visual novel) titles
- **Sample sizes determined** so that statistical estimates can be drawn for the population parameters
- **Manual checking** of sample elements against ground truth or official websites, etc.

Accuracy main findings

HOCHSCHULE DER MEDIEN

- **Enthusiast community databases** have a **very high accuracy**
  - Most errors are only **typographical errors**
- **Wikidata** suffers from a variety of problems
  - Large number of problems with the **title contents**
  - Largest number of **missing data**
  - **Miscategorization** errors
- **Media-Arts Database** errors are mostly due to the **automatic processing** of the material and excess information added to the title field
  - **OCR** results need to be checked and edited manually
  - A lot of **disambiguation** information added to titles

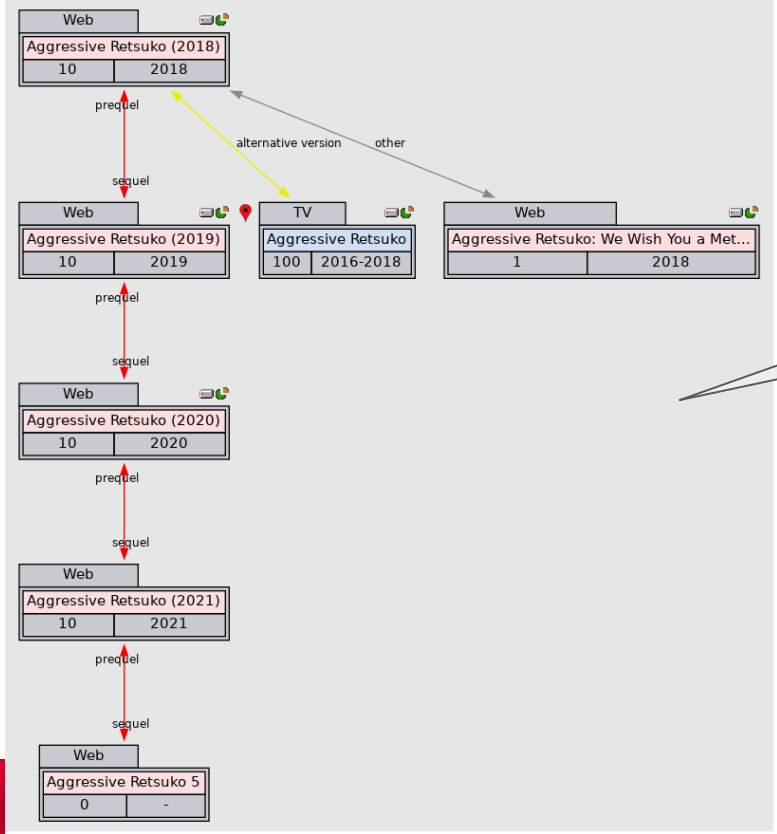
HOCHSCHULE
DER MEDIEN

Data integration and matching

Data integration

- Data is **cleaned and preprocessed**
- All data transformed into **RDF** form
- **Unified JVMG ontology** integrating the individual ontologies
- **Connections** (e.g. matching) between the databases
- **Can be connected** to other linked data sources

Matching: Granularity example



AniDB:
4 web seasons with 10 episodes each
1 announced(?) web season
1 TV series with 100 (short) episodes
1 web christmas special

Image source: anidb.net

Matching: Granularity example

AnimeClick:
4 web seasons with 10 episodes each
1 TV series with 100 (short) episodes
1 web christmas special

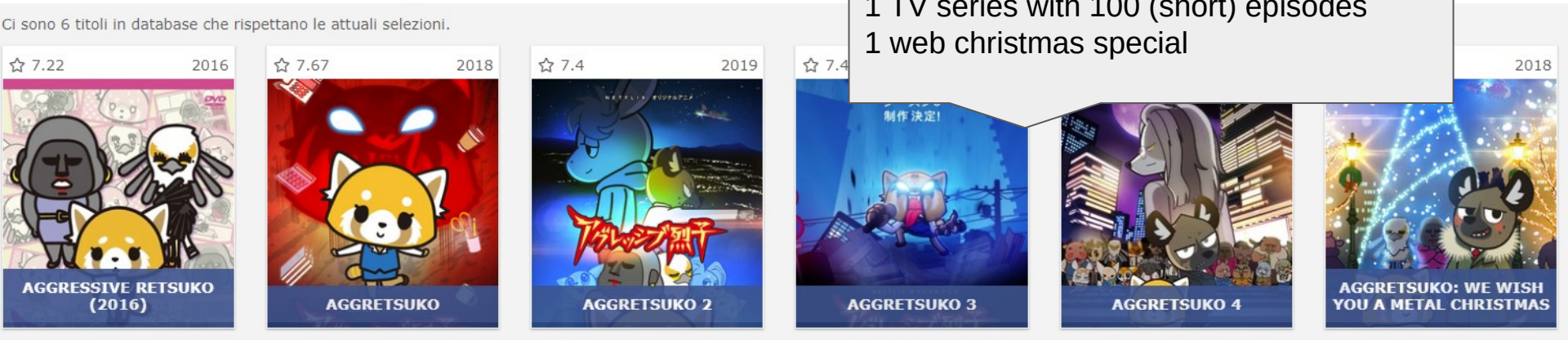
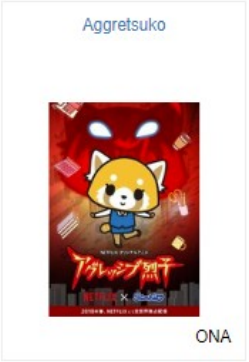


Image source: www.animeclick.it

Matching: Granularity example



Anime Characters Database:
1 TV Anime (no episode information)
1 Web ONA (no episode information)

Image source: www.animecharactersdatabase.com

Matching: Granularity example

HOCHSCHULE DER MEDIEN

	AniDB	AnimeClick	Anime Characters Database
4 web seasons	Yes	Yes	Yes
1 announced(?) web season	Yes	No	Maybe?
1 web christmas special	Yes	Yes	Maybe?
1 TV series	Yes	Yes	Yes

Legal integration of the data

Problems and challenges

HOCHSCHULE DER MEDIEN

- **Licensing practices** of the communities
  - **Lack of awareness** of or disregard for copyright issues
  - Varying and often **incompatible licenses**

- **Concerns** of the communities
  - **Wholesale copying** of their work
  - **Traffic subverted** from their sites
  - **Lack of acknowledgment** of their work

Problems and challenges

HOCHSCHULE DER MEDIEN

- **Licensing needs** of the JVMG project
  - The license has to be **open, allowing all kind of research**
  - Need to find the **lowest common denominator**
  - Have to cover **most jurisdictions**

Overview of JVMG project data sources

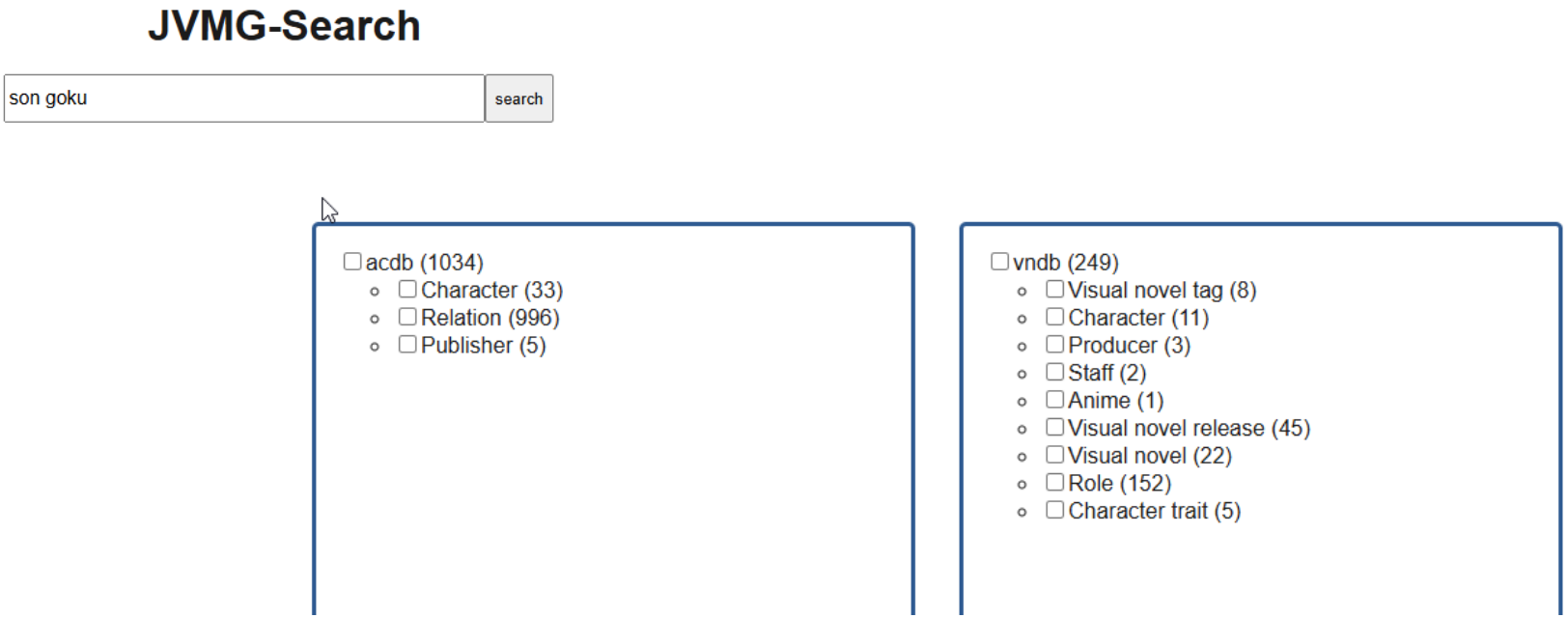
HOCHSCHULE DER MEDIEN

Data source	License	Compatibility with the CC BY-NC-SA 4.0 license
Anime Characters Database	-	CC BY-NC-SA 4.0 license provided for the JVMG project by **individual agreement** for the parts used in each case
AnimeClick	-	
The Visual Novel Database	ODbL	
Media-Arts Database	CC BY 4.0	**yes**
Wikidata	CC0	**yes**
AniDB (publicly available anime titles data dump only)	CC BY-NC-SA 4.0	**identical**

HOCHSCHULE DER MEDIEN

The JVMG web frontend

Web frontend search:
results by subgraph and entity



Web frontend presentation of the data

HOCHSCHULE DER MEDIEN

Hair Color (Hair)

Resource: http://mediagraph.link/vndb/trait/2

Graph: vndb

Property	Value
Alias en	Hair Colour en
Applicable en	false
Default spoil status en	0 en
Group en	Hair en
Name en	Hair Color en
Searchable en	false
Trait parent en	Hair en
is Trait parent en of 15	• Black (Hair) en • Blond (Hair) en • Blue (Hair) en • Brown (Hair) en • Cyan (Hair) en • Dyed (Hair) en • Green (Hair) en • Grey (Hair) en • Multicolored (Hair) en • Orange (Hair) en • Pink (Hair) en • Red (Hair) en • Teal (Hair) en • Violet (Hair) en • White (Hair) en
VNDB link en	https://vndb.org/i2
label	Hair Color (Hair) en
sexual status en	false
type	Character trait en

Dark Mode Search Create Index crosstab languages

Shinpi no Hou 神秘の法 Shinpi no Hou

Resource: http://mediagraph.link/jvmg/mediawork/10

Graph: anidb

hide

Property	Value	
label	Shinpi no Hou x-jat	
AniDB link en	https://anidb.net/anime/9081	
Official title en	神秘の法 ja	
Primary title en	Shinpi no Hou x-jat	
Synonym en	• Happy Science Movie 6 x-unk	• The Mystical Laws en

Graph: http://mediagraph.link/graph/cluster

hide

Property	Value
http://mediagraph.link/jvmg/ont/hasMember	• Shinpi no Hou x-jat • 神秘の法
label	Shinpi no Hou

Graph: madb

hide

Property	Value
https://mediaarts-db.bunka.go.jp/data/property#periodDisplayed	10/6より公開。日活配給。
https://mediaarts-db.bunka.go.jp/data/property#source	アニメアワード投票用データ ／ 年間パーフェクト・データ（AWD+APDY）

Web frontend intro page

Inspired by the DraCor (dracor.org) and GOLEM (golemlab.eu) projects

HOCHSCHULE DER MEDIEN

Overview

Anime Characters Database Graph

Character A character found in ACDB.	131029
Publisher Within ACDB, Publisher refers to a title or series to which a character belongs, and not an entity that is engaged in a publishing role.	12381
Tag A class for descriptive details about an anime or manga.	331
Character Tag A class for descriptive details about a character	1240
Individual A real-world individual associated with a character.	9308

AnimeClick Graph

Character Character of a manga or anime	57579
Author Author of an anime or manga	59502
Staff A person or group involved in the creation or publication of a creative work.	187017
Episode Single episode of an anime	212396
Creative work Creative work includes visual novels and animes	36604
Tag A class for descriptive details about an anime or manga.	2885

Media Arts Database Graph

Agent	60383
Animation TV program	124733
Animation video package	9191
Game package	48118
Manga book	331862
Manga other	5805
Media Art exhibition or performance	14335

Links for 10 random entities of the type

Visual Novel Database Graph

Anime Indicates the anime of a visual novel	1831
Visual novel A specific visual novel.	45612
Character A visual novel character.	126578
Staff An individual staff member that has contributed to the creation of a visual novel.	33267
Visual novel release Physical or online release of a visual novel	107961
Character trait A descriptive character trait found in VNDB	3218

HOCHSCHULE DER MEDIEN

Future directions

Currently in development

HOCHSCHULE DER MEDIEN

- Functions
  - **New search interface** with assistance for **complex queries**
  - Data **export functions** for working with the data in other tools (Pandas, Spreadsheets, Gephi, etc.)
- Data
  - Adding data from **tvtropes.org**
  - **Updating** to the newest version of the Media Arts Database data
  - **Integration** of all data in the shared JVMG graph
- Services
  - Online **documentation** and **tutorials** for typical use cases
  - In-person **JVMG Labs** for researchers in Stuttgart and Kyoto

Possible future directions

- **End of the current project phase** beginning of 2026
  - Further development with own resources or new partners?
- Integrating further data sources
  - Music and audio drama CD publications
  - **Light Novels**
  - Computer games
  - Further international communities
- Deepening of the descriptions
  - Media contents
  - Character attributes
  - **More granular breakdown of works** (For example: individual manga chapters or anime episodes)

The Tiny Use Case workflow methodology

General idea for the Tiny Use Case (TUC)

HOCHSCHULE DER MEDIEN

- Pioneered by the **diggr** (Databased Infrastructure for Global Games Culture Research) research project team
- Inspiration from **agile** software development principles
  - Cycle of continuous incremental innovations and assessments
- Each TUC is **3-4 months** long
- Each TUC confronts the data with a **specific question/aim**

The purpose of TUCs

HOCHSCHULE DER MEDIEN

- Exploring of the **data's usability and adequacy for research**
  - **Limits**, **quality** issues and **fit**
- **Empowering** the researcher
  - Needs of the researchers in relation to the **frontend**
- Finding a **common language**
  - Bridging disciplinary boundaries between library and computer science on the one hand, and humanities and social science on the other

Example TUCs with the JVMG knowledge graph

HOCHSCHULE DER MEDIEN

1. Testing one of the points from Hiroki Azuma's **"Otaku: Japan's Database Animals"**
2. **Exploring recurring patterns in character creation** in visual novel games
3. **Examining the concept of media mix** by looking at networks of co-appearing characters
4. **Census of characters** in Japanese visual media

HOCHSCHULE
DER MEDIEN

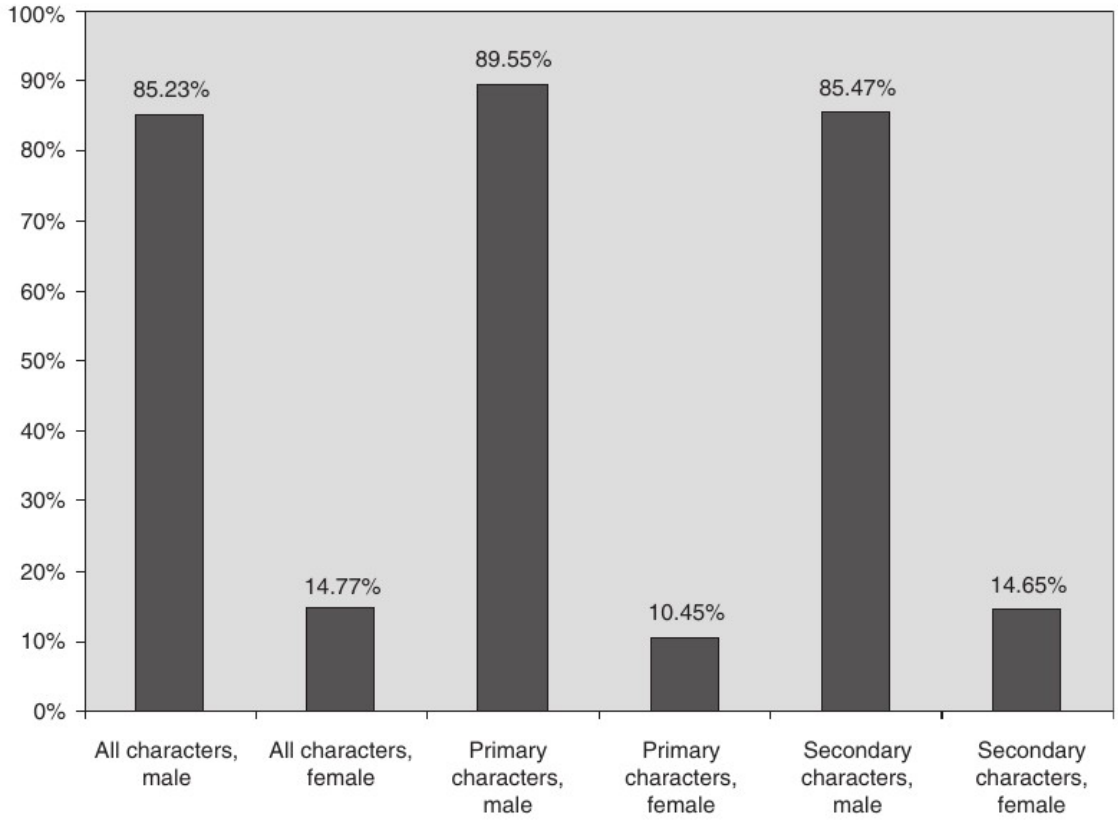
An example TUC:
A virtual census of anime characters

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

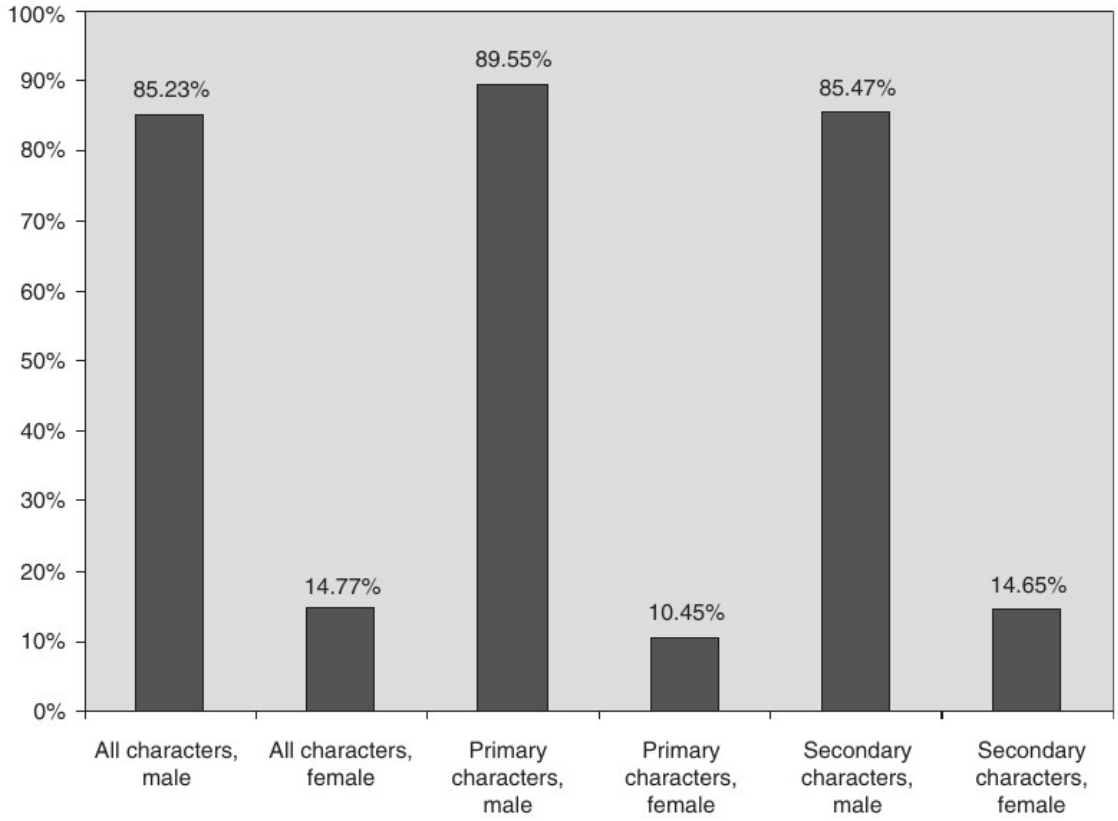
HOCHSCHULE DER MEDIEN

- “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.



Source: Williams et al. 2009, 825.

A recent update of the study (Harrisson et al. 2020) suggests some change:

“similarly to the 2009 study, most characters proportionally are White, male, and adult.” (Harrisson et al. 2020, 2)

“the overall proportion of female characters has nearly doubled from 13% in 2005 to 22% in 2017. [...] this increase was primarily due to non-playable secondary roles” (ibid.)

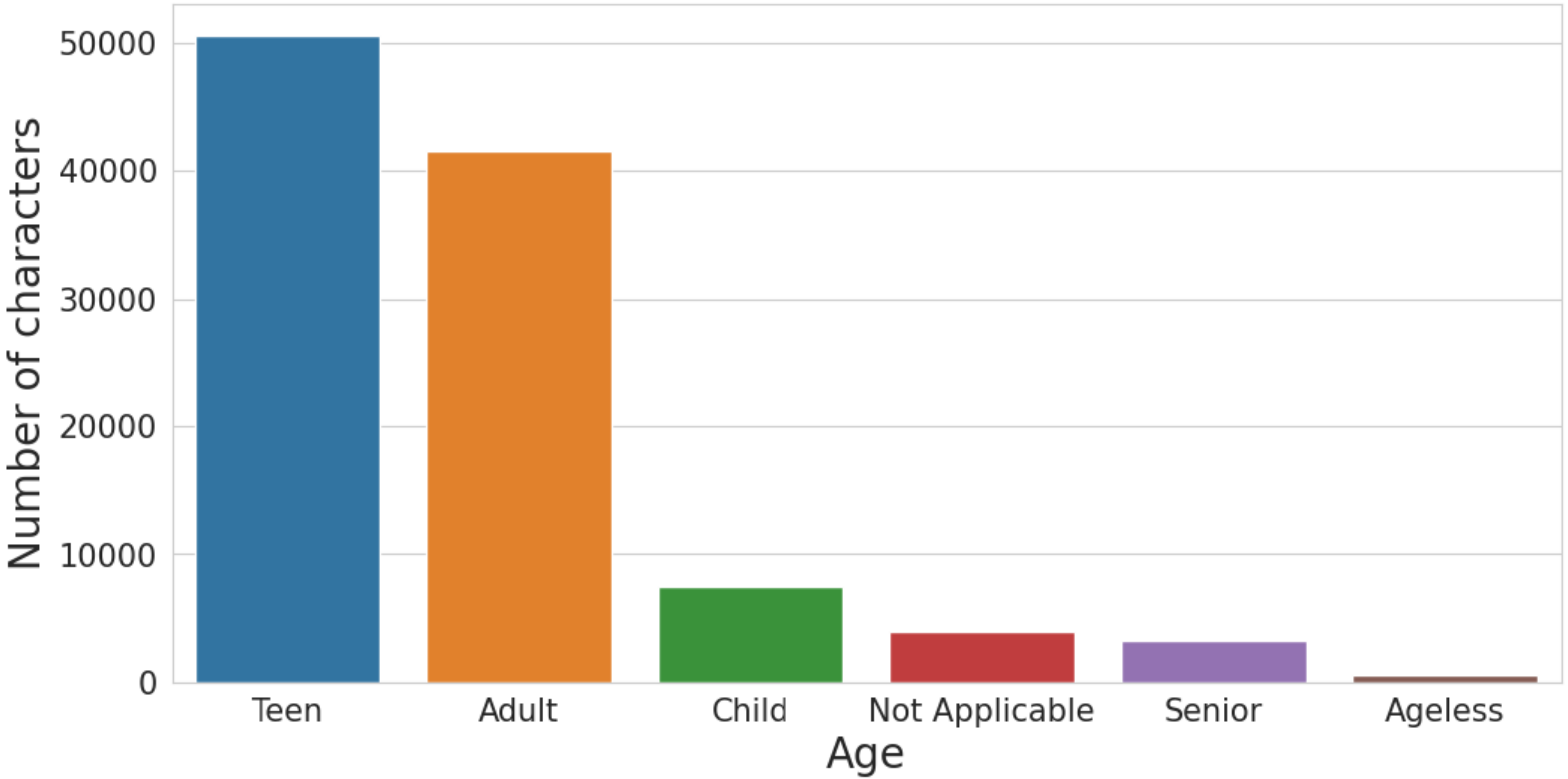
Harrisson, Annie, et al. "THE VIRTUAL CENSUS 2.0: A CONTINUED INVESTIGATION ON THE REPRESENTATIONS OF GENDER, RACE AND AGE IN VIDEOGAMES." AoIR Selected Papers of Internet Research (2020).

HOCHSCHULE
DER MEDIEN

Who populates Anime? Examining the Anime Characters Database

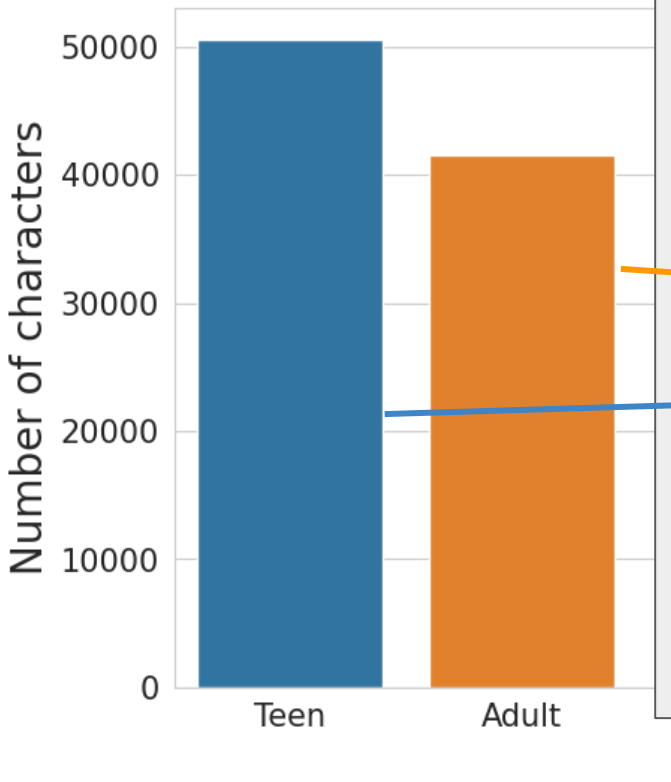
Distribution of all* ACDB characters by age

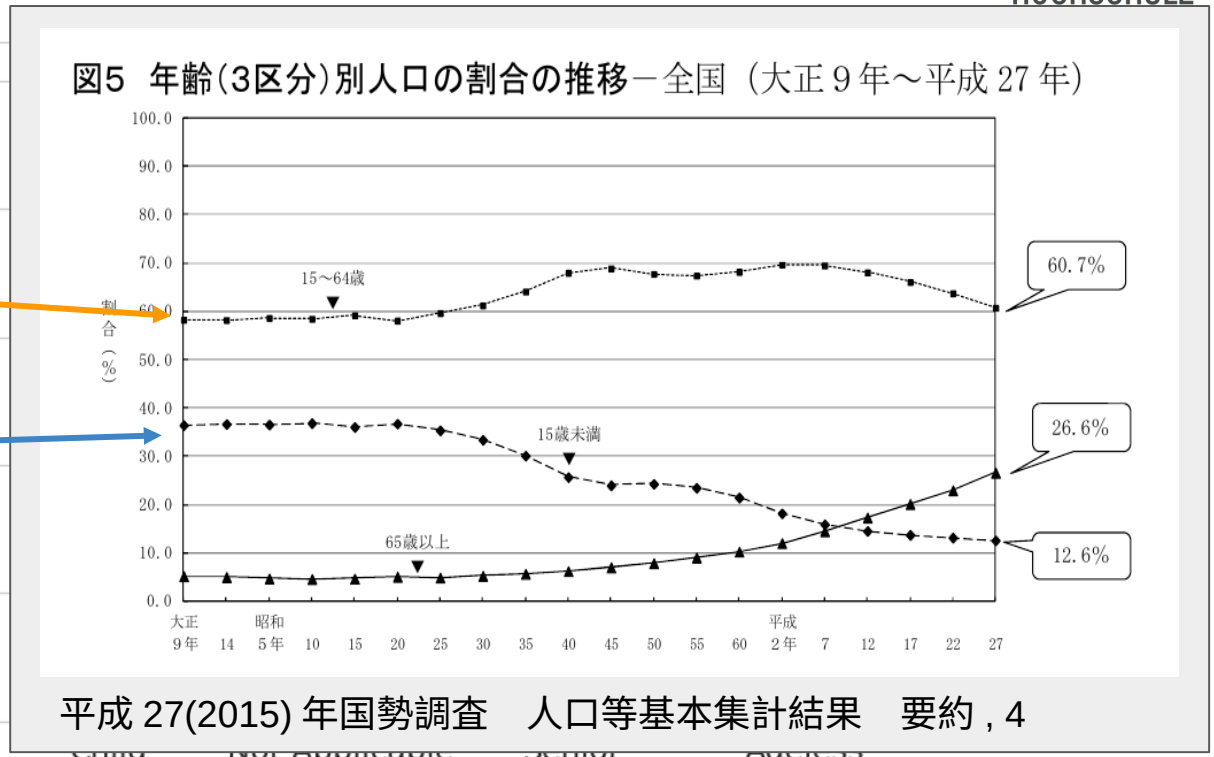
HOCHSCHULE DER MEDIEN



* Excluding characters with no publication year.

Distribution of all* ACDB characters by age

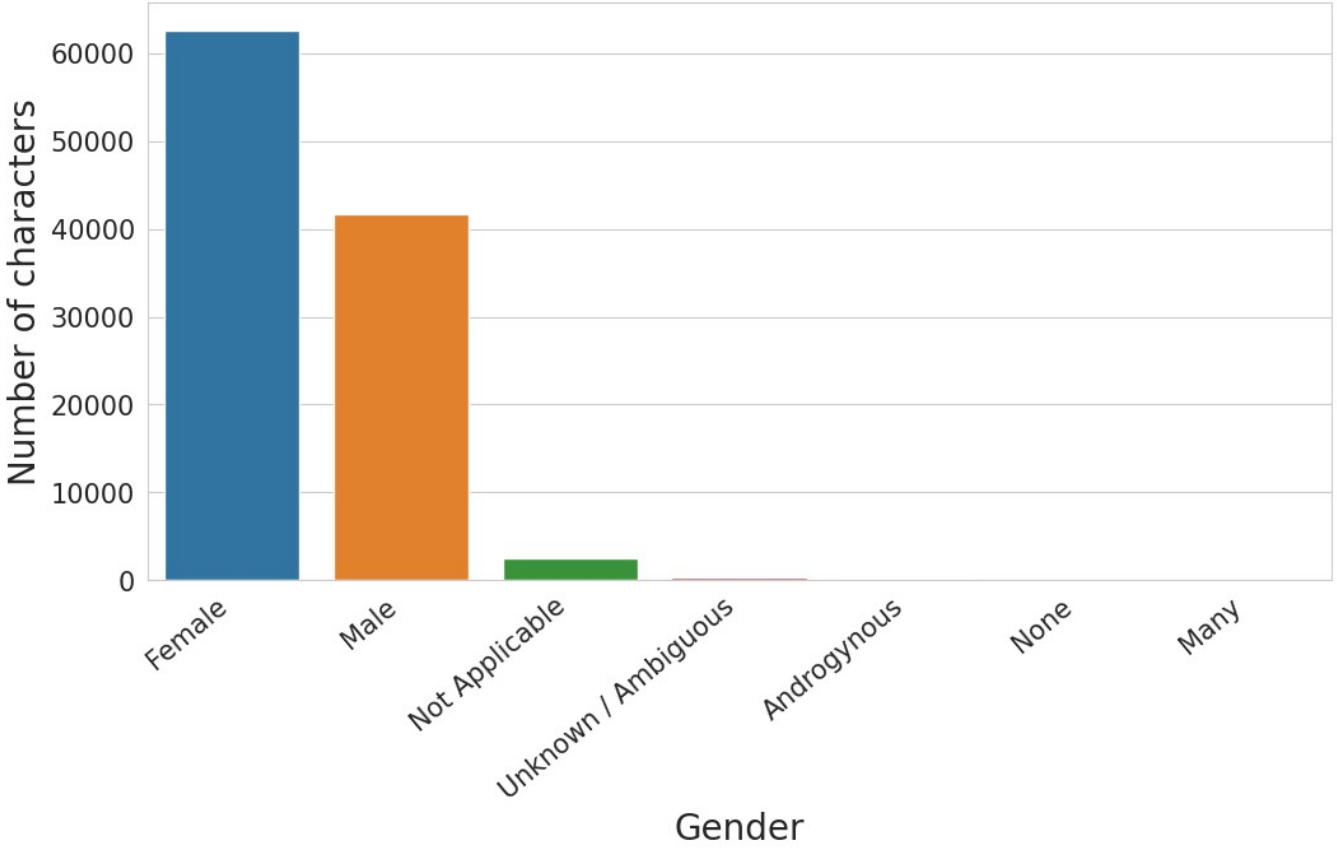




平成 27(2015) 年国勢調査　人口等基本集計結果　要約 , 4

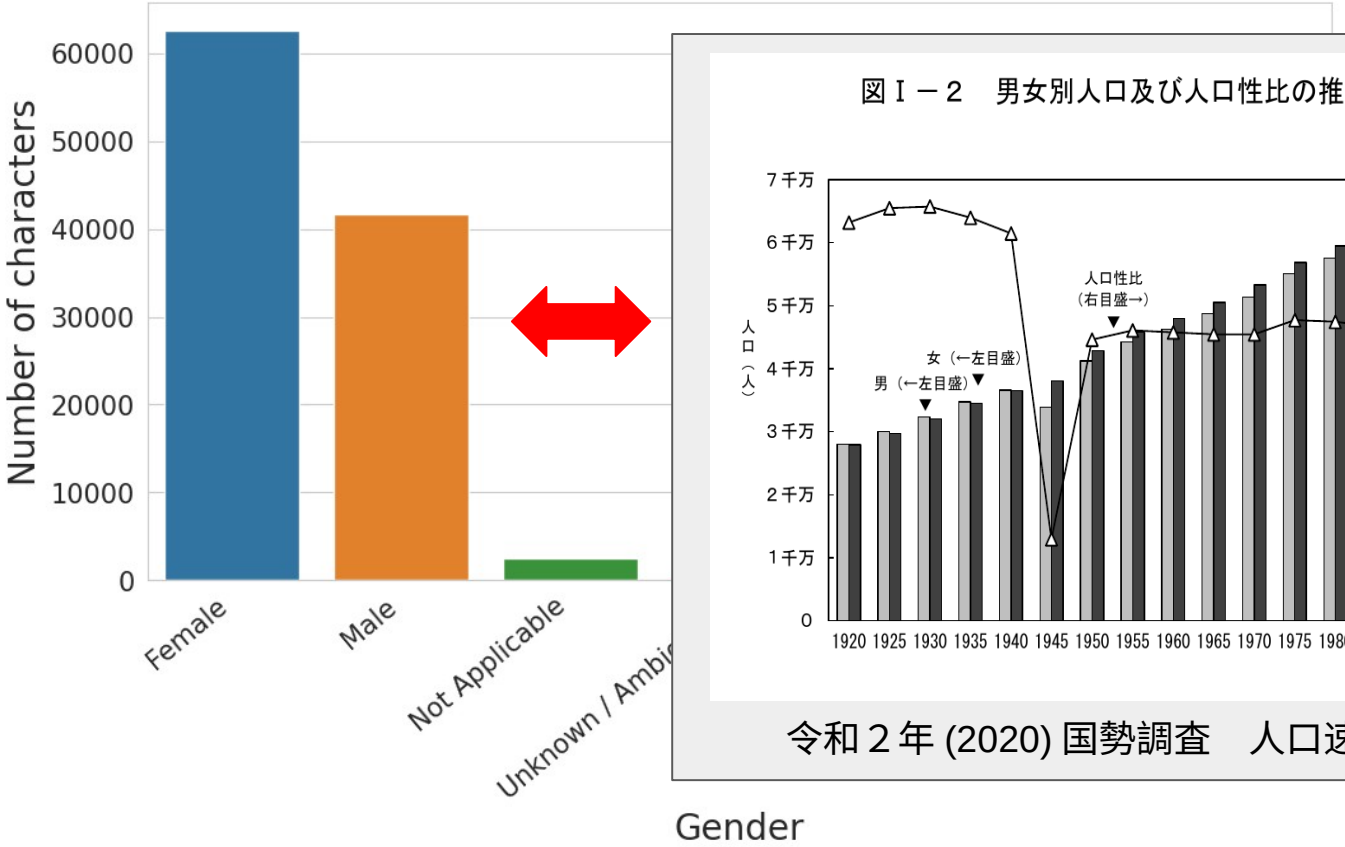
* Excluding characters with no publication year.

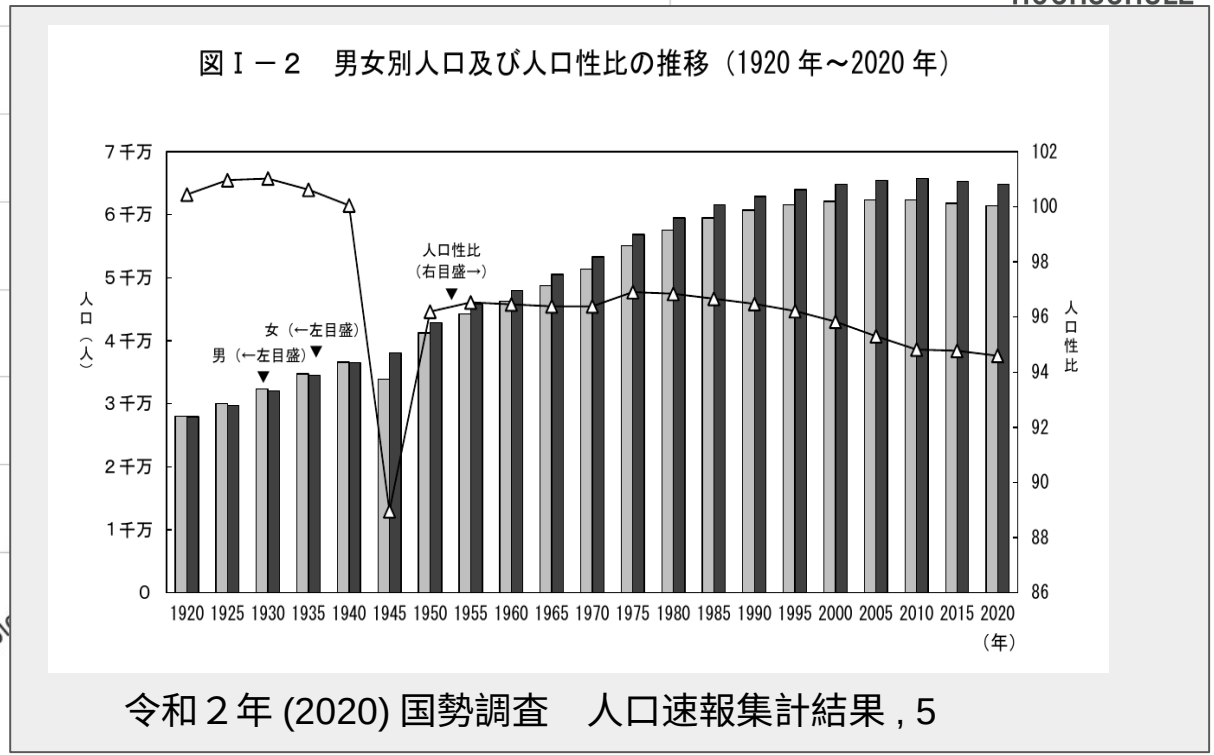
Distribution of all ACDB characters by gender



HOCHSCHULE DER MEDIEN

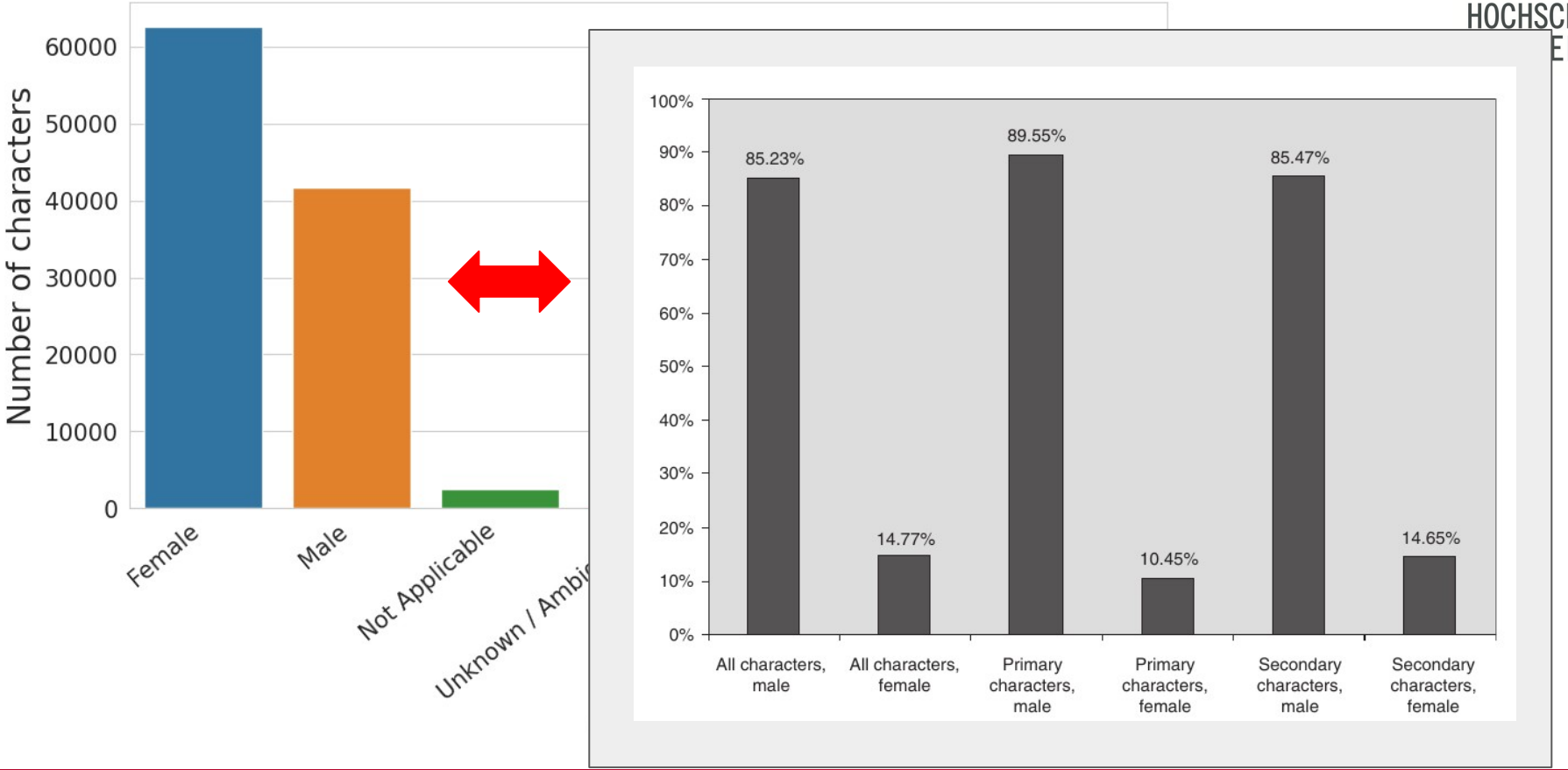
Distribution of all ACDB characters by gender



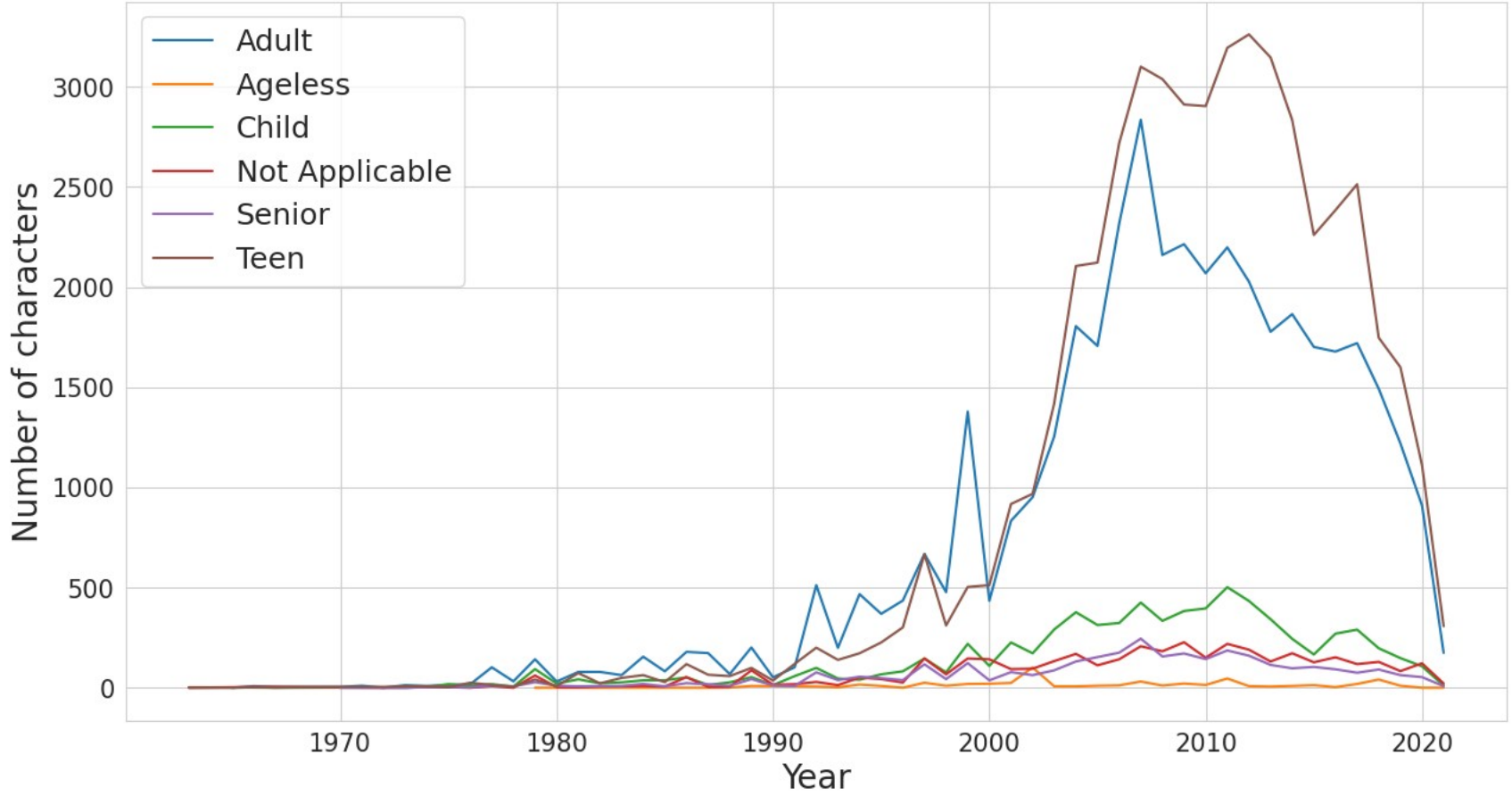


令和２年 (2020) 国勢調査　人口速報集計結果，5

Distribution of all ACDB characters by gender

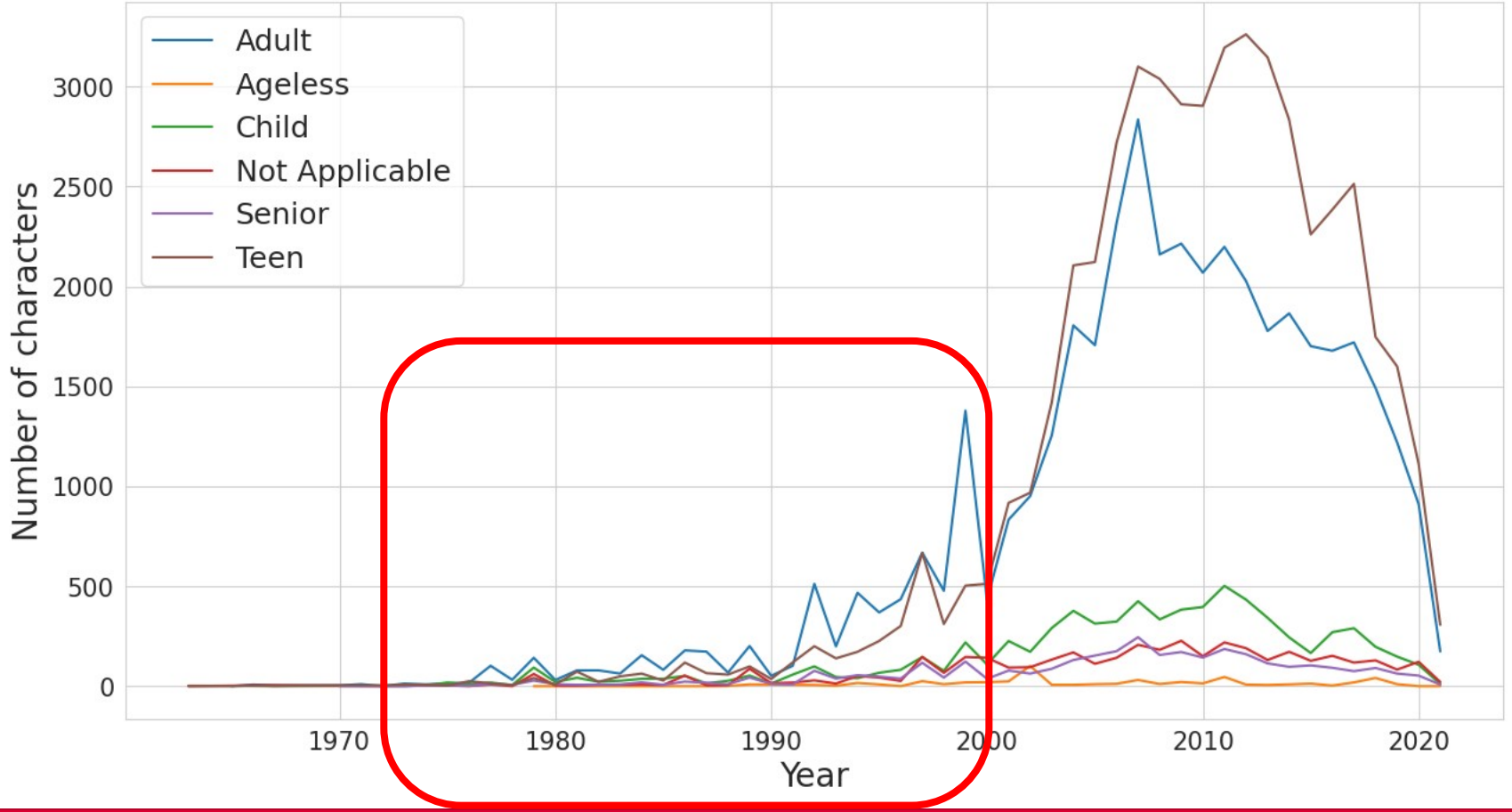


Distribution of all ACDB characters by age and year



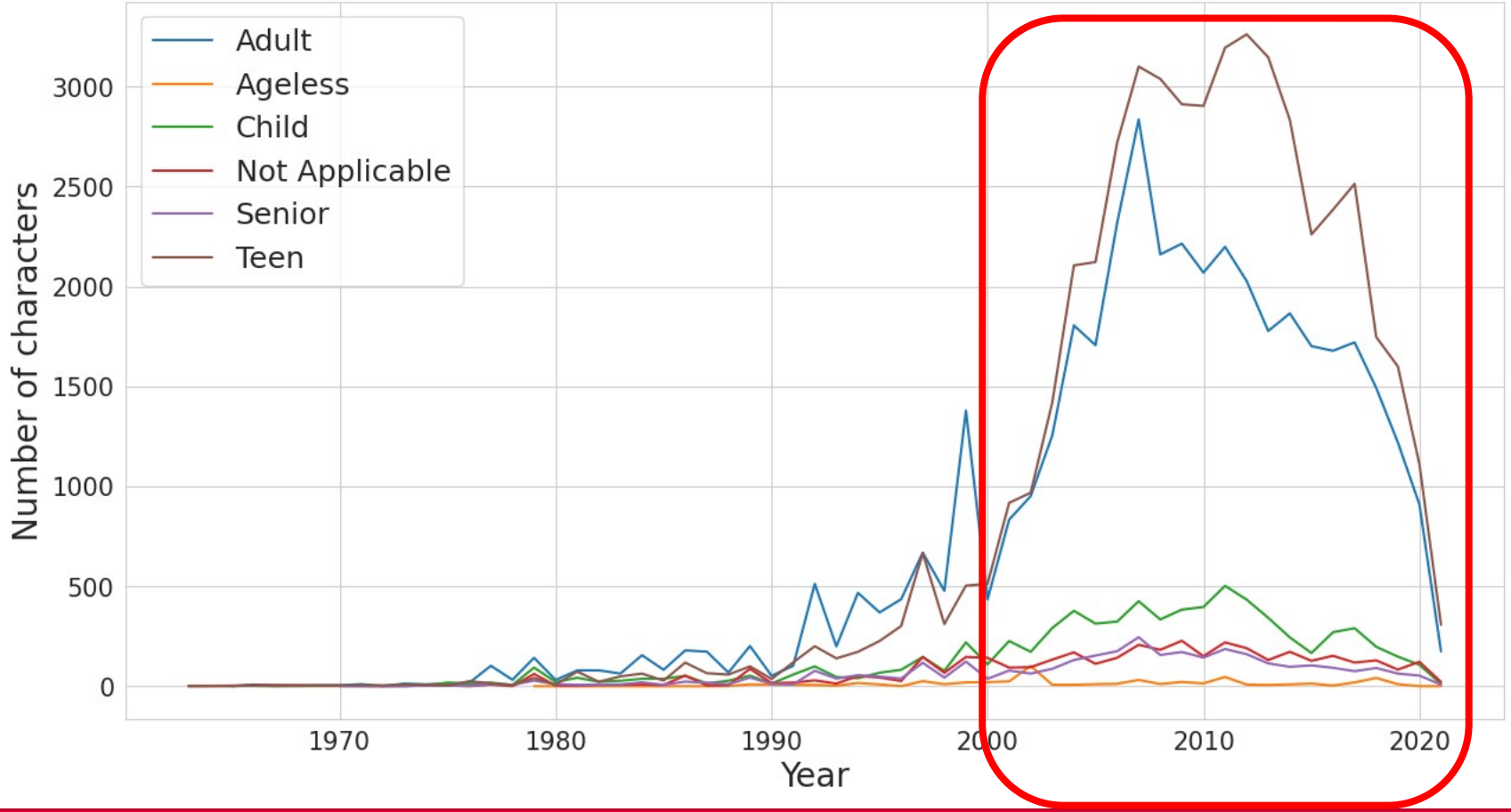
HOCHSCHULE DER MEDIEN

Distribution of all ACDB characters by age and year



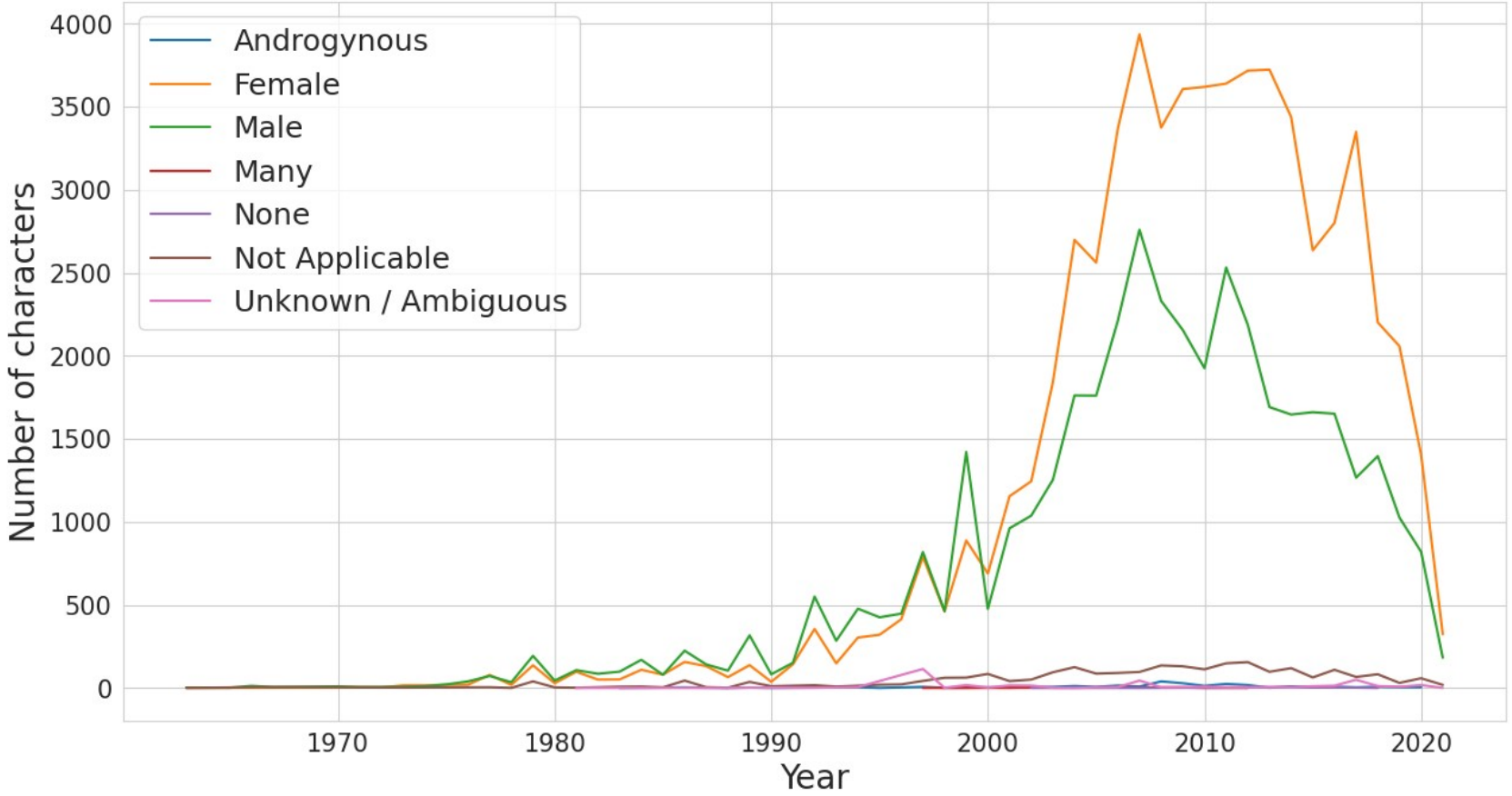
HOCHSCHULE DER MEDIEN

Distribution of all ACDB characters by age and year



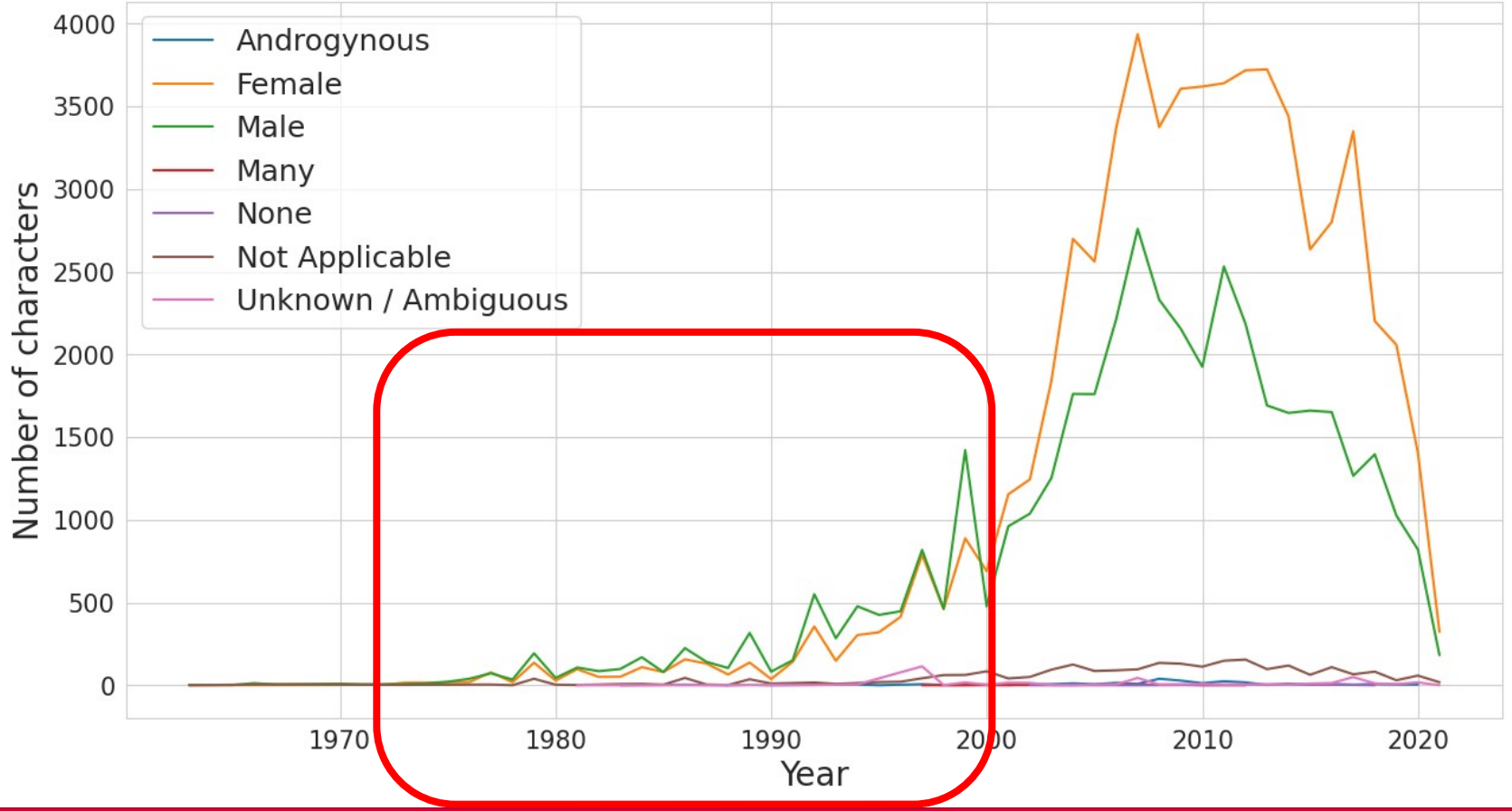
HOCHSCHULE DER MEDIEN

Distribution of all ACDB characters by gender and year



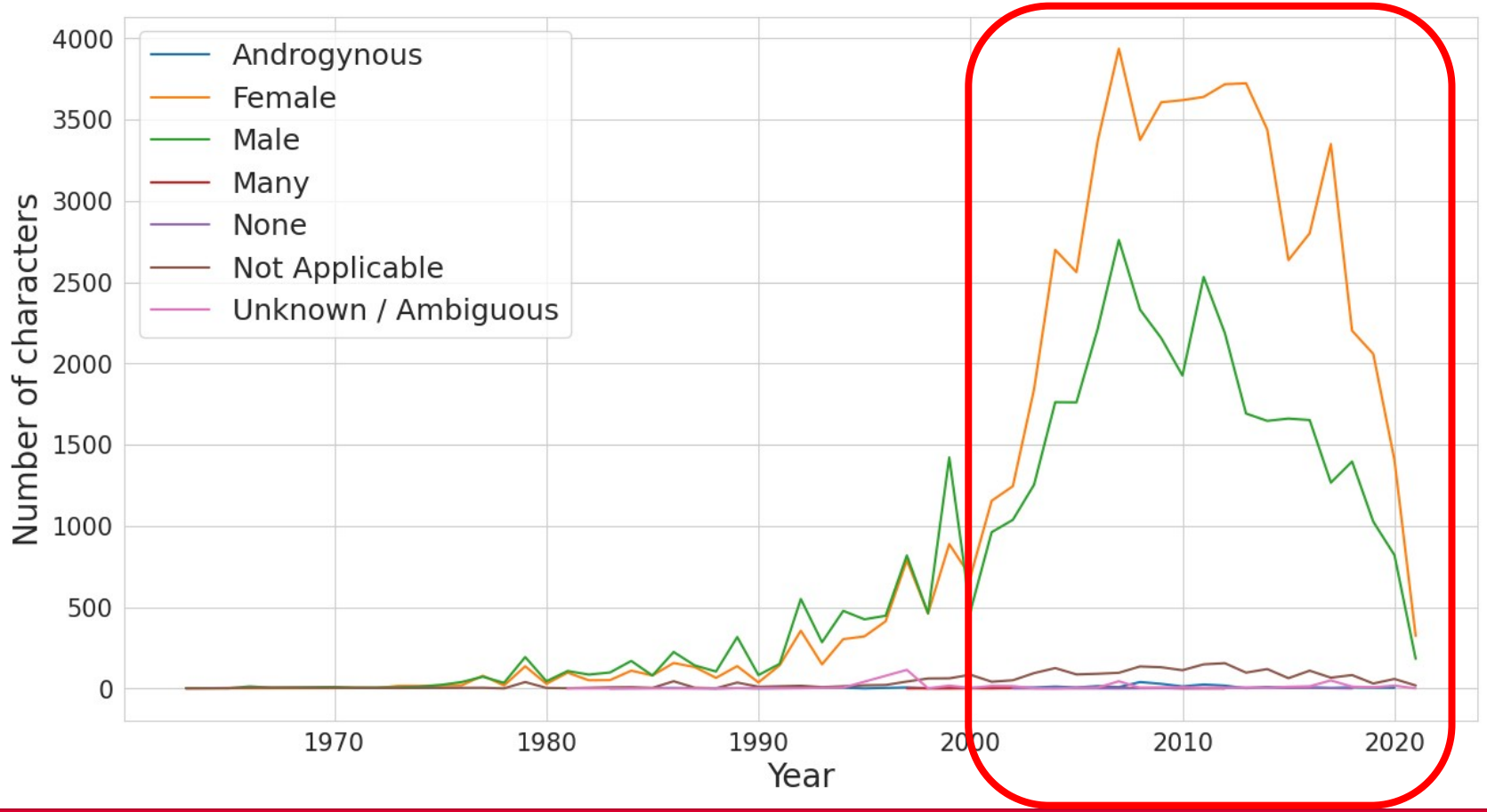
HOCHSCHULE DER MEDIEN

Distribution of all ACDB characters by gender and year



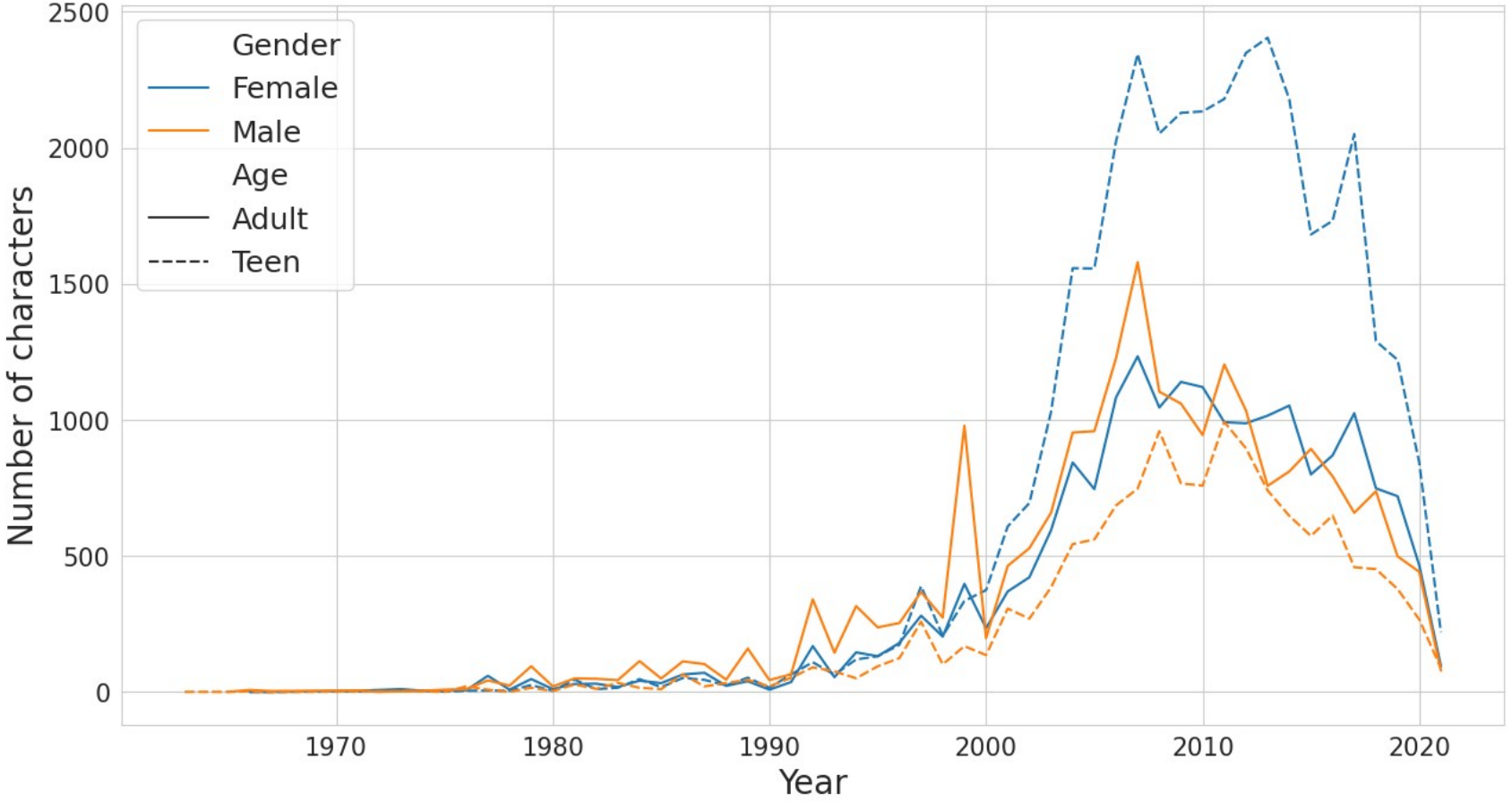
HOCHSCHULE DER MEDIEN

Distribution of all ACDB characters by gender and year



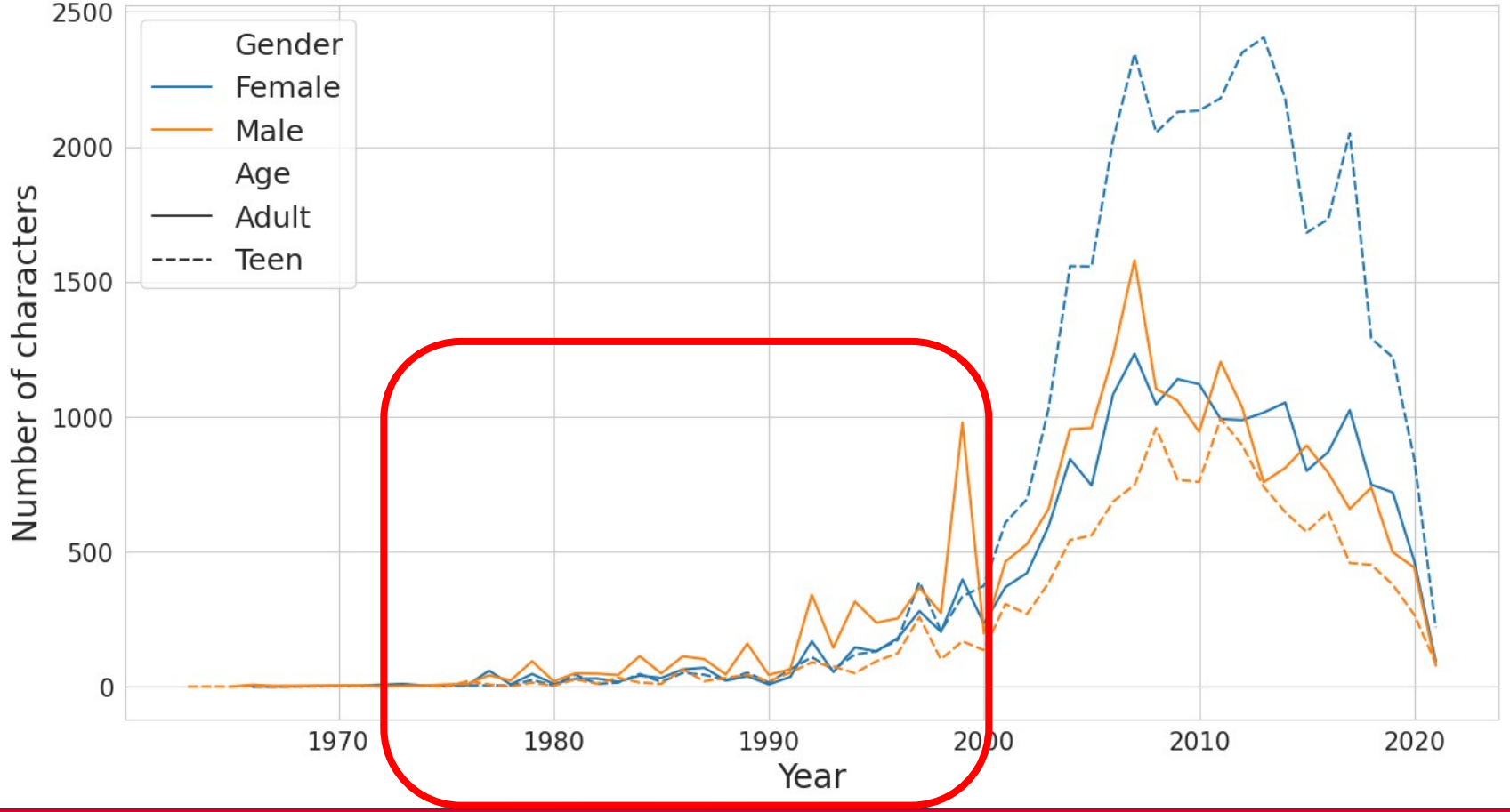
HOCHSCHULE DER MEDIEN

Distribution of ACDB characters by gender, age and year



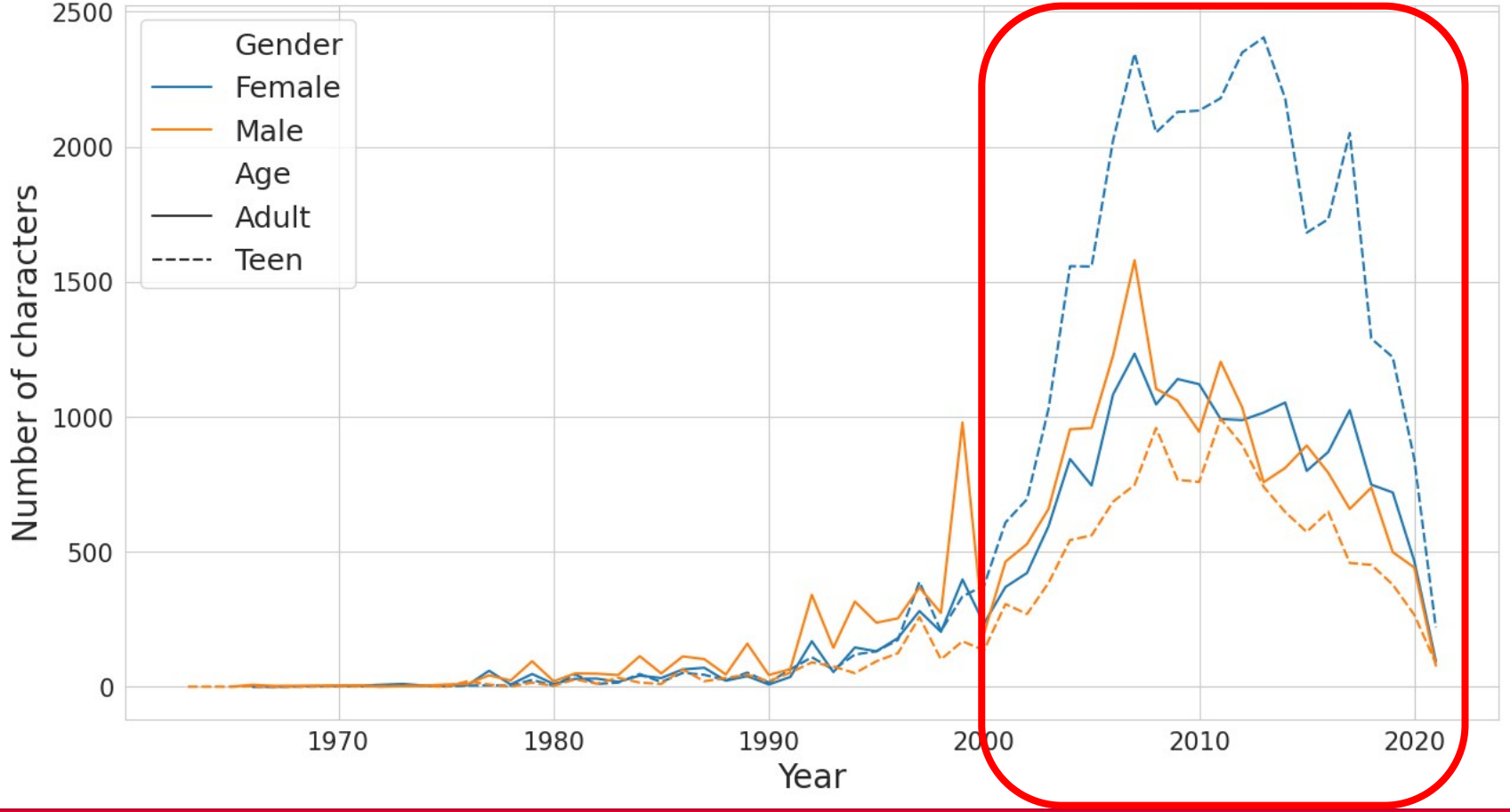
HOCHSCHULE DER MEDIEN

Distribution of ACDB characters by gender, age and year

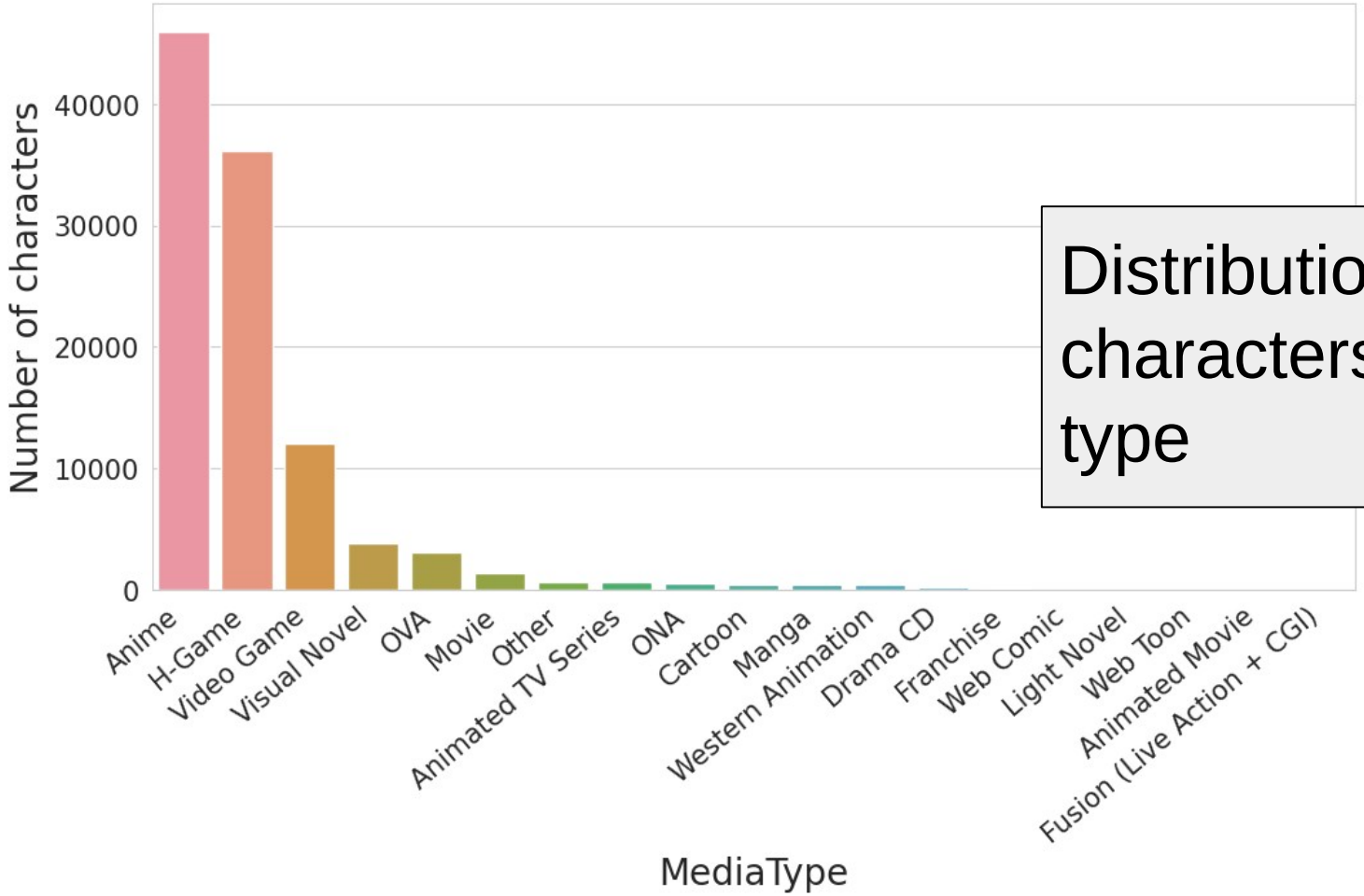


HOCHSCHULE DER MEDIEN

Distribution of ACDB characters by gender, age and year



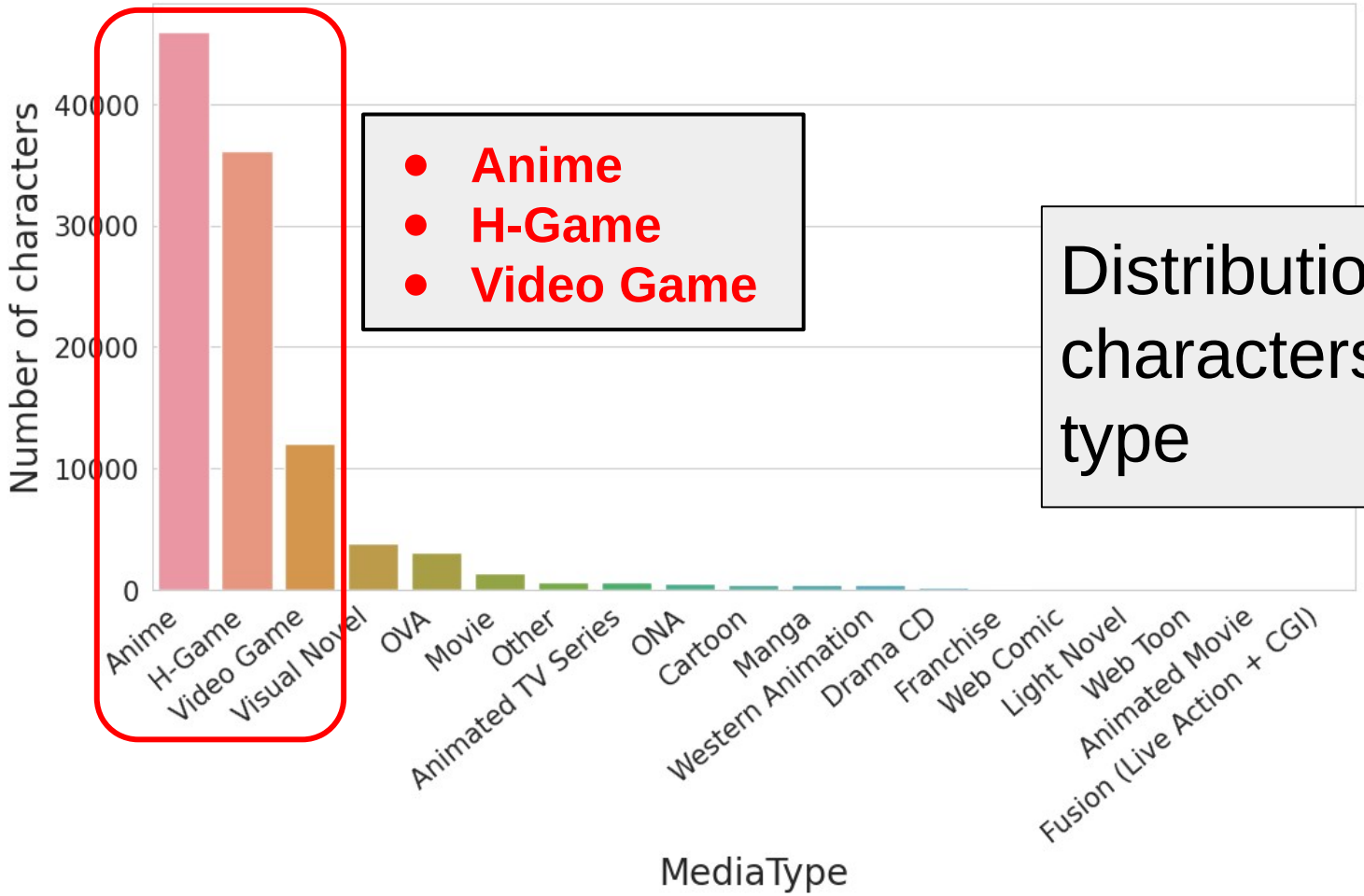
HOCHSCHULE DER MEDIEN



Distribution of ACDB characters* by media type

HOCHSCHULE DER MEDIEN

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



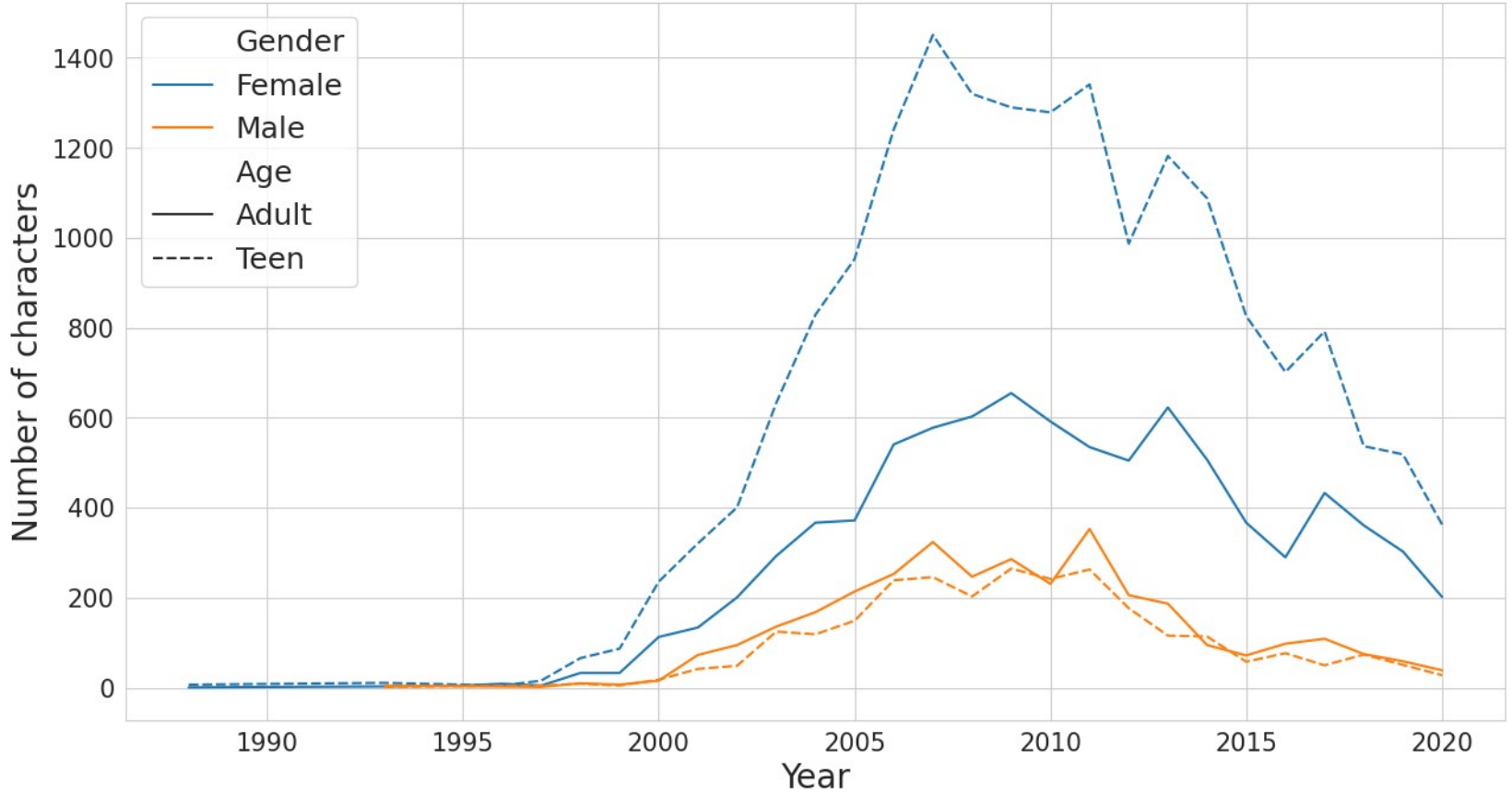
- **Anime**
- **H-Game**
- **Video Game**

Distribution of ACDB characters* by media type

HOCHSCHULE DER MEDIEN

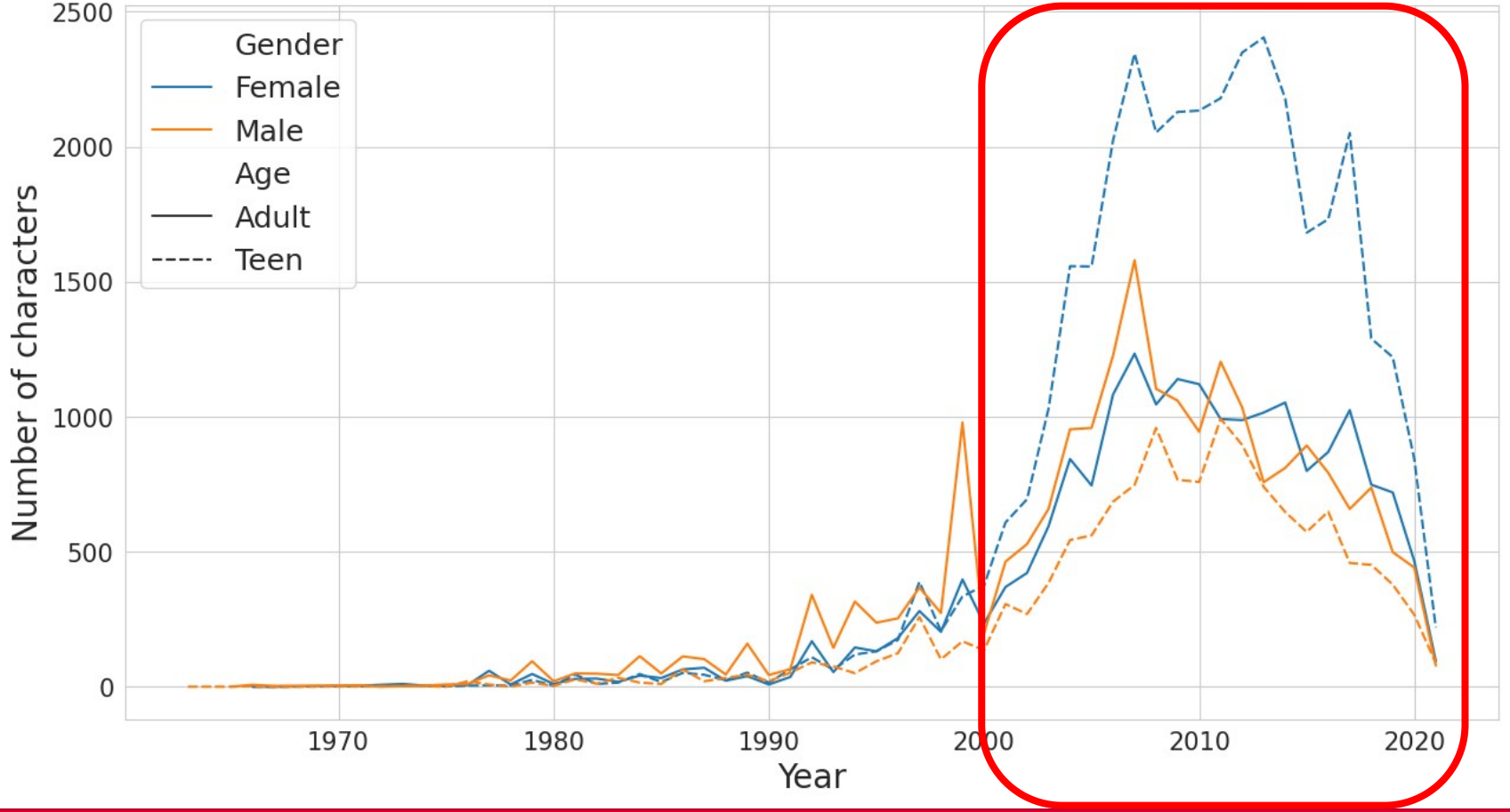
* Removed characters with publication year 2021 and/or with missing media type.

The effect of h-games



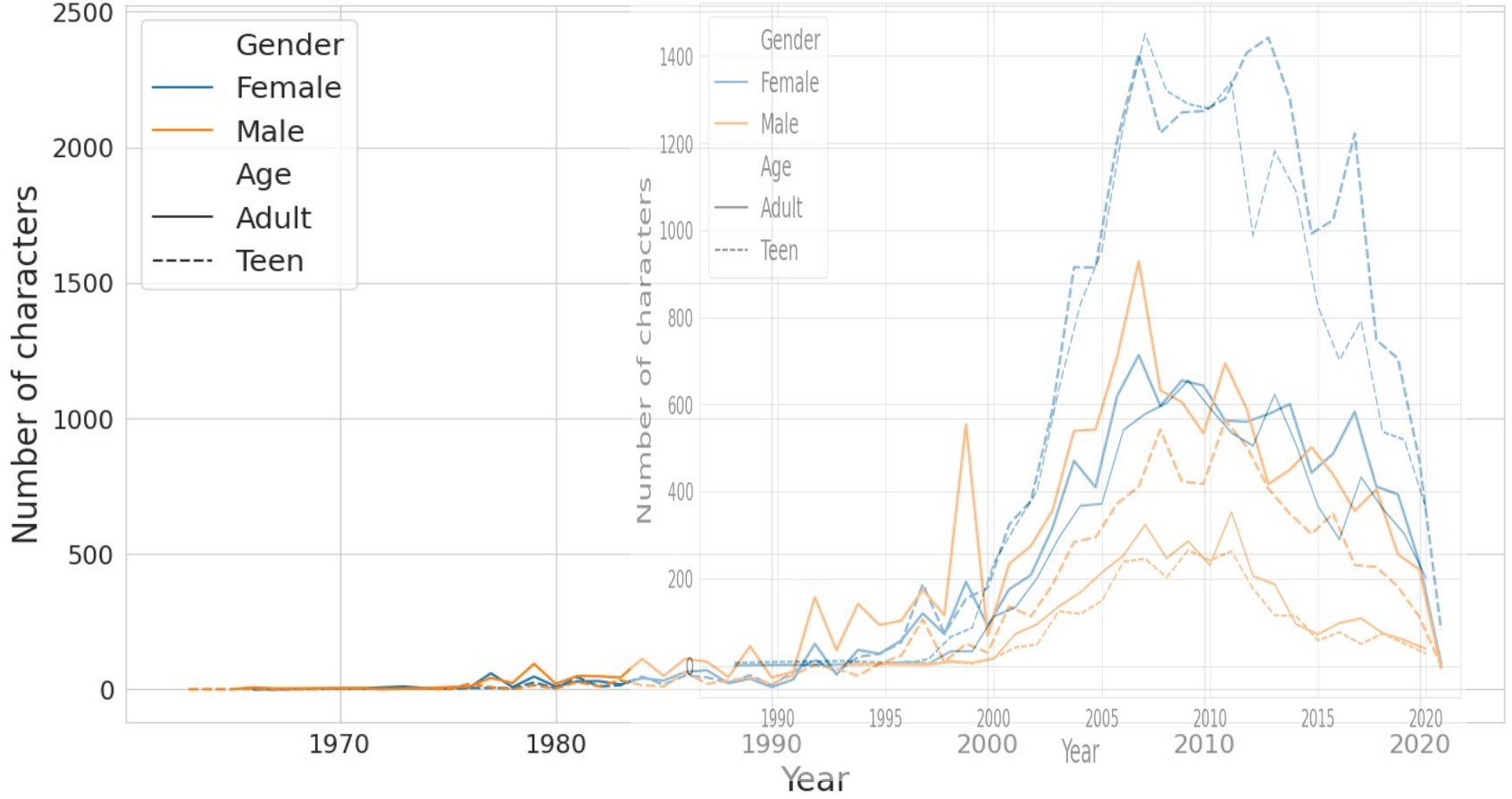
HOCHSCHULE DER MEDIEN

Distribution of ACDB characters by gender, age and year



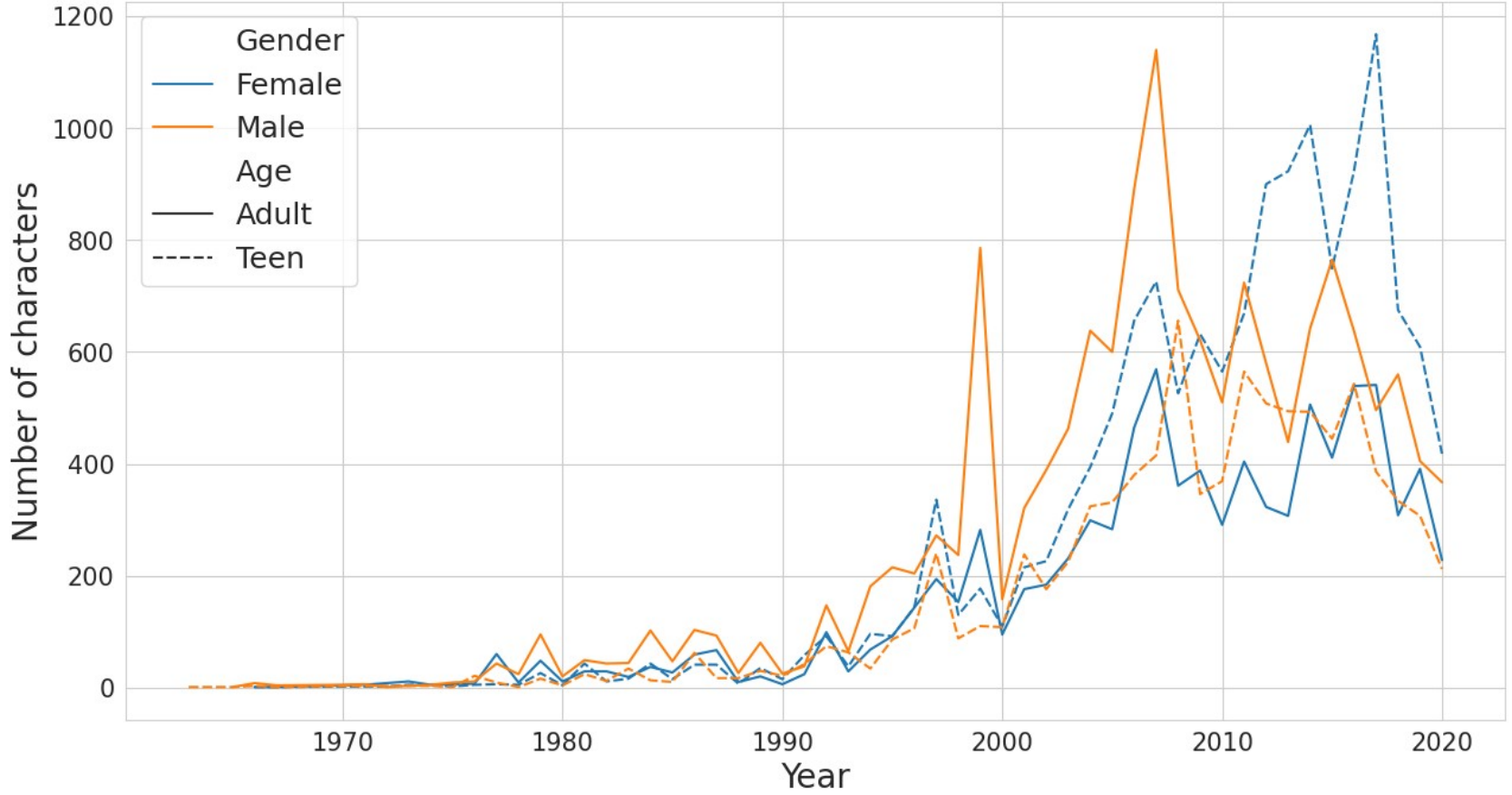
HOCHSCHULE DER MEDIEN

The effect of h-games



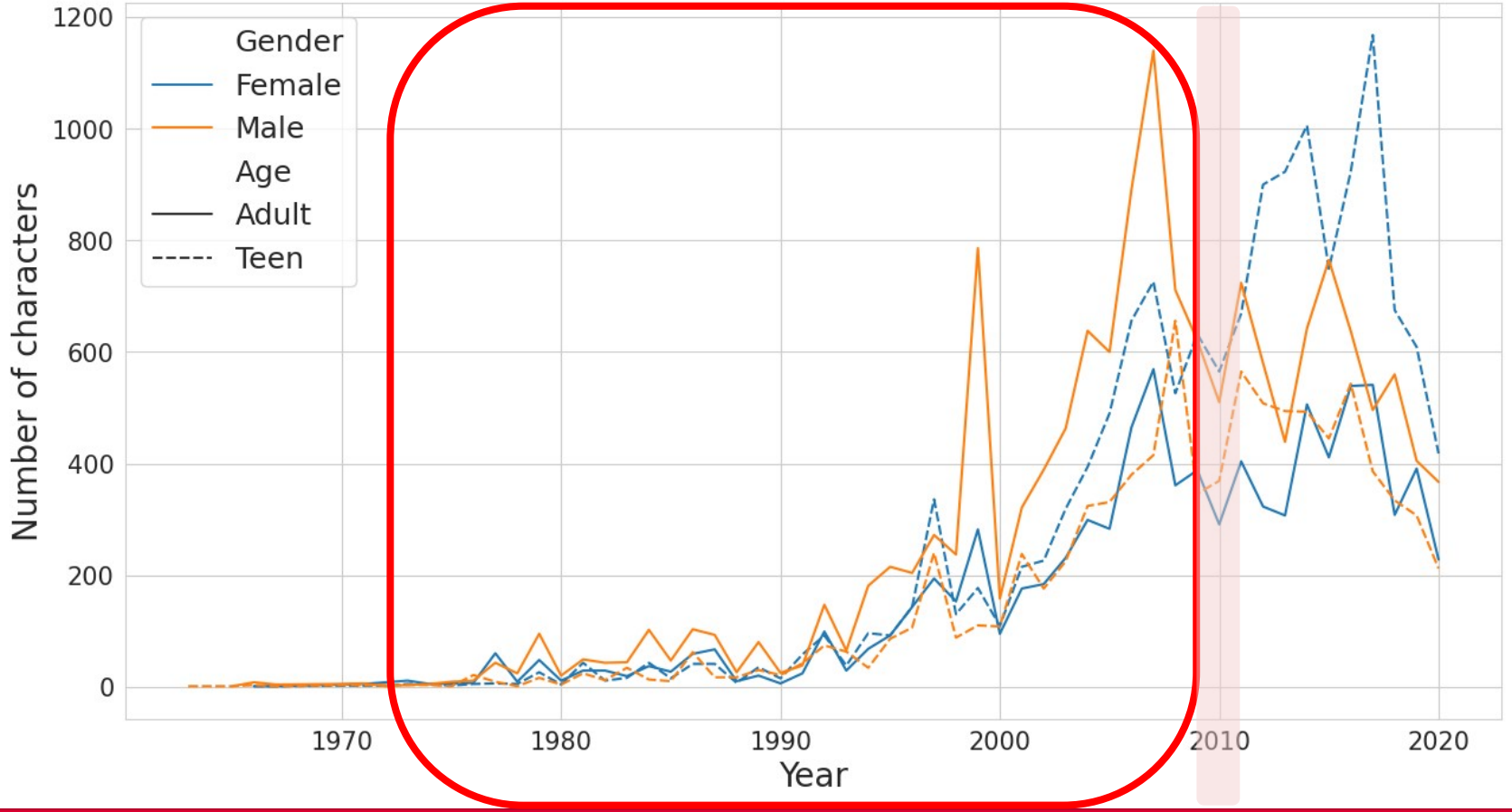
HOCHSCHULE DER MEDIEN

The anime and video games population



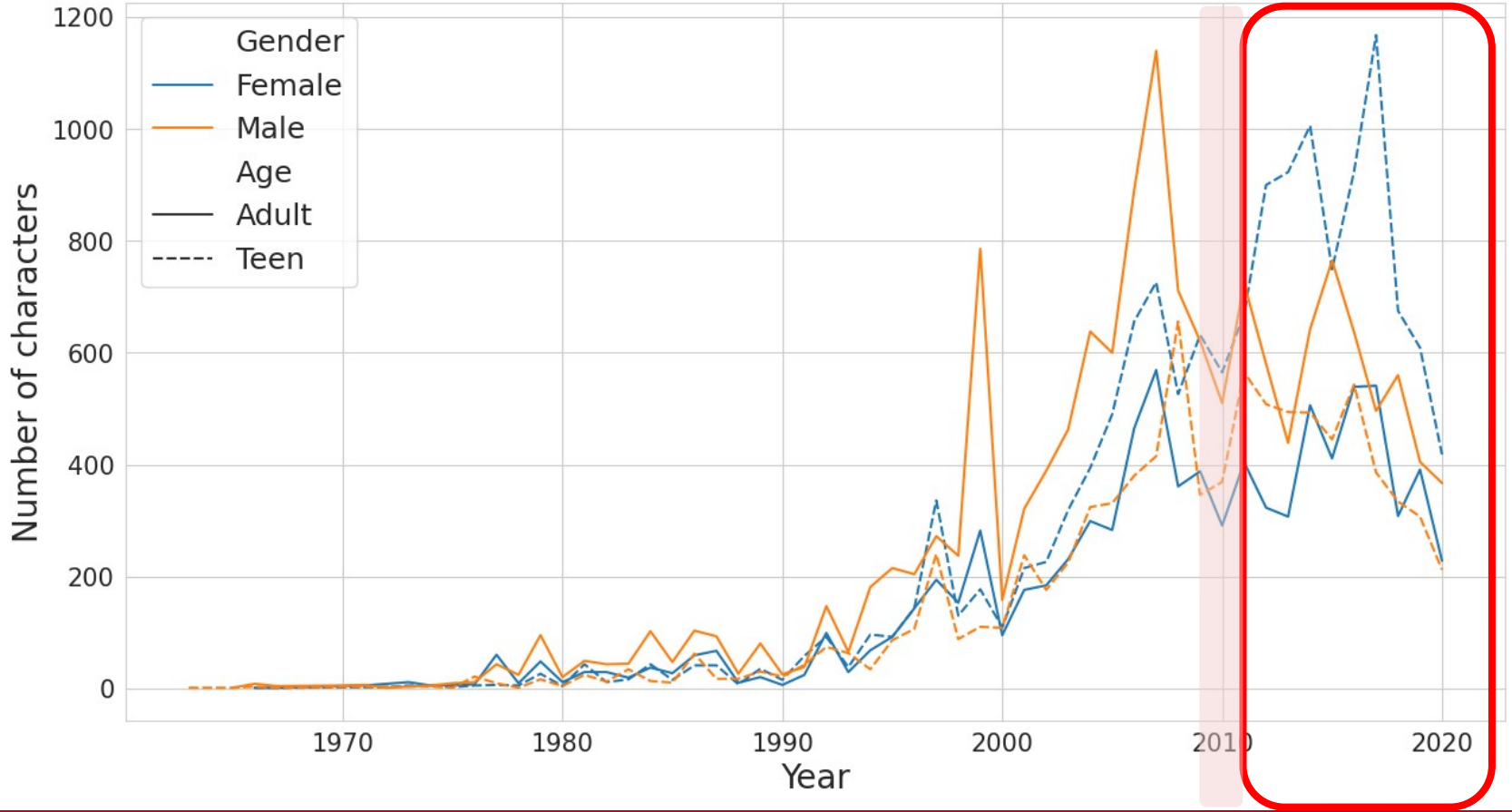
HOCHSCHULE DER MEDIEN

The anime and video games population



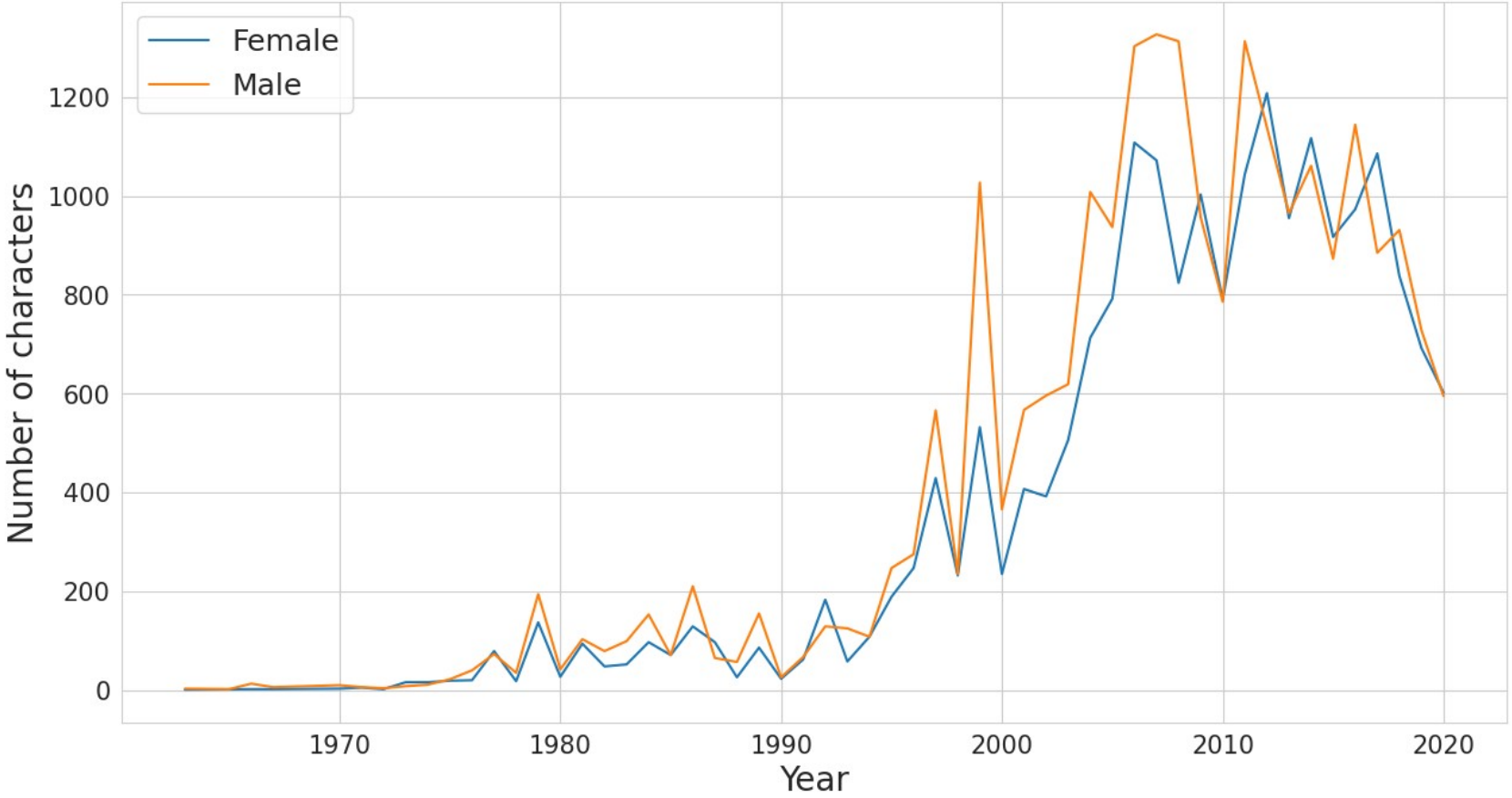
HOCHSCHULE DER MEDIEN

The anime and video games population



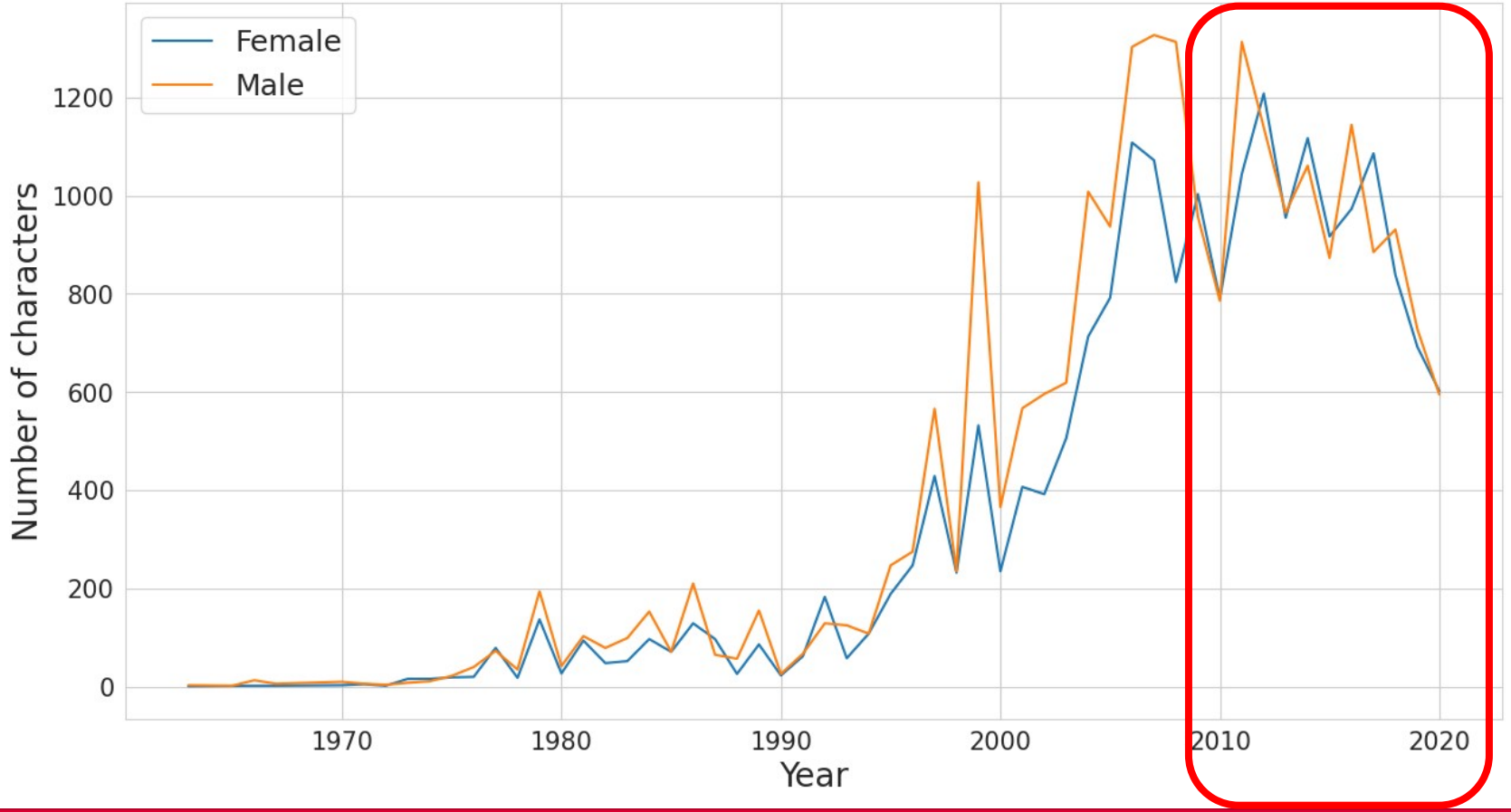
HOCHSCHULE DER MEDIEN

The anime population: gender



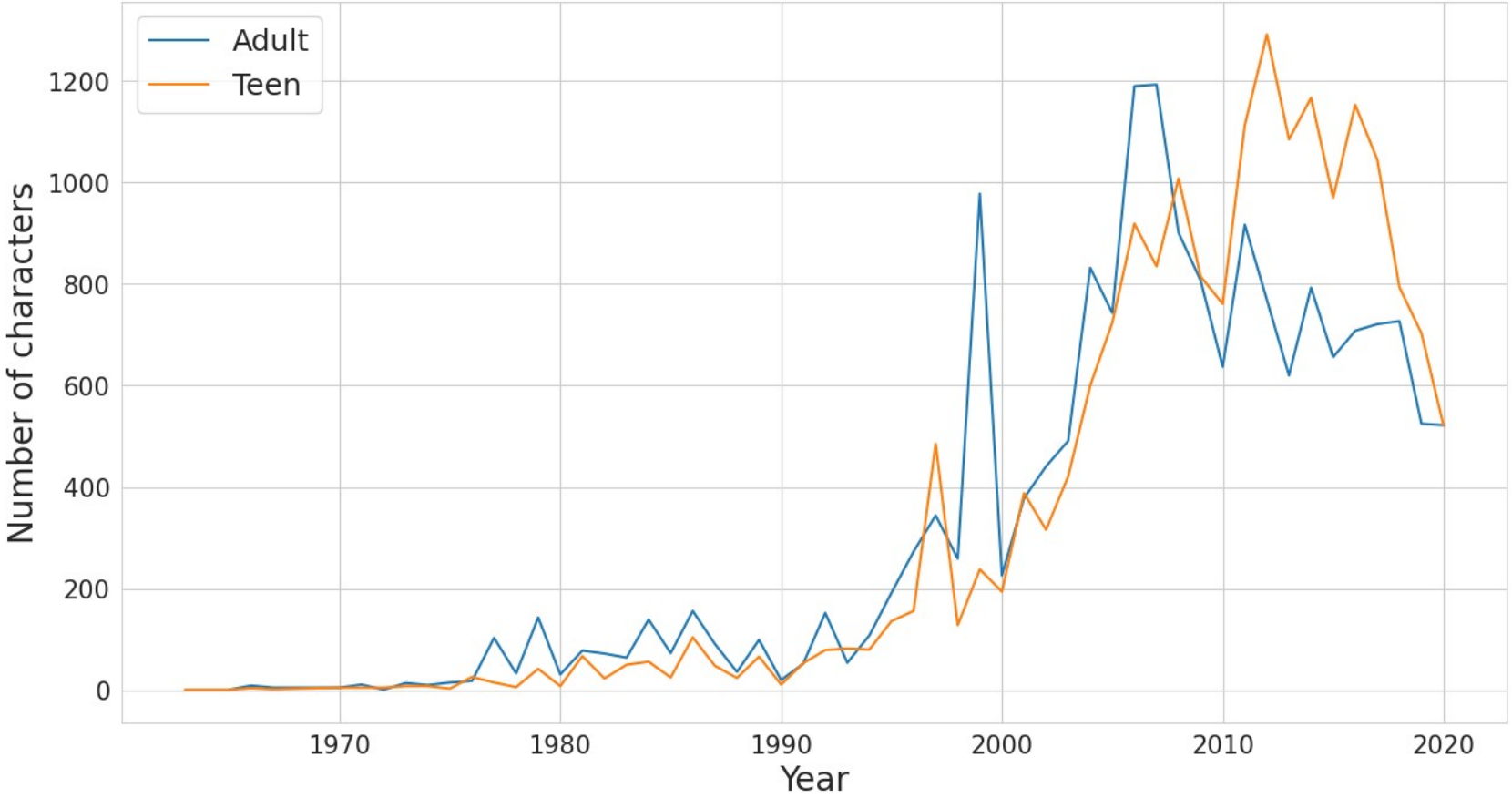
HOCHSCHULE DER MEDIEN

The anime population: gender



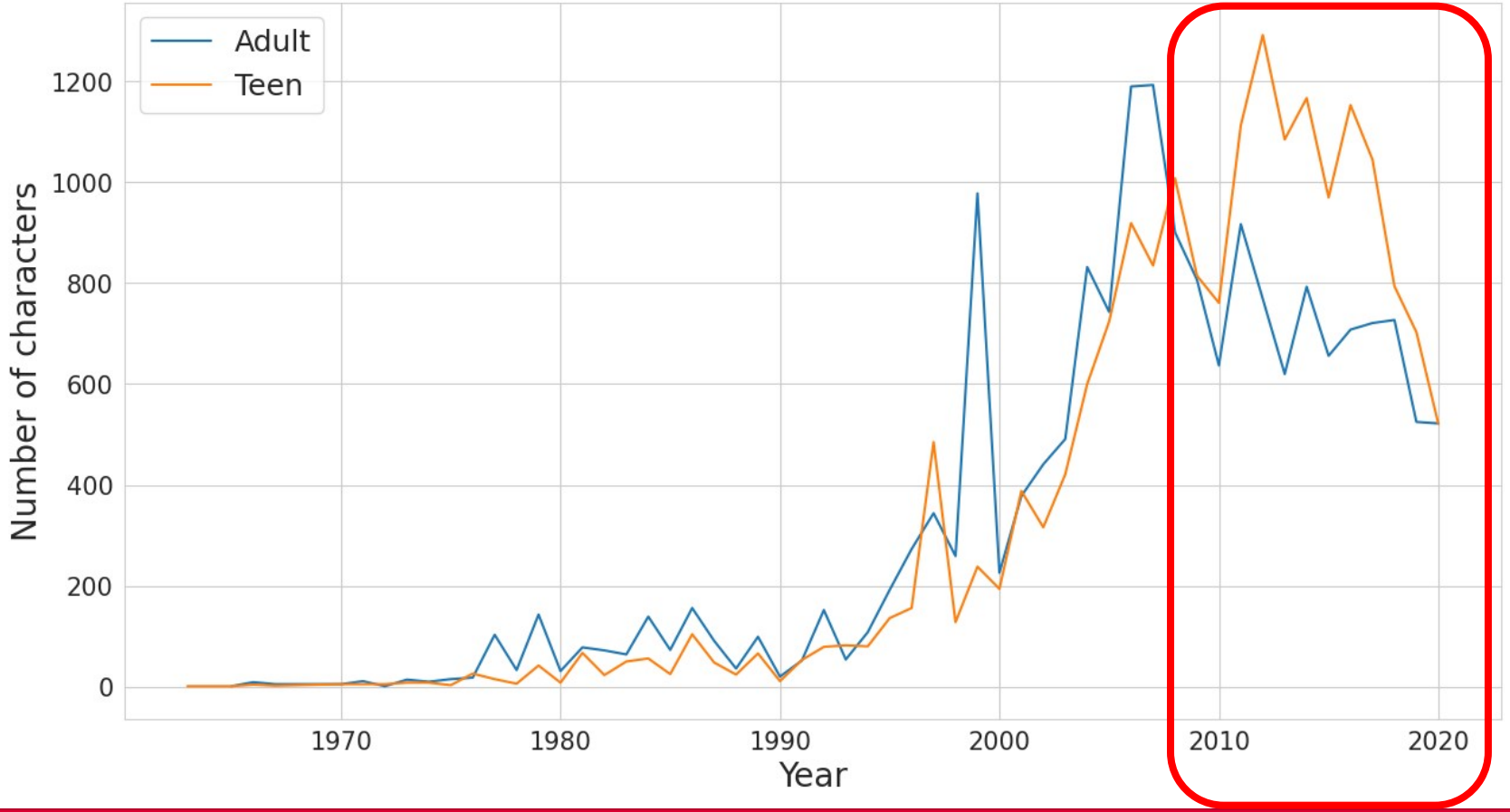
HOCHSCHULE DER MEDIEN

The anime population: age



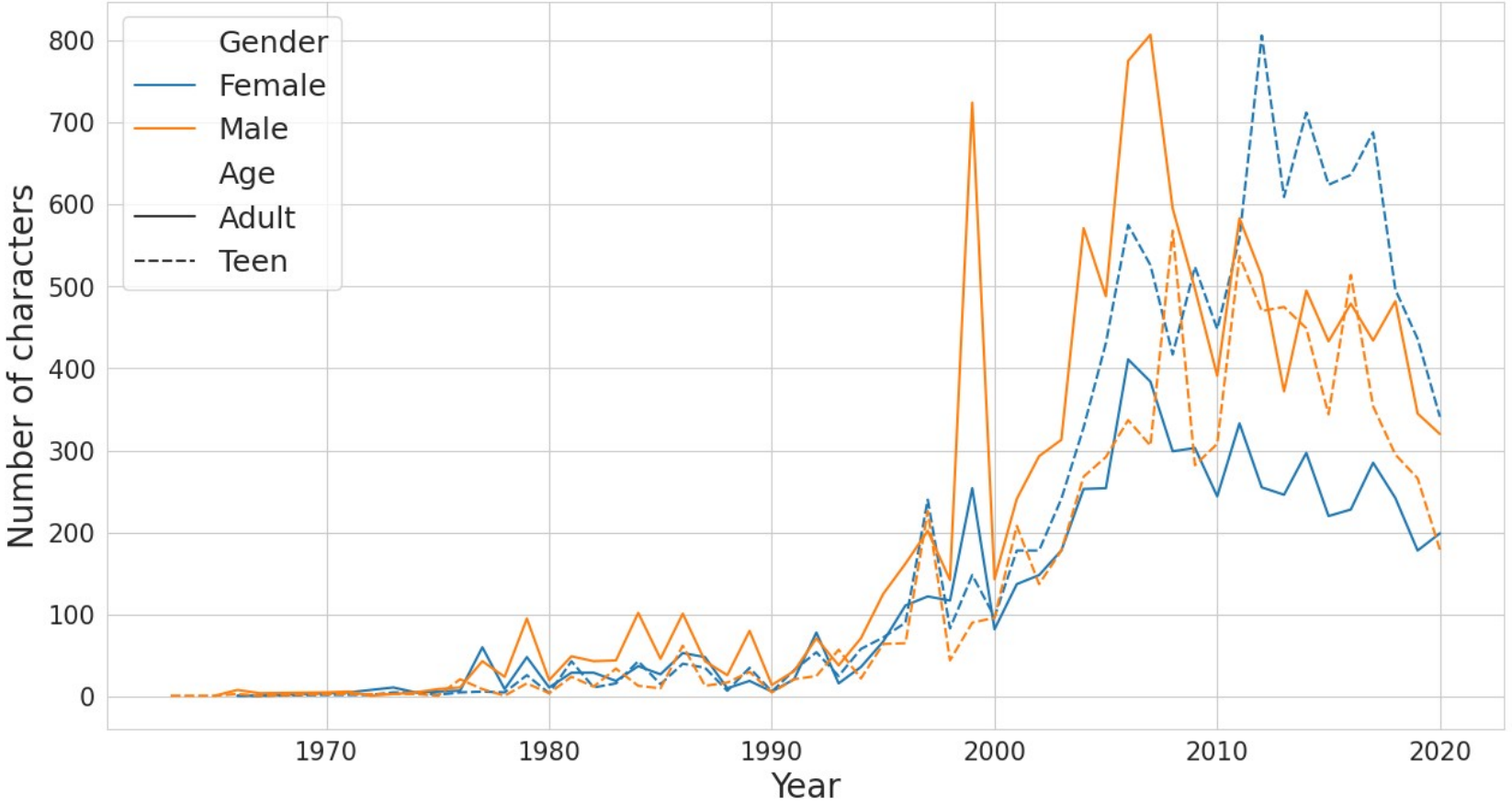
HOCHSCHULE DER MEDIEN

The anime population: age



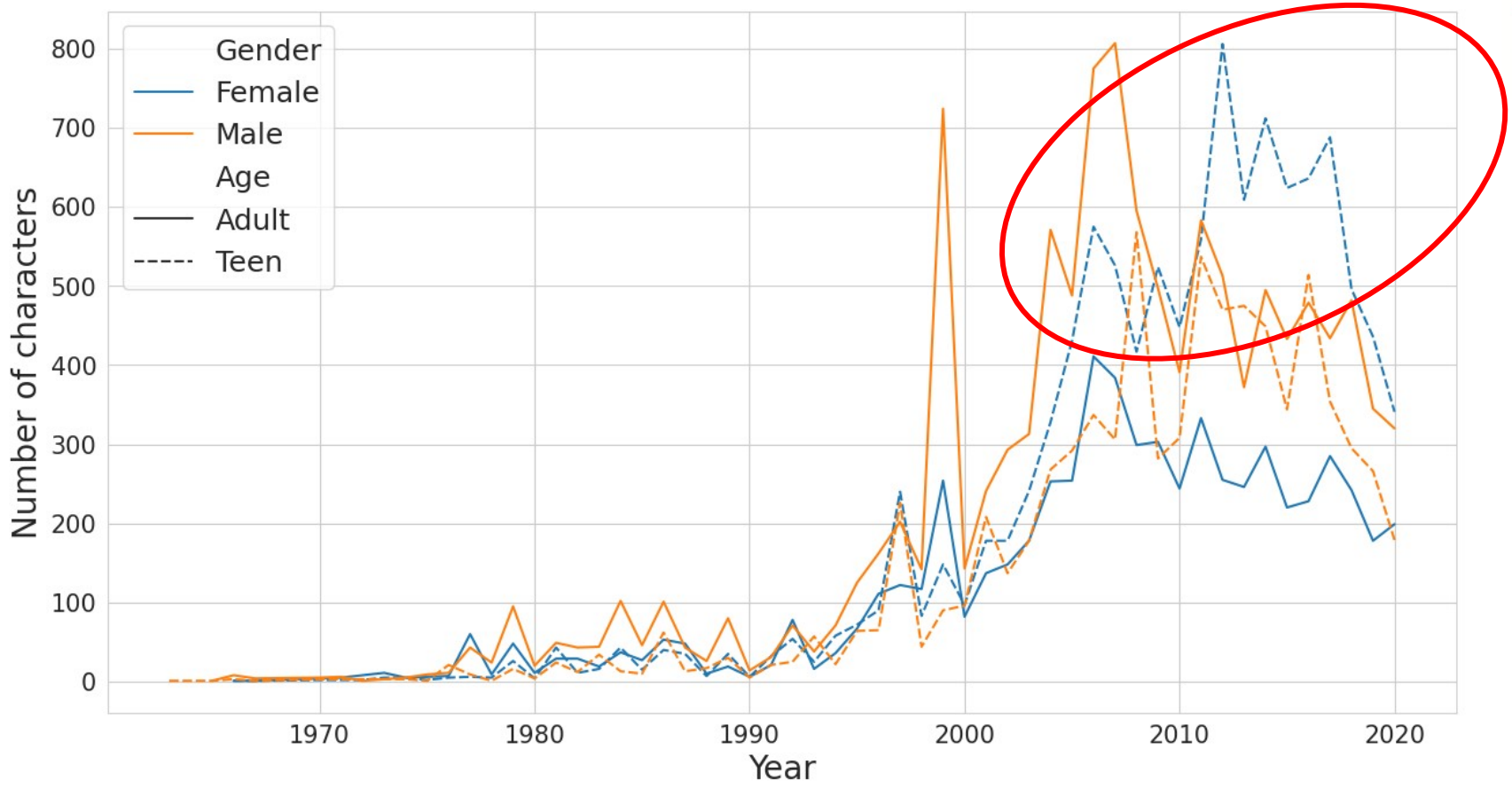
HOCHSCHULE DER MEDIEN

The anime population: gender and age



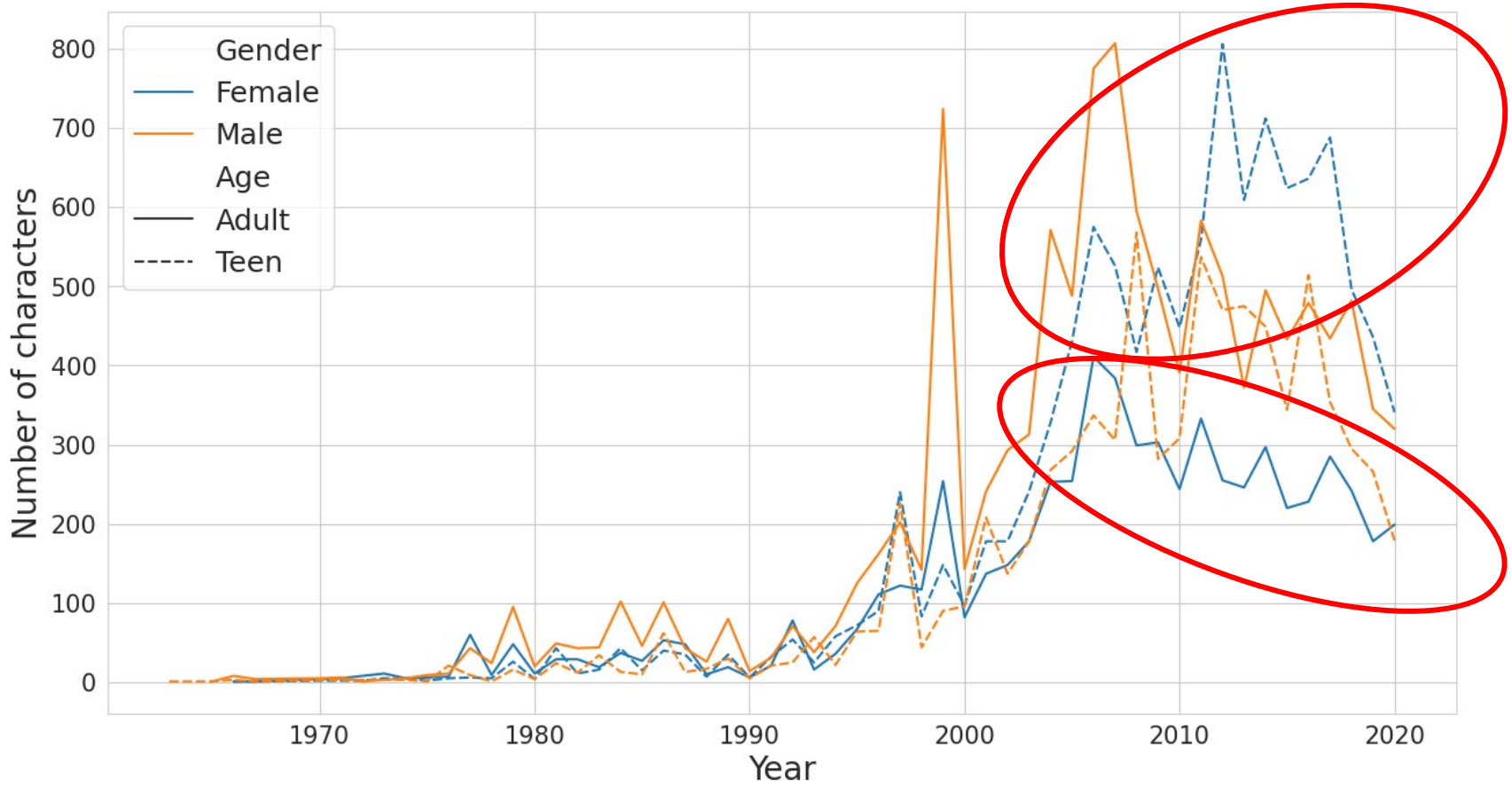
HOCHSCHULE DER MEDIEN

The anime population: gender and age



HOCHSCHULE DER MEDIEN

The anime population: gender and age



HOCHSCHULE DER MEDIEN

Summary of our findings so far

HOCHSCHULE DER MEDIEN

Who populates anime?

- Teen female characters become dominant from ~2010
- Gender and age group diversity is relatively low

What does the fictional population represent?

- “~~Reality~~” vs. Taste vs. Genre conventions vs. …

Moving beyond representation: What does the analysis imply?

HOCHSCHULE DER MEDIEN

- Changes in the **content** being produced
  - Potential influence of “otaku specific media”
- Changes in the **focus of the community** creating the database
  - The example of the video games character population
- An **interplay of both** of the above
  - Further evidence from smaller media type categories (e.g. OVA vs movie populations)

Database bias or actual change in anime?

- **Possible source of database bias 1:** Characters that are left out of the database
  - Random sample of 30-30 anime that are in the database with publication dates 2001-2009 and 2011-2019
- **Possible source of database bias 2:** Anime that are left out of the database
  - Random sample of 30-30 anime that are **NOT** in the database with publication dates 2001-2009 and 2011-2019
- **If the data were complete our results would be even more pronounced**

Actual change in anime

HOCHSCHULE DER MEDIEN

Considering the Japanese critical discourse

Correspondence with the Japanese critical discourse

- Tamaki Saitō (2000): Explosion of **“beautiful fighting girl”** works in the nineties
- Hiroki Azuma (2001): Rise of **database consumption** from the late nineties
- Gō Itō (2005): Gangan manga and the centrality of **kyara** from the nineties
- Satoshi Maejima (2010): **Sekaikei** and post-Evangelion works from the late nineties

Correspondence with the Japanese critical discourse

- Tamaki Saitō (2000): Explosion of **“beautiful fighting girl”** wor
- Hir **n** from the late
- Gō **vara** from the
- Satoshi Maejima (2010): **Sekaikei** and post-Evangelion works from the late nineties

A shift has happened during the mid- to late nineties

Hypothesis for the developments highlighted by both the theorists and the large-scale data

HOCHSCHULE DER MEDIEN

1. **Changes in individual works**, and **new trends** becoming increasingly stronger from the second half of the nineties onward
2. **Strong impact of “otaku specific media”** during the naughties
3. **Shift increasingly felt in the whole of anime production** from the end of the naughties

Large-scale quantitative analysis together with case study based qualitative analysis

HOCHSCHULE DER MEDIEN

- Analyses in the Japanese critical discourse
  - Help **guide quantitative research**
  - Help **interpret results** from quantitative research
- Results from large-scale quantitative analysis
  - Help **test/verify claims** from case-study based analyses
  - Help **provide further starting points** for theorization
- Together the two approaches help paint a better picture of the phenomena being examined

Thank you for your attention!

Get in touch with us:

Email: kacsuk@hdm-stuttgart.de

Website: https://mediagraph.link/