

Revisiting “Otaku: Japan’s Database Animals” twenty years on

Can we use fan created databases to test
one of Hiroki Azuma’s central arguments?

Research funded by:

 Deutsche
Forschungsgemeinschaft
German Research Foundation

Zoltan Kacsuk
JADH2020
Japanese Association for Digital Humanities
Osaka University
November 20, 2020 – November 22, 2020

Outline of the presentation

1. Introducing the **JVMG project**
2. A sample **tiny use case**: Testing one of Hiroki Azuma's points from *Database Animals*
 - i. Hypothesis
 - ii. Data sources
 - iii. Descriptive statistics
 - iv. Regression analysis
3. Lessons learned

Introducing the JVMG project

Introducing the JVMG project

- **Media fans** have always been **compiling data**
- With the internet this **became easier and amplified in its scale**
- Databases by enthusiast communities are the **go to resource for checking information**
- The use of these databases for **large-scale quantitative research**

Introducing the JVMG project

- **Japanese Visual Media Graph (JVMG) project**
- Funded by the **German Research Foundation's** (Deutsche Forschungsgemeinschaft, DFG) e-Research Technologies program

Introducing the JVMG project

- Using **fan created databases for academic research**
- Employing a **linked data** framework
- Using the **Tiny Use Case** workflow methodology (pioneered by diggr.link)

Introducing the JVMG project

- Data is **cleaned and preprocessed**
- **Unified framework** for accessing the databases
- **Unified frontend** for browsing the data
- **Connections** between the databases
- **Can be connected** to other linked data sources

A sample tiny use case:
Testing one of Hiroki Azuma's points
from *Database Animals*

Otaku: Japan's database animals

- 東浩紀、2001: 動物化するポストモダン—オタクから見た日本社会（講談社現代新書）
- Hiroki Azuma, 2009: *Otaku: Japan's database animals* (University Of Minnesota Press)
- On the **production and consumption paradigm** defining Japanese anime, manga, light novels and games in late modernity.



Formulating a hypothesis

- The **metaphor of the database** captures the space of imagination shared by creators and consumers
- “As a result, many of the otaku characters created in recent years [the late nineties to two thousand] are connected to many characters across individual works, rather than emerging from a single author or a work.” (Azuma, 2009: 49)
- The **hypothesis** to test: **The portion of new characters with shared traits should increase over time.**

Data sources

1. *The Visual Novel Database* (VNDB) vndb.org

- only focuses on visual novels
- very rich and rigorously structured ontology of traits
- **79.000** number of characters

2. *Anime Characters Database* (ACDB) www.animecharactersdatabase.com

- collects data on a wide range of characters from various media (although predominantly focusing on visual novels and anime)
- hybrid system for describing characters:
 - a closed ontology for eight flagship traits
 - opportunity for free-form tagging of characters with user created labels
- **100.000** number of characters

ACDB character traits example

Goku from *Dragon Ball* has the tags:

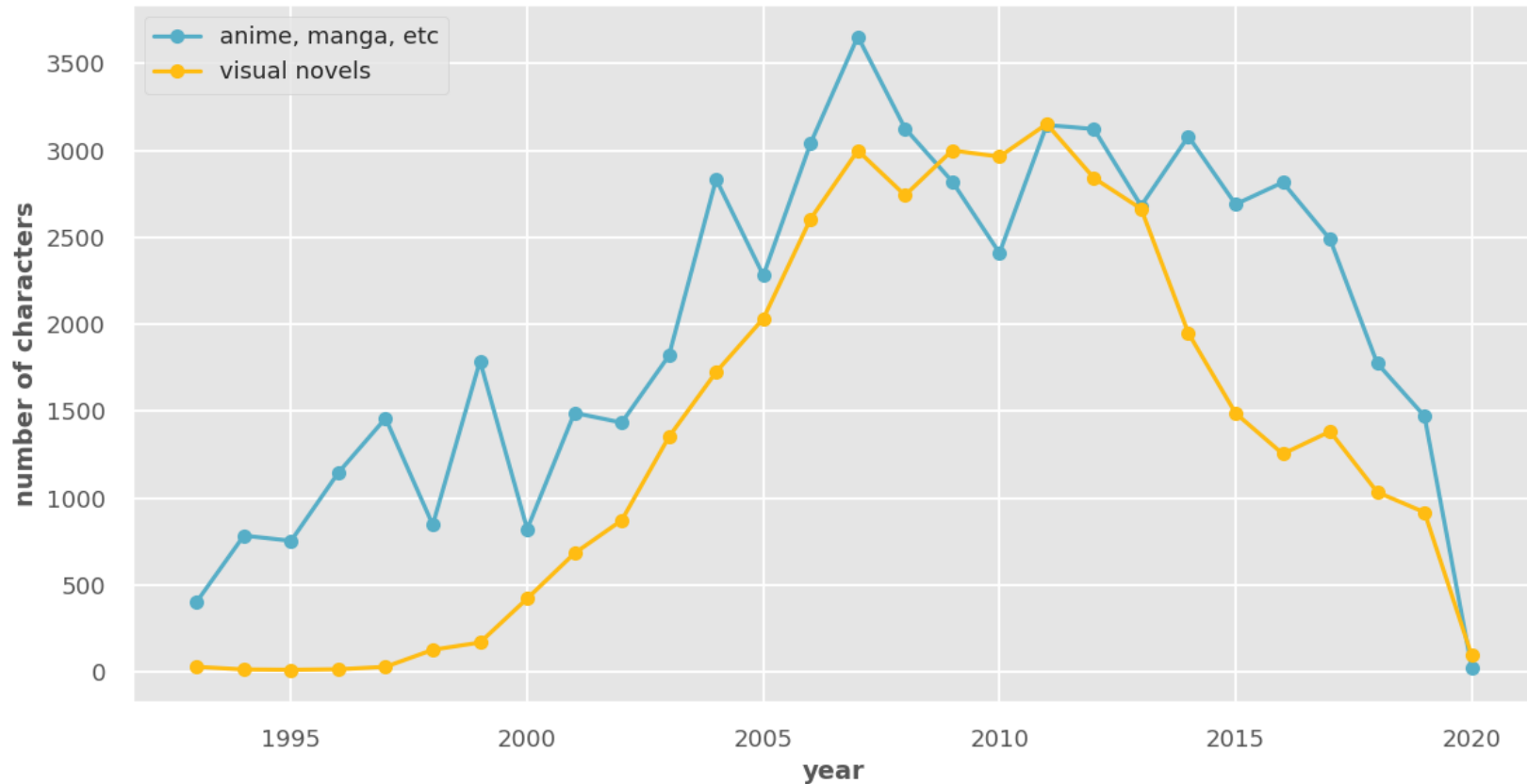
- **Arm guards** (shared with **262** other characters)
- **Karate** (shared with **40** other characters)
- **Spiky hair** (shared with **887** other characters)

Operationalization of our concepts

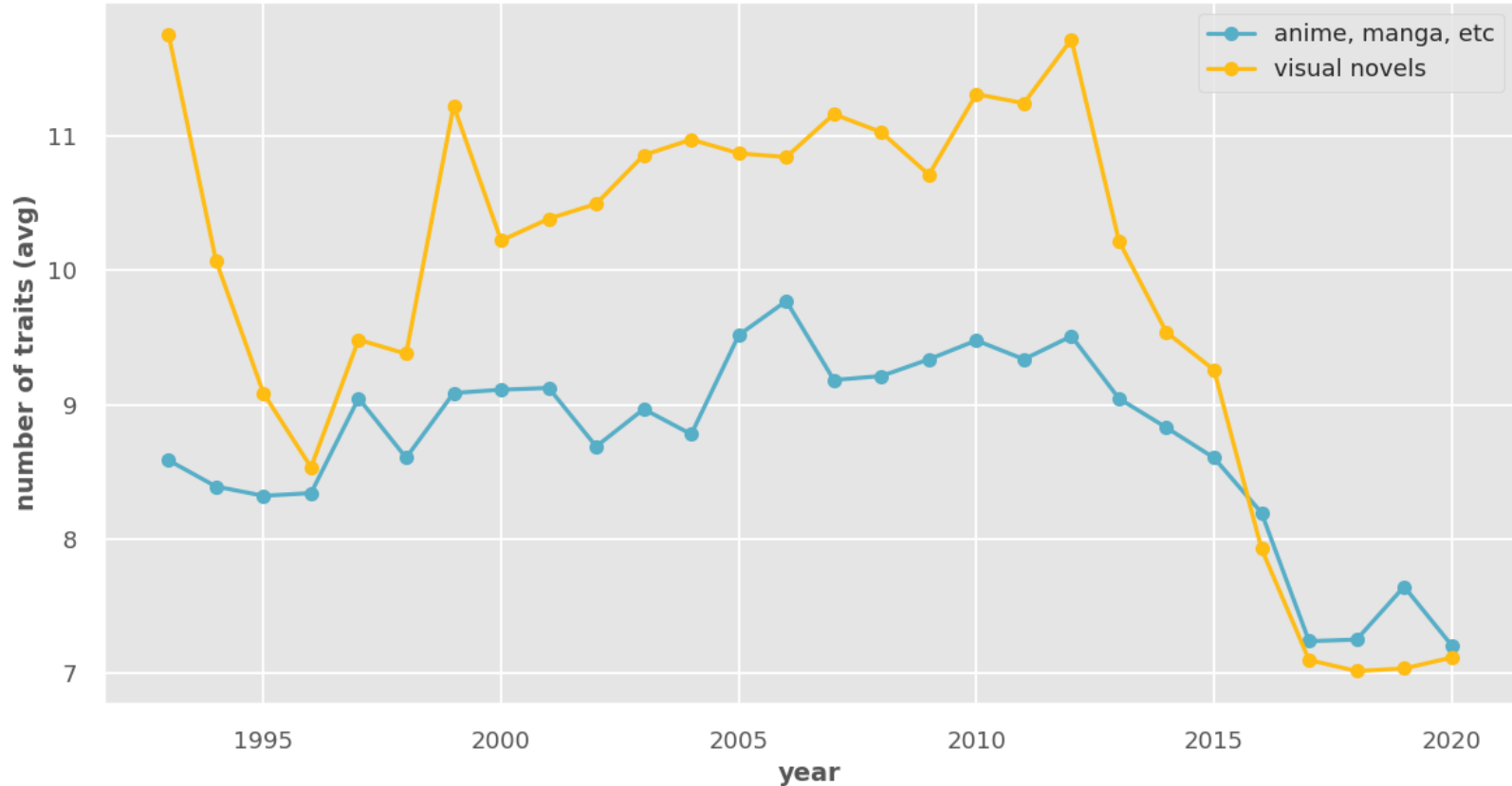
- **minimum number of five shared traits** needed to consider two characters to have shared traits
- **one year window** is the time frame for which we compare characters
- to allow for a better comparison between the VNDB and ACDB data, the **ACDB data was separated into two datasets:**
 - **visual novels**
 - **other works** (the majority is anime)

Descriptive statistics (ACDB only)

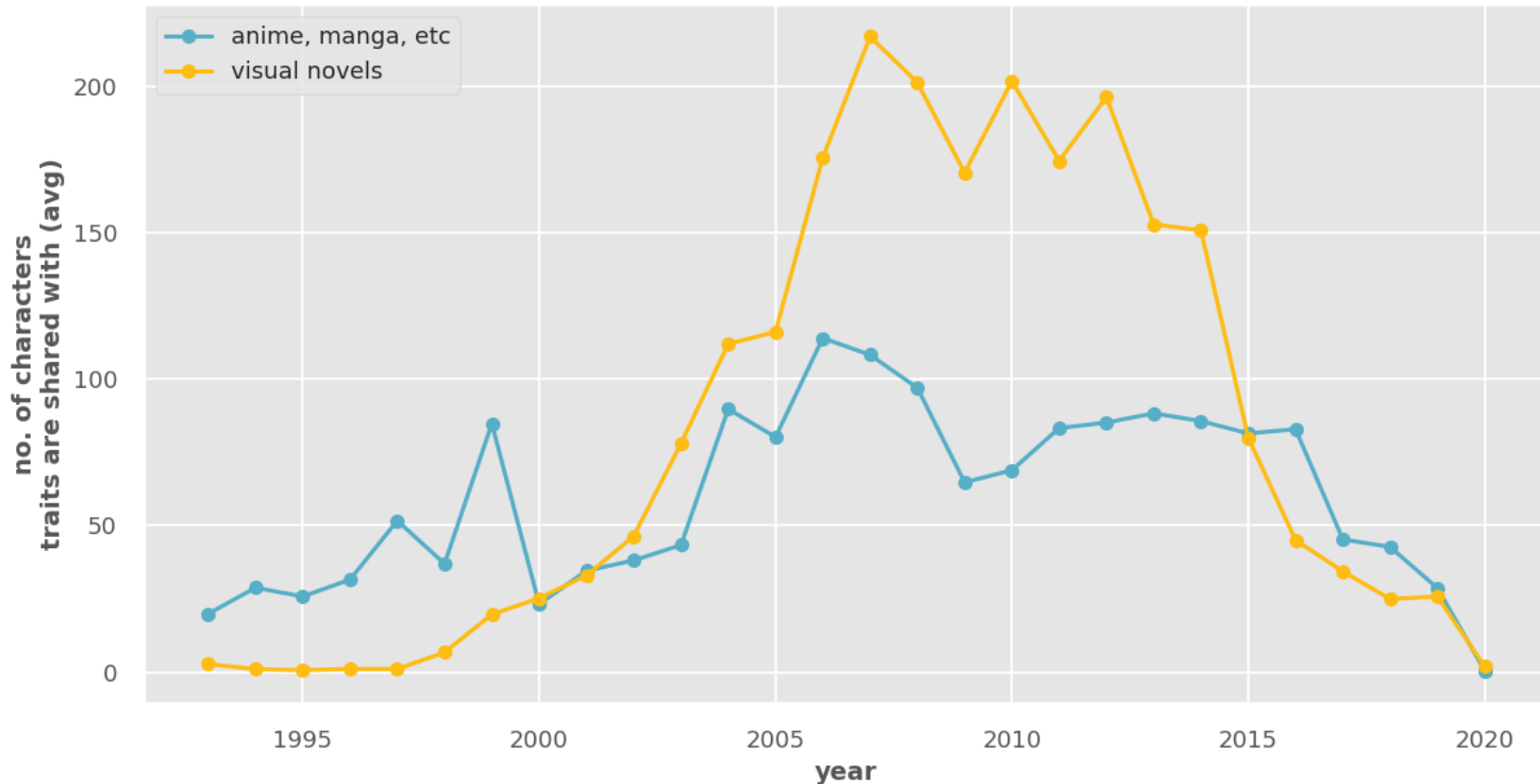
Number of characters in the ACDB dataset by year and type



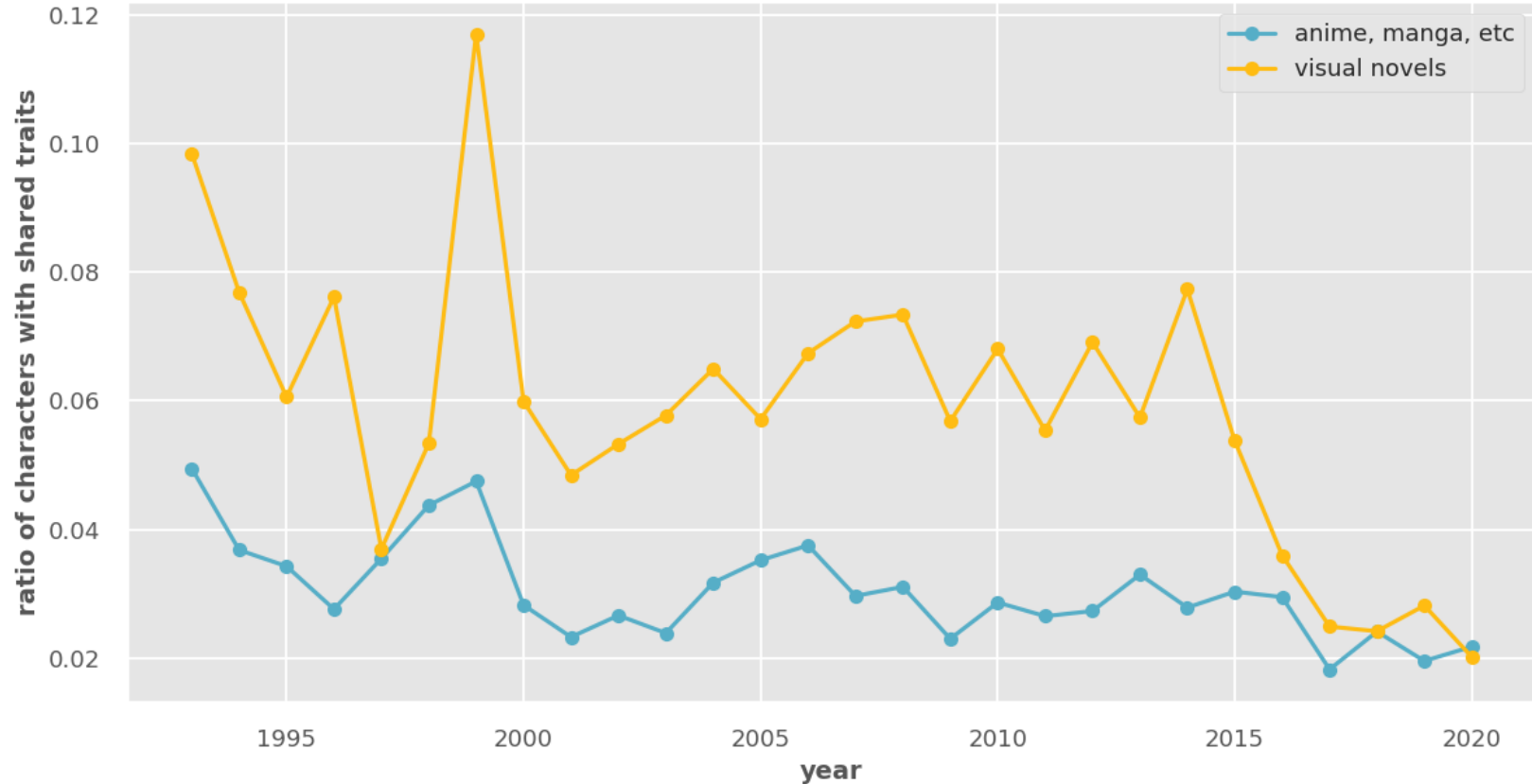
Average number of traits in the ACDB dataset by year and type



Average number of characters traits are shared with in the ACDB dataset by year and type



Ratio of characters with shared traits by no. of all characters in range in the ACDB dataset by year

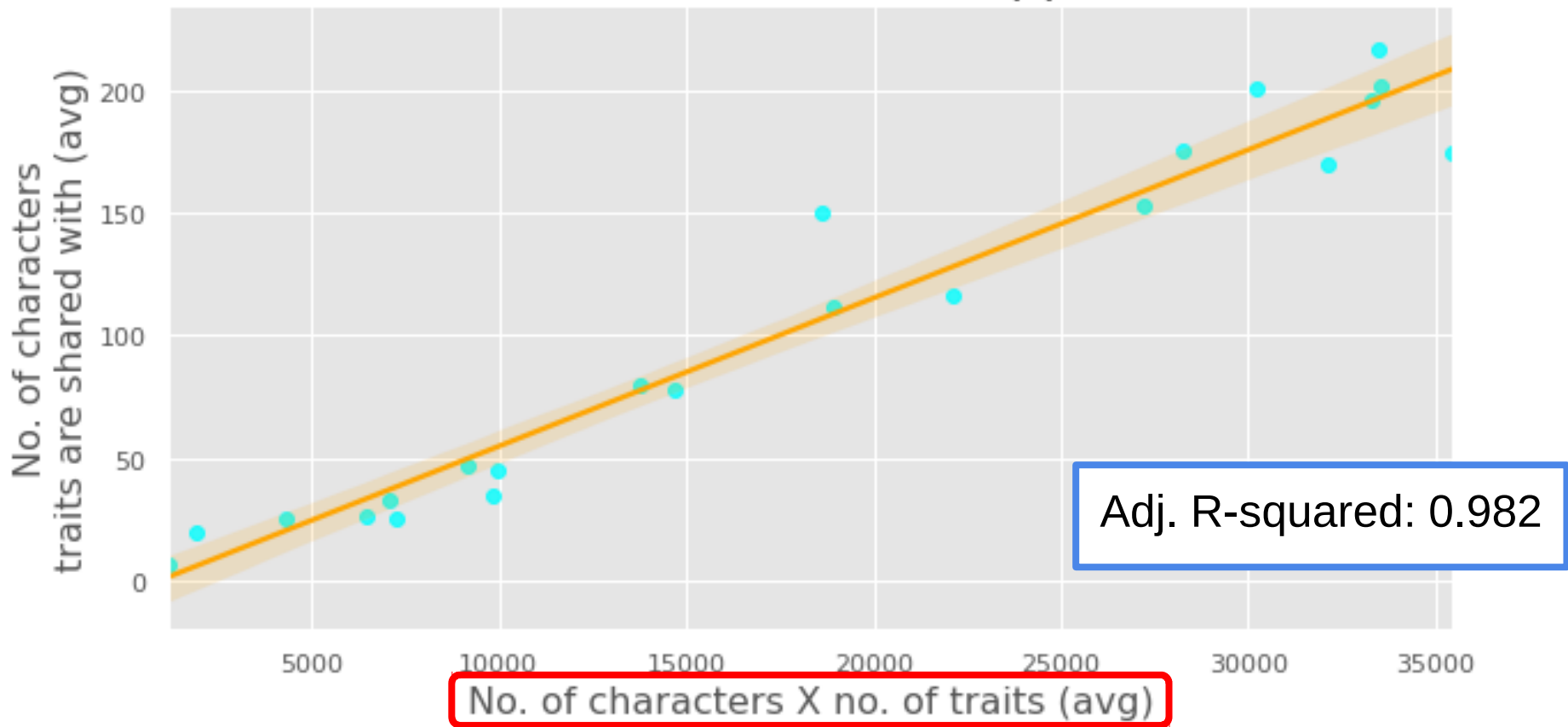


Regression analysis (ACDB only)

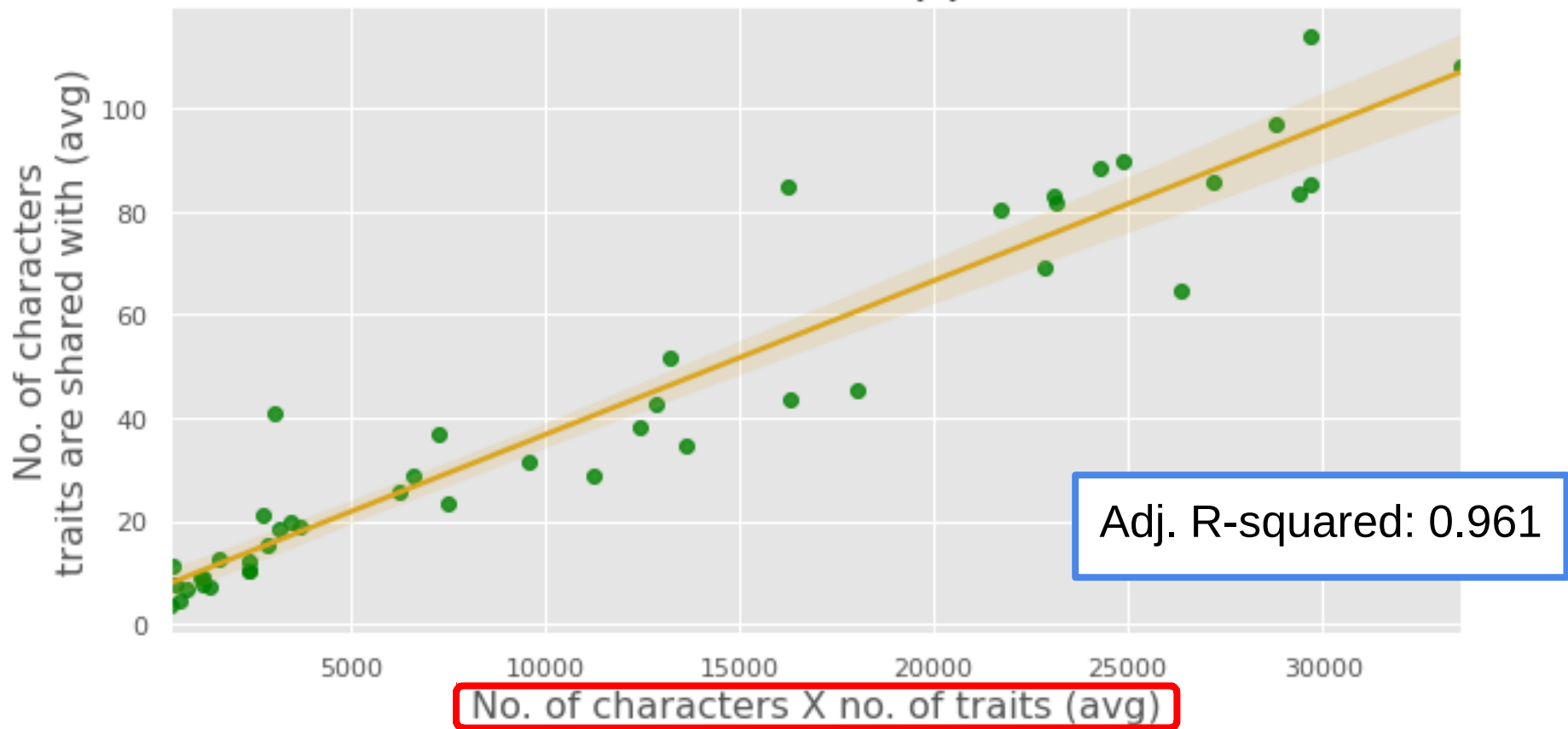
Regression analysis

- **Dependent variable:**
 - *average number of characters traits are shared with*
- **Independent variables:**
 - *number of characters*
 - *average number of traits*
 - *year*
- Also included in the model building and selection process:
 - **Squared terms** for independent variables
 - **Interaction terms** between independent variables

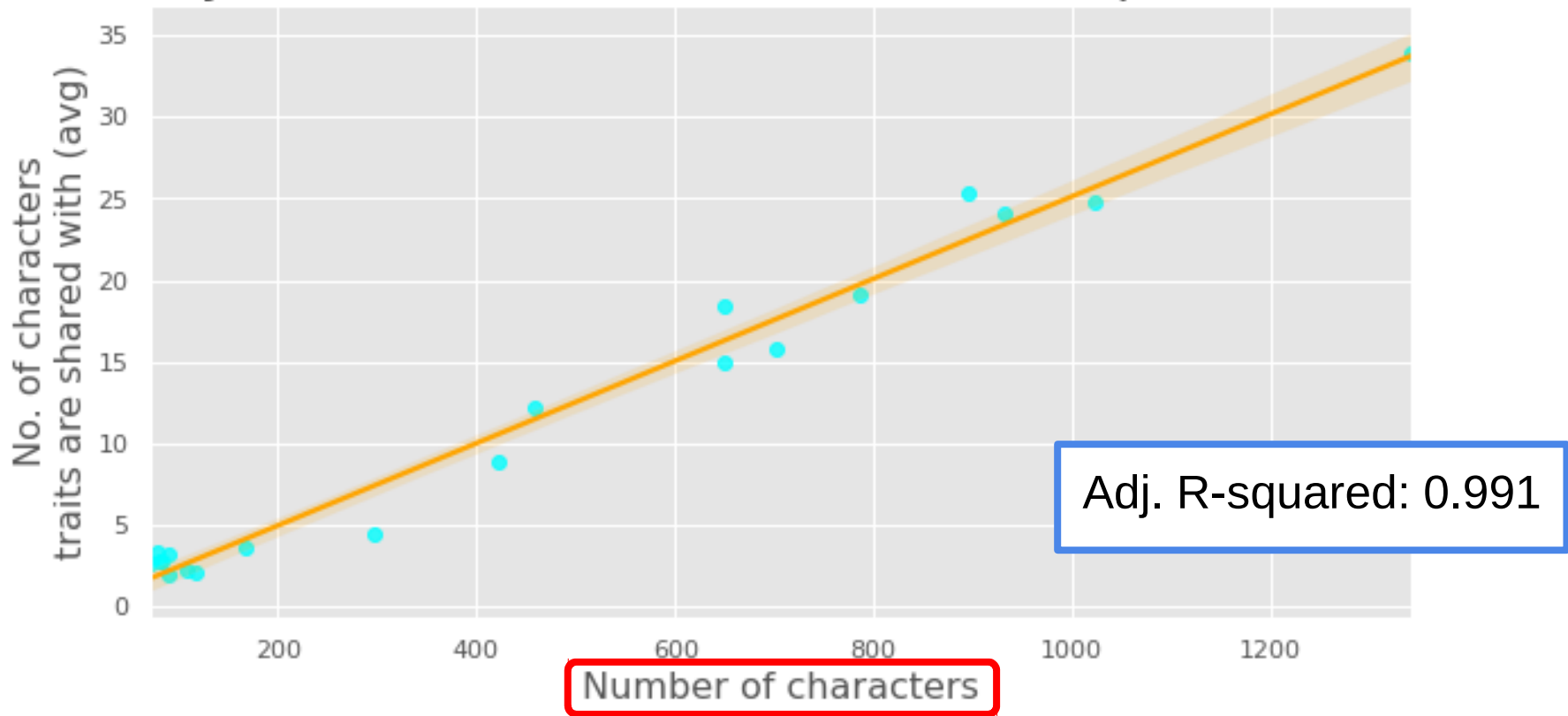
Regression results for ACDB data for visual novel characters with first appearance 1998-2019



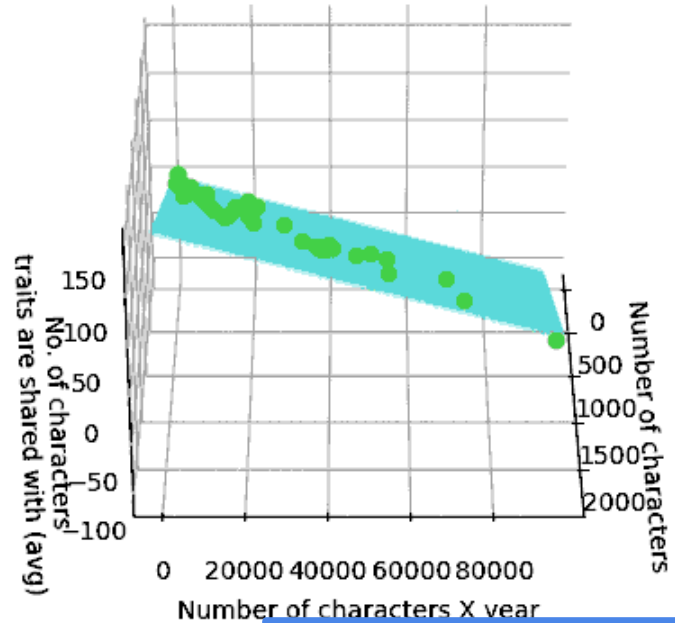
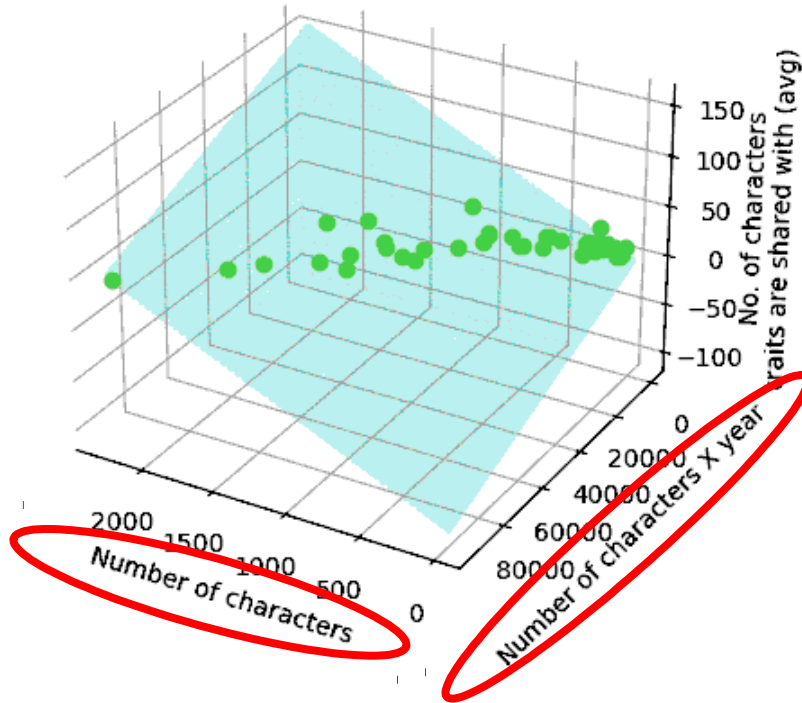
Regression results for ACDB data for anime & other non-VN characters with first appearance 1975-2019



Regression results for ACDB data for visual novel characters with first appearance 2001-2020 (only characters with number of traits equal to seven)



Regression results for ACDB data for anime & other non-VN characters with first appearance 1977-2019 (only characters with number of traits equal to seven)



Adj. R-squared: 0.948

Summary of regression results

The **hypothesis** (the portion of new characters with shared traits should increase over time) **was not substantiated** by our regression analyses:

- In the **ACDB visual novel characters** dataset **no connection** was found with the temporal variable
- In the **ACDB other characters** dataset a **small but negative relationship** with the temporal variable was found
- In the **VNDB dataset** there might be a slight positive relationship with the temporal variable, but **most likely there isn't** one

Implications and lessons learned

Implications of the results for theory

- **Cannot fully rule out that the hypothesis might still be correct**, however, based on the results it is time to seriously consider the alternative
- Without this whole analysis it would not have occurred to me that **there is maybe no paradigm shift going on on the production side**

Lessons learned

- The **datasets are not complete**
- It is **still possible** to use the available data **to test research questions**, thanks to:
 - The **large number of datapoints** in the datasets
 - The availability of **multiple sources of data allows for cross-checking results**, and helps circumvent potential biases that might be present in the datasets

Thank you for your attention!

Get in touch at: kacsuk@hdm-stuttgart.de

Visit our project website:
<https://jvmg.iuk.hdm-stuttgart.de/>