

# Towards an Open Dataset of Visual Language Use in Digital Humanities

---

Rongqian Ma

Dept. Information and Library Science  
Luddy School of Informatics, Computing, and Engineering

IDAH Faculty Fellowship project

April 5, 2024



**LUDDY**

SCHOOL OF INFORMATICS,  
COMPUTING, AND ENGINEERING



# Project Overview

---

**Project goal** is to create an **open-access, large-scale dataset** valuable for quantitative analysis of the use of visualization (or, visual inscriptions and languages) in the digital humanities (DH) field

...which allows analysis from various perspectives, including

- Historical, longitudinal analysis of visual inscription use,
- Cross-sectional analysis of conventions, styles, and use of visual inscriptions across journals and publishers,
- Thematic analysis of topics, ideas, and arguments related to visualization topics or concepts such as visual technologies, methods, and approaches.

# Data collection pipeline

1

DH journal list containing 18 journals (Spinaci et al., 2022)

n=10 open access journals, all accessible through IU library

full-text articles available in XML, HTML, PDF

2

Structure for Article and Figure metadata for each journal

- **Article metadata:** Article ID, Title, Year of publication, authorship info, volume, issue, article type (if available), DOI
- **Figure metadata:** Article ID, Fig. ID, caption, in-text description of the figure

3

Python scripts extracting figures and metadata information; review codes and clean collected data

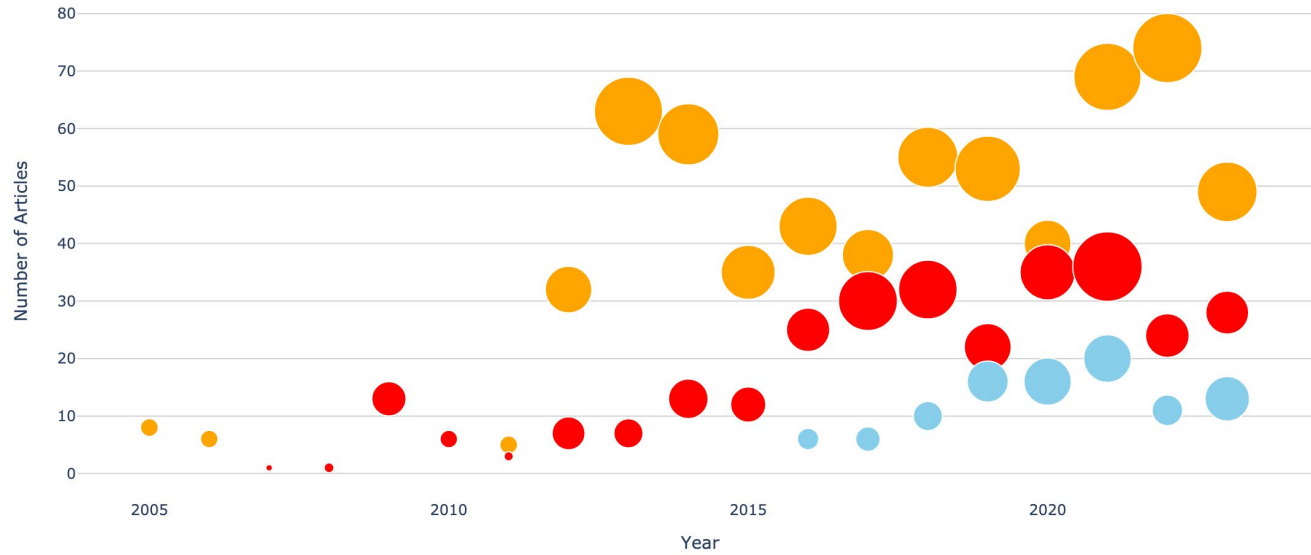


Project Progress

# Overview of the collected data

- Journal of Cultural Analytics
- Digital Scholarship in the Humanities
- Digital Humanities Quarterly

*size of the bubble= number of figures*



# Overview of figure descriptions (using the reference text data)

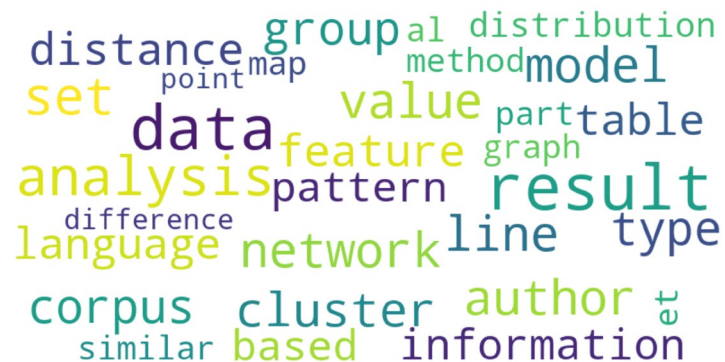
**DHQ**

---



**DSH**

---



**JCA**

---



# ACKNOWLEDGEMENT



Ranvir Singh  
B.S. Computer Science  
Luddy School @ IUB



Amar Ananth  
M.S. Data Science  
Luddy School @ IUB



Sagar Prabhu  
M.S. Computer Science  
Luddy School @ IUB



Dr. Kai Li  
Assistant Professor,  
University of Tennessee



Pei-Ying Chen  
PhD candidate, Information Science  
Luddy School @ IUB

# Why visualizations in DH?

An increasingly “visual” field

- **humanistic visualization** (Druker, 2011)
- **direct visualization** concept (Manovich, 2011)
- **sandcastling** metaphor for visualization (Hinrichs et al., 2019)



# Case studies using the collected dataset

Case study 1: Computational analysis of visual topics and narrative functions of visual inscriptions using DSH full-text publications (2006-2023)

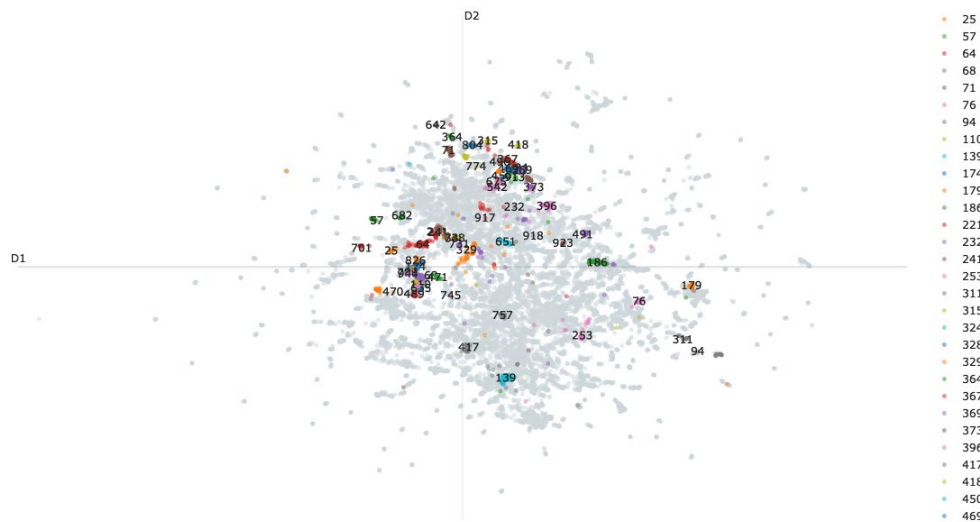


Figure (a). Overview of visual topics in *DSH*.

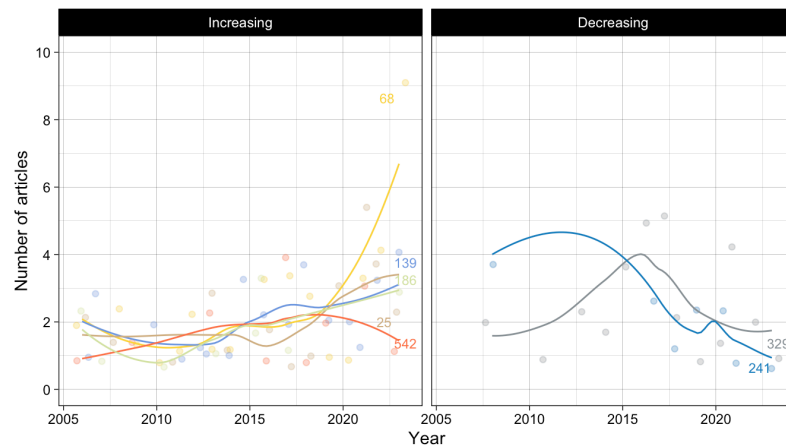


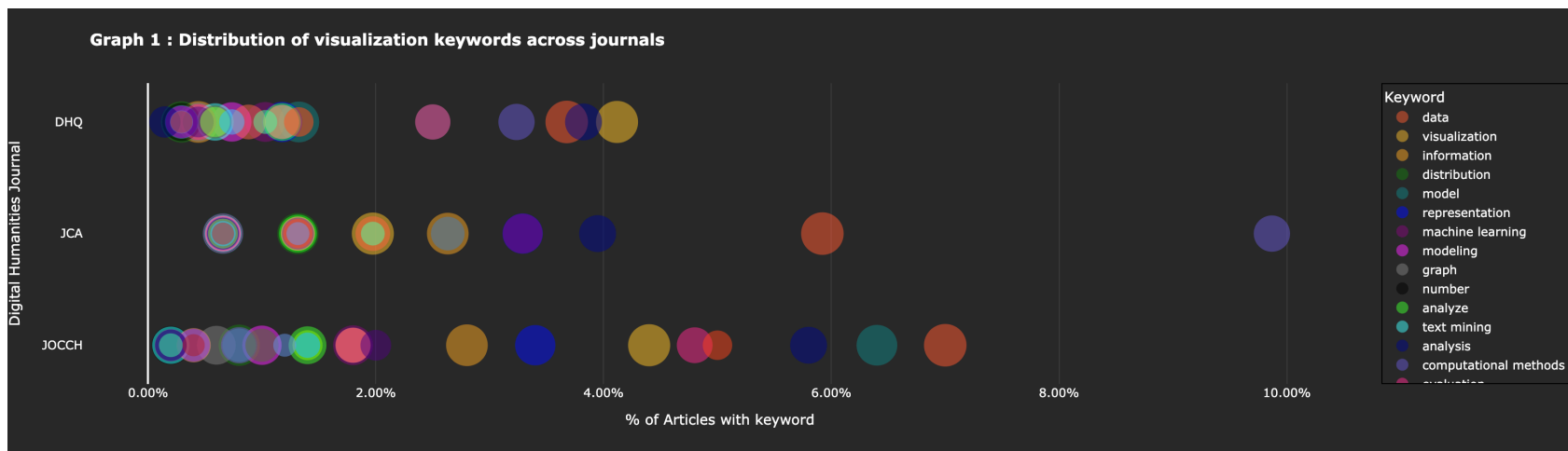
Figure (b). Distribution of 90<sup>th</sup> percentile visual topics in *DSH* from 2006 to 2023.



# Case studies using the collected dataset

Case study 2: Interactive visual dashboard exploring the evolution of key visual concepts across DHQ, JCA, and JOCCH

- supports cross-platform search based on year, journal, and visual concept



screenshot of the beta version of the interactive visual dashboard

# CHALLENGES

---

- To gather full-text data access for all journals;
- Data quality and “messiness” of full-text DH article data;
  - e.g., different “styles” of writing led to difficulty in locating and extracting figures and in-text descriptions of the figures with Python scripts
- Constant modification of Python scripts to extract data from files of various formats and structures

# FUTURE PLAN

---

1. Complete data collection and extraction among other DH journals in the list;
2. Expand the journal sample with other resources (e.g., Scopus keyword search);
3. Curate and release the interactive visual dashboard;
4. Data paper to share the open source dataset;
5. Use the open dataset to conduct empirical analysis of the visual languages use and functions in DH scholarship

# Questions?

---

We would appreciate any suggestions, and opportunities to collaborate! :)

Em a il: [rm56@iu.edu](mailto:rm56@iu.edu)