

Cynthia A. Huang

Munich, Germany

cynthiahqy@gmail.com | cynthia.huang@lmu.de | cynthiahqy.com | github.com/cynthiahqy

Research Interests

I am an interdisciplinary researcher working across statistics, computer science and the social sciences to advance and improve data-driven research. I support the application of novel methods and data sources in the social sciences, and collaborate with leading researchers in computer science and statistics on conceptual frameworks and workflow tools which respect statistical, computational and domain-specific validity, efficiency and usability considerations. Currently, I work on transparent and trustworthy applications of LLMs for social data science, statistical properties of multi-source datasets, and domain specific languages for information visualisation and statistical graphics.

Education

- | | |
|-------------|--|
| 2021 - 2025 | Monash University , Melbourne, Australia
<i>Ph.D. in Mathematics and Statistics</i>
Thesis: Unified Principles and Tools for Complex Datasets and Data-Driven Workflows
Supervisors: Rob J. Hyndman, Simon D. Angus, Sarah Goodwin
Affiliations: Dep. Econometrics and Business Statistics (EBS), Non-Uniform Business Analytics Team (NUMBAT), SoDa Laboratories (SoDa Labs), Monash Data Futures Institute (MDFI), Embodied Visualisation Group (EmVis) |
| 2016 | University of Melbourne , Melbourne, Australia
<i>B.Com Honours in Economics</i> |
| 2012 - 2014 | University of Melbourne , Melbourne, Australia
<i>B.Com in Economics and Management</i> |

Appointments and Research Experience

- | | |
|-------------------|---|
| 01/2026 - Present | Ludwig-Maximillian University, Munich
<i>Postdoctoral Researcher, Chair of Statistics and Social Data Science, Institute of Statistics</i>
Working with Frauke Kreuter in the Social Data and AI Lab (SODA) on AI-Human collaboration in statistical reasoning and data science. Affiliate at Munich Center for Machine Learning (MCML). |
| 03/2025 - 12/2025 | Monash University
<i>Research Fellow, Department of Econometrics and Business Statistics</i>
Worked with Rob J. Hyndman on visualisation principles and software for time-series data. |
| 07/2024 - 11/2024 | University of British Columbia, Vancouver
<i>Visiting Researcher, Department of Computer Science</i>
Hosted by Tamara Munzner in the InfoVis group. Collaborated on reconciling visualisation and provenance approaches from HCL, InfoVis and Statistics. |
| 06/2024 | Ludwig-Maximillian University, Munich
<i>Visiting Researcher, Institute of Statistics</i>
Hosted by Anna-Carolina Haensch in the Social Data and AI Lab (SODA). Collaborated on generating crosswalks for ex-post harmonisation with LLMs. |
| 2023 - 2025 | Monash Addiction Research Center
<i>Research Assistant</i>
Adaption of web-scraped price data from commercial partner for public health policy research. Exploratory analysis and visualisations for research outputs and grant applications. Worked with Tina Lam. |
| 2019 - 2023 | SoDa Labs, Monash Business School
<i>Research Assistant</i>
Supporting multiple empirical economics projects with dataset design, collection, and pre-processing, exploratory visualisation, and preliminary analysis tasks. Worked with Nathaniel Lane, Paul Raschky, Laura Puzzello, Simon D. Angus & Klaus Ackermann |
| 2016 - 2018 | University of Melbourne
<i>Research Assistant, Department of Economics</i>
Data extraction and curation from archival sources such as microfiche and magazine scans, including selecting harmonised variables and coordinating data entry by other team members. Worked with David Byrne & Phillip McCalman. |

Publications

1. Lam, T., **Huang, C. A.**, Torney, A., Callinan, S., Vandenberg, B., Xia, T., Angus, C., Room, R., Ogeil, R., Cowper, A., Pettigrew, S., Rowland, B., Keric, D., Lubman, D. I., & Nielsen, S. (2026). Leveraging web-scraped data to examine alcohol pricing: An Australian feasibility study with retail data. *International Journal of Drug Policy*, 148, 105115. <https://doi.org/10.1016/j.drugpo.2025.105115> (*Journal Article*)
2. Nassar, J. B., Yio, Y. X., Athukorala, N., Simran, Fan, S., **Huang, C. A.**, Bartram, L., Dwyer, T., & Goodwin, S. (2025). Out of the Loop: Enhancing Documentation and Transparency in Collaborative Causal Loop Diagrams to Capture Multiple Perspectives. *2025 IEEE Visualization and Visual Analytics (VIS)*, 111–115. <https://doi.org/10.1109/VIS60296.2025.00028> (*Conference Paper, Best Paper Honorable Mention*)
3. Xiong, K., **Huang, C. A.**, Wybrow, M., & Wu, Y. (2025). TableCanoniser: Interactive Grammar-Powered Transformation of Messy, Non-Relational Tables to Canonical Tables. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–20. <https://doi.org/10.1145/3706598.3714321> (*Conference Paper, Best Paper Honorable Mention*)
4. Ma, B., **Huang, C. A.**, & Haensch, A.-C. (2025). Can Large Language Models Advance Crosswalks? The Case of Danish Occupation Codes. *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop)*, 392–399. <https://doi.org/10.18653/v1/2025.naacl-srw.38> (*Workshop Paper*)
5. **Huang, C. A.** (2023, October 23). Visualising category recoding and numeric redistributions. 1st IEEE Workshop on Visualization and Provenance Across Domains. <https://doi.org/10.48550/arXiv.2308.06535> (*Workshop Paper*)

Preprints and Manuscripts Under Review

1. Solen, M., **Huang, C. A.**, & Munzner, T. (2025). A descriptive framework for visualization context through the lens of influence. (*Conference Paper, under review at Eurographics Conference on Visualization (EuroVis) 2026*).
2. **Huang, C. A.**, O'Hara-Wild, M., Hyndman, R. J., & Kay, M. (2025). ggtime: A Grammar of Temporal Graphics (arXiv:2510.25656). arXiv. <https://doi.org/10.48550/arXiv.2510.25656> (*Preprint*)
3. **Huang, C. A.**, & Lam, T. (2025). PRITES: An integrative framework for investigating and assessing web-scraped HTTP-response datasets for research applications. <https://doi.org/10.48550/arXiv.2511.13773> (*Preprint, invited for submission to Harvard Data Science Review*)
4. **Huang, C. A.** (2024). Crossmaps: A Unified Statistical And Computational Framework For Ex-Post Harmonisation Of Aggregate Statistics. <https://doi.org/10.48550/arXiv.2406.14163> (*Preprint, under review at ACM/IMS Journal of Data Science*)

Presentations

Invited Talks and Panels

- ‘The Past, Present, and Future of ggplot2 Extension Packages’, 2026 Joint Statistical Meetings (JSM), Boston, MA, USA, Aug 2026 (*Invited Panel, forthcoming*)
- ‘The Challenges and Opportunities of AI for Unstructured Data’, Data Quality Workshop: Unstructured Data and the Need for Quality Metrics, Munich, Germany, Jan 2026 (*Invited Panel*)
- ‘Improving ex-post data harmonisation with xmap’, Making Data Work: Tools for Better Statistical Practice, LMU-NYU Workshop, Munich, Germany, Jun 2025 (*Invited Talk*)
- ‘Webscraping in FinEco Research: Risks and Opportunities’, Department of Banking and Finance, Annual Retreat, Inverlock, VIC, Australia, Apr 2024 (*Invited Talk*)
- ‘Web-scraping for Health Policy Research’, Monash Econometrics and Business Statistics Industry Engagement Workshop, Melbourne, Australia, Aug 2023 (*Invited Talk*)

Guest Seminars and Lectures

- ‘An integrative framework for investigating and assessing web-scraped datasets for research applications’, Institute of Service Science, National Tsinghua University, Hsinchu, Taiwan, Apr 2025
- ‘Visualising Category Recoding Graphs in Data Harmonisation’, Department of Computer Science, University of British Columbia, Vancouver, Canada, Oct 2024
- ‘LLMs for Preparing Data in R’, Guest Lecture for Wild-Caught Data (ETC5512), Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia, May 2025

Contributed Talks and Posters

- ‘Scalable self-paced e-learning of statistical programming with fine-grained feedback and assessment’, 2024 Joint Statistical Meetings (JSM), Portland, OR, USA, Aug 2024
- ‘Xmap: Unified Tools for Ex-Post Data Harmonisation’, user! Conference, Salzburg, Austria, Jul 2024
- ‘Leveraging web-scraped retail prices with an industry data partner’, Workshop on Advanced Analytical Methods Using Big Data, Monash Business School and Monash Data Futures Institute, Melbourne, Australia, Apr 2024
- ‘Visualising Category Recoding and Redistributions’, 12th conference of the Asian Regional Section of the International Association for Statistical Computing (IASC-ARS), Sydney, Australia, Dec 2023
- ‘Crossmaps: A principled approach to ex-post data harmonisation and dataset integration’, 2023 Australian Statistical Conference (ASC), Wollongong, Australia, Dec 2023 (*Best Rapid Poster Award*)
- ‘Advances in Ex-Post Harmonisation using Graph Representations of Cross-Taxonomy Transformations’, Monash EBS PhD Contest, International Day of Women in Statistics and Data Science, Online, Oct 2023 (*Best Talk Award*)

Awards, Honours and Funding

2026	LMU Mentoring Program , LMU Munich €1,500 EUR research funds
2026	LMU Postdoc Support Fund , LMU Munich €3,900 EUR relocation and start-up grant
2025	Best Short Paper Honorable Mention , IEEE VIS 2025 Awarded for "Out of the Loop: Enhancing Documentation and Transparency in Collaborative Causal Loop Diagrams..."
2025	Best Full Paper Honorable Mention , ACM CHI 2025 Awarded for "TableCanoniser: Interactive Grammar-Powered Transformation..."
2024	Speaker Support , Posit PBC \$1,800 USD accommodation and travel grant to attend posit::conf(2024)
2023	Enhanced International Research Experience Award , Monash Business School \$8,000 AUD for research visit to UBC, Vancouver
2023	PhD Top-Up Scholarship , Statistical Society of Australia \$2,500 AUD sponsored by Australian Bureau of Statistics
2021 - 2025	PhD Top-Up Scholarship , Monash Data Futures Institute \$8,000 AUD stipend top-up and \$3,000 AUD travel grant per year
2021 - 2025	Monash Graduate Scholarship , Monash University Full doctoral scholarship covering tuition and living stipend

Teaching and Supervision

2020 - 2023	Teaching Associate , Monash Business School, Monash University Led weekly tutorials for undergraduate and masters courses. Received teaching excellence awards for ECC2800 and ETC5512 based on student evaluations. * ETC5523: Communicating with data (2023) * ETC1010: Introduction to data analysis (2022) * ECC2800: Prosperity, poverty and sustainability in a globalised world (2022) * BEX2421: Harnessing big data for business and society (2022) * ETX2250/ETF5922: Data visualisation & analytics (2022) * ETC5512: Wild-caught data (2020)
2014 - 2019	Academic Tutor , Faculty of Business and Economics, University of Melbourne Led weekly tutorials for undergraduate courses in Economics and Management. Developed assignments and marking rubrics. Preparation and delivery of subject review lectures. * ECON10005: Quantitative Methods (2019) * ECON10004: Introductory Microeconomics (2017) * ECON10003: Introductory Macroeconomics (2014, 2016, 2017, 2018, 2019) * ECON20002: Intermediate Microeconomics (2019) * ECON20001: Intermediate Macroeconomics (2018) * ECON20005: Competition and Strategy (2017, 2018) * ECON30025: Computational Economics and Business (2019) * ECON40002: Advanced Macroeconomics (2017) * MGMT20012: Business Case Analysis (2019)

Supervision and Mentoring

2025 | **Research Internship Supervisor**, Monash University
Supervised 2 Masters of Business Analytics students working on LLMs for data cleaning

2018 - 2019 | **Peer Mentor**, Williams Centre for Learning Advancement, University of Melbourne
Peer support for new tutors in the Faculty of Business and Economics

Service, Outreach and Public Engagement

Open-Source Software Contributions

- Contributor, R Dev Day, Melbourne, Australia, Nov 2025
- Contributor, Tidy Dev Day, Seattle, WA, USA, Aug 2024
- R package maintainer: xmap, ggtilecal

Service

- Organising Committee Member, NUMBAT Hackathon, Melbourne, Australia, Nov 2025
- Co-Chair, WOMBAT2025 Conference, Melbourne, Australia, Sep 2025
- Organiser, Monash NUMBATs Research Software Study Group, Monash University, 2023-2024

Research Skills Workshops

Delivered to graduate students and faculty at Monash Business School:

- Reproducibility in Economics Research (May 2024, Sep 2025)
- Meeting AEA replication standards using Quarto (Aug 2023)
- Designing R Packages (Oct 2022)
- Quarto websites as research compendiums (Aug 2022)
- Writing academic papers with Rmarkdown and friends (Aug 2022)

Public Talks and Meetups

- 'Reusing ggplot2 code: how to design better plot helper functions', Statistical Society of Australia (SSA) Vic & Tas Branch and R Users of Melbourne Network (MelbURN) Meetup, Melbourne, Australia, Aug 2025
- 'Quarto for Knowledge Management', posit::conf(2024), Seattle, WA, USA, Aug 2024
- 'Silver Linings: How a Data Harmonisation Mishap Turned into a PhD Thesis', Data Mishaps Night, Online, Mar 2024
- 'Australia's Rising WiDS Stars!', WiDS @ Australian Data Science Network, Online, Mar 2023
- 'Misadventures with Reproducibility in R', R Ladies Melbourne Meetup, Melbourne, Australia, Nov 2022

Science Communication

- Presenter, *3 Minute Thesis Competition*, Monash University, Jun 2023 (*2nd Place in Faculty Heat*)
- Podcast Interview Host, *The Random Sample*, 2022-2025

Media Appearances

- Podcast Guest, 'Why Host a Podcast?', *The Random Sample*, Dec 2023. Available at: <https://www.therandomsample.com.au/podcast/why-host-a-podcast/>
- Featured in 'Crimes against spaghetti and other three minute tales win over judges', *Monash Business School*, Jun 2023. Available at: <https://www.monash.edu/business/news/2023/crimes-against-spaghetti-and-other-three-minute-theses-win-over-judges>
- Interview, 'Shaping Research Software: An Interview with Cynthia Huang', *Australian Research Data Commons*, May 2023. Available at: <https://ardc.edu.au/article/shaping-research-software-an-interview-with-cynthia-huang/>