

Lydia R. Lucchesi

lydia.lucchesi@anu.edu.au

+1 (605) 695-2689

<https://lydialucchesi.github.io>

My research interests are in data provenance, data preprocessing, data visualisation, uncertainty communication, and reproducible research. These interests stem from my formal training in statistics and real-world experiences analysing data. Recently, my work has focused on developing software tools to visualise and communicate data preprocessing decisions, to support reproducibility and increase transparency in data analytics.

EDUCATION

Australian National University, School of Computing, October 2019 – present (Thesis submission date: 8 February 2024)

Ph.D. candidate in computer science

- Thesis: “Visualisation and Software to Communicate Data Preprocessing Decisions”
- Advisors: Dr. Lexing Xie, Dr. Petra Kuhnert, and Dr. Jenny L. Davis
- Fully funded through an ANU Fee Remission Merit Scholarship, ANU PhD Scholarship, and Top-up Scholarship from the Commonwealth Scientific and Industrial Research Organisation (Australia’s national science agency).

University of Washington, Department of Global Health, September 2017 – December 2018

Graduate non-matriculated student

- Completed graduate-level courses in biostatistics, global health, and science & technology studies while a post-bachelor fellow at the Institute for Health Metrics and Evaluation.

University of Missouri, Department of Statistics, August 2014 – May 2017

B.A. in statistics, summa cum laude and departmental honours

- Thesis: “Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation”
- Advisor: Dr. Christopher K. Wikle

Minor in Spanish

Certificate in multicultural studies

RESEARCH & WORK EXPERIENCE

Data61 postgraduate, Commonwealth Scientific and Industrial Research Organisation (CSIRO)

Canberra, Australia, October 2019 – December 2023

Conducted research at Australia’s national science agency in the Statistical Machine Learning group within CSIRO’s Data61. Received PhD supervision from a CSIRO senior research scientist and access to facilities and resources, as a recipient of the Data61 Top-up Scholarship. Maintained the Vizumap software, developed at CSIRO in 2017.

Team member, Humanising Machine Intelligence research initiative

Australian National University, October 2019 – December 2023

Actively engaged with and contributed to the interdisciplinary community consisting of computer scientists, social scientists, philosophers, and lawyers. Integrated learned research techniques from other disciplines into research projects to view and solve problems from new angles.

Post-bachelor fellow, Institute for Health Metrics and Evaluation (IHME)

University of Washington, September 2017 – August 2019

Modelled the health burden of nonfatal injuries from over 30 different causes of injury (e.g., cyclist road injuries, falls, venomous animal contact, etc.) for the Global Burden of Disease Study (2017 & 2019). Modelled progress on several United Nations Sustainable Development Goal indicators. Collated and preprocessed large injury datasets, e.g., health insurance claims. Programmed and managed analytical pipelines in R, Python, and Stata.

This fellowship included a practicum. I was a **visiting research fellow at the Ethiopian Public Health Institute in Addis Ababa, Ethiopia, from July – August 2019**. I developed training materials on metadata and dataset documentation for the new National Data Management Center fellowship program.

Volunteer fellow, CSIRO’s Data61

Canberra, Australia, July – August 2017

Developed the Vizumap R package, which offers four mapping methods to visualise uncertainty in spatial statistics. Focused on environmental applications, e.g., sediment in catchment areas flowing into the Great Barrier Reef. Travel and accommodation fully funded by CSIRO.

Undergraduate research assistant, U.S. National Science Foundation (NSF)-Census Research Network Grant

University of Missouri, August 2016 – June 2017

Developed methods to visualise uncertainty in areal data presented on choropleth maps. Produced example maps, in the R programming language, for American Community Survey statistics.

Researcher and trainee, Big Data Summer Institute in Biostatistics

University of Michigan School of Public Health, June – July 2016

Collaborated with two students to build a statistical model to predict the prevalence of chronic kidney disease and identify potential risk factors. Team presented at a university symposium. Attended career talks and lectures on statistics and computing.

Grader, Department of Statistics

University of Missouri, January – May 2016

Corrected weekly assignments for four sophomore and junior level statistics courses.

Student assistant, Dean's Office at the Harry S Truman School of Public Affairs

University of Missouri, January 2015 – May 2016

Performed clerical tasks. Assisted in assembling the department's 25-page annual report.

PUBLICATIONS

Lucchesi, Lydia R., Petra M. Kuhnert, Jenny L. Davis, and Lexing Xie (2022). Smallset Timelines: A Visual Representation of Data Preprocessing Decisions. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 23 pages.

<https://doi.org/10.1145/3531146.3533175>

Lucchesi, Lydia R., Petra M. Kuhnert, and Christopher K. Wikle. Vizumap: an R package for visualising uncertainty in spatial data. *Journal of Open Source Software* 6.59 (2021): 2409. <https://doi.org/10.21105/joss.02409>

Lucchesi, Lydia R., and Christopher K. Wikle. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat* 6.1 (2017): 292-302. <https://doi.org/10.1002/sta4.150>

SELECTED IHME PUBLICATIONS

I am a co-author on various large collaborative publications led by the Institute for Health Metrics and Evaluation (see post-bachelor fellow in research & work experience). Three examples below.

(2nd author, 44 collaborators)

James, Spencer L., et al. The global burden of falls: global, regional and national estimates of morbidity and mortality from the Global Burden of Disease Study 2017. *Injury prevention* 26.Suppl 2 (2020): i3-i11.

<https://doi.org/10.1136/injuryprev-2019-043286>

(2nd author, 91 collaborators)

James, Spencer L., et al. Morbidity and mortality from road injuries: results from the Global Burden of Disease Study 2017." *Injury prevention* 26.Suppl 2 (2020): i46-i56. <https://doi.org/10.1136/injuryprev-2019-043302>

(>100 collaborators)

James, Spencer L., et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 392.10159 (2018): 1789-1858. [https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)

SOFTWARE

Lucchesi L (2023). smallsets: Visual Documentation for Data Preprocessing. R package version 2.0.0, <https://CRAN.R-project.org/package=smallsets>. Project website: <https://lydialucchesi.github.io/smallsets>

Nelson, S., **Lucchesi, L.** and Kuhnert. P.M. (2022). VizumApp: A Shiny App for visualizing uncertainty in spatial data using the Vizumap R package.

Lucchesi L, Kuhnert P (2021). Vizumap: Visualizing uncertainty in spatial data. R package version 1.2.0. <https://github.com/lydialucchesi/Vizumap>

TALKS

Smallset Timelines: A Visual Representation of Data Preprocessing Decisions

- ACM FAccT 2022 conference in Seoul, Republic of Korea (in-person, June 2022)
- Gradient Institute (Zoom, June 2022)
- University of Chicago HCI Club seminar (in-person, July 2022)
- Midwest Uncertainty Collective at Northwestern University (in-person, July 2022)

Smallset Timelines to Track and Communicate Data Preprocessing, August 2021

Roundtable discussion on “Ethics and Methods for Digital and Quantitative Sociology” at the American Sociological Association (ASA) virtual annual meeting.

Mighty Maps: Showcasing Estimate Uncertainty, June 2020

Joint talk with Dr. Susanna Cramb for the Statistical Society of Australia Queensland branch meeting.

Using the Vizumap package to visualise uncertainty in spatial estimates, November 2019

Workshop organized by the Monash University Business Analytics team in Melbourne, Australia.

The Global Burden of Pediatric Injuries, March 2019

Poster at the International Pediatric Association Congress in Panama City, Panama.

Visualizing uncertainty in American Community Survey estimates with bivariate choropleth maps and map pixelation, April 2017

American Statistical Association Mid-Missouri Chapter student research symposium in Columbia, Missouri.

TEACHING & ACADEMIC SERVICE

Teaching assistant, Statistical Machine Learning (COMP 4670 / 8600)

Australian National University, February – June 2023

Co-led a weekly tutorial on topics such as regression, neural networks, Gaussian processes, and graphical models.

Convener, Humanising Machine Intelligence postgraduate seminar, August – November 2022

Organised bi-weekly reading and discussion activities and two invited talks by postdoctoral fellows.

Supervisor, CSIRO vacation student program, November 2022 – February 2023

Co-supervised an undergraduate student working on software development for the Vizumap R package and VizumApp Shiny app.

Paper session chair, ACM FAccT 2022 conference in Seoul, Republic of Korea, June 2022

Invited to chair the paper session “Use and Abuse of Visual Analysis.”

AWARDS

Travel award (\$500 AUD), Sydney Law School, July 2023

Financial support to attend the PhD Winter Workshop in Critical Technology Studies at the University of Sydney.

SSA Venables Award (\$4,000 AUD), ARDC, June 2022

Won with Sam Nelson and Petra Kuhnert for the Vizumap package and VizumApp app. Award hosted by the Statistical Society of Australia and sponsored by the Australian Research Data Commons to recognise new developers of open-source software for data analytics, in honour of Bill Venables.

Interview: <https://ardc.edu.au/article/shaping-research-software-an-interview-with-lydia-lucchese-and-sam-nelson/>

Early career researcher publication award (\$300 AUD), CSIRO’s Data61, June 2022

Monetary award from the Analytics and Decision Sciences program for one of the best publications in 2021 in the early career researcher category.

Travel grant (\$2,000 USD), ACM FAccT 2022, June 2022

Financial support to attend the conference in Seoul, Republic of Korea.

Award for Academic Distinction, University of Missouri, April 2017

One of 15 undergraduates selected from 70 nominees for the university-wide award for “showing evidence of extraordinary intellectual curiosity, actively seeking knowledge beyond the classroom and striving to share that knowledge with public audiences for a broader impact, and significantly contributing to the academic atmosphere at the University of Missouri.”

Honourable mention (physical sciences & mathematics division), University of Missouri Undergraduate Research & Creative Achievements Forum, April 2017

Poster presentation on honours thesis research about uncertainty visualisation for spatial statistics.