Laura W. Dozal

• Email: lwerthmann@arizona.edu • Website: lwdozal.github.io • LinkedIn: ./laurawerthmann/

EDUCATION

PhD Candidate, iSchool (School of Information); Minor in Statistics and Data Science;
Certificate in Computational Social Science
> **University of Arizona,** *Tucson Arizona, Exp. Spring 2025*

Master of Arts in Communication, Culture, and Technology
> **Georgetown University,** *Washington, D.C. 2016*

Bachelor of Arts in English Literature; Minor in Communication
> **St. Edward's University**, *Austin Texas 2012;* **Recognition**: *Dean's List, 2010*

EXPERTISE

Programming Languages / Tools:
- R - tidyverse, ggplot2, quarto, Shiny, Survival, bnlearn
- Python - NumPy, pandas, matplotlib, OpenCV , scikit-learn, py-torch, spaCy, TensorFlow, keras, langchain

- VSCode, Jupyter Notebooks, Spyder, Anaconda
- SQL, Javascript, D3.js, HTML , Leaflet, ArcGIS
- LaTeX, Github

Bilingual: Fluent in English (native) and Spanish

EXPERIENCE & RESEARCH

Instructor, ISTA 116: Statistical Foundations for the Information Age | College of Information, University of Arizona, 7/2024 – Present
- Instructor on record for undergraduate course that covers understanding uncertainty and variation in modern data: data summarization and description, rules of counting and basic probability, data visualization, graphical data summaries, working with large data sets, prediction of stochastic outputs from quantitative inputs. Operations with statistical computer packages such as R.

Data Science Fellow |Laboratory of Applied Science - Summer Conference of Applied Data Science; North Carolina State University, Raleigh, North Carolina, 06/2024 – 07/2024
- NLP researcher comparing chunk-size methods for text summarization. Used extractive summarization occams and abstractive summarization, Retrieval Augmentation Generation (RAG) to test different text segmentation methods. Used and created vector embedding database for evaluation using RAGAs, an LLM RAG evaluator. Used the 2024 TREC Challenge data-MSMarco docs and QA for analysis.
- Implemented two video object tracking models: zero-shot transformers - OWL-VIT and LMM gpt-4o on noisy data- dashcam footage of a ride in Afghanistan provided by LAS staff. The goal was to explore improvements for tracking object obfuscation and detecting unknown/unconventional objects in the noisy surveillance video data. We found that pre-trained image models provided the most accurate structured data outputs.

Complexity Science Researcher | Winter Workshop on Complex Systems (Santa Fe Institute), Spain, 01/2024
- Ongoing research of spreading infections (SI) phenomenon through honeybee food transmission and communication (trophallaxis) using complex systems methods including autocorrelation networks, temporal networks, burstiness, and social role clustering.

R4R (Roots for Resilience) Dissertation Fellow, 2023 Cohort | *University of Arizona Data Science* Institute, Tucson, AZ, 07/2023 – 01/2024
- *Data Management*: Created and maintained a Data Management Plan using Open Science Framework in collaboration with CyVerse. Use of DataOps and pipelines, Version Control - github, and Reproducibility- Docker.
- *Dissertation Analysis:* Map Building and Image Analysis for social movement group in Mexico Ellastienennombre.org. Maintain multimodal dataset in MySQL, create interactive Leaflet map. Also used Github and docker to apply computer vision methods including localized feature matching, segmentation, homography, recognition and Bayesian modeling.

LEADING Data Science Fellow 2023 Cohort | *University of Rochester, Remote 06/2023 – 01/2024*
- Led creation of an Open-Source Data Pipeline to collect and analyze bibliometric data from University of Rochester Researchers and identify publication behavior. Collected 500K+ metadata using APIs from various repositories, cleaned and processed the data to set-up for analysis and better reproducibility. R and Python.

- Used trend analysis and network analysis methods including regression (decision tree), time series analysis, structural analysis, and prediction analysis for network validation of researcher behavior.

Data Science Grad. Research Assistant | *University Analytics & Institutional Research, UA, Tucson, 08/2021 – 06/2023*
- *NLP Sentiment Analysis:* Analyze 4M+ student course survey responses using Natural language processing to identify sentiment, word frequencies and word synonyms for curriculum in various departments on campus. Increase faculty participation in curriculum review across schools on campus.
- *Statistical Modeling/Machine Learning:* Spearheaded the Women in Science and Engineering Retention report project which uses measurement methodologies, survival and churn analysis, to understand graduation and retention. These analyses review dropout rates of 37K+ students over ten years via large-scale backend and data pipeline connected to an Oracle SQL Database. The report incorporates decision tree and random forest models along with other time-based statistical methods to provide decision making report for campus admin.

Computational Tools and Thinking Teaching Fellow | *University of Texas Austin, 06/2022 - 7/2022*
- Developed ten, two-hour lesson plans for computational tools and thinking survey course that covered R, Python, Social Network Analysis, Design Theory, and coding best practices. Taught to a wide range of experience levels that included 25 high achieving undergraduate students from STEM and Social Sciences.
- Held office hours to support outside research and enabled the development of a 20-month research design process. Participated in community building activities with other distinguished fellows as well as with our students. Maintained mentor level contact with students to aid in coding and analysis applications for research.

Data Science Ambassador | *University Arizona Data Science Institute, Tucson 08/2021 – 08/2022*
- <u>Data Carpentries Instructor as of 2/14/2022</u> -Live coding instructor and helper for Data Analysis and Visualization with R for Social Scientists, Introduction to Python, Open Refine, and other data science skills.
- Data Science community support and resource for the CSS certificate and iSchool. Created and Organized Projects in Progress (PIP). Fostered a Data Science community environment for researchers on campus to collaborate and share projects ideas and progress. Helped consult 3 - 5 projects over the span of two semester (over 20 researchers) using data science methods monthly.

NLP Annotator and Research Reviewer | *Extracting Space Situational Awareness Events from News Text Publication, Publication in LREC -- Tucson, AZ 2022*

- We constructed a corpus of 48.5k news articles spanning all known active satellites between 2009 and 2020. Using a dependency-rule-based extraction system designed to target three high-impact events – spacecraft launches, failures, and decommissionings, we identified 1,787 space-event sentences for human annotation with 15.9k labels for event slots. We empirically demonstrated a state-of-the-art neural extraction system achieves an overall F1 between 53 and 91 per slot for event extraction in this low-resource, high-impact domain.

LEADERSHIP

AI Steering Committee, University of Arizona, Grad-Student Representative, 2023 – Present

PhD Community Manager (University of Arizona iSchool) 2021 - Present
Data Carpentries Instructor – Lead coding workshops in R, Python, Open Refine, and data science skills, 2022 –Present
NLP Reading Group & Machine Learning Reading Group, UofA, 2019-Present
Lecturer- ISTA 130: - Computational Thinking and Doing (Intro to Python),2023 - Present
ASIS&T Publication Reader, Paper Submission Reviewer, 2024
ASIS&T SIG – Visualization, Sound, and Image: Secretary Chair 2021-2022
WiDS-Tucson (Women in Data Science), Panel Sessions Organizer, Data Blitz Committees 2021 -2022
RezBaz - Presenter & Moderator, Network Analysis 2021; R Visualization 2021 and 2022
iSchool Colloquium, Organizer: Moderated Lectures, led school-wide Faculty Research Blitz 2020 – 2021

PUBLICATIONS and CONFERENCES

- ***Visual and Textual Analysis of the Online Anti-Feminicide Movement in Mexico: A Narrative Network Approach***, *Conference of Complexity Science 2024 Poster Acceptance; Exeter, England*
- *(Accepted, not published),* ***Visual and Textual Analysis of the Online Anti-Feminicide Movement in Mexico: A Narrative Network Approach,*** *ASONAM2024, IEEE Conference; and ACM GOODIT2024 Conference*
- ***Visual and Textual Narrative Analysis of the Online Anti-Femicide Movement in Mexico: A Narrative Network Approach,*** *ACL Poster for WOAH workshop, 2024*
- ***Narrative Frames in Topical Images: Mixed Methods Approach to Understanding Social Movements***, *ASIS&T Workshop Panelist: "Turning Social Informatics Research into Action in a Changing Moment", London UK, October 2023*
- ***Mixed Methods Framework for Understanding Visual Frames in Social Movements,*** *ASIS&T Conference, Long Paper Publication, London UK, October 2023*
- ***Posit::Conf2023,*** *Opportunity Scholar, September 2023*
- ***Mixed Methods for Understanding Visual Frames in Social Movements,*** International Conference on Computational Social Science (IS2C2), Poster, Chicago, July 2022
- ***Extracting Space Situational Awareness Events from News Text***, LREC, May 2022
- ***Street Art as Visual Information: Mixed Methods Approach to Analyzing Community Spaces,*** ASIS&T Poster, 11/2021
- ***Detecting the Spread of Propaganda Bias Online: Comparing HAN and BERT Models***, TPRC, Paper Publication, 4/2021
- ***Street Art Dialogue,*** Gnovis Journal*, December 18, 2014 (Peer Reviewed Publication)

RELEVANT COURSEWORK

Advanced Natural Language Processing, CS – Fall 2024
Introduction to Computer Vision, CS - Fall 2023
Theory of Statistics School of Statistics, Spring 2023
Intro to Artificial Intelligence, iSchool - Spring 2022
Probabilistic Graphical Models, CS – Spring 2022
Theory of Probability – School of Statistics, Fall 2021
Neural Networks, iSchool - Fall 2020