Connor Gilroy

cgilroy@uw.edu; ccgilroy.com; github.com/ccgilroy; ccgilroy; 731-225-0001

Summary: Computational social scientist & PhD candidate, with 7+ years in research and data science. I create new analytic insights by linking data across multiple domains – including survey, behavioral, text, spatial, and network data – and building statistical/machine learning models for explanation and prediction.

Skills: programming (R, Python, SQL, and Stan), data collection (webscraping and webcrawling, geocoding, APIs), databases (Hive, PostgreSQL), surveys (stratified survey design and weighting), statistical modeling (GLMs, hierarchical models, regularized regression, random forests, spatial autocorrelation, Bayesian statistics), text analysis (topic modeling, word embeddings), data visualization (ggplot2, matplotlib, altair)

Education

2023 (expected). PhD, Sociology, University of Washington
2018. MA, Sociology, University of Washington
2013. BA, Biology and Geophysical Sciences, University of Chicago

Certificate (2019): Special Concentration in Social Statistics

Employment

University of Washington (2015 – present)
Graduate Research Assistant. Built a Python web crawler and a PostgreSQL database to collect and store text data from nonprofit and government webpages; analyzed Twitter networks in R and SQL to uncover long-distance ties of local US homeless care organizations

Graduate Teaching Assistant & Instructor. Taught technical topics in R, Python, Stata, and Stan to audiences of diverse skill levels, including data and society (undergraduate); social statistics, data science and population processes, Bayesian statistics for the social sciences (graduate); digital demography and word embeddings for social sciences (workshops)

Meta (summer 2020, 2021, 2022)
***Data Science Intern* (2022, Communities Ecosystem team)**

- Quantified feed engagement with reshared Facebook Groups content by creating and comparing interactions-per-view metrics, segmented by content type and genre (with SQL and R). Presented insights to the Director of Data Science for Communities, and met with multiple Communities teams to propose new content-related initiatives based on my findings. Justified and launched a new feed experiment to assess incremental value of reshared content.
- Built a data pipeline in Python and SQL to create a new experimental metric and reanalyzed a collection of experiments in R to quantify tradeoffs between types of content using a quadratic model of diminishing returns.

Quantitative User Experience Research Intern

- (2020, Community Creation team) Designed and conducted two studies (log data analysis using random forest models to rank variable importance, behaviorally-targeted followup survey with questions based on the model results) of factors driving initial participation in new online communities, identifying relational traits that distinguished engaged members from passive ones.
- (2021, Commerce Ecosystem team) Designed, fielded, and analyzed a complex stratified survey of job seeker needs and satisfaction, which quantified key gaps with competitor platforms and identified a use case for career-focused contextual profiles. Wrote a widely-referenced technical guide to creating inverse probability weights for nonresponse using an internal Python tool (now open-sourced as the balance package). Evaluated quality of an external vendor's missing data imputation for a sensitive survey question.

The World Bank Group (2020)
Consultant, Survey Data Analysis. Created statistical models (hierarchical models and GLMs) and data visualizations of gender-based disparities in Central Asia; wrote reproducible R pipeline for updating reports with new waves of survey data (feature story based on my analysis)

Epic Systems Corporation (2013 – 2015)
Quality Assurance. Automated and manual software testing for web APIs and enterprise healthcare software

Selected Publications

Gilroy, Connor and Ridhi Kashyap. 2021. "Digital Traces of Sexualities: Understanding the Salience of Sexual Identity through Disclosure on Social Media." *Socius* 7: 1–18. doi: 10.1177/23780231211029499

- Compiled a dataset through the Python version of the Facebook Marketing API, fit Bayesian loglinear model (a negative-binomial GLM with complex interactions) in R and Stan, and used simulations from the model to produce visualizations of systematic variations in sexual identity disclosure rates by age, gender, and relationship status.

Filippova, Anna, **Connor Gilroy**, Ridhi Kashyap, Antje Kirchner, Allison C. Morgan, Kivan Polimis, Adaner Usmani, and Tong Wang. 2019. "Humans in the Loop: Incorporating Expert and Crowdsourced Knowledge for Predictions using Social Survey Data." *Socius* 5: 1–15. doi: 10.1177/2378023118820157. (corresponding author)

- Fit regularized regression models with glmnet in R to predict life outcomes from survey data, varying the coefficient penalties according to human-derived variable importance ranking from a wikisurvey. Managed the reproducibility of our code base and coordinated the project writeup.

Projects

"Expressions of Community: Understanding Variations in Community through LGBTQ Experiences." (Dissertation)

- Three interrelated projects: 1) link location characteristics to survey responses to show how density and prevalence of same-sex couples contribute to a sense of community for LGBTQ people; 2) compare a locally-trained word embedding model to a generic GloVe model to show how "community" takes on one specific meaning in an LGBTQ Usenet group, as opposed to the dual meanings encoded in the GloVe model (which I disentangle with PCA); 3) connect conversational interaction network features from subreddits to GloVe-derived measures of "community talk".

"How Distinct is Gay Neighborhood Change? Patterns and Variation in Gayborhood Trajectories." (MA thesis; Distinguished Thesis Award in Social Sciences, 2019)

- Built a webscraper to collect gay bar addresses; geocoded addresses within Census tracts to identify gayborhoods and join with Census demographic data; matched via Mahalanobis distance to create synthetic control neighborhoods; and statistically compared trajectories of change in gayborhoods to matched neighborhoods (with linear regression and multilevel models) to show that gayborhoods do not exhibit a common or universal pattern of change.