Statistical Consulting Projects

2022

Stephanie Dickinson
Executive Director, Biostatistics Consulting Center
Development of a data tracking system for COVID-19 test samples.

- Developed an **R Shiny** app to monitor the **cloud database** and track sample collection progress.

Stephanie Dickinson
Executive Director, Biostatistics Consulting Center
Development of a project management system.

- Revised and debugged an **R Shiny** app for accessing cloud data storage.
- Developed a custom **pie-donut plot** function to replace a package incompatible with R Shiny.

2021

Jaroslaw Harezlak, Ph.D.
Professor / Assistant Dean of Research Analytics, School of Public Health
CARE Consortium: Effects of prolonged head impact exposure on concussion in NCAA football players.

- Conducted rigorous **data wrangling and validation** to correct duplications, misentries, and systematic UTC time convert errors in 700,000+ head impact records.
- Developed 4 **R Shiny** apps for descriptive statistics, model selection, and network visualization.
- Conducted **survival analysis** (`survival`) to assess the relationship between prolonged impact exposure and concussion diagnosis.
- Applied **hierarchical clustering** to match concussed players and their comparable counterparts, and conducted **functional data analysis** (`refund`) to evaluate group differences.

Stephanie Dickinson
Executive Director, Biostatistics Consulting Center
*Implementation of **RStudio Connect (Posit Connect)**.*

- Collaborated with the IT team to test system functionality, including package compatibility, cloud data management, and authentication.

2020

Stacey Giroux, Ph.D.
Research Scientist, Ostrom Workshop, Indiana University
Informal vendors and food systems planning in an emerging African city.

- Created network figures to display vendor-supplier connections using **network** (`statnet`) and **mapping** (`leaflet`) packages.

https://doi.org/10.1016/j.foodpol.2020.101997 (Acknowledged)

David Allison, Ph.D. (Research Team)
Dean / Provost Professor, School of Public Health
Protein expression during circadian rhythms in the hippocampus of senescence-accelerated mice.

- Reproduced 100+ GraphPad Prism figures by developing a custom R Plotting function to ensure consistent plot margins and other parameters.
- Reviewed and debugged code from clients and colleagues for **ANOVA Type III** sum of squares.

Carmen D. Tekwe, Ph.D.
Associate Professor, Department of Epidemiology and Biostatistics
ActiGraph accelerometer data processing.

- **Data merging** of SAS demographics data with accelerometer records.
- Provided consulting on data management, activity measure quantification, group comparison, and systematic error handling.

Priscilla A. Barnes, Ph.D.
Associate Professor, Department of Applied Health Science
Communication and partnership behaviors on COVID-19 response among organizations in Indiana.

- Designed partnership **network** structure based on questionnaires for visualization.
- Developed an **R Shiny app** to interactively generate networks based on partnership strength.

Kevin Naaman
Ph.D. Student in Epidemiology and Quantitative Methodology
Transparency and Openness Promotion Guidelines: a theory-based survey of journal editors.

- Consulted on **R Shiny** development and created a Shiny template tailored for data applications.

Shiny application for the TRUST project
https://doi.org/10.1098/rsos.221093 (Acknowledged)

2019

Wasantha P. Jayawardene, MD, PhD
Assistant Research Scientist / Adjunct Professor, Department of Applied Health Science
Activity tracking of humans and their dogs.

- Contributed to study design and data management in the early stages of the project.
- Helped develop a cloud-based data management system for tracking two devices via **R Shiny**.

Kimberly Kelly, Ph.D.
Associate Professor, West Virginia University School of Pharmacy
Survey data analysis for a breast cancer risk communication tool.

- Conducted **data cleaning** and **model selection**.
- Generated reports with statistically significant findings using **logistic regression**.

David Allison, Ph.D.
Dean / Provost Professor, School of Public Health
A randomized, placebo-controlled crossover trial of a decaffeinated energy drink.

- Conducted a rigorous R code review for **generalized linear mixed models**.
- Performed cross-validation of **Bayes factors** between SAS and R.

2018

Cynthia Kroeger, Ph.D.
Postdoctoral Fellow, Department of Epidemiology and Biostatistics
Alternate Day Fasting versus Calorie Restriction for Health Behavior: Randomized Clinical Trial.

- Reviewed R code for a **generalized linear mixed model** to compare behaviors across three groups.
- Conducted cross-validation between SAS and R for **multiple imputation** for missing data, adding post-hoc code to R package (`mice`) to align with SAS output.

Stacey Giroux, Ph.D.
Research Scientist, Ostrom Workshop, Indiana University
Debugging for R package (CCTpack).

- Identified a critical error in the package that mistakenly transposed the data matrix and led to incorrect **clustering** evaluations.

2017

Jon Macy, Ph.D.
Associate Professor, Department of Applied Health Science
Smoking Cessation Trials in Pregnant Women.

- Developed an **R Shiny** app to display variability of data within each subject, suggesting a Gaussian **generalized linear mixed model** (`nlme`, `lme4`) to test group differences with random slopes.

Priscilla A. Barnes, Ph.D.
Associate Professor, Department of Applied Health Science
Visualization of partnerships among hospitals and health organizations.

- Designed an algorithm to construct a **graph adjacency matrix** from survey responses.
- Developed an **R Shiny** app to visualize partnership structure of health organizations as a **network**.

2016

Stephanie Dickinson
Senior Biostatistician & Consulting Manager, Biostatistics Consulting Center
Cancer risk map development in the state of Indiana.

- Developed an **R Shiny** app to visualize colon cancer risk using a mapping (`leaflet`) package.

Marcy Kingsbury, PhD
Senior Scientist, Ecology, Evolution and Behavior, Department of Biology
Development of custom 3D surface plots.

- Used a **convolution** algorithm to show peaks on surface mesh with color gradients.
- Developed an **R Shiny** app to customize mesh resolutions, colors, and peak values.

2015

Adam V. Maltese, Ph.D.
Associate Professor, Science Education, School of Education
Demographic differences in standardized test scores among 8th and 10th-grade students.
- Performed **exploratory data analysis** to identify demographic patterns in the scores.
- Conducted **pseudo-longitudinal analysis** using the **permutation pairing** technique based on demographic variables in cross-sectional data.

Michael S. Willett, Ed.D.
Associate Chair, Department of Kinesiology
Sports financial metrics on medal counts in the Olympics and World Championships.
- Conducted **ANOVA**, finding a significant relationship in track and field.
- Developed an **R Shiny** app to demonstrate data visualization in top 100 performance in swimming.

2014

Paul Bryant, Ph.D.
Associate Professor, Media Psychology, The Media School
Behavioral network in pornography.
- Analyzed the link among sexual behaviors in adult videos by designing an algorithm to construct an adjacency matrix from a **Jaccard similarity matrix**.
- Created **network plots** to visualize behavioral clusters and connections.

Qatrunnada Ismail
Ph.D. Student, Department of Computer Science
Crowdsourced exploration of security configurations.
- Recommended **non-parametric group comparison tests** for user experience survey data.
- Created figures for publication.

https://doi.org/10.1145/2702123.2702370 (Acknowledged)

Olga Scrivner
Ph.D. Student in Computational Linguistics and French Linguistics
Historical change in a language: word order transitions from Object-Verb to Verb-Object.
- Advised on model selection for **logistic linear mixed model** in **Bayesian** framework (JAGS).

2013

Michael S. Willett
Associate Chair, Department of Kinesiology
The effect of implementation of the Revenue Cycle Management system at Indiana University.

- Transformed time-series data to facilitate simple analyses, including **one-sample and two-sample t-tests**, to evaluate differences before and after the implementation of the RCM.
- Developed a custom R plotting function to visualize each of 100+ variables along with its corresponding p-value.