Intermediate R Training: Data Science for Workforce and Economic Development Research

In-person: June 13, 2022 | 9:00am – 4:30pmET
Continues Virtually: June 21 + 23 4:30 – 6:30pmET

Session 1 In-person (must attend in-person to continue in virtual sessions)

Section 1
Refresher
- Refresh on tidyverse and data wrangling.
- Refresh on using blscrapeR and tidycensus packages.
- An API example with PA COVID-19 data

Key Takeaways: Remember basic operations with select, filter, mutate functions. Import data with
Session Materials: See course materials for Week 1 to view slides and data necessary for in-session exercises.

Section 2
Functional programming with R
- Learn how to write your own functions in R.
- Learn running iterative operations with:
  - for loop
  - purrr package
- If time permits: if else loop

Key Takeaway: Learn some key concepts and tools foundational to more scalable and repeatable importing, tidying, and visualization. This content will set us up for the following sessions.

Lunch provided

Section 3
Formal introduction to RMarkdown
- Markdown fundamentals
- R Markdown fundamentals
- YAML and table of contents
- Chunk options
  - Control code execution
  - Adjust plot presentation
- Making plots and tables using kable

Section 4
Build your first Shiny app!
- Shiny essentials and basic principles including
  - basic ui
  - basic reactivity
  - upload
  - download
  - best practices

Key Takeaway: Make your graphs and reporting interactive. Learn basics to allow other users to explore your results interactively.

Session 2: June 21, 2022 (Virtual; building on in-person Session 1)

Section 5

Intro to modeling and feature engineering
- Use previously introduced data sets and data tidying tools too:
  - Explore distributions of key variables
  - Inform feature engineering for rolling metrics
  - Inform feature engineering and model set up for time series data

Key Takeaways: Learn the foundational steps to the most critical part of modeling: the data cleaning, prep, and iterative feature engineering
Session Materials: See course materials for Week 2 to view data necessary for in-session exercises and slides.

Session 3: June 23, 2023 (Virtual; building on previous sessions)

Section 6

Create and run regression model
- Build on feature created to run model
- Learn basics of statistics behind it
- Evaluate the accuracy of the model using key base R plotting features
- Understand how we can expand upon visuals using ggplot2
- Leave with resources needed for self guided investigation of more advanced ML principles (i.e., training, sampling techniques, associated packages).

Key Takeaway: Learn 1st principles of the most foundational and commonly used modeling approach. Learn how to assess model fit and accuracy using a basic, repeatable framework
Session Materials: See course materials for Week 2 to view data necessary for in-session exercises and slides.