

# Introduction to Programming in R

- Week 1
  - Introduction to R
  - RMarkdown
- Week 2
  - Data Manipulation with dplyr
  - Creating Visualizations
- Week 3
  - Reading Data
  - Iterate Over Lists with purrr
  - Reshaping Data
- Week 4
  - Linear Models
  - Generalized Linear Models
  - Assessing Model Quality
- Week 5
  - Cross-Validation
  - Penalized Regression
  - Boosted Trees
- Week 6
  - Shiny Basics
  - Shiny Dashboard
- Instructor Bio
- Textbook

## Syllabus

This class is an intensive introduction to R. It starts with the very basics of assigning variables and reading data. It then progresses to using RMarkdown for document and presentation creation.

## Week 1

Introduction to R

- The RStudio Interface
- Basic Math
- Assigning Variables
- Working Directories
- Relative Paths
- Reading Data
  - Read from text files with readr
  - Read from Excel files with readxl
- Writing Functions

## RMarkdown

- RMarkdown Primer
  - Sections
  - Text Formatting
  - Lists
  - Links
- Integrating R into Markdown
  - Code Chunks
  - Chunk Options
- Including Figures
- Output Formats
  - HTML
  - PDF
  - Word
- Presentations

## Week 2

### Data Manipulation with dplyr

- Understanding a tbl
- Use pipes for cleaner code
- Select columns with select
- Filter rows with filter
- Change and create columns with mutate
- Calculate summary statistics with summarize
- Group data for calculations with group\_by
- Joins with left\_join

### Creating Visualizations

- ggplot2 paradigm
- Aesthetics
- Scatter plots

- Color Coding
- Size
- Shape
- Opacity
- Small multiple plots
- Histograms
- Density Plots
- Combining Layers
- Violin Plots
- Themes

## Week 3

### Reading Data

- CSVs with readr
- Databases with DBI
- JSON with jsonlite
- Web pages with rvest

### Iterate Over Lists with purrr

- Basics of functional programming
- Mapping over a list
- Difference from lapply
- Consistent Data Types
- Mapping to different data types
  - character
  - numeric
  - data.frame
- Mapping functions with multiple arguments

### Reshaping Data

- Convert from wide to long with gather
- Convert from long to wide with spread

## Week 4

### Linear Models

- Simple Linear Model with lm
- The Formula Interface
- Multiple Regression
- Tidying models with broom

- Visualizing models with coefplot

## Generalized Linear Models

- Logistic Regression for Binary Data
- Poisson Regression for Count Data
- Quasipoisson Regression for Overdispersed Count Data

## Assessing Model Quality

- AIC
- BIC

## Week 5

### Cross-Validation

- Use Cross-Validation for Model Assessment

## Penalized Regression

- L1 Penalty (Lasso)
- L2 Penalty (Ridge)
- Implement via the Elastic Net with glmnet
- Tuning Hyperparameters

## Boosted Trees

- Decision Trees
- Boosted Trees
- Fit Model with xgboost

## Week 6

### Shiny Basics

- Inputs
- Outputs
- Reactive Expressions
- HTML Widgets
  - Interactive Plots
  - Interactive Maps
  - Interactive Tables

### Shiny Dashboard

- Server Code

- UI Code
- Dashboard Layout

## Instructor Bio

Jared P. Lander is the Chief Data Scientist of Lander Analytics, a data science and artificial intelligence consulting and training firm based in New York City; the organizer of the New York Open Statistical Programming Meetup—the world's largest R meetup—and the New York R Conference); author of *R for Everyone* and an adjunct professor at Columbia University. With an M.A. from Columbia University in statistics and a B.S. from Muhlenberg College in mathematics, he has experience in both academic research and industry. Very active in the data community, Jared is a frequent speaker at conferences, universities and meetups around the world. His writings on statistics can be found at [jaredlander.com](http://jaredlander.com) and his work has been featured in publications such as *Forbes* and the *Wall Street Journal*.

## Textbook

[R for Everyone](#)