

Syllabus - Intro to R for Journalists: How to Find Great Stories in Data

Introduction Module: R

In this introductory module, you will learn how to configure your computer to work with R. Before you can use it to analyze data, your computer needs the following tools installed:

- A command-line interface to interact with your computer
- The git version control software and a GitHub account
- The latest version of R
- The latest version of RStudio
- An API key from Census.gov (https://api.census.gov/data/key_signup.html)

Module 1: Programming in R

This week you will be introduced to RStudio and learn how to start a new analysis project. You will learn the basics of how to import and explore data with R.

This module will cover:

- A tour of the RStudio IDE
- Syntax for coding in R
- Creating R scripts
- Importing packages
- Good habits for workflow and documentation habits
- How to import data like CSVs, Excel spreadsheets, XML
- Exploring the data's structure

Module 2: Wrangling data

This week you will learn how to transform and analyze data the tidy way using the dplyr package.

This module will cover:

- Filtering, selecting, arranging, mutating, summarizing data
- How to join two data sets for more insight
- Chaining analyses functions with pipes for efficiency and readability

Module 3: [Visualizing data](#)

This week, you'll learn about the grammar of graphics how to use the ggplot2 package to make quick exploratory data visualizations.

This module will cover:

- The aesthetics of data visualizations
- How to create different charts like, bar, box, line, scatterplots
- Grouping for charts
- How to create facets or small multiples with the data
- Labels and titles for visualizations

Module 4: Spatial analysis

This week you will learn how to visualize geographical data and look for neighborhood racial profiling disparities using Census data and traffic stop data from Connecticut.

This module will cover:

- Creating interactive maps with the R Leaflet package
- How to geolocate addresses in R
- Importing and visualizing shapefiles
- Points in a polygon analysis that merges location data and boundaries for deeper insights

Module 5: Publishing for reproducibility

This week you will learn how to use RMarkdown to present your analysis in a narrative format. You'll also learn how to log changes to your project with version-control software and publish your analysis on the Internet.

This module will cover:

- The git version control software and its integration with Github
- How data journalists use GitHub and RMarkdown and other notebooks to publish their work
- How to use the Markdown markup language to annotate RMarkdown
- How to create a new git code repository and start tracking code
- How to connect the repository to GitHub and publish to Github Pages