R Data Course - Syllabus

Ronnie Bailey-Steinitz

2025-09-15

9-Week Intro to Data Analysis & Visualization in R

Pre-Course Workshop (Led by Laban): Welcome to R – Dive into Data

Goal: Set up, open RStudio, load a dataset, and start exploring.
- Install R, RStudio, and packages
- RStudio environment: console, script, environment, project
- Import CSV with `readr::read_csv()`
- Explore with: `glimpse()`, `summary()`, `head()`, `tail()`, `nrow()`
- Basic functions: `mean()`, `median()`, `sd()`, `range()`, `length()`
- **Homework:** Explore your own dataset; apply 3 functions and get summary statistics

---

Week 1: Is My Data Clean? Exploring, Diagnosing, and Visualizing Problems

Dataset: `palmerpenguins` (raw)
- Data types: numeric, character, logical, factor
- Convert types: `as.factor()`, `as.numeric()`
- NA handling: `is.na()`, `sum(is.na())`
- Column checks: `class()`, `str()`, `dplyr::count()`
- **Visualization:** bar plots, histograms
- **Lab Challenge:** Check for NAs, misclassified types, and plot 2 variables

Week 2: Wrangling Basics – Select, Filter, Mutate

Dataset: `palmerpenguins` (raw) or `ratdat::rat_growth`
- Core verbs: `select()`, `filter()`, `mutate()`
- Logical comparisons, `case_when()` recoding
- Pipes: `%>%`
- **Visualization:** histograms or bar plots of filtered subsets
- **Lab Challenge:** Create a new variable, subset, and plot from filtered data

Week 3: Grouping, Summarizing, and Custom Visuals

Dataset: `gapminder` (demographic)
- `group_by()` + `summarise()`, `arrange()`
- Aggregation logic (means, counts)
- **Visualization:** `geom_col()`, `geom_boxplot()`
- **Lab Challenge:** Summarize data by group, visualize differences

Week 4: Joining and Reshaping

Dataset: TBD
- Joins: `left_join()`, `inner_join()`
- Binds: `bind_rows()`, `bind_cols()`
- Reshape: `pivot_longer()`, `pivot_wider()`
- Other: `revalue()` (for incorrect join var inputs)
- Spatial data???
- **Visualization:** faceted plots for reshaped data
- **Lab Challenge:** Join and reshape your data, then visualize

Week 5: Distributions, Skew, and Data Transformations

Dataset: TBD
- Distribution types: normal vs. skewed
- Transform: `log()`, `sqrt()`, z-scores
- **Visualization:** `geom_histogram()`, `geom_density()`, `facet_wrap()`
- **Lab Challenge:** Transform and plot skewed variable (before and after)

Week 6: Relationships – Regression and Correlation

Dataset: TBD
- Scatterplot with `geom_point()` + `geom_smooth(method = "lm")`
- Fit linear models: `lm()`, `summary()`, `cor()`
- **Visualization:** trend lines, colored groups
- **Lab Challenge:** Fit + plot a model, interpret slope and R^2

Week 7: Comparing Groups – Boxplots & Statistical Tests

Dataset: TBD
- Categorical comparisons: `t.test()`, `wilcox.test()`, `aov()`
- Assumptions: normality, variance
- **Visualization:** boxplots, violin plots
- **Lab Challenge:** Compare groups with a test and visualize results

Week 8: Project Lab – Build Your Analysis

- Full RMarkdown pipeline: import → wrangle → summarize → visualize → interpret
- Build from scratch; ask for feedback
- Focus on reproducibility and clarity

Week 9: Final Project Showcase

- Each student gives 5-minute presentation
- Slides optional (HTML from RMarkdown fine)
- Emphasize clean visualizations, clear explanations, and interpretation