Lesson 1: Publication tables and figures in R

Thomas Blakely

2026-09-11

Contents

The papers and our teaching analyses 1
Parts 1
Seminar plan 2
What you need 3

Start with published clinical papers and their public data, then build your own Table 1 and figures. Designed as a **110-minute guided demonstration**, with editable R code running in your browser.

For a beginner workshop with hands-on practice and questions, plan about 3–4 hours.

By the end, you should be able to check a CSV import, build a Table 1 grouped by baseline vision, style it for a report, make boxplots, patient trajectories, and heatmaps, and export editable tables and sharp figures for Word and PowerPoint.

The papers and our teaching analyses

Our Table 1 uses Fasler et al. (2019), *One- and two-year visual outcomes from the Moorfields age-related macular degeneration database*. Look at the paper's Table 1 before starting: it gives us a reason to learn the tools for describing who was studied.

We will create **new teaching outputs**, with our own layout, colours, and explicit data-selection rules. Our Table 1 describes baseline patients by baseline vision category. These are descriptive groups in an observational dataset, not randomised treatment arms.

Part 4 uses Chenoweth et al. (2024), *Gene expression signatures in blood from a West African sepsis cohort define host response phenotypes*. Its public figure source data support boxplots of gene-expression scores, patient trajectories, and heatmaps. The Ghana sepsis data notes document the two teaching CSVs, repeated samples, missing follow-up visits, and attribution to the paper and its Zenodo source data.

The prepared baseline CSV has one row per patient; the follow-up CSV has one row per patient and selected visit. The data notes record the source, preparation, age coarsening, and exclusions. The full preparation script is linked there.

Download the complete export script with the table setup, figure setup, and prepared CSVs in `data/`. Run it from the lesson folder to write the table, sepsis figures, interactive HTML, and Office files into `output/`.

Preview the finished Table 1 before working through the code. Part 4 links to the published sepsis figures for comparison.

Parts

1. **Installing R** — R, an editor, and the `install.packages()` / `library()` distinction that trips up everyone.

2. **Loading data** — read the prepared study files and check types, missing values, category order, and the unit of observation.
3. **Building a Table 1** — `gtsummary`, one verb at a time, then `gt` for a styled header, grouped columns, stripes, and source notes. Compare row and column percentages in a cross table, with an optional extension on formatting regression output.
4. **Basic figures with ggplot2** — Ghana sepsis boxplots, patient trajectories with `ggplotly()`, and a pattern heatmap, with an optional six-panel comparison of the published source data.
5. **Exporting to Word and PowerPoint** — the Moorfields Table 1 and sepsis figures, with PNG, vector, interactive HTML, and optional direct Office exports at sizes suited to each figure.
6. **Now you try: a clinical data project** — use a different published clinical dataset from the Mayo Clinic PBC study to build Table 1, make two figures, and export your report assets independently.

Seminar plan

The core demonstration route takes 110 minutes. Work through the examples together and leave the longer reference sections for later.

Time	Topic	Live focus
0–5 min	Getting started	Introduce the paper, open a browser cell, and explain `library()`.
5–15 min	Loading data	Read the baseline CSV, check one row per patient, and set factor levels.
15–50 min	Table 1	Group by baseline vision, add counts, compare cross-table percentages, then style the table.
50–85 min	Figures	Build sepsis boxplots step by step, patient trajectories, and a pattern heatmap.
85–105 min	Export	Download a Word-readable table and two sepsis figures; choose sizes and formats for Word or PowerPoint.
105–110 min	Questions	Recap the workflow and answer questions.

Choose exercises and demonstrations to fit the session. The plotting exercise takes five minutes; the optional six-panel heatmap takes another ten. Use extra time or follow-up practice for these, the interactive plot, and the direct Office exports. The optional regression-table example adds 5–10 minutes.

Alternative import routes and advanced Office formatting are follow-up reading. Statistical testing and treatment-effect models are outside the core session. The optional regression example demonstrates how to format model output.

After the seminar: allow another 35–45 minutes for the independent clinical project in Part 6, or assign it as take-home practice. It includes a new dataset, hints, worked solutions, and an export checklist.

What you need

A browser and an internet connection. No local R installation is required for the seminar. Open the interactive pages before the session so package downloads can begin; each page starts its own R session and includes its own setup code.

For local R, complete Installing R and put the Moorfields and Ghana CSVs from the linked data notes in your project's `data/` folder. Install `gtsummary`, `gt`, `ggplot2`, `plotly`, `patchwork`, and `svglite`. Optional direct Office exports also use `flextable`, `officer`, and `rvg`. The optional regression table uses `broom` and `broom.helpers`.

← All R tutorials

1. Installing R →