RMRWR chapters

Currently in e-book – outline to Lara Speik in Jan

rmd\_files:

- index.Rmd

- io02-getting-started.Rmd

- io03-tasting.Rmd

- io80-intro-reproducibility.Rmd (04)

- io05-importing.Rmd

- io06a-filter.Rmd (06)

- io05a-columns.Rmd (07)

- io92-mutating\_new\_variables.Rmd (08)

- io80-mutating\_joins.Rmd (09)

- io65-error\_messages.Rmd (10)

- io09-building-blocks.Rmd (11 – or earlier)

- io59-hashtag\_debugging.Rmd (12)

- io60-help\_in\_R.Rmd (13)

- io66-base\_r.Rmd (14)

- io04-updating.Rmd (15)

- io07-major-updates.Rmd (16)

- io81-intermed-reproducibility.Rmd (17)

- io33-\*\* making functions

making-table-one.Rmd (18)

- io30-ttest.Rmd (19)

\*\* chisq here

\*\* design study

- io69-sample\_size.Rmd (21)

- io66-blockrand.rmd (22)

- io48-univariate\_distrib\_plots.Rmd (23)

- io48b-bivariate-scatter-plots.Rmd (24)

- io48j-extension-plots.Rmd (25)

- io48c-scales.Rmd (26)

- io67-ggplot\_helpers.Rmd (27)

- io71-functions\_plots.Rmd (29)

- io76-found\_data.Rmd – importing data 2 (30)

- io23-linear-regression.Rmd (31)

- io23b-logistic-regression.Rmd (32)

- io101-fftrees.Rmd (33)

- io23c-intro-shiny.Rmd (35)

- io24-sharing-shiny.Rmd (36)

- io56-intro\_rmarkdown.Rmd (34)

- io56a-rmarkdown-output.Rmd (38)

- io56b-Rmd-citations-Zotero.Rmd (39)

- io56c-Intro-to-Quarto.Rmd (40)

- io70-r\_cmd\_line.Rmd (41 – other stuff)

- io198-title-holder.Rmd

- io199-references.Rmd

Not in book

Epi like epitools

ROC, diagnostics

Propensity/IPTW

Longitudinal data - Time and time zones with lubridate/lag and lead

Build in consistent callouts

Consistent quick exercises that can be at the bottom of the page

Code exercises/solutions for appendix

Themes

* - intro
* - importing (add redcap)
* - clean/check/wrangle
* - data viz
* - data analysis and modeling
* - communicate results – Rmd, quarto, Shiny
* -repro – projects, here, renv, writing functions, labels, self-documenting code, git/github, code review buddy, targets, deposit in repos, docker/binder, nix

-io2a alt\_intro

-io5a columns

- io8 data validation

-9a timeseries

-10 table one

-11 ggplot singlevar

12 cont vs catvar

13 cont vs cont

14 Rmd

15 repro research

16 study design

17 sample size

18 sources bias

19 randmization blockrand

20 one var 1 group tests

21 on var 2 group tests

22 one var multigrop tests

23 line regression

23b logistic

23c intro shiny

24 sharing shingy

25 logistic again

26 meta-analysis

27 learn consort

28 mockstudy chisq

29a chisq example

30 ttest

30a ttest – add rstatix

31 ttest part 2

32 table ttext broom

33 making table one

New Outline

|  |  |  |  |
| --- | --- | --- | --- |
| Chapter | File | Topic | Theme |
| 1 | index | index.Rmd | Intro/repro |
| 2 | Io02 | Getting started – companion package | intro |
| 3 | Io03 | Tasting menu - the whole game, tables | intro |
| 4 | Io80 | Intro repro | repro |
| 5 | Io05 | Importing Intro | import |
|  |  | DEV – Data Exploration and Validation – dlookr, dataMaid, others  How data obtained, good var names, missing, bad values, validate, diagnose, pointblank package | wrangling |
| 6 | Io06 | filter | wrangling |
| 7 | Io05a | Columns | wrangling |
| 8 | Io09 | building-blocks | intro |
| 9 | Io92 | mutating | wrangling |
| 10 | Io65 | error\_messages | intro |
| 11 |  | Data entry with redcap > excel | repro |
| 12 | Io59 | hashtag\_debugging.Rmd | programming |
| 13 | Io60 | Getting help\_in\_R | intro |
|  |  | Making a reprex, rstudio comm, StackOverflow, etc | Intro help |
|  |  | Labelling data with labeller package for varnames, purr: set\_names | repro |
| 14 | Io66 | base\_r.Rmd | programming |
| 15 | Io04 | updating.Rmd | intro |
|  | Io07 | major-updates.Rmd | intro |
|  | Io81 | intermed-reproducibility aving, naming scripts, project, syntax error red x, rainbow parentheses | repro |
|  | Io33 | making functions | programming |
|  |  | making-table-one.Rmd (18) | studies |
|  | Io33 | ttest.Rmd | stats |
|  |  | chisq here | stats |
|  |  | \*\* design studies | studies |
|  | Io69 | sample\_size.Rmd (21) | studies |
|  | Io66 | blockrand.rmd | studies |
|  | Io48 | univariate\_distrib\_plots.Rmd (23) | dataviz |
|  | Io48b | bivariate-scatter-plots | dataviz |
|  | Io48j | Extension plots | dataviz |
|  | Io48c | scales | dataviz |
|  | Io67 | ggplot\_helpers.Rmd (27) | dataviz |
|  | Io71 | functions\_plots.Rmd (29) | programming |
|  | Io76 | io76-found\_data.Rmd – importing data 2 | import |
|  | Io23 | linear-regression.Rmd | stats |
|  | Io23b | logistic-regression.Rmd | stats |
|  |  | ROC, sens/spec. PPV NPV, etc, code disease as 1/health as 0 | stats |
|  | io101 | fftrees.Rmd (33) | Stats modeling |
|  | Io23c | intro-shiny | comm |
|  | Io24 | sharing-shiny.Rmd (36) | comm |
|  | Io56 | intro\_rmarkdown | comm |
|  | Io56b | Rmd-citations-Zotero.Rmd (39) | comm |
|  | Io56 | Intro-to-Quarto | comm |
|  | Io70 | R\_cmd\_line | programming |
|  |  | Mapping health data with tidycensus, tigris, censustracts | studies |
|  |  | Presentations with PowerPoint | comm |
|  |  | Presentations on Posters | comm |
|  |  | Importing data from REDCap | Import 3 |
|  |  | Lasso | modeling |
|  |  | Marginal effects with models | modeling |
|  |  | Reproducibility with Git & Github | repro |
|  |  | Renv for reproducibility | repro |
|  |  | How to do a code review | repro |
|  |  | Collaboration with Github | repro |
|  |  | Data cleaning with a checklist | repro |
|  |  | Labeling your data, making factors | Self-encoding |
|  |  | Deidentifying data – PHI and synthpop, encryptr | Studies, |
|  |  | Depositing data and code in repositories | repro |
|  |  | Mapping health data with sf, mapboxgl, rayshader | studies |
|  |  | Reusing data and code in Environments – docker, binder, nix | repro |
|  |  | The heemod package and cost-effectiveness | studies |
|  |  | Self-documenting values and code (naming things) | repro |
|  |  | Iteratin of functions with purr | programming |
|  |  | Intro to Tidymodels | modeling |
|  |  | Epitools – epi handbook | studies |
|  |  | Meta-analysis | Studies |
|  |  | CONSORT diagrams | studies |
|  |  | Large surveys with surveyr, tidysurveys | studies |
|  |  | Broom and tidying models | Stat modeling earlier |
|  |  | Using inline code in reports | comm |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Kira Suggestions

-~~Importing SAS datasets (still surprisingly common)~~

-~~Dealing with text (stringr, regular expressions)~~

-~~Longitudinal, clustered, and hierarchical data (random and fixed effects, GLM)~~

~~-Matching (algorithms for matching, propensity score matching)~~

-~~Standardization (e.g. Standardized rates—might fit in epi along with stratification)~~

-~~Missing data (including assessment of missingness and methods for imputation)~~

-Sensitivity analysis

-~~Using AI tools to help write code for you~~

Themes

* - intro
* - importing (add redcap) – importing intro, import > excel, import from REDCap
  + Advanced Import R4DS ch7 control col types, import, bind multiple files from folder, saving to Rd or csv
* - clean/check/wrangle
  + Ch 3 R4DS 2e
  + Rows
  + Cols
  + Rename/reclocate
  + Select, select helpers
  + .keep
  + Piping
  + Group – summarize, .by vs ungroup
  + Rowwise grouping
  + Slice
  + Tidydata in Ch5 R4DS 2e
* - data viz
* - data analysis stats
* - data analysis modeling
* - repro –
  + Include naming things, code-styling, Ch4R4DS, sectioning code, comments,
* programming
  + basics in Cg 2 R3DS 2e
* - communicate results – Rmd, quarto, Shiny, ppt, posters
* -repro – projects, here, renv, writing functions, labels, self-documenting code, git/github, code review buddy, targets, deposit in repos, docker/binder, nix

Chapters by Theme

1. Intro
   1. Index
   2. Getting Started
   3. Tasting
   4. Building blocks, typeof()
   5. Error Messages
   6. Getting Help
   7. View(), Structures matrix/array, objects, recycling. vector/df/list, classes, nesting, and environments, R Fundamentals = <https://csc-ubc-okanagan.github.io/R-Fundamentals/index.html>
   8. Updating
   9. Major Updates
2. Import
   1. Importing intro, including excel, readxl
   2. Found data
   3. Importing data from SAS, Stata, SPSS formats with {haven}
   4. Importing from REDCap
   5. Webscraping
   6. Advanced importing - R4DS ch7 control col types; import, bind multiple files from folder, saving to Rd or csv vs openxlsx to write
3. Reproducibility
   1. Intro to repro
   2. Data Entry – REDCap > Excel, using NLM fields for NIH work
   3. Data cleaning with a checklist
   4. Missing data - (including assessment of missingness and methods for imputation) – {naniar}, {visdat}
   5. Labelling data with labeller package for varnames, purr: set\_names
   6. Codebooks, vars with timepoint, units mg\_dL
   7. Level 2 repro – saving, naming scripts, projects, here, renv
   8. Level 3 repro - naming things, code-styling, Ch4R4DS, sectioning code, comments, git and github, gitignore (version control with git on software-carpentry)
   9. Level 4 repro - labels, self-documenting code, anonymize/deidentify data with encryptr/synthpop, git/github, code review buddy/code checkist (r for repro on software carpentry)
   10. Level 5 repro - writing functions, targets, deposit data/code in repos, docker/binder, nix
4. Wrangling
   1. Intro to DEV
   2. Rows/observations/filter
   3. Columns/select/tidyselect helpers/mutate/rename/relocate/.keep
   4. Labeling your data, making factors
   5. Dates and times, time zones, hms
   6. Time differences, lead and lag, diff functions, padr
   7. Strings and regex, stringr and rebus and websites for regex
   8. Sticking things together – paste, glue, epoxy
   9. Piping
   10. Group – summarize, .by vs ungroup
   11. Rowwise grouping
   12. Slice
   13. Tidydata in Ch5 R4DS 2e
   14. Bigger than RAM data with arrow and duckdb
5. Programming
   1. basics of R in Cg 2 R3DS 2e (programming with R on software-carpentry)
   2. base R vs Tidy dialect
   3. basic hashtag debugging and tidylog
   4. Clean code – styler and linter, outlines and sections
   5. writing functions
   6. functions for plots/ Rmd/Quarto reports with parameters
   7. Iterating with purr
   8. R at the command line (the Unix shell on software-carpentry)
   9. Using AI tools to help write code for you (Github Copilot, others – benefits, costs, security risks)
6. Data Visualization
   1. Univariate distrib plots
   2. Bivariate scatter plots
   3. Extension plots
   4. Scales
   5. Ggplot helpers
7. Communication of Your Results
   1. Rmarkdown
   2. Using Inline code in reports
   3. Adding references to reports
   4. Paramaterized reports
   5. Quarto
   6. Dashboards in Quarto with quarto-pub
   7. Powerpoint and Presentations
   8. Posters with Typst vs Powerpoint
   9. Shiny
   10. Publishing on rpubs, shinyapps.io and quarto-pubs
8. Clinical Studies
   1. Making table one with gtsummary
   2. CONSORT diagrams
   3. Designing studies
   4. Sample size
   5. Randomization
   6. Epidemiology with Epitools and Epi Handbook, standardized rates, stratification
   7. Mapping health data
   8. Monitoring studies in REDCap – Reports and Dashboards, API Keys
   9. Deidentifying data – PHI deletion vs synthpop, encryptr
   10. Heemod and cost-effectiveness
   11. Annual NIH enrollment reports from REDCap with combine\_checkboxes and codified
   12. Large surveys with surveyr, tidysurveys
   13. Meta-analysis in R
9. Stats
   1. Ttest for continuous, rstatix, broom
   2. Categorical with chisq, mcnemars, mosaic
   3. Linear regression, broom, predict
   4. Logistic regression, broom, predict
   5. ROC sens/spec, PPV/NPV, code disease as 1, health as zero
   6. Survival/time to event modeling
   7. Generalized Additive Models (GAM)
   8. Matching (algorithms for matching, propensity score matching), twang package, IPTW
   9. FFTrees
   10. Random Forest
   11. Lasso
   12. Marginal effects
   13. Longitudinal, clustered, and hierarchical data (random and fixed effects, GLM)
   14. Introduction to Tidymodels

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chapter # | Title | Path | file | New filename |
| 1 | Index | Beginning |  |  |
| 2 | Getting started | Beginning |  |  |
| 3 | Tasting | Beginning |  |  |
| 4 | Importing introduction including excel, readxl, readcsv, janitor, write csv, saveRDS | Importing |  |  |
| 5 | Introduction to Reproducibiility: Setting up a project and saving files | Reproducibility |  |  |
| 6 | Introduction to DEV (Data Evaluation and Validation) and Tidy data | Wrangling |  |  |
| 7 | Basics of R | Programming |  |  |
| 8 | Univariate Distrib Plots, layers and structure of ggplot | DataViz |  |  |
| 9 | Making TableOne with gtsummary | Clinical Studies |  |  |
| 10 | Centrality tests – ttest, rstatix, broom | Statistics |  |  |
| 11 | Rmarkdown | Communication |  |  |
| 12 | Building blocks of data in R – types, objects, classes | Beginning |  |  |
| 13 | Data entry – REDCap > Excel, using NLM fields for NIH, Common Data Elements, build dataset | Import |  |  |
| 14 | Wrangling Rows with Filter and Slice | Wrangling |  |  |
| 15 | Wrangling Columns with Select mutate, rename, relocate, .keep, tidyselect helpers | Wrangling |  |  |
| 16 | Base R vs Tidy dialect and literate programming | Programming |  |  |
| 17 | Bivariate Scatter plots and correlation | DataViz |  |  |
| 18 | Using inline code and adding references to Rmd reports | Communication |  |  |
| 19 | CONSORT diagrams with the flowchart package | Clinical Studies |  |  |
| 20 | Labeling your data with {labelled}, purr: set\_names, and making factors | Wrangling |  |  |
| 21 | Analyzing categorical stats with chisq, fisher, mcnemar, mosaic | Statistics |  |  |
| 22 | Error Messages in R | Beginning |  |  |
| 23 | Importing Data from REDCap | Importing |  |  |
| 24 | Data Cleaning with a Checklist | Reproducibility |  |  |
| 25 | Basic Hashtag debugging, tidylog, and reality-checking data step by step | Programming |  |  |
| 26 | GG Extension Plots | DataViz |  |  |
| 27 | Study Designs | Clinical Studies |  |  |
| 28 | Linear regression, test assumptions, broom, predict(skip anova) | Statistics |  |  |
| 29 | Parameterized Reports in Rmarkdown | Communication |  |  |
| 30 | Getting Help in R | Beginning |  |  |
| 31 | Advanced importing – R4DS ch7, control col types, import, bind multiple files from one folder, save to RDS or csv or openxlsx to write | Importing |  |  |
| 32 | Dealing with missing data – naniar, visdat, types of missingness, imputation | Reproducibility |  |  |
| 33 | Writing clean code – styler, linter, outlines, sections, style guides, parent and child docs in R and Rmd | Programming |  |  |
| 34 | Organizing projects, codebooks, varnames with units mg\_dL, without timepoints | Reproducibility |  |  |
| 35 | Quarto and dashboards to quarto-pub | Communication |  |  |
| 36 | Dates and times, lubridate, timezones, hms, time differences, lead and lag, padr | Wrangling |  |  |
| 37 | Sample size calculation in R | Clinical Studies |  |  |
| 38 | Epidemiology with epitools and Epi Handbook | Clinical Studies |  |  |
| 39 | Logistic regression, test assumptions, broom, predict | Statistics |  |  |
| 40 | Structures, View, class, objects, matrix/array, recycling, vector/df/list/ nesting, environments, R fundamentals | Beginning |  |  |
| 41 | How to Learn a New Package (dplyr example) | Beginning |  |  |
| 42 | Found data on the web | Importing |  |  |
| 43 | Level 2 Reproducibility – Projects, saving and naming scripts/Rmd/Qmd, here, renv | Reproducibility |  |  |
| 44 | Writing functions | Programming |  |  |
| 45 | Scales in ggplot | Dataviz |  |  |
| 46 | ROC, sens/spec. PPV/NPV. Code disease as 1/health zero | Statistics |  |  |
| 47 | Strings and Regex, rebus, regex websites.stringr | Wrangling |  |  |
| 48 | Powerpoint and Presentations | Communication |  |  |
| 49 | Webscraping with rvest | Importing |  |  |
| 50 | Level 3 Reproducibilty- naming things, ch4R4DS, sectioning code, comments, git and github, gitignore, (version control from softcarp) | Reproducibility |  |  |
| 51 | Sticking strings together with paste, glue, epoxy | Wrangling |  |  |
| 52 | Piping and grouping =, groupby summarize, rowwise, .keep | Wrangling |  |  |
| 53 | Functions for plots/Iterating with purrr | Programming |  |  |
| 54 | GGplot helpers | DataViz |  |  |
| 55 | Mapping Health Data | Clinical Studies |  |  |
| 56 | Survival/Time to Event Modeling, Cox PH | Statistics |  |  |
| 57 | Updating R and RStudio, packages, minor vs major updates and renv, x.0 updates | Beginning |  |  |
| 58 | Monitoring Studies in REDCap- reports, API keys, and dashboards | Clinical Studies |  |  |
| 59 | Edge case modeling – GAMs, Poisson, ZIP, and ordinal data | Statistics |  |  |
| 60 | Deidentifying data = PHI deletion vs encryptr vs synthpop | Clinical Studies |  |  |
| 61 | Posters with Typst vs Powerpoint/officedown | Communication |  |  |
| 62 | Matching, propensity scores, twang package and IPTW | Statistics |  |  |
| 63 | Tidydata in R4DS ch5 2e vs data table, dtplyr | Wrangling |  |  |
| 64 | R at the command line, unix shell on SoftCarp | Programming |  |  |
| 65 | Shiny to share models interactively | Communication |  |  |
| 66 | Heemod and Cost-effectiveness | Clinical Studies |  |  |
| 67 | Transparent Modeling with FFTrees | Statistics |  |  |
| 68 | Annual NIH Enrollment tables from REDCap with combine\_checkboxes and codified | Clinical Studies |  |  |
| 69 | Random Forest Modeling | Statistics |  |  |
| 70 | Bigger than RAM data in R with dtplyr, arrow and duckdb | Wrangling |  |  |
| 71 | Lasso and | Statistics |  |  |
| 72 | Using AI Tools to help write repeated code for you (Github Copilot, others) – benefits, costs, security risks | Programming |  |  |
| 73 | Using functions in targets, depositing data and code in repos, docker/binder, Nix | Reproducibility |  |  |
| 74 | Longitudinal, clustered, and hierarchical data (random and mixed effects with glm) | Statistics |  |  |
| 75 | Publishing on Rpubs, shinyapps.io, and quarto-pubs | Communication |  |  |
| 76 | Large Surveys with surveyr, tidysurveys | Clinical Studies |  |  |
| 77 | Marginal Effects | Statistics |  |  |
| 78 | Meta-analysis in R | Clinical Studies |  |  |
| 79 | Introduction to Tidymodels | Statistics |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |