
R for Data Science

*Import, Tidy, Transform, Visualize,
and Model Data*

Hadley Wickham and Garrett Grolemund

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

Table of Contents

Preface.....	ix
--------------	----

Part I. Explore

1. Data Visualization with ggplot2.....	3
Introduction	3
First Steps	4
Aesthetic Mappings	7
Common Problems	13
Facets	14
Geometric Objects	16
Statistical Transformations	22
Position Adjustments	27
Coordinate Systems	31
The Layered Grammar of Graphics	34
2. Workflow: Basics.....	37
Coding Basics	37
What's in a Name?	38
Calling Functions	39
3. Data Transformation with dplyr.....	43
Introduction	43
Filter Rows with filter()	45
Arrange Rows with arrange()	50
Select Columns with select()	51

Add New Variables with <code>mutate()</code>	54
Grouped Summaries with <code>summarize()</code>	59
Grouped Mutates (and Filters)	73
4. Workflow: Scripts.....	77
Running Code	78
RStudio Diagnostics	79
5. Exploratory Data Analysis.....	81
Introduction	81
Questions	82
Variation	83
Missing Values	91
Covariation	93
Patterns and Models	105
ggplot2 Calls	108
Learning More	108
6. Workflow: Projects.....	111
What Is Real?	111
Where Does Your Analysis Live?	113
Paths and Directories	113
RStudio Projects	114
Summary	116

Part II. Wrangle

7. Tibbles with tibble.....	119
Introduction	119
Creating Tibbles	119
Tibbles Versus <code>data.frame</code>	121
Interacting with Older Code	123
8. Data Import with readr.....	125
Introduction	125
Getting Started	125
Parsing a Vector	129
Parsing a File	137
Writing to a File	143
Other Types of Data	145

9. Tidy Data with tidyr.....	147
Introduction	147
Tidy Data	148
Spreading and Gathering	151
Separating and Pull	157
Missing Values	161
Case Study	163
Nontidy Data	168
10. Relational Data with dplyr.....	171
Introduction	171
nycflights13	172
Keys	175
Mutating Joins	178
Filtering Joins	188
Join Problems	191
Set Operations	192
11. Strings with stringr.....	195
Introduction	195
String Basics	195
Matching Patterns with Regular Expressions	200
Tools	207
Other Types of Pattern	218
Other Uses of Regular Expressions	221
stringi	222
12. Factors with forcats.....	223
Introduction	223
Creating Factors	224
General Social Survey	225
Modifying Factor Order	227
Modifying Factor Levels	232
13. Dates and Times with lubridate.....	237
Introduction	237
Creating Date/Times	238
Date-Time Components	243
Time Spans	249
Time Zones	254

Part III. Program

14. Pipes with magrittr.....	261
Introduction	261
Piping Alternatives	261
When Not to Use the Pipe	266
Other Tools from magrittr	267
15. Functions.....	269
Introduction	269
When Should You Write a Function?	270
Functions Are for Humans and Computers	273
Conditional Execution	276
Function Arguments	280
Return Values	285
Environment	288
16. Vectors.....	291
Introduction	291
Vector Basics	292
Important Types of Atomic Vector	293
Using Atomic Vectors	296
Recursive Vectors (Lists)	302
Attributes	307
Augmented Vectors	309
17. Iteration with purrr.....	313
Introduction	313
For Loops	314
For Loop Variations	317
For Loops Versus Functionals	322
The Map Functions	325
Dealing with Failure	329
Mapping over Multiple Arguments	332
Walk	335
Other Patterns of For Loops	336

Part IV. Model

18. Model Basics with modelr.....	345
Introduction	345
A Simple Model	346
Visualizing Models	354
Formulas and Model Families	358
Missing Values	371
Other Model Families	372
19. Model Building.....	375
Introduction	375
Why Are Low-Quality Diamonds More Expensive?	376
What Affects the Number of Daily Flights?	384
Learning More About Models	396
20. Many Models with purrr and broom.....	397
Introduction	397
gapminder	398
List-Columns	409
Creating List-Columns	411
Simplifying List-Columns	416
Making Tidy Data with broom	419

Part V. Communicate

21. R Markdown.....	423
Introduction	423
R Markdown Basics	424
Text Formatting with Markdown	427
Code Chunks	428
Troubleshooting	435
YAML Header	435
Learning More	438
22. Graphics for Communication with ggplot2.....	441
Introduction	441
Label	442
Annotations	445

Scales	451
Zooming	461
Themes	462
Saving Your Plots	464
Learning More	467
23. R Markdown Formats.....	469
Introduction	469
Output Options	470
Documents	470
Notebooks	471
Presentations	472
Dashboards	473
Interactivity	474
Websites	477
Other Formats	477
Learning More	478
24. R Markdown Workflow.....	479
Index.....	483