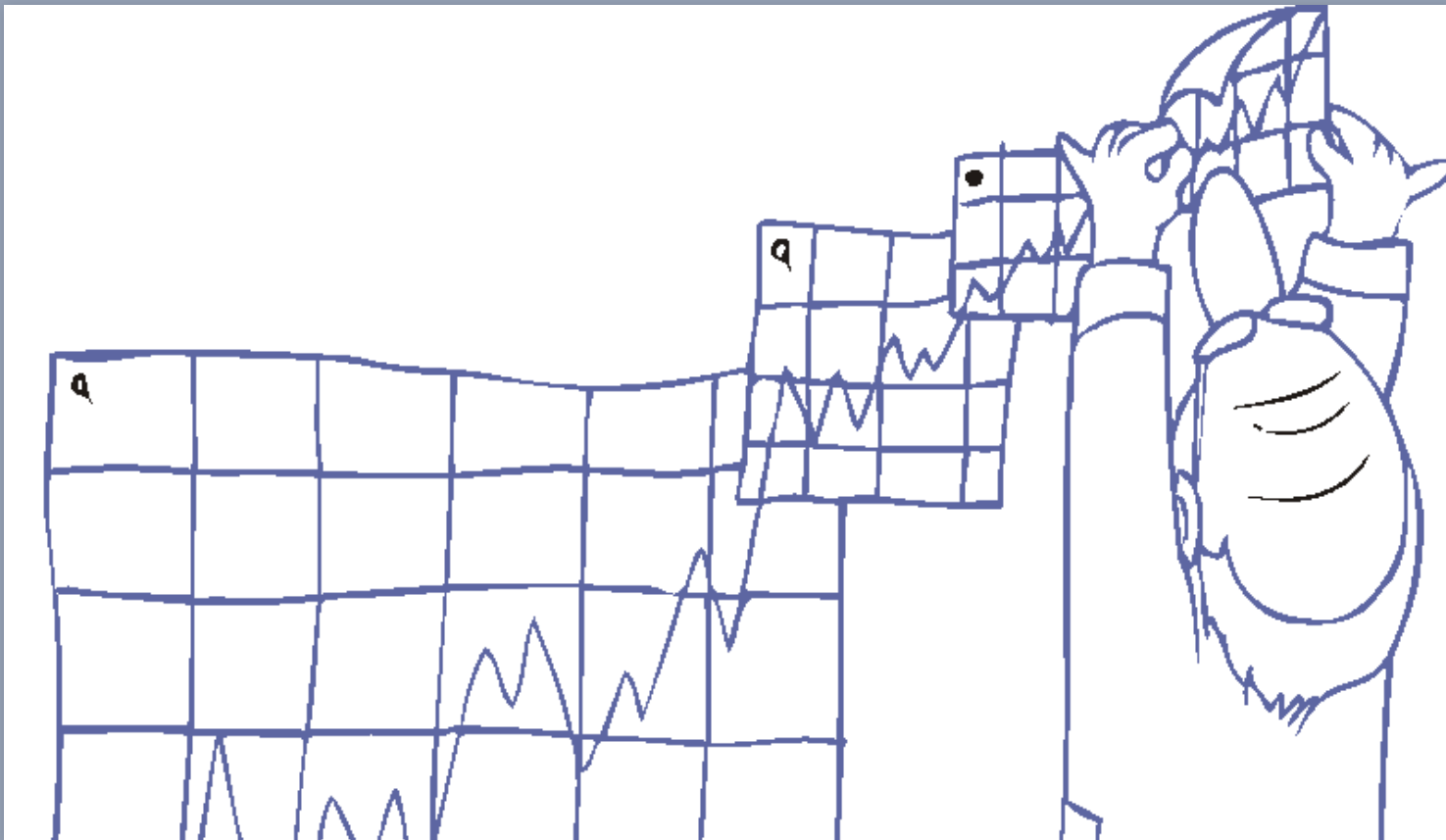


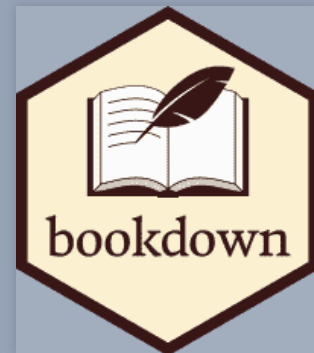
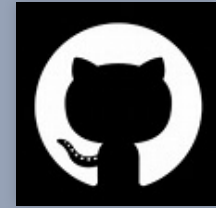
R from Start to Finish: A Case Study



Overview

- GitHub & R project
- feather: “a fast, lightweight, and easy-to-use binary file format for storing data frames”
short-term as part of analysis¹
- flextable
- bookdown @ <https://bookdown.org/>

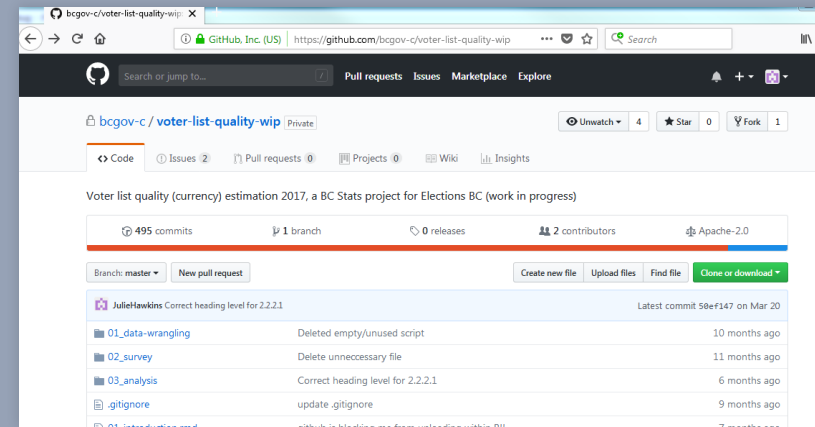
1: <https://blog.cloudera.com/blog/2016/03/feather-a-fast-on-disk-format-for-data-frames-for-r-and-python-powered-by-apache-arrow/>



GitHub & R project

@ <https://github.com>

- connected GitHub repo to an R project to keep files together and organized
- used .gitignore file to ensure that confidential data files did not push to GitHub
- provided version control and issue notification for project with more than one contributor



Feather

- project required working with massive datasets (@1.1 GB text file)
 - needed to sample from $N = 1,168,572$; reading in 450 MB text file with `read_csv` took @9 minutes
 - post-survey, read in 551 MB worked on csv file (took @12 minutes)
 - read in 740 MB text file with `read_csv` (took @15 minutes)
 - merge of these two files with some additional information took HOURS to `write_csv`
- `library(feather)`

• `write_feather(m, path)` where file has feather

Flextable

@ <https://cran.r-project.org/web/packages/flextable/vignettes/overview.html>

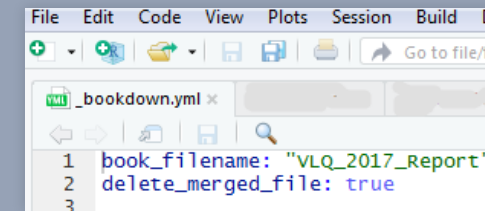
Bookdown

@ <https://bookdown.org/yihui/bookdown/>

- install the development versions of `bookdown` from GitHub:
`devtools::install_github("rstudio/bookdown")`
- `library(bookdown)` built on R Markdown and knitr
- `bookdown` (and `knitr`) package author: Yihui Xie, a software engineer at RStudio
- project required Word document output
 - `bookdown` meant to output formats such as HTML, PDF, and e-books

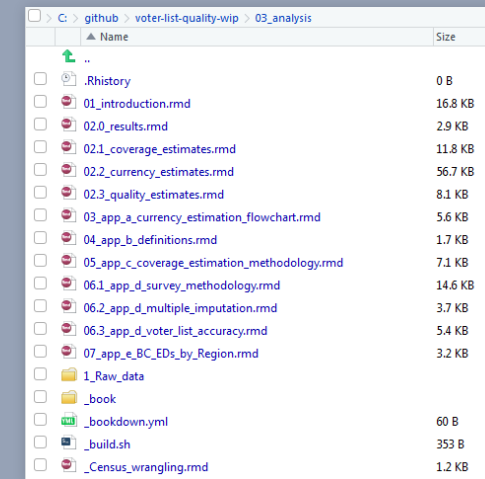
Bookdown Learnings

- site: “bookdown::bookdown_site” is REQUIRED to knit all rmds together
- you must have either an `index.rmd` OR `_bookdown.yml` file for bookdown to work
- bookdown will include all `.rmd` files that are not named starting with an underscore
- number the rmd files for bookdown to knit them in order
- each `.rmd` file that you want in the book must have one and only one chapter (i.e. header with one #)



The screenshot shows the RStudio interface with the `_bookdown.yml` file open in the editor. The file contains the following content:

```
1 book_filename: "VLQ_2017_Report"
2 delete_merged_file: true
3
```



The screenshot shows a file explorer window displaying the contents of the `03_analysis` directory. The files and their sizes are listed in the following table:

Name	Size
..	0 B
.Rhistory	0 B
01_introduction.rmd	16.8 KB
02.0_results.rmd	2.9 KB
02.1_coverage_estimates.rmd	11.8 KB
02.2_currency_estimates.rmd	56.7 KB
02.3_quality_estimates.rmd	8.1 KB
03_app_a_currency_estimation_flowchart.rmd	5.6 KB
04_app_b_definitions.rmd	1.7 KB
05_app_c_coverage_estimation_methodology.rmd	7.1 KB
06.1_app_d_survey_methodology.rmd	14.6 KB
06.2_app_d_multiple_imputation.rmd	3.7 KB
06.3_app_d_voter_list_accuracy.rmd	5.4 KB
07_app_e_BC_EDs_by_Region.rmd	3.2 KB
1_Raw_data	
_book	
_bookdown.yml	60 B
_build.sh	353 B
_Census_wrangling.rmd	1.2 KB

Conclusion

Thanks!



Any questions?

Julie.Hawkins@gov.bc.ca