

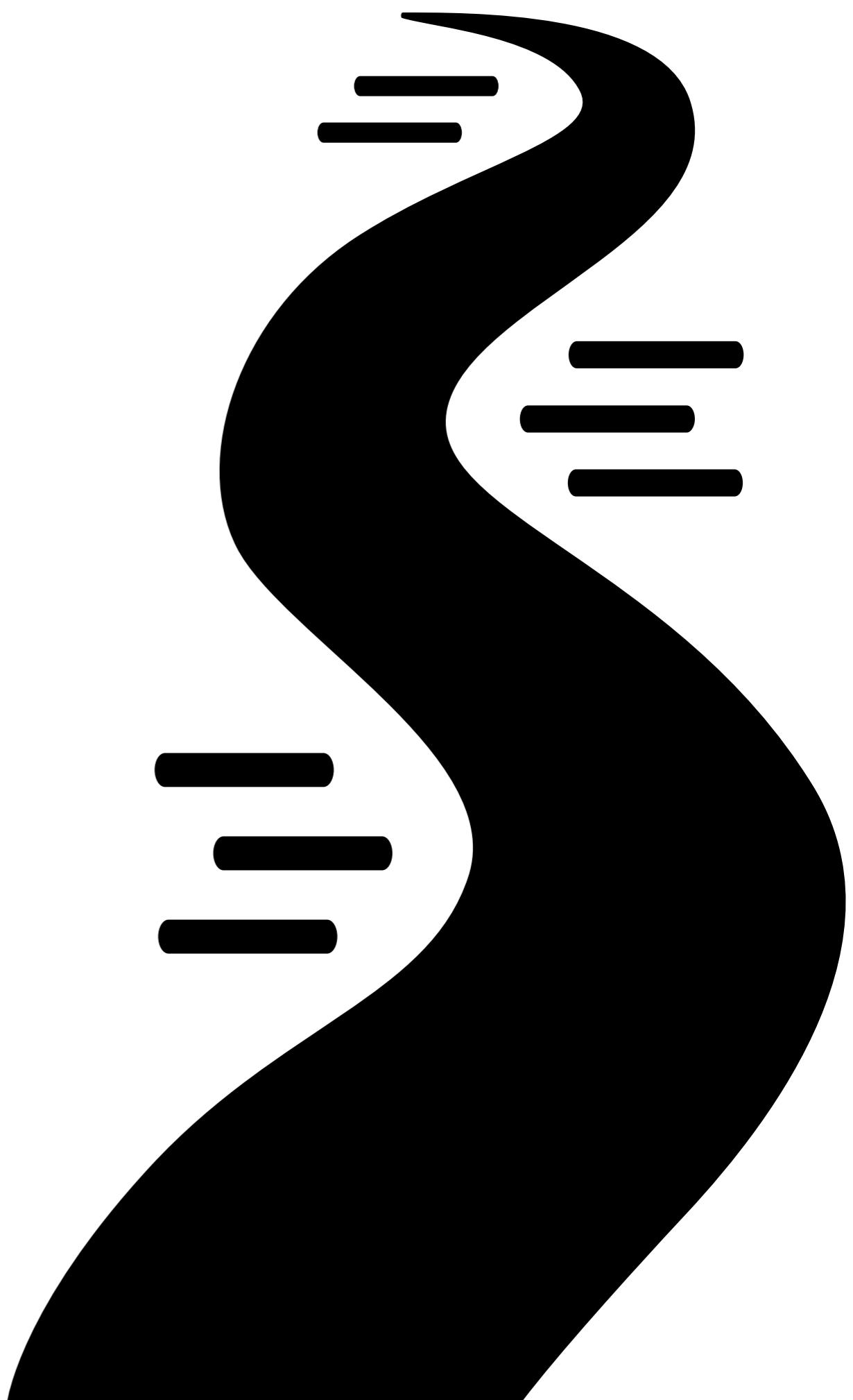
carlisle + matt

on workflow

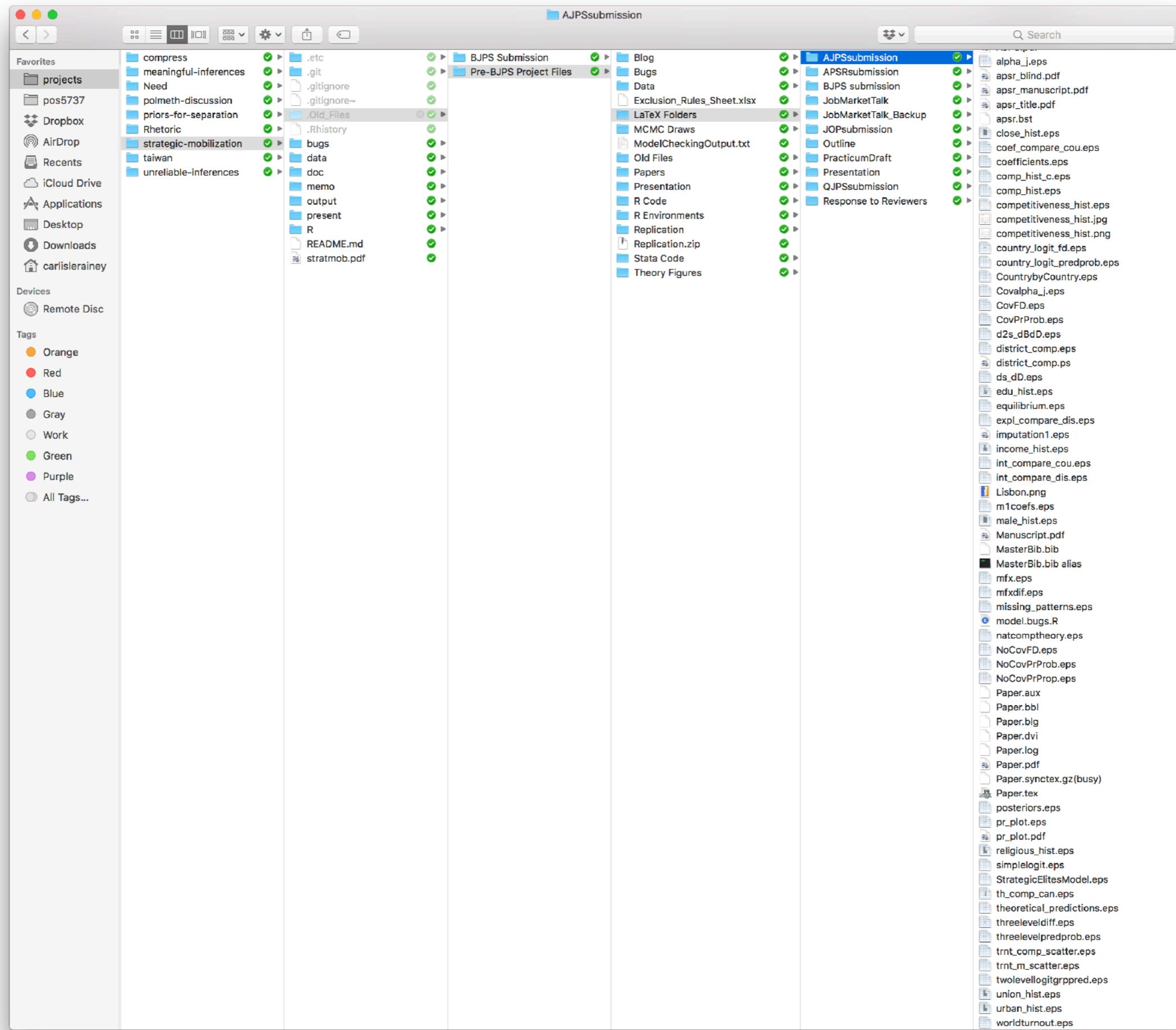


carlisle's 10 commandments of workflow

- I. Plot your data. Use small multiples.
- II. Use plain text.
- III. Use a README to carefully explain the structure of the files.
- IV. Preserve the unchanged, raw data in the original format.
- V. Document your data sources carefully.
- VI. Use comments liberally to explain changes you make to the raw data—what and why.
- VII. Automate.
- VIII. Use open source software.
- IX. Keep a version control history.
- X. Use GitHub! Sharability. Browsability.



What does a research
project look like?



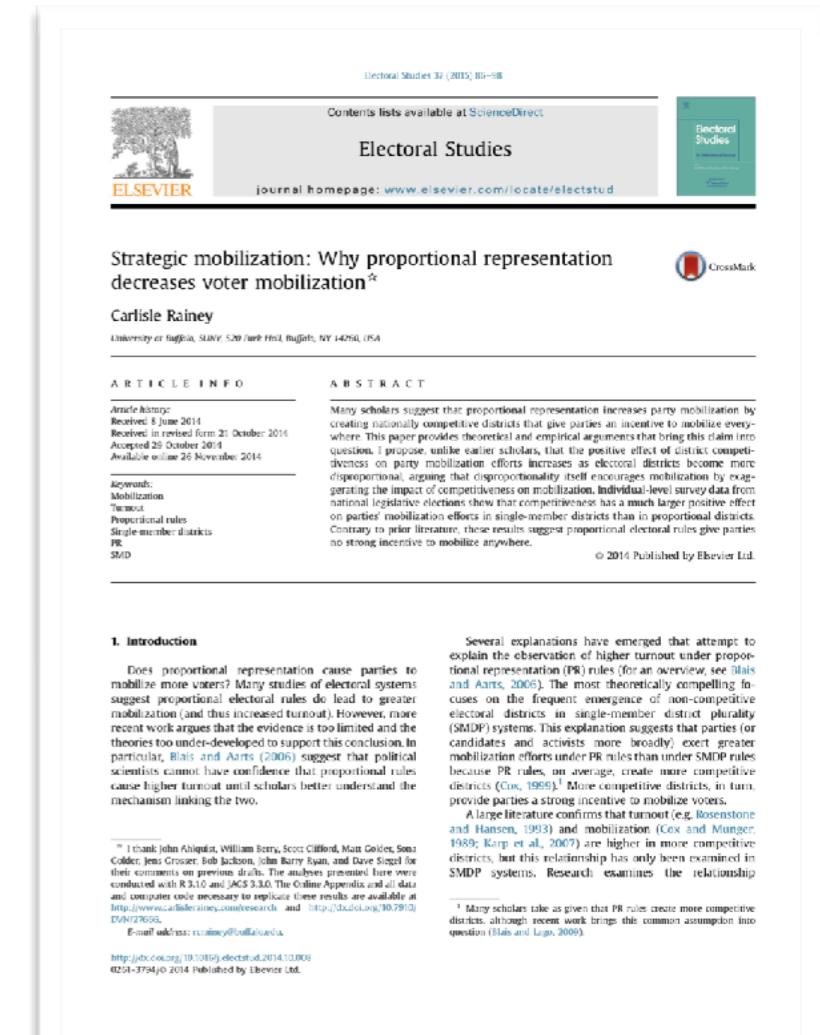
raw data

the process

**100s of decisions
1000s of lines of code**

tidy modeling plotting

**principled
implemented
documented**





The Process

“The process never ends until we die. And the choices we make are ultimately our own responsibility.”

—Eleanor Roosevelt

principled

if you made correct decisions

implemented

if you did what you decided to do

documented

if you can check that you did what you decided to do

principled

multiple imputation or list-wise deletion?

least squares or a robust alternative?

random effects or fixed effects?

excluding post-treatment variables?

computing the quantity of greatest interest?

implemented

two sources of errors

errors in a script

software errors

user errors

```

# Make sure that working directory is set properly
# setwd("~/Dropbox/projects/strategic-mobilization/")

# Clear workspace
rm(list = ls())

# Read in the raw data from the CSES Module 2 data set
cses2 <- read.csv("data/cses2_rawdata.txt")

# Pull out variables of interest
mycses2 <- c("B1004", "B2001", "B2002", "B2003", "B2004", "B2005", "B2020", "B2023", "B2030", "B2031", "B3001_2", "B3002_2",
"B3003", "B3004_1", "B3014", "B3016", "B3028", "B3045", "B3047_1", "B3047_2", "B3047_3", "B4001", "B4002", "B4003", "B4004_A",
"B4004_B", "B4004_C", "B4004_D", "B4004_E", "B4004_F", "B4005", "B5043_1")
cses2 <- cses2[, mycses2]

# Change the variable names
names(cses2) <- c("Alpha.Polity", "Age", "Male", "Education", "Married", "Union.Member", "Household.Income",
"Religious.Attendance", "Urban", "District", "Campaign.Activities", "Freq.Campaign", "Contacted", "Cast.Ballot", "Vote.Matters",
"Cast.Ballot.Previous", "Close.To.Party", "Ideology", "Know1", "Know2", "Know3", "Number.Seats", "Number.Candidates",
"Number.Lists", "VoteA", "VoteB", "VoteC", "VoteD", "VoteE", "VoteF", "District.Turnout", "Electoral.Formula")

# Drop countries for which there is not information about the electoral district
cses2 <- cses2[cses2$District != 99999, ]
cses2 <- cses2[cses2$Number.Seats != 999, ]

#### Recode and Create Variables

# Alpha.Polity

cses2$Alpha.Polity <- as.character(cses2$Alpha.Polity)

cses2$Alpha.Polity[cses2$Alpha.Polity == "CAN_2004"] <- "Canada"
cses2$Alpha.Polity[cses2$Alpha.Polity == "FIN_2003"] <- "Finland"
cses2$Alpha.Polity[cses2$Alpha.Polity == "GBR_2005"] <- "Great Britain"
cses2$Alpha.Polity[cses2$Alpha.Polity == "PRT_2002"] <- "Portugal 2002"
cses2$Alpha.Polity[cses2$Alpha.Polity == "PRT_2005"] <- "Portugal 2005"

cses2 <- cses2[cses2$Alpha.Polity == "Canada" |
               cses2$Alpha.Polity == "Finland" |
               cses2$Alpha.Polity == "Great Britain" |
               cses2$Alpha.Polity == "Portugal 2002" |
               cses2$Alpha.Polity == "Portugal 2005", ]

cses2$Alpha.Polity <- as.factor(cses2$Alpha.Polity)

# Age

```

```

cses2$District.Country <- paste(cses2$Alpha.Polity, cses2$District, sep = "")  

cses2$District.Country <- as.factor(cses2$District.Country)  
  

District.Names <- sort(unique(cses2$District.Country))  

for (i in 1:length(District.Names)) {  

  cses2$District[cses2$District.Country == District.Names[i]] <- i  

}  
  

#####
## Save datasets as .csv files      ##  

#####  

cses2$District <- as.numeric(as.character(cses2$District))  

cses2$Country <- as.numeric(cses2$Alpha.Polity)  
  

# Save a listwise-deleted data set.  

ld.vars <- c("Contacted", "Age", "Male", "Education", "Married", "Union.Member", "Household.Income", "Urban", "Close.To.Party",  

"District.Competitiveness", "ENEP", "PR", "Alpha.Polity", "District", "Country", "District.Country")  

ld.data <- cses2[, ld.vars]  

ld.data <- na.omit(ld.data)  

write.csv(ld.data, "output/ld-data.csv")  
  

# Save a data set with missing values for multiple imputation.  

mi.vars <- c("Alpha.Polity", "Age", "Male", "Education", "Married", "Union.Member", "Household.Income", "Religious.Attendance",  

"Urban", "District.Competitiveness", "ENEP", "PR", "Contacted", "Candidate.Ballot", "Vote.Matters", "Cast.Ballot.Previous",  

"Close.To.Party", "Ideology", "Knowledge1", "Knowledge2", "District.Competitiveness", "SMDP", "Number.Seats", "ENEP", "Country",  

"District")  

mi.data <- cses2[, mi.vars]  

write.csv(mi.data, "output/mi-data.csv")

```

Pr(correct) < 1

```

# Create the district-level data  

get.first <- function(x) {  

  return(x[1])  

}  
  

district.data <- cses2[, c("Alpha.Polity", "Country", "District", "District.Competitiveness", "PR")]  

district.data <- aggregate(district.data, by = list(cses2$District), FUN = get.first)  

district.data$SMDP <- 1 - district.data$PR  

write.csv(district.data, "output/district-data.csv")  
  

# Create the country-level data  

country.data <- cses2[, c("Alpha.Polity", "Country", "PR")]  

country.data <- aggregate(country.data, by = list(cses2$Country), FUN = get.first)  

country.data$SMDP <- 1 - country.data$PR  

write.csv(country.data, "output/country-data.csv")

```

two sources of errors

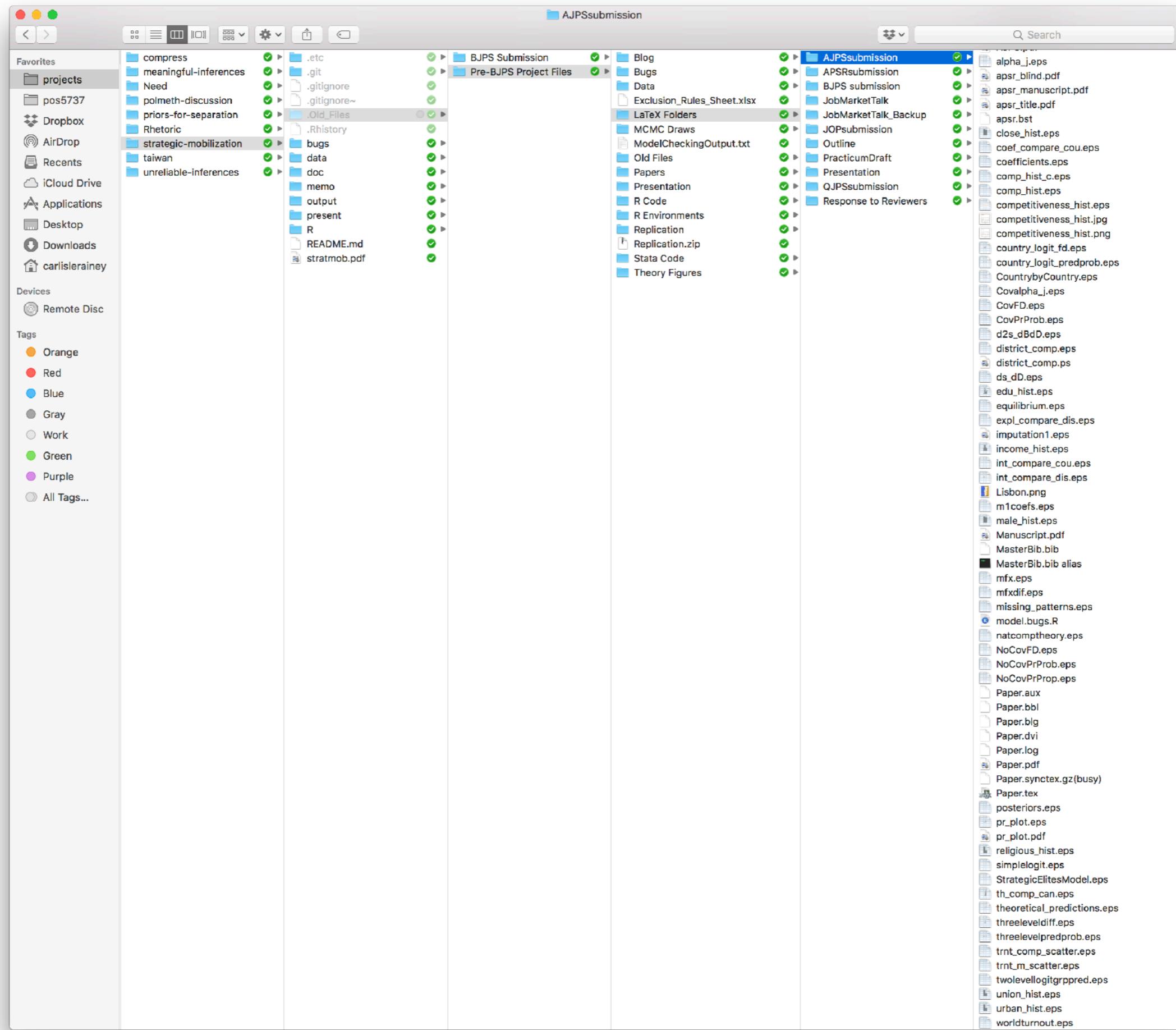
errors in a script

software errors

user errors

mismanage your files

mismanage versions



fit-model-v4.2_withBayes-add_GDP-APSRrevisions-(for Scott).R

two sources of errors

errors in a script

software errors

user errors

mismange your files

mismange versions

mismange dependencies



documented

share your work

Carlisle's Fundamental Theorem of Implementation (CFTI)

The same strategies that allow others to easily check your work (1) allow you to easily check your work and (2) ensure that you implement your decisions correctly in the first place.

My Tools



statistical computing



document creating

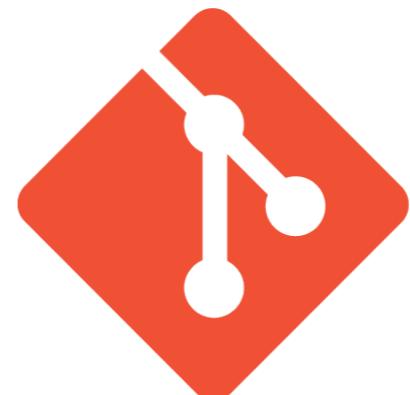
LAT_EX



dependency defining



version controlling



git





Makefile — ~/Dropbox/projects/unnecessary

Project

unnecessary

- old
- .git
- .Rproj.user
- data
- doc
- memo
- present
- R
- .DS_Store
- .gitignore
- .Rapp.history
- .Rhistory
- ~.bash_profile

Makefile

```
1 # phony
2 all: paper dag
3 paper: doc/unnecessary.pdf
4 dag: makefile-dag.png
5
6 # draw makefile dag
7 makefile-dag.png: makefile-dag.R Makefile
8   Rscript $<
9
10 # do and plot poisson simulations
11 doc/figs/poisson-mcs.pdf: R/poisson-mcs.R
12   Rscript $<
13
14 # do intuition and plot simulations
15 doc/figs/intuition-1.pdf doc/figs/intuition-2.pdf doc/figs/intuition-3.pdf doc/figs/intuition-4.pdf doc/fig
16   Rscript $<
17
18 # do logit simulations
19 data/nagler-fd-bias.rds: R/nagler-fd-sims.R data/scobit.dta
20   Rscript $<
21
22 # plot logit simulations
23 doc/figs/nagler-fd-bias.pdf: R/plot-nagler-fd-bias.R data/nagler-fd-bias.rds
24   Rscript $<
25
26 # do ge example
27 doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf: R/ge.R data/ge.csv
28   Rscript R/ge.R
29
30 # compile manuscript
31 doc/unnecessary.pdf: doc/unnecessary.tex doc/bibliography.bib doc/figs/intuition-1.pdf doc/figs/intuition-2
32   # cd into doc so that pdflatex runs in the doc directory
33   cd doc; pdflatex unnecessary
34   cd doc; bibtex unnecessary
35   cd doc; pdflatex unnecessary
36   cd doc; pdflatex unnecessary
37   cd doc; rm -f *.bbl *.log *.synctex.gz *.aux *.out *blg
38 cleanpaper:
39   rm -f doc/unnecessary.pdf
40
41 cleandag:
42   rm -f makefile-dag.png
43
44 cleanALL: cleanpaper
45   rm -f doc/figs/poisson-mcs.pdf
46   rm -f doc/figs/intuition-*.pdf
47   rm -f doc/figs/nagler-fd-bias.pdf data/nagler-fd-bias.rds data/nagler-fd-bias.csv
48   rm -f doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf
```

Makefile 1:1

LF UTF-8 Makefile master Fetch 2 files

Makefile — ~/Dropbox/projects/unnecessary

Project

unnecessary

- old
- .git
- .Rproj.user
- data
- doc
- memo
- present
- R
- .DS_Store
- .gitignore
- .Rapp.history
- .Rhistory
- ~.bash_profile

Makefile

```
1 # phony
2 all: paper dag
3 paper: doc/unnecessary.pdf
4 dag: makefile-dag.png
5
6 # draw makefile dag
7 makefile-dag.png: makefile-dag.R Makefile
8   Rscript $<
9
10 # do and plot poisson simulations
11 doc/figs/poisson-mcs.pdf: R/poisson-mcs.R
12   Rscript $<
13
14 # do intuition and plot simulations
15 doc/figs/intuition-1.pdf doc/figs/intuition-2.pdf doc/figs/intuition-3.pdf doc/figs/intuition-4.pdf doc/fig
16   Rscript $<
17
18 # do logit simulations
19 data/nagler-fd-bias.rds: R/nagler-fd-sims.R data/scobit.dta
20   Rscript $<
21
22 # plot logit simulations
23 doc/figs/nagler-fd-bias.pdf: R/plot-nagler-fd-bias.R data/nagler-fd-bias.rds
24   Rscript $<
25
26 # do ge example
27 doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf: R/ge.R data/ge.csv
28   Rscript R/ge.R
29
30 # compile manuscript
31 doc/unnecessary.pdf: doc/unnecessary.tex doc/bibliography.bib doc/figs/intuition-1.pdf doc/figs/intuition-2
32   # cd into doc so that pdflatex runs in the doc directory
33   cd doc; pdflatex unnecessary
34   cd doc; bibtex unnecessary
35   cd doc; pdflatex unnecessary
36   cd doc; pdflatex unnecessary
37   cd doc; rm -f *.bbl *.log *.synctex.gz *.aux *.out *blg
38 cleanpaper:
39   rm -f doc/unnecessary.pdf
40
41 cleandag:
42   rm -f makefile-dag.png
43
44 cleanALL: cleanpaper
45   rm -f doc/figs/poisson-mcs.pdf
46   rm -f doc/figs/intuition-*.pdf
47   rm -f doc/figs/nagler-fd-bias.pdf data/nagler-fd-bias.rds data/nagler-fd-bias.csv
48   rm -f doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf
```

target

Makefile 1:1

LF UTF-8 Makefile master Fetch 2 files

Makefile — ~/Dropbox/projects/unnecessary

Project

unnecessary

- old
- .git
- .Rproj.user
- data
- doc
- memo
- present
- R
- .DS_Store
- .gitignore
- .Rapp.history
- .Rhistory
- ~.bash_profile

Makefile

```
1 # phony
2 all: paper dag
3 paper: doc/unnecessary.pdf
4 dag: makefile-dag.png
5
6 # draw makefile dag
7 makefile-dag.png: makefile-dag.R Makefile
8   Rscript $<
9
10 # do and plot poisson simulations
11 doc/figs/poisson-mcs.pdf: R/poisson-mcs.R
12   Rscript $<
13
14 # do intuition and plot simulations
15 doc/figs/intuition-1.pdf doc/figs/intuition-2.pdf doc/figs/intuition-3.pdf doc/figs/intuition-4.pdf doc/fig
16   Rscript $<
17
18 # do logit simulations
19 data/nagler-fd-bias.rds: R/nagler-fd-sims.R data/scobit.dta
20   Rscript $<
21
22 # plot logit simulations
23 doc/figs/nagler-fd-bias.pdf: R/plot-nagler-fd-bias.R data/nagler-fd-bias.rds
24   Rscript $<
25
26 # do ge example
27 doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf: R/ge.R data/ge.csv
28   Rscript R/ge.R
29
30 # compile manuscript
31 doc/unnecessary.pdf: doc/unnecessary.tex doc/bibliography.bib doc/figs/intuition-1.pdf doc/figs/intuition-2
32   # cd into doc so that pdflatex runs in the doc directory
33   cd doc; pdflatex unnecessary
34   cd doc; bibtex unnecessary
35   cd doc; pdflatex unnecessary
36   cd doc; pdflatex unnecessary
37   cd doc; rm -f *.bbl *.log *.synctex.gz *.aux *.out *blg
38 cleanpaper:
39   rm -f doc/unnecessary.pdf
40
41 cleandag:
42   rm -f makefile-dag.png
43
44 cleanALL: cleanpaper
45   rm -f doc/figs/poisson-mcs.pdf
46   rm -f doc/figs/intuition-*.pdf
47   rm -f doc/figs/nagler-fd-bias.pdf data/nagler-fd-bias.rds data/nagler-fd-bias.csv
48   rm -f doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf
```

prerequisite

Makefile 1:1

LF UTF-8 Makefile master Fetch 2 files

Makefile — ~/Dropbox/projects/unnecessary

Project

unnecessary

- old
- .git
- .Rproj.user
- data
- doc
- memo
- present
- R
- .DS_Store
- .gitignore
- .Rapp.history
- .Rhistory
- ~.bash_profile

Makefile

```
1 # phony
2 all: paper dag
3 paper: doc/unnecessary.pdf
4 dag: makefile-dag.png
5
6 # draw makefile dag
7 makefile-dag.png: makefile-dag.R Makefile
8   Rscript $<
9
10 # do and plot poisson simulations
11 doc/figs/poisson-mcs.pdf: R/poisson-mcs.R
12   Rscript $< recipe
13
14 # do intuition and plot simulations
15 doc/figs/intuition-1.pdf doc/figs/intuition-2.pdf doc/figs/intuition-3.pdf doc/figs/intuition-4.pdf doc/fig
16   Rscript $<
17
18 # do logit simulations
19 data/nagler-fd-bias.rds: R/nagler-fd-sims.R data/scobit.dta
20   Rscript $<
21
22 # plot logit simulations
23 doc/figs/nagler-fd-bias.pdf: R/plot-nagler-fd-bias.R data/nagler-fd-bias.rds
24   Rscript $<
25
26 # do ge example
27 doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf: R/ge.R data/ge.csv
28   Rscript R/ge.R
29
30 # compile manuscript
31 doc/unnecessary.pdf: doc/unnecessary.tex doc/bibliography.bib doc/figs/intuition-1.pdf doc/figs/intuition-2
32   # cd into doc so that pdflatex runs in the doc directory
33   cd doc; pdflatex unnecessary
34   cd doc; bibtex unnecessary
35   cd doc; pdflatex unnecessary
36   cd doc; pdflatex unnecessary
37   cd doc; rm -f *.bbl *.log *.synctex.gz *.aux *.out *blg
38 cleanpaper:
39   rm -f doc/unnecessary.pdf
40
41 cleandag:
42   rm -f makefile-dag.png
43
44 cleanALL: cleanpaper
45   rm -f doc/figs/poisson-mcs.pdf
46   rm -f doc/figs/intuition-*.pdf
47   rm -f doc/figs/nagler-fd-bias.pdf data/nagler-fd-bias.rds data/nagler-fd-bias.csv
48   rm -f doc/figs/ge-pr.pdf doc/figs/ge-fd.pdf
```

Makefile 1:1

LF UTF-8 Makefile master Fetch 2 files

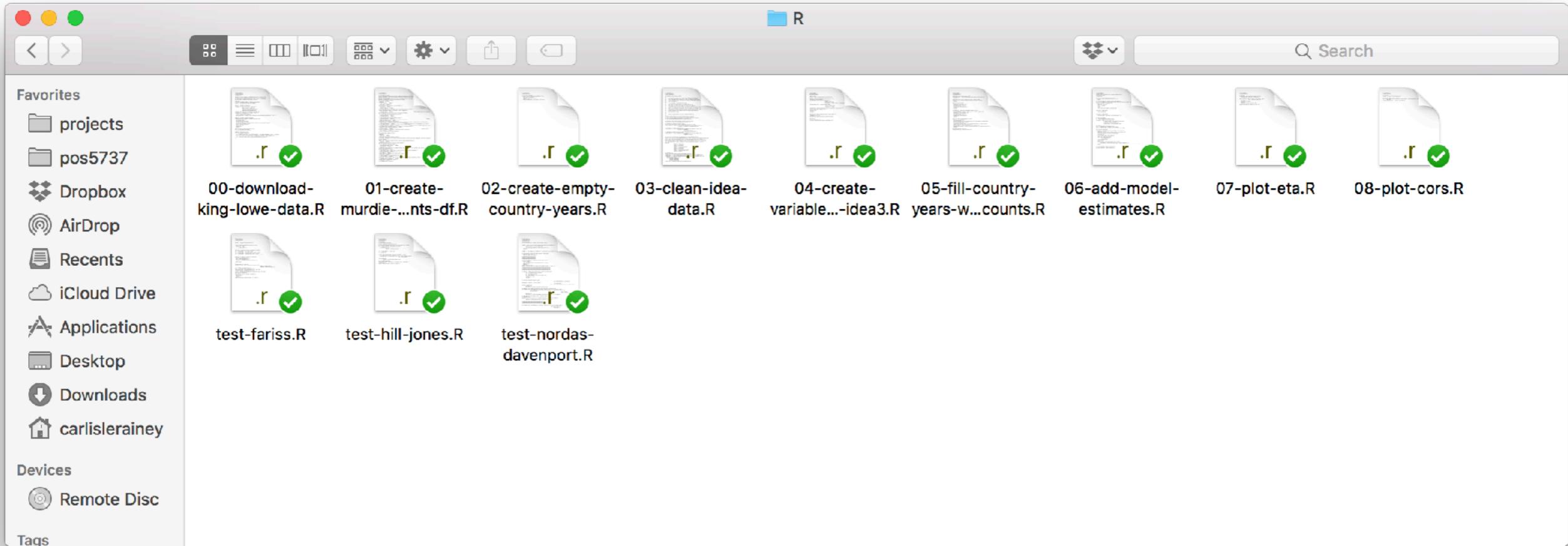
Measuring Dissent



What needs to be done?

principled
implemented
documented

1. Download the raw event data from Dataverse.
2. Tidy it into a country-year data set, with counts of total number of events and the number of dissent events.
3. Fit one model to the data and tidy the estimates into a country-year data set.
4. Do three evaluations.
5. Write the paper.



- 12 R scripts
- a stan model
- 13 raw data sets
- 4 cleaned data sets
- 6 figures

**The dependencies
are complex!**



latent-dissent — -bash — 80x24

```
Last login: Tue Aug  7 16:12:11 on ttys000
latent-dissent > make
```

```
git clone https://github.com/carlislerrainey/latent-dissent.git  
make cleanALL  
make
```

Further Info

- Broman and Woo. 2018. “Data Organization in Spreadsheets.” https://github.com/kbroman/Paper_DataOrg.
- Healy. 2018. *The Plain Person’s Guide to Plain Text Social Science.* <https://kieranhealy.org/publications/plain-person-text/>
- Bryan. 2018. “Excuse Me, Do You Have a Moment to Talk about Version Control?” <https://peerj.com/preprints/3159/>
- Wilson et al. 2017. “Good Enough Practices in Scientific Computing.” <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005510>
- Others at <https://pos5737.github.io/resources.html>.