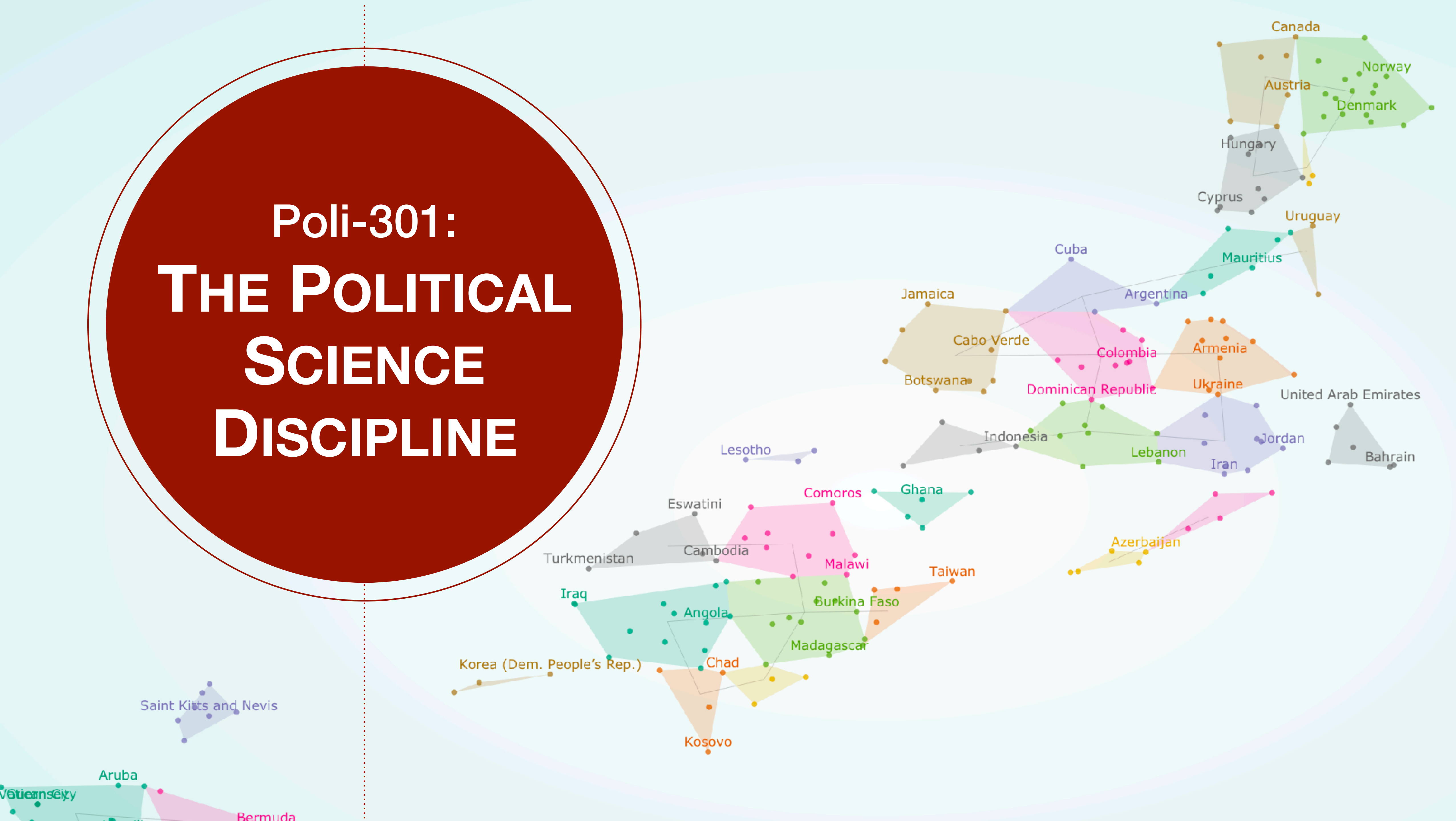


Poli-301: THE POLITICAL SCIENCE DISCIPLINE



TODAY'S AGENDA

1

Gathering data to make it tidy

2

Factors for categories

3

Setting up projects, getting data into R

Naming objects

There are only two hard things in Computer Science: cache invalidation and naming things.

-- Phil Karlton

Two most common ways of using factors

By hand

```
factor(variable, levels = c("first", "second", ...))
```

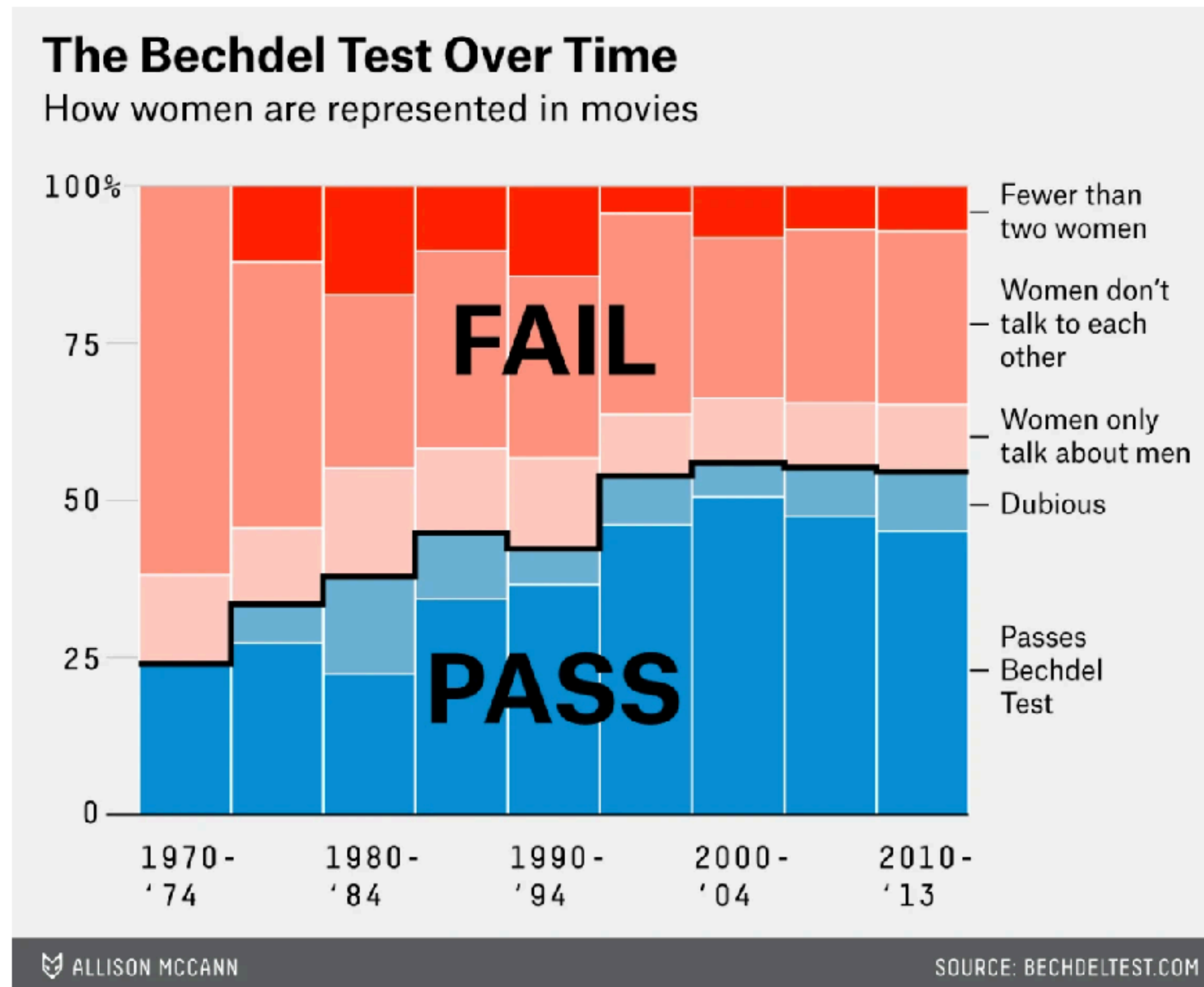
Order by frequency of value

```
fct_infreq(variable)
```

Order by another variable

```
fct_reorder(variable, 2nd variable)
```

Bechdel Test



Practice

Getting data into R

Up to this point we've used data stored in package

Or I've written code to load external data

How do we get data we've downloaded into R?

File paths

Everything on your computer has a location

The location has an “address” called a file path

You can find the path for any file via
right-click, info

```
/Users/JuamnTellez/Dropbox/poli-301/lectures/07-projects.key
```


File extensions

How you read data into R depends on the extension

.csv (comma separated values)

.dta (STATA, a computer program economists use)

.xls or .xlsx (Excel files)

Functions for importing data

.csv

.dta

.xlsx or xls

read_csv()

read_dta()

read_excel()

tidyverse

haven

Readxl

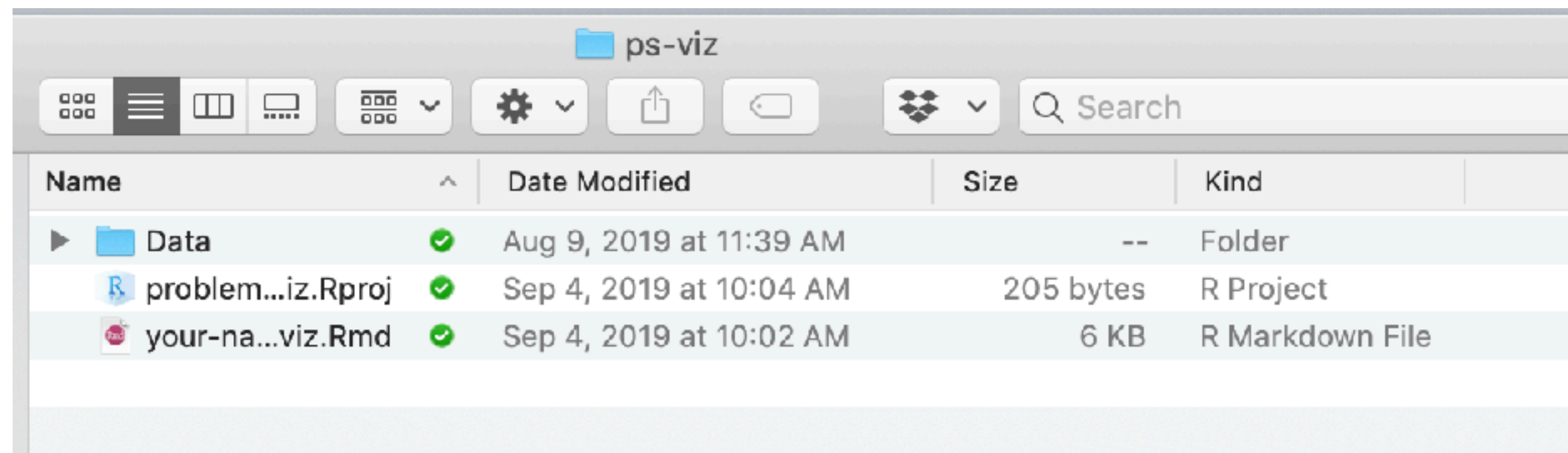
Importing using RStudio

Manually importing data

Projects and working directories

Most times we use data that lives in our computer

How does R know where it is?



```
# read data
movies = read_csv("Data/movies.csv")

# clean up movie data
movies_clean =
  movies %>%
  # exclude movies made before 1980
  filter(year >= 1980) %>%
  # create a column called profit = gross - budget
  mutate(profit = gross - budget) %>%
  # download by million to get profit in millions
  mutate(profit = profit/1000000)
movies_clean
...
```

Projects and working directories

Rstudio projects help keep you organized

Keep projects in one folder

They establish a *working directory* (a point of reference) for your file paths

/Users/JuamnTellez/Dropbox/poli-301/problem-sets/answers/ps1/Data/movies.csv

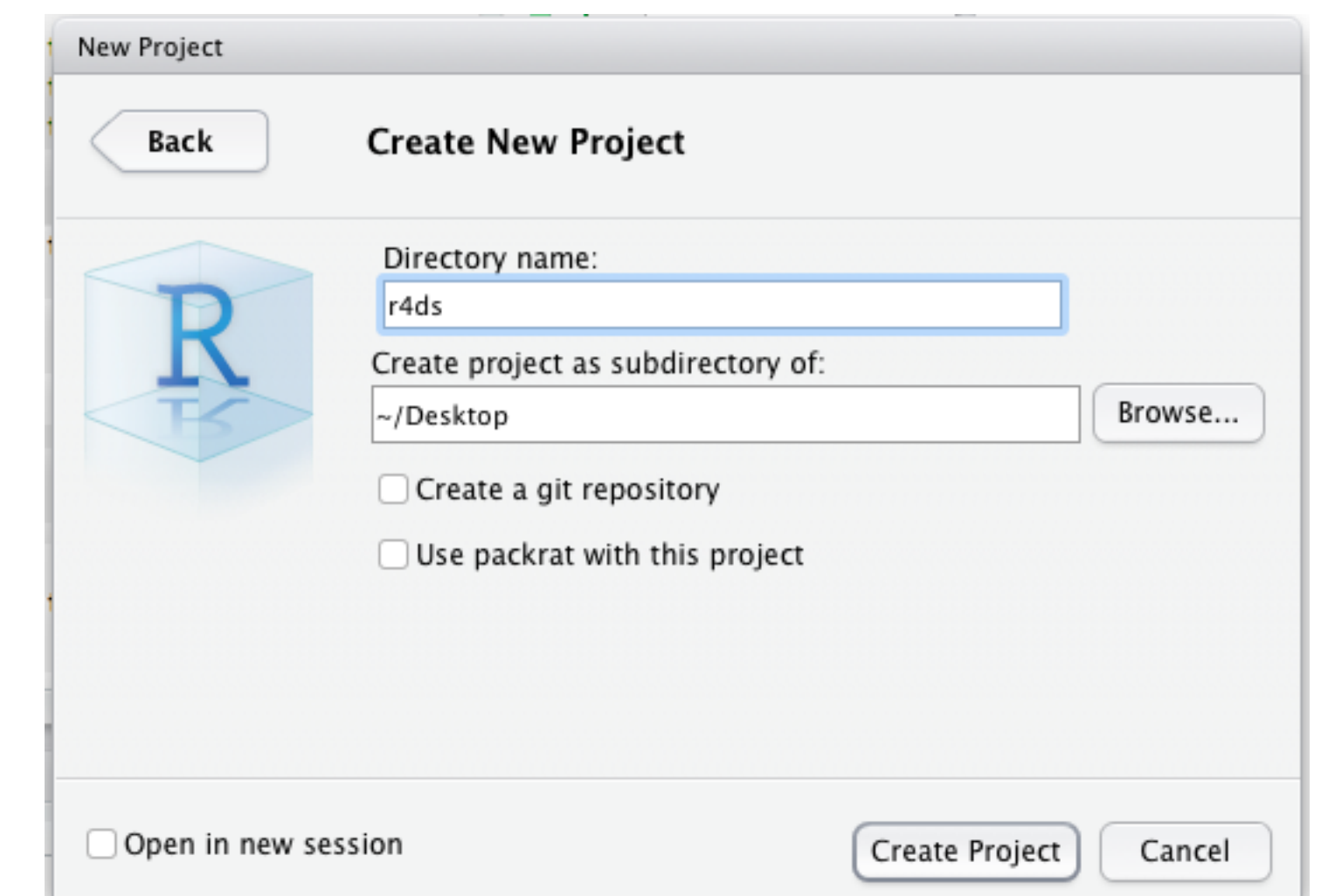
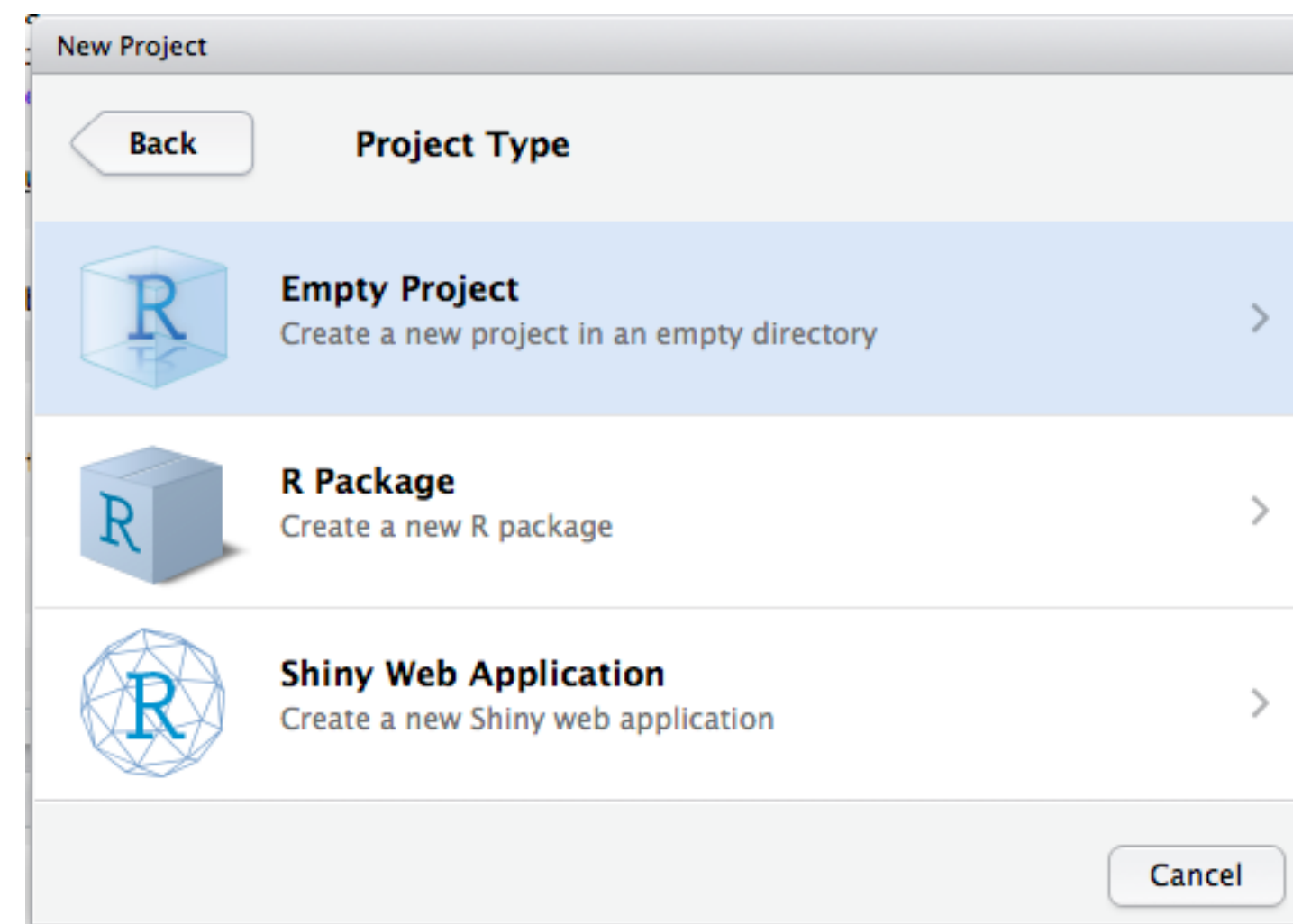
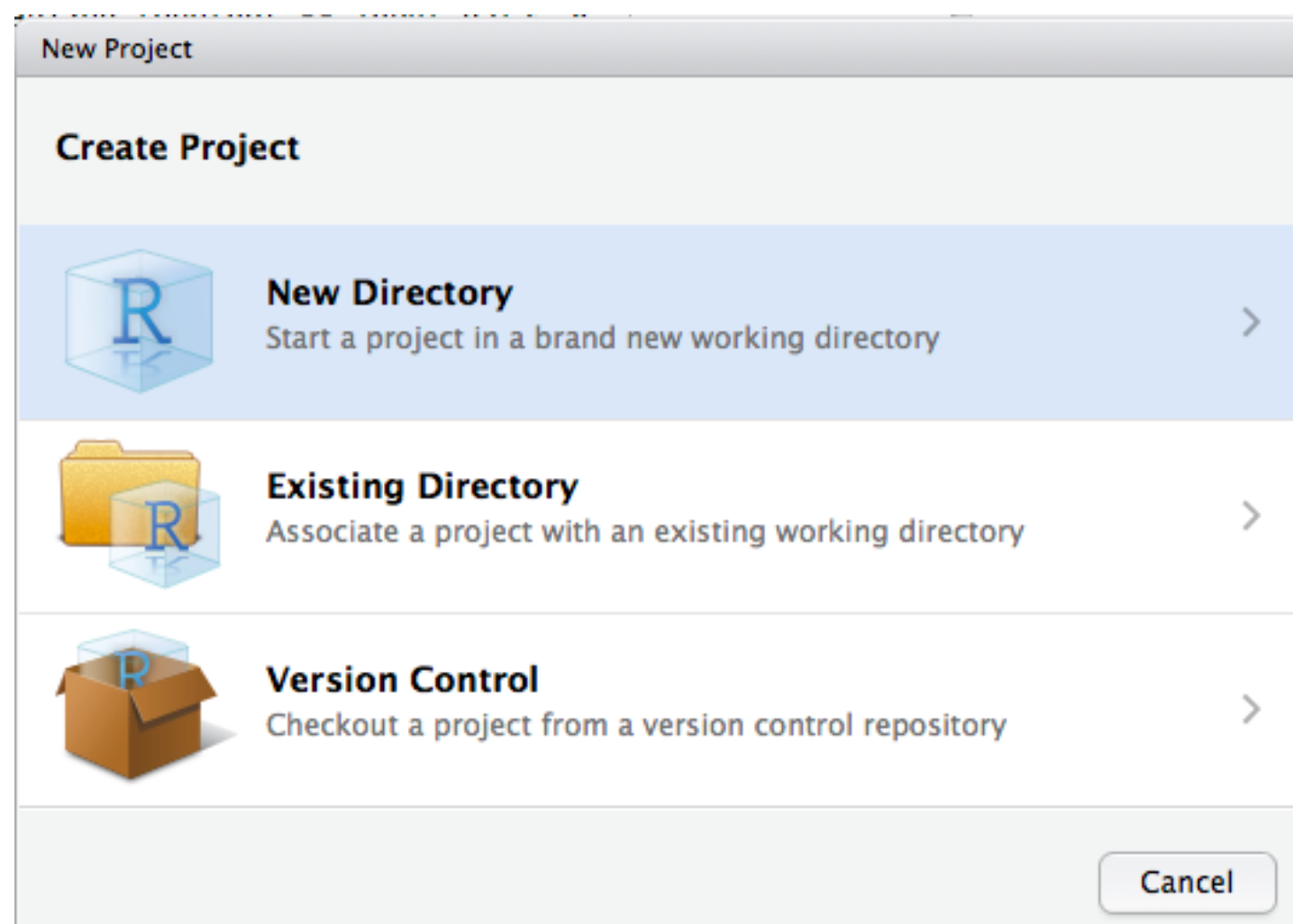
VS

Data/movies.csv

Workflow

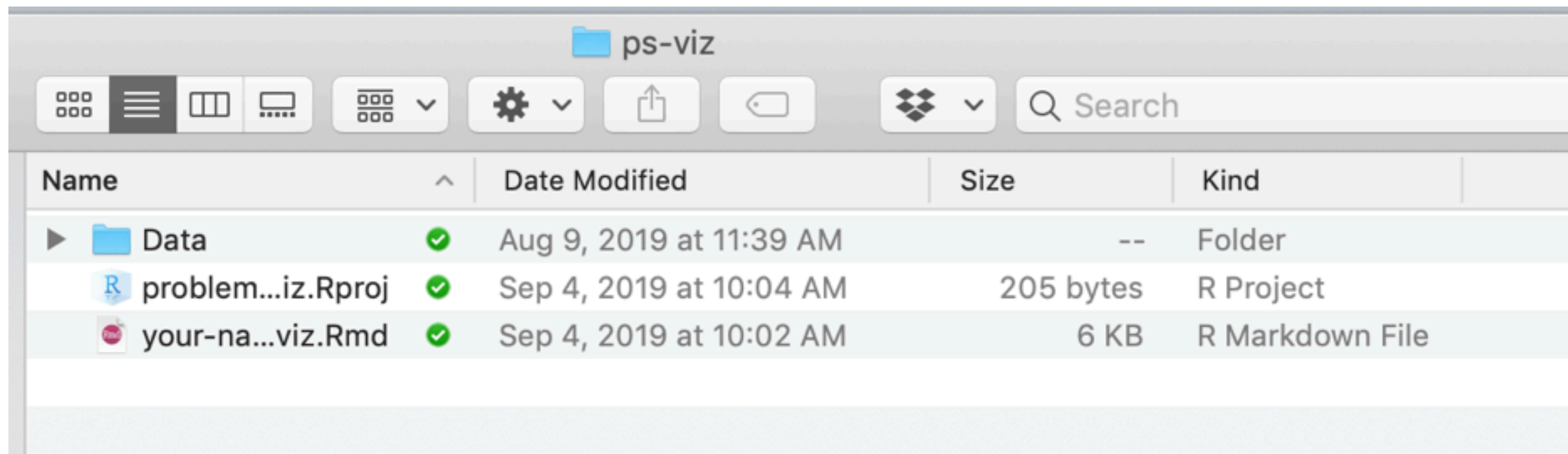
Make new project for each homework

Put it somewhere you can find!



Workflow

Make a folder for data



Double-click on .Rproj file to start coding