R PROGRAMMING

for environmental health research

Brooke Anderson
Colorado State University

April 8, 2019

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'packa
##
##     filter, lag

## The following objects are masked from 'packa
##
##     intersect, setdiff, setequal, union
```

Today's goals

LEARN principles of a key R plotting framework

UNDERSTAND what R can do for visualizations

KNOW what to do next to learn more

Homework?!

https://bit.ly/2WQV6XT

Today's plan

ORGANIZE

TRACK

PACKAGE

COLLECT

PROCESS

PREREQUISITES

Setting up

Install RStudio Desktop

https://www.rstudio.com/

Install git

https://git-scm.com/downloads

Create GitHub account

https://github.com/

Download example project

https://github.com/geanders/columbia_env_health_examples

ORGANIZE

RStudio's **R Projects** for organizing

Common project subdirectories

data-raw Raw data and R scripts to clean the raw data.

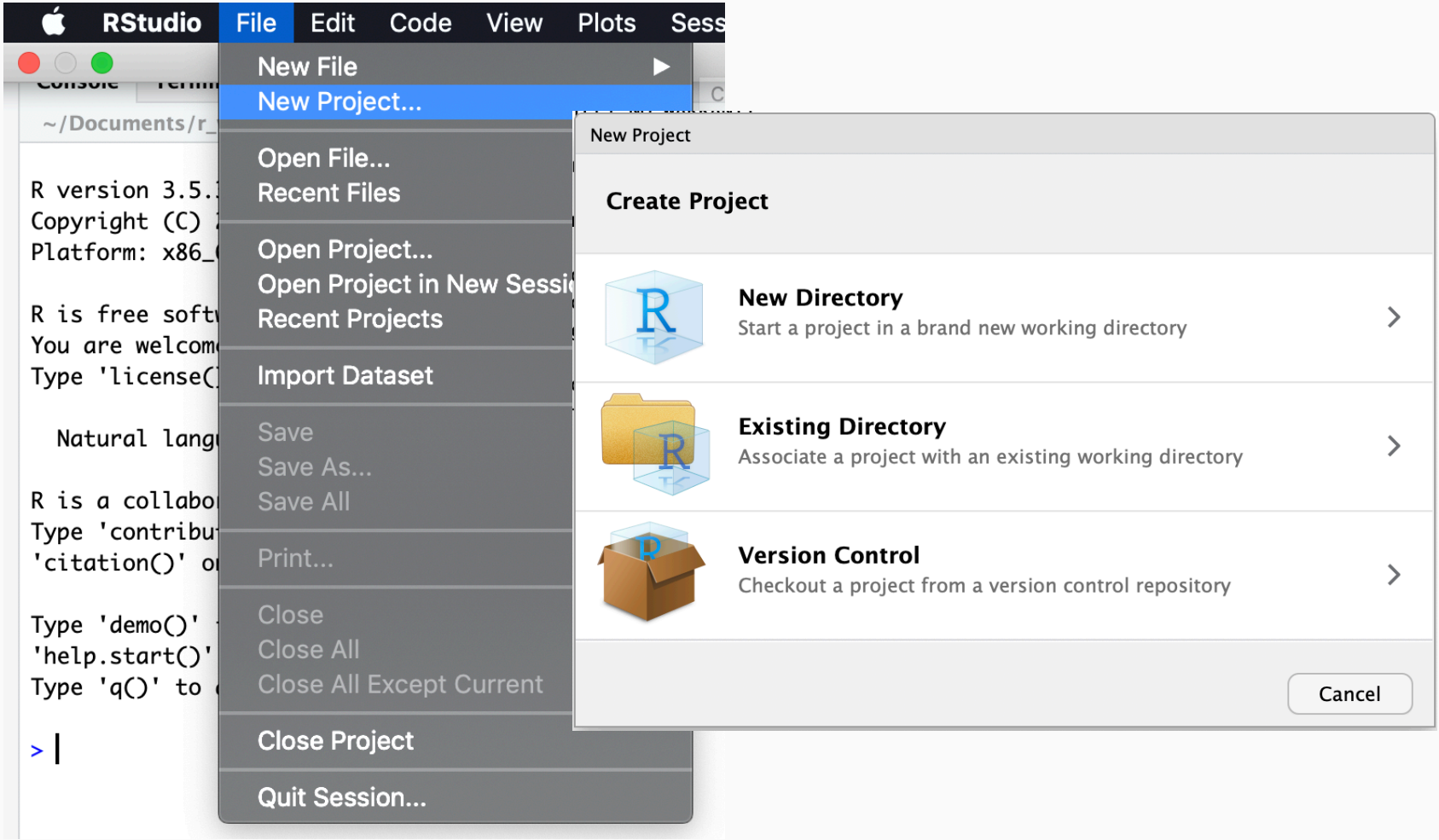
data Cleaned data, often saved as `.RData` after being generated by a script in `data-raw`.

R Code for any functions used in analysis.

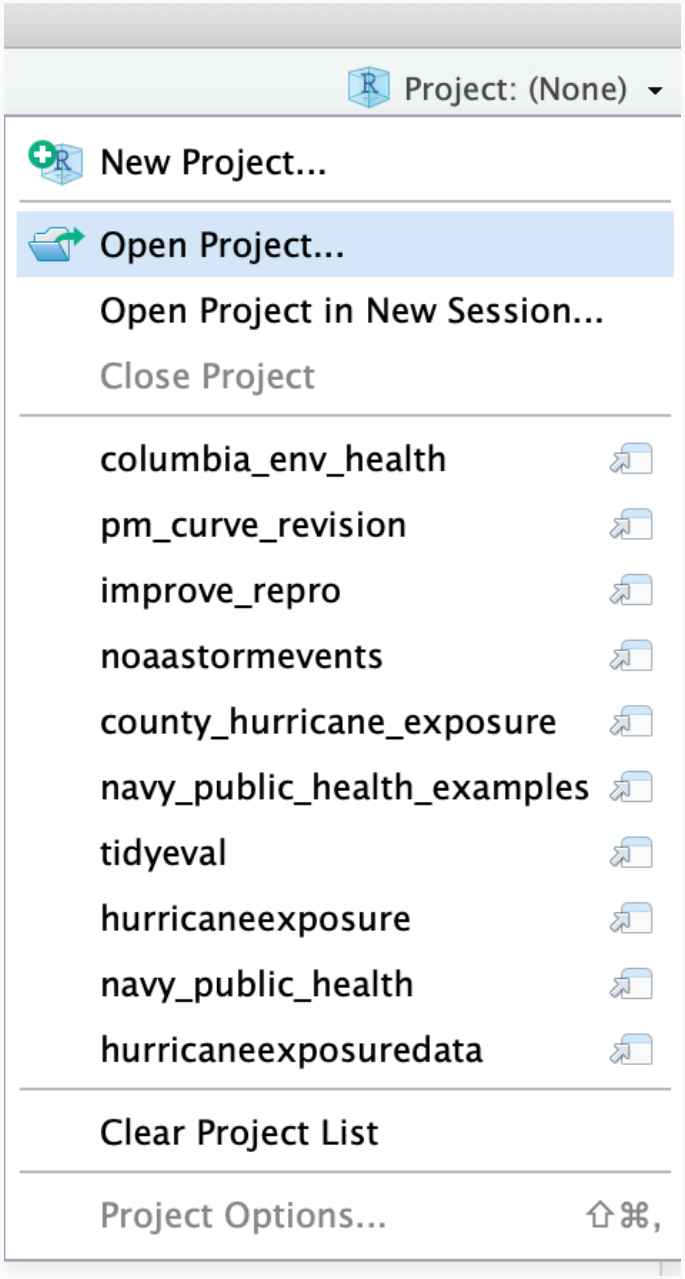
figures Figures created from R code.

reports R Markdown files and products rendered from those files (e.g., paper drafts, presentations).

Create R project



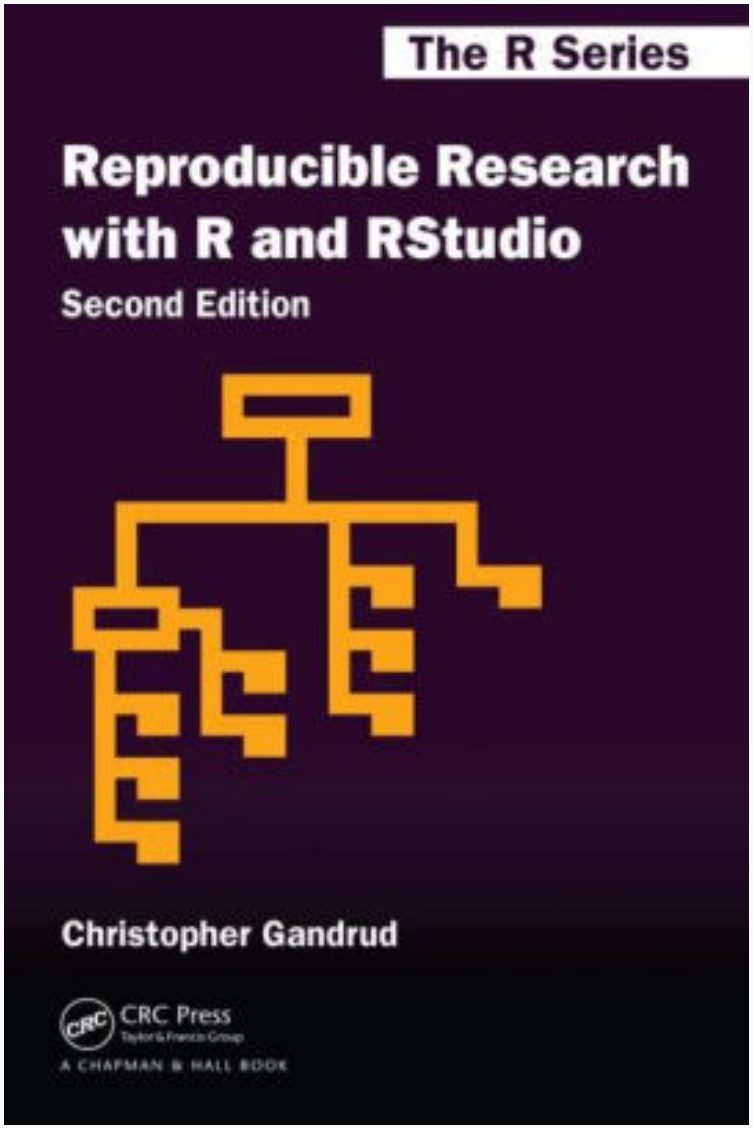
Navigating R Projects



```
Georgianas-MacBook-Pro:columbia_env_health georgianaanderson$ ls -a
.                               _build.sh
..                              _output.yml
.DS_Store                       _workshop_slides
.Rproj.user                     book.bib
.git                            columbia_env_health.Rproj
.gitignore                      columbia_env_health.log
.nojekyll                       data
01-organize.Rmd                 docs
02-track.Rmd                    flexdashboard
03-package.Rmd                  images
04-collect.Rmd                  index.Rmd
05-process.Rmd                  irma_fatalities.pdf
06-summary.Rmd                  now.json
DESCRIPTION                     old_data
LICENSE                         packages.bib
R                               preamble.tex
README.md                       skeleton.bib
_bookdown.yml                   style.css
_bookdown_files                 toc.css
```

.Rproj/

Resources


The R Series
Reproducible Research with R and RStudio
Second Edition
Christopher Gandrud
CRC Press
A CHAPMAN & HALL BOOK

[Live coding example]

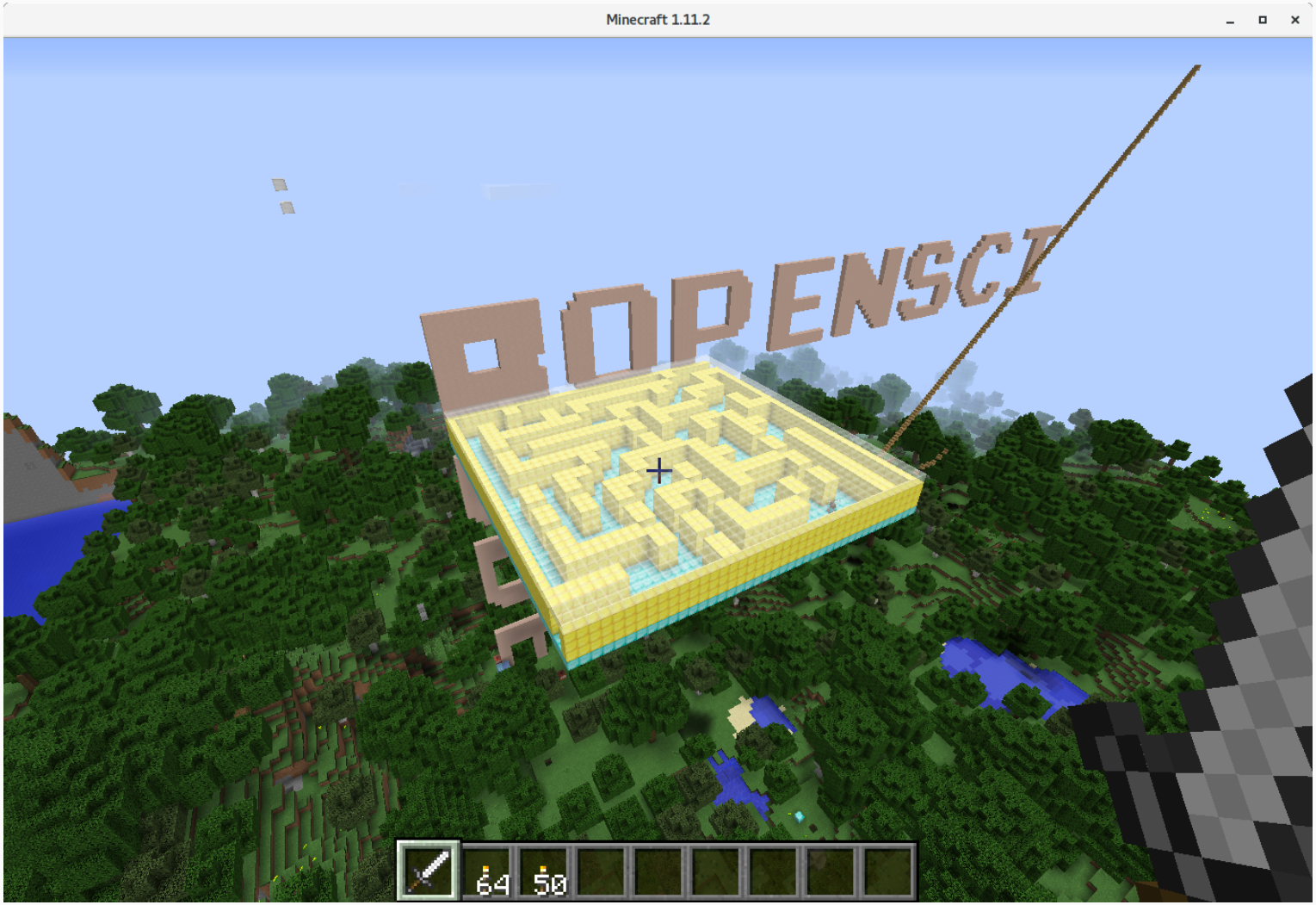
TRACK

git and GitHub for version control

```
Georgianas-MacBook-Pro:columbia_env_health georgianaanderson$ ls -a
.                               _build.sh
..                              _output.yml
.DS_Store                       _workshop_slides
.Rproj.user                     book.bib
.git                            columbia_env_health.Rproj
.gitignore                      columbia_env_health.log
.nojekyll                       cot
01-organize.Rmd                 docs
02-track.Rmd                    flexdashboard
03-package.Rmd                  images
04-collect.Rmd                  index.Rmd
05-process.Rmd                  irma_fatalities.pdf
06-summary.Rmd                  now.json
DESCRIPTION                     old_data
LICENSE                         packages.bib
R                               preamble.tex
README.md                       skeleton.bib
_bookdown.yml                   style.css
_bookdown_files                 toc.css
```

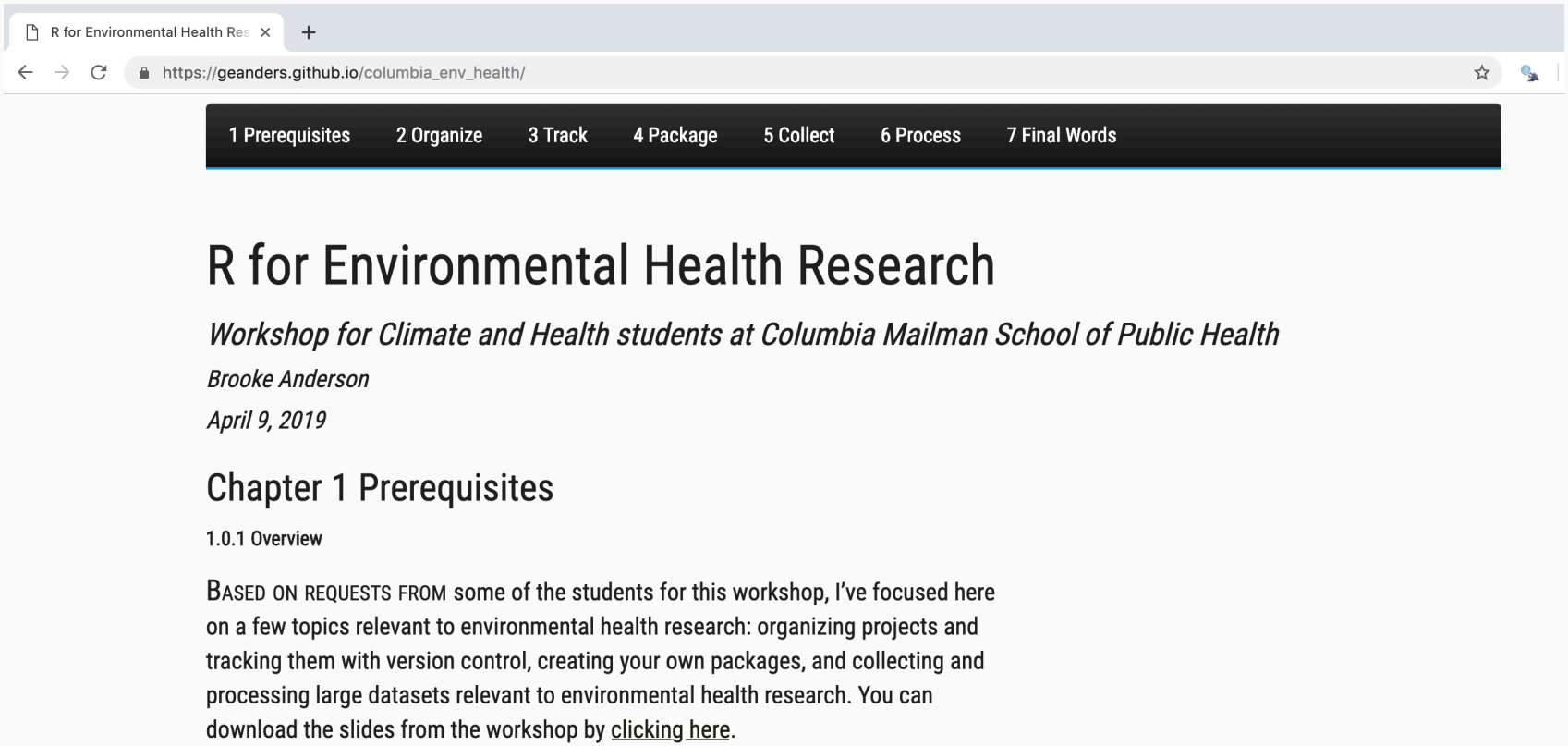
.git/

Using GitHub to collaborate



https://github.com/ropenscilabs/miner

Hosting content with GitHub Pages



[Live coding example]

PACKAGE

Collect R functions in **packages**

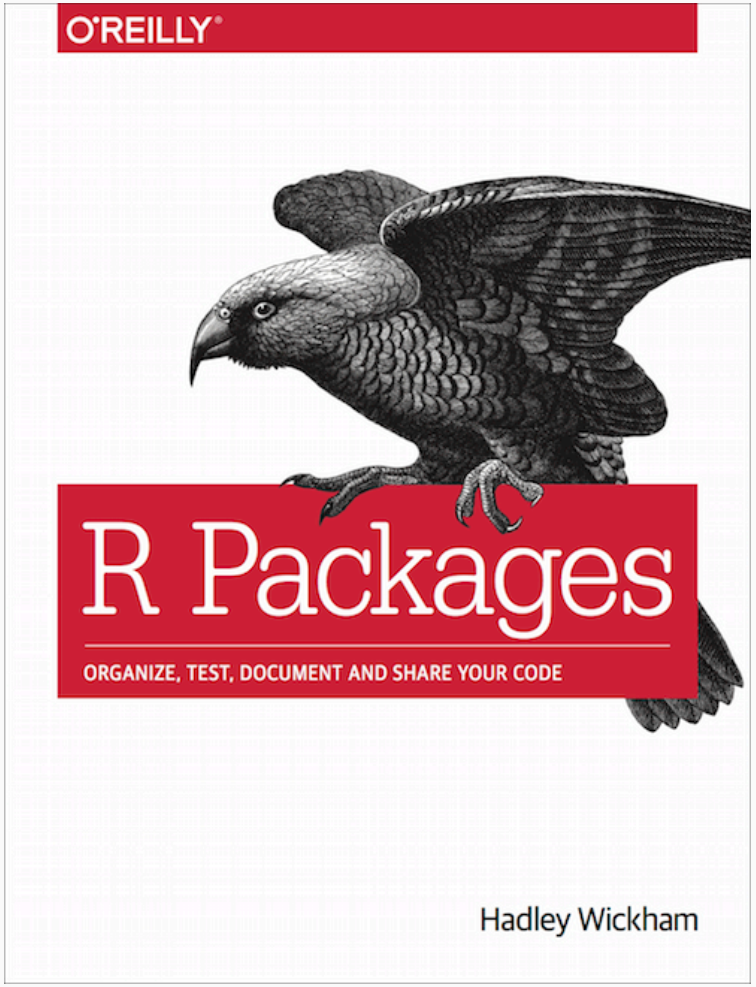
Dirk Eddelbuettel @eddelbuettel · 27 Jan 2017

Big congratulations to @gbwanderson whose new package 'hurricaneexposure' just became package 10,000 on CRAN !!

> **CRAN Package Updates** @CRANberriesFeed
> 9999 packages on CRAN right now, so imagine dozens of R nerds hanging in suspense waiting for the package to make it 10k ...

2 35 93

Resources



http://r-pkgs.had.co.nz/

[Live coding example]

COLLECT

Leverage **open data** tools for collecting data

Data packages

```
library(hurricaneexposure)
county_wind(counties = "36061",
            start_year = 1988, end_year = 2015,
            wind_limit = 17.5) %>%
  select(storm_id, vmax_sust, storm_dist, closest_date)
```

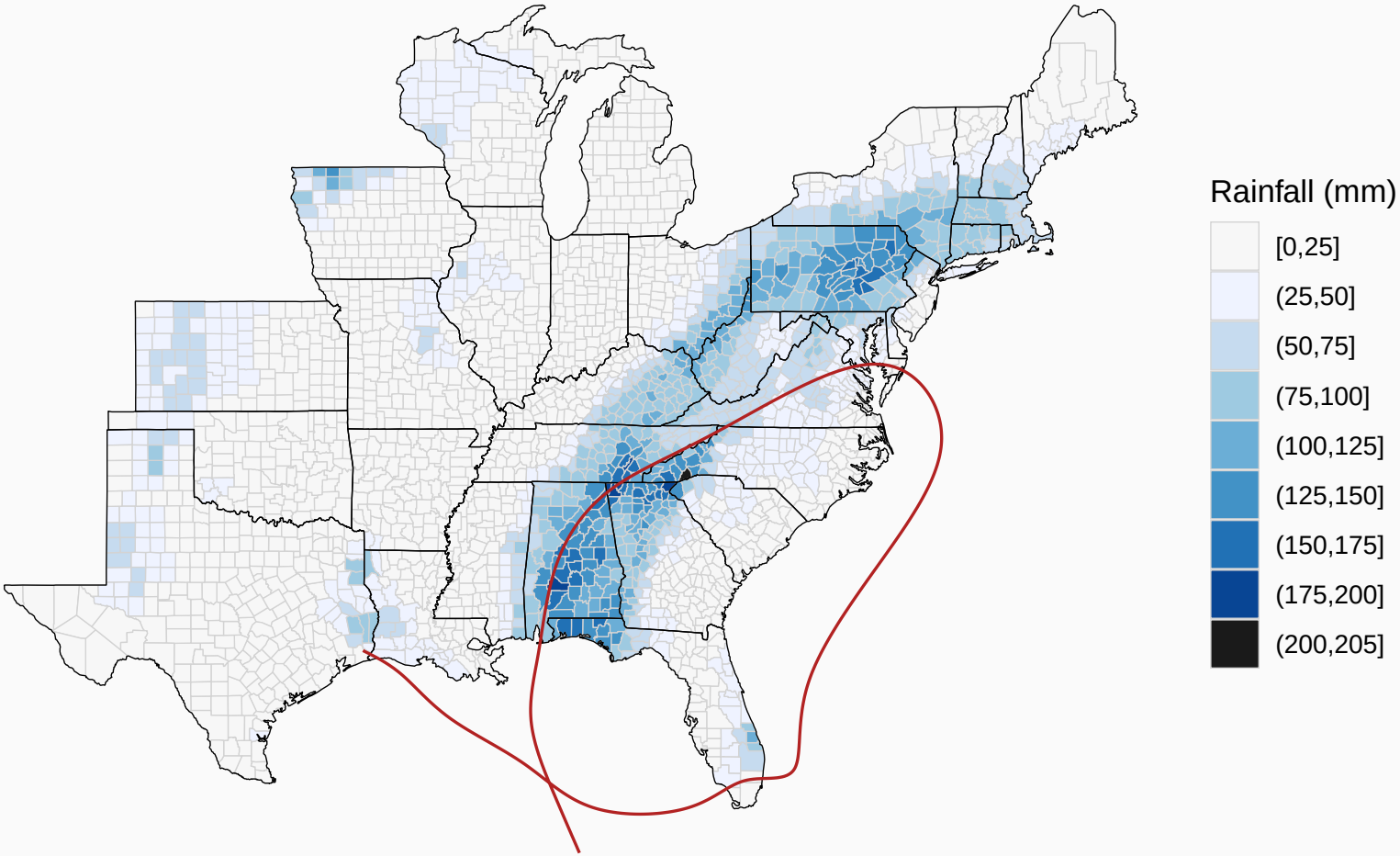
```
##      storm_id vmax_sust storm_dist closest_date
## 1    Bob-1991  18.19559 161.571830   1991-08-19
## 2 Bertha-1996  28.95496  16.966013   1996-07-13
## 3  Floyd-1999  20.50178  45.408483   1999-09-16
## 4  Hanna-2008  19.25390  29.916672   2008-09-06
## 5  Irene-2011  25.68553   5.796733   2011-08-28
## 6  Sandy-2012  21.99213 158.040788   2012-10-29
```

Data packages

```
county_events(counties = "36061",
              start_year = 1988, end_year = 2015,
              event_type = "flood") %>%
  select(storm_id, storm_dist, closest_date)
```

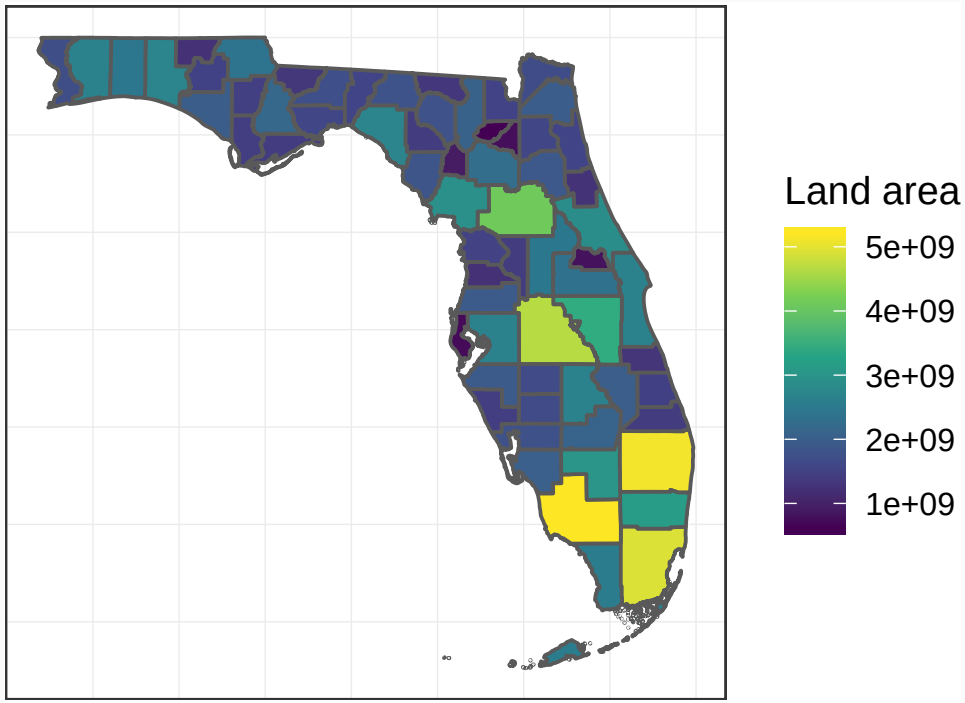
```
##       storm_id storm_dist closest_date
## 1   Floyd-1999  45.408483   1999-09-16
## 2 Allison-2001 158.909890   2001-06-17
## 3 Frances-2004 379.343696   2004-09-09
## 4    Ivan-2004 311.346881   2004-09-18
## 5  Jeanne-2004 222.900157   2004-09-29
## 6   Beryl-2006 207.358443   2006-07-20
## 7   Barry-2007 148.251718   2007-06-04
## 8   Irene-2011   5.796733   2011-08-28
## 9  Andrea-2013  92.381282   2013-06-08
```

```
map_counties(storm = "Ivan-2004", metric = "rainfall")
```



Open Data APIs

```
library(tigris)
fl_counties <- counties(state = "FL",
                        class = "sf")
```



[Live coding example]

PROCESS

Find and make **R packages** for processing data

[Live coding example]

```
irma_week_accs <- fl_accidents %>%
  group_by(fips) %>%
  summarize(fatals = sum(fatals))

irma_accs <- fl_counties %>%
  full_join(irma_week_accs, by = c("GEOID" = "fips")) %>%
  mutate(fatals = ifelse(is.na(fatals), 0, fatals))

fl_accidents <- fl_accidents %>%
  st_as_sf(coords = c("longitud", "latitude")) %>%
  st_set_crs(st_crs(

irma_track <- st_read("data/al112017_best_track",
                      layer = "al112017_lin") %>%
  st_transform(crs = st_crs(irma_accs))
```

st_read(dsn, layer, ...)

Homework!!

https://bit.ly/2WQV6XT