

R PROGRAMMING

**for environmental health
research**

Brooke Anderson

Colorado State University

April 8, 2019

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](https://github.com/geanders/columbia_env_health_examples)

ORGANIZE

Setting up

One project : One directory

Rule #1 of research project file organization

Use consistent names

Rule #2 of research project file organization

Use relative filenames

Rule #3 of research project file organization

Common project subdirectories

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

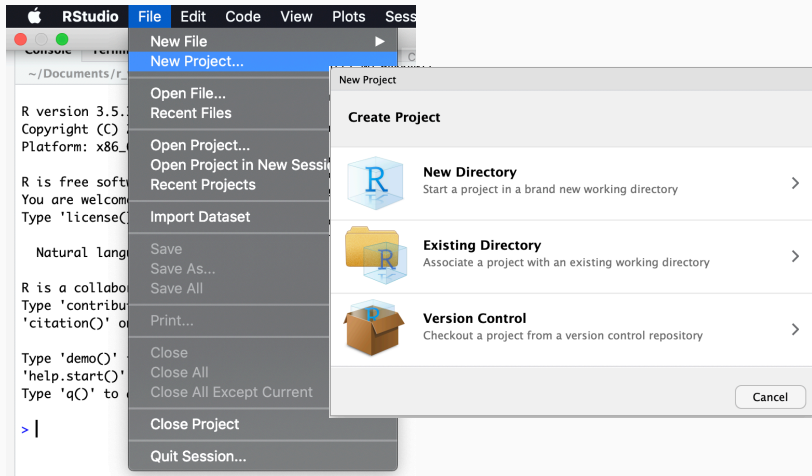
data Cleaned data, often saved as `.RData` after being generated by a script in `data-raw`.

R Code for any functions used in analysis.

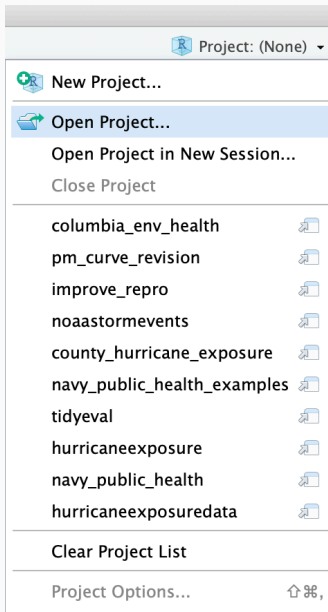
figures Figures created from R code.

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

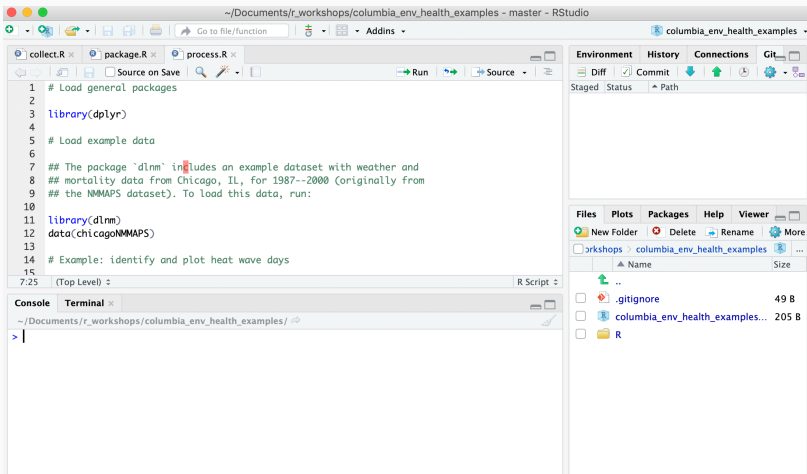
Create R project



Navigating R Projects



Navigating R Projects



The screenshot displays the RStudio environment with the following components:

- Title Bar:** Shows the file path `~/Documents/r_workshops/columbia_env_health_examples - master - RStudio`.
- Toolbar:** Includes icons for file operations and a "Go to file/function" search bar.
- Source Editor:** Contains an R script with the following code:

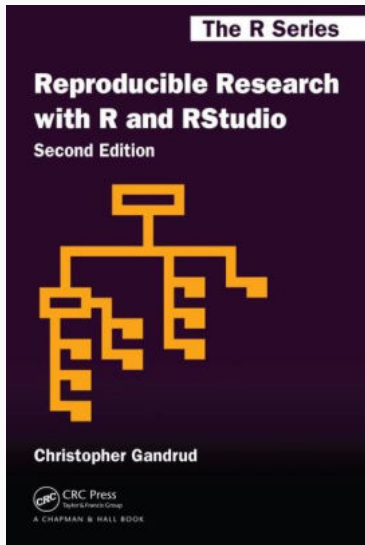
```
1 # Load general packages
2
3 library(dplyr)
4
5 # Load example data
6
7 ## The package 'dlnm' includes an example dataset with weather and
8 ## mortality data from Chicago, IL, for 1987--2000 (originally from
9 ## the NMMAPS dataset). To load this data, run:
10
11 library(dlnm)
12 data(chicagoNMMAPS)
13
14 # Example: identify and plot heat wave days
15
```
- Console:** Shows the current directory `~/Documents/r_workshops/columbia_env_health_examples/` and a prompt `> |`.
- Environment Panel:** Displays tabs for "Diff", "Commit", "Connections", and "Git". The "Staged" tab is active, showing a list of files with their sizes:

Name	Size
<code>..</code>	
<code>.gitignore</code>	49 B
<code>columbia_env_health_examples...</code>	205 B
<code>R</code>	

```
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  flexdashboard
02-track.Rmd     images
03-package.Rmd   index.Rmd
04-collect.Rmd   irma_fatalities.pdf
05-process.Rmd   now.json
06-summary.Rmd   old_data
DESCRIPTION      packages.bib
LICENSE          preamble.tex
R                skeleton.bib
README.md        style.css
_bookdown.yml    toc.css
_bookdown_files
```

.Rproj/

Resources



```
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, fatalities))
```

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitud", "latitude")) %>%  
  st_set_crs(st_crs(st_read(dsn, layer, ...)))
```

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

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        columbia_env_health.Rproj
01-organize.Rmd  flexdashboard
02-track.Rmd     images
03-package.Rmd  index.Rmd
04-collect.Rmd  irma_fatalities.pdf
05-process.Rmd  now.json
06-summary.Rmd  old_data
DESCRIPTION      packages.bib
LICENSE          preamble.tex
R                skeleton.bib
README.md        style.css
_bookdown.yml    toc.css
_bookdown_files
```

.git/

Using GitHub to collaborate



<https://github.com/ropenscilabs/miner>

Hosting content with GitHub Pages



R for Environmental Health Research

Workshop for Climate and Health students at Columbia Mailman School of Public Health

Brooke Anderson

April 9, 2019

Chapter 1 Prerequisites

1.0.1 Overview

BASED ON REQUESTS FROM some of the students for this workshop, I've focused here on a few topics relevant to environmental health research: organizing projects and tracking them with version control, creating your own packages, and collecting and processing large datasets relevant to environmental health research. You can download the slides from the workshop by [clicking here](#).


```
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, fatalities))
```

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitud", "latitude")) %>%  
  st_set_crs(st_crs(st_read(dsn, layer, ...)))
```

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

PACKAGE

Collect R functions in **packages**

Why write R packages

Software development in biostatistics

So I have a new policy when evaluating CV's of candidates for jobs, or when I'm reading a paper as a referee. If the paper is about a new statistical method or machine learning algorithm and there is no software available for that method - I simply mentally cross it off the CV. If I'm reading a data analysis and there isn't code that reproduces their analysis - I mentally cross it off. In my mind, new methods/analyses without software are just [vapor ware](#). Now, you'd definitely have to cross a few papers off my CV, based on this principle. I do that. But I'm trying really hard going forward to make sure nothing gets crossed off.

Source: Jeff Leek, Simply Statistics

Why write R packages
























Research impacts of NMMAPS package (*Source: Barnett, Huang, and Turner, "Benefits of Publicly Available Data", Epidemiology 2012*):

As of November 2011, 67 publications had been published using this data, with 1,781 citations to these papers

Research using NMMAPS has been used by the US EPA in creating regulatory impact statements for air pollution (particulates and ozone)

"Thanks to NMMAPS, there is probably no other country in the world with a greater understanding of the health effects of air pollution and heat waves in its population."

What an R package looks like

Folders	Documents	Developer
 weathermetrics ▶	 cran-comments.md	 data.R
PDF Documents	 NEWS.md	 heat_index.R
 weathermetrics.pdf	 README.md	 moisture_conversions.R
Other	Folders	 rainmeasure_conversion.R
 weathermetrics_1.2.0.tar.gz	 data ▶	 temperature_conversions.R
 weathermetrics_1.2.2.tar.gz	 inst ▶	 weathermetrics.R
	 man ▶	 wind_conversions.R
	 R ▶	
	 vignettes ▶	
	Other	
	 DESCRIPTION	
	 NAMESPACE	
	 README.Rmd	
	 weathermetrics.Rproj	

R package template

The screenshot displays the RStudio interface for a new R package project named 'convertr'. The main editor window shows the 'hello.R' file with the following content:

```
1 #
2 #
3 # You can learn more about package authoring with RStudio at:
4 #
5 # http://r-pkgs.had.co.nz/
6 #
7 # Some useful keyboard shortcuts for package authoring:
8 #
9 # Build and Reload Package: 'Cmd + Shift + B'
10 # Check Package: 'Cmd + Shift + E'
11 # Test Package: 'Cmd + Shift + T'
12 #
13 hello <- function() {
14   print("Hello, world!")
15 }
16
17
18
19
```

The console window at the bottom shows the output of the R session:

```
~/Documents/r_workshops/convertr/
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

The right-hand pane shows the 'Environment' tab, which is currently empty. Below it, the 'Files' tab displays the project structure:

Name	Size
..	
.Rbuildignore	28 B
convertr.Rproj	356 B
DESCRIPTION	369 B
man	
NAMESPACE	31 B
R	

Required files

R/ or **data/** If you don't have one of these, your package won't do anything

DESCRIPTION Needed, but you can't keep the template version as-is

NAMESPACE Needed, but you can't keep the template version as-is

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

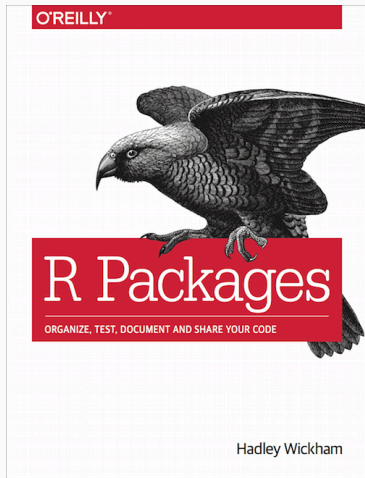
```
irma_accs <- fl_accidents %>%  
  full_join(irma_week_accs, by = c("GEOID" = "fips")) %>%  
  mutate(fatals = ifelse(is_na(fatals), 0, fatalities))
```

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitud", "latitude")) %>%  
  st_set_crs(st_crs(st_read(dsn, layer, ...)))
```

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


Resources



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

Resources



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
```

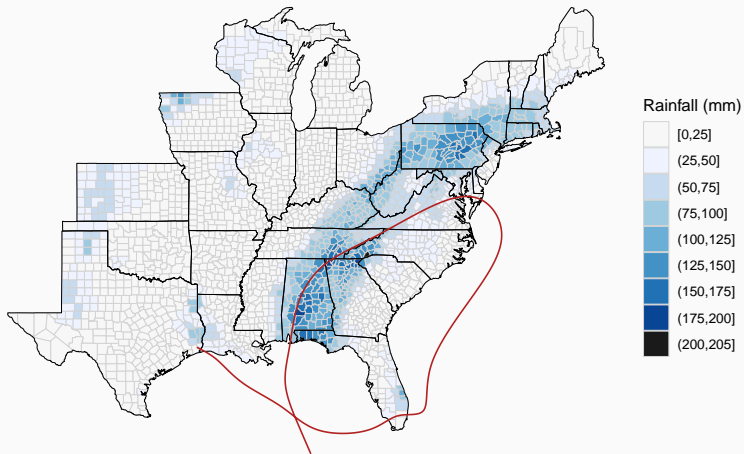
Data packages

```
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

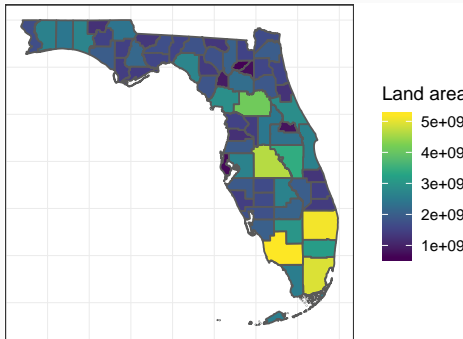
Data packages

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



Open Data APIs

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



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

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

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitud", "latitude")) %>%  
  st_set_crs(st_crs(st_read(dsn, layer, ...)))
```

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


#rstats



Dirk Eddebuettel @eddebuettel · 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



PROCESS

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

```
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, fatalities))
```

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitud", "latitude")) %>%  
  st_set_crs(st_crs(st_read(dsn, layer, ...)))
```

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

ROpenSci

Homework!!

<https://bit.ly/2WQV6XT>