

R PROGRAMMING

for data visualization

Brooke Anderson
Colorado State University

March 28, 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?!

goo.gl/7fPYUx

Today's plan

PLOT

MAP

INTERACT

REPORT

TIDY

PREREQUISITES

Setting up for success

R packages



R packages



Installing R packages

```
install.packages("readr")
```

Use the **install.packages** function to install an R package to your computer.

Loading R packages

```
library("readr")
```

Use the **library** function to load an R package that is installed on your computer.

Hello my name is

<-

Assign an object a name with R's **gets arrow**

Assignment with the gets arrow

You want to read in the “`daily_fatalities.csv`” file, which is in the “`data`” subdirectory.

Assignment with the gets arrow

Assign the filepath of this file to the R object named **fatalities_files**.

Reference that object to read in the data and assign it to the R object named **daily_fatalities**.

```
fatalities_file <- "data/daily_fatalities.csv"  
daily_fatalities <- read_csv(fatalities_file)
```

Hurricane Irma



Hurricane Irma



NWS Key West

@NWSKeyWest

Follow



THIS IS AS REAL AS IT GETS

***NOWHERE IN THE FLORIDA KEYS
WILL BE SAFE***

***YOU STILL HAVE TIME TO
EVACUATE***

Please RT. #Irma

Hurricane Irma

Navy evacuates over 5,000 personnel from Florida base ahead of Hurricane Irma

Published time: 6 Sep, 2017 05:08

Edited time: 7 Sep, 2017 10:56

[Get short URL](#)





Ratings Recalls Risky Driving Road Safety Equipment Technology & Innovation

← RESEARCH & DATA

Fatality Analysis Reporting System (FARS)

<https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>

Example data

daily_fatalities

```
## # A tibble: 28 x 4
##   date       week weekday   fatals
##   <date>     <dbl> <chr>      <dbl>
## 1 2017-08-27     35 Sunday      4
## 2 2017-08-28     35 Monday      5
## 3 2017-08-29     35 Tuesday     6
## 4 2017-08-30     35 Wednesday    6
## 5 2017-08-31     35 Thursday     6
## 6 2017-09-01     35 Friday      9
## 7 2017-09-02     35 Saturday     8
## 8 2017-09-03     36 Sunday     15
## 9 2017-09-04     36 Monday      7
## 10 2017-09-05    36 Tuesday     8
## # ... with 18 more rows
```

PLOT

R's **ggplot2** framework for plotting

ggplot2: Think layers



Plot elements

data Observations shown with the plot

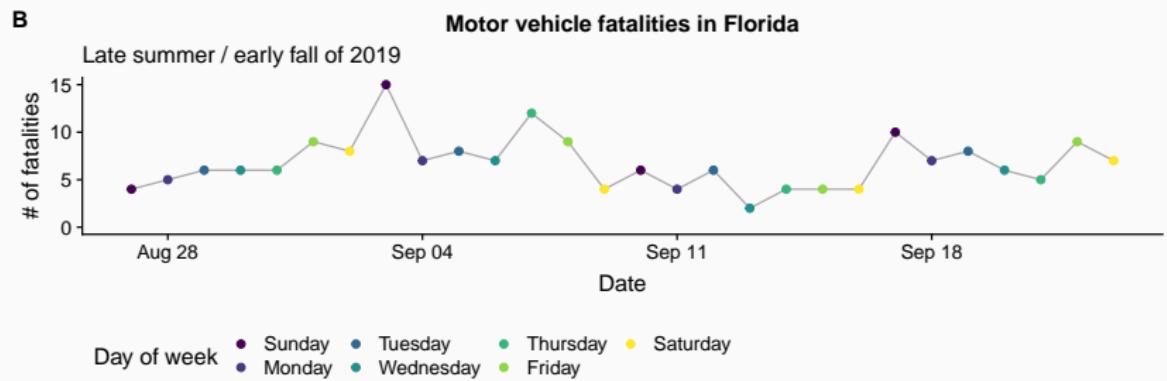
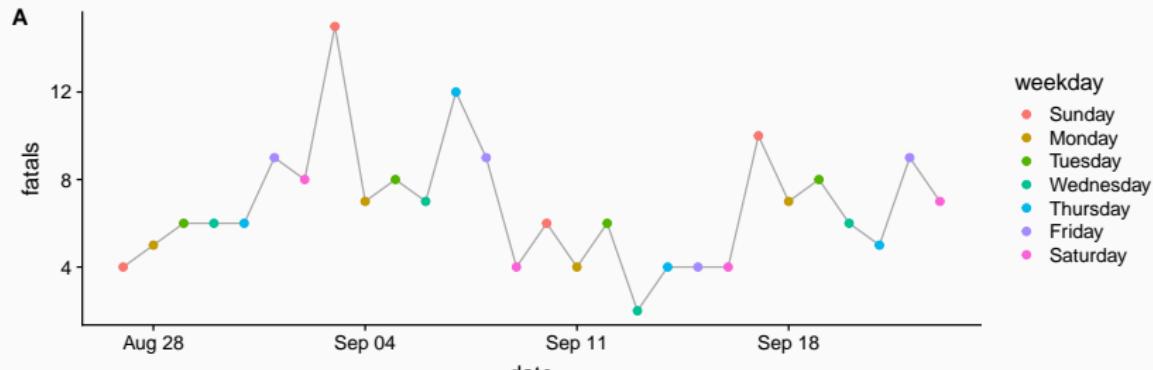
geoms Geometric objects showing the data on the plot, with aesthetics mapped to characteristics of the data or given constant values.

coordinate system Grid system defining how the data in the plot are layed out; often a Cartesian coordinate system

labels Names of axes and other scales, as well as plot titles

themes Background elements of the plot

Spot the differences



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

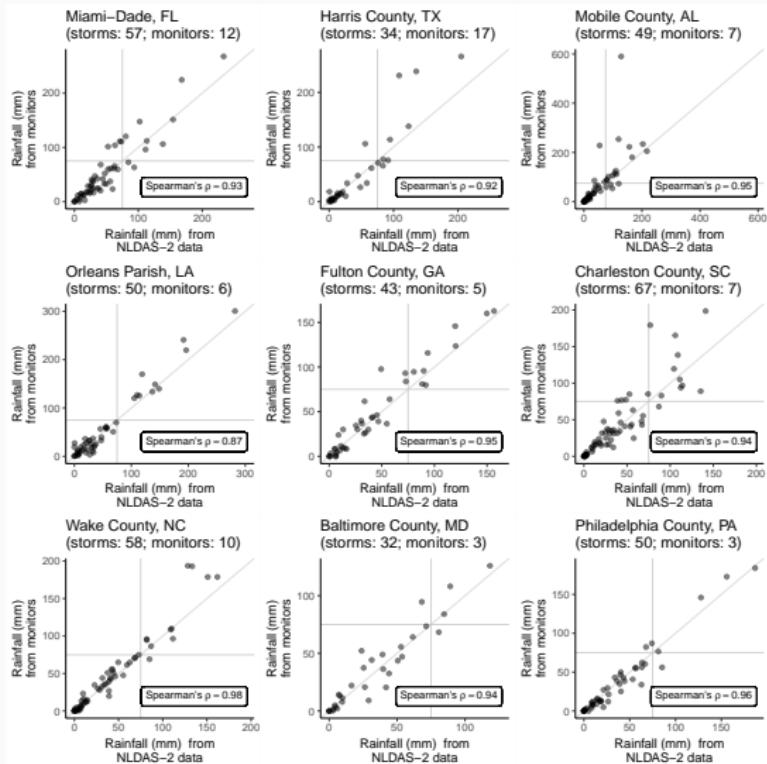
```
irma_accs <- fl_counties %>%  
  full_join(irma_week_accs, by = c("EOID" = "fips")) %>%  
  mutate(fatals = ifelse(is.na(fatals), 0, fatalities))
```

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitude", "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))
```

``Kirk Moment''



MAP

R's **sf** framework for mapping

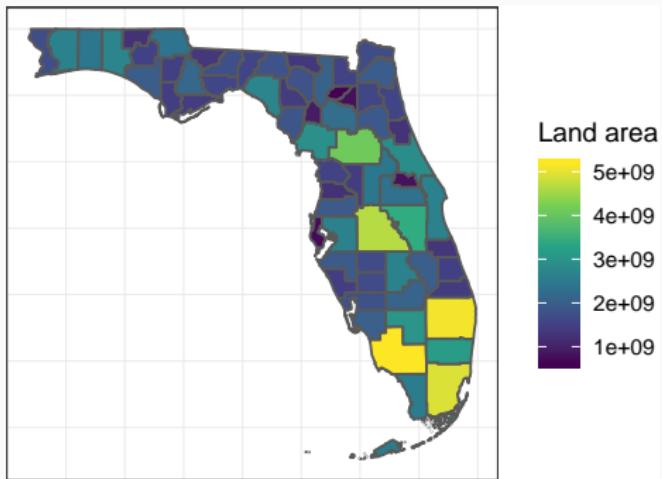
sf: simple features

sf class

```
## Simple feature collection with 37 features and 3 fields
## geometry type:  POINT
## dimension:      XY
## bbox:            xmin: -87.3797 ymin: 25.6876 xmax: -80.32332 ymax: 30.37913
## epsg (SRID):    4326
## proj4string:    +proj=longlat +datum=WGS84 +no_defs
## # A tibble: 37 x 4
##       fips date     fatalities      geometry
##   <dbl> <date>     <dbl>      <POINT [°]>
## 1 12031 2017-09-08     1 (-81.5407 30.2474)
## 2 12095 2017-09-07     1 (-81.41927 28.51796)
## 3 12097 2017-09-08    1 (-81.34221 28.26149)
## 4 12095 2017-09-07    1 (-81.18188 28.57327)
## 5 12031 2017-09-08    1 (-81.84294 30.22846)
## 6 12033 2017-09-07    2 (-87.3797 30.56341)
## 7 12023 2017-09-10    1 (-82.71024 30.13608)
## 8 12075 2017-09-08    1 (-82.87318 29.56196)
## 9 12045 2017-09-09    2 (-85.25815 30.06198)
## 10 12031 2017-09-12   1 (-81.76086 30.37913)
```

Open Data APIs

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



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

[Live coding example]

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

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitude", "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))
```

``Kirk Moment''

The Grisly Murder Case That Could Turn Half of Oklahoma Back Into Tribal Lands

Did Congress fail to legally abolish the Muscogee Creek Nation's reservation? The Supreme Court may have to answer that question.

By **MATT FORD** | March 15, 2018

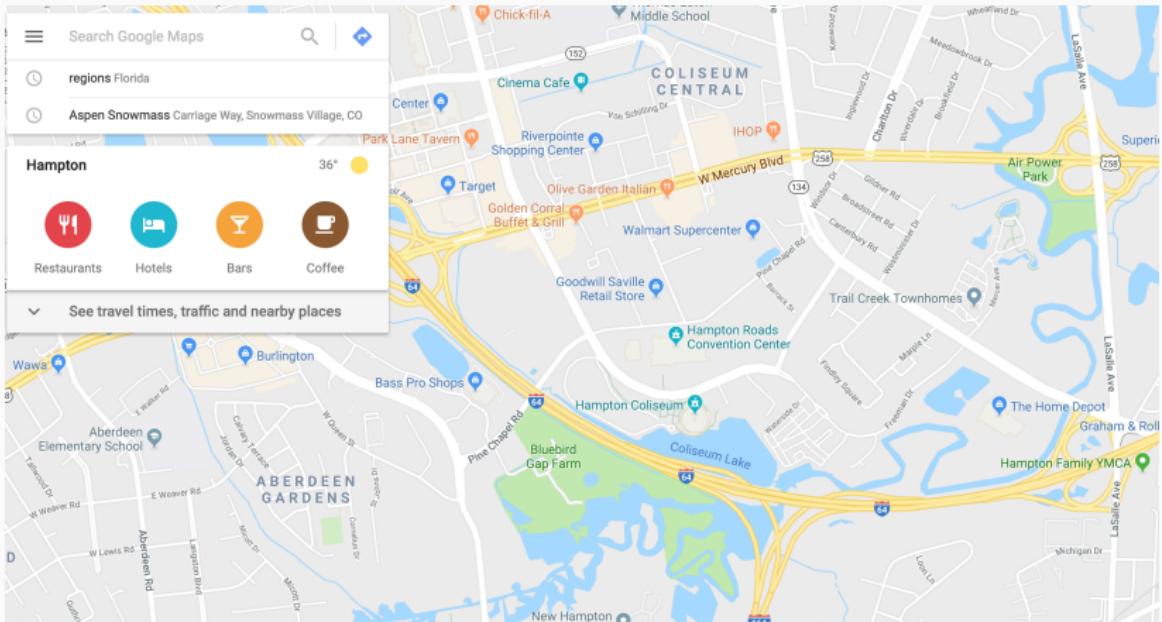
 Add to Pocket



INTERACT

R's **htmlwidgets** framework for interacting

Interactive graphics



JavaScript



htmlwidgets

<https://www.htmlwidgets.org/>

leaflet map

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

[Live coding example]

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

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitude", "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))
```

``Kirk Moment''

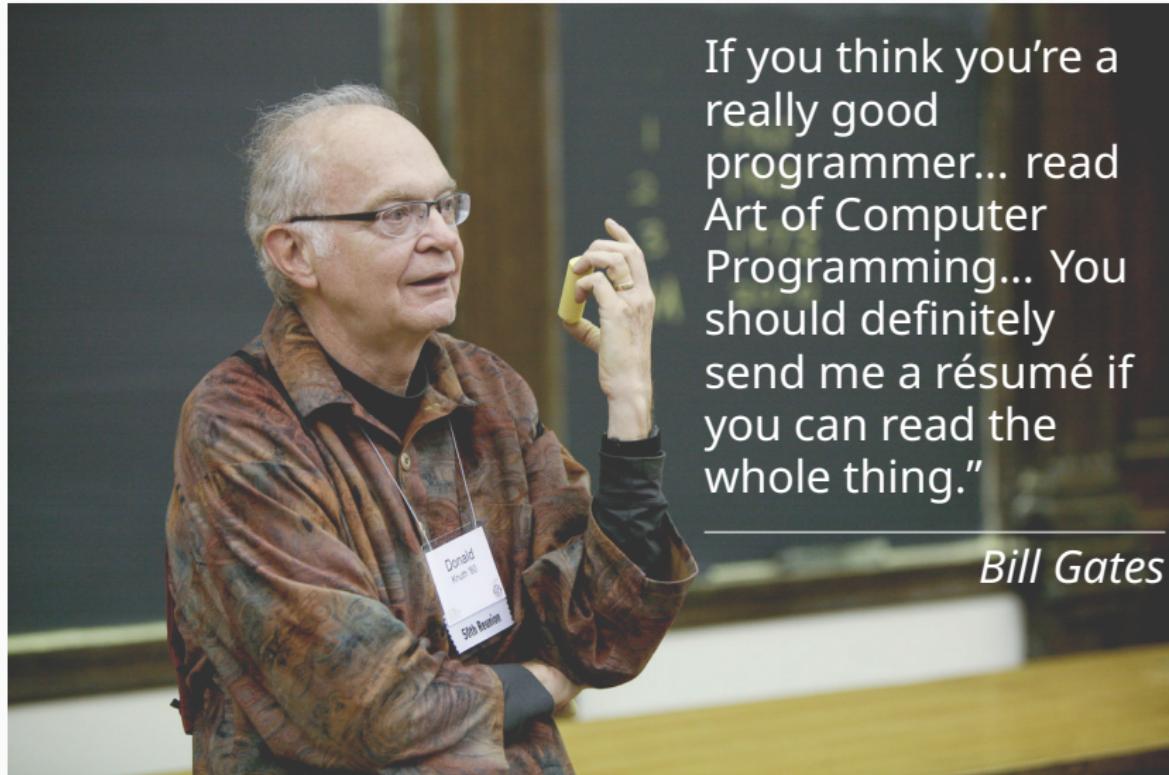


<https://github.com/ropenscilabs/miner>

REPORT

R's **RMarkdown** framework for reporting

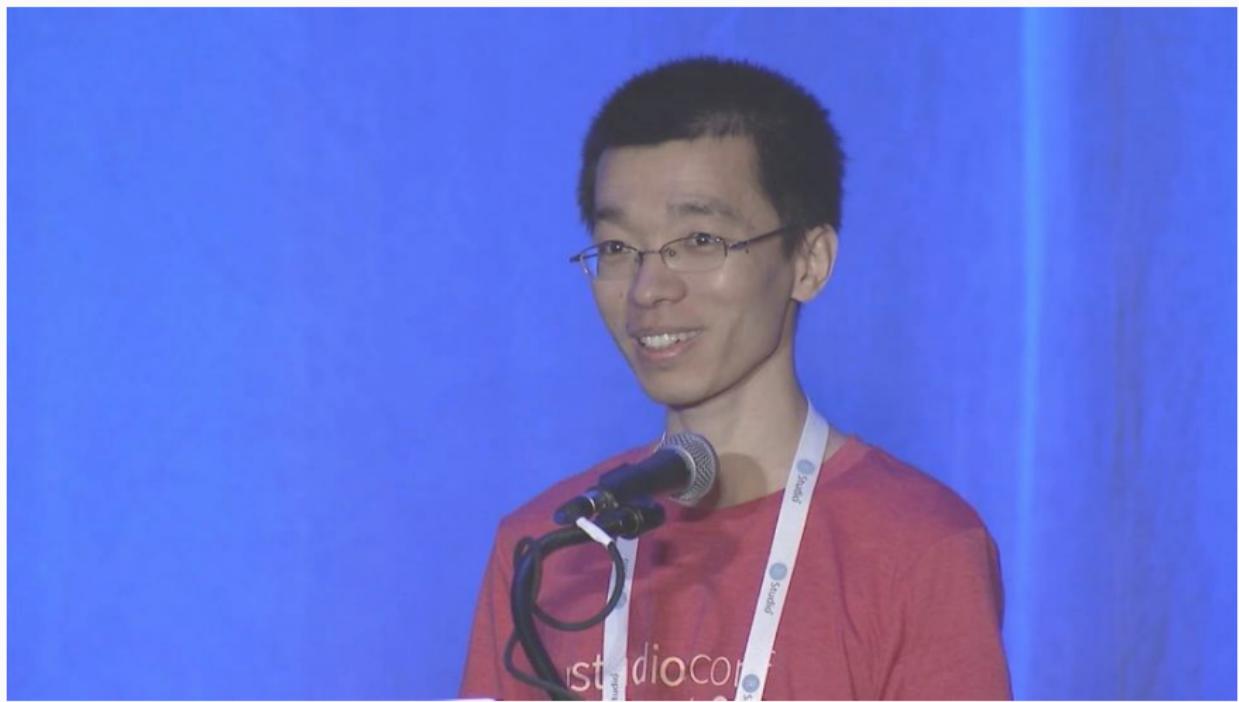
Donald Knuth



If you think you're a
really good
programmer... read
Art of Computer
Programming... You
should definitely
send me a résumé if
you can read the
whole thing."

Bill Gates

Yihui Xie



WYSISYG

What You See Is What You Get

Text of the report, with Markdown **format markers**.

```
```{r}  
number_one <- 1
```

```
number_one
```
```

More text, *also* with Markdown format markers.

And a list:

```
\- Item 1
```

```
\- Item 2
```

Text of the report, with Markdown **format markers**.

```
number_one <- 1
```

```
number_one
```

```
## [1] 1
```

More text, *also* with Markdown format markers. And some items:

Item 1

Item 2

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

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

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitude", "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))
```

``Kirk Moment''

1 Prerequisites 2 Plot 3 Map 4 Interact 5 Report 6 Tidy 7 Final Words

Data Visualization in R

Workshop for the 2019 Navy and Marine Corps Public Health Conference

Brooke Anderson

March 28, 2019

Chapter 1 Prerequisites

I HAVE BASED THIS WORKSHOP on examples for you to try yourself, because you won't be able to learn how to program unless you try it out. I've picked example data that I hope will be interesting to Navy and Marine Corp public health researchers and practitioners. You can download the slides from the workshop by [clicking here](#).

To try out these examples, you need some set-up:

1. Download R
2. Download RStudio
3. Install some R packages
4. Download example R Project

TIDY

R's **tidyverse** framework for tidying

``Tidy" data

| country | year | cases | population |
|-------------|------|-------|------------|
| Afghanistan | 1990 | 15 | 17071 |
| Afghanistan | 2000 | 666 | 2095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 84888 | 17404898 |
| China | 1999 | 21258 | 127215272 |
| China | 2000 | 21666 | 128023583 |

variables

| country | year | cases | population |
|-------------|------|-------|------------|
| Afghanistan | 1990 | 15 | 17071 |
| Afghanistan | 2000 | 666 | 2095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 84888 | 17404898 |
| China | 1999 | 21258 | 127215272 |
| China | 2000 | 21666 | 128023583 |

observations

| country | year | cases | population |
|-------------|------|-------|------------|
| Afghanistan | 1990 | 15 | 17071 |
| Afghanistan | 2000 | 666 | 2095360 |
| Brazil | 1999 | 31737 | 17206362 |
| Brazil | 2000 | 84888 | 17404898 |
| China | 1999 | 21258 | 127215272 |
| China | 2000 | 21666 | 128023583 |

values

Source: *R for Data Science*, Grolemund and Wickham

Untidy data example

| | Rural Male | Rural Female | Urban Male | Urban Female |
|-------|------------|--------------|------------|--------------|
| 50-54 | 11.7 | 8.7 | 15.4 | 8.4 |
| 55-59 | 18.1 | 11.7 | 24.3 | 13.6 |
| 60-64 | 26.9 | 20.3 | 37.0 | 19.3 |
| 65-69 | 41.0 | 30.9 | 54.6 | 35.1 |
| 70-74 | 66.0 | 54.3 | 71.1 | 50.0 |

Why are Legos such a great toy?



Tidyverse---Small functions that play well together

select some columns

slice to certain rows, or **filter** to rows that meet certain conditions

mutate existing columns to create new ones or change the old ones in place

unite separate columns into one

summarize the data, possibly after you **group_by** certain characteristics

Pipe operator



Cleaning data with the tidyverse

```
f1_accidents %>%  
  rename_all(.funs = str_to_lower) %>%  
  select(state, county, day, month, year, latitude, longitud, fatal)  
  filter(state == 12) %>%  
  mutate(county = str_pad(county, width = 3, pad = "0")) %>%  
  unite(col = fips, c(state, county), sep = "") %>%  
  unite(col = date, c(month, day, year), sep = "-") %>%  
  mutate(date = mdy(date)) %>%  
  filter(date >= mdy("9-7-2017") & date <= mdy("9-13-2017"))
```

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

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

[Live coding example]

```
fl_accidents <- fl_accidents %>%  
  st_as_sf(coords = c("longitude", "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))
```

``Kirk Moment''



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



Homework!!

goo.gl/7fPYUx