Getting started with R

R-Ladies Berlin

Lisa Hehnke

@DataPlanes

February 18, 2020

"R is a big part of who I am. I'm very involved in the R community. R is the thing I'm most passionate about."

Gabriela de Queiroz, Founder of R-Ladies

Start with WHY

"Why should I learn R? Isn't it yet another programming language?"

"Why should I learn R? Isn't it yet another programming language?"

1. Supportive, diverse, and inclusive community (one word: DataCamp 👏)

"Why should I learn R? Isn't it yet another programming language?"

1. Supportive, diverse, and inclusive community (one word: DataCamp 👏)
2. Free and open source software with >15k packages on CRAN

"Why should I learn R? Isn't it yet another programming language?"

1. Supportive, diverse, and inclusive community (one word: DataCamp 👏)
2. Free and open source software with >15k packages on CRAN
3. Extensive documentation and community support

"Why should I learn R? Isn't it yet another programming language?"

1. Supportive, diverse, and inclusive community (one word: DataCamp 👏)
2. Free and open source software with >15k packages on CRAN
3. Extensive documentation and community support
4. Go-to programming language for statistical analysis and data visualization

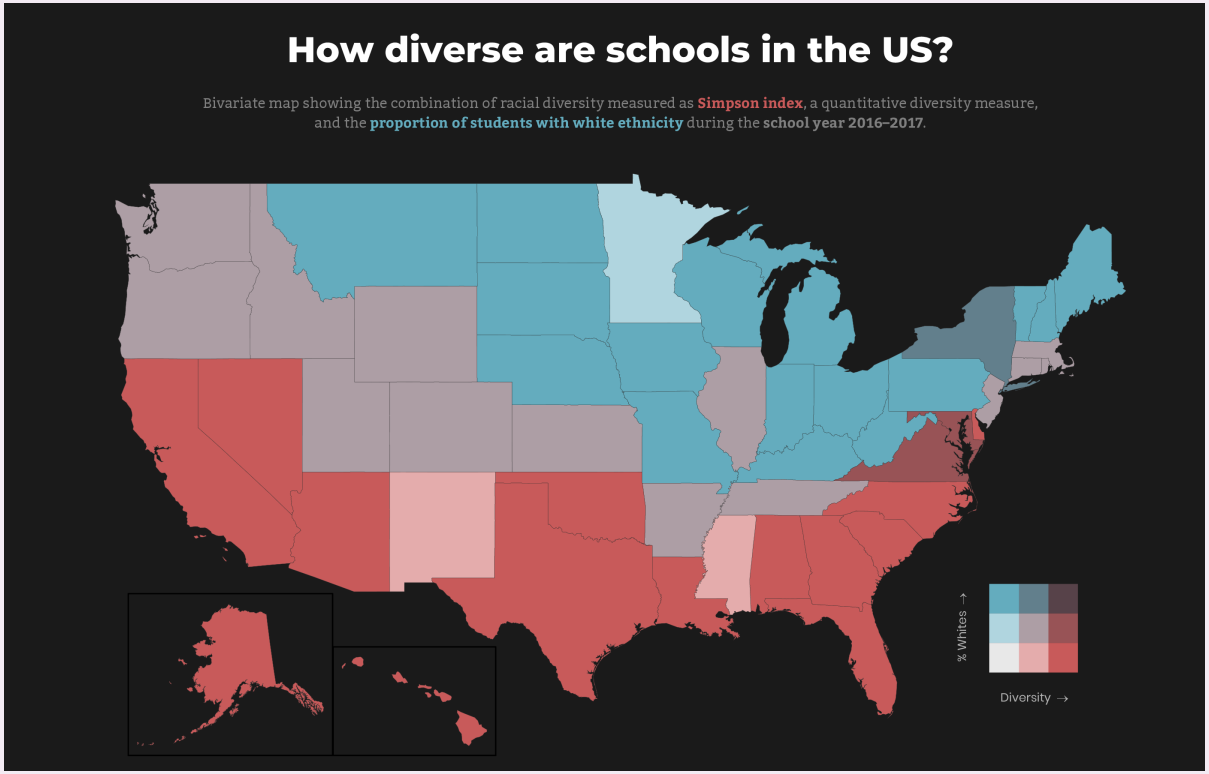
"Why should I learn R? Isn't it yet another programming language?"

1. Supportive, diverse, and inclusive community (one word: DataCamp 👏)
2. Free and open source software with >15k packages on CRAN
3. Extensive documentation and community support
4. Go-to programming language for statistical analysis and data visualization
5. Widely used in academia (no 💶), data-driven journalism (also no 🤑), and the industry (💵)

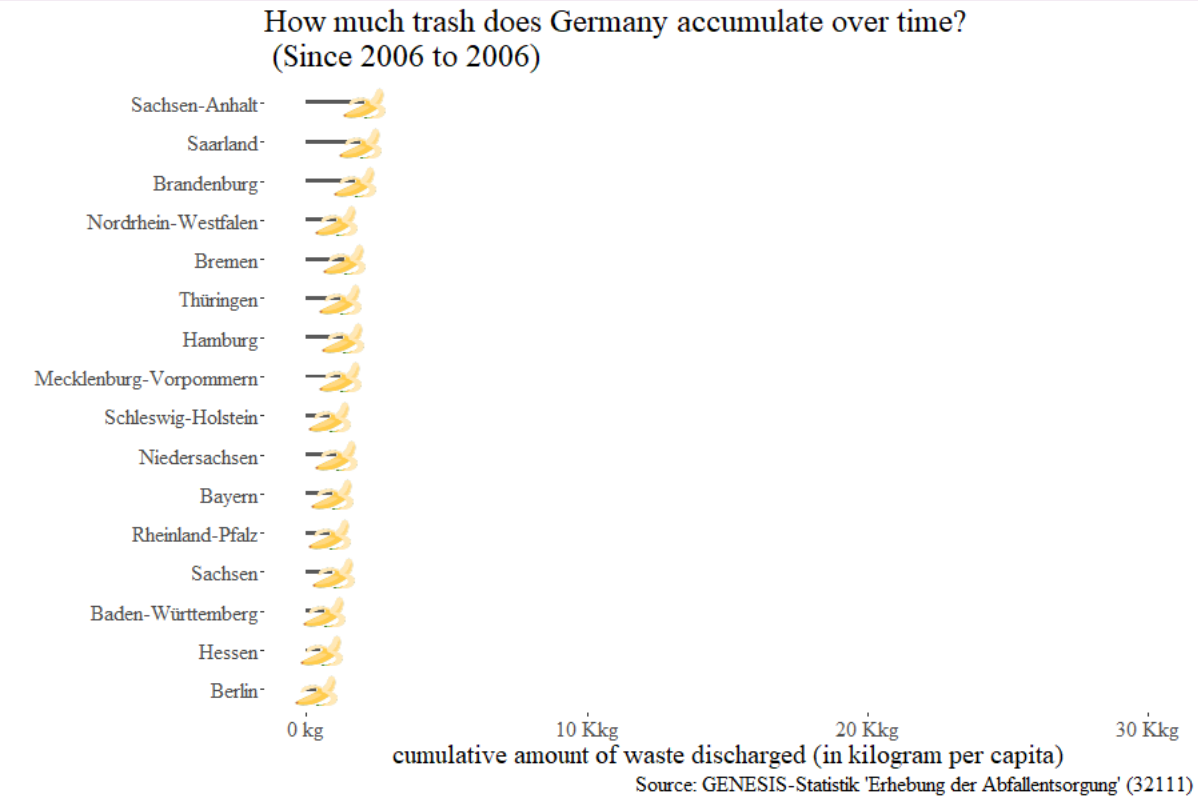
"Why should I learn R? Isn't it yet another programming language?"

1. Supportive, diverse, and inclusive community (one word: DataCamp 👏)
2. Free and open source software with >15k packages on CRAN
3. Extensive documentation and community support
4. Go-to programming language for statistical analysis and data visualization
5. Widely used in academia (no 💷), data-driven journalism (also no 🤑), and the industry (💶)
6. Why I *personally* love R: Learning how to fail, finding out why (thanks to Google and Stack Overflow), trying again, failing better, learning from this experience (aka refining Google searches), and eventually having a moment of instant gratification when the code runs through (or not) 🙏

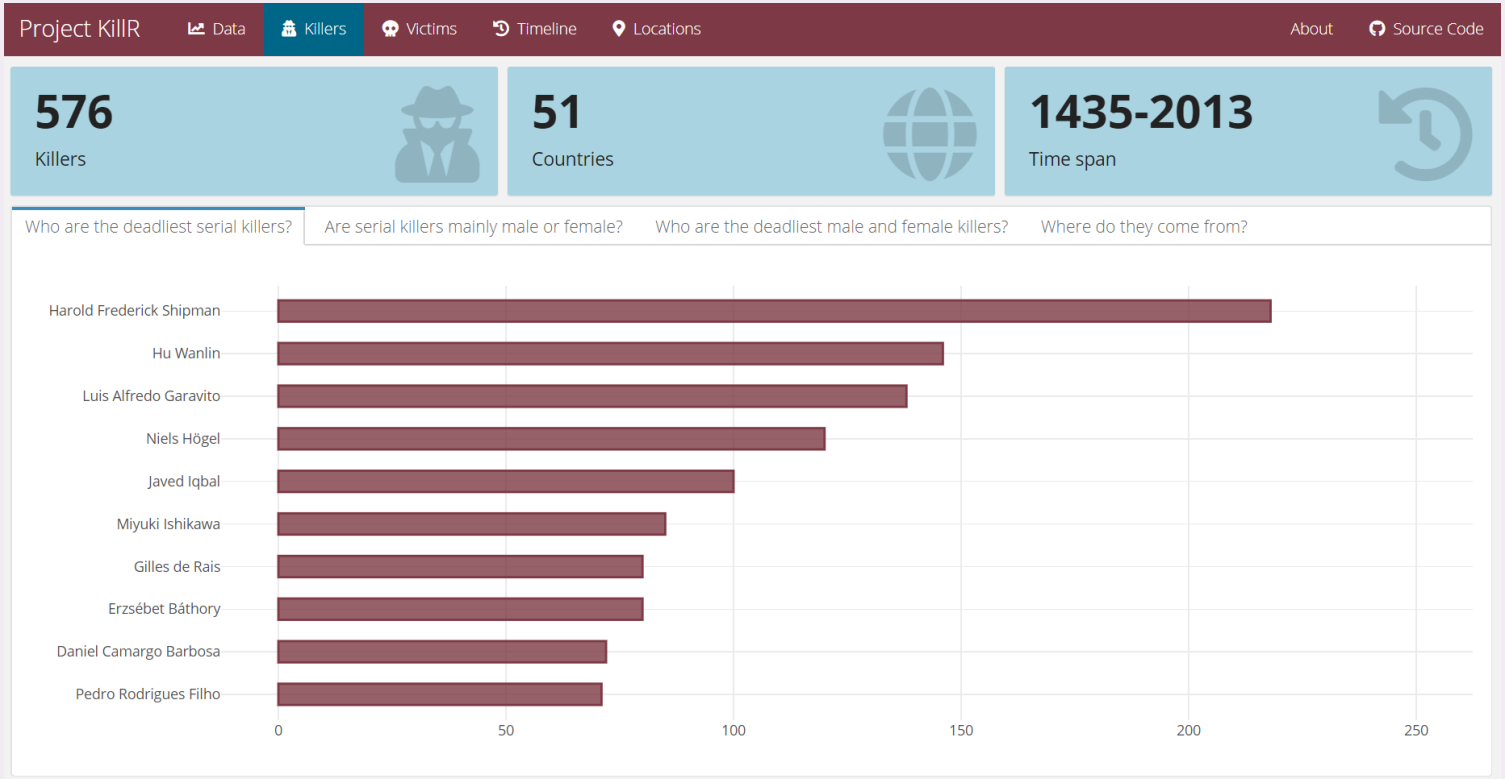
Some example projects: #tidytuesday data viz



Some more example projects: trashy animations



Even more example projects: killR interactive dashboard



Me (sneaky self promo)

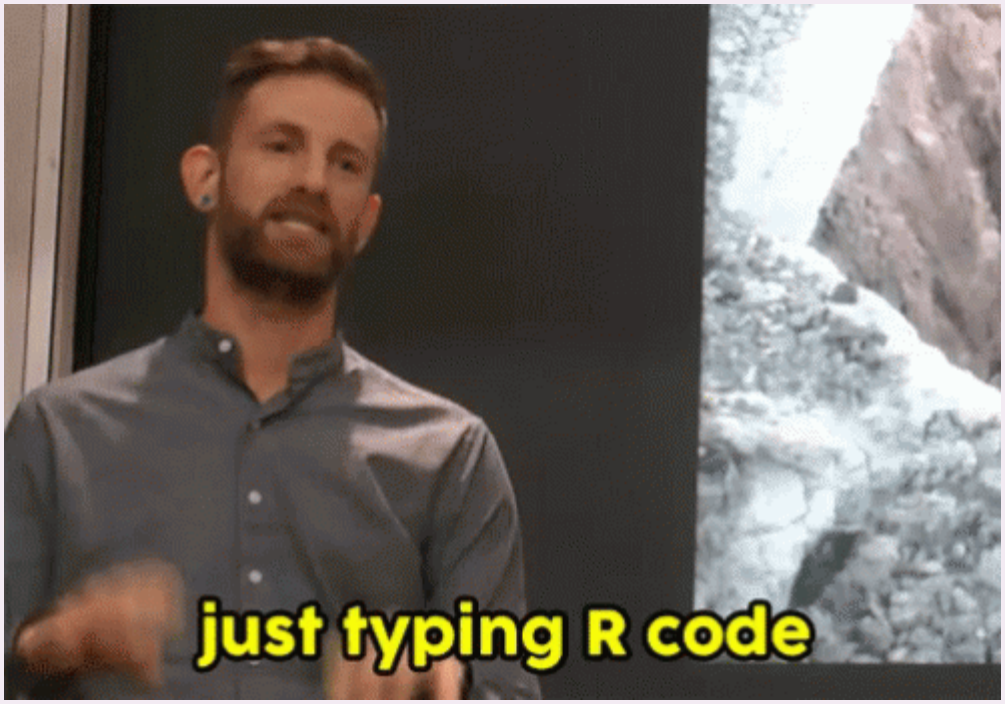
HOW

How do you use R?

It's pretty simple:

How do you use R?

It's pretty simple:

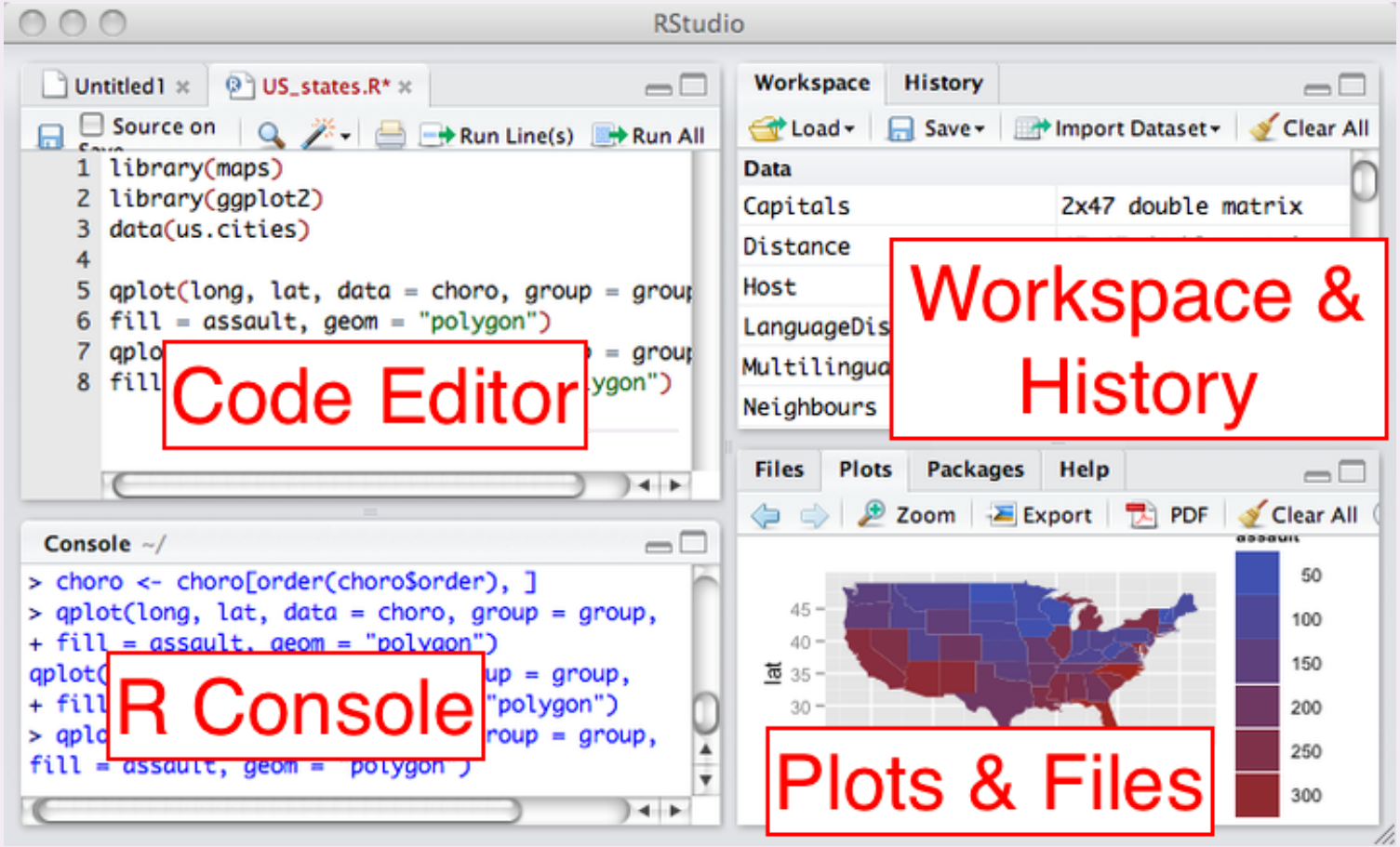


But before you can start typing R code, you first have to install the required software.

- Download **R**: https://cran.r-project.org/
- Download **RStudio**: https://rstudio.com/products/rstudio/

It's your turn to install the software.
If you already did, please help others who aren't there yet!

This is what RStudio looks like:



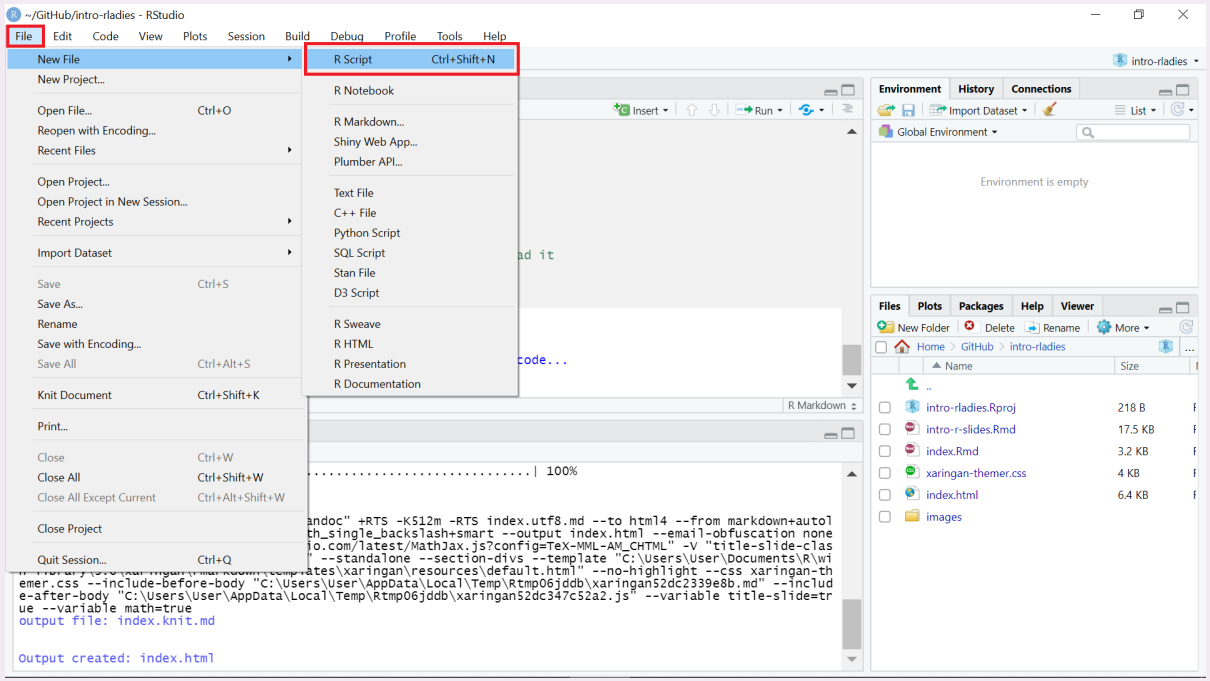
Now that you have the prerequisites ready, let's continue with the extras: packages!

```
# Install {swirl} package
install.packages('swirl')

# See which packages are already installed
library()

# Load the {swirl} package
library(swirl)

# Need help?
help(swirl)
```

"Do I have to type this every time? That's a lot of code..."

In short: no. You can use so-called *R scripts* to store (and rerun) your code:

File -> New File -> R Script (or Ctrl + Shift + N)



Please try it youRself:

Open a new R script, type the solutions to the exercises below, and run it.

If you don't know how to do it (yet), that's perfectly normal! We'll find out together.

Please try it youRself:

Open a new R script, type the solutions to the exercises below, and run it.

If you don't know how to do it (yet), that's perfectly normal! We'll find out together.

1. Add 2 + 2 in R.
2. Print *"Hello, R-Ladies Berlin!"* to the console (hint: *?print*).
3. Install and load the {tidyverse} package collection (trust me, you'll need it!).
4. Try to assign the result of 2 + 2 to a new variable named x (hint: <-).

Please try it youRself:

Open a new R script, type the solutions to the exercises below, and run it.

If you don't know how to do it (yet), that's perfectly normal! We'll find out together.

1. Add 2 + 2 in R.
2. Print *"Hello, R-Ladies Berlin!"* to the console (hint: *?print*).
3. Install and load the {tidyverse} package collection (trust me, you'll need it!).
4. Try to assign the result of 2 + 2 to a new variable named x (hint: <-).

Did it work?

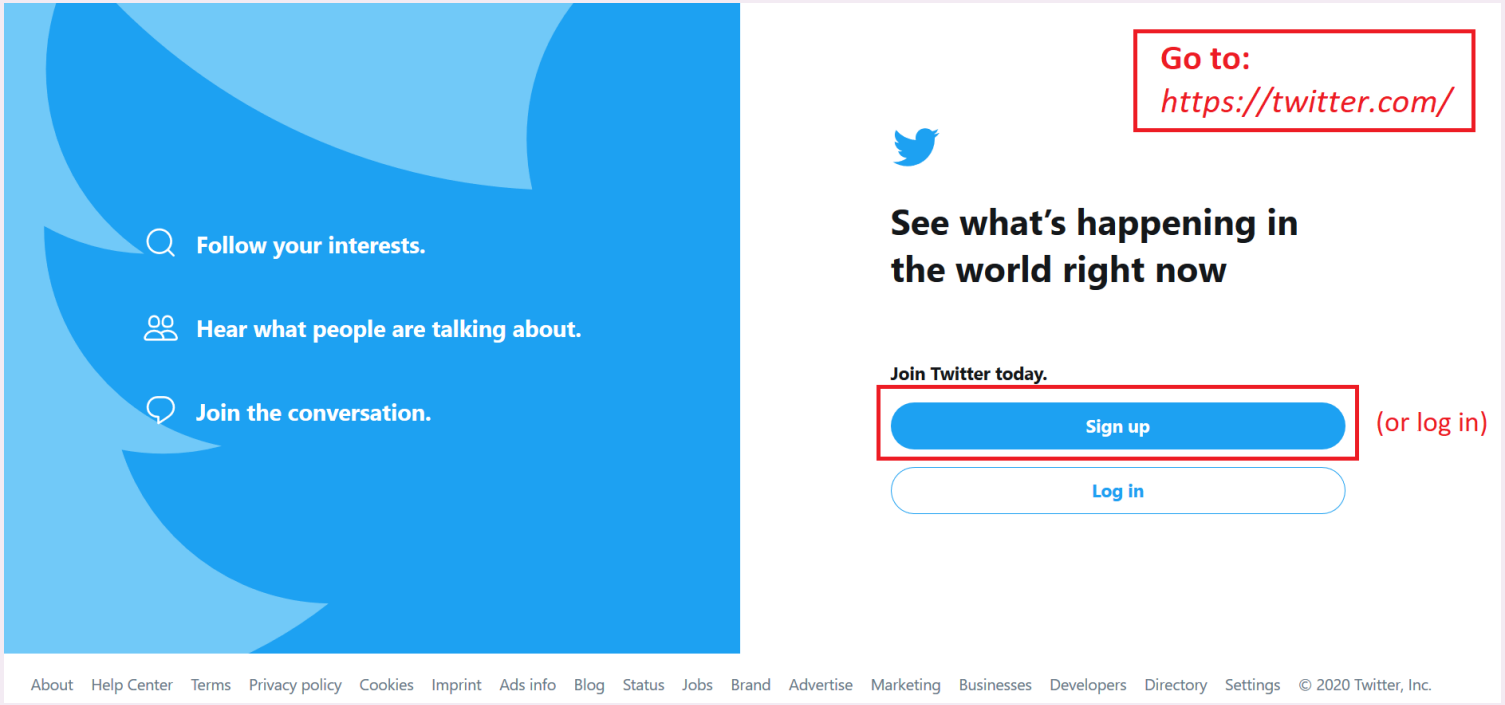
Here's the solution:

```
# Add 2 + 2.
2 + 2

# Print "Hello, R-Ladies Berlin!" to the console.
print("Hello, R-Ladies Berlin!")

# Install and load the {tidyverse}
install.packages('tidyverse')
library(tidyverse)

# Assign the result of 2 + 2 to a new variable named x
x <- 2 + 2
```

Congratulations, you also achieved the WHAT:
You wrote and executed your first R code! 🏆

Now on to the fun stuff:
Twitter data!

For the next steps, you need your own Twitter account.



We're using {rtweet} to download some tweets.

First, we install and load the {rtweet} package.

```
# Install the {rtweet} package
install.packages('rtweet')

# Load the package
library(rtweet)
```

We're using {rtweet} to download some tweets.

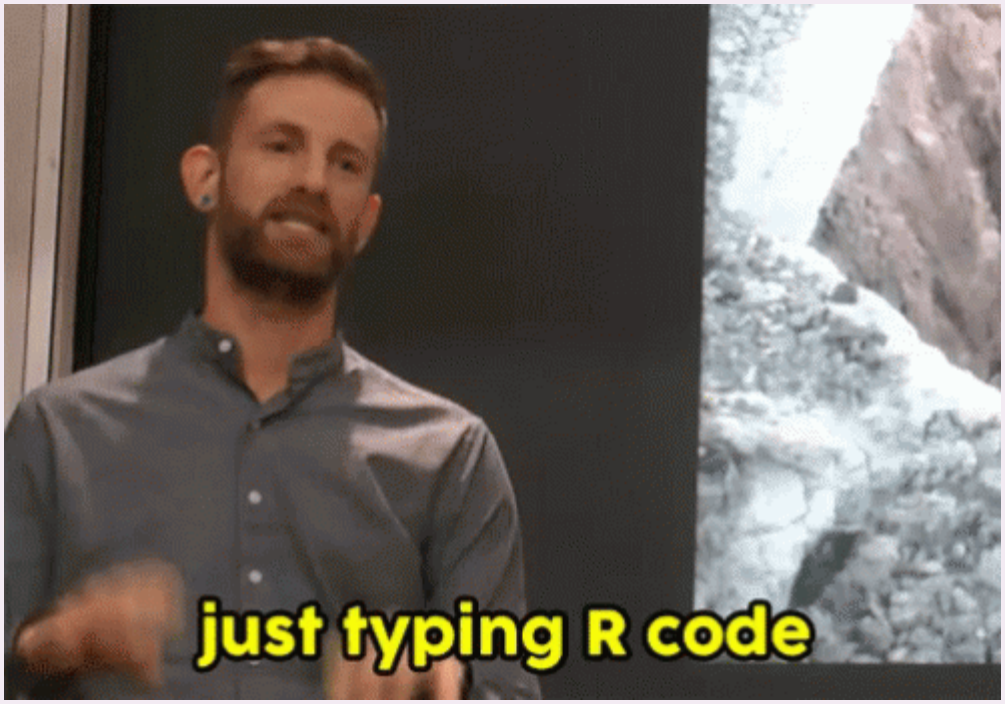
First, we install and load the {rtweet} package.

```
# Install the {rtweet} package
install.packages('rtweet')

# Load the package
library(rtweet)
```

Next, we download our data. Choose whichever option you like best:

```
# A: Search for 1k tweets using the rstats hashtag (w/o retweets)
data_raw <- search_tweets('#rstats', n = 1000, include_rts = FALSE)

# B: Get the most recent 1k tweets posted by R-Ladies Global
data_raw <- get_timeline('rladiesglobal', n = 1000)

# C: Get the 1k most recently favorited statuses by Hadley Wickham
data_raw <- get_favorites('hadleywickham', n = 1000)
```

(Yes, *that's* Hadley Wickham.)



First things first: data wrangling

Let's do the following steps together.

```
# Get a glimpse of the data
glimpse(data_raw)

# Select the variables containing the user handle,
# date, text, and number of favorites and assign
# the result to a new data frame called data_subset
data_subset <- data_raw %>%
  select(screen_name, created_at, text,
         favorite_count, retweet_count)

# Show all tweets with more than 100 favorites
data_subset %>%
  filter(favorite_count > 100)
```

And now some basic stats...

```
# Summary statistics
data_subset %>%
  summary()

# Summary stats using {skimr}
install.packages('skimr')
library(skimr)

data_subset %>%
  skim()
```

...and visualizations...

```
# Plot number of tweets over time using {ggplot2}
data_subset %>%
  ggplot() + geom_histogram(aes(x = created_at)) +
  xlab("Date") + ylab("Number of tweets") +
  ggtitle("Tweets over time") + theme_minimal()
```

...and fancier visualizations.

```r
# Install and load the {tidytext} package
install.packages("tidytext")
library(tidytext)

# Install and load the {wordcloud} package
install.packages("wordcloud")
library(wordcloud)

# Clean tweets using {tidytext}
tweets_tidy <- data_subset %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word)

# Plot wordcloud using {wordcloud} (base R style)
wordcloud(tweets_tidy$word, tweets_tidy$n, colors = "#88398a",
          min.freq = 5, max.words = 200)
```

Thanks for listening and happy coding!

Slides made with ♡ and R *xaringan*.