

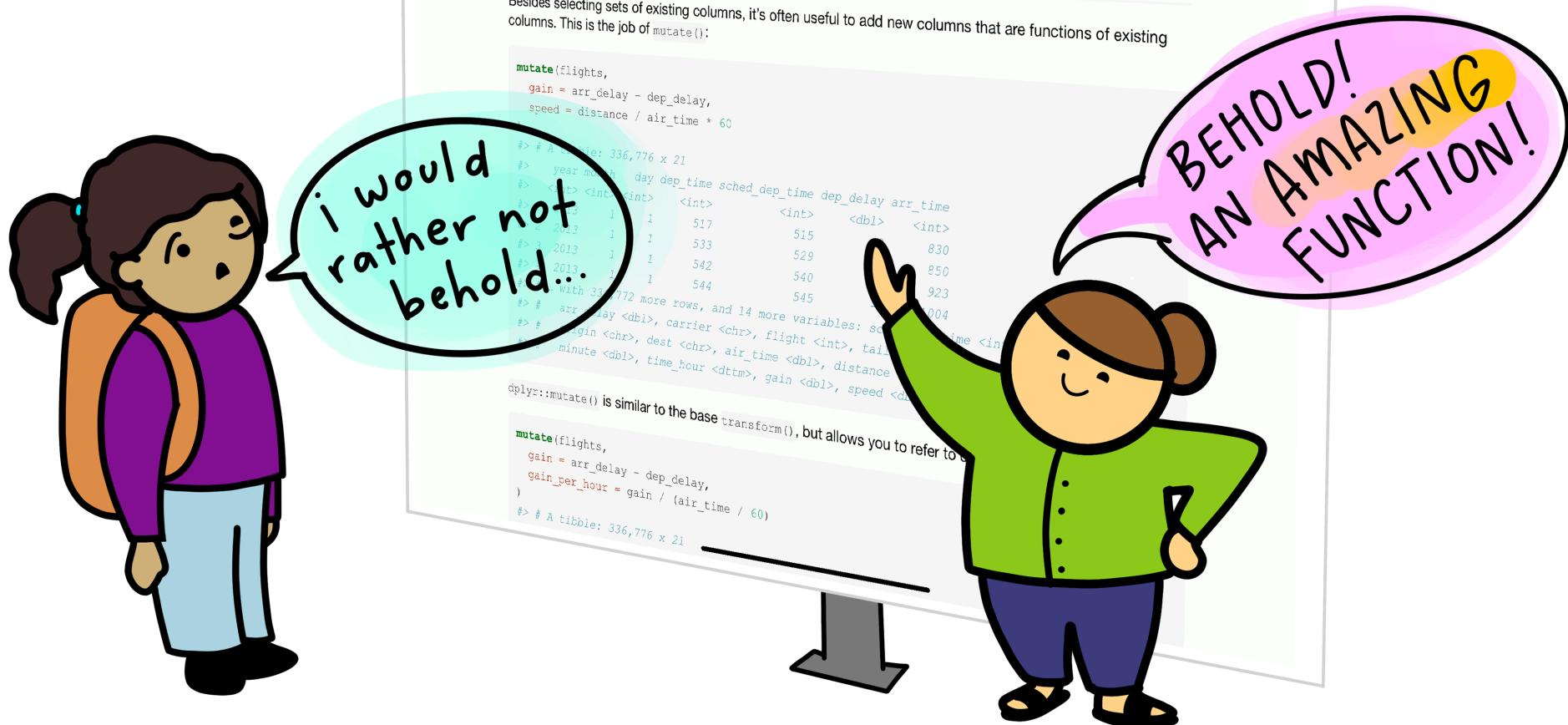
Intro to EPI 590R

Why this class?

About this class

Goal: Learn some best practices to make your life in R easier and your research more reproducible

- Quick! Intense!
 - It will require practice afterward, and time to sink in
 - The goal is to set you up for success and give you resources to learn more
- You don't have to use everything you learn here!
 - Some of these tools I use for *every* project, some just occasionally
 - Experiment with what works for you, a little at a time



About this class

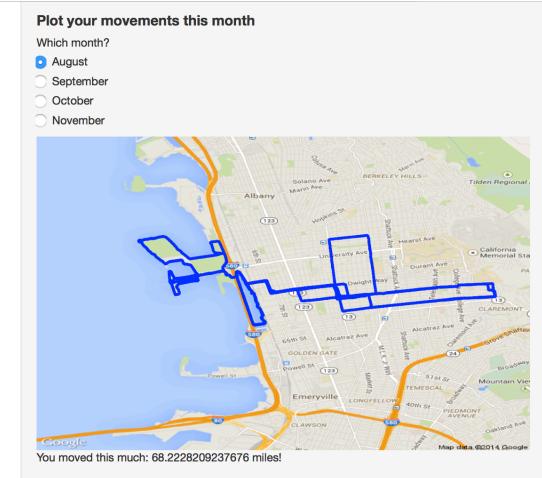
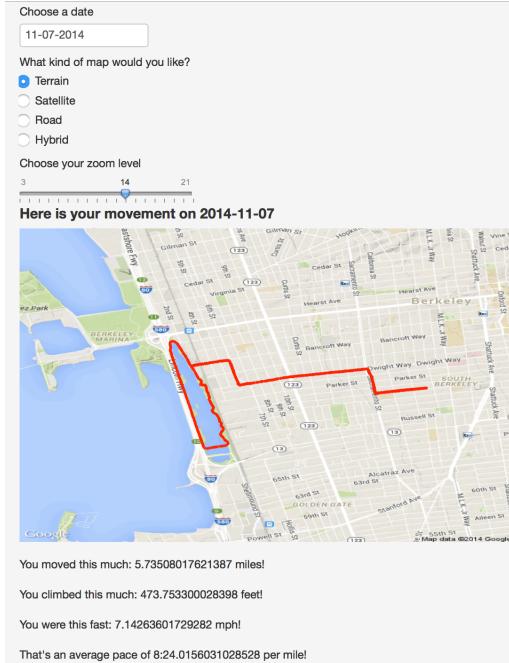
- Everything you need is at
<http://epi590r.louisahsmith.com>
 - Canvas will link you there, but good to bookmark as well
 - The website will be up indefinitely
- General format:
 - Some slides (introduce new content and overview of exercises)
 - I'll demonstrate while you watch
 - Practice on your own/with your classmates

Balance



About Louisa

- Assistant professor at Northeastern University
 - Department of Health Sciences and the Roux Institute (Portland, Maine)
- Started using R during my master's (so just about 10 years of experience)
 - I learned mostly by doing!
 - Twitter, blogs, RStudio::conf videos, meetups
- Basically everything I do is in R!
- Actual epi research in causal inference, pregnancy, lots of other stuff



About Ollie



Why this class?

Economics

Genes

Orangutans

3.2 Spreadsheet coding error

In addition to these deliberate data exclusions by RR, a coding error in the RR working spreadsheet also unintentionally excludes five countries entirely (Australia, Austria, Belgium, Canada and Denmark) from all parts of the analysis.⁹ The error appears in the calculations of both mean and median GDP growth with the 1946–2009 sample as well as with the mean and median GDP growth for the sample over the 220-year period 1790–2009. The omitted countries are selected alphabetically. It is clear from the spreadsheet itself that these are random exclusions. RR have since acknowledged this to be the case ([RR, 2013A](#), [2013B](#), [2013C](#)).

Errors are everywhere

Joana Grave
@joanafqg

A few months ago, I discovered an error in the database of my first paper. And today, the retraction notice is finally out! Please don't be afraid to talk about your errors and to correct them. It's hard, but errors do happen!



sciedirect.com
Retraction notice to "The effects of perceptual load in processing emotional facial expression in..."

3:51 PM · Jul 11, 2021

24 Retweets 6 Quotes 187 Likes 2 Bookmarks

Julia Strand
@juliastrand

I recently found a massive error in one of my published papers. The main finding was the result of a programming bug and was, in fact, completely untrue. THREAD with the very short version below, essay with the full description here:



elemental.medium.com
Scientists Make Mistakes. I Made a Big One.
A researcher learns the right thing to do when the wrong thing happens

2:16 PM · Mar 24, 2020

Leigh Senderowicz
@LSenderowicz

Hi all, today I'm writing a post that no researcher ever wants to write: We've discovered a coding error in the analysis of an article that's already been published.

A short ↗



onlinelibrary.wiley.com
Supply-Side Versus Demand-Side Unmet Need: Implicati...
Despite its central importance to global family planning, the "unmet need for contraception" metric is frequently ...

1:20 PM · Aug 2, 2023 · 62.7K Views



No one and no field is immune from errors in data analysis. Our goal is to make them as unlikely as possible (and report them when we find them!)

Not only improves science but also your life!

- It's really boring to copy lots of numbers into a table
 - And then change a tiny thing in the analysis and do it all over again
- It's really frustrating to lose work when your computer crashes, or completely change an analysis before your advisor forgets what they told you last time and has you change it back
- It's fun when things just work! And you get more time for the fun parts of epidemiology

BEEP BEEP BEEP!!



oh just add
SALT!?!?



@allison_horst

bless this
workflow



@allison_horst

Questions?

Exercises: Connecting to GitHub

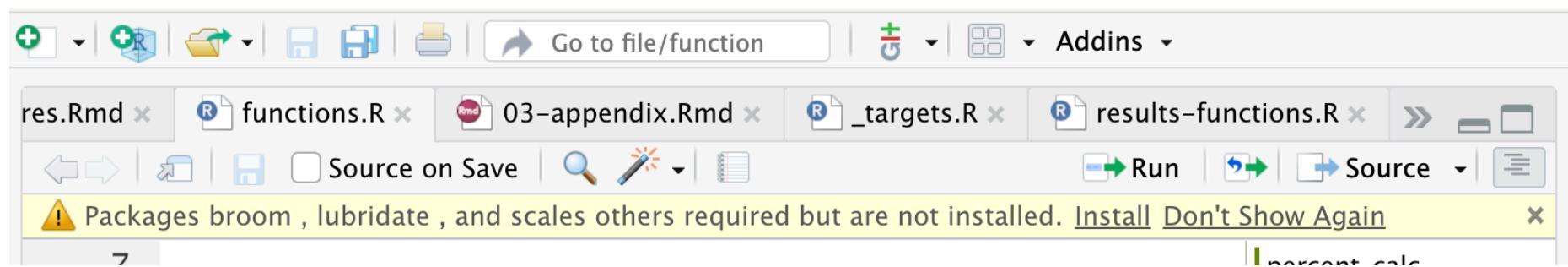
1. Install the `{usethis}` package:

```
install.packages("usethis")
```

Installing packages

If you just updated R to a new “major” version, you will need to reinstall packages

- I tend to do this as I need them rather than try to reinstall them all at once
 - RStudio tries to help!



Possible errors

Spelling the package or function's name wrong, or not installing or loading the package

```
> install.packages("blah")
Warning in install.packages :
  package ‘blah’ is not available for this version of R
```

A version of this package for your version of R might be available elsewhere,
see the ideas at

<https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages>

```
> library(blah)
Error in library(blah) : there is no package called ‘blah’
> blah::blah()
Error in loadNamespace(x) : there is no package called ‘blah’
> blah()
Error in blah() : could not find function "blah"
```

Using packages: library vs. :::

If you are writing a script you will save, and will use several functions from this package

```
1 library(usethis)
2 use_git_config(user.name())
3 user.email()
```

If you are just running some quick code in the console or only need to use the package a few times in a script

```
1 usethis::use_git_config()
```

I try to only run `library(package)` from a script (not the console) so that there's a "record" of me loading the package, or else I might accidentally write code that doesn't work later

Set up git

2. Introduce yourself:

```
use this::use_git_config(user.name = "Louisa Smith",  
user.email = "louisahsmith@gmail.com")
```

When you make changes to your code, they will be associated with this name and email address (this doesn't really matter for our purposes)

- You only need to do this once
- We'll explain all this in a little bit!

Since I only need to run this once, I would probably run this from the console (bottom) rather than a script (top)

The screenshot shows the RStudio interface. At the top is a script editor window titled "Untitled1*". It contains the following R code:

```
1 library(usethis)
2 use_git_config(user.name = "Louisa Smith", user.email = "louisahsmith@gmail.com")
3 # or if I don't plan to use a lot of functions from the usethis package
4 usethis::use_git_config(user.name = "Louisa Smith", user.email = "louisahsmith@gmail.com")
```

The status bar at the bottom of the script editor indicates "3:72 (Top Level)". To the right of the script editor is a "Source" tab. Below the script editor is a "Console" tab, which is currently active. The console window shows the command being run:

```
R 4.3.1 · ~/Documents/Teaching/Emory/epi590r-in-class/
> usethis::use_git_config(user.name = "Louisa Smith", user.email = "louisahsmith@gmail.com")
```

Running from the console is great for `install.packages()`, quick calculations, fiddling with code until you get it right, or scenarios like this – otherwise save your code in a script!

Connect to GitHub

3. Create a github token:

```
usethis::create_github_token()
```

Instead of entering your password every time, this is a secure way to connect to GitHub

- If you are ever asked for your GitHub password in RStudio, you *have* to give this instead

Connect to GitHub

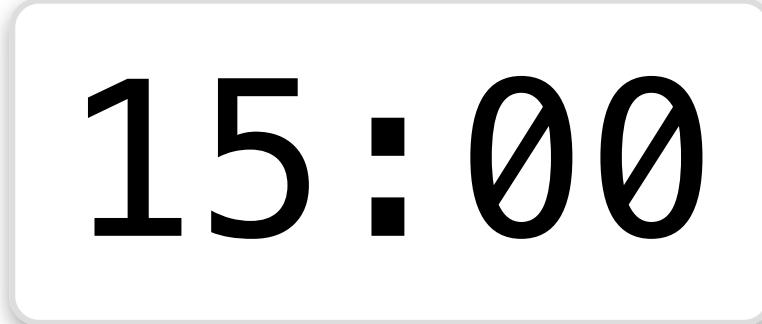
4. Copy the token
5. Back in R, run this code and paste your token in the console where it says “Enter password”:

```
gitcreds::gitcreds_set()
```

You can do this again whenever your token expires or you are using a different device

Exercises

- Refer back to the slides as needed
- Ask a classmate if you're stuck
- Raise your hand for the teaching team
- Done early? Help a friend! Read the resources section!
Play around in R! Check your email!



15:00