Data Wrangling with R: Day 1

Getting started

Presented by Emi Tanaka

Department of Econometrics and Business Statistics

MONASH University

Emi Tanaka
Lecturer in Statistics
Monash University

🌐 https://emitanaka.org/

✉ emi.tanaka@monash.edu

🐦 @statsgen

About your instructor

- I teach R at university courses as well as use it daily for research and hobby 👽

- My research includes mixed models, experimental design, statistical genetics and data visualisation.

- My main application area is bioinformatics and agricultural, particularly plant breeding.

- My childhood dream was to be a comic artist.. it's not too late, is it?

Fan Cheng
PhD Student
Monash University

🌐 https://fancheng.me/

✉ Fan.Cheng@monash.edu

🐦 @fanchengfc

About your helper

- I am a second-year PhD student in Statistics at Monash

- I work for the Monash Data Fluency team to teach workshops on R, Python, Git and Unix

- My research project is about manifold learning based on empirical probability distributions. If you are interested, I am more than happy to talk about it :)

Introduce yourself in the chat!

What's your profession?

Where do you live?

What's your hobby?

What's would you like for people to know about you?

Day 1 Schedule

Tue 1st Dec 2020

Time	Activity
1.00-1.30	Registration + Installation Help
1.30-2.00	**Lesson 1:** Base R and `tidyverse`
2.00-2.15	*Break out session*
2.15-2.45	**Lesson 2:** Data manipulation with `dplyr`
2.45-3.00	*Break out session*
3.00-3.30	*Coffee Break*
3.30-4.00	**Lesson 3:** Manipulating strings with `stringr`
4.00-4.15	*Break out session*
4.15-4.45	**Lesson 4:** Tidying data with `tidyr`
4.45-5.00	*Break out session*

Day 2 Schedule

Wed 2nd Dec 2020

Time	Activity
1.00-1.30	Installation Help
1.30-2.00	**Lesson 1:** Relational data wrangling, starring `janitor` and `broom`
2.00-2.15	*Break out session*
2.15-2.45	**Lesson 2:** Formatting factors with `forcats`
2.45-3.00	*Break out session*
3.00-3.30	*Coffee Break*
3.30-4.00	**Lesson 3:** Dealing with dates with `lubridate`
4.00-4.15	*Break out session*
4.15-4.45	**Lesson 4:** Advanced data wrangling, starring `purrr`
4.45-5.00	*Break out session*

Getting Started

```r
remotes::install_github("emitanaka/dwexercise")
```

> ℹ️
> - Find the latest updated material at 🌐 http://emitanaka.org/datawrangle-workshop-ssavic/
> - All materials are hosted in GitHub repo here
> - All exercises are hosted as an R-package on GitHub repo here
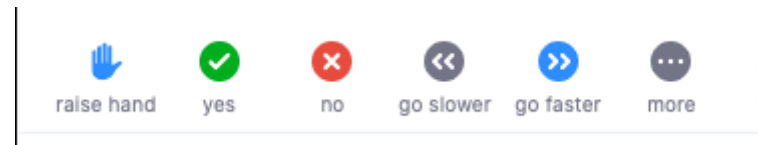> - You can download the current material here

- Some materials may be updated before the workshop is completed so it will be better to re-download the latest material at the conclusion of the workshop.

- I will send you a follow up email with link to download all material

Help me!

You can write in the chat.

You can ask (unmute yourself).

You can raise hand from zoom under Participants.

We'll come around in breakout sessions.

```
devtools::session_info()

## — Session info ─────────────────────────────────────────────────
##  setting  value
##  version  R version 4.0.1 (2020-06-06)
##  os       macOS Catalina 10.15.7
##  system   x86_64, darwin17.0
##  ui       X11
##  language (EN)
##  collate  en_AU.UTF-8
##  ctype    en_AU.UTF-8
##  tz       Australia/Melbourne
##  date     2020-11-30
##
## — Packages ──────────────────────────────────────────────────────
##  package      * version date       lib source
##  anicon         0.1.0   2020-06-21 [1] Github (emitanaka/anicon@0b756df)
##  assertthat     0.2.1   2019-03-21 [2] CRAN (R 4.0.0)
```

These slides are licensed under