Coding 1: Lecture 1

Marc Kaufmann
Central European University

9/10/2019

Coding 1: Data Management and Analysis with R

Basic Admin

- Weeks 1 to 6: Mondays 13:30-16:10
- Weeks 7 to 12: Mondays 17:30-19:10
- Instructor: Marc Kaufmann (call me Marc)
- Teaching Assistant: Júlia Hermann gives 3 sessions
  - 17:30-19:10 on Tuesday October 8th
  - 17:30-19:10 on Tuesday November 5th
  - 17:30-19:10 on Tuesday November 19th

Who am I?

Who am I?

- Assistant Professor in Economics and Business
- Research in Psychology and Economics (aka Behavioral Economics)

Who am I?

- Assistant Professor in Economics and Business
- Research in Psychology and Economics (aka Behavioral Economics)

Relevant for this class:

- Collect my own data in (mostly online) experiments
- Analyze said data: Fairly basic, since design is up to me
- Program in Racket, bash/unix, Python, and R (in that order of competence)

Who am I?

- Assistant Professor in Economics and Business
- Research in Psychology and Economics (aka Behavioral Economics)

Relevant for this class:

- Collect my own data in (mostly online) experiments
- Analyze said data: Fairly basic, since design is up to me
- Program in Racket, bash/unix, Python, and R (in that order of competence)

Most importantly: I am good at getting help.

Goal for the class in one sentence

Goal for the class in one sentence

Generate basic insights

Goal for the class in one sentence

Generate basic insights from existing data

Goal for the class in one sentence

Generate basic insights from existing data that is small

Goal for the class in one sentence

Generate basic insights from existing data that is small and relational

Goal for the class in one sentence

Generate basic insights from existing data that is small and relational in a reproducible and replicable manner.

Goal for the class in one sentence

> *Generate basic insights from existing data that is small and relational in a reproducible and replicable manner.*

Or expressed in R (pseudo-)code:

Goal for the class in one sentence

> *Generate basic insights from existing data that is small and relational in a reproducible and replicable manner.*

Or expressed in R (pseudo-)code:

```
library(msc_ba)

eureka_or_bust <- your_great_ideas %>%
  # We need data to figure this out. Let's...
  collect_data() %>%
  # Apply knowledge from this class...
  code_1()
```

Goals of the Course

```
code_1 <- function(collected_data) {
  collected_data %>%
    read_in() %>%
    explore() %>%
    visualize() %>%
    summarize() %>%
    clean() %>%
    tidy() %>%
    analyze() %>%
    knit()
}
```

Focus of the Class

Since we cover much of the data analysis cycle and have little time:

- Focus on few libraries and commands (80/20 rule)
  - Tidyverse only: coherent set of tools with sane interface
- Focus on correct code; then maintainable; then fast
- Focus on *fluency*
- Focus on teaching you how to learn
  - Clean code; documenting; debugging; communicating
  - How to get help

Focus of the Class

Strive to write code that is correct; maintainable; and fast. The ordering of these adjectives is critical: correct is more important than maintainable; maintainable is more important than fast; and fast is important to include, because nobody wants to live with slow programs.
From "How to Program Racket: a Style Guide", Felleisen et al

Assignments and Grading

- Participation (30%): helping yourself and helping others
  - Includes attendance. Starting week 2, will take attendance via https://www.youhere.org/
    - Let me know if you cannot or do not want to use that.
- Assignments (77-78%): weekly assignments in the form of R Markdown notebooks, and assessing those of your peers
  - Grading is N% in week N ($N = 1, ..., 12$)
- No exam

Assignments and Grading

- Participation (30%): helping yourself and helping others
  - Includes attendance. Starting week 2, will take attendance via https://www.youhere.org/
    - Let me know if you cannot or do not want to use that.
- Assignments (77-78%): weekly assignments in the form of R Markdown notebooks, and assessing those of your peers
  - Grading is N% in week N ($N = 1, ..., 12$)
- No exam

Total: 107-108%

Useful Resources

When you get stuck, the following fantastic books may help:

- Kieran Healy's book on *Data Visualization* (http://socviz.co/)
- Grolemund and Wickham's book *R for Data Science* (https://r4ds.had.co.nz/)

Additionally:

- Tuesdays 17:30-19:10, N13 309: Coding practice with R
- Júlia's sessions
- Ask questions on https://discourse.trichotomy.xyz

Where you should be

You should:

- Have RStudio set up up and working
- Have git set up and working
- Have cloned the repository for the course
- Have an account on https://discourse.trichotomy.xyz

First time I teach R

I tend to experiment quite a bit:

- Early on, still worth figuring out what works
- Especially assignments and how to incentivize teamwork
- After week 3, should be ironed out

First time I teach R

I tend to experiment quite a bit:

- Early on, still worth figuring out what works
- Especially assignments and how to incentivize teamwork
- After week 3, should be ironed out

Any questions?