MPP-E1180 Lecture 1: Introduction to the Course

Christopher Gandrud

11 September 2014

Objectives for the week

- Introduce the course goals, plan, and expectations/assessment
- Introduce Collaborative & Reproducible Data Analysis
- Set up computational research environment

Christopher Gandrud

Contact:

- Public
  - Hertie_Collab_Data_Science/issues
  - @ChrisGandrud
- Private
  - gandrud@hertie-school.org

Official Office Hours:

- Room: 1.52
- Wednesday 15:00-17:00

Objectives for the course

Prerequisites

- **Introductory-level statistics**
  - Basic descriptive statistics (e.g. data types, ways of describing distributions)
  - Basic inferential statistics: (significance testing, linear regression)
  - Exposure to statistics software (e.g. SPSS, STATA)
- Knowledge of particular software or computer programming is **not expected**
- **Patience**

Assessment

- 3 Pair Assignments (Weeks 3, 6, 9)
  - 10% each
- Collaborative Research Project (Presentation: Week 12, Website/Paper: Exam Week)
  - 50%
- Attendance & Active Participation
  - 20%
- No traditional midterm or final exam

Assessment (details)

- **Due**: Midnight on Friday of the week it is due.
- All assignments must be submitted electronically on GitHub.
- All assigments must be completed in **pairs**
- More details will be given on the specific pair assigments/research project in future classes.

Assessment (attendance, participation)

-

Course Outline

Syllabus

`https://github.com/christophergandrud/Hertie_Collab_Data_Science`

- The syllabus will be **updated**. **Check regularly**.
  - Course **difficulty** is **monotonically decreasing** from the original (11 September) baseline.
- Make a pull request if you notice any errors or have suggestions.

Lecture Slides

`https://github.com/christophergandrud/Hertie_Collab_Data_Science/tree/master/LectureSlides`

- Accessible as both HTML (recommended) or PDF.
- Make a pull request if you notice any errors or have suggestions.

Seminar to-do

- Objective: setup software
- **Highly recommend**: use your own laptop

Modern Web browser

- Make sure you have a modern web browser, e.g.:
  - Chrome

GitHub

- Set up (free) GitHub account: `https://github.com/join`
- Install GitHub application:
  - Mac: `https://mac.github.com/`
  - Windows: `https://windows.github.com/`

Statistics software

- **Install** software (all software is free):
  - R: `http://cran.rstudio.com/`
  - RStudio (dev build): `http://www.rstudio.org/download/daily/desktop/`
- Make sure that you can install R packages:

```
install.packages('ggplot2')
```

LaTeX

- Install a LaTeX distribution
  - Mac: `https://tug.org/mactex/`
  - Windows: `http://miktex.org/download`
- This is a large download, so maybe do it in your spare time.

Post-Installation

Play around with the softwared (especially RStudio)