MS DATA SCIENCE 597: DATA WRANGLING AND HUSBANDRY
16:960:597:01, Spring 2021

Instructor: Tirthankar Dasgupta
Email: `td370@stat.rutgers.edu` (typical response within 48 hours)

Class meetings: Mondays 6:40 – 9:30 PM

Meeting Format: Zoom, access through Canvas, first meeting on January 25, 2021. Recordings will be available after class. Attendance is encouraged but not mandatory.

Teaching Fellow: Ziyue Wang.

Sections and Office Hours: Instructor's office hours (tentative): Tuesdays 3-4 PM, Thursdays 3-4 PM (Via Zoom, links on Canvas), or by appointment.

Text:

- R for Data Science, Garrett Grolemund & Hadley Wickham, O'Reilly, http://r4ds.had.co.nz/
- Data Wrangling with R, Bradley C. Boehmke, Springer, https://catalog-libraries-rutgers-edu.proxy.libraries.rutgers.edu/vufind/Record/ 5725290

Note: the first text is available for free on line, while the second is available online to Rutgers students. For convenience of everyone, a pdf version is being uploaded on the course website.

Topics:

- Introduction to the course, R, and RStudio
- Report writing using R Markdown
- Basic data management in R
- Introductory data visualization
- Data manipulation and the split-apply-combine paradigm
- String manipulation
- Text analysis
- Writing R functions
- Project organization, including the use of github
- Getting data off the web
- Basic statistical analysis
- Basics of writing R packages

Course Work and Grading There will be weekly graded assignments and a final project. The homework will collectively count towards 75% of the grade and the final project the remaining 25%. There will be no exams.

Homework submission: Homework will be submitted on Canvas. Instructions on the submission format (e.g., R markdown format) will be provided in the Assignments. Homeworks will be posted every Monday (starting January 25) and typically be due the following Monday at 4:30 PM (with a half an hour grace period).

Late Homework: You will get ONE automatic extension of 72 hours for the semeter's homework. Homework will typically be due every Monday (starting Feb 1 at 4:30 pm), so the extension would go to Thursday at 4:30 pm. With the understanding that issues with technology and internet connections may arise when trying to submit homework, there is a 30 minutes grace period built into the due date. That is, homework assignments are due at 4:30 pm but will not be counted as late until after 5:00pm. Anything submitted after the 5:00 pm cut-off will count as your late extension. If you need more than one extension (due to illness, power outage, or any other reason), please get in touch with the instructor so we can work something out.

Homework collaboration policy: You are welcome to discuss the homework problems with others, but you must write up your codes and solutions yourself. Additionally, you must list the names of the students with whom you collaborated (if any). Copying someone else's solution, or just making trivial changes for the sake of not copying verbatim, is not acceptable and violates Rutgers University's academic integrity policy.

Specific norms for Zoom:

- Classes will be recorded and posted on Canvas. These recordings are only for members of this class.
- Do attend class, if you can.
- Do turn on your camera, if you are able and comfortable.
- Mute yourself once class begins, to avoid background noise.
- Do answer questions, ask questions, and provide feedback.
- We are all learning, please be patient with each other.

ACADEMIC INTEGRITY: Rutgers University takes academic dishonesty very seriously. By enrolling in this course, you assume responsibility for familiarizing yourself with the Academic Integrity Policy. The University's academic integrity policy can be found at

`http://nbacademicintegrity.rutgers.edu/home/academic-integrity-policy/`