

MA213: Lab 1 Logistics

Module 1: Exploratory Data Analysis and Study Design

OpenIntro Statistics, 4th Edition

Based on slides developed by Mine Çetinkaya-Rundel of OpenIntro.
The slides may be copied, edited, and/or shared via the CC BY-SA license.
Some images may be included under fair use guidelines (educational purposes).

Lab Time and Location

- *C1 Lab:*
 - Friday 9:05 AM - 9:55 AM : CA 213
- *C2 Lab:*
 - Friday 10:10 AM - 11:00 AM : COM 215
- *C3 Lab:*
 - Friday 12:20 PM - 1:10 PM : CDS 264
- *C4 Lab:*
 - Friday 1:25 PM - 2:15 PM : COM 215

People

- *Instructor*: Prof. Emily Stephen
- *Labs*: Dr. Yongho Lim
- *Graduate Teaching Fellows (TFs)*:
 - Matt Broe (Discussions)
 - James Zheng Yang (Discussions)
 - Matt Ludwig (Labs)
- *Undergraduate Learning Assistants (LAs)*:
 - Yao Lu (Labs, Lecture)
 - Jack Hincks (Labs, Lecture)

See the course website for office hours and contact info

Course Webpage

- *Course website (Blackboard):* <https://learn.bu.edu>
(Log in with your BU credentials)
- *What's there:*
 - Announcements
 - Course documents: **Syllabus** (with Office Hours and Calendar), GenAI Policy, Learning Objectives
 - Links to Textbook, Edfinity, Gradescope
 - Course Forum
 - Lecture slides and videos
 - Lab materials
 - Your gradebook
 - Gradescope for post-lab activity questions

Structure of Labs

Led by Yongho Lim, with support from TFs and LAs

- *Weekly labs* on Fridays, starting this week
- Practice analyzing data and running simulations in R
- *Skills labs*
 - Work in pairs
 - Analyze data and run simulations in R
 - Post-lab questions to be submitted on Gradescope
- *Lab projects*
 - Work in groups to explore real data
 - Project 1: Data analysis video presentation
 - Project 2: Statistical report

Assignments in Lab

- 7 *Skills Labs*
 - Post-lab activity (R file) to be submitted on Gradescope
 - Due by 10 PM every Thursday, starting this week
 - You can submit as many times as you want before the deadline
 - You can work with your friends, but each person must submit their own work
 - You can use Generative AI tools for **specific uses** with **proper documentation** (see syllabus and GenAI policy document)
- 2 *Lab Projects*
 - Project 1: Data analysis video presentation
 - Project 2: Statistical report
 - Both projects will have deliverables (Outline, Progress Report with R file, Final Submission)
 - Both projects will be group projects (groups of 4)

Grading Structure

	A	B	C
Homeworks	12/12 complete	11/12 complete	10/12 complete
Quizzes (Core)	18/19 passed	15/19 passed	12/19 passed
Quizzes (Aux)	9/11 passed	6/11 passed	0/11 passed
Skills Labs	7/7 passed	Labs 1-6 passed	Labs 1-6 passed
Lab Projects	2 satisfactory	2 satisfactory	2 satisfactory
Lectures	>34 attended	>28 attended	>20 attended

Additional factors:

- Grades between letter grades will be determined by how close you are to the next letter grade
- Each “unsatisfactory” project will drop your course grade by a third of a letter grade (e.g. B becomes B-)

Class Policies

- *Attendance and participation:*
 - Expected to attend all lectures, labs, and discussions.
 - In-class activities and participation will be part of your grade.
 - Email the instructor or your TF if you need to miss class/discussion/lab
- *Academic integrity:*
 - All work must be your own.
 - Collaboration is allowed on homeworks and labs/projects, but **not on quizzes**.
 - See syllabus for details.
- *Use of Generative AI tools (e.g. ChatGPT):*
 - Quizzes: **Not allowed**.
 - Homeworks: Use at your own discretion
 - Labs/Projects: Allowed for **specific uses** with **proper documentation**.
 - See syllabus and GenAI policy document for details.

- **Acceptable Uses** include:
 - Loading data
 - Cleaning, formatting and organizing data tables
 - Generating figures with ggplot
 - Understanding error messages
- **Proper Documentation** include:
 - AI-written code with comments
 - Examples of prompts used to generate code

GenAI Good Examples

Question : Import “mtcars.csv” to the variable carsdata.
Create a new variable subdata containing only the data for automatic transmission cars with weight over 3000 lbs.

- Your Prompt:

```
Give me R code for printing all the rows with automatic transmission cars  
(am = 0) and weight over 3000 lbs (wt>3) from the mtcars dataset.
```

- GenAI Response:

```
Certainly! Here is the R code to print all rows from the mtcars dataset  
where the transmission is automatic (am == 0) and the weight (wt)  
is greater than 3:
```

Code generated by GenAI for the above prompt:

```
# Load the mtcars dataset
data(mtcars)

# Filter rows with automatic transmission (am == 0) and weight > 3
result <- mtcars[mtcars$am == 0 & mtcars$wt > 3, ]

# Print the filtered rows
print(result)
```

GenAI Bad Examples

Question : Import “mtcars.csv” to the variable carsdata.
Create a new variable subdata containing only the data for automatic transmission cars with weight over 3000 lbs.

Your Prompt:

```
Import mtcars.csv to the variable carsdata.  
Create a new variable subdata containing only the data for automatic  
transmission cars with weight over 3000 lbs.
```

GenAI Response:

```
# Import mtcars.csv into the variable carsdata  
carsdata <- read.csv("mtcars.csv")  
  
# Create subdata with automatic transmission cars (am == 0)  
# and weight over 3000 lbs (wt > 3.0)  
subdata <- subset(carsdata, am == 0 & wt > 3.0)
```