Lab 1

STAT 109: Introductory Biostatistics

Lab 1

DATA 322 — Lab 1

Reviewing Data Skills for Machine Learning

Due: Friday, January 30, 2026 at 11:59 PM (Canvas)
Format: Google Colab notebook (template provided)

Purpose of This Lab

This first lab is designed to help you **review data wrangling and visualization in Python** in preparation for machine learning workflows you will use throughout the course. You will review key skills from DATA 271, practice reading and navigating a computational notebook, and begin building a **personal quick reference** you can return to all semester.

This lab is **not** about learning new machine learning algorithms. Instead, the focus is on:

- Translating information into data a machine can use
- Understanding how data is structured for the libraries we'll be using
- Reviewing essential wrangling and visualization skills

In-Lab Activity Overview

During lab, you will work with a partner and complete the following activities:

1. Meet a classmate and think like a model
2. Skim the lab in Ch. 2 from the textbook
3. Begin a personal "Quick Reference Guide"

You are **not expected to finish the full assignment during lab**. Lab time this week is just intended to get you started.

Part 1: Thinking Like a Model

Working with a partner, come up with **one question** you could ask your instructor to help you **predict** their age.

After hearing responses and predictions, we will discuss how to move from real-world information to data a model can use.

Part 2: Skimming the Lab in Ch. 2

You will use a lab from the course textbook as a reference:

Textbook lab:
https://github.com/intro-stat-learning/ISLP_labs/blob/stable/Ch02-statlearn-lab.ipynb

Click the **"Open in Colab"** link at the top of the page.

Your goal is to *skim*, not to understand everything.

As you skim, look for the following:

- How data is loaded
- How object types and shapes are checked
- How a design matrix `X` and response vector `y` are created

- How plots are generated

Part 3: Quick Reference Guide (Start in Lab, Finish at Home)

Create a new notebook in Colab. Your job is to begin building a **personal Quick Reference Guide** for machine learning tasks.

Your guide should include short explanations *in your own words* and example code snippets for the following topics:

Required Sections

1. Loading Data from a .csv

In Colab, you'll have to upload the dataset first using the file manager on the left side bar.

2. Checking Objects and Data Structures

Examples may include:

- Checking object types
- Viewing dimensions or shape
- Inspecting column names

3. Creating a Design Matrix (`X`)

Examples may include:

- Selecting predictor variables
- Ensuring all values are numeric
- Confirming dimensions are appropriate

4. Defining a Response (`y`)

Examples may include:

- Extracting a single variable as a response and checking that it is numeric
- Ensuring shape is appropriate for an `X`

5. Descriptive Statistics

Examples may include:

- measures of center and spread
- histogram
- bar plot
- stacked bar plot
- scatterplot
- boxplot

This quick reference is meant to be useful to **you** later in the course. Clarity and usefulness matter more than length.

What to Submit

Submit your completed work Canvas. Your submission should include:

- Your completed **Quick Reference Guide**
- Brief written responses to the reflection prompt in Canvas

Grading

This lab is graded on **completion and a reflection on what you learned**, not on correctness or aesthetics.