MATH113/CAIS105: Intro to Data Science

Fall 2023

# Homework 05

Homework is DUE before class on the day indicated on the course schedule.

**Learning Objectives:**

* **Practice making Tidy data**

## Overview

Complete this assignment in an R Markdown file.

When you answer the questions below, be sure to include your code and a written answer in your R Markdown file. For example, if I were answering the question: “What were the most popular baby names in the 1990s”, my R Markdown report would look something like:

A screenshot of a computer

Description automatically generated

## Part 1

The HELPfull data within the mosaicData package contains information about the Health Evaluation and Linkage to Primary Care (HELP) randomized trial in tall format.

1. Generate a table of the data for subjects (ID) 1, 2, and 3 that includes the ID variable, the TIME variable, and the DRUGRISK and SEXRISK variables (measures of drug and sex risk-taking behaviors, respectively).
2. The HELP trial was designed to collect information at 0, 6, 12, 18, and 24 month intervals. At which timepoints were measurements available on the \*RISK variables for subject 3?
3. Let’s restrict our attention to the data from the baseline (TIME = 0) and 6-month data. Create a table that looks like the following:

A white background with black text

Description automatically generated

## Part 2

1. Consider the number of home runs hit (HR) and home runs allowed (HRA) for the Chicago Cubs (CHN) baseball team. Reshape the Teams data from the Lahman package into “long” format and plot a time series conditioned on whether the HRs that involved the Cubs were hit by them or allowed by them.
2. Write a function called count\_seasons that, when given a teamID, will count the number of seasons the team played in the Teams data frame from the Lahman package

## Submission

Knit your R Markdown file to a PDF and submit through PLATO.