

## DSCI401 - Homework 2

**Due: September 22, 2024**

Homework should be submitted as an R Markdown file with links to Google colab notes where necessary. Homework should be turned in on Sakai.

1. (50 points) Create an R Markdown file and run the following code:

```
cubs <- read.csv("https://raw.githubusercontent.com/menawhalen/DSCI_401/main/data/cubs_all_time.csv")
```

Answer the following questions using R:

- (a) How many total games have the Cubs won and lost between 1876 and 2022?
  - (b) What's the most number of ties had in a season and which years were those?
  - (c) What year did the Cubs score the most runs? What year did the Cubs score the fewest runs? Do you have any thoughts about the year that the Cubs scored the fewest runs?
  - (d) In how many seasons was the Cubs total attendance (i.e. the variable Attendance) over 3 million?
  - (e) Using the variable Tm how many unique team names are there and how many years was each name used?
2. (50 points) Create a google colab notebook. Run the following code and answer the same question as above using python.

```
import pandas as pd
cubs = pd.read_csv("https://raw.githubusercontent.com/menawhalen/DSCI_401/main/data/cubs_all_time.csv")
```

Answer the following questions using Python:

- (a) How many total games have the Cubs won and lost between 1876 and 2022?
- (b) What's the most number of ties had in a season and which years were those?
- (c) What year did the Cubs score the most runs? What year did the Cubs score the fewest runs? Do you have any thoughts about the year that the Cubs scored the fewest runs?
- (d) In how many seasons was the Cubs total attendance (i.e. the variable Attendance) over 3 million?
- (e) Using the variable Tm how many unique team names are there and how many years was each name used?