# Homework (HW05) Exploring DataFrames

### General Instructions

For this homework you will upload 1 R file into blackboard.

Reminder:

* All HW must start with an Identification Block like this sample…

################################################

# IST 387/687, Standard Homework Heading

#

# Student name:

# Homework number:

# Date due:

#

# Attribution statement: (choose the statements that are true)

# 1. I did this work by myself, with help from the book and the professor

# 2. I did this work with help from the book and the professor and these Internet sources: <provide the urls>

# 3. I did this work with coaching from <Name of another student> but did not cut and paste any code

# Run these three functions to get a clean test of homework code

dev.off() # Clear the graph window

cat('\014') # Clear the console

rm(list=ls()) # Clear all user objects from the environment!!!

# Set working directory

# Change to the folder containing your homework data files

setwd("~/MyDesktop/ISTX87/Homework")

* This homework builds on our JSON and List efforts from the Prep Exercise and depends on a careful read of Chapter 11 of *An Introduction to Data Science*.

### HW05

**Step 1: Understand & Document R code**

1. Document the following R code, with a comment for each line and explain, via a comment, what question the code answers

df$vehicle\_count <- as.numeric(df$vehicle\_count)

value <- df %>%

filter(str\_trim(day\_of\_week)=="THURSDAY") %>%

pull(vehicle\_count) %>%

mean(na.rm=TRUE)

value <- mean( df$vehicle\_count[str\_trim(df$day\_of\_week)=="THURSDAY "], na.rm=TRUE)

**Step 2: Investigating the data frame by answering the following questions with explicit R code, and make sure you handle NAs (in any way you think is appropriate):**

1. What was the total number of accidents with injuries?
2. How many accidents happen on Friday?
3. What is the total number of accidents on Friday where injury=’YES’?
4. What is the total number of accidents on Friday where injury =’NO’?
5. How many injuries occurred each day of the week?
6. Create a new data frame that only includes accidents on Friday.
7. What is the mean number of vehicles involved in accidents on Friday?
8. Use hist( ) to make a histogram of the number of vehicles in accidents on Friday.
9. What is the distribution of the number of vehicles in accidents on Sunday (use a histogram and quantile)
10. Add a block comment explaining how the distribution on Sunday compares with the distribution of the number of vehicles in accidents on Friday

***You must submit all Homework to blackboard prior to the deadline specified for each assignment.***

Late HW assignments will not be accepted for credit.

*Materials provided by the instructors of this course are copyrighted 2018 by Jeffrey Saltz and Jeffrey Stanton, or by the respective instructor that produced them. These materials are provided to enrolled students for their exclusive use during the semester of instruction. Uploading instructor-produced materials to Internet sites (e.g., Coursehero) without permission is considered a breach of professional ethics and will be reported as a possible academic integrity violation.*