

## Homework 3

Prepare your answers as a **single PDF file**.

**Group work:** You may work in groups of 1-3. Include all group member names in the PDF file. Only **one person** in the group should submit to Canvas.

**Due:** check on Canvas.

1. The following questions use the data collected from the anonymous survey collected at the beginning of the course and from previous semesters. The dataset can be downloaded from the Datasets module on Canvas. Load the survey data into a variable called “survey”. Use the function **read\_csv()** that is part of the tidyverse library (instead of the built-in function

```
read.csv())  
# install tidyverse  
library(tidyverse)  
> survey <- read.csv("Downloads/surveydataFall2023.csv")
```

Apply the tidyverse’s data wrangling verbs to answer these questions. For each question, **give only the code (as one data pipeline with multiple “verbs” one after the other) beginning with survey %>% ....**

a. Show only rows where Statistics skill is 10

```
> survey %>% filter(Statistics == 10)
```

b. Show only rows where both Statistics and Math skills are 10

```
> survey %>% filter(Statistics == 10, Math == 10)
```

c. Show only the Semester, Communication, and Visualization columns where both Statistics and Math skills are 10

```
> survey %>% filter(Statistics == 10, Math == 10) %>%  
select(Semester, Communication, Visualization)
```

d. Show only the Semester, Communication, and Visualization columns in decreasing order of Statistics and then Math skills

```
> survey %>% arrange(desc(Statistics), desc(Math)) %>%  
select(Semester, Communication, Visualization)
```

e. Show only rows where Statistics skill is missing

```
> survey %>% filter(is.na(Statistics))
```

- f. Show for every student, only their Semester, Math skill, Statistics skill, and the maximum of their Math and Statistics skills (Hint: mutate)

```
> survey %>% mutate(Max_Skill = pmax(Math, Statistics)) %>%  
select(Semester, Math, Statistics, Max_Skill)
```

- g. Show the median value of Computer Science skills

```
> survey %>% summarize(Median_CS = median(ComputerScience, na.rm =  
TRUE))
```

- h. Show the median value of Computer Science skills in Fall 2023

```
> survey %>% filter(Semester == "Fall2023") %>% summarize(Median_CS =  
median(ComputerScience))
```

- i. Show the median values of (each of) Computer Science, Math, and Statistics skills in every semester

```
> survey %>% group_by(Semester) %>% summarize(Median_CS =  
median(ComputerScience, na.rm = TRUE), Median_Math = median(Math,  
na.rm = TRUE), Median_Statistics = median(Statistics, na.rm = TRUE))
```

- j. Show the median values of (each of) Computer Science, Math, and Statistics skills in every semester of students who have taken CPSC483 (TakenCPSC483 is Yes)

```
> survey %>% filter(TakenCPSC483 == "Yes") %>% group_by(Semester) %>%  
summarize(Median_CS = median(ComputerScience, na.rm = TRUE),  
Median_Math = median(Math, na.rm = TRUE), Median_Statistics =  
median(Statistics, na.rm = TRUE))
```

- k. Show the median values of (each of) Computer Science, Math, and Statistics skills of students for students who have taken and not taken (Hint: group\_by)

```
> survey %>% group_by(TakenCPSC483) %>% summarize(Median_CS =  
median(ComputerScience, na.rm = TRUE), Median_Math = median(Math,  
na.rm = TRUE), Median_Statistics = median(Statistics, na.rm = TRUE))
```

- l. Show the number of students who took the survey in every semester

```
> survey %>% group_by(Semester) %>% summarize(Count = n())
```