DataAssignment\_5

Md. Saifur Rahman

2/20/2021

# **Task-1: Use at least 2 levels of headings (Heading Level 1)**

## \* Setup Working Directory (heading Level 2)

# Setup working directory for the assignment  
setwd("C:/Users/mdsaifur.rahman.1/Desktop/Spring 2021/EFR 535/Code R")

# **Task-3: Mindfully use chunk options (e.g., include = FALSE) to show only necessary code and output.**

# **Task-4: Integrate white space and paragraphs to separate the components of your output**

# **Task-5: Use at least one instance of bold, italics, and bold + italic**

# **Task-6: Make at least one comment inside a code chunk**

## To Load packages do the following. I am loading two packages here.

1. ‘tidyverse’ package
2. ‘tibble’ package

### This is a separate paragraph. I am using white space to make a separate paragraph. Besides, to get an extara space in between paragraph I am using this symbol ‘$~$’ with a white space in before and after this symbol.I am using *“echo = TRUE, results = ‘hide’”* in the code chunk options because I don’t want to see the output portion for this code chunk in my word document. Beside, To get rid of warning, error and message; I am using false for these options in code chunk like (***error=FALSE, warning=FALSE,message=FALSE***). I am using ‘\*' to make italic,’\*\*' to make bold, and ’\*\*\*’ in this paragraph to make those words bold and italic. I am also using ‘#’ symbol to write a comment inside code chunk.

# install and load "tidyverse" package to work with ggplot2  
library(tidyverse)  
#load "tibble" package to see data in tibble format!  
library(tibble)

## \* Load data in R studio from excel file, view the data and attach the dataset in the R studio environment for further use.

library(readxl)  
Data <- read\_excel("EFR 535 Assign 2 - Data Tophat - Jan 18 2018.xls",   
 na = "NA")  
View(Data)  
attach(Data)

## \* Show the the Tophat data set with a tibble

#Examine the Tophat data set with a tibble  
Data

## # A tibble: 129 x 14  
## ID Program\_of\_Study Program\_of\_Stud~ Year Tophat\_particip~  
## <dbl> <chr> <chr> <chr> <dbl>  
## 1 1 Pre-Nursing Pre-Nursing Soph~ 78.8  
## 2 2 General Studies Other Soph~ 78.0  
## 3 3 Pre-Nursing Pre-Nursing Soph~ 99.8  
## 4 4 Pre-Nursing Pre-Nursing Soph~ 72.6  
## 5 5 Pre-Nursing Pre-Nursing Soph~ 96.7  
## 6 6 Pre-Nursing Pre-Nursing Soph~ 96.9  
## 7 7 Pre-Nursing Pre-Nursing Soph~ 92.3  
## 8 8 General Studies Other Seni~ 77.6  
## 9 9 Pre-Nursing Pre-Nursing Soph~ 88.2  
## 10 10 General Studies Other Soph~ 90.0  
## # ... with 119 more rows, and 9 more variables: Tophat\_total\_Qs\_answered <dbl>,  
## # Tophat\_Qs\_Correct <dbl>, Tophat\_Qs\_Incorrect <dbl>,  
## # Tophat\_Qs\_percentcorrect <dbl>, Exam\_1\_outof60 <dbl>, Exam\_2\_outof60 <dbl>,  
## # Exam\_3\_outof60 <dbl>, Exam\_4\_outof100 <dbl>, Final\_Grade\_percent <dbl>

# same previous task but using as\_tibble() function from "tibble" package  
#as\_tibble(Data)

## \* Average and range of Tophat participation percent column

# Average and range of Tophat participation percent column  
mean(Tophat\_participation\_percent)

## [1] 86.91295

range(Tophat\_participation\_percent)

## [1] 24.13 99.81

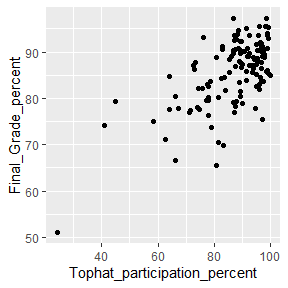
# Average and range of Final Grade percent  
mean(Final\_Grade\_percent)

## [1] 85.6114

range(Final\_Grade\_percent)

## [1] 51.02 97.33

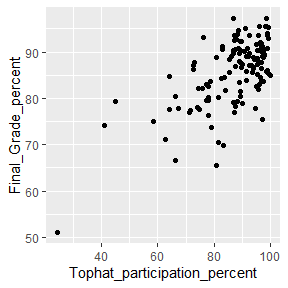
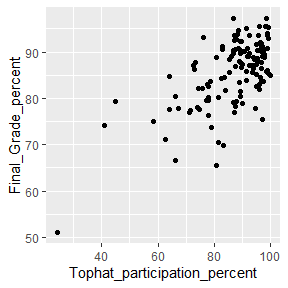
## \* A Scatter plot comparing Tophat participation percent ( in x axis) and Final Grade percent (in y axis)



## **Task-2: Put the questions in text outside of code chunks. You may wish to put any direct answers to assignment questions outside code chunks as well.**

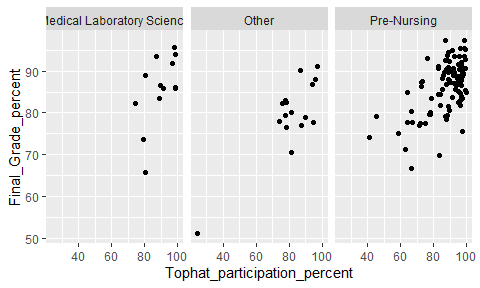
## \* Did you see much difference between the non-jittered and jittered scatterplots? What does that mean?

### Answer: There is a little difference in positioning of the dots in the scatter plot after adding “jitter” that is ignorable.This means the position of these values are mostly unique. There are a few values with the same or almost same position in this coordinate system with the given data. Most have different positioning values. The left plot is without using ‘jitter’ option and the right plot is using ‘jitter’ option.



## \* Create small multiples of the scatterplot based on Program of Study2 variable. Include a comment on what you observed.

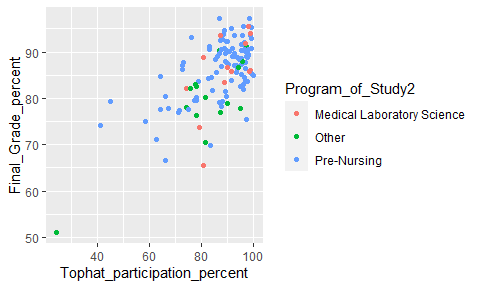
### Answer: This facet\_warp() divide the single scatterplot and present the same data in a group of scatter plots based on number of groups find in the “Program\_of\_Study2” variable. In the Program\_of\_Study2 variable there are 3 groups (i) Medical Laboratory Science, (ii) Pre-Nursing, and (iii) Other. Most students are from the Pre-nursing group and their final grade percent increases according to the increase in the Tophat participation percent.



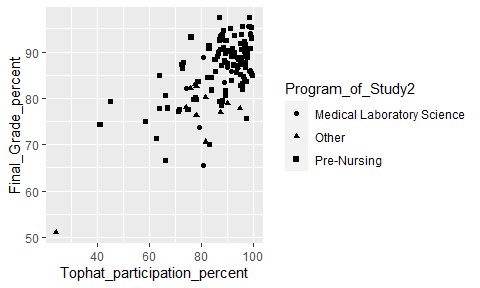
## \* Examine Program of Study2 by including the variable as an aesthetic based on EACH of color, shape, size, and transparency. Which aesthetics did you prefer?

### Answer: From this four aesthetics my preference is “color” (when work with one aesthetic). Color makes this scatterplot more presentable. This also help me to differentiate the type of color(which color means what) by providing a legend.

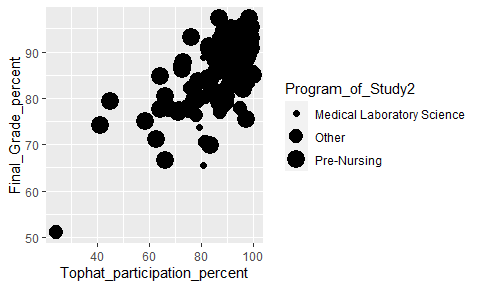
# based on color  
ggplot(data = Data) + geom\_point( mapping = aes(x = Tophat\_participation\_percent, y = Final\_Grade\_percent, color = Program\_of\_Study2))



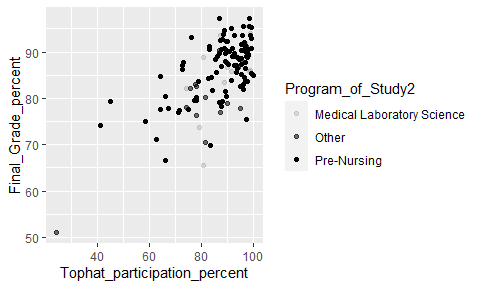
# based on shape  
ggplot(data = Data) + geom\_point( mapping = aes(x = Tophat\_participation\_percent, y = Final\_Grade\_percent, shape = Program\_of\_Study2))



# based on size  
ggplot(data = Data) + geom\_point( mapping = aes(x = Tophat\_participation\_percent, y = Final\_Grade\_percent, size = Program\_of\_Study2))

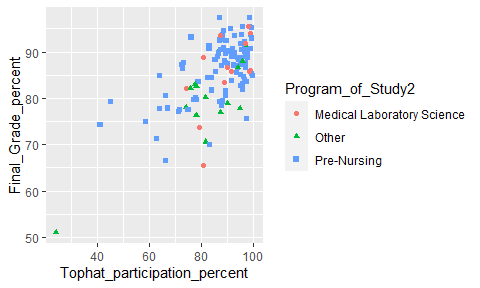


# based on transparency  
ggplot(data = Data) + geom\_point( mapping = aes(x = Tophat\_participation\_percent, y = Final\_Grade\_percent, alpha = Program\_of\_Study2))



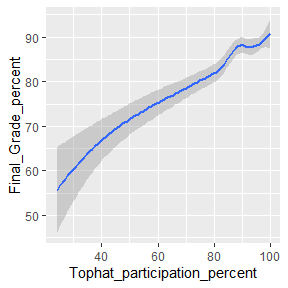
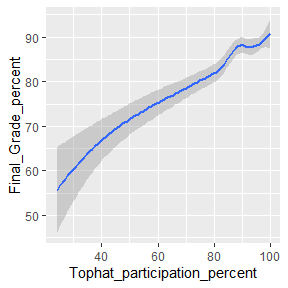
## \* Choose your favorites to make one plot with two aesthetics applied together

### Answer: Sometimes multiple aesthetics make a graph more presentable. I am using the “color” and “shape” aesthetics here. Different color and different shape to plot the data, make this graph more understandable and presentable.



## \* Add a smooth line. See if you can make your code efficient by using a global mapping of the x- and y-axis.

### The left plot present data in a graph by using a smooth line. Make the same code more efficient by using a global mapping of the x- and y-axis in the right plot.



## \* Select a new theme for your graph

### I prefer to use classic theme because this is clean!

