DataAssignment\_11

Md. Saifur Rahman

4/5/2021

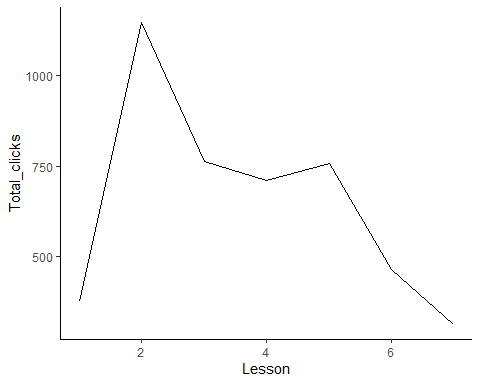
# Task-1: Download the EFR 535 BB data. Review the data and transform it as necessary into data you can use to complete this assignment.

## Answer: I downloaded the EFR 535 BB data file from the lesson 12. I investigate this excel file and finally transform the information from this excel file into a new file named “Necessary data from EFR 535 - 535 Bb data - April 26 2019.xlsx”. I organize data based on the requirements written in this Data Assignment 11. I create 5 columns named (i) Lesson, lesson 1 to lesson 7 is listed in this column; (ii) Total\_clicks, total number of clicks in each lesson is added together and listed here; (iii) Video\_clicks, total number of video clicks is added together and listed against each lesson here; (iv) Data\_asignment, the average of data assignment score from each lesson is listed in this column (Lesson 1 doesn’t has this score); (v) Quiz, the average of quiz score is listed here (Lesson 1 doesn’t has this score). Thus I transform the information from the given file and make this new excel file to use in this assignment.

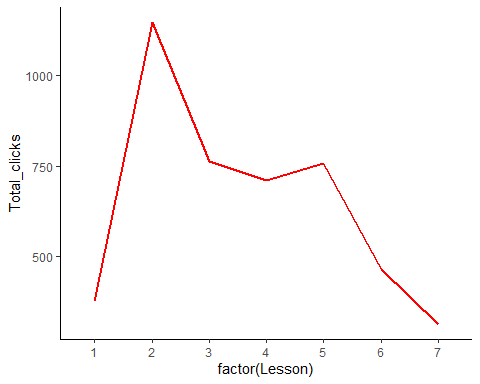
# Load data from excel file  
  
library(readxl)  
Data <- read\_excel("C:/Users/mdsaifur.rahman.1/Desktop/Spring 2021/EFR 535/Code R/Necessary data from EFR 535 - 535 Bb data - April 26 2019.xlsx",   
 na = "NA")  
View(Data)  
attach(Data)

# Task-2: Visualize the total number of clicks over the course of the seven lessons. Create two unique charts (one line chart, one connected scatterplot) across which you uniquely modify the aesthetics.

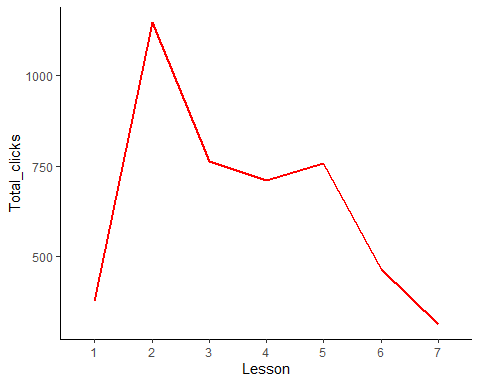
# Visualize the total number of clicks over the course of the seven lessons using line graph  
ggplot(Data, aes(Lesson,Total\_clicks))+ geom\_line()



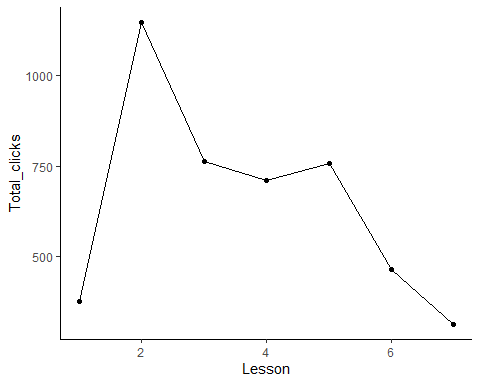
# Convert 'Lesson' from number into factor to visualize each lesson in the line graph. Use group=1 to separately identify each group. Use size and color attributes to make this visually appealing.   
ggplot(Data, aes(factor(Lesson),Total\_clicks, group=1))+ geom\_line(color= "red", size = 1)



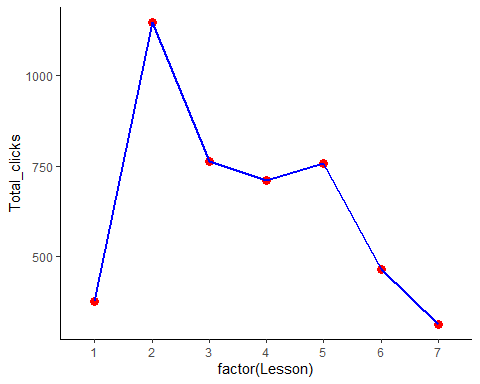
# Use xlab() to rename x label in this graph.  
ggplot(Data, aes(factor(Lesson),Total\_clicks, group=1))+ geom\_line(color= "red", size = 1)+xlab("Lesson")



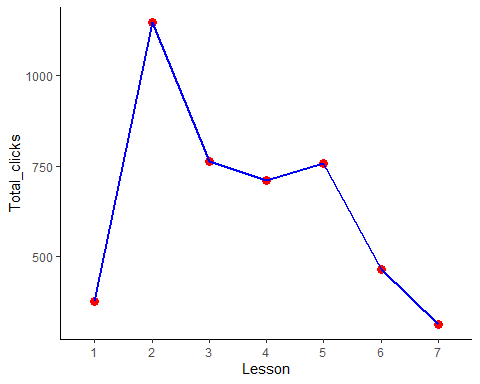
# Visualize the total number of clicks over the course of the seven lessons using connected scatterplot  
ggplot(Data, aes(Lesson,Total\_clicks))+ geom\_point()+geom\_line()



# Convert 'Lesson' from number into factor to visualize each lesson in the line graph. Use group=1 to separately identify each group. Use size and color attributes to make this visually appealing.   
ggplot(Data, aes(factor(Lesson),Total\_clicks, group=1))+ geom\_point(size = 3, color="red")+geom\_line(color= "blue", size = 1)



# Use xlab() to rename x label in this graph.  
ggplot(Data, aes(factor(Lesson),Total\_clicks, group=1))+ geom\_point(size = 3, color="red")+geom\_line(color= "blue", size = 1) + xlab("Lesson")

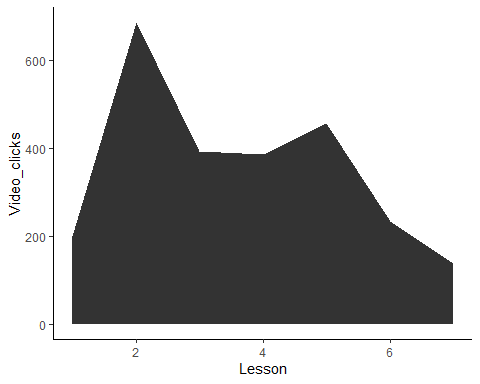


# Task-3: How did the number of video clicks change over the semester across the lessons? Present the visualization as an area chart. Keep in mind, there were multiple videos for each lesson.

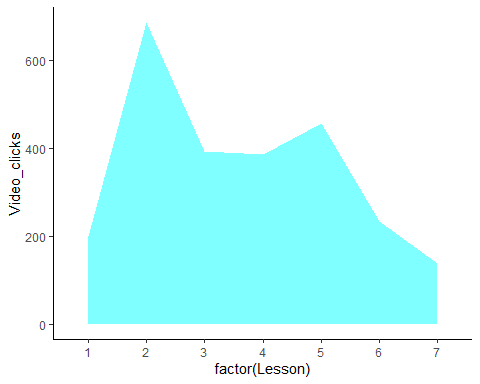
## How did the number of video clicks change over the semester across the lessons?

## Answer: The number of total video clicks increased from 1st lesson. During the 2nd lesson this video clicks was highest with more than 1000 clicks. After that, there was a declining trend throughout and until the lesson 7, except a raise in lesson 5.

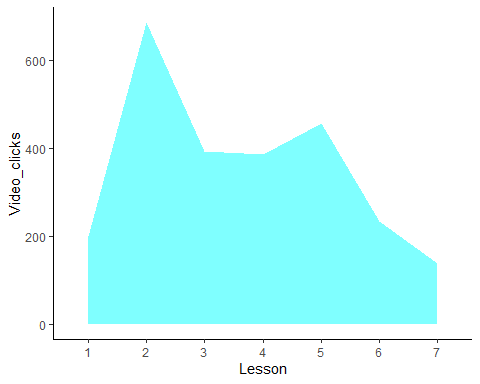
# Visualize the number of video clicks change over the semester across the lessons using area chart  
ggplot(Data, aes(Lesson,Video\_clicks))+ geom\_area()



# Convert 'Lesson' from number into factor to visualize each lesson in the line graph. Use group=1 to separately identify each group. Use fill (fill the area with color) and alpha attributes to make this graph visually appealing.   
ggplot(Data, aes(factor(Lesson),Video\_clicks, group=1))+ geom\_area(fill = "cyan", alpha = 0.5)

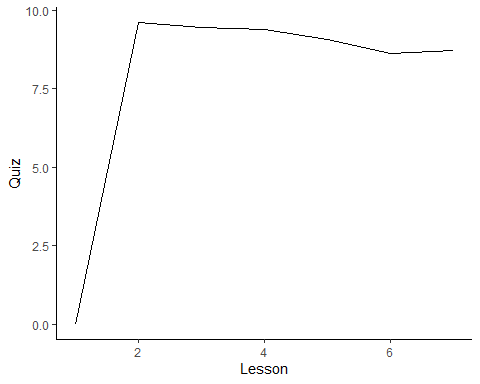


# Use xlab() to rename x label in this graph.  
ggplot(Data, aes(factor(Lesson),Video\_clicks, group=1))+ geom\_area(fill = "cyan", alpha = 0.5)+ xlab("Lesson")

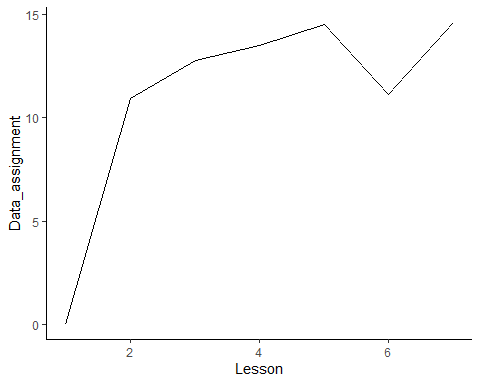


# Task-4: Did quiz and data assignment scores follow unique trends over time? Create plots for each of these variables separately. Add smoothed lines to represent the trends for each variable on each plot. Then, take the appropriate steps to plot them on the same graph. BONUS: Add smoothed lines to represent the trends for each variable on the combined plot.

# Visualize trends on quiz score over the course of the seven lessons using line graph  
ggplot(Data, aes(Lesson,Quiz))+ geom\_line()

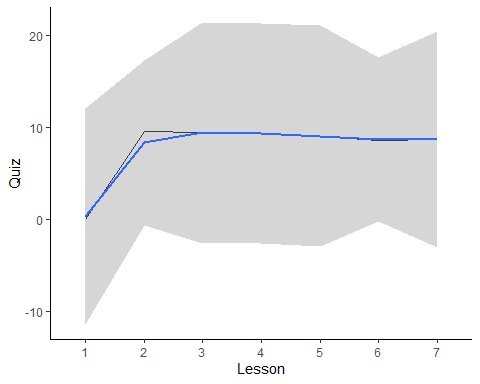


# Visualize trends on data assignment score over the course of the seven lessons using line graph  
ggplot(Data, aes(Lesson,Data\_assignment))+ geom\_line()



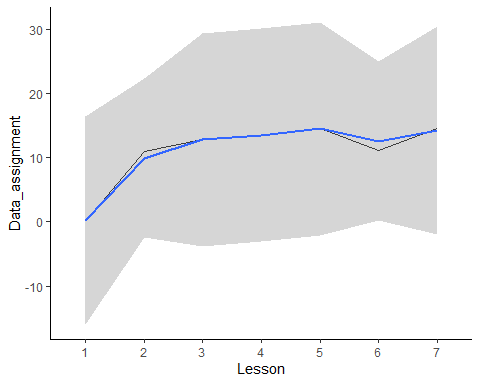
# Add smoothed lines to represent the trends for each variable on each plot.Convert 'Lesson' from number into factor to visualize each lesson in the line graph. Use group=1 to separately identify each group. Use xlab() to rename x label in this graph.  
ggplot(Data, aes(factor(Lesson),Quiz, group = 1))+ geom\_line()+ geom\_smooth() + xlab("Lesson")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



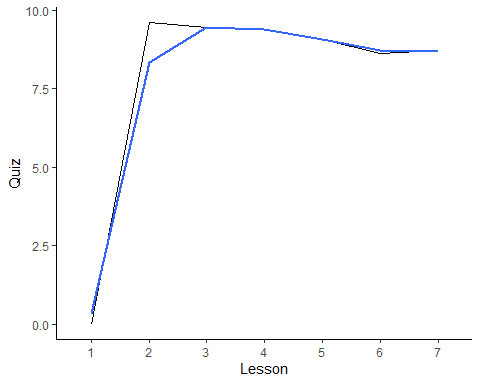
ggplot(Data, aes(factor(Lesson),Data\_assignment, group = 1))+ geom\_line()+ geom\_smooth()+ xlab("Lesson")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



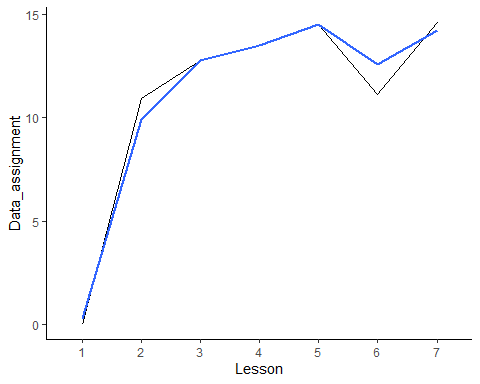
# remove confidence band or standard error  
ggplot(Data, aes(factor(Lesson),Quiz, group = 1))+ geom\_line()+ geom\_smooth(se = FALSE) + xlab("Lesson")

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

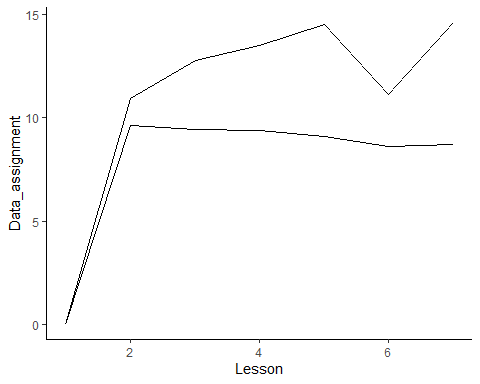


ggplot(Data, aes(factor(Lesson),Data\_assignment, group = 1))+ geom\_line()+ geom\_smooth(se = FALSE) + xlab("Lesson")

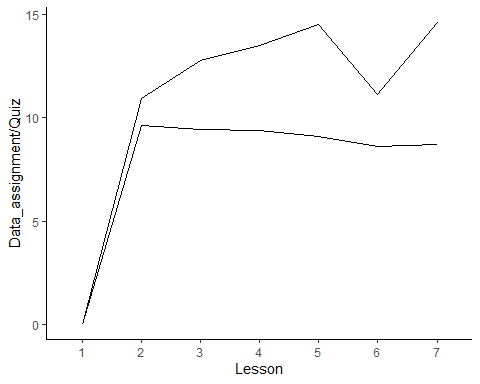
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



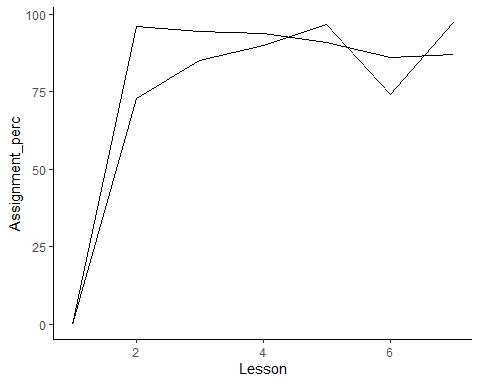
# Visualize trends on quiz score and data assignment score over the course of the seven lessons but in the same plot  
ggplot(Data)+ geom\_line(aes(Lesson,Data\_assignment))+ geom\_line(aes(Lesson,Quiz))



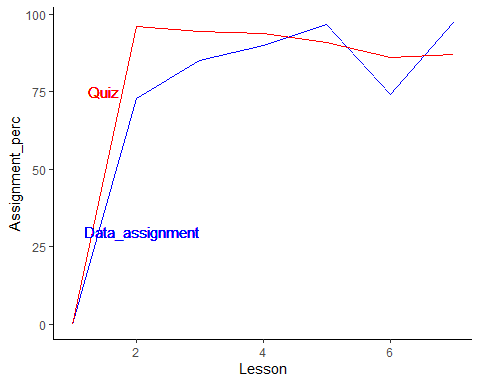
# Convert 'Lesson' from number into factor to visualize each lesson in the line graph. Use group=1 to separately identify each group. Use xlab() to rename x label in this graph. Use ylab() to rename the label in y-axis.  
ggplot(Data)+ geom\_line(aes(factor(Lesson),Data\_assignment, group=1))+ geom\_line(aes(factor(Lesson),Quiz, group =1))+xlab("Lesson")+ylab("Data\_assignment/Quiz")



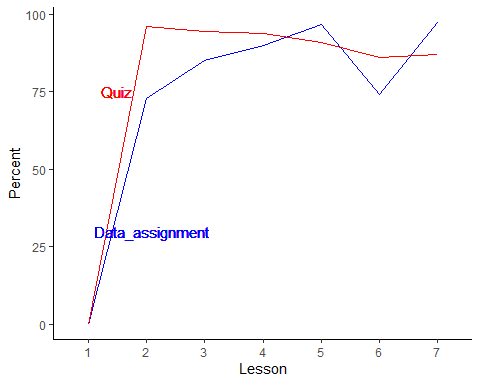
# Still, the scaling is different for these two line graphs! I am rescaling in the following.   
# Make a new data frame with copy of the previous Data frame   
Data\_combine\_plot <- Data  
  
# Investigating the data from the given excel file, I assume that the Quiz is scored by 10 and the data assignment is scored by 15.  
# Rescale in percent convert these two scale into one percent scale. This will help to perform side by side comparison.   
Data\_combine\_plot$Quiz\_perc <- (Data\_combine\_plot$Quiz/10)\*100  
Data\_combine\_plot$Assignment\_perc = (Data\_combine\_plot$Data\_assignment/15)\*100  
  
# Now, plot them in the same graph and scale.   
ggplot(Data\_combine\_plot)+ geom\_line(aes(Lesson,Assignment\_perc))+ geom\_line(aes(Lesson,Quiz\_perc))



# Add text to identify the line and color to make this visually appealing.   
ggplot(Data\_combine\_plot)+ geom\_line(aes(Lesson,Assignment\_perc), color="blue")+ geom\_line(aes(Lesson,Quiz\_perc),color="red")+ geom\_text(aes(2.1,30), label="Data\_assignment", color="blue")+ geom\_text(aes(1.5,75), label="Quiz", color="red")



# Convert 'Lesson' from number into factor to visualize each lesson in the line graph. Use group=1 to separately identify each group.Use ylab() to rename the label of y-axis. Use xlab() to rename the label of x-axis.  
ggplot(Data\_combine\_plot)+ geom\_line(aes(factor(Lesson),Assignment\_perc,group=1), color="blue")+ geom\_line(aes(factor(Lesson),Quiz\_perc, group=1),color="red")+ geom\_text(aes(2.1,30), label="Data\_assignment", color="blue")+ geom\_text(aes(1.5,75), label="Quiz", color="red")+ylab("Percent")+ylab("Percent")+xlab("Lesson")



# BONUS: Add smoothed lines to represent the trends for each variable on the combined plot  
ggplot(Data\_combine\_plot)+ geom\_line(aes(factor(Lesson),Assignment\_perc,group=1), color="darkcyan")+geom\_line(aes(factor(Lesson),Quiz\_perc, group=1),color="red")+geom\_text(aes(2.1,30), label="Data\_assignment", color="darkcyan")+ geom\_text(aes(1.5,75), label="Quiz", color="red")+ylab("Percent")+ylab("Percent")+xlab("Lesson")+geom\_smooth(aes(Lesson,Assignment\_perc), se =F)+ geom\_smooth(aes(Lesson,Quiz\_perc),se=F)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'  
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

