AllDataFinalExplore

2022-06-08

#clean up and set up

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.8  
## v tidyr 1.2.0 v stringr 1.4.0  
## v readr 2.1.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readr)  
library(dplyr)  
library(tidyr)  
#df = read\_csv("AllDataFinalCSV.csv")  
df <- read\_csv("C:/Users/kuipers/Dropbox/MAP\_2022Summer/Analytics/Final Greenhouse Code/CleaningScoringData/AllDataFinalCSV.csv")

## New names:  
## \* `Section 3:Identify your agreement with each of the following: - As long as we can verify that the mathematical calculations are done correctly, we can trust the statistical results.` -> `Section 3:Identify your agreement with each of the following: - As long as we can verify that the mathematical calculations are done correctly, we can trust the statistical results....24`  
## \* `Section 3:Identify your agreement with each of the following: - As long as we can verify that the mathematical calculations are done correctly, we can trust the statistical results.` -> `Section 3:Identify your agreement with each of the following: - As long as we can verify that the mathematical calculations are done correctly, we can trust the statistical results....65`  
## \* `` -> ...90  
## \* `` -> ...91  
## \* `` -> ...92  
## \* ...

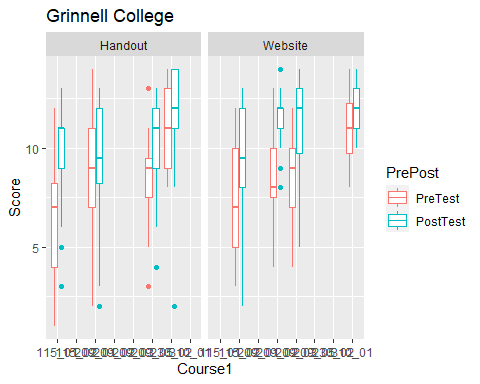
## Rows: 584 Columns: 93  
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (76): Date1, Name1, Pronoun1, Institution1, Instructor1, Course1, Major1...  
## dbl (11): Row, Progress1, Duration1, Residual1, PreTest, PreT, PostT, Progre...  
## lgl (6): Finished1, Finished2, ...90, ...91, ...92, ...93  
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

df <- df %>% mutate\_all(na\_if,"")  
  
NewData = df  
  
NewData = gather(data = df, key = PrePost, value = Score, PreTest, PostTest)  
  
#Cornell was spelt as 'Cornel WITH AN EXTRA SPACE' is some of the rows of Institution2 so it failed some rows. Change is made below  
NewData$Institution2 = ifelse(NewData$Institution2 == "Cornel College", "Cornell College", NewData$Institution2)  
  
NewData$InstitutionMatch = ifelse(NewData$Institution1 == NewData$Institution2, "Success", "Fail")   
NewData$InstructorMatch = ifelse(NewData$Instructor1 == NewData$Instructor2, "Success", "Fail")   
NewData$CourseMatch = ifelse(NewData$Course1 == NewData$Course2, "Success", "Fail")   
  
# A few courses and instructors don't match up.   
  
#renaming some columns  
  
NewData = NewData %>%  
 rename("att1.1" = "I would prefer to avoid courses that use statistics1","att1.2" = "Statistics will help me better understand studies that I read or hear about1", "att1.3" = "Statistics requires critical thinking that helps researchers make better decisions based upon data1", "att1.4" = "Real problems often have obstacles that could be approached in multiple ways1", "att1.5" = "This course will be useful for me in the future1", "att1.6" = "There really is no difference between statistics courses and math courses1")  
  
  
NewData = NewData %>%  
 rename("att2.1" = "I would prefer to avoid courses that use statistics2", "att2.2" = "Statistics will help me better understand studies that I read or hear about2", "att2.3" = "Statistics requires critical thinking that helps researchers make better decisions based upon data2", "att2.4" = "Real problems often have obstacles that could be approached in multiple ways2", "att2.5" = "This course will be useful for me in the future2", "att2.6" = "There really is no difference between statistics courses and math courses2")  
  
NewData = NewData %>%  
 rename("Response2" = "1) What is the response variable?", "EV2" = "2) What is the explanatory variable?", "DPoint2" = "3) What does each point represent?", "SlopeInc2" = "4) Would the slope increase or decrease if we removed the data point for Percent NoseBlack = 3?", "SameSlope2" = "5) Assume that in 2021 the researchers decided to collect data on a new set of lions. If they created a new plot and regression line, would you expect the value of the slope to:", "RegrLine2" = "6) Assume that in 2021 the researchers decided to collect data on a similar set of lions from the same region. Would you expect the slope of the regression line for 2021 data to be:", "Residual2" = "7) Which of the following observed Percent NoseBlack values would have the largest positive residual?", "ModifyData2" = "8) Based on the data shown in the graph above, indicate whether you agree with each research design described below: - The researchers should consider modifying the data (such as removing the data representing Percent NoseBlack = 90) because it would create a better regression model.", "BetterModel2" = "8) Based on the data shown in the graph above, indicate whether you agree with each research design described below: - The researchers should recognize that there may be a better model to explain the patterns in their data.", "Predict2" = "8) Based on the data shown in the graph above, indicate whether you agree with each research design described below: - This graph shows that we can be confident that if Percent NoseBlack = 20, the Age will always be between 2 and 2.5.", "CorrGood2" = "8) Based on the data shown in the graph above, indicate whether you agree with each research design described below: - Since the correlation is 0.89, we can be certain that the prediction from the linear regression model would be reliable when Percent NoseBlack = 94.")  
  
NewData = NewData %>%  
 rename("Know1.1" = "Section 3:Identify your agreement with each of the following: - If a hypothesis test results in a p-value less than 0.05, we can have confidence that the test was properly conducted.", "Know1.2" = "Section 3:Identify your agreement with each of the following: - As long as we can verify that the mathematical calculations are done correctly, we can trust the statistical results....24", "Know1.3" = "Section 3:Identify your agreement with each of the following: - If the statistical conclusions align with our expectations, we can assume the data was collected appropriately.", "Know1.4" = "Section 3:Identify your agreement with each of the following: - For any real dataset, there can be multiple ways to analyze it.")  
  
NewData = NewData %>%  
 rename("Know2.1" = "Section 3:Identify your agreement with each of the following: - If a hypothesis test results in a p-value less than 0.05, we can have confidence that it was properly conducted.", "Know2.2" = "Section 3:Identify your agreement with each of the following: - As long as we can verify that the mathematical calculations are done correctly, we can trust the statistical results....65", "Know2.3" = "Section 3:Identify your agreement with each of the following: - If the statistical conclusions align with our expectations, we can assume that data was collected properly.", "Know2.4" = "Section 3:Identify your agreement with each of the following: - For any real dataset, there can be multiple ways to analyze it.")  
  
#removing columns with greenhouse data and blank columns  
#NewData = NewData[-c(67:76, 88:91)]  
  
#Create an new order for several factor levels  
NewData$PrePost <- factor(NewData$PrePost, levels = c("PreTest", "PostTest"))   
NewData$PriorStats1 <- factor(NewData$PriorStats1, levels = c("none", "one", "two", "three or more", "NA"))  
NewData$PriorStats2 <- factor(NewData$PriorStats2, levels = c("none", "one", "two", "three or more", "NA"))  
  
  
##Are students "White" or "non-white" ##Majors include STATS  
vec1 = c("White","Caucasian", "caucasian", "White (Caucasian)", "caucasion", "White/Caucasian", "Caucasian, non-Hispanic", "White/Non hispanic","white/ Caucasian", "Caucasion" , "White / Caucasian", "White (not hispanic or latino)")  
NewData = mutate(NewData, Ethnicity2=ifelse(str\_detect(Ethnicity2, paste(vec1, collapse = "|" )), "White", "Other"))  
  
##Are students US Citizens  
NewData <- NewData %>%  
 mutate(Citizen2 = if\_else(Citizen2 == "Yes", "Yes","No"))  
   
##Do Studnets Majors include STATS  
# There are strange characters in the Major1 column for 2 specific students with a Gender.... major, this removes the strange characters  
NewData = mutate(NewData,Major1 = if\_else(str\_detect(Major1,"Gender"), "Gender", Major1))  
  
NewData$Major1 <- tolower(NewData$Major1)  
  
vec <- c("ecn", "econ", "stats", "nurs", "engineer", "tech", "bio", "soc", "science", "business", "health", "psych")  
NewData = mutate(NewData, statsmaj=ifelse(str\_detect(Major1, paste(vec, collapse = "|" )), "Yes", "No"))  
  
## Student Gender  
NewData$Pronoun2[NewData$Pronoun2 == "He/Him/His"] <- "He"  
NewData$Pronoun2[NewData$Pronoun2 == "She/Her/Hers"] <- "She"  
NewData$Pronoun2[NewData$Pronoun2 == "They/Them/Theirs"] <- "Other"  
NewData$Pronoun2[NewData$Pronoun2 == "Other"] <- "Other"  
NewData$Pronoun2[NewData$Pronoun2 == "Prefer not to answer"] <- "Other"  
  
NewData$Pronoun2[is.na(NewData$Pronoun2)] = "Other"  
NewData$Ethnicity2[is.na(NewData$Ethnicity2)] = "Other"  
  
# NewData$Score[is.na(NewData$Score)] = 0  
  
###Filtering   
#NewData = filter(NewData, NewData$Finished1 == "TRUE", NewData$Finished2 == "TRUE")  
#dim(NewData)  
###

#visualizations:

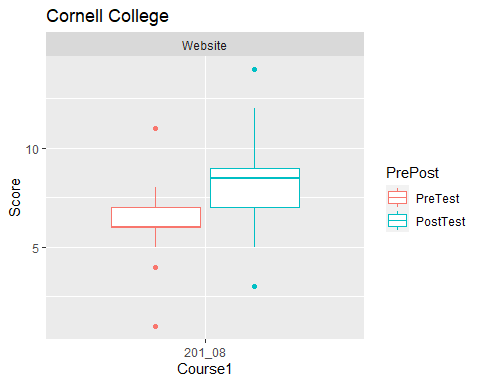
#COURSES  
NewData %>%  
 filter(Institution1 == "Grinnell College") %>%  
 ggplot(aes(x = Course1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Grinnell College")

## Warning: Removed 47 rows containing non-finite values (stat\_boxplot).



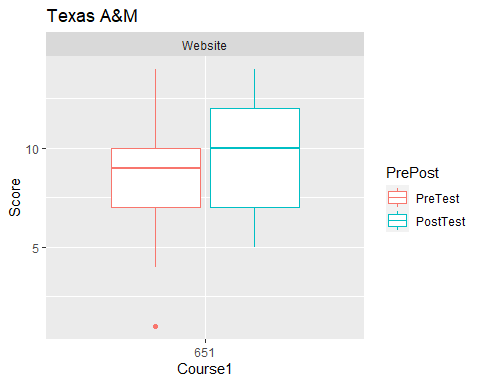
NewData %>%  
 filter(Institution1 == "Cornell College") %>%  
 ggplot(aes(x = Course1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Cornell College")

## Warning: Removed 3 rows containing non-finite values (stat\_boxplot).



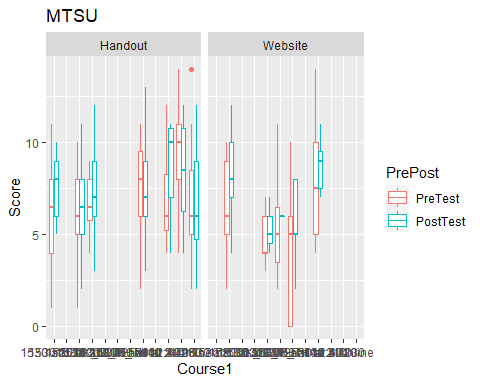
NewData %>%  
 filter(Institution1 == "Texas A&M") %>%  
 ggplot(aes(x = Course1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Texas A&M")

## Warning: Removed 7 rows containing non-finite values (stat\_boxplot).



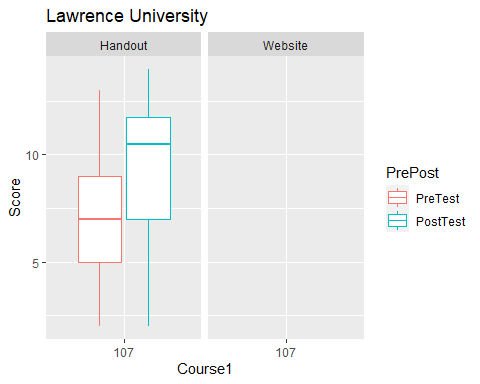
NewData %>%  
 filter(Institution1 == "Middle Tennessee State University") %>%  
 ggplot(aes(x = Course1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "MTSU")

## Warning: Removed 113 rows containing non-finite values (stat\_boxplot).



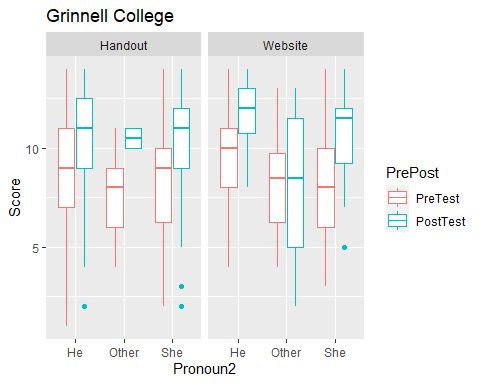
NewData %>%  
 filter(Institution1 == "Lawrence University") %>%  
 ggplot(aes(x = Course1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Lawrence University")

## Warning: Removed 73 rows containing non-finite values (stat\_boxplot).



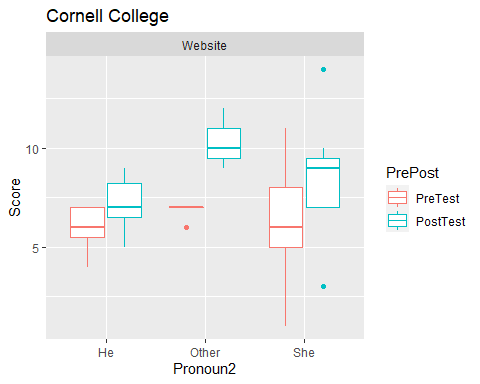
#PRONOUNS  
  
NewData %>%  
 filter(Institution1 == "Grinnell College") %>%  
 ggplot(aes(x = Pronoun2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Grinnell College")

## Warning: Removed 47 rows containing non-finite values (stat\_boxplot).



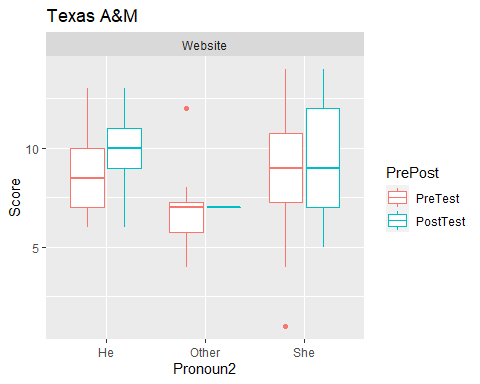
NewData %>%  
 filter(Institution1 == "Cornell College") %>%  
 ggplot(aes(x = Pronoun2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Cornell College")

## Warning: Removed 3 rows containing non-finite values (stat\_boxplot).



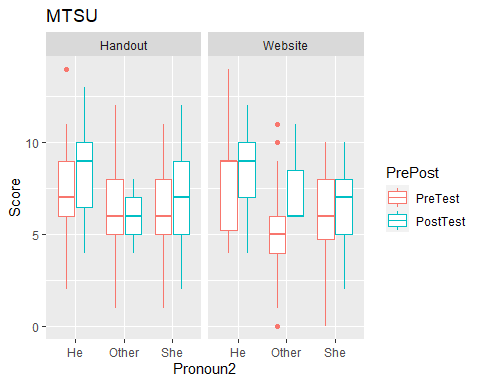
NewData %>%  
 filter(Institution1 == "Texas A&M") %>%  
 ggplot(aes(x = Pronoun2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Texas A&M")

## Warning: Removed 7 rows containing non-finite values (stat\_boxplot).



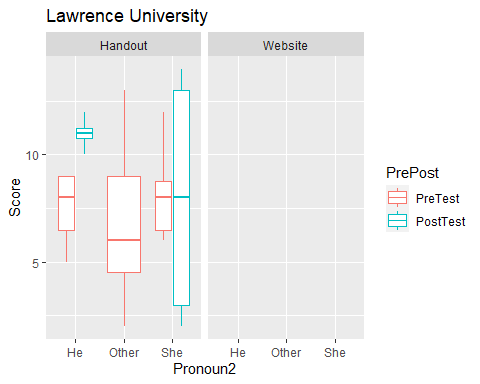
NewData %>%  
 filter(Institution1 == "Middle Tennessee State University") %>%  
 ggplot(aes(x = Pronoun2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "MTSU")

## Warning: Removed 113 rows containing non-finite values (stat\_boxplot).



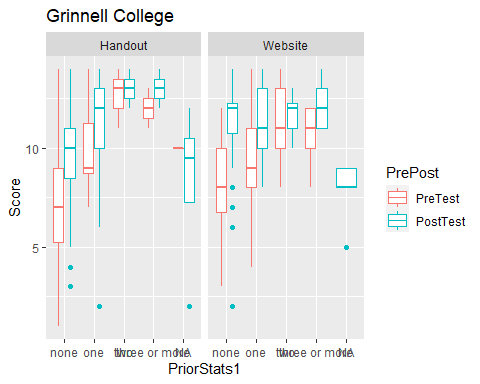
NewData %>%  
 filter(Institution1 == "Lawrence University") %>%  
 ggplot(aes(x = Pronoun2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Lawrence University")

## Warning: Removed 73 rows containing non-finite values (stat\_boxplot).



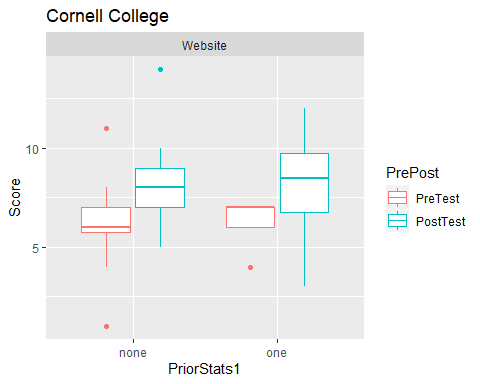
#Prior Stats courses.   
NewData %>%  
 filter(Institution1 == "Grinnell College") %>%  
 ggplot(aes(x = PriorStats1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Grinnell College")

## Warning: Removed 47 rows containing non-finite values (stat\_boxplot).



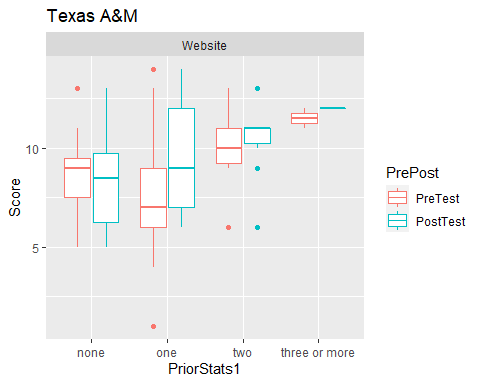
NewData %>%  
 filter(Institution1 == "Cornell College") %>%  
 ggplot(aes(x = PriorStats1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Cornell College")

## Warning: Removed 3 rows containing non-finite values (stat\_boxplot).



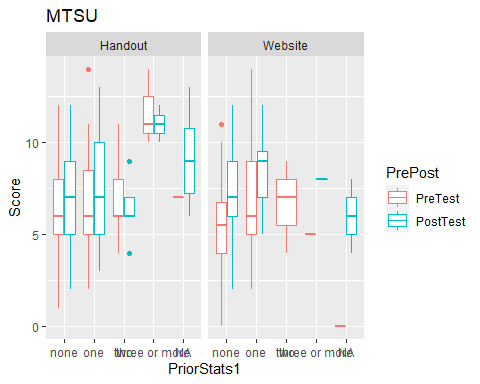
NewData %>%  
 filter(Institution1 == "Texas A&M") %>%  
 ggplot(aes(x = PriorStats1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Texas A&M")

## Warning: Removed 7 rows containing non-finite values (stat\_boxplot).



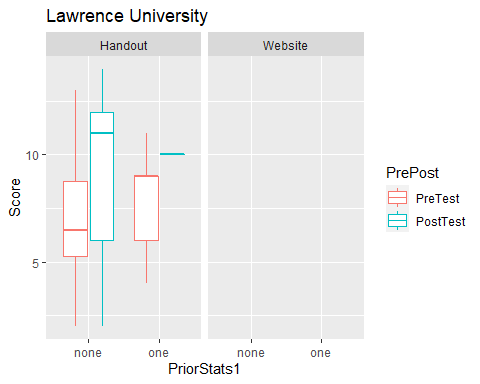
NewData %>%  
 filter(Institution1 == "Middle Tennessee State University") %>%  
 ggplot(aes(x = PriorStats1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "MTSU")

## Warning: Removed 113 rows containing non-finite values (stat\_boxplot).



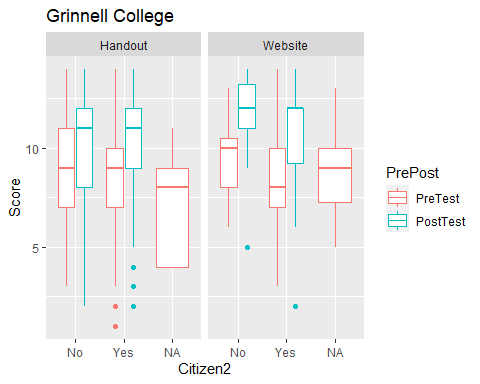
NewData %>%  
 filter(Institution1 == "Lawrence University") %>%  
 ggplot(aes(x = PriorStats1, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Lawrence University")

## Warning: Removed 73 rows containing non-finite values (stat\_boxplot).



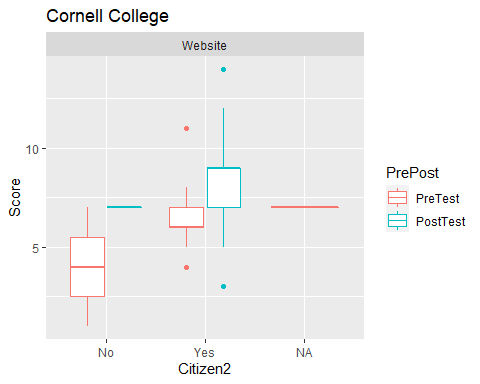
#Citizen  
  
NewData %>%  
 filter(Institution1 == "Grinnell College") %>%  
 ggplot(aes(x = Citizen2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Grinnell College")

## Warning: Removed 47 rows containing non-finite values (stat\_boxplot).



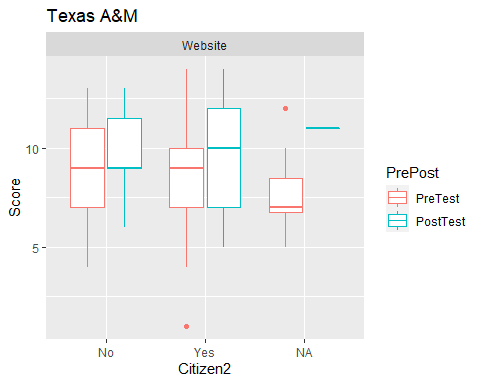
NewData %>%  
 filter(Institution1 == "Cornell College") %>%  
 ggplot(aes(x = Citizen2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Cornell College")

## Warning: Removed 3 rows containing non-finite values (stat\_boxplot).



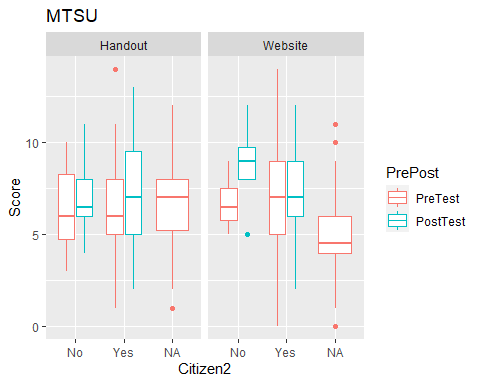
NewData %>%  
 filter(Institution1 == "Texas A&M") %>%  
 ggplot(aes(x = Citizen2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Texas A&M")

## Warning: Removed 7 rows containing non-finite values (stat\_boxplot).



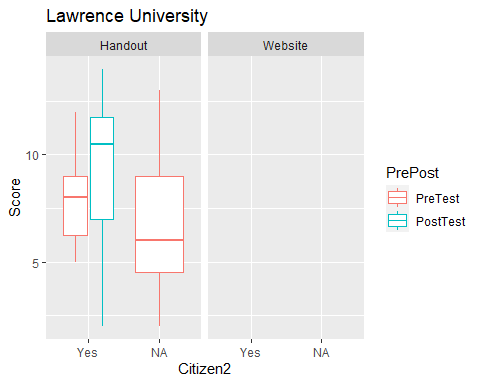
NewData %>%  
 filter(Institution1 == "Middle Tennessee State University") %>%  
 ggplot(aes(x = Citizen2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "MTSU")

## Warning: Removed 113 rows containing non-finite values (stat\_boxplot).



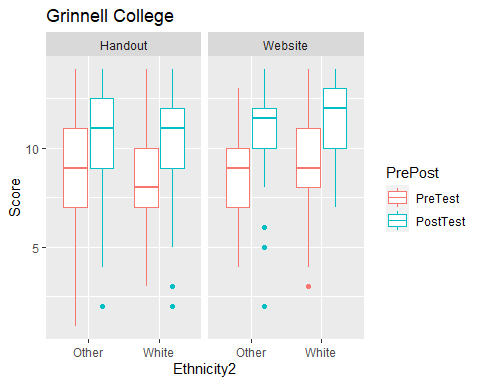
NewData %>%  
 filter(Institution1 == "Lawrence University") %>%  
 ggplot(aes(x = Citizen2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Lawrence University")

## Warning: Removed 73 rows containing non-finite values (stat\_boxplot).



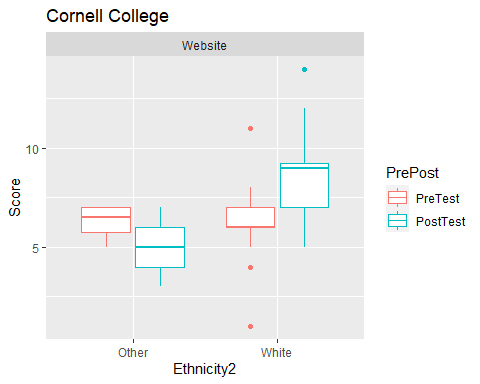
#Ethnicity  
  
  
NewData %>%  
 filter(Institution1 == "Grinnell College") %>%  
 ggplot(aes(x = Ethnicity2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Grinnell College")

## Warning: Removed 47 rows containing non-finite values (stat\_boxplot).



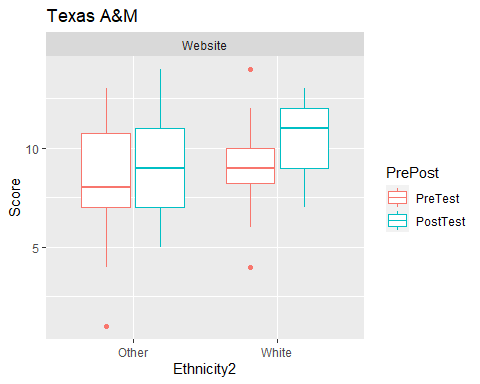
NewData %>%  
 filter(Institution1 == "Cornell College") %>%  
 ggplot(aes(x = Ethnicity2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Cornell College")

## Warning: Removed 3 rows containing non-finite values (stat\_boxplot).



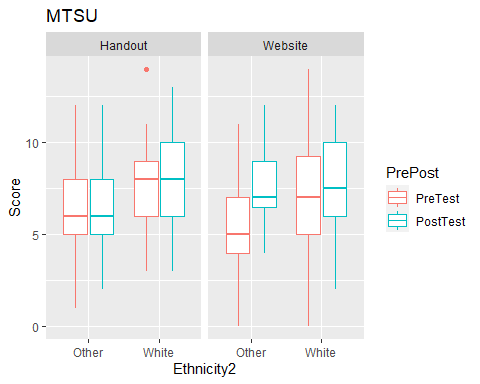
NewData %>%  
 filter(Institution1 == "Texas A&M") %>%  
 ggplot(aes(x = Ethnicity2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Texas A&M")

## Warning: Removed 7 rows containing non-finite values (stat\_boxplot).



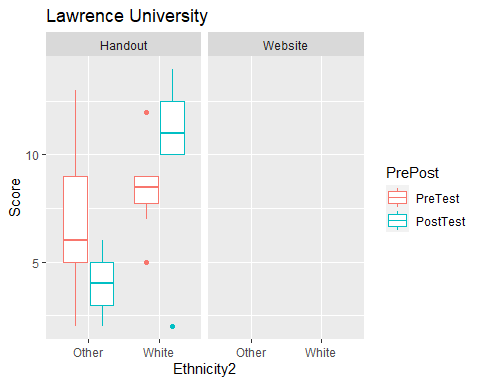
NewData %>%  
 filter(Institution1 == "Middle Tennessee State University") %>%  
 ggplot(aes(x = Ethnicity2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "MTSU")

## Warning: Removed 113 rows containing non-finite values (stat\_boxplot).



NewData %>%  
 filter(Institution1 == "Lawrence University") %>%  
 ggplot(aes(x = Ethnicity2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
 facet\_wrap(~Type) +  
 labs(title = "Lawrence University")

## Warning: Removed 73 rows containing non-finite values (stat\_boxplot).



#optional block - all commented for now

#Attitude questions  
#   
# NewData2 %>%  
# filter(Institution1 == "Grinnell College") %>%  
# ggplot(aes(x = Citizen2, y =Score)) + geom\_boxplot() + aes(color = PrePost) +  
# facet\_wrap(~Type) +  
# labs(title = "Grinnell College")  
#   
# ggplot(data = filter(NewData, NewData$Institution1 == "Grinnell College" & NewData$Institution2 == "Grinnell College"), aes(x = `I would prefer to avoid courses that use statistics1`, y =Score)) + geom\_boxplot() + aes(color = PrePost) + facet\_wrap(~Type) + labs(title = "Grinnell College")  
#   
#   
# ggplot(data = filter(NewData, NewData$Institution1 == "Cornell College" & NewData$Institution2 == "Cornell College"), aes(x = `I would prefer to avoid courses that use statistics1`, y =Score)) + geom\_boxplot() + aes(color = PrePost) + facet\_wrap(~Type) + labs(title = "Cornell College")  
#   
#   
# ggplot(data = filter(NewData, NewData$Institution1 == "Texas A&M" & NewData$Institution2 == "Texas A&M"), aes(x = `I would prefer to avoid courses that use statistics1`, y =Score)) + geom\_boxplot() + aes(color = PrePost) + facet\_wrap(~Type) + labs(title = "Texas A&M")  
#   
#   
# ggplot(data = filter(NewData, NewData$Institution1 == "Middle Tennessee State University" & NewData$Institution2 == "Middle Tennessee State University"), aes(x = `I would prefer to avoid courses that use statistics1`, y =Score)) + geom\_boxplot() + aes(color = PrePost) + facet\_wrap(~Type) + labs(title = "Middle Tennessee State University")  
#   
#   
# ggplot(data = filter(NewData, NewData$Institution1 == "Lawrence University" & NewData$Institution2 == "Lawrence University"), aes(x = `I would prefer to avoid courses that use statistics1`, y =Score)) + geom\_boxplot() + aes(color = PrePost) + facet\_wrap(~Type) + labs(title = "Lawrence University")

#SCORE CALCULATIONS #background score:

#Background1 questions  
  
NewData$PriorStats1 = ifelse(NewData$PriorStats1 == "one", 1, ifelse(NewData$PriorStats1 == "two", 2, ifelse(NewData$PriorStats1 == "three or more", 3, 0)))  
NewData$PriorStats1[is.na(NewData$PriorStats1)] = 0  
  
  
NewData$CalcMean1[is.na(NewData$CalcMean1)] = 0  
NewData$CalcMean1 = ifelse(NewData$CalcMean1 == "Yes", 1, 0)  
  
NewData$CalcSD1[is.na(NewData$CalcSD1)] = 0  
NewData$CalcSD1 = ifelse(NewData$CalcSD1 == "Yes", 1, 0)  
  
NewData$CalcSlope1[is.na(NewData$CalcSlope1)] = 0  
NewData$CalcSlope1 = ifelse(NewData$CalcSlope1 == "Yes", 1, 0)  
  
NewData$CalcPValue1[is.na(NewData$CalcPValue1)] = 0  
NewData$CalcPValue1 = ifelse(NewData$CalcPValue1 == "Yes", 1, 0)  
  
NewData$Background1 = NewData$PriorStats1 + NewData$CalcMean1 + NewData$CalcSD1 + NewData$CalcSlope1 + NewData$CalcPValue1

#attitude score:

#Attitude questions  
NewData$att1.1[NewData$att1.1 == "Somewhat disagree"] = "Disagree"  
NewData$att1.1[NewData$att1.1 == "Strongly disagree"] = "Disagree"  
NewData$att1.1[NewData$att1.1 == "Somewhat agree"] = "Agree"  
NewData$att1.1[NewData$att1.1 == "Strongly agree"] = "Agree"  
  
NewData$att1.2[NewData$att1.2 == "Somewhat disagree"] = "Disagree"  
NewData$att1.2[NewData$att1.2 == "Strongly disagree"] = "Disagree"  
NewData$att1.2[NewData$att1.2 == "Somewhat agree"] = "Agree"  
NewData$att1.2[NewData$att1.2 == "Strongly agree"] = "Agree"  
  
NewData$att1.3[NewData$att1.3 == "Somewhat disagree"] = "Disagree"  
NewData$att1.3[NewData$att1.3 == "Strongly disagree"] = "Disagree"  
NewData$att1.3[NewData$att1.3 == "Somewhat agree"] = "Agree"  
NewData$att1.3[NewData$att1.3 == "Strongly agree"] = "Agree"  
  
NewData$att1.4[NewData$att1.4 == "Somewhat disagree"] = "Disagree"  
NewData$att1.4[NewData$att1.4 == "Strongly disagree"] = "Disagree"  
NewData$att1.4[NewData$att1.4 == "Somewhat agree"] = "Agree"  
NewData$att1.4[NewData$att1.4 == "Strongly agree"] = "Agree"  
  
NewData$att1.5[NewData$att1.5 == "Somewhat disagree"] = "Disagree"  
NewData$att1.5[NewData$att1.5 == "Strongly disagree"] = "Disagree"  
NewData$att1.5[NewData$att1.5 == "Somewhat agree"] = "Agree"  
NewData$att1.5[NewData$att1.5 == "Strongly agree"] = "Agree"  
  
NewData$att1.6[NewData$att1.6 == "Somewhat disagree"] = "Disagree"  
NewData$att1.6[NewData$att1.6 == "Strongly disagree"] = "Disagree"  
NewData$att1.6[NewData$att1.6 == "Somewhat agree"] = "Agree"  
NewData$att1.6[NewData$att1.6 == "Strongly agree"] = "Agree"  
  
NewData$att2.1[NewData$att2.1 == "Somewhat disagree"] = "Strongly disagree"  
NewData$att2.1[NewData$att2.1 == "Somewhat agree"] = "Strongly agree"  
  
NewData$att2.2[NewData$att2.2 == "Somewhat disagree"] = "Strongly disagree"  
NewData$att2.2[NewData$att2.2 == "Somewhat agree"] = "Strongly agree"  
  
NewData$att2.3[NewData$att2.3 == "Somewhat disagree"] = "Strongly disagree"  
NewData$att2.3[NewData$att2.3 == "Somewhat agree"] = "Strongly agree"  
  
NewData$att2.4[NewData$att2.4 == "Somewhat disagree"] = "Strongly disagree"  
NewData$att2.4[NewData$att2.4 == "Somewhat agree"] = "Strongly agree"  
  
NewData$att2.5[NewData$att2.5 == "Somewhat disagree"] = "Strongly disagree"  
NewData$att2.5[NewData$att2.5 == "Somewhat agree"] = "Strongly agree"  
  
NewData$att2.6[NewData$att2.6 == "Somewhat disagree"] = "Strongly disagree"  
NewData$att2.6[NewData$att2.6 == "Somewhat agree"] = "Strongly agree"  
  
NewData$att1.1[is.na(NewData$att1.1)] = 0  
NewData$att1.1 = ifelse(NewData$att1.1 == "Disagree", 1, ifelse(NewData$att1.1 == "Agree", -1,0))  
  
NewData$att1.2[is.na(NewData$att1.2)] = 0  
NewData$att1.2 = ifelse(NewData$att1.2 == "Disagree", -1, ifelse(NewData$att1.2 == "Agree", 1,0))  
  
NewData$att1.3[is.na(NewData$att1.3)] = 0  
NewData$att1.3 = ifelse(NewData$att1.3 == "Disagree", 1, ifelse(NewData$att1.3 == "Agree", -1,0))  
  
NewData$att1.4[is.na(NewData$att1.4)] = 0  
NewData$att1.4 = ifelse(NewData$att1.4 == "Disagree", -1, ifelse(NewData$att1.4 == "Agree", 1,0))  
  
NewData$att1.5[is.na(NewData$att1.5)] = 0  
NewData$att1.5 = ifelse(NewData$att1.5 == "Disagree", -1, ifelse(NewData$att1.5 == "Agree", 1,0))  
  
NewData$att1.6[is.na(NewData$att1.6)] = 0  
NewData$att1.6 = ifelse(NewData$att1.6 == "Disagree", 1, ifelse(NewData$att1.6 == "Agree", -1,0))  
  
NewData$Attd1 = NewData$att1.1 + NewData$att1.2 + NewData$att1.3 + NewData$att1.4 + NewData$att1.5 + NewData$att1.6  
  
NewData$att2.1[is.na(NewData$att2.1)] = 0  
NewData$att2.1 = ifelse(NewData$att2.1 == "Strongly disagree", 1, ifelse(NewData$att2.1 == "Strongly agree", -1,0))  
  
NewData$att2.2[is.na(NewData$att2.2)] = 0  
NewData$att2.2 = ifelse(NewData$att2.2 == "Strongly disagree", -1, ifelse(NewData$att2.2 == "Strongly agree", 1,0))  
  
NewData$att2.3[is.na(NewData$att2.3)] = 0  
NewData$att2.3 = ifelse(NewData$att2.3 == "Strongly disagree", 1, ifelse(NewData$att2.3 == "Strongly agree", -1,0))  
  
NewData$att2.4[is.na(NewData$att2.4)] = 0  
NewData$att2.4 = ifelse(NewData$att2.4 == "Strongly disagree", -1, ifelse(NewData$att2.4 == "Strongly agree", 1,0))  
  
NewData$att2.5[is.na(NewData$att2.5)] = 0  
NewData$att2.5 = ifelse(NewData$att2.5 == "Strongly disagree", -1, ifelse(NewData$att2.5 == "Strongly agree", 1,0))  
  
NewData$att2.6[is.na(NewData$att2.6)] = 0  
NewData$att2.6 = ifelse(NewData$att2.6 == "Strongly disagree", 1, ifelse(NewData$att2.6 == "Strongly agree", -1,0))  
  
NewData$Attd2 = NewData$att2.1 + NewData$att2.2 + NewData$att2.3 + NewData$att2.4 + NewData$att2.5 + NewData$att2.6

#run this block if half scores for knowledge questions

#run this block if half scores for knowledge questions  
NewData$Know1.1[is.na(NewData$Know1.1)] = 0  
NewData$Know1.1 = ifelse(NewData$Know1.1 == "Disagree", 1, ifelse(NewData$Know1.1 == "Unsure", 0.5,0))  
  
NewData$Know1.2[is.na(NewData$Know1.2)] = 0  
NewData$Know1.2 = ifelse(NewData$Know1.2 == "Disagree", 1, ifelse(NewData$Know1.2 == "Unsure", 0.5,0))  
  
NewData$Know1.3[is.na(NewData$Know1.3)] = 0  
NewData$Know1.3 = ifelse(NewData$Know1.3 == "Disagree", 1, ifelse(NewData$Know1.3 == "Unsure", 0.5,0))  
  
NewData$Know1.4[is.na(NewData$Know1.4)] = 0  
NewData$Know1.4 = ifelse(NewData$Know1.4 == "Agree", 1, ifelse(NewData$Know1.4 == "Unsure", 0.5,0))  
  
NewData$ModifyData1[is.na(NewData$ModifyData1)] = 0  
NewData$ModifyData1 = ifelse(NewData$ModifyData1 == "Disagree", 1, ifelse(NewData$ModifyData1 == "Unsure", 0.5,0))  
  
NewData$BetterModel1[is.na(NewData$BetterModel1)] = 0  
NewData$BetterModel1 = ifelse(NewData$BetterModel1 == "Agree", 1, ifelse(NewData$BetterModel1 == "Unsure", 0.5,0))  
  
NewData$Predict1[is.na(NewData$Predict1)] = 0  
NewData$Predict1 = ifelse(NewData$Predict1 == "Disagree", 1, ifelse(NewData$Predict1 == "Unsure", 0.5,0))  
  
NewData$CorrGood1[is.na(NewData$CorrGood1)] = 0  
NewData$CorrGood1 = ifelse(NewData$CorrGood1 == "Disagree", 1, ifelse(NewData$CorrGood1 == "Unsure", 0.5,0))  
  
  
NewData$Know2.1[is.na(NewData$Know2.1)] = 0  
NewData$Know2.1 = ifelse(NewData$Know2.1 == "Disagree", 1, ifelse(NewData$Know2.1 == "Unsure", 0.5,0))  
  
NewData$Know2.2[is.na(NewData$Know2.2)] = 0  
NewData$Know2.2 = ifelse(NewData$Know2.2 == "Disagree", 1, ifelse(NewData$Know2.2 == "Unsure", 0.5,0))  
  
NewData$Know2.3[is.na(NewData$Know2.3)] = 0  
NewData$Know2.3 = ifelse(NewData$Know2.3 == "Disagree", 1, ifelse(NewData$Know2.3 == "Unsure", 0.5,0))  
  
NewData$Know2.4[is.na(NewData$Know2.4)] = 0  
NewData$Know2.4 = ifelse(NewData$Know2.4 == "Agree", 1, ifelse(NewData$Know2.4 == "Unsure", 0.5,0))  
  
NewData$ModifyData2[is.na(NewData$ModifyData2)] = 0  
NewData$ModifyData2 = ifelse(NewData$ModifyData2 == "Disagree", 1, ifelse(NewData$ModifyData2 == "Unsure", 0.5,0))  
  
NewData$BetterModel2[is.na(NewData$BetterModel2)] = 0  
NewData$BetterModel2 = ifelse(NewData$BetterModel2 == "Agree", 1, ifelse(NewData$BetterModel2 == "Unsure", 0.5,0))  
  
NewData$Predict2[is.na(NewData$Predict2)] = 0  
NewData$Predict2 = ifelse(NewData$Predict2 == "Disagree", 1, ifelse(NewData$Predict2 == "Unsure", 0.5,0))  
  
NewData$CorrGood2[is.na(NewData$CorrGood2)] = 0  
NewData$CorrGood2 = ifelse(NewData$CorrGood2 == "Disagree", 1, ifelse(NewData$CorrGood2 == "Unsure", 0.5,0))

# OR run this block if whole scores for knowledge questions (DEFAULT)

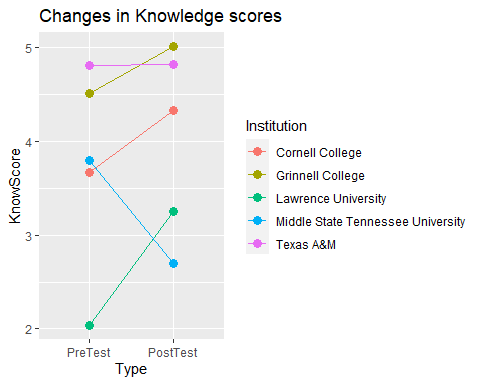
#run this block if whole scores for knowledge questions (DEFAULT)  
  
NewData$Know1.1[is.na(NewData$Know1.1)] = 0  
NewData$Know1.1 = ifelse(NewData$Know1.1 == "Disagree", 1, 0)  
  
NewData$Know1.2[is.na(NewData$Know1.2)] = 0  
NewData$Know1.2 = ifelse(NewData$Know1.2 == "Disagree", 1, 0)  
  
NewData$Know1.3[is.na(NewData$Know1.3)] = 0  
NewData$Know1.3 = ifelse(NewData$Know1.3 == "Disagree", 1,0)  
  
NewData$Know1.4[is.na(NewData$Know1.4)] = 0  
NewData$Know1.4 = ifelse(NewData$Know1.4 == "Agree", 1,0)  
  
NewData$ModifyData1[is.na(NewData$ModifyData1)] = 0  
NewData$ModifyData1 = ifelse(NewData$ModifyData1 == "Disagree", 1, 0)  
  
NewData$BetterModel1[is.na(NewData$BetterModel1)] = 0  
NewData$BetterModel1 = ifelse(NewData$BetterModel1 == "Agree", 1, 0)  
  
NewData$Predict1[is.na(NewData$Predict1)] = 0  
NewData$Predict1 = ifelse(NewData$Predict1 == "Disagree", 1, 0)  
  
NewData$CorrGood1[is.na(NewData$CorrGood1)] = 0  
NewData$CorrGood1 = ifelse(NewData$CorrGood1 == "Disagree", 1, 0)  
  
  
NewData$Know2.1[is.na(NewData$Know2.1)] = 0  
NewData$Know2.1 = ifelse(NewData$Know2.1 == "Disagree", 1,0)  
  
NewData$Know2.2[is.na(NewData$Know2.2)] = 0  
NewData$Know2.2 = ifelse(NewData$Know2.2 == "Disagree", 1,0)  
  
NewData$Know2.3[is.na(NewData$Know2.3)] = 0  
NewData$Know2.3 = ifelse(NewData$Know2.3 == "Disagree", 1,0)  
  
NewData$Know2.4[is.na(NewData$Know2.4)] = 0  
NewData$Know2.4 = ifelse(NewData$Know2.4 == "Agree", 1,0)  
  
NewData$ModifyData2[is.na(NewData$ModifyData2)] = 0  
NewData$ModifyData2 = ifelse(NewData$ModifyData2 == "Disagree", 1, 0)  
  
NewData$BetterModel2[is.na(NewData$BetterModel2)] = 0  
NewData$BetterModel2 = ifelse(NewData$BetterModel2 == "Agree", 1, 0)  
  
NewData$Predict2[is.na(NewData$Predict2)] = 0  
NewData$Predict2 = ifelse(NewData$Predict2 == "Disagree", 1, 0)  
  
NewData$CorrGood2[is.na(NewData$CorrGood2)] = 0  
NewData$CorrGood2 = ifelse(NewData$CorrGood2 == "Disagree", 1, 0)

# run to score rest of the knowledge questions (ALWAYS RUN THIS ONE)

# Rest of the Knowledge questions  
  
NewData$Response1[is.na(NewData$Response1)] = 0  
NewData$Response1 = ifelse(NewData$Response1 == "Age", 1, 0)  
  
NewData$EV1[is.na(NewData$EV1)] = 0  
NewData$EV1 = ifelse(NewData$EV1 == "Percentage NoseBlack", 1,0)  
  
NewData$DPoint1[is.na(NewData$DPoint1)] = 0  
NewData$DPoint1 = ifelse(NewData$DPoint1 == "One Lion", 1, 0)  
  
NewData$SlopeInc1[is.na(NewData$SlopeInc1)] = 0  
NewData$SlopeInc1 = ifelse(NewData$SlopeInc1 == "Increase", 1, 0)  
  
NewData$SameSlope1[is.na(NewData$SameSlope1)] = 0  
NewData$SameSlope1 = ifelse(NewData$SameSlope1 == "Decrease", 1, ifelse(NewData$SameSlope1 == "Stay about the same", 1, 0))  
  
NewData$RegrLine1[is.na(NewData$RegrLine1)] = 0  
NewData$RegrLine1 = ifelse(NewData$RegrLine1 == "Positive", 1, 0)  
  
NewData$Residual1[is.na(NewData$Residual1)] = 0  
NewData$Residual1 = ifelse(NewData$Residual1 == 90, 1, 0)  
  
NewData$Knowled1 = NewData$Response1 + NewData$EV1 + NewData$DPoint1 + NewData$SlopeInc1 + NewData$SameSlope1 + NewData$RegrLine1 + NewData$Residual1 + NewData$ModifyData1 + NewData$BetterModel1 + NewData$Predict1 + NewData$CorrGood1 + NewData$Know1.1 + NewData$Know1.2 + NewData$Know1.3  
  
NewData$Response2[is.na(NewData$Response2)] = 0  
NewData$Response2 = ifelse(NewData$Response2 == "Age", 1, 0)  
  
NewData$EV2[is.na(NewData$EV2)] = 0  
NewData$EV2 = ifelse(NewData$EV2 == "Percentage NoseBlack", 1, 0)  
  
NewData$DPoint2[is.na(NewData$DPoint2)] = 0  
NewData$DPoint2 = ifelse(NewData$DPoint2 == "One Lion", 1, 0)  
  
NewData$SlopeInc2[is.na(NewData$SlopeInc2)] = 0  
NewData$SlopeInc2 = ifelse(NewData$SlopeInc2 == "Increase", 1, 0)  
  
NewData$SameSlope2[is.na(NewData$SameSlope2)] = 0  
NewData$SameSlope2 = ifelse(NewData$SameSlope2 == "Decrease", 1, ifelse(NewData$SameSlope2 == "Stay about the same", 1, 0))  
  
NewData$RegrLine2[is.na(NewData$RegrLine2)] = 0  
NewData$RegrLine2 = ifelse(NewData$RegrLine2 == "Positive", 1, 0)  
  
NewData$Residual2[is.na(NewData$Residual2)] = 0  
NewData$Residual2 = ifelse(NewData$Residual2 == 90, 1, 0)  
  
#add new knowledge cols also  
NewData$Knowled2 = NewData$Response2 + NewData$EV2 + NewData$DPoint2 + NewData$SlopeInc2 + NewData$SameSlope2 + NewData$RegrLine2 + NewData$Residual2 + NewData$ModifyData2 + NewData$BetterModel2 + NewData$Predict2 + NewData$CorrGood2 + NewData$Know2.1 + NewData$Know2.2 + NewData$Know2.3

#visualizations for knowledge comparing pre and post for each institution

dfGrin = filter(NewData, Institution1 == "Grinnell College")  
dfCorn = filter(NewData, Institution1 == "Cornell College")  
dfTex = filter(NewData, Institution1 == "Texas A&M")  
dfMTSU = filter(NewData, Institution1 == "Middle Tennessee State University")  
dfLaw = filter(NewData, Institution1 == "Lawrence University")  
knowGrinPre = mean(dfGrin$Knowled1)  
knowGrinPost = mean(dfGrin$Knowled2)  
knowCornPre = mean(dfCorn$Knowled1)  
knowCornPost = mean(dfCorn$Knowled2)  
knowTexPre = mean(dfTex$Knowled1)  
knowTexPost = mean(dfTex$Knowled2)  
knowMTSUPre = mean(dfMTSU$Knowled1)  
knowMTSUPost = mean(dfMTSU$Knowled2)  
knowLawPre = mean(dfLaw$Knowled1)  
knowLawPost = mean(dfLaw$Knowled2)  
  
  
Institution = c("Grinnell College", "Grinnell College", "Cornell College", "Cornell College", "Texas A&M","Texas A&M", "Middle State Tennessee University", "Middle State Tennessee University", "Lawrence University", "Lawrence University")  
Type = c("PreTest", "PostTest", "PreTest", "PostTest", "PreTest", "PostTest", "PreTest", "PostTest", "PreTest", "PostTest")  
KnowScore = c(knowGrinPre, knowGrinPost, knowCornPre, knowCornPost, knowTexPre, knowTexPost, knowMTSUPre, knowMTSUPost, knowLawPre, knowLawPost)  
  
dfAllKnow = data.frame(Institution, KnowScore, Type)  
dfAllKnow$Type <- factor(dfAllKnow$Type, levels = c("PreTest", "PostTest"))   
  
  
ggplot(dfAllKnow, aes(x = Type, y = KnowScore, colour = Institution, group = Institution)) +  
 geom\_line() +  
 geom\_point(size = 3) +  
 labs(title = "Changes in Knowledge scores", ylab = "Knowledge score")

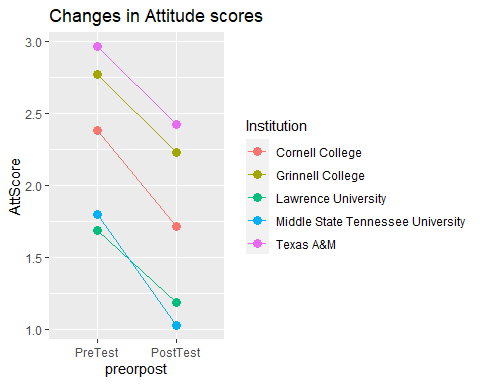


#visualizations for attitude comparing pre and post for each institution

dfGrinAtt = filter(NewData, Institution1 == "Grinnell College")  
dfCornAtt = filter(NewData, Institution1 == "Cornell College")  
dfTexAtt = filter(NewData, Institution1 == "Texas A&M")  
dfMTSUAtt = filter(NewData, Institution1 == "Middle Tennessee State University")  
dfLawAtt = filter(NewData, Institution1 == "Lawrence University")  
attGrinPre = mean(dfGrinAtt$Attd1)  
attGrinPost = mean(dfGrinAtt$Attd2)  
attCornPre = mean(dfCornAtt$Attd1)  
attCornPost = mean(dfCornAtt$Attd2)  
attTexPre = mean(dfTexAtt$Attd1)  
attTexPost = mean(dfTexAtt$Attd2)  
attMTSUPre = mean(dfMTSUAtt$Attd1)  
attMTSUPost = mean(dfMTSUAtt$Attd2)  
attLawPre = mean(dfLawAtt$Attd1)  
attLawPost = mean(dfLawAtt$Attd2)  
  
Institution = c("Grinnell College", "Grinnell College", "Cornell College", "Cornell College", "Texas A&M","Texas A&M", "Middle State Tennessee University", "Middle State Tennessee University", "Lawrence University", "Lawrence University")  
preorpost = c("PreTest", "PostTest", "PreTest", "PostTest", "PreTest", "PostTest", "PreTest", "PostTest", "PreTest", "PostTest")  
AttScore = c(attGrinPre, attGrinPost, attCornPre, attCornPost, attTexPre, attTexPost, attMTSUPre, attMTSUPost, attLawPre, attLawPost)  
  
dfAllAtt = data.frame(Institution, AttScore, preorpost)  
dfAllAtt$preorpost <- factor(dfAllAtt$preorpost, levels = c("PreTest", "PostTest"))   
print (dfAllAtt)

## Institution AttScore preorpost  
## 1 Grinnell College 2.764706 PreTest  
## 2 Grinnell College 2.225490 PostTest  
## 3 Cornell College 2.380952 PreTest  
## 4 Cornell College 1.714286 PostTest  
## 5 Texas A&M 2.961538 PreTest  
## 6 Texas A&M 2.423077 PostTest  
## 7 Middle State Tennessee University 1.799145 PreTest  
## 8 Middle State Tennessee University 1.025641 PostTest  
## 9 Lawrence University 1.685185 PreTest  
## 10 Lawrence University 1.185185 PostTest

ggplot(dfAllAtt, aes(x = preorpost, y = AttScore, colour = Institution, group = Institution)) +  
 geom\_line() +  
 geom\_point(size = 3) +  
 labs(title = "Changes in Attitude scores")



# library(dplyr)  
# colleges\_type <- group\_by(NewData, Institution1, Type)  
# summarize(colleges\_type,  
# mean = mean(Knowled1))

#verification of Score matching Knowledge columns. Doesn’t quite match because of Lawrence?

library(mosaic)

## Registered S3 method overwritten by 'mosaic':  
## method from   
## fortify.SpatialPolygonsDataFrame ggplot2

##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.

##   
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following object is masked from 'package:purrr':  
##   
## cross

## The following object is masked from 'package:ggplot2':  
##   
## stat

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,  
## quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

tmp = NewData %>% filter(Institution1 == "Grinnell College")  
  
tmp$Score[is.na(tmp$Score)] = 0  
  
# paste("Mean for pre Score: " , favstats(tmp$Score))  
# paste("Mean for post Score: " ,favstats(tmp$Score))  
# paste("Mean for pre Know: " , favstats(tmp$Knowled1))  
# paste("Mean for post Know: " ,favstats(tmp$Knowled2))  
  
print(sum(tmp$Score))

## [1] 3476

print(sum(tmp$Knowled1) /2 + sum(tmp$Knowled2)/2)

## [1] 1943

#identify rows where it doesn't match. check if grinnell works.   
  
# write.csv(tmp, "tmp.csv")

#why is attitude scores lowering?

#which questions are making a difference in attitudes. where is the change happening? tables? boxplots? summary?

#In the table, meanatt1.1 means mean score on question1 pretest and meanatt2.1 means mean score on question1 posttest. #“q1” is “I would prefer to avoid courses that use statistics1”, #“q2” is “Statistics will help me better understand studies that I read or hear about1”, #“q3” is “Statistics requires critical thinking that helps researchers make better decisions based upon data1”, #“q4” is “Real problems often have obstacles that could be approached in multiple ways1”, #“q5” is “This course will be useful for me in the future1”, #“q6” is “There really is no difference between statistics courses and math courses1”

NewData2 = NewData  
  
attmeans = group\_by(NewData2, Institution1)  
summarize(attmeans,   
 meanatt1.1 = mean(att1.1),  
 meanatt2.1 = mean(att2.1),  
 meanatt1.2 = mean(att1.2),  
 meanatt2.2 = mean(att2.2),  
 meanatt1.3 = mean(att1.3),  
 meanatt2.3 = mean(att2.3),  
 meanatt1.4 = mean(att1.4),  
 meanatt2.4 = mean(att2.4),  
 meanatt1.5 = mean(att1.5),  
 meanatt2.5 = mean(att2.5),  
 meanatt1.6 = mean(att1.6),  
 meanatt2.6 = mean(att2.6))

## # A tibble: 6 x 13  
## Institution1 meanatt1.1 meanatt2.1 meanatt1.2 meanatt2.2 meanatt1.3 meanatt2.3  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Cornell Col~ 0.0952 0.0952 0.905 0.667 -0.857 -0.810  
## 2 Grinnell Co~ 0.466 0.284 0.858 0.735 -0.848 -0.735  
## 3 Lawrence Un~ -0.204 -0.167 0.815 0.685 -0.907 -0.704  
## 4 Middle Tenn~ 0.132 0.115 0.641 0.376 -0.846 -0.483  
## 5 Texas A&M 0.538 0.462 0.885 0.712 -0.923 -0.692  
## 6 <NA> 0 0 0 0 0 0   
## # ... with 6 more variables: meanatt1.4 <dbl>, meanatt2.4 <dbl>,  
## # meanatt1.5 <dbl>, meanatt2.5 <dbl>, meanatt1.6 <dbl>, meanatt2.6 <dbl>

#Table for PriorStat courses boxplot

Grin = NewData %>% filter(Institution1 == "Grinnell College")  
Corn = NewData %>% filter(Institution1 == "Cornell College")  
Tex = NewData %>% filter(Institution1 == "Texas A&M")  
MTSU = NewData %>% filter(Institution1 == "Middle Tennessee State University")  
Law = NewData %>% filter(Institution1 == "Lawrence University")  
  
  
priorstatgroup1 = group\_by(Grin, Institution1, PriorStats1, Type, PrePost)  
summarise(priorstatgroup1,   
 Gn = n(),  
 Gmean = mean(Score, na.rm = TRUE),   
 Gsd = sd(Score, na.rm= TRUE))

## `summarise()` has grouped output by 'Institution1', 'PriorStats1', 'Type'. You  
## can override using the `.groups` argument.

## # A tibble: 16 x 7  
## # Groups: Institution1, PriorStats1, Type [8]  
## Institution1 PriorStats1 Type PrePost Gn Gmean Gsd  
## <chr> <dbl> <chr> <fct> <int> <dbl> <dbl>  
## 1 Grinnell College 0 Handout PreTest 68 7.37 2.91  
## 2 Grinnell College 0 Handout PostTest 68 9.61 2.79  
## 3 Grinnell College 0 Website PreTest 55 7.90 2.31  
## 4 Grinnell College 0 Website PostTest 55 10.9 2.58  
## 5 Grinnell College 1 Handout PreTest 33 10.0 2.07  
## 6 Grinnell College 1 Handout PostTest 33 10.8 3.06  
## 7 Grinnell College 1 Website PreTest 29 9.31 2.49  
## 8 Grinnell College 1 Website PostTest 29 11.2 1.64  
## 9 Grinnell College 2 Handout PreTest 3 12.7 1.53  
## 10 Grinnell College 2 Handout PostTest 3 13 1   
## 11 Grinnell College 2 Website PreTest 9 11.1 1.90  
## 12 Grinnell College 2 Website PostTest 9 11.8 1.04  
## 13 Grinnell College 3 Handout PreTest 2 12 1.41  
## 14 Grinnell College 3 Handout PostTest 2 13 1.41  
## 15 Grinnell College 3 Website PreTest 5 10.6 1.67  
## 16 Grinnell College 3 Website PostTest 5 12.2 1.30

priorstatgroup2 = group\_by(Corn, Institution1, PriorStats1, Type, PrePost)  
summarise(priorstatgroup2,   
 Cn = n(),  
 Cmean = mean(Score, na.rm = TRUE),   
 Csd = sd(Score, na.rm= TRUE))

## `summarise()` has grouped output by 'Institution1', 'PriorStats1', 'Type'. You  
## can override using the `.groups` argument.

## # A tibble: 4 x 7  
## # Groups: Institution1, PriorStats1, Type [2]  
## Institution1 PriorStats1 Type PrePost Cn Cmean Csd  
## <chr> <dbl> <chr> <fct> <int> <dbl> <dbl>  
## 1 Cornell College 0 Website PreTest 16 6.25 2.11  
## 2 Cornell College 0 Website PostTest 16 8.21 2.33  
## 3 Cornell College 1 Website PreTest 5 6.2 1.30  
## 4 Cornell College 1 Website PostTest 5 8 3.74

priorstatgroup3 = group\_by(Tex, Institution1, PriorStats1, Type, PrePost)  
summarise(priorstatgroup3,   
 Tn = n(),  
 Tmean = mean(Score, na.rm = TRUE),   
 Tsd = sd(Score, na.rm= TRUE))

## `summarise()` has grouped output by 'Institution1', 'PriorStats1', 'Type'. You  
## can override using the `.groups` argument.

## # A tibble: 8 x 7  
## # Groups: Institution1, PriorStats1, Type [4]  
## Institution1 PriorStats1 Type PrePost Tn Tmean Tsd  
## <chr> <dbl> <chr> <fct> <int> <dbl> <dbl>  
## 1 Texas A&M 0 Website PreTest 11 8.82 2.14   
## 2 Texas A&M 0 Website PostTest 11 8.5 2.64   
## 3 Texas A&M 1 Website PreTest 29 7.62 2.77   
## 4 Texas A&M 1 Website PostTest 29 9.62 2.36   
## 5 Texas A&M 2 Website PreTest 10 10.1 1.91   
## 6 Texas A&M 2 Website PostTest 10 10.6 2.01   
## 7 Texas A&M 3 Website PreTest 2 11.5 0.707  
## 8 Texas A&M 3 Website PostTest 2 12 NA

priorstatgroup4 = group\_by(MTSU, Institution1, PriorStats1, Type, PrePost)  
summarise(priorstatgroup4,   
 Mn = n(),  
 Mmean = mean(Score, na.rm = TRUE),   
 Msd = sd(Score, na.rm= TRUE))

## `summarise()` has grouped output by 'Institution1', 'PriorStats1', 'Type'. You  
## can override using the `.groups` argument.

## # A tibble: 16 x 7  
## # Groups: Institution1, PriorStats1, Type [8]  
## Institution1 PriorStats1 Type PrePost Mn Mmean Msd  
## <chr> <dbl> <chr> <fct> <int> <dbl> <dbl>  
## 1 Middle Tennessee State Universi~ 0 Hand~ PreTest 92 6.48 2.40  
## 2 Middle Tennessee State Universi~ 0 Hand~ PostTe~ 92 7.03 2.43  
## 3 Middle Tennessee State Universi~ 0 Webs~ PreTest 50 5.04 2.85  
## 4 Middle Tennessee State Universi~ 0 Webs~ PostTe~ 50 7.05 2.50  
## 5 Middle Tennessee State Universi~ 1 Hand~ PreTest 51 6.84 2.40  
## 6 Middle Tennessee State Universi~ 1 Hand~ PostTe~ 51 7.70 2.90  
## 7 Middle Tennessee State Universi~ 1 Webs~ PreTest 29 6.69 2.80  
## 8 Middle Tennessee State Universi~ 1 Webs~ PostTe~ 29 8.45 2.07  
## 9 Middle Tennessee State Universi~ 2 Hand~ PreTest 5 7 2.65  
## 10 Middle Tennessee State Universi~ 2 Hand~ PostTe~ 5 6.4 1.82  
## 11 Middle Tennessee State Universi~ 2 Webs~ PreTest 3 6.67 2.52  
## 12 Middle Tennessee State Universi~ 2 Webs~ PostTe~ 3 NaN NA   
## 13 Middle Tennessee State Universi~ 3 Hand~ PreTest 3 11.7 2.08  
## 14 Middle Tennessee State Universi~ 3 Hand~ PostTe~ 3 11 1   
## 15 Middle Tennessee State Universi~ 3 Webs~ PreTest 1 5 NA   
## 16 Middle Tennessee State Universi~ 3 Webs~ PostTe~ 1 8 NA

priorstatgroup5 = group\_by(Law, Institution1, PriorStats1, Type, PrePost)  
summarise(priorstatgroup5,   
 Ln = n(),  
 Lmean = mean(Score, na.rm = TRUE),   
 Lsd = sd(Score, na.rm= TRUE))

## `summarise()` has grouped output by 'Institution1', 'PriorStats1', 'Type'. You  
## can override using the `.groups` argument.

## # A tibble: 8 x 7  
## # Groups: Institution1, PriorStats1, Type [4]  
## Institution1 PriorStats1 Type PrePost Ln Lmean Lsd  
## <chr> <dbl> <chr> <fct> <int> <dbl> <dbl>  
## 1 Lawrence University 0 Handout PreTest 18 7.06 2.71  
## 2 Lawrence University 0 Handout PostTest 18 9.11 4.68  
## 3 Lawrence University 0 Website PreTest 23 NaN NA   
## 4 Lawrence University 0 Website PostTest 23 NaN NA   
## 5 Lawrence University 1 Handout PreTest 7 7.71 2.69  
## 6 Lawrence University 1 Handout PostTest 7 10 NA   
## 7 Lawrence University 1 Website PreTest 6 NaN NA   
## 8 Lawrence University 1 Website PostTest 6 NaN NA

write.csv(NewData,"NewData.csv")