Final Project

Jiner Zheng

12/9/2019

### Data Management and Preparation

library(dplyr)  
library(magrittr)  
library(qwraps2)  
library(haven)  
library(ggplot2)  
library(tableone)  
library(Matching)  
MGM\_AER\_P\_P <- read\_dta("~/Documents/R Projects/qac 307/MGM\_AER\_P-P.dta")  
peer\_tutoring <- MGM\_AER\_P\_P[,c(1:17,97:100)]  
peer\_tutoring$dfgpa <- (peer\_tutoring$fgrdnum - peer\_tutoring$ptermcumgpa)/peer\_tutoring$ptermcumgpa   
  
peer\_tutoring$abs\_dfgpa <- abs(peer\_tutoring$dfgpa)  
  
our\_summary1 <-  
 list("Hours Tutored" =  
 list("min" = ~ min(.data$tuthours),  
 "max" = ~ max(.data$tuthours),  
 "mean (sd)" = ~ qwraps2::mean\_sd(.data$tuthours)),  
 "Term Credit Hours Earned" =  
 list("min" = ~ min(.data$termcred),  
 "median" = ~ median(.data$termcred),  
 "max" = ~ max(.data$termcred),  
 "mean (sd)" = ~ qwraps2::mean\_sd(.data$termcred)),  
 "Final Grade on 4.0 scale" =  
 list("min" = ~ min(.data$fgrdnum),  
 "median" = ~ median(.data$fgrdnum),  
 "max" = ~ max(.data$fgrdnum),  
 "mean (sd)" = ~ qwraps2::mean\_sd(.data$fgrdnum)),  
 "Previous Term Cumulative GPA" =  
 list("min" = ~ min(.data$ptermcumgpa),  
 "max" = ~ max(.data$ptermcumgpa),  
 "median" = ~ median(.data$ptermcumgpa),  
 "mean (sd)" = ~qwraps2::mean\_sd(.data$ptermcumgpa)),  
 "Change in GPA compared to Previous Term" =  
 list("min" = ~min(.data$dfgpa),  
 "max" = ~max(.data$dfgpa),  
 "median" = ~median(.data$dfgpa),  
 "mean (sd)" = ~qwraps2::mean\_sd(.data$dfgpa)),  
 "SAT math" =  
 list("min" = ~ min(.data$admsatm),  
 "max" = ~ max(.data$admsatm),  
 "mean (sd)" = ~ qwraps2::mean\_sd(.data$admsatm)),  
 "High School Rank Index" =  
 list("min" = ~ min(.data$admri),  
 "max" = ~ max(.data$admri),  
 "mean (sd)" = ~ qwraps2::mean\_sd(.data$admri)))  
  
demo\_summary1 <-  
 list("Demographics" =   
 list(  
 "Varisity Athlete" = ~ qwraps2::n\_perc0(.data$athpart == 1),  
 "Female" = ~ qwraps2::n\_perc0(.data$demgender == 0),  
 "Male" = ~ qwraps2::n\_perc0(.data$demgender == 1),  
 "African American" = ~ qwraps2::n\_perc0(.data$demafam == 1),  
 "Hispanic" = ~qwraps2::n\_perc0(.data$demhis == 1),  
 "Greek Participant" = ~qwraps2::n\_perc0(.data$demgrkpart == 1),  
 "Freshman" = ~qwraps2::n\_perc0(.data$acafr == 1),  
 "Sophoremore" = ~qwraps2::n\_perc0(.data$acaso == 1)),  
 "Courses Characteristics"=  
 list(  
 "Recitation Course" = ~qwraps2::n\_perc0(.data$acarec == 1),  
 "Lab Course"=~qwraps2::n\_perc0(.data$acalab == 1),  
 "2004-05"=~qwraps2::n\_perc0(.data$y1 == 1),  
 "2005-06"=~qwraps2::n\_perc0(.data$y2 == 1)))  
  
orig\_opt <- options()$qwraps2\_markup  
options(qwraps2\_markup = "markdown")  
peer\_tutoring$tutpart\_factor <- peer\_tutoring$tutpart  
whole <- summary\_table(peer\_tutoring, our\_summary1)  
by\_tutor <- summary\_table(dplyr::group\_by(peer\_tutoring, tutpart\_factor), our\_summary1)  
bind <- cbind(whole, by\_tutor)  
whole2 <- summary\_table(peer\_tutoring, demo\_summary1)  
by\_tutor2 <- summary\_table(dplyr::group\_by(peer\_tutoring,tutpart\_factor), demo\_summary1)  
bind2 <- cbind(whole2,by\_tutor2)

##### regression model (just a glimpse)

mod <- lm(dfgpa~tutpart, data = peer\_tutoring)  
mod2 <- lm(abs\_dfgpa~tutpart, data = peer\_tutoring)

### Summary statistics

##### Academic Performance

print(bind,  
 rtitle = "Summary Statistics of Academic Performance",  
 cnames = c("Total", "Not Tutored", "Tutored"))

##   
##   
## |Summary Statistics of Academic Performance |Total |Not Tutored |Tutored |  
## |:-------------------------------------------|:---------------------|:---------------------|:---------------------|  
## |\*\*Hours Tutored\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; min |0 |0.0 |0.5 |  
## |&nbsp;&nbsp; max |35 |0 |35 |  
## |&nbsp;&nbsp; mean (sd) |0.39 &plusmn; 2.23 |0.00 &plusmn; 0.00 |8.23 &plusmn; 6.28 |  
## |\*\*Term Credit Hours Earned\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; min |0 |0 |4 |  
## |&nbsp;&nbsp; median |16 |16 |15 |  
## |&nbsp;&nbsp; max |22 |22 |21 |  
## |&nbsp;&nbsp; mean (sd) |15.38 &plusmn; 2.53 |15.40 &plusmn; 2.54 |15.02 &plusmn; 2.36 |  
## |\*\*Final Grade on 4.0 scale\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; min |0 |0 |0 |  
## |&nbsp;&nbsp; median |3 |3.0 |2.7 |  
## |&nbsp;&nbsp; max |4 |4 |4 |  
## |&nbsp;&nbsp; mean (sd) |2.72 &plusmn; 1.01 |2.73 &plusmn; 1.01 |2.46 &plusmn; 0.96 |  
## |\*\*Previous Term Cumulative GPA\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; min |0.33 |0.330000 |1.378139 |  
## |&nbsp;&nbsp; max |4.02 |4.02 |4.01 |  
## |&nbsp;&nbsp; median |3.13 |3.138182 |3.002174 |  
## |&nbsp;&nbsp; mean (sd) |3.11 &plusmn; 0.51 |3.11 &plusmn; 0.51 |2.99 &plusmn; 0.49 |  
## |\*\*Change in GPA compared to Previous Term\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; min |-1 |-1 |-1 |  
## |&nbsp;&nbsp; max |4.151515 |4.1515151 |0.8445471 |  
## |&nbsp;&nbsp; median |-0.08410339 |-0.08085422 |-0.14987398 |  
## |&nbsp;&nbsp; mean (sd) |-0.13 &plusmn; 0.30 |-0.13 &plusmn; 0.30 |-0.18 &plusmn; 0.29 |  
## |\*\*SAT math\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; min |420 |420 |460 |  
## |&nbsp;&nbsp; max |800 |800 |800 |  
## |&nbsp;&nbsp; mean (sd) |678.56 &plusmn; 59.42 |680.16 &plusmn; 58.94 |646.82 &plusmn; 60.07 |  
## |\*\*High School Rank Index\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; min |34 |34 |47 |  
## |&nbsp;&nbsp; max |94 |94 |91 |  
## |&nbsp;&nbsp; mean (sd) |67.90 &plusmn; 6.29 |67.92 &plusmn; 6.29 |67.46 &plusmn; 6.12 |

##### Demographics / Courses Characteristics

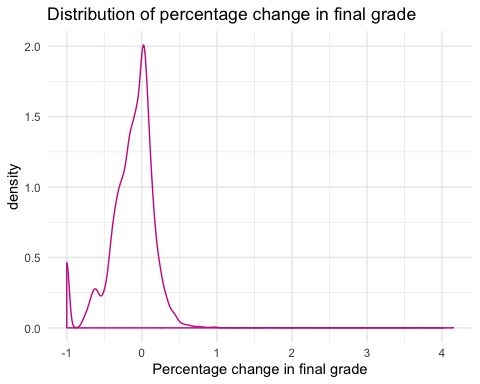
print(bind2,  
 rtitle = "Summary Statistics of Demographics and Courses Characteristics",  
 cnames = c("Total", "Not Tutored", "Tutored"))

##   
##   
## |Summary Statistics of Demographics and Courses Characteristics |Total |Not Tutored |Tutored |  
## |:--------------------------------------------------------------|:------------|:------------|:------------|  
## |\*\*Demographics\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; Varisity Athlete |2,010 (15) |1,854 (14) |156 (23) |  
## |&nbsp;&nbsp; Female |4,472 (32) |4,143 (31) |329 (50) |  
## |&nbsp;&nbsp; Male |9,363 (68) |9,028 (69) |335 (50) |  
## |&nbsp;&nbsp; African American |276 (2) |246 (2) |30 (5) |  
## |&nbsp;&nbsp; Hispanic |340 (2) |307 (2) |33 (5) |  
## |&nbsp;&nbsp; Greek Participant |5,176 (37) |4,885 (37) |291 (44) |  
## |&nbsp;&nbsp; Freshman |2,590 (19) |2,442 (19) |148 (22) |  
## |&nbsp;&nbsp; Sophoremore |7,440 (54) |7,071 (54) |369 (56) |  
## |\*\*Courses Characteristics\*\* |&nbsp;&nbsp; |&nbsp;&nbsp; |&nbsp;&nbsp; |  
## |&nbsp;&nbsp; Recitation Course |7,911 (57) |7,516 (57) |395 (59) |  
## |&nbsp;&nbsp; Lab Course |2,912 (21) |2,737 (21) |175 (26) |  
## |&nbsp;&nbsp; 2004-05 |4,876 (35) |4,685 (36) |191 (29) |  
## |&nbsp;&nbsp; 2005-06 |6,275 (45) |5,916 (45) |359 (54) |

### General Visualizations

#### Univariate graph of percentage change in final grade

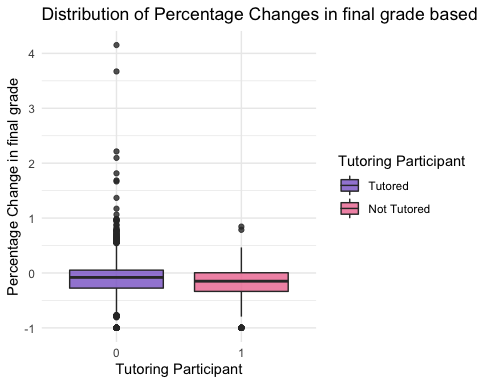
peer\_tutoring %>%  
 ggplot(aes(x=dfgpa))+  
 geom\_density(color="maroon3")+  
 theme\_minimal()+  
 xlab("Percentage change in final grade")+  
 ggtitle("Distribution of percentage change in final grade")



#### Bivariate graph of final grade based on tutoring

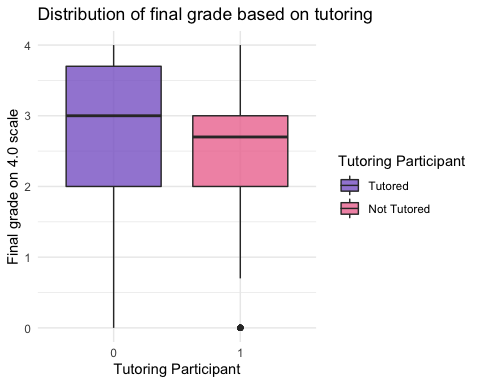
Bivariate graph of Percentage Change in final grade

peer\_tutoring %>%  
 ggplot(aes(y=dfgpa, x=as.factor(tutpart), fill=as.factor(tutpart)))+  
 geom\_boxplot(alpha=0.8)+  
 theme\_minimal()+  
 ylab("Percentage Change in final grade")+  
 xlab("Tutoring Participant")+  
 scale\_fill\_manual("Tutoring Participant",  
 labels=c("Tutored","Not Tutored"),  
 values=c("mediumpurple3","palevioletred2"))+  
 ggtitle("Distribution of Percentage Changes in final grade based on tutoring")



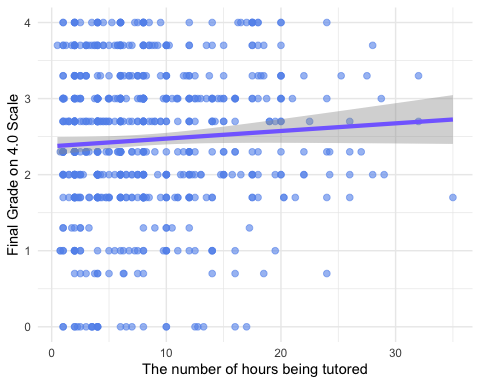
##### boxplot of final grade based on tutoring

peer\_tutoring %>%  
 ggplot(aes(y=fgrdnum, x=as.factor(tutpart), fill=as.factor(tutpart)))+  
 geom\_boxplot(alpha=0.8)+  
 theme\_minimal()+  
 ylab("Final grade on 4.0 scale")+  
 xlab("Tutoring Participant")+  
 scale\_fill\_manual("Tutoring Participant",  
 labels=c("Tutored","Not Tutored"),  
 values=c("mediumpurple3","palevioletred2"))+  
 ggtitle("Distribution of final grade based on tutoring")



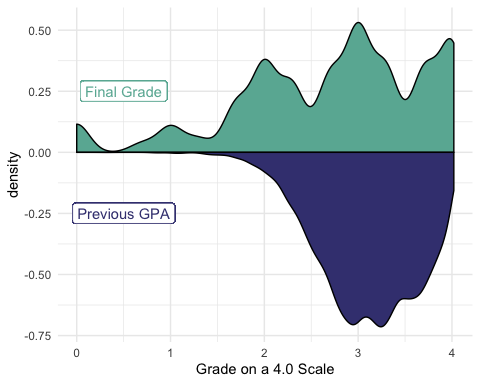
##### bivariate graph of final grade with tutoring hours within participants in the tutoring program

tutored <- peer\_tutoring[peer\_tutoring$tutpart == 1,]  
tutored %>%  
 ggplot(aes(x=tuthours,y=fgrdnum))+  
 geom\_point(color="cornflowerblue",   
 size = 2,   
 alpha = .6) +  
 geom\_smooth(method = "lm",  
 size = 1.5,  
 color = "slateblue1")+  
 xlab("The number of hours being tutored")+  
 ylab("Final Grade on 4.0 Scale")+  
 theme\_minimal()



##### Mirror density plot representing the distribution of previous term culumative gpa (on the bottom) and the distribution of final grade

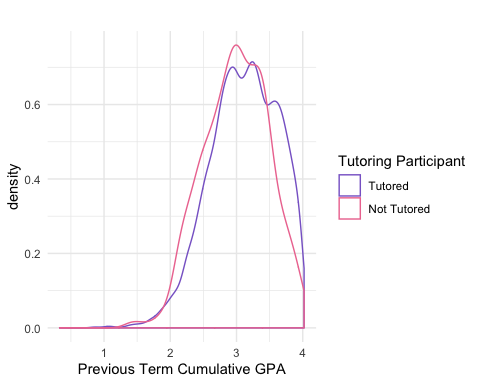
ggplot(peer\_tutoring,aes(x=x)) +  
 # Top  
 geom\_density( aes(x = fgrdnum,y=..density..), fill="#69b3a2" ) +  
 geom\_label( aes(x=0.5, y=0.25, label="Final Grade"), color="#69b3a2") +  
 # Bottom  
 geom\_density( aes(x = ptermcumgpa,y=-..density..), fill= "#404080") +  
 geom\_label( aes(x=0.5, y=-0.25, label="Previous GPA"), color="#404080") +  
 theme\_minimal() +  
 xlab("Grade on a 4.0 Scale")



### Visualizations based on treatment.

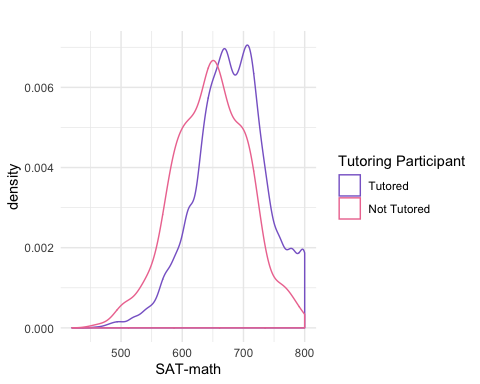
##### Previous Term GPA

peer\_tutoring %>%  
 ggplot(aes(x=ptermcumgpa, group=as.factor(tutpart), color=as.factor(tutpart)))+  
 geom\_density()+  
 theme\_minimal()+  
 xlab("Previous Term Cumulative GPA")+  
 scale\_color\_manual("Tutoring Participant",  
 labels=c("Tutored","Not Tutored"),  
 values=c("mediumpurple3","palevioletred2"))+  
 ggtitle("")



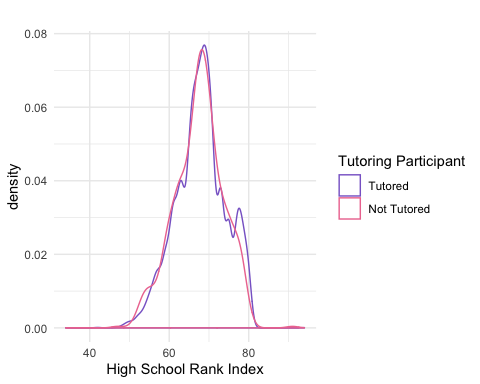
##### SAT-math

peer\_tutoring %>%  
 ggplot(aes(x=admsatm, group=as.factor(tutpart), color=as.factor(tutpart)))+  
 geom\_density()+  
 theme\_minimal()+  
 xlab("SAT-math")+  
 scale\_color\_manual("Tutoring Participant",  
 labels=c("Tutored","Not Tutored"),  
 values=c("mediumpurple3","palevioletred2"))+  
 ggtitle("")



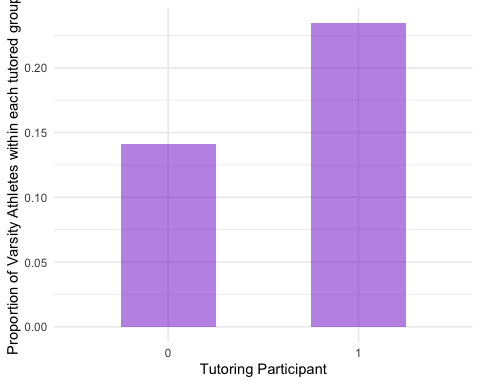
##### High school rank index

peer\_tutoring %>%  
 ggplot(aes(x=admri, group=as.factor(tutpart), color=as.factor(tutpart)))+  
 geom\_density()+  
 theme\_minimal()+  
 xlab("High School Rank Index")+  
 scale\_color\_manual("Tutoring Participant",  
 labels=c("Tutored","Not Tutored"),  
 values=c("mediumpurple3","palevioletred2"))+  
 ggtitle("")



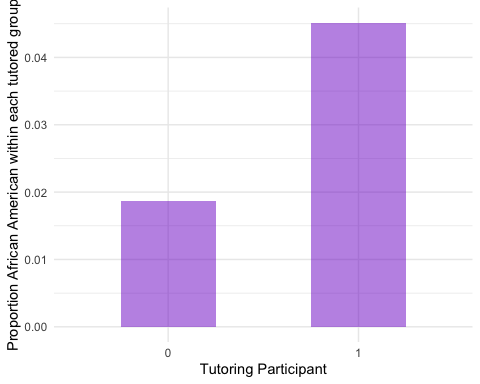
##### Varsity Athlete

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=athpart),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion of Varsity Athletes within each tutored group")+  
 theme\_minimal()



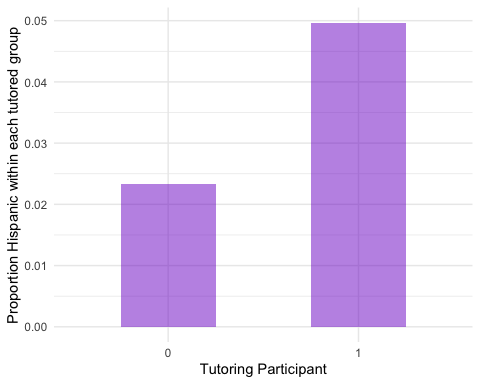
##### African-American

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=demafam),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion African American within each tutored group")+  
 theme\_minimal()



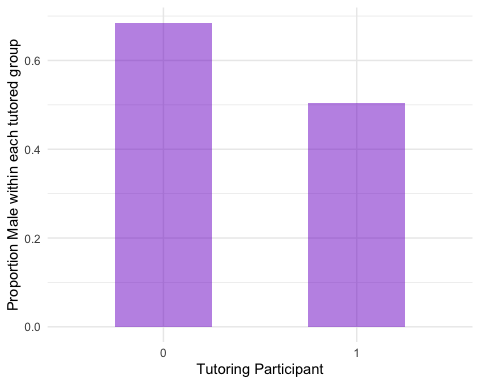
##### Hispanic

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=demhis),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Hispanic within each tutored group")+  
 theme\_minimal()



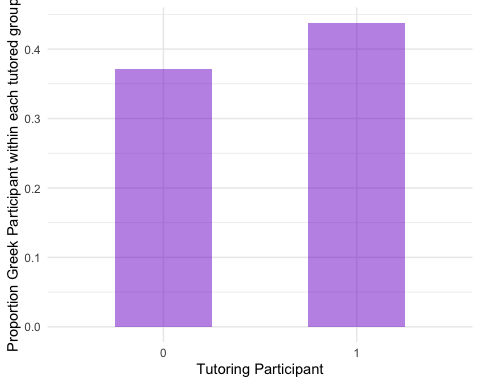
##### Male

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=demgender),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Male within each tutored group")+  
 theme\_minimal()



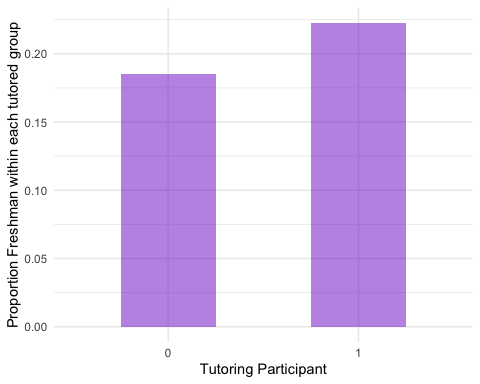
##### Greek Participant

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=demgrkpart),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Greek Participant within each tutored group")+  
 theme\_minimal()



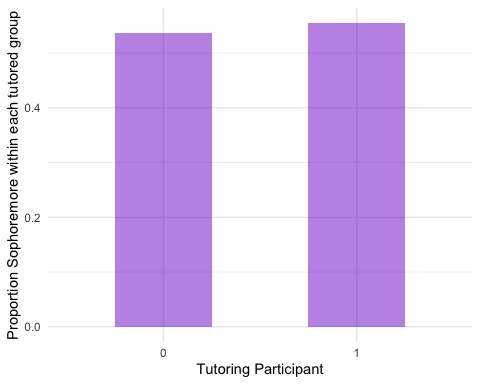
##### Freshman

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=acafr),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Freshman within each tutored group")+  
 theme\_minimal()



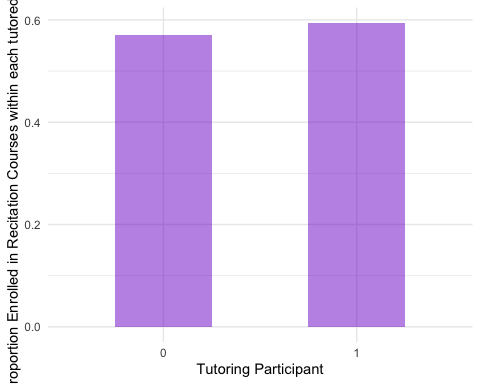
##### Sophomore

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=acaso),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Sophoremore within each tutored group")+  
 theme\_minimal()



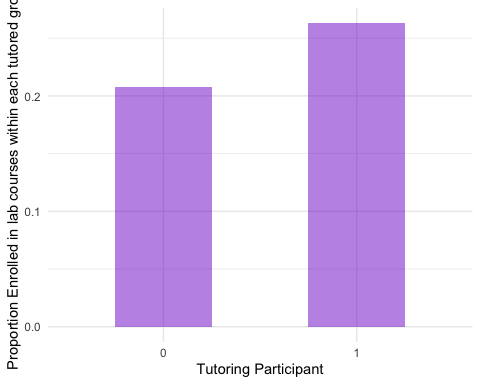
##### Recitation Courses

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=acarec),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Enrolled in Recitation Courses within each tutored group")+  
 theme\_minimal()



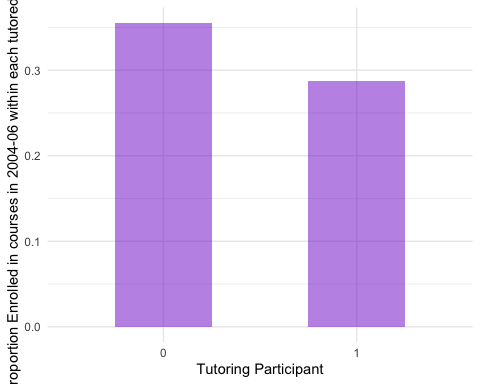
##### Lab courses

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=acalab),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Enrolled in lab courses within each tutored group")+  
 theme\_minimal()



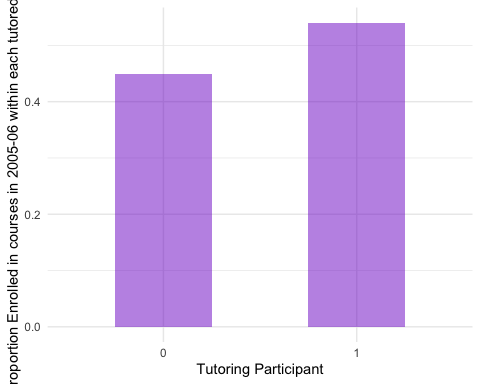
##### 2004-05

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=y1),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Enrolled in courses in 2004-06 within each tutored group")+  
 theme\_minimal()



##### 2005-06

ggplot(data=peer\_tutoring)+  
 stat\_summary(aes(x=factor(tutpart), y=y2),  
 fun.y="mean",   
 geom="bar",   
 fill="purple3",   
 alpha=0.5, width=0.5)+  
 xlab("Tutoring Participant")+  
 ylab("Proportion Enrolled in courses in 2005-06 within each tutored group")+  
 theme\_minimal()



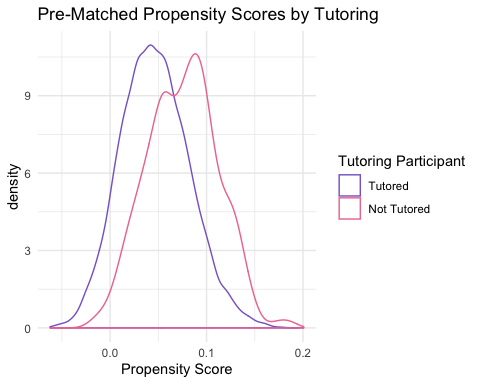
### Perform Matching

mydata <-   
 peer\_tutoring %>%  
 dplyr::select(-c("tutpart\_factor","tuthourssq","tuthourscu"))  
  
psmodel <- glm(tutpart~admsatm+admri+termcred+athpart+demafam+demhis+demgender+demgrkpart+acafr+acaso+ptermcumgpa+acalab+acarec+y1+y2, data=mydata)  
  
summary(psmodel)

##   
## Call:  
## glm(formula = tutpart ~ admsatm + admri + termcred + athpart +   
## demafam + demhis + demgender + demgrkpart + acafr + acaso +   
## ptermcumgpa + acalab + acarec + y1 + y2, data = mydata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.20066 -0.06850 -0.04265 -0.01724 1.01797   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.305e-01 2.863e-02 11.542 < 2e-16 \*\*\*  
## admsatm -3.371e-04 3.408e-05 -9.894 < 2e-16 \*\*\*  
## admri -1.113e-04 3.156e-04 -0.353 0.724325   
## termcred -1.096e-03 7.637e-04 -1.435 0.151406   
## athpart 2.139e-02 5.261e-03 4.066 4.81e-05 \*\*\*  
## demafam 2.445e-02 1.334e-02 1.832 0.066942 .   
## demhis 2.994e-02 1.172e-02 2.554 0.010657 \*   
## demgender -3.081e-02 4.046e-03 -7.616 2.79e-14 \*\*\*  
## demgrkpart 1.432e-02 3.774e-03 3.794 0.000149 \*\*\*  
## acafr 3.089e-03 5.801e-03 0.532 0.594395   
## acaso 1.035e-02 4.233e-03 2.445 0.014488 \*   
## ptermcumgpa -1.132e-02 4.199e-03 -2.696 0.007018 \*\*   
## acalab 2.540e-02 4.474e-03 5.678 1.39e-08 \*\*\*  
## acarec 3.715e-03 3.830e-03 0.970 0.332075   
## y1 -2.165e-02 3.948e-03 -5.484 4.24e-08 \*\*\*  
## y2 2.434e-02 3.872e-03 6.287 3.34e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.04443554)  
##   
## Null deviance: 632.13 on 13834 degrees of freedom  
## Residual deviance: 614.05 on 13819 degrees of freedom  
## AIC: -3798.2  
##   
## Number of Fisher Scoring iterations: 2

##### Propensity Scores before matching

pscore <- psmodel$fitted.values  
  
ggplot(mydata)+  
geom\_density(aes(x=pscore, color=factor(tutpart)))+  
 xlab("Propensity Score")+  
 scale\_color\_manual("Tutoring Participant",  
 labels=c("Tutored","Not Tutored"),  
 values=c("mediumpurple3","palevioletred2"))+  
 ggtitle("Pre-Matched Propensity Scores by Tutoring")+  
 theme\_minimal()



##### Table before matching

xvars <- c("admsatm","admri","termcred","athpart","demafam","demhis","demgender","demgrkpart","acafr","acaso","ptermcumgpa","acarec","acalab","y1","y2")  
#look at a table 1  
table1<- CreateTableOne(vars=xvars,strata="tutpart", data=mydata, test=FALSE)  
## include standardized mean difference (SMD)  
print(table1,smd=TRUE)

## Stratified by tutpart  
## 0 1 SMD   
## n 13171 664   
## admsatm (mean (SD)) 680.16 (58.94) 646.82 (60.07) 0.560  
## admri (mean (SD)) 67.92 (6.29) 67.46 (6.12) 0.074  
## termcred (mean (SD)) 15.40 (2.54) 15.02 (2.36) 0.157  
## athpart (mean (SD)) 0.14 (0.35) 0.23 (0.42) 0.243  
## demafam (mean (SD)) 0.02 (0.14) 0.05 (0.21) 0.151  
## demhis (mean (SD)) 0.02 (0.15) 0.05 (0.22) 0.141  
## demgender (mean (SD)) 0.69 (0.46) 0.50 (0.50) 0.375  
## demgrkpart (mean (SD)) 0.37 (0.48) 0.44 (0.50) 0.138  
## acafr (mean (SD)) 0.19 (0.39) 0.22 (0.42) 0.093  
## acaso (mean (SD)) 0.54 (0.50) 0.56 (0.50) 0.038  
## ptermcumgpa (mean (SD)) 3.11 (0.51) 2.99 (0.49) 0.244  
## acarec (mean (SD)) 0.57 (0.50) 0.59 (0.49) 0.049  
## acalab (mean (SD)) 0.21 (0.41) 0.26 (0.44) 0.132  
## y1 (mean (SD)) 0.36 (0.48) 0.29 (0.45) 0.146  
## y2 (mean (SD)) 0.45 (0.50) 0.54 (0.50) 0.184

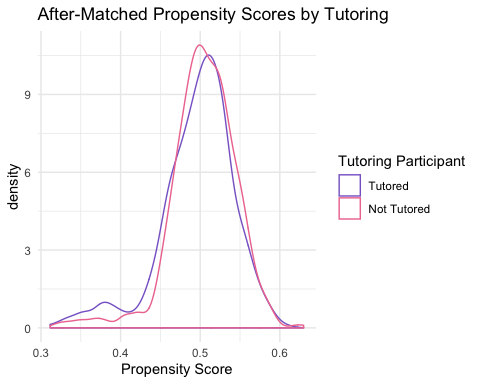
##### Perform greedy matching on propensity score

#do greedy matching on propensity score using Match with a caliper  
psmatch<-Match(Tr=mydata$tutpart,M=1,X=pscore,replace=FALSE,caliper=.2\*sd(pscore))  
matched<-mydata[unlist(psmatch[c("index.treated","index.control")]), ]  
#get standardized differences  
matchedtab1<-CreateTableOne(vars=xvars, strata ="tutpart",  
data=matched, test = FALSE)  
print(matchedtab1, smd = TRUE)

## Stratified by tutpart  
## 0 1 SMD   
## n 657 657   
## admsatm (mean (SD)) 649.67 (61.58) 648.03 (59.18) 0.027  
## admri (mean (SD)) 67.61 (6.17) 67.46 (6.10) 0.025  
## termcred (mean (SD)) 15.02 (2.46) 15.04 (2.37) 0.007  
## athpart (mean (SD)) 0.24 (0.42) 0.23 (0.42) 0.011  
## demafam (mean (SD)) 0.04 (0.18) 0.04 (0.20) 0.039  
## demhis (mean (SD)) 0.07 (0.26) 0.04 (0.20) 0.119  
## demgender (mean (SD)) 0.48 (0.50) 0.51 (0.50) 0.052  
## demgrkpart (mean (SD)) 0.43 (0.50) 0.44 (0.50) 0.025  
## acafr (mean (SD)) 0.19 (0.40) 0.22 (0.42) 0.068  
## acaso (mean (SD)) 0.57 (0.50) 0.56 (0.50) 0.021  
## ptermcumgpa (mean (SD)) 3.00 (0.51) 2.99 (0.49) 0.021  
## acarec (mean (SD)) 0.58 (0.49) 0.60 (0.49) 0.025  
## acalab (mean (SD)) 0.25 (0.43) 0.26 (0.44) 0.014  
## y1 (mean (SD)) 0.28 (0.45) 0.29 (0.45) 0.027  
## y2 (mean (SD)) 0.51 (0.50) 0.54 (0.50) 0.064

##### Propensity Scores after matching

psmodel2 <- glm(tutpart~admsatm+admri+termcred+athpart+demafam+demhis+demgender+demgrkpart+acafr+acaso+ptermcumgpa+acalab+acarec+y1+y2, data=matched)  
pscore2 <- psmodel2$fitted.values  
  
ggplot(matched)+  
geom\_density(aes(x=pscore2, color=factor(tutpart)))+  
 xlab("Propensity Score")+  
 scale\_color\_manual("Tutoring Participant",  
 labels=c("Tutored","Not Tutored"),  
 values=c("mediumpurple3","palevioletred2"))+  
 ggtitle("After-Matched Propensity Scores by Tutoring")+  
 theme\_minimal()



##### Paired wise t-test

y\_trt<-matched$abs\_dfgpa[matched$tutpart==1]  
y\_con<-matched$abs\_dfgpa[matched$tutpart==0]  
t.test(y\_trt, y\_con, paired=TRUE)

##   
## Paired t-test  
##   
## data: y\_trt and y\_con  
## t = -0.041853, df = 656, p-value = 0.9666  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.02654021 0.02543244  
## sample estimates:  
## mean of the differences   
## -0.0005538857

### Sub-analysis examining the asscoiation between tutored hours and changes in final grade within participants who were tutored

tutored <-   
 peer\_tutoring %>%  
 filter(tutpart == 1)  
  
#before controlling for other covariates  
model2 <- lm(dfgpa~tuthours,data = tutored)  
  
#after controlling for covariates  
model3 <- lm(dfgpa~tuthours+admsatm+admri+termcred+athpart+demafam+demhis+demgender+demgrkpart+acafr+acaso+ptermcumgpa+acalab+acarec, data = tutored)  
summary(model2)

##   
## Call:  
## lm(formula = dfgpa ~ tuthours, data = tutored)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.84340 -0.15334 0.03481 0.18936 1.01725   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.211332 0.018755 -11.268 <2e-16 \*\*\*  
## tuthours 0.003219 0.001813 1.776 0.0763 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.293 on 662 degrees of freedom  
## Multiple R-squared: 0.004739, Adjusted R-squared: 0.003236   
## F-statistic: 3.152 on 1 and 662 DF, p-value: 0.07627

summary(model3)

##   
## Call:  
## lm(formula = dfgpa ~ tuthours + admsatm + admri + termcred +   
## athpart + demafam + demhis + demgender + demgrkpart + acafr +   
## acaso + ptermcumgpa + acalab + acarec, data = tutored)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.90154 -0.15591 0.02264 0.18233 1.03824   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.802e-01 1.781e-01 -3.257 0.001185 \*\*   
## tuthours 2.827e-03 1.798e-03 1.573 0.116316   
## admsatm -7.758e-05 2.220e-04 -0.349 0.726843   
## admri -4.102e-04 2.027e-03 -0.202 0.839698   
## termcred 2.456e-02 4.859e-03 5.055 5.61e-07 \*\*\*  
## athpart 1.801e-02 2.848e-02 0.632 0.527298   
## demafam -1.968e-01 5.739e-02 -3.429 0.000645 \*\*\*  
## demhis 6.413e-03 5.276e-02 0.122 0.903304   
## demgender -2.331e-04 2.423e-02 -0.010 0.992326   
## demgrkpart -1.875e-02 2.350e-02 -0.798 0.425410   
## acafr -2.409e-02 3.470e-02 -0.694 0.487781   
## acaso -3.195e-02 2.853e-02 -1.120 0.263140   
## ptermcumgpa 3.372e-02 2.671e-02 1.263 0.207195   
## acalab 5.131e-02 2.639e-02 1.944 0.052272 .   
## acarec 4.301e-03 2.413e-02 0.178 0.858564   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2837 on 649 degrees of freedom  
## Multiple R-squared: 0.08547, Adjusted R-squared: 0.06574   
## F-statistic: 4.332 on 14 and 649 DF, p-value: 2.133e-07