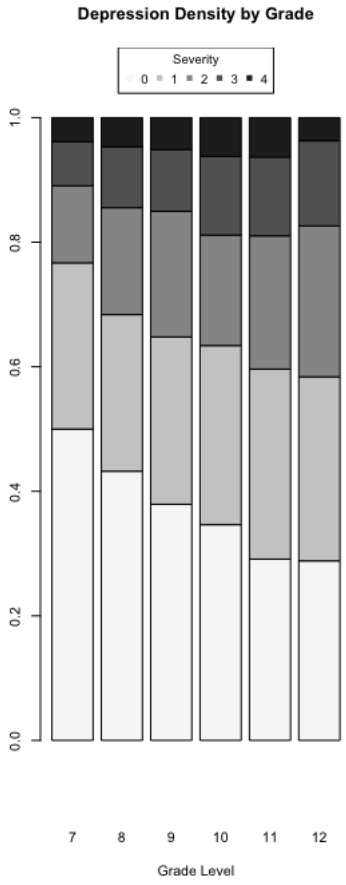
STAT 3280-HW1: He (yh9vhg), Lin (dl2de), Jin (zj5qj)

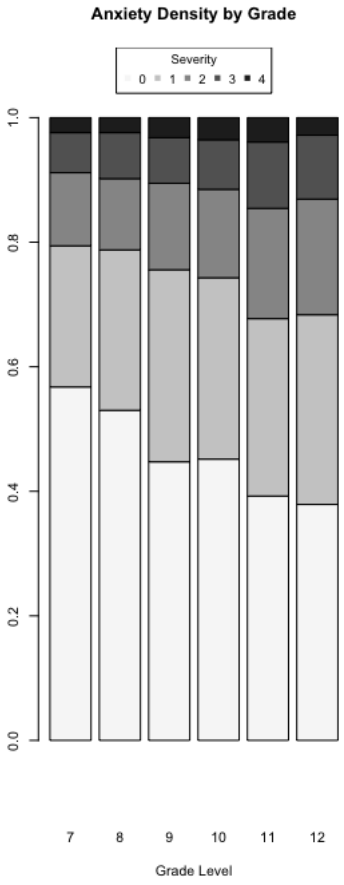
Question 1 Part 1

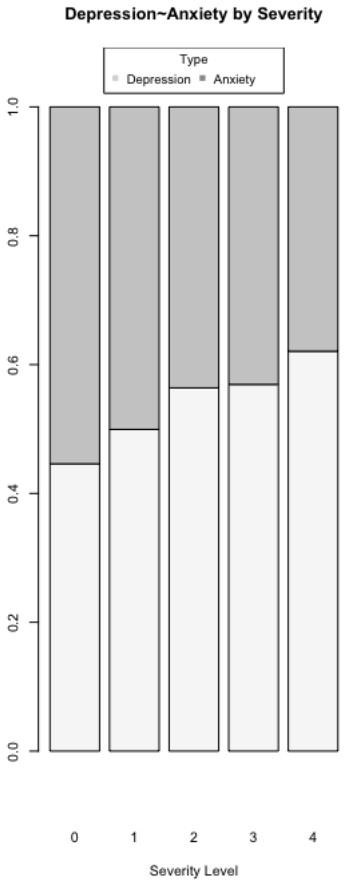
With the specific numbers we have, we decide to use a percentage stacked bar plot, a more straightforward graph instead of a regular bar plot, to demonstrate the percentage of depression and anxiety by level per grade.

As shown below, the first two plots represent the proportions of students with different severity conditions (Depression, Anxiety) from grades 7 through 12. The third plot shows, in each severity level, the proportions of those who have depression and anxiety.

In general, the trend among students with depression and anxiety appears to be similar with respect to grades. The proportion of level 0 students appears to decrease as grade increases, whereas those of levels 1~2 increase. Interestingly, there's a subtle spike in the proportions of levels 3~4 students in grades 10 and 11 and a sudden drop in grade 12. From the third plot, we can see that with the severity level increases, depression's proportion increases, which indicates depression is a more common problem among students who have psychological disorders.



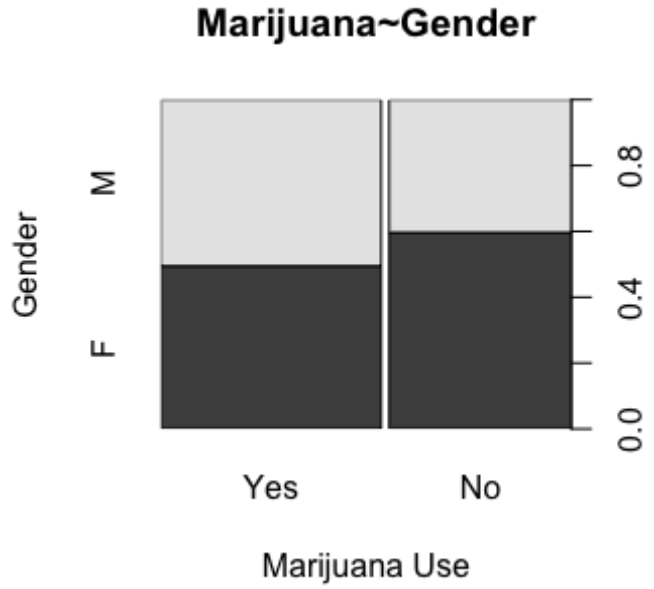


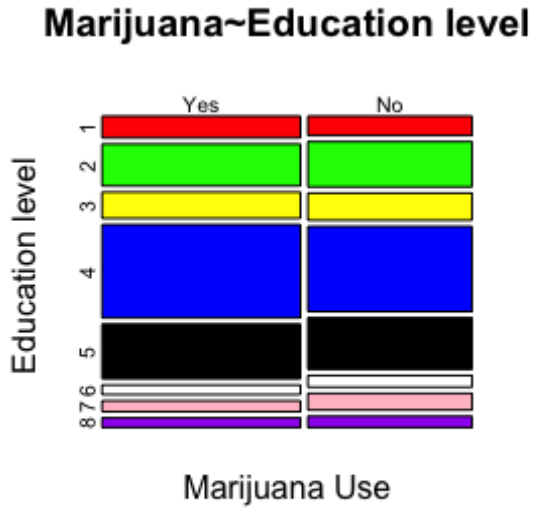


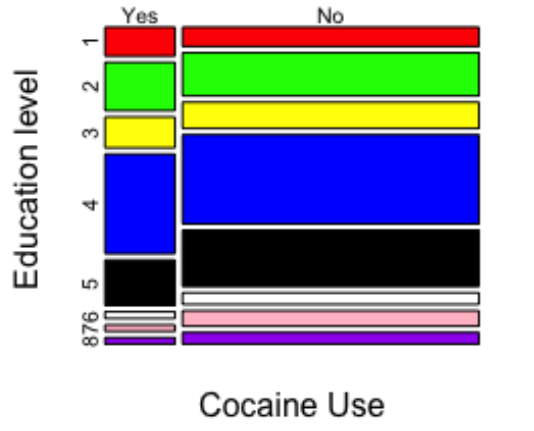
Question 1 Part 2

We want to clearly see the importance of each factor over drug usage, so we used a percentage stacked bar plot. These plots below examine the proportion of marijuana and cocaine usage across gender and education levels.

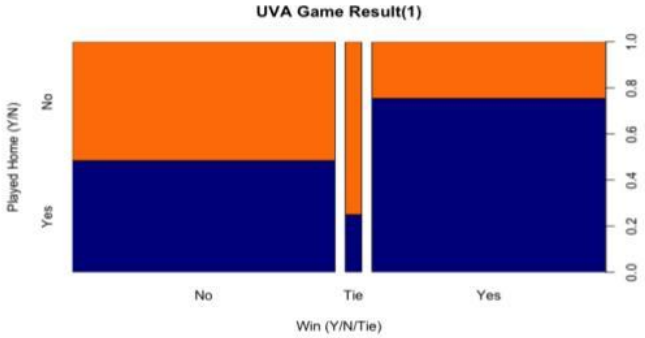
Marijuana appears to be more popular than cocaine within the same education level, and gender preference for marijuana usage is about the same, where male Cocaine users consist of around 60%. For both marijuana and cocaine usage, the distribution in education level amongst users and non-users appears to be similar, with the majority having levels 2, 4, and 5 education.





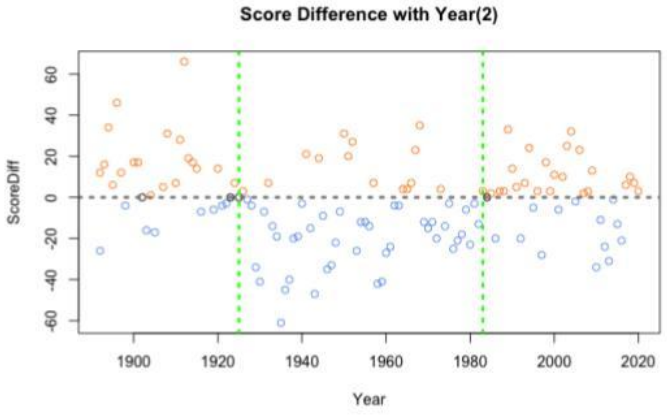


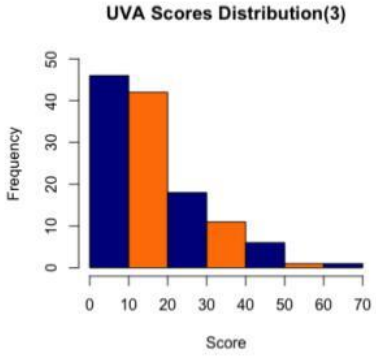
UVA Daily

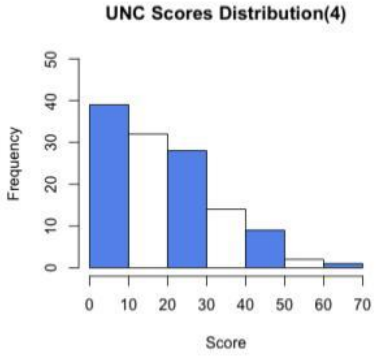


- UVA played most of the tied games away.
- Home/away status had a limited impact on its lost games.
- The majority of UVA's won games were played home.

- From the earliest game to around 1924, UVA dominated the game with much more wins and larger score difference; the same applies to UNC between 1925 and 1983.
- Since 1984, the score difference became smaller and the number of wins amongst the rival schools were balanced.

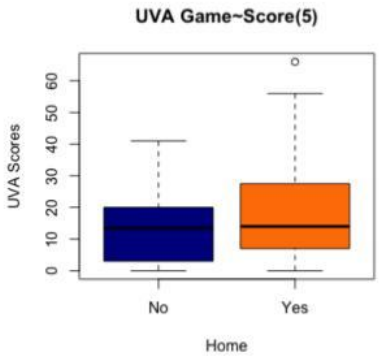


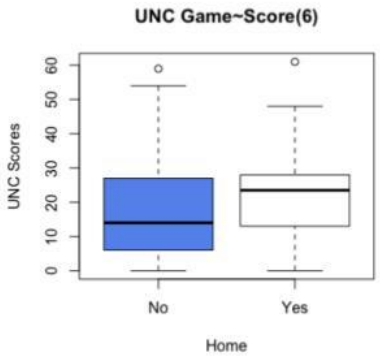




- Both UVA and UNC seem to have a right-tailed distribution of scores, but UVA is more skewed, as it has more games with scores of 20 or less.

- UNC has a smaller variation in score distributions in won and lost games;
- UNC also has a stronger home advantage compared to UVA, as UNC's average score varied significantly in different fields.





Appendix

```
### Question 1a
library(heplots)
data(AddHealth)
dt1 <- heplots::AddHealth

library(RColorBrewer)
colors <- brewer.pal(5,"Greys")
par(mfrow=c(1,3))

# margin=2 columns; margin=1 rows;
tb1 <- prop.table(table(dt1$depression,dt1$grade),margin=2)
tb2 <- prop.table(table(dt1$anxiety,dt1$grade),margin=2)
matrix1 <- rbind(table(dt1$depression),table(dt1$anxiety))
tb3 <- prop.table(as.table(matrix1),margin=2)

barplot(tb1,col=colors,xlab="Grade Level",
    main="Depression Density by Grade",asp=14,ylim=c(0:1))

legend("top",title="Severity",legend=0:4,
      pch=15,col=colors,cex=0.9,ncol=5)

barplot(tb2,col=colors,xlab="Grade Level",
       main="Anxiety Density by Grade",asp=14,ylim=(0:1))

legend("top",title="Severity",legend=0:4,
      pch=15,col=colors,cex=0.9,ncol=5)

barplot(tb3,col=colors,xlab="Severity Level",
       main="Depression~Anxiety by Severity",asp=12,ylim=(0:1))

legend("top",title="Type",legend=c("Depression","Anxiety"),
      pch=15,col=c("grey85","grey60"),cex=0.9,ncol=2)

### Question 1b
library(gscaLCA)
data(AddHealth)
dt2 <- gscaLCA::AddHealth
```

```
par(mfrow=c(1,2))
tbl4=table(dt2$Marijuana,dt2$Edu)
tbl5=table(dt2$Cocaine,dt2$Edu)

# marijuana
plot(dt2$Marijuana,dt2$Gender,xlab="Marijuana Use",ylab="Gender",
    cex.main=1,main="Marijuana~Gender")

mosaicplot(tbl4,xlab="Marijuana Use",ylab="Education level",
         main="Marijuana~Education level",
         col=c("red","green","yellow","blue","black",
             "white","pink","purple","cyan","gray","orange","brown","lightblue"))

# cocaine
plot(dt2$Cocaine,dt2$Gender,xlab="Cocaine Use",ylab="Gender",
    cex.main=1,main="Cocaine~Gender")
mosaicplot(tbl5,xlab="Cocaine Use",ylab="Education level",
         main="Cocaine~Education level",
         col=c("red","green","yellow","blue","black",
             "white","pink","purple","cyan","gray","orange","brown","lightblue"))

# Question 2
rivalry=read.csv("/Users/yhe/Desktop/Homework 1/Rivalry Data.csv")
rivalry1=data.frame(Year=rivalry$Year,City=as.factor(rivalry$City),
                 State=as.factor(rivalry$State),NC=rivalry$North.Carolina,
                 VA=rivalry$Virginia,VA_Win=as.factor(rivalry$VA_Win),
                 NC_Win=as.factor(rivalry$NC_Win),HomeVA=as.factor(rivalry$Home.VA),
                 HomeNC=as.factor(rivalry$Home.NC),ScoreDiff=rivalry$ScoreDiff)
colors <- brewer.pal(5,"Greys")

par(mfrow=c(3,2))

plot(x=rivalry1$VA_Win,y=rivalry1$HomeVA,xlab="Win (Y/N/Tie)",ylab="Played Home
(Y/N)",
    cex.main=1,main="UVA Game Result(1)",col=c("Dark Blue","Dark Orange1"))

plot(rivalry1$Year,rivalry1$ScoreDiff,xlab="Year",ylab="ScoreDiff",
    main="Score Difference with Year(2)",
    col=c("cornflowerblue","Black","Dark Orange1")[as.numeric(rivalry1$VA_Win)])
abline(h=0,col="grey 50",lty=3,lwd=3)
abline(v=1925,col="Green",lty=3,lwd=3)
abline(v=1983,col="Green",lty=3,lwd=3)
```

```
legend("topright",c("VA Won","NC Won","Tie"),
       pch=1,col=c("Dark Orange1","cornflowerblue","Black"),cex=0.7)

hist(rivalry$Virginia,col=c("Dark Blue","Dark Orange1"),
     xlab="Score", main="UVA Scores Distribution(3)",ylim=c(0,50))
hist(rivalry$North.Carolina,col=c("cornflowerblue","white"),
     xlab="Score",main="UNC Scores Distribution(4)",ylim=c(0,50))
plot(rivalry1$HomeVA,rivalry1$VA,xlab="Home",ylab="UVA Scores",
     main="UVA Game~Score(5)",col=c("Dark Blue","Dark Orange1"))
plot(rivalry1$HomeNC,rivalry1$NC,xlab="Home",ylab="UNC Scores",
     main="UNC Game~Score(6)",col=c("cornflowerblue","white"))
```