# 440 Reproducibility and Statistics Assignment Homework

politics<-read.csv("politics.csv")

Check the structure. subject, party, testtime, minwage, and sex need to be factors. optimismscore and income should be integer and numeric vectors. Fix any variables that aren't right. -->

str(politics)

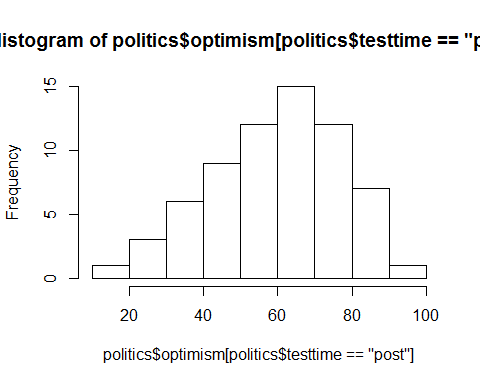
## 'data.frame': 132 obs. of 7 variables:  
## $ subject : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ party : Factor w/ 3 levels "democrat","independent",..: 3 3 2 2 2 3 3 2 3 2 ...  
## $ testtime : Factor w/ 2 levels "post","pre": 2 2 2 2 2 2 2 2 2 2 ...  
## $ optimismscore: int 52 51 69 51 61 31 57 48 42 64 ...  
## $ minwage : Factor w/ 2 levels "no","yes": 1 1 2 1 2 1 1 1 1 1 ...  
## $ sex : Factor w/ 2 levels "female","male": 2 2 1 2 2 2 2 2 2 2 ...  
## $ income : num 37.3 42.3 73 33.8 57.3 ...

politics$subject<-factor(politics$subject)  
politics$testtime<-factor(politics$testtime, levels=c("pre", "post"))

summary(politics$optimism)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 17.00 42.75 59.50 57.40 70.50 94.00

hist(politics$optimism[politics$testtime=="post"])



table(politics$party[politics$testtime=="pre"],politics$sex[politics$testtime=="pre"])

##   
## female male  
## democrat 14 12  
## independent 7 10  
## republican 12 11

chisq.test(politics$party[politics$testtime=="pre"],politics$sex[politics$testtime=="pre"])

##   
## Pearson's Chi-squared test  
##   
## data: politics$party[politics$testtime == "pre"] and politics$sex[politics$testtime == "pre"]  
## X-squared = 0.72674, df = 2, p-value = 0.6953

Political affiliation is independent of one's gender, Chi-square(2)= .72674, p=.6953

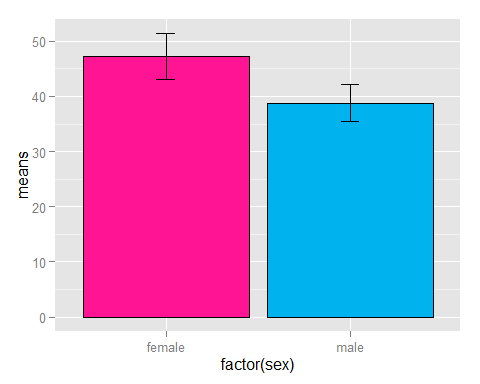
t.test(politics$income[politics$sex=="male"& politics$testtime=="pre"], politics$income[politics$sex=="female" & politics$testtime=="pre"])

##   
## Welch Two Sample t-test  
##   
## data: politics$income[politics$sex == "male" & politics$testtime == and politics$income[politics$sex == "female" & politics$testtime == "pre"] and "pre"]  
## t = -1.5714, df = 61.623, p-value = 0.1212  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -19.23627 2.30508  
## sample estimates:  
## mean of x mean of y   
## 38.80751 47.27310

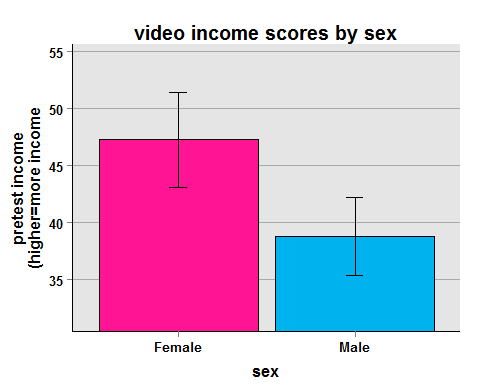
library("dplyr")

##   
## Attaching package: 'dplyr'  
##   
## The following object is masked from 'package:stats':  
##   
## filter  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

temp<-politics[politics$testtime=="pre",]%>%group\_by(sex)%>%summarize(means=mean(income),  
sems=sd(income)/sqrt(length(income)))  
  
library ("ggplot2")  
f<-ggplot(temp,aes(x=factor(sex),y=means))+  
 geom\_bar(stat="identity",color="black",fill=c("deeppink","deepskyblue2"))+geom\_errorbar(aes(ymax=means+sems,ymin=means-sems),width=.1)  
f



f<-f+ggtitle("video income scores by sex")+  
 labs(x="sex", y="pretest income\n(higher=more income")+  
 scale\_x\_discrete(breaks=c("female","male"),labels=c("Female","Male"))+  
 theme(plot.title=element\_text(size=15,face="bold",vjust=.5))+  
 theme(axis.title.x=element\_text(size=12,face="bold",vjust=-.25))+  
 theme(axis.title.y=element\_text(size=12,face="bold",vjust=1))+  
 theme(axis.text.x=element\_text(size=10,face="bold",color="black"))+  
 theme(axis.text.y=element\_text(size=10,face="bold",color="black"))+  
 coord\_cartesian(ylim=c(min(temp$means)-2\*max(temp$sems),max(temp$means)+2\*max(temp$sems)))+  
 theme(panel.border=element\_blank(),axis.line=element\_line())+  
 theme(panel.grid.major.x=element\_blank())+  
 theme(panel.grid.major.y=element\_line(color="darkgrey"))+  
 theme(panel.grid.minor.y=element\_blank())  
f



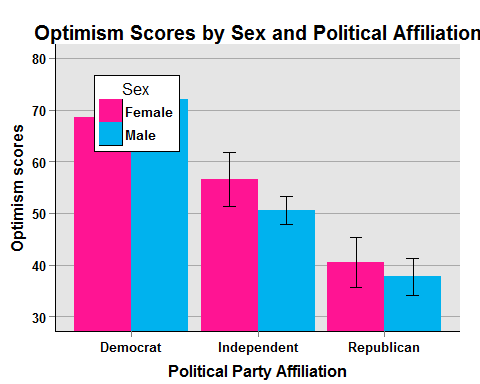
summary(aov(optimismscore~party\*sex,data=politics[politics$testtime=="pre",]))

## Df Sum Sq Mean Sq F value Pr(>F)   
## party 2 11856 5928 34.202 1.22e-10 \*\*\*  
## sex 1 20 20 0.118 0.732   
## party:sex 2 247 123 0.712 0.495   
## Residuals 60 10399 173   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

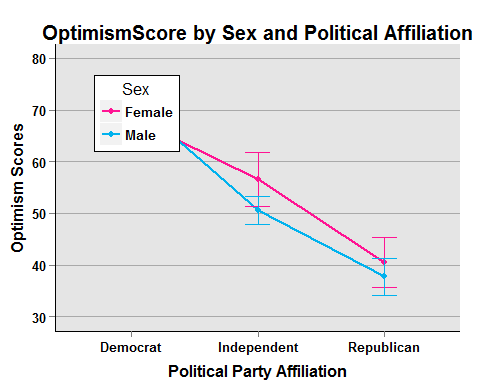
temp<-politics[politics$testtime=="pre",]%>%group\_by(party,sex)%>%  
 summarize(means=mean(optimismscore),sems=sd(optimismscore)/sqrt(length(optimismscore)))  
library("gplots")

##   
## Attaching package: 'gplots'  
##   
## The following object is masked from 'package:stats':  
##   
## lowess

col1=col2hex("deeppink")  
col2=col2hex("deepskyblue2")  
f<-ggplot(temp, aes(x=party, y=means, fill=sex))+  
 geom\_bar(stat="identity",position=position\_dodge())+  
 scale\_fill\_manual(values=c(col1,col2),name="Sex",breaks=c("female","male"),labels=c("Female", "Male"))+  
 theme(legend.key=element\_rect(color="black"))+  
 geom\_errorbar(aes(ymax=means+sems, ymin=means-sems),width=.2,position=position\_dodge(.9))+  
 ggtitle("Optimism Scores by Sex and Political Affiliation")+  
 labs(x="Political Party Affiliation",y="Optimism scores")+  
 scale\_x\_discrete(breaks=c("democrat","independent","republican"),labels=c("Democrat","Independent","Republican"))+  
 theme(plot.title=element\_text(size=15,face="bold",vjust=.5))+  
 theme(axis.title.x=element\_text(size=12,face="bold",vjust=-.25))+  
 theme(axis.title.y=element\_text(size=12,face="bold",vjust=1))+  
 theme(axis.text.x=element\_text(size=10,face="bold",color="black"))+  
 theme(axis.text.y=element\_text(size=10,face="bold",color="black"))+  
 coord\_cartesian(ylim=c(min(temp$means)-2\*max(temp$sems),max(temp$means)+2\*max(temp$sems)))+  
 theme(panel.border=element\_blank(),axis.line=element\_line())+  
 theme(panel.grid.major.x=element\_blank())+  
 theme(panel.grid.major.y=element\_line(color="darkgrey"))+  
 theme(panel.grid.minor.y=element\_blank())+  
 theme(legend.position=c(.2,.76))+  
 theme(legend.background=element\_blank())+  
 theme(legend.background=element\_rect(color="black"))+  
 theme(legend.title=element\_blank())+  
 theme(legend.title=element\_text(size=12))+  
 theme(legend.title.align=.5)+  
 theme(legend.text=element\_text(size=10,face="bold"))  
f



f<-ggplot(temp, aes(x=party, y=means, group=sex, color=sex))+  
 geom\_line(size=1)+  
 geom\_point(size=2)+  
 scale\_color\_manual(values=c(col1,col2),name="Sex",breaks=c("female","male"),labels=c("Female", "Male"))+  
 geom\_errorbar(aes(ymax=means+sems, ymin=means-sems),width=.2)+  
 ggtitle("OptimismScore by Sex and Political Affiliation")+  
 labs(x="Political Party Affiliation",y="Optimism Scores")+  
 scale\_x\_discrete(breaks=c("democrat","independent","republican"),labels=c("Democrat","Independent","Republican"))+  
 theme(plot.title=element\_text(size=15,face="bold",vjust=.5))+  
 theme(axis.title.x=element\_text(size=12,face="bold",vjust=-.25))+  
 theme(axis.title.y=element\_text(size=12,face="bold",vjust=1))+  
 theme(axis.text.x=element\_text(size=10,face="bold",color="black"))+  
 theme(axis.text.y=element\_text(size=10,face="bold",color="black"))+  
 coord\_cartesian(ylim=c(min(temp$means)-2\*max(temp$sems),max(temp$means)+2\*max(temp$sems)))+  
 theme(panel.border=element\_blank(),axis.line=element\_line())+  
 theme(panel.grid.major.x=element\_blank())+  
 theme(panel.grid.major.y=element\_line(color="darkgrey"))+  
 theme(panel.grid.minor.y=element\_blank())+  
 theme(legend.position=c(.2,.76))+  
 theme(legend.background=element\_blank())+  
 theme(legend.background=element\_rect(color="black"))+  
 theme(legend.title=element\_blank())+  
 theme(legend.title=element\_text(size=12))+  
 theme(legend.title.align=.5)+  
 theme(legend.text=element\_text(size=10,face="bold"))  
f



summary(aov(optimismscore~testtime\*sex+Error(subject/testtime),data=politics))

##   
## Error: subject  
## Df Sum Sq Mean Sq F value Pr(>F)  
## sex 1 80 80.4 0.119 0.731  
## Residuals 64 43105 673.5   
##   
## Error: subject:testtime  
## Df Sum Sq Mean Sq F value Pr(>F)   
## testtime 1 770.9 770.9 41.299 1.87e-08 \*\*\*  
## testtime:sex 1 0.9 0.9 0.049 0.825   
## Residuals 64 1194.7 18.7   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(lm(optimismscore~party,data=politics[politics$testtime=="pre",]))

##   
## Call:  
## lm(formula = optimismscore ~ party, data = politics[politics$testtime ==   
## "pre", ])  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -33.231 -7.924 -0.202 10.519 42.826   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 70.231 2.552 27.521 < 2e-16 \*\*\*  
## partyindependent -17.172 4.059 -4.231 7.68e-05 \*\*\*  
## partyrepublican -31.057 3.725 -8.338 9.16e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 13.01 on 63 degrees of freedom  
## Multiple R-squared: 0.5264, Adjusted R-squared: 0.5114   
## F-statistic: 35.01 on 2 and 63 DF, p-value: 5.963e-11

***Fin***