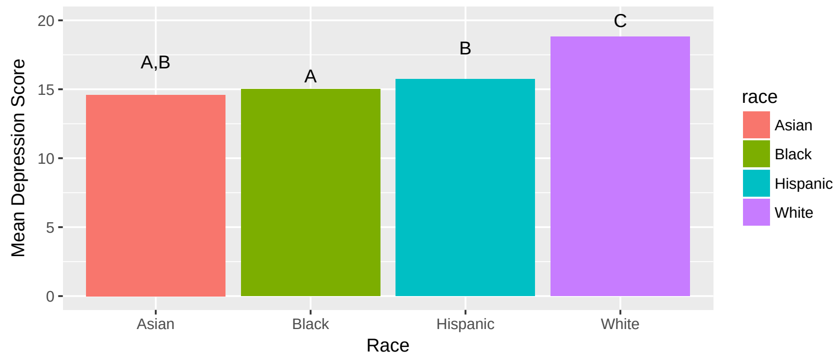
Some Graphing Alternatives

- **Visual Display of Post-Hoc Results**. Suppose you wanted to test the association between depression (quantitative) and race. You run an ANOVA and find that there is significant association between the two variables. You then run post-hoc tests to determine which groups vary. There is a convention that will allow you to do this. Groups with at least one letter in common are considered not significantly different from one another. Groups that have no overlapping letters are considered significantly different from each other.

```
aov.results<-aov(depression~race, data=mydata)
summary(aov.results)

##              Df Sum Sq Mean Sq F value   Pr(>F)
## race          3  543.1  181.02   24.25 2.16e-13 ***
## Residuals   196 1462.9    7.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(aov.results)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = depression ~ race, data = mydata)
##
## $race
##                      diff        lwr      upr     p adj
## Black-Asian    0.4173350 -0.9985203 1.833190 0.8706319
## Hispanic-Asian 1.1351555 -0.2806997 2.551011 0.1641065
## White-Asian    4.2056900  2.7898348 5.621545 0.0000000
## Hispanic-Black 0.7178206 -0.6980347 2.133676 0.5553106
## White-Black    3.7883550  2.3724998 5.204210 0.0000000
## White-Hispanic 3.0705345  1.6546792 4.486390 0.0000004
```

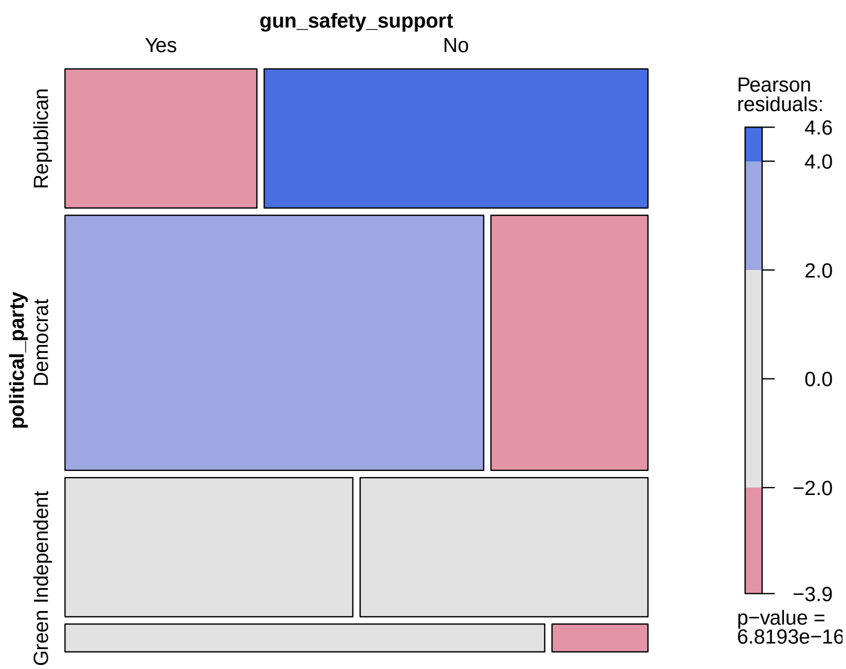
Notice which groups are significantly different from one another. Groups that are not significantly different from one another should have overlapping letters.

```
library(ggplot2)
ggplot()+
  stat_summary(aes(x=race, y=depression, fill=race), fun.y="mean", geom="bar")+
  geom_text(aes(x=c(1,2,3,4), y=c(17,16,18,20), label=c("A,B","A","B","C")))+
  ylab("Mean Depression Score")+
  xlab("Race")
```



- You can also use the code above to show post hoc results of a categorical/categorical relationship. You can do the same things you did above except you would utilize your chi-square post hoc results to determine letter designations.
- Another way of showing post-hoc results of a categorical to categorical relationship would be to would be to plot a mosaic plot. For this illustration suppose that we are examining the relationship between political party affiliation and support of a gun-safety bill. The plot below shows us that there is a significantly large proportion of Republicans who do not support the bill (since the No column is dark blue – likewise, there is a significantly low proportion of Republicans who do support the bill)

```
myChiSq<-chisq.test(final$political_party, final$gun_safety_support)
myChiSq$residuals

##                       final$gun_safety_support
## final$political_party       Yes        No
##           Republican  -3.947959  4.625265
##           Democrat     3.243375 -3.799803
##           Independent -1.264234  1.481124
##           Green        1.835014 -2.149826

library(vcd)
mosaic(~political_party+gun_safety_support, data=final,shade=TRUE, legend=TRUE)
```

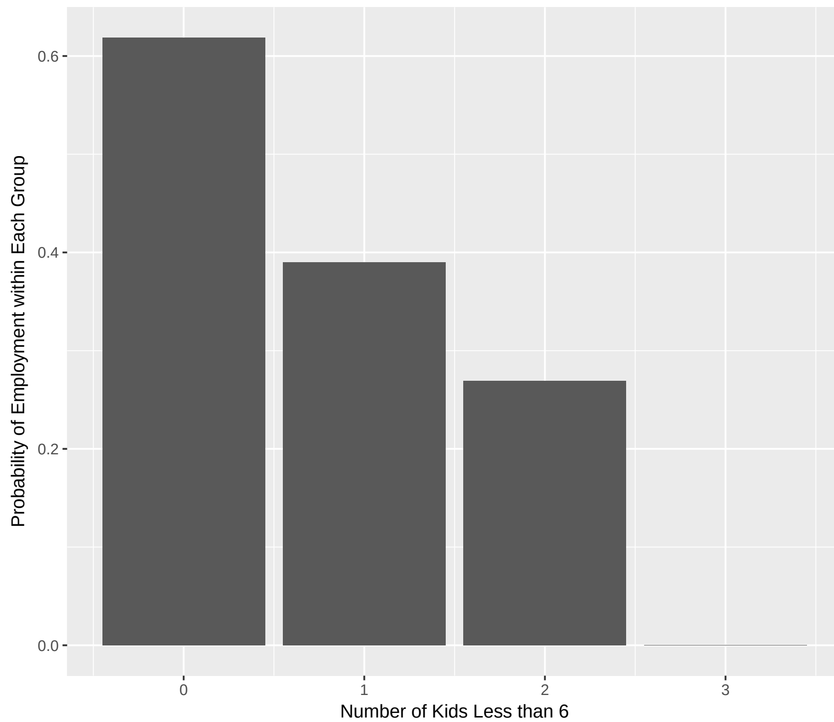


- Show output of logistic regression to obtain predicted probabilities. These are useful since odds ratios can be difficult to interpret.

```
library(foreign)
library(ggplot2)

# import stata data file
mroz<-read.dta("/Users/vnazzaro/Desktop/mroz.dta")


# Logistic regression to predict probability of employment based on
# experience and number of kids less than 6.

mroz.glm<- glm(inlf~ educ +kidslt6,
               data=mroz,
               family=binomial(link="logit"))
```

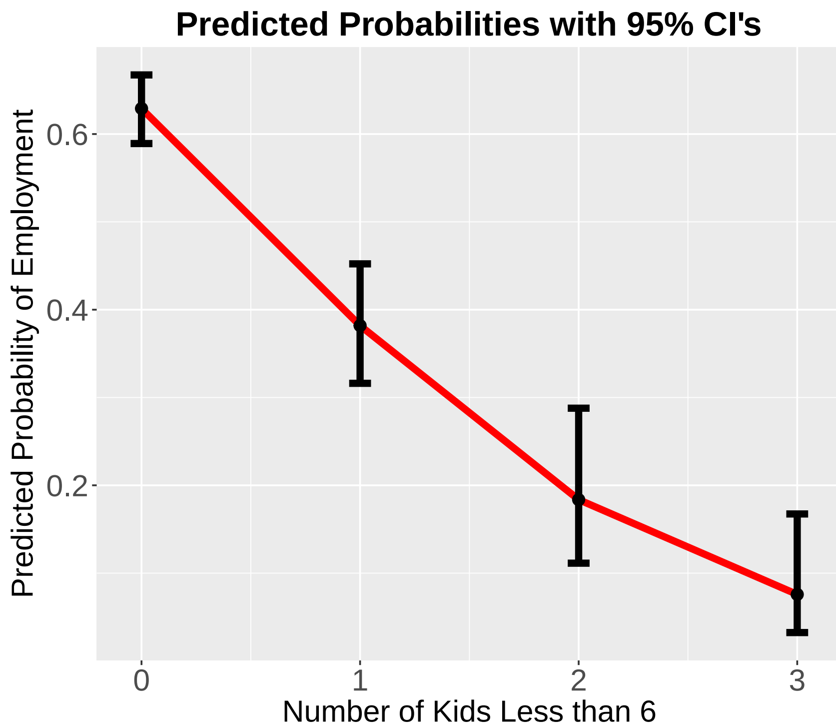
```r
summary(mroz.glm)

##
## Call:
## glm(formula = inlf ~ educ + kidslt6, family = binomial(link = "logit"),
##     data = mroz)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8531  -1.1978   0.6916   0.9861   2.0507
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.05395    0.44414  -4.625 3.75e-06 ***
## educ         0.21017    0.03648   5.761 8.34e-09 ***
## kidslt6     -1.01007    0.16261  -6.212 5.25e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1029.75  on 752  degrees of freedom
## Residual deviance:  958.34  on 750  degrees of freedom
## AIC: 964.34
##
## Number of Fisher Scoring iterations: 4

# Create a new data set that contains values I wish
#to vary along with mean values of other variables.
#To illustrate, I want to show how the number of kids
#less than 6 relates the probability of employment
#To vary only this, I will select a value of
#experience to hold fixed.


margins1 <- data.frame(educ = mean(mroz$educ),
                       kidslt6 = c(0,1, 2, 3))


# Make a dataset with predicted probability of
#response and also lower and upper limits
marginplot1 <- cbind(margins1, predict(mroz.glm, newdata = margins1, type = "link",
                                       se = TRUE))
marginplot1<- cbind(marginplot1, PredictedProb=plogis(marginplot1$fit),
                     LL=plogis(marginplot1$fit-1.96*marginplot1$se.fit),
```

```
                      UL=plogis(marginplot1$fit+1.96*+marginplot1$se.fit))


# Plot predicted probabilities.
# The theme layer is being used to increase font sizes and center title. This
# will make it easier to read on a poster. You can play around with the sizes here.

ggplot(data=mroz)+
  stat_summary(aes(x=kidslt6, y=inlf), geom="bar",fun.y="mean")+
  xlab("Number of Kids Less than 6")+
  ylab("Probability of Employment within Each Group")
```




```
## Plot predicted probabilities with confidence intervals.
ggplot(data=marginplot1,
       aes(x=kidslt6,  y=PredictedProb))+
  geom_line(color="red", size=2)+
  geom_errorbar(aes(ymin=LL, ymax=UL),width=0.1, size=2)+
  geom_point(color="black", size=3)+
  xlab("Number of Kids Less than 6")+
  ylab("Predicted Probability of Employment")+
  ggtitle("Predicted Probabilities with 95% CI's")+
```

```
  theme( plot.title=element_text(size=20,face="bold", hjust=0.5),
         axis.text=element_text(size=18),
         axis.title=element_text(size=18))
```



- If we want to vary two variables (to show how kids1to6 and years of education) both contribute to the model predictions (we would vary those two variables and hold all others fixed - if there were any more variables in our model).

```
# Prepare data to be plotted
margins2 <- data.frame(educ = rep(c(0,5,10,15,20),4),
                       kidslt6 = rep(c(0,1, 2, 3),5))


# Make a dataset with predicted probability of
#response and also lower and upper limits with
#standard errors
marginplot2 <- cbind(margins2,
                     predict(mroz.glm, newdata = margins2, type = "link",se = TRUE))
marginplot2<-cbind(marginplot2,
                   PredictedProb=plogis(marginplot2$fit),
                   LL=plogis(marginplot2$fit-1.96*marginplot2$se.fit),
```

```
                   UL=plogis(marginplot2$fit+1.96*marginplot2$se.fit))

## plot predicted probabilities across educ values for each level of kidslt6

ggplot(data=marginplot2, aes(x=educ,  y=PredictedProb,
                             color=as.factor(kidslt6)))+
  geom_line(size=1.5)+
  geom_errorbar(aes(ymin=LL, ymax=UL),
                width=0.8, size=1.5)+
  geom_point(size=1.5)+
  xlab("Years of Education")+
  ylab("Predicted Probability of Employment")+
  ggtitle("Adjusted Predictions with 95% CI's")+
  theme( plot.title=element_text(size=20,face="bold", hjust=0.5),
         axis.text=element_text(size=18),
         axis.title=element_text(size=18))+
  scale_color_manual("Number of Kids Less than 6",
                     values=c("blue","orange","green","red"))
```