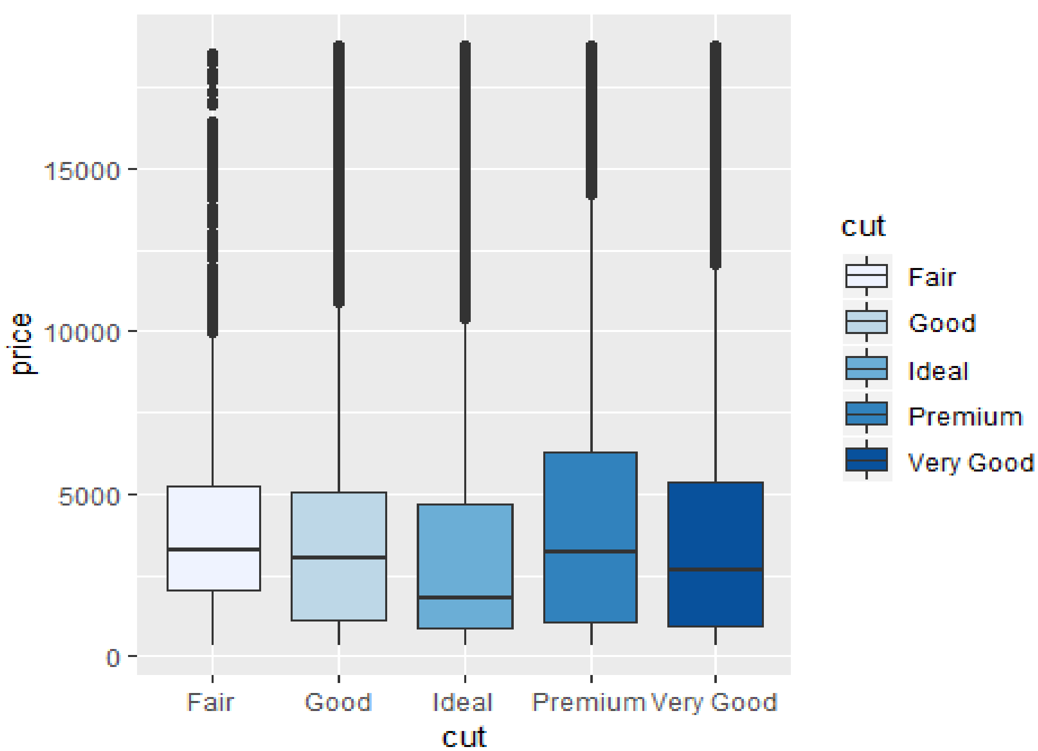
Jamie Burke
Ecostats HW3

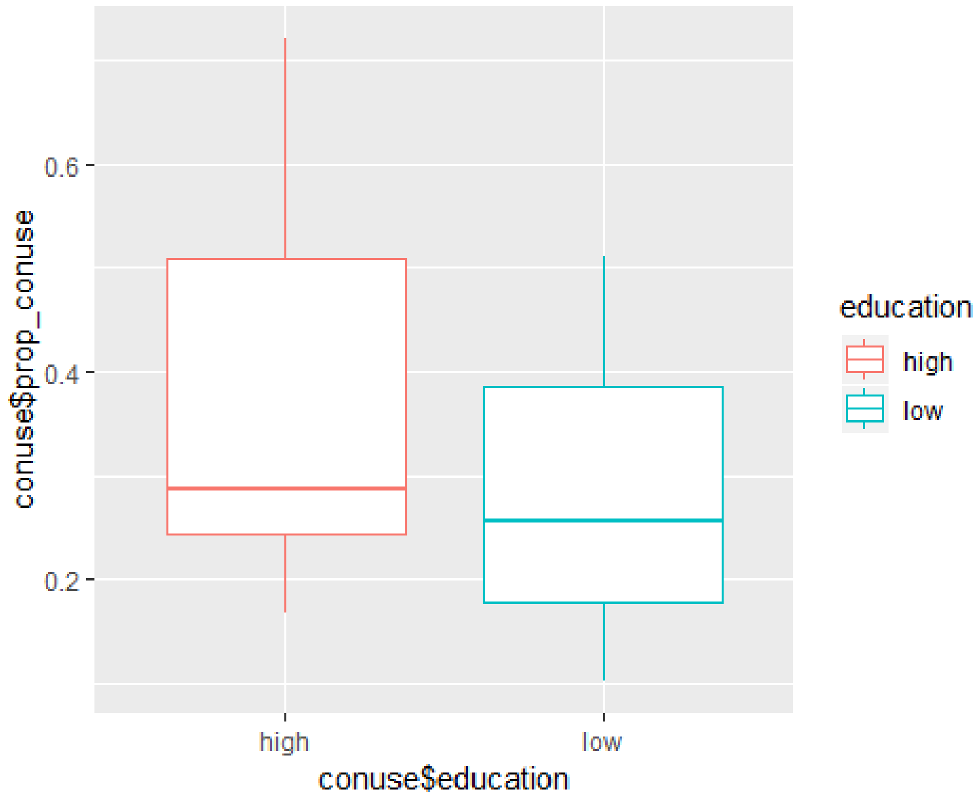
```
####Q1#####

diamonds<-read.csv("diamond.csv")
head(diamonds)

##   price       cut carat
## 1   326     Ideal  0.23
## 2   326   Premium  0.21
## 3   327      Good  0.23
## 4   334   Premium  0.29
## 5   335      Good  0.31
## 6   336 Very Good  0.24

diamond_mod<-glm(diamonds$price~diamonds$cut, family="poisson")
coef(diamond_mod)

##           (Intercept)      diamonds$cutGood     diamonds$cutIdeal
##             8.3799424            -0.1038367            -0.2316292
##   diamonds$cutPremium diamonds$cutVery Good
##             0.0504411            -0.0904632

confint(diamond_mod)

##                             2.5 %      97.5 %
## (Intercept)            8.37920242  8.38068216
## diamonds$cutGood      -0.10470072 -0.10297248
## diamonds$cutIdeal     -0.23240302 -0.23085517
## diamonds$cutPremium    0.04966133  0.05122103
## diamonds$cutVery Good -0.09125511 -0.08967112

exp(8.37994)#average price for fair

## [1] 4358.747

exp(-0.1038367) #good

## [1] 0.9013725 → shows a 9.86% average decrease in price from fair to good, a
$429.77 decrease

exp(-0.2316292)#ideal
```

```
## [1] 0.7932402 → shows a 20.67% average decrease in price from fair to ideal, a $900.95 decrease

exp(0.0504411)#premium

## [1] 1.051735 → shows a 5.17% average increase from fair to premium, a $225.35 increase

exp(-0.0904632)#verygood

## [1] 0.913508 → shows a 8.65% average decrease from fair to very good, a $376.99 decrease

library(ggplot2)

ggplot(diamonds, aes(x=cut, y=price, fill=cut)) +
  geom_boxplot()+
  scale_fill_brewer(palette="Blues")
```



Although the confidence interval does not overlap 0 and so the results show a significance in price between cuts, the average price for a diamond for each cut still seems very close and when looking at the plot, you can see a lot of points in the high price range for each. This leads me to think there is another factor that determines price, and in this dataset, you have another variable of carat size. Carat size should be explored as a variable that determines price as well as cut.

```r
#####Q2#####

conuse<-read.csv("contraception.csv")
head(conuse)


##     age education notUsing using Total
## 1   <25       low       53     6    59
## 2   <25       low       10     4    14
## 3   <25      high      212    52   264
## 4   <25      high       50    10    60
## 5 25-29       low       60    14    74
## 6 25-29       low       19    10    29


#create response variable

conuse$prop_conuse<-conuse$using/conuse$Total
conresponse<-cbind(conuse$using,conuse$notUsing)
conuse_mod<-glm(conresponse~conuse$education, family="binomial")
coef(conuse_mod)


##         (Intercept) conuse$educationlow
##         -0.81020374          0.09248529


confint(conuse_mod)


##                          2.5 %     97.5 %
## (Intercept)         -0.9460962 -0.6766394
## conuse$educationlow -0.1239481  0.3078275


plogis(-0.81020374)


## [1] 0.3078471



plogis(-0.81020374)-plogis(-0.81020374+0.09248529)

## [1] -0.02004851


ggplot(conuse, aes(x=conuse$education, y=conuse$prop_conuse, color=education))+
  geom_boxplot()
```



There is only a 2% average increase in proportion of people using contraception between low and high education. Since the 95% CI overlaps zero, we must assume that education level does not have a significant effect on contraception use.

```
####Q3#####

hurhim<-read.csv("Hurricane.csv")
head(hurhim)
```

```
##   Year     Name  MasFem MinPressure_before Minpressure_Updated.2014
## 1 1950     Easy 6.77778                958                      960
## 2 1950     King 1.38889                955                      955
## 3 1952     Able 3.83333                985                      985
## 4 1953  Barbara 9.83333                987                      987
## 5 1953 Florence 8.33333                985                      985
## 6 1954    Carol 8.11111                960                      960
##   Gender_MF Category alldeaths  NDAM Elapsed.Yrs Source  ZMasFem
## 1         F        3         2  1590          63    MWR -0.00094
## 2         M        3         4  5350          63    MWR -1.67076
## 3         M        1         3   150          61    MWR -0.91331
## 4         F        1         1    58          60    MWR  0.94587
## 5         F        1         0    15          60    MWR  0.48108
## 6         F        3        60 19321          59    MWR  0.41222
##   ZMinPressure_A    ZNDAM
## 1       -0.35636 -0.43913
```

```
## 2       -0.51125 -0.14843
## 3        1.03765 -0.55047
## 4        1.14091 -0.55758
## 5        1.03765 -0.56090
## 6       -0.25310  0.93174
```

```
hurhim_mod<-glm(hurhim$alldeaths~hurhim$Gender_MF, family="poisson")
coef(hurhim_mod)
```

```
##       (Intercept) hurhim$Gender_MFM
##         3.1679220        -0.5123354
```

```
confint(hurhim_mod)
```

```
##                         2.5 %     97.5 %
## (Intercept)        3.1164152  3.2185581
## hurhim$Gender_MFM -0.6211542 -0.4056501
```

```
exp(3.1679220)
```

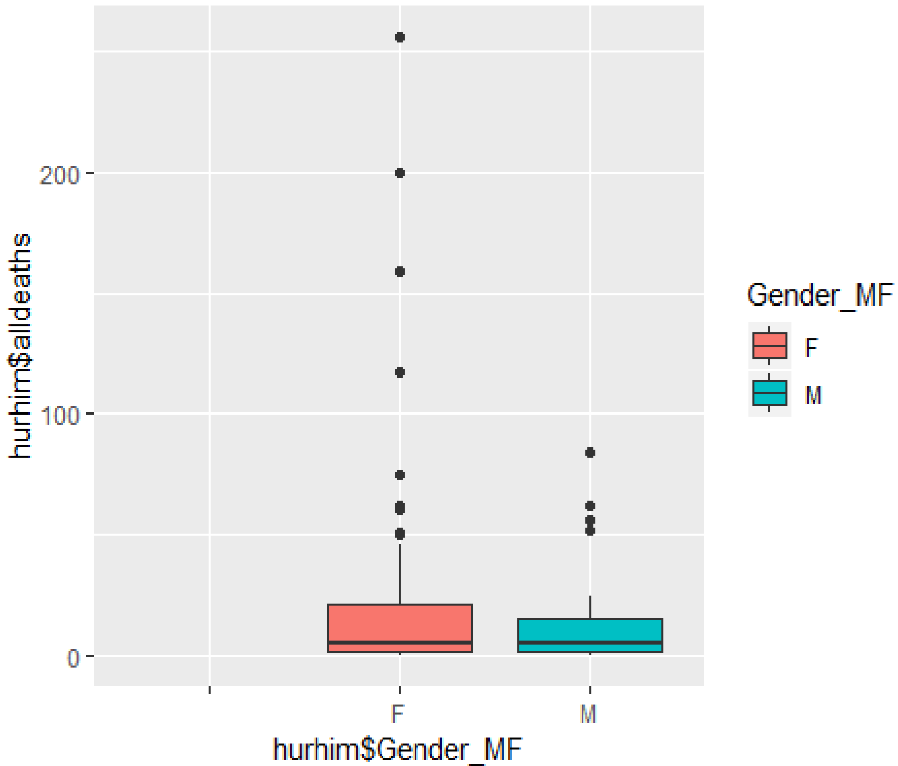
```
## [1] 23.75806
```

```
exp(-0.5123354)
```

```
## [1] 0.5990948 → shows a 40.09% decrease in deaths for male named himmicanes from
female named hurricanes which is on average 10 less deaths.
```

```
ggplot(hurhim, aes(x=hurhim$Gender_MF, y=hurhim$alldeaths, fill=Gender_MF))+
  geom_boxplot()
```

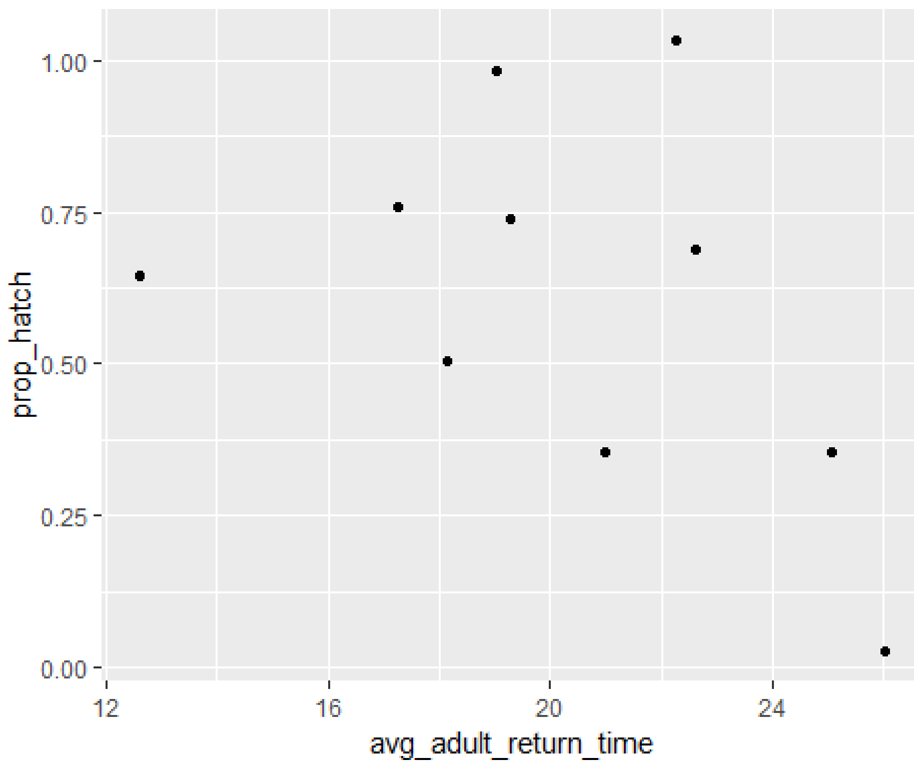
```
## Warning: Removed 6 rows containing non-finite values (stat_boxplot).
```



Although the 95%CI do not overlap zero and show a significant difference in number of deaths between hurricanes and himmicanes, I think the author could reanalyze the data with a better model, possibly a negative binomial model would fit better.

```
#####Q4#####
#does prop of eggs hatch depend on the average time for adult to return to nest

ibis2<-read.csv("ibisnest_summer2018.csv")
head(ibis2)
##   ï..Nest_ID total_eggs hatched first_hatch_date avg_adult_return_time
## 1          1          4       3         6/7/2018                    19
## 2          2          3       0             <NA>                    26
## 3          3          3       1        6/11/2018                    25
## 4          4          3       1        6/11/2018                    21
## 5          5          3       2        6/12/2018                    13
## 6          6          4       3         6/8/2018                    17

ibis2$prop_hatch<-ibis2$hatched/ibis2$total_eggs

response_hatch<-cbind(ibis2$hatched, ibis2$total_eggs-ibis2$hatched)

ibishatch_mod<-glm(response_hatch~ibis2$avg_adult_return_time, family="binomial")
coef(ibishatch_mod)

##                 (Intercept) ibis2$avg_adult_return_time
##                    4.381345                   -0.191110

confint(ibishatch_mod)
```

```
##                                    2.5 %     97.5 %
## (Intercept)                    0.01286491 9.71868150
## ibis2$avg_adult_return_time -0.44209654 0.01897317

plogis(4.381345)

## [1] 0.987646

plogis(4.381345)-plogis(4.381345-0.191110)

## [1] 0.002562851


ggplot(ibis2, aes(x=avg_adult_return_time, y=prop_hatch))+
  geom_jitter()
```



With the 95%CI overlapping zero, we can determine there is no significant difference of the average adult return to nest time on the proportion of eggs hatching. The average proportion hatched with a zero adult return time is 0.9876 but the 95%CI does have a very large range for that proportion. With a 1 min increase in adult return time, on average, the proportion of eggs hatched decreases by 0.25%.