Homework 09 Two Way ANOVA / ANCOVA / GLM

Due by 11:59pm, Friday, April 18, 2025, 11:59pm

S&DS 230/530/ENV 757

This assignment uses data from the International Social Survey Program on Environment from 2020 There are 224 questions from 21718 individuals across 14 countries.

The data you'll need is here. Be aware that it will take a few moments to load this data.

You'll also want the codebook that describes the variables. The final pages of the codebook have information on how to translate variable names to questions in the survey.

1) Data Set creation *(23 pts - 3 pts each section, except part 1.5 which is 5 pts)*

1.1) Read the data into an object called `envdat` (do NOT use the option `as.is = TRUE`). Check the dimension to be sure the data loaded correctly. Then create a new object called `envdat2` which only contains information for the following countries : Austria, Iceland, Japan, New Zealand, Philippines, Russia, and Thailand. The variable that contains country is `country`. You'll need to use the ISO 3166 Standard Book of Country Codes which you can find HERE.

Check the dimensions of your results - you should have 9476 observations.

```
envdat <- read.csv("https://raw.githubusercontent.com/jreuning/sds230_data/refs/heads/main/ISS_2020.csv
dim(envdat)
```

```
## [1] 21718   224
```

```
envdat2 <- envdat[envdat$country %in% c(40, 352, 392, 554, 608, 643, 764),]
dim(envdat2)
```

```
## [1] 9476  224
```

1.2) Create a new variable called `Country` on `envdat2` which has Country names rather than Country numbers. There are several ways to do this, but I suggests you use the `recode()` function in the `car` package. The syntax for this function is something like

```
library(car)
NewVar <- recode(envdat2$country, "1 = 'Transylvania'; 2 = 'Bezerkystan';
                 3 = 'Isle Joyeuse'")
```

Once you're created the variable, make a table of the resulting variable to see how many observations there are from each country.

```
envdat2["Country"] <- recode(envdat2$country,
                             "40 = 'Austria';
                             352 = 'Iceland';
                             392 = 'Japan';
```

```
                              554 = 'New Zealand';
                              608 = 'Phillipines';
                              643 = 'Russia';
                              764 = 'Thailand'")
table(envdat2["Country"])
```

```
## Country
##     Austria     Iceland       Japan New Zealand Phillipines      Russia
##        1261        1150        1491         993        1500        1583
##    Thailand
##        1498
```

1.3) Make a variable `Gender` on `envdat2` that contains gender (which is variable `SEX`). Recode so that 1 becomes 'Male' and 2 becomes 'Female'. Again, make a table of resulting variable to see how many people identify as Male and how many as Female.

```
envdat2["Gender"] <- recode(envdat2$SEX,
                           "1 = 'Male';
                           2 = 'Female'")
table(envdat2["Gender"])
```

```
## Gender
## Female   Male
##   5032   4444
```

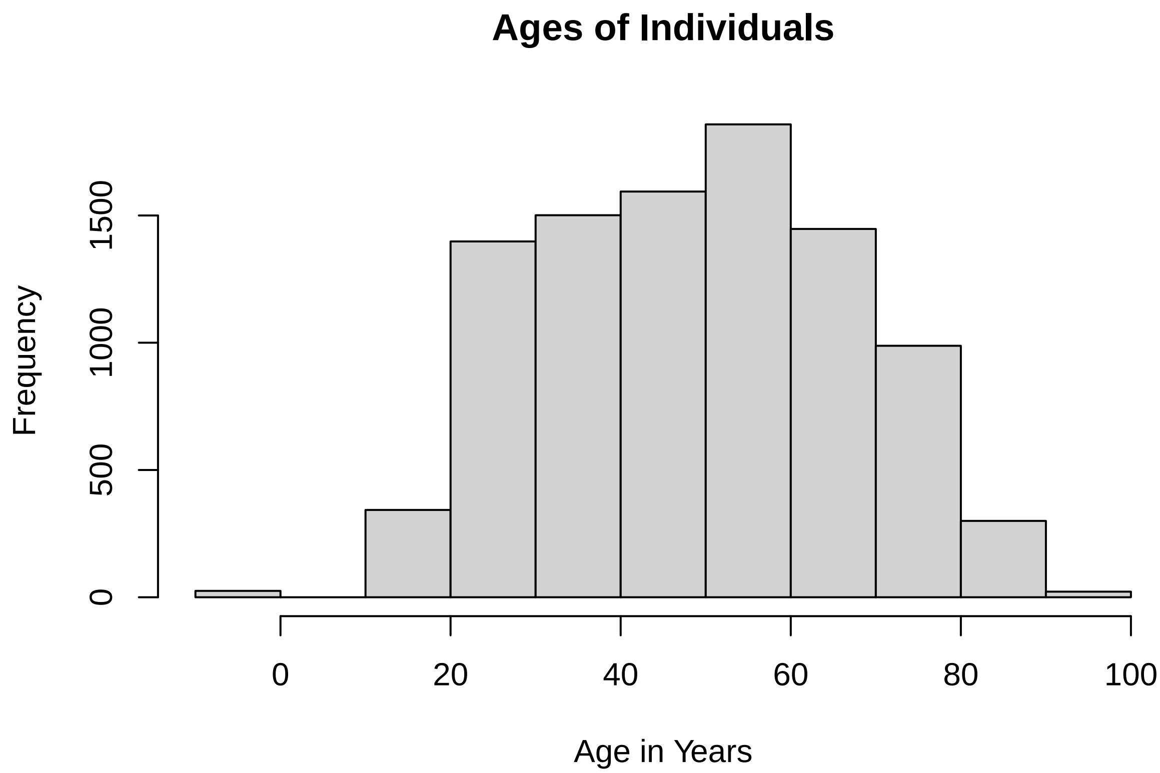
1.4) Create Variables `AgeYears` and `Educ` on `envdat2` from variables `AGE` and `EDUCYRS` which are age in years and years of education, respectively. Get summary information and make a histogram for each variable to see if values seem reasonable.

Modify both variables so that any negative values are changed to `NA`. You'll also discover that some people have more than 30 years of eduction. Assume this is an error and replace these values with NA. Repeat your histograms to make sure your code works.

```
envdat2$AgeYears <- envdat2$AGE
envdat2$Educ <- envdat2$EDUCYRS
summary(envdat2$AgeYears)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -9.00   35.00   50.00   49.35   63.00   97.00
```

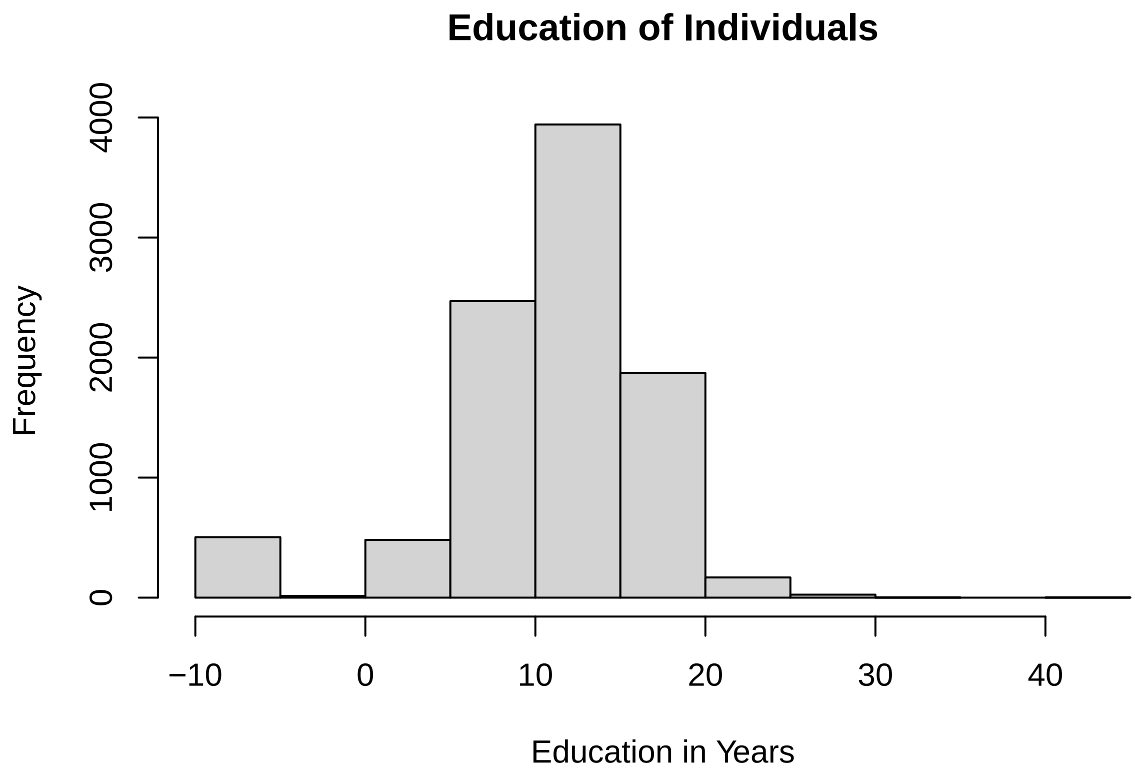
```
hist(envdat2$AgeYears,
     main = "Ages of Individuals",
     xlab = "Age in Years")
```




```
summary(envdat2$Educ)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -9.00   10.00   12.00   11.22   15.00   45.00
```

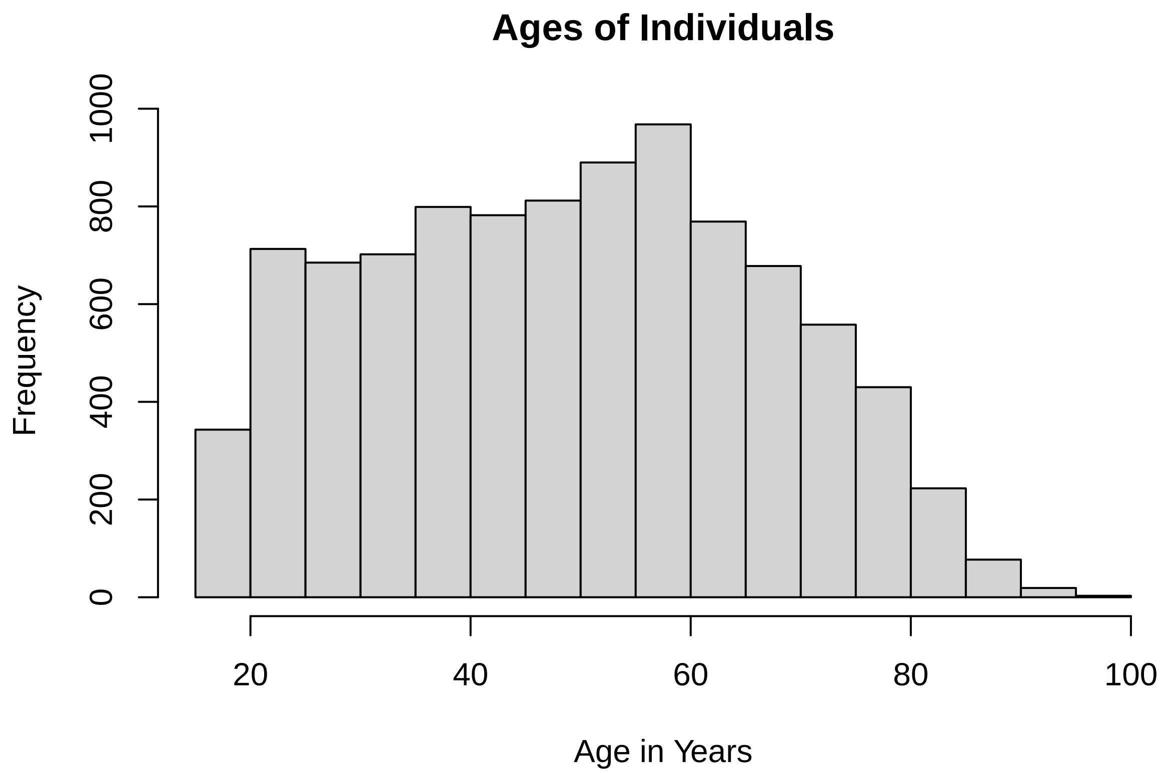
```
hist(envdat2$Educ,
     main = "Education of Individuals",
     xlab = "Education in Years")
```




```
envdat2[envdat2$AgeYears < 0, "AgeYears"] <- NA
envdat2[(envdat2$Educ < 0) | (envdat2$Educ > 30), "Educ"] <- NA
summary(envdat2$AgeYears)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   35.00   50.00   49.51   63.00   97.00      25
```

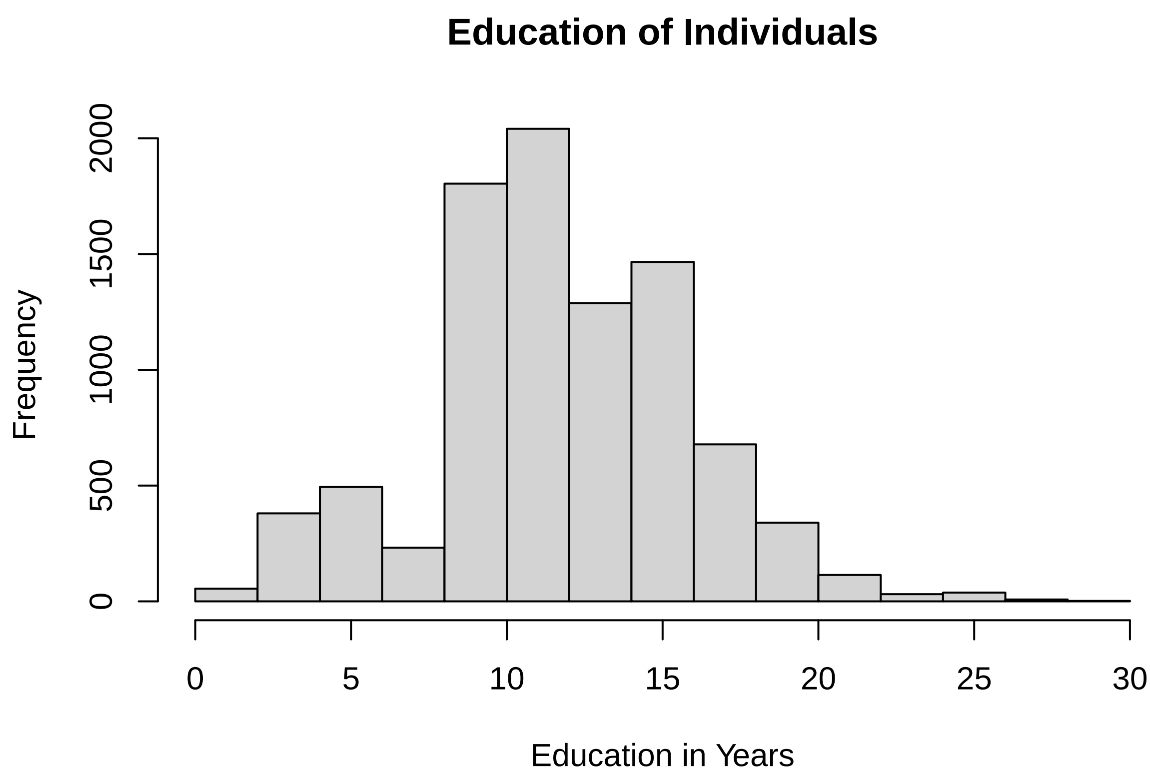
```
hist(envdat2$AgeYears,
     main = "Ages of Individuals",
     xlab = "Age in Years")
```




```r
summary(envdat2$Educ)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   10.00   12.00   12.34   15.00   30.00     505
```

```r
hist(envdat2$Educ,
     main = "Education of Individuals",
     xlab = "Education in Years")
```



1.5) Make a variable `EmpStat` on `envdat2` from variable `MAINSTAT` that contains employment status information. The codebook for this question is listed below:

-9 = No answer -8 = Don't know 1 = In paid work 2 = Unemployed and looking for a job 3 = In education 4 = Apprentice or trainee 5 = Permanently sick or disabled 6 = Retired 7 = Domestic work (meaning working in the home with family) 8 = Incompulsory military service or community service 9 = Other

Code `EmpStat` in the following way : 1, 2, and 8 code as 'Working or Looking', 7 codes as 'At Home', 3 and 4 code as 'Student', 6 codes as 'Retired', -9 and -8 code as NA, and 5, 9 code as 'Other'.

Make a table of your results - look up options for the `table()` function so that it also displays NA values.

```
envdat2$EmpStat <- recode(envdat2$MAINSTAT,
                          "1 = 'Working or Looking';
                          2 = 'Working or Looking';
                          8 = 'Working or Looking';
                          7 = 'At Home';
                          3 = 'Student';
                          4 = 'Student';
                          6 = 'Retired';
                          -9 = NA;
                          -8 = NA;
                          5 = 'Other';
                          9 = 'Other'")
table(envdat2$EmpStat, exclude = NULL)
```

```
##
```

```
##             At Home               Other             Retired             Student
##                 919                 224                1557                 437
## Working or Looking                <NA>
##                5983                 356
```

1.6) If you look in the codebook starting on page 10, you'll notice many questions are on a 5 point scale. We're going to create a composite environmental index score from 12 of these questions. The new variable should be called `EnvAtt` on `envdat2`. It might be tempting to simply add scores together; however, for some questions 'Agree Strongly' (which is coded as 1) would indicate support the environment, while other questions 'Agree Strongly' indicates lack of support for the environment. Furthermore, you want to combined variables so that a higher score indicates more support for the environment.

The following questions should be included : V20, V21, V22, V23; V26, V27, V28, v29; V31, V32, V34, V36. For each quesion, you'll either want to add it directly, OR you'll want to add (6 - question). Using (6 - question) will convert 1's to 5's and 5's to 1's. Remember, for your final composite variable, you want a HIGHER score to indicate GREATER support for the environment.

FIRST - you'll need to replace the values -9 and -8 in each of these questions (which stand for 'No Answer' and 'Can't Choose') with NA. THEN you can proceed with making your composite variable.

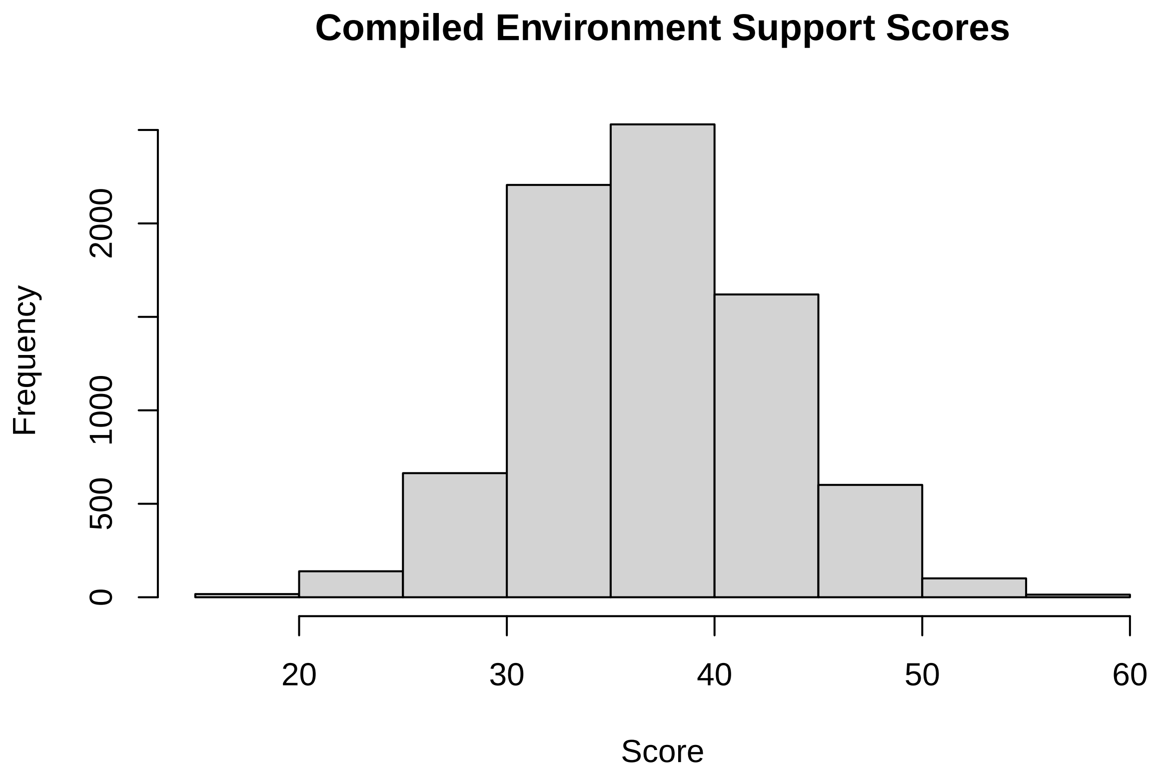
Your new varible should have a minimum of 12 and a maximum of 60. Check this by getting summary information and by making a histogram of `EnvAtt`. Also, make a normal quantile plot of 'Envatt'. You'll also want to determine how many missing values were created in the final index.

The original variables were all on a 5 point scale and were certainly NOT normally distributed. Does your new composite variable seem to be approximately normally distributed?

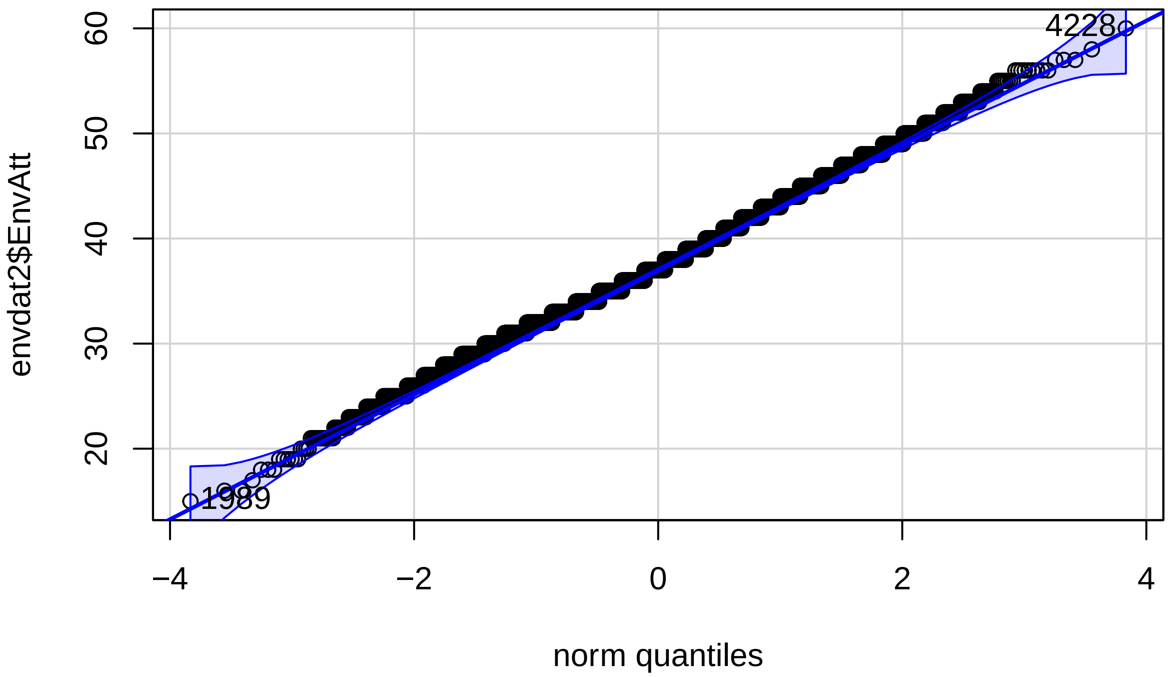
```r
normal_list <- c("v21", "v23", "v29", "v32", "v34")
reverse_list <- c("v20", "v22", "v26", "v27", "v28", "v31", "v36")
envdat2$EnvAtt <- 0
for (i in 1:5) {
  col <- normal_list[i]
  envdat2[(envdat2[col] == -8) | (envdat2[col] == -9),][col] <- NA
  envdat2$EnvAtt <- envdat2$EnvAtt + envdat2[[col]]
}
for (i in 1:7) {
  col <- reverse_list[i]
  envdat2[(envdat2[col] == -8) | (envdat2[col] == -9),][col] <- NA
  envdat2$EnvAtt <- envdat2$EnvAtt + 6 - envdat2[[col]]
}
summary(envdat2$EnvAtt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   15.00   33.00   37.00   37.47   41.00   60.00    1584
```

```r
hist(envdat2$EnvAtt,
     main="Compiled Environment Support Scores",
     xlab="Score")
```




```r
qqPlot(envdat2$EnvAtt)
```




```
## [1] 4228 1989
```

The composite variable does seem normally distributed. The histogram of the composite scores looks normally distributed, and the normal quantile plot shoes that the values are within the confidence interval.

1.7) Finally, create a new dataframe called `envdat3` which contains only the new variables you've created in 1.2 through 1.6. Remove any rows with missing values for any of these variables. Get the dimension of `envdat3`. Lastly, attach `envdat3`.

```
envdat3 <- envdat2[,c("Country", "Gender", "AgeYears", "Educ", "EmpStat", "EnvAtt")]
envdat3 <- na.omit(envdat3)
dim(envdat3)
```

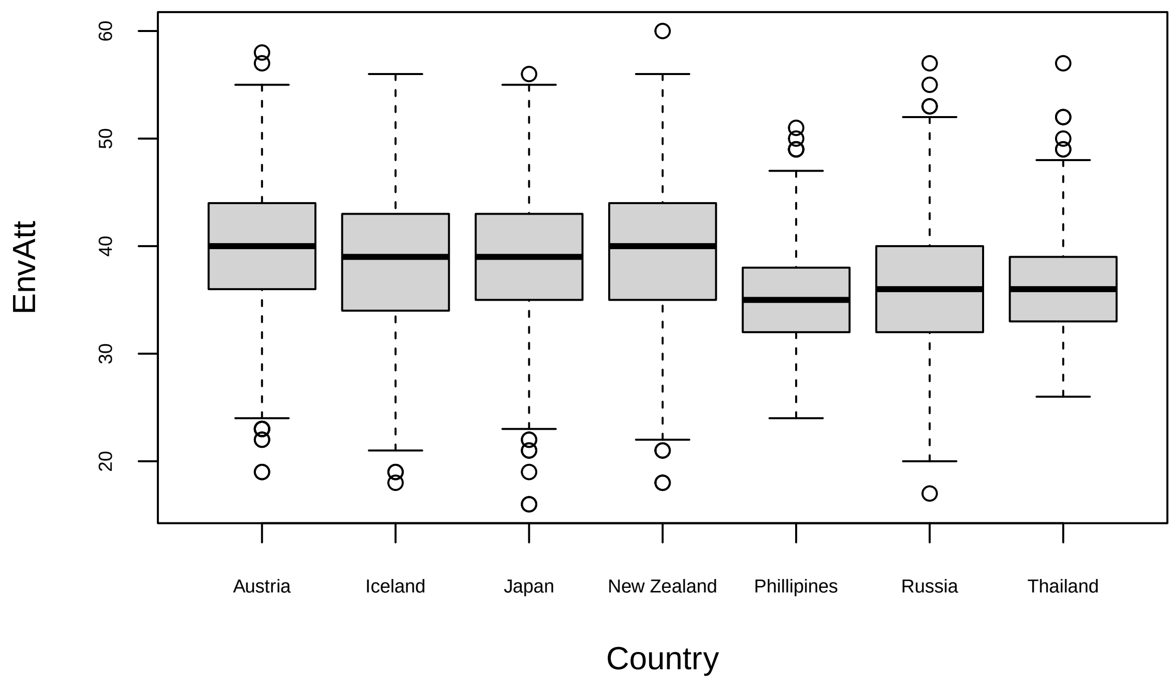
```
## [1] 7394    6
```

```
attach(envdat3)
```

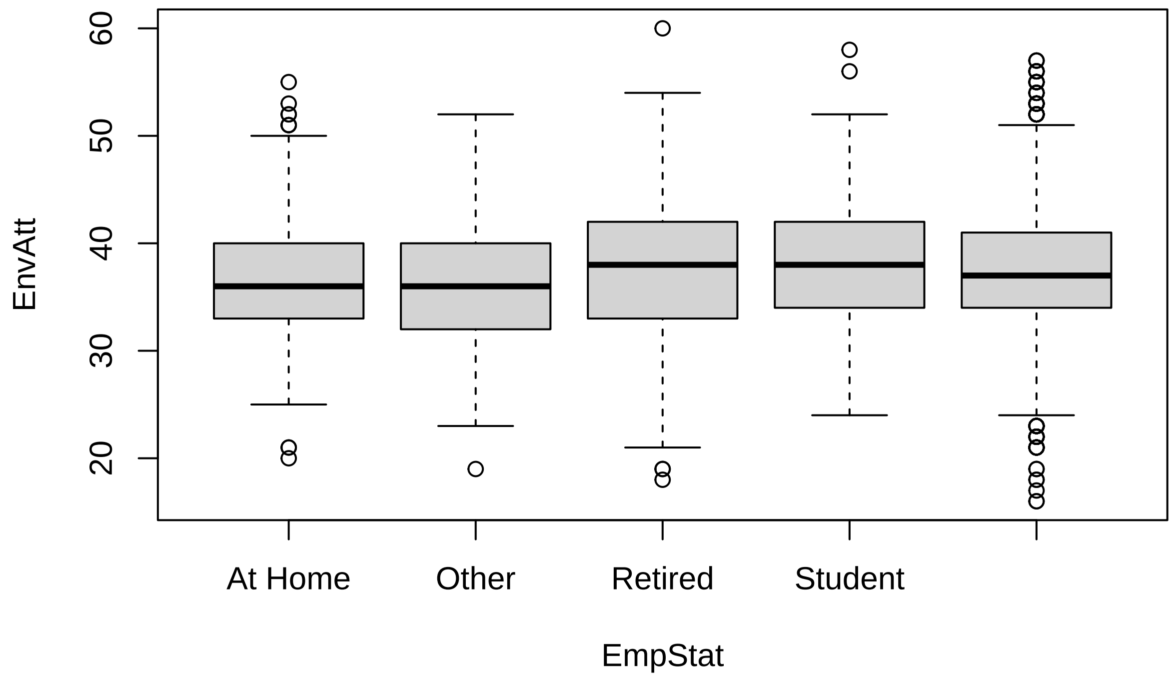
2. Plots and Interactions *(17 pts)*

2.1) Make boxplots of `EnvAtt` by Country, Employment Status, and Gender (that's 3 different boxplots). Additionally, make scatterplots of `EnvAtt` by age and by years of education. Is there evidence visually that our composite environmental index is related to any of these 5 variables (and elaborate)?

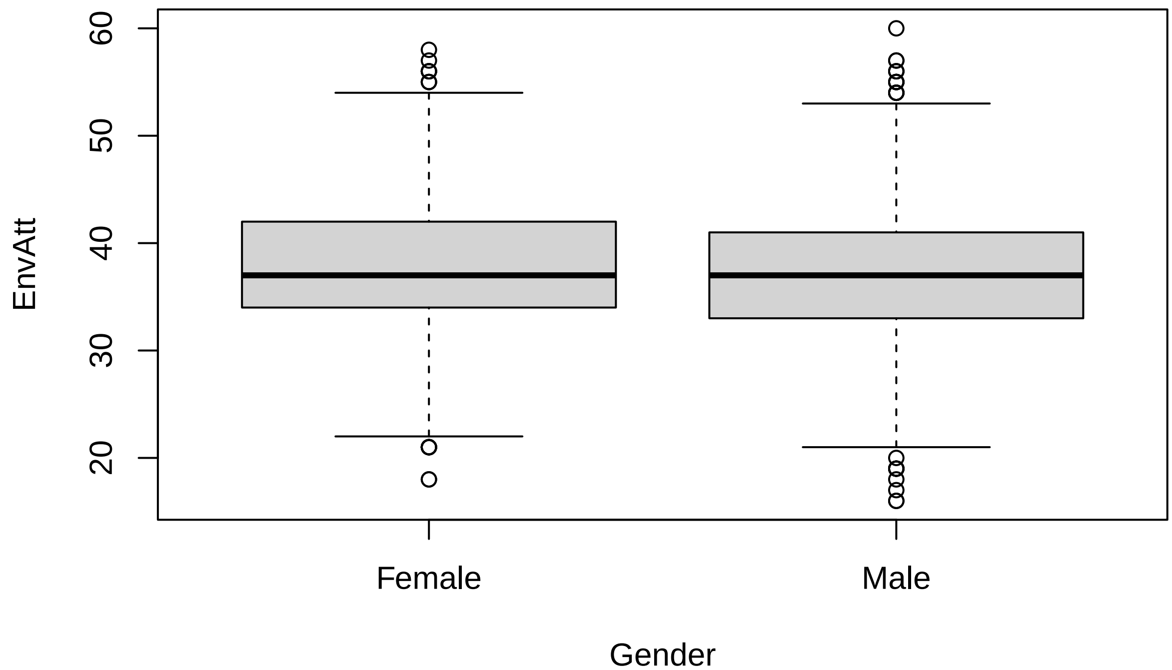
```r
par(cex.axis=0.6)
boxplot(EnvAtt ~ Country)
```




```r
par(cex.axis=1)
boxplot(EnvAtt ~ EmpStat)
```




```r
boxplot(EnvAtt ~ Gender)
```



Visually, there seems to be no significant difference in the composite variable when comparing across these different groups since the boxplots show that the distributions more or less have the same median and variance. There is a perhaps a small difference between the Phillipines, Russia, and Thailand against the other 4 countries; the median for those 3 countries is smaller.

2.2). Examine relationships between the three possible pairs of the variables Gender, Country, EmpStat as they relate to EnvAtt. For each pair of categorical variables, make a table showing counts (i.e. `table(Gender, Country)` for example). This will allow you to see how many individuals exist for each combination of levels of your categorical variables. THEN, make interaction plots for each pair of variables as they relate to `EnvAtt`. Which pairs of categorical variables seem like they might have an interaction as they relate to `EnvAtt`?

```
table(Country, EmpStat)
```

```
##                EmpStat
## Country         At Home Other Retired Student Working or Looking
##   Austria            22    27     405      35                610
##   Iceland             9    45      69      45                587
##   Japan             151    31     116      44                615
##   New Zealand        30     4     199      44                530
##   Phillipines       303    21      81      82                879
##   Russia            113    32     255      52                879
##   Thailand           78     5      65      72                859
```

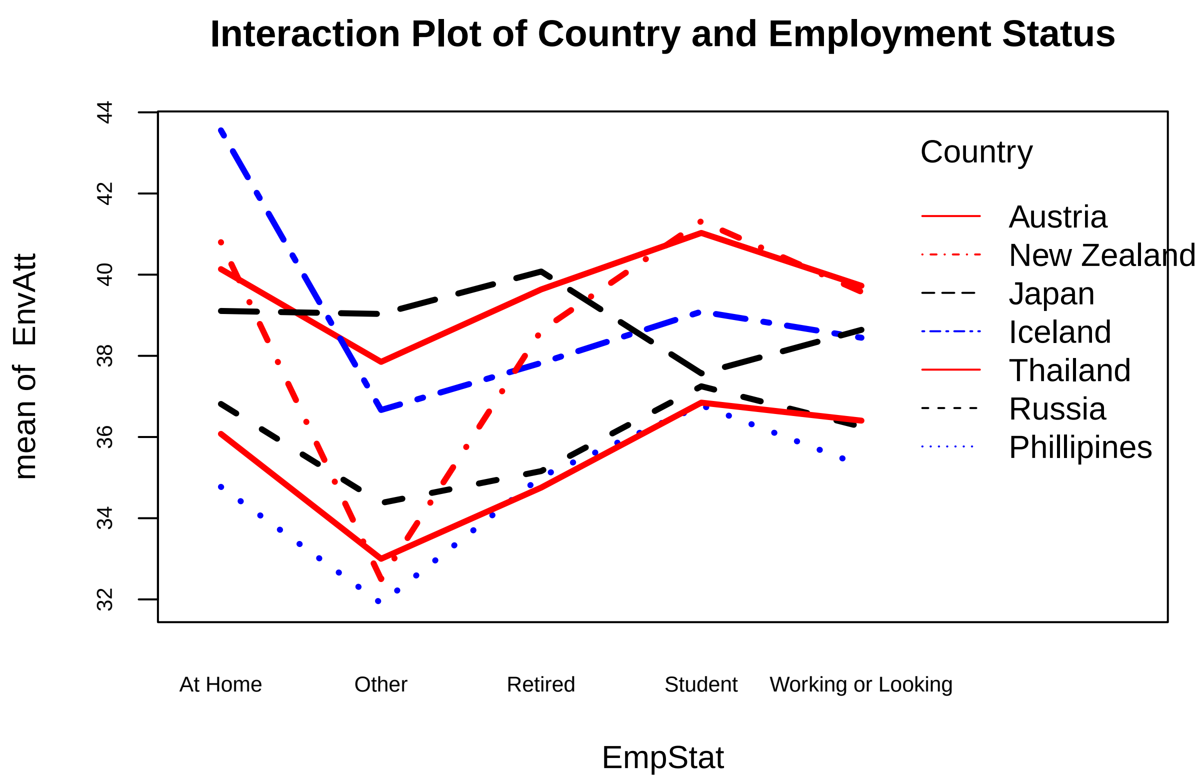
```
table(Country, Gender)
```

```
##                Gender
## Country         Female Male
##    Austria         544  555
##    Iceland         393  362
##    Japan           471  486
##    New Zealand     417  390
##    Phillipines     688  678
##    Russia          725  606
##    Thailand        583  496
```

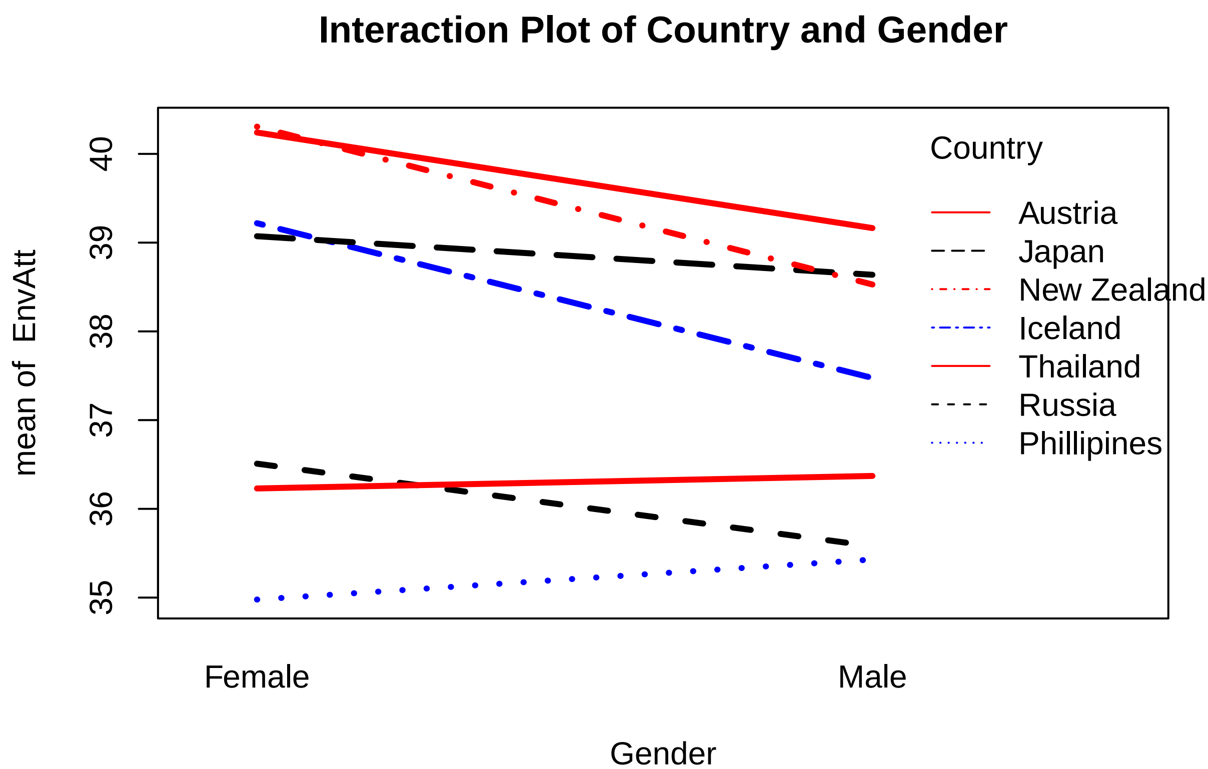
```
table(EmpStat, Gender)
```

```
##                       Gender
## EmpStat                Female Male
##    At Home                615   91
##    Other                   88   77
##    Retired                590  600
##    Student                190  184
##    Working or Looking    2338 2621
```

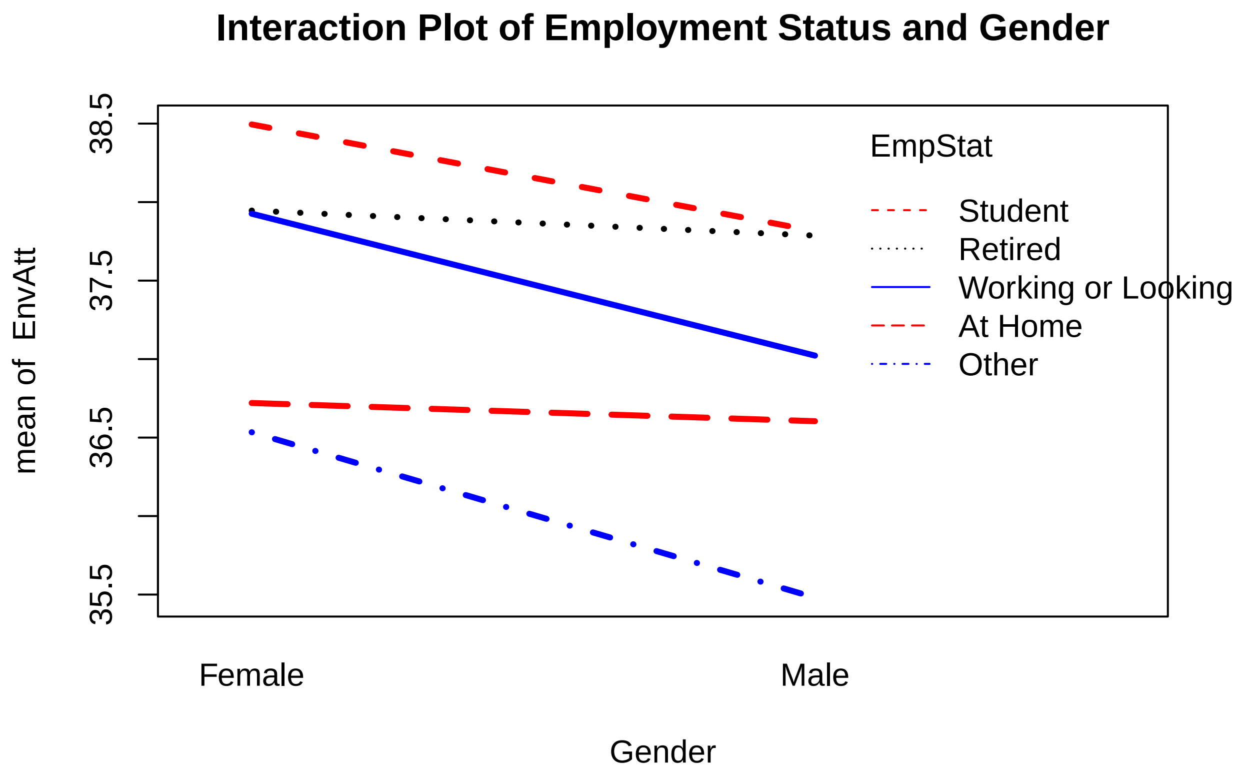
```
par(cex.axis=0.7)
interaction.plot(EmpStat, Country, EnvAtt,
                 lwd = 3,
                 col = c('red','blue','black'),
                 main = "Interaction Plot of Country and Employment Status")
```



```r
par(cex.axis=1)
interaction.plot(Gender, Country, EnvAtt,
                 lwd = 3,
                 col = c('red','blue','black'),
                 main = "Interaction Plot of Country and Gender")
```




```r
interaction.plot(Gender, EmpStat, EnvAtt,
                 lwd = 3,
                 col = c('red','blue','black'),
                 main = "Interaction Plot of Employment Status and Gender")
```



All pairs of categorical variables seem to have interactions with each other. In all three interaction plots the lines are not parallel. However, in the Country vs Employment Status interaction plot, Japan has the most notable interaction by far. The other countries maintain the same sign of slope between the different employment statuses, but Japan's slope sometimes deviates from the others. Also, in the Emplyment Status vs Gender interaction plot, the slope is either flat or negative, which suggests that the interaction is not very strong.

3. Fitting Two-Way ANOVA models *(20 pts, 10 pts each part)*

3.1) Fit three different two-way ANOVA models for `EnvAtt`, one for each pair of your three categorical variables (i.e. Country and EmpStat, Country and Gender, Gender and EmpStat). Include an interaction term in each model. Get summary information for each model using the `Anova()` function with option `type = 'III'`. Which models appear to have significant interaction terms? Does this seem to be consistent with what you observed in the interaction plots above?

```
Anova(aov(EnvAtt ~ Country + EmpStat + Country*EmpStat), type = 'III')
```

```
## Anova Table (Type III tests)
##
## Response: EnvAtt
##                 Sum Sq   Df   F value    Pr(>F)
## (Intercept)      35440    1 1149.7454 < 2.2e-16 ***
## Country           3219    6   17.4043 < 2.2e-16 ***
## EmpStat            160    4    1.2993 0.2677515
## Country:EmpStat   1595   24    2.1562 0.0008702 ***
## Residuals       226838 7359
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(aov(EnvAtt ~ Country + Gender + Country*Gender), type = 'III')
```

```
## Anova Table (Type III tests)
##
## Response: EnvAtt
##                 Sum Sq   Df    F value    Pr(>F)
## (Intercept)    880912    1 28573.2050 < 2.2e-16 ***
## Country         15521    6    83.9065 < 2.2e-16 ***
## Gender            319    1    10.3330  0.001312 **
## Country:Gender   1151    6     6.2214 1.578e-06 ***
## Residuals      227525 7380
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(aov(EnvAtt ~ Gender + EmpStat + Gender*EmpStat), type = 'III')
```

```
## Anova Table (Type III tests)
##
## Response: EnvAtt
##                 Sum Sq   Df    F value    Pr(>F)
## (Intercept)    829255    1 24555.1646 < 2.2e-16 ***
## Gender              1    1     0.0315    0.8590
## EmpStat           981    4     7.2637 7.811e-06 ***
## Gender:EmpStat    169    4     1.2491    0.2878
## Residuals      249366 7384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Country vs EmpStat and Country vs Gender models seem to have significant interaction terms, with p values less than 0.001. The EmpStat vs Gender model does not, with a p value of 0288. This is consistent with what I observed; while all three interaction plots showed different slopes among the categorical variables, the slopes in the EmpStat vs Gender plot were still moved in the same general direction.

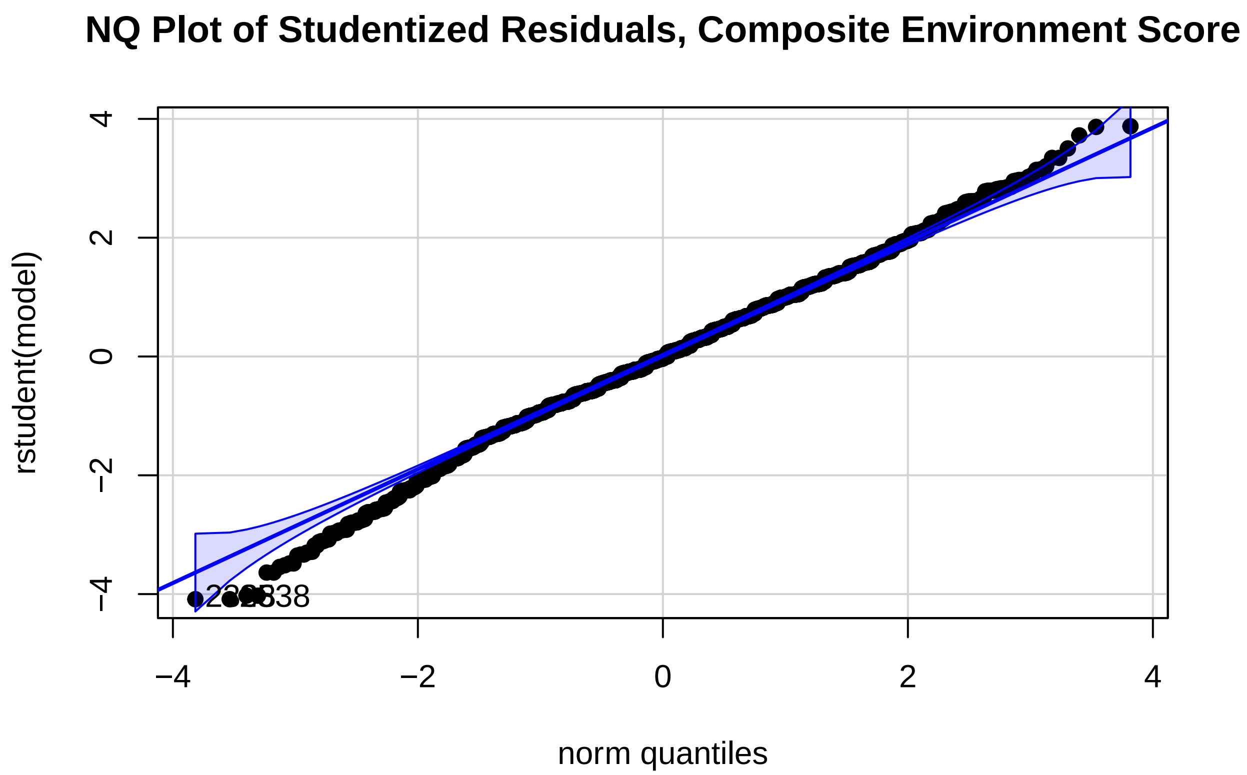
3.2) For the model with Country and Gender you fit in part (3.1), make residual plots. Are the model assumptions reasonably met (write a sentence supporting your answer)?

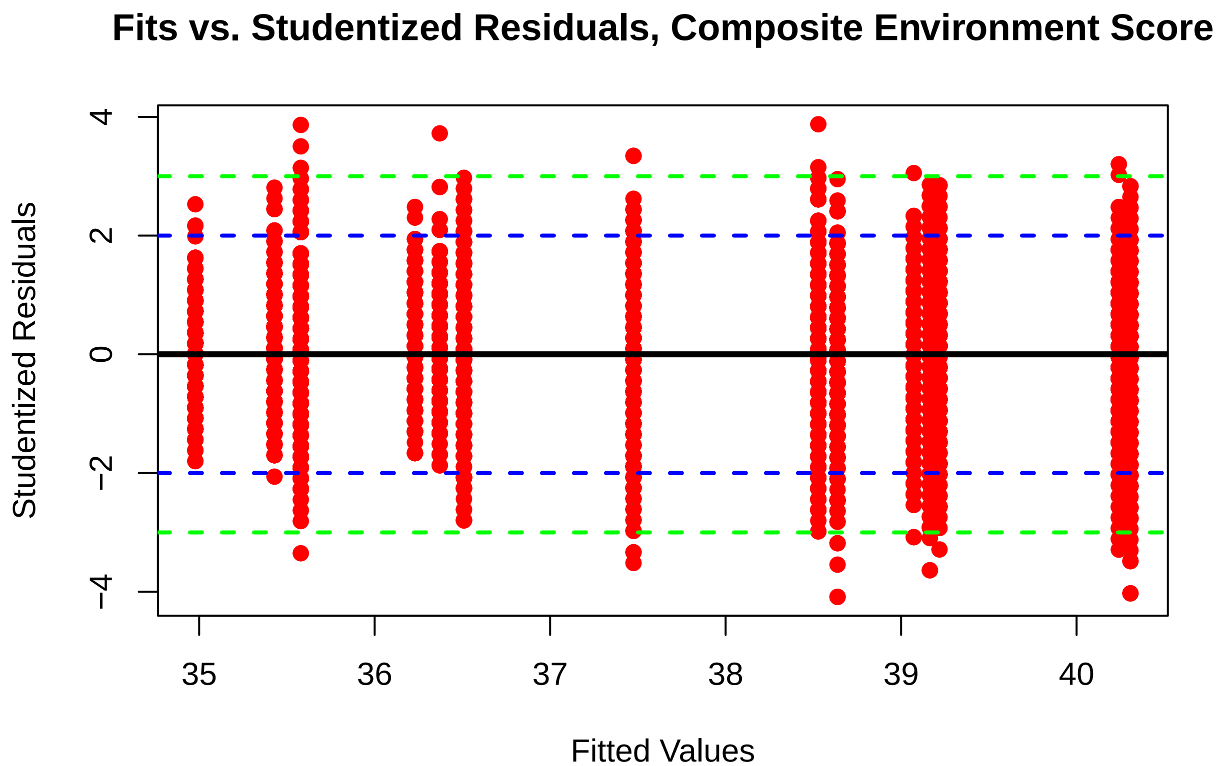
```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.4.3
```

```
myResPlots <- function(model, label){

  #Normal quantile plot of studentized residuals
  qqPlot(rstudent(model), pch = 19,
         main = paste("NQ Plot of Studentized Residuals,", label))

  #plot of fitted vs. studentized residuals
  plot(rstudent(model) ~ model$fitted.values, pch = 19, col = 'red',
       xlab = "Fitted Values",
       ylab = "Studentized Residuals",
```

```
    main = paste("Fits vs. Studentized Residuals,", label))
  abline(h = 0, lwd = 3)
  abline(h = c(2,-2), lty = 2, lwd = 2, col="blue")
  abline(h = c(3,-3), lty = 2, lwd = 2, col="green")

}
myResPlots(aov(EnvAtt ~ Country + Gender + Country*Gender),
           label = "Composite Environment Score")
```





The model assumptions are reasonably met since the composite scores are approximately normally distributed and the residuals follow a uniform distribution.

4. ANCOVA *(20 pts, 10 pts each part)*

4.1) Fit an ANCOVA model predicting `EnvAtt` based on Gender, Education, and the interaction of Gender and Education. Get ANOVA summary information for this model (again, use the `Anova()` function with option `type = 'III'`). Is there a significant interaction between Gender and Education? Also get linear model summary information for this model. For which gender is there a greater increase in Environmental Index Score as Education increases?

```
m1 <- lm(EnvAtt ~ Gender*Educ)
Anova(m1, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: EnvAtt
##               Sum Sq   Df    F value    Pr(>F)
## (Intercept) 380884    1 11732.5017 < 2.2e-16 ***
## Gender         322    1     9.9104   0.00165 **
## Educ          8282    1   255.1285 < 2.2e-16 ***
## Gender:Educ    669    1    20.6000 5.749e-06 ***
## Residuals   239909 7390
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(m1)
```

```
##
## Call:
## lm(formula = EnvAtt ~ Gender * Educ)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.6926  -3.7524  -0.1212   3.7768  22.9409
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     33.08054    0.30541 108.317  < 2e-16 ***
## GenderMale       1.36520    0.43366   3.148  0.00165 **
## Educ             0.36734    0.02300  15.973  < 2e-16 ***
## GenderMale:Educ -0.14956    0.03295  -4.539 5.75e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.698 on 7390 degrees of freedom
## Multiple R-squared:  0.04641,    Adjusted R-squared:  0.04603
## F-statistic: 119.9 on 3 and 7390 DF,  p-value: < 2.2e-16
```

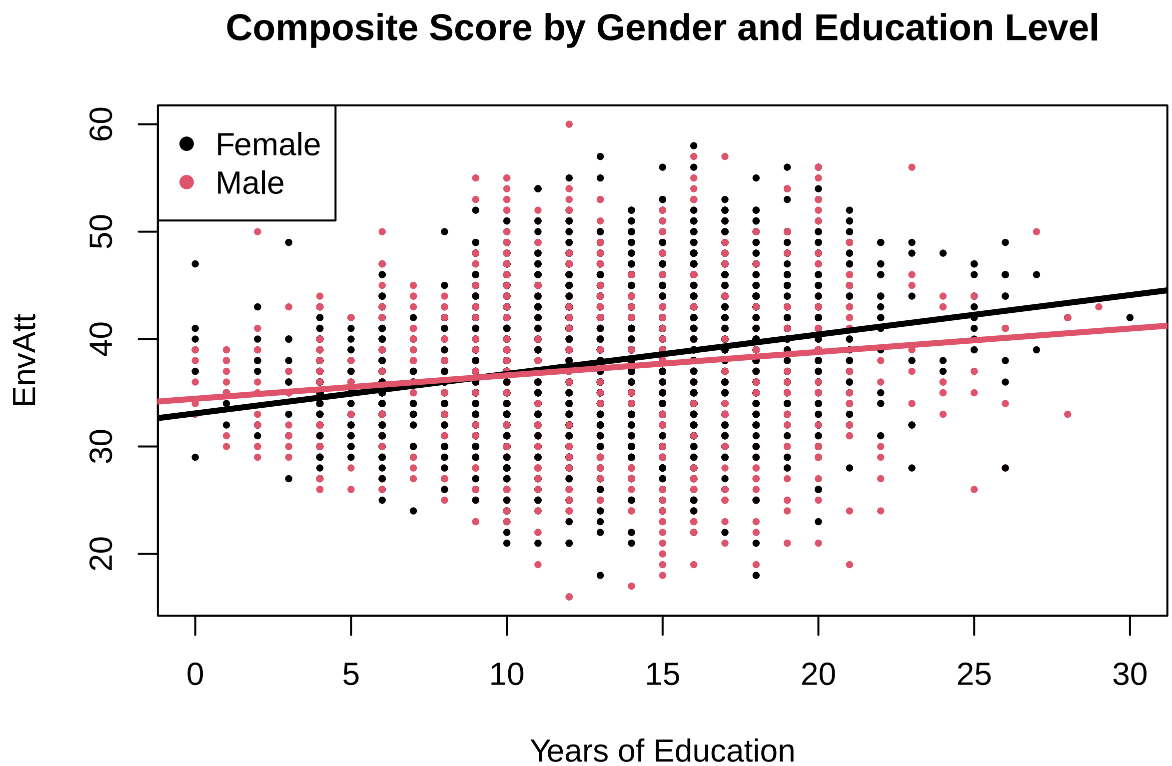
There does seem to be a significant interaction between Gender and Education, with a p value of less than 0.001. Since the slope of GenderMale:Educ is negative, this means that for Female there is a greater increase in Environmental Index Score as Eduation increases compared to Male.

4.2) Make a plot that shows `EnvAtt` as predicted by years of education with separate colors for each gender. Then, superimpose the two predicted regression lines (one for each gender). The plot should be consistent with your results from part a). You'll want to look carefully at similar code in Class 19.

```r
plot(EnvAtt ~ Educ,
     col = factor(Gender),
     pch = 16,
     cex = .5,
     main="Composite Score by Gender and Education Level",
     xlab="Years of Education")
legend("topleft", col = 1:2, legend = levels(factor(Gender)), pch = 16)
coefs <- coef(m1)
coefs
```

```
##     (Intercept)      GenderMale            Educ GenderMale:Educ
##      33.0805350       1.3652012       0.3673366      -0.1495551
```

```r
abline(a = coefs[1], b = coefs[3], col = "black", lwd = 3)
abline(a = coefs[1] + coefs[2], b = coefs[3] + coefs[4], col = 2, lwd = 3)
```



5. GLM *(20 pts, 10 pts each part)*

5.1). Fit a model for `EnvAtt` that includes ALL of the five possible continous and categorical predictors. Also include two-way interactions between Gender and Education, Employment Status and Country, Gender and Country, Employment Status and Gender, Age and Gender. Save to an object called `m1`.

Get ANOVA summary information for this model (again, use `Anova()` with option `type = 'III'`).

THEN, perform manual backwards stepwise regression, removing non-significant terms until all terms have p-values less than 0.05. REMEMBER, you want to remove interactions BEFORE you remove any main effects. You don't need to show every step - just put your final model into an object called `m2`. Get linear model summary information for this model. Finally, check residuals for your final model.

```
m1 <- lm(EnvAtt ~ Gender*Educ +
           EmpStat*Country +
           Gender*Country +
           EmpStat*Gender +
           AgeYears*Gender)
Anova(m1, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: EnvAtt
##                 Sum Sq   Df  F value    Pr(>F)
## (Intercept)      23574    1 784.7398 < 2.2e-16 ***
## Gender               2    1   0.0790   0.77861
## Educ              2801    1  93.2391 < 2.2e-16 ***
```

```
## EmpStat            217    4   1.8026   0.12530
## Country           1955    6  10.8456 4.729e-12 ***
## AgeYears             0    1   0.0052   0.94231
## Gender:Educ        174    1   5.7789   0.01624 *
## EmpStat:Country   1370   24   1.9007   0.00502 **
## Gender:Country     576    6   3.1968   0.00390 **
## Gender:EmpStat      25    4   0.2120   0.93192
## Gender:AgeYears      3    1   0.1065   0.74417
## Residuals       220615 7344
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m1 <- lm(EnvAtt ~ Gender*Educ +
           EmpStat*Country +
           Gender*Country +
           AgeYears)
Anova(m1, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: EnvAtt
##                 Sum Sq   Df  F value    Pr(>F)
## (Intercept)      25626    1 853.4336 < 2.2e-16 ***
## Gender               0    1   0.0078  0.929811
## Educ              3079    1 102.5337 < 2.2e-16 ***
## EmpStat            195    4   1.6199  0.166212
## Country           1971    6  10.9381 3.651e-12 ***
## AgeYears             6    1   0.1844  0.667648
## Gender:Educ        238    1   7.9273  0.004882 **
## EmpStat:Country   1434   24   1.9899  0.002769 **
## Gender:Country     543    6   3.0137  0.006072 **
## Residuals       220670 7349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m2 <- lm(EnvAtt ~ Gender*Educ +
           EmpStat*Country +
           Gender*Country)
Anova(m2, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: EnvAtt
##                 Sum Sq   Df  F value    Pr(>F)
## (Intercept)      28696    1 955.7664 < 2.2e-16 ***
## Gender               0    1   0.0059  0.938581
## Educ              3124    1 104.0433 < 2.2e-16 ***
## EmpStat            191    4   1.5882  0.174440
## Country           2077    6  11.5294 6.965e-13 ***
## Gender:Educ        236    1   7.8595  0.005069 **
## EmpStat:Country   1435   24   1.9920  0.002730 **
## Gender:Country     544    6   3.0183  0.006005 **
## Residuals       220675 7350
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

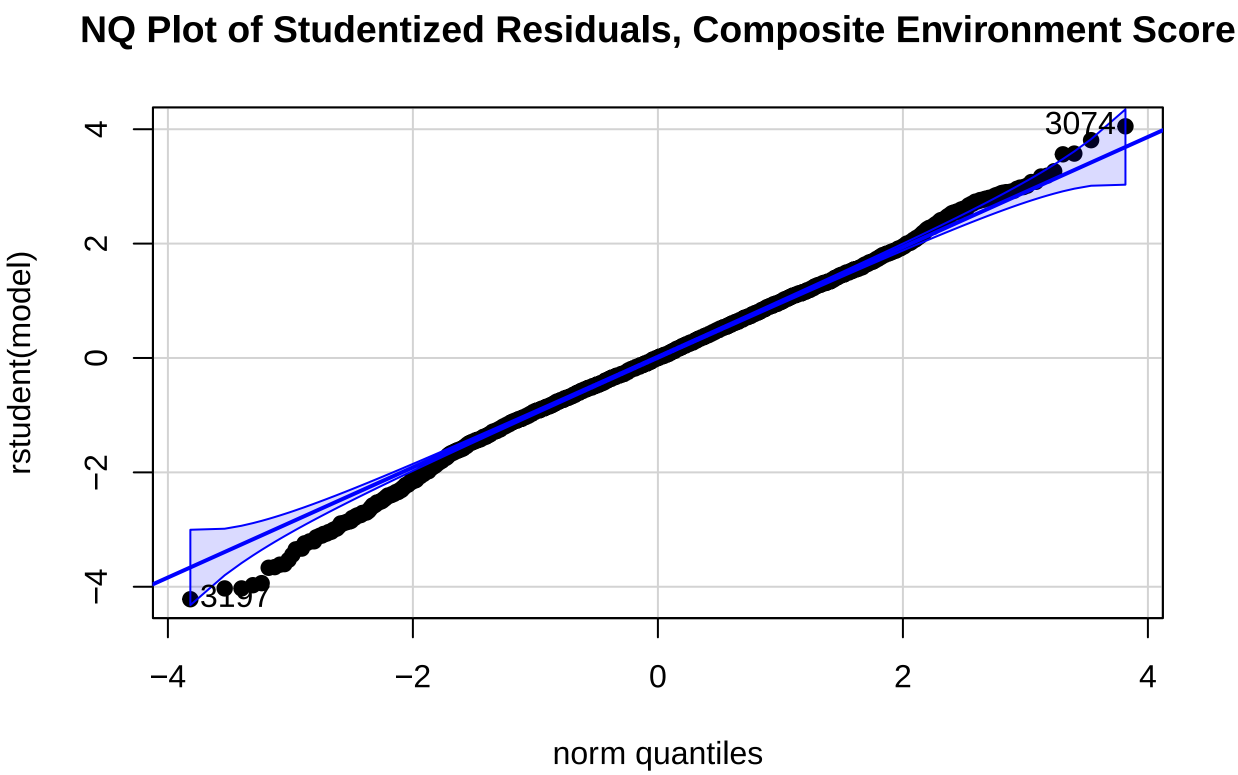
```r
summary(m2)
```

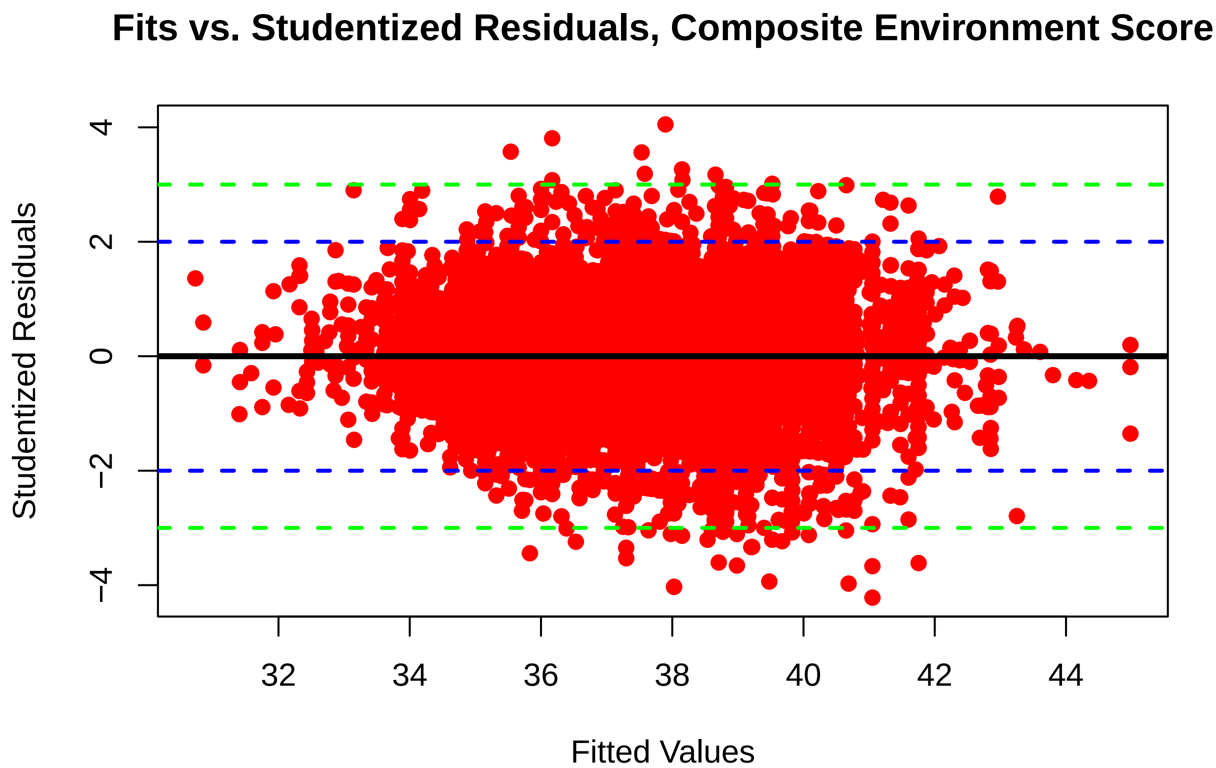
```
##
## Call:
## lm(formula = EnvAtt ~ Gender * Educ + EmpStat * Country + Gender *
##     Country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.0501  -3.4683   0.0107   3.6204  22.1036
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                  37.19781    1.20321  30.915
## GenderMale                                    0.04274    0.55472   0.077
## Educ                                          0.27510    0.02697  10.200
## EmpStatOther                                 -2.19090    1.57564  -1.390
## EmpStatRetired                               -0.12616    1.21021  -0.104
## EmpStatStudent                                1.36325    1.51061   0.902
## EmpStatWorking or Looking                    -0.12089    1.20375  -0.100
## CountryIceland                                2.83208    2.18152   1.298
## CountryJapan                                 -1.51133    1.25188  -1.207
## CountryNew Zealand                           -0.31637    1.54467  -0.205
## CountryPhillipines                           -4.95949    1.21130  -4.094
## CountryRussia                                -3.79782    1.28066  -2.966
## CountryThailand                              -3.58306    1.32418  -2.706
## GenderMale:Educ                              -0.10497    0.03744  -2.803
## EmpStatOther:CountryIceland                  -4.53606    2.54718  -1.781
## EmpStatRetired:CountryIceland                -5.19628    2.28896  -2.270
## EmpStatStudent:CountryIceland                -6.36167    2.50794  -2.537
## EmpStatWorking or Looking:CountryIceland     -5.32941    2.19952  -2.423
## EmpStatOther:CountryJapan                     2.41940    1.92054   1.260
## EmpStatRetired:CountryJapan                   1.36910    1.41210   0.970
## EmpStatStudent:CountryJapan                  -2.78753    1.78851  -1.559
## EmpStatWorking or Looking:CountryJapan       -0.44318    1.31555  -0.337
## EmpStatOther:CountryNew Zealand              -4.52439    3.33116  -1.358
## EmpStatRetired:CountryNew Zealand            -0.85249    1.62590  -0.524
## EmpStatStudent:CountryNew Zealand            -0.38697    1.99708  -0.194
## EmpStatWorking or Looking:CountryNew Zealand -0.66223    1.58758  -0.417
## EmpStatOther:CountryPhillipines              -0.84327    2.01261  -0.419
## EmpStatRetired:CountryPhillipines             0.05759    1.40424   0.041
## EmpStatStudent:CountryPhillipines            -0.29966    1.66221  -0.180
## EmpStatWorking or Looking:CountryPhillipines  0.20209    1.26718   0.159
## EmpStatOther:CountryRussia                   -0.09095    1.92091  -0.047
## EmpStatRetired:CountryRussia                 -1.40857    1.35974  -1.036
## EmpStatStudent:CountryRussia                 -0.61995    1.76899  -0.350
## EmpStatWorking or Looking:CountryRussia      -0.26848    1.32565  -0.203
## EmpStatOther:CountryThailand                 -0.36752    2.98402  -0.123
## EmpStatRetired:CountryThailand               -0.93703    1.53021  -0.612
## EmpStatStudent:CountryThailand               -1.97336    1.76143  -1.120
## EmpStatWorking or Looking:CountryThailand    -0.26761    1.37201  -0.195
```

```
## GenderMale:CountryIceland                      0.12394    0.55187   0.225
## GenderMale:CountryJapan                        0.82099    0.51803   1.585
## GenderMale:CountryNew Zealand                 -0.09063    0.52858  -0.171
## GenderMale:CountryPhillipines                  1.48618    0.47456   3.132
## GenderMale:CountryRussia                       0.39530    0.46230   0.855
## GenderMale:CountryThailand                     1.37238    0.48121   2.852
##                                               Pr(>|t|)
## (Intercept)                                    < 2e-16 ***
## GenderMale                                     0.93858
## Educ                                           < 2e-16 ***
## EmpStatOther                                   0.16442
## EmpStatRetired                                 0.91698
## EmpStatStudent                                 0.36684
## EmpStatWorking or Looking                      0.92001
## CountryIceland                                 0.19426
## CountryJapan                                   0.22738
## CountryNew Zealand                             0.83772
## CountryPhillipines                            4.28e-05 ***
## CountryRussia                                  0.00303 **
## CountryThailand                                0.00683 **
## GenderMale:Educ                                0.00507 **
## EmpStatOther:CountryIceland                    0.07498 .
## EmpStatRetired:CountryIceland                  0.02323 *
## EmpStatStudent:CountryIceland                  0.01121 *
## EmpStatWorking or Looking:CountryIceland       0.01542 *
## EmpStatOther:CountryJapan                      0.20780
## EmpStatRetired:CountryJapan                    0.33230
## EmpStatStudent:CountryJapan                    0.11914
## EmpStatWorking or Looking:CountryJapan         0.73622
## EmpStatOther:CountryNew Zealand                0.17444
## EmpStatRetired:CountryNew Zealand              0.60007
## EmpStatStudent:CountryNew Zealand              0.84636
## EmpStatWorking or Looking:CountryNew Zealand   0.67659
## EmpStatOther:CountryPhillipines                0.67523
## EmpStatRetired:CountryPhillipines              0.96729
## EmpStatStudent:CountryPhillipines              0.85694
## EmpStatWorking or Looking:CountryPhillipines   0.87329
## EmpStatOther:CountryRussia                     0.96224
## EmpStatRetired:CountryRussia                   0.30027
## EmpStatStudent:CountryRussia                   0.72601
## EmpStatWorking or Looking:CountryRussia        0.83951
## EmpStatOther:CountryThailand                   0.90198
## EmpStatRetired:CountryThailand                 0.54033
## EmpStatStudent:CountryThailand                 0.26262
## EmpStatWorking or Looking:CountryThailand      0.84536
## GenderMale:CountryIceland                      0.82231
## GenderMale:CountryJapan                        0.11305
## GenderMale:CountryNew Zealand                  0.86386
## GenderMale:CountryPhillipines                  0.00174 **
## GenderMale:CountryRussia                       0.39253
## GenderMale:CountryThailand                     0.00436 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5.479 on 7350 degrees of freedom
## Multiple R-squared:  0.1229, Adjusted R-squared:  0.1177
## F-statistic: 23.94 on 43 and 7350 DF,  p-value: < 2.2e-16
```

```r
myResPlots(m2, label = "Composite Environment Score")
```





5.2). Write a few sentences describing the overall fit of your final model, the direction of coefficients for each continuous predictor, some discussion of categorical predictors, and the nature of any resulting interactions.

The final model doesn't seem to be that good considering that the adjusted R squared is 0.1177. Years of education is the only continuous predictor; this predictor had a positive slope which suggests that the more educated an individual is, the more likely that they are to support the protection of the environment. The most pro-environment country seems to be Iceland, whereas the least are the Phillipines, Russia, and Thailand. Students and women are more likely to be pro-environment. Only a few of the interactions seem to be statistically significant; those being between employment status and being from Iceland, and being male and being from the Pillipines or Russia.

THE END