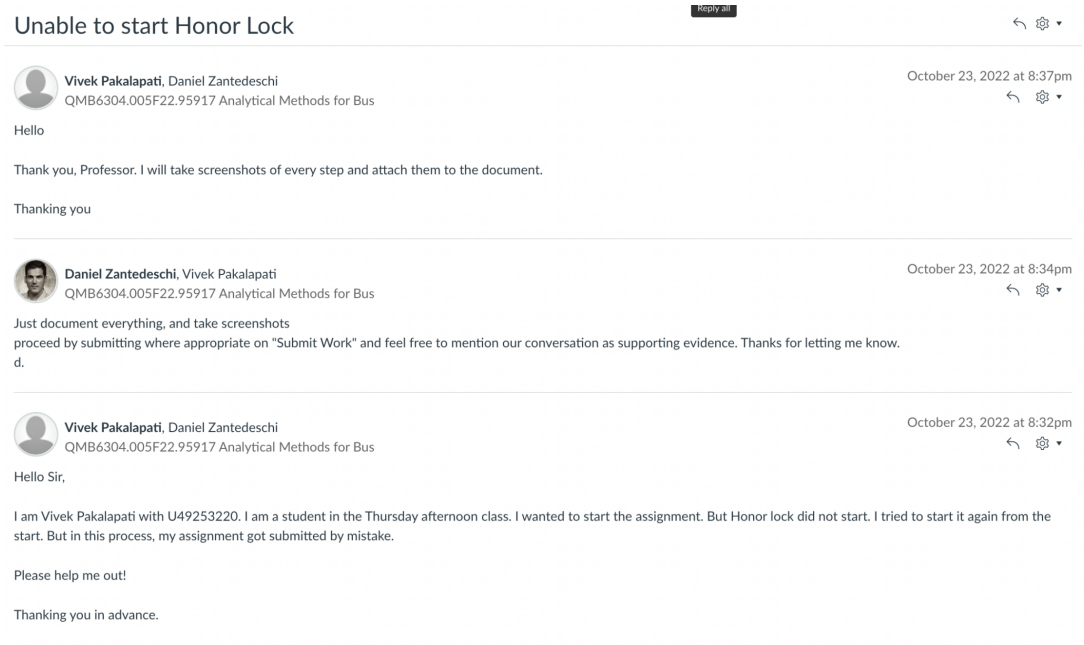
#Vivek Pakalapati
#U49253220

I had an issue with my Honor Lock, So I uploaded the screenshots in every stage in the same document as told by you



Unable to start Honor Lock

Reply all

Vivek Pakalapati, Daniel Zantedeschi
QMB6304.005F22.95917 Analytical Methods for Bus
October 23, 2022 at 8:37pm

Hello

Thank you, Professor. I will take screenshots of every step and attach them to the document.

Thanking you

Daniel Zantedeschi, Vivek Pakalapati
QMB6304.005F22.95917 Analytical Methods for Bus
October 23, 2022 at 8:34pm

Just document everything, and take screenshots
proceed by submitting where appropriate on "Submit Work" and feel free to mention our conversation as supporting evidence. Thanks for letting me know.
d.

Vivek Pakalapati, Daniel Zantedeschi
QMB6304.005F22.95917 Analytical Methods for Bus
October 23, 2022 at 8:32pm

Hello Sir,

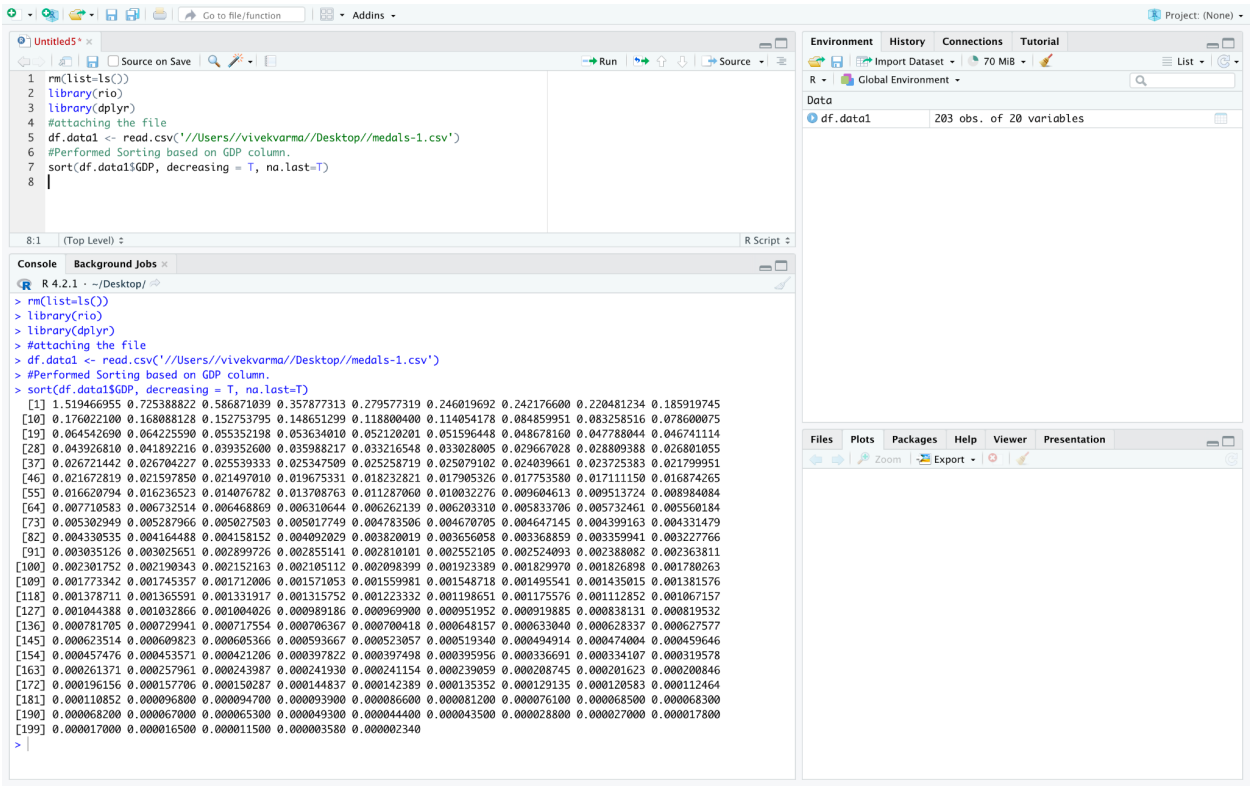
I am Vivek Pakalapati with U49253220. I am a student in the Thursday afternoon class. I wanted to start the assignment. But Honor lock did not start. I tried to start it again from the start. But in this process, my assignment got submitted by mistake.

Please help me out!

Thanking you in advance.

Code:

```
rm(list=ls())
library(rio)
library(dplyr)
#attaching the file
df.data1 <- read.csv('//Users//vivekvarma//Desktop//medals-1.csv')
#Performed Sorting based on GDP column.
sort(df.data1$GDP, decreasing = T, na.last=T)
```



1.) Create a dummy variable for G20 by checking whether the nation is at the top20 for total GDP in 2012.

```
#Created G20 variable based on GDP column.
df.data1$G20 <- ifelse(df.data1$GDP == head(df.data1$GDP,n=20),1,0)
#inspectign the GoldMedals column:
attach(df.data1)
summary(GoldMedals)
```

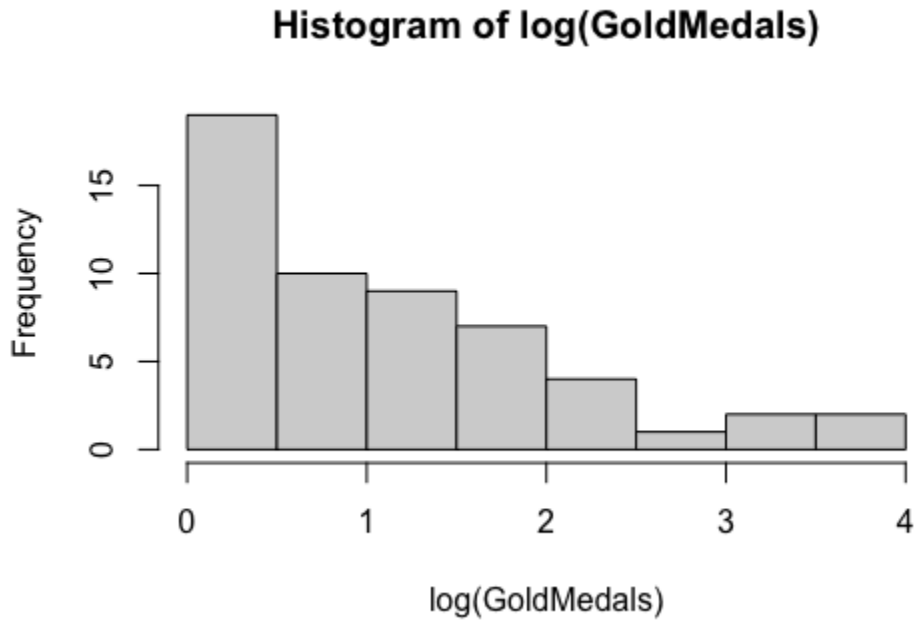
```
> summary(GoldMedals)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   0.000   0.000   1.488   1.000  46.000
```

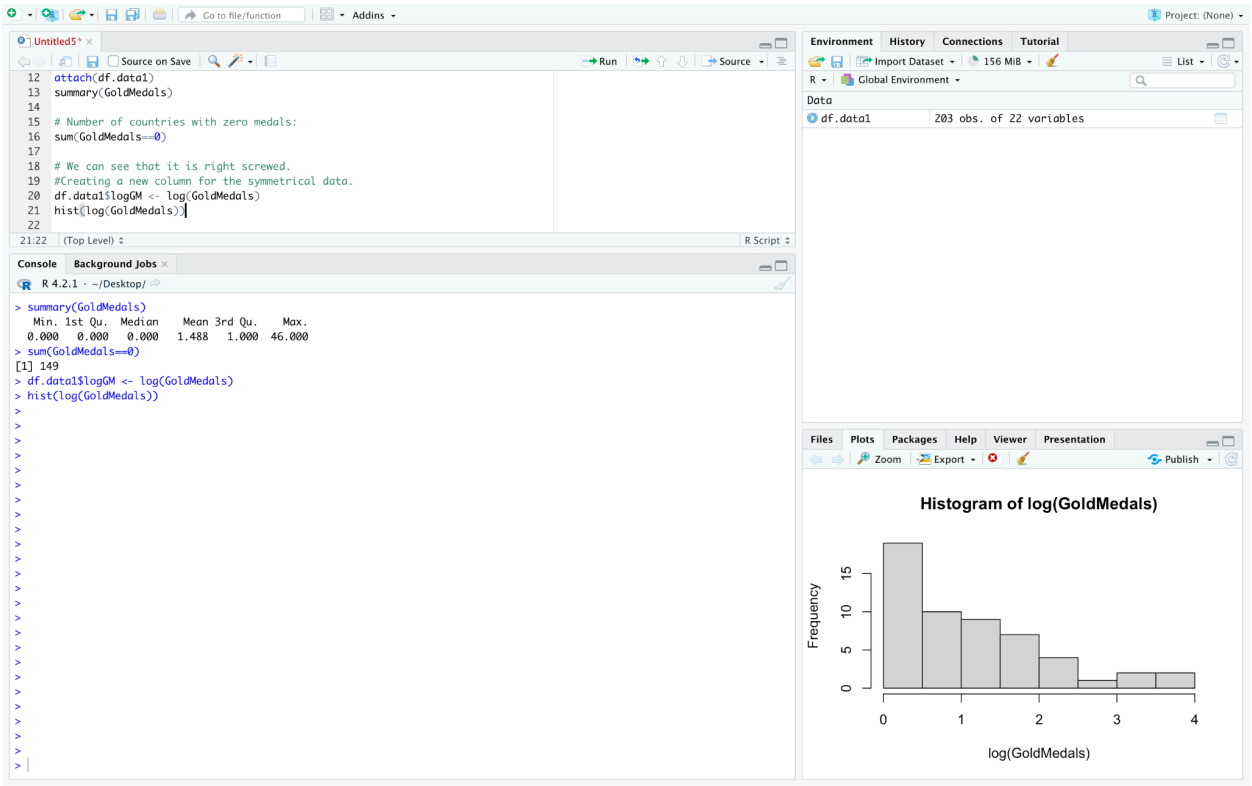
```
# Number of countries with zero medals:
sum(GoldMedals==0)
```

```
> sum(GoldMedals==0)
[1] 149
```

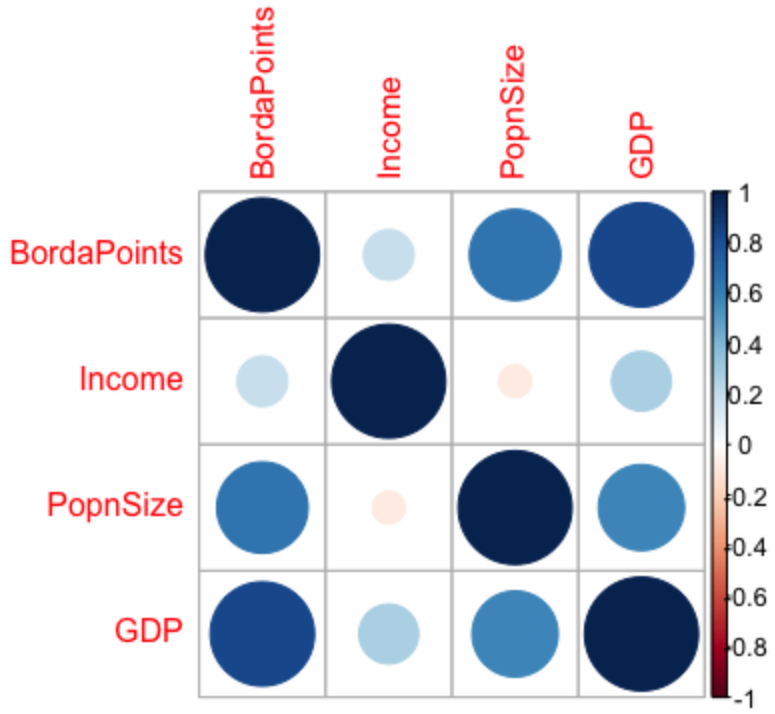
```
# So with the summary we can see that there are 149 countries with zero gold medals,
# Checking if the Goldmedals column is symmetrical or not by generating the Histogram.
hist(GoldMedals)
```

We can see that it is right screwed.
#Creating a new column for the symmetrical data.
df.data1$logGM <- log(GoldMedals)
hist(log(GoldMedals))

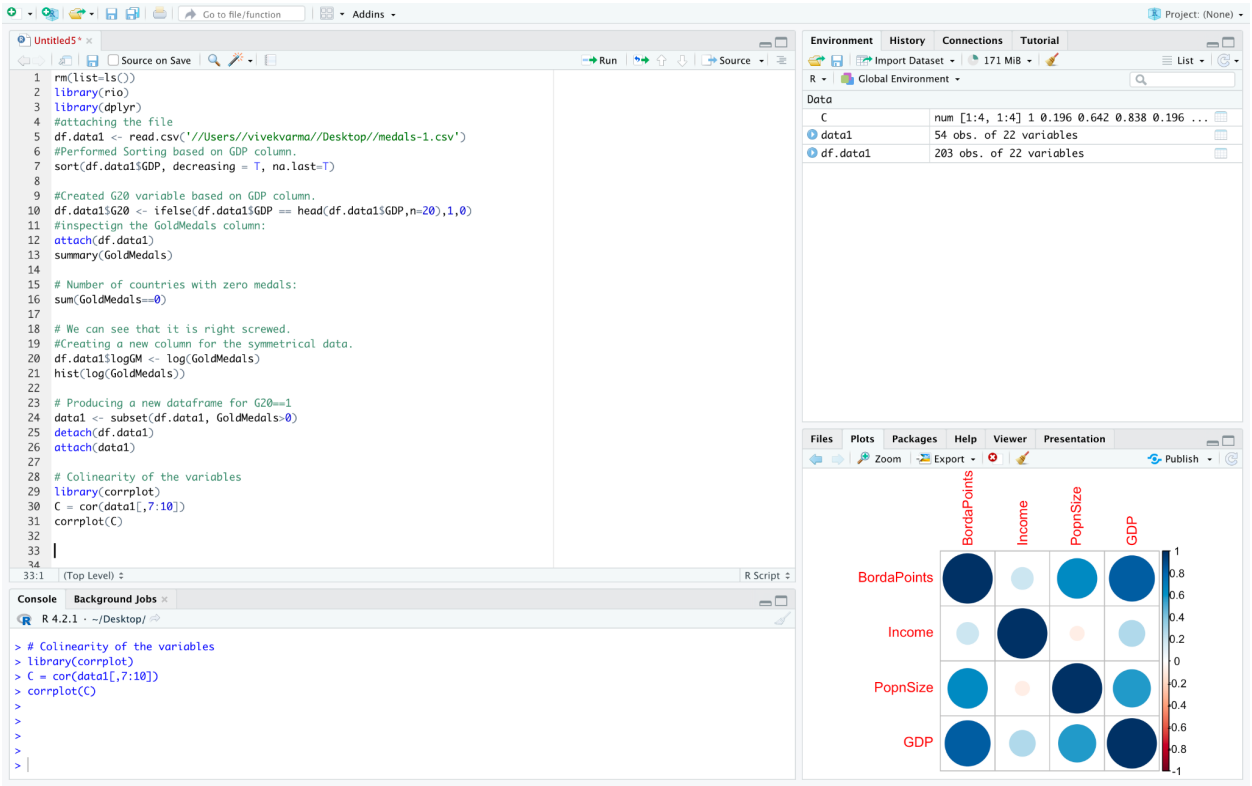





```
# Producing a new dataframe for G20==1
data1 <- subset(df.data1, GoldMedals>0)
detach(df.data1)
attach(data1)
# Colinearity of the variables
library(corrplot)
C = cor(data1[,7:10])
corrplot(C)
```



We can see that the independent variables have some degree of dependency on other independent variables.



#1.2) Investigate models to estimate the number of gold medals won, including the dummy variables you just created. Be careful about how you treat the Nations with no gold medal won. Feel free to use transformations when appropriate. Interpret the results carefully as if you were explaining them to a friend who is not in the MS BAIS program.

```
#ITo check the number of gold medals won
m1 <- lm(logGM ~ G20,data=data1)
summary(m1)
m2 <- lm(logGM ~ GDP+Income+PopnSize,data = data1)
summary(m2)
m3 <- lm(logGM ~ Income,data = data1)
summary(m3)
m4 <- lm(logGM ~ PopnSize,data = data1)
summary(m4)
m5 <- lm(logGM ~ GDP,data=data1)
summary(m5)
m6 <- lm(logGM ~ Income+PopnSize,data = data1)
summary(m6)
m7 <- lm(logGM ~ GDP+PopnSize,data = data1)
summary(m7)
m8 <- lm(logGM ~ GDP+Income,data=data1)
summary(m8)
library(stargazer)
```

```
Call:
lm(formula = logGM ~ G20, data = data1)

Residuals:
   Min     1Q  Median     3Q    Max
-2.0969 -0.4231 -0.1510  0.2700  1.7318

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.4231     0.1169   3.621 0.000665 ***
G20           1.6737     0.1920   8.717 9.54e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6814 on 52 degrees of freedom
Multiple R-squared:  0.5937,            Adjusted R-squared:  0.5859
F-statistic: 75.99 on 1 and 52 DF,  p-value: 9.536e-12

> m2 <- lm(logGM ~ GDP+Income+PopnSize,data = data1)
> summary(m2)

Call:
lm(formula = logGM ~ GDP + Income + PopnSize, data = data1)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.2572 -0.6981 -0.1368  0.7026  1.9760

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.65372    0.16699   3.915 0.000274 ***
GDP          2.18915    0.62574   3.498 0.000993 ***
Income       0.03559    0.05555   0.641 0.524655
PopnSize     1.10794    0.78679   1.408 0.165265
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8299 on 50 degrees of freedom
Multiple R-squared:  0.4205,             Adjusted R-squared:  0.3857
F-statistic: 12.09 on 3 and 50 DF,  p-value: 4.546e-06

> m3 <- lm(logGM ~ Income,data = data1)
> summary(m3)

Call:
lm(formula = logGM ~ Income, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3668 -0.8973 -0.2046  0.5926  2.7486

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.83883    0.20108   4.172 0.000115 ***
Income       0.09270    0.06434   1.441 0.155630
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.048 on 52 degrees of freedom
Multiple R-squared:  0.03839,           Adjusted R-squared:  0.0199
F-statistic: 2.076 on 1 and 52 DF,  p-value: 0.1556

> m4 <- lm(logGM ~ PopnSize,data = data1)
> summary(m4)

Call:
lm(formula = logGM ~ PopnSize, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1717 -0.8787 -0.1982  0.7270  2.3309

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8682     0.1360   6.386 4.71e-08 ***
PopnSize      2.7020     0.6990   3.866  0.00031 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

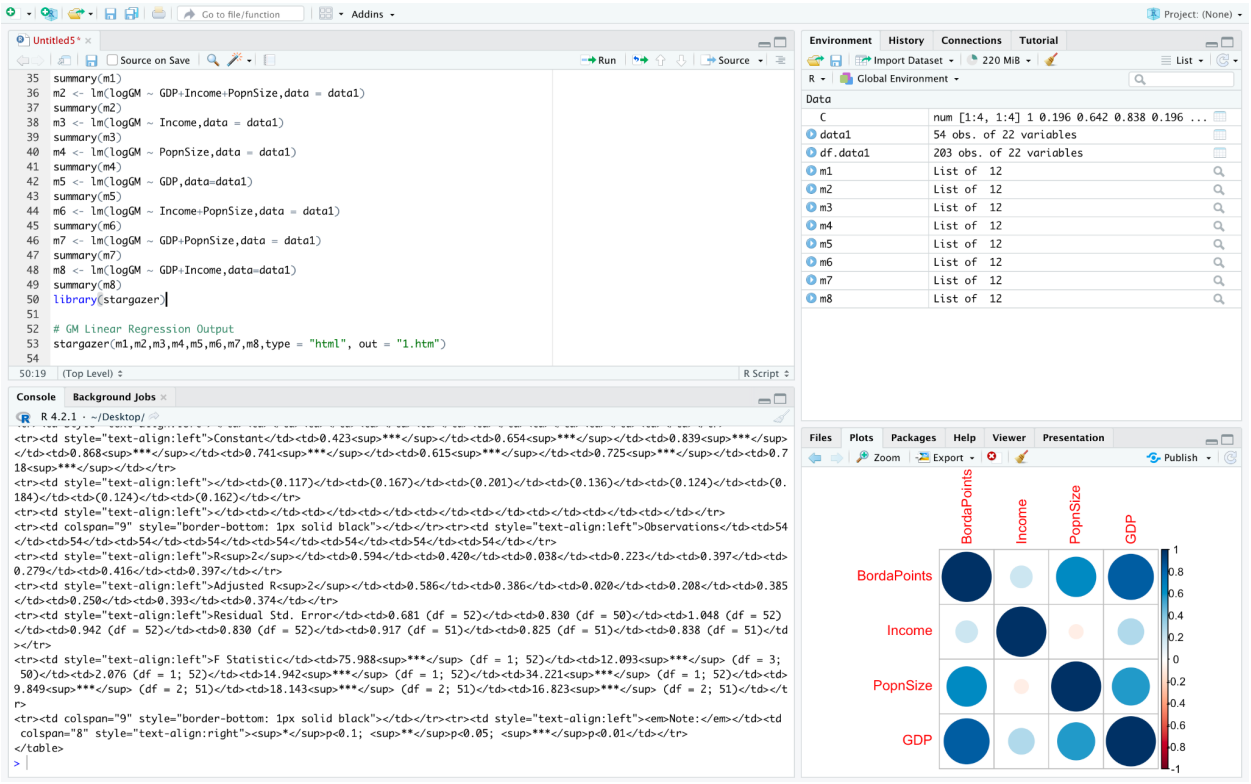
```
Residual standard error: 0.9421 on 52 degrees of freedom
Multiple R-squared:  0.2232,            Adjusted R-squared:  0.2083
F-statistic: 14.94 on 1 and 52 DF,  p-value: 0.0003096

> m5 <- lm(logGM ~ GDP,data=data1)
> summary(m5)

Call:
lm(formula = logGM ~ GDP, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2276 -0.7518 -0.1170  0.6608  1.9567

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.7409     0.1242   5.964 2.19e-07 ***
GDP           2.7653     0.4727   5.850 3.32e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8301 on 52 degrees of freedom
Multiple R-squared:  0.3969,            Adjusted R-squared:  0.3853
F-statistic: 34.22 on 1 and 52 DF,  p-value: 3.32e-07

> m6 <- lm(logGM ~ Income+PopnSize,data = data1)
> summary(m6)

Call:
lm(formula = logGM ~ Income + PopnSize, data = data1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.27777 -0.77455 -0.07552  0.65442  2.14272

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.61490    0.18406   3.341 0.001569 **
Income       0.11173    0.05646   1.979 0.053215 .
PopnSize     2.81249    0.68247   4.121 0.000139 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9168 on 51 degrees of freedom
Multiple R-squared:  0.2786,            Adjusted R-squared:  0.2503
F-statistic: 9.849 on 2 and 51 DF,  p-value: 0.0002417

> m7 <- lm(logGM ~ GDP+PopnSize,data = data1)
> summary(m7)

Call:
lm(formula = logGM ~ GDP + PopnSize, data = data1)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.1711 -0.7410 -0.1022  0.6484  2.0148

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.7248     0.1241   5.841 3.64e-07 ***
GDP           2.3462     0.5724   4.099 0.000149 ***
PopnSize      0.9558     0.7458   1.282 0.205751
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8251 on 51 degrees of freedom
Multiple R-squared:  0.4157,             Adjusted R-squared:  0.3928
F-statistic: 18.14 on 2 and 51 DF,  p-value: 1.119e-06

> m8 <- lm(logGM ~ GDP+Income,data=data1)
> summary(m8)

Call:
lm(formula = logGM ~ GDP + Income, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2596 -0.7385 -0.1324  0.6570  1.9406

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.71779    0.16221   4.425 5.09e-05 ***
GDP          2.73482    0.49604   5.513 1.17e-06 ***
Income       0.01198    0.05347   0.224    0.824
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8378 on 51 degrees of freedom
Multiple R-squared:  0.3975,             Adjusted R-squared:  0.3739
F-statistic: 16.82 on 2 and 51 DF,  p-value: 2.449e-06

> library(stargazer)
```

GM Linear Regression Output

```
stargazer(m1,m2,m3,m4,m5,m6,m7,m8,type = "html", out = "1.htm")
```

	Dependent variable:							
	logGM							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G20	1.674***							
	(0.192)							
GDP		2.189***			2.765***		2.346***	2.735***
		(0.626)			(0.473)		(0.572)	(0.496)
Income		0.036	0.093			0.112*		0.012
		(0.056)	(0.064)			(0.056)		(0.053)
PopnSize		1.108		2.702***		2.812***	0.956	
		(0.787)		(0.699)		(0.682)	(0.746)	
Constant	0.423***	0.654***	0.839***	0.868***	0.741***	0.615***	0.725***	0.718***
	(0.117)	(0.167)	(0.201)	(0.136)	(0.124)	(0.184)	(0.124)	(0.162)
Observations	54	54	54	54	54	54	54	54
R^2	0.594	0.420	0.038	0.223	0.397	0.279	0.416	0.397
Adjusted R^2	0.586	0.386	0.020	0.208	0.385	0.250	0.393	0.374
Residual Std. Error	0.681 (df = 52)	0.830 (df = 50)	1.048 (df = 52)	0.942 (df = 52)	0.830 (df = 52)	0.917 (df = 51)	0.825 (df = 51)	0.838 (df = 51)
F Statistic	75.988*** (df = 1; 52)	12.093*** (df = 3; 50)	2.076 (df = 1; 52)	14.942*** (df = 1; 52)	34.221*** (df = 1; 52)	9.849*** (df = 2; 51)	18.143*** (df = 2; 51)	16.823*** (df = 2; 51)
Note:								*p<0.1; **p<0.05; ***p<0.01



#Q1.3) Investigate the interactions between the G20 variable and other significant predictors in part b). How do you interpret your findings?

```
# Linear regression on the dataset:
mod1 <- lm(G20 ~ GDP ,data=data1)
summary(mod1)
mod2 <- lm(G20 ~ PopnSize ,data=data1)
summary(mod2)
```

```
mod3 <- lm(G20 ~ Income ,data=data1)
summary(mod3)
mod4 <- lm(G20 ~ GDP + PopnSize  ,data=data1)
summary(mod4)
mod5 <- lm(G20 ~ GDP + Income ,data=data1)
summary(mod5)
mod6 <- lm(G20 ~ Income + PopnSize ,data=data1)
summary(mod6)
library(stargazer)
```

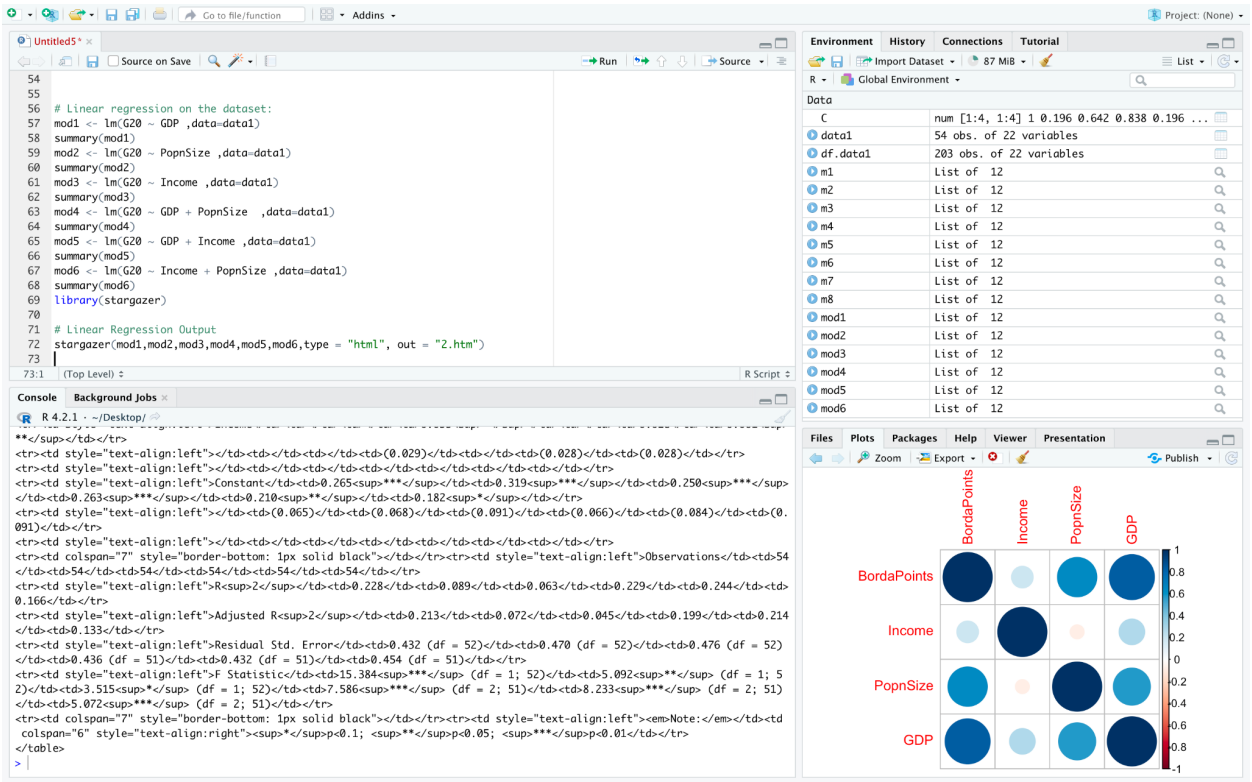
```
Call:
lm(formula = G20 ~ GDP, data = data1)

Residuals:
   Min      1Q  Median      3Q     Max
-0.7319 -0.2853 -0.2681  0.5004  0.7296

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.26487    0.06469   4.095 0.000148 ***
GDP          0.96550    0.24616   3.922 0.000258 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4323 on 52 degrees of freedom
Multiple R-squared:  0.2283,    Adjusted R-squared:  0.2135
F-statistic: 15.38 on 1 and 52 DF,  p-value: 0.0002584

> mod2 <- lm(G20 ~ PopnSize ,data=data1)
> summary(mod2)

Call:
lm(formula = G20 ~ PopnSize, data = data1)

Residuals:
   Min      1Q  Median      3Q     Max
-0.4078 -0.3374 -0.3229  0.6201  0.6770

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.31949    0.06777   4.714 1.86e-05 ***
PopnSize     0.78630    0.34845   2.257   0.0283 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4696 on 52 degrees of freedom
Multiple R-squared:  0.08919,  Adjusted R-squared:  0.07168
F-statistic: 5.092 on 1 and 52 DF,  p-value: 0.02826

> mod3 <- lm(G20 ~ Income ,data=data1)
> summary(mod3)

Call:
lm(formula = G20 ~ Income, data = data1)

Residuals:
   Min      1Q  Median      3Q     Max
```

```
-0.7826 -0.3079 -0.2761  0.5145  0.7305

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.24965    0.09136   2.733  0.00857 **
Income       0.05480    0.02923   1.875  0.06645 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4763 on 52 degrees of freedom
Multiple R-squared:  0.06331,  Adjusted R-squared:  0.0453
F-statistic: 3.515 on 1 and 52 DF,  p-value: 0.06645

> mod4 <- lm(G20 ~ GDP + PopnSize  ,data=data1)
> summary(mod4)

Call:
lm(formula = G20 ~ GDP + PopnSize, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6948 -0.2840 -0.2680  0.5033  0.7304

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.26318    0.06561   4.011 0.000198 ***
GDP          0.92142    0.30263   3.045 0.003679 **
PopnSize     0.10055    0.39431   0.255 0.799740
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4362 on 51 degrees of freedom
Multiple R-squared:  0.2293,    Adjusted R-squared:  0.1991
F-statistic: 7.586 on 2 and 51 DF,  p-value: 0.001306

> mod5 <- lm(G20 ~ GDP + Income ,data=data1)
> summary(mod5)

Call:
lm(formula = G20 ~ GDP + Income, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7050 -0.2677 -0.2289  0.4689  0.7703

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.21011    0.08364   2.512 0.015212 *
GDP          0.89325    0.25578   3.492 0.000999 ***
Income       0.02844    0.02757   1.032 0.307139
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.432 on 51 degrees of freedom
Multiple R-squared:  0.2441,    Adjusted R-squared:  0.2144
F-statistic: 8.233 on 2 and 51 DF,  p-value: 0.0007966

> mod6 <- lm(G20 ~ Income + PopnSize ,data=data1)
> summary(mod6)

Call:
```

```
lm(formula = G20 ~ Income + PopnSize, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7752 -0.2796 -0.2247  0.4949  0.7773

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.18228    0.09112   2.001   0.0508 .
Income       0.06053    0.02795   2.166   0.0350 *
PopnSize     0.84617    0.33783   2.505   0.0155 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4538 on 51 degrees of freedom
Multiple R-squared:  0.1659,    Adjusted R-squared:  0.1332
F-statistic: 5.072 on 2 and 51 DF,  p-value: 0.009792

> library(stargazer)
```

# Linear Regression Output
stargazer(mod1,mod2,mod3,mod4,mod5,mod6,type = "html", out = "2.htm")

	Dependent variable:					
	G20					
	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.966***			0.921***	0.893***	
	(0.246)			(0.303)	(0.256)	
PopnSize		0.786**		0.101		0.846**
		(0.348)		(0.394)		(0.338)
Income			0.055*		0.028	0.061**
			(0.029)		(0.028)	(0.028)
Constant	0.265***	0.319***	0.250***	0.263***	0.210**	0.182*
	(0.065)	(0.068)	(0.091)	(0.066)	(0.084)	(0.091)
Observations	54	54	54	54	54	54
R^2	0.228	0.089	0.063	0.229	0.244	0.166
Adjusted R^2	0.213	0.072	0.045	0.199	0.214	0.133
Residual Std. Error	0.432 (df = 52)	0.470 (df = 52)	0.476 (df = 52)	0.436 (df = 51)	0.432 (df = 51)	0.454 (df = 51)
F Statistic	15.384*** (df = 1; 52)	5.092** (df = 1; 52)	3.515* (df = 1; 52)	7.586*** (df = 2; 51)	8.233*** (df = 2; 51)	5.072*** (df = 2; 51)
Note:						*p<0.1; **p<0.05; ***p<0.01



#Including interaction between indepemdemt variables:
mod7 <- lm(G20 ~ GDP*Income+PopnSize,data=data1)
summary(mod7)
mod8 <- lm(G20 ~ GDP*Income*PopnSize,data = data1)
summary(mod8)
mod9 <- lm(G20 ~ GDP*PopnSize+Income,data=data1)
summary(mod9)

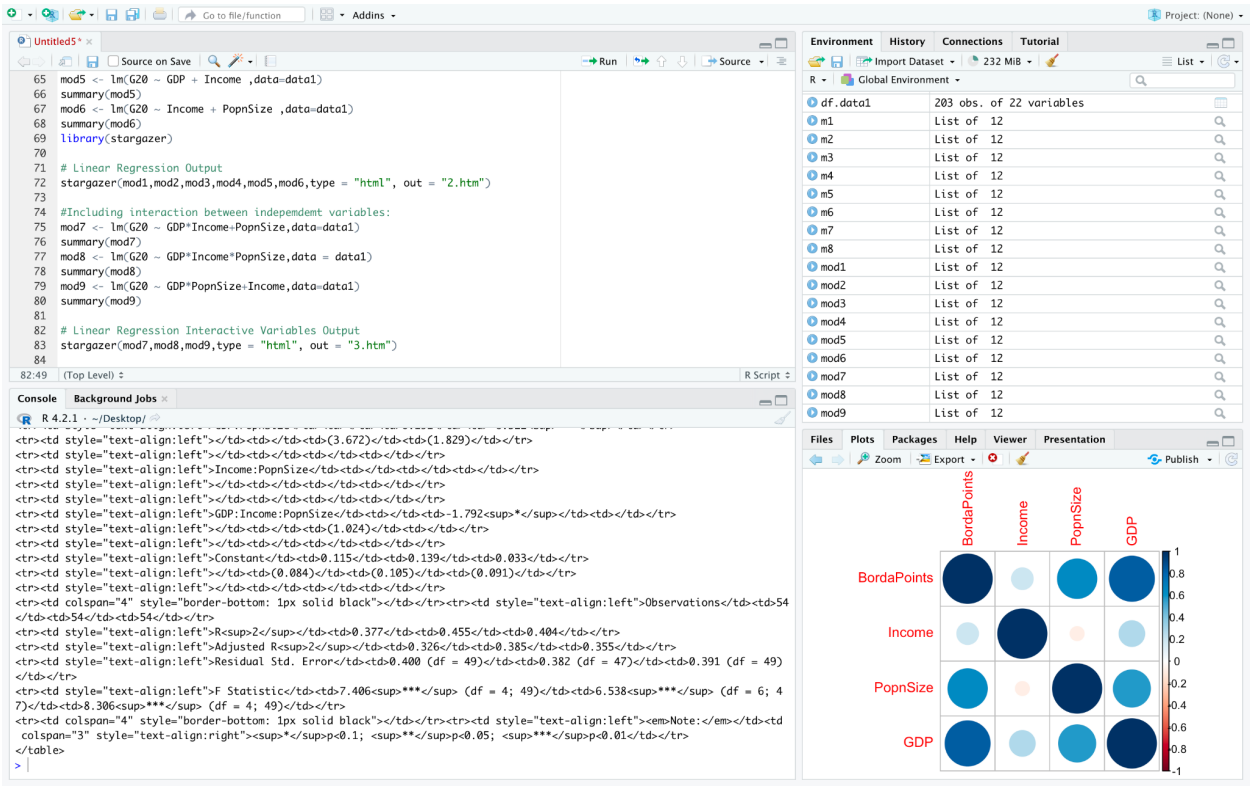
```
Call:
lm(formula = G20 ~ GDP * Income + PopnSize, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5635 -0.2694 -0.1523  0.2108  0.8910

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.11549    0.08445   1.368  0.17769
GDP          8.18153    2.35982   3.467  0.00111 **
Income       0.05846    0.02791   2.095  0.04140 *
PopnSize    -3.37900    1.20590  -2.802  0.00725 **
GDP:Income  -1.42372    0.44977  -3.165  0.00266 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4002 on 49 degrees of freedom
Multiple R-squared:  0.3768,    Adjusted R-squared:  0.3259
F-statistic: 7.406 on 4 and 49 DF,  p-value: 9.516e-05

> mod8 <- lm(G20 ~ GDP*Income*PopnSize,data = data1)
```

```
> summary(mod8)

Call:
lm(formula = G20 ~ GDP * Income * PopnSize, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5291 -0.2242 -0.1465  0.0812  0.8486

Coefficients: (1 not defined because of singularities)
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.138881   0.104969   1.323   0.1922
GDP                 5.808593   2.852097   2.037   0.0473 *
Income              0.005134   0.035346   0.145   0.8851
PopnSize           -1.802965   3.385484  -0.533   0.5968
GDP:Income         -0.467247   0.593474  -0.787   0.4351
GDP:PopnSize        0.190824   3.671827   0.052   0.9588
Income:PopnSize           NA         NA      NA       NA
GDP:Income:PopnSize -1.791889   1.024299  -1.749   0.0868 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3821 on 47 degrees of freedom
Multiple R-squared:  0.4549,    Adjusted R-squared:  0.3854
F-statistic: 6.538 on 6 and 47 DF,  p-value: 4.565e-05

> mod9 <- lm(G20 ~ GDP*PopnSize+Income,data=data1)
> summary(mod9)

Call:
lm(formula = G20 ~ GDP * PopnSize + Income, data = data1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6792 -0.2361 -0.1023  0.1652  0.8917

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03312    0.09103   0.364 0.717562
GDP          1.75423    0.40347   4.348 6.93e-05 ***
PopnSize     4.43583    1.23240   3.599 0.000742 ***
Income       0.03083    0.02621   1.176 0.245080
GDP:PopnSize -6.52178    1.82861  -3.567 0.000820 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3913 on 49 degrees of freedom
Multiple R-squared:  0.4041,    Adjusted R-squared:  0.3554
F-statistic: 8.306 on 4 and 49 DF,  p-value: 3.388e-05
```

Linear Regression Interactive Variables Output

```
stargazer(mod7,mod8,mod9,type = "html", out = "3.htm")
```

	Dependent variable:		
	G20		
	(1)	(2)	(3)
GDP	8.182***	5.809**	1.754***
	(2.360)	(2.852)	(0.403)
Income	0.058**	0.005	0.031
	(0.028)	(0.035)	(0.026)
PopnSize	-3.379***	-1.803	4.436***
	(1.206)	(3.385)	(1.232)
GDP:Income	-1.424***	-0.467	
	(0.450)	(0.593)	
GDP:PopnSize		0.191	-6.522***
		(3.672)	(1.829)
Income:PopnSize			
GDP:Income:PopnSize		-1.792*	
		(1.024)	
Constant	0.115	0.139	0.033
	(0.084)	(0.105)	(0.091)
Observations	54	54	54
R^2	0.377	0.455	0.404
Adjusted R^2	0.326	0.385	0.355
Residual Std. Error	0.400 (df = 49)	0.382 (df = 47)	0.391 (df = 49)
F Statistic	7.406*** (df = 4; 49)	6.538*** (df = 6; 47)	8.306*** (df = 4; 49)
Note:			*p<0.1; **p<0.05; ***p<0.01



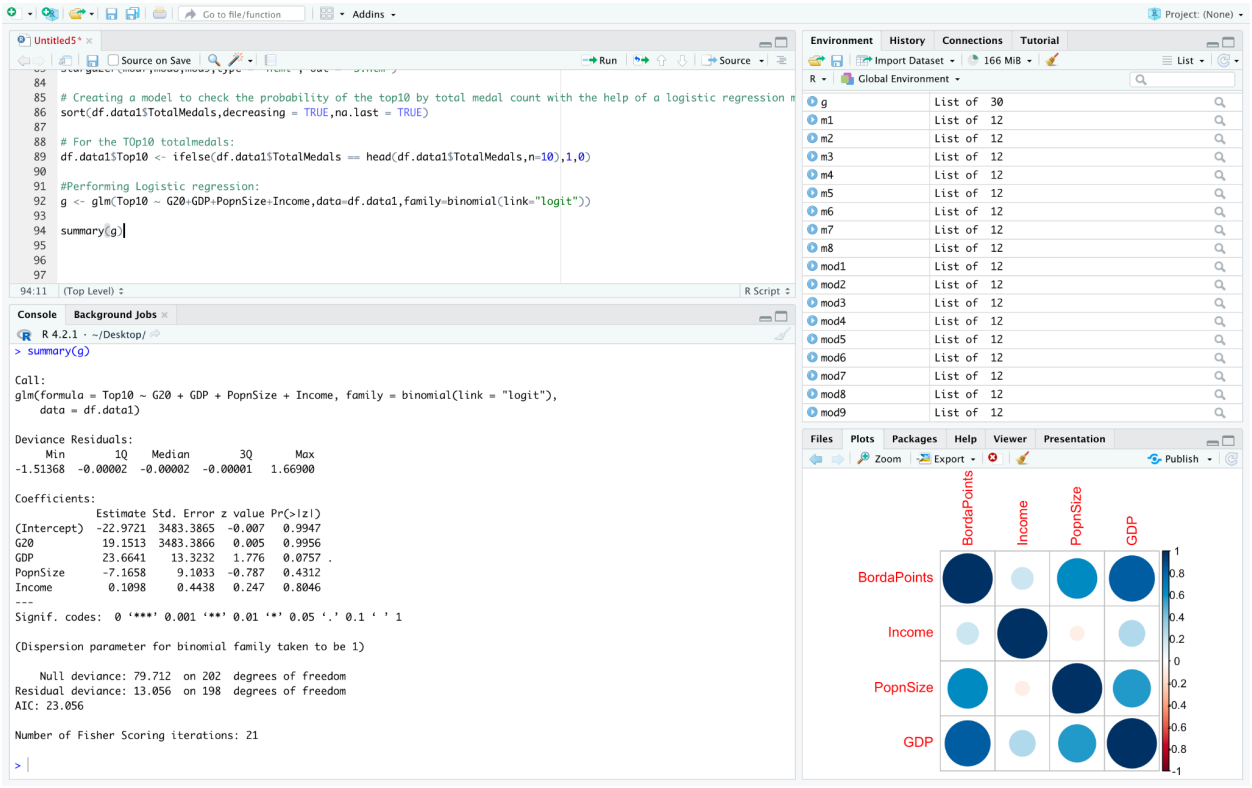
Q1.4) Finally, build a model to predict the probability of being at the top10 by total medal count using a logistic regression model. You may use the same predictors in parts b, and c or include others with clear reasoning. How do you explain the meaning of the significant coefficients in the logistic regression?

# Creating a model to check the probability of the top10 by total medal count with the help of a logistic regression model.
sort(df.data1$TotalMedals,decreasing = TRUE,na.last = TRUE)

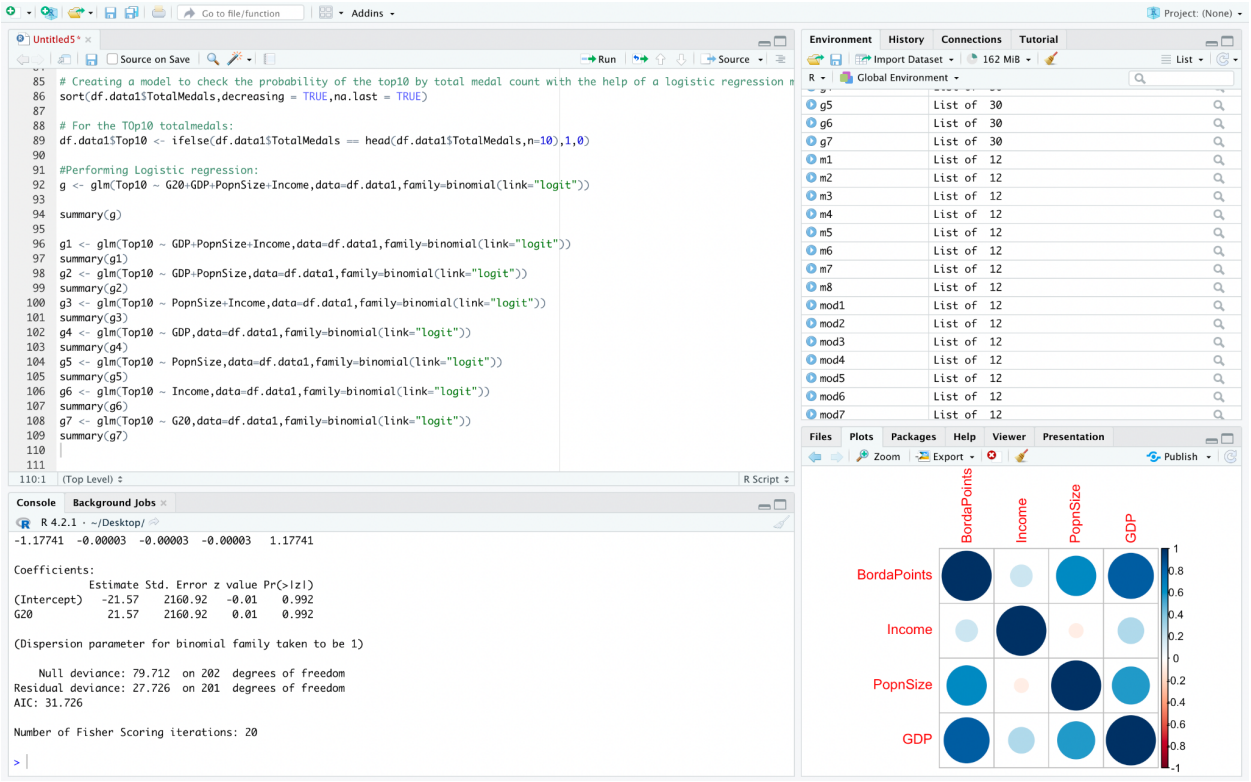
```
> sort(df.data1$TotalMedals,decreasing = TRUE,na.last = TRUE)
  [1] 104  88  82  65  44  38  35  34  28  28  20  20  18  17  17  17  14  13  13  12  12  12  11  10  10  10   9   9
 [29]   8   8   7   7   7   6   6   6   6   5   5   5   5   4   4   4   4   4   4   4   4   3   3   3   3   3   2   2
 [57]   2   2   2   2   2   2   2   2   2   2   2   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
 [85]   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
[113]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
[141]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
[169]   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
[197]   0   0   0   0   0   0   0
>
```

For the TOp10 totalmedals:
df.data1$Top10 <- ifelse(df.data1$TotalMedals == head(df.data1$TotalMedals,n=10),1,0)
#Performing Logistic regression:
g <- glm(Top10 ~ G20+GDP+PopnSize+Income,data=df.data1,family=binomial(link="logit"))
summary(g)

```
Call:
glm(formula = Top10 ~ G20 + GDP + PopnSize + Income, family = binomial(link = "logit"),
    data = df.data1)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.51368  -0.00002  -0.00002  -0.00001   1.66900

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -22.9721  3483.3865  -0.007   0.9947
G20          19.1513  3483.3866   0.005   0.9956
GDP          23.6641    13.3232   1.776   0.0757 .
PopnSize     -7.1658     9.1033  -0.787   0.4312
Income        0.1098     0.4438   0.247   0.8046
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 79.712  on 202  degrees of freedom
Residual deviance: 13.056  on 198  degrees of freedom
AIC: 23.056

Number of Fisher Scoring iterations: 21

>
```



```
g1 <- glm(Top10 ~ GDP+PopnSize+Income,data=df.data1,family=binomial(link="logit"))
summary(g1)
g2 <- glm(Top10 ~ GDP+PopnSize,data=df.data1,family=binomial(link="logit"))
summary(g2)
g3 <- glm(Top10 ~ PopnSize+Income,data=df.data1,family=binomial(link="logit"))
summary(g3)
g4 <- glm(Top10 ~ GDP,data=df.data1,family=binomial(link="logit"))
summary(g4)
g5 <- glm(Top10 ~ PopnSize,data=df.data1,family=binomial(link="logit"))
summary(g5)
g6 <- glm(Top10 ~ Income,data=df.data1,family=binomial(link="logit"))
summary(g6)
g7 <- glm(Top10 ~ G20,data=df.data1,family=binomial(link="logit"))
summary(g7)
```




```
# Logistic Regression Output
stargazer(g1,g2,g3,g4,g5,g6,g7,type = "html", out = "4.htm")
```

	Dependent variable:						
	Top10						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDP	41.735***	42.507***		33.717***			
	(12.310)	(11.782)		(8.468)			
PopnSize	-14.832*	-15.567**	4.001***		3.576**		
	(8.953)	(7.660)	(1.441)		(1.402)		
Income	0.065		0.225***			0.189**	
	(0.323)		(0.083)			(0.080)	
G20							21.566
							(2,160.924)
Constant	-6.501***	-6.339***	-3.793***	-6.275***	-3.219***	-3.394***	-21.566
	(1.731)	(1.467)	(0.481)	(1.359)	(0.362)	(0.415)	(2,160.924)
Observations	203	203	203	203	203	203	203
Log Likelihood	-7.814	-7.833	-32.859	-9.144	-35.847	-37.590	-13.863
Akaike Inf. Crit.	23.627	21.666	71.717	22.288	75.694	79.181	31.726
Note:							*p<0.1; **p<0.05; ***p<0.01

##Creating the confusion matrix:
original<- df.data1$Top10
predicted<-round(predict(g,type="response"))
table(predicted,original)

```
 ##Creating the confusion matrix:
> original<- df.data1$Top10
> predicted<-round(predict(g,type="response"))
> table(predicted,original)
         original
predicted   0   1
        0 191   2
        1   2   8
```

```
g <- glm(Top10 ~ G20+GDP+PopnSize+Income,data=df.data1,family=binomial(link="logit"))

summary(g)

g1 <- glm(Top10 ~ GDP+PopnSize+Income,data=df.data1,family=binomial(link="logit"))
summary(g1)
g2 <- glm(Top10 ~ GDP+PopnSize,data=df.data1,family=binomial(link="logit"))
summary(g2)
g3 <- glm(Top10 ~ PopnSize+Income,data=df.data1,family=binomial(link="logit"))
summary(g3)
g4 <- glm(Top10 ~ GDP,data=df.data1,family=binomial(link="logit"))
summary(g4)
g5 <- glm(Top10 ~ PopnSize,data=df.data1,family=binomial(link="logit"))
summary(g5)
g6 <- glm(Top10 ~ Income,data=df.data1,family=binomial(link="logit"))
summary(g6)
g7 <- glm(Top10 ~ G20,data=df.data1,family=binomial(link="logit"))
summary(g7)

# Logistic Regression Output
stargazer(g1,g2,g3,g4,g5,g6,g7,type = "html", out = "4.htm")

##Creating the confusion matrix:
original<- df.data1$Top10
predicted<-round(predict(g,type="response"))
table(predicted,original)
```

```
<tr><td style="text-align:left">Log Likelihood</td><td>-7.814</td><td>-7.833</td><td>-32.859</td><td>-9.144</td><td>-35.847
</td><td>-37.590</td><td>-13.863</td></tr>
<tr><td style="text-align:left">Akaike Inf. Crit.</td><td>23.627</td><td>21.666</td><td>71.717</td><td>22.288</td><td>75.694
</td><td>79.181</td><td>31.726</td></tr>
<tr><td colspan="8" style="border-bottom: 1px solid black"></td></tr><tr><td style="text-align:left"><em>Note:</em></td><td
 colspan="7" style="text-align:right"><sup>*</sup>p<0.1; <sup>**</sup>p<0.05; <sup>***</sup>p<0.01</td></tr>
</table>
> ##Creating the confusion matrix:
> original<- df.data1$Top10
> predicted<-round(predict(g,type="response"))
> table(predicted,original)
         original
predicted   0   1
        0 191   2
        1   2   8
>
```