MATH 4330 PROJECT

Francis Calingo

ABSTRACT

Despite Canada being a wealthy nation relative to much of the world, inequalities still exist here. Socioeconomic conditions are not uniform and can vary greatly from census division to census division.

What I want to do is collate different data of census subdivisions across Canada, and observe both their demographics and socioeconomic indicators. After selecting specific variables from certain datasets and performing regression analysis on the model I will be creating, I can draw conclusions and formulate answers for the following questions:

(1) Which socioeconomic issue has had the most substantial effect on out-migration?
(2) Could there also be other explanations and factors that explain why out-migration is higher in some regions?

DATASETS AND VARIABLES OF INTEREST

Datasets

Dataset 1:https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E Dataset 1 Selected variables (9): Census Division, Province, 2016 Population, Post secondary credentials attainment rate, Average income, Unemployment rate, Visible Minority Rate, Indigenous Rate, Median Age

Dataset 2: Add/Remove data - Components of population change by census division, 2016 boundaries Dataset 2 Selected variables (12): Mortality (2015-16, 2016-17, 2017-18, 2018-19), Net inter and intra-provincial migration (2015-16, 2016-17, 2017-18, 2018-19)

New variables (9): 4-year average mortality, 4-year average mortality per million people, Net migration (sum of net interprovincial and intraprovincial migration for 2015-16, 2016-17, 2017-18, 2018-19), 4-year average of calculated net migration, 4-year average of calculated net migration per million people, POC rate (sum of Visible Minority and Indigenous rate)

Categorized Variables (7): Mig, Mort, Post.Sec, Un, VM, Ind, POC

Variables of interest:

Mig - 4-year average of calculated net migration per million people Mort - 4-year average mortality per million people Post.Sec - Post secondary credentials attainment rate Avg.Inc - Average income Un - Unemployment rate VM - Visible minority rate Ind - Indigenous rate POC - Combined visible minority and Indigenous rate Med.Age - Median age

Reasoning Behind Variable Selection:

(1) Using yearly average for variables whose data is collected yearly (e.g., mortality) instead of one year will avoid inadvertently using a year where special circumstances occurred.

(2) Speaking of special circumstances, I intentionally avoided using the 2019/20 data to exclude the extraordinary impacts of the COVID-19 pandemic.
(3) I used variables that I felt are related to socioeconomic push factors (Mort and Med.Age could be related to health, Post.Sec could be related to educational opportunities, Avg.Inc and Un are related to each observation's economic conditions)
(4) I included VM and Ind, not because they are direct push factors, but because I felt that due to Canada's ongoing legacy of systemic racism, both groups continue to be socioeconomically disadvantaged compared to the general population. Therefore, I want to explore their relation to the socioeconomic conditions of each observation.

MODELLING APPROACH AND METHODOLOGY:

Mig - categorical response variable 1 if average migration rate is below -5000 2 if average migration rate is between -5000 and -1000 3 if average migration rate is between -1000 and 1000 4 if average migration rate is between 1000 and 5000 5 if average migration rate is greater than 5000

Mort - categorical predictor variable 1 if average mortality rate is below 5000, 5000-7500, 7500-10,000, 10,000-12,500, 12,500+ 2 if average mortality rate is between 5000 and 7500 3 if average mortality rate is between 7500 and 10,000 4 if average mortalityrate is between 10,000 and 12,500 5 if average mortality rate is greater than 12,500

Post.Sec - categorical predictor variable 1 if post secondary credentials attainment rate is below 30% 2 if if post secondary credentials attainment rate is between 30-40% 3 if if post secondary credentials attainment rate is between 40-50% 4 if if post secondary credentials attainment rate is between 50-60% 5 if if post secondary credentials attainment rate is greater than 60%

Avg.Inc - continuous predictor variable

Un - categorical predictor variable 1 if unemployment rate is below 5% 2 if unemployment rate is between 5-10% 3 if unemployment rate is between 10-15% 4 if unemployment rate is between 15-20% 5 if unemployment rate is greater than 20%

VM - categorical predictor variable 1 if Visible Minority rate is below 5% 2 if Visible Minority rate is between 5-10% 3 if Visible Minority rate is between 10-20% 4 if Visible Minority rate is between 20-40% 5 if Visible Minority rate is greater than 40%

Ind - categorical predictor variable 1 if Indigenous rate is below 5% 2 if Indigenous rate is between 5-10% 3 if Indigenous rate is between 10-20% 4 if Indigenous rate is between 20-40% #5 if Indigenous rate is greater than 40%

POC - categorical predictor variable 1 if POC rate is below 5% 2 if POC rate is between 5-10% 3 if POC rate is between 10-20% 4 if POC rate is between 20-40% 5 if POC rate is greater than 40%

Med.Age - continuous predictor variable

I want to test Ind, VM, and POC as interaction variables on each variable of interest.

R ANALYSIS

```
library(car)
```

```
## Loading required package: carData
```

```
library(spida2)
library(effects)
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
library(lattice)
library(latticeExtra)
```

```
Canada <- read.csv("C:\\Users\\francali\\Documents\\Census Division Stats - Sheet1.csv")

summary(Canada$Avg.Income)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   21367   37163   41439   42570   46489   98358
```

```
summary(Canada$MedAge)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.20   40.60   44.80   44.32   49.50   57.50
```

```
Mig<-Canada$Mig
Mort<-Canada$Mort
Post.Sec<-Canada$Post.Sec
Avg.Income<-Canada$Avg.Income
Un<-Canada$Un
VM<-Canada$VM
Ind<-Canada$Ind
MedAge<-Canada$MedAge
POC<-Canada$POC

class(Mig)
```

```
## [1] "integer"
```

```
class(Mort)
```

```
## [1] "integer"
```

```
class(Post.Sec)
```

```
## [1] "integer"
```

```
class(Avg.Income)
```

```
## [1] "integer"
```

```
class(Un)
```

```
## [1] "integer"
```

```
class(VM)
```

```
## [1] "integer"
```

```
class(Ind)
```

```
## [1] "integer"
```

```
class(MedAge)
```

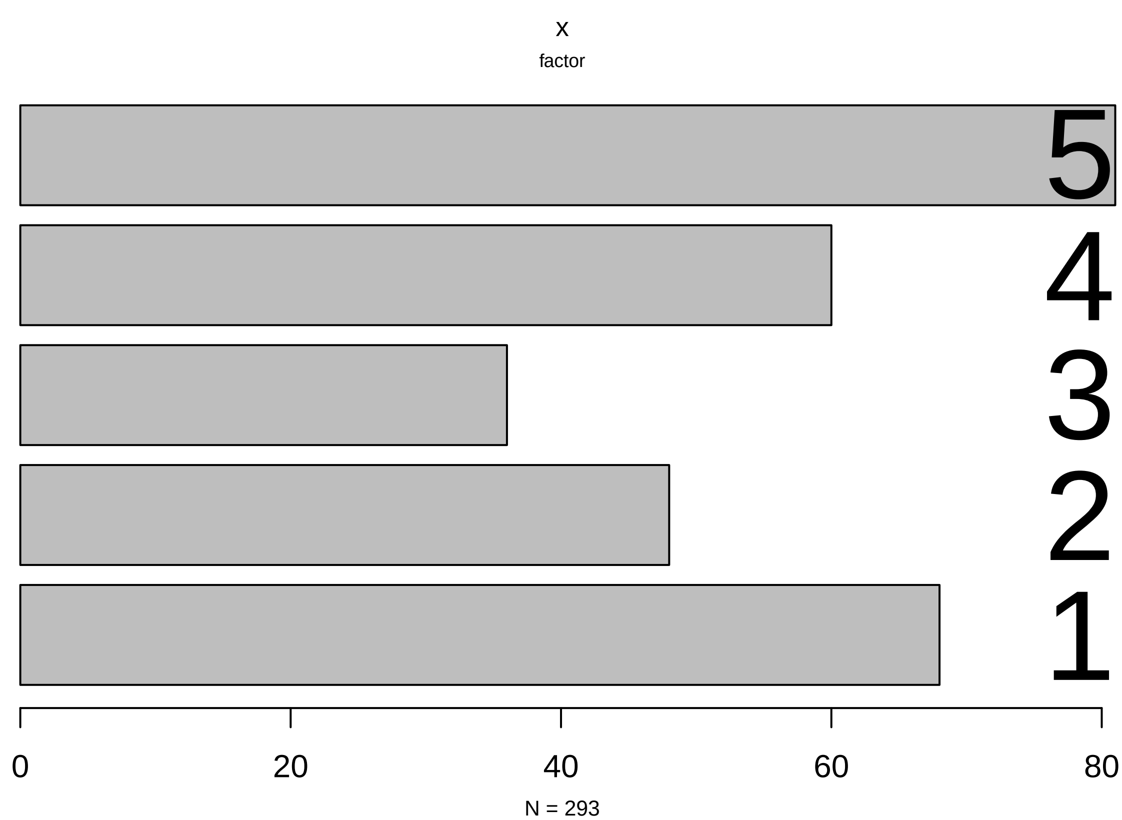
```
## [1] "numeric"
```

```
class(POC)
```

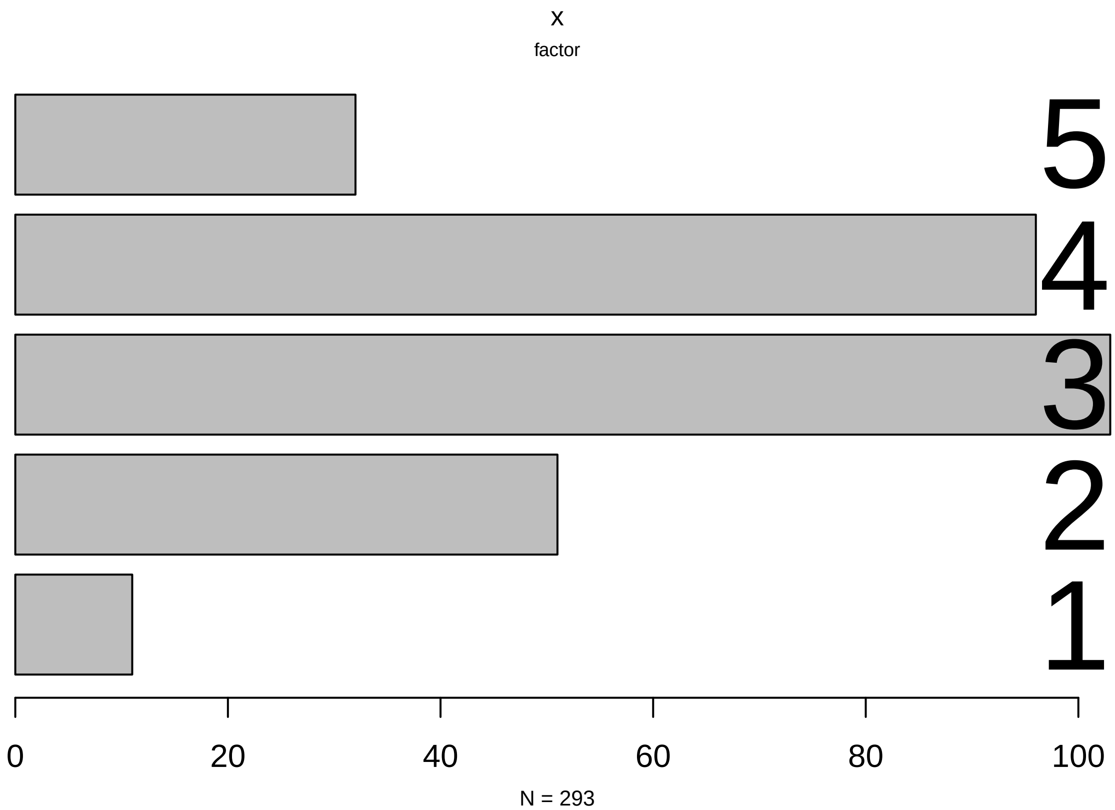
```
## [1] "integer"
```

```
Mig<-as.factor(Mig)
Mort<-as.factor(Mort)
Post.Sec<-as.factor(Post.Sec)
Un<-as.factor(Un)
VM<-as.factor(VM)
Ind<-as.factor(Ind)
POC<-as.factor(POC)
```

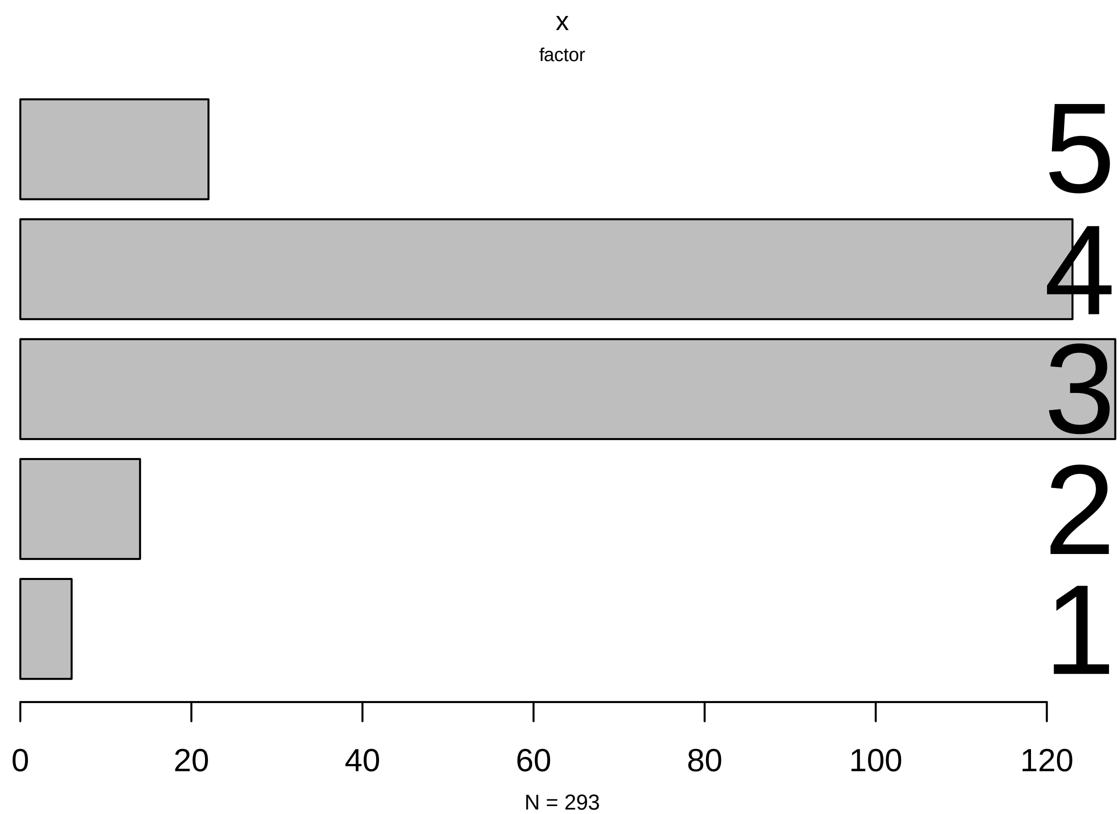
```
xqplot(Mig)
```



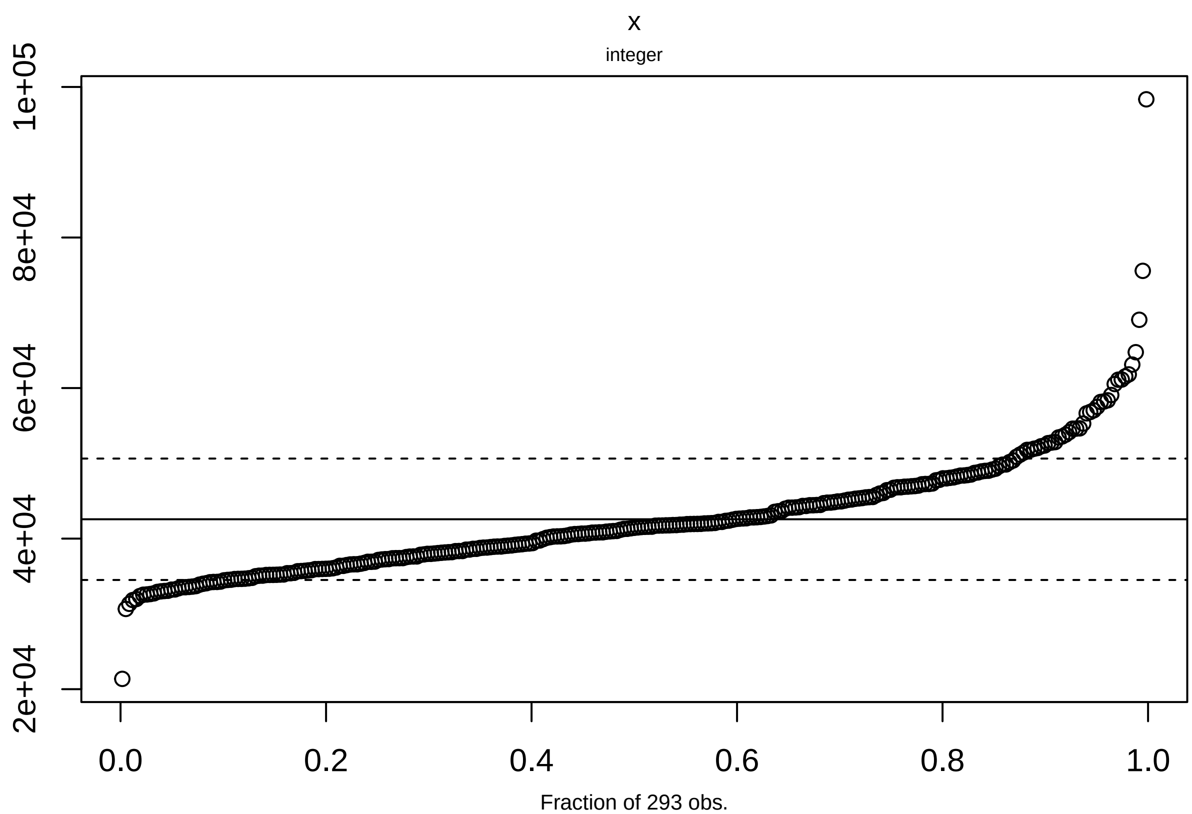
```
xqplot(Mort)
```



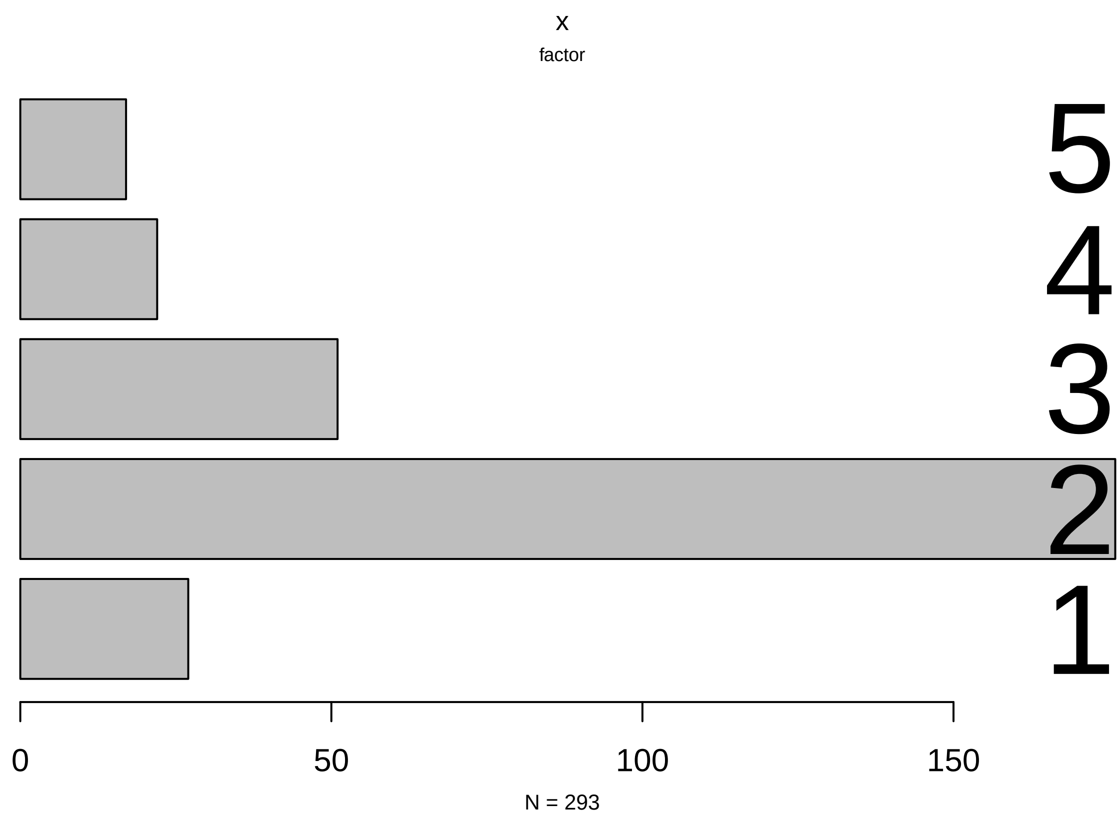

```
xqplot(Post.Sec)
```



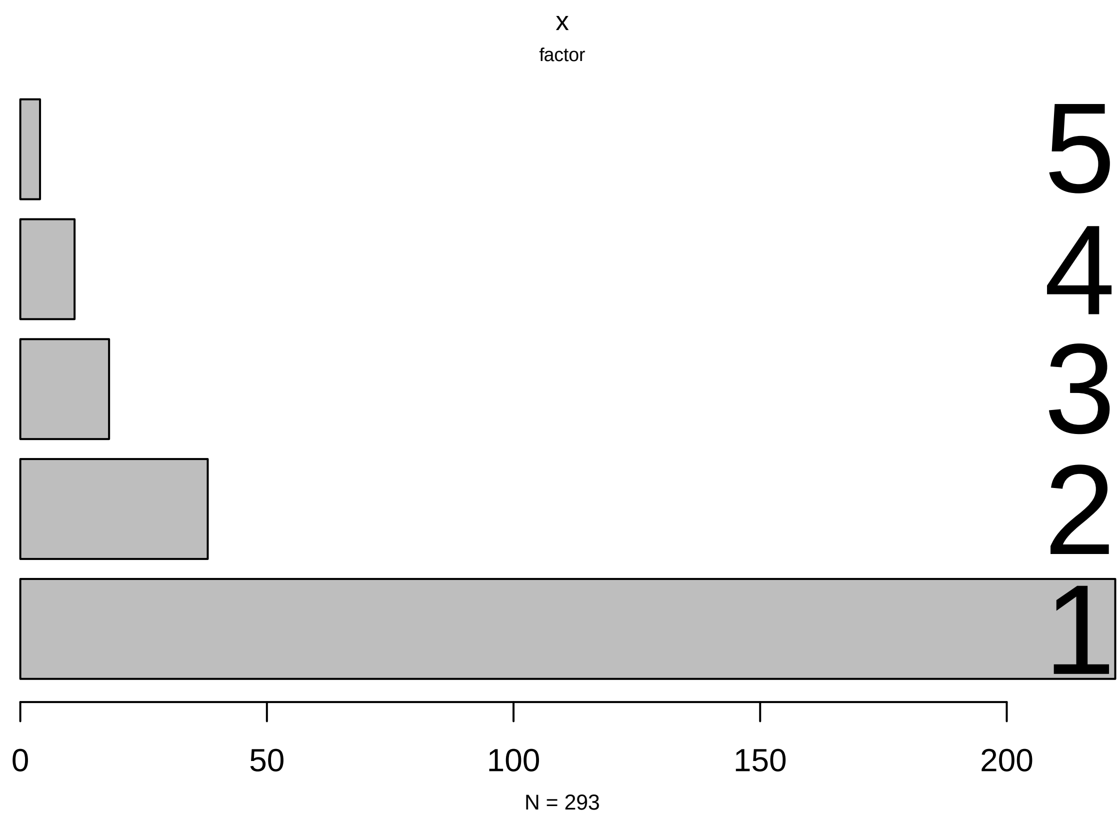

```
xqplot(Avg.Income)
```


x
integer
1e+05
8e+04
6e+04
4e+04
2e+04
0.0
0.2
0.4
0.6
0.8
1.0
Fraction of 293 obs.

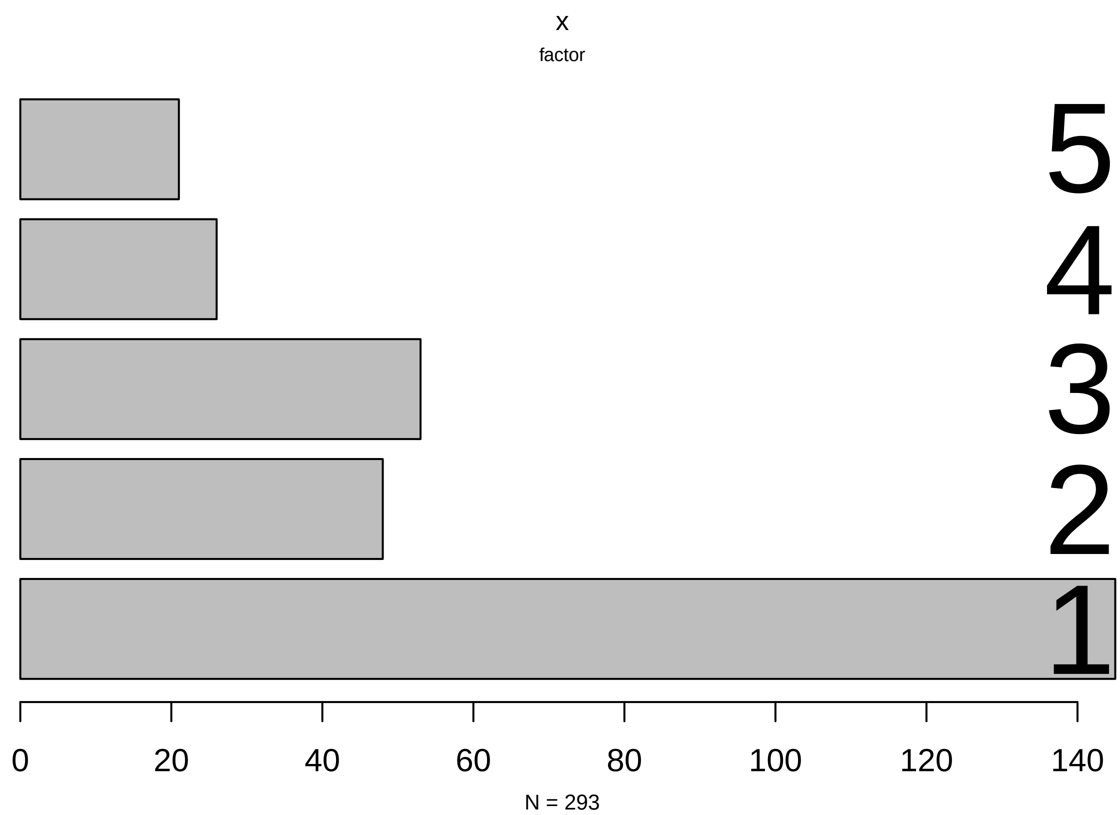
```
xqplot(Un)
```



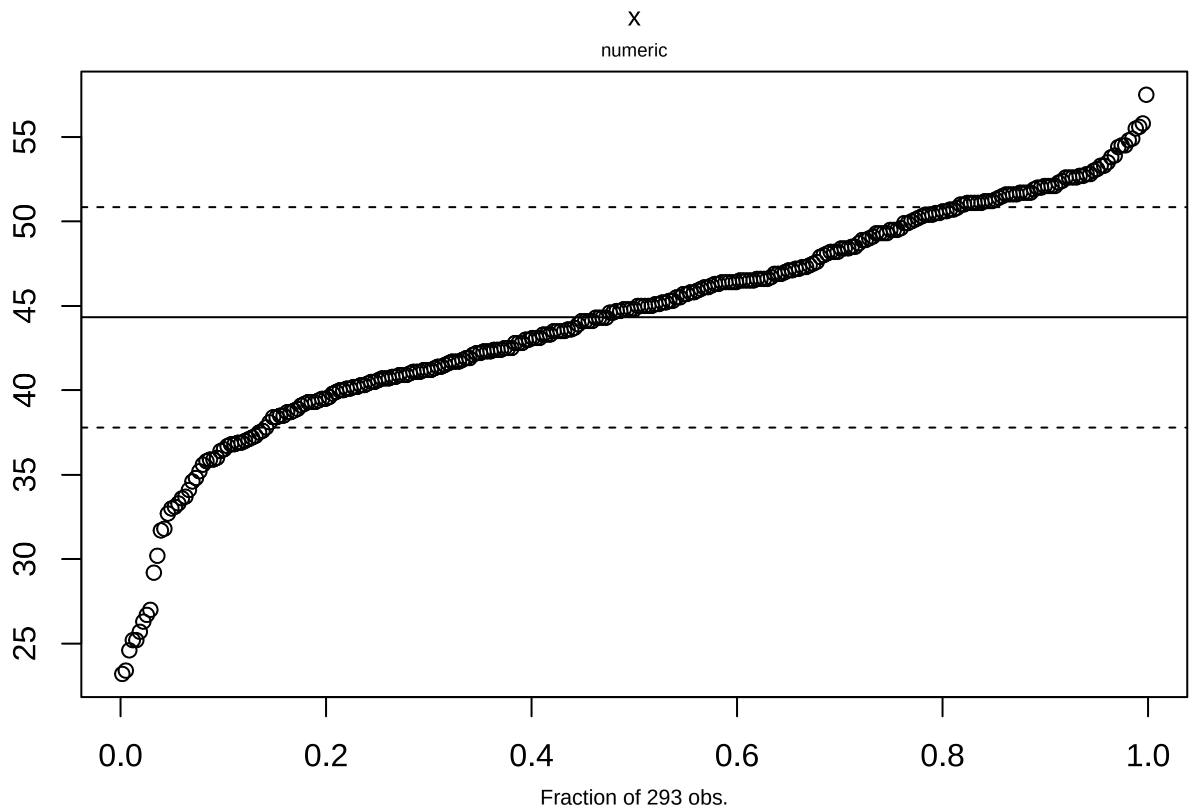

```
xqplot(VM)
```



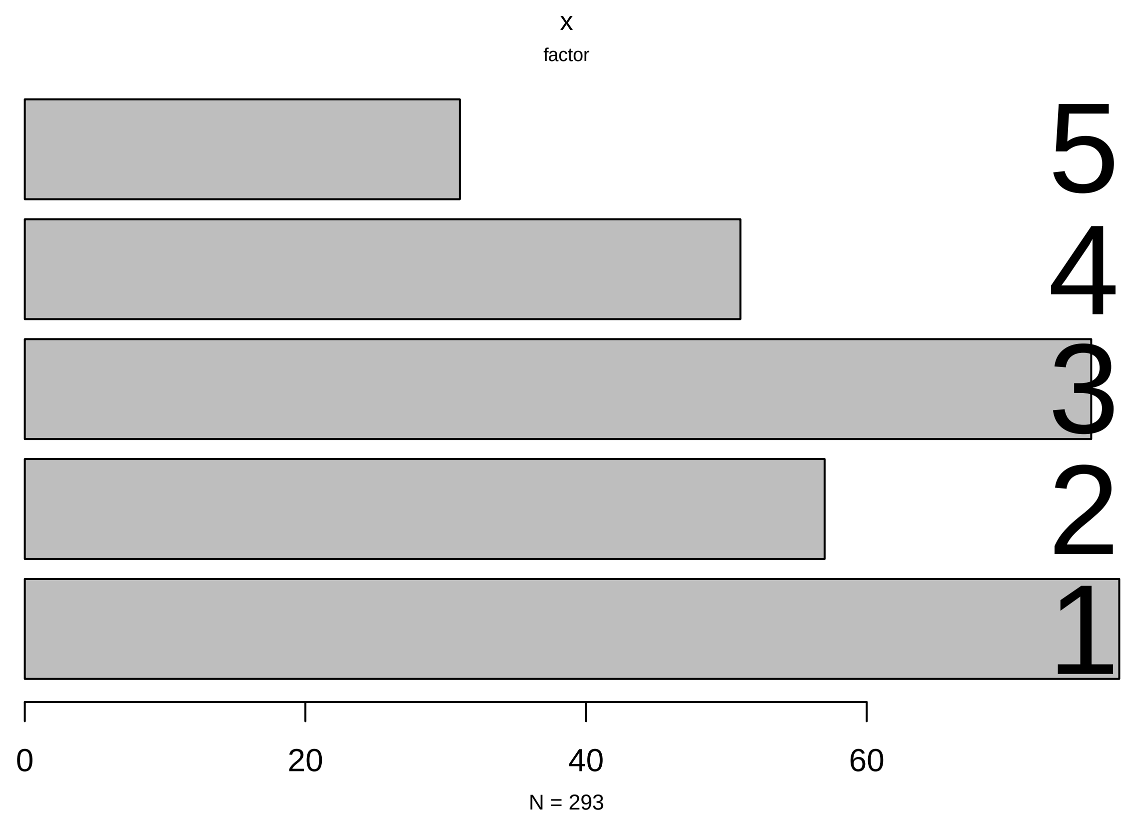

```
xqplot(Ind)
```




```
xqplot(MedAge)
```


x
numeric
55
50
45
40
35
30
25
0.0
0.2
0.4
0.6
0.8
1.0
Fraction of 293 obs.

```
xqplot(POC)
```



Notice that Mig is not normally distributed.

Let's try VM, Ind, and POC as interactions

```
Model.Int1_1 <- glm(Mig ~ Mort*Ind, family="binomial")
summary(Model.Int1_1)
```

```
##
## Call:
## glm(formula = Mig ~ Mort * Ind, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.46760   0.00022   0.40837   0.48535   1.79412
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.0986     1.1547   0.951    0.341
## Mort2            1.3437     1.3700   0.981    0.327
## Mort3            1.8971     1.2974   1.462    0.144
## Mort4            0.9808     1.2483   0.786    0.432
## Mort5            0.8473     1.5736   0.538    0.590
## Ind2            16.4675  3956.1805   0.004    0.997
## Ind3            -0.4055     1.6833  -0.241    0.810
```

```
## Ind4          16.4675  3956.1805   0.004    0.997
## Ind5          -1.0986     1.8257  -0.602    0.547
## Mort2:Ind2    -1.3437  4568.2037   0.000    1.000
## Mort3:Ind2   -18.6159  3956.1806  -0.005    0.996
## Mort4:Ind2   -16.1046  3956.1806  -0.004    0.997
## Mort5:Ind2    -0.8473  4170.1805   0.000    1.000
## Mort2:Ind3    -2.5477     1.9774  -1.288    0.198
## Mort3:Ind3    -2.5903     1.8529  -1.398    0.162
## Mort4:Ind3    -1.4227     1.8200  -0.782    0.434
## Mort5:Ind3    -0.1542     2.1450  -0.072    0.943
## Mort2:Ind4   -20.2961  3956.1807  -0.005    0.996
## Mort3:Ind4   -19.1755  3956.1806  -0.005    0.996
## Mort4:Ind4   -17.2941  3956.1806  -0.004    0.997
## Mort5:Ind4   -19.5120  3956.1808  -0.005    0.996
## Mort2:Ind5   -18.9098  1251.0556  -0.015    0.988
## Mort3:Ind5    -2.8134     2.0936  -1.344    0.179
## Mort4:Ind5    16.5852  3956.1808   0.004    0.997
## Mort5:Ind5    16.7188  3956.1809   0.004    0.997
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 214.41  on 268  degrees of freedom
## AIC: 264.41
##
## Number of Fisher Scoring iterations: 16
```

```
Model.Int1_2 <- glm(Mig ~ Mort*VM, family="binomial")
summary(Model.Int1_2)
```

```
##
## Call:
## glm(formula = Mig ~ Mort * VM, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.09629   0.00036   0.57252   0.63352   1.48230
##
## Coefficients: (7 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.60944    1.09545   1.469   0.1418
## Mort2         -1.94591    1.17108  -1.662   0.0966 .
## Mort3         -0.10536    1.13459  -0.093   0.9260
## Mort4          0.11607    1.13606   0.102   0.9186
## Mort5         -0.04082    1.20069  -0.034   0.9729
## VM2           -0.87547    1.31972  -0.663   0.5071
## VM3           14.95663 1696.73470   0.009   0.9930
## VM4           -1.60944    1.78885  -0.900   0.3683
## VM5          -18.17551 2399.54497  -0.008   0.9940
## Mort2:VM2      1.54841    1.50198   1.031   0.3026
## Mort3:VM2     -0.04082    1.46290  -0.028   0.9777
## Mort4:VM2      0.40272    1.57326   0.256   0.7980
## Mort5:VM2           NA         NA      NA       NA
## Mort2:VM3    -13.01072 1696.73510  -0.008   0.9939
```

```
## Mort3:VM3    -14.38127 1696.73505  -0.008   0.9932
## Mort4:VM3    -33.24821 2938.83031  -0.011   0.9910
## Mort5:VM3           NA         NA      NA       NA
## Mort2:VM4     18.51198  979.61175   0.019   0.9849
## Mort3:VM4     -0.58779    2.18799  -0.269   0.7882
## Mort4:VM4           NA         NA      NA       NA
## Mort5:VM4           NA         NA      NA       NA
## Mort2:VM5     19.20513 2399.54532   0.008   0.9936
## Mort3:VM5           NA         NA      NA       NA
## Mort4:VM5           NA         NA      NA       NA
## Mort5:VM5           NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 280.74  on 275  degrees of freedom
## AIC: 316.74
##
## Number of Fisher Scoring iterations: 15
```

```
Model.Int1_3 <- glm(Mig ~ Mort*POC, family="binomial")
summary(Model.Int1_3)
```

```
##
## Call:
## glm(formula = Mig ~ Mort * POC, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.29741   0.00013   0.38499   0.77754   2.00744
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.857e+01  4.612e+03   0.004    0.997
## Mort2         1.380e-07  5.230e+03   0.000    1.000
## Mort3         2.492e-07  4.742e+03   0.000    1.000
## Mort4        -1.600e+01  4.612e+03  -0.003    0.997
## Mort5        -1.696e+01  4.612e+03  -0.004    0.997
## POC2          6.248e-07  7.989e+03   0.000    1.000
## POC3          3.786e-07  7.989e+03   0.000    1.000
## POC4         -1.787e+01  4.612e+03  -0.004    0.997
## POC5         -1.857e+01  4.612e+03  -0.004    0.997
## Mort2:POC2   -1.747e+01  8.360e+03  -0.002    0.998
## Mort3:POC2   -1.649e+01  8.064e+03  -0.002    0.998
## Mort4:POC2   -7.324e-01  7.989e+03   0.000    1.000
## Mort5:POC2    1.696e+01  8.504e+03   0.002    0.998
## Mort2:POC3   -1.747e+01  8.360e+03  -0.002    0.998
## Mort3:POC3   -1.776e+01  8.064e+03  -0.002    0.998
## Mort4:POC3   -1.523e+00  7.989e+03   0.000    1.000
## Mort5:POC3    9.555e-01  7.989e+03   0.000    1.000
## Mort2:POC4    1.178e-01  5.230e+03   0.000    1.000
## Mort3:POC4   -2.877e-01  4.742e+03   0.000    1.000
```

```
## Mort4:POC4   1.622e+01  4.612e+03   0.004    0.997
## Mort5:POC4   1.557e+01  4.612e+03   0.003    0.997
## Mort2:POC5  -1.872e+00  5.230e+03   0.000    1.000
## Mort3:POC5  -6.931e-01  4.742e+03   0.000    1.000
## Mort4:POC5   3.457e+01  6.523e+03   0.005    0.996
## Mort5:POC5   3.552e+01  7.989e+03   0.004    0.996
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 232.57  on 268  degrees of freedom
## AIC: 282.57
##
## Number of Fisher Scoring iterations: 17
```

There is quite a lot of high p-values, implying that there is lack of statistical significance. However, I still do want to analyze the estimates. For VM, there is a lot of singularities for high-level Mort interacting on VM. Since most visible minorities settle in Canada's biggest urban areas, and many big urban areas are assigned one census division each, most visible minority rates in many census divisions are negligible. Also, the larger urban areas tend to have higher net migration due to better economic opportunities, notwithstanding the contrasting socioeconomic conditions within urban areas. POC, as mentioned before, is the sum of VM and Ind. Therefore, it would be better to use Ind as interaction. There is quite a wide discrepancy when it comes to estimates for different mortality and Indigenous groups, implying that the effect on each group is vastly different. However, the high standard errors among many terms (along with the aforementioned high p-values) demonstrate that Ind would not be the best interaction variable for Mort.

```
Model.Int2_1 <- glm(Mig ~ Post.Sec*Ind, family="binomial")
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(Model.Int2_1)
```

```
##
## Call:
## glm(formula = Mig ~ Post.Sec * Ind, family = "binomial")
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
##  -8.49    0.00    0.00    0.00    8.49
##
## Coefficients: (6 not defined because of singularities)
##                  Estimate Std. Error    z value Pr(>|z|)
## (Intercept)     7.387e+15  5.533e+07  133511359   <2e-16 ***
## Post.Sec2      -5.136e+15  7.289e+07  -70454652   <2e-16 ***
## Post.Sec3      -3.604e+15  5.614e+07  -64204538   <2e-16 ***
## Post.Sec4      -3.002e+15  5.479e+07  -54794158   <2e-16 ***
## Post.Sec5      -3.149e+15  5.768e+07  -54594214   <2e-16 ***
## Ind2            2.647e+14  6.905e+07    3833751   <2e-16 ***
## Ind3           -1.987e+15  5.017e+07  -39605444   <2e-16 ***
## Ind4            2.649e+14  5.017e+07    5280726   <2e-16 ***
```

```
## Ind5           -6.637e+15  4.807e+07 -138057120   <2e-16 ***
## Post.Sec2:Ind2         NA         NA         NA       NA
## Post.Sec3:Ind2 -5.677e+14  7.116e+07   -7978248   <2e-16 ***
## Post.Sec4:Ind2 -1.469e+14  7.077e+07   -2075688   <2e-16 ***
## Post.Sec5:Ind2         NA         NA         NA       NA
## Post.Sec2:Ind3 -3.643e+15  7.678e+07  -47445291   <2e-16 ***
## Post.Sec3:Ind3 -3.748e+15  5.251e+07  -71376250   <2e-16 ***
## Post.Sec4:Ind3 -3.988e+15  5.330e+07  -74816235   <2e-16 ***
## Post.Sec5:Ind3         NA         NA         NA       NA
## Post.Sec2:Ind4 -4.769e+15  7.678e+07  -62110199   <2e-16 ***
## Post.Sec3:Ind4 -1.929e+15  5.359e+07  -35989198   <2e-16 ***
## Post.Sec4:Ind4 -3.144e+15  6.385e+07  -49241551   <2e-16 ***
## Post.Sec5:Ind4         NA         NA         NA       NA
## Post.Sec2:Ind5 -1.380e+14  7.542e+07   -1829081   <2e-16 ***
## Post.Sec3:Ind5 -1.485e+14  5.387e+07   -2757587   <2e-16 ***
## Post.Sec4:Ind5         NA         NA         NA       NA
## Post.Sec5:Ind5         NA         NA         NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance:  317.48  on 292  degrees of freedom
## Residual deviance: 4901.94  on 274  degrees of freedom
## AIC: 4939.9
##
## Number of Fisher Scoring iterations: 25
```

Possible Hauck-Donner phenomenon for Post.Sec

```
Model.Int3_1 <- glm(Mig ~ Avg.Income*Ind, family="binomial")
summary(Model.Int3_1)
```

```
##
## Call:
## glm(formula = Mig ~ Avg.Income * Ind, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5002   0.1937   0.3730   0.4811   1.8145
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.810e+00  1.860e+00   2.586  0.00972 **
## Avg.Income      -5.641e-05  4.205e-05  -1.342  0.17974
## Ind2             4.987e+00  3.846e+00   1.297  0.19475
## Ind3            -2.961e-01  2.615e+00  -0.113  0.90983
## Ind4            -3.142e+00  2.624e+00  -1.197  0.23114
## Ind5            -4.232e+00  2.993e+00  -1.414  0.15738
## Avg.Income:Ind2 -1.139e-04  8.170e-05  -1.394  0.16347
## Avg.Income:Ind3 -3.765e-05  5.839e-05  -0.645  0.51901
## Avg.Income:Ind4  2.168e-05  5.913e-05   0.367  0.71386
## Avg.Income:Ind5  1.387e-05  7.100e-05   0.195  0.84508
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 229.00  on 283  degrees of freedom
## AIC: 249
##
## Number of Fisher Scoring iterations: 5
```

Again there is discrepancy when it comes to each group, but the standard errors are much more acceptable. The p-values, while still not statistically significant, are much lower.

```
Model.Int4_1 <- glm(Mig ~ Un*Ind, family="binomial")
summary(Model.Int4_1)
```

```
##
## Call:
## glm(formula = Mig ~ Un * Ind, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.32725   0.00013   0.37708   0.45118   2.03933
##
## Coefficients: (3 not defined because of singularities)
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.251e+00  7.434e-01   3.028  0.00246 **
## Un2          3.567e-01  8.402e-01   0.424  0.67121
## Un3          3.878e-01  1.274e+00   0.304  0.76092
## Un4          1.631e+01  3.766e+03   0.004  0.99654
## Un5         -2.251e+00  1.246e+00  -1.807  0.07080 .
## Ind2        -2.944e+00  1.433e+00  -2.055  0.03986 *
## Ind3         1.631e+01  3.766e+03   0.004  0.99654
## Ind4        -4.859e-14  1.732e+00   0.000  1.00000
## Ind5        -1.946e+00  1.464e+00  -1.329  0.18375
## Un2:Ind2     2.570e+00  1.605e+00   1.602  0.10925
## Un3:Ind2     1.887e+01  2.465e+03   0.008  0.99389
## Un4:Ind2     2.944e+00  4.501e+03   0.001  0.99948
## Un5:Ind2            NA         NA      NA       NA
## Un2:Ind3    -1.886e+01  3.766e+03  -0.005  0.99600
## Un3:Ind3    -1.880e+01  3.766e+03  -0.005  0.99602
## Un4:Ind3    -1.631e+01  6.523e+03  -0.003  0.99800
## Un5:Ind3    -1.701e+01  3.766e+03  -0.005  0.99640
## Un2:Ind4    -2.048e+00  1.883e+00  -1.088  0.27672
## Un3:Ind4    -2.862e+00  2.126e+00  -1.346  0.17829
## Un4:Ind4    -1.857e+01  3.766e+03  -0.005  0.99607
## Un5:Ind4            NA         NA      NA       NA
## Un2:Ind5    -1.923e+01  6.523e+03  -0.003  0.99765
## Un3:Ind5    -4.055e-01  1.949e+00  -0.208  0.83518
## Un4:Ind5    -3.519e+01  4.763e+03  -0.007  0.99411
## Un5:Ind5            NA         NA      NA       NA
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 216.05  on 271  degrees of freedom
## AIC: 260.05
##
## Number of Fisher Scoring iterations: 17
```

Only Ind2 has positive net mig.

```r
Model.Int5_1 <- glm(Mig ~ MedAge*Ind, family="binomial")
summary(Model.Int5_1)
```

```
##
## Call:
## glm(formula = Mig ~ MedAge * Ind, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6463   0.1833   0.3704   0.5169   2.3928
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.69897    2.95258  -0.914   0.3607
## MedAge       0.11468    0.06754   1.698   0.0895 .
## Ind2        -5.76089    5.57952  -1.033   0.3018
## Ind3        -2.83838    3.81271  -0.744   0.4566
## Ind4         1.29420    3.87884   0.334   0.7386
## Ind5        -3.62569    4.19076  -0.865   0.3870
## MedAge:Ind2  0.12188    0.12890   0.946   0.3444
## MedAge:Ind3  0.01720    0.08685   0.198   0.8430
## MedAge:Ind4 -0.07712    0.09027  -0.854   0.3929
## MedAge:Ind5  0.03578    0.10624   0.337   0.7363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 226.26  on 283  degrees of freedom
## AIC: 246.26
##
## Number of Fisher Scoring iterations: 6
```

Standard errors are low but not as low as in Avg.Income. P-values are still too high.

So Ind would probably not be the best interaction variable to use. Therefore, Ind would have to be included as some other kind of predictor variable.

Let's investigate effects of other variables on Mort.

```
Model.Mort <- glm(Mig ~ Mort, family=binomial)
summary(Model.Mort)
```

```
##
## Call:
## glm(formula = Mig ~ Mort, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8930   0.6039   0.6039   0.6932   1.0302
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.9808     0.6770   1.449    0.147
## Mort2        -0.6242     0.7344  -0.850    0.395
## Mort3         0.3226     0.7184   0.449    0.653
## Mort4         0.6286     0.7303   0.861    0.389
## Mort5         0.4855     0.8145   0.596    0.551
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 306.24  on 288  degrees of freedom
## AIC: 316.24
##
## Number of Fisher Scoring iterations: 4
```

```
Model.Mort1 <- glm(Mig ~ Mort+Post.Sec, family=binomial)
summary(Model.Mort1)
```

```
##
## Call:
## glm(formula = Mig ~ Mort + Post.Sec, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4384   0.2639   0.3856   0.7200   1.8758
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.5703     1.3520  -1.161  0.24547
## Mort2        -0.6252     0.9580  -0.653  0.51401
## Mort3         0.5433     0.9478   0.573  0.56649
## Mort4         1.1594     0.9683   1.197  0.23116
## Mort5         1.3216     1.0546   1.253  0.21015
## Post.Sec2    -0.1291     1.3087  -0.099  0.92141
## Post.Sec3     1.6288     1.1553   1.410  0.15859
## Post.Sec4     3.5885     1.1743   3.056  0.00224 **
## Post.Sec5     3.9473     1.3499   2.924  0.00345 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 248.55  on 284  degrees of freedom
## AIC: 266.55
##
## Number of Fisher Scoring iterations: 5
```

When Post.Sec added, Mort doesn't change SEs much, but changes estimates a lot (except for Mort2)

```
Model.Mort2 <- glm(Mig ~ Mort+Avg.Income, family=binomial)
summary(Model.Mort2)
```

```
##
## Call:
## glm(formula = Mig ~ Mort + Avg.Income, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3653   0.4914   0.6008   0.6961   1.5863
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.638e+00  1.414e+00   3.281  0.00103 **
## Mort2       -1.095e+00  8.420e-01  -1.300  0.19349
## Mort3       -5.220e-01  8.526e-01  -0.612  0.54035
## Mort4       -4.487e-01  8.895e-01  -0.504  0.61397
## Mort5       -7.473e-01  9.804e-01  -0.762  0.44596
## Avg.Income  -6.468e-05  2.097e-05  -3.085  0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 295.80  on 287  degrees of freedom
## AIC: 307.8
##
## Number of Fisher Scoring iterations: 4
```

After adding Avg.Income, SE changes minimally, Estimates greatly change.

```
Model.Mort3 <- glm(Mig ~ Mort+Un, family=binomial)
summary(Model.Mort3)
```

```
##
## Call:
## glm(formula = Mig ~ Mort + Un, family = binomial)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -2.2355   0.4600   0.5451   0.6982   1.6479
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.83136    0.91044   2.012 0.044272 *
## Mort2       -0.76943    0.85049  -0.905 0.365630
## Mort3        0.07629    0.83475   0.091 0.927177
## Mort4        0.58151    0.84701   0.687 0.492372
## Mort5        0.69915    0.94557   0.739 0.459667
## Un2         -0.33768    0.59169  -0.571 0.568201
## Un3         -1.13422    0.64763  -1.751 0.079886 .
## Un4         -1.24320    0.75163  -1.654 0.098128 .
## Un5         -2.89180    0.78772  -3.671 0.000241 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 281.00  on 284  degrees of freedom
## AIC: 299
##
## Number of Fisher Scoring iterations: 4
```

SE and Est minimal change

```
Model.Mort4 <- glm(Mig ~ Mort+MedAge, family=binomial)
summary(Model.Mort4)
```

```
##
## Call:
## glm(formula = Mig ~ Mort + MedAge, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4958   0.2237   0.4674   0.6491   2.9596
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.56076    1.62196  -4.662 3.14e-06 ***
## Mort2       -1.52305    0.91993  -1.656  0.09780 .
## Mort3       -1.81532    0.94525  -1.920  0.05480 .
## Mort4       -2.67437    1.03566  -2.582  0.00982 **
## Mort5       -3.60972    1.17361  -3.076  0.00210 **
## MedAge       0.25293    0.04128   6.127 8.98e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 252.13  on 287  degrees of freedom
```

```
## AIC: 264.13
##
## Number of Fisher Scoring iterations: 5
```

Estimates greatly change, SEs don't.

```
Model.Mort5 <- glm(Mig ~ Mort+Ind, family=binomial)
summary(Model.Mort5)
```

```
##
## Call:
## glm(formula = Mig ~ Mort + Ind, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5018   0.2991   0.4058   0.5626   1.6072
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.8882     0.8458   3.415 0.000639 ***
## Mort2        -1.1250     0.8601  -1.308 0.190860
## Mort3        -0.4326     0.8332  -0.519 0.603627
## Mort4        -0.1754     0.8479  -0.207 0.836080
## Mort5         0.1966     0.9265   0.212 0.831985
## Ind2         -0.4817     0.5778  -0.834 0.404433
## Ind3         -2.2997     0.4264  -5.393 6.92e-08 ***
## Ind4         -2.3683     0.5113  -4.632 3.61e-06 ***
## Ind5         -3.4258     0.6138  -5.581 2.39e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 240.24  on 284  degrees of freedom
## AIC: 258.24
##
## Number of Fisher Scoring iterations: 5
```

Estimates greatly change, SEs don't.

Un independent of Mort (as expected, since unemployment doesn't necessarily cause mortality rates.) Let's look at Post Sec, Avg Income, Ind, and MedAge. Med Age could be mediator for mort (mortality is part of the median age calculation, and . Avg Income possible conf? (due to poorer people have less access to health services) Post.Sec doesn't look independent nor collinear to Mort, but I struggle to find a plausible relationship between them. Ind could be a possible mediator. Of course, a high rate of Indigenous people people inside a community won't be a direct push factor so I doubt it's a confounder. However, it is a possible mediator, given that Indigenous people have a lower life expectancy than the general Canadian population.

Let's investigate effects of other variables on Post Sec

```r
Model.Post <- glm(Mig ~ Post.Sec, family=binomial)
summary(Model.Post)
```

```
##
## Call:
## glm(formula = Mig ~ Post.Sec, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2404   0.4118   0.4118   0.8788   1.8930
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.6094     1.0954  -1.469 0.141776
## Post.Sec2     0.6931     1.2450   0.557 0.577699
## Post.Sec3     2.3618     1.1117   2.124 0.033631 *
## Post.Sec4     4.0342     1.1440   3.526 0.000421 ***
## Post.Sec5     3.9120     1.3229   2.957 0.003104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 265.46  on 288  degrees of freedom
## AIC: 275.46
##
## Number of Fisher Scoring iterations: 5
```

```r
Model.Post1 <- glm(Mig ~ Post.Sec+Mort, family=binomial)
summary(Model.Post1)
```

```
##
## Call:
## glm(formula = Mig ~ Post.Sec + Mort, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4384   0.2639   0.3856   0.7200   1.8758
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.5703     1.3520  -1.161  0.24547
## Post.Sec2    -0.1291     1.3087  -0.099  0.92141
## Post.Sec3     1.6288     1.1553   1.410  0.15859
## Post.Sec4     3.5885     1.1743   3.056  0.00224 **
## Post.Sec5     3.9473     1.3499   2.924  0.00345 **
## Mort2        -0.6252     0.9580  -0.653  0.51401
## Mort3         0.5433     0.9478   0.573  0.56649
## Mort4         1.1594     0.9683   1.197  0.23116
## Mort5         1.3216     1.0546   1.253  0.21015
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 248.55  on 284  degrees of freedom
## AIC: 266.55
##
## Number of Fisher Scoring iterations: 5
```

Minimal change for both Est and SEs, but big change for Post.Sec.2

```r
Model.Post2 <- glm(Mig ~ Post.Sec+Avg.Income, family=binomial)
summary(Model.Post2)
```

```
##
## Call:
## glm(formula = Mig ~ Post.Sec + Avg.Income, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9116   0.1551   0.3127   0.5650   2.8024
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.248e+00  1.436e+00   2.262  0.02368 *
## Post.Sec2    1.648e+00  1.371e+00   1.203  0.22913
## Post.Sec3    3.915e+00  1.274e+00   3.072  0.00213 **
## Post.Sec4    6.100e+00  1.345e+00   4.535 5.75e-06 ***
## Post.Sec5    7.588e+00  1.703e+00   4.455 8.39e-06 ***
## Avg.Income  -1.514e-04  2.587e-05  -5.853 4.82e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 217.68  on 287  degrees of freedom
## AIC: 229.68
##
## Number of Fisher Scoring iterations: 5
```

Est greatly change, not so much for Standard Errors.

```r
Model.Post3 <- glm(Mig ~ Post.Sec+Un, family=binomial)
summary(Model.Post3)
```

```
##
## Call:
## glm(formula = Mig ~ Post.Sec + Un, family = binomial)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -2.2768   0.3875   0.3945   0.8006   2.0341
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.75668    1.32002  -0.573  0.56649
## Post.Sec2    0.31526    1.29811   0.243  0.80811
## Post.Sec3    1.76725    1.18751   1.488  0.13670
## Post.Sec4    3.30811    1.23472   2.679  0.00738 **
## Post.Sec5    3.10927    1.41761   2.193  0.02828 *
## Un2         -0.03731    0.60855  -0.061  0.95111
## Un3         -0.48548    0.66106  -0.734  0.46270
## Un4         -0.35617    0.77644  -0.459  0.64643
## Un5         -1.17713    0.84262  -1.397  0.16242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 261.49  on 284  degrees of freedom
## AIC: 279.49
##
## Number of Fisher Scoring iterations: 5
```

Est change varies, but est mostly unchanged. SEs are also unchanged. assume independent.

```
Model.Post4 <- glm(Mig ~ Post.Sec+MedAge, family=binomial)
Model.Post4
```

```
##
## Call:  glm(formula = Mig ~ Post.Sec + MedAge, family = binomial)
##
## Coefficients:
## (Intercept)    Post.Sec2    Post.Sec3    Post.Sec4    Post.Sec5       MedAge
##     -6.2849      -2.7737      -0.9652       0.7897       1.1707       0.1825
##
## Degrees of Freedom: 292 Total (i.e. Null);  287 Residual
## Null Deviance:       317.5
## Residual Deviance: 225.5     AIC: 237.5
```

Est greatly changes, not so much for SEs.

```
Model.Post5 <- glm(Mig ~ Post.Sec+Ind, family=binomial)
summary(Model.Post5)
```

```
##
## Call:
## glm(formula = Mig ~ Post.Sec + Ind, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5006   0.2996   0.3265   0.5543   1.8930
```

```
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.1993     1.2969   0.925  0.35509
## Post.Sec2    -0.4134     1.3804  -0.299  0.76456
## Post.Sec3     0.5962     1.2629   0.472  0.63685
## Post.Sec4     1.8822     1.2995   1.448  0.14751
## Post.Sec5     1.6622     1.4876   1.117  0.26382
## Ind2         -0.1767     0.5742  -0.308  0.75825
## Ind3         -1.8709     0.4270  -4.381 1.18e-05 ***
## Ind4         -1.6951     0.5248  -3.230  0.00124 **
## Ind5         -2.8087     0.6942  -4.046 5.21e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 230.05  on 284  degrees of freedom
## AIC: 248.05
##
## Number of Fisher Scoring iterations: 5
```

Est greatly changes, not so much for SEs.

Let's look at Avg.Income, MedAge, and Ind. Avg.Income could be a confounding factor for Post.Sec, as income could affect one's ability to access education. For MedAge, it could be a confounder but not a strong one since while lower median age means more college-aged people, that doesn't mean higher post secondary graduation rate. For Avg.Income, it could be a confounder since it could both be a picture of economic conditions for a geographical area (leading it to be a push factor) and could demonstrate that low-income people are less likely to be able to access education. For Ind, it appears to be either a mediator or confounder for Post.Sec.

Let's investigate effects of other variables on Avg Inc

```r
Model.Inc <- glm(Mig ~ Avg.Income, family=binomial)
summary(Model.Inc)
```

```
##
## Call:
## glm(formula = Mig ~ Avg.Income, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3922   0.5074   0.6142   0.7121   1.6781
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.352e+00  8.173e-01   5.324 1.01e-07 ***
## Avg.Income  -7.250e-05  1.822e-05  -3.979 6.93e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 299.47  on 291  degrees of freedom
## AIC: 303.47
##
## Number of Fisher Scoring iterations: 4
```

```
Model.Inc1 <- glm(Mig ~ Avg.Income+Mort, family=binomial)
summary(Model.Inc1)
```

```
##
## Call:
## glm(formula = Mig ~ Avg.Income + Mort, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3653   0.4914   0.6008   0.6961   1.5863
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.638e+00  1.414e+00   3.281  0.00103 **
## Avg.Income  -6.468e-05  2.097e-05  -3.085  0.00204 **
## Mort2       -1.095e+00  8.420e-01  -1.300  0.19349
## Mort3       -5.220e-01  8.526e-01  -0.612  0.54035
## Mort4       -4.487e-01  8.895e-01  -0.504  0.61397
## Mort5       -7.473e-01  9.804e-01  -0.762  0.44596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 295.80  on 287  degrees of freedom
## AIC: 307.8
##
## Number of Fisher Scoring iterations: 4
```

This confirms that Average income and mort are ind.

```
Model.Inc2 <- glm(Mig ~ Avg.Income+Post.Sec, family=binomial)
summary(Model.Inc2)
```

```
##
## Call:
## glm(formula = Mig ~ Avg.Income + Post.Sec, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9116   0.1551   0.3127   0.5650   2.8024
##
```

```
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.248e+00  1.436e+00   2.262  0.02368 *
## Avg.Income  -1.514e-04  2.587e-05  -5.853 4.82e-09 ***
## Post.Sec2    1.648e+00  1.371e+00   1.203  0.22913
## Post.Sec3    3.915e+00  1.274e+00   3.072  0.00213 **
## Post.Sec4    6.100e+00  1.345e+00   4.535 5.75e-06 ***
## Post.Sec5    7.588e+00  1.703e+00   4.455 8.39e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 217.68  on 287  degrees of freedom
## AIC: 229.68
##
## Number of Fisher Scoring iterations: 5
```

Est changes a lot

```
Model.Inc3 <- glm(Mig ~ Avg.Income+Un, family=binomial)
summary(Model.Inc3)
```

```
##
## Call:
## glm(formula = Mig ~ Avg.Income + Un, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2883   0.3385   0.5156   0.6626   2.0566
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.539e+00  1.114e+00   5.871 4.34e-09 ***
## Avg.Income  -1.054e-04  2.069e-05  -5.095 3.49e-07 ***
## Un2         -2.069e-01  5.955e-01  -0.347   0.7283
## Un3         -1.200e+00  6.547e-01  -1.833   0.0668 .
## Un4         -1.549e+00  7.619e-01  -2.033   0.0421 *
## Un5         -3.544e+00  8.132e-01  -4.359 1.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 261.05  on 287  degrees of freedom
## AIC: 273.05
##
## Number of Fisher Scoring iterations: 4
```

Est changes not as much as before. Assume independence.

```r
Model.Inc4 <- glm(Mig ~ Avg.Income+MedAge, family=binomial)
summary(Model.Inc4)
```

```
##
## Call:
## glm(formula = Mig ~ Avg.Income + MedAge, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5384   0.2901   0.4625   0.7056   1.9901
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.264e+00  1.714e+00  -2.488   0.0128 *
## Avg.Income  -2.376e-05  1.958e-05  -1.214   0.2249
## MedAge       1.519e-01  2.806e-02   5.414 6.17e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 263.78  on 290  degrees of freedom
## AIC: 269.78
##
## Number of Fisher Scoring iterations: 5
```

Est changes greatly.

```r
Model.Inc5 <- glm(Mig ~ Avg.Income+Ind, family=binomial)
summary(Model.Inc5)
```

```
##
## Call:
## glm(formula = Mig ~ Avg.Income + Ind, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5833   0.2933   0.3796   0.4968   2.1456
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.538e+00  9.851e-01   5.621 1.89e-08 ***
## Avg.Income  -7.296e-05  2.118e-05  -3.446  0.00057 ***
## Ind2        -2.323e-01  5.690e-01  -0.408  0.68302
## Ind3        -1.969e+00  4.219e-01  -4.667 3.06e-06 ***
## Ind4        -2.220e+00  5.161e-01  -4.302 1.70e-05 ***
## Ind5        -3.761e+00  6.240e-01  -6.028 1.66e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 232.66  on 287  degrees of freedom
## AIC: 244.66
##
## Number of Fisher Scoring iterations: 5
```

Minimal change in est and SEs. Independent.

Let's focus on Post.Sec and MedAge. MedAge could be a possible confounder, though maybe not as strong. Older people would have accumulated more wealth in general than younger people, but that's not always the case. Post.Sec could be a confounder, since it's both a plausible push factor and also illustrates that post-secondary graduation tend to lead to better-paying jobs.

Let's investigate effects of other variables on Un

```
Model.Un <- glm(Mig ~ Un, family=binomial)
summary(Model.Un)
```

```
##
## Call:
## glm(formula = Mig ~ Un, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9543   0.5663   0.6113   0.6113   1.5645
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.7492     0.5417   3.229 0.001243 **
## Un2          -0.1668     0.5776  -0.289 0.772774
## Un3          -0.8737     0.6228  -1.403 0.160667
## Un4          -0.9871     0.7092  -1.392 0.164001
## Un5          -2.6247     0.7595  -3.456 0.000549 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 293.29  on 288  degrees of freedom
## AIC: 303.29
##
## Number of Fisher Scoring iterations: 4
```

```
Model.Un1 <- glm(Mig ~ Un+Mort, family=binomial)
summary(Model.Un1)
```

```
##
## Call:
## glm(formula = Mig ~ Un + Mort, family = binomial)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2355   0.4600   0.5451   0.6982   1.6479
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.83136    0.91044   2.012 0.044272 *
## Un2         -0.33768    0.59169  -0.571 0.568201
## Un3         -1.13422    0.64763  -1.751 0.079886 .
## Un4         -1.24320    0.75163  -1.654 0.098128 .
## Un5         -2.89180    0.78772  -3.671 0.000241 ***
## Mort2       -0.76943    0.85049  -0.905 0.365630
## Mort3        0.07629    0.83475   0.091 0.927177
## Mort4        0.58151    0.84701   0.687 0.492372
## Mort5        0.69915    0.94557   0.739 0.459667
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 281.00  on 284  degrees of freedom
## AIC: 299
##
## Number of Fisher Scoring iterations: 4
```

Again all except Un2 has minimal est change

```
Model.Un2 <- glm(Mig ~ Un+Post.Sec, family=binomial)
summary(Model.Un2)
```

```
##
## Call:
## glm(formula = Mig ~ Un + Post.Sec, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2768   0.3875   0.3945   0.8006   2.0341
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.75668    1.32002  -0.573  0.56649
## Un2         -0.03731    0.60855  -0.061  0.95111
## Un3         -0.48548    0.66106  -0.734  0.46270
## Un4         -0.35617    0.77644  -0.459  0.64643
## Un5         -1.17713    0.84262  -1.397  0.16242
## Post.Sec2    0.31526    1.29811   0.243  0.80811
## Post.Sec3    1.76725    1.18751   1.488  0.13670
## Post.Sec4    3.30811    1.23472   2.679  0.00738 **
## Post.Sec5    3.10927    1.41761   2.193  0.02828 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 261.49  on 284  degrees of freedom
## AIC: 279.49
##
## Number of Fisher Scoring iterations: 5
```

Massive change in Est not so much in SE

```
Model.Un3 <- glm(Mig ~ Un+Avg.Income, family=binomial)
summary(Model.Un3)
```

```
##
## Call:
## glm(formula = Mig ~ Un + Avg.Income, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2883   0.3385   0.5156   0.6626   2.0566
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.539e+00  1.114e+00   5.871 4.34e-09 ***
## Un2         -2.069e-01  5.955e-01  -0.347   0.7283
## Un3         -1.200e+00  6.547e-01  -1.833   0.0668 .
## Un4         -1.549e+00  7.619e-01  -2.033   0.0421 *
## Un5         -3.544e+00  8.132e-01  -4.359 1.31e-05 ***
## Avg.Income  -1.054e-04  2.069e-05  -5.095 3.49e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 261.05  on 287  degrees of freedom
## AIC: 273.05
##
## Number of Fisher Scoring iterations: 4
```

No Significant change in Est

```
Model.Un4 <- glm(Mig ~ Un+MedAge, family=binomial)
summary(Model.Un4)
```

```
##
## Call:
## glm(formula = Mig ~ Un + MedAge, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.4225   0.2422   0.4544   0.6304   3.1696
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.92048    1.38348  -4.279 1.87e-05 ***
## Un2         -0.33042    0.59549  -0.555 0.578983
## Un3         -1.16304    0.66141  -1.758 0.078677 .
## Un4         -0.91511    0.88439  -1.035 0.300788
## Un5         -3.34190    0.87720  -3.810 0.000139 ***
## MedAge       0.18144    0.03051   5.948 2.72e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 242.72  on 287  degrees of freedom
## AIC: 254.72
##
## Number of Fisher Scoring iterations: 5
```

All except Un2 no significant change

```
Model.Un5 <- glm(Mig ~ Un+Ind, family=binomial)
summary(Model.Un5)
```

```
##
## Call:
## glm(formula = Mig ~ Un + Ind, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3414   0.3652   0.3979   0.4752   2.1725
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.2015     0.5969   3.688 0.000226 ***
## Un2           0.2947     0.6290   0.468 0.639445
## Un3           0.4728     0.7175   0.659 0.509889
## Un4           0.4698     0.8410   0.559 0.576375
## Un5          -1.0782     0.9008  -1.197 0.231347
## Ind2         -0.3721     0.5702  -0.653 0.514021
## Ind3         -2.1923     0.4249  -5.159 2.48e-07 ***
## Ind4         -2.3335     0.5342  -4.368 1.25e-05 ***
## Ind5         -3.3839     0.6858  -4.934 8.05e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 241.14  on 284  degrees of freedom
## AIC: 259.14
```

```
##
## Number of Fisher Scoring iterations: 5
```

Massive change in Est not so much in SE

It seems that Un is largely independent of other variables except for Post.Sec and Ind. I can see Post.Sec being a confounder for Un since many job opportunities require post-secondary credentials, and post-secondary opportunities is in itself a push factor. Ind can't be a push factor in itself, so it might be a mediator for Un.

Let's investigate effects of other variables on MedAge

```
Model.Med <- glm(Mig ~ MedAge, family=binomial)
summary(Model.Med)
```

```
##
## Call:
## glm(formula = Mig ~ MedAge, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4747   0.2854   0.4769   0.7143   2.0930
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.92770    1.12542  -5.267 1.39e-07 ***
## MedAge       0.16620    0.02658   6.252 4.05e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 265.29  on 291  degrees of freedom
## AIC: 269.29
##
## Number of Fisher Scoring iterations: 5
```

```
Model.Med1 <- glm(Mig ~ MedAge+Mort, family=binomial)
summary(Model.Med1)
```

```
##
## Call:
## glm(formula = Mig ~ MedAge + Mort, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4958   0.2237   0.4674   0.6491   2.9596
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.56076    1.62196  -4.662 3.14e-06 ***
```

```
## MedAge       0.25293    0.04128   6.127 8.98e-10 ***
## Mort2       -1.52305    0.91993  -1.656  0.09780 .
## Mort3       -1.81532    0.94525  -1.920  0.05480 .
## Mort4       -2.67437    1.03566  -2.582  0.00982 **
## Mort5       -3.60972    1.17361  -3.076  0.00210 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 252.13  on 287  degrees of freedom
## AIC: 264.13
##
## Number of Fisher Scoring iterations: 5
```

Sig in Est change not so much SE

```
Model.Med2 <- glm(Mig ~ MedAge+Post.Sec, family=binomial)
summary(Model.Med2)
```

```
##
## Call:
## glm(formula = Mig ~ MedAge + Post.Sec, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4942   0.1522   0.3639   0.5762   2.1911
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.28487    1.39115  -4.518 6.25e-06 ***
## MedAge       0.18252    0.03321   5.497 3.87e-08 ***
## Post.Sec2   -2.77372    1.48594  -1.867    0.062 .
## Post.Sec3   -0.96517    1.26055  -0.766    0.444
## Post.Sec4    0.78969    1.26977   0.622    0.534
## Post.Sec5    1.17067    1.41085   0.830    0.407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 225.54  on 287  degrees of freedom
## AIC: 237.54
##
## Number of Fisher Scoring iterations: 5
```

Not as big of a change with est

```
Model.Med3 <- glm(Mig ~ MedAge+Avg.Income, family=binomial)
summary(Model.Med3)
```

```
##
## Call:
## glm(formula = Mig ~ MedAge + Avg.Income, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5384   0.2901   0.4625   0.7056   1.9901
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.264e+00  1.714e+00  -2.488   0.0128 *
## MedAge       1.519e-01  2.806e-02   5.414 6.17e-08 ***
## Avg.Income  -2.376e-05  1.958e-05  -1.214   0.2249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 263.78  on 290  degrees of freedom
## AIC: 269.78
##
## Number of Fisher Scoring iterations: 5
```

Little changes

```
Model.Med4 <- glm(Mig ~ MedAge+Un, family=binomial)
summary(Model.Med4)
```

```
##
## Call:
## glm(formula = Mig ~ MedAge + Un, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4225   0.2422   0.4544   0.6304   3.1696
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.92048    1.38348  -4.279 1.87e-05 ***
## MedAge       0.18144    0.03051   5.948 2.72e-09 ***
## Un2         -0.33042    0.59549  -0.555 0.578983
## Un3         -1.16304    0.66141  -1.758 0.078677 .
## Un4         -0.91511    0.88439  -1.035 0.300788
## Un5         -3.34190    0.87720  -3.810 0.000139 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 242.72  on 287  degrees of freedom
## AIC: 254.72
```

```
##
## Number of Fisher Scoring iterations: 5
```

Little changes

```
Model.Med5 <- glm(Mig ~ MedAge+Ind, family=binomial)
summary(Model.Med5)
```

```
##
## Call:
## glm(formula = Mig ~ MedAge + Ind, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6657   0.2327   0.3718   0.5205   2.2452
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.92200    1.37431  -2.126 0.033490 *
## MedAge       0.11981    0.03087   3.881 0.000104 ***
## Ind2        -0.43051    0.57189  -0.753 0.451587
## Ind3        -2.08636    0.42291  -4.933 8.08e-07 ***
## Ind4        -1.89115    0.52292  -3.617 0.000299 ***
## Ind5        -2.31814    0.68101  -3.404 0.000664 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 229.47  on 287  degrees of freedom
## AIC: 241.47
##
## Number of Fisher Scoring iterations: 5
```

Little changes

MedAge and Mort could be conf or mediators of each other. MedAge could've been a mediator for Mort, so it would mean that Mort is a conf for MedAge.

Let's investigate effects of other variables on Ind.

```
Model.Ind <- glm(Mig ~ Ind, family=binomial)
summary(Model.Ind)
```

```
##
## Call:
## glm(formula = Mig ~ Ind, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.2324   0.4157   0.4157   0.4690   1.6942
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.4054     0.3014   7.980 1.46e-15 ***
## Ind2         -0.2537     0.5605  -0.453    0.651
## Ind3         -2.1397     0.4095  -5.226 1.74e-07 ***
## Ind4         -2.2513     0.4956  -4.543 5.56e-06 ***
## Ind5         -3.5686     0.5944  -6.003 1.93e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 246.35  on 288  degrees of freedom
## AIC: 256.35
##
## Number of Fisher Scoring iterations: 5
```

```
Model.Ind1 <- glm(Mig ~ Ind+Mort, family=binomial)
summary(Model.Ind1)
```

```
##
## Call:
## glm(formula = Mig ~ Ind + Mort, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5018   0.2991   0.4058   0.5626   1.6072
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.8882     0.8458   3.415 0.000639 ***
## Ind2         -0.4817     0.5778  -0.834 0.404433
## Ind3         -2.2997     0.4264  -5.393 6.92e-08 ***
## Ind4         -2.3683     0.5113  -4.632 3.61e-06 ***
## Ind5         -3.4258     0.6138  -5.581 2.39e-08 ***
## Mort2        -1.1250     0.8601  -1.308 0.190860
## Mort3        -0.4326     0.8332  -0.519 0.603627
## Mort4        -0.1754     0.8479  -0.207 0.836080
## Mort5         0.1966     0.9265   0.212 0.831985
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 240.24  on 284  degrees of freedom
## AIC: 258.24
##
## Number of Fisher Scoring iterations: 5
```

All except Ind2 sees minimal change in Est.

```r
Model.Ind2 <- glm(Mig ~ Ind+Post.Sec, family=binomial)
summary(Model.Ind2)
```

```
##
## Call:
## glm(formula = Mig ~ Ind + Post.Sec, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5006   0.2996   0.3265   0.5543   1.8930
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.1993     1.2969   0.925  0.35509
## Ind2         -0.1767     0.5742  -0.308  0.75825
## Ind3         -1.8709     0.4270  -4.381 1.18e-05 ***
## Ind4         -1.6951     0.5248  -3.230  0.00124 **
## Ind5         -2.8087     0.6942  -4.046 5.21e-05 ***
## Post.Sec2    -0.4134     1.3804  -0.299  0.76456
## Post.Sec3     0.5962     1.2629   0.472  0.63685
## Post.Sec4     1.8822     1.2995   1.448  0.14751
## Post.Sec5     1.6622     1.4876   1.117  0.26382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 230.05  on 284  degrees of freedom
## AIC: 248.05
##
## Number of Fisher Scoring iterations: 5
```

Minimal change in Est. Independent.

```r
Model.Ind3 <- glm(Mig ~ Ind+Avg.Income, family=binomial)
summary(Model.Ind3)
```

```
##
## Call:
## glm(formula = Mig ~ Ind + Avg.Income, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5833   0.2933   0.3796   0.4968   2.1456
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.538e+00  9.851e-01   5.621 1.89e-08 ***
## Ind2        -2.323e-01  5.690e-01  -0.408  0.68302
## Ind3        -1.969e+00  4.219e-01  -4.667 3.06e-06 ***
## Ind4        -2.220e+00  5.161e-01  -4.302 1.70e-05 ***
```

```
## Ind5        -3.761e+00  6.240e-01  -6.028 1.66e-09 ***
## Avg.Income  -7.296e-05  2.118e-05  -3.446  0.00057 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 232.66  on 287  degrees of freedom
## AIC: 244.66
##
## Number of Fisher Scoring iterations: 5
```

Minimal change in Est. Independent.

```
Model.Ind4 <- glm(Mig ~ Ind+Un, family=binomial)
summary(Model.Ind4)
```

```
##
## Call:
## glm(formula = Mig ~ Ind + Un, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3414   0.3652   0.3979   0.4752   2.1725
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.2015     0.5969   3.688 0.000226 ***
## Ind2         -0.3721     0.5702  -0.653 0.514021
## Ind3         -2.1923     0.4249  -5.159 2.48e-07 ***
## Ind4         -2.3335     0.5342  -4.368 1.25e-05 ***
## Ind5         -3.3839     0.6858  -4.934 8.05e-07 ***
## Un2           0.2947     0.6290   0.468 0.639445
## Un3           0.4728     0.7175   0.659 0.509889
## Un4           0.4698     0.8410   0.559 0.576375
## Un5          -1.0782     0.9008  -1.197 0.231347
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 241.14  on 284  degrees of freedom
## AIC: 259.14
##
## Number of Fisher Scoring iterations: 5
```

All except Ind2 see minimal change in Est.

```
Model.Ind5 <- glm(Mig ~ Ind+MedAge, family=binomial)
summary(Model.Ind5)
```

```
##
## Call:
## glm(formula = Mig ~ Ind + MedAge, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6657   0.2327   0.3718   0.5205   2.2452
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.92200    1.37431  -2.126 0.033490 *
## Ind2        -0.43051    0.57189  -0.753 0.451587
## Ind3        -2.08636    0.42291  -4.933 8.08e-07 ***
## Ind4        -1.89115    0.52292  -3.617 0.000299 ***
## Ind5        -2.31814    0.68101  -3.404 0.000664 ***
## MedAge       0.11981    0.03087   3.881 0.000104 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 229.47  on 287  degrees of freedom
## AIC: 241.47
##
## Number of Fisher Scoring iterations: 5
```

All except Ind2 see minimal change in Est.

It seems that Ind is also not a good variable of interest, which is no surprise since, as mentioned before, Ind can't be a direct push factor.

Mort as variable of interest

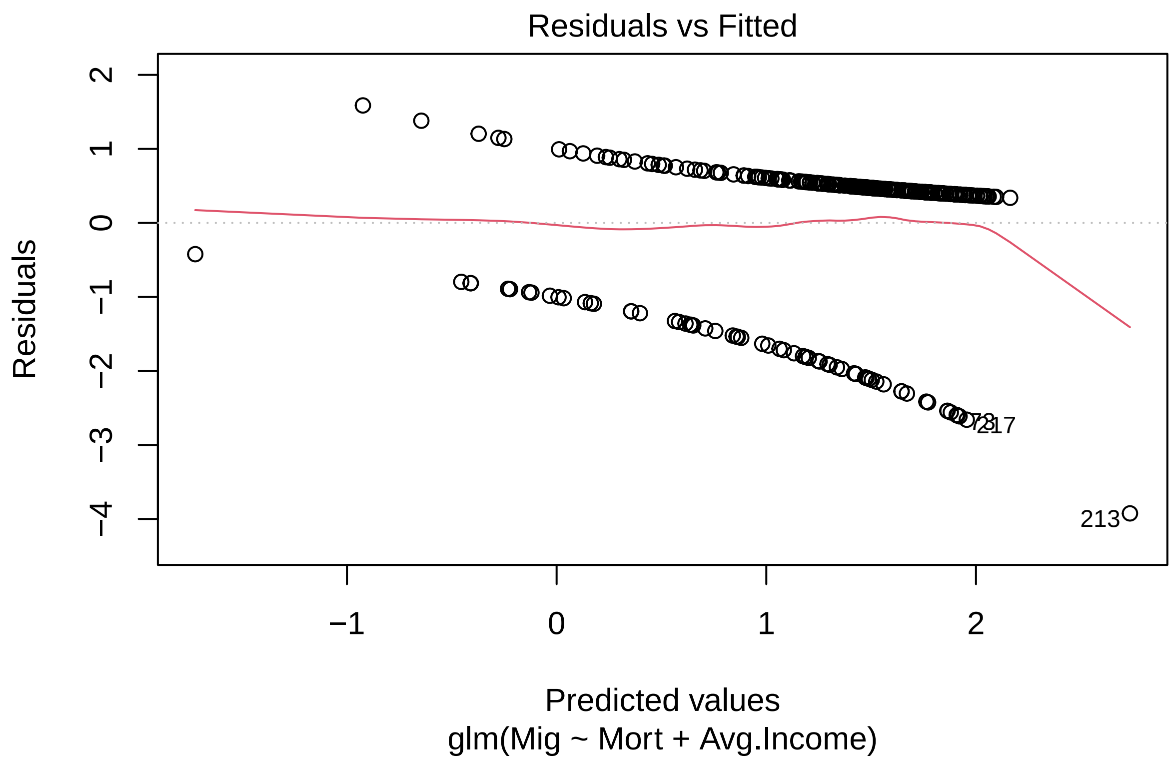
Only seems acceptable to include in the model.

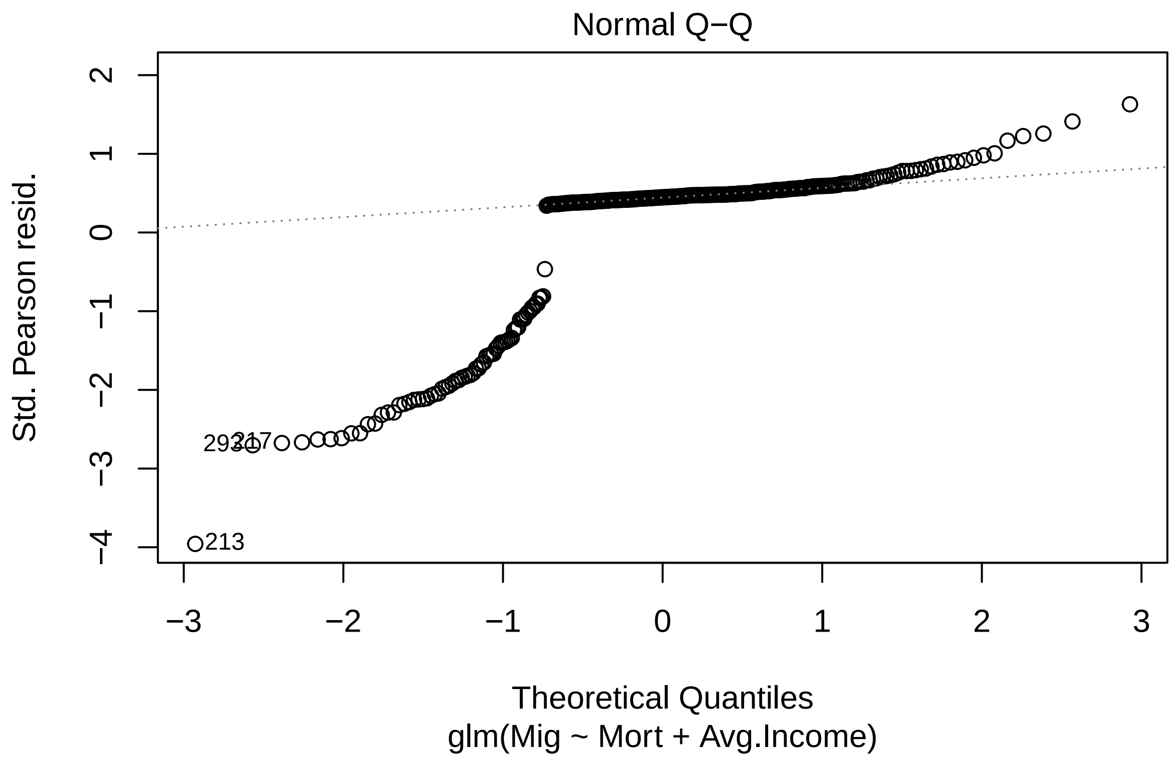
```r
Model1 <- glm(Mig ~ Mort+Avg.Income, family=binomial)
summary(Model1)
```

```
##
## Call:
## glm(formula = Mig ~ Mort + Avg.Income, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3653   0.4914   0.6008   0.6961   1.5863
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.638e+00  1.414e+00   3.281  0.00103 **
## Mort2       -1.095e+00  8.420e-01  -1.300  0.19349
## Mort3       -5.220e-01  8.526e-01  -0.612  0.54035
## Mort4       -4.487e-01  8.895e-01  -0.504  0.61397
```

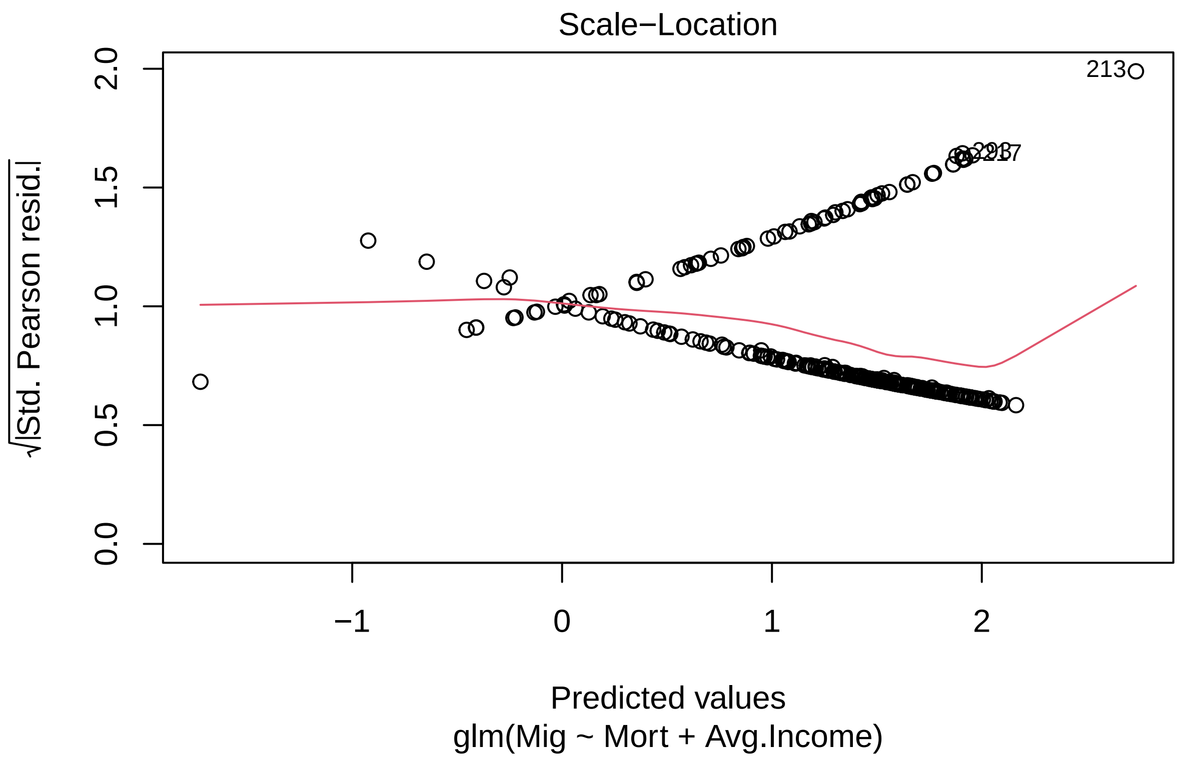
```
## Mort5       -7.473e-01  9.804e-01  -0.762  0.44596
## Avg.Income  -6.468e-05  2.097e-05  -3.085  0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 295.80  on 287  degrees of freedom
## AIC: 307.8
##
## Number of Fisher Scoring iterations: 4
```

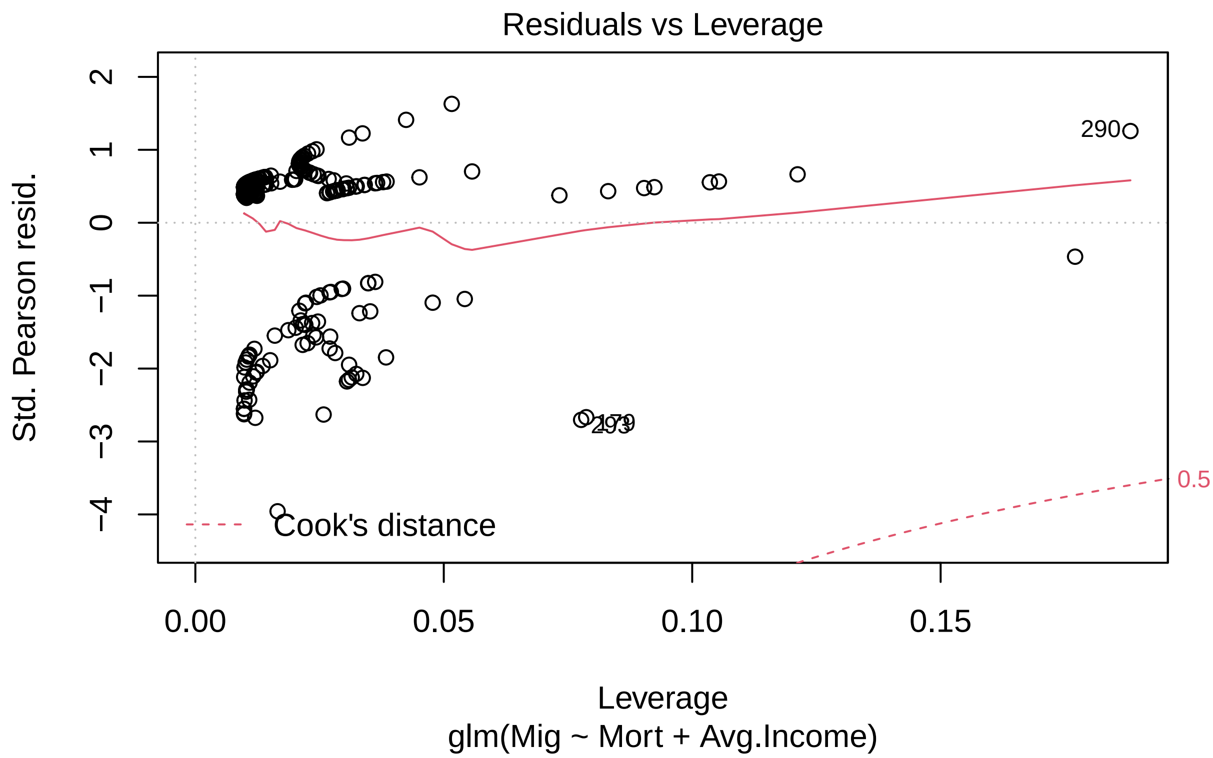
```
plot(Model1)
```





Normal Q−Q
Std. Pearson resid.
Theoretical Quantiles
glm(Mig ~ Mort + Avg.Income)
2
1
0
−1
−2
−3
−4
−3
−2
−1
0
1
2
3
293
217
213


Scale−Location
$\sqrt{|\text{Std. Pearson resid.}|}$
0.0
0.5
1.0
1.5
2.0
−1
0
1
2
Predicted values
glm(Mig ~ Mort + Avg.Income)
213




```r
anova(Model1, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Mig
##
## Terms added sequentially (first to last)
##
##
##            Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                         292     317.48
## Mort        4   11.247       288     306.24 0.023921 *
## Avg.Income  1   10.432       287     295.80 0.001238 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

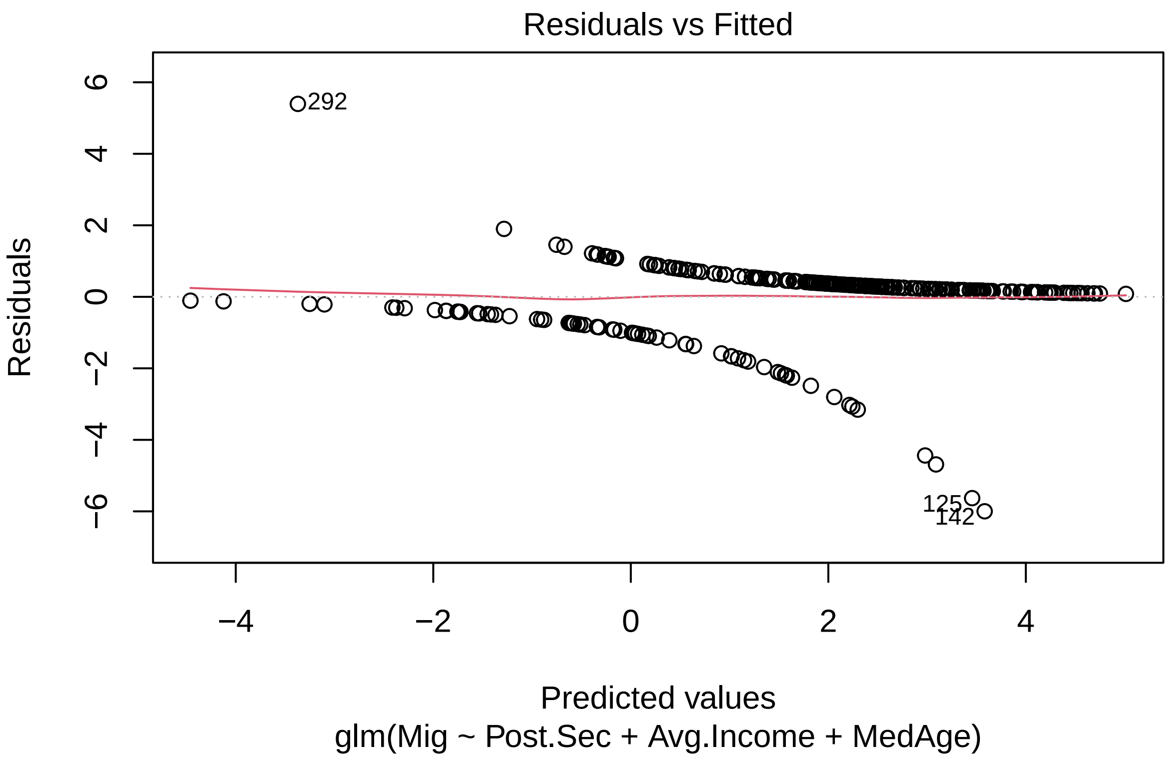
Post Sec as variable of interest

Can only use Avg inc and MedAge

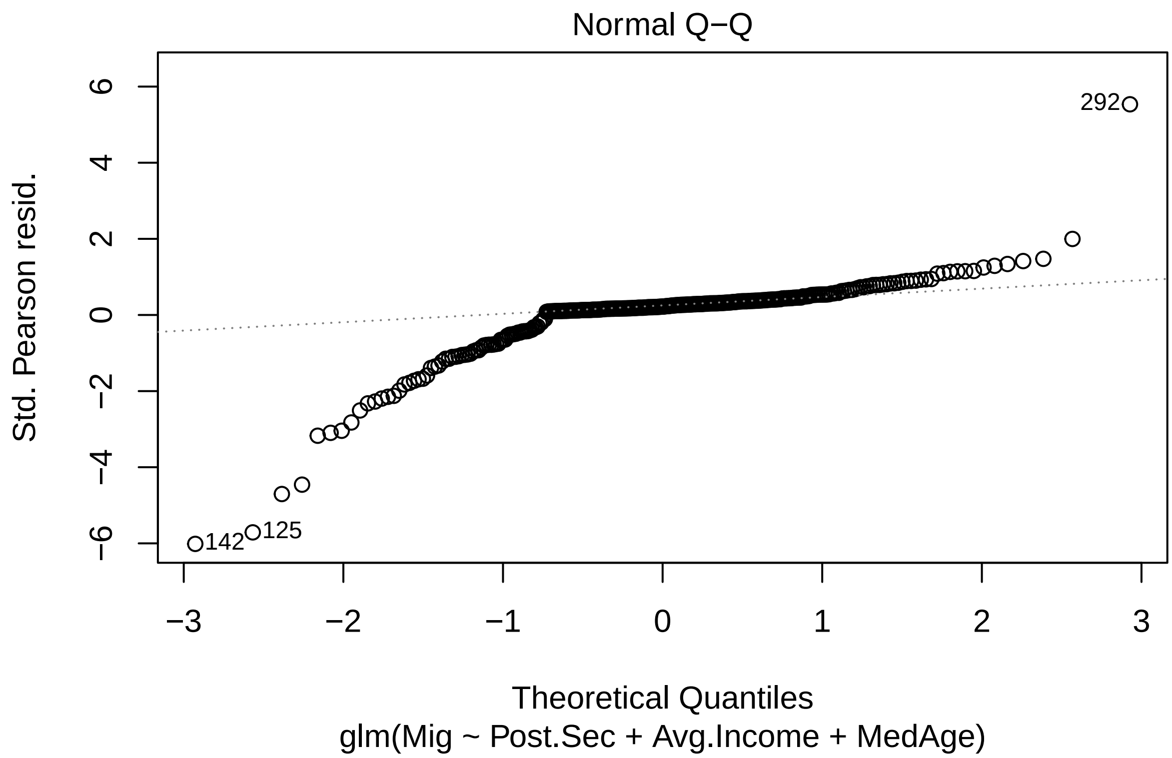
```r
Model2.1 <- glm(Mig ~ Post.Sec+Avg.Income+MedAge, family=binomial)
summary(Model2.1)
```

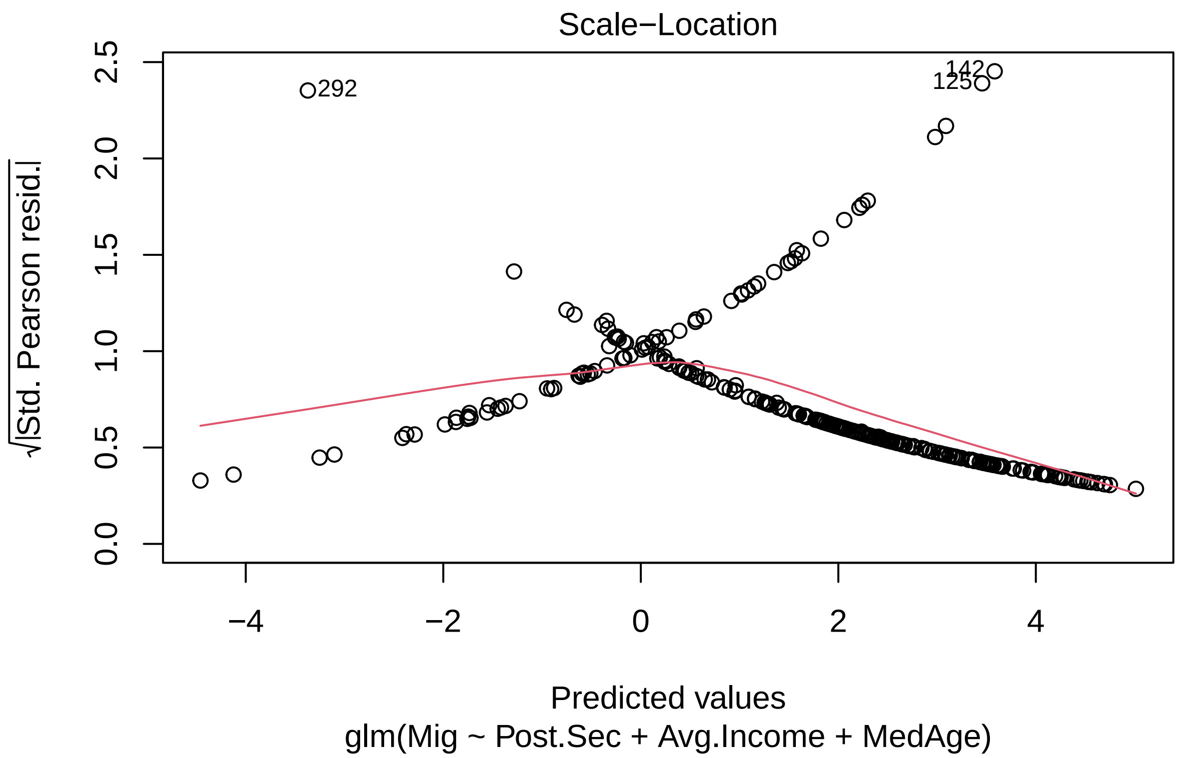
```
##
## Call:
## glm(formula = Mig ~ Post.Sec + Avg.Income + MedAge, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6871   0.1440   0.3065   0.5406   2.6096
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.173e-01  2.046e+00  -0.351 0.725881
## Post.Sec2   -5.303e-01  1.598e+00  -0.332 0.740053
## Post.Sec3    1.593e+00  1.494e+00   1.066 0.286235
## Post.Sec4    3.674e+00  1.558e+00   2.358 0.018382 *
## Post.Sec5    4.895e+00  1.883e+00   2.600 0.009318 **
## Avg.Income  -1.052e-04  3.031e-05  -3.472 0.000516 ***
## MedAge       9.908e-02  3.857e-02   2.569 0.010199 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 210.90  on 286  degrees of freedom
## AIC: 224.9
##
## Number of Fisher Scoring iterations: 5
```

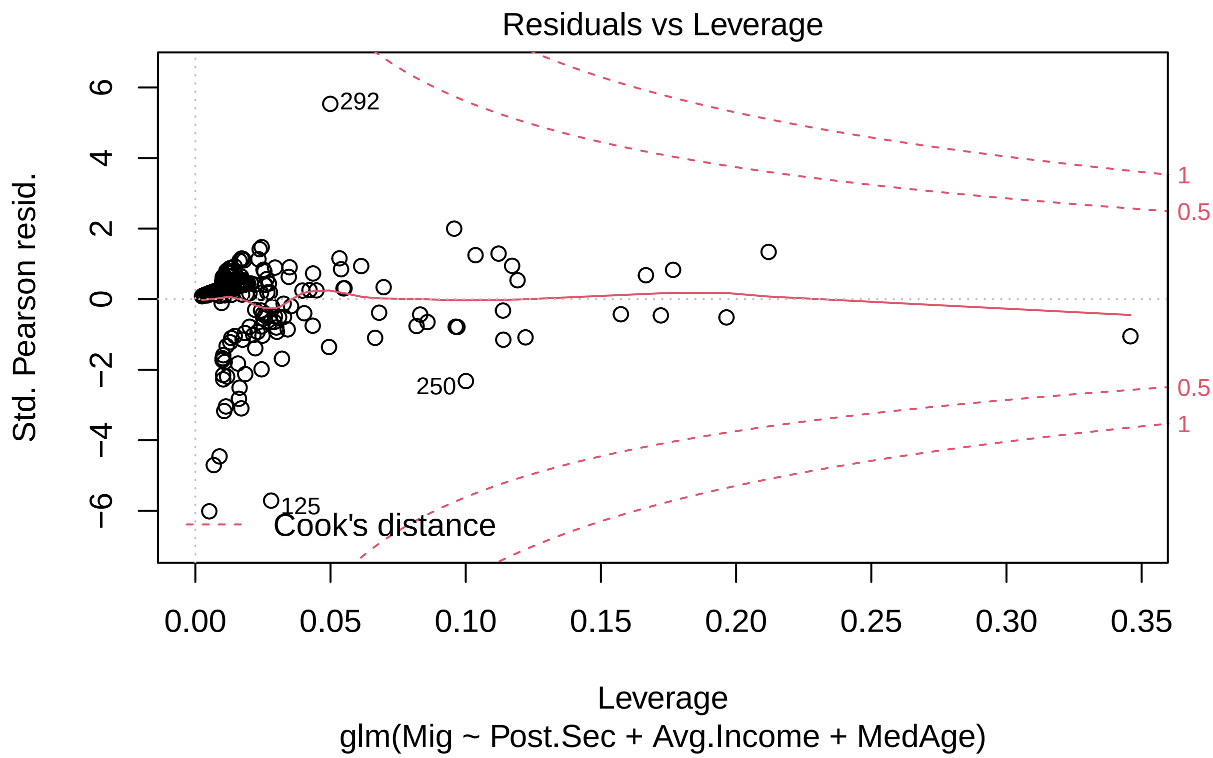
```
plot(Model2.1)
```



Residuals vs Fitted
Residuals
Predicted values
glm(Mig ~ Post.Sec + Avg.Income + MedAge)
292
125
142
6
4
2
0
−2
−4
−6
−4
−2
0
2
4


Normal Q−Q
Std. Pearson resid.
6
4
2
0
−2
−4
−6
292
142
125
−3
−2
−1
0
1
2
3
Theoretical Quantiles
glm(Mig ~ Post.Sec + Avg.Income + MedAge)


Scale−Location
$\sqrt{|\text{Std. Pearson resid.}|}$
0.0
0.5
1.0
1.5
2.0
2.5
292
142
125
−4
−2
0
2
4
Predicted values
glm(Mig ~ Post.Sec + Avg.Income + MedAge)




```
anova(Model2.1, test="LRT")
```

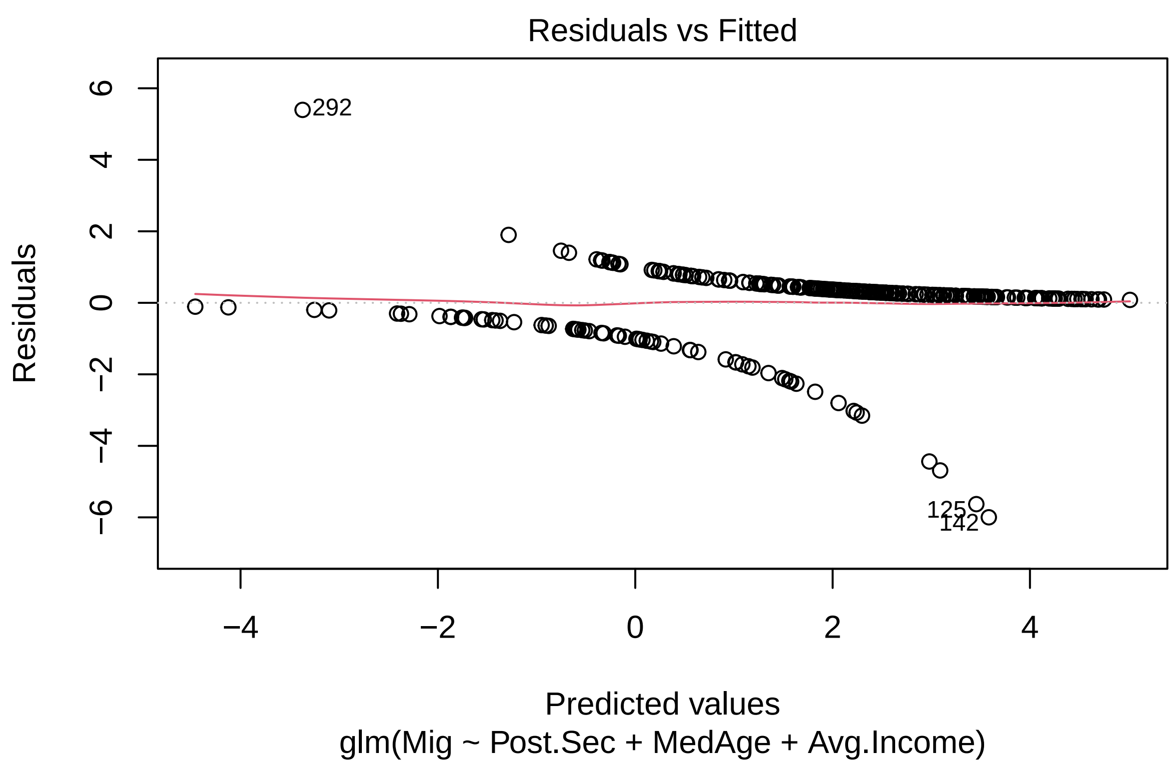
```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Mig
##
## Terms added sequentially (first to last)
##
##
##            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         292     317.48
## Post.Sec    4   52.026       288     265.46 1.362e-10 ***
## Avg.Income  1   47.778       287     217.68 4.773e-12 ***
## MedAge      1    6.777       286     210.90  0.009235 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

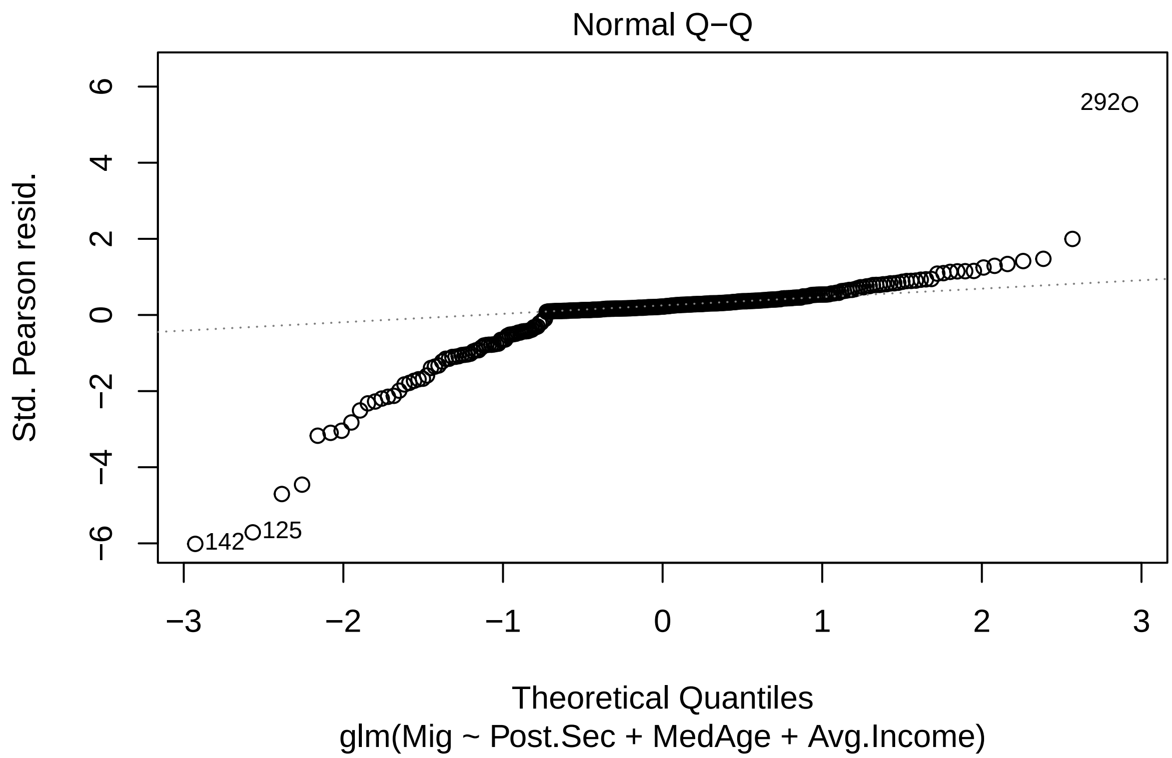
```
Model2.2 <- glm(Mig ~ Post.Sec+MedAge+Avg.Income, family=binomial)
summary(Model2.2)
```

```
##
## Call:
## glm(formula = Mig ~ Post.Sec + MedAge + Avg.Income, family = binomial)
```

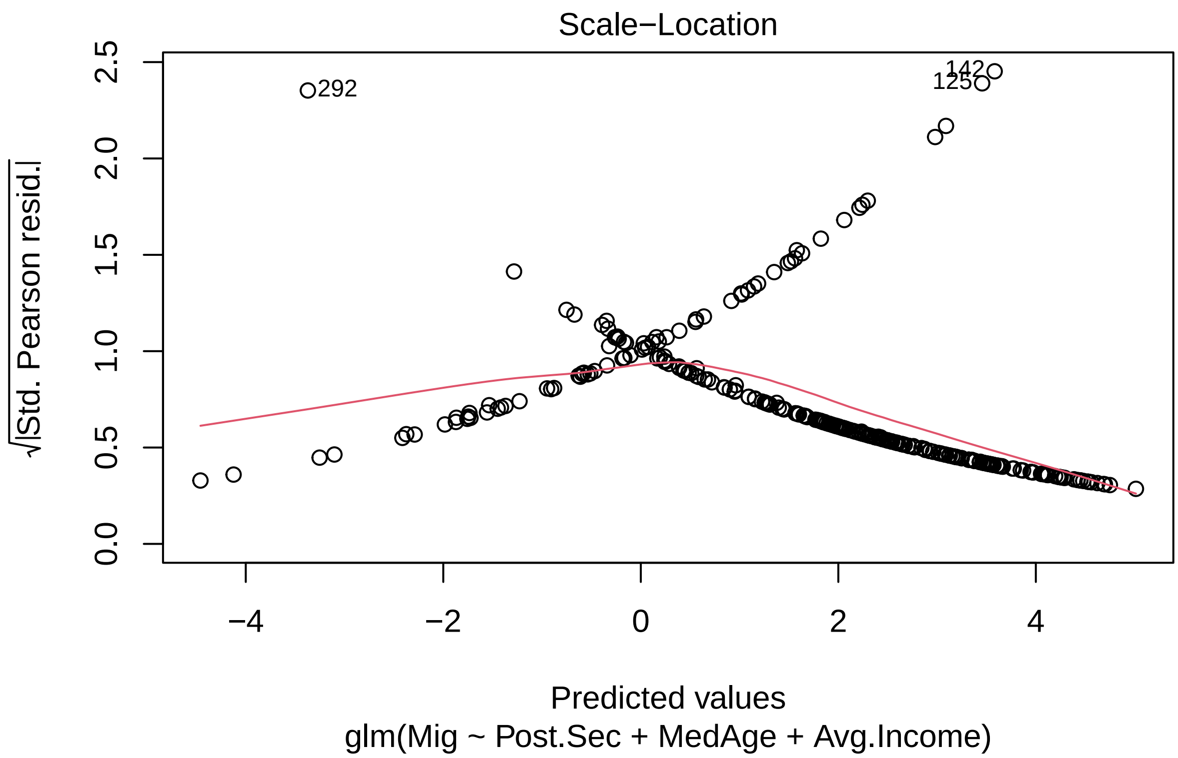
```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6871   0.1440   0.3065   0.5406   2.6096
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.173e-01  2.046e+00  -0.351 0.725881
## Post.Sec2   -5.303e-01  1.598e+00  -0.332 0.740053
## Post.Sec3    1.593e+00  1.494e+00   1.066 0.286235
## Post.Sec4    3.674e+00  1.558e+00   2.358 0.018382 *
## Post.Sec5    4.895e+00  1.883e+00   2.600 0.009318 **
## MedAge       9.908e-02  3.857e-02   2.569 0.010199 *
## Avg.Income  -1.052e-04  3.031e-05  -3.472 0.000516 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 210.90  on 286  degrees of freedom
## AIC: 224.9
##
## Number of Fisher Scoring iterations: 5
```

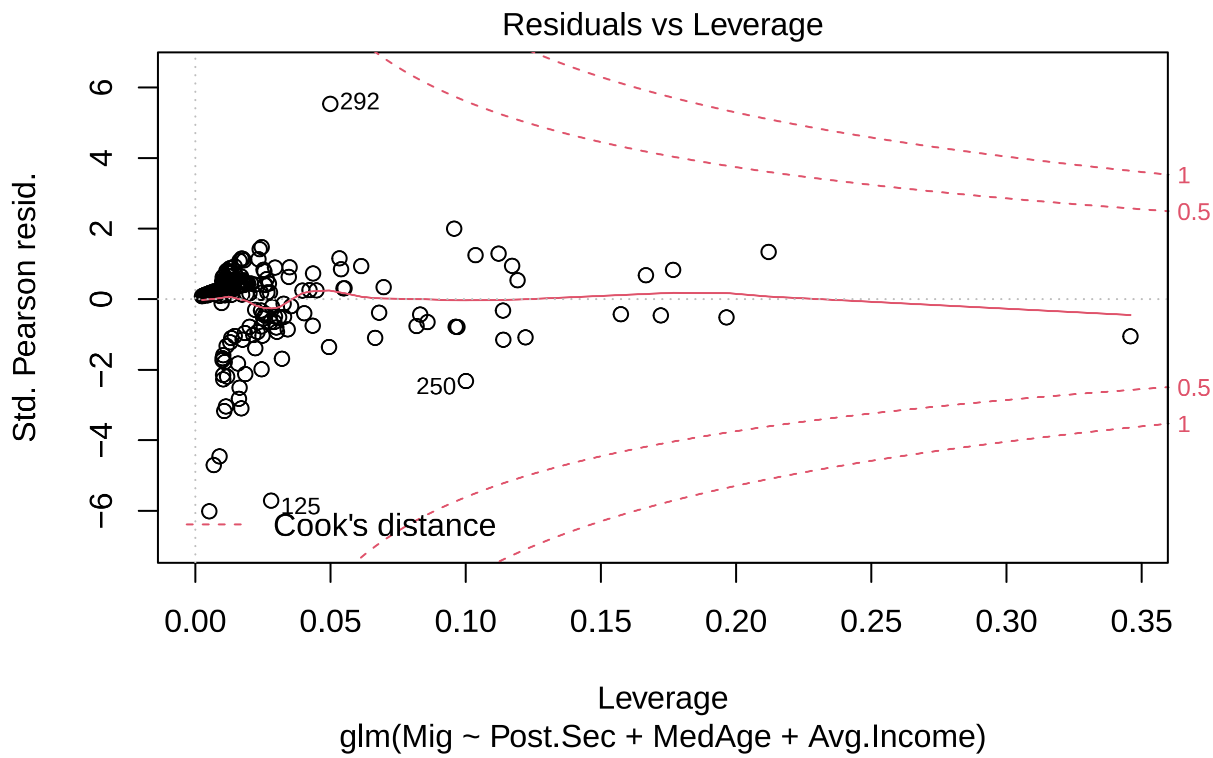
```
plot(Model2.2)
```





Normal Q−Q
Std. Pearson resid.
6
4
2
0
−2
−4
−6
−3
−2
−1
0
1
2
3
292
142
125
Theoretical Quantiles
glm(Mig ~ Post.Sec + MedAge + Avg.Income)


Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
2.5
−4
−2
0
2
4
292
142
125
Predicted values
glm(Mig ~ Post.Sec + MedAge + Avg.Income)




```
anova(Model2.2, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Mig
##
## Terms added sequentially (first to last)
##
##
##            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         292     317.48
## Post.Sec    4   52.026       288     265.46 1.362e-10 ***
## MedAge      1   39.919       287     225.54 2.647e-10 ***
## Avg.Income  1   14.636       286     210.90 0.0001304 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Avg Inc as variable of interest

Use post sec and MedAge

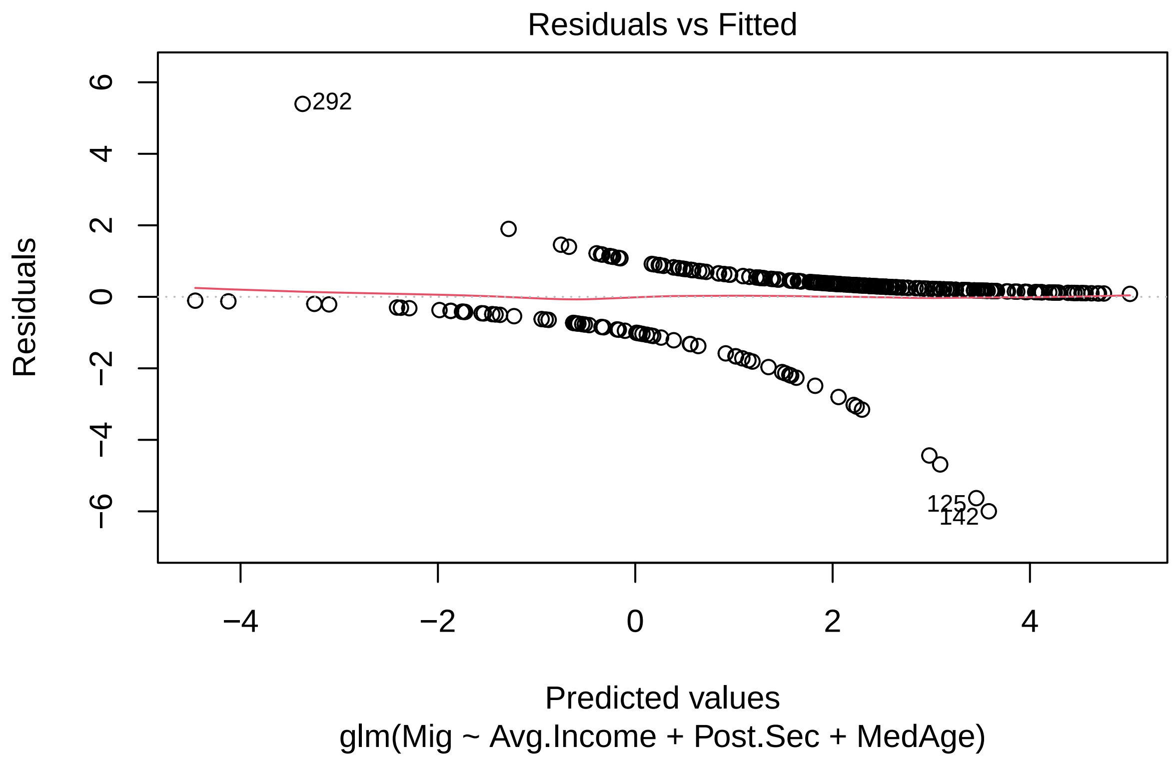
```r
Model3.1 <- glm(Mig ~ Avg.Income+Post.Sec+MedAge, family=binomial)
Model3.1
```

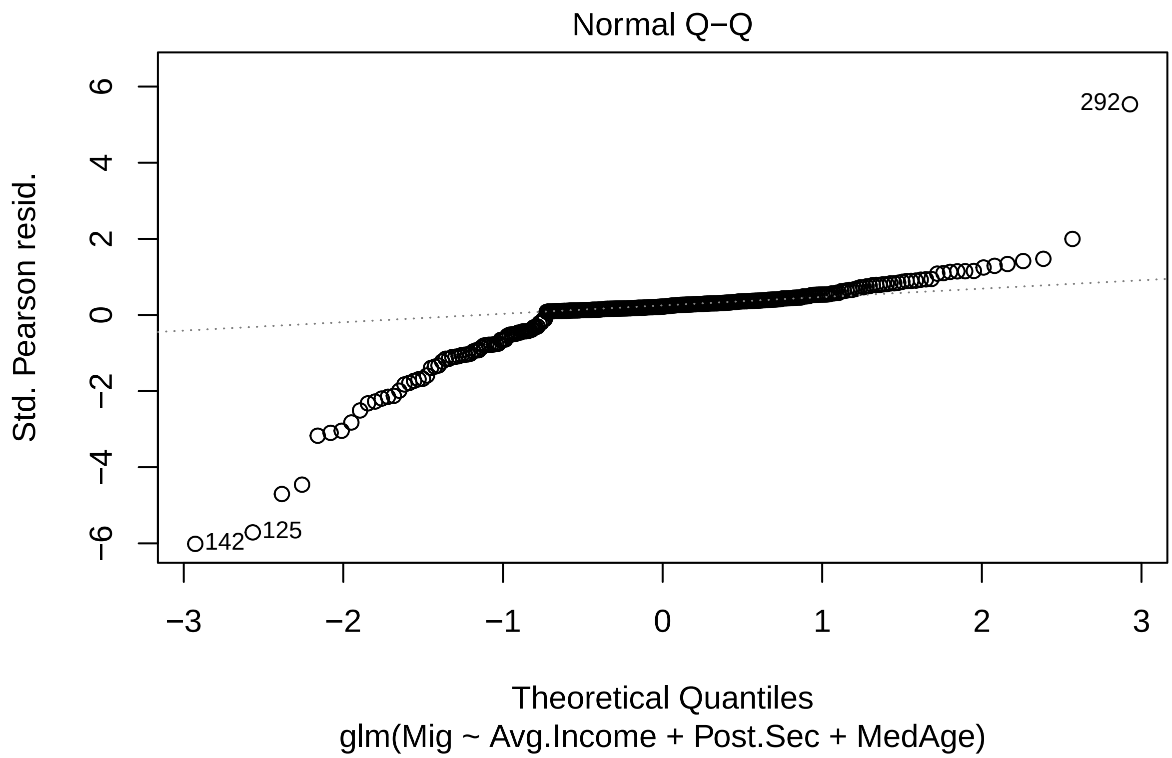
```
##
## Call:  glm(formula = Mig ~ Avg.Income + Post.Sec + MedAge, family = binomial)
##
## Coefficients:
## (Intercept)   Avg.Income    Post.Sec2    Post.Sec3    Post.Sec4    Post.Sec5
##  -0.7173434   -0.0001052   -0.5303339    1.5932386    3.6738848    4.8949600
##      MedAge
##   0.0990790
##
## Degrees of Freedom: 292 Total (i.e. Null);  286 Residual
## Null Deviance:       317.5
## Residual Deviance: 210.9     AIC: 224.9
```

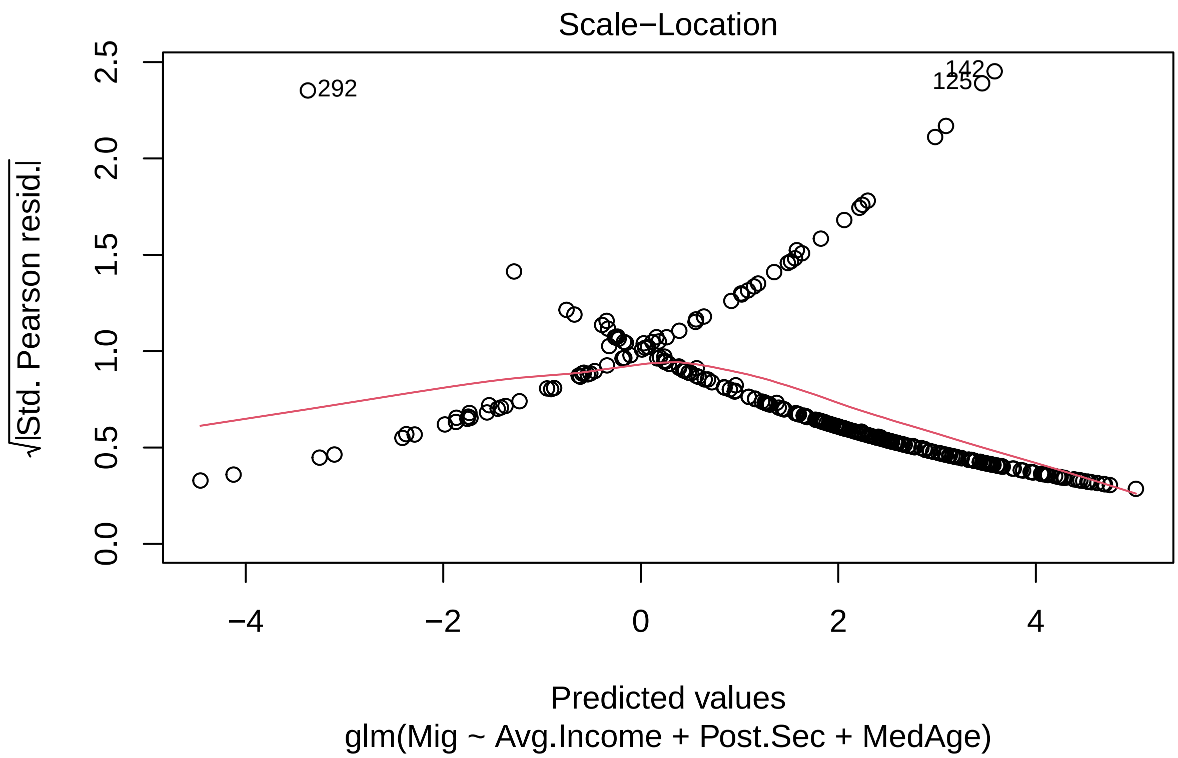
```r
summary(Model3.1)
```

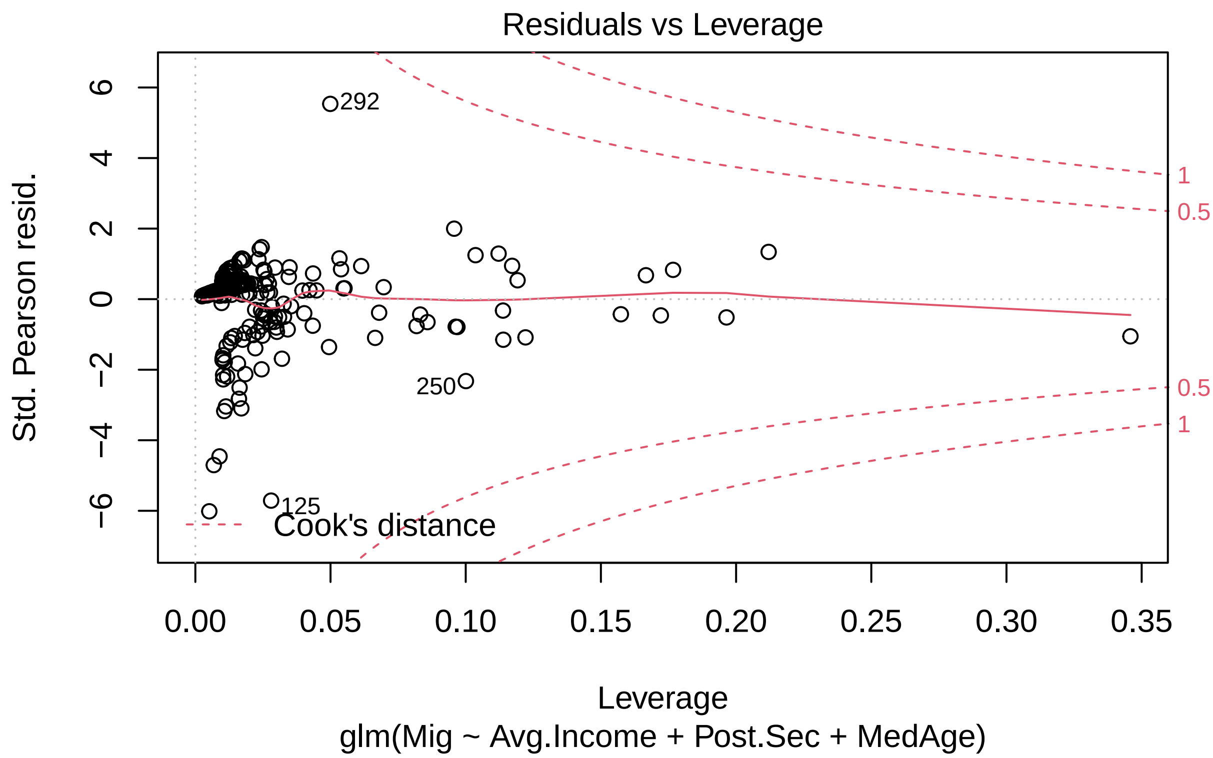
```
##
## Call:
## glm(formula = Mig ~ Avg.Income + Post.Sec + MedAge, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6871   0.1440   0.3065   0.5406   2.6096
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.173e-01  2.046e+00  -0.351 0.725881
## Avg.Income  -1.052e-04  3.031e-05  -3.472 0.000516 ***
## Post.Sec2   -5.303e-01  1.598e+00  -0.332 0.740053
## Post.Sec3    1.593e+00  1.494e+00   1.066 0.286235
## Post.Sec4    3.674e+00  1.558e+00   2.358 0.018382 *
## Post.Sec5    4.895e+00  1.883e+00   2.600 0.009318 **
## MedAge       9.908e-02  3.857e-02   2.569 0.010199 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 210.90  on 286  degrees of freedom
## AIC: 224.9
##
## Number of Fisher Scoring iterations: 5
```

```r
plot(Model3.1)
```


Residuals vs Fitted
Residuals
6
4
2
0
−2
−4
−6
292
125
142
−4
−2
0
2
4
Predicted values
glm(Mig ~ Avg.Income + Post.Sec + MedAge)


Normal Q−Q
Std. Pearson resid.
6
4
2
0
−2
−4
−6
292
142
125
−3
−2
−1
0
1
2
3
Theoretical Quantiles
glm(Mig ~ Avg.Income + Post.Sec + MedAge)


Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
2.5
−4
−2
0
2
4
292
142
125
Predicted values
glm(Mig ~ Avg.Income + Post.Sec + MedAge)




```
anova(Model3.1, test="LRT")
```

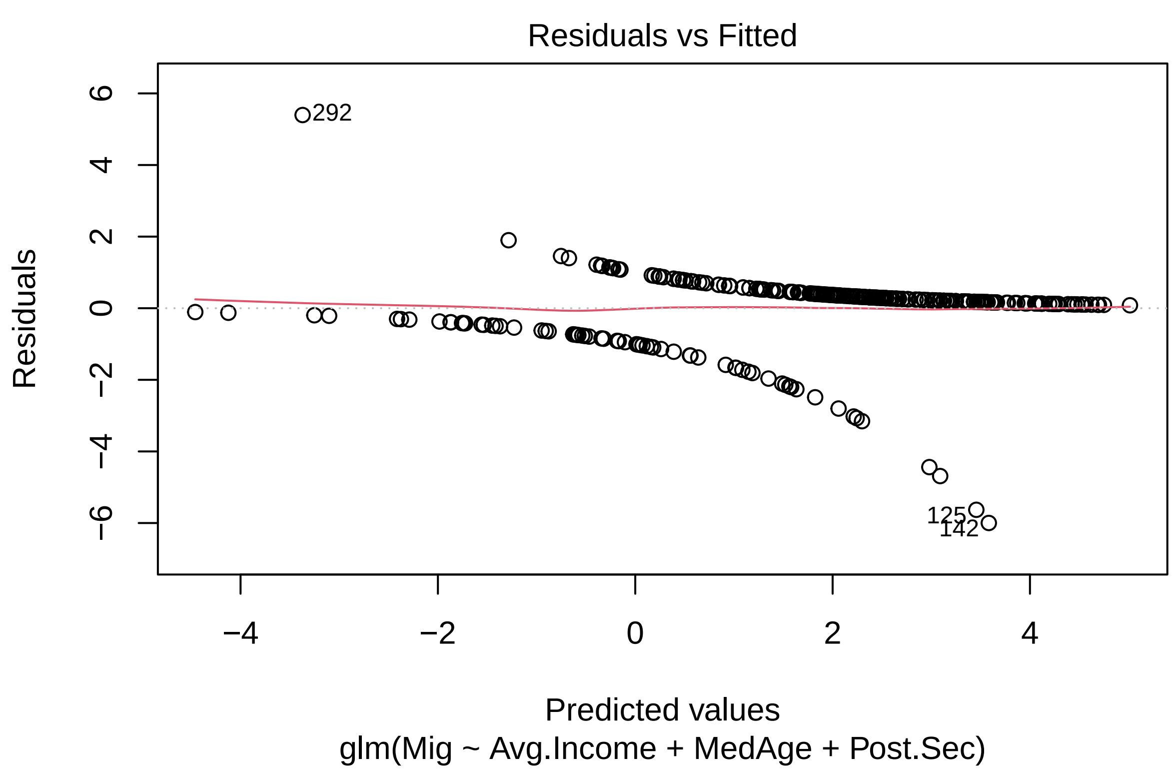
```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Mig
##
## Terms added sequentially (first to last)
##
##
##            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         292     317.48
## Avg.Income  1   18.014       291     299.47 2.193e-05 ***
## Post.Sec    4   81.791       287     217.68 < 2.2e-16 ***
## MedAge      1    6.777       286     210.90  0.009235 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

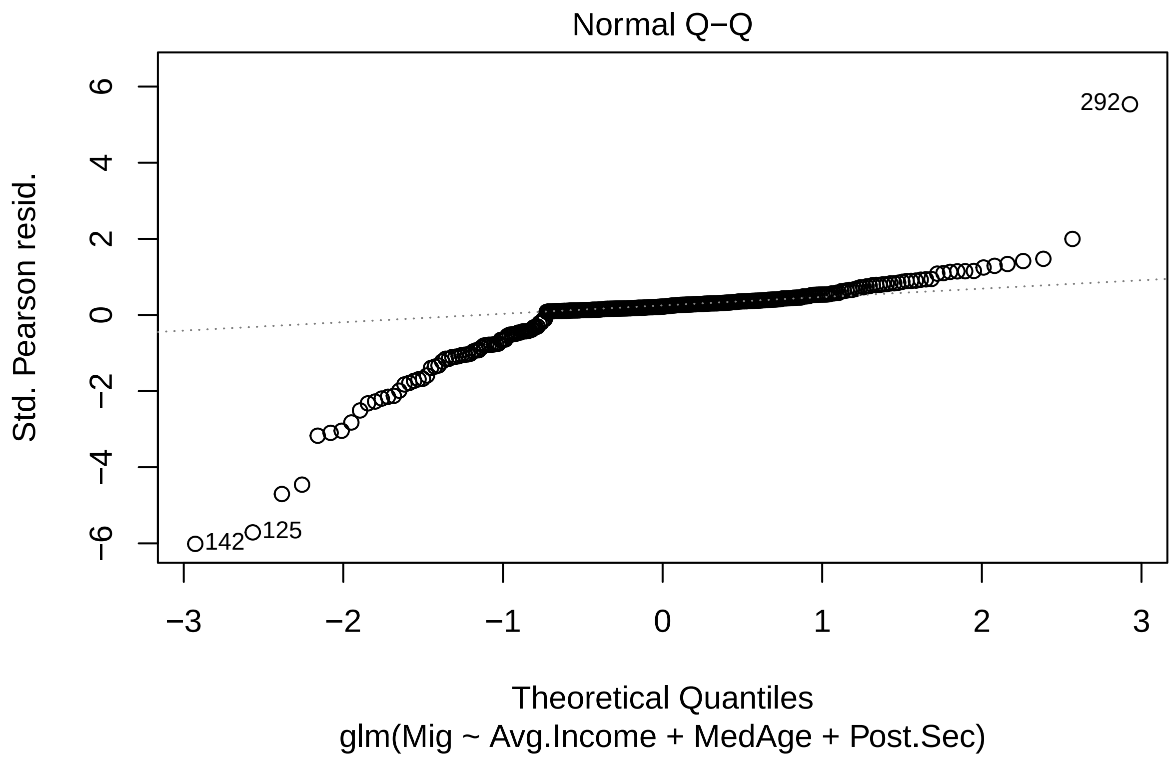
```
Model3.2 <- glm(Mig ~ Avg.Income+MedAge+Post.Sec, family=binomial)
summary(Model3.2)
```

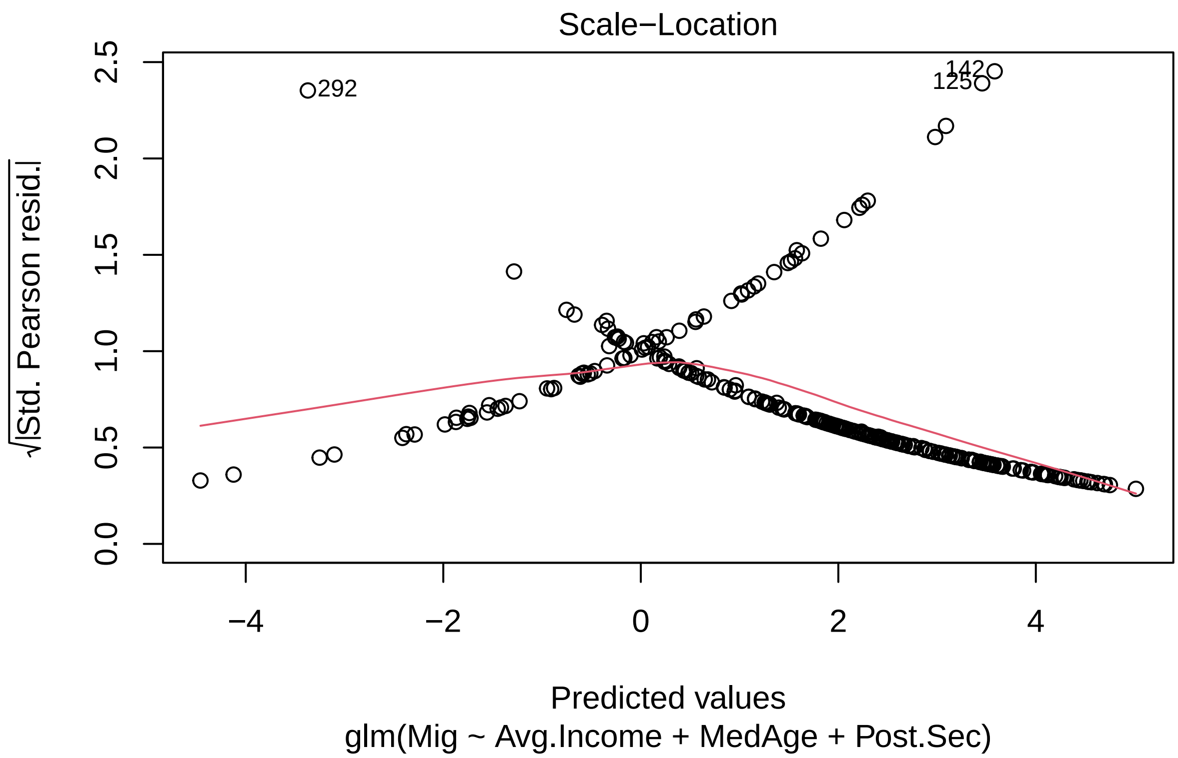
```
##
## Call:
## glm(formula = Mig ~ Avg.Income + MedAge + Post.Sec, family = binomial)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6871   0.1440   0.3065   0.5406   2.6096
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.173e-01  2.046e+00  -0.351 0.725881
## Avg.Income  -1.052e-04  3.031e-05  -3.472 0.000516 ***
## MedAge       9.908e-02  3.857e-02   2.569 0.010199 *
## Post.Sec2   -5.303e-01  1.598e+00  -0.332 0.740053
## Post.Sec3    1.593e+00  1.494e+00   1.066 0.286235
## Post.Sec4    3.674e+00  1.558e+00   2.358 0.018382 *
## Post.Sec5    4.895e+00  1.883e+00   2.600 0.009318 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 210.90  on 286  degrees of freedom
## AIC: 224.9
##
## Number of Fisher Scoring iterations: 5
```

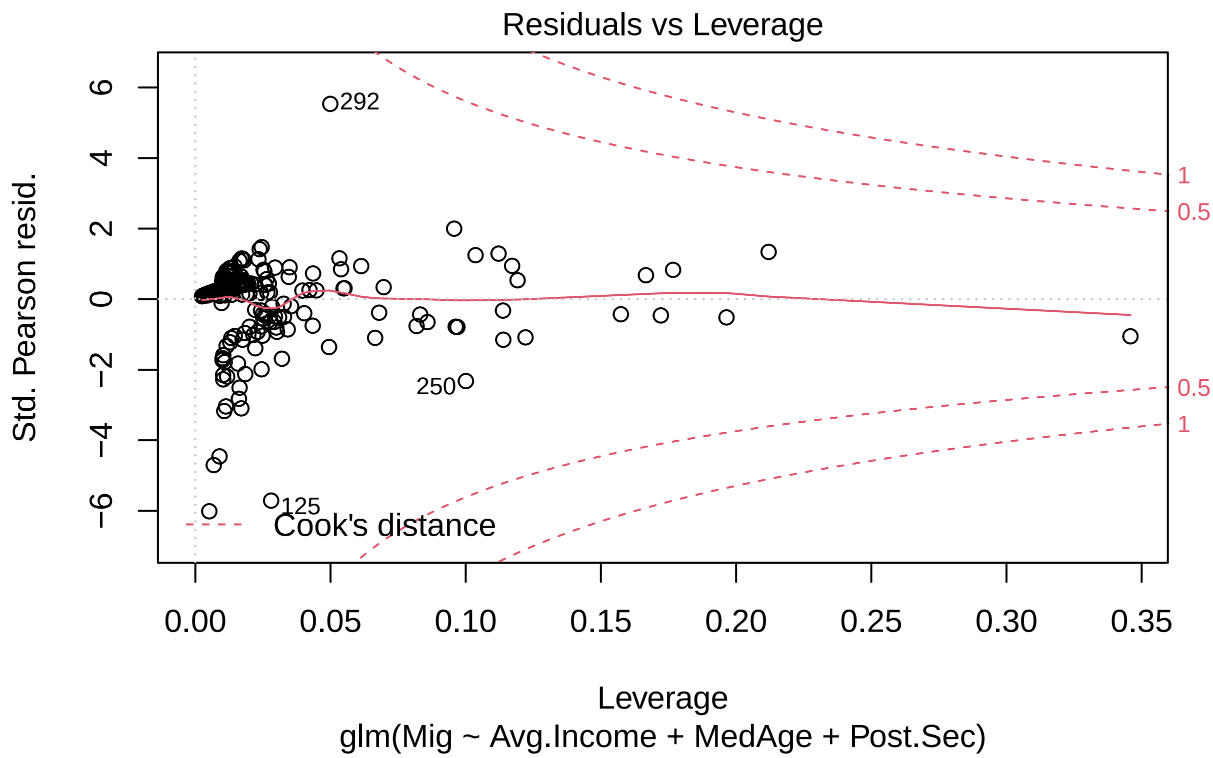
```
plot(Model3.2)
```




Normal Q−Q
Std. Pearson resid.
6
4
2
0
−2
−4
−6
292
142
125
−3
−2
−1
0
1
2
3
Theoretical Quantiles
glm(Mig ~ Avg.Income + MedAge + Post.Sec)



Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
2.5
292
142
125
−4
−2
0
2
4
Predicted values
glm(Mig ~ Avg.Income + MedAge + Post.Sec)




```
anova(Model3.2, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Mig
##
## Terms added sequentially (first to last)
##
##
##            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         292     317.48
## Avg.Income  1   18.014       291     299.47 2.193e-05 ***
## MedAge      1   35.686       290     263.78 2.319e-09 ***
## Post.Sec    4   52.882       286     210.90 9.021e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

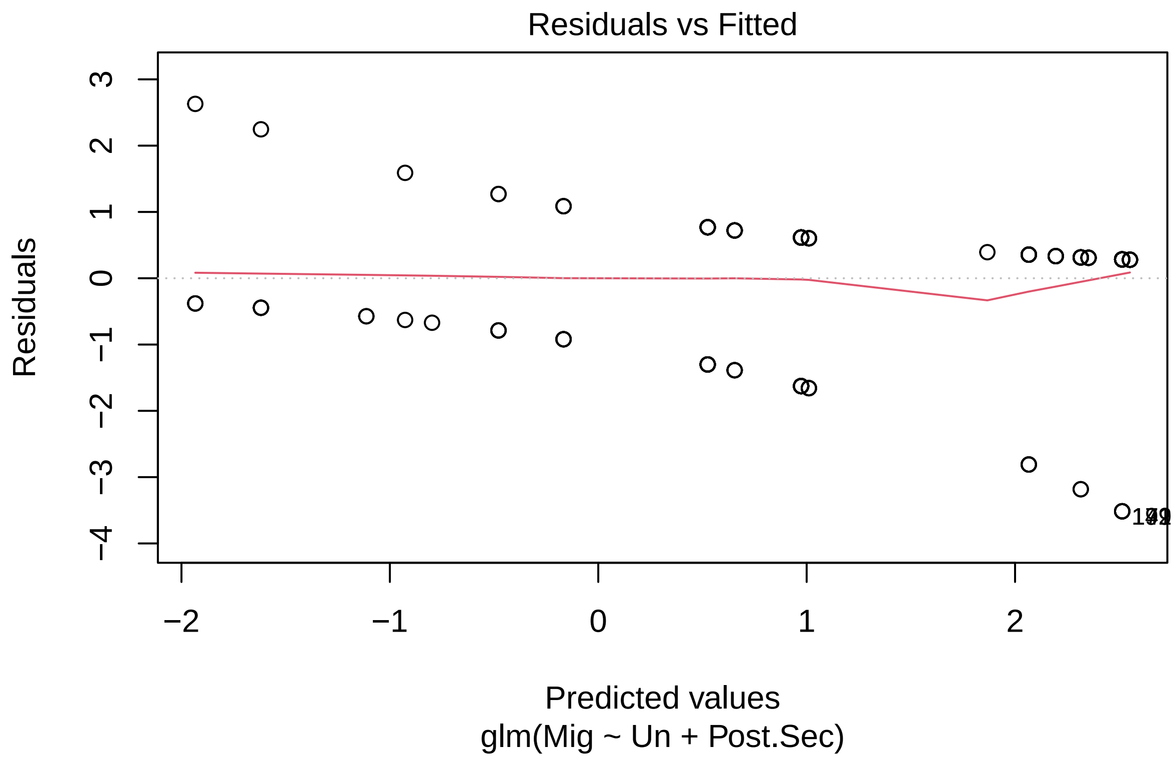
Un as variable of interest

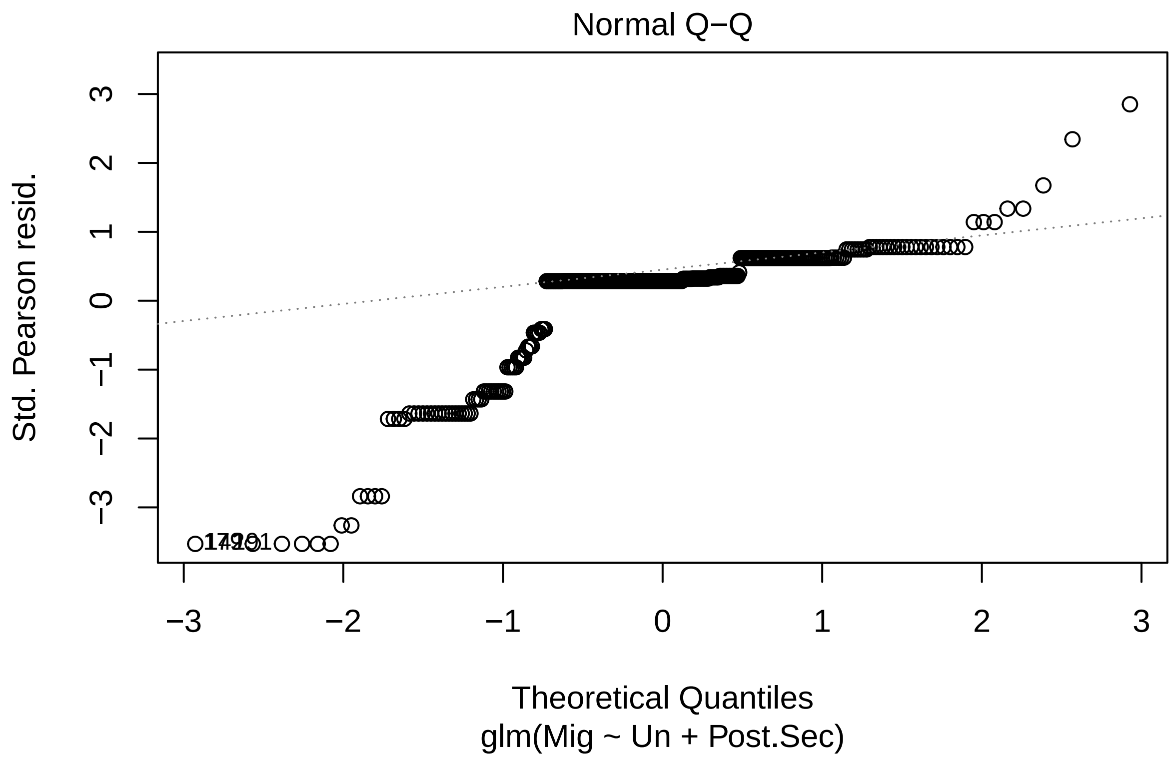
Only Post Sec could be used well

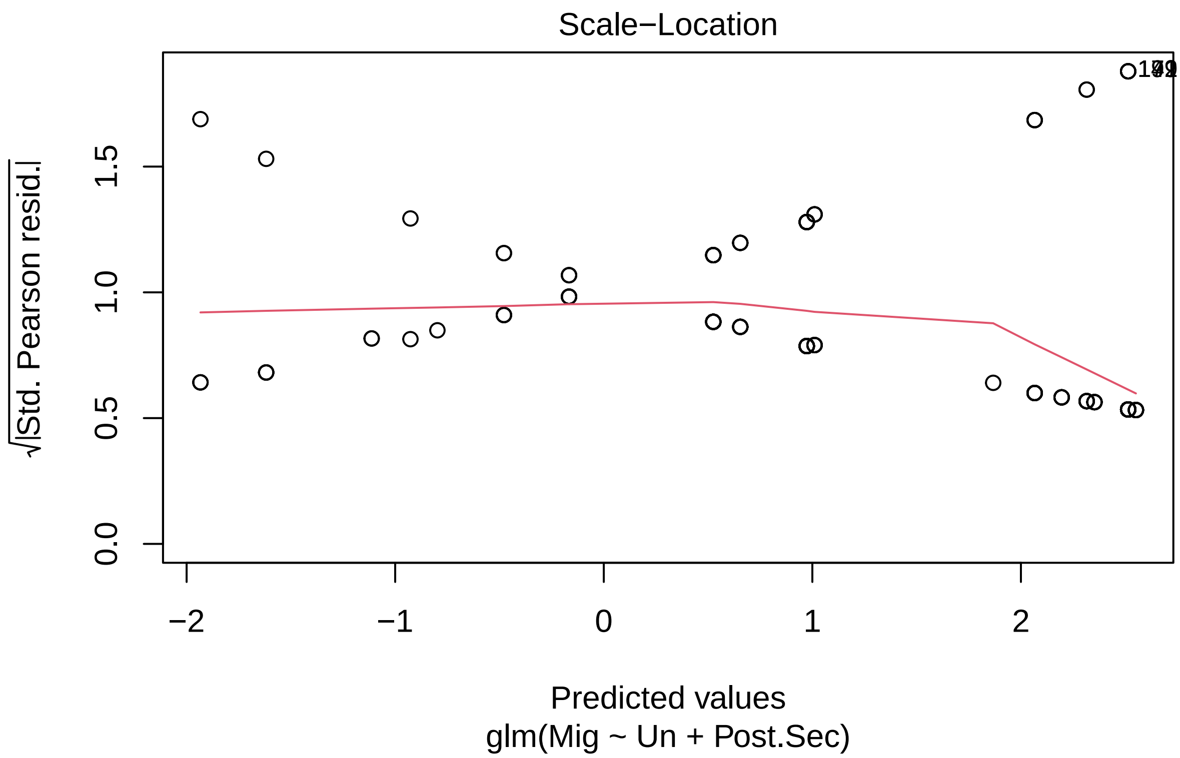
```r
Model4 <- glm(Mig ~ Un+Post.Sec, family=binomial)
summary(Model4)
```

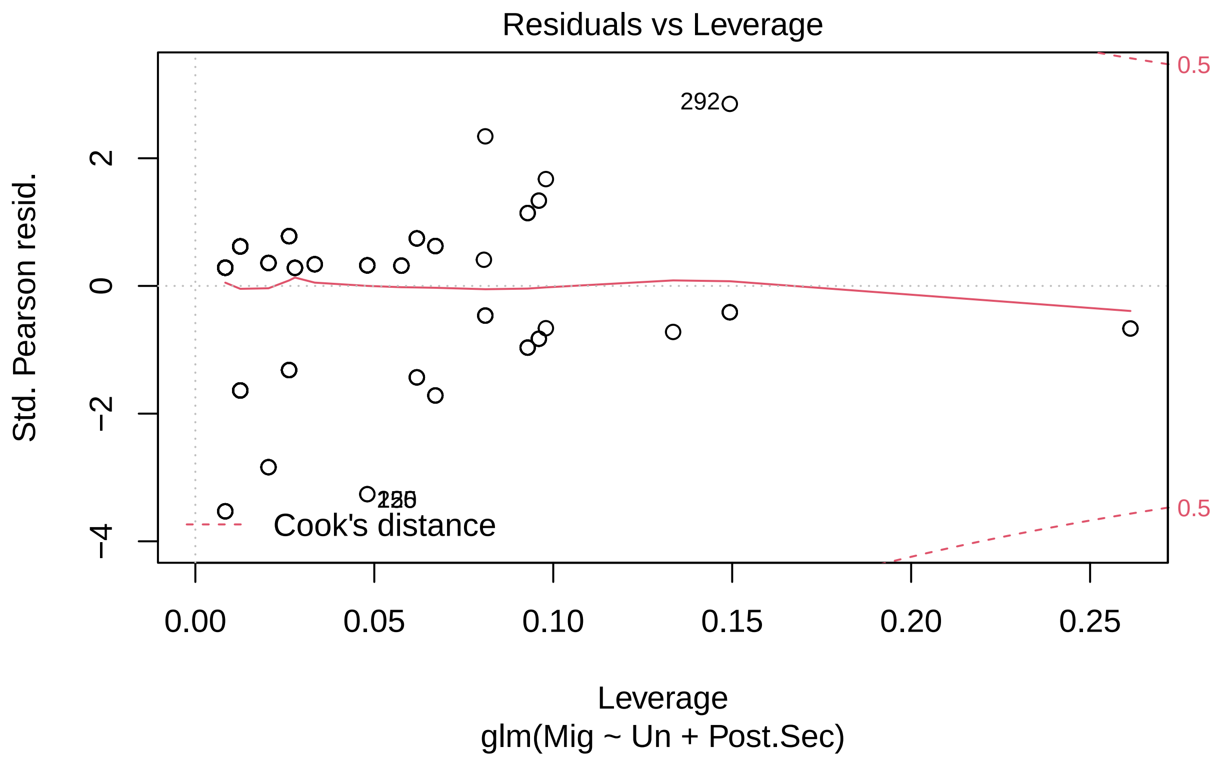
```
##
## Call:
## glm(formula = Mig ~ Un + Post.Sec, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2768   0.3875   0.3945   0.8006   2.0341
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.75668    1.32002  -0.573  0.56649
## Un2         -0.03731    0.60855  -0.061  0.95111
## Un3         -0.48548    0.66106  -0.734  0.46270
## Un4         -0.35617    0.77644  -0.459  0.64643
## Un5         -1.17713    0.84262  -1.397  0.16242
## Post.Sec2    0.31526    1.29811   0.243  0.80811
## Post.Sec3    1.76725    1.18751   1.488  0.13670
## Post.Sec4    3.30811    1.23472   2.679  0.00738 **
## Post.Sec5    3.10927    1.41761   2.193  0.02828 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 261.49  on 284  degrees of freedom
## AIC: 279.49
##
## Number of Fisher Scoring iterations: 5
```

```r
plot(Model4)
```


Residuals vs Fitted
Residuals
3
2
1
0
−1
−2
−3
−4
−2
−1
0
1
2
Predicted values
glm(Mig ~ Un + Post.Sec)


Normal Q−Q
Std. Pearson resid.
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
Theoretical Quantiles
glm(Mig ~ Un + Post.Sec)


Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
−2
−1
0
1
2
Predicted values
glm(Mig ~ Un + Post.Sec)




```
anova(Model4, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Mig
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                       292     317.48
## Un        4   24.195       288     293.29 7.299e-05 ***
## Post.Sec  4   31.798       284     261.49 2.104e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

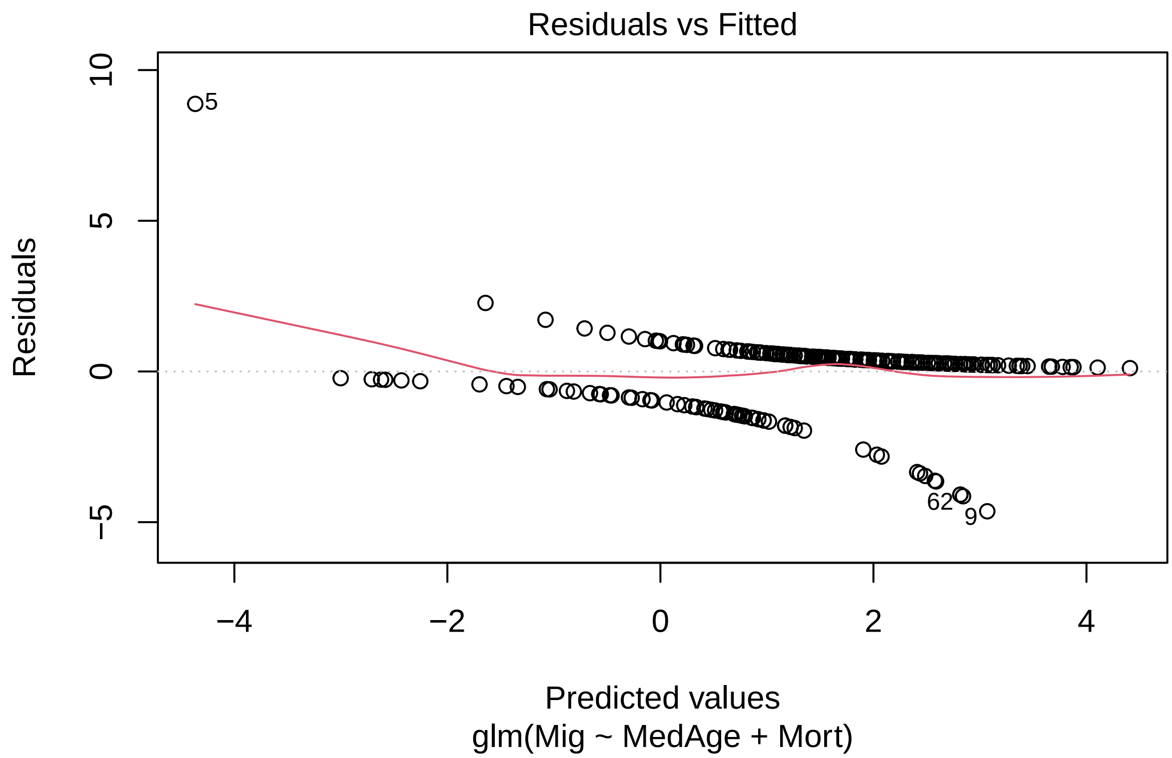
##Med as variable of interest Only Mort could be used well

```
Model5 <- glm(Mig ~ MedAge+Mort, family=binomial)
summary(Model5)
```

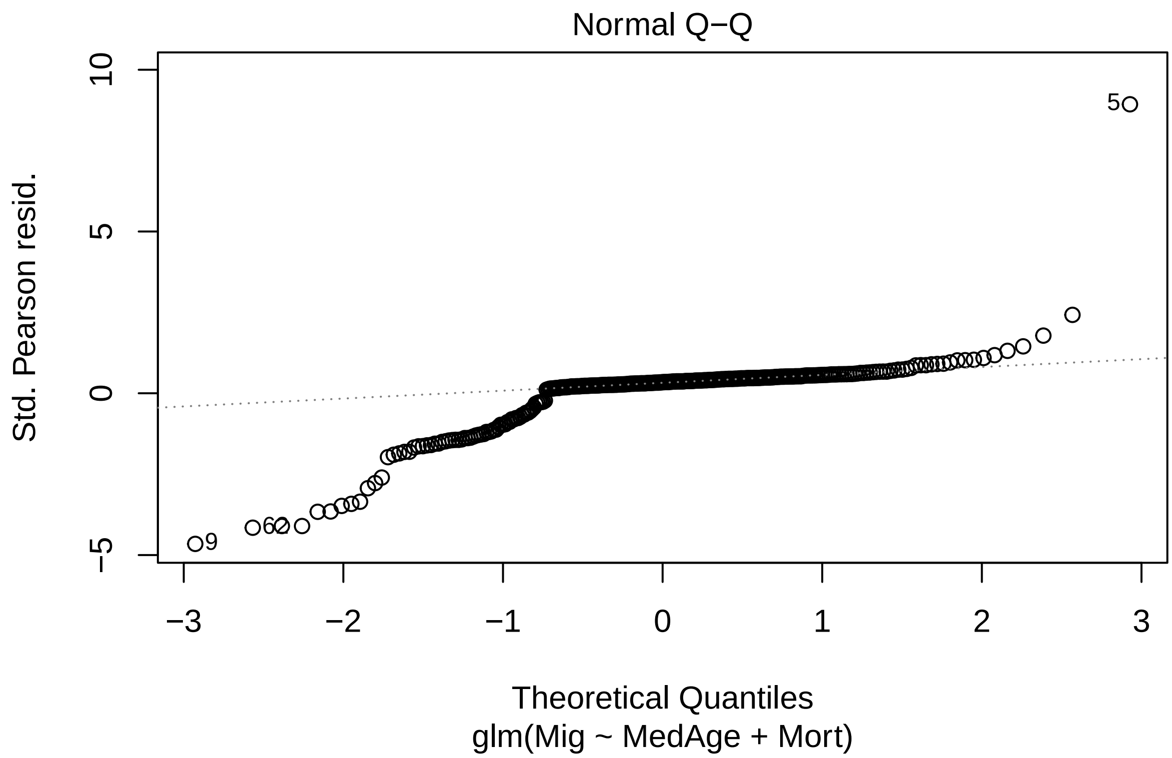
```
##
```

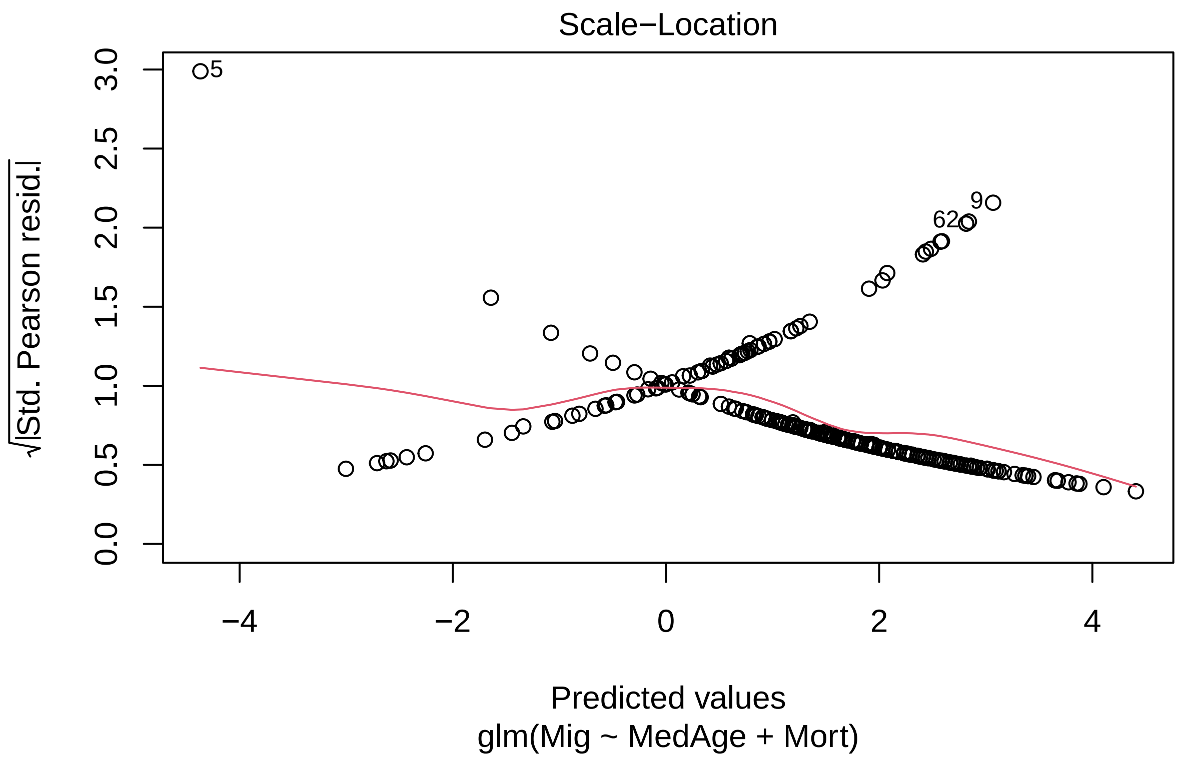
```
## Call:
## glm(formula = Mig ~ MedAge + Mort, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4958   0.2237   0.4674   0.6491   2.9596
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.56076    1.62196  -4.662 3.14e-06 ***
## MedAge       0.25293    0.04128   6.127 8.98e-10 ***
## Mort2       -1.52305    0.91993  -1.656  0.09780 .
## Mort3       -1.81532    0.94525  -1.920  0.05480 .
## Mort4       -2.67437    1.03566  -2.582  0.00982 **
## Mort5       -3.60972    1.17361  -3.076  0.00210 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 317.48  on 292  degrees of freedom
## Residual deviance: 252.13  on 287  degrees of freedom
## AIC: 264.13
##
## Number of Fisher Scoring iterations: 5
```

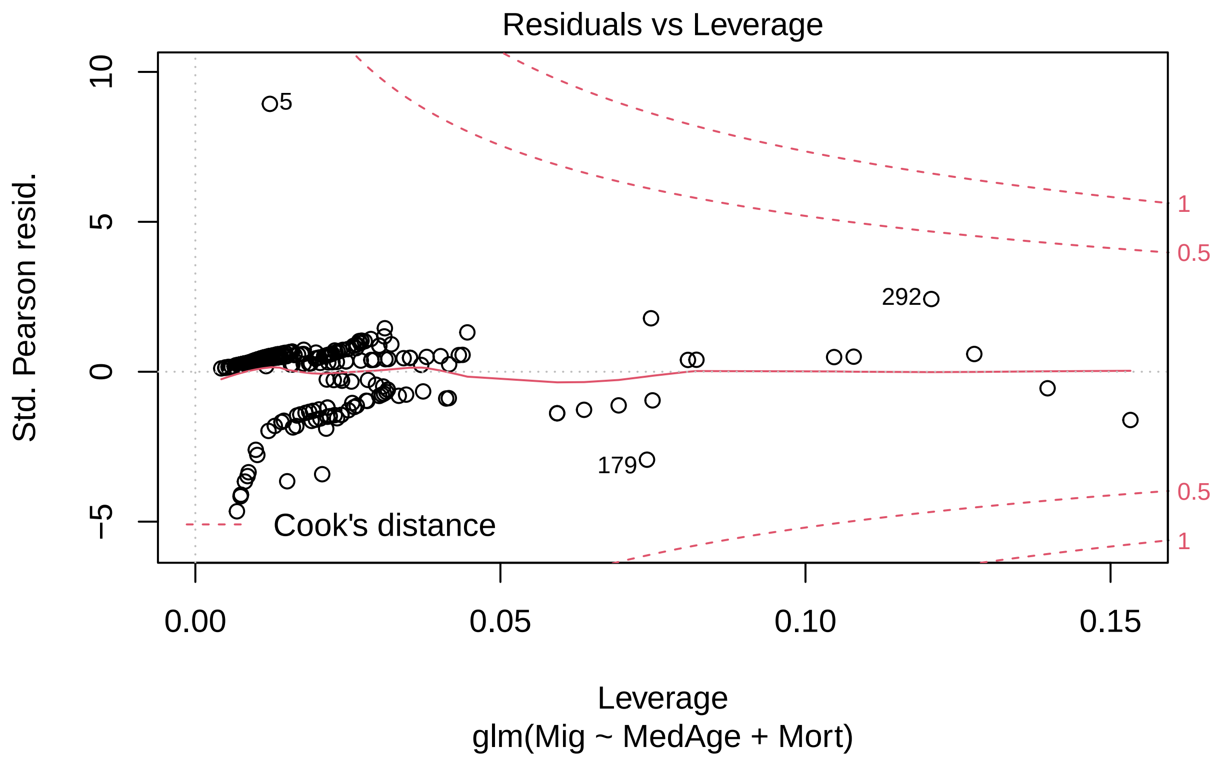
```
plot(Model5)
```



Residuals vs Fitted
Residuals
10
5
0
−5
−4
−2
0
2
4
5
62
9
Predicted values
glm(Mig ~ MedAge + Mort)


Normal Q−Q
Std. Pearson resid.
10
5
0
−5
−3
−2
−1
0
1
2
3
Theoretical Quantiles
glm(Mig ~ MedAge + Mort)
5
9
6


Scale−Location
3.0
2.5
2.0
1.5
1.0
0.5
0.0
√|Std. Pearson resid.|
5
9
62
−4
−2
0
2
4
Predicted values
glm(Mig ~ MedAge + Mort)




```
anova(Model5, test="LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Mig
##
## Terms added sequentially (first to last)
##
##
##        Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                     292     317.48
## MedAge  1   52.198       291     265.29 5.019e-13 ***
## Mort    4   13.157       287     252.13   0.01053 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnostics for Model 3.1: Residuals vs Fitted: Looks linear Normal QQ most observations fit in the line although there are heavy tails. Recall that the histogram of the response variable suggested non-normality. Scale Location: Line not horizontal. Violation of homoscedasticity assumption, which is to be expected given that Migration and Avg.Income has a wide range, but we cannot exclude them from the analysis. Look at Observation 292 and 125. They are the closest to the dashed lines but still do not lie outside Cook's Distance. Therefore, there are no influential points.

CONCLUSIONS

It appears that, based on the p-values and the coefficient estimates, that average income and post-secondary opportunities have the biggest influence on out-migration. That makes sense, since average income is tied to economic opportunities, and post-secondary opportunities are tied to job prospects.

Of course, this is not to say that this is perfect modelling. This is simply just one interpretation out of many. It is possible that different factorizations of categorical variables could yield drastically different results. It is also just as possible that there are socioeconomic indicators not captured by Statistics Canada that would both create better models and also better answer my questions. The main takeaway from this project is that official national data gathering institutions still do not fully capture the socioeconomic picture of Canada's diverse geopolitical divisions. There is lack of specific data on things such as how much systemic racism plays into such as healthcare access, mental health indicators, job prospects, and infrastructure development for communities with higher rates of Visible Minorities and Indigenous communities.