MSSP Consulting Report: Stress, Empathy and Self-care Analysis for Client Evelyn

Armstrong Wang; Tianjian Xie; Haochen Pan; Nuo Chen; Haochen Zhu

2022-12-12

INTRODUCTION

Our client is Evelyn Fisher from Boston University Graduate Medical Sciences, Genetic Counseling department. Her capstone is looking at self-reported levels of stress and empathy in current genetic counseling students. Our client's 1st goal is to characterize stress, empathy, and self-care among genetic counseling students. Their second goal is to identify relationships between these characteristics.

##Data Description

Data is collected through surveys adminstered to Genetic Counseling students throughout the US and Canada. 20 stress-related questions are converted into a score called PSQ (percieved stress questionaire). 28 IRI empathy-related questions are converted into a score called IRI (interpersonal reactivity index). There are 5 questions for self-care and some demographic questions.

PSQ: Perceived Stress Questionnaire, which means levels of stress. IRI: Interpersonal Reactivity Index, which means levels of empathy, there is 1 total score with 4 empathy variables score Self-care Questions: Exercise, therapy, socializing, religion, hobby Other covariance questions

Characterization of Stress, Empathy and Self-Care through Exploratory Data Analysis

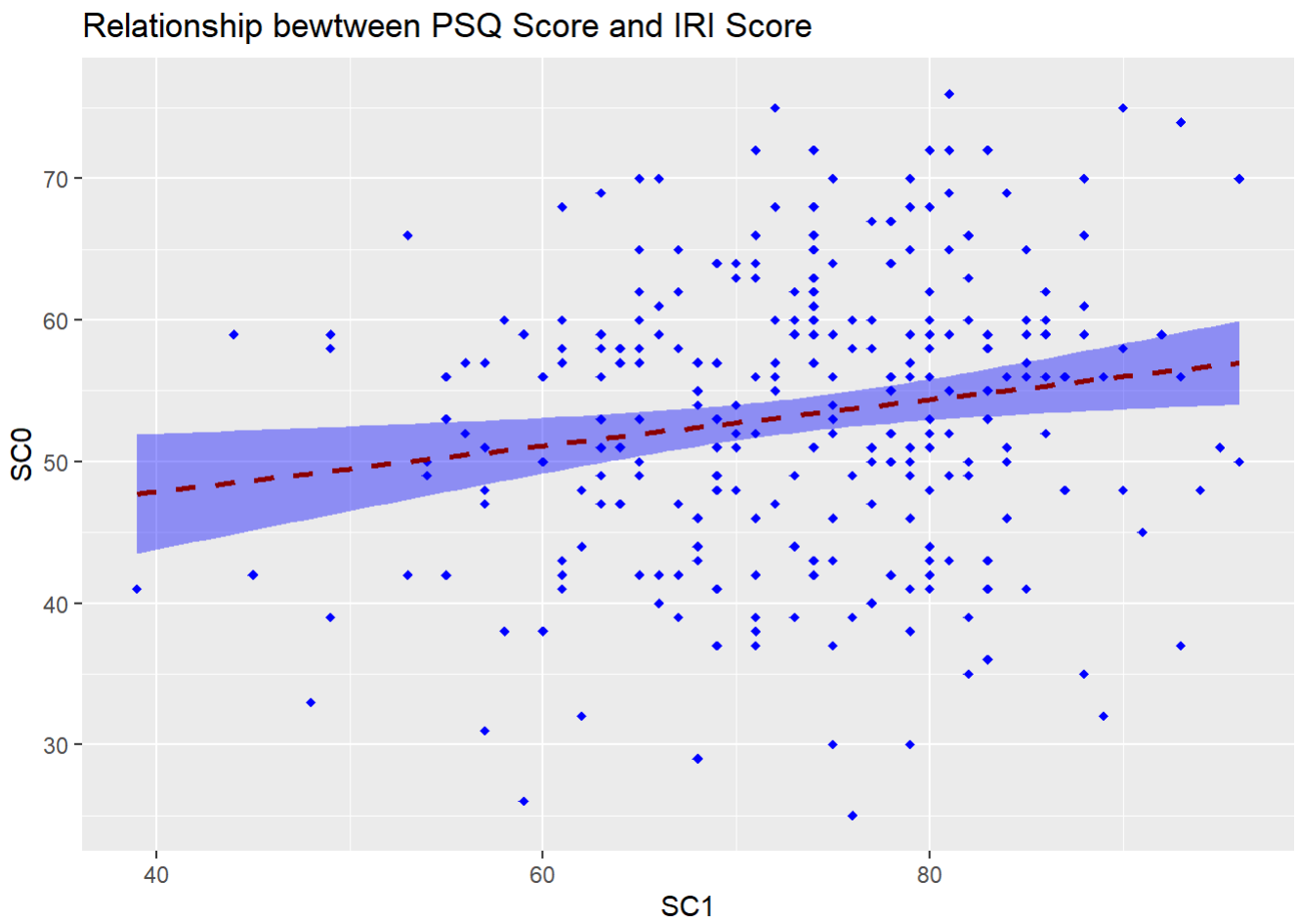
Summary statistics of stress, empathy and self care, including five number summaries, are included in the following figures.

```
##       SC0              SC1              SC2              SC3              SC4
##  Min.   :25.00   Min.   :39.00   Min.   : 7.00   Min.   :10.00   Min.   : 6.0
##  1st Qu.:46.00   1st Qu.:66.00   1st Qu.:16.00   1st Qu.:20.00   1st Qu.:18.0
##  Median :55.00   Median :74.00   Median :19.00   Median :23.00   Median :20.0
##  Mean   :53.26   Mean   :73.14   Mean   :18.61   Mean   :22.39   Mean   :19.9
##  3rd Qu.:60.00   3rd Qu.:80.00   3rd Qu.:21.00   3rd Qu.:25.00   3rd Qu.:22.0
##  Max.   :76.00   Max.   :96.00   Max.   :26.00   Max.   :28.00   Max.   :27.0
##       SC5
##  Min.   : 1.00
##  1st Qu.: 9.00
##  Median :11.00
##  Mean   :11.78
##  3rd Qu.:15.00
##  Max.   :28.00
```

```
##           vars   n mean   sd median trimmed  mad min max range  skew kurtosis
## Q83*         1 268 2.56 1.06      3    2.57 1.48   1   4     3 -0.02    -1.24
## Q85*         2 268 1.56 1.03      1    1.34 0.00   1   4     3  1.48     0.56
## Q86...71*    3 268 2.75 0.86      2    2.70 1.48   1   4     3  0.40    -1.34
## Q87...72*    4 268 1.50 1.02      1    1.26 0.00   1   4     3  1.79     1.54
## Q88...73*    5 268 2.38 1.15      2    2.36 1.48   1   4     3  0.12    -1.43
##              se
## Q83*       0.06
## Q85*       0.06
## Q86...71* 0.05
## Q87...72* 0.06
## Q88...73* 0.07
```

The first figure shows the characteristics of PSQ and IRI, while the second one shows the information about self-care questions.

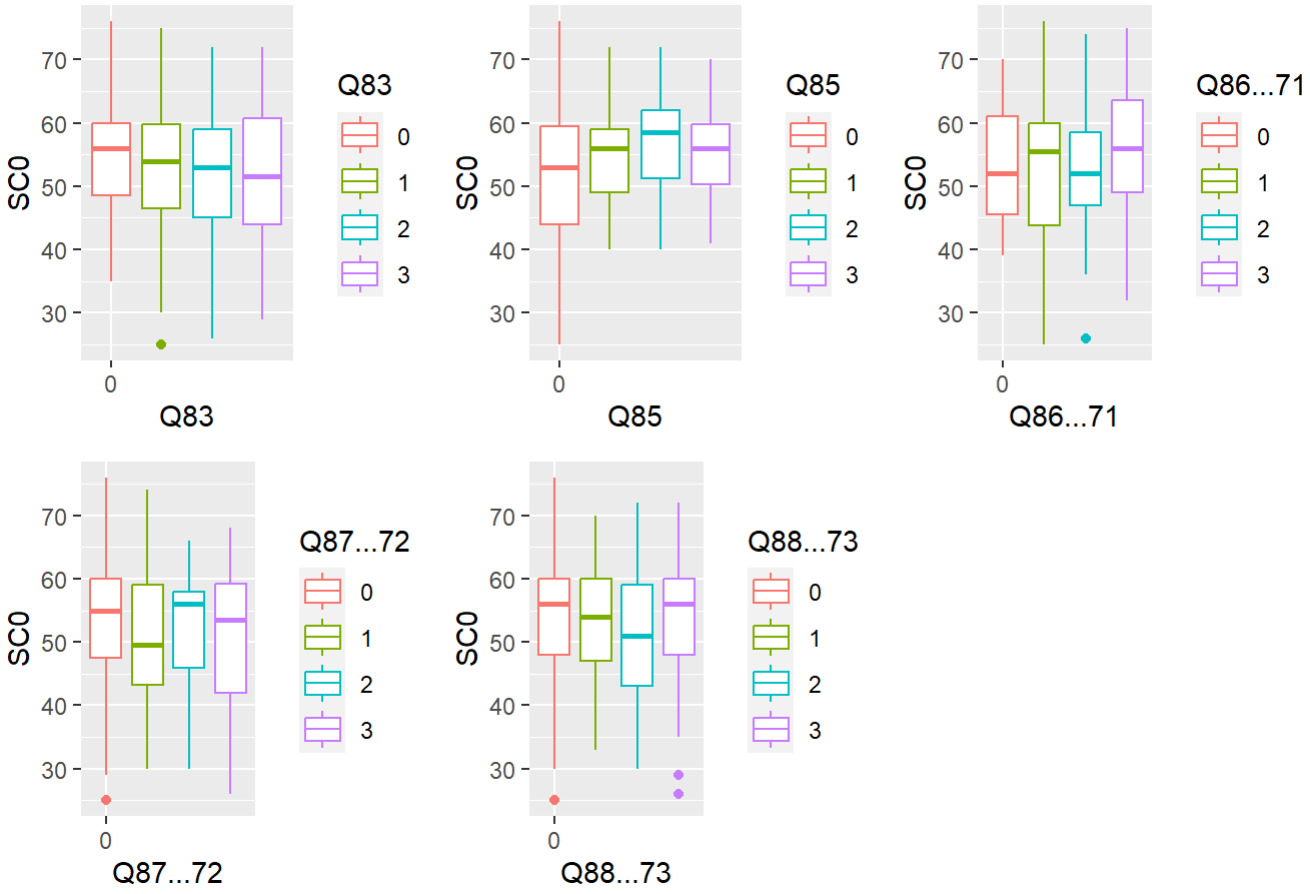
Next, we explore the relationship between PSQ and IRI, which can be seen in the following figure.



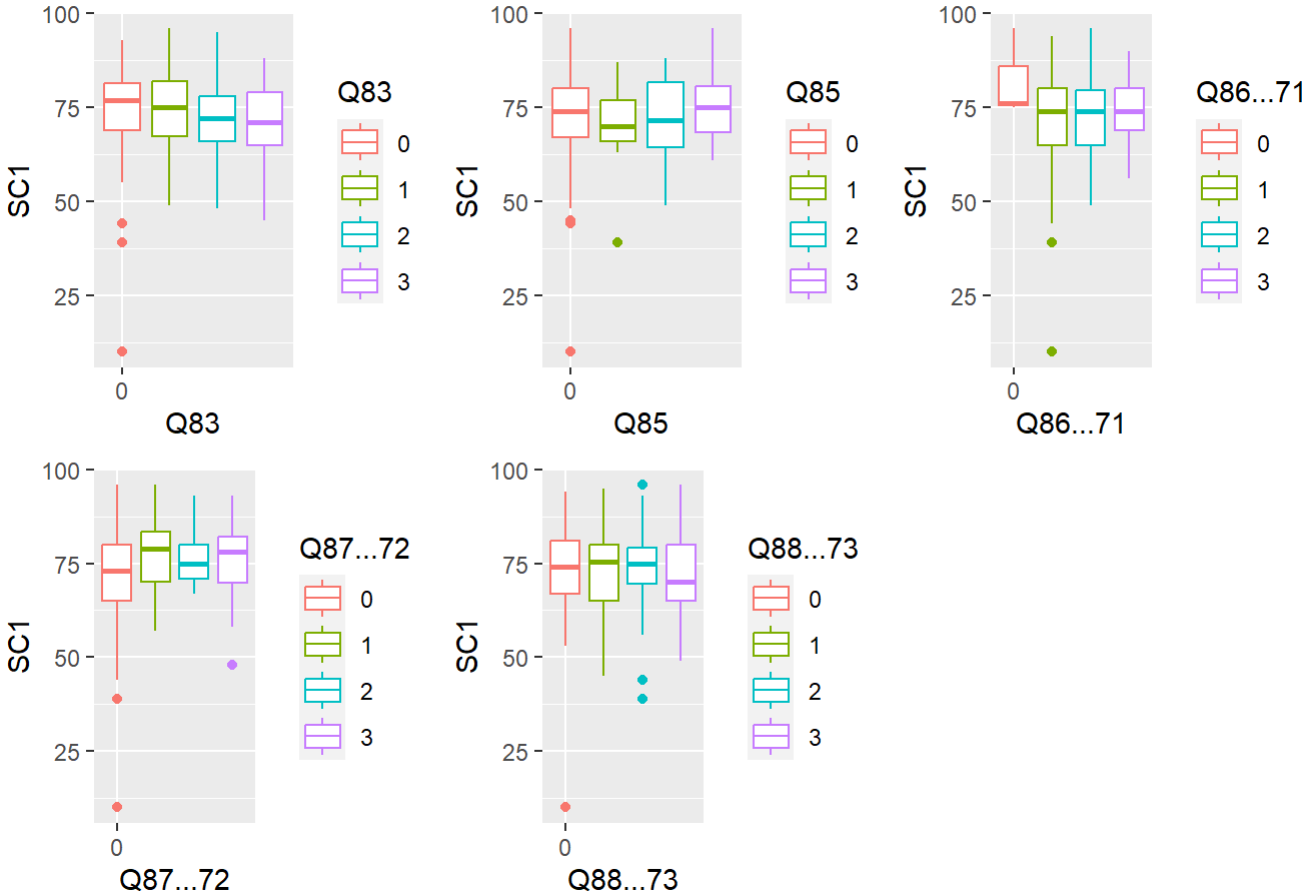
This is the scatter plot between the PSQ score and IRI score after we clean the data. The smooth line is not very steep, and the points are not concentrated close to the line, indicating the relationship between these variables is weak. A similar pattern happens to other LRL variables scores. But that is still able to prove that there is a relationship between stress and empathy.

The second step is the box plot for PSQ/IRI and self-care questions.

PSQ Score vs. self-care variable

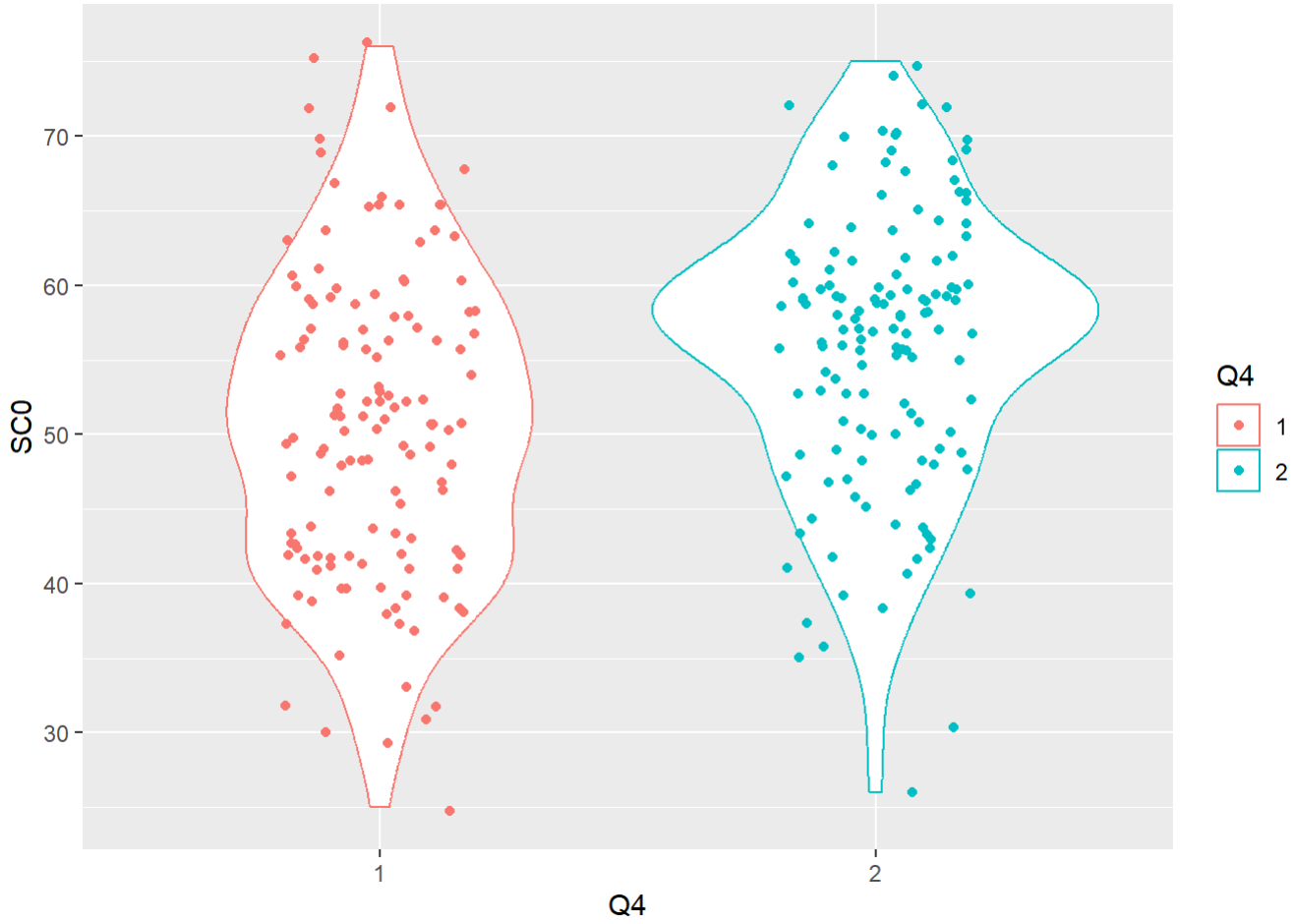


IRI Score vs. self-care variable



We can see that the question about exercise shows some relationship with PSQ and IRI, but a similar thing did not happen to other questions. We will explain it in the next parts.

In the next part, we examine some demographic questions. that can also affect stress/empathy, such as the year of students. So we will put this in our modeling part.



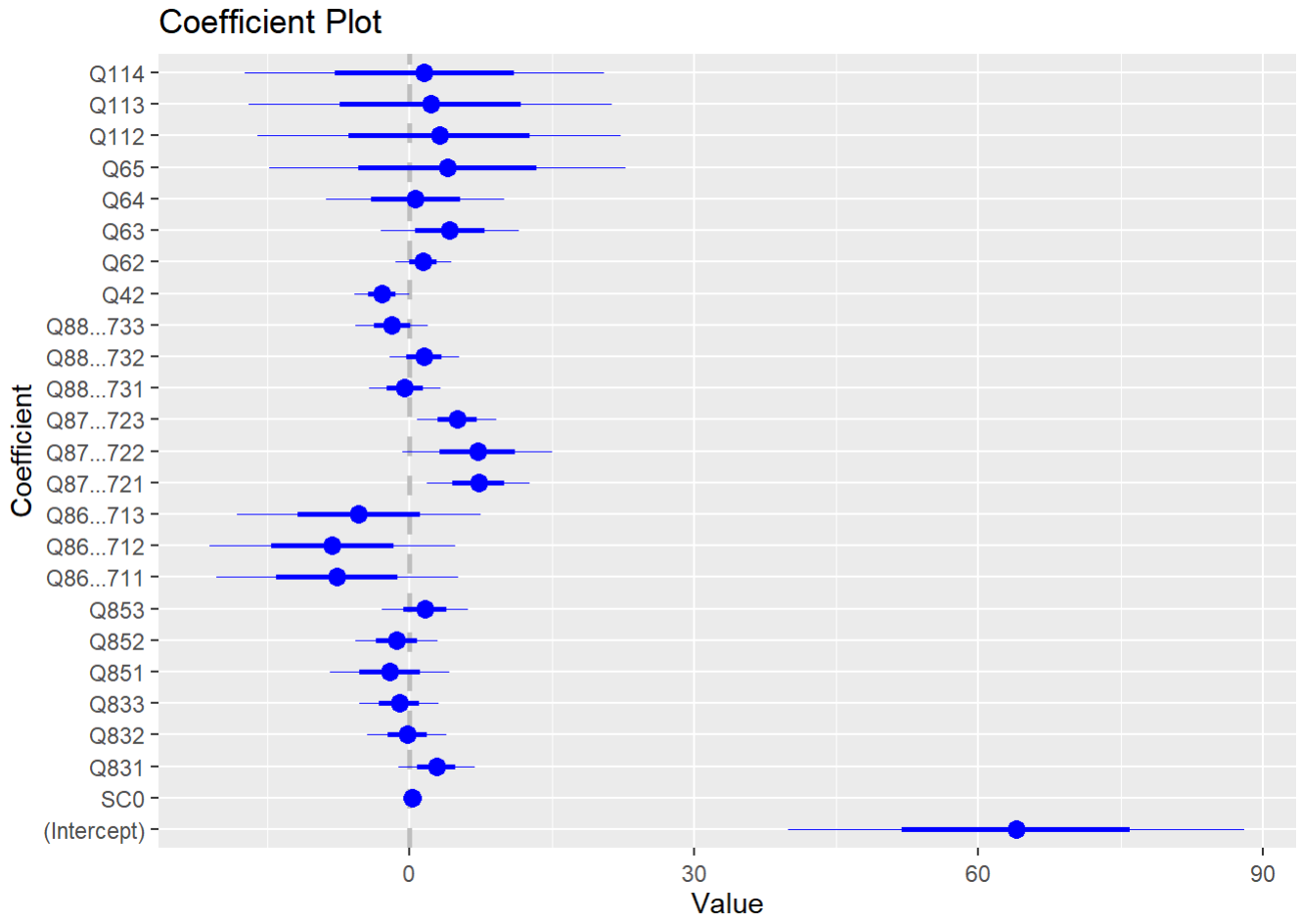
Modeling

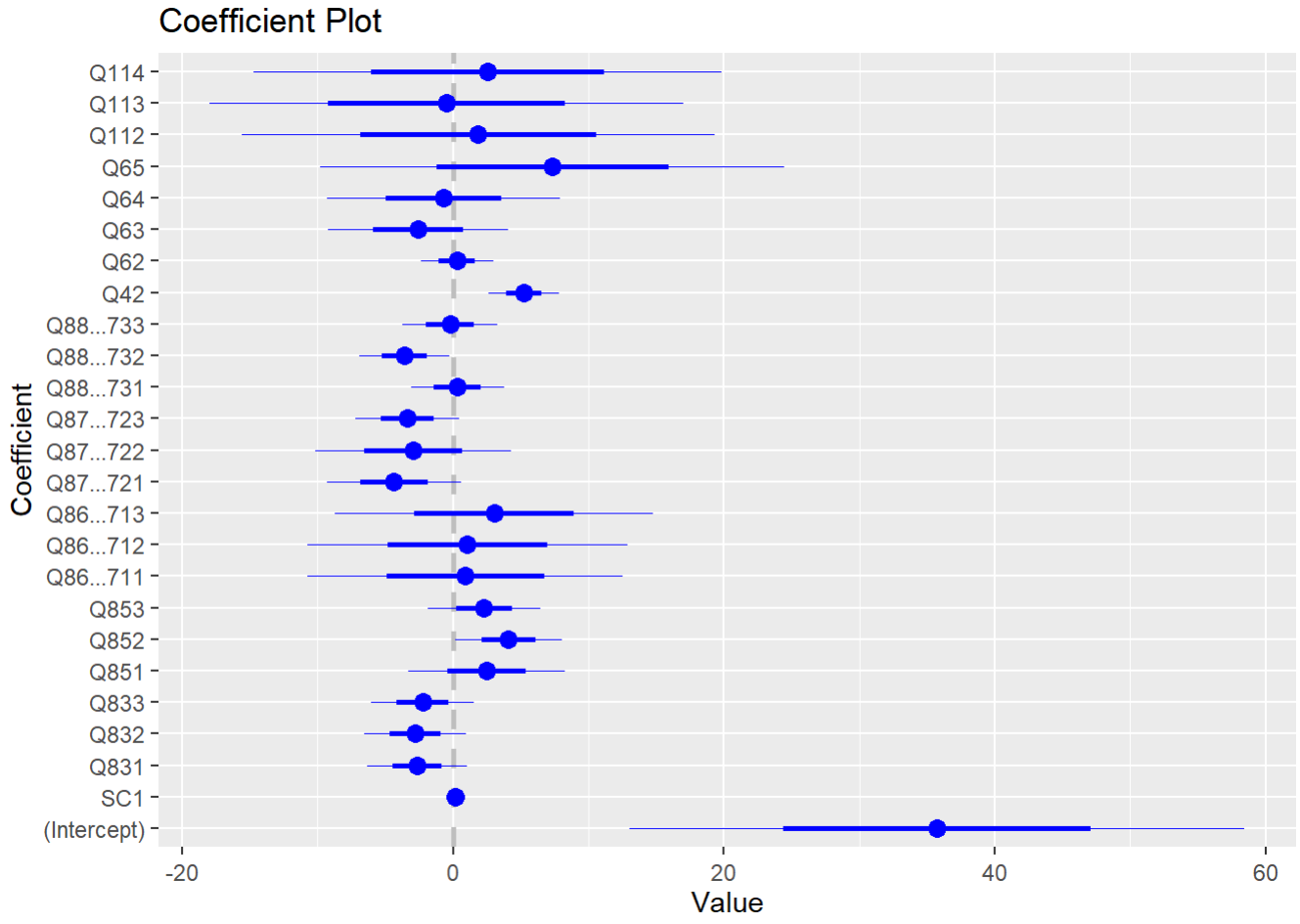
We first use the linear model to find relationships among all the variables with these two formulas:

empathy ~ stress + self_care_variables + Current Student Year + Age + Marrage situation

stress ~ empathy + self_care_variables + Current Student Year + Age + Marrage situation

And in a coefficient plot, we are able to see how these variables affect each other:


Coefficient Plot
Q114
Q113
Q112
Q65
Q64
Q63
Q62
Q42
Q88...733
Q88...732
Q88...731
Q87...723
Q87...722
Q87...721
Q86...713
Q86...712
Q86...711
Q853
Q852
Q851
Q833
Q832
Q831
SC0
(Intercept)
Coefficient
0
30
60
90
Value


Coefficient Plot
Q114
Q113
Q112
Q65
Q64
Q63
Q62
Q42
Q88...733
Q88...732
Q88...731
Q87...723
Q87...722
Q87...721
Q86...713
Q86...712
Q86...711
Q853
Q852
Q851
Q833
Q832
Q831
SC1
(Intercept)
Coefficient
-20
0
20
40
60
Value

In the graph, we find some variables that worth analyzing. For example, in Q4, second year student have more stress increase comparing to first year students; in Q11, people in a romantic relationship will have lower stress than other situations.

Discussion

For this part of the whole project, we are able to find some conclusions but there are still some problems.
1. More data has been collected since this analysis was run. A future analysis may be useful to see if additional data changes the results we see here.
2. As is common in survey data, some participants did not finish the survey, some participants do not have reliable answers, and some questions may not make the relationship be easy to see. Especially since it is conceivable that stress is related to whether or not someone finishes the survey, this could be a source of bias and a limitation of the study.

CONCLUSION

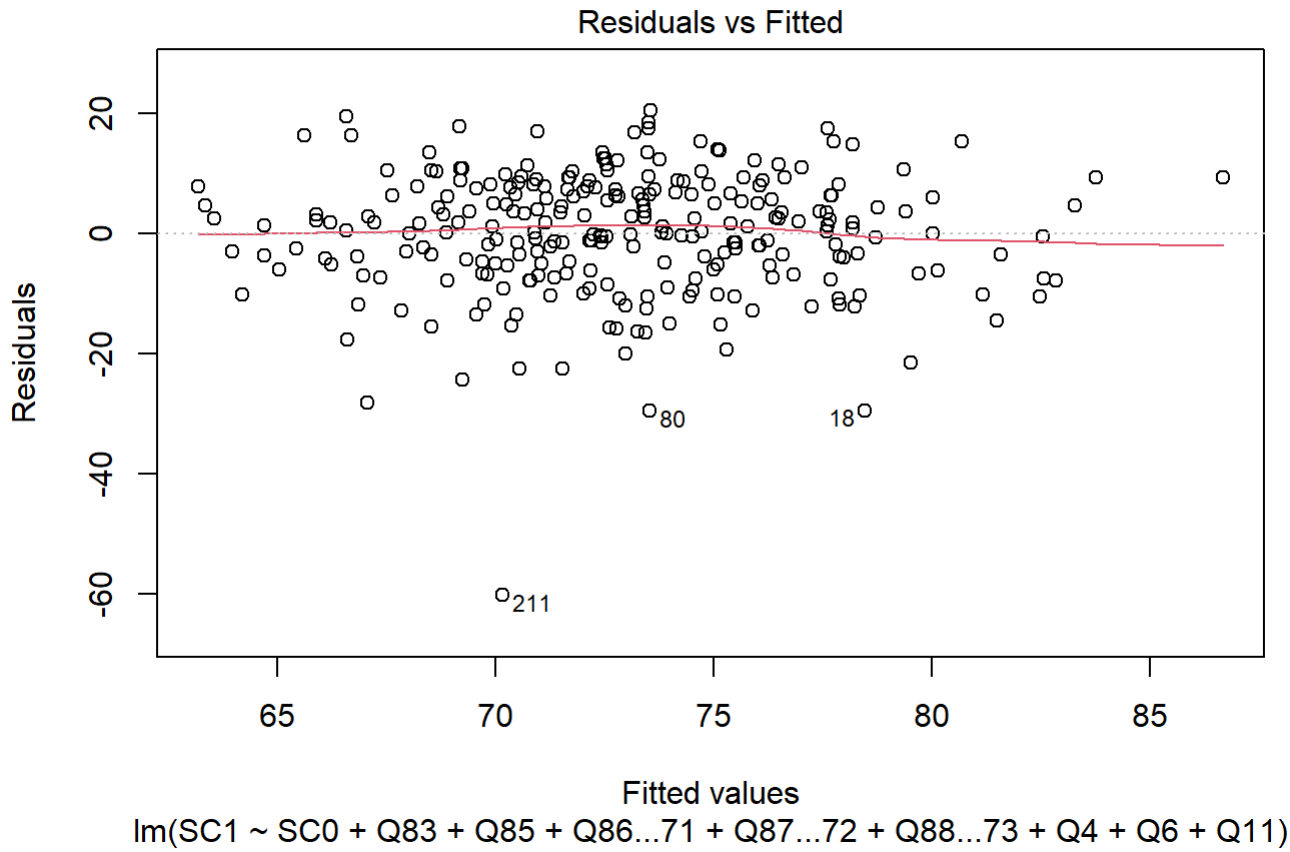
The characterization for each type of question can be seen in the EDA and summary statistics. For the relationship among all the variables, we can see relationships in the visualizations between higher stress and lower empathy, although the relationship is not very strong. Our models suggest that there is not enough evidence to say that there is a relationship between stress and empathy or self care. While for self-care, we should depend on specific self-care problems. Also, some specific covariance questions also make sense for the stress/empathy relationships.

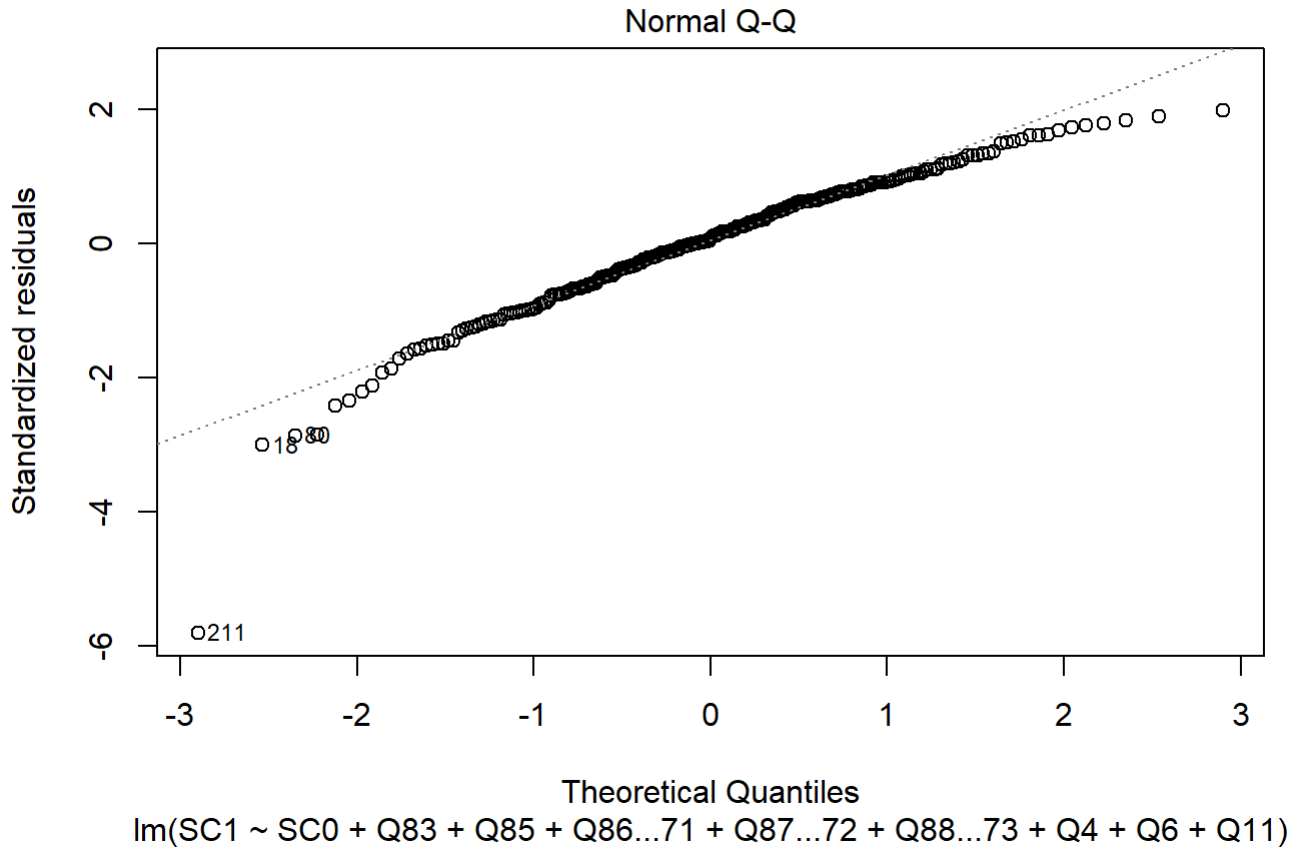
Appendix

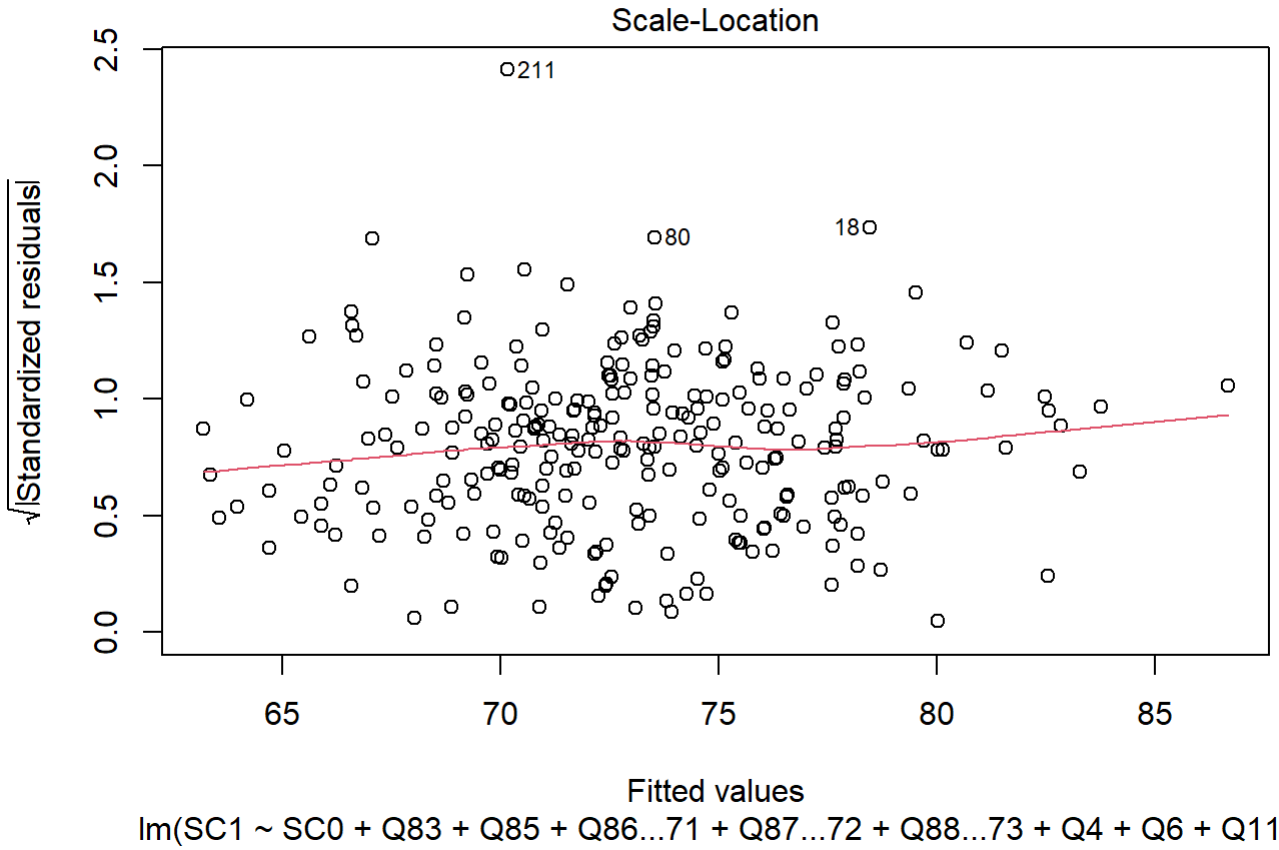
This part is for residual plots. The first 4 are the empathy ~ stress + self_care_variables + Current Student Year + Age + Marrage situation formula and the rest are for the stress ~ empathy + self_care_variables + Current Student Year + Age + Marrage situation. We can also analyze these residual plots.

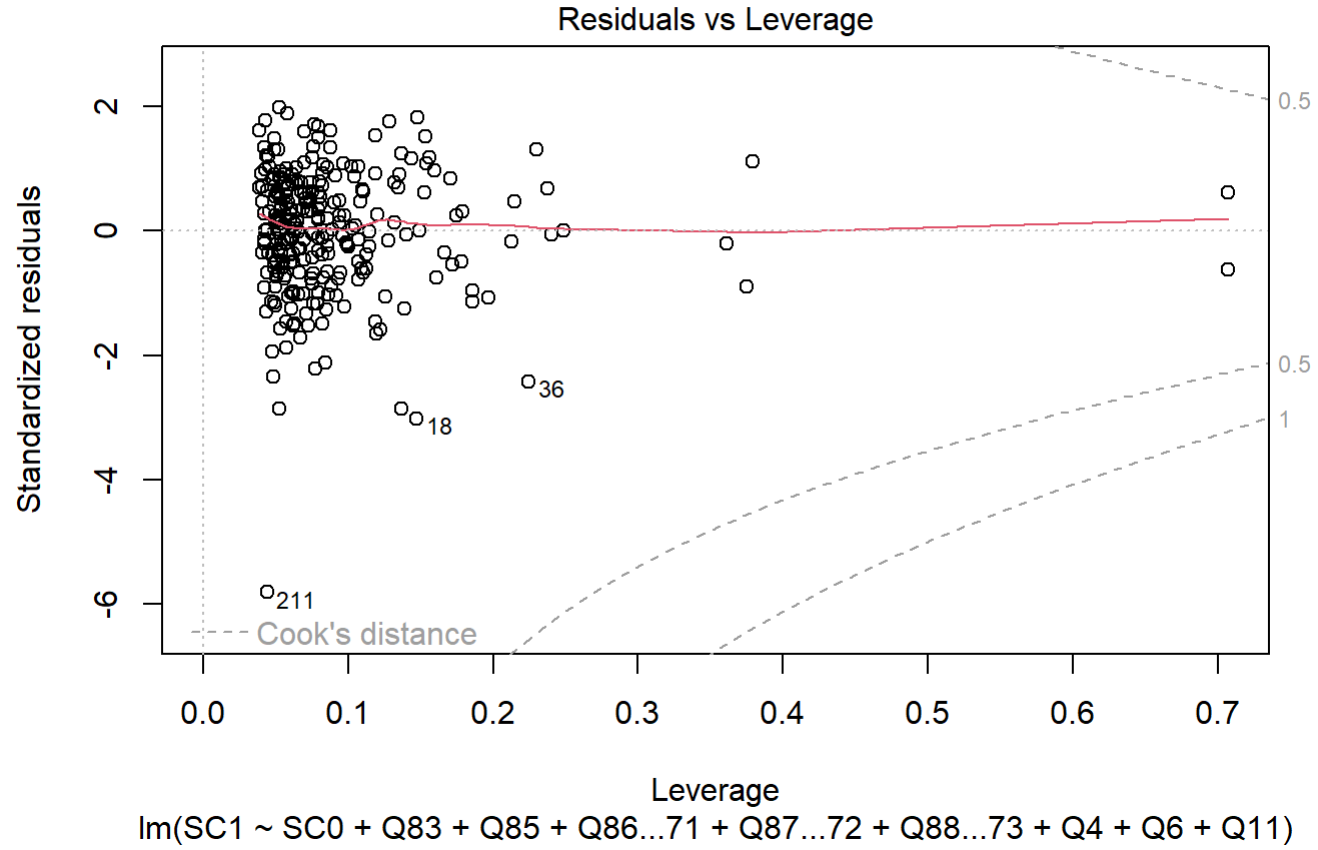
```
##
## Call:
## lm(formula = SC1 ~ SC0 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 +
##     Q4 + Q6 + Q11, data = data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -60.176  -6.074   0.610   7.264  20.438
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 63.96717   12.04165   5.312 2.45e-07 ***
## SC0          0.23890    0.06875   3.475 0.000605 ***
## Q831         2.82453    2.00816   1.407 0.160846
## Q832        -0.27082    2.08086  -0.130 0.896555
## Q833        -1.11282    2.08946  -0.533 0.594804
## Q851        -2.10091    3.17076  -0.663 0.508221
## Q852        -1.39805    2.18780  -0.639 0.523413
## Q853         1.61920    2.27007   0.713 0.476356
## Q86...711   -7.66371    6.38309  -1.201 0.231065
## Q86...712   -8.15093    6.45900  -1.262 0.208177
## Q86...713   -5.37140    6.43583  -0.835 0.404757
## Q87...721    7.23984    2.69754   2.684 0.007778 **
## Q87...722    7.14847    3.94641   1.811 0.071316 .
## Q87...723    4.99996    2.09314   2.389 0.017670 *
## Q88...731   -0.51955    1.89338  -0.274 0.784006
## Q88...732    1.53366    1.84513   0.831 0.406681
## Q88...733   -1.88784    1.91154  -0.988 0.324329
## Q42         -2.93375    1.45915  -2.011 0.045474 *
## Q62          1.42384    1.46504   0.972 0.332078
## Q63          4.24073    3.65537   1.160 0.247131
## Q64          0.60931    4.70783   0.129 0.897129
## Q65          3.99843    9.41426   0.425 0.671416
## Q112         3.12212    9.58321   0.326 0.744863
## Q113         2.18155    9.59353   0.227 0.820306
## Q114         1.54245    9.48026   0.163 0.870889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.58 on 243 degrees of freedom
## Multiple R-squared:  0.1449, Adjusted R-squared:  0.06046
## F-statistic: 1.716 on 24 and 243 DF,  p-value: 0.02288
```

```
## (Intercept)         SC0        Q831        Q832        Q833        Q851
##  63.9671690   0.2389006   2.8245306  -0.2708237  -1.1128241  -2.1009120
##        Q852        Q853   Q86...711   Q86...712   Q86...713   Q87...721
##  -1.3980454   1.6191984  -7.6637072  -8.1509281  -5.3714017   7.2398368
##   Q87...722   Q87...723   Q88...731   Q88...732   Q88...733         Q42
##   7.1484741   4.9999639  -0.5195546   1.5336590  -1.8878438  -2.9337523
##         Q62         Q63         Q64         Q65        Q112        Q113
##   1.4238436   4.2407310   0.6093095   3.9984339   3.1221158   2.1815451
##        Q114
##   1.5424450
```


Residuals vs Fitted
Residuals
20
0
-20
-40
-60
80
18
211
65
70
75
80
85
Fitted values
lm(SC1 ~ SC0 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)

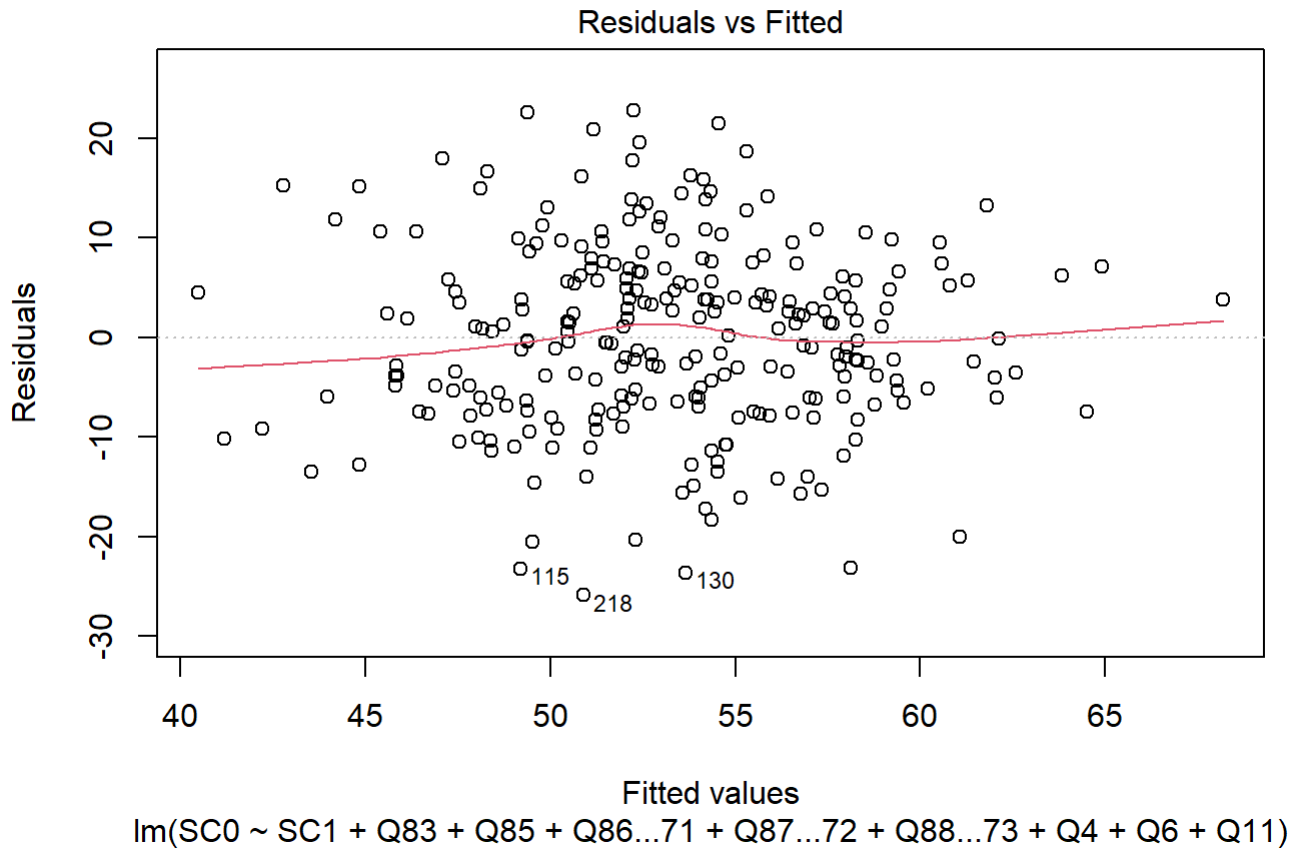

Normal Q-Q
Standardized residuals
2
0
-2
-4
-6
18
80
211
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(SC1 ~ SC0 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)

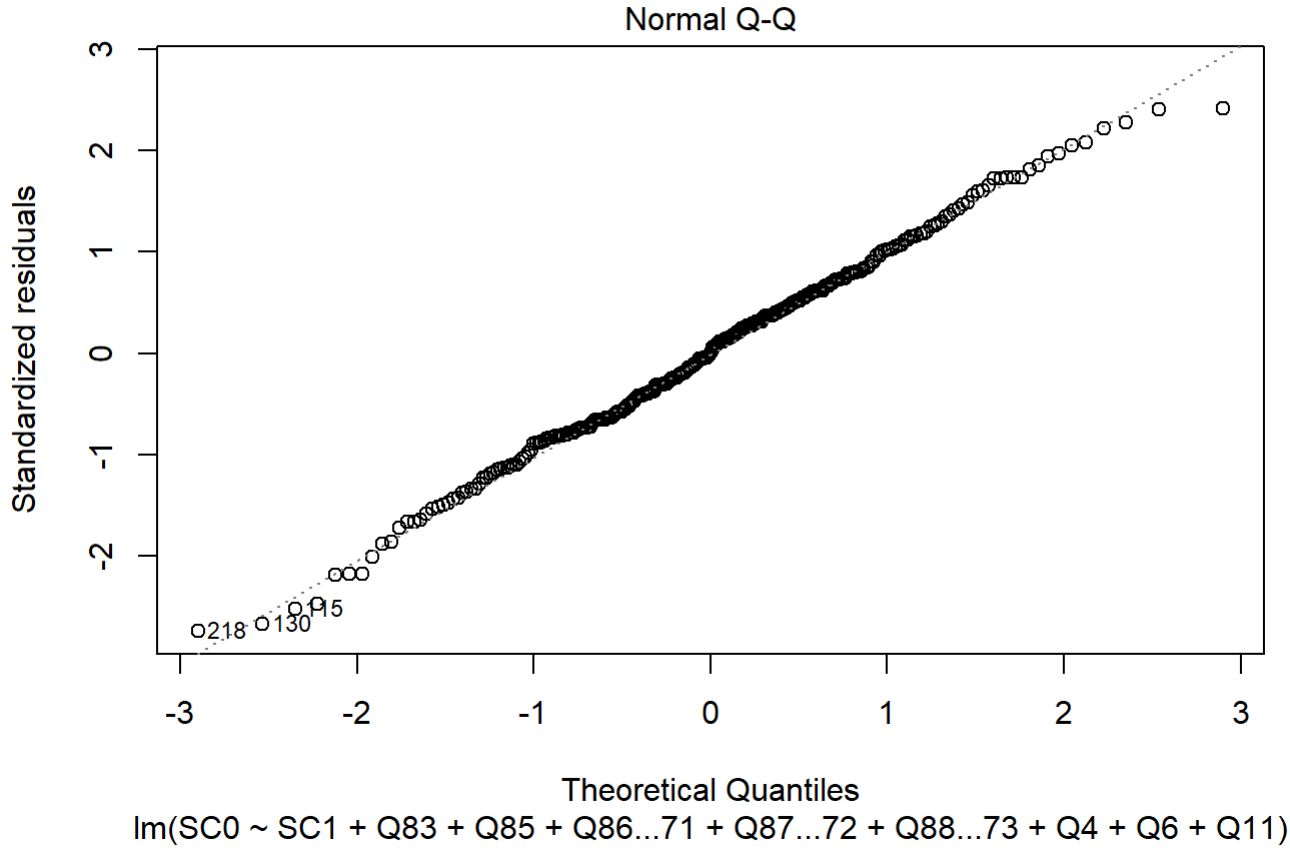

Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
211
80
18
65
70
75
80
85
Fitted values
lm(SC1 ~ SC0 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)

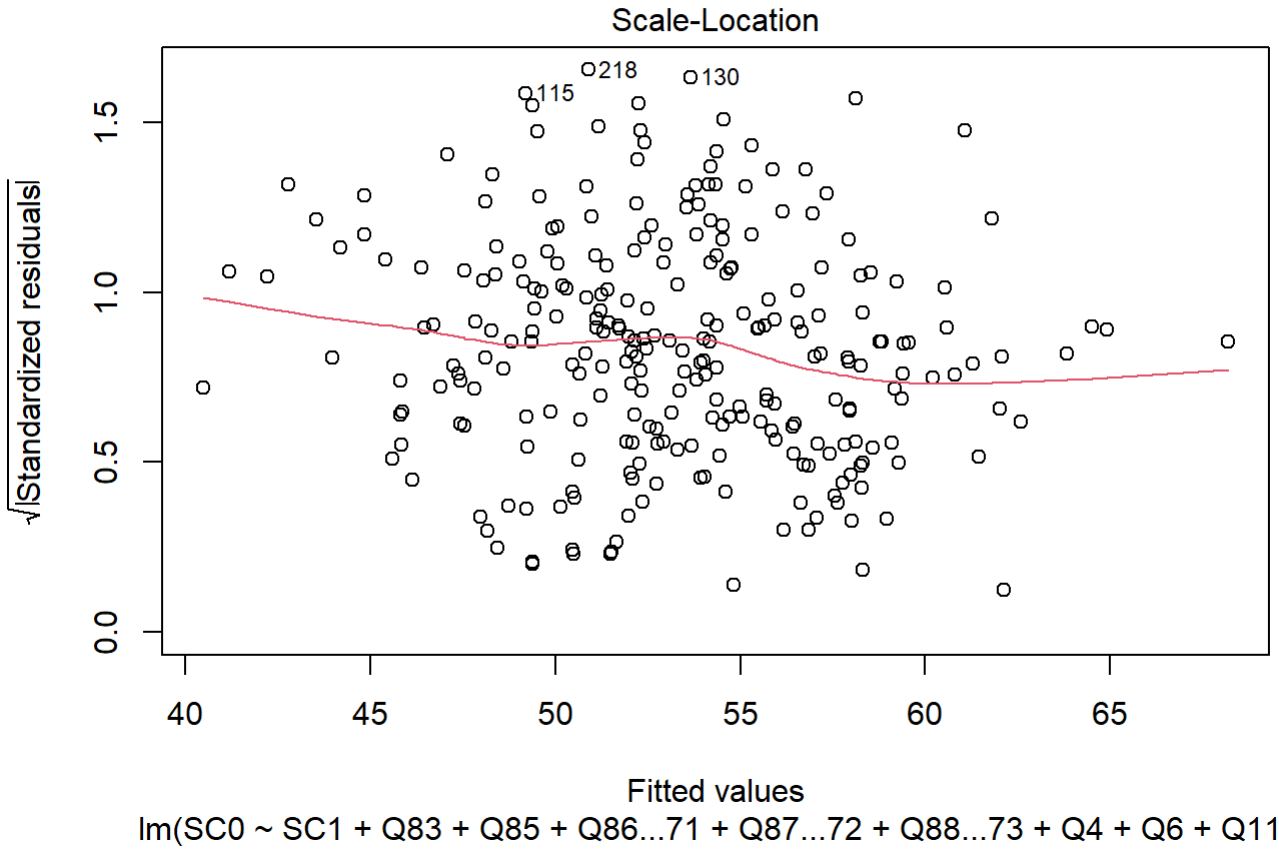

Residuals vs Leverage
Standardized residuals
2
0
-2
-4
-6
0.5
0.5
1
36
18
211
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Leverage
lm(SC1 ~ SC0 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)

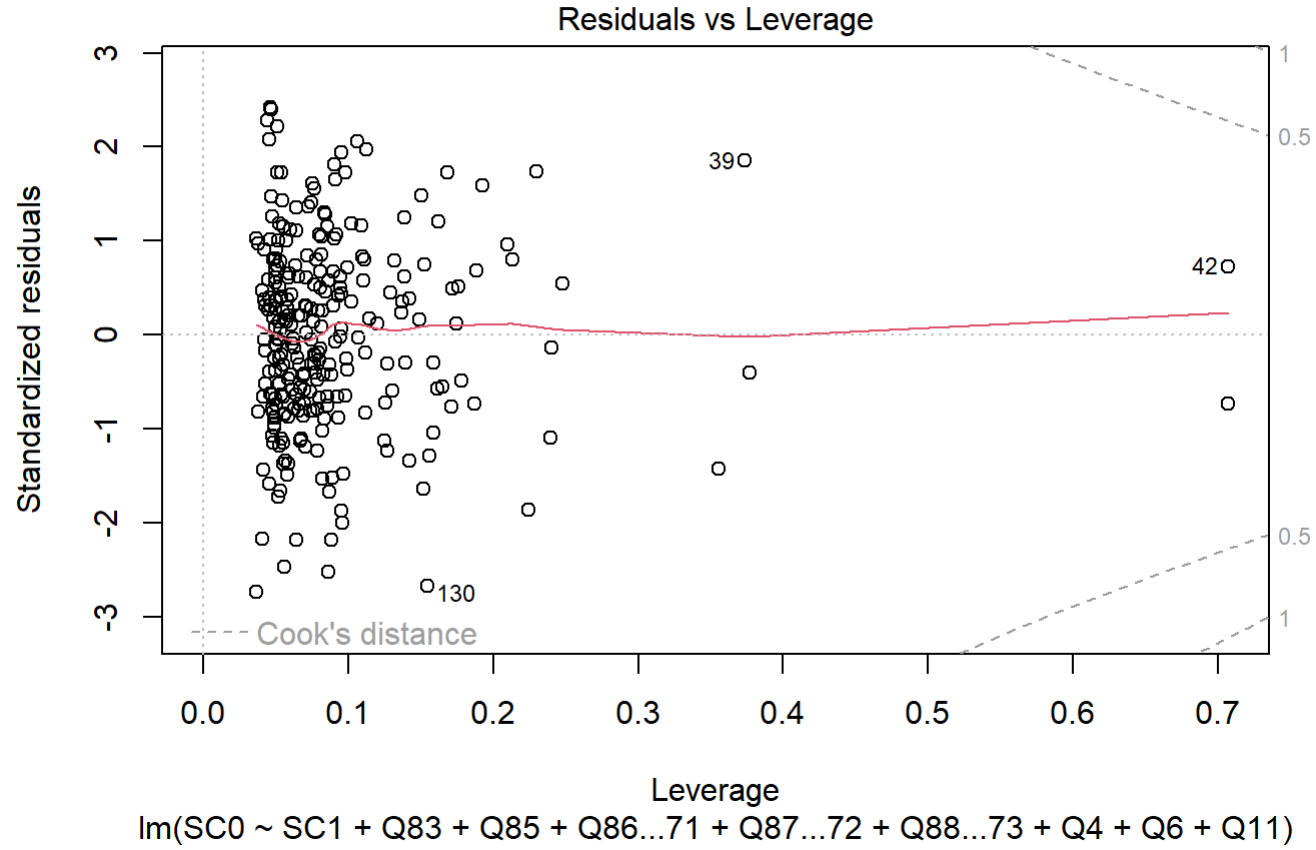
```
##
## Call:
## lm(formula = SC0 ~ SC1 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 +
##     Q4 + Q6 + Q11, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.9113  -6.2425   0.0179   5.9700  22.7574
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.73025   11.35666   3.146 0.001860 **
## SC1          0.19814    0.05702   3.475 0.000605 ***
## Q831        -2.64638    1.82841  -1.447 0.149082
## Q832        -2.78394    1.88668  -1.476 0.141352
## Q833        -2.24479    1.89854  -1.182 0.238211
## Q851         2.50785    2.88575   0.869 0.385680
## Q852         4.10184    1.97668   2.075 0.039028 *
## Q853         2.30578    2.06424   1.117 0.265092
## Q86...711    0.90983    5.83003   0.156 0.876116
## Q86...712    1.07121    5.90109   0.182 0.856105
## Q86...713    3.03360    5.86631   0.517 0.605541
## Q87...721   -4.35952    2.47707  -1.760 0.079674 .
## Q87...722   -2.96547    3.61319  -0.821 0.412603
## Q87...723   -3.36677    1.91635  -1.757 0.080201 .
## Q88...731    0.33524    1.72444   0.194 0.846021
## Q88...732   -3.61140    1.66674  -2.167 0.031226 *
## Q88...733   -0.21586    1.74428  -0.124 0.901614
## Q42          5.21933    1.29736   4.023 7.68e-05 ***
## Q62          0.31128    1.33666   0.233 0.816050
## Q63         -2.56958    3.33410  -0.771 0.441635
## Q64         -0.68748    4.28737  -0.160 0.872739
## Q65          7.34688    8.56383   0.858 0.391795
## Q112         1.85899    8.72857   0.213 0.831523
## Q113        -0.46517    8.73775  -0.053 0.957587
## Q114         2.56087    8.63263   0.297 0.766987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.638 on 243 degrees of freedom
## Multiple R-squared:  0.1925, Adjusted R-squared:  0.1128
## F-statistic: 2.414 on 24 and 243 DF,  p-value: 0.0003933
```

```
## (Intercept)         SC1        Q831        Q832        Q833        Q851
##  35.7302483   0.1981402  -2.6463836  -2.7839407  -2.2447924   2.5078491
##        Q852        Q853   Q86...711   Q86...712   Q86...713   Q87...721
##   4.1018405   2.3057792   0.9098304   1.0712076   3.0336019  -4.3595242
##   Q87...722   Q87...723   Q88...731   Q88...732   Q88...733         Q42
##  -2.9654693  -3.3667676   0.3352397  -3.6114038  -0.2158578   5.2193310
##         Q62         Q63         Q64         Q65        Q112        Q113
##   0.3112839  -2.5695781  -0.6874782   7.3468795   1.8589913  -0.4651668
##        Q114
##   2.5608668
```


Residuals vs Fitted
Residuals
20
10
0
-10
-20
-30
115
218
130
40
45
50
55
60
65
Fitted values
lm(SC0 ~ SC1 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)


Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
218
130
115
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(SC0 ~ SC1 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)


Scale-Location
√|Standardized residuals|
218
115
130
0.0
0.5
1.0
1.5
40
45
50
55
60
65
Fitted values
lm(SC0 ~ SC1 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)


Residuals vs Leverage
Standardized residuals
39
42
130
Cook's distance
1
0.5
0.5
1
-3
-2
-1
0
1
2
3
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Leverage
lm(SC0 ~ SC1 + Q83 + Q85 + Q86...71 + Q87...72 + Q88...73 + Q4 + Q6 + Q11)