Multiple Regression Exercises

Advanced Psychological Research Methods

A clinical psychologist is interested in the effects of a treatment plan on the general wellbeing of clients. She is also interested in whether the level of severity of symptoms when entering treatment or the clients' trust in the psychologist predict wellbeing.

The severity_data.csv dataset contains 5 variables:

- client: an anonymous ID of the client
- treatment_group (control or therapy): Which treatment the client has been assigned to
- level_of_severity (1-20): A rating assigned by clinicians as to the severity of the client's difficulties when they entered treatment (a computed value based on a battery of tests)
- trust_score (0-100): A rating by the client as to the level of trust they have in their psychologist based on their interaction to this point
- wellbeing_after_3_months (0-10): Client's score on a psychometric measure of general wellbeing

To answer the research questions outlined above, conduct the following analyses:

```
library(readr)
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.6.1

## -- Attaching packages ---------------------------------------------------
---------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   0.8.3     v stringr 1.4.0
## v ggplot2 3.2.1     v forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.6.1

## Warning: package 'tibble' was built under R version 3.6.1

## Warning: package 'dplyr' was built under R version 3.6.1

## -- Conflicts ------------------------------------------------------------
---------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

severity_data <- read_csv("Datasets/severity_data.csv")

## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   client = col_double(),
##   treatment_group = col_character(),
##   level_of_severity = col_double(),
##   trust_score = col_double(),
##   wellbeing_after_3_months = col_double()
## )

severity_data$treatment_group <- as.factor(severity_data$treatment_group)
```

1. Calculate if there is a difference in the means of *severity level*, *trust score* and *wellbeing after 3 months* between the 2 treatment groups (Hint: use the describeBy function in the Psych package)

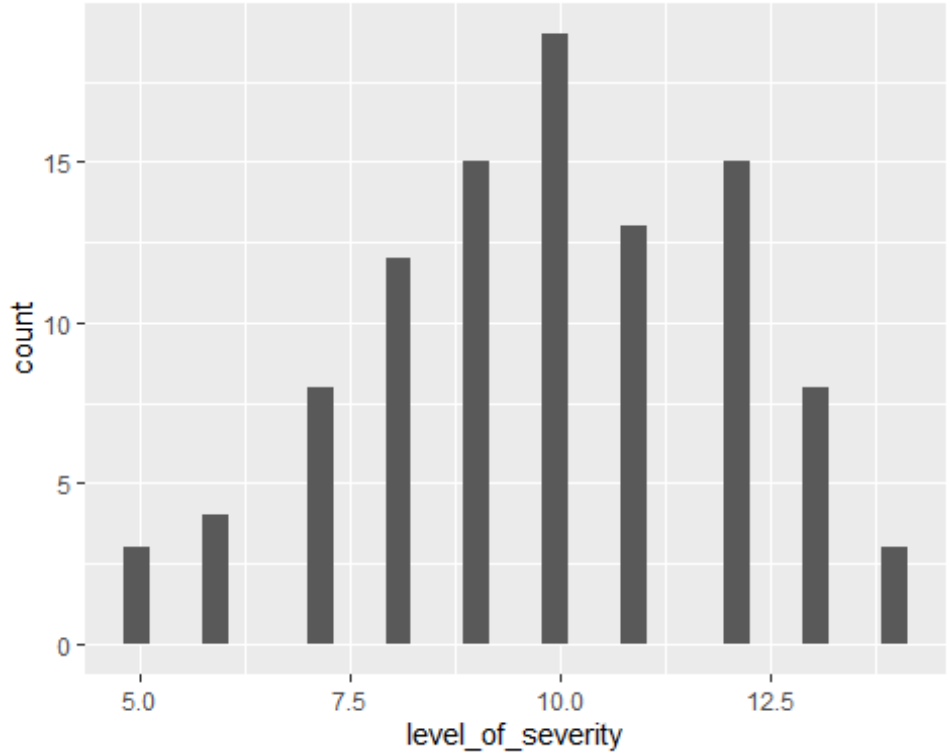
```
library(psych)

## Warning: package 'psych' was built under R version 3.6.1

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

describeBy(severity_data, group = severity_data$treatment_group)

##
##  Descriptive statistics by group
## group: control
##                           vars  n  mean    sd median trimmed   mad min max
## X1                           1 47 50.40 28.45     49   49.82 34.10   3  99
## client                       2 47 50.40 28.45     49   49.82 34.10   3  99
## treatment_group*             3 47  1.00  0.00      1    1.00  0.00   1   1
## level_of_severity            4 47 10.13  2.19     10   10.18  2.97   5  14
## trust_score                  5 47 51.15 28.56     53   51.21 35.58   2  99
## wellbeing_after_3_months     6 47  4.83  2.47      5    4.85  2.97   0  10
##                           range  skew kurtosis   se
## X1                           96  0.23    -1.19 4.15
## client                       96  0.23    -1.19 4.15
## treatment_group*              0   NaN      NaN 0.00
## level_of_severity             9 -0.21    -0.80 0.32
## trust_score                  97 -0.06    -1.02 4.17
## wellbeing_after_3_months     10  0.09    -0.98 0.36
## ------------------------------------------------------------
## group: therapy
##                           vars  n  mean    sd median trimmed   mad min max
## X1                           1 53 50.58 29.77     53   51.14 37.06   1 100
## client                       2 53 50.58 29.77     53   51.14 37.06   1 100
## treatment_group*             3 53  2.00  0.00      2    2.00  0.00   2   2
```

```
## level_of_severity           4 53  9.60  2.13     10    9.67  1.48   5  14
## trust_score                 5 53 48.91 27.98     44   48.51 32.62   2 100
## wellbeing_after_3_months    6 53  5.04  2.41      4    5.00  2.97   0  10
##                           range  skew kurtosis   se
## X1                           99 -0.18    -1.34 4.09
## client                       99 -0.18    -1.34 4.09
## treatment_group*              0   NaN      NaN 0.00
## level_of_severity             9 -0.22    -0.58 0.29
## trust_score                  98  0.12    -1.22 3.84
## wellbeing_after_3_months     10  0.15    -0.80 0.33
```

2. Make a histogram for each of the variables in the previous question and assess the distribution (Hint: Use the ggplot2 package to make plots)

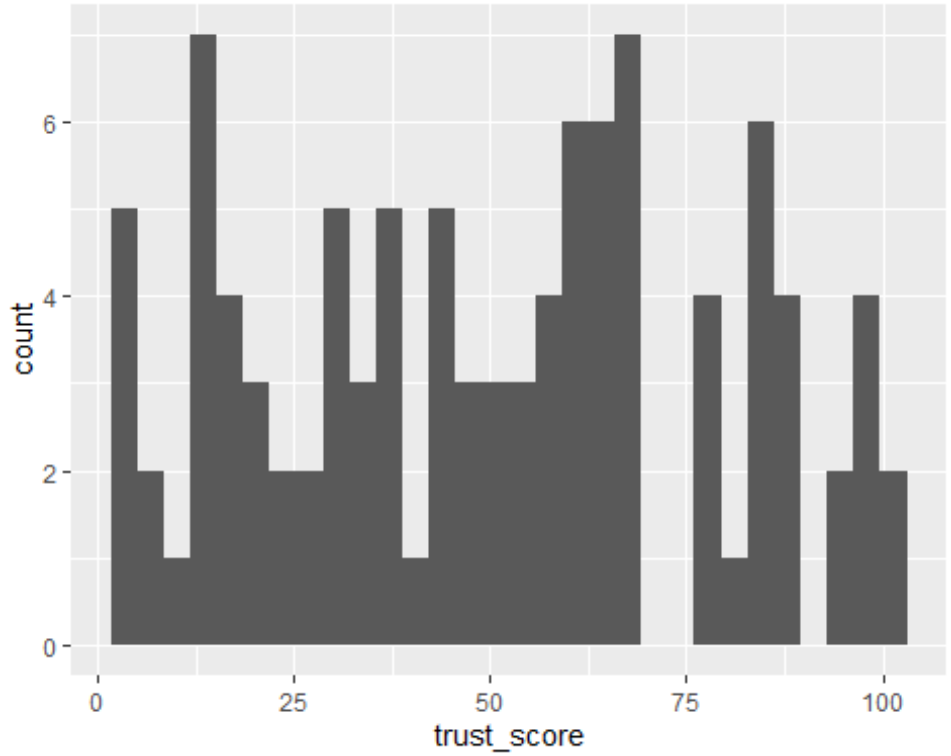
```
severity_data %>% ggplot(aes(x=level_of_severity)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



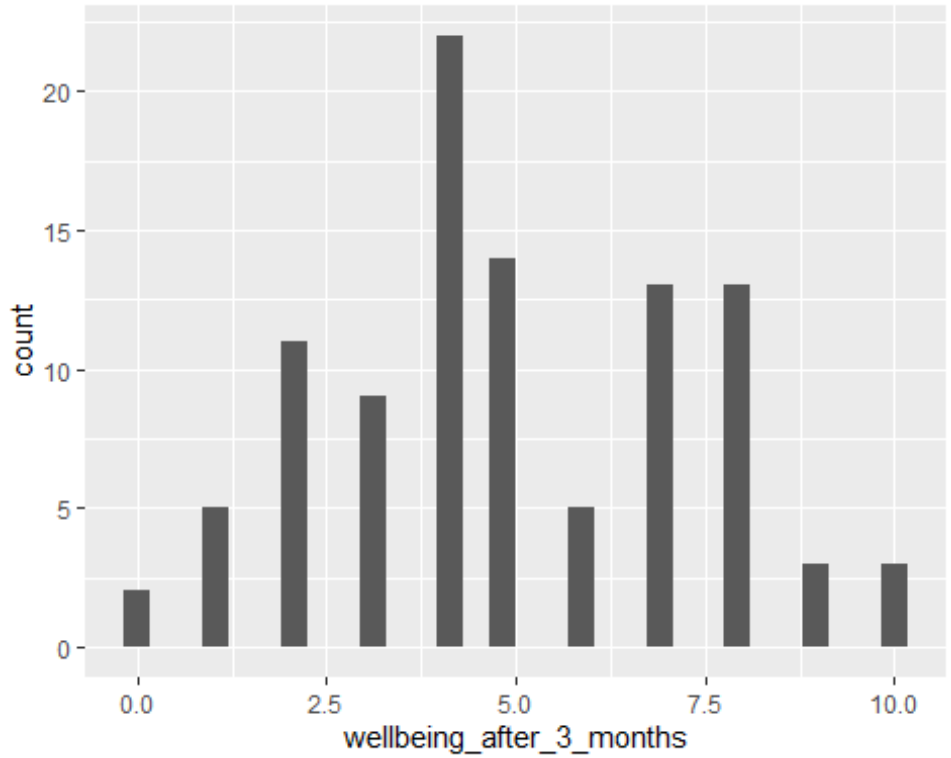

```
severity_data %>% ggplot(aes(x=trust_score)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
severity_data %>% ggplot(aes(x=wellbeing_after_3_months)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



3. Run a heirarchical multiple regression analysis comparing the following models (in each case, *wellbeing after 3 months* is the outcome). At this stage, check main effects only, not interactions:

- Model 0: Constant
- Model 1: treatment_group, severity_level
- Model 2: treatment_group, severity_level, trust_score

```
model0 <- lm(wellbeing_after_3_months ~ 1, data = severity_data)
model1 <- lm(wellbeing_after_3_months ~ treatment_group + level_of_severity,
data = severity_data)
model2 <- lm(wellbeing_after_3_months ~ treatment_group + level_of_severity +
trust_score, data = severity_data)

anova(model0,model1,model2)

## Analysis of Variance Table
##
## Model 1: wellbeing_after_3_months ~ 1
## Model 2: wellbeing_after_3_months ~ treatment_group + level_of_severity
## Model 3: wellbeing_after_3_months ~ treatment_group + level_of_severity +
##     trust_score
##   Res.Df    RSS Df Sum of Sq       F Pr(>F)
## 1     99 583.64
## 2     97 251.79  2    331.85 63.2753 <2e-16 ***
## 3     96 251.74  1      0.05  0.0188 0.8911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

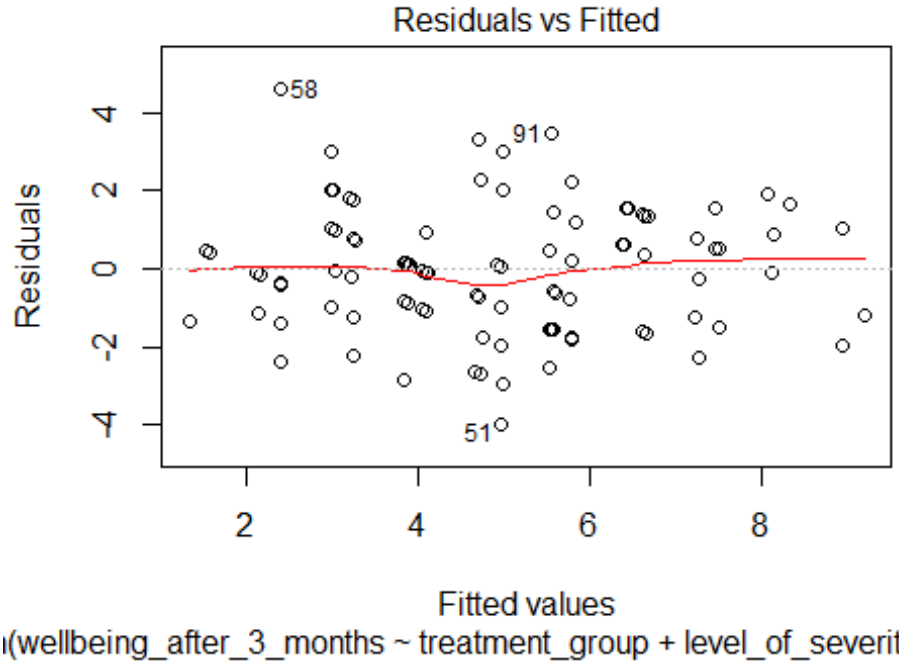
4. Check the assumption of multicoliniearity in Model 2 (Hint: Install and load the *mctest* package)

```
#install.packages("mctest")
library(mctest)

## Warning: package 'mctest' was built under R version 3.6.1

mctest(cbind(severity_data$level_of_severity,severity_data$trust_score),sever
ity_data$wellbeing_after_3_months )

##
## Call:
## omcdiag(x = x, y = y, Inter = TRUE, detr = detr, red = red, conf = conf,
##     theil = theil, cn = cn)
##
##
## Overall Multicollinearity Diagnostics
##
##                        MC Results detection
## Determinant |X'X|:         0.9697         0
## Farrar Chi-Square:         2.9995         0
## Red Indicator:             0.1741         0
```

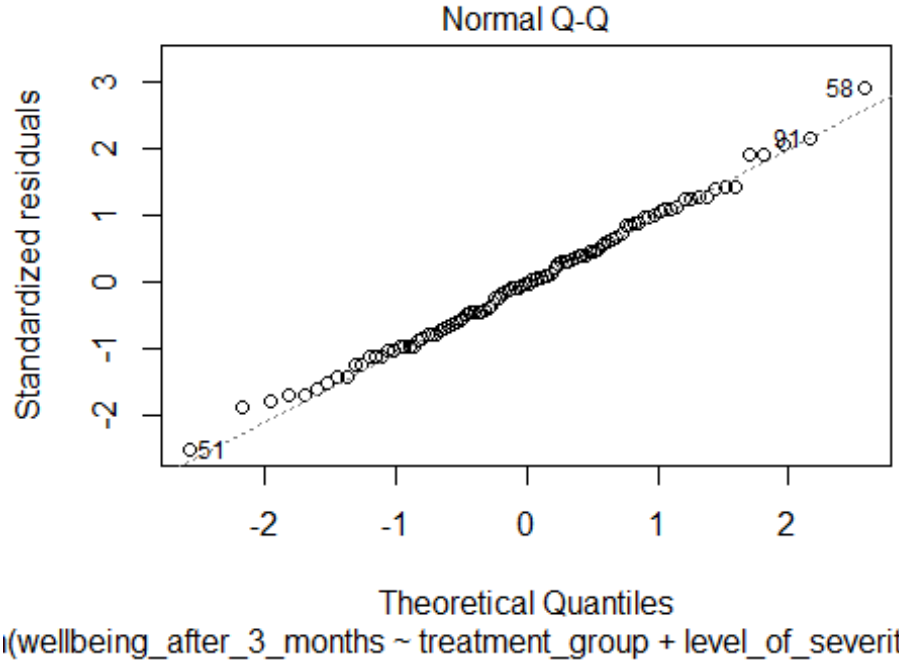
```
## Sum of Lambda Inverse:     2.0625         0
## Theil's Method:           -0.5057         0
## Condition Number:         11.0340         0
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
```

5. Use the plot() function to check other assumptions of Model 2

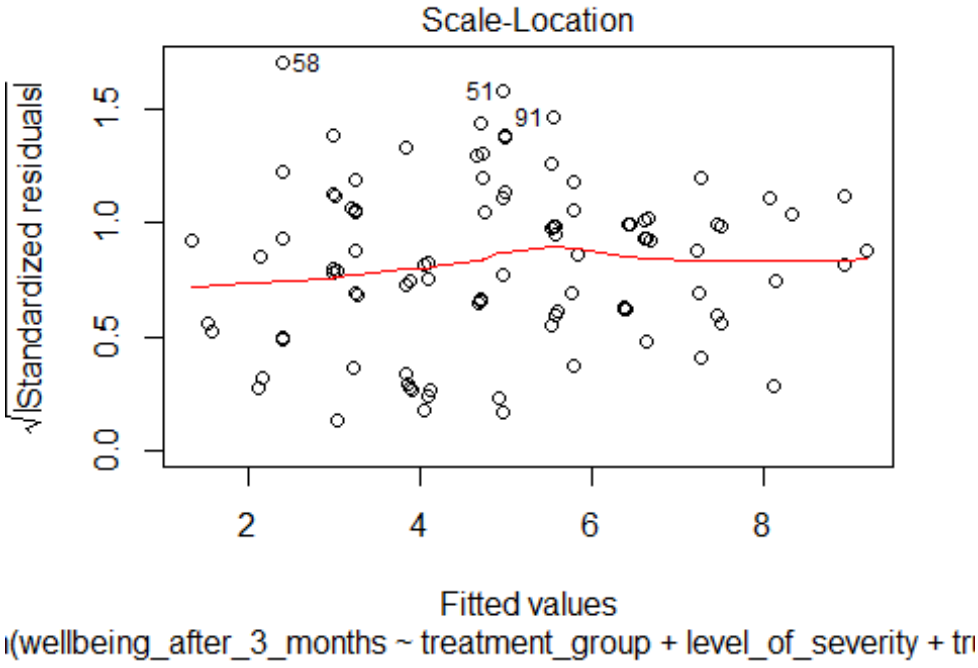
```
plot(model2)
```

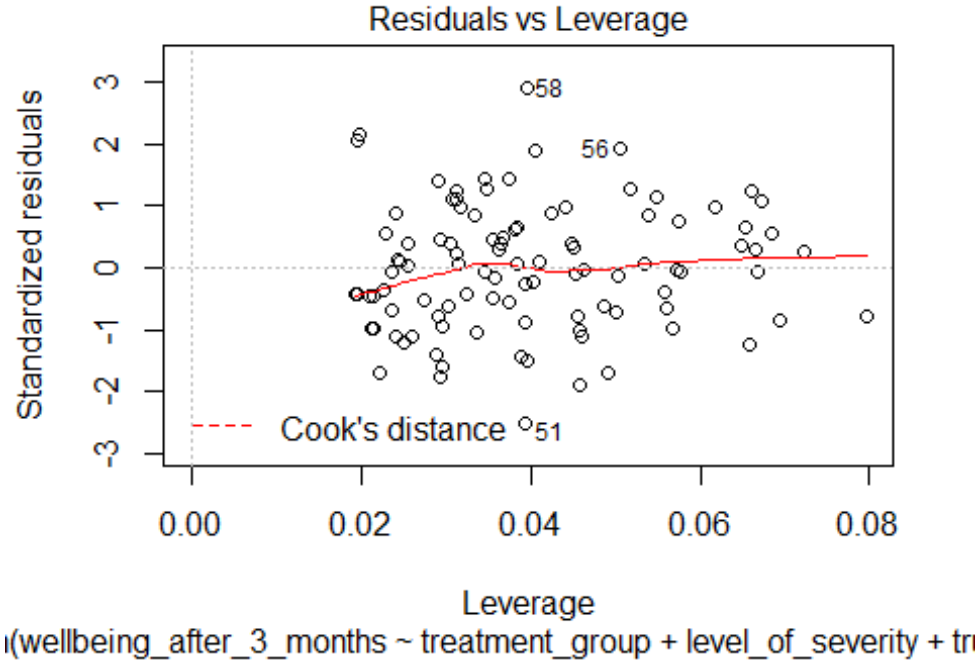


ı(wellbeing_after_3_months ~ treatment_group + level_of_severity + tr



ı(wellbeing_after_3_months ~ treatment_group + level_of_severity + tr





6. Run Model 2 again, including interactions this time.

```
model2 <- lm(wellbeing_after_3_months ~ treatment_group * level_of_severity *
trust_score, data = severity_data)
summary(model2)
```

```
##
## Call:
## lm(formula = wellbeing_after_3_months ~ treatment_group *
level_of_severity *
##     trust_score, data = severity_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7560 -1.0252 -0.0749  1.0017  4.4514
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                             15.160233   2.275438
## treatment_grouptherapy                                  -2.128932   3.096236
## level_of_severity                                       -1.009046   0.232981
## trust_score                                             -0.042651   0.043760
## treatment_grouptherapy:level_of_severity                 0.150063   0.319365
## treatment_grouptherapy:trust_score                       0.050896   0.059523
## level_of_severity:trust_score                            0.003909   0.004354
## treatment_grouptherapy:level_of_severity:trust_score -0.004216   0.005968
##                                                         t value Pr(>|t|)
## (Intercept)                                               6.663 1.94e-09 ***
## treatment_grouptherapy                                   -0.688    0.493
## level_of_severity                                        -4.331 3.78e-05 ***
## trust_score                                              -0.975    0.332
## treatment_grouptherapy:level_of_severity                  0.470    0.640
## treatment_grouptherapy:trust_score                        0.855    0.395
## level_of_severity:trust_score                             0.898    0.372
## treatment_grouptherapy:level_of_severity:trust_score  -0.707    0.482
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.641 on 92 degrees of freedom
## Multiple R-squared:  0.5754, Adjusted R-squared:  0.5431
## F-statistic: 17.81 on 7 and 92 DF,  p-value: 9.521e-15
```

7. What conclusions can you draw from these analyses in relation to your research question?

```
## The model explains 57.5% of the variance in wellbeing and the model is
significant
## Level of severity is a significant predictor of wellbeing after 3 months.
The other predictors are not.
## As level of severity increases by 1 unit, wellbeing decreases by 1.009
```

8. Run Model 2 as a stepwise analysis (Hint: Install and load the *MASS* package first)

```
library(MASS)

##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select

stepwiseModel <- stepAIC(model2, direction = "both", trace = TRUE)

## Start:  AIC=106.74
## wellbeing_after_3_months ~ treatment_group * level_of_severity *
##     trust_score
##
##                                                 Df Sum of Sq    RSS    AIC
## - treatment_group:level_of_severity:trust_score  1    1.3445 249.14 105.29
## <none>                                                      247.80 106.75
##
## Step:  AIC=105.29
## wellbeing_after_3_months ~ treatment_group + level_of_severity +
##     trust_score + treatment_group:level_of_severity +
treatment_group:trust_score +
##     level_of_severity:trust_score
##
##                                                 Df Sum of Sq    RSS    AIC
## - treatment_group:level_of_severity              1   0.24387 249.39 103.38
## - level_of_severity:trust_score                  1   0.84138 249.99 103.62
## - treatment_group:trust_score                    1   1.77716 250.92 104.00
## <none>                                                      249.14 105.29
## + treatment_group:level_of_severity:trust_score  1   1.34450 247.80 106.75
##
## Step:  AIC=103.38
## wellbeing_after_3_months ~ treatment_group + level_of_severity +
##     trust_score + treatment_group:trust_score +
level_of_severity:trust_score
##
##                                      Df Sum of Sq    RSS    AIC
## - level_of_severity:trust_score       1   0.84732 250.23 101.72
## - treatment_group:trust_score         1   1.60561 250.99 102.03
## <none>                                            249.39 103.38
## + treatment_group:level_of_severity   1   0.24387 249.14 105.29
##
## Step:  AIC=101.72
## wellbeing_after_3_months ~ treatment_group + level_of_severity +
##     trust_score + treatment_group:trust_score
##
##                                      Df Sum of Sq    RSS    AIC
## - treatment_group:trust_score         1      1.50 251.74 100.32
## <none>                                            250.23 101.72
## + level_of_severity:trust_score       1      0.85 249.39 103.38
## + treatment_group:level_of_severity   1      0.25 249.98 103.62
## - level_of_severity                   1    321.49 571.72 182.35
##
## Step:  AIC=100.32
```

```
## wellbeing_after_3_months ~ treatment_group + level_of_severity +
##     trust_score
##
##                                      Df Sum of Sq    RSS     AIC
## - trust_score                         1      0.05 251.79  98.342
## - treatment_group                     1      1.37 253.11  98.866
## <none>                                             251.74 100.322
## + treatment_group:trust_score         1      1.50 250.23 101.723
## + level_of_severity:trust_score       1      0.75 250.99 102.025
## + treatment_group:level_of_severity   1      0.08 251.66 102.291
## - level_of_severity                   1    322.51 574.25 180.789
##
## Step:  AIC=98.34
## wellbeing_after_3_months ~ treatment_group + level_of_severity
##
##                                      Df Sum of Sq    RSS     AIC
## - treatment_group                     1      1.38 253.17  96.889
## <none>                                             251.79  98.342
## + treatment_group:level_of_severity   1      0.08 251.71 100.310
## + trust_score                         1      0.05 251.74 100.322
## - level_of_severity                   1    330.77 582.56 180.227
##
## Step:  AIC=96.89
## wellbeing_after_3_months ~ level_of_severity
##
##                      Df Sum of Sq    RSS     AIC
## <none>                             253.17  96.889
## + treatment_group     1      1.38 251.79  98.342
## + trust_score         1      0.06 253.11  98.866
## - level_of_severity   1    330.47 583.64 178.411
```

```
summary(stepwiseModel)
```

```
##
## Call:
## lm(formula = wellbeing_after_3_months ~ level_of_severity, data =
severity_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8135 -1.1270  0.0297  0.9851  4.7162
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       13.24600    0.75176   17.62   <2e-16 ***
## level_of_severity -0.84325    0.07456  -11.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.607 on 98 degrees of freedom
```

```
## Multiple R-squared:  0.5662, Adjusted R-squared:  0.5618
## F-statistic: 127.9 on 1 and 98 DF,  p-value: < 2.2e-16
```

9. Are the results of the stepwise analysis consistent with your previous conclusions?

```
# yes, the final model from the stepwise analysis only includes level of
severity, as it is the only significant predictor
```