Anti_vac Exploratory Data Analysis

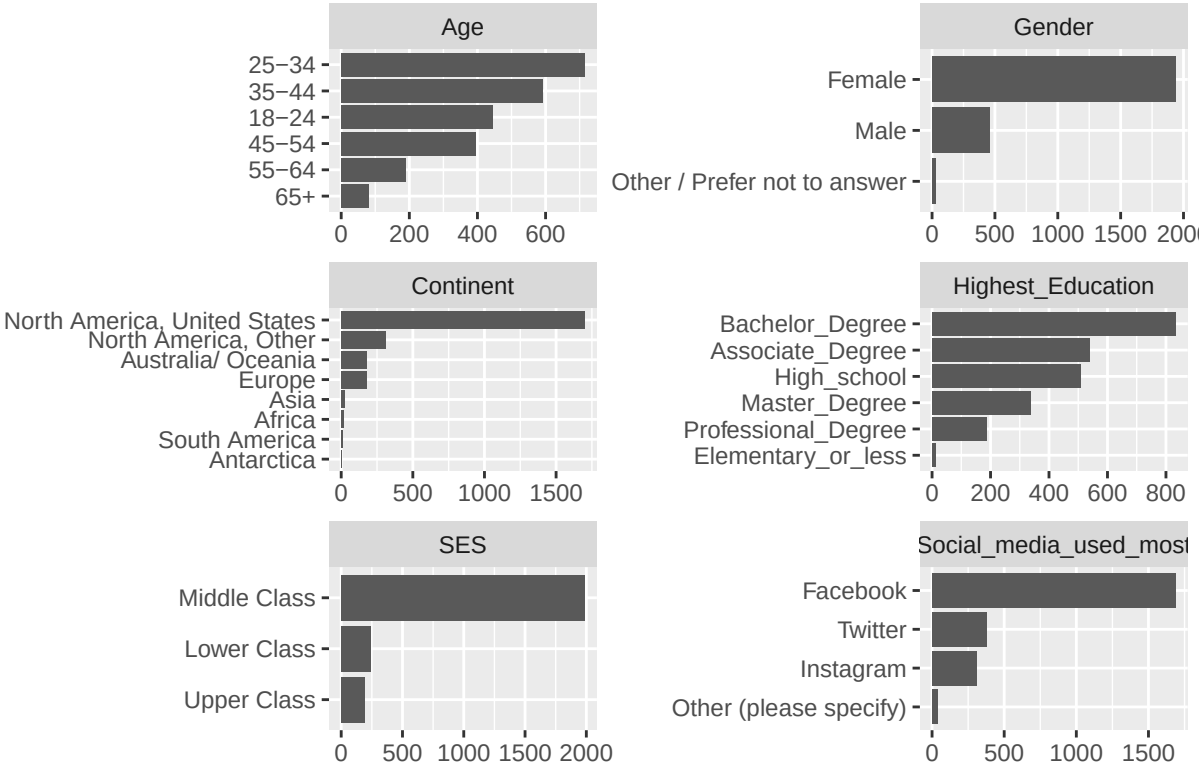
2022-09-10

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## here() starts at /Users/kohei/Dropbox/Mac/Documents/GitHub/Multivariate_Project
##
##
## Attaching package: 'psych'
##
##
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
##
##
## corrplot 0.92 loaded
```

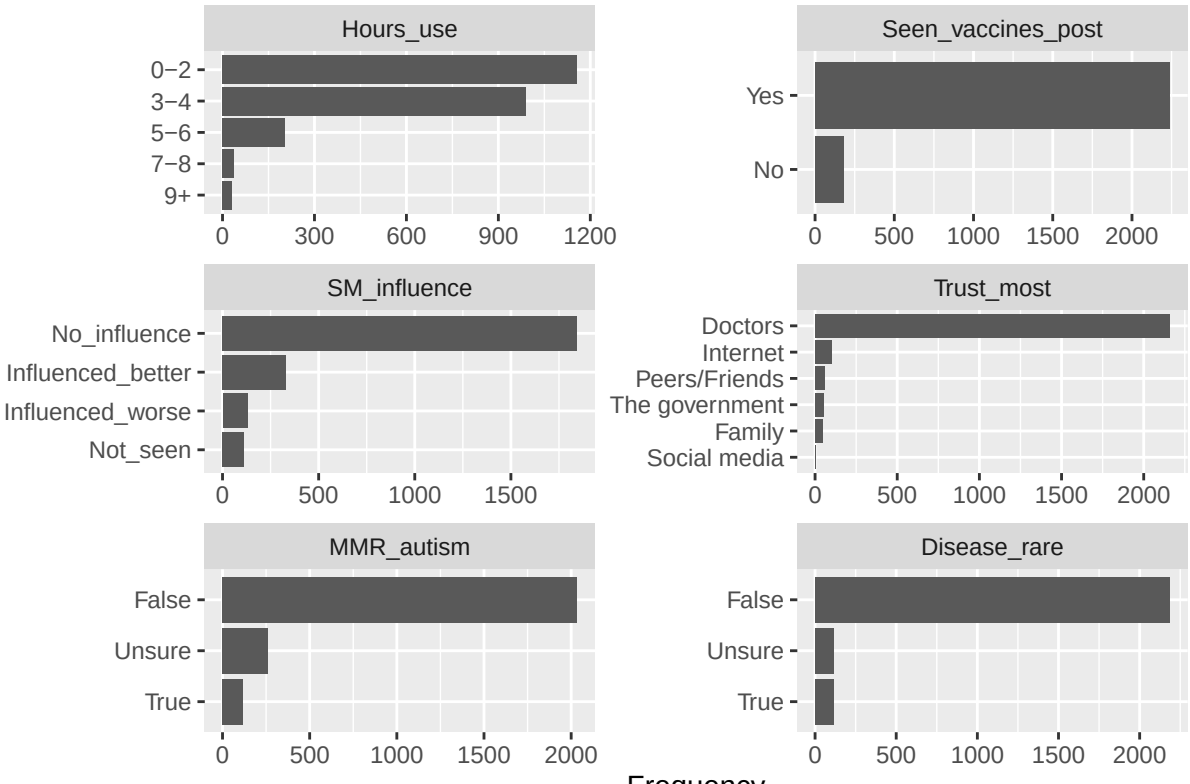
JL: From the original file antivac.csv, I created a new file antivac_renamed.csv, in which I manually shortened the variable names. For original variable names/survey items pls refer to the original dataset

Data cleaning

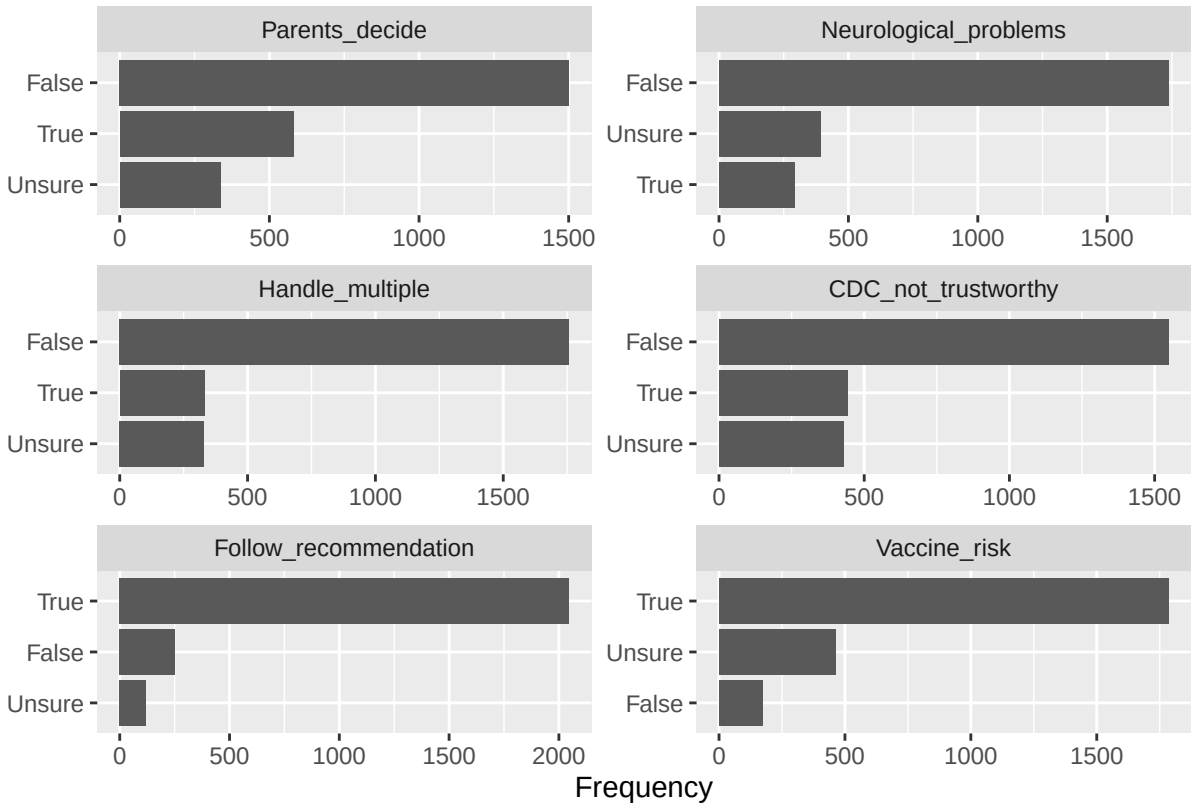
Visualization



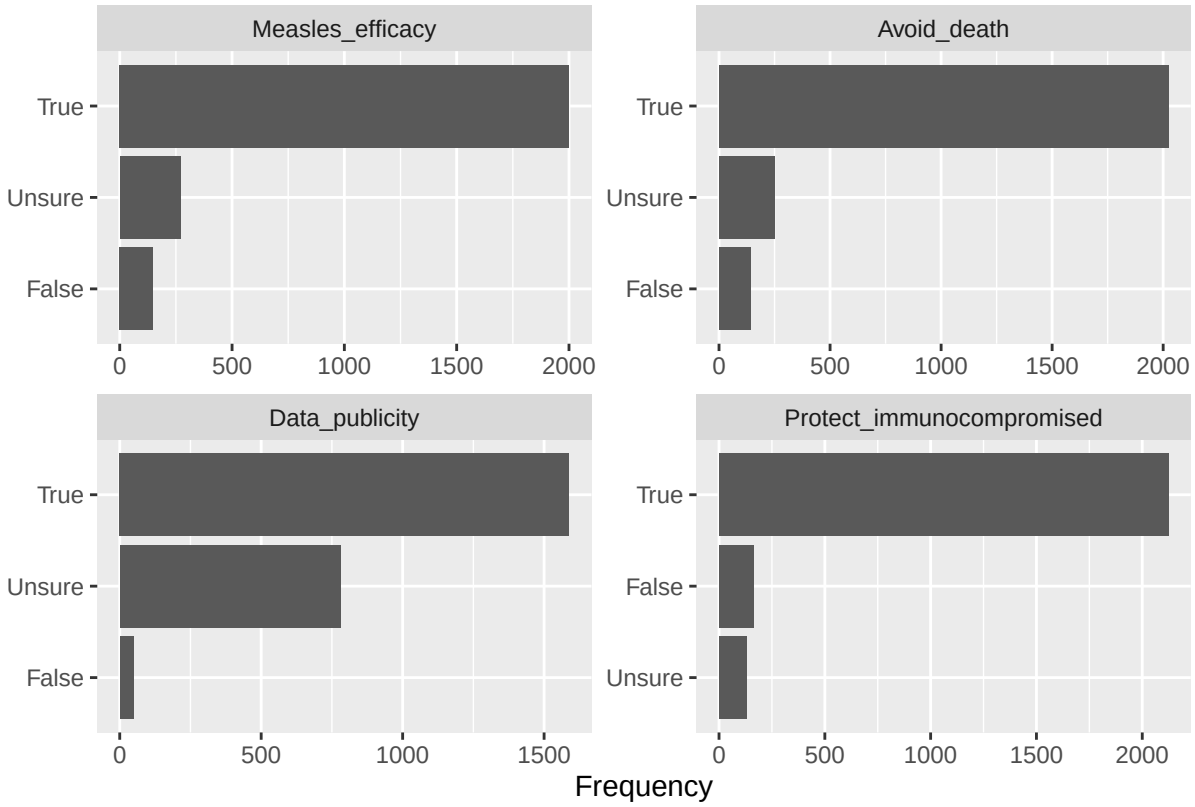
Page 1



Page 2


Parents_decide
False
True
Unsure
0
500
1000
1500
Neurological_problems
False
Unsure
True
0
500
1000
1500
Handle_multiple
False
True
Unsure
0
500
1000
1500
CDC_not_trustworthy
False
True
Unsure
0
500
1000
1500
Follow_recommendation
True
False
Unsure
0
500
1000
1500
2000
Vaccine_risk
True
Unsure
False
0
500
1000
1500
Frequency

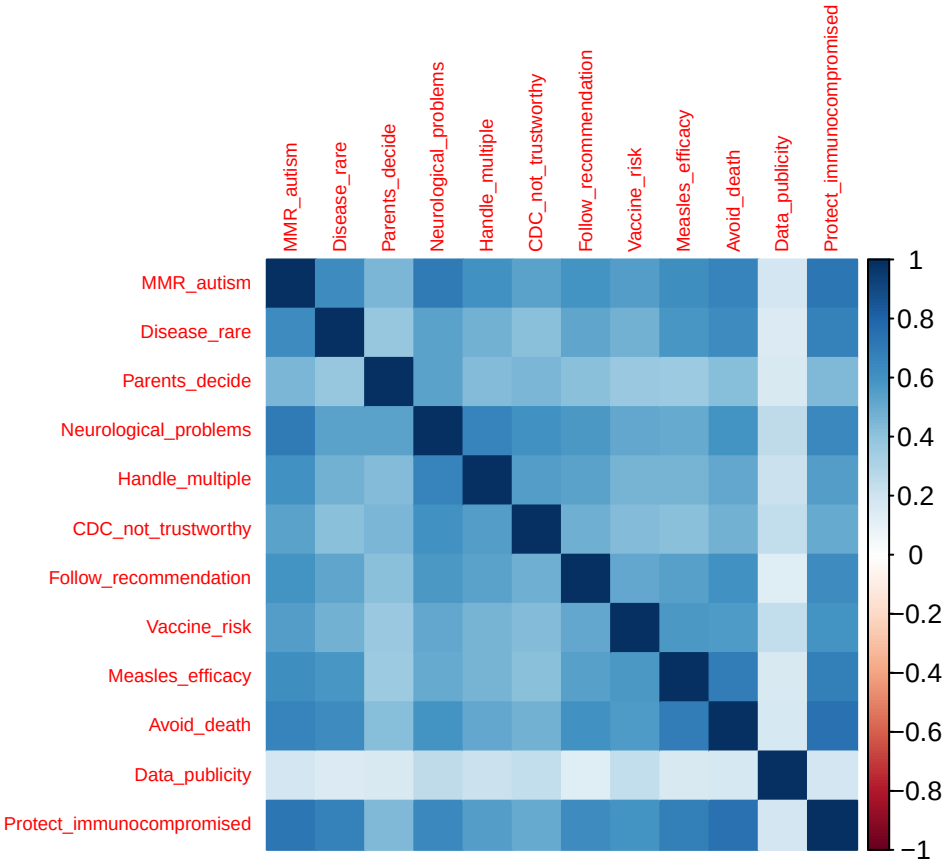
Page 3


Measles_efficacy
True
Unsure
False
0
500
1000
1500
2000
Avoid_death
True
Unsure
False
0
500
1000
1500
2000
Data_publicity
True
Unsure
False
0
500
1000
1500
Protect_immunocompromised
True
False
Unsure
0
500
1000
1500
2000
Frequency

Page 4

Converting outcome variable

Looking at outcome variable



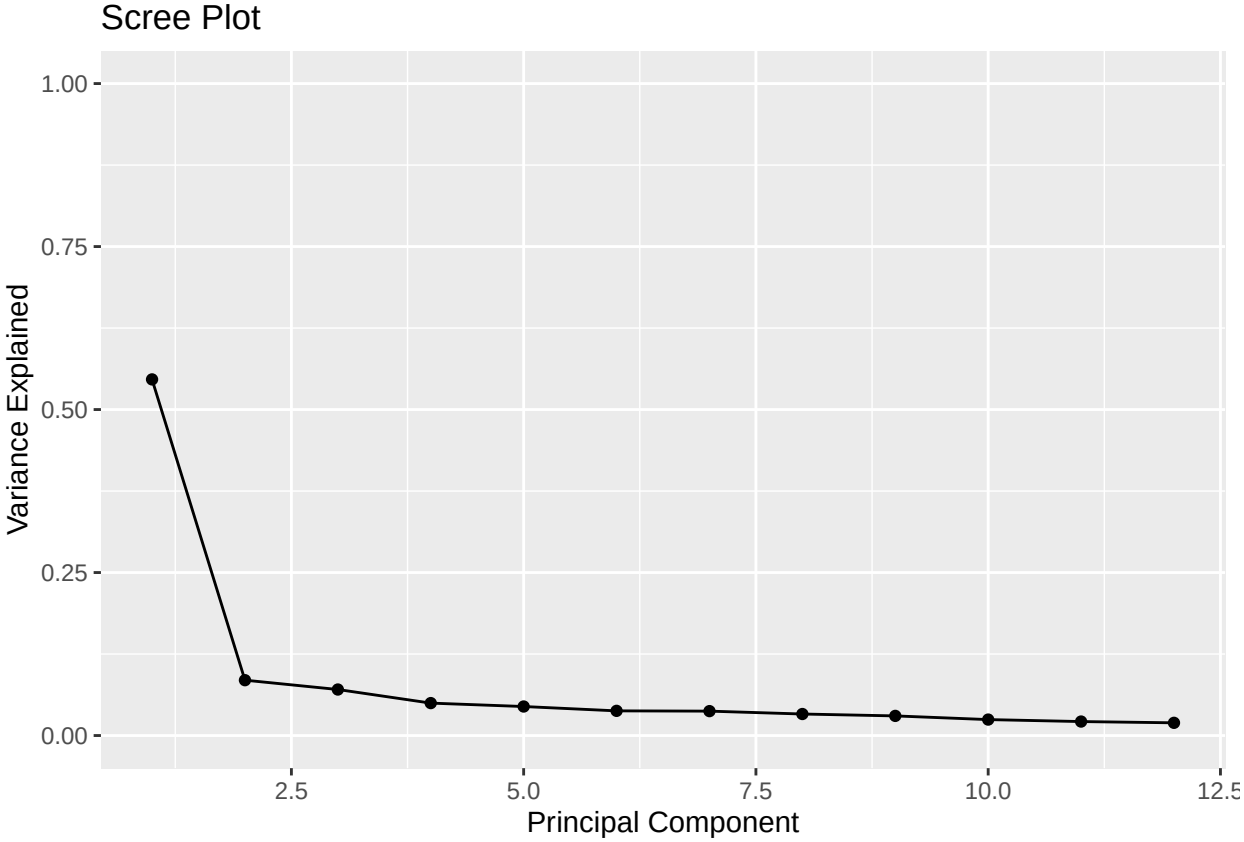
PCA for key-dependent coded

```
##                                PC1         PC2         PC3         PC4
## MMR_autism               0.3297216 -0.07331726  0.04013925 -0.12002494
## Disease_rare             0.2932395 -0.22362902 -0.14557988  0.05002360
## Parents_decide           0.2390695  0.22120981  0.44734958  0.81031113
## Neurological_problems    0.3201582  0.18086398  0.23155845 -0.15530164
## Handle_multiple          0.2901178  0.17117996  0.26772434 -0.44007036
## CDC_not_trustworthy      0.2701117  0.30514880  0.31410682 -0.23686820
## Follow_recommendation    0.2971270 -0.11777624  0.06624697 -0.12144711
## Vaccine_risk             0.2820874 -0.01538217 -0.27286680  0.11883093
## Measles_efficacy         0.2998799 -0.25136541 -0.30718308  0.12246975
## Avoid_death              0.3226695 -0.20093506 -0.17300918  0.07621593
## Data_publicity           0.1140512  0.76807478 -0.57764771  0.04809629
## Protect_immunocompromised 0.3370477 -0.18150626 -0.12766953  0.02468421
##                                PC5         PC6         PC7         PC8
## MMR_autism               0.15608965 -0.06756981  0.23740167 -0.02343652
## Disease_rare             0.54504845 -0.16509283 -0.02846009  0.61262200
## Parents_decide           0.04149231  0.08803154  0.05636432 -0.04468289
## Neurological_problems    0.12541078  0.04754628  0.21673205 -0.03259866
## Handle_multiple          0.02863455  0.20356649  0.43291874 -0.17165607
## CDC_not_trustworthy     -0.23483997 -0.58288350 -0.51766448  0.08473669
## Follow_recommendation   -0.15025291  0.72213514 -0.54108717  0.13546021
```

```
## Vaccine_risk              -0.72124207 -0.03005959  0.34569315  0.42552792
## Measles_efficacy          -0.11340557 -0.18734038 -0.06012208 -0.50942305
## Avoid_death                0.05228362 -0.03101009 -0.12071159 -0.34970531
## Data_publicity             0.18470273  0.12118236 -0.07765789 -0.03764873
## Protect_immunocompromised  0.11253187 -0.06204873 -0.02810378 -0.04234529
##                                   PC9         PC10         PC11         PC12
## MMR_autism                -5.067610e-01  0.386904791  0.507853170  0.345884029
## Disease_rare               3.115933e-01  0.024461306 -0.150276006  0.139943050
## Parents_decide             1.153923e-01  0.018995490  0.092643157  0.016811099
## Neurological_problems     -4.249955e-01  0.039195364 -0.716261812 -0.160737784
## Handle_multiple            5.638475e-01 -0.126155029  0.167214762  0.015724400
## CDC_not_trustworthy        5.699310e-02 -0.007384695  0.068927761  0.032259072
## Follow_recommendation     -1.016323e-02  0.147378205  0.007548868  0.040567941
## Vaccine_risk              -1.378899e-02 -0.080486197 -0.040180230  0.040729715
## Measles_efficacy           3.153863e-01  0.518949140 -0.240073645 -0.005597766
## Avoid_death               -1.230660e-01 -0.704791544 -0.049160395  0.406097367
## Data_publicity             9.805136e-05  0.021299379  0.065398395  0.009179406
## Protect_immunocompromised -1.268342e-01 -0.192738787  0.314922599 -0.815498432
```

```
##  [1] 0.54624025 0.08492084 0.07061426 0.04975854 0.04453283 0.03786050
##  [7] 0.03744852 0.03302765 0.03013049 0.02448054 0.02146945 0.01951612
```

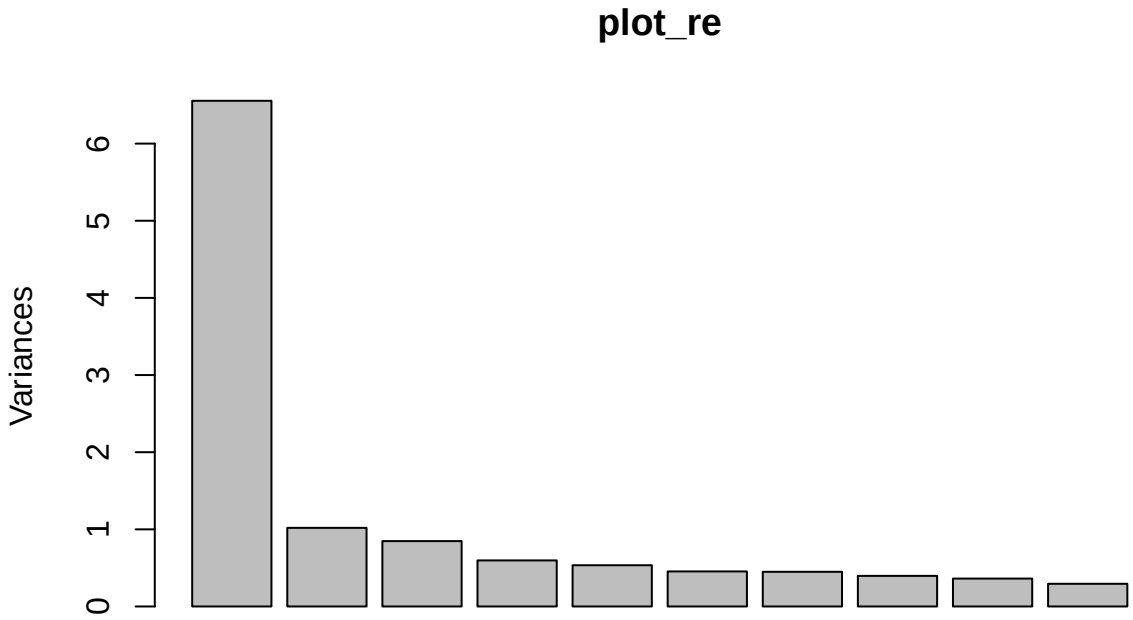
```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
```




```
## Warning in plot.window(xlim, ylim, log = log, ...): "data" is not a graphical
## parameter
```

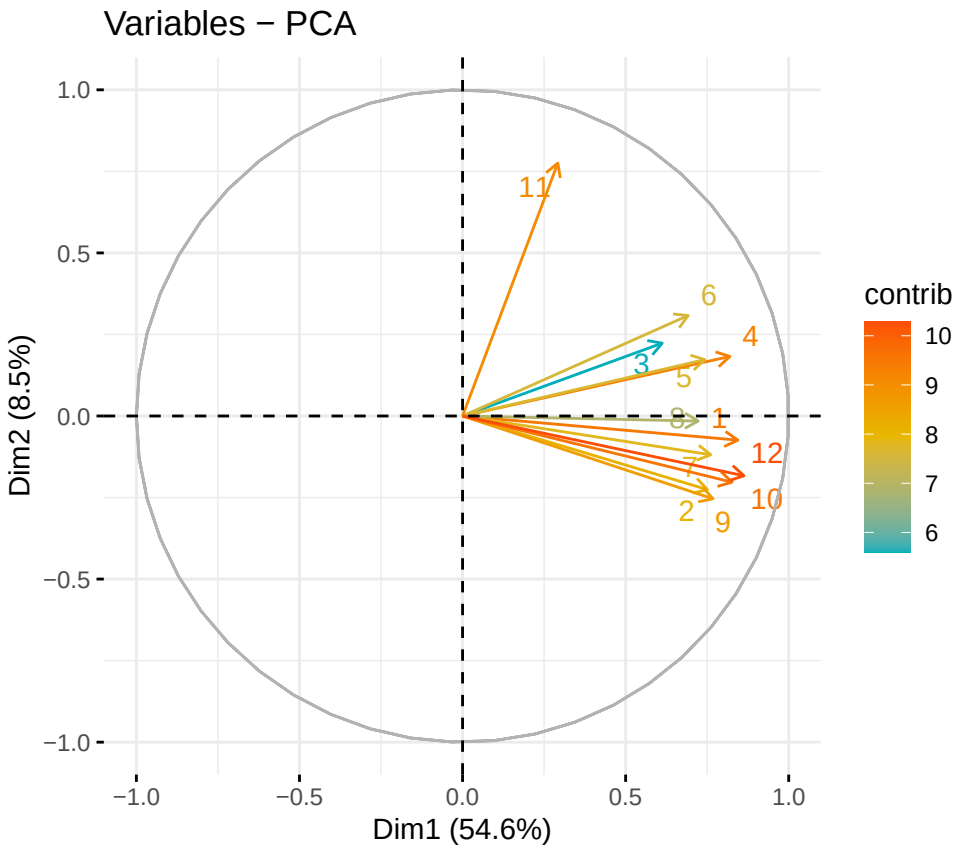
```
## Warning in plot.window(xlim, ylim, log = log, ...): "loadings" is not a
## graphical parameter
```

```
## Warning in plot.window(xlim, ylim, log = log, ...): "loadings.colour" is not a
## graphical parameter

## Warning in plot.window(xlim, ylim, log = log, ...): "loadings.label" is not a
## graphical parameter

## Warning in plot.window(xlim, ylim, log = log, ...): "loadings.label.size" is not
## a graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...): "data"
## is not a graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "loadings" is not a graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "loadings.colour" is not a graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "loadings.label" is not a graphical parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## "loadings.label.size" is not a graphical parameter

## Warning in axis(if (horiz) 1 else 2, cex.axis = cex.axis, ...): "data" is not a
## graphical parameter

## Warning in axis(if (horiz) 1 else 2, cex.axis = cex.axis, ...): "loadings" is
## not a graphical parameter

## Warning in axis(if (horiz) 1 else 2, cex.axis = cex.axis, ...):
## "loadings.colour" is not a graphical parameter

## Warning in axis(if (horiz) 1 else 2, cex.axis = cex.axis, ...): "loadings.label"
## is not a graphical parameter

## Warning in axis(if (horiz) 1 else 2, cex.axis = cex.axis, ...):
## "loadings.label.size" is not a graphical parameter
```




```
##
## Call:
```

```
## factanal(x = variable, factors = 2)
##
## Uniquenesses:
##                MMR_autism              Disease_rare            Parents_decide
##                     0.303                     0.444                     0.639
##     Neurological_problems           Handle_multiple       CDC_not_trustworthy
##                     0.230                     0.422                     0.504
##     Follow_recommendation              Vaccine_risk          Measles_efficacy
##                     0.471                     0.528                     0.349
##               Avoid_death            Data_publicity Protect_immunocompromised
##                     0.284                     0.916                     0.221
##
## Loadings:
##                           Factor1 Factor2
## MMR_autism                0.609   0.571
## Disease_rare              0.657   0.353
## Parents_decide            0.297   0.522
## Neurological_problems     0.400   0.781
## Handle_multiple           0.366   0.666
## CDC_not_trustworthy       0.313   0.631
## Follow_recommendation     0.561   0.463
## Vaccine_risk              0.575   0.375
## Measles_efficacy          0.763   0.264
## Avoid_death               0.761   0.370
## Data_publicity            0.101   0.271
## Protect_immunocompromised 0.774   0.423
##
##                Factor1 Factor2
## SS loadings      3.699   2.991
## Proportion Var   0.308   0.249
## Cumulative Var   0.308   0.557
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 214.65 on 43 degrees of freedom.
## The p-value is 1.16e-24

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```



Chi squared test of independence

```
##
##       Facebook Instagram Other (please specify) Twitter  Sum
##   0-2      887       118                    16     137 1158
##   3-4      660       143                    10     176  989
##   5-6      107        40                     9      48  204
##   7-8       21         3                     3       9   36
##   9+        13         7                     2       8   30
##   Sum     1688       311                    40     378 2417
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  table_hrs_vs_platform
## X-squared = 92.976, df = NA, p-value = 0.0004998
```

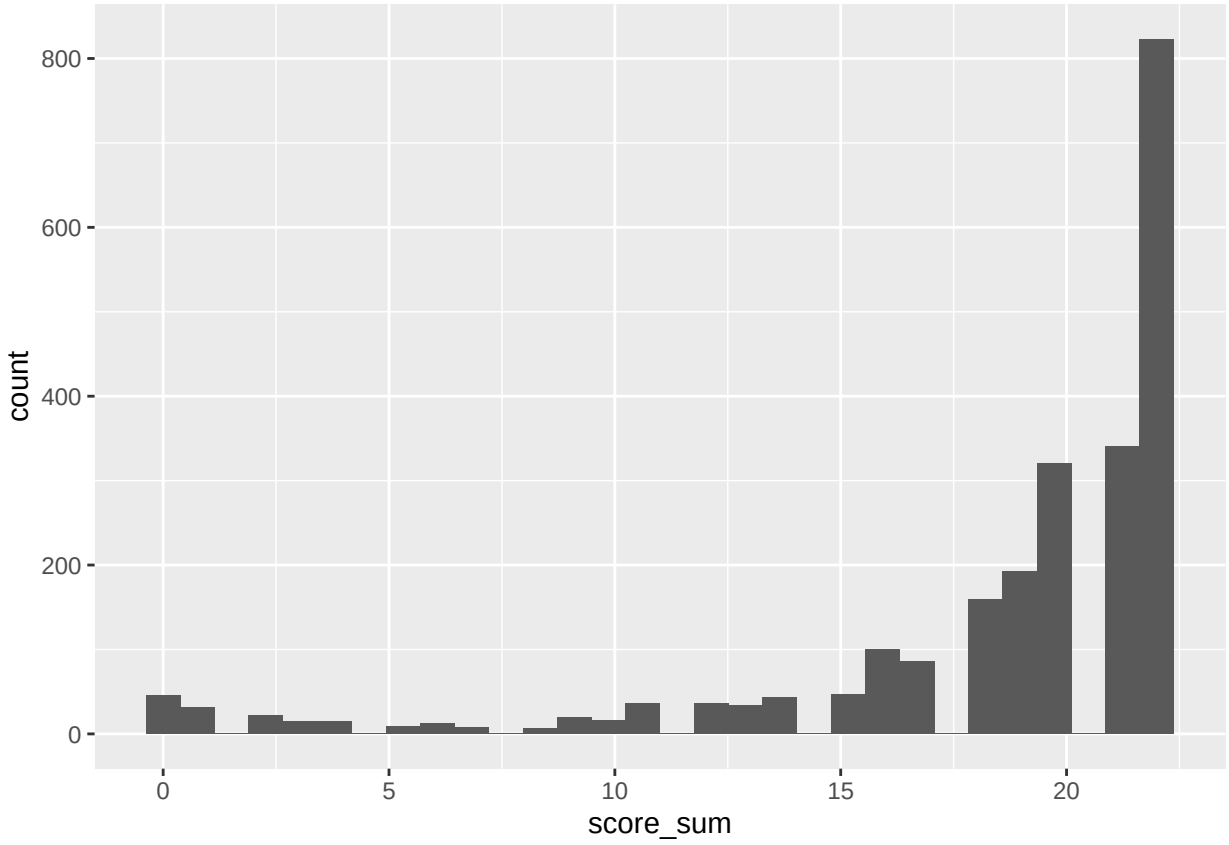
Modeling

```
##
## Reliability analysis
## Call: alpha(x = variable)
##
##   raw_alpha std.alpha G6(smc) average_r S/N    ase mean   sd median_r
##       0.91      0.92    0.92      0.48  11 0.0025  1.7 0.45     0.51
##
```

```
##     95% confidence boundaries
##          lower alpha upper
## Feldt     0.91  0.91  0.92
## Duhachek  0.91  0.91  0.92
##
##  Reliability if an item is dropped:
##                            raw_alpha std.alpha G6(smc) average_r  S/N alpha se
## MMR_autism                      0.90      0.91    0.91      0.47  9.7   0.0029
## Disease_rare                    0.91      0.91    0.92      0.48 10.3   0.0027
## Parents_decide                  0.91      0.92    0.92      0.50 11.0   0.0025
## Neurological_problems           0.90      0.91    0.91      0.47  9.8   0.0030
## Handle_multiple                 0.90      0.91    0.92      0.48 10.2   0.0028
## CDC_not_trustworthy             0.91      0.91    0.92      0.49 10.5   0.0027
## Follow_recommendation           0.90      0.91    0.92      0.48 10.2   0.0028
## Vaccine_risk                    0.91      0.91    0.92      0.49 10.4   0.0027
## Measles_efficacy                0.90      0.91    0.91      0.48 10.1   0.0028
## Avoid_death                     0.90      0.91    0.91      0.47  9.8   0.0028
## Data_publicity                  0.92      0.93    0.93      0.54 13.1   0.0024
## Protect_immunocompromised       0.90      0.91    0.91      0.47  9.6   0.0029
##                             var.r med.r
## MMR_autism                 0.0243  0.50
## Disease_rare               0.0260  0.51
## Parents_decide             0.0269  0.54
## Neurological_problems      0.0262  0.48
## Handle_multiple            0.0275  0.51
## CDC_not_trustworthy        0.0283  0.53
## Follow_recommendation      0.0265  0.51
## Vaccine_risk               0.0280  0.53
## Measles_efficacy           0.0255  0.51
## Avoid_death                0.0244  0.51
## Data_publicity             0.0089  0.54
## Protect_immunocompromised 0.0230  0.50
##
##  Item statistics
##                               n raw.r std.r r.cor r.drop mean   sd
## MMR_autism                 2417  0.82  0.83  0.83   0.79  1.8 0.51
## Disease_rare               2417  0.72  0.74  0.71   0.67  1.9 0.47
## Parents_decide             2417  0.66  0.63  0.57   0.55  1.4 0.85
## Neurological_problems      2417  0.83  0.82  0.81   0.78  1.6 0.69
## Handle_multiple            2417  0.76  0.74  0.72   0.69  1.6 0.72
## CDC_not_trustworthy        2417  0.73  0.70  0.66   0.64  1.5 0.78
## Follow_recommendation      2417  0.75  0.75  0.72   0.69  1.7 0.63
## Vaccine_risk               2417  0.72  0.72  0.69   0.66  1.7 0.60
## Measles_efficacy           2417  0.74  0.76  0.74   0.69  1.8 0.55
## Avoid_death                2417  0.80  0.81  0.81   0.76  1.8 0.54
## Data_publicity             2417  0.35  0.36  0.27   0.26  1.6 0.52
## Protect_immunocompromised 2417  0.83  0.85  0.85   0.80  1.8 0.54
##
## Non missing response frequency for each item
##                               0    1    2 miss
## MMR_autism                 0.05 0.11 0.84    0
## Disease_rare               0.05 0.05 0.90    0
## Parents_decide             0.24 0.14 0.62    0
## Neurological_problems      0.12 0.16 0.72    0
```

```
## Handle_multiple           0.14 0.14 0.73    0
## CDC_not_trustworthy       0.18 0.18 0.64    0
## Follow_recommendation     0.10 0.05 0.85    0
## Vaccine_risk              0.07 0.19 0.74    0
## Measles_efficacy          0.06 0.11 0.83    0
## Avoid_death               0.06 0.10 0.84    0
## Data_publicity            0.02 0.32 0.66    0
## Protect_immunocompromised 0.07 0.05 0.88    0
```

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




```
##
## Call:
## lm(formula = score_sum ~ Hours_use + SM_influence + Highest_Education +
##     SES + Social_media_used_most + Age + Gender + Continent +
##     Trust_most, data = antivax_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2704  -1.2363   0.6828   1.9312  14.2214
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                              21.30143    0.96037  22.180
## Hours_use3-4                              0.65220    0.15463   4.218
## Hours_use5-6                              0.17261    0.27282   0.633
## Hours_use7-8                              0.32056    0.58654   0.547
## Hours_use9+                               0.39549    0.64167   0.616
```

```
## SM_influenceInfluenced_worse                   -8.00814    0.39993 -20.024
## SM_influenceNo_influence                       -0.69781    0.20897  -3.339
## SM_influenceNot_seen                           -3.07867    0.38597  -7.976
## Highest_EducationBachelor_Degree                0.56546    0.19312   2.928
## Highest_EducationElementary_or_less            -3.05444    1.00970  -3.025
## Highest_EducationHigh_school                   -0.58665    0.21725  -2.700
## Highest_EducationMaster_Degree                  0.83656    0.24646   3.394
## Highest_EducationProfessional_Degree            1.80743    0.30575   5.911
## SESMiddle Class                                -0.16629    0.24258  -0.686
## SESUpper Class                                 -0.36118    0.35365  -1.021
## Social_media_used_mostInstagram                -0.75774    0.22711  -3.336
## Social_media_used_mostOther (please specify)    0.24134    0.55707   0.433
## Social_media_used_mostTwitter                   0.12311    0.21038   0.585
## Age25-34                                       -0.73808    0.22648  -3.259
## Age35-44                                       -1.00661    0.24494  -4.110
## Age45-54                                       -0.52032    0.26252  -1.982
## Age55-64                                       -0.78533    0.32160  -2.442
## Age65+                                         -1.46580    0.43548  -3.366
## GenderMale                                      0.02665    0.18581   0.143
## GenderOther / Prefer not to answer              0.12288    0.69266   0.177
## ContinentAntarctica                            -2.65349    3.64465  -0.728
## ContinentAsia                                  -1.11509    1.15364  -0.967
## ContinentAustralia/ Oceania                     0.73142    0.92189   0.793
## ContinentEurope                                 0.19006    0.92283   0.206
## ContinentNorth America, Other                  -0.61154    0.90702  -0.674
## ContinentNorth America, United States          -0.89836    0.89029  -1.009
## ContinentSouth America                         -1.36979    1.36539  -1.003
## Trust_mostFamily                               -9.88099    0.51376 -19.233
## Trust_mostInternet                             -8.52932    0.37433 -22.785
## Trust_mostPeers/Friends                       -10.12334    0.49585 -20.416
## Trust_mostSocial media                         -8.04893    1.56405  -5.146
## Trust_mostThe government                       -0.25170    0.49695  -0.506
##                                               Pr(>|t|)
## (Intercept)                                    < 2e-16 ***
## Hours_use3-4                                  2.56e-05 ***
## Hours_use5-6                                  0.526997
## Hours_use7-8                                  0.584762
## Hours_use9+                                   0.537720
## SM_influenceInfluenced_worse                   < 2e-16 ***
## SM_influenceNo_influence                      0.000853 ***
## SM_influenceNot_seen                          2.32e-15 ***
## Highest_EducationBachelor_Degree              0.003444 **
## Highest_EducationElementary_or_less           0.002512 **
## Highest_EducationHigh_school                  0.006977 **
## Highest_EducationMaster_Degree                0.000699 ***
## Highest_EducationProfessional_Degree          3.88e-09 ***
## SESMiddle Class                               0.493093
## SESUpper Class                                0.307214
## Social_media_used_mostInstagram               0.000862 ***
## Social_media_used_mostOther (please specify) 0.664890
## Social_media_used_mostTwitter                 0.558485
## Age25-34                                      0.001134 **
## Age35-44                                      4.10e-05 ***
## Age45-54                                      0.047591 *
```

```
## Age55-64                                       0.014680 *
## Age65+                                         0.000775 ***
## GenderMale                                     0.885967
## GenderOther / Prefer not to answer             0.859205
## ContinentAntarctica                            0.466654
## ContinentAsia                                  0.333850
## ContinentAustralia/ Oceania                    0.427629
## ContinentEurope                                0.836848
## ContinentNorth America, Other                  0.500229
## ContinentNorth America, United States          0.313045
## ContinentSouth America                         0.315855
## Trust_mostFamily                                < 2e-16 ***
## Trust_mostInternet                              < 2e-16 ***
## Trust_mostPeers/Friends                         < 2e-16 ***
## Trust_mostSocial media                         2.88e-07 ***
## Trust_mostThe government                       0.612558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.413 on 2380 degrees of freedom
## Multiple R-squared:  0.5782, Adjusted R-squared:  0.5718
## F-statistic: 90.61 on 36 and 2380 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = score_sum ~ hours_num * SM_influence + Highest_Education +
##     SES + Social_media_used_most + Age + Gender + trust_other,
##     data = antivax_clean)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.0911  -1.2507   0.7916   1.9469  13.3499
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                 21.31267    0.49505  43.052
## hours_num                                   -0.06187    0.10276  -0.602
## SM_influenceInfluenced_worse                -8.82692    0.72315 -12.206
## SM_influenceNo_influence                    -1.39341    0.39475  -3.530
## SM_influenceNot_seen                        -3.01914    0.64576  -4.675
## Highest_EducationBachelor_Degree             0.62914    0.19436   3.237
## Highest_EducationElementary_or_less         -2.66896    1.01826  -2.621
## Highest_EducationHigh_school                -0.59247    0.21912  -2.704
## Highest_EducationMaster_Degree               0.87385    0.24634   3.547
## Highest_EducationProfessional_Degree         1.88035    0.30778   6.109
## SESMiddle Class                             -0.23959    0.24463  -0.979
## SESUpper Class                              -0.56902    0.35491  -1.603
## Social_media_used_mostInstagram             -0.83577    0.22865  -3.655
## Social_media_used_mostOther (please specify)  0.09296    0.56131   0.166
## Social_media_used_mostTwitter                0.02680    0.21077   0.127
## Age25-34                                    -0.73392    0.22766  -3.224
## Age35-44                                    -0.96921    0.24609  -3.938
## Age45-54                                    -0.56142    0.26333  -2.132
## Age55-64                                    -0.79196    0.32185  -2.461
```

```
## Age65+                                        -1.43360    0.43633  -3.286
## GenderMale                                     0.15528    0.18501   0.839
## GenderOther / Prefer not to answer             0.22300    0.68719   0.325
## trust_other                                   -9.35847    0.28650 -32.665
## hours_num:SM_influenceInfluenced_worse         0.23575    0.22027   1.070
## hours_num:SM_influenceNo_influence             0.19851    0.11073   1.793
## hours_num:SM_influenceNot_seen                -0.19154    0.21180  -0.904
##                                               Pr(>|t|)
## (Intercept)                                    < 2e-16 ***
## hours_num                                     0.547181
## SM_influenceInfluenced_worse                   < 2e-16 ***
## SM_influenceNo_influence                      0.000424 ***
## SM_influenceNot_seen                          3.10e-06 ***
## Highest_EducationBachelor_Degree              0.001224 **
## Highest_EducationElementary_or_less           0.008820 **
## Highest_EducationHigh_school                  0.006903 **
## Highest_EducationMaster_Degree                0.000397 ***
## Highest_EducationProfessional_Degree          1.16e-09 ***
## SESMiddle Class                               0.327492
## SESUpper Class                                0.109001
## Social_media_used_mostInstagram               0.000262 ***
## Social_media_used_mostOther (please specify) 0.868475
## Social_media_used_mostTwitter                 0.898819
## Age25-34                                      0.001282 **
## Age35-44                                      8.44e-05 ***
## Age45-54                                      0.033113 *
## Age55-64                                      0.013940 *
## Age65+                                        0.001032 **
## GenderMale                                    0.401375
## GenderOther / Prefer not to answer            0.745583
## trust_other                                    < 2e-16 ***
## hours_num:SM_influenceInfluenced_worse        0.284606
## hours_num:SM_influenceNo_influence            0.073133 .
## hours_num:SM_influenceNot_seen                0.365893
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.451 on 2391 degrees of freedom
## Multiple R-squared:  0.5667, Adjusted R-squared:  0.5622
## F-statistic: 125.1 on 25 and 2391 DF,  p-value: < 2.2e-16
```

Lasso

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
## Loaded glmnet 4.1-4
## 22 x 1 sparse Matrix of class "dgCMatrix"
```

```
##                                              s0
##                                       .
## hours_num                            0.06483037
## SM_influence_worse                  -7.95236307
## SM_influence_no_influence           -0.77362751
## SM_influence_not_seen               -2.42270242
## Highest_Education_elementary_or_less -2.17073350
## Highest_Education_high              -0.50478161
## Highest_Education_bachelor           0.45595671
## Highest_Education_master             0.62508111
## Highest_Education_professional       1.50605674
## SES_middle                           .
## SES_upper                           -0.12888025
## Social_media_used_most_Instagram    -0.68235271
## Social_media_used_most_Twitter       0.15388387
## Social_media_used_most_other         0.04401002
## age                                 -0.01380191
## Gender_male                          0.07389670
## Gender_other                         .
## hours_sum_SM_worse                   .
## hours_sum_SM_no_influence            0.03976459
## hours_sum_SM_not_seen               -0.28578148
## trust_other                         -9.38086479
```