Citation Analysis

Read data

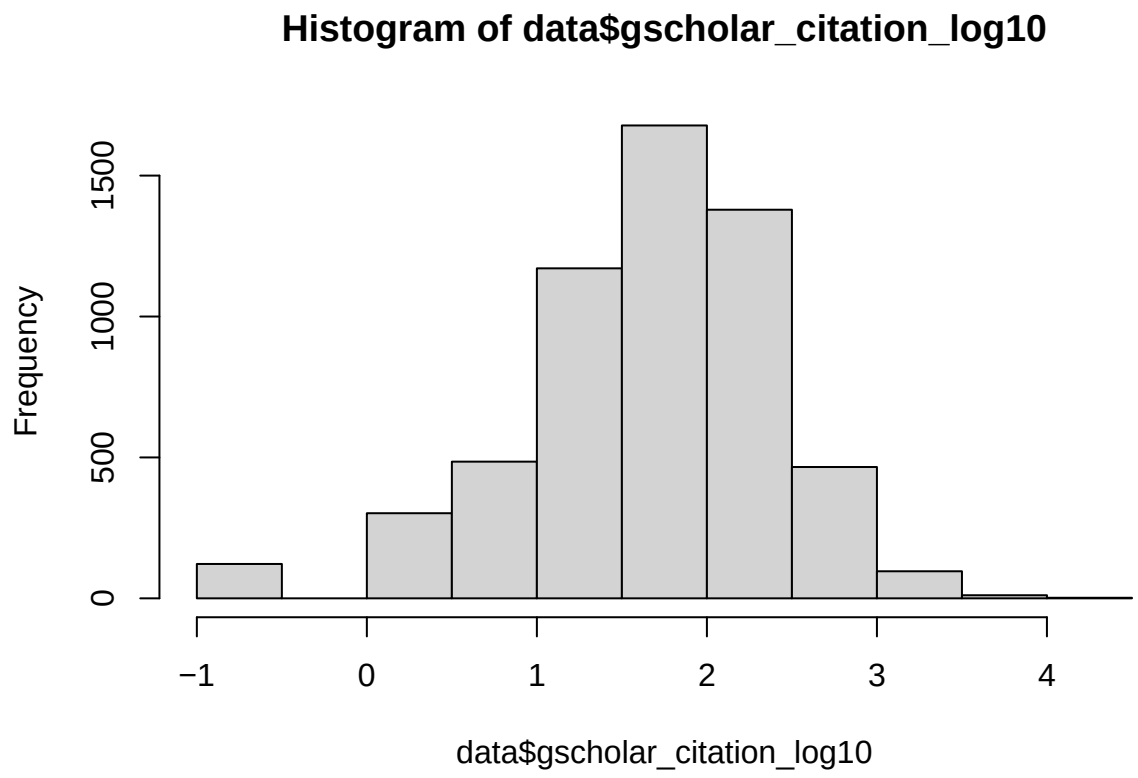
The first step is to read data:

```
df <- read.csv("../data/processed/papers_to_study_expanded.csv")
df$Year.Distance.from.2022 = abs(df$year - 2022)

# do the log10 transformation on citation counts
df$gscholar_citation_log10 <- log10(df$gscholar_citation + 0.1)
var_cols <- c(1, 16:17, 20:33, 35:36)
data <- df[, var_cols]
```

Deal with univariate outliers

```
hist(data$gscholar_citation_log10)
```




```
data$gscholar_citation_log10_stdized <- scale(
  data$gscholar_citation_log10,
  center = T,
```

```
  scale = T
)
data$outlier0 <- ifelse(data$gscholar_citation_log10_stdized > 3.3 |
                          data$gscholar_citation_log10_stdized < -3.3,
                        1, 0
)
table(data$outlier0)
```

```
##
##    0    1
## 5588  124
```

As can be seen, there are 124 outliers.

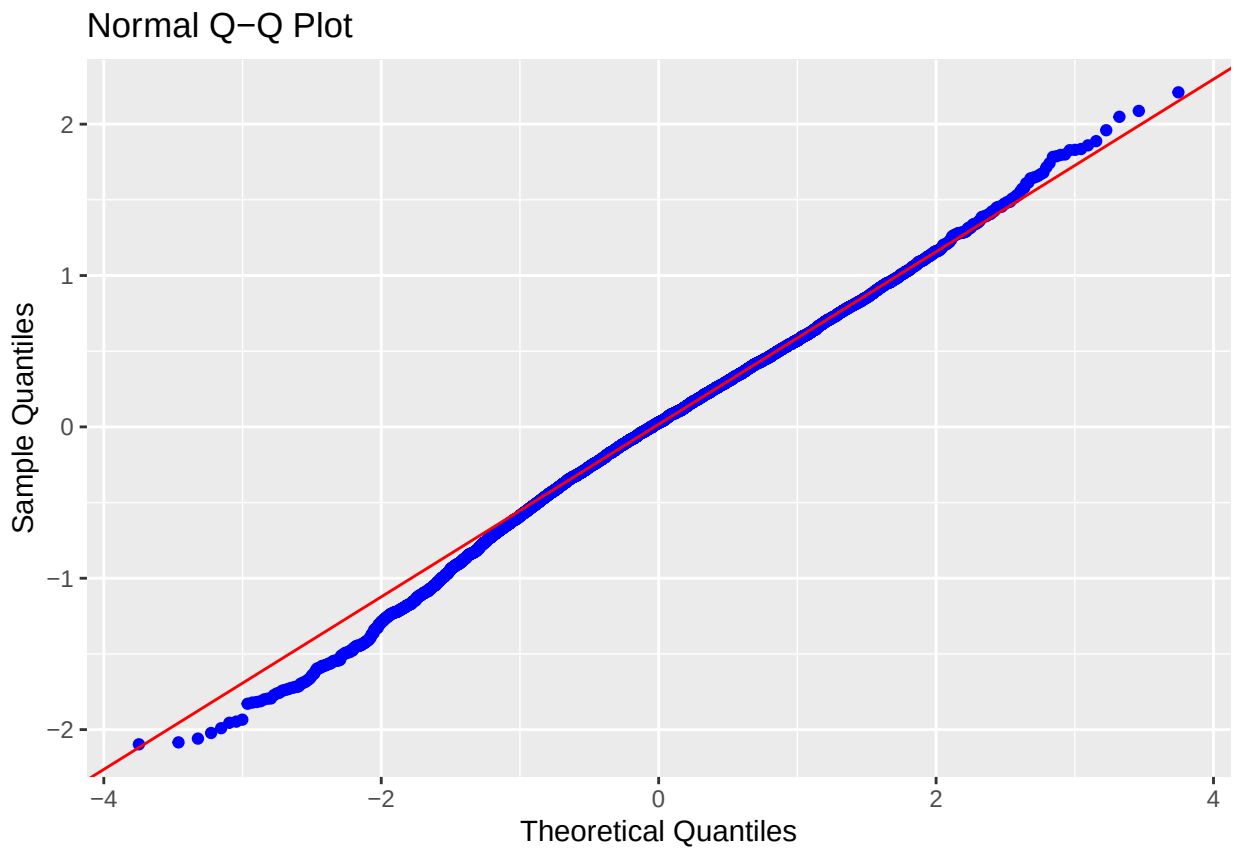
Build model

```
datanew <- subset(data, outlier0 == 0)
datanew$outlier0 <- NULL
datanew$gscholar_citation_log10_stdized <- NULL

model <- lm(gscholar_citation_log10~., datanew)
```

Dianostics

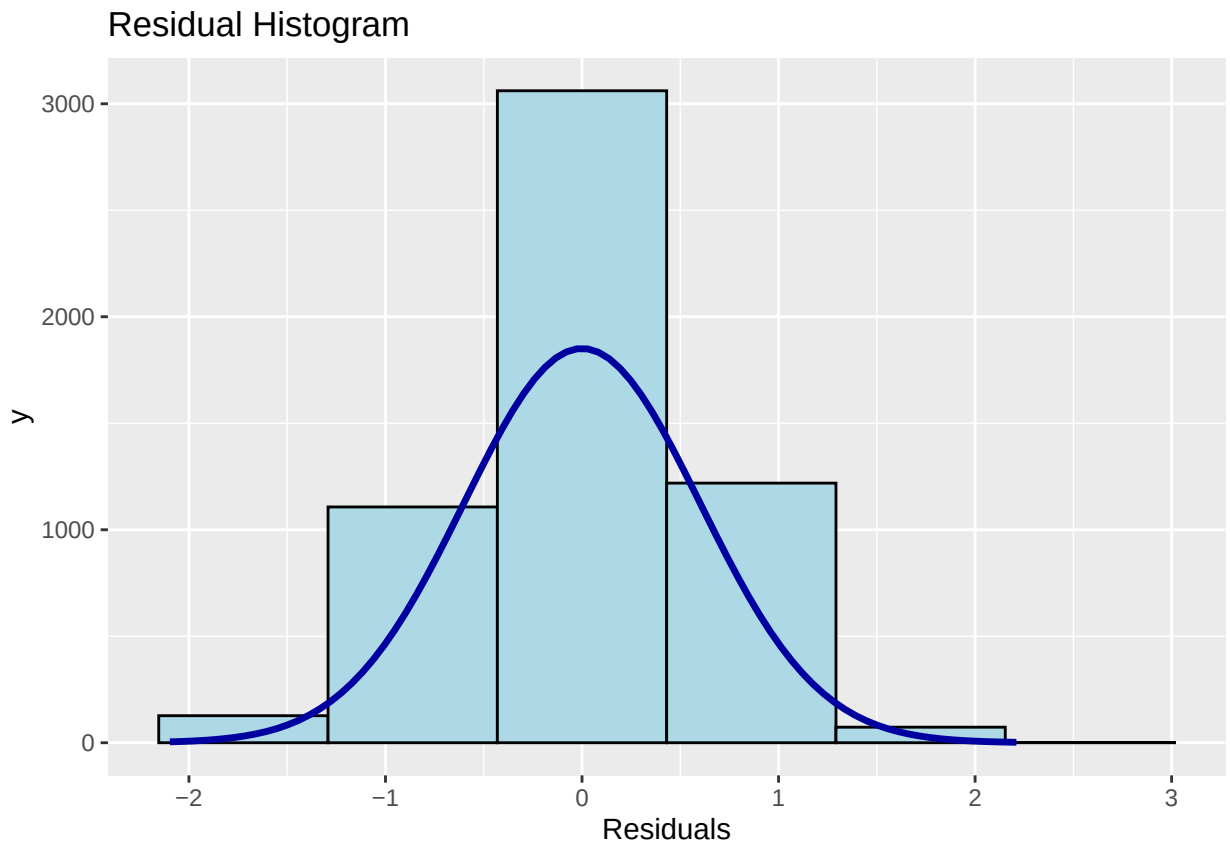
```
## Normality
ols_plot_resid_qq(model)
```



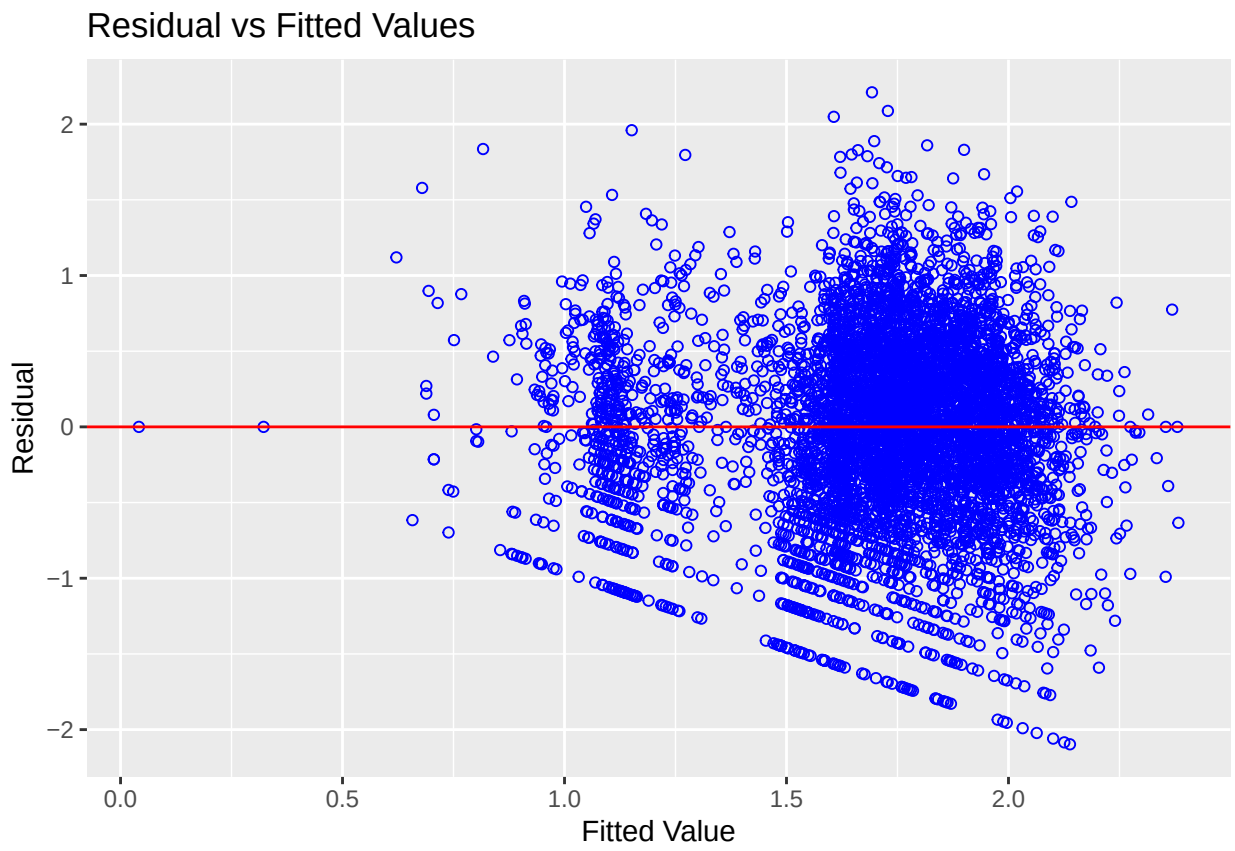
```
#Correlation between observed residuals and expected residuals under normality.
ols_test_correlation(model)
```

```
## [1] 0.9981045
```

```
ols_plot_resid_hist(model)
```




```
## linearity & homoscedasticity
ols_plot_resid_fit(model)
```




```
## collinearity diagnostics
# ols_coll_diag(model)
vif(model)
```

```
##                                    GVIF Df GVIF^(1/(2*Df))
## journal                        1.803091  4        1.076471
## cross_country                  6.648568  1        2.578482
## cross_type                     1.182023  1        1.087209
## num_race                      12.741419  1        3.569512
## num_country                    4.993049  1        2.234513
## cross_gender_and_race          4.116806  1        2.028991
## cross_gender_and_country       4.423884  1        2.103303
## cross_country_and_race         5.699357  1        2.387333
## cross_gender_race_and_country  6.238870  1        2.497773
## numberOfAuthors                2.313605  1        1.521054
## first_author_gender           11.043074  2        1.822941
## first_author_race             88.421658  5        1.565510
## first_author_country          27.371667 64        1.026193
## first_author_afftype           1.204277  1        1.097395
## with_us_authors                8.730568  1        2.954753
## cross_race_details           938.444111  6        1.768889
## cross_gender_details          24.309062  3        1.702007
## Year.Distance.from.2022        1.850935  1        1.360491
```

Run Model

```
summary(model)
```

```
##
## Call:
## lm(formula = gscholar_citation_log10 ~ ., data = datanew)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.09704 -0.36744  0.02661  0.40143  2.21022
##
## Coefficients:
##                                                  Estimate Std. Error t value
## (Intercept)                                     0.9054069  0.4402838   2.056
## journalCommunication, Culture and Critique     -0.6230127  0.0394026 -15.811
## journalHuman Communication Research             0.2030748  0.0321655   6.313
## journalJournal of Communication                 0.0816402  0.0296166   2.757
## journalJournal of Computer-Mediated Communication  0.2675762  0.0334730   7.994
## cross_countryTrue                               0.1314701  0.0684275   1.921
## cross_typeTrue                                  0.0646100  0.0421698   1.532
## num_race                                       -0.1356223  0.0698775  -1.941
## num_country                                     0.0881840  0.0430653   2.048
## cross_gender_and_raceTrue                      -0.1023844  0.0555191  -1.844
## cross_gender_and_countryTrue                   -0.2296005  0.0706195  -3.251
## cross_country_and_raceTrue                     -0.0983420  0.0974330  -1.009
## cross_gender_race_and_countryTrue              -0.0303581  0.1235740  -0.246
## numberOfAuthors                                -0.0041004  0.0087465  -0.469
## first_author_genderM                            0.0470639  0.0306602   1.535
## first_author_genderN                           -0.1497782  0.3551343  -0.422
## first_author_raceBlack                         -0.0459667  0.1365361  -0.337
## first_author_raceHispanic                       0.0343671  0.0825406   0.416
## first_author_raceIndigenous                    -0.2784185  0.3070507  -0.907
## first_author_raceMiddle Eastern                -0.1951185  0.1301925  -1.499
## first_author_raceWhite                          0.0121872  0.0475028   0.257
## first_author_countryAR                          0.6030282  0.6097410   0.989
## first_author_countryAT                          0.7924376  0.4497344   1.762
## first_author_countryAU                          0.7658587  0.4375402   1.750
## first_author_countryBE                          0.8221114  0.4462609   1.842
## first_author_countryBR                          0.6473883  0.4829682   1.340
## first_author_countryBS                          0.4459892  0.7446018   0.599
## first_author_countryBY                          0.5909765  0.7446071   0.794
## first_author_countryCA                          0.9234440  0.4343033   2.126
## first_author_countryCH                          0.9517731  0.4482276   2.123
## first_author_countryCL                          0.9530488  0.4655217   2.047
## first_author_countryCN                          0.7189568  0.4386865   1.639
## first_author_countryCO                          0.9174762  0.5277814   1.738
## first_author_countryCR                          1.4875070  0.7500735   1.983
## first_author_countryCZ                          1.6558278  0.4981715   3.324
## first_author_countryDE                          1.0187567  0.4346193   2.344
## first_author_countryDK                          0.9775656  0.4518420   2.164
## first_author_countryEE                          0.7461639  0.7450854   1.001
## first_author_countryEG                          1.1411127  0.5588135   2.042
## first_author_countryES                          0.9672190  0.4514063   2.143
```

```
## first_author_countryFI                          0.6343883  0.4515546   1.405
## first_author_countryFR                          0.4756327  0.4518587   1.053
## first_author_countryGB                          0.9540089  0.4331178   2.203
## first_author_countryGH                          0.9070180  0.6091975   1.489
## first_author_countryGR                          0.7652742  0.5555281   1.378
## first_author_countryHK                          0.3247295  0.5558951   0.584
## first_author_countryHR                          1.0390913  0.7457074   1.393
## first_author_countryHU                          0.6821178  0.5283550   1.291
## first_author_countryIE                          0.6120480  0.5556880   1.101
## first_author_countryIL                          0.8160946  0.4340386   1.880
## first_author_countryIN                          0.7741964  0.4885663   1.585
## first_author_countryIR                          1.4822005  0.7465686   1.985
## first_author_countryIT                          0.7851227  0.4514413   1.739
## first_author_countryJP                          0.8378602  0.4454740   1.881
## first_author_countryKE                          1.5368749  0.7496814   2.050
## first_author_countryKH                          1.1795531  0.7446038   1.584
## first_author_countryKR                          1.0137829  0.4424729   2.291
## first_author_countryLB                          1.0854016  0.4811275   2.256
## first_author_countryLK                          0.6998536  0.7454291   0.939
## first_author_countryMA                          0.6242003  0.7455965   0.837
## first_author_countryMO                          0.7959764  0.7451396   1.068
## first_author_countryMX                          0.9006917  0.5124692   1.758
## first_author_countryMZ                          0.1767140  0.7456640   0.237
## first_author_countryNG                          1.0490244  0.7446132   1.409
## first_author_countryNL                          0.9264245  0.4339976   2.135
## first_author_countryNO                          0.8896458  0.4624942   1.924
## first_author_countryNZ                          0.9674905  0.4585130   2.110
## first_author_countryPH                          1.0692064  0.4359772   2.452
## first_author_countryPL                          0.9234878  0.4970105   1.858
## first_author_countryPT                          0.5030576  0.6136298   0.820
## first_author_countryQA                          1.1229326  0.4991421   2.250
## first_author_countryRU                          0.4715051  0.5570985   0.846
## first_author_countrySE                          0.9003227  0.4458302   2.019
## first_author_countrySG                          0.7789790  0.4412175   1.766
## first_author_countrySI                          0.6197372  0.4968017   1.247
## first_author_countrySK                          0.8497126  0.7487288   1.135
## first_author_countryTH                          1.1234615  0.6122257   1.835
## first_author_countryTJ                         -0.0234866  0.4812555  -0.049
## first_author_countryTR                          0.9252435  0.5290125   1.749
## first_author_countryTW                          0.6440796  0.4616167   1.395
## first_author_countryUG                         -0.5597109  0.7506651  -0.746
## first_author_countryUK                          0.7585159  0.4551153   1.667
## first_author_countryUS                          0.9188326  0.4343952   2.115
## first_author_countryVN                          1.7658334  0.7456379   2.368
## first_author_countryZA                          0.4577875  0.5559675   0.823
## first_author_afftypeNon Education              -0.3824957  0.0397444  -9.624
## with_us_authorsTrue                             0.0563801  0.0589108   0.957
## cross_race_detailsBlack only                   -0.0813242  0.1679970  -0.484
## cross_race_detailscross race                    0.1990631  0.0923945   2.154
## cross_race_detailsHispanic only                -0.0675993  0.1142456  -0.592
## cross_race_detailsIndigenous only              -0.1667734  0.5290248  -0.315
## cross_race_detailsMiddle Eastern only           0.3183880  0.1755174   1.814
## cross_race_detailsWhite only                    0.0383960  0.0593955   0.646
## cross_gender_detailsF only                     -0.0982156  0.0346245  -2.837
```

```
## cross_gender_detailsM only                          -0.1623034  0.0304370  -5.332
## cross_gender_detailsN only                           0.0928102  0.4249839   0.218
## Year.Distance.from.2022                             -0.0051172  0.0006606  -7.747
##                                                     Pr(>|t|)
## (Intercept)                                         0.039790 *
## journalCommunication, Culture and Critique           < 2e-16 ***
## journalHuman Communication Research                 2.94e-10 ***
## journalJournal of Communication                     0.005860 **
## journalJournal of Computer-Mediated Communication 1.58e-15 ***
## cross_countryTrue                                   0.054745 .
## cross_typeTrue                                      0.125546
## num_race                                            0.052327 .
## num_country                                         0.040639 *
## cross_gender_and_raceTrue                           0.065218 .
## cross_gender_and_countryTrue                        0.001156 **
## cross_country_and_raceTrue                          0.312861
## cross_gender_race_and_countryTrue                   0.805949
## numberOfAuthors                                     0.639230
## first_author_genderM                                0.124838
## first_author_genderN                                0.673223
## first_author_raceBlack                              0.736384
## first_author_raceHispanic                           0.677159
## first_author_raceIndigenous                         0.364578
## first_author_raceMiddle Eastern                     0.134011
## first_author_raceWhite                              0.797529
## first_author_countryAR                              0.322711
## first_author_countryAT                              0.078123 .
## first_author_countryAU                              0.080110 .
## first_author_countryBE                              0.065497 .
## first_author_countryBR                              0.180159
## first_author_countryBS                              0.549222
## first_author_countryBY                              0.427419
## first_author_countryCA                              0.033526 *
## first_author_countryCH                              0.033764 *
## first_author_countryCL                              0.040679 *
## first_author_countryCN                              0.101294
## first_author_countryCO                              0.082203 .
## first_author_countryCR                              0.047401 *
## first_author_countryCZ                              0.000894 ***
## first_author_countryDE                              0.019113 *
## first_author_countryDK                              0.030545 *
## first_author_countryEE                              0.316655
## first_author_countryEG                              0.041196 *
## first_author_countryES                              0.032183 *
## first_author_countryFI                              0.160108
## first_author_countryFR                              0.292564
## first_author_countryGB                              0.027660 *
## first_author_countryGH                              0.136578
## first_author_countryGR                              0.168395
## first_author_countryHK                              0.559139
## first_author_countryHR                              0.163546
## first_author_countryHU                              0.196751
## first_author_countryIE                              0.270761
## first_author_countryIL                              0.060129 .
```

```
## first_author_countryIN                              0.113108
## first_author_countryIR                              0.047155 *
## first_author_countryIT                              0.082065 .
## first_author_countryJP                              0.060048 .
## first_author_countryKE                              0.040408 *
## first_author_countryKH                              0.113221
## first_author_countryKR                              0.021991 *
## first_author_countryLB                              0.024113 *
## first_author_countryLK                              0.347844
## first_author_countryMA                              0.402526
## first_author_countryMO                              0.285466
## first_author_countryMX                              0.078879 .
## first_author_countryMZ                              0.812674
## first_author_countryNG                              0.158946
## first_author_countryNL                              0.032835 *
## first_author_countryNO                              0.054459 .
## first_author_countryNZ                              0.034898 *
## first_author_countryPH                              0.014220 *
## first_author_countryPL                              0.063210 .
## first_author_countryPT                              0.412362
## first_author_countryQA                              0.024506 *
## first_author_countryRU                              0.397390
## first_author_countrySE                              0.043491 *
## first_author_countrySG                              0.077532 .
## first_author_countrySI                              0.212284
## first_author_countrySK                              0.256478
## first_author_countryTH                              0.066553 .
## first_author_countryTJ                              0.961078
## first_author_countryTR                              0.080347 .
## first_author_countryTW                              0.162991
## first_author_countryUG                              0.455929
## first_author_countryUK                              0.095642 .
## first_author_countryUS                              0.034458 *
## first_author_countryVN                              0.017908 *
## first_author_countryZA                              0.410312
## first_author_afftypeNon Education                    < 2e-16 ***
## with_us_authorsTrue                                 0.338589
## cross_race_detailsBlack only                        0.628348
## cross_race_detailscross race                        0.031245 *
## cross_race_detailsHispanic only                     0.554075
## cross_race_detailsIndigenous only                   0.752586
## cross_race_detailsMiddle Eastern only               0.069733 .
## cross_race_detailsWhite only                        0.518017
## cross_gender_detailsF only                          0.004577 **
## cross_gender_detailsM only                          1.01e-07 ***
## cross_gender_detailsN only                          0.827137
## Year.Distance.from.2022                             1.12e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6073 on 5491 degrees of freedom
## Multiple R-squared:  0.1677, Adjusted R-squared:  0.1531
## F-statistic: 11.52 on 96 and 5491 DF,  p-value: < 2.2e-16
```