Easy Way #2

This example uses the ipeds library to carry out a simple regression analysis involving data from about half a dozen different IPEDS surveys from the same year. It is intended to investigate the impact of a quarter calendar system on graduation rates. It is a rather silly analysis, however, and should be regarded as demonstration of what *can* be done with the package, not perhaps what *should* be done with the package.

We'll use these IVs in our analysis:

- Size of graduation cohort
- Selectivity of the institution
- Tuition $
- Control (public/private)
- Locale (city/town/suburb/rural)
- Student:faculty ratio
- Calendar system (semester/quarter)

```
library(ipeds)
```

```
## Loading required package: RCurl

## Loading required package: bitops

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

## Loading required package: httr
```

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------------------------- tidy

## v tibble  2.1.3     v purrr   0.3.2
## v tidyr   0.8.3     v dplyr   0.8.3
## v readr   1.3.1     v stringr 1.4.0
## v tibble  2.1.3     v forcats 0.4.0
## -- Conflicts ----------------------------------------------------------------------------- tidyverse_
## x tidyr::complete()  masks RCurl::complete()
## x dplyr::filter()    masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x dplyr::src()       masks Hmisc::src()
## x dplyr::summarize() masks Hmisc::summarize()
```

```r
library(gvlma)
library(car)
```

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

```
##
## Attaching package: 'car'
```

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

```
## The following object is masked from 'package:purrr':
##
##     some
```

```r
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
dir <- "C:\\Users\\mjs26\\Documents\\data\\downloaded"
min_school_size <- 100
```

What's available?

Helpful for a quick reminder of the various IPEDS surveys and their abbreviations.

```r
data(surveys)
surveys %>% select(c('SurveyID','Survey','Title'))
```

```
##        SurveyID                              Survey
## 1            HD       Institutional Characteristics
## 2            IC       Institutional Characteristics
## 3         IC_AY       Institutional Characteristics
## 4         IC_PY       Institutional Characteristics
## 5         FLAGS       Institutional Characteristics
## 6         EFEST                         Enrollments
## 7           EFA                         Enrollments
## 8         EFANR                         Enrollments
## 9           EFB                         Enrollments
## 10          EFC                         Enrollments
## 11          EFD                         Enrollments
## 12         EFFY                         Enrollments
## 13         EFD1                         Enrollments
## 14         EFIA                         Enrollments
## 15         EFD2                         Enrollments
## 16         EFCP                         Enrollments
## 17        FLAGS                         Enrollments
## 18          C_A                         Completions
## 19         CCIP                         Completions
## 20        FLAGS                         Completions
## 21        SAL_A        Instructional staff/Salaries
## 22        SAL_B        Instructional staff/Salaries
## 23  SAL_FACULTY        Instructional staff/Salaries
## 24    SAL_A_LT9        Instructional staff/Salaries
## 25        FLAGS        Instructional staff/Salaries
## 26        S_ABD                          Fall Staff
## 27          S_F                          Fall Staff
## 28          S_G                          Fall Staff
## 29         S_CN                          Fall Staff
## 30        FLAGS                          Fall Staff
## 31          EAP      Employees by Assigned Position
## 32        FLAGS      Employees by Assigned Position
## 33        F_F1A                             Finance
## 34         F_F2                             Finance
## 35         F_F3                             Finance
## 36           GR                    Graduation Rates
## 37        GR_L2                    Graduation Rates
## 38        GR200                    Graduation Rates
## 39          SFA Student Financial Aid and Net Price
## 40          ADM           Admission and Test Scores
## 61        DRVIC       Institutional Characteristics
## 71    ICMISSION       Institutional Characteristics
## 81  CUSTOMCGIDS       Institutional Characteristics
## 101      DRVADM                          Admissions
## 131     DRVEF12                 12-month Enrollment
## 141          EF                     Fall Enrollment
## 191    EFA_DIST                     Fall Enrollment
## 201       DRVEF                     Fall Enrollment
## 221         C_B                         Completions
## 231         C_C                         Completions
## 241        CDEP                         Completions
## 251        DRVC                         Completions
## 311 GR_PELL_SSL                    Graduation Rates
## 331       DRVGR                    Graduation Rates
## 341          OM                    Outcome Measures
## 351       DRVOM                    Outcome Measures
## 391        DRVF                             Finance
## 41       SAL_IS                     Human Resources
## 42      SAL_NIS                     Human Resources
## 43         S_OC                     Human Resources
## 44        S_SIS                    Human Resources
## 45         S_IS                    Human Resources
## 46         S_NH                    Human Resources
## 47        DRVHR                    Human Resources
## 48           AL                 Academic Libraries
## 49        DRVAL                 Academic Libraries
##
## 1
## 2                                                                                                      I
## 3
## 4
## 5
## 6
## 7
## 8                                                                                                     Ra
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16                                                                                                   Ma
## 17
## 18                                                                                         Awards/degree
## 19                                                                                                 Award
## 20
## 21                                                                                              Salaries
## 22
## 23                                                                    Tenure status of full-time instructi
## 24                                                                                   Number of full-time i
## 25
## 26                                         Employees by primary occupation, salary categories, race/ethnic
## 27  Full-time instruction/research/public service staff, by tenure status, academic rank, race/ethnic
## 28                                                            New hires by primary occupation, race/ethnic
## 29            Employees by primary occupation, race/ethnicity, and gender (Degree-granting instituti
## 30
## 31
## 32
## 33
## 34
## 35
## 36
## 37
## 38
## 39
## 40
## 61
## 71
## 81
## 101
## 131
## 141
## 191
## 201
## 221                                                                           Number of students rec
## 231                                       Number of students receiving awards/degrees, by award level
## 241                                                     Number of programs offered and number of prog
## 251
## 311 Graduation rate data for Pell Grant and Subsidized Stafford loan recipients, 150% of normal time
## 331                                       Frequently used derived variables (GR) 150% of normal time
## 341                         Award and enrollment data at four, six and eight years for four entering
## 351                                                      Frequently used derived variables (OM) Awar
## 391
## 41                                                        Number and salary outlays for full-time no
## 42                                                                      Number and salary outlays fo
## 43                                                                                          Full- an
## 44                                                                Full-time instructional staff, by
## 45                                     Full-time instructional staff, by faculty and tenure status,
## 46                                                                           New hires by occupation
## 47
## 48
## 49
```

Get the data

We're going to grab the survey files one at a time, merging (joining) them together by unit id as we go. The three IC files are first up:

```
directory <- ipeds_survey(table='HD',year=2017, dir=dir)
names(directory) <- tolower(names(directory))

charges <- ipeds_survey('IC_AY', year=2017, dir=dir)
names(charges) <- tolower(names(charges))

charges = charges[,c('unitid',
 'tuition1', 'fee1', 'hrchg1',  #In-district average tuition for full-time undergraduates
 'tuition2', 'fee2', 'hrchg2',  #In-state average tuition for full-time undergraduates
 'tuition3', 'fee3', 'hrchg3',  #Out-of-state average tuition for full-time undergraduates
 'tuition5', 'fee5', 'hrchg5', #In-district average tuition full-time graduates
 'tuition6', 'fee6', 'hrchg6', #In-state average tuition full-time graduates
 'tuition7', 'fee7', 'hrchg7')] #Out-of-state average tuition full-time graduates

dirCharges = merge(charges, directory, by='unitid', all.x=TRUE)

ic <- ipeds_survey(table='IC',year=2017, dir=dir)
names(ic) <- tolower(names(ic))

dirCharges <- merge(dirCharges, ic, by='unitid', all.x=TRUE)
```

Then Admissions:

```
admissions <- ipeds_survey(table='ADM',year=2017, dir=dir)
names(admissions) <- tolower(names(admissions))
```

Graduation rates:

```r
gradrates <- (ipeds_survey('GR',year=2017, dir=dir))
names(gradrates) <- tolower(names(gradrates))
gradrates <- gradrates[which(gradrates$grtype %in% c(2,3)),]

# extract the 150% graduation rate
theRates <- dcast(gradrates, unitid ~ grtype, value.var = 'grtotlt')
names(theRates) <- c('unitid','adjusted_cohort','completers')
theRates$rate <- theRates$completers/theRates$adjusted_cohort
```

Eliminate any schools with missing graduation rates:

```r
theRates <- theRates[which(!is.na(theRates$rate)),]
```

And any with less than 100 in the grad rate cohort

```r
d1 <- merge(dirCharges, theRates, by='unitid', all.y=TRUE)
d1 <- d1[which(d1$calsys %in% c(1,2)),]
d1$calsys <- as.factor(d1$calsys)
levels(d1$calsys) <- c('Semester','Quarter')
d1 <- d1[which(d1$adjusted_cohort > min_school_size),]
```

IPEDS Admissions gives us selectivity.

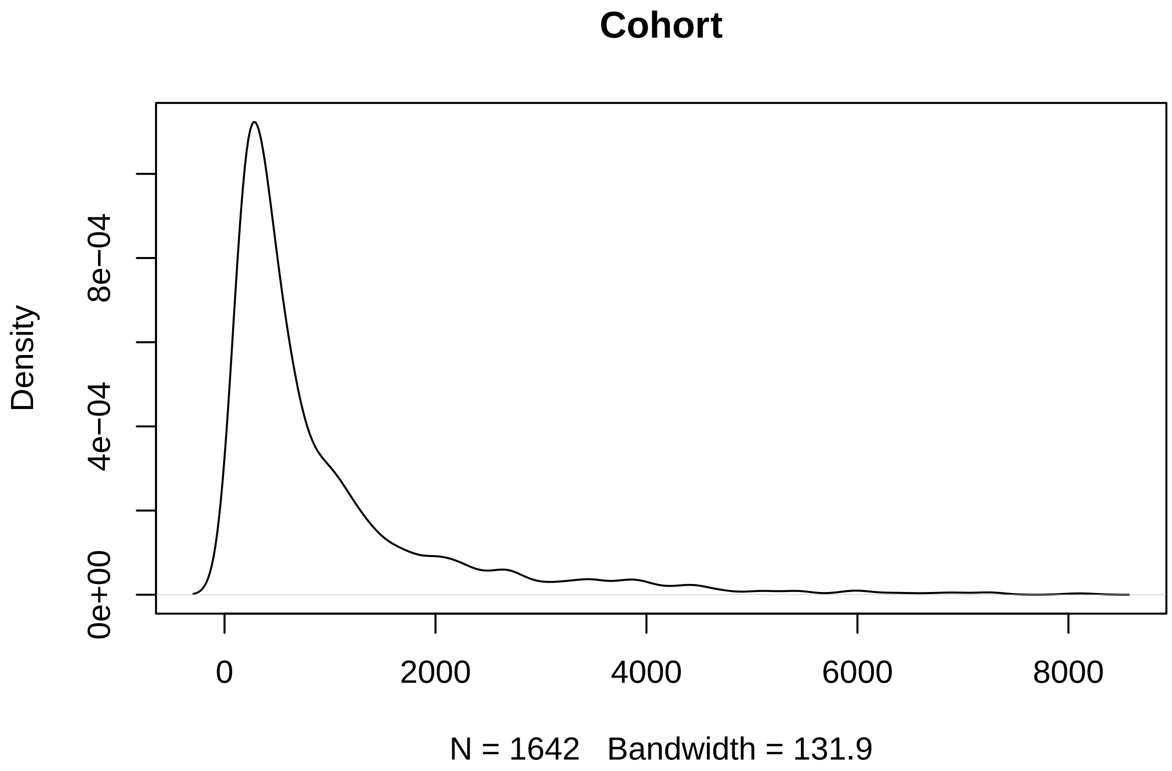
```r
d1 <- merge(d1, admissions, by='unitid', all.x=TRUE)
d1$select <- d1$admssn / d1$applcn
```

Here's Fall Enrollment, which is where student:faculty ratio lives.

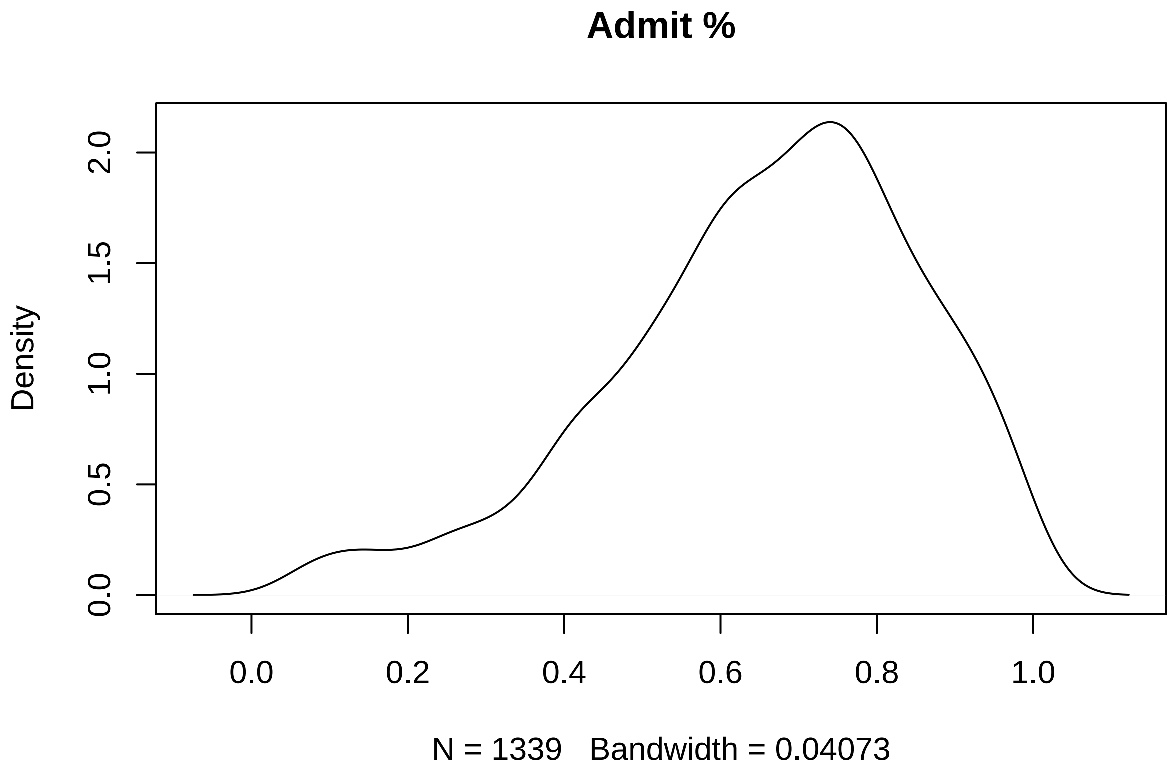
```r
fallenr <- ipeds_survey(table='EFD', year=2017, dir=dir)
names(fallenr) <- tolower(names(fallenr))
d1 <- merge(d1, fallenr, by='unitid', all.x=TRUE)
d1 <- d1[which(!is.na(d1$stufacr)),] # remove any schools with missing s:f ratio
```

That's all the data we need. Do our continuous variables have sensible shapes?

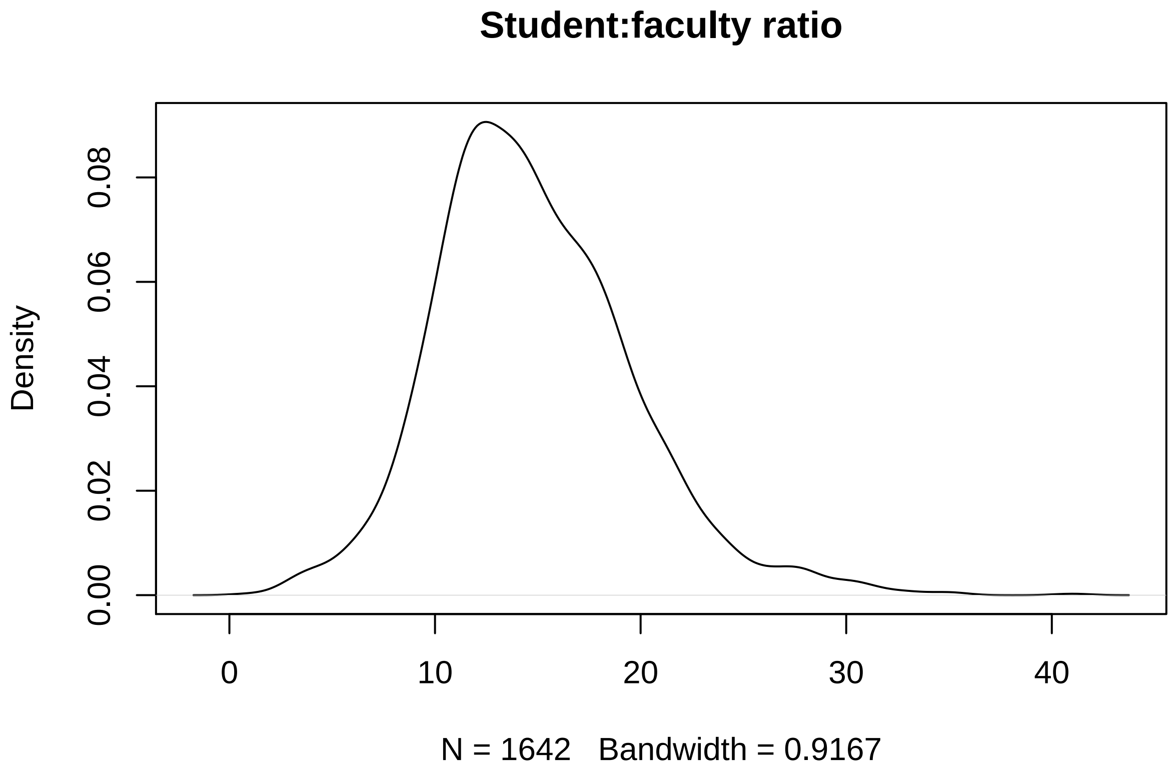
```r
plot(density(d1$adjusted_cohort), main="Cohort")
```



```
plot(density(d1[which(!is.na(d1$select)),]$select), main="Admit %")
```



```
plot(density(d1$stufacr), main="Student:faculty ratio")
```



```r
table(d1$calsys,d1$control)
```

```
## 
##               1   2   3
##   Semester 625 802  52
##   Quarter   50  29  84
```

This code chunk recodes IPEDS' locale codes into something more readable.

```r
d1$locale2 <- substr(d1$locale,1,1)
d1$locale2 <- as.factor(d1$locale2)
levels(d1$locale2) <- c('City','Town','Suburb','Rural')
table(d1$locale2, d1$locale)
```

```
## 
##           11  12  13  21  22  23  31  32  33  41  42  43
##   City   366 209 234   0   0   0   0   0   0   0   0   0
##   Town     0   0   0 315  51  36   0   0   0   0   0   0
##   Suburb   0   0   0   0   0   0  57 163 112   0   0   0
##   Rural    0   0   0   0   0   0   0   0   0  57  25  17
```

Model and output

```
theLM <- lm(rate ~ calsys + as.integer(tuition1) + control + select + stufacr + locale2  + adjusted_coh
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'family' will be disregarded
```

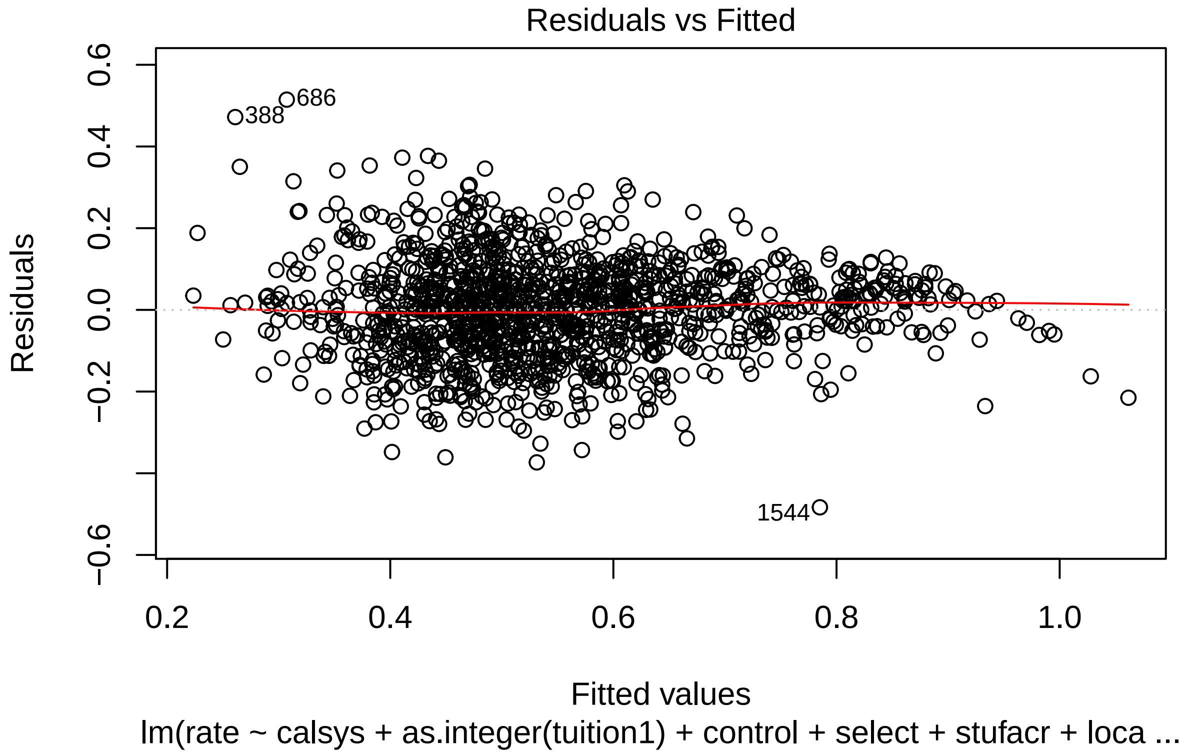
```
summary(theLM)
```

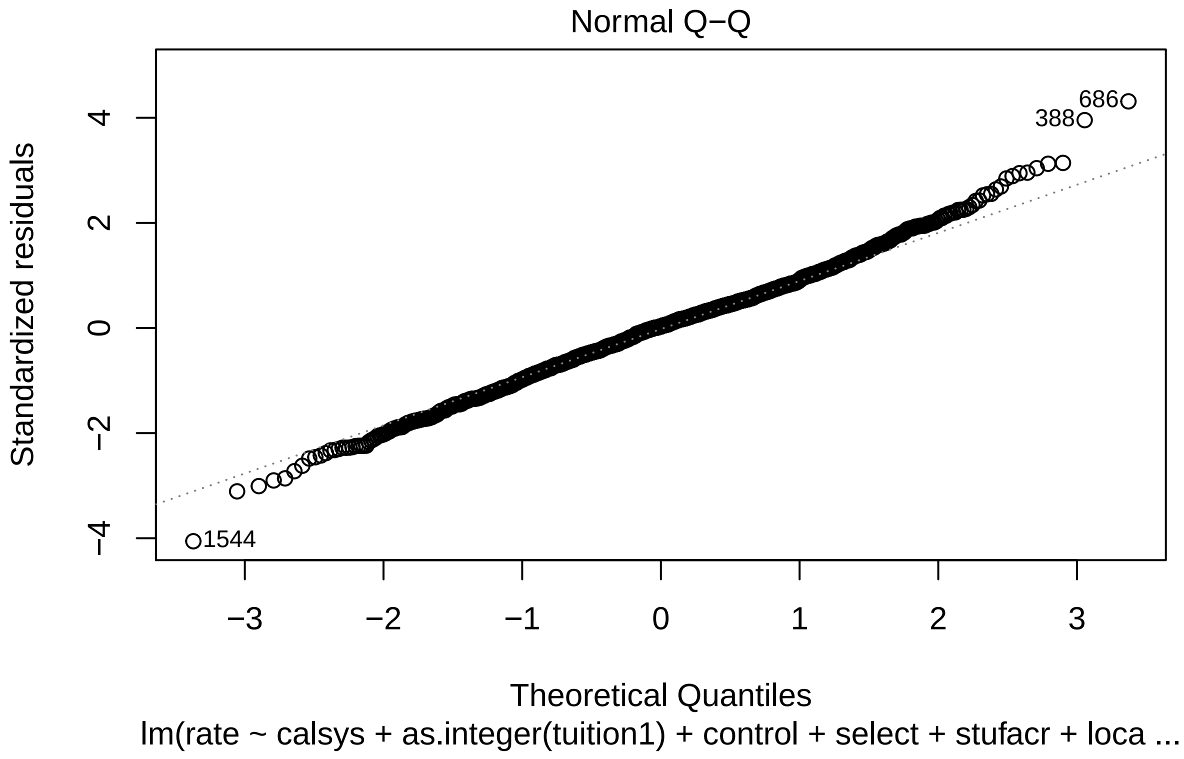
```
##
## Call:
## lm(formula = rate ~ calsys + as.integer(tuition1) + control +
##     select + stufacr + locale2 + adjusted_cohort, data = d1,
##     family = gaussian)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48337 -0.07656  0.00349  0.07163  0.51467
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           5.297e-01  2.572e-02  20.592  < 2e-16 ***
## calsysQuarter         5.952e-02  1.607e-02   3.703 0.000222 ***
## as.integer(tuition1)  9.753e-06  3.946e-07  24.718  < 2e-16 ***
## control              -8.916e-02  9.550e-03  -9.336  < 2e-16 ***
## select               -1.083e-01  1.777e-02  -6.092 1.45e-09 ***
## stufacr              -3.071e-03  1.033e-03  -2.972 0.003016 **
## locale2Town           3.518e-02  8.279e-03   4.249 2.30e-05 ***
## locale2Suburb         1.107e-02  8.895e-03   1.245 0.213466
## locale2Rural         -2.757e-02  1.537e-02  -1.794 0.073044 .
## adjusted_cohort       5.909e-05  3.471e-06  17.025  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1203 on 1326 degrees of freedom
##   (306 observations deleted due to missingness)
## Multiple R-squared:  0.5646, Adjusted R-squared:  0.5617
## F-statistic: 191.1 on 9 and 1326 DF,  p-value: < 2.2e-16
```

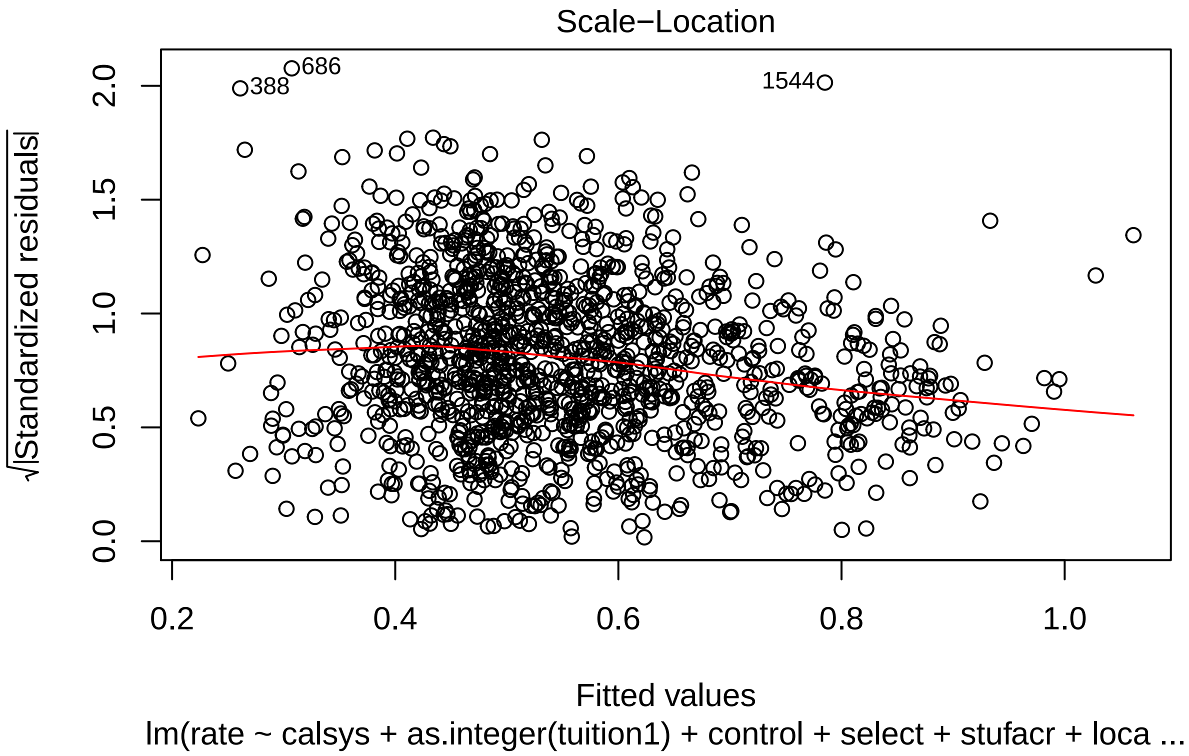
```
library(gvlma)
gvlma(theLM)
```

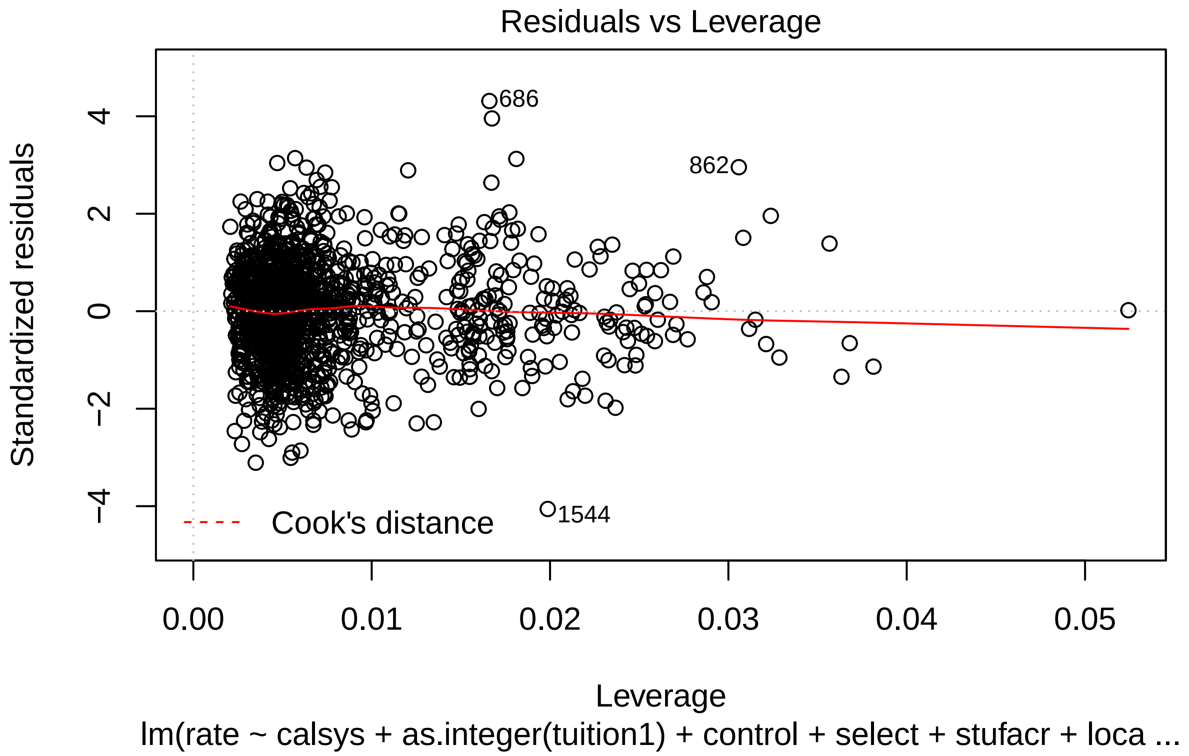
```
##
## Call:
## lm(formula = rate ~ calsys + as.integer(tuition1) + control +
##     select + stufacr + locale2 + adjusted_cohort, data = d1,
##     family = gaussian)
##
## Coefficients:
##          (Intercept)         calsysQuarter  as.integer(tuition1)
##            5.297e-01             5.952e-02             9.753e-06
##                control               select              stufacr
##             -8.916e-02           -1.083e-01           -3.071e-03
##            locale2Town         locale2Suburb         locale2Rural
##              3.518e-02            1.107e-02           -2.757e-02
##        adjusted_cohort
##              5.909e-05
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = theLM)
##
##                      Value   p-value                   Decision
## Global Stat         26.007 3.155e-05 Assumptions NOT satisfied!
## Skewness             1.136 2.864e-01    Assumptions acceptable.
## Kurtosis            21.549 3.449e-06 Assumptions NOT satisfied!
## Link Function        3.132 7.677e-02    Assumptions acceptable.
## Heteroscedasticity   0.189 6.637e-01    Assumptions acceptable.
```

```
plot(theLM)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(rate ~ calsys + as.integer(tuition1) + control + select + stufacr + loca ...
686
388
1544
−4
−2
0
2
4
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
0.2
0.4
0.6
0.8
1.0
388
686
1544
Fitted values
lm(rate ~ calsys + as.integer(tuition1) + control + select + stufacr + loca ...



```r
vif(theLM)
```

```
##                           GVIF Df GVIF^(1/(2*Df))
## calsys                1.022104  1        1.010992
## as.integer(tuition1) 2.943326  1        1.715612
## control               2.451026  1        1.565575
## select                1.117419  1        1.057080
## stufacr               1.871449  1        1.368009
## locale2               1.141902  3        1.022362
## adjusted_cohort       1.592042  1        1.261761
```