MEM5220 - R Econometrics Evaluation

YOUR NAME HERE

This first R econometrics self-evaluation exercise is focused on data cleaning, data manipulation, plotting and estimating and interpreting simple linear regression models. You can use **any** additional packages for answering the questions.

Packages I have used to solve the exercises:

```
library(tidyverse)
library(stargazer)
library(huxtable)
library(broom)
```

Note:

- This assignemtn has to be solved in this R markdown and you should be able to "knit" the document without errors
- Fill out your name in "yaml" - block on top of this document
- Use the R markdown syntax:
  - Write your code in code chunks
  - Write your explanations including the equations in markdown syntax
- If you have an error in your code use # to comment the line out where the error occurs but do not delete the code itself so I

You will be working with the dataset `Caschool` from the **Ecdat** package, the dataset `Wage1` from **wooldrige** package. In the last step, you will be working with a simulated dataset.

Q1:

Load the dataset `Caschool` from the **Ecdat** package.

The Caschooldataset contains the average test scores of 420 elementary schools in California along with some additional information.

```
# install.packages("Ecdat")
library("Ecdat") # Attach the Ecdat library
data("Caschool", package = "Ecdat")
```

What are the dimensions of the `Caschool` dataset?

A:

```r
dim(Caschool)
```

```
[1] 420  17
```

Q2:

Display the structure of the Caschool dataset. Which variable has values encoded as factors?

A:

```r
str(Caschool)
```

```
'data.frame':   420 obs. of  17 variables:
 $ distcod : int  75119 61499 61549 61457 61523 62042 68536 63834 62331 67306 ...
 $ county  : Factor w/ 45 levels "Alameda","Butte",..: 1 2 2 2 2 6 29 11 6 25 ...
 $ district: Factor w/ 409 levels "Ackerman Elementary",..: 362 214 367 132 270 53 152 3
 $ grspan  : Factor w/ 2 levels "KK-06","KK-08": 2 2 2 2 2 2 2 2 2 1 ...
 $ enrltot : int  195 240 1550 243 1335 137 195 888 379 2247 ...
 $ teachers: num  10.9 11.1 82.9 14 71.5 ...
 $ calwpct : num  0.51 15.42 55.03 36.48 33.11 ...
 $ mealpct : num  2.04 47.92 76.32 77.05 78.43 ...
 $ computer: int  67 101 169 85 171 25 28 66 35 0 ...
 $ testscr : num  691 661 644 648 641 ...
 $ compstu : num  0.344 0.421 0.109 0.35 0.128 ...
 $ expnstu : num  6385 5099 5502 7102 5236 ...
 $ str     : num  17.9 21.5 18.7 17.4 18.7 ...
 $ avginc  : num  22.69 9.82 8.98 8.98 9.08 ...
 $ elpct   : num  0 4.58 30 0 13.86 ...
 $ readscr : num  692 660 636 652 642 ...
 $ mathscr : num  690 662 651 644 640 ...
```

County, district and grspan are encoded as factors.

Q3:

Provide a summary statistic of the data

A:

```r
summary(Caschool)
```

```
    distcod              county                               district
 Min.   :61382   Sonoma     : 29   Lakeside Union Elementary:  3
 1st Qu.:64308   Kern       : 27   Mountain View Elementary :  3
 Median :67760   Los Angeles: 27   Jefferson Elementary     :  2
 Mean   :67473   Tulare     : 24   Liberty Elementary       :  2
```

```
 3rd Qu.:70419   San Diego  : 21   Ocean View Elementary    :  2  
 Max.   :75440   Santa Clara: 20   Pacific Union Elementary :  2  
                 (Other)    :272   (Other)                  :406  
  grspan       enrltot          teachers          calwpct      
 KK-06: 61   Min.   :   81.0   Min.   :   4.85   Min.   : 0.000  
 KK-08:359   1st Qu.:  379.0   1st Qu.:  19.66   1st Qu.: 4.395  
             Median :  950.5   Median :  48.56   Median :10.520  
             Mean   : 2628.8   Mean   : 129.07   Mean   :13.246  
             3rd Qu.: 3008.0   3rd Qu.: 146.35   3rd Qu.:18.981  
             Max.   :27176.0   Max.   :1429.00   Max.   :78.994  

    mealpct          computer         testscr         compstu       
 Min.   :  0.00   Min.   :   0.0   Min.   :605.5   Min.   :0.00000  
 1st Qu.: 23.28   1st Qu.:  46.0   1st Qu.:640.0   1st Qu.:0.09377  
 Median : 41.75   Median : 117.5   Median :654.5   Median :0.12546  
 Mean   : 44.71   Mean   : 303.4   Mean   :654.2   Mean   :0.13593  
 3rd Qu.: 66.86   3rd Qu.: 375.2   3rd Qu.:666.7   3rd Qu.:0.16447  
 Max.   :100.00   Max.   :3324.0   Max.   :706.8   Max.   :0.42083  

    expnstu          str            avginc           elpct       
 Min.   :3926   Min.   :14.00   Min.   : 5.335   Min.   : 0.000  
 1st Qu.:4906   1st Qu.:18.58   1st Qu.:10.639   1st Qu.: 1.941  
 Median :5215   Median :19.72   Median :13.728   Median : 8.778  
 Mean   :5312   Mean   :19.64   Mean   :15.317   Mean   :15.768  
 3rd Qu.:5601   3rd Qu.:20.87   3rd Qu.:17.629   3rd Qu.:22.970  
 Max.   :7712   Max.   :25.80   Max.   :55.328   Max.   :85.540  

    readscr         mathscr     
 Min.   :604.5   Min.   :605.4  
 1st Qu.:640.4   1st Qu.:639.4  
 Median :655.8   Median :652.5  
 Mean   :655.0   Mean   :653.3  
 3rd Qu.:668.7   3rd Qu.:665.9  
 Max.   :704.0   Max.   :709.5  
```

Q4:

What are the names of the variables in the dataset?

A:

```r
names(Caschool)
```

```
 [1] "distcod"  "county"   "district" "grspan"   "enrltot"  "teachers"
```

```
 [7] "calwpct"  "mealpct"  "computer" "testscr"  "compstu"  "expnstu"
[13] "str"      "avginc"   "elpct"    "readscr"  "mathscr"
```

Q5:

How many unique observations are available in the variable "county"

A:

```r
unique(Caschool$county)
```

```
 [1] Alameda         Butte           Fresno          San Joaquin
 [5] Kern            Sacramento      Merced          Tulare
 [9] Los Angeles     Imperial        Monterey        San Diego
[13] San Bernardino  San Mateo       Ventura         Riverside
[17] Santa Clara     Madera          Santa Barbara   Orange
[21] Kings           Sonoma          Contra Costa    Humboldt
[25] Siskiyou        Lake            Sutter          Mendocino
[29] San Benito      Shasta          Tehama          Stanislaus
[33] Tuolumne        El Dorado       Placer          Glenn
[37] Lassen          Santa Cruz      Nevada          Calaveras
[41] Marin           San Luis Obispo Inyo            Trinity
[45] Yuba
45 Levels: Alameda Butte Calaveras Contra Costa El Dorado Fresno ... Yuba
```

Q6:

Summarize the mean number of students grouped by county.

A:

```r
Caschool %>%
  group_by(county) %>%
  summarise(mean_count = mean(enrltot)) %>%
  arrange(desc(mean_count))
```

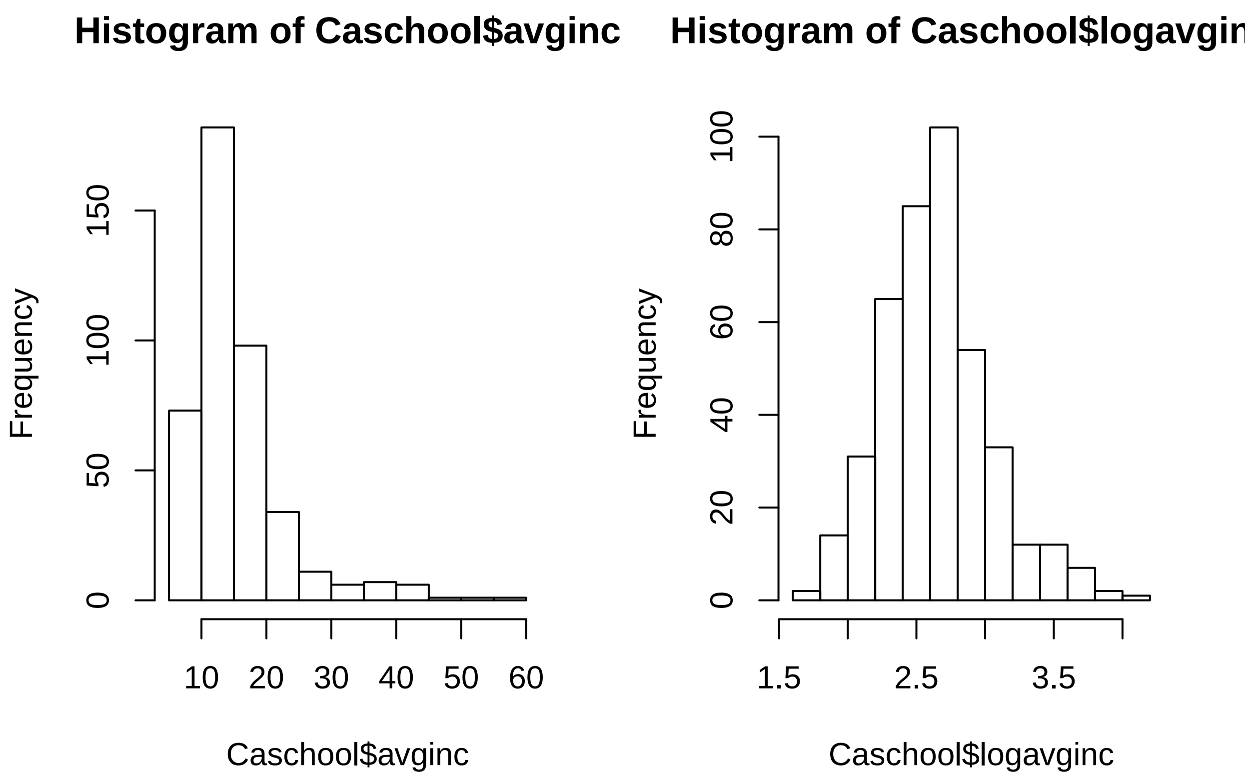
Q7:

Calculate the log of average income from of the Caschool dataset. Call the variable **logavginc** and add this variable to the dataset. Then, plot a histogram of the average income vs. a histogram of log average income. What do you observe?

A:

```
Caschool$logavginc <- log(Caschool$avginc)
```

```
par(mfrow=c(1,2))
hist(Caschool$avginc)
hist(Caschool$logavginc)
```



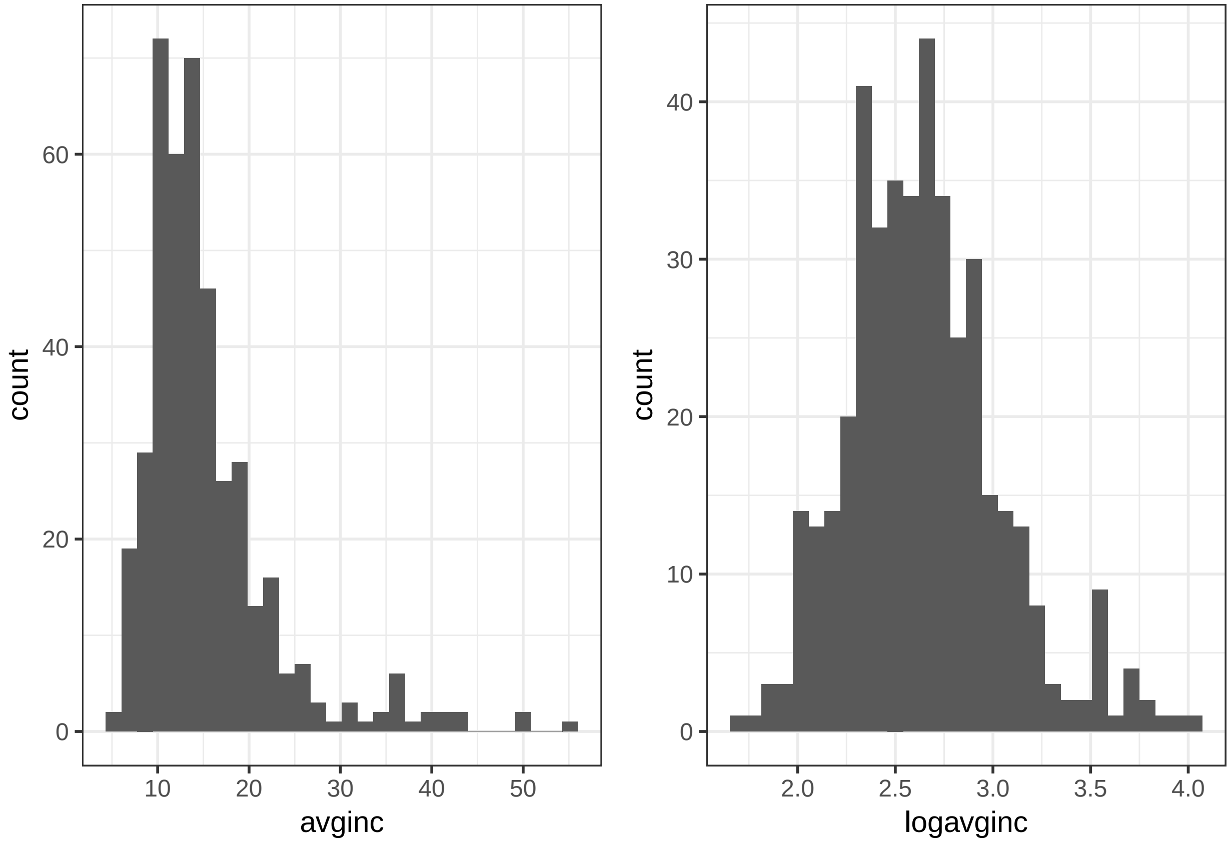

```
library(gridExtra)
```

```
Attaching package: 'gridExtra'

The following object is masked from 'package:dplyr':

    combine
```

```
p1 <- ggplot(Caschool, aes(avginc)) +
  geom_histogram(show.legend = FALSE) +
  theme_bw()
p2 <- ggplot(Caschool, aes(logavginc)) +
  geom_histogram(show.legend = FALSE) +
  theme_bw()
grid.arrange(p1, p2,
             ncol = 2)
```

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

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



Average income is clearly leftward-skewed. The log of averge income looks more like a normal distribution.

Q8:

We want to create now a subset of counties that have the ten highest district average income and that have the ten lowest district average income. Call this subset *Caschool_lowhighincome*.

Hint: One way is the create two subsets (eg. Cascholl_highincome and Caschool_lowincome and the use the `rbind()` function to bind them together.).

A:

```r
Caschool_highincome <- Caschool %>%
  arrange(desc(avginc)) %>%
  head(10)

Caschool_lowincome <- Caschool %>%
  arrange((avginc)) %>%
  head(10)
```

```r
Caschool_lowhighincome <- rbind(Caschool_highincome,Caschool_highincome)
```

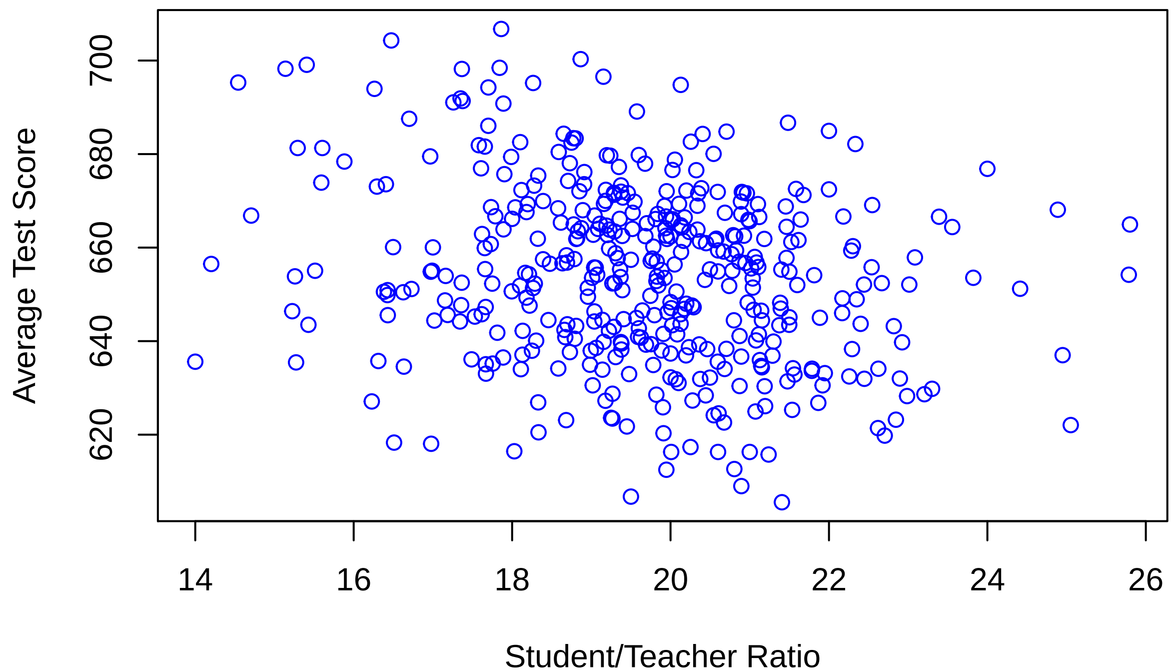
Q9:

Let us test wether a high student/teacher ratio will be associated with higher-than-average test scores for the school? Create a scatter plot for the full dataset (*Caschool*) for the variables **testscr** and **str**. You can plot either in base R or use ggplot2.

A:

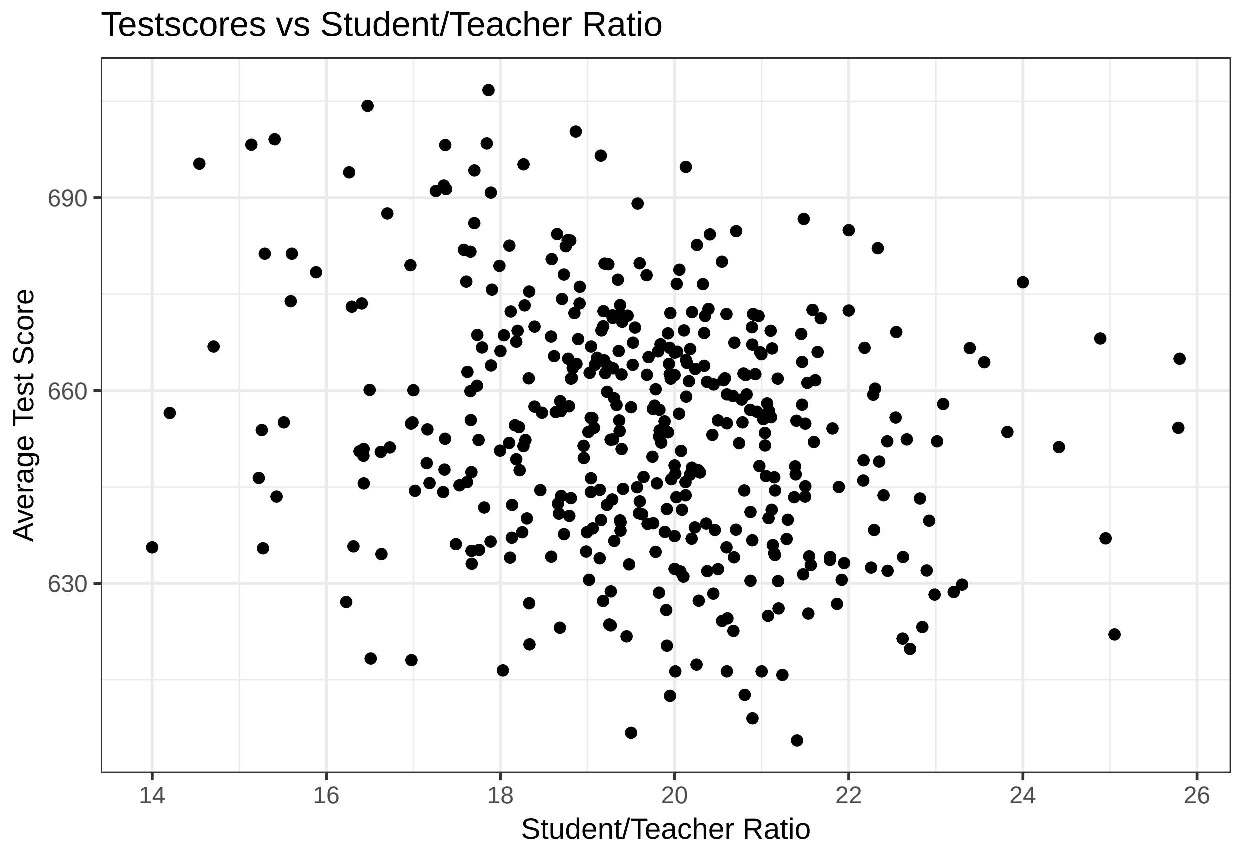
Base R-style

```r
plot(formula = testscr ~ str,
     data = Caschool,
     xlab = "Student/Teacher Ratio",
     ylab = "Average Test Score", pch = 21, col = 'blue')
```



ggplot2-style

```r
ggplot(mapping = aes(x = str, y = testscr), data = Caschool) + # base plot
  geom_point() + # add points
scale_y_continuous(name = "Average Test Score") +
  scale_x_continuous(name = "Student/Teacher Ratio") +
theme_bw() + ggtitle("Testscores vs Student/Teacher Ratio")
```



Q10:

Suppose a policymaker is interested in the following linear model:

$$testscr_i = \beta_0 + \beta_1 \times str_i + \epsilon_i \tag{1}$$

Where $(testscr)_i$ is the average test score for a given school and $(str)_i$ is the Student/Teacher Ratio (i.e. the average number of students per teacher) in the same school i.

Estimate the specified linear model. Is the estimated relationship between a school's Student/Teacher Ratio and its average test results postitive or negative?

A:

```
fit_single <- lm(formula = testscr ~ str, data = Caschool)
summary(fit_single)
```

```
Call:
lm(formula = testscr ~ str, data = Caschool)

Residuals:
```

```
    Min      1Q  Median      3Q     Max 
-47.727 -14.251   0.483  12.822  48.540 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 698.9330     9.4675  73.825  < 2e-16 ***
str          -2.2798     0.4798  -4.751 2.78e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.58 on 418 degrees of freedom
Multiple R-squared:  0.05124,	Adjusted R-squared:  0.04897 
F-statistic: 22.58 on 1 and 418 DF,  p-value: 2.783e-06
```

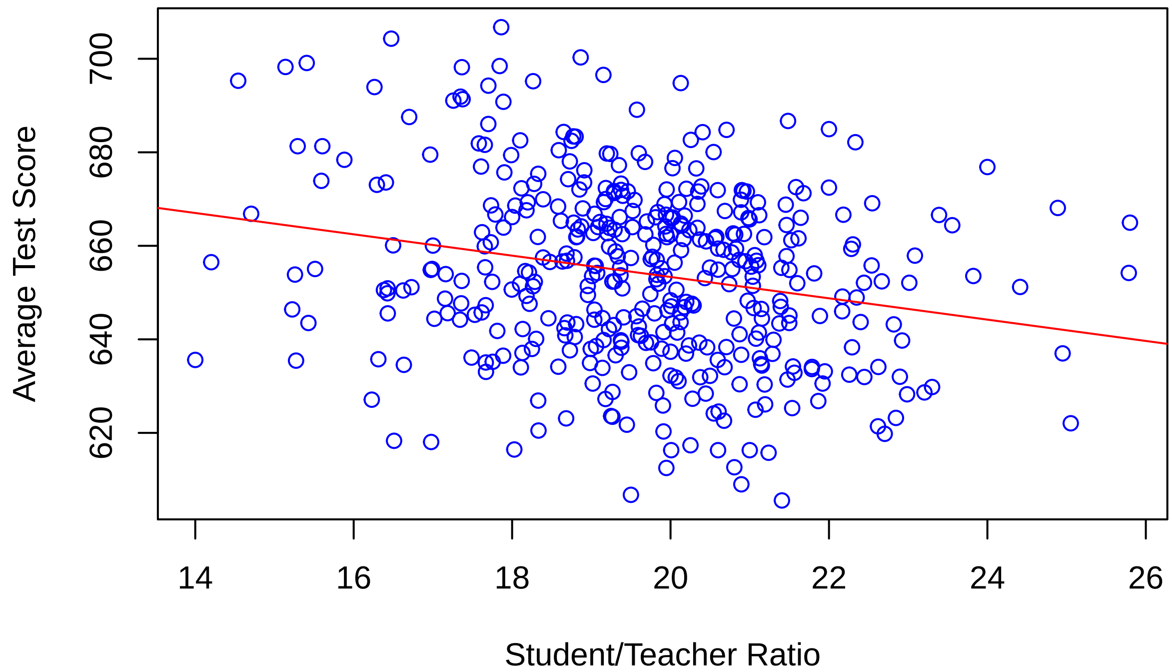
Q11:

Now, plot the regression line for the model we have just estimated. Again, you can use either base R or ggplot2-style.

A:

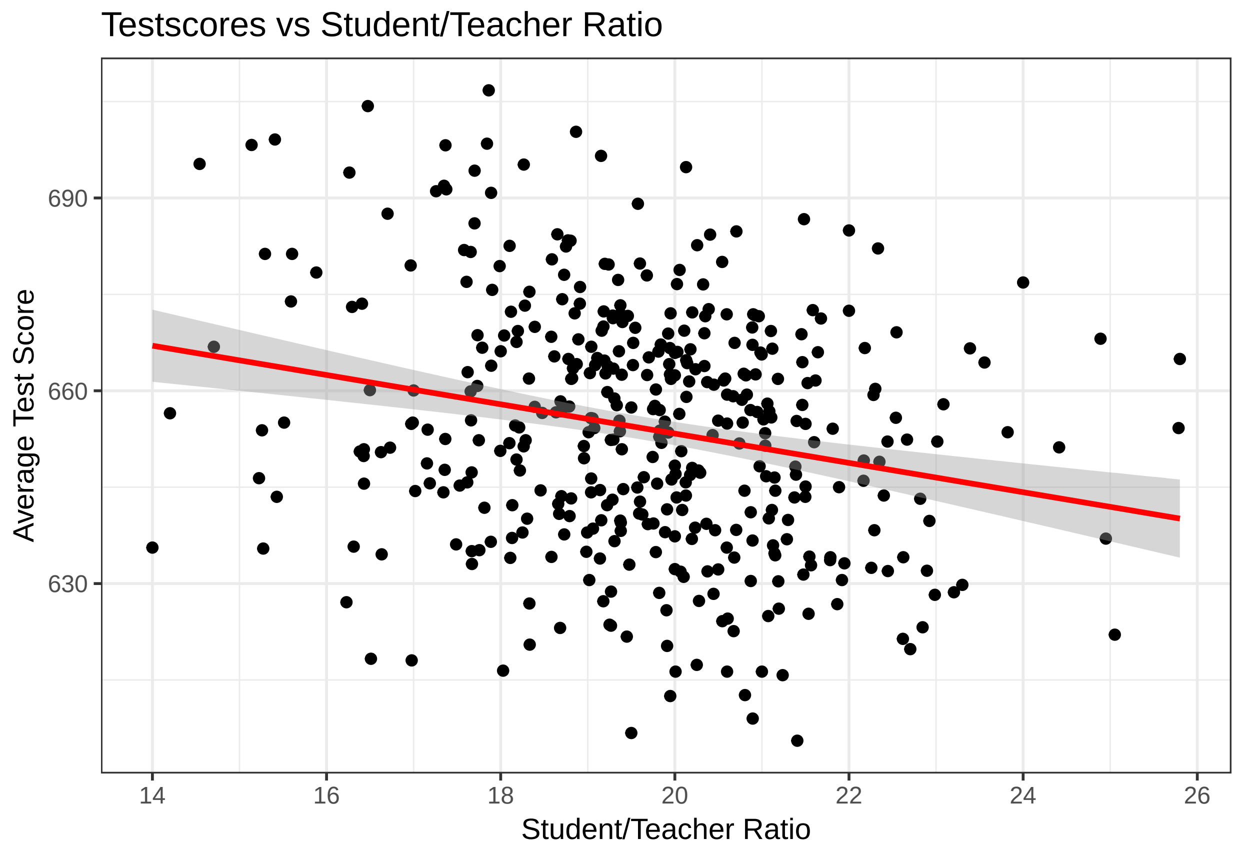
Base R-style

```r
fit_california <- lm(formula = testscr ~ str, data = Caschool)
plot(formula = testscr ~ str,
     data = Caschool,
     xlab = "Student/Teacher Ratio",
     ylab = "Average Test Score", pch = 21, col = 'blue')# same plot as before
abline(fit_california, col = 'red') # add regression line
```



ggplot2-style

```
ggplot(mapping = aes(x = str, y = testscr), data = Caschool) + # base plot
  geom_point() + # add points
  geom_smooth(method = "lm", size=1, color="red") + # add regression line
  scale_y_continuous(name = "Average Test Score") +
  scale_x_continuous(name = "Student/Teacher Ratio") +
  theme_bw() + ggtitle("Testscores vs Student/Teacher Ratio")
```



Q12:

Let us extend our example of student test scores by adding families' average income to our previous model:

$$testscr_i = \beta_0 + \beta_1 \times str_i + \beta_2 \times avginc_i + \epsilon_i \tag{2}$$

A:

```
fit_multivariate <- lm(formula = "testscr ~ str + avginc", data = Caschool)
summary(fit_multivariate)
```

```
Call:
lm(formula = "testscr ~ str + avginc", data = Caschool)

Residuals:
    Min      1Q  Median      3Q     Max 
-39.608  -9.052   0.707   9.259  31.898 

Coefficients:
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 638.72915    7.44908  85.746   <2e-16 ***
str          -0.64874    0.35440  -1.831   0.0679 .
avginc        1.83911    0.09279  19.821   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.35 on 417 degrees of freedom
Multiple R-squared:  0.5115,    Adjusted R-squared:  0.5091
F-statistic: 218.3 on 2 and 417 DF,  p-value: < 2.2e-16
```

Q13:

Assume know that "str" depends also on the value of yet another regressor, "avginc". Estimate the following model. Compare the sign of the estimate of β_2 and β_3. Interpret the results.

$$testscr_i = \beta_0 + \beta_1 \times str_i + \beta_2 \times avginc_i + \beta_3(str_i \times avginc_i) + \epsilon_i \tag{3}$$

A:

```r
fit_inter = lm(formula = testscr ~ str + avginc + str*avginc, data = Caschool)
summary(fit_inter)
```

```
Call:
lm(formula = testscr ~ str + avginc + str * avginc, data = Caschool)

Residuals:
    Min      1Q  Median      3Q     Max
-41.346  -9.260   0.209   8.736  33.368

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 689.47473   14.40894  47.850  < 2e-16 ***
str          -3.40957    0.75980  -4.487 9.34e-06 ***
avginc       -1.62388    0.85214  -1.906   0.0574 .
str:avginc    0.18988    0.04646   4.087 5.24e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.1 on 416 degrees of freedom
Multiple R-squared:  0.5303,    Adjusted R-squared:  0.527
F-statistic: 156.6 on 3 and 416 DF,  p-value: < 2.2e-16
```

We observe also that the estimate of β_2 changes signs and becomes negative, while the interaction effect β_3 is positive.

This means that an increase in str reduces average student scores (more students per teacher make it harder to teach effectively); that an increase in average district income in isolation actually reduces scores; and that the interaction of both increases scores (more students per teacher are actually a good thing for student performance in richer areas).

Q14:

In question 10, 12 and 13, you have fitted 3 models. Report the regression results, the number of observations, the Akaike information criterion and the model fit (adj. R^2) in formatted table regression output table. You can use for example the **stargazer** or the **huxtable** package. Which model fits the data best?

stargazer package:

```
library(stargazer)
invisible(stargazer(
  list(fit_single,
       fit_multivariate,
      fit_inter)
  ,keep.stat = c("n", "adj.rsq", "aic", "bic"), type = "text", style = "ajps"))# to have
```

```
-----------------------------------------------
                  testscr    testscr      NA
                  Model 1    Model 2    Model 3
-----------------------------------------------
str             -2.280***   -0.649*   -3.410***
                 (0.480)    (0.354)    (0.760)
avginc                      1.839***   -1.624*
                            (0.093)    (0.852)
str:avginc                             0.190***
                                       (0.046)
Constant        698.933*** 638.729*** 689.475***
                 (9.467)    (7.449)    (14.409)
N                  420        420        420
Adj. R-squared    0.049      0.509      0.527
-----------------------------------------------
***p < .01; **p < .05; *p < .1
```

huxtable package:

```
library(huxtable)
huxreg(fit_single,
       fit_multivariate,
```

```
        fit_inter,
        statistics = c("nobs", "adj.r.squared", "AIC", "BIC"))
```

The adjusted R^2 is highest for the model 3, the model that includes an interaction term. AIC and BIC, two widely used information criteria, would also select model 3, relative to each of the other models (The relatively quality of the model is maximized when the information criteria is minimized).

Q15:

This exercise focuses on the **collineartiy** problem.

Perform the following commands in R:

```
set.seed(1)
x1 <- runif(100)
x2 <- 0.5 * x1 + rnorm(100)/10
y <- 2 +2*x1 + 0.3 *x2 +rnorm(100)
```

The last line corresponds to creating a linear model in which y is a function of x_1 and x_2. Write out the form of the linear model. What are the regression coefficients?

A:

$y = 2 + 2x_1 + 0.3x_2 + \epsilon$

$\beta_0 = 2$, $\beta_1 = 2$, $\beta_3 = 0.3$

Q15:

What is the correlation between x_1 and x_2? Create a scatterplot displaying the relationship between the variables.

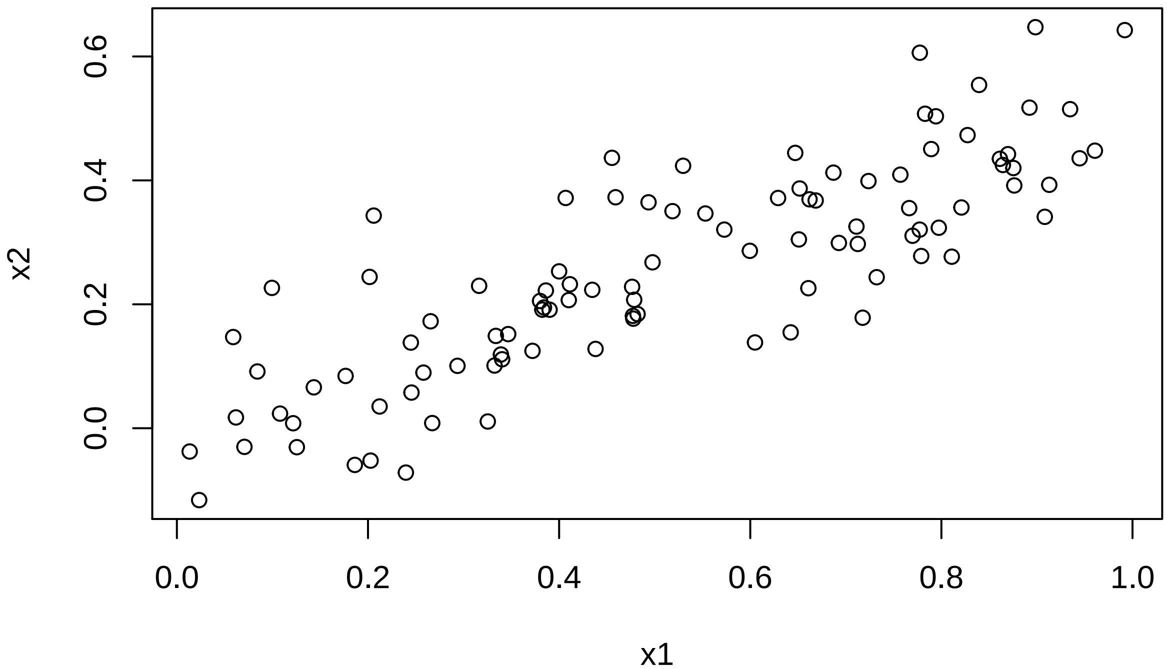
A:

```
cor(x1, x2)
```

```
[1] 0.8351212
```

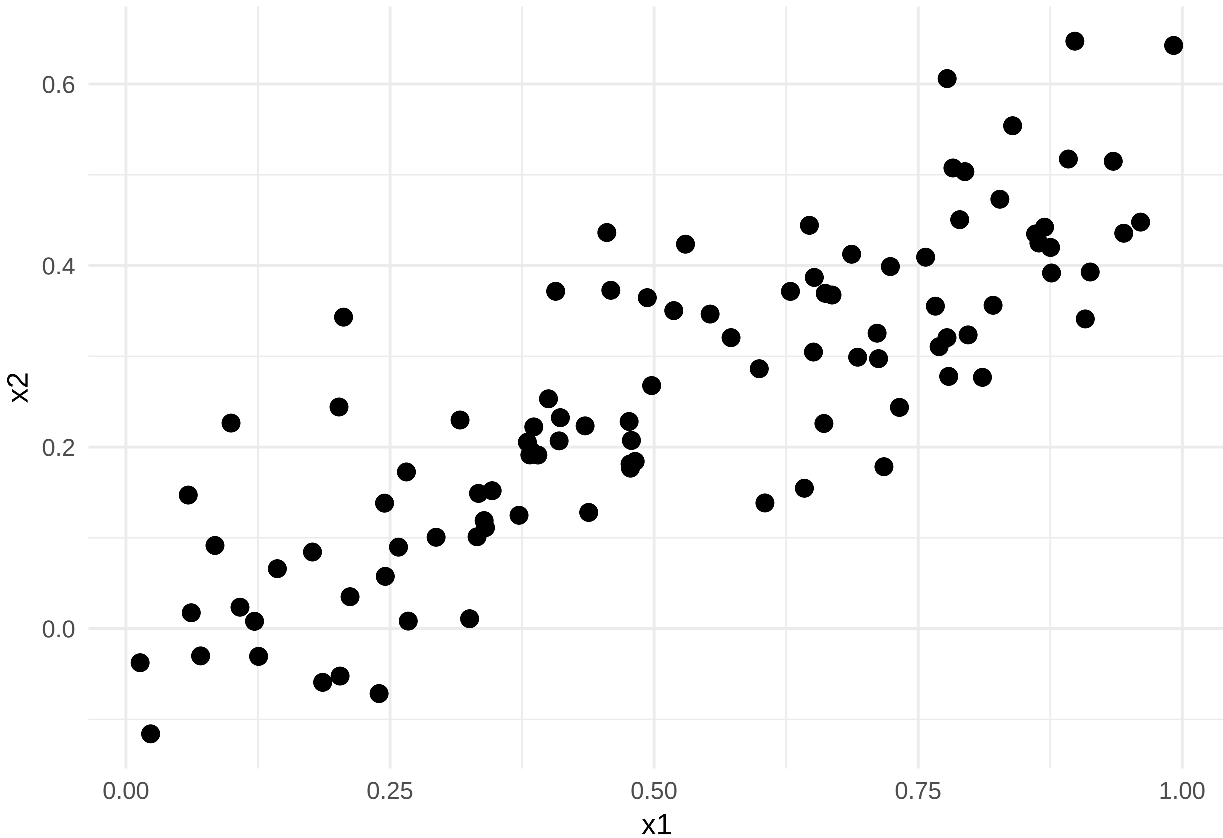
Base R style:

```
plot(x1, x2)
```



ggplot2 style:

```
d <-  data.frame(x1,x2)
ggplot(d, aes(x1, x2)) +
  geom_point(shape = 16, size = 3, show.legend = FALSE) +
  theme_minimal()
```



Q16:

Using this data, fit a least squares regression to predict y using x_1 and x_2. Describe the results obtained. What are $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$? How do these relate to the true β_0, β_1 and β_2? Can you reject the null hypothesis $H_0 : \beta_1 = 0$? How about the null hypothesis $H_0 : \beta_2 = 0$?

A:

```
lm.fit = lm(y~x1+x2)
summary(lm.fit)
```

```

Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.8311 -0.7273 -0.0537  0.6338  2.3359 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
```

```
x1             1.4396     0.7212   1.996   0.0487 *
x2             1.0097     1.1337   0.891   0.3754
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.056 on 97 degrees of freedom
Multiple R-squared:  0.2088,    Adjusted R-squared:  0.1925
F-statistic:  12.8 on 2 and 97 DF,  p-value: 1.164e-05
```

The regression coefficients are close to the true coefficients, although with high standard error. We can reject the null hypothesis for β_1 because its p-value is below 5%. We cannot reject the null hypothesis for β_2 because its p-value is much above the 5% typical cutoff, over 60%.

Q17:

Now fit least squares regression to predict y using only x_1. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

A:

```r
lm.fit = lm(y~x1)
summary(lm.fit)
```

```

Call:
lm(formula = y ~ x1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.89495 -0.66874 -0.07785  0.59221  2.45560

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.1124     0.2307   9.155 8.27e-15 ***
x1            1.9759     0.3963   4.986 2.66e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.055 on 98 degrees of freedom
Multiple R-squared:  0.2024,    Adjusted R-squared:  0.1942
F-statistic: 24.86 on 1 and 98 DF,  p-value: 2.661e-06
```

Yes, we can reject the null hypothesis for the regression coefficient given the p-value for its t-statistic is near zero.

Q18:

Now fit least squares regression to predict y using only x_2. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_2 = 0$?

A:

```
lm.fit = lm(y~x2)
summary(lm.fit)
```

```
Call:
lm(formula = y ~ x2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.62687 -0.75156 -0.03598  0.72383  2.44890 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.3899     0.1949   12.26  < 2e-16 ***
x2            2.8996     0.6330    4.58 1.37e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.072 on 98 degrees of freedom
Multiple R-squared:  0.1763,    Adjusted R-squared:  0.1679 
F-statistic: 20.98 on 1 and 98 DF,  p-value: 1.366e-05
```

Yes, we can reject the null hypothesis for the regression coefficient given the p-value for its t-statistic is near zero.

Q19:

Do the results from the previous questions contradict each other? Explain your answer.

A:

No, because x_1 and x_2 have collinearity, it is hard to distinguish their effects when regressed upon together. When they are regressed upon separately, the linear relationship between y and each predictor is indicated more clearly.

Q20:

Now suppose we obtain one additional observation, which was unfortunately mismeasured.

```r
x1 <- c(x1, 0.1)
x2 <- c(x2, 0.8)
y = c(y,6)
```

Re-fit the linear model using the new data. What effect does this new observation have on the each of the models? In each model, is this observation an outlier?

A:

```r
lm.fit1 = lm(y~x1+x2)
summary(lm.fit1)
```

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.73348 -0.69318 -0.05263  0.66385  2.30619 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.2267     0.2314   9.624 7.91e-16 ***
x1            0.5394     0.5922   0.911  0.36458    
x2            2.5146     0.8977   2.801  0.00614 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.075 on 98 degrees of freedom
Multiple R-squared:  0.2188,    Adjusted R-squared:  0.2029 
F-statistic: 13.72 on 2 and 98 DF,  p-value: 5.564e-06
```

```r
lm.fit2 = lm(y~x1)
summary(lm.fit2)
```

```
Call:
lm(formula = y ~ x1)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.8897 -0.6556 -0.0909  0.5682  3.5665 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.2569     0.2390   9.445 1.78e-15 ***
```

```
x1             1.7657     0.4124   4.282 4.29e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.111 on 99 degrees of freedom
Multiple R-squared:  0.1562,    Adjusted R-squared:  0.1477
F-statistic: 18.33 on 1 and 99 DF,  p-value: 4.295e-05
```

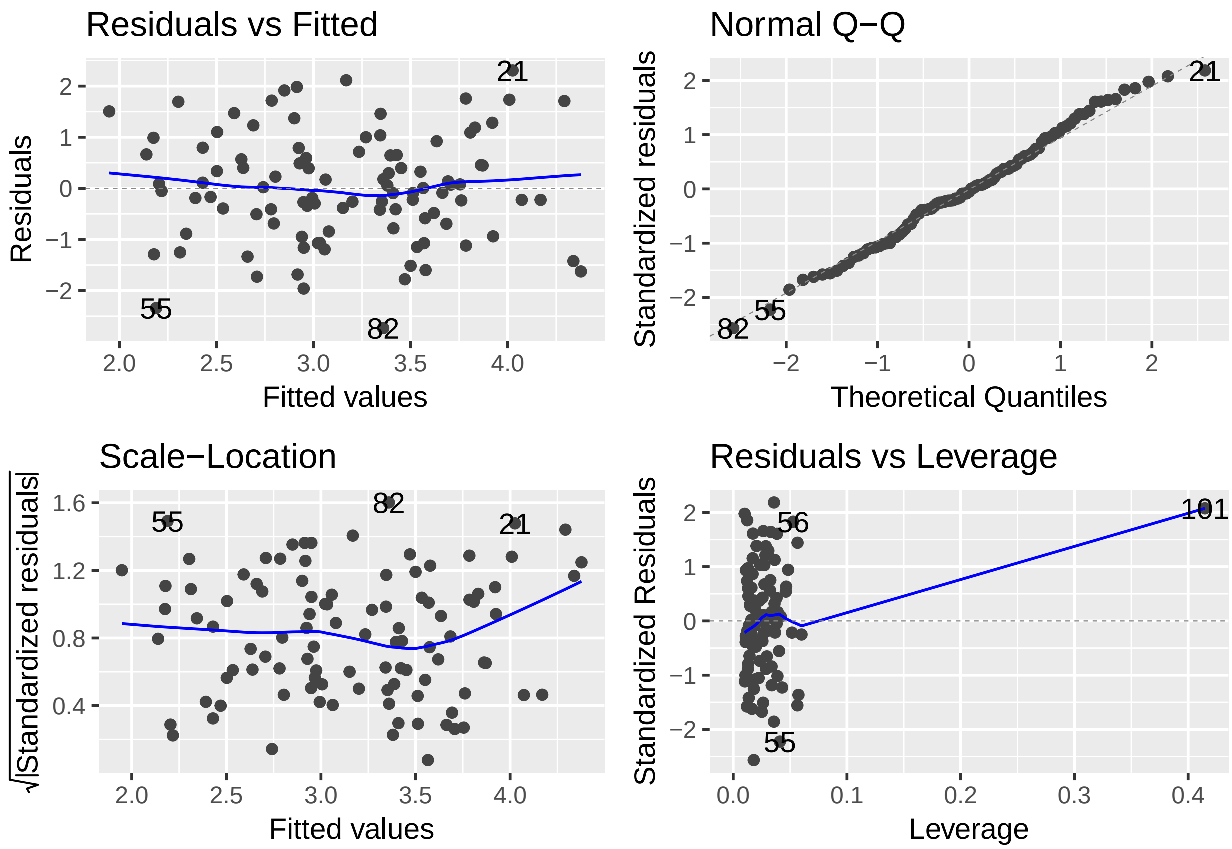
```r
lm.fit3 = lm(y~x2)
summary(lm.fit3)
```

```

Call:
lm(formula = y ~ x2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.64729 -0.71021 -0.06899  0.72699  2.38074

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.3451     0.1912  12.264  < 2e-16 ***
x2            3.1190     0.6040   5.164 1.25e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.074 on 99 degrees of freedom
Multiple R-squared:  0.2122,    Adjusted R-squared:  0.2042
F-statistic: 26.66 on 1 and 99 DF,  p-value: 1.253e-06
```

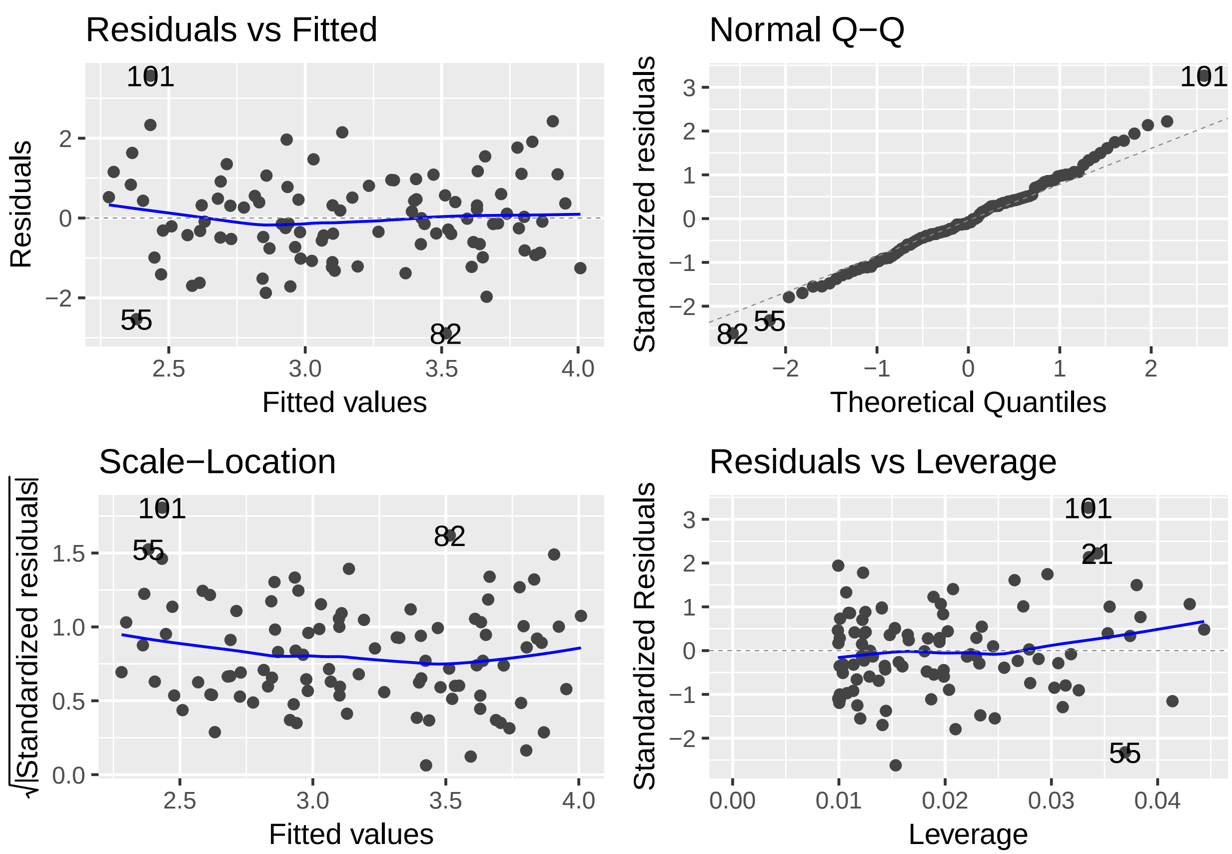
In the first model, it shifts x_1 to statistically insignificance and shifts x_2 to statistiscal significance from the change in p-values between the two linear regressions.

```r
library(ggfortify)
```

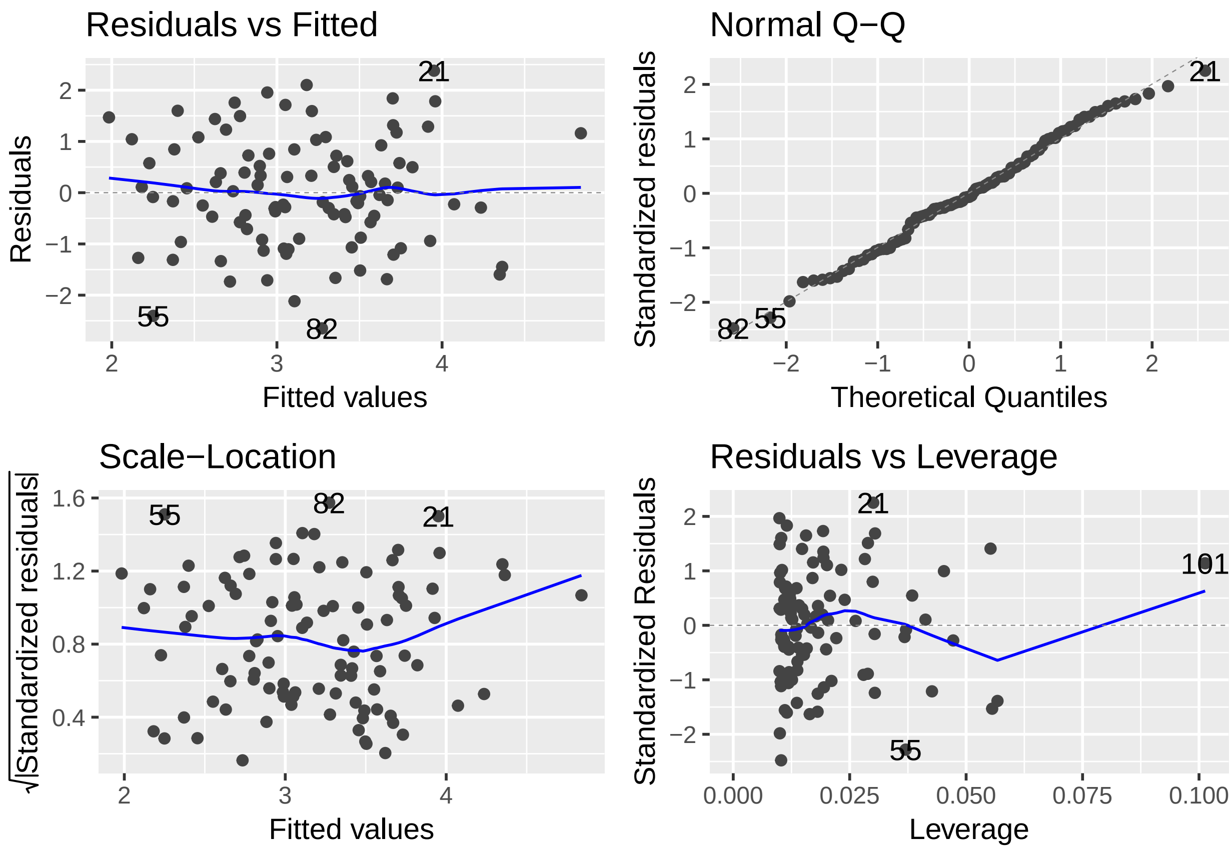
```r
autoplot(lm.fit1)
```




```r
autoplot(lm.fit2)
```




```
autoplot(lm.fit3)
```



The additional observation for x_2 seems to become a high leverage point.

county	mean_count
Orange	8.22e+03
San Bernardino	6.47e+03
San Diego	6.17e+03
Santa Clara	5.93e+03
Los Angeles	5.83e+03
Ventura	4.63e+03
Monterey	3.55e+03
Sacramento	3.51e+03
San Mateo	3.29e+03
Kern	3.11e+03
Stanislaus	3.01e+03
Riverside	2.85e+03
Contra Costa	2.73e+03
Santa Barbara	2.41e+03
San Benito	2.15e+03
Merced	2.11e+03
Imperial	1.94e+03
Placer	1.88e+03
Marin	1.64e+03
Inyo	1.51e+03
Kings	1.51e+03
El Dorado	1.37e+03
Santa Cruz	1.18e+03
Butte	1.15e+03
Madera	1.02e+03
Sonoma	984
Nevada	951
Yuba	940
Shasta	937
Tulare	899
Tehama	859
Calaveras	777

	(1)	(2)	(3)
(Intercept)	698.933 ***	638.729 ***	689.475 ***
	(9.467)	(7.449)	(14.409)
str	-2.280 ***	-0.649	-3.410 ***
	(0.480)	(0.354)	(0.760)
avginc		1.839 ***	-1.624
		(0.093)	(0.852)
str:avginc			0.190 ***
			(0.046)
nobs	420	420	420
adj.r.squared	0.049	0.509	0.527
AIC	3650.499	3373.711	3359.174
BIC	3662.620	3389.872	3379.376

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.