str~testscr

Brian

2019

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

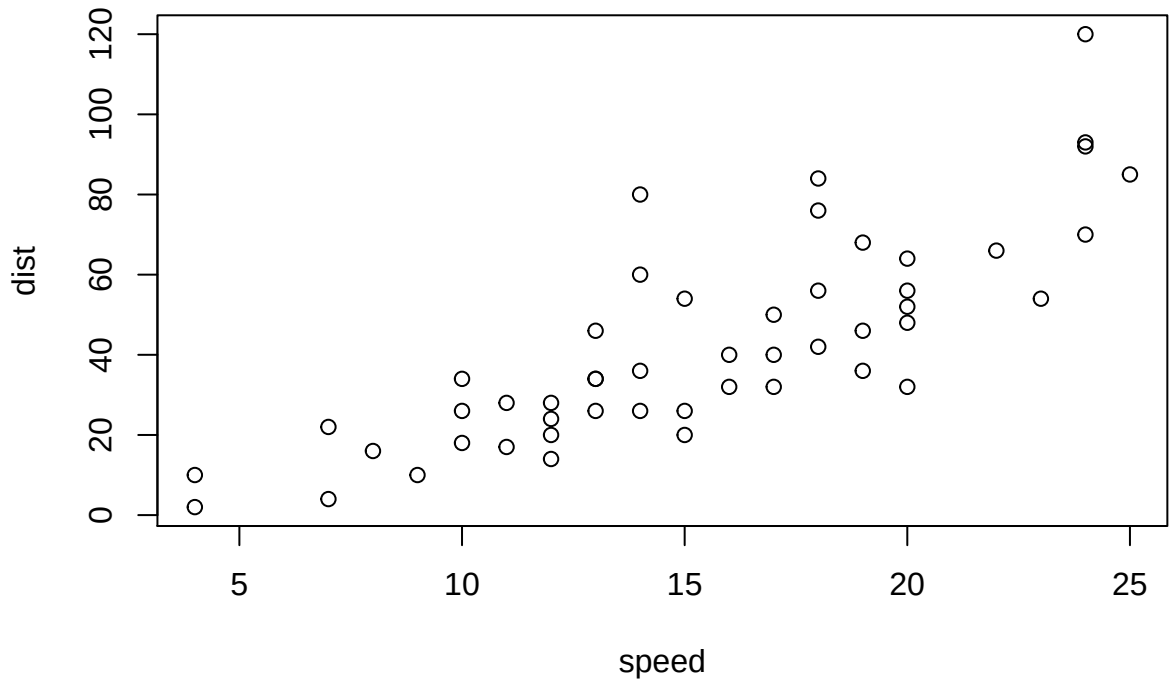
Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

```
tinytex::install_tinytex()
```

```
## Warning: Detected an existing tlmgr at /usr/local/bin/tlmgr. It seems
## TeX Live has been installed (check tinytex::tinytex_root()). You are
## recommended to uninstall it, although TinyTeX should work well alongside
## another LaTeX distribution if a LaTeX document is compiled through
## tinytex::latexmk().

## TinyTeX installed to /Users/Brian/Library/TinyTeX
```

```
plot(cars)
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

```
setwd("/Users/Brian/desktop/R/lab1")
csdata <- read.csv("dataSTR.csv")
head(csdata)
```

```
##   Observation.Number dist_cod  county                      district
## 1                  1    75119 Alameda              Sunol Glen Unified
## 2                  2    61499   Butte            Manzanita Elementary
## 3                  3    61549   Butte     Thermalito Union Elementary
## 4                  4    61457   Butte Golden Feather Union Elementary
## 5                  5    61523   Butte        Palermo Union Elementary
## 6                  6    62042  Fresno         Burrel Union Elementary
##   gr_span enrl_tot teachers calw_pct meal_pct computer testscr  comp_stu
## 1   KK-08      195    10.90   0.5102   2.0408       67  690.80 0.3435898
## 2   KK-08      240    11.15  15.4167  47.9167      101  661.20 0.4208333
## 3   KK-08     1550    82.90  55.0323  76.3226      169  643.60 0.1090323
## 4   KK-08      243    14.00  36.4754  77.0492       85  647.70 0.3497942
## 5   KK-08     1335    71.50  33.1086  78.4270      171  640.85 0.1280899
## 6   KK-08      137     6.40  12.3188  86.9565       25  605.55 0.1824818
##   expn_stu      str    avginc    el_pct read_scr math_scr
## 1 6384.911 17.88991 22.690001  0.000000    691.6    690.0
## 2 5099.381 21.52466  9.824000  4.583333    660.5    661.9
## 3 5501.955 18.69723  8.978000 30.000002    636.3    650.9
## 4 7101.831 17.35714  8.978000  0.000000    651.9    643.5
## 5 5235.988 18.67133  9.080333 13.857677    641.8    639.9
## 6 5580.147 21.40625 10.415000 12.408759    605.7    605.4
```

```
tail(csdata)
```

```
##     Observation.Number dist_cod      county                  district
## 415                415    69682 Santa Clara Saratoga Union Elementary
## 416                416    68957   San Mateo    Las Lomitas Elementary
## 417                417    69518 Santa Clara      Los Altos Elementary
## 418                418    72611     Ventura    Somis Union Elementary
## 419                419    72744        Yuba         Plumas Elementary
## 420                420    72751        Yuba      Wheatland Elementary
##     gr_span enrl_tot teachers calw_pct meal_pct computer testscr  comp_stu
## 415   KK-08     2341   124.09   0.1709   0.5980      286  700.30 0.1221700
## 416   KK-08      984    59.73   0.1016   3.5569      195  704.30 0.1981707
## 417   KK-08     3724   208.48   1.0741   1.5038      721  706.75 0.1936090
## 418   KK-08      441    20.15   3.5635  37.1938       45  645.00 0.1020408
## 419   KK-08      101     5.00  11.8812  59.4059       14  672.20 0.1386139
## 420   KK-08     1778    93.40   6.9235  47.5712      313  655.75 0.1760405
##     expn_stu      str   avginc    el_pct read_scr math_scr
## 415 5392.639 18.86534 40.40200  2.050406    698.9    701.7
## 416 7290.339 16.47413 28.71700  5.995935    700.9    707.7
## 417 5741.463 17.86263 41.73411  4.726101    704.0    709.5
```

```
## 418 4402.832 21.88586 23.73300 24.263039    648.3    641.7
## 419 4776.336 20.20000  9.95200  2.970297    667.9    676.5
## 420 5993.393 19.03640 12.50200  5.005624    660.5    651.0
```

```
summary(csdata)
```

```
##  Observation.Number    dist_cod              county
##  Min.   :  1.0      Min.   :61382   Sonoma     : 29
##  1st Qu.:105.8      1st Qu.:64308   Kern       : 27
##  Median :210.5      Median :67760   Los Angeles: 27
##  Mean   :210.5      Mean   :67473   Tulare     : 24
##  3rd Qu.:315.2      3rd Qu.:70419   San Diego  : 21
##  Max.   :420.0      Max.   :75440   Santa Clara: 20
##                                     (Other)    :272
##                        district    gr_span       enrl_tot
##  Lakeside Union Elementary:  3   KK-06: 61   Min.   :   81.0
##  Mountain View Elementary :  3   KK-08:359   1st Qu.:  379.0
##  Jefferson Elementary     :  2               Median :  950.5
##  Liberty Elementary       :  2               Mean   : 2628.8
##  Ocean View Elementary    :  2               3rd Qu.: 3008.0
##  Pacific Union Elementary :  2               Max.   :27176.0
##  (Other)                  :406
##     teachers          calw_pct         meal_pct         computer
##  Min.   :   4.85   Min.   : 0.000   Min.   :  0.00   Min.   :   0.0
##  1st Qu.:  19.66   1st Qu.: 4.395   1st Qu.: 23.28   1st Qu.:  46.0
##  Median :  48.56   Median :10.520   Median : 41.75   Median : 117.5
##  Mean   : 129.07   Mean   :13.246   Mean   : 44.71   Mean   : 303.4
##  3rd Qu.: 146.35   3rd Qu.:18.981   3rd Qu.: 66.86   3rd Qu.: 375.2
##  Max.   :1429.00   Max.   :78.994   Max.   :100.00   Max.   :3324.0
##
##     testscr         comp_stu          expn_stu         str
##  Min.   :605.5   Min.   :0.00000   Min.   :3926   Min.   :14.00
##  1st Qu.:640.0   1st Qu.:0.09377   1st Qu.:4906   1st Qu.:18.58
##  Median :654.5   Median :0.12546   Median :5215   Median :19.72
##  Mean   :654.2   Mean   :0.13593   Mean   :5312   Mean   :19.64
##  3rd Qu.:666.7   3rd Qu.:0.16447   3rd Qu.:5601   3rd Qu.:20.87
##  Max.   :706.8   Max.   :0.42083   Max.   :7712   Max.   :25.80
##
##      avginc           el_pct          read_scr        math_scr
##  Min.   : 5.335   Min.   : 0.000   Min.   :604.5   Min.   :605.4
##  1st Qu.:10.639   1st Qu.: 1.941   1st Qu.:640.4   1st Qu.:639.4
##  Median :13.728   Median : 8.778   Median :655.8   Median :652.5
##  Mean   :15.317   Mean   :15.768   Mean   :655.0   Mean   :653.3
##  3rd Qu.:17.629   3rd Qu.:22.970   3rd Qu.:668.7   3rd Qu.:665.9
##  Max.   :55.328   Max.   :85.540   Max.   :704.0   Max.   :709.5
##
```

create a subset of the original dataset

```
cssub <- subset(csdata, select=c("testscr","str"))
summary(cssub)
```

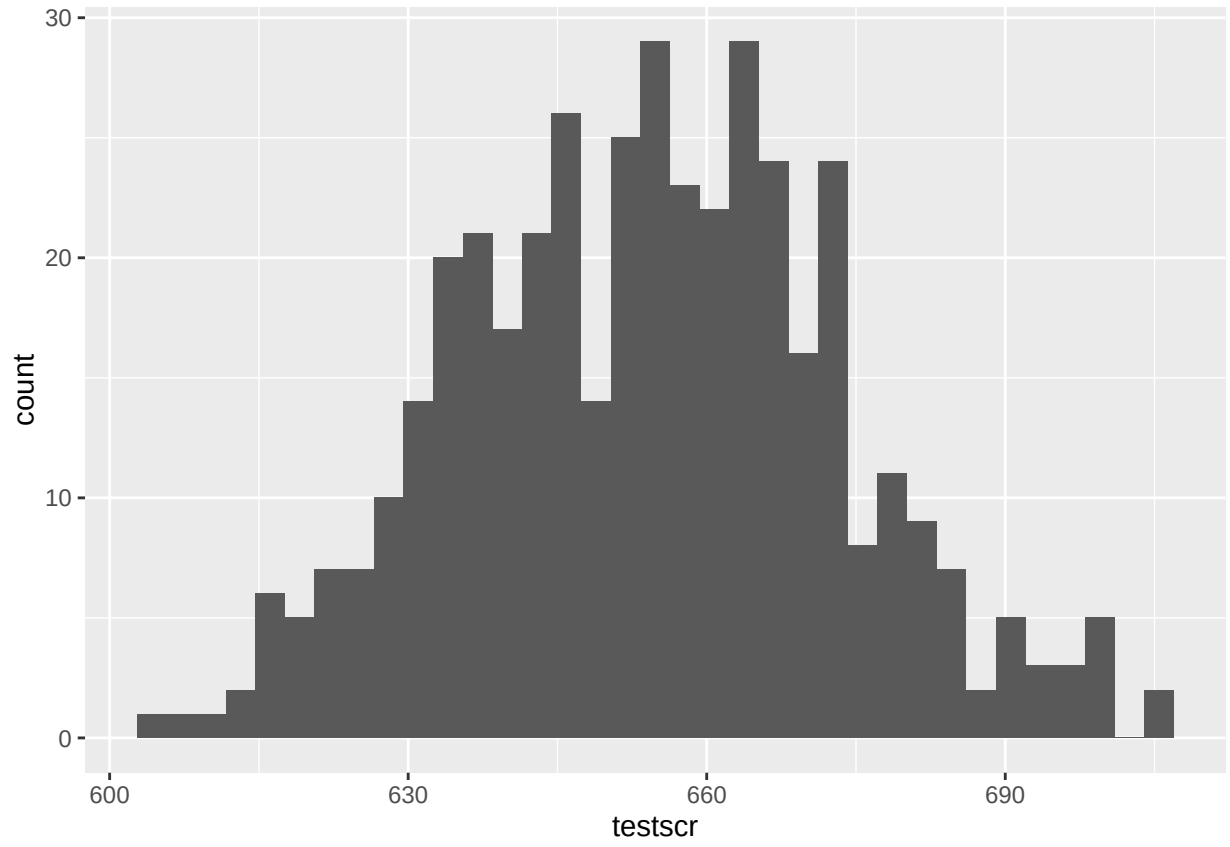
```
##     testscr           str
##  Min.   :605.5   Min.   :14.00
##  1st Qu.:640.0   1st Qu.:18.58
##  Median :654.5   Median :19.72
##  Mean   :654.2   Mean   :19.64
##  3rd Qu.:666.7   3rd Qu.:20.87
##  Max.   :706.8   Max.   :25.80
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

visualize the data in R, here we are trying to find the relationship b/w testscr(testscore) and str(student to teach ratio)

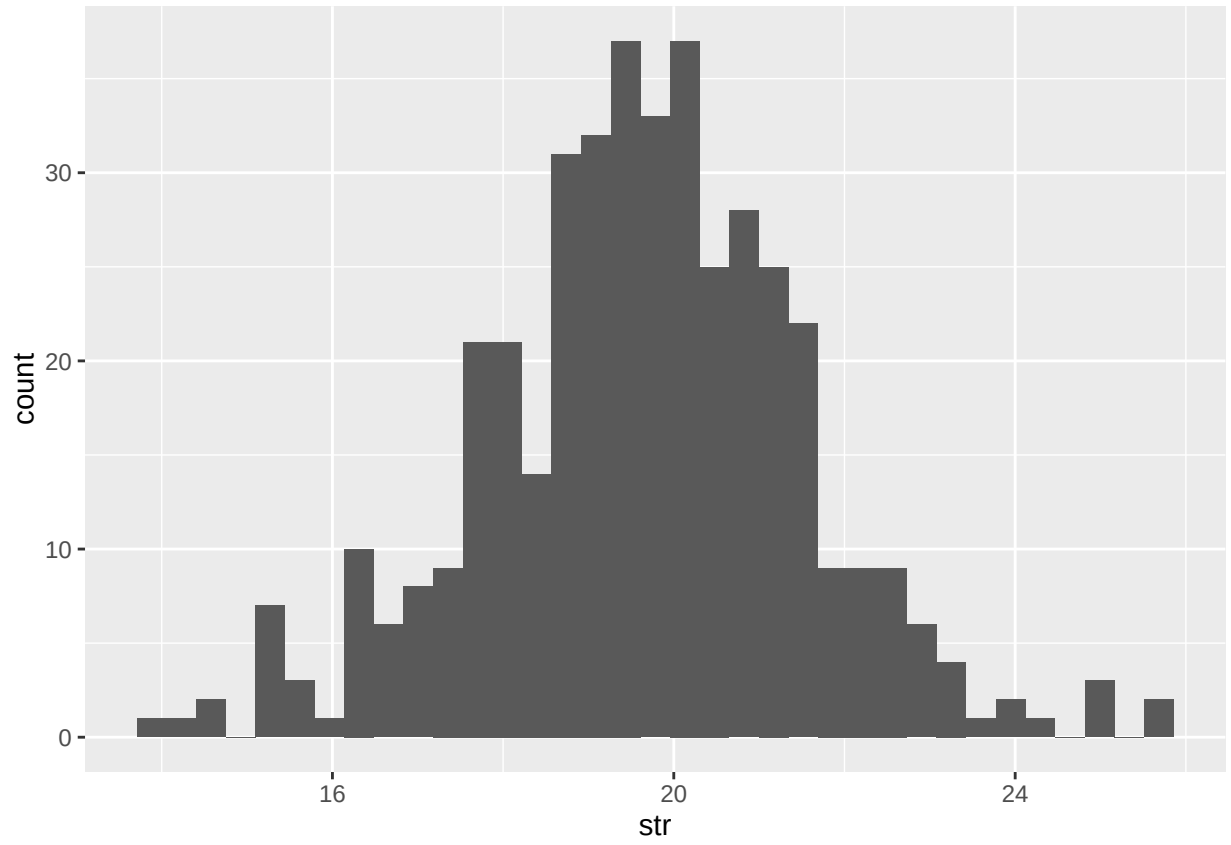
```
qplot(testscr, data=cssub, bins=35)
```




```
ggsave("testscr_plot.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

```
qplot(str, data=cssub, bins=35)
```

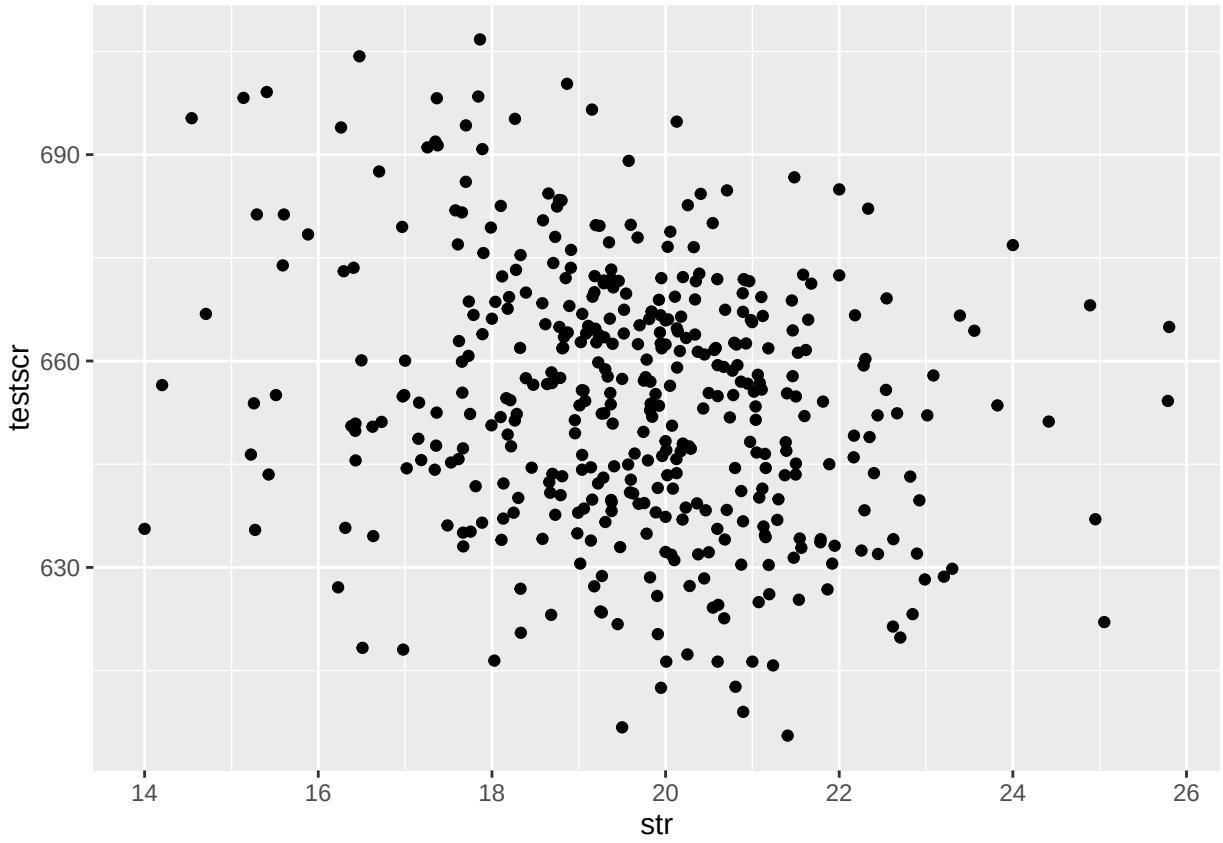



```
ggsave("str_plot.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

#plot tow variables together , form a scatterplot

```
qplot(str, testscr, data=cssub)
```

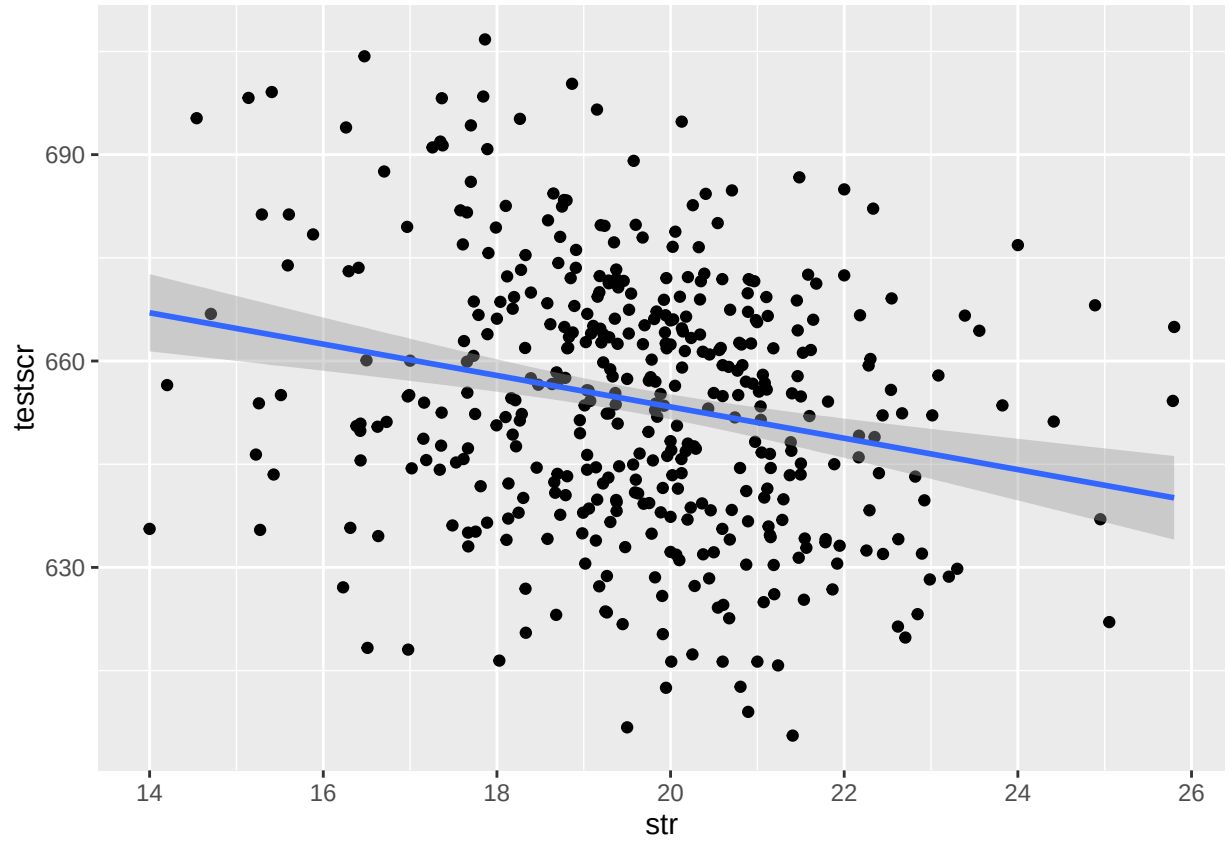



```
ggsave("str_testscr_scatterplot.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

fitting a straight line in the scatterplot

```
ggplot(cssub, aes(str, testscr))+geom_point()+geom_smooth(method="lm")
```



#Fit a line, linear~regression model

```
Lm <- lm(testscr~str, data=cssub)
summary(Lm)
```

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

Conclusion:

Using this model, we find that:

1. increasing the student-teacher-ratio...

ceteris paribus, reduces the performance in terms of

the average 5th grade test score.

2. Our best estimate is that, if str increases by one,

we expect performance to fall by -2.2.

Remember that test scores were between 600 and 700.

Remember:

3. We have no idea where these numbers come from

4. We did not involve any other variables

5. We have assumed linearity (and a lot more!)