Assignment 9–R

Insert your name here

September 5, 2018

Load necessary libraries here.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(ggformula)
```

```
## Loading required package: ggplot2

## Loading required package: ggstance

##
## Attaching package: 'ggstance'

## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh

##
## New to ggformula?  Try the tutorials:
##   learnr::run_tutorial("introduction", package = "ggformula")
##   learnr::run_tutorial("refining", package = "ggformula")
```

Problem 1: Colleges and Universities

In this problem, you will do further cleaning and analysis of the data from the 1995 US News and World Report on colleges and universities in the US.

a. In the Python portion of homework 9, you created a modified version of the data set usnews.csv. Read the modified data into R. Check the first few values of each vector to ensure that they were read correctly.

1

```
usnews <- read_csv("usframe.csv")
```

```
## Parsed with column specification:
## cols(
##    .default = col_integer(),
##    `College Name` = col_character(),
##    State = col_character(),
##    `Student/faculty ratio` = col_double(),
##    Pub_or_Priv = col_character(),
##    `IQR:Math` = col_double(),
##    `IQR:Verbal` = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
head(usnews)
```

```
## # A tibble: 6 x 38
##    `ID Number` `College Name` State `Public/private` `Avg Math SAT`
##          <int> <chr>          <chr>            <int>          <int>
## 1        1061 Alaska Pacifi~ AK                   2            490
## 2        1063 University of~ AK                   1            499
## 3        1065 University of~ AK                   1             NA
## 4       11462 University of~ AK                   1            459
## 5        1002 Alabama Agri.~ AL                   1             NA
## 6        1003 Faulkner Univ~ AL                   2             NA
## # ... with 33 more variables: `Avg Verbal SAT` <int>, `Avg combined
## #   SAT` <int>, `Avg ACT` <int>, `First quartile - Math SAT` <int>, `Third
## #   quartile - Math SAT` <int>, `First quartile - Verbal SAT` <int>,
## #   `Third quartile - Verbal SAT` <int>, `First quartile - ACT` <int>,
## #   `Third quartile - ACT` <int>, `Num applications received` <int>, `Num
## #   applicants accepted` <int>, `Num students enrolled` <int>, `Pct new
## #   students from top 10% of HS class` <int>, `Pct new students from top
## #   25% of HS class` <int>, `Num full-time undergraduates` <int>, `Num
## #   part-time undergraduates` <int>, `In-state tuition` <int>,
## #   `Out-of-state tuition` <int>, `Room and Board costs` <int>, `Room
## #   costs` <int>, `Board costs` <int>, `Additional fees` <int>, `Estimated
## #   book costs` <int>, `Estimated personal spending` <int>, `Pct of
## #   faculty with PhDs` <int>, `Pct of faculty with terminal degree` <int>,
## #   `Student/faculty ratio` <dbl>, `Pct alumni who donate` <int>,
## #   `Instructional expenditure per student` <int>, `Graduation
## #   rate` <int>, Pub_or_Priv <chr>, `IQR:Math` <dbl>, `IQR:Verbal` <dbl>
```

b. Examine the summary of each variable. Identify any unrealistic values and set them to missing.

- It may be helpful to use control flow or functions to help organize your work.

```
summary(usnews)
```

```
##     ID Number     College Name          State           Public/private
##  Min.   : 1002   Length:1302        Length:1302        Min.   :1.000
```

```
##    1st Qu.: 1874    Class :character    Class :character    1st Qu.:1.000
##    Median : 2650    Mode  :character    Mode  :character    Median :2.000
##    Mean   : 3126                                            Mean   :1.639
##    3rd Qu.: 3431                                            3rd Qu.:2.000
##    Max.   :30431                                            Max.   :2.000
##
##     Avg Math SAT    Avg Verbal SAT  Avg combined SAT     Avg ACT
##    Min.   :320.0   Min.   :280.0   Min.   : 600.0   Min.   :11.00
##    1st Qu.:460.0   1st Qu.:422.0   1st Qu.: 884.5   1st Qu.:20.25
##    Median :500.0   Median :457.0   Median : 957.0   Median :22.00
##    Mean   :506.8   Mean   :461.2   Mean   : 968.0   Mean   :22.12
##    3rd Qu.:544.0   3rd Qu.:492.0   3rd Qu.:1038.0   3rd Qu.:24.00
##    Max.   :750.0   Max.   :665.0   Max.   :1410.0   Max.   :31.00
##    NA's   :525     NA's   :525     NA's   :523      NA's   :588
##    First quartile - Math SAT Third quartile - Math SAT
##    Min.   :220.0             Min.   :330.0
##    1st Qu.:410.0             1st Qu.:530.0
##    Median :453.0             Median :580.0
##    Mean   :462.2             Mean   :583.1
##    3rd Qu.:510.0             3rd Qu.:630.0
##    Max.   :740.0             Max.   :785.0
##    NA's   :530               NA's   :530
##    First quartile - Verbal SAT Third quartile - Verbal SAT
##    Min.   :200.0               Min.   :330.0
##    1st Qu.:380.0               1st Qu.:480.0
##    Median :410.0               Median :530.0
##    Mean   :418.5               Mean   :530.5
##    3rd Qu.:450.0               3rd Qu.:570.0
##    Max.   :630.0               Max.   :720.0
##    NA's   :530                 NA's   :530
##    First quartile - ACT Third quartile - ACT Num applications received
##    Min.   :10.00        Min.   :15.00        Min.   :    35.0
##    1st Qu.:18.00        1st Qu.:23.00        1st Qu.:   695.8
##    Median :19.00        Median :25.00        Median :  1470.0
##    Mean   :19.82        Mean   :25.11        Mean   :  2752.1
##    3rd Qu.:22.00        3rd Qu.:27.00        3rd Qu.:  3314.2
##    Max.   :29.00        Max.   :35.00        Max.   :48094.0
##    NA's   :639          NA's   :639          NA's   :10
##    Num applicants accepted Num students enrolled
##    Min.   :    35.0        Min.   :   18.0
##    1st Qu.:   554.5        1st Qu.:  236.0
##    Median :  1095.0        Median :  447.0
##    Mean   :  1870.7        Mean   :  778.9
##    3rd Qu.:  2303.0        3rd Qu.:  984.0
##    Max.   :26330.0        Max.   : 7425.0
##    NA's   :11             NA's   :5
##    Pct new students from top 10% of HS class
##    Min.   : 1.00
##    1st Qu.:13.00
##    Median :21.00
##    Mean   :25.67
##    3rd Qu.:32.00
##    Max.   :98.00
##    NA's   :235
```

```
## Pct new students from top 25% of HS class Num full-time undergraduates
## Min.   :  6.00                            Min.   :   59
## 1st Qu.: 36.75                            1st Qu.:  966
## Median : 50.00                            Median : 1812
## Mean   : 52.35                            Mean   : 3693
## 3rd Qu.: 66.00                            3rd Qu.: 4540
## Max.   :100.00                            Max.   :31643
## NA's   :202                               NA's   :3
## Num part-time undergraduates In-state tuition Out-of-state tuition
## Min.   :    1.0               Min.   :  480    Min.   : 1044
## 1st Qu.:  131.2               1st Qu.: 2580    1st Qu.: 6111
## Median :  472.0               Median : 8050    Median : 8670
## Mean   : 1081.5               Mean   : 7897    Mean   : 9277
## 3rd Qu.: 1313.0               3rd Qu.:11600    3rd Qu.:11659
## Max.   :21836.0               Max.   :25750    Max.   :25750
## NA's   :32                    NA's   :30       NA's   :20
## Room and Board costs   Room costs    Board costs    Additional fees
## Min.   :1260           Min.   : 500   Min.   : 531   Min.   :   9.0
## 1st Qu.:3320           1st Qu.:1710   1st Qu.:1619   1st Qu.: 130.0
## Median :4030           Median :2200   Median :1980   Median : 264.5
## Mean   :4162           Mean   :2515   Mean   :2061   Mean   : 392.0
## 3rd Qu.:4849           3rd Qu.:3040   3rd Qu.:2402   3rd Qu.: 480.0
## Max.   :8700           Max.   :7400   Max.   :6250   Max.   :4374.0
## NA's   :76             NA's   :321    NA's   :498    NA's   :274
## Estimated book costs Estimated personal spending Pct of faculty with PhDs
## Min.   :  90         Min.   :  75                 Min.   :  8.00
## 1st Qu.: 480         1st Qu.: 900                 1st Qu.: 57.00
## Median : 502         Median :1250                 Median : 71.00
## Mean   : 550         Mean   :1389                 Mean   : 68.65
## 3rd Qu.: 600         3rd Qu.:1794                 3rd Qu.: 82.00
## Max.   :2340         Max.   :6900                 Max.   :105.00
## NA's   :48           NA's   :181                  NA's   :32
## Pct of faculty with terminal degree Student/faculty ratio
## Min.   : 20.00                       Min.   : 2.30
## 1st Qu.: 63.00                       1st Qu.:11.80
## Median : 77.00                       Median :14.30
## Mean   : 75.23                       Mean   :14.86
## 3rd Qu.: 90.00                       3rd Qu.:17.60
## Max.   :100.00                       Max.   :91.80
## NA's   :30                           NA's   :2
## Pct alumni who donate Instructional expenditure per student
## Min.   : 0.00         Min.   : 1834
## 1st Qu.:11.00         1st Qu.: 6116
## Median :19.00         Median : 7729
## Mean   :20.91         Mean   : 8988
## 3rd Qu.:29.00         3rd Qu.:10054
## Max.   :81.00         Max.   :62469
## NA's   :222           NA's   :39
## Graduation rate  Pub_or_Priv        IQR:Math        IQR:Verbal
## Min.   : 8.00    Length:1302      Min.   :-10.0   Min.   :  0
## 1st Qu.: 47.00   Class :character 1st Qu.:100.0   1st Qu.:100
## Median : 60.00   Mode  :character Median :120.0   Median :110
## Mean   : 60.41                    Mean   :120.9   Mean   :112
## 3rd Qu.: 74.00                    3rd Qu.:140.0   3rd Qu.:120
```

```
## Max.   :118.00                        Max.   :400.0  Max.    :310
## NA's   :98                            NA's   :530   NA's    :530
```

```r
which(usnews$`IQR:Math` < 0)
```

```
## [1] 674
```

```r
usnews$`IQR:Math`[usnews$`IQR:Math` == -10] <- NA
```

```r
usnews$`Student/faculty ratio`[usnews$`Student/faculty ratio` > 50] <- NA
```

```r
usnews$`Num applications received`[usnews$`Num applications received` > 40000] <- NA
usnews$`Num applicants accepted`[usnews$`Num applicants accepted` > 25000] <- NA
usnews$`Num students enrolled`[usnews$`Num students enrolled` == 4520] <- NA
```

```r
summary(usnews)
```

```
##     ID Number      College Name          State          Public/private
##  Min.   : 1002   Length:1302        Length:1302        Min.   :1.000
##  1st Qu.: 1874   Class :character   Class :character   1st Qu.:1.000
##  Median : 2650   Mode  :character   Mode  :character   Median :2.000
##  Mean   : 3126                                         Mean   :1.639
##  3rd Qu.: 3431                                         3rd Qu.:2.000
##  Max.   :30431                                         Max.   :2.000
##
##   Avg Math SAT    Avg Verbal SAT  Avg combined SAT    Avg ACT
##  Min.   :320.0   Min.   :280.0   Min.   : 600.0   Min.   :11.00
##  1st Qu.:460.0   1st Qu.:422.0   1st Qu.: 884.5   1st Qu.:20.25
##  Median :500.0   Median :457.0   Median : 957.0   Median :22.00
##  Mean   :506.8   Mean   :461.2   Mean   : 968.0   Mean   :22.12
##  3rd Qu.:544.0   3rd Qu.:492.0   3rd Qu.:1038.0   3rd Qu.:24.00
##  Max.   :750.0   Max.   :665.0   Max.   :1410.0   Max.   :31.00
##  NA's   :525     NA's   :525     NA's   :523      NA's   :588
##  First quartile - Math SAT Third quartile - Math SAT
##  Min.   :220.0             Min.   :330.0
##  1st Qu.:410.0             1st Qu.:530.0
##  Median :453.0             Median :580.0
##  Mean   :462.2             Mean   :583.1
##  3rd Qu.:510.0             3rd Qu.:630.0
##  Max.   :740.0             Max.   :785.0
##  NA's   :530               NA's   :530
##  First quartile - Verbal SAT Third quartile - Verbal SAT
##  Min.   :200.0               Min.   :330.0
##  1st Qu.:380.0               1st Qu.:480.0
##  Median :410.0               Median :530.0
##  Mean   :418.5               Mean   :530.5
##  3rd Qu.:450.0               3rd Qu.:570.0
##  Max.   :630.0               Max.   :720.0
##  NA's   :530                 NA's   :530
##  First quartile - ACT Third quartile - ACT Num applications received
##  Min.   :10.00        Min.   :15.00        Min.   :   35.0
##  1st Qu.:18.00        1st Qu.:23.00        1st Qu.:  695.5
```

```
##  Median :19.00        Median :25.00        Median : 1470.0
##  Mean   :19.82        Mean   :25.11        Mean   : 2717.0
##  3rd Qu.:22.00        3rd Qu.:27.00        3rd Qu.: 3310.0
##  Max.   :29.00        Max.   :35.00        Max.   :22165.0
##  NA's   :639          NA's   :639          NA's   :11
##  Num applicants accepted Num students enrolled
##  Min.   :   35.0         Min.   :   18.0
##  1st Qu.:  554.2         1st Qu.:  236.0
##  Median : 1094.0         Median :  446.5
##  Mean   : 1851.7         Mean   :  776.0
##  3rd Qu.: 2299.2         3rd Qu.:  981.0
##  Max.   :18744.0         Max.   : 7425.0
##  NA's   :12              NA's   :6
##  Pct new students from top 10% of HS class
##  Min.   : 1.00
##  1st Qu.:13.00
##  Median :21.00
##  Mean   :25.67
##  3rd Qu.:32.00
##  Max.   :98.00
##  NA's   :235
##  Pct new students from top 25% of HS class Num full-time undergraduates
##  Min.   :  6.00                            Min.   :   59
##  1st Qu.: 36.75                            1st Qu.:  966
##  Median : 50.00                            Median : 1812
##  Mean   : 52.35                            Mean   : 3693
##  3rd Qu.: 66.00                            3rd Qu.: 4540
##  Max.   :100.00                            Max.   :31643
##  NA's   :202                               NA's   :3
##  Num part-time undergraduates In-state tuition Out-of-state tuition
##  Min.   :    1.0              Min.   :  480    Min.   : 1044
##  1st Qu.:  131.2              1st Qu.: 2580    1st Qu.: 6111
##  Median :  472.0             Median : 8050    Median : 8670
##  Mean   : 1081.5             Mean   : 7897    Mean   : 9277
##  3rd Qu.: 1313.0             3rd Qu.:11600    3rd Qu.:11659
##  Max.   :21836.0             Max.   :25750    Max.   :25750
##  NA's   :32                  NA's   :30       NA's   :20
##  Room and Board costs   Room costs      Board costs     Additional fees
##  Min.   :1260           Min.   : 500    Min.   : 531    Min.   :   9.0
##  1st Qu.:3320           1st Qu.:1710    1st Qu.:1619    1st Qu.: 130.0
##  Median :4030           Median :2200    Median :1980    Median : 264.5
##  Mean   :4162           Mean   :2515    Mean   :2061    Mean   : 392.0
##  3rd Qu.:4849           3rd Qu.:3040    3rd Qu.:2402    3rd Qu.: 480.0
##  Max.   :8700           Max.   :7400    Max.   :6250    Max.   :4374.0
##  NA's   :76             NA's   :321     NA's   :498     NA's   :274
##  Estimated book costs Estimated personal spending Pct of faculty with PhDs
##  Min.   :  90         Min.   :  75                Min.   :  8.00
##  1st Qu.: 480         1st Qu.: 900                1st Qu.: 57.00
##  Median : 502         Median :1250                Median : 71.00
##  Mean   : 550         Mean   :1389                Mean   : 68.65
##  3rd Qu.: 600         3rd Qu.:1794                3rd Qu.: 82.00
##  Max.   :2340         Max.   :6900                Max.   :105.00
##  NA's   :48           NA's   :181                 NA's   :32
##  Pct of faculty with terminal degree Student/faculty ratio
```

```
##  Min.   : 20.00                     Min.   : 2.30
##  1st Qu.: 63.00                      1st Qu.:11.80
##  Median : 77.00                      Median :14.30
##  Mean   : 75.23                      Mean   :14.76
##  3rd Qu.: 90.00                      3rd Qu.:17.50
##  Max.   :100.00                      Max.   :42.60
##  NA's   :30                          NA's   :4
##  Pct alumni who donate Instructional expenditure per student
##  Min.   : 0.00         Min.   : 1834
##  1st Qu.:11.00         1st Qu.: 6116
##  Median :19.00         Median : 7729
##  Mean   :20.91         Mean   : 8988
##  3rd Qu.:29.00         3rd Qu.:10054
##  Max.   :81.00         Max.   :62469
##  NA's   :222           NA's   :39
##  Graduation rate  Pub_or_Priv          IQR:Math        IQR:Verbal
##  Min.   : 8.00   Length:1302      Min.   : 0.0   Min.   : 0
##  1st Qu.: 47.00  Class :character 1st Qu.:100.0  1st Qu.:100
##  Median : 60.00  Mode  :character Median :120.0  Median :110
##  Mean   : 60.41                   Mean   :121.1  Mean   :112
##  3rd Qu.: 74.00                   3rd Qu.:140.0  3rd Qu.:120
##  Max.   :118.00                   Max.   :400.0  Max.   :310
##  NA's   :98                       NA's   :531    NA's   :530
```

Write a sentence describing what you did, naming the colleges or universities affected. (For example, "Listed ages less than zero (ABC University, XYZ College) were converted to missing data.")

Pembroke State University had an IQR: Math value of -10 and was converted to missing data.

Northwood University has a student/faculty ration of 91.80 and was converted to missing data

St Leo College has a student/faculty ratio of 72.4 and was converted to missing data

Rutgers at New Brunswich received 48094 applications and accepted 26330 but only has 4520 students enrolled. Converted all three to missing data

  c. Find the mean percentage of alumni who donate, for private and public schools.

```
meanPub <- mean(usnews$`Pct alumni who donate`[usnews$Pub_or_Priv == "Public"], na.rm=T)
meanPub <- as.vector(meanPub)
```

```
meanPriv <- mean(usnews$`Pct alumni who donate`[usnews$Pub_or_Priv == "Private"], na.rm=T)
meanPriv <- as.vector(meanPriv)
```

  d. Test whether there is evidence that a higher percentage of alumni from private schools donate to their schools, compared to alumni from public schools.

  • Hint: In part c, we took the mean of this variable in each group. What does this tell you about what type of hypothesis to use? two sample t-test

```
t1 <- usnews$`Pct alumni who donate`[usnews$Pub_or_Priv == "Public"]
t1 <- t1[!is.na(t1)]
t2 <- usnews$`Pct alumni who donate`[usnews$Pub_or_Priv == "Private"]
t2 <- t2[!is.na(t2)]
```

```
testResult <- t.test(t2, t1, alternative="greater")
testResult
```

```
## 
##  Welch Two Sample t-test
## 
## data:  t2 and t1
## t = 17.317, df = 1018.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  10.07495      Inf
## sample estimates:
## mean of x mean of y
##  24.58287  13.44944
```

State your conclusion in context.

H0: There's not higher percentage of donations from alumni between public and private schools

Ha: There is a higher percentage of donations from alumni between private and public schools

Since the p-value $< 2.2e\text{-}16$ we can reject the null hypothesis and say that there is a higher percentage of alumni from private schools donating to their schools than public schools.

e. Use `write_csv()` or `write.csv()` to save your updated data set.

- If you are using `write.csv()`, consult the R documentation to set the arguments for the function. Your output file should not have row names or row numbers.
- After you save the file, open it in Excel, Notepad++, or a program of your choice to verify that the version you saved contains the updates you made in part b.

```
write.csv(usnews, "usnewsR.csv", row.names = FALSE)
```

Problem 2: Current Population Survey

The data set cps.csv contains data from the 1985 Current Population Survey.

- Dataset: "Wages from the Current Population Survey," http://www.macalester.edu/~kaplan/ism/, from Daniel Kaplan, *Statistical Modeling: A Fresh Approach.* Original source: Berndt, ER. *The Practice of Econometrics 1991.* Addison-Wesley.

- Metadata: cps_metadata.pdf, from p. 418 of *Statistical Modeling: A Fresh Approach* by Daniel Kaplan.

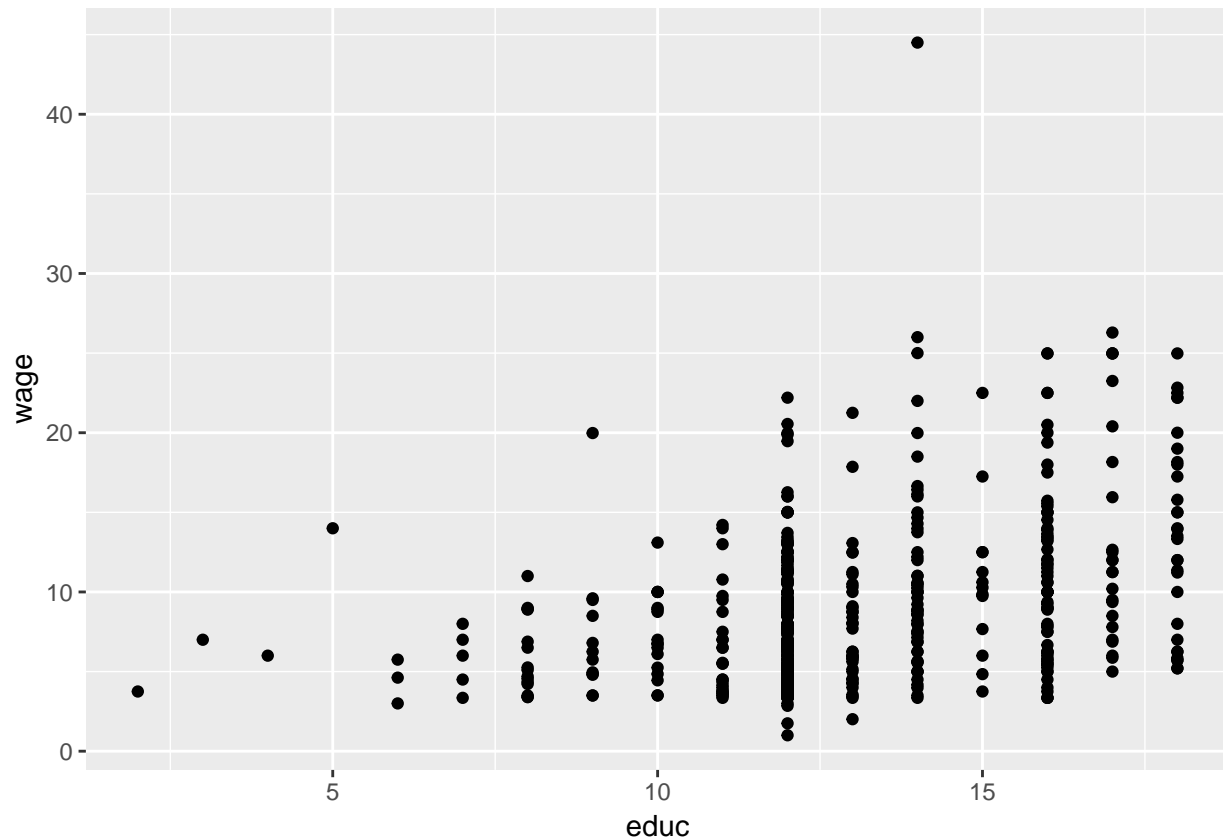a. Read the data into R and plot wages versus education.

```
cps <- read_csv("cps.csv")
```

```
## Parsed with column specification:
## cols(
##   wage = col_double(),
##   educ = col_integer(),
##   race = col_character(),
```

```
##     sex = col_character(),
##     hispanic = col_character(),
##     south = col_character(),
##     married = col_character(),
##     exper = col_integer(),
##     union = col_character(),
##     age = col_integer(),
##     sector = col_character()
## )
```

```
gf_point(wage ~ educ, data = cps)
```


Comment on the appropriateness of linear regression. The wage vs education plot seems like it may have a linear relationship and we're looking at how a variable only compares to one other variable
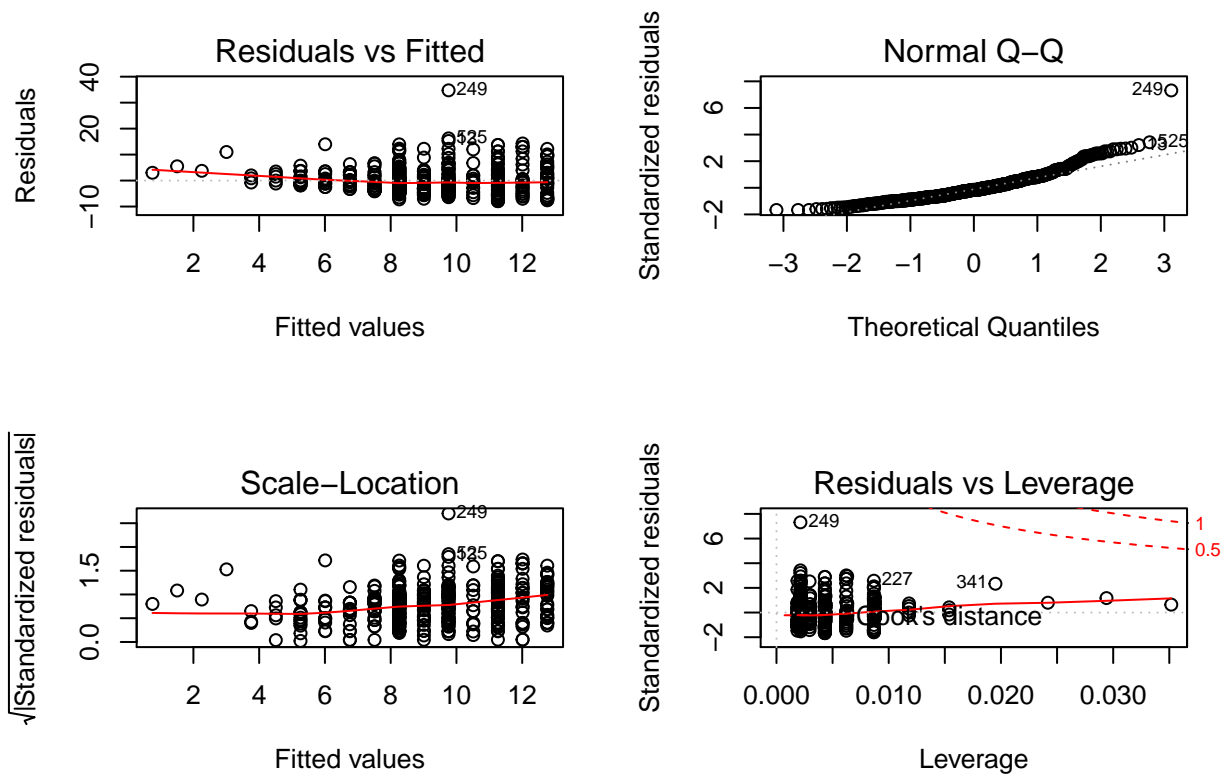
  b. Perform the linear regression and examine the diagnostic plots.

```
cpsreg <- lm(wage ~ educ, data=cps)
summary(cpsreg)
```

```
##
## Call:
## lm(formula = wage ~ educ, data = cps)
##
## Residuals:
```

```
##     Min     1Q Median     3Q    Max
## -7.911 -3.260 -0.760  2.240 34.740
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.74598    1.04545  -0.714    0.476
## educ         0.75046    0.07873   9.532   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.754 on 532 degrees of freedom
## Multiple R-squared:  0.1459, Adjusted R-squared:  0.1443
## F-statistic: 90.85 on 1 and 532 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(cpsreg)
```


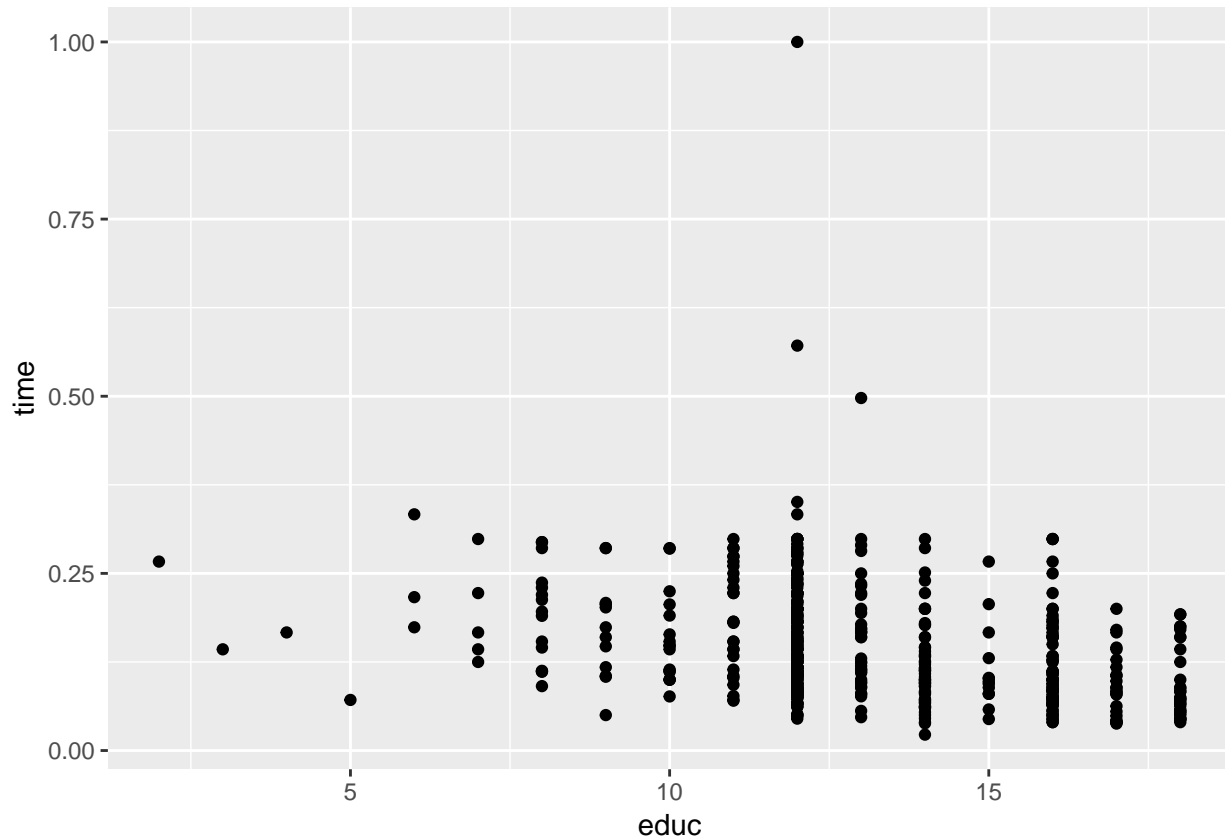**Explain** why transforming the wages variable is a good idea in this case.

There's a lot of changes in variance in the Residuals vs Fitted plot so a transformation is a good idea.

  c. The variable `wage` has units of dollars/hour. Create a new variable, `time`, equal to 1/`wage`. (So `time` has units of hours/dollar, or the length of time a person must work to earn $1.00.)

10

```
cps$time <- 1/cps$wage
```

    d. Plot time versus education.

```
gf_point(time ~ educ, data = cps)
```


Comment on the appropriateness of linear regression.

    e. Perform the linear regression.

```
timereg <- lm(time ~ educ, data = cps)
summary(timereg)
```

```
##
## Call:
## lm(formula = time ~ educ, data = cps)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15393 -0.05180 -0.02021  0.04361  0.84371
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.274700   0.017251  15.924  < 2e-16 ***
```
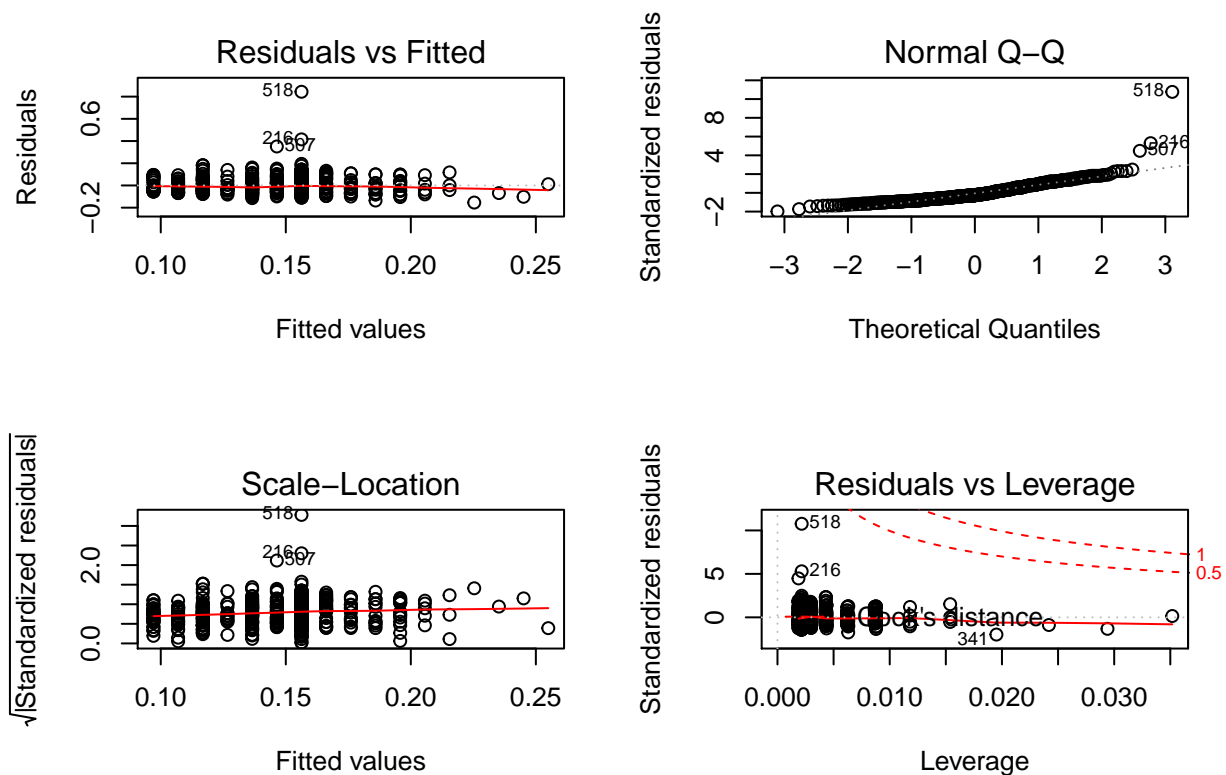
```
## educ           -0.009867   0.001299  -7.595 1.39e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07845 on 532 degrees of freedom
## Multiple R-squared:  0.09782,    Adjusted R-squared:  0.09613
## F-statistic: 57.68 on 1 and 532 DF,  p-value: 1.393e-13
```

Based on these results, are you happy with your decision to pursue a master's degree? **Explain.**

The regression makes it seem like the more education you have the more money you will end up making. I guess from that standpoint I'm glad I'm pursuing my master's even though its not really about the money for me.

f. Examine the diagnostic plots.

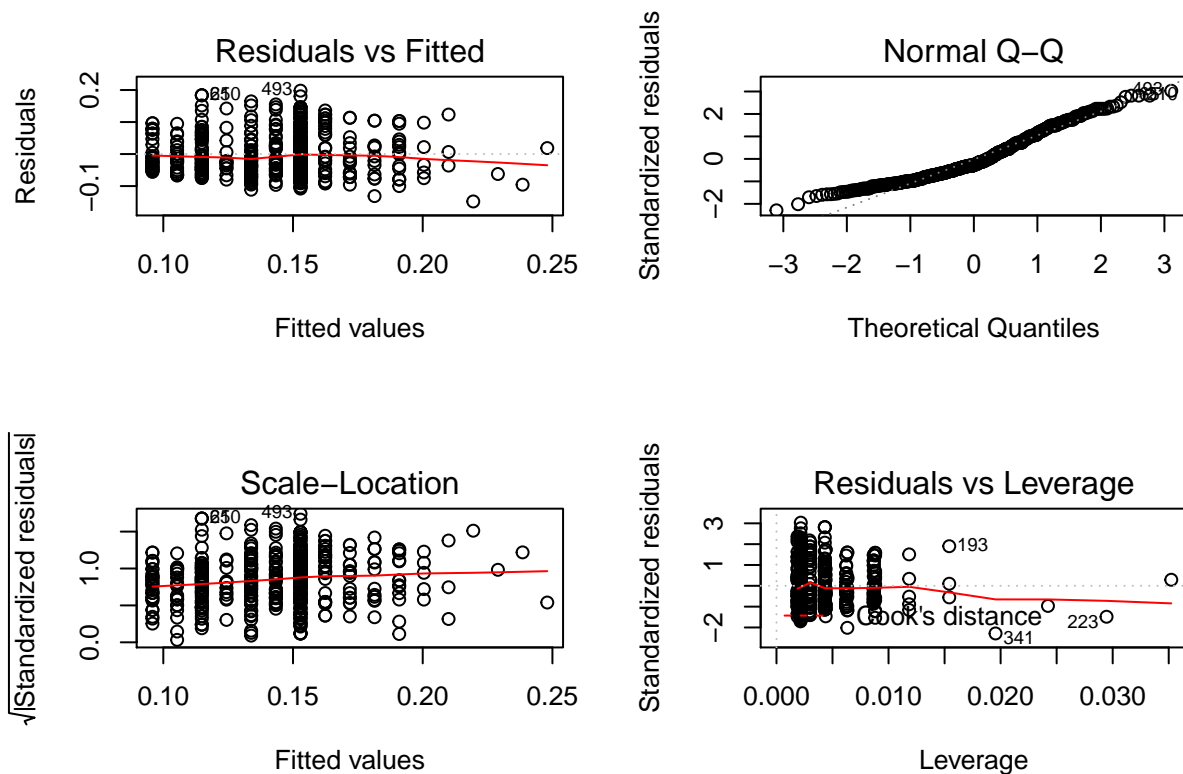```
par(mfrow = c(2,2))
plot(timereg)
```



**Which individuals** appear to be outliers on the residual vs. predicted plot? 216, 507, 518 Re-do the regression without those individuals.

```
cps2 <- cps
cps2[216,] <- NA
cps2[507,] <- NA
cps2[518,] <- NA
```

```
timereg2 <- lm(time ~ educ, data = cps2)
summary(timereg2)
```

```
##
## Call:
## lm(formula = time ~ educ, data = cps2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.14805 -0.04856 -0.01807  0.04557  0.19798
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.267039   0.014387  18.561   <2e-16 ***
## educ        -0.009512   0.001083  -8.783   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06538 on 529 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.1273, Adjusted R-squared:  0.1256
## F-statistic: 77.13 on 1 and 529 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(timereg2)
```

Did excluding the outliers change your conclusion?

No it did not. It still looks like the more education you have, the more money you'll end up making.

- Submit your .Rmd file, your knitted .docx or .pd f (.pdf preferred), and your updated data set from problem 1e to GitHub.