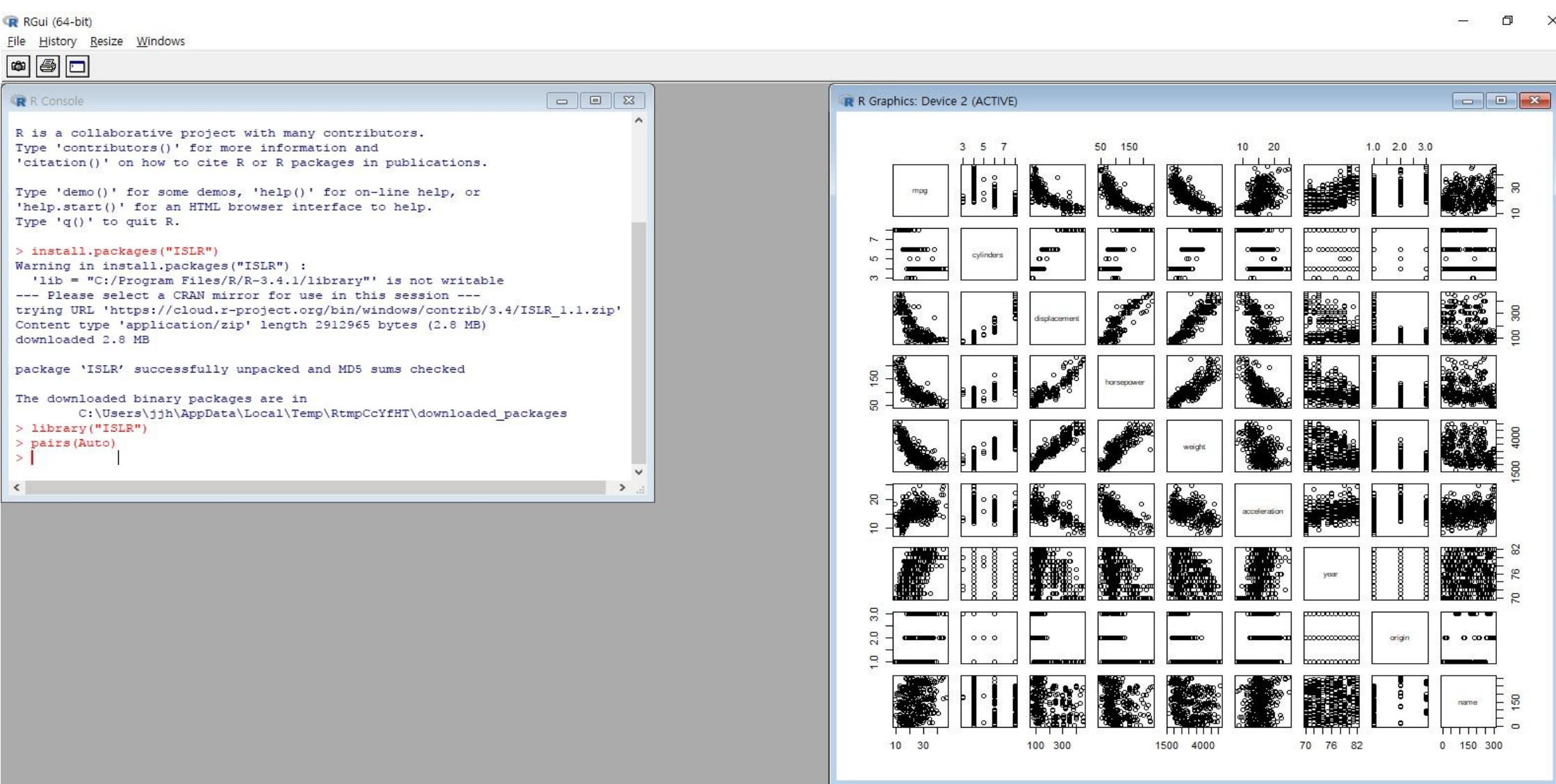


Assignment 2

20131750 전지훈/Jeehun Jeon

1.



2.

```
R Console
C:\Users\jjh\AppData\Local\Temp\RtmpCcYfHT\downloaded_packages
> library("ISLR")
> pairs(Auto)
> cor(Auto[, names(Auto) != "name"])
```

	mpg	cylinders	displacement	horsepower	weight
mpg	1.0000000	-0.7776175	-0.8051269	-0.7784268	-0.8322442
cylinders	-0.7776175	1.0000000	0.9508233	0.8429834	0.8975273
displacement	-0.8051269	0.9508233	1.0000000	0.8972570	0.9329944
horsepower	-0.7784268	0.8429834	0.8972570	1.0000000	0.8645377
weight	-0.8322442	0.8975273	0.9329944	0.8645377	1.0000000
acceleration	0.4233285	-0.5046834	-0.5438005	-0.6891955	-0.4168392
year	0.5805410	-0.3456474	-0.3698552	-0.4163615	-0.3091199
origin	0.5652088	-0.5689316	-0.6145351	-0.4551715	-0.5850054

	acceleration	year	origin
mpg	0.4233285	0.5805410	0.5652088
cylinders	-0.5046834	-0.3456474	-0.5689316
displacement	-0.5438005	-0.3698552	-0.6145351
horsepower	-0.6891955	-0.4163615	-0.4551715
weight	-0.4168392	-0.3091199	-0.5850054
acceleration	1.0000000	0.2903161	0.2127458
year	0.2903161	1.0000000	0.1815277
origin	0.2127458	0.1815277	1.0000000

```
> |
```

3-1.

- Simple linear regression model (mpg ~ cylinders).
- Cylinders has statistically significant relationship with mpg.
- 60.37% of variability of mpg.

```
R Console

origin      0.2127458  0.1815277  1.0000000
> model = lm(mpg ~ cylinders, data = Auto)
> summary(model)

Call:
lm(formula = mpg ~ cylinders, data = Auto)

Residuals:
      Min       1Q   Median       3Q      Max
-14.2413  -3.1832  -0.6332   2.5491  17.9168

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.9155     0.8349   51.40  <2e-16 ***
cylinders    -3.5581     0.1457  -24.43  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.914 on 390 degrees of freedom
Multiple R-squared:  0.6047,    Adjusted R-squared:  0.6037
F-statistic: 596.6 on 1 and 390 DF,  p-value: < 2.2e-16

> |
```


3-2.

- Multiple linear regression model ($\text{mpg} \sim . - \text{name}$).
- Cylinders, Horsepower, Weight have negative relationship.
- Displacement, Acceleration, Year and Origin have positive relationship.
- Displacement, Weight, Year and Origin have significant.
- 81.82% of variability of mpg.

```
R Console

acceleration    1.0000000  0.2903161  0.2127458
year            0.2903161  1.0000000  0.1815277
origin          0.2127458  0.1815277  1.0000000
> model= lm(mpg~. -name, data= Auto)
> summary(model)

Call:
lm(formula = mpg ~ . - name, data = Auto)

Residuals:
    Min       1Q   Median       3Q      Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604

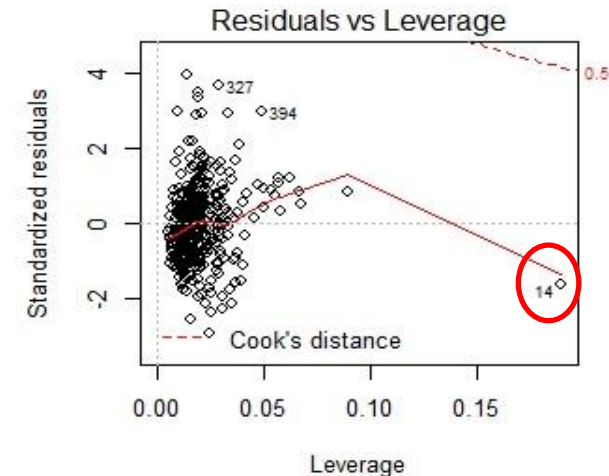
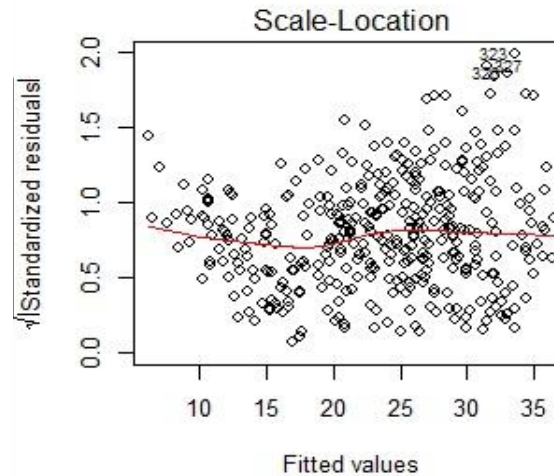
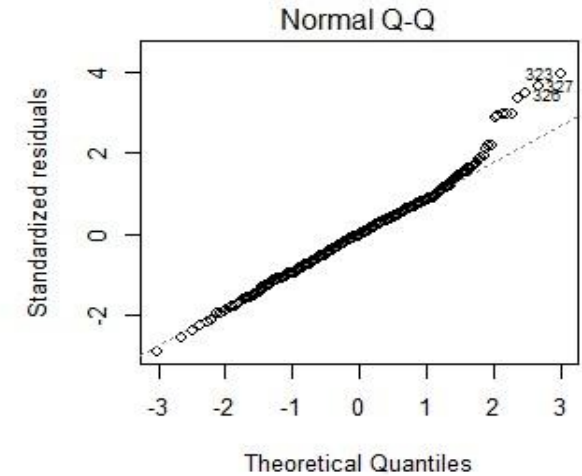
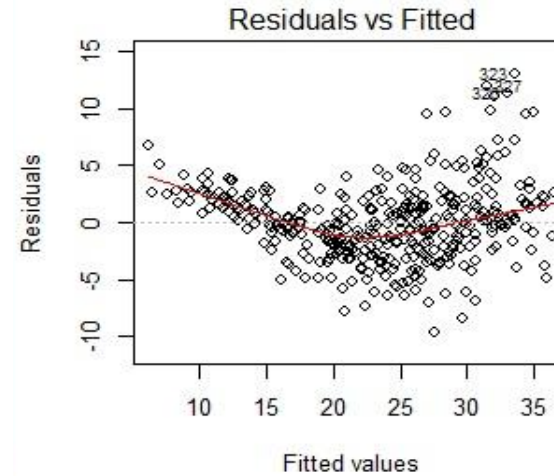
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.218435   4.644294  -3.707  0.00024 ***
cylinders    -0.493376   0.323282  -1.526  0.12780
displacement  0.019896   0.007515   2.647  0.00844 **
horsepower   -0.016951   0.013787  -1.230  0.21963
weight       -0.006474   0.000652  -9.929 < 2e-16 ***
acceleration  0.080576   0.098845   0.815  0.41548
year          0.750773   0.050973  14.729 < 2e-16 ***
origin        1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

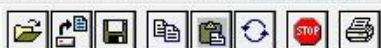
Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16

> |
```

4.

- Residual vs Fitted shows that mpg has non-linear relationship with other variables.
- Normal Q-Q shows there are some outliers.
- In residual vs leverage, it indicates unusually large outliers which is larger than 2.
- And unusually high leverage point is existed.





Untitled - R Editor

```
install.packages("ISLR")
library("ISLR")
pairs(Auto)

cor(Auto[, names(Auto) != "name"])

model = lm(mpg ~ cylinders, data = Auto)
summary(model)

model = lm(mpg ~ . - name, data = Auto)
summary(model)

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

R Console

```
      Min      1Q  Median      3Q      Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-17.218435	4.644294	-3.707	0.00024 ***
cylinders	-0.493376	0.323282	-1.526	0.12780
displacement	0.019896	0.007515	2.647	0.00844 **
horsepower	-0.016951	0.013787	-1.230	0.21963
weight	-0.006474	0.000652	-9.929	< 2e-16 ***
acceleration	0.080576	0.098845	0.815	0.41548
year	0.750773	0.050973	14.729	< 2e-16 ***
origin	1.426141	0.278136	5.127	4.67e-07 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

```
> par(mfrow = c(2,2))
> plot(model)
> |
```

R Graphics: Device 2 (ACTIVE)

