Chapter 8

8.1

For the following data, use R to verify that the least squares regresion line is $\hat{Y} = 1.8X - 8.5$

X: 5, 8, 9, 7, 14 Y: 3, 1, 6, 7, 19

```r
dat <- data.table(
    X = c(5, 8, 9, 7, 14),
    Y = c(3, 1, 6, 7, 19))

lm(Y ~ X, data = dat)
```

```
Call:
lm(formula = Y ~ X, data = dat)

Coefficients:
(Intercept)            X
    -8.478        1.823
```

Also verify that the Theil-Sen estimator, the slope is estimated to be 1.746 and the intercept is estimated to be -7.968.

```r
tsreg(dat$X, dat$Y)$coef
```

```
Intercept
-5.730159   1.746032
```

8.2

Using the R function *lsfit*, compute the residuals using the data in E1,

Verify that if you square and sum the residusls, you get 46.585.

```r
res <- lsfit(dat$X, dat$Y)$resid

sum(res^2)
```

```
[1] 46.58407
```

8.3

Verify that for the data in E1, if you use $\hat{Y} = 2X - 9$, the sum of the squared residuals is greater than 46.584.

```r
Yhat <- 2*dat$X - 9
res <- sum( (dat$Y - Yhat)^2 )
```

```
stopifnot(res > 46.583)
res
```

[1] 53

Why would you expect a value greater than 46.584?

The x coefficent increased.

8.4

Suppose that based on $n = 25$ values, $s_x^2 = 12$ and $\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 144$.

What is the slope of the least squares regression?

$$A = 144, C = (n-1)s_x^2 = 288, b_1 = A/C = 144/288 = .5$$

8.5

The following table reports breast cancer rates plus levels of solar radiation (in calories per day) for various cities in the United States. The data are stored in the file cancer_rate_dat.txt.

```
dat <- data.table::fread(paste0(data.dir, "cancer_rate_dat.txt"), fill = T, sep = "&")

dat
```

	City	Rate	calories
1:	New York	32.75	300
2:	Chicago	30.75	275
3:	Pittsburgh	28.00	280
4:	Seattle	27.25	270
5:	Boston	30.75	305
6:	Cleveland	31.00	335
7:	Columbus	29.00	340
8:	Indianapolis	26.50	342
9:	New Orleans	27.00	348
10:	Nashville	23.50	354
11:	Washington, DC	31.20	357
12:	Salt Lake City	22.70	394
13:	Omaha	27.00	380
14:	San Diego	25.80	383
15:	Atlanta	27.00	397
16:	Los Angeles	27.80	450
17:	Miami	23.50	453
18:	Fort Worth	21.50	446
19:	Tampa	21.00	456
20:	Albuquerque	22.50	513

```
21:       Las Vegas 21.50        510
22:       Honolulu 20.60        520
23:       El Paso 22.80         535
24:       Phoenix 21.00         520
            City   Rate calories
```
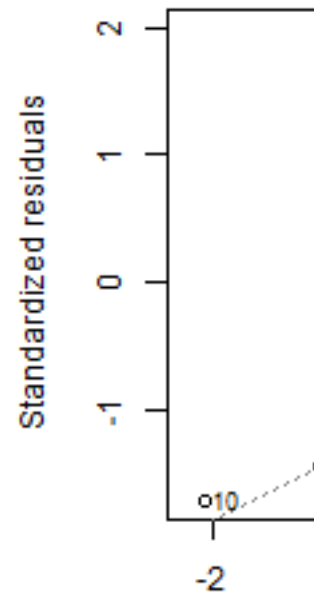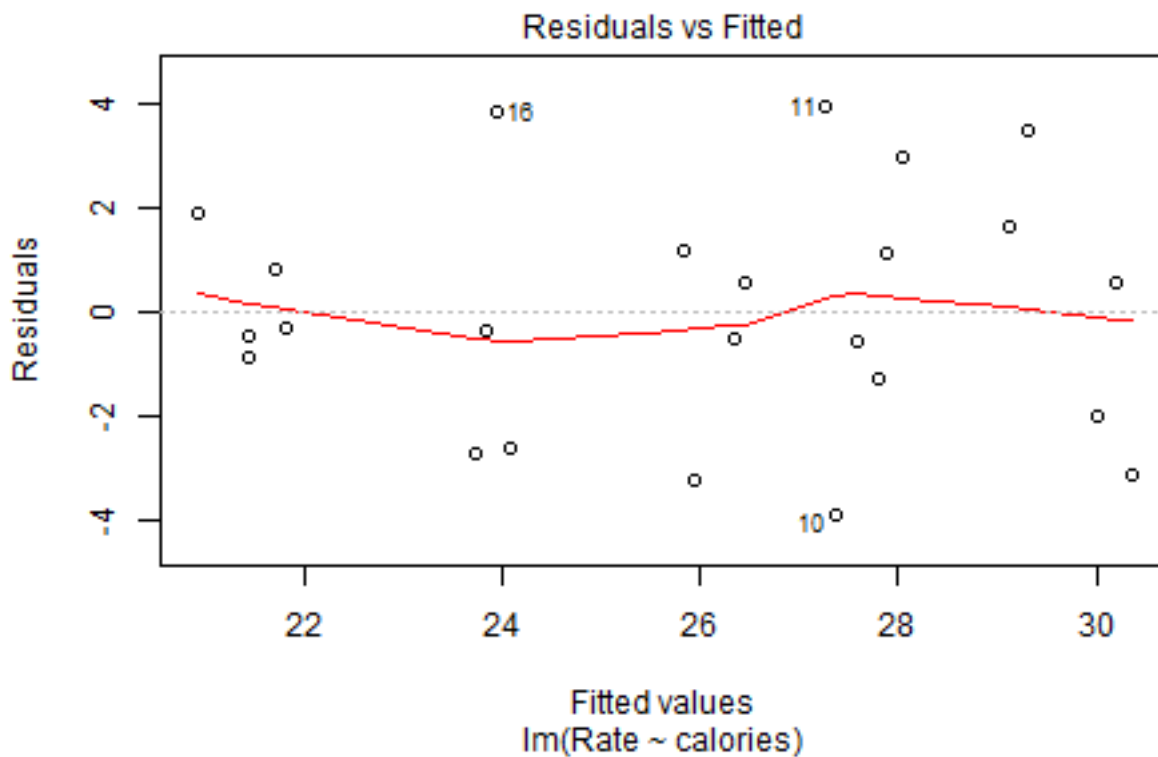
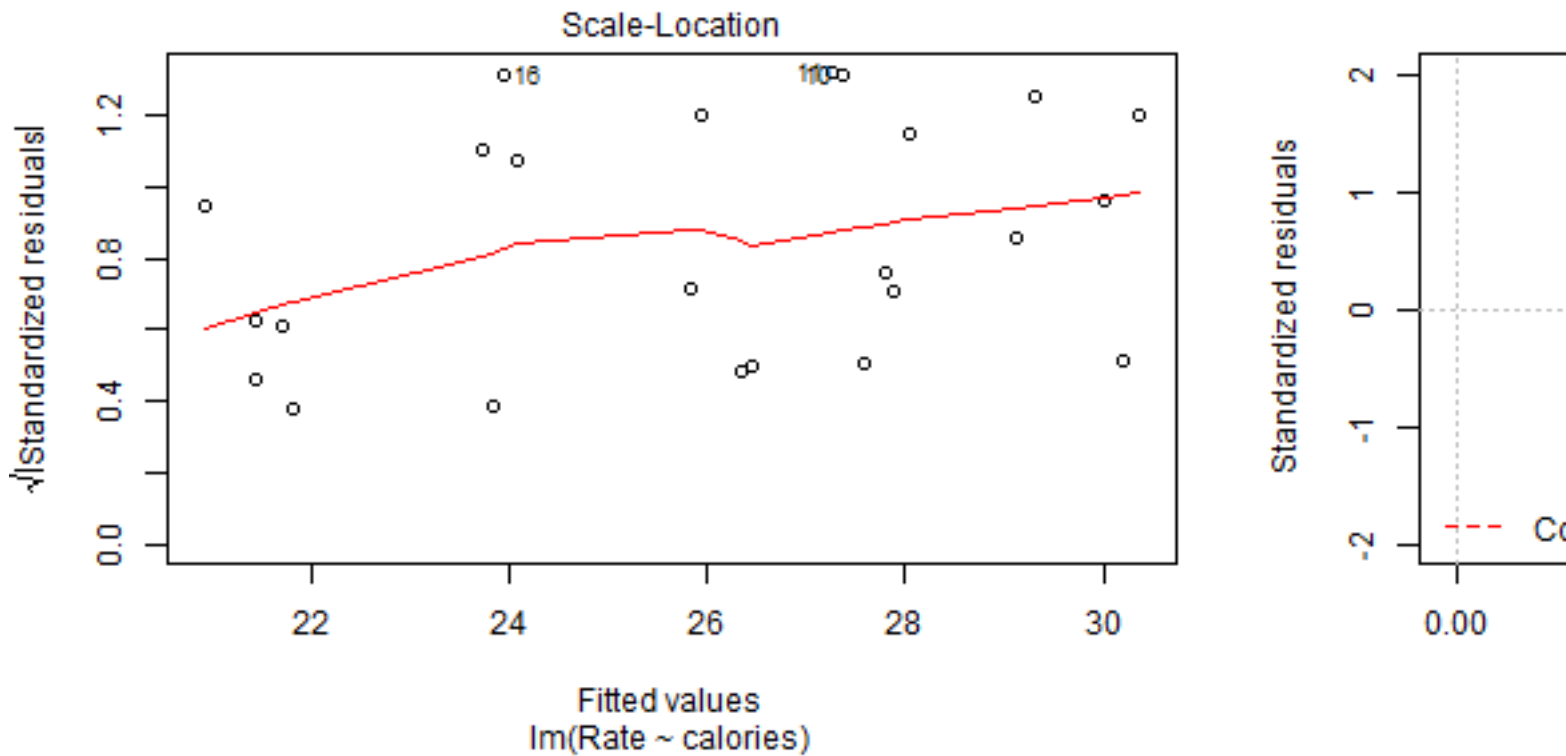Fit a OLS regression to predict cancer rates and comment on what this suggests.

```
fit <- lm(Rate ~ calories, data = dat)

plot(fit)
```

Residuals vs Fitted

Scale-Location

Fitted values
lm(Rate ~ calories)

8.6

For the following data, use R to compute the least squares regresion line for predicting GPA given SAT.

SAT: 500, 530, 590, 660, 610, 700, 570, 640 GPA: 2.3, 3.1, 2.6, 3.0, 2.4, 3.3, 2.6, 3.5

```
dat <- data.table(
    SAT = c(500, 530, 590, 660, 610, 700, 570, 640),
    GPA = c(2.3, 3.1, 2.6, 3.0, 2.4, 3.3, 2.6, 3.5)
)

fit <- lm(GPA ~ SAT, data = dat)
coef(fit)
```

```
(Intercept)         SAT
0.484615385 0.003942308
```

8.7

Compute the residuals for the data used in the previous problem and verify that sum to zero.

```r
round(sum((dat$GPA - fit$fitted.values)), 5)
```

```
[1] 0
```

8.8

For the following data, use R to compute the least squares regression line for predicting Y from X.

X: 40, 41, 42, 43, 44, 45, 46 Y: 1.62, 1.63, 1.90, 2.64, 2.05, 2.13, 1.94

```r
dat <- data.table(
    X = c(40, 41, 42, 43, 44, 45, 46),
    Y = c(1.62, 1.63, 1.90, 2.64, 2.05, 2.13, 1.94)
)

summary(fit <- lm(Y ~ X, data = dat))
```

```
Call:
lm(formula = Y ~ X, data = dat)

Residuals:
        1         2         3         4         5         6         7
-0.141071 -0.206429 -0.011786  0.652857 -0.012500 -0.007857 -0.273214

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.25321    2.73157  -0.459    0.666
X            0.07536    0.06346   1.188    0.288

Residual standard error: 0.3358 on 5 degrees of freedom
Multiple R-squared:   0.22, Adjusted R-squared:   0.064
F-statistic:  1.41 on 1 and 5 DF,  p-value: 0.2883
```
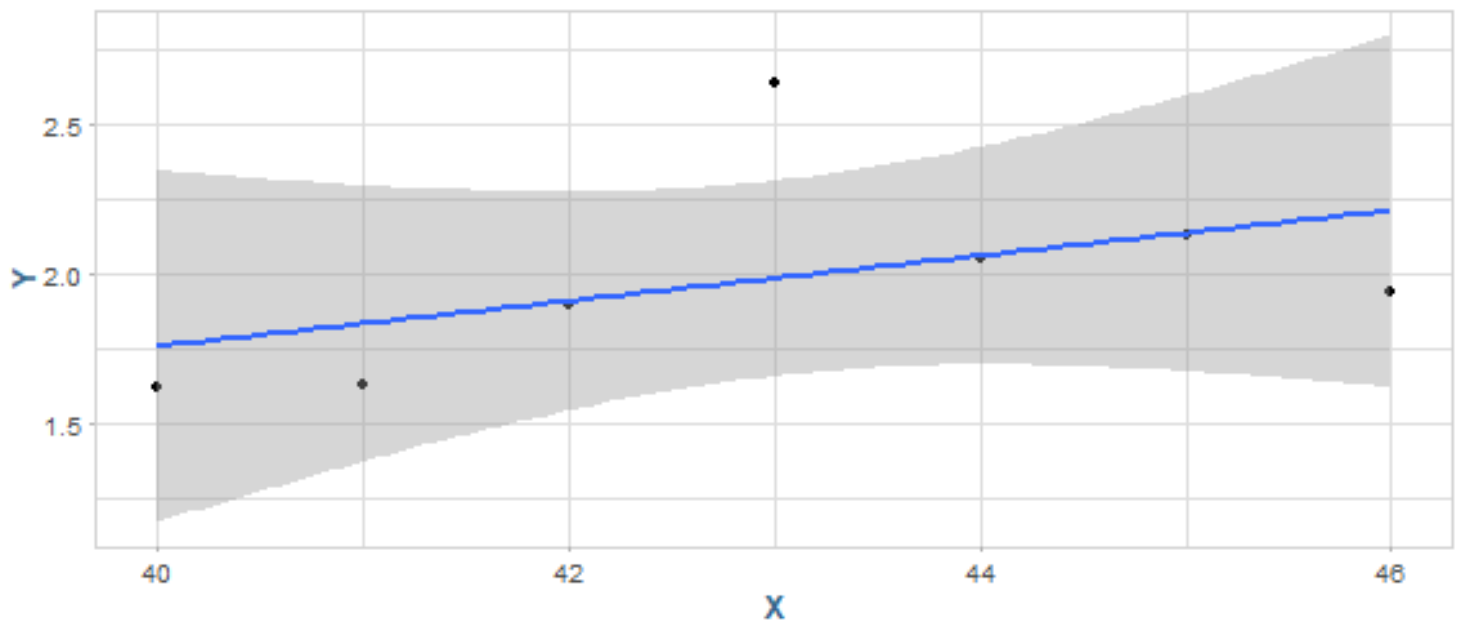
```r
ggplot(dat, aes(X, Y)) +
    geom_point() +
    geom_smooth(method = "lm")
```

8.9

8.10

8.21

8.22

8.23

8.24

8.25

8.26

8.27

8.28

8.29

8.30

8.31

8.32

8.33

8.34

8.35

8.36