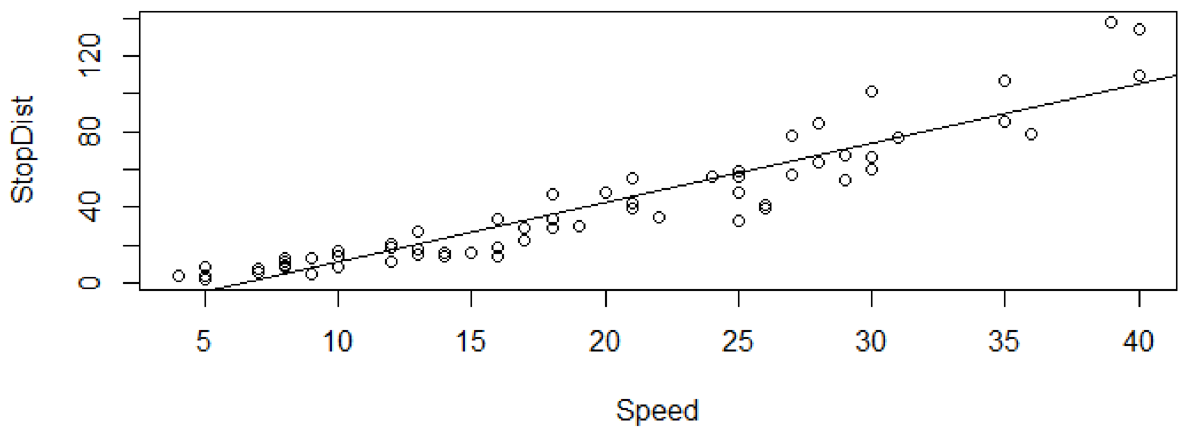
1. (a) Draw a scatterplot and add a regression line on the scatterplot.

```
> attach(CarStopDistance)
> plot(StopDist ~ Speed)
> abline( lm(StopDist ~ Speed) )
```



(b) Fit a linear model to the data. That is, obtain the regression analysis results.

```
> summary( lm(StopDist ~ Speed) )

Call:
lm(formula = StopDist ~ Speed)

Residuals:
    Min      1Q  Median      3Q     Max 
-25.141  -7.300  -2.141   6.044  35.946 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -20.2734     3.2384   -6.26 4.25e-08 ***
Speed         3.1366     0.1517   20.68  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.8 on 61 degrees of freedom
Multiple R-squared:  0.8752,	Adjusted R-squared:  0.8731 
F-statistic: 427.7 on 1 and 61 DF,  p-value: < 2.2e-16
```

(c) Using the results from (b), write the equation of the regression line $\hat{b} = b_0 + b_1 x$

$$\hat{b} = b_0 + b_1 x = -20.2734 + 3.1366x$$

(d) Interpret the slope of the regression line in the context of the problem. **Note**: You need to use the value b_1 in your answer to receive full credit.

Every 3.1366 mile per hour speed increase, the predict of stopping distance increase 1 foot.

(e) Interpret the R-squared value. i.e. Does car speed explain a large portion (how much variability in the average stopping distance? **Note**: You need to use the R-squared value answer to receive full credit.

87.52% of variation/variability in stopping-distance is explained by the model

(f) Predict the average stopping distance of a car when its speed is 32 miles per hour.

$$\hat{b} = b_0 + b_1 x = -20.2734 + 3.1366 \times 32 = 80.0978$$

Predicted average stopping distance is 80.0978 feet

(g) Find the correlation coefficient between the average stopping distance and the car's speed.

- Paste the R command and their result inside a text box.

```
> cor(Speed, StopDist)
[1] 0.9355037
```

(h) Convert the stopping distance in feet to inches. (1 feet = 12 inches)

- Paste the following command in RStudio.

```
> StopDist.inch = StopDist * 12
```

```
> head( cbind(StopDist, StopDist.inch) )
     StopDist StopDist.inch
[1,]        4            48
[2,]        2            24
[3,]        8            96
[4,]        8            96
[5,]        4            48
[6,]        6            72
```

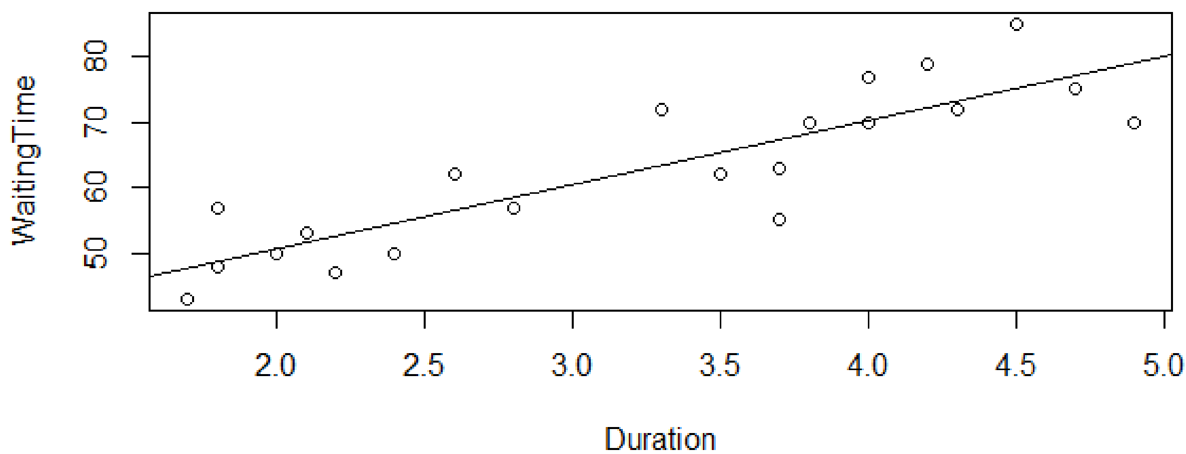
(i) Find the correlation coefficient between the average stopping distance (in inches) and the car's speed. **Note**: You'll notice that any linear transformation of x and/or y won't affect the correlation coefficient.

```
> cor(Speed, StopDist.inch)
[1] 0.9355037
```

2. Duration and Waiting Time (Data: Eruption)

(a) Draw a scatterplot and add a regression line on the scatterplot

```
> attach(CarStopDistance)
> plot(StopDist ~ Speed)
> abline( lm(StopDist ~ Speed) )
```



(b) Fit a linear model to the data. That is, obtain the regression analysis results

```
> summary( lm(WaitingTime ~ Duration) )

Call:
lm(formula = WaitingTime ~ Duration)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.2364  -4.2364  -0.6352   5.5327   9.9316 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   31.013      4.417   7.022 1.10e-06 ***
Duration       9.790      1.300   7.531 4.06e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 6.129 on 19 degrees of freedom
Multiple R-squared:  0.7491,   Adjusted R-squared:  0.7359 
F-statistic: 56.72 on 1 and 19 DF,  p-value: 4.059e-07
```

(c) Using the results from (b), write the equation of the regression line $\widehat{b} = b_0 + b_1 x$

$$\widehat{b} = b_0 + b_1 x = 31.013 + 9.790x$$

(d) Interpret the slope of the regression line in the context of the problem. **Note**: You need to use the value b_1 in your answer to receive full credit.

Every 9.790 minutes length of time of last eruption, the predict of time until the next eruption increase 1 minute.

(e) Is the y-intercept of the regression line meaningful in this case?

$$b_0 = 31.013$$

Predict time in minutes until the next eruption when we don't have any eruption. Not reasonable.

(f) Interpret the R-squared value. i.e. Does the length of time of the last eruption explain a large portion (how much) of the variability in the time until the next eruption?

$$R^2 = 0.7491$$

74.91% of variation in time in minutes until the next eruption is explained by the model

(g) You just missed an eruption, but you know it lasted 3.3 minutes. In how many minutes do you expect to see the next eruption? **Note**: Make your prediction using the regression model.

$$0.7491 \times 3.3 = 2.47203\ mins$$

(h) One of the eruptions in the data set lasted 3.3 minutes and the next eruption occurred 62 minutes later. Find the residual of this eruption and interpret the residual.

$$\hat{y} = b_0 + b_1 x = 31.013 + 9.790 \times 3.3 = 63.32$$

$$Residual = y - \hat{y} = 62 - 63.32 = -1.32$$

The eruption in the data set lasted 3.3 minutes is below the Regression line 1.32 mins.

3. Angle and Distance **(Data: Baseball)**

1) Let's fit a linear model (straight line) to the data.

(a) Using the results, write the regression equation.

$$y \sim 1.55526x + 169.514$$

(b) Report the correlation coefficient (r) and the R-squared value(r^2).

$$r = 0.5125$$

$$r^2 = 0.2627$$

2) Let's fit a quadratic model (parabola) to the data.

(a) Using the results, write the regression equation.

$$y \sim -0.173714x^2 + 14.5212x - 21.8977$$

(b) Report the R-squared value (R^2).

$$R^2 = 0.9812$$

(c) The left field fence is 280 feet from home plate. At what angles, to the nearest degree, will the ball be hit past the left field fence? **Note**: Look at the data set or the scatterplot to answer this.

At 40 degree of angle, the ball goes 284.8 feet which hit past the left field fence at 280 feet.