In [1]:

```
# The default car dataset that comes with R
# It has 50 rows and two columns. One for speed and the other one for distance r
equired to stop
cars
```

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10
10	18
10	26
10	34
11	17
11	28
12	14
12	20
12	24
12	28
13	26
13	34
13	34
13	46
14	26
14	36
14	60
14	80
15	20
15	26
15	54
16	32
16	40
17	32
17	40
17	50
18	42
18	56
18	76
18	84
19	36
19	46
19	68
20	32

speed	dist
20	48
20	52
20	56
20	64
22	66
23	54
24	70
24	92
24	93
24	120
25	85

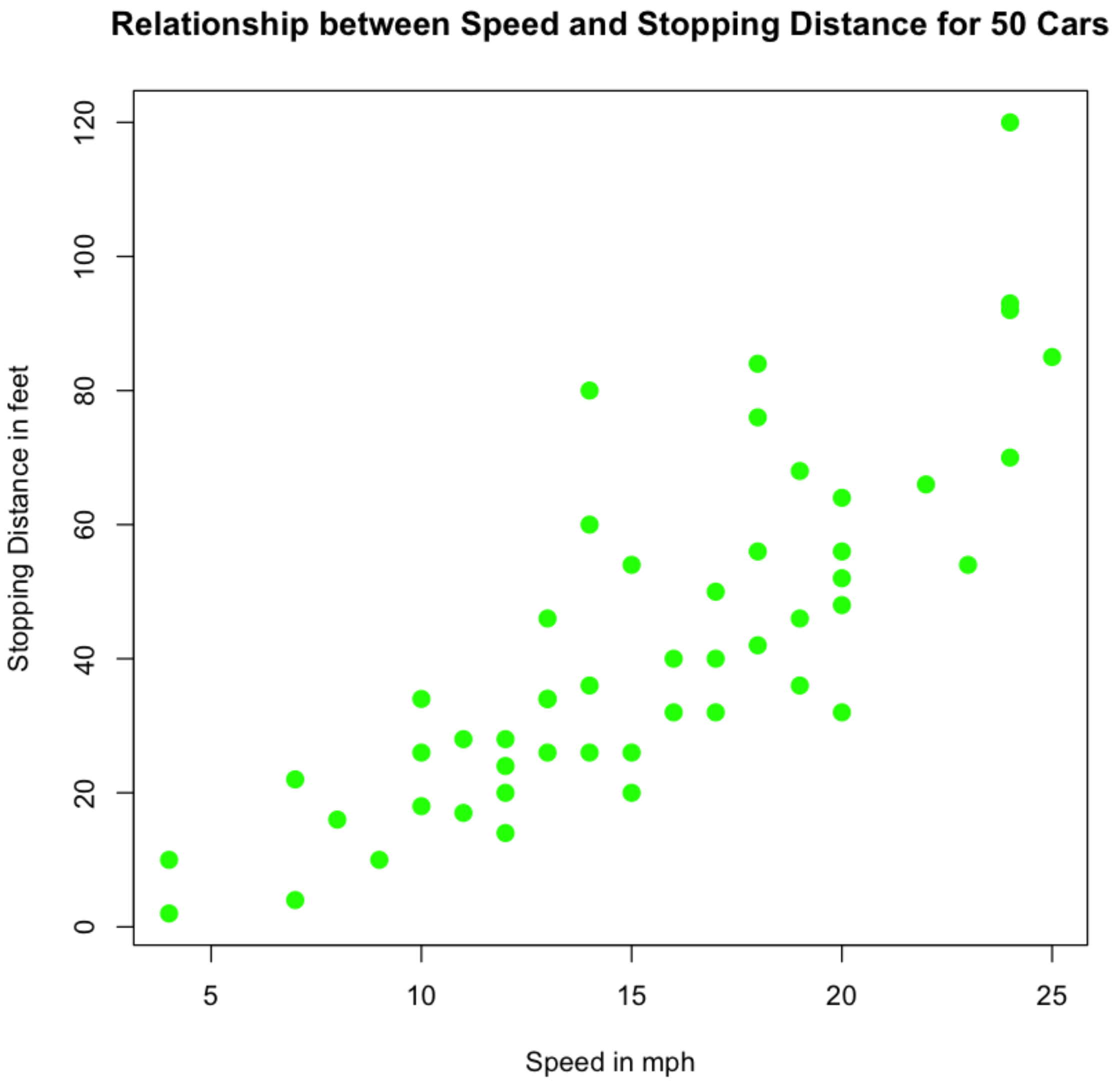
In [2]:

```
summary(cars)
```

```
     speed           dist       
 Min.   : 4.0   Min.   :  2.00  
 1st Qu.:12.0   1st Qu.: 26.00  
 Median :15.0   Median : 36.00  
 Mean   :15.4   Mean   : 42.98  
 3rd Qu.:19.0   3rd Qu.: 56.00  
 Max.   :25.0   Max.   :120.00  
```

In [4]:

```
plot(cars, col='green', pch=20, cex=2, main="Relationship between Speed and Stop
ping Distance",
xlab="Speed in mph", ylab="Stopping Distance in feet")
```


Relationship between Speed and Stopping Distance for 50 Cars
Stopping Distance in feet
0
20
40
60
80
100
120
5
10
15
20
25
Speed in mph

In [ ]:

```
# We know for a fact that there was a relationship between speed and stoping dis
tance.
# The scatter plot above just proves it.
# The higher the speed it takes more distance for a car to stop.
```

In [21]:

```
# The linear Regression Model
speed_scaled = scale(cars$speed, center=TRUE, scale=FALSE)
linear_regression_model=lm(dist~speed_scaled,data=cars)
```

In [22]:

```
summary(linear_regression_model)
```

```
Call:
lm(formula = dist ~ speed_scaled, data = cars)

Residuals:
    Min      1Q  Median      3Q     Max 
-29.069  -9.525  -2.272   9.215  43.201 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   42.9800     2.1750  19.761  < 2e-16 ***
speed_scaled   3.9324     0.4155   9.464 1.49e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.38 on 48 degrees of freedom
Multiple R-squared:  0.6511,	Adjusted R-squared:  0.6438 
F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

In [19]:

```
# This is how we interpret the model
# If speed is increase by 1 unit it takes 42.98+(3.93*1) ft to stop the car
# Standard error means the car can stop between +-3.17 ft from  42.98 ft when th
e speed decreases by 3.9 & +- 0.42mph
# t value tells us how far we are from 0 with respect to estimates.
# px(>|t|) means what is the proability of model output being > t value. < 5% is
accepted
# Residual standard error means how many times the model will be correct and it
 happens to be close to 40% [15.38/42.98]
# Multiple R-squared: this tells us how well the model is fitting the data. It h
as to be far away from 0
#F-statistic: This simply tells us the relationship between input and output. Th
e higher the better
```

In [ ]:

```
# What can we do with this interpretation
```

In [ ]:

```
# If someone says that they were in a car and they came to a clear stop at 100ft
after hitting the brake
# and ask you what thier initial speed was. You can plug in the information and
 get your answer.
# You will be accurate 2/3 of the time with 95% confidence
```

In [ ]:

```

```