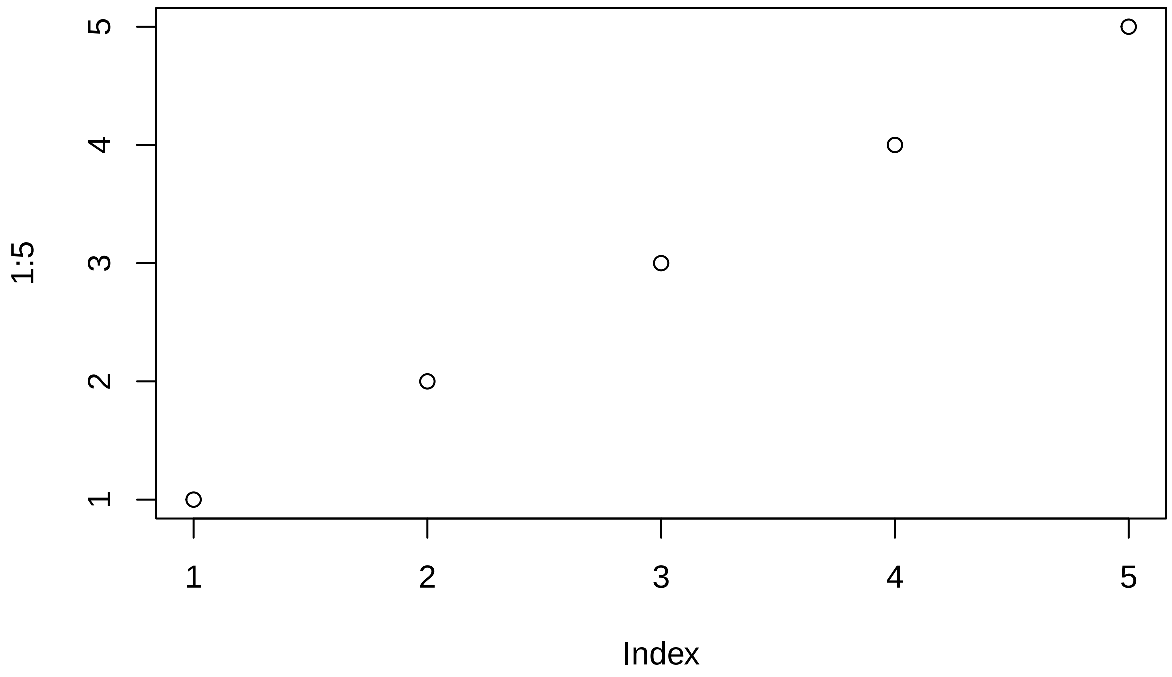
class5script.R

Rachel

2022-02-02

```
#Class 5 Data Visualization

plot(1:5)

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

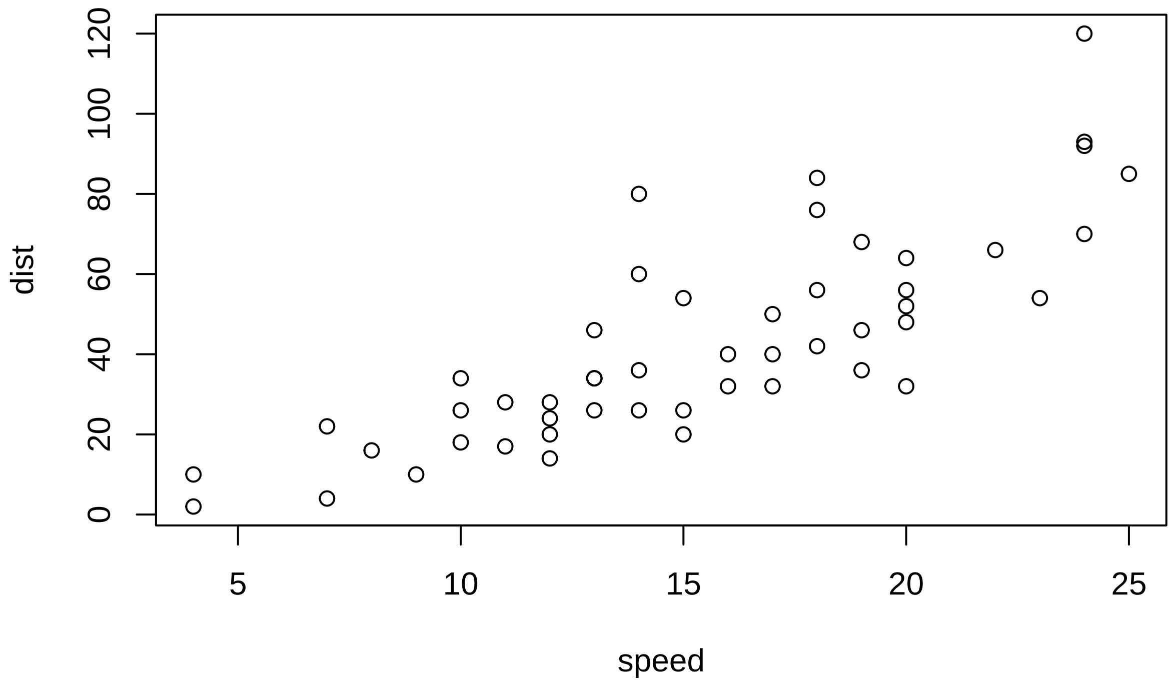


```
#Built-in dataset in R: cars
cars
```

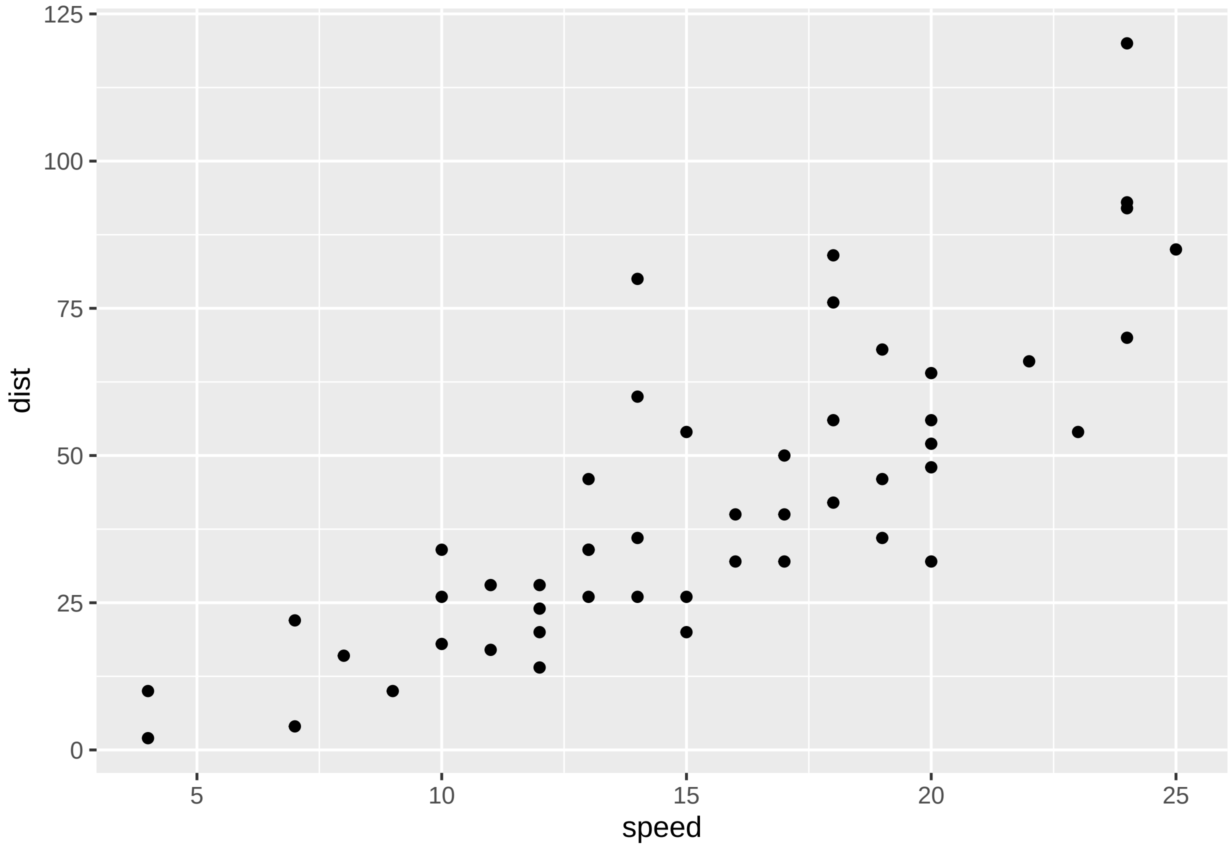
```
##    speed dist
## 1      4    2
## 2      4   10
## 3      7    4
## 4      7   22
## 5      8   16
## 6      9   10
## 7     10   18
## 8     10   26
## 9     10   34
## 10    11   17
## 11    11   28
## 12    12   14
## 13    12   20
## 14    12   24
## 15    12   28
## 16    13   26
## 17    13   34
## 18    13   34
## 19    13   46
## 20    14   26
## 21    14   36
## 22    14   60
## 23    14   80
## 24    15   20
## 25    15   26
## 26    15   54
## 27    16   32
## 28    16   40
## 29    17   32
## 30    17   40
## 31    17   50
## 32    18   42
## 33    18   56
## 34    18   76
## 35    18   84
## 36    19   36
## 37    19   46
## 38    19   68
## 39    20   32
## 40    20   48
## 41    20   52
## 42    20   56
## 43    20   64
## 44    22   66
## 45    23   54
## 46    24   70
## 47    24   92
## 48    24   93
## 49    24  120
## 50    25   85
```

```
#Function View() allows you to visualize the actual table in a separate R tab
View(cars)
```

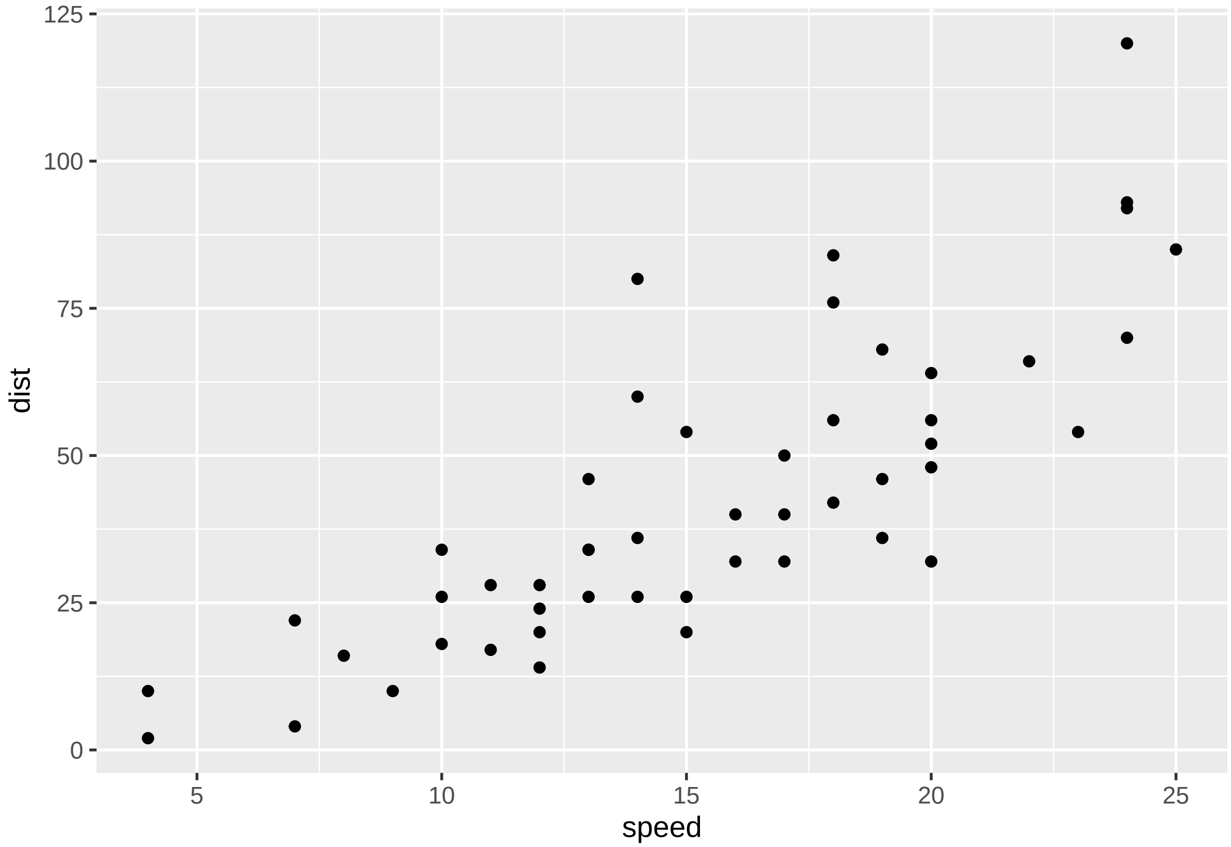
```
plot(cars)
```



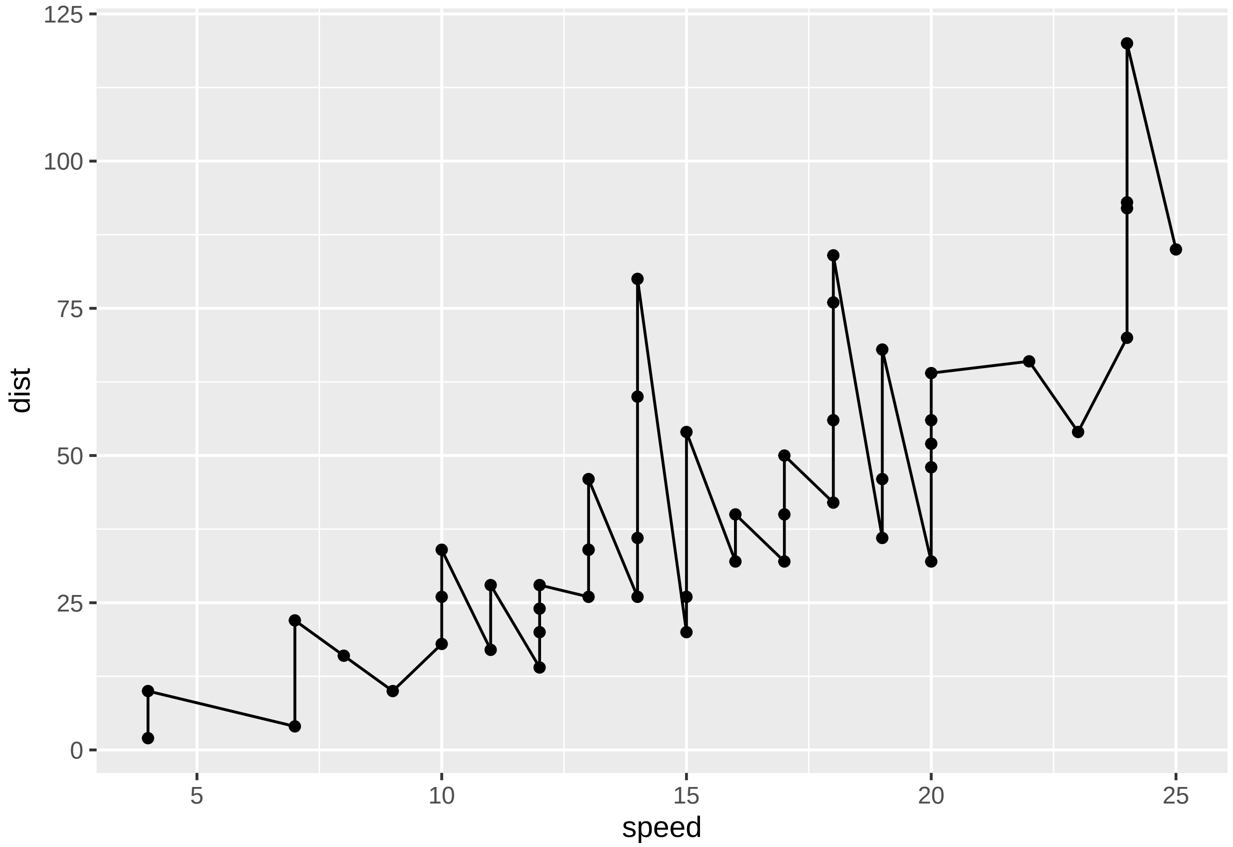
```
#Each ggplot funtion needs at least 3 layers: data, aesthetic, geometry
#ggplot(data= ) + aes() + geom_something()

ggplot(data=cars) + aes(x=speed, y=dist) + geom_point()
```



```
#Common aesthetics: position, size, line type, line width, color, shape...
#i.e. geom_line(), geom_col()
#SEE GGPLOT CHEAT SHEET FOR OTHER AESTHETICS AND GEOMETRIES!!!

p <- ggplot(data=cars) + aes(x=speed, y=dist) + geom_point()
p
```

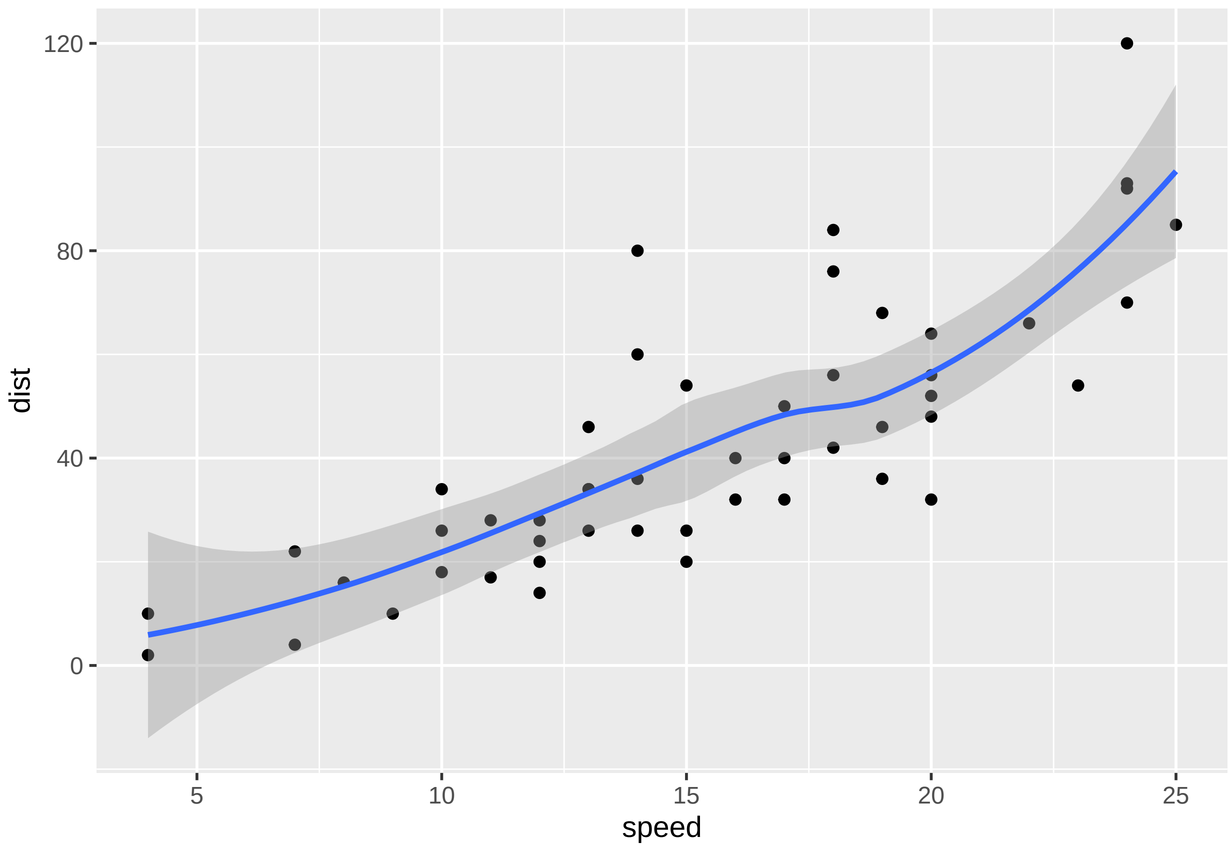


```
p + geom_line()
```



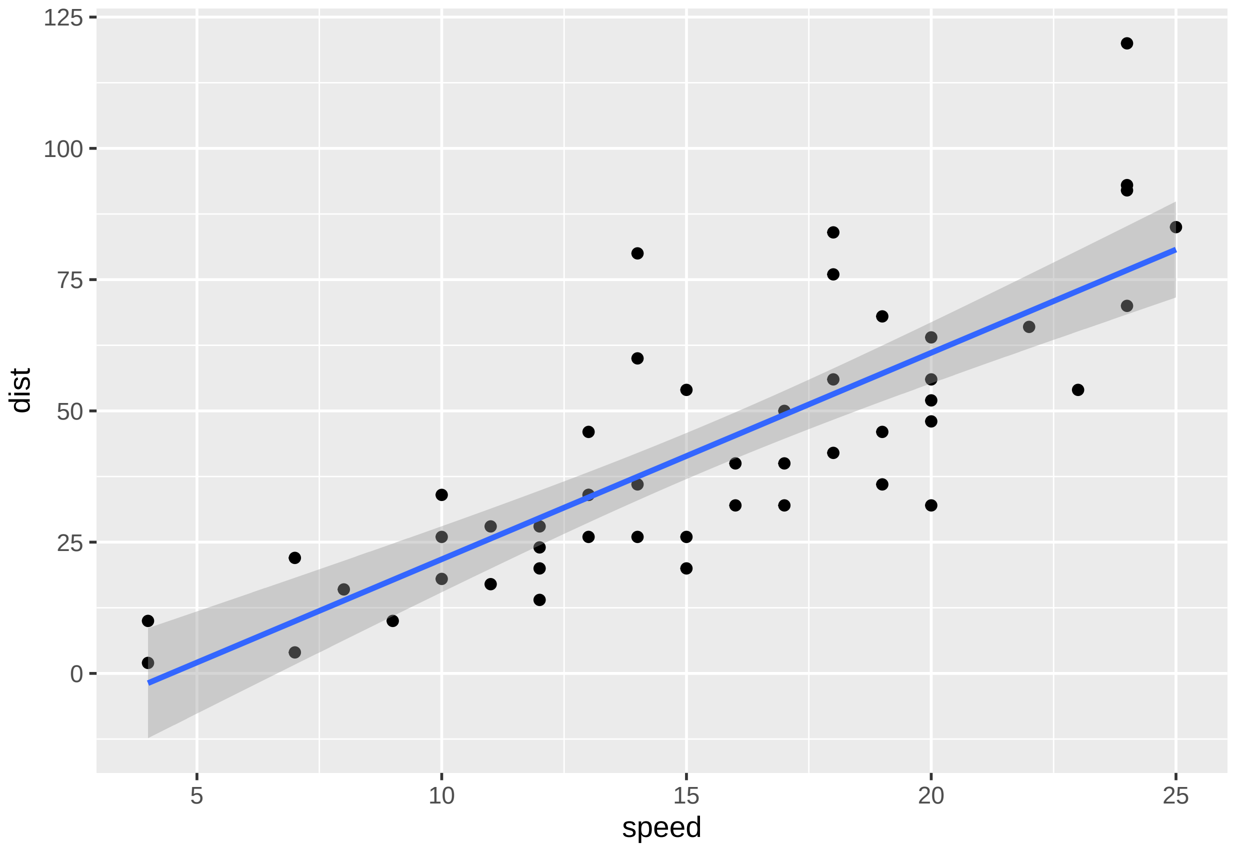
```
#this would be the same as adding + geom_line() in addition to geom_point() in the
#original ggplot line of code

p + geom_smooth() #adds a best fit curve to the graph (smoooths things out)
```

```
## ‘geom_smooth()‘ using method = ’loess’ and formula ’y ~ x’
```

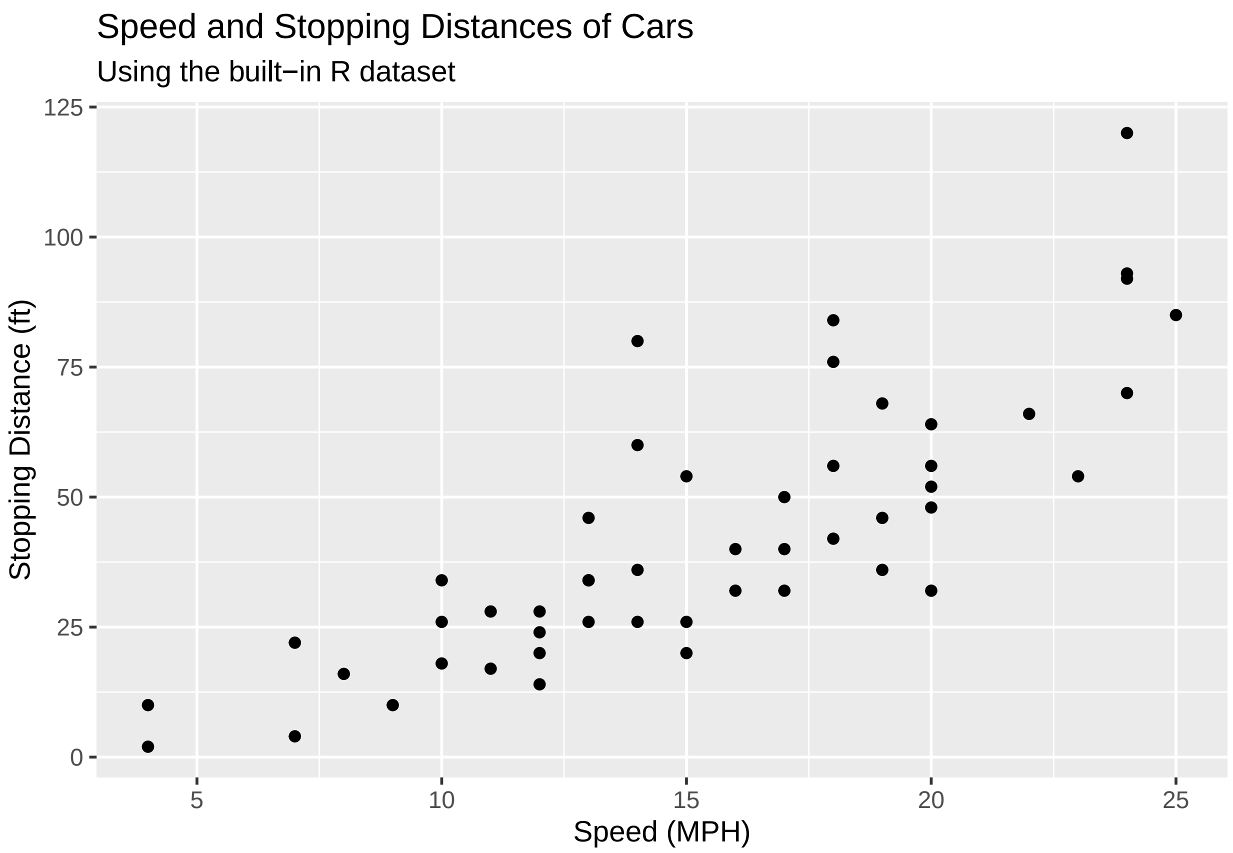


```
p + geom_smooth(method='lm') #this makes it into a best fit LINE
```

```
## ‘geom_smooth()‘ using formula ’y ~ x’
```



```
#Add labels
#Use shift+enter to make multi-line code!
p + labs(title = 'Speed and Stopping Distances of Cars',
         subtitle = 'Using the built-in R dataset') +
  xlab('Speed (MPH)') + ylab('Stopping Distance (ft)')
```



```
#NOW DO THE RNA-SEQ DATA
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
head(genes)
```

```
##          Gene Condition1 Condition2      State
## 1       A4GNT -3.6808610 -3.4401355 unchanging
## 2        AAAS  4.5479580  4.3864126 unchanging
## 3       AASDH  3.7190695  3.4787276 unchanging
## 4        AATF  5.0784720  5.0151916 unchanging
## 5        AATK  0.4711421  0.5598642 unchanging
## 6 AB015752.4 -3.6808610 -3.5921390 unchanging
```

```
View(genes)

#Number of genes in the dataset? - number of rows with nrow()
nrow(genes) #5196 genes
```

```
## [1] 5196
```

```
colnames(genes)
```

```
## [1] "Gene"       "Condition1" "Condition2" "State"
```

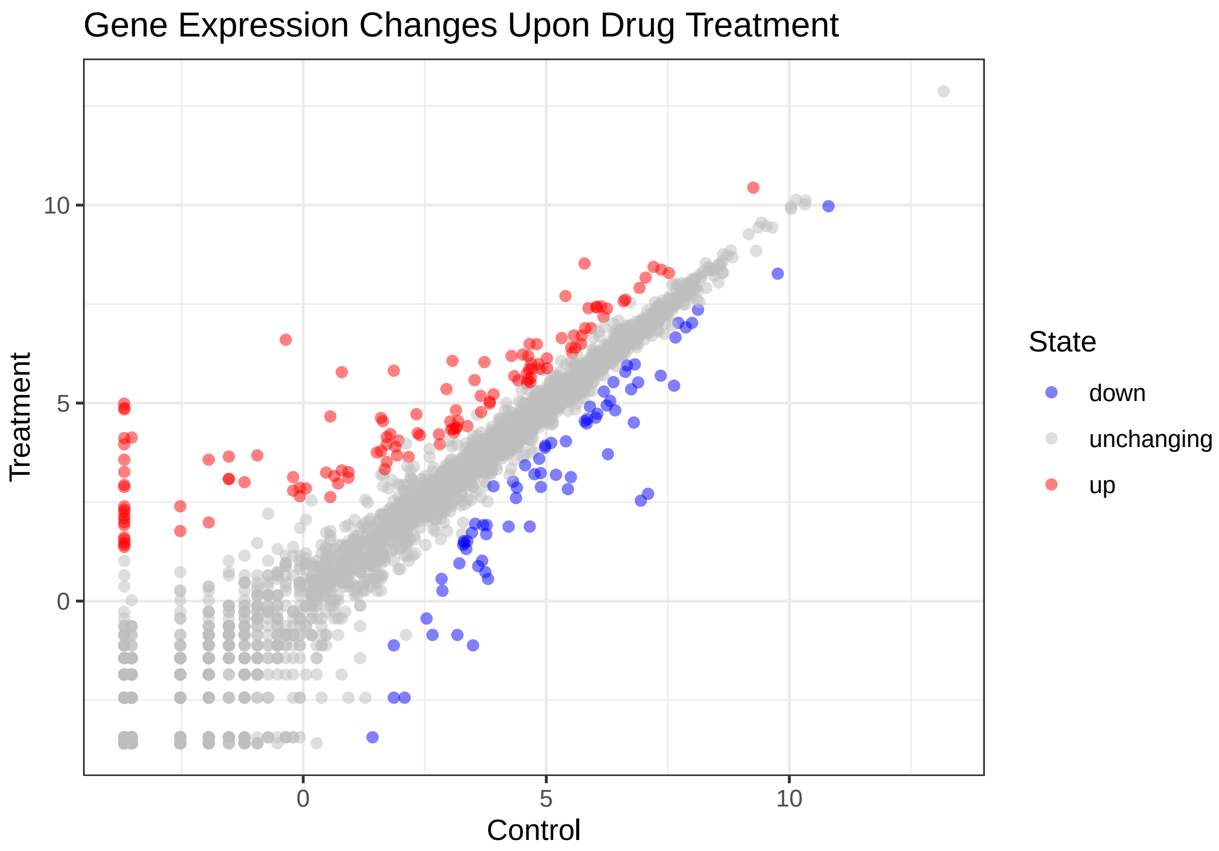
```
ncol(genes) #4
```

```
## [1] 4
```

```
#Tabulate the number of genes belonging to the categories
table(genes['State'])
```

```
##
##        down unchanging         up
##          72       4997        127
```

```
#Now graph condition1 vs condition2 expression values
rnaseq <- ggplot(data = genes) + aes(x = Condition1, y = Condition2) + geom_point()

#Change color based on another categorical variable on the same data (State!)
#can also change transparency of the points (fill) with geom_point(alpha=[insert value])
rnaseq_col <- ggplot(data = genes) + aes(x = Condition1, y = Condition2, col = State) + geom_point(alpha

#Default colors are ugly, change them!
#Red will be what's upregulated, blue will be what's downregulated
final <- rnaseq_col + scale_colour_manual(values=c("blue","gray","red"))

#Now add annotations
final + labs(title = 'Gene Expression Changes Upon Drug Treatment') +
  xlab('Control') + ylab('Treatment') + theme_bw()
```


Gene Expression Changes Upon Drug Treatment
Treatment
Control
State
● down
● unchanging
● up