class05: Data Visualization Lab

Saba Heydari Seradj

Packages were installed, and then left commented.

```
# Install the package ggplot2
# install.packages("ggplot2")
```

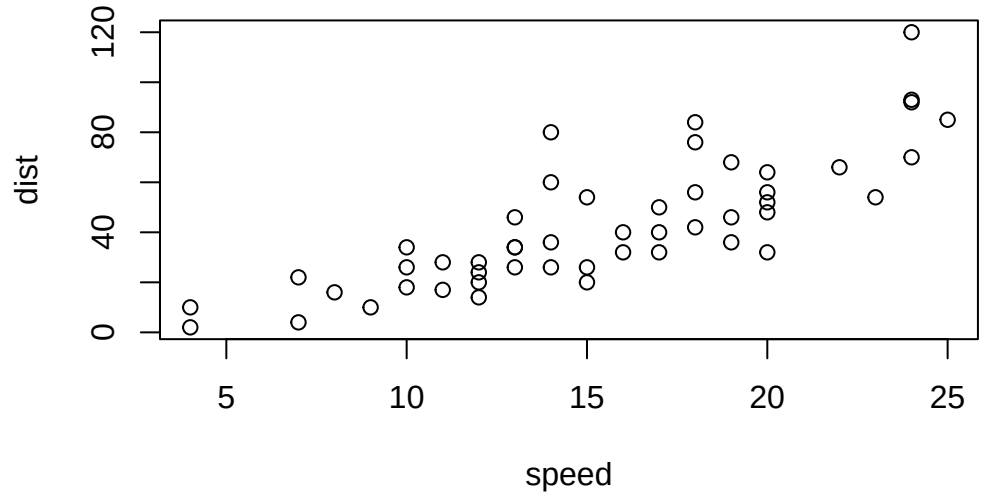
Any time I want to use this package I need to load it.

```
library(ggplot2)
```

```
View(cars)
```

A quick base R plot (not ggplot):

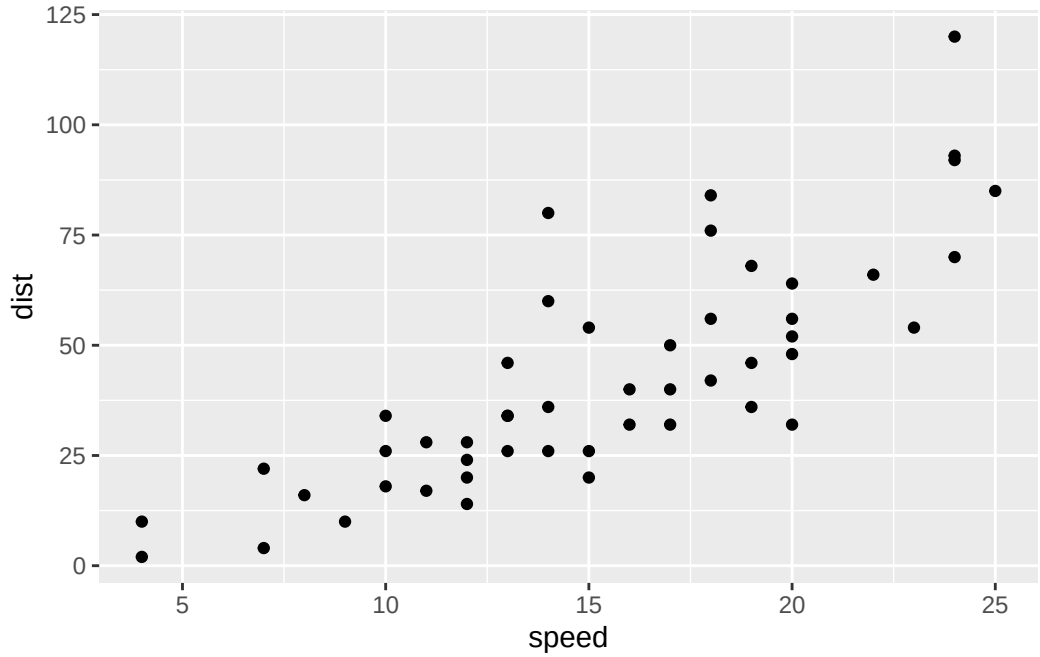
```
plot(cars)
```



Our first ggplot!

We need data + aes + geoms

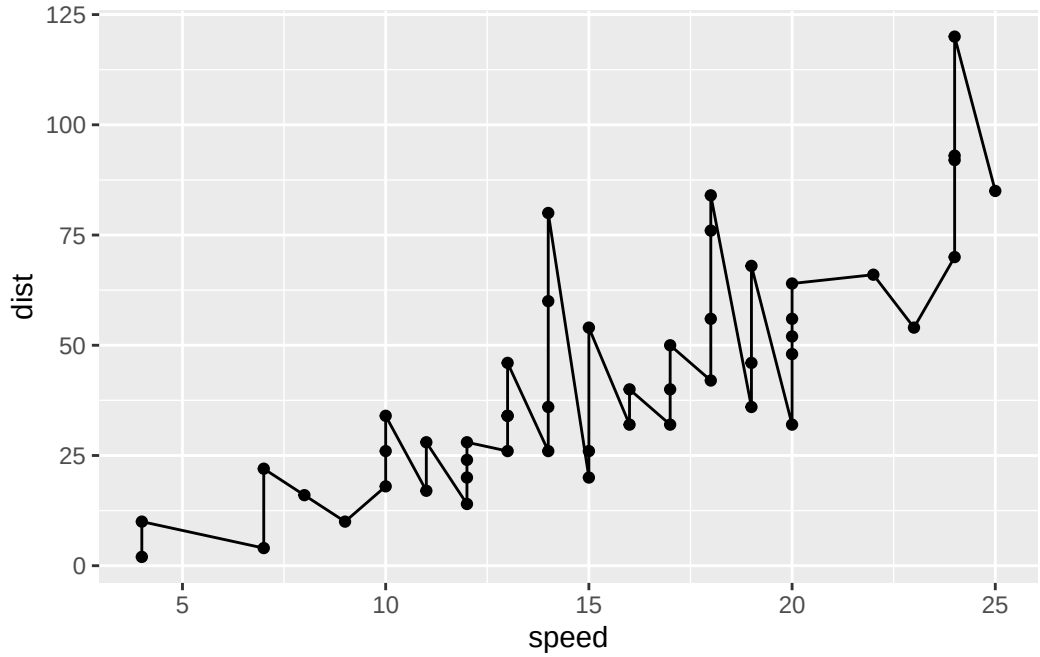
```
ggplot(data=cars) +
  aes(x=speed, y=dist) +
  geom_point()
```




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

Add a line geom with geom_line() which connects the points

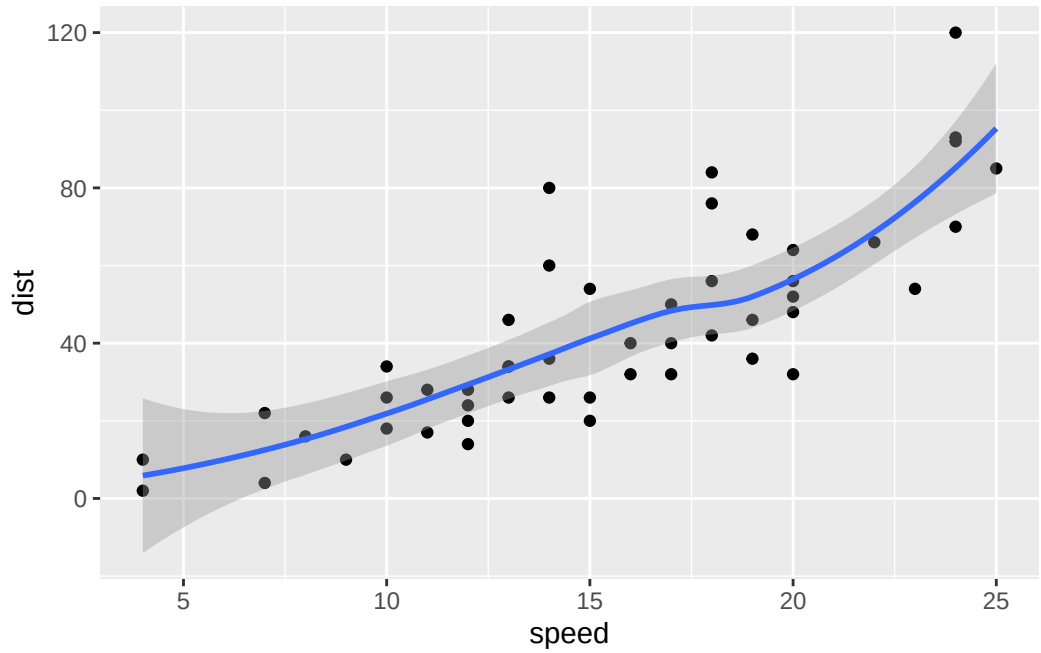
```
p + geom_line()
```



Add a trend line to the data:

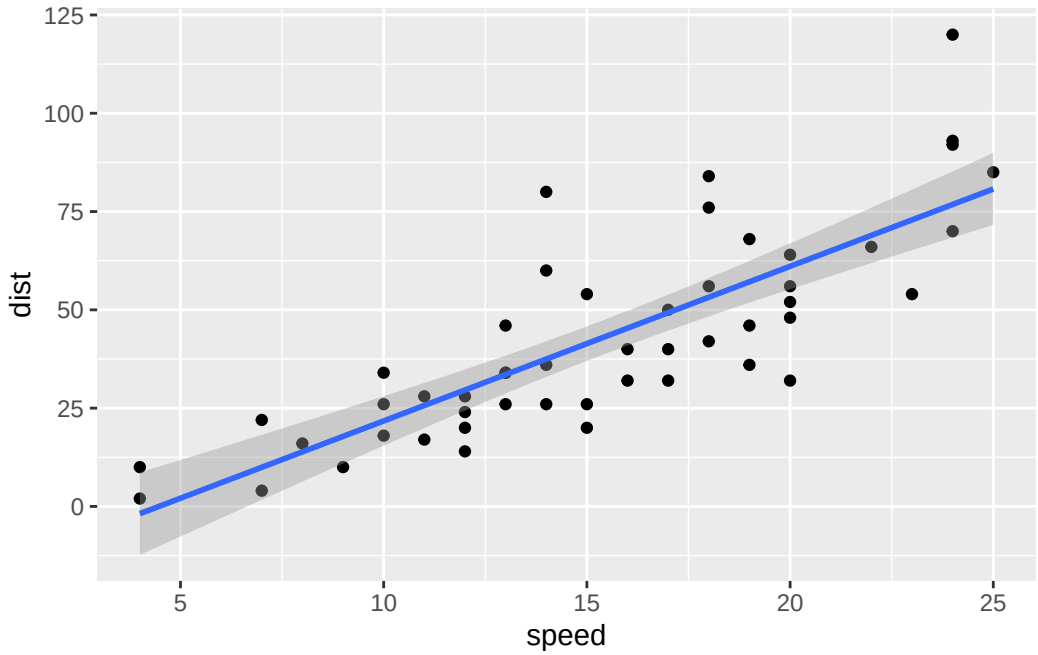
```
p + geom_smooth()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
# trendline using linear model
p + geom_smooth(method='lm')
```

```
`geom_smooth()` using formula = 'y ~ x'
```



Reading in our drug expression data:

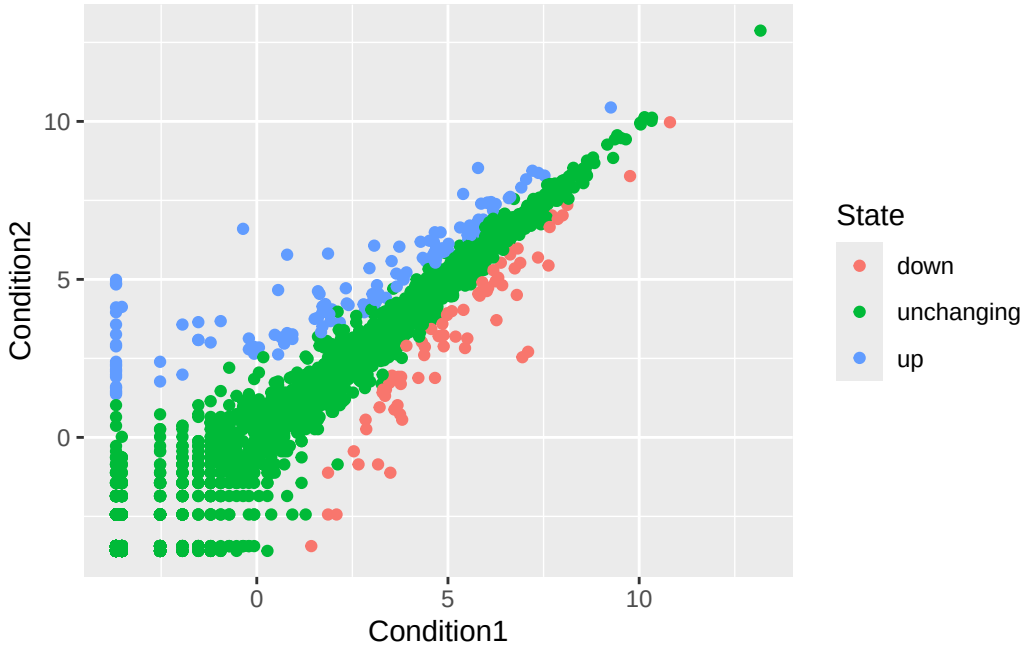
```
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
head(genes)
```

```
        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
```

Let's make a first plot attempt:

```
g <- ggplot(data=genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point()

g
```



Q. Use the nrow() function to find out how many genes are in this dataset. What is your answer?

```
nrow(genes)
```

```
[1] 5196
```

There are 5196 genes in the dataset.

Q. Use the colnames() function and the ncol() function on the genes data frame to find out what the column names are (we will need these later) and how many columns there are. How many columns did you find?

```
ncol(genes)
```

```
[1] 4
```

```
colnames(genes)
```

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

Q. Use the table() function on the State column of this data.frame to find out how many 'up' regulated genes there are. What is your answer?

```
table(genes$State)
```

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

Q. Using your values above and 2 significant figures. What fraction of total genes is up-regulated in this dataset?

```
round(table(genes$State)/nrow(genes)*100,2)
```

```
      down unchanging         up
      1.39      96.17       2.44
```

Let's add some color with custom settings:

```
g+ scale_color_manual(values=c('blue','gray','red'))+
  labs(title='Gene Expression changes upon drug treatment',x='Control(no drug',y='Treatment')
  theme_bw()
```