

Class 5: Data Visualization

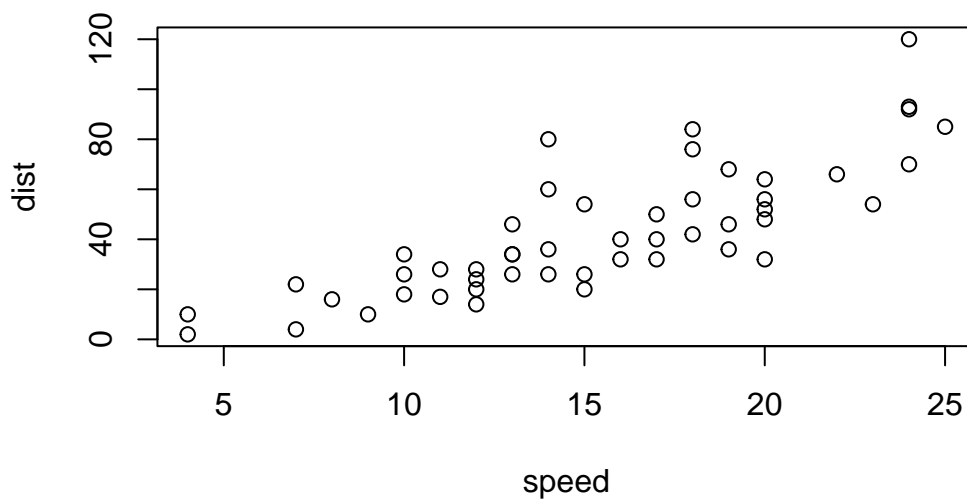
Tomi Lane Timmins

Plotting in R

R has multiple plotting and graphics systems. The most popular of which is **ggplot2**.

We have already played with “base” R graphics. This comes along with R “out of the box”.

```
plot(cars)
```



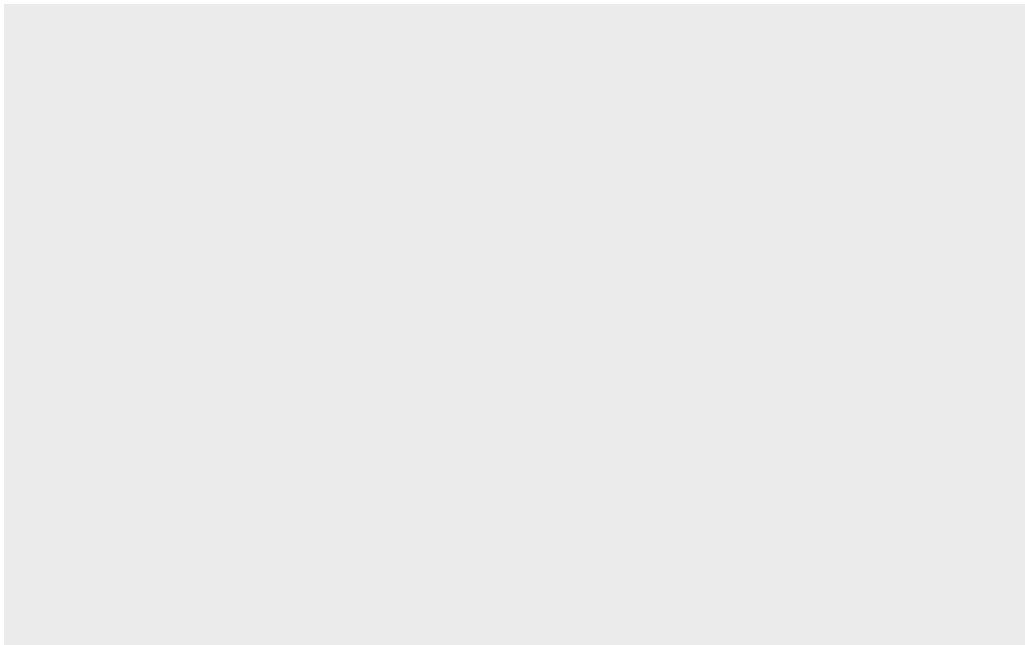
Compared to base R plots, ggplot is much more verbose - I need to write more code to get simple plots like the one above.

To use ggplot, I need to first install the ggplot2 package. To install a package in R I use the 'install.packages()' command along with the package name. (This was already completed beforehand)

The install above is a one time only requirement. The package is now on my computer - no need to re-install it. However, I can't just use it without loading it up with a 'library()' call.

```
library(ggplot2)
```

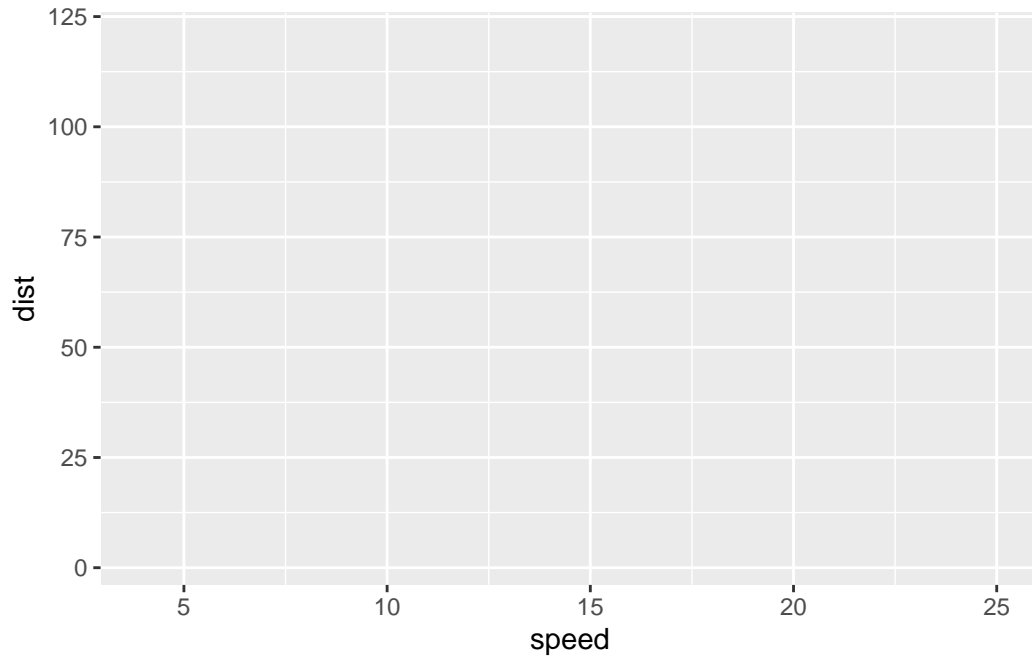
```
ggplot(cars)
```



all ggplot figures need at least 3 things:

- data (this is the data.frame with our numbers, characters, etc) + aesthetics ("aes", how our data maps to the plot - what is on x axis, y axis, what color, etc) + geoms (do you want lines, points, columns, etc?)

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



With only data + aesthetic, you have the start of the plot, but R doesn't know how you want it (point, line, etc)

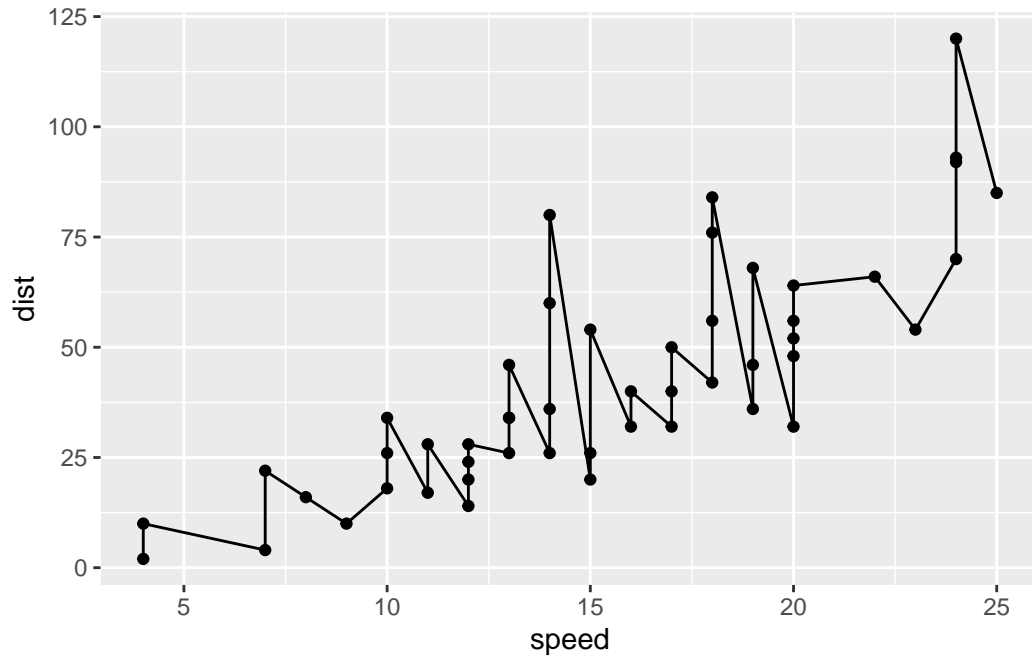
```
bb <- ggplot(data = cars) + aes(x=speed, y=dist) + geom_point()
```

```
bb
```



We want to see a line to show a clear relationship between speed and distance.

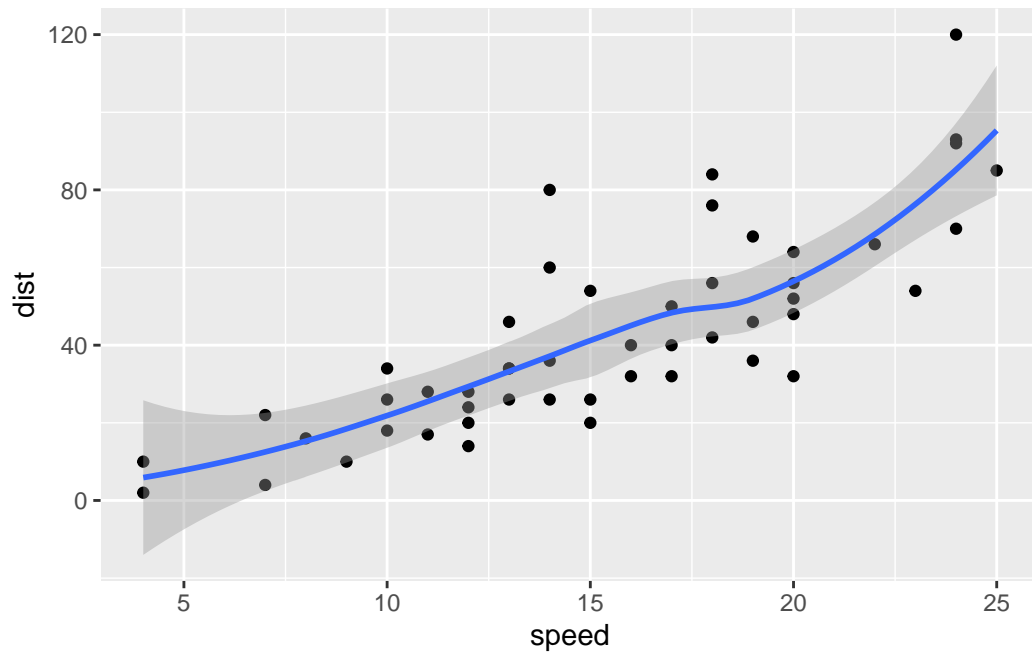
```
bb + geom_line()
```



This is not what we want.

```
bb + geom_smooth()
```

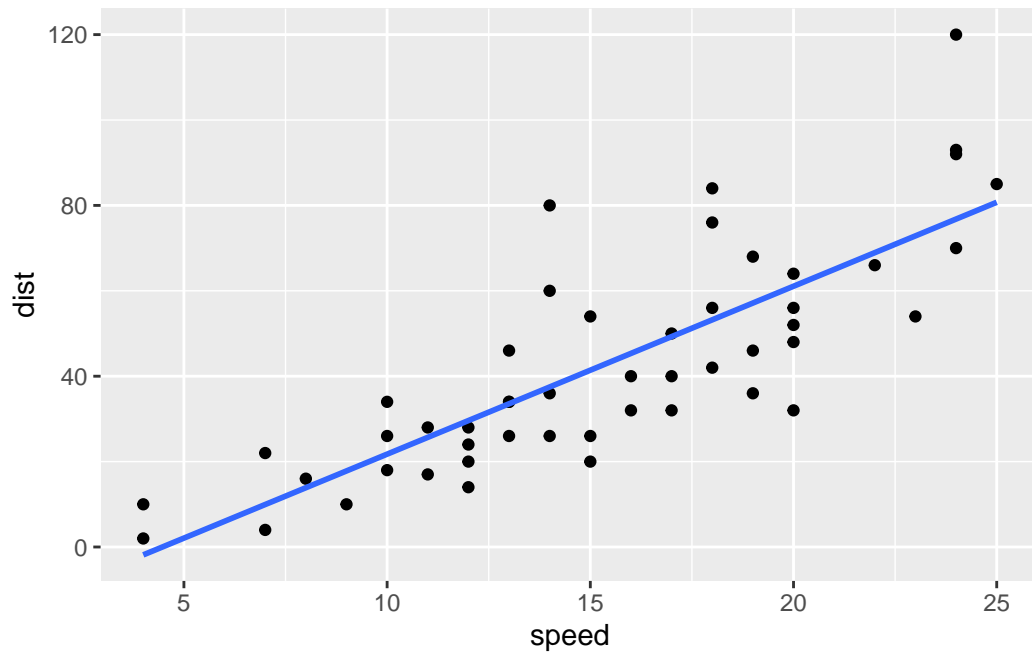
`geom_smooth()` using `method = 'loess'` and `formula = 'y ~ x'`



Remove the standard error from the plot.

```
bb + geom_smooth(method = "lm", se = FALSE)
```

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

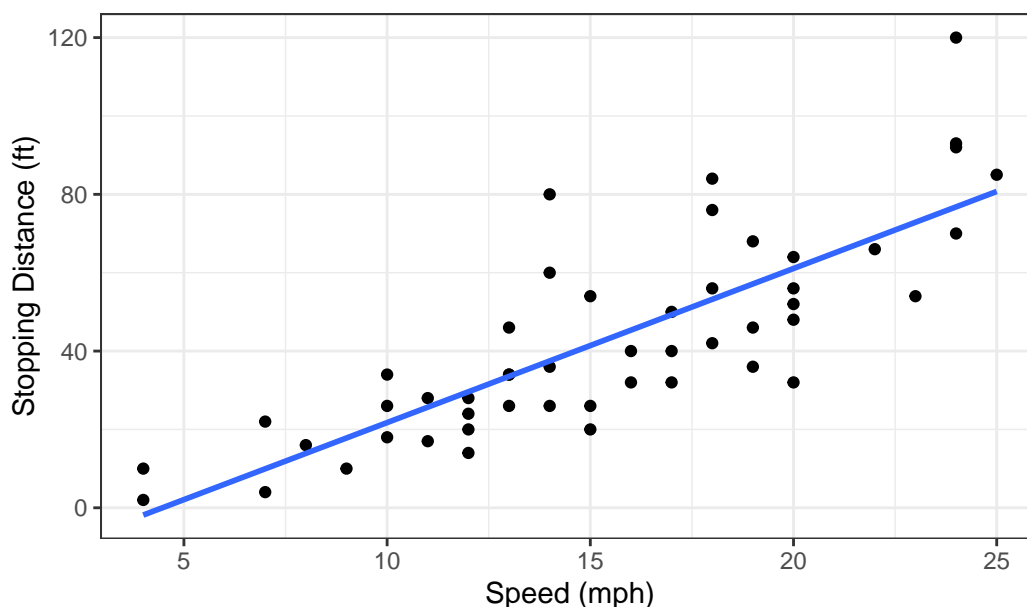


Give the plot labels and make it look more conservative.

```
bb + geom_smooth(method = "lm", se = FALSE) + labs(title = "The Speed and the Stopping Dis
```

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

The Speed and the Stopping Distance of Cars



Moving on to another dataset to show gene expression. Make sure to include 'head(genes)' - otherwise will have 5000 genes on the report. The 'head()' function will print out just the first few rows, (6 by default)

```
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
head(genes, 10)
```

	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
7	ABCA7	3.4484220	3.8266509	unchanging
8	ABCA9-AS1	-3.6808610	-3.5921390	unchanging
9	ABCC11	-3.5288580	-1.8551732	unchanging
10	ABCC3	0.9305738	3.2603040	up

The nrow() function shows how many rows are in the dataset and therefore how many genes are in the dataset.


```
nrow(genes)
```

```
[1] 5196
```

The `colnames()` function tells us the names of the columns in the ‘genes’ dataset. The `ncol()` gives the number of columns.

```
colnames(genes)
```

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

```
ncol(genes)
```

```
[1] 4
```

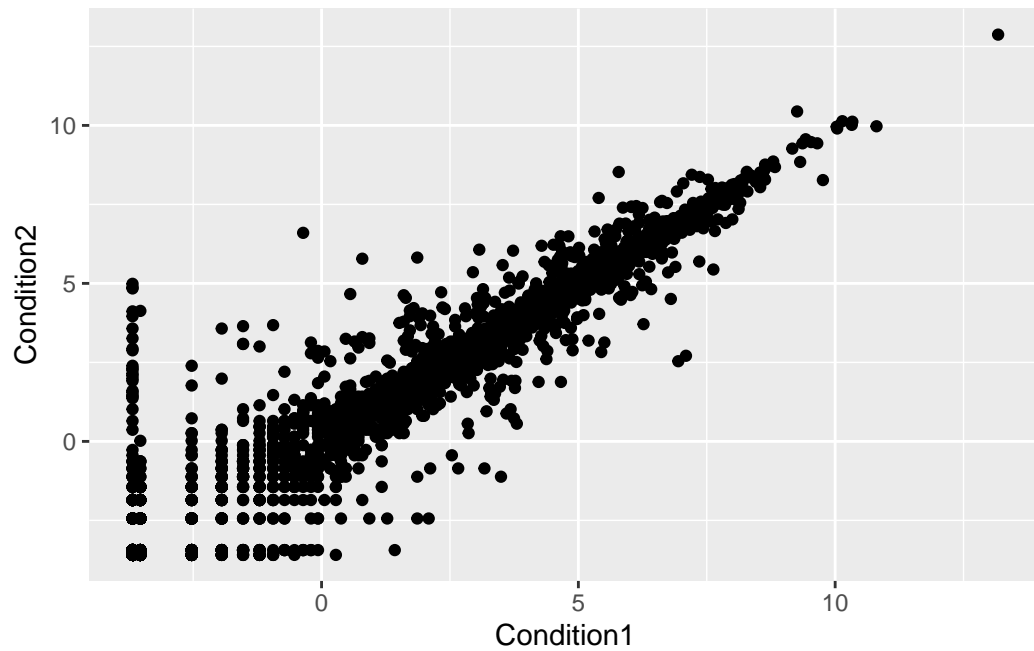
The `table()` function specifically with the `State` column tells us the number of genes in each state.

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

down	unchanging	up
72	4997	127

Plot the gene dataset.

```
ggplot(genes) + aes(x = Condition1, y = Condition2) + geom_point()
```

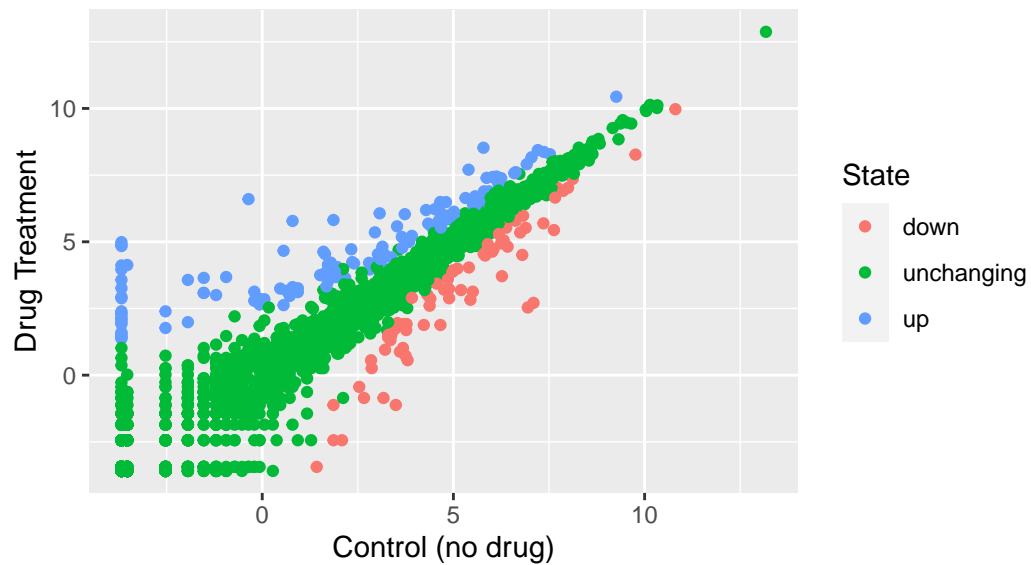


Plot to include the data about the State of the genes (up/down/unchanging). The 'color = State' automatically colors the data from the State vector onto the plot.

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

Gene Expression Changes upon Drug Treatment

The up/down regulation of varying genes under Condition1 and Condition2



We want to change the colors of the plot.

```
p + scale_colour_manual(values = c("blue", "gray", "red"))
```

Gene Expression Changes upon Drug Treatment

The up/down regulation of varying genes under Condition1 and Condition2

