

Class05

Mahsa Naeimi

Base R plotting

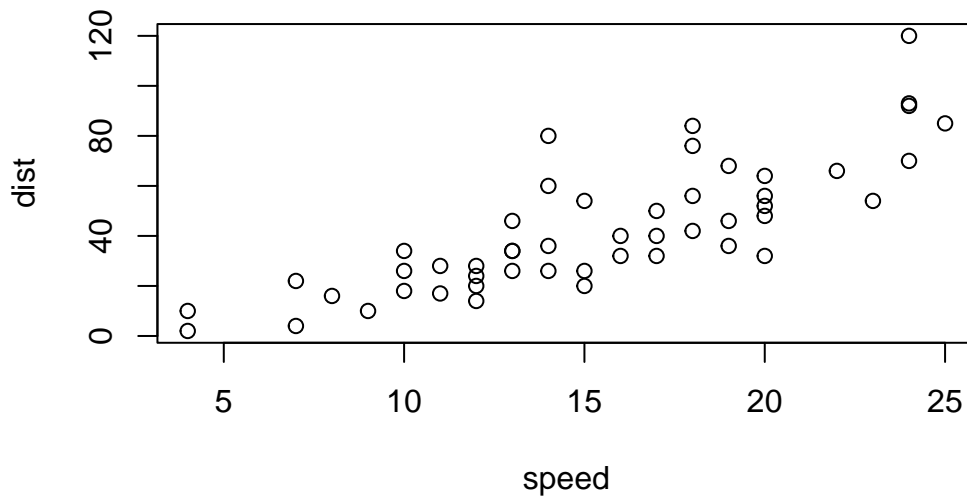
Ggplot2

First, we need to install the package. We do this by using the `install.packages` command.

```
#install.packages('ggplot2')
```

We are going to start by generating the plot of class 04. This code is plotting the cars dataset.

```
plot(cars)
```

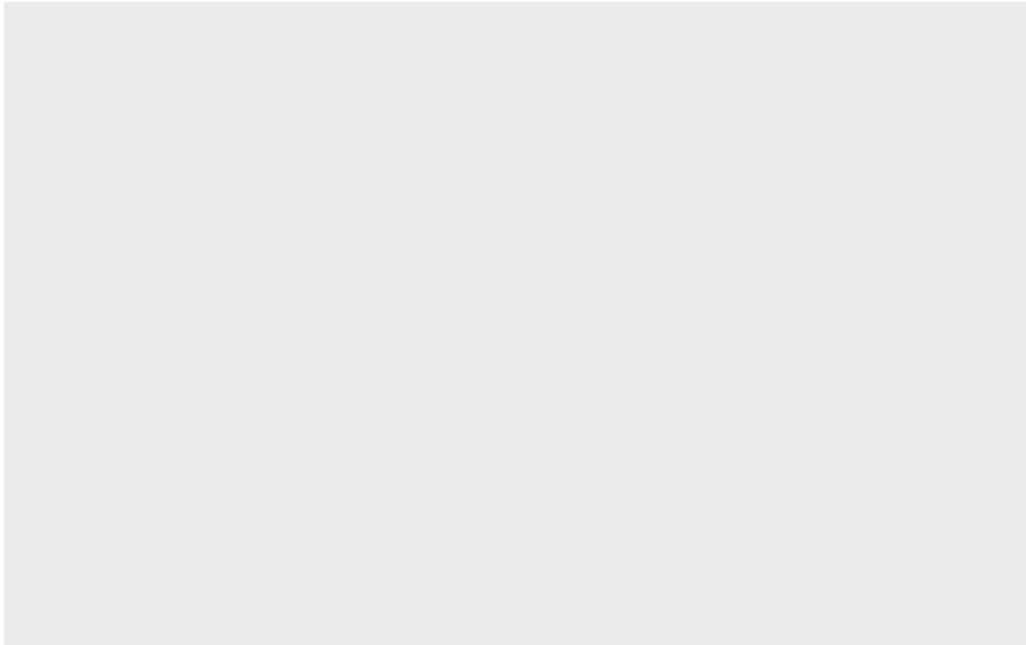


To start, we need to load the package.

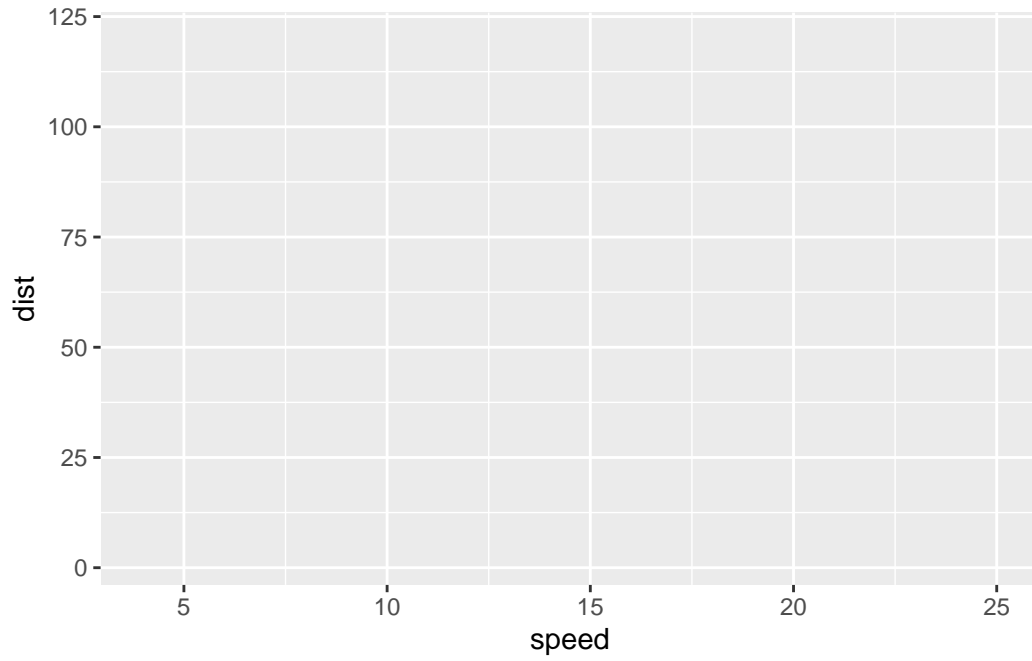
```
library(ggplot2)
```

We are going to build the plot of the cars dataframe by using ggplot2.

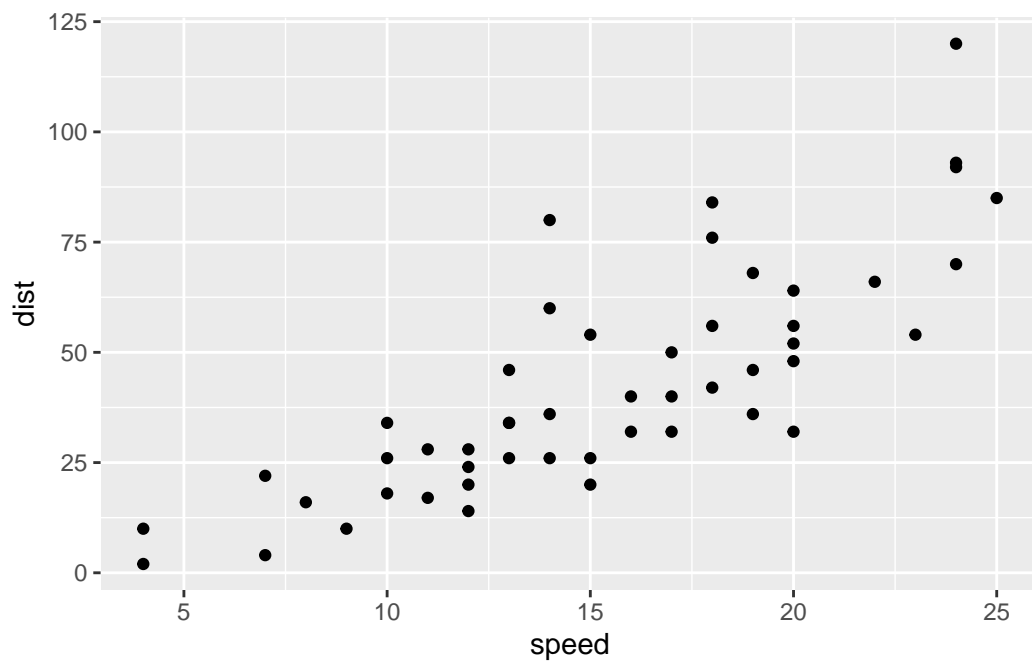
```
ggplot(cars)
```



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

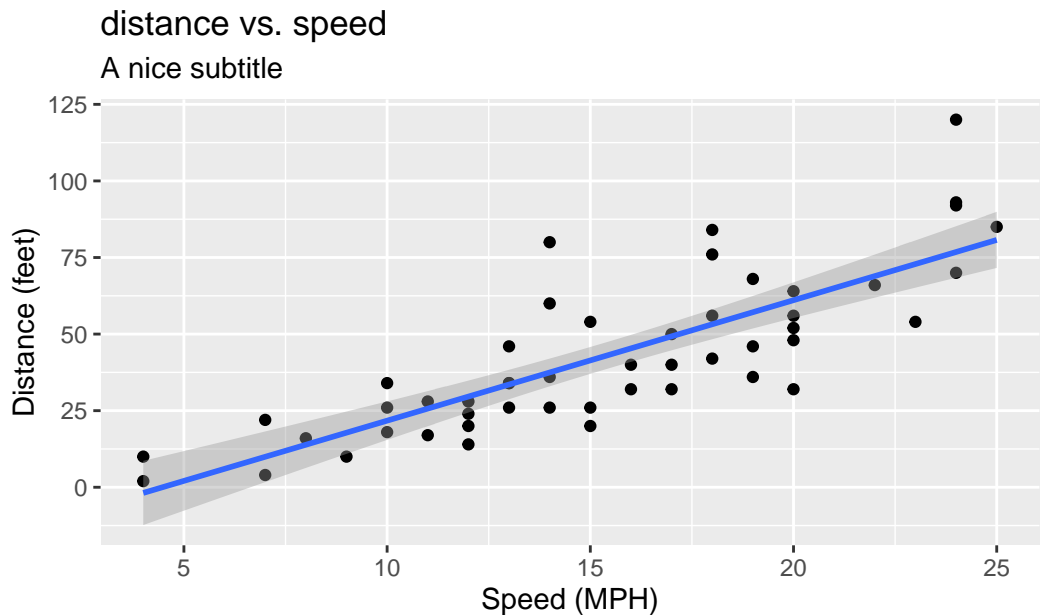


```
ggplot(data = cars) + aes(x=speed, y=dist) +geom_point()
```



```
ggplot(data = cars) + aes(x=speed, y=dist) +geom_point()+geom_smooth(method = 'lm')+ labs(
```

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



BIMM143

Q1. For which phases is data visualization important in our scientific workflows?

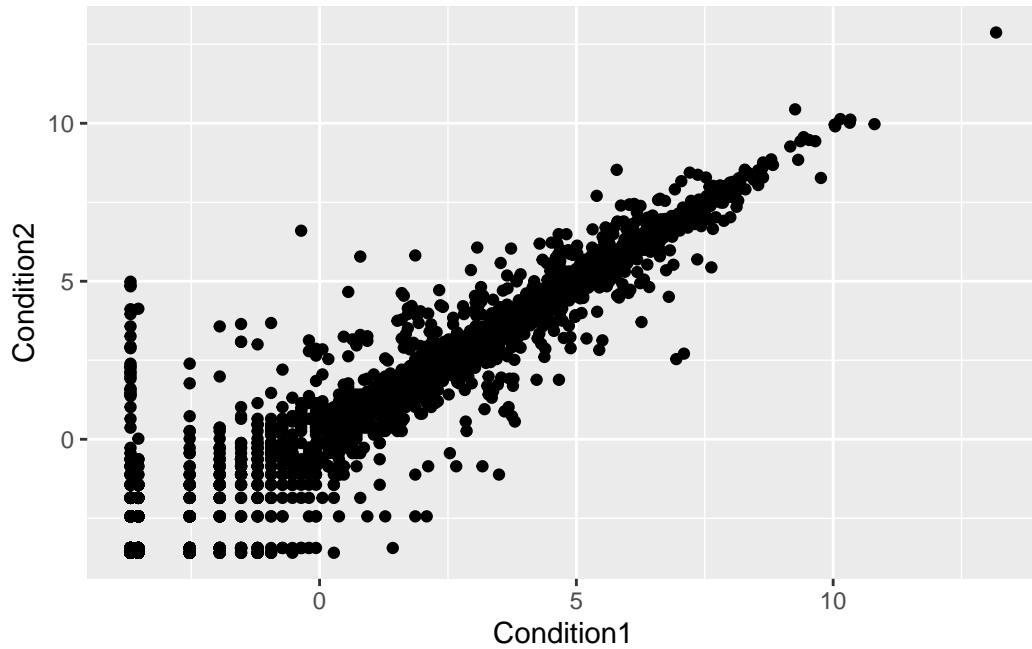
For exploratory Data Analysis, Plotting Gene Expression data, and Loading the data from the URL.

```
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

Initial ggplot

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



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

```
nrow(genes)
```

```
[1] 5196
```

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)
```

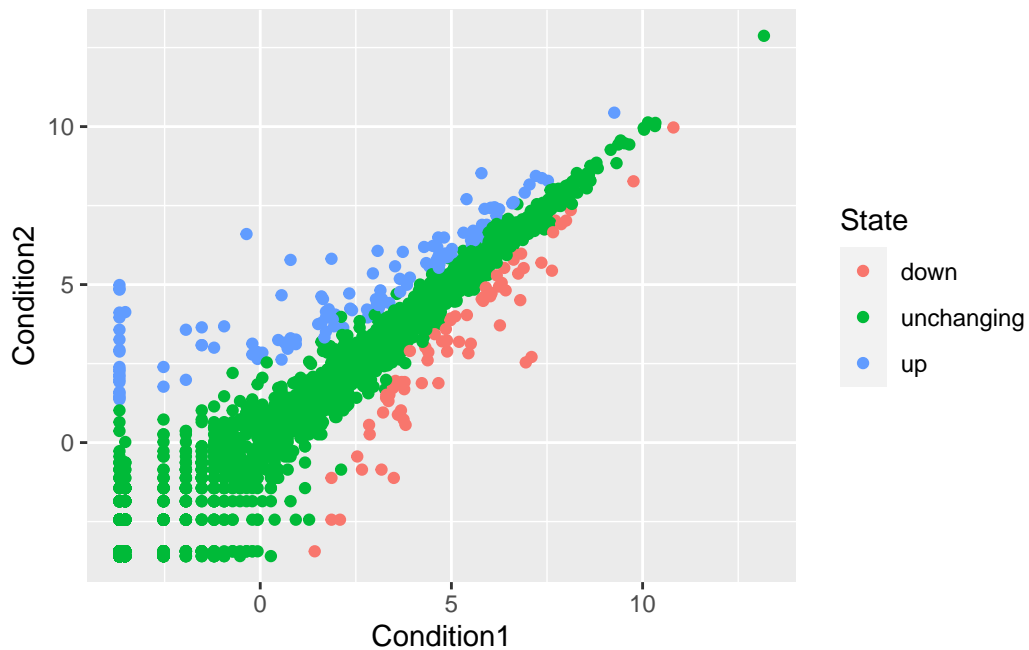
```
[1] "Gene"          "Condition1" "Condition2" "State"
```

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'])
```

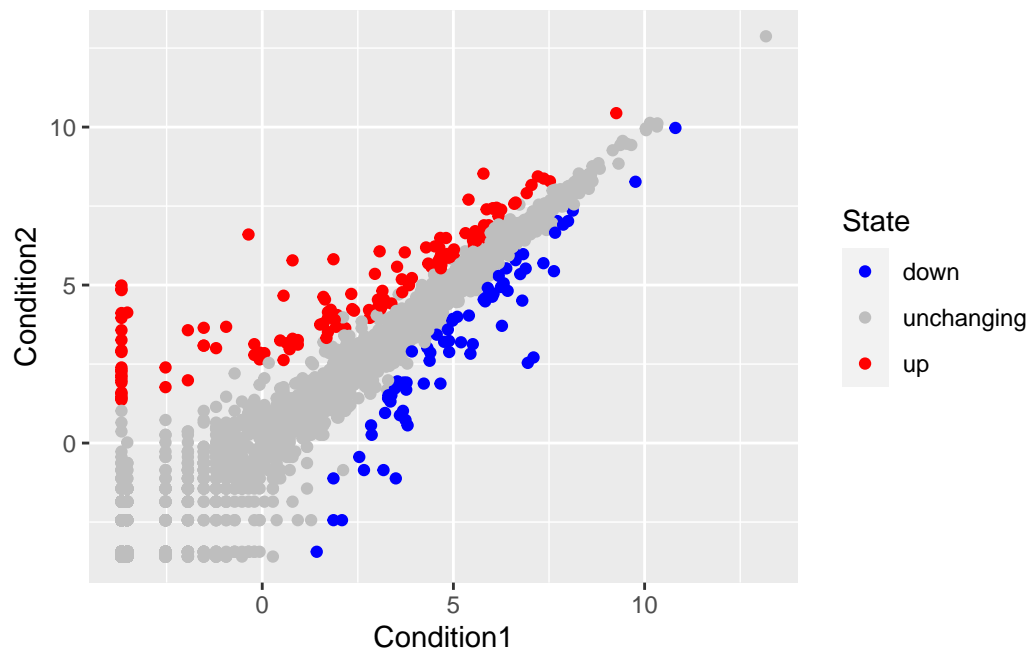
```
down  unchanging    up
   72      4997    127
```

```
p1 <- ggplot(data = genes) + aes(x=Condition1, y=Condition2, col=State) + geom_point()
p1
```



Let's change the color scheme:

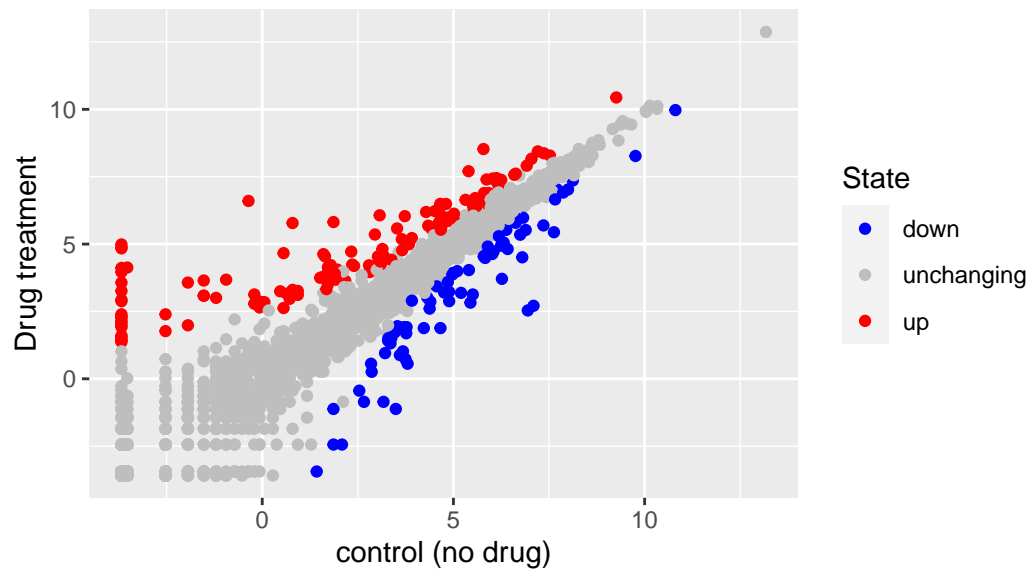
```
p2 <- p1 + scale_color_manual(values = c('blue', 'grey', 'red'))
p2
```



Let's add some labels:

```
p2 + labs(title = 'Differential gene expression', x='control (no drug)', y= 'Drug treatment')
```

Differential gene expression



BIMM 143 – Class 05