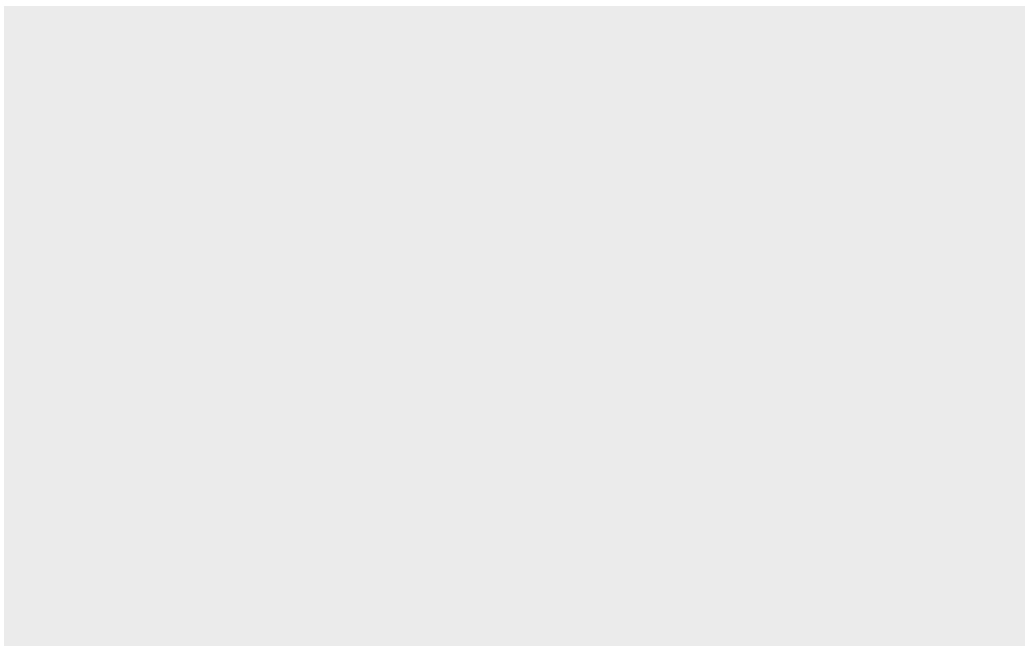# Class 05: Data Visualization with ggplot2
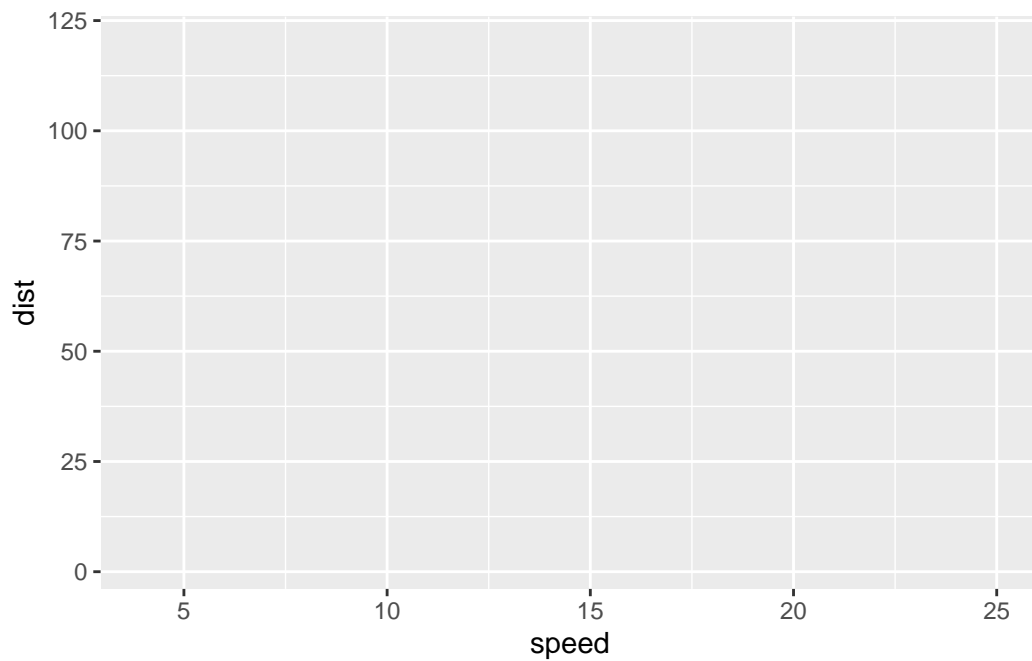
Pagna Hout

4/19/23

Load package

```
library(ggplot2)
ggplot(cars)
```

Specifying aesthetic mappings with aes ()

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

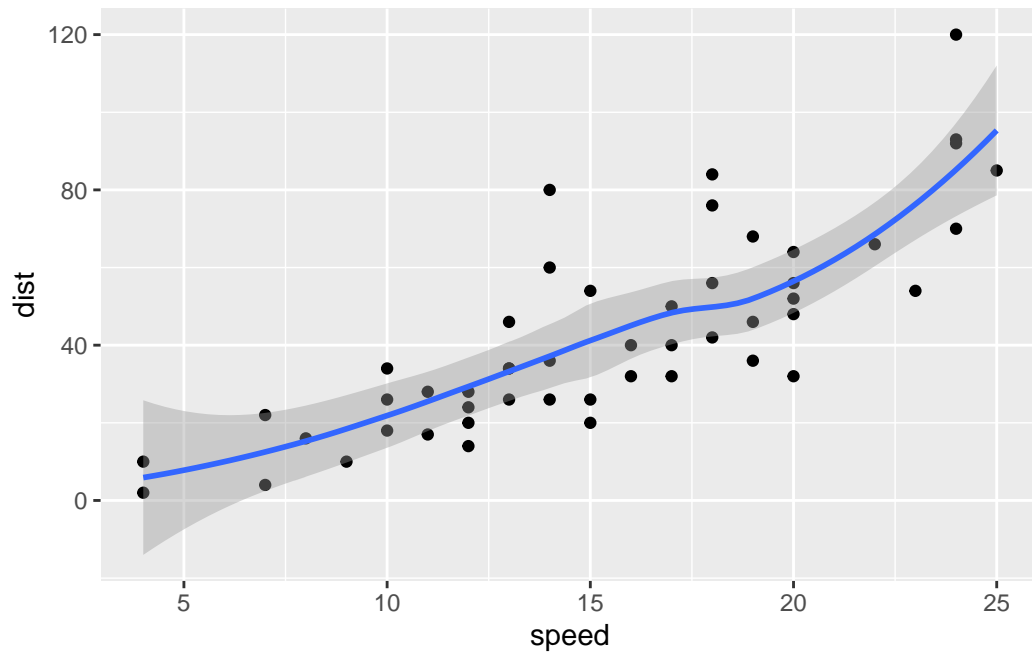Specifying a geom layer with geom_point()

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

Adding trend line layer to help show the relationship between the plot variables with the geom_smooth() function

```
ggplot(cars) +
  aes (x=speed, y=dist) +
  geom_point() +
  geom_smooth()
```
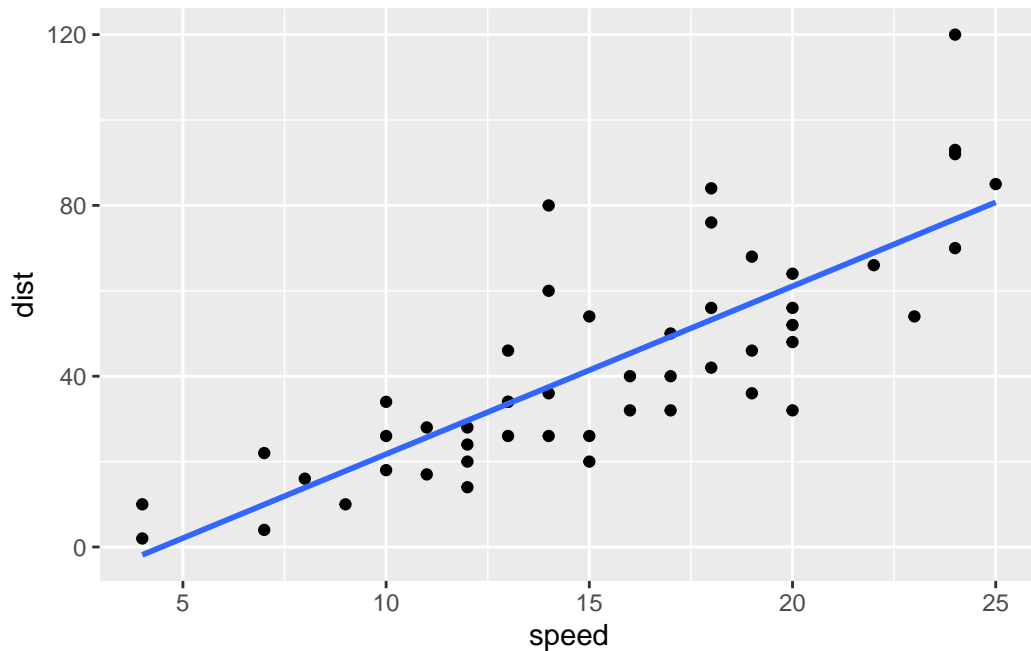
```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

Adding straight line from linear model without the shaded standard error region

```
ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE)
```

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

Adding various label annotations with the labs() function and changing the plot to look more conservative "black & white" theme by adding the theme.bw() function

```
ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point() +
  labs(title="Speed and Stopping Distances of Cars",
       x="Speed (MPH)",
       y="Stopping Distance (ft)",
       subtitle = "Your informative subtitle text here",
       caption="Dataset: 'cars'") +
  geom_smooth(method="lm", se=FALSE) +
  theme_bw()
```

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

Speed and Stopping Distances of Cars
Your informative subtitle text here
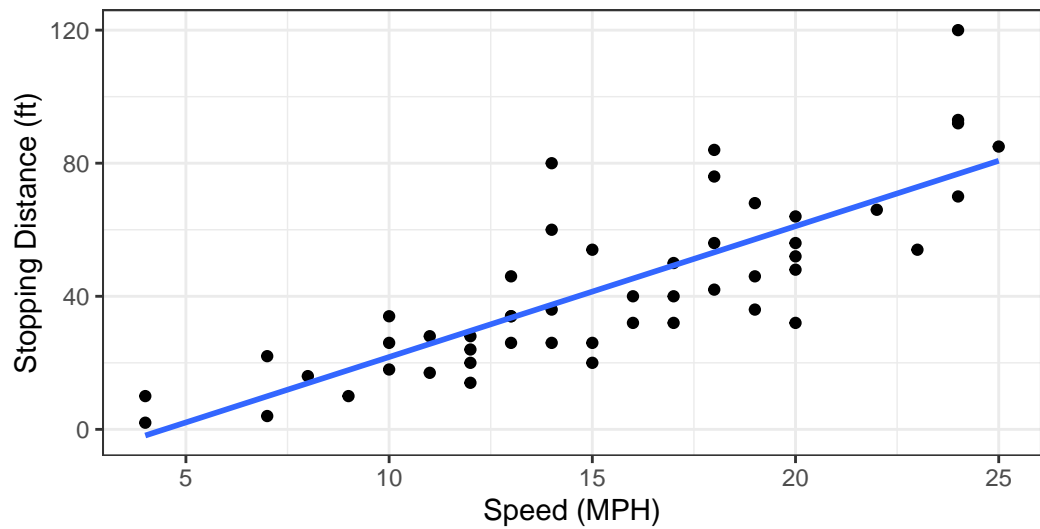
Dataset: 'cars'

Adding more plot aesthetics through aes()

```
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
```

```
nrow(genes)
```

```
[1] 5196
```

```
ncol(genes)
```

```
[1] 4
```

```
colnames(genes)
```

```
[1] "Gene"       "Condition1" "Condition2" "State"
```
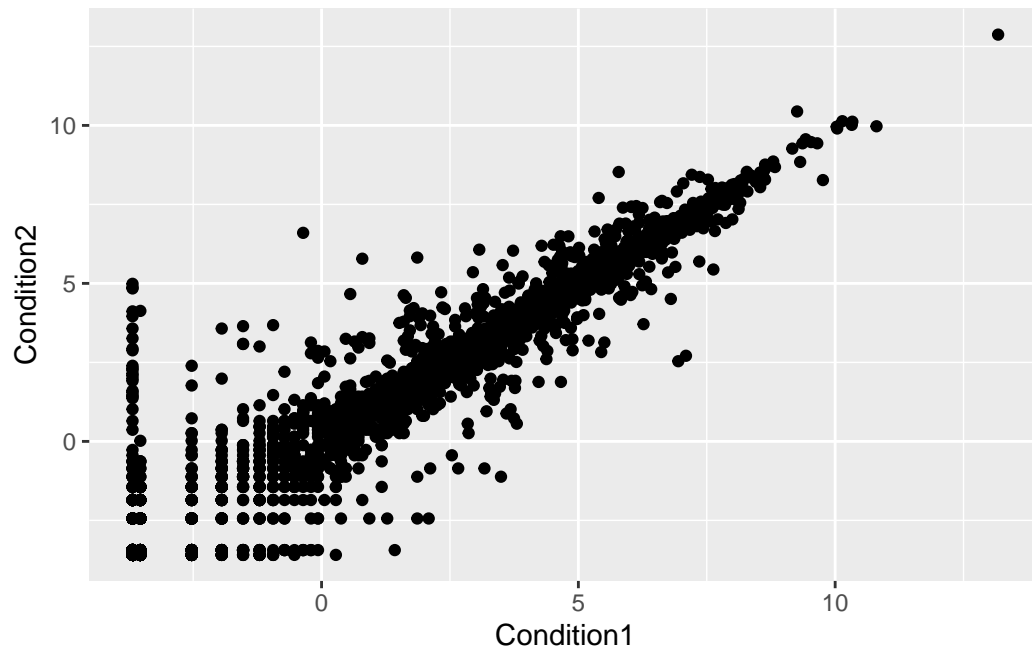
```
table(genes$State)
```

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

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

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

First basic scatter plot of the dataset:

```
ggplot(genes) +
    aes(x=Condition1, y=Condition2) +
    geom_point()
```
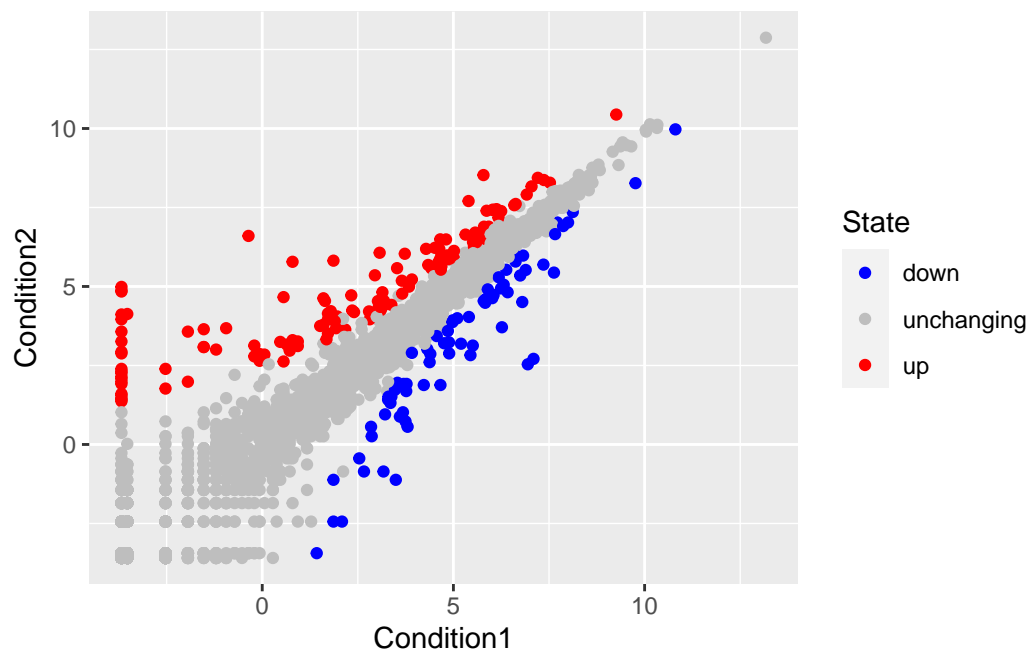
There is extra information in this dataset, namely the `State` column, which tells us whether the difference in expression values between conditions is statistically significant. Let's map this column to point color:

```
p <- ggplot(genes) +
    aes(x=Condition1, y=Condition2, col=State) +
    geom_point()
p
```
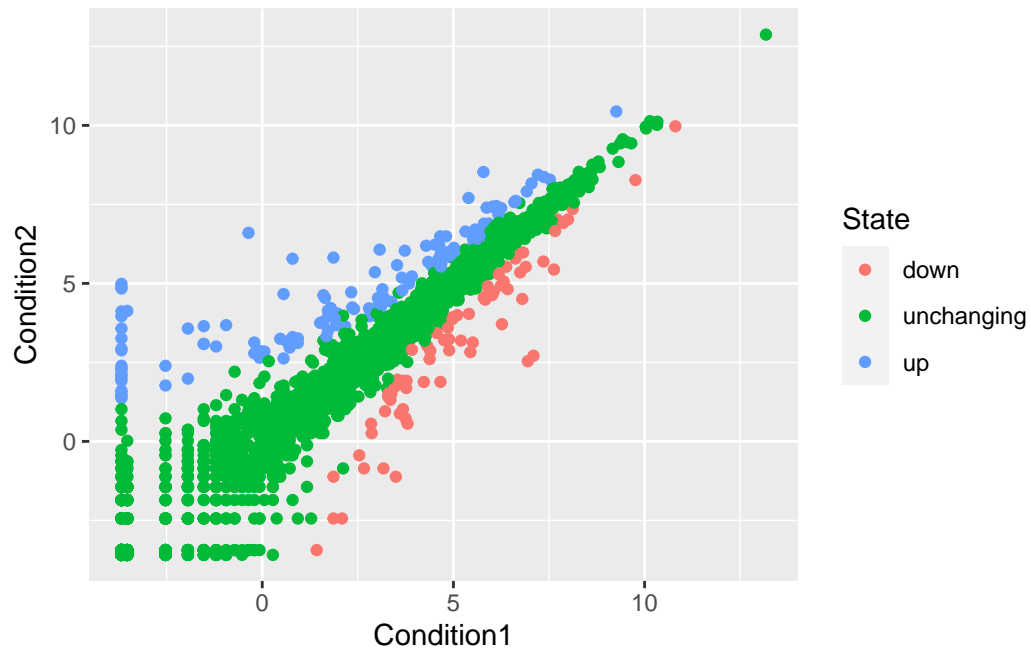
```
p + scale_colour_manual( values=c("blue","gray","red") )
```

```
p + scale_colour_manual(values=c("blue","gray","red")) +
    labs(title="Gene Expression Changes Upon Drug Treatment",
        x="Control (no drug) ",
        y="Drug Treatment")
```


Gene Expresion Changes Upon Drug Treatment