

# Class 05 Data Visualization

Ethan Harding (PID A15468670)

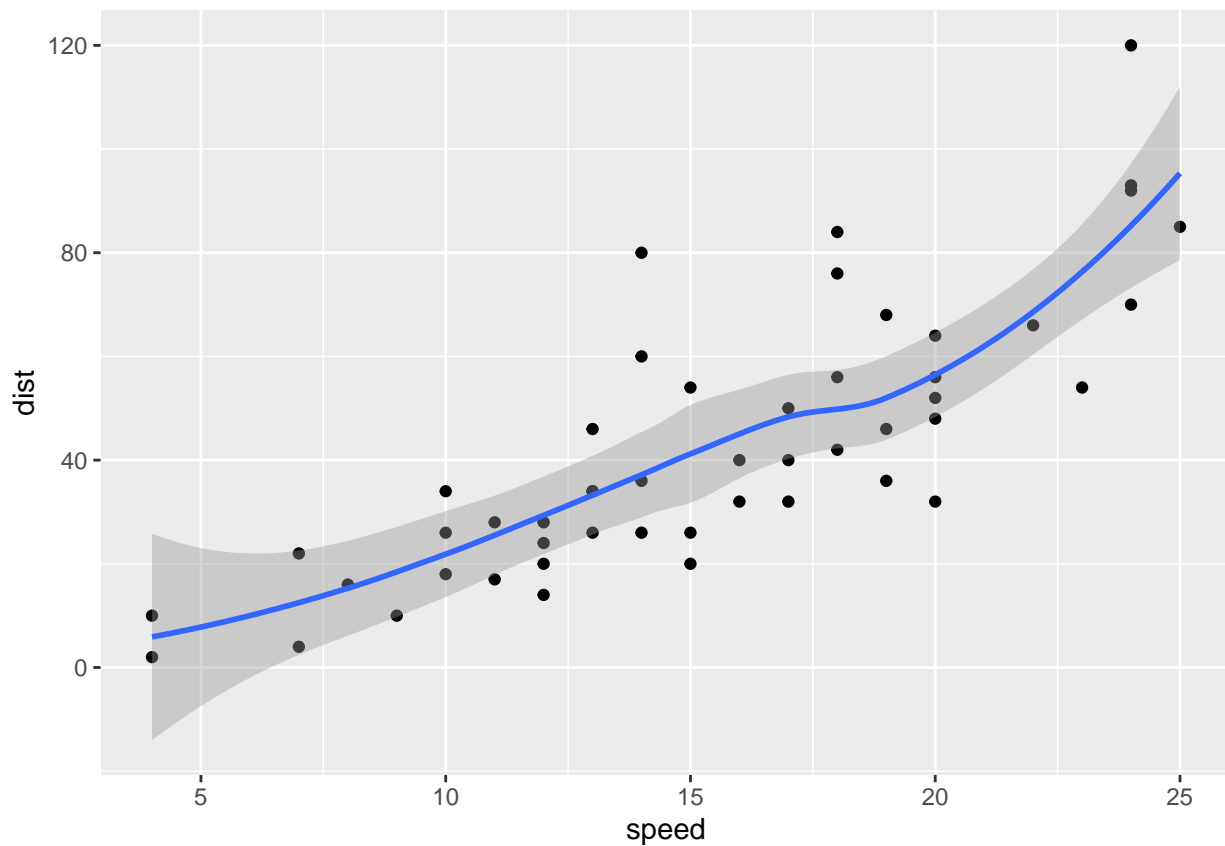
2021-10-12

```
#Lets start with a scatterplot  
#install ggplot by typing "install.packages("ggplot2")" into the R console  
#Before we use ggplot, we need to load it with "library(ggplot2)"  
#Speed = miles/hour and Dist = stopping distance in feet
```

```
library(ggplot2)
```

```
#Every ggplot has a data + aes + geoms  
ggplot(cars) + aes(x=speed, y=dist) + geom_point() +  
  geom_smooth()
```

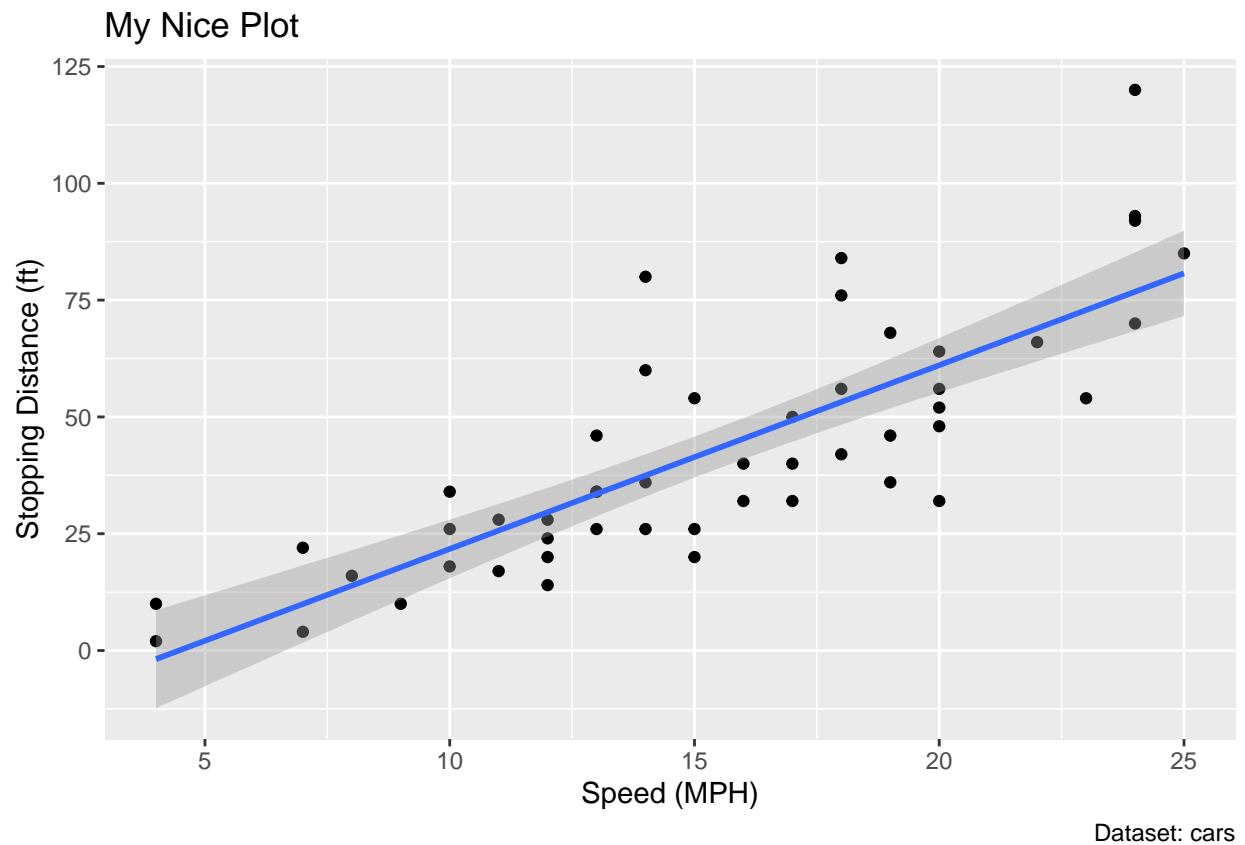
```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



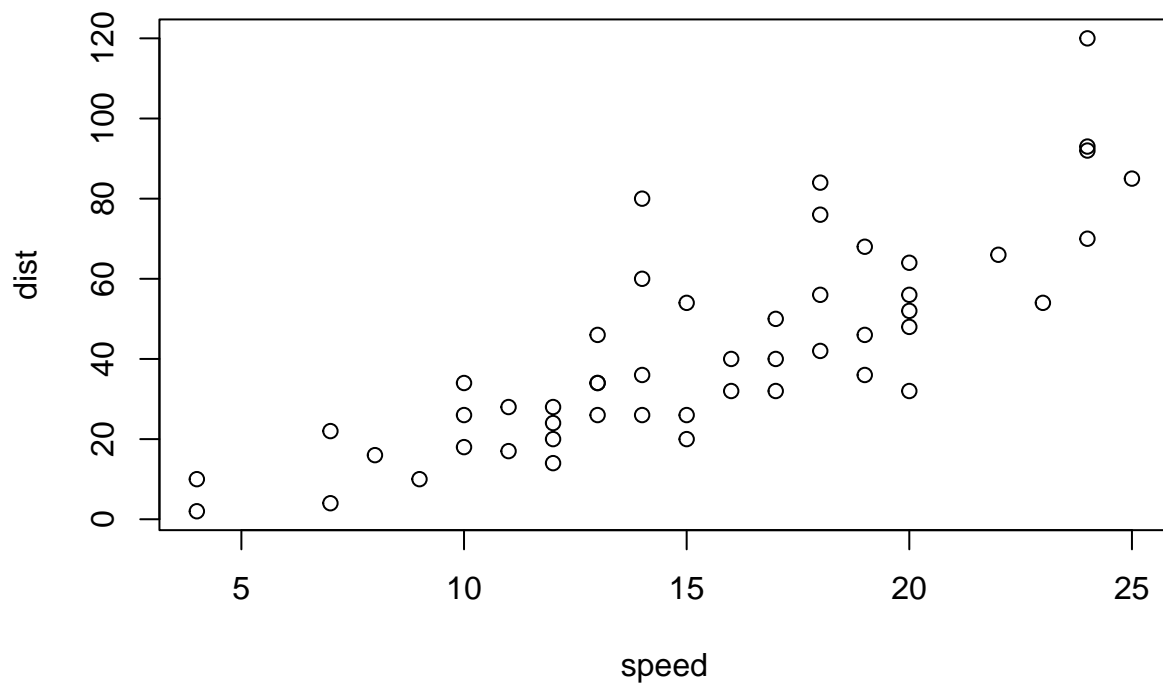
```
#Change to a linear model
p <- ggplot(cars) + aes(x=speed, y=dist) + geom_point() +
  geom_smooth(method="lm")

p + labs(title="My Nice Plot",
          x="Speed (MPH)", y="Stopping Distance (ft)", caption="Dataset: cars")

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



```
#Base graphics is shorter
plot(cars)
```



```
#Let's try a more complicated dataset of gene expression
#First read the dataset
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
```

```
#use nrow() to find how many rows there are
nrow(genes)
```

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

```
#colnames() to find out column names and ncol() to find how many columns there are
colnames(genes)
```

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

```
ncol(genes)
```

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

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

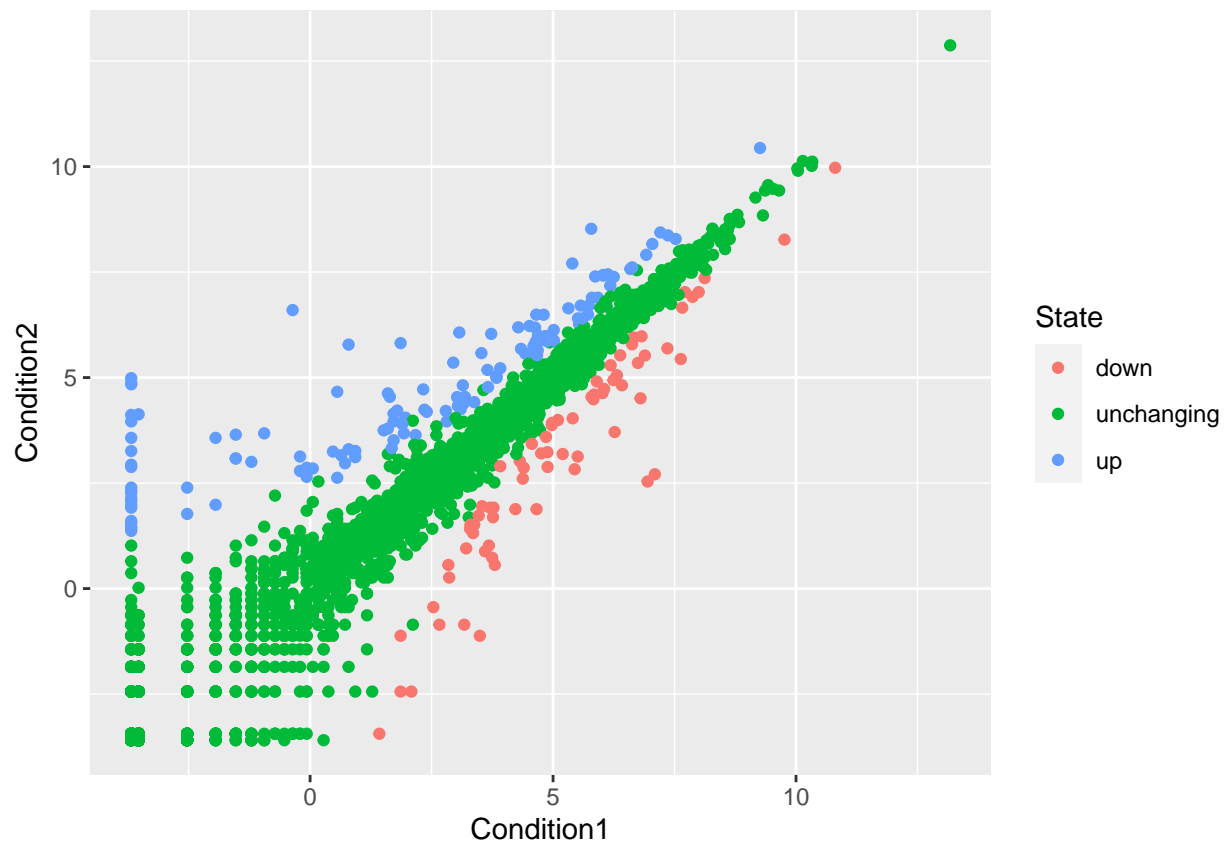
```
##  
##      down  unchanged      up  
##       72      4997     127
```

```
#What % of genes are up/down regulated  
prec <- table( genes$state ) / nrow(genes) * 100
```

```
#use round() to round the numbers to the amount of sig figs that we want  
#we want to find the total fraction of upregulated genes to 2 sig figs  
round( table(genes$State) / nrow(genes) * 100, 2)
```

```
##  
##      down  unchanged      up  
##     1.39     96.17     2.44
```

```
#Make a ggplot of the gene expression data  
g <- ggplot(genes) + aes(x=Condition1, y=Condition2,col=State) + geom_point()  
g
```



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

