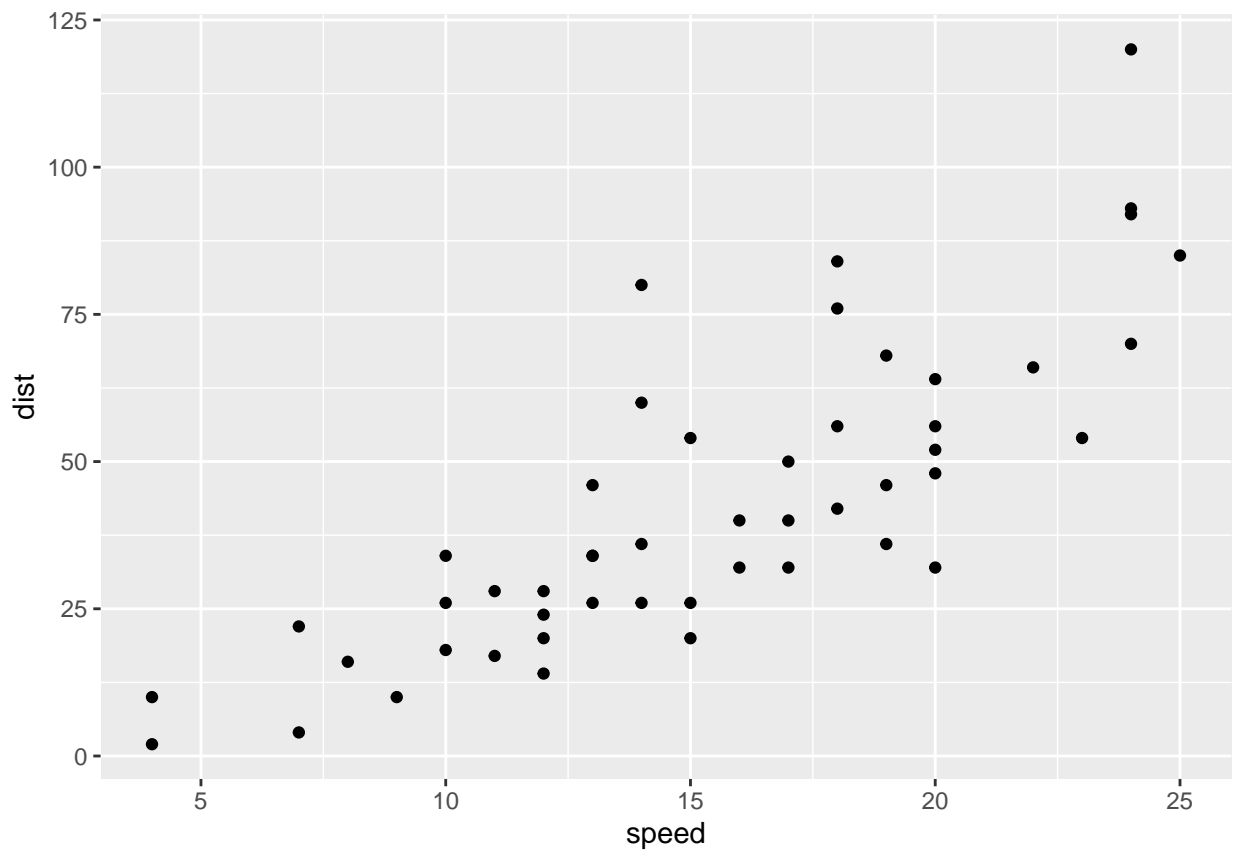# Class 05 Data Visualization

Bangyan Hu (PID: A15540189)

Oct 12, 2021

```r
#Lets start with a scatterplot
#Before we can use it we need to load it up!
# > install.packages("ggplot2")
# > library("ggplot2")
library(ggplot2)

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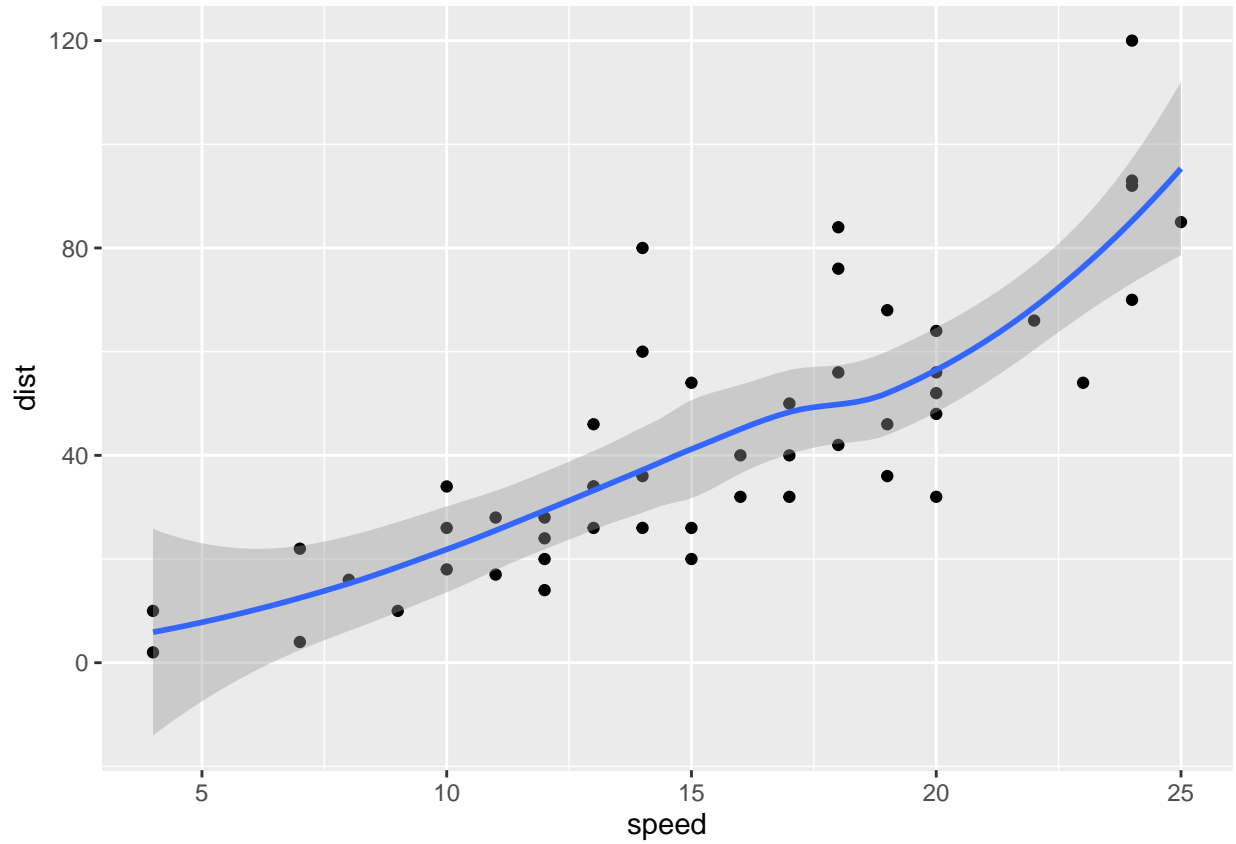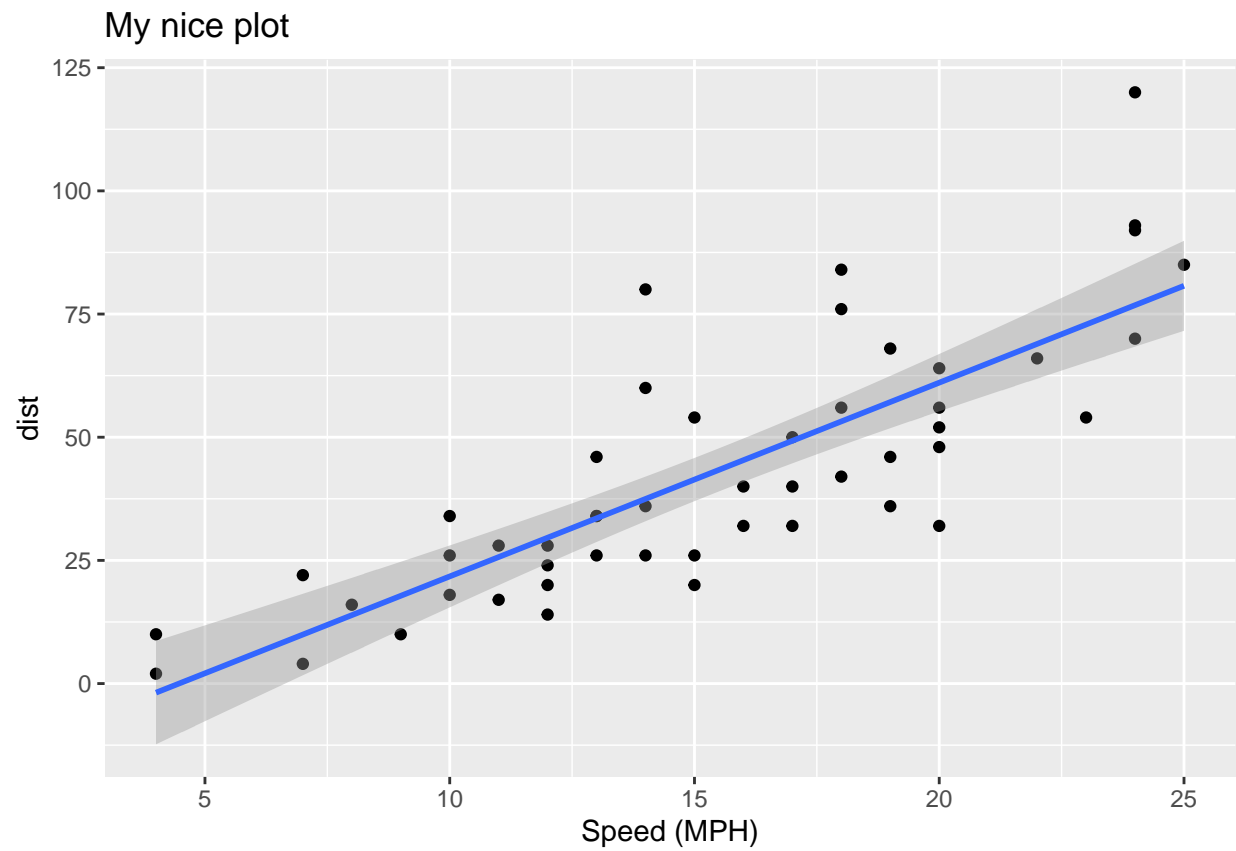


```r
ggplot(data=cars) +
  aes(x=speed, y=dist) +
```

```
  geom_point() +
  geom_smooth()
```

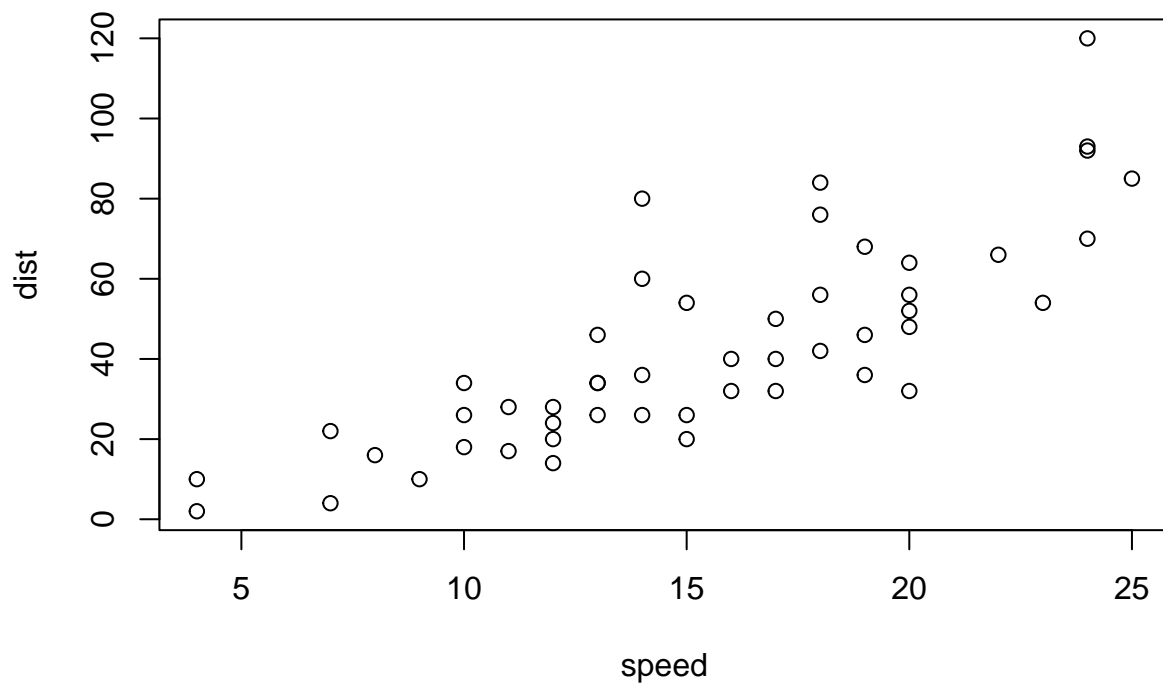## `geom_smooth()` using method = 'loess' and formula 'y ~ x'


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

p + labs(title="My nice plot",
         x="Speed (MPH)")
```

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

My nice plot


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

```
#Lets try a more complicated data set 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
```

```
#Q. How many genes
nrow(genes)
```

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

```
#Q. How many columns
colnames(genes)
```

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

```
ncol(genes)
```

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

```
#Q. How to access State col
table( genes$State )
```

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

```
#Q. what % are up/down (2 significant figures)
prec <- table( genes$State ) / nrow(genes) * 100
round( prec, 2 )
```

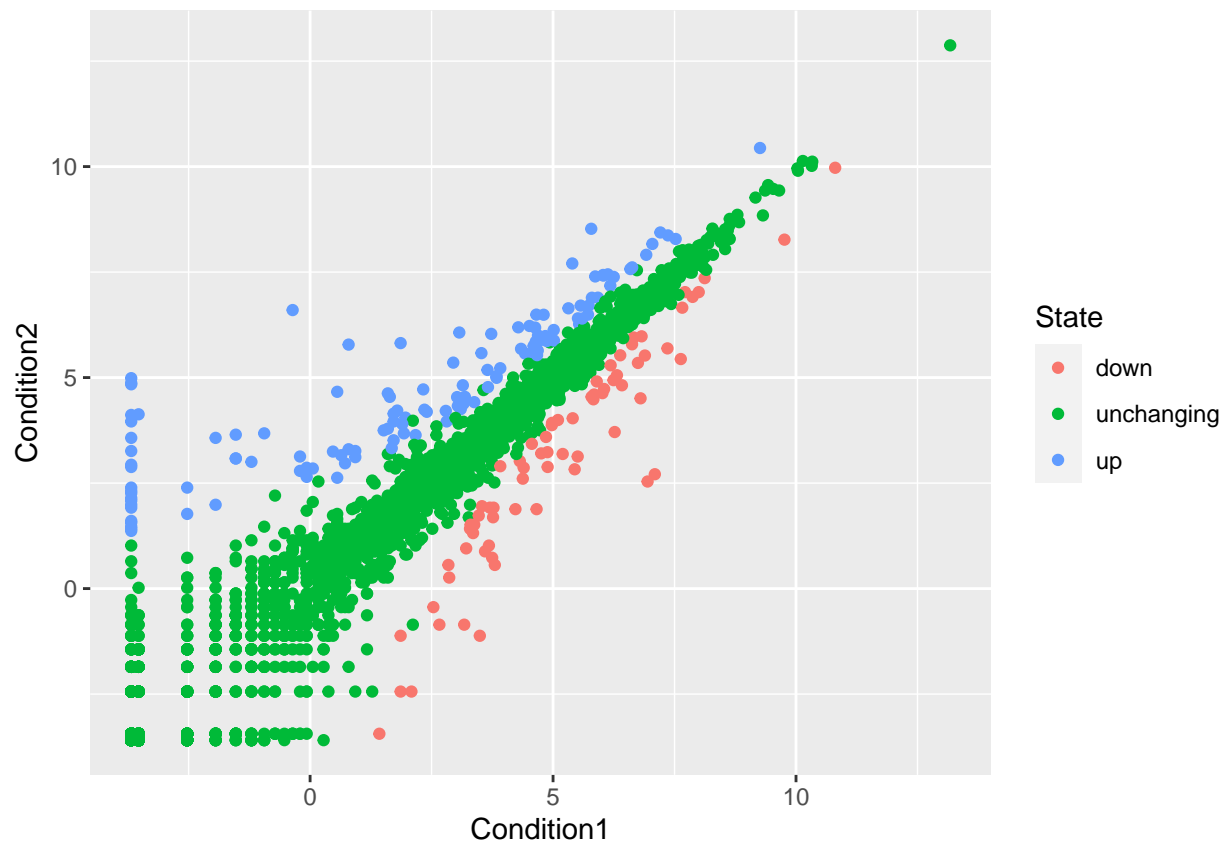```
##
down unchanging   up
1.39   96.17   2.44
```

```
# Time to plot
ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point()
```

```
# Polish the plot
Pg <- ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point()

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

Gene Expression Changes Upon Drug Treatment