Class 05 Data Visualization

Tasnia Sharia PID: A15931128

2021-10-12
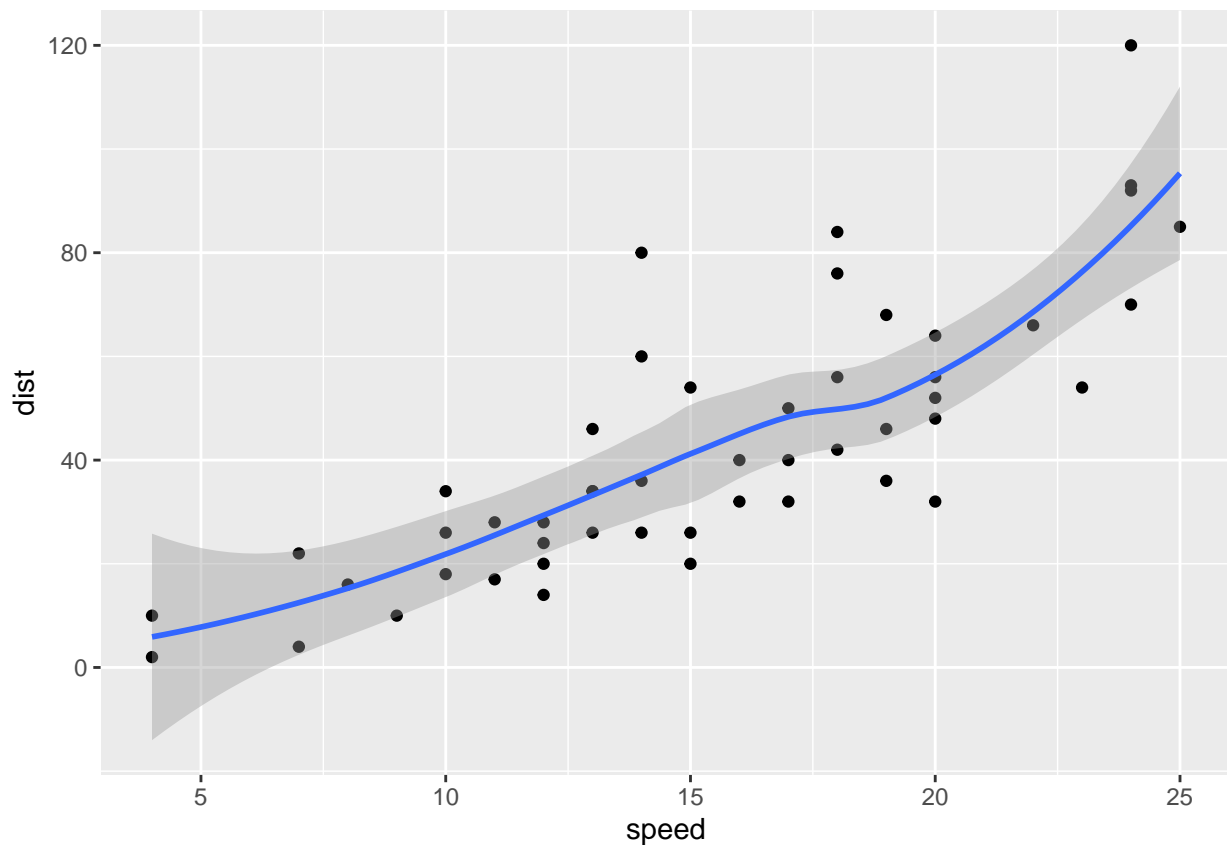
```r
# Class 05 Data Visualization

# Scatter plot
# Before we use the data, needs to be loaded

library(ggplot2)

# Every ggplot has a date + aes + geoms
# rather than geom_line, used geom_smooth to show trend line
ggplot(data = cars) +
  aes(x = speed, y = dist) +
  geom_point() + geom_smooth()
```

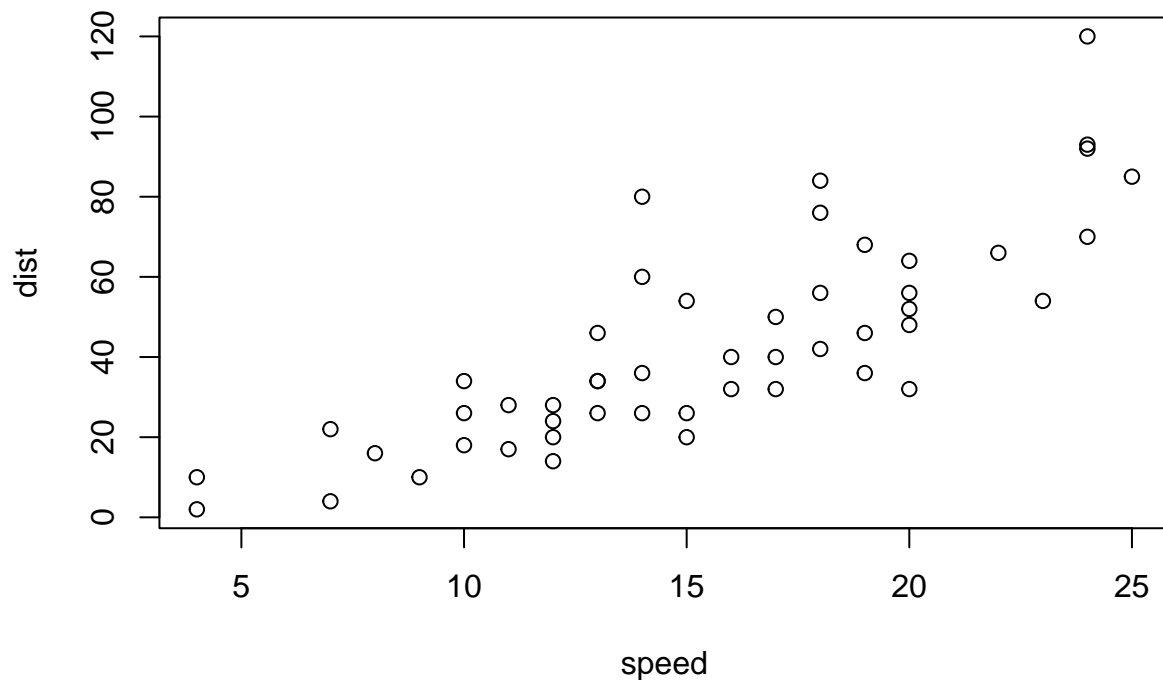## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'

```r
# change to linear model
p <- ggplot(data = cars) +
  aes(x = speed, y = dist) +
  geom_point() + geom_smooth(method="lm", se=FALSE) +
  labs(title="Speed and Stopping Distance of Cars",
       x="Speed (MPH)",
       y="Stopping distance (ft)",
       caption="Dataset: 'cars'")

# use theme_bw for more simple black and white look

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



```r
# Trying more complicated data set of genes expressed
# First read the data set
# Need to run each line from the script!

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
```

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

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

```r
# Q. How to access State column?
table(genes$State)
```

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

```r
# Q. What are % up/down?
prec <- table(genes$State) / nrow(genes)
round(prec * 100, 2) # (round(data, sig figs))
```

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

```r
# Q. names of columns?
colnames(genes)
```

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

```r
#Q. # of columns?
ncol(genes)
```

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

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