Class 5: Data Viz with ggplot

Emily Chen (A16925878)

Table of contents

Background

There are many graphic systems available in R. These include "base" R and tones and tones of added packages like **ggplot2**.
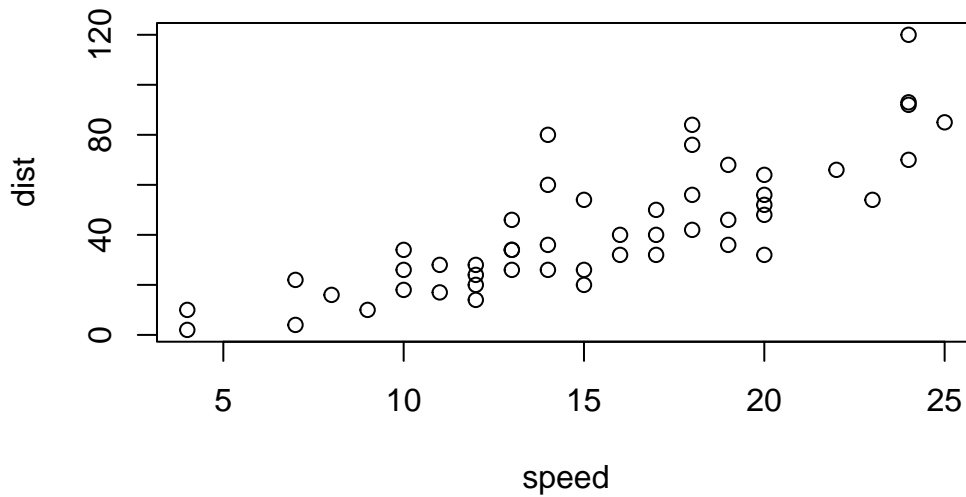
Let's compare "base" and **ggplot2** briefly. We can use some example data that is built-in with R called `cars`:

```
head(cars)
```

```
  speed dist
1     4    2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
```

In base R I can just call `plot()`

```
plot(cars)
```


How can we do this with **ggplot2**

First, we need to install the package. We do this `install.packages("ggplot2")`. I only need to do this once, and then it will be available on my computer from then on.

> Key Points: I only install packages in the R console, not writing quarto docs or R scripts

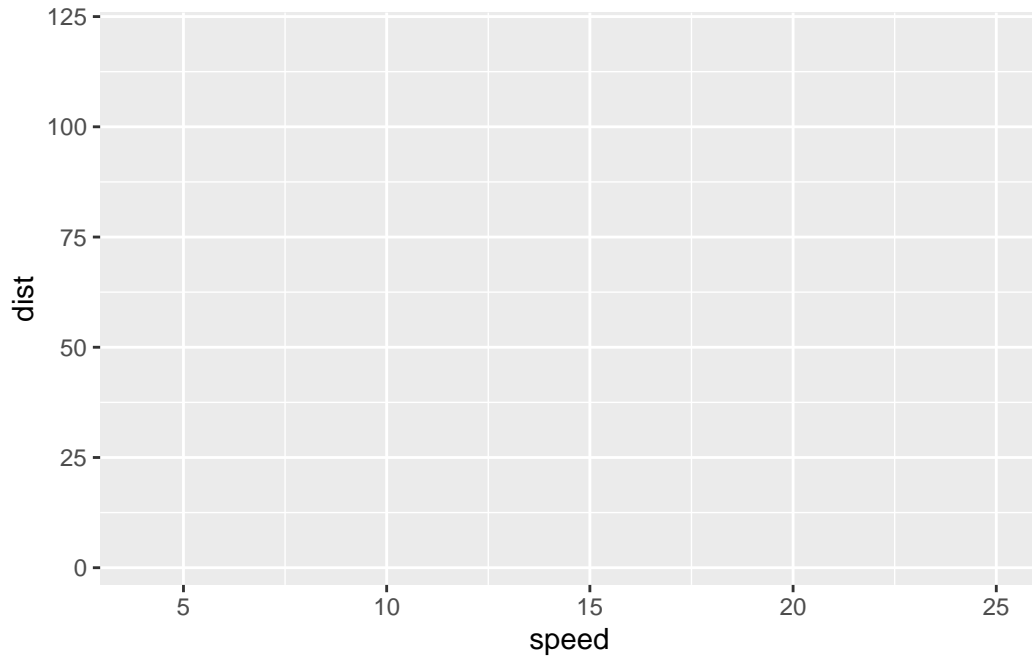Before I use any add-on packages, I must load ut up with a call to `library()`

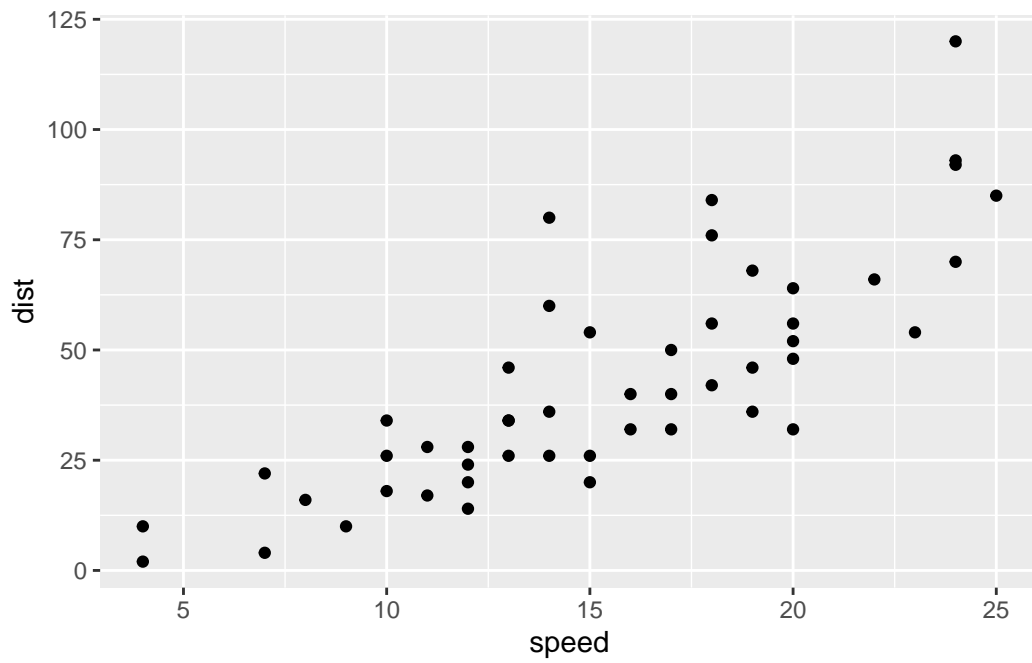```
library(ggplot2)
ggplot(cars)
```

Every ggplot has at least 3 things:

- the **data** (in our case `cars`)

- the **aes**thetics (how the data map to the plot)

- the **geom**s that determine how the plot is drawn (lines, points, columns, etc.)

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

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

For "simple" plots, ggplots is much more verbose than base R, but the defaults are nicer, and for complicated plots, it becomes much more efficient adn structured.
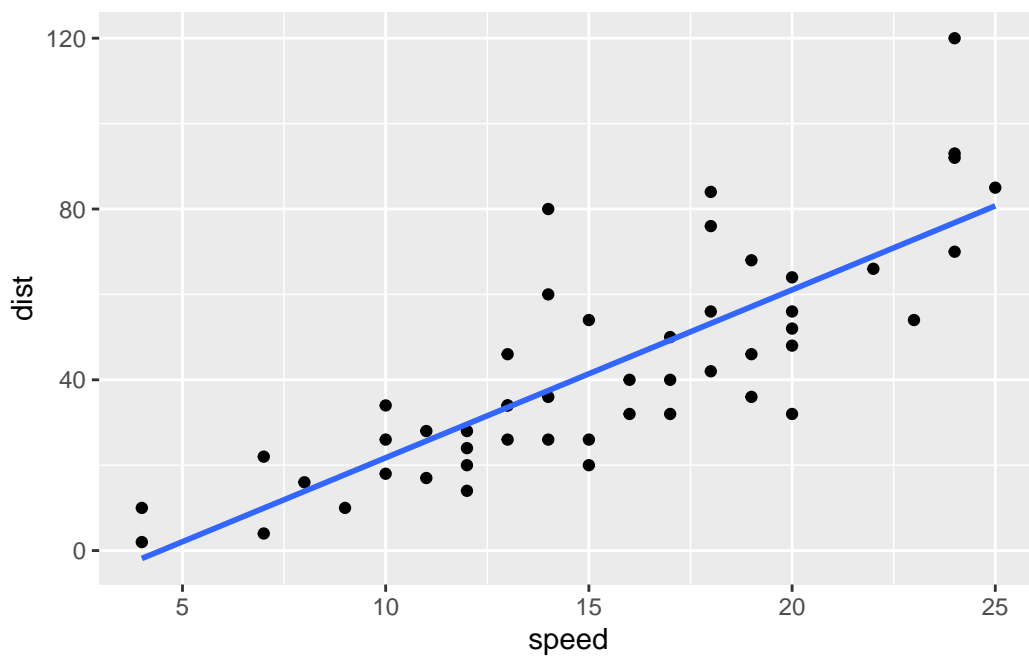
> Add a line to show relationship pf speed to stopping distance (i.e. add another "layer")

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

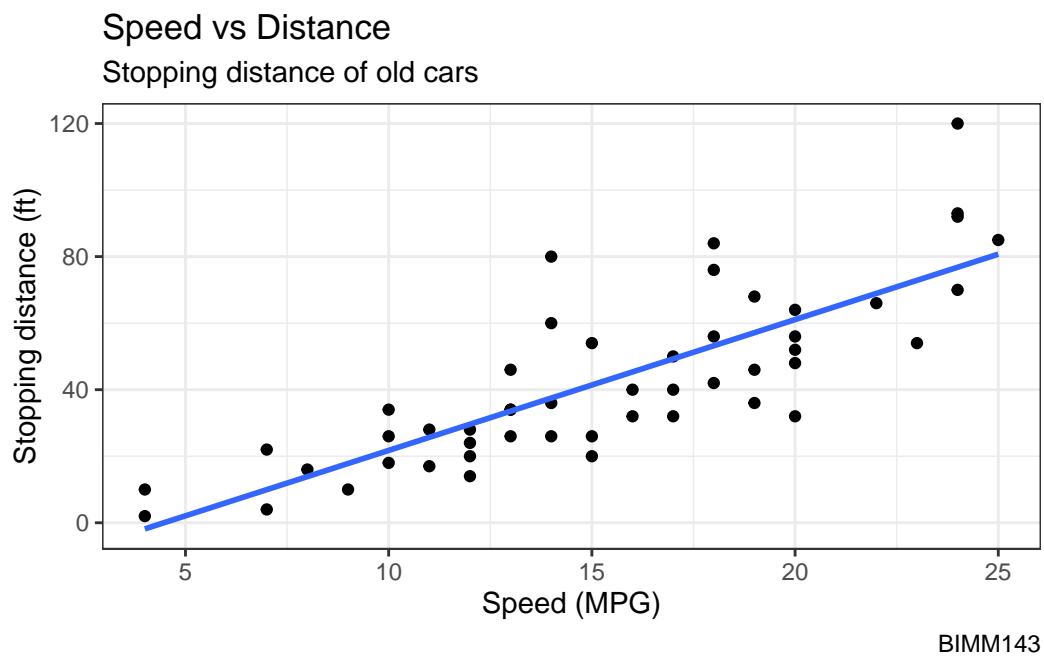I can always save any ggplot object (i.e. plot) and then use it laterfor adding more layers

```r
p
```

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


Q. Add a title and subtitle to the plot

```
p + labs(
  title="Speed vs Distance",
  subtitle="Stopping distance of old cars",
  captions="BIMM143",
  x= "Speed (MPG)",
  y= "Stopping distance (ft)")+
theme_bw()
```

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

Using different aes and geoms

Gene expression plot

Read input data

```
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 are on the data set?

```r
nrow(genes)
```

```
[1] 5196
```

```r
ncol(genes)
```

```
[1] 4
```

Q. What are the column names

```r
colnames(genes)
```

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

Q. How many "up" and "down" regulated genes are there ?

```r
head(genes$State)
```

```
[1] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
[6] "unchanging"
```
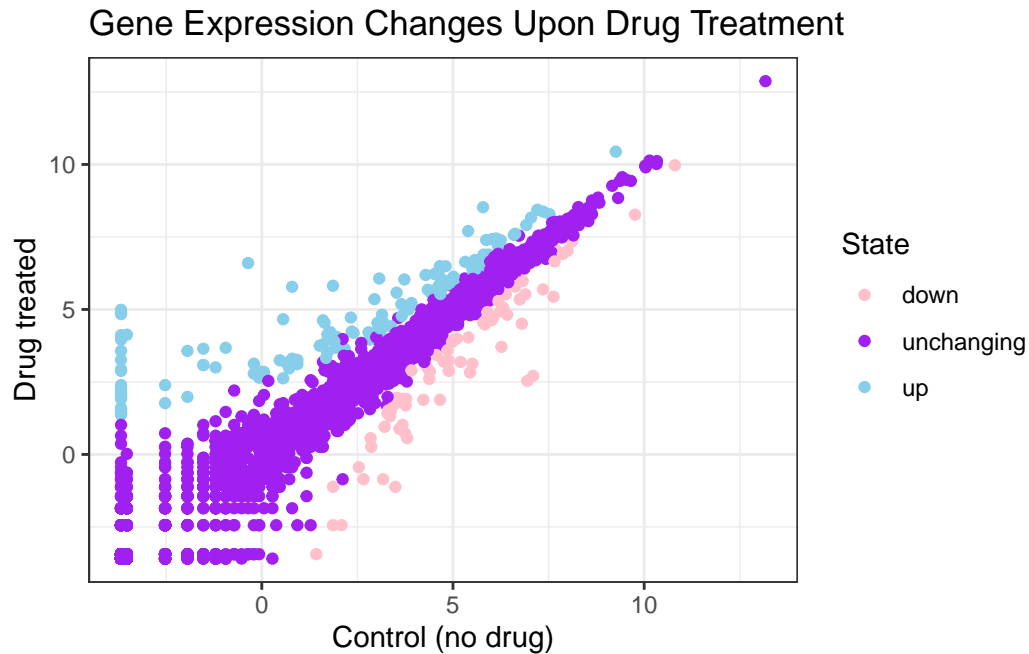
```r
table(genes$State)
```

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

Custom color plot

Q. Make a first plot of this data

```
ggplot(genes)+
  aes(x=Condition1, y=Condition2, col=State)+
  scale_color_manual(values= c("pink", "purple", "skyblue"))+
  geom_point()+
  labs(title= "Gene Expression Changes Upon Drug Treatment", x= "Control (no drug)", y="Drug
  theme_bw()
```


Gene Expression Changes Upon Drug Treatment

Using Differnt geoms

Lets plot some aspect of the in-built `mtcars` dataset

```
head(mtcars)
```

```
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Q. Scatter plot of `mpg` vs `disp`

```
p1<- ggplot(mtcars)+
  aes(x=mpg, y=disp)+
  geom_point()+
  labs(title= "Miles per Gallon vs Displacement", x= "Miles per Gallon", y="Distance")+
  theme_bw()
```

Q. boxplot of `gear` vs `disp`

```
p2<-ggplot(mtcars)+
  aes(x=gear, y=disp, group=gear)+
  geom_boxplot()+
  labs(title= "Gear vs Displacement", x= "gear", y="Distance")+
  theme_bw()
```

Q. barplot of `carb`

```
p3<-ggplot(mtcars)+
  aes(carb)+
  geom_bar()+
  theme_bw()
```
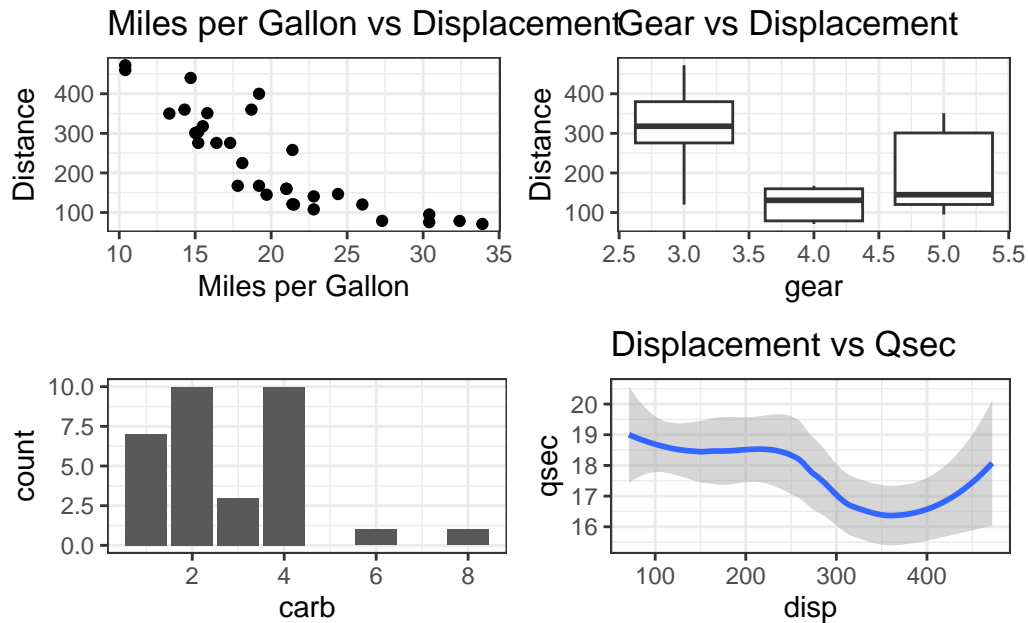
Smooth of `disp` vs `qsec`

```
p4<-ggplot(mtcars)+
  aes(x=disp, y=qsec)+
  geom_smooth()+
  labs(title= "Displacement vs Qsec")+
  theme_bw()
```

I want to combine all these plots into one figure with multiple pannels

We can use the **patchwork** package to do this.

```
library(patchwork)

((p1| p2) / (p3 |p4))
```

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

```r
ggsave(filename= "myplot.png", width=10,height=10)
```

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

File location online

url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"
gapminder <- read.delim(url)

```r
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv
gapminder <- read.delim(url)
```

And a wee peak

Q. How many countries are in this dataset?

```r
length(table(gapminder$country))
```

```
[1] 142
```

Q. Plot gdpPercap vs lifeExp coloe by continent

```
ggplot(gapminder)+
  aes(gdpPercap,lifeExp, color=continent)+
  geom_point(alpha=0.3) +
  facet_wrap(~continent)+
  theme_bw()
```