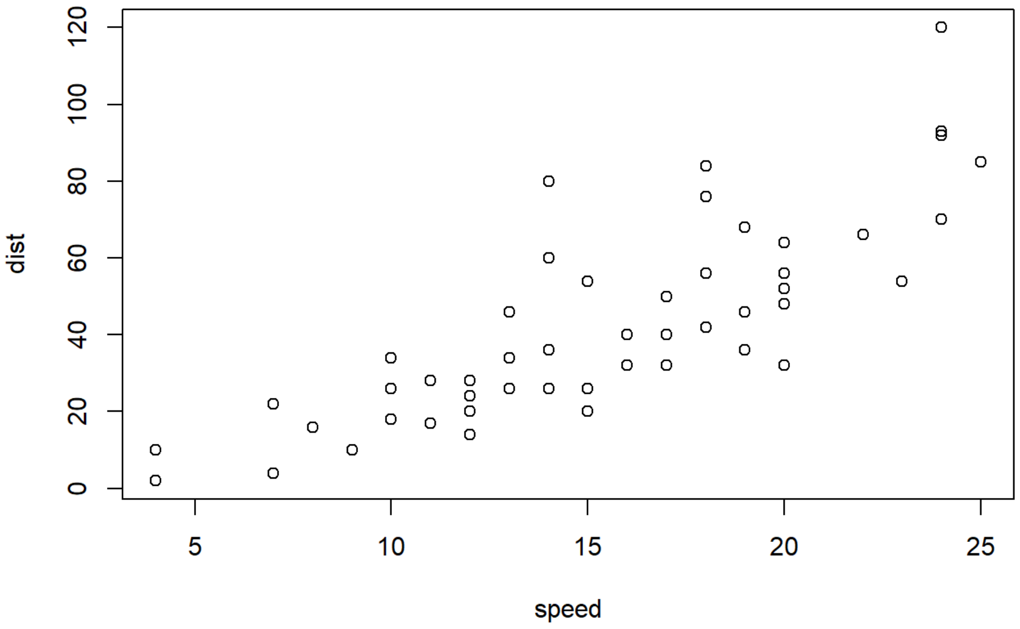
Class 5: Data Visualization with ggplot

AUTHOR
Justin R (PID: A16307501)

Today we will have our first play play with the **ggplot** package - one of the most popular graphics packages on the plante

There are many plotting systems in R. These include so-called *"base"* plotting/ graphics

```
plot(cars)
```



Base plot is gerenally rather short code and somewhat dull plots - but it is always there for you and is fast for big datasets

If I want to use **ggplot2** it takes some work

```
#ggplot(cars)
```

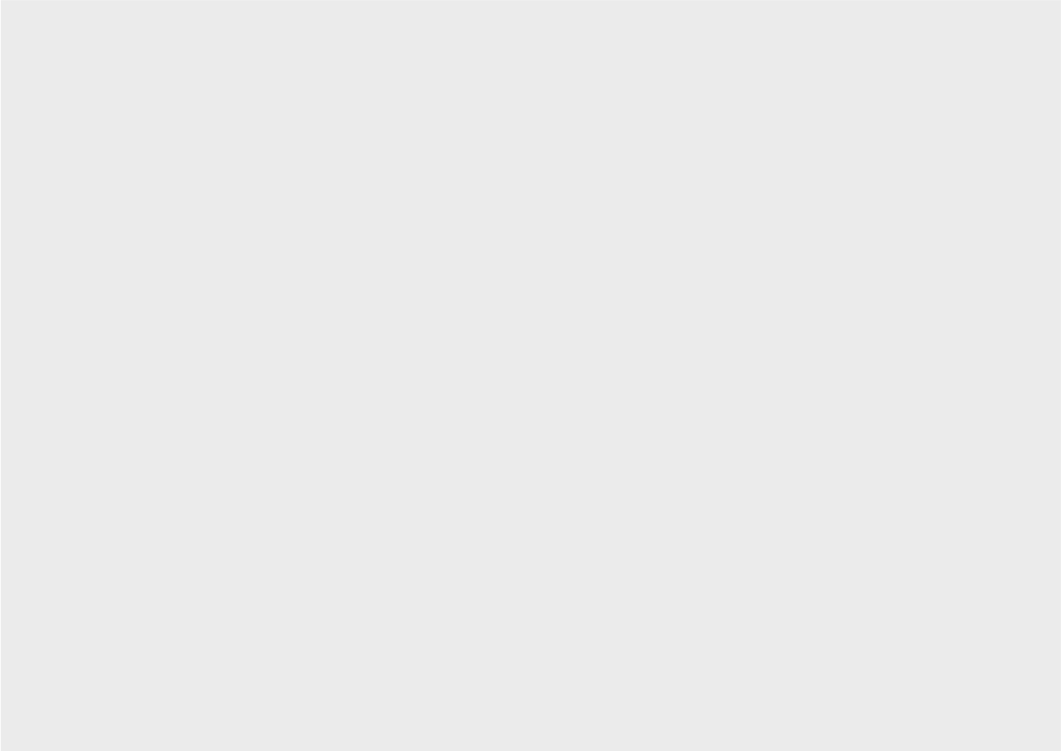
I need to install the package first to my computer. To do this I can use the function `install.packages("ggplot2")`

Every time I want to use a package I need to load it up with a `library()` call.

```
library(ggplot2)
```

Now finally use ggplot

```
ggplot(cars)
```



Every ggplot has at least 3 things:

data the data frame withe the data you want to plot, **aes** the aesthetic mapping, **geom** the way we visualize the dataset

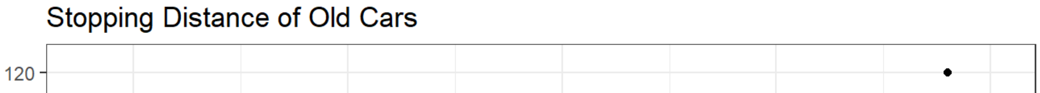
```
head(cars)
```

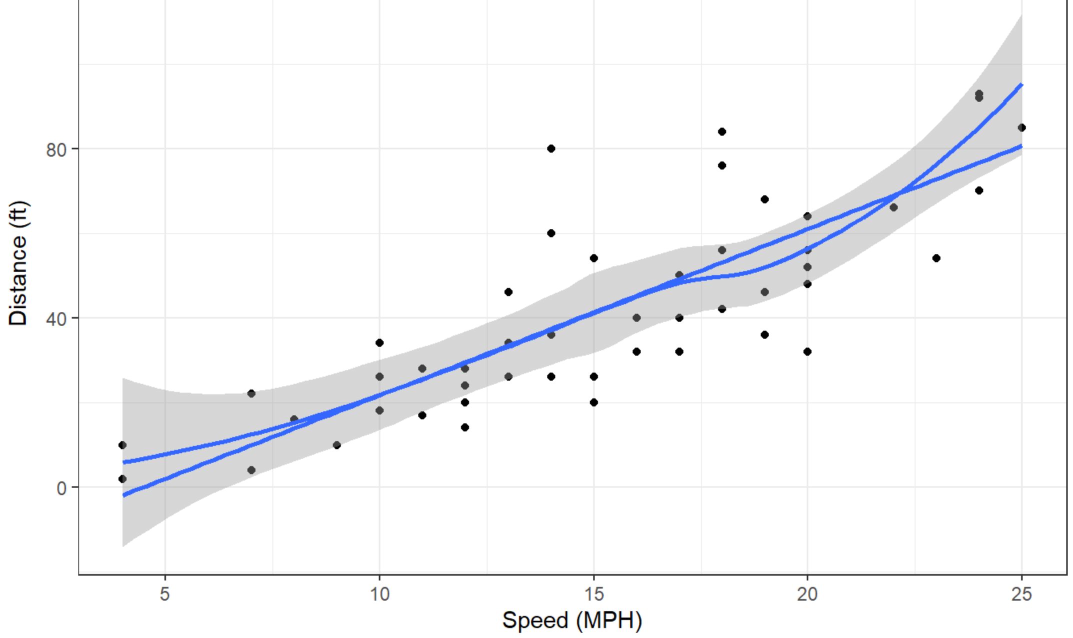
```
  speed dist
1     4    2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
```

```
bp <- ggplot(cars)+
  aes(x=speed, y=dist)+
  geom_point()+
  geom_smooth()
```

```
bp + geom_smooth(se=FALSE, method="lm")+
  labs(title="Stopping Distance of Old Cars",
       x="Speed (MPH)",
       y="Distance (ft)",
       caption="From the cars dataset") +
  theme_bw()
```

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





From the cars dataset

A more complicated scatterplot

```
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_express
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
```

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

```

      down unchanging         up 
        72       4997        127 
```

```
round(sum(genes$State == "up") / nrow(genes) * 100 , 2)
```

```
[1] 2.44
```

```
n.gene <- nrow(genes)
n.up <- sum(genes$State == "up")
```

```
up.precent <- n.up/n.gene * 100
round(up.precent, 2)
```

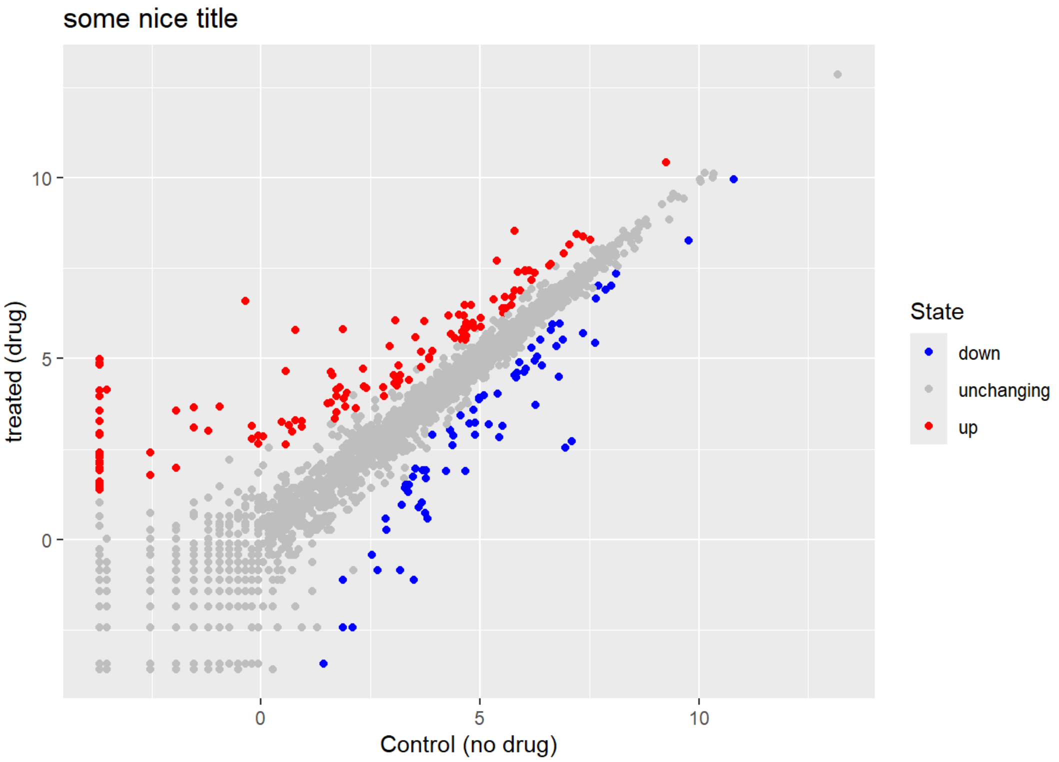
```
[1] 2.44
```

```
head(genes, 2)
```

```
   Gene Condition1 Condition2      State
1 A4GNT  -3.680861  -3.440135 unchanging
2  AAAS   4.547958   4.386413 unchanging
```

```
p <- ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State) +
  geom_point()
```

```
p + labs(title="some nice title",
         x = "Control (no drug)",
         y = "treated (drug)") +
  scale_colour_manual(values=c("blue", "gray", "red"))
```



exploring the gapminder dataset

Here we will load up the gapminder dataset to get practice with different aes mappings

```
# File location online
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extc

gapminder <- read.delim(url)
```

How amny entries rows are in this dataset?

```
nrow(gapminder)
```

```
[1] 1704
```

How many columns?

```
ncol(gapminder)
```

```
[1] 6
```

```
head(gapminder)
```

```
      country continent year lifeExp      pop gdpPercap
1 Afghanistan      Asia 1952  28.801  8425333  779.4453
2 Afghanistan      Asia 1957  30.332  9240934  820.8530
3 Afghanistan      Asia 1962  31.997 10267083  853.1007
4 Afghanistan      Asia 1967  34.020 11537966  836.1971
5 Afghanistan      Asia 1972  36.088 13079460  739.9811
6 Afghanistan      Asia 1977  38.438 14880372  786.1134
```

```
table(gapminder$year)
```

```

1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 2002 2007 
 142  142  142  142  142  142  142  142  142  142  142  142 
```

how many continents?

```
table(gapminder$continent)
```

```

  Africa Americas     Asia   Europe  Oceania 
     624      300      396      360       24 
```

```
length(unique(gapminder$continent))
```

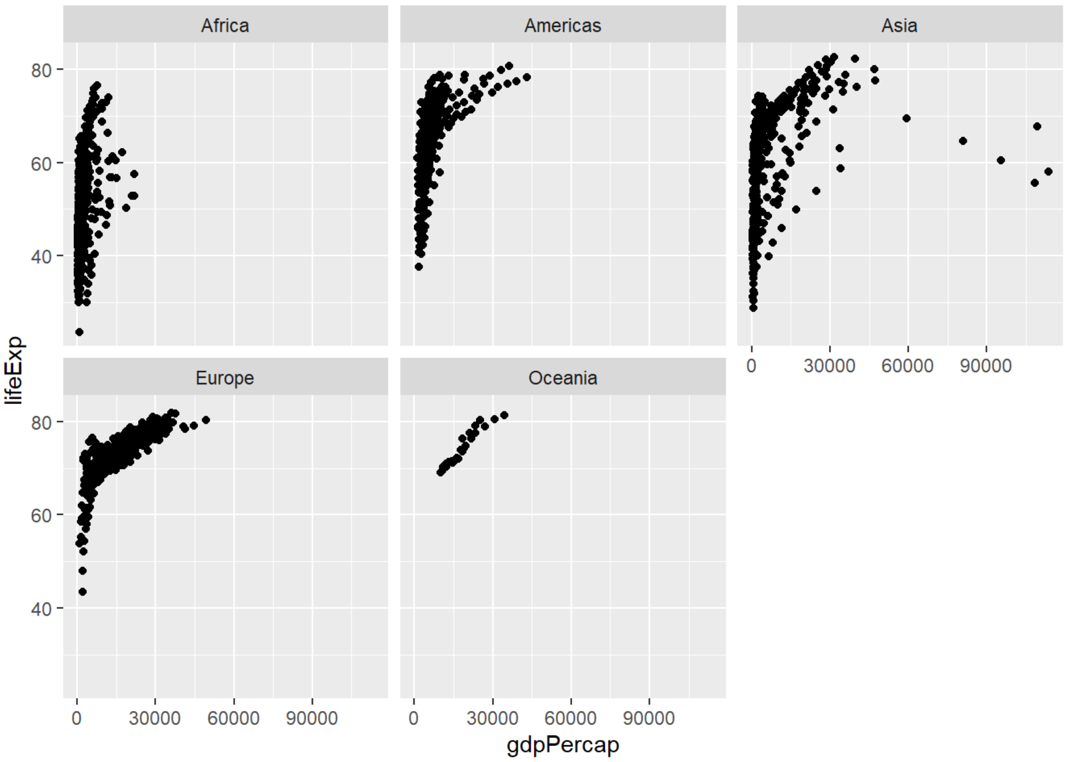
```
[1] 5
```

how many countries in this dataset?

```
length(unique(gapminder$country))
```

```
[1] 142
```

```
ggplot(gapminder)+
  aes(x=gdpPercap, y=lifeExp)+
  geom_point()+
  facet_wrap(~continent)
```

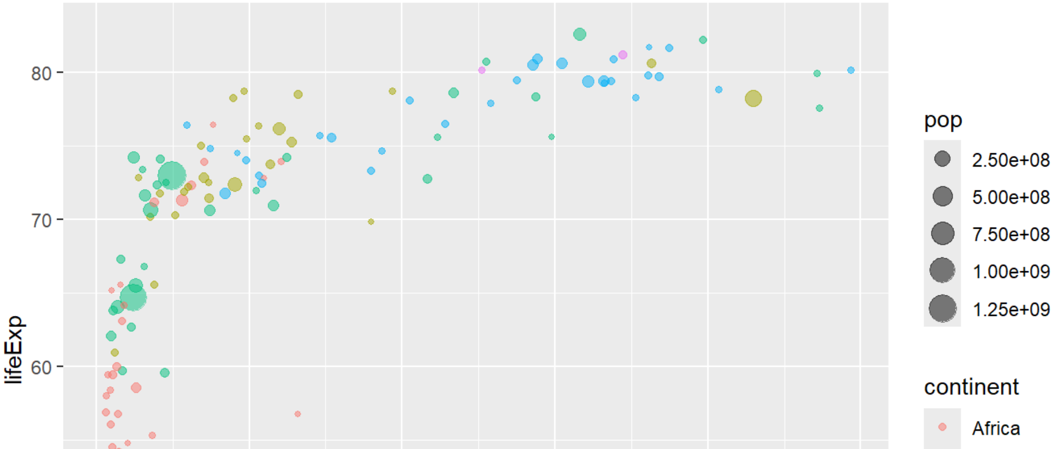


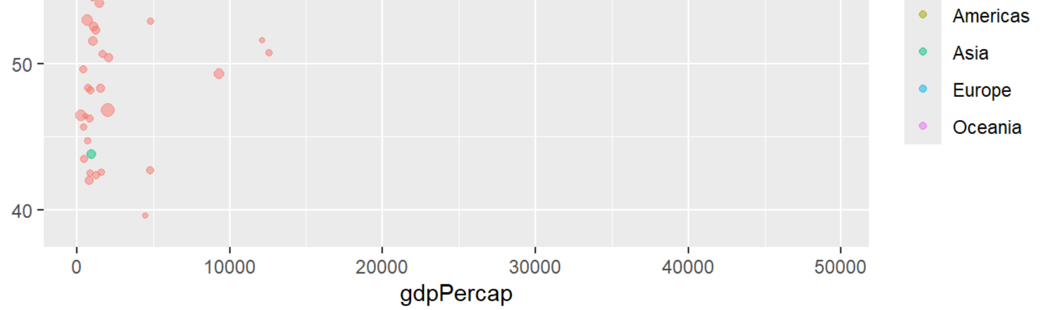
```
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
gapminder_2007 <- gapminder %>% filter(year==2007)
```

```
ggplot(gapminder_2007)+
  aes(x=gdpPercap, y=lifeExp, col=continent, size=pop)+
  geom_point(alpha=0.5)
```




50
40
0
10000
20000
30000
40000
50000
gdpPercap
Americas
Asia
Europe
Oceania