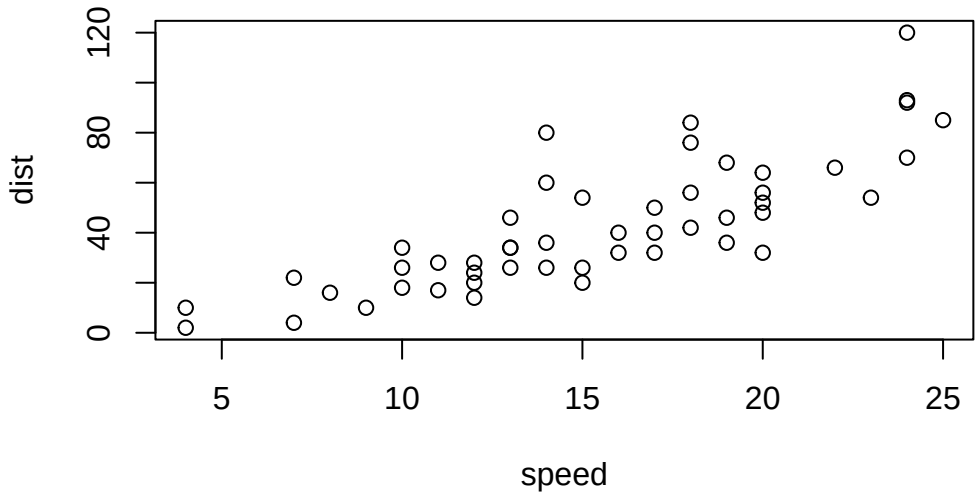
Class 5: Data Visualization with ggplot

Bernice Lozada (PID: A16297973)

Today we will have our first play with the **ggplot2** package - one of the most popular graphics packages on the planet.

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

```
plot(cars)
```



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

```
# install.packages("ggplot2")
library(ggplot2)

ggplot(cars)
```

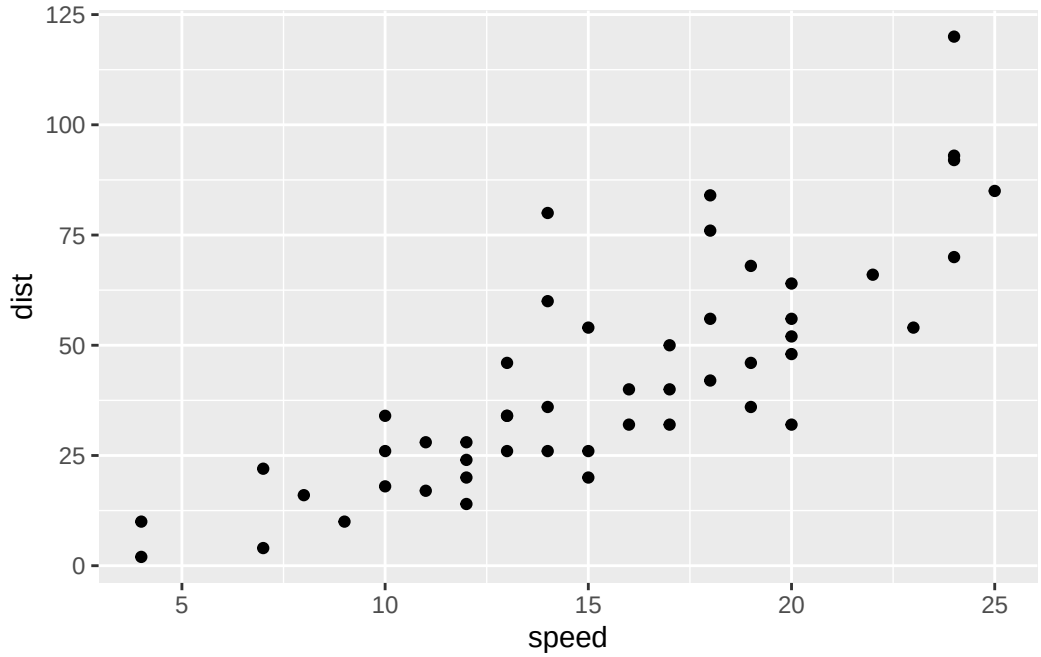

The command to install the package first using `install.packages()` in the R console to make it permanent.

To use a package, it needs to be loaded up with a `library()` call.

Every ggplot has at least three things:

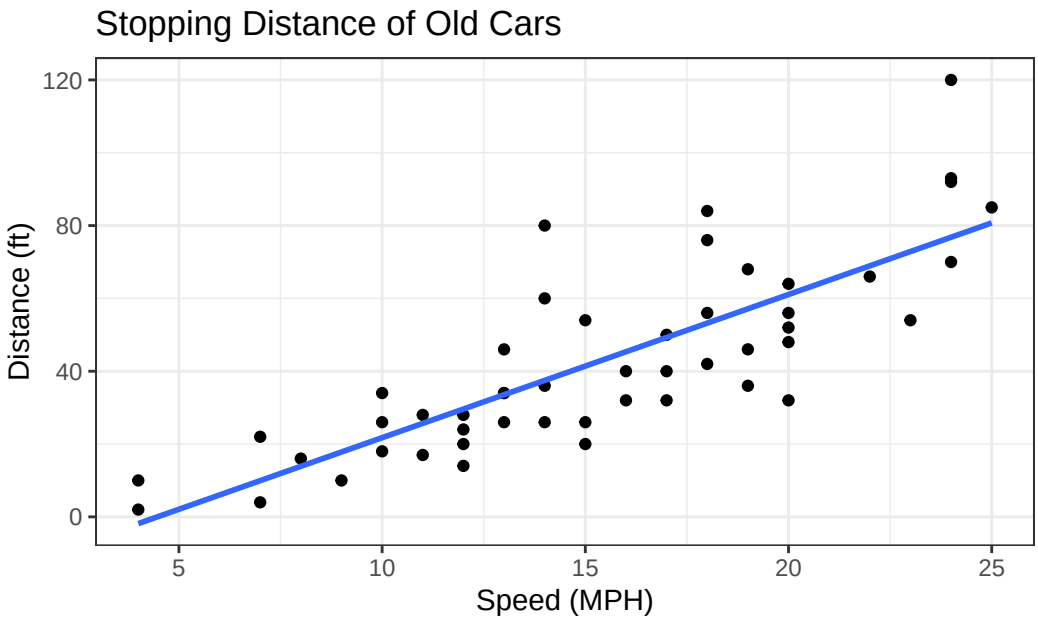
- **data** (the data.frame with the plotting data)
- **aes** (aesthetic mapping of data to plot)
- **geom** (how you want plot to look - points, lines, etc.)

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




```
bp_new <- 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()
bp_new
```

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



A more complicated scatter plot

Here we make a plot of gene expression 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
```

```
nrow(genes)
```

```
[1] 5196
```

```
colnames(genes)
```

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

```
ncol(genes)
```

```
[1] 4
```

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

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

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

```
[1] 2.44
```

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

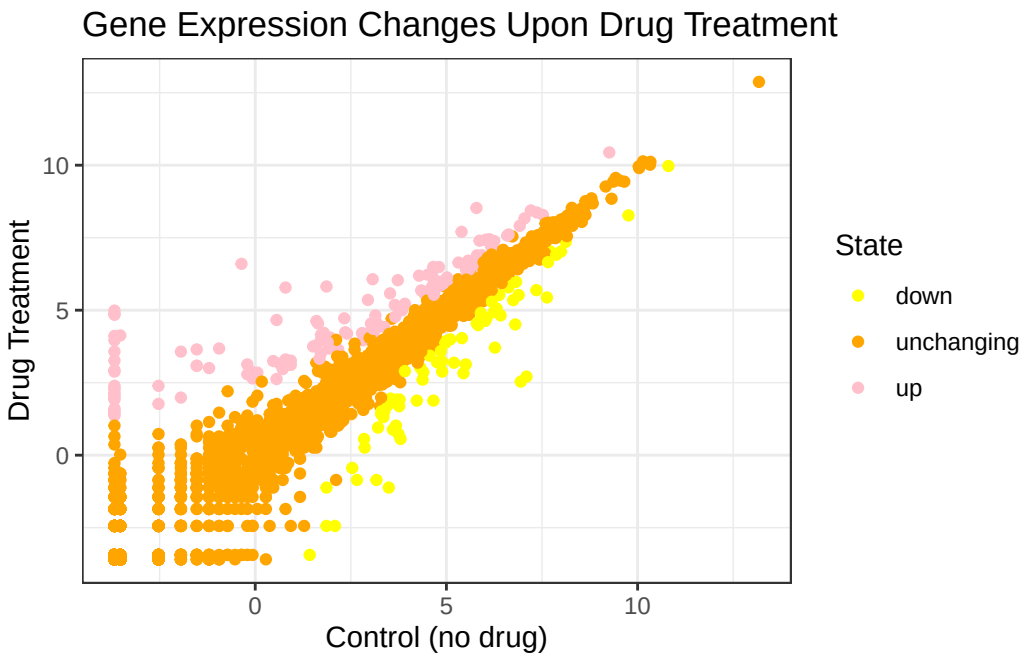
up.percent <- n.up/n.gene * 100
round(up.percent, 2)
```

```
[1] 2.44
```

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

```
# Changing Colors
p + scale_colour_manual(values = c("yellow","orange","pink")) +
  labs(x = "Control (no drug)", y = "Drug Treatment", title = "Gene Expression Changes Upo
```

```
  theme_bw()
```



Exploring the gapmider dataset

Load up the gapminder dataset for practice with different aes mappings.

Find number of countries in databasse

```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.

gapminder <- read.delim(url)

table(gapminder$continent)
```

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

```
# Can use unique() function
length(unique(gapminder$continent))
```

```
[1] 5
```

```
# number of countries
length(unique(gapminder$country))
```

```
[1] 142
```

```
library(dplyr)
```

```
Attaching package: 'dplyr'

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

    filter, lag

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

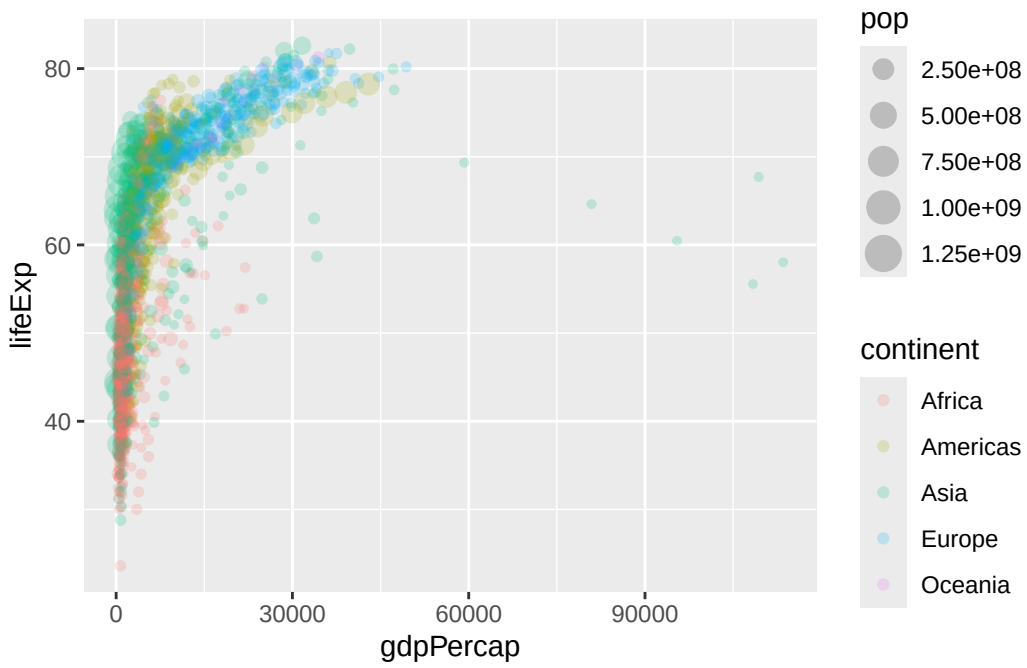
    intersect, setdiff, setequal, union
```

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

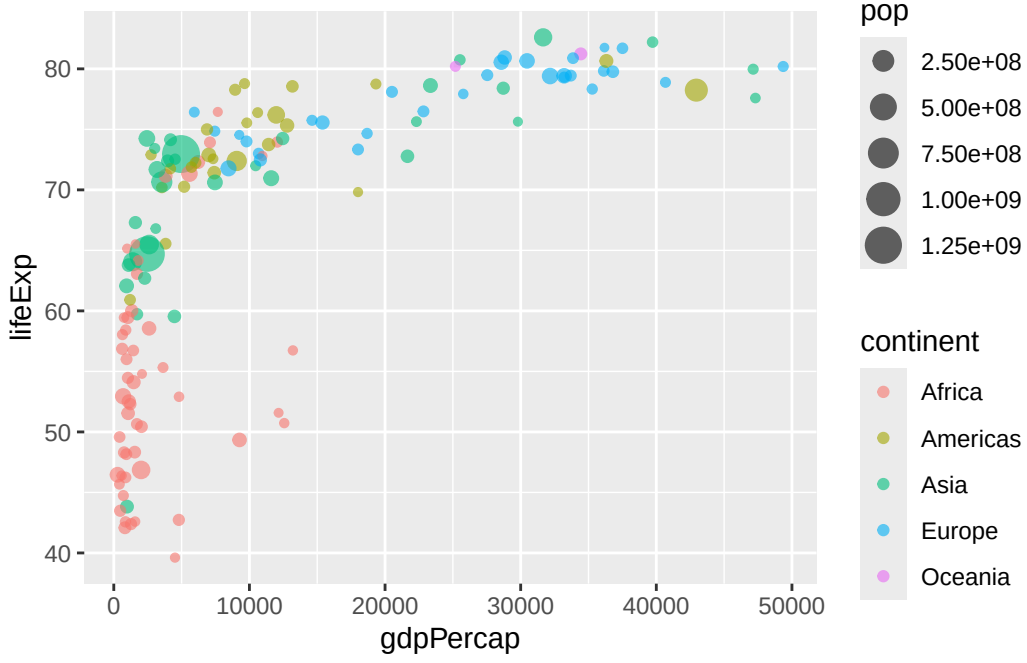
head(gapminder_2007)
```

```
      country continent year lifeExp      pop  gdpPercap
1 Afghanistan      Asia 2007  43.828 31889923   974.5803
2     Albania    Europe 2007  76.423  3600523  5937.0295
3     Algeria    Africa 2007  72.301 33333216  6223.3675
4      Angola    Africa 2007  42.731 12420476  4797.2313
5   Argentina  Americas 2007  75.320 40301927 12779.3796
6   Australia   Oceania 2007  81.235 20434176 34435.3674
```

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




```
## for 2007
ggplot(gapminder_2007) + aes(x=gdpPercap, y = lifeExp, col=continent, size = pop) + geom_p
```



With dyplr

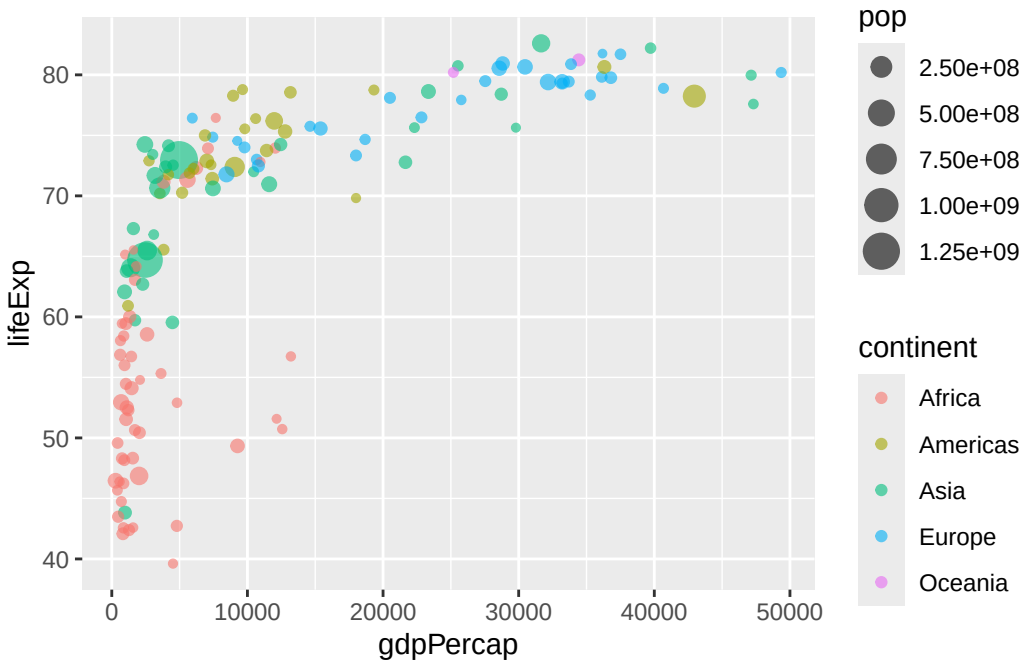
```
#install.packages("dplyr")
library(dplyr)
gapminder_2007 <- filter(gapminder, year == 2007)
head(gapminder_2007)
```

```
      country continent year lifeExp      pop  gdpPercap
1 Afghanistan      Asia 2007  43.828 31889923   974.5803
2     Albania    Europe 2007  76.423  3600523  5937.0295
3     Algeria    Africa 2007  72.301 33333216  6223.3675
4      Angola    Africa 2007  42.731 12420476  4797.2313
5   Argentina  Americas 2007  75.320 40301927 12779.3796
6   Australia   Oceania 2007  81.235 20434176 34435.3674
```

Plot of 2007 with population and continent data

```
::: {.cell}

```{.r .cell-code}
ggplot(gapminder_2007) + aes(x=gdpPercap, y = lifeExp, col=continent, size = pop) + geom_poi
```


```

:::

Facet_wrap data to compare 1957 and 2007

```
gapminder_1957 <- filter(gapminder, year == 1957 | year == 2007)
ggplot(gapminder_1957) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color=continent,
                 size = pop), alpha=0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```