CLASS 5 DATA VIZ WITH GGPLOT

KENNY DANG (PID:A18544481)

Table of contents

Background

There are lots of ways to make plots in R. These include so-called "base R" (like the `plot()`) and add on packages like **ggplot2**

Lets make the same plot with these two graphics systems. We can use the inbuilt `cars` dataset:

```
head(cars)
```

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

With "base R" we can simply:

```
plot(cars)
```

Now lets try ggplot. First I need to install the package using `install.packages("ggplot2")'`
> **N.B.** We never run an`install.packages()'` in a code chunk that otherwise we will re-install needlessly every time we render our document

every time we want to use an add-on package we need to load it up with a call to 'library()'

```
library(ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.4.3
```
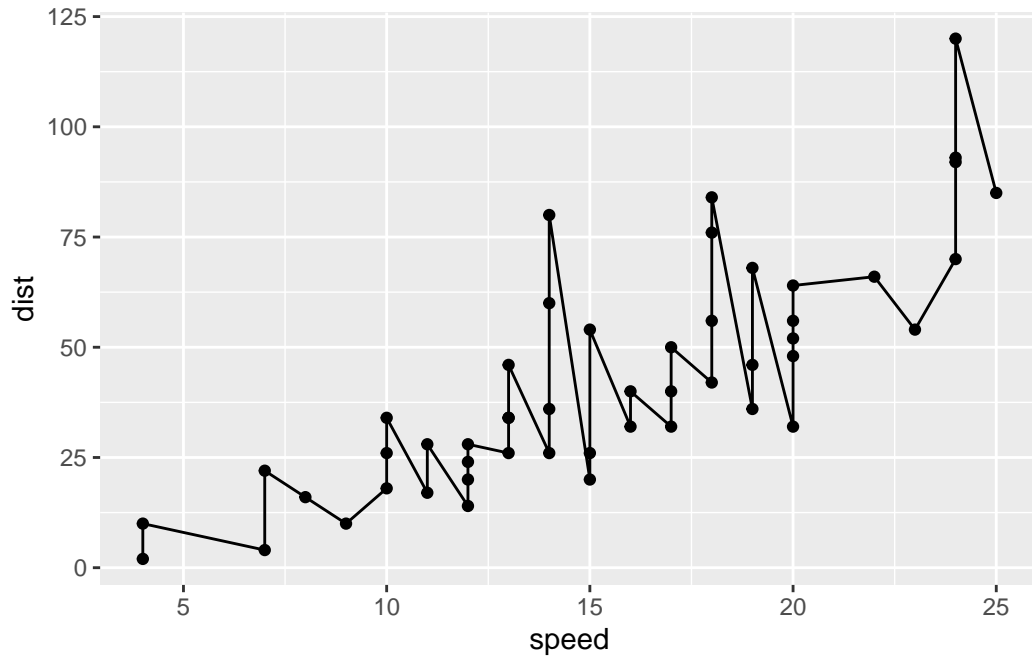
```
ggplot(cars)
```

Every ggplot needs at least three things:

1. the **data** i.e. stuff to plot as a data.frame
2. the **aes** or aesthetics that map the data to the plot
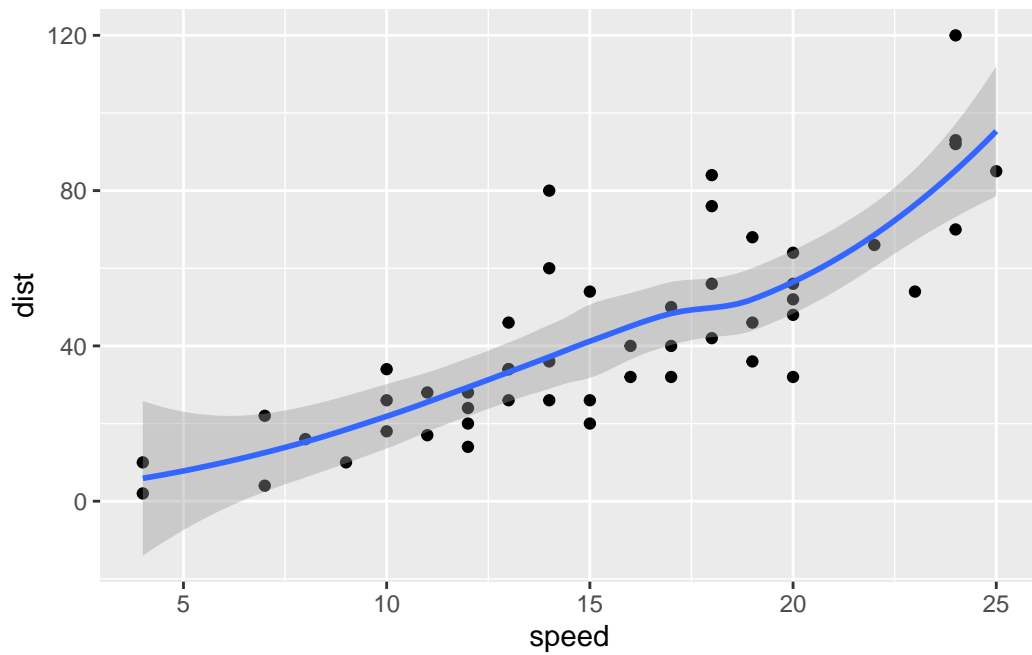3. the **geom__** or geometry i.e. the plot type such as points, lines etc.

install.packages("ggplot2")

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

```
ggplot(cars) + aes(x=speed, y=dist) + geom_point() +geom_smooth ()
```
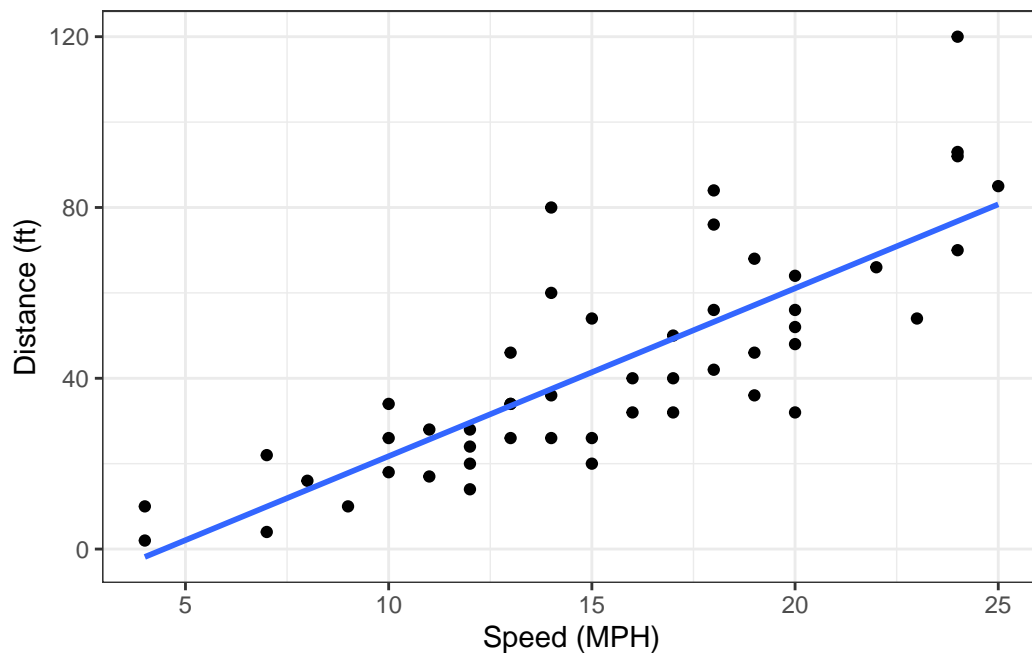
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'

4

```r
ggplot(cars) +
  aes(x=speed, y=dist) + geom_point()+
  geom_smooth(method = "lm", se=FALSE) +
  labs(x="Speed (MPH)", y ="Distance (ft)", Title = "stopping Distance of old Cars") + theme_
```

```
Ignoring unknown labels:
* Title : "stopping Distance of old Cars"
`geom_smooth()` using formula = 'y ~ x'
```


gene expression plot

read some data on the effects of GLP-1 inhibitor (drug) on gene expression values:

```r
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
```
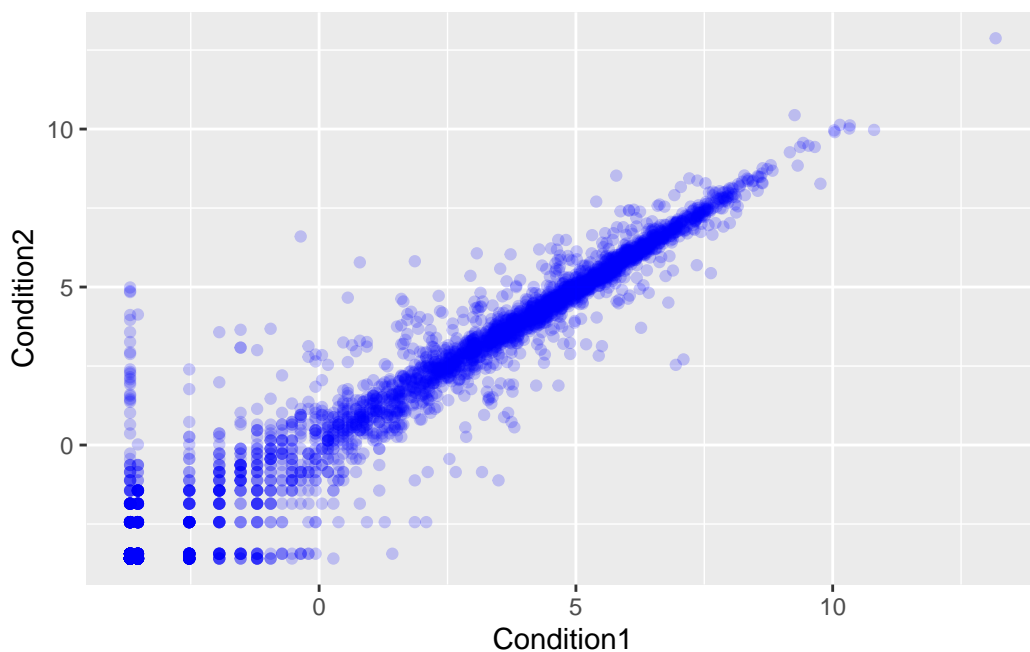
Version 1 plot - start simple by getting some ink on the page

```
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
ggplot(genes) +
  aes(x= Condition1, y=Condition2) +
  geom_point(col="blue", alpha = 0.2)
```


Let's color by State up, down or no change
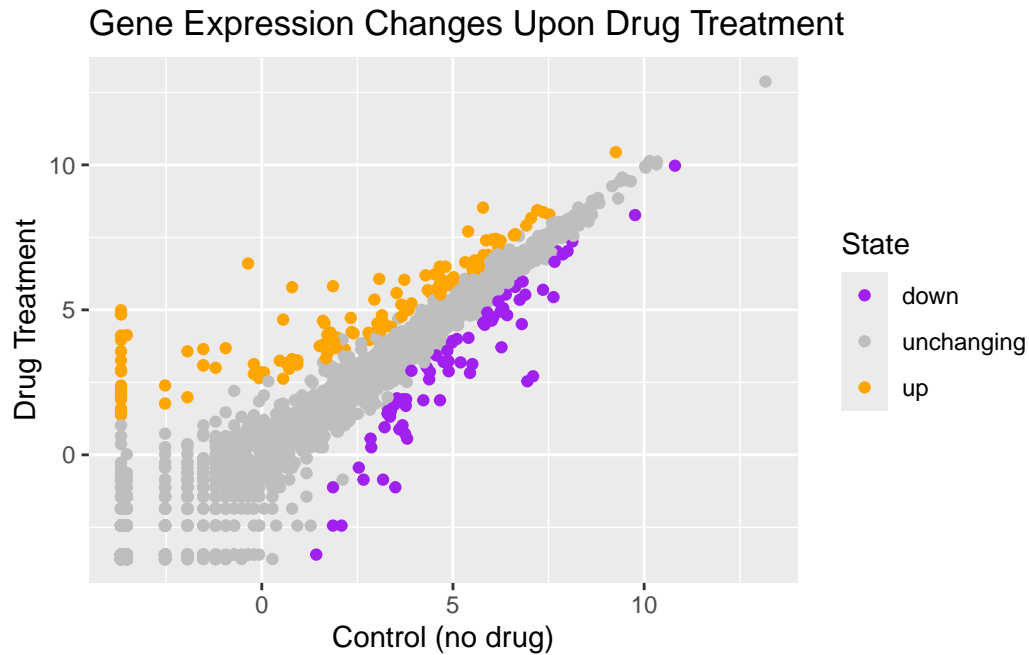
```
table(genes$State)
```

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

library(ggplot2)

6

```
url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expression.txt"
genes <- read.delim(url)
ggplot(genes) + aes(x=Condition1, y=Condition2, col=State) + geom_point() + scale_color_manua
```

Gene Expression Changes Upon Drug Treatment

going further with gapminder

here we explore the famous `gapminder`

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

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

Q. How many rows does this dataset have?

```
nrow(gapminder)
```

```
[1] 1704
```

How many different continents are in this dataset

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

```
  Africa Americas     Asia   Europe  Oceania
     624      300      396      360       24
```
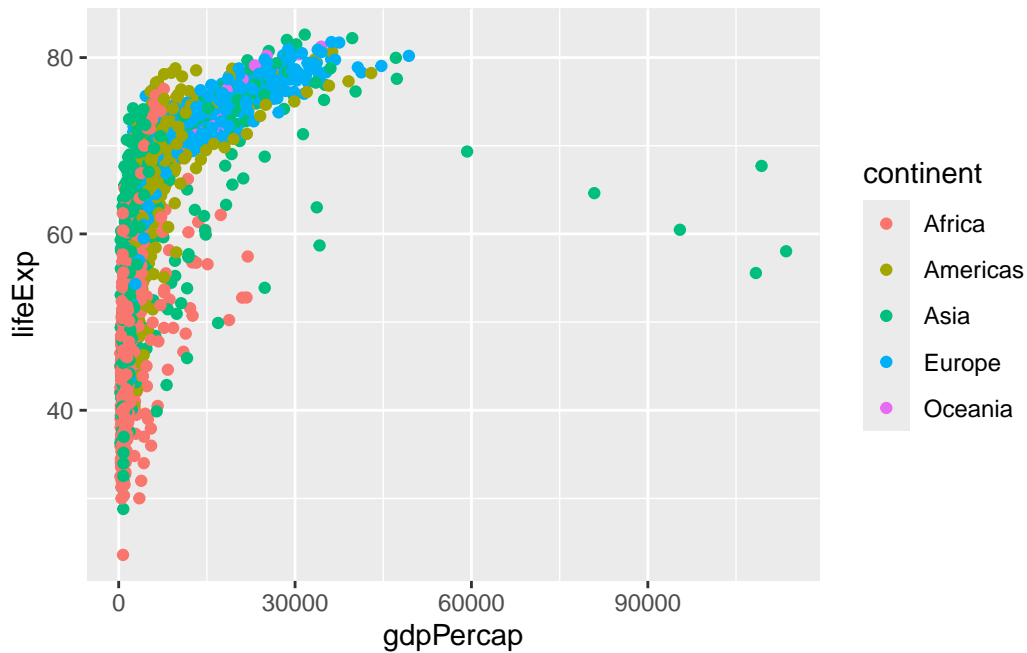
Version 1 plot gdpPercap vs. LifeExp for all rows

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

```
ggplot(gapminder) + aes(x = gdpPercap, lifeExp, col=continent) + geom_point()
```
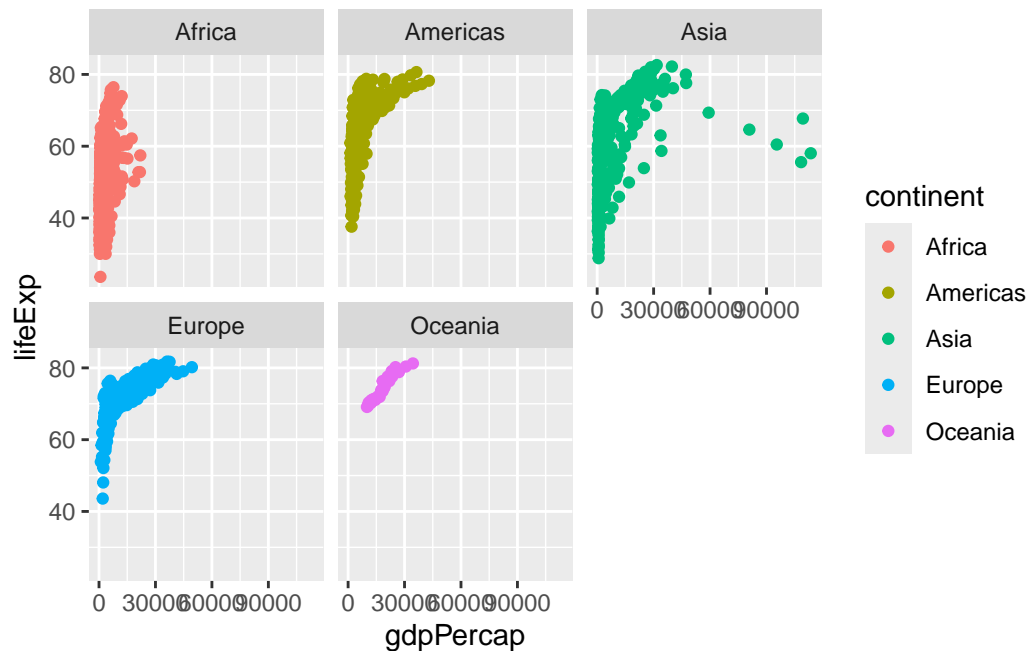
I want to see a plot for each continent – in ggplot language this is called "faceting"

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

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

First look at the dplyr package

Another add on package with a function called 'filter()' that we want to use

install.packages("dplyr")

```r
library(dplyr)
```

```
Warning: package 'dplyr' was built under R version 4.4.3
```

```
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
```

```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv
gapminder <- read.delim(url)
filter(gapminder, year == 2007, country == "Ireland")
```

```
  country continent year lifeExp     pop gdpPercap
1 Ireland    Europe 2007  78.885 4109086     40676
```

```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv
gapminder <- read.delim(url)
filter(gapminder, year == 2007, country == "United States")
```

```
        country continent year lifeExp       pop gdpPercap
1 United States  Americas 2007  78.242 301139947  42951.65
```

```
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv
gapminder <- read.delim(url)
input <- filter(gapminder, year == 2007 | year == 1977)
```

```
library(ggplot2)
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv
gapminder <- read.delim(url)
input <- filter(gapminder, year == 2007 | year == 1977)
ggplot(input) +
  geom_point(aes(x = gdpPercap, y= lifeExp, color=continent)) +
  facet_wrap(~year)
```