Class05: Data Vis with ggplot

Julia (PID: A16950824)

Graphics systems in R

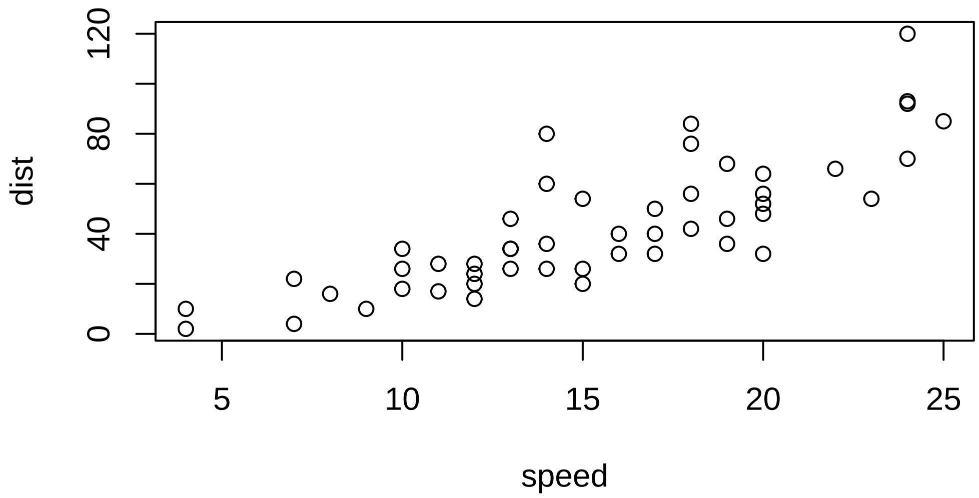
There are many graphics systems in R for making plots and figures.

We have already played a little bit with **"base R"** graphics and the `plot()` function.

Today we will start learning about a popular graphics package called `ggplot2()`

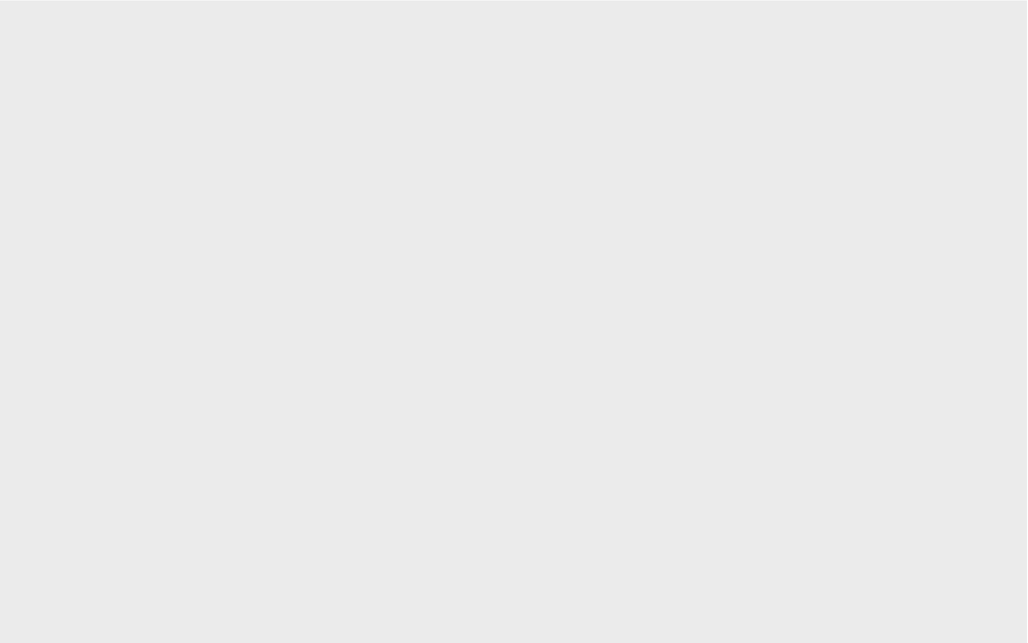
This is an add package – i.e. we need to install it. I install it (like I install any package) with the `install.packages()` function.

```
plot(cars)
```



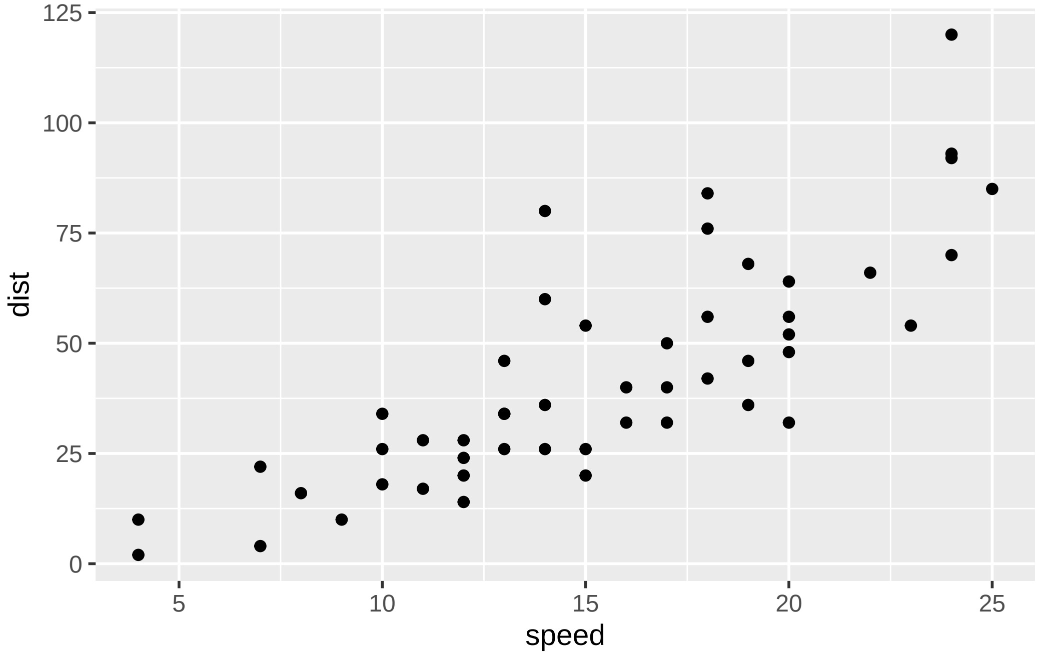
Before I can use the functions from a package I have to look up the package from my "library". We use the `library(ggplot2)` command to load it up.

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



Every ggplot is made up of at least 3 things: - data (the numbers etc. that will go into your plot) - aes (how the columns of data map to the plot aesthetics) - geoms (how the plot actually looks, points, bars, lines, ect.)

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

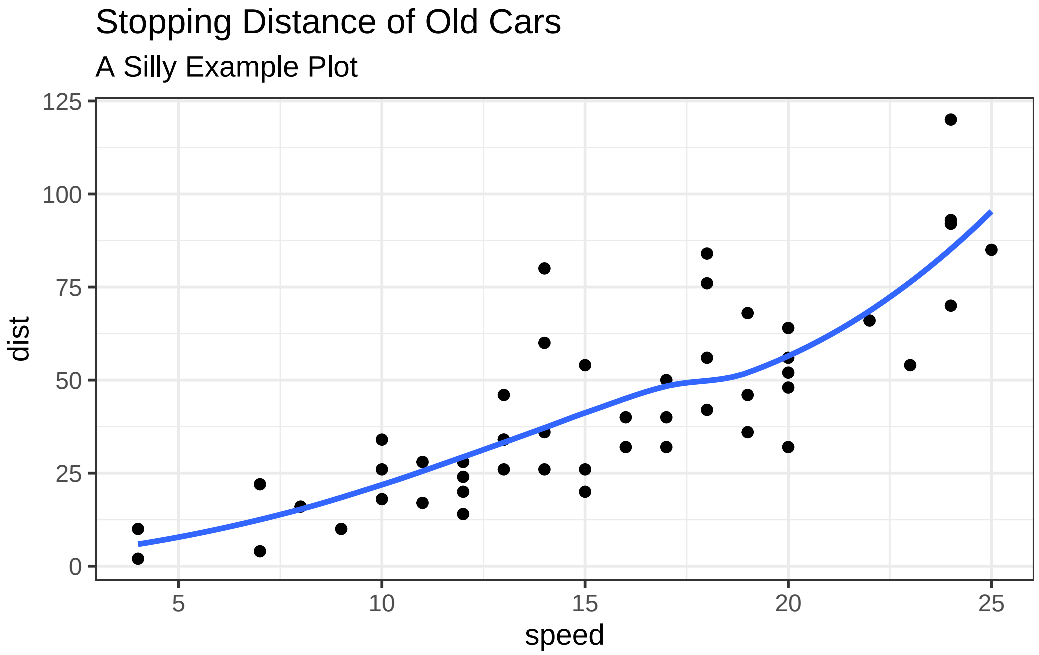


For simple plots ggplot is more verbose - it takes more code - than base R plot.

Add some more layers to our ggplot:

```r
ggplot(cars) + aes(x=speed, y=dist) + geom_point() + geom_smooth(se=F) + labs(title="Stopp
```

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



Calling up **Gene Data Set** from class hand-out

```
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
```

Getting information from data set

```
nrow(genes)
```

```
[1] 5196
```

```
ncol(genes)
```

```
[1] 4
```

```
colnames(genes)
```

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

```
num_upreg<-sum(genes$State == "up")
num_upreg
```

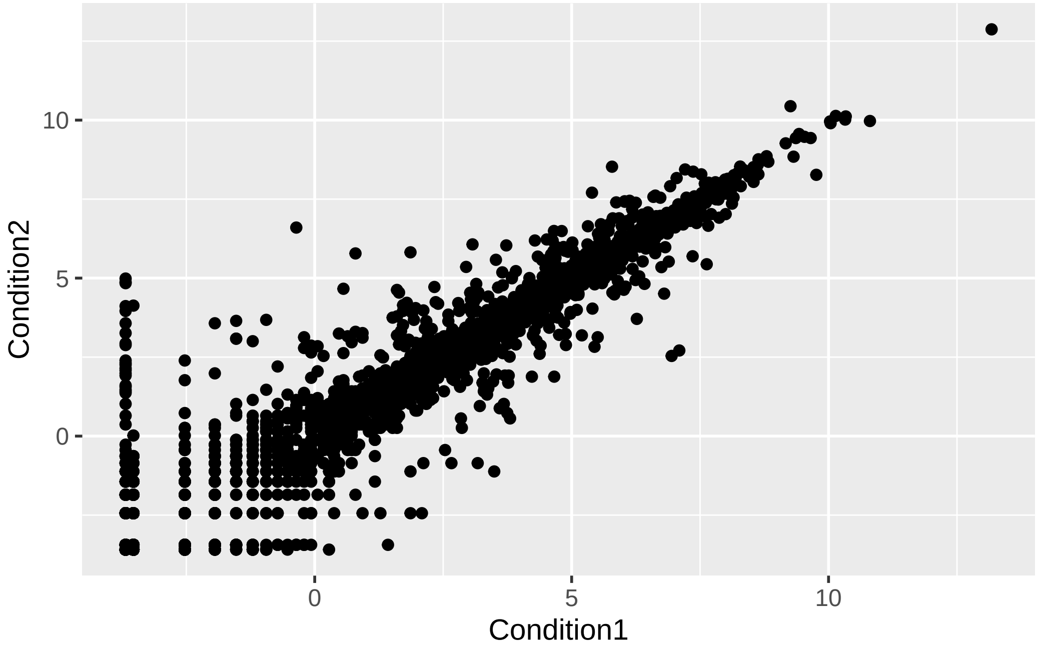
```
[1] 127
```

```
(num_upreg) / (nrow(genes)) * 100
```

```
[1] 2.444188
```

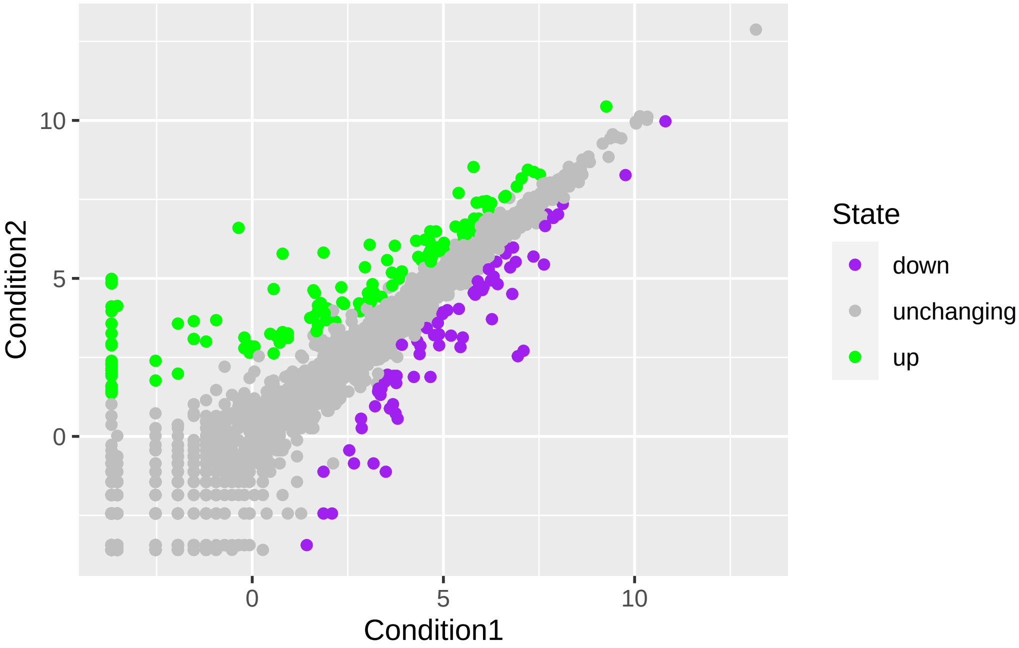
Making a simple plot of "Genes"

```
ggplot(genes) + aes(x=Condition1, y=Condition2) + geom_point()
```



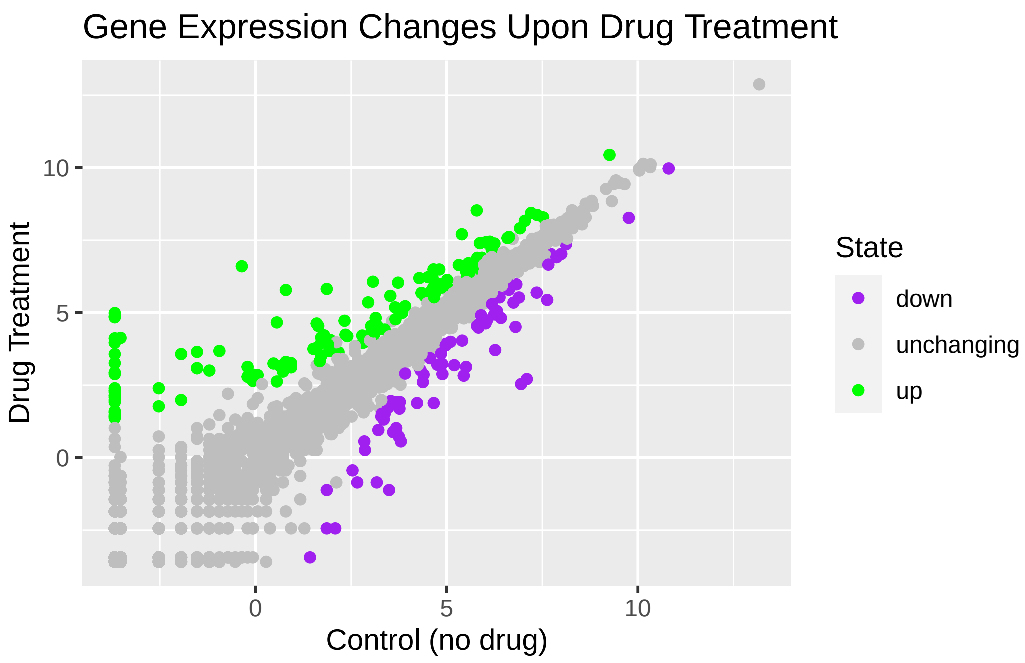
Adding some complexities to the plot

```
gene_plot<-ggplot(genes) + aes(x=Condition1, y=Condition2, col=State) + geom_point()
gene_plot + scale_colour_manual( values=c("purple", "gray", "green") )
```



Adding Labels to graph

```
gene_plot + labs(title="Gene Expression Changes Upon Drug Treatment", x="Control (no drug)
```



Installed `gapminder` and `dplyr` packages

```
# install.packages(gapminder)
# install.packages(dplyr)
library(gapminder)
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
```

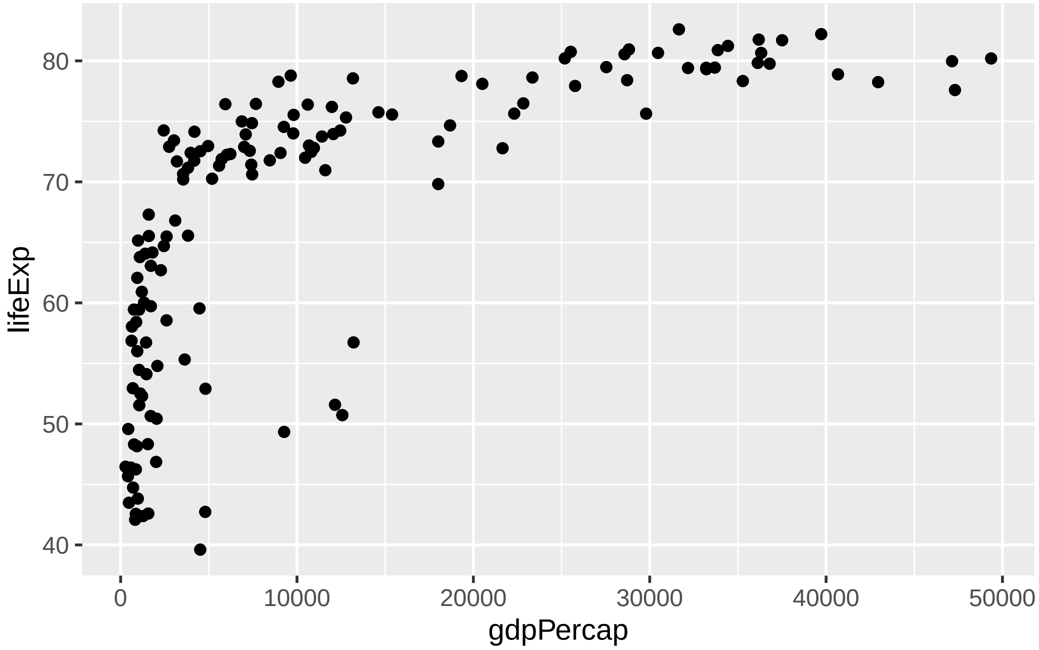
Loading in **Gapminder Data Set**

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

```
# A tibble: 142 x 6
   country     continent  year lifeExp       pop gdpPercap
   <fct>       <fct>     <int>   <dbl>     <int>     <dbl>
 1 Afghanistan Asia       2007    43.8  31889923      975.
 2 Albania     Europe     2007    76.4   3600523     5937.
 3 Algeria     Africa     2007    72.3  33333216     6223.
 4 Angola      Africa     2007    42.7  12420476     4797.
 5 Argentina   Americas   2007    75.3  40301927    12779.
 6 Australia   Oceania    2007    81.2  20434176    34435.
 7 Austria     Europe     2007    79.8   8199783    36126.
 8 Bahrain     Asia       2007    75.6    708573    29796.
 9 Bangladesh  Asia       2007    64.1 150448339     1391.
10 Belgium     Europe     2007    79.4  10392226    33693.
# i 132 more rows
```

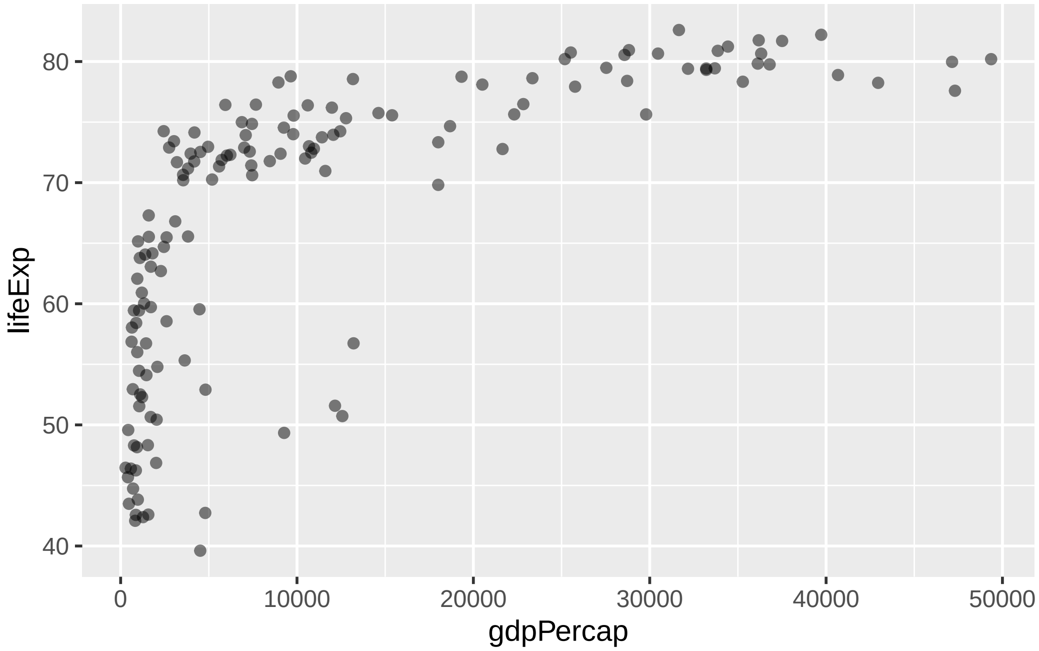
Making a basic scatter plot with data

```r
ggplot(gapminder_2007) + aes(x=gdpPercap, y=lifeExp) + geom_point()
```



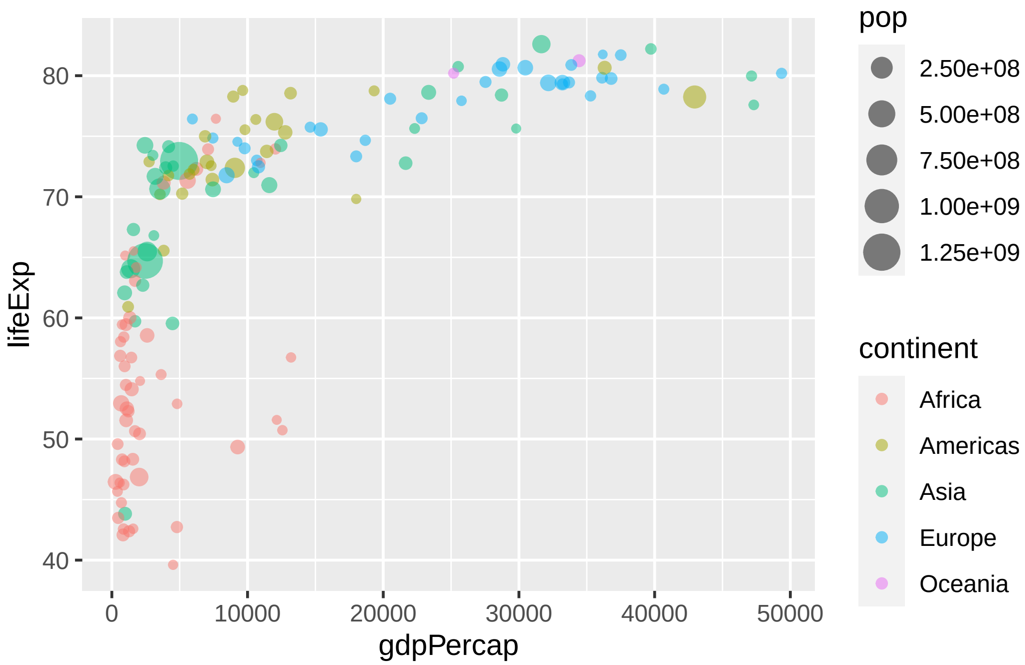
Making points slightly transparent because many of them are overlapping

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



Adding more variables

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


pop
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09
continent
Africa
Americas
Asia
Europe
Oceania
lifeExp
80
70
60
50
40
0
10000
20000
30000
40000
50000
gdpPercap