# Lab5_Data_Viz_with_ggplot
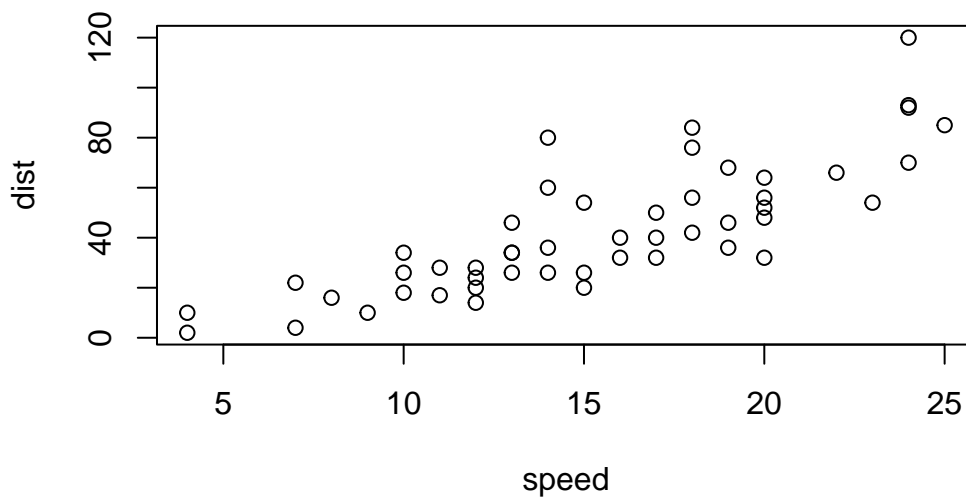
Sung Lien A16628474

**Intro to ggplot**

There are many graphics systems in R (ways to make plots and figures). These include "base" R plots. Today we will focus mostly on **ggplot2** package.

Let's start with a plot of a simple in-built dataset called `cars`.

```
plot(cars)
```



Let's see how we make this figure using **ggplot**. First I need to install this package on my computer. To install any R package I use the function `install.packages()`.

I will run 'install.packages("ggplots2") in my R console not this quarto document.

Before I can use any functions from add on packages I need to load package from my "library()" with the `library(ggplot2)` call.
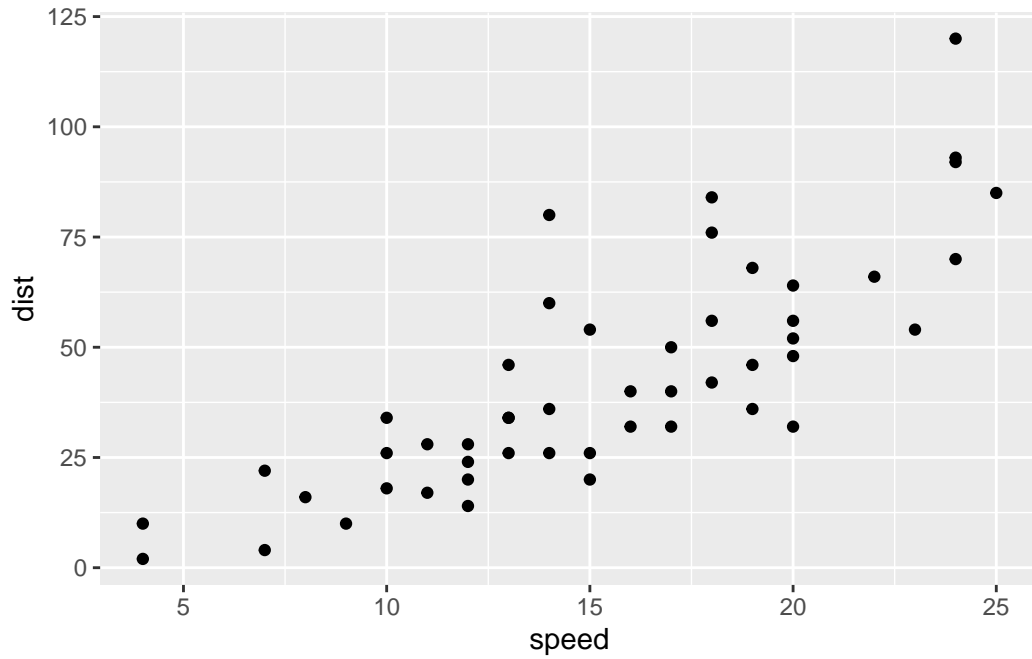
```r
library(ggplot2)
ggplot(cars)
```
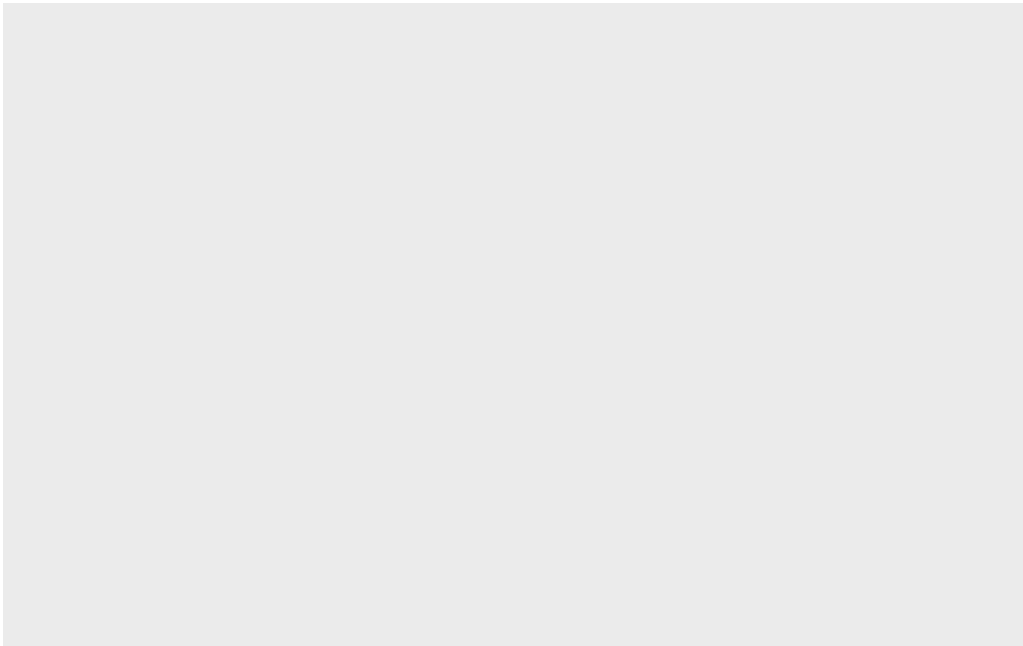
All ggplot figures have at least 3 things (called layers). These include:

-**data** (the input dataset I want to plot from), -**aes** (the aesthetic mapping of the data to my plot), -**geoms** (the geom_point(), geom_line()).
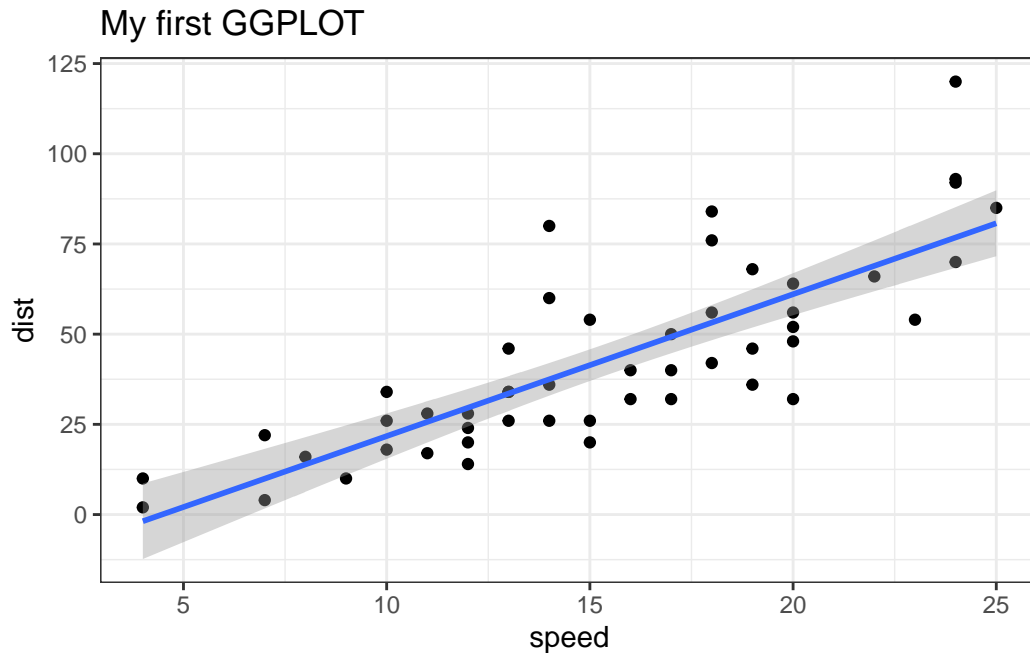
```r
ggplot(cars) + aes(x=speed, y=dist) + geom_point()
```

Let's add a line to show the relationship here:

```
ggplot(cars) + aes(x=speed, y=dist) + geom_point() + geom_smooth(method = "lm") + theme_bw()
```

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

## My first GGPLOT



Q For which phases is data visualization important in our scientific workflows? communication of results, exploratory data analysis. detection of outliers, all of the above Answer: all of the above

Q True or False? The ggplot2 package comes already installed with R? Answer: False

Q Which plot types are typically NOT used to compare distributions of numeric variables? density plots, network graphs, histograms, violin plots, box plots Answer: network graphs

Q Which statement about data visualization with ggplot2 is incorrect? Answer:ggplot is the only way to create plots in R

Q Which gemotetic layer should be used to create scatter plots

geom_point

```
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
```
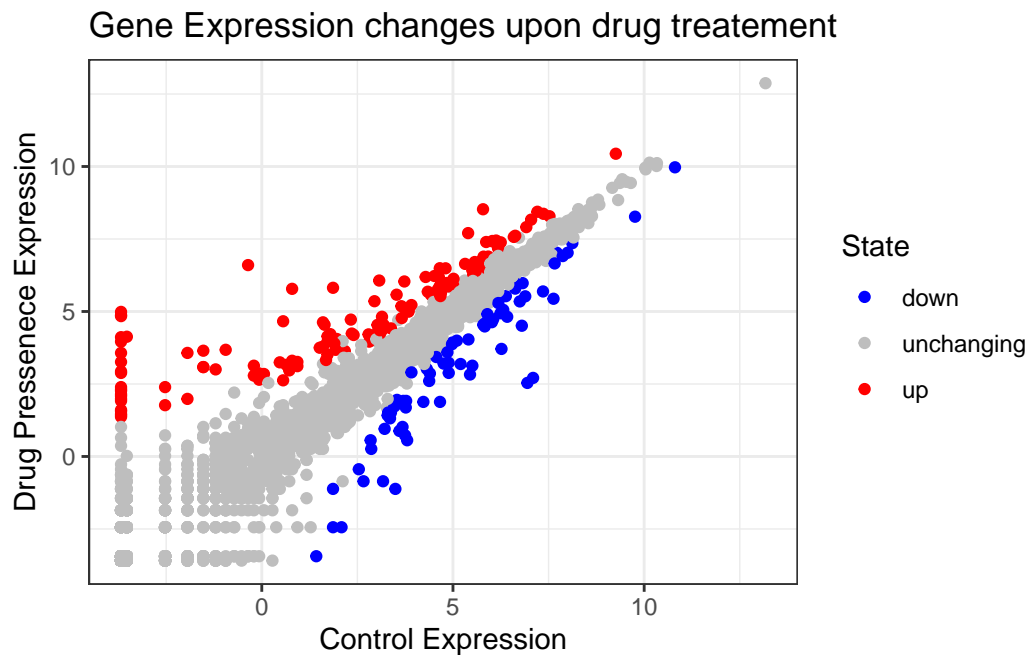
How many genes are in this dataset?

```
nrow(genes)
```

```
[1] 5196
```

A first plot of this

```
ggplot(genes) + aes(x=Condition1, y=Condition2, col=State)+ geom_point() + theme_bw() + labs
```



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

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

Q.Using your values above and 2 significant figures. What fraction of total genes is up-regulated in this dataset?

```r
round(table(genes$State)/nrow(genes), 4)
```

```
    down unchanging          up
  0.0139    0.9617      0.0244
```

```r
n.tot <- nrow(genes)

vals <- table(genes$State)

vals.percent <- vals/n.tot * 100
round(vals.percent, 2)
```

```
    down unchanging          up
    1.39      96.17        2.44
```

```r
colnames(genes)
```
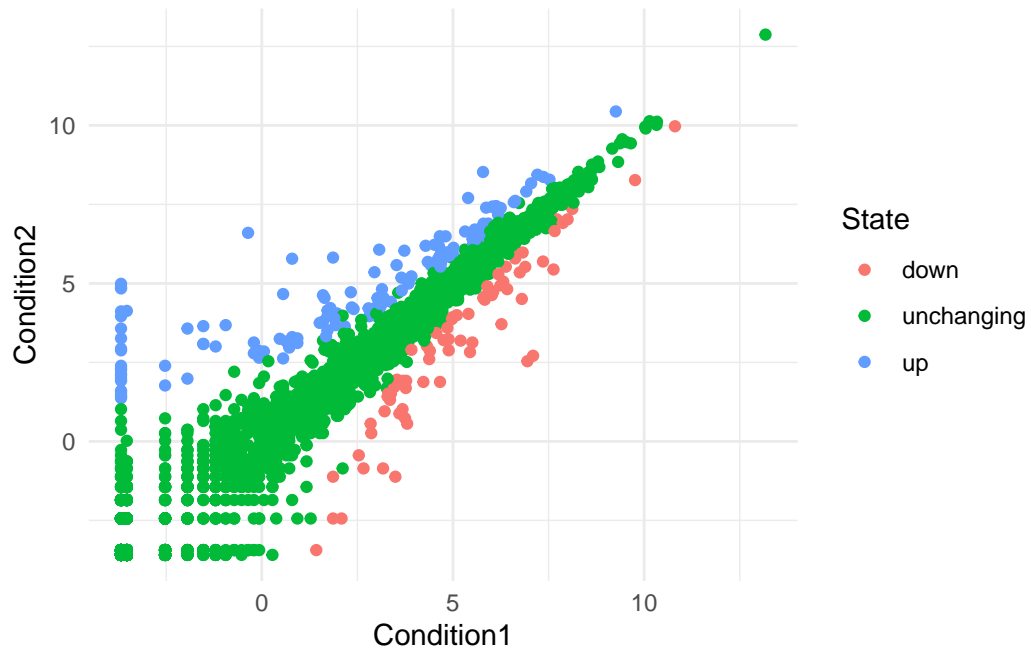
```
[1] "Gene"      "Condition1" "Condition2" "State"
```

```r
round(table(genes$State)/nrow(genes), 2)
```
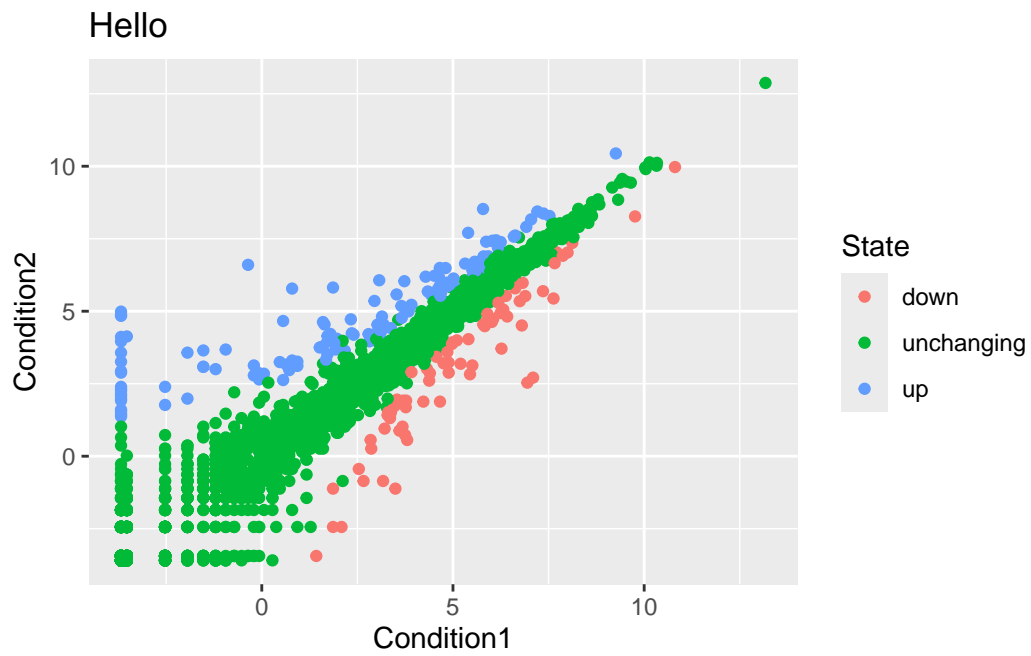
```
    down unchanging          up
    0.01       0.96        0.02
```

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

```r
p + theme_minimal()
```

```
p + labs(title = "Hello")
```

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

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

```
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
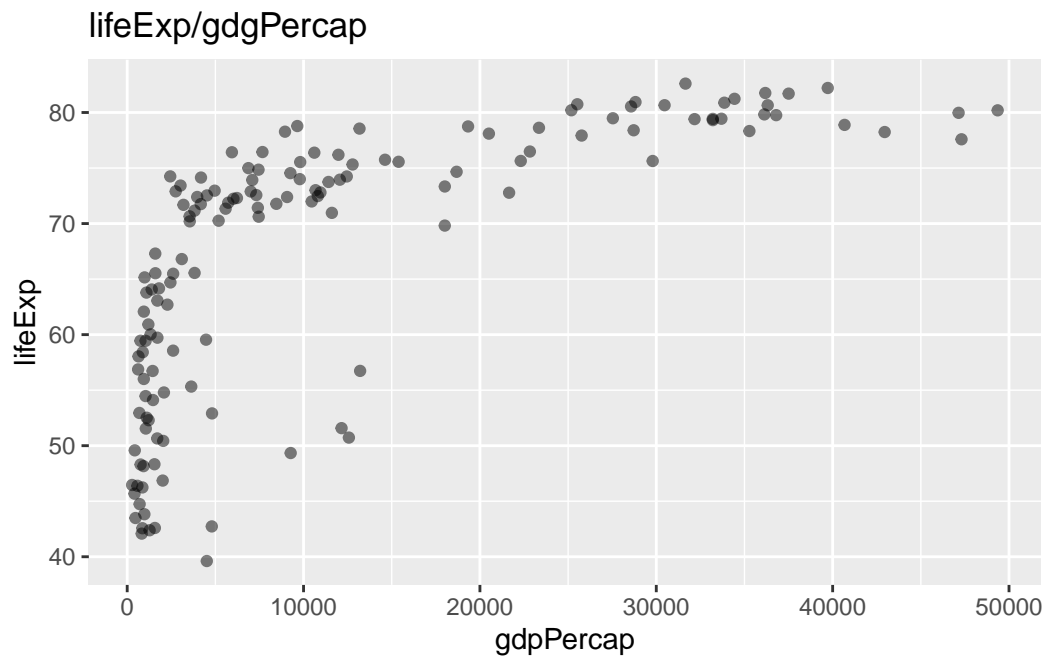```

```
library(gapminder)
```

```
Attaching package: 'gapminder'
```

```
The following object is masked _by_ '.GlobalEnv':

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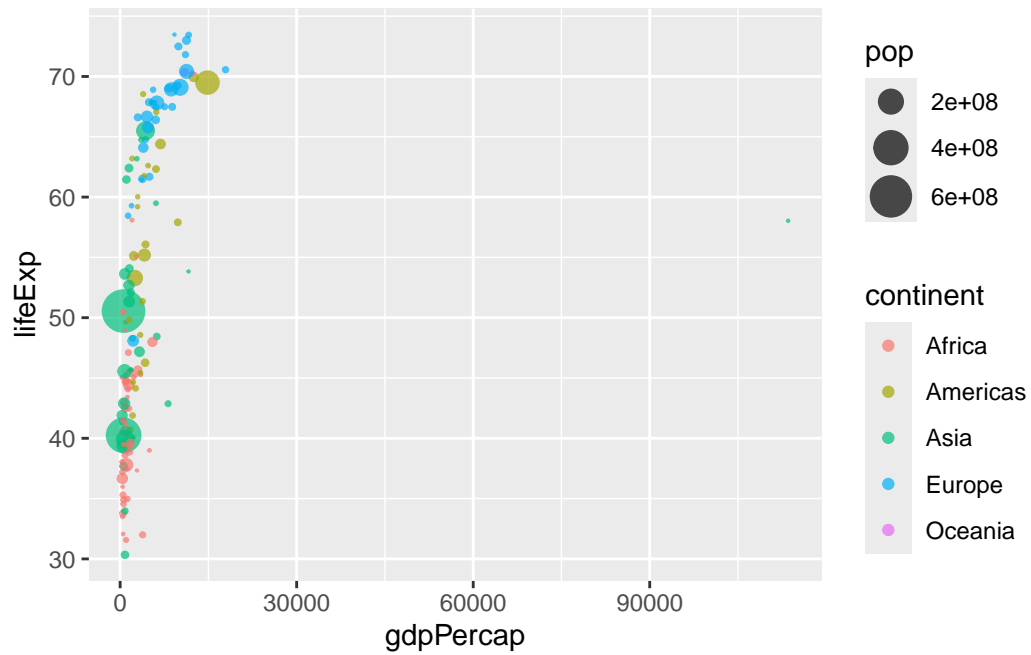
```
gapminder_2007 <-gapminder %>% filter(year==2007)
```

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

## lifeExp/gdgPercap
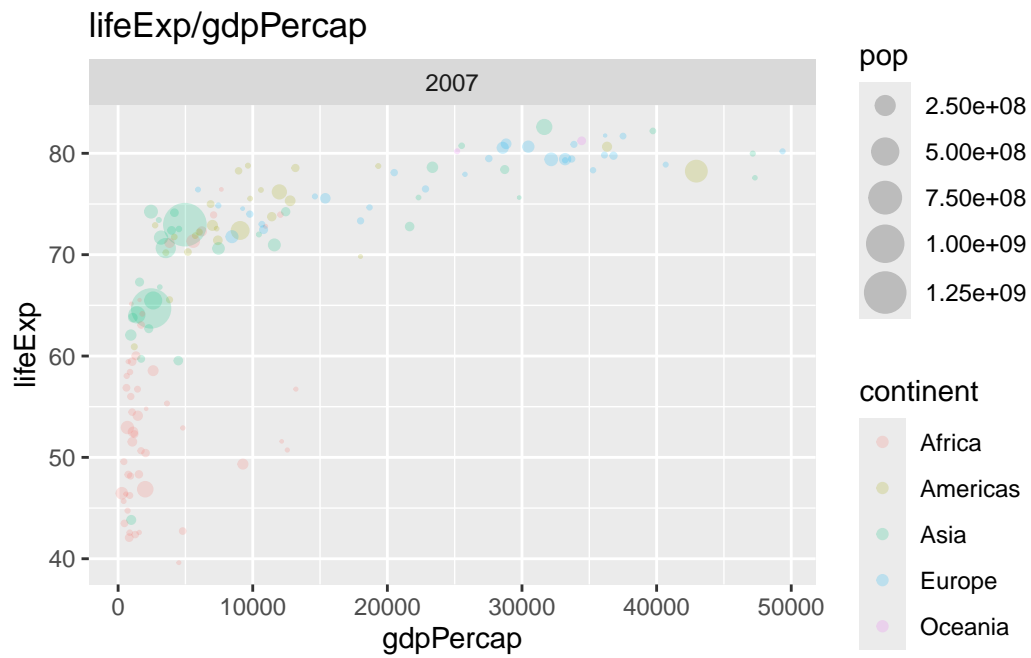


```
gapminder_1957<-gapminder %>% filter(year==1957)
```

```
ggplot(gapminder_1957) +
  aes(x = gdpPercap, y = lifeExp, color=continent,
                 size = pop) +
  geom_point(alpha=0.7) +
  scale_size_area(max_size = 7)
```

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

```
gapminder_1957 <- gapminder %>% filter(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 = 8) +
  facet_wrap(~year)
```