

# Class 05: Data Visualization with GGPLOT

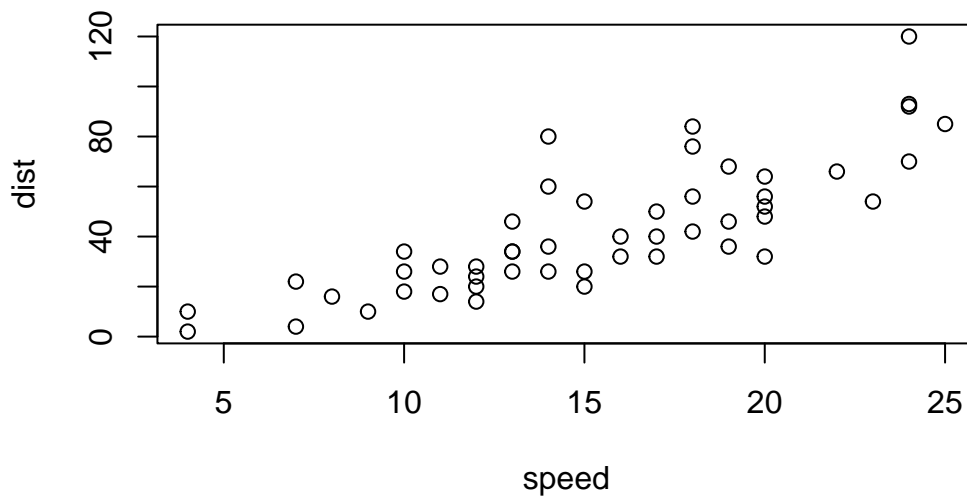
Victor Yu

#Plotting in R

R has multiple plotting and graphics systems. The most popular of which is **ggplot2**.

We have already played with “base” R graphics. This comes along with R “out of the box”

```
plot(cars)
```



Compared to base R plots ggplot is much more verbose - I need to write more code to get simple plots like the above.

To use ggplot I need to first install the ggplot2 package. To install any package in R I use the `install.packages()` command along with the package name.

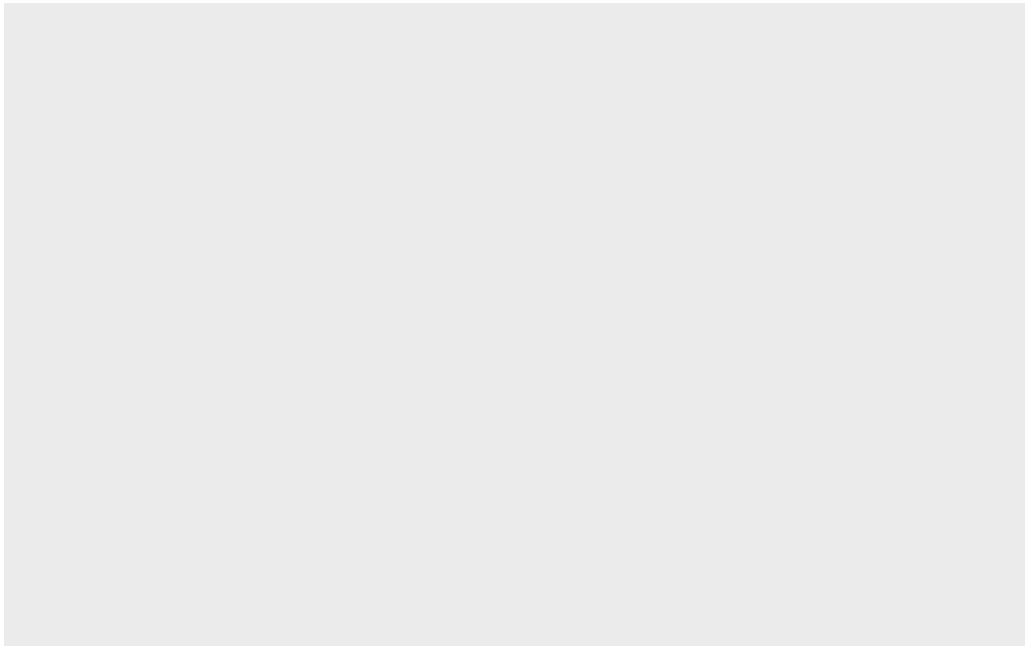
The install is a one time only requirement. The package is now on our computer. I don't need to re-install it.

However, I can't just use it without loading it up with a `library()` call.

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

```
library(ggplot2)
```

```
ggplot(cars)
```

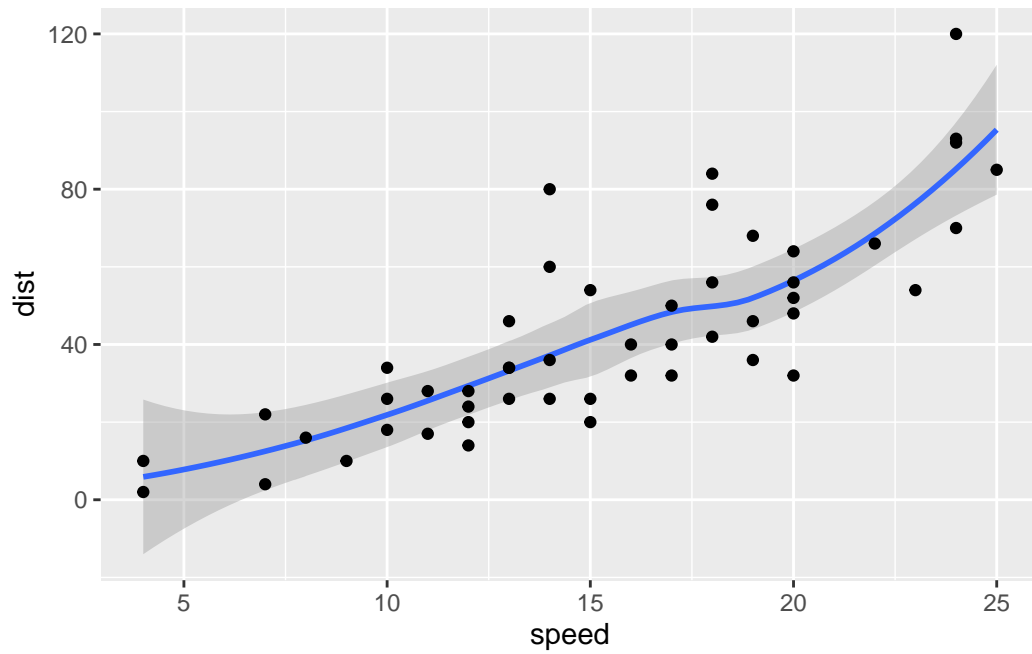


All ggplot figures need at least 3 things:

-data (this is the data.frame with our numbers) -aesthetics () ("aes", how our data maps to the plot) -geoms (do we want lines, points, columns, etc...)

```
#Store to BB as a variable.  
bb <- ggplot(data=cars) +  
  aes(x=speed, y=dist) +  
  geom_smooth() + geom_point()  
  
bb
```

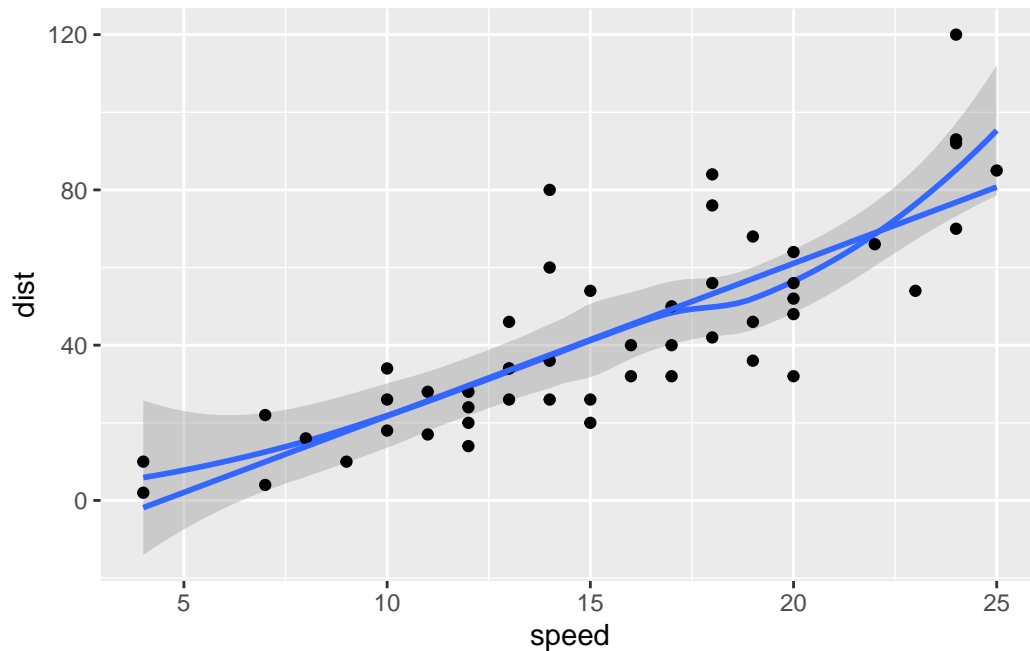
```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



I want a trend line to show the relationship between speed and stopping (smooth)

```
# "bb is your base plot. Add stuff onto it  
bb + geom_smooth(method = "lm", se=FALSE)
```

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



```
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

```
ncol(genes)
```

```
[1] 4
```

```
colnames(genes)
```

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

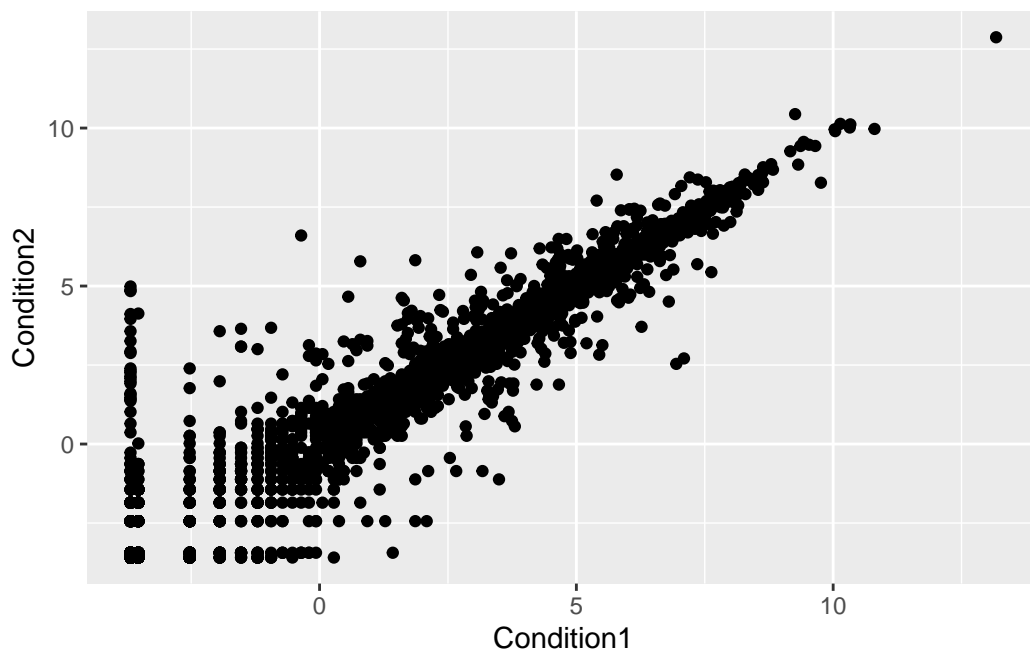
```
table(genes$State)
```

down	unchanging	up
72	4997	127

```
round ( table(genes$State) / nrow(genes)*100, 2)
```

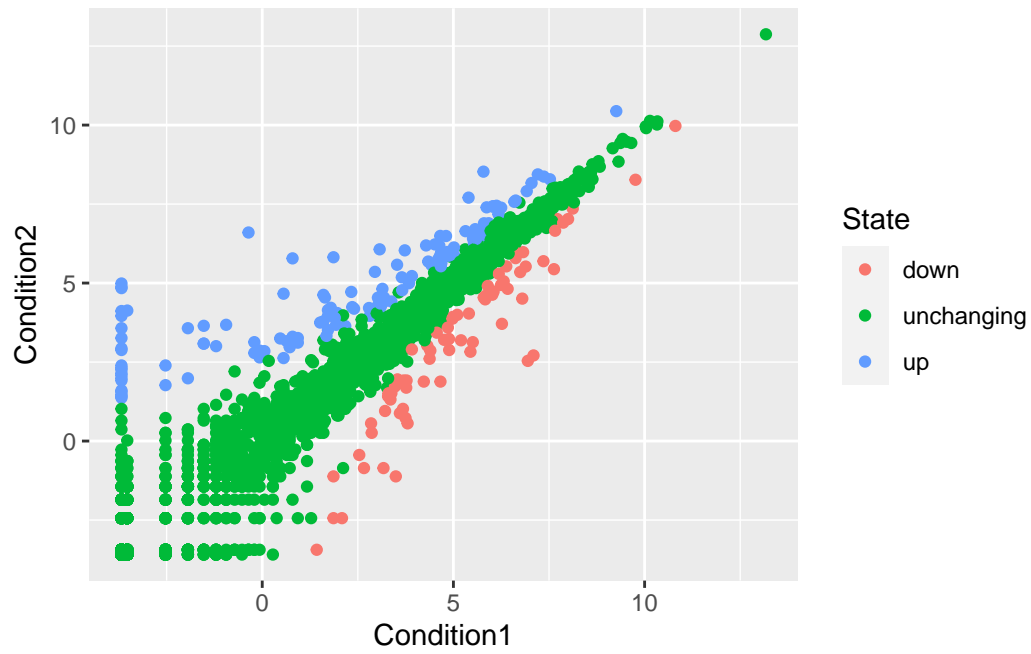
down	unchanging	up
1.39	96.17	2.44

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



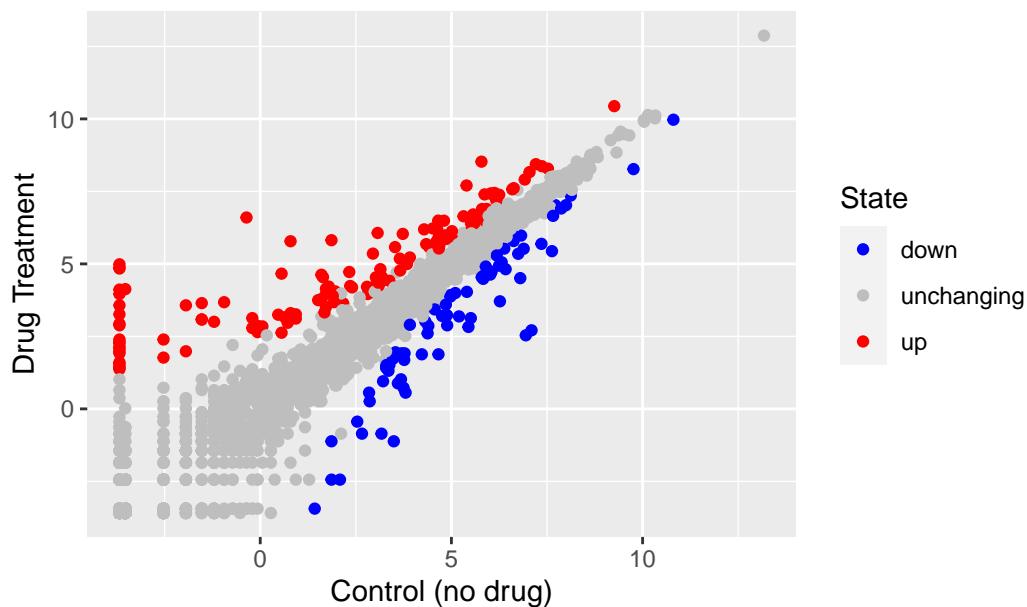
```
p <- ggplot(genes) +  
  aes(x=Condition1, y=Condition2, col=State) +
```

```
geom_point()  
p
```



```
p + scale_colour_manual(values = c("blue", "gray","red")) +  
  labs(title = "Gene Expression Changes Upon Drug Treatment", x="Control (no drug)", y="Drug Treatment")
```

## Gene Expression Changes Upon Drug Treatment



```
#install.packages ("dplyr") for DPLYR
#install.packages("gapminder") for gapminder
library(gapminder)
library(dplyr)
```

Attaching package: 'dplyr'

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

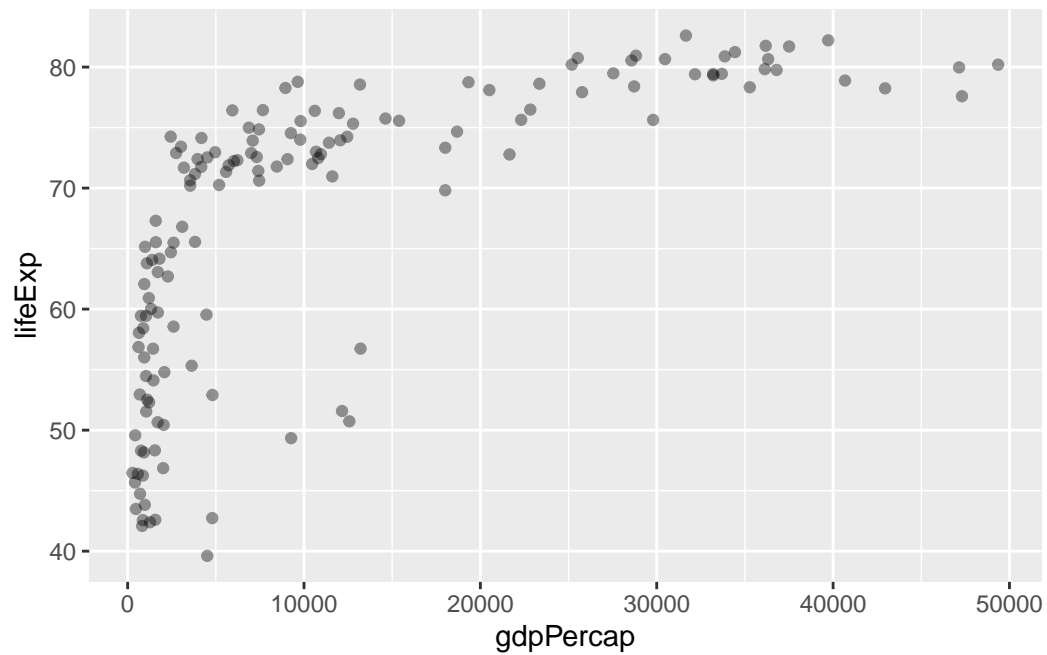
filter, lag

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

intersect, setdiff, setequal, union

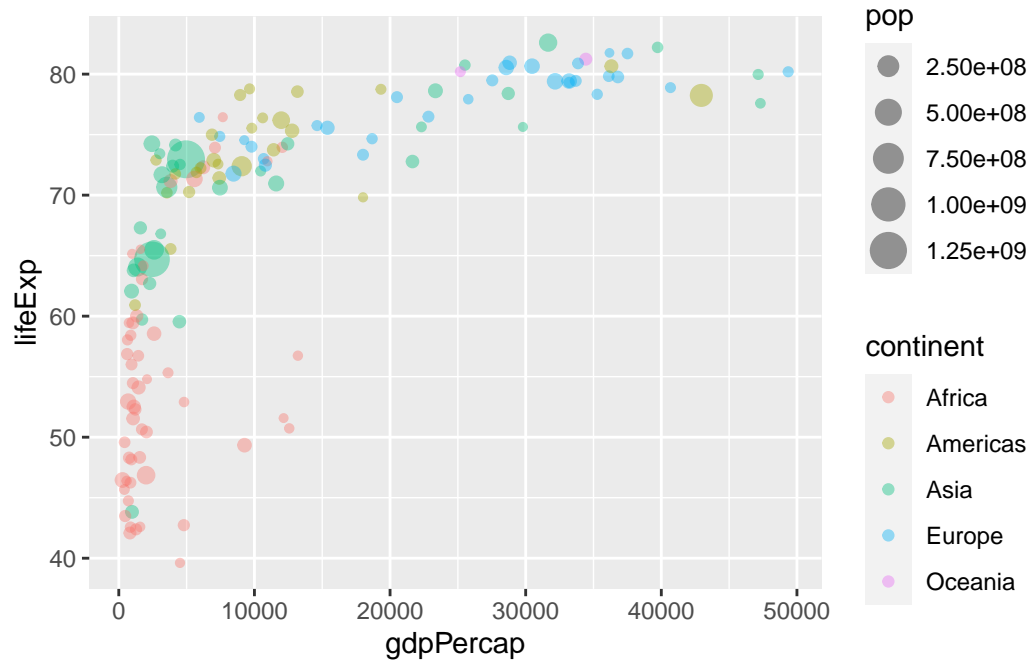
```
# File location online
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder."
gapminder <- read.delim(url)
gapminder_2007 <- gapminder %>% filter(year==2007)
```

```
ggplot(gapminder_2007) +
  aes(x=gdpPerCap, y=lifeExp) +
  geom_point(alpha=0.4)
```

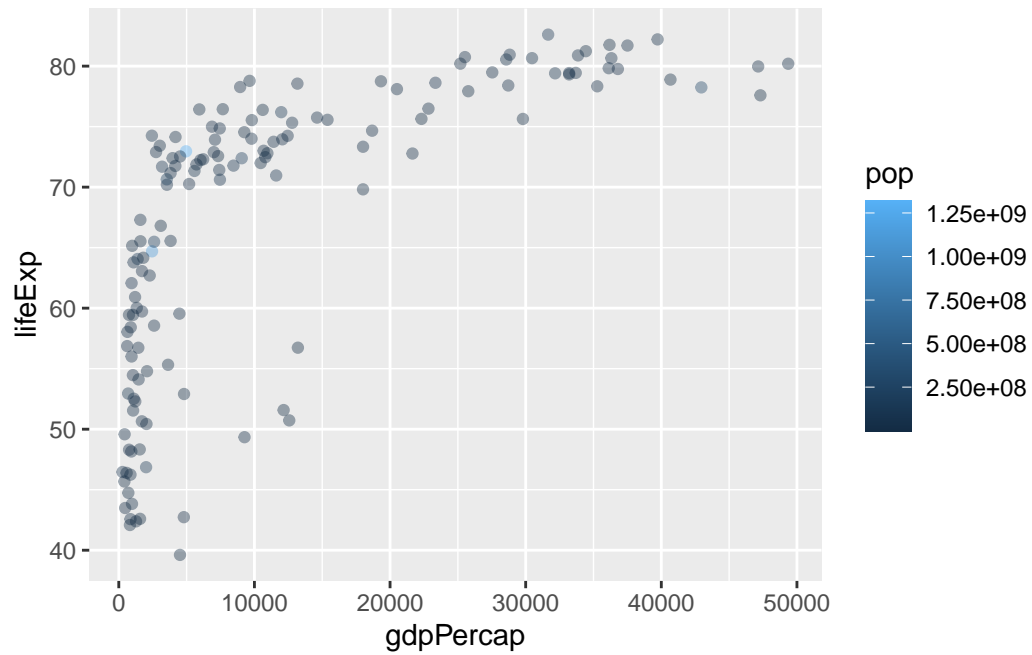


```
ggplot(gapminder_2007) +
  aes(x=gdpPerCap, y=lifeExp, color= continent, size = pop) +
  geom_point(alpha=0.4)
```

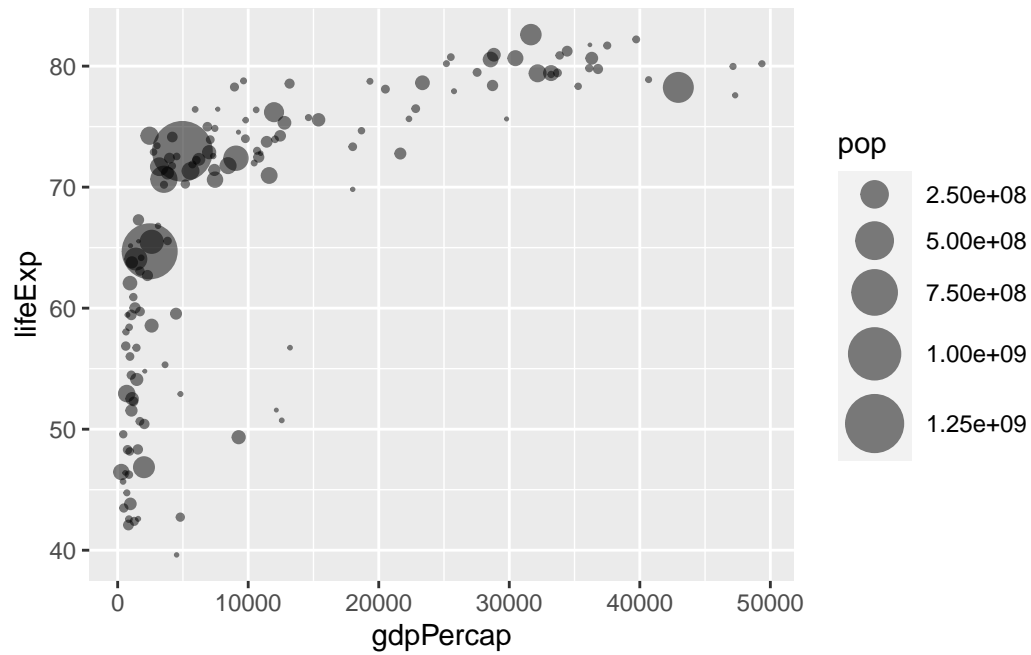




```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color= pop) +  
  geom_point(alpha=0.4)
```

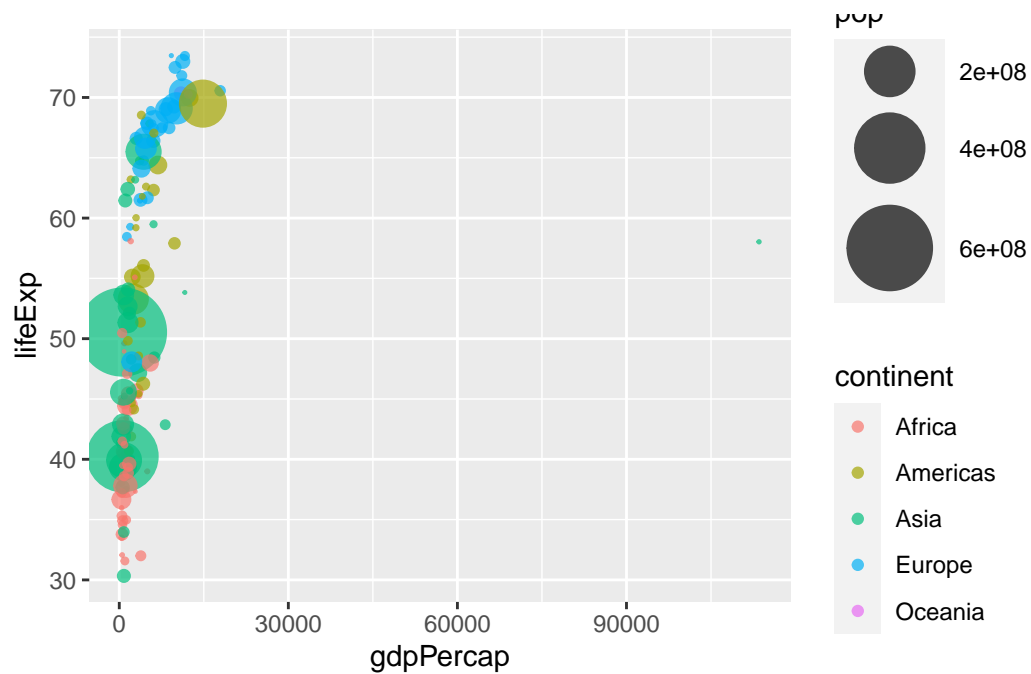


```
ggplot(gapminder_2007) +  
  geom_point(aes(x = gdpPerCap, y = lifeExp, size = pop), alpha = 0.5) +  
  scale_size_area (max_size=10)
```



```
gapminder <- read.delim(url)
gapminder_1957 <- gapminder %>%filter(year==1957)

ggplot(gapminder_1957) +
  geom_point(aes(x=gdpPercap, y=lifeExp, color = continent, size = pop), alpha = 0.7) +
  scale_size_area(max_size = 15)
```



```
gapminder <- read.delim(url)
gapminder_1957_2007 <- gapminder %>%filter(year==1957 | year ==2007)

ggplot(gapminder_1957_2007) +
  geom_point(aes(x=gdpPercap, y=lifeExp, color = continent, size = pop), alpha = 0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```

