

Class 5: Data Vis w/ ggplot

Xinyu Wen (PID: A17115443)

Table of contents

Intro to ggplot	1
Gene Expression Figure	6
7. Going Further	11
gapminder_2007	11
gapminder_1957	14

Intro to ggplot

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

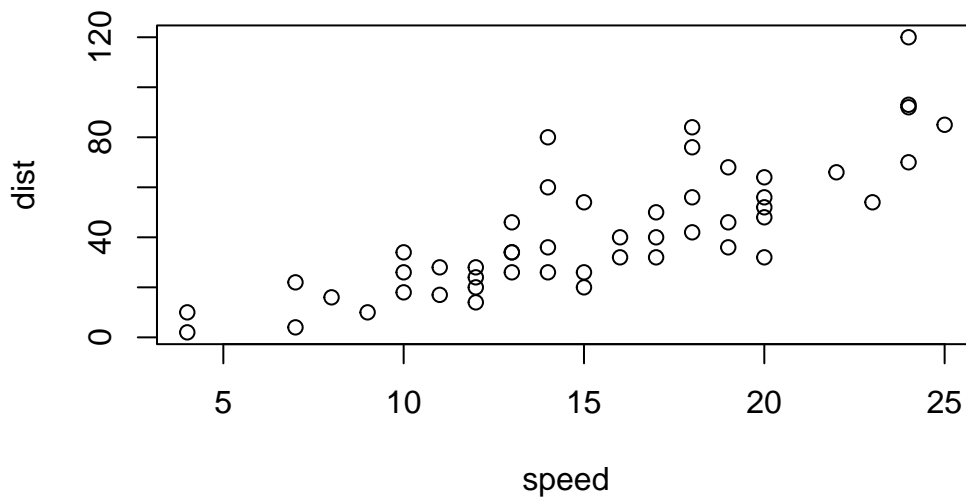
Let’s start with a plot of simple in-built dataset called **cars**.

```
cars
```

```
      speed dist
1         4    2
2         4   10
3         7    4
4         7   22
5         8   16
6         9   10
7        10   18
8        10   26
9        10   34
10       11   17
11       11   28
12       12   14
13       12   20
```

14	12	24
15	12	28
16	13	26
17	13	34
18	13	34
19	13	46
20	14	26
21	14	36
22	14	60
23	14	80
24	15	20
25	15	26
26	15	54
27	16	32
28	16	40
29	17	32
30	17	40
31	17	50
32	18	42
33	18	56
34	18	76
35	18	84
36	19	36
37	19	46
38	19	68
39	20	32
40	20	48
41	20	52
42	20	56
43	20	64
44	22	66
45	23	54
46	24	70
47	24	92
48	24	93
49	24	120
50	25	85

```
plot(cars)
```

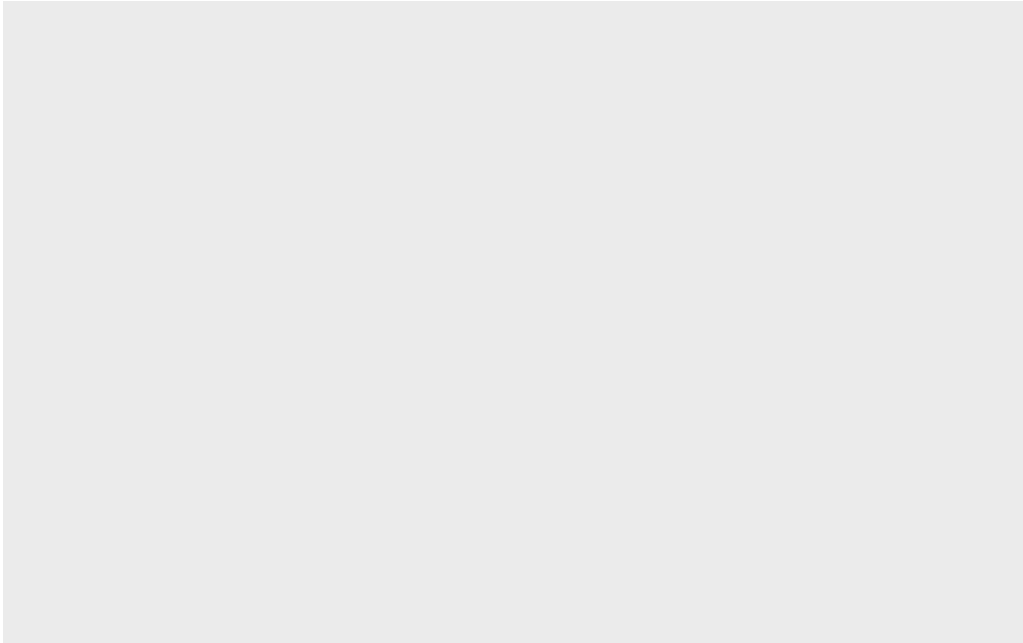


Let's see how we can make this figure using **ggplot2**. First let's install this package on my computer. To install any R package, use the function `install.packages()`

I will run `install.packages("ggplot2")` in my R console, not this quarto document, so that it will not be downloaded everytime I do Render.

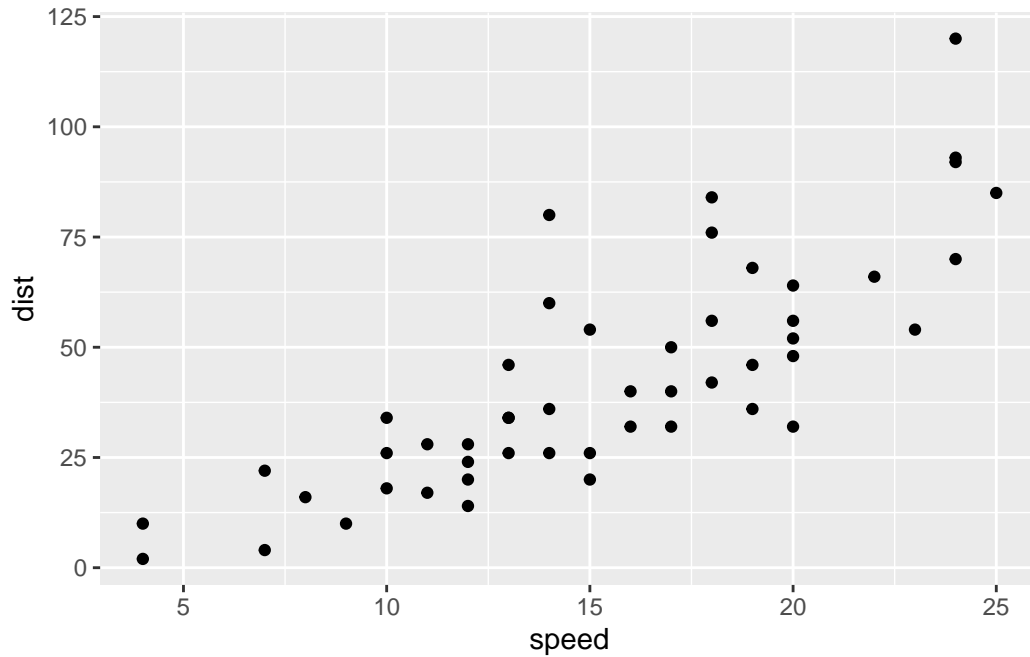
Before I can use any functions from add on packages, I need to load the package from my "library()" with the `library(ggplot2)`.

```
library(ggplot2)
ggplot(cars)
```



All ggplot have at least 3 layers: - **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()`, `geom_col()` etc.)

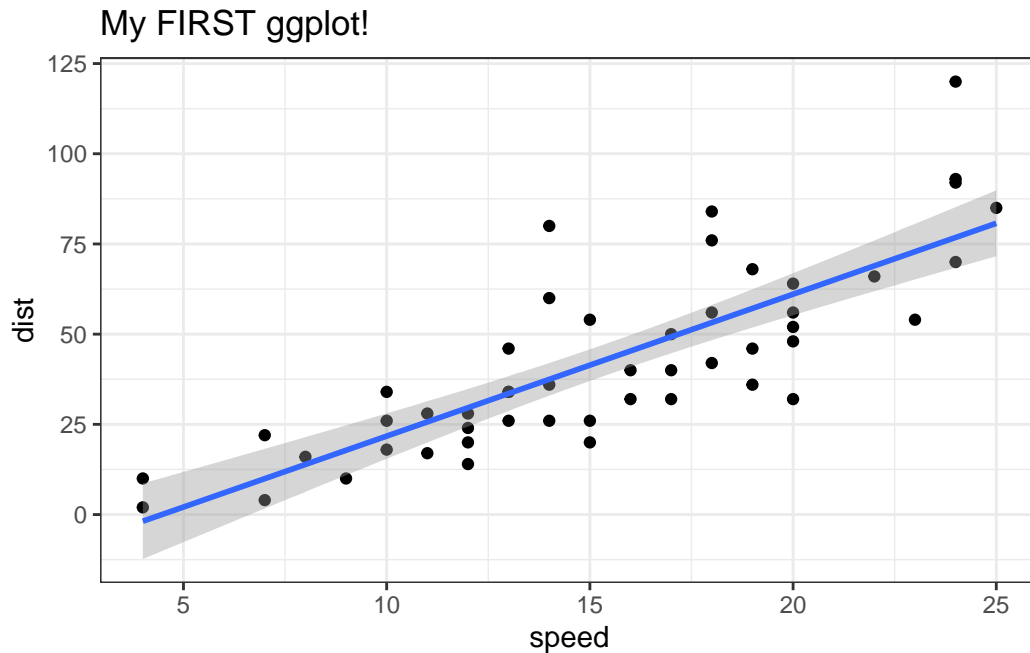
```
ggplot(cars) +  
  aes(x=speed, y=dist) +  
  geom_point()
```



Let's add a line to show the relationship:

```
ggplot(cars) +  
  aes(x=speed, y=dist) +  
  geom_point() +  
  geom_smooth(method = "lm") +  
  theme_bw() +  
  labs(title="My FIRST ggplot!")
```

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



Q1. Which geometric layer should be used to create scatter plots in ggplot2?

`geom_point()`

Gene Expression Figure

The code to read the dataset:

```
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

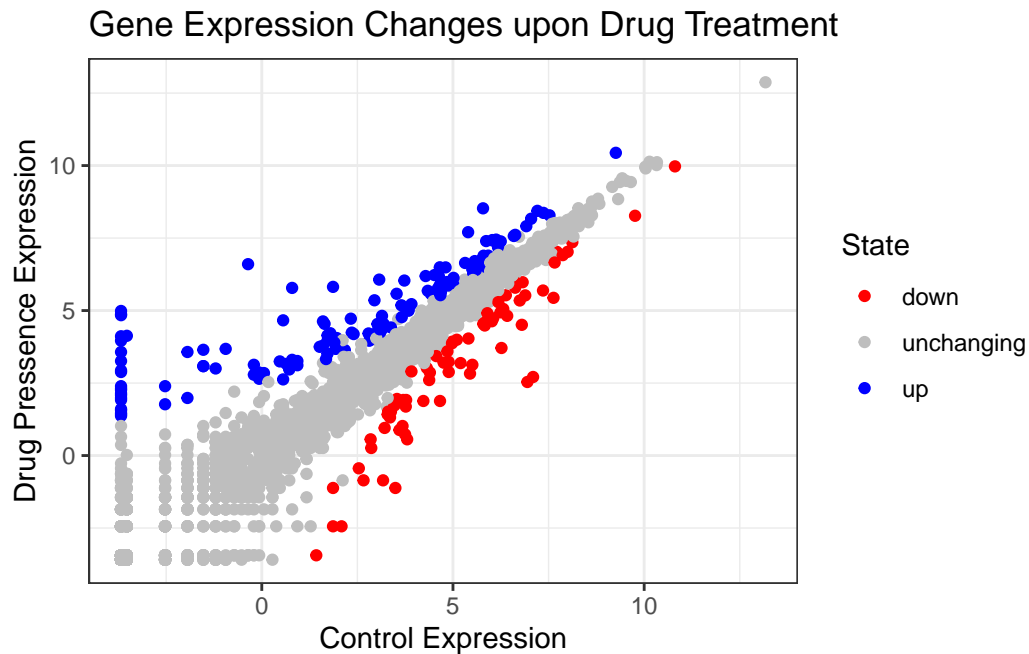
Q. Use the `nrow()` function to find out how many genes are in this dataset.

```
nrow(genes)
```

```
[1] 5196
```

A first plot of this dataset:

```
ggplot(genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point() +  
  theme_bw() +  
  labs(title="Gene Expression Changes upon Drug Treatment",  
        x="Control Expression",  
        y="Drug Pressence Expression") +  
  scale_colour_manual(values=c("red","gray","blue"))
```



Q. Use the `colnames()` function and the `ncol()` function on the `genes` data frame to find out what the column names are (we will need these later) and how many columns there are. How many columns did you find?

```
colnames(genes)
```

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

```
ncol(genes)
```

```
[1] 4
```

Q. Use the `table()` function on the `State` column of this data.frame to find out how many 'up' regulated genes there are. What is your answer?

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

down	unchanging	up
72	4997	127

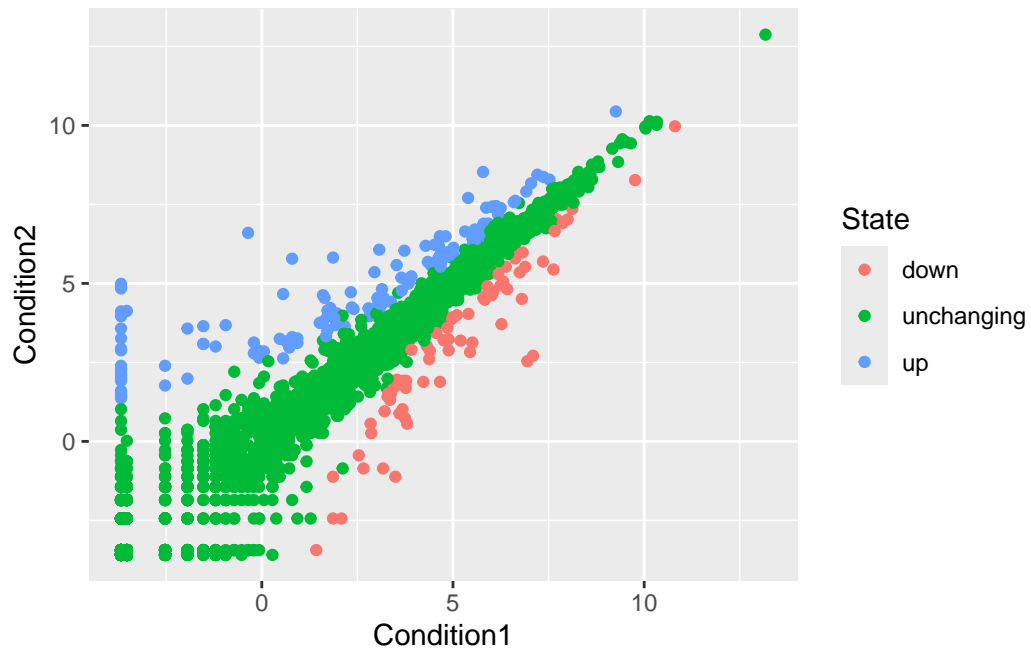
Q. Using your values above and 2 significant figures. What fraction of total genes is up-regulated in this dataset?

```
round(table(genes$State)/nrow(genes), 4) *100
```

down	unchanging	up
1.39	96.17	2.44

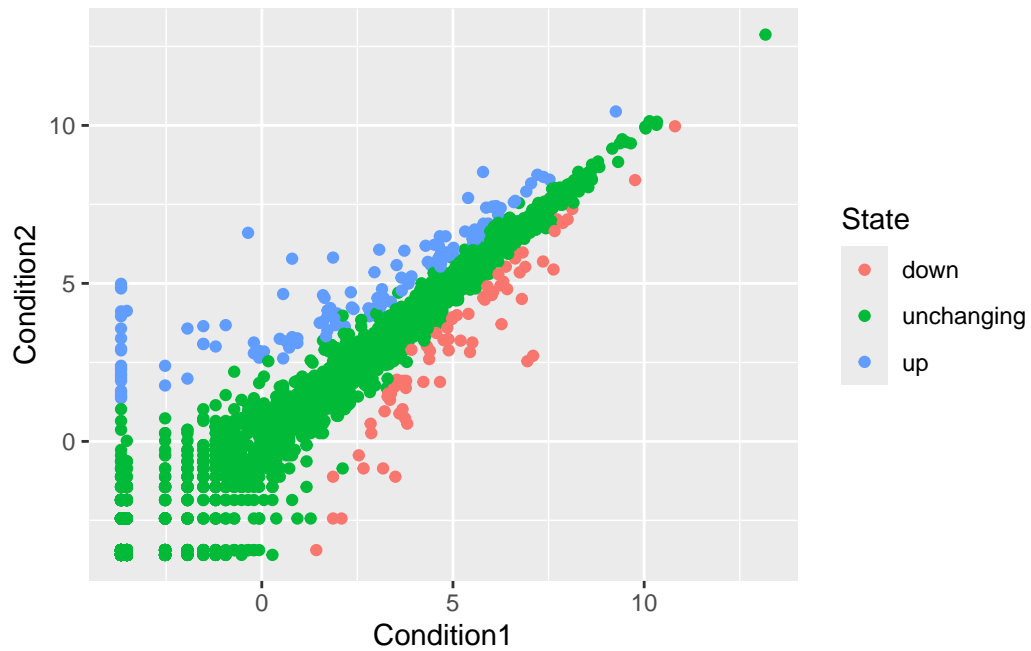
```
p <- ggplot(genes) +  
  aes(x=Condition1, y=Condition2, col=State) +  
  geom_point()
```

```
p
```

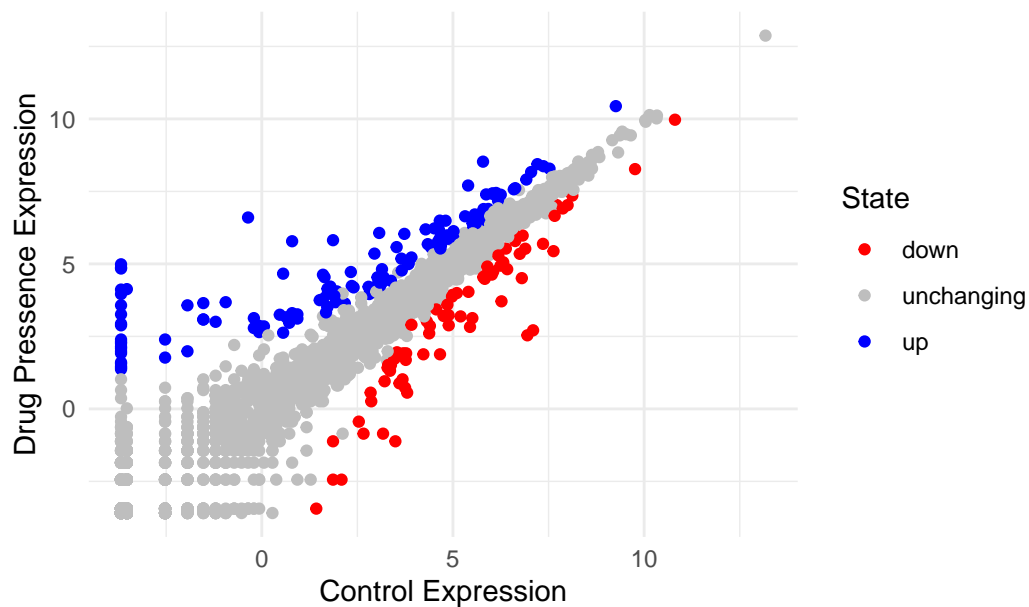
`ggplot(mtcars) + aes(x=mpg, y=disp) + geom_point()` ABOVE PLOT `ggplot(mtcars, aes(mpg, disp)) + geom_point()` BELOW PLOT THEY GENERATE THE SAME PLOT

```
ggplot(genes, aes(Condition1, Condition2, col=State)) +
  geom_point()
```



```
p + scale_colour_manual(values=c("red","gray","blue")) +
  theme_minimal() +
  labs(title="Gene Expression Changes upon Drug Treatment",
        x="Control Expression",
        y="Drug Pressence Expression")
```

Gene Expression Changes upon Drug Treatment



7. Going Further

gapminder_2007

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

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

```
library(dplyr)
```

Attaching package: 'dplyr'

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

filter, lag

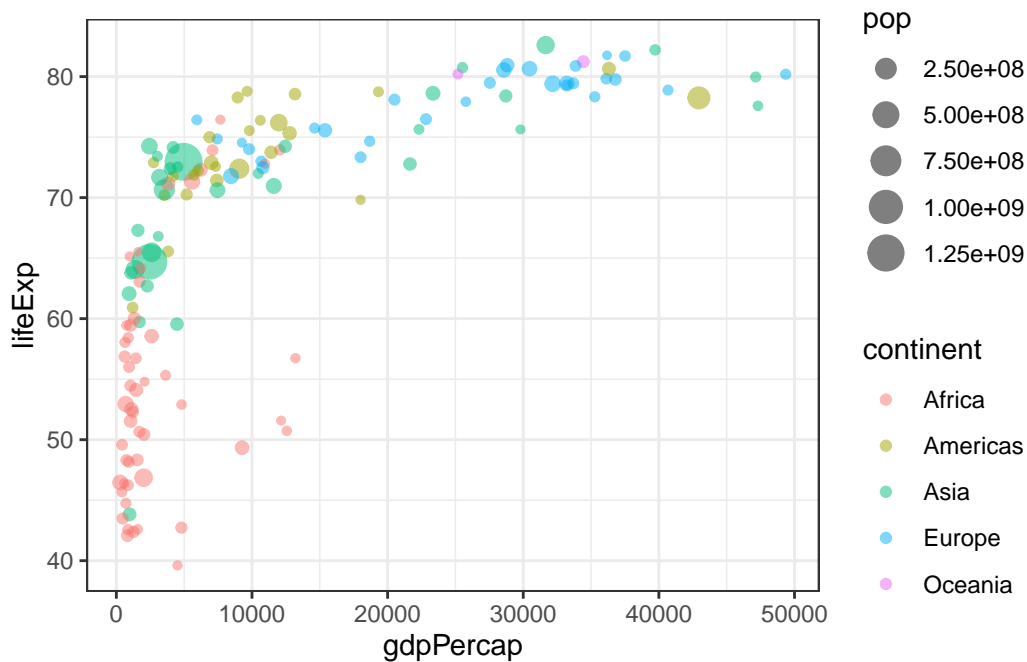
The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

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

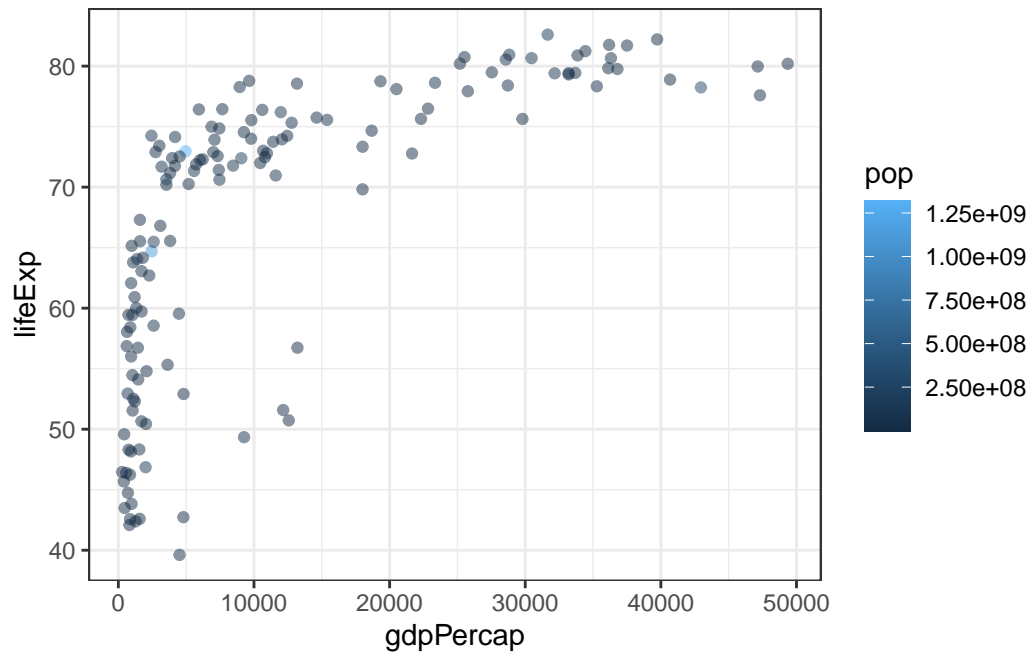
Q. Complete the code below to produce a first basic scatter plot of this gapminder_2007 dataset:

```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +  
  geom_point(alpha=0.5) +  
  theme_bw()
```

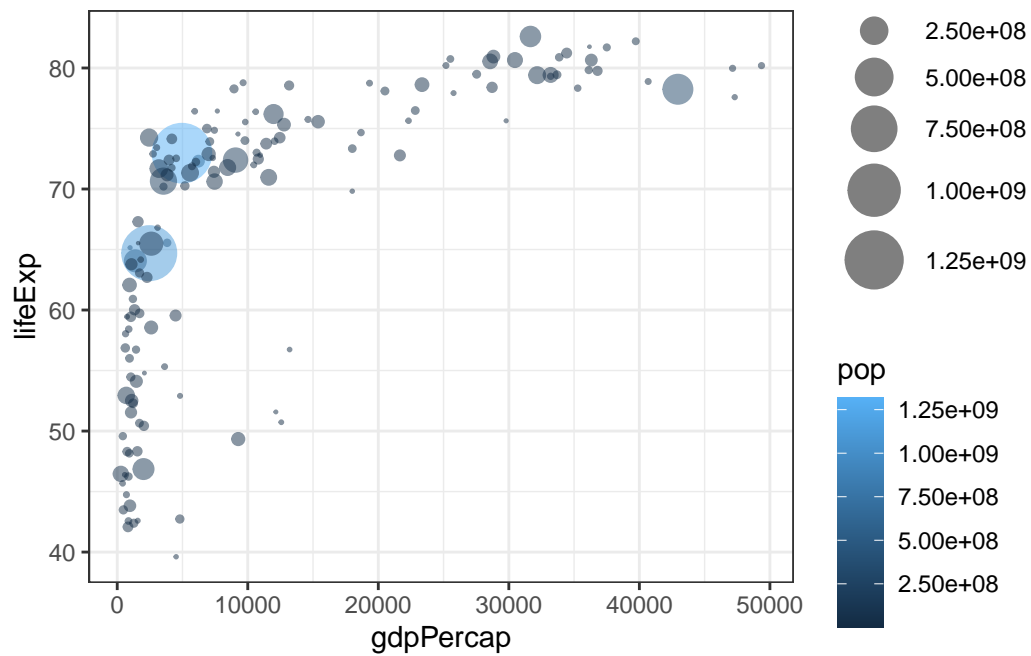


```
gap2007 <- ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp) +  
  geom_point(alpha=0.5) +  
  theme_bw()
```

```
gap2007 + aes(color = pop)
```



```
gap2007 + aes(color = pop, size = pop) +
  scale_size_area(max_size = 10)
```

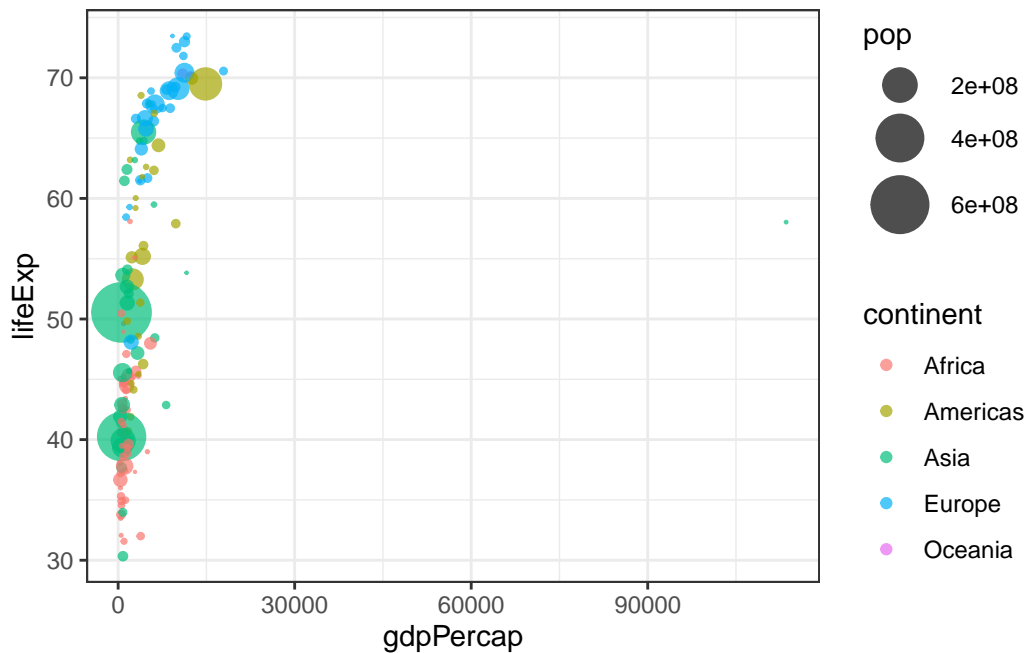


gapminder_1957

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

```
gap1957 <- ggplot(gapminder_1957) +  
  aes(x=gdpPercap, y=lifeExp) +  
  geom_point(alpha=0.7) +  
  theme_bw()
```

```
gap1957 + aes(color = continent, size = pop) +  
  scale_size_area(max_size = 10)
```



```
gap19572007 <- gapminder %>% filter(year==1957 | year==2007)  
ggplot(gap19572007) +  
  geom_point(aes(x = gdpPercap, y = lifeExp, color=continent,  
                 size = pop), alpha=0.7) +  
  theme_bw() +  
  scale_size_area(max_size = 10) +  
  facet_wrap(~year)
```

