Class 5: Data Viz with ggplot

Yaniv Iny (PID:A18090586)

inro to ggplot

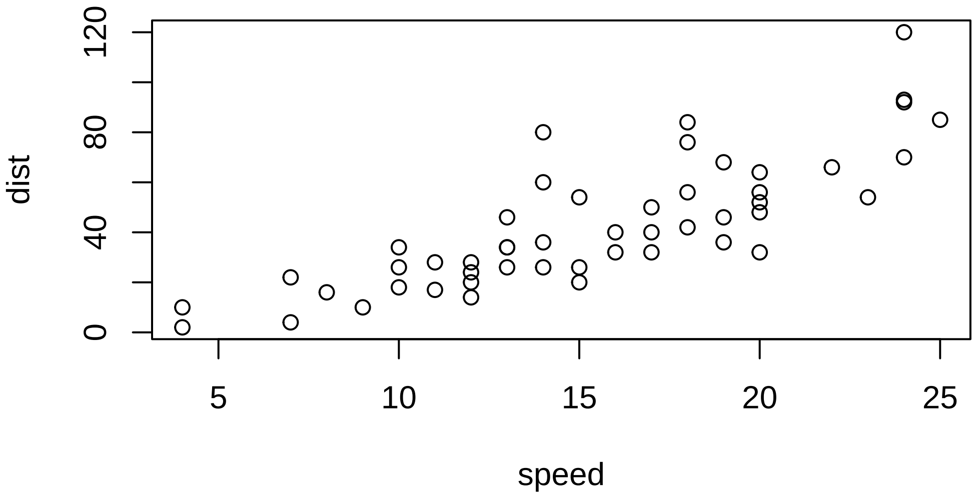
There are many graphic systems in R (ways to make plots and figures). These include "base" R plots. Today we will focus mos=tly on the **ggplot2** package

Lets start with a plot of a simple built in dataset called `cars`.

```
head(cars)
```

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

```
plot(cars)
```

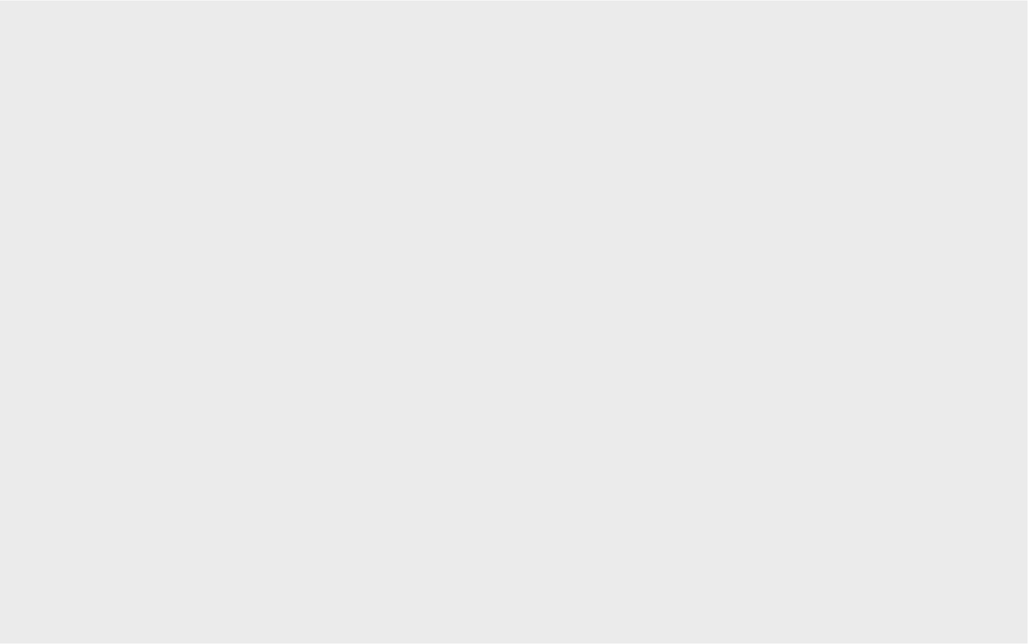


Let's see how we can make this figure using **ggplot**. First I need to install this package on my computer. To install any R package I use the function `install.packages()`

> I will run 'install.packages("ggplot2") in my R console not this quarto document so that it doesn't reinstall every time I render.

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

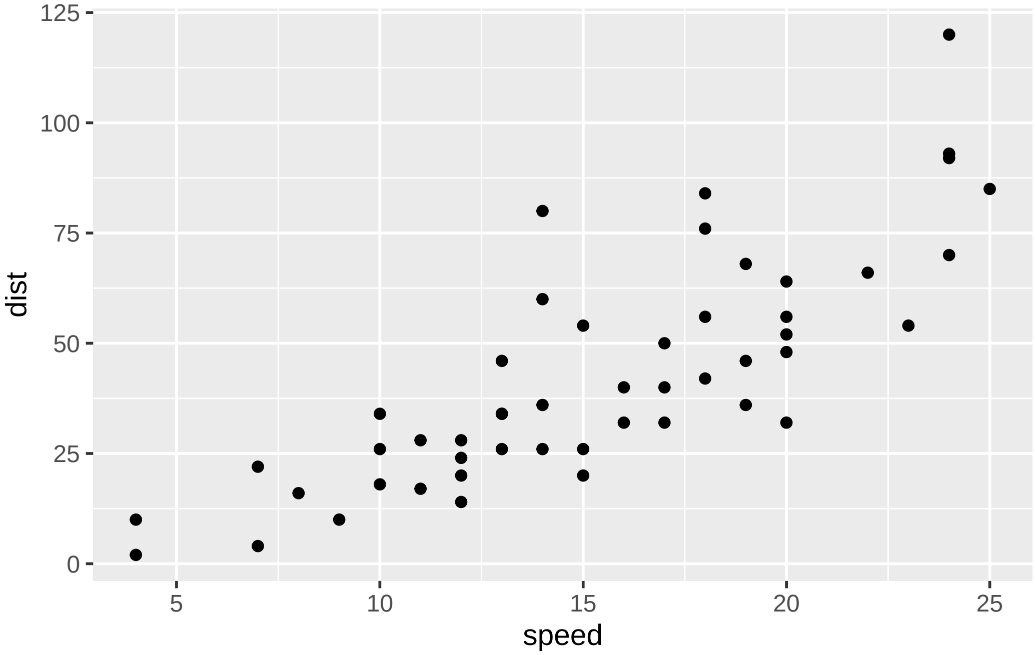
```
library(ggplot2)
ggplot(cars)
```



All ggplot 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() etc, that I want to draw)

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



Q. Which plot types are typically NOT used to compare distributions of numeric variables? **Network graphs**

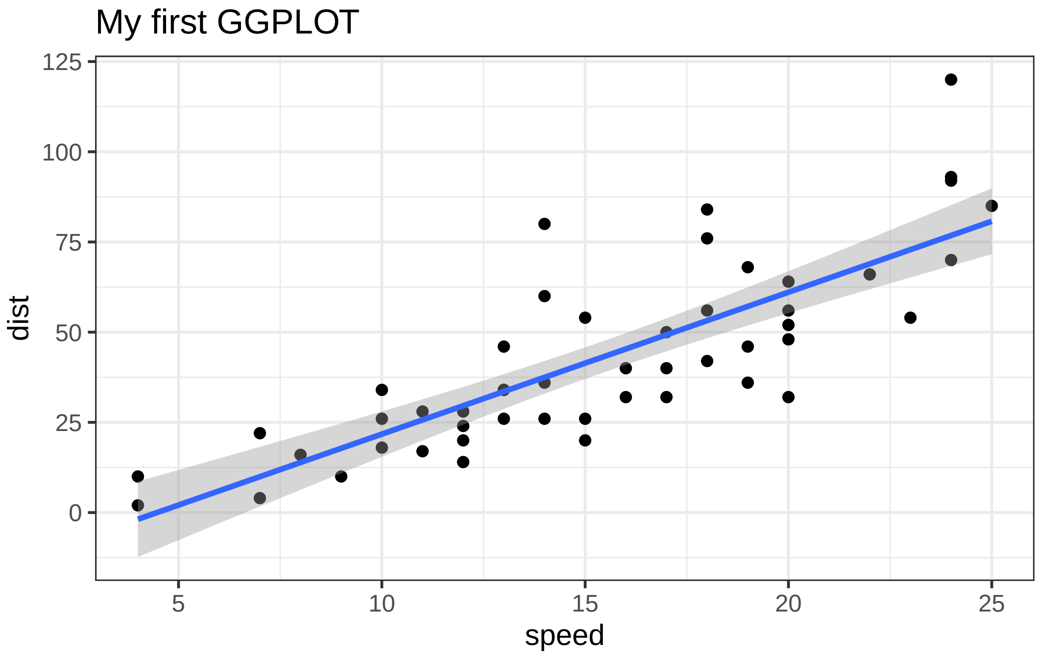
Q. Which statement about data visualization with ggplot2 is incorrect? **ggplot2 is the only way to create plots in R**

Q. Which geometric layer should be used to create scatter plots in ggplot2? **geom_point()**

Lets add a line to show the relationship here:

```r
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

The code to read the datatset

```
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. What is your answer? **5196**

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? **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? **127**

Q. Using your values above and 2 significant figures. What fraction of total genes is up-regulated in this dataset? **2.44**

how many genes are in this datset?

```
nrow(genes)
```

```
[1] 5196
```

```
ncol(genes)
```

```
[1] 4
```

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

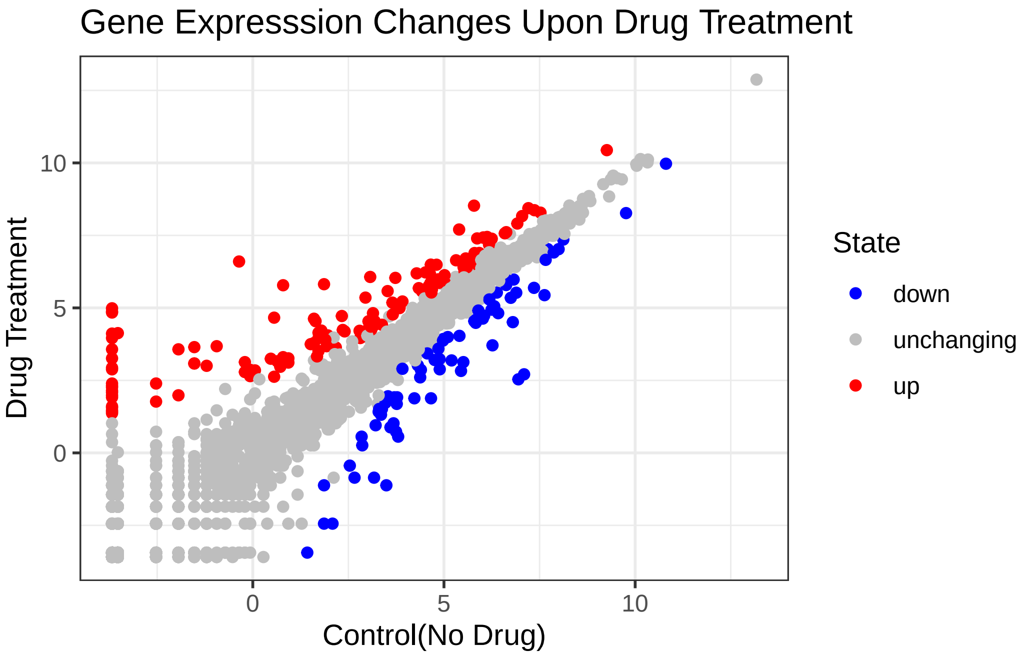
```

      down unchanging         up 
        72       4997        127 
```

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

```

      down unchanging         up 
      1.39      96.17       2.44 
```

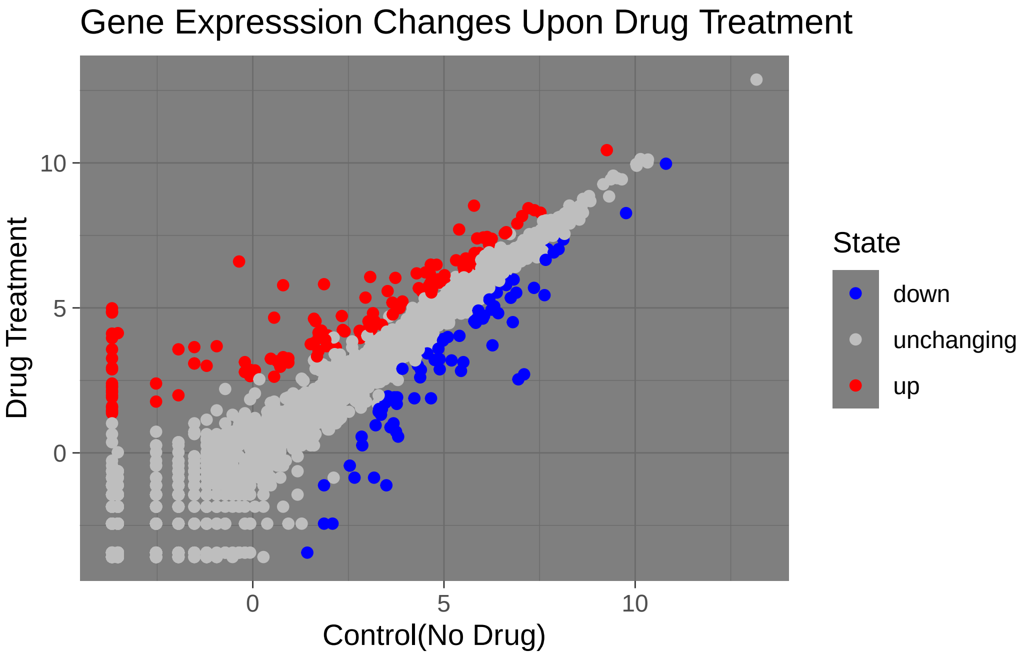
A first plot of this dataset

```
ggplot(genes)+
  aes(x=Condition1, y=Condition2, col=State) +
geom_point()+
  labs(title="Gene Expresssion Changes Upon Drug Treatment",
       x="Control(No Drug)",
       y="Drug Treatment") +
  theme_bw()+
  scale_colour_manual(values=c("blue","gray","red"))
```



```
p <- ggplot(genes)+
  aes(x=Condition1, y=Condition2, col=State) +
geom_point()+
  labs(title="Gene Expresssion Changes Upon Drug Treatment",
       x="Control(No Drug)",
       y="Drug Treatment") +
  theme_bw()+
  scale_colour_manual(values=c("blue","gray","red"))
```

```
p + theme_dark()
```



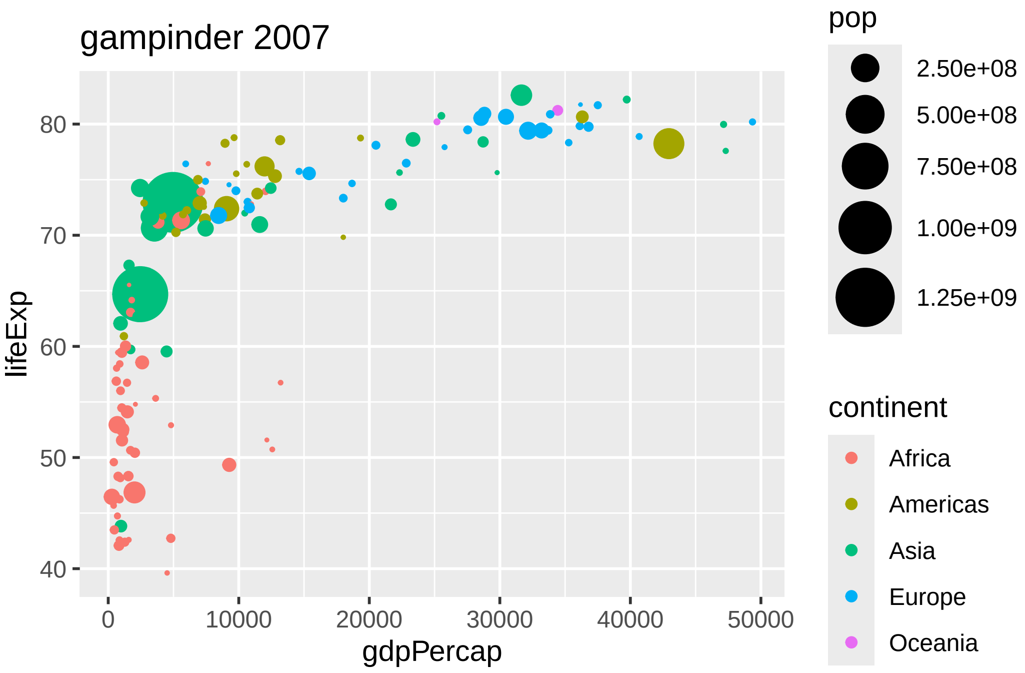
```r
# File location online
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv

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

```r
# install.packages("dplyr")  ## un-comment to install if needed
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)
```

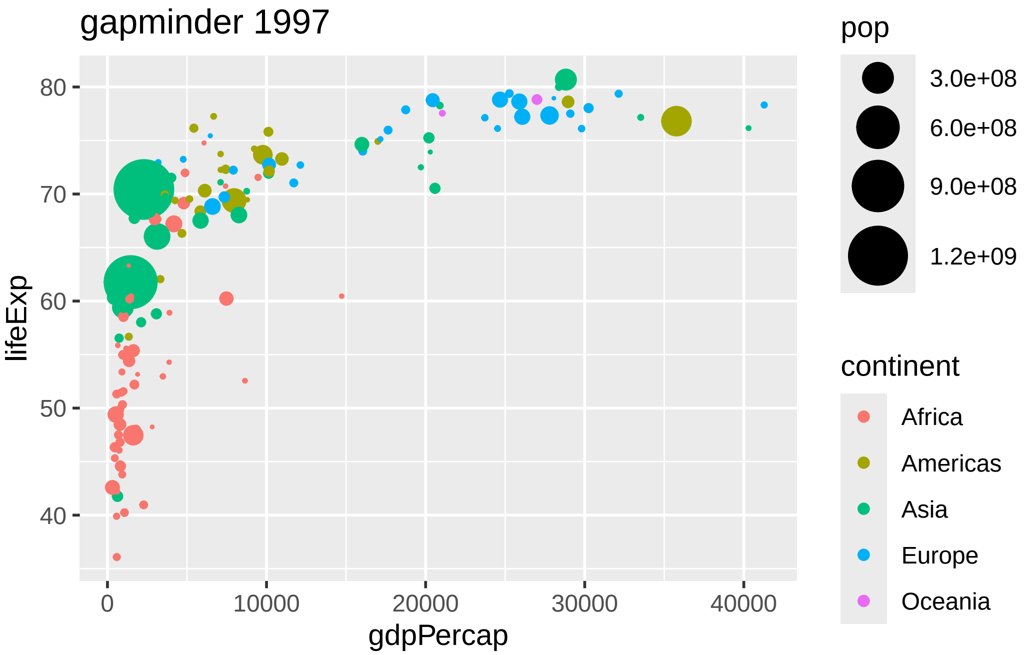
```
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp,color=continent, size=pop) +
  geom_point(alpha=1.0)+
scale_size_area(max_size = 10)+ labs(title= "gampinder 2007")
```



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

```
gapminder_1997 <- gapminder %>% filter(year==1997)
```

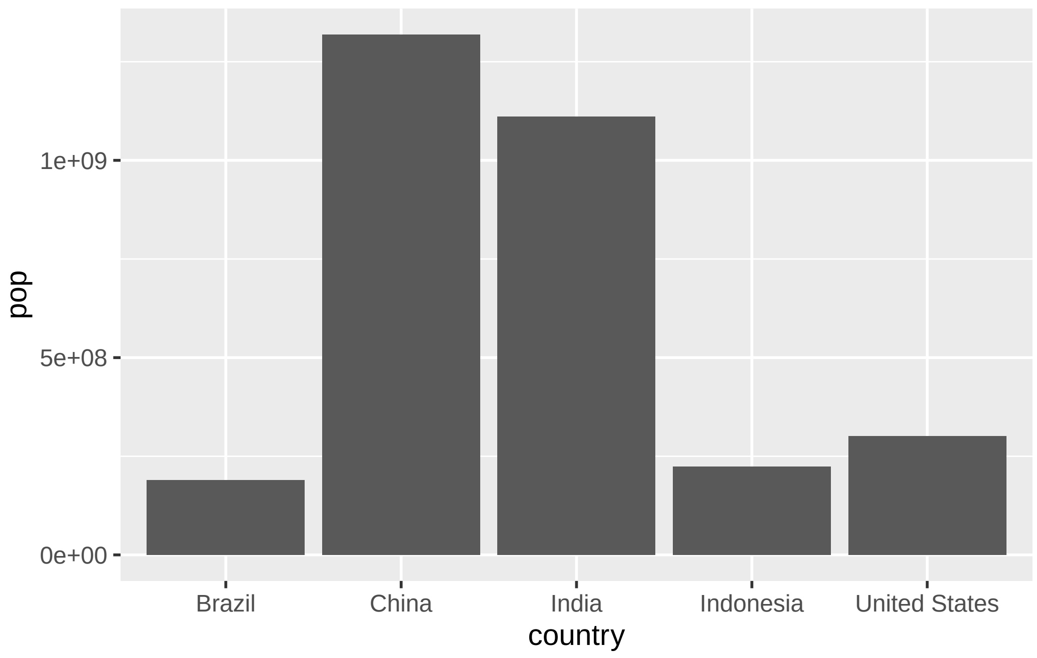
```
ggplot(gapminder_1997) +
  aes(x=gdpPercap, y=lifeExp,color=continent, size=pop) +
  geom_point(alpha=1.0)+
scale_size_area(max_size = 10)+ labs(title="gapminder 1997")
```



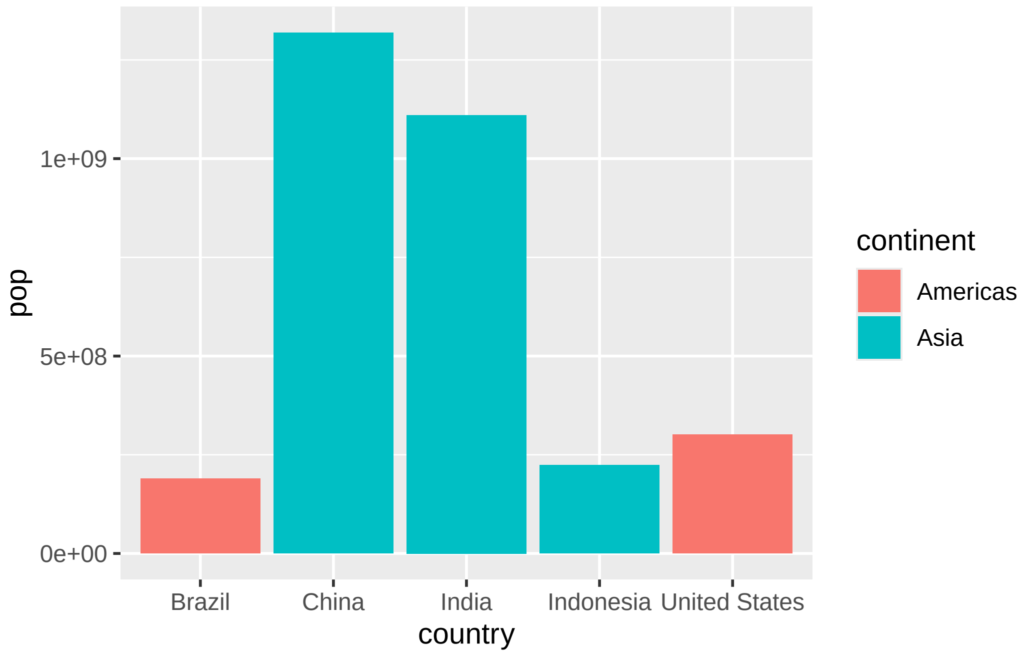
```
gapminder_top5 <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(pop)) %>%
  top_n(5, pop)

gapminder_top5
```

```
        country continent year lifeExp        pop gdpPercap
1         China      Asia 2007  72.961 1318683096  4959.115
2         India      Asia 2007  64.698 1110396331  2452.210
3 United States  Americas 2007  78.242  301139947 42951.653
4     Indonesia      Asia 2007  70.650  223547000  3540.652
5        Brazil  Americas 2007  72.390  190010647  9065.801
```

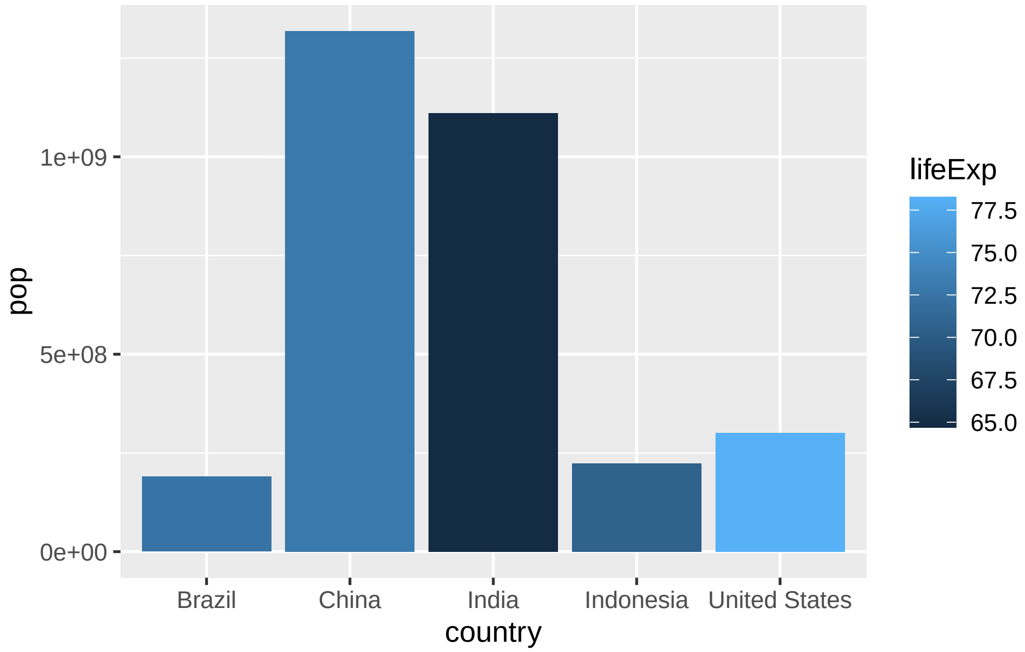
```
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop))
```



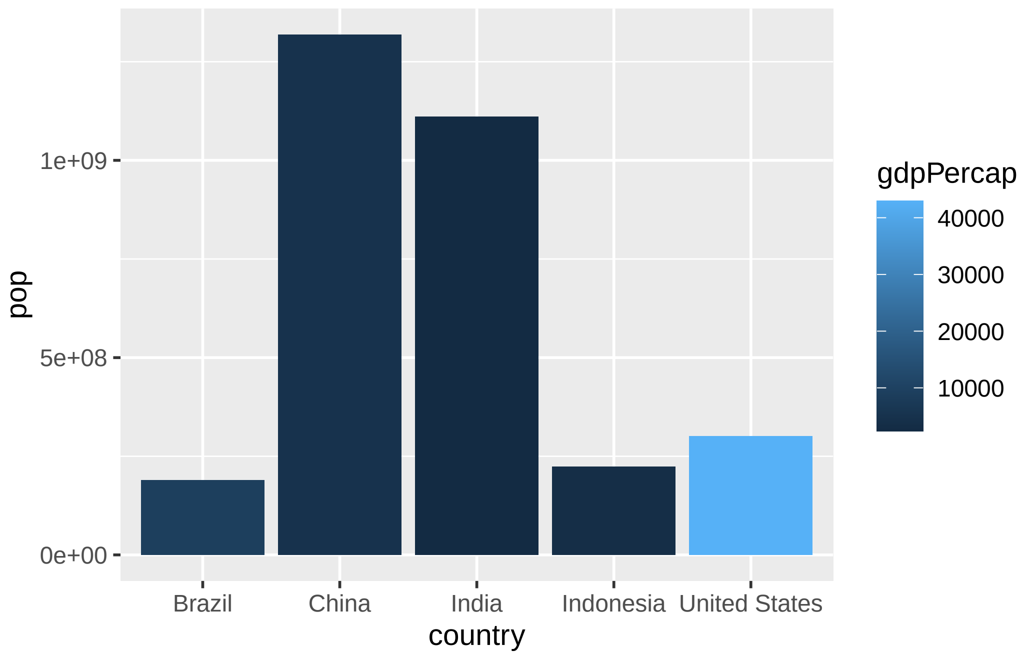
```r
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = continent))
```



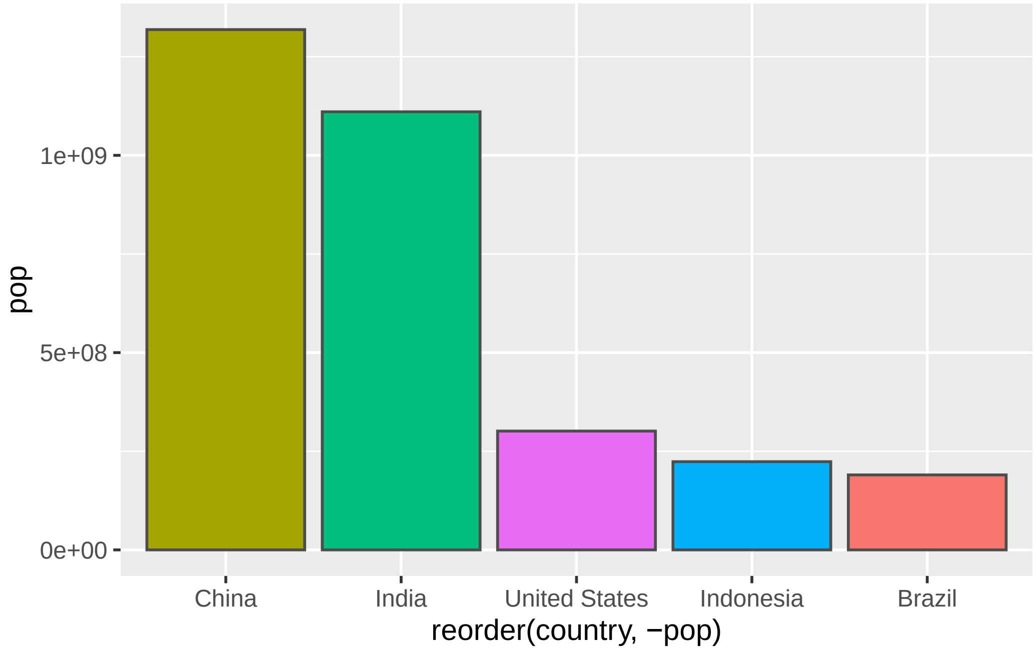
```
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = lifeExp))
```



```
ggplot(gapminder_top5) +
  aes(x=country, y=pop, fill=gdpPercap) +
  geom_col()
```



```
ggplot(gapminder_top5) +
  aes(x=reorder(country, -pop), y=pop, fill=country) +
  geom_col(col="gray30") +
  guides(fill="none")
```


1e+09
5e+08
0e+00
pop
China
India
United States
Indonesia
Brazil
reorder(country, −pop)