class05: Data Visualization with ggplot

Jiachen Fan (A17662703)

Use ggplot

To use ggplot we first need to install it. Use 'install.packages()'. Before I use any package functions I have to load them up with a 'library()'.

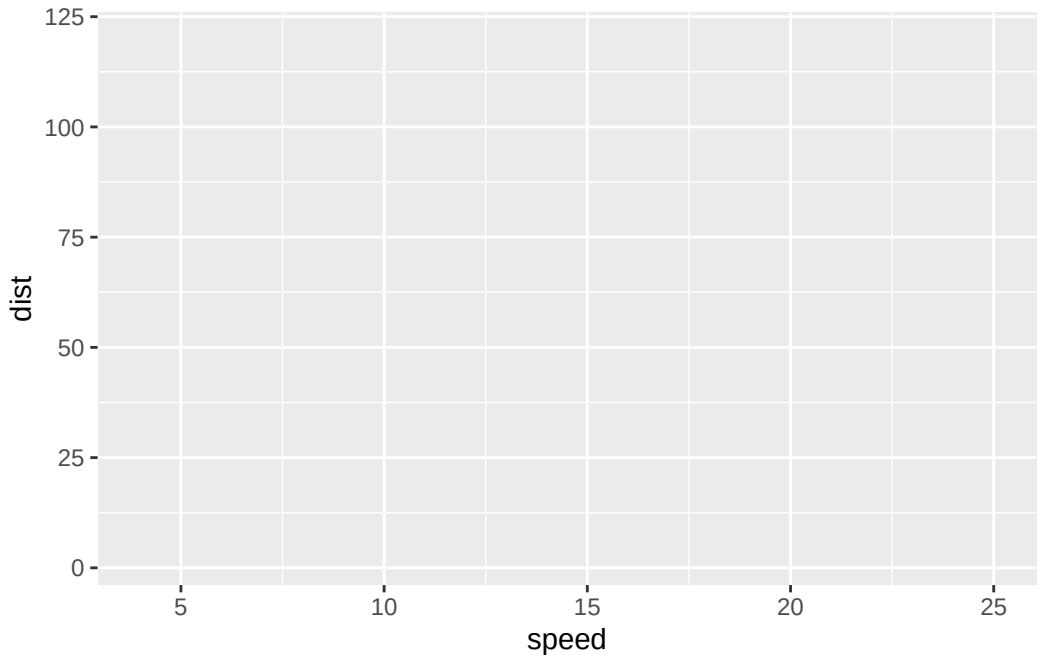
```
library(ggplot2)

ggplot(cars)
```



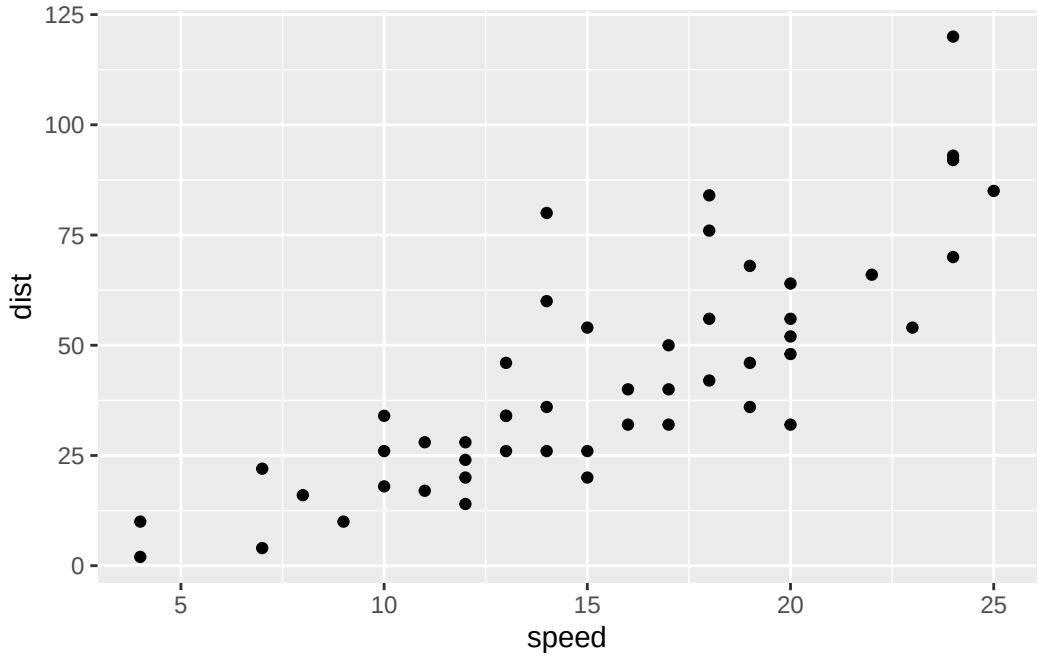
```
tail(cars,6)
```

```
   speed dist
45    23   54
46    24   70
47    24   92
48    24   93
49    24  120
50    25   85
```

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



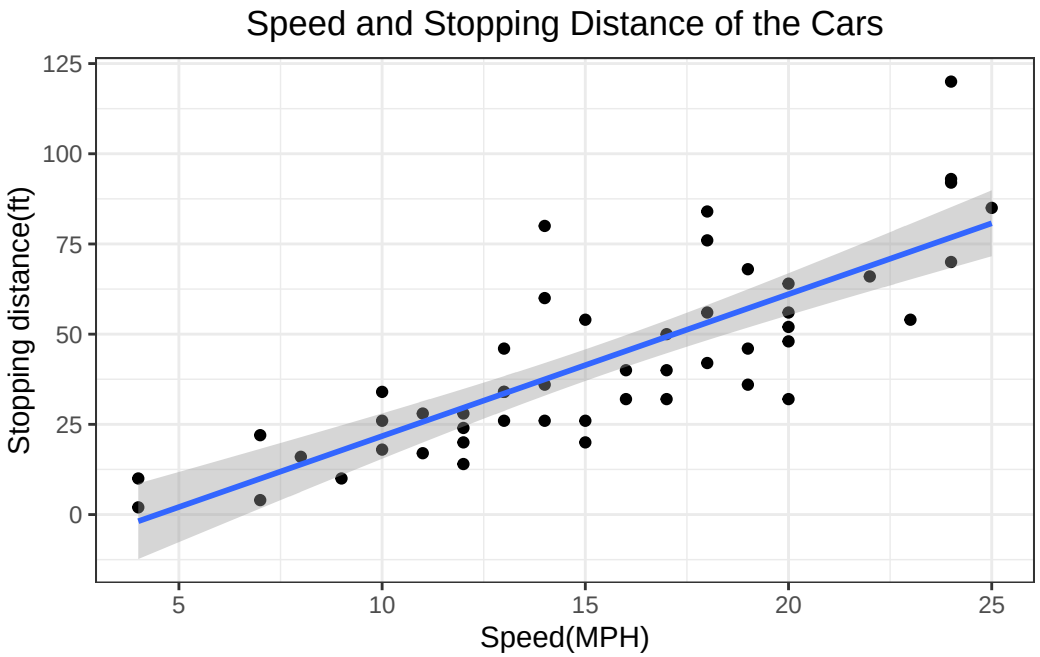

```
ggplot(cars) +
  aes(x=speed, y=dist) +
  geom_point()
```



To use ggplot, three things matter -data(data.frame) -aesthetics (how the data map to the plot) -geoms (how I want things drawn)

```
ggplot(cars)+
  aes(x=speed, y=dist)+
  labs(title ='Speed and Stopping Distance of the Cars',
       x = 'Speed(MPH)',
       y = 'Stopping distance(ft)') +
  geom_point()+
  geom_smooth(method ='lm') +
  theme_bw() +
  theme(plot.title = element_text(hjust=0.5))
```

```
`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
```

```
nrow(genes)
```

```
[1] 5196
```

```
colnames(genes)
```

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

```r
ncol(genes)
```

```
[1] 4
```

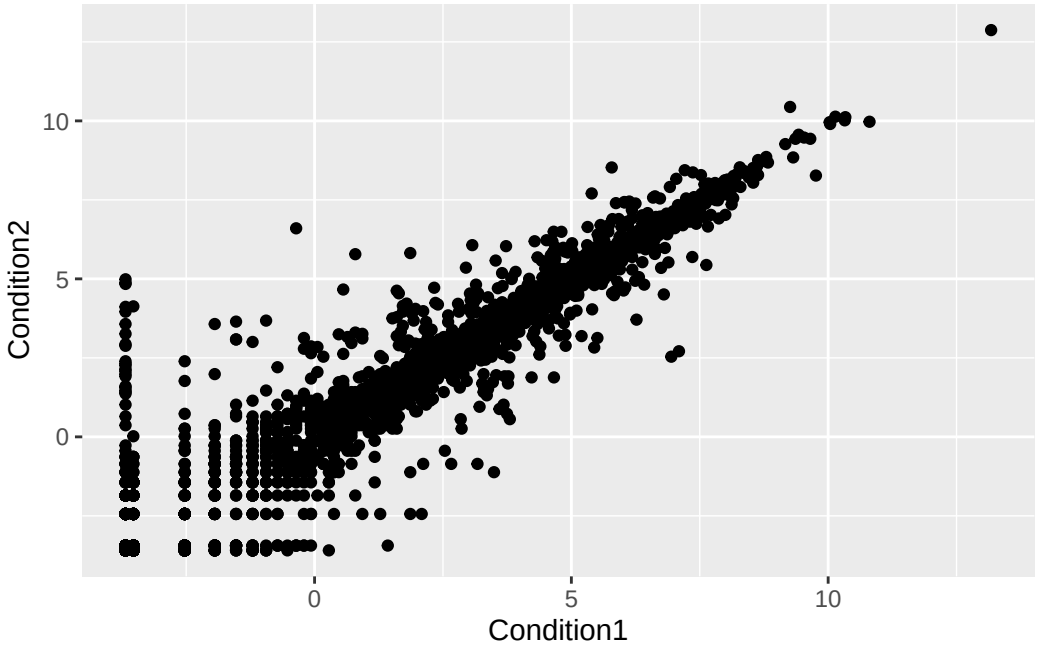
```r
table(genes$State)
```

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

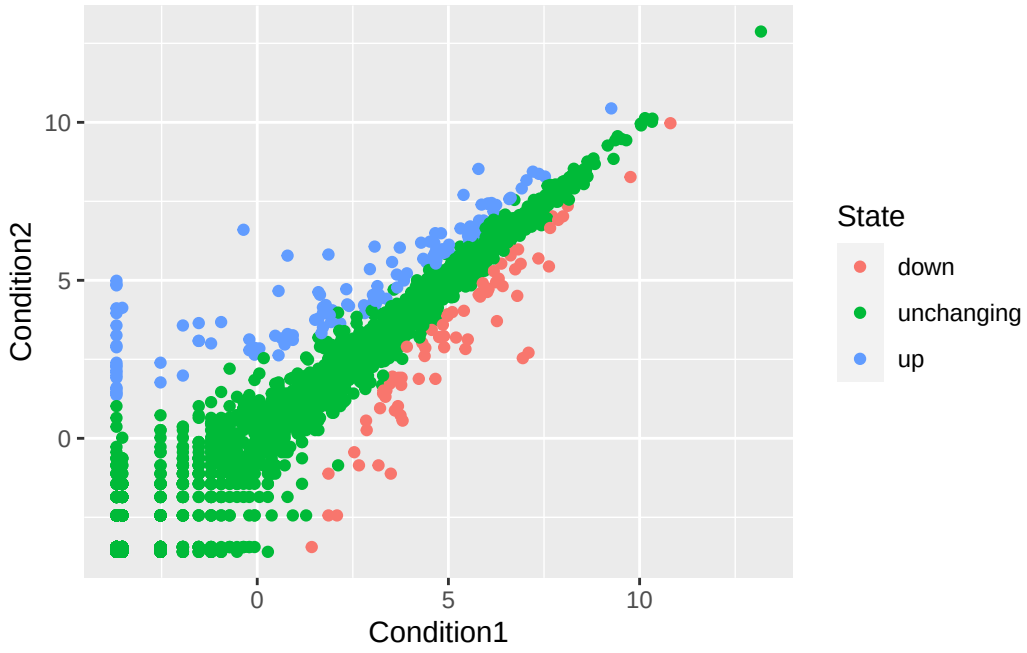
```r
round(table(genes$State)/nrow(genes)*100,2)
```

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

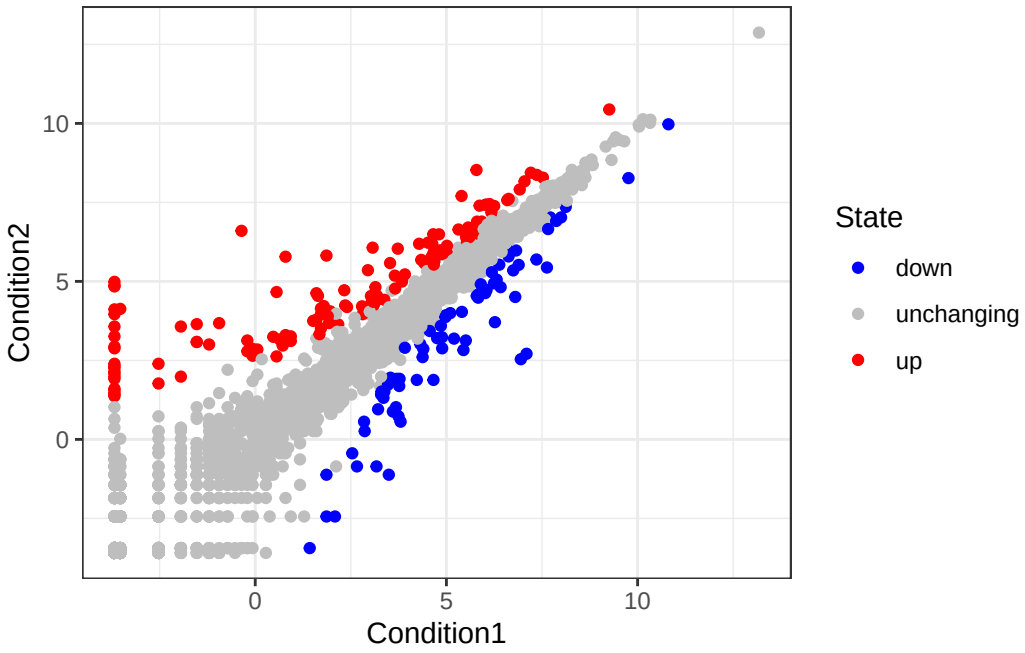
```r
ggplot(genes)+
  aes(x= Condition1, y=Condition2) +
  geom_point()
```



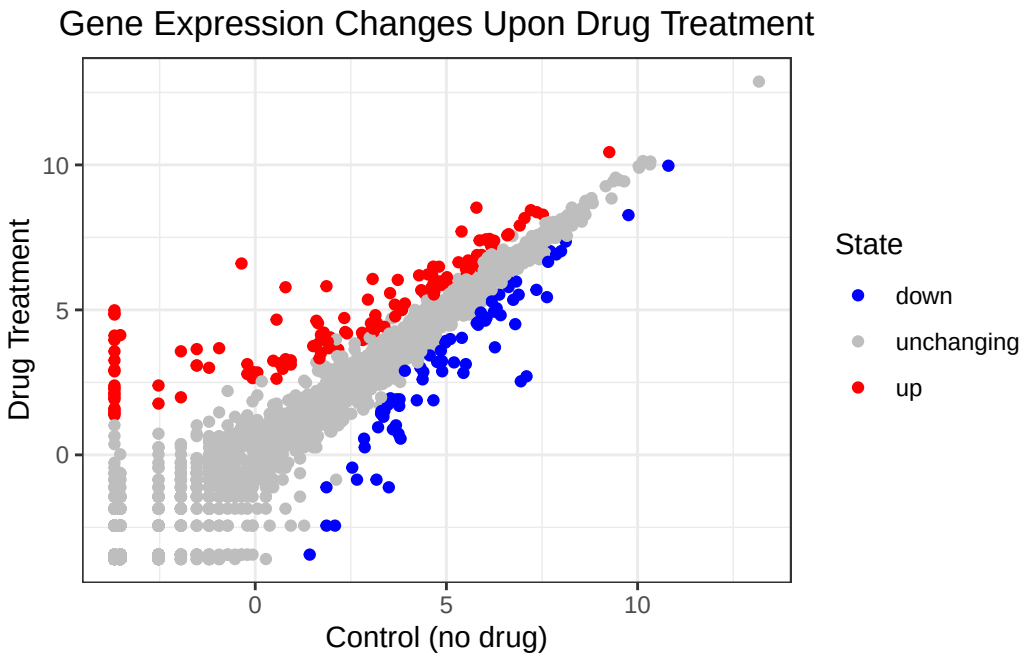
```
p <- ggplot(genes) +
    aes(x=Condition1, y=Condition2, col=State) +
    geom_point()
p
```




```
p + scale_colour_manual( values=c("blue","gray","red") )+
  theme_bw()
```




```
p + scale_colour_manual( values=c("blue","gray","red") )+
  labs(title = 'Gene Expression Changes Upon Drug Treatment',
       x = 'Control (no drug)',
       y = 'Drug Treatment') +
  theme_bw() +
  theme(plot.title = element_text(hjust=0.5))
```



7.Going Further

```
library(gapminder)
```

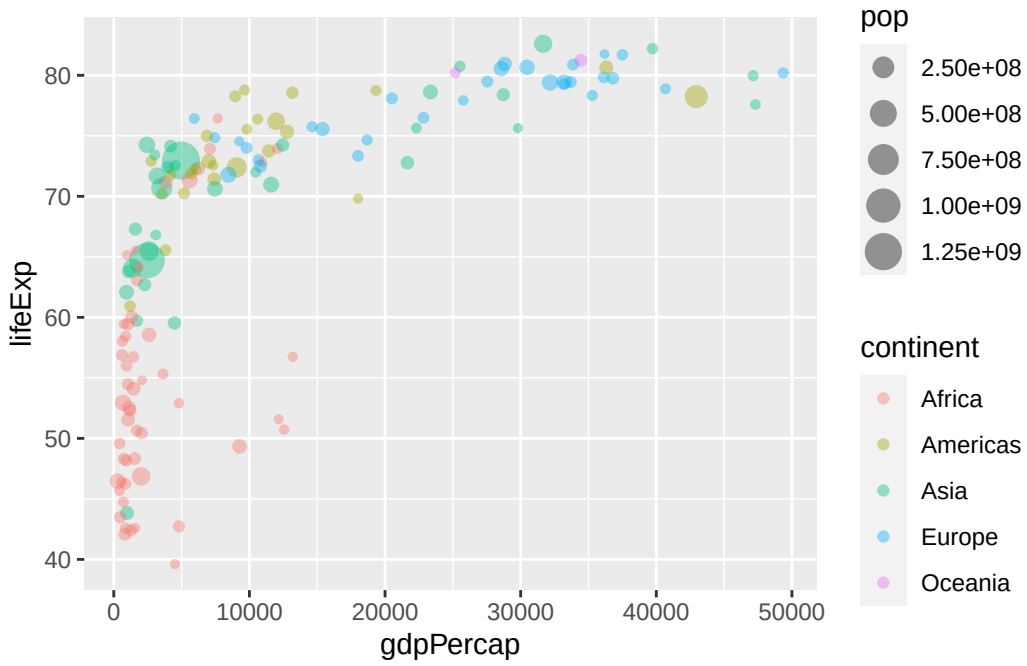
```
# File location online
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.

gapminder <- read.delim(url)

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

```
Attaching package: 'dplyr'

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

    filter, lag
```

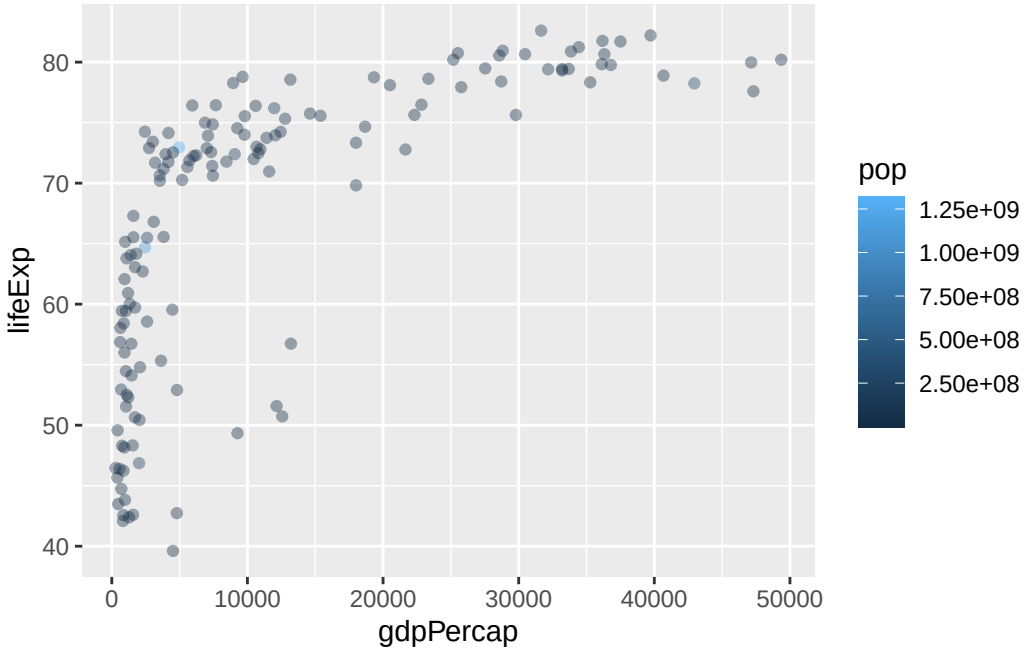
```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

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

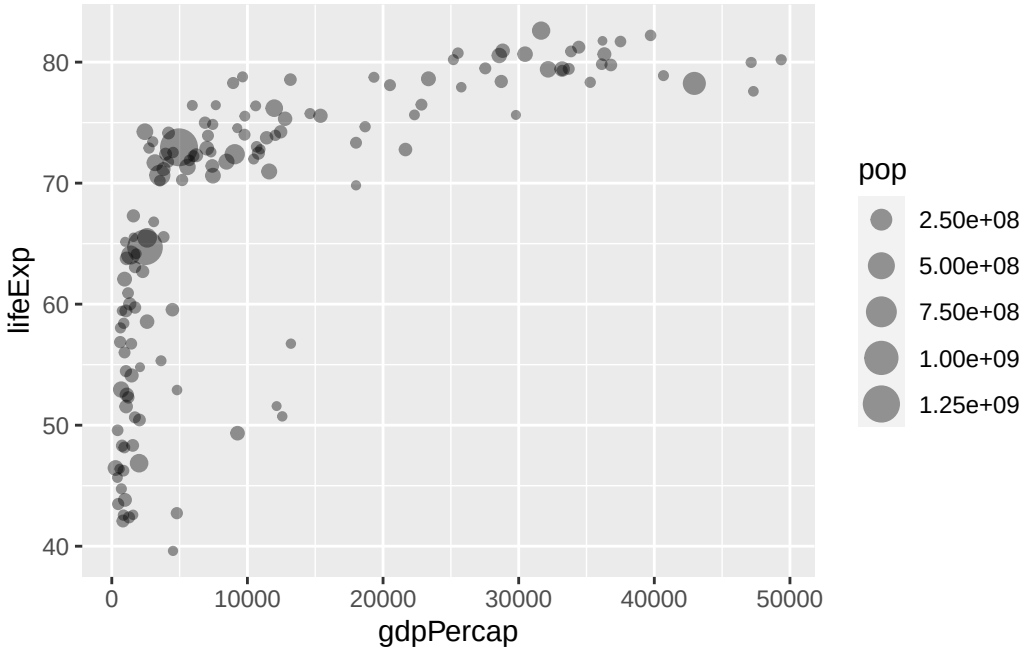
```
ggplot(gapminder_2007)+
  aes(x = gdpPercap, y = lifeExp, color=continent, size=pop) +
  geom_point(alpha=0.4)
```



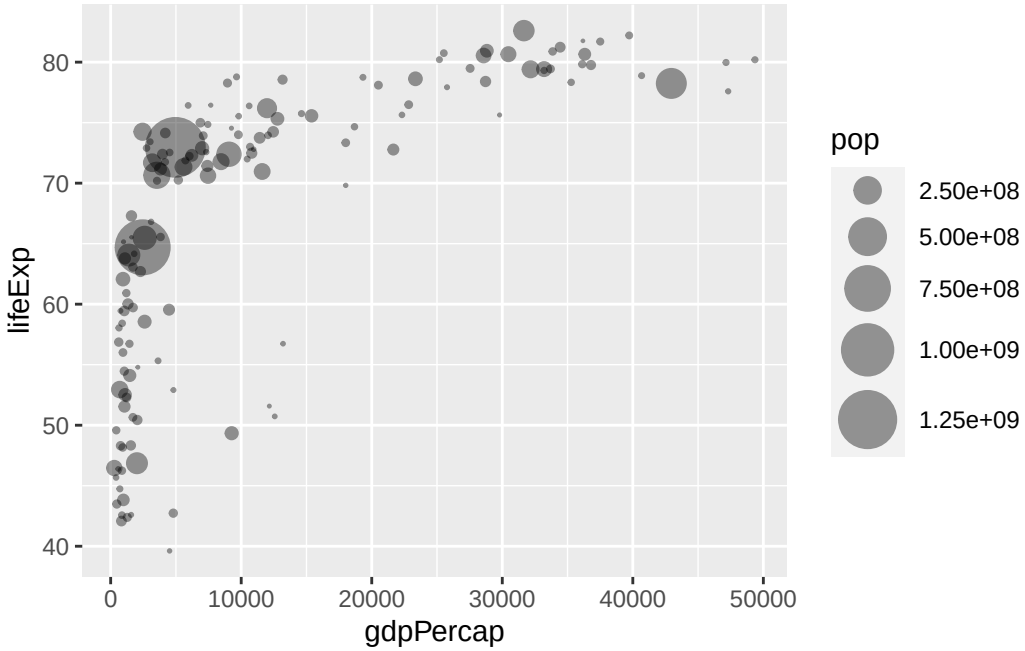

```
ggplot(gapminder_2007)+
  aes(x = gdpPercap, y = lifeExp, color=pop) +
  geom_point(alpha=0.4)
```



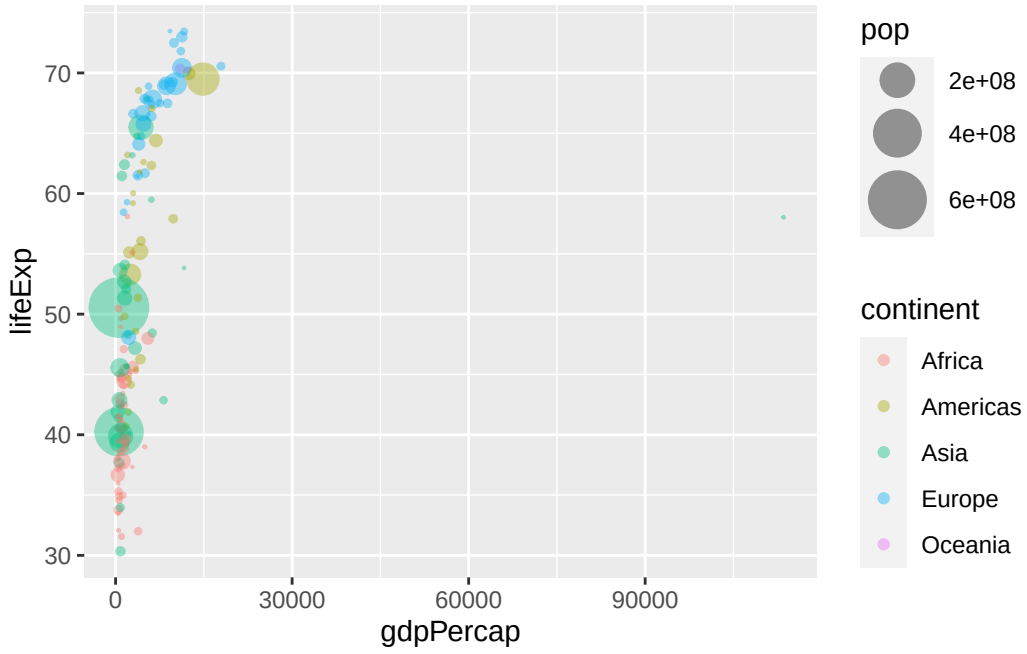

```
q <- ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp, size = pop) +
  geom_point(alpha=0.4)
q
```



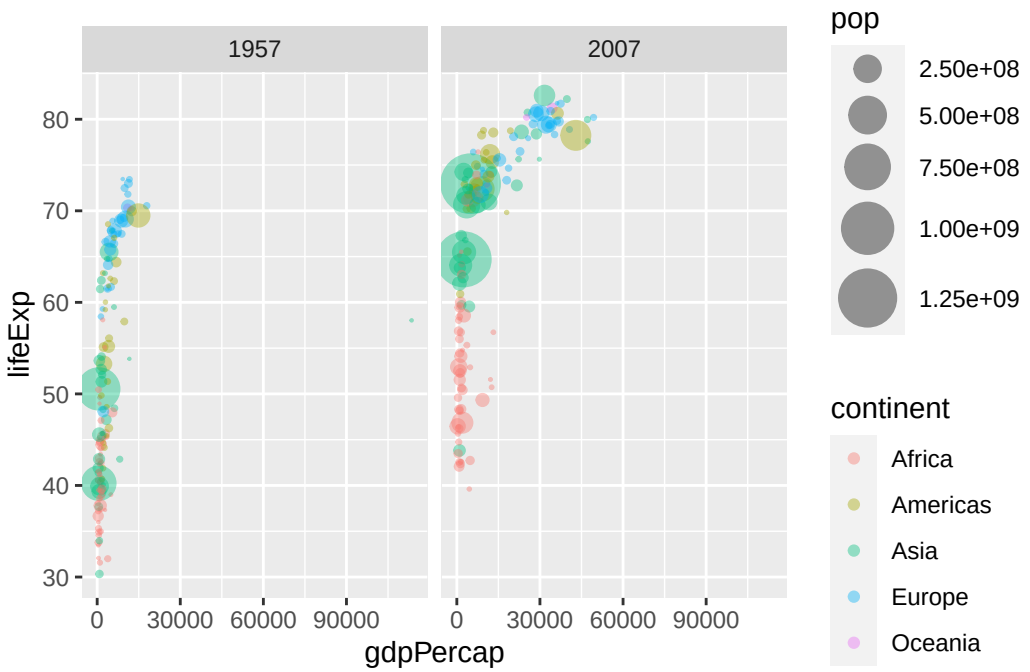

```
q +  scale_size_area(max_size = 10)
```



```
gapminder_1957 <- gapminder %>% filter(year==1957)
ggplot(gapminder_1957)+
  aes(x = gdpPercap, y = lifeExp, color = continent, size = pop) +
  geom_point(alpha = 0.4) +
  scale_size_area(max_size = 10)
```



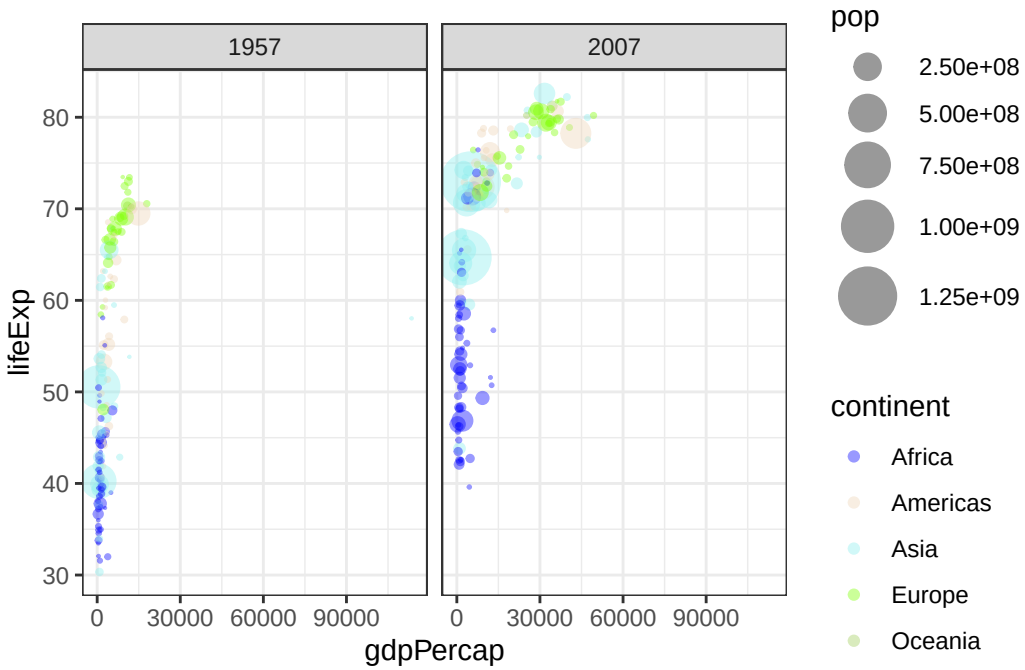

```
gapminder_1957_2007 <- gapminder %>% filter(year==1957|year==2007)
s <-ggplot(gapminder_1957_2007)+
  aes(x = gdpPercap, y = lifeExp, color = continent, size = pop) +
  geom_point(alpha = 0.4) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
s
```




```
table(gapminder$continent)
```

```
Africa Americas     Asia   Europe  Oceania
   624      300      396      360       24
```

```
s + scale_colour_manual( values=c("blue","bisque2",
                                  "darkslategray2","chartreuse1","darkolivegreen3") ) +
  theme_bw()
```

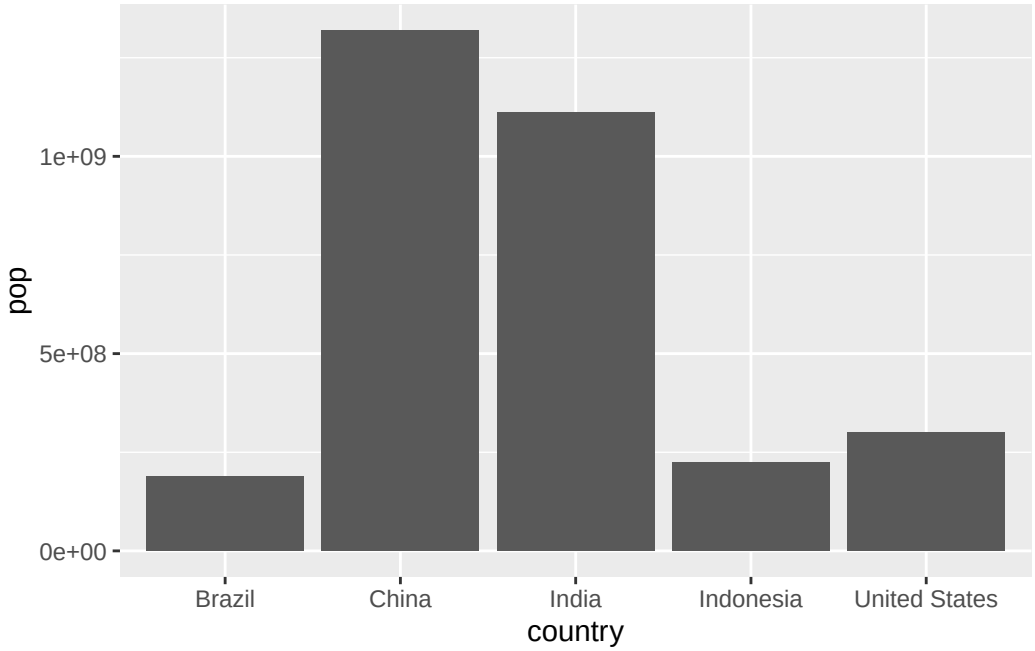


8.Bar Charts

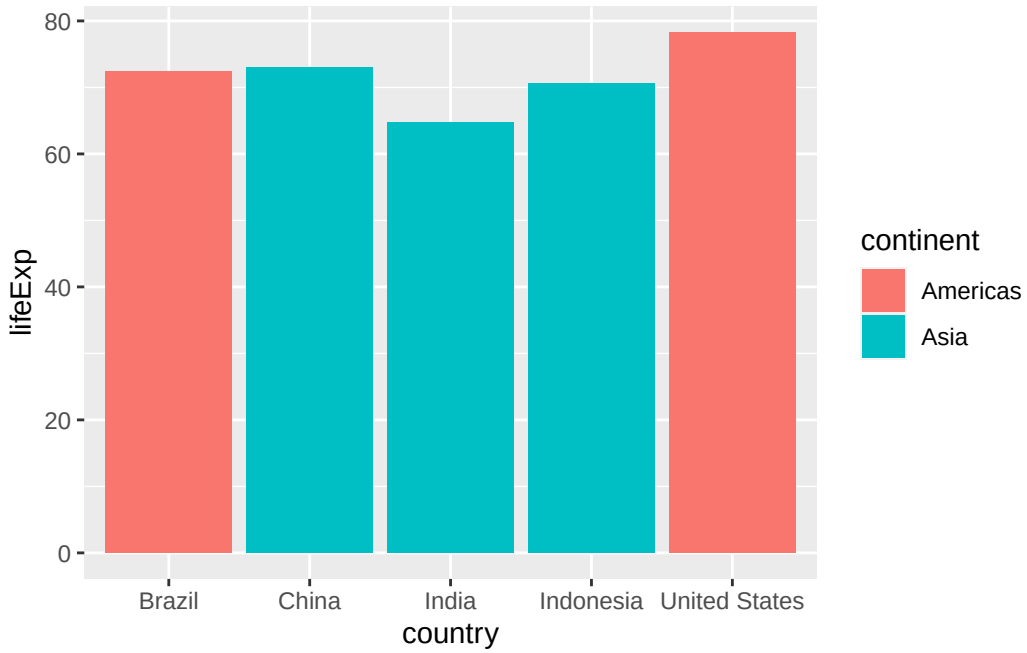
```
gapminder_top5 <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(pop)) %>%
  top_n(5, pop)

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
```

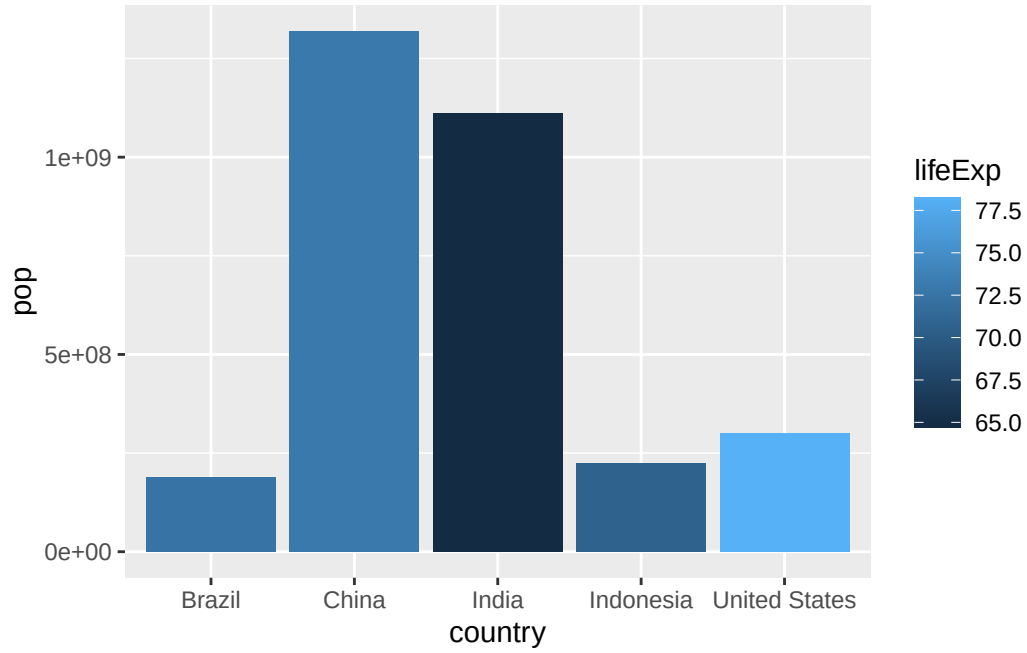
```
ggplot(gapminder_top5) +
  aes(x = country, y = pop) +
  geom_col()
```



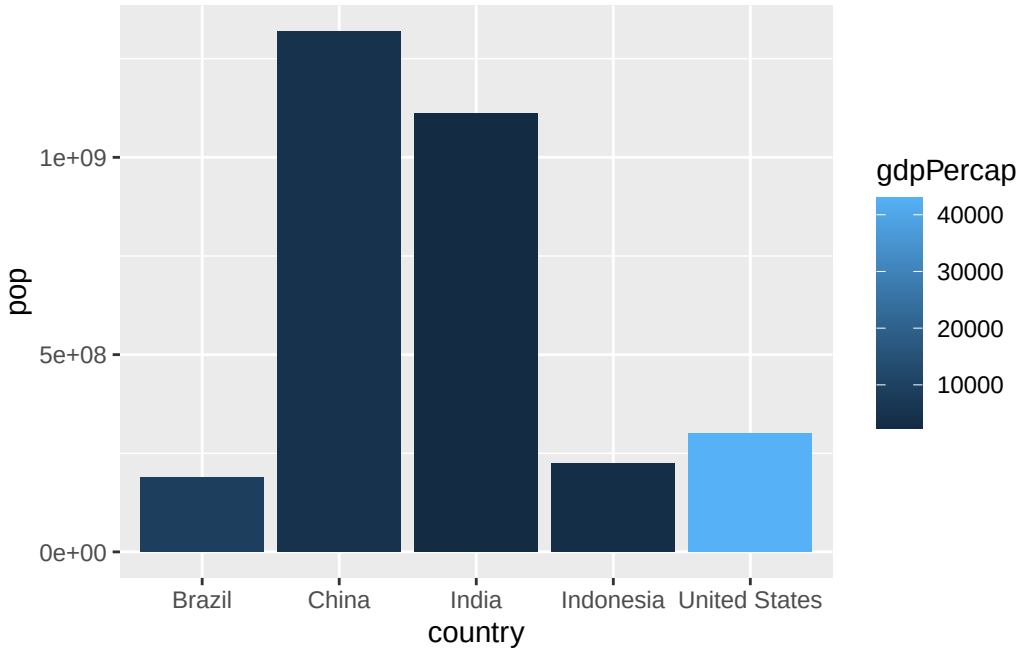

```
ggplot(gapminder_top5) +
  aes(x = country, y = lifeExp, fill = continent) +
  geom_col()
```



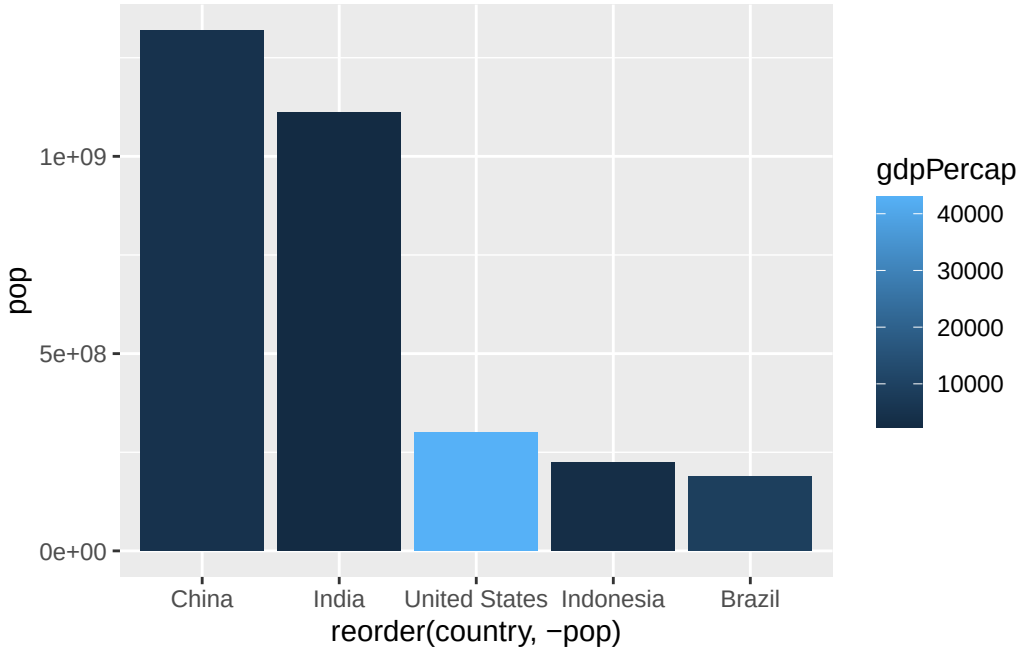

```
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = lifeExp))
```



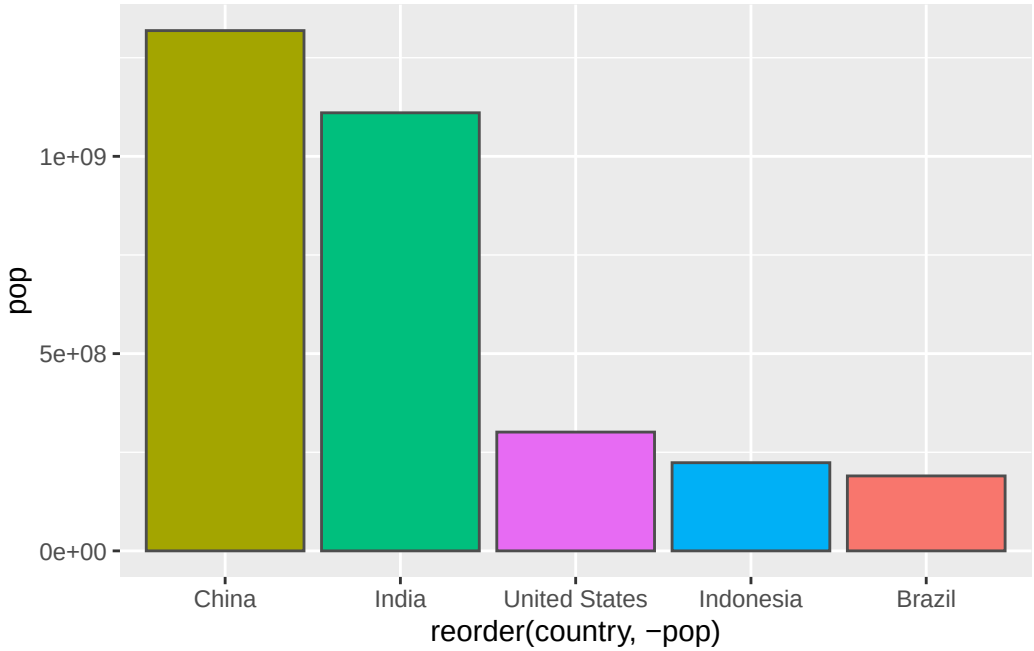
```
ggplot(gapminder_top5) +
  aes(x=country, y=pop, fill=gdpPercap) +
  geom_col()
```




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



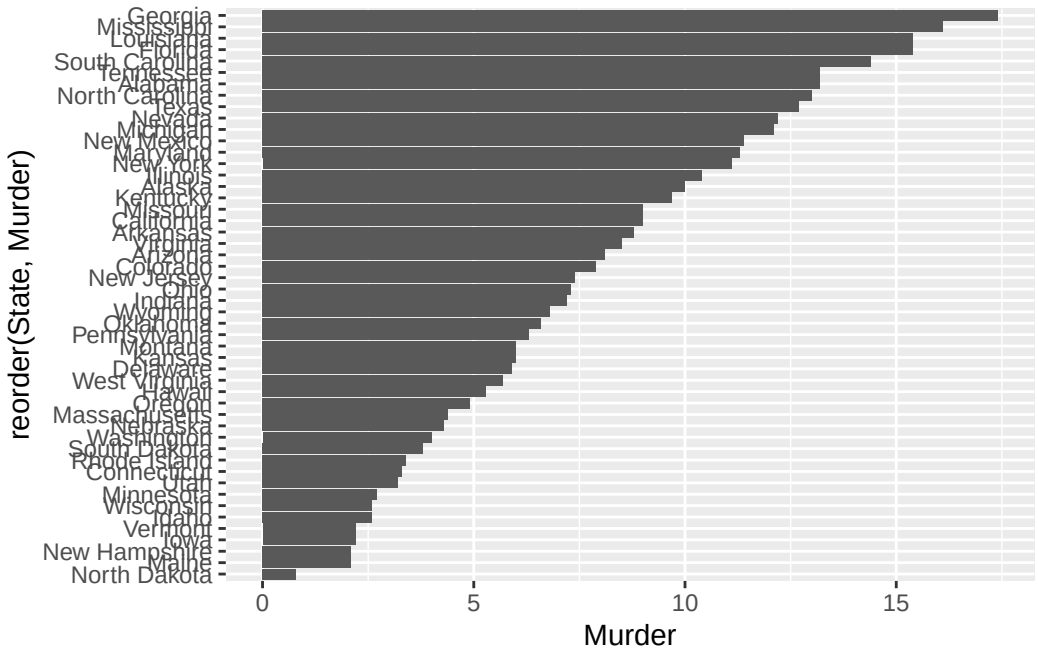

```
ggplot(gapminder_top5) +
  aes(x=reorder(country, -pop), y=pop, fill=country) +
  geom_col(col="gray30") +
  guides(fill="none")
```



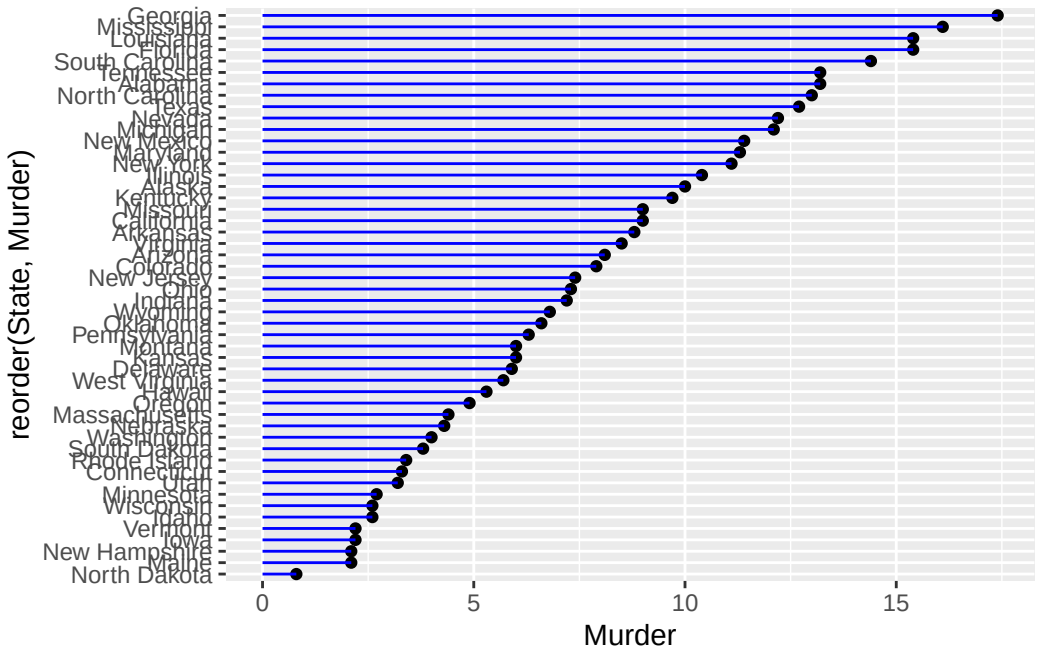

```
head(USArrests)
```

```
           Murder Assault UrbanPop Rape
Alabama      13.2     236       58 21.2
Alaska       10.0     263       48 44.5
Arizona       8.1     294       80 31.0
Arkansas      8.8     190       50 19.5
California    9.0     276       91 40.6
Colorado      7.9     204       78 38.7
```

```
USArrests$State <- rownames(USArrests)
ggplot(USArrests) +
  aes(x=reorder(State,Murder), y=Murder) +
  geom_col() +
  coord_flip()
```




```
ggplot(USArrests) +
  aes(x=reorder(State,Murder), y=Murder) +
  geom_point() +
  geom_segment(aes(x=State,
                   xend=State,
                   y=0,
                   yend=Murder), color="blue") +
  coord_flip()
```



9.Extensions: Animation

First, install.packages("gifski") and install.packages("gganimate") Second, use '#| eval:false' to skip the execution of animation.

```
library(gapminder)
library(gganimate)

ggplot(gapminder, aes(gdpPercap, lifeExp, size = pop, colour = country)) +
  geom_point(alpha = 0.7, show.legend = FALSE) +
  scale_colour_manual(values = country_colors) +
  scale_size(range = c(2, 12)) +
  scale_x_log10() +
  facet_wrap(~continent) +
  # gganimate
  labs(title = 'Year: {frame_time}', x = 'GDP per capita', y = 'life expectancy') +
  transition_time(year) +
  shadow_wake(wake_length = 0.1, alpha = FALSE)
```

10.Combining plots

Install 'patchwork'

```
library(patchwork)

p1 <- ggplot(mtcars) + geom_point(aes(mpg, disp))
p2 <- ggplot(mtcars) + geom_boxplot(aes(gear, disp, group = gear))
p3 <- ggplot(mtcars) + geom_smooth(aes(disp, qsec))
p4 <- ggplot(mtcars) + geom_bar(aes(carb))

(p1 | p2 | p3) /
      p4
```

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