

week05.R

HP

2022-02-02

```
# Week 5 Data Visualization Lab
```

```
# Install the package ggplot2
```

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

```
#Anytime I want to use this package
```

```
# I need to load it
```

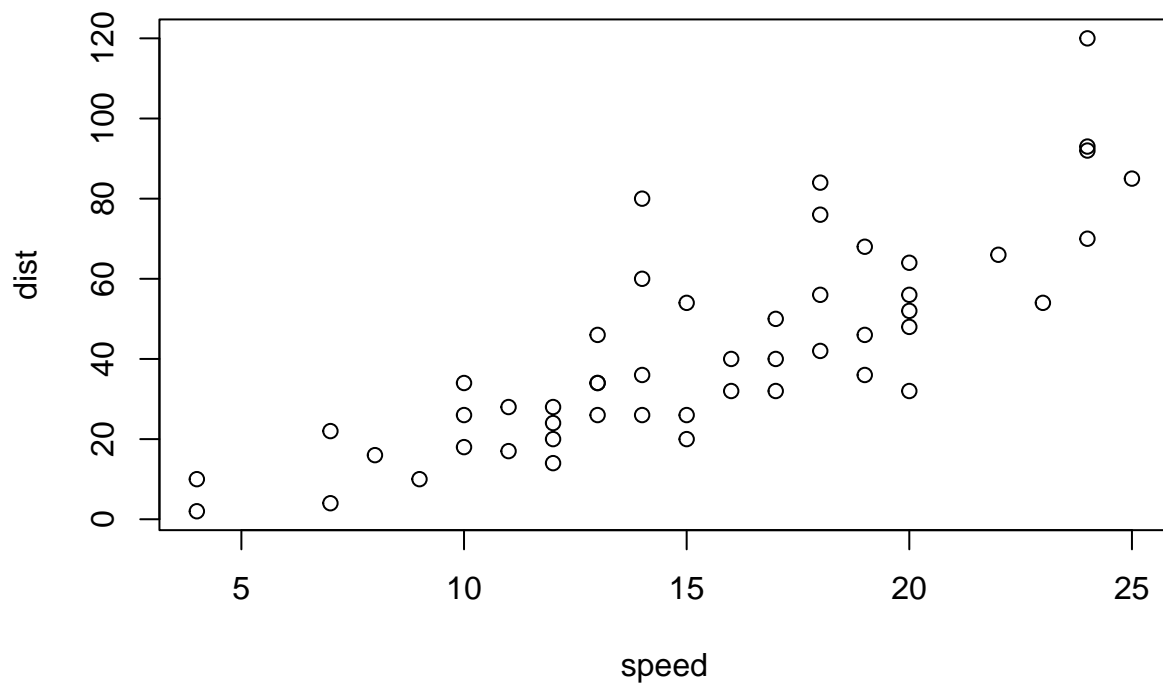
```
library(ggplot2)
```

```
## Warning in register(): Can't find generic 'scale_type' in package ggplot2 to  
## register S3 method.
```

```
View(cars)
```

```
# Make a quick base R plot- this is not a ggplot
```

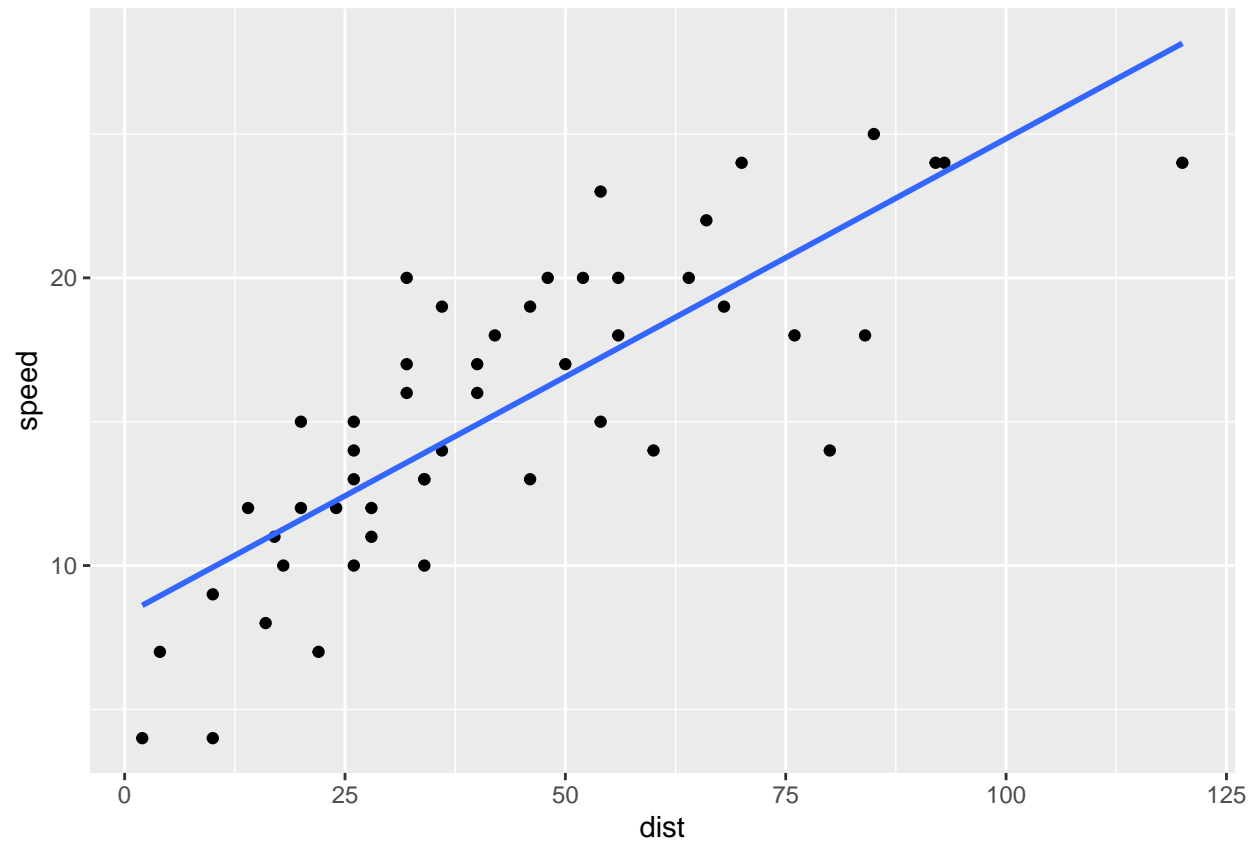
```
plot(cars)
```



```
# Our first ggplot
# we need data + aes + geoms
p <- ggplot(data= cars) +
  aes(x=dist, y=speed) +
  geom_point()

#One last thing. let's add a line to the data
p + geom_smooth(method= "lm", se= FALSE)
```

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

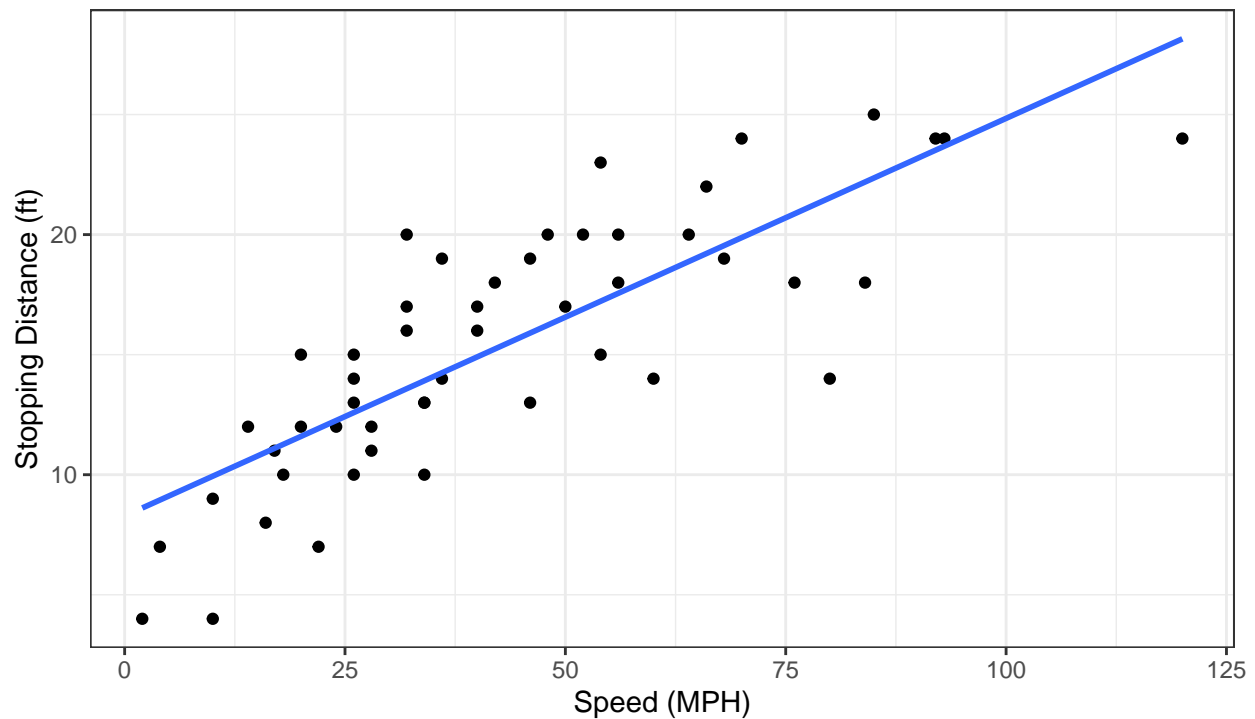


```
# Add lab fxn and bw theme
p + labs(x= "Speed (MPH)" , y= "Stopping Distance (ft)" ,
         title= "Speed and Stopping Distance" ,
         subtitle = "Analyzing correlation" ,
         caption = "Dataset: Cars") +
  geom_smooth(method= "lm", se = FALSE) +
  theme_bw()
```

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

Speed and Stopping Distance

Analyzing correlation



Dataset: Cars

```
# RNAseq experiment dataset
```

```
# Read the data into R
```

```
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 in dataset?
```

```
nrow(genes)
```

```
## [1] 5196
```

```
# Column genes and how many?
```

```
colnames(genes)
```

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

```
ncol(genes)

## [1] 4

# how many "up" regulated genes?
table(genes$State)

##
##      down  unchanged      up
##      72      4997      127

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

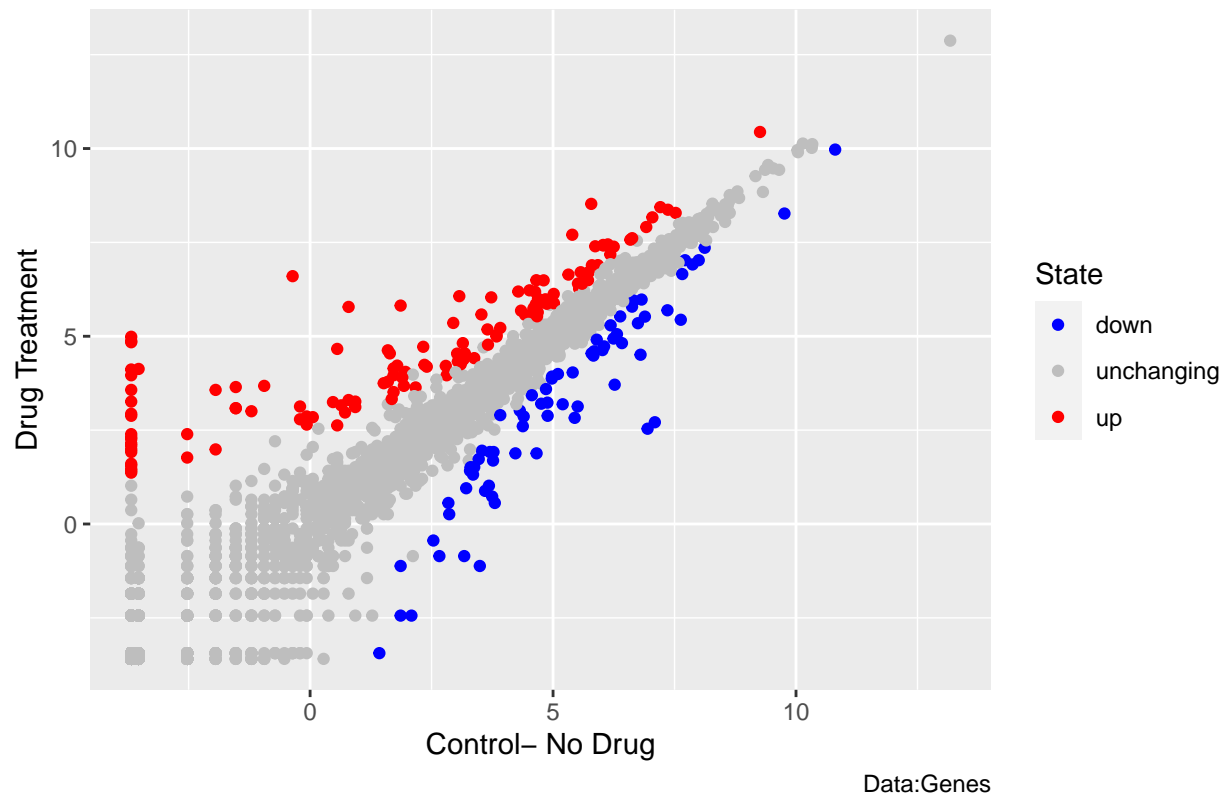
##
##      down  unchanged      up
##      1.39      96.17      2.44

# Make a ggplot
ggp <- ggplot(genes) + aes(x= Condition1, y= Condition2) +
  geom_point()

#Make a ggplot w/ color
ggp2 <- ggplot(genes) + aes(x= Condition1, y= Condition2, col= State) +
  geom_point() +
  labs(x = "Control- No Drug", y = "Drug Treatment",
       title = "Gene Expression Changes Upon Drug Treatment",
       caption = "Data:Genes")

#Change color
ggp2 + scale_color_manual(values= c("blue", "gray", "red"))
```

Gene Expression Changes Upon Drug Treatment



```
#Optional gapminder
library(gapminder)

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

gapminder <- read.delim(url)

# install.packages("dplyr")
library(dplyr)

##
## Attaching package: 'dplyr'

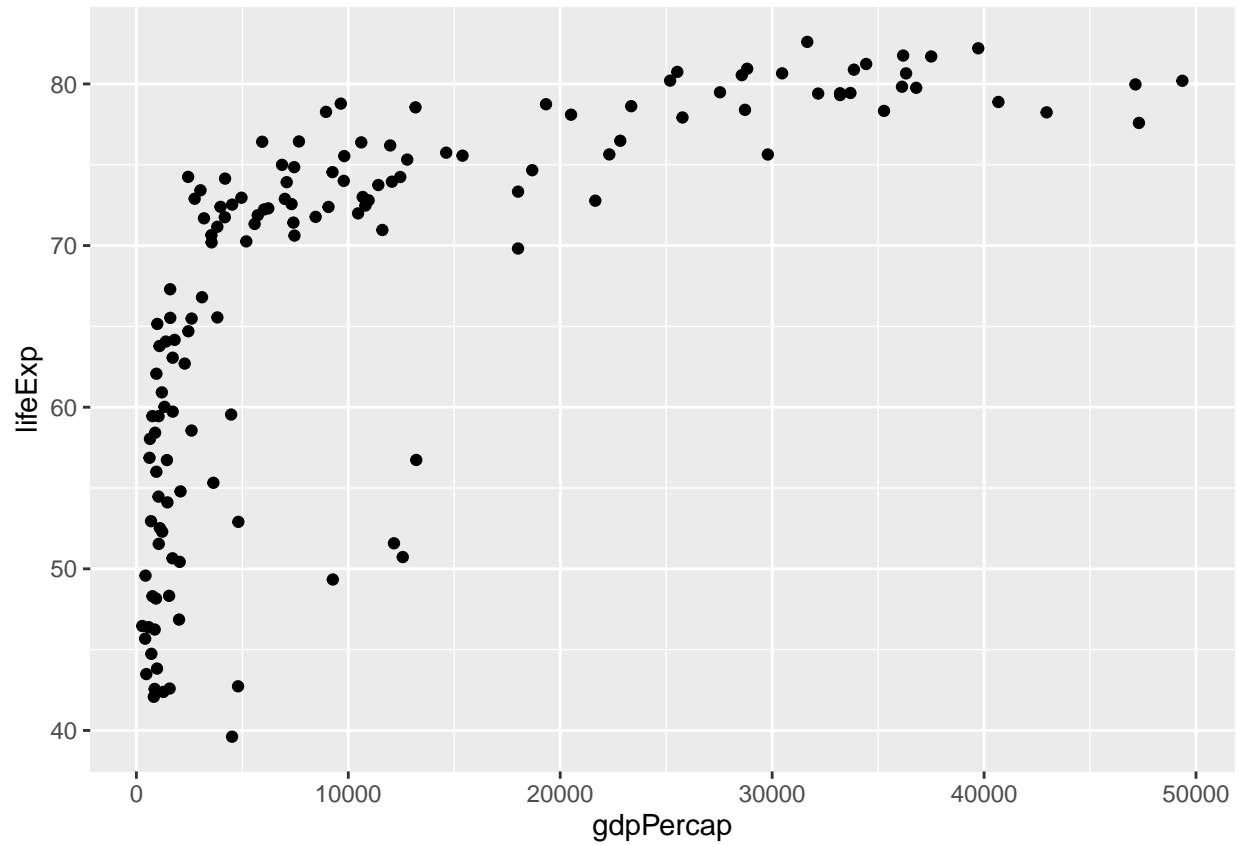
## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

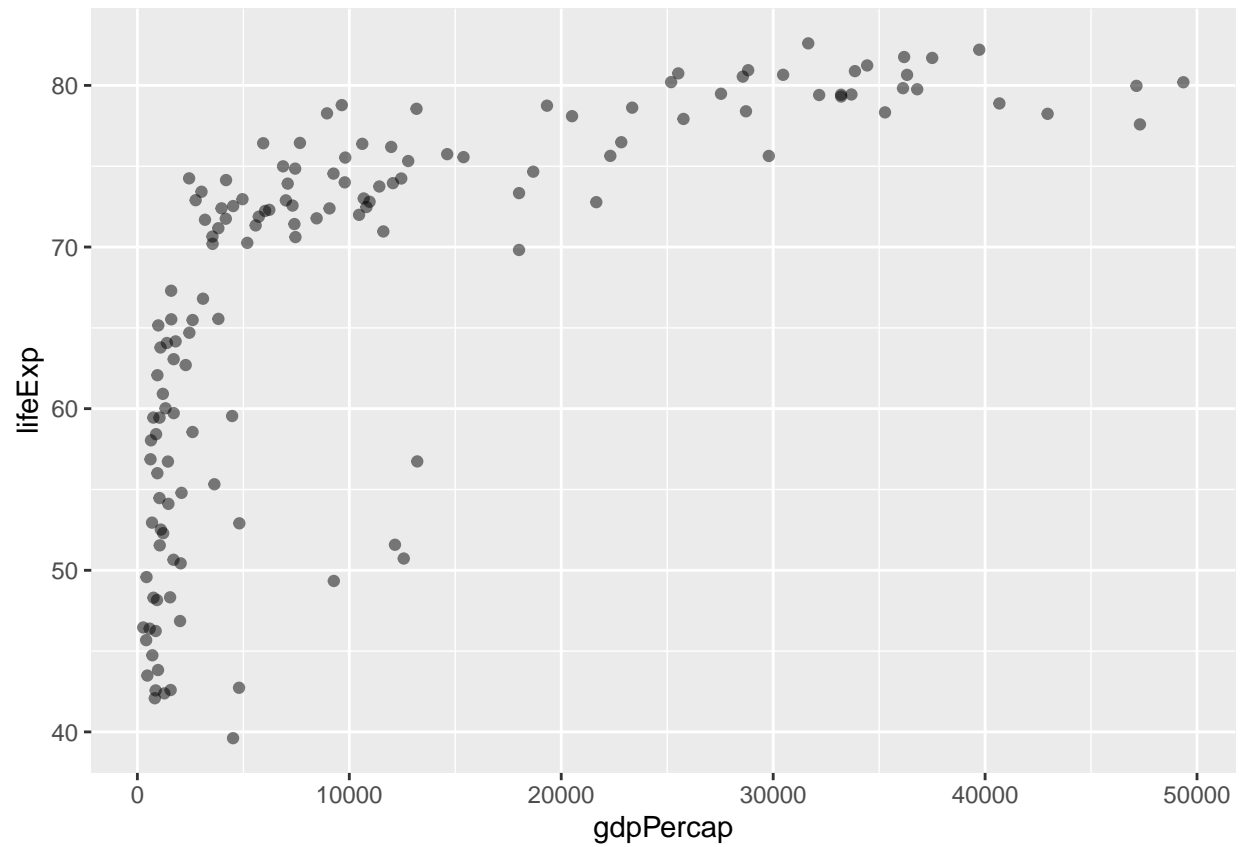
gapminder_2007 <- gapminder %>% filter (year==2007)

# scatter plot
```

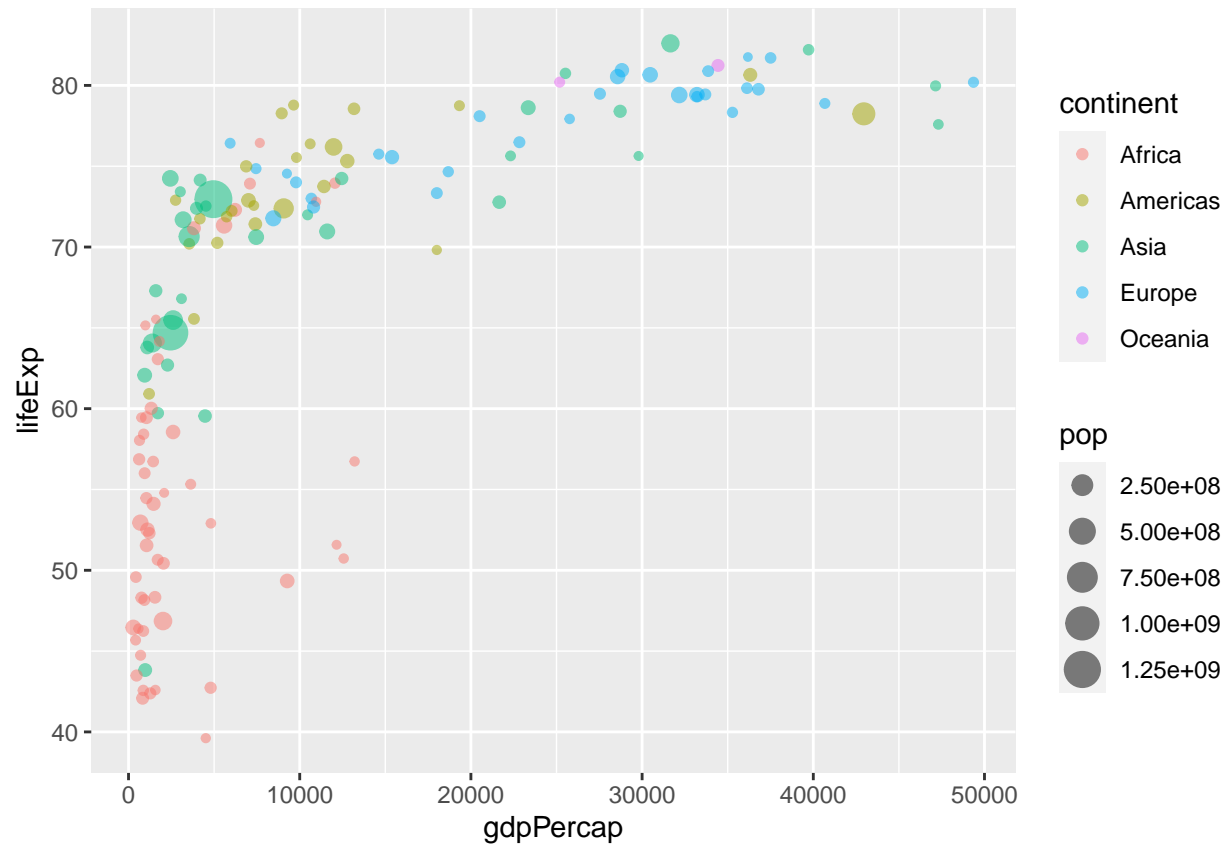
```
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp) +  
  geom_point()
```



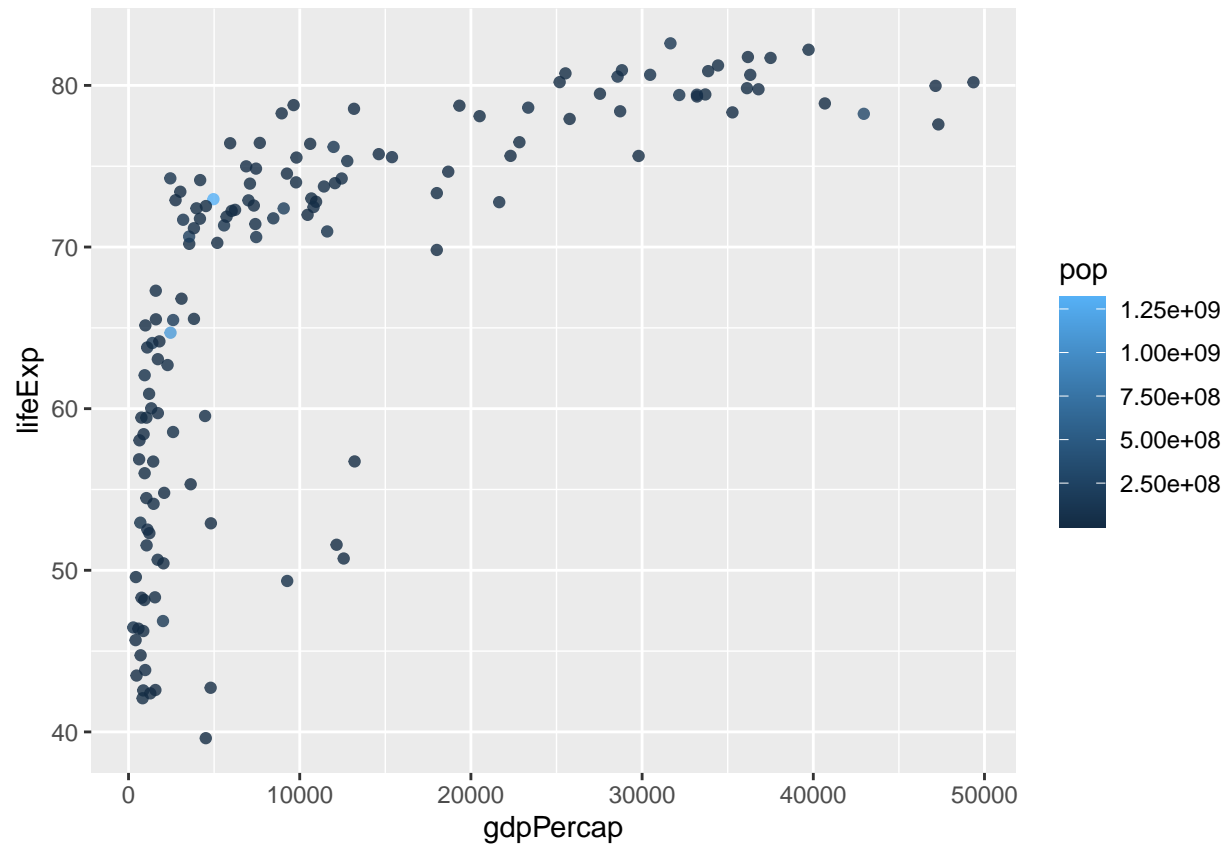
```
# Help us see better scatter plot  
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp) +  
  geom_point(alpha=0.5)
```



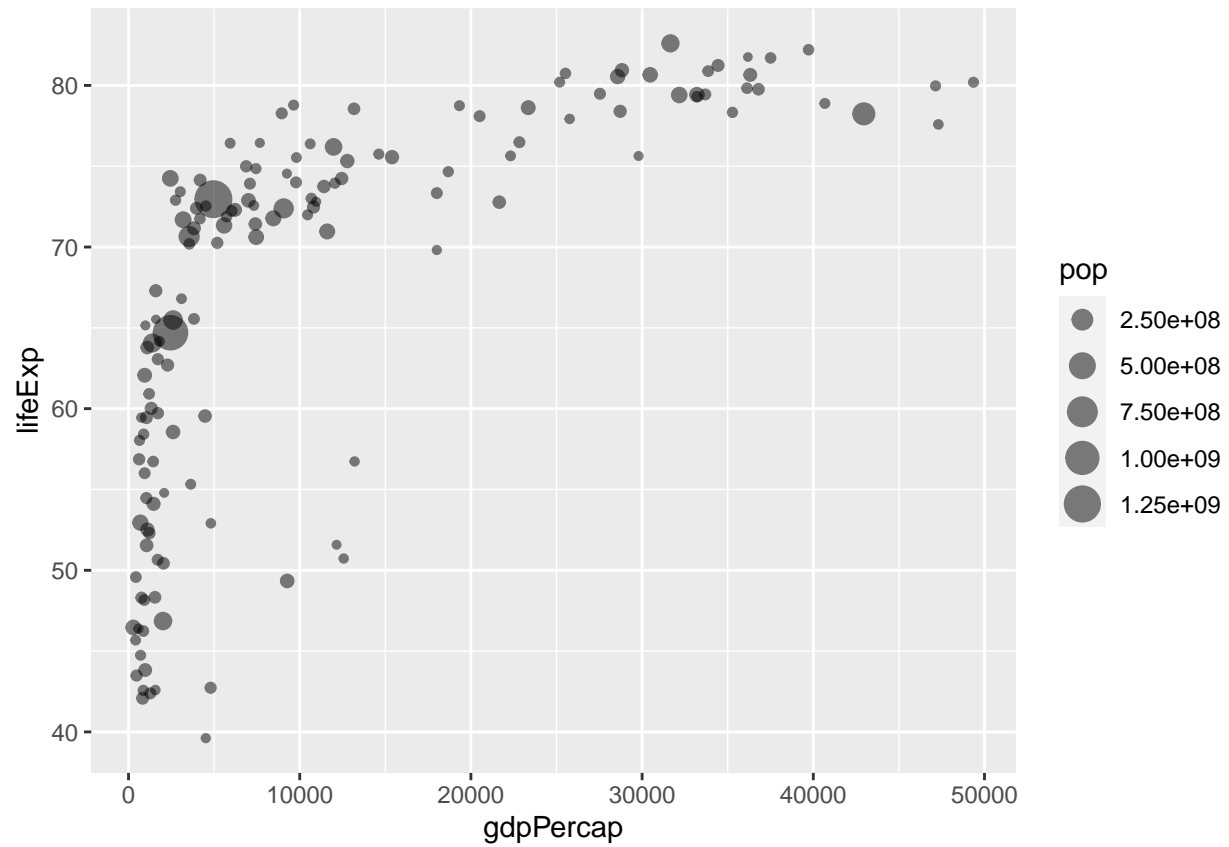
```
# Add more variables
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp,
      color= continent, size = pop) +
  geom_point(alpha=0.5)
```

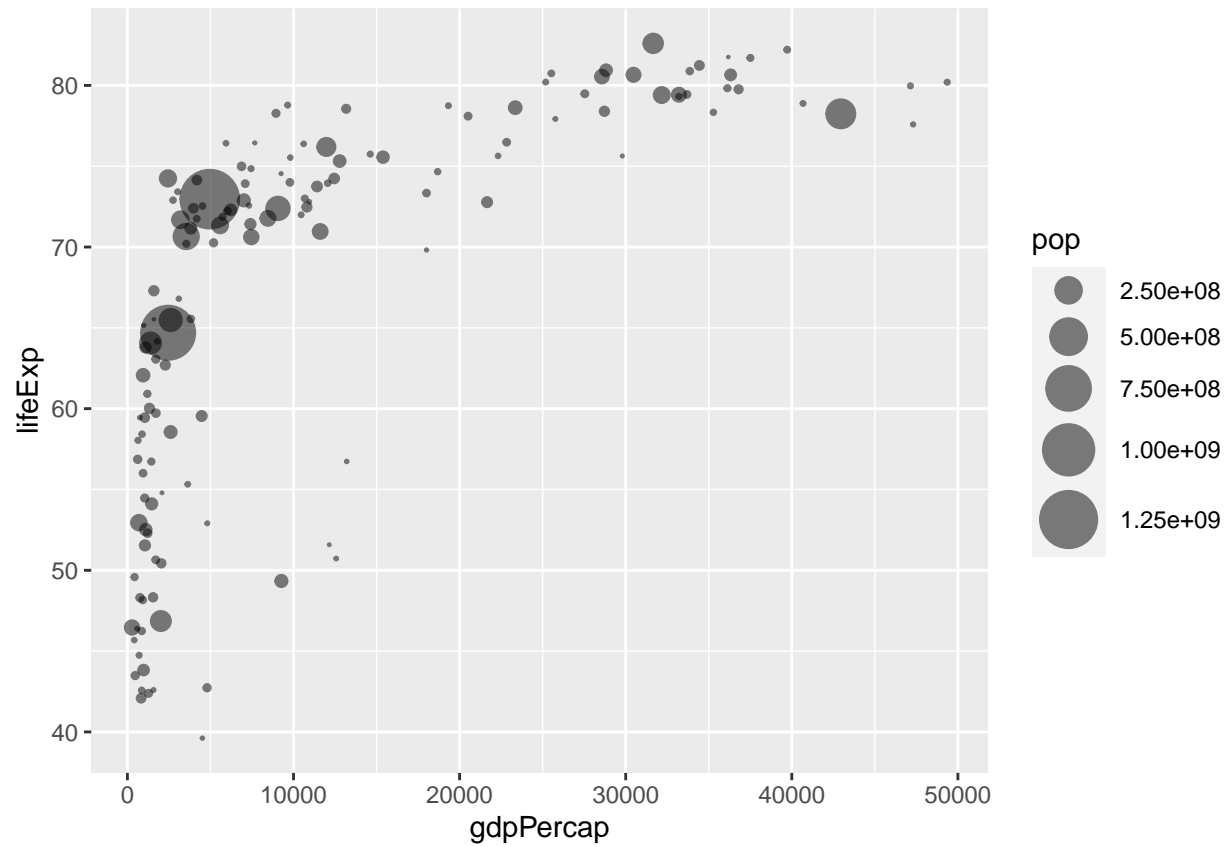
```
# numeric variable population
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp,
      color = pop) +
  geom_point(alpha= 0.8)
```



```
# Adjusting point size  
ggplot(gapminder_2007) +  
  aes(x=gdpPercap, y=lifeExp,  
       size = pop) +  
  geom_point(alpha= 0.5)
```

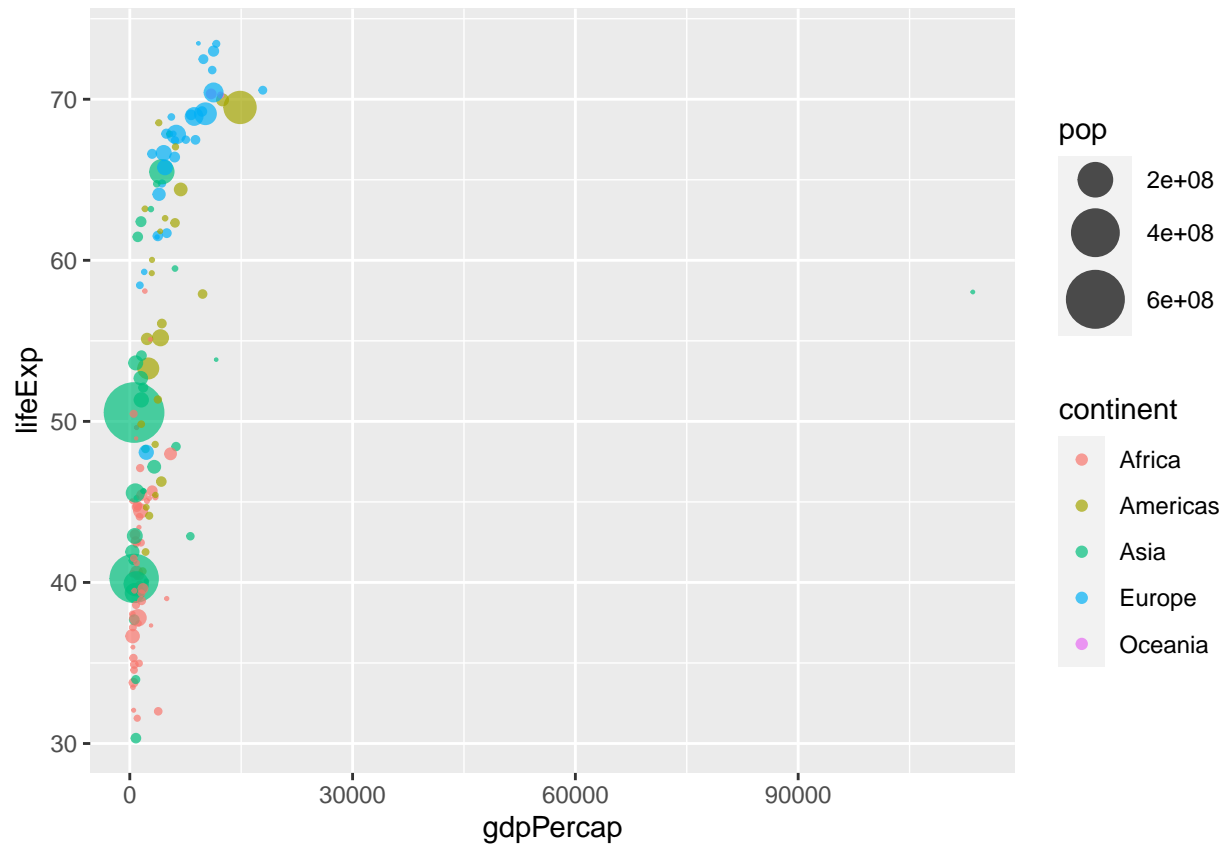


```
# Scale Size area  
ggplot(gapminder_2007) +  
  geom_point(aes(x=gdpPercap, y=lifeExp,  
    size = pop), alpha= 0.5) +  
  scale_size_area(max_size = 10)
```



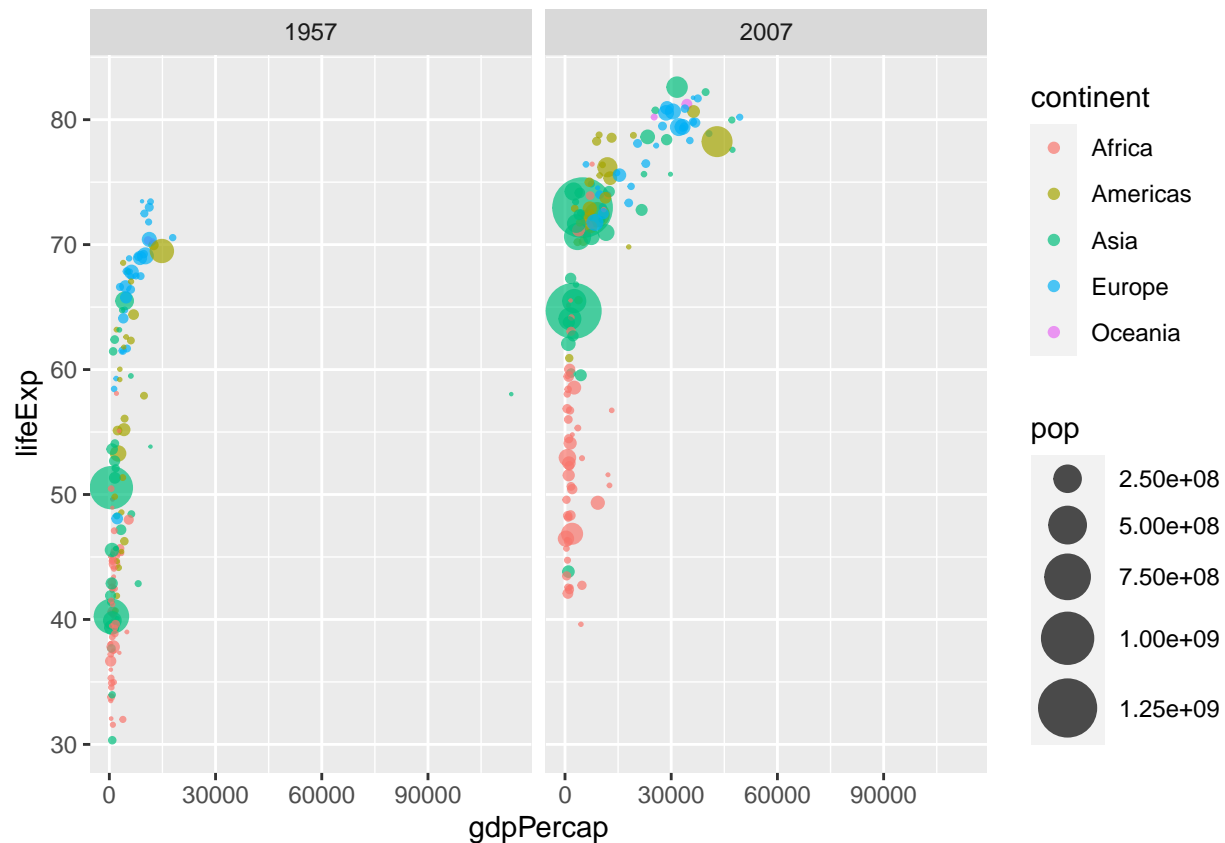
```
#ggplot 1957
gapminder_1957 <- gapminder %>% filter(year==1957)

ggplot(gapminder_1957) +
  aes(x = gdpPercap, y = lifeExp,
      color = continent, size = pop) +
  geom_point(alpha=0.7) +
  scale_size_area(max_size = 10)
```



```
# both 1957 & 2007
gapminder_57_07 <- gapminder %>% filter(year==1957 | year==2007)

ggplot(gapminder_57_07) +
  geom_point(aes(x = gdpPercap, y = lifeExp,
                 color=continent,size = pop),
             alpha=0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```

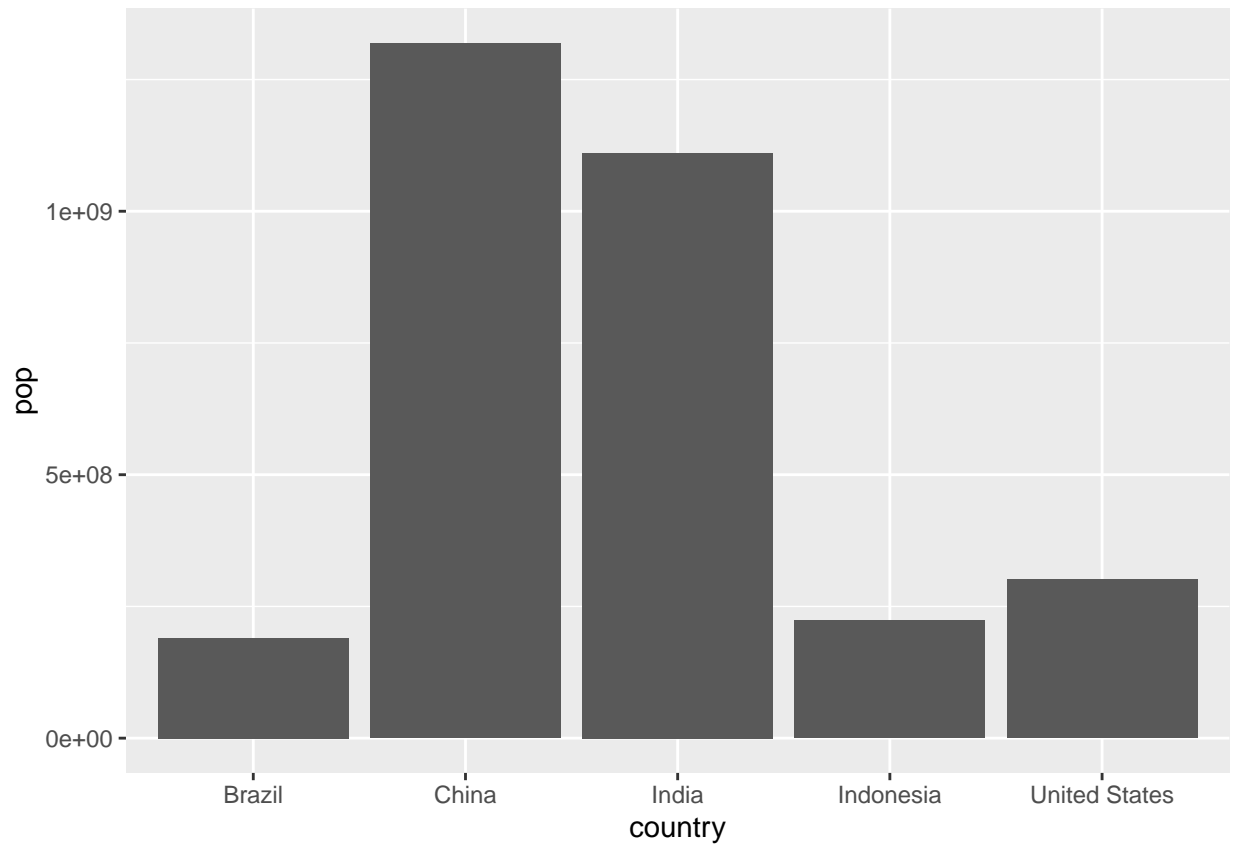


```
# Optional- Bar Chart
gapminder_top5 <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(pop)) %>%
  top_n(5, pop)

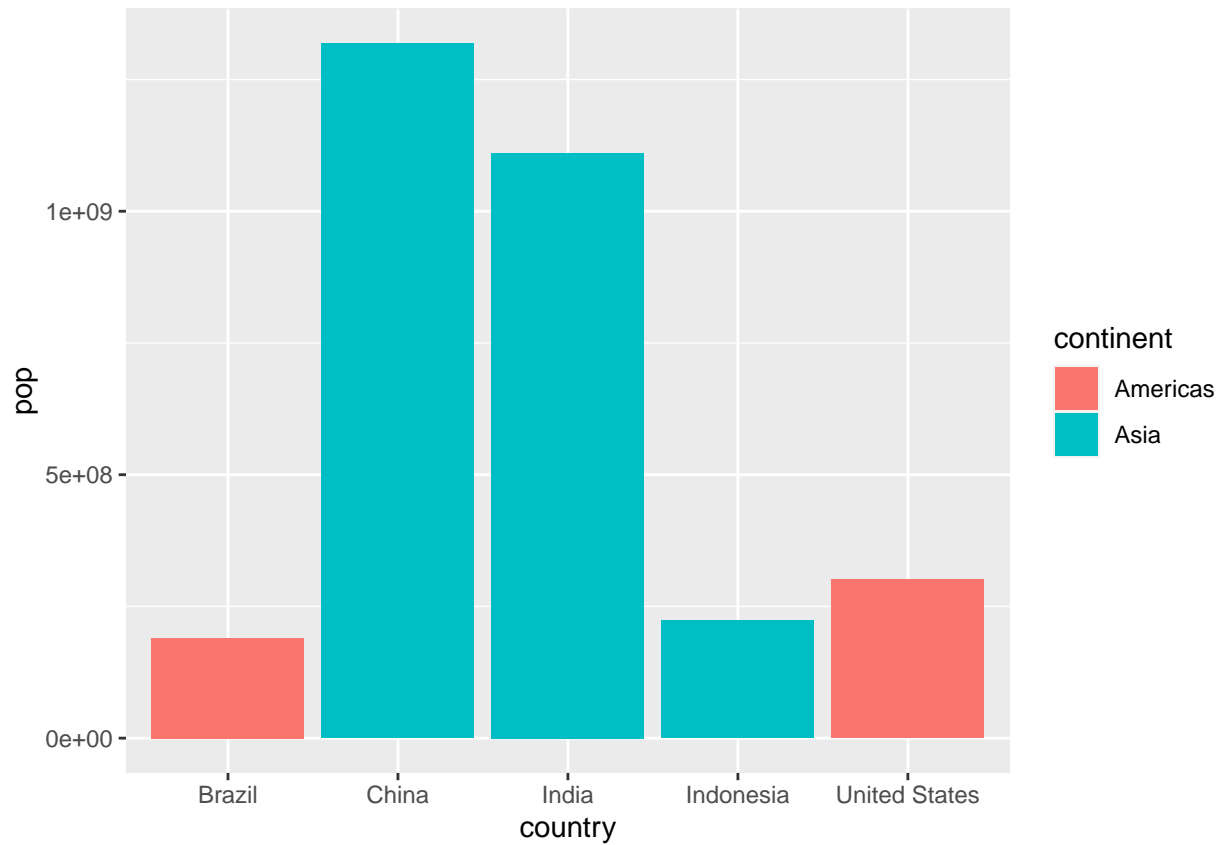
gapminder_top5
```

```
##      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
```

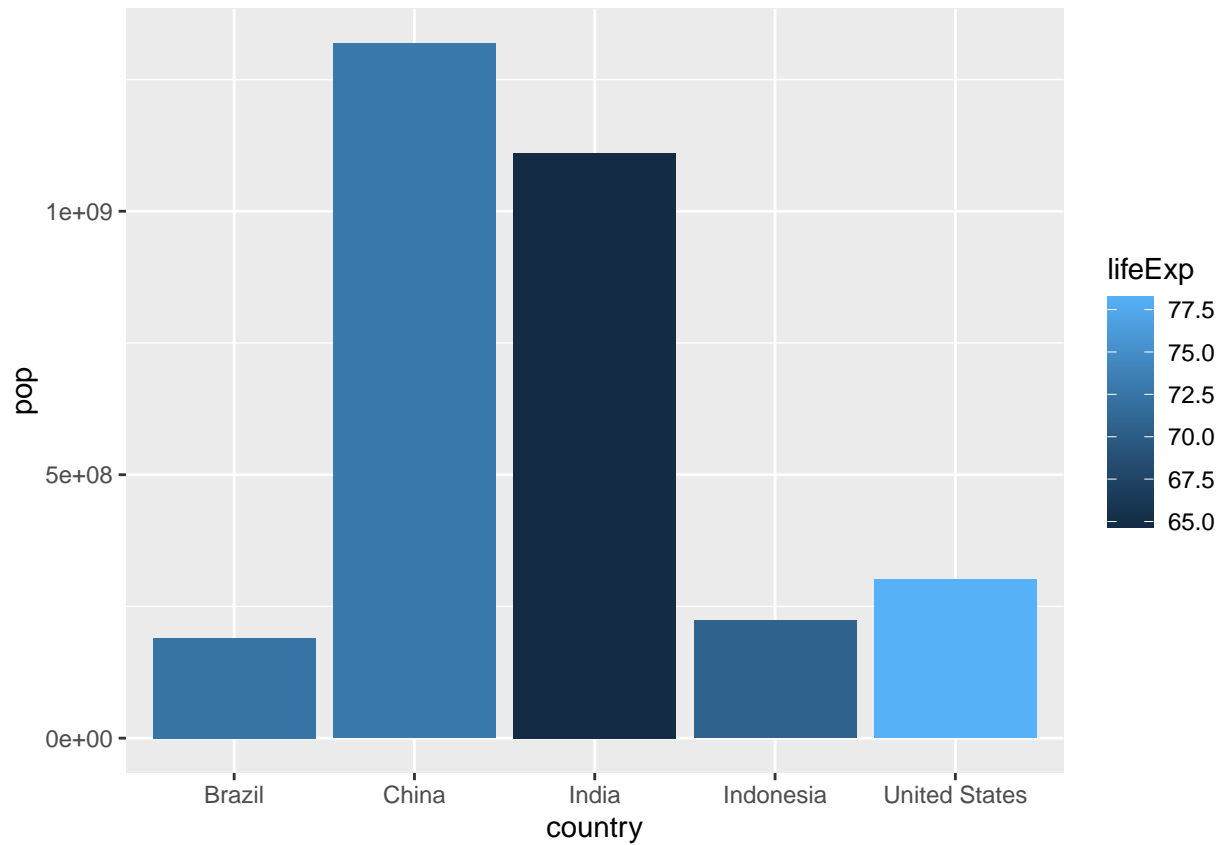
```
# Create bar charts
ggplot(gapminder_top5)+
  geom_col(aes(x=country,y=pop))
```



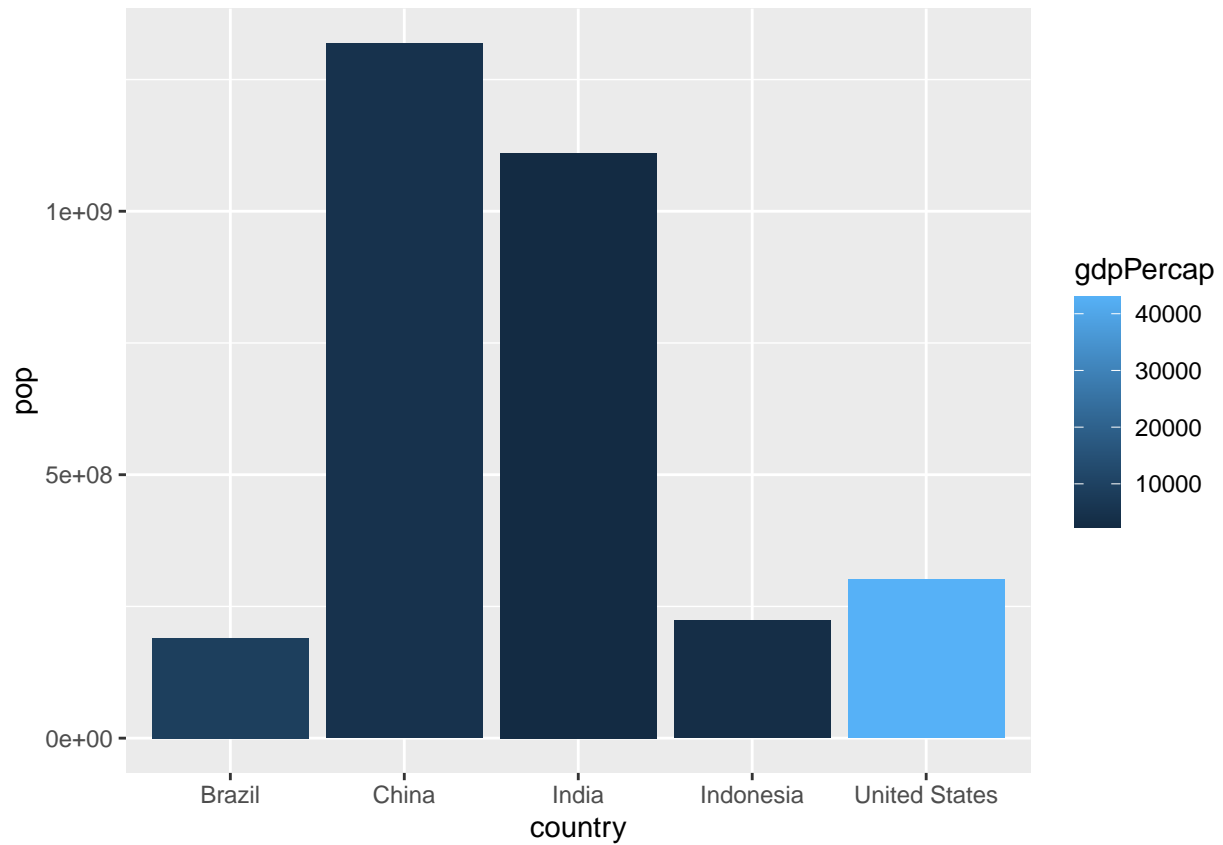
```
# Fill with color  
ggplot(gapminder_top5) +  
  geom_col(aes(x = country, y = pop, fill= continent))
```



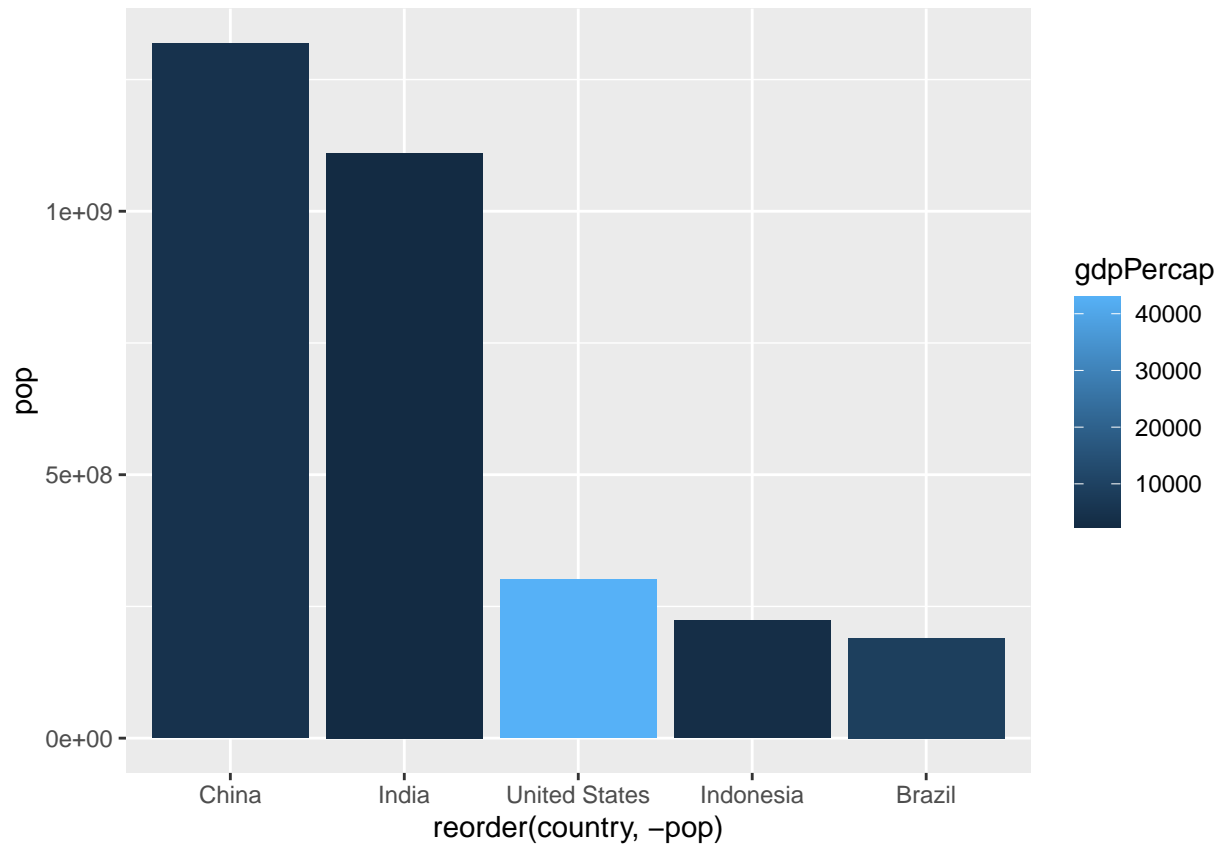
```
# Use numeric variable like Life Expectancy instead  
ggplot(gapminder_top5) +  
  geom_col(aes(x = country, y = pop, fill= lifeExp))
```

```
# Population size by Country  
ggplot(gapminder_top5) +  
  aes(x = country, y = pop, fill=gdpPercap) +  
  geom_col()
```

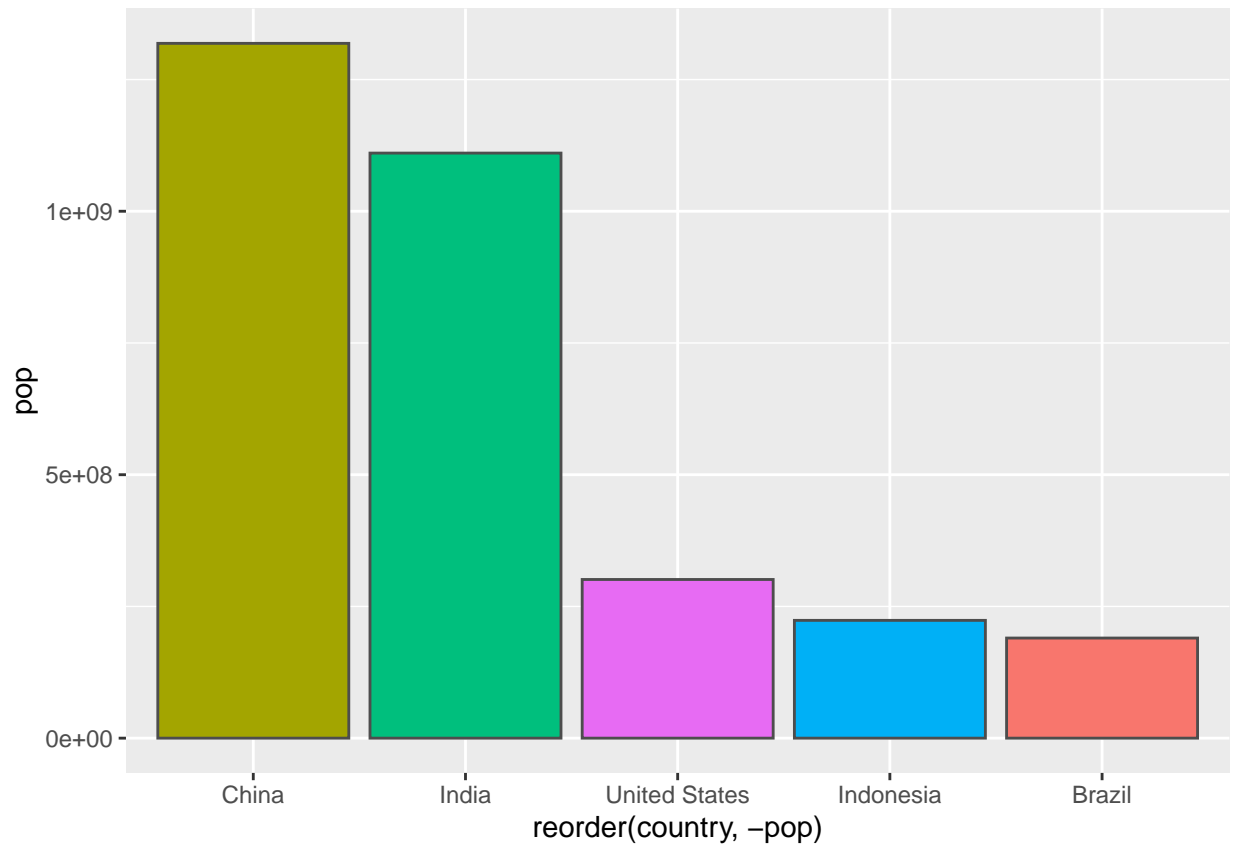


```
# Reorder bars
ggplot(gapminder_top5) +
  aes(x = reorder(country, -pop), y = pop, fill=gdpPercap) +
  geom_col()
```



```
# Fill by countries
ggplot(gapminder_top5) +
  aes(x = reorder(country, -pop), y = pop, fill=country) +
  geom_col(col="gray30") +
  guides(fill=FALSE)
```

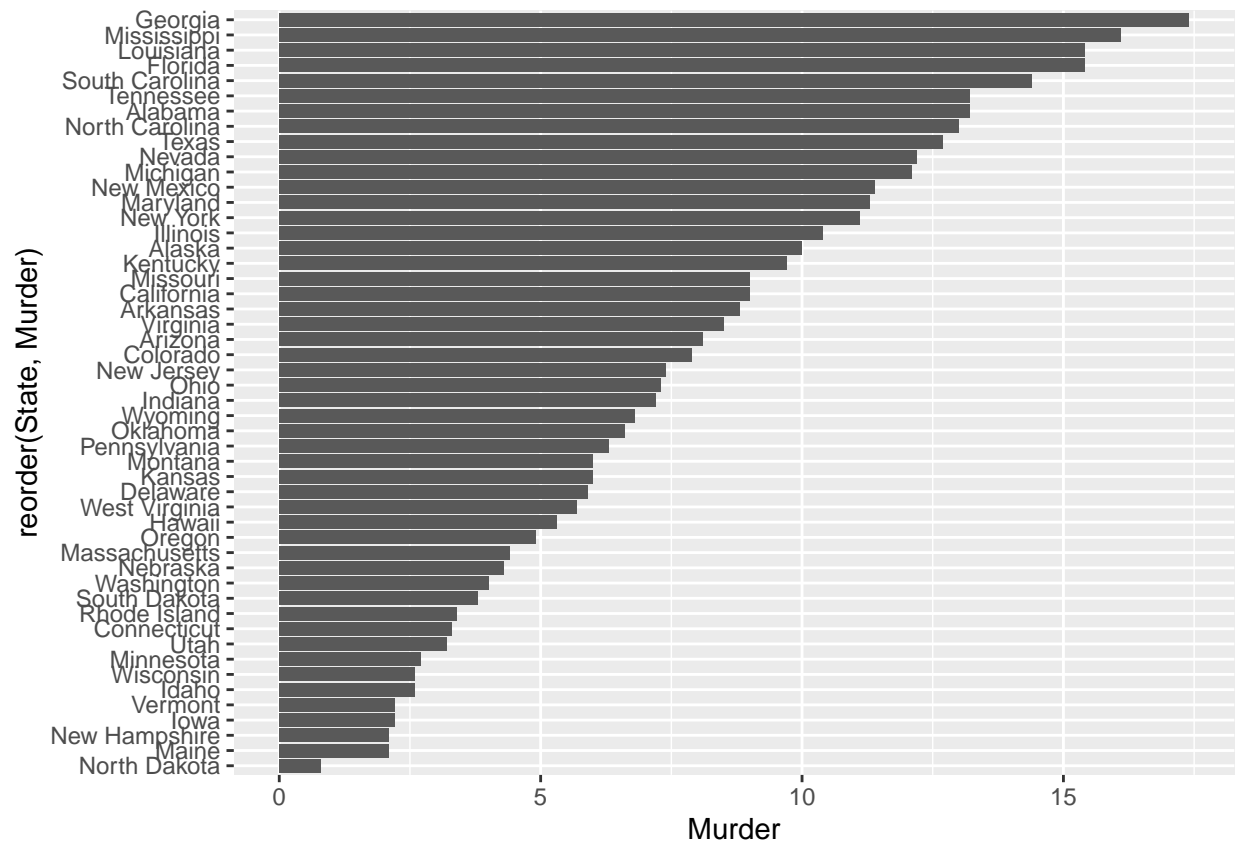
```
## Warning: 'guides(<scale> = FALSE)' is deprecated. Please use 'guides(<scale> =
## "none")' instead.
```



```
# Flipping bar charts
# Let's look at the inbuilt dataset USArrests
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()
```



```
# Seems crowded, let's fix by combining
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()
```

