Class05: Data visualization

Morgan Farrell (PID A59014314)

2022-02-02

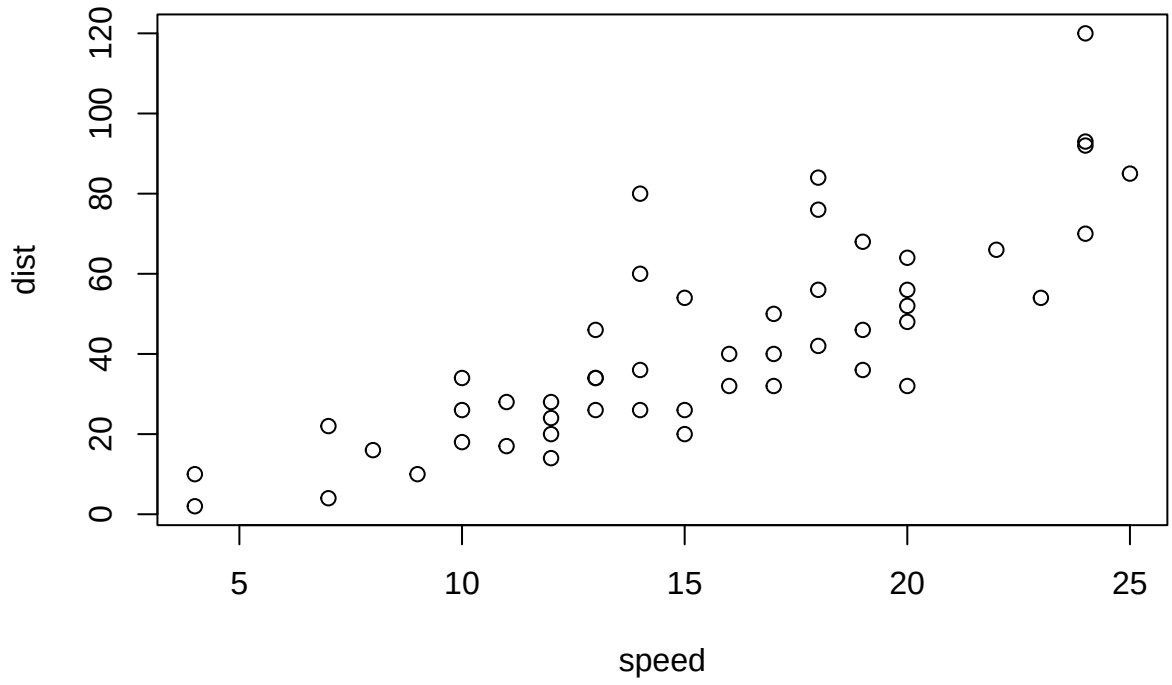
```
#install.packages("ggplot2")
library(ggplot2)
```

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

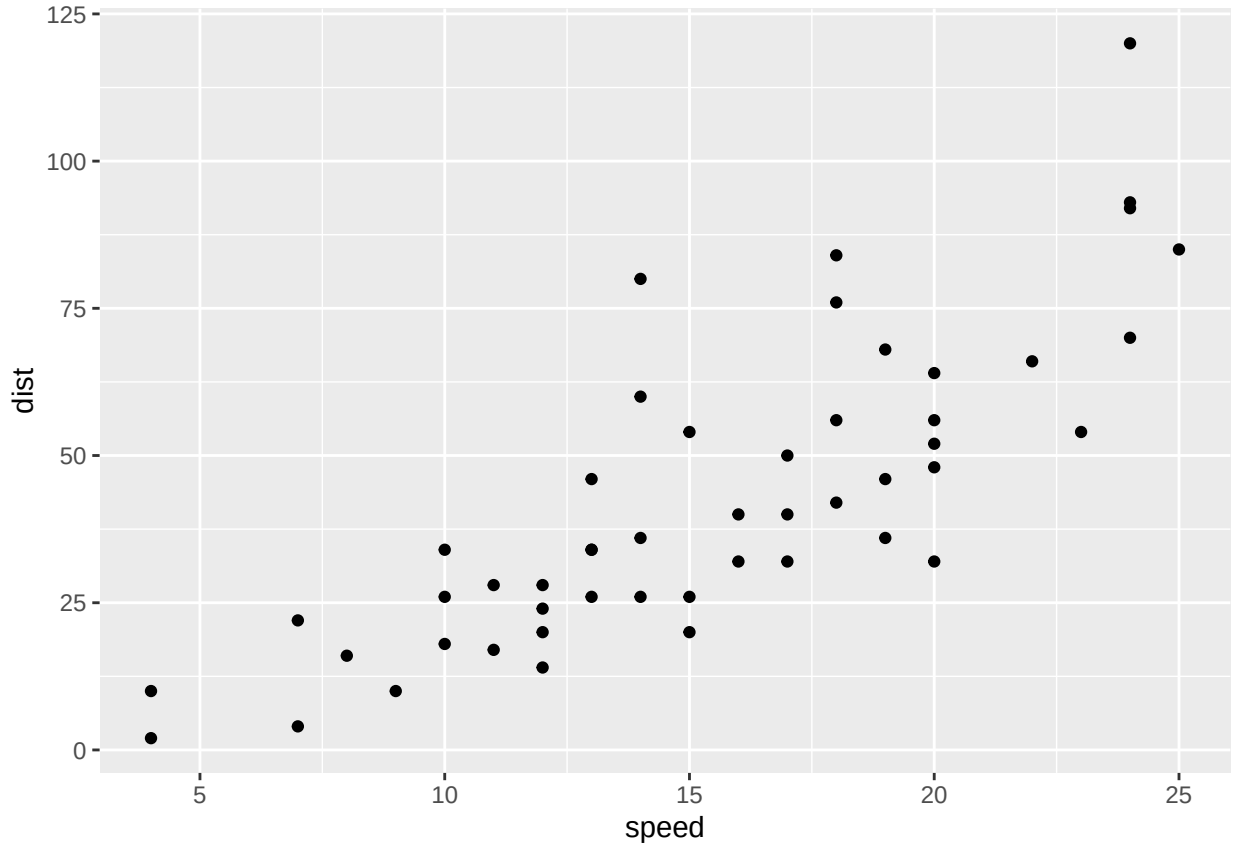
```
ggplot(cars)
```



```
#View(cars)

#Rs built in plot
plot(cars)
```

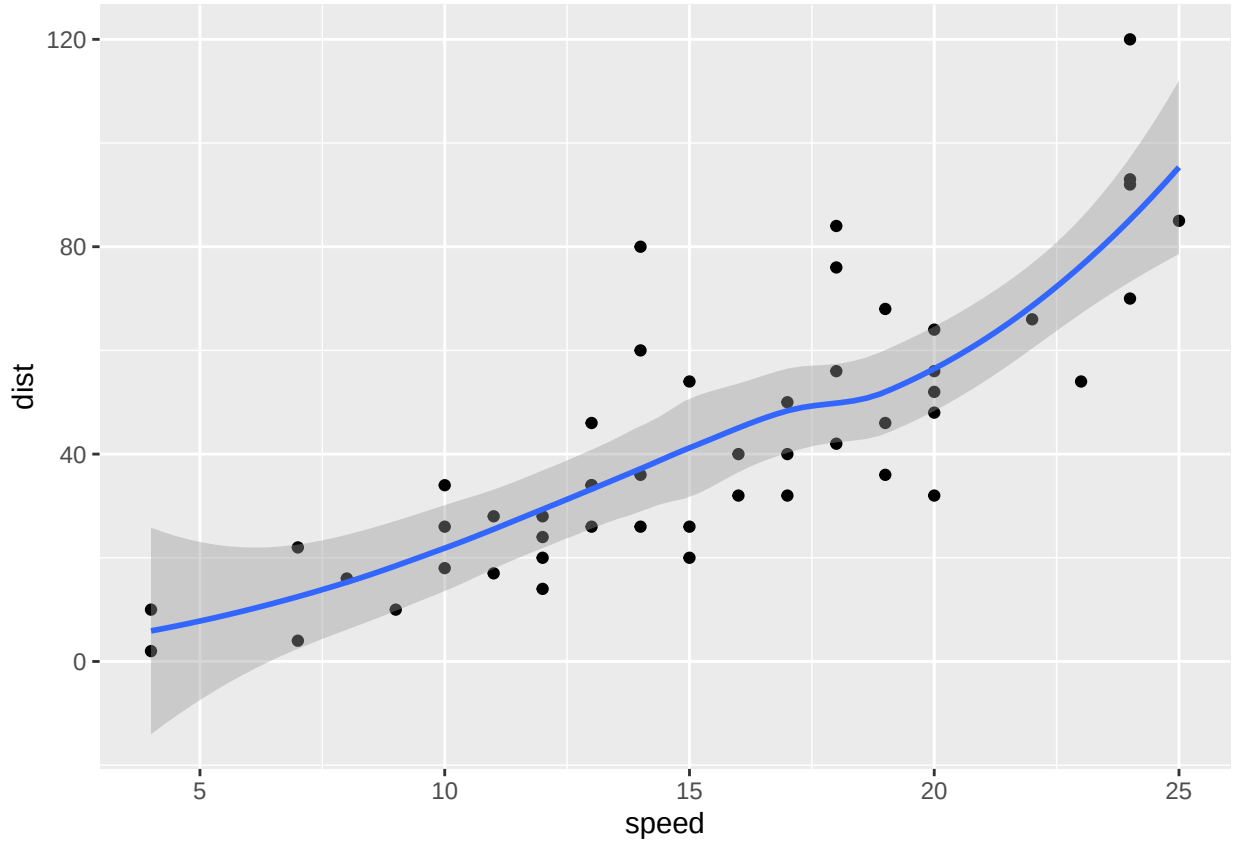



```
#ggplot version
ggplot(cars, aes(speed, dist))+
  geom_point()
```



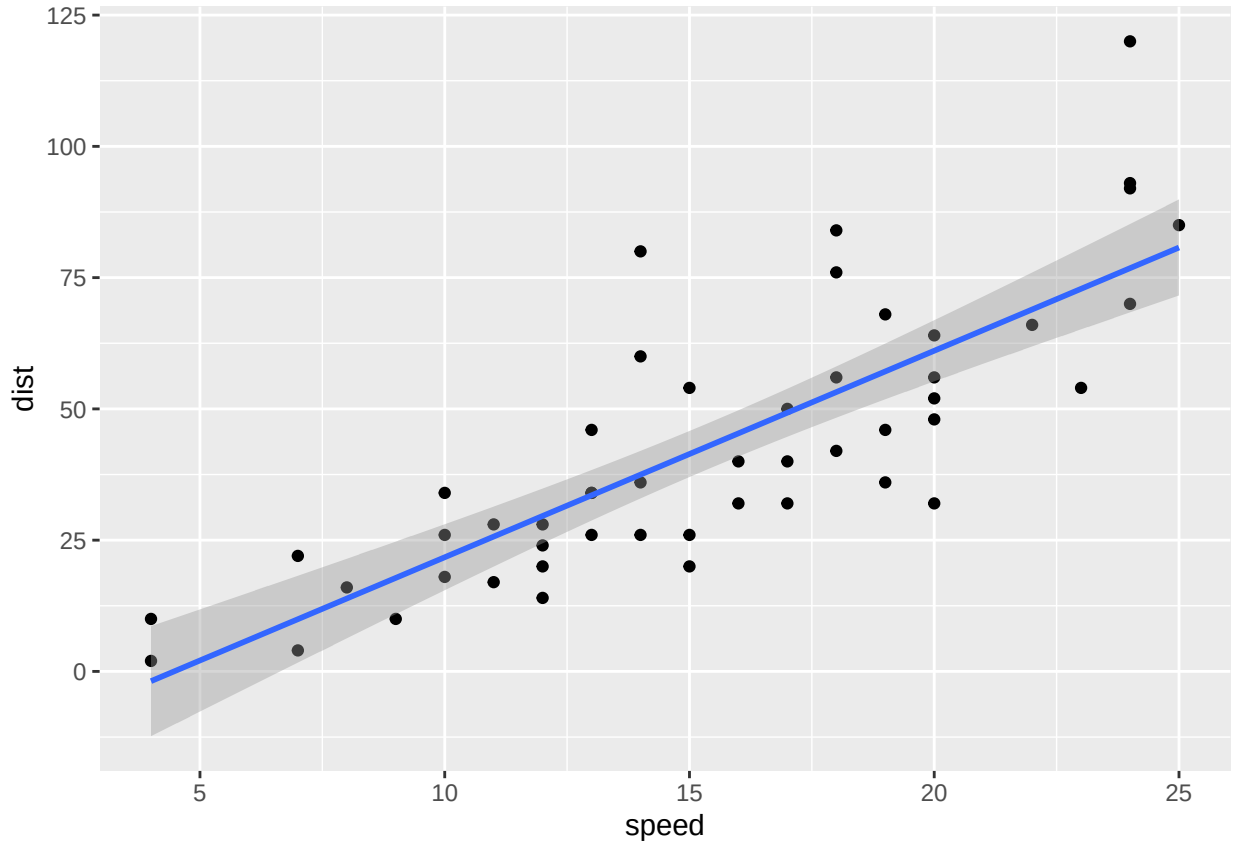

```
#Add trendline
ggplot(cars, aes(speed, dist))+
  geom_point()+
  geom_smooth()
```

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



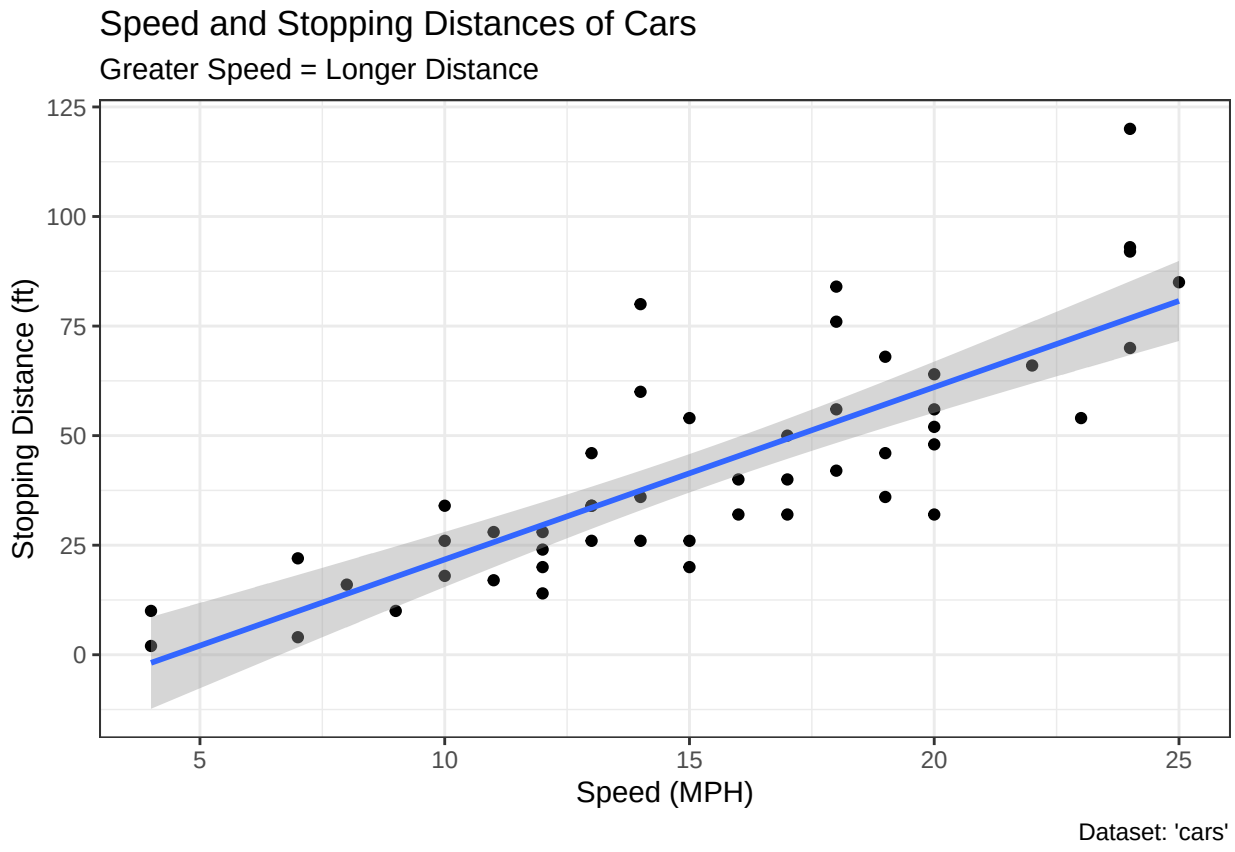

```
#fit a linear model
ggplot(cars, aes(speed, dist))+
  geom_point()+
  geom_smooth(method="lm")
```

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

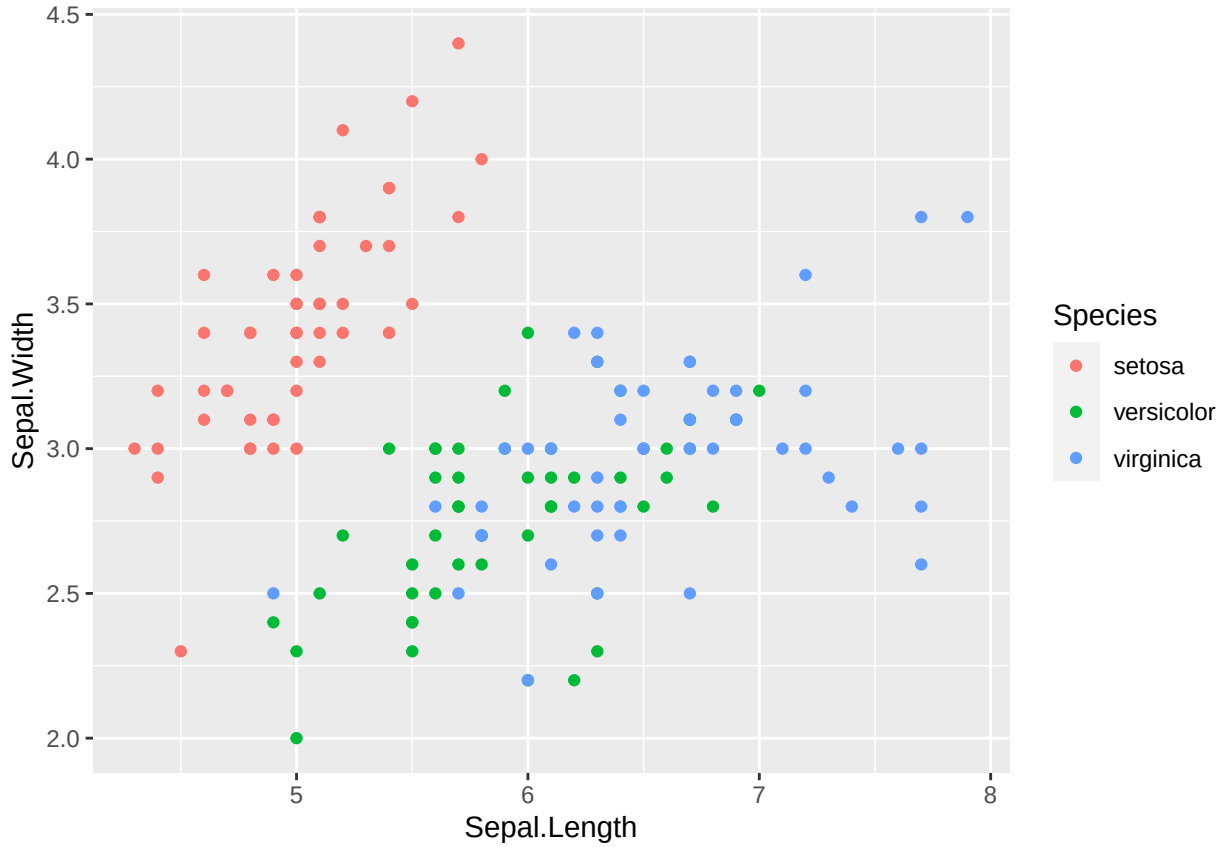



```
#Add labels and theme black and white
ggplot(cars, aes(speed, dist))+
  geom_point()+
  geom_smooth(method="lm")+
  labs(title="Speed and Stopping Distances of Cars",
       x="Speed (MPH)",
       y="Stopping Distance (ft)",
       subtitle = "Greater Speed = Longer Distance",
       caption = "Dataset: 'cars'") +
  theme_bw()
```

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




```
#######################################
#From Data to Viz Site: Scatterplot
ggplot(iris, aes(x=Sepal.Length, y=Sepal.Width, col=Species)) +
  geom_point()
```




```
#View(iris)
########################################
#Working with Genes data set
#RNAseq drug vs no drug treament

#load data and look at it
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"
```

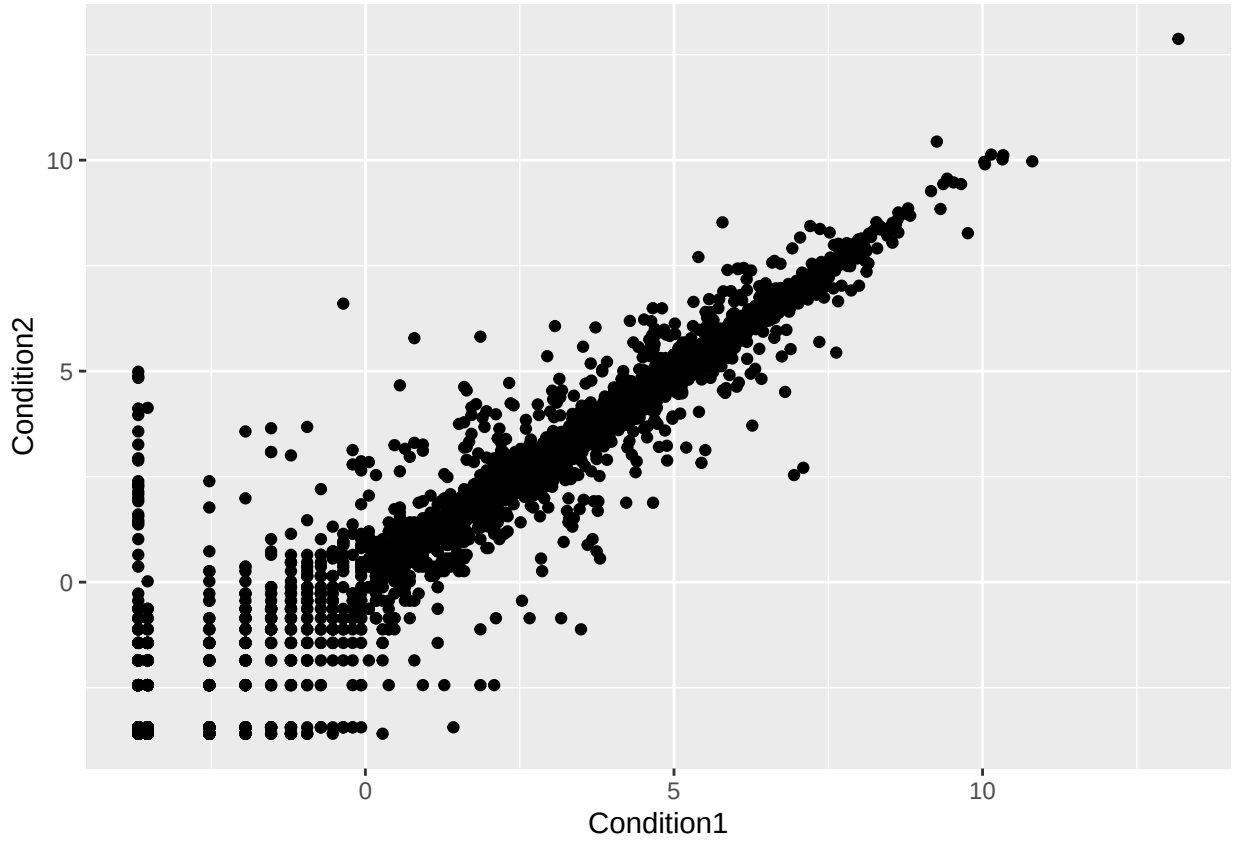
```
table(genes$State)
```

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

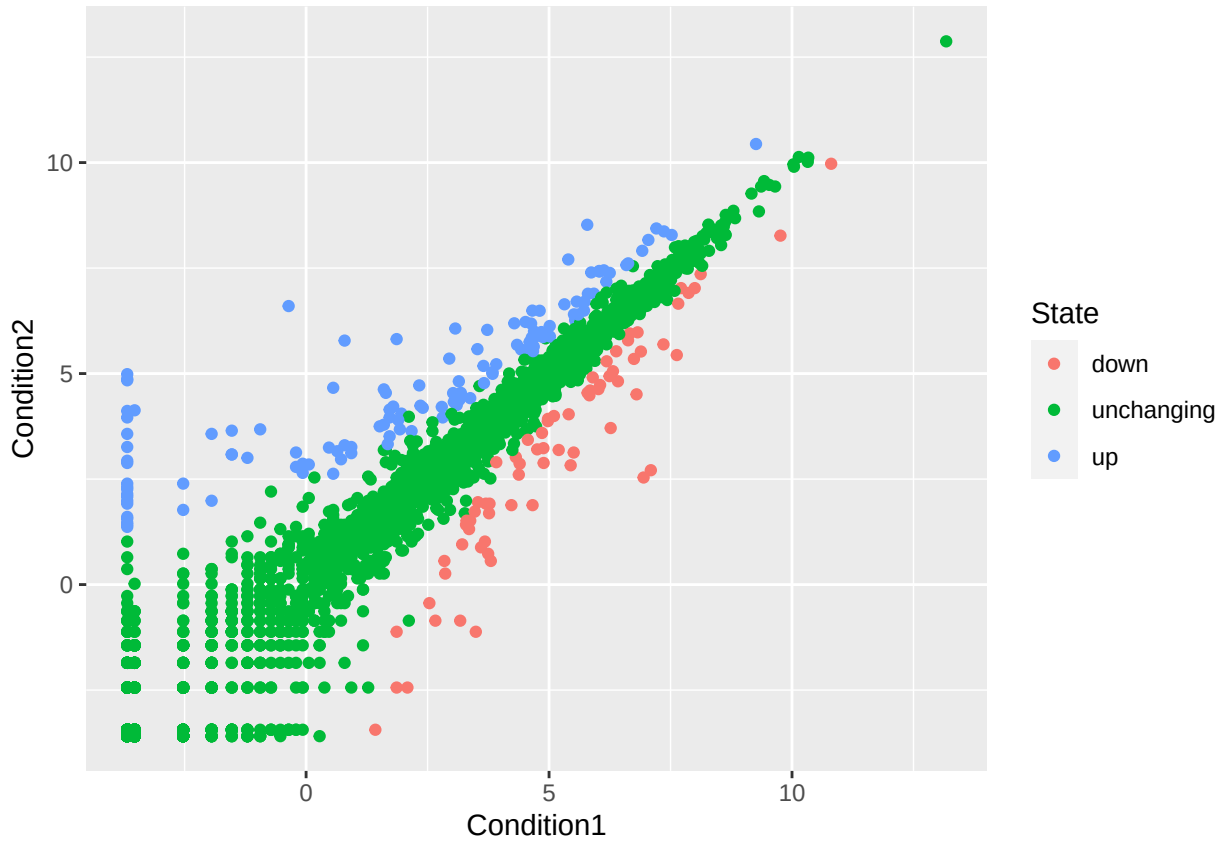
```
#Fraction of genes up , down, etc.
#Round function and final number is number of sig figs
round(table(genes$State)/nrow(genes)*100, 2)
```

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

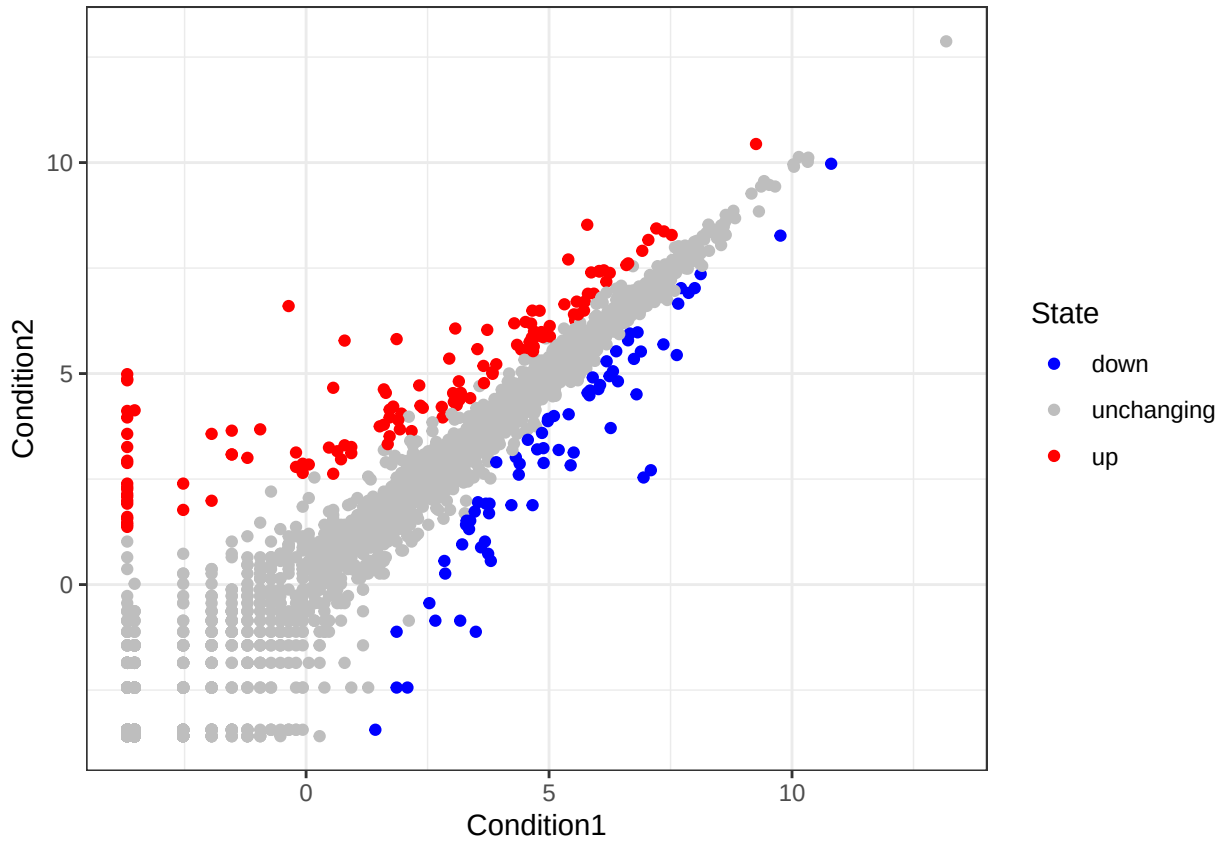
```
#Ploting the data frame genes
ggplot(genes, aes(Condition1, Condition2))+
  geom_point()
```



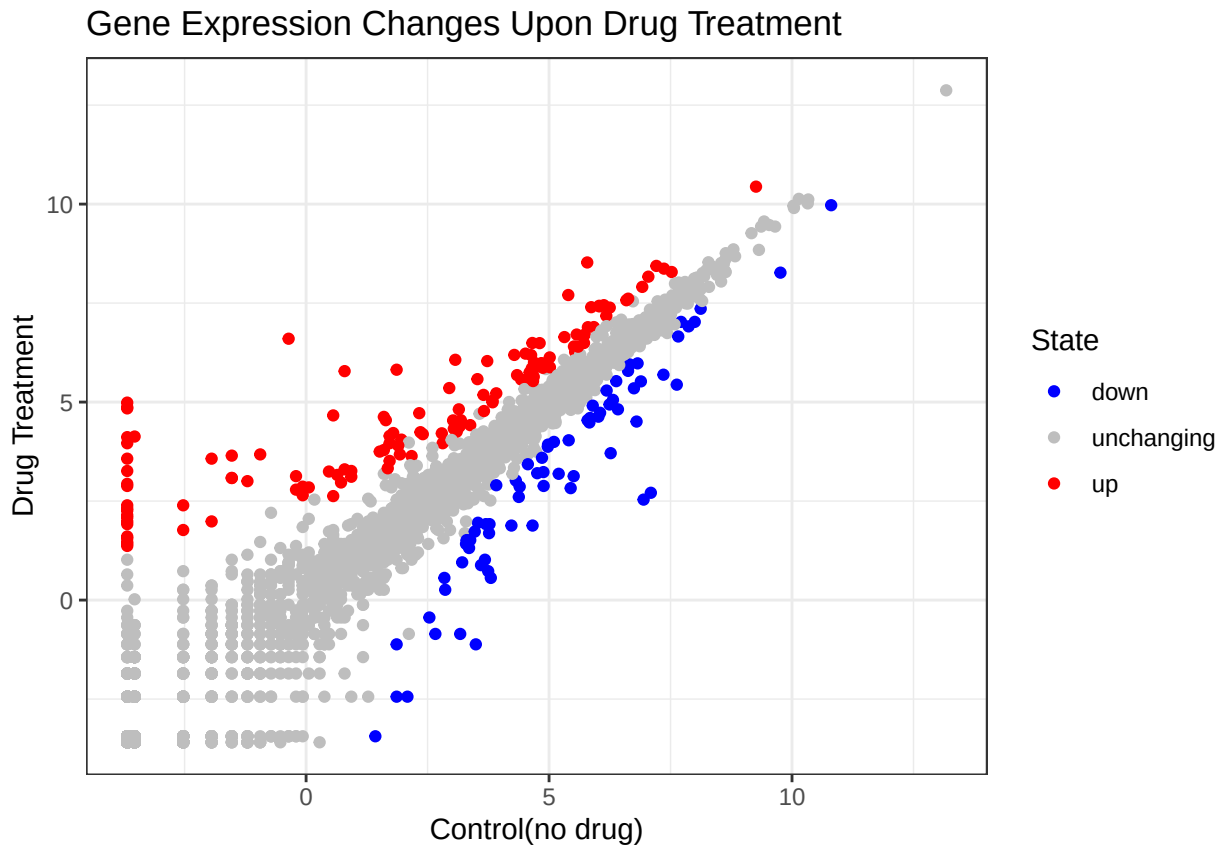

```
#Add column for state which tells up the difference in expression, color coded for state
ggplot(genes, aes(Condition1, Condition2, col=State))+
  geom_point()
```




```
# Add our own colors
ggplot(genes, aes(Condition1, Condition2, col=State))+
  geom_point()+
  theme_bw()+
  scale_color_manual(values=c("blue", "gray", "red"))
```



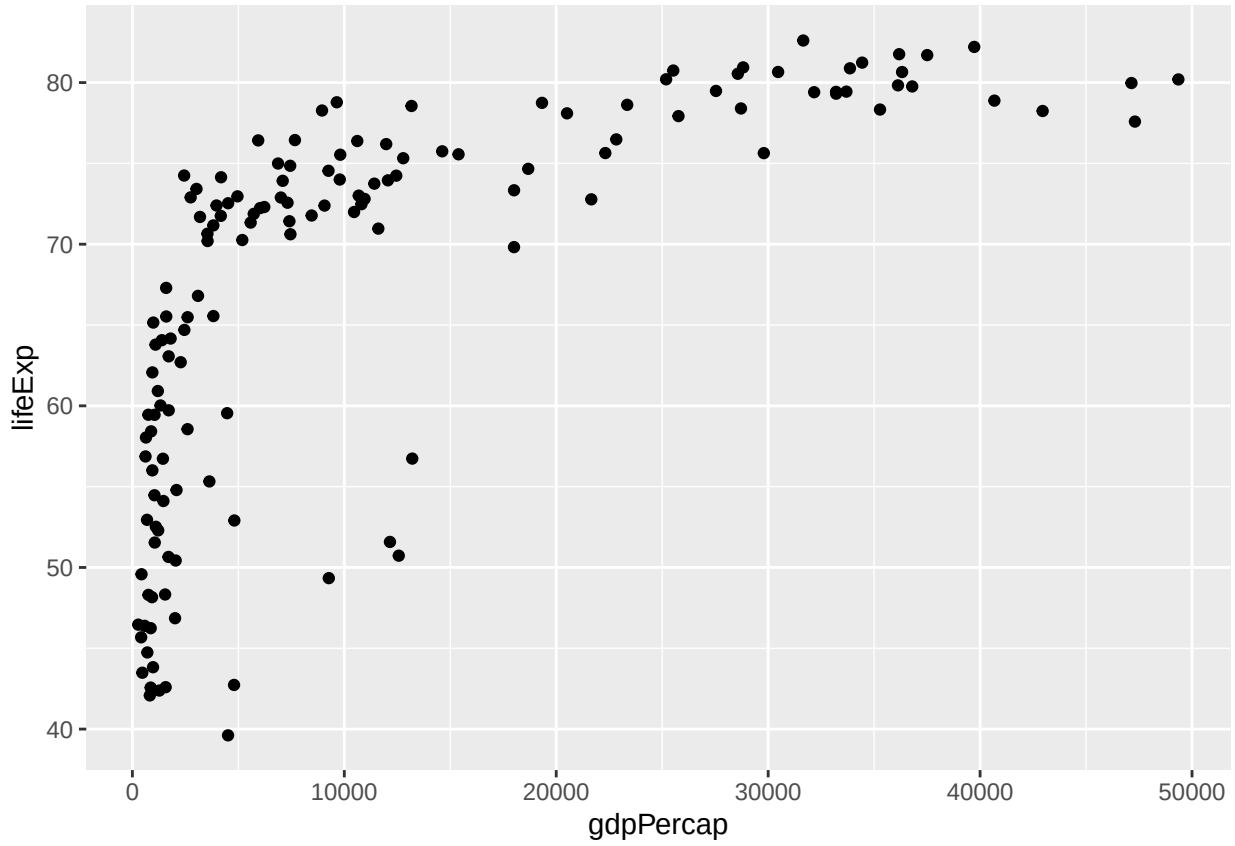

```
#Add labels
ggplot(genes, aes(Condition1, Condition2, col=State))+
  geom_point()+
  theme_bw()+
  scale_color_manual(values=c("blue", "gray", "red"))+
  labs(title="Gene Expression Changes Upon Drug Treatment",
       x="Control(no drug)",
       y="Drug Treatment")
```



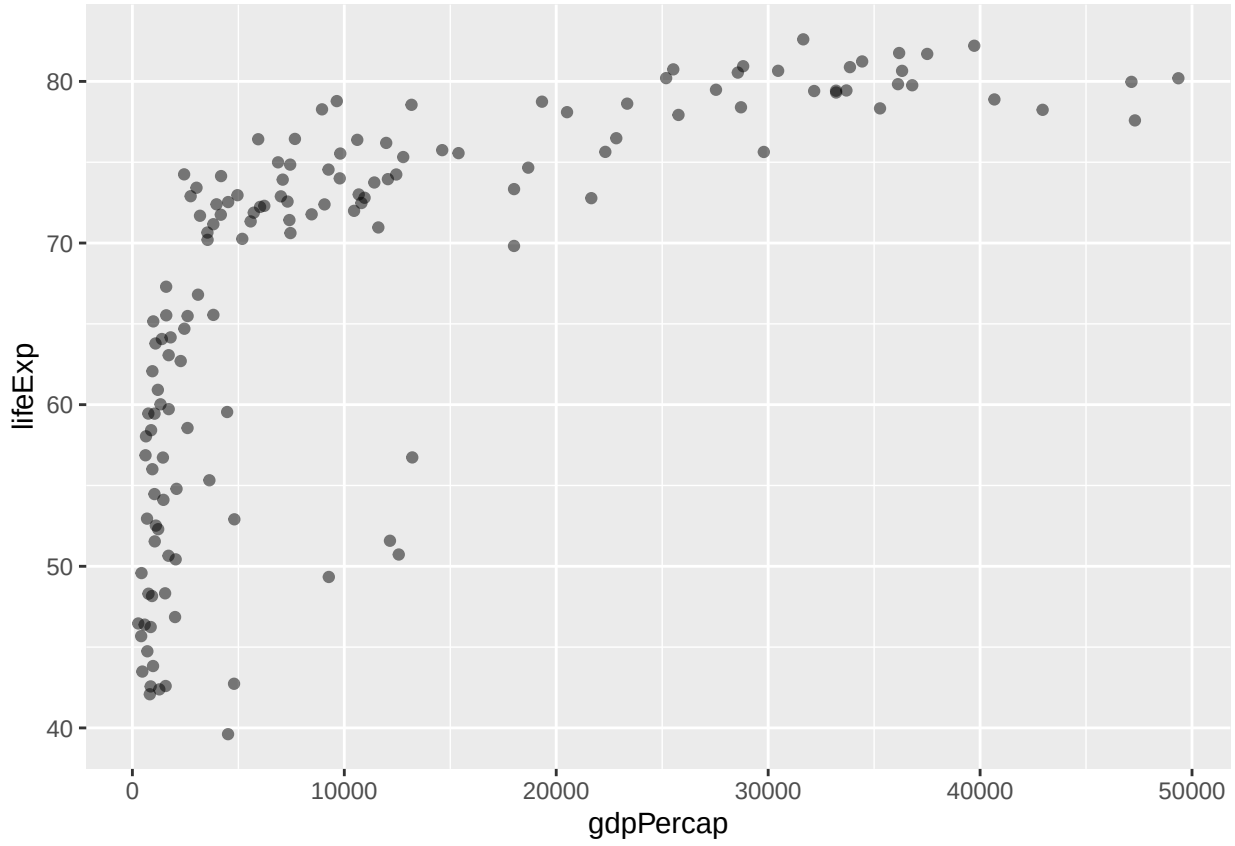

```
################################
# Gapminder package
#install.packages("gapminder")
library(gapminder)

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
```

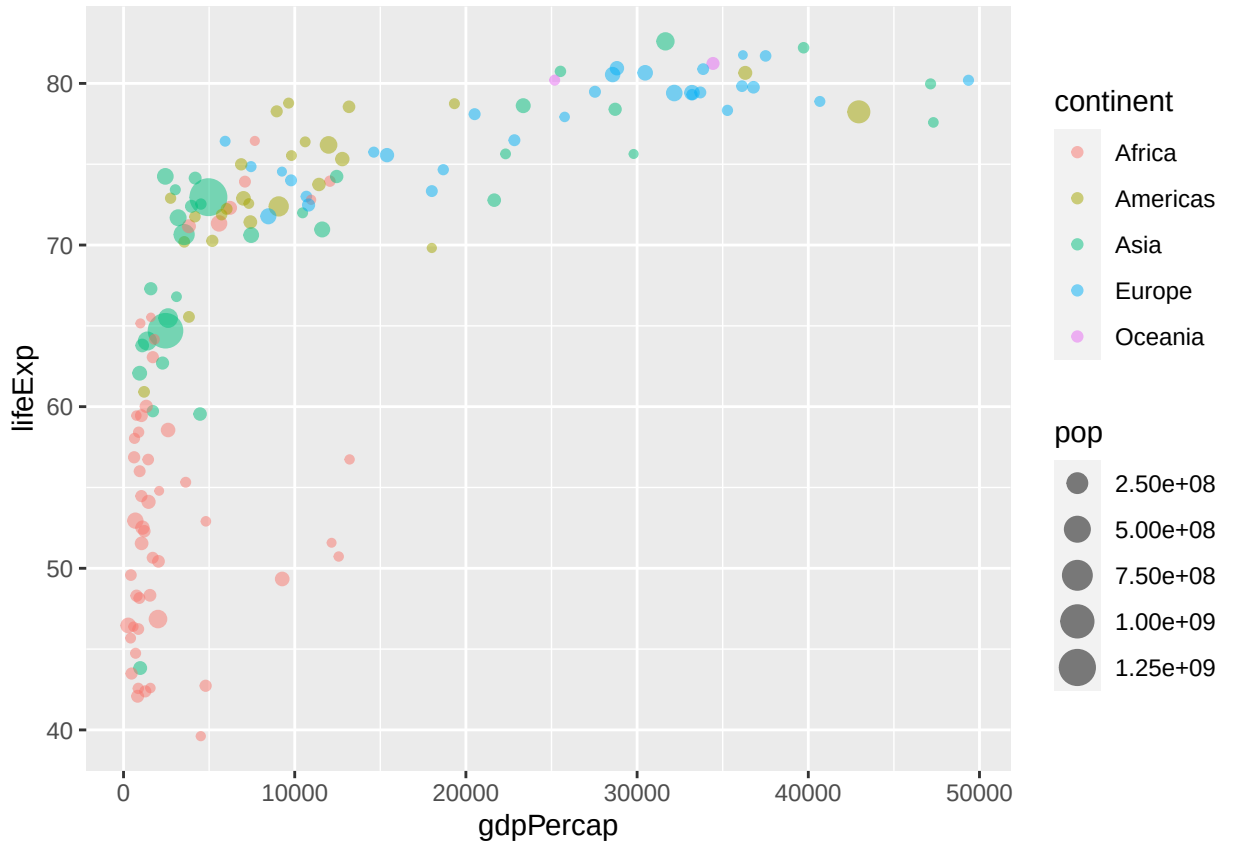
```
gapminder_2007 <- gapminder %>% filter(year==2007)

#First basic scatter plot
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp) +
  geom_point()
```



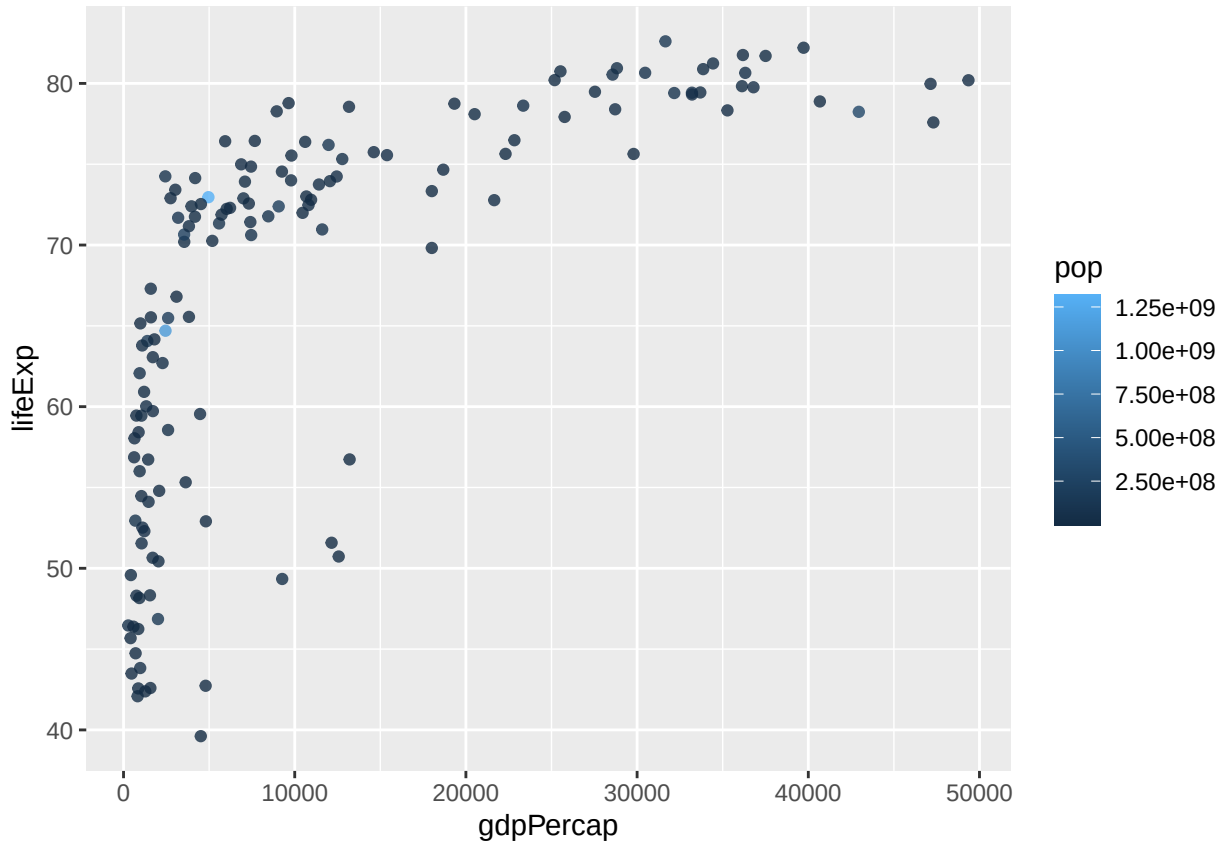

```
#Add transparancy to points
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp) +
  geom_point(alpha=0.5)
```



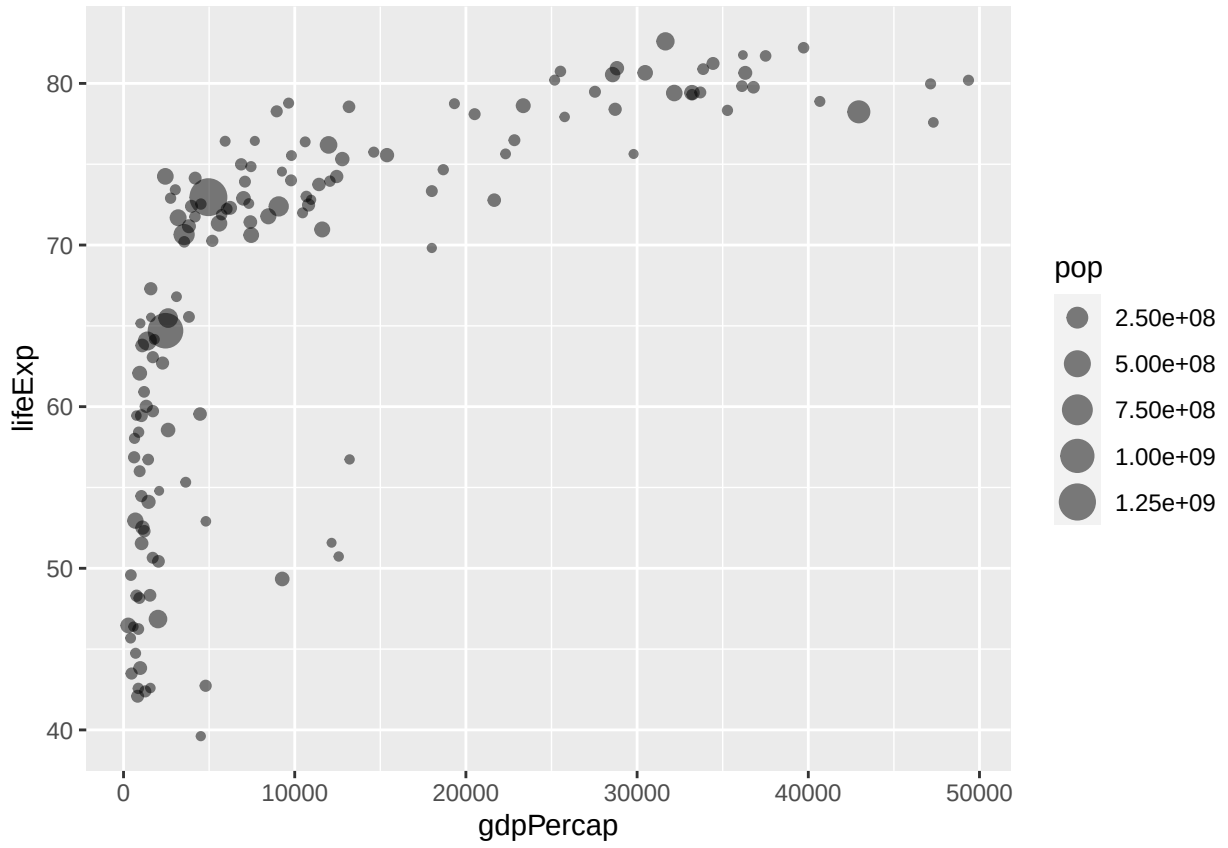

```
#Adding more aes
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +
  geom_point(alpha=0.5)
```



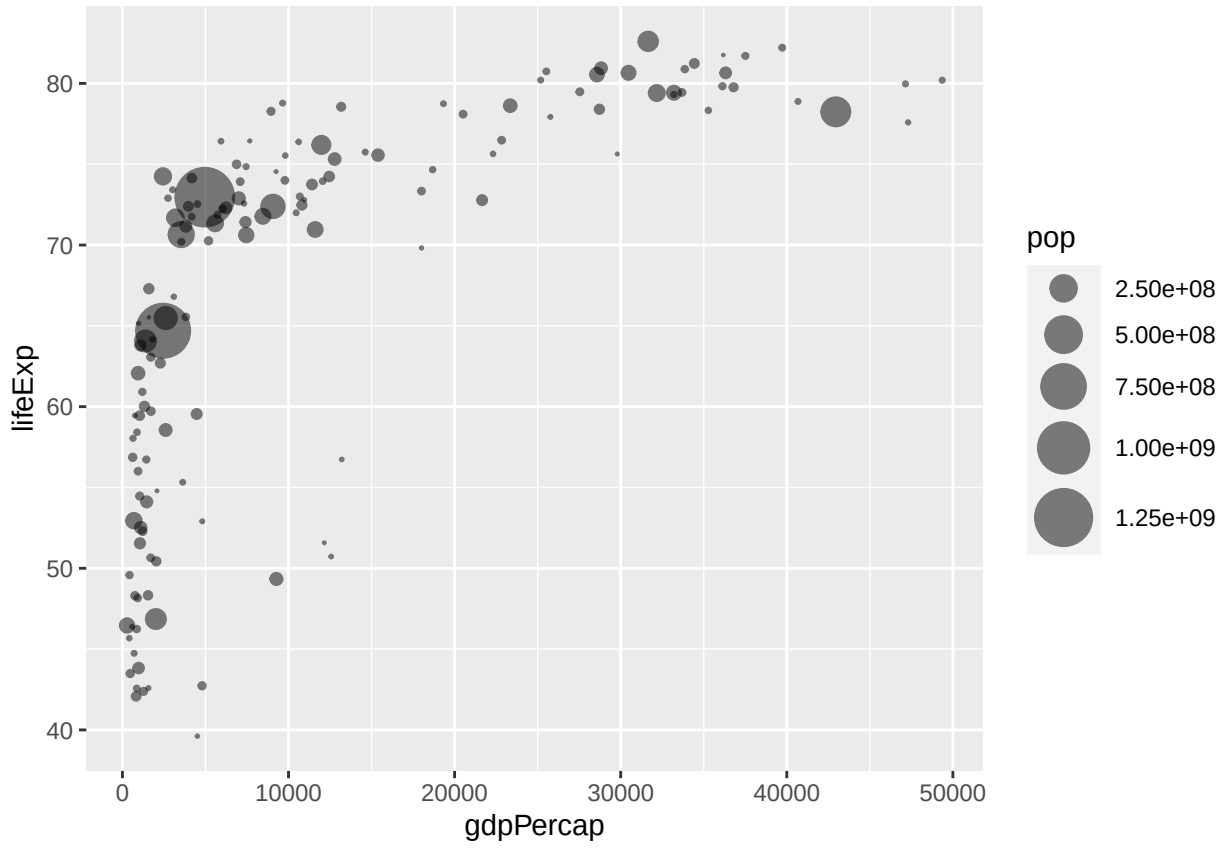

```
#Instead color by contrast
ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp, color = pop) +
  geom_point(alpha=0.8)
```



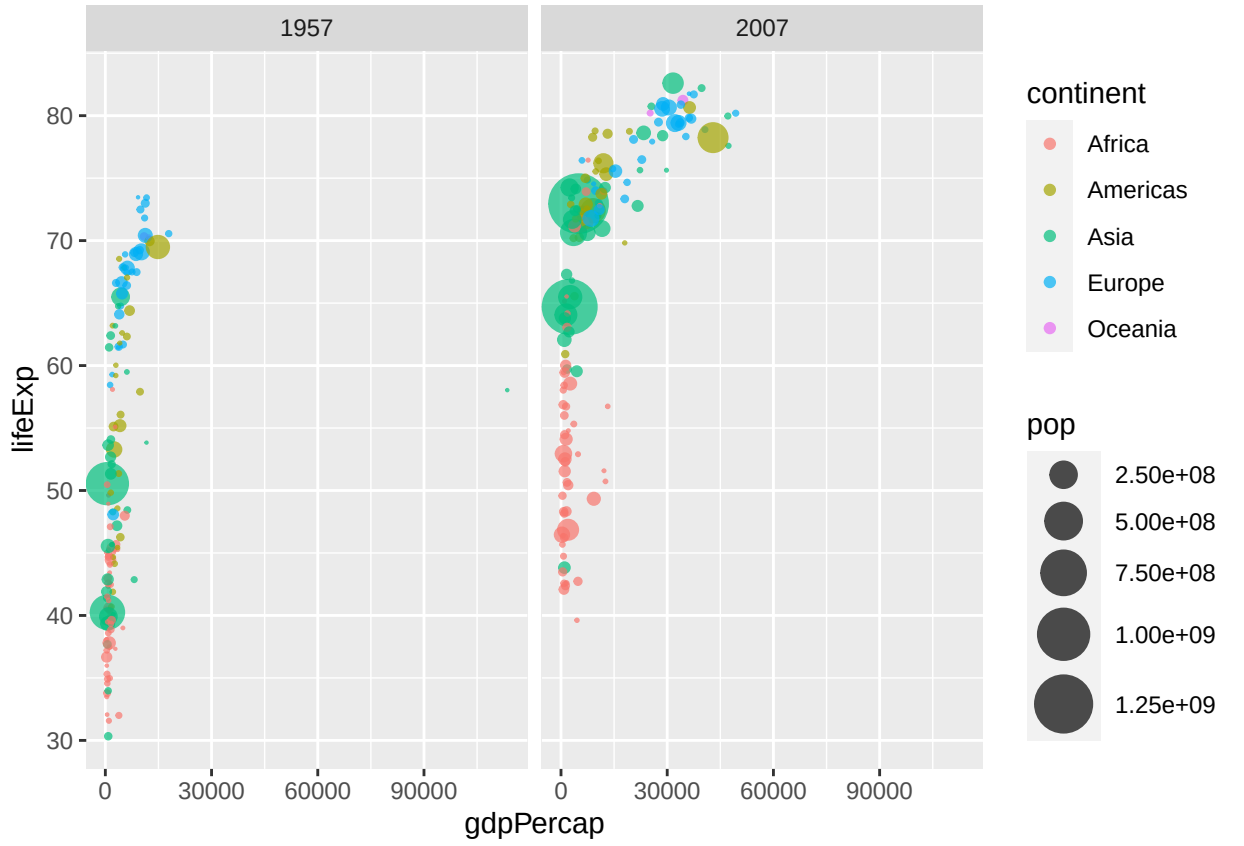

```
#Point size as a proportion to pop size
ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp, size = pop) +
  geom_point(alpha=0.5)
```




```
#A better more acurate point scaling using scale size function
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPercap, y = lifeExp,
                 size = pop), alpha=0.5) +
  scale_size_area(max_size = 10)
```



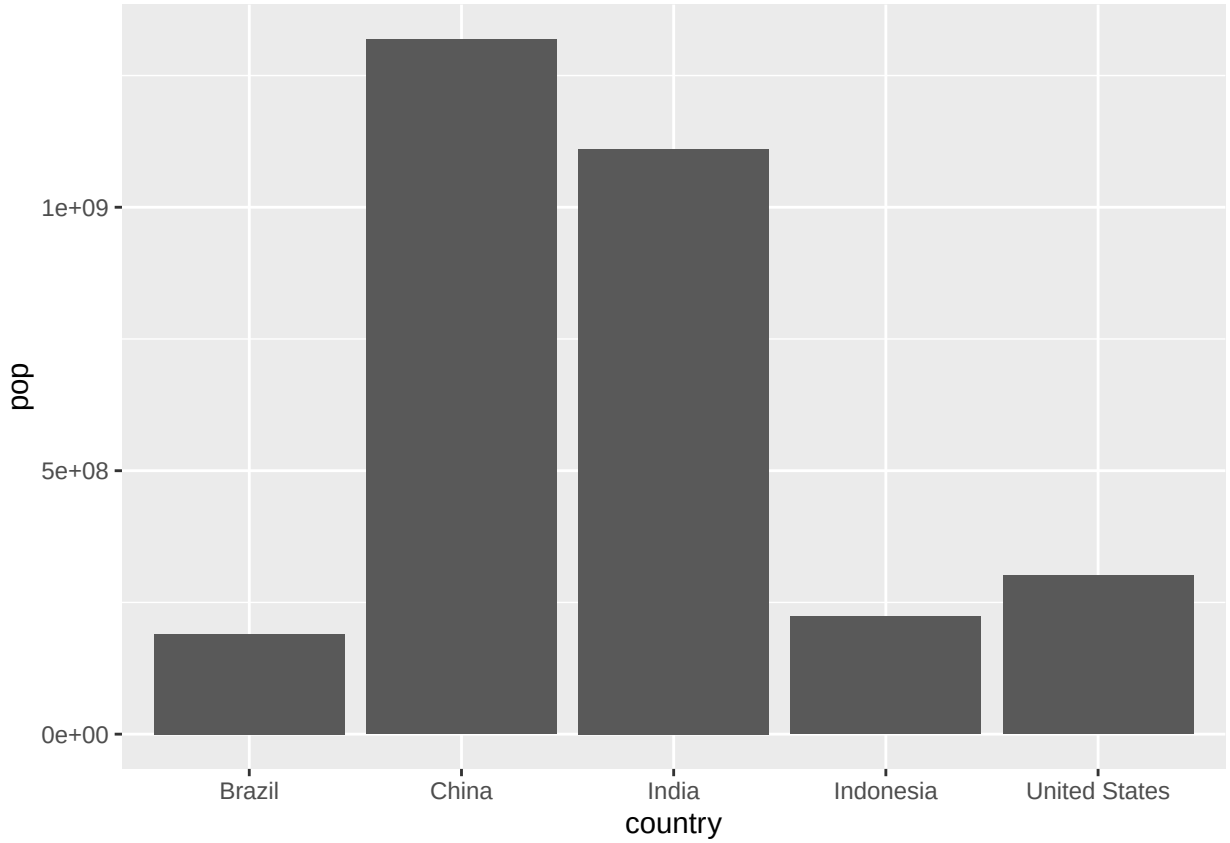

```
#Subset the gapminder dataset for 1957-2007
gapminder_1957 <- gapminder %>% filter(year==1957 | year==2007)

#Plot new data set with color for continent and size for population again
#add facet wrat to seperate 1957 vs 2007
ggplot(gapminder_1957) +
  geom_point(aes(x = gdpPercap, y = lifeExp, color=continent,
                 size = pop), alpha=0.7) +
  scale_size_area(max_size = 10) +
  facet_wrap(~year)
```



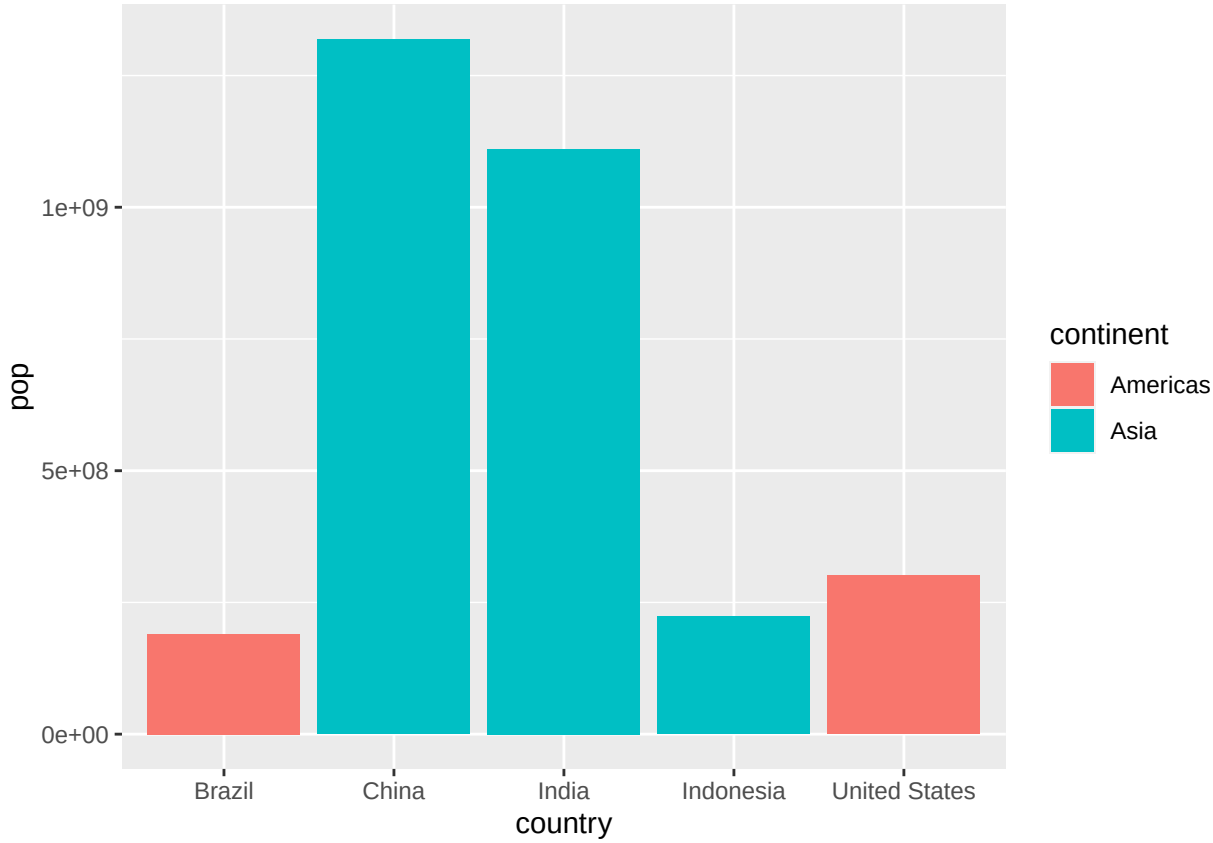

```
################################
# Barcharts

#create vector for the gapminder data set separating the largest population countries
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
```

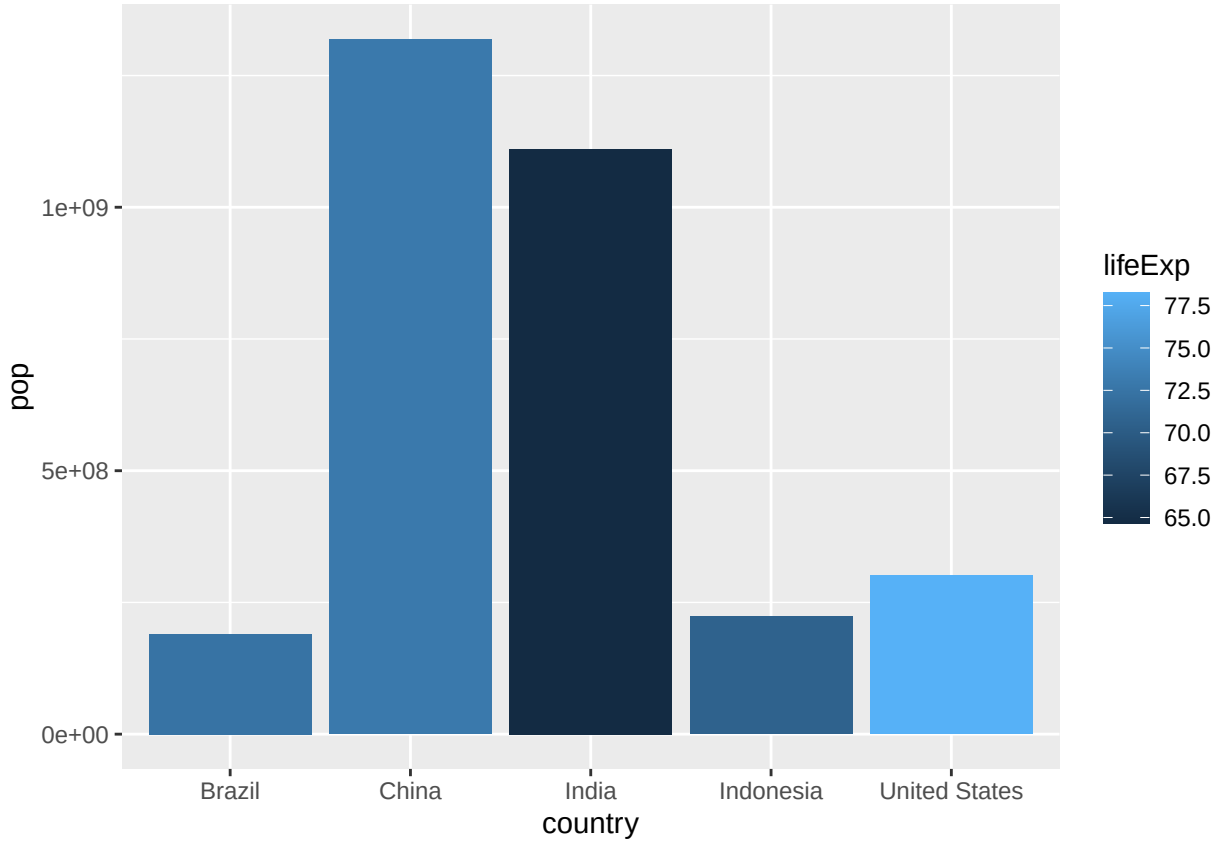
```
#basic barchart
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop))
```



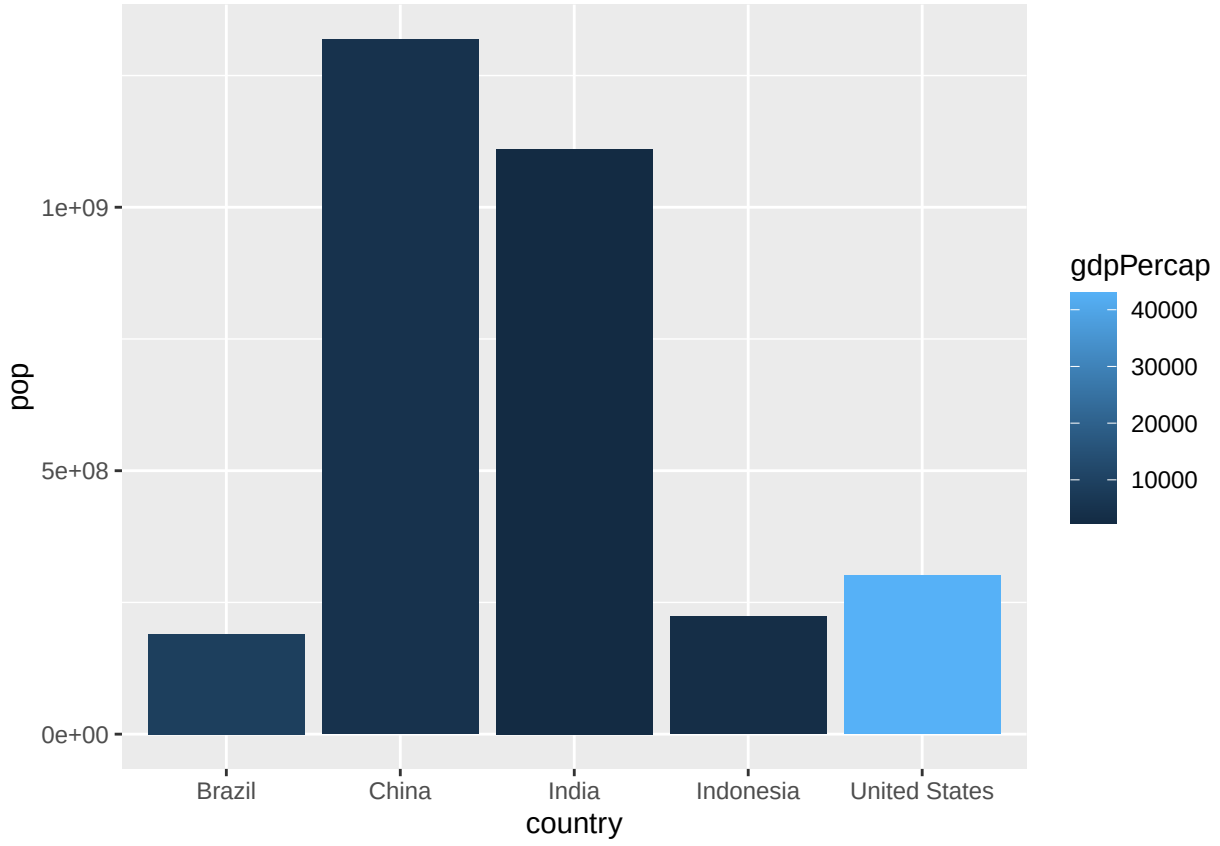

```
#Color!
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = continent))
```



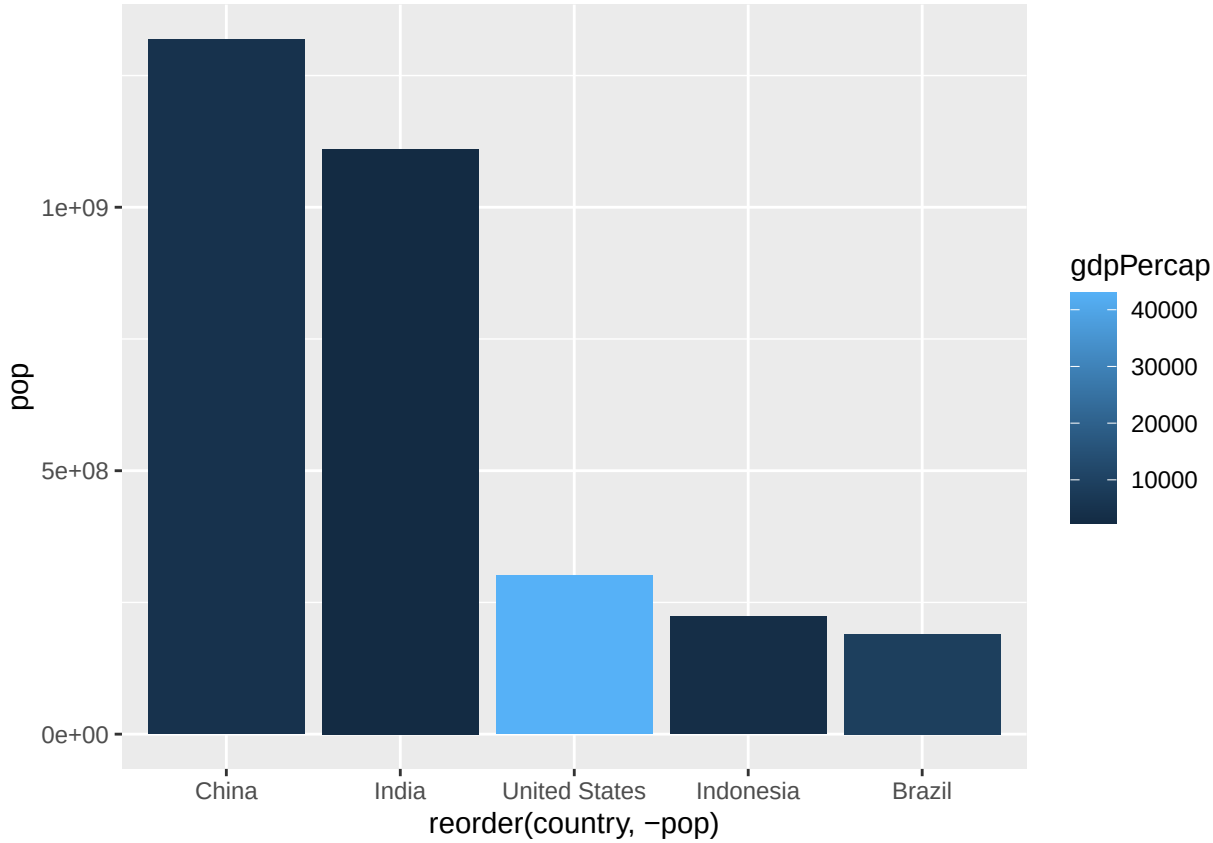

```
#color fill is a scale of the life expectancy
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = lifeExp))
```




```
#color fill as a scale for gdp percapita
ggplot(gapminder_top5) +
  aes(x=country, y=pop, fill=gdpPercap) +
  geom_col()
```

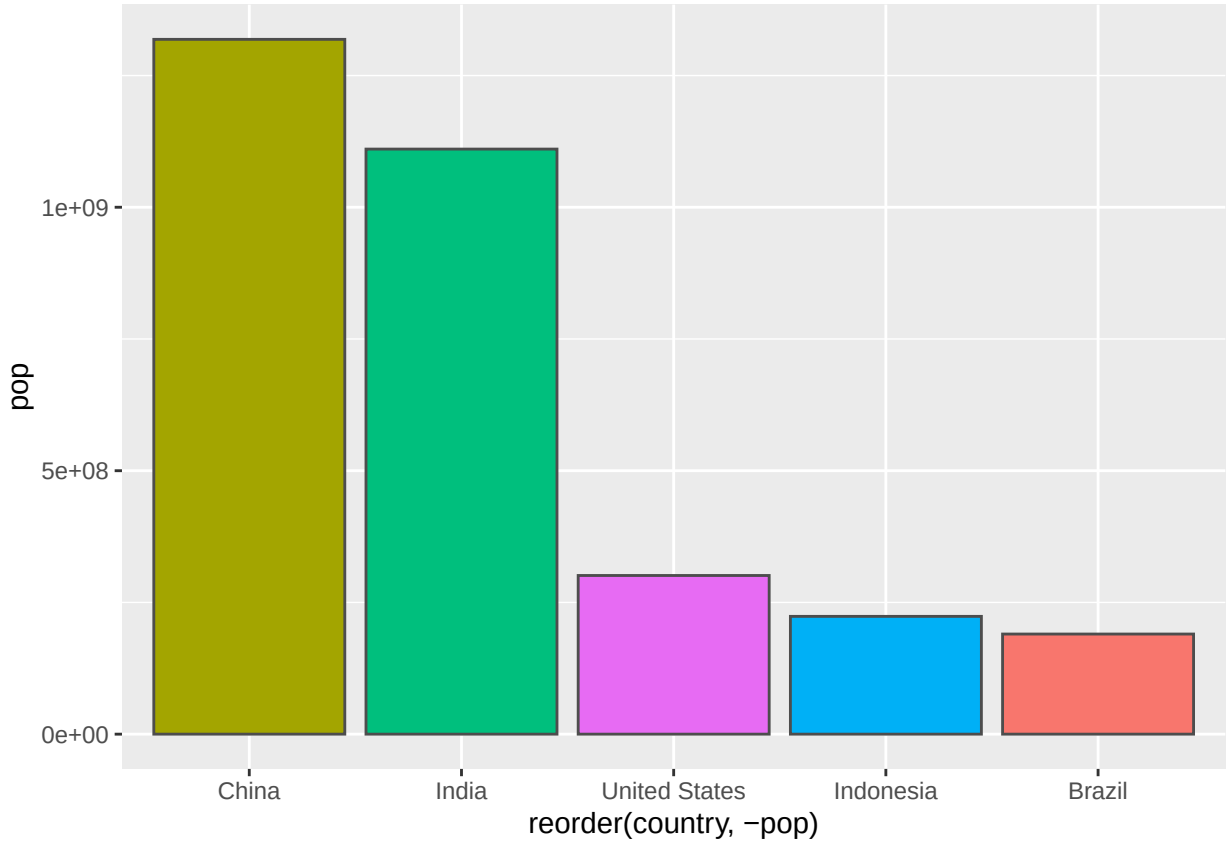



```
#Remove the alphabetical sorting and sort by largest to smallest pop
ggplot(gapminder_top5) +
  aes(x=reorder(country, -pop), y=pop, fill=gdpPercap) +
  geom_col()
```




```
#fill by country
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.
```


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