

Week 5: Data Visualization Lab

Divya Shetty (A15390408)

```
# installing ggplot package...
# install.packages("ggplot2")

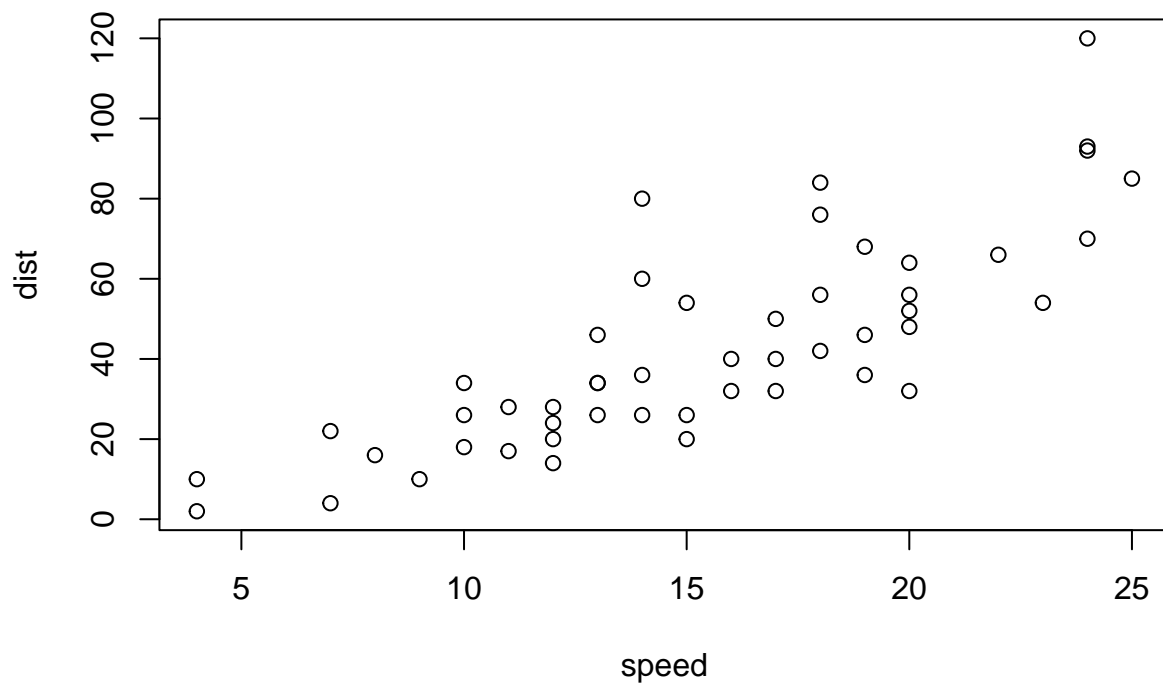
# loading ggplot package...
library(ggplot2)

#1: which plot types are typically NOT used to compare distributions of numeric variables?
#NETWORK GRAPHS

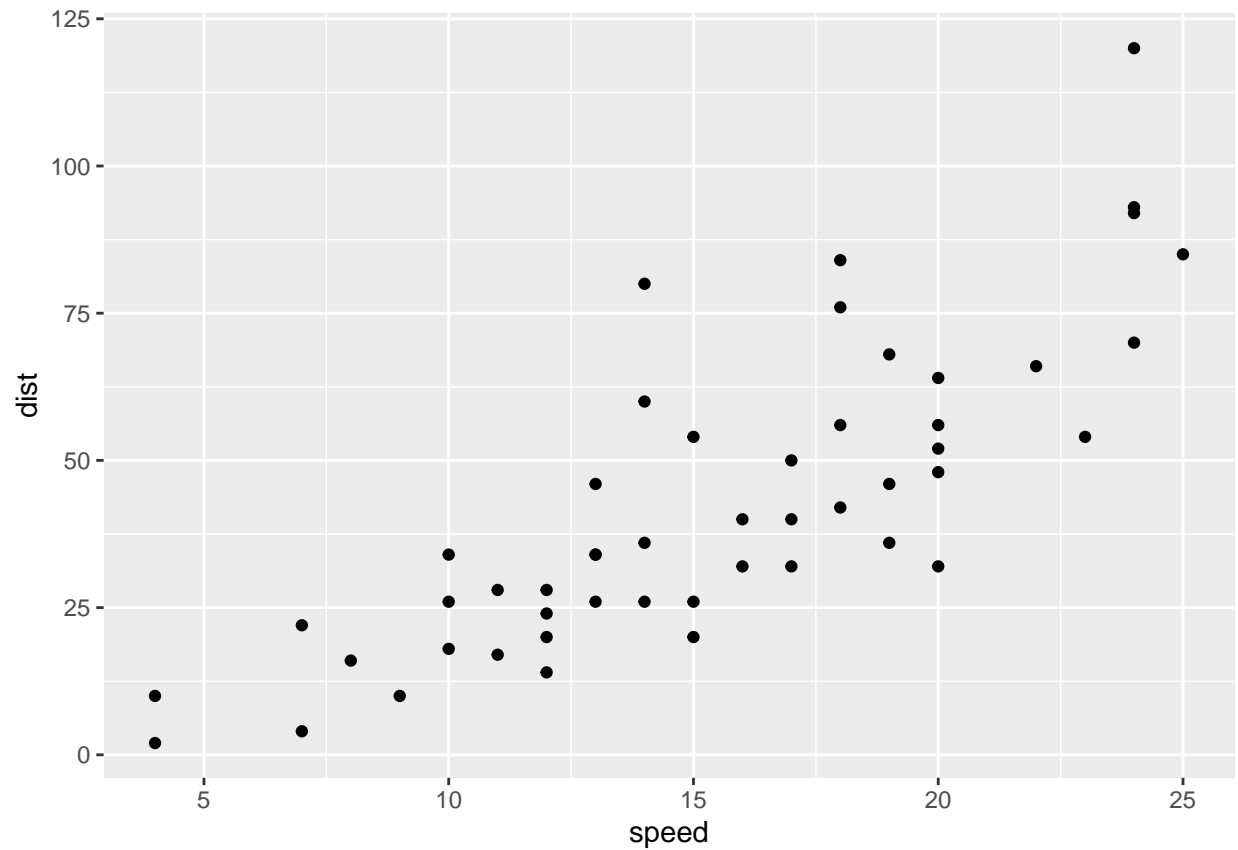
#2: which statement about ggplot2 with R is incorrect?
#GGPLOT2 IS THE ONLY WAY TO CREATE PLOTS IN R

#-----CARS DATASET-----

# base R plot - NOT ggplot
plot(cars)
```

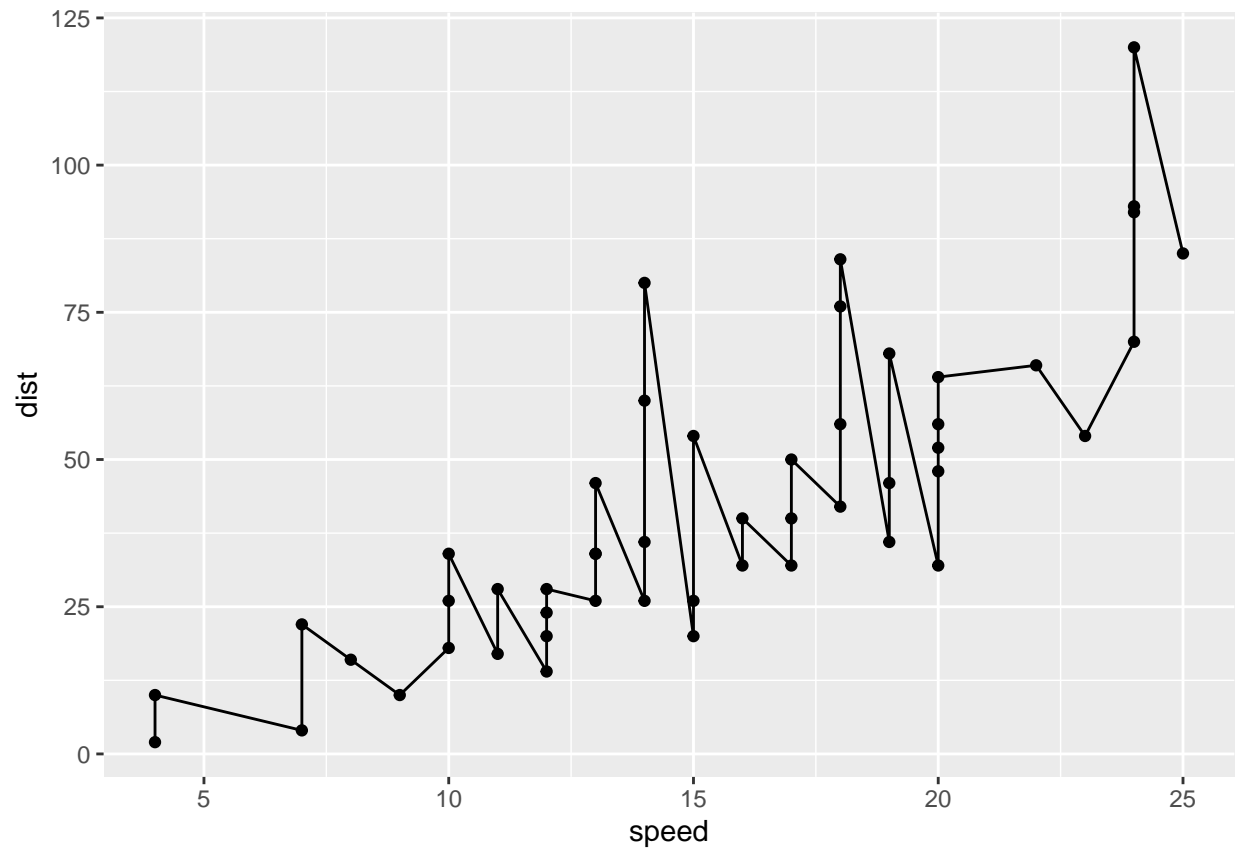


```
# this IS a ggplot
ggplot(data = cars) +
  aes(x = speed, y = dist) +
  geom_point()
```



```
#same ggplot, stored in variable
p <- ggplot(data = cars) +
  aes(x = speed, y = dist) +
  geom_point()

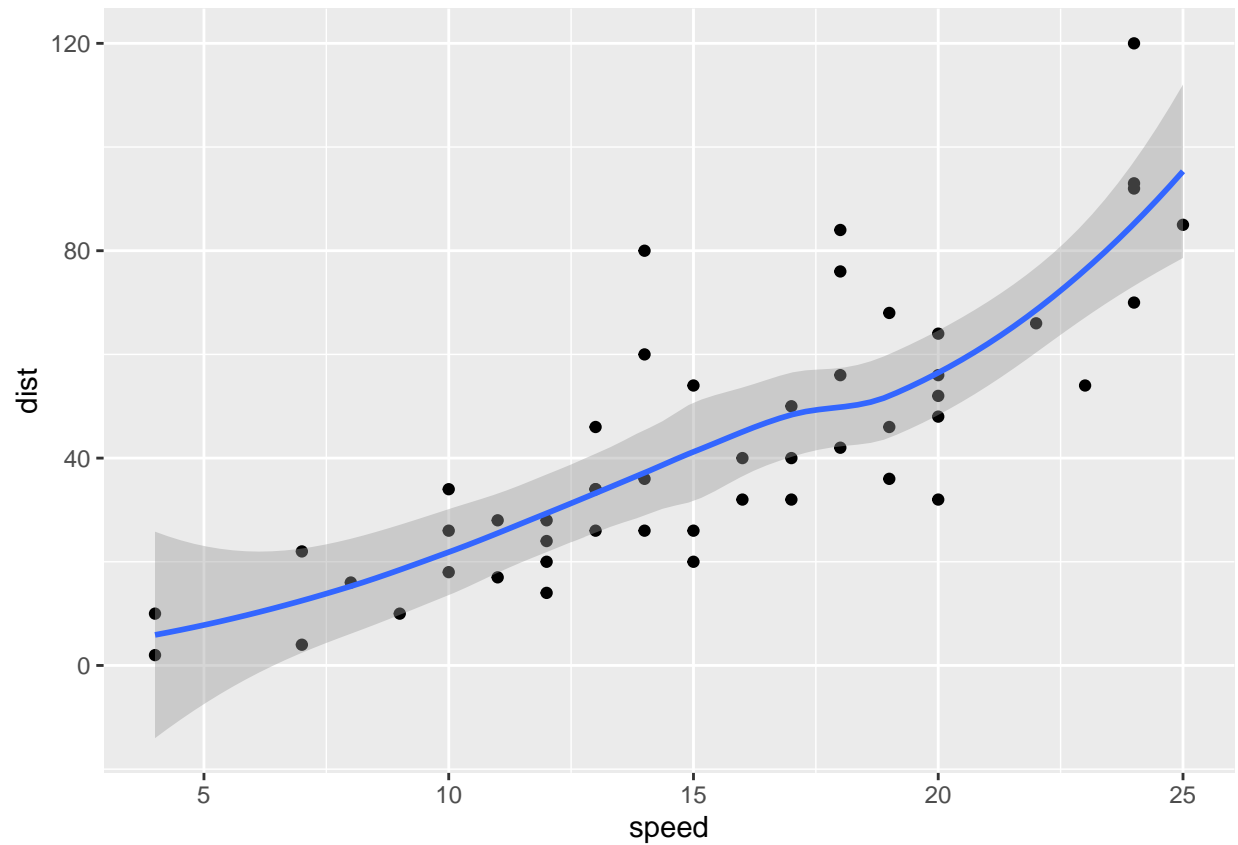
#add line geom
p + geom_line()
```



```
#3: which geom layer should be used to create scatter plots in ggplot2?  
#geom_point()
```

```
#4: add a trendline close to data  
p + geom_smooth()
```

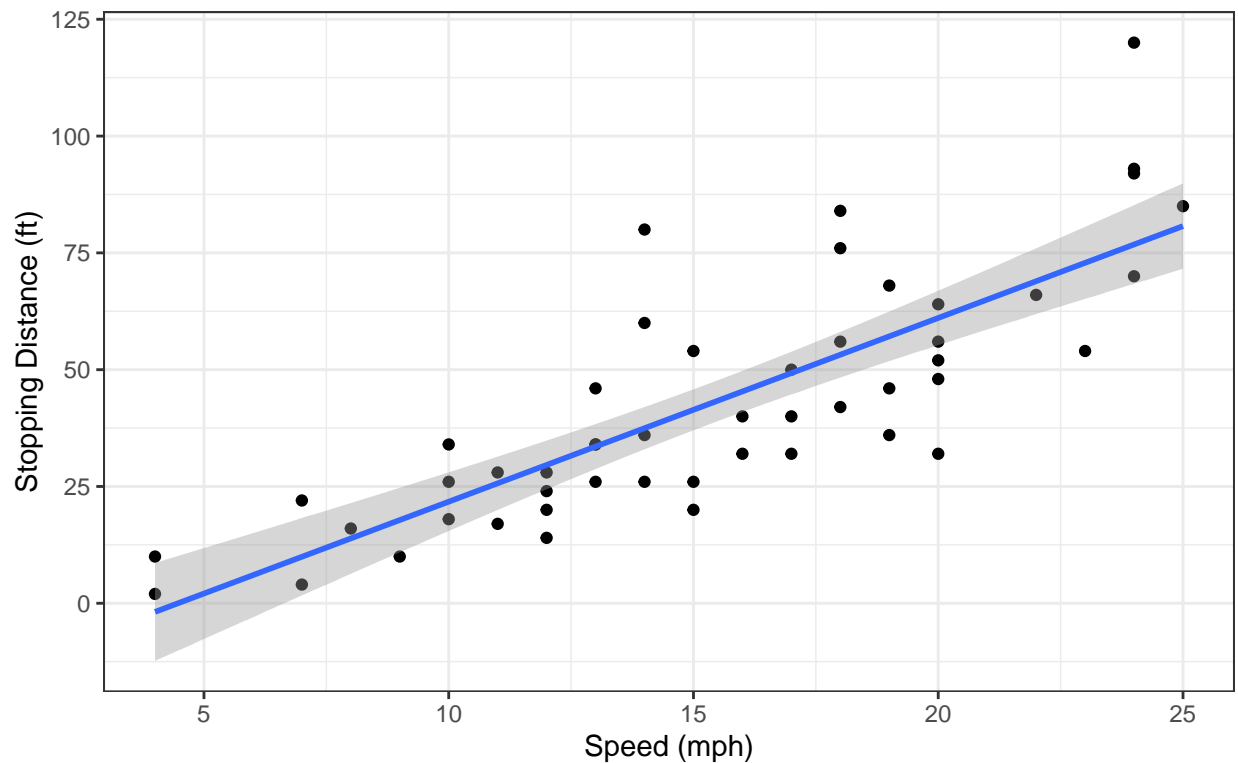
```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



```
#5: add labels and a theme
p + labs(title = "Speed and Stopping Dist. of Cars",
         x = "Speed (mph)", y = "Stopping Distance (ft)",
         caption = "Dataset: 'cars'") +
  geom_smooth(method = "lm") +
  theme_bw()
```

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

Speed and Stopping Dist. of Cars



Dataset: 'cars'

#-----GENES DATASET-----

#read drug expression data

```
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
```

#6: how many genes are in this dataset?

```
nrow(genes)
```

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

#7: num of col and col names

```
ncol(genes)
```

```
## [1] 4
```

```
colnames(genes)
```

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

```
#8: how many up-regulated genes are there?
```

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

```
##
```

```
##      down  unchanging      up
```

```
##      72      4997      127
```

```
#9: what fraction of total genes are up-regulated?
```

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

```
##
```

```
##      down  unchanging      up
```

```
##      1.39      96.17      2.44
```

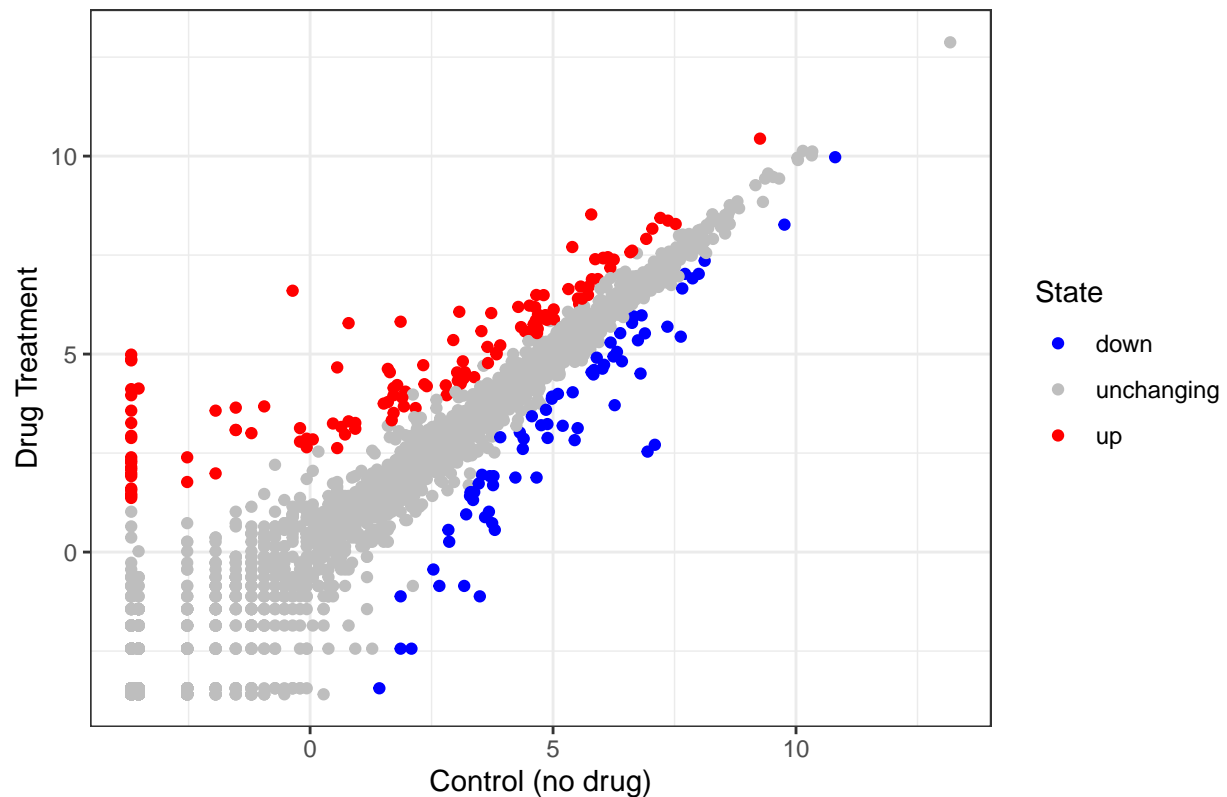
```
#10: ggplot for genes
```

```
g <- ggplot(data = genes) +  
  aes(x = Condition1, y = Condition2, color = State) +  
  geom_point()
```

```
#11: add color and labels
```

```
g + scale_color_manual(values = c("blue", "gray", "red")) +  
  labs(title = "Gene Expression Changes with Drug Treatment",  
       x = "Control (no drug)", y = "Drug Treatment") +  
  theme_bw()
```

Gene Expression Changes with Drug Treatment

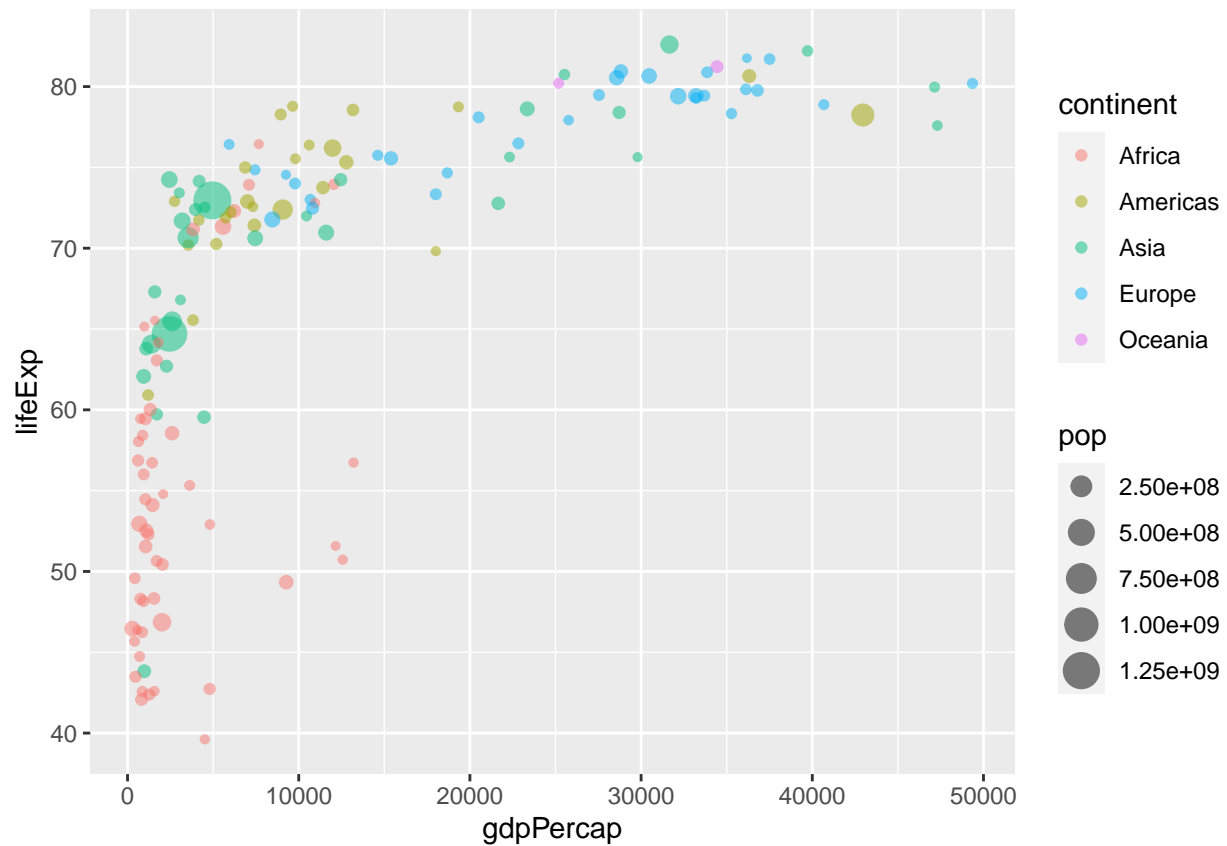


```
#-----GAPMINDER DATASET-----  
  
# File location online  
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"  
  
gapminder <- read.delim(url)  
  
#installing dplyr package...  
#install.packages("dplyr")  
  
library(dplyr)  
  
##  
## Attaching package: '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)
```

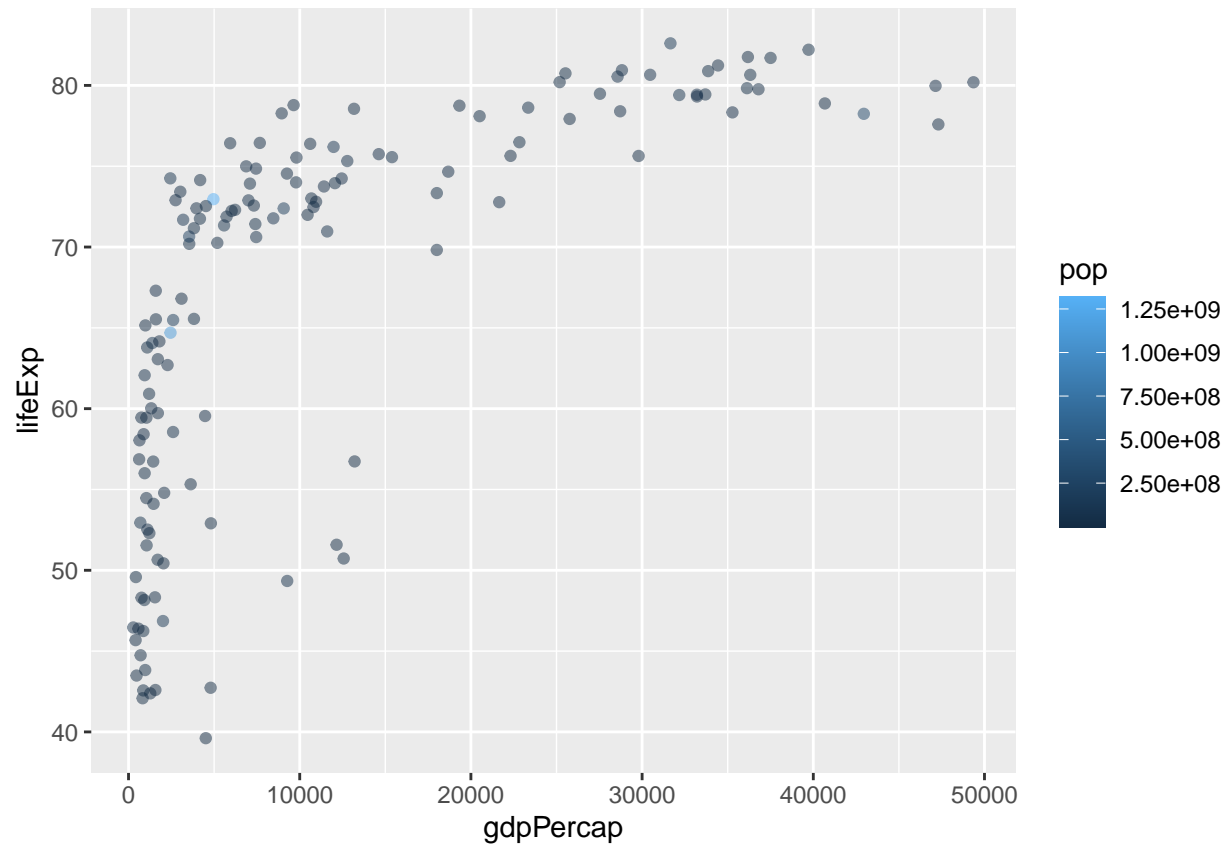
```
#12: ggplot for gapminder
```

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

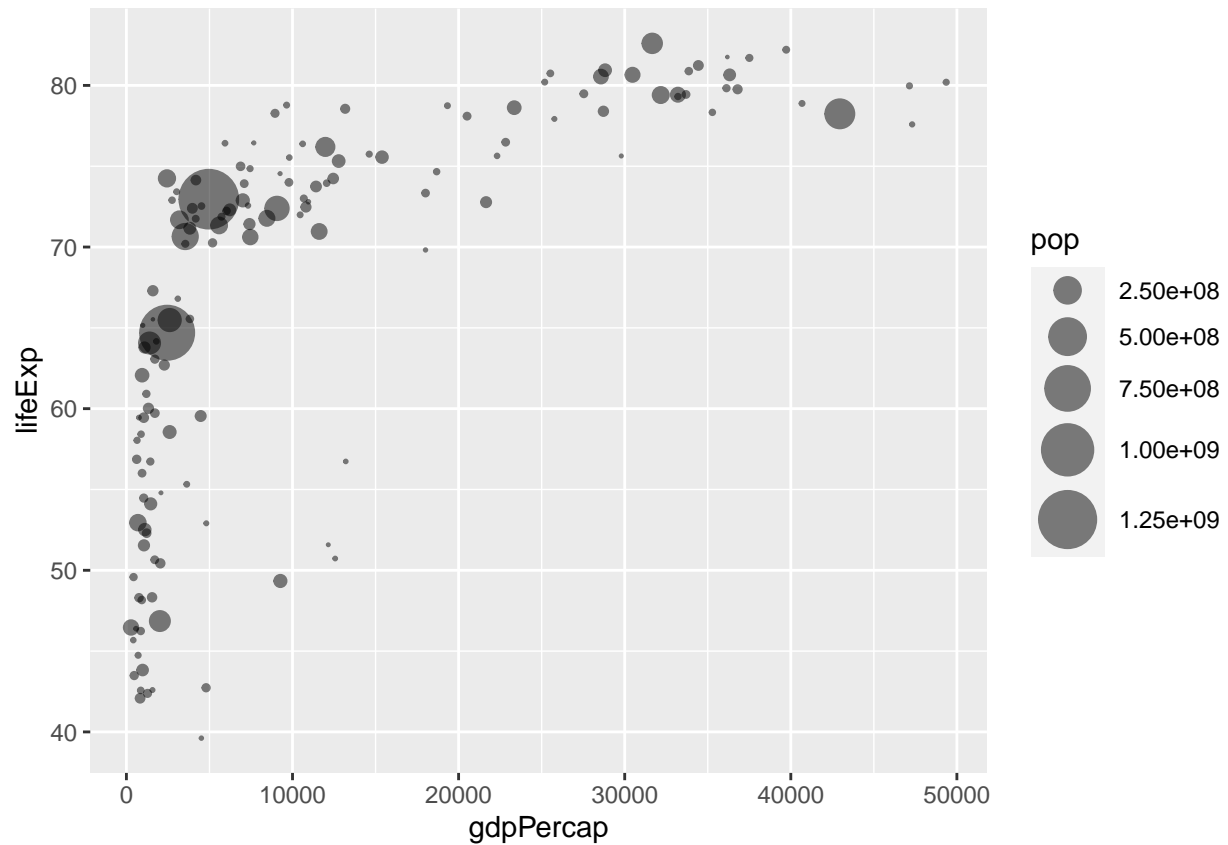


```
#color by population
```

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

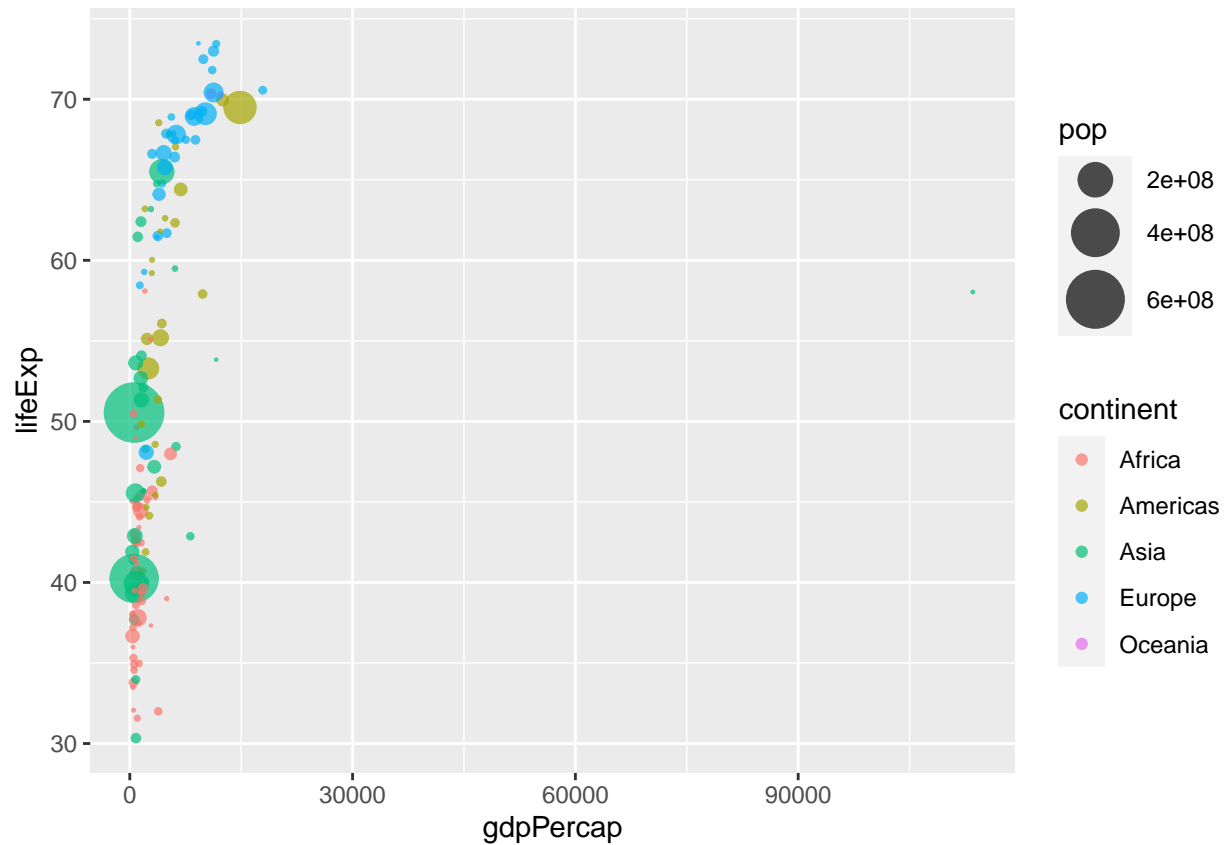



```
#scaling point size  
ggplot(gapminder_2007) +  
  geom_point(aes(x = gdpPerCap, y = lifeExp, size = pop), alpha = 0.5) +  
  scale_size_area(max_size = 10)
```



```
#13: ggplot for 1957 data, combined with 2007
gapminder_1957 <- gapminder %>% filter(year==1957)
```

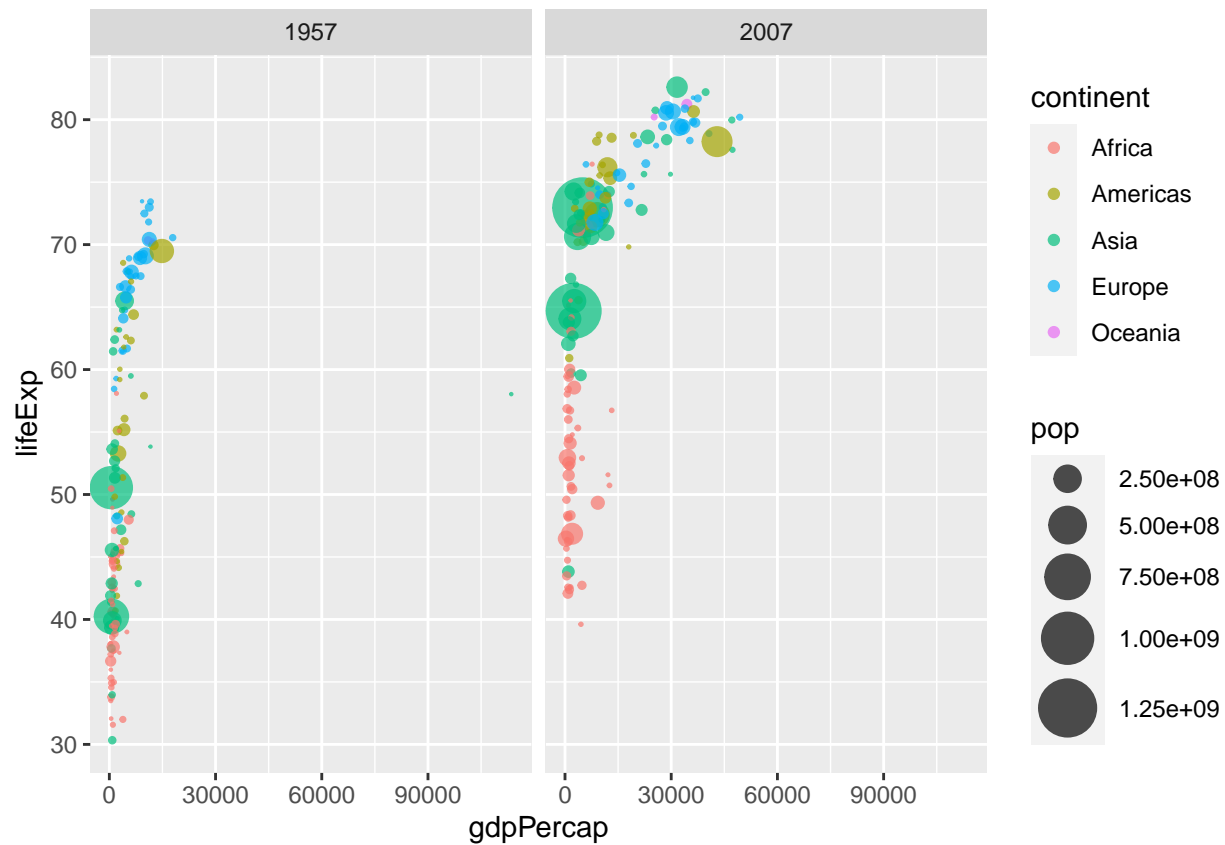
```
ggplot(gapminder_1957) +
  aes(x = gdpPercap, y = lifeExp, color = continent, size = pop) +
  geom_point(alpha = 0.7) +
  scale_size_area(max_size = 10)
```



#14: combine 1957 and 2007!

```
gapminder_1957_2007 <- gapminder %>% filter(year==1957 | year == 2007)
```

```
ggplot(gapminder_1957_2007) +  
  aes(x = gdpPercap, y = lifeExp, color = continent, size = pop) +  
  geom_point(alpha = 0.7) +  
  scale_size_area(max_size = 10) +  
  facet_wrap(~year)
```

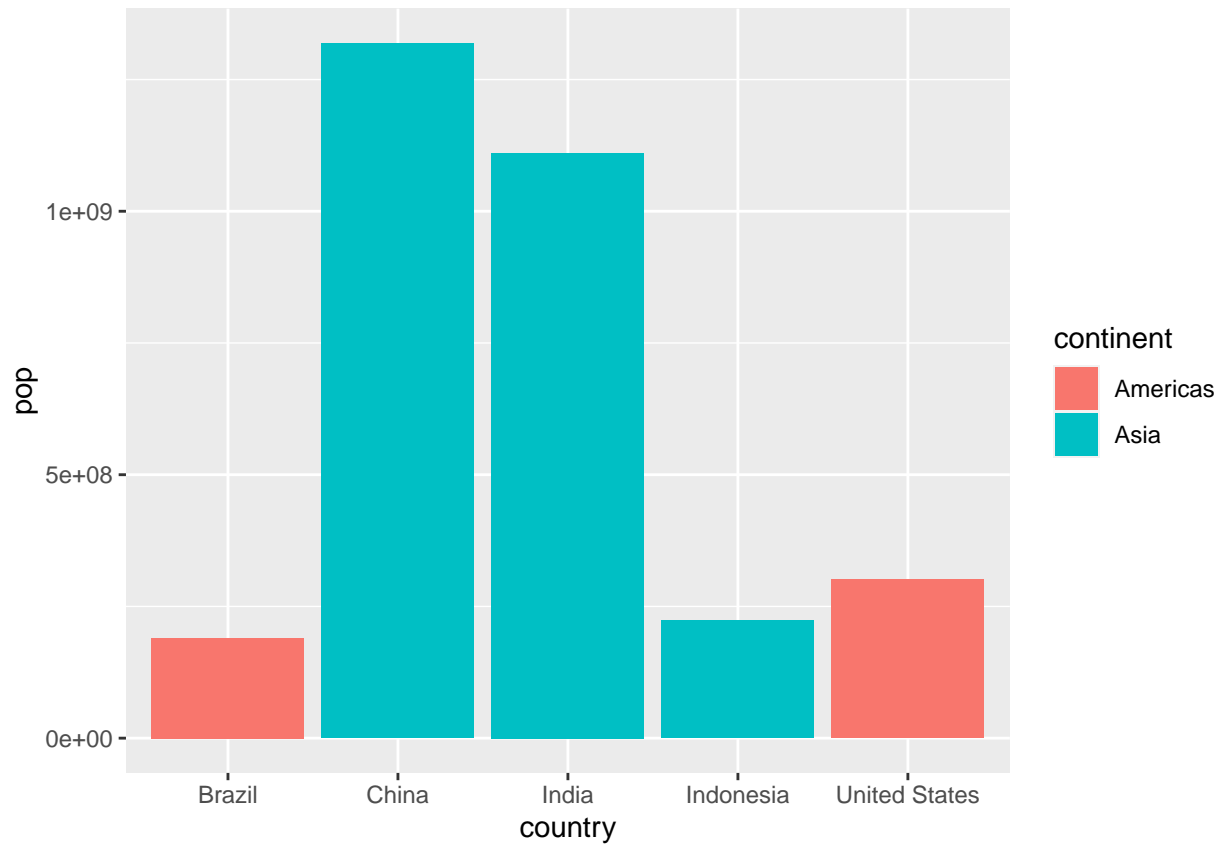


#-----BAR CHARTS-----

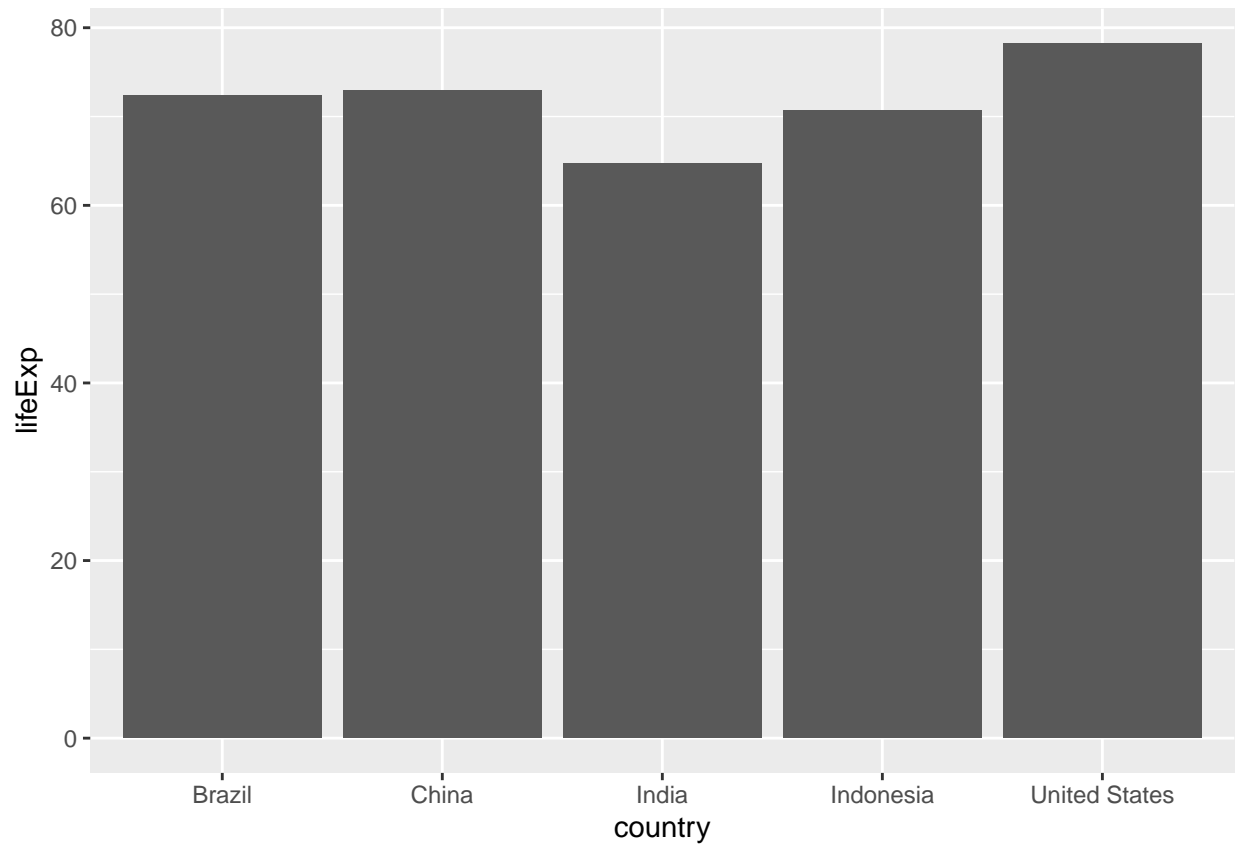
```
gapminder_top5 <- gapminder %>%
  filter(year==2007) %>%
  arrange(desc(pop)) %>%
  top_n(5, pop)
```

#making a bar chart, pop. by country

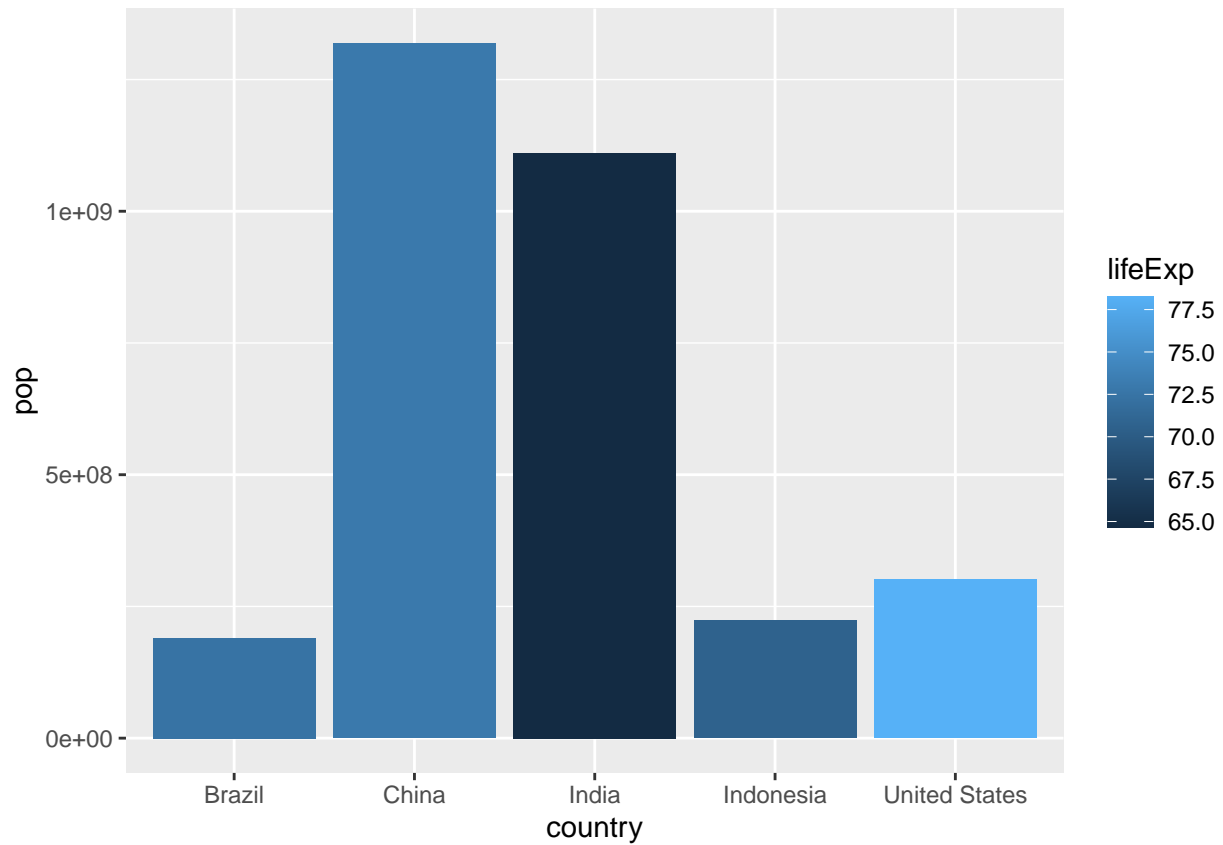
```
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = continent))
```



```
#15: another bar chart, life exp. by country  
ggplot(gapminder_top5) +  
  geom_col(aes(x = country, y = lifeExp))
```



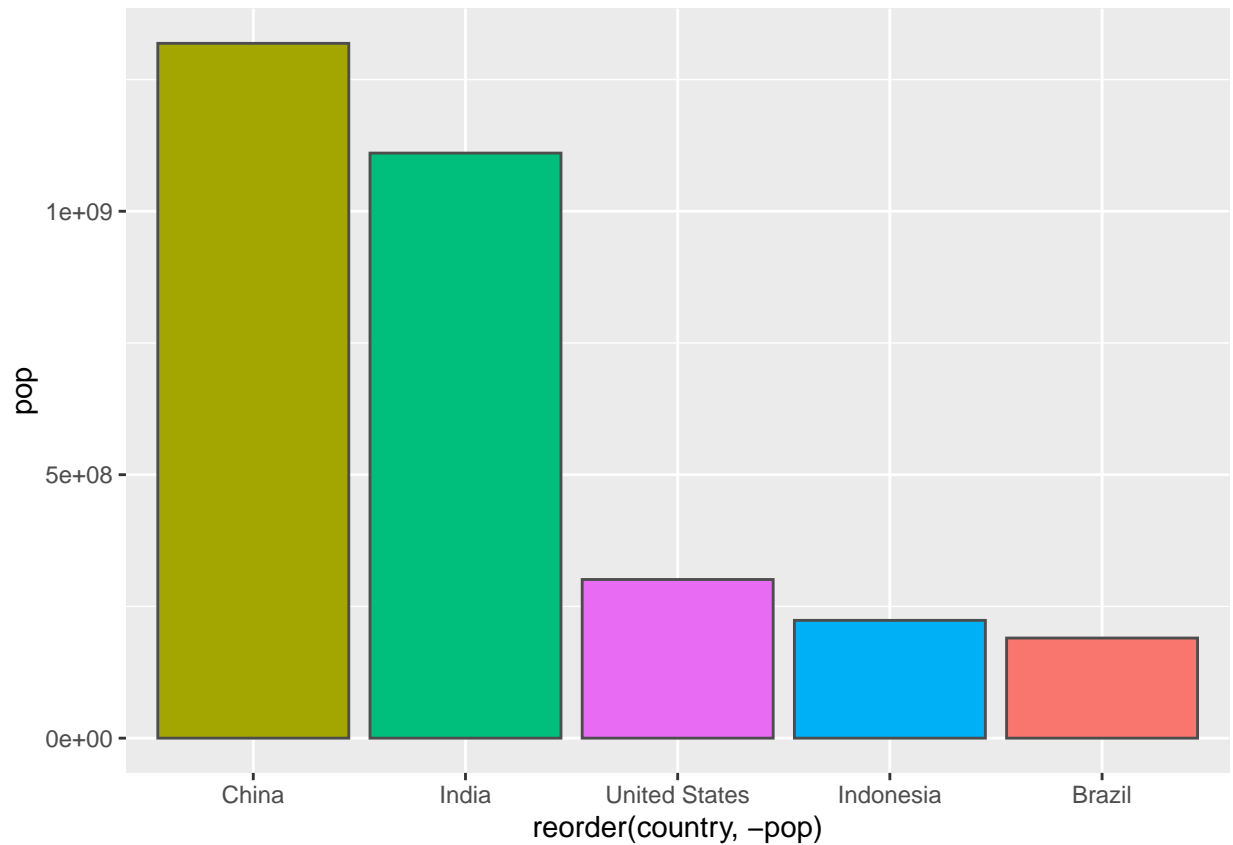
```
#color by lifeExp instead of continent  
ggplot(gapminder_top5) +  
  geom_col(aes(x = country, y = pop, fill = lifeExp))
```



#16: pop. size by country

```
ggplot(gapminder_top5) +  
  aes(x = reorder(country, -pop), y = pop, fill = country) +  
  geom_col(color = "gray30") +  
  guides(fill = FALSE)
```

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

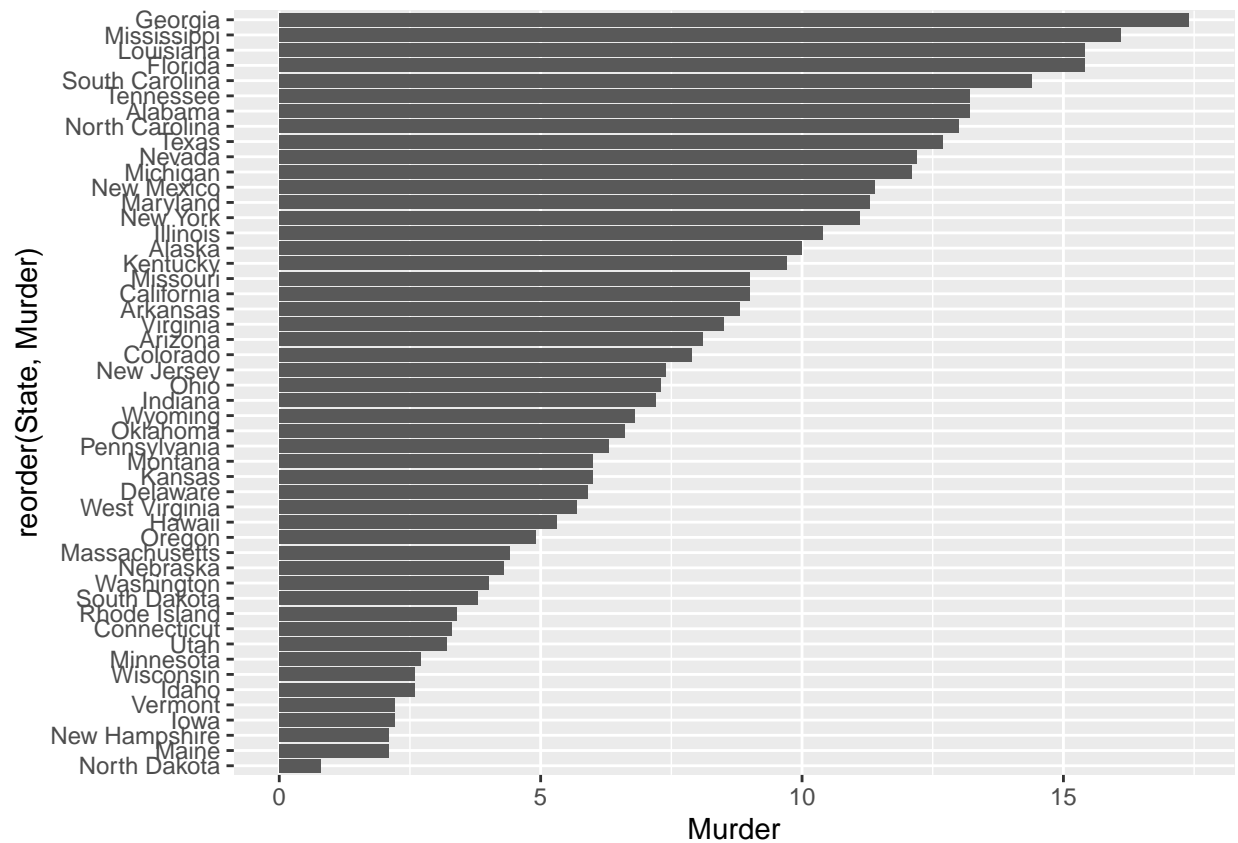


#flipping a bar chart:
 head(USArrests)

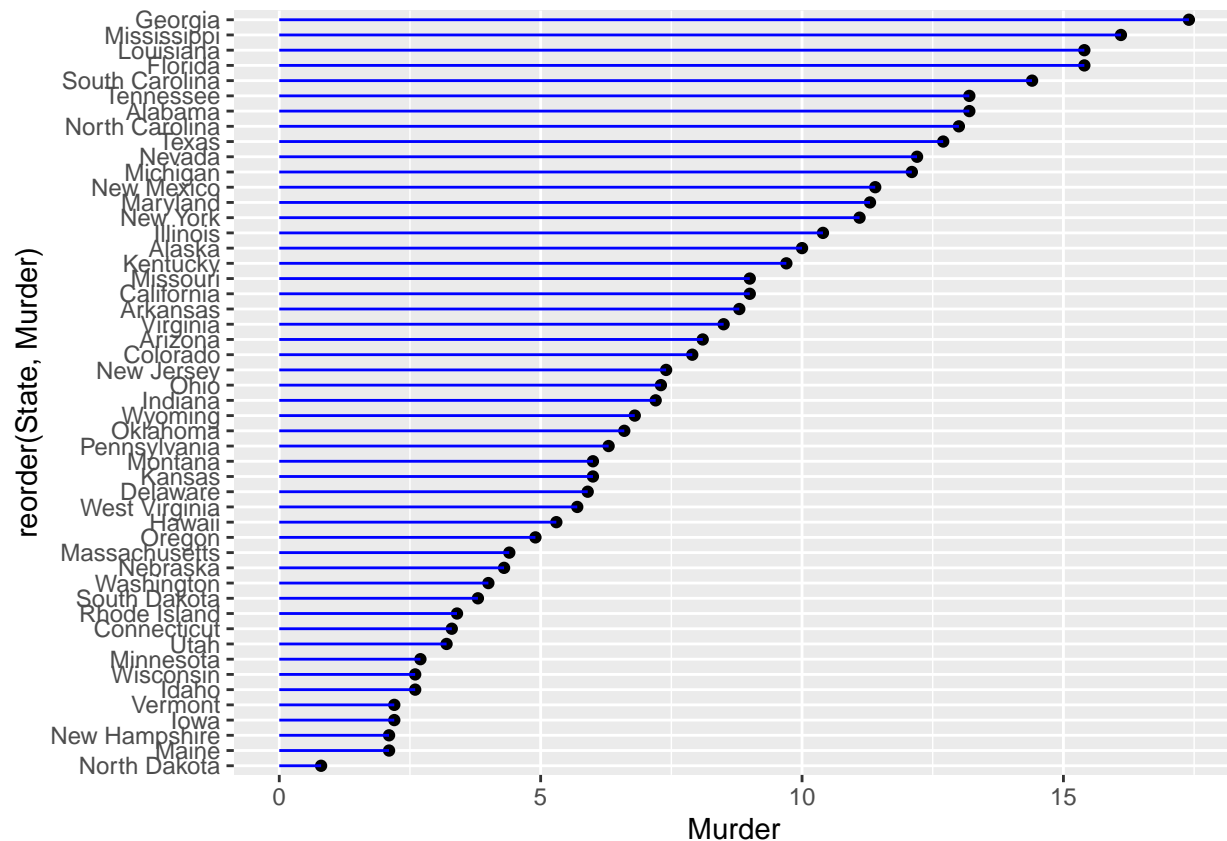
```
##           Murder Assault UrbanPop Rape
## Alabama      13.2     236      58 21.2
## Alaska       10.0     263      48 44.5
## Arizona       8.1     294      80 31.0
## Arkansas      8.8     190      50 19.5
## California    9.0     276      91 40.6
## Colorado      7.9     204      78 38.7
```

```
USArrests$State <- rownames(USArrests)

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

```
#using geom_segment() for cleaner visualization
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()
```



```
#-----PLOT ANIMATION-----
```

```
library(gapminder)
```

```
##
```

```
## Attaching package: 'gapminder'
```

```
## The following object is masked _by_ '.GlobalEnv':
```

```
##
```

```
## gapminder
```

```
library(gganimate)
```

```
###normal ggplot set-up
```

```
#ggplot(gapminder, aes(x = gdpPercap, y = lifeExp, size = pop, color = 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 by continent
```

```
# 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)
```

```
#-----COMBINING PLOTS-----
```

```
library(patchwork)
```

```
# some example plots
```

```
p1 <- ggplot(mtcars) + geom_point(aes(mpg, disp))
```

```
p2 <- ggplot(mtcars) + geom_boxplot(aes(gear, disp, group = gear))
```

```
p3 <- ggplot(mtcars) + geom_smooth(aes(dis, qsec))
```

```
p4 <- ggplot(mtcars) + geom_bar(aes(carb))
```

```
# patchwork to combine:
```

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

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

