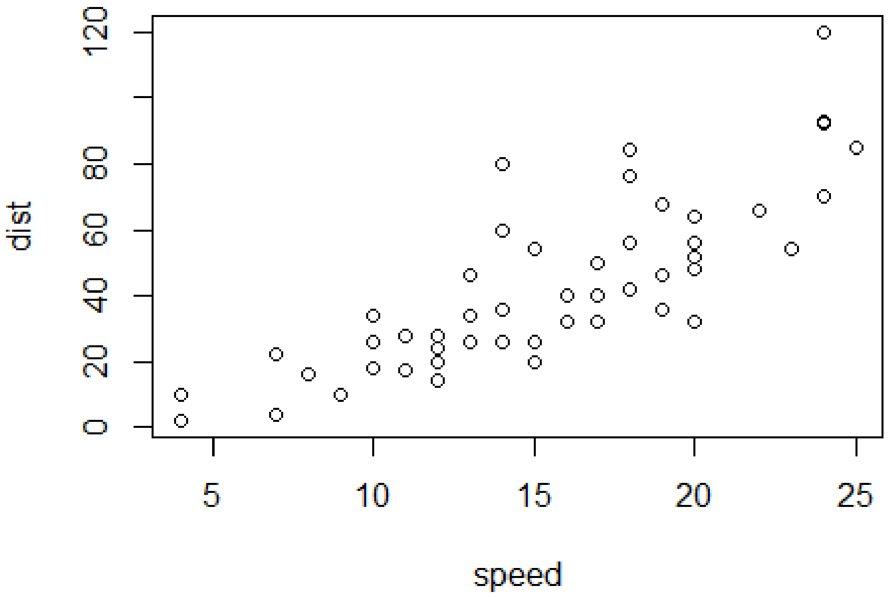
# Class 5 Data Visualization Lab

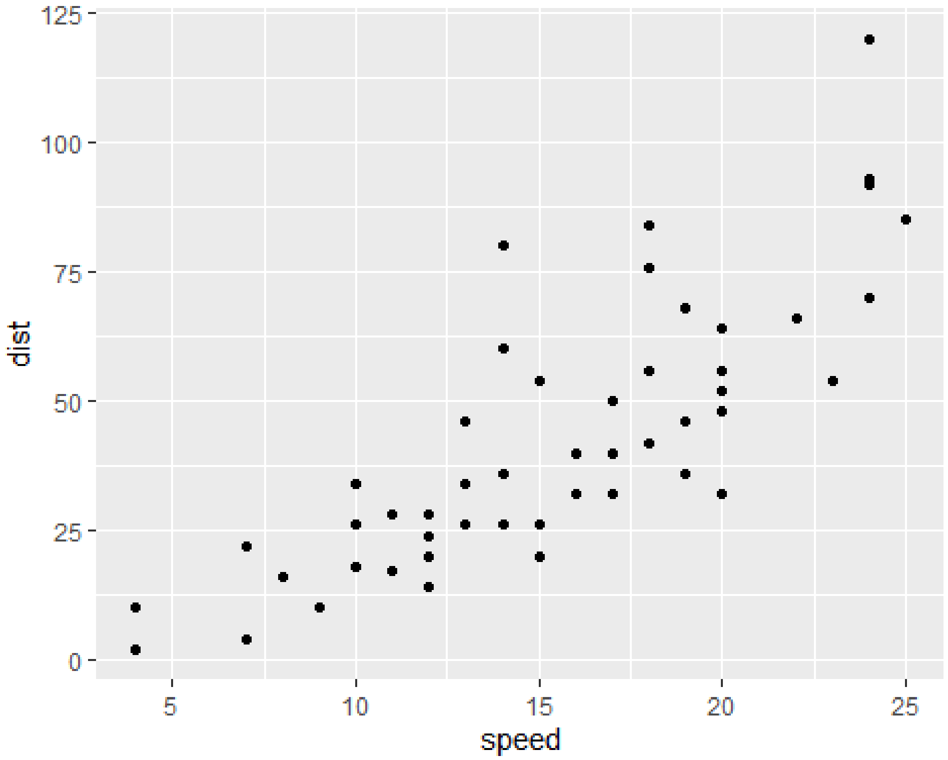
Cindy Tran (A15830581)

Winter 2022

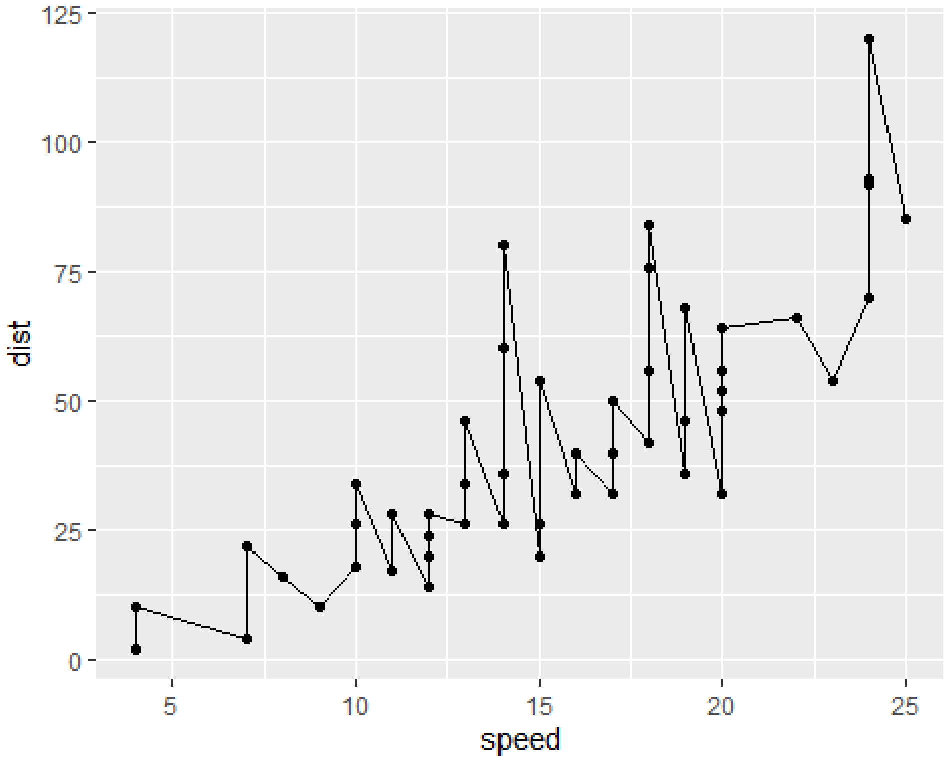
```
# Week 4 Data Visualization Lab

# Install the package ggplot2
#install.packages("ggplot2")

# Any time I want to use this package, I need to load it
library(ggplot2)

View(cars)

# A quick baseR plot - this is not ggplot
plot(cars)
```



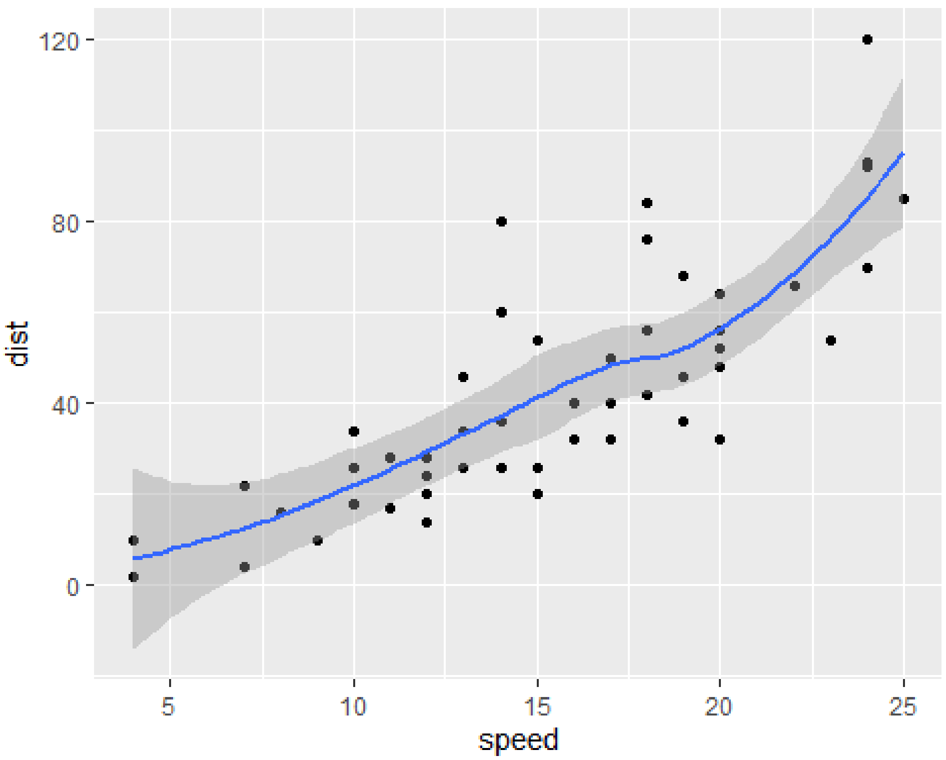
```
# Our first ggplot
#We need data + aes + geom
ggplot(data = cars) +
  aes(x = speed, y = dist) +
  geom_point()
```



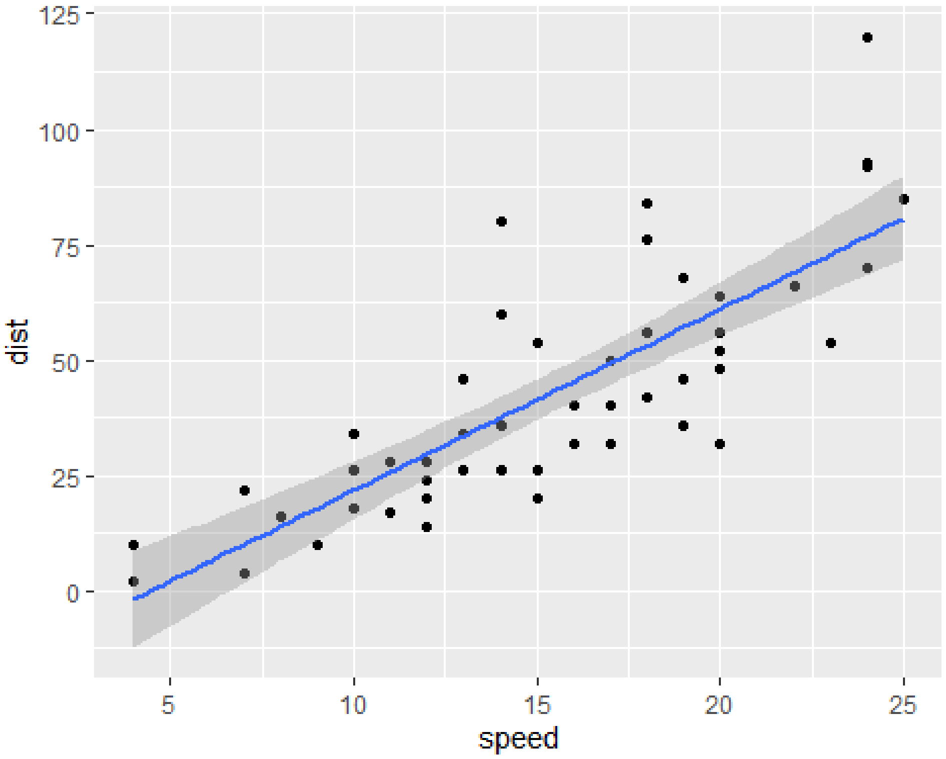
```
p <- ggplot(data = cars) +
  aes(x = speed, y = dist) +
  geom_point()

# Add a line geom with geom_line()
p + geom_line()
```



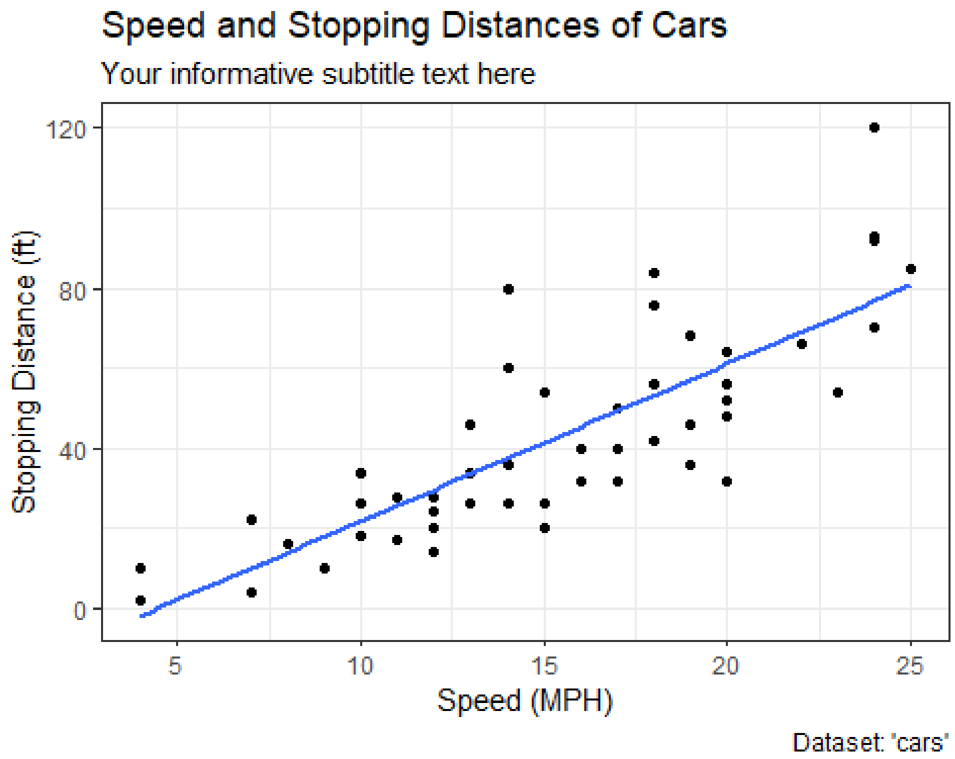
```
# Add a trend line close to the data

p + geom_smooth()

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



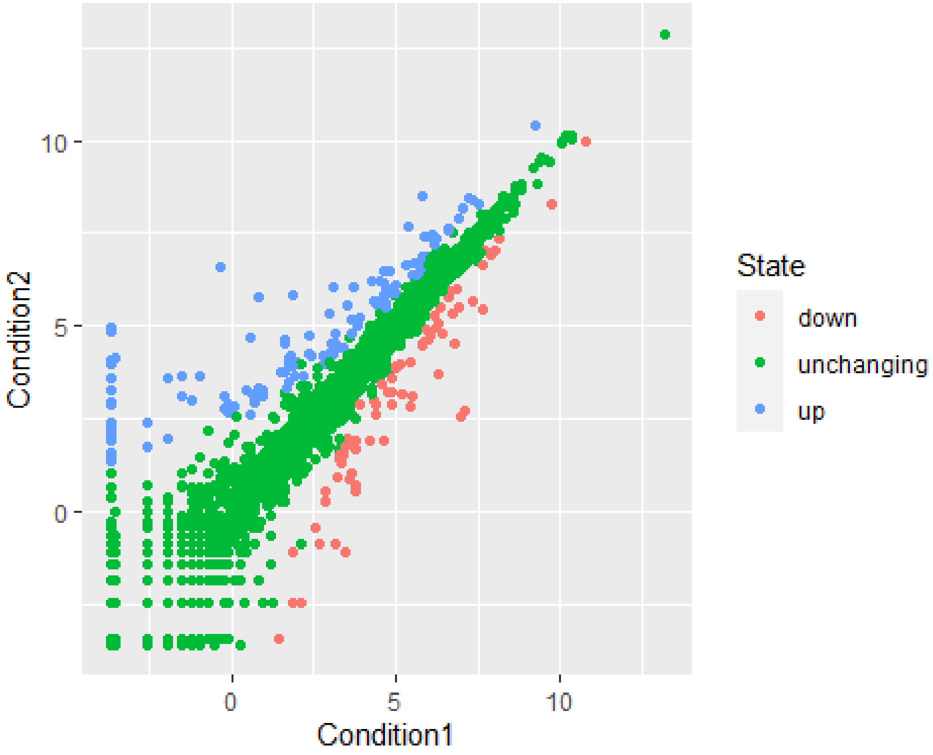
```
p + geom_smooth(method = "lm")

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



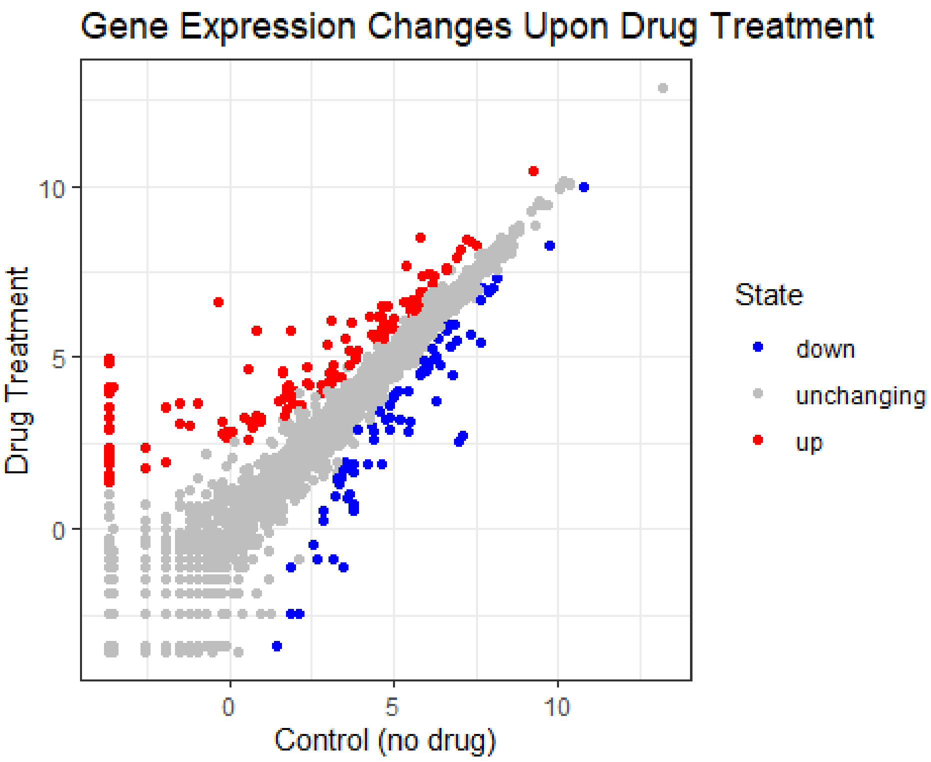
```
# Adding labels
p + geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Speed and Stopping Distances of Cars",
       x = "Speed (MPH)",
       y = "Stopping Distance (ft)",
       subtitle = "Your informative subtitle text here",
       caption = "Dataset: 'cars'") +
  theme_bw()

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



```
##

#Read in drug expression data

url <- "https://bioboot.github.io/bimm143_S20/class-material/up_down_expressi
on.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
nrow(genes)

## [1] 5196

# Column names and number of columns
colnames(genes)

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

ncol(genes)

## [1] 4

#How many upreglated genes
table(genes$State)

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

#Fraction of genes up-regulated
round ( (table(genes$State) / nrow(genes)) * 100, 2)

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

# Let's make a first plot attempt
ggplot(data = genes) +
  aes(x = Condition1, y = Condition2, col = State) +
  geom_point()
```

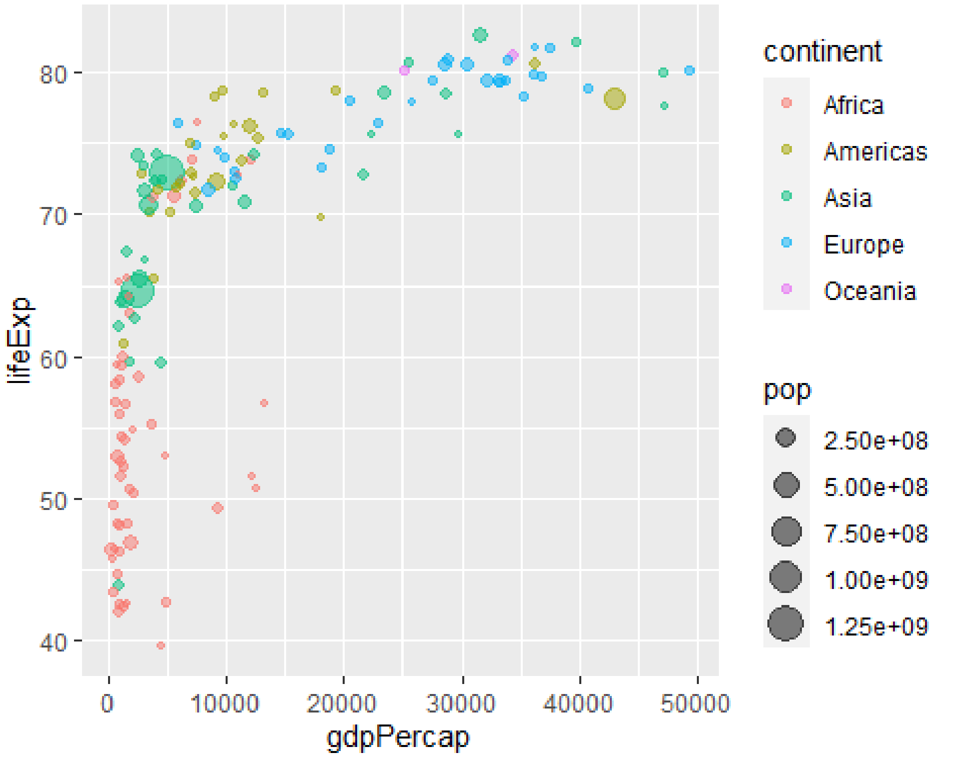


```
#Change colors
ggplot(data = genes) +
  aes(x = Condition1, y = Condition2, col = State) +
  geom_point() +
  scale_color_manual(values = c("blue", "gray", "red")) +
  labs(title = "Gene Expression Changes Upon Drug Treatment",
       x = "Control (no drug)",
       y = "Drug Treatment") +
  theme_bw()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

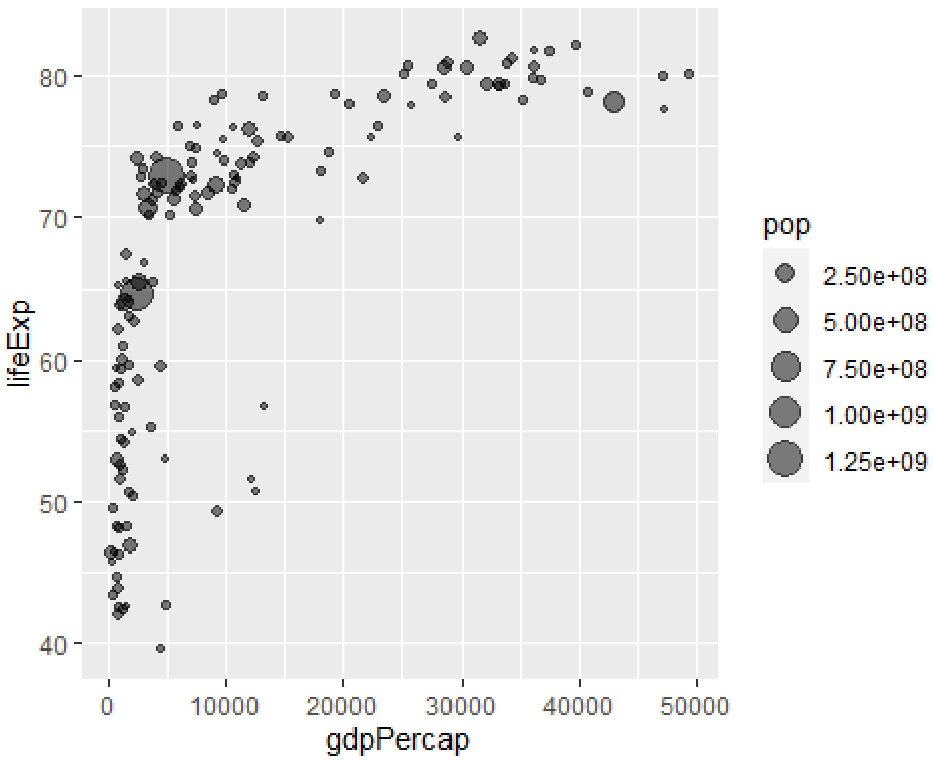


```
##

#Optional Part 6

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

# File location online
url2 <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extd
ata/gapminder.tsv"
gapminder <- read.delim(url2)

#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
```

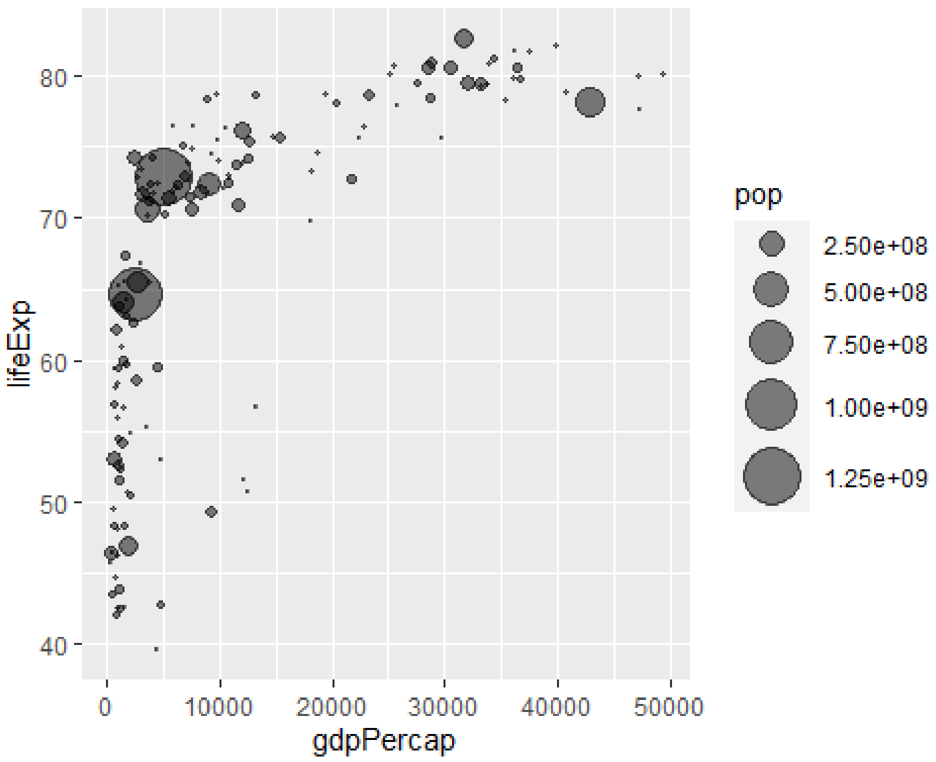
```
gapminder_2007 <- gapminder %>%
  filter(year == 2007)

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



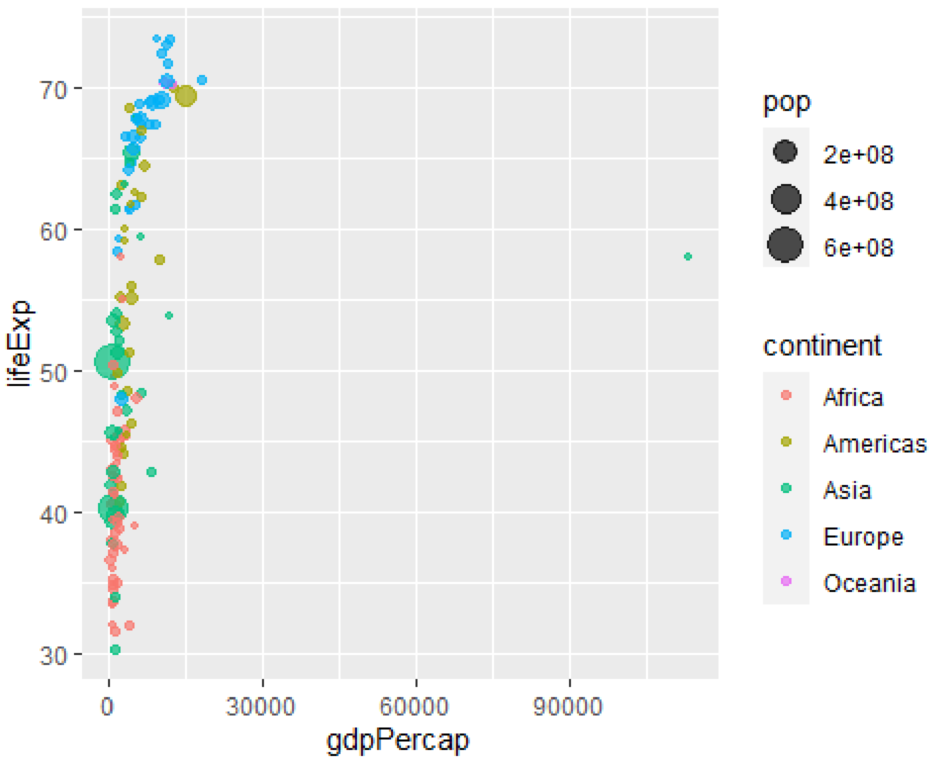
```
# Color by pop
ggplot(gapminder_2007) +
  aes(x = gdpPercap, y = lifeExp,
      size = pop) +
  geom_point(alpha = 0.5)
```



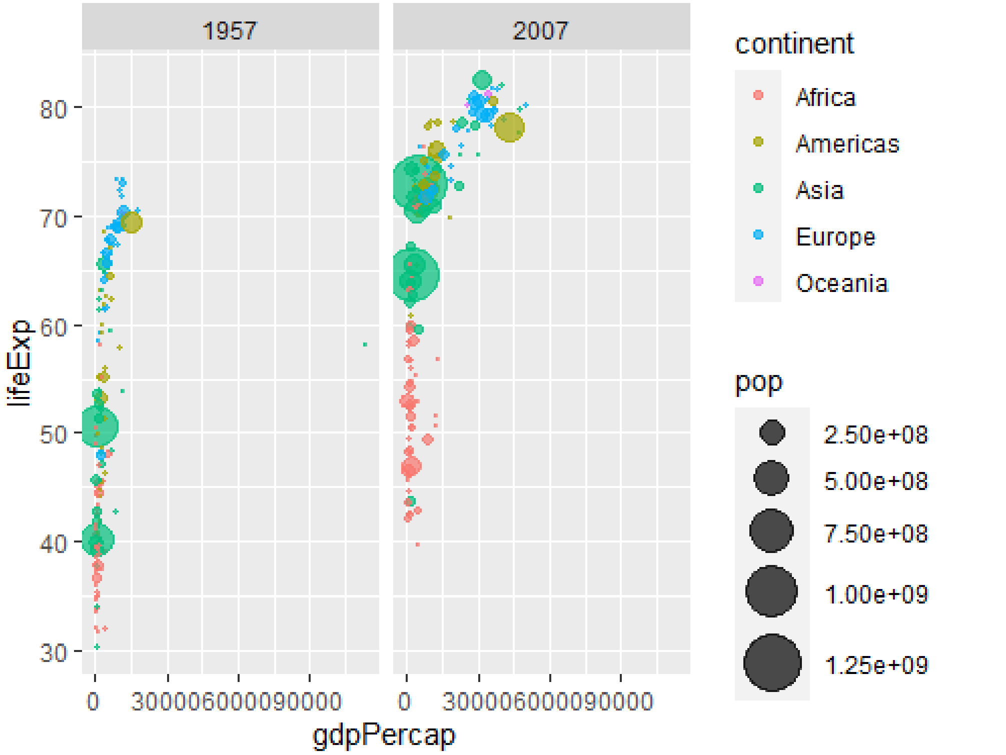
```
#Scale to reflect actual population differences
ggplot(gapminder_2007) +
  geom_point(aes(x = gdpPercap, y = lifeExp,
                 size = pop),
             alpha = 0.5) +
  scale_size_area(max_size = 10)
```



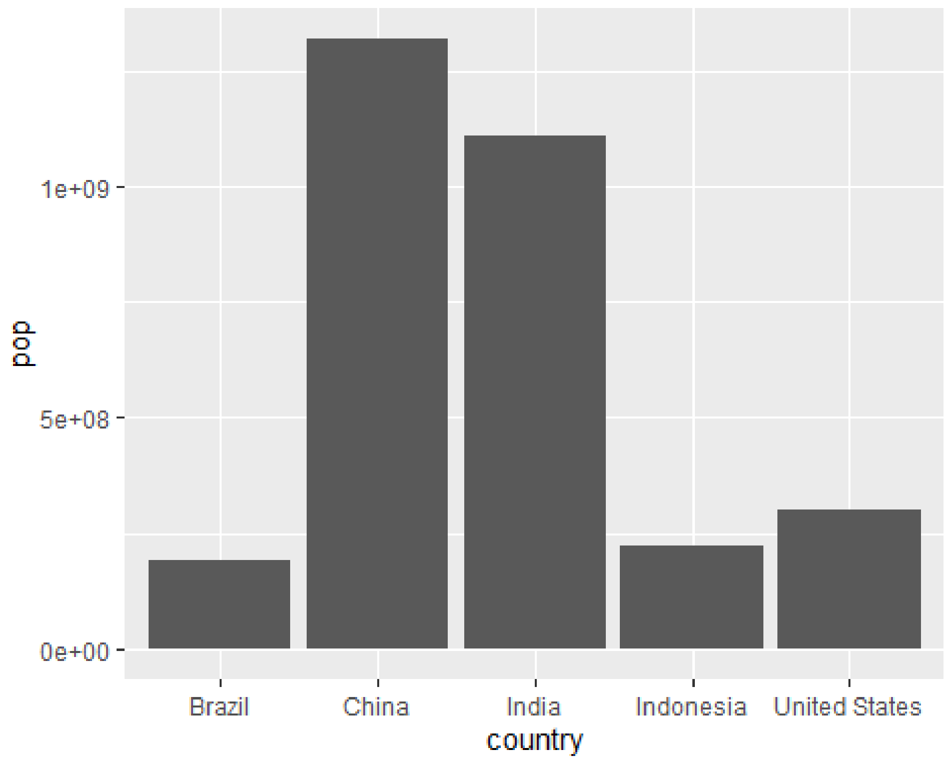
```
#1957 Plot
gapminder_1957 <- gapminder %>%
  filter(year == 1957)

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



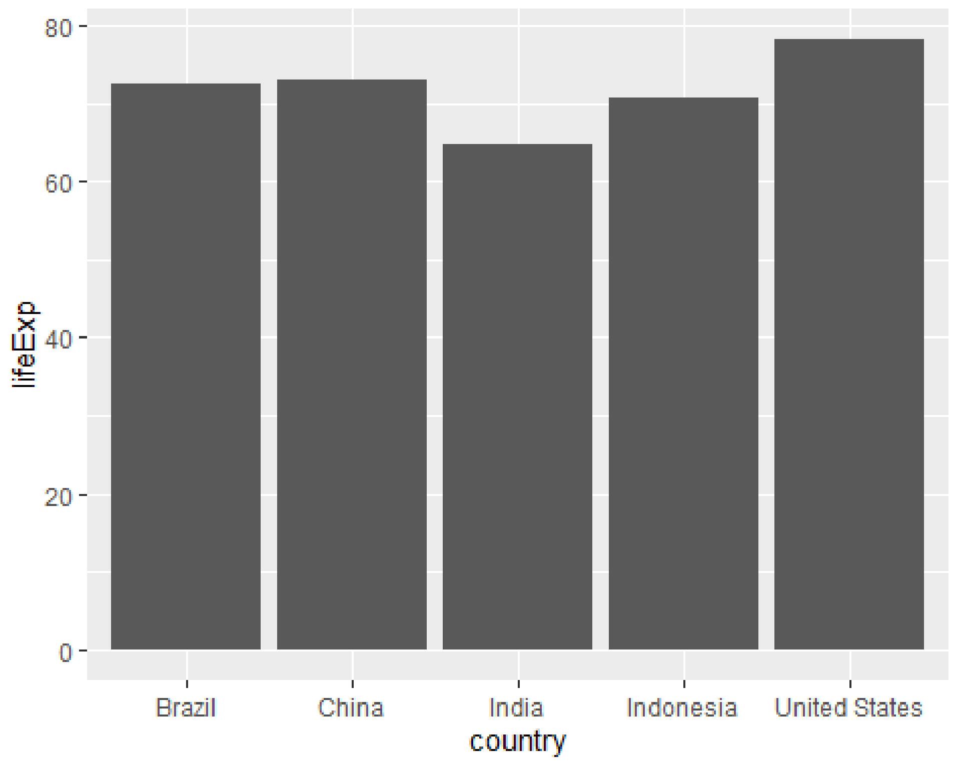
```
scale_size_area(max_size = 10)

## <ScaleContinuous>
##  Range:
##  Limits:    0 --    1

#Combine 1957 and 2007
gapminder_combined <- gapminder %>%
  filter(year == 1957 | year == 2007)

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



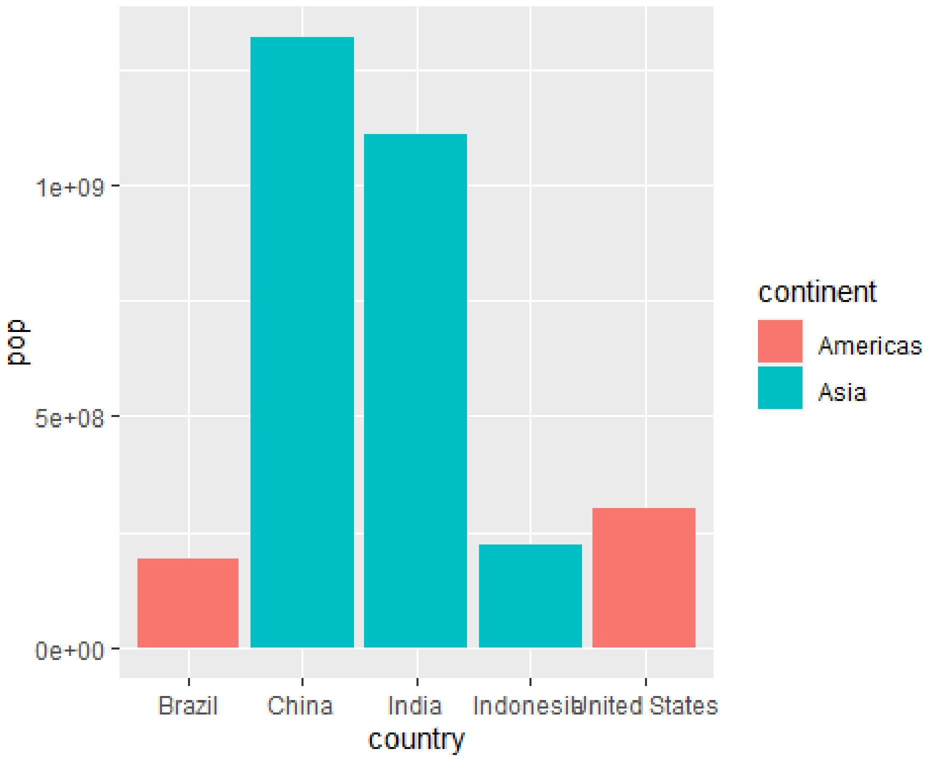
```
##

#Optional Part 7

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

#Creating a bar chart
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop))
```



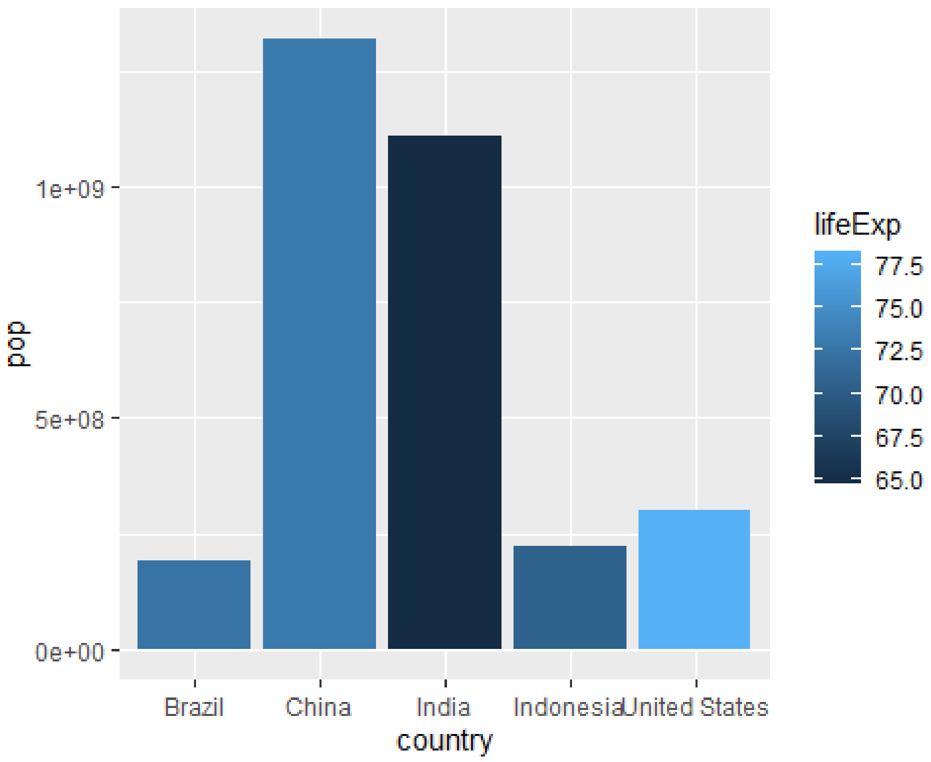
```
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = lifeExp))
```



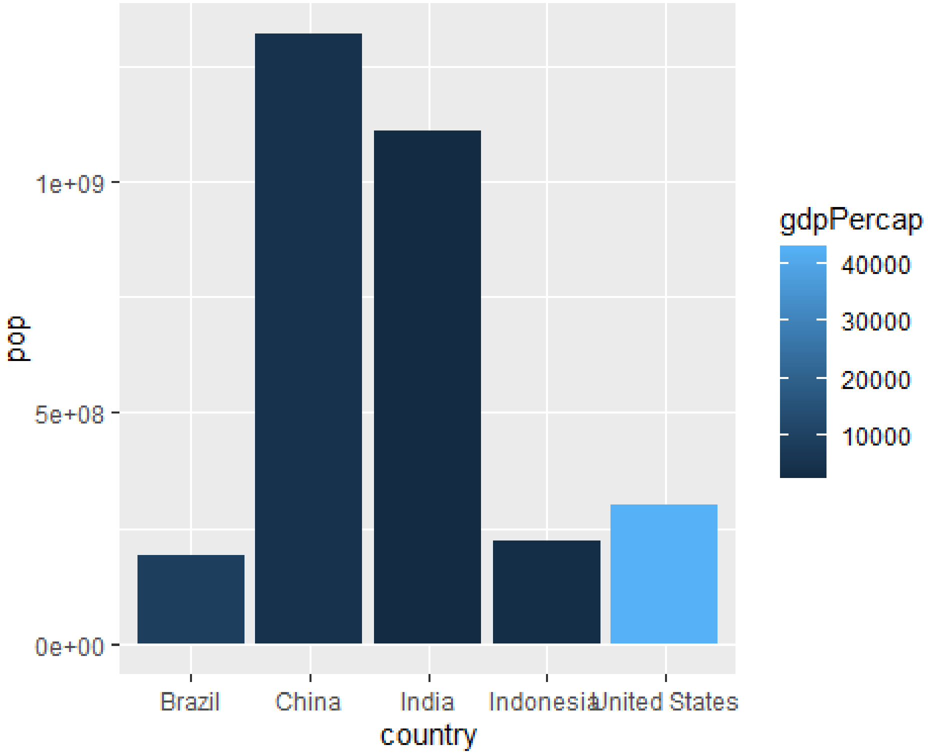
```r
#Filling bars with color
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = continent))
```



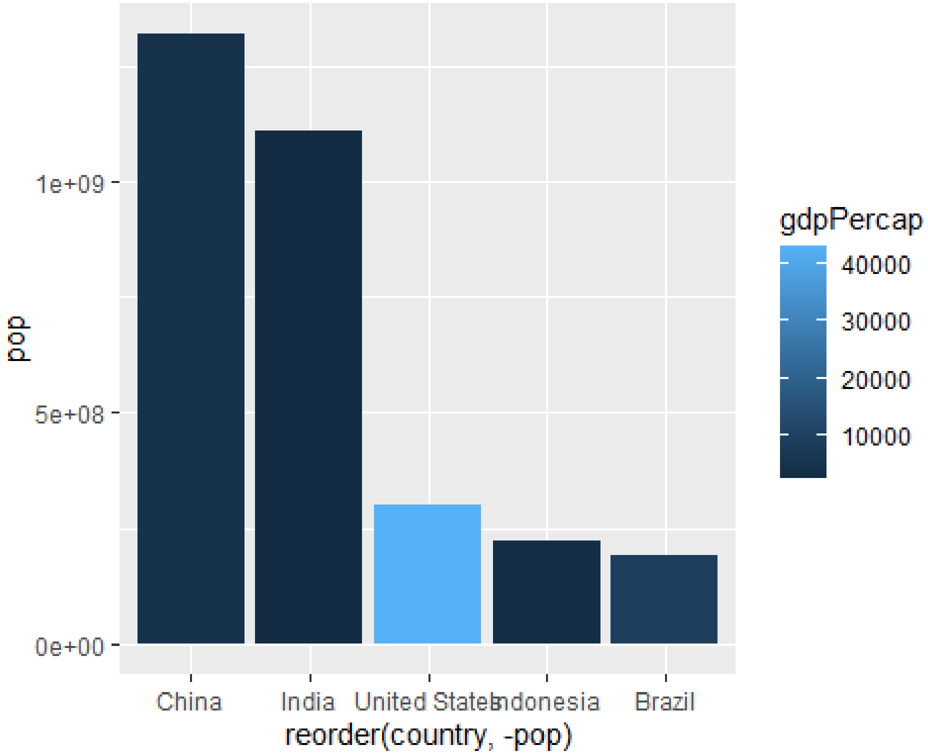
```r
ggplot(gapminder_top5) +
  geom_col(aes(x = country, y = pop, fill = lifeExp))
```



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

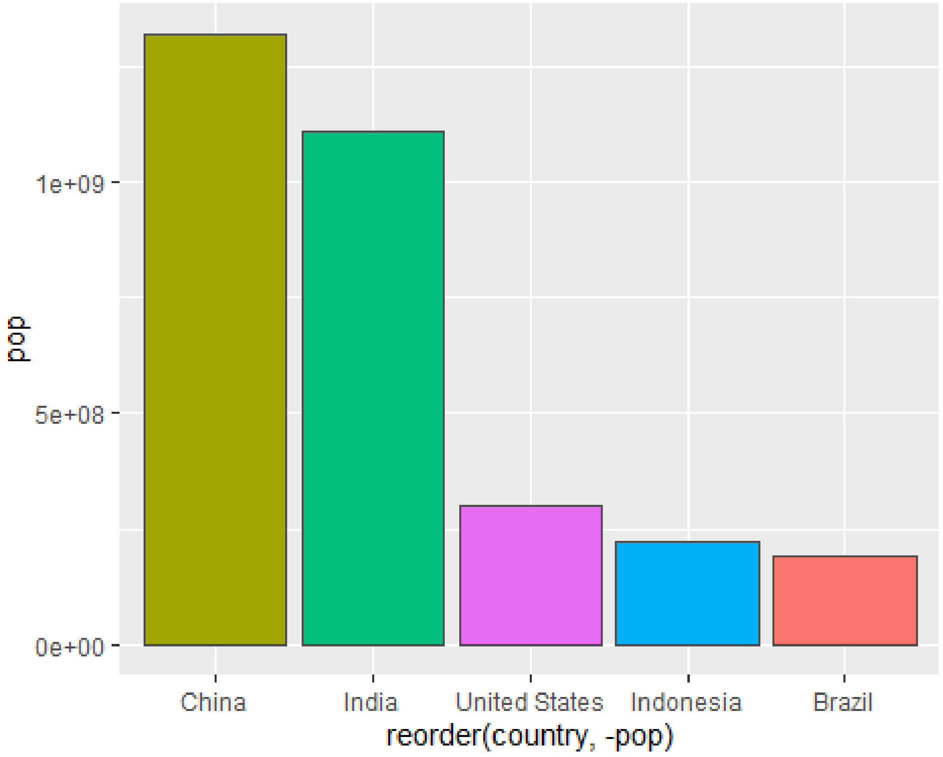


```
#Change order of bars
ggplot(gapminder_top5) +
  aes(x = reorder(country, -pop), y=pop,
      fill = gdpPercap) +
  geom_col()
```

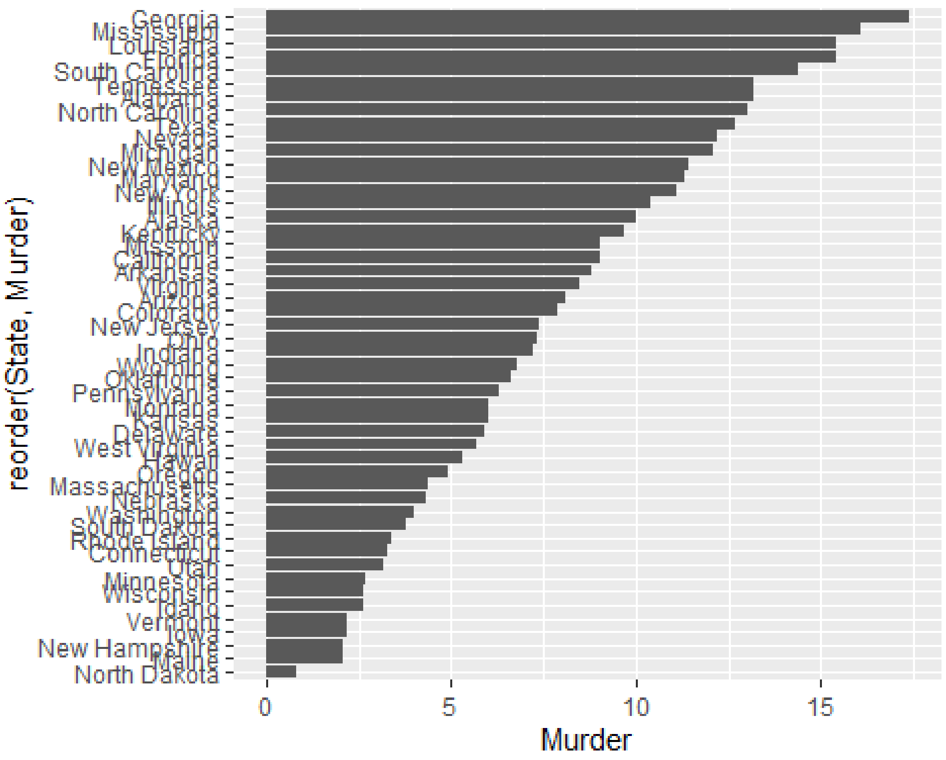


```
ggplot(gapminder_top5) +
  aes(x = reorder(country, -pop), y = pop,
      fill = country) +
  geom_col(col = "gray30") +
  guides(fill = FALSE)
```

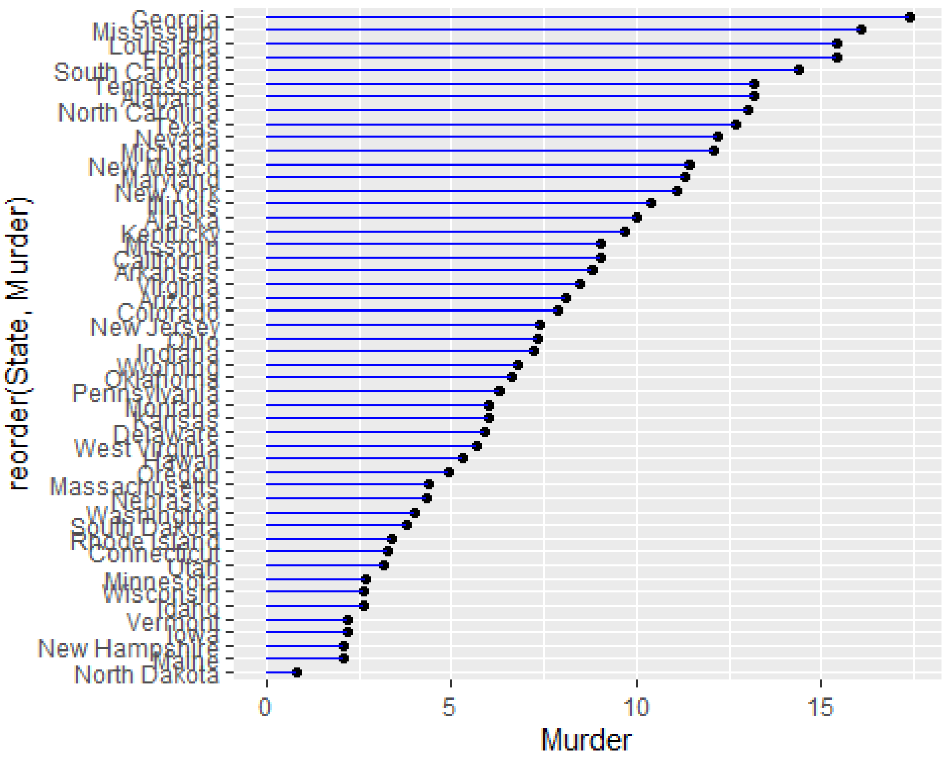
```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scal
e> =
## "none")` instead.
```



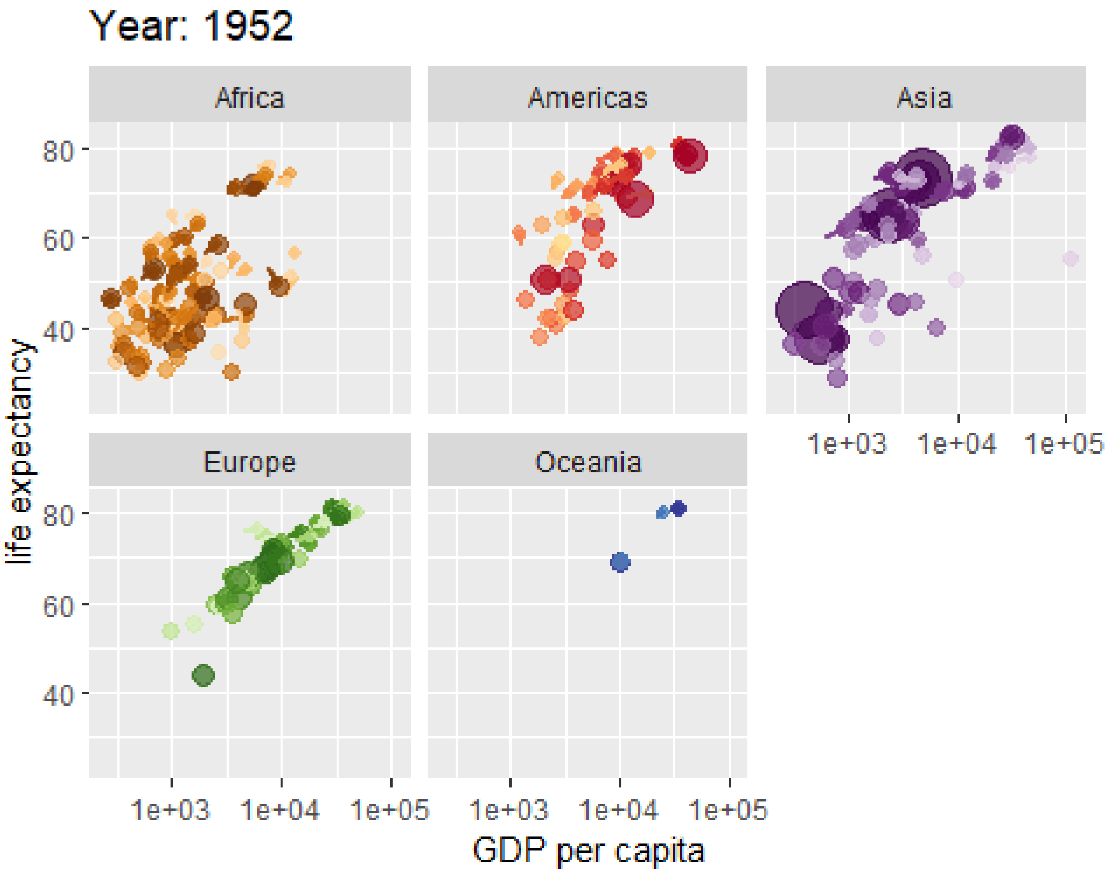
```
#Flipping bar charts
head(USArrests)

##            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()
```



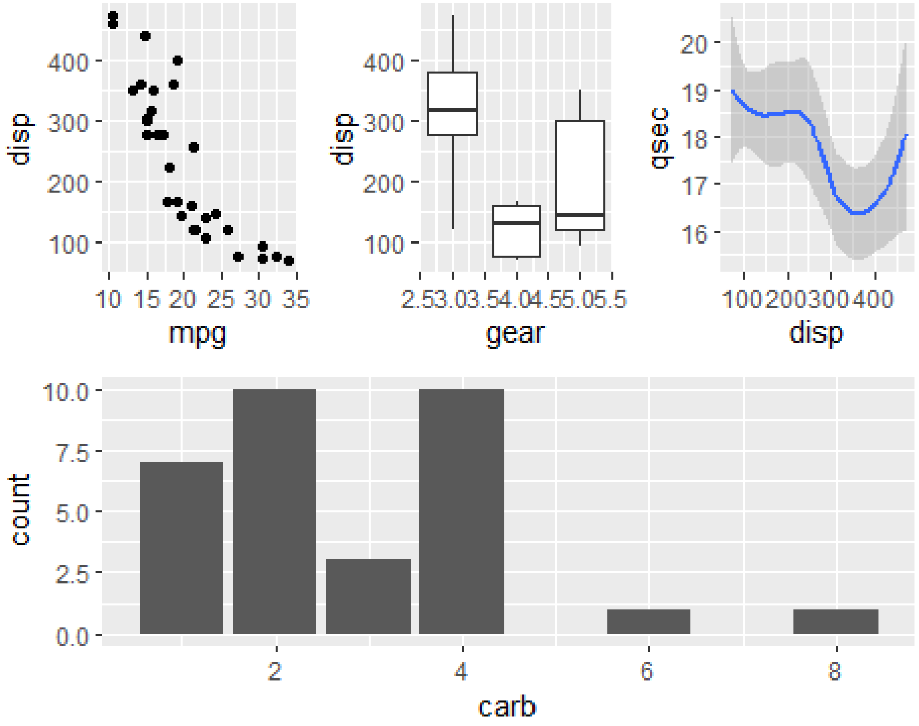
```
ggplot(USArrests) +
  aes(x = reorder(State, Murder), y = Murder) +
  geom_point() +
  geom_segment(aes(x=State,
                   xend=State,
                   y=0,
                   yend=Murder),
               color = "blue") +
  coord_flip()
```



```
##

#Optional Part 8

#install.packages("gifski")
#install.packages("gganimate")

library(gapminder)
library(gganimate)

# Setup nice regular ggplot of the gapminder data
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 by continent
  facet_wrap(~continent) +
  # Here comes the gganimate specific bits
  labs(title = 'Year: {frame_time}', x = 'GDP per capita', y = 'life expectan
cy') +
  transition_time(year) +
  shadow_wake(wake_length = 0.1, alpha = FALSE)
```



```
##

#Optional Part 9

#Combining Plots

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

# Setup 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(disp, qsec))
p4 <- ggplot(mtcars) + geom_bar(aes(carb))

# Use patchwork to combine them here:
(p1 | p2 | p3) /
  p4

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



```
sessionInfo()

## R version 4.1.2 (2021-11-01)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19043)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] patchwork_1.1.1 gganimate_1.0.7 dplyr_1.0.7     gapminder_0.3.0
## [5] ggplot2_3.3.5
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.8        plyr_1.8.6        pillar_1.6.4      compiler_4.1.2
##  [5] highr_0.9         prettyunits_1.1.1 progress_1.2.2    tools_4.1.2
##  [9] digest_0.6.27     lattice_0.20-45   nlme_3.1-155      evaluate_0.14
## [13] lifecycle_1.0.1   tibble_3.1.6      gtable_0.3.0      mgcv_1.8-38
## [17] pkgconfig_2.0.3   rlang_0.4.11      Matrix_1.4-0      DBI_1.1.2
## [21] yaml_2.2.1        xfun_0.29         withr_2.4.3       stringr_1.4.0
## [25] knitr_1.37        hms_1.1.1         generics_0.1.1    vctrs_0.3.8
## [29] grid_4.1.2        tidyselect_1.1.1  glue_1.6.0        R6_2.5.1
## [33] gifski_1.4.3-1    fansi_0.5.0       rmarkdown_2.11    tweenr_1.0.2
## [37] purrr_0.3.4       farver_2.1.0      magrittr_2.0.1    splines_4.1.2
## [41] scales_1.1.1      ellipsis_0.3.2    htmltools_0.5.1.1 assertthat_0.2.
1
## [45] colorspace_2.0-2  labeling_0.4.2    utf8_1.2.2        stringi_1.7.6
## [49] munsell_0.5.0     crayon_1.4.2
```