Class 05 Data Visualization

Nadia Haghani

2021-10-12

```r
# Lets start with a scatterplot

#install.packages("ggplot2")
library(ggplot2) #load before using

# Every ggplot has a data + aes + geoms
p <- ggplot(data = cars) +
  aes(x=speed, y=dist) +
  geom_point() +
  geom_smooth(method="lm")

p + labs(title = "My nice plot",
        x= "Speed (MPH)",
        y = "Distance (ft)") +
  theme_bw()
```
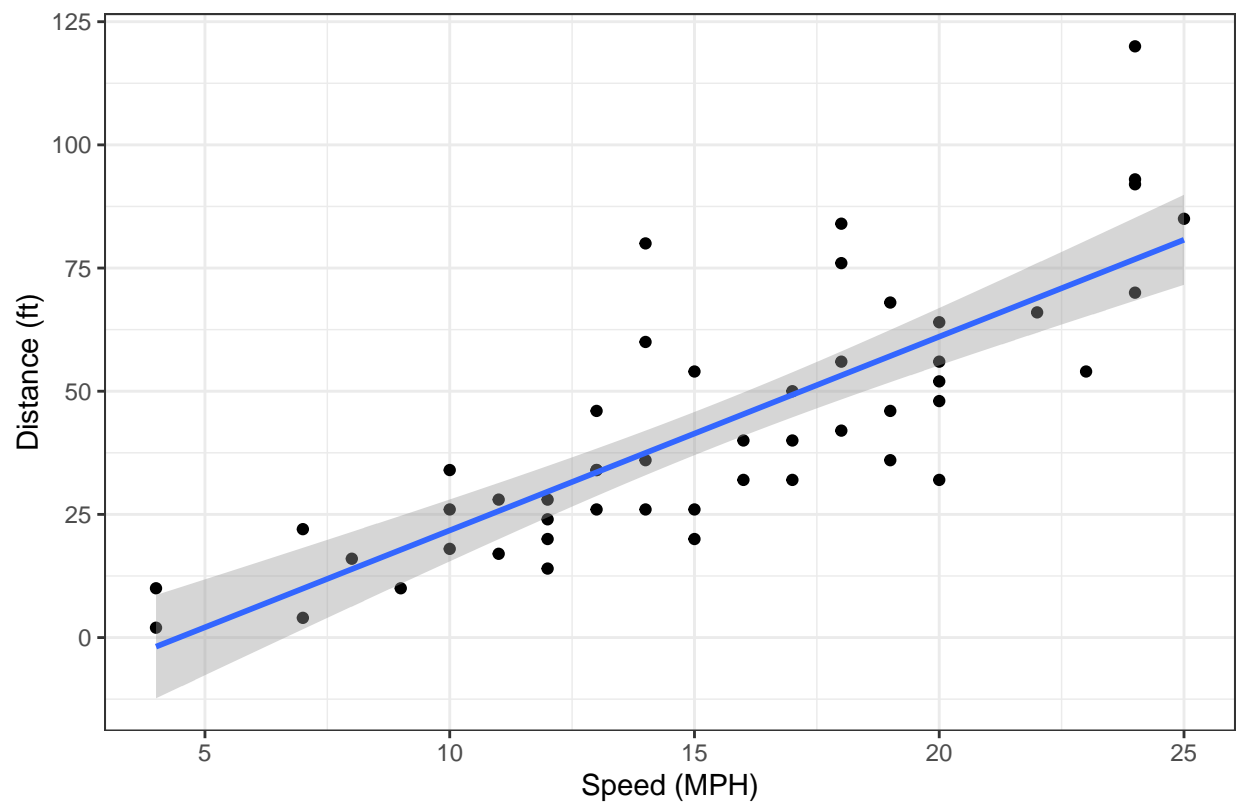
```
## `geom_smooth()` using formula 'y ~ x'
```

```r
# A More complicated dataset

# First read the dataset
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
```

```r
# How many genes in dataset?
nrow(genes)
```

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

```r
# How to access State col
table(genes$State)
```

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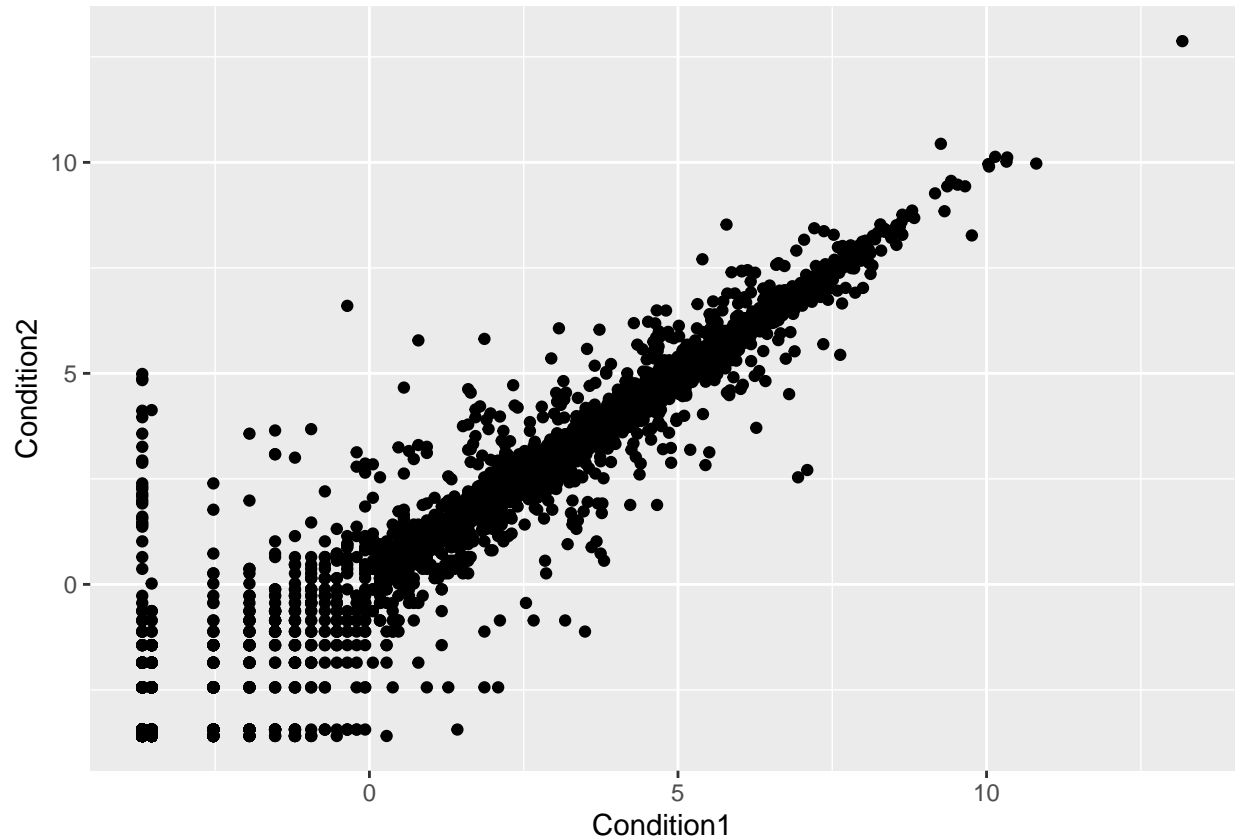
```
##      down unchanging         up
##        72       4997        127
```

```
# What % are up/down
prec <- table(genes$State) / nrow(genes) * 100
round(prec, 2)
```

```
##
##      down unchanging         up
##      1.39      96.17       2.44
```
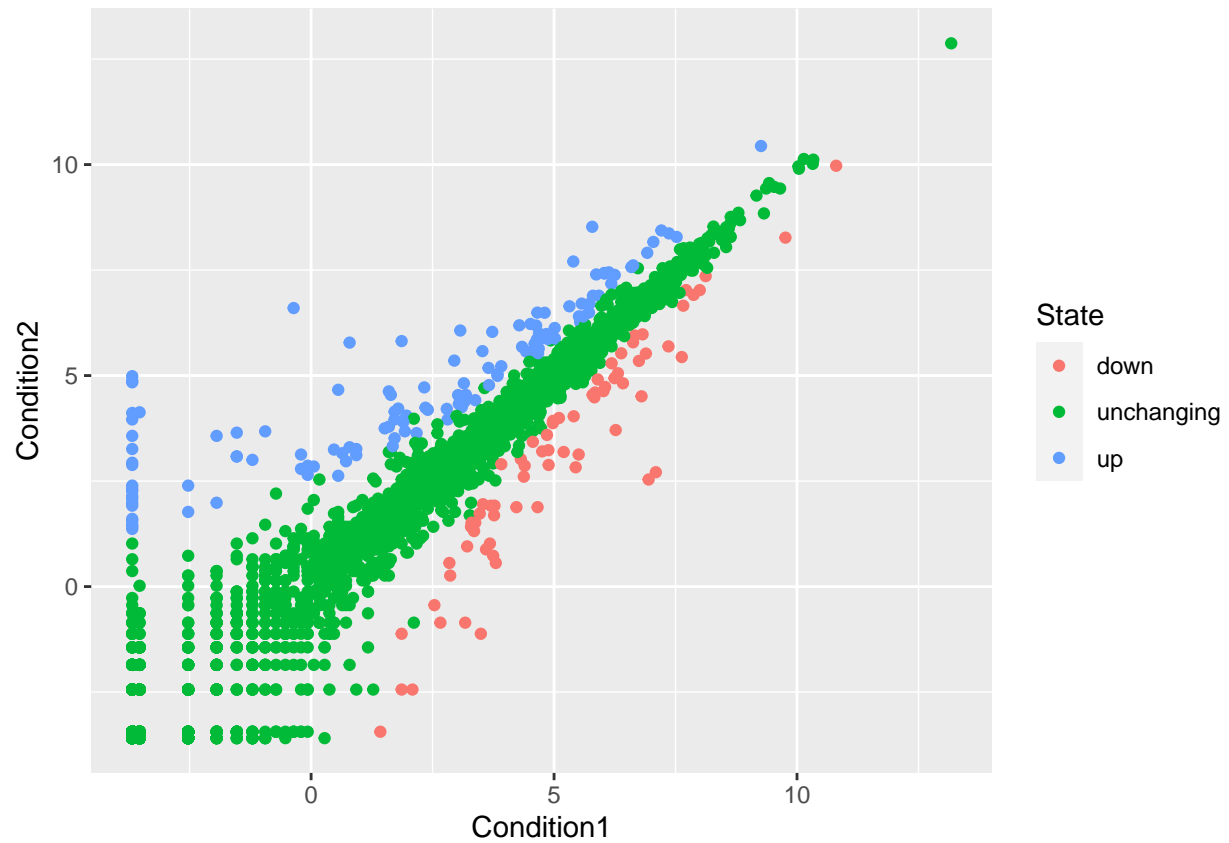
```
# Make a basic scatterplot of genes dataset

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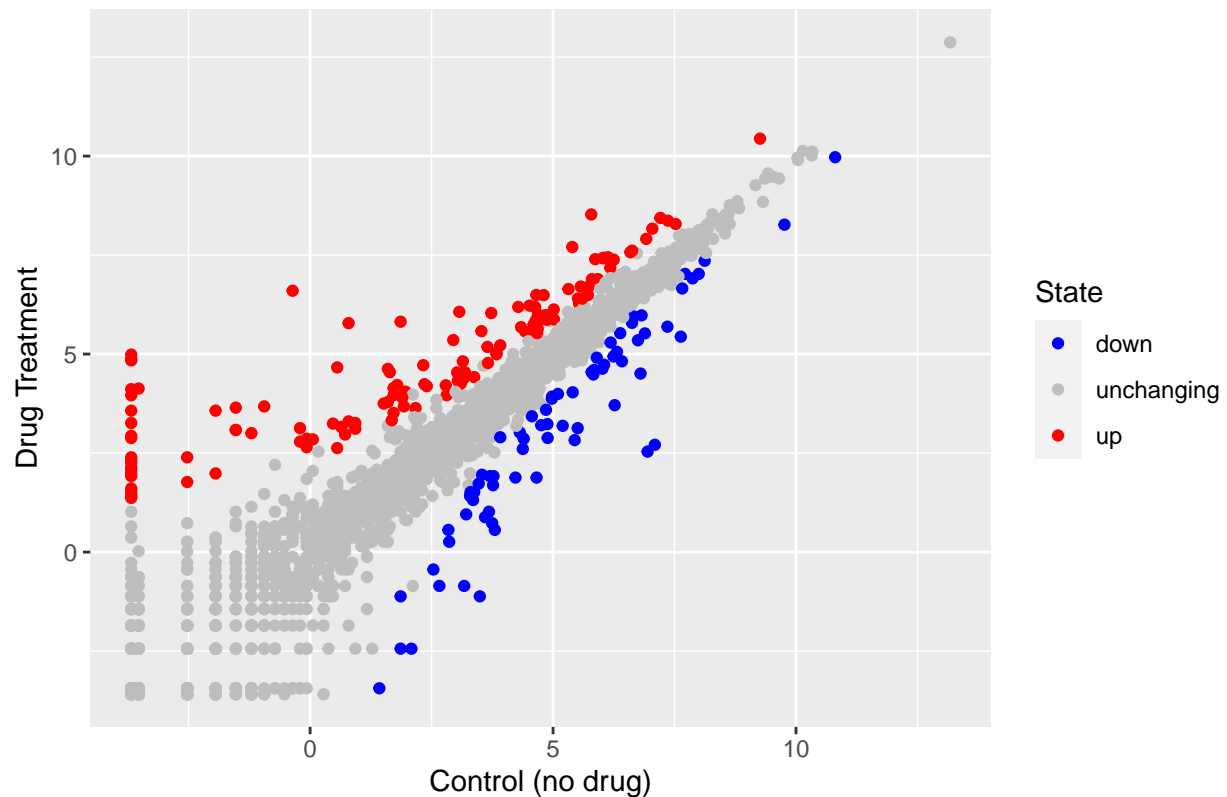


```
# Adding color to the state of the gene

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

3

```
# Change color

p + scale_colour_manual(values= c("blue", "grey", "red")) +
  labs(title = "Gene Expression Changes Upon Drug Treatment",
      x = "Control (no drug)",
      y = "Drug Treatment")
```

Gene Expression Changes Upon Drug Treatment


```
# OPTIONAL: going forward with gapmider dataset

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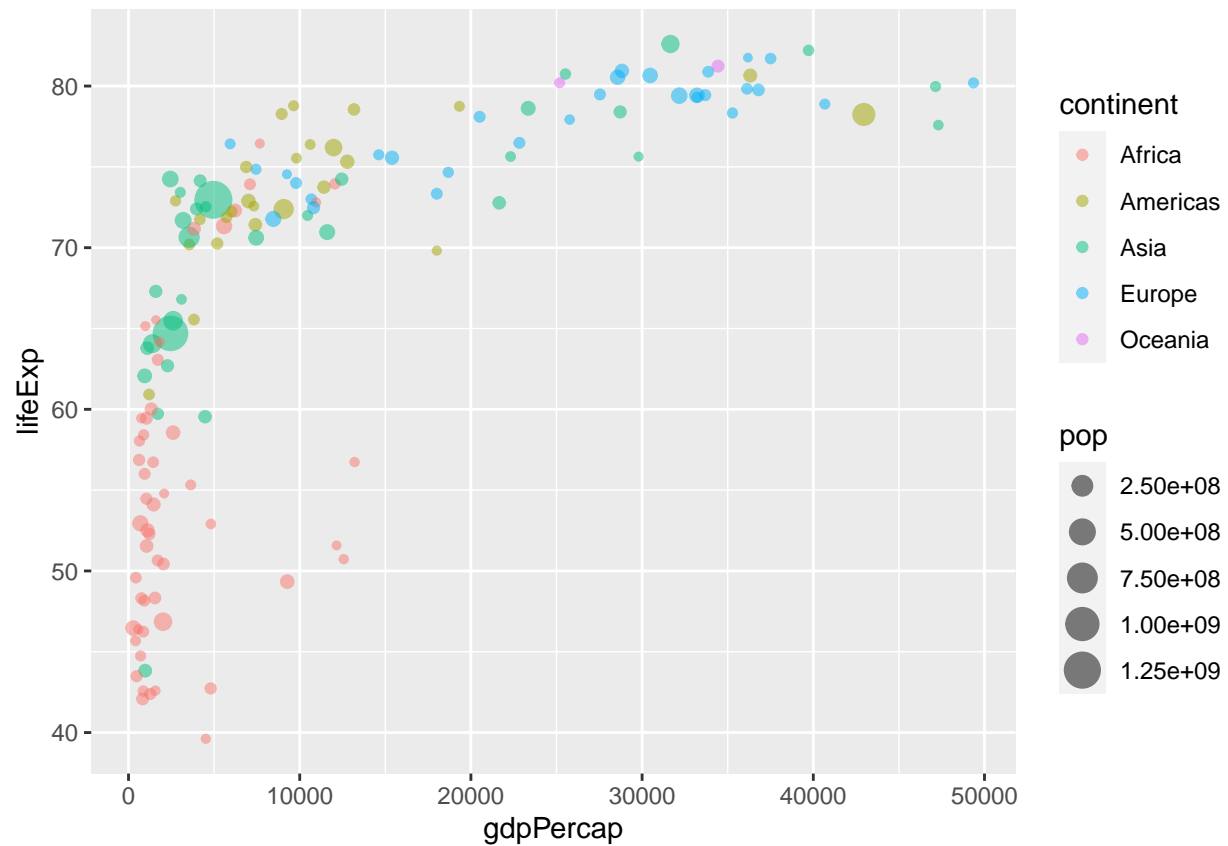
```
gapminder_2007 <- gapminder %>% filter(year==2007)

# Basic scatterplot of gapminder_2007 dataset
ggplot(gapminder_2007) +
```

```
aes(x=gdpPercap, y=lifeExp, color=continent, size=pop) +
geom_point(alpha=0.5) #alpha argument makes points slightly transparent
```
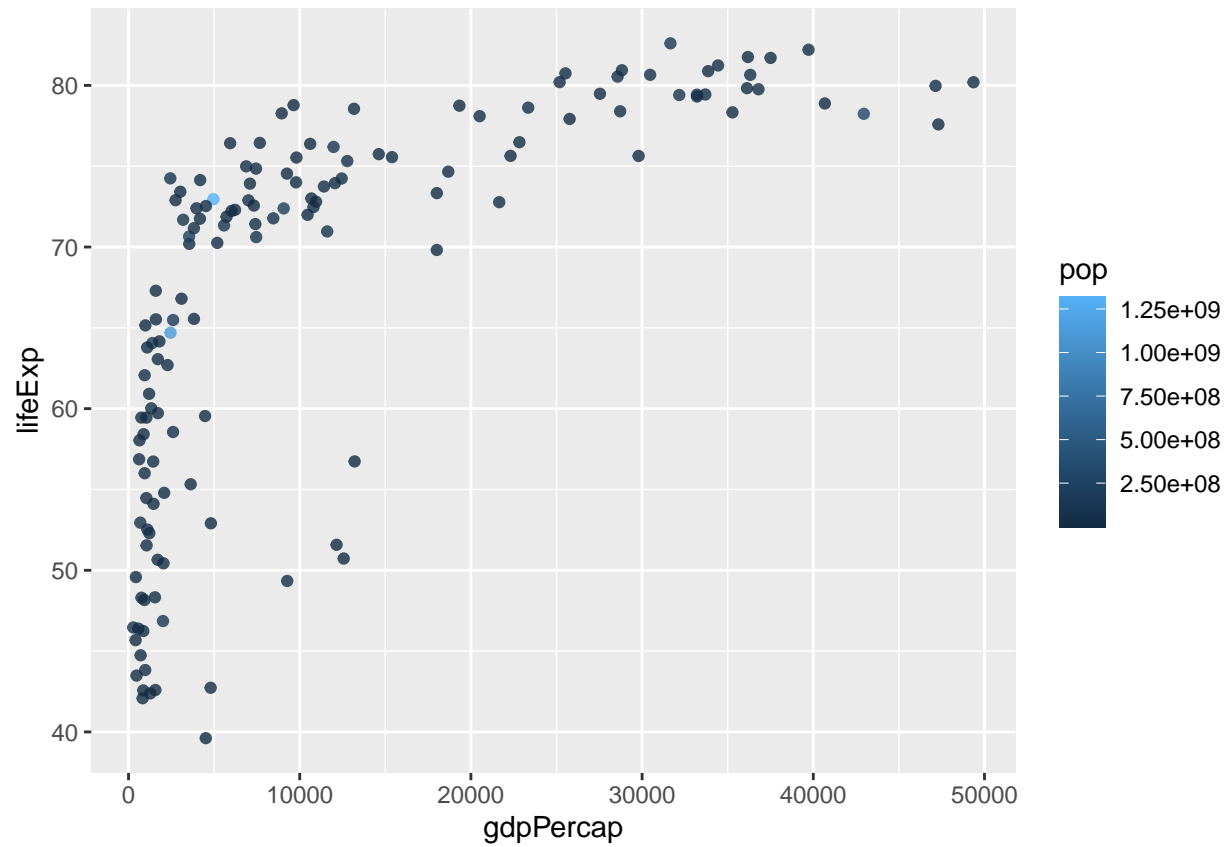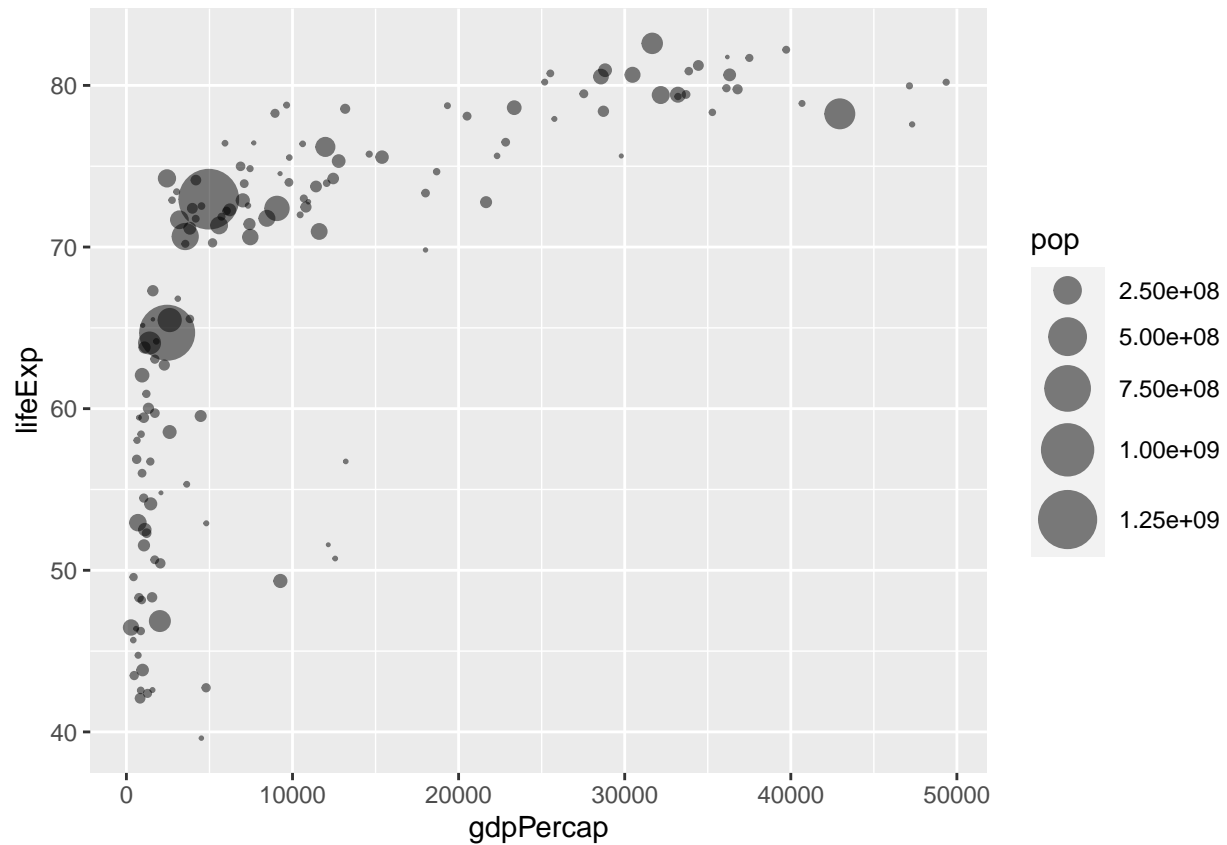


```
ggplot(gapminder_2007) +
  aes(x=gdpPercap, y=lifeExp, color=pop) + # color=pop changes to continuous scale
  geom_point(alpha=0.8)
```

```
# Adjust point size

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

```r
# Bar Plots attempted

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
```
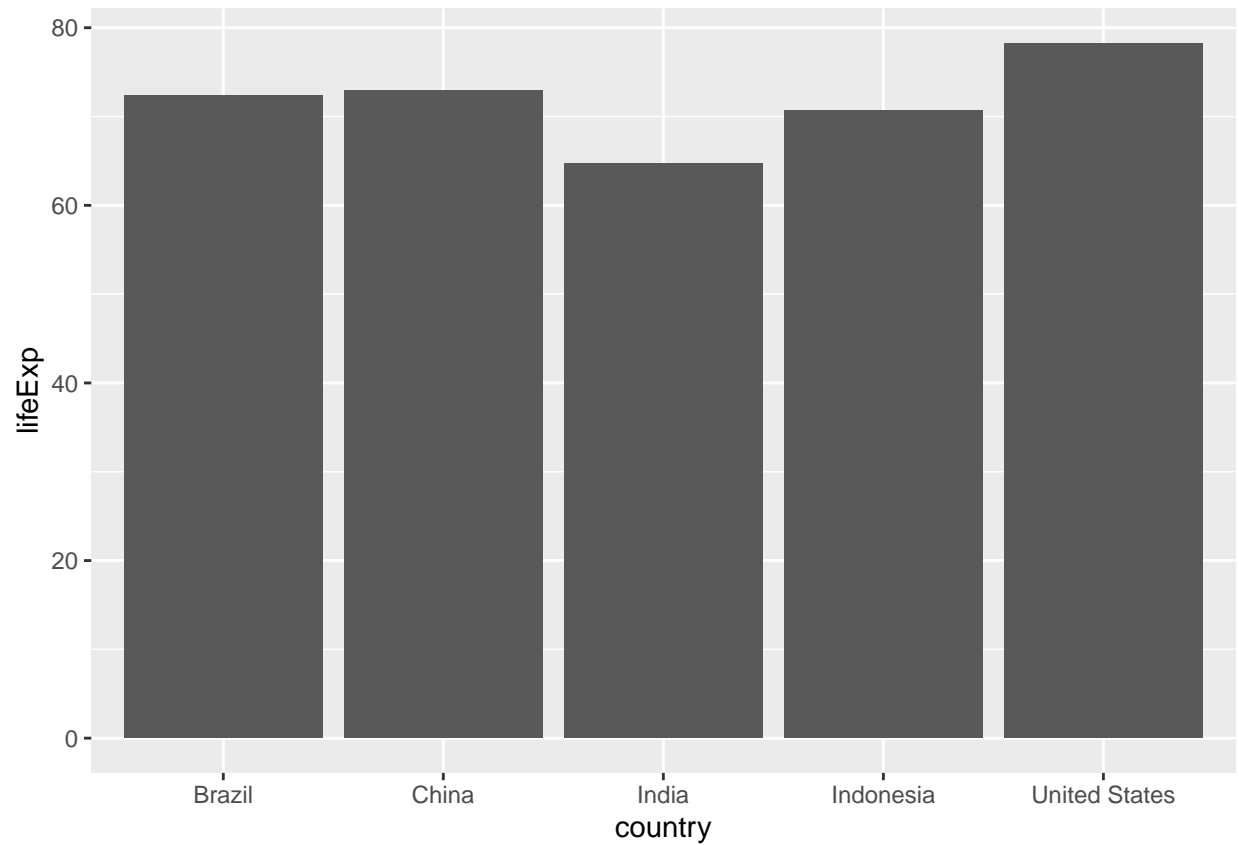
```r
ggplot(gapminder_top5) +
  geom_col(aes(x=country, y=pop))
```
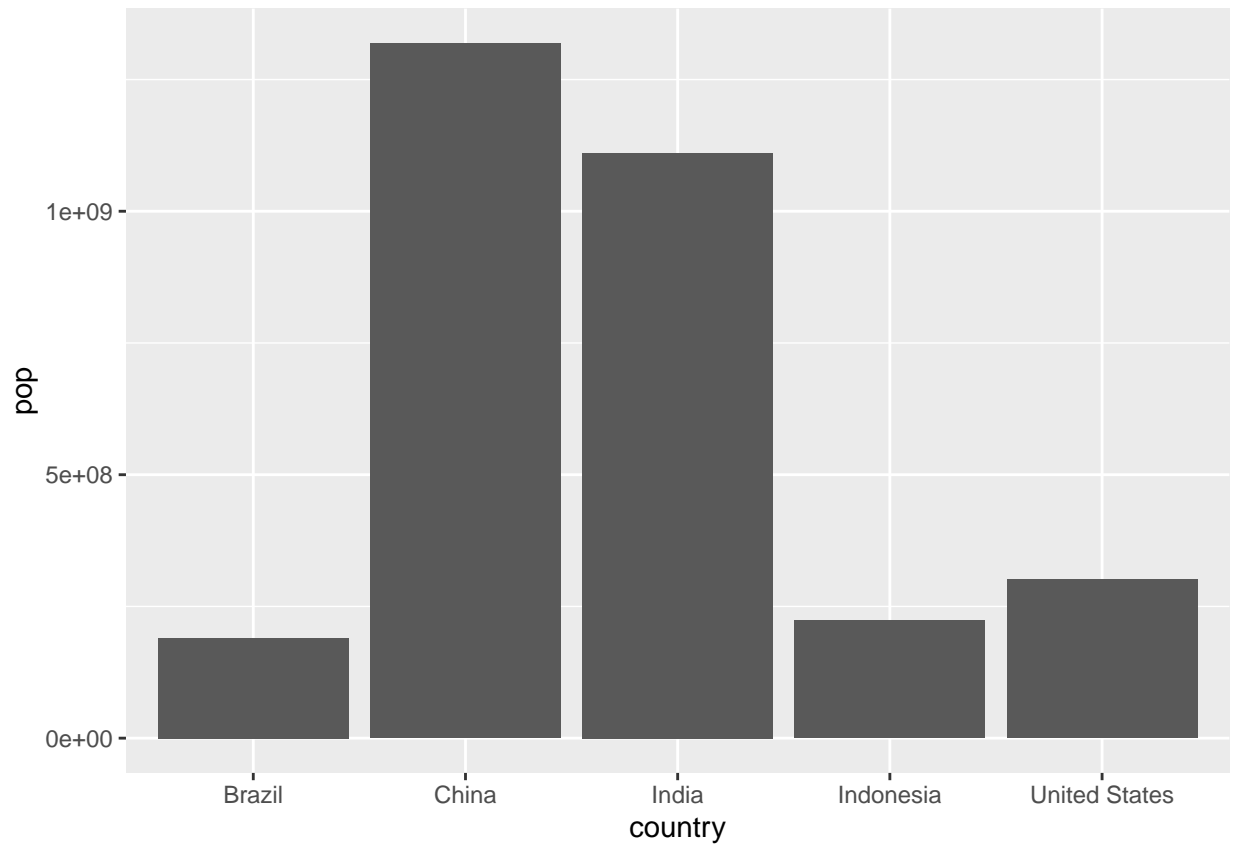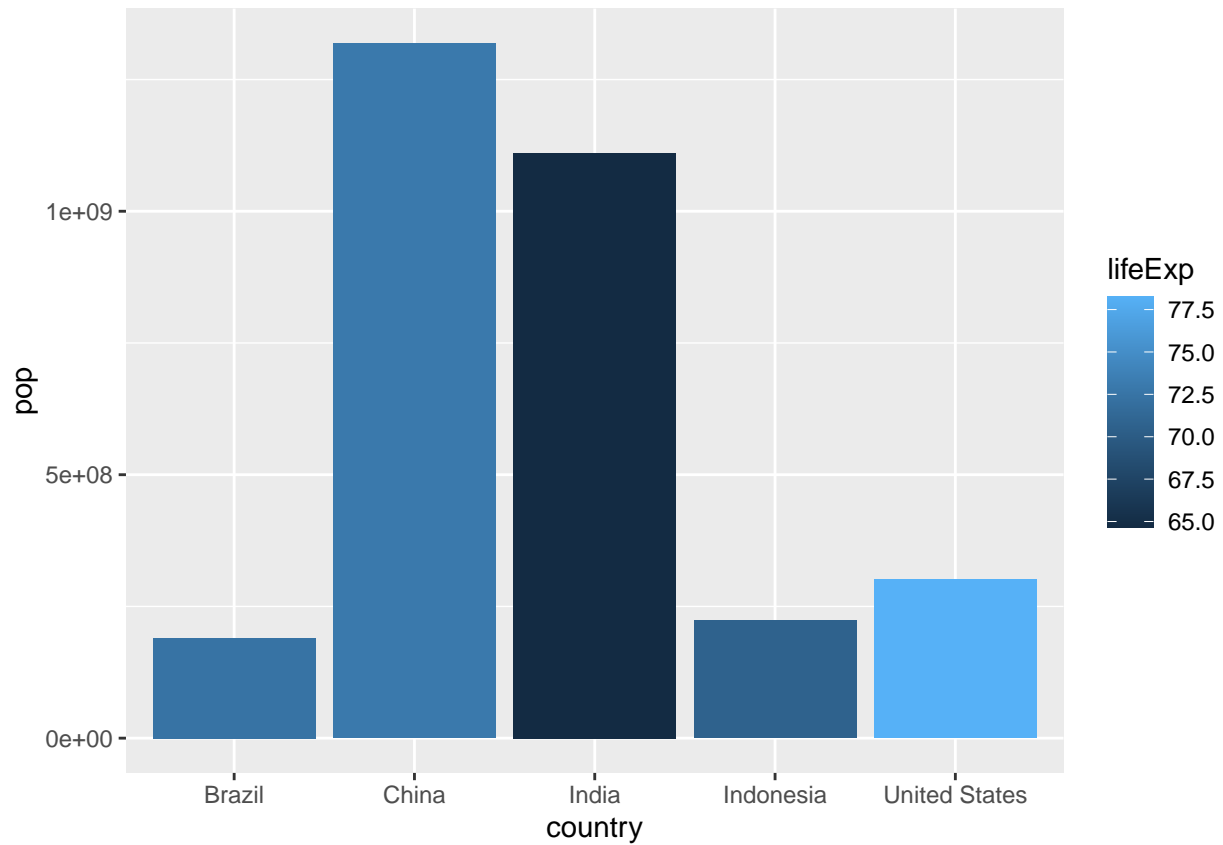
```r
# Q. Plot life expectancy by country

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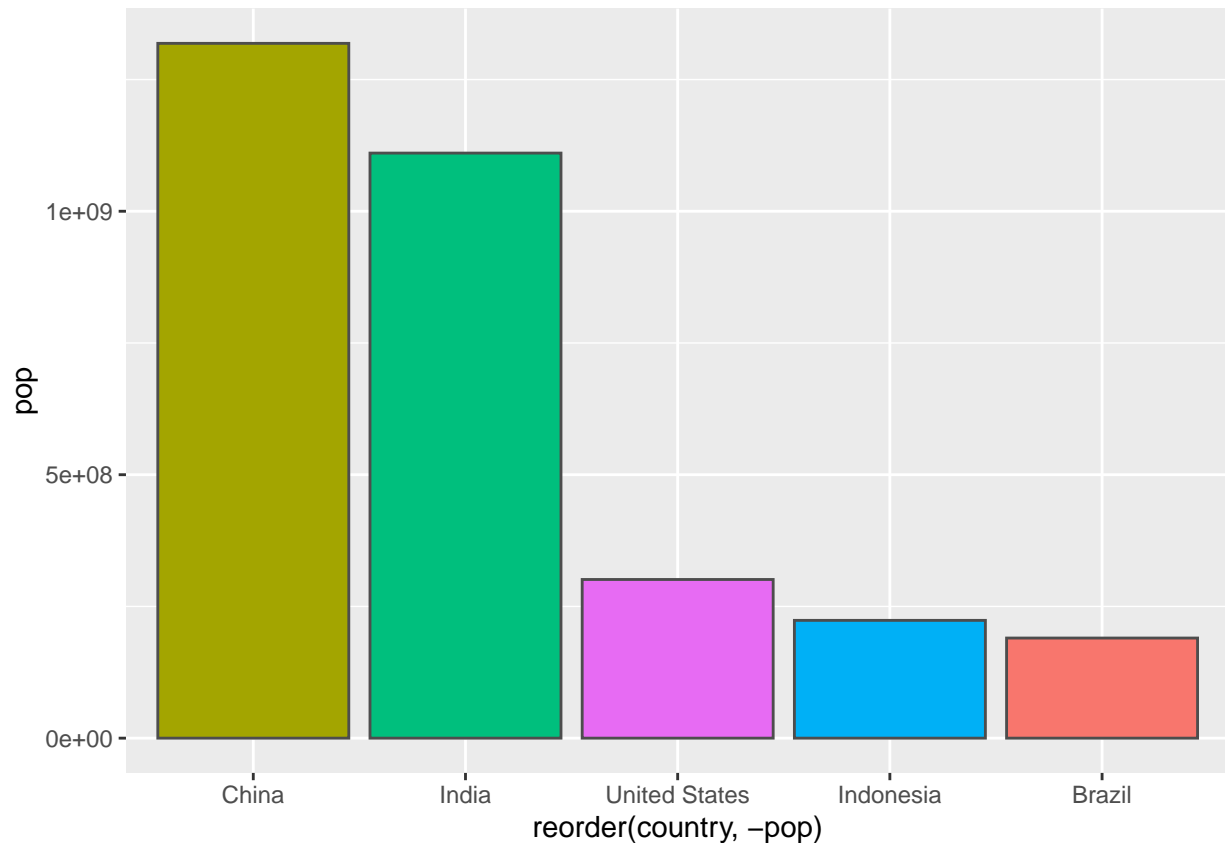
```r
# Filling bars with color

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

```
# Q. Plot population size 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.
```

```
# Advanced plot animations

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

library(gapminder)
```

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

```
## The following object is masked _by_ '.GlobalEnv':
##
##      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 expectancy') +
#   transition_time(year) +
#   shadow_wake(wake_length = 0.1, alpha = FALSE)
```