Class5.R

l.cruz

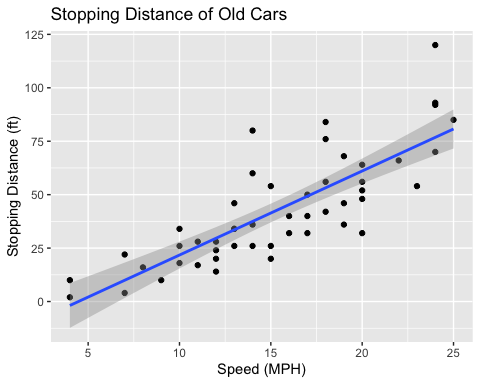
2021-10-13

# Class5: Data visualization   
  
# Today we are going to use ggplot2 packages to help us visualize data  
  
#First we need to load the packages!   
library(ggplot2)  
  
  
# We will use this inbiult "cars" dataset first  
head(cars)

## speed dist  
## 1 4 2  
## 2 4 10  
## 3 7 4  
## 4 7 22  
## 5 8 16  
## 6 9 10

# All ggplots have at least 3 layers,   
# data + aes + geoms  
ggplot(cars) +   
 aes(x=speed, y=dist) +  
 #geom\_points develops scatter plot   
 #do different plots with geom\_(type of graph you want)  
 #lm= linear model (linearizes data sets)  
 #smooth adds line of best regression   
 #method is used to argue what methods we want to see to visualize data   
 geom\_point() +   
 geom\_smooth(method="lm") +  
 #labs= labels that we will be adding to the data set  
 labs(title="Stopping Distance of Old Cars",   
 x="Speed (MPH)", y="Stopping Distance (ft)")

## `geom\_smooth()` using formula 'y ~ x'



#Side-not: ggplot is not the only graphics system   
#a very popular one is good old "base" R graphcs   
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"

ncol(genes)

## [1] 4

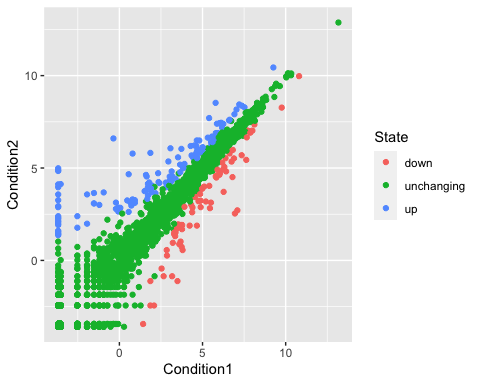
table(genes$State)

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

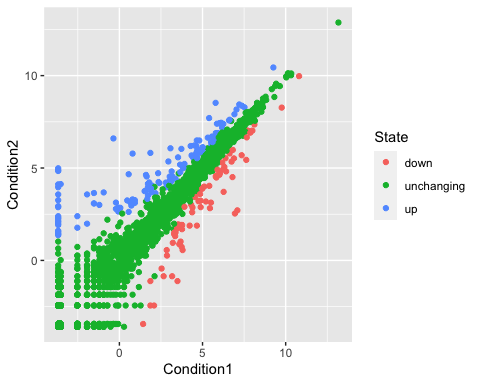
round(table(genes$State)/nrow(genes) \* 100, 2)

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

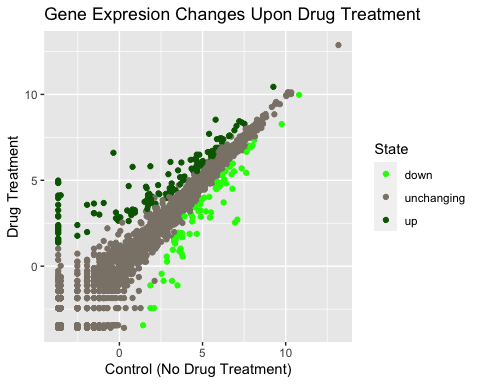
ggplot(genes) +   
 aes(x=Condition1, y=Condition2, col=State) +  
 geom\_point()



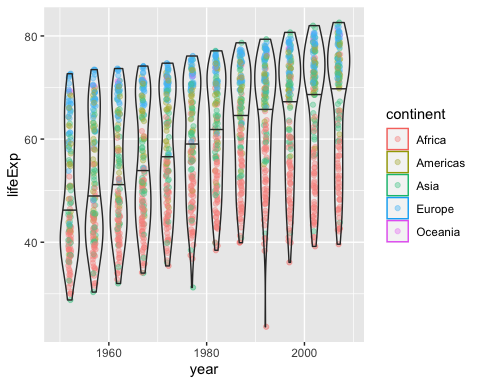
p <- ggplot(genes) +   
 aes(x=Condition1, y=Condition2, col=State) +  
 geom\_point()  
p



p+scale\_color\_manual(values = c("green", "antiquewhite4", "dark green")) +   
 labs(x="Control (No Drug Treatment)", y="Drug Treatment", title="Gene Expresion Changes Upon Drug Treatment")



#install.packages("gapminder")  
library(gapminder)  
  
  
ggplot(gapminder) + aes(x=year, y=lifeExp, col=continent) + geom\_jitter(width=0.3, alpha=0.4) + geom\_violin(aes(group=year), alpha=0.2, draw\_quantiles=0.5)



#install.packages("plotly")  
#library (plotly)  
#ggplotly()  
  
# install.packages("dplyr") ## uncoment to install if needed  
#install.packages("dplyr")  
library(dplyr)

##   
## Attaching package: 'dplyr'

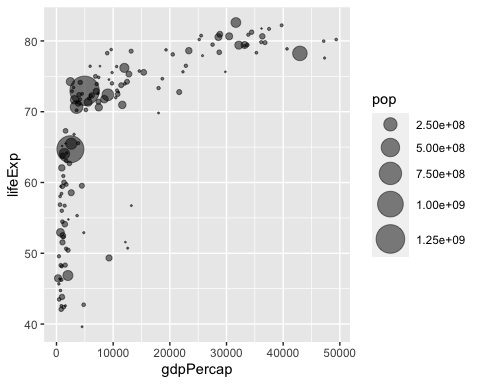
## The following objects are masked from 'package:stats':  
##   
## filter, lag

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

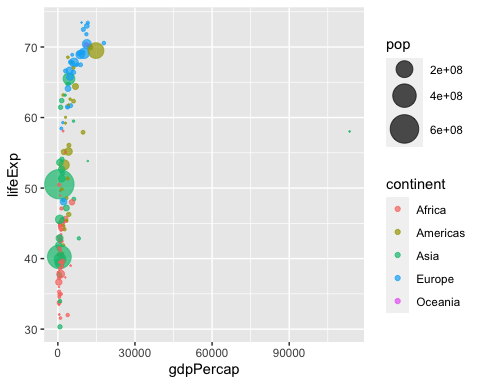
gapminder\_2007 <- gapminder %>% filter(year==2007)  
gapminder\_2007

## # A tibble: 142 × 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 Afghanistan Asia 2007 43.8 31889923 975.  
## 2 Albania Europe 2007 76.4 3600523 5937.  
## 3 Algeria Africa 2007 72.3 33333216 6223.  
## 4 Angola Africa 2007 42.7 12420476 4797.  
## 5 Argentina Americas 2007 75.3 40301927 12779.  
## 6 Australia Oceania 2007 81.2 20434176 34435.  
## 7 Austria Europe 2007 79.8 8199783 36126.  
## 8 Bahrain Asia 2007 75.6 708573 29796.  
## 9 Bangladesh Asia 2007 64.1 150448339 1391.  
## 10 Belgium Europe 2007 79.4 10392226 33693.  
## # … with 132 more rows

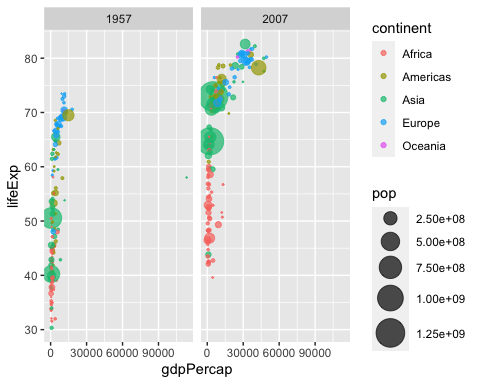
ggplot(gapminder\_2007) +  
 aes(x=gdpPercap, y=lifeExp, size=pop) +  
 #aplha makes the points transparent  
 geom\_point(alpha=0.5) +   
 scale\_size\_area(max\_size = 10)



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)



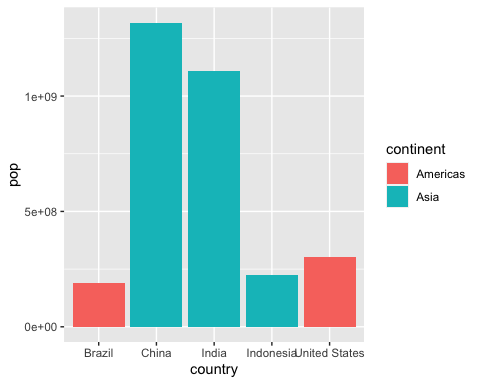
gapminder\_1957 <- gapminder %>% filter(year==1957 | year==2007)  
# the |year ==2007 will add the data from 2007 next to 1957 data sets  
ggplot(gapminder\_1957) +   
 aes(x = gdpPercap, y = lifeExp, color=continent,size = pop) +  
 geom\_point(alpha=0.7) +   
 scale\_size\_area(max\_size = 10) +   
 facet\_wrap(~year)



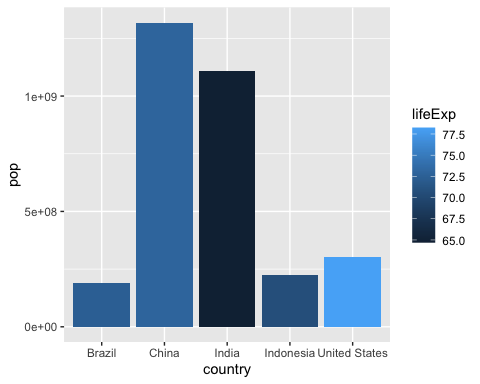
#You should now include the layer facet\_wrap(~year) to produce the following plot:  
  
gapminder\_top5 <- gapminder %>%   
 filter(year==2007) %>%   
 arrange(desc(pop)) %>%   
 top\_n(5, pop)  
  
gapminder\_top5

## # A tibble: 5 × 6  
## country continent year lifeExp pop gdpPercap  
## <fct> <fct> <int> <dbl> <int> <dbl>  
## 1 China Asia 2007 73.0 1318683096 4959.  
## 2 India Asia 2007 64.7 1110396331 2452.  
## 3 United States Americas 2007 78.2 301139947 42952.  
## 4 Indonesia Asia 2007 70.6 223547000 3541.  
## 5 Brazil Americas 2007 72.4 190010647 9066.

ggplot(gapminder\_top5) +   
 geom\_col(aes(x = country, y = pop, fill=continent))

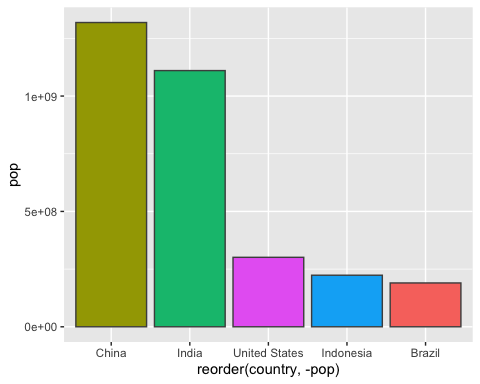


ggplot(gapminder\_top5) +   
 geom\_col(aes(x = country, y = pop, fill=lifeExp))



# Plot population size by country  
ggplot(gapminder\_top5) +  
 aes(x=reorder (country, -pop), y=pop, fill=country) +  
 geom\_col(col="gray30") +  
 guides(fill=FALSE)

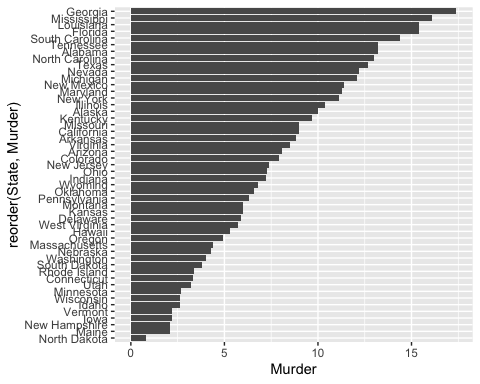
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =  
## "none")` instead.



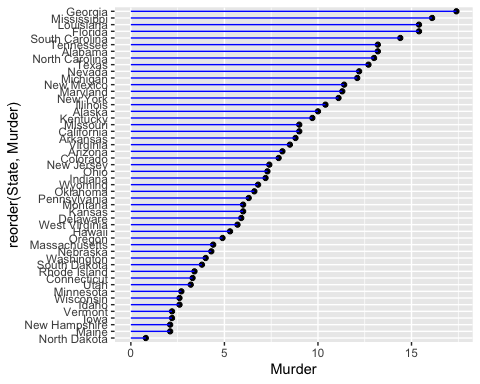
#FLIPPING CHARTS  
head(USArrests)

## 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()

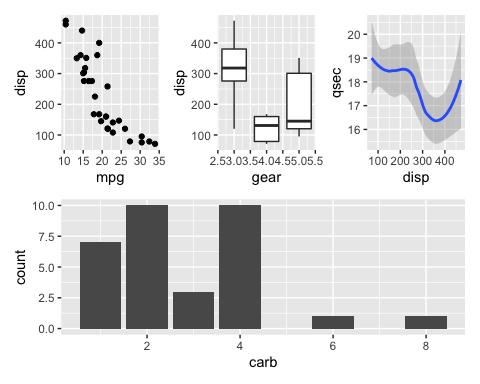


ggplot(USArrests) +  
 aes(x=reorder(State,Murder), y=Murder) +  
 geom\_point() +  
 geom\_segment(aes(x=State,   
 xend=State,   
 y=0,   
 yend=Murder), color="blue") +  
 coord\_flip()



#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 expectancy') +  
 #transition\_time(year) +  
 #shadow\_wake(wake\_length = 0.1, alpha = FALSE)  
  
  
#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'



tinytex::install\_tinytex()

## The directory /usr/local/bin is not writable. I recommend that you make it writable. See https://github.com/yihui/tinytex/issues/24 for more info.