class05sript.R

zaidarodriguez

2021-10-13

```r
#class 05: data visualization (ggplot2)

library(ggplot2)

#will use inbuilt "cars dataset
head(cars)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```
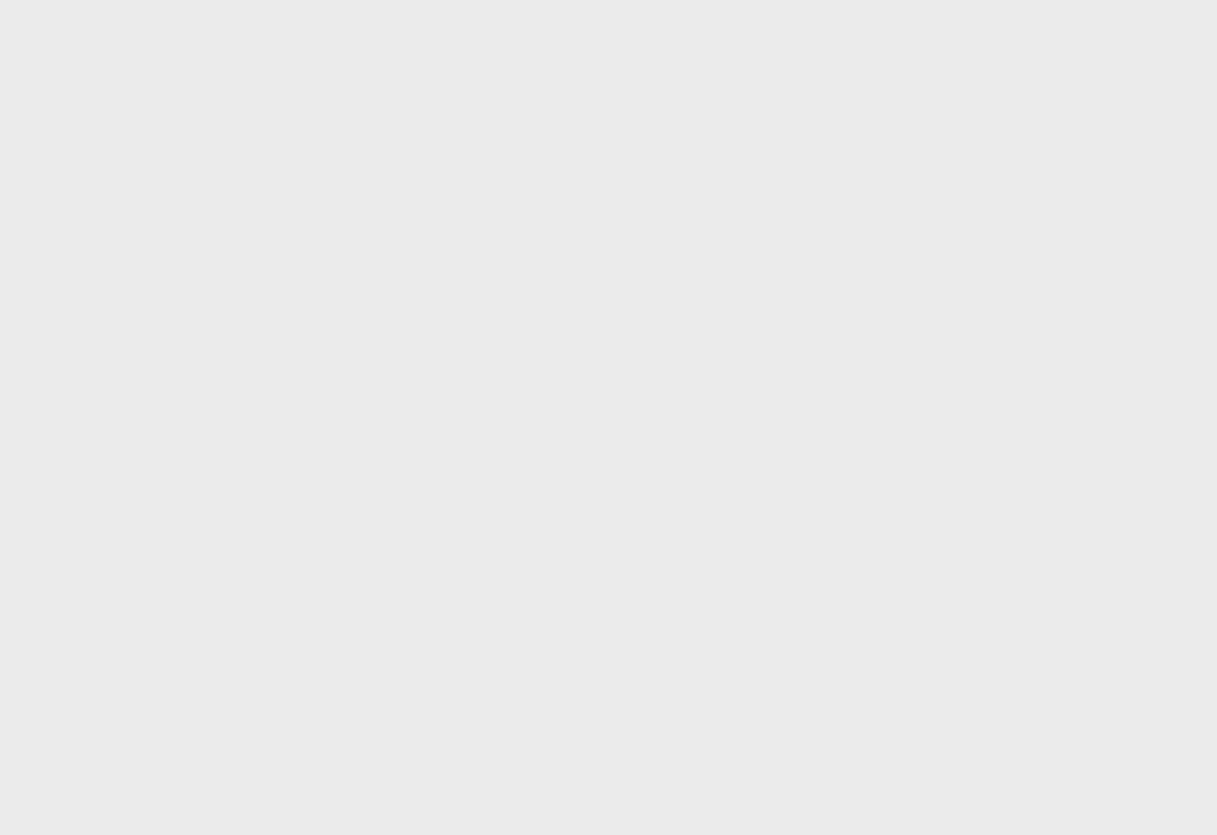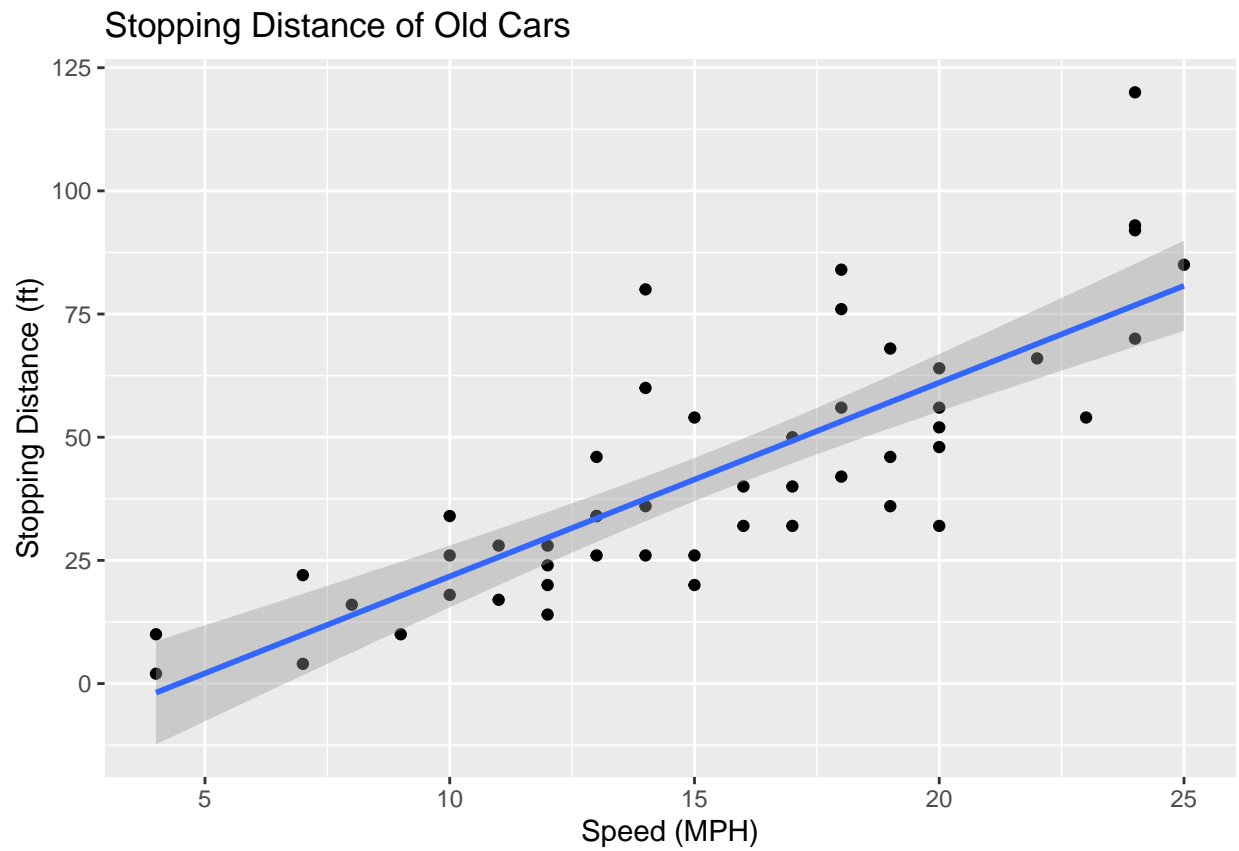
```r
# all ggplots have min. 3 layers
ggplot(cars)
```
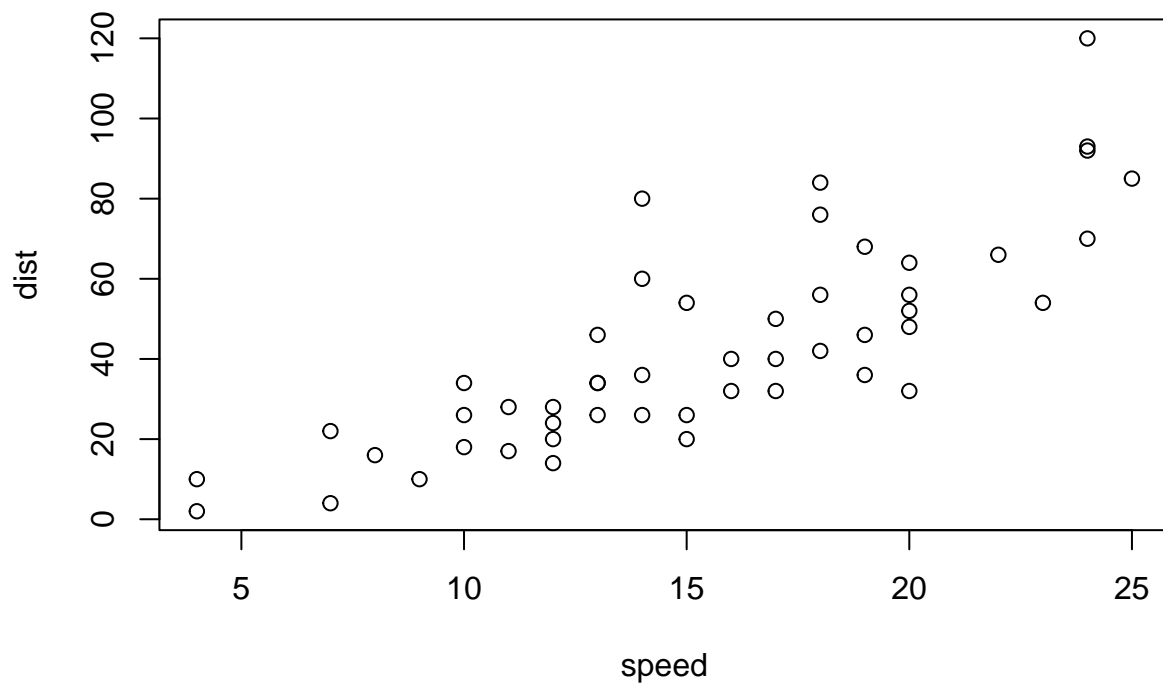
```
# need to add the other layers
ggplot(cars, aes(x=speed, y=dist)) + geom_point() + geom_smooth(method = "lm") +
  labs(title = "Stopping Distance of Old Cars",
       x= "Speed (MPH)", y= "Stopping Distance (ft)")
```

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

Stopping Distance of Old Cars

```
# sidenote: other graphic systems available in R ("base" R graphics)

plot(cars)
```

```r
# workbook 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
```

```r
# how many genes in this dataset?
nrow(genes)
```

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

```r
# how many genes are "up"?
#this looks at the state column
genes$State
```

```
##     [1] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
##     [6] "unchanging" "unchanging" "unchanging" "unchanging" "up"
```

```
## [5141] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5146] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5151] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5156] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5161] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5166] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5171] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5176] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5181] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5186] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5191] "unchanging" "unchanging" "unchanging" "unchanging" "unchanging"
## [5196] "unchanging"
```

```r
#to look at the specific numbers
table(genes$State)
```

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

```r
# to see what percentage of genes are up?
table(genes$State)/nrow(genes) * 100
```

```
##
##       down unchanging         up
##   1.385681  96.170131   2.444188
```
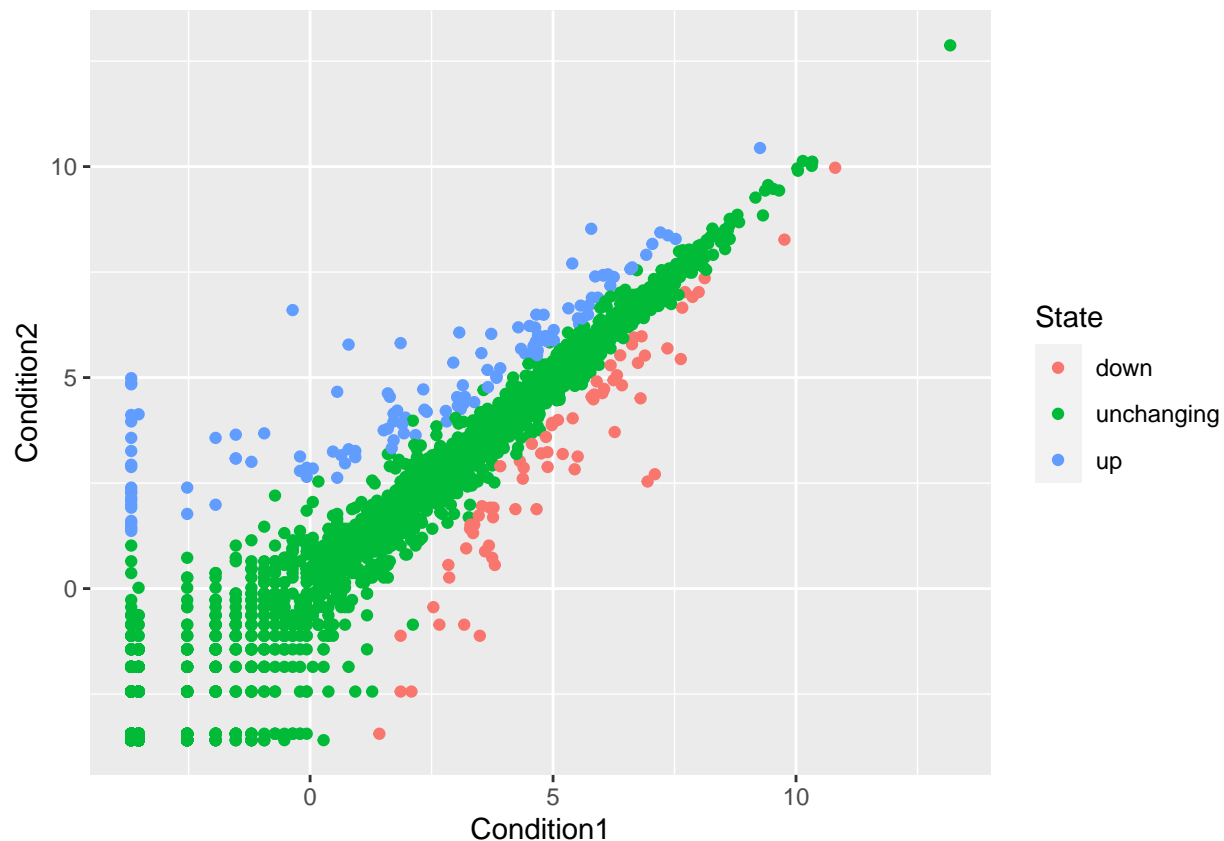
```r
# to round up:
round(table(genes$State)/nrow(genes) * 100, 2)
```

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

```r
# make a figure
p <- ggplot(genes) +
  aes(x=Condition1, y=Condition2, col=State)+
  geom_point()
p
```

```
# change the colors
p <- p + scale_color_manual(values = c("orchid", "orange", "lightblue")) +
  labs(title = "Gene Expression Changes Upon Drug Treatment",
       x= "Control(no drug)", y= "Drug Treatment")

# time to work with gapminder
library(gapminder)
url <- "https://raw.githubusercontent.com/jennybc/gapminder/master/inst/extdata/gapminder.tsv"
gapminder <- read.delim(url)
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
```

```
head(gapminder)
```

```
##       country continent year lifeExp      pop gdpPercap
## 1 Afghanistan      Asia 1952  28.801  8425333  779.4453
## 2 Afghanistan      Asia 1957  30.332  9240934  820.8530
## 3 Afghanistan      Asia 1962  31.997 10267083  853.1007
## 4 Afghanistan      Asia 1967  34.020 11537966  836.1971
## 5 Afghanistan      Asia 1972  36.088 13079460  739.9811
## 6 Afghanistan      Asia 1977  38.438 14880372  786.1134
```

```r
# %>% is called a pipe. We will tell R to filter the gapminder dataset to year 2007
gapminder_2007 <- gapminder %>% filter(year==2007)

# lets make a figure of year vs lifeExp
# add a box plot on top of the jitters
# alpha changes the transprency of the dot/box
ggplot(gapminder) +
  aes(x=year, y=lifeExp, col=continent) +
  geom_jitter(width= 0.3, alpha=0.4) +
  geom_boxplot(aes(group=year, alpha=0.4))
```
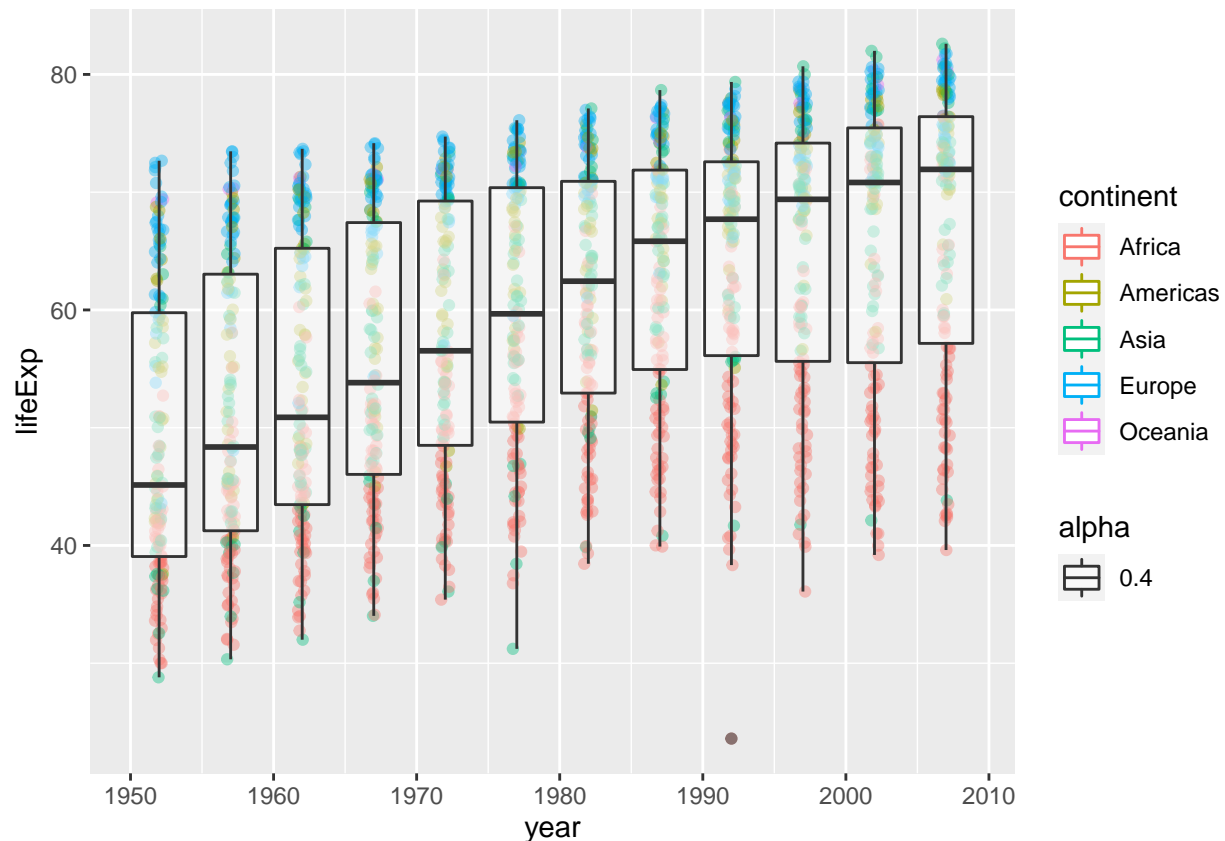


```r
# try a vln plot instead of a box plot
g <- ggplot(gapminder) +
  aes(x=year, y=lifeExp, col=continent) +
  geom_jitter(width= 0.3, alpha=0.4) +
  geom_violin(aes(group=year), alpha=0.2, draw_quantiles =0.5)
g
```
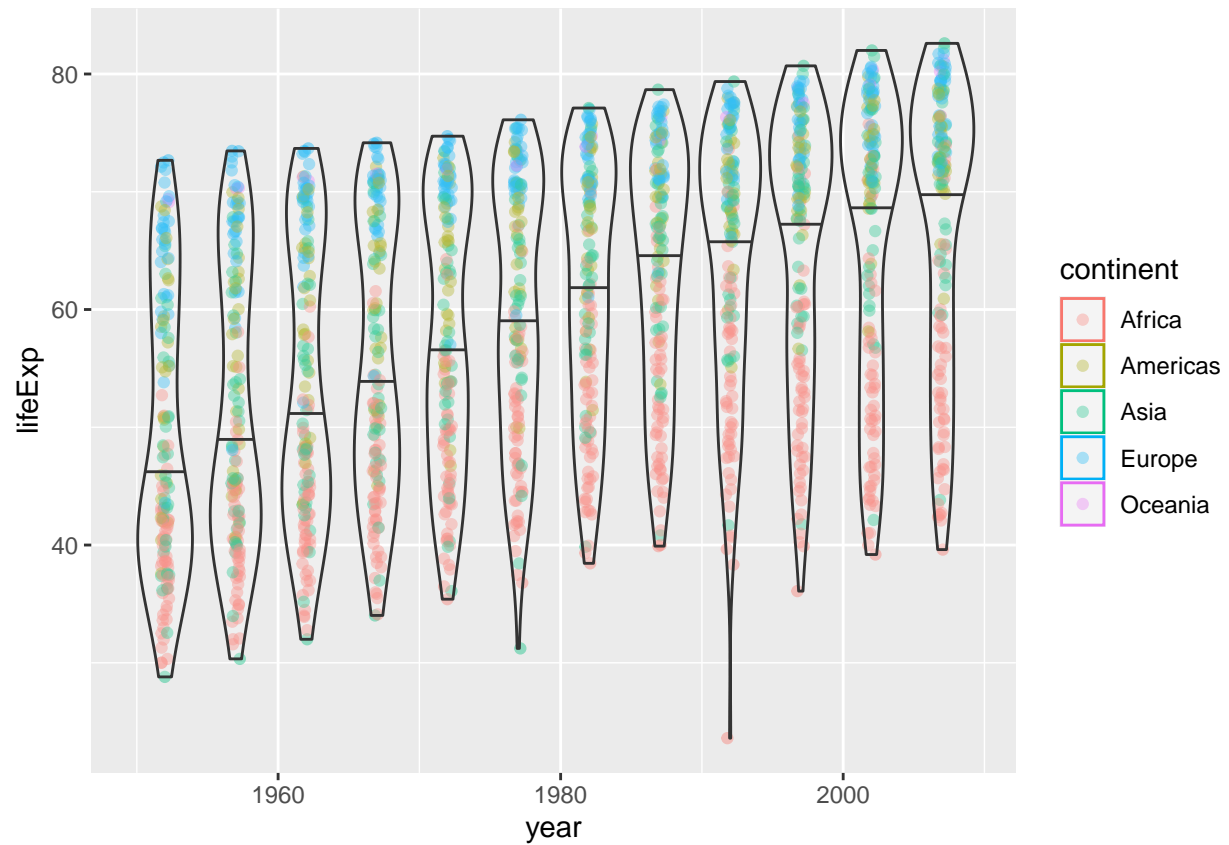
```
#install.packages("plotly")
#install the plotly
#plotly lets you hover over the data and see specifics of the points
#library(plotly)
#ggplotly()
```