class05.R

nicolejacobson

2021-10-13

```
# Class 5: Data visualization Nicole Jacobson

#will learn how to use an array of plots using the ggplot2 package using prebuilt cars dataset

#looks at the top of the data set

head(cars)
```

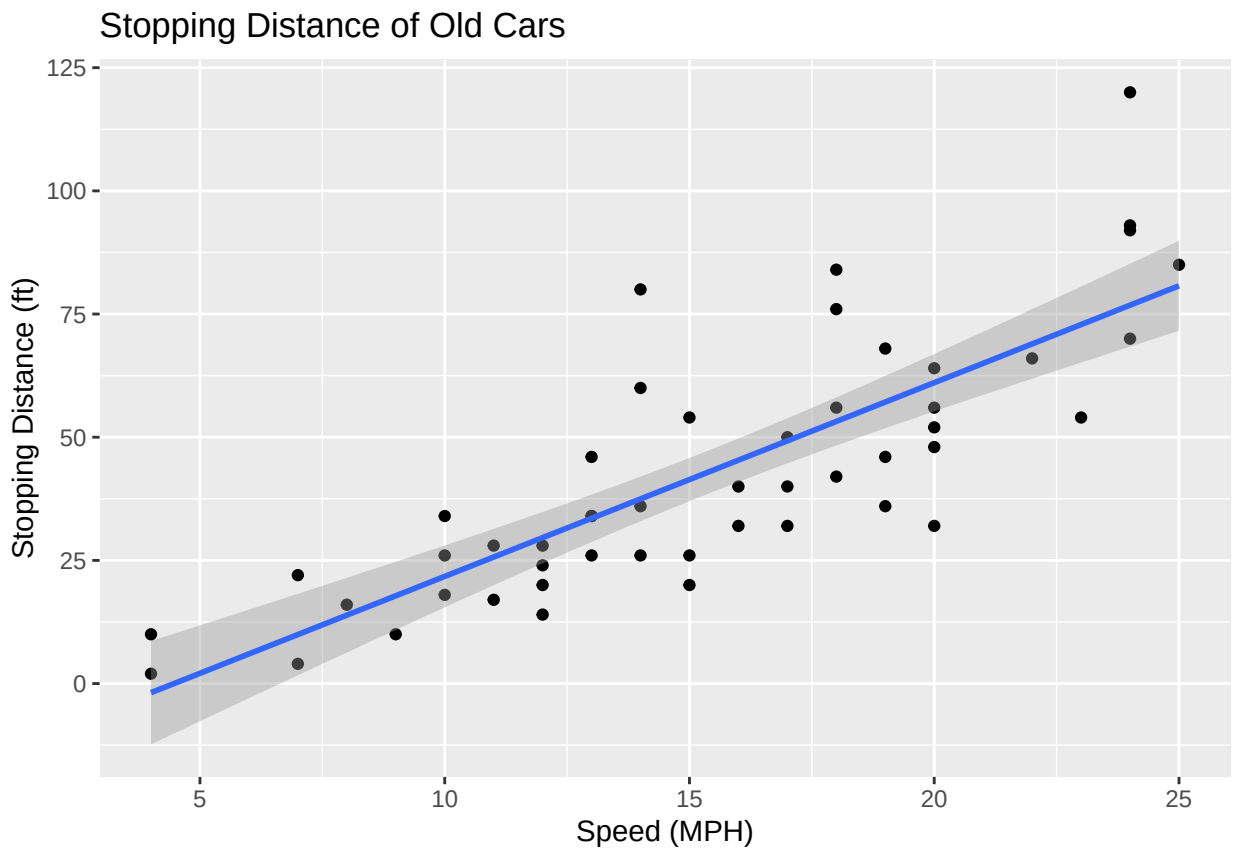
```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

```
# all ggplots have at least 3 layers (data, ascetics -> x/y axis, labels ine color/thickness etc.,
# geometry -> type of graph ie scatter, bar, etc.)

#method=lm linear model will linearize the smooth line of best fit line, otherwise it will be curved
library(ggplot2)

ggplot(cars) + aes(x=speed, y=dist) + geom_point() + geom_smooth(method="lm") +
  ggtitle("Stopping Distance of Old Cars") + labs( x="Speed (MPH)", y="Stopping Distance (ft)")
```

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




```
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
```

```
ncol(genes)
```

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

```
# calls a  particular column from that dataset. Tells me how many ups, downs, and unchanging there are

table(genes$State)
```

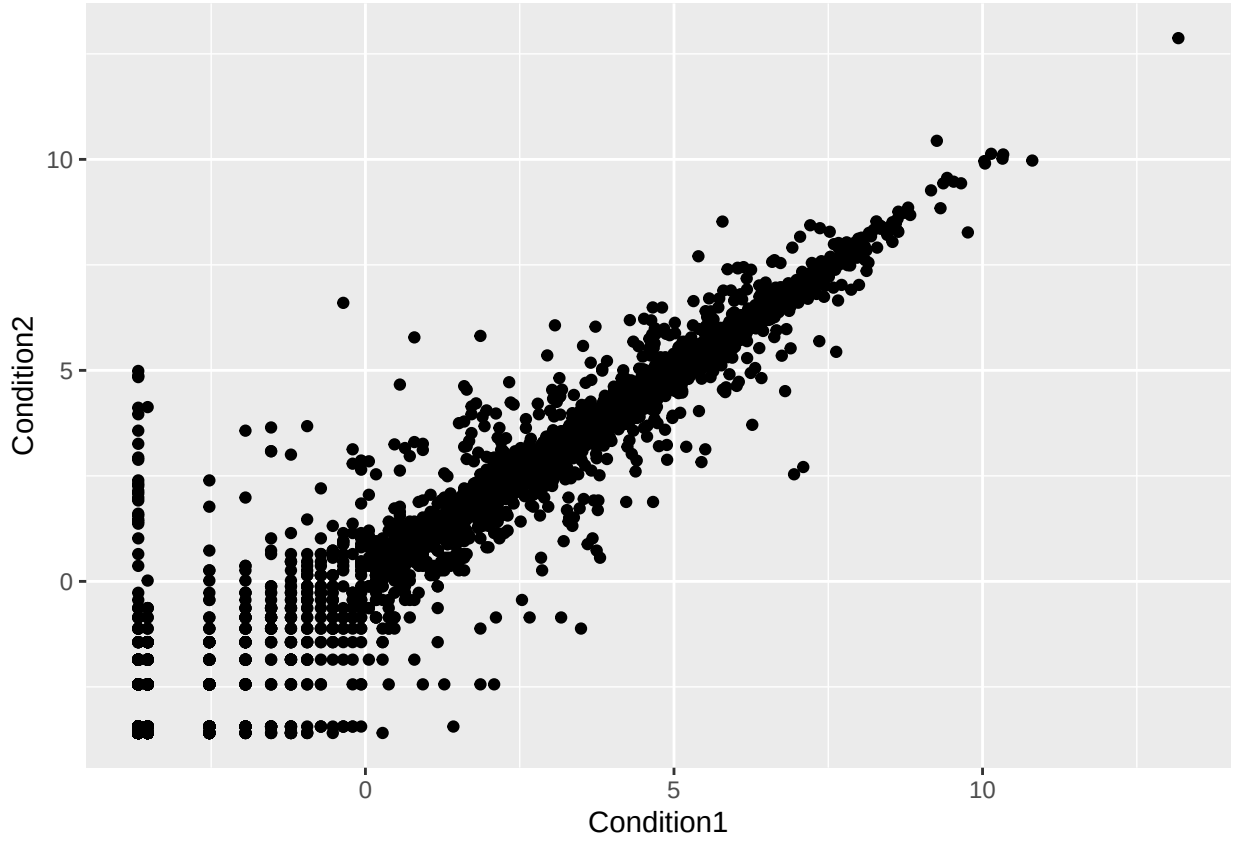
```
##
##       down unchanging         up
##         72       4997        127
```

```
#what percent of total genes are up-regulated in this dataset? Here is the percent of genes that are up
# by dividing table of the column State  by the total number of entries which is nrow(genes)
#  x100 is to turn it into percent, 2 is how many significant digits
# round will round the number to how many significant digits we want

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

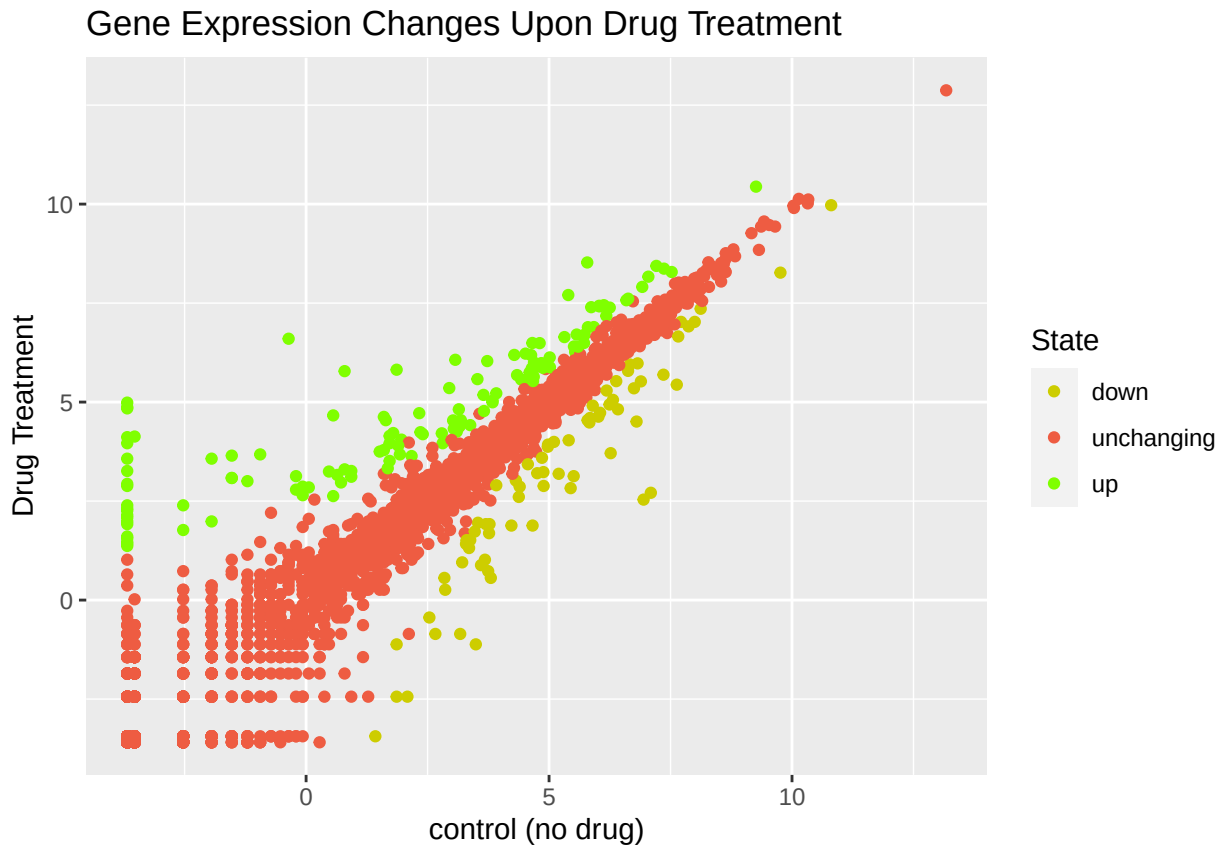
```
##
##       down unchanging         up
##       1.39      96.17       2.44
```

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




```
p <- ggplot(genes) + aes(x=Condition1, y=Condition2, col=State) + geom_point() + labs(x="control (no dr

p + scale_colour_manual(values=c("yellow3", "tomato2", "chartreuse"))
```

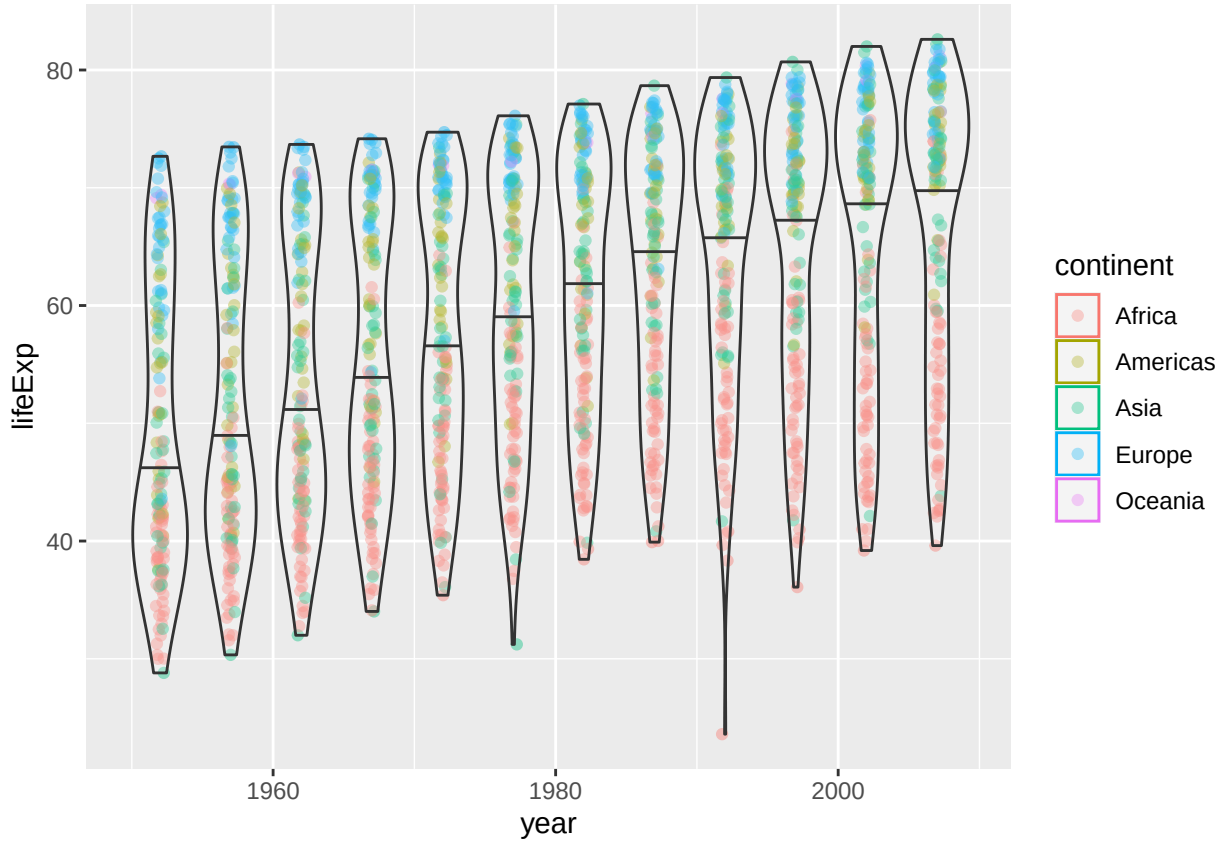



```
#now let's explore the gapminder dataset

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

```
## # A tibble: 6 x 6
##   country     continent  year lifeExp      pop gdpPercap
##   <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Afghanistan Asia       1952    28.8  8425333      779.
## 2 Afghanistan Asia       1957    30.3  9240934      821.
## 3 Afghanistan Asia       1962    32.0 10267083      853.
## 4 Afghanistan Asia       1967    34.0 11537966      836.
## 5 Afghanistan Asia       1972    36.1 13079460      740.
## 6 Afghanistan Asia       1977    38.4 14880372      786.
```

```
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("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 x 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) + geom_point(alpha=0.5)
```