Lab 2

Michael Egle and John Chandara; GitHub: michaelegle and mrpotatofactory

2/5/2020

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
choco <- read.csv("https://xdaiisu.github.io/ds202materials/hwlabs/choco.csv")
```

1. What is the overall number of chocolate bars rated?

Amount of duplicated rows.

```
nrow(choco[! duplicated(choco), ]) - nrow(choco)
```

```
## [1] 0
```

```
length(choco$Rating)
```

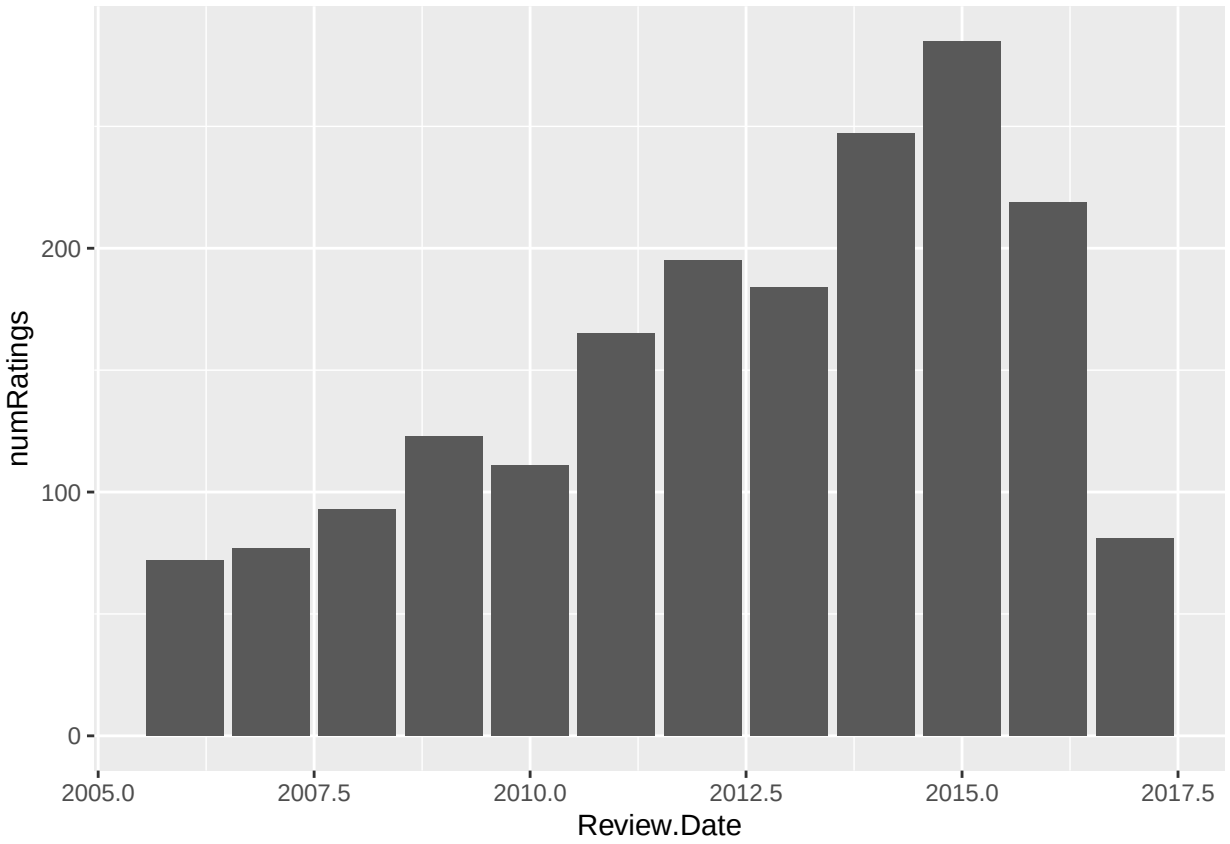
```
## [1] 1852
```

Because there are no duplicates, we can conclude there are 1852 rated bars.

2. How does the number of ratings depend on the year? Draw a bar chart of the number of reports.

```
choco %>%
  group_by(Review.Date) %>%
  summarize(numRatings = n()) -> choco_ratings

choco_ratings %>%
  ggplot(aes(x = Review.Date, y = numRatings)) +
  geom_bar(stat = "identity")
```



The number of ratings gradually increases with a peak in 2015 and then a decrease in 2016 and 2017.

3. Which are the three locations with the most ratings? How do ratings compare across these company locations?

```
AggRatings <- choco %>%
  group_by(Company.Location) %>%
    summarize(AggRatings.Location = n()) %>%
      arrange(desc(AggRatings.Location))

MostPopular <- head(AggRatings, 3)

print(MostPopular)
```

```
## # A tibble: 3 x 2
##   Company.Location AggRatings.Location
##   <fct>                          <int>
## 1 U.S.A.                           787
## 2 France                           158
## 3 Canada                           132
```

```
choco %>%
  filter(Company.Location == 'U.S.A.') %>%
    pull(Rating) %>%
      summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##   1.500   2.750   3.250   3.162   3.500   4.000
```