

Assignment 2

2022-11-22

```
library(jsonlite)

## Warning: package 'jsonlite' was built under R version 4.2.2

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.2

## — Attaching packages ————— tidyverse
1.3.2 —
## ✓ ggplot2 3.4.0      ✓ purrr  0.3.5
## ✓ tibble  3.1.8      ✓ dplyr  1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.4.1
## ✓ readr   2.1.3      ✓ forcats 0.5.2

## Warning: package 'ggplot2' was built under R version 4.2.2
## Warning: package 'tidyr' was built under R version 4.2.2
## Warning: package 'readr' was built under R version 4.2.2
## Warning: package 'purrr' was built under R version 4.2.2
## Warning: package 'dplyr' was built under R version 4.2.2
## Warning: package 'stringr' was built under R version 4.2.2
## Warning: package 'forcats' was built under R version 4.2.2

## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ purrr::flatten() masks jsonlite::flatten()
## ✗ dplyr::lag() masks stats::lag()

library(tidytext)

## Warning: package 'tidytext' was built under R version 4.2.2

#Q1: Import the data
Hotels<-read_csv("C:/Users/HP/Downloads/Hotels.csv")

## Warning: One or more parsing issues, call `problems()` on your data frame
for details,
## e.g.:
```

```

## dat <- vroom(...)
## problems(dat)

## Rows: 6413 Columns: 58
## — Column specification

```

```

## Delimiter: ","
## chr (29): business_id, name, address, city, state, postal_code,
categories, ...
## dbl (6): latitude, longitude, stars, review_count, is_open,
attributes.Rest...
## lgl (23): attributes.BusinessAcceptsCreditCards, attributes.GoodForKids,
att...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.

HotelReviews<-read_csv("C:/Users/HP/Downloads/HotelReviews.csv")

## Rows: 3647 Columns: 23
## — Column specification

```

```

## Delimiter: ","
## chr (8): review_id, user_id, business_id, text, date, name,
yelping_since, ...
## dbl (14): review_count, fans, average_stars, compliment_hot,
compliment_more...
## num (1): elite
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.

glimpse(Hotels)

## Rows: 6,413
## Columns: 58
## $ business_id      <chr> "mKTq1T_IAp1DpHUcMzOXkw",
"FrMRH..."
## $ name              <chr> "Topsy Ryde", "Sin City
Party Bu..."
## $ address           <chr> NA, "4005 W Reno Ave, Ste
C", "3..."
## $ city              <chr> "Gastonia", "Las Vegas",
"Montr  ..."
## $ state             <chr> "NC", "NV", "QC", "QC",
"AZ", "A..."
## $ postal_code       <chr> "28054", "89118", "H2X 1H6",
"H2..."
## $ latitude          <dbl> 35.25284, 36.09688,

```

```

45.51495, 45...
## $ longitude                <dbl> -81.15270, -115.19290, -
73.56571...
## $ stars                    <dbl> 3.5, 4.0, 4.5, 3.5, 4.0,
4.5, 2...
## $ review_count             <dbl> 3, 18, 4, 45, 3, 12, 10, 17,
441...
## $ is_open                  <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1,...
## $ categories               <chr> "Hotels & Travel,
Transportation...
## $ attributes.BusinessAcceptsCreditCards <lgl> TRUE, TRUE, NA, NA, TRUE,
NA, NA...
## $ attributes.BikeParking   <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.GoodForKids    <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.BusinessParking <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.ByAppointmentOnly <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.RestaurantsPriceRange2 <dbl> NA, NA, 2, 2, NA, NA, NA, 1,
3, ...
## $ attributes.DogsAllowed    <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.WiFi           <chr> NA, NA, "'free'", "u'free'",
NA,...
## $ attributes.RestaurantsAttire <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.RestaurantsTakeOut <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.NoiseLevel     <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.RestaurantsReservations <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.RestaurantsGoodForGroups <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.HasTV          <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.Alcohol        <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.RestaurantsDelivery <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.OutdoorSeating <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.Caters         <lgl> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.WheelchairAccessible <chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...
## $ attributes.AcceptsInsurance <lgl> NA, NA, NA, NA, NA, NA, NA,

```

NA, ...	
## \$ attributes.RestaurantsTableService	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.Ambience	<chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.GoodForMeal	<chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.HappyHour	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.BusinessAcceptsBitcoin	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.BYOB	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.Corkage	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.GoodForDancing	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.CoatCheck	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.BestNights	<chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.Music	<chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.Smoking	<chr> NA, NA, NA, NA, NA, "u'no'",
NA,...	
## \$ attributes.DietaryRestrictions	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.DriveThru	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.HairSpecializesIn	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.BYOBCorkage	<chr> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.AgesAllowed	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.RestaurantsCounterService	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ attributes.Open24Hours	<lg1> NA, NA, NA, NA, NA, NA, NA,
NA, ...	
## \$ hours.Monday	<chr> "0:0-0:0", "8:0-20:0", NA,
"0:0-...	
## \$ hours.Tuesday	<chr> "9:0-0:0", "8:0-20:0", NA,
"0:0-...	
## \$ hours.Wednesday	<chr> "9:0-0:0", "8:0-20:0", NA,
"0:0-...	
## \$ hours.Thursday	<chr> "9:0-0:0", "8:0-20:0", NA,
"0:0-...	
## \$ hours.Friday	<chr> "9:0-2:30", "8:0-20:0", NA,
"0:0...	
## \$ hours.Saturday	<chr> "9:0-2:30", "10:0-20:0", NA,

```
"0:...
## $ hours.Sunday          <chr> "12:0-0:0", "10:0-20:0", NA,
"0:...
```

#Q2: Creating tibble that excludes columns

```
Hotels_tbl<-as_tibble(Hotels)
reviews_tbl<-as_tibble(HotelReviews)
Hotels_tbl=Hotels_tbl %>%
  select(-starts_with("hours"))
```

#3.1

```
Hotels_tbl %>%
  filter(str_detect(state,"NV"))%>%
  filter(str_detect(categories,"Car Rental"))%>%
  count()
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    204
```

#3.2

```
Hotels_tbl %>%
  filter(str_detect(categories,"Car Rental"))%>%
  group_by(state)%>%
  count()
```

```
## # A tibble: 11 × 2
## # Groups:   state [11]
##   state      n
##   <chr> <int>
## 1 AB         58
## 2 AZ        243
## 3 IL          6
## 4 NC         81
## 5 NV        204
## 6 OH        100
## 7 ON        198
## 8 PA         59
## 9 QC         39
## 10 SC          5
## 11 WI         18
```

#3.3

```
Hotels_tbl %>%
  filter(str_detect(categories,"Car Rental"))%>%
  filter(is_open == 1 )%>%
  count()
```

```
## # A tibble: 1 × 1
##       n
```

```
## <int>
## 1 911

Hotels_tbl %>%
  filter(str_detect(categories, "Car Rental"))%>%
  summarise(proportion=sum(is_open/length(is_open)))
```

```
## # A tibble: 1 × 1
##   proportion
##   <dbl>
## 1 0.901
```

#3.4

```
Hotels_tbl %>%
  select(state, review_count, stars)%>%
  group_by(state)%>%
  summarise(Average_stars = mean(stars))%>%
  arrange(desc(Average_stars))%>%
  head(5)
```

```
## # A tibble: 5 × 2
##   state Average_stars
##   <chr>         <dbl>
## 1 VT             5
## 2 QC            3.32
## 3 WI            3.28
## 4 NV            3.24
## 5 AZ            3.22
```

#3.5

```
Hotels_tbl %>%
  select(name, stars)%>%
  filter(stars==5)%>%
  group_by(name)%>%
  summarise(Five_star_reviews = n())%>%
  arrange(desc(Five_star_reviews))%>%
  head(10)
```

```
## # A tibble: 10 × 2
##   name                    Five_star_reviews
##   <chr>                    <int>
## 1 Enterprise Rent-A-Car      7
## 2 Diplomat Exotic Rentals   2
## 3 'Round Table Tours        1
## 4 2 Fly Us Hot Air Balloon Rides 1
## 5 24/7 Vegas VIP            1
## 6 360 Adventures            1
## 7 702 Housing                1
## 8 8th Wonder Studio          1
## 9 A-1 Vans Rental & Leasing  1
## 10 À la Carte                1
```

#3.6

```
Hotels_tbl %>%
  select(name)%>%
  group_by(name)%>%
  summarise(No_of_times = n())%>%
  arrange(desc(No_of_times))%>%
  head(10)
```

```
## # A tibble: 10 × 2
##   name                      No_of_times
##   <chr>                    <int>
## 1 Enterprise Rent-A-Car      235
## 2 Hertz Rent A Car          86
## 3 Avis Rent A Car           41
## 4 Bob Evans                 38
## 5 Motel 6                   37
## 6 Budget Car Rental         21
## 7 National Car Rental       16
## 8 Thrifty Car Rental        16
## 9 Alamo Rent A Car          15
## 10 Dollar Car Rental        15
```

#3.7

```
erac=Hotels_tbl %>%
  filter(str_detect(name,"Enterprise Rent-A-Car"))
```

#3.8

```
business_reviews<-erac%>%
  left_join(reviews_tbl, by="business_id")
```

#3.9

```
options(dplyr.summarise.inform = FALSE)
business_reviews%>%
  select(business_id,address,city,state,review_count.x)%>%
  group_by(business_id,address,city,state)%>%
  summarise(No_of_reviews=sum(review_count.x))%>%
  arrange(desc(No_of_reviews))%>%
  head(10)
```

```
## # A tibble: 10 × 5
## # Groups:   business_id, address, city [10]
##   business_id      address          city      state
##   <chr>          <chr>          <chr>    <chr>
##   <dbl>
## 1 _BELd09Ppp0aMUTW-ItJfw 7135 GilesPie St  Las Vegas  NV
## 658
## 2 r3Xx9Qk3xAMzYGIGFBckpQ 1805 E Sky Harbor Cir S  Phoenix    AZ
## 339
## 3 pkscmWKfr5ad6L1M6Jn9eA 1050 W Warm Springs Rd  Henderson  NV
## 119
```

```
## 4 IzOAFrYWu0Q7bFVQApZRBA 5489 R C Josh Birmingham Pkwy Charlotte NC
116
## 5 lz_TRWwKRAzHo_BzX_j7ng 7150 S Durango, Ste 180 Las Vegas NV
92
## 6 iRCgRxw33qEg4G1KcFpKvQ 8201 S Priest Dr Tempe AZ
88
## 7 q4_dkV1S4KEAJ0EJ1Nv5Pg 1000 Airport Blvd Pittsburgh PA
82
## 8 9g1GP5prqzQKw4zMpvkuQw 15075 N Hayden Rd Scottsdale AZ
73
## 9 NQCib9ceJhaIdfQ6ILwBJA 4635 N 7th St Phoenix AZ
73
## 10 4lpvbMfXQw2rUz16N_KtsA 1461 N Arizona Ave Chandler AZ
72
## # ... with abbreviated variable name ^No_of_reviews
```

#3.10

```
business_reviews%>%
  mutate(text_new = strsplit(text, " "))%>%
  unnest(text_new)%>%
  select(text_new)%>%
  mutate(text_new = str_trim(str_squish(text_new)))%>%
  filter(!text_new %in% c("enterprise", "rental", "car", "cars", NA, ""))%>%
  count(text_new)%>%
  arrange(desc(n))%>%
  head(10)
```

```
## # A tibble: 10 × 2
##   text_new      n
##   <chr>    <int>
## 1 the      124
## 2 to       98
## 3 I        91
## 4 and      84
## 5 a        64
## 6 was      57
## 7 we       34
## 8 in       31
## 9 had      28
## 10 for     26
```

#3.11

```
reviews_sentiment<-business_reviews %>%
  unnest_tokens(word, text)%>%
  inner_join(get_sentiments("afinn"), by="word")%>%
  group_by(review_id)%>%
  summarise(sentiment= mean(value))%>%
  ungroup()

reviews_sentiment%>%
```



```

arrange(sentiment)%>%
top_n(-10,sentiment)%>%
left_join(business_reviews,by="review_id")%>%
select(address,city,date,sentiment,text)

```

```

## # A tibble: 11 x 5
##   address          city      date      sentiment text
##   <chr>          <chr>    <chr>      <dbl> <chr>
## 1 947 Dovercourt Rd Toronto 1/21/14 16:35    -3    "The
servic...
## 2 6821 Wilkinson Blvd Belmont 9/11/14 15:22   -1.2    "I
rented a...
## 3 2810 E Bell Rd Phoenix 4/6/18 21:15    -1    "Just
had t...
## 4 6920 E Cave Creek Rd Cave Creek 8/29/17 16:48  -0.889 "The
car we...
## 5 303 Cleveland St Elyria 11/28/19 4:07  -0.556 "Awful
expe...
## 6 303 Cleveland St Elyria 4/22/18 14:26    0
"Customer s...
## 7 1140 N Gilbert Rd, Ste 111 Gilbert 9/24/19 16:43   0.375 "Just
drove...
## 8 13740 W Van Buren St Goodyear 9/8/15 15:59   0.727
"Service wa...
## 9 4742 W Glendale Ave Glendale 4/4/15 19:56   0.75 "I had
a re...
## 10 303 Cleveland St Elyria 5/10/19 19:58   0.8 "Wow
just w...
## 11 27521 Lorain Rd North Olmsted 9/8/15 12:21   0.8 "The
repres...

```

#3.12

```

reviews_sentiment%>%
  arrange(desc(sentiment))%>%
  top_n(10,sentiment)%>%
  left_join(business_reviews,by="review_id")%>%
  select(address,city,date,sentiment,text)

```

```

## # A tibble: 11 x 5
##   address          city      date      sentiment
##   <chr>          <chr>    <chr>      <dbl>
## 1 10301 N Scottsdale Rd Scottsdale 3/10/16 12:41    3
leas...
## 2 3441 Dundas Street W Toronto 2/9/17 14:12    3
Great ...
## 3 7135 Gillespie St Las Vegas 12/26/14 1:09    3
good e...
## 4 7150 S Durango, Ste 180 Las Vegas 9/28/18 1:33   2.5
I

```

```

was ...
## 5 7340 Smiths Corners Blvd, Ste 100 Charlotte 8/21/15 12:43 2.33
The se...
## 6 4517 W Flamingo Rd Las Vegas 10/19/15 23:33 2.22 I
came...
## 7 534 S Boulder Hwy Henderson 3/30/16 14:59 2.17
Enterp...
## 8 7A - 1199 Kingston Road Pickering 9/12/18 20:53 2.11 I
book...
## 9 7150 S Durango, Ste 180 Las Vegas 6/30/19 0:50 2.1
I've r...
## 10 7150 S Durango, Ste 180 Las Vegas 11/18/19 21:35 2
Wanted...
## 11 7150 S Durango, Ste 180 Las Vegas 8/5/17 6:29 2
I've r...

```

#4.1

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.2.2
```

```
## Loading required package: timechange
```

```
## Warning: package 'timechange' was built under R version 4.2.2
```

```
##
```

```
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## date, intersect, setdiff, union
```

```
business_reviewsDateFormatted<-business_reviews%>%
```

```
  mutate(date_formatted = as_date(strptime(date,"%m/%d/%Y %H:%M")),
         month_formatted = month(strptime(date,"%m/%d/%Y %H:%M",)),
         day_formatted = day(strptime(date,"%M/%m/%d %H:%M")),
         year_formatted = year(strptime(date,"%m/%d/%Y %H:%M")),
         hour_formatted = hour(strptime(date,"%m/%d/%Y %H:%M")))
```

```
business_reviewsDateFormatted
```

```
## # A tibble: 242 × 78
```

```
##   business_id name.x address city state posta...1 latit...2 longi...3 stars
revie...4
```

```
##   <chr>         <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>
<dbl>
```

```
## 1 xo0_NBXWjRW... Enter... 4000 I... Madi... WI 53704 43.1 -89.3 3
23
```

```
## 2 AucmLoCNGDk... Enter... 9773 S... Stre... OH 44241 41.3 -81.4 3.5
4
```

```
## 3 LuVF0lyyZ23... Enter... 1000 R... Wood... ON L4L 5X3 43.8 -79.6 2
4
```

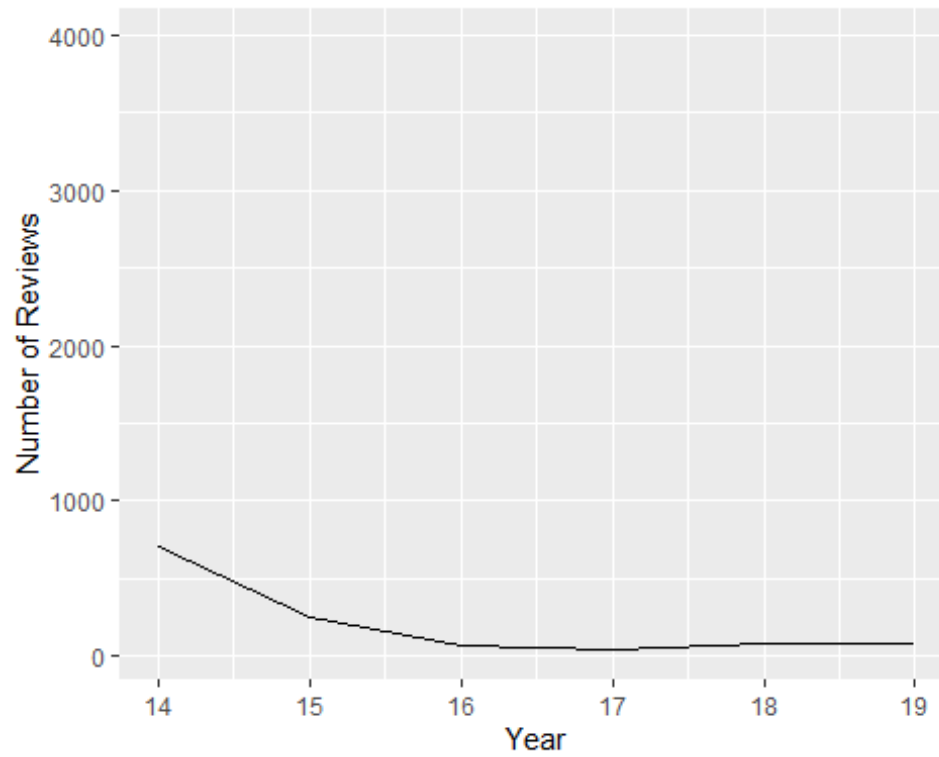
```

## 4 7A9CrMSBbXu... Enter... 1131 W... Tempe AZ      85282      33.4  -112.    4
33
## 5 jvKGjH5gTzE... Enter... 6646 O... Madi... WI      53719      43.1  -89.5    4
8
## 6 q4_dkVlS4KE... Enter... 1000 A... Pitt... PA      15231      40.5  -80.3    3.5
82
## 7 qSkGFtWm2yw... Enter... 8635 S... Las ... NV      89139      36.0  -115.    4
41
## 8 hPlNFqGNge-... Enter... 303 Cl... Elyr... OH      44035      41.4  -82.1    1
3
## 9 hPlNFqGNge-... Enter... 303 Cl... Elyr... OH      44035      41.4  -82.1    1
3
## 10 hPlNFqGNge-... Enter... 303 Cl... Elyr... OH      44035      41.4  -82.1    1
3
## # ... with 232 more rows, 68 more variables: is_open <dbl>, categories
<chr>,
## #   attributes.BusinessAcceptsCreditCards <lgl>, attributes.BikeParking
<chr>,
## #   attributes.GoodForKids <lgl>, attributes.BusinessParking <chr>,
## #   attributes.ByAppointmentOnly <lgl>,
## #   attributes.RestaurantsPriceRange2 <dbl>, attributes.DogsAllowed <chr>,
## #   attributes.WiFi <chr>, attributes.RestaurantsAttire <chr>,
## #   attributes.RestaurantsTakeOut <lgl>, attributes.NoiseLevel <chr>, ...

business_reviewsDateFormatted%>%
  select(year_formatted,review_count.x)%>%
  group_by(year_formatted)%>%
  summarise(No_of_reviews = sum(review_count.x))%>%
  ggplot(aes(x=year_formatted,y=No_of_reviews))+
  geom_line()+
  labs(y= "Number of Reviews", x = "Year")

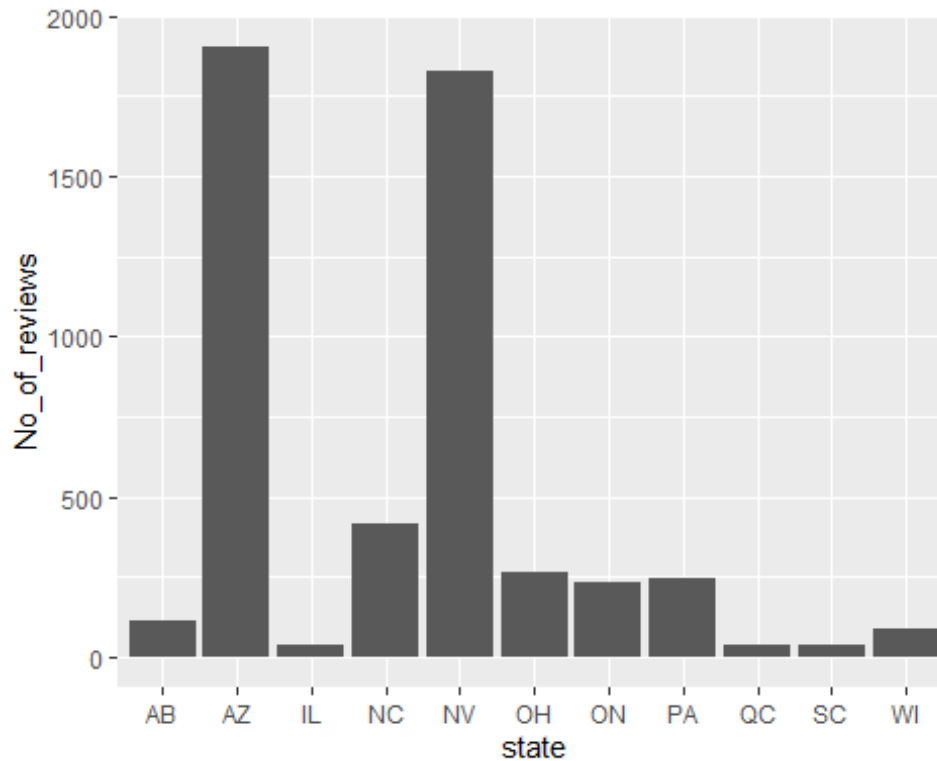
## Warning: Removed 1 row containing missing values (`geom_line()`).

```



#4.2

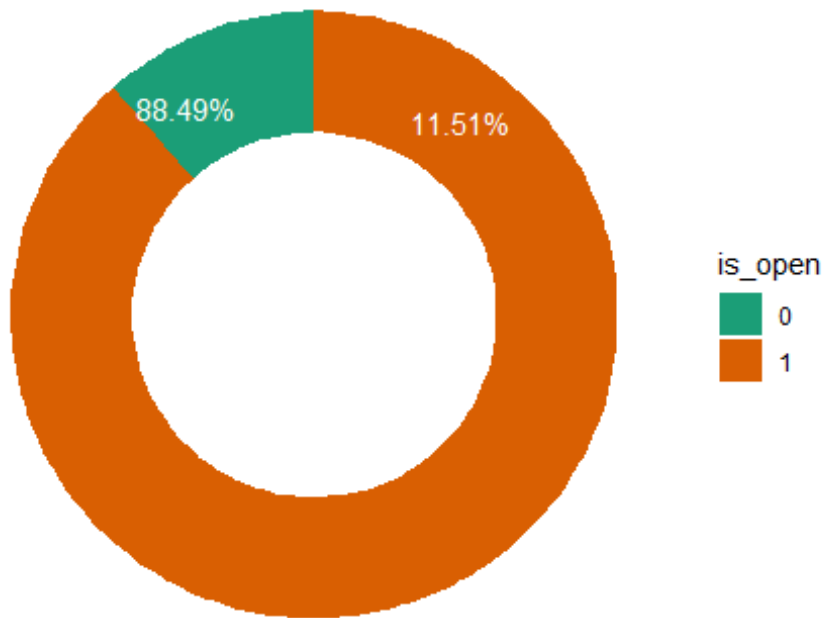
```
business_reviewsDateFormatted%>%  
  select(state,review_count.x)%>%  
  group_by(state)%>%  
  summarise(No_of_reviews = sum(review_count.x))%>%  
  ggplot(aes(x=state,y=No_of_reviews))+  
  geom_col()
```



#4.3

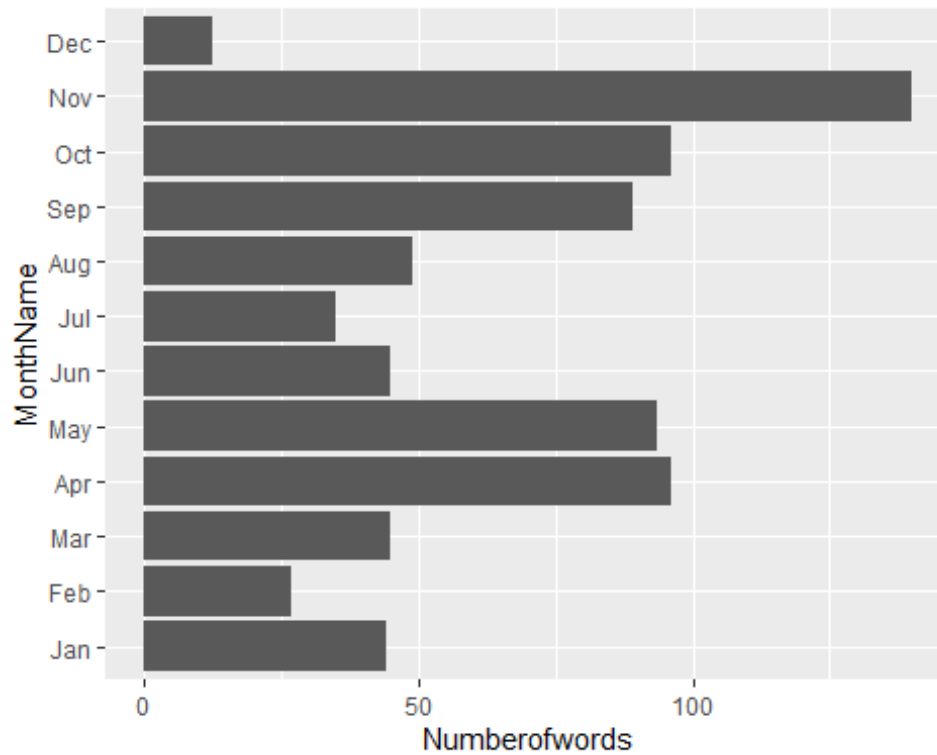
```
Hotels_tbl %>%
  group_by(is_open) %>%
  summarise(Count = n()) %>%

mutate(is_open = as.factor(is_open), percentage = round(Count/sum(Count)*100, 2), LabelPosition = cumsum(percentage) - .1*percentage) %>%
  ggplot(aes(x=2, y=percentage, fill=is_open)) +
  geom_bar(stat = "identity") +
  coord_polar("y") +
  geom_text(aes(y= LabelPosition, label=
paste(percentage, "%", sep="")), col="white") +
  theme_void() +
  scale_fill_brewer(palette = "Dark2") +
  xlim(.2, 2.5)
```



#4.4

```
business_reviewsDateFormatted%>%  
  mutate(NumberOfwords = str_count(text,boundary("word")))%>%  
  mutate(MonthName=month(ymd(date_formatted),label=TRUE))%>%  
  filter(!MonthName %in% NA)%>%  
  group_by(MonthName)%>%  
  summarize(NumberOfwords=median(NumberOfwords))%>%  
  ggplot(aes(x=MonthName ,y=NumberOfwords))+  
  geom_col()+  
  coord_flip()
```



#4.5

```
business_reviewsDateFormatted%>%
```

```
  ggplot(aes(x=compliment_funny, y=compliment_hot))+
```

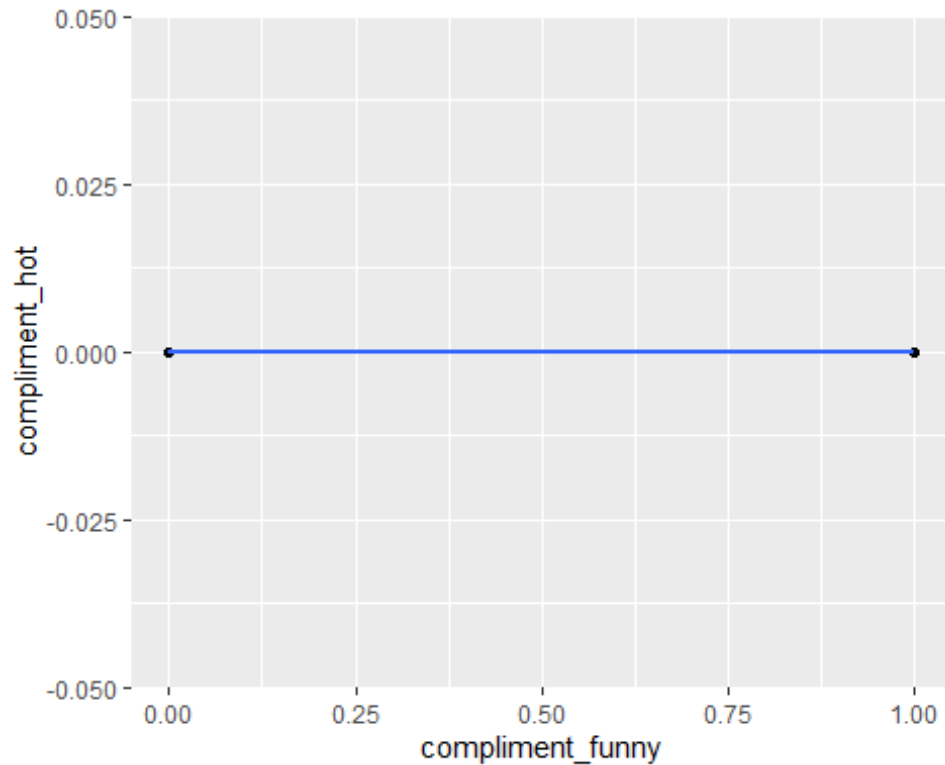
```
  geom_point()+
```

```
  geom_smooth(method=lm)
```

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

```
## Warning: Removed 213 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 213 rows containing missing values (`geom_point()`).
```



#4.6

```
library(wordcloud)

## Warning: package 'wordcloud' was built under R version 4.2.2
## Loading required package: RColorBrewer

createWordCloud = function(train)
{
  business_reviewsDateFormatted%>%
    unnest_tokens(word,text)%>%
    filter(!word %in% stop_words$word, !word %in% c("enterprise", "rental" ,
"car", "cars", NA))%>%
    count(word,sort = TRUE)%>%
    ungroup() %>%
    head(30) %>%
    with(wordcloud(word,n,max.words =30,colors=brewer.pal(8,"Dark2")))
}
createWordCloud(text)
```

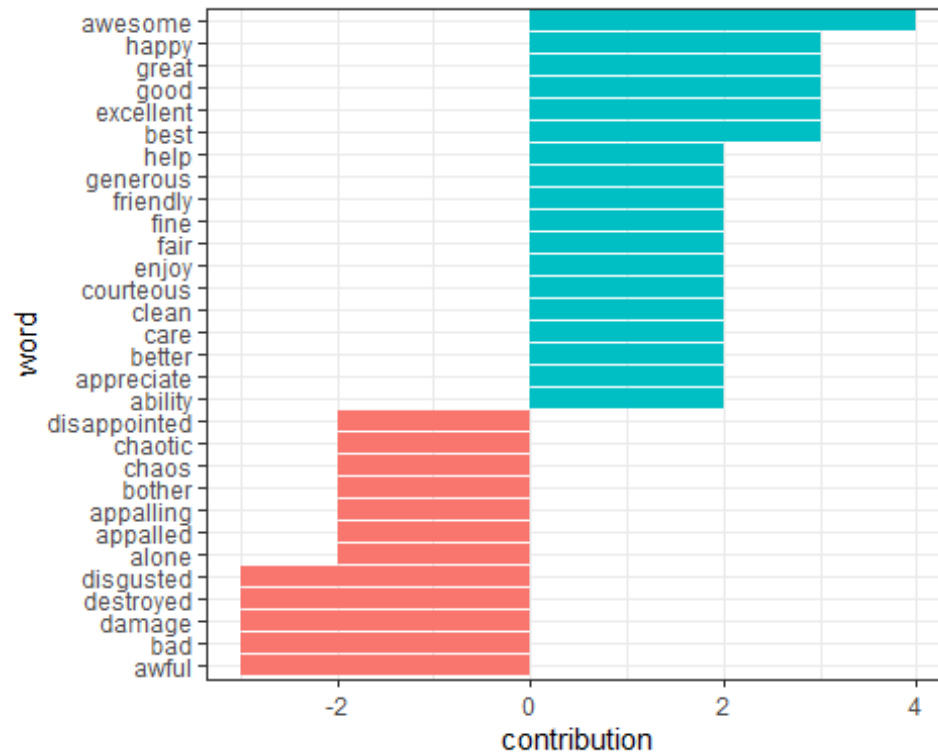



#4.7

```
positiveWordsBarGraph<- function(SC)
{
  contributions <- SC %>%
unnest_tokens (word, text) %>%
count(word, sort=TRUE) %>%
ungroup()%>%
inner_join(get_sentiments("afinn"), by="word") %>%
group_by (word) %>%
summarize(contribution=sum(value), n=n())

contributions %>%
top_n(30, abs(contribution)) %>%
mutate (word = reorder(word, contribution)) %>%
head (30) %>%
ggplot(aes (word, contribution, fill = contribution > 0)) +
geom_col(show.legend=FALSE) +
coord_flip() + theme_bw()
}

positiveWordsBarGraph(business_reviews)
```



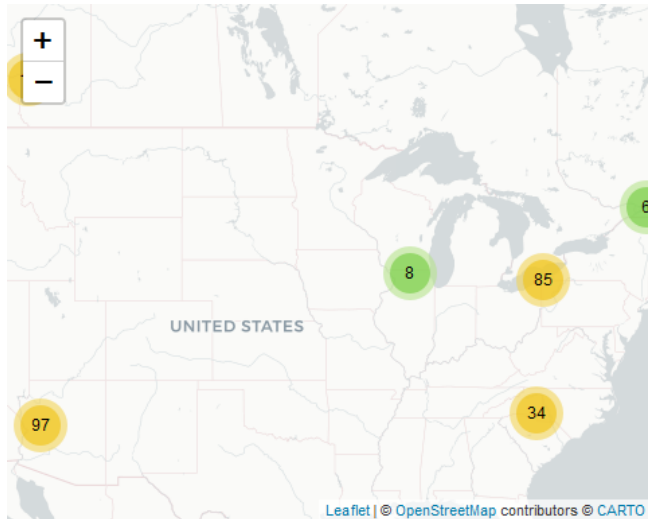
#4.8

```
library(leaflet)
```

```
## Warning: package 'leaflet' was built under R version 4.2.2
```

```
pal <- colorFactor(c("purple", "red", "orange", "black", "blue"),
  domain=unique(business_reviews$stars))
```

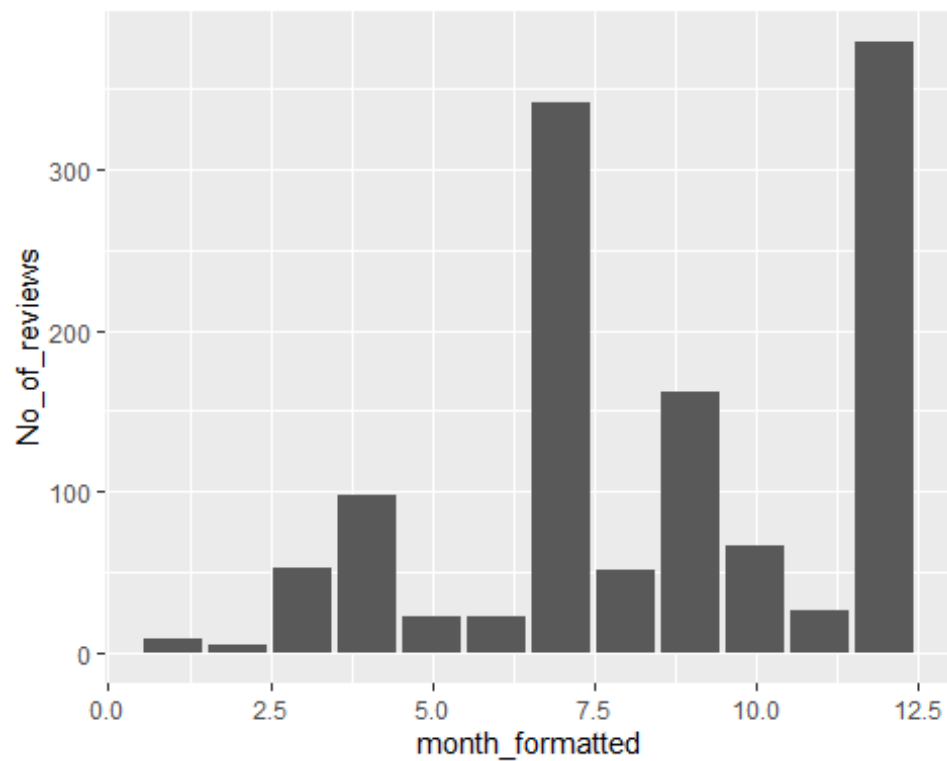
```
map<-leaflet(business_reviews) %>%
addProviderTiles ("CartoDB.Positron") %>%
addCircleMarkers (
  color= ~pal(business_reviews$stars),
  stroke = FALSE, fillOpacity = 0.5,
  lat=business_reviews$latitude,
  lng=business_reviews$longitude,
  clusterOptions = markerClusterOptions(),
  popup=as.character(business_reviews$address))
map
```



#4.9

#which month has maximum number of review count

```
business_reviewsDateFormatted%>%
  select(month_formatted,review_count.x)%>%
  filter(!month_formatted %in% NA)%>%
  group_by(month_formatted)%>%
  summarise(No_of_reviews = sum(review_count.x))%>%
  arrange(desc(No_of_reviews))%>%
  ggplot(aes(x=month_formatted,y=No_of_reviews))+
  geom_col()
```



#which month has maximum store opened

```
business_reviewsDateFormatted%>%  
  select(month_formatted,is_open)%>%  
  filter(!month_formatted %in% NA)%>%  
  group_by(month_formatted)%>%  
  summarise(total_open = sum(is_open))%>%  
  ggplot(aes(x=month_formatted,y=total_open))+  
  geom_col()
```

