# Stat-295 HW1

Sabahattin Alp Kocabaş, Başak Kabaloğlu, Efe Örencik

2024-04-20

# Question 1

Codes for Linux

- **To work on a file.**

```
#cd C:/Users/kocab/Desktop/Stat_295_HW1
```

- **(i) To read the data.**

```
#wget --content-disposition https://raw.githubusercontent.com/dhavalpotdar/College-Score
card-Data-Analysis/master/MERGED_2017_2018_cleaned.csv
```

- **(i) To change the name of data set.**

```
#mv MERGED_2017_2018_cleaned.csv college_score.csv
```

- **(i) To print the lines 10 through 60.**

```
#head -n 60 college_score.csv | tail -n 51
```

- **(ii) To create sub-sample of data set with appropriate conditions.**

```
#grep "Public" college_score.csv | grep ",Montgomery," > subsample.csv
```

- **(iii) To obtain the frequencies of each cities.**

```
#cut -d ',' -f2 college_score.csv | sort | uniq -c
```

# Outputs for 1

- **(i)The lines 10 through 60.**

```
kocab@AlpKocabas /cygdrive/c/Users/kocab/Desktop/Stat_295_HW1
$ head -n 60 college_score.csv | tail -n 51
100830,Auburn University at Montgomery,Montgomery,AL,1,Graduate's,Bachelor's,Public,No,8709.0,7518.0,0.8253839935327401,490.0,565.0,475.0,545.0,19.0,24.0,4211.0,0.2244,0.5769,0.115
100858,Auburn University,Auburn,AL,1,Graduate's,Bachelor's,Public,No,15479.0,10137.0,0.83930942895086,570.0,650.0,560.0,660.0,24.0,30.0,23391.0,0.0377,0.3198,0.03
100937,Birmingham Southern College,Birmingham,AL,1,Bachelor's,Bachelor's,Private Nonprofit,No,10088.0,7774.0,0.6186010160218 8,520.0,630.0,510.0,630.0,23.0,29.0,1283.0,0.0133,0.4467,0.066
101028,Chattahoochee Valley Community College,Phenix City,AL,1,Associate's,Associate's,Public,No,2035.0,5961.0,,,,,,,1370.0,0.3574,0.2271,0.133
101073,Concordia College Alabama,Selma,AL,1,Bachelor's,Associate's,Private Nonprofit,No,5301.0,5409.0,,,,,,,389.0,0.2288,0.8118,
101116,South University-Montgomery,Montgomery,AL,15,Graduate's,Bachelor's,Private For-Profit,No,16005.0,4830.0,,,,,,369.0,0.6206,0.6726,0.145
101143,Enterprise State Community College,Enterprise,AL,1,Associate's,Associate's,Public,No,2606.0,6337.0,,,,,,,1365.0,0.2402,0.1647,0.209
101161,Coastal Alabama Community College,Bay Minette,AL,3,Associate's,Associate's,Public,No,1949.0,5932.0,,,,,,,4101.0,0.2262,0.4446,0.185
101189,Faulkner University,Montgomery,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,11037.0,5702.0,0.51006423982869,460.0,590.0,460.0,560.0,18.0,22.0,2272.0,0.4188,0.6469,0.122
101240,Gadsden State Community College,Gadsden,AL,1,Associate's,Certificate,Public,No,1789.0,6236.0,,,,,,,4349.0,0.2926,0.0,
101277,New Beginning College of Cosmetology,Albertville,AL,1,Certificate,Certificate,Private For-Profit,No,7119.0,,,,,,,,83.0,0.4217,0.3305,0.047
101286,George C Wallace Community College-Dothan,Dothan,AL,1,Associate's,Associate's,Public,No,2313.0,5866.0,,,,,,,3836.0,0.3053,0.0,
101295,George C Wallace State Community College-Hanceville,Hanceville,AL,1,Associate's,Associate's,Public,No,2962.0,5790.0,,,,,,,4805.0,0.2541,0.3243,0.154
101301,George C Wallace State Community College-Selma,Selma,AL,1,Associate's,Associate's,Public,No,1279.0,4775.0,,,,,,,1217.0,0.2345,0.0,
101365,Herzing University-Birmingham,Birmingham,AL,11,Graduate's,Certificate,Private Nonprofit,No,10525.0,5360.0,0.7940828402366801,,,,,,537.0,0.7598,0.7692,0.136
101435,Huntingdon College,Montgomery,AL,1,Bachelor's,Bachelor's,Private Nonprofit,No,11118.0,6851.0,0.55978784956605,490.0,595.0,500.0,580.0,19.0,24.0,1099.0,0.1833,0.7064,0.10800000000000001
101453,Heritage Christian University,Florence,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,10311.0,4812.0,1.0,,,,19.0,21.0,59.0,0.5085,0.2,0.166
101462,J. F. Drake State Community and Technical College,Huntsville,AL,1,Associate's,Certificate,Public,No,1785.0,5981.0,,,,,,,685.0,0.5,0.0,
101471,J F Ingram State Technical College,Deatsville,AL,1,Certificate,Certificate,Public,No,32.0,5623.0,,,,,,,449.0,0.8285,0.0,
101480,Jacksonville State University,Jacksonville,AL,1,Graduate's,Bachelor's,Public,No,6700.0,6960.0,0.5402164862614 4,420.0,530.0,420.0,520.0,19.0,26.0,6622.0,0.1605,0.6466,0.11800000000000001
101499,Jefferson Davis Community College,Brewton,AL,3,Associate's,Associate's,Public,No,803.0,6173.0,,,,,,,720.0,0.4088,0.0,0.185
101505,Jefferson State Community College,Birmingham,AL,1,Associate's,Associate's,Public,No,3130.0,6451.0,,,,,,,7052.0,0.2778,0.1889,0.13
101514,John C Calhoun State Community College,Tanner,AL,1,Associate's,Associate's,Public,No,3451.0,6479.0,,,,,,,8247.0,0.3238,0.2379,0.17300000000000001
101541,Judson College,Marion,AL,1,Bachelor's,Bachelor's,Private Nonprofit,No,8067.0,5932.0,0.50704225352112,,,,18.0,25.0,315.0,0.2635,0.6626,0.159
101569,Lawson State Community College,Birmingham,AL,1,Associate's,Certificate,Public,No,1775.0,6035.0,,,,,,,3034.0,0.3107,0.0,
101587,University of West Alabama,Livington,AL,1,Graduate's,Bachelor's,Public,No,8192.0,6711.0,0.28207904710341,440.0,520.0,420.0,500.0,18.0,23.0,2106.0,0.1525,0.6461,0.078
101602,Lurleen B Wallace Community College,Andalusia,AL,1,Associate's,Associate's,Public,No,1666.0,6303.0,,,,,,,1354.0,0.1823,0.0,
101648,Marion Military Institute,Marion,AL,1,Associate's,Associate's,Public,No,2364.0,4647.0,0.5586544600938,530.0,630.0,540.0,640.0,19.0,26.0,444.0,0.0045,0.4706,0.107
101675,Miles College,Fairfield,AL,1,Bachelor's,Bachelor's,Private Nonprofit,No,5300.0,4444.0,,,,,,,1647.0,0.1503,0.9253,0.11199999999999999
101693,University of Mobile,Mobile,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,9537.0,4938.0,0.47019423978566,460.0,560.0,450.0,540.0,20.0,25.0,1276.0,0.17600000000000002,0.7616,0.067
101709,University of Montevallo,Montevallo,AL,1,Graduate's,Bachelor's,Public,No,8422.0,7451.0,0.644547325102 88,440.0,620.0,460.0,580.0,20.0,26.0,2329.0,0.0874,0.5791,0.091
101736,Northwest-Shoals Community College,Muscle Shoals,AL,1,Associate's,Associate's,Public,No,2422.0,4895.0,,,,,,,2698.0,0.1959,0.2481,0.21899999999999997
101879,University of North Alabama,Florence,AL,1,Graduate's,Bachelor's,Public,No,6997.0,7567.0,0.7021919870627,515.0,610.0,455.0,580.0,19.0,25.0,5771.0,0.1087,0.7101,0.096
101897,Northeast Alabama Community College,Rainsville,AL,1,Associate's,Certificate,Public,No,2032.0,4898.0,,,,,,,1983.0,0.2036,0.0,0.151
101912,Oakwood University,Huntsville,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,12014.0,4902.0,0.4706833805095,460.0,570.0,430.0,530.0,17.0,22.0,1623.0,0.1497,0.7273,0.13699999999999998
101949,Alabama Southern Community College,Monroeville,AL,3,Associate's,Associate's,Public,No,1467.0,6213.0,,,,,,,765.0,0.1069,0.0,0.185
101994,Reid State Technical College,Evergreen,AL,1,Associate's,Certificate,Public,No,1856.0,6177.0,,,,,,,285.0,0.3342,0.0,
102030,Bishop State Community College,Mobile,AL,1,Associate's,Certificate,Public,No,1912.0,6258.0,,,,,,,2623.0,0.317,0.0,
102049,Samford University,Birmingham,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,18867.0,8580.0,0.831932773109 24,550.0,650.0,530.0,630.0,23.0,29.0,3364.0,0.0415,0.3245,0.019
102058,Selma University,Selma,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,13530.0,2900.0,,,,,,,311.0,0.5595,0.9295,0.31
102067,Shelton State Community College,Tuscaloosa,AL,1,Associate's,Associate's,Public,No,2496.0,4690.0,,,,,,,4230.0,0.2123,0.0,
102076,Snead State Community College,Boaz,AL,1,Associate's,Associate's,Public,No,2530.0,6934.0,,,,,,,1891.0,0.2198,0.24600000000000002,0.258
102094,University of South Alabama,Mobile,AL,1,Graduate's,Bachelor's,Public,No,9198.0,7203.0,0.8207867073584201,510.0,610.0,500.0,590.0,20.0,26.0,10712.0,0.159,0.5335,0.084
102234,Spring Hill College,Mobile,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,12554.0,6643.0,0.4568117977528,515.0,600.0,510.0,585.0,21.0,26.0,1375.0,0.0145,0.7581,0.062
102270,Stillman College,Tuscaloosa,AL,1,Bachelor's,Bachelor's,Private Nonprofit,No,8697.0,4641.0,0.48116338083154,,,,,,633.0,0.1167,0.8631,0.22899999999999998
102298,Talladega College,Talladega,AL,1,Bachelor's,Bachelor's,Private Nonprofit,No,8182.0,5471.0,,,,,,,782.0,0.1036,0.9704,0.32
102313,H Councill Trenholm State Community College,Montgomery,AL,1,Associate's,Certificate,Public,No,2590.0,5811.0,,,,,,,1610.0,0.2997,0.0,
102368,Troy University,Troy,AL,4,Graduate's,Bachelor's,Public,No,9357.0,6444.0,0.881227093796 66,480.0,600.0,465.0,605.0,18.0,25.0,13634.0,0.3411,0.6182,0.10400000000000001
102377,Tuskegee University,Tuskegee,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,15268.0,4753.0,0.5031934306569 3,440.0,540.0,420.0,520.0,18.0,24.0,2833.0,0.0999,0.652,0.115
102395,United States Sports Academy,Daphne,AL,1,Graduate's,Bachelor's,Private Nonprofit,No,1598.0,5732.0,,,,,,,88.0,0.8058,0.4267,0.078
102429,Bevill State Community College,Jasper,AL,1,Associate's,Associate's,Public,No,2122.0,6476.0,,,,,,,3029.0,0.2283,0.0,
```

- **(iii)Head of obtained frequencies of each cities.**

```
kocab@AlpKocabas /cygdrive/c/Users/kocab/Desktop/Stat_295_HW1
$ cut -d ',' -f2 college_score.csv | sort | uniq -c | head
      1 A T Still University of Health Sciences
      1 ABCO Technology
      1 AI Miami International University of Art and Design
      1 AMG School of Licensed Practical Nursing
      1 AOMA Graduate School of Integrative Medicine
      1 ASA College
      1 ASI Career Institute
      1 ASM Beauty World Academy
      1 ATA Career Education
      1 ATA College
```

# Question 2

- **Loading necessary packages.**

```
library(dplyr)
library(tidyverse)
```

- **(i) Reading the data set, assigning name as chocolate. Printing first 6 rows.**

```
chocolate <- read.csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2022/2022-01-18/chocolate.csv")
head(chocolate)
```

```
##    ref company_manufacturer company_location review_date country_of_bean_origin
## 1 2454                 5150           U.S.A.        2019               Tanzania
## 2 2458                 5150           U.S.A.        2019     Dominican Republic
## 3 2454                 5150           U.S.A.        2019             Madagascar
## 4 2542                 5150           U.S.A.        2021                   Fiji
## 5 2546                 5150           U.S.A.        2021              Venezuela
## 6 2546                 5150           U.S.A.        2021                 Uganda
##   specific_bean_origin_or_bar_name cocoa_percent ingredients
## 1            Kokoa Kamili, batch 1           76%    3- B,S,C
## 2                  Zorzal, batch 1           76%    3- B,S,C
## 3           Bejofo Estate, batch 1           76%    3- B,S,C
## 4            Matasawalevu, batch 1           68%    3- B,S,C
## 5            Sur del Lago, batch 1           72%    3- B,S,C
## 6         Semuliki Forest, batch 1           80%    3- B,S,C
##      most_memorable_characteristics rating
## 1         rich cocoa, fatty, bready   3.25
## 2            cocoa, vegetal, savory   3.50
## 3      cocoa, blackberry, full body   3.75
## 4               chewy, off, rubbery   3.00
## 5 fatty, earthy, moss, nutty,chalky   3.00
## 6 mildly bitter, basic cocoa, fatty   3.25
```

- **(i) Examining the structure and comments.**

```
str(chocolate)
```

```
## 'data.frame':    2530 obs. of  10 variables:
##  $ ref                            : int  2454 2458 2454 2542 2546 2546 2542 797 797
1011 ...
##  $ company_manufacturer           : chr  "5150" "5150" "5150" "5150" ...
##  $ company_location               : chr  "U.S.A." "U.S.A." "U.S.A." "U.S.A." ...
##  $ review_date                    : int  2019 2019 2019 2021 2021 2021 2021 2012 201
2 2013 ...
##  $ country_of_bean_origin         : chr  "Tanzania" "Dominican Republic" "Madagasca
r" "Fiji" ...
##  $ specific_bean_origin_or_bar_name: chr  "Kokoa Kamili, batch 1" "Zorzal, batch 1"
"Bejofo Estate, batch 1" "Matasawalevu, batch 1" ...
##  $ cocoa_percent                  : chr  "76%" "76%" "76%" "68%" ...
##  $ ingredients                    : chr  "3- B,S,C" "3- B,S,C" "3- B,S,C" "3- B,S,C"
...
##  $ most_memorable_characteristics  : chr  "rich cocoa, fatty, bready" "cocoa, vegeta
l, savory" "cocoa, blackberry, full body" "chewy, off, rubbery" ...
##  $ rating                         : num  3.25 3.5 3.75 3 3 3.25 3.5 3.5 3.75 2.75
...
```

The data set likely contains information about chocolate products, including details like manufacturer, origin, ingredients, characteristics, and ratings.

- **(ii) Converting all the characters into factors.**

```
chocolate <- chocolate %>%
  mutate_if(is.character, factor)
```

- **(iii) Obtaining some statistics with respect to different company locations. Printing the 10 observations. And comments.**

```
chocolate %>%
  group_by(company_location) %>%
  summarise(mean_rating = mean(rating),
            sd_rating = sd(rating),
            median_rating = median(rating),
            range_rating = diff(range(rating))) %>%
  head(10)
```

```
## # A tibble: 10 × 5
##    company_location mean_rating sd_rating median_rating range_rating
##    <fct>                  <dbl>     <dbl>         <dbl>        <dbl>
##  1 Amsterdam               3.31     0.264          3.25         0.75
##  2 Argentina               3.31     0.349          3.5          1
##  3 Australia               3.36     0.409          3.5          1.5
##  4 Austria                 3.26     0.325          3.25         1
##  5 Belgium                 3.10     0.661          3            3
##  6 Bolivia                 3.25     0.707          3.25         1
##  7 Brazil                  3.28     0.356          3.25         1.5
##  8 Canada                  3.30     0.416          3.25         2
##  9 Chile                   3.75     0              3.75         0
## 10 Colombia                3.20     0.425          3.25         1.75
```

If we consider 10 rows of statistics, mean ratings vary 3.1 to 3.75 , standart deviatons is also vary and there are differences among locations. Median ratings are between 3.00 to 3.75 for this 10 examination. And there are range ratings that varies from location to location between 0 to 3.

- **(iv) Finding the chocolates that its review date is equal to 2020 and country of bean origin is equal to Colombia.**

```
chocolate %>%
  filter(review_date == 2020 & country_of_bean_origin == "Colombia")
```

```
##    ref company_manufacturer company_location review_date country_of_bean_origin
## 1 2466        Crow and Moss           U.S.A.        2020               Colombia
## 2 2534              El Buen           U.S.A.        2020               Colombia
## 3 2482               Finnia           Canada        2020               Colombia
## 4 2478           Kin + Pod            Canada        2020               Colombia
## 5 2498             Odyssey            U.S.A.        2020               Colombia
##   specific_bean_origin_or_bar_name cocoa_percent ingredients
## 1                Aruaca, batch 39           70%      2- B,S
## 2                          Tumaco           70%      2- B,S
## 3            Chigorodo, batch 001           70%    3- B,S,C
## 4               Tumaco, batch 113           70%    3- B,S,C
## 5                         Arhuaca           70%      2- B,S
##   most_memorable_characteristics rating
## 1                 nutty, citrus   3.25
## 2   cocoa, spice, alcohol, dirty   3.00
## 3          nutty, melon, vinegar   2.75
## 4     poor finish, cocoa, grassy   3.25
## 5        walnut, tobacco, grassy   3.25
```

- **(v) Taking the mean of chocolate rating and cocoa percent according to the company location.**

```
chocolate %>%
  group_by(company_location) %>%
  summarise(mean_rating = mean(rating),
            mean_cocoa_percent = mean(as.numeric(factor(cocoa_percent))))
```

```
## # A tibble: 67 × 3
##    company_location mean_rating mean_cocoa_percent
##    <fct>                  <dbl>              <dbl>
##  1 Amsterdam               3.31               3
##  2 Argentina               3.31               1.67
##  3 Australia               3.36               5.89
##  4 Austria                 3.26               6.47
##  5 Belgium                 3.10               7.95
##  6 Bolivia                 3.25               1.5
##  7 Brazil                  3.28               6
##  8 Canada                  3.30               9.84
##  9 Chile                   3.75               1
## 10 Colombia                3.20               6.90
## # ℹ 57 more rows
```

- **(vi) Selecting company manufacturer, company location and country of bean origin shortly. Printing the first 10 rows of the data frame.**

```
chocolate %>%
 select(starts_with("c")) %>%
  head(10)
```

```
##    company_manufacturer company_location country_of_bean_origin cocoa_percent
## 1                  5150           U.S.A.               Tanzania           76%
## 2                  5150           U.S.A.     Dominican Republic           76%
## 3                  5150           U.S.A.             Madagascar           76%
## 4                  5150           U.S.A.                   Fiji           68%
## 5                  5150           U.S.A.              Venezuela           72%
## 6                  5150           U.S.A.                 Uganda           80%
## 7                  5150           U.S.A.                  India           68%
## 8              A. Morin           France                Bolivia           70%
## 9              A. Morin           France                   Peru           63%
## 10             A. Morin           France                 Panama           70%
```

- **(vii) Filtering that company location in Switzerland whose rating between 3.25 and 3.5. Taking the five observations.**

```
chocolate %>%
  filter(company_location == "Switzerland" & rating >= 3.25 & rating <= 3.5) %>%
  head(5)
```

```
##   ref company_manufacturer company_location review_date country_of_bean_origin
## 1 508   Beschle (Felchlin)      Switzerland        2010              Venezuela
## 2 508   Beschle (Felchlin)      Switzerland        2010              Venezuela
## 3 508   Beschle (Felchlin)      Switzerland        2010              Indonesia
## 4 508   Beschle (Felchlin)      Switzerland        2010              Venezuela
## 5 636   Beschle (Felchlin)      Switzerland        2011              Venezuela
##       specific_bean_origin_or_bar_name cocoa_percent ingredients
## 1   Carenero S., Barlovento, Grand Cru           70%    3- B,S,C
## 2 Porcelana, Premier Cru, Quizas No. 1           74%    3- B,S,C
## 3                      Java, Grand Cru           64%    3- B,S,C
## 4   Ocumare, Premier Cru, Quizas No. 2           72%    3- B,S,C
## 5 Indigena Amazonia, Grand Cru, Quizas           72%    3- B,S,C
##    most_memorable_characteristics rating
## 1       creamy, macadamia, pepper   3.25
## 2 nutty, light toffee, mild musty   3.25
## 3        ham-like, smokey, banana   3.50
## 4        dark cocoa, spicy pepper   3.50
## 5            creamy, banana, rich   3.50
```

- **(viii) Mean of the rating column for each company locations that ordered by descending.**

```
chocolate %>%
  group_by(company_location) %>%
  summarise(mean_rating = mean(rating)) %>%
  arrange(desc(mean_rating))
```

```
## # A tibble: 67 × 2
##    company_location mean_rating
##    <fct>                  <dbl>
##  1 Chile                   3.75
##  2 U.A.E.                  3.4
##  3 Poland                  3.38
##  4 Vietnam                 3.36
##  5 Australia               3.36
##  6 Guatemala               3.35
##  7 Denmark                 3.34
##  8 Norway                  3.33
##  9 Switzerland             3.32
## 10 Amsterdam               3.31
## # ℹ 57 more rows
```

- **(ix) Counting the observations are assigned Bonnat for each country of bean origin.**

```
chocolate %>%
  filter(company_manufacturer == "Bonnat") %>%
  count(country_of_bean_origin, sort = TRUE)
```

```
##    country_of_bean_origin n
## 1                   Blend 4
## 2                  Brazil 4
## 3                    Peru 4
## 4               Venezuela 4
## 5                  Mexico 3
## 6              Madagascar 2
## 7                    Cuba 1
## 8      Dominican Republic 1
## 9                 Ecuador 1
## 10                  Gabon 1
## 11                  Haiti 1
## 12            Ivory Coast 1
## 13                Jamaica 1
## 14              Nicaragua 1
## 15              Sri Lanka 1
```

- **(x) Creating a new column called Rating Percentage, which is percentage version of the rating column & rating the chocolates.**

```
chocolate <- chocolate %>%
  mutate(Rating_Percentage = rating / 4 * 100,
         Class = case_when(
           Rating_Percentage < 25 ~ "Low",
           Rating_Percentage < 50  ~ "Medium",
           Rating_Percentage <= 87.5 ~ "Tasty",
           Rating_Percentage >87.5 ~ "Excellent"
         ))
```

# Question 3

- **Calling necessary libraries.**

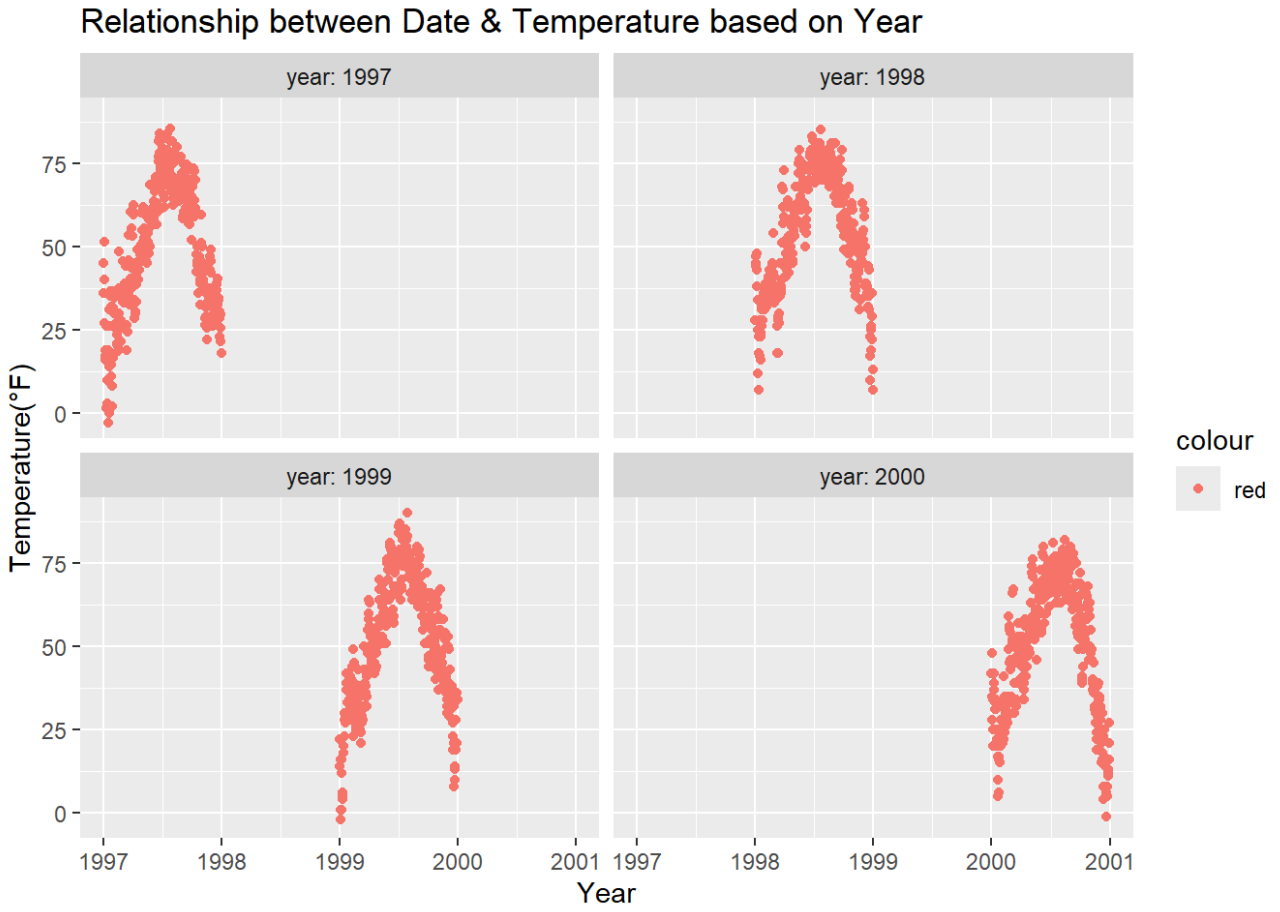
```
library(ggplot2)
library(dplyr)
library(tidyverse)
```

- **(i) Reading data and named it as nmmaps.**

```
nmmaps<-read.csv("https://www.cedricscherer.com/data/chicago-nmmaps-custom.csv")
```

- **(i)Examining the relationship between date and temp based on the year.**

```
ggplot(nmmaps,aes(as.Date(date),temp,color="red"))+
  geom_point()+
  labs(x = "Year", y = "Temperature(°F)",
       title = "Relationship between Date & Temperature based on Year")+
  facet_wrap(~year, labeller = label_both)
```




- **Interpretation for plot 1.**

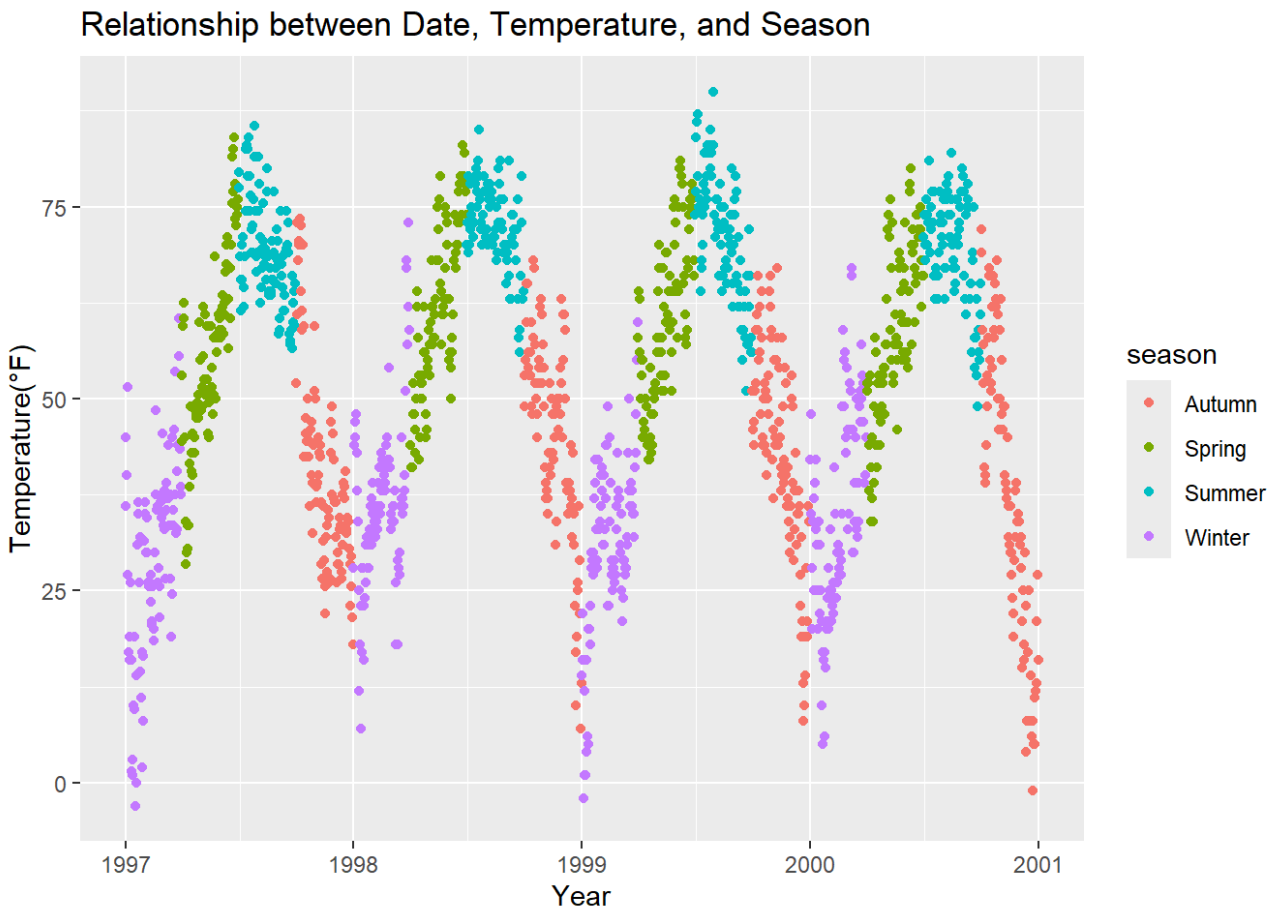
  Although there are no major changes in general, temperatures in the last months of the year have started to decrease over the years. The lowest temperature was reached in early 1997, and the highest temperature was reached in mid-1999. In general, the temperature first increased and then decreased within a year.

- **(ii) Factoring season variables.**

```
nmmaps$season<-factor(nmmaps$season,labels = c("Autumn", "Spring", "Summer", "Winter"))
```

- **(ii) Examining the relationship between date, temp and season.**

```
ggplot(nmmaps,aes(as.Date(date),temp,color=season))+
  geom_point()+
  labs(x = "Year", y = "Temperature(°F)", title = "Relationship between Date, Temperatur
e, and Season" )
```

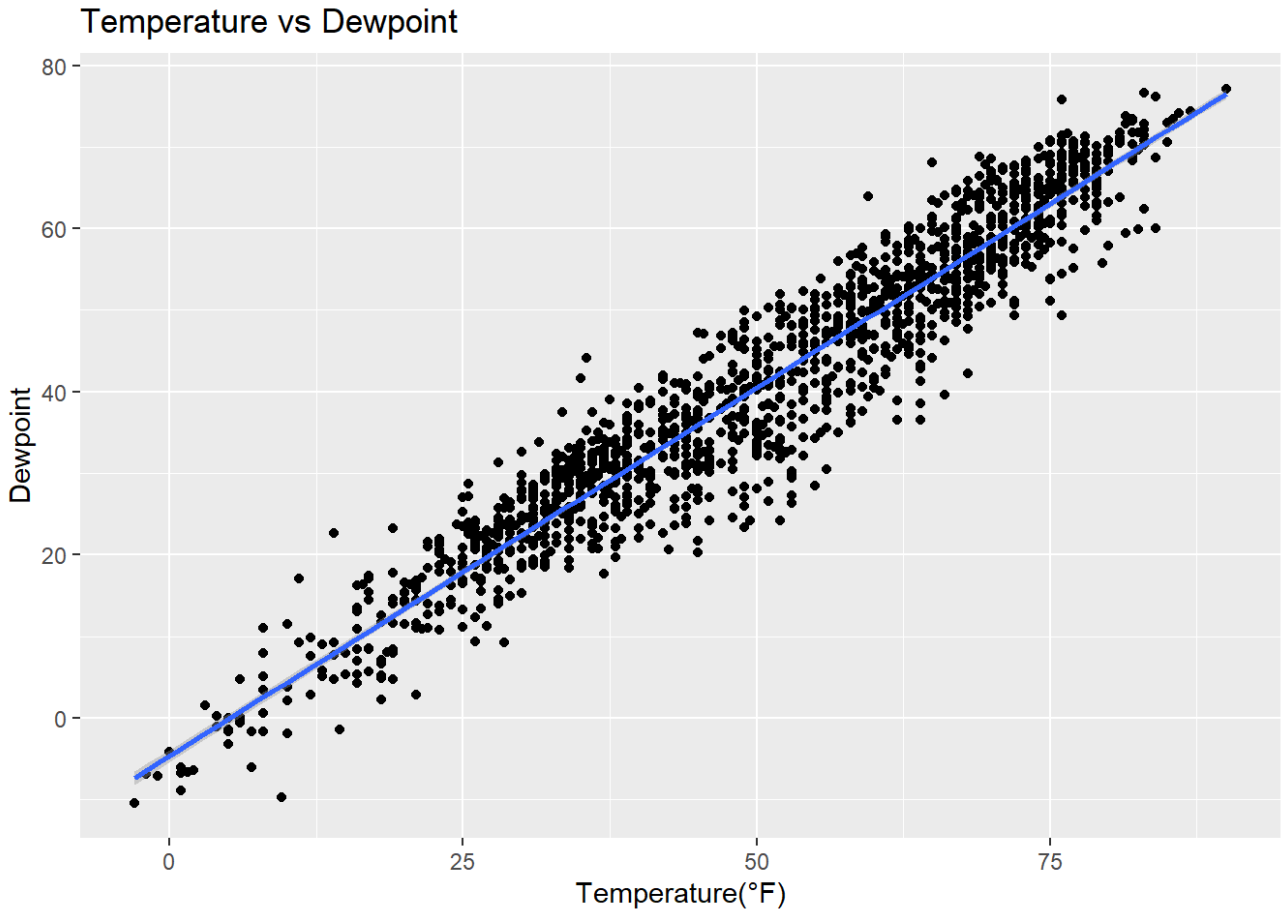



- **Interpretation for plot 2.**

  The temperature change generally followed a regular pattern. Temperature change was more vary in winter and autumn. As we see in the first graph, the temperature reached its lowest value at the beginning of 1997 and its peak in the middle of 1999.

- **(iii) The relationship between temperature and dewpoint, and the correlation.**

```
ggplot(nmmaps,aes(temp,dewpoint))+
  geom_point()+
  labs(x = "Temperature(°F)",
       y = "Dewpoint",
       title = "Temperature vs Dewpoint") +
  geom_smooth(method="lm")
```




- **Interpretation for plot 3.**

  In this graph we observe the relationship between temperature and dewpoint. There is a strong linear relationship between temperature and dewpoint with small deviations. Simply we can conclude that the dewpoint increases as the temperature increases.

# Question 4

- In this part, we tried to fetch earthquake data from the Terremoti website, analyzed the data, categorized the earthquakes into different classes based on their magnitudes, and visualized it on a Leaflet map.
- **Calling necessary libraries.**

```
libraries <- c("ggplot2","sf","rworldmap","tidyverse","magrittr",
               "leaflet", "dplyr", "rvest", "xml2","rvest",
               "maps","mapdata","RgoogleMaps","lubridate","rnaturalearth","dplyr","rnatu
ralearthdata","RColorBrewer","httr")
lapply(libraries, require, character.only = TRUE)
```

- **Scraping data from web.**

```
url <-"https://terremoti.ingv.it/en/events?starttime=2024-04-12%2B00%253A00%253A00&endti
me=2024-04-19%2B23%253A59%253A59&last_nd=7&minmag=2&maxmag=10&mindepth=-10&maxdepth=1000
&minlat=-90&maxlat=90&minlon=-180&maxlon=180&minversion=100&limit=30&orderby=mag-asc&lat
=0&lon=0&maxradiuskm=-1&wheretype=area&box_search=Mondo&timezone=UTC&page=2"
res <- GET(url)
html_con <- content(res, "text")
```

- **Reading the html content.**

```
html_data <- read_html(html_con)
```

```
tables <- html_data %>%
  html_nodes("table") %>%
  html_table()
```

```
earthquake <- as.data.frame(tables)
str(earthquake)
```

```
## 'data.frame':    30 obs. of  6 variables:
##  $ Origin.time..UTC.: chr  "2024-04-13 14:59:00" "2024-04-16 10:17:55" "2024-04-16 0
2:11:32" "2024-04-14 04:20:06" ...
##  $ Magnitude        : chr  "ML 2.3" "Md 2.3" "Md 2.3" "ML 2.4" ...
##  $ Region           : chr  "3 km SW Arpino (FR)" "Campi Flegrei" "Campi Flegrei" "6 k
m SE Mistretta (ME)" ...
##  $ Depth            : int  9 3 3 4 126 15 106 4 1 11 ...
##  $ Latitude         : num  41.6 40.8 40.8 37.9 38.5 ...
##  $ Longitude        : num  13.6 14.1 14.1 14.4 15.6 ...
```

- **Removing non-numeric characters from the Magnitude column and converting it to numeric.**

```
earthquake$Magnitude <- as.numeric(gsub("[^0-9.]", "", earthquake$Magnitude))
```

- **Adding a new column Class based on earthquake magnitudes, categorizing them as "Minor","Light",and"Major".**

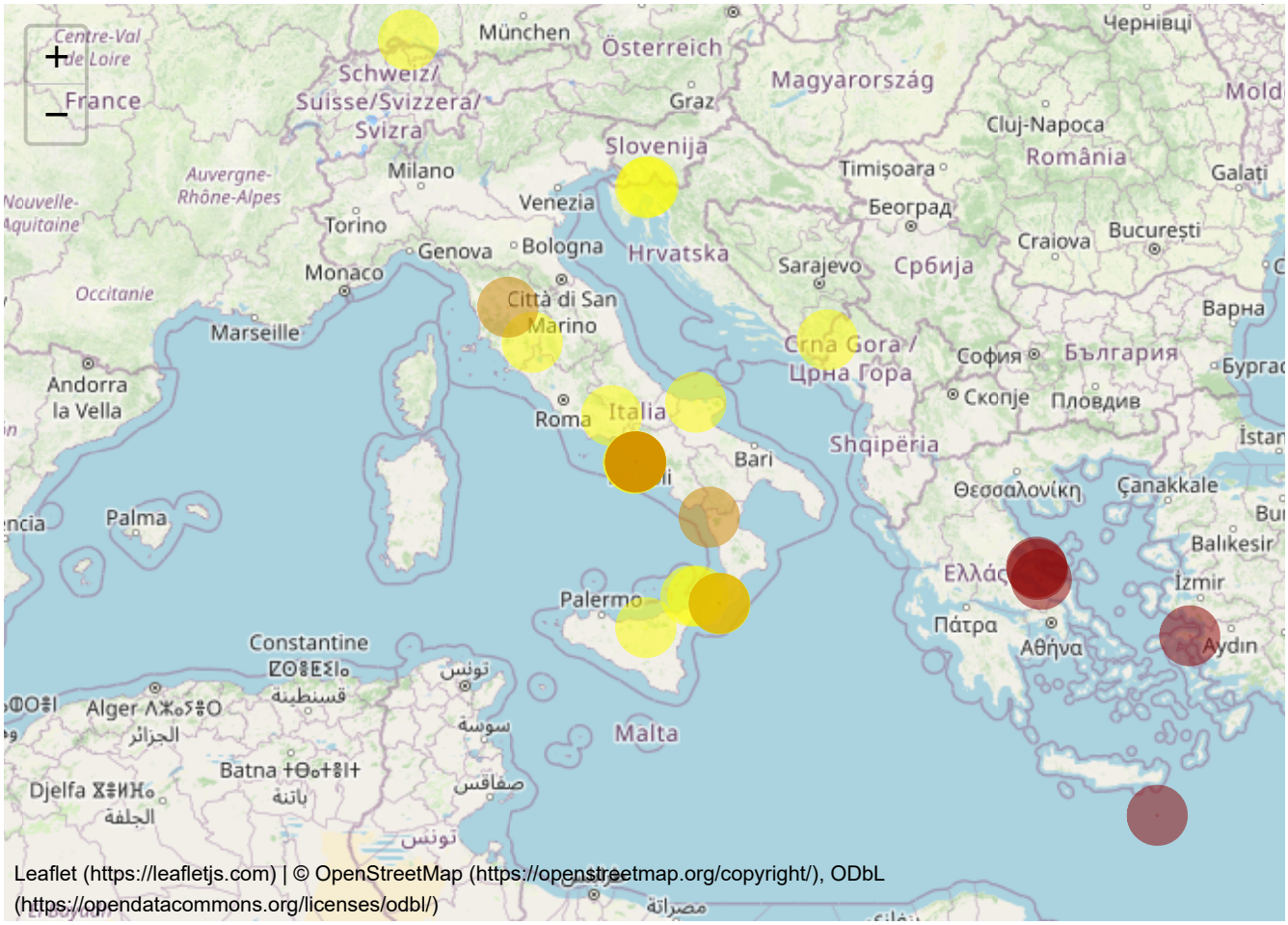
```
earthquake <- earthquake %>%
  mutate(Class = ifelse(Magnitude < 3, "Minor",
                        ifelse(Magnitude<=3.7, "Light","Major")))
```

- **Defining colors for different earthquake classes.**

```
colors <- c("yellow","orange3", "darkred")
color_vector <- colorFactor(colors, levels = c("Minor","Light","Major"))
```

- **Creating Leaflet map. (Please click the circles to open popups.)**

```
leaflet() %>%
  addTiles() %>%
  addCircles(data = earthquake,
             ~Longitude, ~Latitude,
             weight = 30,
             radius = 100,
             popup = paste0(
               "<b>Date: </b>", earthquake$Origin.time..UTC.,
               "<br>",
               "<b>Place: </b>", earthquake$Region,
               "<br>",
               "<b>Depth in km: </b>", earthquake$Depth,
               "<br>",
               "<b>Magnitude: </b>", earthquake$Magnitude),
             label = ~Region,
             color = ~color_vector(Class)) %>%
  setView(lng = median(earthquake$Longitude),
          lat = median(earthquake$Latitude),
          zoom = 5)
```




**End Of The Work.**

Note: Dear teacher, even though we used resources on the internet when we got stuck while answering the questions, our goal for each question was to learn, and it was a very educational assignment for us. Thank you, have a good day.