Zeru-Zhou-project12

December 1, 2021

1 Project 12 – Zeru Zhou

TA Help: NA

Collaboration: NA

- Get help from Dr. Ward's video

1.1 Question 1

```
[2]: library(data.table)
```

```
[3]: liquor <- fread("/depot/datamine/data/iowa_liquor_sales/clean_sample.csv")
```

```
[4]: dim(liquor)
```

1. 5999999 2. 24

```
[5]: liquor$Date[1] - liquor$Date[2]
```

```
Error in liquor$Date[1] - liquor$Date[2]: non-numeric argument to binary operator
Traceback:
```

```
[4]: liquor$date <- as.Date(liquor$Date, format="%m/%d/%Y")
```

```
[11]: class(liquor$date)
```

'Date'

```
[9]: head(liquor$date)
```

1. 2015-11-20 2. 2015-11-21 3. 2015-11-16 4. 2015-11-04 5. 2015-11-17 6. 2015-11-11

```
[10]: liquor$date[1] - liquor$date[2]
```

```
Time difference of -1 days
```

We inserted a new column that is in the "Date" class.

1.2 Question 2

```
[5]: liquor$year <- format(liquor$date, "%Y")
```

```
[6]: liquor$month <- format(liquor$date, "%m")
```

```
[6]: table(liquor$month,liquor$year)
```

```
      2012   2013   2014   2015
  01 108290 119072 115816 116038
  02 114902 110401 109683 116015
  03 117510 113427 117634 130953
  04 120966 130008 129591 132087
  05 140596 130791 120785 124099
  06 123756 115001 135112 150338
  07 132524 142773 137450 133481
  08 140060 124087 126628 130932
  09 109169 121248 125338 138860
  10 144372 138849 132034 127839
  11 123508 109391 114983 128989
  12 149395 155762 169456      0
```

2012, 2013, 2014, and 2015 are covered in this dataset. All month are covered except for Dec, 2015.

1.3 Question 3

```
[15]: liquor <- fread("/depot/datamine/data/iowa_liquor_sales/clean_sample.csv")
```

```
[16]: library(lubridate)
```

```
Attaching package: 'lubridate'


The following objects are masked from 'package:data.table':

    hour, isoweek, mday, minute, month, quarter, second, wday, week,
    yday, year


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

    date, intersect, setdiff, union
```

```
[17]: liquor$date_b <- mdy(liquor$Date)
```

```
[18]: liquor$month_b <- month(mdy(liquor$Date))
```

```
[19]: liquor$year_b <- year(mdy(liquor$Date))
```

```
[20]: table(liquor$month_b,liquor$year_b)
```

```
        2012    2013    2014    2015
  1   108290  119072  115816  116038
  2   114902  110401  109683  116015
  3   117510  113427  117634  130953
  4   120966  130008  129591  132087
  5   140596  130791  120785  124099
  6   123756  115001  135112  150338
  7   132524  142773  137450  133481
  8   140060  124087  126628  130932
  9   109169  121248  125338  138860
 10   144372  138849  132034  127839
 11   123508  109391  114983  128989
 12   149395  155762  169456       0
```

```
[21]: head(liquor)
```

A data.table: 6 x 27

Invoice/Item Number	Date	Store Number	Store Name
<chr>	<chr>	<int>	<chr>
S29198800001	11/20/2015	2191	Keokuk Spirits
S29195400002	11/21/2015	2205	Ding's Honk And Holler
S29050300001	11/16/2015	3549	Quicker Liquor Store
S28867700001	11/04/2015	2513	Hy-Vee Food Store #2 / Iowa City
S29050800001	11/17/2015	3942	Twin Town Liquor
S28869200001	11/11/2015	3650	Spirits, Stogies and Stuff

I got the same result as previous questions. I prefer lubridate package because it is always good to simplify the code.
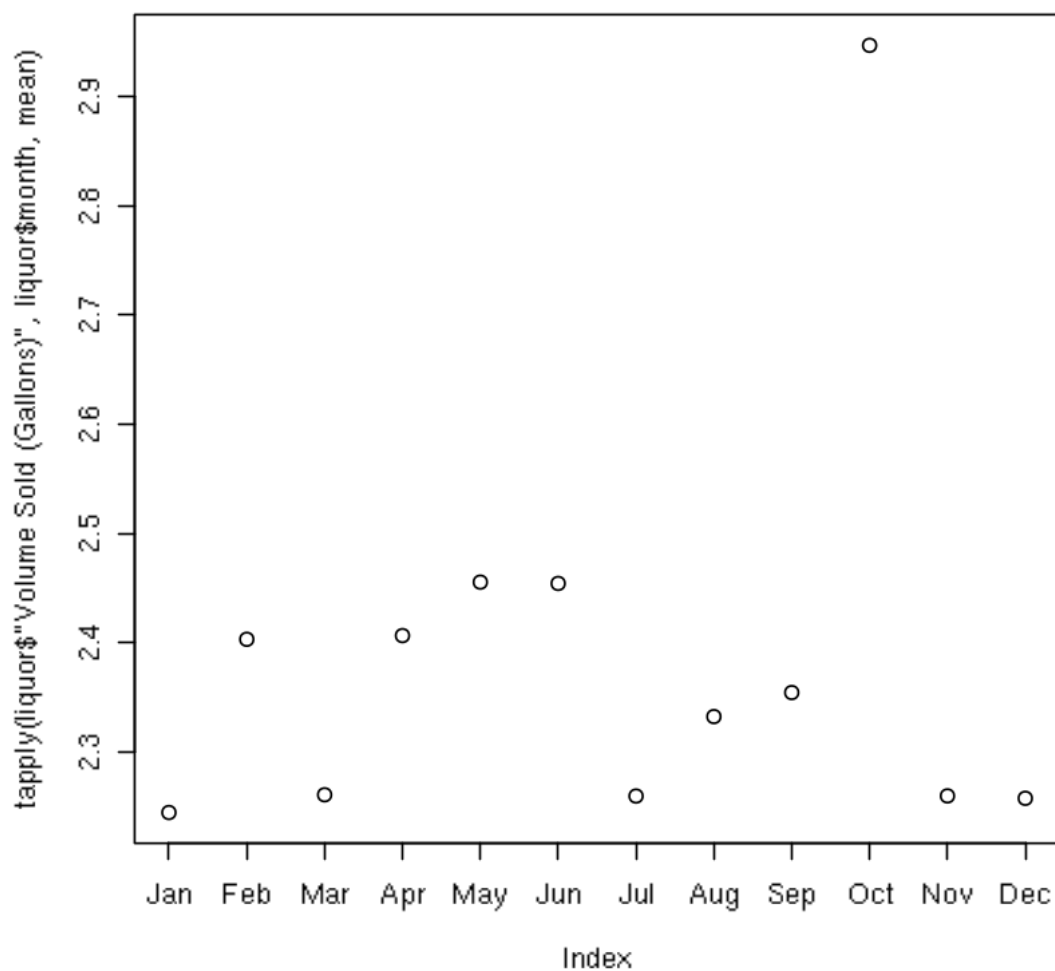
1.4   Question 4

```
[7]: tapply(liquor$'Volume Sold (Gallons)', liquor$month, mean)
```

01 2.24426820929584 **02** 2.40295622404385 **03** 2.26050548043476 **04** 2.40633398484742 **05** 2.45531439883317 **06** 2.45412745346781 **07** 2.25934767166824 **08** 2.33215458101961 **09** 2.35416469375171 **10** 2.94707691486188 **11** 2.25948166275576 **12** 2.25729010794057

```
[23]: plot(tapply(liquor$'Volume Sold (Gallons)', liquor$month, mean), xaxt="n")
      axis(side=1, at=1:12, labels=month.abb)
```

As we can see, January has the lowest average sold per month. This is a little bit surprising because the temperature in Jan is extremely cold so liquor should be in great need.

1.5 Question 5

```
[8]: myDF <- tapply(liquor$'Volume Sold (Gallons)', list(liquor$month, liquor$year),␣
     ↪mean)
```

```
[9]: head(myDF)
```

A matrix: 6 x 4 of type dbl		2012	2013	2014	2015
	01	2.164335	2.152848	2.456344	2.201005
	02	2.348675	2.319316	2.447075	2.494598
	03	2.240106	2.206151	2.278171	2.310022
	04	2.390600	2.451803	2.445837	2.337234
	05	2.347184	2.507527	2.468052	2.510394
	06	2.540984	2.396912	2.459515	2.421554

```
[19]: plot(myDF[, 1], xaxt="n", col="red", type="b", xlab="Month", ylab="Sale")
      axis(side=1, at=1:12, labels=month.abb)
      lines(myDF[, 2], col="blue", type="b")
      lines(myDF[, 3], col="green", type="b")
      lines(myDF[, 4], col="black", type="b")
      legend("top", legend=c("2012","2013","2014","2015"),␣
       ↪col=c("red","blue","green","black"), lty=1 )
```
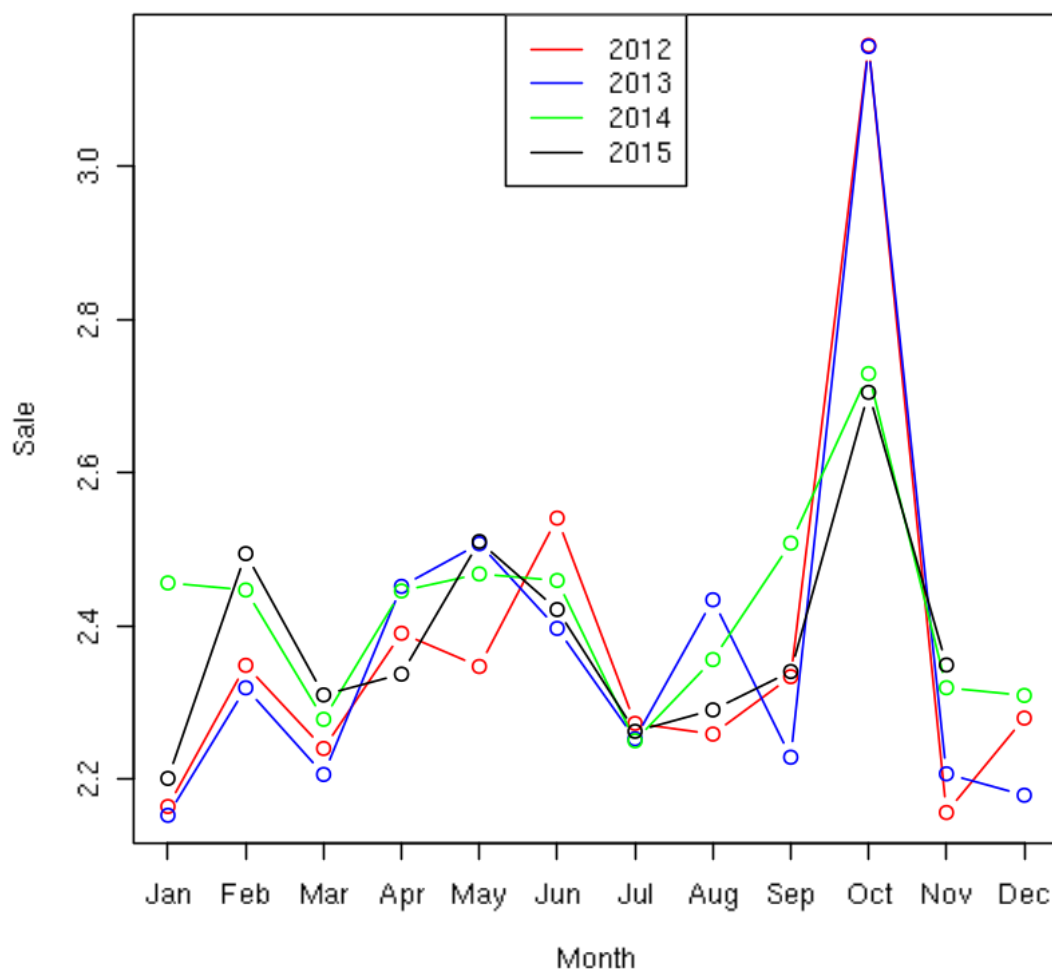
As we can see, For different years, the trend of sale of liquor is different with respect to months. Especially in October, the data from 2012 and 2013 are very close, while the data from 2014 and 2015 are very close. Maybe 2014 is a turning point.

1.6 Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

> As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together – We are Purdue.