ISE 599 HW 1

Shaoqian Chen

1/28/2020

```
#library(readr)
library(MASS)
library(help=MASS)
library(ISLR)
library(lubridate)
```

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

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(rlang)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
data1 = read.csv("crime.csv")
```

```
#data2 is the updated dataset without column "Precinct", "Sector", "Beat"
names(data1)<-c("Report.Number","ODate","OTime","RDate","RTime","category","Description"
,"Precinct","Sector","Beat","Neighborhood")
keep <- c("Report.Number","ODate","OTime","RDate","RTime","category","Description","Neig
hborhood")
data2 <- data1[keep]
```

```
summary(data2)
```

```
##  Report.Number              ODate            OTime
##  Min.   :2.008e+08   07/01/2017:   199   Min.   :   0
##  1st Qu.:2.008e+13   05/26/2017:   192   1st Qu.: 900
##  Median :2.012e+13   01/20/2016:   186   Median :1500
##  Mean   :1.633e+13   12/01/2015:   184   Mean   :1362
##  3rd Qu.:2.015e+13   11/25/2015:   182   3rd Qu.:1923
##  Max.   :2.011e+15   09/23/2014:   181   Max.   :2359
##                      (Other)   :480252   NA's   :2
##         RDate            RTime                        category
##  12/31/2008:   238   Min.   :   0   CAR PROWL           :137766
##  03/31/2014:   196   1st Qu.: 950   THEFT-ALL OTHER     : 49624
##  06/18/2018:   195   Median :1408   THEFT-SHOPLIFT      : 44768
##  05/12/2014:   193   Mean   :1355   BURGLARY-RESIDENTIAL: 43908
##  07/05/2016:   193   3rd Qu.:1818   MOTOR VEHICLE THEFT : 40362
##  01/21/2014:   192   Max.   :2359   BURGLARY-COMMERCIAL : 21274
##  (Other)   :480169   NA's   :2      (Other)             :143674
##              Description                 Neighborhood
##  THEFT-CARPROWL    :122025   DOWNTOWN COMMERCIAL: 45127
##  THEFT-SHOPLIFT    : 44768   NORTHGATE          : 28480
##  THEFT-OTH         : 43164   CAPITOL HILL       : 28296
##  VEH-THEFT-AUTO    : 35202   QUEEN ANNE         : 25172
##  BURGLARY-FORCE-RES: 26417   SLU/CASCADE        : 21630
##  THEFT-BUILDING    : 19718   UNIVERSITY         : 19167
##  (Other)           :190082   (Other)            :313504
```

Question 1

Find the number of neighborhoods and crime categories in the dataset

```
#Return the number of neighborhoods
length(table(data2$Neighborhood))
```

```
## [1] 59
```

```
#Return the numnber of crime categories
length(table(data2$category))
```

```
## [1] 31
```

59 Neighborhood;

31 Crime Categories

Question 2

Report the total number of crimes in each neighborhood. What neighborhood is most dangerous?

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:rlang':
##
##     set_names
```

```
dataQ2<-data2[c("Report.Number","Neighborhood")]
table(data2$Neighborhood)[table(data2$Neighborhood) %>% which.max]
```

```
## DOWNTOWN COMMERCIAL
##               45127
```

The most dangerous neighborhood is DOWNTOWN COMMERCIAL, it has 45127 crimes

Question 3

```
#Question 1 using pipe
library(magrittr)
data2$Neighborhood %>%
  unique %>%
  length
```

```
## [1] 59
```

```
data2$category %>%
  unique %>%
  length
```

```
## [1] 31
```

Same result as the previous questions.

```
dataQ3<-data2 %>%
  group_by(Neighborhood = tolower(Neighborhood)) %>%
  summarise(count = n())
dataQ3
```

Neighborhood <chr>	count <int>
alaska junction	6378
alki	2335
ballard north	10155
ballard south	14031
belltown	14437
bitterlake	9227
brighton/dunlap	6608
capitol hill	28296
central area/squire park	11361
chinatown/international district	13627

1-10 of 59 rows Previous **1** 2 3 4 5 6 Next

According to the table above Downtown Commercial(45127 Crimes) is more dangerous than other area

Question 4

```
QueenAnne <- data2$Neighborhood=="QUEEN ANNE"
dataQ4 <- data2[QueenAnne,]
max(summary(dataQ4$category))
```

```
## [1] 10115
```

```
summary(dataQ4$category)
```

```
##                                                 AGGRAVATED ASSAULT
##                                 17                                402
##                AGGRAVATED ASSAULT-DV                              ARSON
##                                189                                 45
##                  BURGLARY-COMMERCIAL  BURGLARY-COMMERCIAL-SECURE PARKING
##                               1359                                125
##                 BURGLARY-RESIDENTIAL BURGLARY-RESIDENTIAL-SECURE PARKING
##                               2193                                991
##                          CAR PROWL                 DISORDERLY CONDUCT
##                              10115                                  7
##                                DUI           FAMILY OFFENSE-NONVIOLENT
##                                793                                180
##                             GAMBLE                            HOMICIDE
##                                  0                                  9
##                  LIQUOR LAW VIOLATION                           LOITERING
##                                 57                                  1
##                   MOTOR VEHICLE THEFT                            NARCOTIC
##                               2284                                236
##                        PORNOGRAPHY                        PROSTITUTION
##                                  2                                 24
##                               RAPE                  ROBBERY-COMMERCIAL
##                                 62                                129
##                  ROBBERY-RESIDENTIAL                      ROBBERY-STREET
##                                 25                                200
##                    SEX OFFENSE-OTHER                     THEFT-ALL OTHER
##                                199                               2225
##                       THEFT-BICYCLE                      THEFT-BUILDING
##                                585                                881
##                      THEFT-SHOPLIFT                            TRESPASS
##                               1265                                470
##                             WEAPON
##                                102
```

```
which.max(table(dataQ4$category))
```

```
## CAR PROWL
##         9
```

According to the summary, the most frequent crime category in the Queen Anne neighborhood is “Car Prowl”

Question 5

Report a two-column table comparing the number of crimes per month(from RDate) What month seems to be more dangerous?

```
library(lubridate)
head(data2)
```

	Report.Number <dbl>	ODate <fctr>	OT… <int>	RDate <fctr>	RT… <int>	category <fctr>	Description <fctr>
1	2.008e+13	12/13/1908	2114	12/13/2008	2114	DUI	DUI-LIQUOR
2	2.010e+13	06/15/1964	0	06/15/2010	1031	FAMILY OFFENSE-NONVIOLENT	CHILD-OTH
3	2.012e+12	01/01/1973	0	01/25/2012	1048	SEX OFFENSE-OTHER	SEXOFF-OT
4	2.013e+13	06/01/1974	0	09/09/2013	1117	SEX OFFENSE-OTHER	SEXOFF-OT
5	2.016e+13	01/01/1975	0	08/11/2016	1054	SEX OFFENSE-OTHER	SEXOFF-OT
6	1.975e+12	12/16/1975	900	12/16/1975	1500	BURGLARY-RESIDENTIAL	BURGLARY-RES

6 rows | 1-8 of 9 columns

```
dataQ5 <- data2[c("RDate")]
month <- c(1,2,3,4,5,6,7,8,9,10,11,12)
dataQ5_1 <- data.frame("Month" = month,"Total_Num"=c(0))
head(dataQ5)
```

	RDate <fctr>
1	12/13/2008
2	06/15/2010
3	01/25/2012
4	09/09/2013
5	08/11/2016
6	12/16/1975

6 rows

```
library(stringr)
a1 = str_split(dataQ5$RDate,"/")
month <- rep(0,length(a1))
```

```
for(i in 1:length(a1))
  {
    month[i] = a1[[i]][1]
}
dataQ5$Month = month
head(dataQ5)
```

	RDate <fctr>	Month <chr>
1	12/13/2008	12
2	06/15/2010	06
3	01/25/2012	01
4	09/09/2013	09
5	08/11/2016	08
6	12/16/1975	12

6 rows

```
Q5_ans<-as.data.frame(table(dataQ5$Month))
names(Q5_ans)<-c("Month","Number_Crimes")
Q5_ans
```

Month <fctr>	Number_Crimes <int>
01	43006
02	37302
03	40860
04	40770
05	43672
06	42479
07	39214
08	39223
09	38967
10	40248

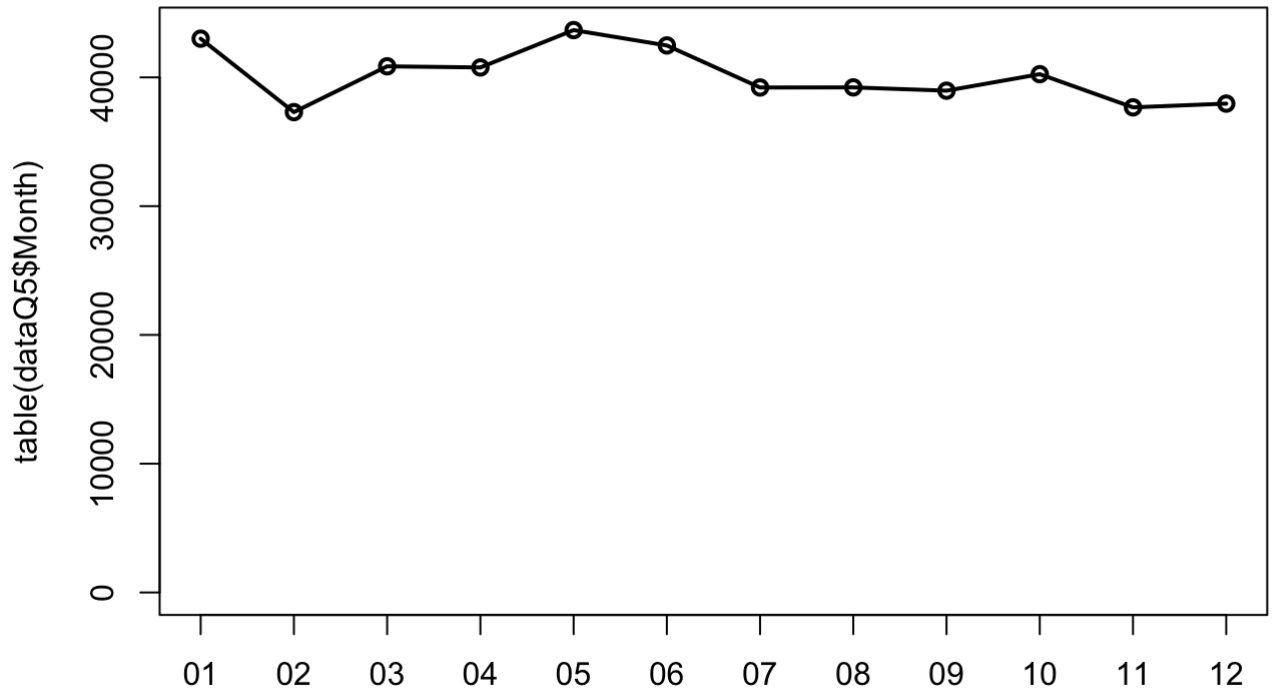
1-10 of 12 rows Previous **1** 2 Next

May seems to be more dangerous.

Question 6

Draw a lineplot of the number of crimes as a function of month (in RDate)

```
plot(table(dataQ5$Month),type="o")
```



Question 7

Use vertical barplots to compare the number of crimes by category in Seattle.

```
dataQ7 <- data2$category
as.data.frame(table(dataQ7))
```

dataQ7 <fctr>	Freq <int>
	262
AGGRAVATED ASSAULT	13954
AGGRAVATED ASSAULT-DV	6307
ARSON	1009
BURGLARY-COMMERCIAL	21274
BURGLARY-COMMERCIAL-SECURE PARKING	1042
BURGLARY-RESIDENTIAL	43908
BURGLARY-RESIDENTIAL-SECURE PARKING	7667
CAR PROWL	137766
DISORDERLY CONDUCT	245

1-10 of 31 rows Previous **1** 2 3 4 Next

```
barplot(table(dataQ7), main="Crime Distribution",
   xlab="Type of Crimes",cex.axis =0.4,cex.names = 0.3,las=2)
```