Hmk4-Q3

Question 3

Begin the analysis of one variable in the dataset you are using the final project. As this is an individual homework assignment, each group member should choose a different variable. Choose three visualizations as appropriate to show the distribution of the variable, conditioned on another variable if desired (for example, the distribution of income by region). Write a few sentences describing what you found and what new questions your visualizations have generated. (Faceted graphs count as one graph; graphs put together with grid.arrange() or similar count as multiple graphs.)

Read data

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.7
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

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

```
data = read.csv('Rodent_Inspection_77K_Sample_Clean 1.csv',
                header = TRUE,na.strings = "n/a")
head(data,20)
```

```
##         X1 INSPECTION_TYPE JOB_TICKET_OR_WORK_ORDER_ID    JOB_ID
## 1   444994      COMPLIANCE                     1752369 PO1711595
## 2  1291918         INITIAL                     1421971 PO1421971
## 3   225709            BAIT                      430548 PO1661724
## 4  1095170         INITIAL                     1072981 PO1072981
## 5   151814            BAIT                      301336 PO1265443
## 6   421002      COMPLIANCE                     1544595 PO1497160
## 7  1429651         INITIAL                     1650231 PO1650231
## 8   660953         INITIAL                      350376  PO350376
## 9   619873         INITIAL                      285038  PO285038
## 10 1012477         INITIAL                      920606  PO920606
## 11 1420440         INITIAL                     1636365 PO1636365
## 12  643877         INITIAL                      322667  PO322667
## 13  891397         INITIAL                      714303  PO714303
## 14  409916      COMPLIANCE                     1434806 PO1410199
## 15   24149            BAIT                       53173  PO238978
## 16  768633         INITIAL                      506225  PO506225
## 17  888998         INITIAL                      710311  PO710311
## 18 1145567         INITIAL                     1160570 PO1160570
## 19 1188602         INITIAL                     1246050 PO1246050
## 20  720151         INITIAL                      441936  PO441936
##    JOB_PROGRESS        BBL BORO_CODE BLOCK LOT HOUSE_NUMBER
## 1             2 1016460016         1  1646  16          225
## 2             1 4085650001         4  8565   1       252-12
## 3             4 2033260001         2  3326   1          115
```

```
## 4             1 1019070104         1  1907 104          165
## 5             3 4101620052         4 10162  52       104-59
## 6             2 4020900047         4  2090  47        97-25
## 7             1 1003550052         1   355  52          251
## 8             1 1011990009         1  1199   9           63
## 9             1 4036920001         4  3692   1        71-05
## 10            1 1017830038         1  1783  38          218
## 11            1 2029700006         2  2970   6          857
## 12            1 1013720026         1  1372  26            1
## 13            1 4099790036         4  9979  36        90-19
## 14            2 2033470050         2  3347  50         3271
## 15            5 1021450049         1  2145  49           82
## 16            1 2028080032         2  2808  32         1990
## 17            1 3033380031         3  3338  31          296
## 18            1 4034790068         4  3479  68         1911
## 19            1 1015230109         1  1523 109          125
## 20            1 3032560006         3  3256   6         1351
##                    STREET_NAME ZIP_CODE X_COORD Y_COORD         LATITUDE
## 1             EAST   96 STREET    10128  998582  225184  40.784736075117
## 2               UNION TURNPIKE    11426 1062308  210336  40.743647756671
## 3  EAST MOSHOLU PARKWAY NORTH    10467 1016868  258955  40.877379807432
## 4              WEST 122 STREET    10027  998340  233299 40.8070183298669
## 5                   164 STREET    11433 1041818  194231  40.699601134813
## 6                    64 AVENUE    11374 1023084  205223  40.729874028523
## 7          EAST HOUSTON STREET    10002  988311  202190 40.7216415852987
## 8               WEST 85 STREET    10024  992077  225470 40.7856379272557
## 9                MYRTLE AVENUE    11385 1017568  195067  40.702020771578
## 10             EAST 119 STREET    10035 1001483  230481 40.7992767516485
## 11             EAST 169 STREET    10459 1012795  241430 40.8292864948313
## 12                SUTTON PLACE    10022  995249  215202 40.7573899382086
## 13                  139 STREET    11435 1036391  194709 40.7009552422394
## 14                HULL AVENUE    10467 1018759  258528 40.8762090956968
## 15            WADSWORTH AVENUE    10033 1001573  247904  40.847102528429
## 16              CRESTON AVENUE    10453 1010229  249310 40.8509232057752
## 17          ST NICHOLAS AVENUE    11237 1008790  194770 40.7012408100934
## 18             WOODBINE STREET    11385 1010650  196087 40.7048512050519
## 19              EAST 94 STREET    10128  997417  225298 40.7850584014609
## 20               MYRTLE AVENUE    11221 1005548  193769 40.6984988728113
##             LONGITUDE   BOROUGH INSPECTION_DATE             RESULT
## 1   -73.948246192798 Manhattan      2018-07-11   Active Rat Signs
## 2   -73.718301394645    Queens      2017-05-19  Passed Inspection
## 3   -73.882049446771     Bronx      2018-09-04       Bait applied
## 4  -73.9491028928087 Manhattan      2015-11-30  Passed Inspection
## 5    -73.79238442044    Queens      2017-01-04       Bait applied
## 6   -73.859883682979    Queens      2017-11-21  Passed Inspection
## 7  -73.9853470218584 Manhattan      2018-03-14  Passed Inspection
## 8  -73.9716880802003 Manhattan      2011-10-20  Passed Inspection
## 9    -73.879836305361    Queens      2011-04-29 Problem Conditions
## 10 -73.9377564304275 Manhattan      2015-01-22  Passed Inspection
## 11 -73.8968615568375     Bronx      2018-02-22  Passed Inspection
## 12 -73.9603027158538 Manhattan      2011-08-15  Passed Inspection
## 13 -73.8119548568868    Queens      2013-11-08 Problem Conditions
## 14 -73.8752138556471     Bronx      2017-07-05  Passed Inspection
## 15 -73.9373251458908 Manhattan      2011-05-14       Bait applied
```

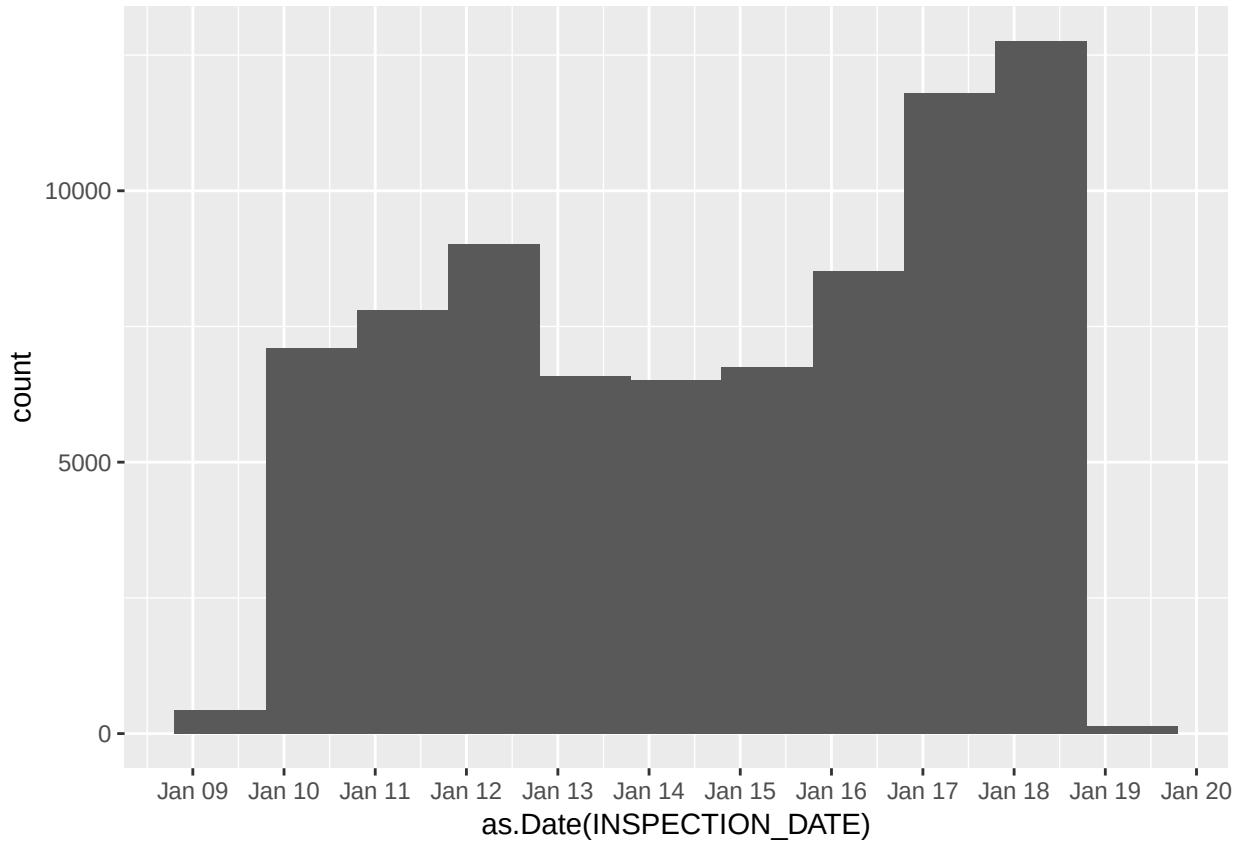
```
## 16 -73.9060686340153     Bronx      2012-06-28   Active Rat Signs
## 17 -73.9114949734671  Brooklyn      2013-10-08  Passed Inspection
## 18 -73.9047792778727    Queens      2016-05-10  Passed Inspection
## 19 -73.9524531112948 Manhattan      2016-10-04  Passed Inspection
## 20 -73.9231900493702  Brooklyn      2012-04-09  Passed Inspection
##    APPROVED_DATE                             LOCATION
## 1     2018-07-12   (40.784736075117, -73.948246192798)
## 2     2017-05-25   (40.743647756671, -73.718301394645)
## 3     2018-09-05   (40.877379807432, -73.882049446771)
## 4     2015-12-01 (40.8070183298669, -73.9491028928087)
## 5     2017-01-05    (40.699601134813, -73.79238442044)
## 6     2017-11-27   (40.729874028523, -73.859883682979)
## 7     2018-03-20 (40.7216415852987, -73.9853470218584)
## 8     2011-10-27 (40.7856379272557, -73.9716880802003)
## 9     2011-05-02   (40.702020771578, -73.879836305361)
## 10    2015-01-26 (40.7992767516485, -73.9377564304275)
## 11    2018-02-28 (40.8292864948313, -73.8968615568375)
## 12    2011-08-17 (40.7573899382086, -73.9603027158538)
## 13    2013-11-13 (40.7009552422394, -73.8119548568868)
## 14    2017-07-07 (40.8762090956968, -73.8752138556471)
## 15    2011-05-23  (40.847102528429, -73.9373251458908)
## 16    2012-07-03 (40.8509232057752, -73.9060686340153)
## 17    2013-10-11 (40.7012408100934, -73.9114949734671)
## 18    2016-05-12 (40.7048512050519, -73.9047792778727)
## 19    2016-10-05 (40.7850584014609, -73.9524531112948)
## 20    2012-04-11 (40.6984988728113, -73.9231900493702)
```

Because the data are huge and contain . To make visualization simpler and more direct I decide to manipulate time data to year-month data at first.

```
data$INSPECTION_DATE = format(as.Date(data$INSPECTION_DATE),format='%y-%m')
data$INSPECTION_DATE = paste(data$INSPECTION_DATE, "01", sep="-")
```

And I then start exploring from year to find out the overall trend.

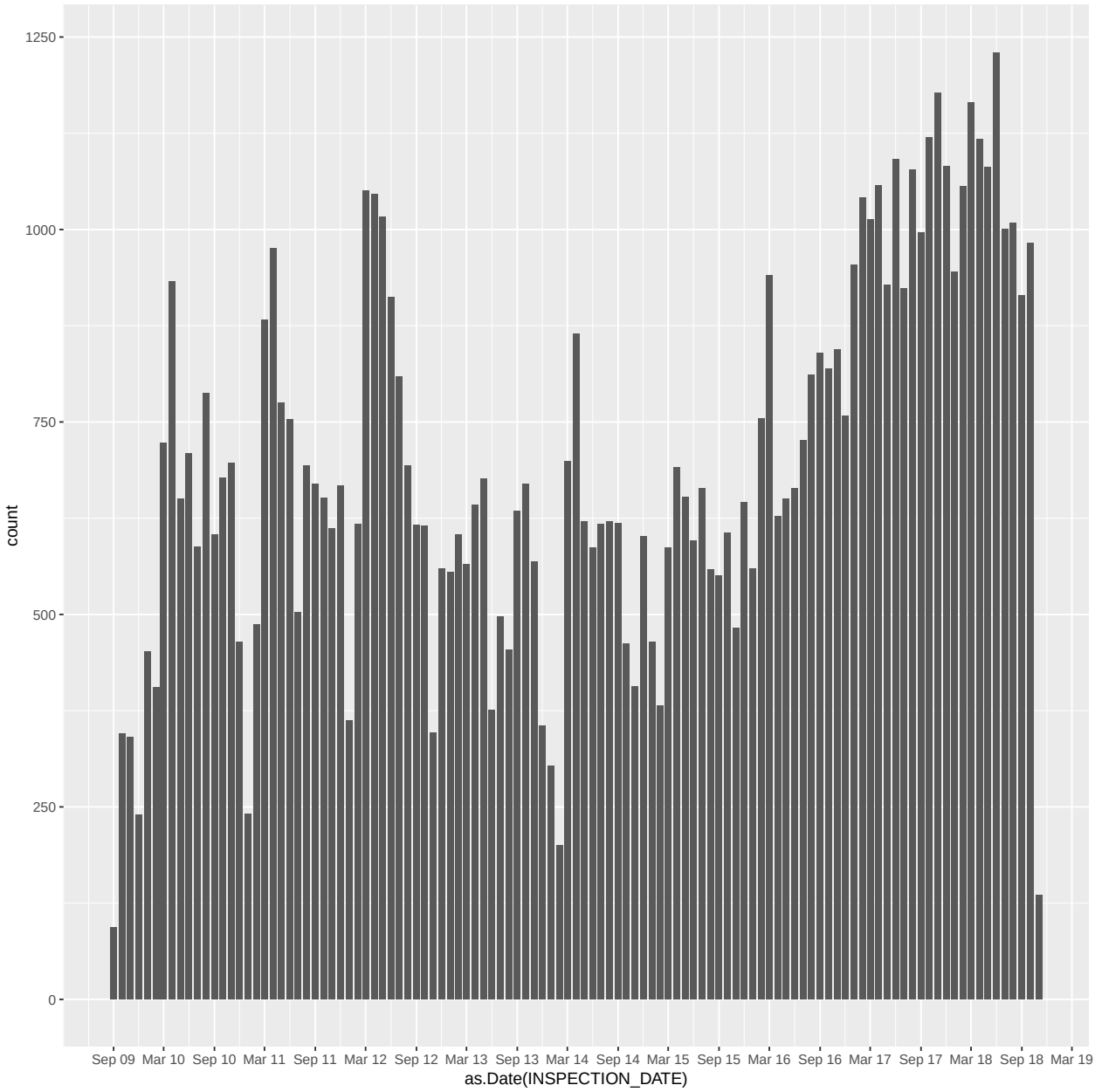
```
ggplot(data, aes(as.Date(INSPECTION_DATE))) +
 geom_histogram(binwidth = 365) +
 scale_x_date(date_breaks = "1 year", date_labels = "%b %y")
```



Based on the graph above, the number of rodent inspection is generally increasing but it dropped down in 2013.

And then I checked the monthly change of the data.

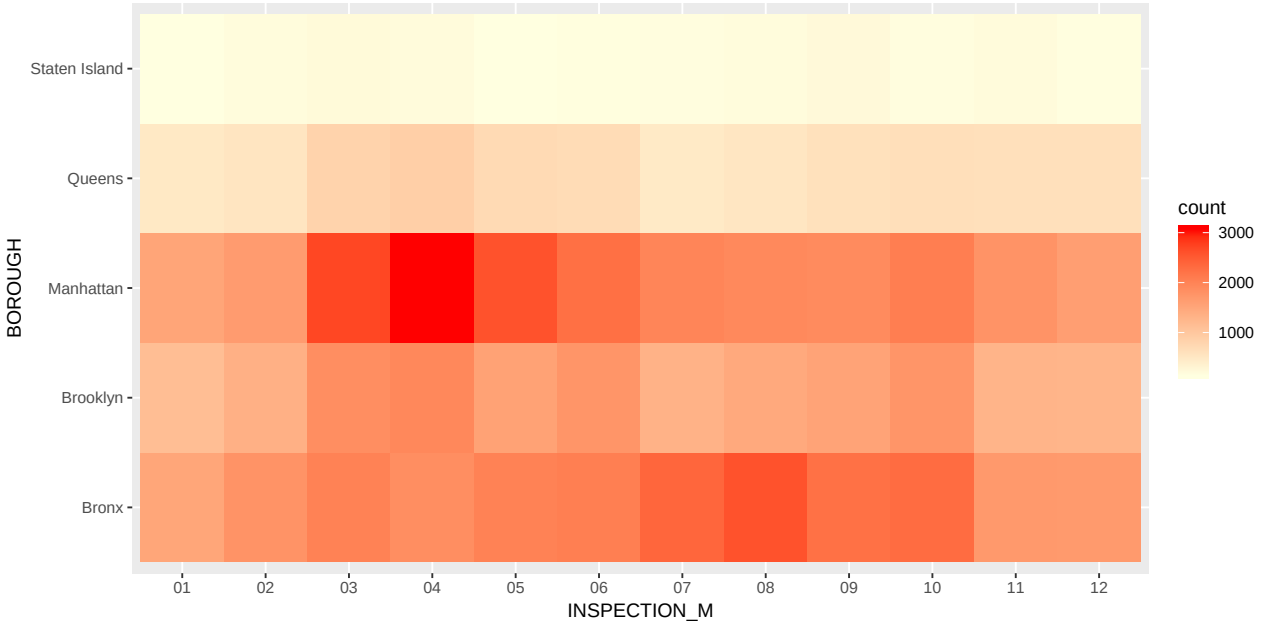
```
ggplot(data, aes(as.Date(INSPECTION_DATE))) +
  geom_bar(position = position_dodge(width = 1/3)) +
  scale_x_date(date_breaks = "6 months", date_labels = "%b %y")
```



From the Figure above, we can find that in most years, rodent inspection happened fewer from November to Feburary which are usually cold.

To further explore the how different month influences the the number of rodent inspection, I chose to use heat map to find the relationship between month and borough.

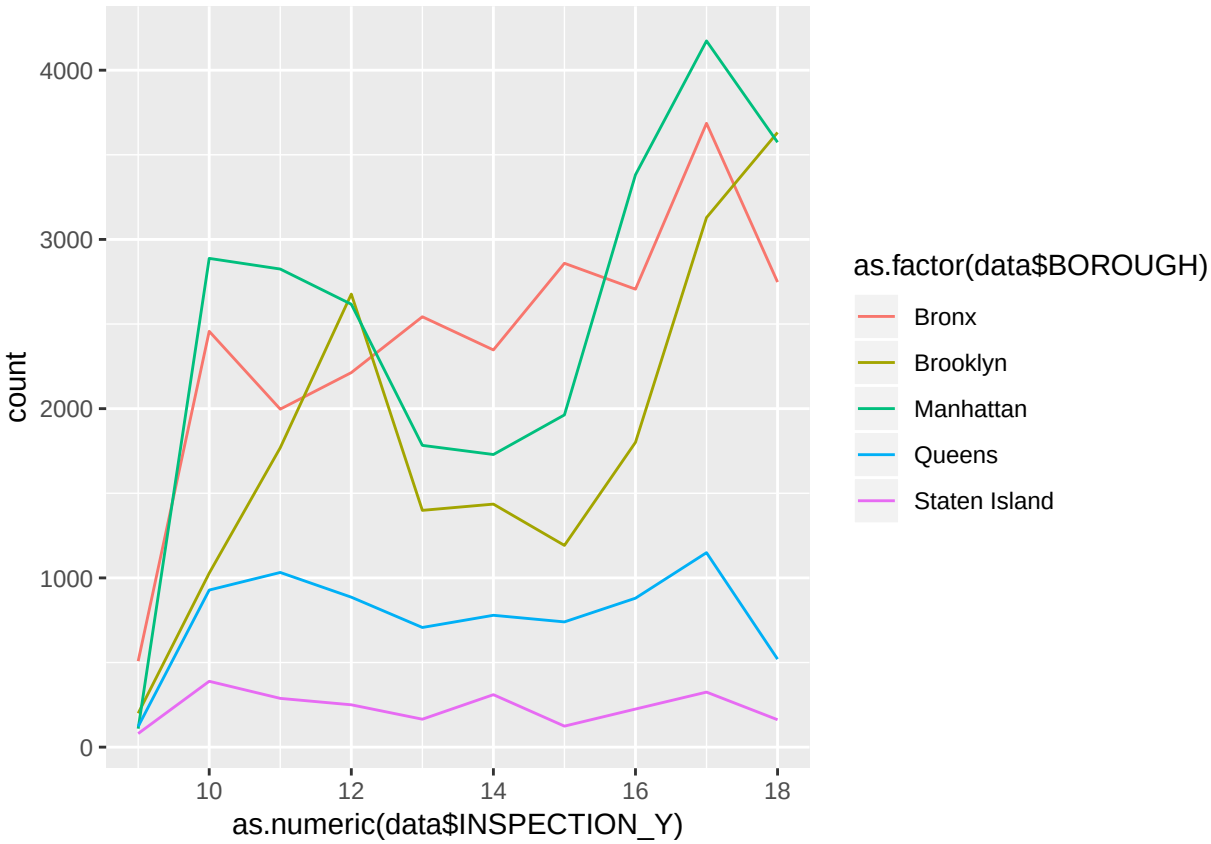
```
data$INSPECTION_M = format(as.Date(data$INSPECTION_DATE),format='%m')
ggplot(data,aes(x=INSPECTION_M, y = BOROUGH)) +
  geom_bin2d(binwidth = c(3, 1)) +
  scale_fill_gradient(low = 'lightyellow',high = 'red')
```



As shown above, we can easily draw the conclusion that most rodent inspections happened in Manhattan and Bronx from March to October.

```
data$INSPECTION_Y = format(as.Date(data$INSPECTION_DATE),format='%y')


ggplot(data, aes(x = as.numeric(data$INSPECTION_Y),
                 group = data$BOROUGH, colour = as.factor(data$BOROUGH))) +
  geom_line(stat = 'count')
```



Based on the graph above and first histogram, we can find that Queens and Staten Island have rather steady Rodent Inspection while Manhattan and Brooklyn mainly lead to decreasing of Rodent Inspection from 2013 to 2015. And the number of Rodent Inspection in Bronx is generally increasing.

According to the graphs above focusing on the time data, we can find the main trend of Rodent Inspection is increasing. And different borough has different distribution of Rodent Inspection. Further more, we can find that the month is related to Rodent Inspection, especially in Manhattan and Bronx.