**Assignment #3**

Complete the following tasks:

• Using the dataset of your choice, do the following

• Import the data using code

library(readxl)

TxBorder <- read\_excel("C:/Users/Nancy/Downloads/Border-Crossings.xls.xlsx",

col\_types = c("text", "text", "text",

"numeric", "numeric", "numeric",

"numeric", "numeric"))

• Create a subset of your dataset with only the variables you want to use for this assignment

recentTxBB<-TxBorder[c(1:36), c(6, 7, 8)]

• Choose three variables of interest and complete the following tasks:

o Describe the central tendency of the variables

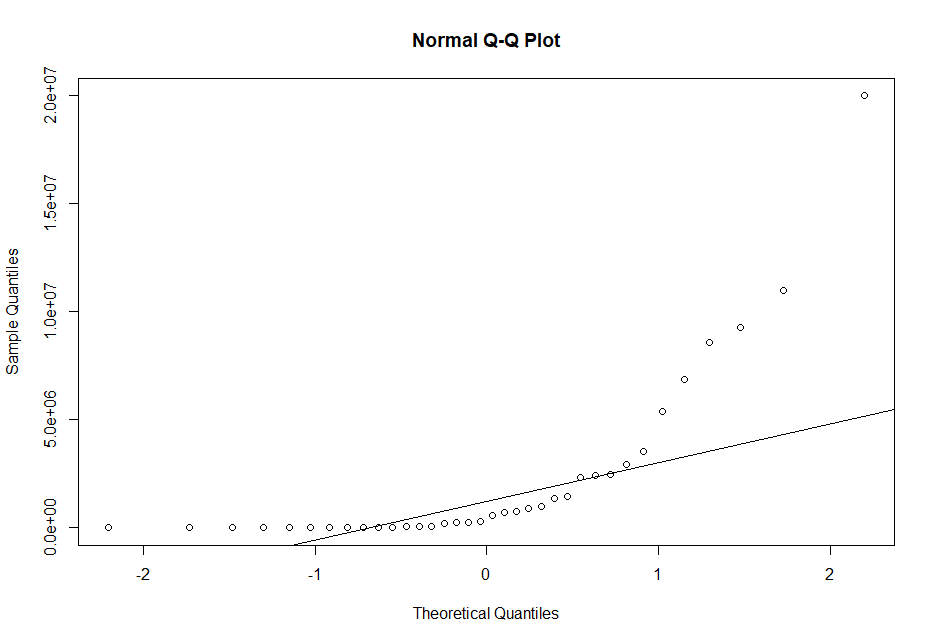
|  |  |  |
| --- | --- | --- |
| 2015 | 2016 | 2017 |
| Median: 430044 | Median: 471336 | Median: 456142 |
| Mean: 2297787 | Mean: 2380660 | Mean: 2394135 |
| Mode: 0 | Mode: 0 | Mode: 0 |

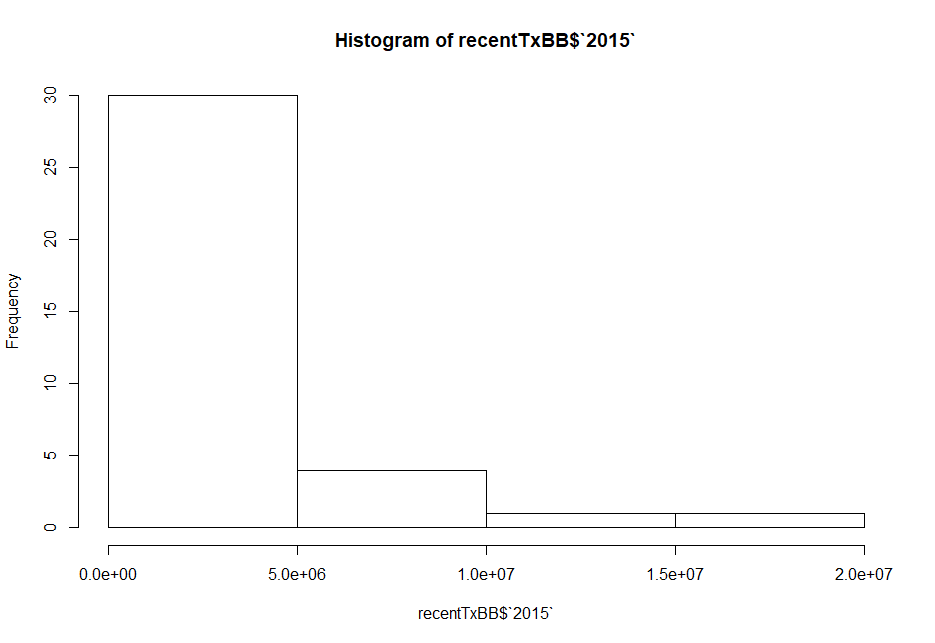
o Describe the variables through variation

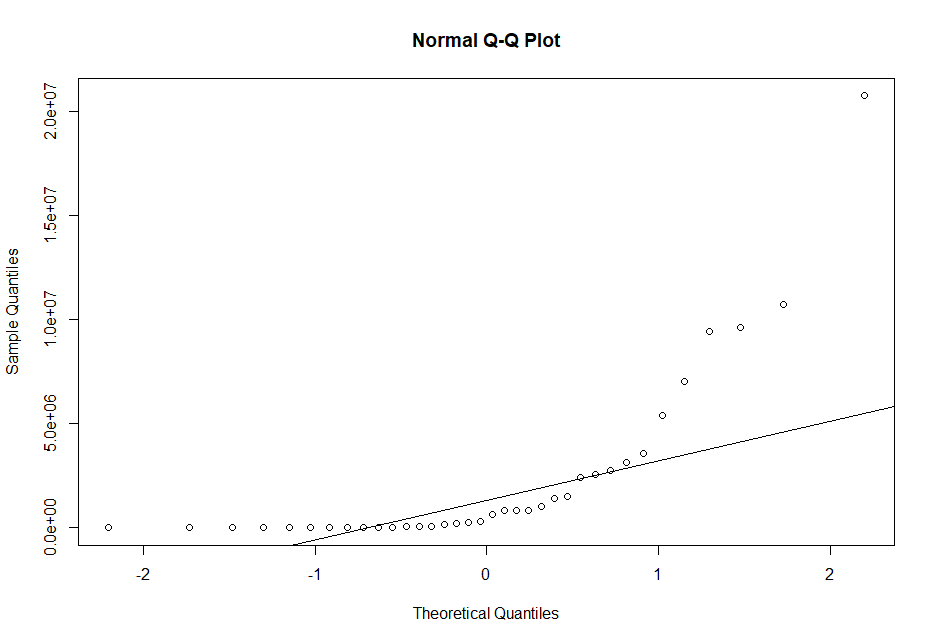
|  |  |  |
| --- | --- | --- |
| 2015 | 2016 | 2017 |
| Range: 19982407 | Range: 20767737 | Range: 22046772 |
| SD: 4178219 | SD: 4322183 | SD: 4472522 |
| Varience: 1.745751e+13 | Varience: 1.868126e+13 | Varience: 2.000345e+13 |

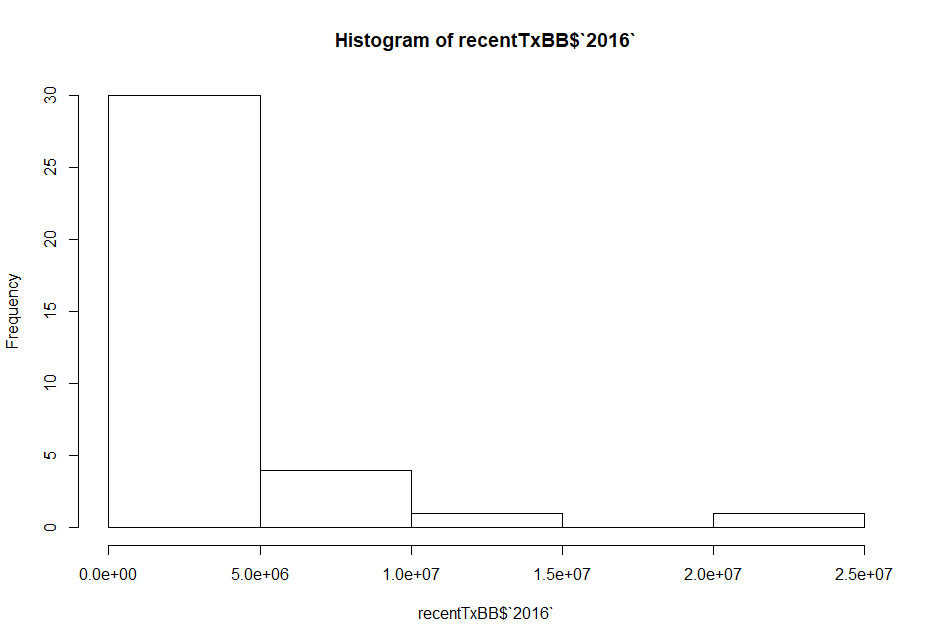
o Discuss normality. Are your variables approximating normality? What are ways that you can show that they are or are not?

By using a histogram and a qqplot for each, outiers exist in the variables. While the graphs are hard to read due to the large numbers, some normality can be seen.

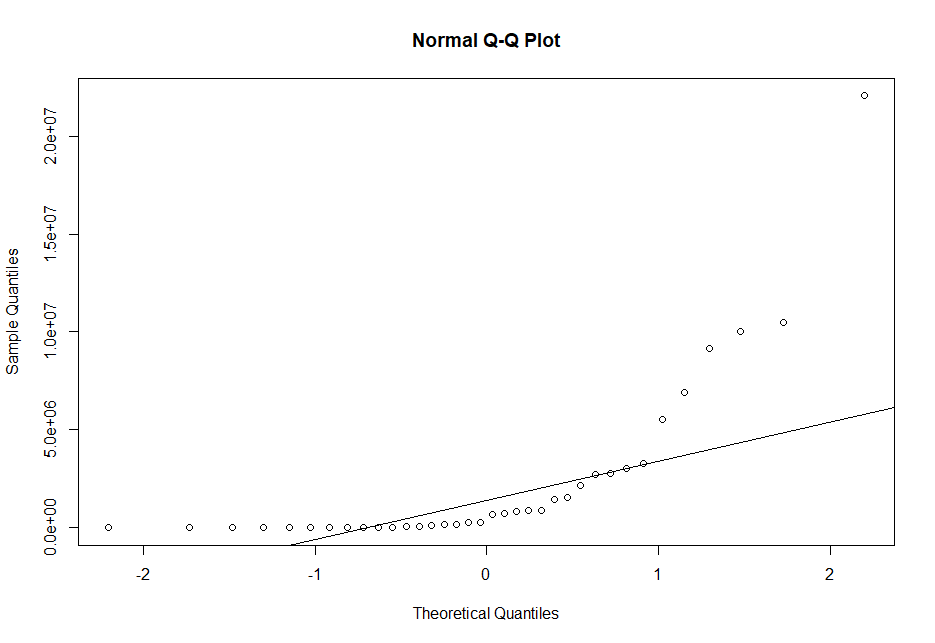
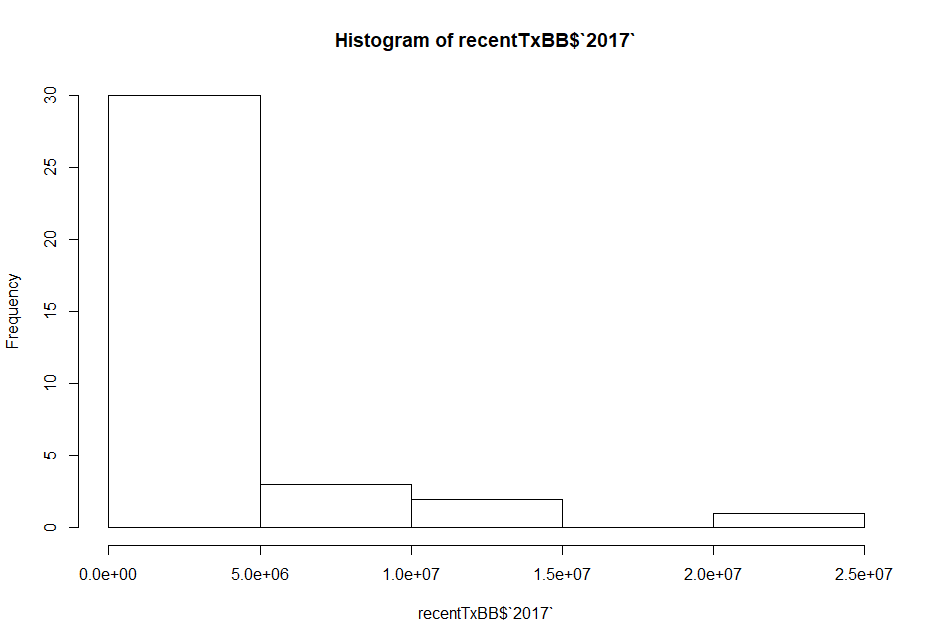
2015



2016



2017



What to turn in

• Publish your homework to GitHub under your user account, the appropriate format (check syllabus for instructions). This means turn in a separate document from your code that answers the questions above.

• Upload the code you used to GitHub. • Provide me the links of these two files via e-mail no later than Thursday, October 4th at 6:00pm.