Brazil House Rent

Danuka

2023-12-19

## After that Attach the dataset to R and start the Analysis.

BrazilianHousestoRent <- readxl::read\_xlsx("BrazilianHousestoRent.xlsx")  
str(BrazilianHousestoRent) #column names and data type

## tibble [10,692 × 13] (S3: tbl\_df/tbl/data.frame)  
## $ city : chr [1:10692] "SÃ£o Paulo" "SÃ£o Paulo" "Porto Alegre" "Porto Alegre" ...  
## $ area : num [1:10692] 70 320 80 51 25 376 72 213 152 35 ...  
## $ rooms : num [1:10692] 2 4 1 2 1 3 2 4 2 1 ...  
## $ bathroom : num [1:10692] 1 4 1 1 1 3 1 4 2 1 ...  
## $ parking spaces: num [1:10692] 1 0 1 0 0 7 0 4 1 0 ...  
## $ floor : chr [1:10692] "7" "20" "6" "2" ...  
## $ keeping animal: chr [1:10692] "acept" "acept" "acept" "acept" ...  
## $ furniture : chr [1:10692] "furnished" "not furnished" "not furnished" "not furnished" ...  
## $ hoa : num [1:10692] 2065 1200 1000 270 0 ...  
## $ rent amount : num [1:10692] 3300 4960 2800 1112 800 ...  
## $ property tax : num [1:10692] 211 1750 0 22 25 ...  
## $ fire insurance: num [1:10692] 42 63 41 17 11 121 25 41 191 30 ...  
## $ total rent : num [1:10692] 5618 7973 3841 1421 836 ...

colnames(BrazilianHousestoRent) #only column names

## [1] "city" "area" "rooms" "bathroom"   
## [5] "parking spaces" "floor" "keeping animal" "furniture"   
## [9] "hoa" "rent amount" "property tax" "fire insurance"  
## [13] "total rent"

# 

# DATA PREPOSESING

### First, we have to check there is any missing values. In our data set there is no any missing values. After that, we have to handle outliers. i used to numerical method to do this

## Now I replace the 0 in floor variable instead of “-” symbol.

BrazilHouserent=clean\_dataFile  
str(BrazilHouserent)

## tibble [8,873 × 13] (S3: tbl\_df/tbl/data.frame)  
## $ city : chr [1:8873] "SÃ£o Paulo" "Porto Alegre" "Porto Alegre" "SÃ£o Paulo" ...  
## $ area : num [1:8873] 70 80 51 25 72 35 26 46 36 55 ...  
## $ rooms : num [1:8873] 2 1 2 1 2 1 1 1 1 1 ...  
## $ bathroom : num [1:8873] 1 1 1 1 1 1 1 1 1 1 ...  
## $ parking spaces: num [1:8873] 1 1 0 0 0 0 0 1 0 1 ...  
## $ floor : chr [1:8873] "7" "6" "2" "1" ...  
## $ keeping animal: chr [1:8873] "acept" "acept" "acept" "not acept" ...  
## $ furniture : chr [1:8873] "furnished" "not furnished" "not furnished" "not furnished" ...  
## $ hoa : num [1:8873] 2065 1000 270 0 740 ...  
## $ rent amount : num [1:8873] 3300 2800 1112 800 1900 ...  
## $ property tax : num [1:8873] 211 0 22 25 85 35 150 43 70 224 ...  
## $ fire insurance: num [1:8873] 42 41 17 11 25 30 27 8 27 54 ...  
## $ total rent : num [1:8873] 5618 3841 1421 836 2750 ...  
## - attr(\*, "na.action")= 'omit' Named int [1:1819] 2 6 8 9 20 38 41 45 64 76 ...  
## ..- attr(\*, "names")= chr [1:1819] "2" "6" "8" "9" ...

glimpse(BrazilHouserent)

## Rows: 8,873  
## Columns: 13  
## $ city <chr> "SÃ£o Paulo", "Porto Alegre", "Porto Alegre", "SÃ£o P…  
## $ area <dbl> 70, 80, 51, 25, 72, 35, 26, 46, 36, 55, 100, 330, 110…  
## $ rooms <dbl> 2, 1, 2, 1, 2, 1, 1, 1, 1, 1, 2, 4, 2, 2, 2, 7, 1, 3,…  
## $ bathroom <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 6, 2, 3, 1, 4, 1, 2,…  
## $ `parking spaces` <dbl> 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 2, 6, 1, 1, 0, 0, 1, 2,…  
## $ floor <chr> "7", "6", "2", "1", "7", "2", "2", "10", "11", "2", "…  
## $ `keeping animal` <chr> "acept", "acept", "acept", "not acept", "acept", "ace…  
## $ furniture <chr> "furnished", "not furnished", "not furnished", "not f…  
## $ hoa <dbl> 2065, 1000, 270, 0, 740, 590, 470, 550, 359, 790, 900…  
## $ `rent amount` <dbl> 3300, 2800, 1112, 800, 1900, 2300, 2100, 580, 2100, 4…  
## $ `property tax` <dbl> 211, 0, 22, 25, 85, 35, 150, 43, 70, 224, 17, 328, 12…  
## $ `fire insurance` <dbl> 42, 41, 17, 11, 25, 30, 27, 8, 27, 54, 56, 121, 39, 1…  
## $ `total rent` <dbl> 5618, 3841, 1421, 836, 2750, 2955, 2747, 1181, 2556, …

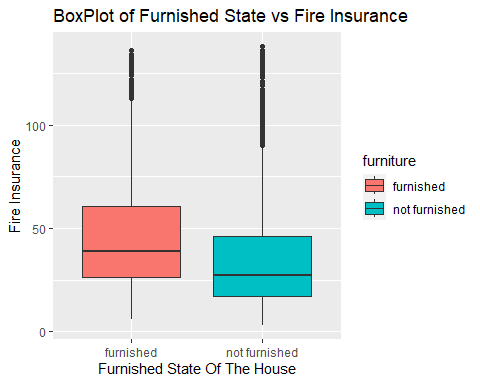
BrazilHouserent <- BrazilHouserent %>%  
 mutate(floor = ifelse(floor == "-", 0, floor))  
View(BrazilHouserent)

# 

# DATA ANALYSIS

## 1. Does the fire insurance relate to when the house is furnished or not ?

ggplot(data = BrazilHouserent)+  
 geom\_boxplot(mapping =aes(x=furniture,y=`fire insurance`,fill=furniture) )+  
 labs(x="Furnished State Of The House",y="Fire Insurance")+  
 ggtitle("BoxPlot of Furnished State vs Fire Insurance")+  
 scale\_fill\_discrete(name = "furniture")



summary\_by\_category <- by(BrazilHouserent$`fire insurance`, BrazilHouserent$furniture, fivenum)  
summary\_by\_category

## BrazilHouserent$furniture: furnished  
## [1] 6 26 39 61 136  
## ------------------------------------------------------------   
## BrazilHouserent$furniture: not furnished  
## [1] 3 17 27 46 138

kruskal.test(BrazilHouserent$`fire insurance` ~BrazilHouserent$furniture, data=BrazilHouserent)

##   
## Kruskal-Wallis rank sum test  
##   
## data: BrazilHouserent$`fire insurance` by BrazilHouserent$furniture  
## Kruskal-Wallis chi-squared = 434.54, df = 1, p-value < 2.2e-16

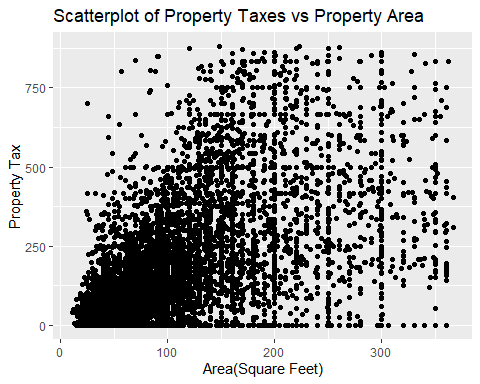
qchisq(0.95,1,lower.tail = TRUE)

## [1] 3.841459

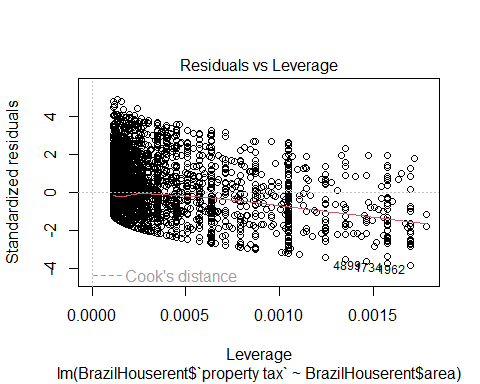
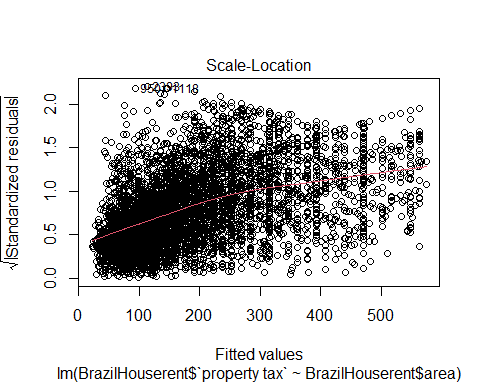
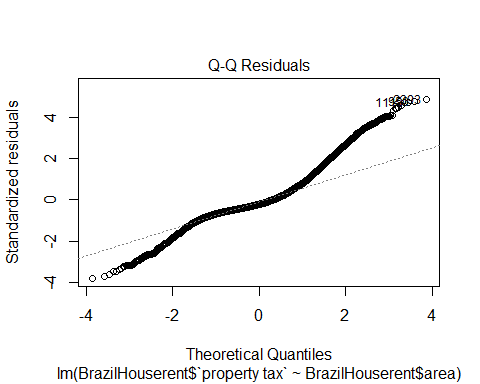
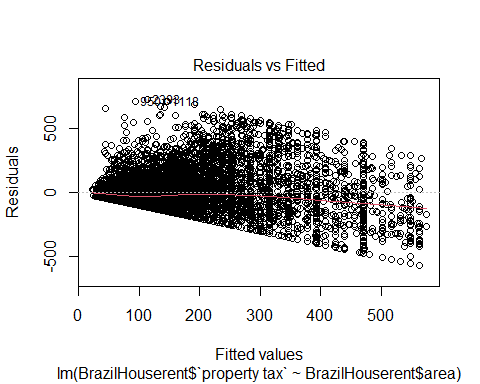
## 

## 2. Does the property taxes relate to the area ?

ggplot(data = BrazilHouserent)+  
 geom\_point(mapping =aes(x=area,y=`property tax`) )+  
 labs(x="Area(Square Feet)",y="Property Tax")+  
 ggtitle("Scatterplot of Property Taxes vs Property Area")



lm1=lm(BrazilHouserent$`property tax`~BrazilHouserent$area)  
plot(lm1)

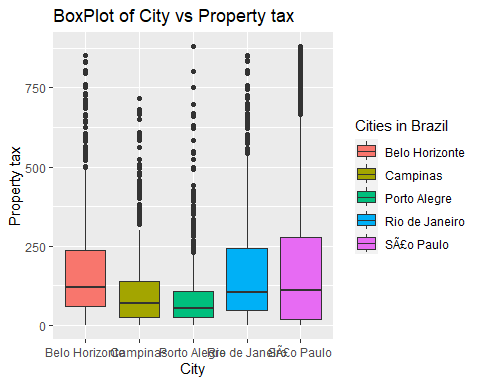


crr=cor(BrazilHouserent$area,BrazilHouserent$`property tax`,method = "spearman")  
crr

## [1] 0.5861278

## 3. Does the location relate to the Property taxes ?

ggplot(data = BrazilHouserent)+  
 geom\_boxplot(mapping =aes(x=city,y=`property tax`,fill=city) )+  
 labs(x="City",y="Property tax")+  
 ggtitle("BoxPlot of City vs Property tax")+  
 scale\_fill\_discrete(name = "Cities in Brazil")



summary\_by\_category <- by(BrazilHouserent$`property tax`, BrazilHouserent$city, fivenum)  
summary\_by\_category

## BrazilHouserent$city: Belo Horizonte  
## [1] 0 61 119 236 852  
## ------------------------------------------------------------   
## BrazilHouserent$city: Campinas  
## [1] 0.0 25.0 70.0 138.5 715.0  
## ------------------------------------------------------------   
## BrazilHouserent$city: Porto Alegre  
## [1] 0 24 54 106 880  
## ------------------------------------------------------------   
## BrazilHouserent$city: Rio de Janeiro  
## [1] 0.0 46.5 103.0 244.5 853.0  
## ------------------------------------------------------------   
## BrazilHouserent$city: SÃ£o Paulo  
## [1] 0.0 19.0 109.0 276.5 880.0

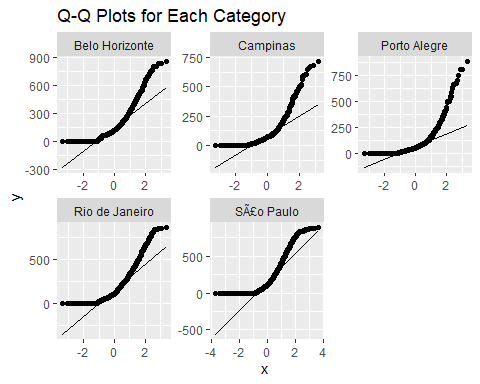
city\_tax\_summary <- BrazilHouserent %>%  
 group\_by(city) %>%  
 summarize(mean\_tax = mean(`property tax`, na.rm = TRUE),  
 median\_tax = median(`property tax`, na.rm = TRUE),  
 max\_tax = max(`property tax`, na.rm = TRUE),  
 min\_tax = min(`property tax`, na.rm = TRUE))  
   
city\_tax\_summary

## # A tibble: 5 × 5  
## city mean\_tax median\_tax max\_tax min\_tax  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Belo Horizonte 169. 119 852 0  
## 2 Campinas 109. 70 715 0  
## 3 Porto Alegre 90.6 54 880 0  
## 4 Rio de Janeiro 166. 103 853 0  
## 5 SÃ£o Paulo 183. 109 880 0

kruskal.test(BrazilHouserent$`property tax` ~BrazilHouserent$city, data=BrazilHouserent)

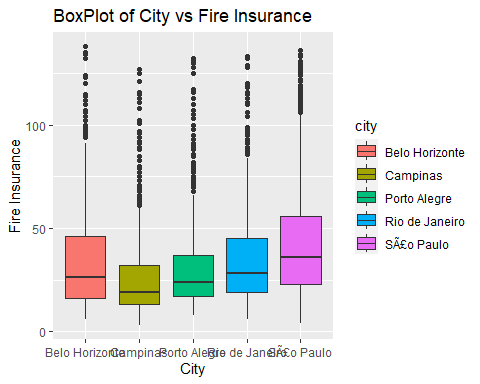
##   
## Kruskal-Wallis rank sum test  
##   
## data: BrazilHouserent$`property tax` by BrazilHouserent$city  
## Kruskal-Wallis chi-squared = 226.15, df = 4, p-value < 2.2e-16

ggplot(BrazilHouserent, aes(sample = `property tax`)) +  
 geom\_qq() +  
 facet\_wrap(~ city, scales = "free") +  
 stat\_qq\_line() +  
 labs(title = "Q-Q Plots for Each Category")



## 4 .Does the fire insurance price relate to the location ?

ggplot(data = BrazilHouserent)+  
 geom\_boxplot(mapping =aes(x=city,y=`fire insurance`,fill=city) )+  
 labs(x="City",y="Fire Insurance")+  
 ggtitle("BoxPlot of City vs Fire Insurance")+  
 scale\_fill\_discrete(name = "city")



city\_FireInsu\_summary <- BrazilHouserent %>%  
 group\_by(city) %>%  
 summarize(mean\_Insurance = mean(`fire insurance`, na.rm = TRUE),  
 median\_Insurance = median(`fire insurance`, na.rm = TRUE),  
 max\_Insurance = max(`fire insurance`, na.rm = TRUE),  
 min\_Insurance = min(`fire insurance`, na.rm = TRUE))  
   
city\_FireInsu\_summary

## # A tibble: 5 × 5  
## city mean\_Insurance median\_Insurance max\_Insurance min\_Insurance  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Belo Horizonte 35.1 26 138 6  
## 2 Campinas 26.5 19 127 3  
## 3 Porto Alegre 30.5 24 132 8  
## 4 Rio de Janeiro 34.9 28 133 6  
## 5 SÃ£o Paulo 43.4 36 136 4

city\_counts <- BrazilHouserent %>%  
 count(city)  
city\_counts

## # A tibble: 5 × 2  
## city n  
## <chr> <int>  
## 1 Belo Horizonte 1047  
## 2 Campinas 796  
## 3 Porto Alegre 1127  
## 4 Rio de Janeiro 1379  
## 5 SÃ£o Paulo 4524

kruskal.test(BrazilHouserent$`fire insurance` ~BrazilHouserent$city, data=BrazilHouserent)

##   
## Kruskal-Wallis rank sum test  
##   
## data: BrazilHouserent$`fire insurance` by BrazilHouserent$city  
## Kruskal-Wallis chi-squared = 732.47, df = 4, p-value < 2.2e-16

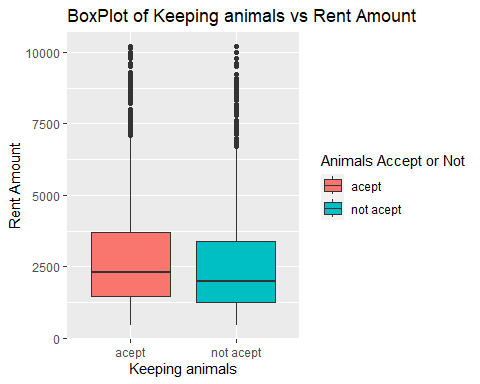
qchisq(0.95,4,lower.tail = TRUE)

## [1] 9.487729

## 

## 5 .Does the rent amount relate to the keeping animal?

ggplot(data = BrazilHouserent)+  
 geom\_boxplot(mapping =aes(x=`keeping animal`,y=`rent amount`,fill=`keeping animal` ))+  
 labs(x="Keeping animals",y="Rent Amount")+  
 ggtitle("BoxPlot of Keeping animals vs Rent Amount")+  
 scale\_fill\_discrete(name = "Animals Accept or Not")



summary\_by\_category <- by(BrazilHouserent$`rent amount`, BrazilHouserent$`keeping animal`, fivenum)  
summary\_by\_category

## BrazilHouserent$`keeping animal`: acept  
## [1] 450 1450 2300 3700 10200  
## ------------------------------------------------------------   
## BrazilHouserent$`keeping animal`: not acept  
## [1] 460 1250 2000 3400 10200

kruskal.test(BrazilHouserent$`rent amount` ~BrazilHouserent$`keeping animal`, data=BrazilHouserent)

##   
## Kruskal-Wallis rank sum test  
##   
## data: BrazilHouserent$`rent amount` by BrazilHouserent$`keeping animal`  
## Kruskal-Wallis chi-squared = 38.714, df = 1, p-value = 4.906e-10

qchisq(0.95,4,lower.tail = TRUE)

## [1] 9.487729