Rental Pricing Data Analysis

Rental Pricing Trend within Klang Valley

Dataset was taken from Kaggle, webscraped from Mudah.com by a public user. This project aims to provide visual insights on the trend of monthly rent with given category (provided in Mudah website) such as location, size, number of bedrooms, parking or furnished. This dataset is chosen as I am also actively looking for a new unit to move in Klang Valley. Therefore having this insights will hopefully help me compare my options and choices that I already had, to compare it within the actual market trend.

Installing data packages and library

```
options(repos = "https://cloud.r-project.org")
```

```
install.packages("stringr")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'stringr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpU58XcE\downloaded_packages
```

```
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 4.3.1
```

```
install.packages("ggplot2")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpU58XcE\downloaded_packages
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/1/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\1\AppData\Local\Temp\RtmpU58XcE\downloaded_packages
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
install.packages("dplyr")
```

```
## Warning: package 'dplyr' is in use and will not be installed
```

```
library(dplyr)
library(readr)
```

```
rental_df = read.csv("Rental Pricing.csv")
```

Data Exploration

To describe the whole dataset (rows and columns), also checking on how many empty cells are within the dataset.

```
colnames(rental_df)
```

```
##  [1] "ads_id"                "property_type"         "prop_name"
##  [4] "completion_year"       "monthly_rent"          "district"
##  [7] "parking"               "bathroom"              "size"
## [10] "furnished"             "facilities"            "additional_facilities"
## [13] "region"
```

```r
str(rental_df)
```

```
## 'data.frame':    19991 obs. of  13 variables:
##  $ ads_id               : int  100323185 100203973 100323128 100191767 97022692 100322897 100322962 
##  $ property_type        : chr  "Condominium" "Condominium" "Apartment" "Apartment" ...
##  $ prop_name            : chr  "The Hipster @ Taman Desa" "Segar Courts" "Pangsapuri Teratak Muhibbal
##  $ completion_year      : int  2022 NA NA 2020 NA NA NA 2018 2014 NA ...
##  $ monthly_rent         : num  4200 2300 1000 1700 1299 ...
##  $ district             : chr  " Taman Desa" " Cheras" " Taman Desa" " Sentul" ...
##  $ parking              : int  2 1 NA 1 1 1 2 1 1 1 ...
##  $ bathroom             : int  6 2 2 2 1 2 2 1 1 2 ...
##  $ size                 : int  1842 1170 650 743 494 884 982 700 750 862 ...
##  $ furnished            : chr  "Fully Furnished" "Partially Furnished" "Fully Furnished" "Partially F
##  $ facilities           : chr  "Minimart, Gymnasium, Security, Playground, Swimming Pool, Parking, Li
##  $ additional_facilities: chr  "Air-Cond, Cooking Allowed, Washing Machine" "Air-Cond, Cooking Allowe
##  $ region               : chr  "Kuala Lumpur" "Kuala Lumpur" "Kuala Lumpur" "Kuala Lumpur" ...
```

```r
head(rental_df)
```

```
##      ads_id     property_type                    prop_name completion_year
## 1 100323185       Condominium      The Hipster @ Taman Desa            2022
## 2 100203973       Condominium                   Segar Courts              NA
## 3 100323128         Apartment Pangsapuri Teratak Muhibbah 2              NA
## 4 100191767         Apartment  Sentul Point Suite Apartment            2020
## 5  97022692 Service Residence               Arte Mont Kiara              NA
## 6 100322897         Apartment      Residensi Vista Wirajaya              NA
##   monthly_rent    district parking bathroom size           furnished
## 1         4200  Taman Desa       2        6 1842     Fully Furnished
## 2         2300      Cheras       1        2 1170 Partially Furnished
## 3         1000  Taman Desa      NA        2  650     Fully Furnished
## 4         1700      Sentul       1        2  743 Partially Furnished
## 5         1299  Mont Kiara       1        1  494       Not Furnished
## 6         1500     Setapak       1        2  884 Partially Furnished
##
## 1            Minimart, Gymnasium, Security, Playground, Swimming Pool, Parking, Lift, Barbeque area
## 2                                 Playground, Parking, Barbeque area, Security, Jogging Track, Swi
## 3                                                                                          Minimart, 
## 4                                              Parking, Playground, Swimming Pool, Squash Court, Se
## 5 Parking, Security, Lift, Swimming Pool, Playground, Gymnasium, Barbeque area, Minimart, Multipurpos
## 6                                                          Parking, Security, Lift, Swimming Po
##                           additional_facilities       region
## 1     Air-Cond, Cooking Allowed, Washing Machine Kuala Lumpur
## 2        Air-Cond, Cooking Allowed, Near KTM/LRT Kuala Lumpur
## 3                                                Kuala Lumpur
## 4 Cooking Allowed, Near KTM/LRT, Washing Machine Kuala Lumpur
## 5                                     Air-Cond Kuala Lumpur
## 6                   Cooking Allowed, Near KTM/LRT Kuala Lumpur
```

```r
emptycell = colSums(is.na(rental_df))
print(emptycell)
```

```
##                ads_id         property_type             prop_name
```

```
##                     0                     0                     0
##       completion_year          monthly_rent              district
##                  9185                     2                     0
##               parking              bathroom                  size
##                  5702                     6                     0
##             furnished            facilities additional_facilities
##                     0                     0                     0
##                region
##                     0
```

Dataset is already good as it is. Empty cells in column 'parking' and 'completion_year' will be ignored as it will not be used as much in the upcoming analysis

Data Preparation

Checking for numerical outliers in the dataset; size and monthly rent. Plotting graph with outliers will make our plots skewed.

```
#ggplot(data = rental_df) +
  #geom_point(mapping = aes(x = size, y = monthly_rent, color = property_type)) +
  #labs(title = "Plotted chart with presence of outliers")
```

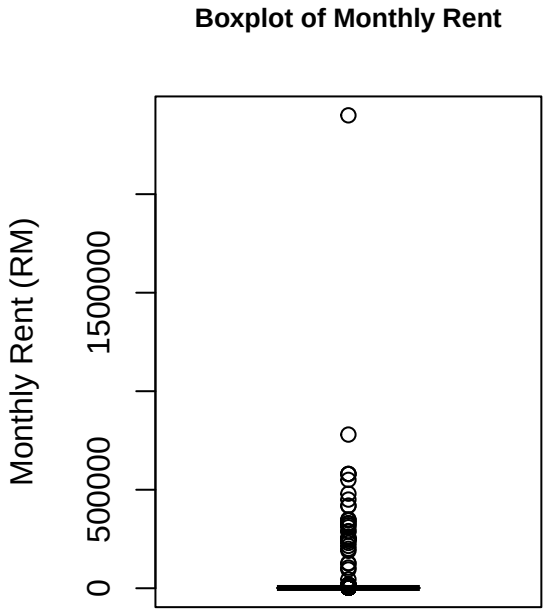
```
rentsummary = summary(rental_df$monthly_rent)
q1 = rentsummary["1st Qu."]
median = rentsummary['Median']
q3 = rentsummary["3rd Qu."]

lower_limit = q1 - 1.5 * IQR(rental_df$monthly_rent, na.rm = TRUE)
upper_limit = q3 + 1.5 * IQR(rental_df$monthly_rent, na.rm = TRUE)


print(lower_limit)
```

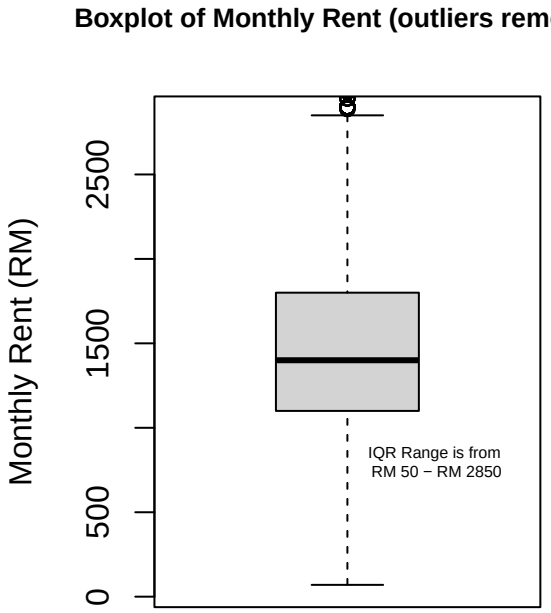
```
## 1st Qu.
##      50
```

```
print(upper_limit)
```

```
## 3rd Qu.
##    2850
```

```
par(mfrow = c(1, 2))
boxplot(rental_df$monthly_rent, main = "Boxplot of Monthly Rent", ylab = "Monthly Rent (RM)", cex.main =
boxplot(rental_df$monthly_rent, main = "Boxplot of Monthly Rent (outliers removed)", ylab = "Monthly Re
text(x = 1.25, y = 800, labels = "IQR Range is from \nRM 50 - RM 2850", cex = 0.5)
```





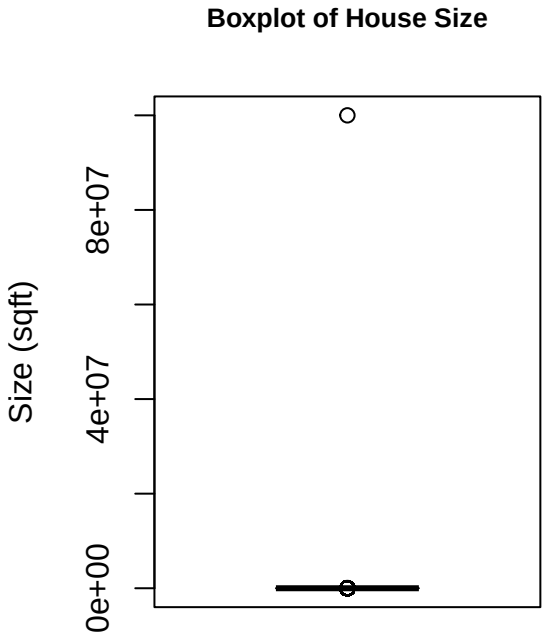

```
sizesummary = summary(rental_df$size)
q1_size = sizesummary["1st Qu."]
median_size = sizesummary['Median']
q3_size = sizesummary["3rd Qu."]

size_lowerlimit = q1 - 1.5 * IQR(rental_df$size, na.rm = TRUE)
size_upperlimit = q3 + 1.5 * IQR(rental_df$size, na.rm = TRUE)


print(size_lowerlimit)
```

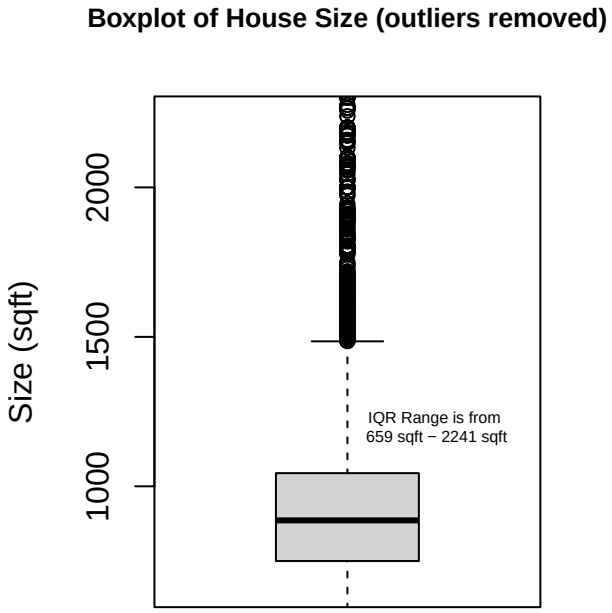
```
## 1st Qu.
##     659
```

```
print(size_upperlimit)
```

```
## 3rd Qu.
##    2241
```

```
par(mfrow = c(1, 2))
boxplot(rental_df$size, main = "Boxplot of House Size", ylab = "Size (sqft)", cex.main = 0.8)
boxplot(rental_df$size, main = "Boxplot of House Size (outliers removed)", ylab = "Size (sqft)", ylim =
text(x = 1.25, y = 1200, labels = "IQR Range is from \n659 sqft - 2241 sqft", cex = 0.5)
```





Creating new dataframe after removing outliers in our numerical column: monthly rent and size.

```
filteredrental = rental_df %>%
  filter(monthly_rent <= 2850 & size <= 2241 & monthly_rent > 100 & size > 100)

dim(filteredrental)
```

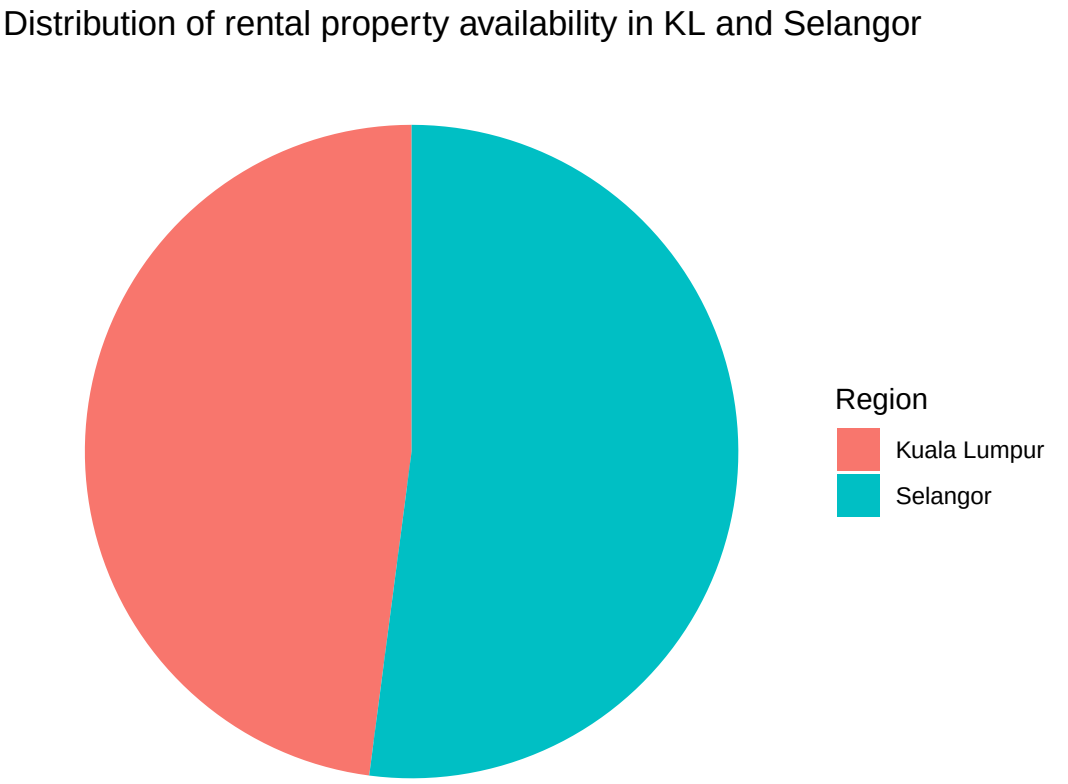
```
## [1] 18626    13
```

```
summary(filteredrental)
```

```
##      ads_id          property_type       prop_name          completion_year
##  Min.   : 16525511   Length:18626       Length:18626       Min.   :1977
##  1st Qu.: 99850470   Class :character   Class :character   1st Qu.:2013
##  Median :100228333   Mode  :character   Mode  :character   Median :2017
##  Mean   : 99723326                                         Mean   :2015
##  3rd Qu.:100627338                                         3rd Qu.:2020
##  Max.   :100854617                                         Max.   :2025
##                                                            NA's   :8693
##   monthly_rent    district           parking          bathroom
##  Min.   : 120   Length:18626       Min.   : 1.000   Min.   :1.000
##  1st Qu.:1100   Class :character   1st Qu.: 1.000   1st Qu.:2.000
##  Median :1400   Mode  :character   Median : 1.000   Median :2.000
##  Mean   :1447                      Mean   : 1.391   Mean   :1.849
##  3rd Qu.:1750                      3rd Qu.: 2.000   3rd Qu.:2.000
```

```
##  Max.   :2850                     Max.   :10.000   Max.   :8.000
##                                   NA's   :5356     NA's   :2
##       size         furnished          facilities        additional_facilities
##  Min.   : 104.0   Length:18626       Length:18626       Length:18626
##  1st Qu.: 737.0   Class :character   Class :character   Class :character
##  Median : 867.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 884.6
##  3rd Qu.:1012.0
##  Max.   :2238.0
##
##     region
##  Length:18626
##  Class :character
##  Mode  :character
##
##
##
##
```

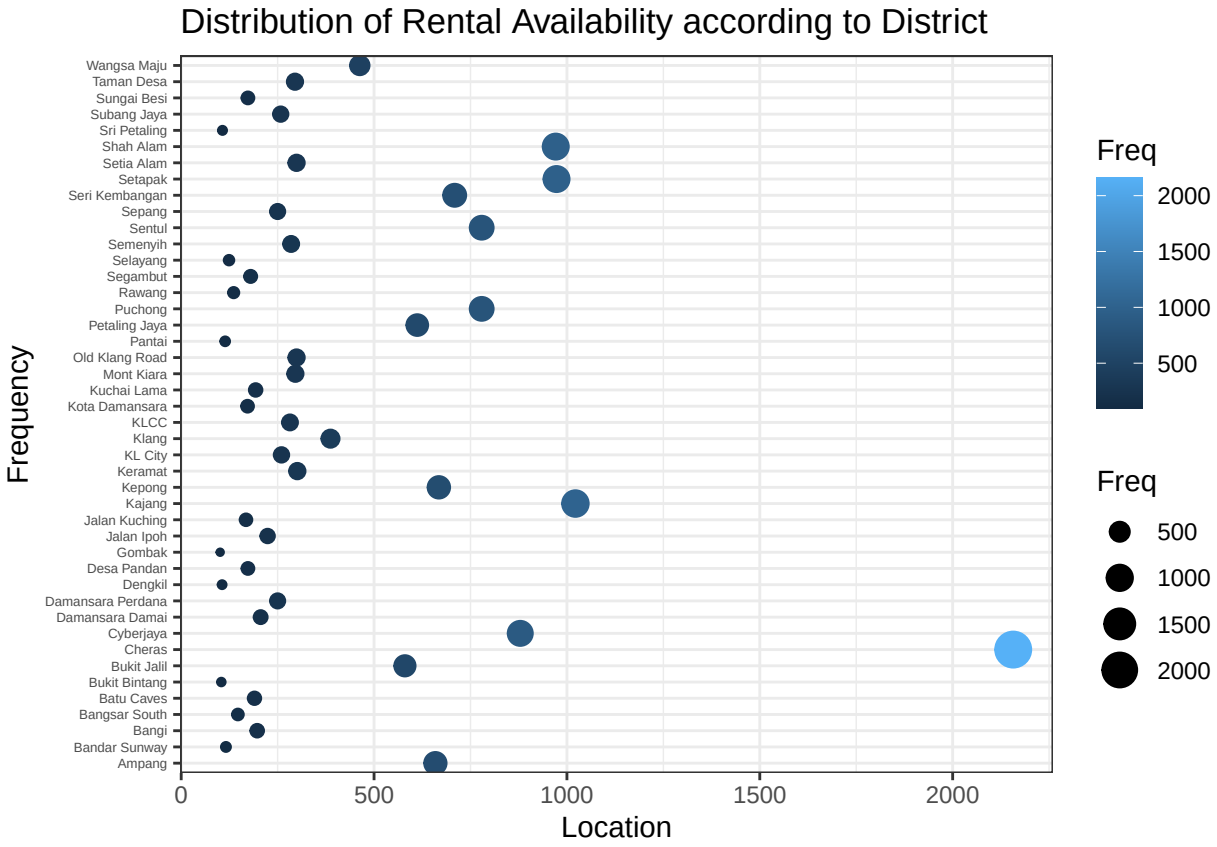
```
ggplot(data = filteredrental, aes(x ="", fill = region)) +
  geom_bar(width = 1) +
  coord_polar("y", start = 0) +
  theme_void() +
  labs(title = "Distribution of rental property availability in KL and Selangor", fill = "Region")
```



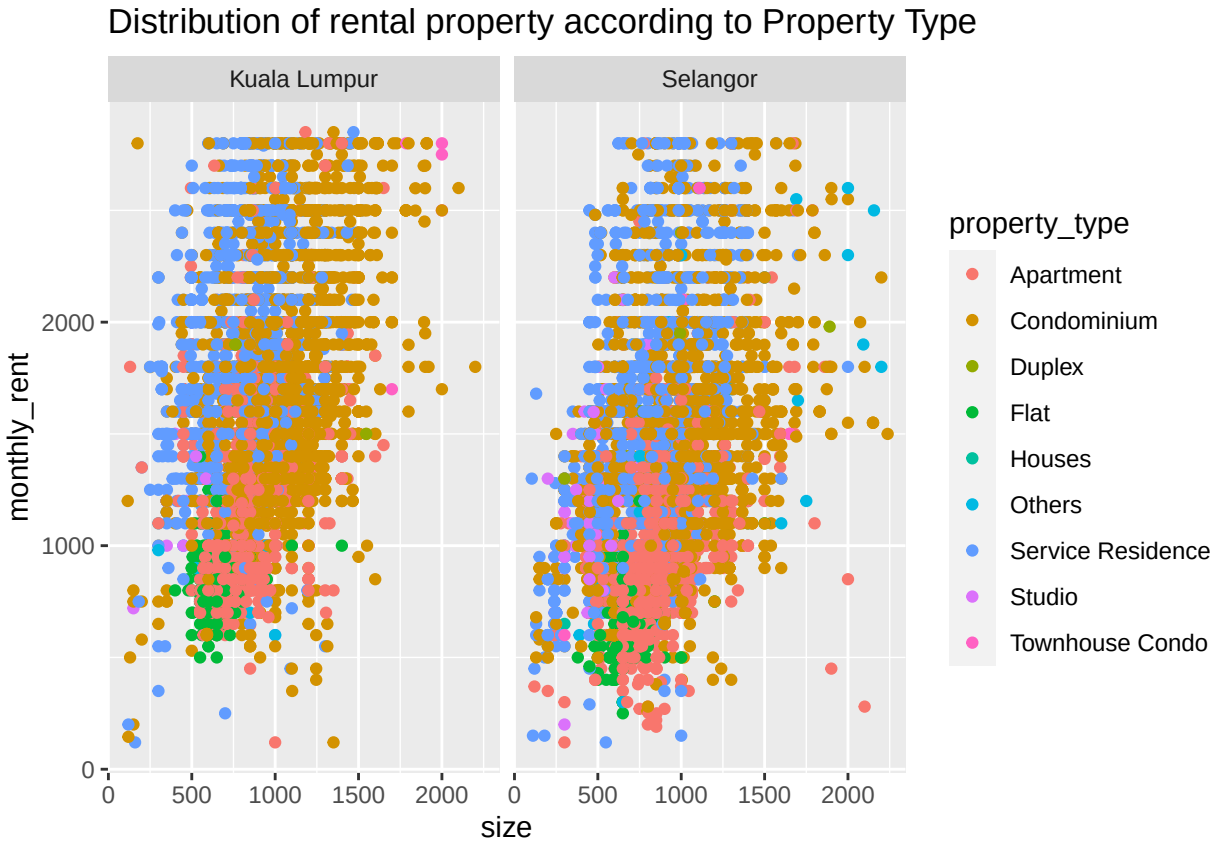
Data Visualization

```
location_count = (table(rental_df$district)) %>%
  as.data.frame() %>%
  arrange(desc(Freq))
```

```
location_count %>%
  filter(Freq >= 100) %>%
  ggplot(aes(x = Freq, y = Var1, size = Freq, color = Freq)) +
  geom_point() +
  labs(title = "Distribution of Rental Availability according to District", x = "Location", y = "Freque
  theme_bw() +
  theme(axis.text.y = element_text(size = 5))
```




```
ggplot (data = filteredrental) +
  geom_point(mapping = aes(x = size, y = monthly_rent, color = property_type)) +
  facet_wrap(~region) +
  labs(title = "Distribution of rental property according to Property Type")
```

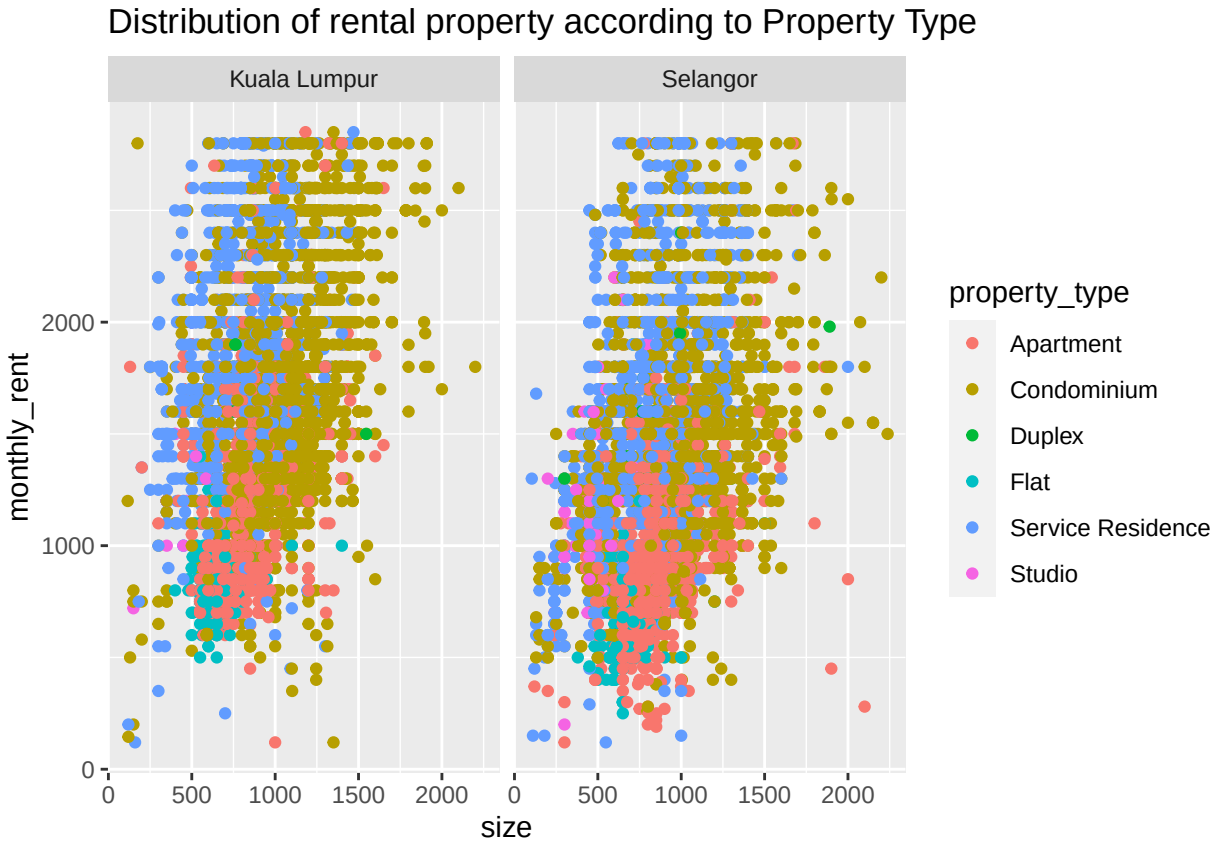


Removing property type with too few data, to visually improving the plot.

```
table(filteredrental$property_type)
```

```
##
##         Apartment       Condominium            Duplex              Flat
##              5194              7681                69               577
##            Houses            Others Service Residence            Studio
##                 1                78              4808               187
##   Townhouse Condo
##                31
```

```
apt_rental = filteredrental %>%
  filter(property_type %in%  c("Apartment", "Condominium", "Service Residence", "Flat", "Studio", "Dupl

ggplot (data = apt_rental) +
  geom_point(mapping = aes(x = size, y = monthly_rent, color = property_type)) +
  facet_wrap(~region) +
  labs(title = "Distribution of rental property according to Property Type")
```

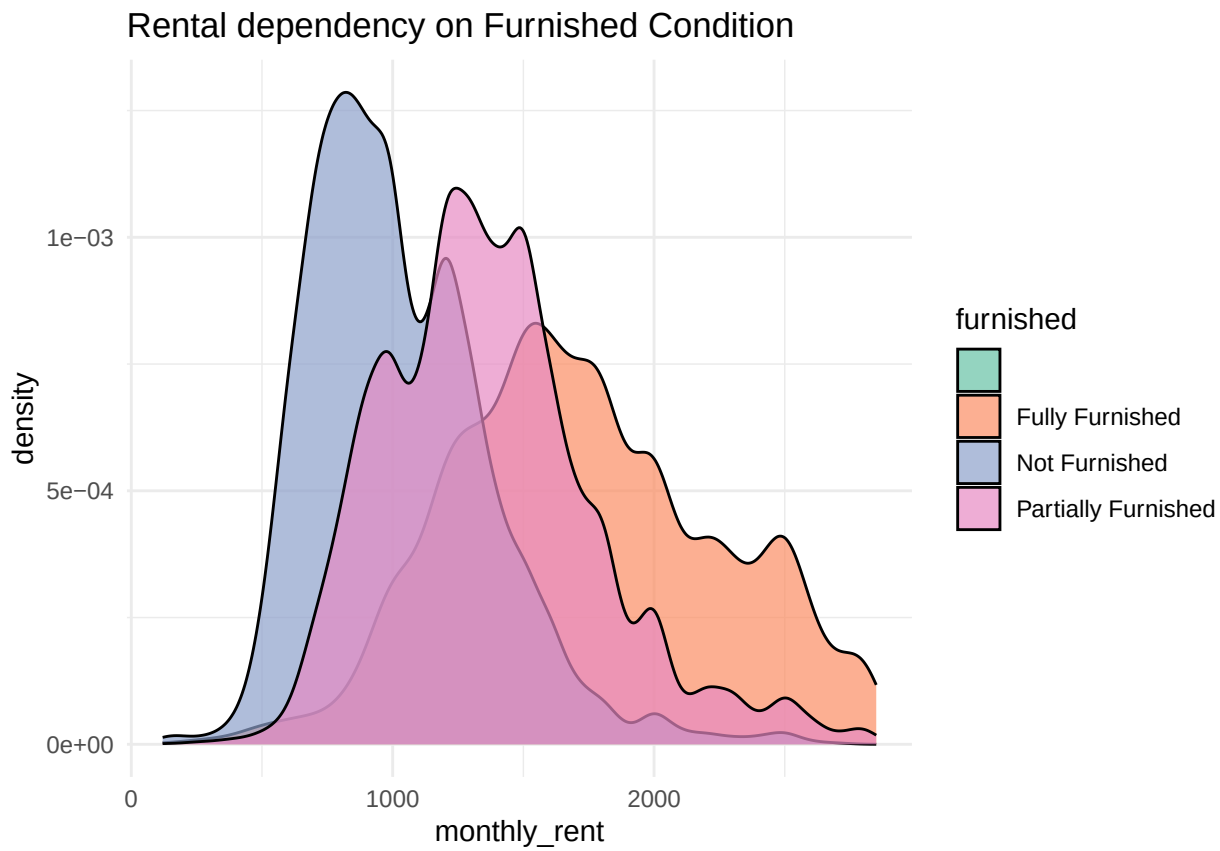


Analysing rental dependency on other variable

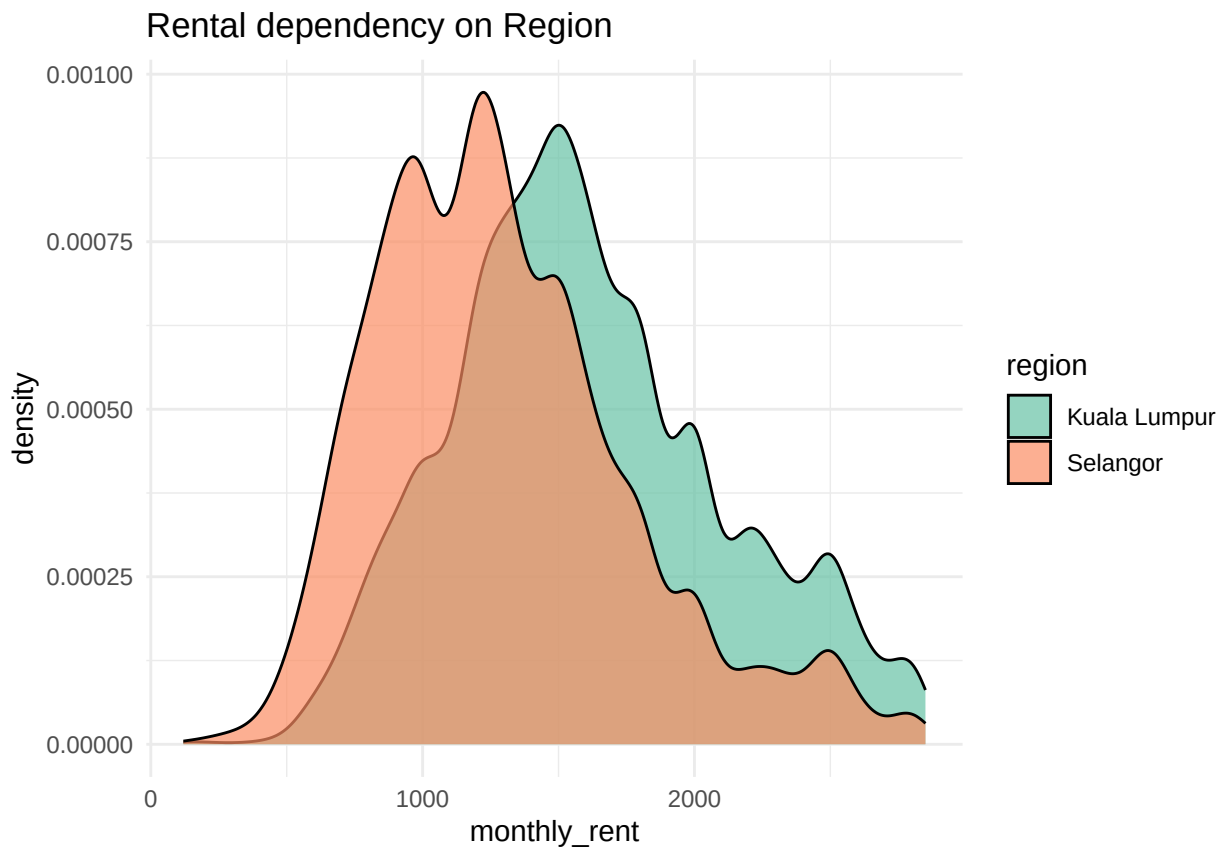
```
ggplot(filteredrental, aes(x = monthly_rent, fill = furnished)) +
  geom_density(alpha = 0.7) +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Rental dependency on Furnished Condition") +
  theme_minimal()
```

```
## Warning: Groups with fewer than two data points have been dropped.
```

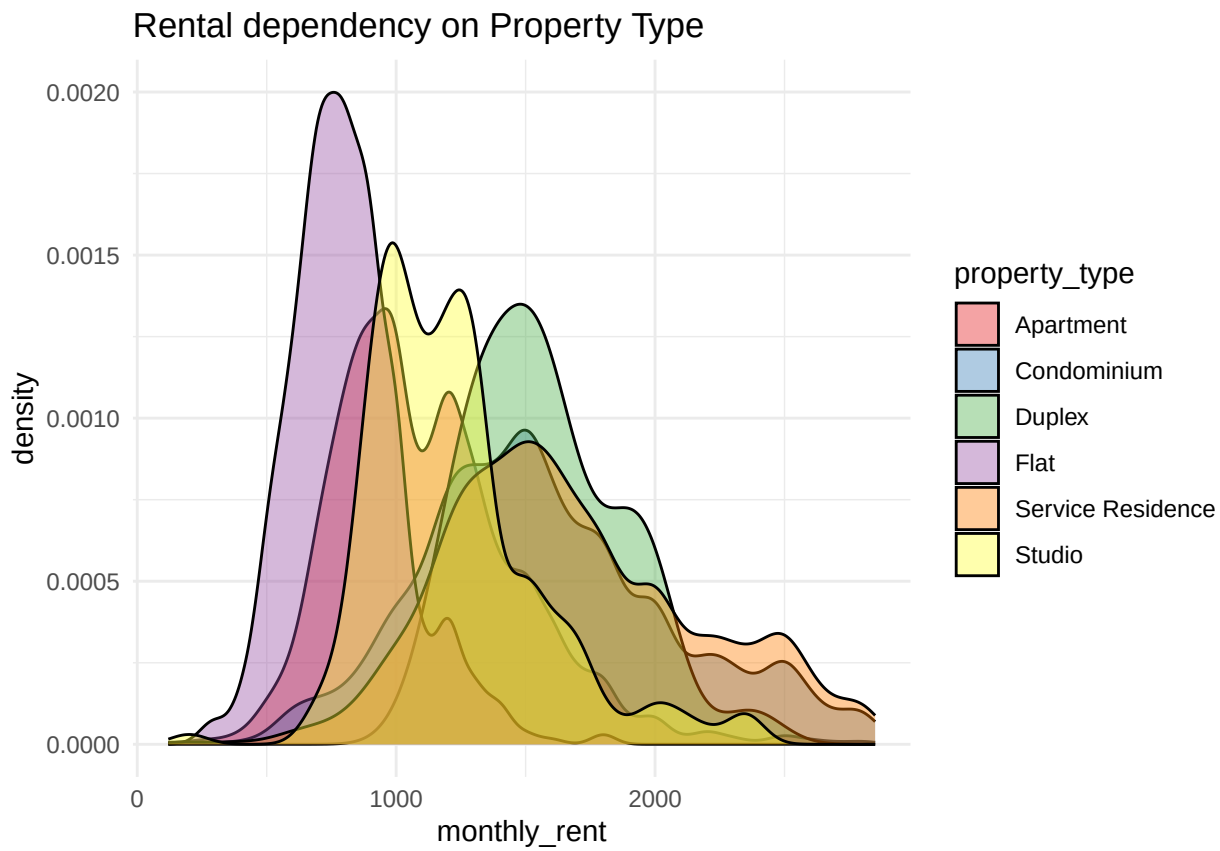
```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```




```
ggplot(filteredrental, aes(x = monthly_rent, fill = region)) +
  geom_density(alpha = 0.7) +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Rental dependency on Region") +
  theme_minimal()
```



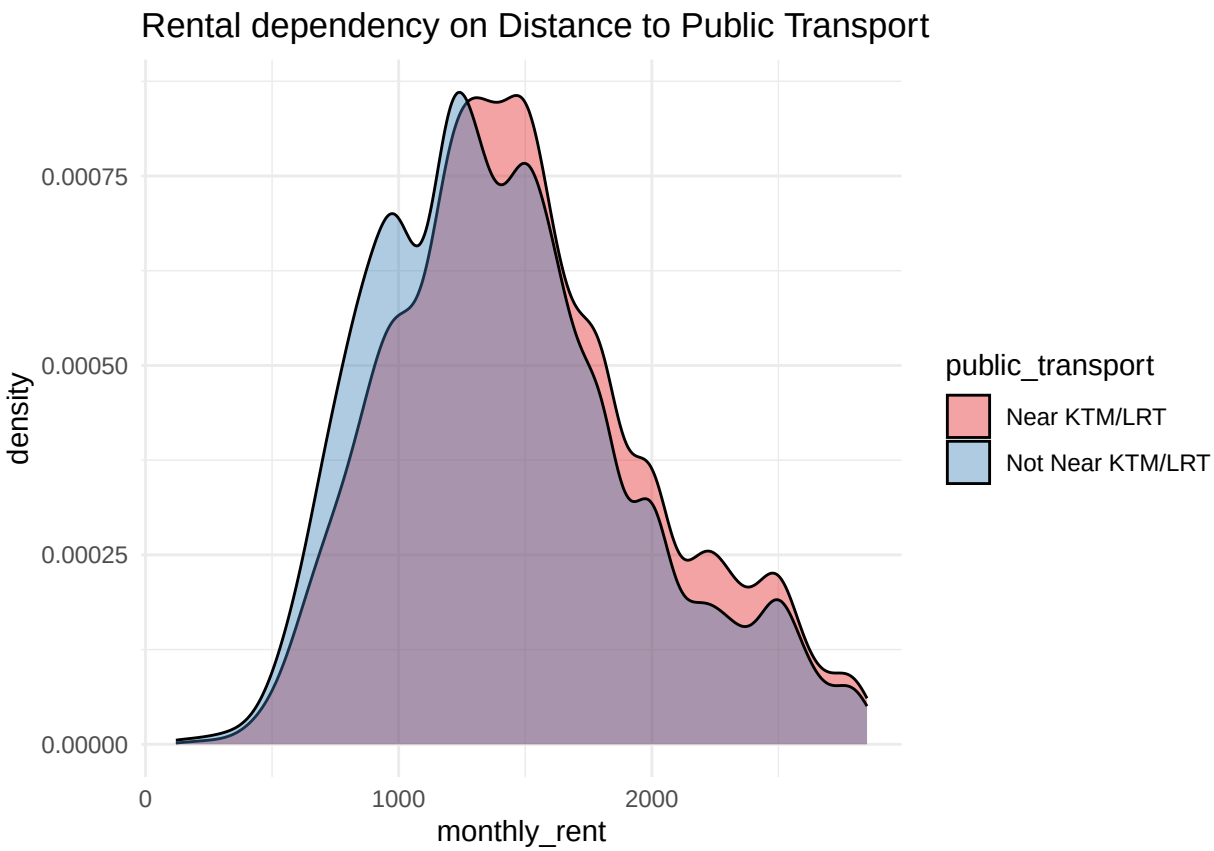

```
ggplot(apt_rental, aes(x = monthly_rent, fill = property_type)) +
  geom_density(alpha = 0.4) +
  scale_fill_brewer(palette = "Set1") +
  labs(title = "Rental dependency on Property Type") +
  theme_minimal()
```



Extracting "Near KTM/LRT" information into a new column, added onto dataframe.

```
rental_publictransport = filteredrental %>%
  mutate(
    public_transport = str_extract(additional_facilities, "Near KTM/LRT"),
    public_transport = replace(public_transport, is.na(public_transport), "Not Near KTM/LRT"))
```

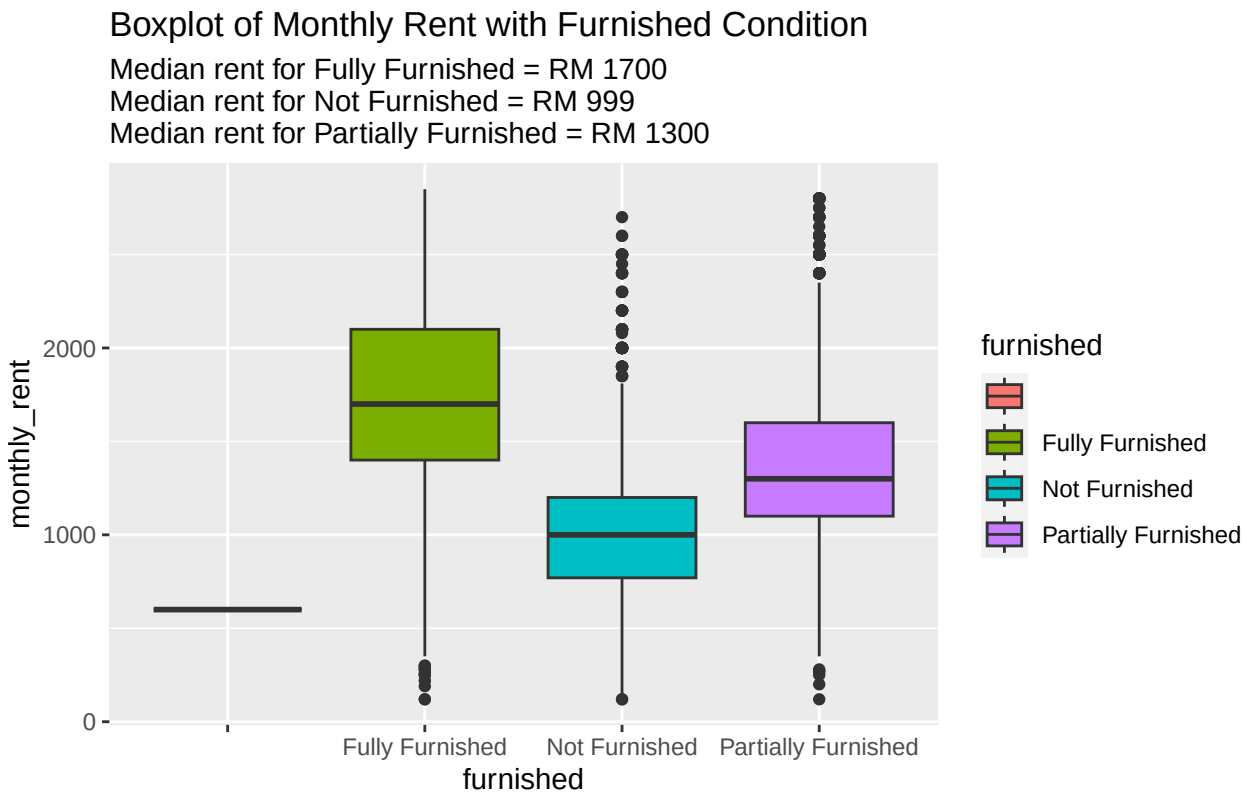
```
ggplot(rental_publictransport, aes(x = monthly_rent, fill = public_transport)) +
  geom_density(alpha = 0.4) +
  scale_fill_brewer(palette = "Set1") +
  labs(title = "Rental dependency on Distance to Public Transport") +
  theme_minimal()
```



These graphs may conclude that: 1) Location and Distance to Public Transport do not have major affect on rental distribution. 2) Furnished condition and Property type displays significant differences on rental distribution.

Further analysis to be done to gain insights on rental against furnished condition and property type.

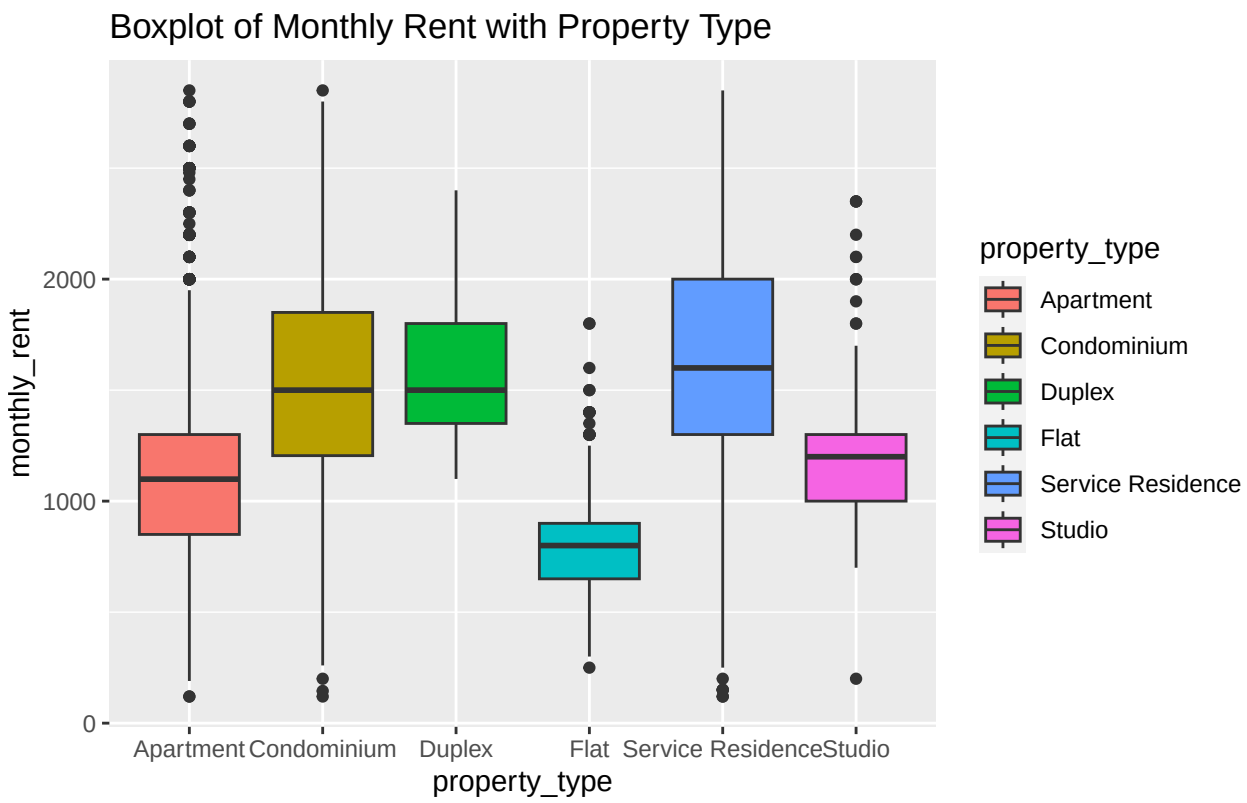
```
filteredrental %>%
  select(furnished, monthly_rent) %>%
  ggplot(aes(x = furnished, y = monthly_rent, fill = furnished)) +
  geom_boxplot() +
  labs(title = "Boxplot of Monthly Rent with Furnished Condition",
       subtitle = "Median rent for Fully Furnished = RM 1700 \nMedian rent for Not Furnished = RM 999 \n
  theme(plot.margin = margin(b=30))
```




```
median_data = filteredrental %>%
  group_by(furnished) %>%
  summarize(median_rent = median(monthly_rent))
print(median_data)
```

```
## # A tibble: 4 x 2
##   furnished             median_rent
##   <chr>                       <dbl>
## 1 ""                            600
## 2 "Fully Furnished"            1700
## 3 "Not Furnished"              1000
## 4 "Partially Furnished"        1300
```

```
apt_rental %>%
  select(property_type, monthly_rent) %>%
  ggplot(aes(x = property_type, y = monthly_rent, fill = property_type)) +
  geom_boxplot() +
  labs(title = "Boxplot of Monthly Rent with Property Type") +
  theme(plot.margin = margin(b=30))
```




```
propertymedian_data = apt_rental %>%
  group_by(property_type) %>%
  summarize(propertymedian_rent = median(monthly_rent))
print(propertymedian_data)
```

```
## # A tibble: 6 x 2
##   property_type     propertymedian_rent
##   <chr>                           <dbl>
## 1 Apartment                        1099
## 2 Condominium                      1500
## 3 Duplex                           1500
## 4 Flat                              800
## 5 Service Residence                1600
## 6 Studio                           1200
```

Summary

As a potential tenant, I'd always consider an affordable rent of below RM 1.5K as my budget, before deciding on a house. From this analysis, RM1.5K per month offers a lot of potential unit such as: 1) Apartment, flat, and condominium shows a median of monthyl rent close to RM 1.5k. However, service residence can be considered out of the budget. 2)Not furnished and partially furnished rental are still within RM 1.5k budget, however a fully furnished unit median is around RM1.7K. 3) Monthyl rental for units in Kl are slighly higher that units in Selangor. 4) Vicinity of public transport, whether it is near KTM/LRT or not, does not significantly affect monthly rental as much.