QMB Exercise 1 - Exploring Housing Rents

Dino Nienhold

Wednesday, February 18, 2015

Introduction

The following report is based on the QMB Exercise 1 - Exploring Housing Rents. The task description pdf file is bis_ex1-HousingRents.pdf

Requirements

Please make sure that you the following packages loaded in your workspace.

```
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("ggplot2")
```

Data Set

Please make sure you have the file housingrents.csv in the subdirectoy Data in your workspace.

```
housingrents <- read.csv("./Data/housingrents.csv",sep=";")
```

Task 1

There are `dim(housingrents)[1]` observations and `dim(housingrents)[2]` variables in the dataset. The str command gives an overview of the variable types:

```
str(housingrents)
```

```
## 'data.frame':    152 obs. of  7 variables:
##  $ id     : int  1 2 3 4 6 7 8 10 11 13 ...
##  $ rooms  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ area   : int  34 35 50 45 35 40 43 45 37 60 ...
##  $ rent   : int  310 749 281 483 515 530 480 560 580 510 ...
##  $ nre    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ econage: int  24 40 34 30 27 31 30 28 50 42 ...
##  $ balcony: Factor w/ 2 levels "no","yes": 2 1 2 1 1 NA 1 1 1 1 ...
```

There are 14 NA values in the balcony variable.

```
summary(housingrents)
```

```
##        id            rooms           area            rent
##  Min.   :  1.00   Min.   :1.000   Min.   : 18.00   Min.   : 250.0
##  1st Qu.: 38.75   1st Qu.:2.000   1st Qu.: 60.00   1st Qu.: 793.8
##  Median : 76.50   Median :3.000   Median : 83.00   Median :1046.0
##  Mean   : 76.50   Mean   :3.171   Mean   : 86.84   Mean   :1240.3
##  3rd Qu.:114.25   3rd Qu.:4.000   3rd Qu.:105.00   3rd Qu.:1552.8
##  Max.   :152.00   Max.   :6.000   Max.   :250.00   Max.   :4725.0
##       nre            econage      balcony
##  Min.   :0.0000   Min.   : 0.00   no  :61
##  1st Qu.:0.0000   1st Qu.:22.00   yes :77
##  Median :0.0000   Median :31.00   NA's:14
##  Mean   :0.3355   Mean   :30.18
##  3rd Qu.:1.0000   3rd Qu.:39.00
##  Max.   :1.0000   Max.   :60.00
```

Task 2

Data Processing

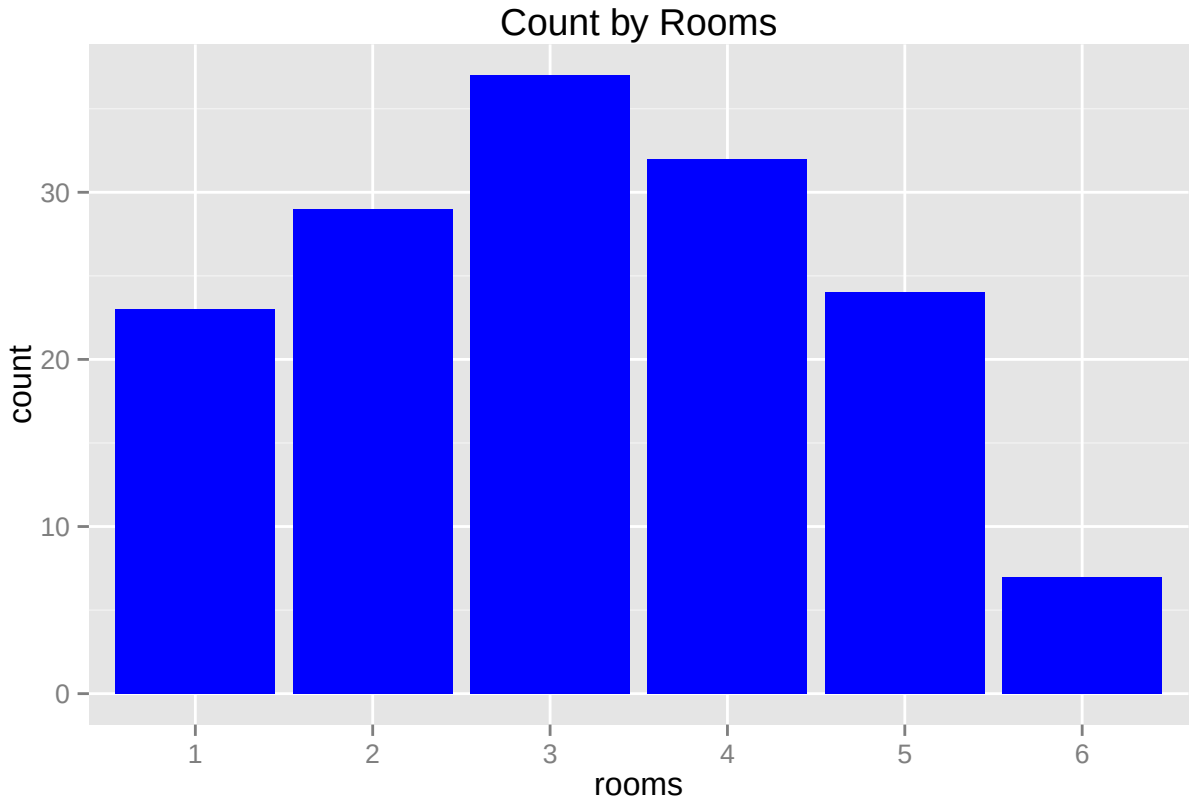
For analysis purposes it is necessary to convert the rooms and nre variable to a factor.

```
housingrents <- mutate(housingrents, rooms = factor(rooms), nre = factor(nre,levels=c(0,1),labels=c("no
```

Plotting

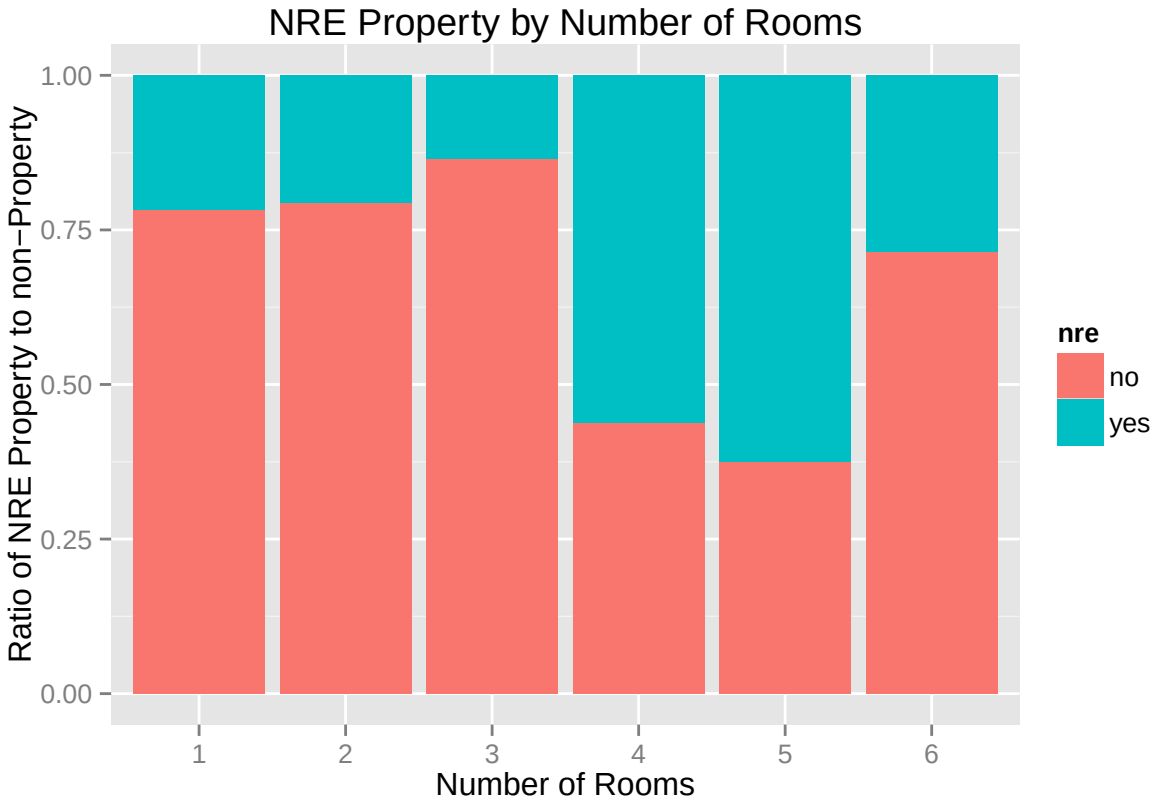
The following chart shows the frequency of appartments according to their numberof rooms.

```
ggplot(data=housingrents, aes(x=rooms,label=rooms))  + geom_bar(fill="blue") + ggtitle("Count by Rooms"
```



Task 3

```
rooms2nre <- xtabs(~rooms+nre, data=housingrents)
rooms2nre <- prop.table(rooms2nre,1)
ggplot(data.frame(rooms2nre), aes(x=rooms, y=Freq, fill=nre)) + geom_bar(stat="identity") +
  xlab("Number of Rooms") + ylab("Ratio of NRE Property to non-Property") +
  ggtitle("NRE Property by Number of Rooms")
```




```
addmargins(prop.table(rooms2nre,1))
```

```
##      nre
## rooms        no       yes       Sum
##   1   0.7826087 0.2173913 1.0000000
##   2   0.7931034 0.2068966 1.0000000
##   3   0.8648649 0.1351351 1.0000000
##   4   0.4375000 0.5625000 1.0000000
##   5   0.3750000 0.6250000 1.0000000
##   6   0.7142857 0.2857143 1.0000000
##   Sum 3.9673627 2.0326373 6.0000000
```