Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

**SHORT REPORT**

# **Prologue**

# First, I must stress that the work was an excellent opportunity to combat my natural tendency towards a linear programming philosophy using loops and logic controls. The problem is simple and yet an excellent opportunity for getting your feet wet in a new language.

# **Ι**

Load the file using the command read.csv() .

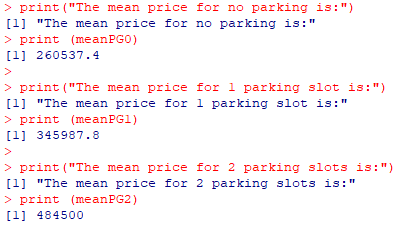
# **ΙΙ**

Using table() selecting the garage column to count how many instances there are. With the following results:



# **ΙIΙ**

I create a list named pg which contains the totalprice and garage columns, then I create 3 vectors pg0,pg1,pg2 that each contain the prices of the houses that have 0,1,2 garages and I use mean() to get the average values. With the following results:



# **ΙV**

Here I completely forgot about the existence of frq() from sjmisc. So I ended up creating it from scratch by computing 4 non-zero integers. Houses with 0 garages and 0 lifts, 0 garages and 1 lifts, 1 garage and 1 lifts, 2 garages and 1 lifts. Making extensive use of length(which()).

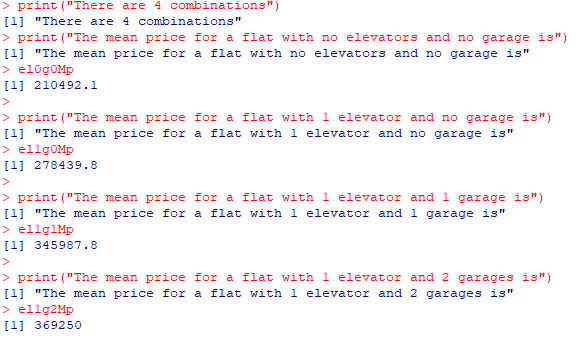
Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

For the frequencies I simply divided by the sum.

# **V**

Being more familiar with this query I simply created a list for each of the above 4 cases with 3 columns, this time including the price and used mean(el1g1p[,"totalprice"]) to get the average of column. With the following results:



# **VΙ**

I don't consider the need for an extensive explanation. It simply selects where the value in the garage column is greater than or equal to 1. And then displays the results for review.

# **VΙΙ**

Εικόνα που περιέχει πίνακας

Περιγραφή που δημιουργήθηκε αυτόματα

Taking all the elements with 3BA we notice that they all have an elevator already. Therefore, it is unnecessary to carry out filtering. We can either count them since they are so few, however alternatively for larger sizes lengths(data.c) can be used which gives a result of 10. Also the file can be saved for future use although this would be more important for more complex answers query calculations with the command:



# **VΙII**

We can access rds files with the command:



From the lists I retrieved the average values by finding the mean of the columns for price and squared with the following results.

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

# The price and square footage of category 3B are significantly higher than the average.

# **ΙΧ**

For this query I created list with only the out column which I split into three vectors one for each case, whose length I calculated with length().

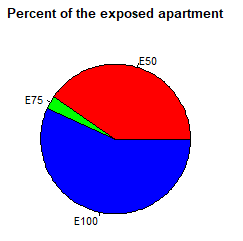
Εικόνα που περιέχει κείμενο, πίνακας

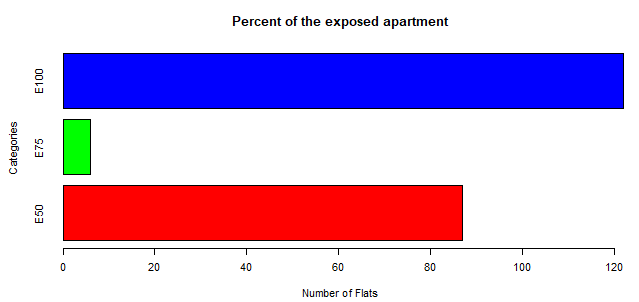
Περιγραφή που δημιουργήθηκε αυτόματα

In order to fit all the vectors of the work, I used par() in retrospect.



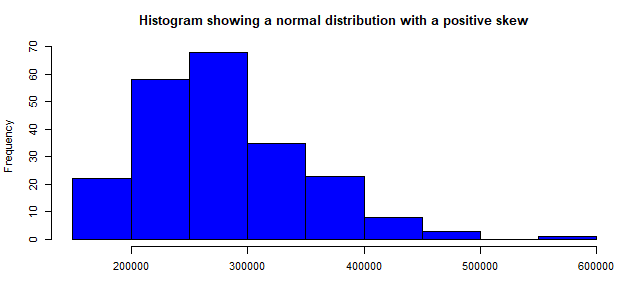
And then I used Pie() and barplot() .





We can say that the histogram renders the data with a higher degree of clarity.

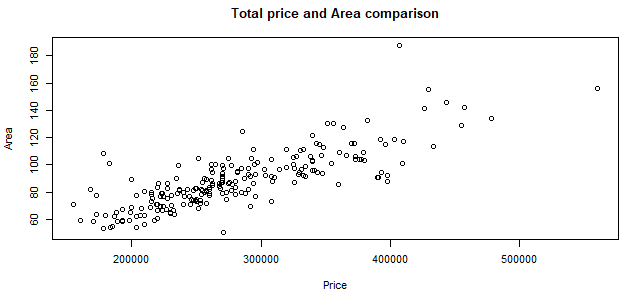
# **Χ**



Using hist(), whose output shows that the distribution of values is normal with positive/right symmetry. Also, it would not be completely unreasonable to characterize buildings with a price of more than €550,000 as extreme prices.

# **ΧΙ**

In order to be able to confidently characterize the price relationship with quadratics I created a scatter plot and calculated the correlation.

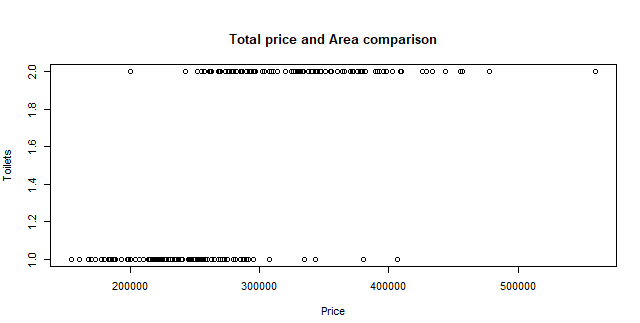


Εικόνα που περιέχει κείμενο

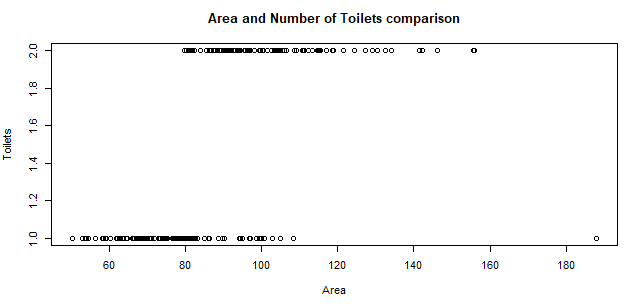
Περιγραφή που δημιουργήθηκε αυτόματα

Their relationship is one of dependence. In fact, they have a strong positive correlation, which is, by the way, expected.

# **ΧΙΙ**

Accordingly, in order to confidently characterize the relationship between price and square footage and the number of bathrooms, I created a scatter diagram and calculated the correlation.







There is a strong positive relationship in both 2 and more between price and bathrooms but weaker than that between square footage and price. However, it is logical that due to the previous question, it was obvious that from the moment either the price or the squares presented a dependence, the other would follow.

# **ΧΙΙΙ**

Here, based on the valuable information provided by the diagrams, it is enough to calculate the price of the most expensive house with one toilet and the cheapest one with 2. This was achieved by creating 2 lists one for each category and sorting them. The range was:

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

# **ΧIV**

It is enough to compare the average price of apartments with 1 toilet with those that have 2.



I also wanted to calculate how much should be given for the acquisition of another toilet, what percentage is on the average price of houses with one toilet.





Roughly speaking, we are talking about an increase of one third of the purchase amount.

# **ΧV**

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

You call the function with inputs a data frame and a vector. The content of the vector is passed as character type to 2 variables that feed the order function in order first sorting by values and then by squares.

**I wish you a productive academic year!**