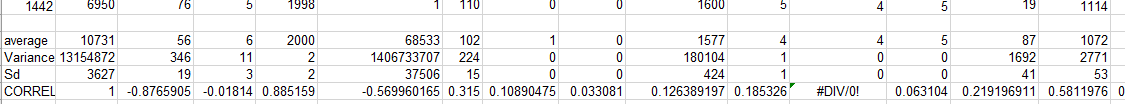
**Problem 2.11.**

**a.**

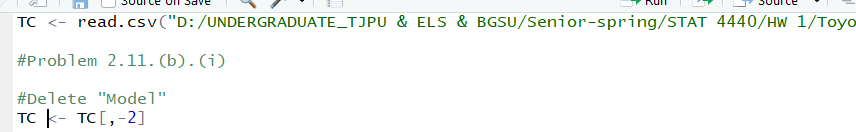


**The price of car is negatively correlated with the age which is the second column (Age\_08\_04). So as the year of car (Mfg\_Year). Lager the age, lower the price.**

b.

(i)

**Delete the column “Model”**



**We decide to use numbers to instead of words, like**

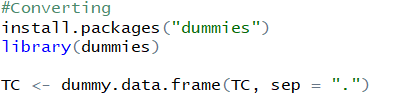
**For *Fuel Type***

|  |  |  |  |
| --- | --- | --- | --- |
|  | **D1** | **D2** | **D3** |
| **CNG** | **1** | **0** | **0** |
| **Diesel** | **0** | **1** | **0** |
| **Petrol** | **0** | **0** | **1** |

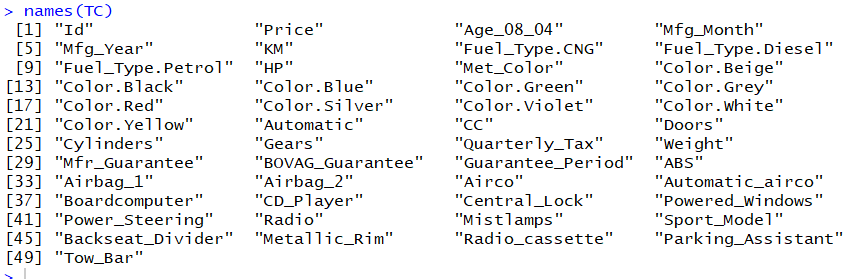
**For color**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 |
| Beige | **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Black | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Blue | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Green | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Grey | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Red | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Silver | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Violet | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| White | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Yellow | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

**Using the dummies package to transform categorical data to dummies.**



Result would be like

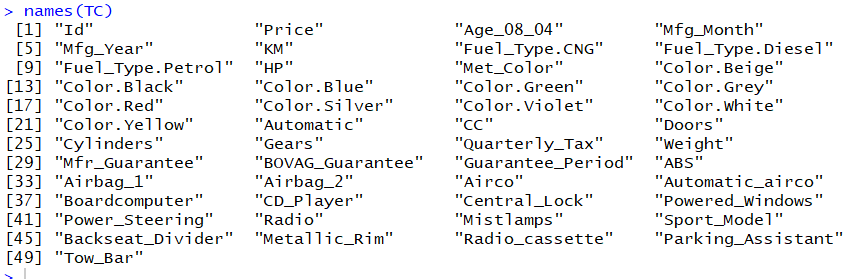


(ii)

(We still don’t understand the meaning yet)

**Problem 4.3.**

1. **All categorical variables**



**b.**

**Same with the part in 2.11.b.i**

**The series of binary dummy are decided by categorical variable as 1 or 0.**

**c.**

**N – 1, dummy binary variables are required to capture the information in a categorical variable with n categories**

**d.**

**Same with the part in 2.11.b.i.**

**In our work, when there have n categories in a variable (“color” category has 10 different colors as variables), we divide it to n dummy variables (we haven’t use all 0s as dummies here), and we put 0 to all of the spaces except the one the category belongs to.**

**e.**

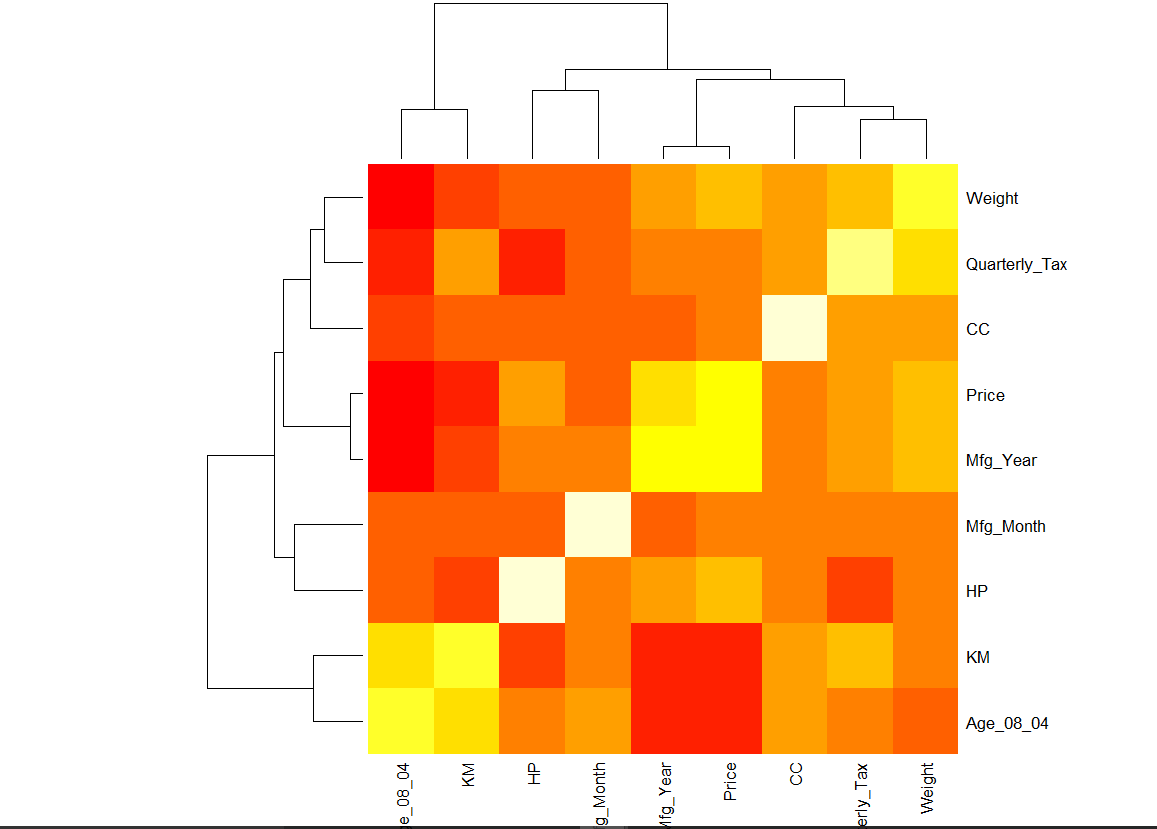
**Correlation matrix and matrix plot.**

**A heatmap of the result**

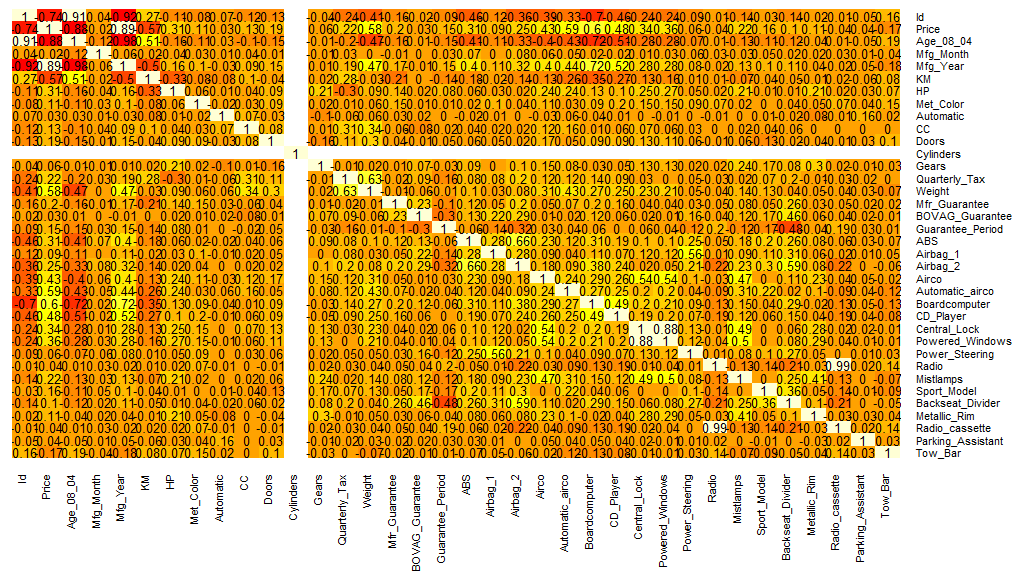
1. Simple heatmap

**For this part, we choose the categories that we think it might useful for the plot.  
(Price, Age\_08\_04, Mfg\_Month, Mfg\_Year, KM, HP, CC, Quarterly\_Tax , Weight)**

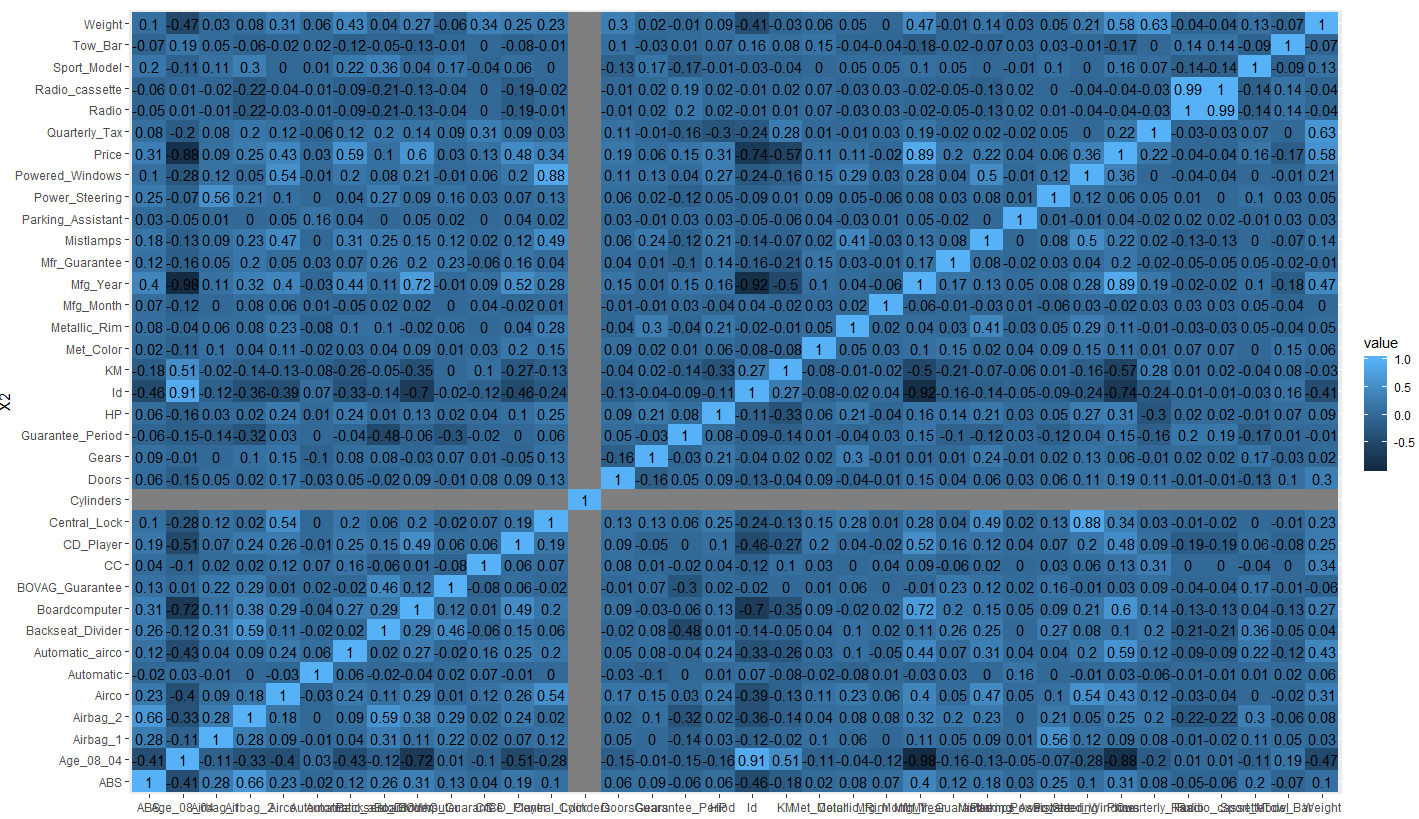
**Because the system said the Stdev is 0.**



2. heatmap with values



1. alternative plot with ggplot



**Matrix scatterplot.**

**For this part, we choose the categories that we think it might useful for the plot.  
(Price, Age\_08\_04, Mfg\_Month, Mfg\_Year, KM, HP, CC, Quarterly\_Tax , Weight)**

