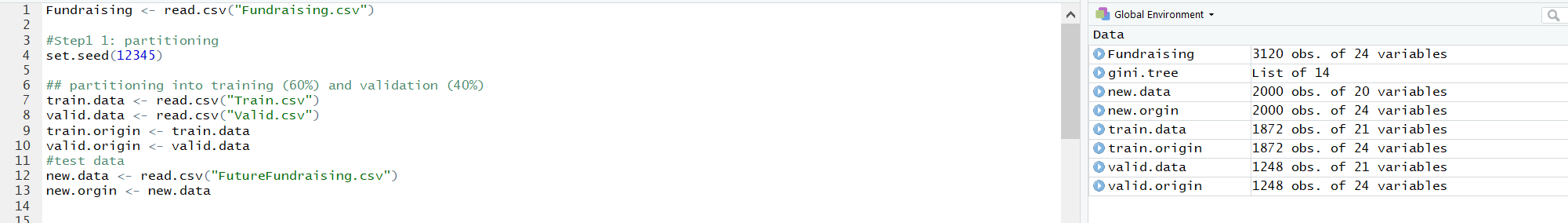
**Ch21.7**

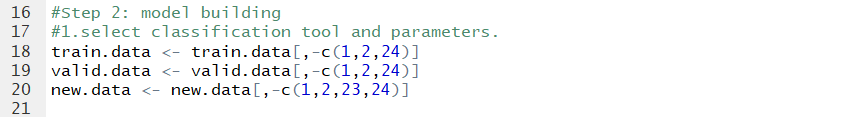
**Assignment**

**Step 1: partitioning**

We made partition by hand. There are 1872 records (936 1s, 936 0s) in Train and 1248 records (624 1s, 624 0s) in Valid.



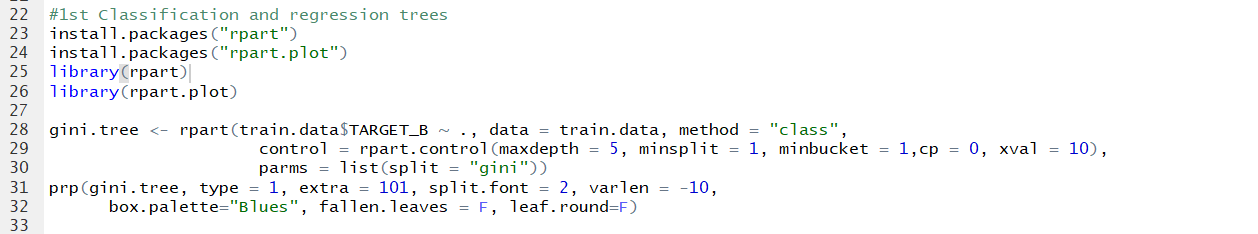
**Step 2: model building**

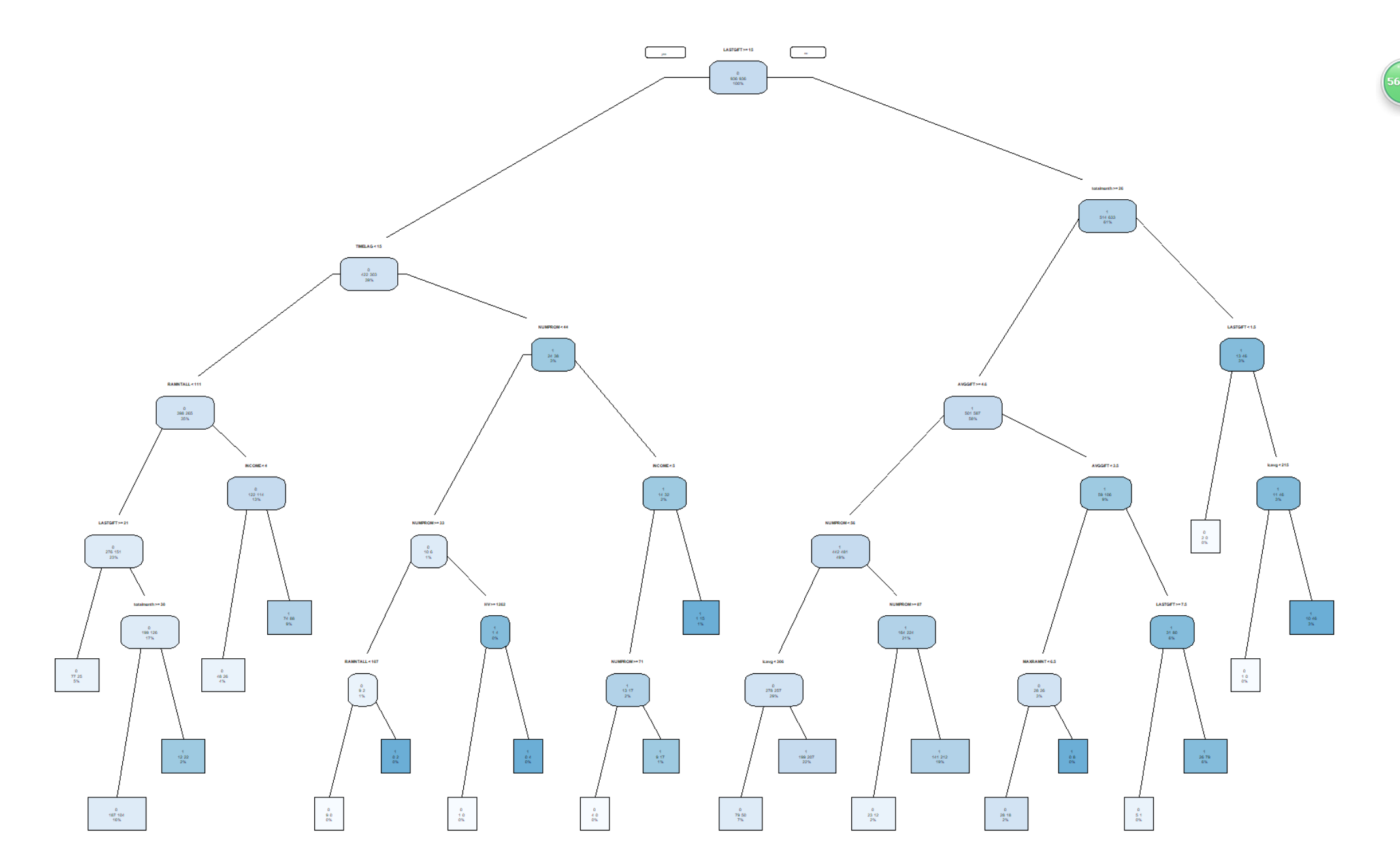


**#1. Select classification tool and parameters.**

May be we need to do PCA or DA firstly for reducing the dimensions cause troubles in the 2nd and 3rd method?

**1st Classification and regression trees**





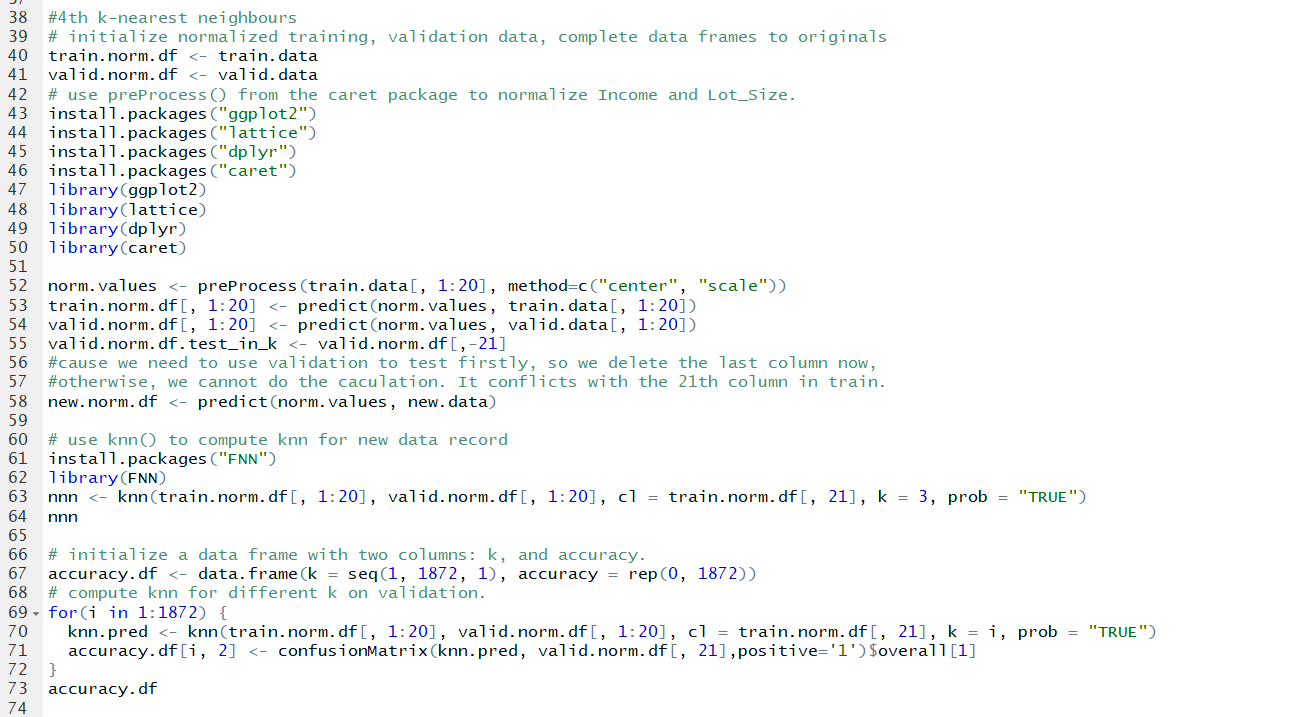
**2nd Logistic Regression**

We need to select variables from the 20 variables and then, test them one by one. That needs a lot of works. So, pass this.

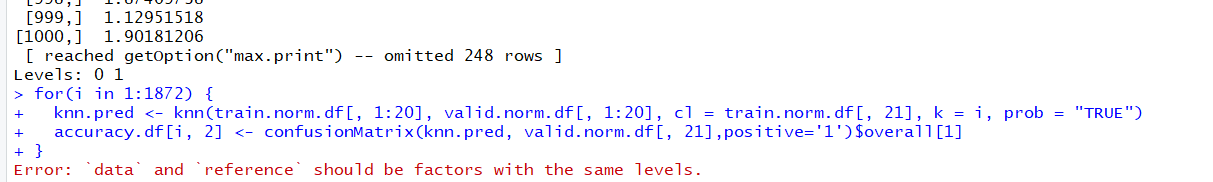
**3rd Naive Bayes**

The reason is the same one.

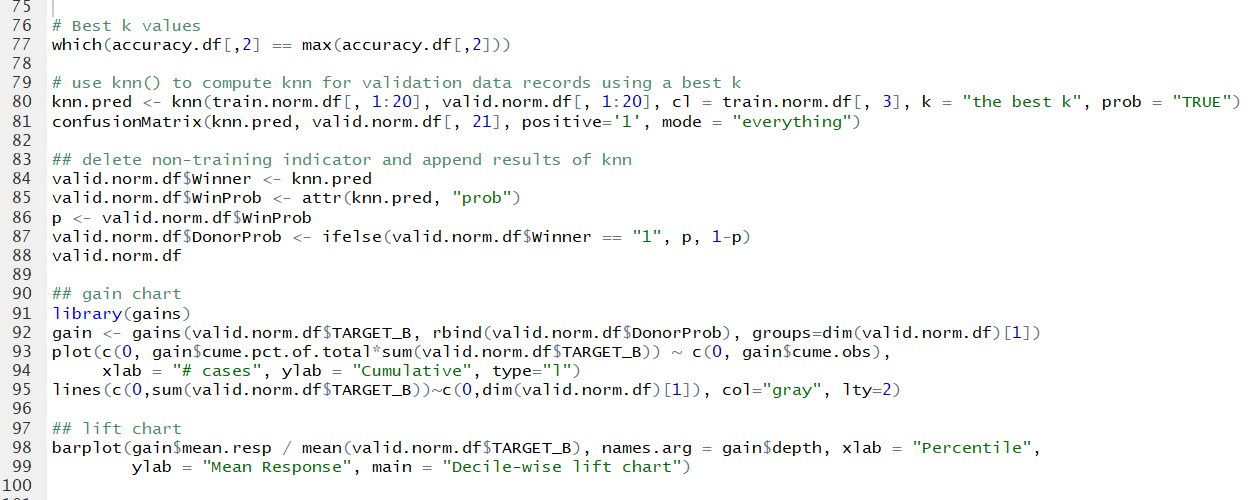
**4th k-nearest neighbours**



We met a problem at the end of the last graph.



The “knn.pred” and “valid.norm.df [,21]” are both 1248 rows, I don’t know why does the software told me that. Anyway, the below is the following steps even we cannot run it because of the problem.



Based on these steps, we could create a matrix, do adjustment, and calculate net profit.

**#2 Classification under asymmetric response and cost.**

**#3 Calculate net profit**

**#4 Draw lift curves.**

**#5 Select best model.**

**Step 3: Testing**

**#6**