project

Liu Yuzhou / Lu Zhoudao

2023-11-27

As students from the Department of Computing, we are more familiar with data mining and machine learning. Machine learning need large-scale data, and its inherent difficulty of interpretation prevent it from being a complete replacement for the linear models. Linear models still have a high status in many basic tasks.

Customer Personality Analysis

Customer Personality Analysis is a detailed analysis of a company's ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers.

Customer personality analysis helps a business to modify its product based on its target customers from different types of customer segments. For example, instead of spending money to market a new product to every customer in the company's database, a company can analyze which customer segment is most likely to buy the product and then market the product only on that particular segment.(From Kaggle)

We get the dataset from Kaggle, the link is here: https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis/data

We also do the data cleaning. because some data is missed, we just delete those rows. We also attached our final excel.

```
data<-read.csv("marketing.csv",header = TRUE)
head(data)
```

```
##     ID Year_Birth  Education Marital_Status Income Kidhome Teenhome Dt_Customer
## 1 5524       1957 Graduation         Single  58138       0        0  04-09-2012
## 2 2174       1954 Graduation         Single  46344       1        1  08-03-2014
## 3 4141       1965 Graduation       Together  71613       0        0  21-08-2013
## 4 6182       1984 Graduation       Together  26646       1        0  10-02-2014
## 5 5324       1981        PhD        Married  58293       1        0  19-01-2014
## 6 7446       1967     Master       Together  62513       0        1  09-09-2013
##   Recency MntWines MntFruits MntMeatProducts MntFishProducts MntSweetProducts
## 1      58      635        88             546             172               88
## 2      38       11         1               6               2                1
## 3      26      426        49             127             111               21
## 4      26       11         4              20              10                3
## 5      94      173        43             118              46               27
## 6      16      520        42              98               0               42
##   MntGoldProds NumDealsPurchases NumWebPurchases NumCatalogPurchases
## 1           88                 3               8                  10
## 2            6                 2               1                   1
```

```
## 3           42                 1               8                   2
## 4            5                 2               2                   0
## 5           15                 5               5                   3
## 6           14                 2               6                   4
##   NumStorePurchases NumWebVisitsMonth AcceptedCmp3 AcceptedCmp4 AcceptedCmp5
## 1                 4                 7            0            0            0
## 2                 2                 5            0            0            0
## 3                10                 4            0            0            0
## 4                 4                 6            0            0            0
## 5                 6                 5            0            0            0
## 6                10                 6            0            0            0
##   AcceptedCmp1 AcceptedCmp2 Complain Z_CostContact Z_Revenue Response
## 1            0            0        0             3        11        1
## 2            0            0        0             3        11        0
## 3            0            0        0             3        11        0
## 4            0            0        0             3        11        0
## 5            0            0        0             3        11        0
## 6            0            0        0             3        11        0
```

```r
attach(data)
```

```r
summary(data)
```

```
##        ID          Year_Birth    Education         Marital_Status    
##  Min.   :    0   Min.   :1893   Length:2216        Length:2216       
##  1st Qu.: 2815   1st Qu.:1959   Class :character   Class :character  
##  Median : 5458   Median :1970   Mode  :character   Mode  :character  
##  Mean   : 5588   Mean   :1969                                        
##  3rd Qu.: 8422   3rd Qu.:1977                                        
##  Max.   :11191   Max.   :1996                                        
##      Income          Kidhome          Teenhome      Dt_Customer       
##  Min.   :  1730   Min.   :0.0000   Min.   :0.0000   Length:2216       
##  1st Qu.: 35303   1st Qu.:0.0000   1st Qu.:0.0000   Class :character  
##  Median : 51382   Median :0.0000   Median :0.0000   Mode  :character  
##  Mean   : 52247   Mean   :0.4418   Mean   :0.5054                     
##  3rd Qu.: 68522   3rd Qu.:1.0000   3rd Qu.:1.0000                     
##  Max.   :666666   Max.   :2.0000   Max.   :2.0000                     
##     Recency         MntWines        MntFruits      MntMeatProducts 
##  Min.   : 0.00   Min.   :   0.0   Min.   :  0.00   Min.   :   0.0  
##  1st Qu.:24.00   1st Qu.:  24.0   1st Qu.:  2.00   1st Qu.:  16.0  
##  Median :49.00   Median : 174.5   Median :  8.00   Median :  68.0  
##  Mean   :49.01   Mean   : 305.1   Mean   : 26.36   Mean   : 167.0  
##  3rd Qu.:74.00   3rd Qu.: 505.0   3rd Qu.: 33.00   3rd Qu.: 232.2  
##  Max.   :99.00   Max.   :1493.0   Max.   :199.00   Max.   :1725.0  
##  MntFishProducts  MntSweetProducts  MntGoldProds    NumDealsPurchases
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   : 0.000   
##  1st Qu.:  3.00   1st Qu.:  1.00   1st Qu.:  9.00   1st Qu.: 1.000   
##  Median : 12.00   Median :  8.00   Median : 24.50   Median : 2.000   
##  Mean   : 37.64   Mean   : 27.03   Mean   : 43.97   Mean   : 2.324   
##  3rd Qu.: 50.00   3rd Qu.: 33.00   3rd Qu.: 56.00   3rd Qu.: 3.000   
##  Max.   :259.00   Max.   :262.00   Max.   :321.00   Max.   :15.000   
##  NumWebPurchases  NumCatalogPurchases NumStorePurchases NumWebVisitsMonth
##  Min.   : 0.000   Min.   : 0.000      Min.   : 0.000    Min.   : 0.000   
```

```
##  1st Qu.: 2.000   1st Qu.: 0.000      1st Qu.: 3.000    1st Qu.: 3.000
##  Median : 4.000   Median : 2.000      Median : 5.000    Median : 6.000
##  Mean   : 4.085   Mean   : 2.671      Mean   : 5.801    Mean   : 5.319
##  3rd Qu.: 6.000   3rd Qu.: 4.000      3rd Qu.: 8.000    3rd Qu.: 7.000
##  Max.   :27.000   Max.   :28.000      Max.   :13.000    Max.   :20.000
##   AcceptedCmp3      AcceptedCmp4      AcceptedCmp5     AcceptedCmp1
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.07356   Mean   :0.07401   Mean   :0.0731   Mean   :0.06408
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.0000   Max.   :1.00000
##   AcceptedCmp2        Complain        Z_CostContact   Z_Revenue
##  Min.   :0.00000   Min.   :0.000000   Min.   :3     Min.   :11
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:3     1st Qu.:11
##  Median :0.00000   Median :0.000000   Median :3     Median :11
##  Mean   :0.01354   Mean   :0.009477   Mean   :3     Mean   :11
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:3     3rd Qu.:11
##  Max.   :1.00000   Max.   :1.000000   Max.   :3     Max.   :11
##     Response
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.1503
##  3rd Qu.:0.0000
##  Max.   :1.0000
```

Attributes

#People

ID: Customer's unique identifier Year_Birth: Customer's birth year Education: Customer's education level Marital_Status: Customer's marital status Income: Customer's yearly household income Kidhome: Number of children in customer's household Teenhome: Number of teenagers in customer's household Dt_Customer: Date of customer's enrollment with the company Recency: Number of days since customer's last purchase Complain: 1 if the customer complained in the last 2 years, 0 otherwise #Products

MntWines: Amount spent on wine in last 2 years MntFruits: Amount spent on fruits in last 2 years MntMeatProducts: Amount spent on meat in last 2 years MntFishProducts: Amount spent on fish in last 2 years MntSweetProducts: Amount spent on sweets in last 2 years MntGoldProds: Amount spent on gold in last 2 years #Promotion

NumDealsPurchases: Number of purchases made with a discount AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise Response: 1 if customer accepted the offer in the last campaign, 0 otherwise #Place

NumWebPurchases: Number of purchases made through the company's website NumCatalogPurchases: Number of purchases made using a catalogue NumStorePurchases: Number of purchases made directly in stores NumWebVisitsMonth: Number of visits to company's website in the last month

As a company, it is impolite to collect information about a customers' incomes. However, income is a important attribute. For example, company can recommend latest products and services to higher-income groups, while recommending discounted services to lower-income groups.

The company may know other information, like the number of purchases made with a discount, the number of purchases made through the company's website, which can be got by on-line transcation processing (OLTP), whcih is used to deal with everyday running of one aspect of an enterprise; customers also may fill in some personal information before becoming a member. Customer's birth year and education level may also can be got.

However, most of the attributes don't fluctuate very much (Most variables are related to frequency). But the income gap will be huge. The prediction may be difficult.

In this project, we want to use random intercept and random slope model to build a model to predict income based on other features. Based on our understanding, fixed models aren't easy to use categorical features which volatile very little to build models. However, this kind of features may be good to group. Random effect models is helpful by introducing more variables that are less convenient to add in a fixed model, in prder to improve prediction accuracy.

```r
library(car)
```

```
##      carData
```

```r
library(corrplot)
```

```
## Warning:   'corrplot' R 4.3.2
```

```
## corrplot 0.92 loaded
```

```r
library(readr)
```

```
## Warning:   'readr' R 4.3.2
```

```r
library(dplyr)
```

```
## Warning:   'dplyr' R 4.3.2
```

```
##
##    'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(lme4)
```

```
##      Matrix
```

```r
library(mlmRev)
```

```
## Warning:   'mlmRev' R 4.3.2
```

```r
library(sjstats)
```

```
## Warning:   'sjstats' R 4.3.2
```

```r
library(lattice)
library(dplyr)
```

We put the installation package at the top for easy management.

We want to start simple.

Random Interect Model: no x's

Customers' incomes are closely related to the education levels.In most situation, the higher the education level, the higher the income.

There are two characteristics:

Income: Customer's yearly household income Education: Customer's education level

Consider the follow level-1 and level-2 models:

$$Income_{ij} = \alpha_{0i} + \varepsilon_{ij}$$

$$\alpha_{0i} = \gamma_{00} + u_{0i}$$

In Laird-Ware form:

$$y_{ij} = \beta_0 + b_0 + \varepsilon_{ij}$$

This us a random-effects one-way ANOVA model with one fixed effect, β_1, representing the general population mean of the customers' yearly household income, and two random effects:

b_{0i} , representing the deviation of of income for the education level i from the general mean

ε_{ij}, representing the deviation of indivdual j's income at the education level i from the education mean

there are two variance components for this models:

- Var(b_{0i}) $= d^2$: the variance among education level means
- Var(ε_{ij}) $= \delta^2$: the variance among individuals at the same education level

since b_{0i} and ε_{ij} are assumed to be independent, the variation in incomes among individuals can be decomposed into these two variance components:

$$Var(y_{ij}) = d^2 + \delta^2$$

```r
fit.model1 <- lmer(Income~1+(1|Education), data=data)
s1 <- summary(fit.model1)
s1
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Income ~ 1 + (1 | Education)
##    Data: data
##
## REML criterion at convergence: 51098.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1174 -0.6748 -0.0369  0.6203 24.9710
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  Education (Intercept) 200365017 14155
##  Residual              604522974 24587
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    46187       6385   7.234
```

```r
ranef(fit.model1)
```

```
## $Education
##             (Intercept)
## 2n Cycle       1424.535
## Basic        -24511.396
## Graduation     6515.594
## Master         6675.192
## PhD            9896.075
##
## with conditional variances for "Education"
```

NULL Model

Parameter	Value/variance	SD
γ_{00}	46187	6385
d^2	200365017	14155
δ^2	604522974	24587

```r
performance::icc(fit.model1)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.249
##   Unadjusted ICC: 0.249
```

the intra-class correlation is the proportion of variation in individuals' income due to different education levels:

$$\frac{d^2}{Var(y_{ij})} = \frac{d^2}{d^2 + \delta^2} = \rho$$

ρ may also be interpreted as the correlation between the incomes of two individuals at the same education level:

$$Cor(y_{ij}, y_{ij'}) = \rho$$

$\rho = 0.249$, about 24.9 percent of the variation in customers' incomes is "attribute" to differences at the same education levels

and $(\hat{Income})_{ij} = 46187$, is the overall mean, The overall variance of incomes is $s^2 =$ (200365017+ 604522974) = 804887991. s = 28370.55

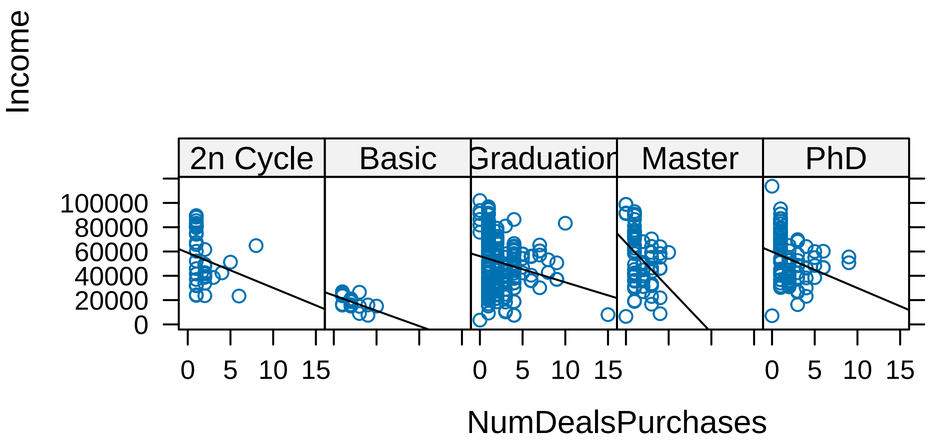
Add NumDealsPurchases as Level 1 Explanatory?

```
set.seed(1234)
#randomly select 4 single customers' educations
cat<-unique(Education[Marital_Status=="Single"],4)
cat
```

```
## [1] "Graduation" "PhD"        "2n Cycle"   "Master"     "Basic"
```

```
#check whether it matches with data in cat
cat.4<-data[is.element(Education,cat),]
cat.4 <- cat.4 %>%
filter(Marital_Status=="Single")
plot<-xyplot(Income~NumDealsPurchases|Education,data=cat.4,main="Single",xlab="NumDealsPurchases",ylab=
    panel.xyplot(x,y)
    panel.lmline(x,y)
  })
plot
```

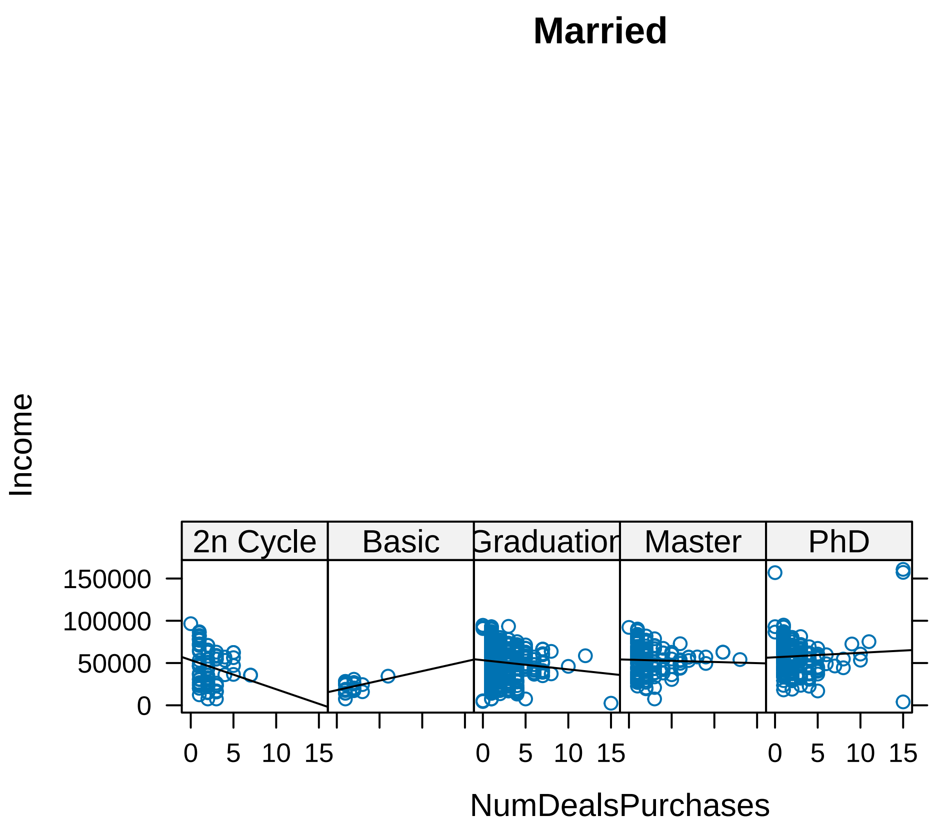
Single




```
#randomly select 4 Married customers' educations
cat<-unique(Education[Marital_Status=="Married"],4)
cat
```

```
## [1] "PhD"        "Basic"      "Graduation" "Master"     "2n Cycle"
```

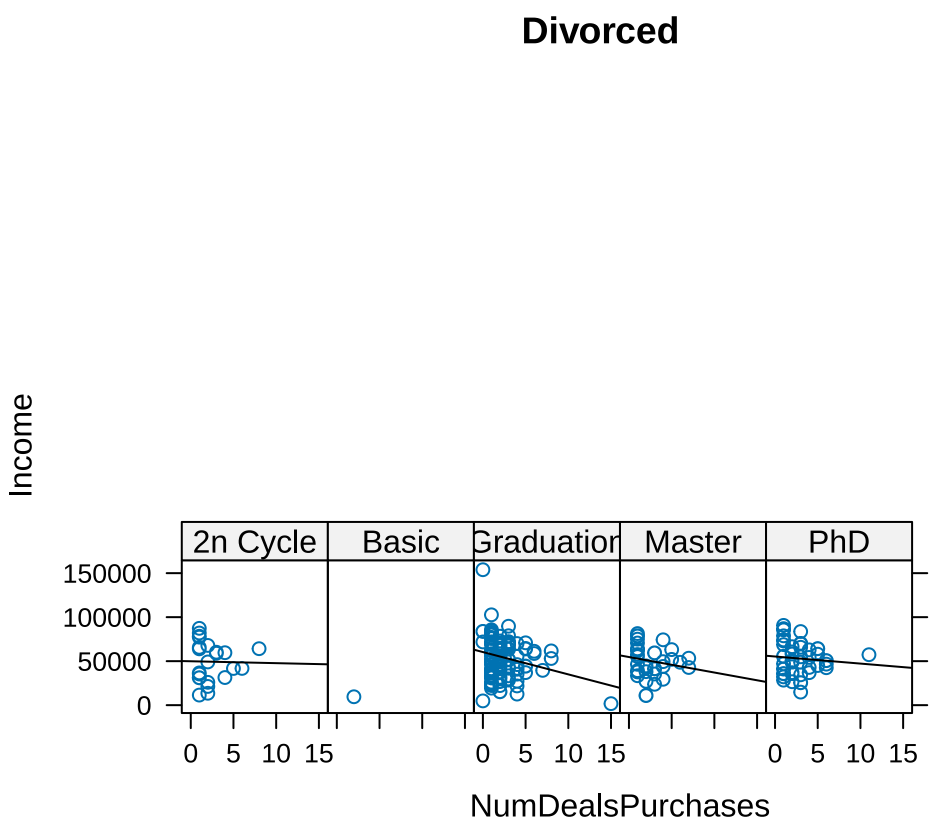
```
#check whether it matches with data in cat
cat.4<-data[is.element(Education,cat),]
cat.4 <- cat.4 %>%
filter(Marital_Status=="Married")
plot<-xyplot(Income~NumDealsPurchases|Education,data=cat.4,main="Married",xlab="NumDealsPurchases",ylab=
    panel.xyplot(x,y)
    panel.lmline(x,y)
  })
plot
```




```
#randomly select 4 divorced customers' educations
cat<-unique(Education[Marital_Status=="Divorced"],4)
cat
```

```
## [1] "Graduation" "Master"     "PhD"        "Basic"      "2n Cycle"
```

```
#check whether it matches with data in cat
cat.4<-data[is.element(Education,cat),]
cat.4 <- cat.4 %>%
filter(Marital_Status=="Divorced")
plot<-xyplot(Income~NumDealsPurchases|Education,data=cat.4,main="Divorced",xlab="NumDealsPurchases",ylal
    panel.xyplot(x,y)
    panel.lmline(x,y)
  })
plot
```



Here We can observe from the graph that NumDealsPurchases seems highly related to Income, as the income get higher,the value of NumDealsPurchases tends to be first increase and then decrease for different clusters, and for graduate,basic, it seems to be much more dense in the 0-2.5 region then master and phd classes. Here we try to explain the reason behind: the low income family maybe not good at manage their money, so they are less likely to use discount, but if the Income is high for that family, they won't pay attention to the discount. The popular which pay more attention to discount is those mid-Income falimilies. And for Masters and Phds, they maybe better in manage their salary, so they tend to make use of disaount better.

NumDealsPurchases to help explain some of the variability of Y_{ij}.

a random-effects one-way ANCOVA

–1 level-1 predictor (NumDealsPurchases, centered with education level), no level-2 predictors –random intercept, no random slops –model for the first (individual) level:

$$y_{ij} = \alpha_{0i} + \alpha_{1i} NumDealsPurchases_{ij} + \varepsilon_{ij}$$

–model for the second(education level) level: $\alpha_{0i} = \gamma_{00} + u_{0i}$ (the random intercept) $\alpha_{1i} = \gamma_{10}$ (the constant slope)

–the combined model and the Laird-Ware form:

$$y_{ij} = (\gamma_{00} + u_{0i}) + \gamma_{01} NumDealsPurchases_{ij} + \varepsilon_{ij}$$

$$= \gamma_{00} + \gamma_{01} NumDealsPurchases_{ij} + u_{0i} + \varepsilon_{ij} = \beta_0 + \beta_1 x_{1ij} + b_{0i} + \varepsilon_{ij}$$

the fixed-effect coefficients β_0 and β_1 represent the average within-education-levels population intercept and slope respectively

```r
fit.model2 <- lmer(Income~1+NumDealsPurchases+(1|Education), data=data)
summary(fit.model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Income ~ 1 + NumDealsPurchases + (1 | Education)
##    Data: data
##
## REML criterion at convergence: 51065.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1289 -0.6847 -0.0016  0.6167 25.1634
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  Education (Intercept) 209329613 14468
##  Residual              599296388 24481
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        48888.1     6551.2   7.462
## NumDealsPurchases  -1214.1      270.7  -4.485
##
## Correlation of Fixed Effects:
##             (Intr)
## NmDlsPrchss -0.092
```

```r
performance::icc(fit.model2)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.259
##   Unadjusted ICC: 0.257
```

While the adjusted ICC only relates to the random effects, the unadjusted ICC also takes the fixed effects variances into account.In our analysis, we just use adjusted ICC to ignore the effects of fixed variables.

NULL Model Add NumDealsPurchases

	Value/variance	SD	Value/variance	SD
Fixed effects				
γ_{00}	46187	6385	48888.1	6551.2
γ_{01}	-	-	-1214.1	270.7
Random effects				
d^2	200365017	14155	209329613	14468
δ^2	604522974	24587	599296388	24481

Residual intra-class correlation: $\hat{\rho}(Income|NumDealsPurchases) = \frac{209329613}{209329613+599296388} = 0.259$, is larger than the original one! That means a larger variation in customers' incomes is "attribute" to differences at the same education levels! Our try is better than the example about high schools because its ICC is from 0.18 to 0.11.

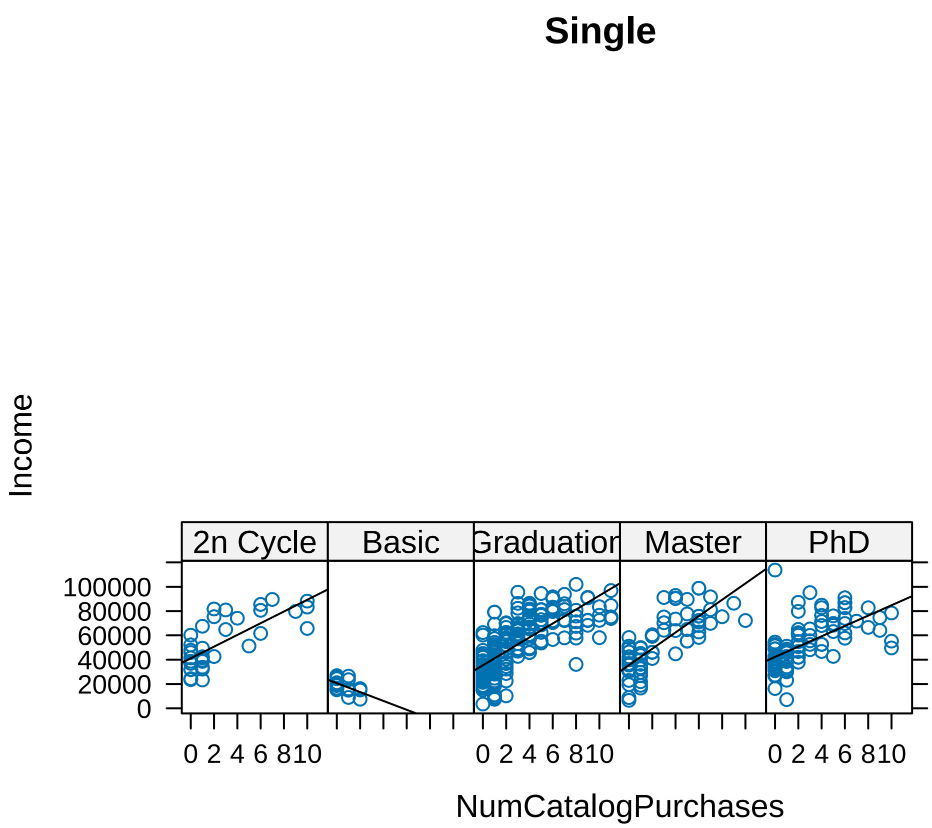
However, ICC can explain the difference among groups, it is not an indicator about how good a model is. We cannot just focus on ICC.

Groups variance estimate $\hat{\gamma_0}^2$is larger, from 200365017 to 209329613. This isn't good for our model. In that case, we'd like to change to a better variable. There is a very unique variable called 'NumCatalogPurchases', which is the number of purchases made using a catalogue. We believe that this variable is not related too much to educational levels, and more in favor of individual consumption habits. It may be good to be a variable for the fixed model.

```
set.seed(1234)
#randomly select 4 single customers' education
cat<-unique(Education[Marital_Status=="Single"],4)
cat
```

```
## [1] "Graduation" "PhD"        "2n Cycle"   "Master"     "Basic"
```

```
#check whether it matches with data in cat
cat.4<-data[is.element(Education,cat),]
cat.4 <- cat.4 %>%
filter(Marital_Status=="Single")
plot<-xyplot(Income~NumCatalogPurchases|Education,data=cat.4,main="Single",xlab="NumCatalogPurchases",yl
    panel.xyplot(x,y)
    panel.lmline(x,y)
  })
plot
```

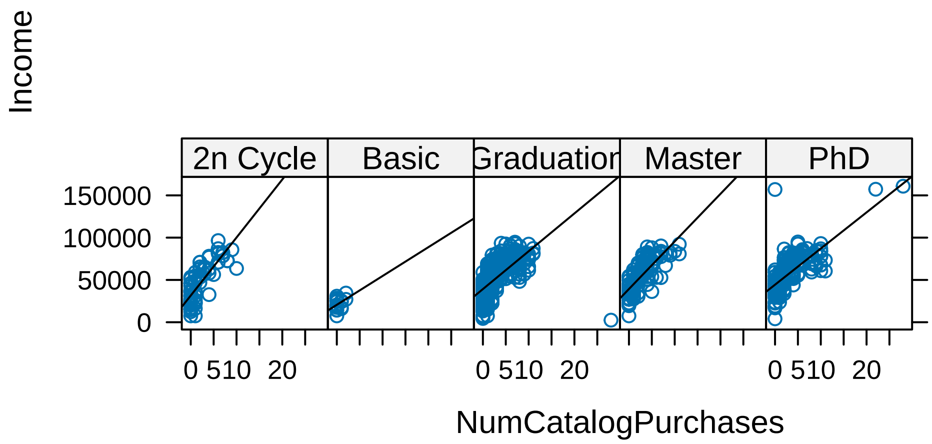



```
#randomly select 18 married customers' education
cat<-unique(Education[Marital_Status=="Married"],4)
cat
```

```
## [1] "PhD"        "Basic"      "Graduation" "Master"     "2n Cycle"
```

```
#check whether it matches with data in cat
cat.4<-data[is.element(Education,cat),]
cat.4 <- cat.4 %>%
filter(Marital_Status=="Married")
plot<-xyplot(Income~NumCatalogPurchases|Education,data=cat.4,main="Married",xlab="NumCatalogPurchases",
    panel.xyplot(x,y)
    panel.lmline(x,y)
  })
plot
```

Married




```r
fit.model3 <- lmer(Income~1+NumCatalogPurchases+NumDealsPurchases+(1|Education), data=data)
summary(fit.model3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Income ~ 1 + NumCatalogPurchases + NumDealsPurchases + (1 | Education)
##    Data: data
##
## REML criterion at convergence: 50141.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.0471 -0.4987 -0.0073  0.4529 31.3180
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  Education (Intercept)  91385387  9560
##  Residual              397394718 19935
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)          37699.8     4368.5   8.630
## NumCatalogPurchases   4905.0      146.0  33.599
## NumDealsPurchases    -1083.7      220.5  -4.915
##
```

```
## Correlation of Fixed Effects:
##             (Intr) NmCtlP
## NmCtlgPrchs -0.077
## NmDlsPrchss -0.114  0.017
```

```
performance::icc(fit.model3)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.187
##   Unadjusted ICC: 0.131
```

Although the ICC of this new model decreases from 0.259 to 0.176, the variances decrease so much.

Drop in between groups variance estimate $\hat{\gamma_0}^2$:

$$\frac{91385387}{209329613} = 0.4365622$$

or$(1 - 0.4365622) * 100 = 56.36$ decrease

Drop in within groups variance estimate$\hat{\hat{\delta}}^2$:

$$\frac{397394718}{604522974} = 0.6573691$$

or $(1 - 0.6573691) * 100 = 34$ decrease

Interpretation some what problematic because NumCatalogPurchases helps to explain both the between and within groups variance of Y_{ij}:

$$NumCatalogPurchases_{ij} = \overline{NumCatalogPurchases_j} + (NumCatalogPurchases_{ij} - \overline{NumCatalogPurchases_j})$$

Regarding different mean levels of NumCatalogPurchases between and within education levels:

"Group mean centered" variable, e.g.:

$$x_{ij} = (NumCatalogPurchases_{ij} - \overline{NumCatalogPurchases_j})$$

to model with group variability of Y_{ij} w/rt education level

Group mean as a level 2 (education level) for NumCatalogPurchases:

$$z_{ij} = \overline{NumCatalogPurchases_j}$$

The corresponding linear mixed model:

$$Income_{ij} = \gamma_{00} + \gamma_{01}(NumCatalogPurchases_{ij} - \overline{NumCatalogPurchases_j}) + \gamma_{02}NumDealsPurchases_{ij} + U_{0j} + \varepsilon_{ij}$$

```
mean_value <- mean(NumCatalogPurchases, na.rm = TRUE)
data$NumCatalogPurchases2 <- NumCatalogPurchases - mean_value
```

```
fit.model4 <- lmer(Income~1+NumDealsPurchases+NumCatalogPurchases2+(1|Education), data=data)
summary(fit.model4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Income ~ 1 + NumDealsPurchases + NumCatalogPurchases2 + (1 |
##     Education)
##    Data: data
##
## REML criterion at convergence: 50141.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.0471 -0.4987 -0.0073  0.4529 31.3180
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  Education (Intercept)  91385387  9560
##  Residual              397394718 19935
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)           50801.2     4355.9  11.663
## NumDealsPurchases     -1083.7      220.5  -4.915
## NumCatalogPurchases2   4905.0      146.0  33.599
##
## Correlation of Fixed Effects:
##             (Intr) NmDlsP
## NmDlsPrchss -0.113
## NmCtlgPrch2  0.012  0.017
```

```r
performance::icc(fit.model4)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.187
##   Unadjusted ICC: 0.131
```

Maybe because our fixed effects are just the counts, which have very little fluctuation. Hierarchical model with centered NumCatalogPurchases may not as strong as the high school example. We wanted to do more analysis for this part, but decided to move on to other models because of the poor result.

Can we add more fixed variables?

```r
fit.model5 <- lmer(Income~1+NumCatalogPurchases+NumDealsPurchases+Recency+(1|Education), data=data)
summary(fit.model5)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Income ~ 1 + NumCatalogPurchases + NumDealsPurchases + Recency +
##     (1 | Education)
##    Data: data
##
## REML criterion at convergence: 50133.4
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.0581 -0.4997 -0.0007  0.4555 31.2983
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  Education (Intercept)  91317199  9556
##  Residual              397383430 19934
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)         38424.38    4423.00   8.687
## NumCatalogPurchases  4908.53     146.03  33.614
## NumDealsPurchases   -1082.97     220.47  -4.912
## Recency               -15.12      14.64  -1.032
##
## Correlation of Fixed Effects:
##             (Intr) NmCtlP NmDlsP
## NmCtlgPrchs -0.072
## NmDlsPrchss -0.112  0.017
## Recency     -0.159 -0.023 -0.003
```

```r
drop1(fit.model5, test="Chisq")
```

```
## Single term deletions
##
## Model:
## Income ~ 1 + NumCatalogPurchases + NumDealsPurchases + Recency +
##     (1 | Education)
##                     npar   AIC    LRT   Pr(Chi)
## <none>                   50195
## NumCatalogPurchases    1 51108 914.47 < 2.2e-16 ***
## NumDealsPurchases      1 50217  23.96  9.85e-07 ***
## Recency                1 50195   1.07    0.3015
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We try to add 'recency' as an example. However, we found that the AIC is the same whether or not the model contains a recency (both are 50195). the predictors do a sufficiently good job of accounting for differences in slopes that the variance component for slopes is no longer needed.

For the fixed effects, can we employ something we learned before like variable selection or multicollinearity to get good predictors? We will try to select the best predictors later. Let us analyze more about the random effect models now!

Try random slope?

Will random slopes also help to improve the model? It deserves to be studied.

```r
fit.model5 <- lmer(Income~1+NumCatalogPurchases+NumDealsPurchases+(1+NumCatalogPurchases|Education), da
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
summary(fit.model5)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Income ~ 1 + NumCatalogPurchases + NumDealsPurchases + (1 + NumCatalogPurchases |
##     Education)
##    Data: data
##
## REML criterion at convergence: 50136.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.1791 -0.4981  0.0088  0.4607 31.3460
##
## Random effects:
##  Groups    Name                Variance  Std.Dev. Corr
##  Education (Intercept)         104594149 10227
##            NumCatalogPurchases   2005644  1416    -1.00
##  Residual                      396664790 19916
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)          37197.7     4667.8   7.969
## NumCatalogPurchases   5607.0      654.0   8.574
## NumDealsPurchases    -1069.0      220.5  -4.848
##
## Correlation of Fixed Effects:
##             (Intr) NmCtlP
## NmCtlgPrchs -0.980
## NmDlsPrchss -0.105 -0.002
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```r
isSingular(fit.model5)
```

```
## [1] TRUE
```

Singular models mean there are linear dependencies or collinearity among the predictor variables.If the model is singular, the result is easy to get bias so it isn't good enough. In our project, we try many times but easy to get singular models. We just list one as an example try here.

Can we add more random intercepts?

So we decided to consider more intercepts at first.

```r
fit.modelr2 <- lmer(Income~1+NumDealsPurchases+NumCatalogPurchases+(1|Education)+(1|Year_Birth), data=da
summary(fit.modelr2)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula:
## Income ~ 1 + NumDealsPurchases + NumCatalogPurchases + (1 | Education) +
##     (1 | Year_Birth)
##    Data: data
##
## REML criterion at convergence: 50139.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.9869 -0.5019 -0.0092  0.4603 31.2869
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Year_Birth (Intercept)   2811412  1677
##  Education  (Intercept)  88459361  9405
##  Residual               394764476 19869
## Number of obs: 2216, groups:  Year_Birth, 59; Education, 5
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)           37848.0     4308.3   8.785
## NumDealsPurchases     -1101.2      220.8  -4.987
## NumCatalogPurchases    4886.2      146.3  33.402
##
## Correlation of Fixed Effects:
##             (Intr) NmDlsP
## NmDlsPrchss -0.115
## NmCtlgPrchs -0.079  0.017
```

Customer's birth year may also be closely related to income levels. People born earlier are more likely to earn higher incomes because they will have more experience. We add a new random intercept named 'year_birth'.

```
performance::icc(fit.modelr2)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.188
##   Unadjusted ICC: 0.131
```

The adjusted ICC still increases! Residual also decreases from 397394718 to 394764476!

```
ranef(fit.modelr2)
```

```
## $Year_Birth
##        (Intercept)
## 1893   159.9311475
## 1899    79.5747030
## 1900   -41.0953848
## 1940    23.2855851
## 1941     0.7224783
## 1943  -370.9075785
## 1944   122.4241281
```

```
## 1945   526.6810176
## 1946   447.8032841
## 1947   697.2208996
## 1948    42.8194102
## 1949   701.2453942
## 1950  -244.5367085
## 1951   608.3184858
## 1952   443.8316993
## 1953   322.2265835
## 1954   797.4421429
## 1955   421.9594489
## 1956   731.6309477
## 1957    -3.1552432
## 1958   548.7976836
## 1959   367.9328903
## 1960    74.8510639
## 1961  1028.9143049
## 1962  1290.3374466
## 1963 -1179.2940449
## 1964   -78.5959356
## 1965  1352.2134447
## 1966   525.5863360
## 1967   427.3390385
## 1968   -60.6153396
## 1969  -641.7618176
## 1970  -254.7279031
## 1971  -366.0309973
## 1972   227.9841910
## 1973  -446.8327941
## 1974  -119.5999126
## 1975   358.3996316
## 1976  -916.7886155
## 1977  2684.9501476
## 1978  -553.5378578
## 1979 -1029.2376831
## 1980   283.0891749
## 1981   -72.1172427
## 1982  -976.8406576
## 1983  -370.8510196
## 1984 -1942.1506521
## 1985 -1192.5842219
## 1986  -798.3597487
## 1987  -971.4220868
## 1988  -799.5504054
## 1989 -1055.9564712
## 1990  -762.8101921
## 1991    57.3217278
## 1992  -226.2978342
## 1993     4.9736335
## 1994   457.6511829
## 1995   -68.9154672
## 1996  -272.8854380
##
## $Education
```

```
##              (Intercept)
## 2n Cycle        818.5622
## Basic        -16211.8996
## Graduation     4022.3967
## Master         5155.8156
## PhD            6215.1250
##
## with conditional variances for "Year_Birth" "Education"
```

We use 'ranef()' to make a list containing the estimated random effects for each level of the grouping factor.

```
age = 2023-Year_Birth
fit.modelr3 <- lmer(Income~1+NumDealsPurchases+NumCatalogPurchases+(1|Education)+(1|age), data=data)
summary(fit.modelr3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Income ~ 1 + NumDealsPurchases + NumCatalogPurchases + (1 | Education) +
##     (1 | age)
##    Data: data
##
## REML criterion at convergence: 50139.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.9869 -0.5019 -0.0092  0.4603 31.2869
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  age       (Intercept)   2811412  1677
##  Education (Intercept)  88459360  9405
##  Residual              394764476 19869
## Number of obs: 2216, groups:  age, 59; Education, 5
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)           37848.0     4308.3   8.785
## NumDealsPurchases     -1101.2      220.8  -4.987
## NumCatalogPurchases    4886.2      146.3  33.402
##
## Correlation of Fixed Effects:
##             (Intr) NmDlsP
## NmDlsPrchss -0.115
## NmCtlgPrchs -0.079  0.017
```

We also want to try use age instead of year_birth

```
performance::icc(fit.modelr3)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.188
##   Unadjusted ICC: 0.131
```

The effect doesn't change too much. This is of course because the variables did not acquire large changes that would affect the variance, but rather the intercepts.

More random intercepts?

```r
fit.modelr4 <- lmer(Income~1+NumCatalogPurchases+NumDealsPurchases+(1|Education)+(1|Year_Birth)+(1|Compl
summary(fit.modelr4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Income ~ 1 + NumCatalogPurchases + NumDealsPurchases + (1 | Education) +
##     (1 | Year_Birth) + (1 | Complain)
##    Data: data
##
## REML criterion at convergence: 50139.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.9869 -0.5019 -0.0092  0.4603 31.2869
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Year_Birth (Intercept) 2.811e+06  1676.667
##  Education  (Intercept) 8.848e+07  9406.444
##  Complain   (Intercept) 5.203e+01     7.213
##  Residual               3.948e+08 19868.680
## Number of obs: 2216, groups:  Year_Birth, 59; Education, 5; Complain, 2
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)          37847.9     4308.8   8.784
## NumCatalogPurchases   4886.2      146.3  33.402
## NumDealsPurchases    -1101.2      220.8  -4.987
##
## Correlation of Fixed Effects:
##             (Intr) NmCtlP
## NmCtlgPrchs -0.079
## NmDlsPrchss -0.115  0.017
```

```r
performance::icc(fit.modelr4)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.188
##   Unadjusted ICC: 0.131
```

```r
fit.modelr5 <- lmer(Income~1+NumCatalogPurchases+(1|Education)+(1|Year_Birth)+(1|Complain)+(1|Recency),
summary(fit.modelr5)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula:
## Income ~ 1 + NumCatalogPurchases + (1 | Education) + (1 | Year_Birth) +
##     (1 | Complain) + (1 | Recency)
##    Data: data
##
## REML criterion at convergence: 50174
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.6787 -0.4783 -0.0220  0.4513 30.8752
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Recency    (Intercept) 4.643e+06  2154.772
##  Year_Birth (Intercept) 2.177e+06  1475.579
##  Education  (Intercept) 8.383e+07  9155.960
##  Complain   (Intercept) 3.703e+01     6.086
##  Residual               3.949e+08 19871.641
## Number of obs: 2216, groups:
## Recency, 100; Year_Birth, 59; Education, 5; Complain, 2
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)            35381       4175   8.475
## NumCatalogPurchases     4898        147  33.328
##
## Correlation of Fixed Effects:
##             (Intr)
## NmCtlgPrchs -0.080
```

```
performance::icc(fit.modelr5)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.187
##   Unadjusted ICC: 0.131
```

```
ranef(fit.modelr5)
```

```
## $Recency
##    (Intercept)
## 0   -316.04421
## 1    497.83799
## 2    579.32160
## 3   1027.42458
## 4   1087.87717
## 5  -1013.47881
## 6   -314.22432
## 7   -981.58083
## 8   -487.68230
## 9    878.97247
## 10    63.92461
## 11  -947.43442
```

```
## 12  -357.91541
## 13   605.59649
## 14  -409.89894
## 15 -1081.08778
## 16 -1057.50029
## 17  1075.53454
## 18   422.47113
## 19  -315.96358
## 20  -141.65590
## 21    45.11222
## 22   174.08252
## 23  6739.05267
## 24 -2147.41787
## 25   586.00327
## 26   471.98061
## 27   644.13219
## 28  -275.46711
## 29   929.85923
## 30  -162.83095
## 31  1134.86420
## 32  -217.08540
## 33   528.57234
## 34  -827.05996
## 35   -78.30016
## 36   591.98538
## 37   579.12894
## 38  -489.34193
## 39  -806.34215
## 40  -489.68822
## 41    59.67913
## 42  -657.20352
## 43  -448.27764
## 44   467.12862
## 45   -43.74652
## 46  -716.85550
## 47  -453.71145
## 48   474.51619
## 49  -364.89798
## 50  -176.92575
## 51  -608.40750
## 52 -1055.22601
## 53  -579.01963
## 54   692.80178
## 55   -37.54573
## 56  -620.22449
## 57  1347.16846
## 58  -119.97111
## 59   -24.16263
## 60  -430.32512
## 61   165.36870
## 62  -409.23219
## 63  -976.73131
## 64   141.86558
## 65  -309.02635
```

```
## 66  1008.24408
## 67   119.19913
## 68  -344.50421
## 69   212.20930
## 70   -13.56031
## 71  -428.31288
## 72   958.41460
## 73  1373.77503
## 74  -233.68416
## 75  -922.36105
## 76 -1143.64193
## 77 -1143.26141
## 78   726.58479
## 79   250.22301
## 80  -879.25528
## 81   462.79226
## 82  -544.85277
## 83   923.05261
## 84   -92.80162
## 85  1701.15223
## 86  -844.92076
## 87   633.21622
## 88  -416.97821
## 89  -307.29112
## 90 -1167.16073
## 91   -25.04538
## 92 -1096.11809
## 93   297.95148
## 94   -26.74668
## 95  -164.65902
## 96  -468.13729
## 97  -179.66843
## 98   -49.31488
## 99   762.68987
##
## $Year_Birth
##       (Intercept)
## 1893    94.418143
## 1899    66.716000
## 1900   -28.763368
## 1940    29.198011
## 1941    20.867161
## 1943  -256.379518
## 1944   136.912113
## 1945   429.070550
## 1946   333.480437
## 1947   638.507763
## 1948    62.612296
## 1949   566.949810
## 1950  -158.799751
## 1951   501.984647
## 1952   277.956272
## 1953   232.429719
## 1954   601.931023
```

```
## 1955   286.351265
## 1956   433.965215
## 1957    41.762131
## 1958   497.757682
## 1959   378.814650
## 1960   -65.901487
## 1961   831.622457
## 1962   973.846276
## 1963 -1073.710641
## 1964  -124.631452
## 1965  1093.775481
## 1966   354.759893
## 1967   353.505770
## 1968  -205.215820
## 1969  -452.271840
## 1970  -297.964287
## 1971  -502.586388
## 1972   191.582977
## 1973  -605.974913
## 1974   -89.959095
## 1975   224.900793
## 1976  -738.653110
## 1977  2198.558130
## 1978  -456.744056
## 1979  -863.711262
## 1980   226.896634
## 1981   -92.931669
## 1982  -740.157683
## 1983   -76.204482
## 1984 -1521.480980
## 1985  -889.955555
## 1986  -529.316302
## 1987  -708.746262
## 1988  -549.633776
## 1989  -721.479277
## 1990  -491.170279
## 1991    65.802823
## 1992  -108.185366
## 1993    32.549904
## 1994   379.495821
## 1995    -6.028857
## 1996  -202.424375
##
## $Education
##            (Intercept)
## 2n Cycle      843.7444
## Basic      -15781.4344
## Graduation   3908.8244
## Master       5023.1726
## PhD          6005.6931
##
## $Complain
##    (Intercept)
## 0  0.006775673
```

```
## 1 -0.006775673
##
## with conditional variances for "Recency" "Year_Birth" "Education" "Complain"
```

```r
fit.modelr6 <- lmer(Income~1+NumDealsPurchases+(1|Education)+(1|Year_Birth)+(1|Complain)+(1|Recency)+(1
summary(fit.modelr5)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Income ~ 1 + NumCatalogPurchases + (1 | Education) + (1 | Year_Birth) +
##     (1 | Complain) + (1 | Recency)
##    Data: data
##
## REML criterion at convergence: 50174
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.6787 -0.4783 -0.0220  0.4513 30.8752
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Recency    (Intercept) 4.643e+06  2154.772
##  Year_Birth (Intercept) 2.177e+06  1475.579
##  Education  (Intercept) 8.383e+07  9155.960
##  Complain   (Intercept) 3.703e+01     6.086
##  Residual               3.949e+08 19871.641
## Number of obs: 2216, groups:
## Recency, 100; Year_Birth, 59; Education, 5; Complain, 2
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)            35381       4175   8.475
## NumCatalogPurchases     4898        147  33.328
##
## Correlation of Fixed Effects:
##             (Intr)
## NmCtlgPrchs -0.080
```

```r
performance::icc(fit.modelr6)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.307
##   Unadjusted ICC: 0.305
```

We found we can add at most 5 random intercepts. The ICC increases to 0.307 which is very high. However, it's not true that more random intercepts are better. The residual is 394900000 now, it is higher than the previous ones, which tends to lead to large errors. You also may find we don't add NumCatalogPurchases in our 'fit.modelr6' because it will make our model singular.

```r
fit.modelr6ML <- lmer(Income~1+NumDealsPurchases+(1|Education)+(1|Year_Birth)+(1|Complain)+(1|Recency)+
summary(fit.modelr6ML)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Income ~ 1 + NumDealsPurchases + (1 | Education) + (1 | Year_Birth) +
##     (1 | Complain) + (1 | Recency) + (1 | Response)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##  51063.4  51109.0 -25523.7  51047.4     2208
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1615 -0.6464  0.0045  0.6146 25.3533
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Recency    (Intercept)   8089565  2844
##  Year_Birth (Intercept)  17777938  4216
##  Education  (Intercept) 153724974 12399
##  Response   (Intercept)  27099822  5206
##  Complain   (Intercept)   4701873  2168
##  Residual               566940426 23811
## Number of obs: 2216, groups:
## Recency, 100; Year_Birth, 59; Education, 5; Response, 2; Complain, 2
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        51819.3     7076.0   7.323
## NumDealsPurchases  -1244.3      268.3  -4.638
##
## Correlation of Fixed Effects:
##             (Intr)
## NmDlsPrchss -0.084
```

ML method does not account for the random effects' estimation uncertainty,it can result in a downward bias in the fixed-effects estimates. The residual is 566940426 but still very high. The AIC is 51063.4. Can we make it smaller?

Try random slopes again!

```
fit.modelr7 <- lmer(Income~1++NumDealsPurchases+Recency+(1|Education)+(1|Year_Birth)+(1+AcceptedCmp1|Re
summary(fit.modelr7)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Income ~ 1 + +NumDealsPurchases + Recency + (1 | Education) +
##     (1 | Year_Birth) + (1 + AcceptedCmp1 | Recency)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##  51041.1  51092.4 -25511.5  51023.1     2207
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.1547 -0.6145 -0.0095  0.5633 25.8220
##
## Random effects:
##  Groups     Name         Variance  Std.Dev. Corr
##  Recency    (Intercept)   14927504  3864
##             AcceptedCmp1 528161500 22982    -0.77
##  Year_Birth (Intercept)   17673222  4204
##  Education  (Intercept)  140918471 11871
##  Residual                543929144 23322
## Number of obs: 2216, groups:  Recency, 100; Year_Birth, 59; Education, 5
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)       49477.330   5530.328   8.947
## NumDealsPurchases  -923.392    266.051  -3.471
## Recency               1.463     20.243   0.072
##
## Correlation of Fixed Effects:
##             (Intr) NmDlsP
## NmDlsPrchss -0.104
## Recency     -0.178  0.002
```

```
fit.modelr8 <- lmer(Income~1+NumDealsPurchases+Recency+(1|Education)+(1|Year_Birth)+(1+AcceptedCmp5+Acce
summary(fit.modelr8)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Income ~ 1 + NumDealsPurchases + Recency + (1 | Education) +
##     (1 | Year_Birth) + (1 + AcceptedCmp5 + AcceptedCmp1 | Recency) +
##     (1 | Response)
##    Data: data
##
##      AIC      BIC   logLik deviance df.resid
##  50973.9  51048.1 -25474.0  50947.9     2203
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1162 -0.6002 -0.0123  0.5429 26.4856
##
## Random effects:
##  Groups     Name         Variance  Std.Dev. Corr
##  Recency    (Intercept)   26191423  5118
##             AcceptedCmp5 451157072 21240    -0.86
##             AcceptedCmp1 252477669 15890    -0.86  1.00
##  Year_Birth (Intercept)   17937076  4235
##  Education  (Intercept)  127456059 11290
##  Response   (Intercept)    1557970  1248
##  Residual                513756382 22666
## Number of obs: 2216, groups:
## Recency, 100; Year_Birth, 59; Education, 5; Response, 2
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)       49780.891   5354.824   9.296
## NumDealsPurchases  -582.799    261.445  -2.229
```

```
## Recency              8.695     20.200   0.430
##
## Correlation of Fixed Effects:
##             (Intr) NmDlsP
## NmDlsPrchss -0.105
## Recency     -0.172 -0.009
```

If we want to add more random slopes, we must drop some existing variables. However, the AIC is from 51063.4 to 50973.9, which increases a little.This try may not good enough but very interesting: AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise Recency: Number of days since customer's last purchase

Only the offer of the first campaign and the last campaign can be used as random slopes, otherwise the models will be singular! Customers may pay more attention to the first campaign because of the freshness. different recencies have different changes to accept the offers. If they always purchase goods from this company, they are more likely to be concerned about campaigns; customers may also care about the last campaign, because customers may not get a chance after that. Other campaigns are not as attractive.

```r
anova(fit.modelr6ML, fit.modelr7,fit.modelr8)
```

```
## Data: data
## Models:
## fit.modelr6ML: Income ~ 1 + NumDealsPurchases + (1 | Education) + (1 | Year_Birth) + (1 | Complain) +
## fit.modelr7: Income ~ 1 + +NumDealsPurchases + Recency + (1 | Education) + (1 | Year_Birth) + (1 + Ac
## fit.modelr8: Income ~ 1 + NumDealsPurchases + Recency + (1 | Education) + (1 | Year_Birth) + (1 + Acc
##               npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit.modelr6ML    8 51063 51109 -25524    51047
## fit.modelr7      9 51041 51092 -25512    51023 24.324  1  8.140e-07 ***
## fit.modelr8     13 50974 51048 -25474    50948 75.121  4  1.879e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We use 'anova()' to compare the effect of random slopes. As you can see, adding more random slopes can make AIC smaller! However, it's not growing very fast.

We have tried many kinds of random effects. However, we didn't systematically learn how to compare random effect models, we try many models but don't know how to choose the best one. Overall, we have deepened our understanding in these attempts, and understand more about linear models!

Now , let us focus on fixed oredictors!!!

We want to try the knowledgeabout variable selection.

```r
#backward method to select variables
full<-lm(Income~.,data=data)
back_data<-step(full,data=data,direction="backward",k=2)#AIC method
```

```
## Start:  AIC=43257.67
## Income ~ ID + Year_Birth + Education + Marital_Status + Kidhome +
##     Teenhome + Dt_Customer + Recency + MntWines + MntFruits +
##     MntMeatProducts + MntFishProducts + MntSweetProducts + MntGoldProds +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
```

```
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Z_CostContact +
##     Z_Revenue + Response + NumCatalogPurchases2
##
##
## Step:  AIC=43257.67
## Income ~ ID + Year_Birth + Education + Marital_Status + Kidhome +
##     Teenhome + Dt_Customer + Recency + MntWines + MntFruits +
##     MntMeatProducts + MntFishProducts + MntSweetProducts + MntGoldProds +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Z_CostContact +
##     Z_Revenue + Response
##
##
## Step:  AIC=43257.67
## Income ~ ID + Year_Birth + Education + Marital_Status + Kidhome +
##     Teenhome + Dt_Customer + Recency + MntWines + MntFruits +
##     MntMeatProducts + MntFishProducts + MntSweetProducts + MntGoldProds +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Z_CostContact +
##     Response
##
##
## Step:  AIC=43257.67
## Income ~ ID + Year_Birth + Education + Marital_Status + Kidhome +
##     Teenhome + Dt_Customer + Recency + MntWines + MntFruits +
##     MntMeatProducts + MntFishProducts + MntSweetProducts + MntGoldProds +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Response
##
##                       Df  Sum of Sq        RSS   AIC
## - Dt_Customer        661 2.7115e+11 6.2634e+11 43193
## - Complain             1 6.9754e+05 3.5519e+11 43256
## - MntFishProducts      1 3.3814e+06 3.5519e+11 43256
## - Response             1 6.1068e+06 3.5519e+11 43256
## - AcceptedCmp2         1 1.0597e+07 3.5520e+11 43256
## - Year_Birth           1 3.4232e+07 3.5522e+11 43256
## - ID                   1 6.8809e+07 3.5525e+11 43256
## - AcceptedCmp1         1 1.0210e+08 3.5529e+11 43256
## - AcceptedCmp3         1 2.4553e+08 3.5543e+11 43257
## <none>                              3.5519e+11 43258
## - MntGoldProds         1 3.3623e+08 3.5552e+11 43258
## - Kidhome              1 3.6704e+08 3.5555e+11 43258
## - AcceptedCmp4         1 5.1692e+08 3.5570e+11 43259
## - Recency              1 5.5765e+08 3.5574e+11 43259
## - NumStorePurchases    1 5.8941e+08 3.5578e+11 43259
## - NumDealsPurchases    1 6.0084e+08 3.5579e+11 43259
## - MntSweetProducts     1 8.5790e+08 3.5604e+11 43261
## - Marital_Status       7 3.0098e+09 3.5820e+11 43262
## - AcceptedCmp5         1 1.1731e+09 3.5636e+11 43263
## - MntFruits            1 1.5235e+09 3.5671e+11 43265
```

```
## - NumCatalogPurchases   1 4.6856e+09 3.5987e+11 43285
## - Education             4 5.8007e+09 3.6099e+11 43286
## - NumWebPurchases       1 4.8892e+09 3.6008e+11 43286
## - MntMeatProducts       1 5.9445e+09 3.6113e+11 43292
## - Teenhome              1 6.8499e+09 3.6204e+11 43298
## - MntWines              1 9.1380e+09 3.6432e+11 43312
## - NumWebVisitsMonth     1 3.0626e+10 3.8581e+11 43439
##
## Step:  AIC=43192.68
## Income ~ ID + Year_Birth + Education + Marital_Status + Kidhome +
##     Teenhome + Recency + MntWines + MntFruits + MntMeatProducts +
##     MntFishProducts + MntSweetProducts + MntGoldProds + NumDealsPurchases +
##     NumWebPurchases + NumCatalogPurchases + NumStorePurchases +
##     NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
##     AcceptedCmp1 + AcceptedCmp2 + Complain + Response
##
##                       Df  Sum of Sq        RSS   AIC
## - Marital_Status       7 5.6378e+08 6.2690e+11 43181
## - Complain             1 3.6645e+07 6.2637e+11 43191
## - AcceptedCmp2         1 6.7213e+07 6.2640e+11 43191
## - MntGoldProds         1 8.2168e+07 6.2642e+11 43191
## - Response             1 1.1293e+08 6.2645e+11 43191
## - MntFishProducts      1 1.7149e+08 6.2651e+11 43191
## - Year_Birth           1 1.8396e+08 6.2652e+11 43191
## - AcceptedCmp3         1 3.8073e+08 6.2672e+11 43192
## - ID                   1 4.0987e+08 6.2675e+11 43192
## <none>                              6.2634e+11 43193
## - Recency              1 6.4371e+08 6.2698e+11 43193
## - NumDealsPurchases    1 6.4444e+08 6.2698e+11 43193
## - MntFruits            1 9.2021e+08 6.2726e+11 43194
## - AcceptedCmp1         1 9.4438e+08 6.2728e+11 43194
## - MntSweetProducts     1 1.1921e+09 6.2753e+11 43195
## - AcceptedCmp5         1 1.1952e+09 6.2753e+11 43195
## - AcceptedCmp4         1 1.2070e+09 6.2754e+11 43195
## - Kidhome              1 1.4567e+09 6.2779e+11 43196
## - NumStorePurchases    1 1.5729e+09 6.2791e+11 43196
## - NumCatalogPurchases  1 4.1688e+09 6.3051e+11 43205
## - Education            4 8.3651e+09 6.3470e+11 43214
## - Teenhome             1 1.0249e+10 6.3659e+11 43227
## - NumWebPurchases      1 1.0482e+10 6.3682e+11 43227
## - MntWines             1 1.0515e+10 6.3685e+11 43228
## - MntMeatProducts      1 1.2794e+10 6.3913e+11 43235
## - NumWebVisitsMonth    1 5.4616e+10 6.8095e+11 43376
##
## Step:  AIC=43180.68
## Income ~ ID + Year_Birth + Education + Kidhome + Teenhome + Recency +
##     MntWines + MntFruits + MntMeatProducts + MntFishProducts +
##     MntSweetProducts + MntGoldProds + NumDealsPurchases + NumWebPurchases +
##     NumCatalogPurchases + NumStorePurchases + NumWebVisitsMonth +
##     AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 + AcceptedCmp1 +
##     AcceptedCmp2 + Complain + Response
##
##                       Df  Sum of Sq        RSS   AIC
## - Complain             1 3.8714e+07 6.2694e+11 43179
```

```
## - MntGoldProds         1 8.7231e+07 6.2699e+11 43179
## - AcceptedCmp2         1 9.0701e+07 6.2699e+11 43179
## - Response             1 1.4487e+08 6.2705e+11 43179
## - MntFishProducts      1 1.8554e+08 6.2709e+11 43179
## - Year_Birth           1 2.1501e+08 6.2712e+11 43179
## - AcceptedCmp3         1 3.8232e+08 6.2728e+11 43180
## - ID                   1 4.0523e+08 6.2731e+11 43180
## <none>                             6.2690e+11 43181
## - NumDealsPurchases    1 6.3772e+08 6.2754e+11 43181
## - Recency              1 6.5340e+08 6.2755e+11 43181
## - AcceptedCmp1         1 9.2365e+08 6.2782e+11 43182
## - MntFruits            1 9.2468e+08 6.2783e+11 43182
## - MntSweetProducts     1 1.1648e+09 6.2807e+11 43183
## - AcceptedCmp4         1 1.1656e+09 6.2807e+11 43183
## - AcceptedCmp5         1 1.2324e+09 6.2813e+11 43183
## - Kidhome              1 1.4893e+09 6.2839e+11 43184
## - NumStorePurchases    1 1.5176e+09 6.2842e+11 43184
## - NumCatalogPurchases  1 4.1834e+09 6.3108e+11 43193
## - Education            4 8.4143e+09 6.3532e+11 43202
## - Teenhome             1 1.0349e+10 6.3725e+11 43215
## - NumWebPurchases      1 1.0634e+10 6.3753e+11 43216
## - MntWines             1 1.0659e+10 6.3756e+11 43216
## - MntMeatProducts      1 1.2832e+10 6.3973e+11 43224
## - NumWebVisitsMonth    1 5.4710e+10 6.8161e+11 43364
##
## Step:  AIC=43178.82
## Income ~ ID + Year_Birth + Education + Kidhome + Teenhome + Recency +
##     MntWines + MntFruits + MntMeatProducts + MntFishProducts +
##     MntSweetProducts + MntGoldProds + NumDealsPurchases + NumWebPurchases +
##     NumCatalogPurchases + NumStorePurchases + NumWebVisitsMonth +
##     AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 + AcceptedCmp1 +
##     AcceptedCmp2 + Response
##
##                       Df  Sum of Sq        RSS   AIC
## - MntGoldProds         1 8.3833e+07 6.2702e+11 43177
## - AcceptedCmp2         1 9.1071e+07 6.2703e+11 43177
## - Response             1 1.4743e+08 6.2709e+11 43177
## - MntFishProducts      1 1.8709e+08 6.2713e+11 43177
## - Year_Birth           1 2.0644e+08 6.2715e+11 43178
## - AcceptedCmp3         1 3.8502e+08 6.2732e+11 43178
## - ID                   1 3.9751e+08 6.2734e+11 43178
## <none>                             6.2694e+11 43179
## - NumDealsPurchases    1 6.3360e+08 6.2757e+11 43179
## - Recency              1 6.5938e+08 6.2760e+11 43179
## - MntFruits            1 9.1904e+08 6.2786e+11 43180
## - AcceptedCmp1         1 9.3076e+08 6.2787e+11 43180
## - MntSweetProducts     1 1.1725e+09 6.2811e+11 43181
## - AcceptedCmp4         1 1.1734e+09 6.2811e+11 43181
## - AcceptedCmp5         1 1.2259e+09 6.2817e+11 43181
## - Kidhome              1 1.4750e+09 6.2841e+11 43182
## - NumStorePurchases    1 1.5071e+09 6.2845e+11 43182
## - NumCatalogPurchases  1 4.1714e+09 6.3111e+11 43192
## - Education            4 8.4213e+09 6.3536e+11 43200
## - Teenhome             1 1.0355e+10 6.3729e+11 43213
```

```
## - NumWebPurchases      1 1.0623e+10 6.3756e+11 43214
## - MntWines             1 1.0685e+10 6.3762e+11 43214
## - MntMeatProducts      1 1.2839e+10 6.3978e+11 43222
## - NumWebVisitsMonth    1 5.4742e+10 6.8168e+11 43362
##
## Step:  AIC=43177.11
## Income ~ ID + Year_Birth + Education + Kidhome + Teenhome + Recency +
##     MntWines + MntFruits + MntMeatProducts + MntFishProducts +
##     MntSweetProducts + NumDealsPurchases + NumWebPurchases +
##     NumCatalogPurchases + NumStorePurchases + NumWebVisitsMonth +
##     AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 + AcceptedCmp1 +
##     AcceptedCmp2 + Response
##
##                       Df  Sum of Sq        RSS   AIC
## - AcceptedCmp2         1 8.7550e+07 6.2711e+11 43175
## - MntFishProducts      1 1.5658e+08 6.2718e+11 43176
## - Response             1 1.5715e+08 6.2718e+11 43176
## - Year_Birth           1 2.0804e+08 6.2723e+11 43176
## - ID                   1 3.9617e+08 6.2742e+11 43177
## - AcceptedCmp3         1 4.2621e+08 6.2745e+11 43177
## <none>                              6.2702e+11 43177
## - NumDealsPurchases    1 6.7381e+08 6.2770e+11 43177
## - Recency              1 6.7534e+08 6.2770e+11 43177
## - MntFruits            1 8.7531e+08 6.2790e+11 43178
## - AcceptedCmp1         1 9.3493e+08 6.2796e+11 43178
## - MntSweetProducts     1 1.1754e+09 6.2820e+11 43179
## - AcceptedCmp5         1 1.2071e+09 6.2823e+11 43179
## - AcceptedCmp4         1 1.2405e+09 6.2826e+11 43179
## - NumStorePurchases    1 1.4900e+09 6.2851e+11 43180
## - Kidhome              1 1.5572e+09 6.2858e+11 43181
## - NumCatalogPurchases  1 4.0920e+09 6.3112e+11 43190
## - Education            4 8.5322e+09 6.3556e+11 43199
## - Teenhome             1 1.0363e+10 6.3739e+11 43211
## - NumWebPurchases      1 1.0601e+10 6.3762e+11 43212
## - MntWines             1 1.0614e+10 6.3764e+11 43212
## - MntMeatProducts      1 1.3038e+10 6.4006e+11 43221
## - NumWebVisitsMonth    1 5.4664e+10 6.8169e+11 43360
##
## Step:  AIC=43175.42
## Income ~ ID + Year_Birth + Education + Kidhome + Teenhome + Recency +
##     MntWines + MntFruits + MntMeatProducts + MntFishProducts +
##     MntSweetProducts + NumDealsPurchases + NumWebPurchases +
##     NumCatalogPurchases + NumStorePurchases + NumWebVisitsMonth +
##     AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 + AcceptedCmp1 +
##     Response
##
##                       Df  Sum of Sq        RSS   AIC
## - Response             1 1.4132e+08 6.2725e+11 43174
## - MntFishProducts      1 1.5287e+08 6.2726e+11 43174
## - Year_Birth           1 2.0559e+08 6.2732e+11 43174
## - ID                   1 3.9377e+08 6.2750e+11 43175
## - AcceptedCmp3         1 4.0885e+08 6.2752e+11 43175
## <none>                              6.2711e+11 43175
## - Recency              1 6.6926e+08 6.2778e+11 43176
```

```
## - NumDealsPurchases    1 6.8982e+08 6.2780e+11 43176
## - MntFruits            1 8.5772e+08 6.2797e+11 43176
## - AcceptedCmp1         1 9.6737e+08 6.2808e+11 43177
## - MntSweetProducts     1 1.1654e+09 6.2828e+11 43178
## - AcceptedCmp5         1 1.2642e+09 6.2837e+11 43178
## - AcceptedCmp4         1 1.4171e+09 6.2853e+11 43178
## - NumStorePurchases    1 1.5265e+09 6.2864e+11 43179
## - Kidhome              1 1.5448e+09 6.2866e+11 43179
## - NumCatalogPurchases  1 4.1492e+09 6.3126e+11 43188
## - Education            4 8.5264e+09 6.3564e+11 43197
## - Teenhome             1 1.0373e+10 6.3748e+11 43210
## - NumWebPurchases      1 1.0516e+10 6.3763e+11 43210
## - MntWines             1 1.0797e+10 6.3791e+11 43211
## - MntMeatProducts      1 1.2955e+10 6.4007e+11 43219
## - NumWebVisitsMonth    1 5.4577e+10 6.8169e+11 43358
##
## Step:  AIC=43173.92
## Income ~ ID + Year_Birth + Education + Kidhome + Teenhome + Recency +
##     MntWines + MntFruits + MntMeatProducts + MntFishProducts +
##     MntSweetProducts + NumDealsPurchases + NumWebPurchases +
##     NumCatalogPurchases + NumStorePurchases + NumWebVisitsMonth +
##     AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 + AcceptedCmp1
##
##                       Df  Sum of Sq        RSS   AIC
## - MntFishProducts      1 1.5553e+08 6.2741e+11 43172
## - Year_Birth           1 2.0178e+08 6.2745e+11 43173
## - ID                   1 4.0130e+08 6.2765e+11 43173
## - AcceptedCmp3         1 5.4573e+08 6.2780e+11 43174
## <none>                              6.2725e+11 43174
## - Recency              1 5.6798e+08 6.2782e+11 43174
## - NumDealsPurchases    1 7.3424e+08 6.2799e+11 43175
## - MntFruits            1 8.3422e+08 6.2809e+11 43175
## - AcceptedCmp1         1 8.8389e+08 6.2814e+11 43175
## - AcceptedCmp5         1 1.1589e+09 6.2841e+11 43176
## - MntSweetProducts     1 1.1610e+09 6.2841e+11 43176
## - AcceptedCmp4         1 1.3408e+09 6.2859e+11 43177
## - Kidhome              1 1.5555e+09 6.2881e+11 43177
## - NumStorePurchases    1 1.6477e+09 6.2890e+11 43178
## - NumCatalogPurchases  1 4.1046e+09 6.3136e+11 43186
## - Education            4 8.4073e+09 6.3566e+11 43195
## - NumWebPurchases      1 1.0414e+10 6.3767e+11 43208
## - Teenhome             1 1.0725e+10 6.3798e+11 43209
## - MntWines             1 1.0792e+10 6.3804e+11 43210
## - MntMeatProducts      1 1.2814e+10 6.4007e+11 43217
## - NumWebVisitsMonth    1 5.5743e+10 6.8299e+11 43361
##
## Step:  AIC=43172.47
## Income ~ ID + Year_Birth + Education + Kidhome + Teenhome + Recency +
##     MntWines + MntFruits + MntMeatProducts + MntSweetProducts +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1
##
##                       Df  Sum of Sq        RSS   AIC
```

```
## - Year_Birth           1 2.1416e+08 6.2762e+11 43171
## - ID                   1 3.8400e+08 6.2779e+11 43172
## - AcceptedCmp3         1 5.5954e+08 6.2797e+11 43172
## <none>                             6.2741e+11 43172
## - Recency              1 5.7685e+08 6.2798e+11 43173
## - NumDealsPurchases    1 7.6283e+08 6.2817e+11 43173
## - AcceptedCmp1         1 9.6166e+08 6.2837e+11 43174
## - MntFruits            1 1.0801e+09 6.2849e+11 43174
## - AcceptedCmp5         1 1.0905e+09 6.2850e+11 43174
## - AcceptedCmp4         1 1.2999e+09 6.2871e+11 43175
## - MntSweetProducts     1 1.4185e+09 6.2883e+11 43175
## - Kidhome              1 1.5393e+09 6.2895e+11 43176
## - NumStorePurchases    1 1.7347e+09 6.2914e+11 43177
## - NumCatalogPurchases  1 4.2774e+09 6.3169e+11 43186
## - Education            4 8.3083e+09 6.3572e+11 43194
## - NumWebPurchases      1 1.0567e+10 6.3797e+11 43207
## - Teenhome             1 1.0583e+10 6.3799e+11 43208
## - MntWines             1 1.0821e+10 6.3823e+11 43208
## - MntMeatProducts      1 1.3352e+10 6.4076e+11 43217
## - NumWebVisitsMonth    1 5.6388e+10 6.8380e+11 43361
##
## Step:  AIC=43171.23
## Income ~ ID + Education + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntSweetProducts + NumDealsPurchases +
##     NumWebPurchases + NumCatalogPurchases + NumStorePurchases +
##     NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
##     AcceptedCmp1
##
##                       Df  Sum of Sq        RSS   AIC
## - ID                   1 3.8267e+08 6.2800e+11 43171
## <none>                             6.2762e+11 43171
## - Recency              1 5.6793e+08 6.2819e+11 43171
## - AcceptedCmp3         1 6.0188e+08 6.2822e+11 43171
## - NumDealsPurchases    1 7.7949e+08 6.2840e+11 43172
## - AcceptedCmp1         1 9.7421e+08 6.2860e+11 43173
## - AcceptedCmp5         1 1.0644e+09 6.2869e+11 43173
## - MntFruits            1 1.0722e+09 6.2869e+11 43173
## - AcceptedCmp4         1 1.3002e+09 6.2892e+11 43174
## - MntSweetProducts     1 1.4026e+09 6.2902e+11 43174
## - Kidhome              1 1.4149e+09 6.2904e+11 43174
## - NumStorePurchases    1 1.6723e+09 6.2929e+11 43175
## - NumCatalogPurchases  1 4.3560e+09 6.3198e+11 43185
## - Education            4 8.5887e+09 6.3621e+11 43193
## - NumWebPurchases      1 1.0669e+10 6.3829e+11 43207
## - MntWines             1 1.0946e+10 6.3857e+11 43208
## - Teenhome             1 1.2599e+10 6.4022e+11 43213
## - MntMeatProducts      1 1.3273e+10 6.4090e+11 43216
## - NumWebVisitsMonth    1 5.7211e+10 6.8483e+11 43363
##
## Step:  AIC=43170.58
## Income ~ Education + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntSweetProducts + NumDealsPurchases +
##     NumWebPurchases + NumCatalogPurchases + NumStorePurchases +
##     NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
```

```
##     AcceptedCmp1
##
##                       Df  Sum of Sq        RSS   AIC
## <none>                              6.2800e+11 43171
## - Recency              1 6.1379e+08 6.2862e+11 43171
## - AcceptedCmp3         1 6.4367e+08 6.2865e+11 43171
## - NumDealsPurchases    1 8.2969e+08 6.2883e+11 43172
## - AcceptedCmp1         1 9.5624e+08 6.2896e+11 43172
## - AcceptedCmp5         1 1.0721e+09 6.2908e+11 43172
## - MntFruits            1 1.0855e+09 6.2909e+11 43172
## - AcceptedCmp4         1 1.2754e+09 6.2928e+11 43173
## - MntSweetProducts     1 1.3893e+09 6.2939e+11 43173
## - Kidhome              1 1.4290e+09 6.2943e+11 43174
## - NumStorePurchases    1 1.6675e+09 6.2967e+11 43174
## - NumCatalogPurchases  1 4.4202e+09 6.3242e+11 43184
## - Education            4 8.6324e+09 6.3664e+11 43193
## - NumWebPurchases      1 1.0670e+10 6.3867e+11 43206
## - MntWines             1 1.0915e+10 6.3892e+11 43207
## - Teenhome             1 1.2664e+10 6.4067e+11 43213
## - MntMeatProducts      1 1.3241e+10 6.4125e+11 43215
## - NumWebVisitsMonth    1 5.7130e+10 6.8513e+11 43362
```

```
#forward method to select variables
null<-lm(Income~1,data=data)
full<-lm(Income~.,data=data)
forward_data<-step(null,scope=list(lower=null,upper=full),direction="forward",k=2)#AIC method
```

```
## Start:  AIC=44912.81
## Income ~ 1
##
##                         Df  Sum of Sq        RSS   AIC
## + NumCatalogPurchases    1 4.8721e+11 9.1640e+11 43970
## + NumCatalogPurchases2   1 4.8721e+11 9.1640e+11 43970
## + MntMeatProducts        1 4.7975e+11 9.2386e+11 43988
## + MntWines               1 4.6998e+11 9.3363e+11 44011
## + NumWebVisitsMonth      1 4.2937e+11 9.7424e+11 44106
## + NumStorePurchases      1 3.9333e+11 1.0103e+12 44186
## + MntSweetProducts       1 2.7266e+11 1.1310e+12 44436
## + MntFishProducts        1 2.7035e+11 1.1333e+12 44441
## + MntFruits              1 2.6054e+11 1.1431e+12 44460
## + Kidhome                1 2.5792e+11 1.1457e+12 44465
## + NumWebPurchases        1 2.1117e+11 1.1924e+12 44553
## + AcceptedCmp5           1 1.5841e+11 1.2452e+12 44649
## + MntGoldProds           1 1.4909e+11 1.2545e+12 44666
## + AcceptedCmp1           1 1.0756e+11 1.2961e+12 44738
## + Education              4 6.7073e+10 1.3365e+12 44812
## + AcceptedCmp4           1 4.7728e+10 1.3559e+12 44838
## + Year_Birth             1 3.6742e+10 1.3669e+12 44856
## + Response               1 2.4846e+10 1.3788e+12 44875
## + AcceptedCmp2           1 1.0757e+10 1.3929e+12 44898
## + NumDealsPurchases      1 9.6930e+09 1.3939e+12 44899
## <none>                                1.4036e+12 44913
## + Complain               1 1.0403e+09 1.4026e+12 44913
## + Teenhome               1 5.1384e+08 1.4031e+12 44914
```

```
## + AcceptedCmp3           1 3.6720e+08 1.4032e+12 44914
## + ID                     1 2.4071e+08 1.4034e+12 44914
## + Recency                1 2.2119e+07 1.4036e+12 44915
## + Marital_Status         7 4.0388e+09 1.3996e+12 44920
## + Dt_Customer          661 5.1925e+11 8.8436e+11 45211
##
## Step:  AIC=43970.01
## Income ~ NumCatalogPurchases
##
##                       Df  Sum of Sq        RSS   AIC
## + NumWebVisitsMonth    1 1.1632e+11 8.0008e+11 43671
## + MntWines             1 9.8480e+10 8.1792e+11 43720
## + NumStorePurchases    1 9.6465e+10 8.1993e+11 43726
## + MntMeatProducts      1 7.0441e+10 8.4596e+11 43795
## + NumWebPurchases      1 4.2230e+10 8.7417e+11 43867
## + MntSweetProducts     1 4.1298e+10 8.7510e+11 43870
## + MntFruits            1 3.8306e+10 8.7809e+11 43877
## + AcceptedCmp5         1 3.3371e+10 8.8303e+11 43890
## + Kidhome              1 3.2527e+10 8.8387e+11 43892
## + MntFishProducts      1 3.0619e+10 8.8578e+11 43897
## + Education            4 2.8924e+10 8.8748e+11 43907
## + AcceptedCmp4         1 1.4841e+10 9.0156e+11 43936
## + AcceptedCmp1         1 1.3932e+10 9.0247e+11 43938
## + Year_Birth           1 1.1554e+10 9.0485e+11 43944
## + Teenhome             1 1.0399e+10 9.0600e+11 43947
## + AcceptedCmp3         1 8.5564e+09 9.0784e+11 43951
## + NumDealsPurchases    1 8.1002e+09 9.0830e+11 43952
## + MntGoldProds         1 7.4314e+09 9.0897e+11 43954
## + AcceptedCmp2         1 1.1660e+09 9.1523e+11 43969
## <none>                              9.1640e+11 43970
## + Recency              1 4.6304e+08 9.1594e+11 43971
## + Complain             1 3.1371e+08 9.1609e+11 43971
## + ID                   1 2.9249e+08 9.1611e+11 43971
## + Response             1 1.7881e+07 9.1638e+11 43972
## + Marital_Status       7 1.3874e+09 9.1501e+11 43981
## + Dt_Customer        661 3.7124e+11 5.4516e+11 44141
##
## Step:  AIC=43671.21
## Income ~ NumCatalogPurchases + NumWebVisitsMonth
##
##                       Df  Sum of Sq        RSS   AIC
## + MntWines             1 1.0157e+11 6.9851e+11 43372
## + NumWebPurchases      1 7.6160e+10 7.2392e+11 43452
## + NumStorePurchases    1 5.8007e+10 7.4207e+11 43506
## + MntMeatProducts      1 3.2422e+10 7.6766e+11 43582
## + Education            4 2.4486e+10 7.7559e+11 43610
## + AcceptedCmp4         1 1.9598e+10 7.8048e+11 43618
## + AcceptedCmp5         1 1.8954e+10 7.8112e+11 43620
## + Teenhome             1 1.7319e+10 7.8276e+11 43625
## + MntSweetProducts     1 1.7168e+10 7.8291e+11 43625
## + MntFruits            1 1.5207e+10 7.8487e+11 43631
## + AcceptedCmp1         1 1.0828e+10 7.8925e+11 43643
## + Kidhome              1 9.6429e+09 7.9043e+11 43646
## + MntFishProducts      1 9.6346e+09 7.9044e+11 43646
```

```
## + Year_Birth          1 6.9541e+09 7.9312e+11 43654
## + MntGoldProds        1 6.2031e+09 7.9387e+11 43656
## + AcceptedCmp2        1 2.7271e+09 7.9735e+11 43666
## + Response            1 2.5852e+09 7.9749e+11 43666
## + NumDealsPurchases   1 2.4988e+09 7.9758e+11 43666
## + AcceptedCmp3        1 2.1517e+09 7.9793e+11 43667
## <none>                             8.0008e+11 43671
## + Recency             1 5.7197e+08 7.9951e+11 43672
## + Complain            1 2.0024e+08 7.9988e+11 43673
## + ID                  1 1.8257e+08 7.9989e+11 43673
## + Marital_Status      7 2.0285e+09 7.9805e+11 43680
## + Dt_Customer       661 3.3572e+11 4.6435e+11 43788
##
## Step:  AIC=43372.37
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines
##
##                      Df  Sum of Sq        RSS   AIC
## + NumWebPurchases     1 2.3016e+10 6.7550e+11 43300
## + MntMeatProducts     1 1.3556e+10 6.8496e+11 43331
## + Education           4 1.3643e+10 6.8487e+11 43337
## + Teenhome            1 1.0222e+10 6.8829e+11 43342
## + NumStorePurchases   1 9.6568e+09 6.8886e+11 43344
## + MntSweetProducts    1 8.7711e+09 6.8974e+11 43346
## + MntFruits           1 7.2240e+09 6.9129e+11 43351
## + MntFishProducts     1 4.5616e+09 6.9395e+11 43360
## + Year_Birth          1 2.4134e+09 6.9610e+11 43367
## + NumDealsPurchases   1 1.9808e+09 6.9653e+11 43368
## + AcceptedCmp3        1 1.9155e+09 6.9660e+11 43368
## + AcceptedCmp1        1 1.3930e+09 6.9712e+11 43370
## + MntGoldProds        1 7.5873e+08 6.9775e+11 43372
## <none>                             6.9851e+11 43372
## + Recency             1 5.8213e+08 6.9793e+11 43373
## + AcceptedCmp4        1 5.4550e+08 6.9797e+11 43373
## + ID                  1 4.6537e+08 6.9805e+11 43373
## + AcceptedCmp5        1 4.1068e+08 6.9810e+11 43373
## + Kidhome             1 1.0085e+08 6.9841e+11 43374
## + AcceptedCmp2        1 4.7936e+07 6.9846e+11 43374
## + Response            1 3.5828e+07 6.9848e+11 43374
## + Complain            1 1.0744e+07 6.9850e+11 43374
## + Marital_Status      7 1.0549e+09 6.9746e+11 43383
## + Dt_Customer       661 3.0046e+11 3.9805e+11 43448
##
## Step:  AIC=43300.12
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases
##
##                      Df  Sum of Sq        RSS   AIC
## + MntMeatProducts     1 1.3445e+10 6.6205e+11 43258
## + Education           4 1.1390e+10 6.6411e+11 43270
## + Teenhome            1 5.4143e+09 6.7008e+11 43284
## + MntSweetProducts    1 4.1926e+09 6.7130e+11 43288
## + MntFruits           1 3.8695e+09 6.7163e+11 43289
## + NumStorePurchases   1 2.7810e+09 6.7272e+11 43293
## + MntFishProducts     1 2.1154e+09 6.7338e+11 43295
```

```
## + AcceptedCmp1        1 1.9389e+09 6.7356e+11 43296
## + AcceptedCmp5        1 1.7650e+09 6.7373e+11 43296
## + AcceptedCmp3        1 1.6270e+09 6.7387e+11 43297
## + Year_Birth          1 1.2184e+09 6.7428e+11 43298
## + AcceptedCmp4        1 1.0735e+09 6.7442e+11 43299
## <none>                             6.7550e+11 43300
## + ID                  1 5.1256e+08 6.7498e+11 43300
## + Recency             1 4.6325e+08 6.7503e+11 43301
## + Kidhome             1 3.5426e+08 6.7514e+11 43301
## + MntGoldProds        1 1.2413e+08 6.7537e+11 43302
## + NumDealsPurchases   1 1.0637e+08 6.7539e+11 43302
## + AcceptedCmp2        1 9.1520e+07 6.7540e+11 43302
## + Response            1 6.6018e+07 6.7543e+11 43302
## + Complain            1 1.4868e+07 6.7548e+11 43302
## + Marital_Status      7 7.1474e+08 6.7478e+11 43312
## + Dt_Customer       661 2.8898e+11 3.8651e+11 43385
##
## Step:  AIC=43257.57
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts
##
##                     Df  Sum of Sq        RSS   AIC
## + Teenhome            1 1.2372e+10 6.4968e+11 43218
## + Education           4 1.1376e+10 6.5067e+11 43227
## + Year_Birth          1 2.5799e+09 6.5947e+11 43251
## + NumStorePurchases   1 2.3715e+09 6.5968e+11 43252
## + AcceptedCmp4        1 1.8584e+09 6.6019e+11 43253
## + MntSweetProducts    1 1.5133e+09 6.6054e+11 43254
## + AcceptedCmp3        1 1.1877e+09 6.6086e+11 43256
## + AcceptedCmp1        1 1.1594e+09 6.6089e+11 43256
## + MntFruits           1 1.0737e+09 6.6098e+11 43256
## + AcceptedCmp5        1 7.6366e+08 6.6129e+11 43257
## <none>                             6.6205e+11 43258
## + ID                  1 5.3512e+08 6.6152e+11 43258
## + Recency             1 4.9097e+08 6.6156e+11 43258
## + NumDealsPurchases   1 3.5691e+08 6.6169e+11 43258
## + Kidhome             1 3.3302e+08 6.6172e+11 43258
## + AcceptedCmp2        1 3.2451e+08 6.6173e+11 43258
## + MntFishProducts     1 2.6586e+08 6.6179e+11 43259
## + MntGoldProds        1 1.8098e+08 6.6187e+11 43259
## + Response            1 5.6004e+07 6.6199e+11 43259
## + Complain            1 1.2268e+07 6.6204e+11 43260
## + Marital_Status      7 1.1088e+09 6.6094e+11 43268
## + Dt_Customer       661 2.8247e+11 3.7958e+11 43347
##
## Step:  AIC=43217.77
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome
##
##                     Df  Sum of Sq        RSS   AIC
## + Education           4 8.4063e+09 6.4127e+11 43197
## + MntSweetProducts    1 2.6073e+09 6.4707e+11 43211
## + AcceptedCmp1        1 2.1925e+09 6.4749e+11 43212
## + AcceptedCmp5        1 2.0848e+09 6.4759e+11 43213
```

```
## + MntFruits           1 2.0080e+09 6.4767e+11 43213
## + AcceptedCmp4        1 1.8261e+09 6.4785e+11 43214
## + NumStorePurchases   1 1.6326e+09 6.4805e+11 43214
## + MntFishProducts     1 1.0222e+09 6.4866e+11 43216
## + AcceptedCmp3        1 7.9899e+08 6.4888e+11 43217
## + Kidhome             1 7.8308e+08 6.4890e+11 43217
## + Recency             1 6.1417e+08 6.4906e+11 43218
## <none>                             6.4968e+11 43218
## + ID                  1 5.3409e+08 6.4914e+11 43218
## + NumDealsPurchases   1 4.8006e+08 6.4920e+11 43218
## + AcceptedCmp2        1 4.5124e+08 6.4923e+11 43218
## + Year_Birth          1 1.8814e+08 6.4949e+11 43219
## + MntGoldProds        1 1.1490e+08 6.4956e+11 43219
## + Response            1 5.1835e+07 6.4963e+11 43220
## + Complain            1 1.3830e+07 6.4966e+11 43220
## + Marital_Status      7 6.4948e+08 6.4903e+11 43230
## + Dt_Customer       661 2.7762e+11 3.7206e+11 43305
##
## Step:  AIC=43196.91
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education
##
##                      Df  Sum of Sq        RSS   AIC
## + MntSweetProducts    1 3.1731e+09 6.3810e+11 43188
## + MntFruits           1 2.3443e+09 6.3893e+11 43191
## + AcceptedCmp1        1 2.2486e+09 6.3902e+11 43191
## + AcceptedCmp5        1 2.1812e+09 6.3909e+11 43191
## + AcceptedCmp4        1 1.8378e+09 6.3943e+11 43193
## + NumStorePurchases   1 1.5595e+09 6.3971e+11 43194
## + MntFishProducts     1 1.3564e+09 6.3992e+11 43194
## + AcceptedCmp3        1 7.0692e+08 6.4057e+11 43196
## + Recency             1 6.1314e+08 6.4066e+11 43197
## <none>                             6.4127e+11 43197
## + NumDealsPurchases   1 4.9162e+08 6.4078e+11 43197
## + ID                  1 4.6879e+08 6.4080e+11 43197
## + Kidhome             1 4.5983e+08 6.4081e+11 43197
## + AcceptedCmp2        1 4.2999e+08 6.4084e+11 43197
## + Year_Birth          1 7.6200e+07 6.4120e+11 43199
## + MntGoldProds        1 5.2451e+07 6.4122e+11 43199
## + Complain            1 2.5275e+07 6.4125e+11 43199
## + Response            1 8.0319e+06 6.4126e+11 43199
## + Marital_Status      7 5.7199e+08 6.4070e+11 43209
## + Dt_Customer       661 2.7446e+11 3.6681e+11 43281
##
## Step:  AIC=43187.91
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts
##
##                      Df  Sum of Sq        RSS   AIC
## + AcceptedCmp4        1 2.1777e+09 6.3592e+11 43182
## + AcceptedCmp5        1 2.0210e+09 6.3608e+11 43183
## + AcceptedCmp1        1 1.9839e+09 6.3612e+11 43183
## + MntFruits           1 1.0535e+09 6.3705e+11 43186
```

```
## + NumStorePurchases   1 1.0191e+09 6.3708e+11 43186
## + Recency             1 6.5746e+08 6.3744e+11 43188
## + AcceptedCmp3        1 6.5440e+08 6.3744e+11 43188
## + Kidhome             1 6.2639e+08 6.3747e+11 43188
## <none>                             6.3810e+11 43188
## + AcceptedCmp2        1 5.0854e+08 6.3759e+11 43188
## + ID                  1 4.7910e+08 6.3762e+11 43188
## + MntFishProducts     1 4.2153e+08 6.3768e+11 43188
## + NumDealsPurchases   1 3.6599e+08 6.3773e+11 43189
## + MntGoldProds        1 1.4270e+08 6.3796e+11 43189
## + Year_Birth          1 9.2111e+07 6.3801e+11 43190
## + Complain            1 1.6467e+07 6.3808e+11 43190
## + Response            1 9.6725e+06 6.3809e+11 43190
## + Marital_Status      7 6.1742e+08 6.3748e+11 43200
## + Dt_Customer       661 2.7337e+11 3.6473e+11 43270
##
## Step:  AIC=43182.34
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4
##
##                      Df  Sum of Sq        RSS   AIC
## + AcceptedCmp1        1 1.4120e+09 6.3451e+11 43179
## + AcceptedCmp5        1 1.3484e+09 6.3457e+11 43180
## + MntFruits           1 1.3160e+09 6.3461e+11 43180
## + NumStorePurchases   1 1.0923e+09 6.3483e+11 43181
## + Kidhome             1 7.2124e+08 6.3520e+11 43182
## + Recency             1 6.9444e+08 6.3523e+11 43182
## <none>                             6.3592e+11 43182
## + MntFishProducts     1 5.6450e+08 6.3536e+11 43182
## + ID                  1 5.1341e+08 6.3541e+11 43183
## + AcceptedCmp3        1 4.1143e+08 6.3551e+11 43183
## + NumDealsPurchases   1 3.5098e+08 6.3557e+11 43183
## + AcceptedCmp2        1 1.5481e+08 6.3577e+11 43184
## + Year_Birth          1 8.2047e+07 6.3584e+11 43184
## + MntGoldProds        1 5.0952e+07 6.3587e+11 43184
## + Complain            1 1.0612e+07 6.3591e+11 43184
## + Response            1 7.0705e+06 6.3591e+11 43184
## + Marital_Status      7 6.6161e+08 6.3526e+11 43194
## + Dt_Customer       661 2.7208e+11 3.6384e+11 43267
##
## Step:  AIC=43179.41
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1
##
##                      Df  Sum of Sq        RSS   AIC
## + MntFruits           1 1.3959e+09 6.3311e+11 43177
## + NumStorePurchases   1 1.3763e+09 6.3313e+11 43177
## + AcceptedCmp5        1 8.1897e+08 6.3369e+11 43179
## + Kidhome             1 6.3304e+08 6.3388e+11 43179
## + Recency             1 6.2630e+08 6.3388e+11 43179
## + AcceptedCmp3        1 5.7879e+08 6.3393e+11 43179
## <none>                             6.3451e+11 43179
```

```
## + ID                  1 5.3437e+08 6.3397e+11 43180
## + MntFishProducts     1 4.7026e+08 6.3404e+11 43180
## + NumDealsPurchases   1 2.4721e+08 6.3426e+11 43181
## + Response            1 1.1074e+08 6.3440e+11 43181
## + AcceptedCmp2        1 8.2678e+07 6.3443e+11 43181
## + Year_Birth          1 8.0389e+07 6.3443e+11 43181
## + MntGoldProds        1 5.7235e+07 6.3445e+11 43181
## + Complain            1 7.8156e+06 6.3450e+11 43181
## + Marital_Status      7 6.7209e+08 6.3384e+11 43191
## + Dt_Customer       661 2.7103e+11 3.6348e+11 43267
##
## Step:  AIC=43176.53
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits
##
##                     Df  Sum of Sq        RSS   AIC
## + NumStorePurchases   1 1.0640e+09 6.3205e+11 43175
## + AcceptedCmp5        1 9.0664e+08 6.3221e+11 43175
## + Kidhome             1 7.4694e+08 6.3237e+11 43176
## + AcceptedCmp3        1 5.7974e+08 6.3253e+11 43177
## <none>                            6.3311e+11 43177
## + Recency             1 5.7071e+08 6.3254e+11 43177
## + ID                  1 5.1264e+08 6.3260e+11 43177
## + NumDealsPurchases   1 1.9469e+08 6.3292e+11 43178
## + MntGoldProds        1 1.6239e+08 6.3295e+11 43178
## + MntFishProducts     1 1.5694e+08 6.3296e+11 43178
## + Response            1 1.2393e+08 6.3299e+11 43178
## + AcceptedCmp2        1 1.0638e+08 6.3301e+11 43178
## + Year_Birth          1 8.4413e+07 6.3303e+11 43178
## + Complain            1 1.0202e+07 6.3310e+11 43178
## + Marital_Status      7 6.7725e+08 6.3244e+11 43188
## + Dt_Customer       661 2.7098e+11 3.6213e+11 43261
##
## Step:  AIC=43174.8
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits +
##     NumStorePurchases
##
##                     Df  Sum of Sq        RSS   AIC
## + AcceptedCmp5        1 1.2176e+09 6.3083e+11 43173
## + Kidhome             1 9.5770e+08 6.3109e+11 43173
## <none>                            6.3205e+11 43175
## + Recency             1 5.3920e+08 6.3151e+11 43175
## + ID                  1 5.2127e+08 6.3153e+11 43175
## + AcceptedCmp3        1 4.1373e+08 6.3164e+11 43175
## + NumDealsPurchases   1 3.5734e+08 6.3169e+11 43176
## + MntGoldProds        1 1.9588e+08 6.3185e+11 43176
## + Year_Birth          1 1.0691e+08 6.3194e+11 43176
## + AcceptedCmp2        1 9.5841e+07 6.3195e+11 43176
## + MntFishProducts     1 9.3910e+07 6.3196e+11 43176
## + Response            1 4.4966e+07 6.3200e+11 43177
## + Complain            1 1.1972e+07 6.3204e+11 43177
```

```
## + Marital_Status      7 7.1849e+08 6.3133e+11 43186
## + Dt_Customer       661 2.7024e+11 3.6181e+11 43261
##
## Step:  AIC=43172.53
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits +
##     NumStorePurchases + AcceptedCmp5
##
##                      Df  Sum of Sq        RSS   AIC
## + Kidhome             1 8.6149e+08 6.2997e+11 43172
## <none>                             6.3083e+11 43173
## + Recency             1 5.2032e+08 6.3031e+11 43173
## + AcceptedCmp3        1 5.1410e+08 6.3032e+11 43173
## + ID                  1 5.1087e+08 6.3032e+11 43173
## + NumDealsPurchases   1 2.9568e+08 6.3054e+11 43173
## + MntGoldProds        1 2.2169e+08 6.3061e+11 43174
## + MntFishProducts     1 1.7082e+08 6.3066e+11 43174
## + Response            1 1.6818e+08 6.3066e+11 43174
## + Year_Birth          1 1.3567e+08 6.3070e+11 43174
## + AcceptedCmp2        1 4.1806e+07 6.3079e+11 43174
## + Complain            1 1.7376e+07 6.3081e+11 43174
## + Marital_Status      7 6.9078e+08 6.3014e+11 43184
## + Dt_Customer       661 2.7041e+11 3.6042e+11 43254
##
## Step:  AIC=43171.5
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits +
##     NumStorePurchases + AcceptedCmp5 + Kidhome
##
##                      Df  Sum of Sq        RSS   AIC
## + NumDealsPurchases   1 7.4610e+08 6.2922e+11 43171
## <none>                             6.2997e+11 43172
## + Recency             1 5.6529e+08 6.2940e+11 43172
## + AcceptedCmp3        1 5.3609e+08 6.2943e+11 43172
## + ID                  1 5.1902e+08 6.2945e+11 43172
## + Year_Birth          1 2.6009e+08 6.2971e+11 43173
## + MntFishProducts     1 2.0008e+08 6.2977e+11 43173
## + Response            1 1.7096e+08 6.2980e+11 43173
## + MntGoldProds        1 1.6024e+08 6.2981e+11 43173
## + AcceptedCmp2        1 5.1810e+07 6.2992e+11 43173
## + Complain            1 2.4695e+07 6.2995e+11 43173
## + Marital_Status      7 6.7721e+08 6.2929e+11 43183
## + Dt_Customer       661 2.6976e+11 3.6021e+11 43255
##
## Step:  AIC=43170.88
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits +
##     NumStorePurchases + AcceptedCmp5 + Kidhome + NumDealsPurchases
##
##                    Df  Sum of Sq        RSS   AIC
## + AcceptedCmp3      1 6.0586e+08 6.2862e+11 43171
```

```
## + Recency           1 5.7598e+08 6.2865e+11 43171
## <none>                           6.2922e+11 43171
## + ID                1 4.6976e+08 6.2875e+11 43171
## + Year_Birth        1 2.4528e+08 6.2898e+11 43172
## + MntFishProducts   1 1.7210e+08 6.2905e+11 43172
## + Response          1 1.3516e+08 6.2909e+11 43172
## + MntGoldProds      1 1.1756e+08 6.2911e+11 43172
## + AcceptedCmp2      1 4.1088e+07 6.2918e+11 43173
## + Complain          1 2.9611e+07 6.2919e+11 43173
## + Marital_Status    7 6.7968e+08 6.2854e+11 43182
## + Dt_Customer     661 2.6957e+11 3.5966e+11 43253
##
## Step:  AIC=43170.74
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits +
##     NumStorePurchases + AcceptedCmp5 + Kidhome + NumDealsPurchases +
##     AcceptedCmp3
##
##                    Df  Sum of Sq        RSS   AIC
## + Recency           1 6.1379e+08 6.2800e+11 43171
## <none>                           6.2862e+11 43171
## + ID                1 4.2853e+08 6.2819e+11 43171
## + Year_Birth        1 2.0351e+08 6.2841e+11 43172
## + MntFishProducts   1 1.5758e+08 6.2846e+11 43172
## + MntGoldProds      1 6.7457e+07 6.2855e+11 43173
## + AcceptedCmp2      1 6.5035e+07 6.2855e+11 43173
## + Response          1 4.0729e+07 6.2858e+11 43173
## + Complain          1 2.7771e+07 6.2859e+11 43173
## + Marital_Status    7 6.5764e+08 6.2796e+11 43182
## + Dt_Customer     661 2.6930e+11 3.5932e+11 43253
##
## Step:  AIC=43170.58
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + Education +
##     MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits +
##     NumStorePurchases + AcceptedCmp5 + Kidhome + NumDealsPurchases +
##     AcceptedCmp3 + Recency
##
##                    Df  Sum of Sq        RSS   AIC
## <none>                           6.2800e+11 43171
## + ID                1 3.8267e+08 6.2762e+11 43171
## + Year_Birth        1 2.1283e+08 6.2779e+11 43172
## + MntFishProducts   1 1.4991e+08 6.2785e+11 43172
## + Response          1 1.4730e+08 6.2786e+11 43172
## + AcceptedCmp2      1 6.3860e+07 6.2794e+11 43172
## + MntGoldProds      1 5.9088e+07 6.2795e+11 43172
## + Complain          1 2.4396e+07 6.2798e+11 43172
## + Marital_Status    7 6.6590e+08 6.2734e+11 43182
## + Dt_Customer     661 2.6938e+11 3.5862e+11 43251
```

Even though the system helped to select a lot of predictors for us, most of the predictors don't have much impacts because AIC is pretty close. For example:

Step: AIC=43196.91 Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines + NumWebPurchases + MntMeatProducts + Teenhome + Education

```
                 Df  Sum of Sq        RSS   AIC
```

- MntSweetProducts 1 3.1731e+09 6.3810e+11 43188
- MntFruits 1 2.3443e+09 6.3893e+11 43191
- AcceptedCmp1 1 2.2486e+09 6.3902e+11 43191
- AcceptedCmp5 1 2.1812e+09 6.3909e+11 43191
- AcceptedCmp4 1 1.8378e+09 6.3943e+11 43193
- NumStorePurchases 1 1.5595e+09 6.3971e+11 43194
- MntFishProducts 1 1.3564e+09 6.3992e+11 43194
- AcceptedCmp3 1 7.0692e+08 6.4057e+11 43196
- Recency 1 6.1314e+08 6.4066e+11 43197 6.4127e+11 43197
- NumDealsPurchases 1 4.9162e+08 6.4078e+11 43197
- NumDealsPurchases2 1 4.9162e+08 6.4078e+11 43197
- ID 1 4.6879e+08 6.4080e+11 43197
- Kidhome 1 4.5983e+08 6.4081e+11 43197
- AcceptedCmp2 1 4.2999e+08 6.4084e+11 43197
- Year_Birth 1 7.6200e+07 6.4120e+11 43199
- MntGoldProds 1 5.2451e+07 6.4122e+11 43199
- Complain 1 2.5275e+07 6.4125e+11 43199

when the seventh variable is selected, the AIC is very close. AIC without more predictors is 43197, is just a little larger than the smallest AIC, 43188. Meanwhile, the AICs of many new predictors is the same (MntFruits, AcceptedCmp1, AcceptedCmp5 are all 43191).

Forward analysis

Here we observed that the NumCatalogPurchase attribute is so important, as it was selected during the first round when the model was null, which can minimize the AIC value at that time, so we consider to use it in our model.

```r
fit.model6 <- lmer(Income~1+NumCatalogPurchases+(1|Education), data=data)
summary(fit.model6)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Income ~ 1 + NumCatalogPurchases + (1 | Education)
##    Data: data
##
## REML criterion at convergence: 50178.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.7073 -0.4799 -0.0196  0.4529 31.0634
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  Education (Intercept)  85702768  9258
##  Residual              401602496 20040
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
```

```
##                      Estimate Std. Error t value
## (Intercept)          35263.9     4207.6   8.381
## NumCatalogPurchases   4918.0      146.7  33.517
##
## Correlation of Fixed Effects:
##             (Intr)
## NmCtlgPrchs -0.078
```

Because most variables have little effect, we just choose the best four predictors and end.

```
fit.model7 <- lmer(Income~1+NumCatalogPurchases+NumWebVisitsMonth+MntWines+NumWebPurchases+(1|Education)
summary(fit.model7)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Income ~ 1 + NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + (1 | Education)
##    Data: data
##
## REML criterion at convergence: 49507.2
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -5.237 -0.365 -0.012  0.334 36.087
##
## Random effects:
##  Groups    Name        Variance  Std.Dev.
##  Education (Intercept)  33042440  5748
##  Residual              300954683 17348
## Number of obs: 2216, groups:  Education, 5
##
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)         54741.107   2903.926  18.851
## NumCatalogPurchases  1403.169    183.417   7.650
## NumWebVisitsMonth   -3822.143    182.494 -20.944
## MntWines               19.399      1.604  12.093
## NumWebPurchases      1385.874    165.841   8.357
##
## Correlation of Fixed Effects:
##             (Intr) NmCtlP NmWbVM MntWns
## NmCtlgPrchs -0.208
## NmWbVstsMnt -0.381  0.447
## MntWines    -0.009 -0.457  0.075
## NumWbPrchss -0.058 -0.135 -0.208 -0.434
```

```
drop1(fit.model7, test="Chisq")
```

```
## Single term deletions
##
## Model:
## Income ~ 1 + NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + (1 | Education)
```

```
##                      npar   AIC    LRT   Pr(Chi)    
## <none>                    49577                     
## NumCatalogPurchases    1 49633  57.86 2.815e-14 ***
## NumWebVisitsMonth      1 49976 401.20 < 2.2e-16 ***
## MntWines               1 49717 142.07 < 2.2e-16 ***
## NumWebPurchases        1 49644  69.03 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We use 'drop1()' to test should we keep them in the model. We should keep the fixed effects for 4 variables since the p-values are less than 0.05.

```r
fit.model8 <- lmer(Income~1+NumCatalogPurchases+NumWebVisitsMonth+(1|Education)+(1|Year_Birth), data=da
summary(fit.model8)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: 
## Income ~ 1 + NumCatalogPurchases + NumWebVisitsMonth + (1 | Education) +  
##     (1 | Year_Birth)
##    Data: data
## 
## REML criterion at convergence: 49865.9
## 
## Scaled residuals: 
##    Min     1Q Median     3Q    Max 
## -8.055 -0.453 -0.011  0.425 33.168 
## 
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Year_Birth (Intercept)   2177090  1475   
##  Education  (Intercept)  66644611  8164   
##  Residual               349091394 18684   
## Number of obs: 2216, groups:  Year_Birth, 59; Education, 5
## 
## Fixed effects:
##                      Estimate Std. Error t value
## (Intercept)           57957.9     3932.2   14.74
## NumCatalogPurchases    3428.7      160.2   21.40
## NumWebVisitsMonth     -3447.9      192.8  -17.89
## 
## Correlation of Fixed Effects:
##             (Intr) NmCtlP
## NmCtlgPrchs -0.233       
## NmWbVstsMnt -0.321  0.514
```

```r
performance::icc(fit.model8)
```

```
## # Intraclass Correlation Coefficient
## 
##     Adjusted ICC: 0.165
##   Unadjusted ICC: 0.102
```

However, more random predictors may make the model singular. It is difficult to add all of fixed effects and fixed effects to a model at the same time. In this case, we try to use only two fixed predictors and two random intercepts to build a new model. Models with more random effects are better, models with more fixed effects are better, or a combination of both are better? We aren't sure. We will compare the effects later. Then we want to analyse the multicollinearity first to see whrther we can find a better model.

Select only the text columns for encoding

Because in vif and eigen value analysis, we cannot have those labeled columns exist in the dataset, so here we decide to encode the data into numeric value first. And then we will apply those three different stepwise method in order to select the attributes we need for further analysis.

```
text_columns <- select_if(data, is.character)

# Encode text columns
encoded <- text_columns %>%
  mutate(across(everything(), as.factor)) %>%
  mutate(across(everything(), as.numeric))

# Select numeric columns
numeric_columns <- select_if(data, is.numeric)

# Combine non-text and numeric columns
new_data <- bind_cols(numeric_columns, encoded)
#backward method to select variables
full<-lm(Income~.,data=new_data)
back_data<-step(full,data=new_data,direction="backward",k=2)#AIC method
```

```
## Start:  AIC=43199.55
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     MntGoldProds + NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Z_CostContact +
##     Z_Revenue + Response + NumCatalogPurchases2 + Education +
##     Marital_Status + Dt_Customer
##
##
## Step:  AIC=43199.55
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     MntGoldProds + NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Z_CostContact +
##     Z_Revenue + Response + Education + Marital_Status + Dt_Customer
##
##
## Step:  AIC=43199.55
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     MntGoldProds + NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Z_CostContact +
```

```
##     Response + Education + Marital_Status + Dt_Customer
##
##
## Step:  AIC=43199.55
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     MntGoldProds + NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Response +
##     Education + Marital_Status + Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - Marital_Status       1 8.7875e+06 6.3284e+11 43198
## - Complain             1 1.8801e+07 6.3285e+11 43198
## - MntGoldProds         1 5.9292e+07 6.3289e+11 43198
## - Response             1 9.2140e+07 6.3293e+11 43198
## - AcceptedCmp2         1 1.1777e+08 6.3295e+11 43198
## - MntFishProducts      1 1.8554e+08 6.3302e+11 43198
## - Year_Birth           1 2.9028e+08 6.3312e+11 43199
## - ID                   1 4.2368e+08 6.3326e+11 43199
## - Dt_Customer          1 4.5488e+08 6.3329e+11 43199
## - AcceptedCmp3         1 4.6610e+08 6.3330e+11 43199
## <none>                              6.3283e+11 43200
## - Recency              1 6.1141e+08 6.3344e+11 43200
## - NumDealsPurchases    1 7.4197e+08 6.3358e+11 43200
## - MntFruits            1 9.2007e+08 6.3375e+11 43201
## - AcceptedCmp1         1 9.8868e+08 6.3382e+11 43201
## - AcceptedCmp5         1 1.1547e+09 6.3399e+11 43202
## - AcceptedCmp4         1 1.1960e+09 6.3403e+11 43202
## - MntSweetProducts     1 1.2046e+09 6.3404e+11 43202
## - NumStorePurchases    1 1.8284e+09 6.3466e+11 43204
## - Kidhome              1 1.9594e+09 6.3479e+11 43204
## - Education            1 2.0440e+09 6.3488e+11 43205
## - NumCatalogPurchases  1 4.2636e+09 6.3710e+11 43212
## - MntWines             1 1.0559e+10 6.4339e+11 43234
## - NumWebPurchases      1 1.1390e+10 6.4422e+11 43237
## - Teenhome             1 1.2197e+10 6.4503e+11 43240
## - MntMeatProducts      1 1.3690e+10 6.4652e+11 43245
## - NumWebVisitsMonth    1 5.6085e+10 6.8892e+11 43386
##
## Step:  AIC=43197.58
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     MntGoldProds + NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Complain + Response +
##     Education + Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - Complain             1 1.8987e+07 6.3286e+11 43196
## - MntGoldProds         1 5.9481e+07 6.3290e+11 43196
## - Response             1 9.2751e+07 6.3293e+11 43196
## - AcceptedCmp2         1 1.1932e+08 6.3296e+11 43196
## - MntFishProducts      1 1.8855e+08 6.3303e+11 43196
```

```
## - Year_Birth           1 2.9703e+08 6.3314e+11 43197
## - ID                   1 4.2607e+08 6.3327e+11 43197
## - Dt_Customer          1 4.5717e+08 6.3330e+11 43197
## - AcceptedCmp3         1 4.6814e+08 6.3331e+11 43197
## <none>                              6.3284e+11 43198
## - Recency              1 6.1047e+08 6.3345e+11 43198
## - NumDealsPurchases    1 7.4311e+08 6.3359e+11 43198
## - MntFruits            1 9.1637e+08 6.3376e+11 43199
## - AcceptedCmp1         1 9.8397e+08 6.3383e+11 43199
## - AcceptedCmp5         1 1.1563e+09 6.3400e+11 43200
## - AcceptedCmp4         1 1.1988e+09 6.3404e+11 43200
## - MntSweetProducts     1 1.2055e+09 6.3405e+11 43200
## - NumStorePurchases    1 1.8248e+09 6.3467e+11 43202
## - Kidhome              1 1.9592e+09 6.3480e+11 43202
## - Education            1 2.0445e+09 6.3489e+11 43203
## - NumCatalogPurchases  1 4.2607e+09 6.3710e+11 43210
## - MntWines             1 1.0556e+10 6.4340e+11 43232
## - NumWebPurchases      1 1.1392e+10 6.4423e+11 43235
## - Teenhome             1 1.2191e+10 6.4503e+11 43238
## - MntMeatProducts      1 1.3714e+10 6.4656e+11 43243
## - NumWebVisitsMonth    1 5.6090e+10 6.8893e+11 43384
##
## Step:  AIC=43195.65
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     MntGoldProds + NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Response + Education +
##     Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - MntGoldProds         1 5.7603e+07 6.3292e+11 43194
## - Response             1 9.4221e+07 6.3296e+11 43194
## - AcceptedCmp2         1 1.1955e+08 6.3298e+11 43194
## - MntFishProducts      1 1.8964e+08 6.3305e+11 43194
## - Year_Birth           1 2.9018e+08 6.3315e+11 43195
## - ID                   1 4.2061e+08 6.3328e+11 43195
## - Dt_Customer          1 4.5549e+08 6.3332e+11 43195
## - AcceptedCmp3         1 4.7010e+08 6.3333e+11 43195
## <none>                              6.3286e+11 43196
## - Recency              1 6.1470e+08 6.3348e+11 43196
## - NumDealsPurchases    1 7.3985e+08 6.3360e+11 43196
## - MntFruits            1 9.1240e+08 6.3377e+11 43197
## - AcceptedCmp1         1 9.8906e+08 6.3385e+11 43197
## - AcceptedCmp5         1 1.1519e+09 6.3401e+11 43198
## - AcceptedCmp4         1 1.2043e+09 6.3407e+11 43198
## - MntSweetProducts     1 1.2110e+09 6.3407e+11 43198
## - NumStorePurchases    1 1.8168e+09 6.3468e+11 43200
## - Kidhome              1 1.9480e+09 6.3481e+11 43200
## - Education            1 2.0741e+09 6.3494e+11 43201
## - NumCatalogPurchases  1 4.2525e+09 6.3711e+11 43208
## - MntWines             1 1.0577e+10 6.4344e+11 43230
## - NumWebPurchases      1 1.1384e+10 6.4425e+11 43233
## - Teenhome             1 1.2193e+10 6.4505e+11 43236
```

```
## - MntMeatProducts      1 1.3717e+10 6.4658e+11 43241
## - NumWebVisitsMonth    1 5.6113e+10 6.8897e+11 43382
##
## Step:  AIC=43193.85
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Response + Education +
##     Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - Response             1 1.0007e+08 6.3302e+11 43192
## - AcceptedCmp2         1 1.1622e+08 6.3303e+11 43192
## - MntFishProducts      1 1.6436e+08 6.3308e+11 43192
## - Year_Birth           1 2.9185e+08 6.3321e+11 43193
## - ID                   1 4.1911e+08 6.3334e+11 43193
## - Dt_Customer          1 4.6026e+08 6.3338e+11 43193
## - AcceptedCmp3         1 5.0849e+08 6.3343e+11 43194
## <none>                              6.3292e+11 43194
## - Recency              1 6.2774e+08 6.3355e+11 43194
## - NumDealsPurchases    1 7.7738e+08 6.3370e+11 43195
## - MntFruits            1 8.7634e+08 6.3380e+11 43195
## - AcceptedCmp1         1 9.9247e+08 6.3391e+11 43195
## - AcceptedCmp5         1 1.1355e+09 6.3405e+11 43196
## - MntSweetProducts     1 1.2136e+09 6.3413e+11 43196
## - AcceptedCmp4         1 1.2599e+09 6.3418e+11 43196
## - NumStorePurchases    1 1.8012e+09 6.3472e+11 43198
## - Kidhome              1 2.0286e+09 6.3495e+11 43199
## - Education            1 2.2041e+09 6.3512e+11 43200
## - NumCatalogPurchases  1 4.1953e+09 6.3711e+11 43206
## - MntWines             1 1.0527e+10 6.4345e+11 43228
## - NumWebPurchases      1 1.1429e+10 6.4435e+11 43232
## - Teenhome             1 1.2195e+10 6.4511e+11 43234
## - MntMeatProducts      1 1.3876e+10 6.4680e+11 43240
## - NumWebVisitsMonth    1 5.6056e+10 6.8897e+11 43380
##
## Step:  AIC=43192.2
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + AcceptedCmp2 + Education +
##     Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - AcceptedCmp2         1 1.0151e+08 6.3312e+11 43191
## - MntFishProducts      1 1.6656e+08 6.3319e+11 43191
## - Year_Birth           1 2.8731e+08 6.3331e+11 43191
## - ID                   1 4.2545e+08 6.3344e+11 43192
## - Dt_Customer          1 4.5678e+08 6.3348e+11 43192
## - Recency              1 5.4851e+08 6.3357e+11 43192
## <none>                              6.3302e+11 43192
## - AcceptedCmp3         1 6.3767e+08 6.3366e+11 43192
```

```
## - NumDealsPurchases    1 8.1824e+08 6.3384e+11 43193
## - MntFruits            1 8.5624e+08 6.3387e+11 43193
## - AcceptedCmp1         1 9.2566e+08 6.3394e+11 43193
## - AcceptedCmp5         1 1.0573e+09 6.3408e+11 43194
## - AcceptedCmp4         1 1.2084e+09 6.3423e+11 43194
## - MntSweetProducts     1 1.2089e+09 6.3423e+11 43194
## - NumStorePurchases    1 1.9184e+09 6.3494e+11 43197
## - Kidhome              1 2.0360e+09 6.3505e+11 43197
## - Education            1 2.1342e+09 6.3515e+11 43198
## - NumCatalogPurchases  1 4.1602e+09 6.3718e+11 43205
## - MntWines             1 1.0528e+10 6.4355e+11 43227
## - NumWebPurchases      1 1.1341e+10 6.4436e+11 43230
## - Teenhome             1 1.2534e+10 6.4555e+11 43234
## - MntMeatProducts      1 1.3782e+10 6.4680e+11 43238
## - NumWebVisitsMonth    1 5.7071e+10 6.9009e+11 43381
##
## Step:  AIC=43190.56
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntFishProducts + MntSweetProducts +
##     NumDealsPurchases + NumWebPurchases + NumCatalogPurchases +
##     NumStorePurchases + NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 +
##     AcceptedCmp5 + AcceptedCmp1 + Education + Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - MntFishProducts      1 1.6241e+08 6.3328e+11 43189
## - Year_Birth           1 2.8401e+08 6.3340e+11 43190
## - ID                   1 4.2319e+08 6.3354e+11 43190
## - Dt_Customer          1 4.4592e+08 6.3357e+11 43190
## - Recency              1 5.4972e+08 6.3367e+11 43190
## <none>                            6.3312e+11 43191
## - AcceptedCmp3         1 6.0763e+08 6.3373e+11 43191
## - NumDealsPurchases    1 8.3471e+08 6.3395e+11 43191
## - MntFruits            1 8.3962e+08 6.3396e+11 43191
## - AcceptedCmp1         1 9.6833e+08 6.3409e+11 43192
## - AcceptedCmp5         1 1.1262e+09 6.3425e+11 43192
## - MntSweetProducts     1 1.1986e+09 6.3432e+11 43193
## - AcceptedCmp4         1 1.4024e+09 6.3452e+11 43193
## - NumStorePurchases    1 1.9550e+09 6.3508e+11 43195
## - Kidhome              1 2.0204e+09 6.3514e+11 43196
## - Education            1 2.1176e+09 6.3524e+11 43196
## - NumCatalogPurchases  1 4.2292e+09 6.3735e+11 43203
## - MntWines             1 1.0715e+10 6.4384e+11 43226
## - NumWebPurchases      1 1.1246e+10 6.4437e+11 43228
## - Teenhome             1 1.2534e+10 6.4565e+11 43232
## - MntMeatProducts      1 1.3694e+10 6.4681e+11 43236
## - NumWebVisitsMonth    1 5.6974e+10 6.9009e+11 43380
##
## Step:  AIC=43189.12
## Income ~ ID + Year_Birth + Kidhome + Teenhome + Recency + MntWines +
##     MntFruits + MntMeatProducts + MntSweetProducts + NumDealsPurchases +
##     NumWebPurchases + NumCatalogPurchases + NumStorePurchases +
##     NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
##     AcceptedCmp1 + Education + Dt_Customer
##
```

```
##                       Df  Sum of Sq        RSS   AIC
## - Year_Birth           1 2.9914e+08 6.3358e+11 43188
## - ID                   1 4.0490e+08 6.3369e+11 43189
## - Dt_Customer          1 4.5481e+08 6.3374e+11 43189
## - Recency              1 5.5832e+08 6.3384e+11 43189
## <none>                             6.3328e+11 43189
## - AcceptedCmp3         1 6.2275e+08 6.3391e+11 43189
## - NumDealsPurchases    1 8.6615e+08 6.3415e+11 43190
## - AcceptedCmp1         1 1.0520e+09 6.3433e+11 43191
## - AcceptedCmp5         1 1.0570e+09 6.3434e+11 43191
## - MntFruits            1 1.0924e+09 6.3438e+11 43191
## - AcceptedCmp4         1 1.3602e+09 6.3464e+11 43192
## - MntSweetProducts     1 1.4663e+09 6.3475e+11 43192
## - Kidhome              1 2.0022e+09 6.3528e+11 43194
## - Education            1 2.0063e+09 6.3529e+11 43194
## - NumStorePurchases    1 2.0533e+09 6.3534e+11 43194
## - NumCatalogPurchases  1 4.4065e+09 6.3769e+11 43202
## - MntWines             1 1.0742e+10 6.4402e+11 43224
## - NumWebPurchases      1 1.1410e+10 6.4469e+11 43227
## - Teenhome             1 1.2383e+10 6.4567e+11 43230
## - MntMeatProducts      1 1.4279e+10 6.4756e+11 43237
## - NumWebVisitsMonth    1 5.7635e+10 6.9092e+11 43380
##
## Step:  AIC=43188.17
## Income ~ ID + Kidhome + Teenhome + Recency + MntWines + MntFruits +
##     MntMeatProducts + MntSweetProducts + NumDealsPurchases +
##     NumWebPurchases + NumCatalogPurchases + NumStorePurchases +
##     NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
##     AcceptedCmp1 + Education + Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - ID                   1 4.0279e+08 6.3398e+11 43188
## - Dt_Customer          1 4.5059e+08 6.3403e+11 43188
## - Recency              1 5.4905e+08 6.3413e+11 43188
## <none>                             6.3358e+11 43188
## - AcceptedCmp3         1 6.7599e+08 6.3426e+11 43189
## - NumDealsPurchases    1 8.8622e+08 6.3447e+11 43189
## - AcceptedCmp5         1 1.0263e+09 6.3461e+11 43190
## - AcceptedCmp1         1 1.0671e+09 6.3465e+11 43190
## - MntFruits            1 1.0814e+09 6.3466e+11 43190
## - AcceptedCmp4         1 1.3606e+09 6.3494e+11 43191
## - MntSweetProducts     1 1.4455e+09 6.3503e+11 43191
## - Kidhome              1 1.8357e+09 6.3542e+11 43193
## - NumStorePurchases    1 1.9750e+09 6.3556e+11 43193
## - Education            1 2.2086e+09 6.3579e+11 43194
## - NumCatalogPurchases  1 4.4983e+09 6.3808e+11 43202
## - MntWines             1 1.0893e+10 6.4448e+11 43224
## - NumWebPurchases      1 1.1539e+10 6.4512e+11 43226
## - MntMeatProducts      1 1.4183e+10 6.4776e+11 43235
## - Teenhome             1 1.4906e+10 6.4849e+11 43238
## - NumWebVisitsMonth    1 5.8630e+10 6.9221e+11 43382
##
## Step:  AIC=43187.58
## Income ~ Kidhome + Teenhome + Recency + MntWines + MntFruits +
```

```
##     MntMeatProducts + MntSweetProducts + NumDealsPurchases +
##     NumWebPurchases + NumCatalogPurchases + NumStorePurchases +
##     NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
##     AcceptedCmp1 + Education + Dt_Customer
##
##                       Df  Sum of Sq        RSS   AIC
## - Dt_Customer          1 4.7732e+08 6.3446e+11 43187
## <none>                               6.3398e+11 43188
## - Recency              1 5.9397e+08 6.3458e+11 43188
## - AcceptedCmp3         1 7.2171e+08 6.3471e+11 43188
## - NumDealsPurchases    1 9.4355e+08 6.3493e+11 43189
## - AcceptedCmp5         1 1.0331e+09 6.3502e+11 43189
## - AcceptedCmp1         1 1.0492e+09 6.3503e+11 43189
## - MntFruits            1 1.0960e+09 6.3508e+11 43189
## - AcceptedCmp4         1 1.3351e+09 6.3532e+11 43190
## - MntSweetProducts     1 1.4338e+09 6.3542e+11 43191
## - Kidhome              1 1.8531e+09 6.3584e+11 43192
## - NumStorePurchases    1 1.9703e+09 6.3595e+11 43192
## - Education            1 2.2064e+09 6.3619e+11 43193
## - NumCatalogPurchases  1 4.5665e+09 6.3855e+11 43201
## - MntWines             1 1.0857e+10 6.4484e+11 43223
## - NumWebPurchases      1 1.1545e+10 6.4553e+11 43226
## - MntMeatProducts      1 1.4155e+10 6.4814e+11 43235
## - Teenhome             1 1.4991e+10 6.4898e+11 43237
## - NumWebVisitsMonth    1 5.8531e+10 6.9252e+11 43381
##
## Step:  AIC=43187.25
## Income ~ Kidhome + Teenhome + Recency + MntWines + MntFruits +
##     MntMeatProducts + MntSweetProducts + NumDealsPurchases +
##     NumWebPurchases + NumCatalogPurchases + NumStorePurchases +
##     NumWebVisitsMonth + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5 +
##     AcceptedCmp1 + Education
##
##                       Df  Sum of Sq        RSS   AIC
## <none>                               6.3446e+11 43187
## - Recency              1 6.0736e+08 6.3507e+11 43187
## - AcceptedCmp3         1 7.2110e+08 6.3518e+11 43188
## - NumDealsPurchases    1 9.1452e+08 6.3538e+11 43188
## - AcceptedCmp1         1 1.0282e+09 6.3549e+11 43189
## - AcceptedCmp5         1 1.0827e+09 6.3554e+11 43189
## - MntFruits            1 1.1245e+09 6.3559e+11 43189
## - AcceptedCmp4         1 1.3263e+09 6.3579e+11 43190
## - MntSweetProducts     1 1.4061e+09 6.3587e+11 43190
## - Kidhome              1 1.8688e+09 6.3633e+11 43192
## - NumStorePurchases    1 1.9662e+09 6.3643e+11 43192
## - Education            1 2.1750e+09 6.3664e+11 43193
## - NumCatalogPurchases  1 4.6485e+09 6.3911e+11 43201
## - MntWines             1 1.0801e+10 6.4526e+11 43223
## - NumWebPurchases      1 1.1529e+10 6.4599e+11 43225
## - MntMeatProducts      1 1.4079e+10 6.4854e+11 43234
## - Teenhome             1 1.4974e+10 6.4944e+11 43237
## - NumWebVisitsMonth    1 5.8796e+10 6.9326e+11 43382
```

```r
#forward method to select variables
null<-lm(Income~1,data=new_data)
full<-lm(Income~.,data=new_data)
b_model<-step(null,scope=list(upper=full),data=data,direction = "both")
```

```
## Start:  AIC=44912.81
## Income ~ 1
##
##                        Df  Sum of Sq        RSS   AIC
## + NumCatalogPurchases   1 4.8721e+11 9.1640e+11 43970
## + NumCatalogPurchases2  1 4.8721e+11 9.1640e+11 43970
## + MntMeatProducts       1 4.7975e+11 9.2386e+11 43988
## + MntWines              1 4.6998e+11 9.3363e+11 44011
## + NumWebVisitsMonth     1 4.2937e+11 9.7424e+11 44106
## + NumStorePurchases     1 3.9333e+11 1.0103e+12 44186
## + MntSweetProducts      1 2.7266e+11 1.1310e+12 44436
## + MntFishProducts       1 2.7035e+11 1.1333e+12 44441
## + MntFruits             1 2.6054e+11 1.1431e+12 44460
## + Kidhome               1 2.5792e+11 1.1457e+12 44465
## + NumWebPurchases       1 2.1117e+11 1.1924e+12 44553
## + AcceptedCmp5          1 1.5841e+11 1.2452e+12 44649
## + MntGoldProds          1 1.4909e+11 1.2545e+12 44666
## + AcceptedCmp1          1 1.0756e+11 1.2961e+12 44738
## + AcceptedCmp4          1 4.7728e+10 1.3559e+12 44838
## + Year_Birth            1 3.6742e+10 1.3669e+12 44856
## + Response              1 2.4846e+10 1.3788e+12 44875
## + Education             1 2.0446e+10 1.3832e+12 44882
## + AcceptedCmp2          1 1.0757e+10 1.3929e+12 44898
## + NumDealsPurchases     1 9.6930e+09 1.3939e+12 44899
## + Dt_Customer           1 1.3889e+09 1.4022e+12 44913
## <none>                               1.4036e+12 44913
## + Complain              1 1.0403e+09 1.4026e+12 44913
## + Marital_Status        1 6.3997e+08 1.4030e+12 44914
## + Teenhome              1 5.1384e+08 1.4031e+12 44914
## + AcceptedCmp3          1 3.6720e+08 1.4032e+12 44914
## + ID                    1 2.4071e+08 1.4034e+12 44914
## + Recency               1 2.2119e+07 1.4036e+12 44915
##
## Step:  AIC=43970.01
## Income ~ NumCatalogPurchases
##
##                       Df  Sum of Sq        RSS   AIC
## + NumWebVisitsMonth    1 1.1632e+11 8.0008e+11 43671
## + MntWines             1 9.8480e+10 8.1792e+11 43720
## + NumStorePurchases    1 9.6465e+10 8.1993e+11 43726
## + MntMeatProducts      1 7.0441e+10 8.4596e+11 43795
## + NumWebPurchases      1 4.2230e+10 8.7417e+11 43867
## + MntSweetProducts     1 4.1298e+10 8.7510e+11 43870
## + MntFruits            1 3.8306e+10 8.7809e+11 43877
## + AcceptedCmp5         1 3.3371e+10 8.8303e+11 43890
## + Kidhome              1 3.2527e+10 8.8387e+11 43892
## + MntFishProducts      1 3.0619e+10 8.8578e+11 43897
## + AcceptedCmp4         1 1.4841e+10 9.0156e+11 43936
```

```
## + AcceptedCmp1         1 1.3932e+10 9.0247e+11 43938
## + Year_Birth           1 1.1554e+10 9.0485e+11 43944
## + Teenhome             1 1.0399e+10 9.0600e+11 43947
## + Education            1 9.0287e+09 9.0737e+11 43950
## + AcceptedCmp3         1 8.5564e+09 9.0784e+11 43951
## + NumDealsPurchases    1 8.1002e+09 9.0830e+11 43952
## + MntGoldProds         1 7.4314e+09 9.0897e+11 43954
## + AcceptedCmp2         1 1.1660e+09 9.1523e+11 43969
## <none>                              9.1640e+11 43970
## + Recency              1 4.6304e+08 9.1594e+11 43971
## + Dt_Customer          1 3.2911e+08 9.1607e+11 43971
## + Complain             1 3.1371e+08 9.1609e+11 43971
## + ID                   1 2.9249e+08 9.1611e+11 43971
## + Marital_Status       1 2.2134e+08 9.1618e+11 43971
## + Response             1 1.7881e+07 9.1638e+11 43972
## - NumCatalogPurchases  1 4.8721e+11 1.4036e+12 44913
##
## Step:  AIC=43671.21
## Income ~ NumCatalogPurchases + NumWebVisitsMonth
##
##                       Df  Sum of Sq        RSS   AIC
## + MntWines             1 1.0157e+11 6.9851e+11 43372
## + NumWebPurchases      1 7.6160e+10 7.2392e+11 43452
## + NumStorePurchases    1 5.8007e+10 7.4207e+11 43506
## + MntMeatProducts      1 3.2422e+10 7.6766e+11 43582
## + AcceptedCmp4         1 1.9598e+10 7.8048e+11 43618
## + AcceptedCmp5         1 1.8954e+10 7.8112e+11 43620
## + Teenhome             1 1.7319e+10 7.8276e+11 43625
## + MntSweetProducts     1 1.7168e+10 7.8291e+11 43625
## + MntFruits            1 1.5207e+10 7.8487e+11 43631
## + AcceptedCmp1         1 1.0828e+10 7.8925e+11 43643
## + Kidhome              1 9.6429e+09 7.9043e+11 43646
## + MntFishProducts      1 9.6346e+09 7.9044e+11 43646
## + Education            1 8.6687e+09 7.9141e+11 43649
## + Year_Birth           1 6.9541e+09 7.9312e+11 43654
## + MntGoldProds         1 6.2031e+09 7.9387e+11 43656
## + AcceptedCmp2         1 2.7271e+09 7.9735e+11 43666
## + Response             1 2.5852e+09 7.9749e+11 43666
## + NumDealsPurchases    1 2.4988e+09 7.9758e+11 43666
## + AcceptedCmp3         1 2.1517e+09 7.9793e+11 43667
## <none>                              8.0008e+11 43671
## + Recency              1 5.7197e+08 7.9951e+11 43672
## + Complain             1 2.0024e+08 7.9988e+11 43673
## + ID                   1 1.8257e+08 7.9989e+11 43673
## + Dt_Customer          1 1.6494e+08 7.9991e+11 43673
## + Marital_Status       1 5.3812e+07 8.0002e+11 43673
## - NumWebVisitsMonth    1 1.1632e+11 9.1640e+11 43970
## - NumCatalogPurchases  1 1.7416e+11 9.7424e+11 44106
##
## Step:  AIC=43372.37
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines
##
##                       Df  Sum of Sq        RSS   AIC
## + NumWebPurchases      1 2.3016e+10 6.7550e+11 43300
```

```
## + MntMeatProducts      1 1.3556e+10 6.8496e+11 43331
## + Teenhome             1 1.0222e+10 6.8829e+11 43342
## + NumStorePurchases    1 9.6568e+09 6.8886e+11 43344
## + MntSweetProducts     1 8.7711e+09 6.8974e+11 43346
## + MntFruits            1 7.2240e+09 6.9129e+11 43351
## + MntFishProducts      1 4.5616e+09 6.9395e+11 43360
## + Year_Birth           1 2.4134e+09 6.9610e+11 43367
## + NumDealsPurchases    1 1.9808e+09 6.9653e+11 43368
## + AcceptedCmp3         1 1.9155e+09 6.9660e+11 43368
## + AcceptedCmp1         1 1.3930e+09 6.9712e+11 43370
## + Education            1 8.9930e+08 6.9761e+11 43372
## + MntGoldProds         1 7.5873e+08 6.9775e+11 43372
## <none>                              6.9851e+11 43372
## + Recency              1 5.8213e+08 6.9793e+11 43373
## + AcceptedCmp4         1 5.4550e+08 6.9797e+11 43373
## + ID                   1 4.6537e+08 6.9805e+11 43373
## + AcceptedCmp5         1 4.1068e+08 6.9810e+11 43373
## + Dt_Customer          1 3.8383e+08 6.9813e+11 43373
## + Kidhome              1 1.0085e+08 6.9841e+11 43374
## + Marital_Status       1 5.3846e+07 6.9846e+11 43374
## + AcceptedCmp2         1 4.7936e+07 6.9846e+11 43374
## + Response             1 3.5828e+07 6.9848e+11 43374
## + Complain             1 1.0744e+07 6.9850e+11 43374
## - NumCatalogPurchases  1 2.4461e+10 7.2297e+11 43447
## - MntWines             1 1.0157e+11 8.0008e+11 43671
## - NumWebVisitsMonth    1 1.1941e+11 8.1792e+11 43720
##
## Step:  AIC=43300.12
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases
##
##                       Df  Sum of Sq        RSS   AIC
## + MntMeatProducts      1 1.3445e+10 6.6205e+11 43258
## + Teenhome             1 5.4143e+09 6.7008e+11 43284
## + MntSweetProducts     1 4.1926e+09 6.7130e+11 43288
## + MntFruits            1 3.8695e+09 6.7163e+11 43289
## + NumStorePurchases    1 2.7810e+09 6.7272e+11 43293
## + MntFishProducts      1 2.1154e+09 6.7338e+11 43295
## + AcceptedCmp1         1 1.9389e+09 6.7356e+11 43296
## + AcceptedCmp5         1 1.7650e+09 6.7373e+11 43296
## + AcceptedCmp3         1 1.6270e+09 6.7387e+11 43297
## + Year_Birth           1 1.2184e+09 6.7428e+11 43298
## + Education            1 1.1744e+09 6.7432e+11 43298
## + AcceptedCmp4         1 1.0735e+09 6.7442e+11 43299
## <none>                              6.7550e+11 43300
## + ID                   1 5.1256e+08 6.7498e+11 43300
## + Recency              1 4.6325e+08 6.7503e+11 43301
## + Dt_Customer          1 4.1575e+08 6.7508e+11 43301
## + Kidhome              1 3.5426e+08 6.7514e+11 43301
## + MntGoldProds         1 1.2413e+08 6.7537e+11 43302
## + NumDealsPurchases    1 1.0637e+08 6.7539e+11 43302
## + AcceptedCmp2         1 9.1520e+07 6.7540e+11 43302
## + Response             1 6.6018e+07 6.7543e+11 43302
## + Marital_Status       1 5.8736e+07 6.7544e+11 43302
```

```
## + Complain             1 1.4868e+07 6.7548e+11 43302
## - NumCatalogPurchases  1 1.7830e+10 6.9333e+11 43356
## - NumWebPurchases      1 2.3016e+10 6.9851e+11 43372
## - MntWines             1 4.8421e+10 7.2392e+11 43452
## - NumWebVisitsMonth    1 1.3644e+11 8.1193e+11 43706
##
## Step:  AIC=43257.57
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts
##
##                       Df  Sum of Sq        RSS   AIC
## + Teenhome             1 1.2372e+10 6.4968e+11 43218
## + Year_Birth           1 2.5799e+09 6.5947e+11 43251
## + NumStorePurchases    1 2.3715e+09 6.5968e+11 43252
## + AcceptedCmp4         1 1.8584e+09 6.6019e+11 43253
## + Education            1 1.7260e+09 6.6032e+11 43254
## + MntSweetProducts     1 1.5133e+09 6.6054e+11 43254
## + AcceptedCmp3         1 1.1877e+09 6.6086e+11 43256
## + AcceptedCmp1         1 1.1594e+09 6.6089e+11 43256
## + MntFruits            1 1.0737e+09 6.6098e+11 43256
## + AcceptedCmp5         1 7.6366e+08 6.6129e+11 43257
## <none>                              6.6205e+11 43258
## + ID                   1 5.3512e+08 6.6152e+11 43258
## + Recency              1 4.9097e+08 6.6156e+11 43258
## + Dt_Customer          1 4.7136e+08 6.6158e+11 43258
## + NumDealsPurchases    1 3.5691e+08 6.6169e+11 43258
## + Kidhome              1 3.3302e+08 6.6172e+11 43258
## + AcceptedCmp2         1 3.2451e+08 6.6173e+11 43258
## + MntFishProducts      1 2.6586e+08 6.6179e+11 43259
## + MntGoldProds         1 1.8098e+08 6.6187e+11 43259
## + Response             1 5.6004e+07 6.6199e+11 43259
## + Marital_Status       1 2.8549e+07 6.6202e+11 43259
## + Complain             1 1.2268e+07 6.6204e+11 43260
## - NumCatalogPurchases  1 3.7612e+09 6.6581e+11 43268
## - MntMeatProducts      1 1.3445e+10 6.7550e+11 43300
## - NumWebPurchases      1 2.2905e+10 6.8496e+11 43331
## - MntWines             1 3.7879e+10 6.9993e+11 43379
## - NumWebVisitsMonth    1 1.0476e+11 7.6681e+11 43581
##
## Step:  AIC=43217.77
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome
##
##                       Df  Sum of Sq        RSS   AIC
## + MntSweetProducts     1 2.6073e+09 6.4707e+11 43211
## + AcceptedCmp1         1 2.1925e+09 6.4749e+11 43212
## + AcceptedCmp5         1 2.0848e+09 6.4759e+11 43213
## + MntFruits            1 2.0080e+09 6.4767e+11 43213
## + AcceptedCmp4         1 1.8261e+09 6.4785e+11 43214
## + NumStorePurchases    1 1.6326e+09 6.4805e+11 43214
## + MntFishProducts      1 1.0222e+09 6.4866e+11 43216
## + Education            1 8.8255e+08 6.4880e+11 43217
## + AcceptedCmp3         1 7.9899e+08 6.4888e+11 43217
## + Kidhome              1 7.8308e+08 6.4890e+11 43217
```

```
## + Recency             1 6.1417e+08 6.4906e+11 43218
## <none>                            6.4968e+11 43218
## + ID                  1 5.3409e+08 6.4914e+11 43218
## + Dt_Customer         1 4.8077e+08 6.4920e+11 43218
## + NumDealsPurchases   1 4.8006e+08 6.4920e+11 43218
## + AcceptedCmp2        1 4.5124e+08 6.4923e+11 43218
## + Year_Birth          1 1.8814e+08 6.4949e+11 43219
## + MntGoldProds        1 1.1490e+08 6.4956e+11 43219
## + Response            1 5.1835e+07 6.4963e+11 43220
## + Marital_Status      1 2.1110e+07 6.4966e+11 43220
## + Complain            1 1.3830e+07 6.4966e+11 43220
## - NumCatalogPurchases  1 3.4078e+09 6.5309e+11 43227
## - Teenhome            1 1.2372e+10 6.6205e+11 43258
## - NumWebPurchases     1 1.5920e+10 6.6560e+11 43269
## - MntMeatProducts     1 2.0403e+10 6.7008e+11 43284
## - MntWines            1 3.5200e+10 6.8488e+11 43333
## - NumWebVisitsMonth   1 1.0202e+11 7.5170e+11 43539
##
## Step:  AIC=43210.85
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts
##
##                      Df  Sum of Sq        RSS   AIC
## + AcceptedCmp4        1 2.0941e+09 6.4498e+11 43206
## + AcceptedCmp1        1 1.9061e+09 6.4516e+11 43206
## + AcceptedCmp5        1 1.9001e+09 6.4517e+11 43206
## + Education           1 1.5539e+09 6.4552e+11 43208
## + NumStorePurchases   1 1.0974e+09 6.4597e+11 43209
## + Kidhome             1 1.0046e+09 6.4607e+11 43209
## + MntFruits           1 8.7398e+08 6.4620e+11 43210
## + AcceptedCmp3        1 7.4725e+08 6.4632e+11 43210
## + Recency             1 6.6503e+08 6.4641e+11 43211
## <none>                            6.4707e+11 43211
## + ID                  1 5.4084e+08 6.4653e+11 43211
## + AcceptedCmp2        1 5.0665e+08 6.4656e+11 43211
## + Dt_Customer         1 5.0437e+08 6.4657e+11 43211
## + NumDealsPurchases   1 3.6801e+08 6.4670e+11 43212
## + MntGoldProds        1 2.7633e+08 6.4679e+11 43212
## + MntFishProducts     1 2.5510e+08 6.4682e+11 43212
## + Year_Birth          1 2.4379e+08 6.4683e+11 43212
## + Response            1 6.6582e+07 6.4700e+11 43213
## + Marital_Status      1 2.0459e+07 6.4705e+11 43213
## + Complain            1 1.0798e+07 6.4706e+11 43213
## - MntSweetProducts    1 2.6073e+09 6.4968e+11 43218
## - NumCatalogPurchases  1 2.9047e+09 6.4998e+11 43219
## - NumWebPurchases     1 1.2495e+10 6.5957e+11 43251
## - Teenhome            1 1.3466e+10 6.6054e+11 43254
## - MntMeatProducts     1 1.6925e+10 6.6400e+11 43266
## - MntWines            1 3.5508e+10 6.8258e+11 43327
## - NumWebVisitsMonth   1 9.1369e+10 7.3844e+11 43502
##
## Step:  AIC=43205.67
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
```

```
##     AcceptedCmp4
##
##                       Df  Sum of Sq        RSS   AIC
## + Education            1 1.7261e+09 6.4325e+11 43202
## + AcceptedCmp1         1 1.3497e+09 6.4363e+11 43203
## + AcceptedCmp5         1 1.2543e+09 6.4372e+11 43203
## + NumStorePurchases    1 1.1634e+09 6.4381e+11 43204
## + Kidhome              1 1.1244e+09 6.4385e+11 43204
## + MntFruits            1 1.0893e+09 6.4389e+11 43204
## + Recency              1 7.0398e+08 6.4427e+11 43205
## <none>                              6.4498e+11 43206
## + ID                   1 5.7584e+08 6.4440e+11 43206
## + Dt_Customer          1 5.1010e+08 6.4447e+11 43206
## + AcceptedCmp3         1 4.9037e+08 6.4449e+11 43206
## + NumDealsPurchases    1 3.5506e+08 6.4462e+11 43206
## + MntFishProducts      1 3.5167e+08 6.4463e+11 43206
## + Year_Birth           1 2.3475e+08 6.4474e+11 43207
## + AcceptedCmp2         1 1.5762e+08 6.4482e+11 43207
## + MntGoldProds         1 1.5574e+08 6.4482e+11 43207
## + Marital_Status       1 1.4258e+07 6.4496e+11 43208
## + Response             1 7.5964e+06 6.4497e+11 43208
## + Complain             1 6.8699e+06 6.4497e+11 43208
## - AcceptedCmp4         1 2.0941e+09 6.4707e+11 43211
## - MntSweetProducts     1 2.8753e+09 6.4785e+11 43214
## - NumCatalogPurchases  1 3.0662e+09 6.4804e+11 43214
## - NumWebPurchases      1 1.2951e+10 6.5793e+11 43248
## - Teenhome             1 1.3496e+10 6.5847e+11 43250
## - MntMeatProducts      1 1.7688e+10 6.6267e+11 43264
## - MntWines             1 2.5045e+10 6.7002e+11 43288
## - NumWebVisitsMonth    1 9.1986e+10 7.3696e+11 43499
##
## Step:  AIC=43201.73
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education
##
##                       Df  Sum of Sq        RSS   AIC
## + AcceptedCmp1         1 1.4993e+09 6.4175e+11 43199
## + MntFruits            1 1.3792e+09 6.4187e+11 43199
## + AcceptedCmp5         1 1.3694e+09 6.4188e+11 43199
## + NumStorePurchases    1 1.3029e+09 6.4195e+11 43199
## + Kidhome              1 1.0194e+09 6.4223e+11 43200
## + Recency              1 6.7704e+08 6.4257e+11 43201
## + MntFishProducts      1 6.2032e+08 6.4263e+11 43202
## <none>                              6.4325e+11 43202
## + ID                   1 5.7613e+08 6.4267e+11 43202
## + Dt_Customer          1 5.4460e+08 6.4271e+11 43202
## + AcceptedCmp3         1 4.8065e+08 6.4277e+11 43202
## + NumDealsPurchases    1 3.2926e+08 6.4292e+11 43203
## + AcceptedCmp2         1 1.8272e+08 6.4307e+11 43203
## + Year_Birth           1 1.1755e+08 6.4313e+11 43203
## + MntGoldProds         1 3.2054e+07 6.4322e+11 43204
## + Marital_Status       1 1.2273e+07 6.4324e+11 43204
## + Complain             1 4.3670e+05 6.4325e+11 43204
```

```
## + Response            1 3.4369e+05 6.4325e+11 43204
## - Education           1 1.7261e+09 6.4498e+11 43206
## - AcceptedCmp4        1 2.2662e+09 6.4552e+11 43208
## - NumCatalogPurchases  1 3.1503e+09 6.4640e+11 43211
## - MntSweetProducts    1 3.6280e+09 6.4688e+11 43212
## - Teenhome            1 1.2535e+10 6.5579e+11 43243
## - NumWebPurchases     1 1.3053e+10 6.5630e+11 43244
## - MntMeatProducts     1 1.7653e+10 6.6090e+11 43260
## - MntWines            1 2.1581e+10 6.6483e+11 43273
## - NumWebVisitsMonth   1 9.0455e+10 7.3371e+11 43491
##
## Step:  AIC=43198.56
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1
##
##                      Df  Sum of Sq        RSS   AIC
## + NumStorePurchases   1 1.6213e+09 6.4013e+11 43195
## + MntFruits           1 1.4609e+09 6.4029e+11 43196
## + Kidhome             1 9.1040e+08 6.4084e+11 43197
## + AcceptedCmp5        1 8.1939e+08 6.4093e+11 43198
## + AcceptedCmp3        1 6.6621e+08 6.4109e+11 43198
## + Recency             1 6.0889e+08 6.4114e+11 43198
## + ID                  1 5.9741e+08 6.4115e+11 43198
## <none>                           6.4175e+11 43199
## + Dt_Customer         1 5.6029e+08 6.4119e+11 43199
## + MntFishProducts     1 5.1797e+08 6.4123e+11 43199
## + NumDealsPurchases   1 2.2547e+08 6.4153e+11 43200
## + Year_Birth          1 1.1595e+08 6.4164e+11 43200
## + AcceptedCmp2        1 1.0073e+08 6.4165e+11 43200
## + Response            1 7.4721e+07 6.4168e+11 43200
## + MntGoldProds        1 3.7442e+07 6.4171e+11 43200
## + Marital_Status      1 2.0671e+07 6.4173e+11 43200
## + Complain            1 4.2102e+04 6.4175e+11 43201
## - AcceptedCmp1        1 1.4993e+09 6.4325e+11 43202
## - AcceptedCmp4        1 1.6651e+09 6.4342e+11 43202
## - Education           1 1.8757e+09 6.4363e+11 43203
## - NumCatalogPurchases  1 2.9599e+09 6.4471e+11 43207
## - MntSweetProducts    1 3.3196e+09 6.4507e+11 43208
## - NumWebPurchases     1 1.3294e+10 6.5505e+11 43242
## - Teenhome            1 1.3297e+10 6.5505e+11 43242
## - MntMeatProducts     1 1.7086e+10 6.5884e+11 43255
## - MntWines            1 1.9653e+10 6.6140e+11 43263
## - NumWebVisitsMonth   1 9.0276e+10 7.3203e+11 43488
##
## Step:  AIC=43194.96
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1 + NumStorePurchases
##
##                      Df  Sum of Sq        RSS   AIC
## + Kidhome             1 1.2085e+09 6.3892e+11 43193
## + AcceptedCmp5        1 1.1980e+09 6.3893e+11 43193
## + MntFruits           1 1.1141e+09 6.3902e+11 43193
```

```
## + ID                   1 6.0516e+08 6.3953e+11 43195
## <none>                              6.4013e+11 43195
## + Dt_Customer          1 5.6409e+08 6.3957e+11 43195
## + Recency              1 5.6161e+08 6.3957e+11 43195
## + AcceptedCmp3         1 4.4854e+08 6.3968e+11 43195
## + NumDealsPurchases    1 4.3299e+08 6.3970e+11 43195
## + MntFishProducts      1 3.2956e+08 6.3980e+11 43196
## + Year_Birth           1 1.4775e+08 6.3998e+11 43196
## + AcceptedCmp2         1 9.1495e+07 6.4004e+11 43197
## + MntGoldProds         1 7.0064e+07 6.4006e+11 43197
## + Marital_Status       1 2.7215e+07 6.4010e+11 43197
## + Response             1 1.0841e+07 6.4012e+11 43197
## + Complain             1 3.9578e+05 6.4013e+11 43197
## - NumStorePurchases    1 1.6213e+09 6.4175e+11 43199
## - AcceptedCmp4         1 1.6875e+09 6.4182e+11 43199
## - AcceptedCmp1         1 1.8177e+09 6.4195e+11 43199
## - Education            1 2.0557e+09 6.4219e+11 43200
## - MntSweetProducts     1 2.6188e+09 6.4275e+11 43202
## - NumCatalogPurchases  1 3.0188e+09 6.4315e+11 43203
## - NumWebPurchases      1 1.0111e+10 6.5024e+11 43228
## - Teenhome             1 1.2420e+10 6.5255e+11 43236
## - MntWines             1 1.3854e+10 6.5398e+11 43240
## - MntMeatProducts      1 1.6766e+10 6.5690e+11 43250
## - NumWebVisitsMonth    1 7.6816e+10 7.1695e+11 43444
##
## Step:  AIC=43192.77
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1 + NumStorePurchases +
##     Kidhome
##
##                       Df  Sum of Sq        RSS   AIC
## + MntFruits            1 1.2183e+09 6.3770e+11 43191
## + AcceptedCmp5         1 1.0881e+09 6.3783e+11 43191
## + NumDealsPurchases    1 1.0741e+09 6.3785e+11 43191
## + Recency              1 6.1770e+08 6.3830e+11 43193
## + ID                   1 6.1540e+08 6.3831e+11 43193
## <none>                              6.3892e+11 43193
## + Dt_Customer          1 5.3796e+08 6.3838e+11 43193
## + AcceptedCmp3         1 4.7448e+08 6.3845e+11 43193
## + MntFishProducts      1 3.9531e+08 6.3853e+11 43193
## + Year_Birth           1 3.0233e+08 6.3862e+11 43194
## + AcceptedCmp2         1 1.0450e+08 6.3882e+11 43194
## + MntGoldProds         1 3.2101e+07 6.3889e+11 43195
## + Marital_Status       1 3.1798e+07 6.3889e+11 43195
## + Response             1 1.3654e+07 6.3891e+11 43195
## + Complain             1 2.7365e+06 6.3892e+11 43195
## - Kidhome              1 1.2085e+09 6.4013e+11 43195
## - AcceptedCmp1         1 1.7133e+09 6.4064e+11 43197
## - AcceptedCmp4         1 1.8169e+09 6.4074e+11 43197
## - NumStorePurchases    1 1.9194e+09 6.4084e+11 43197
## - Education            1 1.9415e+09 6.4086e+11 43197
## - MntSweetProducts     1 2.7996e+09 6.4172e+11 43200
## - NumCatalogPurchases  1 3.4227e+09 6.4234e+11 43203
```

```
## - NumWebPurchases      1 1.0804e+10 6.4973e+11 43228
## - Teenhome             1 1.2984e+10 6.5191e+11 43235
## - MntWines             1 1.4603e+10 6.5352e+11 43241
## - MntMeatProducts      1 1.6836e+10 6.5576e+11 43248
## - NumWebVisitsMonth    1 7.7042e+10 7.1596e+11 43443
##
## Step:  AIC=43190.54
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1 + NumStorePurchases +
##     Kidhome + MntFruits
##
##                       Df  Sum of Sq        RSS   AIC
## + AcceptedCmp5         1 1.1316e+09 6.3657e+11 43189
## + NumDealsPurchases    1 9.5103e+08 6.3675e+11 43189
## + ID                   1 5.9083e+08 6.3711e+11 43190
## + Recency              1 5.7561e+08 6.3713e+11 43191
## <none>                              6.3770e+11 43191
## + Dt_Customer          1 5.0981e+08 6.3719e+11 43191
## + AcceptedCmp3         1 4.9956e+08 6.3720e+11 43191
## + Year_Birth           1 3.1393e+08 6.3739e+11 43191
## + MntFishProducts      1 1.3690e+08 6.3757e+11 43192
## + AcceptedCmp2         1 1.3020e+08 6.3757e+11 43192
## + MntGoldProds         1 1.0676e+08 6.3760e+11 43192
## + Marital_Status       1 3.9009e+07 6.3766e+11 43192
## + Response             1 2.3336e+07 6.3768e+11 43192
## + Complain             1 4.2145e+06 6.3770e+11 43193
## - MntFruits            1 1.2183e+09 6.3892e+11 43193
## - Kidhome              1 1.3126e+09 6.3902e+11 43193
## - NumStorePurchases    1 1.5394e+09 6.3924e+11 43194
## - MntSweetProducts     1 1.6331e+09 6.3934e+11 43194
## - AcceptedCmp1         1 1.7511e+09 6.3945e+11 43195
## - AcceptedCmp4         1 2.0337e+09 6.3974e+11 43196
## - Education            1 2.2090e+09 6.3991e+11 43196
## - NumCatalogPurchases  1 3.2947e+09 6.4100e+11 43200
## - NumWebPurchases      1 1.0314e+10 6.4802e+11 43224
## - Teenhome             1 1.3564e+10 6.5127e+11 43235
## - MntWines             1 1.4586e+10 6.5229e+11 43239
## - MntMeatProducts      1 1.4921e+10 6.5262e+11 43240
## - NumWebVisitsMonth    1 7.5115e+10 7.1282e+11 43435
##
## Step:  AIC=43188.6
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1 + NumStorePurchases +
##     Kidhome + MntFruits + AcceptedCmp5
##
##                       Df  Sum of Sq        RSS   AIC
## + NumDealsPurchases    1 8.2153e+08 6.3575e+11 43188
## + AcceptedCmp3         1 6.0367e+08 6.3597e+11 43189
## + ID                   1 5.7913e+08 6.3599e+11 43189
## <none>                              6.3657e+11 43189
## + Recency              1 5.5530e+08 6.3602e+11 43189
## + Dt_Customer          1 4.6161e+08 6.3611e+11 43189
```

```
## + Year_Birth           1 3.5247e+08 6.3622e+11 43189
## + MntFishProducts      1 2.2180e+08 6.3635e+11 43190
## + MntGoldProds         1 1.2972e+08 6.3644e+11 43190
## + Response             1 1.1640e+08 6.3646e+11 43190
## + AcceptedCmp2         1 6.6784e+07 6.3651e+11 43190
## - AcceptedCmp1         1 1.1062e+09 6.3768e+11 43190
## + Marital_Status       1 3.7860e+07 6.3653e+11 43190
## - AcceptedCmp5         1 1.1316e+09 6.3770e+11 43191
## + Complain             1 7.2861e+06 6.3656e+11 43191
## - Kidhome              1 1.1976e+09 6.3777e+11 43191
## - MntFruits            1 1.2618e+09 6.3783e+11 43191
## - MntSweetProducts     1 1.4762e+09 6.3805e+11 43192
## - AcceptedCmp4         1 1.5371e+09 6.3811e+11 43192
## - NumStorePurchases    1 1.8759e+09 6.3845e+11 43193
## - Education            1 2.3278e+09 6.3890e+11 43195
## - NumCatalogPurchases  1 3.5910e+09 6.4016e+11 43199
## - NumWebPurchases      1 1.0687e+10 6.4726e+11 43223
## - MntWines             1 1.0831e+10 6.4740e+11 43224
## - MntMeatProducts      1 1.4172e+10 6.5074e+11 43235
## - Teenhome             1 1.4299e+10 6.5087e+11 43236
## - NumWebVisitsMonth    1 7.1335e+10 7.0791e+11 43422
##
## Step:  AIC=43187.74
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1 + NumStorePurchases +
##     Kidhome + MntFruits + AcceptedCmp5 + NumDealsPurchases
##
##                       Df  Sum of Sq        RSS   AIC
## + AcceptedCmp3         1 6.8117e+08 6.3507e+11 43187
## <none>                              6.3575e+11 43188
## + Recency              1 5.6743e+08 6.3518e+11 43188
## + ID                   1 5.2313e+08 6.3523e+11 43188
## + Dt_Customer          1 4.8966e+08 6.3526e+11 43188
## + Year_Birth           1 3.3516e+08 6.3542e+11 43189
## - NumDealsPurchases    1 8.2153e+08 6.3657e+11 43189
## + MntFishProducts      1 1.9100e+08 6.3556e+11 43189
## - AcceptedCmp1         1 9.7823e+08 6.3673e+11 43189
## - AcceptedCmp5         1 1.0021e+09 6.3675e+11 43189
## + MntGoldProds         1 8.9637e+07 6.3566e+11 43189
## + Response             1 8.6241e+07 6.3566e+11 43189
## + AcceptedCmp2         1 5.3380e+07 6.3570e+11 43190
## + Marital_Status       1 3.4339e+07 6.3572e+11 43190
## + Complain             1 1.0312e+07 6.3574e+11 43190
## - MntFruits            1 1.1419e+09 6.3689e+11 43190
## - MntSweetProducts     1 1.3969e+09 6.3715e+11 43191
## - AcceptedCmp4         1 1.5969e+09 6.3735e+11 43191
## - Kidhome              1 1.7340e+09 6.3748e+11 43192
## - Education            1 2.2344e+09 6.3798e+11 43194
## - NumStorePurchases    1 2.2688e+09 6.3802e+11 43194
## - NumCatalogPurchases  1 4.1559e+09 6.3991e+11 43200
## - MntWines             1 1.0422e+10 6.4617e+11 43222
## - NumWebPurchases      1 1.1402e+10 6.4715e+11 43225
## - MntMeatProducts      1 1.4538e+10 6.5029e+11 43236
```

```
## - Teenhome             1 1.4876e+10 6.5063e+11 43237
## - NumWebVisitsMonth    1 6.0415e+10 6.9617e+11 43387
##
## Step:  AIC=43187.37
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1 + NumStorePurchases +
##     Kidhome + MntFruits + AcceptedCmp5 + NumDealsPurchases +
##     AcceptedCmp3
##
##                       Df  Sum of Sq        RSS   AIC
## + Recency              1 6.0736e+08 6.3446e+11 43187
## <none>                              6.3507e+11 43187
## + Dt_Customer          1 4.9072e+08 6.3458e+11 43188
## + ID                   1 4.7674e+08 6.3459e+11 43188
## - AcceptedCmp3         1 6.8117e+08 6.3575e+11 43188
## + Year_Birth           1 2.8280e+08 6.3479e+11 43188
## - NumDealsPurchases    1 8.9902e+08 6.3597e+11 43189
## + MntFishProducts      1 1.7454e+08 6.3489e+11 43189
## + AcceptedCmp2         1 8.2022e+07 6.3499e+11 43189
## - AcceptedCmp1         1 1.0830e+09 6.3615e+11 43189
## - AcceptedCmp5         1 1.1014e+09 6.3617e+11 43189
## + MntGoldProds         1 4.5065e+07 6.3502e+11 43189
## + Marital_Status       1 2.7949e+07 6.3504e+11 43189
## + Response             1 1.3418e+07 6.3506e+11 43189
## + Complain             1 9.2623e+06 6.3506e+11 43189
## - MntFruits            1 1.1672e+09 6.3624e+11 43189
## - AcceptedCmp4         1 1.2947e+09 6.3636e+11 43190
## - MntSweetProducts     1 1.3507e+09 6.3642e+11 43190
## - Kidhome              1 1.7983e+09 6.3687e+11 43192
## - NumStorePurchases    1 2.0115e+09 6.3708e+11 43192
## - Education            1 2.2210e+09 6.3729e+11 43193
## - NumCatalogPurchases  1 4.5842e+09 6.3965e+11 43201
## - MntWines             1 1.0691e+10 6.4576e+11 43222
## - NumWebPurchases      1 1.1587e+10 6.4666e+11 43225
## - MntMeatProducts      1 1.3995e+10 6.4906e+11 43234
## - Teenhome             1 1.4832e+10 6.4990e+11 43237
## - NumWebVisitsMonth    1 5.8686e+10 6.9376e+11 43381
##
## Step:  AIC=43187.25
## Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
##     NumWebPurchases + MntMeatProducts + Teenhome + MntSweetProducts +
##     AcceptedCmp4 + Education + AcceptedCmp1 + NumStorePurchases +
##     Kidhome + MntFruits + AcceptedCmp5 + NumDealsPurchases +
##     AcceptedCmp3 + Recency
##
##                       Df  Sum of Sq        RSS   AIC
## <none>                              6.3446e+11 43187
## - Recency              1 6.0736e+08 6.3507e+11 43187
## + Dt_Customer          1 4.7732e+08 6.3398e+11 43188
## + ID                   1 4.2951e+08 6.3403e+11 43188
## - AcceptedCmp3         1 7.2110e+08 6.3518e+11 43188
## + Year_Birth           1 2.9263e+08 6.3417e+11 43188
## - NumDealsPurchases    1 9.1452e+08 6.3538e+11 43188
```

```
## + MntFishProducts      1 1.6666e+08 6.3430e+11 43189
## - AcceptedCmp1         1 1.0282e+09 6.3549e+11 43189
## + Response             1 8.7028e+07 6.3437e+11 43189
## + AcceptedCmp2         1 8.1180e+07 6.3438e+11 43189
## - AcceptedCmp5         1 1.0827e+09 6.3554e+11 43189
## + MntGoldProds         1 3.7874e+07 6.3442e+11 43189
## + Marital_Status       1 3.0083e+07 6.3443e+11 43189
## - MntFruits            1 1.1245e+09 6.3559e+11 43189
## + Complain             1 7.3323e+06 6.3445e+11 43189
## - AcceptedCmp4         1 1.3263e+09 6.3579e+11 43190
## - MntSweetProducts     1 1.4061e+09 6.3587e+11 43190
## - Kidhome              1 1.8688e+09 6.3633e+11 43192
## - NumStorePurchases    1 1.9662e+09 6.3643e+11 43192
## - Education            1 2.1750e+09 6.3664e+11 43193
## - NumCatalogPurchases  1 4.6485e+09 6.3911e+11 43201
## - MntWines             1 1.0801e+10 6.4526e+11 43223
## - NumWebPurchases      1 1.1529e+10 6.4599e+11 43225
## - MntMeatProducts      1 1.4079e+10 6.4854e+11 43234
## - Teenhome             1 1.4974e+10 6.4944e+11 43237
## - NumWebVisitsMonth    1 5.8796e+10 6.9326e+11 43382
```

Select attributes

Because from backward, forward, and both stepwise result, finally we reached the same decision, so here we just choose the varibles that obtained from the model we obtained from them.We first take a look at all the attribute we have here. All the vif value are less then 5.

```r
VIF<-vif(lm(Income ~ NumCatalogPurchases + NumWebVisitsMonth + MntWines +
    NumWebPurchases + MntMeatProducts + Teenhome + Education +
    MntSweetProducts + AcceptedCmp4 + AcceptedCmp1 + MntFruits +
    NumStorePurchases + AcceptedCmp5 + Kidhome + NumDealsPurchases +
    AcceptedCmp3 + Recency,data=data))
# detach(data)
VIF<5
```

```
##                     GVIF   Df GVIF^(1/(2*Df))
## NumCatalogPurchases TRUE TRUE            TRUE
## NumWebVisitsMonth   TRUE TRUE            TRUE
## MntWines            TRUE TRUE            TRUE
## NumWebPurchases     TRUE TRUE            TRUE
## MntMeatProducts     TRUE TRUE            TRUE
## Teenhome            TRUE TRUE            TRUE
## Education           TRUE TRUE            TRUE
## MntSweetProducts    TRUE TRUE            TRUE
## AcceptedCmp4        TRUE TRUE            TRUE
## AcceptedCmp1        TRUE TRUE            TRUE
## MntFruits           TRUE TRUE            TRUE
## NumStorePurchases   TRUE TRUE            TRUE
## AcceptedCmp5        TRUE TRUE            TRUE
## Kidhome             TRUE TRUE            TRUE
## NumDealsPurchases   TRUE TRUE            TRUE
## AcceptedCmp3        TRUE TRUE            TRUE
## Recency             TRUE TRUE            TRUE
```

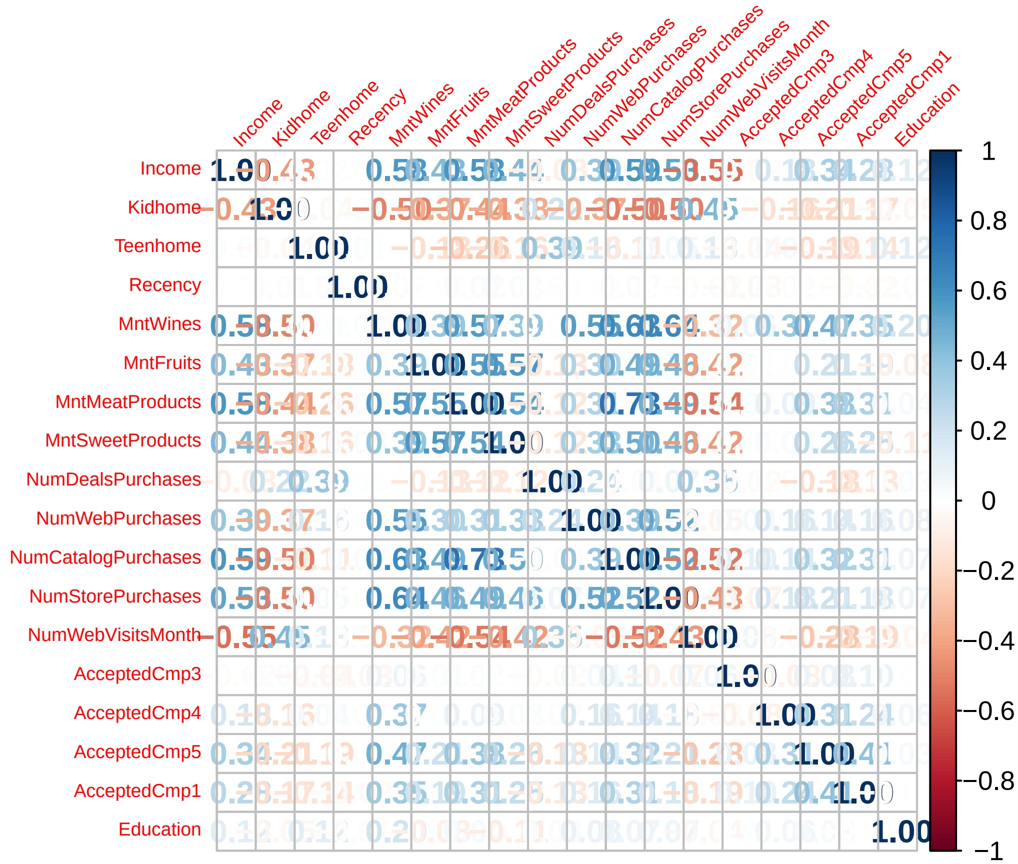
Here we try to use the remaining data to make a combined_data set for further analysis, and by plot the data, we can find that some attributes have a high multicollinearity coefficient which we need to deal with later.

```r
attribute_names <- names(coef(back_data))
encode_data <- data["Income"]
for(col in attribute_names[2:18]){
  encode_data<-bind_cols(encode_data,data[col])
  # break
}
# Select only the text columns for encoding
text_columns <- select_if(encode_data, is.character)

# Encode text columns
encoded <- text_columns %>%
  mutate(across(everything(), as.factor)) %>%
  mutate(across(everything(), as.numeric))

# Select numeric columns
numeric_columns <- select_if(encode_data, is.numeric)

# Combine non-text and numeric columns
combined_data <- bind_cols(numeric_columns, encoded)
par(cex.lab = 1.5, cex.main = 1.5, cex.axis = 7)
corrplot(cor(combined_data),tl.srt = 45, tl.cex = 0.6,cex.lab = 0.1,method= "number")
```



Overall analysis based on Eigenvalue analysis

We first use the eigen method to diagose the multicollinearity,find it is severe, since K»1000,which means in the dataset, there exist strong evidence of multicollinearity.

```
X<-as.matrix(combined_data)
lambda<-eigen(t(X)%*%X)$values
k<-max(lambda)/min(lambda)
k
```

```
## [1] 93524601051
```

```
k>1000
```

```
## [1] TRUE
```

Overall analyse based on vif method

But vif method find the data don't have numeric value over 5 or 10, but since there are lot of value among 2-3, so the multicollinearity is still exist.Then we consider to drop some of the variables in next step.

```
vif(lm(Income~.,data=combined_data))
```

```
##            Kidhome            Teenhome             Recency            MntWines
##           1.833060            1.420622            1.007075            3.400804
##          MntFruits     MntMeatProducts    MntSweetProducts   NumDealsPurchases
##           1.816859            2.915512            1.845349            1.649160
##    NumWebPurchases NumCatalogPurchases   NumStorePurchases   NumWebVisitsMonth
##           1.880926            3.017223            2.380047            2.126153
##       AcceptedCmp3        AcceptedCmp4        AcceptedCmp5        AcceptedCmp1
##           1.085782            1.306010            1.598613            1.318077
##          Education
##           1.120015
```

By look into both side we find that eigen method is more straight forward to show the multicolinearity, and can show how severe it is.

Variable selection

Then we try to drop some of the features, I try to drop the features which have bigest vif value, and ket the k value drop to less than 1000,Until we droped the eighth element "AcceptedCmp5","AcceptedCmp1",even though their vif value is not that huge, the k value droped a significantly, seems that other attribute don't have a significant influence on k value. And we also discovered that the vif value is highly related to sample size, so if we only sample 50 samples from the dataset, some of the attributes attain huge vif value more than 6, so we decided to remove them.

```
set.seed(1234)
# Specify the name of the column to drop
column_to_drop <- c("AcceptedCmp5","AcceptedCmp1","NumCatalogPurchases","NumWebVisitsMonth","MntWines")
```

```r
# Drop the column from the data frame
combined_data <- combined_data[, !names(combined_data) %in% column_to_drop]
vif(lm(Income~.,data=combined_data[sample(nrow(combined_data), 50), ]))
```

```
##          Kidhome          Teenhome           Recency         MntFruits
##         2.512393          1.830165          1.099868          2.935560
##   MntMeatProducts  MntSweetProducts NumDealsPurchases   NumWebPurchases
##         1.969900          3.045241          1.786998          2.562186
## NumStorePurchases      AcceptedCmp3      AcceptedCmp4         Education
##         3.265888          1.278072          1.555344          1.515536
```

```r
X<-as.matrix(combined_data)
lambda<-eigen(t(X)%*%X)$values
k<-max(lambda)/min(lambda)
k
```

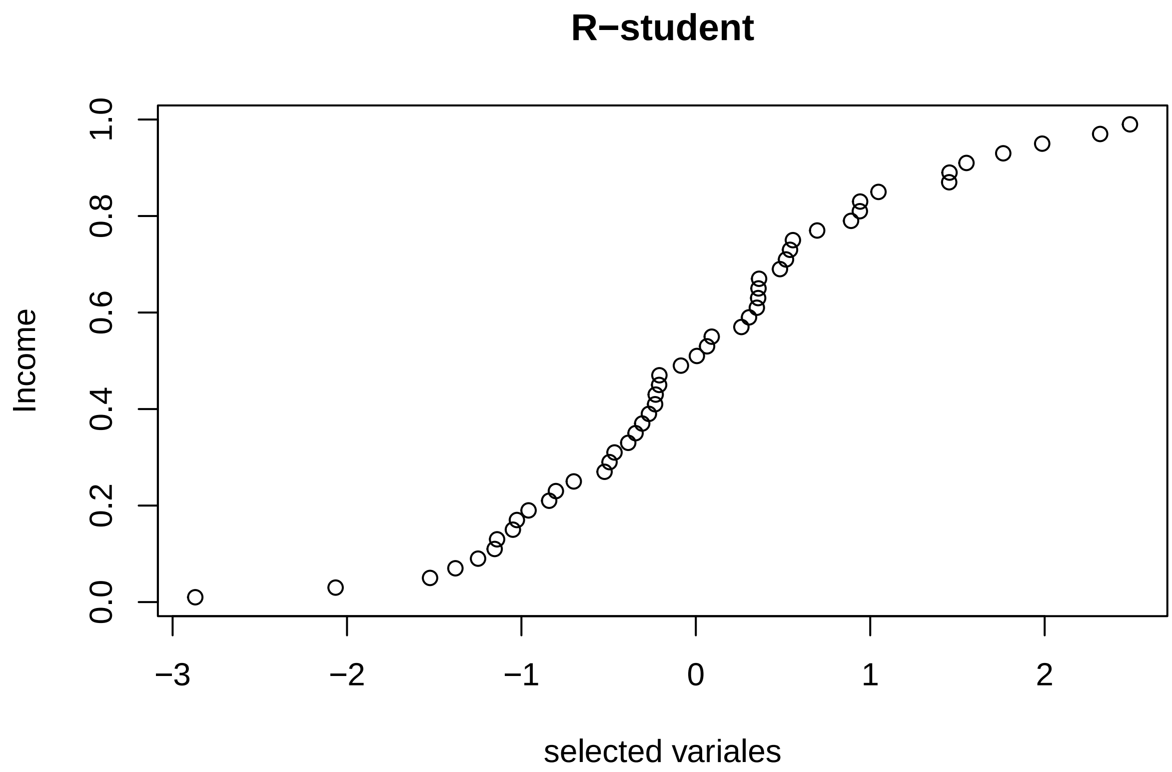
```
## [1] 55961075705
```

```r
# corrplot(cor(combined_data),tl.srt = 45, tl.cex = 0.6,cex.lab = 0.1,method= "number")
```

Analyse by R-student plot

here we sampled 50 samples from the dataset, and we can see the R-student looks normal, which means the evidence of multicolinearity is not really severe.

```r
set.seed(1234)
# detach(data)
# fit.modelr8 <- lmer(Income~MntMeatProduct(1|Education)+(1|Year_Birth)+(1|Complain)+(1+data$AcceptedCm
# AIC(fit.modelr8)
res<-lm(Income~.,data=combined_data[sample(nrow(combined_data), 50), ])
r<-rstudent(res)
qqplot(sort(r),ppoints(res$fit),ylab = "Income",xlab="selected variales",main="R-student")
```



So we by now, we learned that, we can combine both method to deal with existence of multicolliearity, seems vif gives more detail on each variables, and based on the observation on the change of both vif and K value, we can deal with muticollieartity much more efficiently.

Comparision

```
# New data for prediction
fit.modelr7 <- lmer(Income~1+NumDealsPurchases+Recency+(1|Education)+(1|Year_Birth)+(1+AcceptedCmp1|Rec
fit.model8 <- lmer(Income~1+NumCatalogPurchases+NumWebVisitsMonth+(1|Education)+(1|Year_Birth), data=da
fit.model7 <- lmer(Income~1+NumCatalogPurchases+NumWebVisitsMonth+MntWines+NumWebPurchases+(1|Education
new_data_r7 <- data.frame(NumDealsPurchases = c(1,2,5),
                         Recency=c(26,26,94),
                         Education=c("Graduation","Graduation","PhD"),
                         Year_Birth=c(1965,1984,1981),
                         AcceptedCmp1=c(0,0,0))
new_data_8<-data.frame(NumCatalogPurchases = c(2,0,3),
                         NumWebVisitsMonth=c(4,6,5),
                         Education=c("Graduation","Graduation","PhD"),
                         Year_Birth=c(1965,1984,1981))
new_data_7<-data.frame(NumCatalogPurchases = c(2,0,3),
                         NumWebVisitsMonth=c(4,6,5),
                         Education=c("Graduation","Graduation","PhD"),
                         MntWines=c(426,11,173),
                       NumWebPurchases=c(8,2,5))
```

```r
print("actua7 value:71612.0 26646.0 58293.0")
```

```
## [1] "actua7 value:71612.0 26646.0 58293.0"
```

```r
predictions_r7<- predict(fit.modelr7, newdata = new_data_r7)
predictions_r7
```

```
##        1        2        3
## 54673.68 43994.52 48524.37
```

```r
predictions_8<- predict(fit.model8, newdata = new_data_8)
predictions_8
```

```
##        1        2        3
## 55738.55 39418.71 56951.70
```

```r
predictions_7<- predict(fit.model7, newdata = new_data_7)
predictions_7
```

```
##        1        2        3
## 64688.09 37871.79 53307.05
```

Models with more random effects are better, models with more fixed effects are better, or a combination of both are better? Let us take a try!

You can find, our models with more fixed effects perform the best, its results are closer to the real income, but it still not very good. However, in the last prediction. the mixed model predict the best. Different groups perhaps have different best models and combine them may improve the accuracy. Due to our ability isn't good enough, we stop our analyses here.

Our considerations are still very limited, and there are still significant gaps in the accurate prediction of data. However, performing clustering to summarize customer segments instead predict income may be enough for the company. In that case, using our linear models and then clustering them may be powerful!

Anyway, we enjoyed this project!