

# Project\_final.R

Nick

2021-05-06

```
library(naniar)
```

```
## Warning: package 'naniar' was built under R version 3.6.3
```

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 3.6.3
```

```
##  
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:stats':  
##  
##      filter
```

```
## The following objects are masked from 'package:base':  
##  
##      cbind, rbind
```

```
library(VIM)
```

```
## Warning: package 'VIM' was built under R version 3.6.3
```

```
## Loading required package: colorspace
```

```
## Warning: package 'colorspace' was built under R version 3.6.3
```

```
## Loading required package: grid
```

```
## Loading required package: data.table
```

```
## VIM is ready to use.  
## Since version 4.0.0 the GUI is in its own package VIMGUI.  
##  
##      Please use the package to use the new (and old) GUI.
```

```
## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues
```

```
##  
## Attaching package: 'VIM'
```

```
## The following object is masked from 'package:datasets':  
##  
##     sleep
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 3.6.3
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.6.3
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 3.6.3
```

```
library(lattice)  
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.6.3
```

```
IDEgroup <- read_csv("C:/Users/Nick/OneDrive/Documents/Spring 2021/AMS 578 Regr/Project/IDEgroup355429.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   ID = col_double(),
##   E1 = col_double(),
##   E2 = col_double(),
##   E3 = col_double(),
##   E4 = col_double(),
##   E5 = col_double(),
##   E6 = col_double()
## )
```

```
IDGgroup <- read_csv("C:/Users/Nick/OneDrive/Documents/Spring 2021/AMS 578 Regr/Proje
ct/IDGgroup355429.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification -----
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

```
IDYgroup <- read_csv("C:/Users/Nick/OneDrive/Documents/Spring 2021/AMS 578 Regr/Proje
ct/IDYgroup355429.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   ID = col_double(),
##   Y = col_double()
## )
```

```
IDE <- subset(IDEgroup[order(IDEgroup$ID),], select = -c(X1,ID))
IDG <- subset(IDGgroup[order(IDGgroup$ID),], select = -c(X1,ID))
IDY <- subset(IDYgroup[order(IDYgroup$ID),], select = -c(X1,ID))
dataset <- cbind(IDE,IDG,IDY)

sum.stat = summary(dataset)
sum.stat
```

```

##          E1          E2          E3          E4
## Min.    : 14.51   Min.    : -232.8   Min.    : 453.7   Min.    : 400.4
## 1st Qu.: 455.52   1st Qu.: 207.3   1st Qu.: 828.7   1st Qu.: 758.2
## Median : 566.37   Median : 316.0   Median : 932.4   Median : 866.1
## Mean    : 566.15   Mean    : 313.7   Mean    : 934.0   Mean    : 867.4
## 3rd Qu.: 679.59   3rd Qu.: 415.5   3rd Qu.:1035.3   3rd Qu.: 975.5
## Max.    :1083.97   Max.    : 759.6   Max.    :1451.3   Max.    :1329.5
## NA's    :20       NA's    :30       NA's    :20       NA's    :20
##          E5          E6          R1          R2
## Min.    : 490.4   Min.    : 17.02   Min.    :0.000   Min.    :0.0000
## 1st Qu.: 965.1   1st Qu.: 378.26   1st Qu.:0.000   1st Qu.:0.0000
## Median :1063.4   Median : 490.49   Median :0.000   Median :0.0000
## Mean    :1067.3   Mean    : 485.51   Mean    :0.478   Mean    :0.4922
## 3rd Qu.:1171.1   3rd Qu.: 591.73   3rd Qu.:1.000   3rd Qu.:1.0000
## Max.    :1574.9   Max.    :1057.54   Max.    :1.000   Max.    :1.0000
## NA's    :30       NA's    :20       NA's    :30       NA's    :30
##          R3          R4          R5          R6
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :1.0000   Median :1.0000   Median :1.0000   Median :0.0000
## Mean    :0.5078   Mean    :0.5078   Mean    :0.5058   Mean    :0.4966
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
## NA's    :30       NA's    :30
##          R7          R8          R9          R10
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :0.0000   Median :1.0000   Median :0.0000   Median :1.0000
## Mean    :0.4948   Mean    :0.5168   Mean    :0.4936   Mean    :0.5125
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##          R11         R12         R13         R14
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :1.0000   Median :0.0000   Median :0.0000   Median :0.0000
## Mean    :0.5047   Mean    :0.4879   Mean    :0.4966   Mean    :0.4936
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
## NA's    :30       NA's    :30
##          R15         R16         R17         R18
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :1.0000   Median :1.0000   Median :1.0000   Median :1.0000
## Mean    :0.5168   Mean    :0.5168   Mean    :0.5196   Mean    :0.5082
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##
##          NA's    :30
##          R19         R20         R21         R22
## Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000

```

```
## Median :1.0000 Median :0.0000 Median :1.0000 Median :1.0000
## Mean :0.5009 Mean :0.4954 Mean :0.5034 Mean :0.5155
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000
##
## NA's :30
## R23 R24 R25 Y
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :2418
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:2817
## Median :0.0000 Median :0.0000 Median :1.0000 Median :2910
## Mean :0.4984 Mean :0.4839 Mean :0.5133 Mean :2909
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:3009
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :3320
## NA's :30 NA's :30 NA's :30
```

```
#Missing Value Analysis
data.na =is.na(dataset)
row.sums = rep(NA,length(data.na[,1]))
for(i in 1:length(data.na[,1])){
  row.sums[i] = sum(data.na[i,])
}
col.sums = rep(NA,length(data.na[1,]))
col.sd = rep(NA,length(data.na[1,]))
for(i in 1:length(data.na[1,])){
  col.sums[i] = sum(data.na[,i])
  col.sd[i] = sd(dataset[,i],na.rm=TRUE)
}
col.sums = data.frame(col.sums,row.names = colnames(dataset))
col.sddf = data.frame(col.sd, col.names = colnames(dataset))
col.sddf
```

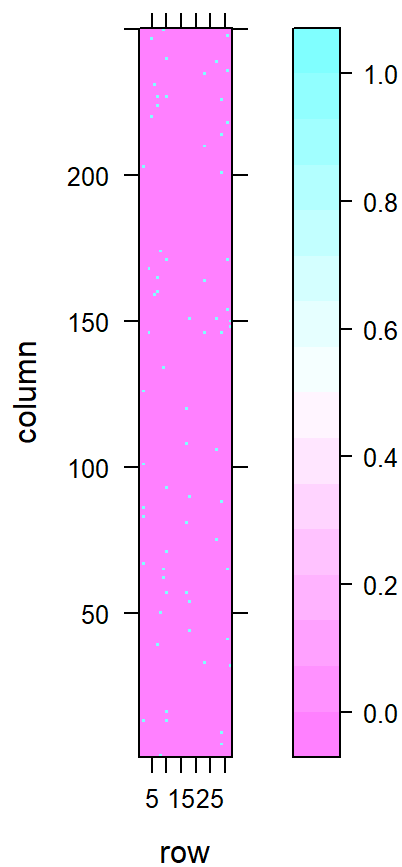
```
##          col.sd col.names
## 1  161.2249346      E1
## 2  158.5766711      E2
## 3  154.7397274      E3
## 4  163.8501074      E4
## 5  152.4820444      E5
## 6  160.0223517      E6
## 7    0.4996693      R1
## 8    0.5000950      R2
## 9    0.5000950      R3
## 10   0.5000950      R4
## 11   0.5001189      R5
## 12   0.5001412      R6
## 13   0.5001256      R7
## 14   0.4998714      R8
## 15   0.5001115      R9
## 16   0.4999963     R10
## 17   0.5001336     R11
## 18   0.5000087     R12
## 19   0.5001412     R13
## 20   0.5001115     R14
## 21   0.4998714     R15
## 22   0.4998714     R16
## 23   0.4997727     R17
## 24   0.5000847     R18
## 25   0.5001516     R19
## 26   0.5001315     R20
## 27   0.5001436     R21
## 28   0.4999108     R22
## 29   0.5001528     R23
## 30   0.4998915     R24
## 31   0.4999771     R25
## 32  142.4371304      Y
```

```
col.sums
```

```
##      col.sums
## E1      20
## E2      30
## E3      20
## E4      20
## E5      30
## E6      20
## R1      30
## R2      30
## R3      30
## R4      30
## R5       0
## R6       0
## R7       0
## R8       0
## R9       0
## R10      0
## R11     30
## R12     30
## R13       0
## R14       0
## R15       0
## R16       0
## R17     30
## R18       0
## R19       0
## R20       0
## R21     30
## R22       0
## R23     30
## R24       0
## R25     30
## Y       30
```

```
data.na.num = matrix(lapply(data.na, as.numeric), ncol= 32)
```

```
levelplot(t(data.na.num[1:250,]))
```



```
mat.cor = cor(dataset, use = "complete.obs")  
mat.cor
```



##		E1	E2	E3	E4	E5
## E1	1.0000000000	-0.0291753478	-0.006065138	-0.0389056554	0.0426528038	
## E2	-0.0291753478	1.0000000000	0.020354047	-0.0127121839	0.0112033533	
## E3	-0.0060651384	0.0203540469	1.000000000	-0.0205047920	0.0493808563	
## E4	-0.0389056554	-0.0127121839	-0.020504792	1.0000000000	0.0005160968	
## E5	0.0426528038	0.0112033533	0.049380856	0.0005160968	1.0000000000	
## E6	-0.0078457204	0.0182615603	-0.013094895	0.0060440104	-0.0034001148	
## R1	-0.0205042069	-0.0009266508	-0.001217854	0.0519644273	-0.0002269269	
## R2	0.0161067633	0.0028404237	-0.020510608	0.0062259978	-0.0401870094	
## R3	-0.0373520361	0.0020911566	0.015576871	0.0145900558	-0.0645906647	
## R4	0.0450281549	0.0119811526	-0.003080114	0.0313708062	0.0091559559	
## R5	-0.0068572834	-0.0601443172	0.046228679	0.0228108168	0.0142355006	
## R6	-0.0231359947	-0.0130929486	0.005053902	0.0091885580	0.0298159164	
## R7	-0.0276520664	-0.0844360261	0.014073503	-0.0452685965	0.0079501590	
## R8	-0.0243719641	0.0214746071	0.001994774	-0.0357130768	-0.0180463713	
## R9	-0.0488273599	0.0220481423	-0.019913107	-0.0222814309	-0.0110126051	
## R10	-0.0592221735	0.0219222124	-0.004523055	-0.0047857271	-0.0358133281	
## R11	-0.0575297706	-0.0491277992	-0.022785569	-0.0142969947	-0.0443035015	
## R12	-0.0117368052	-0.0030640723	0.027005391	-0.0115114596	0.0092338043	
## R13	0.0221535728	-0.0004878878	-0.009244051	0.0213580875	0.0202016682	
## R14	-0.0531519139	0.0162749727	-0.011220250	-0.0104517536	-0.0111767040	
## R15	0.0563133208	0.0352460861	0.001169642	-0.0553210151	0.0418573734	
## R16	0.0004655953	0.0212049646	-0.042726343	0.0191318702	0.0006096463	
## R17	0.0110371367	-0.0024362077	0.050642537	0.0116180092	-0.0042509172	
## R18	0.0323014094	0.0040156729	-0.022597028	-0.0333290417	0.0263536826	
## R19	0.0136536301	-0.0266219562	-0.009439145	0.0143203902	0.0094114522	
## R20	0.0231998673	-0.0214993667	0.019431763	-0.0554234455	0.0196029294	
## R21	-0.0056427557	-0.0652096030	-0.038602705	-0.0136955963	-0.0235180685	
## R22	0.0073031601	0.0021968647	-0.016670613	0.0039178242	-0.0452226619	
## R23	-0.0079243072	-0.0064684336	0.008572679	0.0016282966	0.0211345844	
## R24	0.0342083140	0.0092863933	0.048267811	0.0626945893	0.0393944640	
## R25	-0.0399251227	0.0377308299	-0.014014450	-0.0139384480	0.0187574656	
## Y	0.4128738791	-0.0386130082	0.210757608	0.4899262083	0.4259518615	
##		E6	R1	R2	R3	R4
## E1	-0.0078457204	-0.0205042069	0.0161067633	-0.037352036	0.0450281549	
## E2	0.0182615603	-0.0009266508	0.0028404237	0.002091157	0.0119811526	
## E3	-0.0130948945	-0.0012178538	-0.0205106080	0.015576871	-0.0030801141	
## E4	0.0060440104	0.0519644273	0.0062259978	0.014590056	0.0313708062	
## E5	-0.0034001148	-0.0002269269	-0.0401870094	-0.064590665	0.0091559559	
## E6	1.0000000000	-0.0020358305	0.0103374298	0.019844597	-0.0300014764	
## R1	-0.0020358305	1.0000000000	-0.0175833732	0.056470504	-0.0253966759	
## R2	0.0103374298	-0.0175833732	1.0000000000	0.012388288	-0.0041663467	
## R3	0.0198445969	0.0564705044	0.0123882883	1.000000000	-0.0056358470	
## R4	-0.0300014764	-0.0253966759	-0.0041663467	-0.005635847	1.0000000000	
## R5	0.0356397839	-0.0230768625	0.0206681519	-0.014033582	0.0058033894	
## R6	-0.0301525750	0.0207460644	-0.0168547808	-0.035139437	0.0041475973	
## R7	0.0074477451	0.0026043118	-0.0055992733	0.012145674	-0.0203099593	
## R8	-0.0271400529	-0.0539696178	0.0711619517	-0.048234357	0.0009297373	
## R9	-0.0072378351	0.0305457924	-0.0194722279	0.025705268	-0.0206068693	
## R10	0.0491774671	-0.0079161411	-0.0034455163	-0.006087414	0.0401059047	
## R11	-0.0283741709	0.0577708237	-0.0284965107	-0.013824555	-0.0170919800	

##	R12	0.0218504927	0.0100505844	0.0216562602	0.004267290	0.0137284196
##	R13	0.0141070945	0.0100505844	-0.0011338079	-0.064096757	-0.0285904999
##	R14	0.0166286950	0.0522079734	0.0150357017	-0.031478083	0.0071817617
##	R15	-0.0047155335	0.0012081999	0.0287217332	-0.009098060	-0.0284028308
##	R16	0.0183812145	-0.0458728493	-0.0379348802	0.012009864	-0.0137284196
##	R17	-0.0064522515	0.0184351760	-0.0260904954	0.013356919	0.0271449100
##	R18	-0.0168481525	0.0123724678	0.0399736197	0.028420241	-0.0203099593
##	R19	0.0124494306	0.0118170431	-0.0352326638	-0.049495038	-0.0106576829
##	R20	-0.0421138261	0.0252108208	0.0073101933	-0.013824555	0.0187100772
##	R21	0.0004276921	0.0333971844	-0.0333514326	0.004059970	-0.0284742746
##	R22	-0.0280483295	0.0160957799	-0.0118342412	0.015231869	0.0090778333
##	R23	-0.0289163934	-0.0155355869	-0.0008555068	0.010604151	0.0203282283
##	R24	0.0192830573	0.0376862525	0.0102337361	0.012423479	0.0511567425
##	R25	0.0015927031	0.0309897532	0.0387801303	0.023289829	-0.0185516388
##	Y	-0.0137955430	0.0403968565	0.0039786941	-0.027975465	0.0268654230
##		R5	R6	R7	R8	R9
##	E1	-0.006857283	-0.0231359947	-0.0276520664	-0.0243719641	-0.048827360
##	E2	-0.060144317	-0.0130929486	-0.0844360261	0.0214746071	0.022048142
##	E3	0.046228679	0.0050539021	0.0140735031	0.0019947742	-0.019913107
##	E4	0.022810817	0.0091885580	-0.0452685965	-0.0357130768	-0.022281431
##	E5	0.014235501	0.0298159164	0.0079501590	-0.0180463713	-0.011012605
##	E6	0.035639784	-0.0301525750	0.0074477451	-0.0271400529	-0.007237835
##	R1	-0.023076863	0.0207460644	0.0026043118	-0.0539696178	0.030545792
##	R2	0.020668152	-0.0168547808	-0.0055992733	0.0711619517	-0.019472228
##	R3	-0.014033582	-0.0351394374	0.0121456739	-0.0482343566	0.025705268
##	R4	0.005803389	0.0041475973	-0.0203099593	0.0009297373	-0.020606869
##	R5	1.000000000	0.0347383368	-0.0025480879	0.0118288800	-0.004894048
##	R6	0.034738337	1.000000000	0.0121276168	0.0330890925	-0.019765504
##	R7	-0.002548088	0.0121276168	1.000000000	-0.0172067622	-0.002197465
##	R8	0.011828880	0.0330890925	-0.0172067622	1.000000000	-0.009711520
##	R9	-0.004894048	-0.0197655037	-0.0021974652	-0.0097115204	1.000000000
##	R10	0.005014697	0.0263648237	-0.0107936666	0.0033264511	-0.035092281
##	R11	-0.021957706	-0.0203324382	-0.0349837396	0.0187348361	0.020325568
##	R12	-0.039533430	-0.0575272579	0.0383564043	0.0467777231	-0.017913911
##	R13	-0.029765973	-0.0184605360	-0.0104731181	-0.0215966471	0.021185447
##	R14	-0.026392069	0.0011630557	-0.0657855727	-0.0279800326	0.004613684
##	R15	0.002182735	-0.0351858621	-0.0106576829	-0.0027232077	0.003040643
##	R16	0.023254335	-0.0401395468	0.0332602285	0.0118288800	0.004880792
##	R17	0.014610043	0.0327573016	0.0217286534	-0.0164242252	-0.015117301
##	R18	0.007217817	-0.0041475973	0.0105456067	-0.0367391921	0.014091345
##	R19	0.030372866	-0.0201542626	-0.0251477880	0.0515581625	0.006733420
##	R20	0.036636790	0.0187274548	0.0203467124	0.0415223089	0.023583278
##	R21	0.002426207	0.0008024917	-0.0008183176	-0.0089687763	0.031797797
##	R22	0.013426837	0.0086493962	-0.0286109360	0.0085072899	-0.040597303
##	R23	0.022018684	-0.0089174809	0.0284928880	0.0334171241	-0.072589612
##	R24	-0.005317035	0.0222571896	0.0269729772	0.0321602103	-0.022872314
##	R25	0.018162402	-0.0126284113	-0.0009856055	0.0067150555	0.039595537
##	Y	0.005226562	0.0161334440	-0.0280234341	-0.0110401946	-0.046392020
##		R10	R11	R12	R13	R14
##	E1	-0.059222174	-0.0575297706	-0.011736805	0.0221535728	-0.0531519139
##	E2	0.021922212	-0.0491277992	-0.003064072	-0.0004878878	0.0162749727

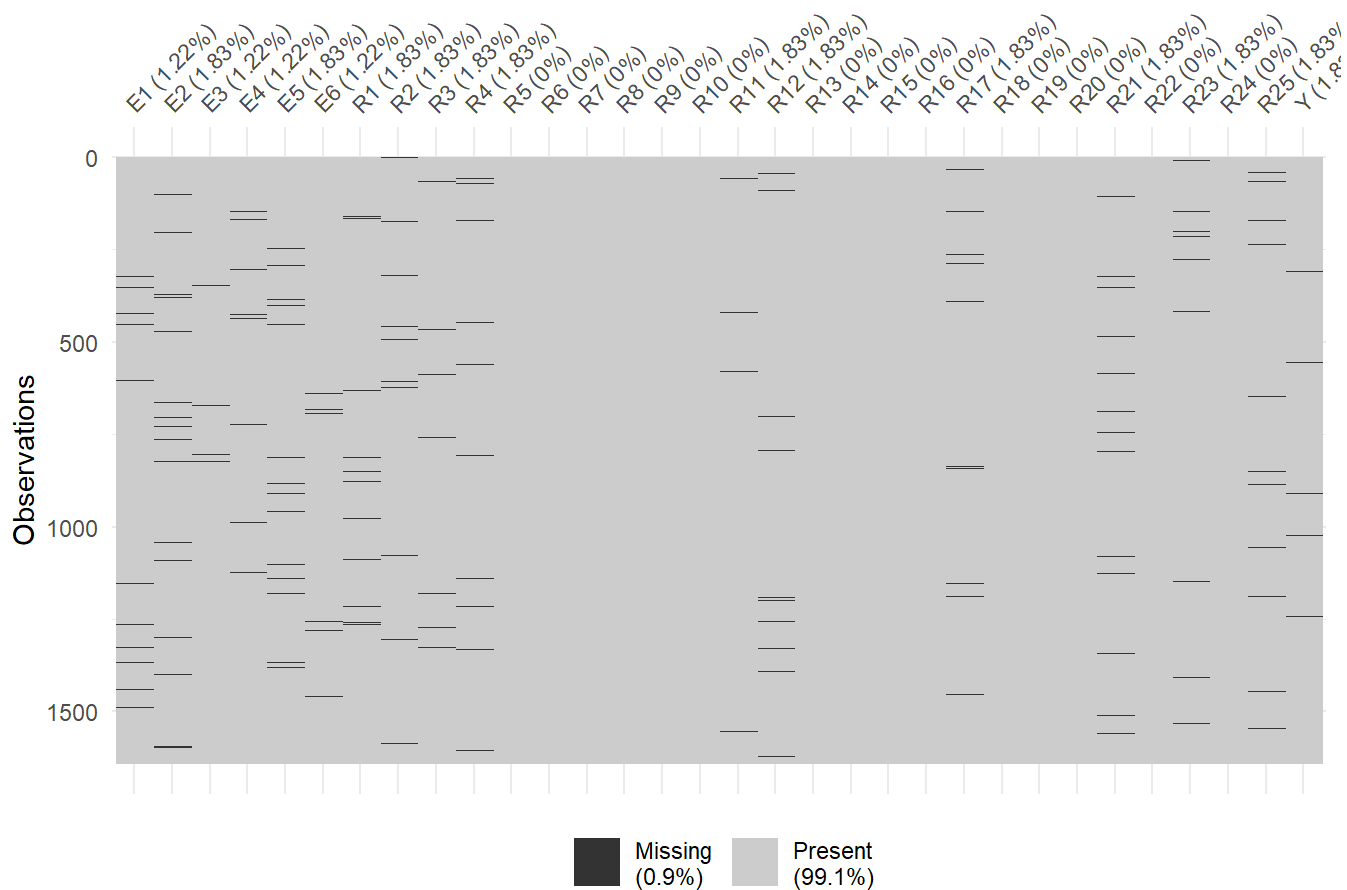
##	E3	-0.004523055	-0.0227855685	0.027005391	-0.0092440510	-0.0112202503
##	E4	-0.004785727	-0.0142969947	-0.011511460	0.0213580875	-0.0104517536
##	E5	-0.035813328	-0.0443035015	0.009233804	0.0202016682	-0.0111767040
##	E6	0.049177467	-0.0283741709	0.021850493	0.0141070945	0.0166286950
##	R1	-0.007916141	0.0577708237	0.010050584	0.0100505844	0.0522079734
##	R2	-0.003445516	-0.0284965107	0.021656260	-0.0011338079	0.0150357017
##	R3	-0.006087414	-0.0138245553	0.004267290	-0.0640967575	-0.0314780827
##	R4	0.040105905	-0.0170919800	0.013728420	-0.0285904999	0.0071817617
##	R5	0.005014697	-0.0219577059	-0.039533430	-0.0297659731	-0.0263920686
##	R6	0.026364824	-0.0203324382	-0.057527258	-0.0184605360	0.0011630557
##	R7	-0.010793667	-0.0349837396	0.038356404	-0.0104731181	-0.0657855727
##	R8	0.003326451	0.0187348361	0.046777723	-0.0215966471	-0.0279800326
##	R9	-0.035092281	0.0203255676	-0.017913911	0.0211854469	0.0046136844
##	R10	1.000000000	0.0285273691	0.043846797	0.0112724675	-0.0015189391
##	R11	0.028527369	1.0000000000	-0.010594792	0.0187024562	-0.0008344592
##	R12	0.043846797	-0.0105947918	1.000000000	0.0265101542	0.0035980754
##	R13	0.011272467	0.0187024562	0.026510154	1.0000000000	-0.0582713347
##	R14	-0.001518939	-0.0008344592	0.003598075	-0.0582713347	1.0000000000
##	R15	-0.022492788	-0.0073124739	-0.002182735	0.0075839453	0.0402353674
##	R16	0.031074161	0.0105947918	-0.026510154	-0.0004636032	-0.0329046381
##	R17	0.034978024	0.0252887178	-0.017868798	-0.0276450642	-0.0175690348
##	R18	-0.040105905	0.0561487696	0.012313992	-0.0072178166	0.0090970747
##	R19	-0.010078953	0.0040574828	0.005438295	-0.0140950656	0.0020929703
##	R20	0.028527369	-0.0024416726	0.018702456	0.0284682055	-0.0236244681
##	R21	0.039866766	0.0447525224	-0.028468205	-0.0089367069	-0.0121884030
##	R22	-0.008134845	-0.0056791845	0.058205922	-0.0004026991	-0.0002596721
##	R23	-0.044692555	-0.0073250373	-0.002487134	0.0365759667	0.0040080640
##	R24	0.006446159	-0.0122241296	0.037876259	0.0508999486	-0.0208657914
##	R25	-0.078489129	0.0220035578	-0.008392228	0.0046346720	0.0048314557
##	Y	-0.035517125	-0.0375238023	0.020574221	0.0233505340	-0.0336172394
##		R15	R16	R17	R18	R19
##	E1	0.056313321	0.0004655953	0.011037137	0.032301409	0.013653630
##	E2	0.035246086	0.0212049646	-0.002436208	0.004015673	-0.026621956
##	E3	0.001169642	-0.0427263431	0.050642537	-0.022597028	-0.009439145
##	E4	-0.055321015	0.0191318702	0.011618009	-0.033329042	0.014320390
##	E5	0.041857373	0.0006096463	-0.004250917	0.026353683	0.009411452
##	E6	-0.004715533	0.0183812145	-0.006452251	-0.016848152	0.012449431
##	R1	0.001208200	-0.0458728493	0.018435176	0.012372468	0.011817043
##	R2	0.028721733	-0.0379348802	-0.026090495	0.039973620	-0.035232664
##	R3	-0.009098060	0.0120098644	0.013356919	0.028420241	-0.049495038
##	R4	-0.028402831	-0.0137284196	0.027144910	-0.020309959	-0.010657683
##	R5	0.002182735	0.0232543354	0.014610043	0.007217817	0.030372866
##	R6	-0.035185862	-0.0401395468	0.032757302	-0.004147597	-0.020154263
##	R7	-0.010657683	0.0332602285	0.021728653	0.010545607	-0.025147788
##	R8	-0.002723208	0.0118288800	-0.016424225	-0.036739192	0.051558163
##	R9	0.003040643	0.0048807919	-0.015117301	0.014091345	0.006733420
##	R10	-0.022492788	0.0310741614	0.034978024	-0.040105905	-0.010078953
##	R11	-0.007312474	0.0105947918	0.025288718	0.056148770	0.004057483
##	R12	-0.002182735	-0.0265101542	-0.017868798	0.012313992	0.005438295
##	R13	0.007583945	-0.0004636032	-0.027645064	-0.007217817	-0.014095066
##	R14	0.040235367	-0.0329046381	-0.017569035	0.009097075	0.002092970

```
## R15 1.000000000 -0.0238617461 -0.016120145 -0.039953068 -0.013643660
## R16 -0.023861746 1.0000000000 -0.034271289 0.003962515 -0.008693856
## R17 -0.016120145 -0.0342712886 1.000000000 -0.023886672 0.025895635
## R18 -0.039953068 0.0039625151 -0.023886672 1.000000000 -0.057698216
## R19 -0.013643660 -0.0086938555 0.025895635 -0.057698216 1.000000000
## R20 0.054532357 0.0138500415 -0.033358626 -0.038238472 -0.028492428
## R21 -0.021982446 0.0349787050 0.043133184 -0.030110910 0.028492428
## R22 -0.059721895 -0.0061093699 -0.005093966 0.003944235 -0.021672499
## R23 0.030142516 0.0057423923 0.010702466 0.025238147 -0.013867517
## R24 -0.020066438 -0.0150848023 0.016424225 -0.008836478 -0.015745862
## R25 -0.009371945 0.0051355027 -0.023401139 -0.010754228 -0.016679784
## Y -0.007838996 -0.0220443279 0.030335120 -0.018107955 -0.015917939
## R20 R21 R22 R23 R24
## E1 0.023199867 -0.0056427557 0.0073031601 -0.0079243072 0.034208314
## E2 -0.021499367 -0.0652096030 0.0021968647 -0.0064684336 0.009286393
## E3 0.019431763 -0.0386027046 -0.0166706126 0.0085726788 0.048267811
## E4 -0.055423445 -0.0136955963 0.0039178242 0.0016282966 0.062694589
## E5 0.019602929 -0.0235180685 -0.0452226619 0.0211345844 0.039394464
## E6 -0.042113826 0.0004276921 -0.0280483295 -0.0289163934 0.019283057
## R1 0.025210821 0.0333971844 0.0160957799 -0.0155355869 0.037686252
## R2 0.007310193 -0.0333514326 -0.0118342412 -0.0008555068 0.010233736
## R3 -0.013824555 0.0040599700 0.0152318690 0.0106041506 0.012423479
## R4 0.018710077 -0.0284742746 0.0090778333 0.0203282283 0.051156742
## R5 0.036636790 0.0024262073 0.0134268373 0.0220186843 -0.005317035
## R6 0.018727455 0.0008024917 0.0086493962 -0.0089174809 0.022257190
## R7 0.020346712 -0.0008183176 -0.0286109360 0.0284928880 0.026972977
## R8 0.041522309 -0.0089687763 0.0085072899 0.0334171241 0.032160210
## R9 0.023583278 0.0317977972 -0.0405973032 -0.0725896118 -0.022872314
## R10 0.028527369 0.0398667664 -0.0081348446 -0.0446925549 0.006446159
## R11 -0.002441673 0.0447525224 -0.0056791845 -0.0073250373 -0.012224130
## R12 0.018702456 -0.0284682055 0.0582059225 -0.0024871340 0.037876259
## R13 0.028468205 -0.0089367069 -0.0004026991 0.0365759667 0.050899949
## R14 -0.023624468 -0.0121884030 -0.0002596721 0.0040080640 -0.020865791
## R15 0.054532357 -0.0219824459 -0.0597218954 0.0301425156 -0.020066438
## R16 0.013850042 0.0349787050 -0.0061093699 0.0057423923 -0.015084802
## R17 -0.033358626 0.0431331837 -0.0050939657 0.0107024661 0.016424225
## R18 -0.038238472 -0.0301109098 0.0039442351 0.0252381469 -0.008836478
## R19 -0.028492428 0.0284924281 -0.0216724994 -0.0138675171 -0.015745862
## R20 1.000000000 -0.0203410927 -0.0284674422 -0.0268531735 -0.021990189
## R21 -0.020341093 1.0000000000 -0.0268754693 0.0333625523 -0.013818696
## R22 -0.028467442 -0.0268754693 1.0000000000 -0.0398259167 0.063127747
## R23 -0.026853174 0.0333625523 -0.0398259167 1.0000000000 0.008902580
## R24 -0.021990189 -0.0138186965 0.0631277472 0.0089025804 1.000000000
## R25 0.002466624 -0.0252597134 -0.0112469129 -0.0414425723 -0.016485541
## Y -0.002910912 -0.0325571243 -0.0019256183 -0.0212862441 0.068907284
## R25 Y
## E1 -0.0399251227 0.412873879
## E2 0.0377308299 -0.038613008
## E3 -0.0140144500 0.210757608
## E4 -0.0139384480 0.489926208
## E5 0.0187574656 0.425951862
```

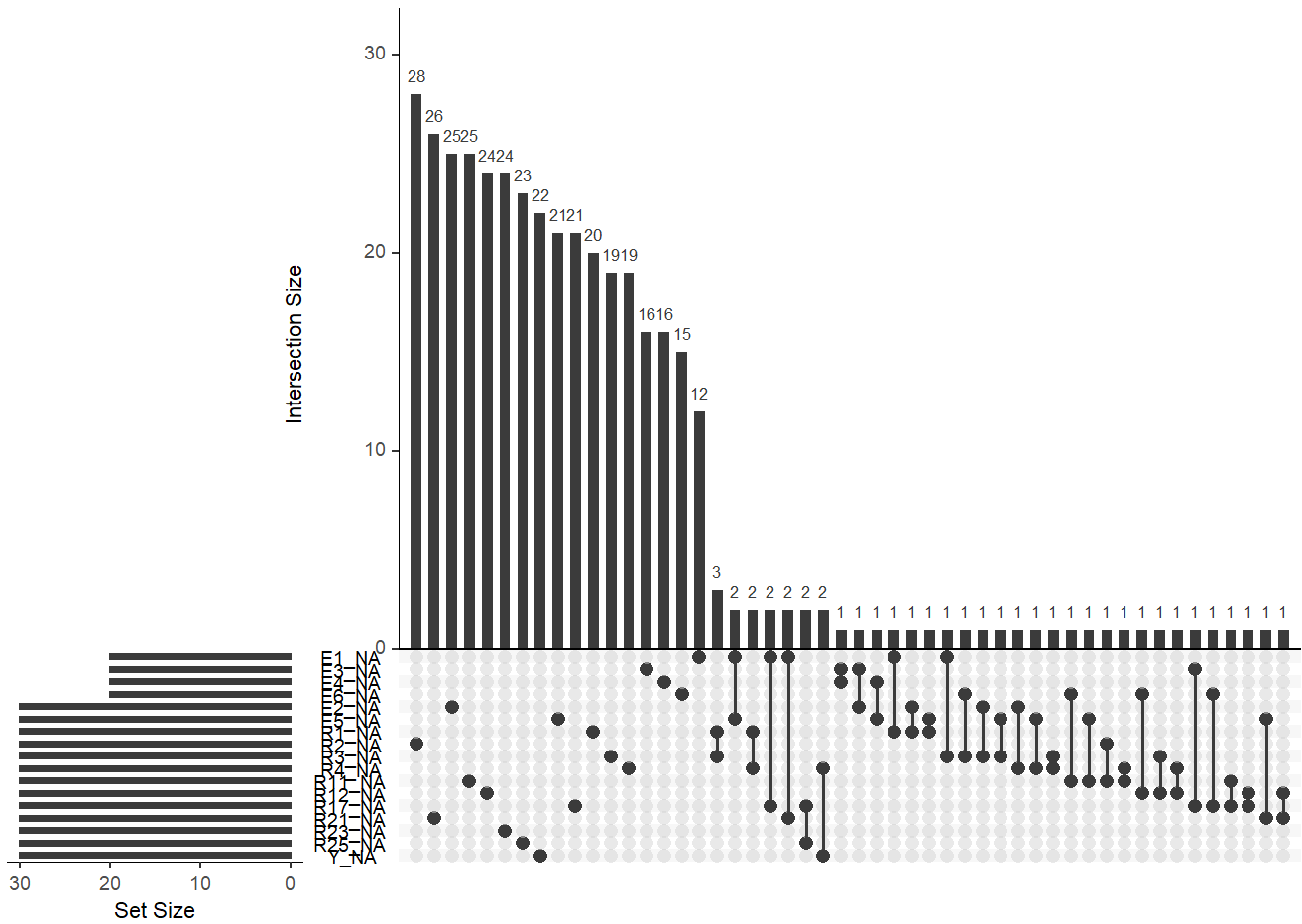
```
## E6    0.0015927031 -0.013795543
## R1     0.0309897532  0.040396856
## R2     0.0387801303  0.003978694
## R3     0.0232898290 -0.027975465
## R4    -0.0185516388  0.026865423
## R5     0.0181624024  0.005226562
## R6    -0.0126284113  0.016133444
## R7    -0.0009856055 -0.028023434
## R8     0.0067150555 -0.011040195
## R9     0.0395955366 -0.046392020
## R10   -0.0784891292 -0.035517125
## R11    0.0220035578 -0.037523802
## R12   -0.0083922276  0.020574221
## R13    0.0046346720  0.023350534
## R14    0.0048314557 -0.033617239
## R15   -0.0093719452 -0.007838996
## R16    0.0051355027 -0.022044328
## R17   -0.0234011392  0.030335120
## R18   -0.0107542276 -0.018107955
## R19   -0.0166797837 -0.015917939
## R20    0.0024666240 -0.002910912
## R21   -0.0252597134 -0.032557124
## R22   -0.0112469129 -0.001925618
## R23   -0.0414425723 -0.021286244
## R24   -0.0164855409  0.068907284
## R25    1.0000000000 -0.048812037
## Y     -0.0488120368  1.000000000
```

```
mat.cor.e = cor(IDE,use = "complete.obs")
mat.cor.g = cor(IDG,use = "complete.obs")

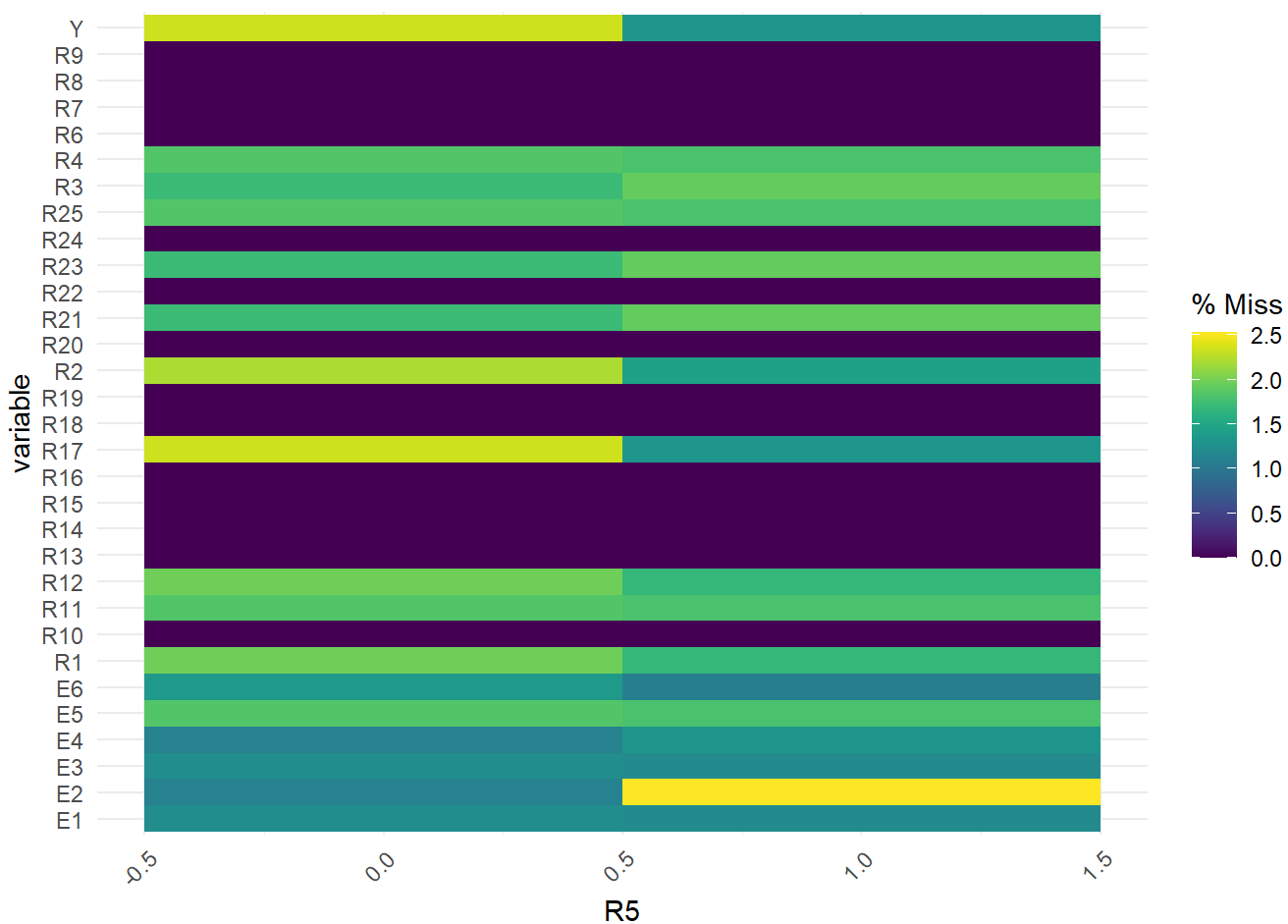
vis_miss(dataset)
```



```
gg_miss_upset(subset(dataset, select=-c(R5, R6, R7, R8, R9, R10, R13, R14, R15, R16, R18, R19, R20, R22, R24)), nsets=40, nintersects=50)
```

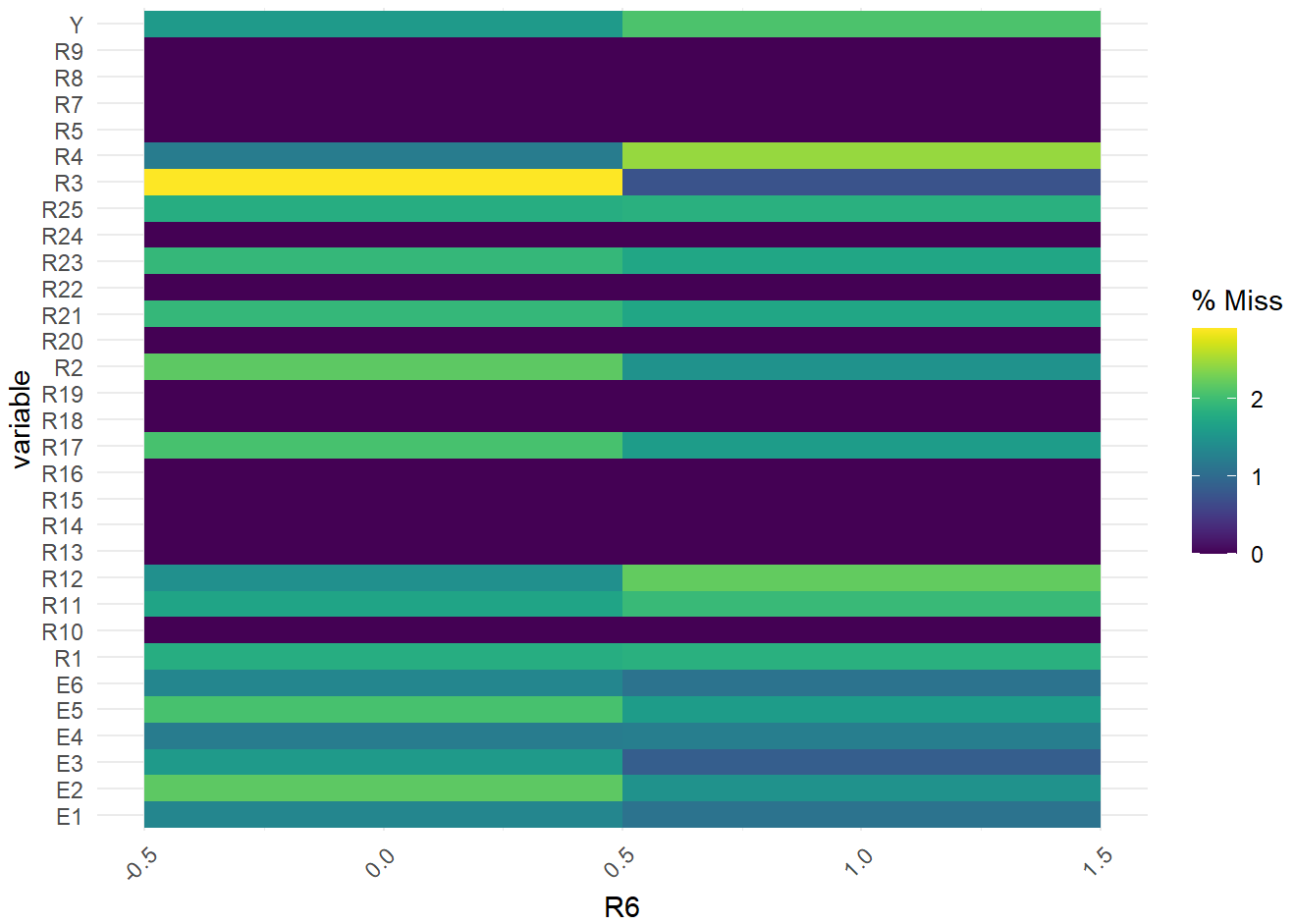


```
gg_miss_fct (dataset, fct=R5)
```

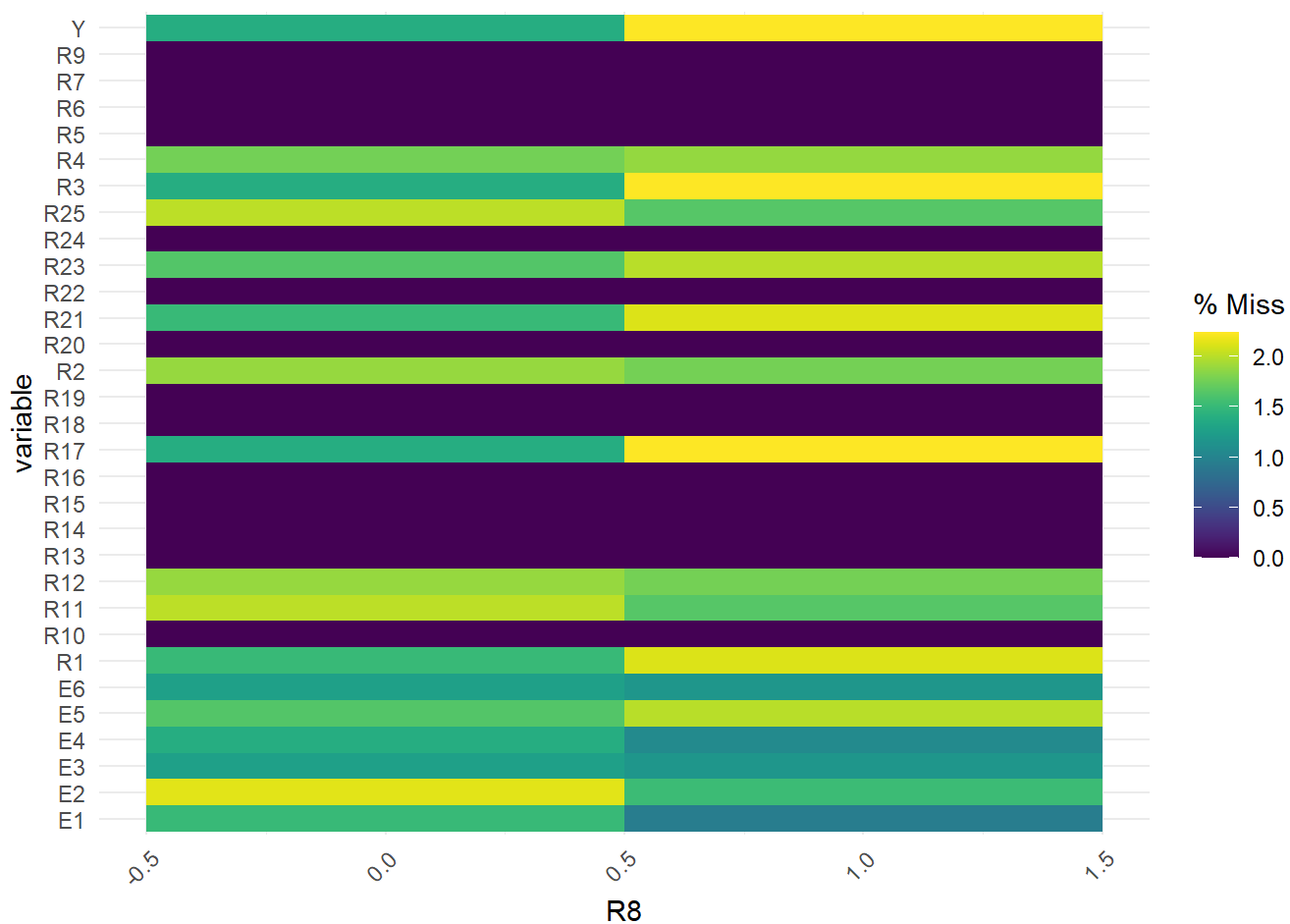


```
gg_miss_fct(dataset, fct=R6)
```



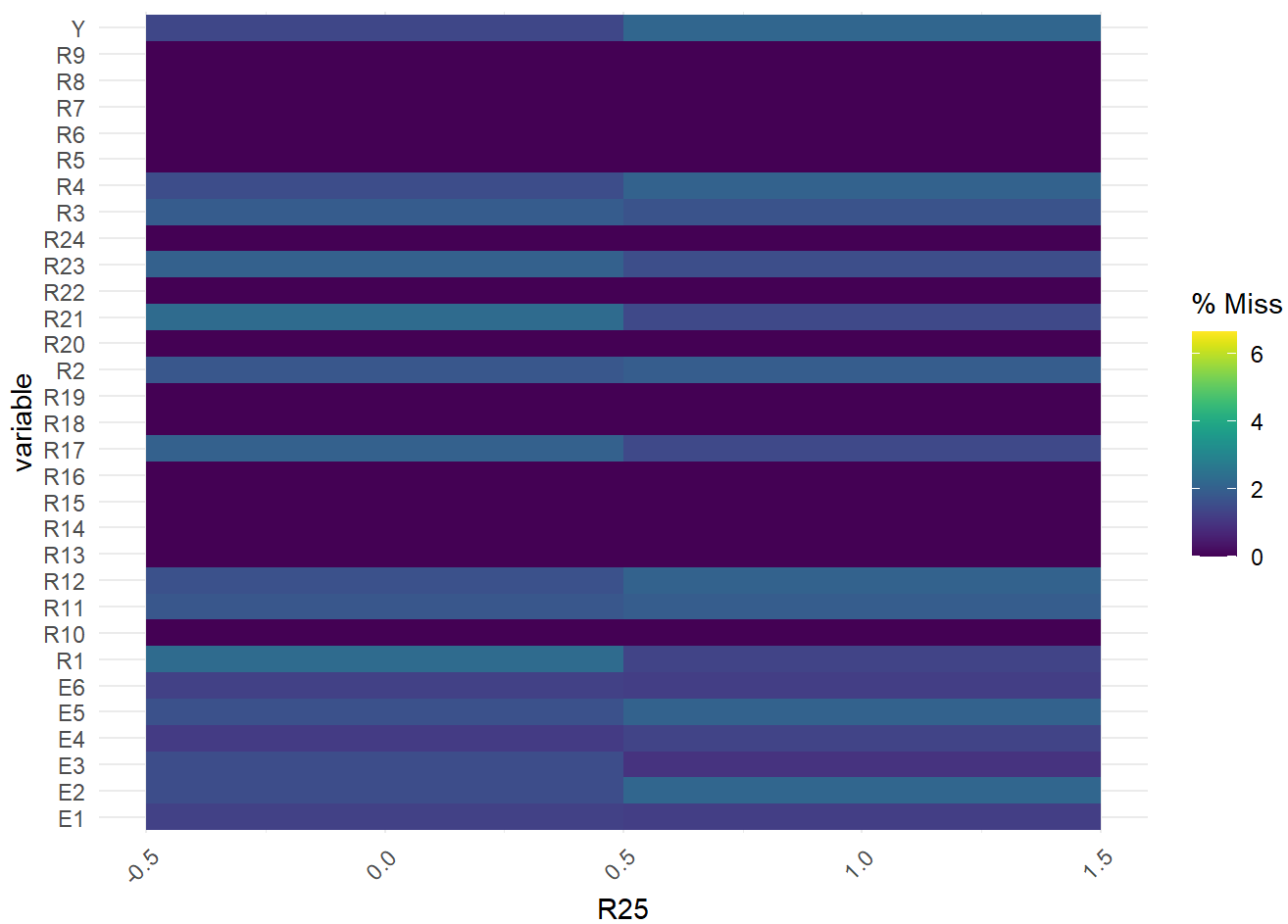


```
gg_miss_fct(dataset, fct=R8)
```



```
gg_miss_fct(dataset, fct=R25)
```

```
## Warning: Removed 31 rows containing missing values (geom_tile).
```



```
#Imputation
```

```
cart.impute<- mice(dataset, m=6, maxit = 10, method = 'cart', seed = 500)
```

##	iter	imp	variable																		
##	1	1	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	1	2	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	1	3	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	1	4	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	1	5	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	1	6	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	2	1	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	2	2	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	2	3	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	2	4	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	2	5	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	2	6	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	3	1	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	3	2	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	3	3	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	3	4	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	3	5	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	3	6	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	4	1	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	4	2	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	4	3	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	4	4	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	4	5	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	4	6	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	5	1	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	5	2	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	5	3	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	5	4	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	5	5	E1	E2	E3	E4	E5	E6	R1	R2	R3	R4	R11	R12	R17	R21	R23	R25	Y		
##	5	6	E1																		

```
## 9 2 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 9 3 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 9 4 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 9 5 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 9 6 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 10 1 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 10 2 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 10 3 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 10 4 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 10 5 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
## 10 6 E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R11 R12 R17 R21 R23 R25 Y
```

```
summary(cart.impute)
```

```
## Class: mids
## Number of multiple imputations: 6
## Imputation methods:
##      E1      E2      E3      E4      E5      E6      R1      R2      R3      R4      R5
## "cart" "cart" "cart" "cart" "cart" "cart" "cart" "cart" "cart" "cart" ""
##      R6      R7      R8      R9      R10     R11     R12     R13     R14     R15     R16
##      ""      ""      ""      ""      ""      "cart" "cart"      ""      ""      ""      ""
##      R17     R18     R19     R20     R21     R22     R23     R24     R25     Y
## "cart"      ""      ""      ""      "cart"      ""      "cart"      ""      "cart" "cart"
## PredictorMatrix:
##      E1 E2 E3 E4 E5 E6 R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11 R12 R13 R14 R15 R16 R17
## E1  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## E2  1  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## E3  1  1  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## E4  1  1  1  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## E5  1  1  1  1  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## E6  1  1  1  1  1  0  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##      R18 R19 R20 R21 R22 R23 R24 R25 Y
## E1  1  1  1  1  1  1  1  1  1
## E2  1  1  1  1  1  1  1  1  1
## E3  1  1  1  1  1  1  1  1  1
## E4  1  1  1  1  1  1  1  1  1
## E5  1  1  1  1  1  1  1  1  1
## E6  1  1  1  1  1  1  1  1  1
```

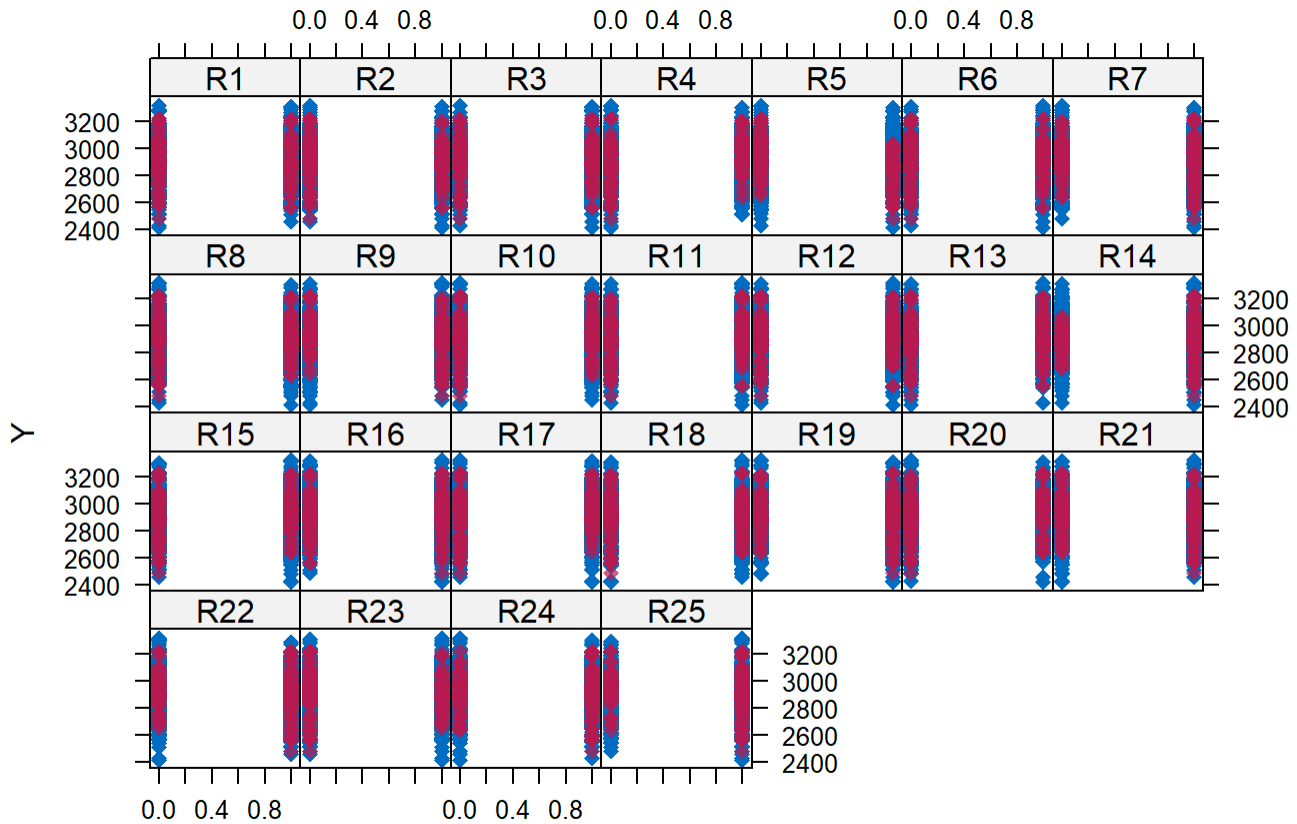
```
#This generates 16 different data sets using the cart method,
#we will take the 15th but return to pool them all after
#model building
data.cart <- complete(cart.impute,5)

#Inspection
summary(data.cart)
```

##	E1	E2	E3	E4
##	Min. : 14.51	Min. : -232.8	Min. : 453.7	Min. : 400.4
##	1st Qu.: 455.37	1st Qu.: 206.7	1st Qu.: 828.4	1st Qu.: 758.3
##	Median : 566.36	Median : 315.3	Median : 933.3	Median : 865.9
##	Mean : 566.11	Mean : 312.9	Mean : 933.9	Mean : 866.9
##	3rd Qu.: 679.59	3rd Qu.: 415.4	3rd Qu.: 1035.7	3rd Qu.: 973.5
##	Max. : 1083.97	Max. : 759.6	Max. : 1451.3	Max. : 1329.5
##	E5	E6	R1	R2
##	Min. : 490.4	Min. : 17.02	Min. : 0.0000	Min. : 0.000
##	1st Qu.: 966.1	1st Qu.: 378.54	1st Qu.: 0.0000	1st Qu.: 0.000
##	Median : 1064.0	Median : 490.74	Median : 0.0000	Median : 0.000
##	Mean : 1068.1	Mean : 485.88	Mean : 0.4784	Mean : 0.493
##	3rd Qu.: 1171.5	3rd Qu.: 591.73	3rd Qu.: 1.0000	3rd Qu.: 1.000
##	Max. : 1574.9	Max. : 1057.54	Max. : 1.0000	Max. : 1.000
##	R3	R4	R5	R6
##	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
##	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
##	Median : 1.0000	Median : 1.0000	Median : 1.0000	Median : 0.0000
##	Mean : 0.5094	Mean : 0.5046	Mean : 0.5058	Mean : 0.4966
##	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000
##	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000
##	R7	R8	R9	R10
##	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
##	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
##	Median : 0.0000	Median : 1.0000	Median : 0.0000	Median : 1.0000
##	Mean : 0.4948	Mean : 0.5168	Mean : 0.4936	Mean : 0.5125
##	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000
##	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000
##	R11	R12	R13	R14
##	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
##	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
##	Median : 1.0000	Median : 0.0000	Median : 0.0000	Median : 0.0000
##	Mean : 0.5034	Mean : 0.4881	Mean : 0.4966	Mean : 0.4936
##	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000
##	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000
##	R15	R16	R17	R18
##	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
##	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
##	Median : 1.0000	Median : 1.0000	Median : 1.0000	Median : 1.0000
##	Mean : 0.5168	Mean : 0.5168	Mean : 0.5198	Mean : 0.5082
##	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000
##	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000
##	R19	R20	R21	R22
##	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
##	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
##	Median : 1.0000	Median : 0.0000	Median : 1.0000	Median : 1.0000
##	Mean : 0.5009	Mean : 0.4954	Mean : 0.5034	Mean : 0.5155
##	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000	3rd Qu.: 1.0000
##	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000	Max. : 1.0000
##	R23	R24	R25	Y
##	Min. : 0.000	Min. : 0.0000	Min. : 0.0000	Min. : 2418

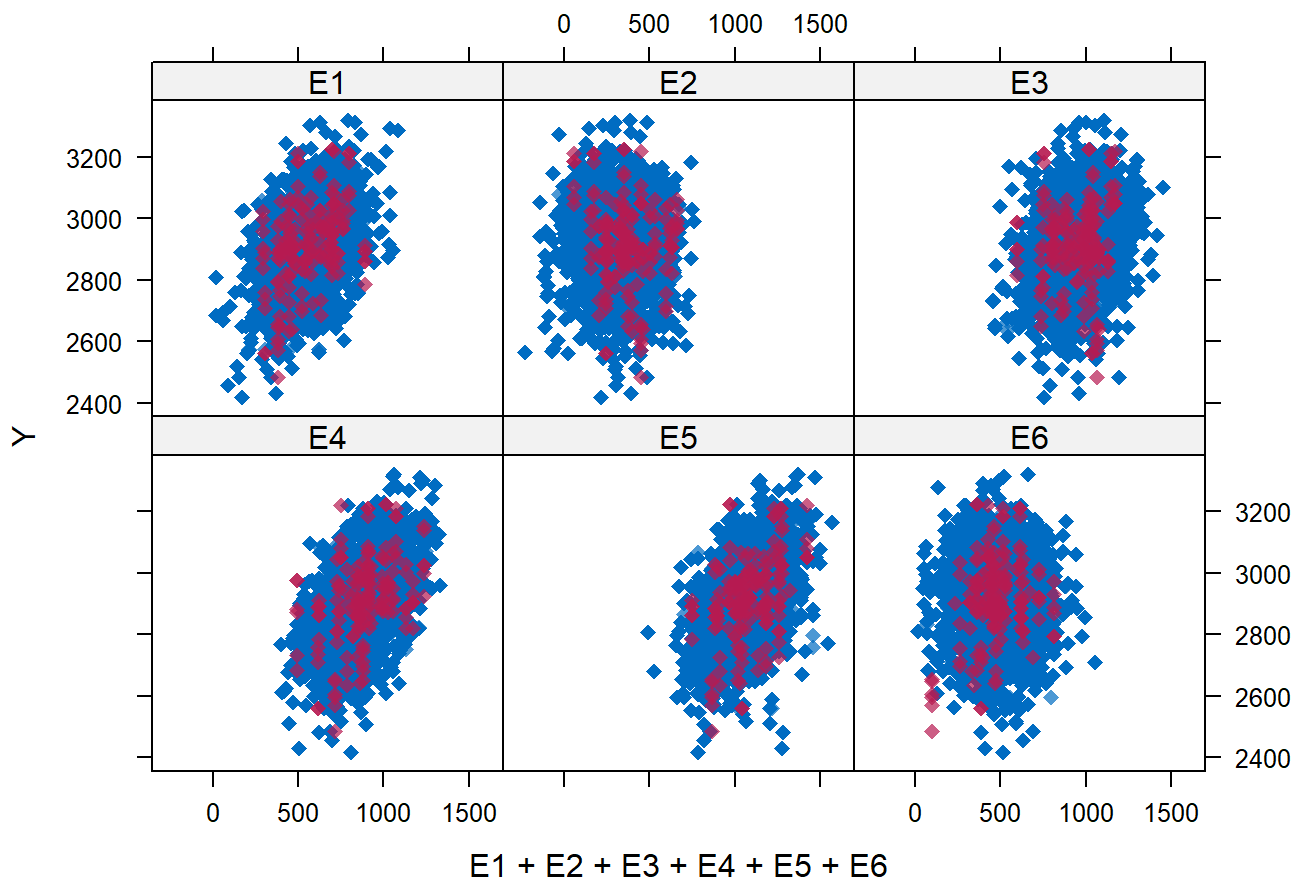
```
## 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:2819
## Median :0.000 Median :0.0000 Median :1.0000 Median :2910
## Mean :0.496 Mean :0.4839 Mean :0.5131 Mean :2910
## 3rd Qu.:1.000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:3008
## Max. :1.000 Max. :1.0000 Max. :1.0000 Max. :3320
```

```
#boxplot(log(data.cart$Y))
xyplot(cart.impute,Y~ R1+R2+R3+R4+R5+R6+R7+R8+R9+R10+R11+R12+R13+R14+R15+R16+R17+R18+
R19+R20+R21+R22+R23+R24+R25,pch=18,cex=1)
```



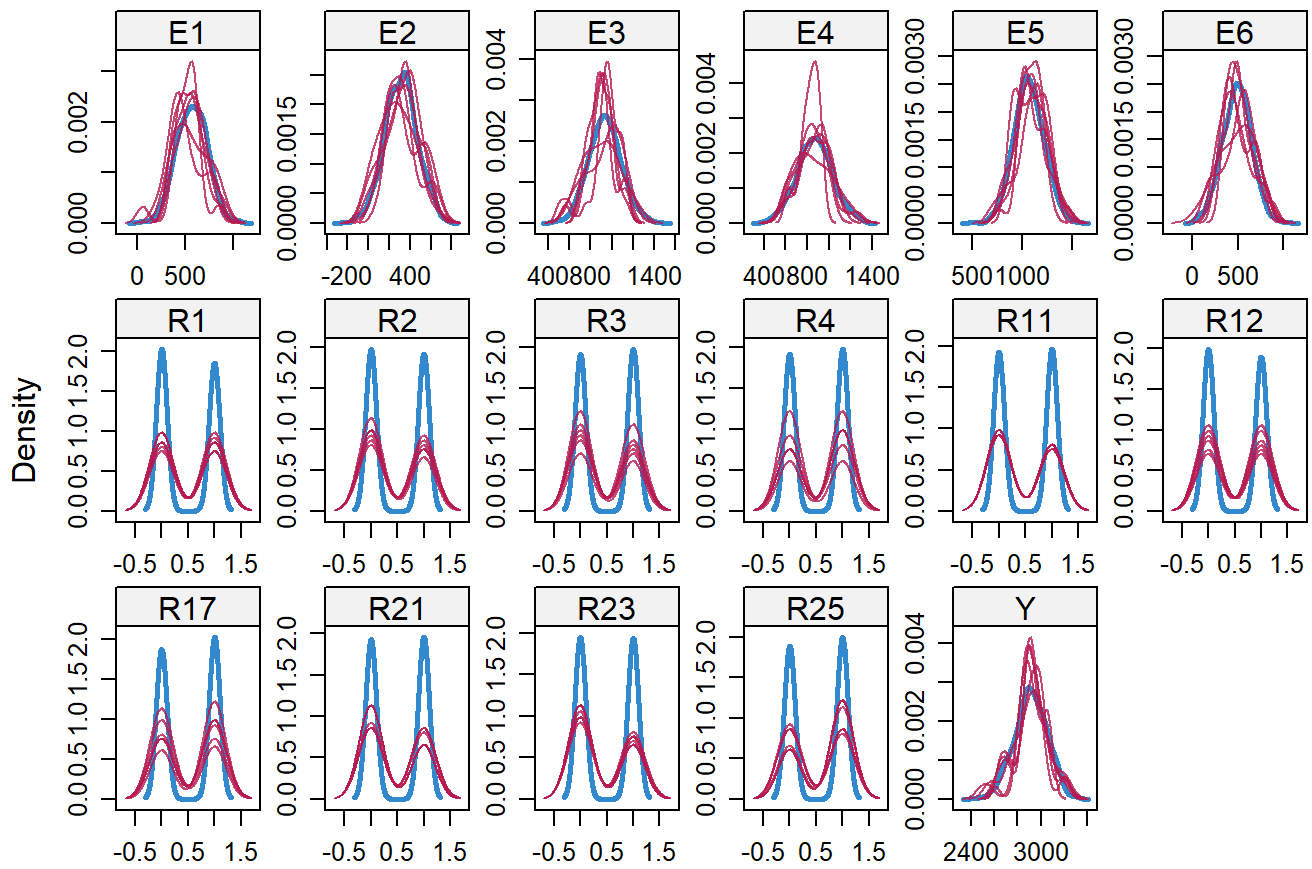
· R7 + R8 + R9 + R10 + R11 + R12 + R13 + R14 + R15 + R16 + R17 + R18 + R19 + R20 + |

```
xyplot(cart.impute,Y~ E1+E2+E3+E4+E5+E6,pch=18,cex=1)
```

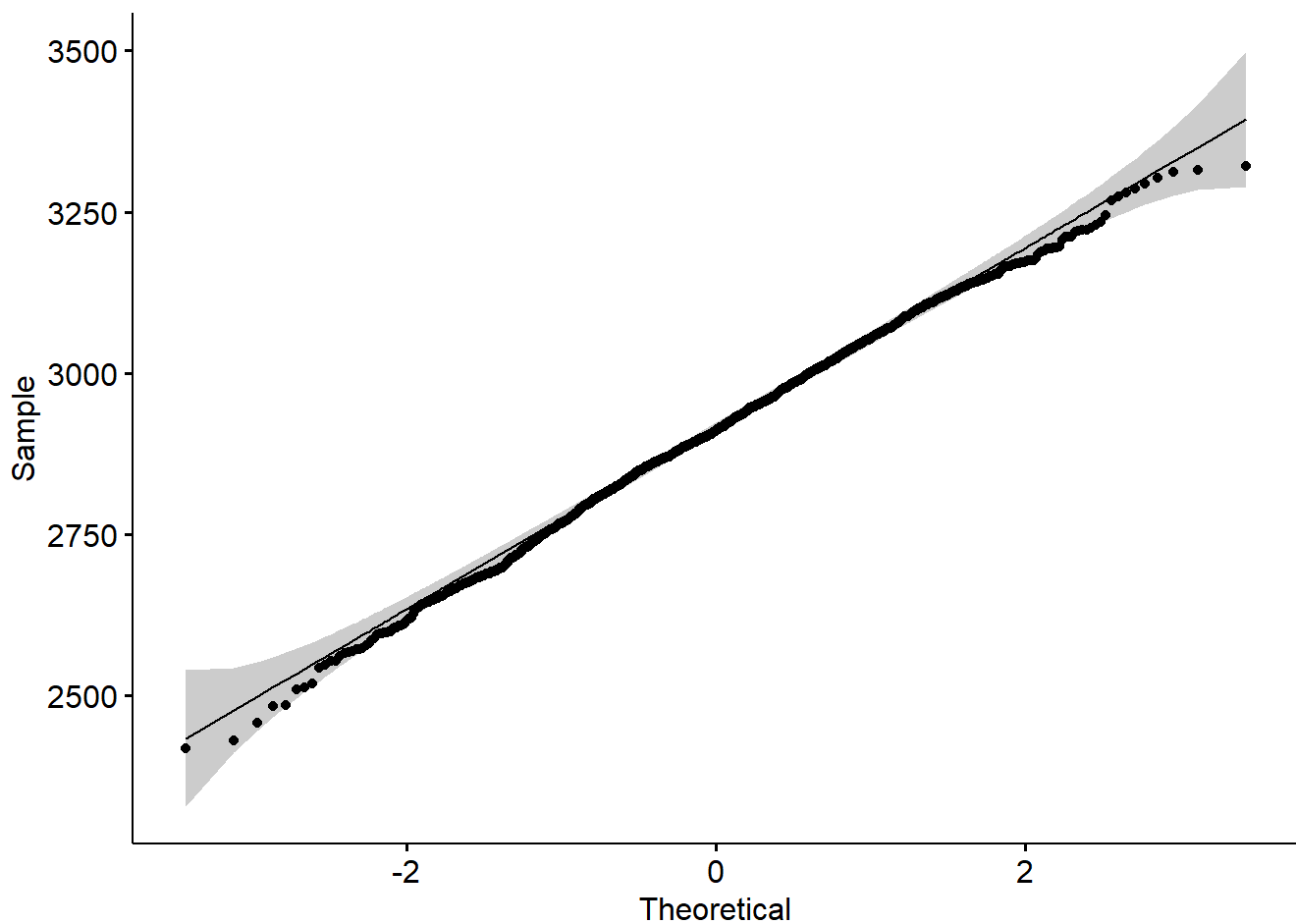


```
densityplot(cart.impute)
```





```
ggqqplot(data.cart$Y)
```



```
#model building
```

```
test = data.cart
```

```
M1 = lm( log(Y) ~(.)^3,data=data.cart)
```

```
leaps_int = regsubsets(model.matrix(M1)[,-1], I(log(data.cart$Y)),really.big = TRUE,  
                      method = "forward", nbest = 1, intercept = TRUE, nvmax = 500)
```

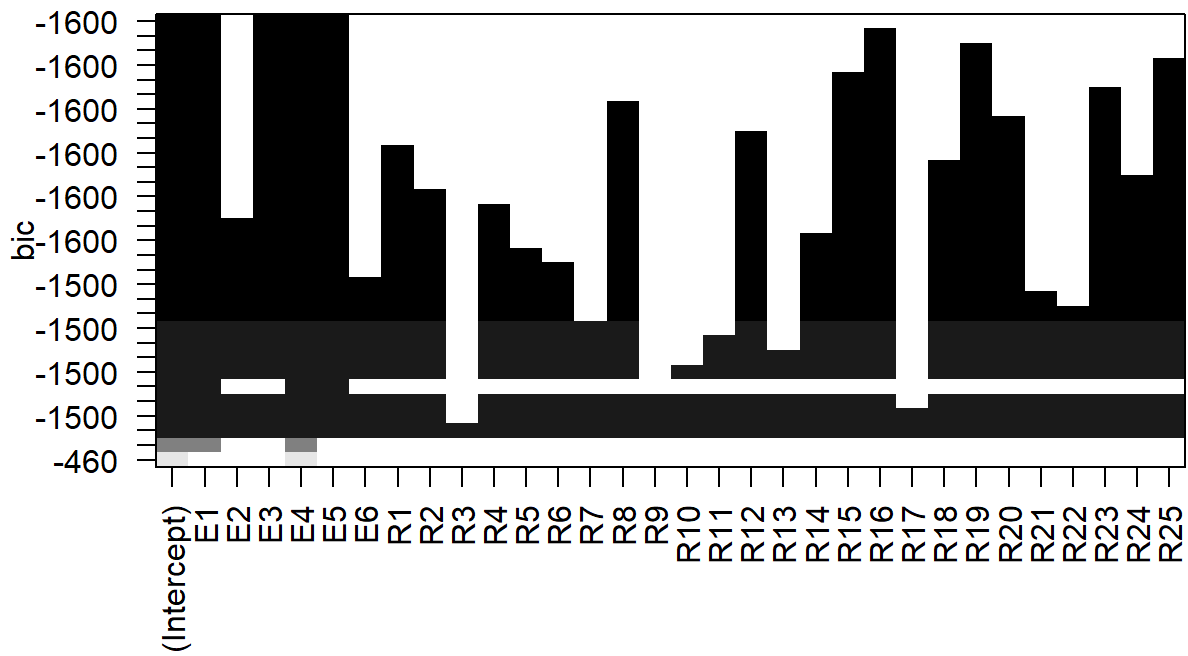
```
## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 3351  
## linear dependencies found
```

```
leaps_noint = regsubsets(Y~.,data=data.cart,nbest=1,nvmax = 500,method = "exhaustiv  
e",intercept = TRUE)
```

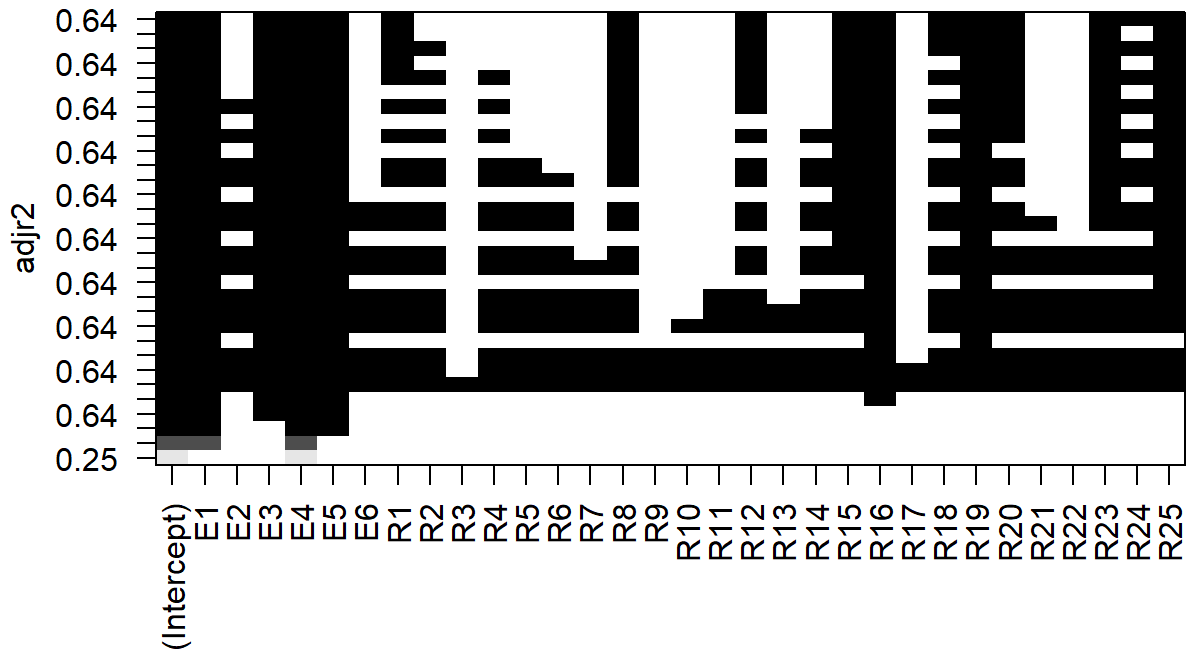
```
final_result_int <- summary(leaps_int)
```

```
#No interactions plot
```

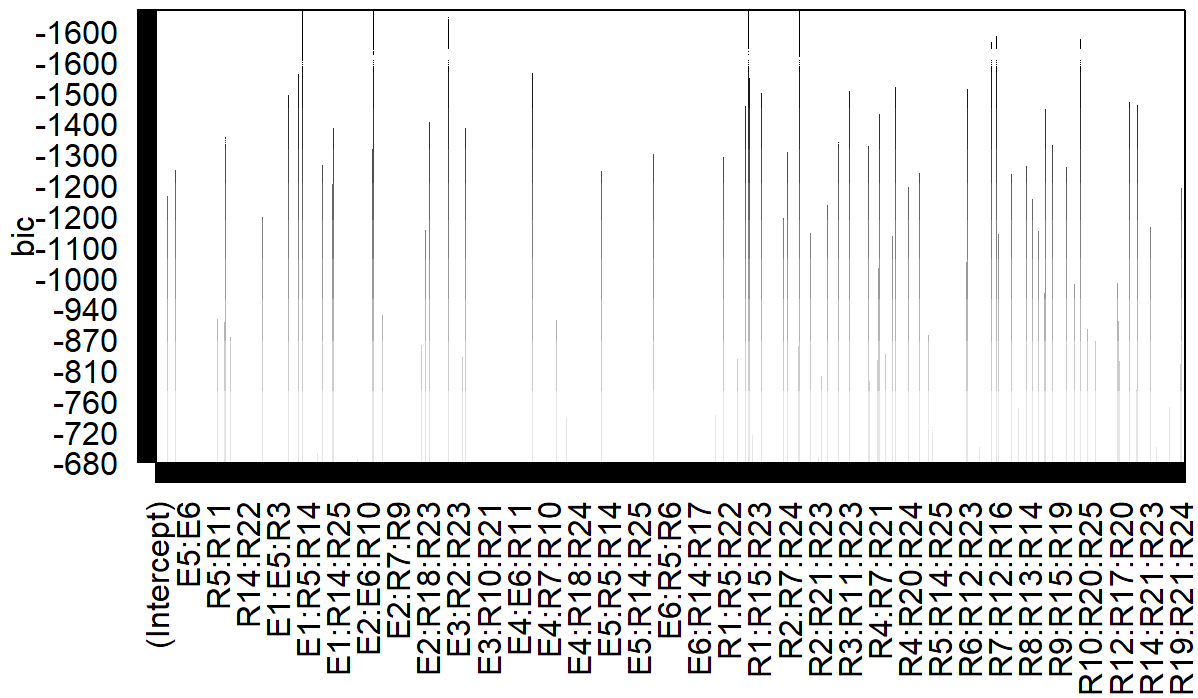
```
plot(leaps_noint,scale="bic")
```



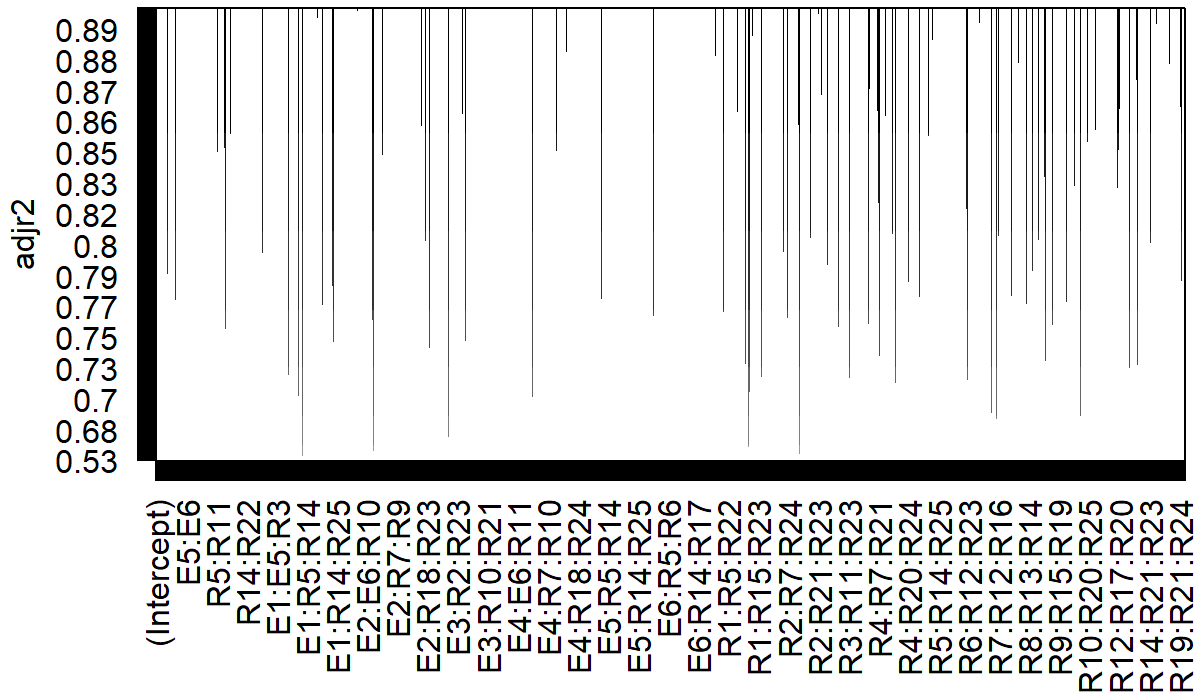
```
plot(leaps_noint,scale="adjr2")
```



```
plot(leaps_int,scale="bic")
```



```
plot(leaps_int,scale="adjr2")
```



```
final_result_int$rsq[(1:20)]
```

```
## [1] 0.5266572 0.6169806 0.6196188 0.6220374 0.6238594 0.6255790 0.6272967
## [8] 0.6289869 0.6310652 0.6324978 0.6341217 0.6359305 0.6375963 0.6393270
## [15] 0.6408752 0.6424717 0.6440483 0.6601121 0.6618080 0.6635097
```

```
final_result_int$bic[1:20]
```

```
## [1] -1212.556 -1552.609 -1556.548 -1559.612 -1560.139 -1560.255 -1560.397
## [8] -1560.453 -1562.268 -1561.250 -1561.114 -1561.844 -1561.966 -1562.419
## [15] -1562.075 -1561.983 -1561.833 -1630.210 -1631.015 -1631.890
```

```
colnames(final_result_int$which)[final_result_int$which[5,]]
```

```
## [1] "(Intercept)" "E1:E4:E5"      "E3:E4:E5"      "R3:R18:R19"    "R4:R7:R16"
## [6] "R16:R23:R25"
```

```
#look at models with 10 or less terms
```

```
var_chose <- colnames(final_result_int$which)[final_result_int$which[10,]]
formula_select <- paste0('log(Y) ~ ', paste(var_chose[-1], collapse = '+') )
M <- lm(formula_select, data=data.cart)
summary(M)
```

```
##
## Call:
## lm(formula = formula_select, data = data.cart)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.118560	-0.018963	0.001494	0.021123	0.081908

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.849e+00	2.775e-03	2827.929	< 2e-16 ***
E1:E4:E5	1.303e-10	4.410e-12	29.552	< 2e-16 ***
E1:R4:R12	1.021e-05	3.024e-06	3.376	0.000753 ***
E4:E5:E3	6.757e-11	3.425e-12	19.728	< 2e-16 ***
E3:R1:R5	7.457e-06	2.397e-06	3.111	0.001899 **
R4:E6:R19	-9.020e-06	3.578e-06	-2.521	0.011807 *
R1:R5:R15	-1.309e-02	3.052e-03	-4.289	1.9e-05 ***
R15:R2:R11	6.941e-03	2.291e-03	3.030	0.002486 **
R19:R3:R18	-6.628e-03	2.329e-03	-2.846	0.004485 **
R4:R7:R16	-6.070e-03	2.302e-03	-2.637	0.008442 **
R16:R23:R25	-7.225e-03	2.272e-03	-3.180	0.001501 **

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02988 on 1630 degrees of freedom
## Multiple R-squared:  0.6325, Adjusted R-squared:  0.6302
## F-statistic: 280.5 on 10 and 1630 DF,  p-value: < 2.2e-16
```

```
anova(M)
```

```
## Analysis of Variance Table
##
## Response: log(Y)
##
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
E1:E4:E5	1	2.08500	2.08500	2335.9073	< 2.2e-16	***
E1:R4:R12	1	0.00094	0.00094	1.0487	0.3059542	
E4:E5:E3	1	0.36115	0.36115	404.6090	< 2.2e-16	***
E3:R1:R5	1	0.00003	0.00003	0.0355	0.8504797	
R4:E6:R19	1	0.01303	0.01303	14.5965	0.0001382	***
R1:R5:R15	1	0.01244	0.01244	13.9331	0.0001959	***
R15:R2:R11	1	0.00666	0.00666	7.4629	0.0063664	**
R19:R3:R18	1	0.00688	0.00688	7.7081	0.0055603	**
R4:R7:R16	1	0.00887	0.00887	9.9342	0.0016519	**
R16:R23:R25	1	0.00903	0.00903	10.1117	0.0015010	**
Residuals	1630	1.45492	0.00089			

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Keep terms with F > 16
```

```
#Hierarchical model
```

```
M2 = lm(log(Y) ~ 1 + E1+ E3 + E4 + E5 + E1:E3 + E4:E5 + R12 +R11 + R22 + R12:R11:R22
+ R11:R22 + R22:R12 + R12:R11 ,data=data.cart)
summary(M2)
```



```
##
## Call:
## lm(formula = log(Y) ~ 1 + E1 + E3 + E4 + E5 + E1:E3 + E4:E5 +
##       R12 + R11 + R22 + R12:R11:R22 + R11:R22 + R22:R12 + R12:R11,
##       data = data.cart)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.118122 -0.019089  0.001645  0.020021  0.084881
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.712e+00  3.318e-02 232.468  < 2e-16 ***
## E1           7.616e-05  2.821e-05   2.700 0.007014 **
## E3           2.918e-05  1.738e-05   1.678 0.093455 .
## E4           3.323e-05  3.313e-05   1.003 0.316024
## E5           3.061e-05  2.672e-05   1.146 0.252090
## R12          3.521e-03  3.004e-03   1.172 0.241349
## R11         -1.021e-03  2.849e-03  -0.358 0.720066
## R22         -9.802e-04  2.933e-03  -0.334 0.738262
## E1:E3        5.607e-08  2.996e-08   1.871 0.061458 .
## E4:E5        1.083e-07  3.039e-08   3.564 0.000376 ***
## R11:R22      1.938e-03  4.087e-03   0.474 0.635431
## R12:R22     -3.300e-03  4.156e-03  -0.794 0.427276
## R12:R11     -4.578e-03  4.218e-03  -1.085 0.277950
## R12:R11:R22  9.796e-03  5.861e-03   1.671 0.094833 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02953 on 1627 degrees of freedom
## Multiple R-squared:  0.6415, Adjusted R-squared:  0.6386
## F-statistic: 224 on 13 and 1627 DF,  p-value: < 2.2e-16
```

```
anova(M2)
```

```
## Analysis of Variance Table
##
## Response: log(Y)
##           Df Sum Sq Mean Sq    F value    Pr(>F)
## E1           1  0.75453  0.75453   864.9780 < 2.2e-16 ***
## E3           1  0.17928  0.17928   205.5243 < 2.2e-16 ***
## E4           1  0.99650  0.99650  1142.3654 < 2.2e-16 ***
## E5           1  0.58616  0.58616   671.9587 < 2.2e-16 ***
## R12          1  0.00160  0.00160     1.8370  0.175496
## R11          1  0.00001  0.00001     0.0168  0.896937
## R22          1  0.00021  0.00021     0.2437  0.621604
## E1:E3        1  0.00320  0.00320     3.6687  0.055617 .
## E4:E5        1  0.01089  0.01089    12.4838  0.000422 ***
## R11:R22      1  0.00458  0.00458     5.2448  0.022140 *
## R12:R22      1  0.00027  0.00027     0.3074  0.579332
## R12:R11      1  0.00003  0.00003     0.0287  0.865400
## R12:R11:R22  1  0.00244  0.00244     2.7936  0.094833 .
## Residuals   1627  1.41925  0.00087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

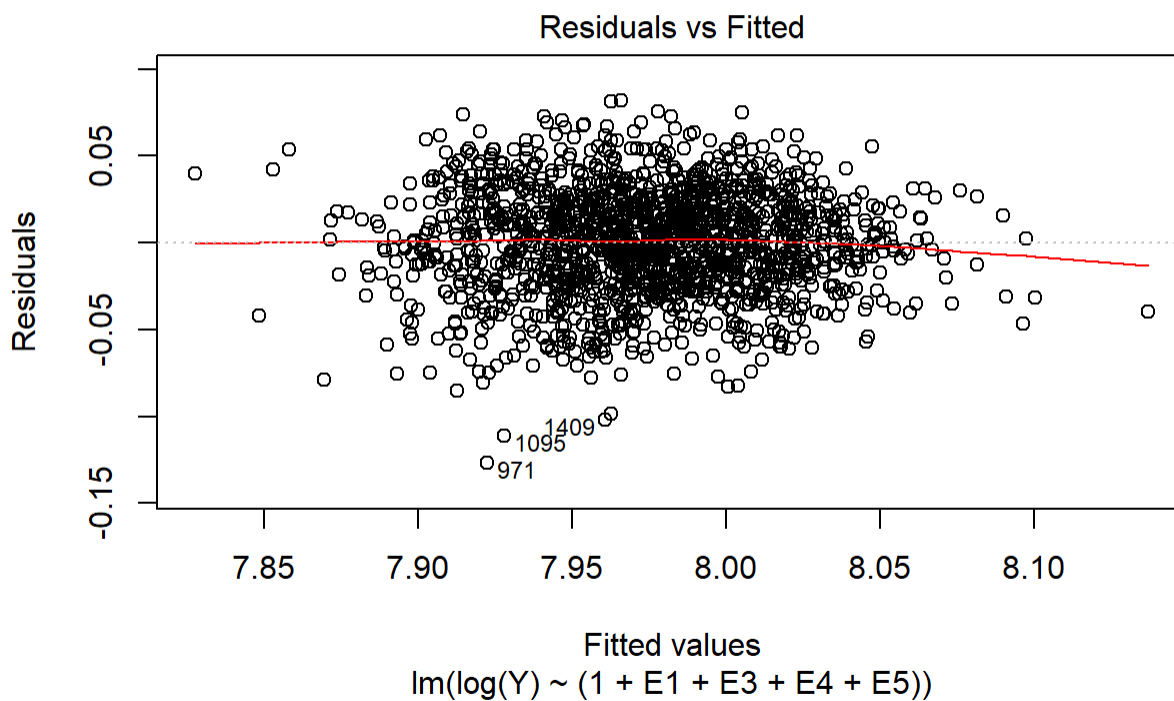
```
M3 = lm(log(Y) ~ (1 + E1 + E3 + E4 + E5),data=data.cart)
summary(M3)
```

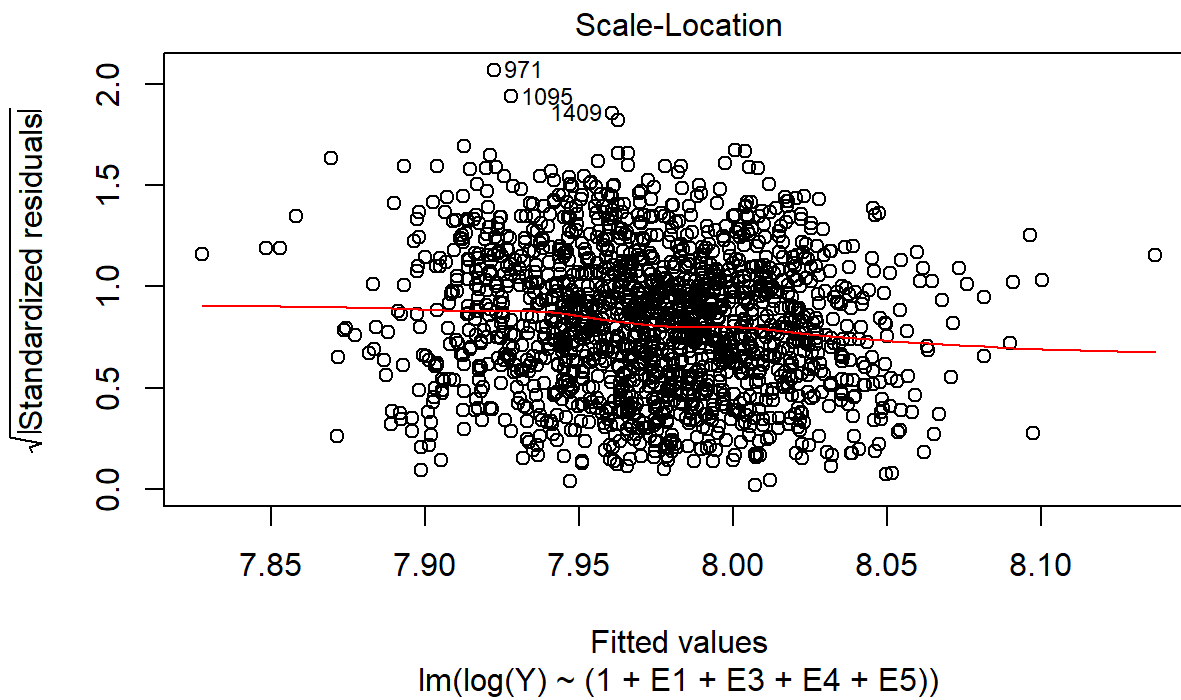
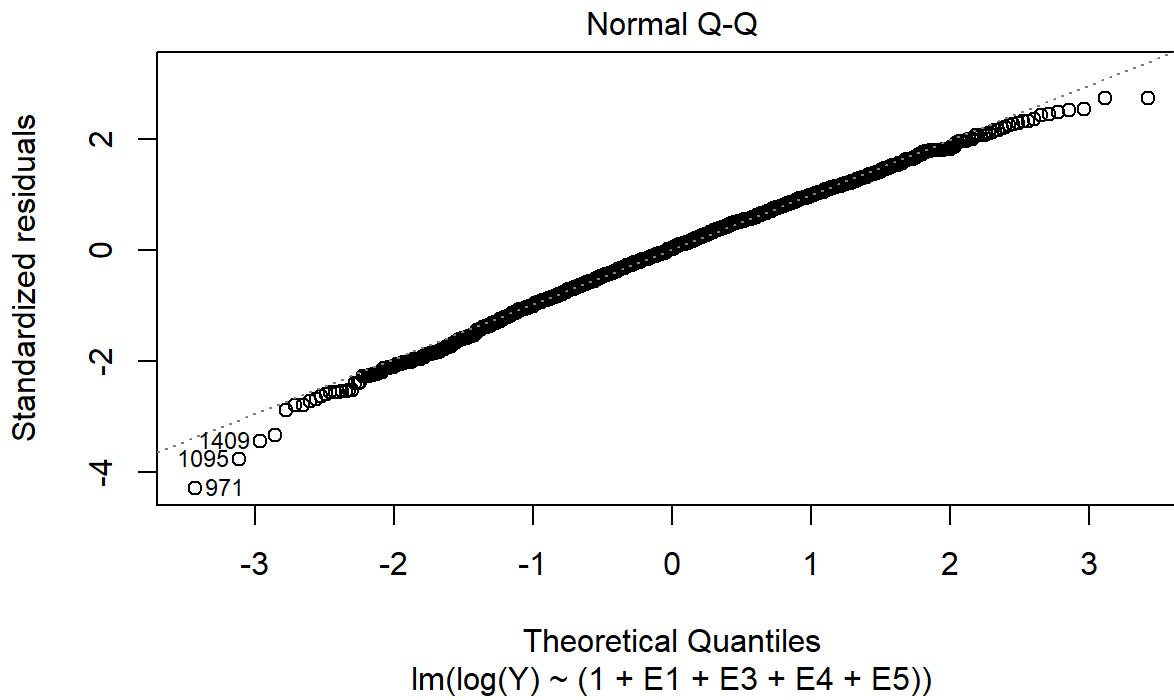
```
##
## Call:
## lm(formula = log(Y) ~ (1 + E1 + E3 + E4 + E5), data = data.cart)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.126830 -0.019227  0.000952  0.020264  0.081616
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.582e+00  8.040e-03  943.13  <2e-16 ***
## E1           1.279e-04  4.552e-06   28.10  <2e-16 ***
## E3           6.037e-05  4.741e-06   12.73  <2e-16 ***
## E4           1.502e-04  4.494e-06   33.43  <2e-16 ***
## E5           1.247e-04  4.835e-06   25.78  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02969 on 1636 degrees of freedom
## Multiple R-squared:  0.6356, Adjusted R-squared:  0.6348
## F-statistic: 713.5 on 4 and 1636 DF, p-value: < 2.2e-16
```

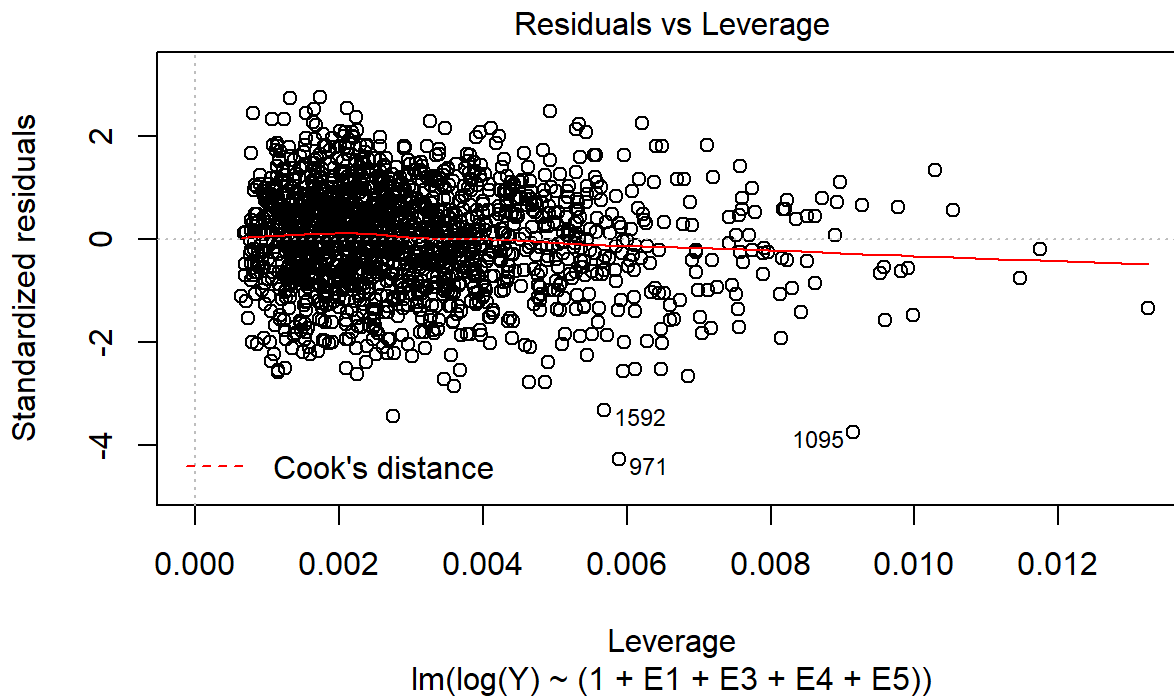
```
anova(M3)
```

```
## Analysis of Variance Table
##
## Response: log(Y)
##           Df Sum Sq Mean Sq F value    Pr(>F)
## E1           1  0.75453  0.75453   855.76 < 2.2e-16 ***
## E3           1  0.17928  0.17928   203.33 < 2.2e-16 ***
## E4           1  0.99650  0.99650  1130.19 < 2.2e-16 ***
## E5           1  0.58616  0.58616   664.80 < 2.2e-16 ***
## Residuals 1636  1.44247  0.00088
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(M3)
```







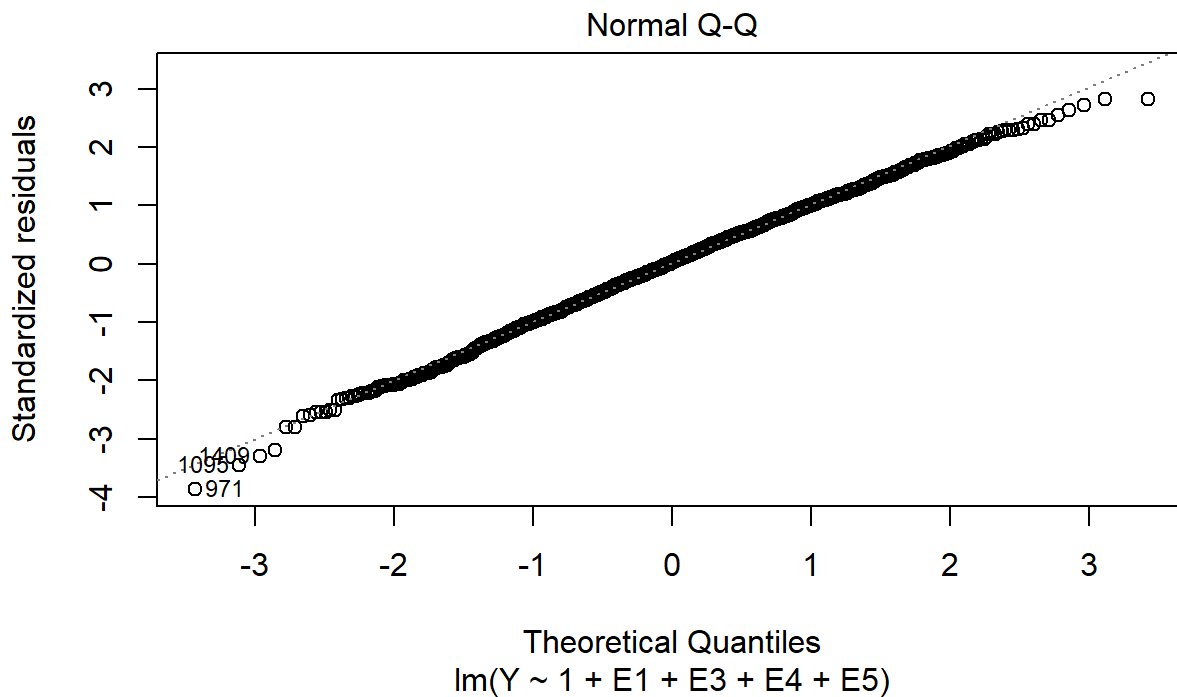
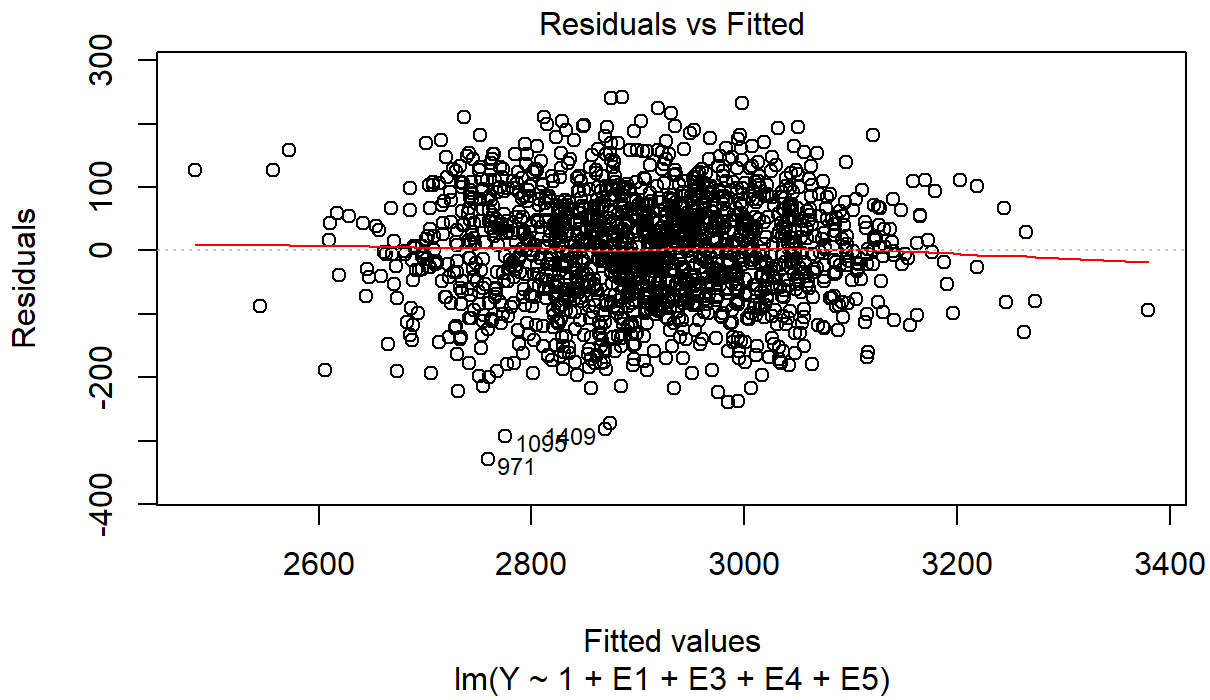
```
#non log model  
M3 = lm(Y ~ 1 + E1 + E3 + E4 + E5 ,data=data.cart)  
summary(M3)
```

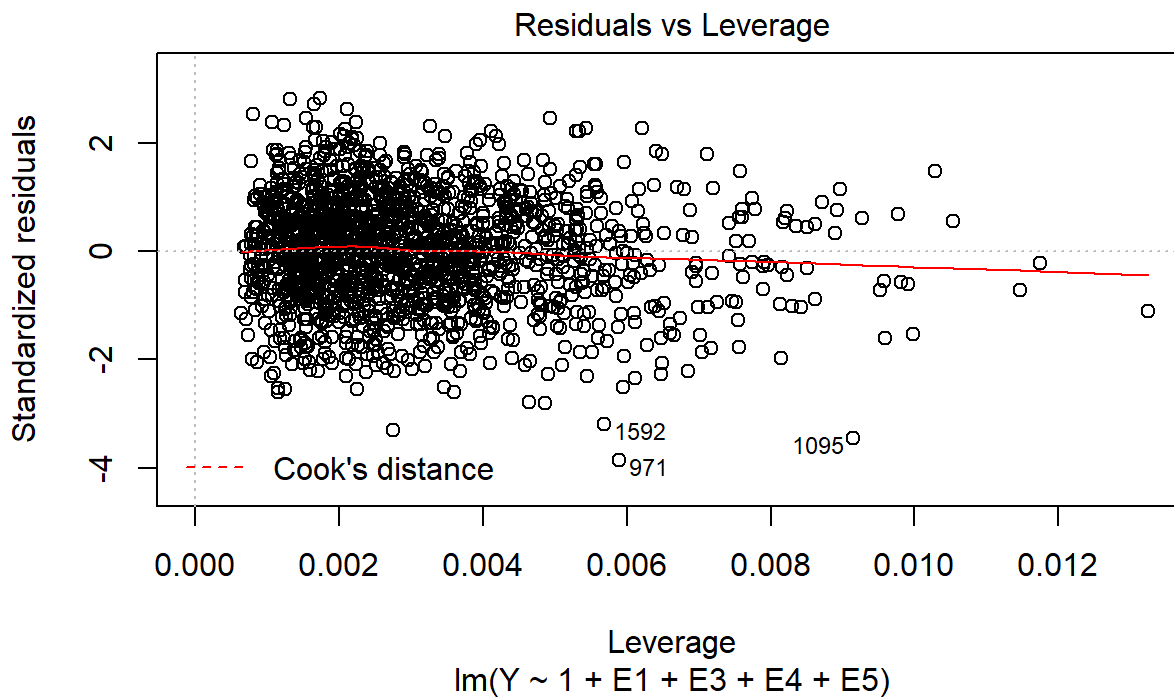
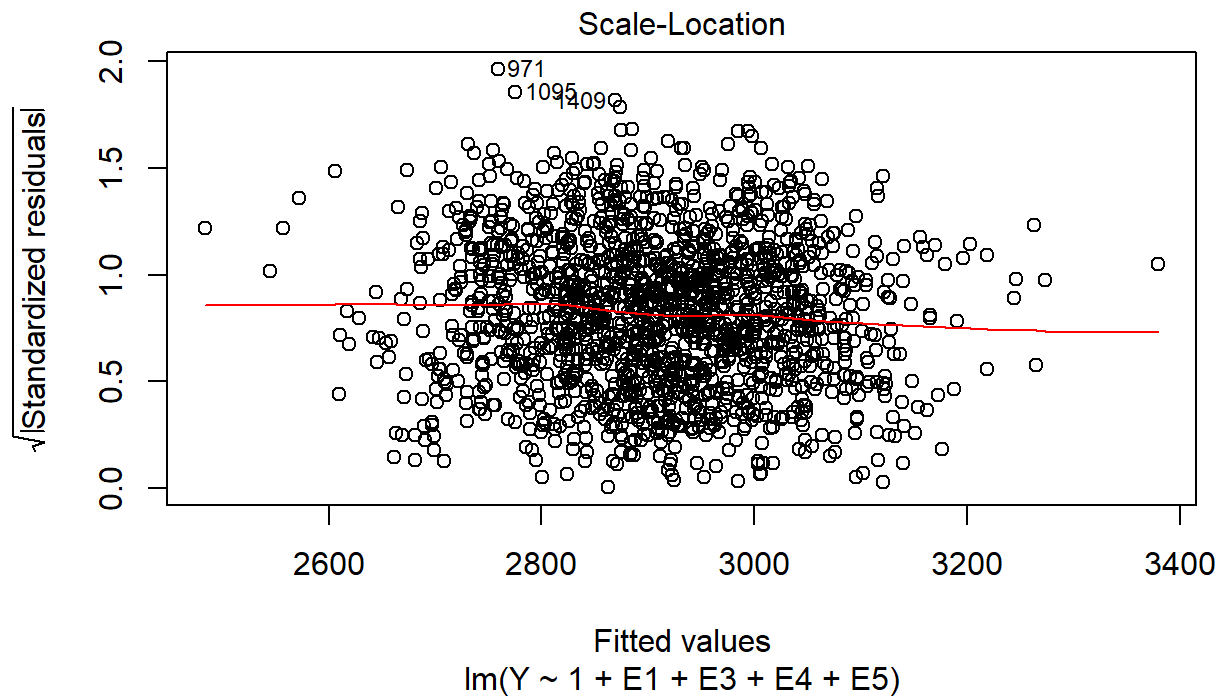
```
##
## Call:
## lm(formula = Y ~ 1 + E1 + E3 + E4 + E5, data = data.cart)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -329.01  -56.90    2.36   58.68  241.15
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.772e+03  2.313e+01   76.63  <2e-16 ***
## E1           3.689e-01  1.310e-02   28.17  <2e-16 ***
## E3           1.751e-01  1.364e-02   12.83  <2e-16 ***
## E4           4.360e-01  1.293e-02   33.72  <2e-16 ***
## E5           3.625e-01  1.391e-02   26.06  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 85.43 on 1636 degrees of freedom
## Multiple R-squared:  0.6391, Adjusted R-squared:  0.6382
## F-statistic: 724.1 on 4 and 1636 DF, p-value: < 2.2e-16
```

```
anova(M3)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df    Sum Sq Mean Sq F value    Pr(>F)
## E1           1  6279122  6279122   860.43 < 2.2e-16 ***
## E3           1  1508111  1508111   206.66 < 2.2e-16 ***
## E4           1  8395795  8395795  1150.48 < 2.2e-16 ***
## E5           1  4955164  4955164   679.01 < 2.2e-16 ***
## Residuals 1636 11938966    7298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(M3)
```

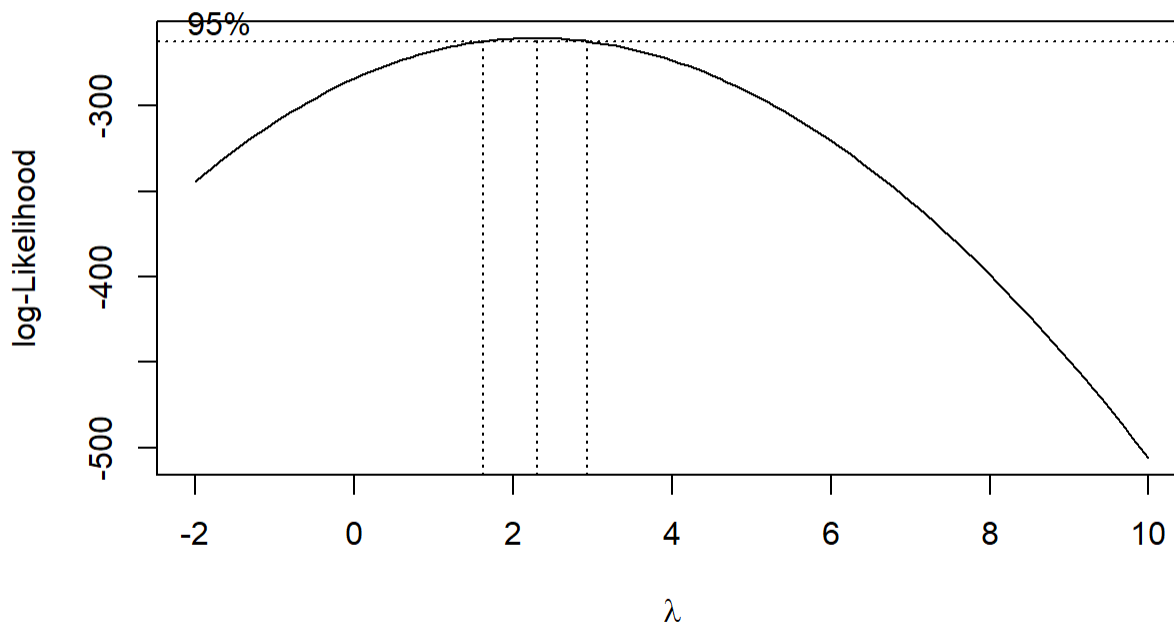






```
#

#box cox
bc <- boxcox(Y ~., data=test, lambda = seq(-2, 10, 1/10))
```



```
(lambda <- bc$x[which.max(bc$y)])
```

```
## [1] 2.3
```

```
new_model <- lm((Y^lambda-1)/lambda) ~(.)^3,data=test)
leaps_bc = regsubsets(model.matrix(new_model)[,-1], I((data.cart$Y^lambda-1)/lambda),
really.big = TRUE,
                      method = "forward", nbest = 1, intercept = TRUE, nvmax = 500)
```

```
## Warning in leaps.setup(x, y, wt = weights, nbest = nbest, nvmax = nvmax, : 3351
## linear dependencies found
```

```

final_result_bc <- summary(leaps_bc)

var_chose <- colnames(final_result_bc$which)[final_result_bc$which[10,]]
formula_select <- paste0('(Y^2.2-1)/2.2 ~ ', paste(var_chose[-1], collapse = '+') )
M <- lm(formula_select, data=data.cart)
summary(M)

```

```

##
## Call:
## lm(formula = formula_select, data = data.cart)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3767728  -776384    1405    832905   3705470
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.375e+07  1.121e+05 122.674 < 2e-16 ***
## E1:E4:E5      5.446e-03  1.778e-04  30.635 < 2e-16 ***
## E1:R4:R12     4.656e+02  1.225e+02   3.799 0.000150 ***
## E4:E5:E3      2.864e-03  1.375e-04  20.829 < 2e-16 ***
## R4:E6:R19    -4.632e+02  1.479e+02  -3.131 0.001772 **
## R1:R5:R15    -3.275e+05  9.576e+04  -3.420 0.000641 ***
## R19:R1:R17    3.072e+05  9.429e+04   3.258 0.001143 **
## R15:R2:R11    2.682e+05  9.224e+04   2.907 0.003696 **
## R19:R3:R18   -2.905e+05  9.480e+04  -3.065 0.002214 **
## R4:R7:R16    -2.944e+05  9.203e+04  -3.199 0.001406 **
## R15:R23:R25  -2.851e+05  9.126e+04  -3.124 0.001813 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1208000 on 1630 degrees of freedom
## Multiple R-squared:  0.6481, Adjusted R-squared:  0.646
## F-statistic: 300.3 on 10 and 1630 DF,  p-value: < 2.2e-16

```

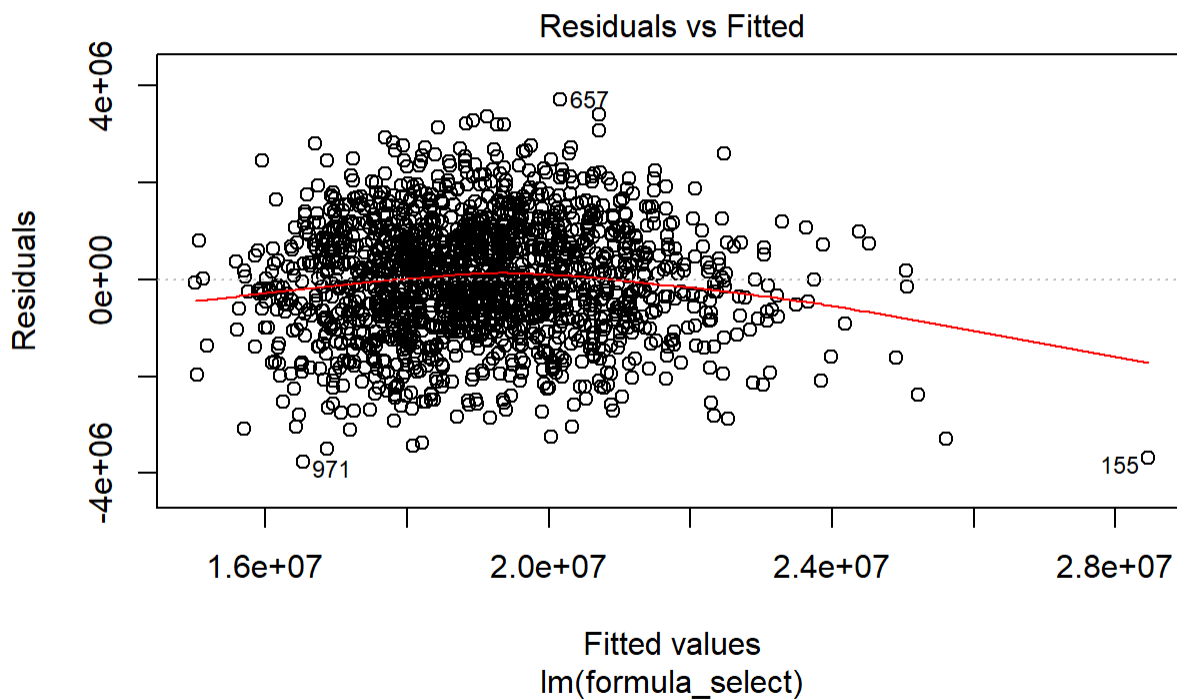
```
anova(M)
```

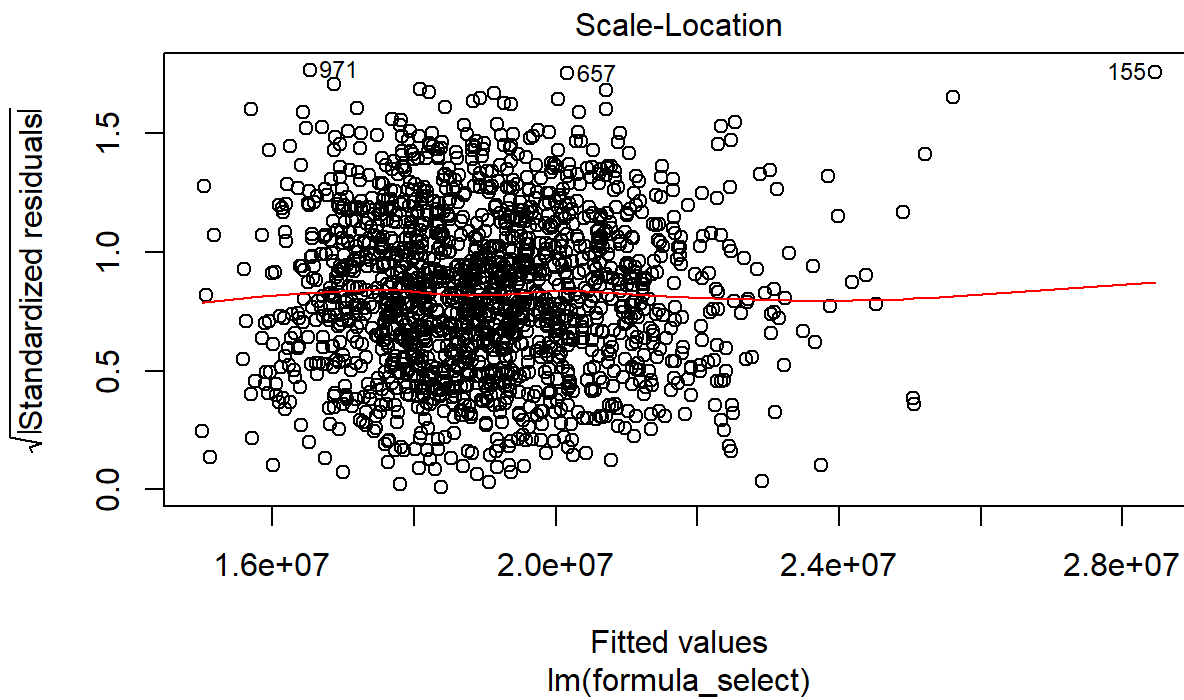
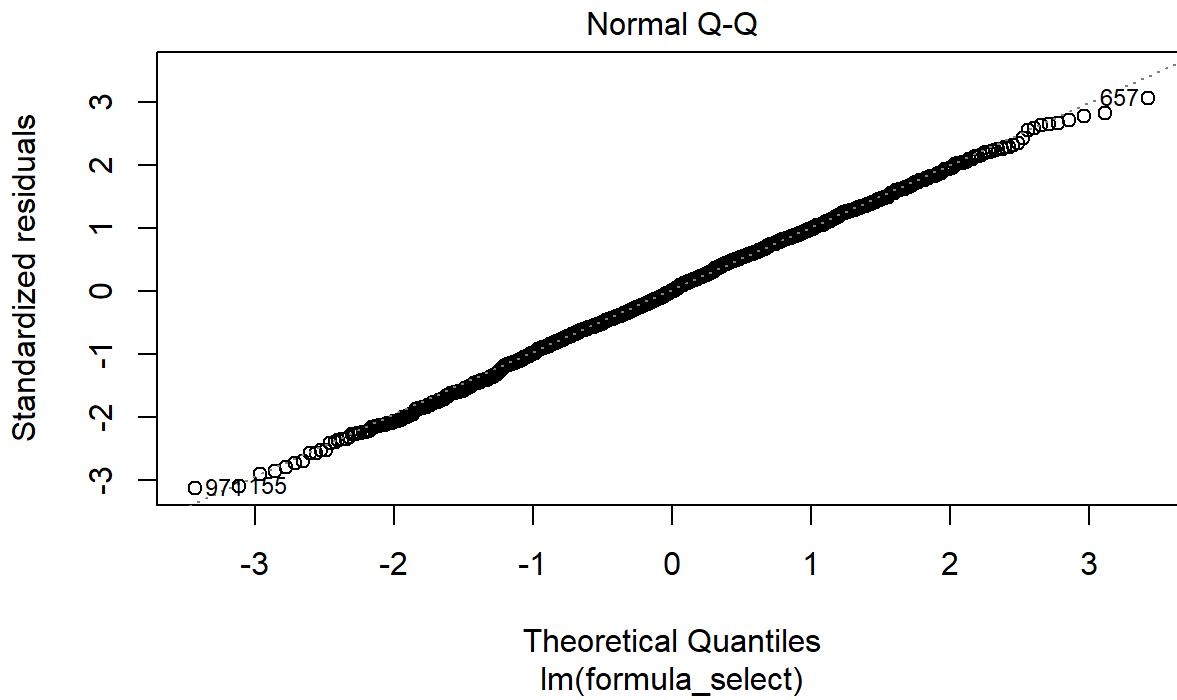
```
## Analysis of Variance Table
##
## Response: (Y^2.2 - 1)/2.2
##
```

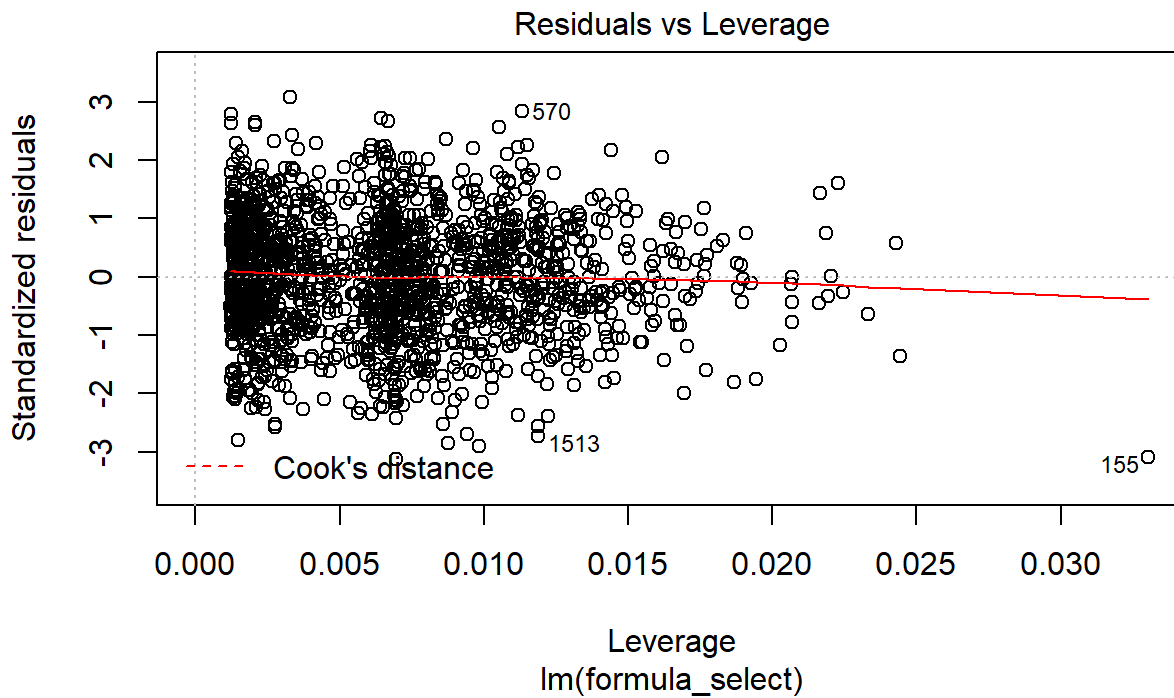
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
E1:E4:E5	1	3.6450e+15	3.6450e+15	2498.0851	< 2.2e-16	***
E1:R4:R12	1	1.2240e+12	1.2240e+12	0.8388	0.3598668	
E4:E5:E3	1	6.3519e+14	6.3519e+14	435.3211	< 2.2e-16	***
R4:E6:R19	1	2.2194e+13	2.2194e+13	15.2102	0.0001001	***
R1:R5:R15	1	1.2742e+13	1.2742e+13	8.7327	0.0031704	**
R19:R1:R17	1	1.4362e+13	1.4362e+13	9.8428	0.0017355	**
R15:R2:R11	1	8.8671e+12	8.8671e+12	6.0770	0.0137975	*
R19:R3:R18	1	1.2556e+13	1.2556e+13	8.6052	0.0033990	**
R4:R7:R16	1	1.4681e+13	1.4681e+13	10.0614	0.0015423	**
R15:R23:R25	1	1.4244e+13	1.4244e+13	9.7622	0.0018128	**
Residuals	1630	2.3784e+15	1.4591e+12			

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#comparable to the analysis before
plot(M)
```







```
#merging of the data
fit <- with(cart.impute, lm(log(Y) ~ E1 + E3 + E4 + E5 + 1))
combine <- pool(fit)
summary(combine)
```

##	term	estimate	std.error	statistic	df	p.value
## 1	(Intercept)	7.579729e+00	8.257453e-03	917.92580	990.8617	0
## 2	E1	1.287289e-04	4.607526e-06	27.93884	1446.5014	0
## 3	E3	6.209577e-05	4.885846e-06	12.70932	837.1574	0
## 4	E4	1.505683e-04	4.591086e-06	32.79579	989.5125	0
## 5	E5	1.249340e-04	4.836470e-06	25.83165	1628.6876	0

```
#aov(combine)

#Here are some of the other methods i had tried
#
#null = lm(Y~1,data=data.cart)
#null
#full = lm(Y~.,data=data.cart)
#full
#step(null,scope=list(lower=null,upper=full),direction="forward")

#leaps
#summary.out <- summary(leaps)
#as.data.frame(summary.out$outmat)

#library(car)
#subsets(leaps,statistic="bic",max.size = 10,)

#library(MASS)
#fit.test = lm(Y~., data= data.cart)
#step <- stepAIC(fit.test,scope =list(upper= . ~ .^2, lower= ~1),direction="both")
#step$anova

#library(rFSA)
#fsa.fit = FSA(Y~.,data=data.cart,fitfunc = lm,m=3,numrs = 50,criterion = BIC)
#print(fsa.fit)

#library(bestglm)
#bestglm(data.cart,IC="BIC",family=binomial)

#library(glmnet)
#f <- as.formula(Y ~ .^3)
#x <- model.matrix(f,data.cart)[,-1]
#x

#glmnet(x,data.cart$Y)
#M_LASS <- glmnet::cv.glmnet(x, data.cart$Y,nfolds = 5, alpha=1, grouped = TRUE)
#coef_select <- as.matrix(coef(M_LASS, s="lambda.1se"))
#cbind(rownames(coef_select)[coef_select > 0], coef_select[coef_select > 0])
#M_LASS$lambda.1se
#coef_select
```