

# PA\_Project\_2

*Chuan Du (Sophie)*

*11/17/2018*

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.4
```

```
##  
## Attaching package: 'dplyr'
```

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

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

```
# Load in all raw data sets  
book <- read.csv("book.csv")  
book.test <- read.csv("booktest.csv")  
book.train <- read.csv("booktrain.csv")  
orders <- read.csv("ordersall.csv", stringsAsFactors = F)
```

```
# Add respond column to indicate if a customer responded to promotion  
#book$respond <- ifelse(book$logtargamt > 0, 1, 0)
```

```
# Drop empty malformat column  
book.train$X <- NULL
```

```
# Orders with orddate as Date and net as total price of order  
orders$orddate <- lubridate::dmy(orders$orddate)
```

```
## Warning in as.POSIXlt.POSIXct(x, tz): unknown timezone 'default/America/  
## Chicago'
```

```
orders$net <- orders$qty * orders$price
```

```
# Orders of customers who respond to promotion
```

```
orders.responders <- book.train %>%  
  filter(logtargamt > 0) %>%  
  merge(orders, by = 'id')
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
# All information about the customers in training set
```

```
book.train.merge <- merge(book.train, book, by = 'id')  
book.train.merge$logtargamt.y <- NULL  
book.train.merge <- rename(book.train.merge, logtargamt = logtargamt.x)  
book.train.merge$respond <- ifelse(book.train.merge$logtargamt > 0, 1, 0)
```

```
# All information about customers in training set with addition of avgNetOrder
```

```
book.train.full <- orders %>%  
  group_by(id) %>%  
  #could try sum(net) to see difference  
  summarise(avgNetOrder = mean(net)) %>%  
  merge(book.train.merge, by = 'id')
```

```
head(book.train.full, 25)
```

```
##      id avgNetOrder logtargamt recency frequency  amount  tof  
## 1  2062    6.699425  0.000000    19      16 1078.60742 2117  
## 2  2232   16.335338  0.000000   142     40 1453.84473 2442  
## 3  2623    9.610636  0.000000   345      5  144.15955 2378  
## 4  3000   11.756627  0.000000   380     13  258.64575 2460  
## 5  4693    6.182206  0.000000   283      9  117.46191 2379  
## 6  5010    7.238901  0.000000  1152     15  224.40588 2239  
## 7  5533    9.404574  0.000000  1620      5   75.23657 2454  
## 8  6130    9.662655  0.000000   272      8  212.57837 2014  
## 9  6653   10.112772  0.000000   135     35  707.89355 2390  
## 10 6831    5.530025  0.000000   326      4  132.72058 1025  
## 11 7161   11.909453  0.000000    58     24  428.74023 2456  
## 12 7315   12.054311  0.000000  1244     13  253.14050 2348  
## 13 7897    6.626340  0.000000  1567      3   33.13168 2379  
## 14 9270    7.065624  0.000000   846      6  120.11560 2119  
## 15 9539   10.174706  0.000000  2218      1   10.17471 2218  
## 16 11282   9.314032  3.362108  1418      2   27.94209 1803  
## 17 12955   10.487708  0.000000     6     26  388.04517 2428  
## 18 14010    8.643997  0.000000    46      4   86.43994  270  
## 19 14508    9.602059  0.000000  1424      5  144.03076 2244  
## 20 15253   10.077100  0.000000     6     28  745.70508 2365  
## 21 16276   10.821150  0.000000   788      2   32.46344 1070  
## 22 17221   13.184934  0.000000  1668      4  105.47943 2295
```

##	23	17507	8.151781	0.000000	153	9	171.18738	2063
##	24	17639	13.155701	0.000000	164	4	197.33545	2371
##	25	20567	10.614178	0.000000	150	8	498.86621	1543
##		Ffiction1	Fclassics3	Fcartoons5	Flegends6	Fphilosophy7	Freligion8	
##	1	3	0	2	1	0	4	
##	2	0	0	0	0	0	0	
##	3	1	0	1	1	0	0	
##	4	0	0	1	0	0	0	
##	5	3	0	1	0	1	1	
##	6	1	0	0	0	0	2	
##	7	0	0	0	0	0	3	
##	8	0	0	1	0	0	0	
##	9	0	0	1	0	0	1	
##	10	1	0	0	0	0	0	
##	11	5	2	3	1	0	0	
##	12	0	0	0	0	0	0	
##	13	0	0	0	0	0	0	
##	14	2	1	2	0	0	2	
##	15	0	0	0	0	0	0	
##	16	0	0	0	0	0	0	
##	17	0	0	0	0	0	0	
##	18	0	0	0	0	0	1	
##	19	0	0	0	0	1	2	
##	20	2	0	1	1	3	9	
##	21	0	0	0	0	0	0	
##	22	0	0	1	0	0	0	
##	23	1	0	0	0	0	0	
##	24	0	0	0	0	0	0	
##	25	0	1	0	0	0	16	
##		Fpsychology9	Flinguistics10	Fart12	Fmusic14	Ffacsimile17	Fhistory19	
##	1	0	0	0	1	0	1	
##	2	0	1	7	12	3	17	
##	3	0	0	0	2	0	0	
##	4	0	1	0	0	0	2	
##	5	0	0	0	0	0	0	
##	6	0	0	0	7	0	2	
##	7	0	0	0	0	0	1	
##	8	0	0	0	2	0	4	
##	9	0	0	2	6	0	9	
##	10	0	0	0	3	0	3	
##	11	0	0	3	6	1	5	
##	12	0	0	0	0	0	0	
##	13	0	0	0	0	0	0	
##	14	0	0	0	0	0	2	
##	15	0	0	0	0	0	0	
##	16	0	0	0	0	0	0	
##	17	1	2	1	0	0	12	
##	18	0	0	0	0	0	0	
##	19	0	0	1	1	0	2	
##	20	0	1	0	0	0	4	

##	21	0	0	0	1	0	0
##	22	0	0	1	0	0	1
##	23	0	0	0	3	0	5
##	24	0	0	1	0	0	3
##	25	0	0	0	1	0	4
##		Fconthist20	Feconomy21	Fpolitics22	Fscience23	Fcompsci26	Frailroads27
##	1	0	0	0	5	3	0
##	2	0	2	0	0	1	0
##	3	3	0	0	0	0	1
##	4	0	0	0	0	5	0
##	5	0	0	0	0	0	0
##	6	5	0	0	0	0	0
##	7	0	0	0	0	0	0
##	8	2	1	0	0	0	0
##	9	19	1	0	5	2	1
##	10	0	0	0	1	0	0
##	11	1	0	0	0	1	2
##	12	17	0	0	0	0	1
##	13	2	0	0	0	0	1
##	14	0	0	0	0	0	0
##	15	0	0	0	0	0	1
##	16	0	0	0	0	0	0
##	17	5	0	0	7	1	1
##	18	6	0	0	0	0	0
##	19	1	0	0	0	0	0
##	20	0	2	0	2	0	0
##	21	0	0	0	0	0	0
##	22	0	0	0	0	0	0
##	23	2	1	0	0	1	0
##	24	0	0	0	0	0	3
##	25	22	0	0	0	0	1
##		Fmaps30	Ftravelguides31	Fhealth35	Fcooking36	Flearning37	
##	1	0	4	95	4	10	
##	2	0	2	12	2	1	
##	3	0	1	0	0	4	
##	4	0	0	1	0	10	
##	5	0	2	5	2	0	
##	6	0	5	4	1	0	
##	7	0	0	0	0	0	
##	8	0	1	3	2	0	
##	9	1	1	10	1	1	
##	10	0	2	0	7	2	
##	11	0	2	0	1	0	
##	12	1	0	0	0	0	
##	13	0	0	0	1	0	
##	14	0	0	1	0	0	
##	15	0	0	0	0	0	
##	16	0	0	1	0	0	
##	17	0	5	1	0	0	
##	18	0	0	1	0	2	

##	19	0	1	0	1	0	
##	20	0	1	44	0	0	
##	21	0	0	2	0	0	
##	22	0	1	1	1	0	
##	23	0	2	1	2	1	
##	24	0	5	0	0	0	
##	25	0	1	0	0	1	
##		FGamesRiddles38	Fsports39	Fhobby40	Fnature41	Fencyclopedia44	Fvideos50
##	1	0	0	10	14	3	0
##	2	0	1	16	8	4	0
##	3	0	0	0	0	1	0
##	4	0	0	0	0	2	0
##	5	1	0	2	1	0	0
##	6	0	0	3	1	0	0
##	7	0	0	0	0	0	0
##	8	0	0	4	1	0	1
##	9	0	0	4	4	1	0
##	10	0	0	2	2	1	0
##	11	0	0	1	0	0	2
##	12	0	0	0	0	0	2
##	13	0	0	0	0	0	0
##	14	0	0	3	1	2	0
##	15	0	0	0	0	0	0
##	16	0	0	0	0	0	2
##	17	0	0	1	0	0	0
##	18	0	0	0	0	0	0
##	19	0	0	2	1	2	0
##	20	0	0	3	0	1	0
##	21	0	0	0	0	0	0
##	22	0	0	0	1	1	0
##	23	0	0	2	0	0	0
##	24	0	0	2	1	0	0
##	25	0	0	0	0	0	0
##		Fnonbooks99	Mfiction1	Mclassics3	Mcartoons5	Mlegends6	Mphilosophy7
##	1	1	13.109299	0.00000	10.000000	9.949997	0.00000
##	2	0	0.000000	0.00000	0.000000	0.000000	0.00000
##	3	0	10.123573	0.00000	0.000000	15.236496	0.00000
##	4	0	0.000000	0.00000	17.383911	0.000000	0.00000
##	5	0	9.449997	0.00000	7.158085	0.000000	0.00000
##	6	0	6.902439	0.00000	0.000000	0.000000	0.00000
##	7	4	0.000000	0.00000	0.000000	0.000000	0.00000
##	8	0	0.000000	0.00000	10.123573	0.000000	0.00000
##	9	0	0.000000	0.00000	7.158085	0.000000	0.00000
##	10	0	3.949999	0.00000	0.000000	0.000000	0.00000
##	11	0	43.296265	29.63004	20.349991	15.000000	0.00000
##	12	0	0.000000	0.00000	0.000000	0.000000	0.00000
##	13	1	0.000000	0.00000	0.000000	0.000000	0.00000
##	14	1	14.073570	7.64381	13.666389	0.000000	0.00000
##	15	0	0.000000	0.00000	0.000000	0.000000	0.00000
##	16	0	0.000000	0.00000	0.000000	0.000000	0.00000

##	17	0	0.000000	0.000000	0.000000	0.000000	0.000000
##	18	0	0.000000	0.000000	0.000000	0.000000	0.000000
##	19	0	0.000000	0.000000	0.000000	0.000000	7.64381
##	20	0	19.736496	0.000000	0.000000	8.500000	36.81299
##	21	0	0.000000	0.000000	0.000000	0.000000	0.000000
##	22	0	0.000000	0.000000	12.680038	0.000000	0.000000
##	23	0	7.669376	0.000000	0.000000	0.000000	0.000000
##	24	0	0.000000	0.000000	0.000000	0.000000	0.000000
##	25	0	0.000000	10.20027	0.000000	0.000000	0.000000
##		Mreligion8	Mpsychology9	Mlinguistics10	Mart12	Mmusic14	
##	1	31.981293	0.000000	0.000000	0.000000	0.000000	
##	2	0.000000	0.000000	8.180664	89.389893	308.375977	
##	3	0.000000	0.000000	0.000000	0.000000	13.099998	
##	4	0.000000	0.000000	7.669376	0.000000	0.000000	
##	5	6.949997	0.000000	0.000000	0.000000	0.000000	
##	6	15.210922	0.000000	0.000000	0.000000	44.738007	
##	7	33.080566	0.000000	0.000000	0.000000	0.000000	
##	8	0.000000	0.000000	0.000000	0.000000	5.949997	
##	9	3.949999	0.000000	0.000000	21.494522	42.729279	
##	10	0.000000	0.000000	0.000000	0.000000	16.899994	
##	11	0.000000	0.000000	0.000000	71.367859	65.174927	
##	12	0.000000	0.000000	0.000000	0.000000	0.000000	
##	13	0.000000	0.000000	0.000000	0.000000	0.000000	
##	14	14.899994	0.000000	0.000000	0.000000	0.000000	
##	15	0.000000	0.000000	0.000000	0.000000	0.000000	
##	16	0.000000	0.000000	0.000000	0.000000	0.000000	
##	17	0.000000	9.899994	15.124702	6.033241	0.000000	
##	18	4.949997	0.000000	0.000000	0.000000	0.000000	
##	19	17.690689	0.000000	0.000000	10.225838	15.236496	
##	20	60.373993	0.000000	9.899994	0.000000	0.000000	
##	21	0.000000	0.000000	0.000000	0.000000	5.087353	
##	22	0.000000	0.000000	0.000000	25.539017	0.000000	
##	23	0.000000	0.000000	0.000000	0.000000	27.699982	
##	24	0.000000	0.000000	0.000000	20.349411	0.000000	
##	25	190.951538	0.000000	0.000000	0.000000	4.949997	
##		Mfacsimile17	Mhistory19	Mconthist20	Meconomy21	Mpolitics22	Mscience23
##	1	0.000000	8.589699	0.000000	0.000000	0	30.639954
##	2	105.33563	381.341553	0.000000	15.26205	0	0.000000
##	3	0.000000	0.000000	44.749969	0.000000	0	0.000000
##	4	0.000000	17.869644	0.000000	0.000000	0	0.000000
##	5	0.000000	0.000000	0.000000	0.000000	0	0.000000
##	6	0.000000	12.015358	48.905029	0.000000	0	0.000000
##	7	0.000000	10.200272	0.000000	0.000000	0	0.000000
##	8	0.000000	37.569275	19.899994	10.12357	0	0.000000
##	9	0.000000	101.644775	250.679932	10.12357	0	59.565460
##	10	0.000000	21.899994	0.000000	0.000000	0	4.949997
##	11	49.94998	50.749969	8.666389	0.000000	0	0.000000
##	12	0.000000	0.000000	204.746704	0.000000	0	0.000000
##	13	0.000000	0.000000	17.844070	0.000000	0	0.000000
##	14	0.000000	25.073563	0.000000	0.000000	0	0.000000

##	15	0.000000	0.000000	0.000000	0.000000	0	0.000000
##	16	0.000000	0.000000	0.000000	0.000000	0	0.000000
##	17	0.000000	116.527588	60.110901	0.000000	0	73.861450
##	18	0.000000	0.000000	50.639954	0.000000	0	0.000000
##	19	0.000000	12.756729	10.123573	0.000000	0	0.000000
##	20	0.000000	38.463440	0.000000	14.54082	0	16.898453
##	21	0.000000	0.000000	0.000000	0.000000	0	0.000000
##	22	0.000000	6.595665	0.000000	0.000000	0	0.000000
##	23	0.000000	43.015045	8.899994	10.12357	0	0.000000
##	24	0.000000	65.828796	0.000000	0.000000	0	0.000000
##	25	0.000000	40.889984	234.837036	0.000000	0	0.000000
##		Mcompsci26	Mrailroads27	Mmaps30	Mtravelguides31	Mhealth35	Mcooking36
##	1	19.183502	0.000000	0.000000	35.074615	678.758301	16.568237
##	2	5.112919	0.000000	0.000000	39.899994	121.057129	12.674706
##	3	0.000000	11.452934	0.000000	8.899994	0.000000	0.000000
##	4	48.163666	0.000000	0.000000	0.000000	18.866669	0.000000
##	5	0.000000	0.000000	0.000000	13.702614	36.945923	13.779312
##	6	0.000000	0.000000	0.000000	22.931427	26.305954	6.621227
##	7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	8	0.000000	0.000000	0.000000	5.949997	21.618103	19.923569
##	9	17.150269	20.426102	9.949997	5.061787	47.830017	4.949997
##	10	0.000000	0.000000	0.000000	9.449997	0.000000	32.025818
##	11	6.949997	20.247147	0.000000	14.450645	0.000000	9.949997
##	12	0.000000	10.200272	5.087353	0.000000	0.000000	0.000000
##	13	0.000000	7.643810	0.000000	0.000000	0.000000	7.643810
##	14	0.000000	0.000000	0.000000	0.000000	2.500000	0.000000
##	15	0.000000	10.174706	0.000000	0.000000	0.000000	0.000000
##	16	0.000000	0.000000	0.000000	0.000000	5.061787	0.000000
##	17	10.200272	12.680038	0.000000	63.706940	6.949997	0.000000
##	18	0.000000	0.000000	0.000000	0.000000	2.949999	0.000000
##	19	0.000000	0.000000	0.000000	9.203247	0.000000	10.174706
##	20	0.000000	0.000000	0.000000	5.000000	504.852783	0.000000
##	21	0.000000	0.000000	0.000000	0.000000	27.376099	0.000000
##	22	0.000000	0.000000	0.000000	8.666389	20.426102	10.174706
##	23	10.200272	0.000000	0.000000	12.731163	6.544533	15.262054
##	24	0.000000	30.447418	0.000000	50.185730	0.000000	0.000000
##	25	0.000000	5.087353	0.000000	2.949999	0.000000	0.000000
##		Mlearning37	MGamesRiddles38	Msports39	Mhobby40	Mnature41	
##	1	48.994843	0.000000	0.0000	72.573059	79.945618	
##	2	5.949997	0.000000	12.7823	160.018311	87.980957	
##	3	32.978302	0.000000	0.0000	0.000000	0.000000	
##	4	72.101013	0.000000	0.0000	0.000000	0.000000	
##	5	0.000000	7.56712	0.0000	15.287621	6.621227	
##	6	0.000000	0.000000	0.0000	30.575241	10.200272	
##	7	0.000000	0.000000	0.0000	0.000000	0.000000	
##	8	0.000000	0.000000	0.0000	54.270020	16.949997	
##	9	8.640831	0.000000	0.0000	52.586334	38.839966	
##	10	7.899998	0.000000	0.0000	12.949997	16.150269	
##	11	0.000000	0.000000	0.0000	6.135502	0.000000	
##	12	0.000000	0.000000	0.0000	0.000000	0.000000	

##	13	0.000000	0.00000	0.0000	0.000000	0.000000
##	14	0.000000	0.00000	0.0000	25.462326	6.595665
##	15	0.000000	0.00000	0.0000	0.000000	0.000000
##	16	0.000000	0.00000	0.0000	0.000000	0.000000
##	17	0.000000	0.00000	0.0000	12.949997	0.000000
##	18	27.899994	0.00000	0.0000	0.000000	0.000000
##	19	0.000000	0.00000	0.0000	20.400543	15.313187
##	20	0.000000	0.00000	0.0000	25.513458	0.000000
##	21	0.000000	0.00000	0.0000	0.000000	0.000000
##	22	0.000000	0.00000	0.0000	0.000000	12.731163
##	23	7.618248	0.00000	0.0000	21.423111	0.000000
##	24	0.000000	0.00000	0.0000	20.400543	10.123573
##	25	9.000000	0.00000	0.0000	0.000000	0.000000
##		Mencyclopaedia44	Mvideos50	Mnonbooks99	respond	
##	1	23.238739	0.00000	0.00000	0	
##	2	100.482666	0.00000	0.00000	0	
##	3	7.618248	0.00000	0.00000	0	
##	4	76.591492	0.00000	0.00000	0	
##	5	0.000000	0.00000	0.00000	0	
##	6	0.000000	0.00000	0.00000	0	
##	7	0.000000	0.00000	31.95573	0	
##	8	0.000000	10.20027	0.00000	0	
##	9	5.112915	0.00000	0.00000	0	
##	10	6.544533	0.00000	0.00000	0	
##	11	0.000000	16.82149	0.00000	0	
##	12	0.000000	33.10614	0.00000	0	
##	13	0.000000	0.00000	0.00000	0	
##	14	10.200264	0.00000	0.00000	0	
##	15	0.000000	0.00000	0.00000	0	
##	16	0.000000	22.88031	0.00000	1	
##	17	0.000000	0.00000	0.00000	0	
##	18	0.000000	0.00000	0.00000	0	
##	19	15.262054	0.00000	0.00000	0	
##	20	5.112915	0.00000	0.00000	0	
##	21	0.000000	0.00000	0.00000	0	
##	22	8.666389	0.00000	0.00000	0	
##	23	0.000000	0.00000	0.00000	0	
##	24	0.000000	0.00000	0.00000	0	
##	25	0.000000	0.00000	0.00000	0	

```

train_final = book.train.full[c(1:7, 68)]
colnames(train_final) = c("ID", "AvgCost", "logtargamt", "Recency", "Frequency", "Amount", "Time", "Respond")
train_final$AvgCost = round(train_final$AvgCost, 2)
train_final$Amount = round(train_final$Amount, 2)
train_final$Respond = as.factor(train_final$Respond)
#head(train_final, 10)

```

# Outlier Detections



```

#repeated part for running
train_final = book.train.full[c(1:7, 68)]
colnames(train_final) = c("ID", "AvgCost", "logtargamt", "Recency", "Frequency", "Amount", "Time", "Respond")
train_final$AvgCost = round(train_final$AvgCost, 2)
train_final$Amount = round(train_final$Amount, 2)
train_final$Respond = as.factor(train_final$Respond)

#avgcost
train_final1 = subset(train_final, train_final$AvgCost >= mean(train_final$AvgCost) -
3*sd(train_final$AvgCost) & train_final$AvgCost <= mean(train_final$AvgCost) +3*sd(tr
ain_final$AvgCost))

#recency
train_final2 = subset(train_final1, train_final1$Recency >= mean(train_final1$Recency
) - 3*sd(train_final1$Recency) & train_final1$Recency <= mean(train_final1$Recency) +
3*sd(train_final1$Recency))

#frequency
train_final3 = subset(train_final2, train_final2$Frequency >= mean(train_final2$Freque
ncy) - 3*sd(train_final2$Frequency) & train_final2$Frequency <= mean(train_final2$Fr
equency) +3*sd(train_final2$Frequency))

#amount
train_final4 = subset(train_final3, train_final3$Amount >= mean(train_final3$Amount)
- 3*sd(train_final3$Amount) & train_final3$Amount <= mean(train_final3$Amount) +3*sd(
train_final3$Amount))

#refine final train
train_final = train_final4

```

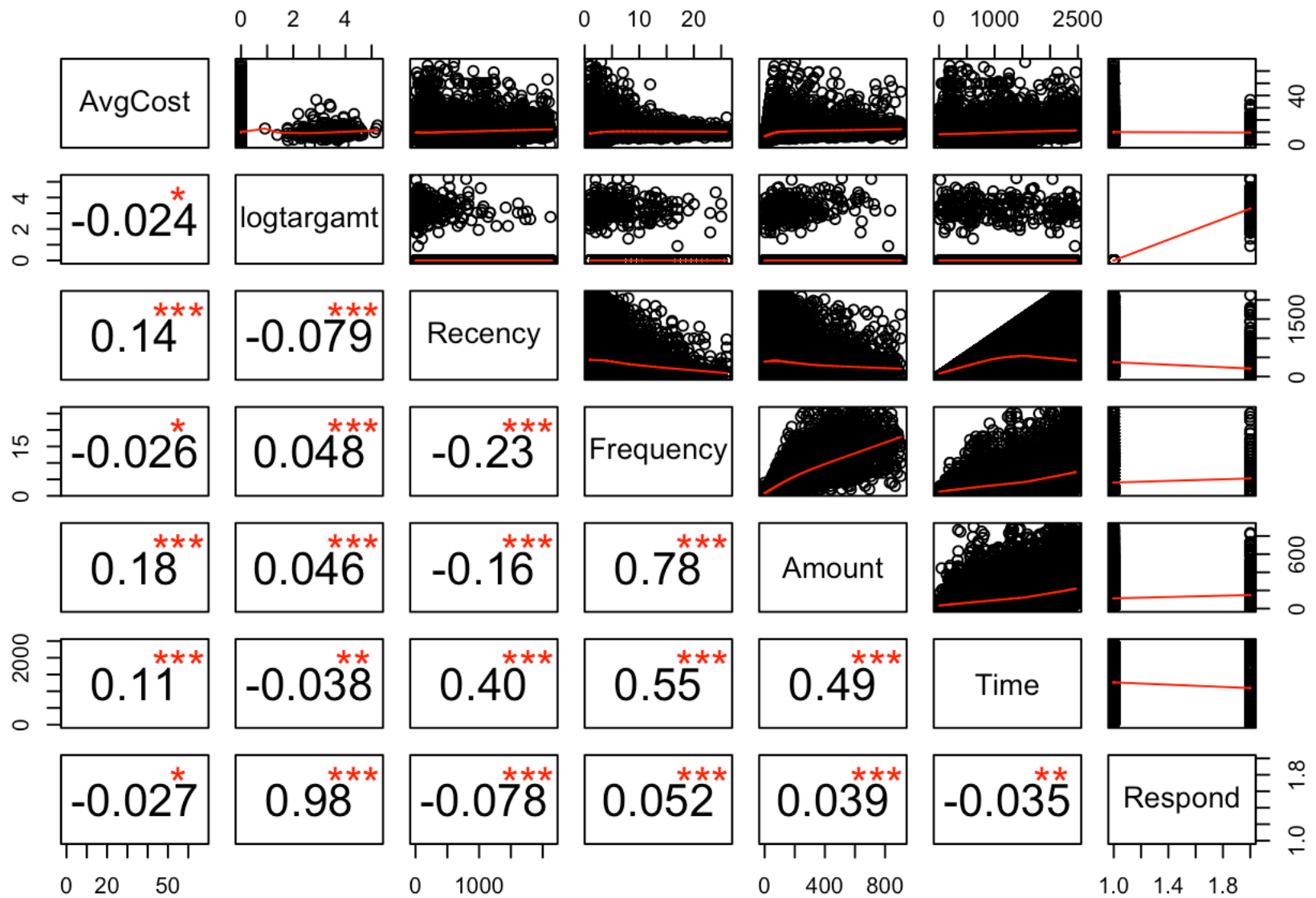
```
summary(train_final)
```

```

##          ID          AvgCost      logtargamt      Recency
##  Min.      :    2623  Min.      : 0.00      Min.      :0.0000  Min.      :    1.0
##  1st Qu.: 4428389  1st Qu.:  7.85      1st Qu.:0.0000  1st Qu.: 128.0
##  Median : 8069611  Median :10.09      Median :0.0000  Median : 304.0
##  Mean    : 8360258  Mean    :11.52      Mean    :0.1101  Mean    : 495.4
##  3rd Qu.:12614986  3rd Qu.:13.09      3rd Qu.:0.0000  3rd Qu.: 703.5
##  Max.    :16252071  Max.    :67.27      Max.    :5.2282  Max.    :2148.0
##  Frequency      Amount      Time      Respond
##  Min.      : 1.000  Min.      :  0.0  Min.      :    2  0:7407
##  1st Qu.: 2.000  1st Qu.: 39.8  1st Qu.: 496  1: 256
##  Median : 3.000  Median : 98.8  Median :1291
##  Mean    : 5.035  Mean    :163.3  Mean    :1260
##  3rd Qu.: 7.000  3rd Qu.:224.0  3rd Qu.:2064
##  Max.    :26.000  Max.    :907.2  Max.    :2462

```

```
source("panelfxns.R")
pairs(train_final[, -1], upper.panel = panel.smooth, lower.panel = panel.cor, cex.lab
els=1.3)
```



```
mod1_logistic = glm(Respond ~ AvgCost + Recency + Frequency + Amount + Time, data = t
rain_final, family = binomial)
summary(mod1_logistic)
```

```
##
## Call:
## glm(formula = Respond ~ AvgCost + Recency + Frequency + Amount +
##      Time, family = binomial, data = train_final)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5950  -0.2983  -0.2538  -0.1893   3.3150
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.8273251  0.1639257 -17.248  < 2e-16 ***
## AvgCost      -0.0176537  0.0119541  -1.477  0.139731
## Recency      -0.0008340  0.0002308  -3.614  0.000302 ***
## Frequency     0.0531199  0.0218328   2.433  0.014973 *
## Amount        0.0005639  0.0005487   1.028  0.304145
## Time         -0.0003660  0.0001235  -2.963  0.003042 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2243.6  on 7662  degrees of freedom
## Residual deviance: 2164.7  on 7657  degrees of freedom
## AIC: 2176.7
##
## Number of Fisher Scoring iterations: 7
```

Amount is the most non-significant predictor, so we first remove it.

```
mod2_logistic = glm(Respond ~ AvgCost + Recency + Frequency + Time, data = train_fina
l, family = binomial)
summary(mod2_logistic)
```

```
##
## Call:
## glm(formula = Respond ~ AvgCost + Recency + Frequency + Time,
##      family = binomial, data = train_final)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5677  -0.2985  -0.2542  -0.1902   3.3118
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.8536104  0.1612429 -17.698  < 2e-16 ***
## AvgCost      -0.0142075  0.0113059  -1.257  0.208883
## Recency      -0.0008478  0.0002302  -3.683  0.000231 ***
## Frequency     0.0672736  0.0166381   4.043  5.27e-05 ***
## Time         -0.0003526  0.0001225  -2.878  0.004004 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2243.6  on 7662  degrees of freedom
## Residual deviance: 2165.7  on 7658  degrees of freedom
## AIC: 2175.7
##
## Number of Fisher Scoring iterations: 7
```

Average Cost is the most non-significant predictor, so we then remove it.

```
mod3_logistic = glm(Respond ~ Recency + Frequency + Time, data = train_final, family
= binomial)
summary(mod3_logistic)
```

```
##
## Call:
## glm(formula = Respond ~ Recency + Frequency + Time, family = binomial,
##      data = train_final)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5642  -0.2972  -0.2560  -0.1907   3.3269
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.9948877  0.1191238 -25.141  < 2e-16 ***
## Recency      -0.0008580  0.0002300  -3.730  0.000191 ***
## Frequency     0.0686628  0.0166166   4.132  3.59e-05 ***
## Time         -0.0003685  0.0001219  -3.024  0.002492 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2243.6  on 7662  degrees of freedom
## Residual deviance: 2167.4  on 7659  degrees of freedom
## AIC: 2175.4
##
## Number of Fisher Scoring iterations: 7
```

## Data Cleaning for Testing Data

```
# Add respond column to indicate if a customer responded to promotion
book$respond <- ifelse(book$logtargamt > 0, 1, 0)
```

```
# Drop empty malformat column
book.test$X <- NULL
```

```
# Orders with orddate as Date and net as total price of order
orders$orddate <- lubridate::dmy(orders$orddate)
```

```
## Warning: All formats failed to parse. No formats found.
```

```
orders$net <- orders$qty * orders$price
```

```
# Orders of customers who respond to promotion
```

```
orders.responders <- book.test %>%
```

```
  filter(logtargamt > 0) %>%
```

```
  merge(orders, by = 'id')
```

```
# All information about the customers in training set
```

```
book.test.merge <- merge(book.test, book, by = 'id')
```

```
book.test.merge$logtargamt.y <- NULL
```

```
book.test.merge <- rename(book.test.merge, logtargamt = logtargamt.x)
```

```
book.test.merge$respond <- ifelse(book.test.merge$logtargamt > 0, 1, 0)
```

```
# All information about customers in training set with addition of avgNetOrder
```

```
book.test.full <- orders %>%
```

```
  group_by(id) %>%
```

```
  #could try sum(net) to see difference
```

```
  summarise(avgNetOrder = mean(net)) %>%
```

```
  merge(book.test.merge, by = 'id')
```

```
head(book.test.full, 25)
```

##	id	avgNetOrder	logtargamt	recency	frequency	amount	tof	Ffiction1
## 1	914	9.965463	0	194	7	318.89478	1703	1
## 2	957	10.825067	0	3	14	368.05225	2364	1
## 3	1406	12.094241	0	1489	15	423.29834	2371	0
## 4	1414	10.173871	0	155	4	71.21704	1290	0
## 5	1546	11.349698	0	194	6	442.63818	2188	0
## 6	1651	11.983398	0	1797	2	47.93359	1808	0
## 7	2046	10.479935	0	383	3	52.39966	2186	0
## 8	2534	9.175542	0	80	7	119.28204	2063	0
## 9	2801	13.693696	0	96	9	164.32434	2453	1
## 10	3689	9.143172	0	705	5	137.14746	1839	0
## 11	4308	7.676315	0	290	9	115.14471	2321	0
## 12	4480	11.961920	0	76	13	370.81934	2462	0
## 13	4847	11.536343	0	181	9	392.23560	2252	0
## 14	4871	9.989156	0	208	7	319.65283	2371	0
## 15	5665	7.811709	0	9	5	124.98730	1480	0
## 16	6157	8.494238	0	75	7	220.85010	437	1
## 17	6190	8.392484	0	83	10	117.49475	2123	0
## 18	6882	13.716762	0	342	25	603.53711	2368	2
## 19	7501	4.908400	0	1474	1	14.72520	1474	0
## 20	8109	13.519214	0	256	7	365.01855	2166	1
## 21	8303	16.820646	0	919	2	100.92383	1011	0
## 22	8699	8.051407	0	643	6	112.71967	1437	0
## 23	8907	11.853447	0	2229	1	35.56033	2229	0
## 24	9512	12.388102	0	109	36	1152.09277	2397	0
## 25	9520	8.958932	0	22	33	1281.12695	2391	1

##	Fclassics3	Fcartoons5	Flegends6	Fphilosophy7	Freligion8	Fpsychology9
## 1	0	1	0	0	0	0
## 2	0	2	0	0	0	0
## 3	0	0	0	0	0	0
## 4	0	0	0	0	0	0
## 5	0	1	0	3	1	0
## 6	0	0	0	0	0	0
## 7	0	0	0	0	0	0
## 8	0	0	0	0	0	0
## 9	1	0	0	0	0	0
## 10	0	0	0	0	1	0
## 11	0	0	0	0	0	0
## 12	0	0	0	1	1	0
## 13	0	0	0	1	0	0
## 14	0	0	0	0	0	0
## 15	0	0	0	0	2	0
## 16	0	0	0	0	0	0
## 17	0	0	0	0	0	0
## 18	3	0	0	1	0	0
## 19	0	0	0	0	0	0
## 20	0	1	0	0	0	0
## 21	0	0	0	0	1	0
## 22	0	2	0	0	1	0
## 23	0	0	0	0	0	0
## 24	0	0	0	0	2	0
## 25	1	1	2	6	21	0
##	Flinguistics10	Fart12	Fmusic14	Ffacsimile17	Fhistory19	Fconthist20
## 1	0	1	7	0	0	17
## 2	0	0	4	0	4	4
## 3	0	0	0	0	1	34
## 4	0	0	0	0	0	7
## 5	1	1	3	0	6	7
## 6	0	0	0	0	2	1
## 7	0	0	3	0	0	0
## 8	0	0	1	0	0	11
## 9	0	0	0	0	7	2
## 10	0	0	0	0	3	0
## 11	0	1	10	0	0	0
## 12	0	0	3	0	5	11
## 13	1	0	2	0	5	19
## 14	0	0	0	0	1	18
## 15	0	0	13	0	0	0
## 16	0	0	0	0	4	17
## 17	0	0	0	0	4	10
## 18	0	0	0	0	22	10
## 19	0	0	0	0	0	0
## 20	0	1	1	0	1	2
## 21	0	0	0	0	2	0
## 22	0	0	1	0	4	0
## 23	0	1	0	0	0	0

##	24	0	0	0	0	3	31
##	25	0	1	3	0	6	1
##		Feconomy21	Fpolitics22	Fscience23	Fcompsci26	Frailroads27	Fmaps30
##	1	0	0	0	0	0	0
##	2	0	0	0	1	0	0
##	3	0	0	0	0	0	0
##	4	0	0	0	0	0	0
##	5	1	0	2	0	2	0
##	6	0	0	0	0	0	0
##	7	0	0	0	0	0	0
##	8	0	0	0	0	0	0
##	9	0	0	0	0	0	0
##	10	0	1	0	0	0	0
##	11	0	0	0	0	0	0
##	12	0	0	7	1	0	0
##	13	0	0	0	0	2	0
##	14	0	0	0	0	5	0
##	15	0	0	0	0	0	0
##	16	0	0	0	0	0	0
##	17	0	0	0	0	0	0
##	18	0	0	0	0	0	0
##	19	0	0	0	0	0	0
##	20	0	0	1	0	0	0
##	21	0	0	0	0	0	0
##	22	0	0	0	0	0	0
##	23	0	0	0	0	0	0
##	24	0	0	17	1	1	0
##	25	0	0	2	1	0	1
##		Ftravelguides31	Fhealth35	Fcooking36	Flearning37	FGamesRiddles38	
##	1	2	2	1	0	0	
##	2	0	7	0	4	0	
##	3	0	0	0	0	0	
##	4	0	0	0	0	0	
##	5	1	3	0	2	0	
##	6	0	0	1	0	0	
##	7	0	0	0	0	1	
##	8	1	0	0	0	0	
##	9	1	0	0	0	0	
##	10	3	0	1	0	0	
##	11	0	2	1	0	0	
##	12	0	0	0	0	0	
##	13	1	1	0	0	0	
##	14	7	1	0	0	0	
##	15	0	0	0	0	0	
##	16	3	0	0	1	0	
##	17	0	0	0	0	0	
##	18	5	0	0	0	0	
##	19	2	0	0	0	0	
##	20	3	4	4	1	0	
##	21	0	1	0	1	0	



##	22	0	1	1	0	0	
##	23	0	1	0	0	0	
##	24	0	11	0	14	0	
##	25	3	50	3	19	1	
##		Fsports39	Fhobby40	Fnature41	Fencyclopaedia44	Fvideos50	Fnonbooks99
##	1	0	0	0	0	0	0
##	2	0	4	2	1	0	0
##	3	0	0	0	0	0	0
##	4	0	0	0	0	0	0
##	5	0	2	0	0	3	0
##	6	0	0	0	0	0	0
##	7	0	0	1	0	0	0
##	8	0	0	0	0	0	0
##	9	0	0	0	0	0	0
##	10	0	2	1	2	0	1
##	11	0	0	0	0	0	1
##	12	0	2	0	0	0	0
##	13	0	0	0	1	0	1
##	14	0	0	0	0	0	0
##	15	0	0	0	0	1	0
##	16	0	0	0	0	0	0
##	17	0	0	0	0	0	0
##	18	0	0	0	0	0	1
##	19	0	0	0	0	0	1
##	20	0	3	4	0	0	0
##	21	0	0	1	0	0	0
##	22	0	1	0	1	2	0
##	23	0	0	1	0	0	0
##	24	0	0	2	6	5	0
##	25	0	8	3	8	1	0
##		Mfiction1	Mclassics3	Mcartoons5	Mlegends6	Mphilosophy7	Mreligion8
##	1	5.949997	0.000000	7.643810	0.00000	0.00000	0.000000
##	2	12.680038	0.000000	20.236496	0.00000	0.00000	0.000000
##	3	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	4	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	5	0.000000	0.000000	8.899994	0.00000	26.19841	19.899994
##	6	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	7	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	8	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	9	0.000000	25.539017	0.000000	0.00000	0.00000	0.000000
##	10	0.000000	0.000000	0.000000	0.00000	0.00000	15.236496
##	11	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	12	0.000000	0.000000	0.000000	0.00000	10.12357	7.643810
##	13	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	14	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	15	0.000000	0.000000	0.000000	0.00000	0.00000	17.767380
##	16	3.949999	0.000000	0.000000	0.00000	0.00000	0.000000
##	17	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000
##	18	23.287613	28.990234	0.000000	0.00000	7.64381	0.000000
##	19	0.000000	0.000000	0.000000	0.00000	0.00000	0.000000

##	20	8.640831	0.000000	8.899994	0.000000	0.000000	0.000000
##	21	0.000000	0.000000	0.000000	0.000000	0.000000	12.782295
##	22	0.000000	0.000000	15.067116	0.000000	0.000000	6.949997
##	23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	24	0.000000	0.000000	0.000000	0.000000	0.000000	24.899994
##	25	8.589699	2.949999	7.643810	41.38904	61.20639	185.334228
##		Mpsychology9	Mlinguistics10	Mart12	Mmusic14	Mfacsimile17	Mhistory19
##	1	0	0.000000	8.589699	61.429962	0	0.000000
##	2	0	0.000000	0.000000	29.414169	0	53.839020
##	3	0	0.000000	0.000000	0.000000	0	10.123573
##	4	0	0.000000	0.000000	0.000000	0	0.000000
##	5	0	6.621227	15.313187	39.804047	0	64.744934
##	6	0	0.000000	0.000000	0.000000	0	17.332779
##	7	0	0.000000	0.000000	27.136490	0	0.000000
##	8	0	0.000000	0.000000	6.544533	0	0.000000
##	9	0	0.000000	0.000000	0.000000	0	103.224976
##	10	0	0.000000	0.000000	0.000000	0	37.753601
##	11	0	0.000000	8.589699	84.419617	0	0.000000
##	12	0	0.000000	0.000000	72.682251	0	55.909729
##	13	0	10.174706	0.000000	15.037346	0	60.699921
##	14	0	0.000000	0.000000	0.000000	0	4.949997
##	15	0	0.000000	0.000000	90.319946	0	0.000000
##	16	0	0.000000	0.000000	0.000000	0	32.749969
##	17	0	0.000000	0.000000	0.000000	0	31.383774
##	18	0	0.000000	0.000000	0.000000	0	393.441162
##	19	0	0.000000	0.000000	0.000000	0	0.000000
##	20	0	0.000000	15.287621	9.949997	0	14.899994
##	21	0	0.000000	0.000000	0.000000	0	19.403519
##	22	0	0.000000	0.000000	5.989998	0	41.438995
##	23	0	0.000000	15.236496	0.000000	0	0.000000
##	24	0	0.000000	0.000000	0.000000	0	70.256836
##	25	0	0.000000	15.313187	15.287621	0	67.541626
##		Mconthist20	Meconomy21	Mpolitics22	Mscience23	Mcompsci26	Mrailroads27
##	1	197.93555	0.000000	0.000000	0.000000	0.000000	0.000000
##	2	60.74997	0.000000	0.000000	0.000000	5.899998	0.000000
##	3	413.17480	0.000000	0.000000	0.000000	0.000000	0.000000
##	4	71.21704	0.000000	0.000000	0.000000	0.000000	0.000000
##	5	86.94574	5.112919	0.000000	19.84999	0.000000	9.960655
##	6	10.20027	0.000000	0.000000	0.000000	0.000000	0.000000
##	7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	8	108.67273	0.000000	0.000000	0.000000	0.000000	0.000000
##	9	20.32384	0.000000	0.000000	0.000000	0.000000	0.000000
##	10	0.000000	0.000000	18.89999	0.000000	0.000000	0.000000
##	11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	12	116.60895	0.000000	0.000000	78.86084	10.123573	0.000000
##	13	253.31567	0.000000	0.000000	0.000000	0.000000	22.123566
##	14	185.88293	0.000000	0.000000	0.000000	0.000000	58.517334
##	15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	16	144.40015	0.000000	0.000000	0.000000	0.000000	0.000000
##	17	86.11096	0.000000	0.000000	0.000000	0.000000	0.000000

##	18	105.87128	0.000000	0.000000	0.000000	0.000000	0.000000
##	19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	20	39.84998	0.000000	0.000000	15.28762	0.000000	0.000000
##	21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
##	24	401.64624	0.000000	0.000000	317.01514	8.666389	10.200272
##	25	25.46233	0.000000	0.000000	22.93143	5.010658	0.000000
##	Mmaps30 Mtravelguides31 Mhealth35 Mcooking36 Mlearning37						
##	1	0.000000	12.271004	14.899994	10.174706	0.000000	
##	2	0.000000	0.000000	74.156250	0.000000	22.749985	
##	3	0.000000	0.000000	0.000000	0.000000	0.000000	
##	4	0.000000	0.000000	0.000000	0.000000	0.000000	
##	5	0.000000	10.123573	41.518219	0.000000	7.949997	
##	6	0.000000	0.000000	0.000000	20.400543	0.000000	
##	7	0.000000	0.000000	0.000000	0.000000	0.000000	
##	8	0.000000	4.064770	0.000000	0.000000	0.000000	
##	9	0.000000	15.236496	0.000000	0.000000	0.000000	
##	10	0.000000	26.791672	0.000000	6.949997	0.000000	
##	11	0.000000	0.000000	15.185364	6.949997	0.000000	
##	12	0.000000	0.000000	0.000000	0.000000	0.000000	
##	13	0.000000	6.621227	8.949997	0.000000	0.000000	
##	14	0.000000	63.706940	6.595665	0.000000	0.000000	
##	15	0.000000	0.000000	0.000000	0.000000	0.000000	
##	16	0.000000	29.849991	0.000000	0.000000	9.899994	
##	17	0.000000	0.000000	0.000000	0.000000	0.000000	
##	18	0.000000	44.303406	0.000000	0.000000	0.000000	
##	19	0.000000	14.725197	0.000000	0.000000	0.000000	
##	20	0.000000	33.182800	55.909729	66.263367	2.990000	
##	21	0.000000	0.000000	8.589699	0.000000	49.948090	
##	22	0.000000	0.000000	2.500000	10.123573	0.000000	
##	23	0.000000	0.000000	5.010658	0.000000	0.000000	
##	24	0.000000	0.000000	78.270447	0.000000	108.036743	
##	25	5.010658	35.100159	367.708496	22.880295	220.015259	
##	MGamesRiddles38 Msports39 Mhobby40 Mnature41 Mencyclopaedia44 Mvideos50						
##	1	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	2	0.000000	0	61.17603	17.15027	10.000000	0.000000
##	3	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	4	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	5	0.000000	0	17.89520	0.000000	0.000000	61.79999
##	6	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	7	9.949997	0	0.000000	15.31319	0.000000	0.000000
##	8	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	9	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	10	0.000000	0	12.72584	10.12357	8.666389	0.000000
##	11	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	12	0.000000	0	18.86665	0.000000	0.000000	0.000000
##	13	0.000000	0	0.000000	0.000000	15.313187	0.000000
##	14	0.000000	0	0.000000	0.000000	0.000000	0.000000
##	15	0.000000	0	0.000000	0.000000	0.000000	16.89999

##	16	0.000000	0	0.00000	0.00000	0.000000	0.00000
##	17	0.000000	0	0.00000	0.00000	0.000000	0.00000
##	18	0.000000	0	0.00000	0.00000	0.000000	0.00000
##	19	0.000000	0	0.00000	0.00000	0.000000	0.00000
##	20	0.000000	0	44.78915	49.06757	0.000000	0.00000
##	21	0.000000	0	0.00000	10.20027	0.000000	0.00000
##	22	0.000000	0	14.95000	0.00000	1.950000	13.74999
##	23	0.000000	0	0.00000	15.31319	0.000000	0.00000
##	24	0.000000	0	0.00000	15.03735	31.923569	86.14008
##	25	10.174706	0	60.61362	35.63702	55.126373	10.20027
##	Mnonbooks99 respond						
##	1	0	0				
##	2	0	0				
##	3	0	0				
##	4	0	0				
##	5	0	0				
##	6	0	0				
##	7	0	0				
##	8	0	0				
##	9	0	0				
##	10	0	0				
##	11	0	0				
##	12	0	0				
##	13	0	0				
##	14	0	0				
##	15	0	0				
##	16	0	0				
##	17	0	0				
##	18	0	0				
##	19	0	0				
##	20	0	0				
##	21	0	0				
##	22	0	0				
##	23	0	0				
##	24	0	0				
##	25	0	0				

```
test_final = book.test.full[c(1:7, 68)]
colnames(test_final) = c("ID", "AvgCost", "logtargamt", "Recency", "Frequency", "Amount", "Time", "Respond")
test_final$AvgCost = round(test_final$AvgCost, 2)
test_final$Amount = round(test_final$Amount, 2)
test_final$Respond = as.factor(test_final$Respond)
head(test_final, 10)
```

##	ID	AvgCost	logtargamt	Recency	Frequency	Amount	Time	Respond
## 1	914	9.97	0	194	7	318.89	1703	0
## 2	957	10.83	0	3	14	368.05	2364	0
## 3	1406	12.09	0	1489	15	423.30	2371	0
## 4	1414	10.17	0	155	4	71.22	1290	0
## 5	1546	11.35	0	194	6	442.64	2188	0
## 6	1651	11.98	0	1797	2	47.93	1808	0
## 7	2046	10.48	0	383	3	52.40	2186	0
## 8	2534	9.18	0	80	7	119.28	2063	0
## 9	2801	13.69	0	96	9	164.32	2453	0
## 10	3689	9.14	0	705	5	137.15	1839	0

```
predicted = predict(mod3_logistic, test_final, type = "response")
```

```
SSE_test1 = sum((as.numeric(test_final$Respond) - predicted)^2); SSE_test1
```

```
## [1] 25726.65
```