Marketing Analytics: Amazon Review Sentiment & Customer Segmentation

Hanzhe Zhang

Step 1

```
#Set seed
set.seed(2600)

# Install pacman.
if(!"pacman" %in% installed.packages()[,"Package"]) install.packages("pacman")

# Install and load the required packages.
pacman::p_load(rvest, dplyr, tidyr, stringr, DT, RCurl, XML, purrr)
```

Scraping reviews of iLife V5s

Pruduct page of iLife V5s

https://www.amazon.com/ILIFE-Automatically-Sweeping-Scrubbing-Cleaning/dp/B06X1F3HXG/ref=sr_1_4?keywords=iLIFE&qid=1557164153&s=gateway&sr=8-4&th=1 (https://www.amazon.com/ILIFE-Automatically-Sweeping-Scrubbing-Cleaning/dp/B06X1F3HXG/ref=sr_1_4?keywords=iLIFE&qid=1557164153&s=gateway&sr=8-4&th=1)

Amazon Product Code is B06X1F3HXG

```
prod_code <- "B06X1F3HXG"
```

```
# Start by getting the product information using the URL
url <- paste0("https://www.amazon.com/dp/", prod_code)
doc <- read_html(url)

# Obtain the text in the node, remove "\n", remove white spaces from the text
prod <- html_nodes(doc, "#productTitle") %>% html_text() %>% gsub("\n", "", .) %>% trimws()
prod
```

```
## [1] "ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping Scrubbing Mopping Fl
oor Cleaning Robot"
```

```
# Scrape elements from Amazon reviews
scrape_amazon <- function(url, throttle = 0){

  # Install / Load relevant packages

  if(!"pacman" %in% installed.packages()[,"Package"]) install.packages("pacman")
  pacman::p_load(RCurl, XML, dplyr, stringr, rvest, purrr)

  # Set throttle between URL calls
  sec = 0
  if(throttle < 0) warning("throttle was less than 0: set to 0")
  if(throttle > 0) sec = max(0, throttle + runif(1, -1, 1))

  # obtain HTML of URL
  doc <- read_html(url)

  # Parse relevant elements from HTML, you can add more if needed
  title <- doc %>%
    html_nodes("#cm_cr-review_list .a-color-base") %>%
    html_text() %>% gsub("\n", "", .) %>% trimws()

  author <- doc %>%
    html_nodes("#cm_cr-review_list .a-profile-name") %>%
    html_text()

  date <- doc %>%
    html_nodes("#cm_cr-review_list .review-date") %>%
    html_text() %>% gsub(".*on ", "", .)

  review_format <- doc %>%
    html_nodes(".review-format-strip") %>% html_text()

  stars <- doc %>%
    html_nodes("#cm_cr-review_list  .review-rating") %>%
    html_text() %>% str_extract("\\d") %>% as.numeric()

  comments <- doc %>%
    html_nodes("#cm_cr-review_list .review-text") %>%
    html_text() %>% gsub("\n", "", .) %>% trimws()

  suppressWarnings(n_helpful <- doc %>%
                     html_nodes(".a-expander-inline-container") %>%
                     html_text() %>%
                     gsub("\n\n \\s*|found this helpful.*", "", .) %>%
                     gsub("One", "1", .) %>%
                     map_chr(~ str_split(string = .x, pattern = " ")[[1]][1]) %>%
                     as.numeric())

  # Combine attributes into a single data frame
  dfiLIFE <- data.frame(title, author, date, review_format, stars, comments, n_helpful, stringsAsFactors =
F)

  return(dfiLIFE)
}

#Do a test run, get the first page of reviews for a test and see if we succeed.
urltesting<- "http://www.amazon.com/product-reviews/B06X1F3HXG/?pageNumber=1"
```

```
reviewstesting <- scrape_amazon(urltesting)
str(reviewstesting)
```

```
## 'data.frame':    10 obs. of  7 variables:
##  $ title        : chr  "Our vacuum committed suicide" "this has been a fun mission, to say the least" "P
romising, but disappointing" "Best Cleaning Robot EVER!" ...
##  $ author       : chr  "Lizz" "Jami James" "Review Panda" "Adrian" ...
##  $ date         : chr  "July 29, 2018" "April 22, 2018" "June 23, 2017" "May 10, 2016" ...
##  $ review_format: chr  "Size: V5s ProVerified Purchase" "Size: V5s ProVerified Purchase" "Size: V5sVerif
ied Purchase" "Size: V5s" ...
##  $ stars        : num  1 1 2 5 3 5 5 5 1 5
##  $ comments     : chr  "I wish that I could say that everything was going fine until Rosie, our vacuumin
g unit, hurled herself down the"| __truncated__ "Ok, this has been a fun mission, to say the least. I have
been on a journey to find the best robot vacuum they "| __truncated__ " I read a lot about robot vacuums an
d was swayed by the Amazon reviews on the iLife V5s. Unfortunately, my expe"| __truncated__ "=== Original R
eview Written May 10, 2016 === (An update can be found below)This is the best cleaning robot ever"| __trunc
ated__ ...
##  $ n_helpful    : num  649 431 178 752 50 51 28 16 15 15
```

```
# set appropriate number of pages
pages <- 100

# reviews_iLIFE will store all our reviews extracted, start with null
reviews_iLIFE <- NULL

# loop through the pages
for(page_num in 1:pages){

  url <- paste0("http://www.amazon.com/product-reviews/",prod_code,"/?pageNumber=", page_num)
  reviews <- scrape_amazon(url, throttle = 2)
  reviews_iLIFE <- rbind(reviews_iLIFE, cbind(prod, reviews))
}

#examine the structure of the reviews extracted
str(reviews_iLIFE)
```

```
## 'data.frame':    1000 obs. of  8 variables:
##  $ prod         : Factor w/ 1 level "ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatical
ly Sweeping Scrubbing Mopping Floor Cleaning Robot": 1 1 1 1 1 1 1 1 1 1 ...
##  $ title        : chr  "Our vacuum committed suicide" "this has been a fun mission, to say the least" "P
romising, but disappointing" "Best Cleaning Robot EVER!" ...
##  $ author       : chr  "Lizz" "Jami James" "Review Panda" "Adrian" ...
##  $ date         : chr  "July 29, 2018" "April 22, 2018" "June 23, 2017" "May 10, 2016" ...
##  $ review_format: chr  "Size: V5s ProVerified Purchase" "Size: V5s ProVerified Purchase" "Size: V5sVerif
ied Purchase" "Size: V5s" ...
##  $ stars        : num  1 1 2 5 3 5 5 5 1 5 ...
##  $ comments     : chr  "I wish that I could say that everything was going fine until Rosie, our vacuumin
g unit, hurled herself down the"| __truncated__ "Ok, this has been a fun mission, to say the least. I have
been on a journey to find the best robot vacuum they "| __truncated__ " I read a lot about robot vacuums an
d was swayed by the Amazon reviews on the iLife V5s. Unfortunately, my expe"| __truncated__ "=== Original R
eview Written May 10, 2016 === (An update can be found below)This is the best cleaning robot ever"| __trunc
ated__ ...
##  $ n_helpful    : num  649 431 178 752 50 51 28 16 15 15 ...
```

```
# save the reviews collected as a csv file
write.csv(reviews_iLIFE, file=" B06X1F3HXG_Reviews.csv")
```

Sentence sentiment analysis for iLife V5s

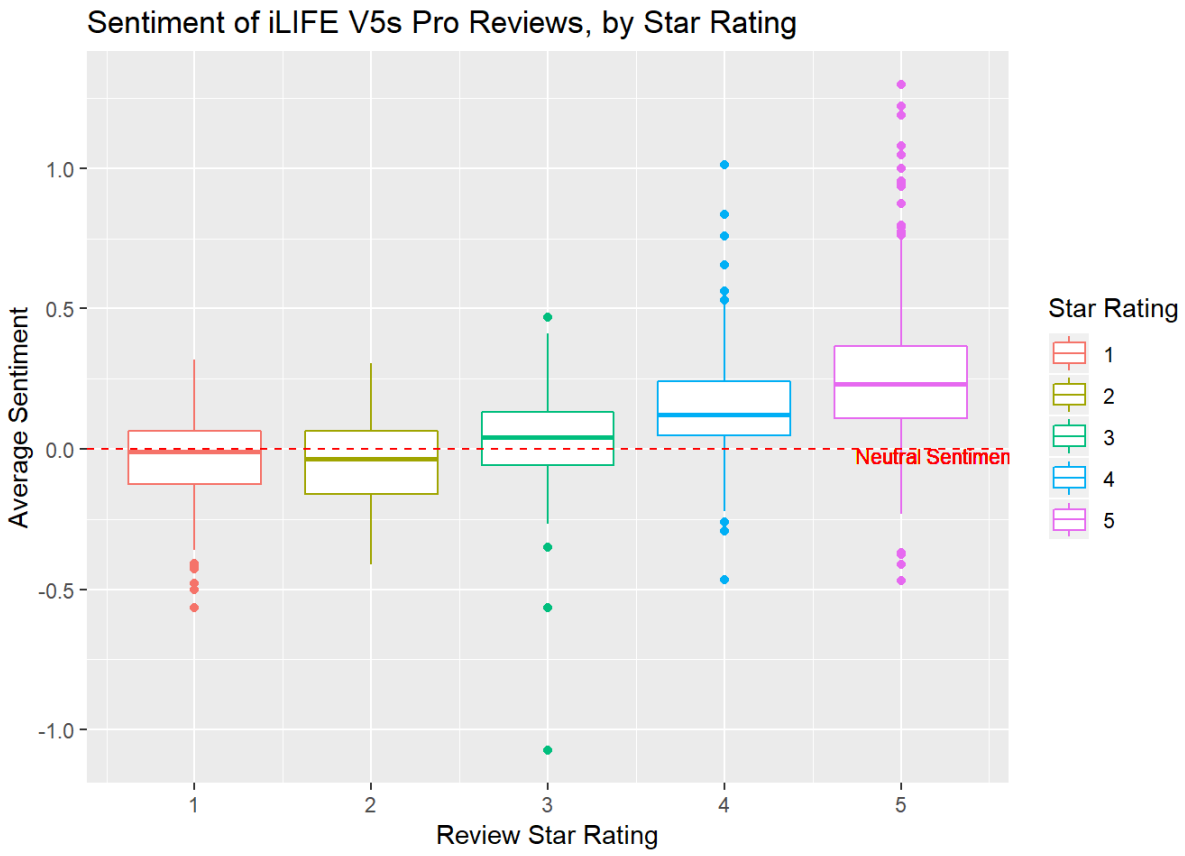
```
# Install and load the required packages.
pacman::p_load(tidyr, dplyr, stringr, data.table, sentimentr, ggplot2)

# load the data
reviews_iLIFE = read.csv(file.choose(), stringsAsFactors = F)

# create a rowid for the reviews
review_dfiLIFE <- reviews_iLIFE %>% mutate(id = row_number())

# examine the structure
str(review_dfiLIFE)
```

```
## 'data.frame':    1000 obs. of  10 variables:
##  $ X            : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ prod         : chr  "ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping S
crubbing Mopping Floor Cleaning Robot" "ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatical
ly Sweeping Scrubbing Mopping Floor Cleaning Robot" "ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tan
k, Automatically Sweeping Scrubbing Mopping Floor Cleaning Robot" "ILIFE V5s Pro Robot Vacuum Mop Cleaner w
ith Water Tank, Automatically Sweeping Scrubbing Mopping Floor Cleaning Robot" ...
##  $ title        : chr  "Our vacuum committed suicide" "this has been a fun mission, to say the least" "P
romising, but disappointing" "Best Cleaning Robot EVER!" ...
##  $ author       : chr  "Lizz" "Jami James" "Review Panda" "Adrian" ...
##  $ date         : chr  "July 29, 2018" "April 22, 2018" "June 23, 2017" "May 10, 2016" ...
##  $ review_format: chr  "Size: V5s ProVerified Purchase" "Size: V5s ProVerified Purchase" "Size: V5sVerif
ied Purchase" "Size: V5s" ...
##  $ stars        : int  1 1 2 5 3 5 5 5 1 5 ...
##  $ comments     : chr  "I wish that I could say that everything was going fine until Rosie, our vacuumin
g unit, hurled herself down the"| __truncated__ "Ok, this has been a fun mission, to say the least. I have
been on a journey to find the best robot vacuum they "| __truncated__ " I read a lot about robot vacuums an
d was swayed by the Amazon reviews on the iLife V5s. Unfortunately, my expe"| __truncated__ "=== Original R
eview Written May 10, 2016 === (An update can be found below)This is the best cleaning robot ever"| __trunc
ated__ ...
##  $ n_helpful    : int  649 431 178 752 50 51 28 16 15 15 ...
##  $ id           : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
# define the lexicon and any changes needed for our context
# get n rows - to see what we have in the lexicon -
# Tyler Rinker is the author of sentimentr
nrow(lexicon::hash_sentiment_jockers_rinker)
```

```
## [1] 11710
```

```
# words appearing in product title to replace.
replace_in_lexicon1 <- tribble(
  ~x, ~y,
  "ILIFE", 0,
  "Pro", 0,
  "Robot", 0,
  "Mop", 0,
  "Cleaner", 0,
  "Water", 0,
  "Tank", 0,
  "Automatically", 0,
  "Sweeping", 0,
  "Scrubbing", 0,
  "floor", 0,
  "cleaning", 0,
)
# create a new lexicon with modified sentiment
review_lexicon1 <- lexicon::hash_sentiment_jockers_rinker %>%
  filter(!x %in% replace_in_lexicon1$x) %>%
  bind_rows(replace_in_lexicon1) %>%
  setDT() %>%
  setkey("x")

# start by getting the sentence level sentiment for testing
# get sentence-level sentiment
sent_dfiLIFE <- review_dfiLIFE %>%
  get_sentences() %>%
  sentiment_by(by = c('id', 'author', 'date', 'stars', 'review_format'),  polarity_dt = review_lexicon1)

# start by getting the sentence level sentiment for testing
# check the relationship between star rating and sentiment
ggplot(sent_dfiLIFE, aes(x = stars, y = ave_sentiment, color = factor(stars), group = stars)) +
  geom_boxplot() +
  geom_hline(yintercept=0, linetype="dashed", color = "red") +
  geom_text(aes(5.2, -0.05, label = "Neutral Sentiment", vjust = 0), size = 3, color = "red") +
  guides(color = guide_legend(title="Star Rating")) +
  ylab("Average Sentiment") +
  xlab("Review Star Rating") +
  ggtitle("Sentiment of iLIFE V5s Pro Reviews, by Star Rating")
```



Sentence sentiment analysis for iRobot Roomba 650 and 880

```
# load the data
reviews_iRobot = read.csv(file.choose(), stringsAsFactors = F)
review_iRobot650 <- reviews_iRobot[which(reviews_iRobot$Product== 'iRobot Roomba 650 for Pets'), ]
review_iRobot880 <- reviews_iRobot[which(reviews_iRobot$Product== 'iRobot Roomba 880 for Pets and Allergie
s'), ]
# create a rowid for the reviews
review_df650 <- review_iRobot650 %>% mutate(id = row_number())
review_df880 <- review_iRobot880 %>% mutate(id = row_number())
# examine the structure
str(review_df650)
```

```
## 'data.frame':    633 obs. of  6 variables:
##  $ Date   : chr  "2/28/15" "1/12/15" "12/26/13" "8/4/13" ...
##  $ Product: chr  "iRobot Roomba 650 for Pets" "iRobot Roomba 650 for Pets" "iRobot Roomba 650 for Pets"
"iRobot Roomba 650 for Pets" ...
##  $ Stars  : int  5 4 5 3 5 5 1 5 5 5 ...
##  $ Title  : chr  "Five Stars" "Four Stars" "Awesome love it." "Love-hate this vaccuum" ...
##  $ Review : chr  "You would not believe how well this works" "You just walk away and it does the rest"
"You have to Roomba proof your house but once that is done it is awesome. It's like having a pet that clean
s." "Yes, it's a fascinating, albeit expensive, bit of technology. That said, it is a pain in the butt to c
lean - ev"| __truncated__ ...
##  $ id     : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
str(review_df880)
```

```
## 'data.frame':    1200 obs. of  6 variables:
##  $ Date   : chr  "8/10/15" "9/9/15" "11/18/15" "10/31/14" ...
##  $ Product: chr  "iRobot Roomba 880 for Pets and Allergies" "iRobot Roomba 880 for Pets and Allergies"
"iRobot Roomba 880 for Pets and Allergies" "iRobot Roomba 880 for Pets and Allergies" ...
##  $ Stars  : int  5 5 4 5 5 5 5 5 5 2 ...
##  $ Title  : chr  "Your house isn't \"Roomba Clean.\"" "Yes, it works.Ê It works really, really well." "L
OVE our roomba" "My previous Roomba was working fine and I was reluctant to shell out the cash ..." ...
##  $ Review : chr  "You think your floors are clean?Ê They're probably not. The first time I unleashed by
Roomba 880 on what I thou"| __truncated__ "Yes, it really does work.Ê We have dogs who shed tremendously, a
nd running the Roomba a couple of times per wee"| __truncated__ "Yes to everything everyone else said. LOVE
Ê our roomba. But- be forewarned, if the battery dies- you have to re"| __truncated__ "Yes there is a diffe
rence!Ê I have had every model of Roomba since the very first model came out.Ê My previous "| __truncated__
...
##  $ id     : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
# define the lexicon and any changes needed for our context
# get n rows – to see what we have in the lexicon –
# Tyler Rinker is the author of sentimentr
nrow(lexicon::hash_sentiment_jockers_rinker)
```

```
## [1] 11710
```

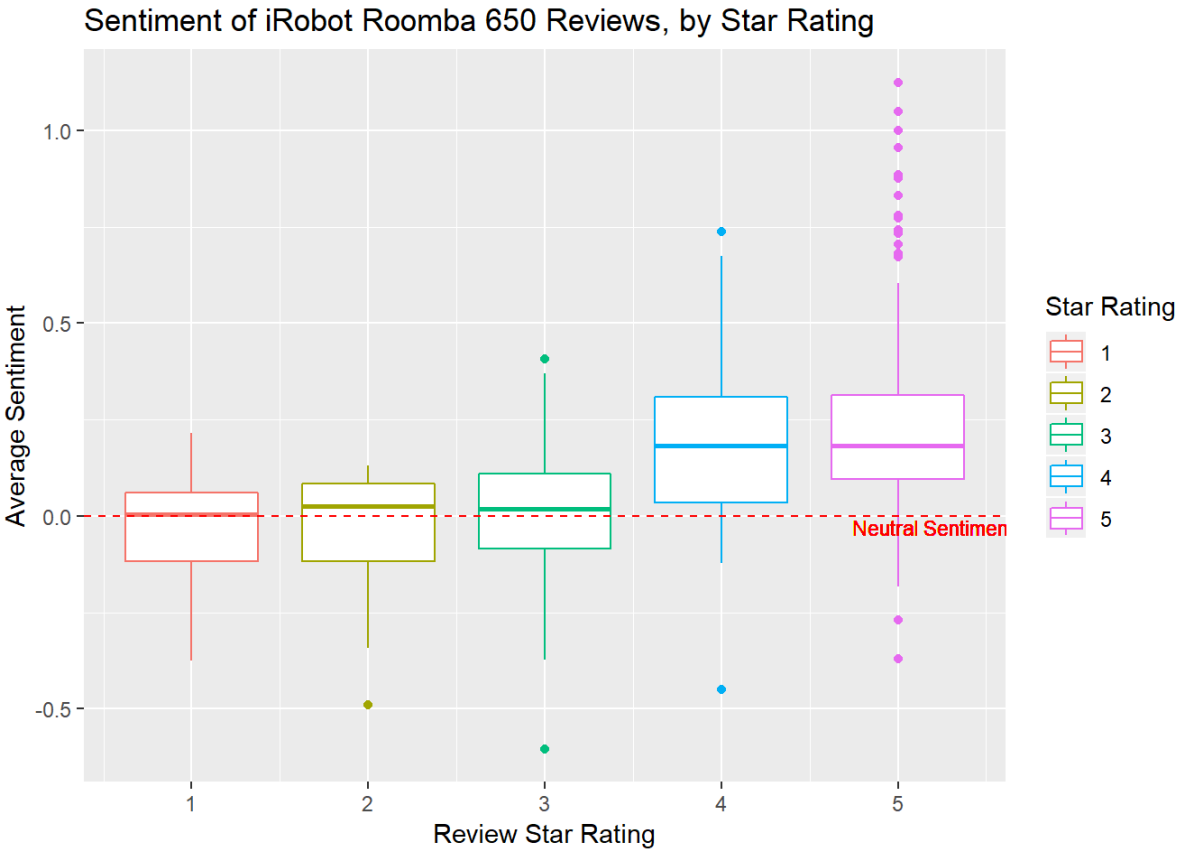
```
# words appearing in product title to replace.
replace_in_lexicon2 <- tribble(
  ~x, ~y,
  "iRobot", 0,
  "Roomba", 0,
  "Pets", 0,
)
# create a new lexicon with modified sentiment
review_lexicon2 <- lexicon::hash_sentiment_jockers_rinker %>%
  filter(!x %in% replace_in_lexicon2$x) %>%
  bind_rows(replace_in_lexicon2) %>%
  setDT() %>%
  setkey("x")

# start by getting the sentence level sentiment for testing
# get sentence-level sentiment
sent_df650 <- review_df650 %>%
  get_sentences() %>%
  sentiment_by(by = c('id', 'Date', 'Stars'),  polarity_dt = review_lexicon2)

sent_df880 <- review_df880 %>%
  get_sentences() %>%
  sentiment_by(by = c('id', 'Date', 'Stars'),  polarity_dt = review_lexicon2)
```

start by getting the sentence level sentiment for testing

check the relationship between star rating and sentiment

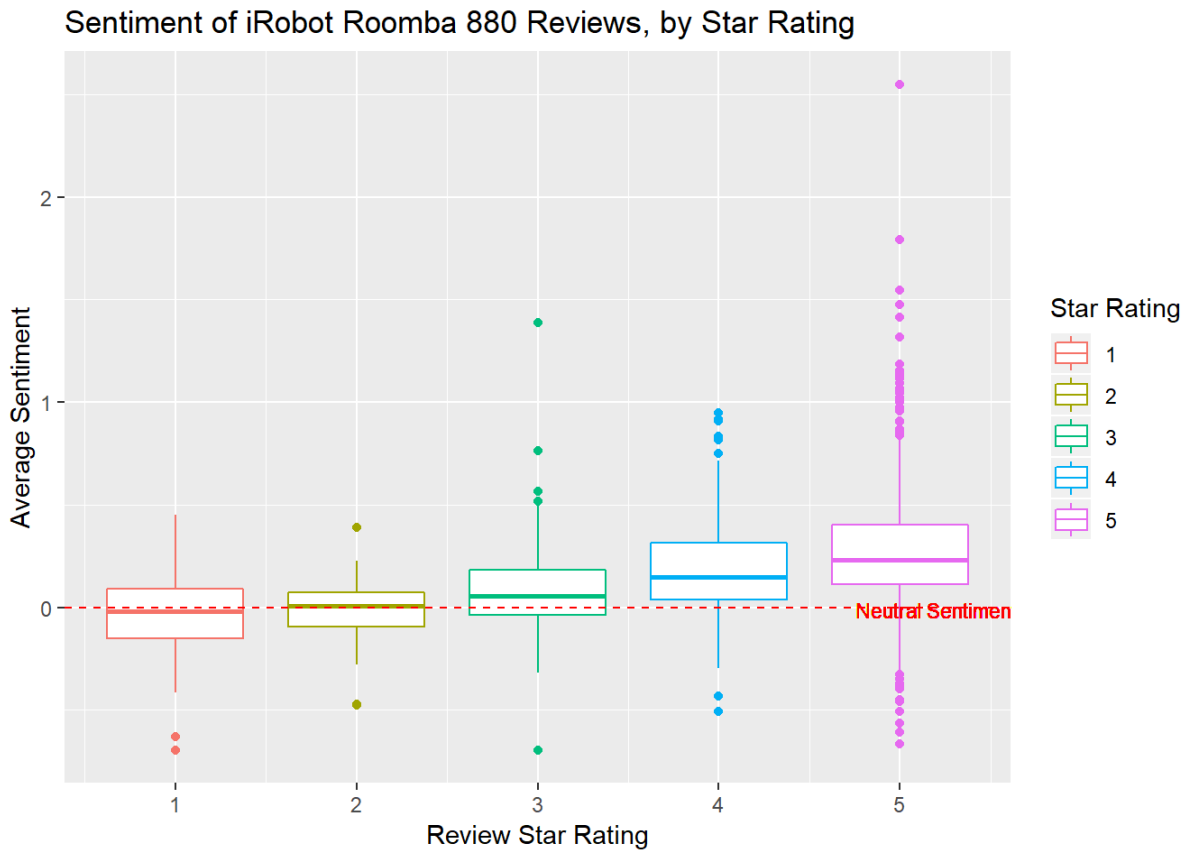
Sentiment of iRobot Roomba 650 Reviews, by Star Rating

```
ggplot(sent_df650, aes(x = Stars, y = ave_sentiment, color = factor(Stars), group = Stars)) +
  geom_boxplot() +
  geom_hline(yintercept=0, linetype="dashed", color = "red") +
  geom_text(aes(5.2, -0.05, label = "Neutral Sentiment", vjust = 0), size = 3, color = "red") +
  guides(color = guide_legend(title="Star Rating")) +
  ylab("Average Sentiment") +
  xlab("Review Star Rating") +
  ggtitle("Sentiment of iRobot Roomba 650 Reviews, by Star Rating")
```



Sentiment of iRobot Roomba 880 Reviews, by Star Rating

```
ggplot(sent_df880, aes(x = Stars, y = ave_sentiment, color = factor(Stars), group = Stars)) +
  geom_boxplot() +
  geom_hline(yintercept=0, linetype="dashed", color = "red") +
  geom_text(aes(5.2, -0.05, label = "Neutral Sentiment", vjust = 0), size = 3, color = "red") +
  guides(color = guide_legend(title="Star Rating")) +
  ylab("Average Sentiment") +
  xlab("Review Star Rating") +
  ggtitle("Sentiment of iRobot Roomba 880 Reviews, by Star Rating")
```



The plots that I got above are the sentence sentiment analysis for iLIFE V5s, iRobot Roomba 650 and iRobott Roomba 880. They shows the relation between review star rating and average sentiment of customers' comments. As the plots show, overall, positive reviews are related to higher star ratings and vice versa. However, for iLIFE V5s, the average sentiment of 2 stars rating is lower than the average sentiment of 1 star rating, even lower than 0.

Step 2

#Word Cloud for iLIFE V5s

```
# Install and load the required packages.
pacman::p_load(dplyr, ggplot2, tidytext, wordcloud2)
```

```
names(reviews_iLIFE)
```

```
## [1] "X"             "prod"          "title"         "author"
## [5] "date"          "review_format" "stars"         "comments"
## [9] "n_helpful"
```

```
# select only the comment.
iLIFE<-reviews_iLIFE %>% select(comment = comments, X)
glimpse(iLIFE)
```

```
## Observations: 1,000
## Variables: 2
## $ comment <chr> "I wish that I could say that everything was going fin...
## $ X       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,...
```

```
# delete all undesirable words, here we only delete things that may bias analyses
# adjust this list as you need it, basically eliminate all undesirable words
undesirable_words_iLife <- c("iLIFE", "pro", "robot", "theres", "tank", "water", "cleaner",
                                  "wanna", "gonna", "what", "gotta", "make",
                                  "automatically", "sweeping", "vacuum", "scrubbing",
                                  "then", "those", "when")

# check out a small sample of stop words, randomly
head(sample(stop_words$word, 15), 15)
```

```
##  [1] "a's"       "i"         "until"     "being"     "say"
##  [6] "et"        "gone"      "among"     "just"      "really"
## [11] "yet"       "they're"   "on"        "thanks"    "therefore"
```

```
# unnest the comment, remove all stop and undesirable words and words smaller than 3 characters and examine
the result
# unnest and remove stop, undesirable and short words
iLIFE_words_filtered <- iLIFE%>%
  unnest_tokens(word, comment) %>%
  anti_join(stop_words) %>%
  distinct() %>%
  filter(!word %in% undesirable_words_iLife) %>%
  filter(nchar(word) > 3)
```

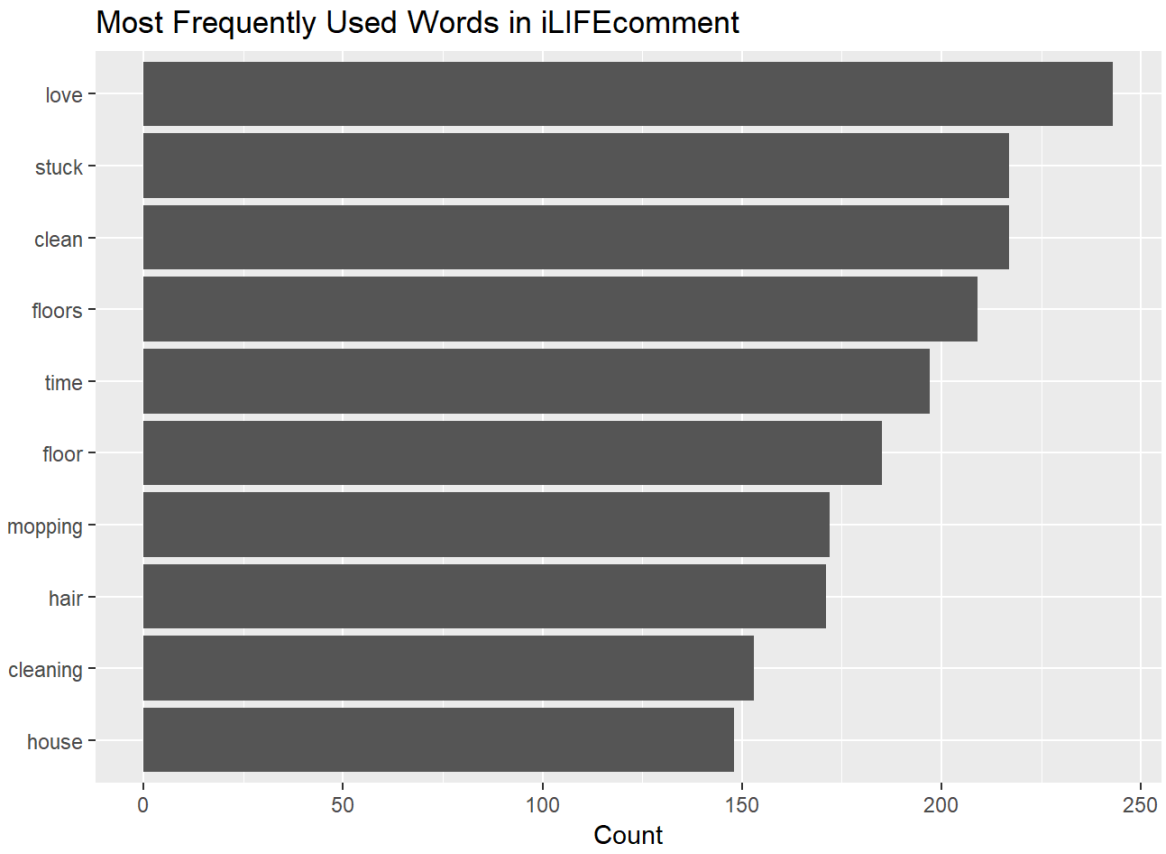
```
## Joining, by = "word"
```

```
dim(iLIFE_words_filtered)
```

```
## [1] 20036     2
```

```
# get the full word count from the comment and quickly examine the results
full_word_count_iLife <- iLIFE%>%
  unnest_tokens(word, comment) %>%
  group_by(X) %>%
  summarise(num_words = n()) %>%
  arrange(desc(num_words))

# plot the most commonly used words in the comment
iLIFE_words_filtered %>%
  count(word, sort = TRUE) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot() +
  geom_col(aes(word, n)) +
  xlab("") +
  ylab("Count") +
  ggtitle("Most Frequently Used Words in iLIFEcomment") +
  coord_flip()
```

```
## Selecting by n
```




```
# create a cool wordcloud of the words in the comment
iLIFE_word_counts <- iLIFE_words_filtered %>% count(word, sort = TRUE)
wordcloud2(iLIFE_word_counts[1:300, ], size = .5)
```

#Word Cloud for iRobot Roomba 650

```
#iRobot Roomba 650
names(review_df650)
```

```
## [1] "Date"    "Product" "Stars"   "Title"   "Review"  "id"
```

```
# select only the comment and song titles from the data and examine it
iRobot650<-review_df650 %>% select(comment = Review, X = id)
glimpse(iRobot650)
```

```
## Observations: 633
## Variables: 2
## $ comment <chr> "You would not believe how well this works", "You just...
## $ X       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,...
```

```
# delete all undesirable words, here we only delete things that may bias analyses
# adjust this list as you need it, basically eliminate all undesirable words
undesirable_words_iRobot650 <- c("iRobot", "roomba", "robot", "theres",
                                 "floor", "wanna", "gonna", "what", "gotta", "make", "vacuum",
                                 "then", "those", "when", "him", "how", "whether", "as")

# check out a small sample of stop words, randomly
head(sample(stop_words$word, 15), 15)
```

```
##  [1] "moreover" "myself"   "right"    "new"      "most"     "near"
##  [7] "younger"  "with"     "himself"  "for"      "on"       "my"
## [13] "saying"   "th"       "still"
```

```
# unnest the comment, remove all stop and undesirable words and words smaller than 3 characters and examine
the result
# unnest and remove stop, undesirable and short words
iRobot650_words_filtered <- iRobot650%>%
  unnest_tokens(word, comment) %>%
  anti_join(stop_words) %>%
  distinct() %>%
  filter(!word %in% undesirable_words_iRobot650) %>%
  filter(nchar(word) > 3)
```

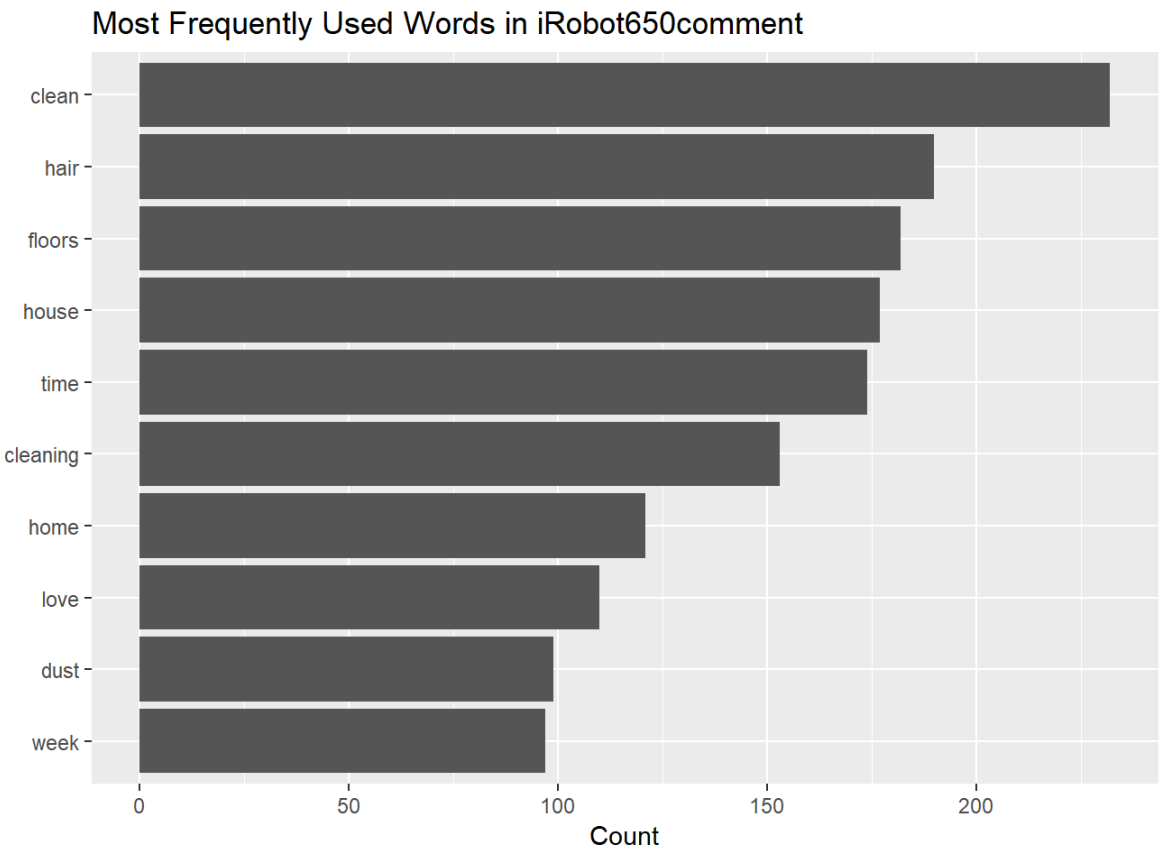
```
## Joining, by = "word"
```

```
dim(iRobot650_words_filtered)
```

```
## [1] 17225     2
```

```
# get the full word count from the comment and quickly examine the results
full_word_count_iRobot650 <- iRobot650%>%
  unnest_tokens(word, comment) %>%
  group_by(X) %>%
  summarise(num_words = n()) %>%
  arrange(desc(num_words))

# plot the most commonly used words in the comment
iRobot650_words_filtered %>%
  count(word, sort = TRUE) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot() +
  geom_col(aes(word, n)) +
  xlab("") +
  ylab("Count") +
  ggtitle("Most Frequently Used Words in iRobot650comment") +
  coord_flip()
```

```
## Selecting by n
```




```
# create a cool wordcloud of the words in the comment
iRobot650_word_counts <- iRobot650_words_filtered %>% count(word, sort = TRUE)
wordcloud2(iRobot650_word_counts[1:300, ], size = .5)
```

#Word Cloud for iRobot Roomba 880

```
#iRobot Roomba 880
names(review_df880)
```

```
## [1] "Date"    "Product" "Stars"   "Title"   "Review"  "id"
```

```
# select only the comment and song titles from the data and examine it
iRobot880<-review_df880 %>% select(comment = Review, X = id)
glimpse(iRobot880)
```

```
## Observations: 1,200
## Variables: 2
## $ comment <chr> "You think your floors are clean?Ê They're probably no...
## $ X       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,...
```

```
# delete all undesirable words, here we only delete things that may bias analyses
# adjust this list as you need it, basically eliminate all undesirable words
undesirable_words_iRobot880 <- c("iRobot", "roomba", "robot", "theres",
                                 "floor", "wanna", "gonna", "what", "gotta", "make", "vacuum",
                                 "then", "those", "when", "him", "how", "whether", "as")

# check out a small sample of stop words, randomly
head(sample(stop_words$word, 15), 15)
```

```
##  [1] "room"    "i'll"    "last"    "came"    "mostly"  "his"     "see"
##  [8] "further" "five"    "anyways" "ordered" "too"     "alone"   "what"
## [15] "work"
```

```
# unnest the comment, remove all stop and undesirable words and words smaller than 3 characters and examine the result
# unnest and remove stop, undesirable and short words
iRobot880_words_filtered <- iRobot880%>%
  unnest_tokens(word, comment) %>%
  anti_join(stop_words) %>%
  distinct() %>%
  filter(!word %in% undesirable_words_iRobot880) %>%
  filter(nchar(word) > 3)
```

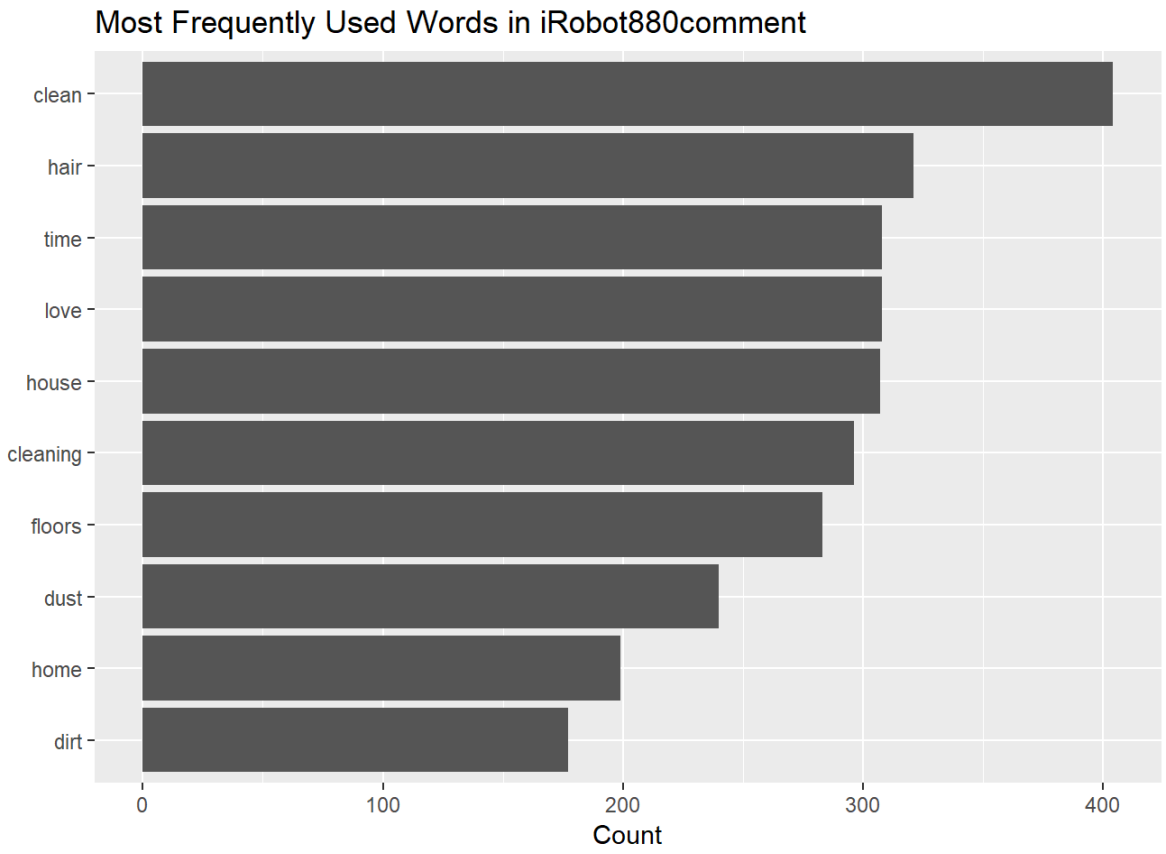
```
## Joining, by = "word"
```

```
dim(iRobot880_words_filtered)
```

```
## [1] 37842     2
```

```
# get the full word count from the comment and quickly examine the results
full_word_count_iRobot880 <- iRobot880%>%
  unnest_tokens(word, comment) %>%
  group_by(X) %>%
  summarise(num_words = n()) %>%
  arrange(desc(num_words))

# plot the most commonly used words in the comment
iRobot880_words_filtered %>%
  count(word, sort = TRUE) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot() +
  geom_col(aes(word, n)) +
  xlab("") +
  ylab("Count") +
  ggtitle("Most Frequently Used Words in iRobot880comment") +
  coord_flip()
```

```
## Selecting by n
```




```
# create a cool wordcloud of the words in the comment
iRobot880_word_counts <- iRobot880_words_filtered %>% count(word, sort = TRUE)
wordcloud2(iRobot880_word_counts[1:300, ], size = .5)
```

The plots that I got above are word cloud for these three products. They shows the frequency of words appearing in customers' reviews. As we can see, words appearing most frequently are basically positive, such as clean, love, recommend. There are also some difference among them. The word "price" appeared a lot in the reviews of iLIFE, which reflected that it is cheaper than other robot vacuum including iRobot 650 and 880. And the word " dog" and "pet" appeared more in the reviews of iRobot 650 and 880 because their main characteristic which is mentioned in their title is for pet.

Step 3

```
# Install and load the required packages.
pacman::p_load(dplyr, ggplot2, stringr, udpipe, lattice)
```

#iLIFE V5s

```
head(reviews_iLIFE)
```

```
##   X
## 1 1
## 2 2
## 3 3
## 4 4
## 5 5
## 6 6
##
prod
## 1 ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping Scrubbing Mopping Floor
Cleaning Robot
## 2 ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping Scrubbing Mopping Floor
Cleaning Robot
## 3 ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping Scrubbing Mopping Floor
Cleaning Robot
## 4 ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping Scrubbing Mopping Floor
Cleaning Robot
## 5 ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping Scrubbing Mopping Floor
Cleaning Robot
## 6 ILIFE V5s Pro Robot Vacuum Mop Cleaner with Water Tank, Automatically Sweeping Scrubbing Mopping Floor
Cleaning Robot
##                                          title           author
## 1                  Our vacuum committed suicide             Lizz
## 2 this has been a fun mission, to say the least       Jami James
## 3                  Promising, but disappointing     Review Panda
## 4                     Best Cleaning Robot EVER!           Adrian
## 5                     Vacuums well, mops poorly Kypros Nighthawk
## 6         ILIFE V5s is Great - See Video Review            Dylan
##               date                  review_format stars
## 1    July 29, 2018 Size: V5s ProVerified Purchase     1
## 2   April 22, 2018 Size: V5s ProVerified Purchase     1
## 3    June 23, 2017     Size: V5sVerified Purchase     2
## 4     May 10, 2016                      Size: V5s     5
## 5     May 16, 2018 Size: V5s ProVerified Purchase     3
## 6 January 10, 2017     Size: V5sVerified Purchase     5
##
comments
## 1
I wish that I could say that everything was going fine until Rosie, our vacuuming unit, hurled herself down
the stairs, but I feel I must be truthful. In reality, Rosie missed large swaths of the rooms she had been
assigned to, as well as seeming to obsess over areas she had cleaned several times already.Although the des
criptions here imply that she needs no supervision, she required our assistance in scaling even the smalles
t hurdles (the sloping 1/2" barriers that divide our rooms derailed her completely) and needed to be picked
up and carried over them.The final blow came when, in the middle of a routine sweep of the living room, she
reached the edge of the stairs. Normally she would come just over the edge and stop, beep and retreat, but
on this fateful evening Rosie had other plans. Without so much as a glance backward at us, she sailed over
the lip of the top step and crashed down all 17, landing with a rattle at the bottom. We rushed to her aid
but alas, peered over the banister only to find her in pieces by the front door, one wheel still spinning i
dly.In addition to the trauma caused by losing a valued member of our household, we find ourselves up to ou
r ankles in furballs and dustbunnies. Rosie, you are gone, but your careless legacy lives on.
## 2
Ok, this has been a fun mission, to say the least. I have been on a journey to find the best robot vacuum t
hey make and I think I'm making headway. So far I have tried The Shark 720, Ecovacs N79S, iLife V5, iRobot
Jet 240.The iLife V5-Noise Level: 3rd loudestFunctionality: TERRIBLECleaning Duration: cleaned for almost 1
00 minutesEmptying Convenience: TERRIBLENot worth it at all, had it for 1 hour and sent it backThe Ecovac N
79S-Noise Level: This was a tad bit louder than the shark, but not by muchFunctionality: cleaned well but w
ith no kind of plan, just zigzagged here and there, picked up more than the shark.Cleaning Duration: cleane
d for almost 2 hoursEmptying Convenience: easy to empty had a back-end tray that slid out easily.Amazing Va
```

lue for the cost - may end up with this oneThe iClebo Omega-Noise Level: is the loudest so far, sounds like you are using a regular vacuum cleaner for 80 minutes, that's because of the amount of pick up power this has.Functionality: Cleaned on a mission and was very strong, great mapping technology, picks up the bigger stuff more than all of the bots so far.Cleaning Duration: cleaned for 80 minutesEmptying Convenience: personally, I do not like the design in this one, tray stores in the top of the robot tray seems a lot small than all of the rest of them.Mop future is uselessOverpriced - Sending back for its little brotheriRobot Jet 240 Noise Level: pretty quiteFunctionality: Cleaned on a mission but does not do what I felt it shouldCleaning Duration: cleaned for 100 minutesEmptying Convenience: the process was fine but it will never outdo a regular mop. Cleans a quarter of the floor and then is empty. Felt like I had to babysit the thing all the time. Mop future is uselessWould not recommend to anyone.OverpricedThe Shark 720-Noise Level: This was the quietest so farFunctionality: cleaned well but with no kind of plan, just zigzagged here and there. Did not fin under the couch as good as the Ecovac.Cleaning Duration: cleaned for almost 2 hoursEmptying Convenience: easy to empty had a back-end tray that slid out easily.Great Value
3
I read a lot about robot vacuums and was swayed by the Amazon reviews on the iLife V5s. Unfortunately, my experience has been less than stellar and I'm considering returning this and trying out another brand. There are a couple of things you should consider before purchasing this vacuum:* The vacuum does not follow any kind of logical route that means that it will go over some areas multiple times and others only once. If you vacuum a big room you'll have dirty areas left over, which is a big minus. If you vacuum a small room then it will do OK.* This vacuum gets stuck very easily. I ran it 3-4 times before I figured out what furniture/shoes/chairs and other items I had to move to another room for the vacuum to not get stuck. Even then it sometimes would just randomly get stuck on the wall (see video). Once it gets stuck, it doesn't realize it, so prepare for it to spin in place until it runs out of battery.* The "plan" feature on the vacuum is unreliable. I always schedule the vacuum to do its work at night. However, half the time, fully charged, it never actually turns on.* The dirt bin is really small and once it fills up then it will stop picking things up. Not a big deal breaker, but annoying.* Don't buy this for the mop. The pad barely gets wet as it moves around and basically doesn't actually clean.I reached out to customer support and it took a week to get a response, which wasn't super helpful.
4 === Original Review Written May 10, 2016 === (An update can be found below)This is the best cleaning robot ever. I don't personally own a Roomba but have seen them work. This thing is almost identical, but at the fraction of the cost. This thing is definitely worth it! It vacuums and it mops (with some adjustments). This review is based off of non-carpeted flooring only!I have the robot vacuum work about once a day in my medium sized room and it will always have in its canister a bunch of hair and at least 2 2-3 inch in diameter dust bunnies. It tends to get wedged under arch type furniture (near the ground of course). The robot seems to be afraid of its charging base; it won't go near it (approximately 0.5-1 meters / 2-3 feet), which kinda sucks for me cause it fits perfectly under my desk with the charging base, but now he (yes it is a he!) won't clean under except when he leaves the base. All in all, his vacuuming skills are pretty good! He'll also return to his base with no problems unless he gets stuck on something.I had it mop today, and it behaves practically the same as vacuuming except it is much quieter and works kinda funny. You essentially fill up the water tank and (the instructions say only add water) I added a little Mr. Clean because it feels cleaner and smells better. It mops almost identically to how it vacuums.Other thoughts:- Dang good for about $140 (can easily find it for that price at other retailers if it isn't that much on Amazon).- Mops decently. The mop can also be taken off and thrown in the wash if you like it clean clean. Might need a steam mop to follow this puppy when he's done.- Vacuums like crazy! It's dang good.- He can push heavy objects. Probably up to 30-40 pound objects if he gets stuck or wedged in them. He got wedged under a portable AC and he moved it a good 2-4 inches to free himself.- Seems to vacuum for about 2-3 hours on max on a full charge.- I usually put the charging base on top of something once in a while so that he will vacuum everywhere, since he doesn't vacuum near the base. He will usually roam around for about 30-60 minutes when he's done vacuuming looking for the base, he will eventually find it no problem unless he gets stuck on something.- Never put the charging base near an intersection of a pathway, otherwise he will never cross it and be stuck in one quadrant while vacuuming.===== Update: August 29, 2016 ===== (the last paragraph is a tl;dr)So, how has the ILIFE V5s working for me after about 3.5 months now. I have to say that it has its quirks (will talk about it below) when compared to a Roomba. Overall, I'd still rate the ILIFE V5s a 5/5. Why?- It cleans my floor regularly. Has worked predictably and reliably since the day I got it.- The replacement parts can be bought affordably from Amazon. (4 brushes, 2 filters, and 2 mop pads all for $14.99 at the time of this review).- It was much cheaper than a Roomba. I think the cheapest Roomba on sale can be bought at about $400 (at the time of this review). I picked the ILIFE V5s up from a Chinese e-retailer for about $140 (after shipping).- I can send the V5s to vacuum and forget it. Usually, before I leave my room, I set it and forget i

t. When I come back, 9/10 times I'll find it on the charging base. Usually it's user error that'll prevent him from getting back to the base (i.e., A cable is lying on the floor and he gets stuck on it, or I block the base with furniture).Some comments:- Navigation: I know some of my friends will say a Roomba navigates with intelligence (i.e., it can navigate your house and it kind of memorizes it), the V5s, seems to work wi th heuristics. That is, it tends to roam the room a little first, then clean the outside, roam, outside, ro am etc. until the battery dies. It works effectively but if it is charged with cleaning a WHOLE floor, I th ink it might forget where its base is in a larger apartment / floor of a house.- Getting stuck: he only see ms to get stuck on arch type furniture. He doesn't get stuck easily except with mopping (see next comment). If he gets stuck he often will stay there and attempt to free himself a little, but then just let his batte ry die after that, if you don't rescue him in the next few hours. He will also repeatedly attempt to enter his base, but if there is something off to the side that triggers his bumper while he tries to park he'll r everse and attempt to park in the base repeatedly. Sometimes he will attempt to correct by going to the lef t or right but most the time he will fail to navigate around the object and will eventually die trying to r eturn to the base (reversing and trying to park until battery dies).- Mopping: it mops well and there isn't much to say about it. The only problem as the V5s ages when he mops is that his wheels seem to lose grip, s o when he mops on a wet floor, he gets stuck much more easily than normal. For instance, if he has mopped t he area and makes a second pass while the floor is still wet, he can get stuck on the wall because his whee ls don't provide him with enough grip to hit the wall and activate his bumper; thereby, getting stuck on th e wall (or furniture, or pretty much anything he hits gently).- Bumping things: he can hit objects quite ha rd if his proximity sensors don't pick it up. For example if he hits a chair at an angle (his proximity sen sors appear to be at about the 11 and 1 o'clock positions if the 12 is the path he is traveling forward i n), and the chair is at his 12, he will smack the chair leg with full force. While most furniture this wo n't matter but if you have delicate items or something heavy on a night stand that isn't against a wall whe n he hits it, it can knock over some things. This does pose as a potential risk but one will learn over tim e that the V5s can be left unattended (I don't recommend doing this till you stay nearby the first handful of times you send the V5s to work).tl;dr:I still love my V5s. I haven't had any battery problems or functio ning problems. Most issues I lined out are usually human error and normal wear and tear of the item. Few qu irks with getting him to work, but once you learn them the V5s is definitely worth the investment. I have s een other Chinese Roomba-like robots, and I have to say the V5s was definitely worth every penny. I've seen other Chinese Roomba-like robots go for about $100-120, and they are definitely not worth even half their p rice in headaches and how inefficient they are at vacuuming. So, I have to say despite the VERY MINOR quirk s I've had with the V5s, it is definitely worth it for $140. I would highly recommend this item and would r ecommend anyone who is considering a cheaper alternative to the Roomba pick this up.

5

I decided to get this guy more for the mopping feature than anything else, but I'll get to that in a bit an d start with the good first.PROS: The unit is fairly heavy which is good, it does a pretty good job of navi gating and transitions well from hard-wood floor to carpet/rugs. The auto-home feature is really nice to h ave since it automatically docks itself into the charger. It comes with normal and max vacuum mode and it picks up quite a good bit of dirt (It's no Dyson, so don't expect it to compete with an upright vac). It c ame with plenty of accessories as well as replacement parts. It also comes with a mopping feature as well as a reservoir for the mop. The remote control is a nice touch and allows you to program an auto-run timer as well as offers limited directional control.CONS: The mopping feature is pathetic. The tank constantly l eaks when filling it and when mounted in the machine, doesn't uniformly distribute the solution onto the mo pping pad so as it mops it leaves streaks on the floor because the middle is try and the edges are damp. T he AI is not very bright, it'll repeatedly go over the same areas multiple times while completely avoiding and missing other spots. Due to it's ability to only turn at 20degree angles or so, it zig-zags. Another annoying issue is in mop mode, it will still go over area rugs, which ends up causing the unit to deposit a ll the collected dirt/dust on the mop right onto the rug. I've also had the unit get stuck on raised (woo d) vent grills.THOUGHTS: The mopping issue, along with the water tank leaking all over the place is really disappointing, and I wish the AI knew not to go over carpets/rugs with the mop attached because it just mak es the carpet messy. I also wish it had a better way to map areas, such as a phone app or something that w ould let you limit the unit to certain rooms or distances so it'll actually clean a whole room, rather than wander around doing half a job.

6

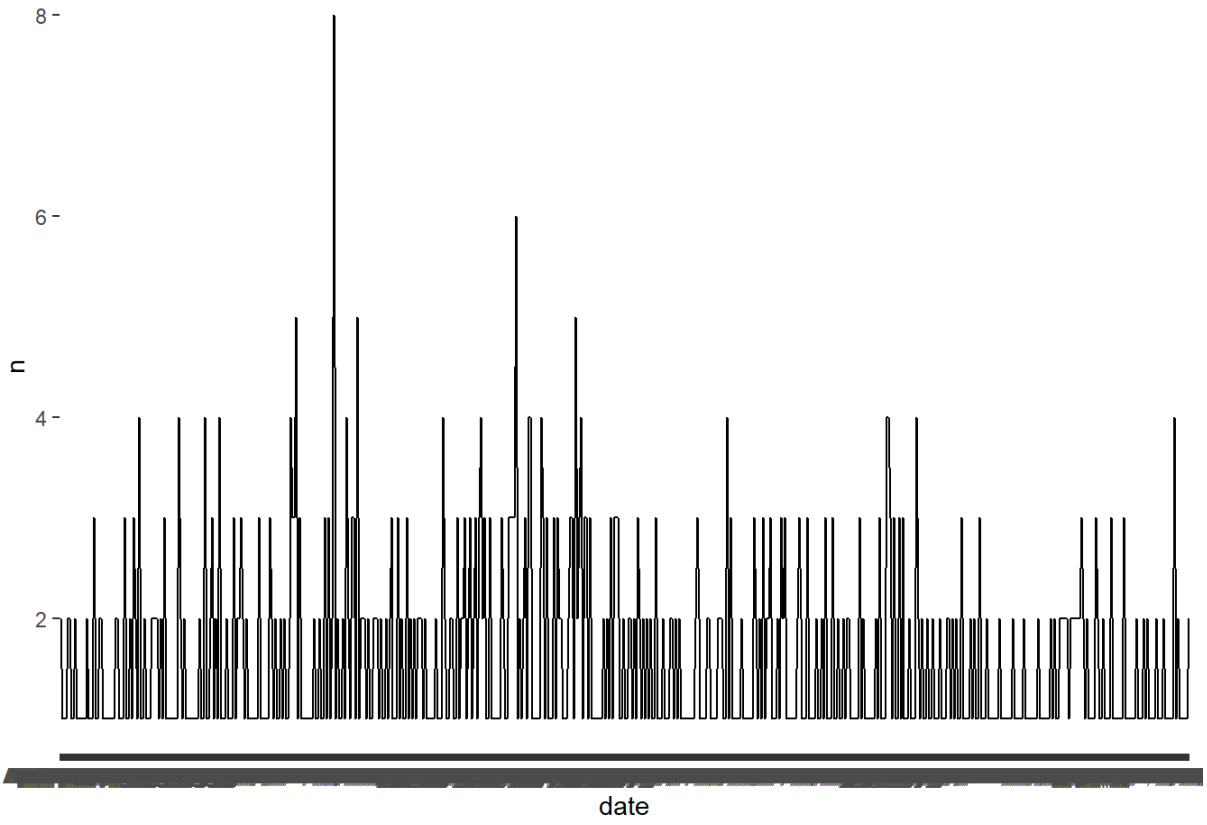
 The V5s robot vacuum by iLife is a fantastic product for a number of reasons: price, functionality, and co nvenience, just to name three.We like it so much, we even named this little guy. Even though his big brothe r might be more powerful, he does a good job on most surfaces. We have hardwood in most of our home, with t hrow rugs and tile. There are times he gets stuck on cords, but 99% of the time he does what he is supposed

```
to.We only tried the wet cycle once, and it seemed to work well, but our main purpose for this unit is to c
ut down on the amount of housework we need to do. You are not supposed to add soaps to the water tank, beca
use while it can still function with soap, it isn't going to clean the soap up unless you then run it 2-3x
over the same area with a fresh tank of water. Otherwise, your floor will be sticky. If you're looking for
a wet mop primarily (we weren't because we have hardwood floors), you probably need something like this rob
ot wet mop instead: http://amzn.to/2jbfFOCCons: He isn't always great at finding his charging station on hi
s own, but that isn't an issue for us, and it is probably our fault because we tucked it away so it is inco
nspicuous.We've told our friends about this more than once, and would definitely recommend the iLife Robot
Vacuum Cleaner. Feel free to ask me questions below...happy to help.
##   n_helpful
## 1       649
## 2       431
## 3       178
## 4       752
## 5        50
## 6        51
```

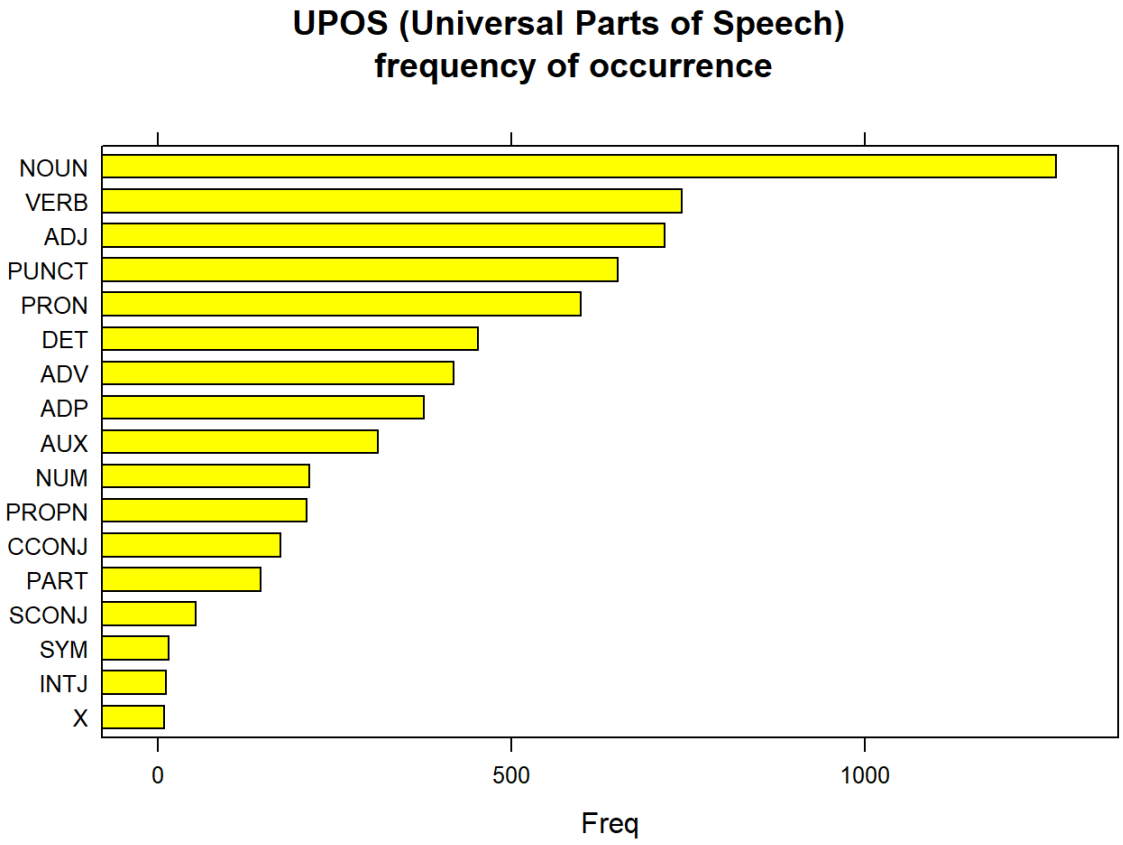
```
udmodel_english <- udpipe_load_model(file = "english-ewt-ud-2.3-181115.udpipe")

reviews_iLIFE %>% group_by(date) %>% count() %>% arrange(desc(n))
```

```
## # A tibble: 632 x 2
## # Groups:   date [632]
##    date                 n
##    <chr>            <int>
##  1 December 27, 2016     8
##  2 January 19, 2017      6
##  3 December 17, 2017     5
##  4 December 31, 2017     5
##  5 January 5, 2017       5
##  6 April 28, 2019        4
##  7 August 12, 2018       4
##  8 August 2, 2018        4
##  9 August 23, 2018       4
## 10 December 15, 2016     4
## # ... with 622 more rows
```

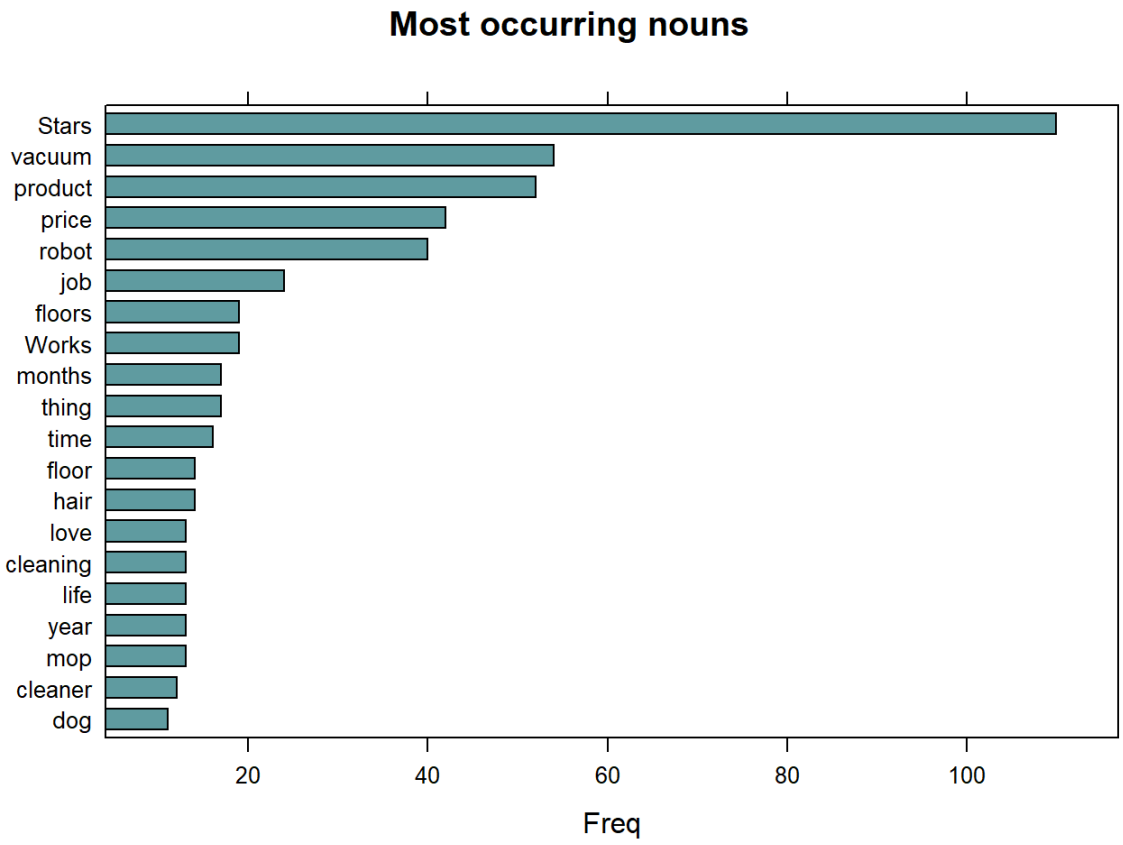
```
reviews_iLIFE %>% group_by(date) %>% count() %>% ggplot() + geom_line(aes(date,n, group = 1))
```



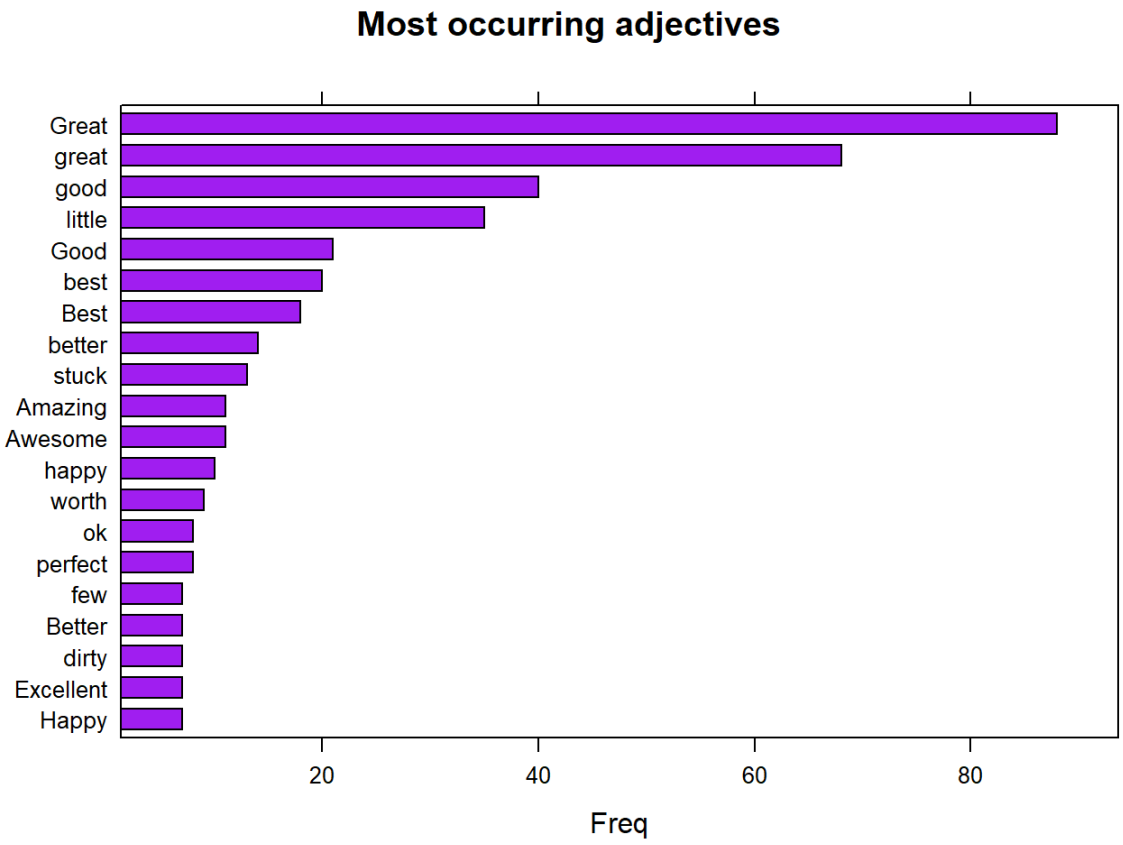

```
s <- udpipe_annotate(udmodel_english, reviews_iLIFE$title)
x <- data.frame(s)

stats <- txt_freq(x$upos)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = stats, col = "yellow",
         main = "UPOS (Universal Parts of Speech)\n frequency of occurrence",
         xlab = "Freq")
```



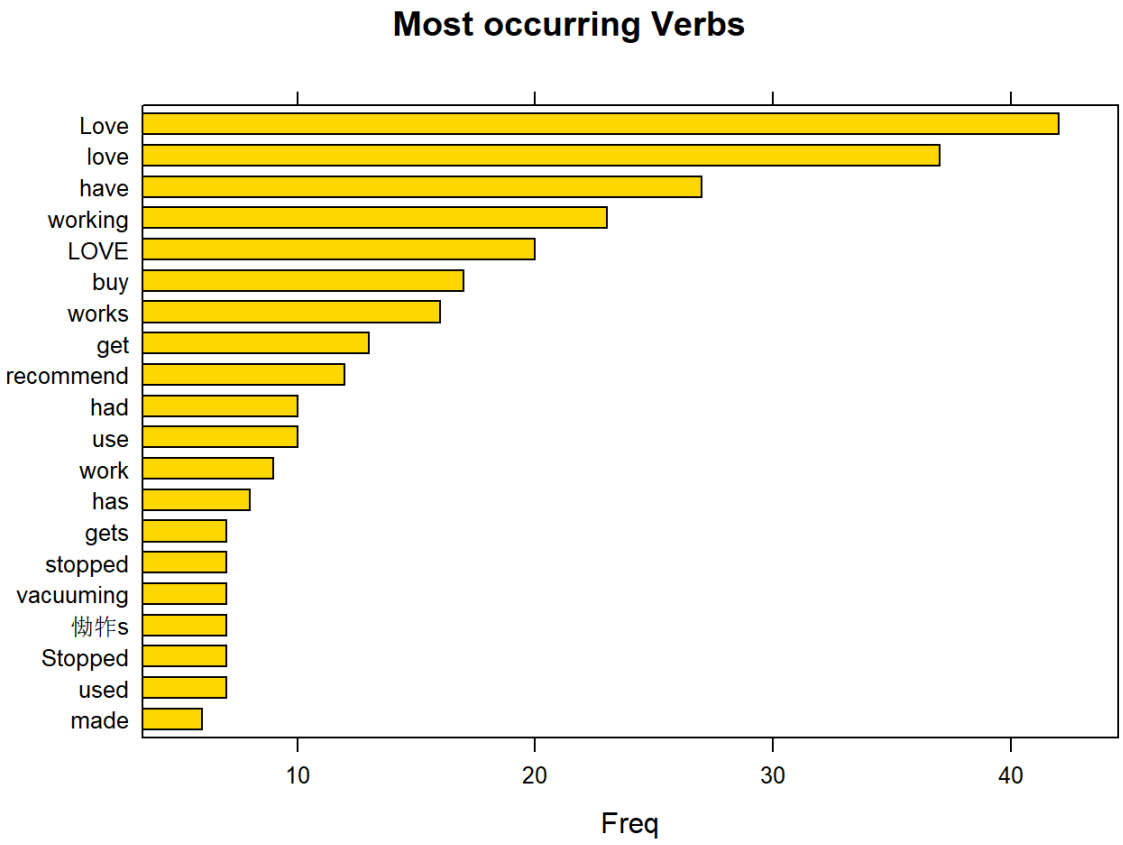

```
## NOUNS
stats <- subset(x, upos %in% c("NOUN"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "cadetblue",
         main = "Most occurring nouns", xlab = "Freq")
```



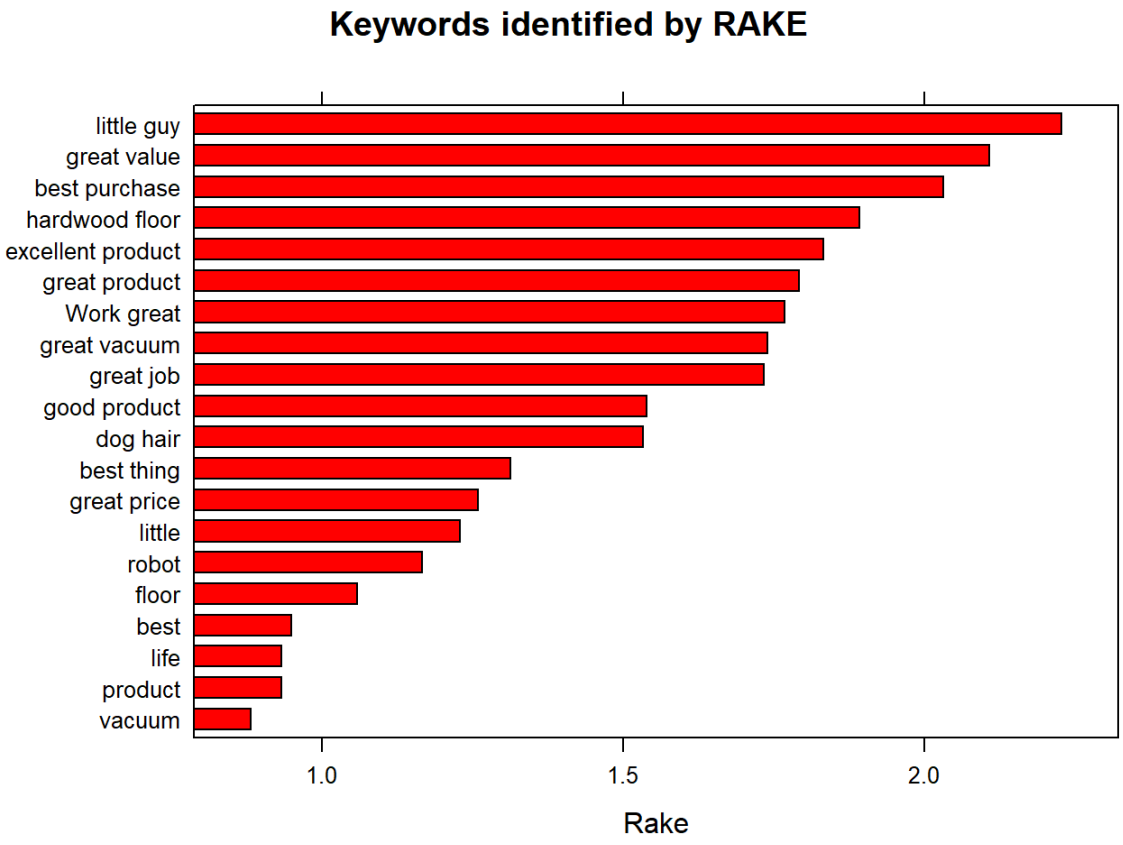

```
## ADJECTIVES
stats <- subset(x, upos %in% c("ADJ"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "purple",
         main = "Most occurring adjectives", xlab = "Freq")
```



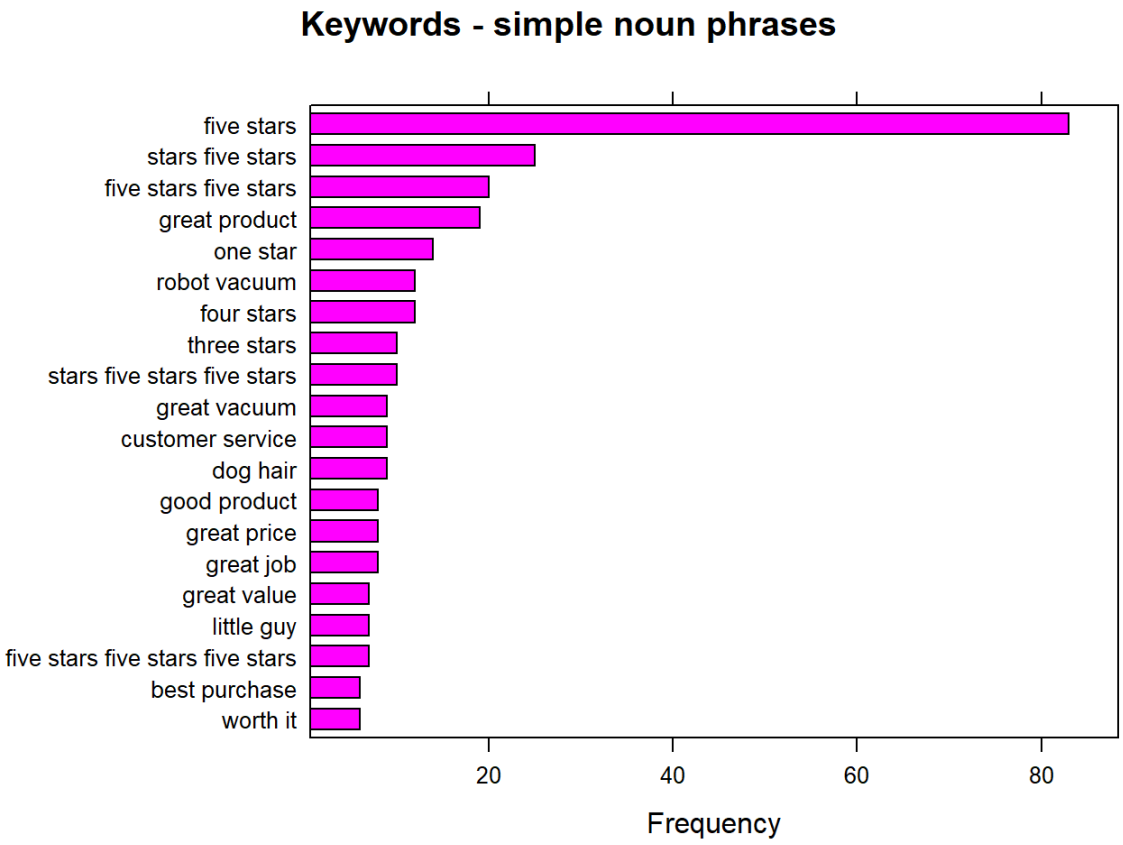

```
## VERBS
stats <- subset(x, upos %in% c("VERB"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "gold",
         main = "Most occurring Verbs", xlab = "Freq")
```




```
## RAKE
stats <- keywords_rake(x = x, term = "lemma", group = "doc_id",
                       relevant = x$upos %in% c("NOUN", "ADJ"))
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ rake, data = head(subset(stats, freq > 3), 20), col = "red",
         main = "Keywords identified by RAKE",
         xlab = "Rake")
```




```
## display by plot a sequence of POS tags (noun phrases / verb phrases)
x$phrase_tag <- as_phrasemachine(x$upos, type = "upos")
stats <- keywords_phrases(x = x$phrase_tag, term = tolower(x$token),
                          pattern = "(A|N)*N(P+D*(A|N)*N)*",
                          is_regex = TRUE, detailed = FALSE)
stats <- subset(stats, ngram > 1 & freq > 3)
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ freq, data = head(stats, 20), col = "magenta",
         main = "Keywords - simple noun phrases", xlab = "Frequency")
```



#iRobot Roomba 650

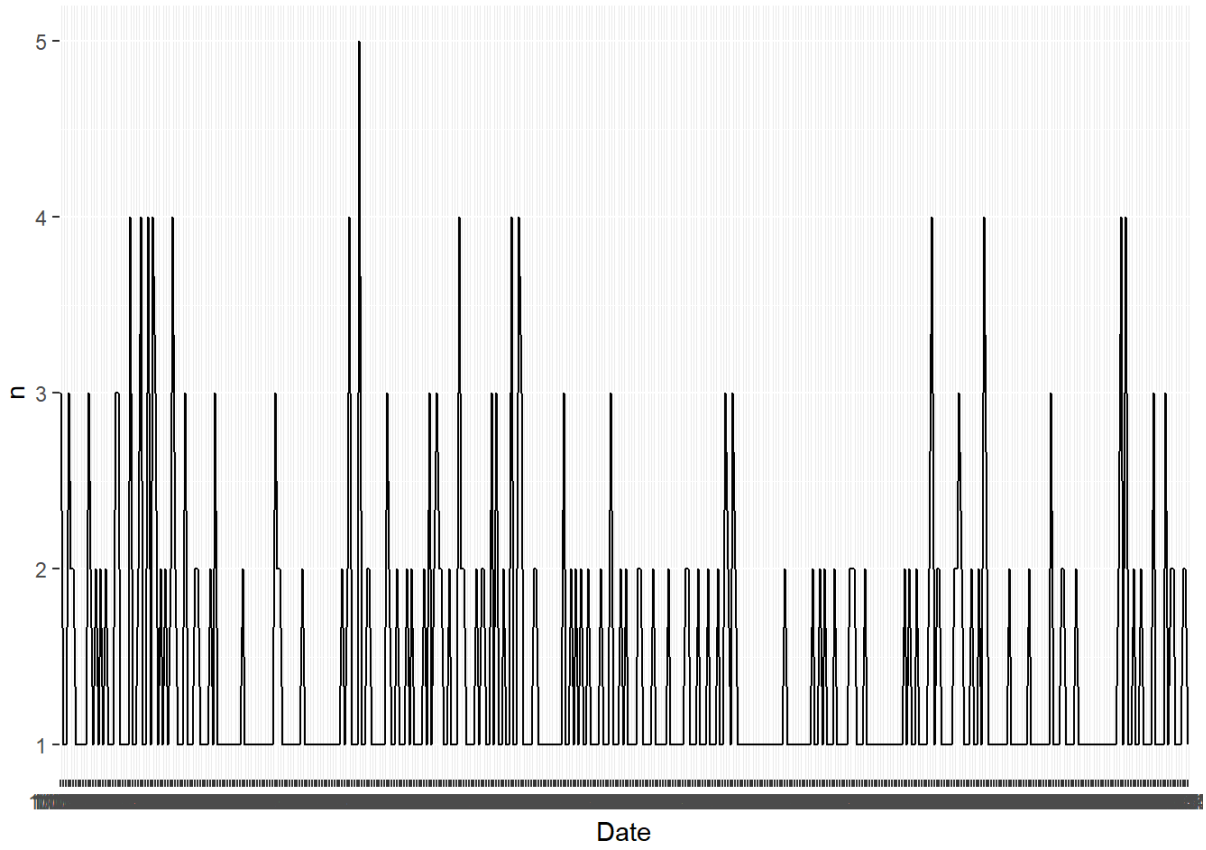
```
head(review_df650)
```

```
##       Date                    Product Stars                      Title
## 1  2/28/15 iRobot Roomba 650 for Pets     5                 Five Stars
## 2  1/12/15 iRobot Roomba 650 for Pets     4                 Four Stars
## 3 12/26/13 iRobot Roomba 650 for Pets     5           Awesome love it.
## 4   8/4/13 iRobot Roomba 650 for Pets     3     Love-hate this vaccuum
## 5 12/22/15 iRobot Roomba 650 for Pets     5 This vacuum is fantastic!!
## 6 12/27/15 iRobot Roomba 650 for Pets     5                       Wow!
##
Review
## 1
You would not believe how well this works
## 2
You just walk away and it does the rest
## 3
You have to Roomba proof your house but once that is done it is awesome. It's like having a pet that clean
s.
## 4
Yes, it's a fascinating, albeit expensive, bit of technology. That said, it is a pain in the butt to clean
- every time, if you have the model for pets. Cat hair will clog it in one small room. On the plus side, it
gets under the bed, where there could be more pet hair than I care to admit. Can two short haired cats shed
faster than Roomba can clean? Will my patience wear out, before my picks, and tweezers and screwdriver? Sta
y tuned...
## 5 Years ago I bought one of the original Roomba's and was underwhelmed - it was slow and got stuck in co
rners.  But I'm lazy at heart, so decided to try again - and this model is amazing!!  First - no problem in
corners - it spins and backs out.  (It also has a spinning brush on the side which whisks all those dust de
vils out of the corner into the vacuum).  Also, it goes easily from carpet to hardwood floor to doormat and
back.  If it encounters a cord on the floor, it backs up and comes back in from another angle.  And I know
it's weird to think about - how the heck does it figure out your floor plan, especially when it seems to go
in random directions?  It all gets down to math, my friends, and the engineers at Roomba worked it out.  We
use it for individual rooms (like our muddy utility room), and for entire house runs.  It's great if you ha
ve a dog (you wouldn't believe how much dog hair is under your couch).  Easy to clean, easy to set up and d
efinately easy to use.  (You push a button and walk away!)  Great buy.
## 6
Wow.Wow.  I never knew my floors were so dirty before this!  I've set my "Room by" to clean every day.  I a
m astounded each day when I empty the bin.  Where did all this come from??  I have 3 small dogs and can't b
elieve I've spent years living in what I would now consider gross based on what I dump out each day.  I hav
en't had to vacuum in weeks!  I am going to get this for my 70+ parents. Best invention ever!!!
##   id
## 1  1
## 2  2
## 3  3
## 4  4
## 5  5
## 6  6
```

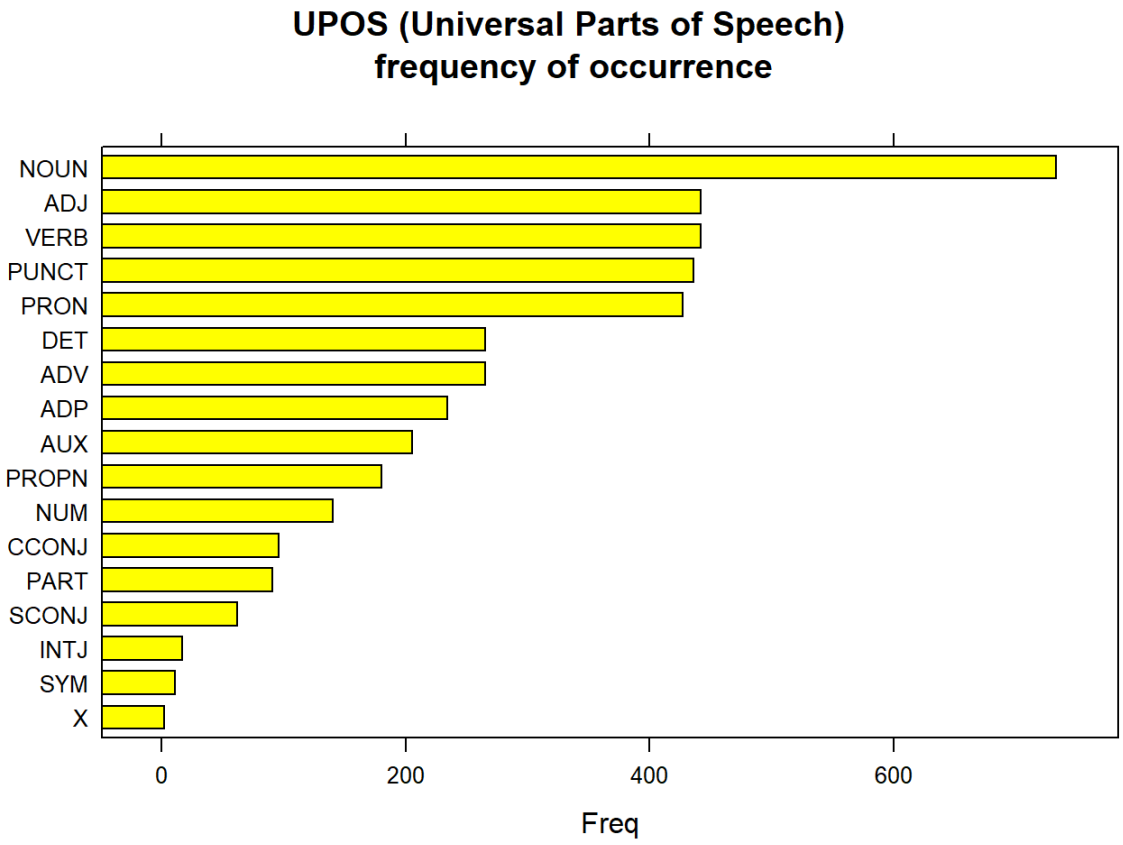
```
udmodel_english <- udpipe_load_model(file = "english-ewt-ud-2.3-181115.udpipe")

review_df650 %>% group_by(Date) %>% count() %>% arrange(desc(n))
```

```
## # A tibble: 454 x 2
## # Groups:   Date [454]
##    Date         n
##    <chr>    <int>
##  1 11/30/15     5
##  2 1/28/15      4
##  3 1/3/15       4
##  4 1/31/15      4
##  5 1/4/15       4
##  6 1/6/16       4
##  7 11/29/15     4
##  8 12/22/15     4
##  9 12/31/14     4
## 10 12/4/15      4
## # ... with 444 more rows
```

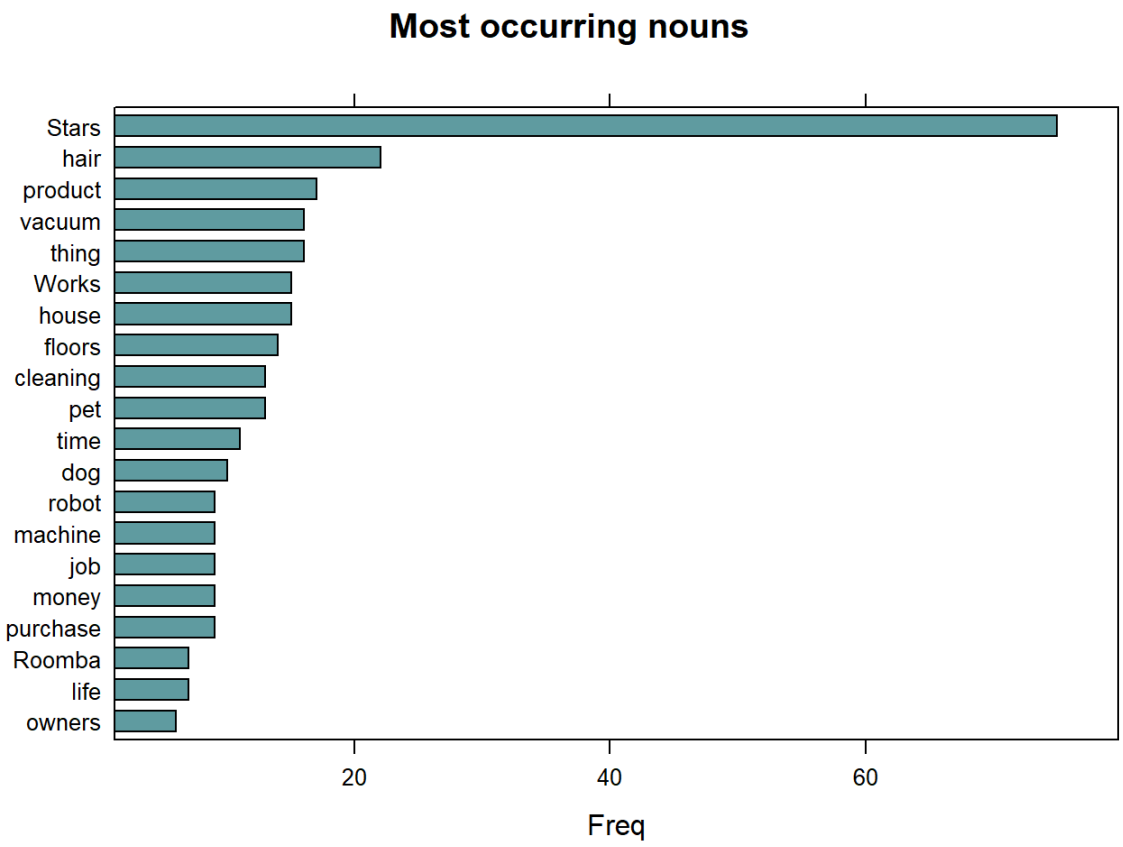
```
review_df650 %>% group_by(Date) %>% count() %>% ggplot() + geom_line(aes(Date,n, group = 1))
```



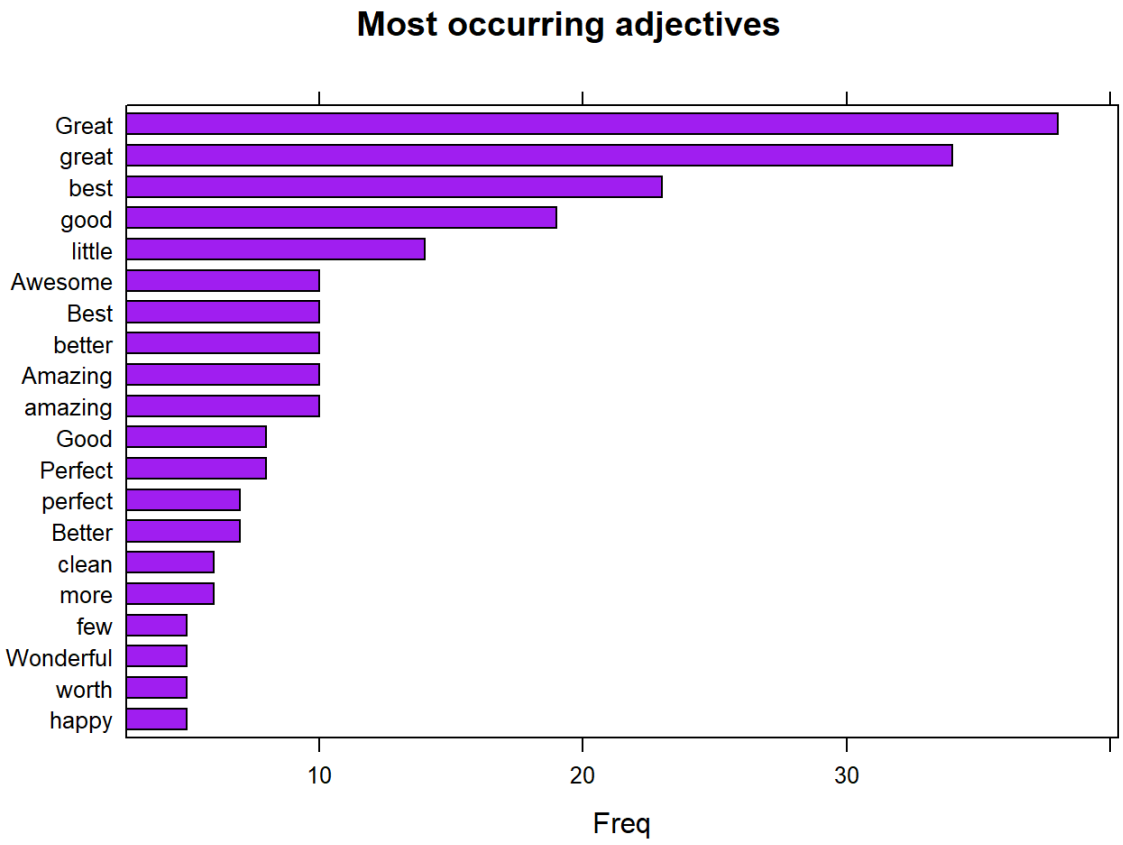

```
s <- udpipe_annotate(udmodel_english, review_df650$Title)
x <- data.frame(s)

stats <- txt_freq(x$upos)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = stats, col = "yellow",
         main = "UPOS (Universal Parts of Speech)\n frequency of occurrence",
         xlab = "Freq")
```



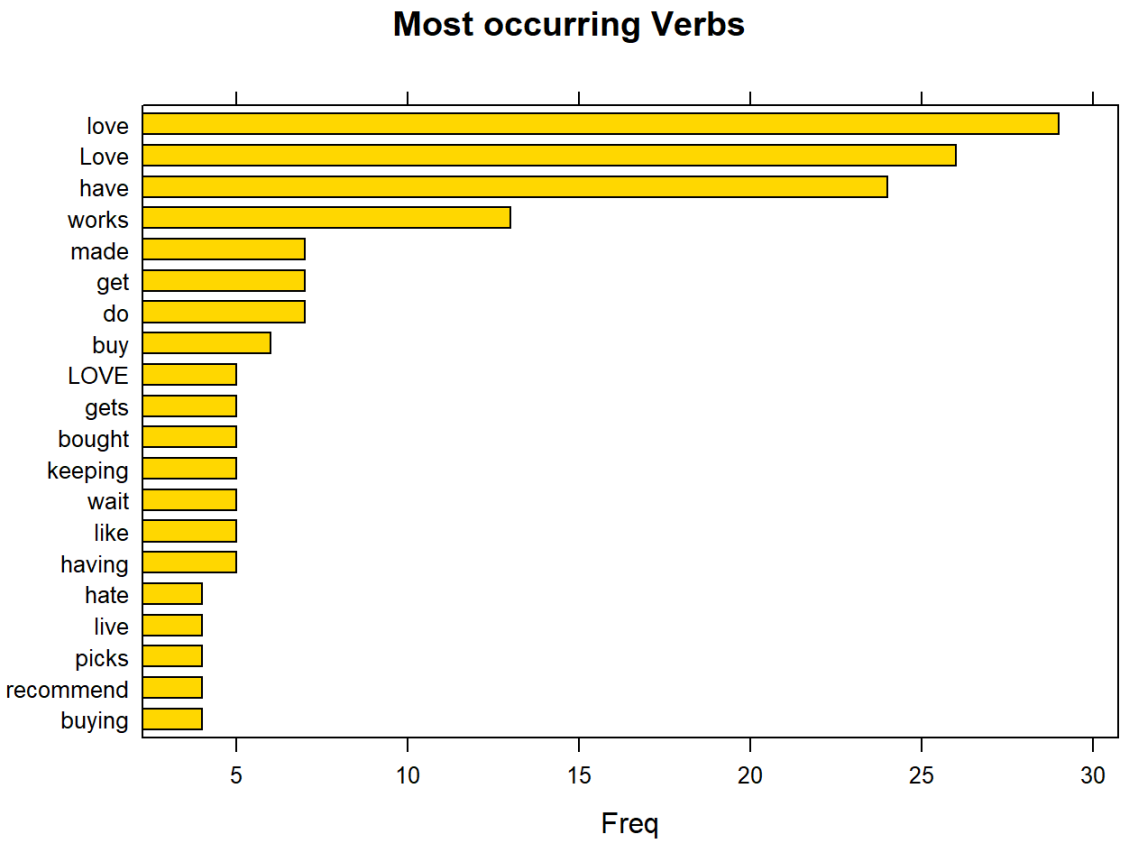

```
## NOUNS
stats <- subset(x, upos %in% c("NOUN"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "cadetblue",
         main = "Most occurring nouns", xlab = "Freq")
```



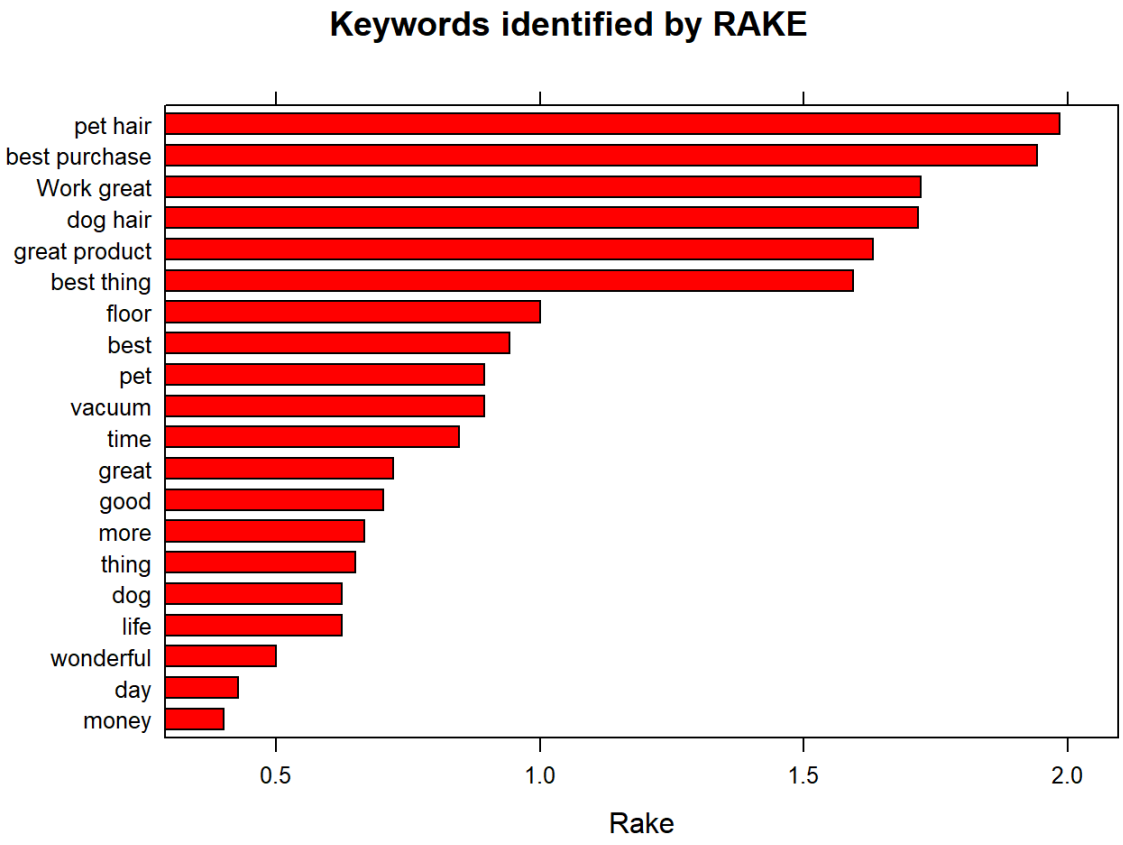

```
## ADJECTIVES
stats <- subset(x, upos %in% c("ADJ"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "purple",
         main = "Most occurring adjectives", xlab = "Freq")
```



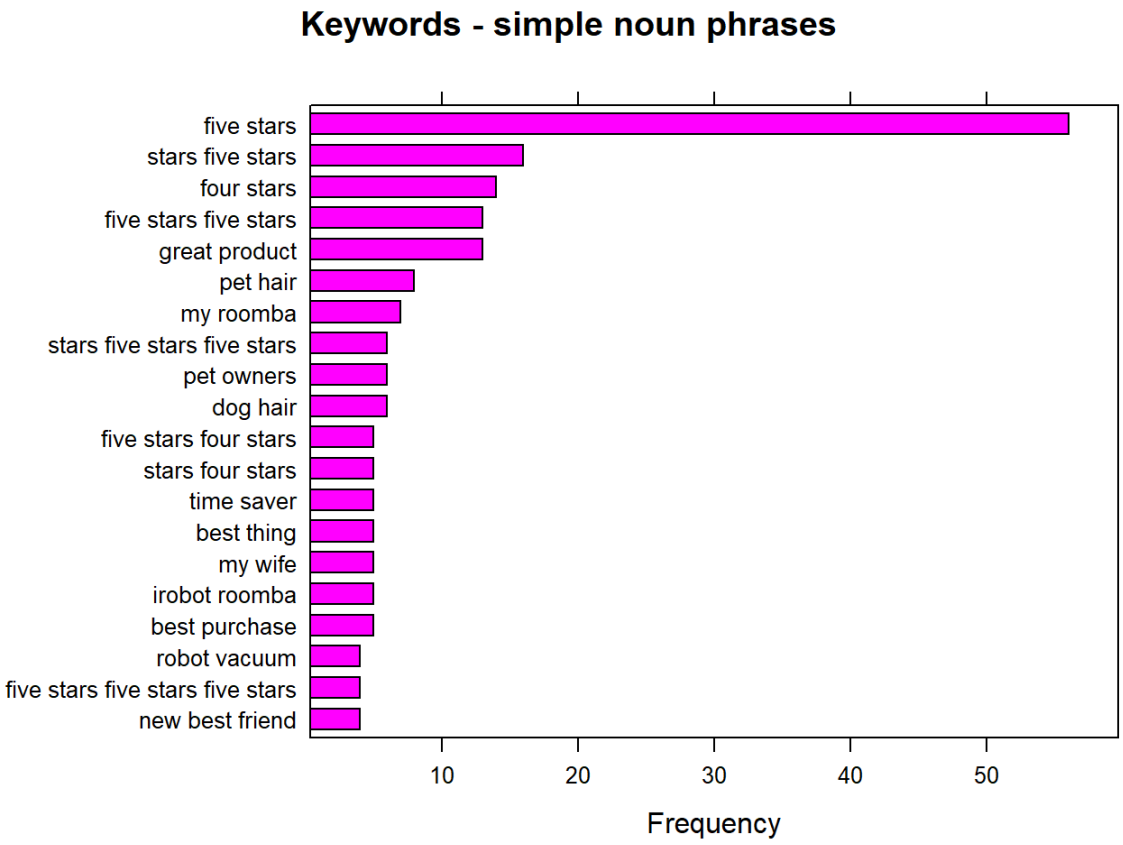

```
## VERBS
stats <- subset(x, upos %in% c("VERB"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "gold",
         main = "Most occurring Verbs", xlab = "Freq")
```




```
## RAKE
stats <- keywords_rake(x = x, term = "lemma", group = "doc_id",
                       relevant = x$upos %in% c("NOUN", "ADJ"))
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ rake, data = head(subset(stats, freq > 3), 20), col = "red",
         main = "Keywords identified by RAKE",
         xlab = "Rake")
```




```
## display by plot a sequence of POS tags (noun phrases / verb phrases)
x$phrase_tag <- as_phrasemachine(x$upos, type = "upos")
stats <- keywords_phrases(x = x$phrase_tag, term = tolower(x$token),
                          pattern = "(A|N)*N(P+D*(A|N)*N)*",
                          is_regex = TRUE, detailed = FALSE)
stats <- subset(stats, ngram > 1 & freq > 3)
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ freq, data = head(stats, 20), col = "magenta",
         main = "Keywords - simple noun phrases", xlab = "Frequency")
```



#iRobot Roomba 880

```
head(review_df880)
```

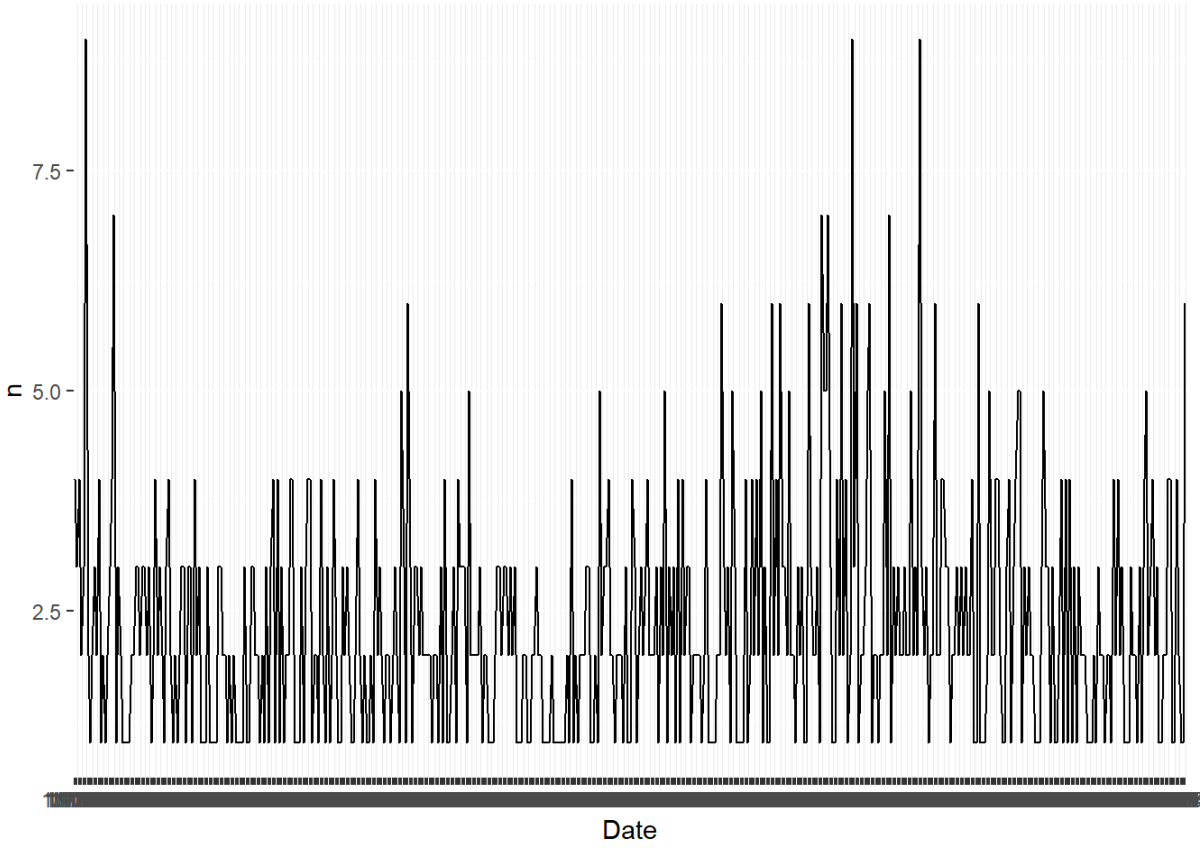
```
##       Date                                    Product Stars
## 1  8/10/15 iRobot Roomba 880 for Pets and Allergies     5
## 2   9/9/15 iRobot Roomba 880 for Pets and Allergies     5
## 3 11/18/15 iRobot Roomba 880 for Pets and Allergies     4
## 4 10/31/14 iRobot Roomba 880 for Pets and Allergies     5
## 5  6/16/14 iRobot Roomba 880 for Pets and Allergies     5
## 6   8/6/15 iRobot Roomba 880 for Pets and Allergies     5
##                                                                          Title
## 1                                                 Your house isn't "Roomba Clean."
## 2                                      Yes, it works.Ê It works really, really well.
## 3                                                              LOVE our roomba
## 4 My previous Roomba was working fine and I was reluctant to shell out the cash ...
## 5                                                  Great product, does a fine job.
## 6                                    I LOVE it! It is much more advanced in its ...
##
Review
## 1
You think your floors are clean?Ê They're probably not. The first time I unleashed by Roomba 880 on what I
thought were clean hardwood floors and large area rugs, I had to follow it around and continually empty the
bin and clean the filter. I was shocked. What appeared to be a spotless home simply was not. No wonder my a
llergies and sinuses were problematic.I scheduled it to run everyday for a week to vacuum a large open area
(kitchen, dining room and sitting room) and put a lighthouse to the master bedroom. The bin emptying became
less frequent. The hardwood glows. The carpets feel soft and new. Roomba even dusted the tops of the floor
based lights. My allergies are under control and I imagine my Flonase purchases will slow. It now vacuums e
very other day and I only empty the bin when it finishes and is docked.Advice. Shim your doors open. Roomba
is so intent on getting at dust along walls, it will push doors closed and imprison itself in a room.
## 2 Yes, it really does work.Ê We have dogs who shed tremendously, and running the Roomba a couple of time
s per week seems to keep up with the mess very well.Ê The instructions recommend running the Roomba several
times in succession over the same area to get embedded dirt and hair out, as the bin will fill up quickly.Ê
Once we did that, we found that the Roomba could cover a large zone in the house before needing to return f
or a charge and to have its bin emptied.The extraction is excellent, even compared to my Hoover upright.Ê T
he longer I use the Roomba, the cleaner and brighter my rugs look and feel.Ê It's also rather fun to watch
it bumbling around the room, like a bumblebee in a bottle.Ê It does a surprisingly good job of getting unde
r furniture and into tight corners.We have a house with a semi-open floor plan.Ê It took a little figuring,
but it was fairly easy to figure out a layout that divided the house into zones that were within the capaci
ty of the battery and bin.Ê The Virtual Wall and Lighthouse units confine the robot into these zones, along
with closing some doors. I find it easiest to run the Roomba in a zone, then empty the bin and let it recha
rge before moving on to the next zone.Ê Emptying the bin and replacing the HEPA filter are both easy tasks.
It takes me about ten minutes to pick up things like dog toys, shoes, small scatter rugs and to move chairs
the Roomba can't get under.Ê Then I turn it loose and go work on something else.Ê It tends to rumple up sma
ll rugs, but will ride right over larger rugs.Ê While it does not damage fringe, it does rumple it up a lo
t, so I don't run it over my expensive wool rugs. I do tuck electrical cords out of the way, but I have see
n it encounter them several times, and it does not tangle itself in them.Ê I would be picking up and moving
all of these things in order to vacuum anyway, so using the Roomba does not add any work.If you have a lot
of open doorways, you may find it useful to get a third Lighthouse.Ê I was very confused about how to turn
them off and on, until I realized that they turn themselves on and off when the Roomba is not active.Ê My i
nstruction manual said they had an on/off switch - the later models don't have it.I have not used the sched
ule feature.Ê Our house is large enough that the Roomba and its charging station have to be moved from zone
to zone.Ê Plus, with pets in the house I would prefer to run it when I am there to supervise it.Ê It has go
tten itself stuck twice in the two weeks I have owned it.All other good things being said, it is not a subs
titute for a good upright with cleaning tools.Ê It can't be used on drapes or upholstery, or to clean cobwe
bs from the ceiling, and it certainly won't mop the kitchen floor (that's another IRobot product).Ê It is a
huge help in keeping up with the day to day cleaning, and keeping the dust and pollen down.Ê My husband and
one of our dogs have dust allergies, and they are both doing better since we started using the Roomba regul
arly.
## 3
Yes to everything everyone else said. LOVEÊ our roomba. But- be forewarned, if the battery dies- you have t
o reset all your programming! Completely ridiculous out of a $600 piece of equipment. It can't keep the clo
```

```
ck set if it can't find its way back to the docking station. Customer service says I'm trying to clean too
large of an area and not using my virtual walls (well my toddler runs off with them!). So, until my kid all
grow out of toddler age, I will reset the entire schedule probably on a weekly basis. Ugh!
## 4
Yes there is a difference!Ê I have had every model of Roomba since the very first model came out.Ê My previ
ous Roomba was working fine and I was reluctant to shell out the cash for the newest model, but I just had
to!Ê It really has a lot more sucking power, and it is nice to not have to clean all that hair off brushes.
Ê If you been wondering if there is a huge difference between this your previous Roombas, trust me, there i
s!Update, we dirtied 2 spots on the floor, a 2000 dollar boat anchor Kirby vs. Roomba.Ê Would you believe R
oomba did as good if not better a job? When we figure out a way to do a more scientific test of these 2 wil
l post a video.Ê The Kirby has been relegated to the garage.Ê The old advisory w that a Roomba does not sub
stitute for deep cleaning upgright vacuum cleaners no longer applies.Ê In fact the New York Times I see sta
ted the same thing, cleans as well or better then uprights.
## 5
Yes the dust bin is a bit too small but what can you do, other than make the roomba gigantic (and this is a
lready pretty big).Ê That being said, this thing works great.Ê I even brought it into my work and had it va
cuum the office my boss and I are in and he was very impressed with this little guy (and he doesn't impress
easy).Ê The pathing on this is fine, but yes it will miss an area now and then but I have it cleaning three
main rooms plus 4 small areas (bathrooms, kitchen, etc)...I have a child and two dogs (small and medium), a
nd it's great.Ê They need to come up with a stair roomba next though.Now, a couple things...it has closed i
tself in the bathroom a few times now so I have to lock it out of areas where the doors aren't secured open
(magnetic door stops) when I'm not home, but that's hardly the devices fault.Ê I had to buy small risers fo
r my couch so it would go under there (the first day I ran it the sensor on the top wouldn't allow it under
but the roomba tried, scratched the heck out of the top of the front bumper too.)Ê I recommend picking up r
ugs off the floor if you can.Ê It doesn't have an issue specifically but it doesn't seem to be terribly kin
d on the edges of them, lighter rugs will probably get jammed up in it (maybe).
## 6
Years ago my son and daughter-in-law gave me one of the first roombas for Christmas.Ê It wore out several y
ears ago and I sorely missed it.Ê This year I ordered the Roomba 880 for myself.Ê I LOVE it!Ê It is much mo
re advanced in its cleaning power.Ê I didn't realize my floors could be so dirty and I don't even have pet
s!Ê I do have grandchildren though and they can be messy.Ê I did wait for the price to go down because I co
uld not afford the full price.Ê Great purchase and clean floors!
##   id
## 1  1
## 2  2
## 3  3
## 4  4
## 5  5
## 6  6
```

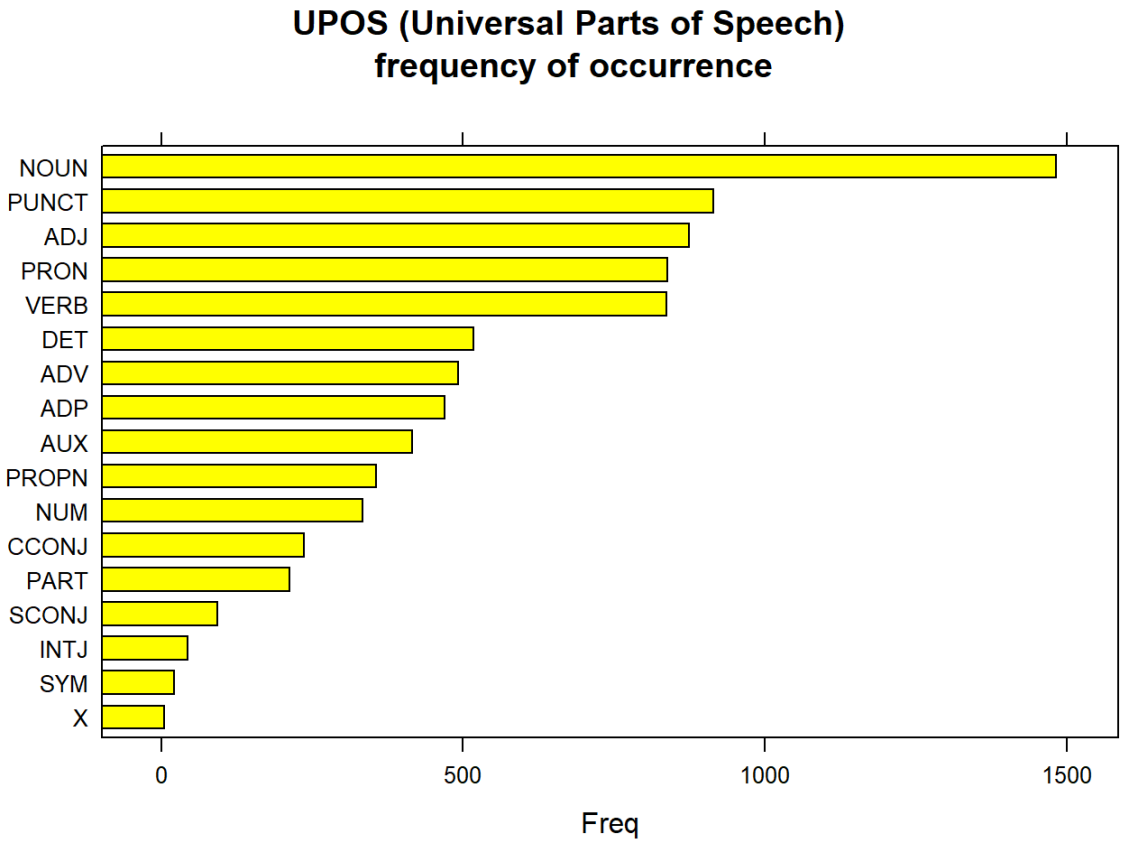
```
udmodel_english <- udpipe_load_model(file = "english-ewt-ud-2.3-181115.udpipe")

review_df880 %>% group_by(Date) %>% count() %>% arrange(desc(n))
```

```
## # A tibble: 511 x 2
## # Groups:   Date [511]
##    Date         n
##    <chr>    <int>
##  1 1/14/15      9
##  2 6/8/15       9
##  3 7/23/15      9
##  4 1/3/16       7
##  5 6/27/15      7
##  6 6/3/15       7
##  7 7/17/15      7
##  8 12/27/15     6
##  9 5/30/15      6
## 10 6/15/15      6
## # ... with 501 more rows
```

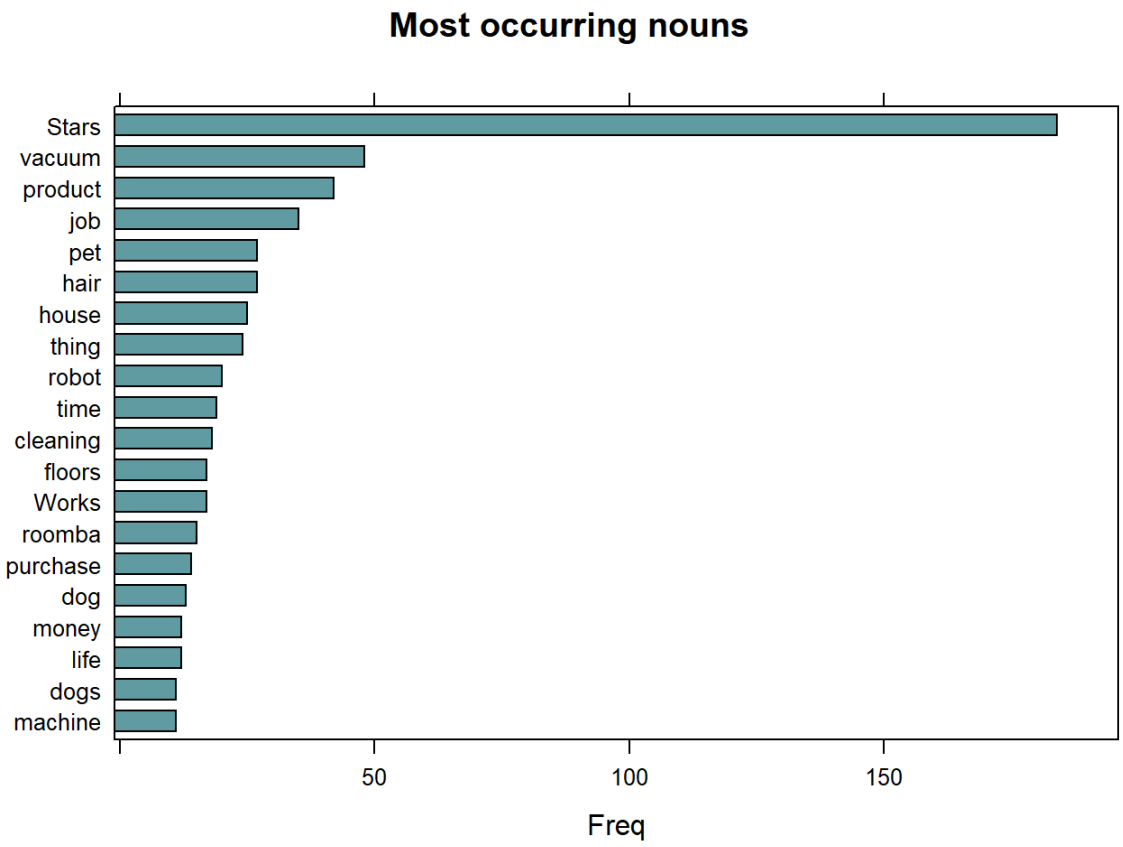
```
review_df880 %>% group_by(Date) %>% count() %>% ggplot() + geom_line(aes(Date,n, group = 1))
```



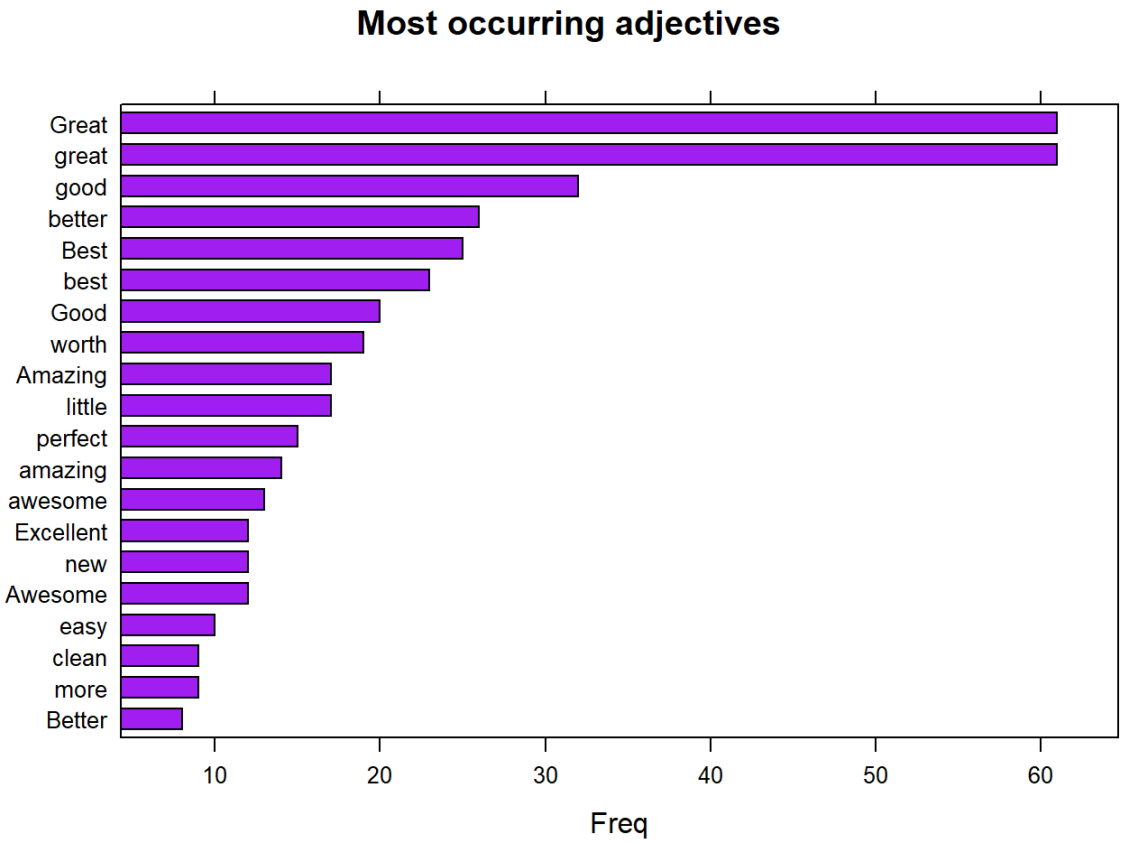

```
s <- udpipe_annotate(udmodel_english, review_df880$Title)
x <- data.frame(s)

stats <- txt_freq(x$upos)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = stats, col = "yellow",
         main = "UPOS (Universal Parts of Speech)\n frequency of occurrence",
         xlab = "Freq")
```



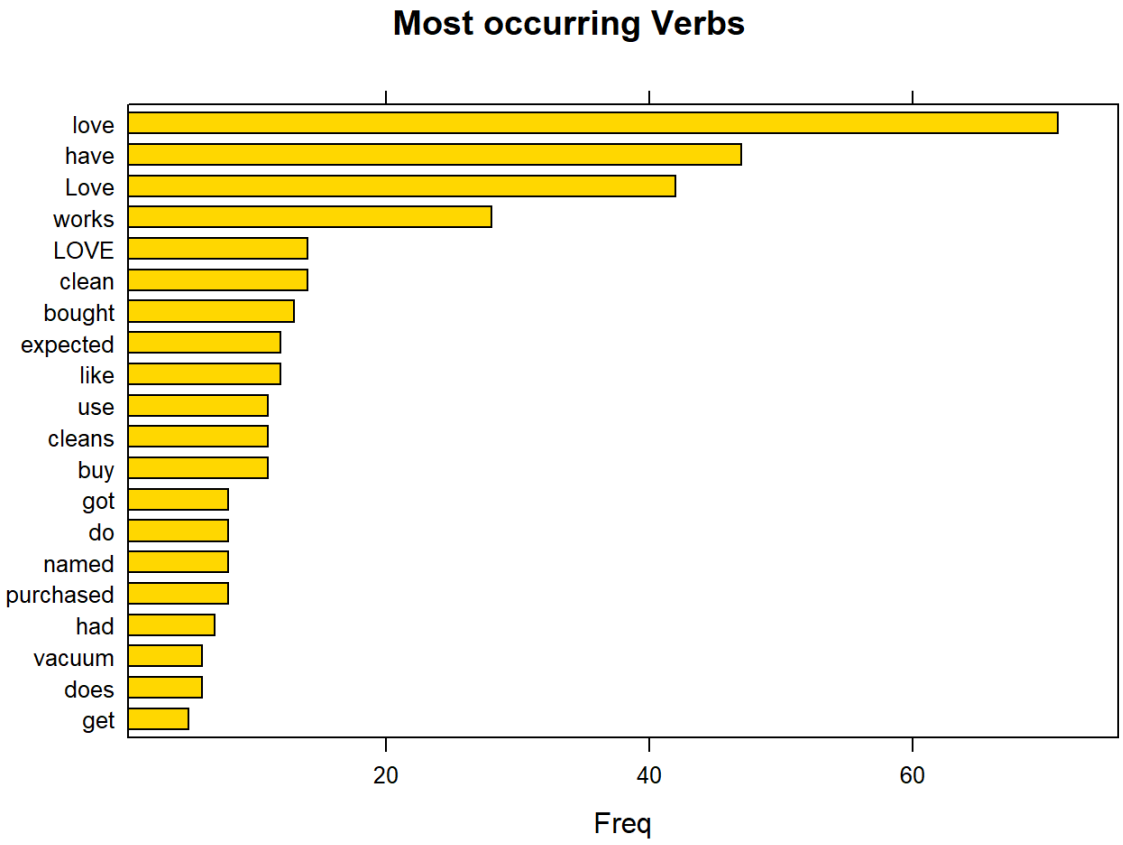

```
## NOUNS
stats <- subset(x, upos %in% c("NOUN"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "cadetblue",
         main = "Most occurring nouns", xlab = "Freq")
```



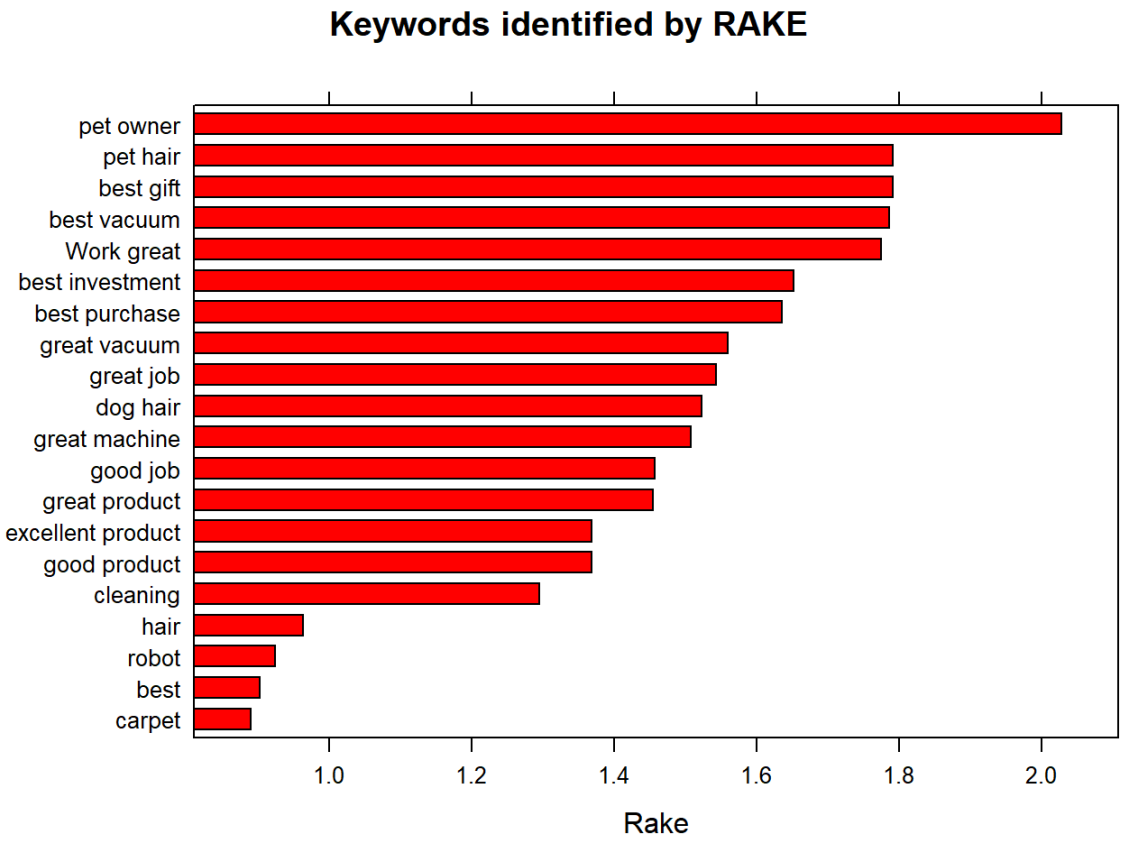

```
## ADJECTIVES
stats <- subset(x, upos %in% c("ADJ"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "purple",
         main = "Most occurring adjectives", xlab = "Freq")
```



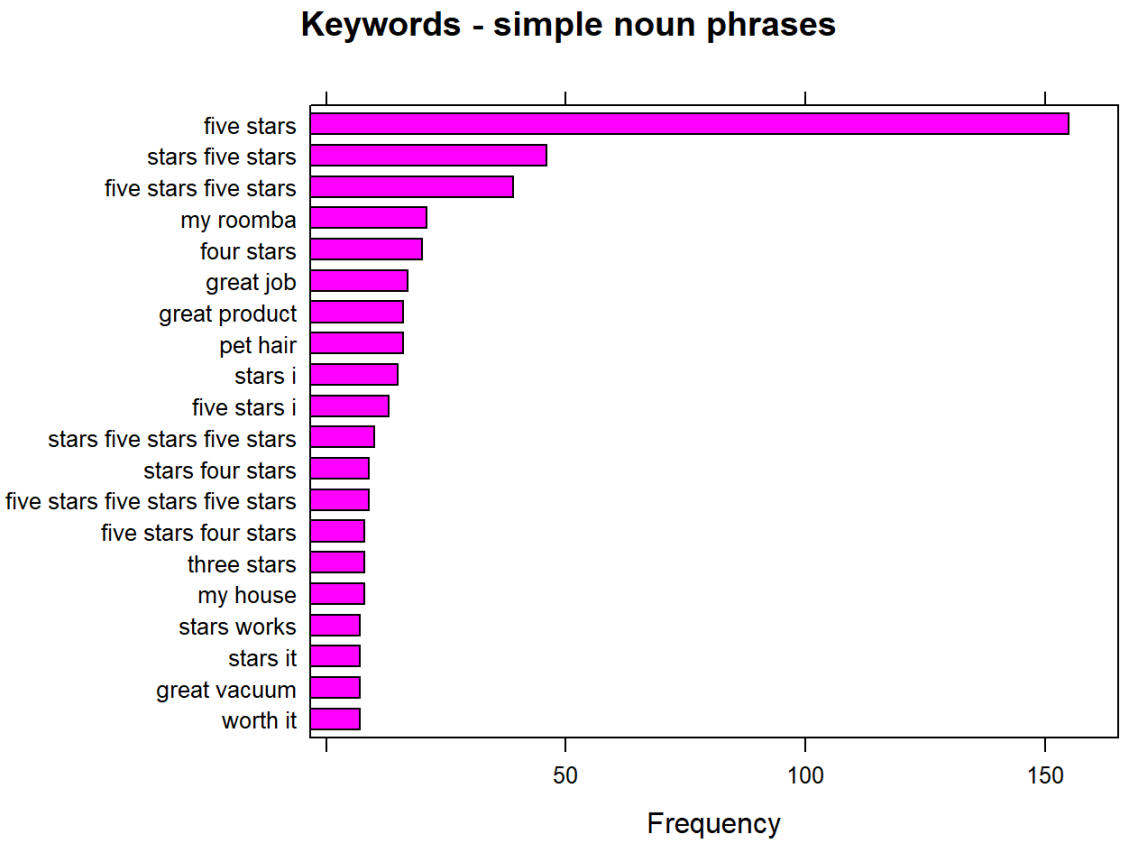

```
## VERBS
stats <- subset(x, upos %in% c("VERB"))
stats <- txt_freq(stats$token)
stats$key <- factor(stats$key, levels = rev(stats$key))
barchart(key ~ freq, data = head(stats, 20), col = "gold",
         main = "Most occurring Verbs", xlab = "Freq")
```




```
## RAKE
stats <- keywords_rake(x = x, term = "lemma", group = "doc_id",
                       relevant = x$upos %in% c("NOUN", "ADJ"))
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ rake, data = head(subset(stats, freq > 3), 20), col = "red",
         main = "Keywords identified by RAKE",
         xlab = "Rake")
```




```
## display by plot a sequence of POS tags (noun phrases / verb phrases)
x$phrase_tag <- as_phrasemachine(x$upos, type = "upos")
stats <- keywords_phrases(x = x$phrase_tag, term = tolower(x$token),
                          pattern = "(A|N)*N(P+D*(A|N)*N)*",
                          is_regex = TRUE, detailed = FALSE)
stats <- subset(stats, ngram > 1 & freq > 3)
stats$key <- factor(stats$keyword, levels = rev(stats$keyword))
barchart(key ~ freq, data = head(stats, 20), col = "magenta",
         main = "Keywords - simple noun phrases", xlab = "Frequency")
```



The plots that I got above are the analysis of verbs, nouns, noun verb and parts of speech in customers' reviews for these three products. Because the rating of them is quite good, verbs, nouns and noun phreses that appear most frequently are similar. However, in keywords identified by RAKE, there are some difference. For iLIFE V5s, "great value" and "best purchase" appear a lot because of its lower price, and "pet hair""pet owner" appear a lot for iRobot because of their product charactoristic.

Step 4

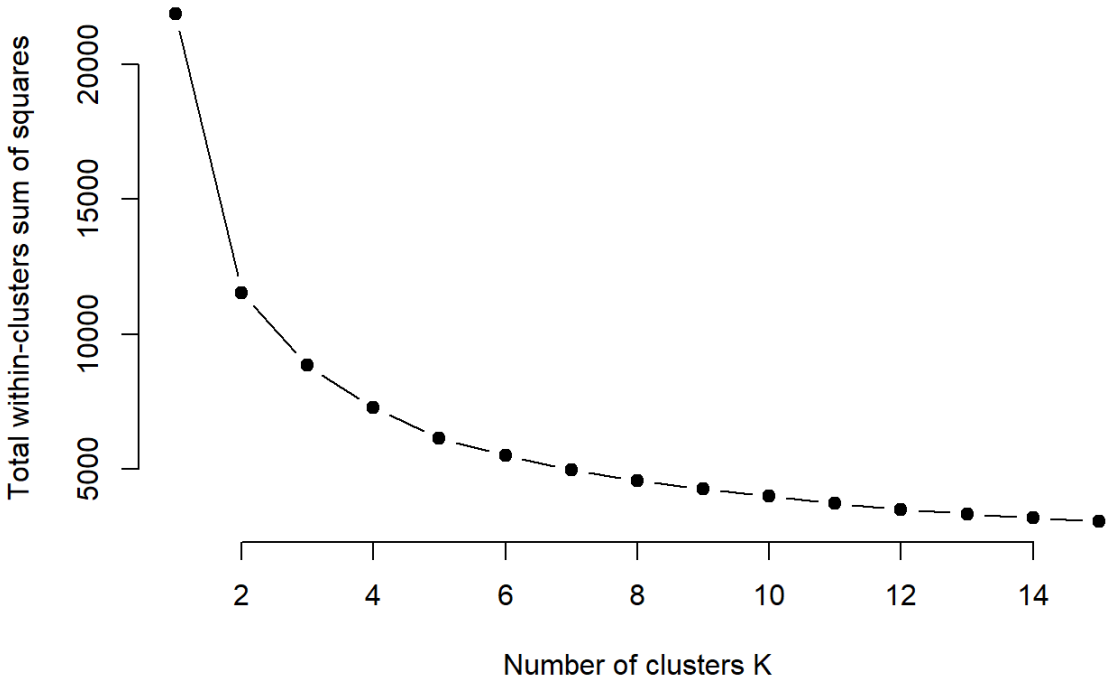
```
library(animation)
library(ggplot2)
library(dplyr)
df <- read.csv (file.choose())
subscribermodel <- select(df, -CustomerID)
summary(subscribermodel)
```

```
##     Recency           Frequency        MonetaryValue
##  Min.   :-2.81452   Min.   :-1.79473   Min.   :-4.0935
##  1st Qu.:-0.63739   1st Qu.:-0.64918   1st Qu.:-0.6579
##  Median : 0.09268   Median : 0.02092   Median :-0.0149
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.83404   3rd Qu.: 0.72108   3rd Qu.: 0.6693
##  Max.   : 1.54788   Max.   : 4.24683   Max.   : 4.4595
```

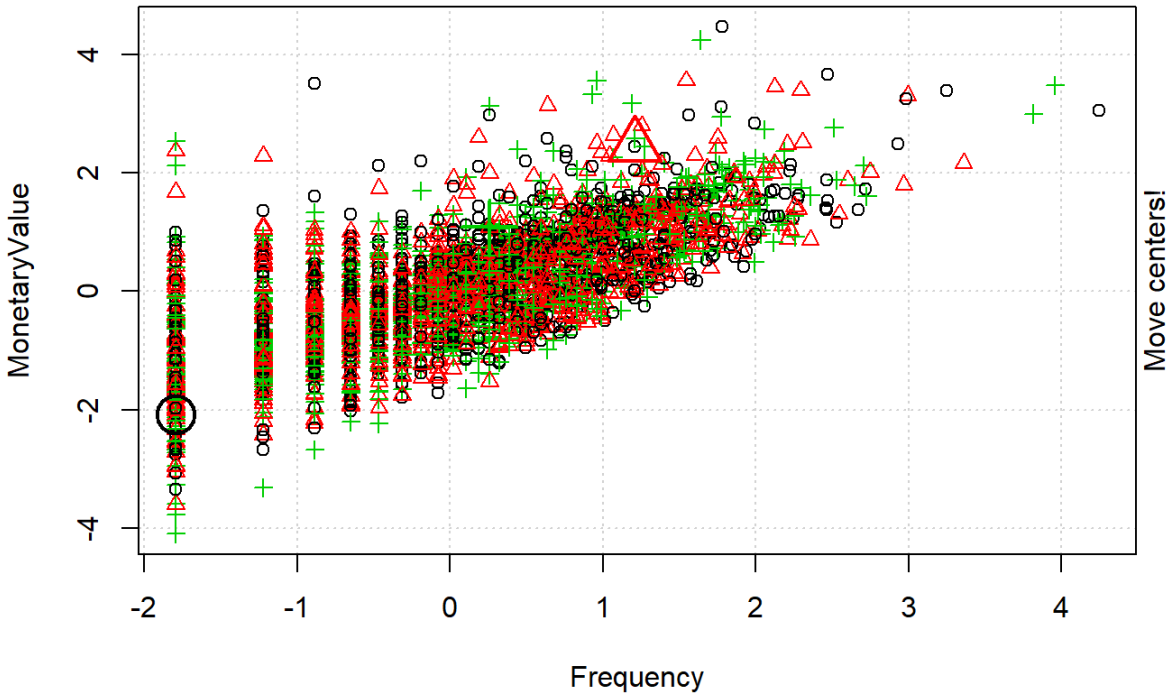
```
# Rescale the numerical items for clustering
rescale_subscribermodel <-subscribermodel %>%  mutate(recency_scaled = scale(Recency), frequency_scaled = s
cale(Frequency),monetary_scaled = scale(MonetaryValue))

# omit the data that is missing
rescale_subscribermodel <- na.omit(rescale_subscribermodel)

# Set a seed and then use ELbow Method to identify the number of cluster
set.seed(2600)
k.max <- 15
data <- rescale_subscribermodel
wss <- sapply(1:k.max,
              function(k){kmeans(data, k, nstart=50,iter.max = 15 )$tot.withinss})
wss
```

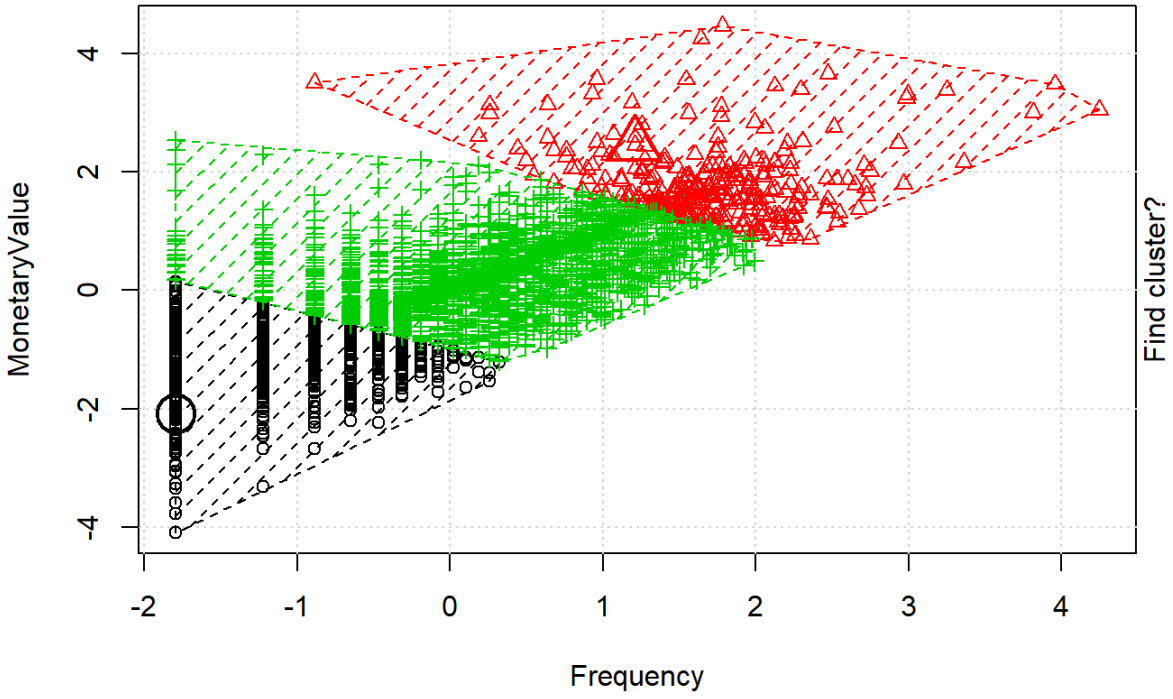
```
##  [1] 21855.000 11545.924  8841.711  7283.901  6158.630  5511.079  4986.825
##  [8]  4578.912  4276.815  4013.862  3749.450  3512.030  3353.338  3208.479
## [15]  3063.915
```

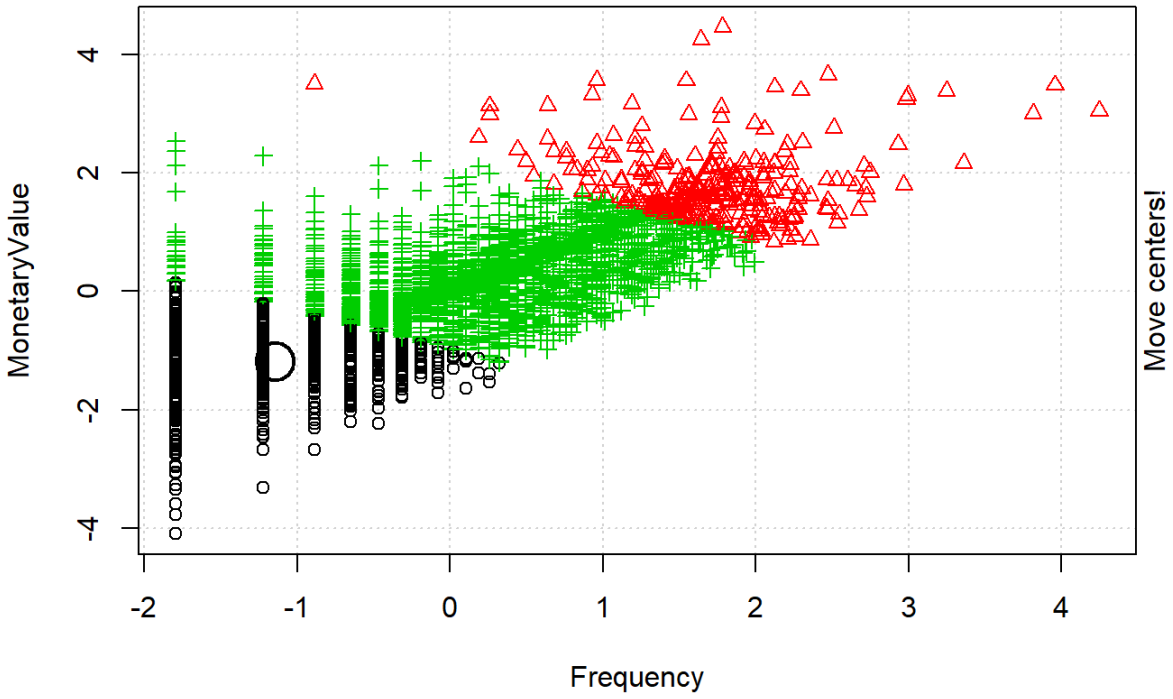
```
plot(1:k.max, wss,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")
```

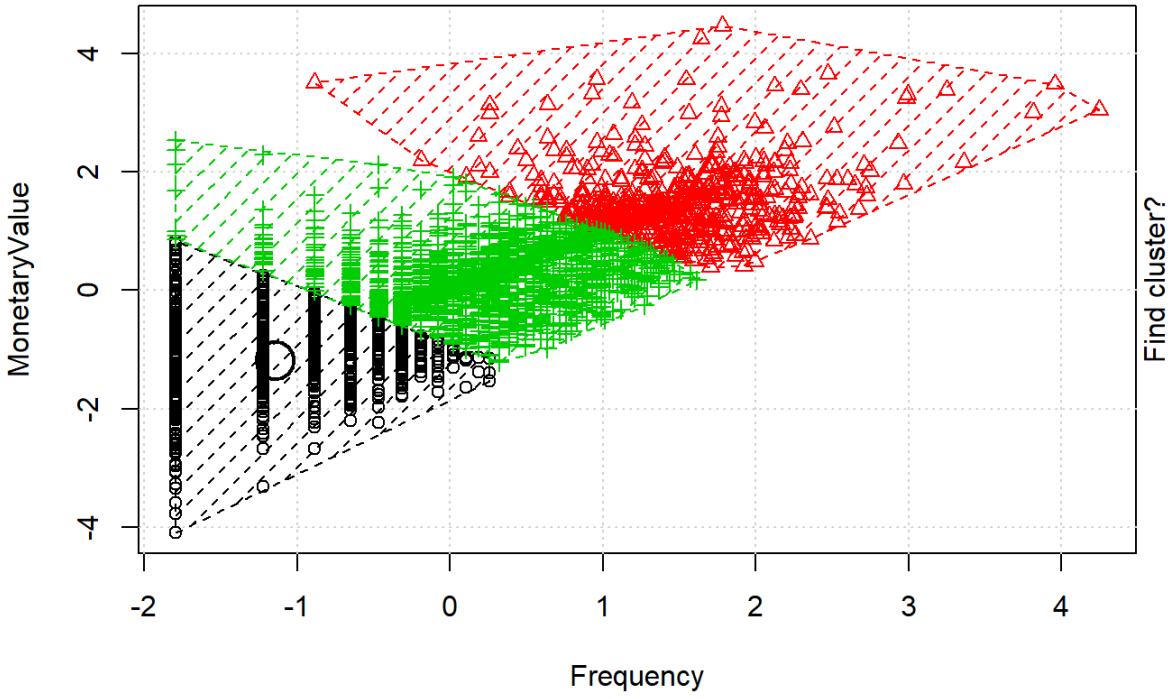


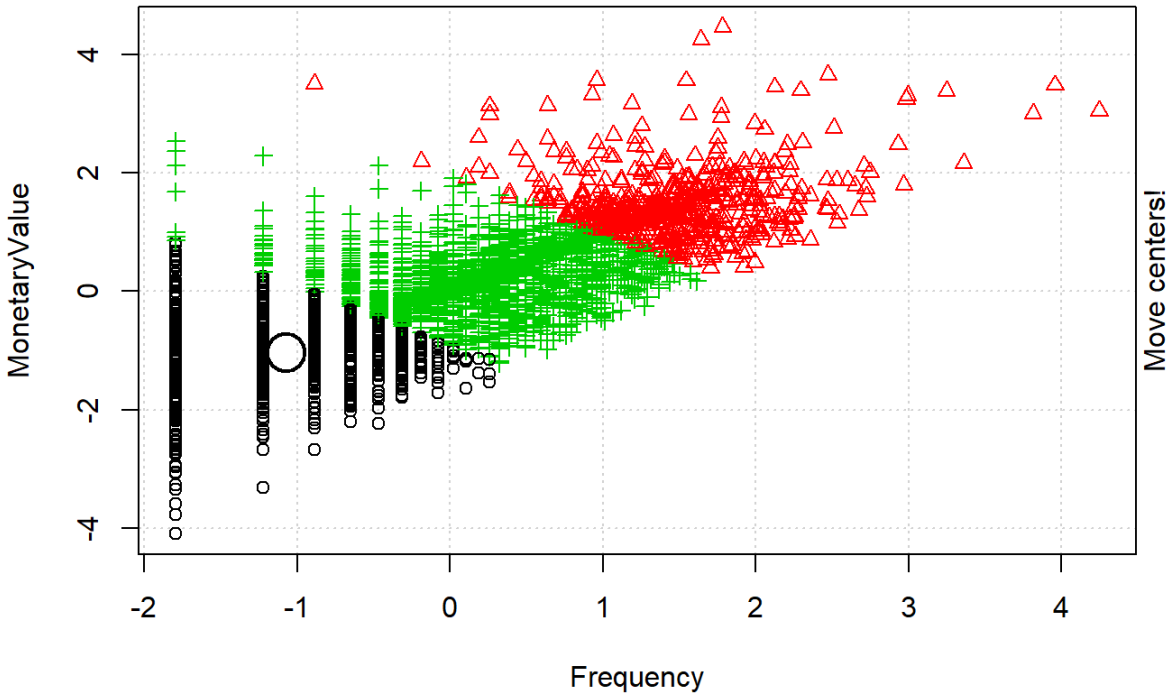

```
kmeans.ani(rescale_subscribermodel[2:3], 3)
```

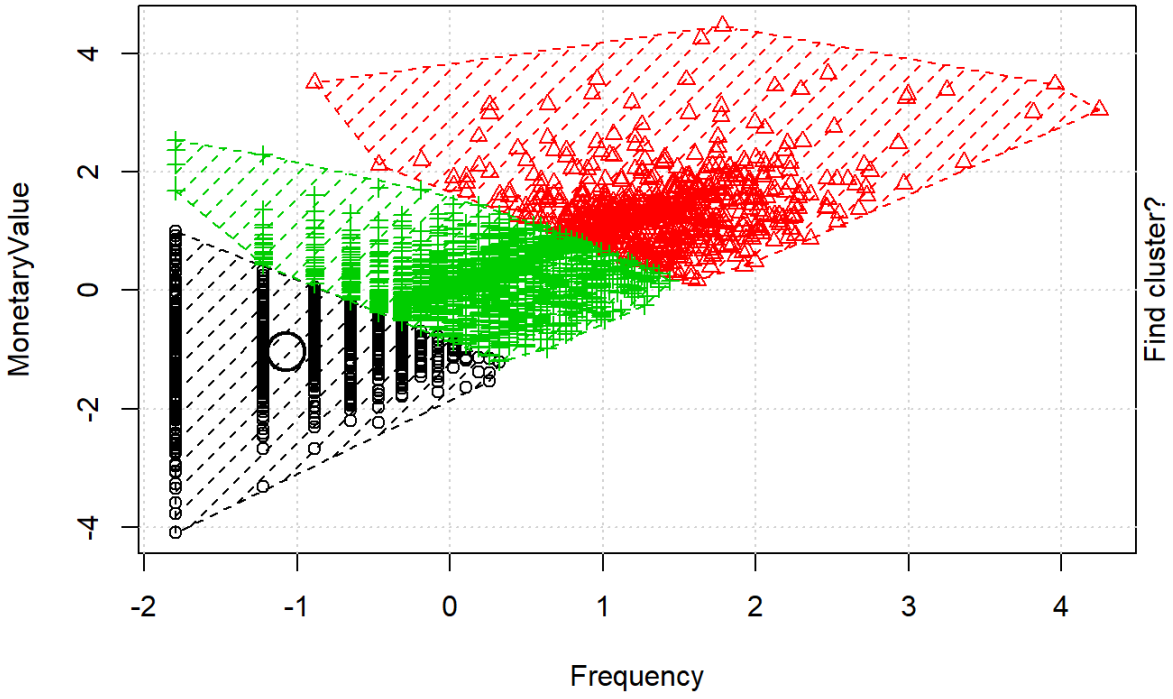

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

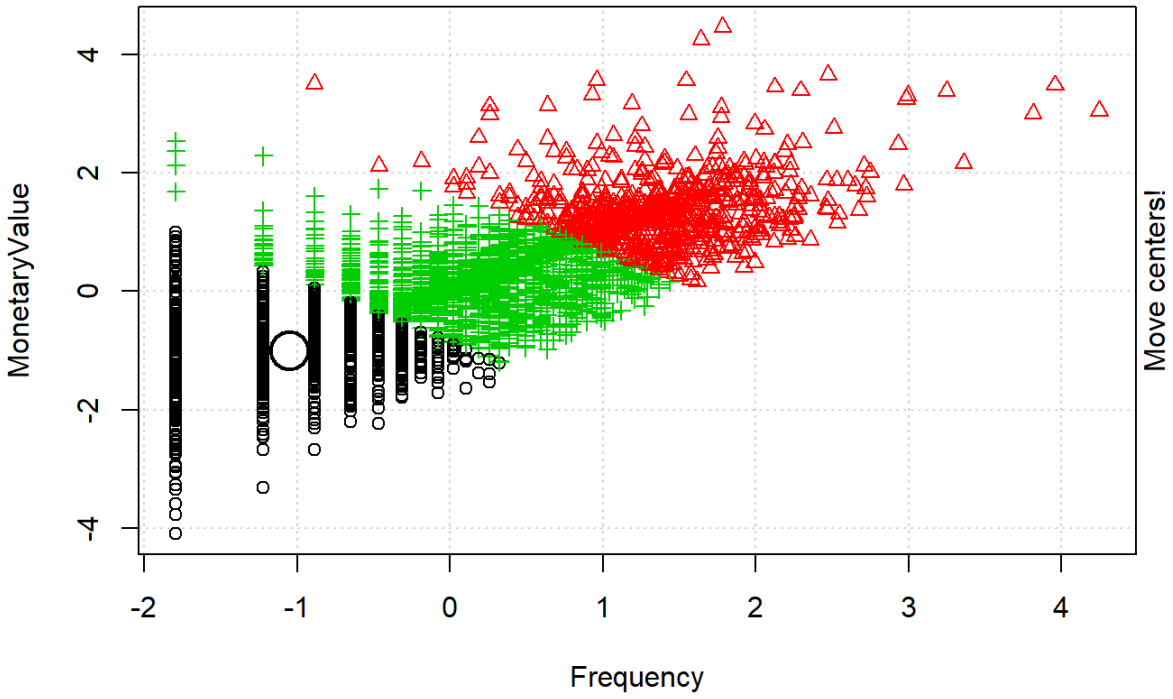

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

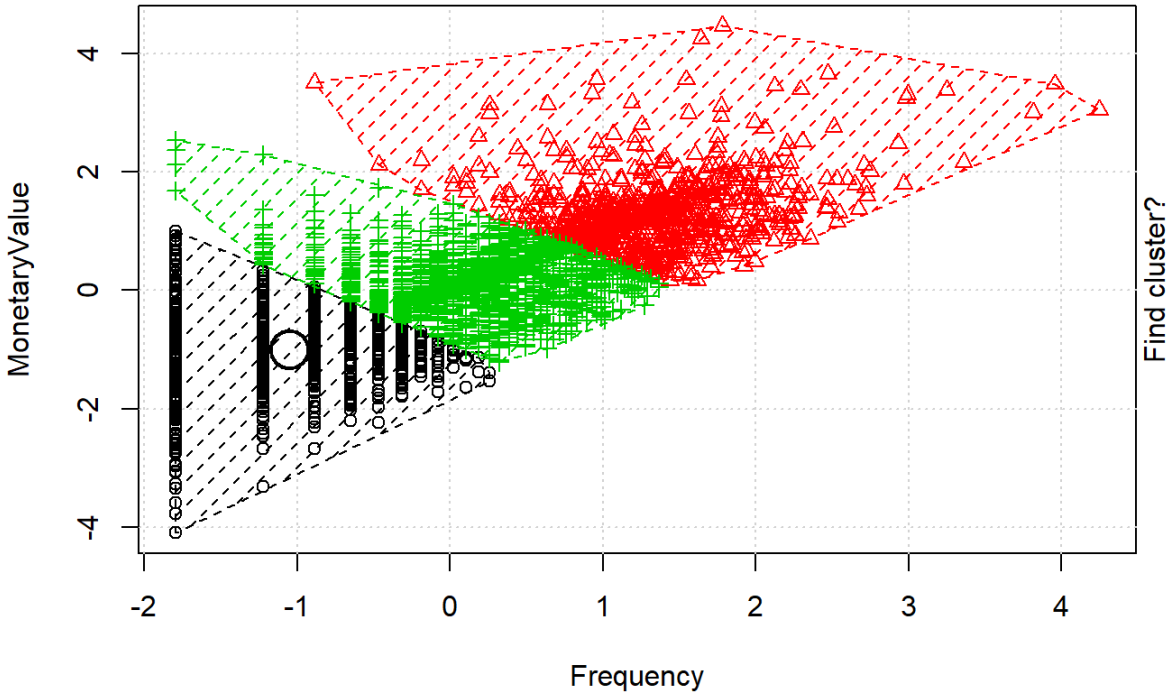

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

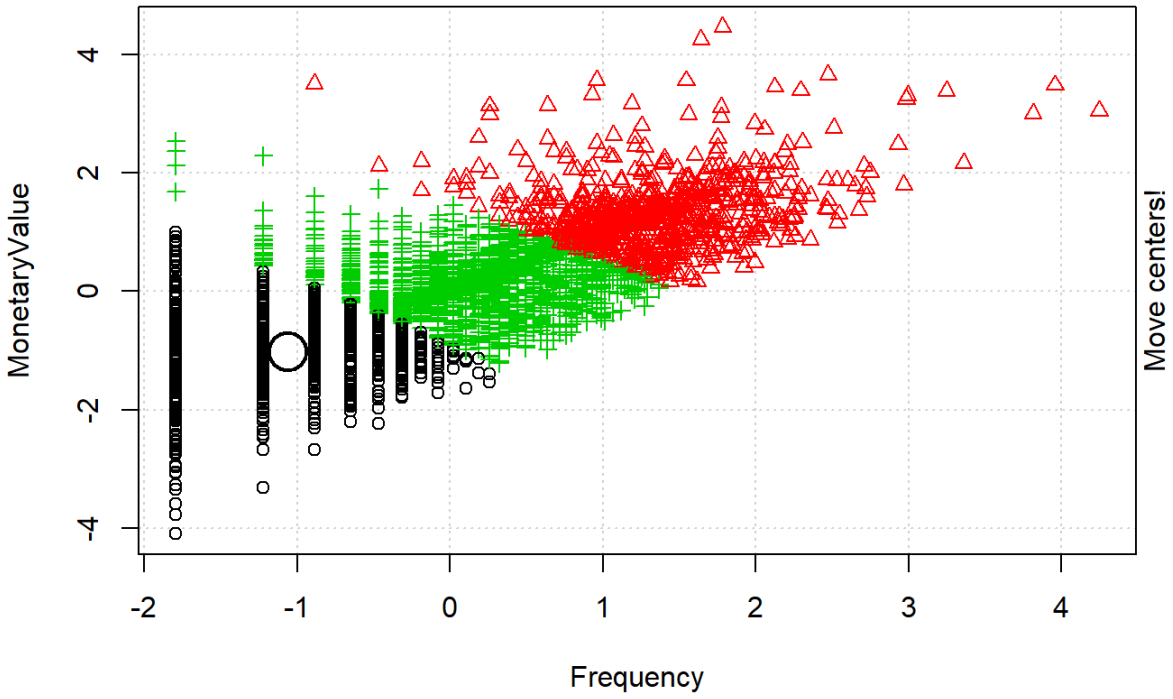

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

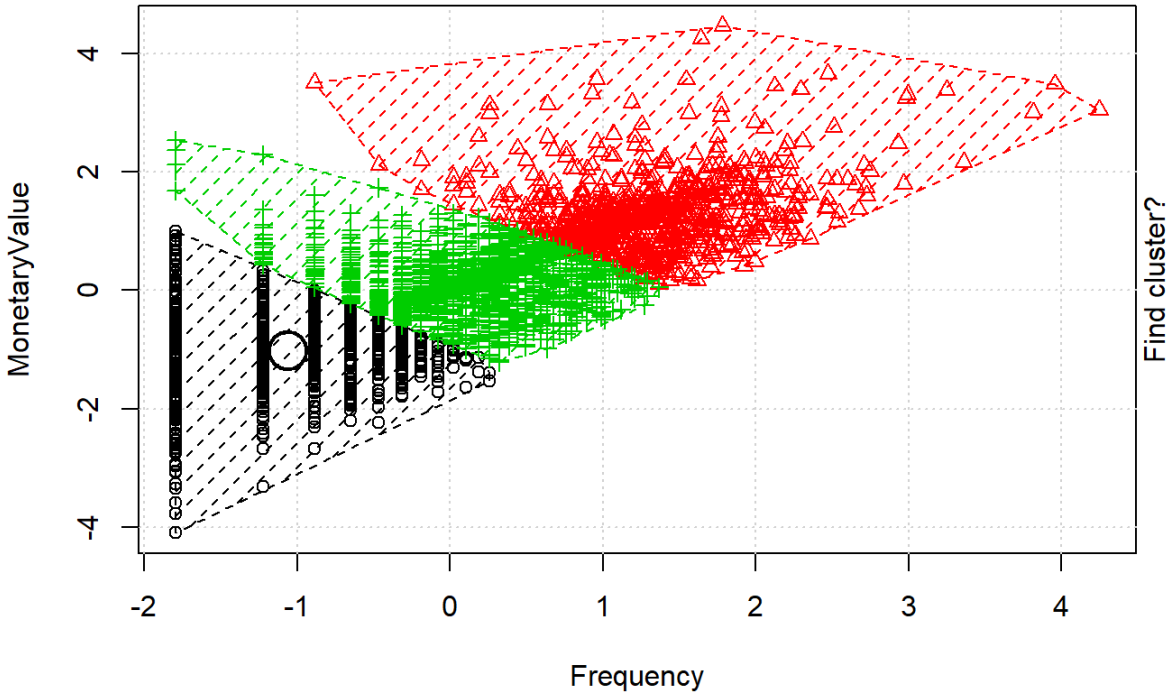

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

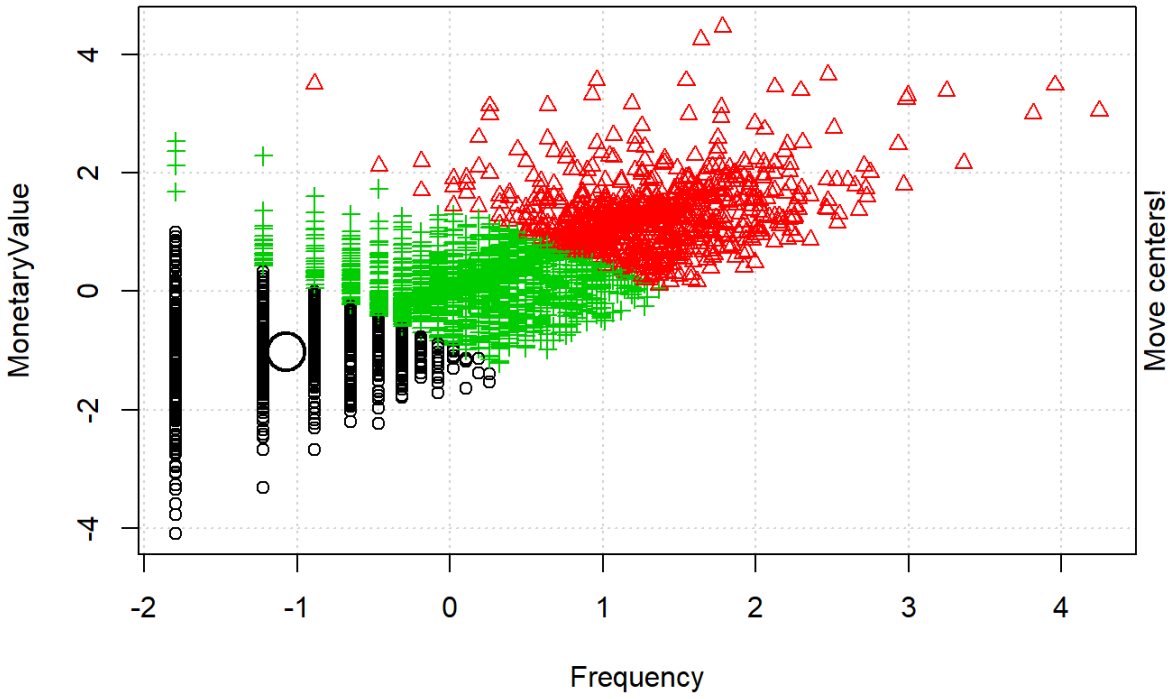

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

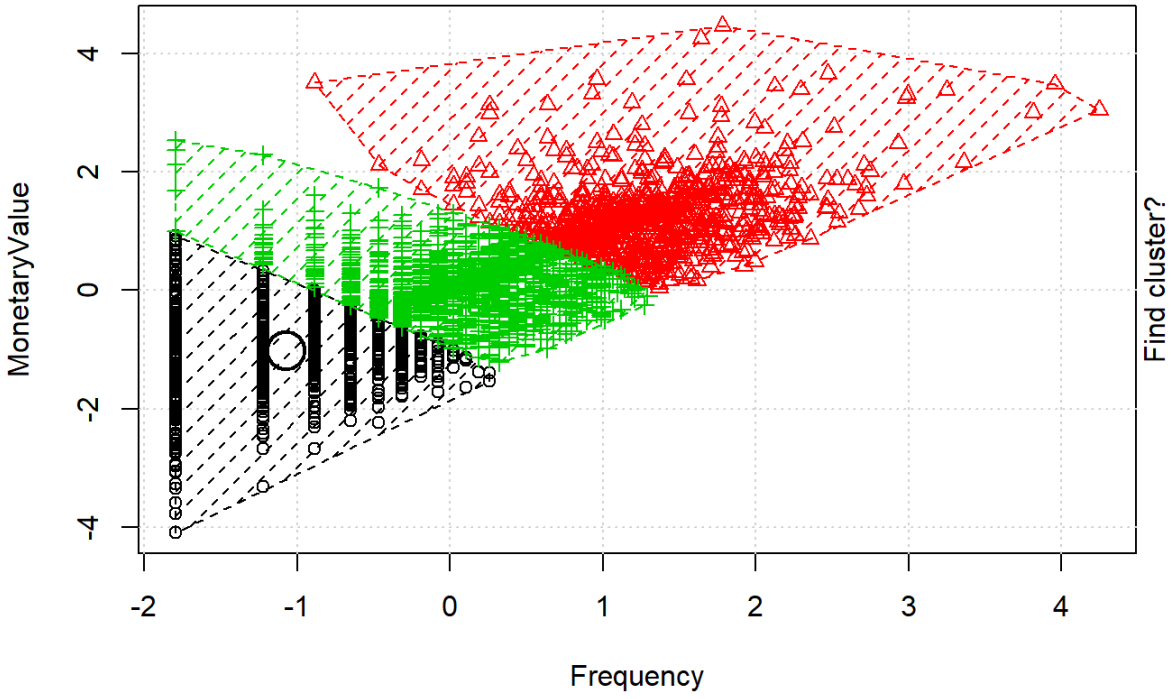

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

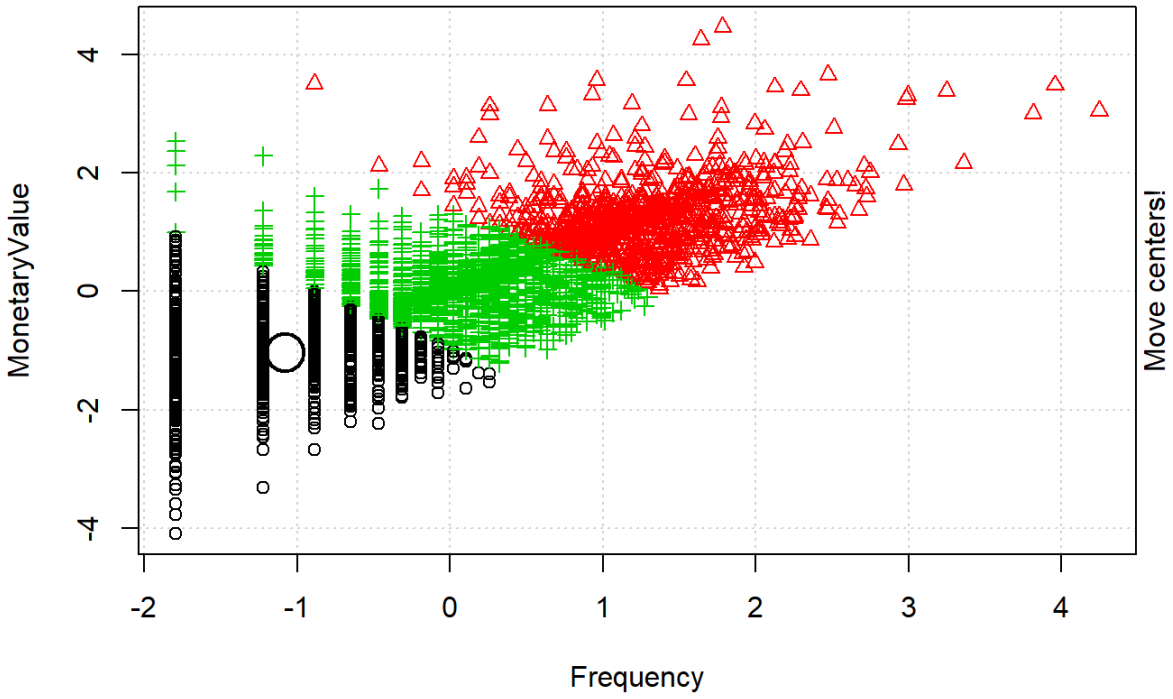

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

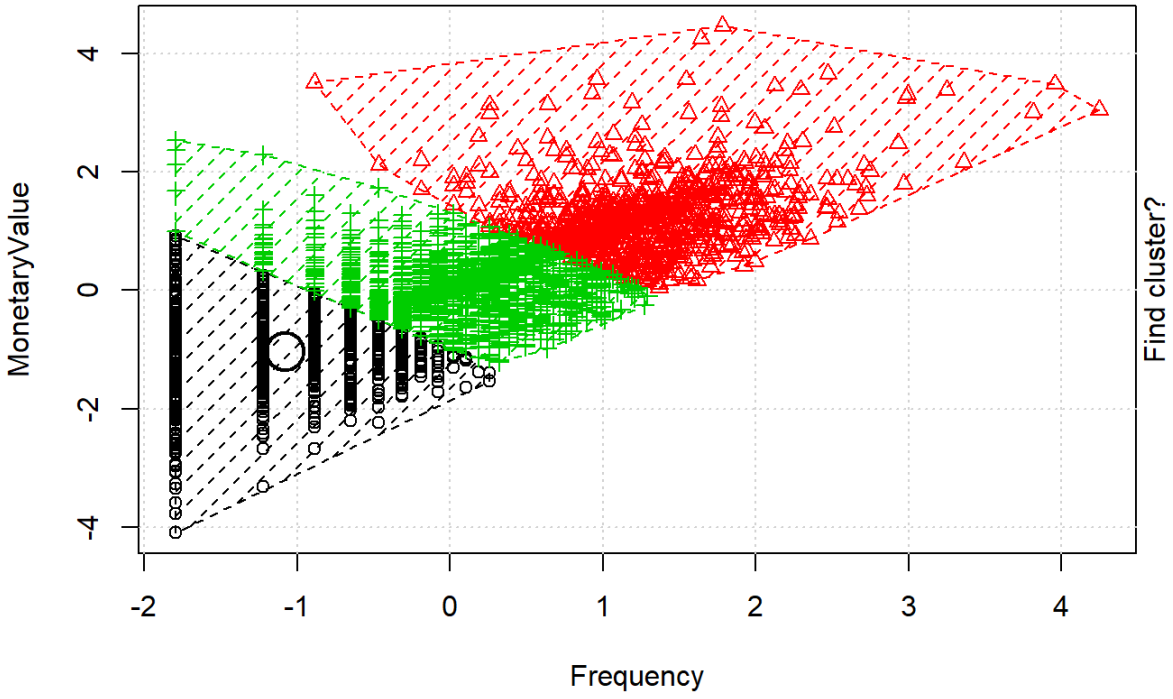

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

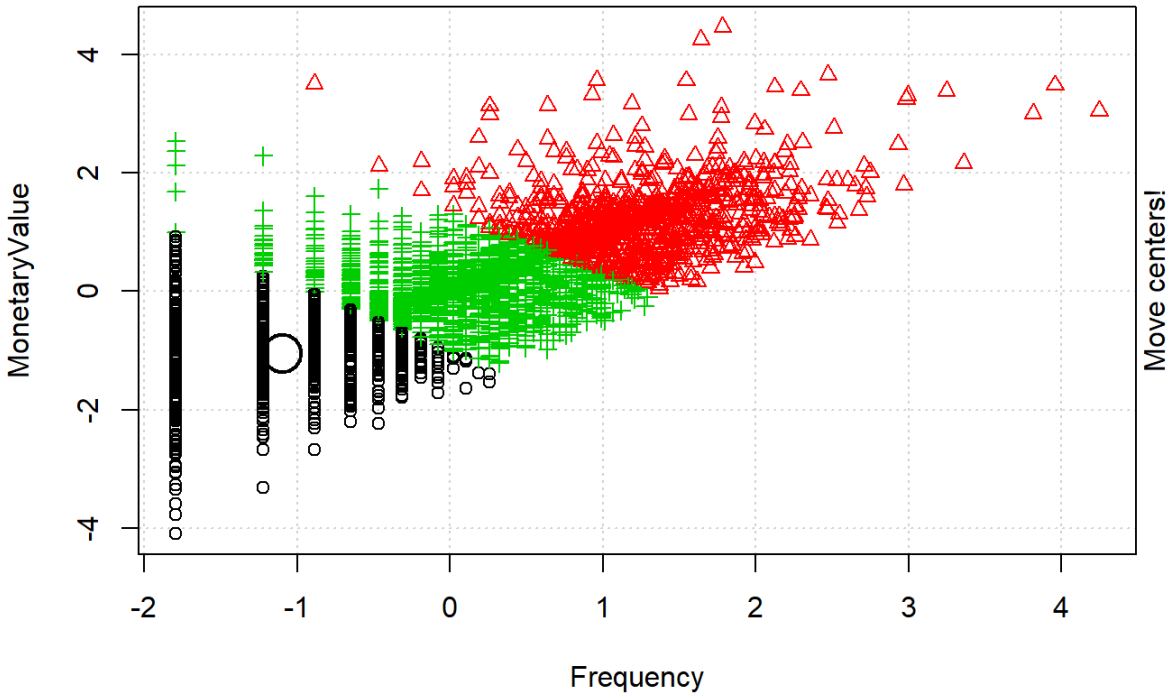

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

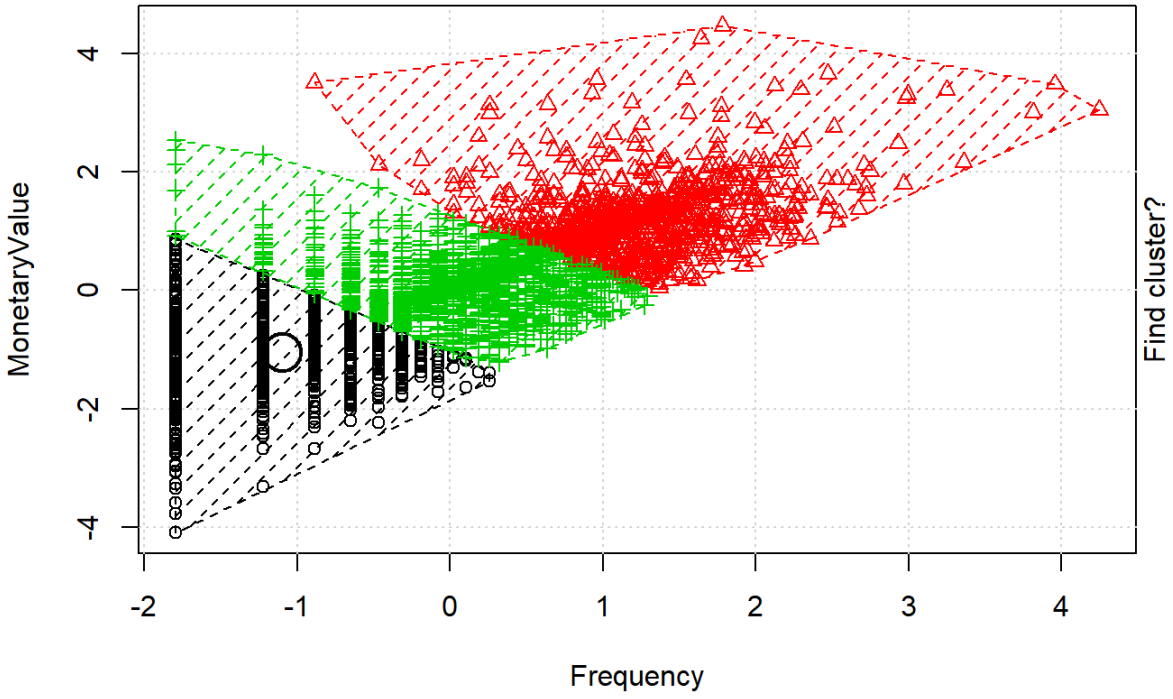

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

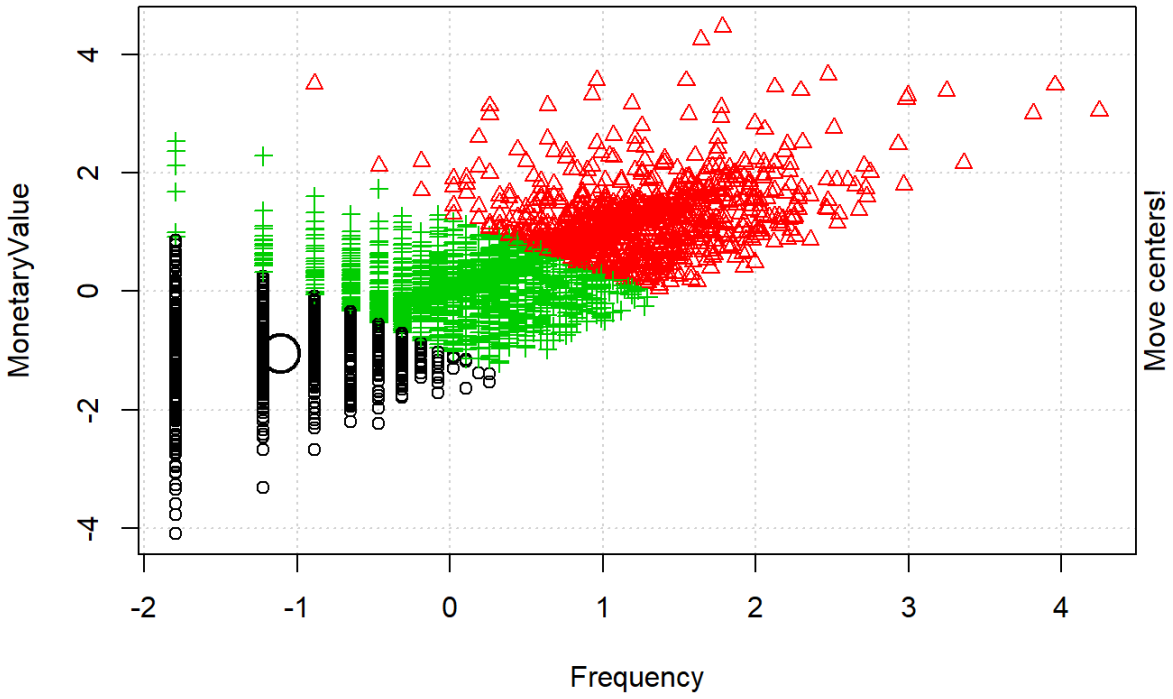

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

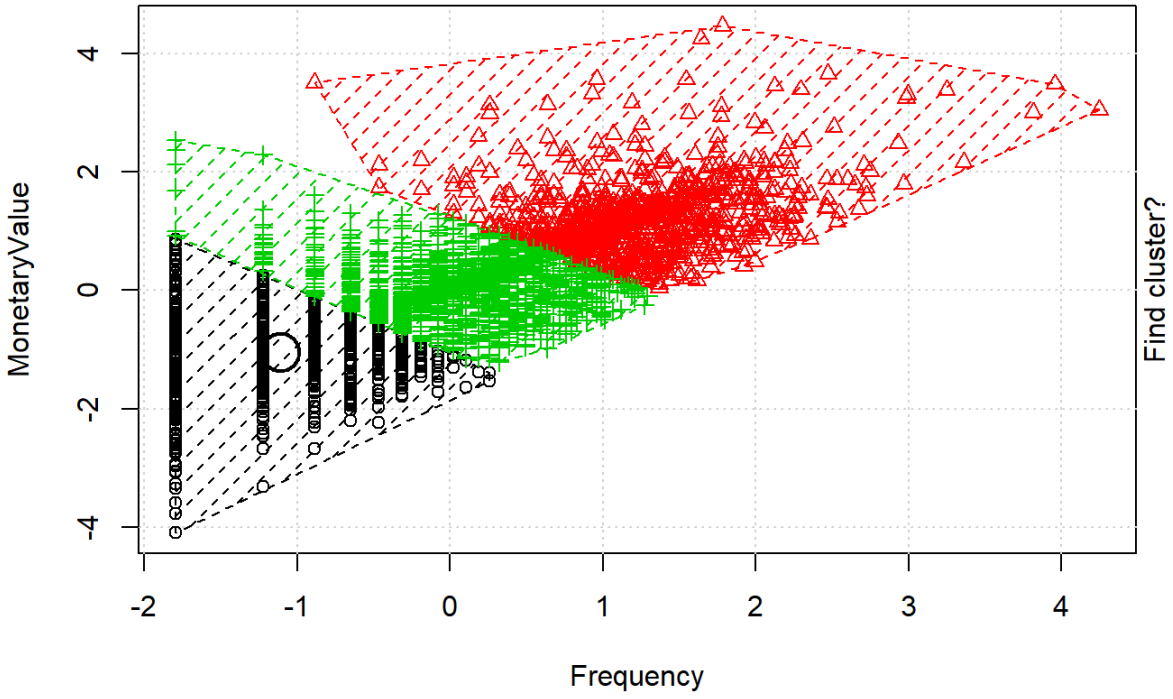

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

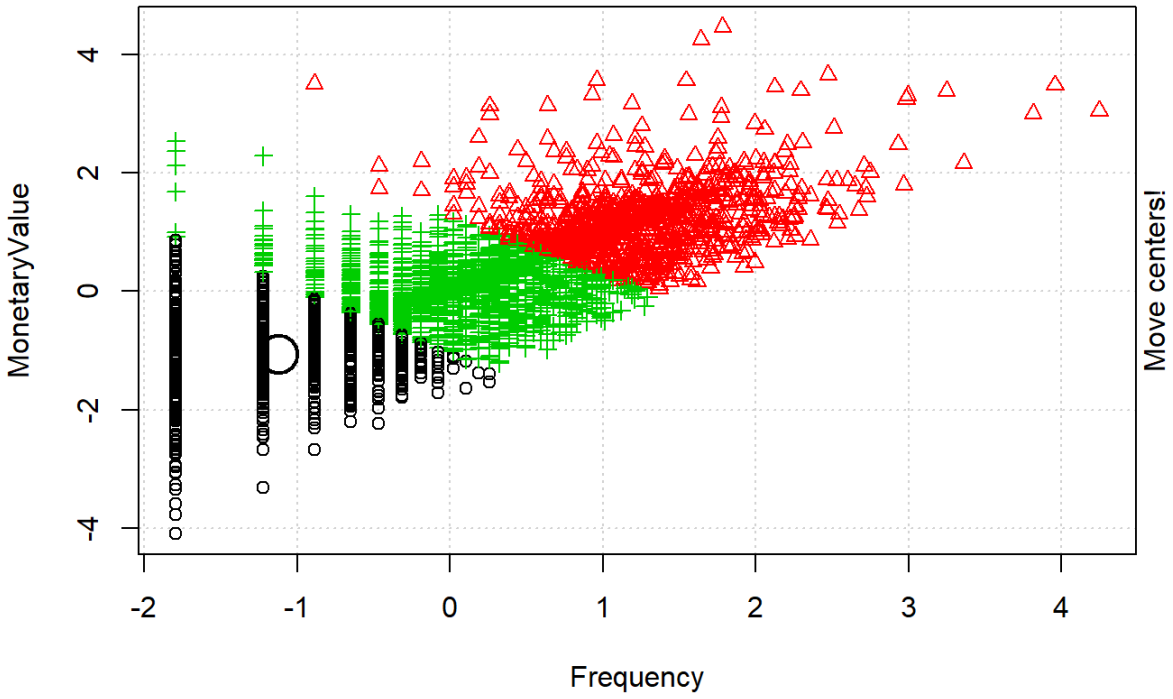

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

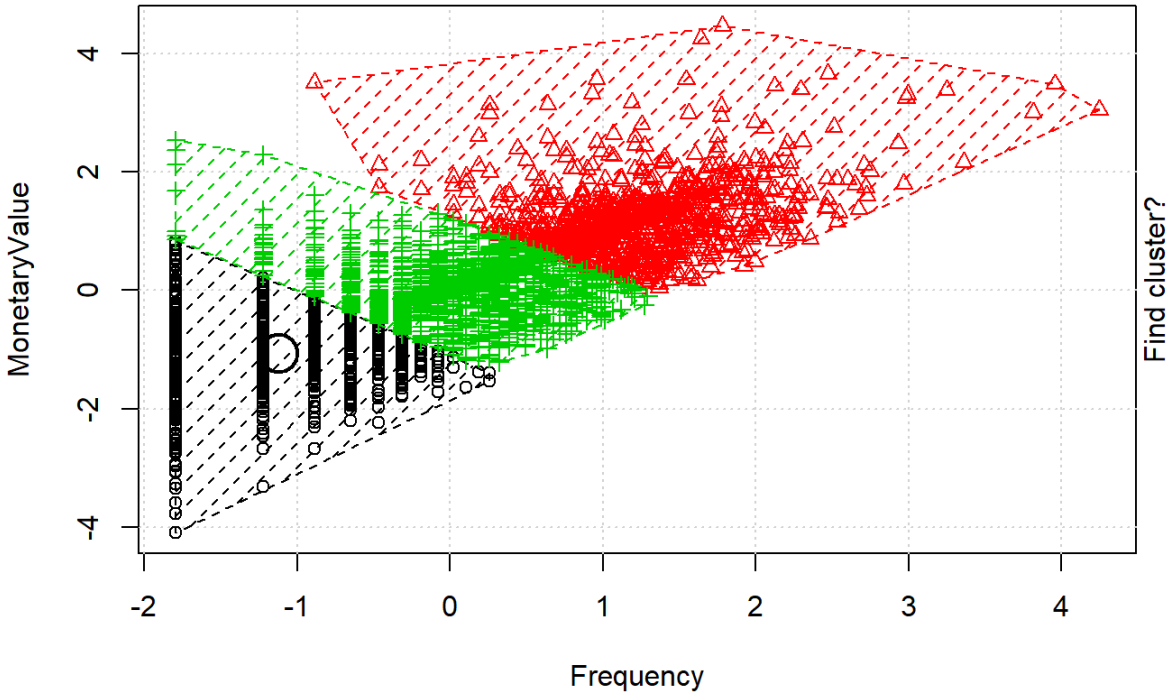

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

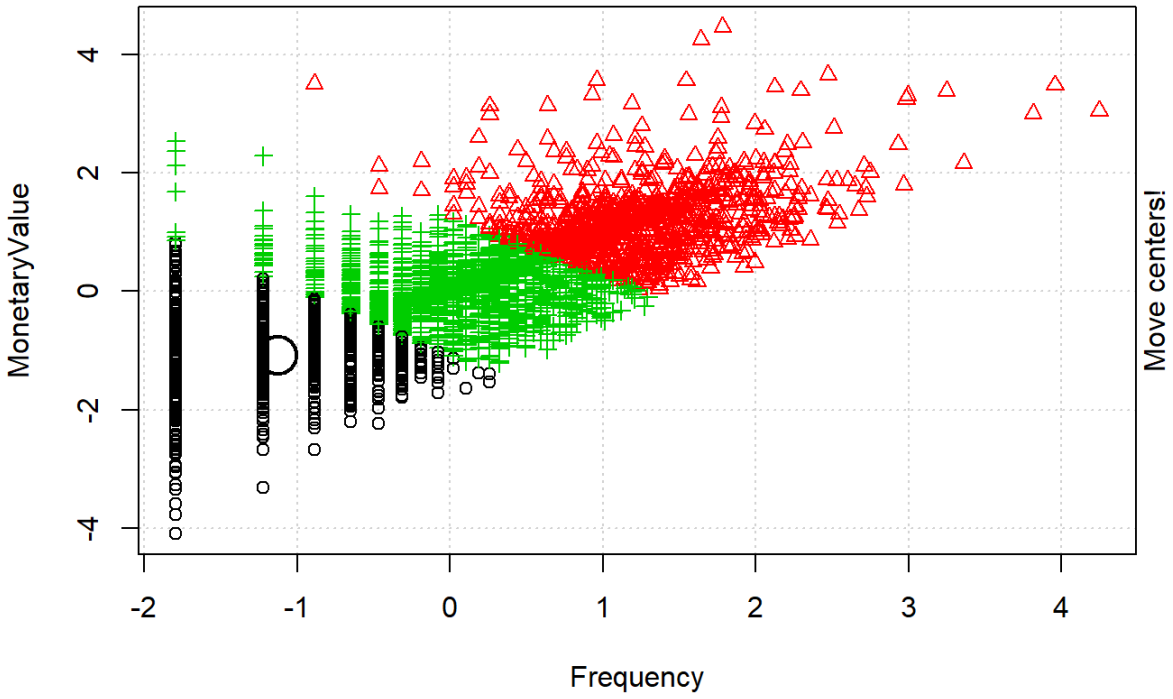

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

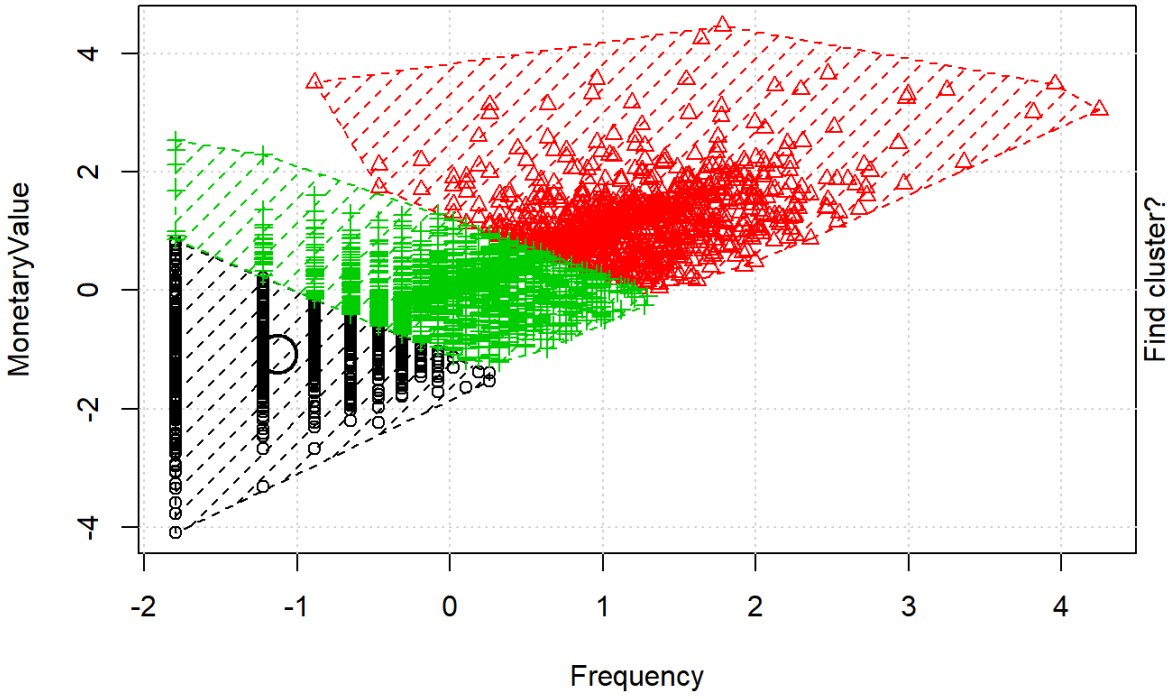

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

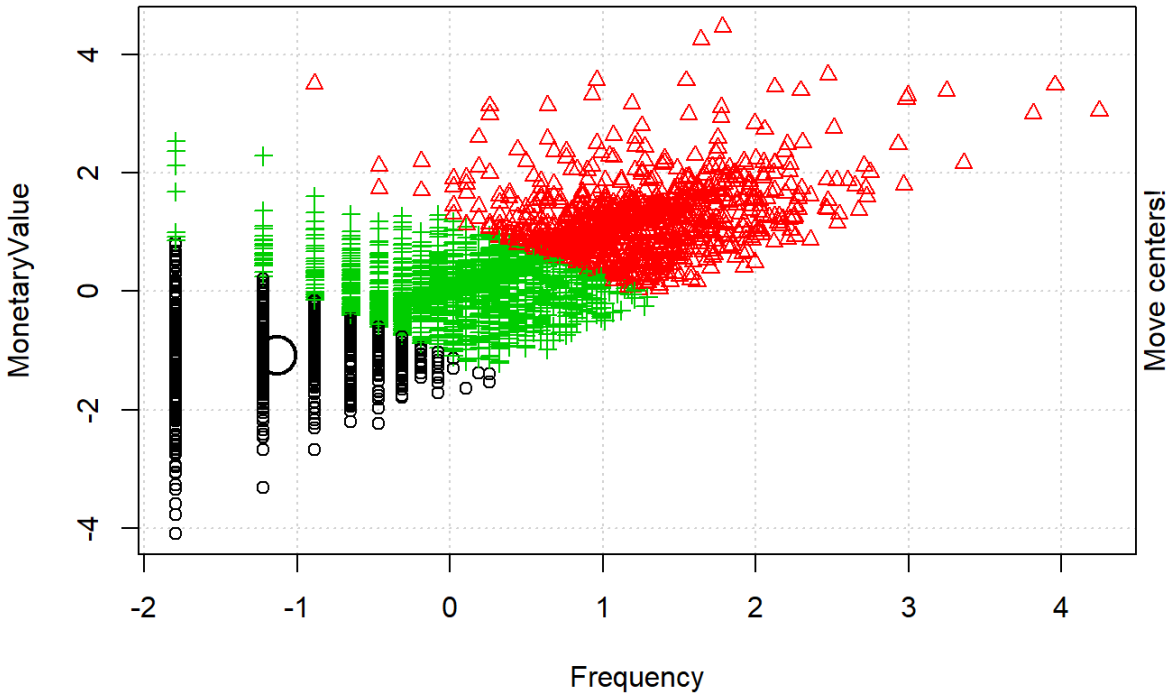

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

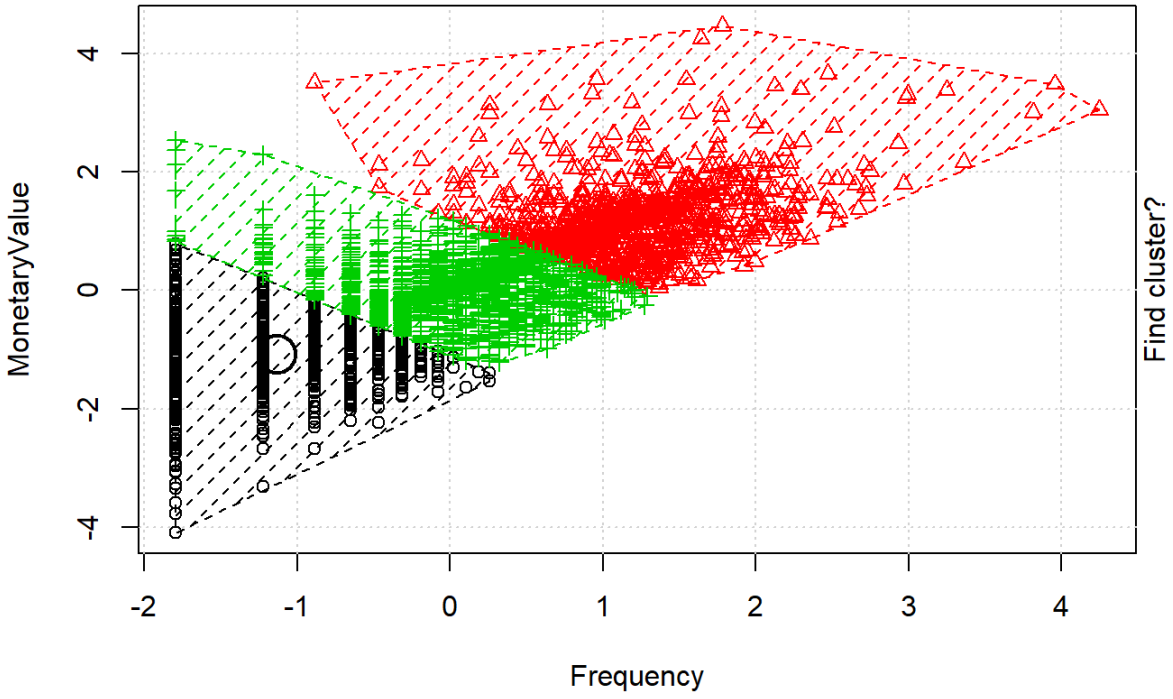

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

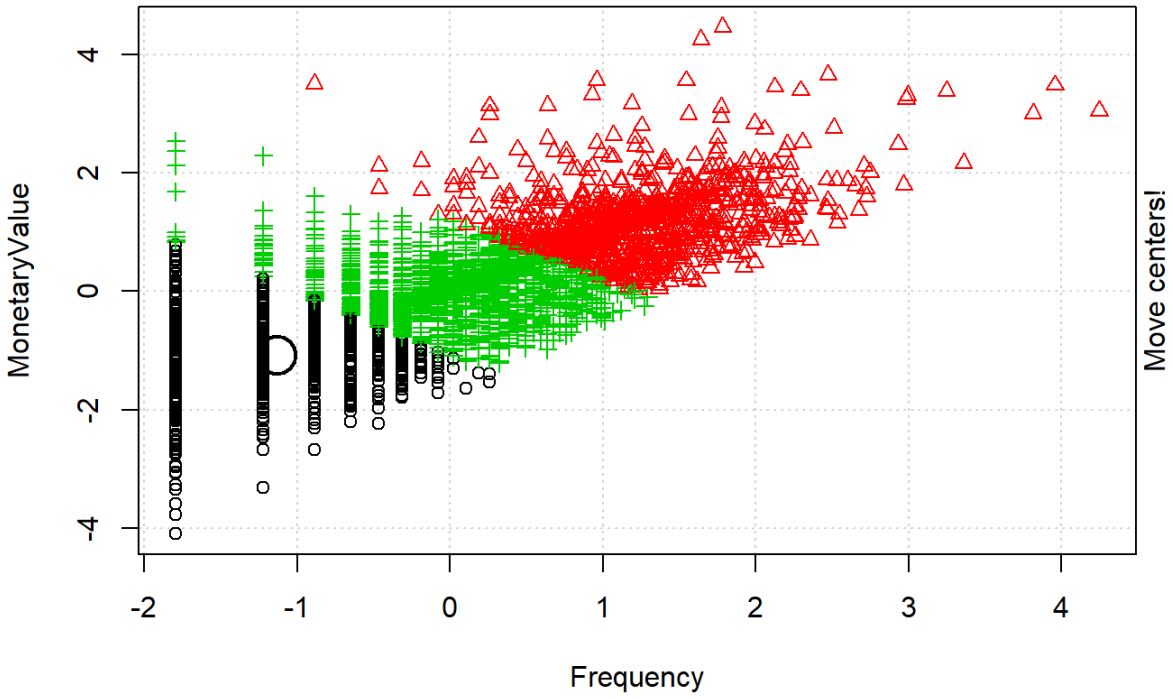

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

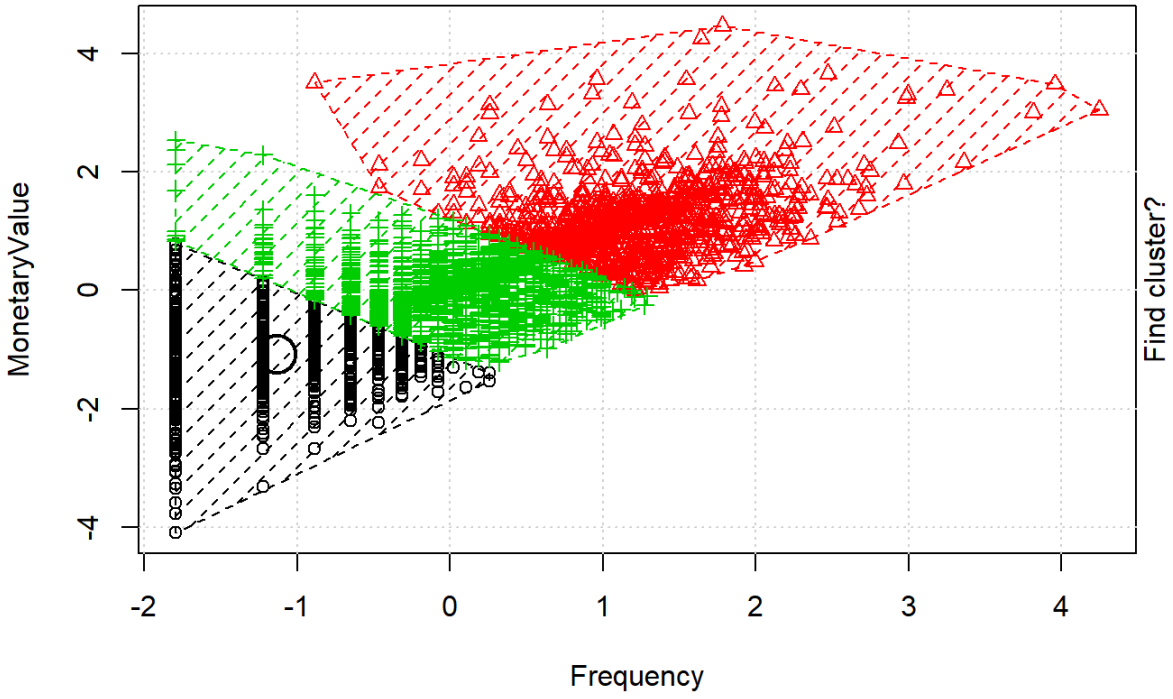

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

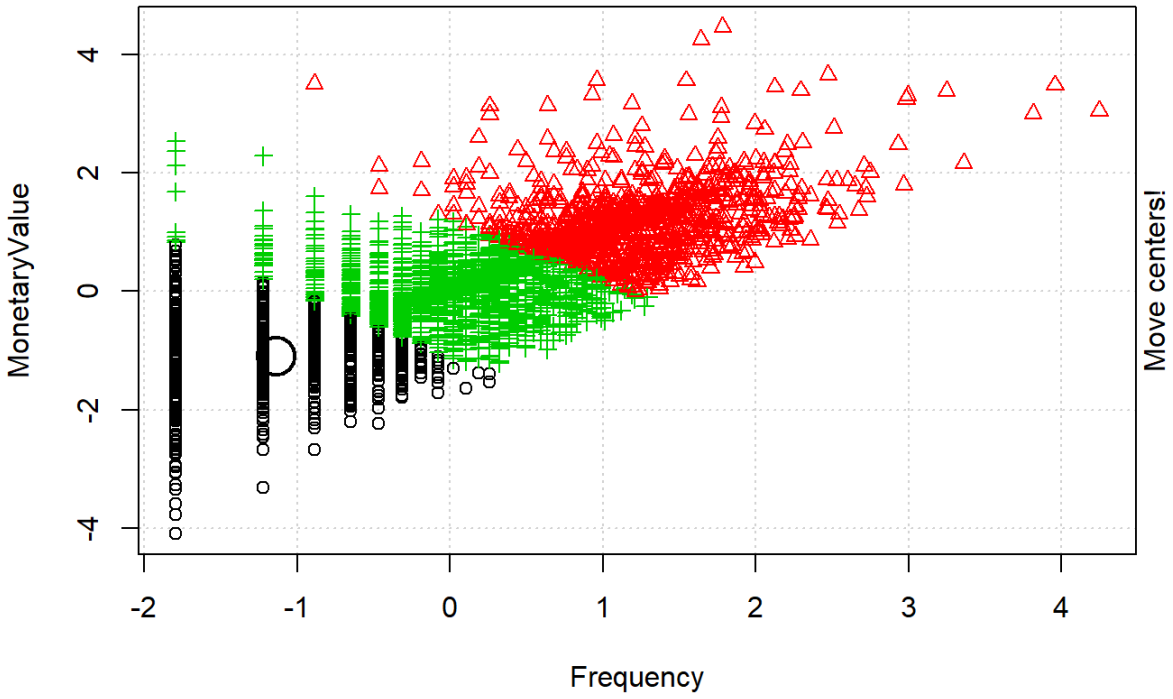

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

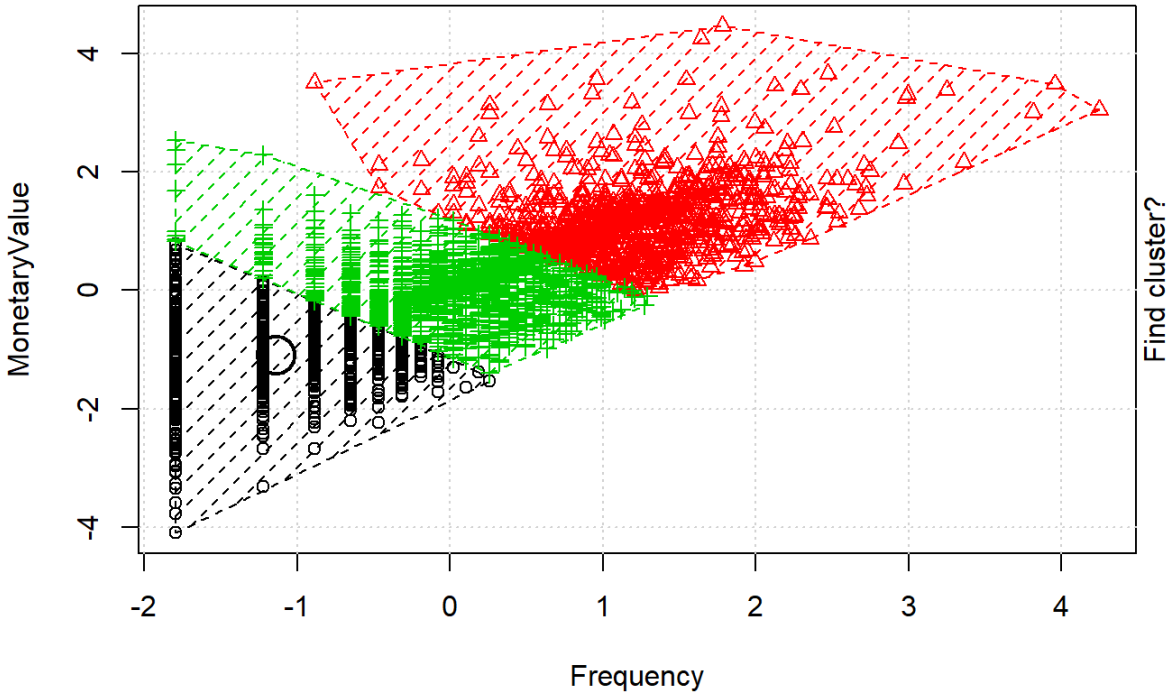

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

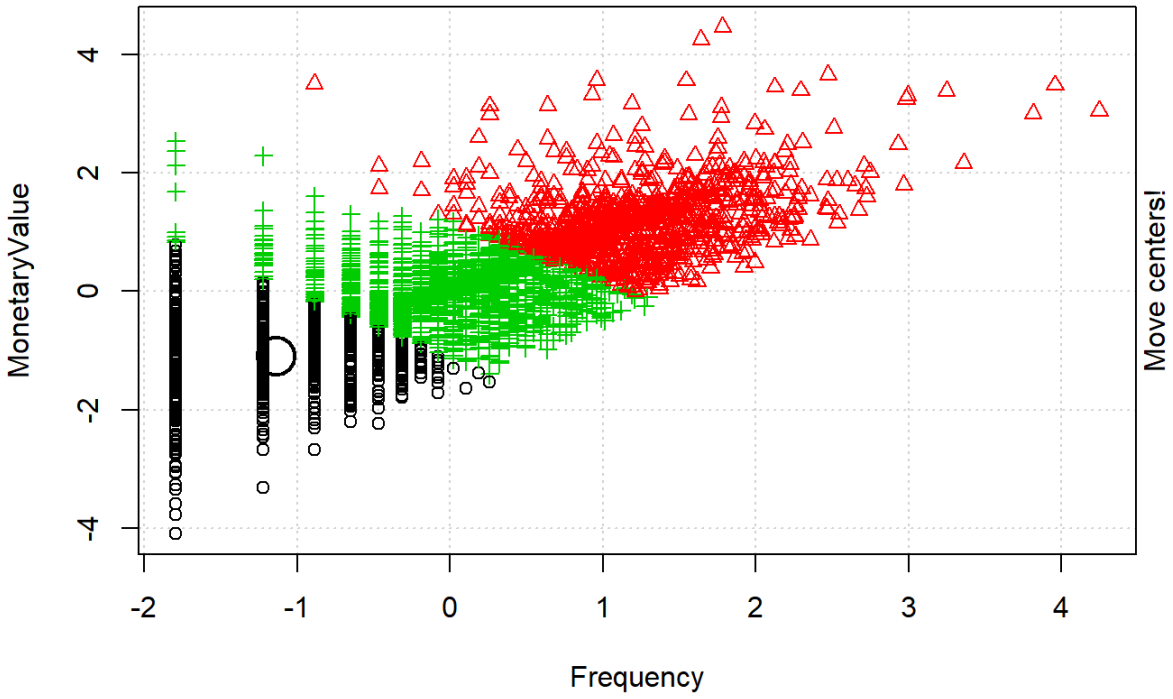

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

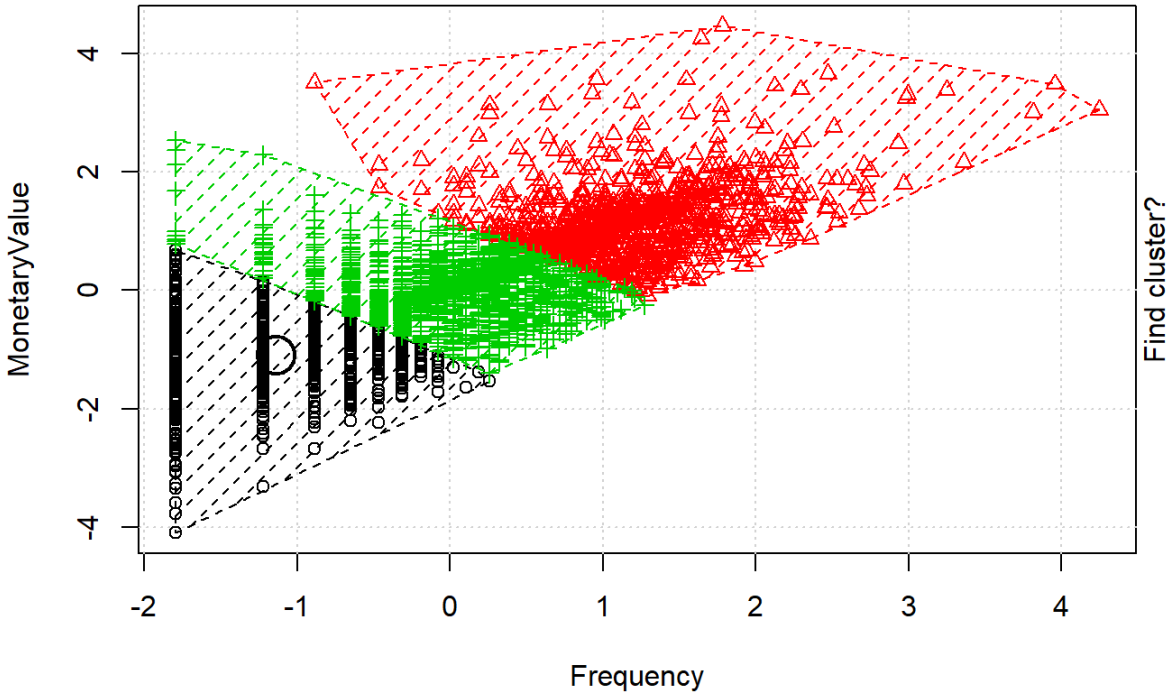

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

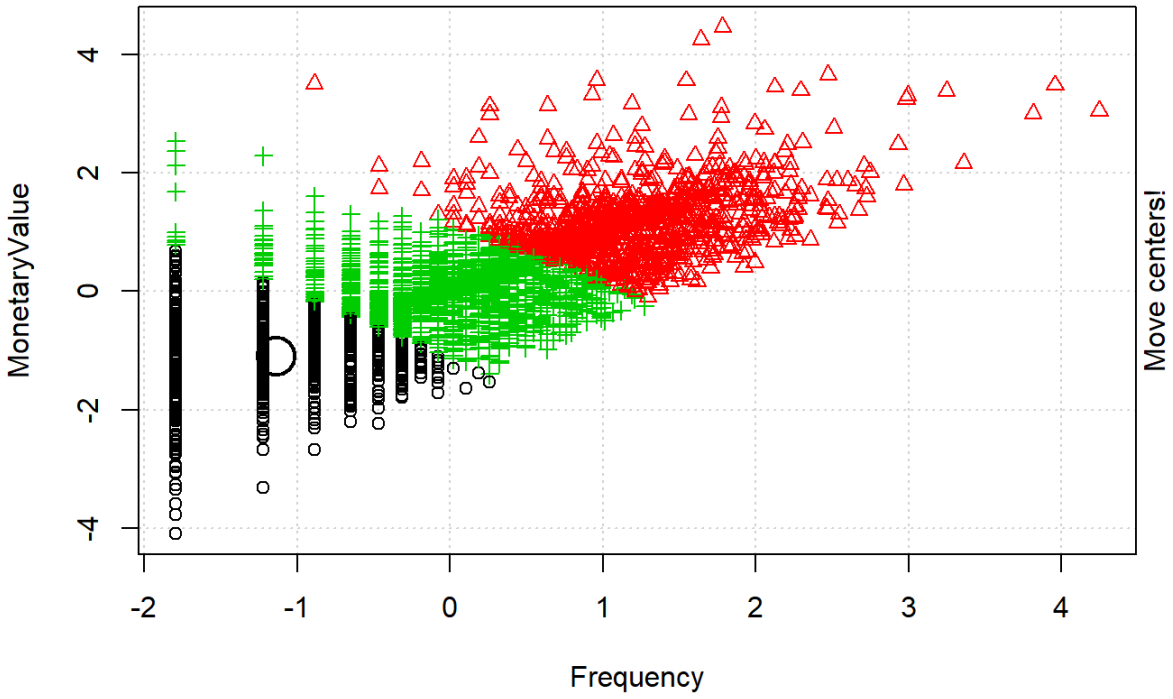

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

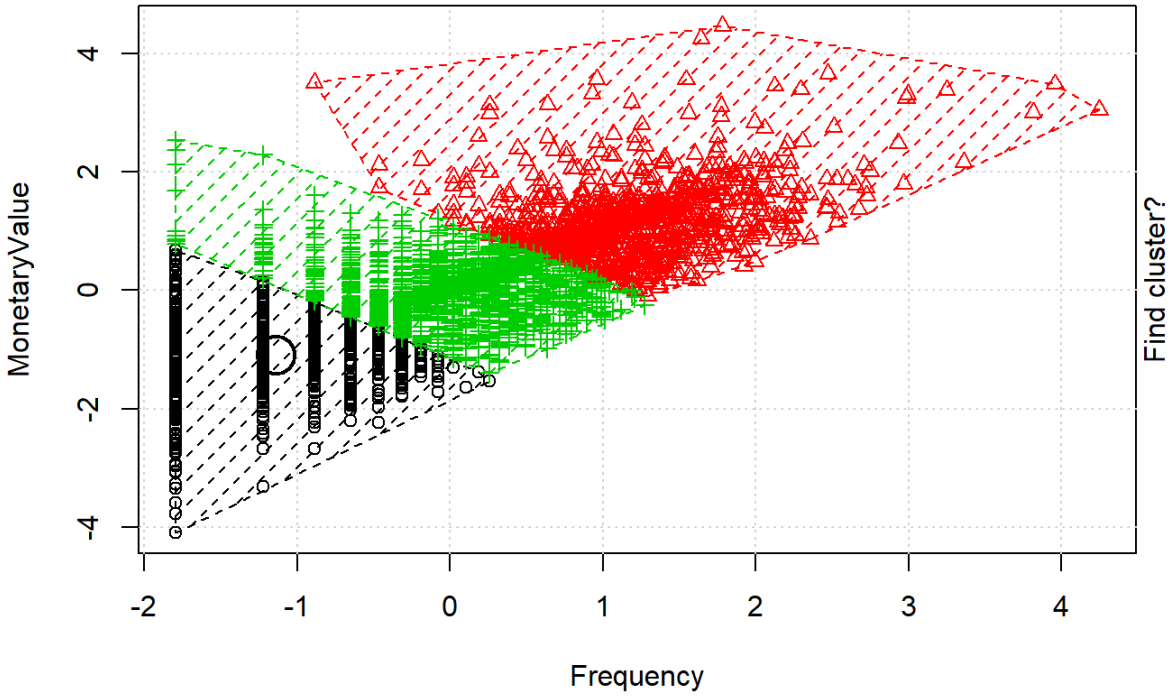

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

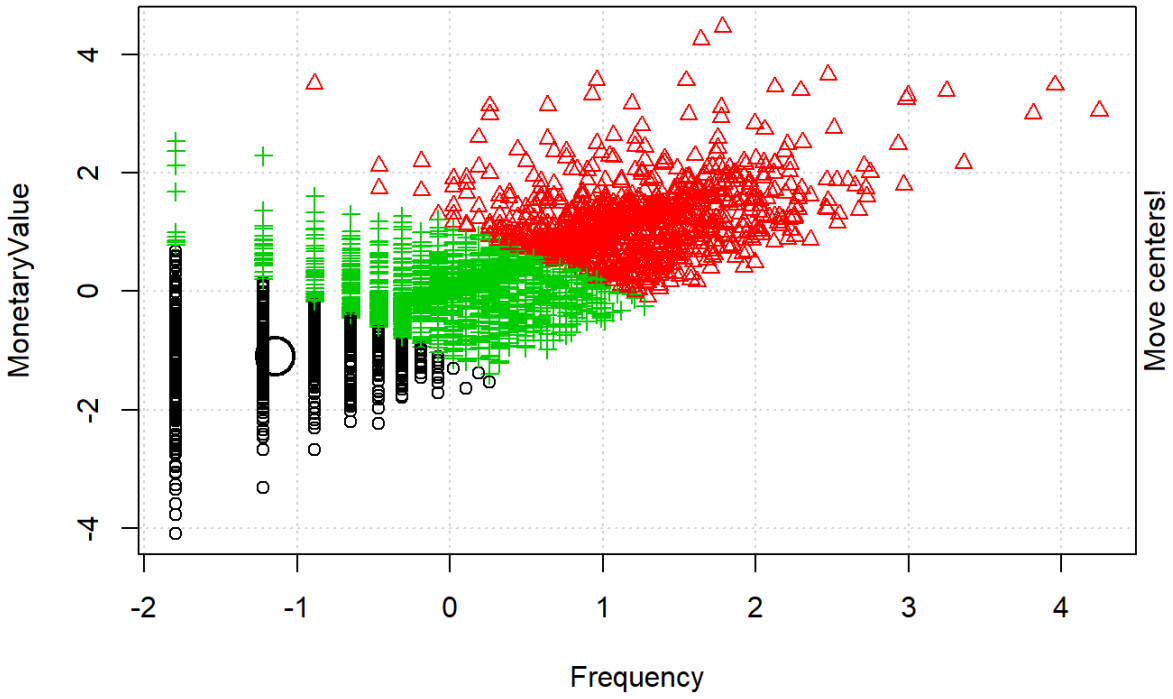

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

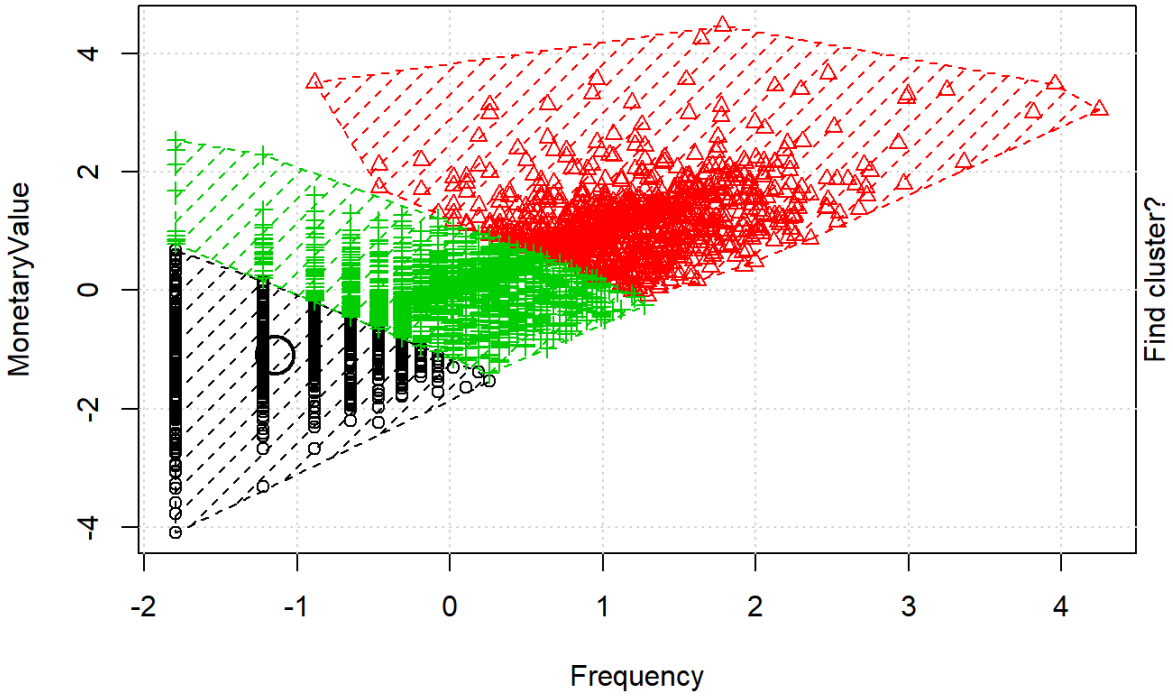

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

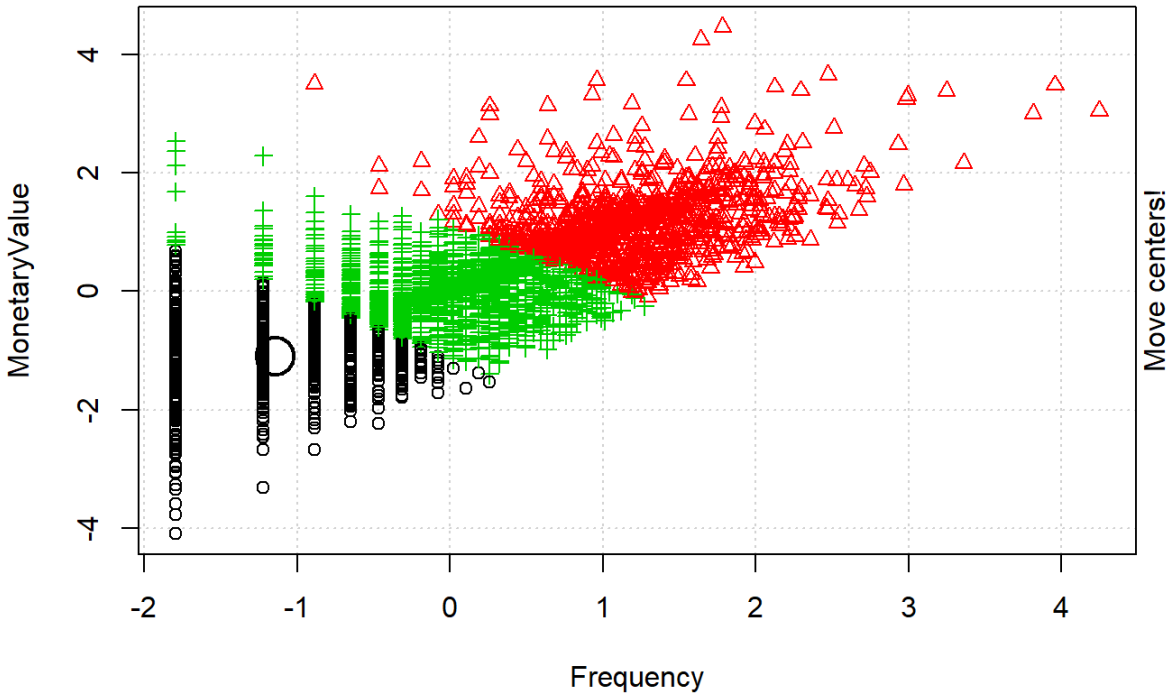

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

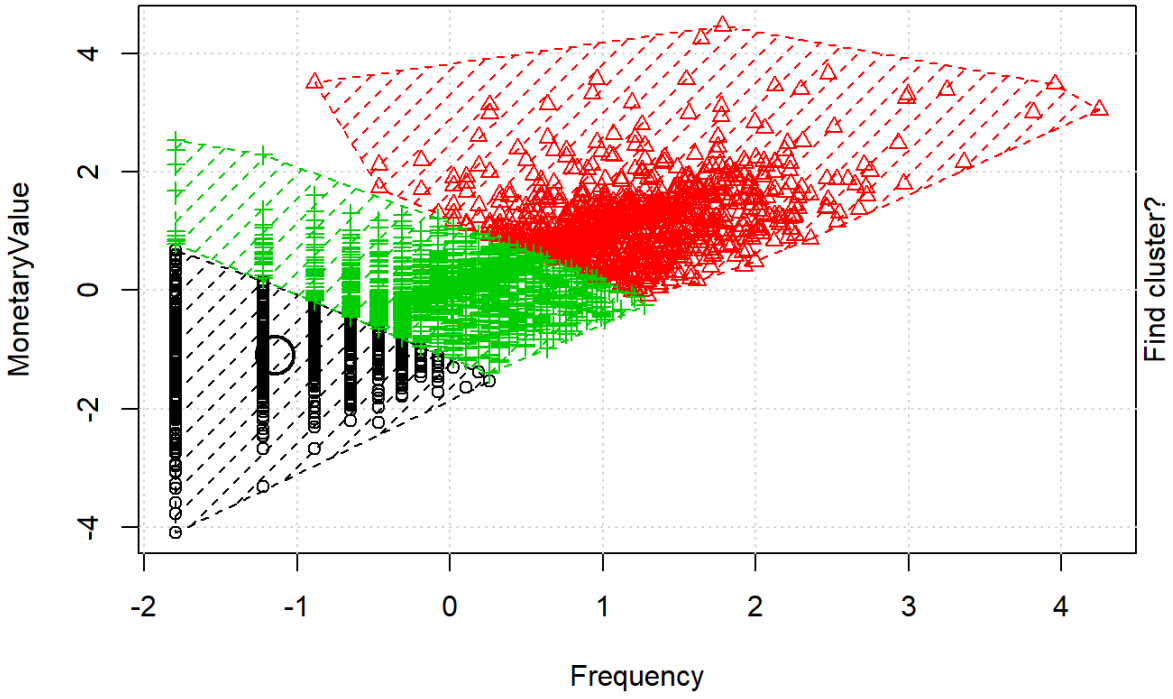

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

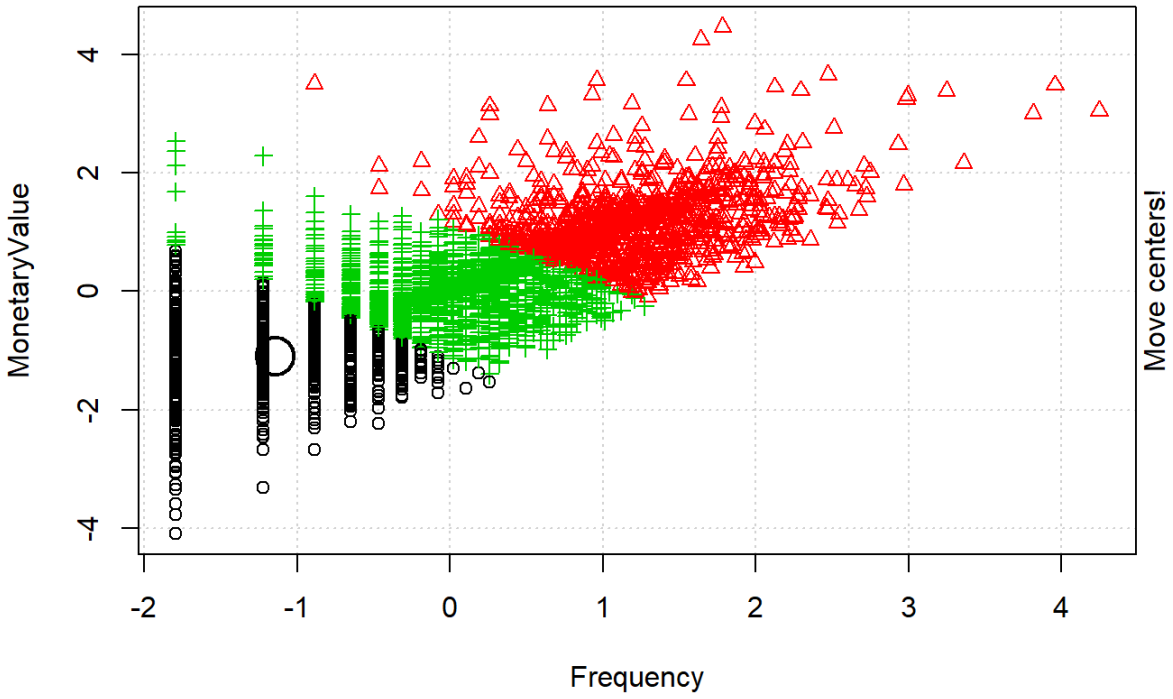

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

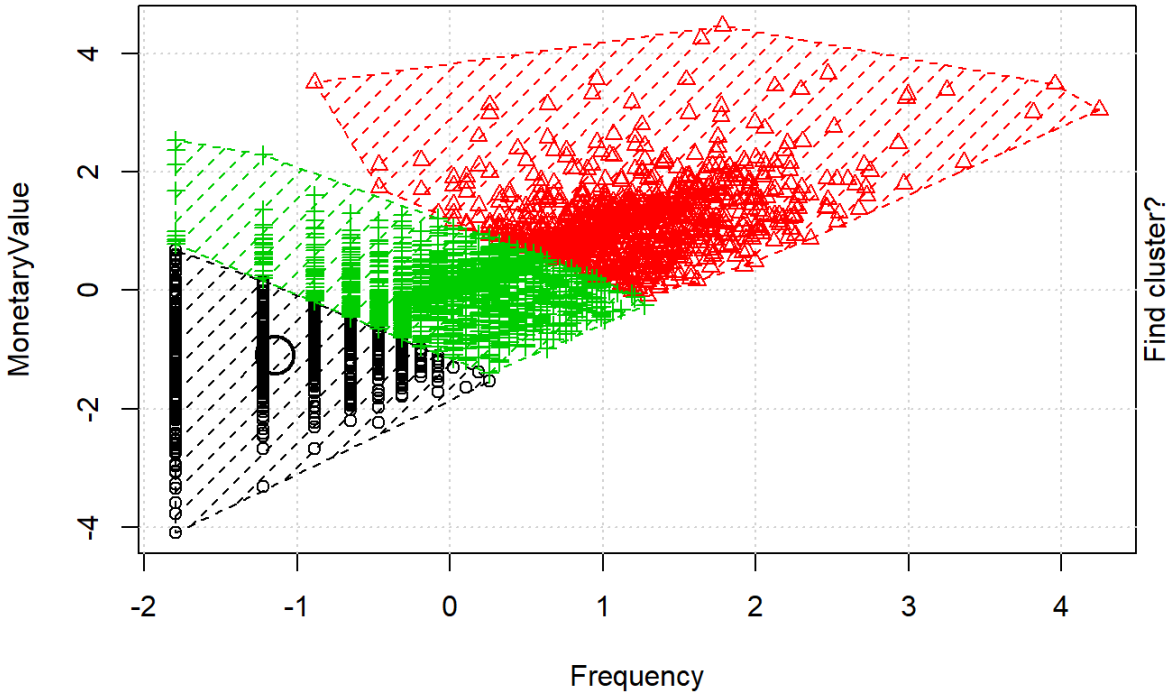

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

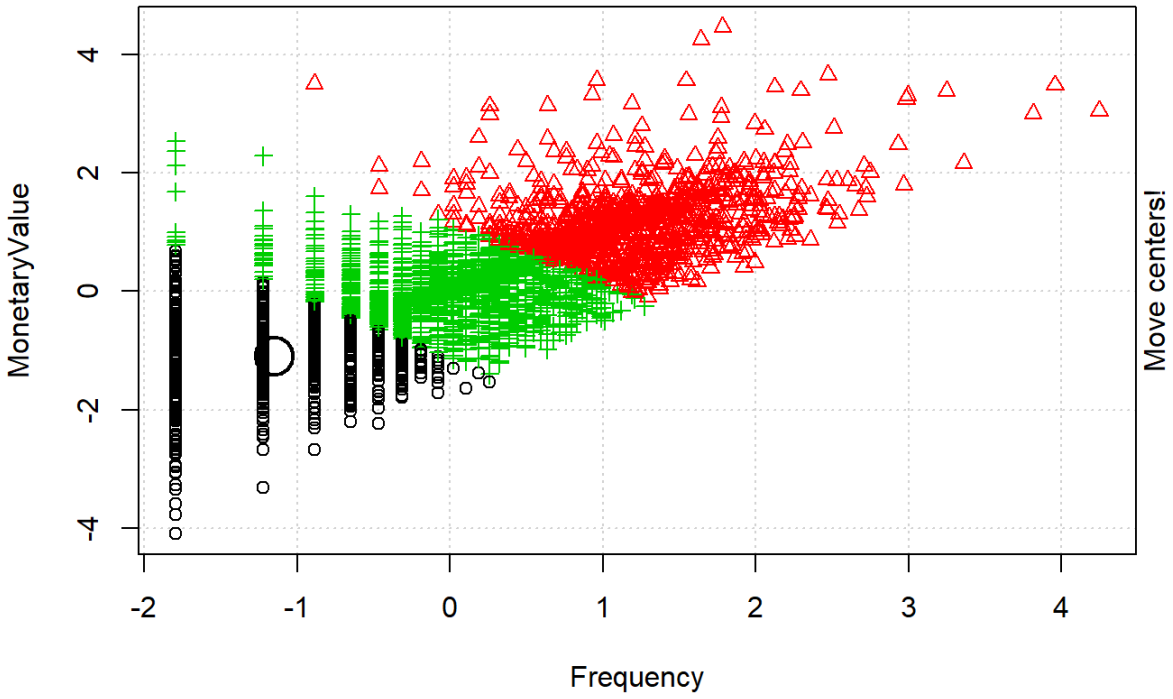

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

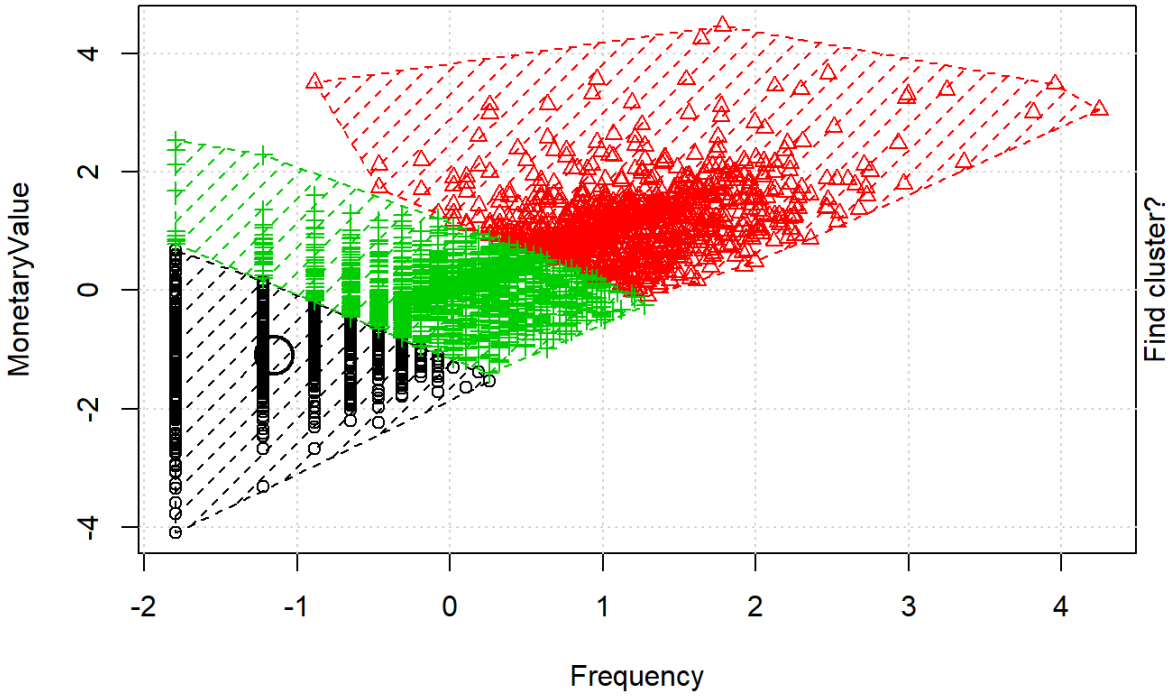

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

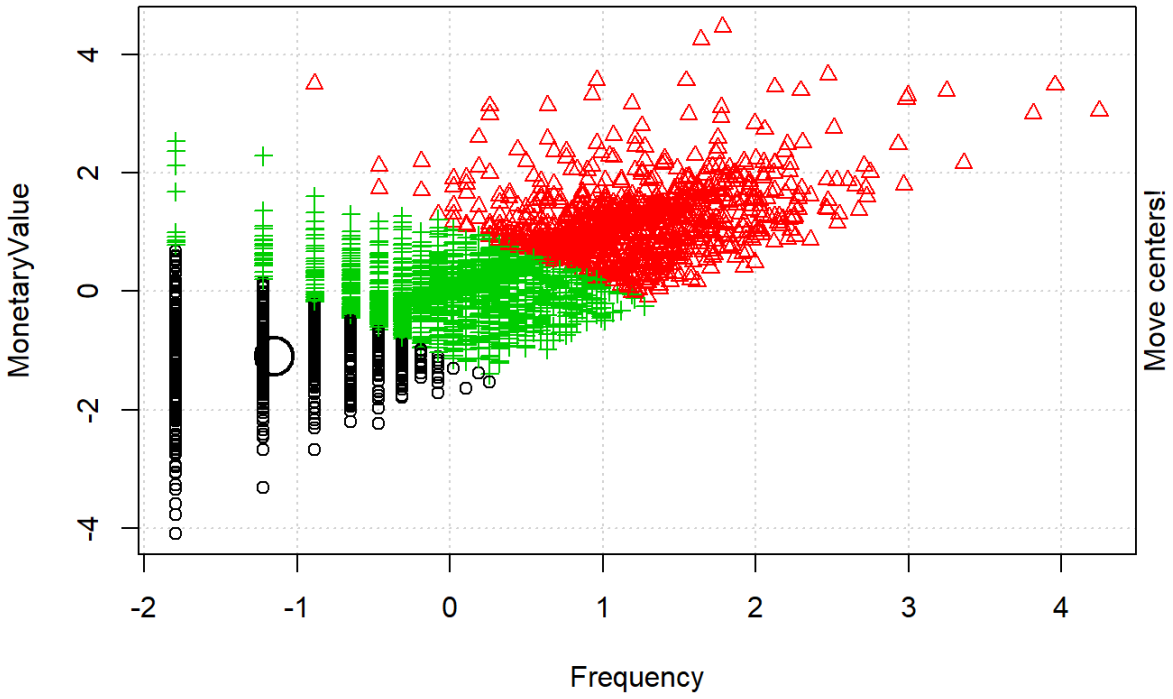

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

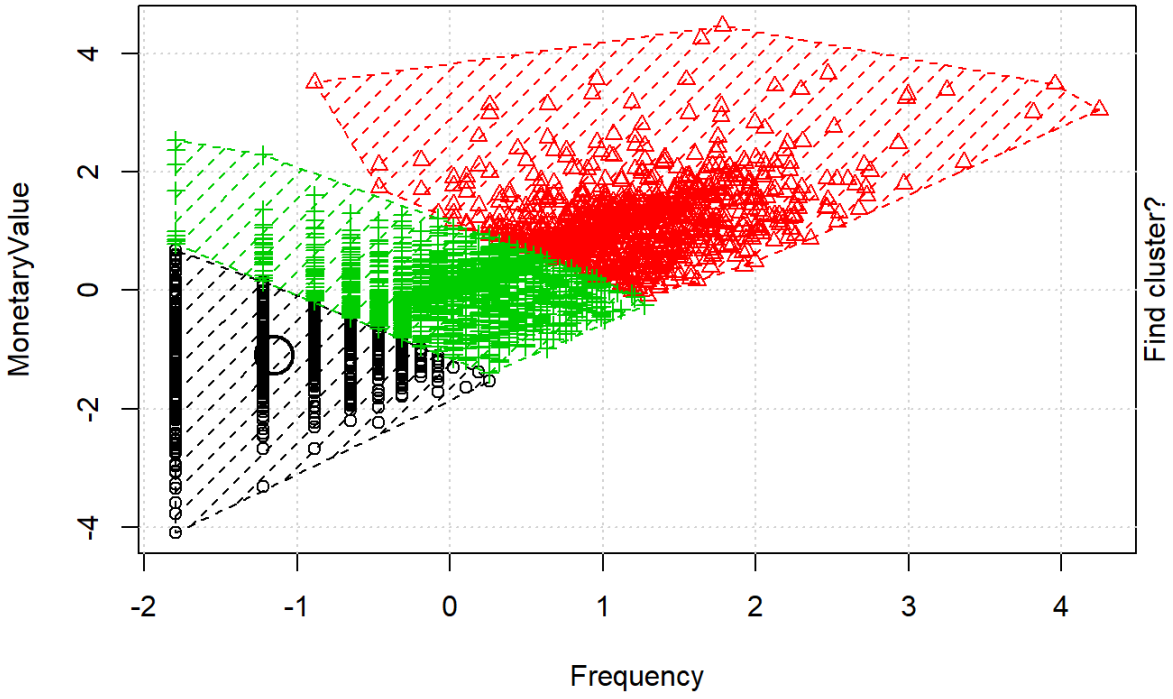

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

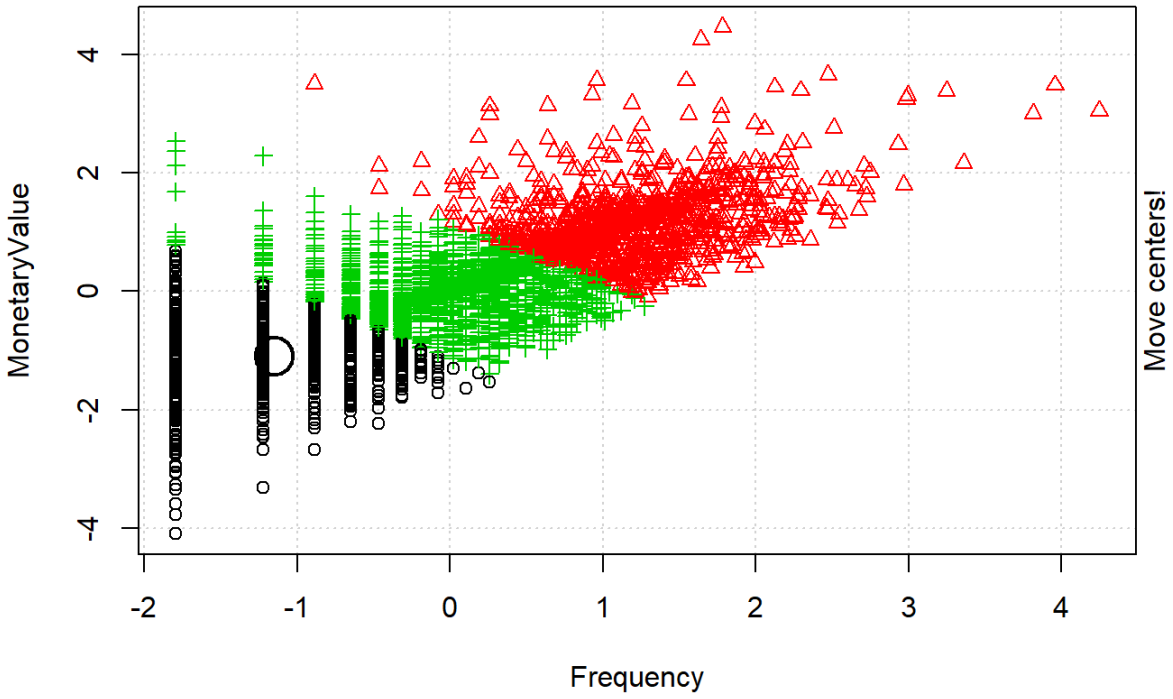

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

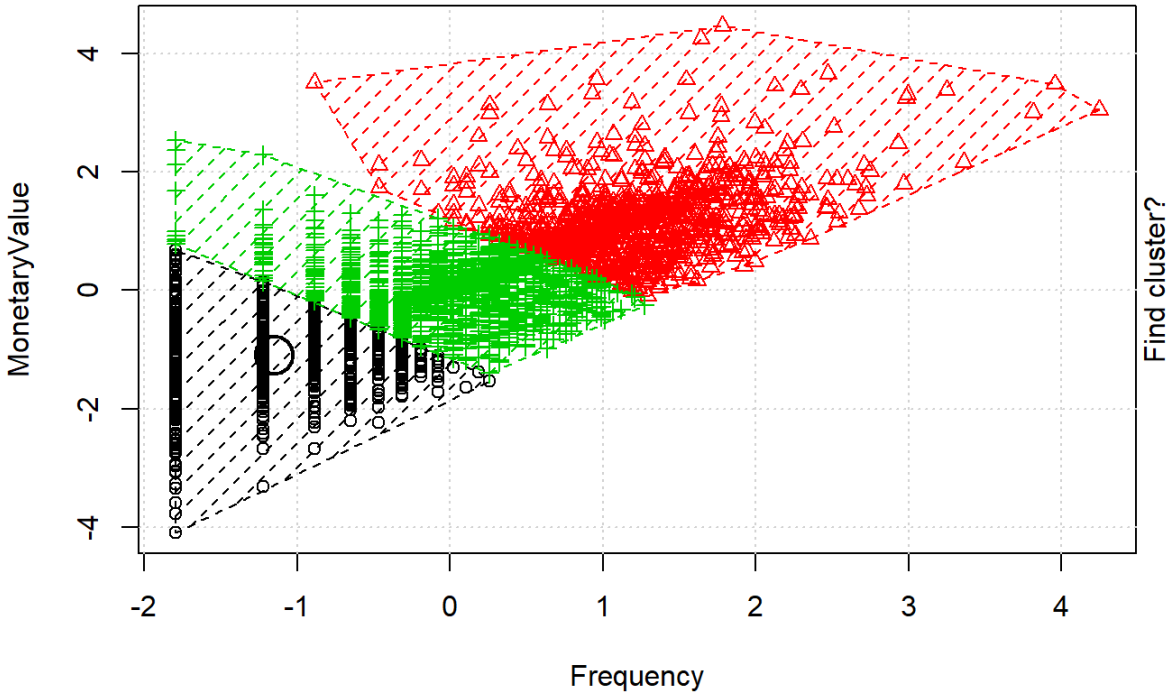

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

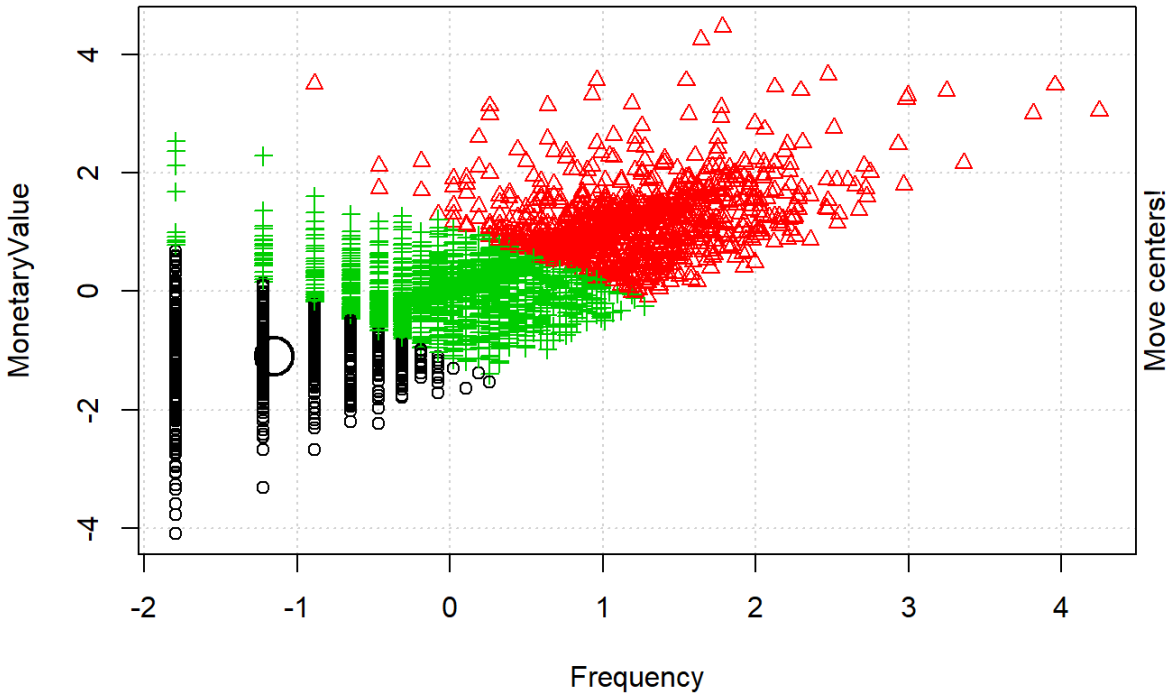

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

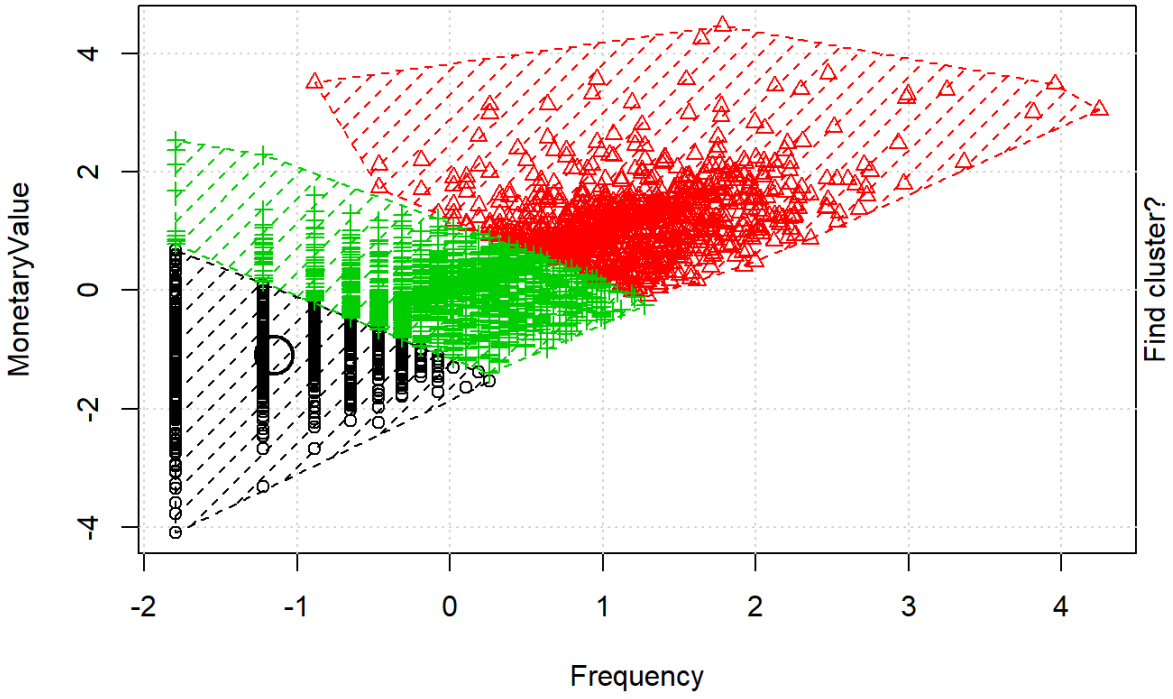

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

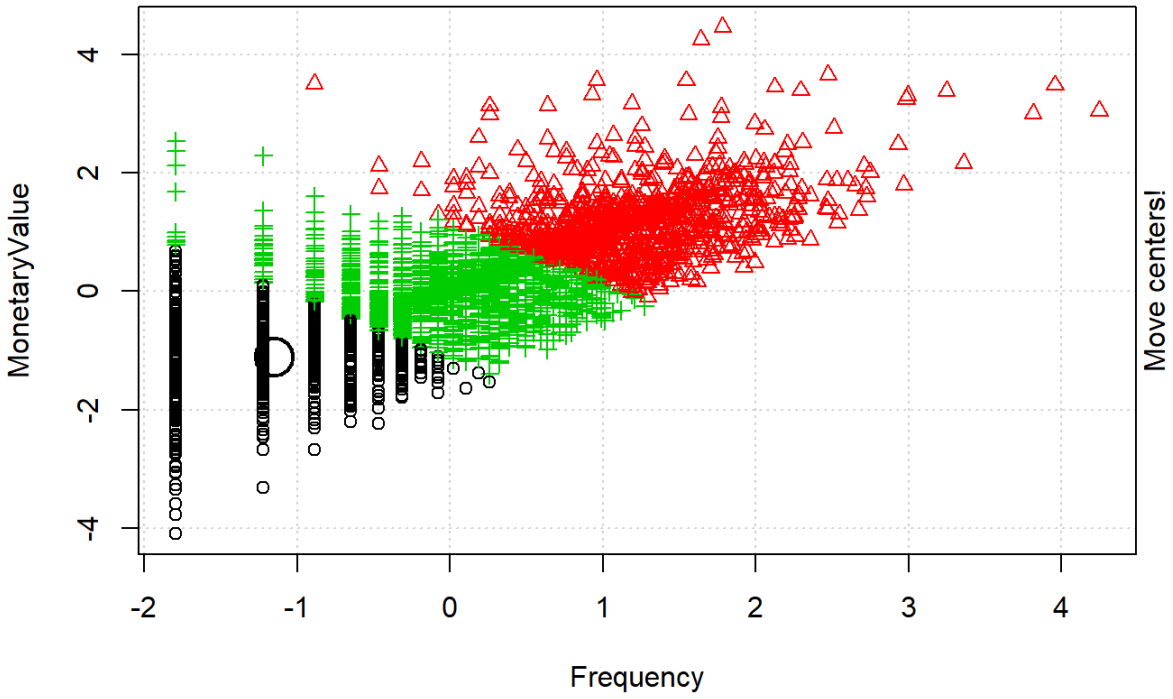

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

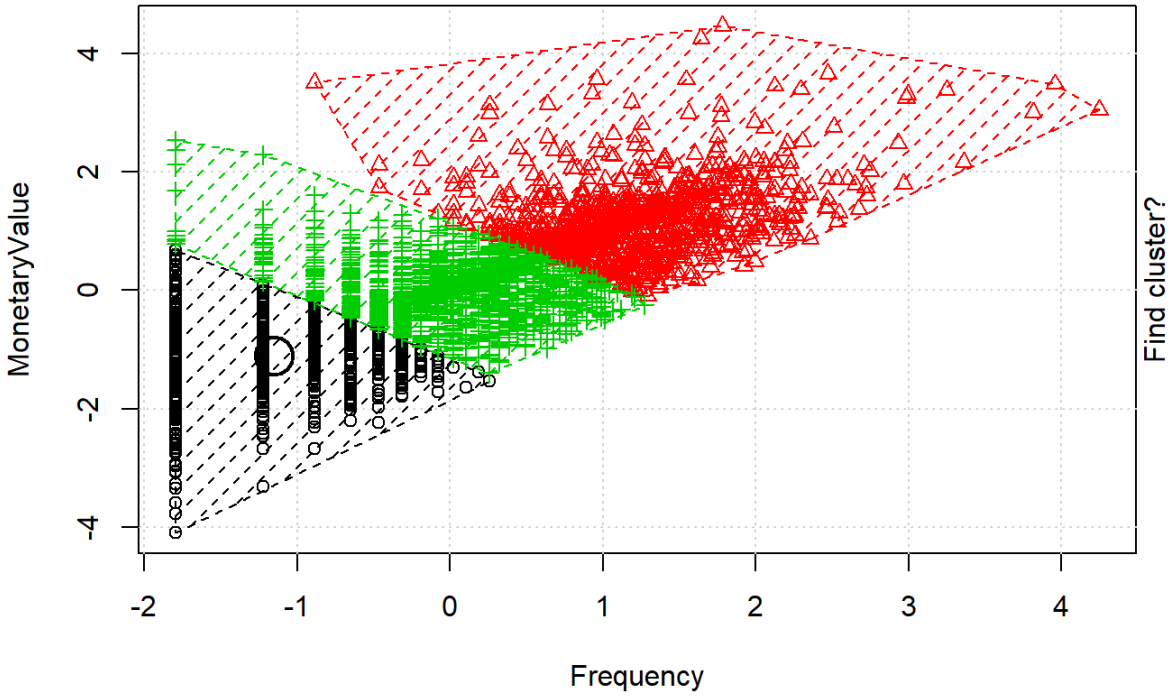

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4


MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

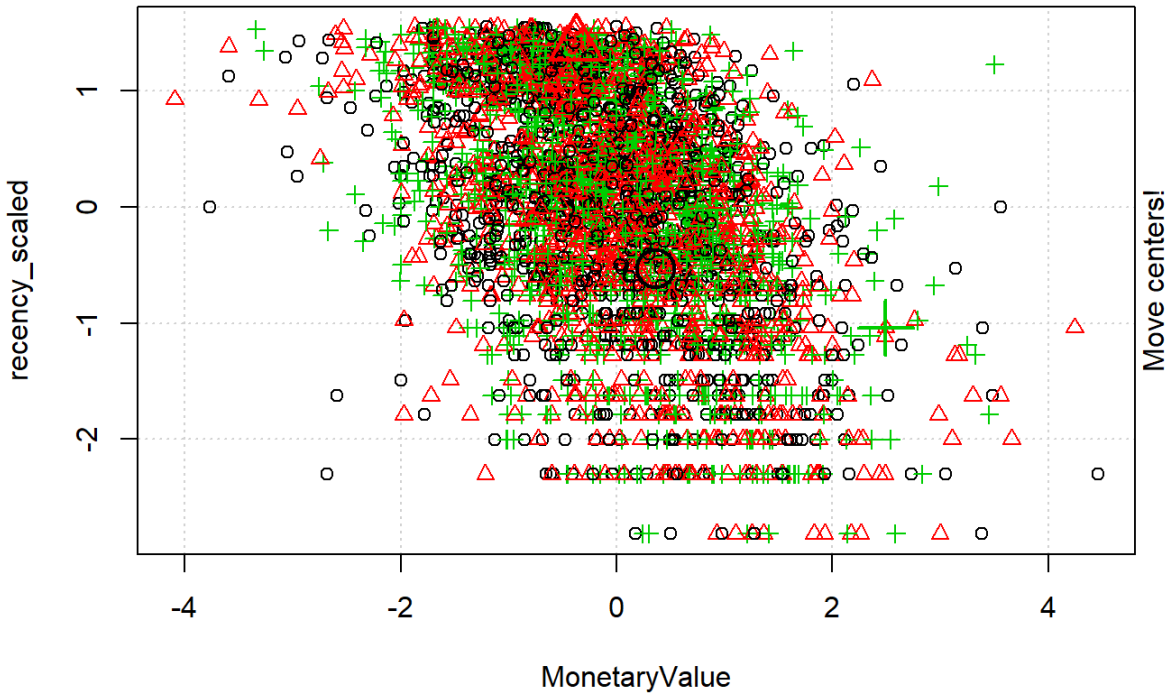

MonetaryValue
Frequency
Move centers!
-2
-1
0
1
2
3
4
-4
-2
0
2
4

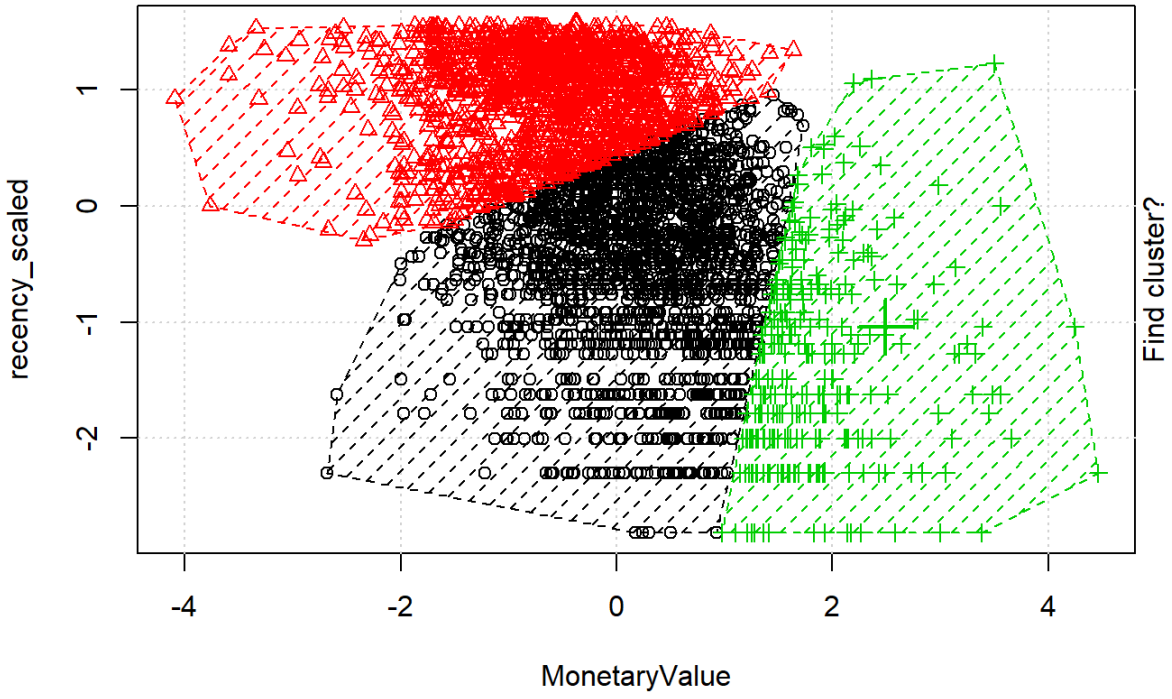

MonetaryValue
Frequency
Find cluster?
-2
-1
0
1
2
3
4
-4
-2
0
2
4

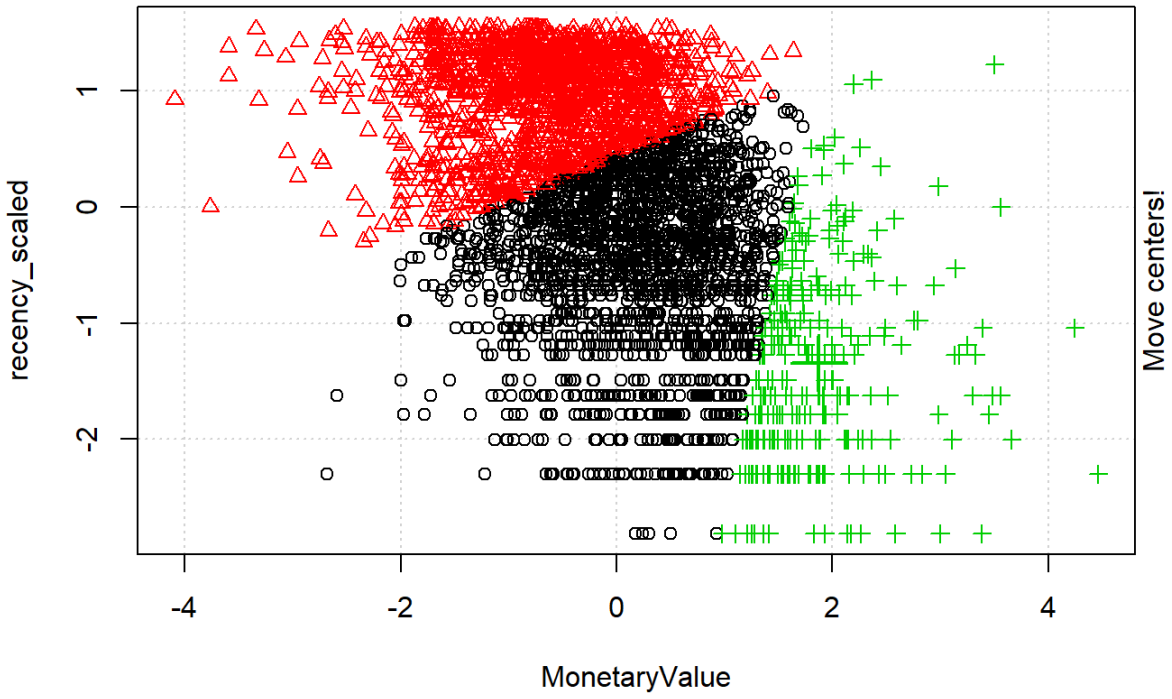


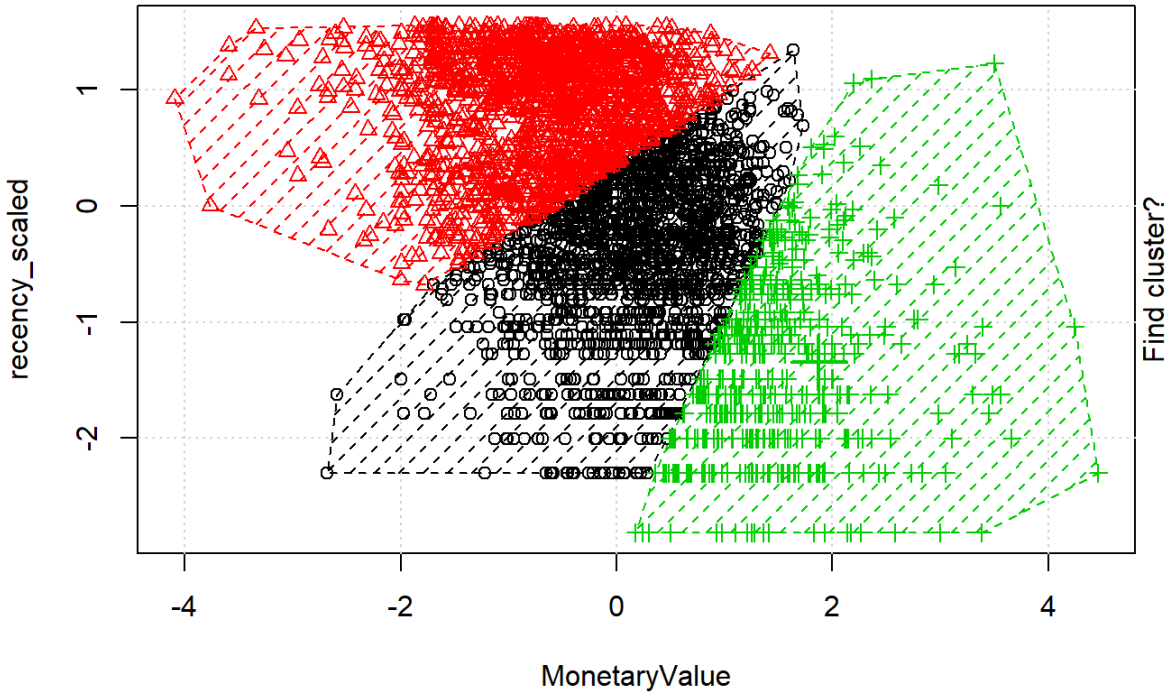


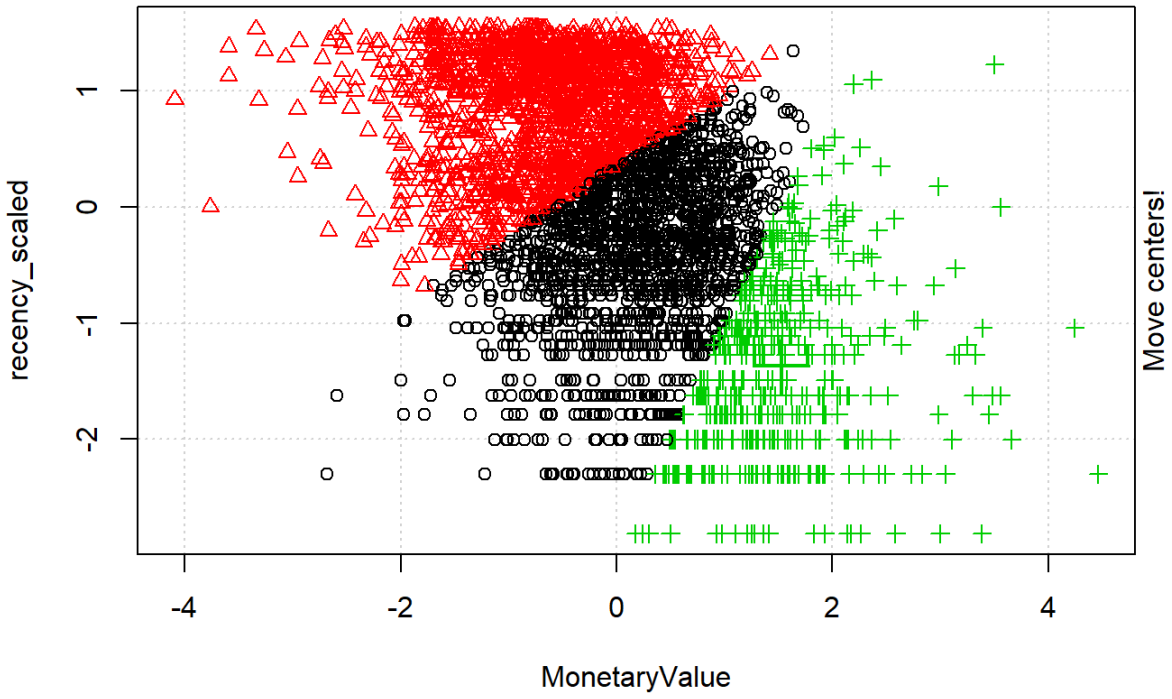

```
kmeans.ani(rescale_subscribermodel[3:4], 3)
```

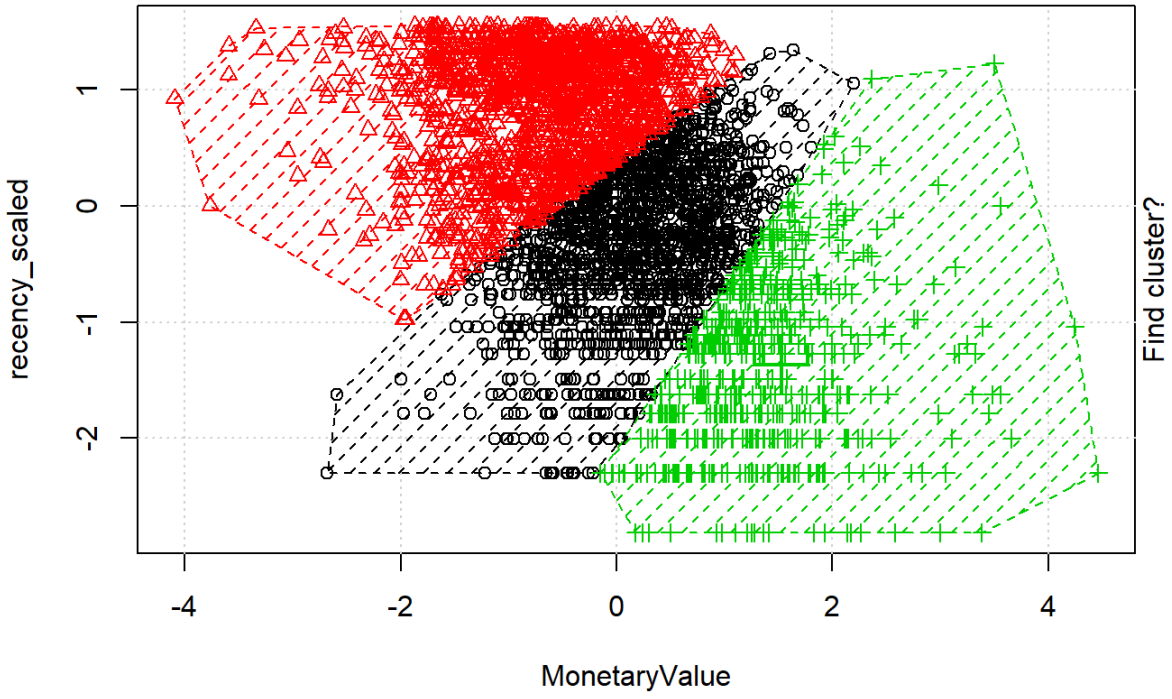

recency_scaled
MonetaryValue
Move centers!
1
0
-1
-2
-4
-2
0
2
4

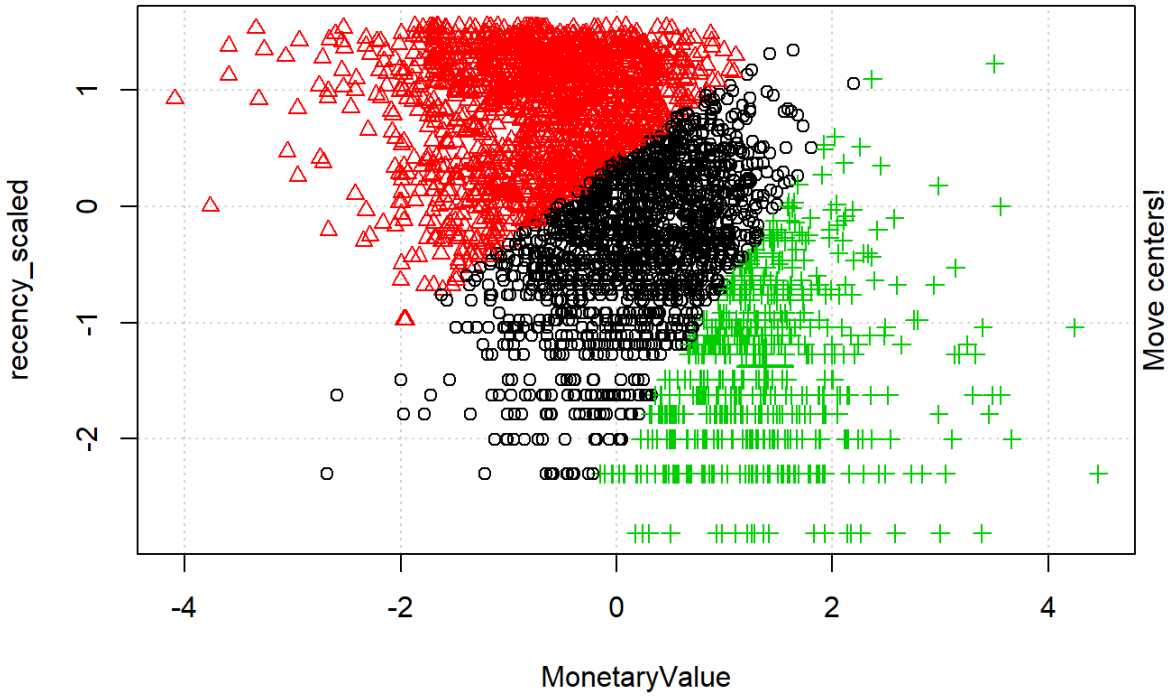

recency_scaled
MonetaryValue
Find cluster?
1
0
-1
-2
-4
-2
0
2
4

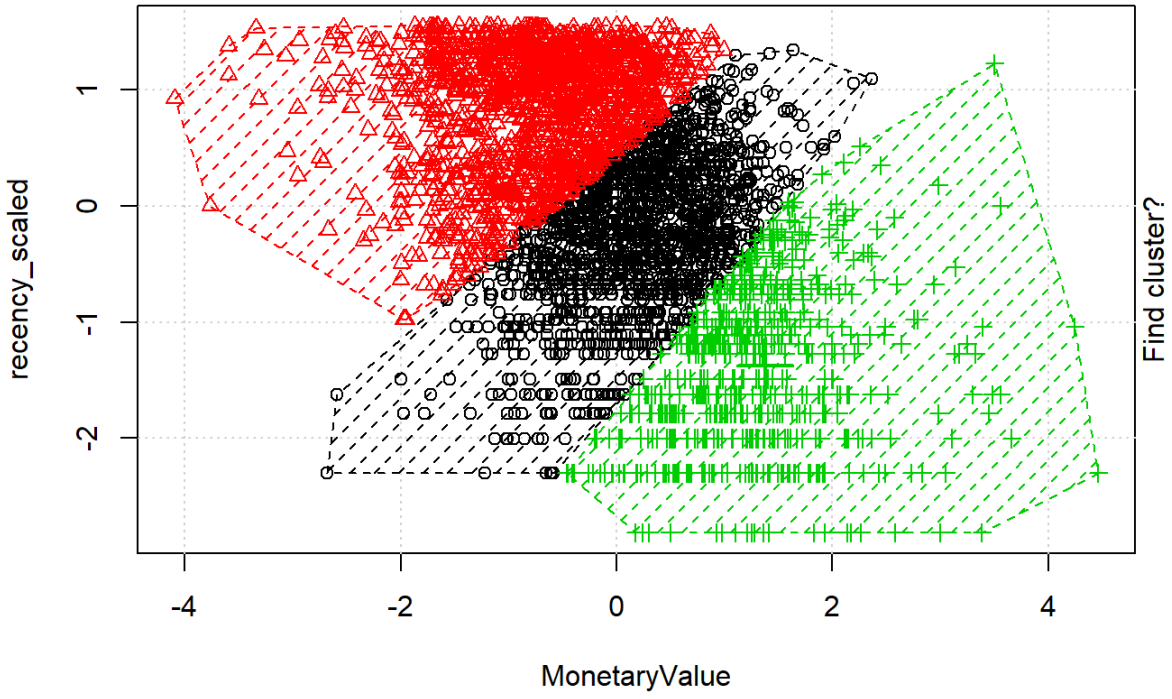

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

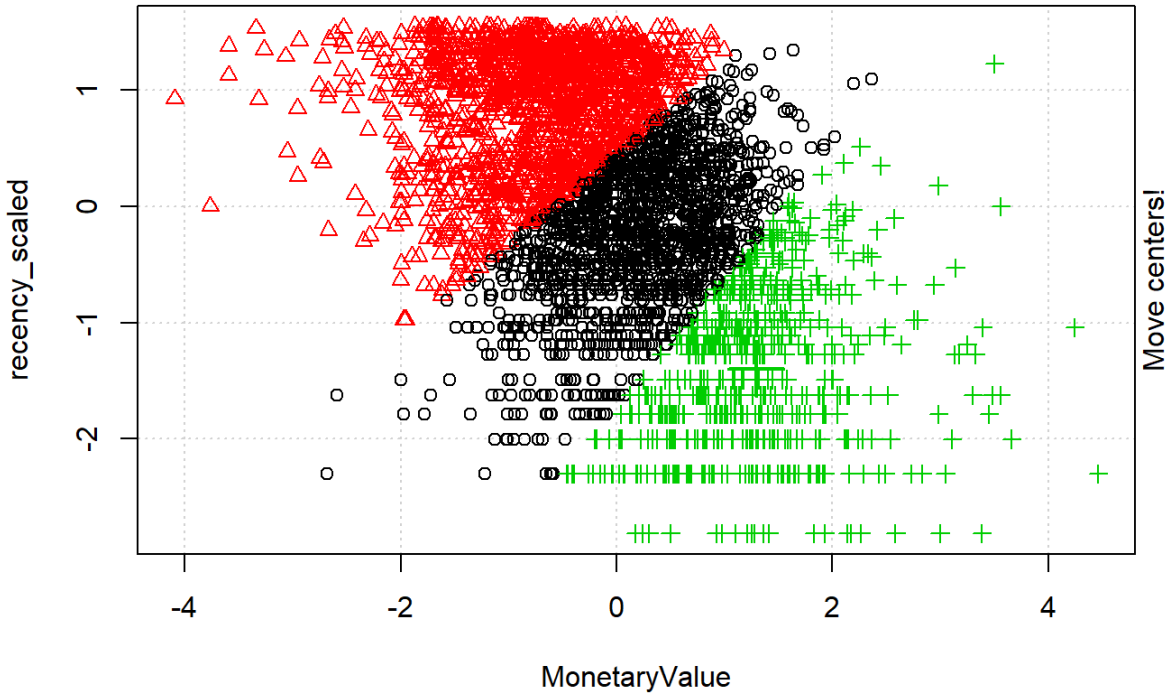

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

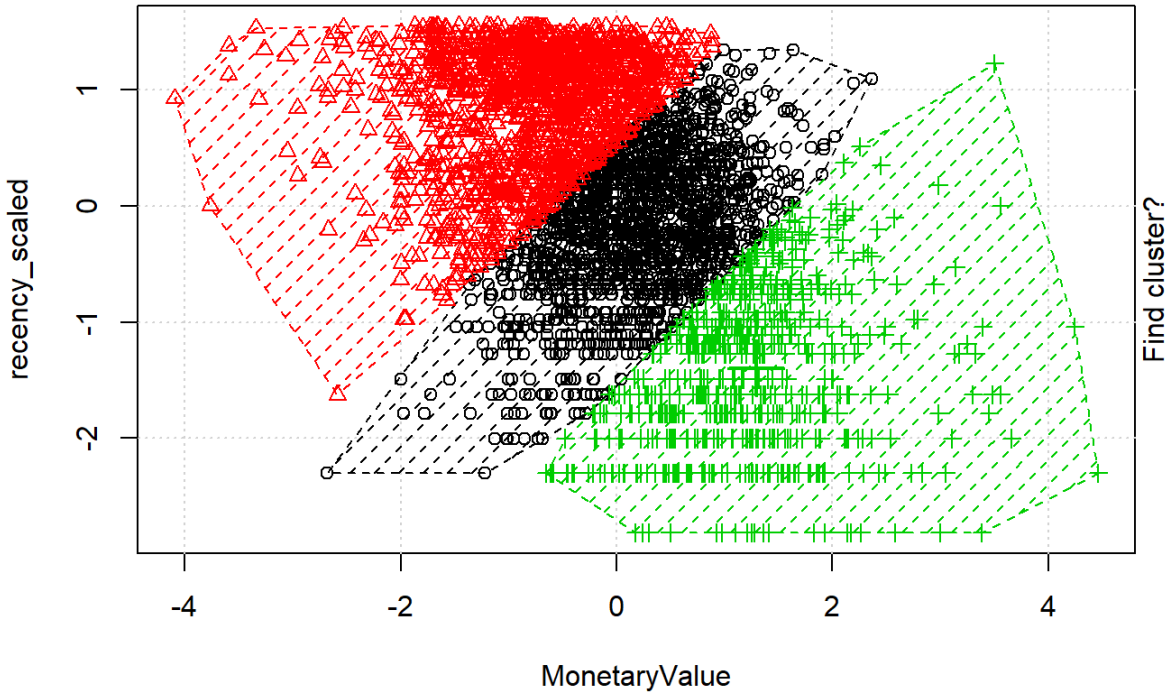

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

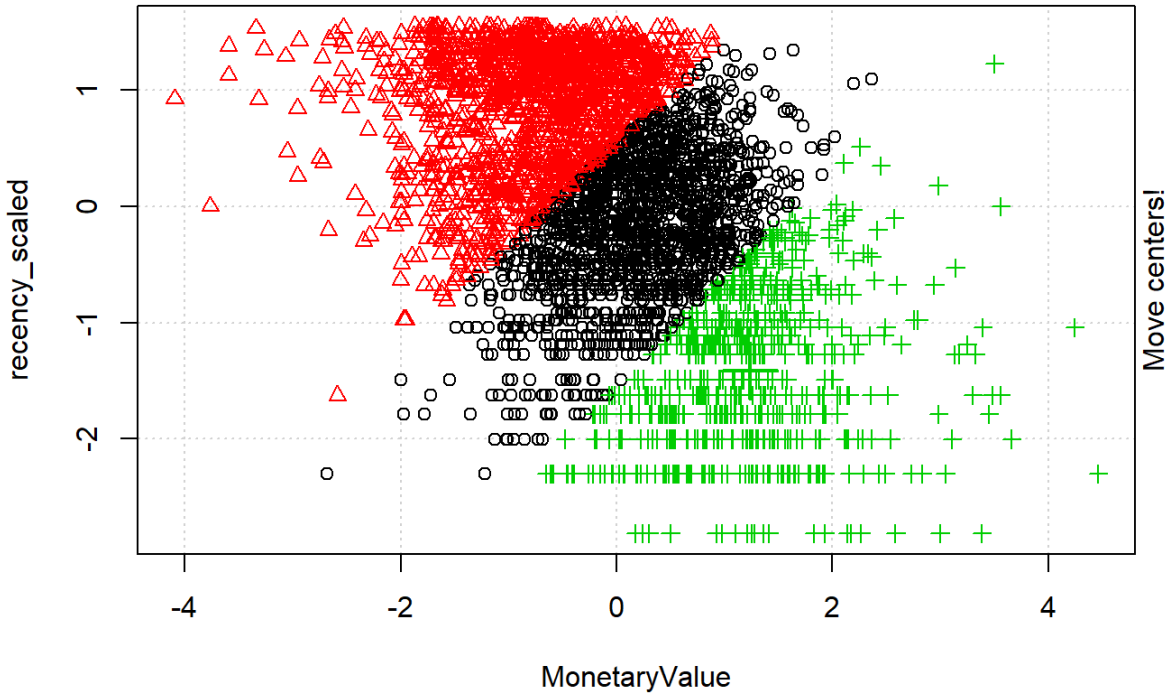

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

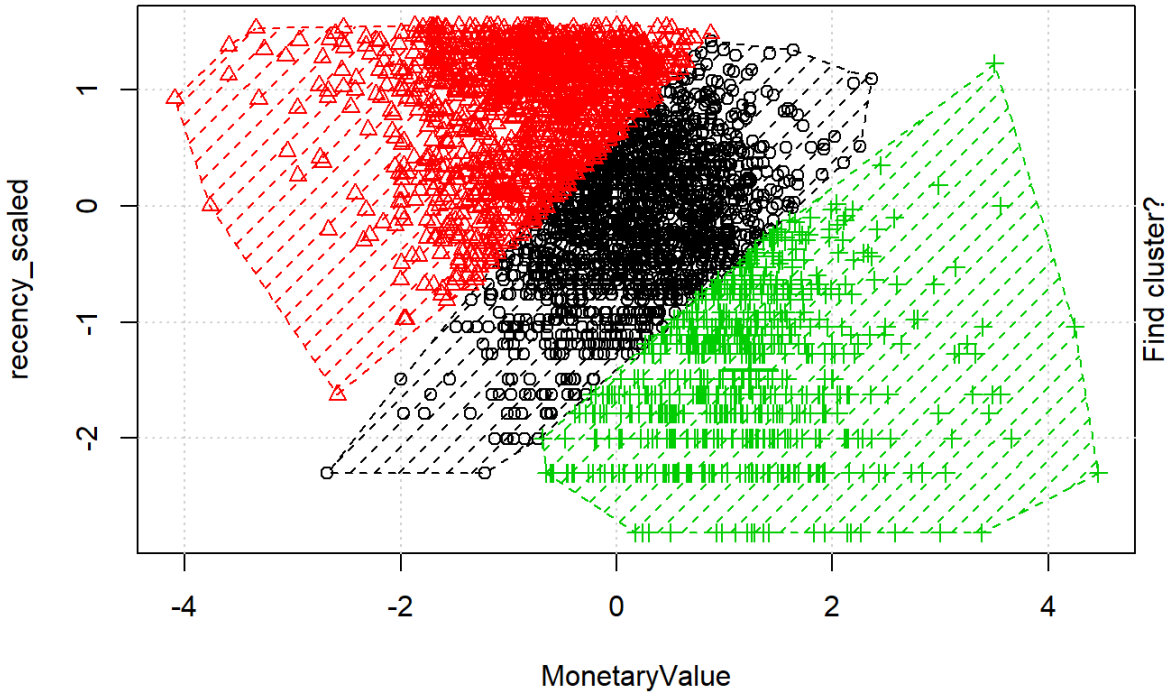

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

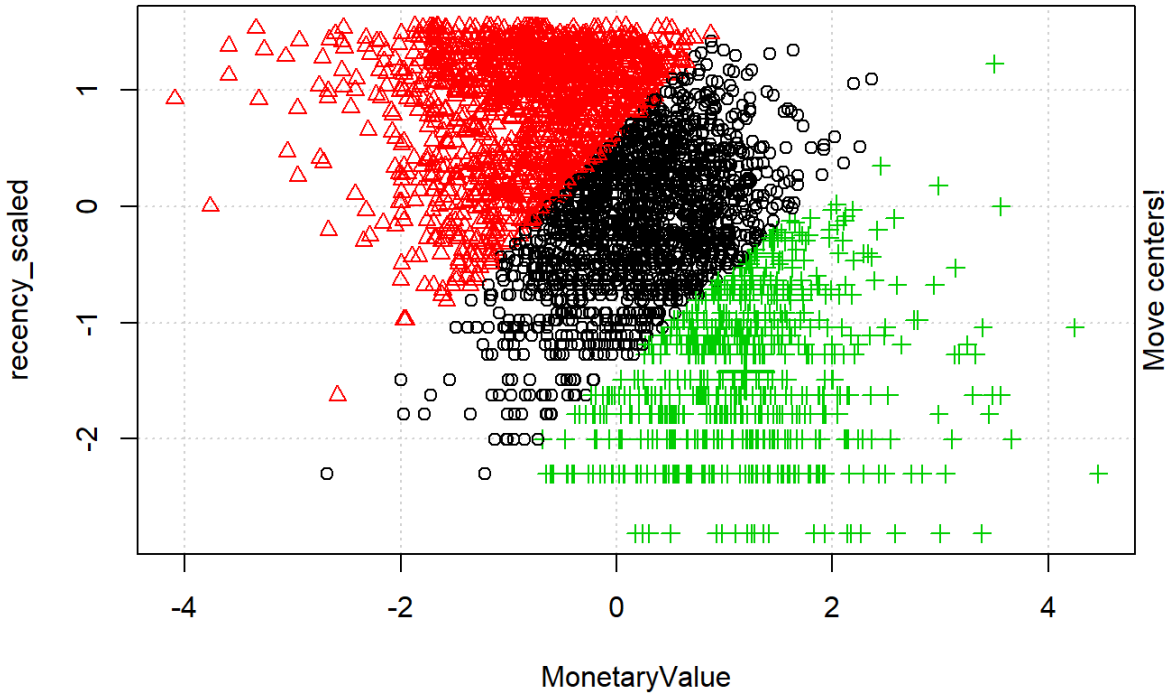

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

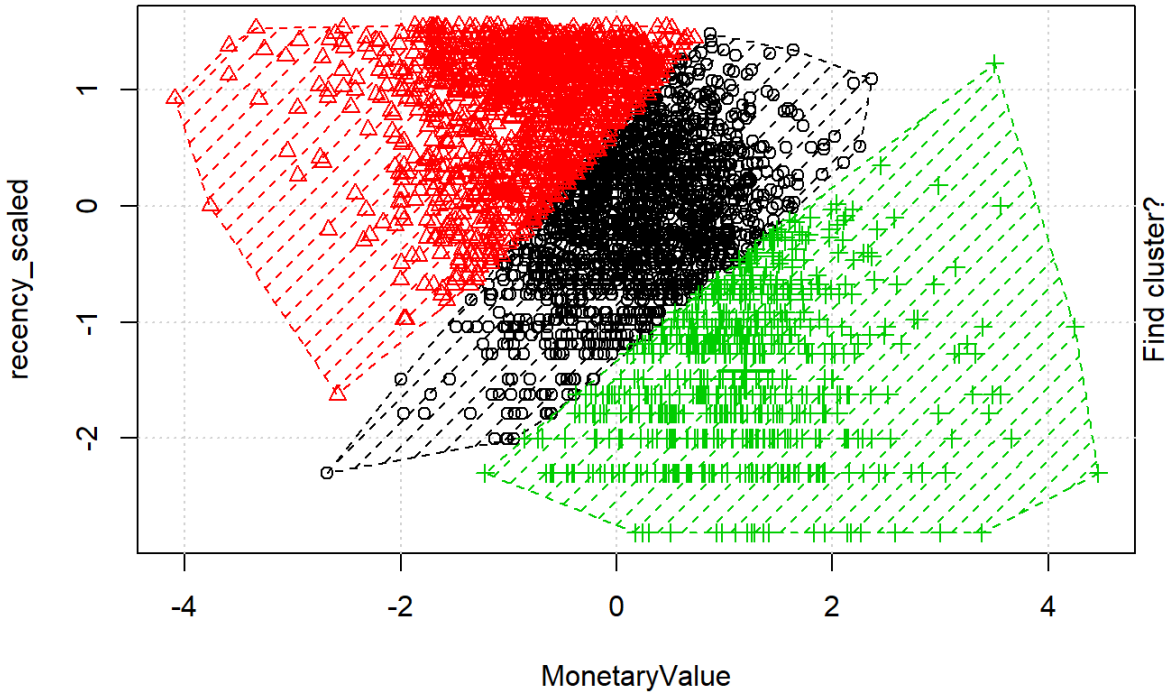

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

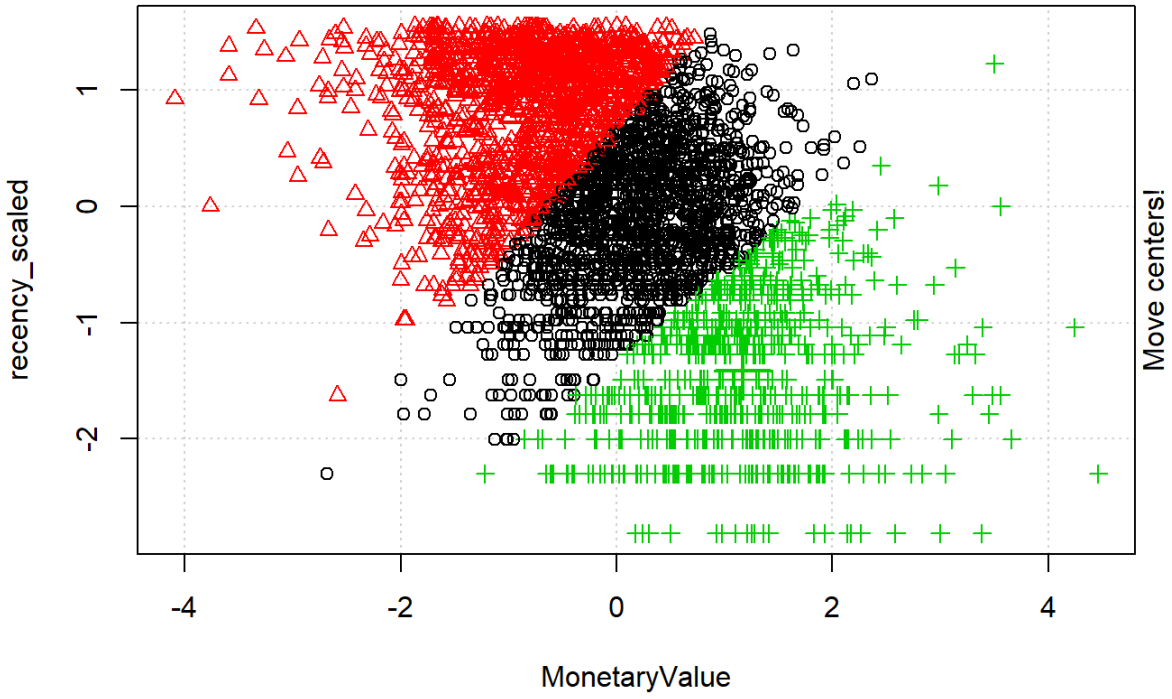

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

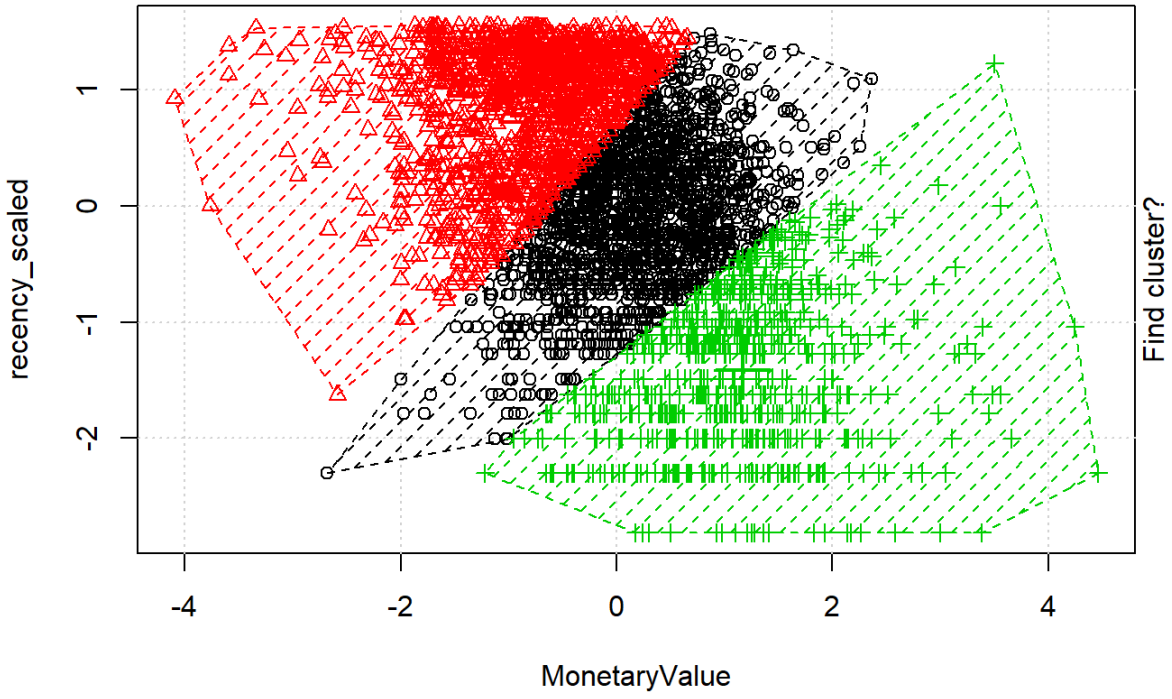

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

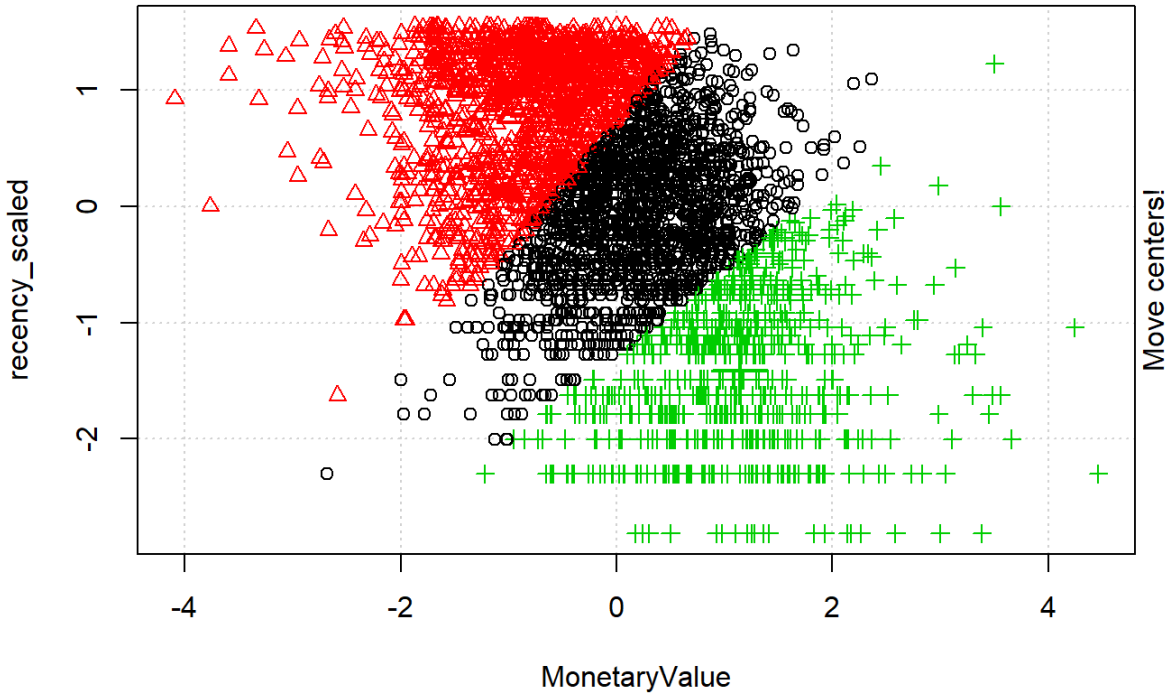

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

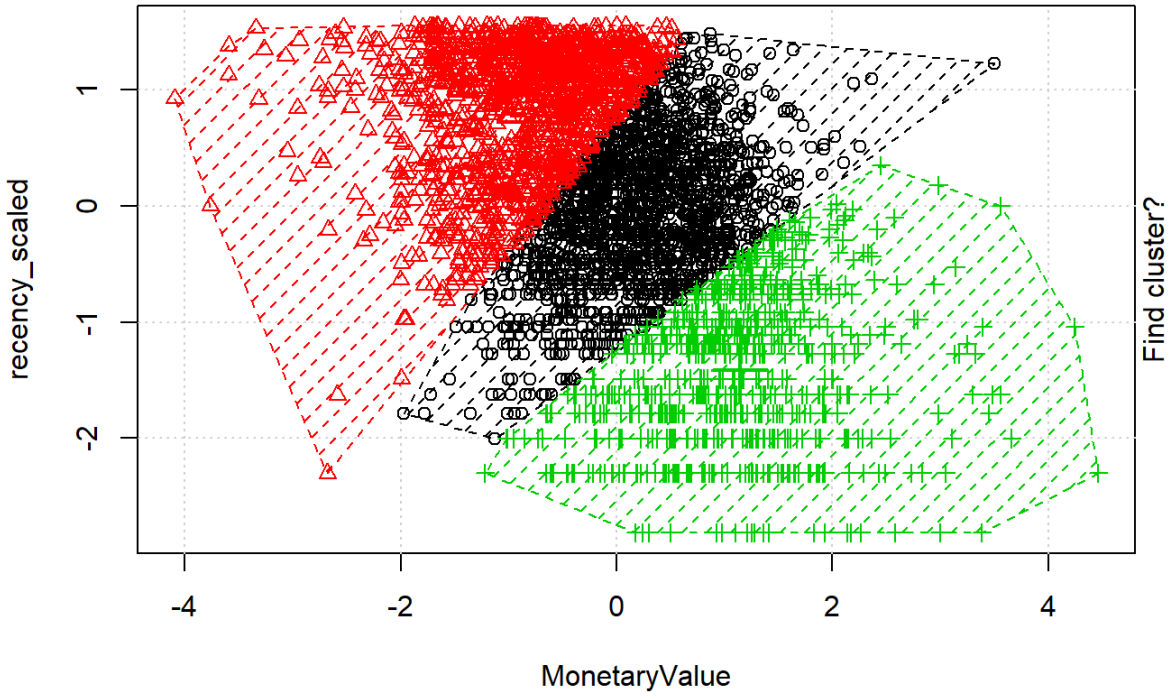

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
1
0
-1
-2

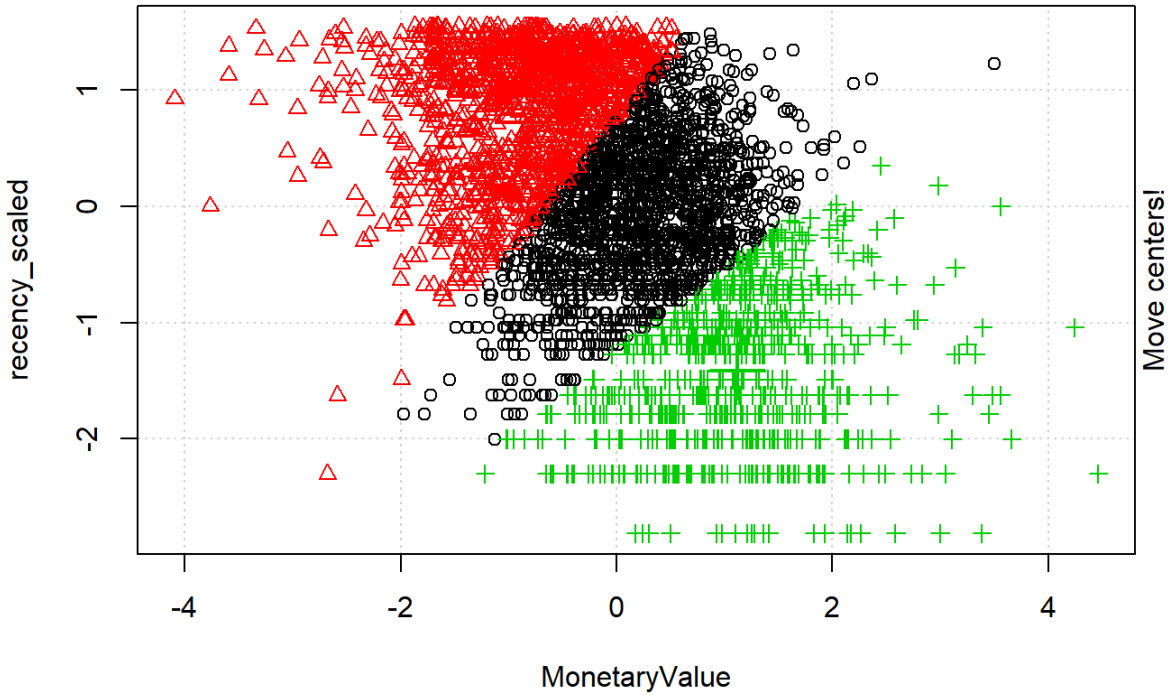

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
1
0
-1
-2

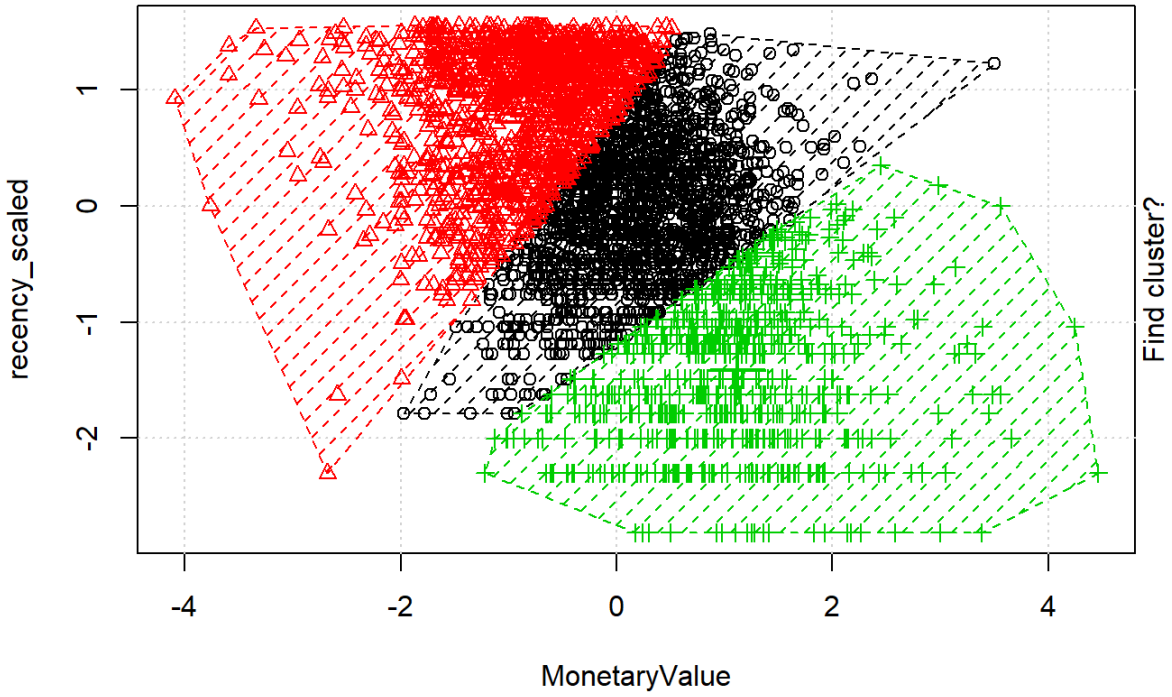

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
1
0
-1
-2

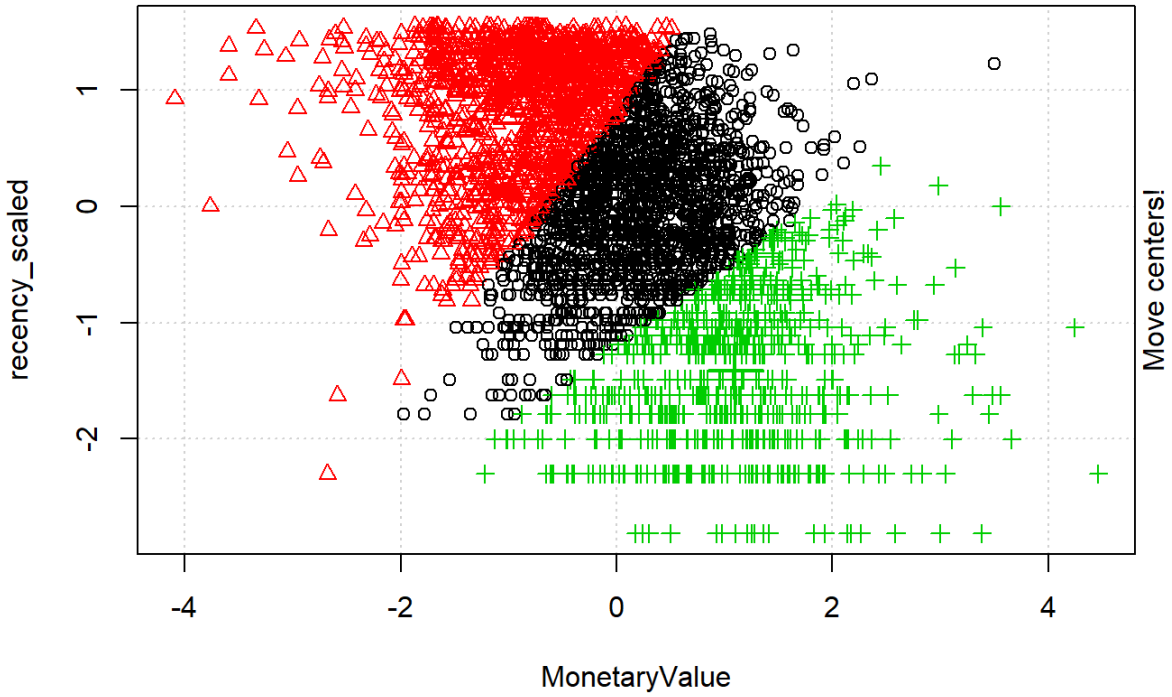

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
1
0
-1
-2

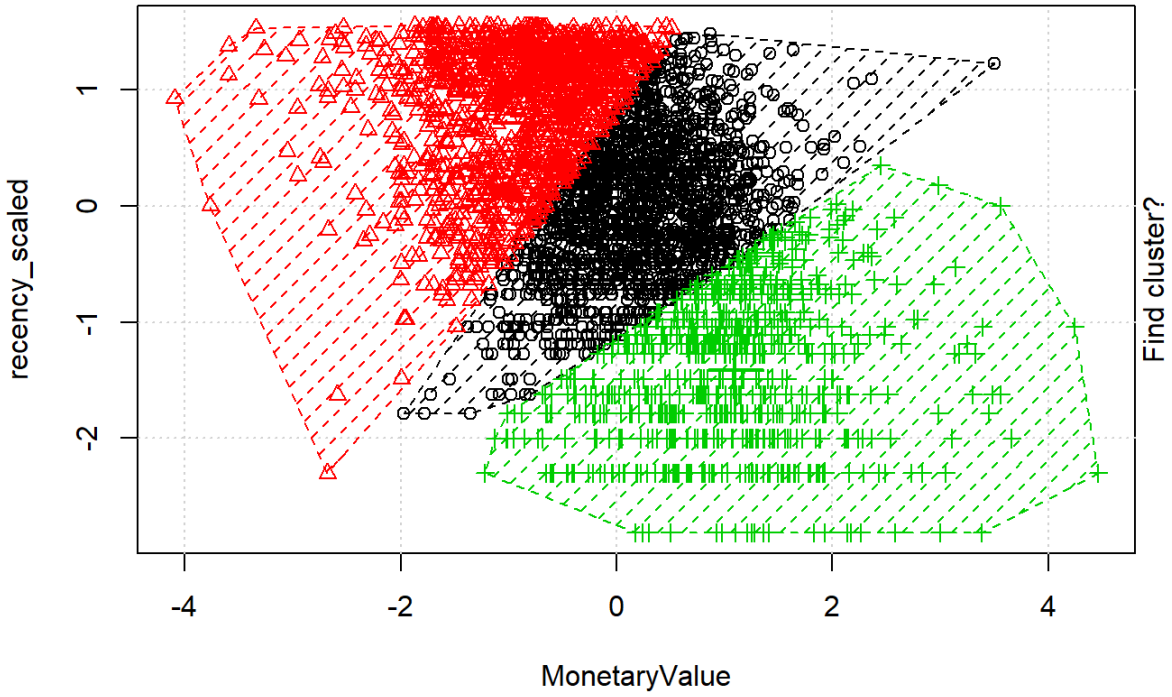

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
1
0
-1
-2

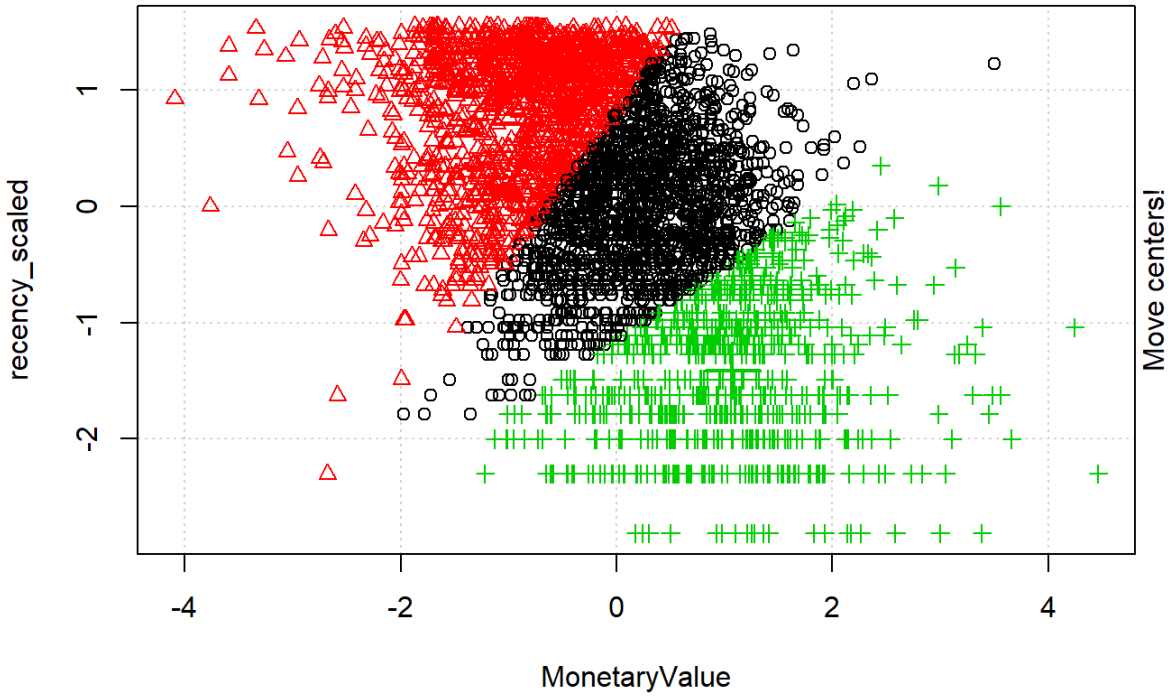

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
1
0
-1
-2

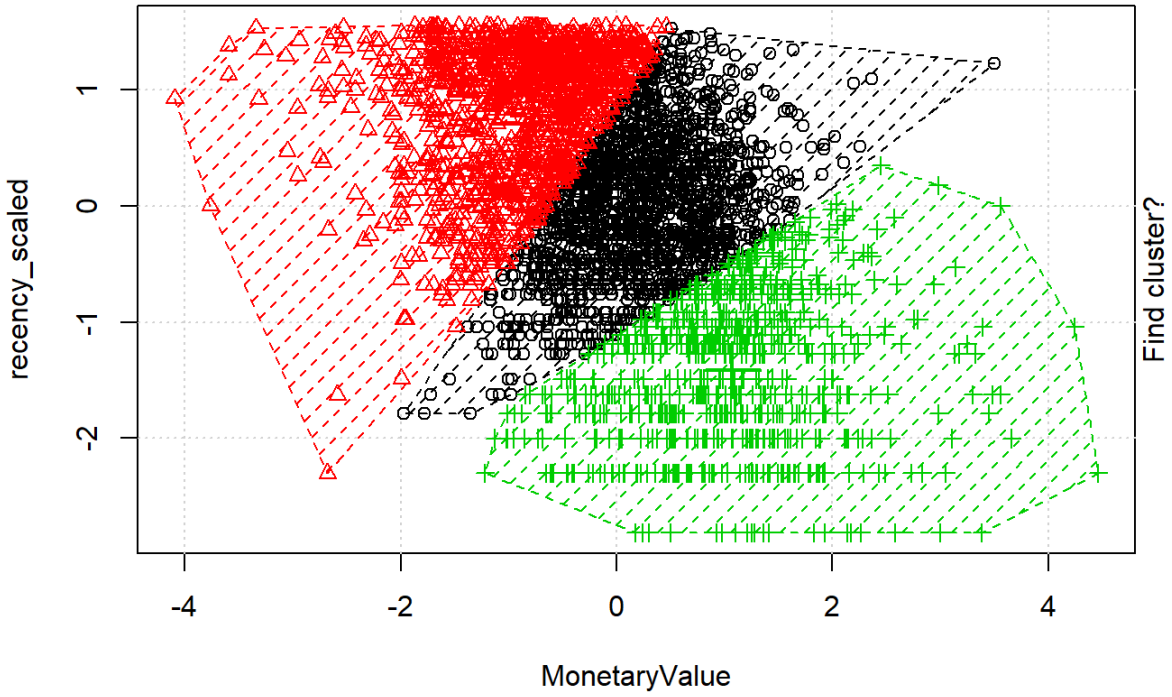

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

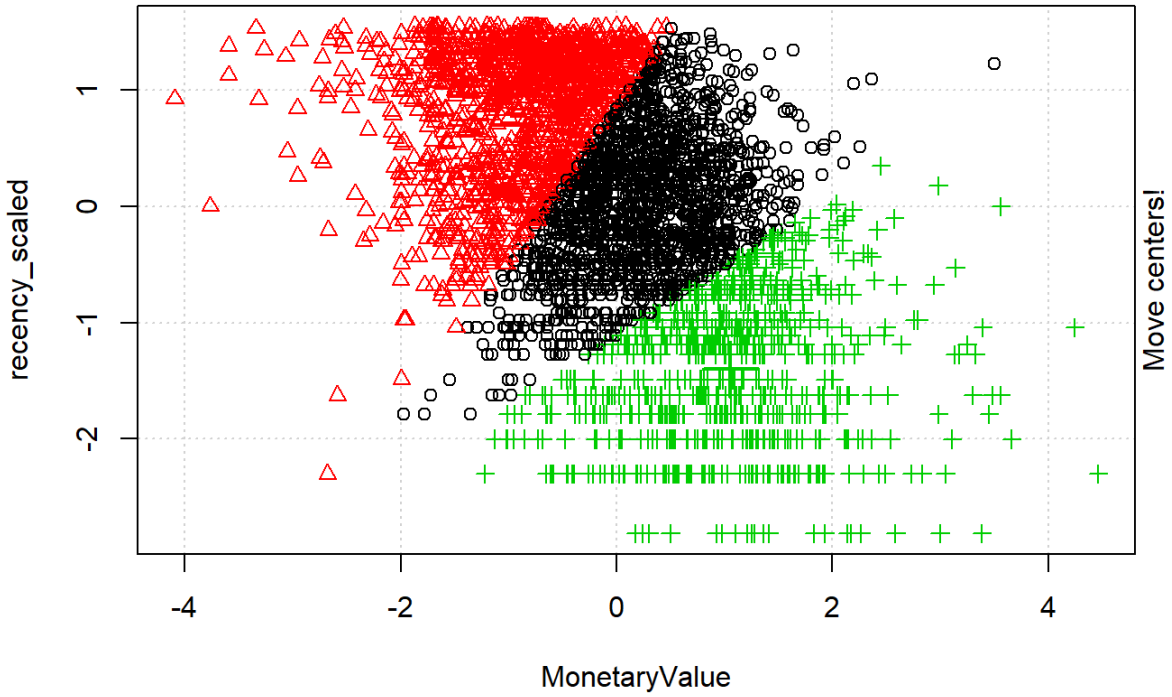

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

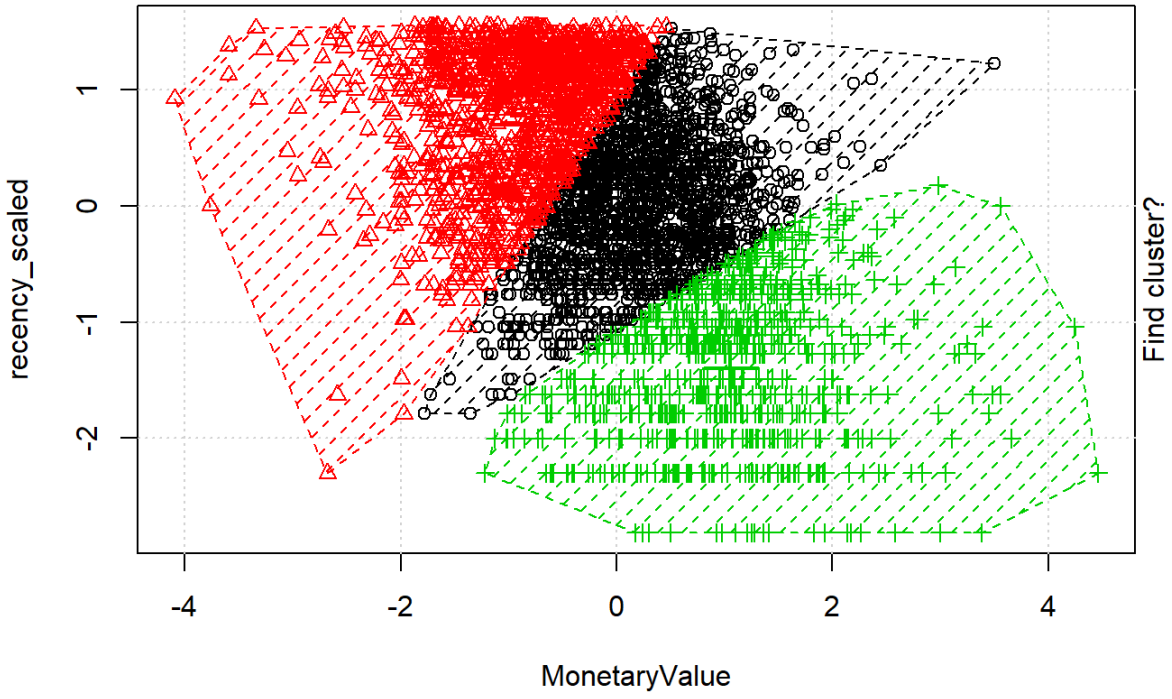

recency_scaled
1
0
-1
-2
-4
-2
0
2
4
MonetaryValue
Move centers!

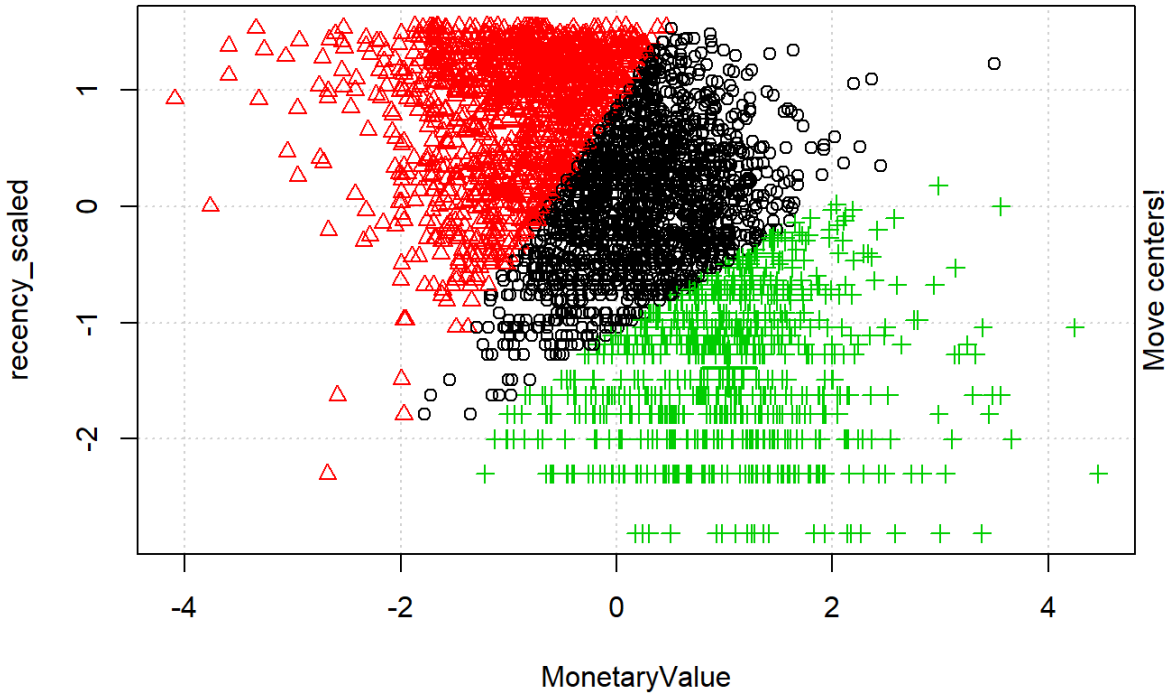

recency_scaled
1
0
-1
-2
-4
-2
0
2
4
MonetaryValue
Find cluster?

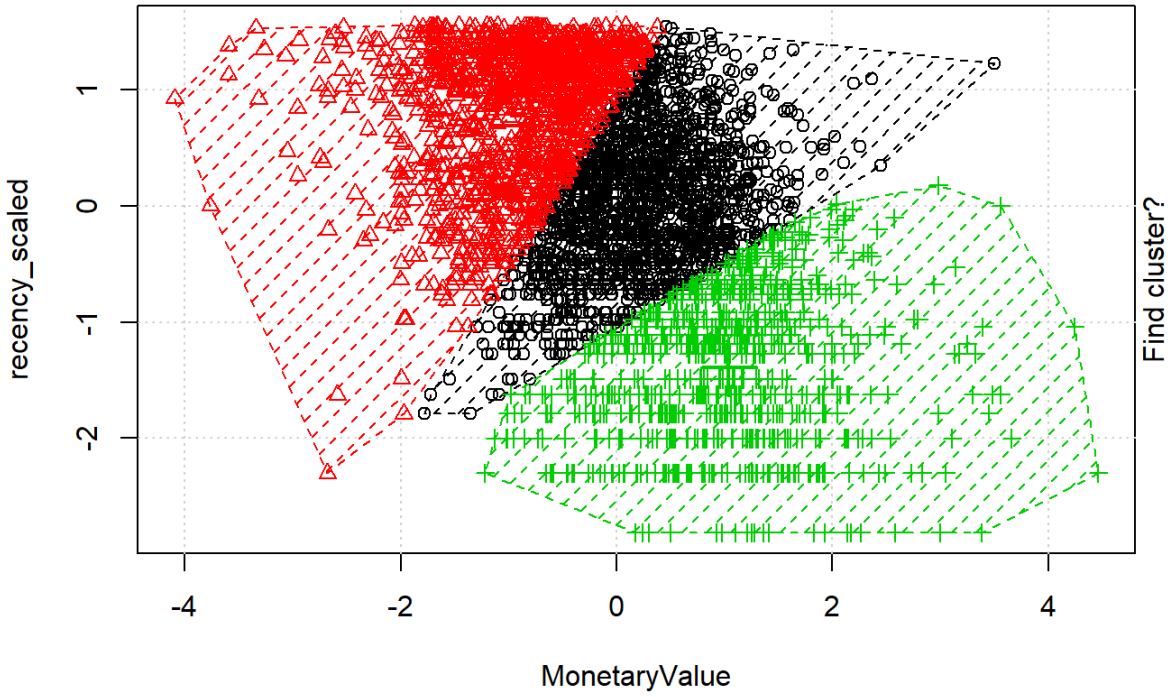

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
1
0
-1
-2

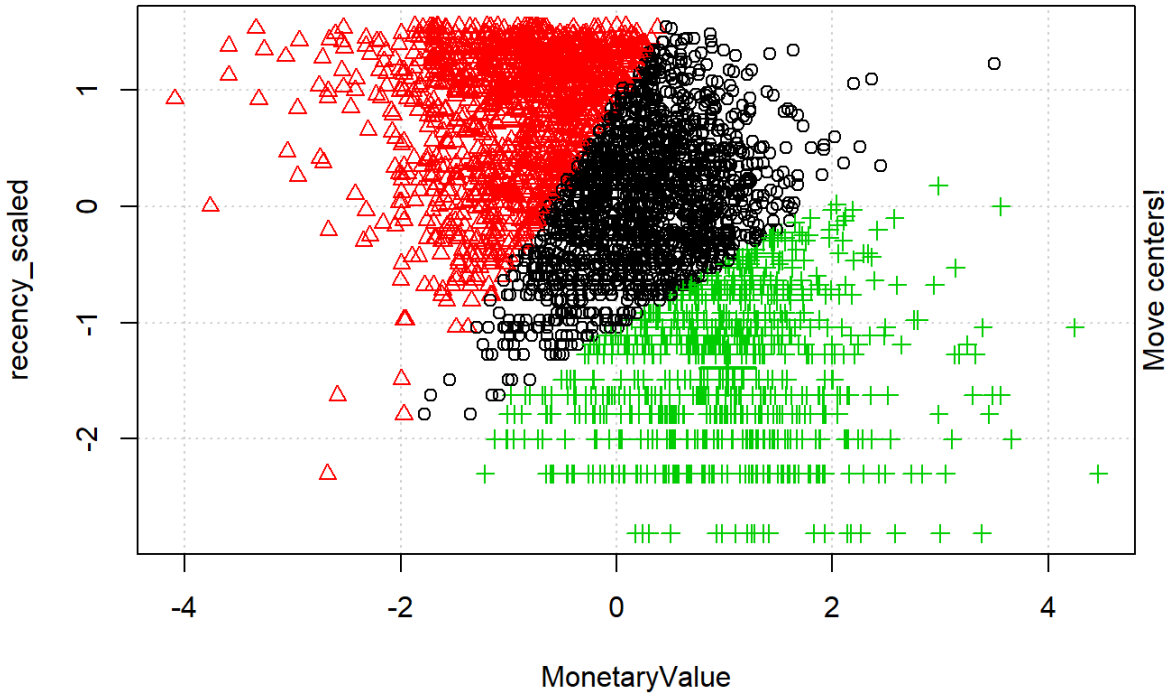

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
1
0
-1
-2

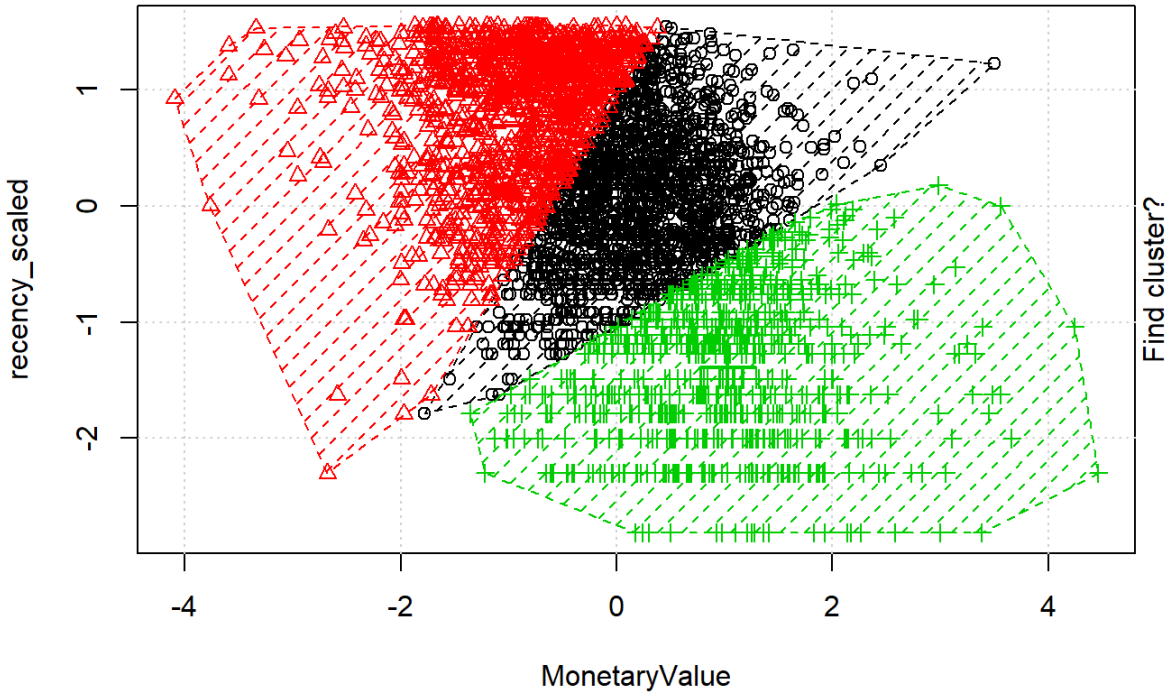

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

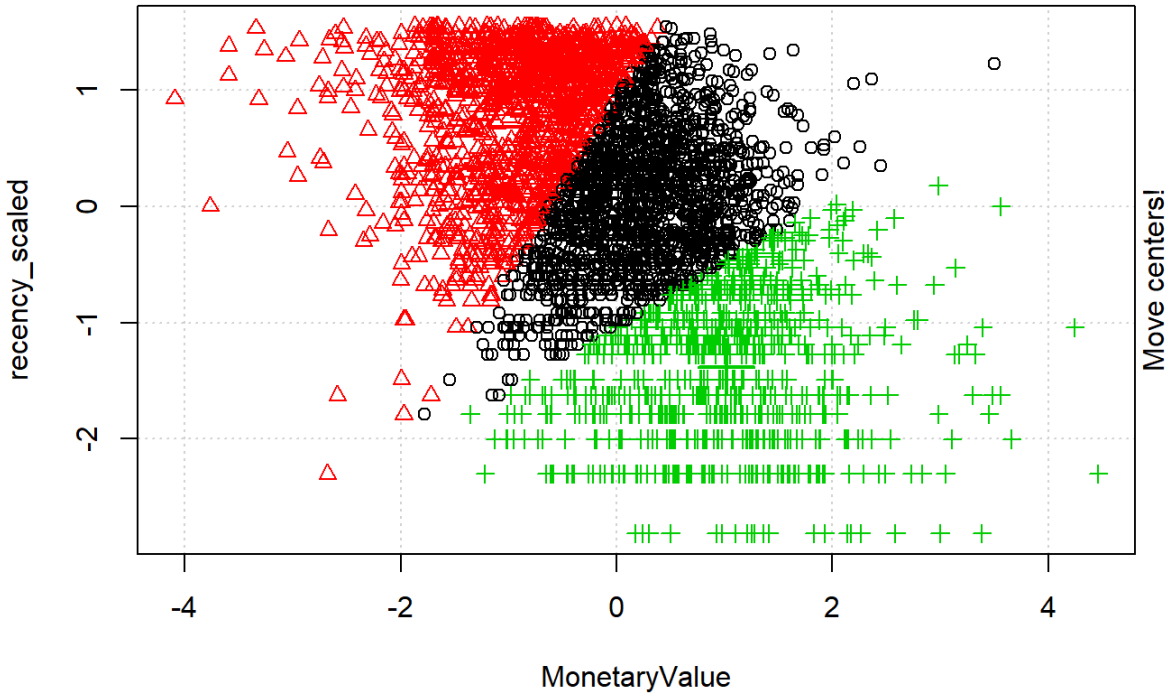

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

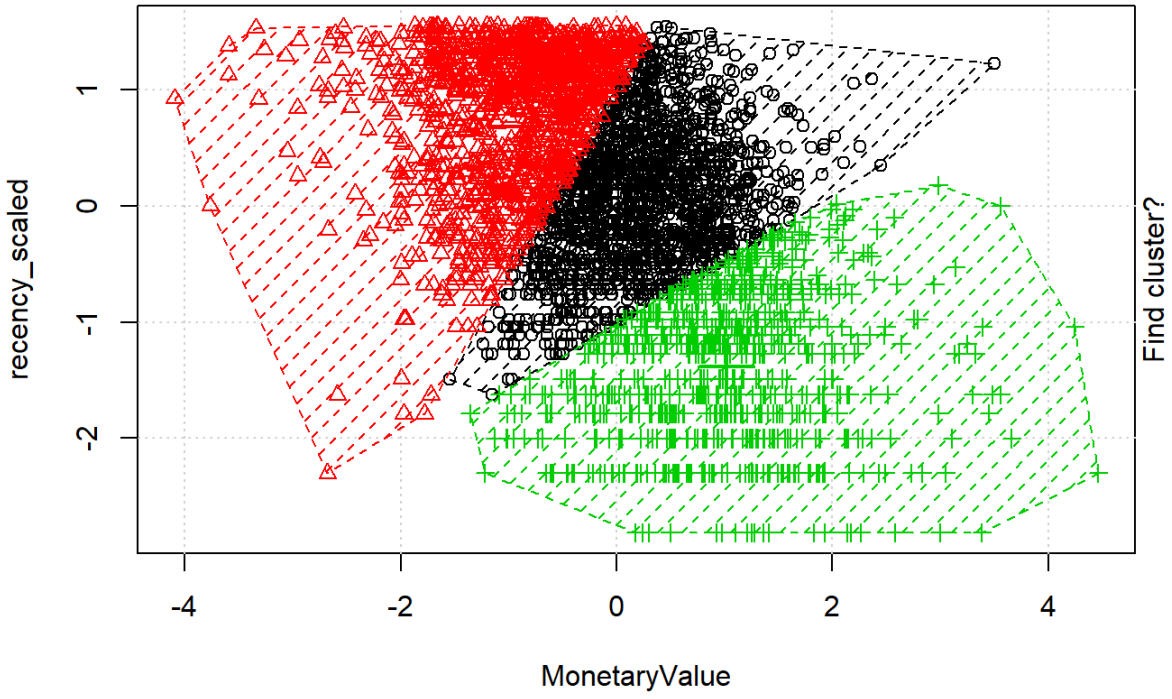

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

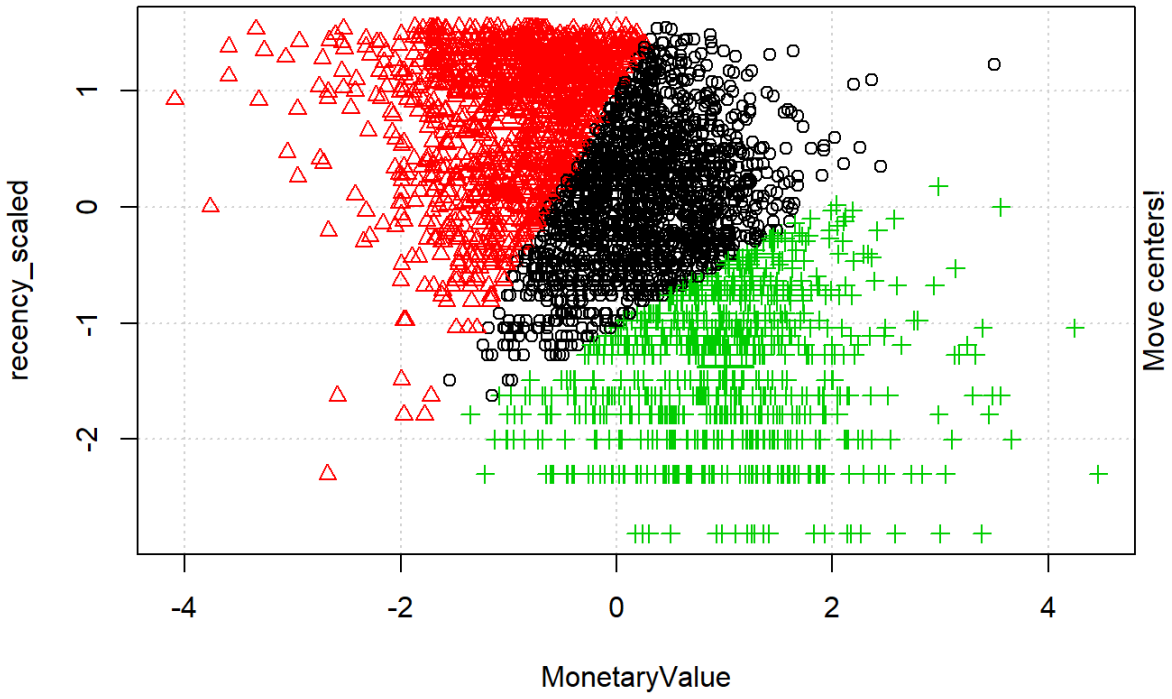

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

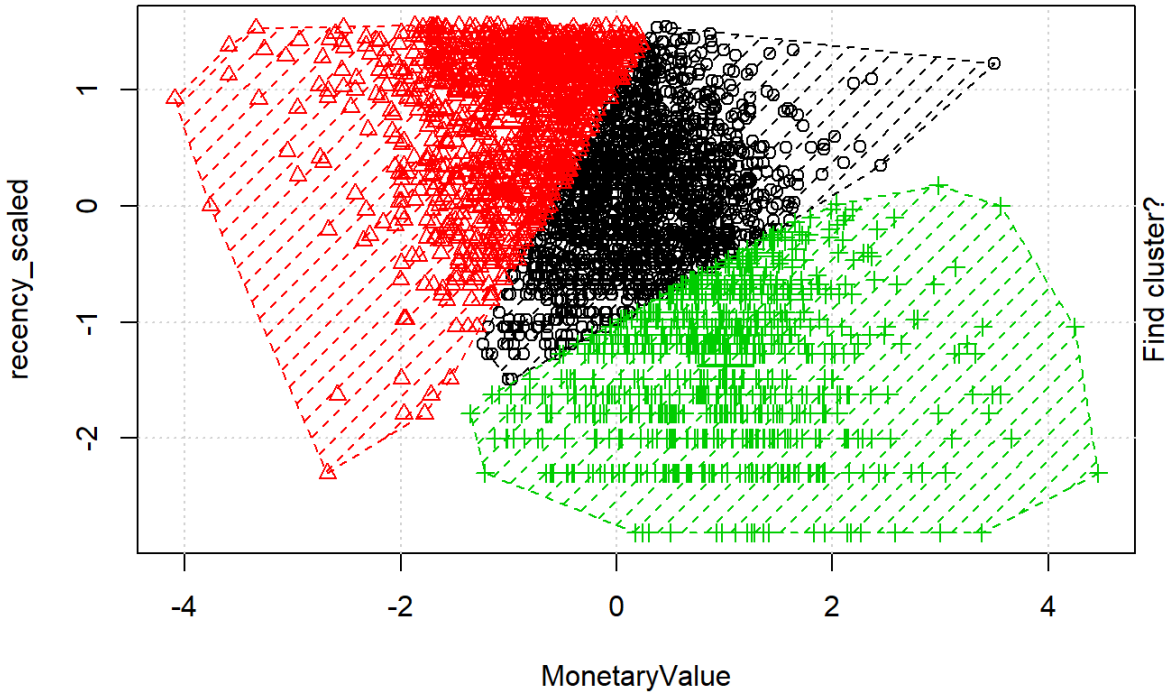

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

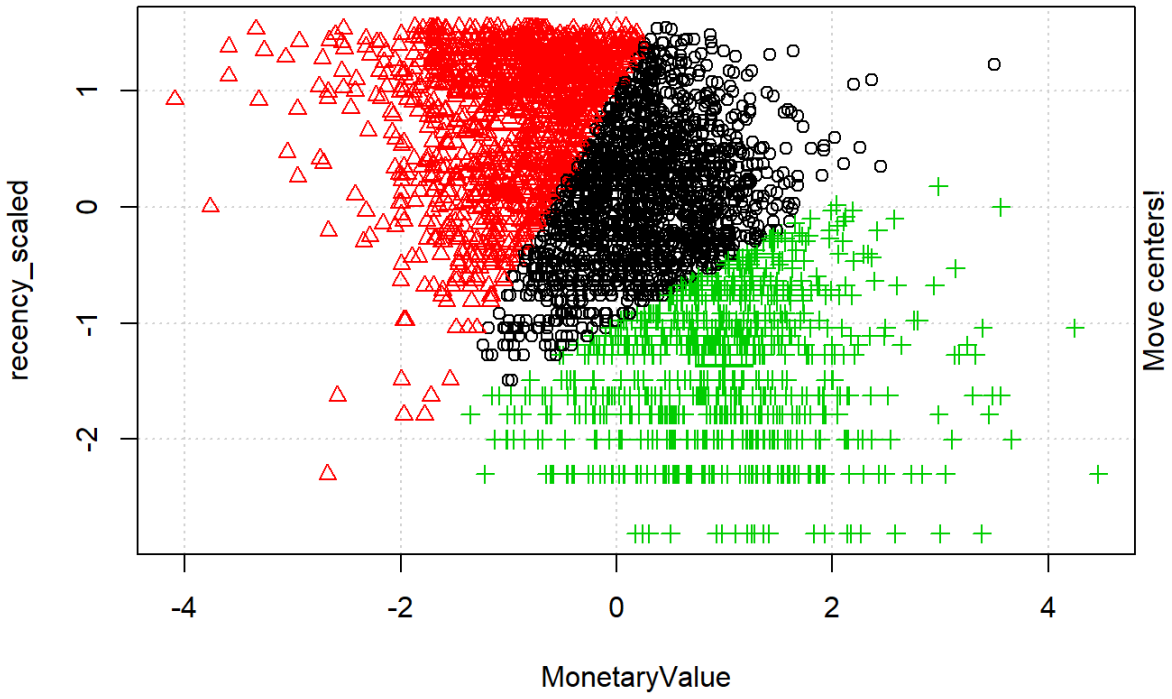

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

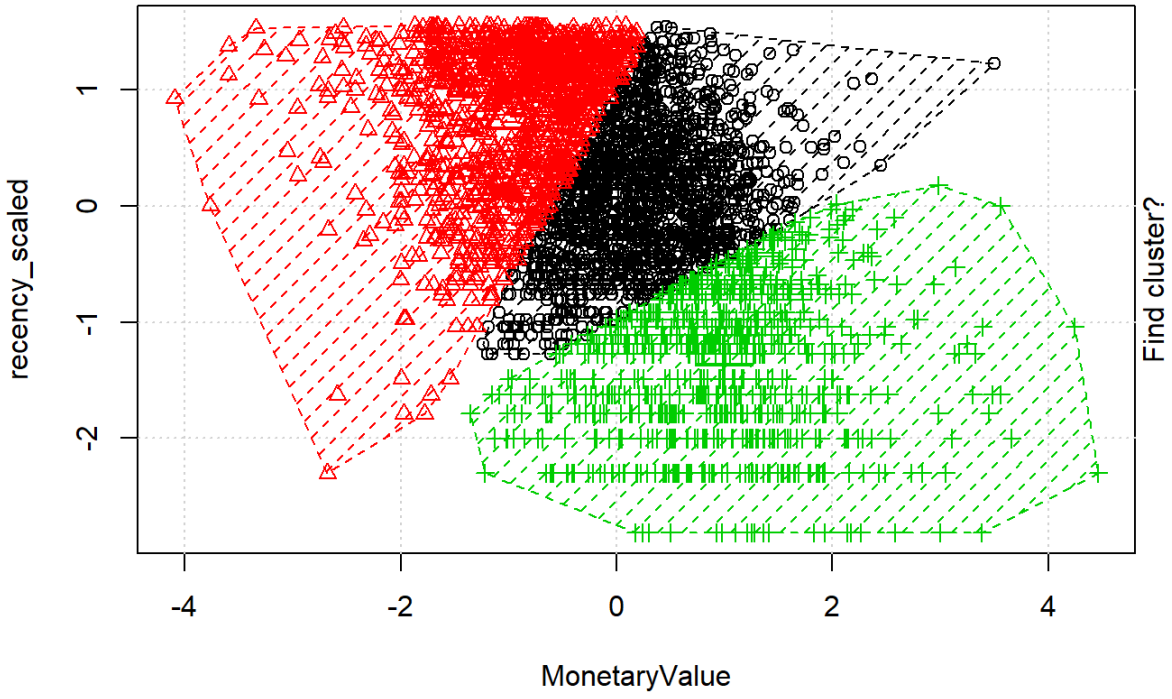

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
1
0
-1
-2

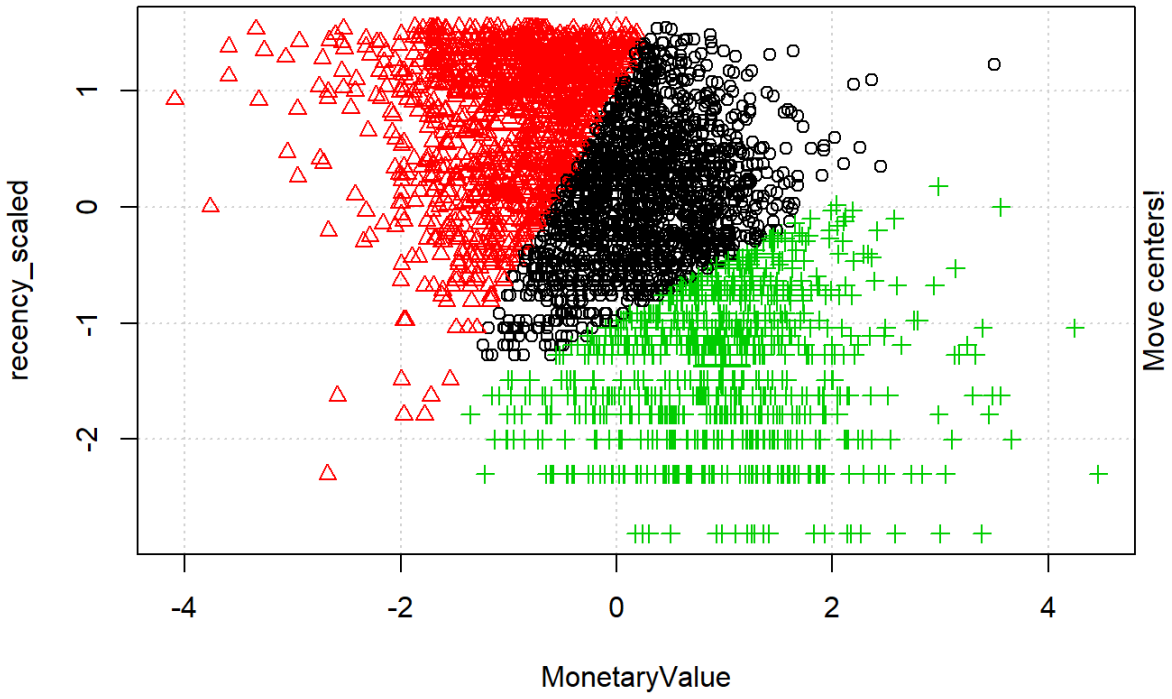

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
1
0
-1
-2

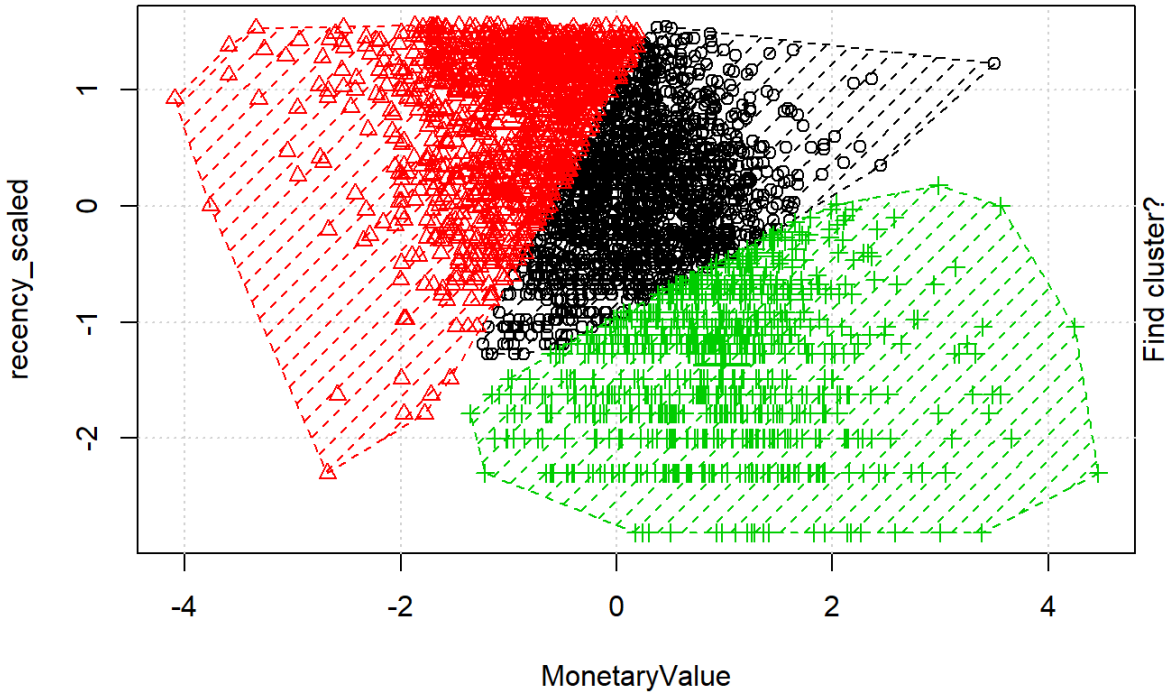

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
1
0
-1
-2

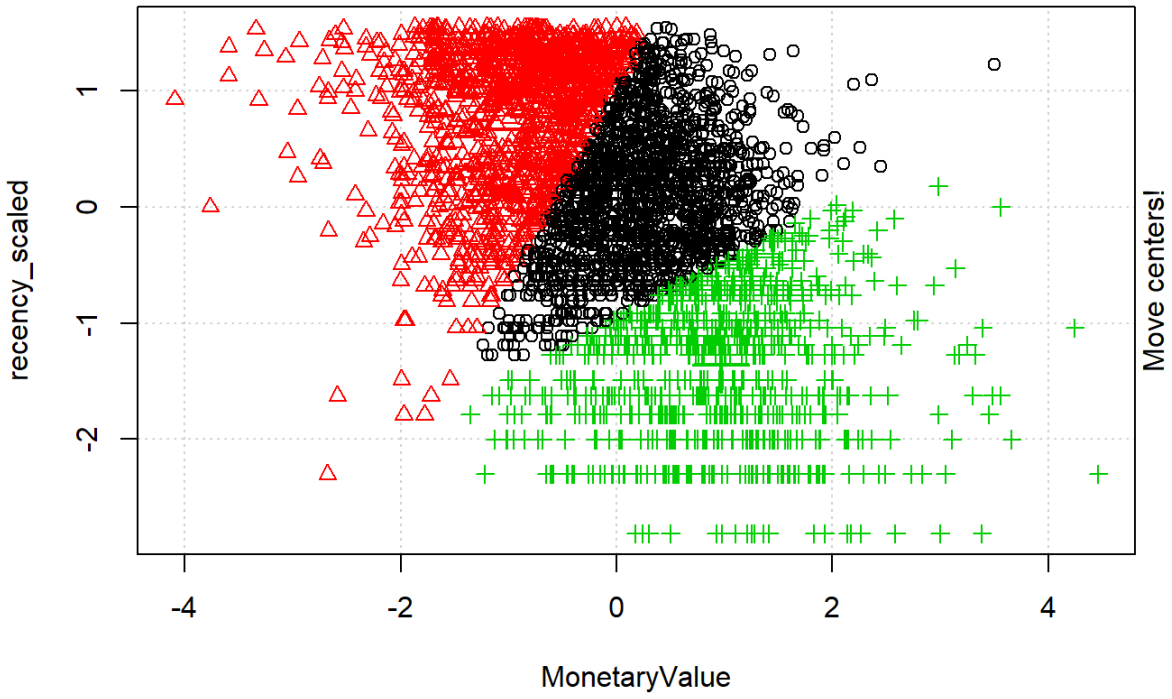

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
1
0
-1
-2

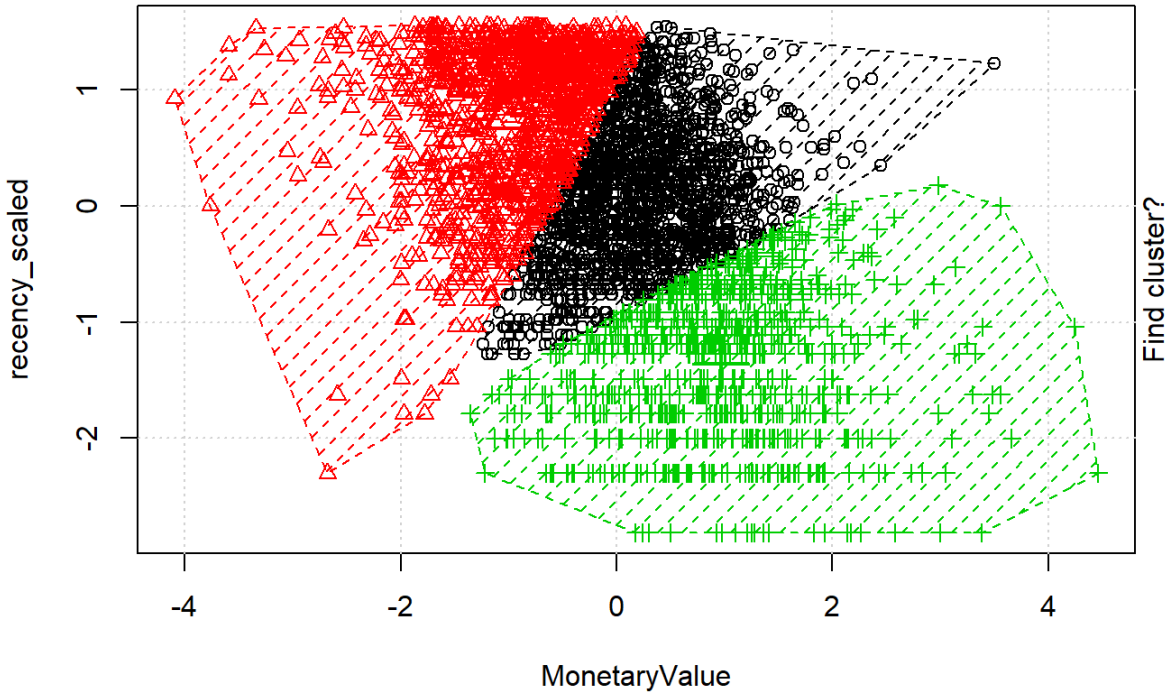

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

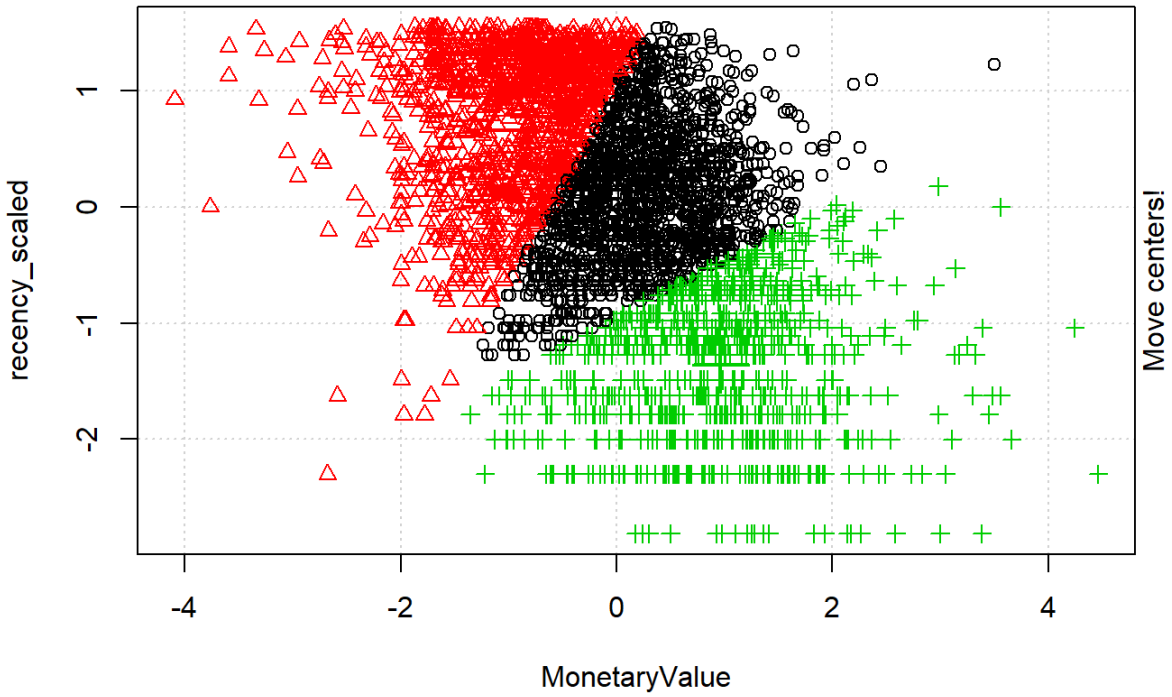

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

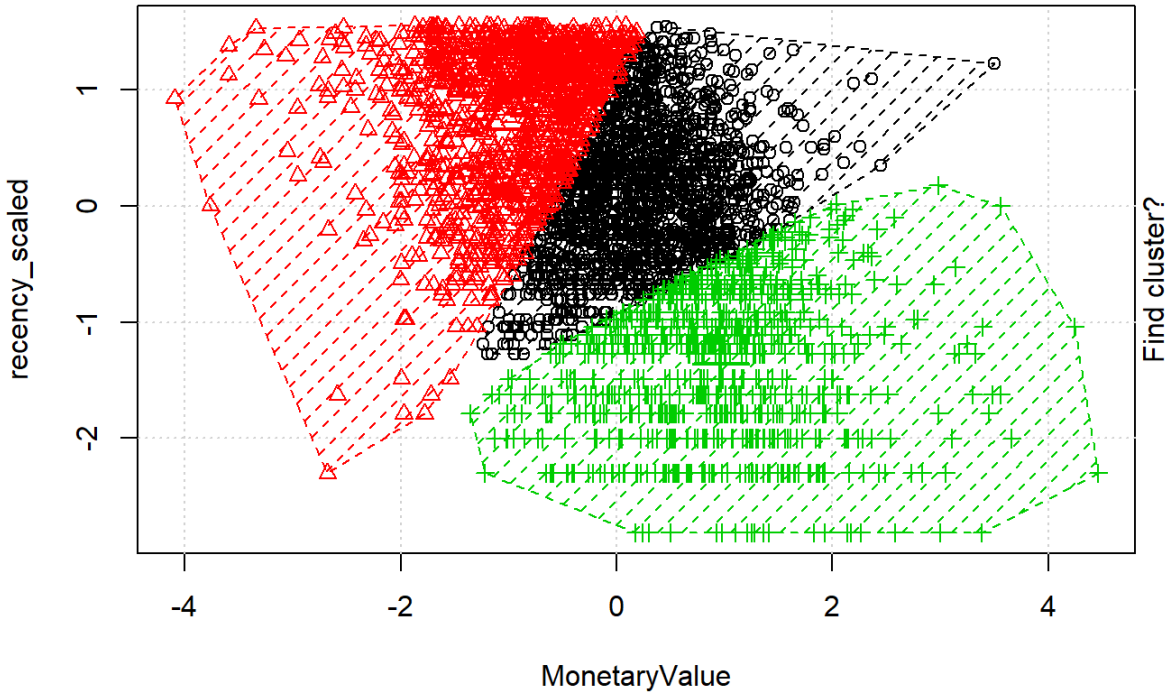

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
1
0
-1
-2

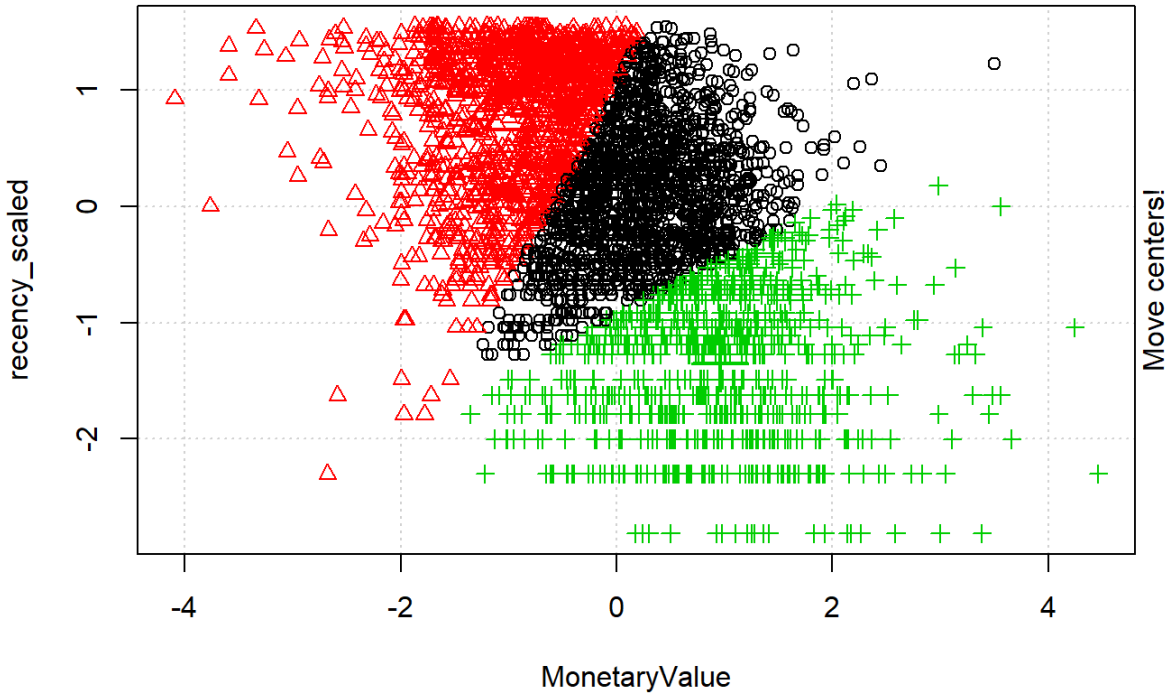

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
1
0
-1
-2

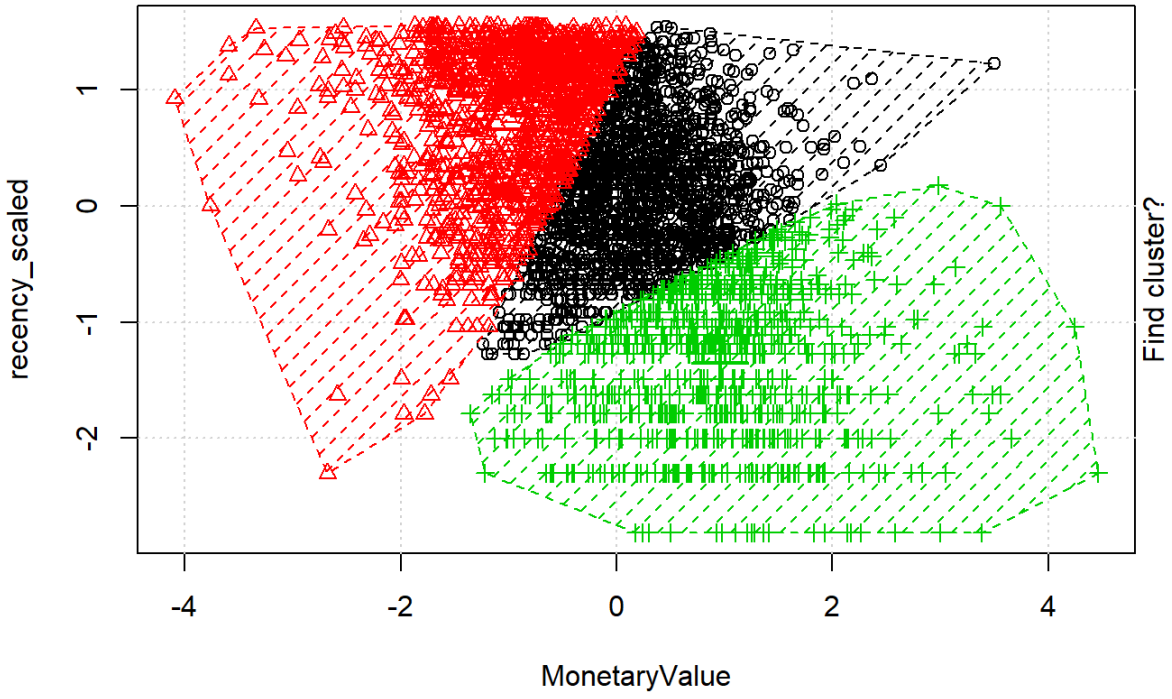

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

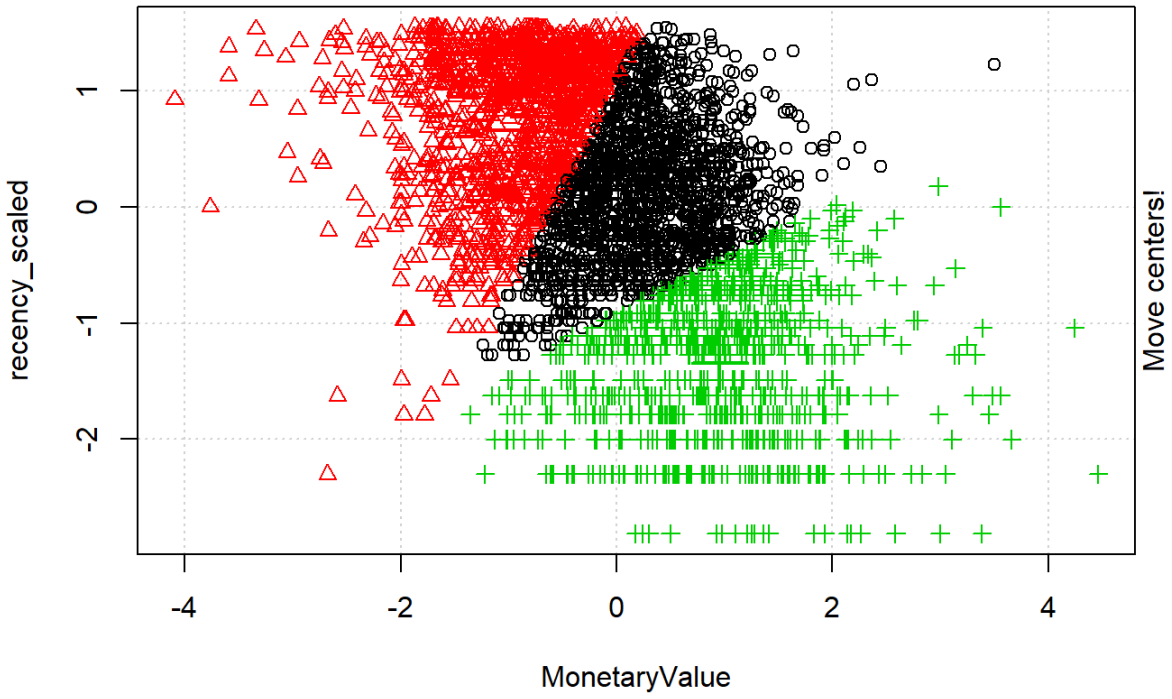

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

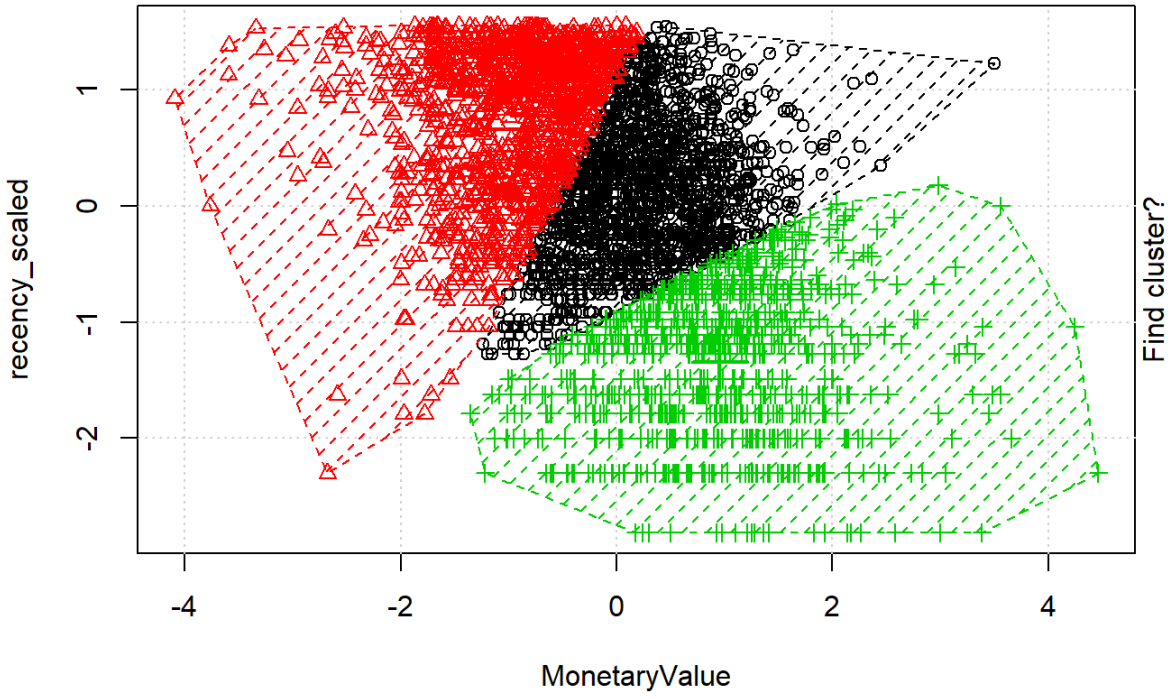

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

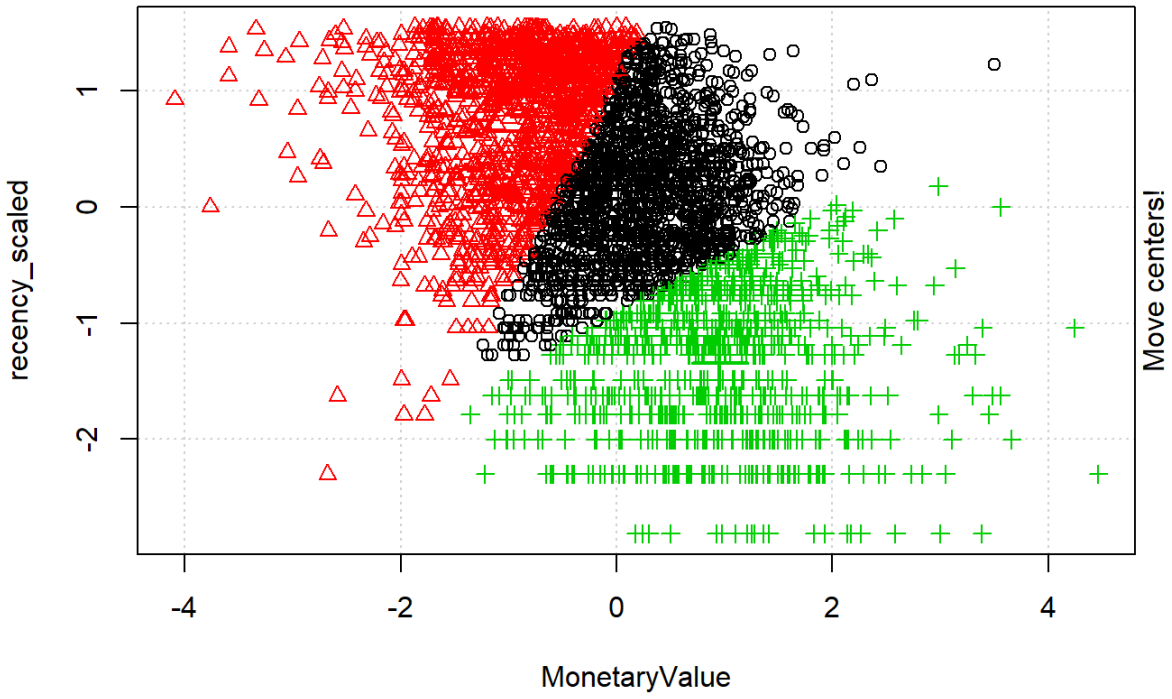

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

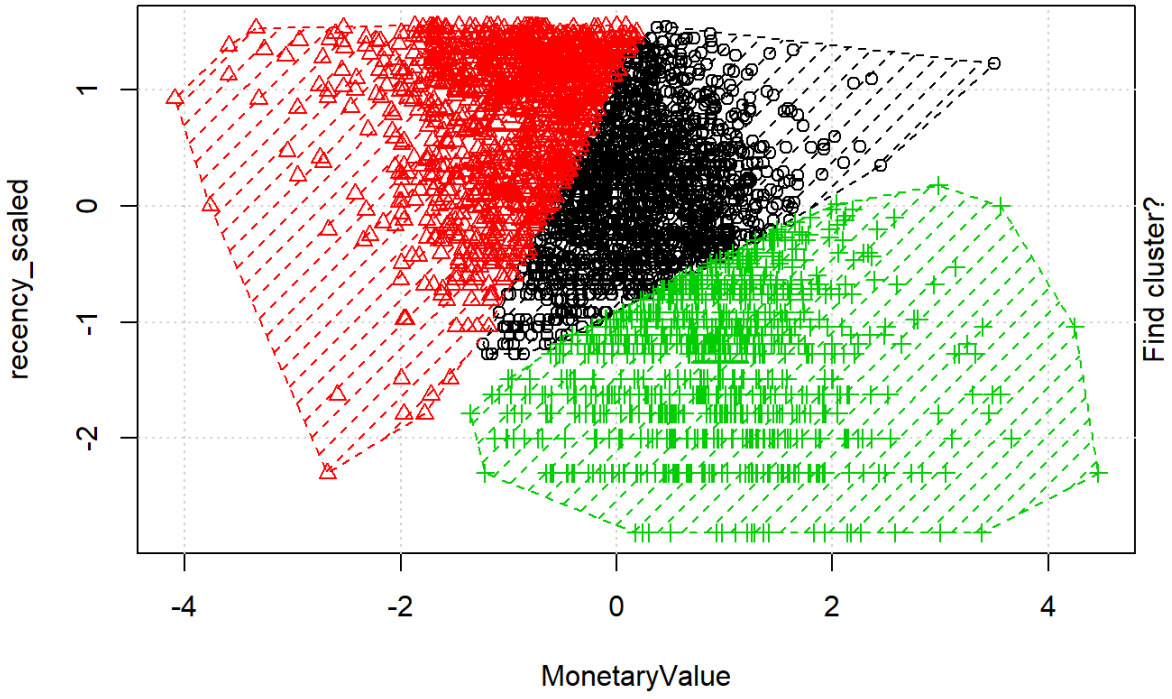

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

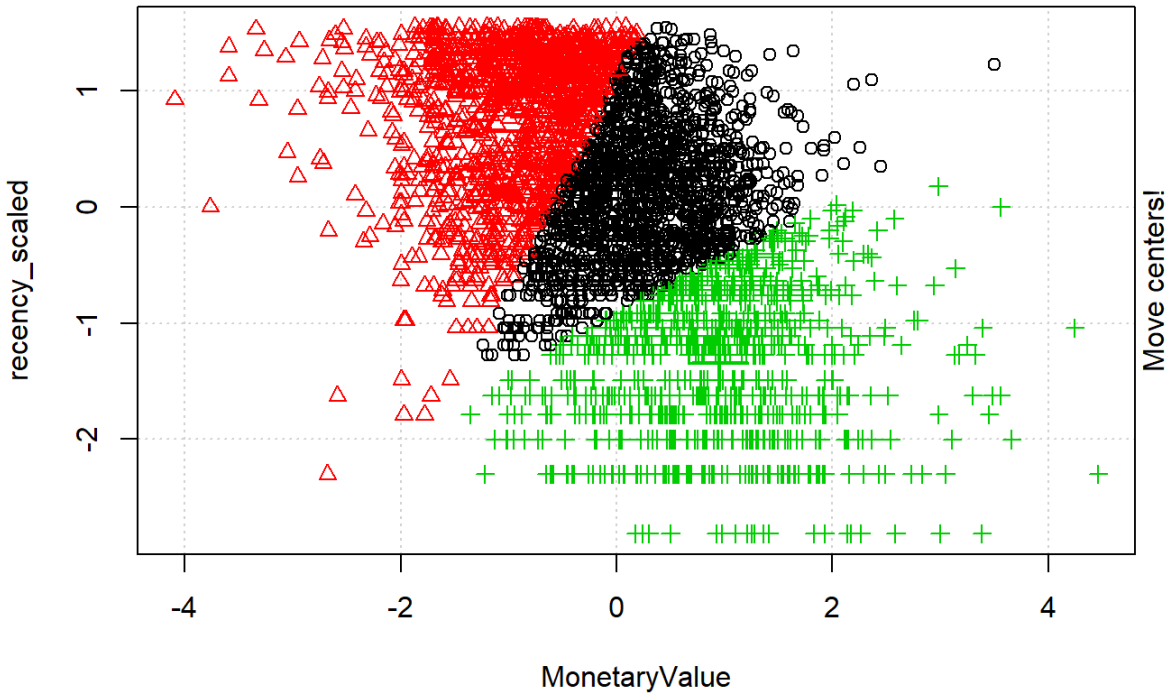

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

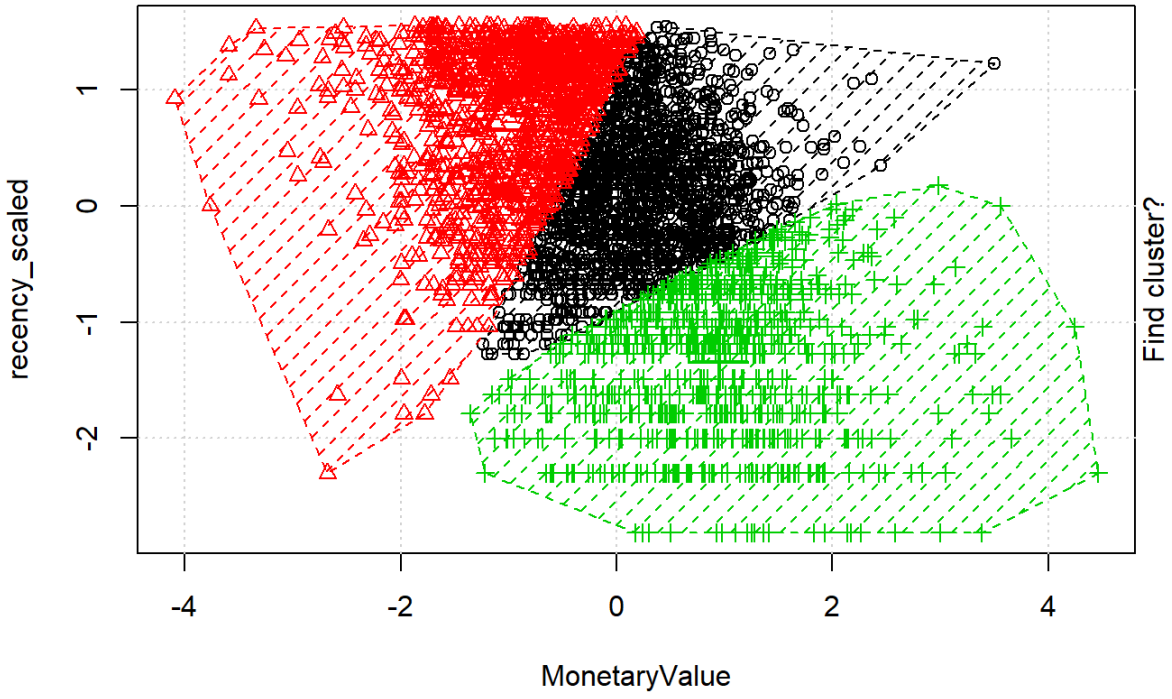

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

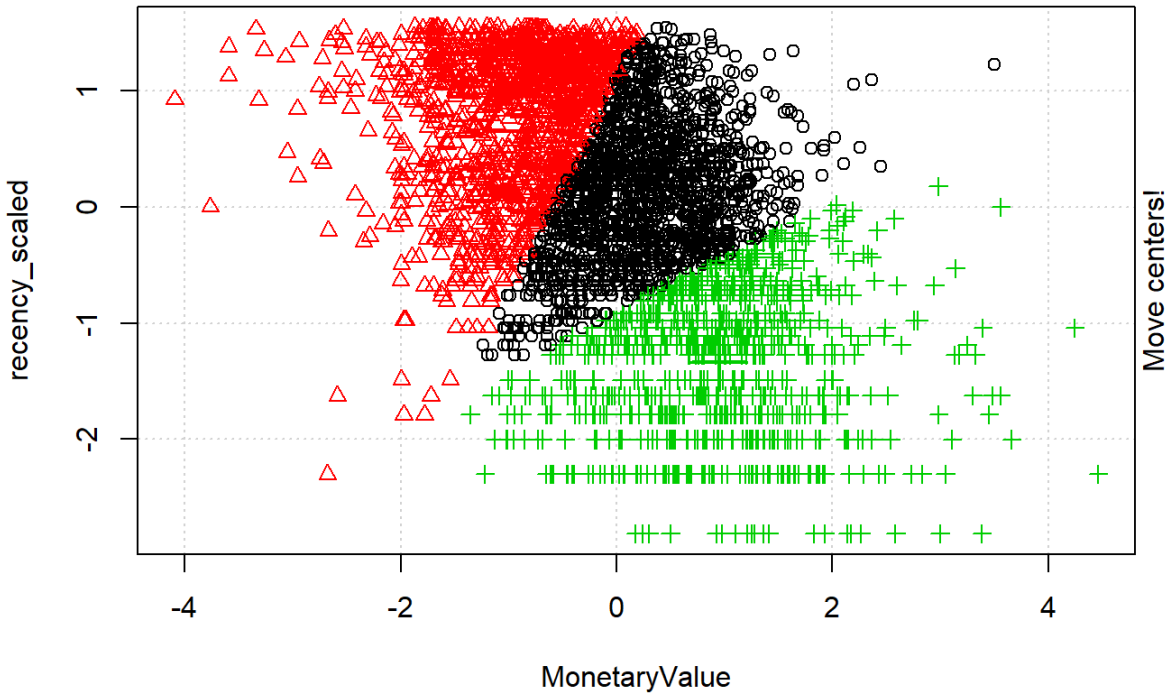

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

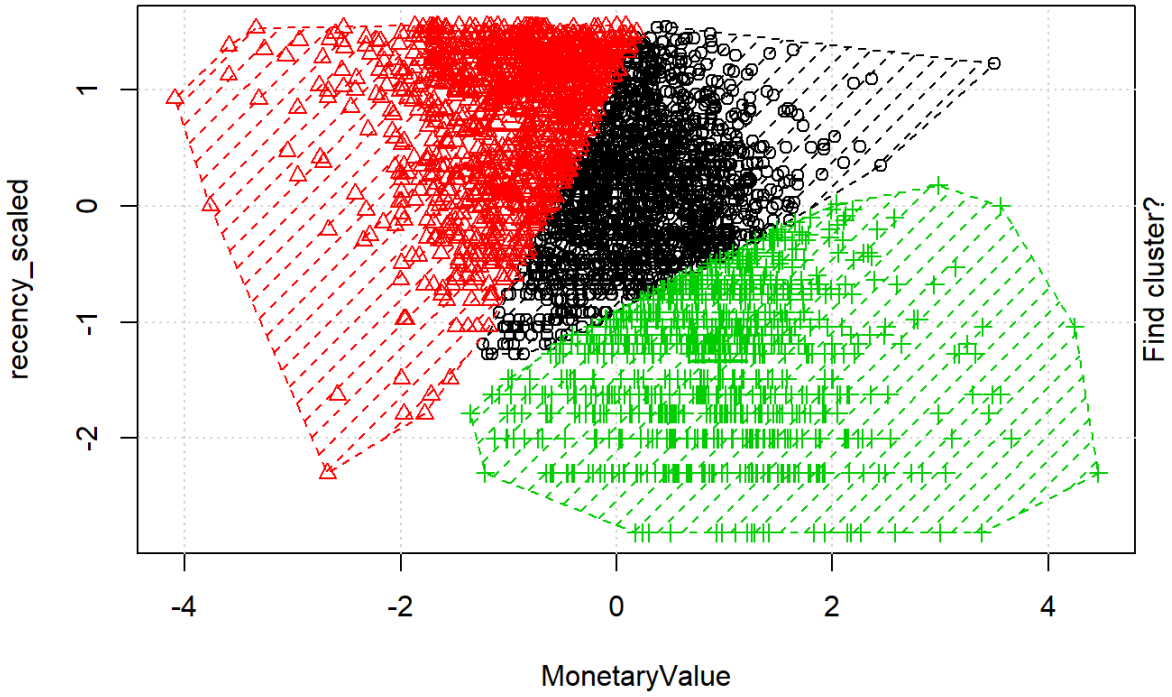

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

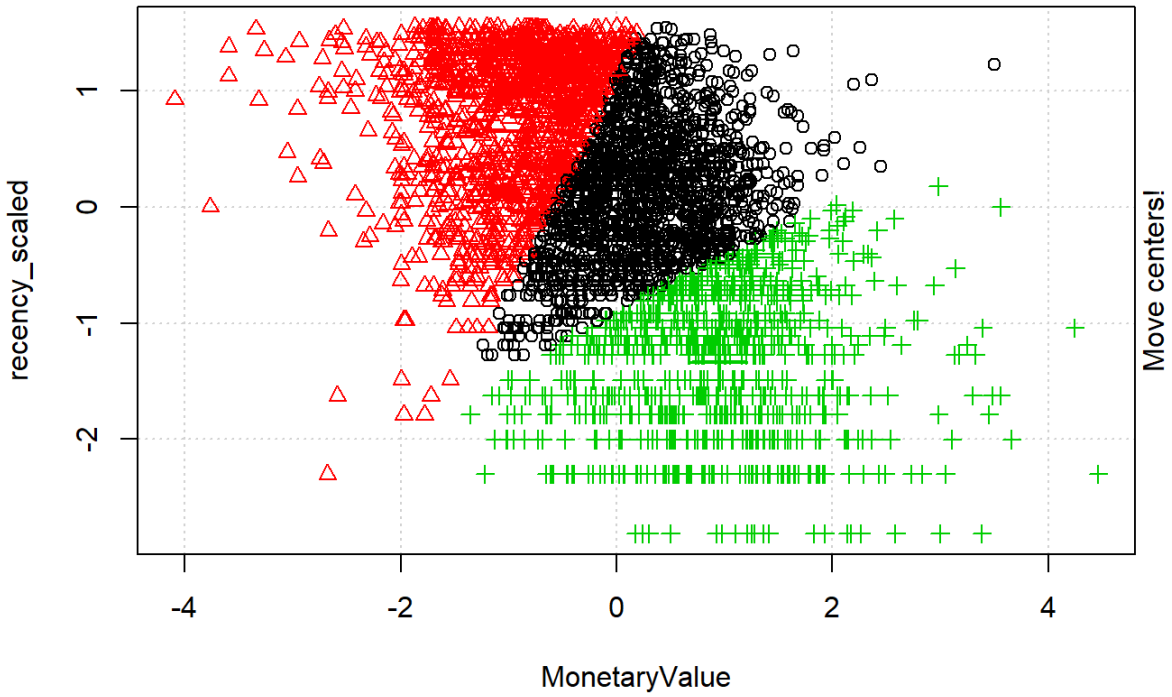

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

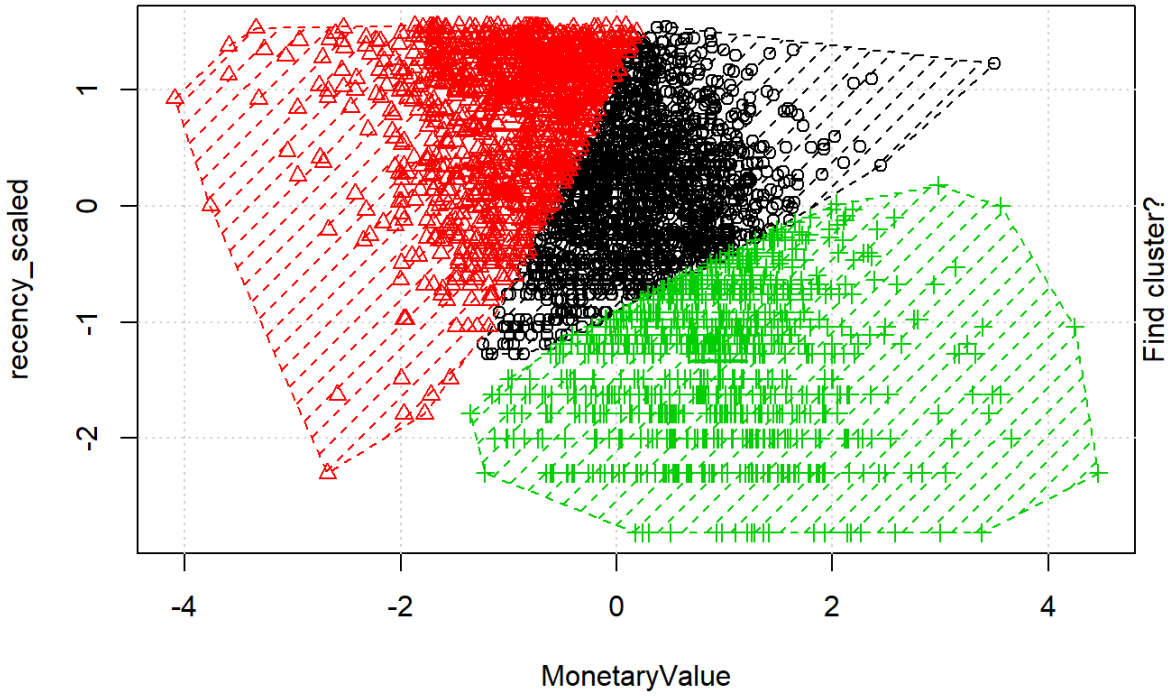

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

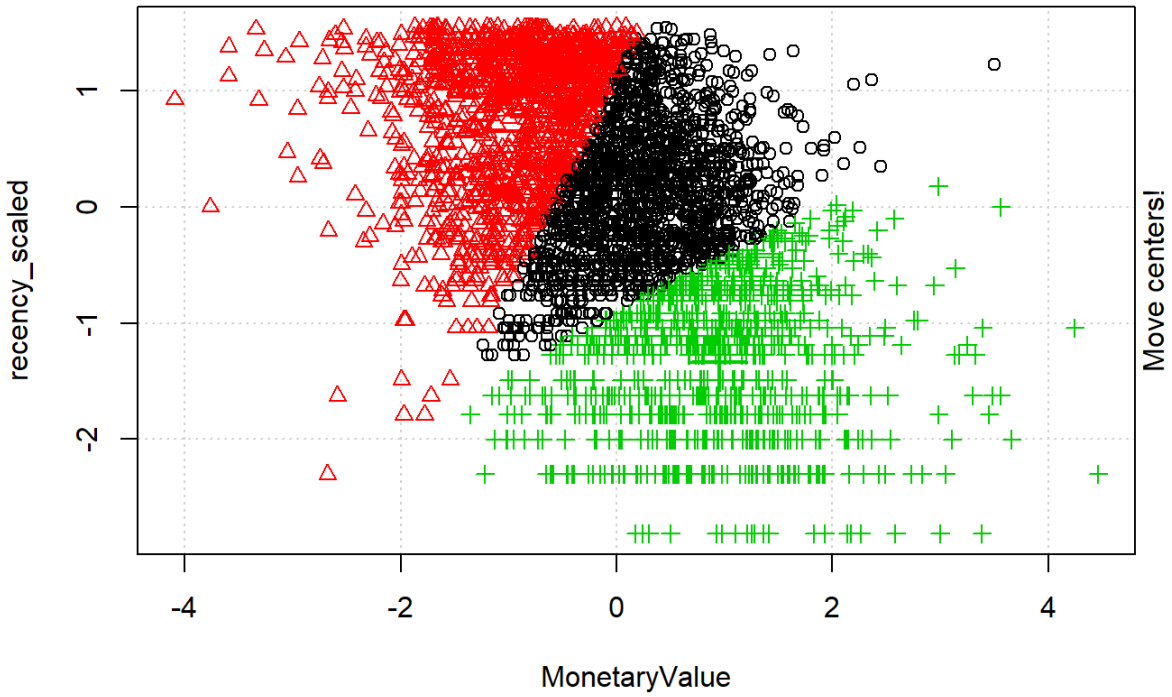

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

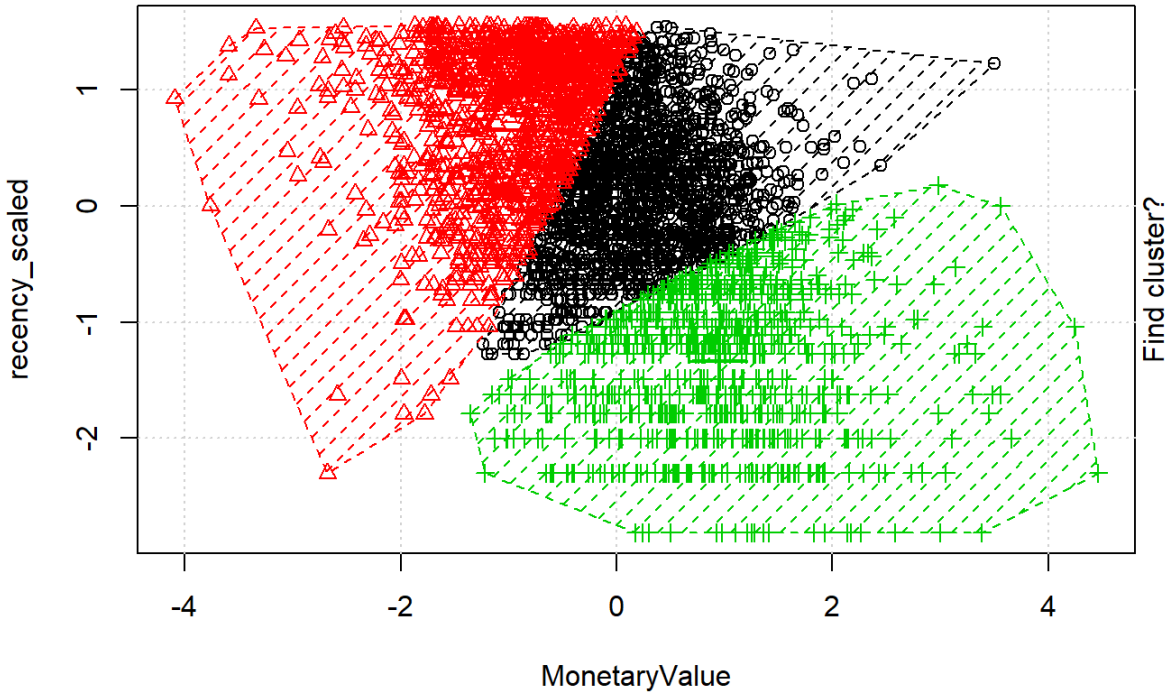

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1


recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

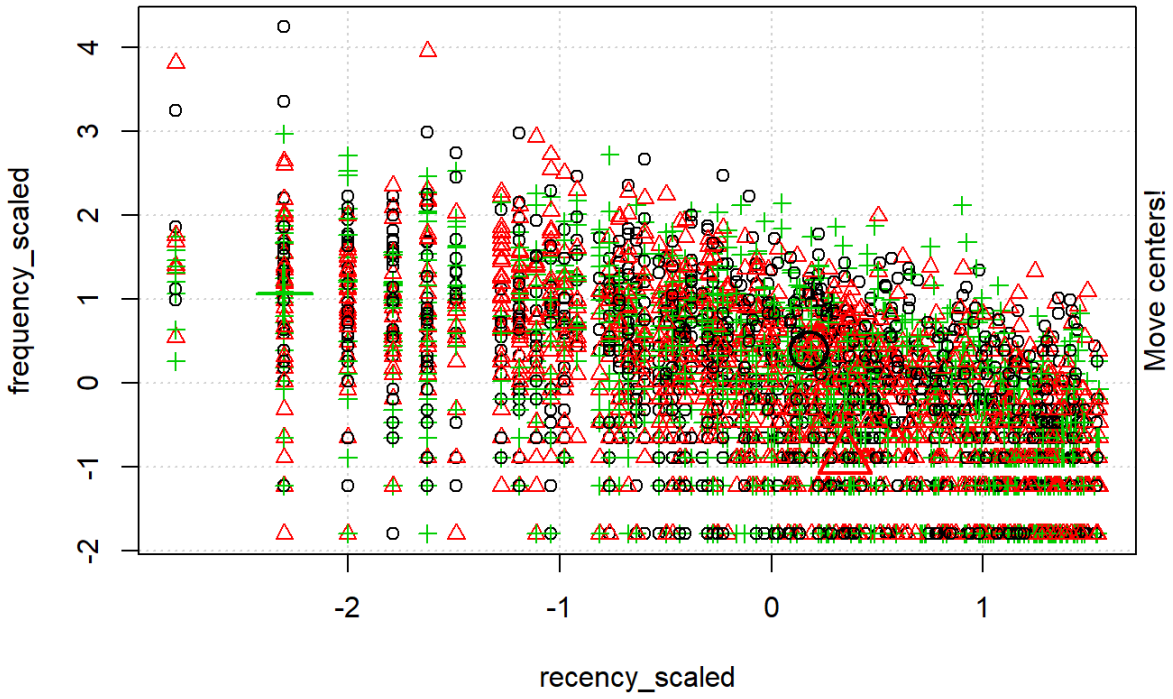

recency_scaled
MonetaryValue
Move centers!
-4
-2
0
2
4
-2
-1
0
1

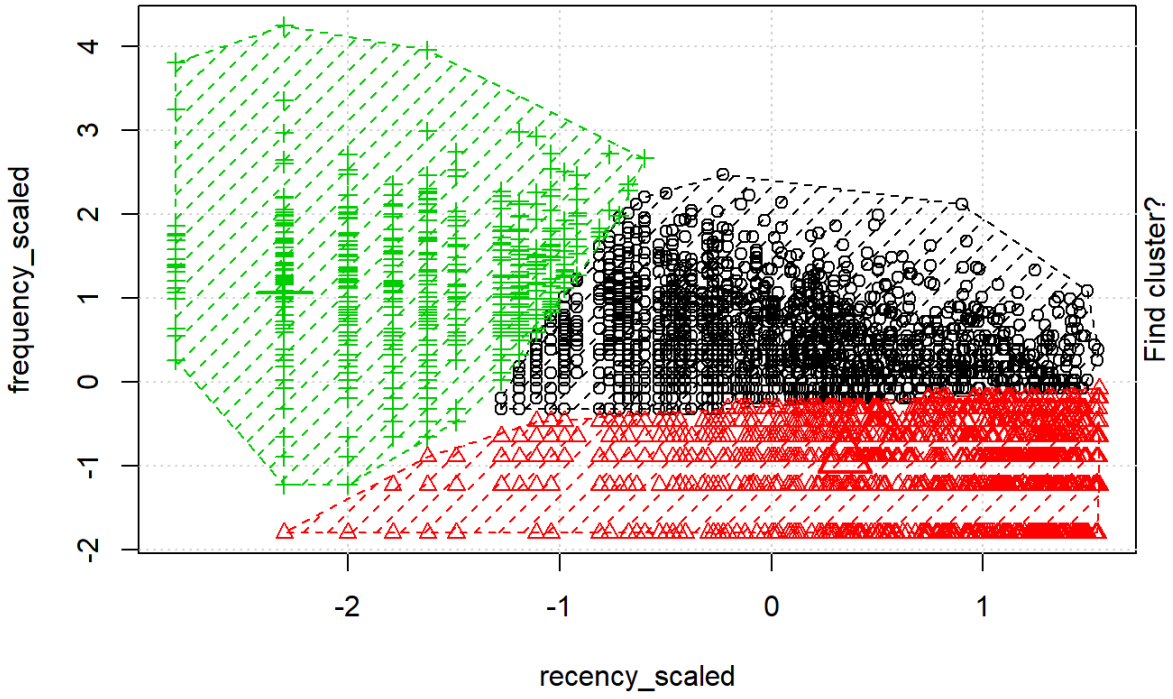

recency_scaled
MonetaryValue
Find cluster?
-4
-2
0
2
4
-2
-1
0
1

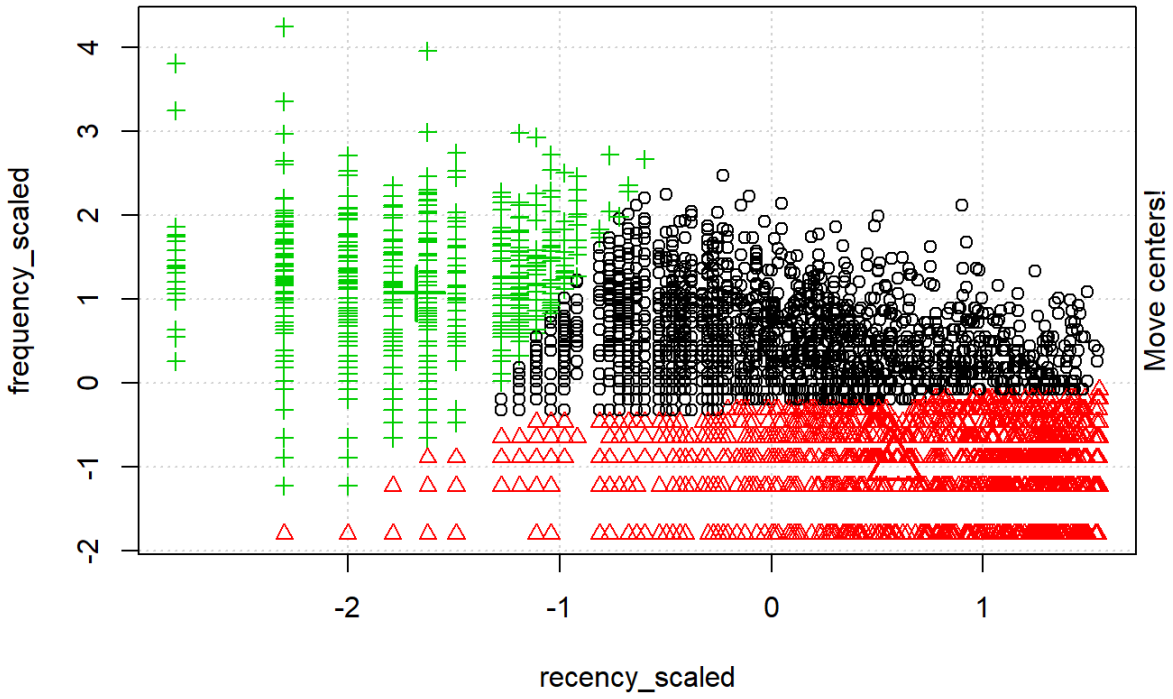


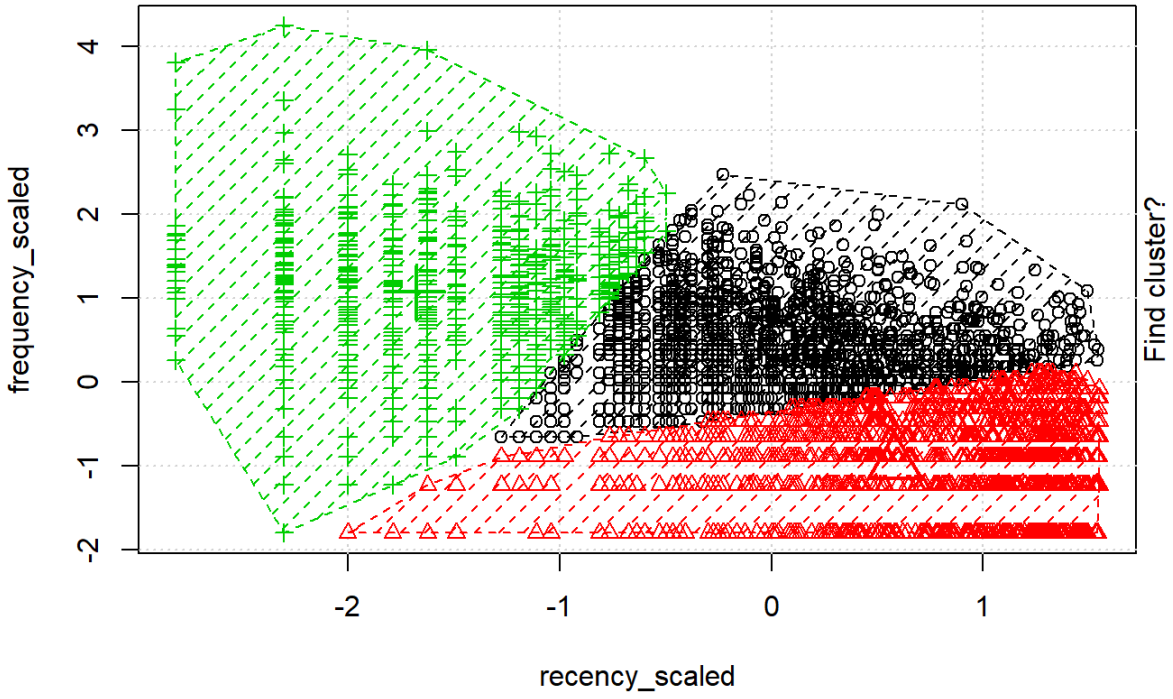


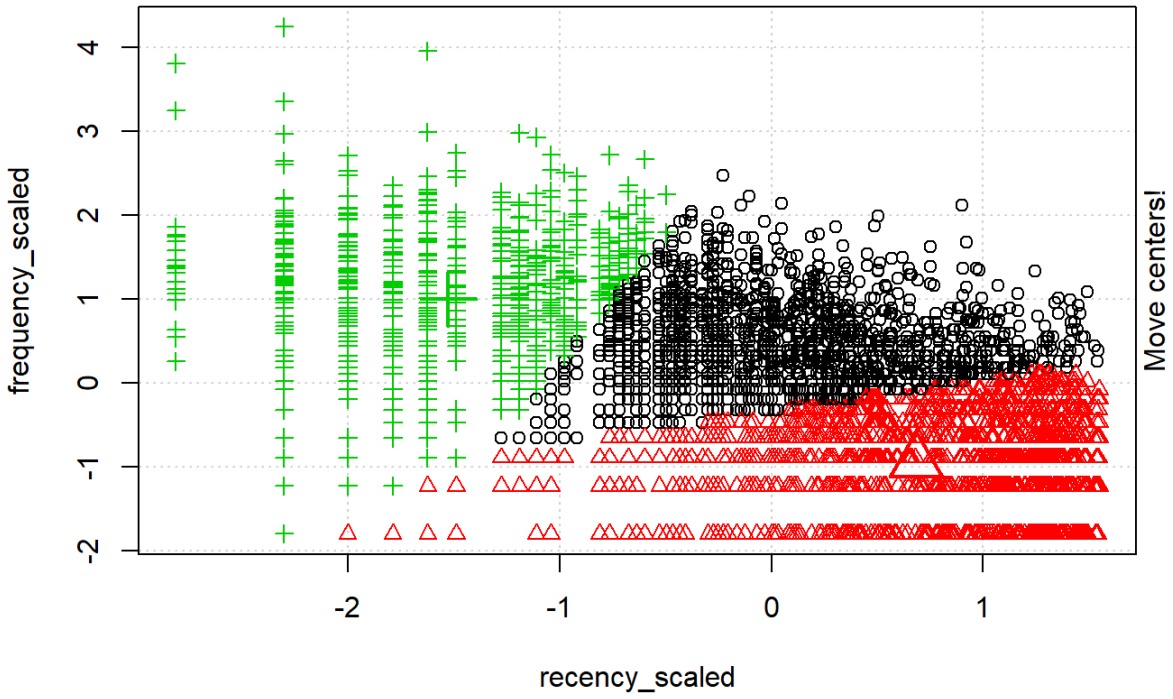

```
kmeans.ani(rescale_subscribermodel[4:5], 3)
```

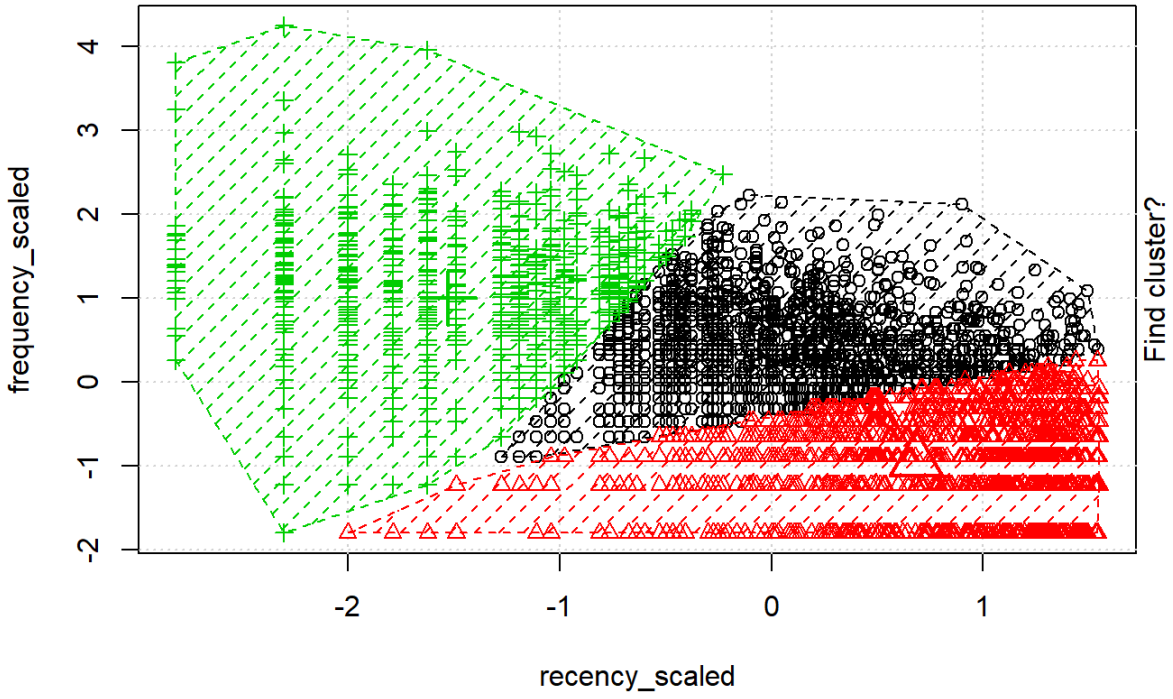

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

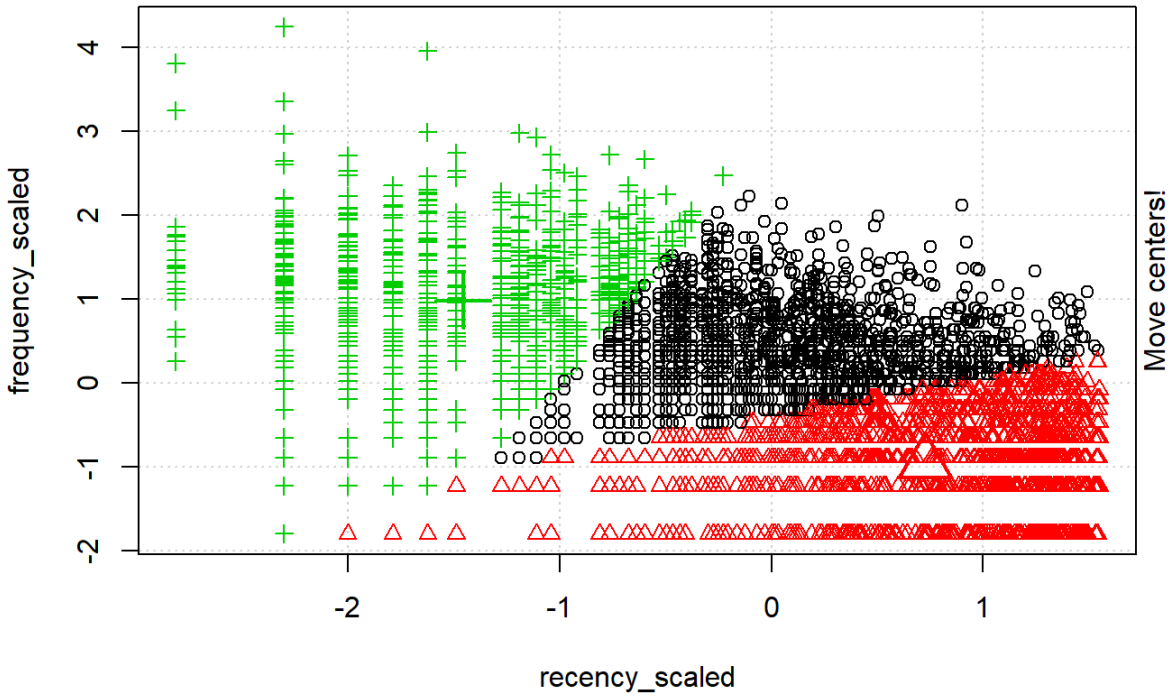

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

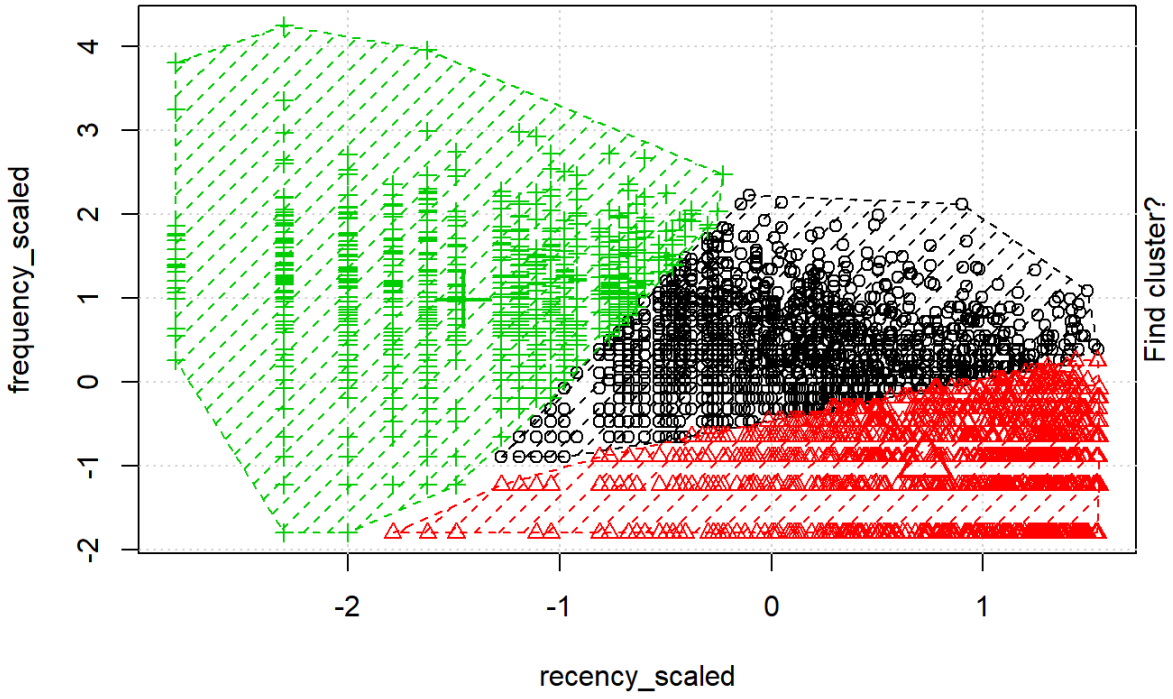

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

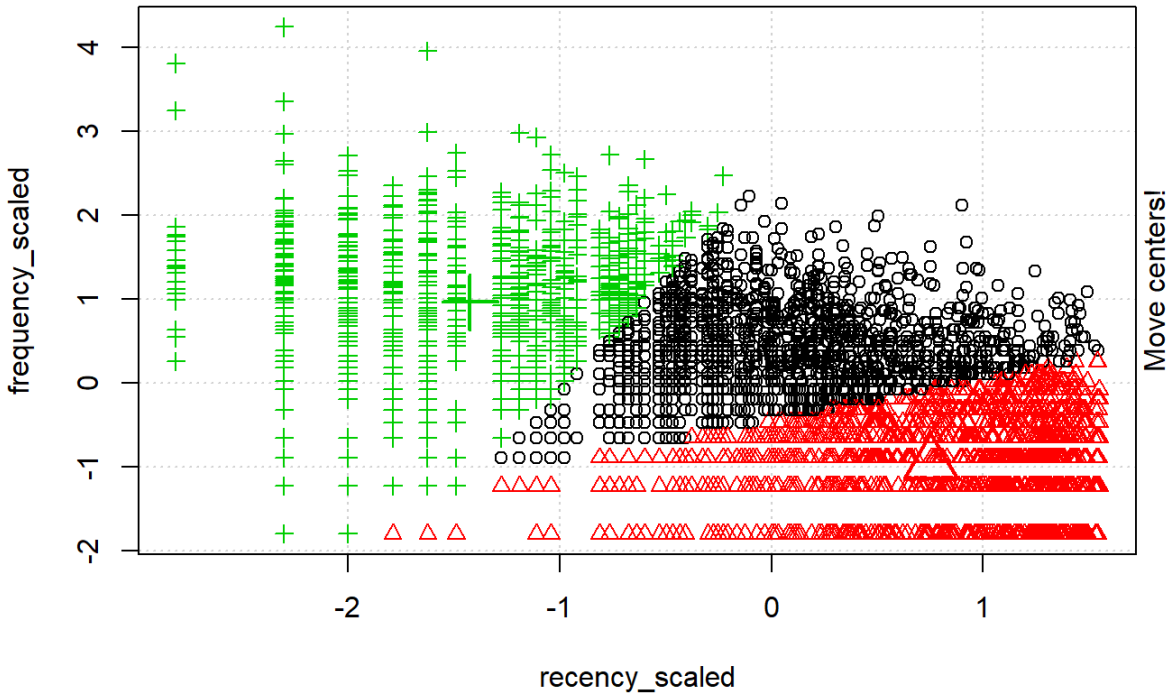

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

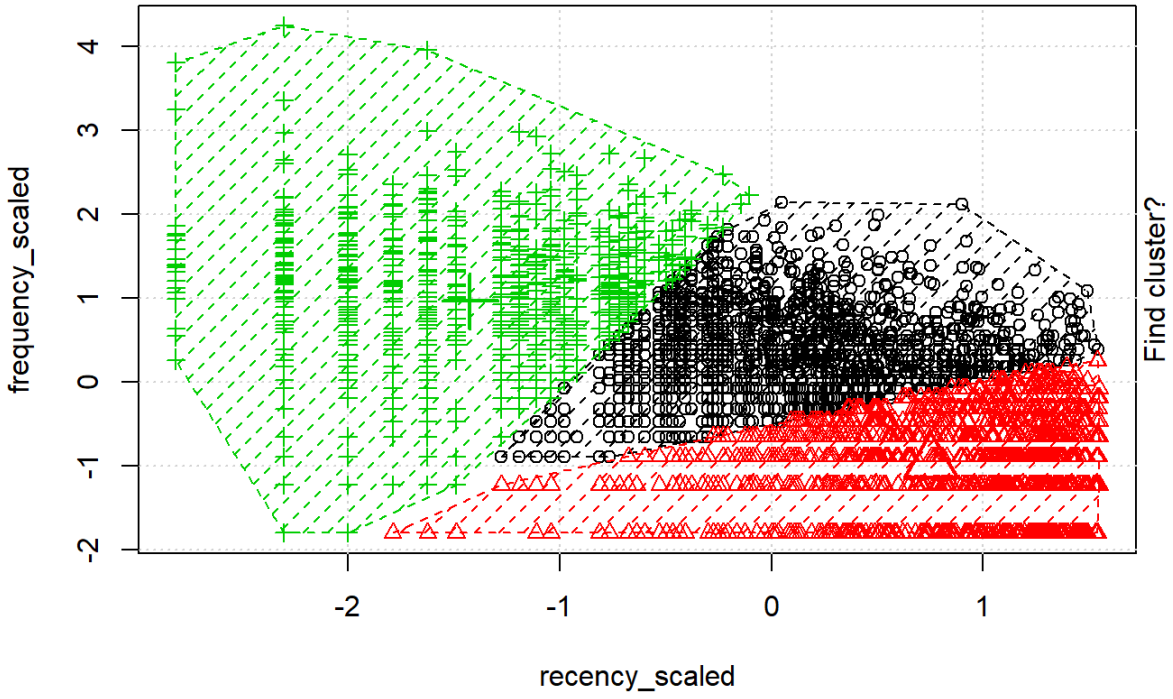

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

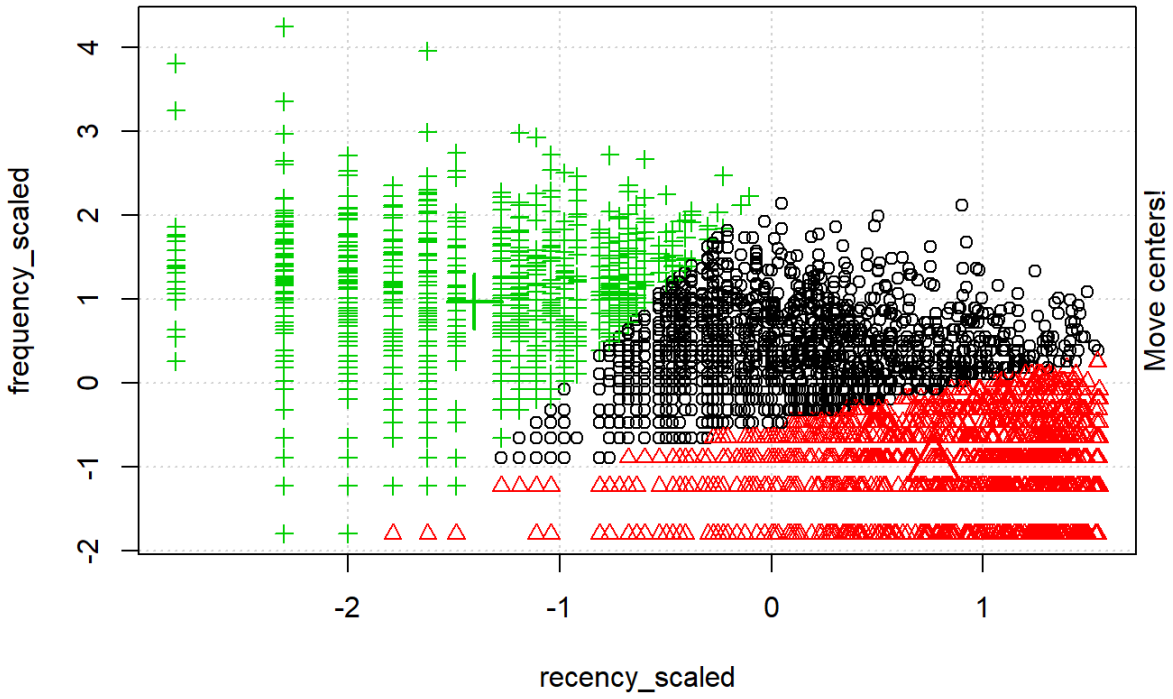

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

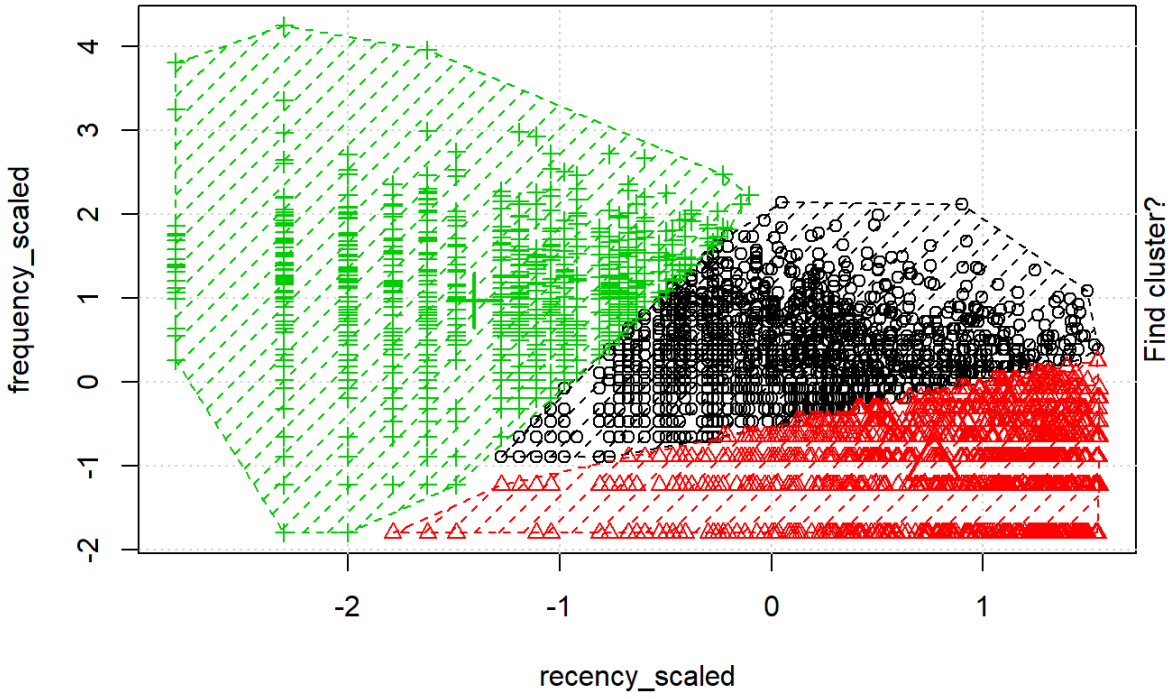

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

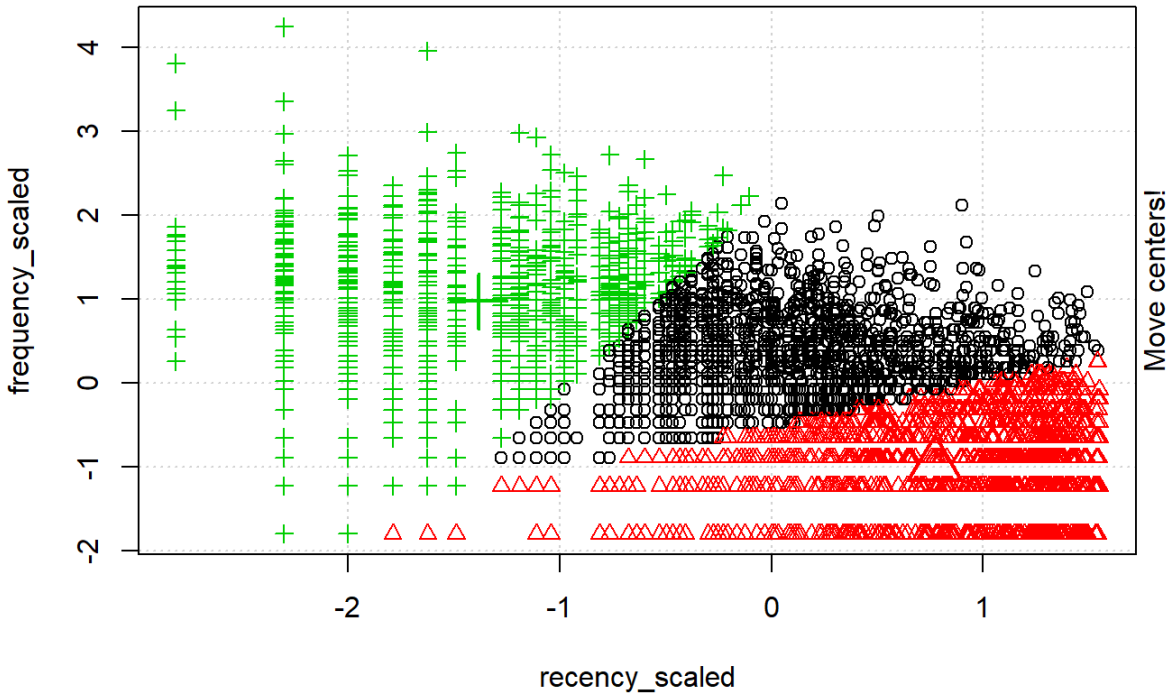

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

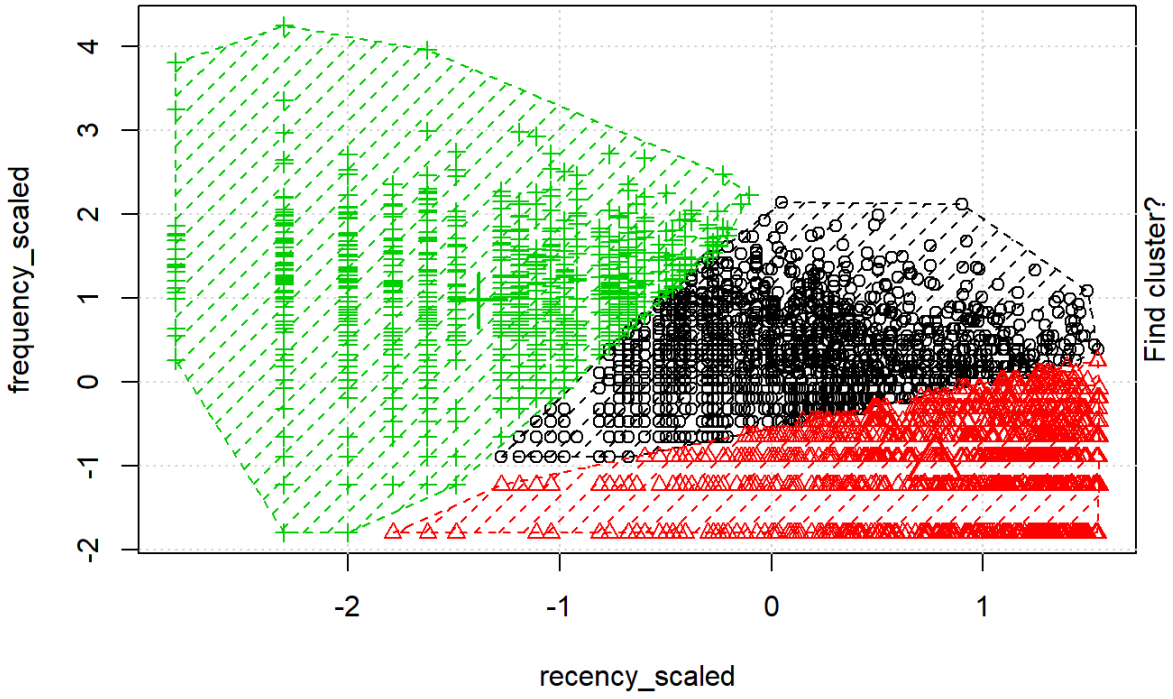

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

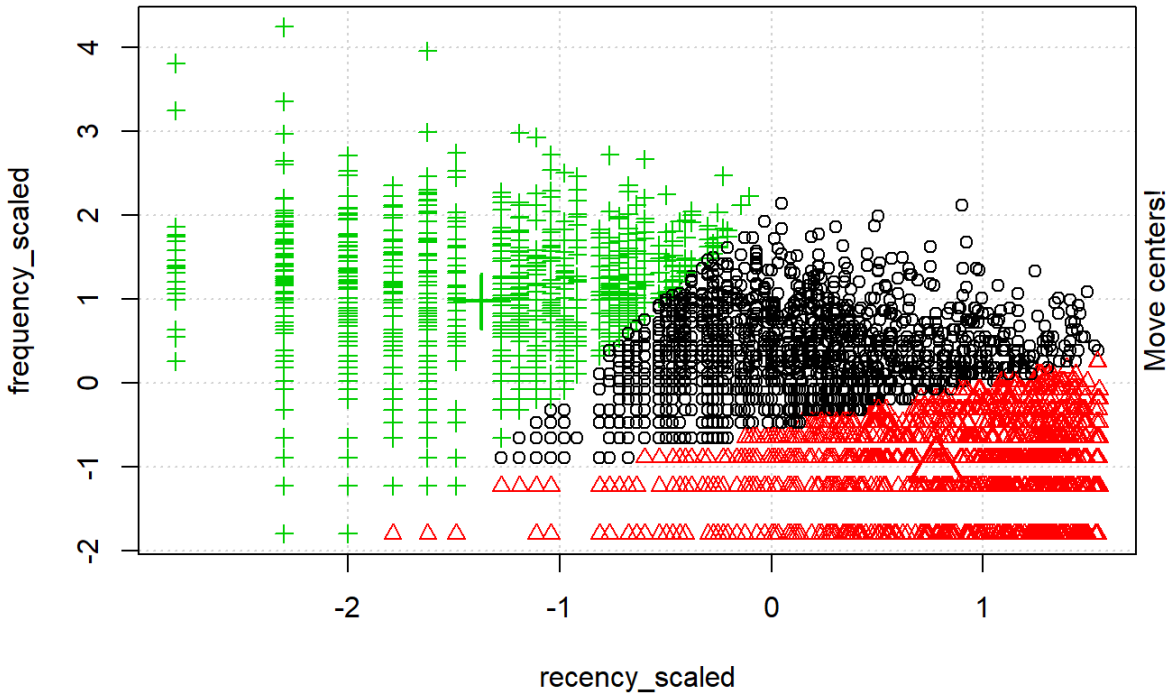

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

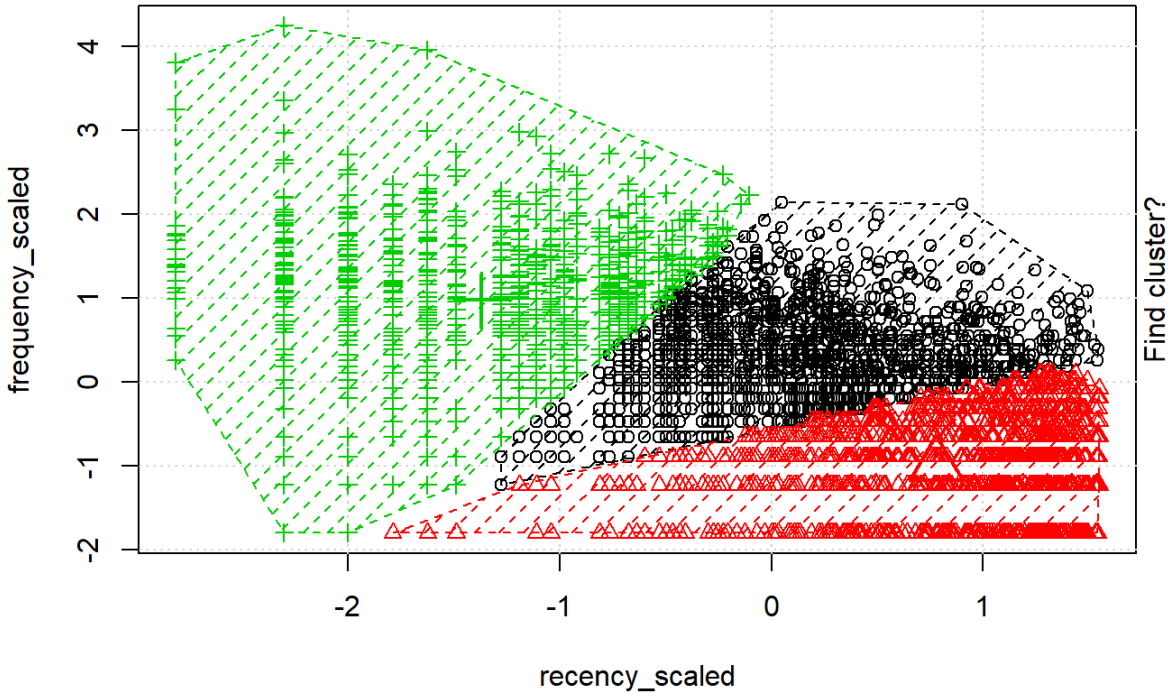

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

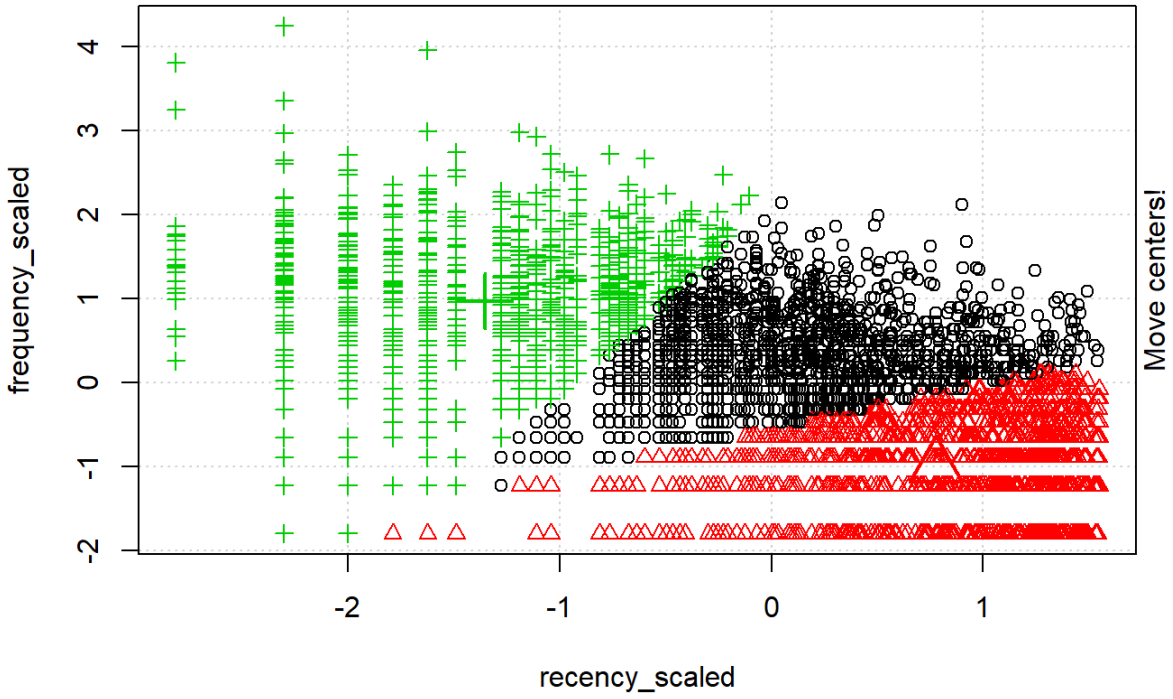

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

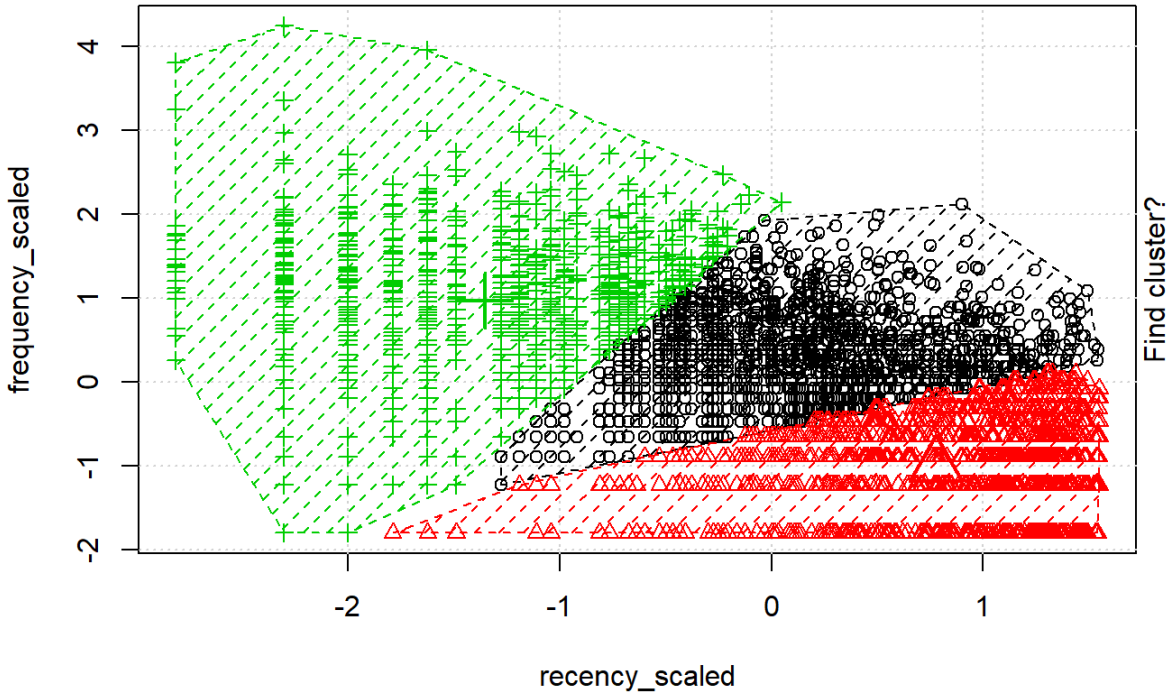

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

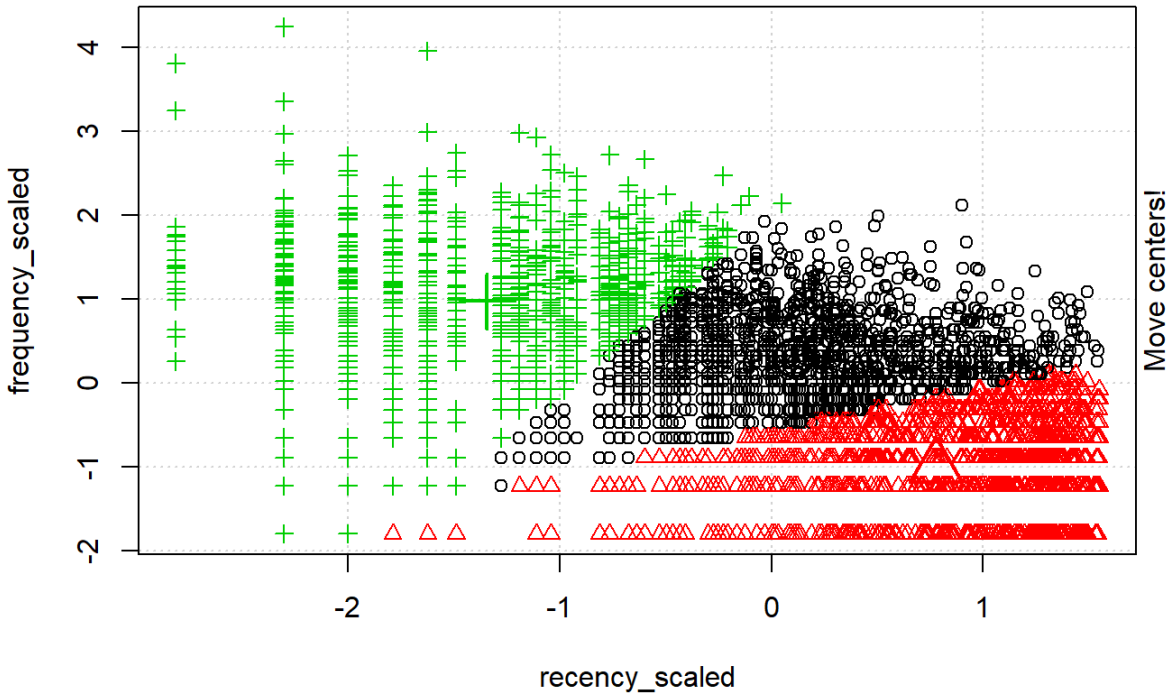

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

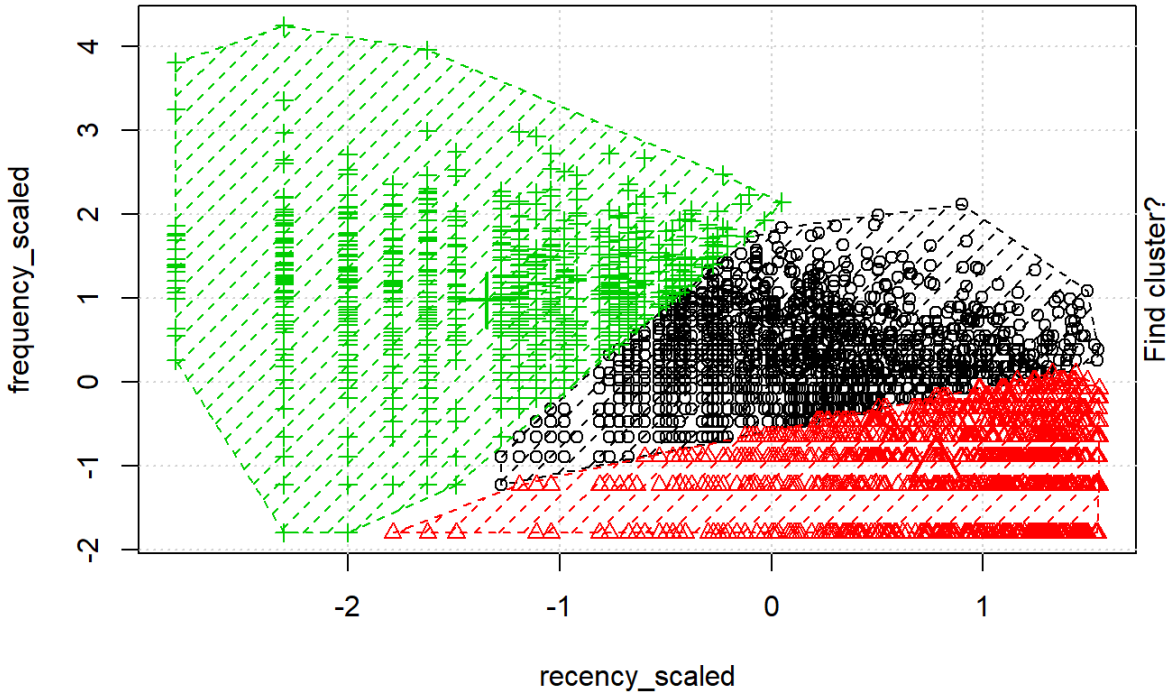

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

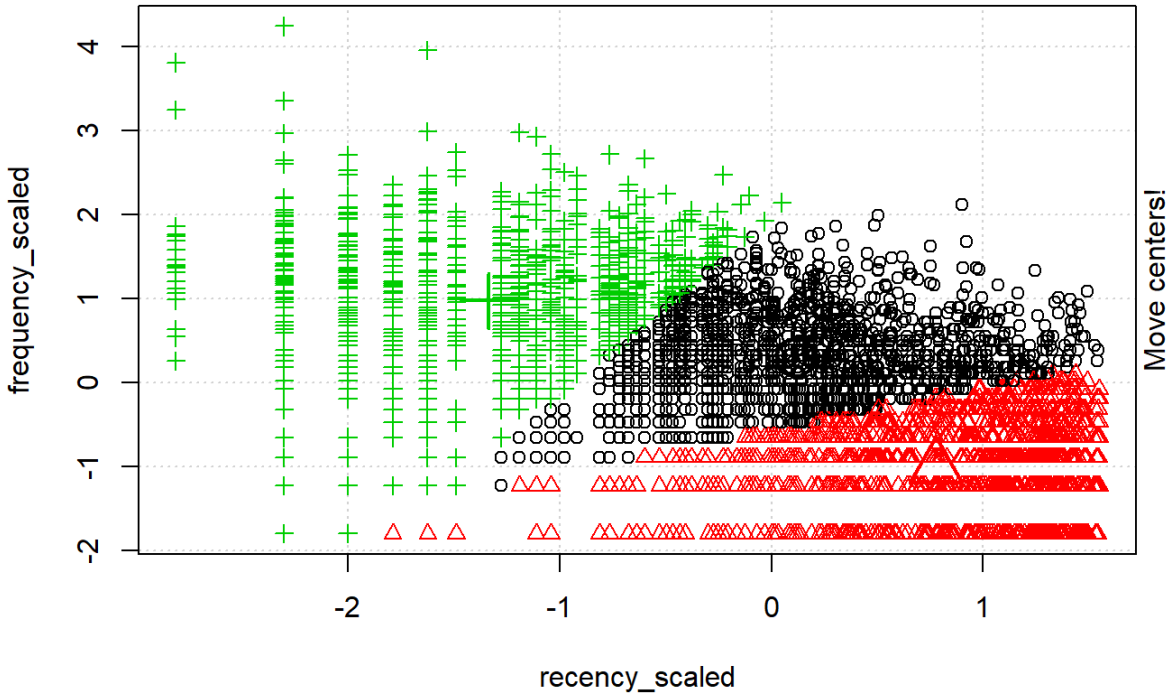

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

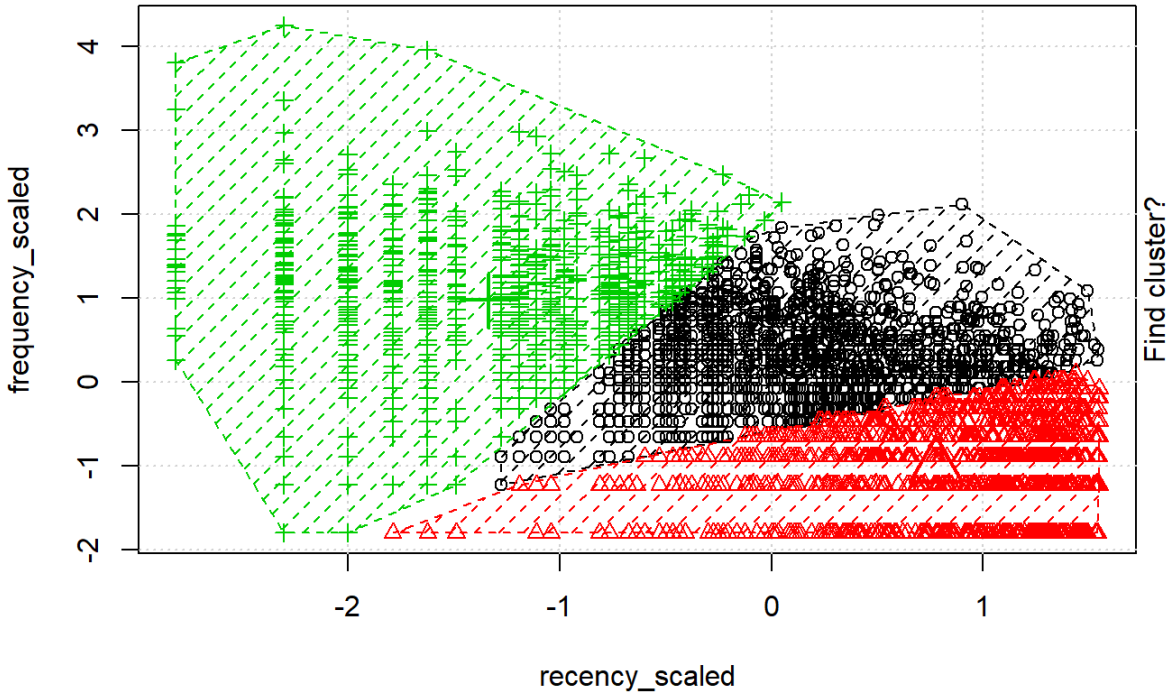

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

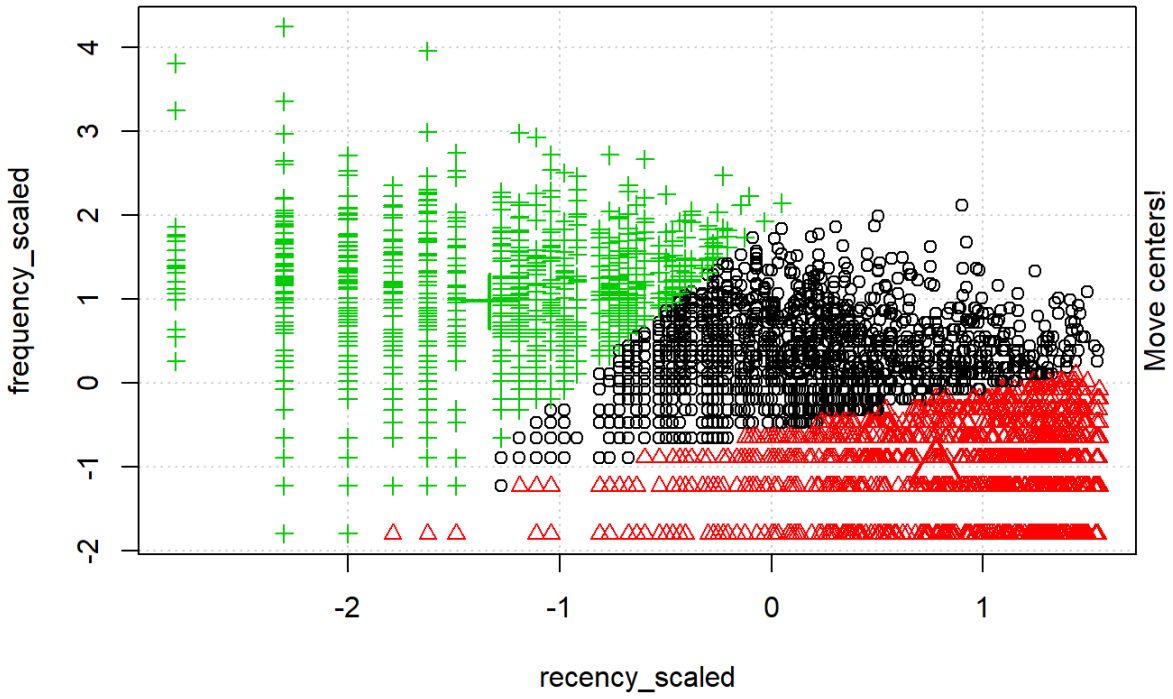

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

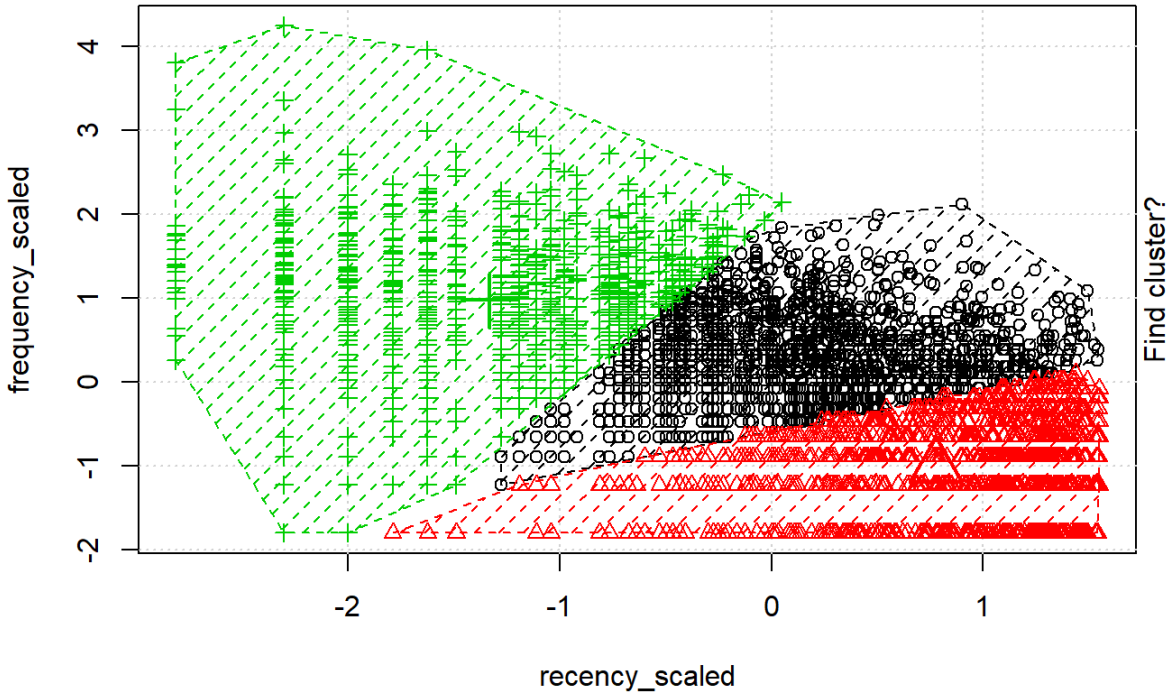

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

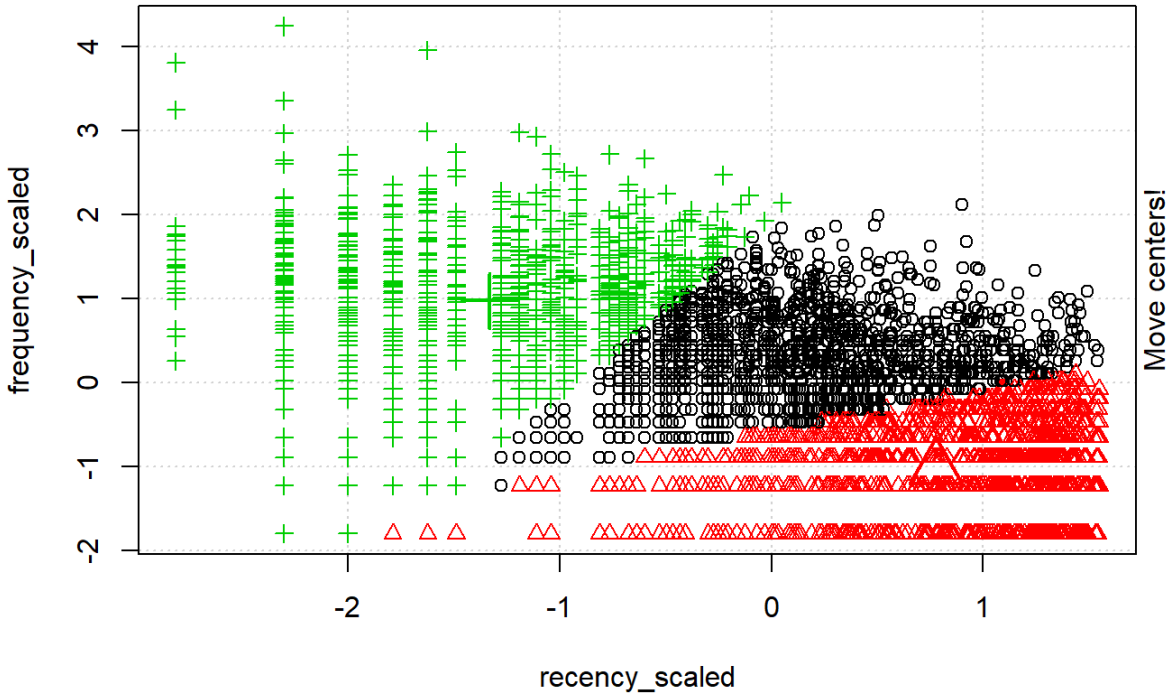

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

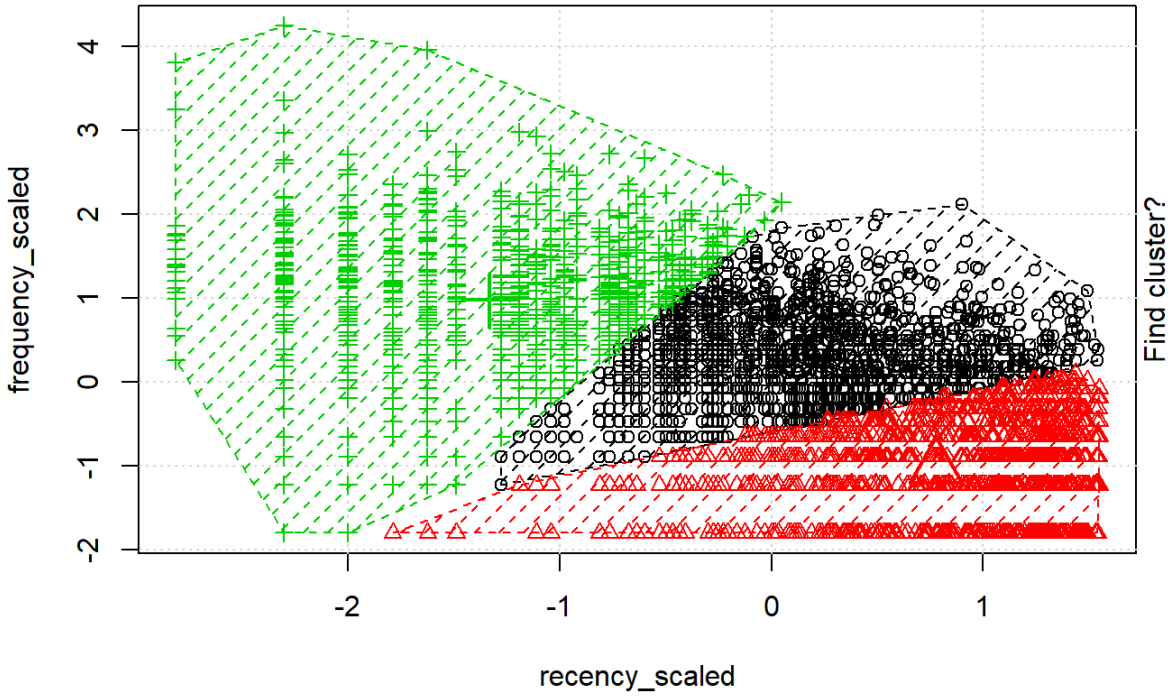

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

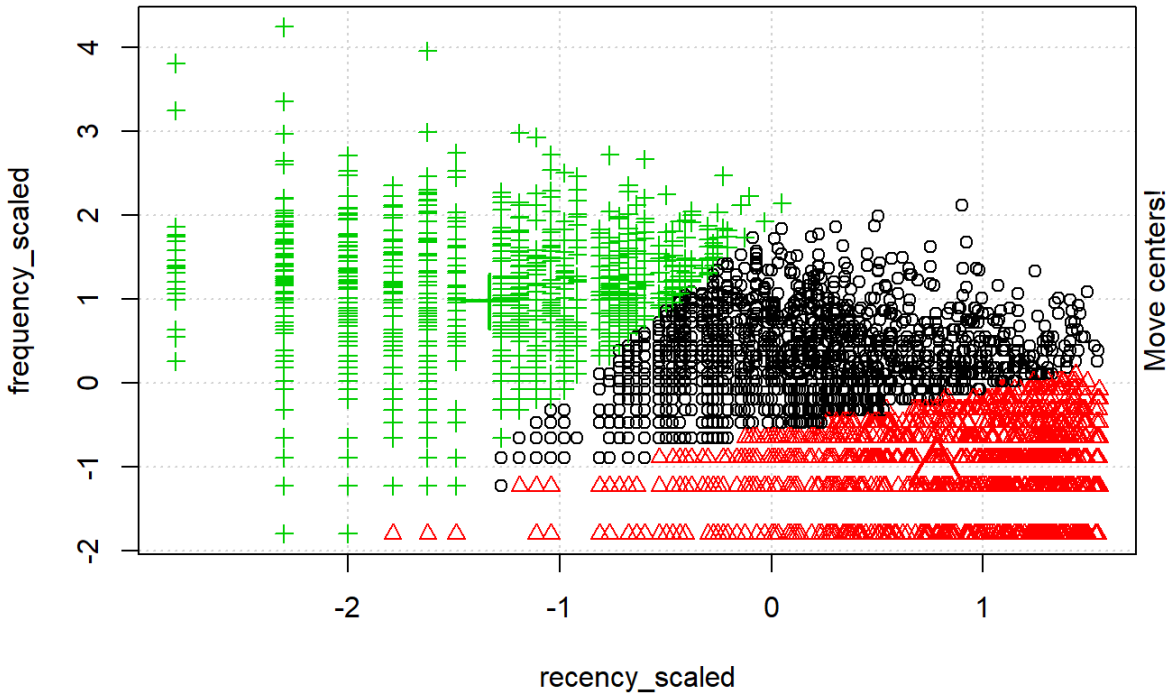

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

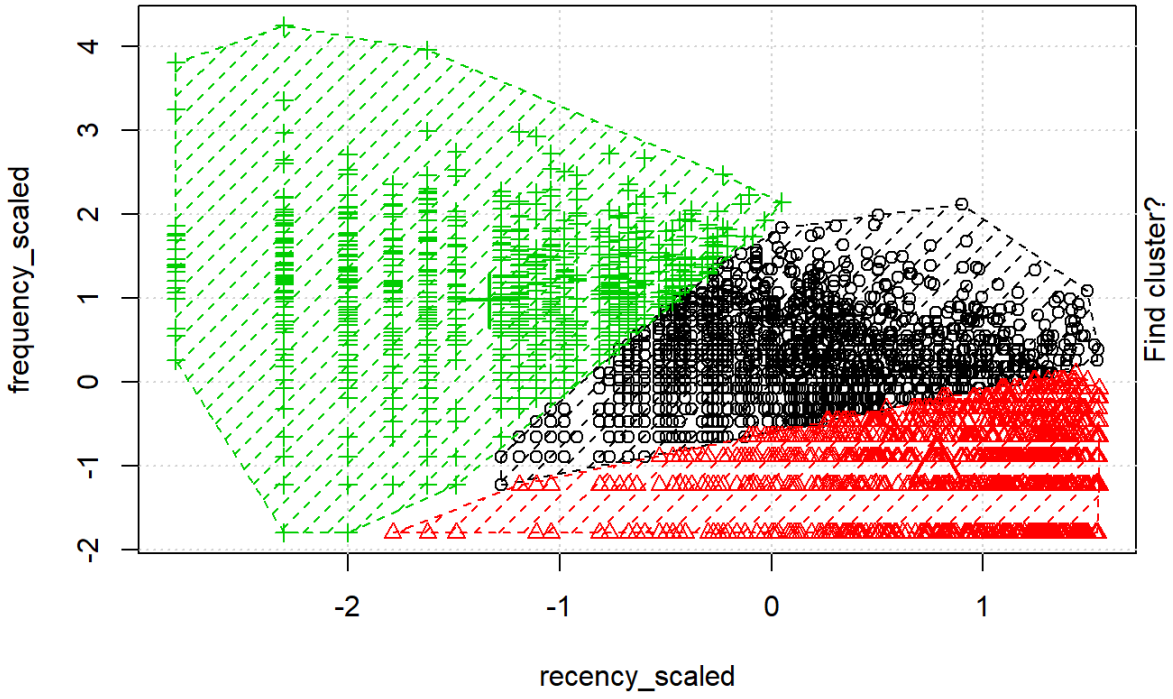

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

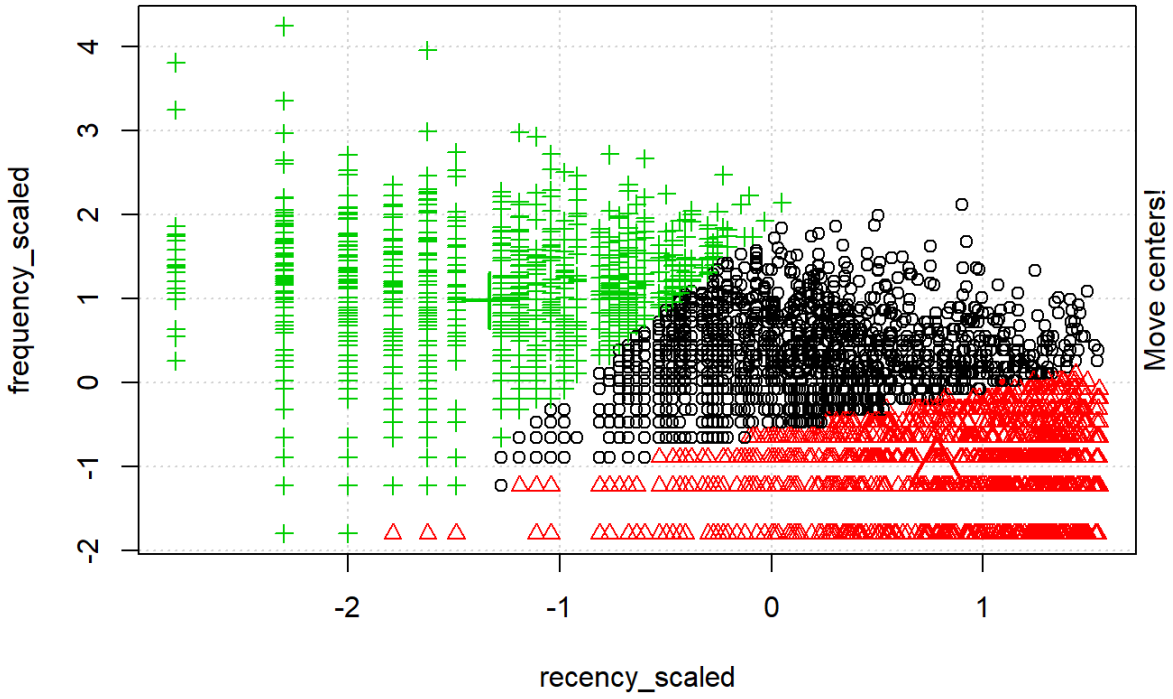

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

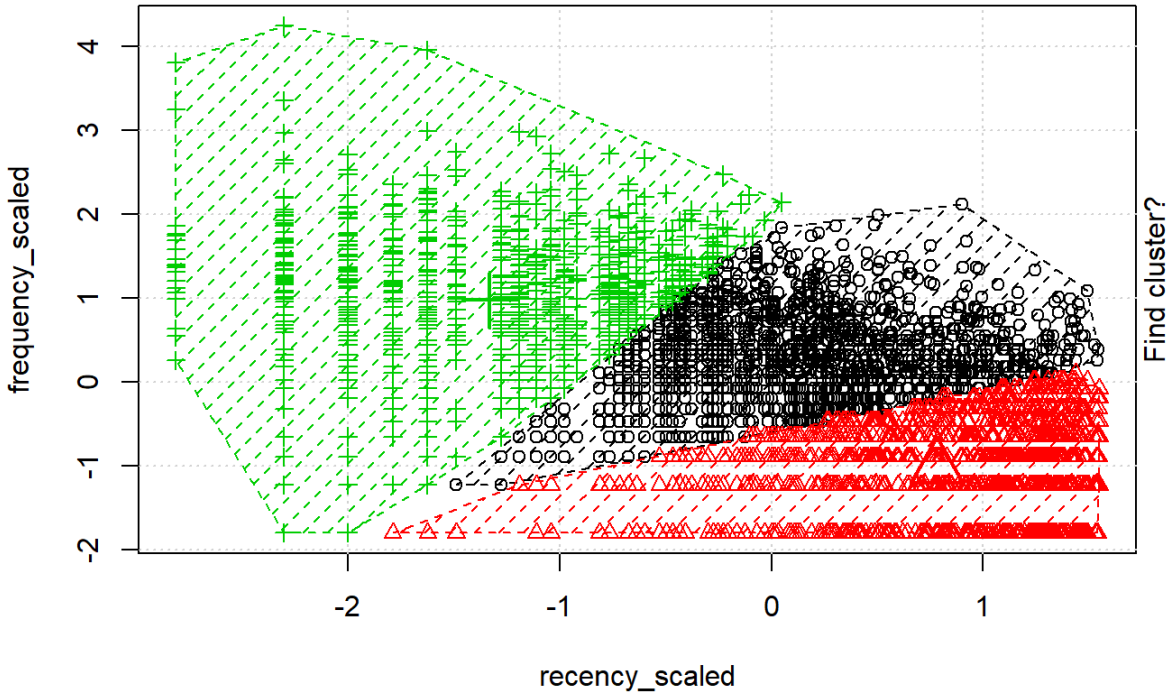

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4

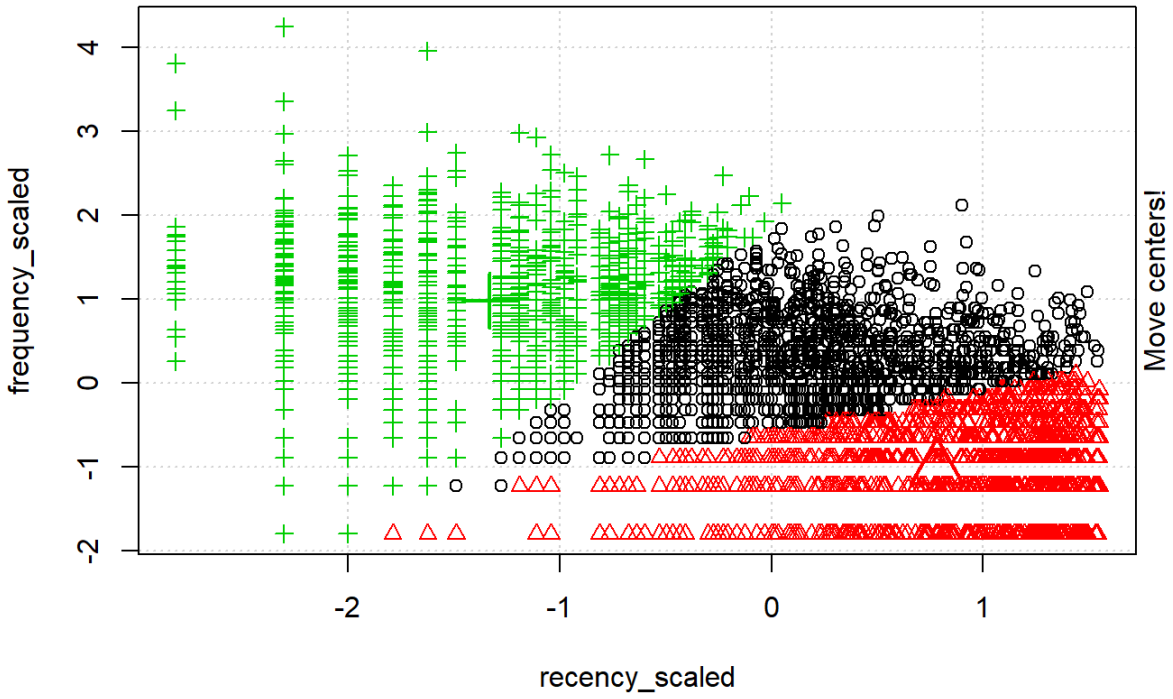

frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4

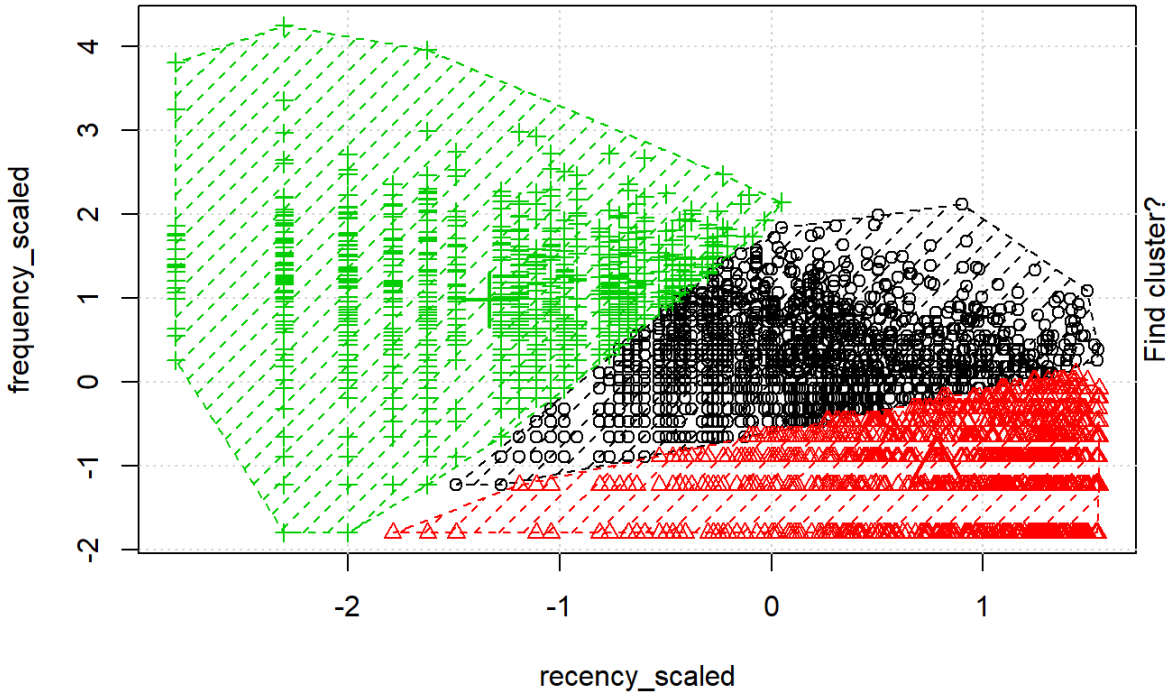

frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4


frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4


frequency_scaled
recency_scaled
Move centers!
-2
-1
0
1
2
3
4


frequency_scaled
recency_scaled
Find cluster?
-2
-1
0
1
2
3
4





To cluster the customers into manageable groups of customers so that Roomba can start its new Roomba subscription model, I used k-means clustering with the dataset containing Recency, Frequency and Moneary values of customers. Firstly, I used Elbow method to identity the number of clusters. From the plot, we can see that 3 is a good choice. Then I used k-means to cluster the customoers. As the result shows above, there are three colours (black, green, red) in the plots, which means it forms 3 clusters.