# E-commerce platform customers segmentation

Jaime Rubio (40425150)

May 09, 2024

- · Data preparation
  - Hierarchical clustering
- · RFM analysis
  - Kmeans based on RFM results

## **Data preparation**

```
#remove observations which purchase was cancelled
data <- subset(data, !startsWith(InvoiceNo, "C"))

#check how many values are higher that 1
filtered_data <- data[data$ReturnRate > 1, ]

# Calculate the percentage of the filtered data set compared to the original datas et
percentage_higher_than_1 <- (nrow(filtered_data) / nrow(data)) * 100

# Print the percentage
print(percentage_higher_than_1)</pre>
```

```
## [1] 1.506361
```

```
#remove rows where returnrate is higher than 1
data <- data[data$ReturnRate <= 1, ]

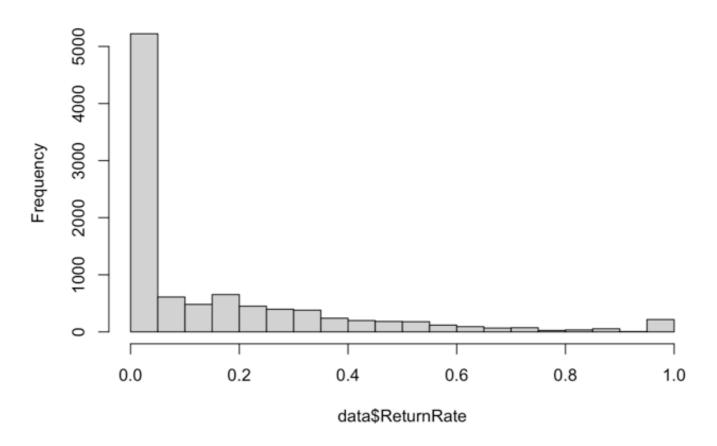
#convert invocedate into the correct format
data$InvoiceDate <- as.POSIXct(data$InvoiceDate, format = "%Y-%m-%dT%H:%M")

#mutate work levels
data <- data %>%
    mutate(Work = case_when(
        Work == 1 ~ "Health services",
        Work == 2 ~ "Financial services",
        Work == 3 ~ "Sales",
        Work == 4 ~ "Advertising/PR",
        Work == 5 ~ "Education",
        Work == 6 ~ "Industrial Sector",
        Work == 7 ~ "Engineering",
```

```
Work == 8 ~ "Technology",
    Work == 9 ~ "Retail & Services",
    Work == 10 ~ "Self-Employed",
    Work == 11 ~ "Other"
  ))
#mutate Education levels
data <- data %>%
  mutate(Edcation = case_when(
    Edcation == 1 ~ "High School",
    Edcation == 2 ~ "Undergraduate",
    Edcation == 3 ~ "Postgraduate"
  ))
### Marriage
data <- data %>%
  mutate(Married = case_when(
    Married == 1 ~ "Married",
    Married == 0 ~ "Single"
  ))
##convert categorical variables into factors
data$Work <- as.factor(data$Work)</pre>
data$Edcation <- as.factor(data$Edcation)</pre>
data$Married <- as.factor(data$Married)</pre>
data$ZipCode <- as.factor(data$ZipCode)</pre>
## imputation of missing values in customer ID
#create a data set with all missing values
data_na <- data %>%
  filter(is.na(CustomerID))
#filter the original data for non missing values
data <- data %>%
  filter(!is.na(CustomerID))
##same InvoceNo same customer
data na <- data na %>%
  group_by(InvoiceNo) %>%
  mutate(CustomerID = cur_group_id())
# join both data sets
data <- bind_rows(data, data_na)</pre>
#mode for categorical variables
```

```
get mode <- function(v) {</pre>
  uniqv <- unique(v)</pre>
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
#group by customer id w mean for numerical and mode for categorical
new_data <- data %>%
  group by(CustomerID) %>%
  summarise(Age = mean(Age),
            Work = get_mode(Work),
            Avg_Quantity = mean(Quantity),
            Total_Quantity = sum(Quantity),
            total_value = sum(Quantity * UnitPrice),
            Avg UnitPrice = mean(UnitPrice),
            Married = get_mode(Married),
            total_invoice = n_distinct(InvoiceNo),
            Avg ReturnRate = mean(ReturnRate),
            Income = mean(Income),
            Edcation = get_mode(Edcation),
            zipcode = get mode(ZipCode))%>% filter(
              total value >= quantile(total value, 0.025),
              total value <= quantile(total value, 0.975),
              Total Quantity >= quantile(Total Quantity, 0.025),
              Total_Quantity <= quantile(Total_Quantity, 0.975))</pre>
hist(data$ReturnRate)
```

### Histogram of data\$ReturnRate



# Clustering

## Hierarchical clustering

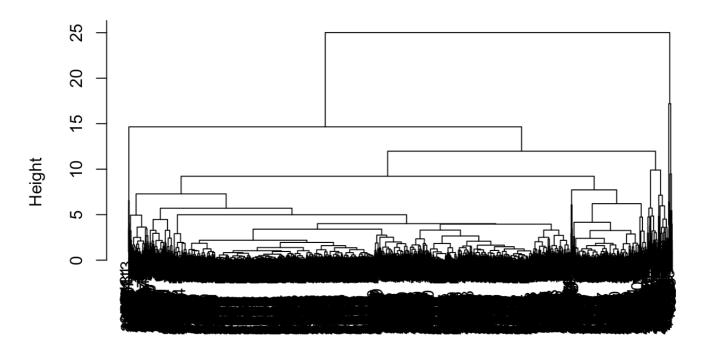
#hierarchical clustering whit 4 linkage methods ---hclust<- hclust(dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity, n
ew\_data\$total\_value, new\_data\$Avg\_UnitPrice, new\_data\$Avg\_ReturnRate))), method =
"complete")
hclust1<- hclust(dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity,
new\_data\$total\_value, new\_data\$Avg\_UnitPrice, new\_data\$Avg\_ReturnRate))), method =
"single")
hclust2<- hclust(dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity,
new\_data\$total\_value, new\_data\$Avg\_UnitPrice, new\_data\$Avg\_ReturnRate))), method =
"centroid")
hclust3<- hclust(dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity,
new\_data\$total\_value, new\_data\$Avg\_UnitPrice, new\_data\$Avg\_ReturnRate))), method =
"average")</pre>

```
#different nstart values
nstart_values <- c(10, 50, 100)

x <- c(1:10)

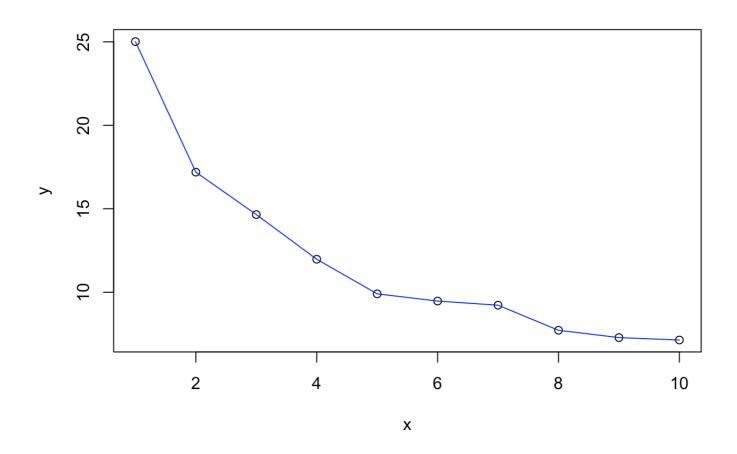
#for complete method---
plot(hclust)</pre>
```

#### **Cluster Dendrogram**



dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity, new data\$total\_value, new local tast (%); d'colmitte new data\$Avg\_ReturnRate)))

```
y <- sort(hclust$height, decreasing = TRUE)[1:10]
plot(x,y); lines(x,y, col= "blue")</pre>
```



```
results <- vector("list", length = 3)
for (i in 1:length(nstart_values)) {
    seg_kmeans <- kmeans(x = data.frame(new_data$Avg_Quantity, new_data$Total_Quanti
    ty, new_data$total_value, new_data$Avg_UnitPrice, new_data$Avg_ReturnRate), center
    s = 8, nstart = nstart_values[i])
    results[[i]] <- seg_kmeans
}

# Comparing results
for (i in 1:length(results)) {
    cat("Results for nstart =", nstart_values[i], ":\n")
    print(results[[i]])
    cat("\n")
}</pre>
```

```
## Results for nstart = 10 :
## K-means clustering with 8 clusters of sizes 62, 1278, 71, 46, 612, 187, 237, 37
7
##
## Cluster means:
## new_data.Avg_Quantity new_data.Total_Quantity new_data.total_value
## 1 34.182397 108.951613 108.64871
```

```
## 2
               4.345405
                                   5.668232
                                                     11.82960
## 3
              11.161622
                                   42.591549
                                                     175.12662
## 4
              30.022314
                                  119.326087
                                                     232.29000
## 5
               6.099323
                                   14.369281
                                                     34.67489
## 6
              13.388517
                                   49.074866
                                                     104.37406
## 7
              25.430058
                                   40.654008
                                                     28.10646
## 8
               8.704253
                                   28.448276
                                                     66.20294
##
    new data.Avg UnitPrice new data.Avg ReturnRate
## 1
                1.696336
                                   0.1274866
## 2
                3.372286
                                   0.1428583
## 3
               19.085038
                                   0.1754068
## 4
                                   0.1747545
                3.116224
## 5
                4.156557
                                   0.1577674
##
  6
                3.396130
                                   0.1557103
##
  7
                1.043599
                                   0.1429953
## 8
                5.121569
                                   0.1465778
##
## Clustering vector:
##
     [1] \ 3 \ 5 \ 2 \ 2 \ 2 \ 5 \ 6 \ 5 \ 3 \ 2 \ 7 \ 2 \ 4 \ 5 \ 2 \ 2 \ 2 \ 6 \ 8 \ 6 \ 2 \ 5 \ 2 \ 8 \ 2 \ 8 \ 2 \ 2 \ 3 \ 6 \ 4 \ 2 \ 2 \ 2 \ 2
2
##
    [38] 6 3 5 8 8 6 2 5 8 2 2 2 2 2 2 2 2 2 2 2 2 2 2 8 6 5 2 2 5 3 2 6 5 7 8
7
##
    [75] 5 2 8 6 6 5 5 5 2 2 2 3 5 2 6 2 5 8 2 2 6 2 2 5 8 8 8 3 2 2 8 8 2 5 2 2
8
##
   2
##
   [149] 8 2 2 2 2 5 5 5 2 5 2 5 2 5 5 5 5 2 2 2 5 6 5 3 2 2 2 8 8 2 2 2 2 2 2 2 5 5
2
##
   2
##
   [223] 2 2 2 2 2 5 2 5 2 5 5 5 5 5 8 2 8 2 3 2 2 2 2 5 6 2 2 2 2 2 2 5 2 2 2 2
2
##
   5
   [297] 5 2 5 2 2 7 7 5 8 5 2 2 2 8 2 2 2 5 2 2 2 5 5 2 3 2 2 2 2 2 2 2 8 2 2 8
##
2
##
   [334] 2 2 5 5 2 2 2 2 5 5 2 2 2 2 2 8 6 2 2 2 6 5 2 8 2 3 2 2 2 5 5 2 8 2 1
8
   ##
8
   ##
2
##
   [445] 2 2 2 7 5 5 2 8 2 2 2 5 5 2 7 2 2 5 2 2 2 6 2 5 5 2 6 2 8 2 8 5 3 2 5 2
2
##
   [482] 5 5 5 8 5 2 3 2 7 5 5 6 8 2 8 5 2 8 2 6 8 2 7 8 8 2 2 6 8 5 2 2 2 2 8 5
2
##
   [519] 5 2 2 2 2 5 5 2 2 8 2 2 8 2 2 3 8 2 2 5 5 5 6 2 5 2 2 5 2 5 2 8 5 2 2
2
##
   [556] 5 6 2 8 5 5 2 2 5 2 2 5 5 5 3 2 2 2 2 5 5 6 2 2 2 2 5 5 5 2 5 5 5 6 2
2
```

```
## [2443] 7 7 6 5 6 8 2 5 2 7 2 4 2 2 2 5 8 2 2 8 2 3 5 2 5 8 4 2 8 5 2 7 6 2 2 2
## [2480] 6 1 8 2 6 7 5 6 2 5 8 2 5 2 6 2 1 2 7 5 2 8 5 3 2 2 2 2 2 6 2 2 7 5 1 5
## [2517] 5 2 1 2 2 2 2 8 1 8 2 5 2 8 8 3 6 2 2 2 5 7 8 2 2 2 8 5 7 7 8 8 2 2 5 2
## [2554] 5 5 2 2 4 5 2 5 6 7 7 2 2 2 5 5 5 7 2 2 2 2 2 7 8 2 5 2 2 2 5 2 5 6 2
## [2591] 2 7 2 6 2 2 5 2 7 2 2 2 8 2 2 7 5 2 2 2 5 2 7 2 5 4 2 5 7 2 5 3 5 7 8
## [2628] 8 7 6 5 2 3 5 8 8 8 5 8 5 2 1 5 6 2 5 1 8 5 5 5 2 7 6 4 7 5 8 6 7 5 8 6
8
## [2665] 5 2 7 8 2 2 8 3 3 7 2 5 6 7 5 2 2 4 2 5 6 2 5 5 7 3 5 2 2 2 2 5 8 2 2 5
7
## [2702] 2 2 5 2 2 5 8 2 2 6 7 6 2 2 2 2 2 5 2 2 2 8 3 2 2 2 6 8 2 5 2 7 7 2 3 2
## [2739] 2 2 7 2 2 2 2 2 2 2 5 2 8 5 2 5 7 6 7 5 2 7 5 2 8 2 2 7 5 2 2 2 2 2 7 2
2
## [2776] 2 8 2 1 2 5 7 6 5 8 2 5 5 5 2 5 5 2 7 8 7 2 2 2 2 6 2 5 5 6 7 2 5 2 7
## [2813] 2 2 2 1 8 7 5 7 2 2 2 2 2 3 5 2 3 5 8 7 5 2 5 5 2 5 2 2 2 2 2 2 7 2 8
## [2850] 5 6 1 7 3 5 7 7 8 5 7 2 2 2 2 5 2 6 2 5 2
##
## Within cluster sum of squares by cluster:
## [1] 157993.41 110007.63 213283.07 167846.51 96291.63 126062.73 152693.52
## [8] 175488.73
   (between_SS / total_SS = 85.6 %)
##
##
## Available components:
## [1] "cluster"
                                      "totss"
                      "centers"
                                                     "withinss"
                                                                    "tot.withinss"
## [6] "betweenss"
                      "size"
                                      "iter"
                                                     "ifault"
##
## Results for nstart = 50 :
## K-means clustering with 8 clusters of sizes 191, 1227, 68, 363, 743, 175, 46, 5
7
##
## Cluster means:
##
     new_data.Avg_Quantity new_data.Total_Quantity new_data.total_value
## 1
                                         47.324607
                 12.360219
                                                               104.85775
## 2
                  4.158012
                                          5.358598
                                                                11.45873
## 3
                 10.999074
                                         41.514706
                                                               176.40353
## 4
                  7.246180
                                         23.000000
                                                                64.31471
## 5
                  9.051475
                                         16.939435
                                                                30.28202
## 6
                                         55.440000
                                                                46.51509
                 25.966610
## 7
                 30.022314
                                         119.326087
                                                               232.29000
## 8
                 33.327929
                                                               115.40544
                                         110.175439
##
     new data.Avg UnitPrice new data.Avg ReturnRate
## 1
                   3.552565
                                          0.1490001
```

##						3	3.4	47!	583	36								0.	. 14	153	369	9															
##									812										. 17																		
##									36!										. 15																		
##									782										. 14																		
##									37										. 15																		
##									622										. 17																		
##	8						L • .	/5	974	43								0.	. 13	3 <b>1</b> ]	L 3 :	9															
##	Cluste	rir	ıg	ve	ect	:01	c <b>:</b>																														
##	[1]	3	4	2	2	2	2	4	1	5	3	2	5	2	7	5	2	2	2	1	4	1	2	5	2	4	2	4	2	2	3	1	7	2	2	2	2
2																																					
##	[38]	1	3	5	4	4	1	2	4	4	2	2	2	2	2	2	2	2	2	2	2	2	5	7	2	4	1	5	2	2	5	3	2	1	5	6	4
##	[75]	5	2	4	1	1	5	5	4	2	2	2	3	5	2	1	2	5	4	2	2	1	2	2	5	4	4	4	3	2	2	6	4	2	5	2	2
4	[,0]	Ū	_	-	_	_		Ū	-	_	_	_			_	_	_	Ū	-		_	_	_	_		-	-	-	•	_	_	·	-	_	Ū	_	_
##	[112]	2	4	2	4	5	2	2	2	2	4	2	2	2	2	2	5	2	2	2	2	2	5	2	2	2	2	2	2	2	2	2	2	2	2	2	1
2 ##	[149]	4	2	2	2	2	5	5	2	5	2	5	2	5	5	4	2	2	2	5	1	5	3	2	2	2	4	4	2	2	2	2	2	2	2	2	5
2																																					
##	[186]	2	1	2	5	5	2	2	8	5	2	2	5	2	5	2	2	2	5	2	5	2	5	5	2	2	5	2	2	5	2	2	2	5	2	3	2
##	[223]	2	2	2	2	2	5	2	4	2	2	5	4	5	5	4	2	4	2	3	2	2	2	2	5	1	2	2	2	2	2	2	5	2	2	2	2
2 ##	[260]	5	2	2	2	4	2	2	2	2	4	4	2	2	5	2	5	2	2	2	2	2	3	2	3	5	2	2	2	2	5	2	2	2	2	2	2
5																																					
##	[297]	5	2	5	2	2	6	6	5	4	5	2	2	2	4	2	2	2	2	5	2	2	2	5	5	2	3	2	2	2	2	2	2	4	2	2	4
##	[334]	2	2	2	5	2	2	2	2	5	5	2	2	2	2	2	4	1	2	2	2	1	2	2	4	2	3	2	2	2	2	4	5	2	4	2	8
4 ##	[371]	5	5	5	5	2	2	2	2	2	2	2	5	5	5	2	5	2	5	2	2	2	2	2	2	5	5	1	2	2	2	2	5	5	5	2	2
4	[3/1]	J	J	,	J	2	2	_	۷	2	2	2	J	J	J	2	J	2	J	,	۷	2	2	2	2	J	J	7	۷	۷	2	2	,	J	,	2	2
##	[408]	5	2	2	2	2	5	2	4	2	2	5	1	5	2	2	2	2	4	5	2	4	2	4	5	2	2	2	4	1	2	4	5	4	5	3	2
2 ##	[445]	2	2	2	6	<b>-</b>	_	2	1	2	2	2	_	<b>-</b>	2	6	2	2	1	2	2	2	1	2	_	5	2	1	2	1	2	1	<b>-</b>	2	2	5	2
2	[445]	2	2	2	O	5	5	2	4	2	2	۷	5	5	۷	O	2	2	4	2	2	2	1	2	5	5	2	1	2	4	2	4	5	3	2	5	2
##	[482]	5	5	5	4	5	2	3	2	5	5	5	1	4	2	4	5	2	4	2	1	4	2	6	4	4	2	2	1	4	5	2	2	2	2	4	5
2																																					
##	[519]	5	2	2	2	2	5	5	2	2	4	2	2	4	2	2	3	4	2	2	5	5	5	5	1	2	5	2	2	5	2	5	2	4	5	2	2
2																																					
##	[556]	5	1	2	4	5	5	2	2	5	2	2	5	5	5	3	2	2	2	5	5	5	1	2	2	2	2	4	5	5	2	5	5	5	5	1	2
2																																					
##	[593]	5	5	5	2	4	5	4	6	7	2	5	5	2	2	2	2	4	5	5	4	4	3	5	5	2	4	5	2	2	2	2	2	2	5	4	5
5		_	_			_	_	_		_		_		_	_	_		_	_		_	_	_	_			_		_	_	_			_	_	_	
##	[630]	2	2	4	4	2	2	5	4	5	1	2	4	2	2	5	4	5	3	Τ	5	5	2	2	4	Τ	3	Τ	3	2	5	4	Τ	5	5	5	4
1   ##	[667]	E	2	2	2	0	1	1	6	1	2	2	6	_	E	E	E	2	1	1	E	2	2	2	1	1	2	1	2	E	2	2	2	2	2	<b>E</b>	5
7	[00/]	5	2	2	2	Ö	4	Т	O	4	2	2	О	5	3	3	Э	2	Т	4	5	3	2	2	Т	Т	2	Т	2	3	2	2	2	2	2	S	S
##	[704]	4	3	5	6	7	1	1	2	3	8	5	8	2	5	5	1	1	6	1	2	1	4	1	5	4	6	5	2	7	2	1	5	4	1	2	4
5		_	1	F	2	2	_	2	1	•	Λ	F	1	2	1	-	F	2	0	1	F	_	F	7	F	1	2	2	F	_	F	F	2	1	1	2	_
##	[741]	О	Ţ	Э	2	2	О	2	4	2	4	Э	Τ	2	Τ	Э	Э	2	g	4	Э	О	Э	/	Э	4	2	2	Э	О	Э	Э	2	Τ	1	2	Э

```
5
## [2628] 4 6 1 4 2 3 5 4 4 6 5 4 5 2 8 5 1 2 5 6 4 5 4 5 2 5 1 7 6 5 4 1 5 5 4 1
## [2665] 5 2 6 4 2 2 4 3 3 6 5 5 6 5 5 2 2 7 2 5 1 5 5 5 5 3 5 2 2 2 2 5 4 2 2 5
## [2702] 2 2 5 2 2 5 4 2 2 1 5 1 2 2 2 2 2 5 2 2 4 3 2 2 2 1 1 2 4 2 6 5 2 3 2
## [2739] 2 2 6 2 2 2 2 2 2 2 5 5 4 5 2 5 6 6 5 2 6 5 2 4 2 2 5 5 2 2 2 2 2 5 5 2
## [2776] 2 6 2 8 2 4 5 1 5 4 5 5 5 5 2 4 5 5 2 2 4 6 2 2 2 2 1 2 5 5 1 6 2 4 2 6
2
## [2813] 2 2 2 8 4 6 5 6 2 2 5 2 2 3 5 2 3 5 4 5 5 2 5 5 2 5 5 2 2 2 2 2 5 2 4
## [2850] 5 1 8 6 3 5 5 6 4 5 5 2 2 2 2 5 2 1 2 5 2
##
## Within cluster sum of squares by cluster:
## [1] 119383.5 103846.7 206730.0 146454.8 162252.0 141391.3 167846.5 141191.8
##
   (between SS / total SS = 85.7 %)
##
## Available components:
##
                                                                     "tot.withinss"
## [1] "cluster"
                      "centers"
                                      "totss"
                                                      "withinss"
## [6] "betweenss"
                      "size"
                                      "iter"
                                                     "ifault"
## Results for nstart = 100:
## K-means clustering with 8 clusters of sizes 59, 192, 43, 68, 175, 743, 363, 122
7
##
## Cluster means:
##
     new data. Avg Quantity new data. Total Quantity new data. total value
## 1
                 36.994779
                                         110.084746
                                                                119.05492
## 2
                 12.384385
                                          47.520833
                                                                104.78490
## 3
                 25.140149
                                         120.674419
                                                                236.00791
## 4
                 10.999074
                                          41.514706
                                                                176.40353
## 5
                 25.966610
                                          55.440000
                                                                 46.51509
## 6
                  9.051475
                                          16.939435
                                                                 30.28202
## 7
                  7.246180
                                          23.000000
                                                                 64.31471
## 8
                  4.158012
                                           5.358598
                                                                 11.45873
##
     new data.Avg UnitPrice new data.Avg ReturnRate
## 1
                   1.754701
                                           0.1357395
## 2
                   3.545249
                                           0.1486092
## 3
                   3.208751
                                           0.1728114
## 4
                  19.798128
                                           0.1770462
## 5
                   1.423715
                                           0.1557503
## 6
                   3.077822
                                           0.1449525
## 7
                   6.063650
                                           0.1555215
## 8
                   3.475836
                                           0.1453699
##
## Clustering vector:
##
      [1] 4 7 8 8 8 8 7 2 6 4 8 6 8 3 6 8 8 8 2 7 2 8 6 8 7 8 7 8 8 4 2 3 8 8 8 8
```

8

```
8
##
  [2813] 8 8 8 1 7 5 6 5 8 8 6 8 8 4 6 8 4 6 7 6 6 8 6 6 6 6 8 8 8 8 8 8 6 8 7
6
## [2850] 6 2 1 5 4 6 6 5 7 6 6 8 8 8 8 6 8 2 8 6 8
##
## Within cluster sum of squares by cluster:
  [1] 163888.7 121013.6 141812.7 206730.0 141391.3 162252.0 146454.8 103846.7
    (between SS / total SS = 85.7 %)
##
## Available components:
##
## [1] "cluster"
                                      "totss"
                                                                     "tot.withinss"
                      "centers"
                                                     "withinss"
## [6] "betweenss"
                      "size"
                                      "iter"
                                                     "ifault"
```

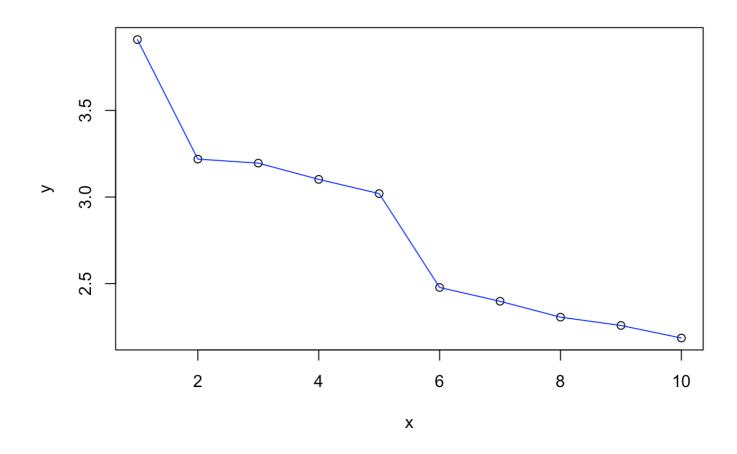
```
#for single method
plot(hclust1)
```

## **Cluster Dendrogram**



dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity, new\_data\$total\_value, new\_data\$Avg\_ReturnRate)))

```
y <- sort(hclust1$height, decreasing = TRUE)[1:10]
plot(x,y); lines(x,y, col= "blue")</pre>
```



```
results1 <- vector("list", length = 3)
for (i in 1:length(nstart_values)) {
    seg_kmeans1 <- kmeans(x = data.frame(new_data$Avg_Quantity, new_data$Total_Quant
    ity, new_data$total_value, new_data$Avg_UnitPrice, new_data$Avg_ReturnRate), cente
rs = 6, nstart = nstart_values[i])
    results1[[i]] <- seg_kmeans1
}

# Comparing results
for (i in 1:length(results1)) {
    cat("Results for nstart =", nstart_values[i], ":\n")
    print(results1[[i]])
    cat("\n")
    cat("Results for nstart =", nstart_values[i], ":\n")
    cat("Results for nstart =", nstart_values[i], ":\n")
    cat("tot.withinss:", results1[[i]]$tot.withinss, "\n\n")
}</pre>
```

```
## Results for nstart = 10 :
## K-means clustering with 6 clusters of sizes 131, 1449, 334, 154, 740, 62
##
## Cluster means:
## new_data.Avg_Quantity new_data.Total_Quantity new_data.total_value
```

```
## 1
          33.557189
                        75.832061
                                     67.06481
## 2
          4.876893
                         6.418219
                                     13.16578
## 3
          8.511582
                        29.140719
                                     78.47461
## 4
          12.075180
                        59.480519
                                     138.69299
## 5
          10.230337
                        22.090541
                                     37.49609
## 6
          26.703191
                       102.725806
                                     228.28145
##
   new_data.Avg_UnitPrice new_data.Avg_ReturnRate
## 1
           1.519733
                         0.1389463
## 2
           3.438270
                         0.1445157
## 3
           6.424934
                         0.1495671
## 4
           6.302775
                         0.1515568
## 5
           3.221781
                         0.1534203
## 6
           9.493247
                         0.1832171
##
## Clustering vector:
##
   [1] 4 5 2 2 2 2 5 3 5 4 2 5 2 6 5 2 2 2 3 3 4 2 5 2 3 2 3 2 2 4 3 6 2 2 2 2
2
##
   5
   [75] 5 2 5 3 4 5 5 5 2 2 2 3 5 2 4 2 5 3 2 2 3 2 2 5 3 3 3 6 2 2 5 3 2 5 2 2
##
3
##
  2
##
  2
  ##
2
##
  2
##
  5
##
  [297] 5 2 5 2 2 5 5 2 3 5 2 2 2 3 2 2 2 2 5 5 2 6 2 2 2 2 2 3 2 2 5
2
  ##
3
##
  3
##
  [408] 5 2 2 2 2 2 2 5 2 2 5 4 5 2 2 2 2 3 2 2 3 2 2 2 2 2 3 3 2 5 5 3 5 4 2
2
##
  2
##
  [482] 5 5 2 3 2 2 4 2 5 2 5 3 3 2 2 2 5 2 4 3 2 5 5 3 2 2 4 5 2 2 2 2 3 2
2
##
  2
##
  [556] 5 3 2 3 5 2 2 2 5 2 2 2 5 4 2 2 2 2 5 3 2 2 2 2 5 5 5 5 5 5 5 5 3 2
2
  [593] 2 5 2 2 5 5 3 1 6 2 5 5 2 2 2 2 3 5 5 5 4 2 5 2 3 2 2 2 2 2 2 3 5
##
5
##
  [630] 2 2 3 5 2 2 5 3 5 4 2 5 2 2 5 3 5 4 4 2 5 2 2 3 4 4 3 4 2 5 5 4 5 5 5 5
```

```
2
## [2517] 5 2 1 2 2 2 2 3 1 5 2 5 2 3 3 4 3 2 2 2 5 5 3 2 2 2 3 5 5 1 5 5 2 2 5 2
## [2554] 5 5 2 2 6 5 2 5 3 5 1 2 2 2 5 2 5 2 2 2 2 2 2 5 3 2 5 2 2 2 5 4 2
5
## [2628] 3 1 4 5 2 4 5 3 5 3 5 5 5 2 1 2 1 2 5 1 3 5 5 5 2 2 3 6 5 5 5 4 5 5 5 3
3
## [2665] 5 2 1 5 2 2 5 4 4 1 2 5 1 2 5 2 2 6 2 5 3 2 5 5 2 4 5 2 2 2 2 5 3 2 2 2
## [2702] 2 2 5 2 2 5 3 2 2 1 5 3 2 2 2 2 2 2 2 5 2 2 2 5 4 2 2 2 4 3 2 5 2 1 5 2 4 2
## [2739] 2 2 5 2 2 2 2 2 2 2 2 2 3 2 2 5 5 1 1 2 2 5 2 2 3 2 2 5 5 2 2 2 2 2 5 5 2
## [2850] 5 3 4 5 4 2 2 5 3 5 5 2 2 2 5 5 2 4 2 5 2
##
## Within cluster sum of squares by cluster:
## [1] 254354.4 172570.2 221392.8 261627.3 253138.8 314066.2
##
   (between SS / total SS = 82.2 %)
##
## Available components:
##
## [1] "cluster"
                   "centers"
                                                           "tot.withinss"
                                "totss"
                                              "withinss"
## [6] "betweenss"
                   "size"
                                "iter"
                                              "ifault"
##
## Results for nstart = 10:
## tot.withinss: 1477150
##
## Results for nstart = 50:
## K-means clustering with 6 clusters of sizes 740, 142, 1419, 57, 137, 375
##
## Cluster means:
    new data. Avg Quantity new data. Total Quantity new data. total value
##
                                    22.214865
## 1
               10.979155
                                                       35.66677
               30.058724
                                    79.690141
## 2
                                                       79.44380
## 3
               4.668089
                                     6.193798
                                                       13.00441
## 4
               28.678617
                                   110.842105
                                                      228.15930
## 5
               10.923347
                                    51.708029
                                                      146.36314
## 6
               8.647979
                                    29.058667
                                                       75.06128
##
    new data.Avg UnitPrice new data.Avg ReturnRate
## 1
                3.096915
                                     0.1516628
## 2
                1.690646
                                     0.1403995
## 3
                3.476512
                                     0.1454914
## 4
                4.891120
                                     0.1797195
## 5
                10.836077
                                     0.1549551
```

##	6					į	5.4	445	559	93								0 .	. 14	185	529	93															
##	<b>a</b> 1						. <b>.</b> .																														
##	Cluste		_					1	c	1	_	2	1	2	1	1	2	2	2	c	c	_	2	1	2	c	2	c	2	2	_	c	1	2	2	2	2
3	[1]	Э	Τ	3	3	3	3	Τ	О	Τ	Э	3	1	3	4	1	3	3	3	О	О	Э	3	Τ	3	О	3	О	3	3	Э	О	4	3	3	3	3
##	[38]	5	5	1	6	6	6	3	1	6	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	6	5	1	3	3	3	5	3	5	1	2	6
1	[30]	J	J	_	U	U	U	5	_	U	5	5	5	5	J	5	5	5	J	J	5	5	5	7	5	U	J	_	J	5	J	J	J	J	_	_	U
##	[75]	1	3	1	6	5	1	1	1	3	3	3	5	1	3	5	3	1	6	3	3	6	3	3	1	6	6	6	4	3	3	1	6	3	1	3	3
6	[,3]	-	•	-	Ü	J	-	-	-	•	•	•	,	-	J	,	•	-	Ū	J	•	Ū	•	J	_	Ü	Ü	Ü	•	J	J	_	Ū	•	-	J	J
##	[112]	3	6	3	6	1	3	3	3	3	6	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	6
3																																					
##	[149]	6	3	3	3	3	1	3	3	1	3	1	3	3	3	1	3	3	3	1	2	3	5	3	3	3	6	6	3	3	3	3	3	3	3	3	1
3																																					
##	[186]	3	6	3	3	1	3	3	2	1	3	3	1	3	3	3	3	3	1	3	1	3	1	1	3	3	3	3	3	3	3	3	3	1	3	5	3
3																																					
##	[223]	3	3	3	3	3	3	3	1	3	3	1	1	1	1	6	3	6	3	5	3	3	3	3	3	6	3	3	3	3	3	3	1	3	3	3	3
3																																					
##	[260]	3	3	3	3	6	3	3	3	3	6	1	3	3	1	3	1	3	3	3	3	3	4	3	5	1	3	3	3	3	3	3	3	3	3	3	3
1																																					
##	[297]	1	3	1	3	3	1	1	3	6	1	3	3	3	6	3	3	3	3	1	3	3	3	1	1	3	5	3	3	3	3	3	3	6	3	3	6
3										_							_	_				_			_		_				_					_	
##	[334]	3	3	3	1	3	3	3	3	1	3	3	3	3	3	3	6	6	3	3	3	5	3	3	6	3	5	3	3	3	3	1	3	3	6	3	2
6		1	1	1	1	2	2	2	2	2	2	2	1	1	2	2	2	2	1	_	2	2	2	2	2	1	1	1	2	2	2	2	2	1	2	2	2
##	[371]	1	Τ	1	1	3	3	3	3	3	3	3	1	1	3	3	3	3	Τ	Э	3	3	3	3	3	1	T	Τ	3	3	3	3	3	Τ	3	3	3
##	[408]	1	3	3	3	3	3	3	6	3	3	1	5	1	3	3	3	3	6	3	3	6	3	6	3	3	3	3	6	6	3	6	1	6	1	5	3
3	[100]	-	•	•	J	•	•	•	Ū	•	•	-	,	-	J	•	•	•	Ū	J	•	Ü	•	Ū	•	J	•	•	Ū	Ü	J	Ü	_	Ü	-	,	J
##	[445]	3	3	3	1	1	1	3	6	3	3	3	3	1	3	1	3	3	1	3	3	3	5	3	1	1	3	6	3	6	3	1	1	5	3	1	3
3																																					
##	[482]	1	1	3	6	3	3	5	3	1	3	1	6	6	3	6	3	3	1	3	5	6	3	1	6	6	3	3	5	6	3	3	3	3	3	6	3
3																																					
##	[519]	3	3	3	3	3	1	1	3	3	6	3	3	6	3	3	5	6	3	3	1	3	1	3	6	3	3	3	3	1	3	1	3	1	1	3	3
3																																					
##	[556]	1	6	3	6	1	3	3	3	1	3	3	3	3	1	5	3	3	3	3	3	1	6	3	3	3	3	1	1	1	3	1	1	1	1	5	3
3																							_														
##	[593]	3	1	3	3	1	1	6	2	4	3	1	1	3	3	3	3	6	1	1	6	1	5	3	1	3	6	3	3	3	3	3	3	3	3	6	1
1		_	_	_		_	_		_		_	_		_	2		_		_	_	_		_	2	_	_	_	_	_	_	-	-	_				
##	[630]	3	3	6	1	3	3	T	6	1	5	3	1	3	3	1	6	1	5	5	3	1	3	3	6	5	5	6	5	3	1	1	5	1	1	1	1
)   ##	[667]	1	2	2	2	2	6	5	1	6	2	2	6	1	1	1	1	2	6	6	1	5	2	2	6	6	2	6	2	2	2	2	2	2	2	1	1
4	[007]	_	5	5	J	_	U	J	_	U	5	5	U	_	_	_	_	5	U	U	_	J	5	J	U	U	J	U	J	5	J	J	J	5	5	1	1
##	[704]	6	5	1	2	4	5	6	3	5	2	3	5	3	1	3	2	2	6	5	3	5	6	2	1	6	1	1	3	4	3	6	1	6	6	3	6
3	[, • 1]	•	J	_	_	_	_	•	-	_	_	-	,	-	-	٥	_	_	9	٦	-	,	•	_	-	•	-	-	-	-	٥	•	-	•	•	٥	•
##	[741]	1	2	1	3	3	2	3	6	3	1	3	6	3	5	1	1	3	2	6	1	2	1	4	1	1	3	3	1	6	1	1	3	6	6	3	3
1	. ,																																				
##	[778]	3	3	1	5	1	3	1	1	1	1	6	3	3	2	4	3	3	1	1	3	4	2	6	6	3	5	4	4	6	4	3	1	3	3	3	1
6																																					
##	[815]	1	2	3	6	4	3	3	5	6	3	2	1	3	3	4	6	4	2	4	1	1	2	2	1	2	6	1	1	1	1	1	1	3	1	6	1
3																																					
##	[852]	3	1	6	1	3	3	1	1	2	1	3	2	3	5	1	6	3	6	3	3	3	4	1	1	1	2	6	2	3	2	6	1	3	1	6	3

```
3
3
3
## [2813] 3 3 3 2 6 2 1 1 3 3 3 3 3 4 3 3 5 1 6 1 1 3 3 3 3 1 3 3 3 3 3 3 3 3 3 6
1
## [2850] 1 6 5 1 5 3 3 1 6 1 1 3 3 3 3 1 3 5 3 1 3
##
## Within cluster sum of squares by cluster:
## [1] 269419.3 287639.5 161229.3 242192.3 302653.7 212654.4
   (between_SS / total_SS = 82.2 %)
##
##
## Available components:
##
## [1] "cluster"
                  "centers"
                               "totss"
                                            "withinss"
                                                         "tot.withinss"
## [6] "betweenss"
                  "size"
                               "iter"
                                            "ifault"
##
## Results for nstart = 50:
## tot.withinss: 1475788
##
## Results for nstart = 100:
## K-means clustering with 6 clusters of sizes 142, 1419, 740, 57, 137, 375
##
## Cluster means:
    new data. Avg Quantity new data. Total Quantity new data. total value
##
## 1
              30.058724
                                  79.690141
                                                     79.44380
## 2
               4.668089
                                   6.193798
                                                     13.00441
## 3
              10.979155
                                  22.214865
                                                     35.66677
## 4
              28.678617
                                 110.842105
                                                    228.15930
## 5
              10.923347
                                  51.708029
                                                    146.36314
## 6
               8.647979
                                  29.058667
                                                     75.06128
##
    new_data.Avg_UnitPrice new_data.Avg_ReturnRate
## 1
                1.690646
                                   0.1403995
## 2
                3.476512
                                   0.1454914
## 3
                3.096915
                                   0.1516628
##
  4
                4.891120
                                   0.1797195
## 5
               10.836077
                                   0.1549551
## 6
                5.445593
                                   0.1485293
##
## Clustering vector:
     [1] 5 3 2 2 2 2 3 6 3 5 2 3 2 4 3 2 2 2 6 6 5 2 3 2 6 2 6 2 2 5 6 4 2 2 2 2
##
2
##
    3
##
    [75] 3 2 3 6 5 3 3 3 2 2 2 5 3 2 5 2 3 6 2 2 6 2 2 3 6 6 6 4 2 2 3 6 2 3 2 2
6
   ##
2
##
   [149] 6 2 2 2 2 3 2 2 3 2 3 2 2 3 2 2 3 2 2 3 2 2 2 3 1 2 5 2 2 2 6 6 2 2 2 2 2 2 2 2 3
```

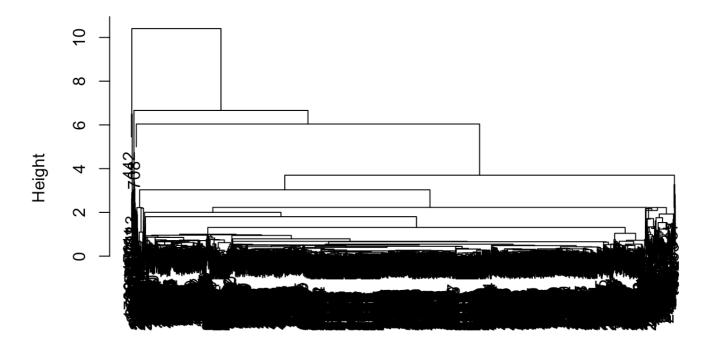
2																																					
##	[186]	2	6	2	2	3	2	2	1	3	2	2	3	2	2	2	2	2	3	2	3	2	3	3	2	2	2	2	2	2	2	2	2	3	2	5	2
##	[223]	2	2	2	2	2	2	2	3	2	2	3	3	3	3	6	2	6	2	5	2	2	2	2	2	6	2	2	2	2	2	2	3	2	2	2	2
##	[260]	2	2	2	2	6	2	2	2	2	6	3	2	2	3	2	3	2	2	2	2	2	4	2	5	3	2	2	2	2	2	2	2	2	2	2	2
##	[297]	3	2	3	2	2	3	3	2	6	3	2	2	2	6	2	2	2	2	3	2	2	2	3	3	2	5	2	2	2	2	2	2	6	2	2	6
##	[334]	2	2	2	3	2	2	2	2	3	2	2	2	2	2	2	6	6	2	2	2	5	2	2	6	2	5	2	2	2	2	3	2	2	6	2	1
##	[371]	3	3	3	3	2	2	2	2	2	2	2	3	3	2	2	2	2	3	5	2	2	2	2	2	3	3	3	2	2	2	2	2	3	2	2	2
##	[408]	3	2	2	2	2	2	2	6	2	2	3	5	3	2	2	2	2	6	2	2	6	2	6	2	2	2	2	6	6	2	6	3	6	3	5	2
2 ##	[445]	2	2	2	3	3	3	2	6	2	2	2	2	3	2	3	2	2	3	2	2	2	5	2	3	3	2	6	2	6	2	3	3	5	2	3	2
2 ##	[482]	3	3	2	6	2	2	5	2	3	2	3	6	6	2	6	2	2	3	2	5	6	2	3	6	6	2	2	5	6	2	2	2	2	2	6	2
2 ##	[519]	2	2	2	2	2	3	3	2	2	6	2	2	6	2	2	5	6	2	2	3	2	3	2	6	2	2	2	2	3	2	3	2	3	3	2	2
2 ##	[556]	3	6	2	6	3	2	2	2	3	2	2	2	2	3	5	2	2	2	2	2	3	6	2	2	2	2	3	3	3	2	3	3	3	3	5	2
2		_	_	_	_		_	_	_		_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_		_	
## 3	[593]																																				
##	[630]	2	2	6	3	2	2	3	6	3	5	2	3	2	2	3	6	3	5	5	2	3	2	2	6	5	5	6	5	2	3	3	5	3	3	3	3
##	[667]	3	2	2	2	1	6	5	3	6	2	2	6	3	3	3	3	2	6	6	3	5	2	2	6	6	2	6	2	2	2	2	2	2	2	3	3
##	[704]	6	5	3	1	4	5	6	2	5	1	2	5	2	3	2	1	1	6	5	2	5	6	1	3	6	3	3	2	4	2	6	3	6	6	2	6
##	[741]	3	1	3	2	2	1	2	6	2	3	2	6	2	5	3	3	2	1	6	3	1	3	4	3	3	2	2	3	6	3	3	2	6	6	2	2
##	[778]	2	2	3	5	3	2	3	3	3	3	6	2	2	1	4	2	2	3	3	2	4	1	6	6	2	5	4	4	6	4	2	3	2	2	2	3
##	[815]	3	1	2	6	4	2	2	5	6	2	1	3	2	2	4	6	4	1	4	3	3	1	1	3	1	6	3	3	3	3	3	3	2	3	6	3
##	[852]	2	3	6	3	2	2	3	3	1	3	2	1	2	5	3	6	2	6	2	2	2	4	3	3	3	1	6	1	2	1	6	3	2	3	6	2
##	[889]	1	2	3	3	3	2	6	2	2	2	2	2	6	3	3	2	2	6	2	3	1	2	5	2	6	5	2	2	2	2	2	6	2	1	2	2
##	[926]	6	3	5	3	3	3	6	2	2	3	2	3	6	3	2	3	2	6	3	6	3	4	3	5	2	1	2	5	4	3	2	4	3	2	3	3
2 ##	[963]	2	3	2	2	6	6	6	2	1	2	6	2	2	2	1	2	3	2	2	6	2	3	3	3	2	2	2	2	1	6	3	2	5	2	2	3
	[1000]	3	2	6	2	3	2	2	2	2	3	2	2	5	3	3	5	1	1	2	3	2	2	2	6	3	3	2	1	2	6	3	2	3	2	2	5
	[1037]	4	2	1	3	6	6	3	6	6	2	3	3	2	3	6	3	2	2	3	2	5	2	6	2	2	2	6	2	6	1	3	2	1	3	2	6
2 ##	[1074]	2	2	3	2	3	3	3	3	3	2	2	2	6	2	3	1	1	2	2	5	3	2	3	2	3	3	2	2	3	4	2	2	3	2	5	1

```
3
## [2036] 2 2 3 2 3 2 2 2 2 6 2 6 2 2 2 5 2 1 3 2 3 3 2 2 3 6 3 6 2 2 3 2 6 2 3 4
3
## [2073] 2 3 3 6 2 2 3 3 3 3 2 2 2 1 3 6 1 1 3 1 2 6 2 1 2 2 6 1 3 2 2 3 2 2 6 2
## [2110] 2 1 2 2 2 2 3 3 2 2 3 2 2 3 3 4 2 5 2 2 2 3 2 2 6 2 6 2 2 3 3 2 2 6 6
## [2147] 2 6 3 2 6 3 2 2 2 3 2 2 6 2 2 2 3 3 2 2 3 2 2 2 3 1 6 3 2 2 2 2 2 3 3
## [2184] 2 2 2 2 2 2 3 6 2 2 6 2 6 2 2 2 2 3 5 2 6 2 2 3 2 3 3 2 2 3 2 2 2 3 3
2
## [2221] 2 3 2 4 2 2 2 2 2 3 3 1 2 4 2 5 3 3 1 3 2 2 2 2 3 3 2 2 2 6 2 3 5 2 2 2
3
## [2258] 6 3 2 6 3 3 2 3 2 3 2 2 2 2 3 2 2 5 6 2 3 2 2 6 3 2 1 5 2 2 3 3 6 5 2 1
##
  [2295] 3 3 2 3 2 2 3 2 4 3 2 3 3 6 2 5 6 2 2 2 2 3 6 3 6 1 2 2 2 3 3 2 3 2 4 2
3
## [2332] 3 2 6 2 3 2 3 2 2 2 2 6 2 3 3 5 2 3 5 3 3 6 3 1 6 3 2 2 3 5 3 2 6 2 3 2
## [2369] 2 3 2 6 2 2 5 2 2 2 2 2 5 2 6 2 6 2 3 2 2 3 2 3 2 6 2 6 2 3 1 2 2 2 2
5
## [2443] 3 3 5 3 6 6 2 3 2 2 2 4 2 2 2 3 3 2 2 1 2 5 2 2 3 6 4 2 3 2 2 3 6 2 2 2
6
## [2480] 6 1 1 2 6 3 3 1 2 3 6 2 3 2 1 2 1 2 3 3 2 6 3 5 2 2 2 2 2 1 2 2 1 2 1 3
2
## [2517] 3 2 1 2 2 2 2 6 1 3 2 3 2 6 6 5 6 2 2 2 3 3 6 2 2 2 6 3 3 3 6 3 2 2 3 2
## [2554] 3 3 2 2 4 3 2 3 6 3 3 2 2 2 3 2 3 3 3 2 2 2 2 2 3 6 2 3 2 2 2 3 1 2
6
3
## [2628] 6 3 5 3 2 5 3 6 6 6 3 3 3 2 1 3 1 2 3 1 6 3 3 3 2 2 6 4 3 3 3 5 3 3 6 6
6
## [2665] 3 2 1 6 2 2 6 5 5 1 2 3 1 2 3 2 2 4 2 3 6 2 3 3 2 5 3 2 2 2 2 3 6 2 2 2
3
## [2702] 2 2 3 2 2 3 6 2 2 1 3 1 2 2 2 2 2 3 2 2 2 6 5 2 2 2 5 6 2 3 2 1 3 2 5 2
## [2739] 2 2 3 2 2 2 2 2 2 2 2 2 6 2 2 3 3 1 3 2 2 3 2 2 6 2 2 3 3 2 2 2 2 2 2 3 2
2
2
## [2813] 2 2 2 1 6 1 3 3 2 2 2 2 2 4 2 2 5 3 6 3 3 2 2 2 2 3 2 2 2 2 2 2 3 2 6
## [2850] 3 6 5 3 5 2 2 3 6 3 3 2 2 2 2 3 2 5 2 3 2
##
## Within cluster sum of squares by cluster:
  [1] 287639.5 161229.3 269419.3 242192.3 302653.7 212654.4
   (between_SS / total_SS = 82.2 %)
##
```

```
##
## Available components:
##
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"
## [6] "betweenss" "size" "iter" "ifault"
##
## Results for nstart = 100:
## tot.withinss: 1475788
```

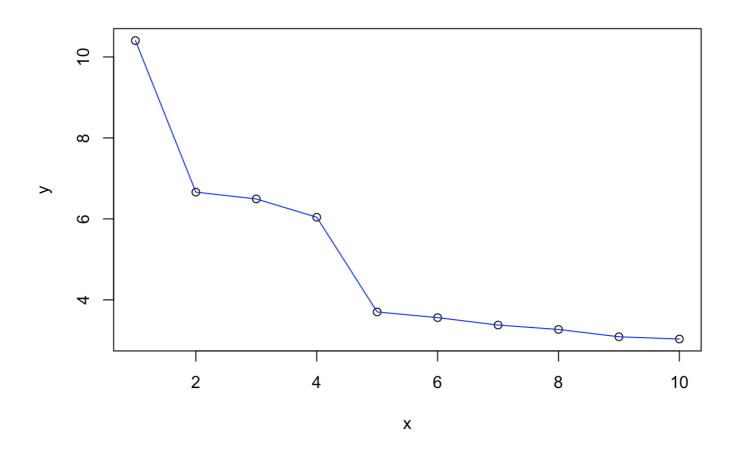
```
## for centroid method
plot(hclust2)
```

## **Cluster Dendrogram**



dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity, new\_data\$total\_value, new\_data\$t\(\daggar{\psi}\) new\_data\$Avg\_ReturnRate)))

```
y <- sort(hclust2$height, decreasing = TRUE)[1:10]
plot(x,y); lines(x,y, col= "blue")</pre>
```



```
results2 <- vector("list", length = 3)
for (i in 1:length(nstart_values)) {
    seg_kmeans2 <- kmeans(x = data.frame(new_data$Avg_Quantity, new_data$Total_Quant
    ity, new_data$total_value, new_data$Avg_UnitPrice, new_data$Avg_ReturnRate), cente
rs = 5, nstart = nstart_values[i])
    results2[[i]] <- seg_kmeans2
}

# Comparing results
for (i in 1:length(results2)) {
    cat("Results for nstart =", nstart_values[i], ":\n")
    print(results2[[i]])
    cat("\n")
    cat("Results for nstart =", nstart_values[i], ":\n")
    cat("Results for nstart =", nstart_values[i], ":\n")
    cat("tot.withinss:", results2[[i]]$tot.withinss, "\n\n")
}</pre>
```

```
## Results for nstart = 10 :
## K-means clustering with 5 clusters of sizes 752, 1706, 194, 146, 72
##
## Cluster means:
## new_data.Avg_Quantity new_data.Total_Quantity new_data.total_value
```

## 2
## 4
## 5
## new_data.Avg_UnitPrice new_data.Avg_ReturnRate ## 1
## 1
## 2
## 3
## 4
## 5
## ## Clustering vector: ## [1] 3 1 2 2 2 2 1 3 2 3 2 2 2 5 1 2 2 2 2 3 1 3 2 2 2 1 2 1 2 1 2 2 3 3 5 2 2 2 2  ## [38] 3 3 1 1 1 3 2 1 1 2 2 2 2 3 1 2 3 2 3 2
## Clustering vector: ## [1] 3 1 2 2 2 2 1 3 2 3 2 2 2 5 1 2 2 2 3 1 3 2 2 2 1 2 1 2 1 2 2 3 3 5 2 2 2 2 2 ## [38] 3 3 1 1 1 3 2 1 1 2 2 2 2 2 2 2 2 2 2 2
## [1] 3 1 2 2 2 2 1 3 2 3 2 2 2 5 1 2 2 2 2 3 1 3 2 2 2 2 1 2 1 2 2 2 3 3 5 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [38] 3 3 1 1 1 1 3 2 1 1 2 2 2 2 2 2 2 2 2 2
## [38] 3 3 1 1 1 3 2 1 1 2 2 2 2 2 2 2 2 2 2 2
1 ## [75] 1 2 1 3 3 1 2 1 2 2 2 3 1 2 3 2 1 1 2 2 2 3 2 2 2 1 1 1 5 2 2 1 1 2 2 2 2  1 ## [112] 2 1 2 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2
## [75] 1 2 1 3 3 1 2 1 2 2 2 3 1 2 3 2 1 1 2 2 3 2 2 1 1 1 5 2 2 1 1 2 2 2 2 1 1 2 2 2 2
## [112] 2 1 2 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2
## [149] 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2
## [149] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 ## [186] 2 3 2 2 1 2 2 4 1 2 2 1 1 2 2 2 2 2 2 2 2 2
## [223] 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 1 2 2 3 2 1 2 5 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [223] 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 1 2 2 3 2 1 2 5 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 ## [260] 2 2 2 2 3 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2
## [260] 2 2 2 2 3 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2
## [297] 1 2 1 2 2 1 1 2 1 1 2 2 1 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2
## [297] 1 2 1 2 2 1 1 2 1 1 2 1 1 2 2 2 1 2
2 ## [334] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
## [334] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 ## [371] 1 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 1 2
## [371] 1 2 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2
3 ## [408] 1 2 2 2 2 2 1 2 2 2 3 1 2 2 2 2 1 2 2 2 1 1 1 3 2  ## [445] 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 1 2 2 2 1 2 1 2 2 2 2 1 1 2 2 2 2 2 2 1 1 1 2 3 2 1 2 3 2 1 2  ## [482] 2 1 2 1 2 1 2 2 3 2 1 2 2 3 1 2 2 3 1 2 1 2
## [408] 1 2 2 2 2 2 2 1 2 2 2 3 1 2 2 2 2 1 2 2 2 3 1 2 2 2 2
2 ## [445] 2 2 2 1 2 1 2 1 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 3 2 1 1 2 2 3 2 1 2 2 3 2 1 2 2 3 2 1 2 1
2 ## [445] 2 2 2 1 2 1 2 1 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 1 2
2 ## [482] 2 1 2 1 2 2 3 2 1 2 2 3 1 2 1 2 2 3 1 2 1 2
2 ## [482] 2 1 2 1 2 2 3 2 1 2 2 3 1 2 1 2 2 3 1 2 1 2
2 ## [519] 2 2 2 2 1 1 2 2 1 2 2 1 2 2 3 1 2 2 1 2 2 3 2 2 2 2
2 ## [519] 2 2 2 2 1 1 2 2 1 2 2 1 2 2 3 1 2 2 1 2 2 3 2 2 2 2
## [519] 2 2 2 2 2 1 1 2 2 1 2 2 1 2 2 3 1 2 2 1 2 2 3 2 2 2 2
## [556] 2 3 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2
## [593] 2 1 2 2 1 2 1 4 5 2 2 1 2 2 2 2 1 1 1 1 1 3 2 2 2 1 2 2 2 2
2
## [630] 2 2 1 1 2 2 2 1 1 3 2 1 2 2 1 1 1 3 3 2 1 2 2 1 3 3 3 3
3
## [667] 1 2 2 2 4 1 3 1 1 2 2 1 2 1 1 1 2 3 1 1 5 2 2 3 3 2 3 2 2 2 2 2 2 2 1

```
1
## [2554] 2 2 2 2 5 2 2 3 2 1 2 2 2 1 2 1 1 2 2 2 2 2 1 1 2 2 2 2 2 3 2
3
1
## [2628] 1 1 3 1 2 3 2 1 1 1 1 1 2 2 4 2 4 2 2 4 1 1 1 2 2 2 3 5 1 2 1 3 2 2 1 3
1
## [2665] 1 2 4 1 2 2 1 3 3 4 2 2 4 2 2 2 2 5 2 2 3 2 1 2 2 3 1 2 2 2 2 1 1 2 2 2
## [2702] 2 2 1 2 2 2 1 2 2 4 2 4 2 2 2 2 2 1 2 2 2 1 3 2 2 2 3 3 2 1 2 4 1 2 3 2
2
1
## [2850] 1 3 3 1 3 2 2 1 1 1 1 2 2 2 2 2 2 3 2 1 2
##
## Within cluster sum of squares by cluster:
## [1] 421942.6 298199.4 320358.9 303098.1 352226.5
   (between SS / total SS = 79.6 %)
##
## Available components:
##
## [1] "cluster"
                  "centers"
                              "totss"
                                           "withinss"
                                                       "tot.withinss"
## [6] "betweenss"
                  "size"
                              "iter"
                                           "ifault"
##
## Results for nstart = 10 :
## tot.withinss: 1695826
## Results for nstart = 50:
## K-means clustering with 5 clusters of sizes 161, 739, 1697, 69, 204
##
## Cluster means:
##
    new data. Avg Quantity new data. Total Quantity new data. total value
## 1
              29.710687
                                 73.881988
                                                    71.67006
## 2
              9.614883
                                 25.093369
                                                    52.11141
## 3
                                                   15.30493
              5.581723
                                  8.017678
## 4
              26.261759
                                 99.115942
                                                   223.87638
## 5
              10.872743
                                 48.818627
                                                   126.78118
##
    new data.Avg UnitPrice new data.Avg ReturnRate
## 1
               1.662493
                                  0.1381651
## 2
               4.076984
                                  0.1496542
## 3
               3.367038
                                  0.1469543
## 4
               8.977743
                                  0.1957062
## 5
               8.120159
                                  0.1473854
##
## Clustering vector:
##
     [1] 5 2 3 3 3 3 2 5 3 5 3 3 3 4 2 3 3 3 5 2 5 3 3 3 2 3 2 3 3 5 5 4 3 3 3 3
```

3		_	_			_		_		_		_			_		_				_		_				_					_		_		_	
##	[38]	5	5	2	2	2	5	3	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	4	3	1	5	3	3	3	3	5	3	5	2	1	2
##	[75]	2	3	2	5	5	2	3	2	3	3	3	5	2	3	5	3	2	2	3	3	5	3	3	3	2	2	2	4	3	3	2	2	3	3	3	3
##	[112]	3	2	3	2	2	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
3 ##	[149]	2	3	3	3	3	3	3	3	3	3	2	3	3	3	2	3	3	3	3	5	3	5	3	3	3	2	2	3	3	3	3	3	3	3	3	3
3																																					
##	[186]	3	5	3	3	2	3	3	1	2	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	5	3
##	[223]	3	3	3	3	3	3	3	2	3	3	2	2	3	3	2	3	2	3	4	3	3	3	3	3	5	3	3	3	3	3	3	3	3	3	3	3
##	[260]	3	3	3	3	5	3	3	3	3	2	2	3	3	3	3	3	3	3	3	3	3	4	3	4	3	3	3	3	3	3	3	3	3	3	3	3
3 ##	[297]	2	3	2	3	3	2	2	3	2	2	3	3	3	2	3	3	3	3	2	3	3	3	3	2	3	4	3	3	3	3	3	3	2	3	3	2
3																																					
##	[334]	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	2	5	3	3	3	5	3	3	2	3	5	3	3	3	3	2	3	3	5	3	1
##	[371]	2	3	2	3	3	3	3	3	3	3	3	2	2	3	3	3	3	3	5	3	3	3	3	3	3	2	2	3	3	3	3	3	2	3	3	3
##	[408]	2	3	3	3	3	3	3	2	3	3	3	5	2	3	3	3	3	2	3	3	2	3	2	3	3	3	3	2	2	3	2	2	2	2	5	3
3 ##	[445]	3	3	3	3	3	2	3	2	3	3	3	3	2	3	2	3	3	2	3	3	3	5	3	2	2	3	5	3	2	3	2	3	5	3	2	3
3	. 4001	2	2	2	2	2	2	_	2	_	2	2	_	_	2	_	2	2	_	2	_	2	2	2	2	2	2	2	_	2	2	2	2	2	2	2	2
##	[482]	3	2	3	2	3	3	5	3	2	3	3	5	2	3	2	3	3	2	3	5	2	3	2	2	2	3	3	5	2	3	3	3	3	3	2	3
##	[519]	3	3	3	3	3	2	2	3	3	2	3	3	2	3	3	5	2	3	3	2	3	3	3	5	3	3	3	3	3	3	3	3	2	3	3	3
##	[556]	3	5	3	2	2	3	3	3	2	3	3	3	3	3	5	3	3	3	3	3	2	5	3	3	3	3	2	3	2	3	3	3	3	2	5	3
3 ##	[593]	3	2	3	3	2	3	2	1	4	3	3	2	3	3	3	3	2	2	2	2	2	5	3	3	3	2	3	3	3	3	3	3	3	3	2	2
3 ##	[630]	2	2	2	2	2	2	2	2	2	5	2	2	2	2	2	2	2	5	5	2	2	2	2	2	5	5	5	5	2	2	2	5	2	2	2	2
5																																					
##	[667]	2	3	3	3	1	2	5	2	2	3	3	2	3	2	2	2	3	5	2	2	4	3	3	5	5	3	5	3	3	3	3	3	3	3	3	2
##	[704]	2	5	2	1	4	5	2	3	4	1	3	5	3	2	3	1	5	1	5	3	5	2	1	3	2	1	2	3	4	3	1	2	2	2	3	2
3 ##	[741]	2	1	2	3	3	1	3	2	3	2	3	5	3	5	2	2	3	5	2	3	1	2	4	2	2	3	3	2	1	2	3	3	5	1	3	3
3 ##	[778]	3	3	2	5	2	3	3	2	3	2	2	3	3	1	4	3	3	3	3	3	4	5	2	2	3	5	4	4	2	4	3	2	3	3	3	2
2																																					
##	[815]	2	1	3	5	4	3	3	5	2	3	1	2	3	3	4	2	4	1	4	2	2	1	1	2	1	2	2	2	2	2	3	2	3	2	1	2
##	[852]	3	3	2	3	3	3	2	2	1	2	3	1	3	5	2	1	3	2	3	3	3	4	2	2	2	1	5	1	3	1	1	2	3	3	2	3
##	[889]	1	3	2	2	2	3	5	3	3	3	3	3	2	3	2	3	3	2	3	3	1	3	5	3	2	5	3	3	3	3	3	2	3	1	3	3
2 ##	[926]	2	3	5	3	2	3	1	3	3	2	3	3	2	2	3	2	3	2	3	2	2	4	2	5	3	1	3	5	4	2	3	4	2	3	3	2
	•																																				

```
3
## [2850] 2 5 5 2 5 3 3 2 2 2 2 3 3 3 3 3 5 3 2 3
##
## Within cluster sum of squares by cluster:
## [1] 310536.7 381525.4 303007.4 341124.0 357693.7
  (between SS / total SS = 79.6 %)
##
## Available components:
##
                         "totss"
## [1] "cluster"
                                    "withinss"
                                              "tot.withinss"
               "centers"
                                    "ifault"
## [6] "betweenss"
               "size"
                         "iter"
##
## Results for nstart = 50:
## tot.withinss: 1693887
##
## Results for nstart = 100:
## K-means clustering with 5 clusters of sizes 739, 161, 1697, 204, 69
##
## Cluster means:
##
   new data. Avg Quantity new data. Total Quantity new data. total value
## 1
            9.614883
                            25.093369
                                           52.11141
## 2
           29.710687
                            73.881988
                                           71.67006
            5.581723
## 3
                             8.017678
                                           15.30493
## 4
           10.872743
                            48.818627
                                           126.78118
## 5
                                           223.87638
           26.261759
                            99.115942
##
   new_data.Avg_UnitPrice new_data.Avg_ReturnRate
## 1
             4.076984
                             0.1496542
## 2
             1.662493
                             0.1381651
## 3
             3.367038
                             0.1469543
## 4
             8.120159
                             0.1473854
## 5
             8.977743
                             0.1957062
##
## Clustering vector:
    [1] 4 1 3 3 3 3 1 4 3 4 3 3 3 5 1 3 3 3 4 1 4 3 3 3 1 3 1 3 3 4 4 5 3 3 3 3
##
3
##
   1
##
   [75] 1 3 1 4 4 1 3 1 3 3 3 4 1 3 4 3 1 1 3 3 4 3 3 1 1 1 5 3 3 1 1 3 3 3 3
1
##
  3
##
  3
  ##
3
  ##
3
##
```

```
1
1
## [2221] 3 1 3 5 3 3 3 3 3 1 1 2 3 5 3 4 1 3 2 1 3 3 3 3 1 3 3 3 1 3 1 4 3 3 3
3
## [2295] 1 3 3 1 3 3 3 3 5 1 3 1 1 1 3 4 1 3 3 3 3 1 4 3 4 2 3 3 3 1 1 3 3 3 5 3
1
## [2332] 3 3 1 3 3 3 1 3 3 3 1 3 3 3 1 3 2 3 4 3 1 4 1 3 1 3 2 1 3 3 3 1 4 1 3 1 3 1 3
## [2443] 1 1 4 1 4 1 3 3 3 3 3 5 3 3 3 1 1 3 3 2 3 4 3 3 3 1 5 3 1 3 3 3 4 3 3 3
1
## [2480] 4 2 2 3 2 3 1 4 3 1 1 3 3 3 2 3 4 3 1 1 3 1 3 4 3 3 3 3 3 2 3 3 2 3 2 1
3
## [2517] 3 3 2 3 3 3 3 1 2 1 3 3 3 1 1 4 4 3 3 3 1 1 1 3 3 3 1 1 3 1 1 1 3 3 1 3
1
1
## [2628] 1 2 4 1 3 4 3 1 1 1 1 1 3 3 2 3 2 3 3 2 1 1 1 3 3 3 4 5 1 3 1 4 3 3 1 4
1
## [2665] 1 3 2 1 3 3 1 4 4 2 3 3 2 3 3 3 3 5 3 3 4 3 1 3 3 4 1 3 3 3 3 1 1 3 3 3
3
1
## [2850] 1 4 4 1 4 3 3 1 1 1 1 3 3 3 3 3 3 4 3 1 3
##
## Within cluster sum of squares by cluster:
## [1] 381525.4 310536.7 303007.4 357693.7 341124.0
##
  (between SS / total SS = 79.6 %)
##
## Available components:
##
## [1] "cluster"
            "centers"
                    "totss"
                            "withinss"
                                    "tot.withinss"
## [6] "betweenss"
            "size"
                    "iter"
                            "ifault"
##
```

```
## Results for nstart = 100 :
## tot.withinss: 1693887
```

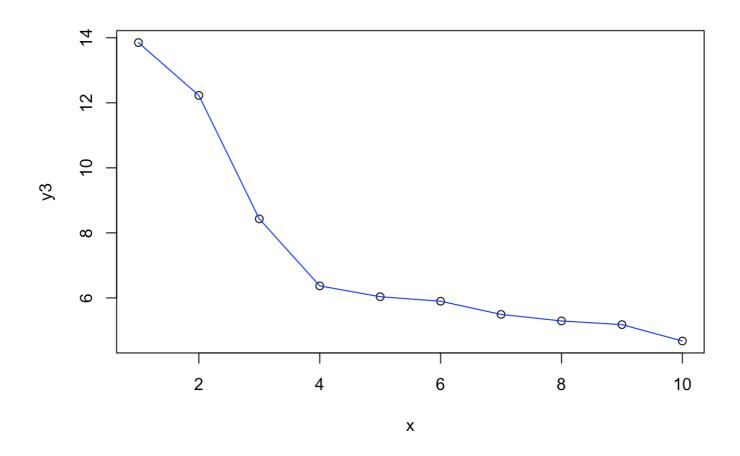
```
## for average method
plot(hclust3)
```

## **Cluster Dendrogram**



dist(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_Quantity, new\_data\$total\_value, new\_data\$t4(\*g"al/heitagie\*), new\_data\$Avg\_ReturnRate)))

```
y3 <- sort(hclust3$height, decreasing = TRUE)[1:10]
plot(x,y3); lines(x,y3, col= "blue")</pre>
```



```
results3 <- vector("list", length = 3)
for (i in 1:length(nstart_values)) {
   seg_kmeans3 <- kmeans(x = data.frame(new_data$Avg_Quantity, new_data$Total_Quant
   ity, new_data$total_value, new_data$Avg_UnitPrice, new_data$Avg_ReturnRate), cente
   rs = 4, nstart = nstart_values[i])
   results3[[i]] <- seg_kmeans3
}
seg_kmeans3$tot.withinss</pre>
```

#### ## [1] 1933954

```
# Comparing results
for (i in 1:length(results3)) {
  cat("Results for nstart =", nstart_values[i], ":\n")
  print(results3[[i]])
  cat("\n")
  cat("Results for nstart =", nstart_values[i], ":\n")
  cat("tot.withinss:", results3[[i]]$tot.withinss, "\n\n")
}
```

```
## Results for nstart = 10 :
## K-means clustering with 4 clusters of sizes 74, 761, 1746, 289
##
## Cluster means:
##
  new data. Avg Quantity new data. Total Quantity new data. total value
      25.056588
                95.743243
                        221.05297
## 1
## 2
      11.836407
                29.561104
                         52.97485
## 3
       5.532533
                8.128866
                         15.92806
## 4
      16.849857
                59.062284
                        113.66893
##
  new_data.Avg_UnitPrice new_data.Avg_ReturnRate
## 1
       9.195324
                0.1917935
## 2
       3.959407
                0.1505239
## 3
       3.384698
                0.1457553
## 4
       5.768671
                0.1472687
##
## Clustering vector:
##
  [1] 4 2 3 3 3 3 3 2 4 3 4 3 3 3 1 2 3 3 3 4 2 4 3 3 3 2 3 2 3 3 4 4 1 3 3 3 3
3
##
  2
##
  [75] 3 3 2 4 4 2 3 2 3 3 3 4 2 3 4 3 2 2 3 3 4 3 3 2 2 2 1 3 3 2 2 3 3 3 3
2
 ##
3
 ##
3
 ##
3
##
 3
##
 3
##
 3
 ##
2
##
 2
##
 3
 ##
3
##
 3
##
 3
##
 3
 ##
3
```

##	[630]	3	3	2	2	3	3	3	2	2	4	3	2	3	3	2	2	3	4	4	3	2	3	3	2	4	4	4	4	3	3	2	4	2	2	3	2
##	[667]	2	3	3	3	4	2	4	2	2	3	3	2	3	2	2	2	3	4	2	2	1	3	3	4	4	3	4	3	3	3	3	3	3	3	3	2
##	[704]	2	4	2	2	1	4	2	3	1	4	3	4	3	2	3	4	4	2	4	3	4	2	4	3	2	2	2	3	1	3	2	2	2	2	3	2
##	[741]	2	4	2	3	3	2	3	2	3	2	3	4	3	4	2	2	3	4	2	3	2	3	1	2	2	3	3	2	2	2	3	3	4	4	3	3
##	[778]	3	3	2	4	2	3	3	2	3	2	2	3	3	4	1	3	3	3	3	3	1	4	2	2	3	4	1	1	2	1	3	2	3	3	3	2
##	[815]	3	4	3	4	1	3	3	4	2	3	4	2	3	3	1	2	1	4	1	2	2	4	4	2	2	2	2	2	2	2	3	2	3	2	2	2
##	[852]	3	3	2	3	3	3	2	2	4	2	3	4	3	4	2	2	3	2	3	3	3	1	2	2	2	4	4	4	3	4	2	2	3	3	2	3
##	[889]	4	3	2	2	2	3	4	3	3	3	3	3	2	3	2	3	3	2	3	3	2	3	4	3	2	4	3	3	3	3	3	2	3	4	3	3
##	[926]	2	3	4	3	2	3	2	3	3	3	3	3	2	2	3	2	3	2	3	2	2	1	2	4	3	4	3	4	1	2	3	1	2	3	3	2
##	[963]	3	2	3	3	4	2	2	3	4	3	2	3	3	3	4	3	3	3	3	2	3	2	2	2	3	3	3	3	4	2	2	3	4	3	3	2
##	[1000]	2	3	2	3	3	3	3	3	3	3	3	3	4	2	2	4	2	2	3	3	3	3	3	2	2	3	3	4	3	2	3	3	3	3	3	4
##	[1037]	1	3	4	2	2	4	2	2	2	3	3	2	3	3	2	2	3	3	2	3	1	3	2	3	3	3	2	3	2	2	2	3	2	3	3	2
##	[1074]	3	3	3	3	2	3	2	2	2	3	3	3	2	3	2	4	2	3	3	4	2	3	3	3	2	2	3	3	3	1	3	3	3	3	4	4
##	[1111]	2	3	3	3	2	2	3	3	3	2	3	3	3	2	3	3	2	2	3	3	3	3	2	3	2	3	3	3	2	3	2	3	3	3	3	2
## 2	[1148]	3	3	3	2	4	2	3	3	3	3	2	3	3	2	3	3	3	2	3	3	2	3	3	3	3	4	3	2	3	3	2	4	3	3	3	2
## 2	[1185]	3	3	3	2	2	2	3	3	4	3	2	2	2	3	3	3	4	2	3	3	3	2	2	3	3	3	3	3	2	3	1	3	2	2	3	3
##	[1222]	3	2	2	3	3	3	2	3	3	3	3	3	2	3	2	4	3	3	2	3	2	3	4	3	2	3	2	2	4	3	2	3	3	3	2	2
##	[1259]	3	3	2	3	2	2	3	3	3	3	3	3	3	2	3	3	3	2	3	3	2	3	3	3	3	3	3	4	4	3	3	2	3	3	2	2
##	[1296]	2	3	3	3	3	3	2	3	3	3	3	3	2	3	3	2	2	2	4	2	4	3	3	3	3	3	3	3	2	4	2	3	3	3	3	2
##	[1333]	2	3	2	3	2	3	3	2	4	2	3	3	3	2	3	4	2	3	3	3	2	3	2	2	2	4	1	2	2	3	2	2	3	3	2	3
##	[1370]	3	2	3	3	4	3	2	2	3	1	2	3	1	3	3	2	3	2	3	2	2	3	4	3	2	2	3	3	3	4	2	3	2	3	3	3
##	[1407]	2	3	2	3	2	2	2	3	3	3	3	3	3	2	4	3	3	3	2	2	3	2	3	4	2	3	4	2	3	2	2	4	2	3	4	1
##	[1444]	3	3	2	3	2	3	2	3	3	3	3	2	4	3	3	3	3	3	3	2	3	3	1	3	3	3	3	4	2	3	3	3	3	4	3	3
##	[1481]	3	2	2	2	4	3	2	3	3	2	2	3	3	2	3	3	3	2	3	3	3	2	4	2	4	4	3	3	3	3	3	3	3	2	3	3
##	[1518]	3	2	3	3	3	3	3	3	2	2	2	3	3	3	3	3	3	2	3	3	4	3	3	3	3	3	2	4	3	3	3	3	3	3	2	3

## [1555]	2 2	3 3	2	2	3	3	3	3	3	2	3	2	1	4	3	3	3	2	2	2	3	3	3	4	3	2	3	3	2	3	4	3	4	3
## [1592] 2	2 3	3 2	3	3	3	4	3	3	3	3	3	1	3	3	4	3	4	3	2	3	2	2	3	2	2	3	3	3	1	3	3	3	2	3
## [1629]	4 3	2 2	4	3	4	2	3	3	3	1	3	4	3	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	3	3	3
3 ## [1666]	3 3	2 3	1	3	3	2	2	3	3	3	3	4	4	3	1	3	3	2	3	4	3	3	3	3	4	2	3	4	3	2	2	3	3	3
2 ## [1703]	2 3	2 2	3	3	2	3	3	3	3	1	3	4	2	2	1	4	3	2	2	3	3	3	3	2	2	1	4	4	3	2	2	3	3	2
3 ## [1740]	4 4	2 4	2	2	2	3	4	3	4	3	3	3	2	4	3	3	3	3	3	2	4	2	2	2	3	3	3	4	2	3	3	3	3	3
3 ## [1777]	3 3	3 2	1	4	3	2	3	2	4	4	2	3	3	2	4	3	3	3	3	2	3	3	2	2	3	2	3	3	3	3	3	3	3	3
3 ## [1814]	2 3	3 3	3	4	3	3	3	3	3	3	2	2	4	3	3	3	3	4	2	3	3	2	3	3	3	2	2	3	3	4	3	4	2	3
3 ## [1851]																																		
2																																		
## [1888] 3																																		
## [1925] 3	2 3	4 3	3	2	3	2	3	2	3	3	3	3	4	3	3	4	3	2	2	3	2	4	4	3	3	2	3	3	3	3	2	2	2	2
## [1962] 3	3 3	3 3	3	2	3	3	3	1	3	2	3	3	2	2	3	4	3	2	1	3	3	3	3	2	3	3	3	3	3	2	3	3	3	2
## [1999] 3	3 1	1 3	3	4	3	3	4	2	2	2	2	3	3	3	4	2	3	3	2	3	2	2	2	2	3	2	1	3	3	3	3	2	2	3
## [2036] 2	3 3	3 3	3	3	3	3	3	2	3	4	3	3	3	4	3	4	3	3	2	3	3	3	2	4	3	2	3	3	3	3	2	3	3	1
## [2073]	3 2	2 2	3	3	2	2	3	3	3	3	3	4	2	2	4	4	3	4	3	2	3	4	3	3	2	2	3	3	3	2	3	3	2	3
## [2110]	3 4	3 3	3	3	2	2	3	3	2	3	3	2	2	1	3	4	3	3	3	3	2	3	3	2	3	2	3	3	3	3	3	3	2	4
2 ## [2147]	3 2	2 3	2	2	3	3	3	2	3	3	2	3	3	3	3	3	3	3	2	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3
2 ## [2184]	3 3	3 3	3	3	2	2	3	3	2	3	4	3	3	3	3	3	3	4	3	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3 ## [2221]	3 2	3 1	3	3	3	3	3	3	2	2	2	3	1	3	4	2	3	4	2	3	3	3	3	2	3	3	3	2	3	2	4	3	3	3
3 ## [2258]	2 2	3 2	2	3	3	3	3	3	3	3	3	3	2	3	3	4	2	3	2	3	3	2	3	3	4	4	3	3	2	2	2	1	3	4
2 ## [2295]																																		
2 ## [2332]																																		
3																																		
## [2369] 4																																		
## [2406] 3	1 3	3 2	3	3	2	1	2	2	3	2	3	3	3	3	3	3	3	2	3	2	3	2	3	3	2	3	2	3	3	4	3	3	3	3
## [2443] 2	2 2	4 2	4	2	3	3	3	3	3	1	3	3	3	2	2	3	3	2	3	4	3	3	3	2	1	3	2	3	3	2	4	3	3	3

```
## [2665] 2 3 2 2 3 3 2 1 4 2 3 3 2 3 3 3 1 3 3 4 3 2 3 3 4 2 3 3 3 3 2 2 3 3 3
2
## [2739] 3 3 2 3 3 3 3 3 3 3 3 3 3 2 3 3 2 4 2 3 3 2 3 3 2 3 3 2 2 3 3 3 2 2 3 3 3 2 3
3
## [2850] 2 4 4 2 4 3 3 2 2 2 2 3 3 3 3 3 3 4 3 2 3
##
## Within cluster sum of squares by cluster:
## [1] 368717.0 562686.5 320013.5 682536.7
##
  (between_SS / total_SS = 76.7 %)
##
## Available components:
##
## [1] "cluster"
              "centers"
                       "totss"
                                 "withinss"
                                          "tot.withinss"
## [6] "betweenss"
              "size"
                       "iter"
                                "ifault"
## Results for nstart = 10:
## tot.withinss: 1933954
##
## Results for nstart = 50 :
## K-means clustering with 4 clusters of sizes 74, 1746, 761, 289
##
## Cluster means:
   new data. Avg Quantity new data. Total Quantity new data. total value
##
## 1
          25.056588
                         95.743243
                                       221.05297
## 2
           5.532533
                          8.128866
                                       15.92806
## 3
          11.836407
                         29.561104
                                       52.97485
## 4
          16.849857
                         59.062284
                                       113.66893
##
   new data.Avg UnitPrice new data.Avg ReturnRate
## 1
            9.195324
                          0.1917935
## 2
            3.384698
                          0.1457553
## 3
            3.959407
                          0.1505239
## 4
            5.768671
                          0.1472687
##
## Clustering vector:
```

##	[1]	4	3	2	2	2	2	3	4	2	4	2	2	2	1	3	2	2	2	4	3	4	2	2	2	3	2	3	2	2	4	4	1	2	2	2	2
##	[38]	4	4	3	3	3	4	2	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	3	4	2	2	2	2	4	2	4	3	3	3
##	[75]	2	2	3	4	4	3	2	3	2	2	2	4	3	2	4	2	3	3	2	2	4	2	2	2	3	3	3	1	2	2	3	3	2	2	2	2
##	[112]	2	3	2	3	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3
##	[149]	3	2	2	2	2	2	2	2	2	2	3	2	2	2	3	2	2	2	2	4	2	4	2	2	2	3	3	2	2	2	2	2	2	2	2	2
##	[186]	2	4	2	2	2	2	2	4	3	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	4	2
##	[223]	2	2	2	2	2	2	2	3	2	2	3	3	2	2	3	2	3	2	1	2	2	2	2	2	4	2	2	2	2	2	2	2	2	2	2	2
##	[260]	2	2	2	2	4	2	2	2	2	3	3	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2
##	[297]	3	2	3	2	2	3	3	2	3	2	2	2	2	3	2	2	2	2	2	2	2	2	2	3	2	1	2	2	2	2	2	2	3	2	2	3
##	[334]	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	3	4	2	2	2	4	2	2	3	2	4	2	2	2	2	3	2	2	3	2	4
##	[371]	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	4	2	2	2	2	2	2	3	3	2	2	2	2	2	2	2	2	2
##	[408]	3	2	2	2	2	2	2	3	2	2	2	4	2	2	2	2	2	3	2	2	3	2	3	2	2	2	2	3	3	2	3	3	3	3	4	2
##	[445]	2	2	2	3	2	3	2	3	2	2	2	2	3	2	3	2	2	3	2	2	2	4	2	3	3	2	4	2	3	2	3	2	4	2	3	2
##	[482]	2	2	2	3	2	2	1	2	3	2	2	4	3	2	3	2	2	3	2	4	3	2	3	3	3	2	2	4	3	2	2	2	2	2	3	2
## 2	[519]	2	2	2	2	2	3	2	2	2	3	2	2	3	2	2	4	3	2	2	3	2	2	2	4	2	2	2	2	2	2	2	2	3	2	2	2
##	[556]	2	4	2	3	3	2	2	2	2	2	2	2	2	2	4	2	2	2	2	2	3	4	2	2	2	2	3	2	3	2	2	2	2	3	4	2
## 2	[593]	2	3	2	2	3	2	3	3	1	2	2	3	2	2	2	2	3	3	3	3	3	4	2	2	2	3	2	2	2	2	2	2	2	2	3	3
## 4	[630]	2	2	3	3	2	2	2	3	3	4	2	3	2	2	3	3	2	4	4	2	3	2	2	3	4	4	4	4	2	2	3	4	3	3	2	3
##	[667]	3	2	2	2	4	3	4	3	3	2	2	3	2	3	3	3	2	4	3	3	1	2	2	4	4	2	4	2	2	2	2	2	2	2	2	3
## 2	[704]	3	4	3	3	1	4	3	2	1	4	2	4	2	3	2	4	4	3	4	2	4	3	4	2	3	3	3	2	1	2	3	3	3	3	2	3
## 2	[741]	3	4	3	2	2	3	2	3	2	3	2	4	2	4	3	3	2	4	3	2	3	2	1	3	3	2	2	3	3	3	2	2	4	4	2	2
##	[778]	2	2	3	4	3	2	2	3	2	3	3	2	2	4	1	2	2	2	2	2	1	4	3	3	2	4	1	1	3	1	2	3	2	2	2	3
##	[815]	2	4	2	4	1	2	2	4	3	2	4	3	2	2	1	3	1	4	1	3	3	4	4	3	3	3	3	3	3	3	2	3	2	3	3	3
##	[852]																																				
##	[889]	4	2	3	3	3	2	4	2	2	2	2	2	3	2	3	2	2	3	2	2	3	2	4	2	3	4	2	2	2	2	2	3	2	4	2	2

## [ 2	926]	3	2	4	2	3	2	3	2	2	2	2	2	3	3	2	3	2	3	2	3	3	1	3	4	2	4	2	4	1	3	2	1	3	2	2	3
	963]	2	3	2	2	4	3	3	2	4	2	3	2	2	2	4	2	2	2	2	3	2	3	3	3	2	2	2	2	4	3	3	2	4	2	2	3
## [1 3	.000]	3	2	3	2	2	2	2	2	2	2	2	2	4	3	3	4	3	3	2	2	2	2	2	3	3	2	2	4	2	3	2	2	2	2	2	4
## [1	.037]	1	2	4	3	3	4	3	3	3	2	2	3	2	2	3	3	2	2	3	2	1	2	3	2	2	2	3	2	3	3	3	2	3	2	2	3
## [1 2	.074]	2	2	2	2	3	2	3	3	3	2	2	2	3	2	3	4	3	2	2	4	3	2	2	2	3	3	2	2	2	1	2	2	2	2	4	4
## [1	111]	3	2	2	2	3	3	2	2	2	3	2	2	2	3	2	2	3	3	2	2	2	2	3	2	3	2	2	2	3	2	3	2	2	2	2	3
## [1	148]	2	2	2	3	4	3	2	2	2	2	3	2	2	3	2	2	2	3	2	2	3	2	2	2	2	4	2	3	2	2	3	4	2	2	2	3
## [1 3	.185]	2	2	2	3	3	3	2	2	4	2	3	3	3	2	2	2	4	3	2	2	2	3	3	2	2	2	2	2	3	2	1	2	3	3	2	2
## [1 3	.222]	2	3	3	2	2	2	3	2	2	2	2	2	3	2	3	4	2	2	3	2	3	2	4	2	3	2	3	3	4	2	3	2	2	2	3	3
## [1 1	.259]	2	2	3	2	3	3	2	2	2	2	2	2	2	3	2	2	2	3	2	2	3	2	2	2	2	2	2	4	4	2	2	3	2	2	3	3
## [1	.296]	3	2	2	2	2	2	3	2	2	2	2	2	3	2	2	3	3	3	4	3	4	2	2	2	2	2	2	2	3	4	3	2	2	2	2	3
## [1 2	.333]	3	2	3	2	3	2	2	3	4	3	2	2	2	3	2	4	3	2	2	2	3	2	3	3	3	4	1	3	3	2	3	3	2	2	3	2
## [1	.370]	2	3	2	2	4	2	3	3	2	1	3	2	1	2	2	3	2	3	2	3	3	2	4	2	3	3	2	2	2	4	3	2	3	2	2	2
## [1 2	407]	3	2	3	2	3	3	3	2	2	2	2	2	2	3	4	2	2	2	3	3	2	3	2	4	3	2	4	3	2	3	3	4	3	2	4	1
## [1 2	444]	2	2	3	2	3	2	3	2	2	2	2	3	4	2	2	2	2	2	2	3	2	2	1	2	2	2	2	4	3	2	2	2	2	4	2	2
## [1 2	.481]	2	3	3	3	4	2	3	2	2	3	3	2	2	3	2	2	2	3	2	2	2	3	4	3	4	4	2	2	2	2	2	2	2	3	2	2
## [1	.518]	2	3	2	2	2	2	2	2	3	3	3	2	2	2	2	2	2	3	2	2	4	2	2	2	2	2	3	4	2	2	2	2	2	2	3	2
## [1 2	.555]	3	3	2	2	3	3	2	2	2	2	2	3	2	3	1	4	2	2	2	3	3	3	2	2	2	4	2	3	2	2	3	2	4	2	4	2
## [1 3	.592]	3	2	2	3	2	2	2	4	2	2	2	2	2	1	2	2	4	2	4	2	3	2	3	3	2	3	3	2	2	2	1	2	2	2	3	2
## [1 2	.629]	4	2	3	3	4	2	4	3	2	2	2	1	2	4	2	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	2	2	2
## [1	.666]	2	2	3	2	1	2	2	3	3	2	2	2	2	4	4	2	1	2	2	3	2	4	2	2	2	2	4	3	2	4	2	3	3	2	2	2
## [1	.703]	3	2	3	3	2	2	3	2	2	2	2	1	2	4	3	3	1	4	2	3	3	2	2	2	2	3	3	1	4	4	2	3	3	2	2	3
## [1 2	.740]	4	4	3	4	3	3	3	2	4	2	4	2	2	2	3	4	2	2	2	2	2	3	4	3	3	3	2	2	2	4	3	2	2	2	2	2
## [1 2	.777]	2	2	2	3	1	4	2	3	2	3	4	4	3	2	2	3	4	2	2	2	2	3	2	2	3	3	2	3	2	2	2	2	2	2	2	2
## [1 2	.814]	3	2	2	2	2	4	2	2	2	2	2	2	3	3	4	2	2	2	2	4	3	2	2	3	2	2	2	3	3	2	2	4	2	4	3	2

## [1851] 3	2 2 2	2 3	3	2	2	2	2	2	2	3	2	4	3	3	2	2	2	3	3	2	4	3	2	2	2	2	2	1	3	2	4	2	2	3
## [1888] 2	4 4	2 1	2	2	2	2	2	2	2	3	3	3	3	2	4	2	2	2	2	2	3	2	2	2	3	2	3	3	2	2	4	2	3	2
## [1925]	3 2	4 2	2	3	2	3	2	3	2	2	2	2	4	2	2	4	2	3	3	2	3	4	4	2	2	3	2	2	2	2	3	3	3	3
2 ## [1962]	2 2 2	2 2	2	3	2	2	2	1	2	3	2	2	3	3	2	4	2	3	1	2	2	2	2	3	2	2	2	2	2	3	2	2	2	3
2 ## [1999]	2 1	1 2	2	4	2	2	4	3	3	3	3	2	2	2	4	3	2	2	3	2	3	3	3	3	2	3	1	2	2	2	2	3	3	2
2 ## [2036]	2 2 :	2 2	2	2	2	2	2	3	2	4	2	2	2	4	2	4	2	2	3	2	2	2	3	4	2	3	2	2	2	2	3	2	2	1
3 ## [2073]																																		
2																																		
## [2110] 3	2 4 2	22	2	2	3	3	2	2	3	2	2	3	3	1	2	4	2	2	2	2	3	2	2	3	2	3	2	2	2	2	2	2	3	4
## [2147] 3	2 3	3 2	3	3	2	2	2	3	2	2	3	2	2	2	2	2	2	2	3	2	2	2	2	3	3	3	3	2	2	2	2	2	2	2
## [2184] 2	2 2 2	2 2	2	2	3	3	2	2	3	2	4	2	2	2	2	2	2	4	2	4	2	2	2	2	2	2	2	2	2	2	2	2	2	2
## [2221] 2	2 3 2	2 1	2	2	2	2	2	2	3	3	3	2	1	2	4	3	2	4	3	2	2	2	2	3	2	2	2	3	2	3	4	2	2	2
## [2258]	3 3 2	2 3	3	2	2	2	2	2	2	2	2	2	3	2	2	4	3	2	3	2	2	3	2	2	4	4	2	2	3	3	3	1	2	4
3 ## [2295]	3 2 2	2 3	2	2	2	2	1	3	2	3	3	3	2	4	3	2	2	2	2	2	4	2	4	4	2	2	2	3	3	2	2	2	1	2
3 ## [2332]	2 2 3	3 2	2	2	3	2	2	2	2	3	2	3	2	4	2	3	4	3	2	3	2	3	3	2	2	2	3	4	3	2	3	2	3	2
2 ## [2369]	2 2 :	2 4	2	2	4	2	2	2	2	2	2	4	2	3	2	3	2	2	2	2	3	2	3	2	3	2	3	2	2	4	2	2	2	2
## [2406]																																		
2																																		
## [2443] 3	3 3	4 3	4	3	2	2	2	2	2	1	2	2	2	3	3	2	2	3	2	4	2	2	2	3	1	2	3	2	2	3	4	2	2	2
## [2480] 2	4 4	3 2	4	2	2	4	2	3	3	2	2	2	4	2	4	2	3	2	2	3	2	4	2	2	2	2	2	4	2	2	3	2	4	3
## [2517]	2 2	4 2	2	2	2	3	4	3	2	2	2	3	3	4	4	2	2	2	2	3	3	2	2	2	3	2	2	3	3	3	2	2	3	2
## [2554]	2 2 2	2 2	1	2	2	2	4	2	3	2	2	2	3	2	3	3	2	2	2	2	2	2	3	3	2	2	2	2	2	2	2	2	4	2
## [2591]	2 2 2	2 4	2	2	3	2	3	2	2	2	3	2	2	2	3	2	2	2	2	2	2	2	3	1	2	2	3	2	3	3	4	2	2	3
2 ## [2628]	3 3	4 3	2	4	2	3	3	3	3	3	2	2	4	2	4	2	2	4	3	3	3	2	2	2	4	1	3	2	3	4	2	2	3	4
3 ## [2665]	3 2	3 3	2	2	3	1	4	3	2	2	3	2	2	2	2	1	2	2	4	2	3	2	2	4	3	2	2	2	2	3	3	2	2	2
3 ## [2702]	2 2 3	2 2	2	2	3	2	2	4	2	4	2	2	2	2	2	3	2	2	2	3	4	2	2	2	4	4	2	3	2	3	3	2	4	2
## [2739]																																		
## [2/39] 2	2 2 .	J	۷	۷	۷	۷	۷	۷	۷	2	3	۷	2	۷	3	4	J	2	۷	3	۷	2	3	۷	۷	J	3	۷	۷	۷	۷	۷	3	۷

```
## [2850] 3 4 4 3 4 2 2 3 3 3 3 2 2 2 2 2 2 4 2 3 2
##
## Within cluster sum of squares by cluster:
## [1] 368717.0 320013.5 562686.5 682536.7
  (between SS / total SS = 76.7 %)
##
## Available components:
##
                        "totss"
                                 "withinss"
                                           "tot.withinss"
## [1] "cluster"
              "centers"
              "size"
                        "iter"
                                 "ifault"
## [6] "betweenss"
##
## Results for nstart = 50:
## tot.withinss: 1933954
##
## Results for nstart = 100:
## K-means clustering with 4 clusters of sizes 74, 289, 1746, 761
##
## Cluster means:
   new data. Avg Quantity new data. Total Quantity new data. total value
## 1
           25.056588
                          95.743243
                                        221.05297
## 2
                          59.062284
           16.849857
                                        113.66893
## 3
           5.532533
                           8.128866
                                        15.92806
## 4
                                        52.97485
           11.836407
                          29.561104
##
   new_data.Avg_UnitPrice new_data.Avg_ReturnRate
## 1
            9.195324
                           0.1917935
## 2
            5.768671
                           0.1472687
## 3
            3.384698
                           0.1457553
## 4
            3.959407
                           0.1505239
##
## Clustering vector:
   [1] 2 4 3 3 3 3 4 2 3 2 3 3 3 1 4 3 3 3 2 4 2 3 3 3 4 3 4 3 3 2 2 1 3 3 3 3
##
3
##
   4
   [75] 3 3 4 2 2 4 3 4 3 3 3 2 4 3 2 3 4 4 3 3 2 3 3 3 4 4 4 1 3 3 4 4 3 3 3 3
##
4
##
  3
##
  3
  ##
3
##
  3
##
  3
```

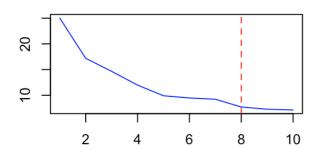
##	[297]	4	3	4	3	3	4	4	3	4	3	3	3	3	4	3	3	3	3	3	3	3	3	3	4	3	1	3	3	3	3	3	3	4	3	3	4
##	[334]	3	3	3	3	3	3	3	3	4	3	3	3	3	3	3	4	2	3	3	3	2	3	3	4	3	2	3	3	3	3	4	3	3	4	3	2
##	[371]	3	3	3	3	3	3	3	3	3	3	3	4	3	3	3	3	3	3	2	3	3	3	3	3	3	4	4	3	3	3	3	3	3	3	3	3
4 ##	[408]	4	3	3	3	3	3	3	4	3	3	3	2	3	3	3	3	3	4	3	3	4	3	4	3	3	3	3	4	4	3	4	4	4	4	2	3
3 ##	[445]	3	3	3	4	3	4	3	4	3	3	3	3	4	3	4	3	3	4	3	3	3	2	3	4	4	3	2	3	4	3	4	3	2	3	4	3
3 ##	[482]	3	3	3	4	3	3	1	3	4	3	3	2	4	3	4	3	3	4	3	2	4	3	4	4	4	3	3	2	4	3	3	3	3	3	4	3
3 ##	[519]	3	3	3	3	3	4	3	3	3	4	3	3	4	3	3	2	4	3	3	4	3	3	3	2	3	3	3	3	3	3	3	3	4	3	3	3
3 ##	[556]	3	2	3	4	4	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	4	2	3	3	3	3	4	3	4	3	3	3	3	4	2	3
3 ##	[593]	3	4	3	3	4	3	4	4	1	3	3	4	3	3	3	3	4	4	4	4	4	2	3	3	3	4	3	3	3	3	3	3	3	3	4	4
3 ##	[630]	3	3	4	4	3	3	3	4	4	2	3	4	3	3	4	4	3	2	2	3	4	3	3	4	2	2	2	2	3	3	4	2	4	4	3	4
2 ##	[667]	4	3	3	3	2	4	2	4	4	3	3	4	3	4	4	4	3	2	4	4	1	3	3	2	2	3	2	3	3	3	3	3	3	3	3	4
1 ##	[704]	4	2	4	4	1	2	4	3	1	2	3	2	3	4	3	2	2	4	2	3	2	4	2	3	4	4	4	3	1	3	4	4	4	4	3	4
3 ##	[741]	4	2	4	3	3	4	3	4	3	4	3	2	3	2	4	4	3	2	4	3	4	3	1	4	4	3	3	4	4	4	3	3	2	2	3	3
3 ##	[778]	3	3	4	2	4	3	3	4	3	4	4	3	3	2	1	3	3	3	3	3	1	2	4	4	3	2	1	1	4	1	3	4	3	3	3	4
4 ##	[815]	3	2	3	2	1	3	3	2	4	3	2	4	3	3	1	4	1	2	1	4	4	2	2	4	4	4	4	4	4	4	3	4	3	4	4	4
3 ##	[852]	3	3	4	3	3	3	4	4	2	4	3	2	3	2	4	4	3	4	3	3	3	1	4	4	4	2	2	2	3	2	4	4	3	3	4	3
3 ##	[889]	2	3	4	4	4	3	2	3	3	3	3	3	4	3	4	3	3	4	3	3	4	3	2	3	4	2	3	3	3	3	3	4	3	2	3	3
4 ##	[926]	4	3	2	3	4	3	4	3	3	3	3	3	4	4	3	4	3	4	3	4	4	1	4	2	3	2	3	2	1	4	3	1	4	3	3	4
3 ##	[963]	3	4	3	3	2	4	4	3	2	3	4	3	3	3	2	3	3	3	3	4	3	4	4	4	3	3	3	3	2	4	4	3	2	3	3	4
3 ##	[1000]	4	3	4	3	3	3	3	3	3	3	3	3	2	4	4	2	4	4	3	3	3	3	3	4	4	3	3	2	3	4	3	3	3	3	3	2
4	[1037]																																				
3	[1074]																																				
3	[1111]																																				
3	[1148]																																				
4																																					
##	[1185]	3	3	3	4	4	4	3	3	2	3	4	4	4	3	3	3	2	4	3	3	3	4	4	3	3	3	3	3	4	3	1	3	4	4	3	3

```
3
## [2221] 3 4 3 1 3 3 3 3 3 4 4 4 4 3 1 3 2 4 3 2 4 3 3 3 3 4 3 3 3 4 3 4 2 3 3 3
## [2295] 4 3 3 4 3 3 3 3 1 4 3 4 4 4 3 2 4 3 3 3 3 3 2 3 2 2 3 3 3 4 4 3 3 3 1 3
3
## [2443] 4 4 2 4 2 4 3 3 3 3 3 1 3 3 3 4 4 3 3 4 3 2 3 3 3 4 1 3 4 3 3 4 2 3 3 3
4
3
## [2665] 4 3 4 4 3 3 4 1 2 4 3 3 4 3 3 3 3 1 3 3 2 3 4 3 3 2 4 3 3 3 3 4 4 3 3 3
3
3
## [2850] 4 2 2 4 2 3 3 4 4 4 4 3 3 3 3 3 3 2 3 4 3
##
## Within cluster sum of squares by cluster:
## [1] 368717.0 682536.7 320013.5 562686.5
 (between SS / total SS = 76.7 %)
##
##
## Available components:
##
## [1] "cluster"
                        "tot.withinss"
        "centers"
             "totss"
                   "withinss"
## [6] "betweenss"
        "size"
             "iter"
                   "ifault"
##
## Results for nstart = 100 :
```

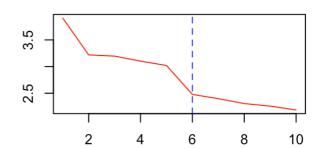
## tot.withinss: 1933954

```
optimal_clusters_complete <- 8
optimal clusters single <- 6
optimal clusters centroid <- 5
optimal_clusters_average <- 4
# Create elbow plots
par(mfrow=c(2,2))
# elbow plot for "complete"
y <- sort(hclust$height, decreasing = TRUE)[1:10]</pre>
plot(x, y, type = "1", col = "blue", main = "Method: Complete", xlab = "", ylab =
"")
abline(v = optimal_clusters_complete, col = "red", lty = 2)
# elbow plot for "single"
y1 <- sort(hclust1$height, decreasing = TRUE)[1:10]</pre>
plot(x, y1, type = "1", col = "red", main = "Method: Single", xlab = "", ylab = ""
abline(v = optimal_clusters_single, col = "blue", lty = 2)
# elbow plot for "centroid"
y2 <- sort(hclust2$height, decreasing = TRUE)[1:10]
plot(x, y2, type = "l", col = "green", main = "Method: Centroid", xlab = "", ylab
abline(v = optimal clusters centroid, col = "blue", lty = 2)
# elbow plot for "average"
y3 <- sort(hclust3$height, decreasing = TRUE)[1:10]
plot(x, y3, type = "1", col = "orange", main = "Method: Average", xlab = "", ylab
= "")
abline(v = optimal_clusters_average, col = "blue", lty = 2)
```

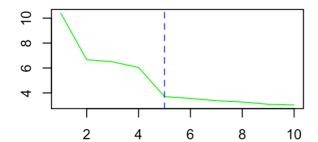
## **Method: Complete**



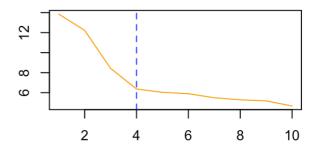
Method: Single



**Method: Centroid** 



## Method: Average



#endogram

plot(hclust)

plot(hclust1)

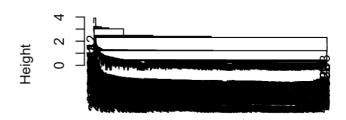
plot(hclust2)

plot(hclust3)

#### **Cluster Dendrogram**

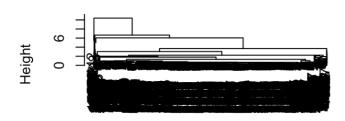
#### **Cluster Dendrogram**





### **Cluster Dendrogram**

## **Cluster Dendrogram**





:(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_C(scale(cbind(new\_data\$Avg\_Quantity, new\_data\$Total\_C
ata\$total\_value, new\_data\$P(\daggerightarrow g"data\$P(\daggerightarrow g"data\$P(\daggerightarr

##final k mean selection ---

 $seg_kmeans_final <- kmeans(x = data.frame(new_data$Avg_Quantity, new_data$Total_Quantity, new_data$total_value, new_data$Avg_UnitPrice, new_data$Avg_ReturnRate), centers = 6, nstart = 50) \\ seg_kmeans_final$tot.withinss$ 

## [1] 1477079

segment <- seg\_kmeans\_final\$cluster
segmentation <- cbind(new\_data, segment)
table(segmentation\$segment)</pre>

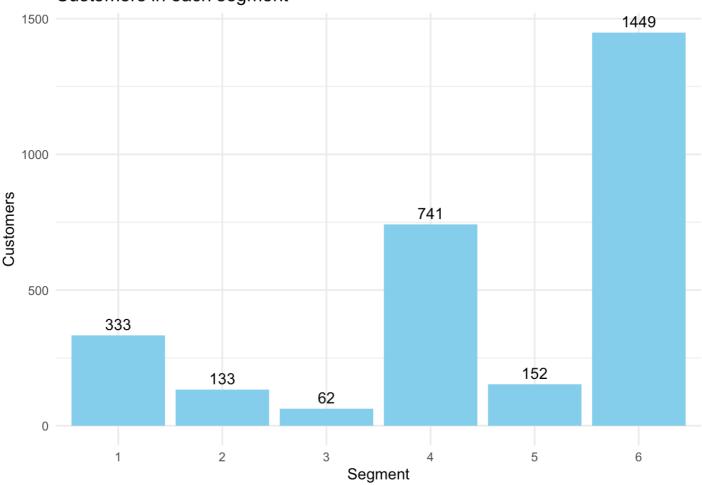
##
## 1 2 3 4 5 6
## 333 133 62 741 152 1449

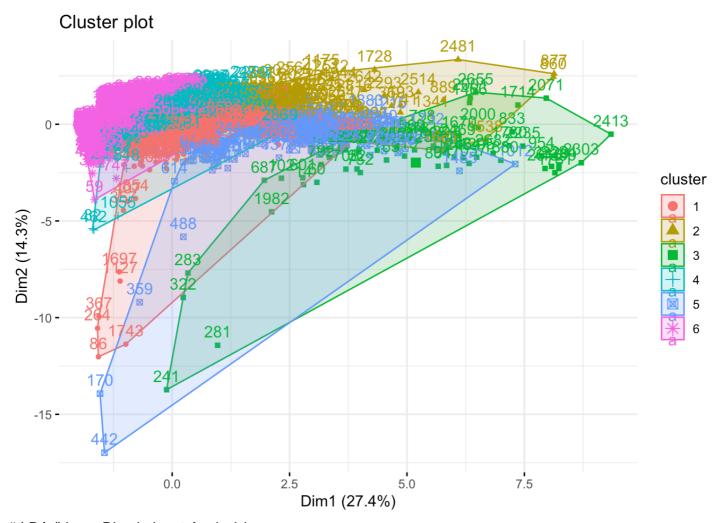
```
#visualize the segments
segment_counts <- table(segmentation$segment)
segment_data <- as.data.frame(segment_counts)

names(segment_data) <- c("Segment", "Count")

# Create bar plot
ggplot(segment_data, aes(x = Segment, y = Count)) +
geom_bar(stat = "identity", fill = "skyblue") +
geom_text(aes(label = Count), vjust = -0.5) +
labs(title = "Customers in each segment", x = "Segment", y = "Customers") +
theme_minimal()</pre>
```

## Customers in each segment





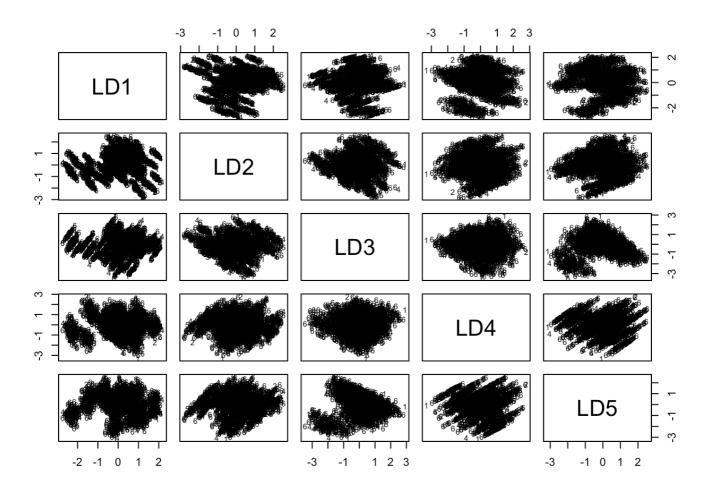
#### # LDA (Linear Discriminant Analysis)

```
##duplicate the data set to group the

library(MASS)

segmentation$Work <- as.factor(segmentation$Work)
segmentation$segment <- as.factor(segmentation$segment)

fit <- lda(segment ~ Married + Age + Income + Edcation + Work , data = segmentation)
plot(fit)</pre>
```



ldapred <- predict(fit, segmentation)</pre>

ld <- ldapred\$x</pre>

anova(lm(ld[,1]~segmentation\$segment))

	<b>Df</b> <int></int>	Sum Sq <dbl></dbl>	Mean Sq <dbl></dbl>	<b>F value</b> <dbl></dbl>	Pr(>F) <dbl></dbl>
segmentation\$segment	5	21.03867	4.207734	4.207734	0.0008197218
Residuals	2864	2864.00000	1.000000	NA	NA
2 rows					

anova(lm(ld[,2]~segmentation\$segment))

	<b>Df</b> <int></int>	Sum Sq <dbl></dbl>	Mean Sq <dbl></dbl>	F value <dbl></dbl>	<b>Pr(&gt;F)</b> <dbl></dbl>
segmentation\$segment	5	19.57101	3.914202	3.914202	0.001540761

Residuals	2864	2864.00000	1.000000	NA	NA
2 rows					

anova(lm(ld[,3]~segmentation\$segment))

	<b>Df</b> <int></int>	Sum Sq <dbl></dbl>	<b>Mean Sq</b> <dbl></dbl>	<b>F value</b> <dbl></dbl>	<b>Pr(&gt;F)</b> <dbl></dbl>
segmentation\$segment	5	8.567743	1.713549	1.713549	0.1279808
Residuals	2864	2864.000000	1.000000	NA	NA
2 rows					

anova(lm(ld[,4]~segmentation\$segment))

	<b>Df</b> <int></int>	Sum Sq <dbl></dbl>	<b>Mean Sq</b> <dbl></dbl>	<b>F value</b> <dbl></dbl>	<b>Pr(&gt;F)</b> <dbl></dbl>
segmentation\$segment	5	7.026431	1.405286	1.405286	0.2190462
Residuals	2864	2864.000000	1.000000	NA	NA
2 rows					

anova(lm(ld[,5]~segmentation\$segment))

	<b>Df</b> <int></int>	Sum Sq <dbl></dbl>	Mean Sq <dbl></dbl>	<b>F value</b> <dbl></dbl>	<b>Pr(&gt;F)</b> <dbl></dbl>
segmentation\$segment	5	4.292582	0.8585163	0.8585163	0.5082389
Residuals	2864	2864.000000	1.0000000	NA	NA
2 rows					

pred.seg <- predict(fit)\$class</pre>

cf<- table(segmentation\$segment, ldapred\$class)
cf</pre>

```
##
##
##
             0
                          0 333
        0
                 0
                 0
##
    2
        0
             0
                          0 133
             0
                 0
                      0
                         0 62
##
        0
             0
                0
                      0 0 741
##
##
                          0 152
                          0 1449
```

```
#overal accuracy of the predicting model
sum(diag(cf))/nrow(segmentation)
```

```
## [1] 0.504878
```

# RFM analysis

```
data <- data %>%
 mutate(revenue = Quantity * UnitPrice)
rfm <- data
data2 <- data %>%
  filter(!is.na(CustomerID))
rfm <- data %>%
  group_by(CustomerID) %>%
  summarise(
    revenue = sum(revenue),
    number of orders = n distinct(InvoiceNo),
    recency_days = round(as.numeric(difftime(as.POSIXct("2021-11-24 17:06:00 UTC",
format = "%Y-%m-%d %H:%M:%S", tz = "UTC"), max(InvoiceDate), units = "days"))),
    purchase = 1,
    zip_code = get_mode(ZipCode))
groups <- 5
## 5.3 Run RFM Analysis with Independent Sort
rfm$recency_score_indep <- ntile(rfm$recency_days*-1, groups)
rfm$frequency_score_indep <- ntile(rfm$number_of_orders, groups)
rfm$monetary score indep <- ntile(rfm$revenue, groups)</pre>
rfm$rfm score indep <- paste(rfm$recency score indep*100 + rfm$frequency score ind
ep * 10 + rfm$monetary score indep)
rfm$recency_score_seq <- ntile(rfm$recency_days*-1, groups)
r_groups <- NULL; rf_groups <- NULL; temp <- NULL ## Initialize empty matrices
for (r in 1:groups) {
  r_groups[[r]] <- filter(rfm, rfm$recency_score_seq == r)</pre>
```

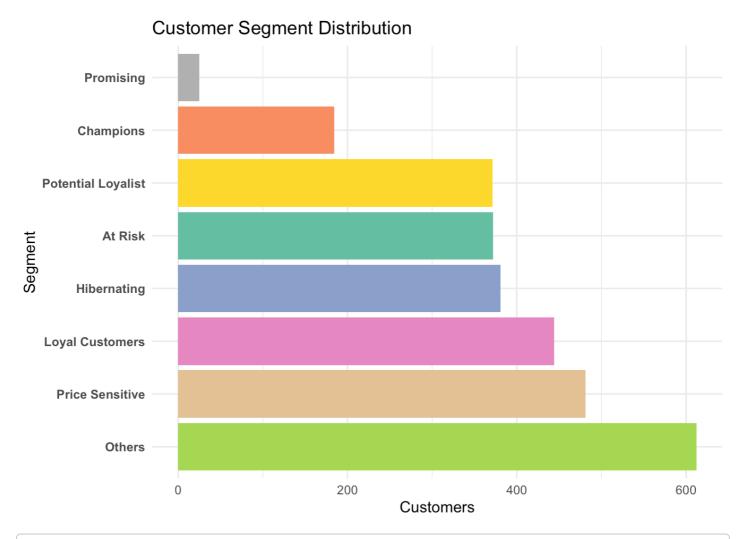
```
r groups[[r]]$frequency score seq <- ntile(r groups[[r]]$number of orders, group
s)
  for (m in 1:groups) {
    rf groups[[m]] <- filter(r groups[[r]], r groups[[r]]$frequency score seq == m</pre>
)
    rf_groups[[m]]$monetary_score_seq <- ntile(rf_groups[[m]]$revenue, groups)</pre>
    temp <- bind_rows(temp, rf_groups[[m]])</pre>
  }
}
rfm_result <- temp[order(temp$CustomerID),]</pre>
rfm result$rfm_score_seq <- paste(rfm_result$recency_score_seq*100 + rfm_result$fr
equency_score_seq * 10 + rfm_result$monetary_score_seq)
## Export RFM Results with Independent and Sequential Sort
#write.csv(rfm result, "Q:/Marketing Analytics/rfm results.csv", row.names = FALSE
) ## Name file rfm result.csv
rfm result <- data.frame(rfm result)</pre>
##customer segmentation for rfm results
rfm_result <- rfm_result %>%
 mutate(
    Segment2 = case when(
      recency score seq <= 2 & frequency score seq >= 4 & monetary score seq >= 4
                              # Best customers
~ "Champions",
      recency score seq <= 3 & frequency score seq >= 3 & monetary score seq >= 3
~ "Loyal Customers",
                            # Consistently good customers
      recency_score_seq <= 2 & frequency_score_seq <= 3 & monetary_score_seq <= 3</pre>
~ "Potential Loyalist",
                            # Newer customers with potential
      recency score seq >= 4 & frequency score seq >= 3 & monetary score seq >= 3
~ "At Risk",
                             # Good customers at risk of leaving
      recency_score_seq == 1 & frequency_score_seq <= 2 & monetary_score_seq <= 2</pre>
~ "New Customers",
                              # Newest customers
      recency score seq <= 3 & frequency score seq == 2 & monetary score seq == 2
~ "Promising",
                              # Showing potential but needs nurturing
      recency score seq >= 4 & frequency score seq <= 2 & monetary score seq >= 2
~ "Hibernating",
                              # Low engagement but still spending
      frequency_score_seq >= 4 & monetary_score_seq <= 2 ~ "Price Sensitive",
# Frequent but low spending
      TRUE ~ "Others"
# Catch-all for any that don't fit above categories
  )
#join rfm table and segmentation table
join <- inner join(rfm result, segmentation, by = "CustomerID")</pre>
```

```
segment_counts <- join %>%
  group_by(Segment2) %>%
  summarise(Count = n())

print(segment_counts)
```

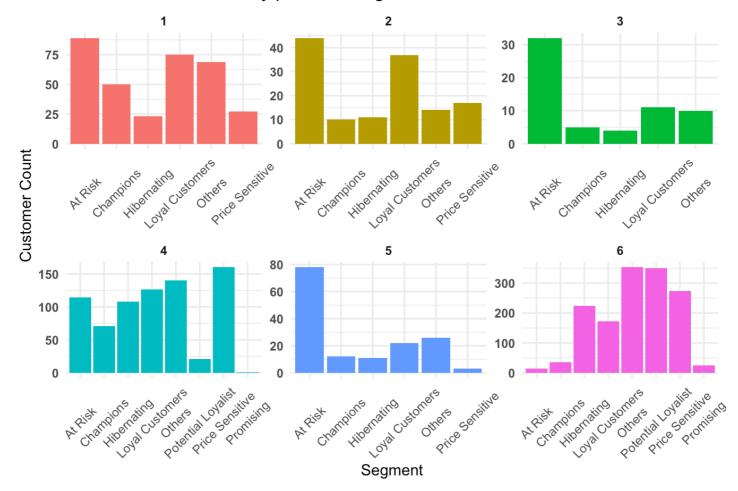
```
## # A tibble: 8 × 2
     Segment2
##
                        Count
##
    <chr>
                        <int>
## 1 At Risk
                          372
## 2 Champions
                          184
## 3 Hibernating
                          381
## 4 Loyal Customers
                         444
## 5 Others
                          612
## 6 Potential Loyalist
                          371
## 7 Price Sensitive
                          481
## 8 Promising
                           25
```

```
#bar plot of customer segmentation with RFM
ggplot(segment_counts, aes(x = reorder(Segment2, -Count), y = Count, fill = Segmen
t2)) +
   geom_bar(stat = "identity", show.legend = FALSE) +
   theme_minimal() +
   labs(x = "Segment", y = "Customers", title = "Customer Segment Distribution") +
   coord_flip() +
   scale_fill_brewer(palette = "Set2") +
   theme(
      axis.text.y = element_text(face = "bold")
)
```



```
#bar plot of customer combining both types of segmentation done
ggplot(join, aes(x = Segment2, fill = segment)) +
  geom_bar(show.legend = FALSE) +
  theme_minimal() +
  labs(x = "Segment", y = "Customer Count", title = "Customer Distribution by prev
ious Segments") +
  facet_wrap(~ segment, scales = "free") +
  theme(
    strip.text = element_text(face = "bold"),
    axis.text.y = element_text(face = "bold"),
    axis.text.x = element_text(angle = 45, vjust = 0.5)
)
```

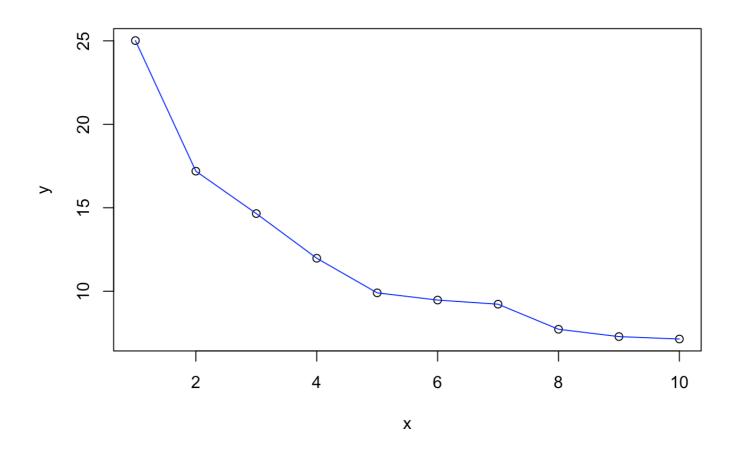
### Customer Distribution by previous Segments



# Kmeans based on RFM results

```
hclust4 <- hclust(dist(scale(cbind(rfm_result$recency_score_seq, rfm_result$freque
ncy_score_seq, rfm_result$monetary_score_seq))), method = "complete")

y <- sort(hclust$height, decreasing = TRUE)[1:10]
plot(x,y); lines(x,y, col= "blue")</pre>
```



```
kmeans_rfm <- kmeans(x = data.frame(rfm_result$recency_score_seq, rfm_result$frequ
ency_score_seq, rfm_result$monetary_score_seq), centers = 8, nstart = 50)

segmentrfm <- kmeans_rfm$cluster
segmentationrfm <- cbind(rfm_result, segmentrfm)
segment_countsrfm <- table(segmentationrfm$segmentrfm)

segment_datarfm <- as.data.frame(segment_countsrfm)

names(segment_datarfm) <- c("Segment", "Count")

ggplot(segment_datarfm, aes(x = Segment, y = Count)) +
    geom_bar(stat = "identity", fill = "skyblue") +
    geom_text(aes(label = Count), vjust = 0) + # Añadir etiquetas de conteo encima d
e las barras
    labs(title = "Customers in each segment", x = "Segment", y = "Customers") +
    theme_minimal()</pre>
```

