MKTG-352 Asssignment 2

MinJae Jo

2025-10-02

Part 1

Step 0

```
suppressPackageStartupMessages({
  library(dplyr)
  library(cluster)
})
```

Step 1

```
Smartwatch <- read.csv("SmartwatchData.csv", stringsAsFactors = FALSE)

summary_stats <- Smartwatch %>%
  summarise(
    Age_mean = mean(Age, na.rm = TRUE), Age_min = min(Age, na.rm = TRUE), Age_max = max(Age, na
    Price_mean = mean(Price, na.rm = TRUE), Price_min = min(Price, na.rm = TRUE), Price_max = m
    Innov_mean = mean(Innov, na.rm = TRUE), Innov_min = min(Innov, na.rm = TRUE), Innov_max = m
    ConstCom_mean = mean(ConstCom, na.rm = TRUE), ConstCom_min = min(ConstCom, na.rm = TRUE), C
    SaveM_T_mean = mean(SaveM_T, na.rm = TRUE), SaveM_T_min = min(SaveM_T, na.rm = TRUE), Save
  )

summary_stats
```

```
## # A tibble: 1 x 15
##   Age_mean Age_min Age_max Price_mean Price_min Price_max Innov_mean Innov_min
##      <dbl>   <int>   <int>      <dbl>     <int>     <int>      <dbl>     <int>
## 1     35.5      24      47       213.       100       390       4.10         1
## # i 7 more variables: Innov_max <int>, ConstCom_mean <dbl>, ConstCom_min <int>,
## #   ConstCom_max <int>, SaveM_T_mean <dbl>, SaveM_T_min <int>,
## #   SaveM_T_max <int>
```

Step 2

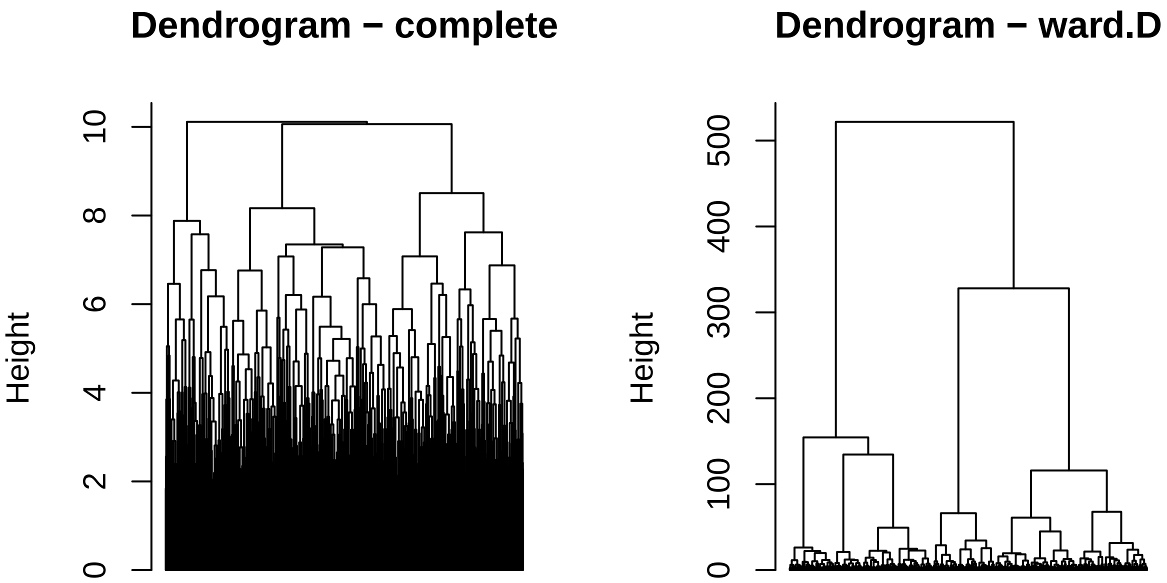
```
seg_vars <- c("Innov","ConstCom","CreatCom","TimelyInf","SaveM_T","SaveM_L",
              "TaskMgm","DeviceSt","Photo","Wellness","Athlete","Style","Price")

SegmentationData <- Smartwatch %>%
  select(1, all_of(seg_vars)) %>%
  rename(Observations = 1)

SegmentationDataScaled <- SegmentationData
SegmentationDataScaled[, -1] <- scale(SegmentationDataScaled[, -1])
```

Step 3

```
dissimilarity <- dist(SegmentationDataScaled[, -1], method = "euclidean")
```

Step 4

```
hc_complete <- hclust(dissimilarity, method = "complete")
hc_ward     <- hclust(dissimilarity, method = "ward.D")

par(mfrow = c(1,2))
plot(hc_complete, labels = FALSE, hang = -1, main = "Dendrogram - complete", xlab = "", sub = "
plot(hc_ward,     labels = FALSE, hang = -1, main = "Dendrogram - ward.D",   xlab = "", sub = "
```

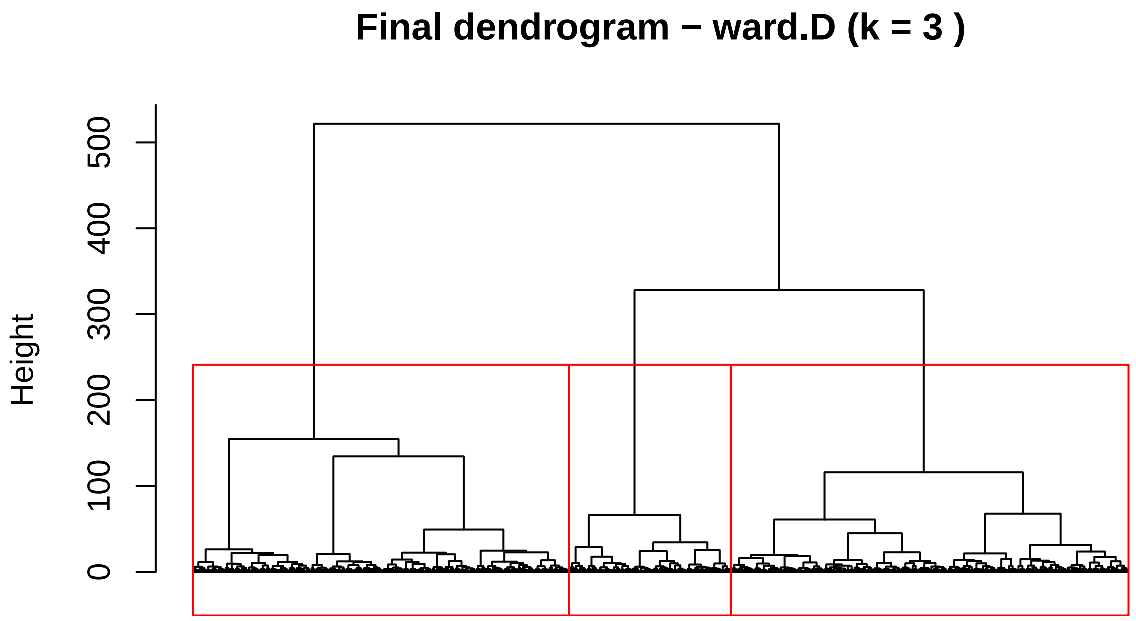


```r
par(mfrow = c(1,1))
```

Step 5

```r
ks <- 2:10

avg_sil <- function(hc, d, k) {
  cl <- cutree(hc, k = k)
  sil <- cluster::silhouette(cl, d)
  mean(sil[, 3])
}

sil_complete <- sapply(ks, function(k) avg_sil(hc_complete, dissimilarity, k))
sil_ward     <- sapply(ks, function(k) avg_sil(hc_ward,     dissimilarity, k))

best_k_complete <- ks[which.max(sil_complete)]
best_k_ward     <- ks[which.max(sil_ward)]

best_method <- if (max(sil_complete) >= max(sil_ward)) "complete" else "ward.D"
best_k      <- if (best_method == "complete") best_k_complete else best_k_ward
hc_final    <- if (best_method == "complete") hc_complete else hc_ward
```

```
plot(hc_final, labels = FALSE, hang = -1,
     main = paste("Final dendrogram -", best_method, "(k =", best_k, ")"),
     xlab = "", sub = "")
rect.hclust(hc_final, k = best_k, border = "red")
```



```
best_method; best_k
```

```
## [1] "ward.D"
```

```
## [1] 3
```

Step 6

```
final_clusters <- cutree(hc_final, k = best_k)
Smartwatch$Cluster <- factor(final_clusters)
head(Smartwatch$Cluster)
```

```
## [1] 1 2 2 2 2 1
## Levels: 1 2 3
```

Step 7

```r
cluster_counts <- table(Smartwatch$Cluster)
cluster_counts
```

```
##
##   1   2   3
## 425 173 402
```

Step 8

```r
cluster_summary <- Smartwatch %>%
  group_by(Cluster) %>%
  summarise(across(where(is.numeric), ~ mean(.x, na.rm = TRUE)), .groups = "drop")
cluster_summary
```

```
## # A tibble: 3 x 39
##   Cluster Observations Innov ConstCom CreatCom TimelyInf SaveM_T SaveM_L TaskMgm
##   <fct>          <dbl> <dbl>    <dbl>    <dbl>     <dbl>   <dbl>   <dbl>   <dbl>
## 1 1               511.  3.69     4.09     4.19      3.67    3.45    3.00    3.27
## 2 2               484.  4.17     5.46     2.27      5.64    2.03    3.77    5.06
## 3 3               496.  4.50     4.99     5.20      4.36    4.07    5.10    4.79
## # i 30 more variables: DeviceSt <dbl>, Photo <dbl>, Wellness <dbl>,
## #   Athlete <dbl>, Style <dbl>, Price <dbl>, iPhone <dbl>, CompBuy <dbl>,
## #   Health <dbl>, Finc <dbl>, Sales <dbl>, Advt <dbl>, Edu <dbl>, Cons <dbl>,
## #   Eng <dbl>, Tech <dbl>, Retail <dbl>, SMB <dbl>, FB_Insta <dbl>, Twit <dbl>,
## #   Snap <dbl>, YouTube <dbl>, Pod_radio <dbl>, TV <dbl>, NewsP <dbl>,
## #   AmznP <dbl>, Age <dbl>, Female <dbl>, Degree <dbl>, Income <dbl>
```

Step 9

```r
write.csv(cluster_summary, "Segment_summary.csv", row.names = FALSE)
```