**Assignment 3**

**Author:** Keerthy Raaj Shanmugam, 8779954

**Course:** PROG8430

**1 Data Transformation,**

* 1. **Standardize all of the variables using either of the two functions demonstrated in class. Describe why you chose the method you did.**

We used **𝑧𝑖 = 𝑥𝑖 − min(𝑥) / max (𝑥) − min(𝑥)** method to standardize data because the data doesn’t have outliers and data are tightly clustered.

Food Entr Educ Tran

1 0.2547170 0.16226415 0.34150943 0.25283019

2 0.1849057 0.19433962 0.01509434 0.36981132

3 0.2566038 0.02075472 0.33396226 0.01509434

4 0.2773585 0.01698113 0.36981132 0.02075472

5 0.3471698 0.04528302 0.24339623 0.06226415

6 0.3113208 0.02452830 0.38679245 0.01886792

Work Hous Othr

1 0.05471698 0.6358491 0.06603774

2 0.02452830 0.8716981 0.10754717

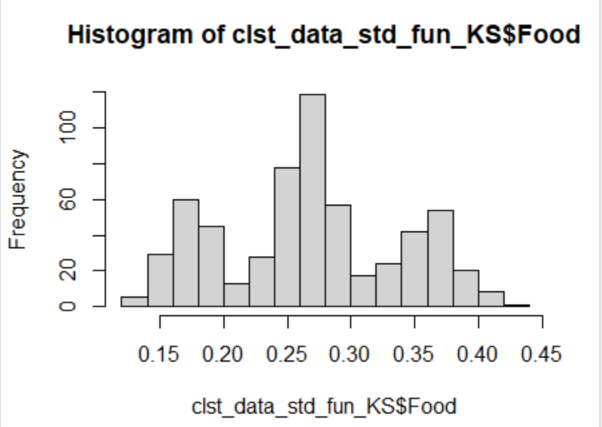
3 0.15471698 0.9716981 0.01509434

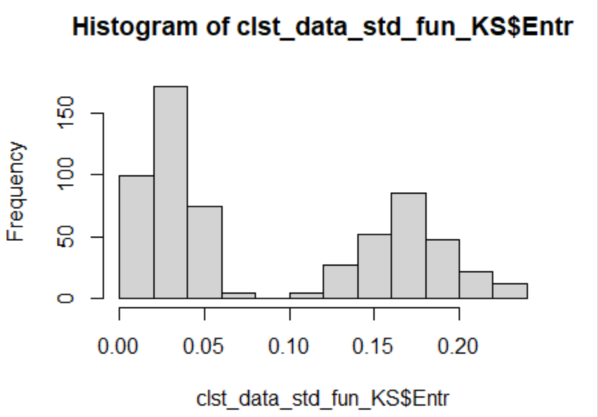
4 0.11320755 0.9490566 0.02075472

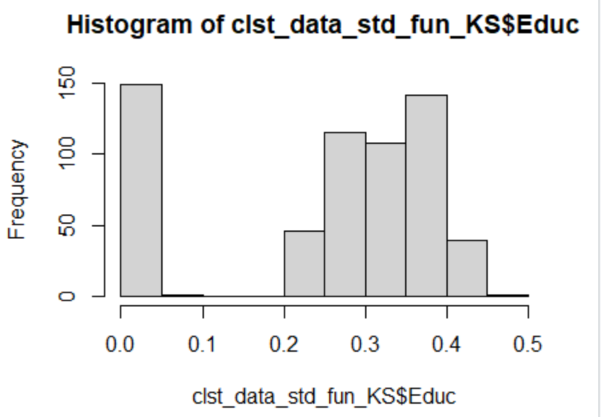
5 0.03962264 0.9792453 0.05094340

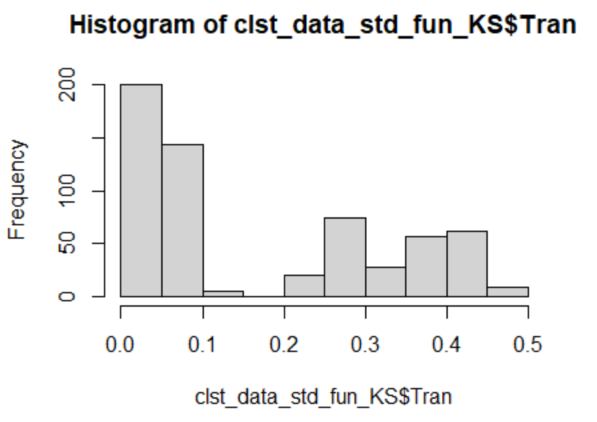
6 0.12452830 0.8830189 0.01886792

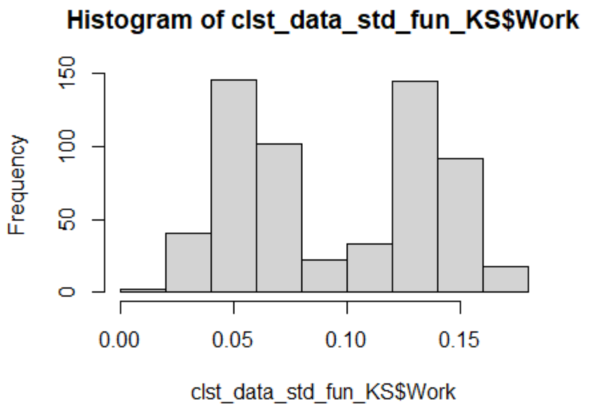
1. **Descriptive Data Analysis,**
   1. **Create graphical summaries of the data (as demonstrated in class: boxplots or histograms) and comment on any observations you make.**

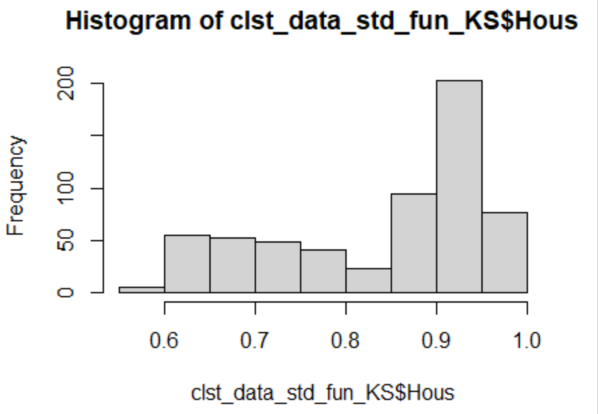
****

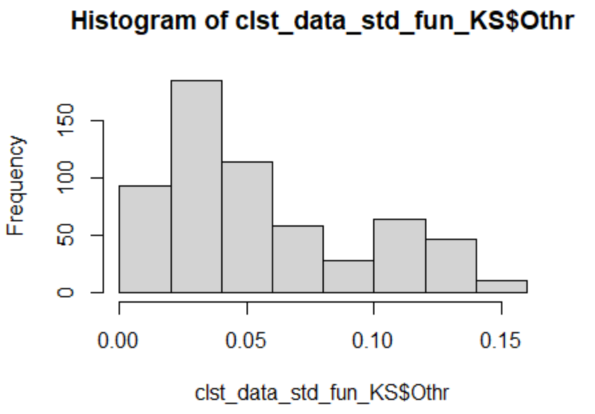
****

****

****

****

****

****

**3 Clustering,**

**Using the K-Means procedure as demonstrated in class, create clusters with k=3,4,5,6,7. You will be using only two variables as your centroids (Food and Work)**

**3.1 Create segmentation/cluster schemes for k=3,4,5,6,7.**

**K=3**

K-means clustering with 3 clusters of sizes 150, 350, 100

Cluster means:

Food Entr Educ Tran

1 0.1741887 0.18272956 0.02040252 0.39651572

2 0.3075849 0.02891105 0.32941779 0.04467385

3 0.2649245 0.15232075 0.32050943 0.26871698

Work Hous Othr

1 0.10333333 0.7752201 0.11553459

2 0.10219407 0.9256927 0.02945013

3 0.05684906 0.6449057 0.05969811

Clustering vector:

[1] 3 1 2 2 2 2 2 3 2 2 2 2 2 2 3 2 2 2 2 2 2 2 2 2 2

[26] 2 2 2 2 2 2 3 3 2 2 3 3 2 2 3 2 2 1 1 2 2 1 1 3 1

[51] 2 1 2 2 2 1 2 1 2 1 2 3 3 2 3 1 3 2 3 2 1 2 2 1 1

[76] 1 3 2 2 2 1 2 2 1 3 3 2 2 2 2 1 1 2 2 3 1 2 2 2 2

[101] 2 2 3 1 2 1 1 2 3 2 1 1 2 3 3 3 1 2 2 1 1 3 2 1 2

[126] 2 1 2 2 2 1 3 2 2 2 1 2 2 2 2 2 3 2 2 3 1 3 1 2 2

[151] 1 2 2 2 1 1 1 2 2 2 3 3 2 2 2 2 3 2 2 1 3 2 1 1 2

[176] 2 2 1 1 1 2 3 2 2 2 1 2 2 2 3 1 2 2 2 2 1 1 2 2 1

[201] 2 1 2 2 1 1 1 2 1 2 2 2 1 1 2 2 2 1 2 1 3 2 3 1 2

[226] 3 2 1 1 3 3 2 1 2 2 2 1 1 2 2 1 3 2 2 2 2 1 1 2 2

[251] 2 3 2 2 3 1 2 3 1 2 1 2 2 3 3 2 1 2 3 1 2 3 2 1 3

[276] 3 2 1 2 2 3 2 2 2 2 2 1 2 3 2 2 1 2 2 2 2 2 2 2 2

[301] 1 2 2 3 2 1 2 2 2 2 1 1 1 2 2 3 2 2 3 1 2 2 1 2 3

[326] 2 2 2 2 2 1 1 2 2 3 2 2 3 2 3 1 2 1 2 2 2 2 2 1 2

[351] 3 2 1 1 2 3 2 2 1 2 2 2 1 3 2 2 2 2 2 1 2 2 2 3 1

[376] 1 3 1 2 2 1 2 2 3 2 1 1 1 2 2 2 2 3 2 2 2 3 2 1 2

[401] 2 2 2 1 1 2 2 2 2 2 2 3 2 1 2 2 3 3 2 1 1 2 3 2 1

[426] 2 3 1 2 1 3 3 2 3 1 3 2 1 2 2 2 1 3 2 2 2 2 2 2 3

[451] 2 2 2 2 2 2 2 3 1 2 1 1 2 2 1 3 1 2 2 1 1 2 2 3 2

[476] 2 2 2 3 3 2 2 2 2 3 2 3 1 2 3 2 2 1 2 2 1 1 2 2 2

[501] 3 2 3 1 2 2 2 2 2 2 3 1 1 2 3 2 2 2 1 1 1 3 3 1 2

[526] 2 2 3 2 3 2 1 2 2 1 2 2 3 2 2 1 2 1 2 2 1 1 2 2 2

[551] 3 2 1 2 2 3 1 2 2 2 2 2 2 2 1 2 2 1 3 2 1 2 1 2 1

[576] 2 2 2 3 2 1 2 3 1 1 1 2 2 2 2 1 2 2 1 2 2 2 3 2 1

Within cluster sum of squares by cluster:

[1] 1.2749212 3.6176584 0.2642503

(between\_SS / total\_SS = 87.9 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

**K=4**

K-means clustering with 4 clusters of sizes 200, 150, 100, 150

Cluster means:

Food Entr Educ Tran

1 0.2665849 0.02020755 0.37933019 0.02100000

2 0.1741887 0.18272956 0.02040252 0.39651572

3 0.2649245 0.15232075 0.32050943 0.26871698

4 0.3622516 0.04051572 0.26286792 0.07623899

Work Hous Othr

1 0.13495283 0.9248208 0.02102830

2 0.10333333 0.7752201 0.11553459

3 0.05684906 0.6449057 0.05969811

4 0.05851572 0.9268553 0.04067925

Clustering vector:

[1] 3 2 1 1 4 1 1 3 4 1 1 1 1 4 3 1 1 1 1 4 4 1 1 1 1

[26] 4 4 4 4 1 4 3 3 4 4 3 3 4 1 3 1 1 2 2 1 1 2 2 3 2

[51] 4 2 4 4 1 2 1 2 1 2 4 3 3 1 3 2 3 1 3 4 2 1 4 2 2

[76] 2 3 1 4 4 2 1 1 2 3 3 4 4 4 4 2 2 1 1 3 2 1 4 4 4

[101] 1 1 3 2 1 2 2 4 3 1 2 2 4 3 3 3 2 1 1 2 2 3 4 2 1

[126] 1 2 1 4 4 2 3 4 4 4 2 4 1 1 4 1 3 4 4 3 2 3 2 1 4

[151] 2 1 1 4 2 2 2 1 1 1 3 3 4 1 1 4 3 1 1 2 3 4 2 2 4

[176] 4 1 2 2 2 4 3 4 1 1 2 1 1 4 3 2 1 1 1 4 2 2 1 4 2

[201] 1 2 4 4 2 2 2 4 2 1 1 1 2 2 4 1 4 2 4 2 3 4 3 2 4

[226] 3 4 2 2 3 3 1 2 4 4 4 2 2 4 4 2 3 4 4 1 4 2 2 1 4

[251] 1 3 1 1 3 2 1 3 2 1 2 4 1 3 3 1 2 1 3 2 1 3 1 2 3

[276] 3 1 2 1 4 3 1 4 4 1 4 2 1 3 4 4 2 1 1 4 1 4 1 1 1

[301] 2 1 1 3 1 2 4 1 4 1 2 2 2 1 4 3 1 4 3 2 1 1 2 1 3

[326] 4 4 1 4 1 2 2 4 1 3 4 1 3 1 3 2 4 2 1 1 1 1 4 2 1

[351] 3 1 2 2 4 3 1 1 2 4 1 4 2 3 4 4 1 1 1 2 1 4 4 3 2

[376] 2 3 2 1 4 2 1 1 3 4 2 2 2 1 1 1 4 3 4 1 4 3 1 2 4

[401] 1 1 4 2 2 1 4 4 1 1 4 3 1 2 1 4 3 3 1 2 2 4 3 1 2

[426] 4 3 2 4 2 3 3 4 3 2 3 4 2 1 1 1 2 3 4 1 1 1 1 1 3

[451] 1 1 4 4 4 4 1 3 2 1 2 2 1 4 2 3 2 1 1 2 2 4 4 3 1

[476] 4 1 4 3 3 1 1 1 1 3 1 3 2 4 3 1 1 2 4 1 2 2 4 4 1

[501] 3 1 3 2 4 4 1 1 1 4 3 2 2 1 3 1 1 4 2 2 2 3 3 2 4

[526] 1 4 3 1 3 4 2 1 4 2 1 1 3 1 4 2 1 2 1 1 2 2 1 4 4

[551] 3 1 2 1 4 3 2 1 4 1 4 1 4 1 2 1 1 2 3 1 2 1 2 4 2

[576] 1 4 4 3 1 2 4 3 2 2 2 1 1 4 4 2 4 4 2 1 1 1 3 1 2

Within cluster sum of squares by cluster:

[1] 0.4893857 1.2749212 0.2642503 0.3500783

(between\_SS / total\_SS = 94.4 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

**K=5**

K-means clustering with 5 clusters of sizes 150, 50, 100, 200, 100

Cluster means:

Food Entr Educ Tran

1 0.3622516 0.04051572 0.26286792 0.07623899

2 0.1770189 0.20464151 0.01905660 0.35958491

3 0.2649245 0.15232075 0.32050943 0.26871698

4 0.2665849 0.02020755 0.37933019 0.02100000

5 0.1727736 0.17177358 0.02107547 0.41498113

Work Hous Othr

1 0.05851572 0.9268553 0.04067925

2 0.03966038 0.8512075 0.11675472

3 0.05684906 0.6449057 0.05969811

4 0.13495283 0.9248208 0.02102830

5 0.13516981 0.7372264 0.11492453

Clustering vector:

[1] 3 2 4 4 1 4 4 3 1 4 4 4 4 1 3 4 4 4 4 1 1 4 4 4 4

[26] 1 1 1 1 4 1 3 3 1 1 3 3 1 4 3 4 4 5 2 4 4 2 5 3 5

[51] 1 5 1 1 4 5 4 5 4 5 1 3 3 4 3 5 3 4 3 1 5 4 1 5 2

[76] 2 3 4 1 1 5 4 4 5 3 3 1 1 1 1 5 5 4 4 3 5 4 1 1 1

[101] 4 4 3 5 4 2 5 1 3 4 5 5 1 3 3 3 5 4 4 5 5 3 1 5 4

[126] 4 5 4 1 1 5 3 1 1 1 5 1 4 4 1 4 3 1 1 3 5 3 2 4 1

[151] 5 4 4 1 2 5 5 4 4 4 3 3 1 4 4 1 3 4 4 5 3 1 5 5 1

[176] 1 4 2 5 5 1 3 1 4 4 5 4 4 1 3 5 4 4 4 1 5 5 4 1 2

[201] 4 2 1 1 5 5 5 1 5 4 4 4 5 2 1 4 1 2 1 5 3 1 3 2 1

[226] 3 1 2 5 3 3 4 5 1 1 1 2 5 1 1 2 3 1 1 4 1 5 2 4 1

[251] 4 3 4 4 3 5 4 3 5 4 5 1 4 3 3 4 5 4 3 2 4 3 4 5 3

[276] 3 4 5 4 1 3 4 1 1 4 1 5 4 3 1 1 5 4 4 1 4 1 4 4 4

[301] 2 4 4 3 4 5 1 4 1 4 5 2 5 4 1 3 4 1 3 5 4 4 5 4 3

[326] 1 1 4 1 4 5 2 1 4 3 1 4 3 4 3 2 1 2 4 4 4 4 1 2 4

[351] 3 4 5 5 1 3 4 4 2 1 4 1 5 3 1 1 4 4 4 5 4 1 1 3 5

[376] 2 3 5 4 1 5 4 4 3 1 2 5 2 4 4 4 1 3 1 4 1 3 4 5 1

[401] 4 4 1 2 5 4 1 1 4 4 1 3 4 5 4 1 3 3 4 5 2 1 3 4 5

[426] 1 3 2 1 5 3 3 1 3 5 3 1 5 4 4 4 5 3 1 4 4 4 4 4 3

[451] 4 4 1 1 1 1 4 3 5 4 2 2 4 1 2 3 5 4 4 5 2 1 1 3 4

[476] 1 4 1 3 3 4 4 4 4 3 4 3 2 1 3 4 4 5 1 4 2 2 1 1 4

[501] 3 4 3 5 1 1 4 4 4 1 3 2 5 4 3 4 4 1 5 5 2 3 3 2 1

[526] 4 1 3 4 3 1 2 4 1 5 4 4 3 4 1 5 4 2 4 4 5 2 4 1 1

[551] 3 4 2 4 1 3 5 4 1 4 1 4 1 4 2 4 4 5 3 4 5 4 2 1 2

[576] 4 1 1 3 4 5 1 3 5 2 5 4 4 1 1 5 1 1 5 4 4 4 3 4 5

Within cluster sum of squares by cluster:

[1] 0.3500783 0.1276852 0.2642503 0.4893857 0.2709613

(between\_SS / total\_SS = 96.5 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

**K=6**

K-means clustering with 6 clusters of sizes 100, 200, 50, 77, 100, 73

Cluster means:

Food Entr Educ Tran

1 0.1727736 0.17177358 0.02107547 0.41498113

2 0.2665849 0.02020755 0.37933019 0.02100000

3 0.1770189 0.20464151 0.01905660 0.35958491

4 0.3732664 0.03908356 0.27282529 0.07547170

5 0.2649245 0.15232075 0.32050943 0.26871698

6 0.3506332 0.04202636 0.25236495 0.07704833

Work Hous Othr

1 0.13516981 0.7372264 0.11492453

2 0.13495283 0.9248208 0.02102830

3 0.03966038 0.8512075 0.11675472

4 0.06049988 0.9048763 0.04190149

5 0.05684906 0.6449057 0.05969811

6 0.05642285 0.9500388 0.03939002

Clustering vector:

[1] 5 3 2 2 6 2 2 5 6 2 2 2 2 6 5 2 2 2 2 6 4 2 2 2 2

[26] 6 4 4 4 2 6 5 5 6 6 5 5 4 2 5 2 2 1 3 2 2 3 1 5 1

[51] 4 1 4 6 2 1 2 1 2 1 6 5 5 2 5 1 5 2 5 6 1 2 4 1 3

[76] 3 5 2 6 4 1 2 2 1 5 5 6 6 4 4 1 1 2 2 5 1 2 6 4 6

[101] 2 2 5 1 2 3 1 4 5 2 1 1 4 5 5 5 1 2 2 1 1 5 6 1 2

[126] 2 1 2 4 6 1 5 6 6 6 1 4 2 2 4 2 5 4 4 5 1 5 3 2 6

[151] 1 2 2 4 3 1 1 2 2 2 5 5 6 2 2 4 5 2 2 1 5 6 1 1 6

[176] 6 2 3 1 1 4 5 4 2 2 1 2 2 4 5 1 2 2 2 4 1 1 2 6 3

[201] 2 3 6 4 1 1 1 4 1 2 2 2 1 3 6 2 4 3 4 1 5 6 5 3 4

[226] 5 4 3 1 5 5 2 1 6 4 4 3 1 6 4 3 5 4 6 2 6 1 3 2 6

[251] 2 5 2 2 5 1 2 5 1 2 1 4 2 5 5 2 1 2 5 3 2 5 2 1 5

[276] 5 2 1 2 6 5 2 4 6 2 6 1 2 5 6 6 1 2 2 4 2 6 2 2 2

[301] 3 2 2 5 2 1 4 2 4 2 1 3 1 2 4 5 2 4 5 1 2 2 1 2 5

[326] 6 4 2 6 2 1 3 4 2 5 6 2 5 2 5 3 6 3 2 2 2 2 4 3 2

[351] 5 2 1 1 4 5 2 2 3 4 2 6 1 5 6 6 2 2 2 1 2 6 4 5 1

[376] 3 5 1 2 6 1 2 2 5 4 3 1 3 2 2 2 6 5 4 2 4 5 2 1 6

[401] 2 2 6 3 1 2 4 6 2 2 4 5 2 1 2 4 5 5 2 1 3 6 5 2 1

[426] 6 5 3 4 1 5 5 4 5 1 5 4 1 2 2 2 1 5 6 2 2 2 2 2 5

[451] 2 2 6 6 6 4 2 5 1 2 3 3 2 4 3 5 1 2 2 1 3 6 4 5 2

[476] 4 2 4 5 5 2 2 2 2 5 2 5 3 4 5 2 2 1 6 2 3 3 4 4 2

[501] 5 2 5 1 4 6 2 2 2 6 5 3 1 2 5 2 2 4 1 1 3 5 5 3 4

[526] 2 6 5 2 5 4 3 2 4 1 2 2 5 2 6 1 2 3 2 2 1 3 2 4 6

[551] 5 2 3 2 4 5 1 2 6 2 6 2 4 2 3 2 2 1 5 2 1 2 3 4 3

[576] 2 6 4 5 2 1 6 5 1 3 1 2 2 6 6 1 4 4 1 2 2 2 5 2 1

Within cluster sum of squares by cluster:

[1] 0.2709613 0.4893857 0.1276852 0.1298010 0.2642503

[6] 0.1076845

(between\_SS / total\_SS = 96.7 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

**K=7**

K-means clustering with 7 clusters of sizes 66, 100, 84, 100, 50, 100, 100

Cluster means:

Food Entr Educ Tran

1 0.3499428 0.04239565 0.25048599 0.07758719

2 0.1727736 0.17177358 0.02107547 0.41498113

3 0.3719227 0.03903863 0.27259659 0.07517969

4 0.2649245 0.15232075 0.32050943 0.26871698

5 0.1770189 0.20464151 0.01905660 0.35958491

6 0.2617547 0.01839623 0.36537736 0.02145283

7 0.2714151 0.02201887 0.39328302 0.02054717

Work Hous Othr

1 0.05568897 0.9521727 0.03965123

2 0.13516981 0.7372264 0.11492453

3 0.06073675 0.9069632 0.04148697

4 0.05684906 0.6449057 0.05969811

5 0.03966038 0.8512075 0.11675472

6 0.13264151 0.9469057 0.02139623

7 0.13726415 0.9027358 0.02066038

Clustering vector:

[1] 4 5 6 6 1 7 6 4 1 6 7 6 7 1 4 6 7 6 6 1 3 7 7 7 6

[26] 1 3 3 3 7 1 4 4 1 1 4 4 3 6 4 6 6 2 5 7 6 5 2 4 2

[51] 3 2 3 1 7 2 7 2 7 2 3 4 4 7 4 2 4 6 4 1 2 6 3 2 5

[76] 5 4 6 1 3 2 6 7 2 4 4 3 1 3 3 2 2 7 7 4 2 7 3 3 1

[101] 7 7 4 2 7 5 2 3 4 7 2 2 3 4 4 4 2 6 6 2 2 4 1 2 7

[126] 7 2 6 3 1 2 4 1 1 1 2 3 7 6 3 6 4 3 3 4 2 4 5 7 1

[151] 2 6 6 3 5 2 2 7 7 7 4 4 1 6 6 3 4 7 7 2 4 1 2 2 1

[176] 1 7 5 2 2 3 4 3 7 7 2 6 7 3 4 2 6 7 6 3 2 2 7 1 5

[201] 6 5 1 3 2 2 2 3 2 6 6 6 2 5 1 7 3 5 3 2 4 1 4 5 3

[226] 4 3 5 2 4 4 6 2 1 3 3 5 2 3 3 5 4 3 1 6 1 2 5 7 1

[251] 6 4 6 7 4 2 6 4 2 7 2 3 7 4 4 6 2 7 4 5 6 4 6 2 4

[276] 4 7 2 6 1 4 6 3 1 7 1 2 7 4 3 3 2 6 6 3 6 1 7 7 7

[301] 5 7 7 4 6 2 3 6 3 7 2 5 2 6 3 4 6 3 4 2 6 7 2 7 4

[326] 1 3 7 1 6 2 5 3 6 4 1 7 4 6 4 5 1 5 7 6 6 6 3 5 7

[351] 4 7 2 2 3 4 6 7 5 3 7 1 2 4 1 1 7 6 6 2 7 1 3 4 2

[376] 5 4 2 6 1 2 7 6 4 3 5 2 5 6 6 7 1 4 3 7 3 4 6 2 1

[401] 6 6 3 5 2 7 3 1 6 6 3 4 7 2 6 3 4 4 7 2 5 1 4 6 2

[426] 1 4 5 3 2 4 4 3 4 2 4 3 2 6 6 7 2 4 1 7 7 6 6 7 4

[451] 6 7 1 1 1 3 7 4 2 6 5 5 7 3 5 4 2 7 7 2 5 1 3 4 6

[476] 3 6 3 4 4 6 7 6 7 4 6 4 5 3 4 7 6 2 1 6 5 5 3 3 6

[501] 4 6 4 2 3 1 6 7 7 1 4 5 2 7 4 7 6 3 2 2 5 4 4 5 3

[526] 6 1 4 7 4 3 5 7 3 2 7 7 4 7 1 2 6 5 6 6 2 5 7 3 1

[551] 4 7 5 7 3 4 2 6 1 7 1 6 3 6 5 7 7 2 4 7 2 7 5 3 5

[576] 6 1 3 4 7 2 1 4 2 5 2 6 6 1 1 2 3 3 2 6 6 7 4 7 2

Within cluster sum of squares by cluster:

[1] 0.09483975 0.27096134 0.14207403 0.26425034

[5] 0.12768515 0.17747241 0.16896953

(between\_SS / total\_SS = 97.1 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

**Overall k= 3,4,5,6,7**

Food Entr Educ Tran Work Hous Othr cluster\_3

1 0.144 0.095 0.190 0.143 0.038 0.346 0.044 3

2 0.107 0.112 0.017 0.205 0.022 0.471 0.066 1

3 0.145 0.020 0.186 0.017 0.091 0.524 0.017 2

4 0.156 0.018 0.205 0.020 0.069 0.512 0.020 2

5 0.193 0.033 0.138 0.042 0.030 0.528 0.036 2

6 0.174 0.022 0.214 0.019 0.075 0.477 0.019 2

cluster\_4 cluster\_5 cluster\_6 cluster\_7

1 3 3 5 4

2 2 2 3 5

3 1 4 2 6

4 1 4 2 6

5 4 1 6 1

6 1 4 2 7

**3.2** **Create the WSS plots as demonstrated in class and select a suitable k value based on the “elbow”. [NOTE – It is easiest to create this chart in Excel or some other spreadsheet program]**

****

**K=5,** is the suitable value based on “Elbow” because between to total value increases and total within SS decreases gradually once the K value reaches 5.

**4 Evaluation of Clusters,**

**4.1 Based on the “k” chosen above, create a scatter plot showing the clusters and colour-coded datapoints for each of “k-1”, “k”, “k+1”. For example, if you think the “elbow” is at k=4 create the charts for k=3, k=4 and k=5.**

**K=5**

****

**K-1 = 4**

****

**K+1 = 6**

****

**4.2 Based on the WSS plot (3.2) and the charts (4.1) choose one set of clusters that best describes the data.**

Based on the WSS plot (3.2) and the charts (4.1), **5** **clusters (K = 5)** will best to describe the data because the clusters are tightly packed and there are some distances between each clusters and cluster centers.

**4.3 Create summary tables for the segmentation/clustering scheme (selected in step 4.2).**

**K = 5**

K-means clustering with 5 clusters of sizes 150, 100, 100, 100, 150

Cluster means:

Food Entr Educ Tran

1 0.1741887 0.18272956 0.02040252 0.39651572

2 0.2617547 0.01839623 0.36537736 0.02145283

3 0.2714151 0.02201887 0.39328302 0.02054717

4 0.2649245 0.15232075 0.32050943 0.26871698

5 0.3622516 0.04051572 0.26286792 0.07623899

Work Hous Othr

1 0.10333333 0.7752201 0.11553459

2 0.13264151 0.9469057 0.02139623

3 0.13726415 0.9027358 0.02066038

4 0.05684906 0.6449057 0.05969811

5 0.05851572 0.9268553 0.04067925

Clustering vector:

[1] 4 1 2 2 5 3 2 4 5 2 3 2 3 5 4 2 3 2 2 5 5 3 3 3 2

[26] 5 5 5 5 3 5 4 4 5 5 4 4 5 2 4 2 2 1 1 3 2 1 1 4 1

[51] 5 1 5 5 3 1 3 1 3 1 5 4 4 3 4 1 4 2 4 5 1 2 5 1 1

[76] 1 4 2 5 5 1 2 3 1 4 4 5 5 5 5 1 1 3 3 4 1 3 5 5 5

[101] 3 3 4 1 3 1 1 5 4 3 1 1 5 4 4 4 1 2 2 1 1 4 5 1 3

[126] 3 1 2 5 5 1 4 5 5 5 1 5 3 2 5 2 4 5 5 4 1 4 1 3 5

[151] 1 2 2 5 1 1 1 3 3 3 4 4 5 2 2 5 4 3 3 1 4 5 1 1 5

[176] 5 3 1 1 1 5 4 5 3 3 1 2 3 5 4 1 2 3 2 5 1 1 3 5 1

[201] 2 1 5 5 1 1 1 5 1 2 2 2 1 1 5 3 5 1 5 1 4 5 4 1 5

[226] 4 5 1 1 4 4 2 1 5 5 5 1 1 5 5 1 4 5 5 2 5 1 1 3 5

[251] 2 4 2 3 4 1 2 4 1 3 1 5 3 4 4 2 1 3 4 1 2 4 2 1 4

[276] 4 3 1 2 5 4 2 5 5 3 5 1 3 4 5 5 1 2 2 5 2 5 3 3 3

[301] 1 3 3 4 2 1 5 2 5 3 1 1 1 2 5 4 2 5 4 1 2 3 1 3 4

[326] 5 5 3 5 2 1 1 5 2 4 5 3 4 2 4 1 5 1 3 2 2 2 5 1 3

[351] 4 3 1 1 5 4 2 3 1 5 3 5 1 4 5 5 3 2 2 1 3 5 5 4 1

[376] 1 4 1 2 5 1 3 2 4 5 1 1 1 2 2 3 5 4 5 3 5 4 2 1 5

[401] 2 2 5 1 1 3 5 5 2 2 5 4 3 1 2 5 4 4 3 1 1 5 4 2 1

[426] 5 4 1 5 1 4 4 5 4 1 4 5 1 2 2 3 1 4 5 3 3 2 2 3 4

[451] 2 3 5 5 5 5 3 4 1 2 1 1 3 5 1 4 1 3 3 1 1 5 5 4 2

[476] 5 2 5 4 4 2 3 2 3 4 2 4 1 5 4 3 2 1 5 2 1 1 5 5 2

[501] 4 2 4 1 5 5 2 3 3 5 4 1 1 3 4 3 2 5 1 1 1 4 4 1 5

[526] 2 5 4 3 4 5 1 3 5 1 3 3 4 3 5 1 2 1 2 2 1 1 3 5 5

[551] 4 3 1 3 5 4 1 2 5 3 5 2 5 2 1 3 3 1 4 3 1 3 1 5 1

[576] 2 5 5 4 3 1 5 4 1 1 1 2 2 5 5 1 5 5 1 2 2 3 4 3 1

Within cluster sum of squares by cluster:

[1] 1.2749212 0.1774724 0.1689695 0.2642503 0.3500783

(between\_SS / total\_SS = 94.8 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

**4.4 Create suitable descriptive names for each cluster.**

**CLUSTER\_3\_KS,**

CLUSTER\_3\_KS <- kmeans(clst\_data\_std\_fun\_KS, iter.max=10, centers=3, nstart=10)

CLUSTER\_3\_KS

**CLUSTER\_4\_KS**

CLUSTER\_4\_KS <- kmeans(clst\_data\_std\_fun\_KS, iter.max=10, centers=4, nstart=10)

CLUSTER\_4\_KS

**CLUSTER\_5\_KS**

CLUSTER\_5\_KS <- kmeans(clst\_data\_std\_fun\_KS, iter.max=10, centers=5, nstart=10)

CLUSTER\_5\_KS

**CLUSTER\_6\_KS**

CLUSTER\_6\_KS <- kmeans(clst\_data\_std\_fun\_KS, iter.max=10, centers=6, nstart=10)

CLUSTER\_6\_KS

**CLUSTER\_7\_KS**

CLUSTER\_7\_KS <- kmeans(clst\_data\_std\_fun\_KS, iter.max=10, centers=7, nstart=10)

CLUSTER\_7\_KS

**4.5 Suggest possible uses for this clustering scheme.**

The clustering used here is used for finding out the amount spent on different things.

**5 Professionalism and Clarity.**