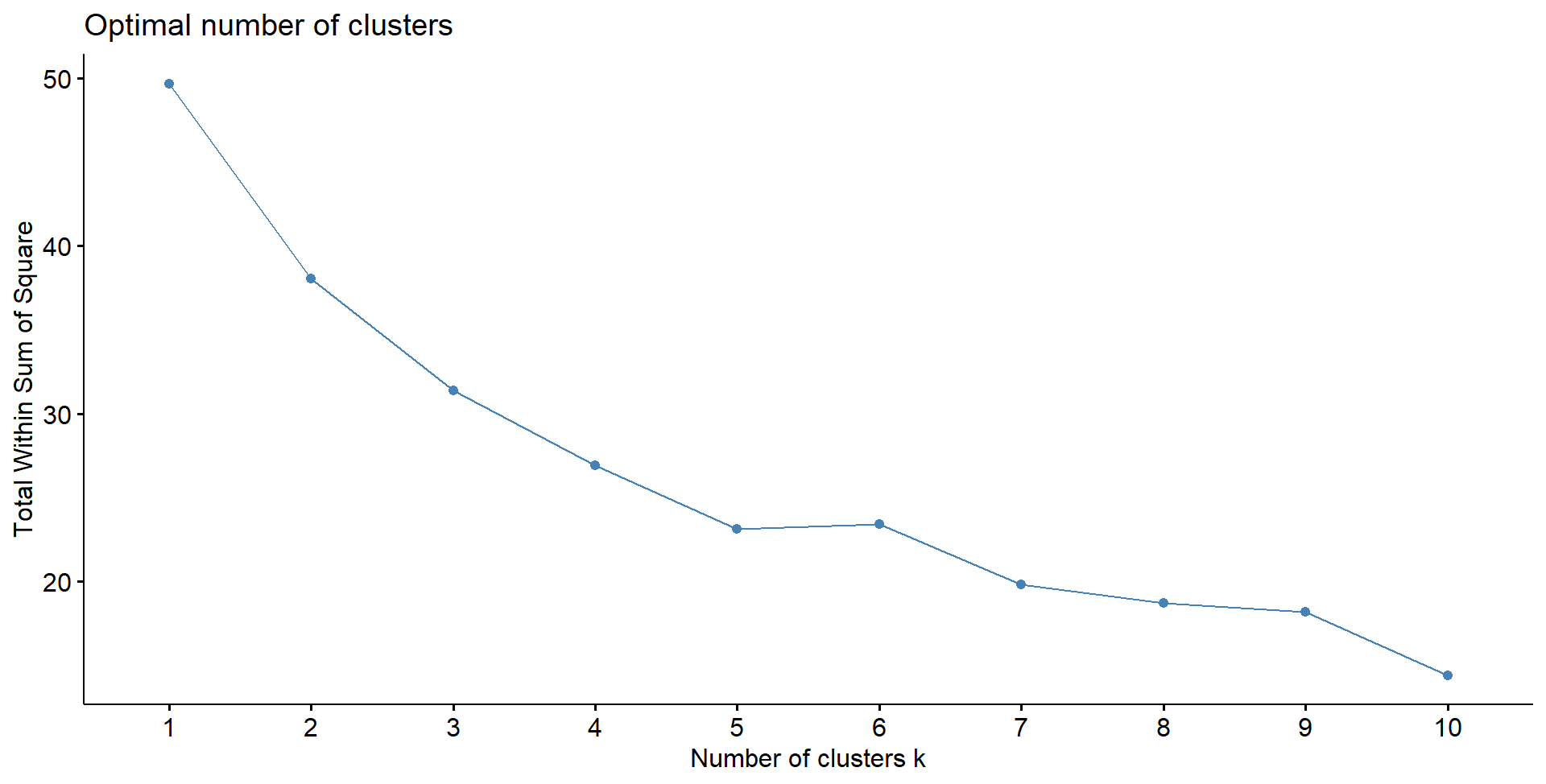
1. **Using all the variables, except name and rating, run the k-means algorithm with k=5**

**to identify clusters within the data.**

Cluster plot when number of clusters k=5

Graphical user interface, chart, scatter chart

Description automatically generated



1. **Develop clustering profiles that clearly describe the characteristics of the cereals within**

**the cluster.**

**K= 5**, Within cluster sum of squares by cluster:

[1] 1.173452 6.335689 9.503646 1.287030 7.522469

(between\_SS / total\_SS = 48.0 %)

Graphical user interface, text, application, email

Description automatically generated

Cluster 1 has shelf 3 and one shelf 2

Cluster 2 has shelf 1 and one shelf 2

Cluster 3 has shelf 3 and one shelf 2

Cluster 4 has a mix of shelf 1 and 2

Cluster 5 has a mix of shelf 1,2 and 3

**Based on shelf and potassium**

Shelf 2 which has potassium >220 belong to cluster 1

Shelf 2 which has potassium between 40-90 belongs to cluster 2,3

Shelf 2 which has potassium between 25-65 belongs to cluster 4

Shelf 2 which has potassium more than 90 belongs to cluster 5

The clusters are formed based on the shelf. The next best differentiator after shelf is potassium level.

1. **Rerun the k-means algorithm with k=3**

Graphical user interface, chart

Description automatically generated

1. **Which clustering solution do you prefer, and why?**

Cluster formed using k-means, where k=5 will be better suited for analysis as the variance between the clusters are higher (48% for K=5) in comparison 3 clusters where the variance between the clusters is lesser (33.8% for K=3)

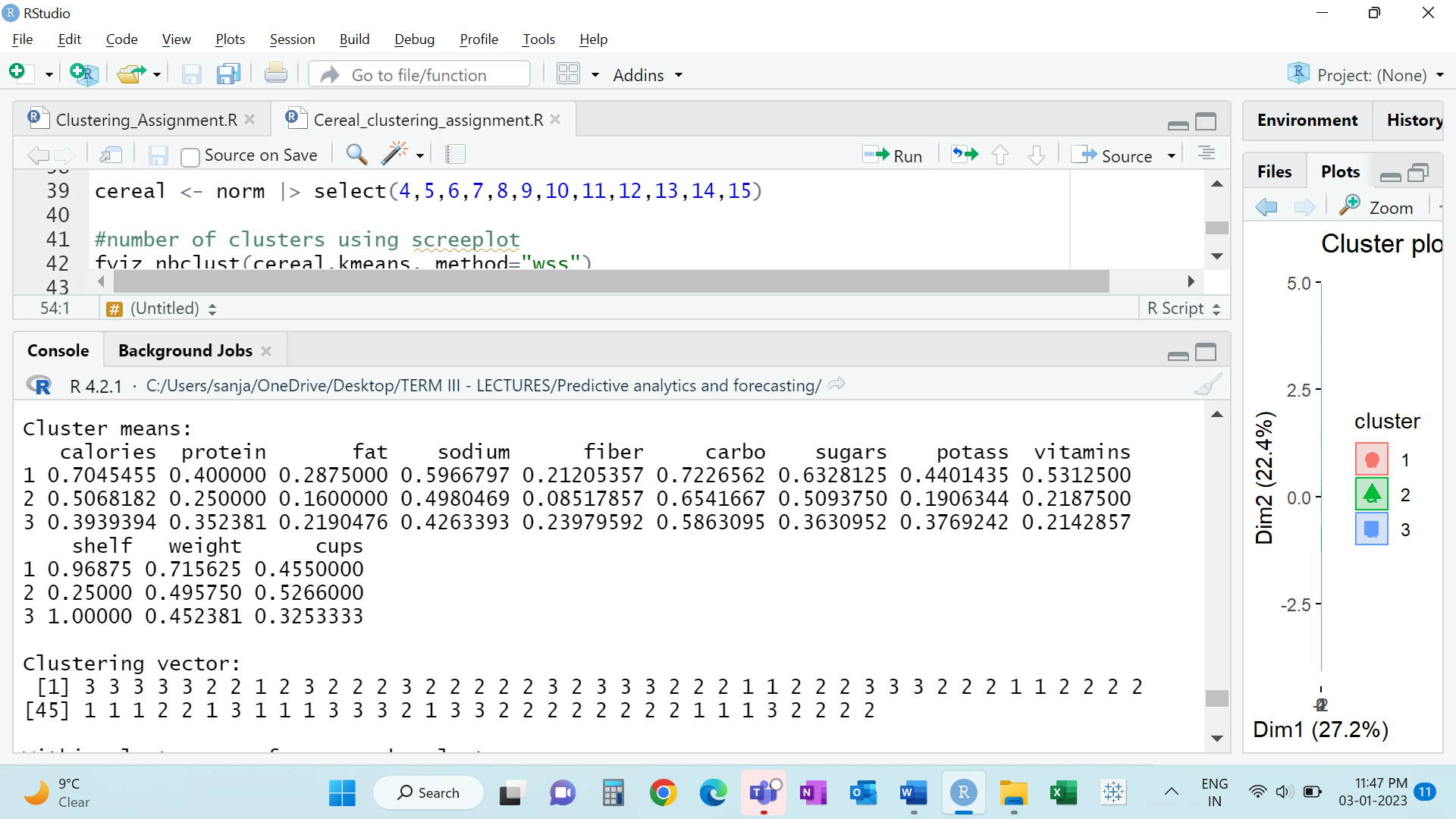
1. **Develop clustering profiles that clearly describe the characteristics of the cereals within**

**the cluster.**

**K=3**, Within cluster sum of squares by cluster:

[1] 6.50960 16.83362 9.52699

(between\_SS / total\_SS = 33.8 %)



Graphical user interface, application

Description automatically generated

Cluster 3 has only shelf 3

Cluster 2 – 50% of Shelf 2 and 50% of Shelf 1

Cluster 1 has shelf 3 and one shelf 2

1. **Use cluster membership to predict rating. One way to do this would be to construct a**

**histogram of rating based on cluster membership alone. Describe how the relationship**

**you uncovered makes sense, based on your earlier profiles.**

Graphical user interface, application

Description automatically generated with medium confidence

**Cereal with average rating of 35.9 belongs to cluster 1**

**Cereal with average rating of 43.9 belongs to cluster 2**

**Cereal s with average rating of 48.7 belongs to cluster 3**

**Cereal with average rating of 27.6 belongs to cluster 4**

**Cereal with average rating of 62.7 belongs to cluster 5**