Sports Product Recommendation system on Young people survey

By:
Gayathri. G.S,
Dataset-young people survey
Kaggle

Exploring the data

- The young people who were surveyed belong to Slovakian nationality.
- The data set is from Comenius University, in Bratislava, capital of Slovakia.
- The university's national rank is #1, and world rank is #637 according to CWUR(Center for World University Rankings) 2016.

Dataset Details

- Music preferences (19 items)
- Movie preferences (12 items)
- Hobbies & interests (32 items)
- Phobias (10 items)
- Health habits (3 items)
- Personality traits, views on life, & opinions (57 items)
- Spending habits (7 items)
- Demographics (10 items)
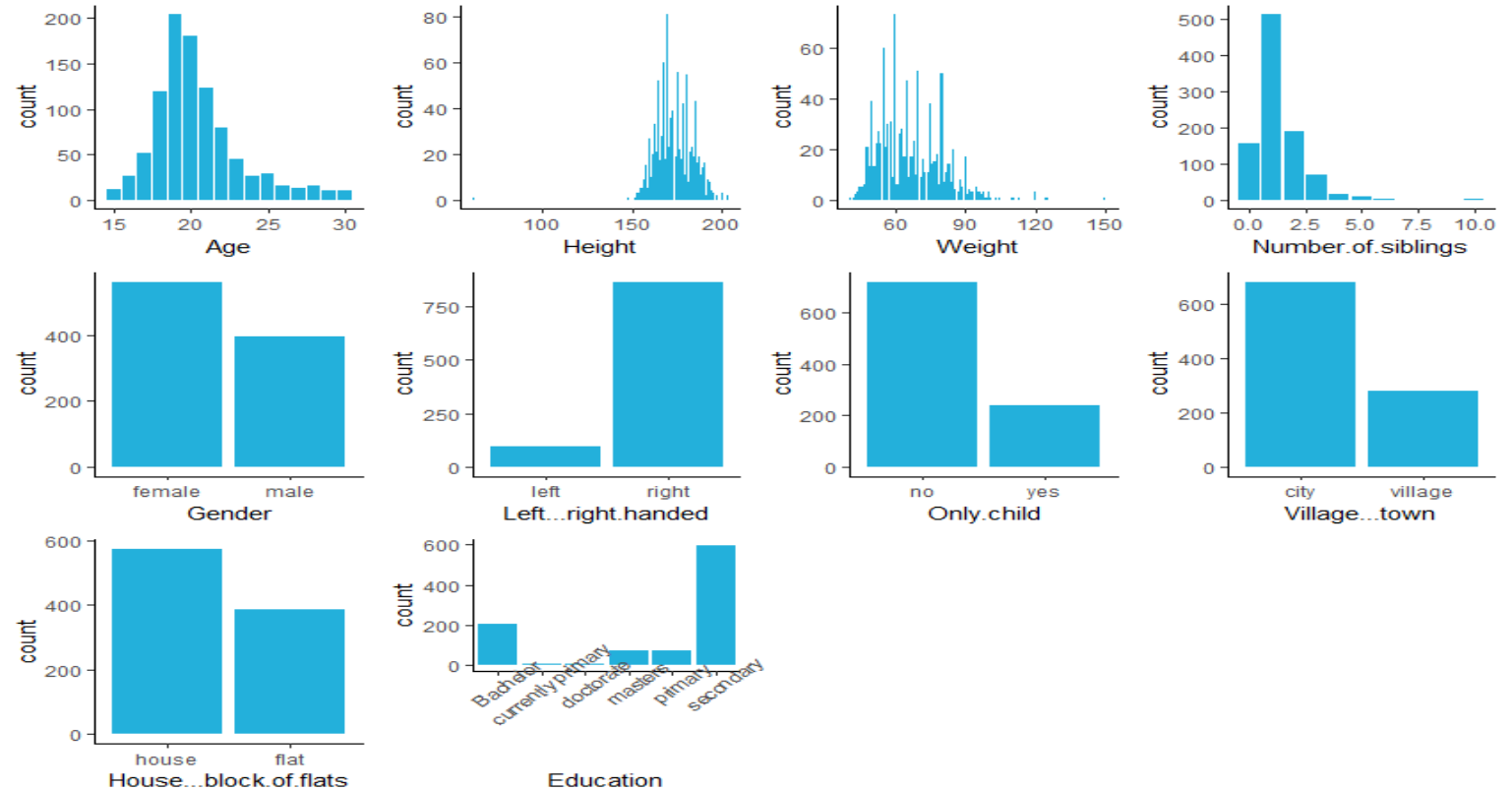
- Young- 15-30,
- City dwellers are twice as much as villagers
- Females are more in numbers
- Most of them are in secondary Education
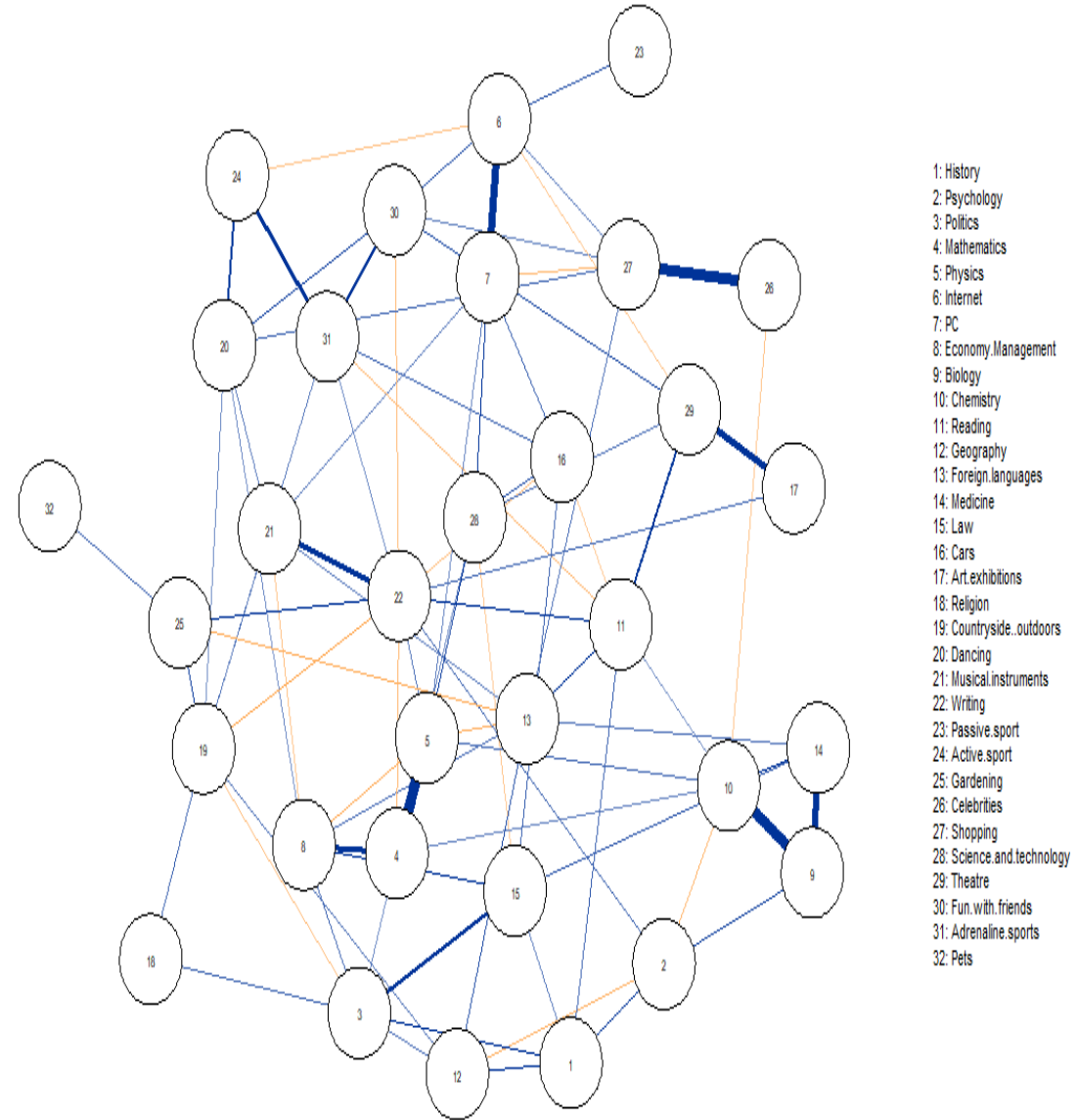
Objective

- Who are the people interested in sports ?
- Which is the right place to Advertise them ?
- If decided to place adds on movies ,what kind of movies to choose ?
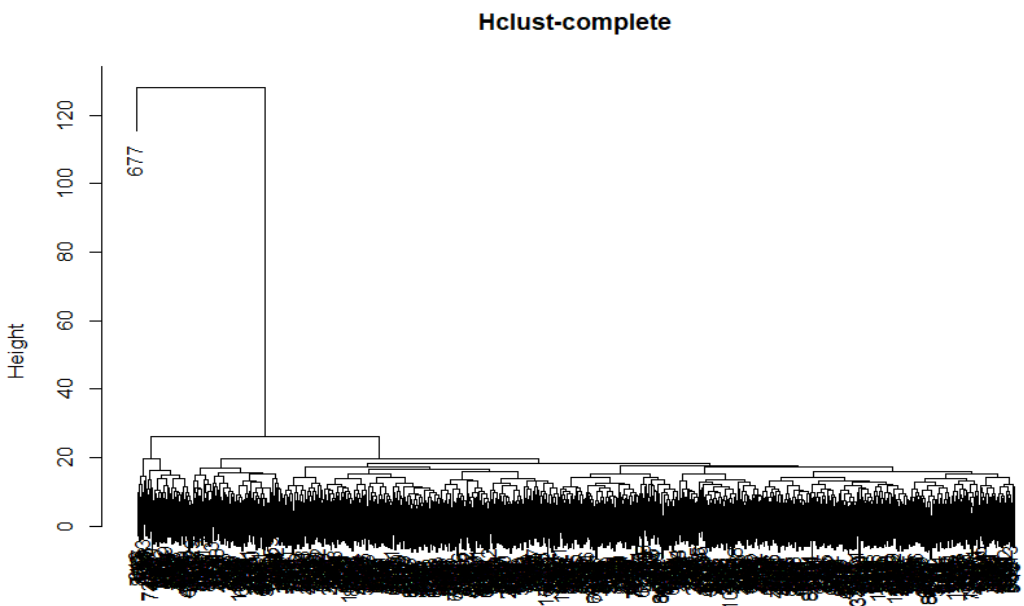
Analytical Methods used

1.Clustering
2.Factoring

Data prep & correlations

- Columns related to Hobbies and interests and demographics are only used.

- Few columns like Gender , Village town etc which had yes/no converted into numeric.

- Height and weight taken as BMI
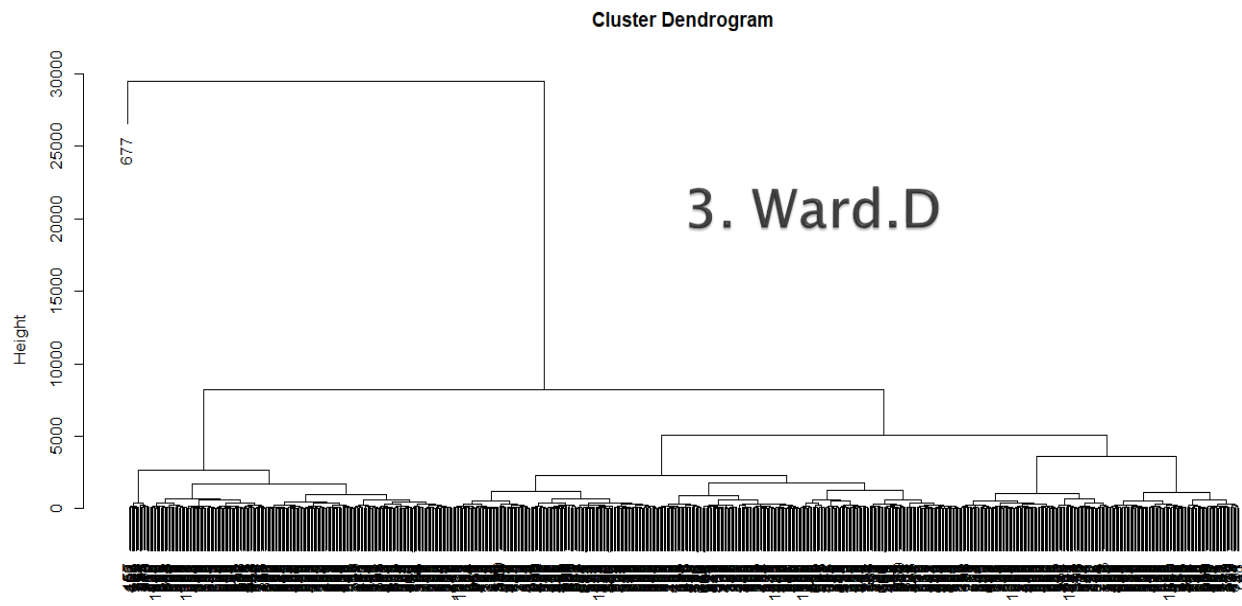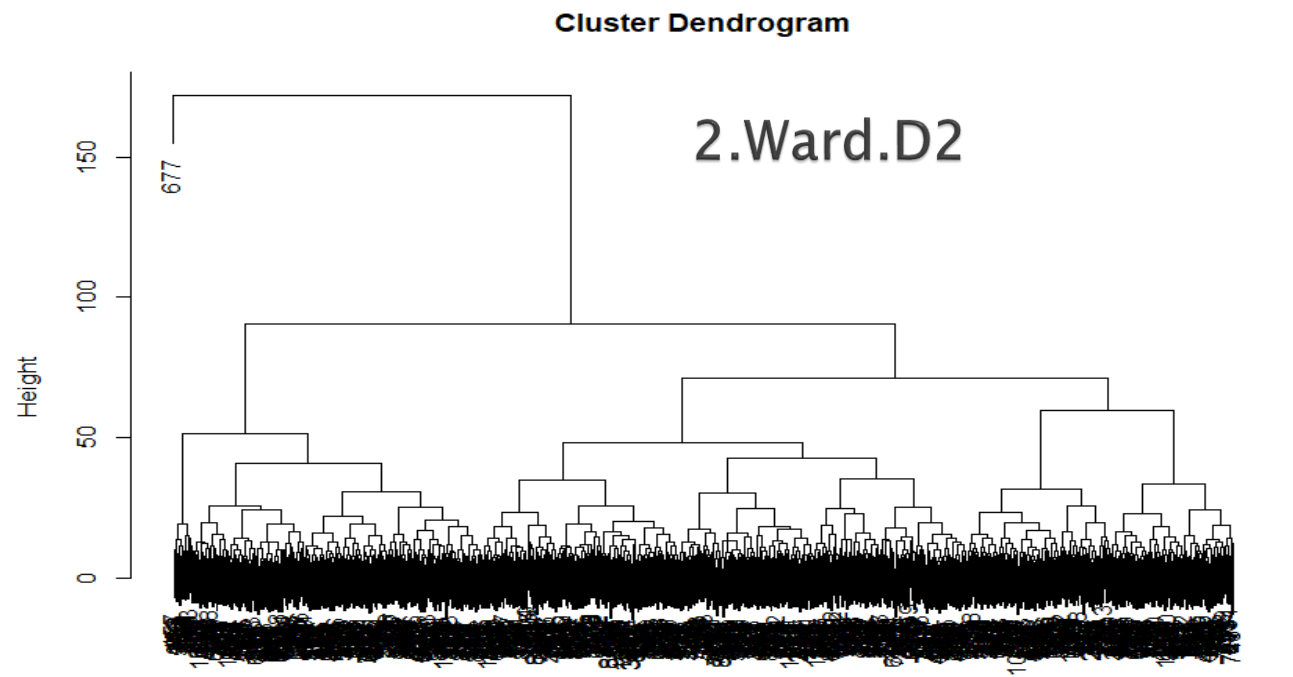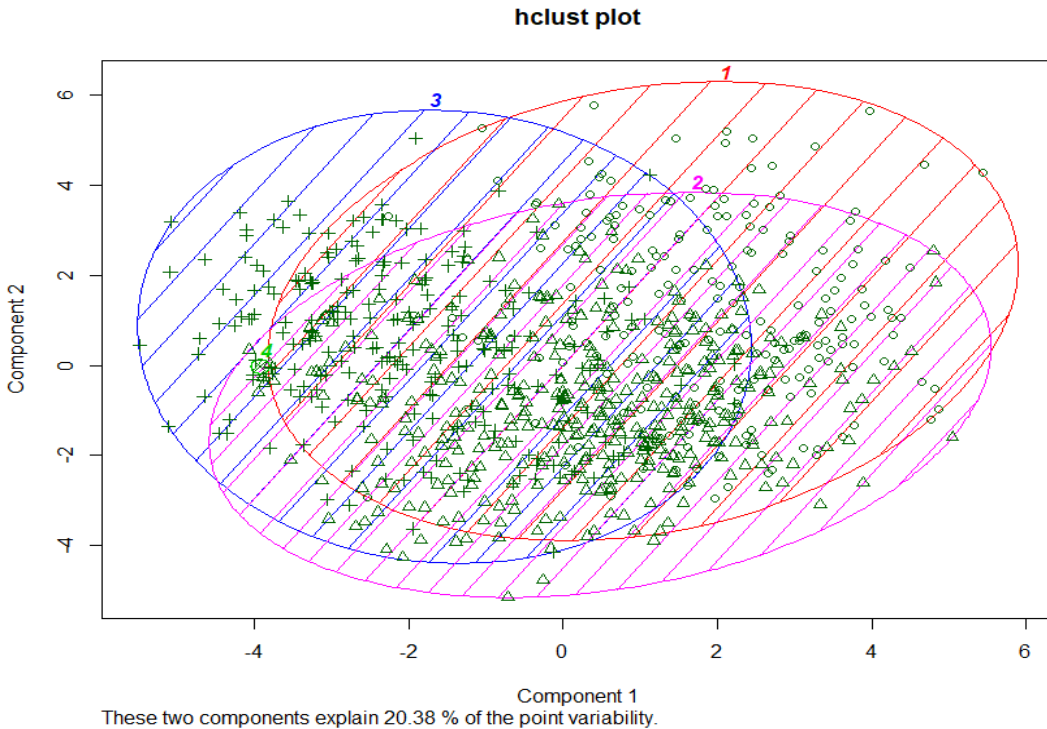
- Binning was done on age and education.

1: History
2: Psychology
3: Politics
4: Mathematics
5: Physics
6: Internet
7: PC
8: Economy.Management
9: Biology
10: Chemistry
11: Reading
12: Geography
13: Foreign.languages
14: Medicine
15: Law
16: Cars
17: Art.exhibitions
18: Religion
19: Countryside..outdoors
20: Dancing
21: Musical.instruments
22: Writing
23: Passive.sport
24: Active.sport
25: Gardening
26: Celebrities
27: Shopping
28: Science.and.technology
29: Theatre
30: Fun.with.friends
31: Adrenaline.sports
32: Pets

Hierarchical-methods

Hclust-complete

disma
hclust (*, "complete")

1.Complete

Cluster Dendrogram

2.Ward.D2

Cluster Dendrogram

3. Ward.D

(disma * disma)

K-means method

hclust plot

These two components explain 20.38 % of the point variability.
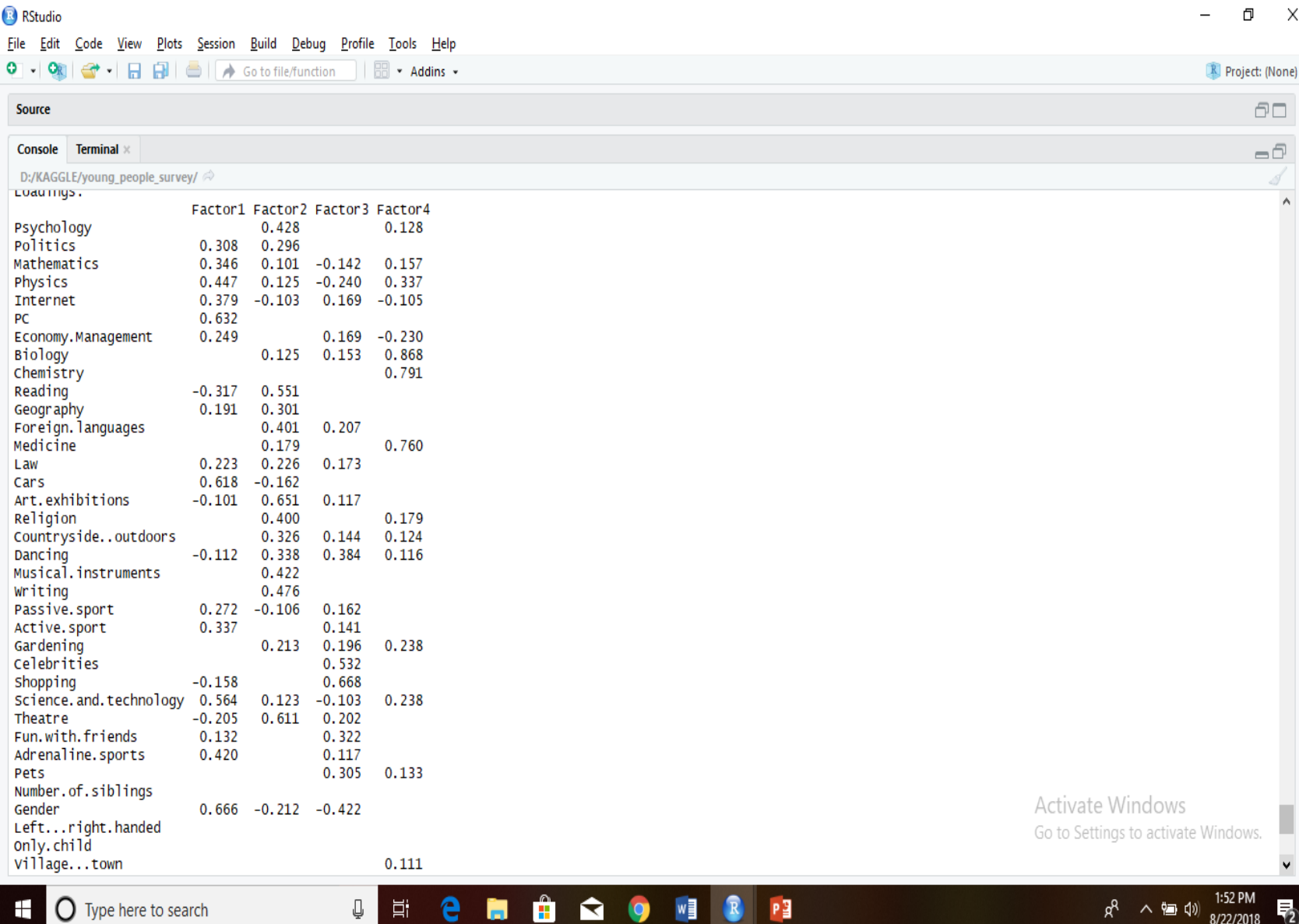
Cluster plot of ward method

Optimal number of clusters

K-means cluster plot

These two components explain 20.2 % of the point variability.
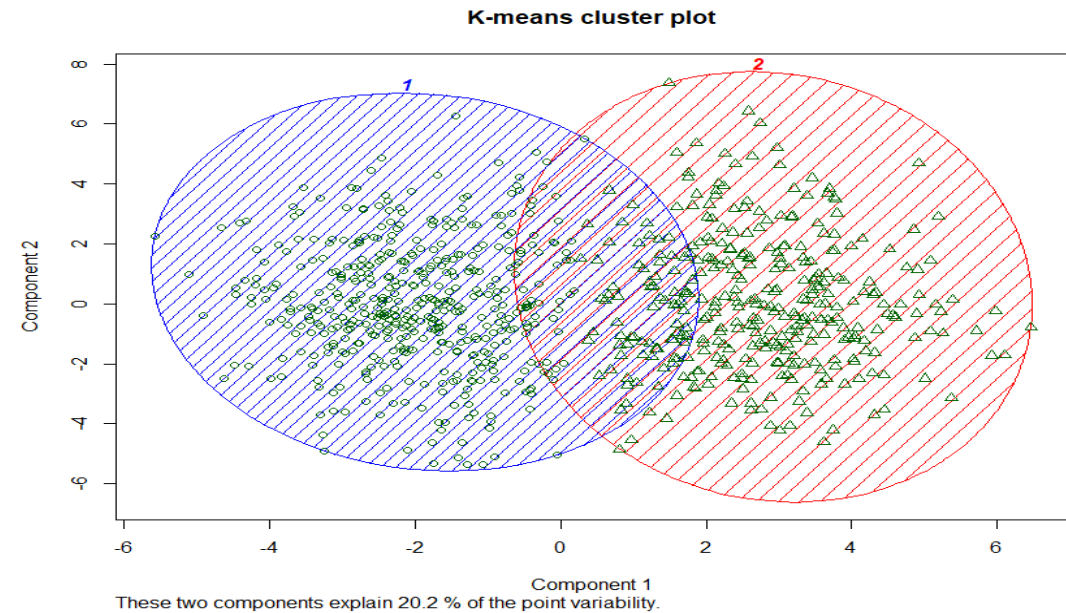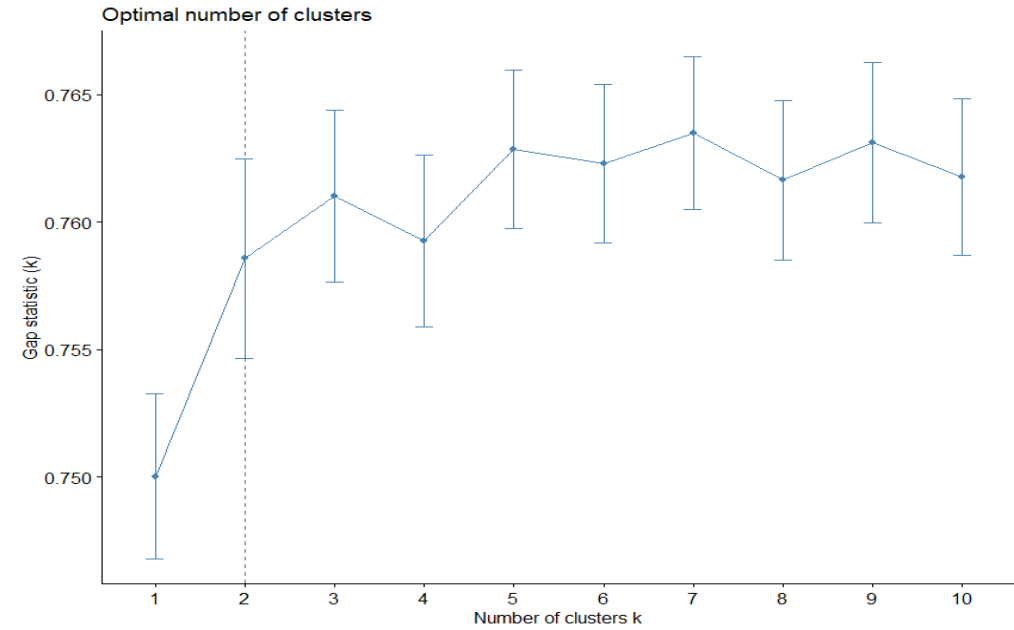
Cluster plot of K-means method

Factors influencing clusters

FACTOR1:

Active sports, adrenaline sports, PC, internet, science and technology, action, thriller, war & western.

FACTOR2:

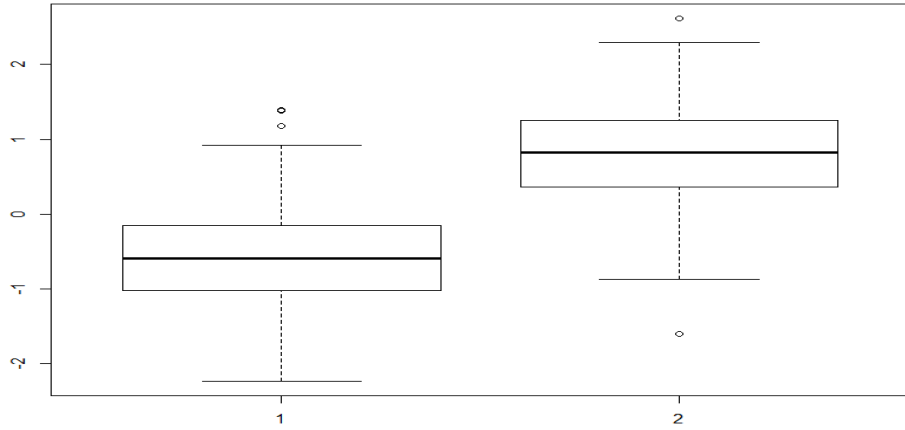Psychology, reading, Art.exhibitions , writing ,theater

FACTOR3:

Celebrities , shopping ,comedy , romantic, fairy tales
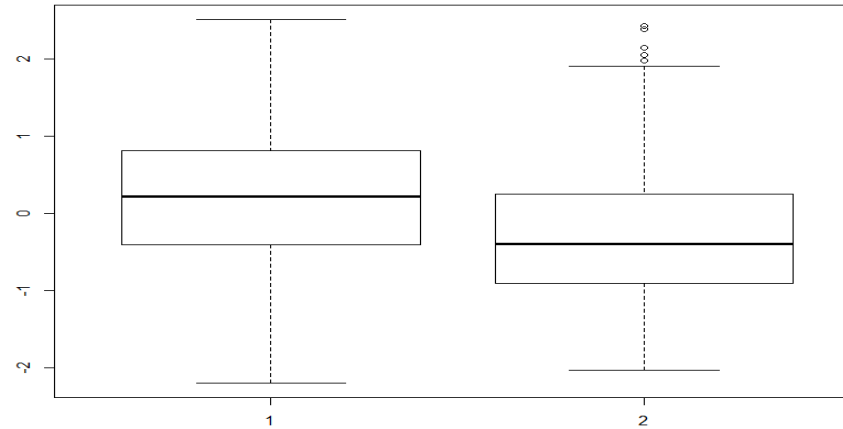
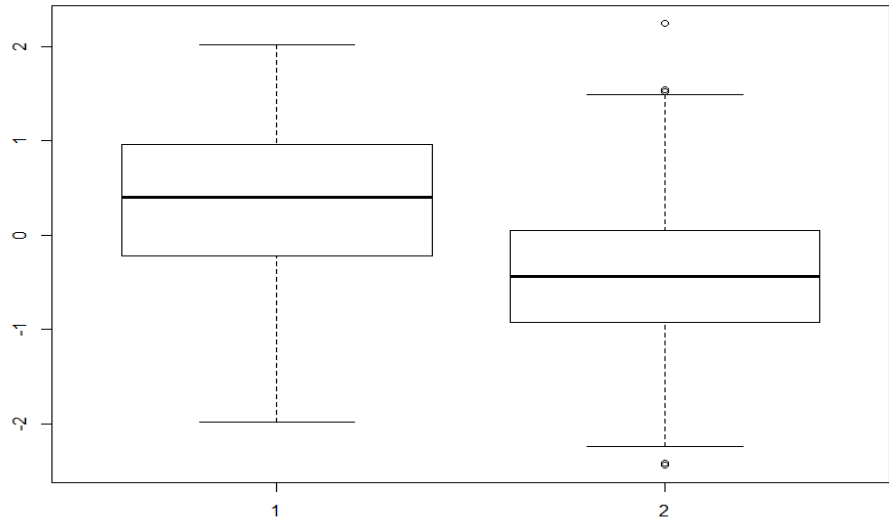**FACTOR4**:

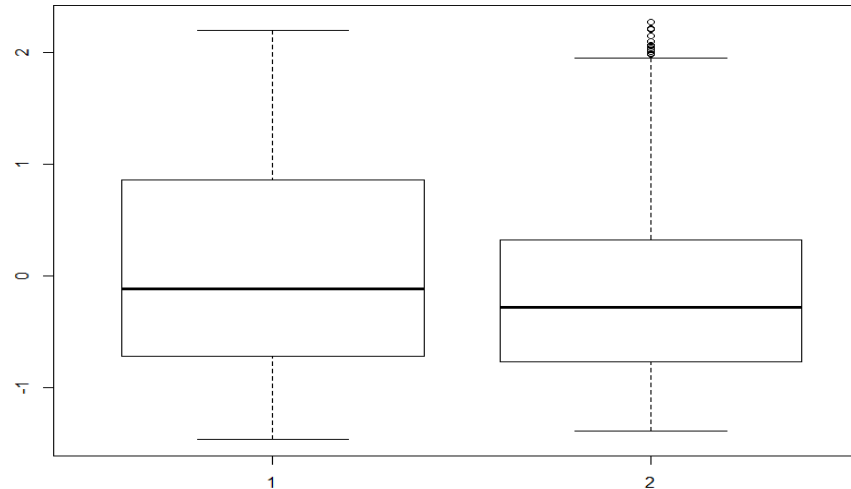Biology,chemistry,Medicine.

Profiling

Cluster2 is mostly influenced by factor1

Cluster1 is mostly influenced by factor2

Cluster 1 again influenced by factor3

Cluster 1 again influenced by factor4

POINTS NOTED

❖ The people who are interested in sports belong to cluster2.

❖ Cluster2 people also like PC , internet , science and technology

❖ They watch Action , thriller, war , western movies.

Summary of findings

- Who the people interested in sports ?
    The people who belong to cluster2 are interested in sports.

- Which is the right place to Advertise them ?
    Cluster2 is highly influenced by factor 1
  - Meaning- people who are interested in sports are also interested in Internet, PC,
  - So u can advertise them on social media , apps like share chat , insta also on the above mentioned movies.

    If decided to place adds on movies ,what kind of movies to choose ?
    Watch Action ,thriller ,war , western movies

  - cluster2 has mostly males.
  - cluster2 belong to secondary education
  - cluster2 people are of age 19-23

- Final insight--Males of age 19-23 in secondary education are interested in sports and u can advertise them on social media, apps like share chat, insta, also the on the movies like action, thriller, war, western.