Assignment 11.1

1. Use the below given data set

Data Set

Answer:

temp<-tempfile()

urlarchive=" https://www.kaggle.com/hugodarwood/epirecipes/data"

download.file(urlarchive,temp)

csvfile=unz(temp," epi\_r.csv")

EPI<-read.csv(csvfile)

EPI

2. Perform the below given activities:

a. Apply PCA to the dataset and show proportion of variance

> km = kmeans(df\_train1,1)

> km$withinss

[1] 1833

> km$tot.withinss

[1] 1833

> km = kmeans(df\_train1,2)

> km$withinss

[1] 600.9150 685.8406

> km$tot.withinss

[1] 1286.756

> km = kmeans(df\_train1,3)

> km$withinss

[1] 346.0231 336.7207 298.1925

> km$tot.withinss

[1] 980.9363

> km = kmeans(df\_train1,4)

> km$withinss

[1] 298.1925 216.7761 169.1202 231.8171

> km$tot.withinss

[1] 915.9058

> km = kmeans(df\_train1,4)

> km$withinss

[1] 90.7904 298.1925 211.4158 301.2184

> km$tot.withinss

[1] 901.6171

>

> km = kmeans(df\_train1,5)

> km$withinss

[1] 94.49965 165.73861 168.75390 309.23224 112.04907

> km$tot.withinss

[1] 850.2735

>

> km = kmeans(df\_train1,6)

> km$withinss

[1] 84.21473 86.00161 153.99935 78.06204 298.19248 94.11996

> km$tot.withinss

[1] 794.5902

>

> km = kmeans(df\_train1,7)

> km$withinss

[1] 69.91732 151.56160 94.11996 153.99935 86.00161 84.21473 94.49965

> km$tot.withinss

[1] 734.3142

>

> km = kmeans(df\_train1,8)

> km$withinss

[1] 120.03869 76.66393 94.11996 118.50099 55.00107 69.91732 78.78610 86.00161

> km$tot.withinss

[1] 699.0297

>

> km = kmeans(df\_train1,9)

> km$withinss

[1] 41.54627 86.00161 87.40893 58.01375 78.78610 78.00943 109.30016 75.27350

[9] 57.36509

> km$tot.withinss

[1] 671.7048

>

> km = kmeans(df\_train1,10)

> km$withinss

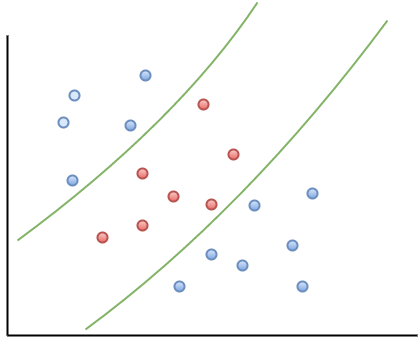
[1] 33.42481 148.18838 43.72160 72.07943 43.90180 57.36509 84.21473 54.41516

[9] 34.73034 71.15537

> km$tot.withinss

[1] 643.1967

b. Perform PCA using SVD approach



c. Show the graphs of PCA components

