Results and Interpretation

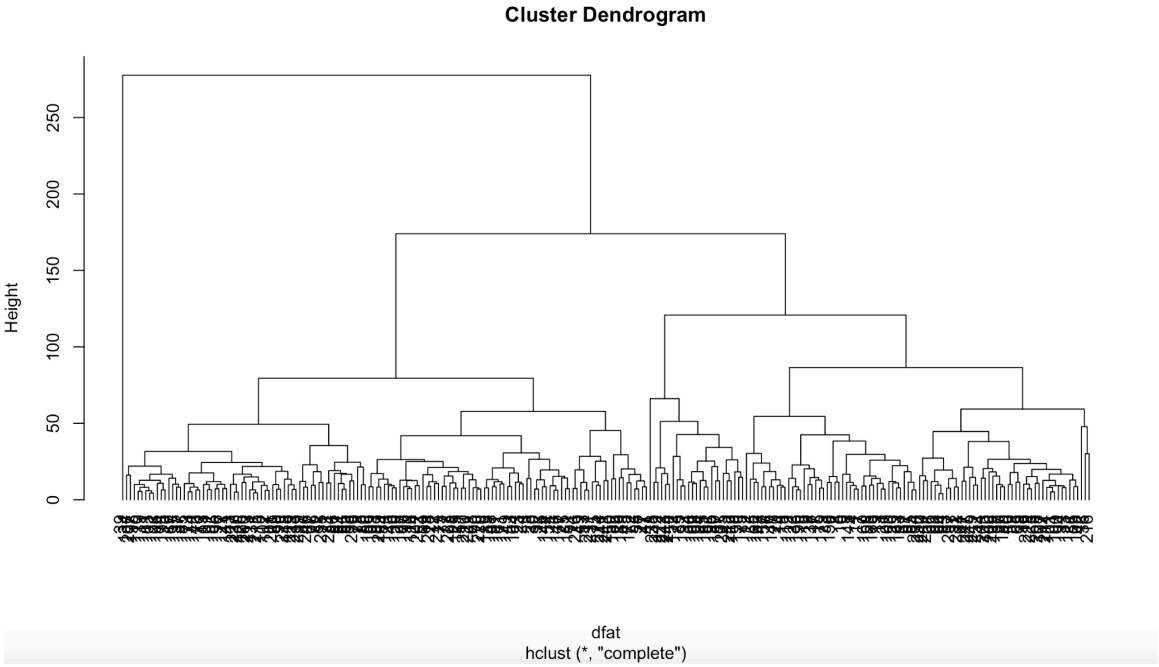

Diagram-1 “complete” cluster of body fat

```
> table(a,b)
   b
a     1   2   3   4
  1 136   0   0   0
  2   0  90  25   0
  3   0   0   0   1
```

Diagram-2 cut tree table data

If we cut two lines of height 150 and height 100 in Diagram-1, basically we can get 3 and 4 clusters, then using cut tree function to compare the difference between cluster membership-a and membership-b, Diagram-2 shows that the only change is the second cluster in membership-a is divided into two clusters and became the second and third cluster in membership-b.

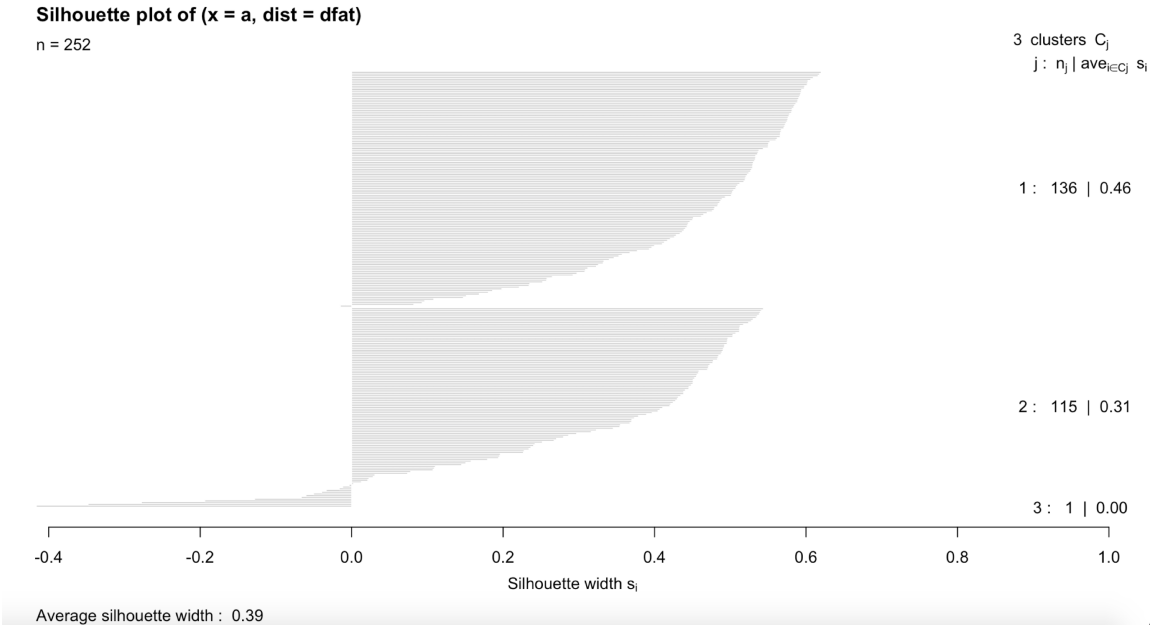

Diagram-3

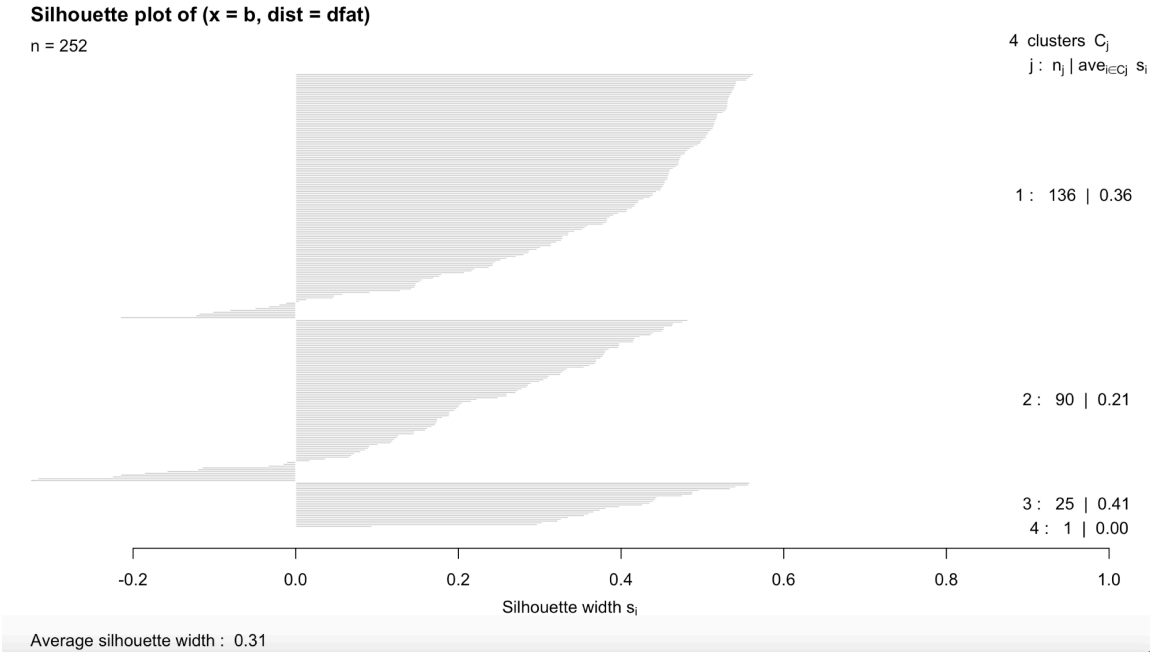

Diagram-4

Diagram-3 and Diagram-4 are the silhouette plot of membership-a and membership-b. when the cluster only have one point, the value of S(i) is defined as 0, so from the diagrams, we can see the third cluster in membership-a and the fourth cluster in membership-b's S(i) value are both 0, because there is only one observation in these clusters. When the value of silhouette width more close to 1, it means the points in that cluster are more close to each other,

negative value means that the outliers in that cluster.

```
> kfat
K-means clustering with 3 clusters of sizes 89, 55, 108
```

Diagram-5

```
> summary(silfat)
Silhouette of 252 units in 3 clusters from silhouette.default(x = kfat$cluster, dist = dfat) :
 Cluster sizes and average silhouette widths:
       89        55       108
0.3392535 0.2528142 0.3202237
Individual silhouette widths:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.07601  0.19640  0.34690  0.31220  0.44190  0.54380
```

Diagram-6

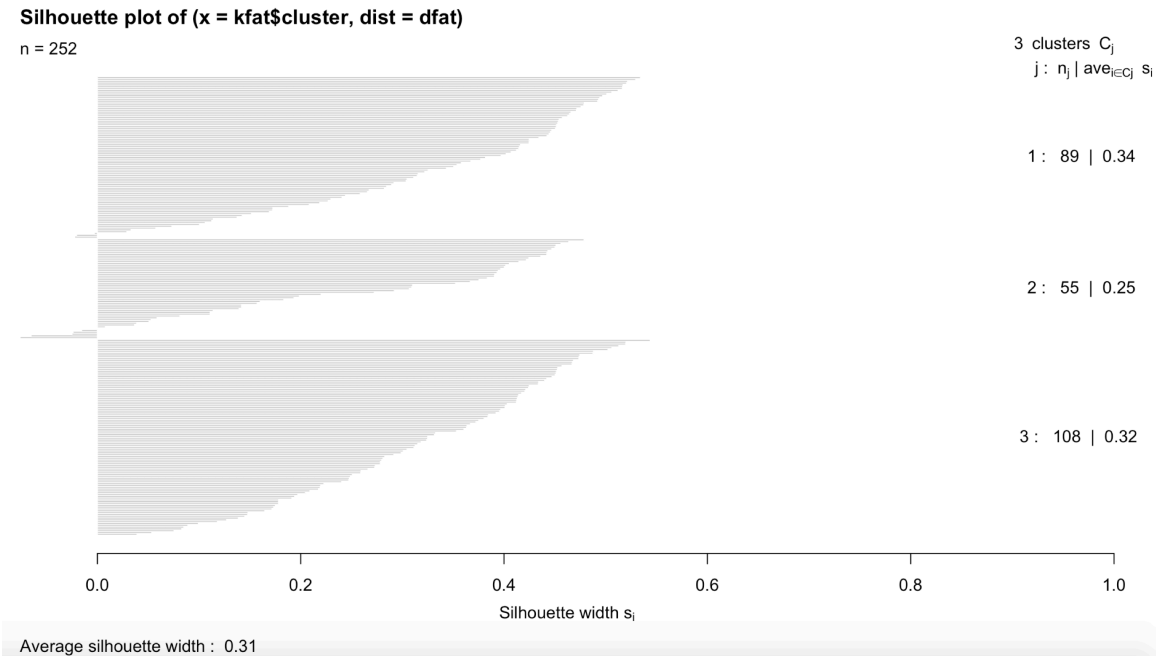

Diagram-7

Diagram-5, D-6, and D-7 show the results of K-means cluster. It divided the data into three clusters, respectively contains 89, 55 and 108 points in each cluster. It is quite different with H-cluster analysis. K-means cluster the data from high to low, but H-cluster cluster the data from low to high. I think that is the reason why these two cluster functions have different cluster results with the same data.

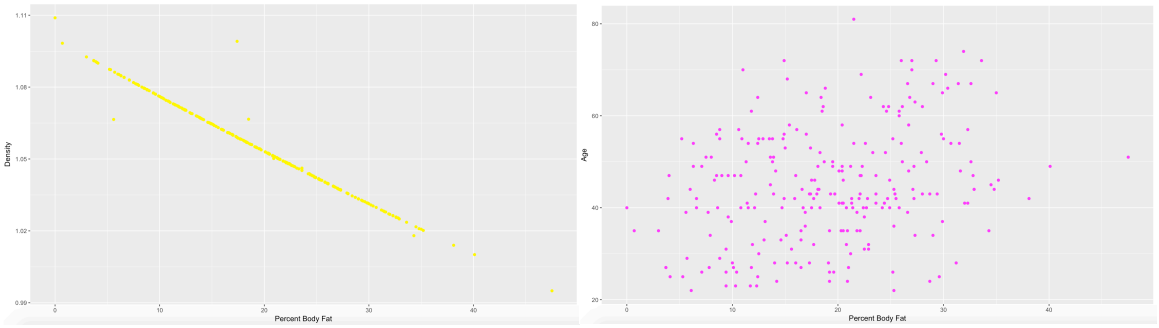

Diagram-8

Because the fat percentage is calculated by the body density, from the yellow line in Diagram-8, we can see that they are perfectly linear related, then I tested the relationship between age and body fat, the result shows that they are random and do not have apparent linear relationship. So age won't be someone's excuse for high body fat percentage, every age group should do workouts hard to keep their body fat percentage in a normal level. Below is the table from the American Council on Exercise shows how average percentages differ according to the specified groups and categories.[i]

Description	Women	Men
Essential fat	10–15%	0–5%
Very lean	15–20%	5–15%
Lean	20–25%	15%
Normal	25–30%	20%
Overweight	30–40%	25–30%
Obese	40% or more	30% or more

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.54803   17.43111  -1.121   0.2632
weight       -0.10974    0.05266  -2.084   0.0382 *
height       -0.09410    0.09569  -0.983   0.3264
neck         -0.42995    0.23280  -1.847   0.0660 .
chest        -0.01728    0.09964  -0.173   0.8625
forearm       0.38964    0.19756   1.972   0.0497 *
wrist        -1.27227    0.50602  -2.514   0.0126 *
abdomen       1.02953    0.07761  13.266   <2e-16 ***
biceps        0.20538    0.17163   1.197   0.2326
hip          -0.22995    0.14626  -1.572   0.1172
thigh         0.13476    0.13510   0.997   0.3195
knee          0.13187    0.23554   0.560   0.5761
ankle         0.12974    0.22150   0.586   0.5586
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> lm.beta(regfat)
    weight      height        neck       chest     forearm       wrist     abdomen
-0.38539409 -0.04118607 -0.12488912 -0.01740807  0.09408035 -0.14192926  1.32654670
    biceps         hip       thigh        knee       ankle
 0.07414761 -0.19684611  0.08453970  0.03800521  0.02627674
```

Diagram-9

Diagram-9 shows the result of multiple linear regression on body fat percentage with various body circumference. Except age and density, abdomen is the most significant variable on fat percent, which means holding other variables fixed, 1cm increase in abdomen, the percentage of body fat will increase 1.02%. for beta coefficients, if we increase abdomen by 1 standard deviation, the body fat percent will increase by 1.33 standard deviation.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.25488    2.36172  -0.955 0.339699
hip          0.17308    0.05043   3.432 0.000599 ***
thigh       -0.35955    0.07126  -5.045 4.53e-07 ***
knee         0.23205    0.11014   2.107 0.035129 *
ankle       -0.12210    0.10185  -1.199 0.230593
```

```
> new<-data.frame(hip=94,thigh=57,knee=42,ankle=20)
> predict(logitfat, new, type="response")
        1
0.6951334
```

```
> exp(coef(logitfat))
(Intercept)         hip       thigh        knee       ankle
  0.1048861   1.1889589   0.6979925   1.2611769   0.8850618
```

Diagram-11

Diagram-11 shows the result of logit analysis, I logit the lower body circumference parameters on class, and predicted a new point (hip=94cm, thigh=57cm, knee=42cm, ankle=20cm) has 69% probability belongs to class 1, then I did the logit odds, holding the other variables constant, as hip circumference increase 1 cm, the odds of class increased by 1.189.

Summary

In this dataset, except density, other variables do not have a clear relationship with body fat percentage, which means we cannot modify our body fat percentage to normal level by exercise specific parts of our body. We should do full body workouts and different cardio exercise to keep us being fitness.

---

[i] ACE (2009) What are the guidelines for percentage of body fat loss? American Council on Exercise (ACE). Ask the Expert Blog. December 2, 2009.