

STT 811

In-Class Assignment 18

This problem will use the Heart dataset, with no target.

(1) Scale the numerical data (not including y)

```
heart <- read.csv("data/Heart.csv")  
heart <- na.omit(heart)  
heart_sc <- scale(heart[,c(2,3,5:13)])
```

(2) Use the kmeans function in R to create 3 clusters. Look at the results.

```
heart_km <- kmeans(heart_sc, centers = 3, nstart = 25)
```

(3) Repeat (2) a couple times. Are the clusters consistent?

```
heart_km$size
```

```
149 105 43  
149 105 43  
149 43 105  
105 149 43
```

Clusters are consistent, just ordered different.

(4) Repeat (2) and (3) for different numbers of clusters. Is there a value for which the clusters are consistent?

Three and four clusters were consistent, but five+ was not