Kmeans clustering analysis

Helton et.al., 2018

Install package dependencies

```
library(amap)
library(ggfortify)
library(ggplot2)
library(misc3d)
library(rgl)
library(heatmap3)
library(jamba)
```

Import morphometric data and scale data per column

```
Data <- read.table("CA2_Morphometric_Data.txt", sep = "\t", header=TRUE)
Data_scaled <- apply(Data[,-1], 2, function(i){i1<- scale(i, center=FALSE);
i2 <-i1- mean(i1);})
row.names(Data_scaled) <- makeNames(Data$Cell_Number)
```

Visualize scaled data and cluster into 3 groups using rank-based distance

```
Hm <- heatmap.3(Data_scaled, farbeLim=2, colLensFactor=5, RowvCutreeK = 3,
distMethod="pearson",
                hclustfun = hclust, margins=c(15, 4), cexCol=0.8, cexRow=0.8,
lhei=c(2.0,4.0),
                lwid=c(1.5,4.0), keysize=0.65, cex.key.main = 0.65);
```
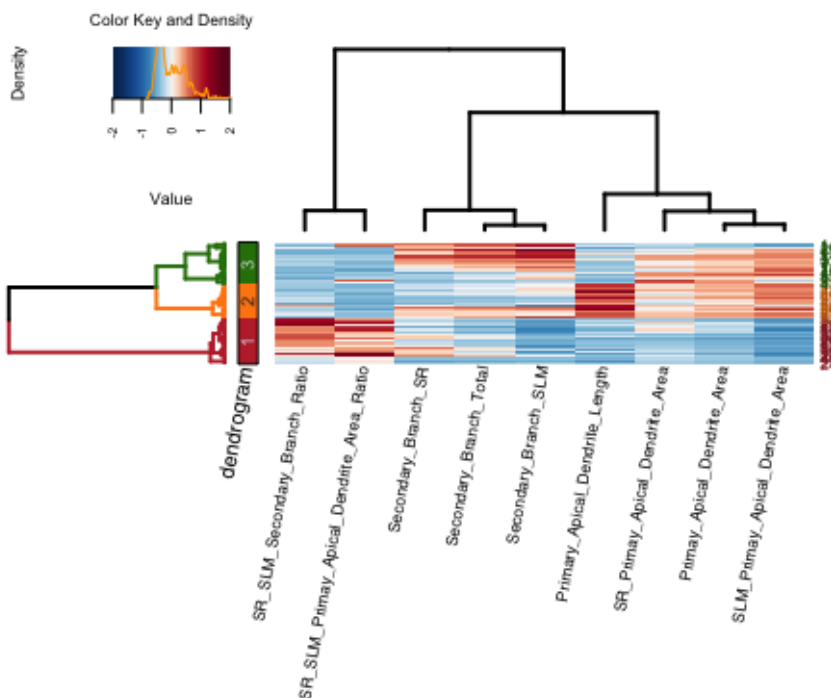
Run Kmeans analysis

```
Data_scaled_df<- as.data.frame(Data_scaled);
Km <- Kmeans(Data_scaled_df[,c(1,4,5)], centers=3, nstart=20,
method="pearson")
Km

## K-means clustering with 3 clusters of sizes 16, 15, 24
##
## Cluster means:
##   Primary_Apical_Dendrite_Length Secondary_Branch_SLM
## 1                          0.848               0.0461
## 2                         -0.269               0.6811
## 3                         -0.397              -0.4564
##   SR_SLM_Secondary_Branch_Ratio
## 1                        -0.280
## 2                        -0.395
## 3                         0.433
##
## Clustering vector:
##   7 30   5 42 16 32 19 44 45   4 20 36 27 35 21   3 12 18 37 14 26 29 13 23 10
##   2  3   2  1  2  3  3  1  1   3  3  3  3  3  3   2  2  3  3  2  3  3  2  3  2
## 28 50   1 15 51 17   6 55 39 31 52 46 11 33 38 34 24 54 43 25   9 41   8  2 48
##   3  1   2  2  1  2   2  1  3  3  1  1  2  3  1  3  3  1  1  3  2  1  2  3  1
## 49 40 47 53 22
##   1  3  1  1  3
##
## Within cluster sum of squares by cluster:
## [1] 0.000405 0.058811 0.040903
##
## Available components:
##
## [1] "cluster"  "centers"  "withinss" "size"
```

## Add Kmeans cluster column to color plots

```
Data_scaled_df$Cluster <- Km$cluster;
levels(Km$cluster) <- c("1","2","3")
col <- c("#FF8711FF", "#BF2D37FF" , "#2C7E0AFF")
table(Km$cluster, Data_scaled_df$Cluster)

##
##      1  2  3
##   1 16  0  0
##   2  0 15  0
##   3  0  0 24
```

Plot some parameters

```
plot1<- plot(Data_scaled_df$Primary_Apical_Dendrite_Length,
             Data_scaled_df$SR_SLM_Secondary_Branch_Ratio,
             xlab = "Primary_Apical_Dendrite_Length", ylab =
"SR_SLM_Secondary_Branch_Ratio",
```

```
        col=group2colors(Data_scaled_df$Cluster));
legend("topright", legend = paste("cluster", levels(Km$cluster)),  col= col,
pch = 1)
```
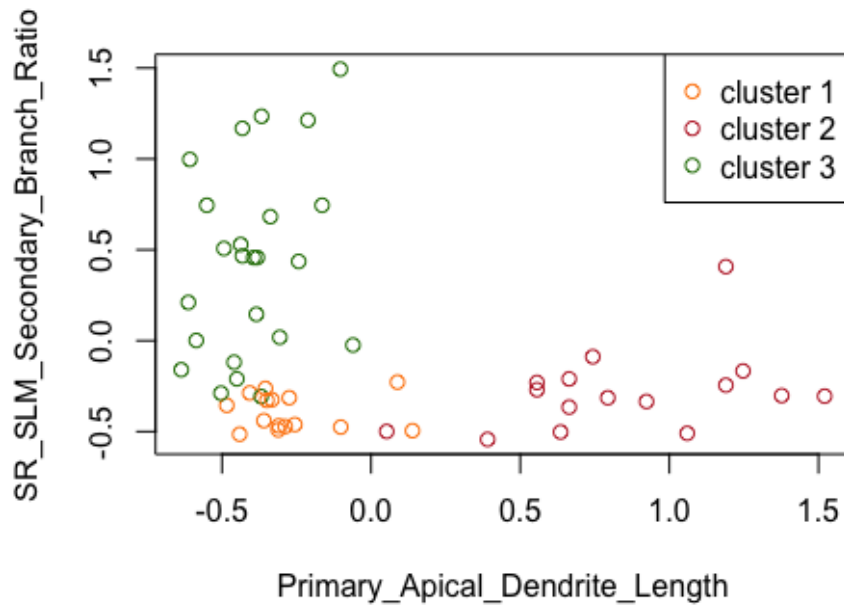


```
plot2 <- plot(Data_scaled_df$Primary_Apical_Dendrite_Length,
Data_scaled_df$Secondary_Branch_SLM,
           xlab="Primary_Apical_Dendrite_Length",
ylab="Secondary_Branch_SLM",
           col= group2colors(Data_scaled_df$Cluster));
           legend("topright", legend = paste("cluster",
levels(Km$cluster)), col= col, pch = 1)
```
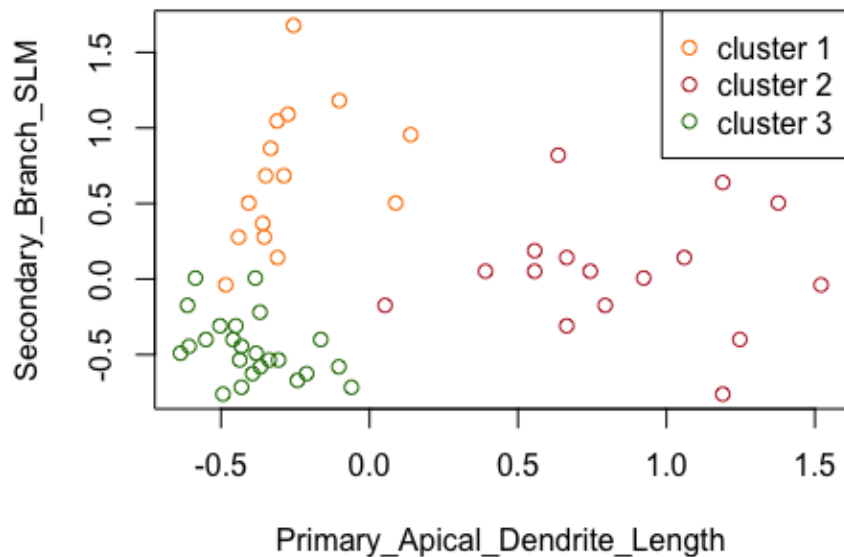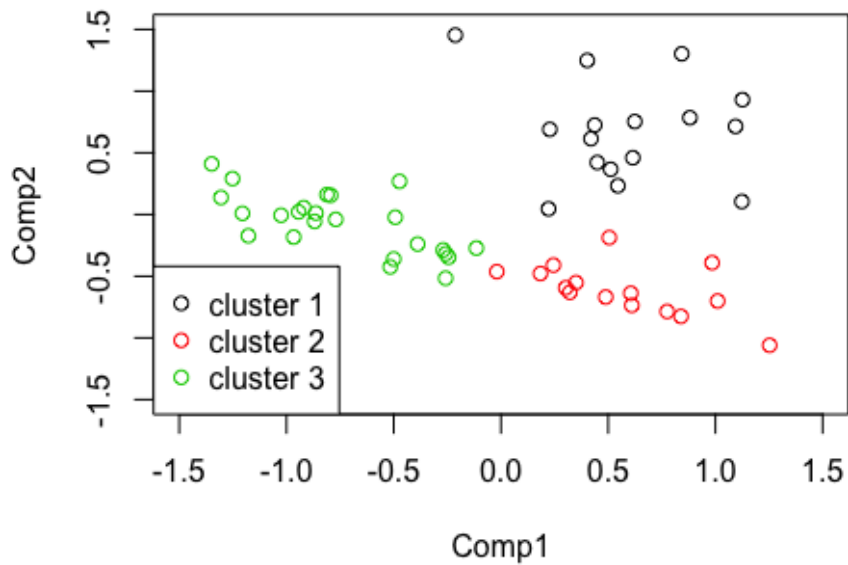
PCA

```
pca <- pca(Data_scaled_df[,c(1,4,5)], center=TRUE, reduce=FALSE);
```
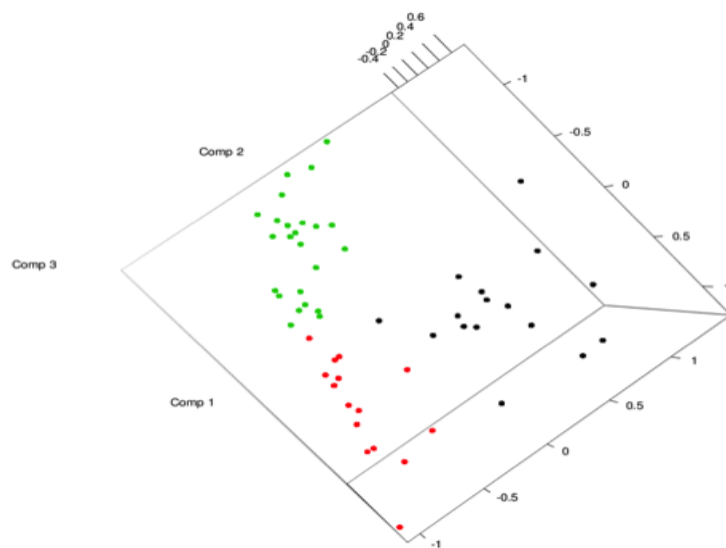
Plot PCA colored by Kmeans cluster

```
scores<- pca$scores
plot(scores[,1:2], col=Km$cluster, xlab = "Comp1", ylab = "Comp2",
     xlim=c(-1.5,1.5), ylim=c(-1.5,1.5));
     legend("bottomleft", col = levels(Km$cluster), legend = paste("cluster",
levels(Km$cluster)),
     pch = 1, cex = 1.0);
```


3D plot PCA with components colored by Kmeans cluster

```
plot3d(scores[,1:3], col = Km$cluster, size=7);
```

Session Info

- R version 3.4.2 (2017-09-28)
- Platform: x86_64-apple-darwin15.6.0 (64-bit)
- Running under: OS X El Capitan 10.11.6

Attached base packages	Other attached packages	
stats	heatmap3_1.1.1	jamba_0.0.5.900
graphics	lubridate_1.7.2	rgl_0.98.1
utils	colorspace_1.3-2	misc3d_0.8-4
methods	dendextend_1.6.0	ggfortify_0.4.1
grDevices	matrixStats_0.52.2	ggplot2_2.2.1
datasets	gtools_3.5.0	amap_0.8-14
base	RColorBrewer_1.1-2	fastcluster_1.1.24