

PCA.R

rob_c

Mon Apr 24 13:57:30 2017

```
# SUMMARY:
#### Al-Yotm distinct from Dead Sea and Kufranjah valley sites
#### Dead Sea and Kufranjah are different ON AVERAGE, but a lot of overlap

# ANALYSIS:
## Libraries
library(ggplot2)
library(ggfortify)
```

```
## Warning: package 'ggfortify' was built under R version 3.3.3
```

```
## Import Data
MorphData<-read.csv("MuhaidatEtAl_RawData.csv",header=T)
str(MorphData)
```

```
## 'data.frame':    57 obs. of  15 variables:
## $ Loc           : Factor w/ 3 levels "AlYotm","DeadSea",...: 3 3 3 3 3 3 3 3 3 3
## $ BLen          : num  4.8 5 5.4 5.6 5.5 5.2 5.9 5.7 4.9 5.9 ...
## $ BWidth        : num  0.4 0.5 0.4 0.4 0.4 0.5 0.4 0.5 0.4 0.5 ...
## $ BVeins         : int   3 5 NA 3 3 3 3 5 3 5 ...
## $ BSpines        : int   2 2 2 2 2 2 3 2 2 2 ...
## $ SpLen          : num   1 1 1 1.2 1.1 0.9 0.9 1.3 1 1.5 ...
## $ NodeLen        : num   1.3 1.6 1.5 1.6 1.3 1.8 1.8 1.6 1.6 2 ...
## $ LLen           : num    7 7 9 8.5 7.5 10 8 8 9.1 10 ...
## $ LWidth         : num    1 1 1.1 0.9 0.7 1.2 1 1.2 0.8 1.2 ...
## $ LTeeth         : int    4 5 4 4 5 3 4 NA 5 4 ...
## $ ASFillLen      : num    1.3 1.3 1.2 1.3 1.2 1.3 1.3 1.2 1.3 1.3 ...
## $ ASAnthLen      : num    0.5 0.5 0.5 0.5 0.4 0.5 0.5 0.5 0.6 0.5 ...
## $ ASApLen        : num    0.5 0.5 0.6 0.5 0.5 0.5 0.6 0.5 0.6 0.5 ...
## $ PSFillLen      : num    1.2 NA 1.1 1.2 1.1 1.1 1.1 1.4 1.2 1.2 ...
## $ PSAnthLen      : num    0.5 0.5 0.5 0.5 0.5 NA 0.6 0.5 0.6 0.4 ...
```

```

## Recode integer as numeric
MorphData$BVeins<-as.numeric(MorphData$BVeins)
MorphData$BSpines<-as.numeric(MorphData$BSpines)
MorphData$LTeeth<-as.numeric(MorphData$LTeeth)

## Add midvalue for missing data
for(Row in 1:nrow(MorphData)){
  for(Col in 2:ncol(MorphData)){
    if(is.na(MorphData[Row,Col])){
      MorphData[Row,Col]<-mean(MorphData[MorphData$Loc==MorphData$Loc[Row],Col],na.rm=T)
    }
  }
}

## Principal Components Analysis
PC<-prcomp(MorphData[,2:ncol(MorphData)],scale=T,center=T)
### Summary
summary(PC)

```

```

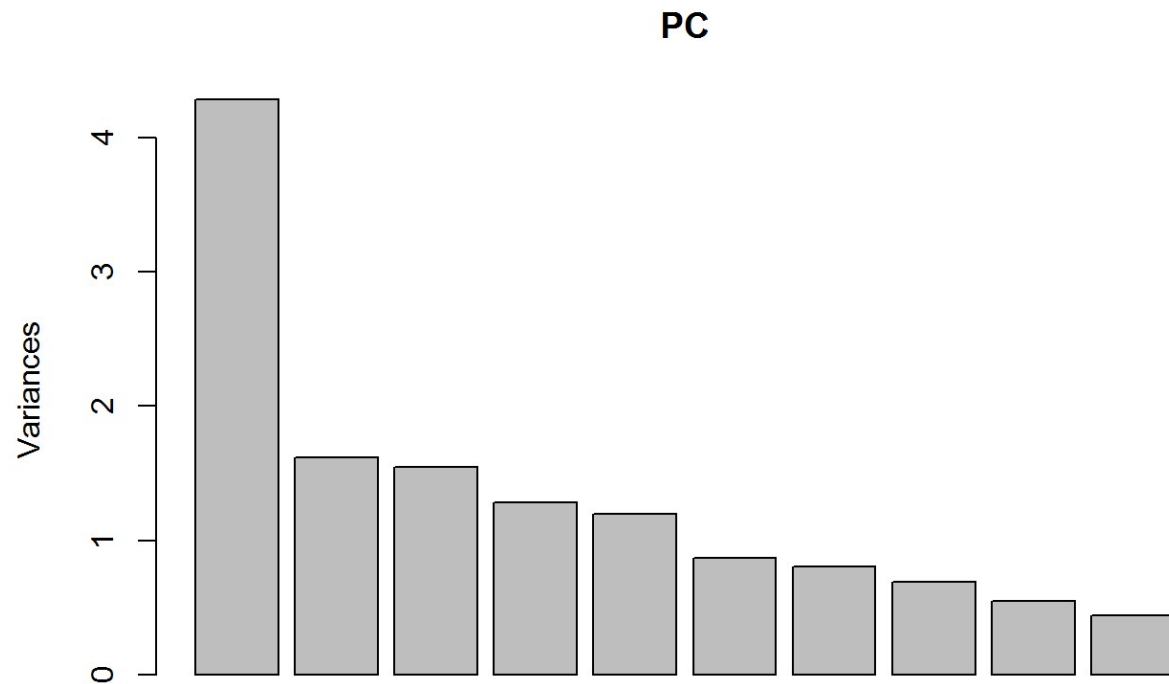
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6
## Standard deviation      2.0704 1.2724 1.2439 1.13256 1.09294 0.93352
## Proportion of Variance 0.3062 0.1156 0.1105 0.09162 0.08532 0.06225
## Cumulative Proportion 0.3062 0.4218 0.5323 0.62396 0.70928 0.77153
##              PC7      PC8      PC9      PC10      PC11      PC12
## Standard deviation      0.89550 0.82883 0.73852 0.66564 0.59368 0.42020
## Proportion of Variance 0.05728 0.04907 0.03896 0.03165 0.02518 0.01261
## Cumulative Proportion 0.82881 0.87788 0.91683 0.94848 0.97366 0.98627
##              PC13      PC14
## Standard deviation      0.35956 0.2509
## Proportion of Variance 0.00923 0.0045
## Cumulative Proportion 0.99550 1.0000

```

```

### ScreePlot
screeplot(PC)

```



```
### % Variation explained by PC1 alone  
100*sum(summary(PC) [[1]] [1])/sum(summary(PC) [[1]])
```

```
## [1] 16.56535
```

```
### % Variation explained by PC1 & PC2  
100*sum(summary(PC) [[1]] [1:2])/sum(summary(PC) [[1]])
```

```
## [1] 26.74563
```

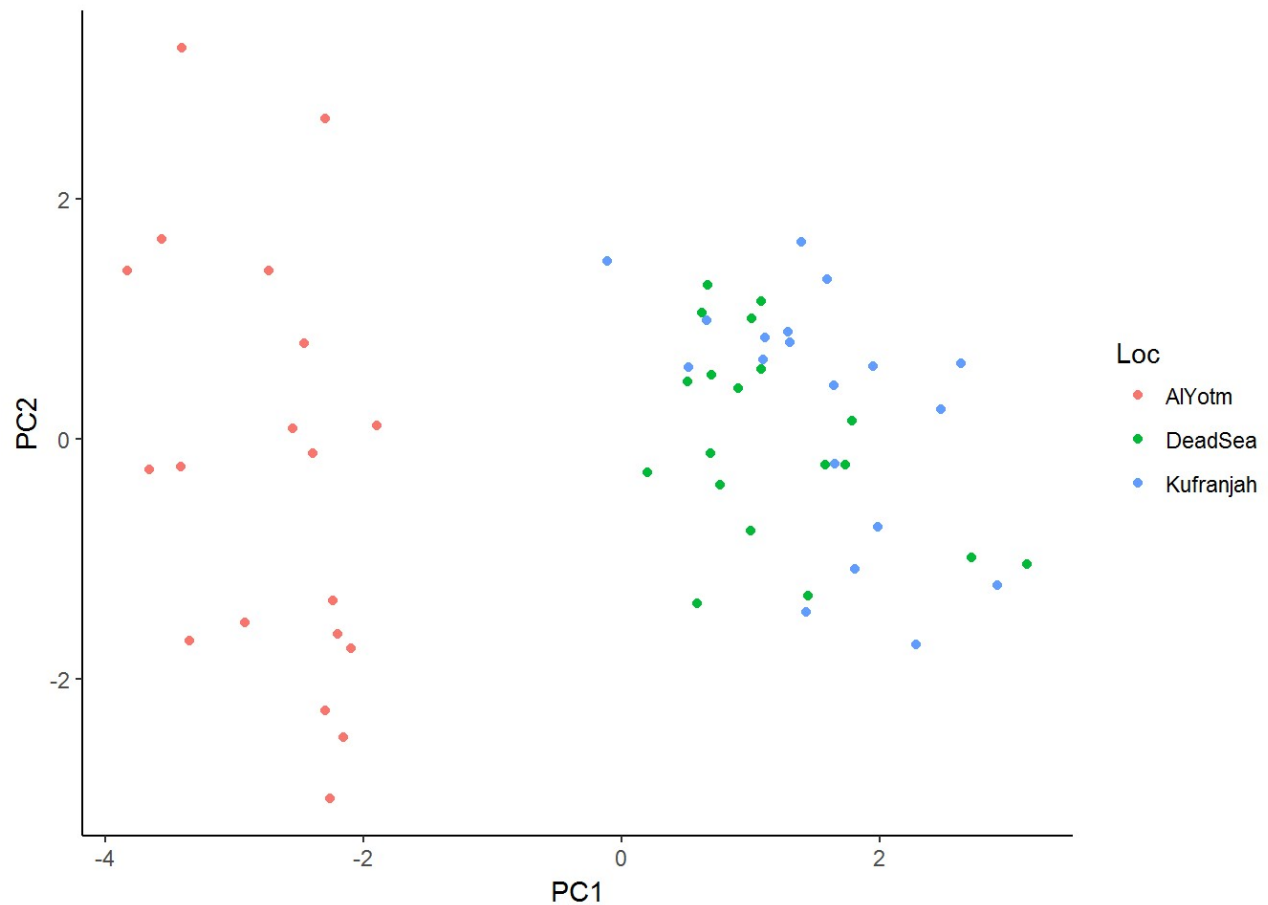
```
### % Variation explained by first n PCs  
n<-10  
100*sum(summary(PC) [[1]] [1:n])/sum(summary(PC) [[1]])
```

```
## [1] 87.00394
```

```
### Factor Loadings  
PC$Rotation
```

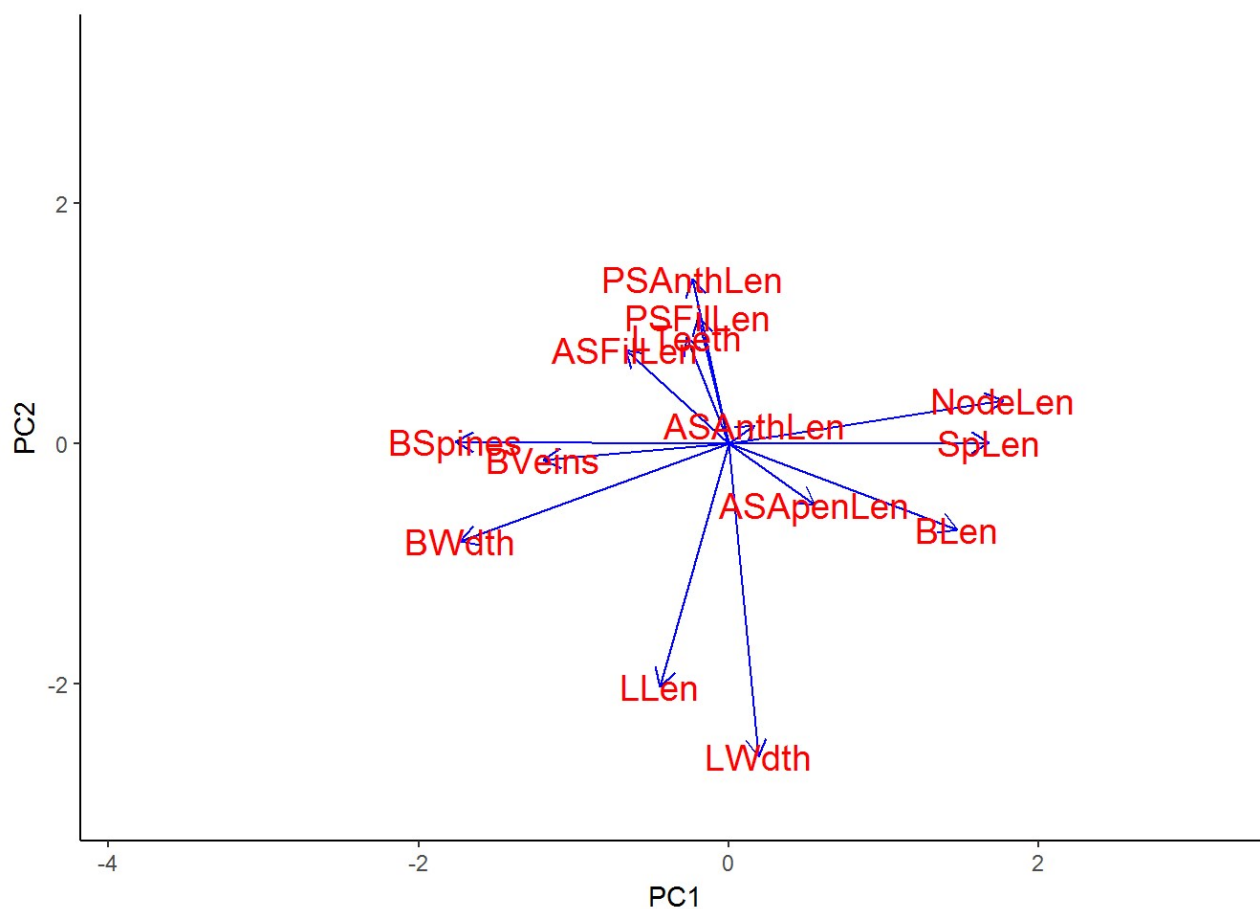
```
## NULL
```

```
### Plot of PC1 & 2  
#### Individuals  
autoplot(PC,data=MorphData,colour="Loc",scale=0)+theme_classic()
```



```
#### Loadings  
autoplot(PC,data=MorphData,colour=NA,  
         loadings = TRUE, loadings.colour = 'blue',  
         loadings.label = TRUE, loadings.label.size = 5,scale=0)+theme_classic  
( )
```

```
## Warning: Removed 57 rows containing missing values (geom_point).
```



```
## Linear Model to test ability of PCs to
MorphPCs<-cbind(MorphData,PC$x)
anova(lm(PC1 ~ Loc,data=MorphPCs)) ## PC1 only significant factor
```

```
## Analysis of Variance Table
##
## Response: PC1
##          Df  Sum Sq Mean Sq F value    Pr(>F)
## Loc         2 212.884  106.442   211.58 < 2.2e-16 ***
## Residuals  54   27.166    0.503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(PC2 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC2
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  2.448  1.2240  0.7493 0.4776
## Residuals  54 88.213  1.6336
```

```
anova(lm(PC3 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC3
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  1.970  0.98523  0.6283 0.5373
## Residuals  54 84.672  1.56800
```

```
anova(lm(PC4 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC4
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  2.401  1.2006  0.9338 0.3993
## Residuals  54 69.429  1.2857
```

```
anova(lm(PC5 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC5
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  0.658  0.32906  0.2683 0.7657
## Residuals  54 66.234  1.22656
```

```
anova(lm(PC6 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC6
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  0.180  0.08992  0.0999 0.9051
## Residuals  54 48.622  0.90040
```

```
anova(lm(PC7 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC7
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  1.067  0.53371   0.6574 0.5223
## Residuals  54 43.840  0.81186
```

```
anova(lm(PC8 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC8
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Loc         2   3.192   1.5958   2.4427 0.09648 .
## Residuals  54 35.278   0.6533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(PC9 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC9
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  0.0355  0.01777   0.0315  0.969
## Residuals  54 30.5074  0.56495
```

```
anova(lm(PC10 ~ Loc,data=MorphPCs))
```

```
## Analysis of Variance Table
##
## Response: PC10
##           Df Sum Sq Mean Sq F value Pr(>F)
## Loc         2  0.8703  0.43516   0.9815 0.3813
## Residuals  54 23.9420  0.44337
```