# PCA.R

*rob_c*

*Mon Jun 05 13:05:25 2017*

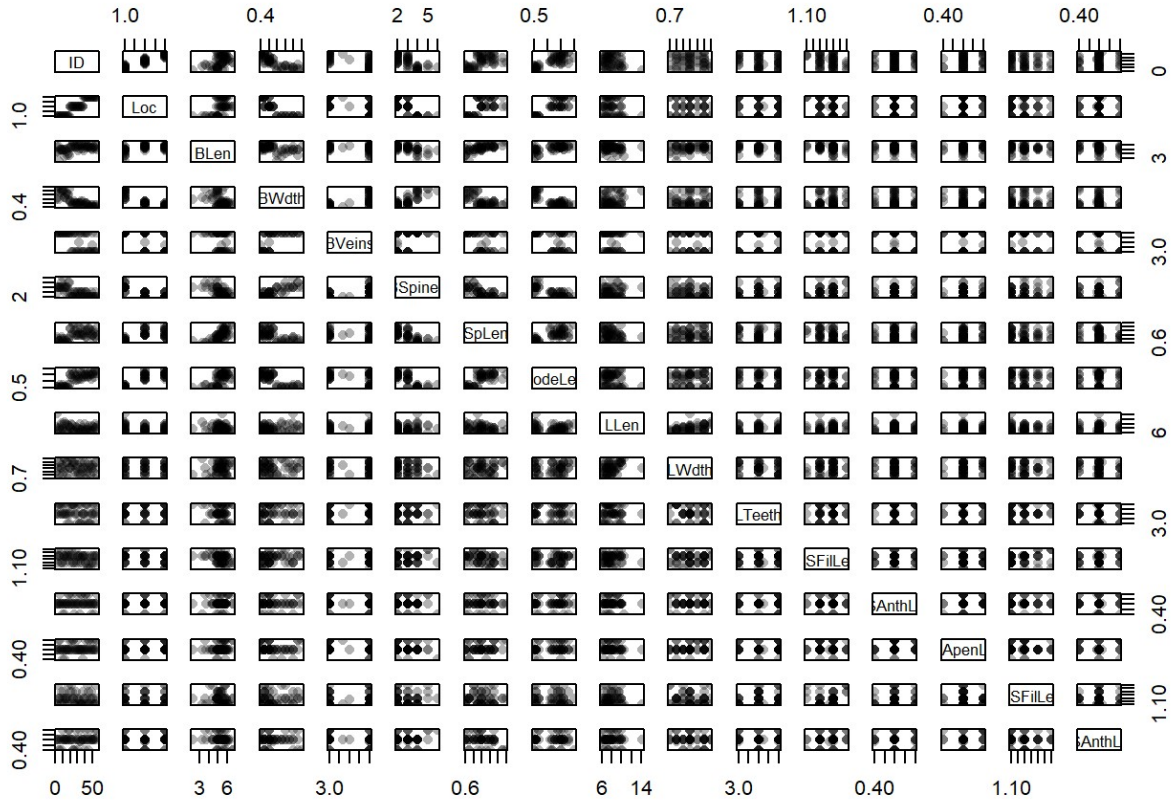
```
# SUMMARY:
## Al-Yotm distinct from Dead Sea and Kufranjah valley sites
## Dead Sea and Kufranjah are different ON AVERAGE, but a lot of overlap

# NOTES:
## Plotting PC results using ggplot2
## https://cran.r-project.org/web/packages/ggfortify/vignettes/plot_pca.html


# 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  16 variables:
##  $ ID       : Factor w/ 57 levels "A1","A10","A11",..: 39 50 51 52 53 54 55 56 57 4
0 ...
##  $ 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 ...
##  $ BWdth    : 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 ...
##  $ LWdth    : 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 ...
##  $ ASFilLen : 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 ...
##  $ ASApenLen: num  0.5 0.5 0.6 0.5 0.5 0.5 0.6 0.5 0.6 0.5 ...
##  $ PSFilLen : 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 ...
```

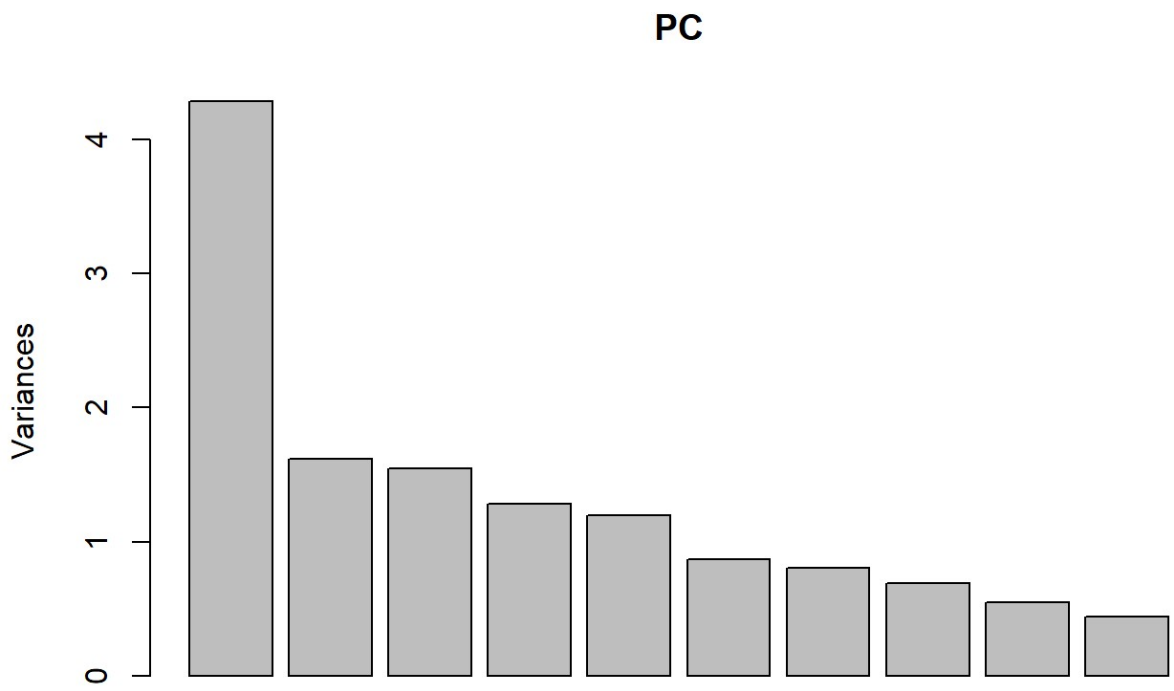
```
## Add midvalue for missing data
for(Row in 1:nrow(MorphData)){
  for(Col in 2:ncol(MorphData)){
    if(is.na(MorphData[Row,Col])){
      MorphData[Row,Col]<-mean(MorphData[MorphData$Loc==MorphData$Loc[Row],Col],na.rm=
T)
    }
  }
}

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

## Inspect pairwise scatterplots
pairs(MorphData,col=rgb(0,0,0,0.3),pch=16)
```




```
## Principal Components Analysis
PC<-prcomp(MorphData[,grep("ID|Loc",names(MorphData),invert=T)],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
```

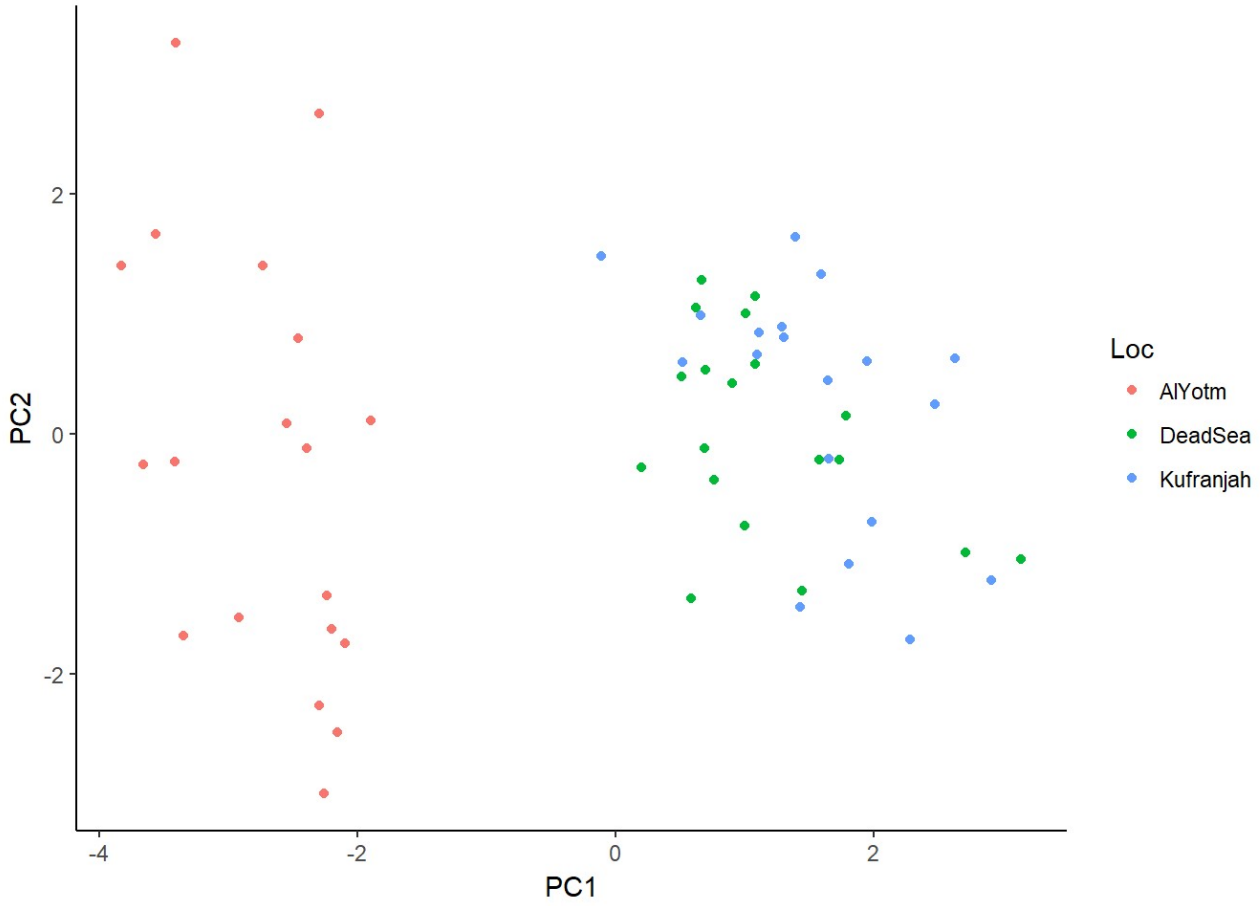
```
### % Variation explained by first n PCs
n<-10
100*sum(summary(PC)[[1]][1:n])/sum(summary(PC)[[1]])
```

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

```
### Factor Loadings
PC$rotation
```

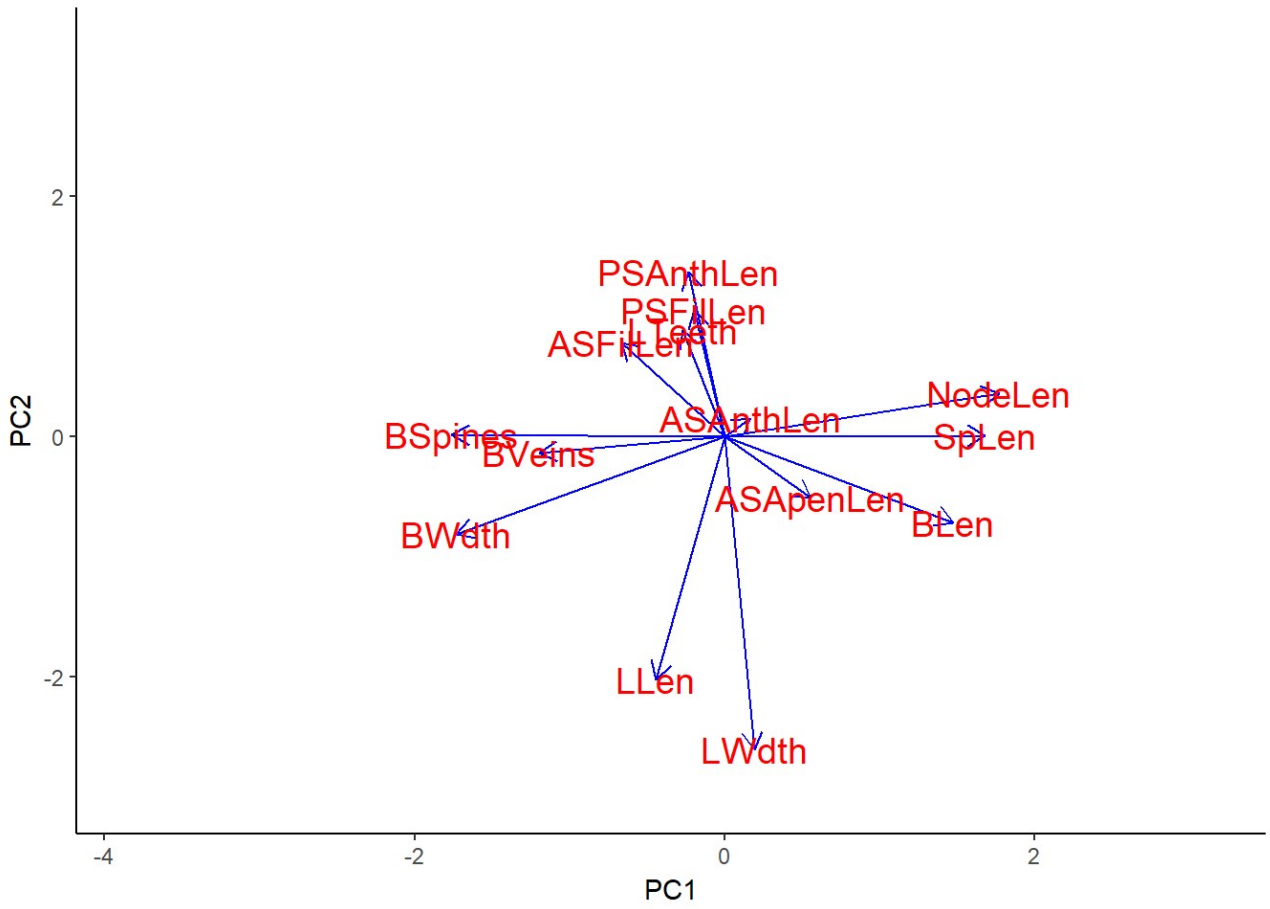
```
##                  PC1          PC2         PC3           PC4         PC5
## BLen       0.35997360 -0.173275624  0.17213646 -0.0652022915  0.14794461
## BWdth     -0.42184542 -0.198254844  0.08404247 -0.0789167022  0.20702657
## BVeins    -0.29081503 -0.032374416  0.44349371 -0.0832740890  0.10233086
## BSpines   -0.42894470  0.004044228  0.03156947  0.1534411553  0.19712624
## SpLen      0.40924571  0.002741427  0.23287070 -0.0917593889  0.07247810
## NodeLen    0.43190176  0.088658295 -0.05751611  0.0004966984 -0.12813240
## LLen      -0.10731076 -0.491422477 -0.28621740 -0.0853446129 -0.27674924
## LWdth      0.04619705 -0.634673057 -0.18553366 -0.1895568467 -0.18314304
## LTeeth    -0.06608094  0.215041388  0.15159130 -0.5087609024 -0.13979283
## ASFilLen  -0.16247572  0.190731523 -0.34276514 -0.3518117334 -0.23678918
## ASAnthLen  0.04058475  0.038290520 -0.31640986 -0.1777418634  0.62771079
## ASApenLen  0.13522523 -0.124002049 -0.30841104 -0.1594884237  0.51550228
## PSFilLen  -0.04717467  0.256857708 -0.07426098 -0.5813604990 -0.06190961
## PSAnthLen -0.05748107  0.334652807 -0.50052719  0.3602618649 -0.12211527
##                  PC6         PC7         PC8         PC9        PC10
## BLen      -0.159721642  0.03019045 -0.19635700  0.36654607 -0.50827442
## BWdth     -0.101420311  0.06612352 -0.09821015  0.19125787 -0.19099882
## BVeins    -0.150762162  0.14174064 -0.34401679  0.33267101  0.36121332
## BSpines    0.019717535  0.01255632  0.05322868  0.16843691 -0.14662435
## SpLen     -0.177819123  0.06689560 -0.07335054  0.24336683  0.25636486
## NodeLen    0.021890197 -0.02132991 -0.27119949  0.12576193  0.11809152
## LLen      -0.034619876 -0.46603032 -0.03246415  0.26821717  0.43306239
## LWdth     -0.004676473  0.22292691 -0.01157547 -0.03445754 -0.34936497
## LTeeth     0.683356046 -0.27046465 -0.07238972  0.22485320 -0.18691967
## ASFilLen  -0.086575677  0.47608369 -0.51301736 -0.16618478  0.07415894
## ASAnthLen -0.132092583 -0.45564637 -0.38031849 -0.22571090 -0.04568098
## ASApenLen  0.348188582  0.43414010  0.31619690  0.23044471  0.30228266
## PSFilLen  -0.540220874 -0.08160538  0.48328915  0.13495641 -0.05052305
## PSAnthLen -0.052654676 -0.01000058 -0.06230853  0.58430507 -0.16577904
##                 PC11         PC12        PC13        PC14
## BLen       0.42671994  0.248520884  0.12146873 -0.26141926
## BWdth      0.16888643 -0.048802392  0.03805911  0.77257616
## BVeins    -0.40224983  0.264089386  0.13121161 -0.22183052
## BSpines    0.13336242 -0.699797997  0.11554523 -0.42247795
## SpLen      0.02350697 -0.425193998 -0.63159677  0.13388989
## NodeLen   -0.15906207 -0.385386443  0.66917514  0.24497043
## LLen       0.30690736  0.027437814  0.02215501 -0.08269281
## LWdth     -0.53109131 -0.139356154 -0.10521318 -0.07427402
## LTeeth    -0.06608973 -0.034955796 -0.10733800  0.00740825
## ASFilLen   0.30105581 -0.045089551 -0.08550321 -0.09726942
## ASAnthLen -0.19695089 -0.004192662 -0.08098495 -0.05528404
## ASApenLen  0.06225634  0.092574492  0.11295387 -0.02576185
## PSFilLen  -0.08928241 -0.027723814  0.14779853 -0.03247371
## PSAnthLen -0.24910695  0.119576082 -0.18699029  0.05007138
```

```
### 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
```

```
## Software version info
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 15063)
##
## locale:
## [1] LC_COLLATE=English_Canada.1252  LC_CTYPE=English_Canada.1252
## [3] LC_MONETARY=English_Canada.1252 LC_NUMERIC=C
## [5] LC_TIME=English_Canada.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggfortify_0.4.1 ggplot2_2.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.9      knitr_1.15.1     magrittr_1.5     munsell_0.4.3
##  [5] colorspace_1.3-2 R6_2.2.0         stringr_1.1.0    plyr_1.8.4
##  [9] dplyr_0.5.0      tools_3.3.2      grid_3.3.2       gtable_0.2.0
## [13] DBI_0.6-1        htmltools_0.3.5  lazyeval_0.2.0   rprojroot_1.2
## [17] digest_0.6.12    assertthat_0.1   tibble_1.2       gridExtra_2.2.1
## [21] tidyr_0.6.1      evaluate_0.10    rmarkdown_1.3    labeling_0.3
## [25] stringi_1.1.2    scales_0.4.1     backports_1.0.5
```