

Local vs Imported Honeybee Stocks - Multivariate Analyses

P. Alexander Burnham

November 12, 2017

```
# Preliminaries:

# Clear memory of characters:
ls()
rm(list=ls())

# Set Working Directory:
setwd("~/BurnhamAlexPrivate/LocalCaliforniaStudy_Hamilton")

# Read in Data:
data <- read.table("LoCalMV.csv", header=TRUE, sep = ",", stringsAsFactors = FALSE)

data$logDWW <- log(data$DWW + 1)
data$logBQCV <- log(data$BQCV + 1)
data$logIAPV <- log(data$IAPV + 1)
data$logNosema <- log(data$Nosema + 1)

# split data into two time points:
x <- split(data, data$Time)
Time1 <- x$`1`
Time2 <- x$`2`

# required packages:
library(plyr)
library(ggplot2)
library(dplyr)
library(lme4)
library(car)
library(MASS)
library(vegan)
library(factoextra)
library(candisc)
```

Linear Classification Analysis for all Vars on Grouping into Treatments (2 time points):

LDA for Time Point 1:

```
# USING CANDISC
lm.mod <- lm(cbind(Pollen, logBQCV, logIAPV, logDWW, Varroa, logNosema, Mass, Brood)~Origin, data=Time1,
anova(lm.mod, test="Wilks")
```

```
## Analysis of Variance Table
##
##           Df    Wilks approx F num Df den Df Pr(>F)
## (Intercept) 1 0.00571    588.03      8    27 <2e-16 ***
## Origin      1 0.76003      1.07      8    27 0.4151
## Residuals   34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

can.mod <- candisc(lm.mod, term="Origin")
can.mod
```

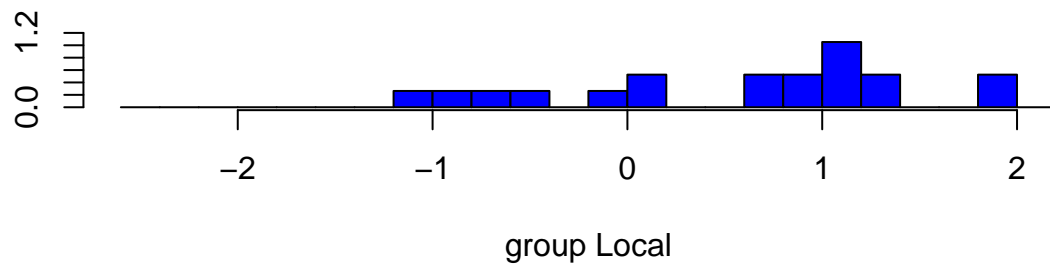
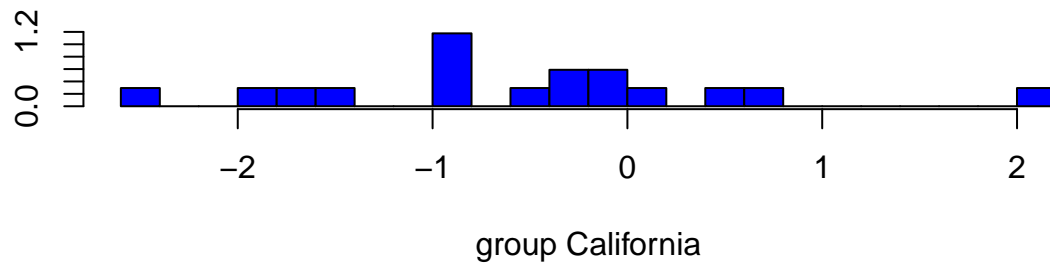
```
##
## Canonical Discriminant Analysis for Origin:
##
##      CanRsqr Eigenvalue Difference Percent Cumulative
## 1 0.23997    0.31573                100          100
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
##
##      LR test stat approx F numDF denDF Pr(> F)
## 1      0.76003    1.0656      8    27 0.4151
```

```
summary(can.mod)
```

```
##
## Canonical Discriminant Analysis for Origin:
##
##      CanRsqr Eigenvalue Difference Percent Cumulative
## 1 0.23997    0.31573                100          100
##
## Class means:
##
## [1] -0.57730  0.51653
##
## std coefficients:
##      Pollen  logBQCV  logIAPV  logDWV  Varroa logNosema      Mass
## -0.71298 -0.31384 -0.38623 -0.32532 -0.85068 -0.39364  0.43165
##      Brood
##      0.67796
```

Classification Analysis for Time point 1:

```
# run LDA
time1 <- lda(Origin~ Pollen + logBQCV + logIAPV + logDWV + Varroa + logNosema + Mass + Brood, data=Time)
plot(time1, col = c("blue"))
```



```
# create data set that includes only variables of interest:
x <- dplyr::select(Time1, Origin, Pollen, logBQCV, logIAPV, logDWV, Varroa, logNosema, Mass, Brood)

# create predictions based on confusion matrix
predictions <- predict(time1, x[,2:9])$class

# summarize accuracy
cm <- table(predictions, x$Origin)
prop.table(cm,1)

##
## predictions  California      Local
## California  0.7058824 0.2941176
## Local       0.2631579 0.7368421
```

PERMANOVA - Time point 1

```
# create dissimilarity matrix
# create matrix for PERM
LoCal1x <- na.omit(Time1)
Dis1 <- dplyr::select(LoCal1x, Pollen, logBQCV, logIAPV, logDWV, Varroa, logNosema, Mass, Brood)

# run PERMANOVA
envdist1 <- vegdist(Dis1, method= "jaccard", na.rm=TRUE)
AD1 <- adonis(envdist1~Origin, data=LoCal1x)
AD1

##
## Call:
## adonis(formula = envdist1 ~ Origin, data = LoCal1x)
##
```

```
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs  MeanSqs F.Model      R2 Pr(>F)
## Origin      1    0.0498 0.049801 0.80121 0.02302 0.577
## Residuals  34    2.1133 0.062157      0.97698
## Total      35    2.1631      1.00000
```

LDA for Time Point 2:

```
# USEING CANDISC
lm.mod2 <- lm(cbind(Pollen, logBQCV, logIAPV, logDWV, Varroa, logNosema, Mass, Brood)~Origin, data=Time2)
anova(lm.mod2, test="Wilks")

## Analysis of Variance Table
##
##           Df   Wilks approx F num Df den Df Pr(>F)
## (Intercept) 1 0.00411   515.09      8    17 <2e-16 ***
## Origin       1 0.45218     2.57      8    17 0.0482 *
## Residuals    24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

can.mod2 <- candisc(lm.mod2, term="Origin")
can.mod2

##
## Canonical Discriminant Analysis for Origin:
##
##      CanRsqr Eigenvalue Difference Percent Cumulative
## 1 0.54782      1.2115              100          100
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
##
##      LR test stat approx F numDF denDF Pr(> F)
## 1      0.45218    2.5744      8    17 0.0482 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

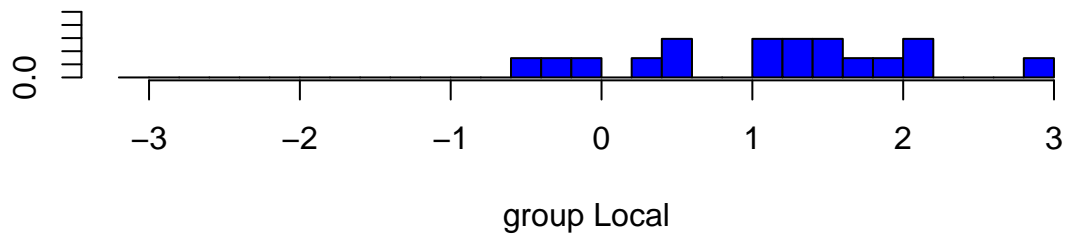
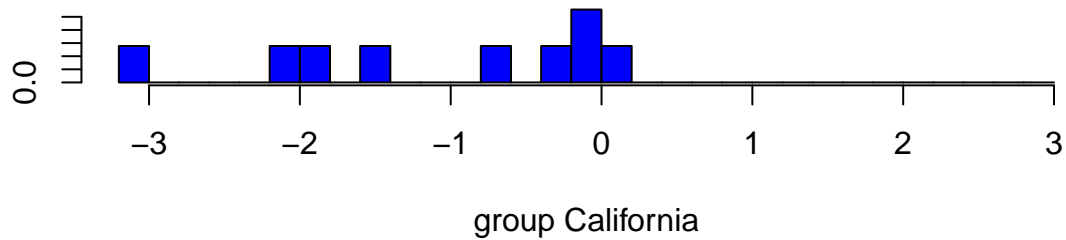
summary(can.mod2)

##
## Canonical Discriminant Analysis for Origin:
##
##      CanRsqr Eigenvalue Difference Percent Cumulative
## 1 0.54782      1.2115              100          100
##
## Class means:
##
## [1] -1.45339 0.76944
##
## std coefficients:
```

```
##      Pollen      logBQCV      logIAPV      logDWV      Varroa      logNosema
## -0.3549737  0.9598486 -0.4989998 -0.0028338 -0.3482053 -0.4730648
##      Mass      Brood
##  0.7766441  0.2267392
```

Classification Analysis for Time point 1:

```
# run LDA
time2 <- lda(Origin ~ Pollen + logBQCV + logIAPV + logDWV + Varroa + logNosema + Mass + Brood, data=Time2)
plot(time2, col = c("blue"))
```



```
x1 <- dplyr::select(Time2, Origin, Pollen, logBQCV, logIAPV, logDWV, Varroa, logNosema, Mass, Brood)
predictions2 <- predict(time2, x1[,2:9])$class
# summarize accuracy
cm1 <- table(predictions2, x1$Origin)
prop.table(cm1, 1)
```

```
##
## predictions2 California    Local
## California  0.8571429 0.1428571
## Local       0.1578947 0.8421053
```

PERMANOVA - Time point 2

```
# create dissimilarity matrix
LoCal2x <- na.omit(Time2)
Dis2 <- dplyr::select(LoCal2x, Pollen, logBQCV, logIAPV, logDWV, Varroa, logNosema, Mass, Brood)
# run PERMANOVA
```

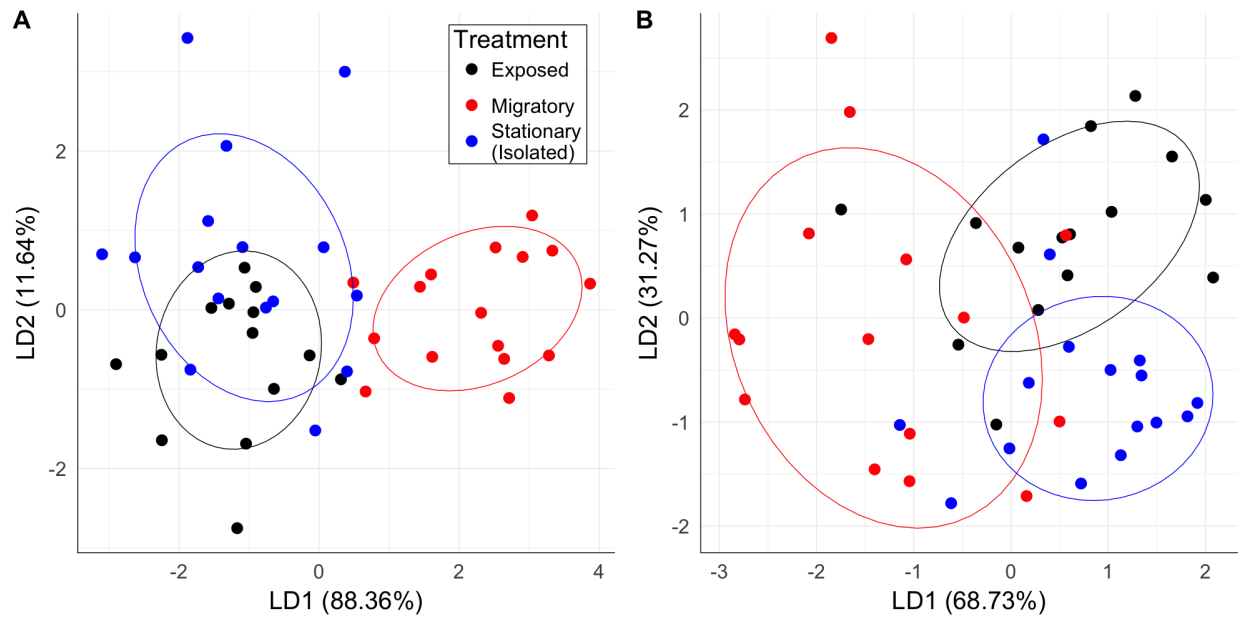


Figure 1: Figure for a Project I did with three groups:

```
envdist <- vegdist(Dis2, method= "jaccard", na.rm=TRUE)
AD2 <- adonis(envdist~Origin, data=LoCal2x)
AD2
```

```
##
## Call:
## adonis(formula = envdist ~ Origin, data = LoCal2x)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##          Df SumsOfSqs  MeanSqs F.Model    R2 Pr(>F)
## Origin    1  0.23516 0.235159  3.5773 0.12972 0.021 *
## Residuals 24  1.57769 0.065737          0.87028
## Total    25  1.81285                1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

NMDS (Non-metric Multidimensional Scaling)

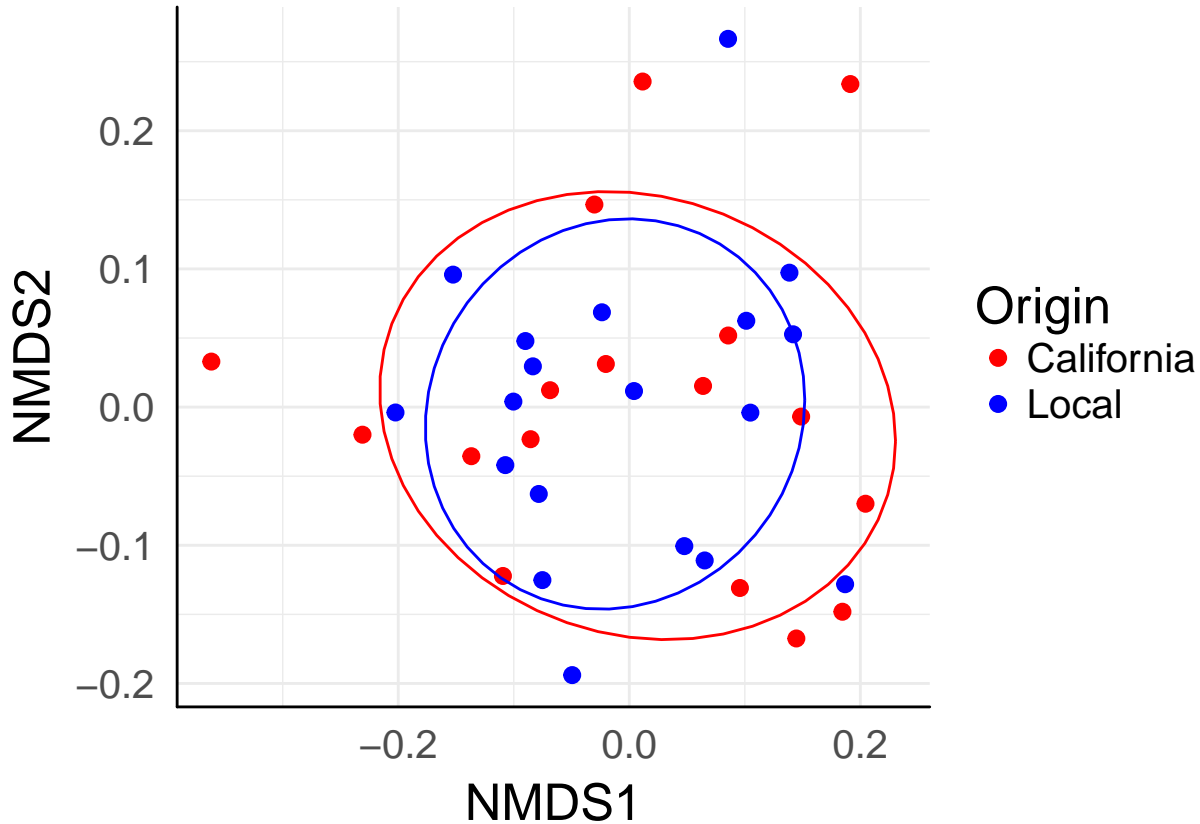
nMDS for Time Point 1:

```
# run nMDS
MDS1 <- metaMDS(envdist1, center=TRUE, autotransform = FALSE)

#Using the scores function from vegan to extract the site scores and convert to a data.frame
```

```
data.scores1 <- as.data.frame(vegan::scores(MDS1))
data.scores1$Origin <- LoCal1x$Origin

NMDS2 <- ggplot(data.scores1, aes(NMDS1, NMDS2, color=Origin))+geom_point(size=2.5) + theme_minimal(base_
NMDS2
```



nMDS for Time Point 2:

```
# run nMDS model:
MDS <- metaMDS(envdist, center=TRUE, autotransform = FALSE)

#Using the scores function from vegan to extract the site scores and convert to a data.frame
data.scores <- as.data.frame(vegan::scores(MDS))
data.scores$Origin <- LoCal2x$Origin

# graph for nMDA
NMDS3 <- ggplot(data.scores, aes(NMDS1, NMDS2, color=Origin))+geom_point(size=2.5) + theme_minimal(base_
NMDS3
```

