Ecological geometry of Africa

John M. Drake

June 12, 2014

1   Introduction

My studies of mosquito distributions in Africa have routinely yielded two interesting phenomena: (1) at least two species (*Anopheles arabiensis* and *Culex pipiens*) are climate generalists in the space of the first two principal components of environmental space, and (2) the set of environmental conditions appears as a "ring" in the first two principal components and a "sheet" in three principal components.

The purpose of this study is to seek to understand this phenomena better.

2   Exploratory plots

First we load data from the ongoing study of *Cx. pipiens.*

```
> load('culex-v4.RData')
```

Now we plot the familar pca plot of environmental conditions in Africa where the scatterplot is resutricted to 2000 observations to better show the "ring". To ensure that these are not ordered in any way, these points are chosen randomly.

```
> par(mfrow=c(2,1))
> plot(pca, main='Scree plot', cex.main=0.5)
> set.seed(10281979)
> n <- dim(pca$x)[1]
> pca.sample <- pca$x[sample(seq(1,50000),5000),]
> plot(pca.sample[,1:2], main='Principal Components Analysis (n=2,000)',
+      pch=19, cex=0.02, col='grey', cex.main=0.5, xlab='PC1', ylab='PC2')
```

Scree plot
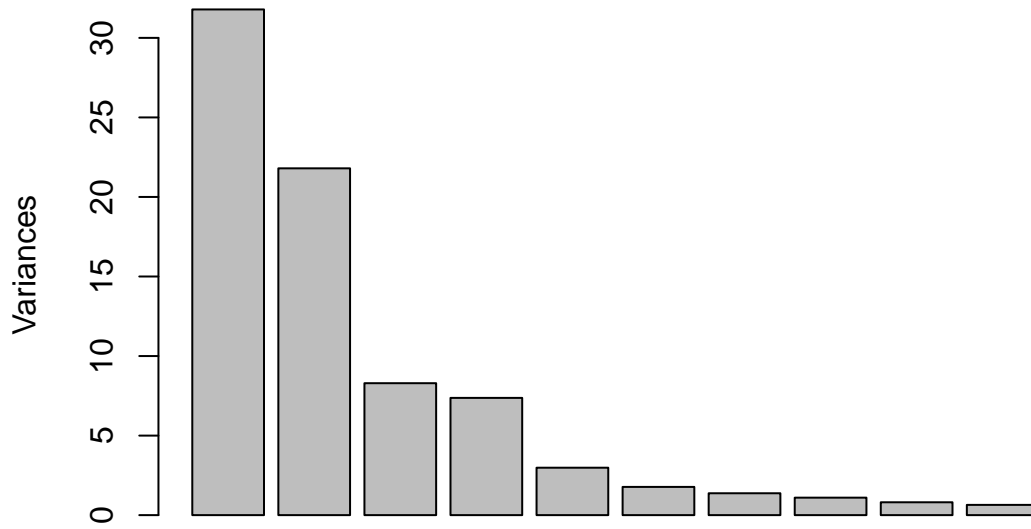
Principal Components Analysis (n=2,000)
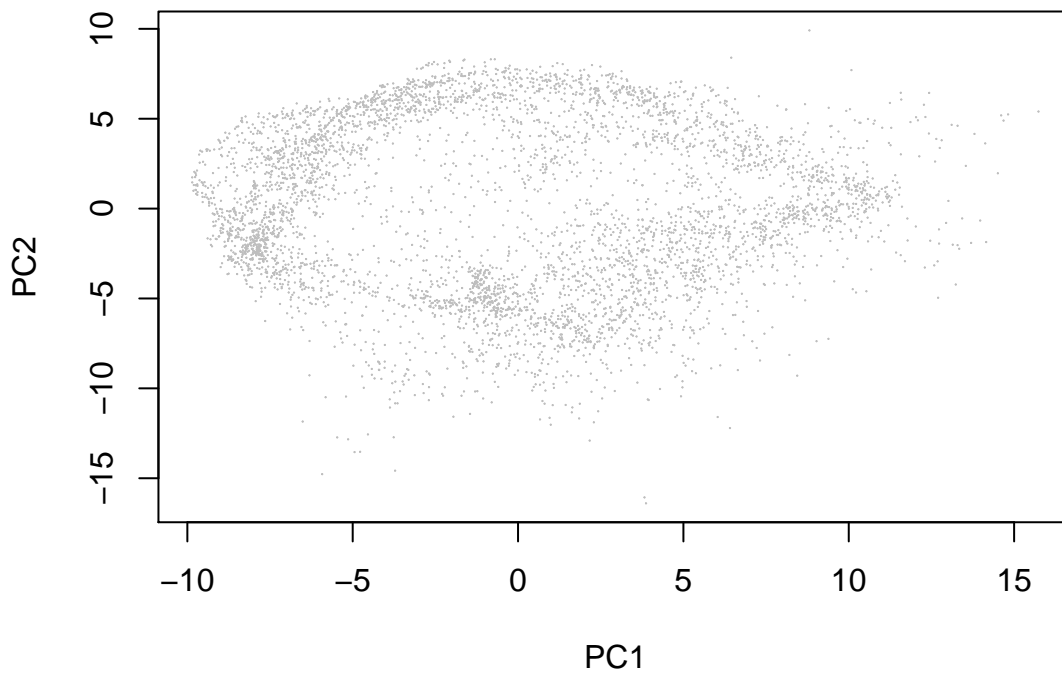
Now, we again inspect the plot in three dimensions

```
> require(scatterplot3d)
> par(mfrow=c(3,1))
> scatterplot <- scatterplot3d(pca.sample[,1:3], pch=19, cex.symbols=0.2, type='p',
```

```
+                                 highlight.3d=TRUE, xlab='PC1', ylab='PC2', zlab='PC3', angle=45)
> scatterplot <- scatterplot3d(pca.sample[,1:3], pch=19, cex.symbols=0.2, type='p',
+                                 highlight.3d=TRUE, xlab='PC1', ylab='PC2', zlab='PC3', angle=100)
> scatterplot <- scatterplot3d(pca.sample[,1:3], pch=19, cex.symbols=0.2, type='p',
+                                 highlight.3d=TRUE, xlab='PC1', ylab='PC2', zlab='PC3', angle=-20)
```
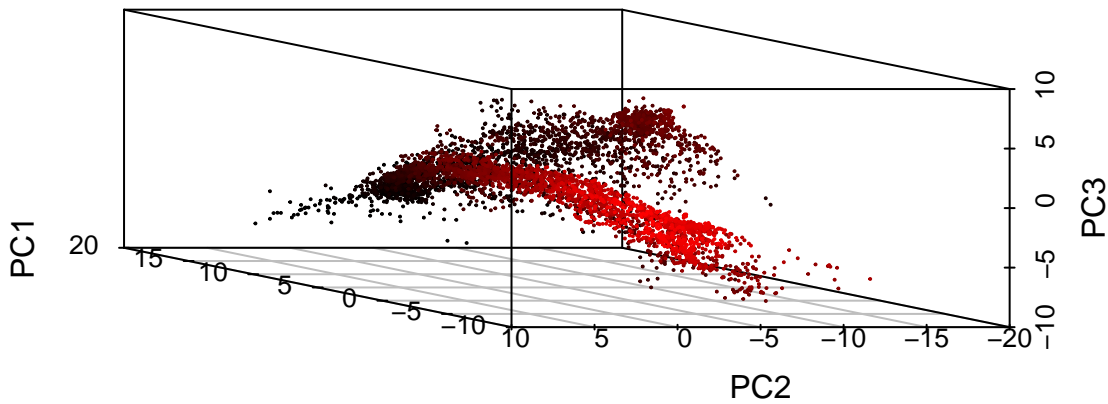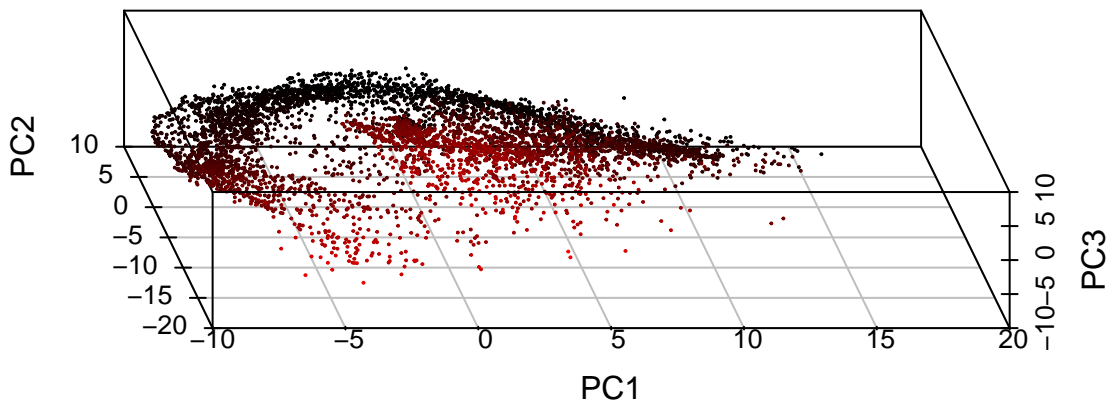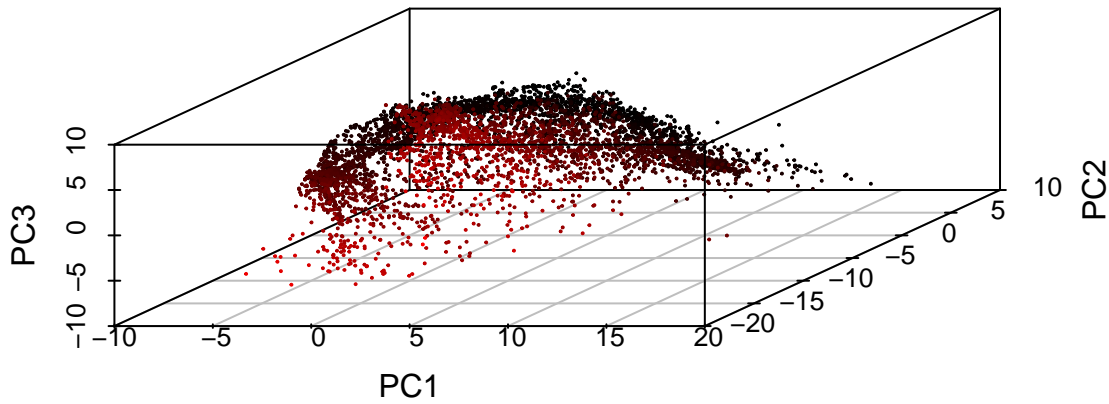
These plots show that there are a few different clusters entangled with each other. At this stage I think some kind of clustering or nonlinear component decomposition is warranted. For now, I'm going to see if I can go learn something about kernel pca and related techniques.