Final Assignment

Ryan Avery

June 13, 2018

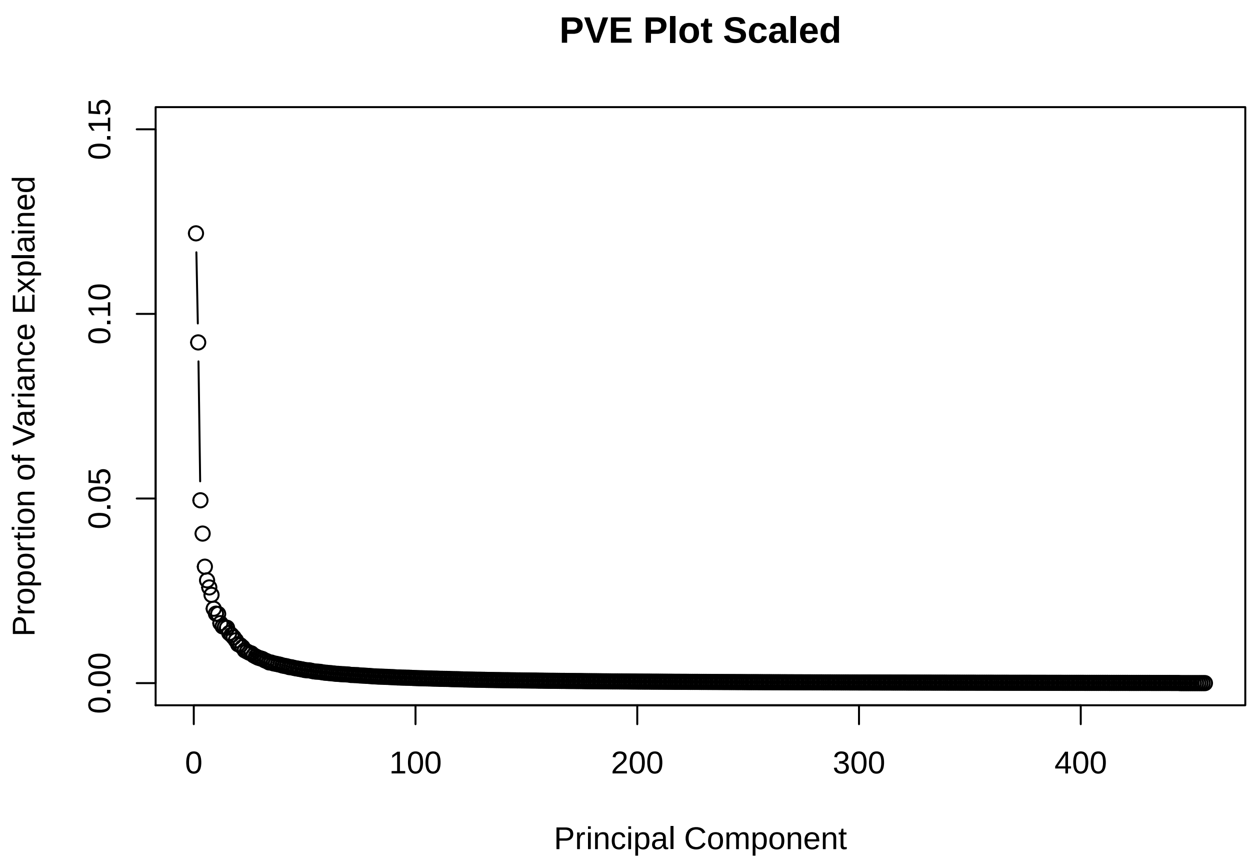
To start I read in the .mat file into a list of matrices and load our libraries.

Below is a plot of the temperature data within the domain of -30 to 30 latitude and 0 to 357 lon1 for a single time observation. I flattened the 3D array into a 2D array, one dimension is space and the other is time. Defined a function to reshape a given observation back to a 2D matrix.

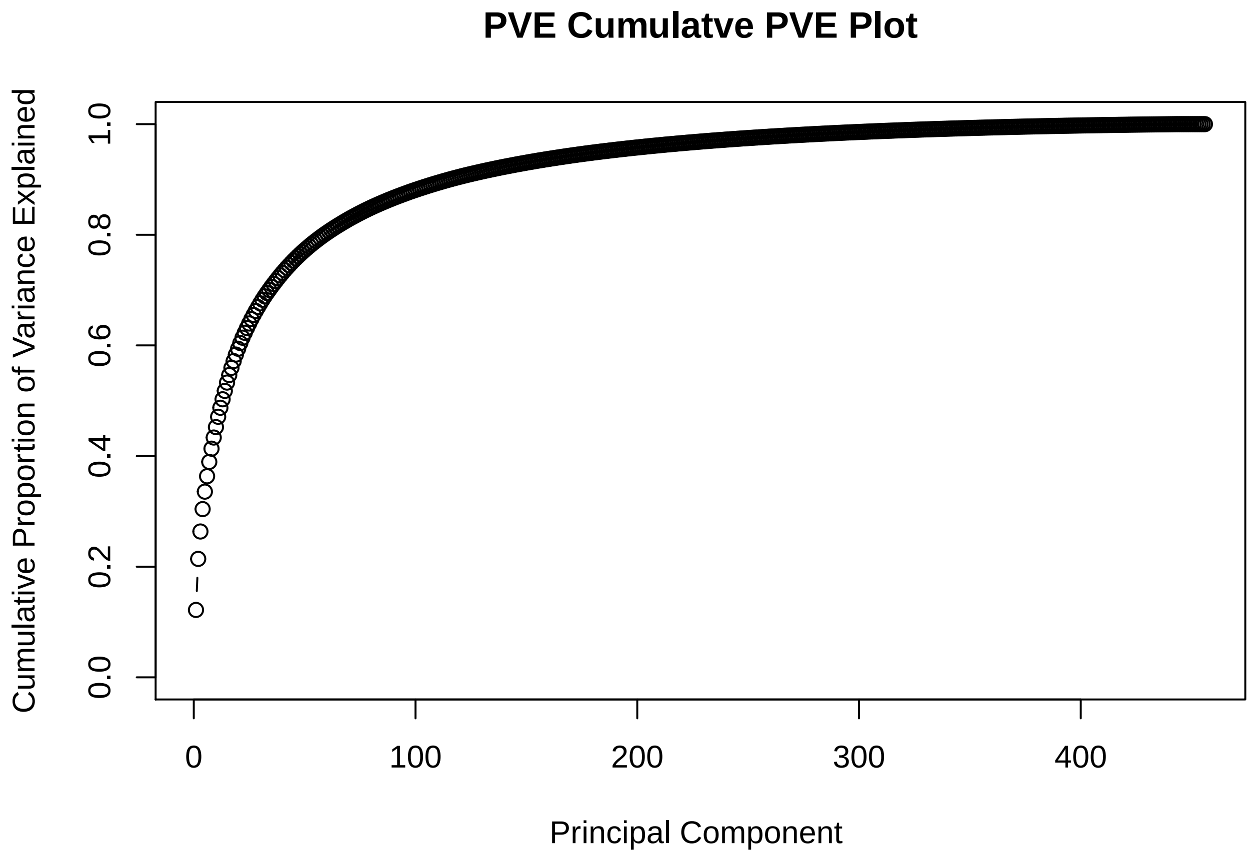
```
var_mat <- sapply(1:456, function(i) as.numeric(mat_listmatrix$var[, 21:41, i]))
plot_var <- function(image_vector) {
plot(as.cimg(matrix(image_vector, ncol=21)), axes=FALSE, asp=1)
}
plot_var(var_mat[,81])
```

1. Plotting PVE by each EOF. After the 116th EOF, 90 percent of the variance is explained. With scaling, the answer would be 125th. Scaling converts values to unit variance before PCA.

```
var.out.noscale=prcomp(t(var_mat), scale=FALSE)
var_var = var.out.noscale$sdev^2
pve = var_var/sum(var_var)

plot(pve, xlab="Principal Component", ylab="Proportion of Variance Explained",ylim= c(0,.15), type='b')
title('PVE Plot Scaled')
```

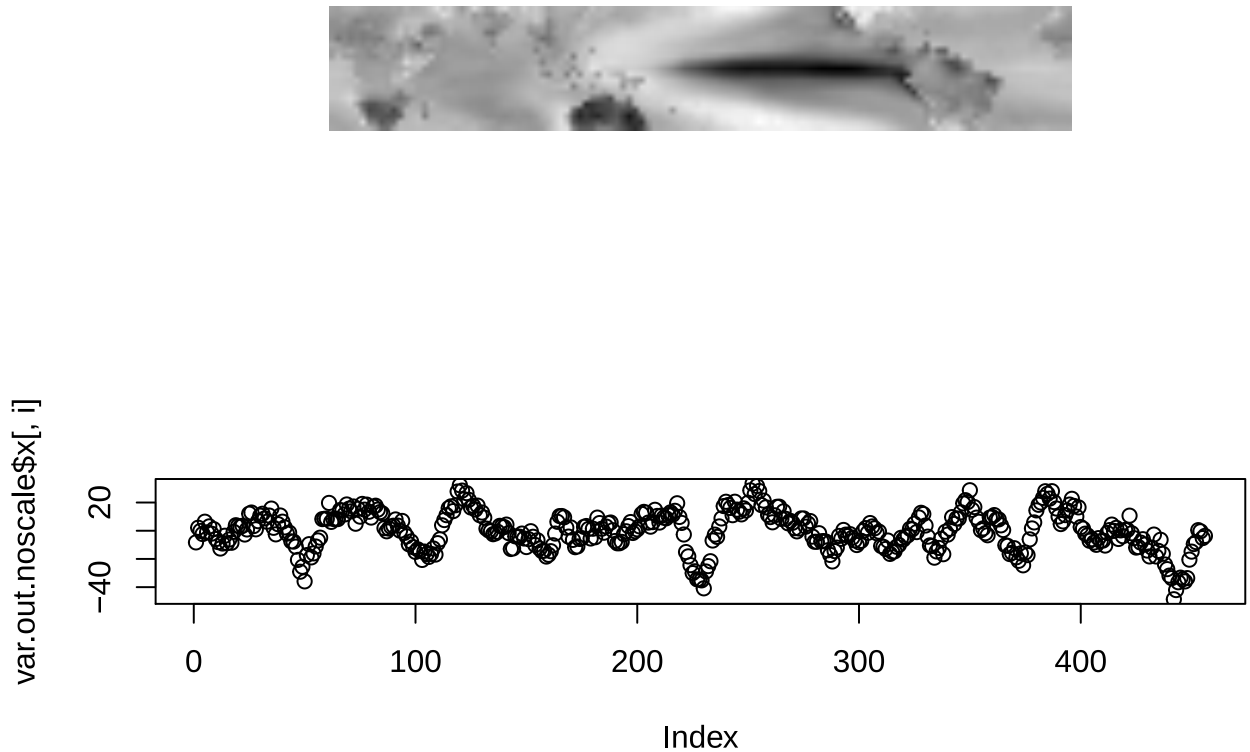



```
plot(cumsum(pve), xlab="Principal Component ",
ylab=" Cumulative Proportion of Variance Explained ", ylim=c(0,1), type='b')
title('PVE Cumulatve PVE Plot')
```

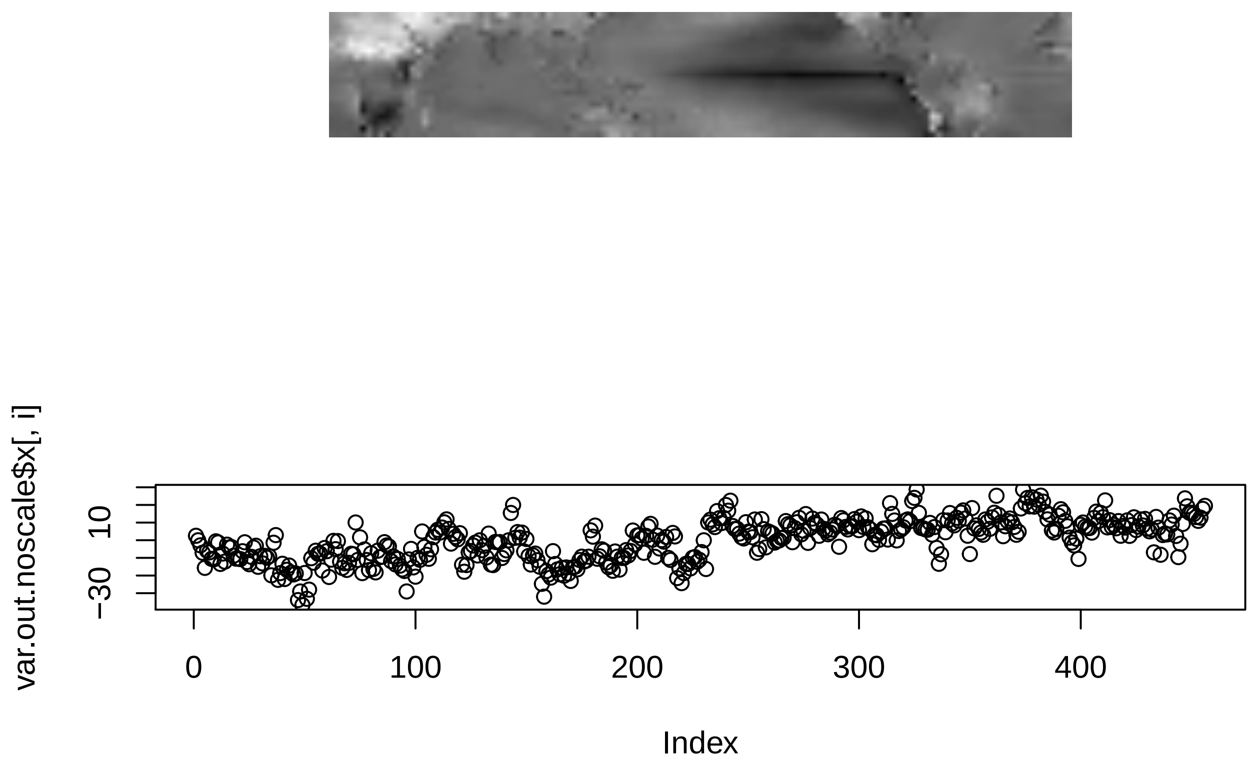


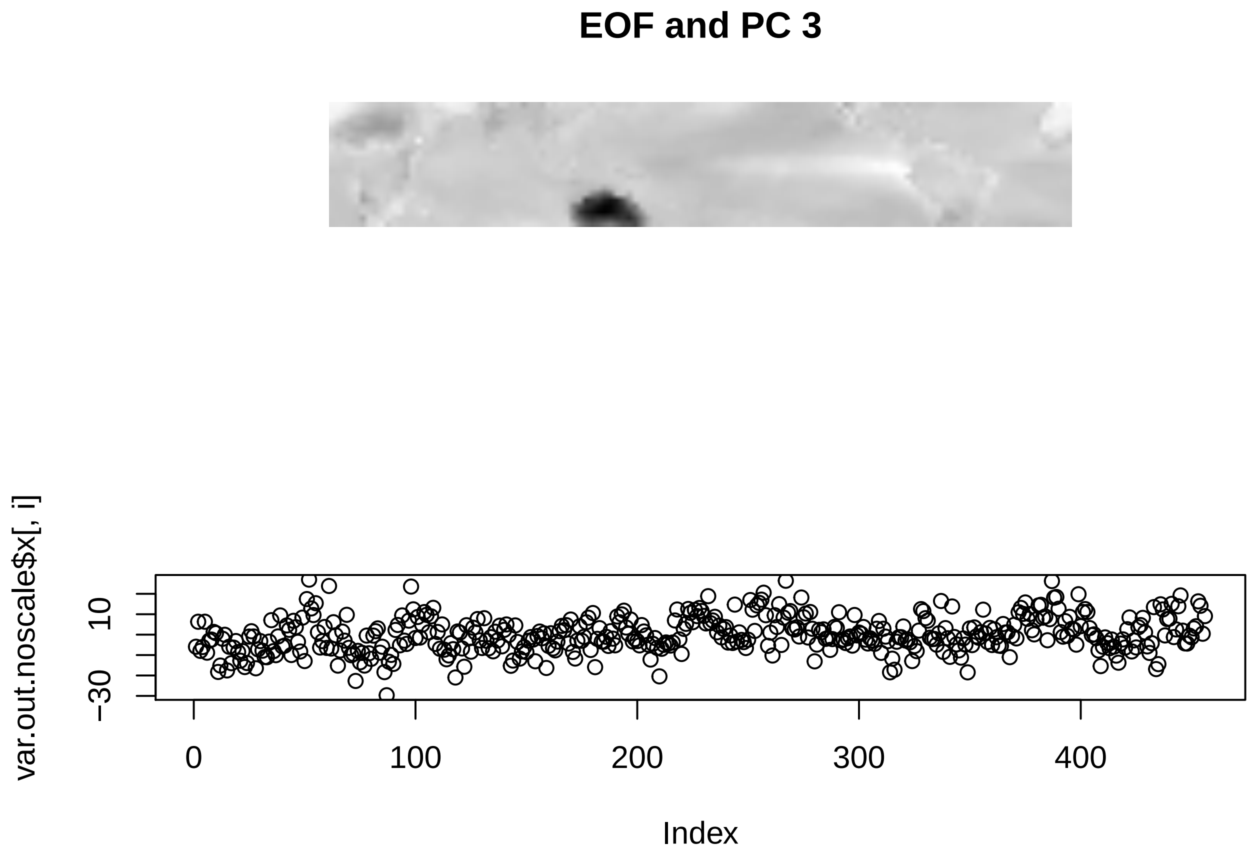

```
for (i in 1:4) {
  par(mfrow=2:1)
  plot_var(var.out.noscale$rotation[,i])
  title(paste('EOF and PC', i))
  plot(var.out.noscale$x[,i])


}
```

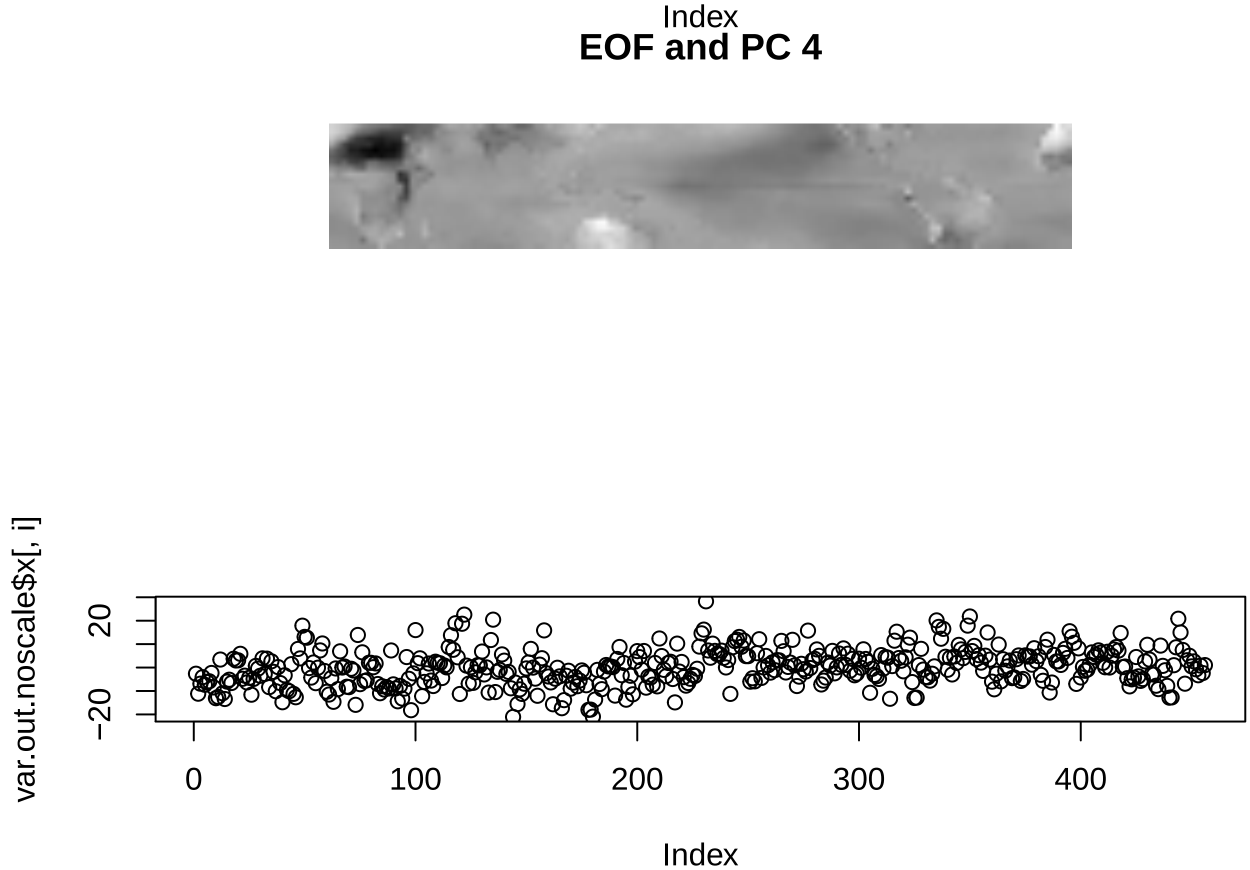
EOF and PC 1



EOF and PC 2




EOF and PC 3
var.out.noscale$x[, i]
10
−30
0
100
200
300
400
Index


EOF and PC 4
var.out.noscale$x[, i]
20
−20
0
100
200
300
400
Index

```r
tibble(EOF = c(1, 2, 3, 4), PVE = pve[1:4])
```

```
## # A tibble: 4 x 2
##     EOF     PVE
##   <dbl>   <dbl>
## 1     1 0.12182
## 2     2 0.09228
## 3     3 0.04952
## 4     4 0.04050
```

2.

```r
min(which(cumsum(pve) > .9))
```

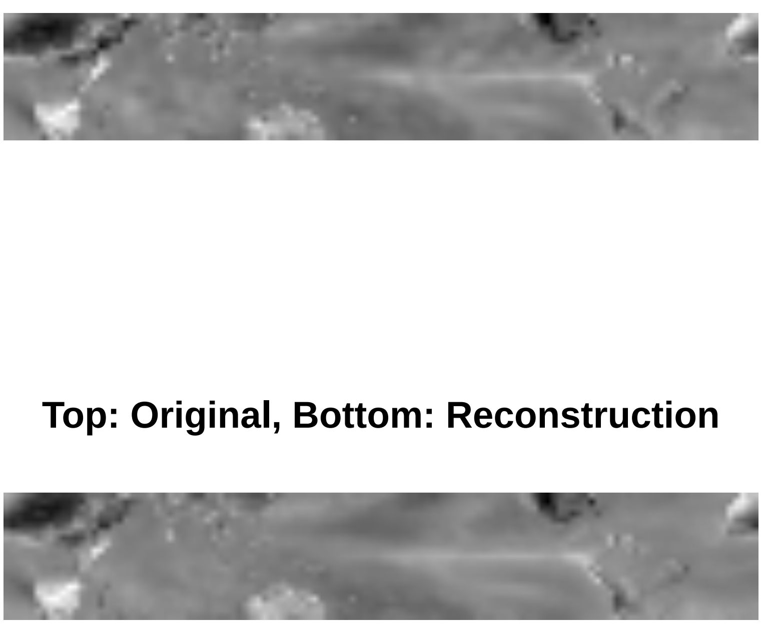
```
## [1] 116
```

```r
print("after the 116th Principal Component, 90 percent of the variance is explained")
```

```
## [1] "after the 116th Principal Component, 90 percent of the variance is explained"
```

PCA reconstruction with 116 PCs. Top plot is original, bottom is reconstruction. Reconstruction is less noisy and still maintains organized variation from ENSO, differences between coastal temperatures and open sea temps, etc.

```r
# PCA reconstruction
mu = colMeans(t(var_mat))
nComp = 116
re_var_mat = var.out.noscale$x[,1:nComp] %*% t(var.out.noscale$rotation[,1:nComp])
re_var_mat = scale(re_var_mat, center = -mu, scale = FALSE)
par(mfrow=2:1)
plot_var(var_mat[,100])
plot_var(t(re_var_mat)[,100])
title('Top: Original, Bottom: Reconstruction')
```




```r
plot_var(var_mat[,100]-t(re_var_mat)[,100])
title('Difference between original and reconstruction')
```

Difference between original and reconstruction

any pair take inner product of PC1 PC2 or EOF1 EOF2

These are very close to 0, showing that the functions are orthagonal.

```
PC1 = var.out.noscale$x[,1]
PC2 = var.out.noscale$x[,2]

PC1 %*% PC2 # inner product not divided by standard deviation
```

```
##            [,1]
## [1,] -3.149e-11
```

```
EOF1 = var.out.noscale$rotation[,1]
EOF2 = var.out.noscale$rotation[,2]

EOF1 %*% EOF2 # inner product not divided by standard deviation
```

```
##            [,1]
## [1,] -2.674e-16
```

4. The power spectrum of the first PC shows that there is more variation at frequencies of about .01*456 years, or around 4 years. This demonstrates tha tthe first PC is associated with ENSO. The first PC also captures some variation at other frequencies but not too much.

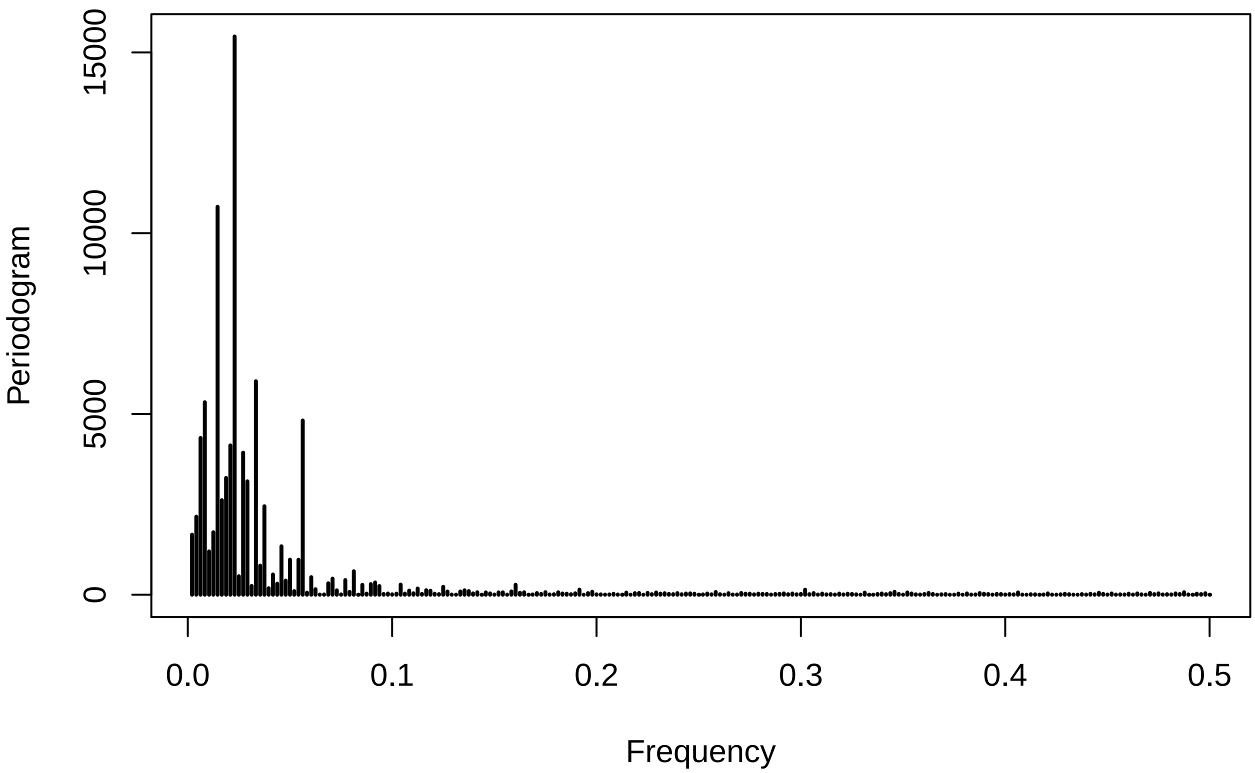
```
library(TSA)
```

```
## Loading required package: leaps

## Loading required package: locfit

## locfit 1.5-9.1    2013-03-22

## Loading required package: mgcv

## Loading required package: nlme

##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
##     collapse

## This is mgcv 1.8-23. For overview type 'help("mgcv-package")'.

## Loading required package: tseries

##
## Attaching package: 'TSA'

## The following object is masked from 'package:readr':
##
##     spec
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
periodogram(var.out.noscale$x[,1])
```



Here we that the 1st principal component (associated with ENSO), is correlated with regions not just between -30 to 30 latitude but all over, particularly in the Pacific, Northern atlantic, southern atlantic, and the atlantic near the equator off the coast of Brazil. Also off the west coast of Australia there is high correlation.

```
var_mat_whole_globe <- sapply(1:456, function(i) as.numeric(mat_listmatrix$var[, , i]))
corr_map = apply(var_mat_whole_globe, 1, function(x) cor(x, PC1))
plot_var2 <- function(image_vector) {
plot(as.cimg(matrix(image_vector, ncol=61)), axes=FALSE, asp=1)
}
plot_var2(corr_map)
```