Supplemental Information: PCA

This is an R markdown document to work through completing a PCA on the acoustic recordings for all four bays. ##Required Libraries

```
library("markdown")
library("rmarkdown")
library("knitr")
```

Toy Problem with a subset of data

```
##    Bay   Date     hh    X16    X20    X25  X16.1  X20.1  X25.1  X31_5
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
##    X40    X50    X63    X80   X100   X125   X160   X200   X250   X315
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
##   X400   X500   X630   X800  X1000  X1250  X1600  X2000  X2500  X3150
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
##  X4000  X5000  X6300  X8000 X10000 X12500 X16000 X20000  rmsdB
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4    PC5     PC6     PC7
## Standard deviation     12.865 6.9697 6.1565 3.78035 2.4069 2.29764 1.89537
## Proportion of Variance  0.563 0.1652 0.1289 0.04861 0.0197 0.01796 0.01222
## Cumulative Proportion   0.563 0.7282 0.8571 0.90571 0.9254 0.94337 0.95559
##                            PC8     PC9    PC10    PC11    PC12    PC13
## Standard deviation     1.59982 1.35754 1.28526 1.06139 1.05316 0.90944
## Proportion of Variance 0.00871 0.00627 0.00562 0.00383 0.00377 0.00281
## Cumulative Proportion  0.96429 0.97056 0.97618 0.98001 0.98378 0.98660
##                           PC14    PC15    PC16    PC17    PC18   PC19
## Standard deviation     0.81305 0.75234 0.73461 0.65025 0.60688 0.5692
## Proportion of Variance 0.00225 0.00193 0.00184 0.00144 0.00125 0.0011
## Cumulative Proportion  0.98885 0.99077 0.99261 0.99404 0.99530 0.9964
##                           PC20    PC21    PC22    PC23    PC24    PC25
## Standard deviation     0.45662 0.44336 0.40714 0.37633 0.30645 0.25854
## Proportion of Variance 0.00071 0.00067 0.00056 0.00048 0.00032 0.00023
## Cumulative Proportion  0.99711 0.99778 0.99834 0.99882 0.99914 0.99937
##                           PC26    PC27   PC28    PC29    PC30    PC31
## Standard deviation     0.22404 0.20210 0.1729 0.12340 0.11474 0.10906
## Proportion of Variance 0.00017 0.00014 0.0001 0.00005 0.00004 0.00004
## Cumulative Proportion  0.99954 0.99968 0.9998 0.99983 0.99988 0.99992
##                           PC32    PC33    PC34      PC35      PC36
## Standard deviation     0.10109 0.09172 0.07378 5.725e-15 1.148e-15
## Proportion of Variance 0.00003 0.00003 0.00002 0.000e+00 0.000e+00
## Cumulative Proportion  0.99995 0.99998 1.00000 1.000e+00 1.000e+00
##                             PC37      PC38      PC39
## Standard deviation     1.148e-15 1.148e-15 1.087e-15
## Proportion of Variance 0.000e+00 0.000e+00 0.000e+00
## Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

Bay 1: Kauhako

Read in the files using read.csv then make it into a data frame. Note that the csv file needs to be in the same directory as the R markdown file. For R to recognize the Date column I first had R recognize it is as a character and then as a Date. Then I told R to recognize this date column as numeric.

Note that these files are ones that have already been cleaned to remove logger servicing and days with malfunctions.

```r
Lzeq1<-read.csv("Kauhako_clean_LZeq.csv")
Lzeq1<-as.data.frame(Lzeq1)
Lzeq1$Date<-as.character(Lzeq1$Date)
Lzeq1$Date<-as.Date(Lzeq1$Date,"%m/%d/%Y")
Lzeq1$Date<-as.numeric(Lzeq1$Date)
```

Now I want to check the data to make sure they imported correctly.

```r
head(Lzeq1);
```

```
##   Bay MSL_Lzeq X.YY MM DD hh min sec  Date     Time      Date.Time BE.M
## 1   1     Lzeq 2011  1  9  0   0   0 14983 12:00 AM 1/9/2011 0:00   NA
## 2   1     Lzeq 2011  1  9  0   4   0 14983 12:04 AM 1/9/2011 0:04   NA
## 3   1     Lzeq 2011  1  9  0   8   0 14983 12:08 AM 1/9/2011 0:08   NA
## 4   1     Lzeq 2011  1  9  0  12   0 14983 12:12 AM 1/9/2011 0:12   NA
## 5   1     Lzeq 2011  1  9  0  16   0 14983 12:16 AM 1/9/2011 0:16   NA
## 6   1     Lzeq 2011  1  9  0  20   0 14983 12:20 AM 1/9/2011 0:20   NA
##        X16      X20      X25    X31_5      X40      X50      X63      X80
## 1 83.69925 84.61776 82.10266 79.44302 80.07610 79.77297 84.37819 85.75966
## 2 83.61594 81.64740 81.98719 80.63694 83.18519 83.03494 84.97215 87.49585
## 3 86.89447 80.31000 87.24537 86.07523 83.88466 85.05075 84.71442 85.26494
## 4 80.53118 78.93966 77.74705 78.35543 80.51754 79.48920 82.04881 83.93940
## 5 80.84436 80.77175 81.03813 79.66289 80.63618 79.22085 82.10349 84.36140
## 6 82.57454 82.25581 81.69774 79.67575 81.36271 79.30994 82.09144 83.61829
##       X100     X125     X160     X200     X250     X315     X400     X500
## 1 79.38824 81.54961 78.93689 79.48846 78.81932 78.64437 79.75469 81.62373
## 2 79.82543 80.37480 77.68743 78.52243 78.83760 79.32380 79.97674 81.72101
## 3 79.20652 79.16684 77.70287 78.79541 78.66482 78.71364 80.05055 82.27485
## 4 77.52544 78.29855 77.62225 78.02077 77.77009 78.13745 79.80925 81.89413
## 5 77.11042 79.20388 78.33863 78.84368 78.97665 78.88854 80.21085 82.18003
## 6 77.45858 79.09539 78.81040 79.04204 78.84296 80.28467 79.81520 82.09605
##       X630     X800    X1000    X1250    X1600    X2000    X2500    X3150
## 1 83.34658 87.19071 82.14246 85.07748 94.14769 97.28539 98.69525 103.3008
## 2 83.81800 87.75812 82.62615 85.54004 94.49245 97.55348 98.85866 103.4371
## 3 83.97100 87.90292 82.51554 85.96418 94.60260 97.65596 99.18504 103.6927
## 4 83.82941 87.51515 82.08434 85.64691 94.25489 97.53704 98.96296 103.5570
## 5 84.04466 87.91065 82.21599 85.57978 94.29226 97.29830 98.82339 103.2960
## 6 83.87601 87.51786 81.75113 85.19473 94.07210 97.28325 98.72953 103.3153
##      X4000    X5000    X6300    X8000   X10000   X12500   X16000   X20000
## 1 105.1586 104.5767 106.0619 106.5193 109.4723 115.2509 117.7202 106.4402
## 2 105.3363 104.7928 106.2842 106.7141 109.7551 115.5190 117.8937 106.6128
```

```
## 3 105.6301 105.1628 106.6009 107.1454 110.1472 115.8672 118.1588 106.8864
## 4 105.3433 104.8958 106.3253 106.8269 109.7389 115.5147 117.8344 106.6686
## 5 105.2539 104.7206 106.2134 106.6074 109.6059 115.2975 117.7899 106.4953
## 6 105.2953 104.7295 106.2888 106.7906 109.6789 115.4043 117.8442 106.5558
##   X1010    rmsdB
## 1     0 120.9985
## 2     0 121.2132
## 3     0 121.5276
## 4     0 121.1921
## 5     0 121.0698
## 6     0 121.1486
```

I am now interested in pulling out all of the 1/3rd-octave band columns from the other columns in the spreadsheet. I will use these columns and the rms dB to run the PCA.

```
Lzeq1s<-Lzeq1[,c("X16","X20","X25","X16", "X20","X25","X31_5","X40", "X50","X63","X80","X100","X125","X
```

Check and make sure the columns being used in the PCA are all numeric.

```
sapply(Lzeq1s,is.numeric)
```

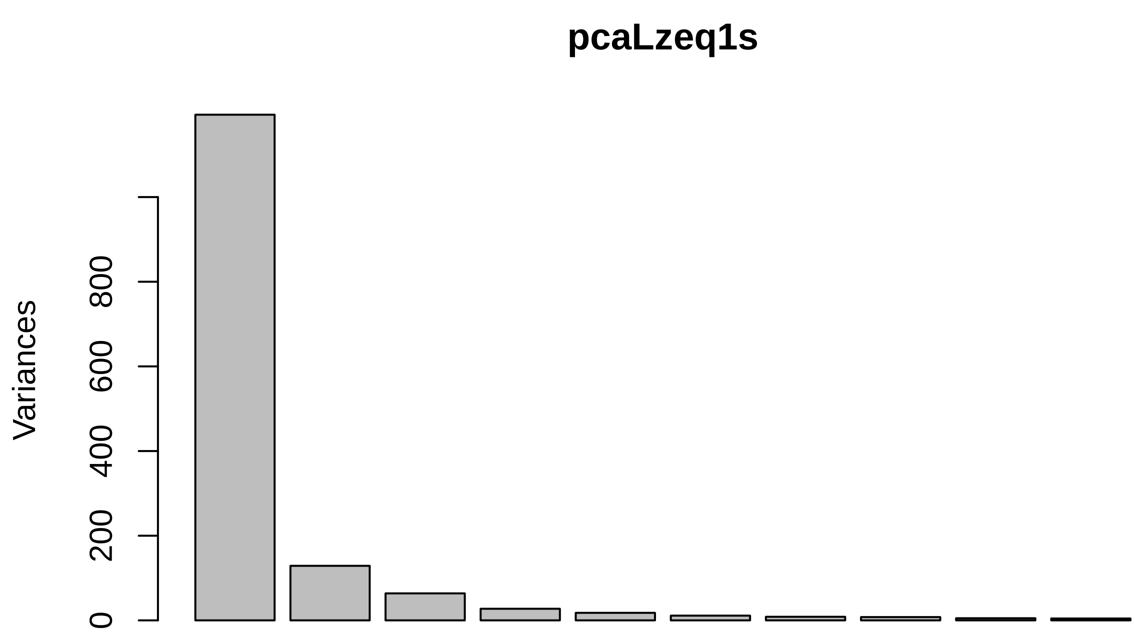
```
##    X16    X20    X25  X16.1  X20.1  X25.1  X31_5    X40    X50    X63
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
##    X80   X100   X125   X160   X200   X250   X315   X400   X500   X630
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
##   X800  X1000  X1250  X1600  X2000  X2500  X3150  X4000  X5000  X6300
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
##  X8000 X10000 X12500 X16000 X20000  rmsdB
##   TRUE   TRUE   TRUE   TRUE   TRUE   TRUE
```

Run the PCA using prcomp.

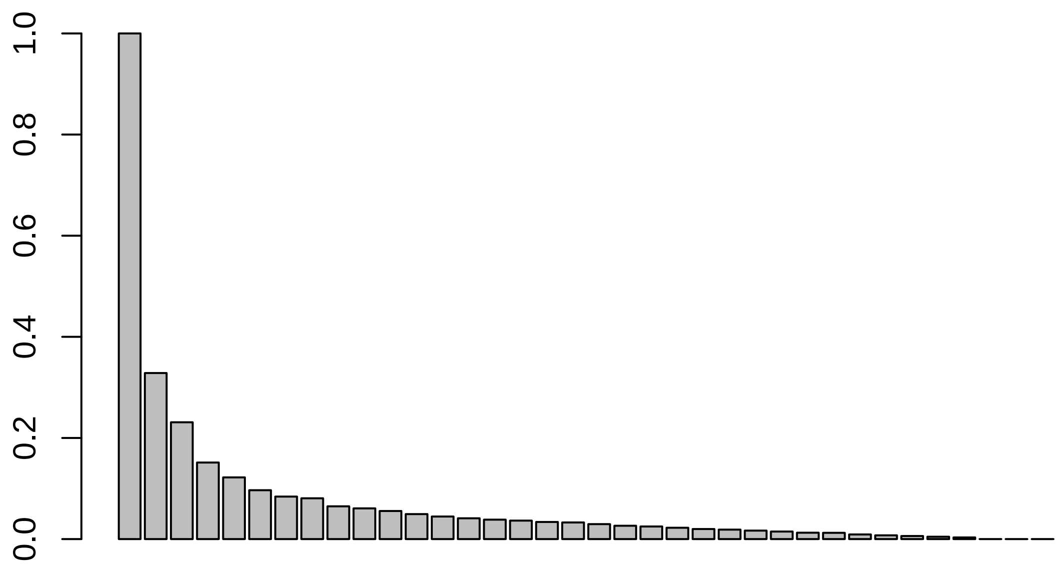
```
pcaLzeq1s<-prcomp(Lzeq1s)
```

I now want to look at the results of the PCA and make some plots to visualize those results. I used http://stackoverflow.com/questions/22403753/dots-instead-of-labels-for-biplot-prcomp to remove the labels from the biplots and to make each data point a dot. I also used calls to cex to reduce the size of the labels for easier reading.

```r
plot(pcaLzeq1s)
```




```r
barplot(pcaLzeq1s$sdev/pcaLzeq1s$sdev[1])
```

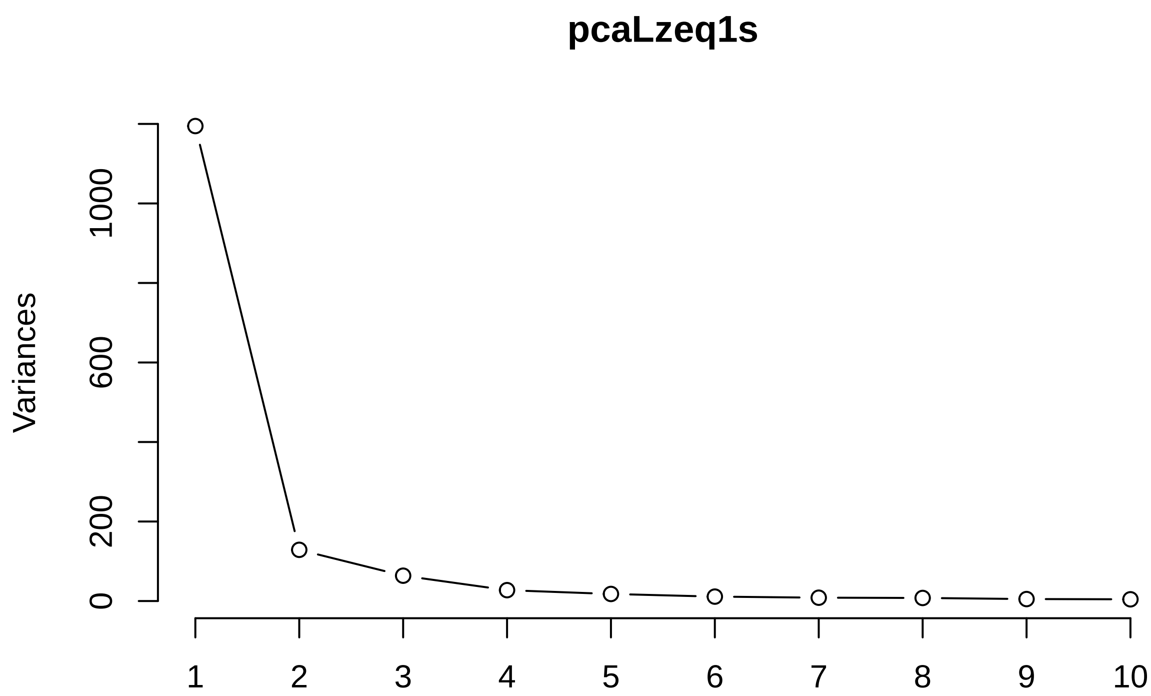



```r
summary(pcaLzeq1s)
```

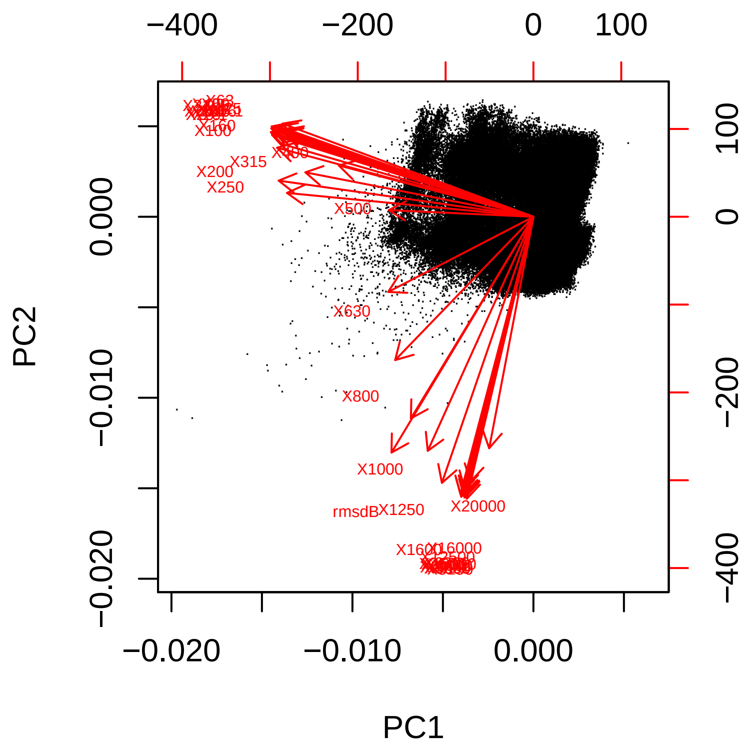
```
## Importance of components:
##                            PC1      PC2     PC3     PC4     PC5     PC6
## Standard deviation     34.5630 11.34961 7.98539 5.23216 4.21839 3.34147
## Proportion of Variance  0.8009  0.08636 0.04275 0.01835 0.01193 0.00749
## Cumulative Proportion   0.8009  0.88729 0.93004 0.94840 0.96033 0.96781
##                            PC7     PC8     PC9    PC10    PC11    PC12
## Standard deviation     2.90606 2.78797 2.23930 2.09946 1.91980 1.70726
## Proportion of Variance 0.00566 0.00521 0.00336 0.00296 0.00247 0.00195
## Cumulative Proportion  0.97348 0.97869 0.98205 0.98500 0.98748 0.98943
##                           PC13    PC14    PC15    PC16    PC17    PC18
## Standard deviation     1.54148 1.42227 1.32959 1.26382 1.16961 1.13797
## Proportion of Variance 0.00159 0.00136 0.00119 0.00107 0.00092 0.00087
## Cumulative Proportion  0.99102 0.99238 0.99356 0.99464 0.99555 0.99642
##                          PC19    PC20   PC21   PC22    PC23    PC24
## Standard deviation     1.0217 0.91213 0.8621 0.7754 0.68763 0.64766
## Proportion of Variance 0.0007 0.00056 0.0005 0.0004 0.00032 0.00028
## Cumulative Proportion  0.9971 0.99768 0.9982 0.9986 0.99890 0.99918
##                           PC25    PC26    PC27    PC28    PC29    PC30
## Standard deviation     0.57933 0.51621 0.43651 0.42862 0.31800 0.25677
## Proportion of Variance 0.00023 0.00018 0.00013 0.00012 0.00007 0.00004
## Cumulative Proportion  0.99940 0.99958 0.99971 0.99983 0.99990 0.99994
##                           PC31    PC32    PC33      PC34      PC35
## Standard deviation     0.21049 0.16095 0.11261 8.021e-14 1.682e-14
```

```
## Proportion of Variance 0.00003 0.00002 0.00001 0.000e+00 0.000e+00
## Cumulative Proportion  0.99997 0.99999 1.00000 1.000e+00 1.000e+00
##                             PC36
## Standard deviation     1.499e-14
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

```
plot(pcaLzeq1s,type="lines")
```




```
biplot(prcomp(Lzeq1s,scale.=T),xlabs=rep(".",nrow(Lzeq1s)),cex=.5)
```

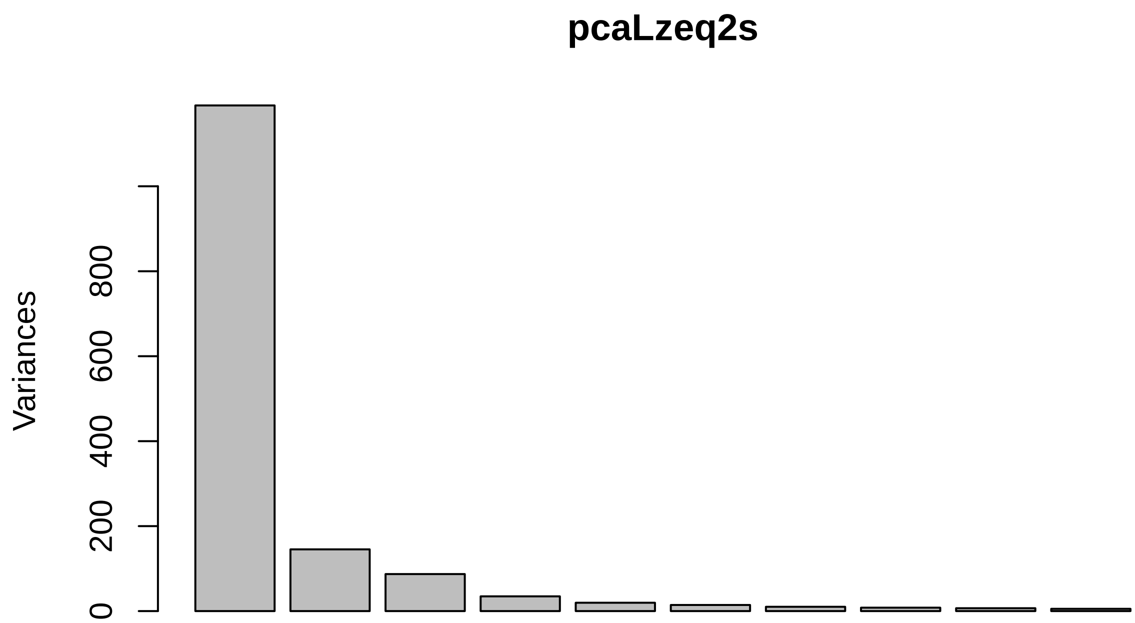


Bay 2: Honaunau

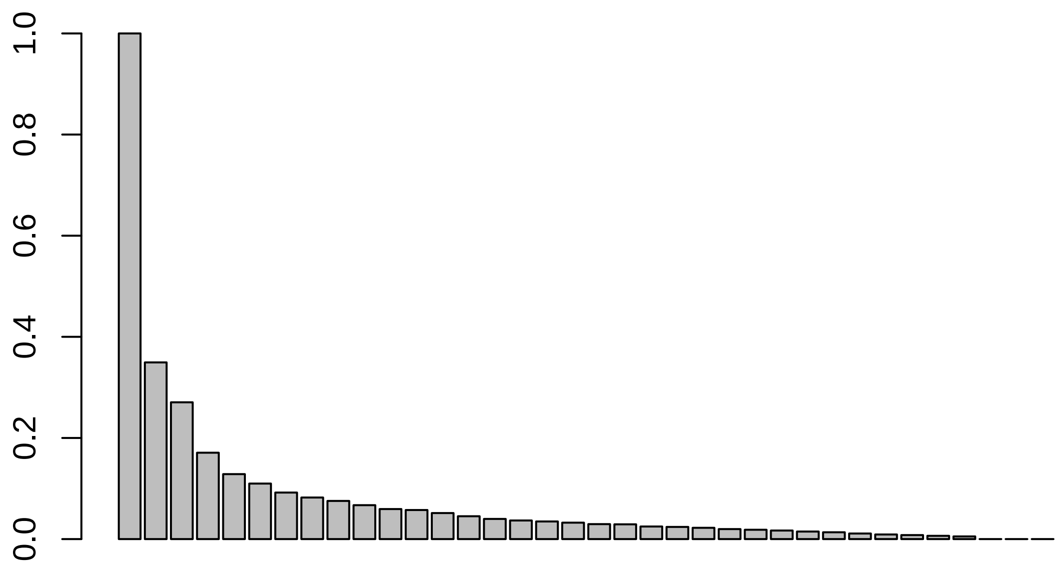
Repeat the process for Bay 2. I will only show the resulting plots and not the code since it was the same as Bay 1.

Plots

```
plot(pcaLzeq2s)
```




```r
barplot(pcaLzeq2s$sdev/pcaLzeq2s$sdev[1])
```

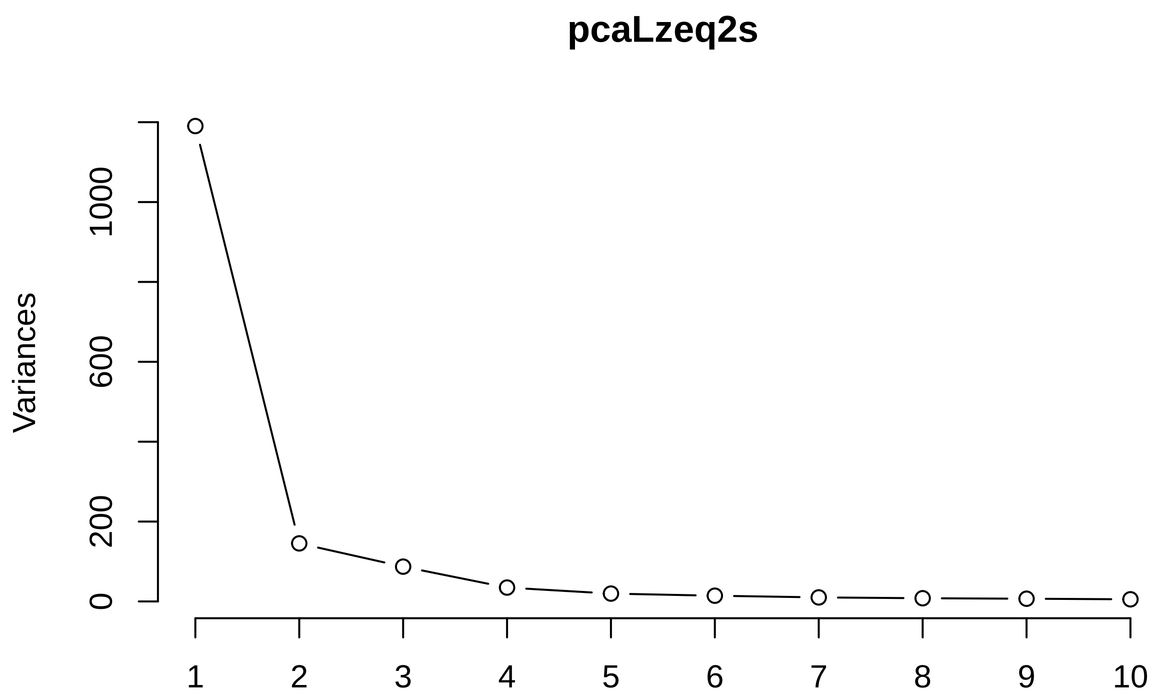



```
summary(pcaLzeq2s)
```

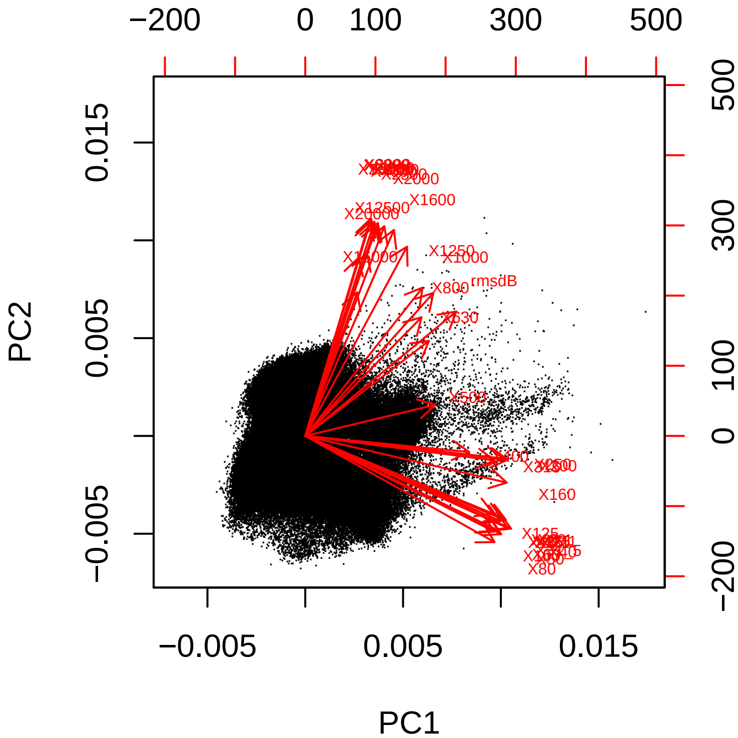
```
## Importance of components:
##                            PC1     PC2     PC3     PC4     PC5     PC6
## Standard deviation     34.501 12.0554 9.33347 5.89227 4.43195 3.78788
## Proportion of Variance  0.769  0.0939 0.05628 0.02243 0.01269 0.00927
## Cumulative Proportion   0.769  0.8629 0.91920 0.94163 0.95432 0.96359
##                            PC7     PC8     PC9    PC10    PC11    PC12
## Standard deviation     3.17466 2.83558 2.60422 2.31495 2.04917 1.98311
## Proportion of Variance 0.00651 0.00519 0.00438 0.00346 0.00271 0.00254
## Cumulative Proportion  0.97010 0.97530 0.97968 0.98314 0.98585 0.98839
##                           PC13    PC14    PC15    PC16    PC17    PC18
## Standard deviation     1.77638 1.55996 1.37507 1.26934 1.20244 1.11998
## Proportion of Variance 0.00204 0.00157 0.00122 0.00104 0.00093 0.00081
## Cumulative Proportion  0.99043 0.99201 0.99323 0.99427 0.99520 0.99601
##                           PC19    PC20    PC21    PC22    PC23   PC24
## Standard deviation     1.02135 1.00848 0.85954 0.82984 0.76911 0.6815
## Proportion of Variance 0.00067 0.00066 0.00048 0.00044 0.00038 0.0003
## Cumulative Proportion  0.99669 0.99734 0.99782 0.99827 0.99865 0.9990
##                           PC25    PC26    PC27    PC28    PC29    PC30
## Standard deviation     0.63593 0.58390 0.51515 0.46554 0.38096 0.31413
## Proportion of Variance 0.00026 0.00022 0.00017 0.00014 0.00009 0.00006
## Cumulative Proportion  0.99921 0.99943 0.99960 0.99974 0.99983 0.99990
##                           PC31    PC32    PC33     PC34      PC35
## Standard deviation     0.27293 0.22281 0.18157 1.26e-13 7.357e-14
```

```
## Proportion of Variance 0.00005 0.00003 0.00002 0.00e+00 0.000e+00
## Cumulative Proportion  0.99995 0.99998 1.00000 1.00e+00 1.000e+00
##                             PC36
## Standard deviation     1.045e-14
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

```r
plot(pcaLzeq2s,type="lines")
```




```r
biplot(prcomp(Lzeq2s,scale.=T),xlabs=rep(".",nrow(Lzeq2s)),cex=.5)
```

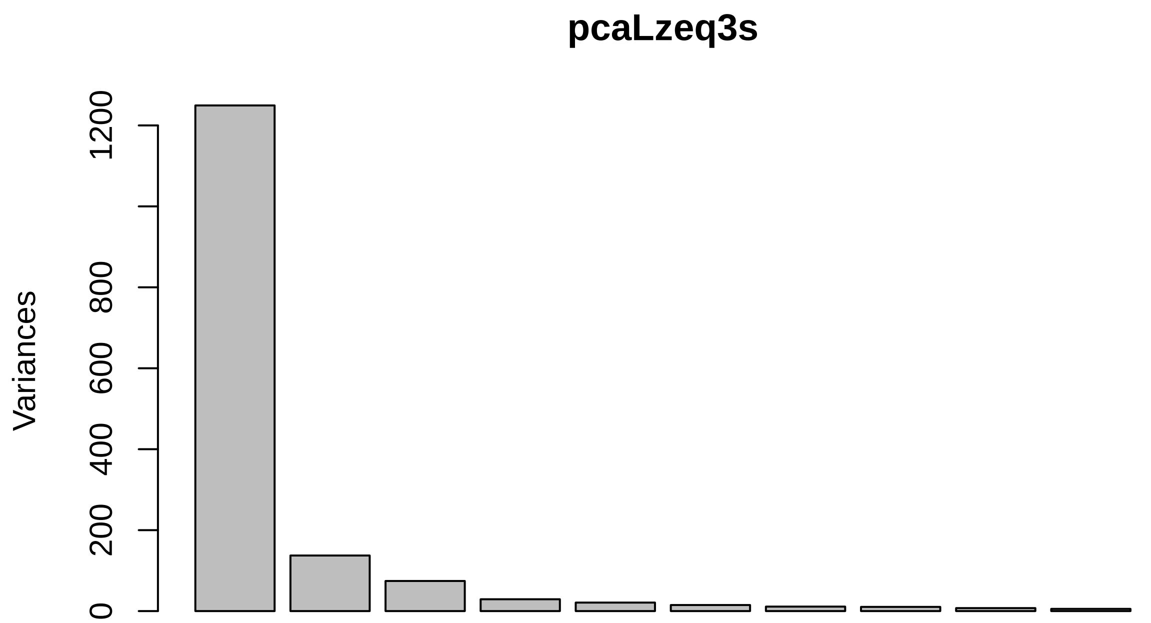


Bay 3: Kealakekua

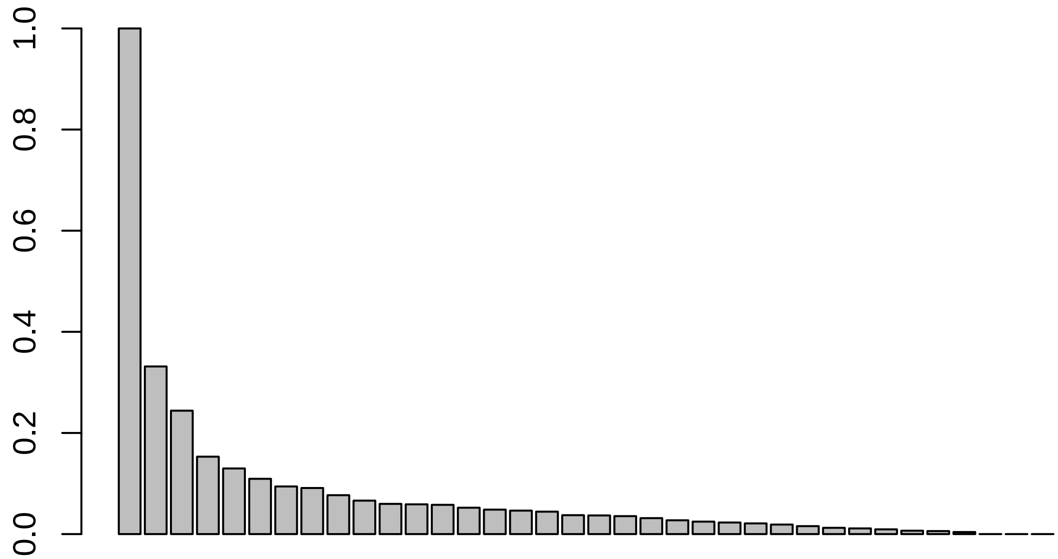
Repeat the process for Bay 3. I will only show the resulting plots and not the code since it was the same as Bay 1.

Plots

```
plot(pcaLzeq3s)
```




```r
barplot(pcaLzeq3s$sdev/pcaLzeq3s$sdev[1])
```

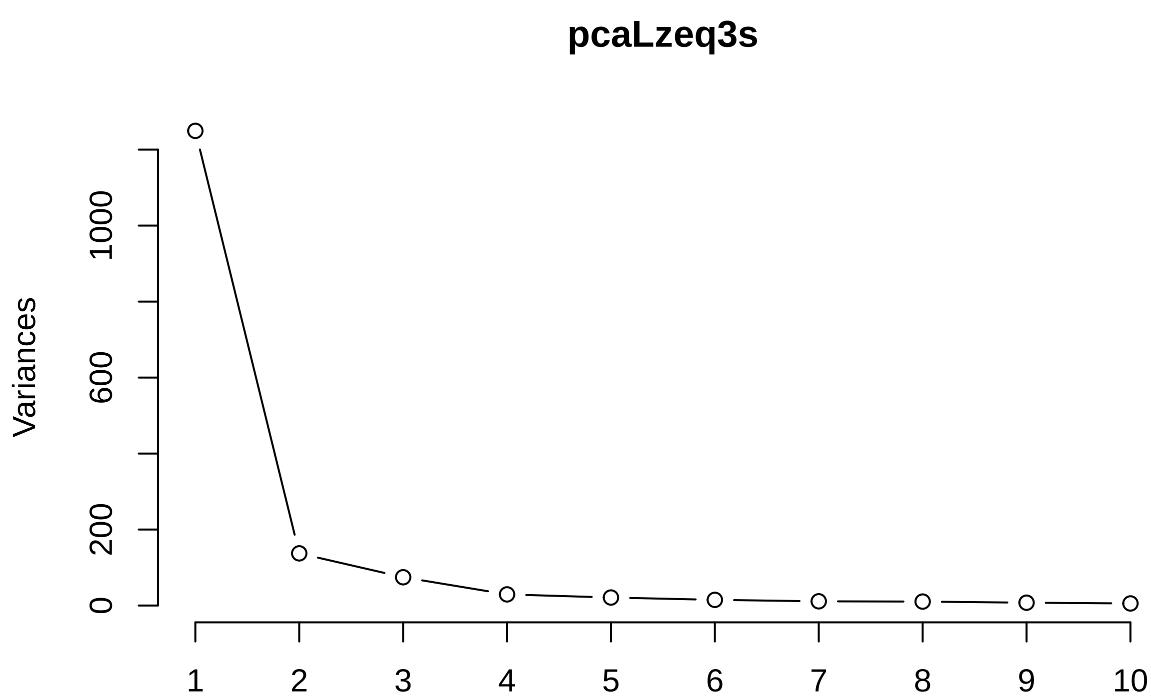



```r
summary(pcaLzeq3s)
```

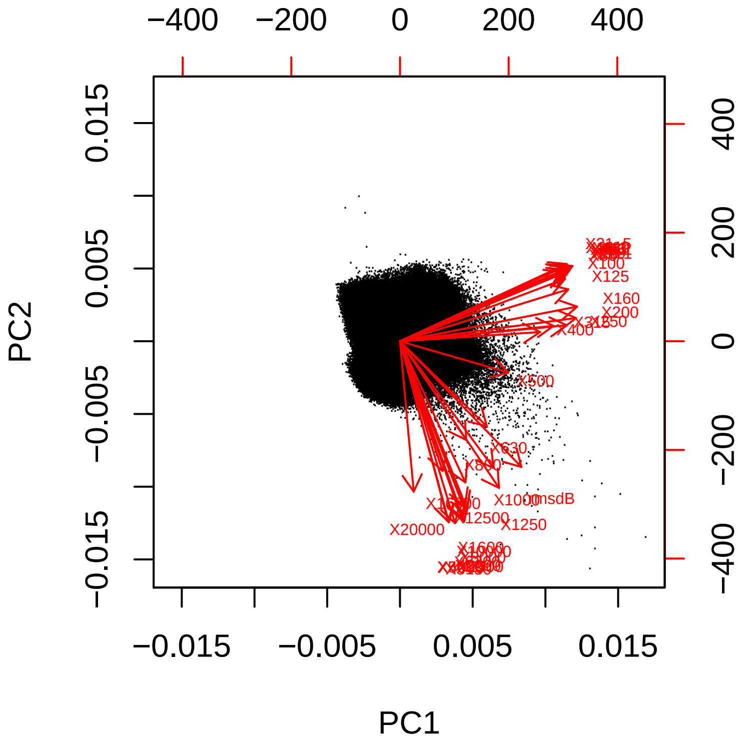
```
## Importance of components:
##                           PC1      PC2     PC3     PC4     PC5     PC6
## Standard deviation     35.346 11.71762 8.62604 5.41058 4.58777 3.86372
## Proportion of Variance  0.783  0.08605 0.04663 0.01835 0.01319 0.00936
## Cumulative Proportion   0.783  0.86907 0.91570 0.93405 0.94724 0.95660
##                            PC7    PC8     PC9    PC10    PC11    PC12
## Standard deviation     3.32875 3.2201 2.71919 2.33931 2.11171 2.08114
## Proportion of Variance 0.00694 0.0065 0.00463 0.00343 0.00279 0.00271
## Cumulative Proportion  0.96354 0.9700 0.97467 0.97810 0.98090 0.98361
##                           PC13    PC14    PC15    PC16    PC17   PC18
## Standard deviation     2.04567 1.84305 1.70904 1.63942 1.56821 1.3225
## Proportion of Variance 0.00262 0.00213 0.00183 0.00168 0.00154 0.0011
## Cumulative Proportion  0.98624 0.98837 0.99020 0.99188 0.99342 0.9945
##                           PC19    PC20    PC21    PC22    PC23    PC24
## Standard deviation     1.30245 1.25497 1.11398 0.96636 0.87036 0.80854
## Proportion of Variance 0.00106 0.00099 0.00078 0.00059 0.00047 0.00041
## Cumulative Proportion  0.99558 0.99657 0.99735 0.99793 0.99841 0.99882
##                           PC25    PC26    PC27    PC28   PC29    PC30
## Standard deviation     0.74807 0.66564 0.55477 0.43743 0.3962 0.33173
## Proportion of Variance 0.00035 0.00028 0.00019 0.00012 0.0001 0.00007
## Cumulative Proportion  0.99917 0.99944 0.99964 0.99976 0.9999 0.99992
##                           PC31    PC32    PC33      PC34      PC35
## Standard deviation     0.23744 0.21093 0.14303 3.625e-13 5.116e-14
```

```
## Proportion of Variance 0.00004 0.00003 0.00001 0.000e+00 0.000e+00
## Cumulative Proportion  0.99996 0.99999 1.00000 1.000e+00 1.000e+00
##                             PC36
## Standard deviation     2.117e-14
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

```
plot(pcaLzeq3s,type="lines")
```




```
biplot(prcomp(Lzeq3s,scale.=T),xlabs=rep(".",nrow(Lzeq3s)),cex=.5)
```

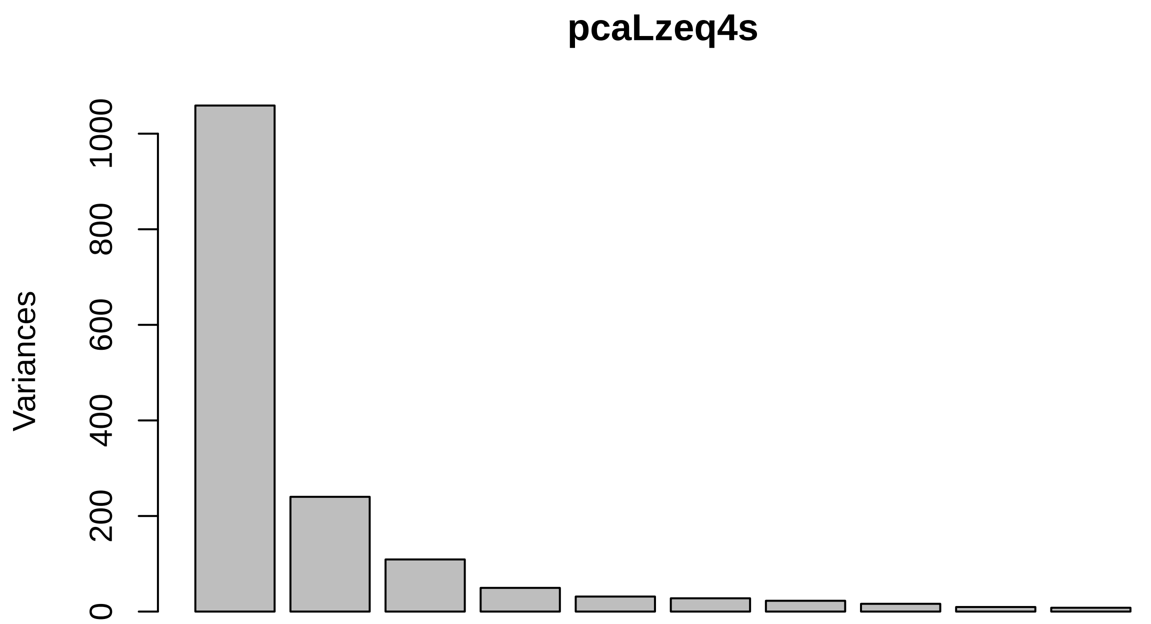


Bay 4: Makako

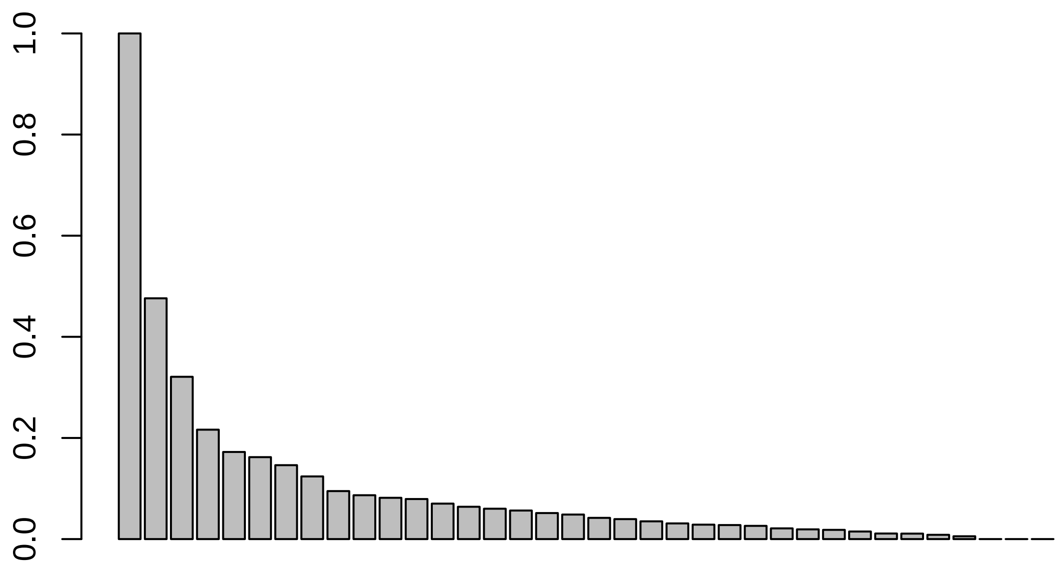
Repeat the process for Bay 4. I will only show the resulting plots and not the code since it was the same as Bay 1.

Plots

```
plot(pcaLzeq4s)
```




```r
barplot(pcaLzeq4s$sdev/pcaLzeq4s$sdev[1])
```

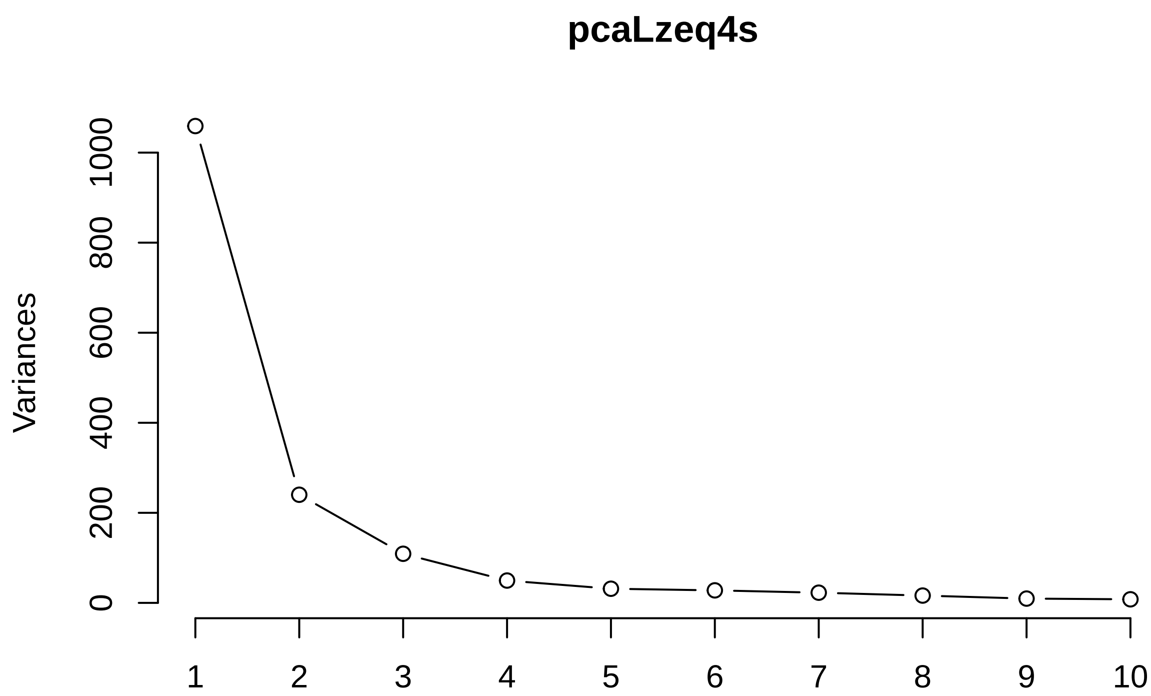



```r
summary(pcaLzeq4s)
```

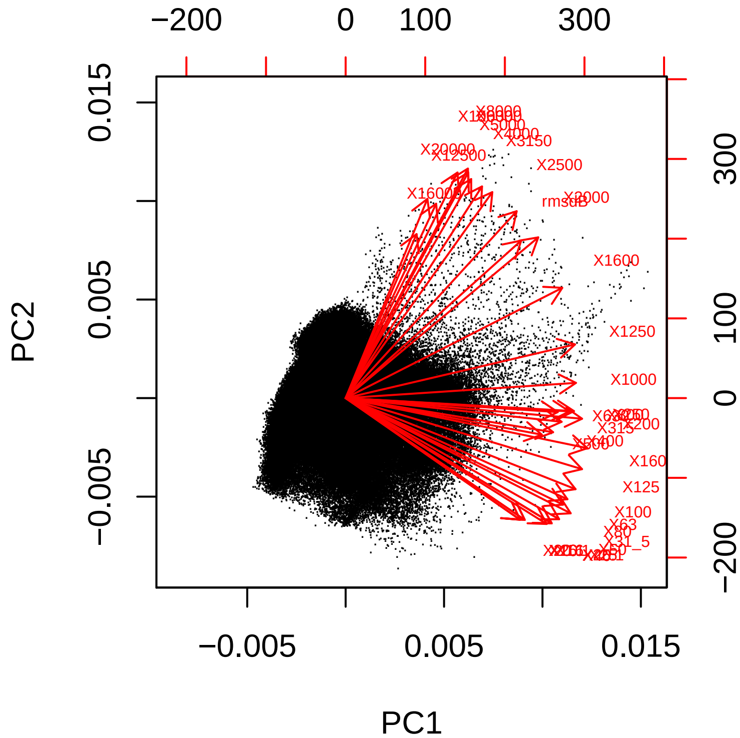
```
## Importance of components:
##                            PC1     PC2      PC3     PC4     PC5     PC6
## Standard deviation     32.541 15.4967 10.44379 7.04166 5.60610 5.27325
## Proportion of Variance  0.654  0.1483  0.06736 0.03062 0.01941 0.01717
## Cumulative Proportion   0.654  0.8023  0.86968 0.90031 0.91972 0.93689
##                            PC7     PC8     PC9    PC10    PC11   PC12
## Standard deviation     4.75676 4.03238 3.08851 2.82280 2.65582 2.5769
## Proportion of Variance 0.01397 0.01004 0.00589 0.00492 0.00436 0.0041
## Cumulative Proportion  0.95086 0.96091 0.96680 0.97172 0.97607 0.9802
##                           PC13    PC14    PC15    PC16    PC17    PC18
## Standard deviation     2.28317 2.08038 1.95413 1.83770 1.67462 1.57831
## Proportion of Variance 0.00322 0.00267 0.00236 0.00209 0.00173 0.00154
## Cumulative Proportion  0.98340 0.98607 0.98843 0.99051 0.99224 0.99378
##                           PC19    PC20    PC21    PC22    PC23   PC24
## Standard deviation     1.36629 1.28547 1.14319 1.00944 0.92847 0.9023
## Proportion of Variance 0.00115 0.00102 0.00081 0.00063 0.00053 0.0005
## Cumulative Proportion  0.99494 0.99596 0.99676 0.99739 0.99793 0.9984
##                           PC25    PC26    PC27    PC28    PC29    PC30
## Standard deviation     0.85155 0.69043 0.62608 0.59334 0.48731 0.35835
## Proportion of Variance 0.00045 0.00029 0.00024 0.00022 0.00015 0.00008
## Cumulative Proportion  0.99888 0.99917 0.99941 0.99963 0.99978 0.99986
##                           PC31    PC32    PC33      PC34      PC35
## Standard deviation     0.35222 0.27536 0.18267 1.025e-13 3.179e-14
```

```
## Proportion of Variance 0.00008 0.00005 0.00002 0.000e+00 0.000e+00
## Cumulative Proportion  0.99993 0.99998 1.00000 1.000e+00 1.000e+00
##                             PC36
## Standard deviation     9.071e-15
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

```r
plot(pcaLzeq4s,type="lines")
```




```r
biplot(prcomp(Lzeq4s,scale.=T),xlabs=rep(".",nrow(Lzeq4s)),cex=.5)
```



PC1
PC2
X20000
X12500
X16000
X5000
X4000
X3150
X2500
rmsdB
X2000
X1600
X1250
X1000
X200
X160
X125
X100
X63
X80
X31_5