PCA Analysis

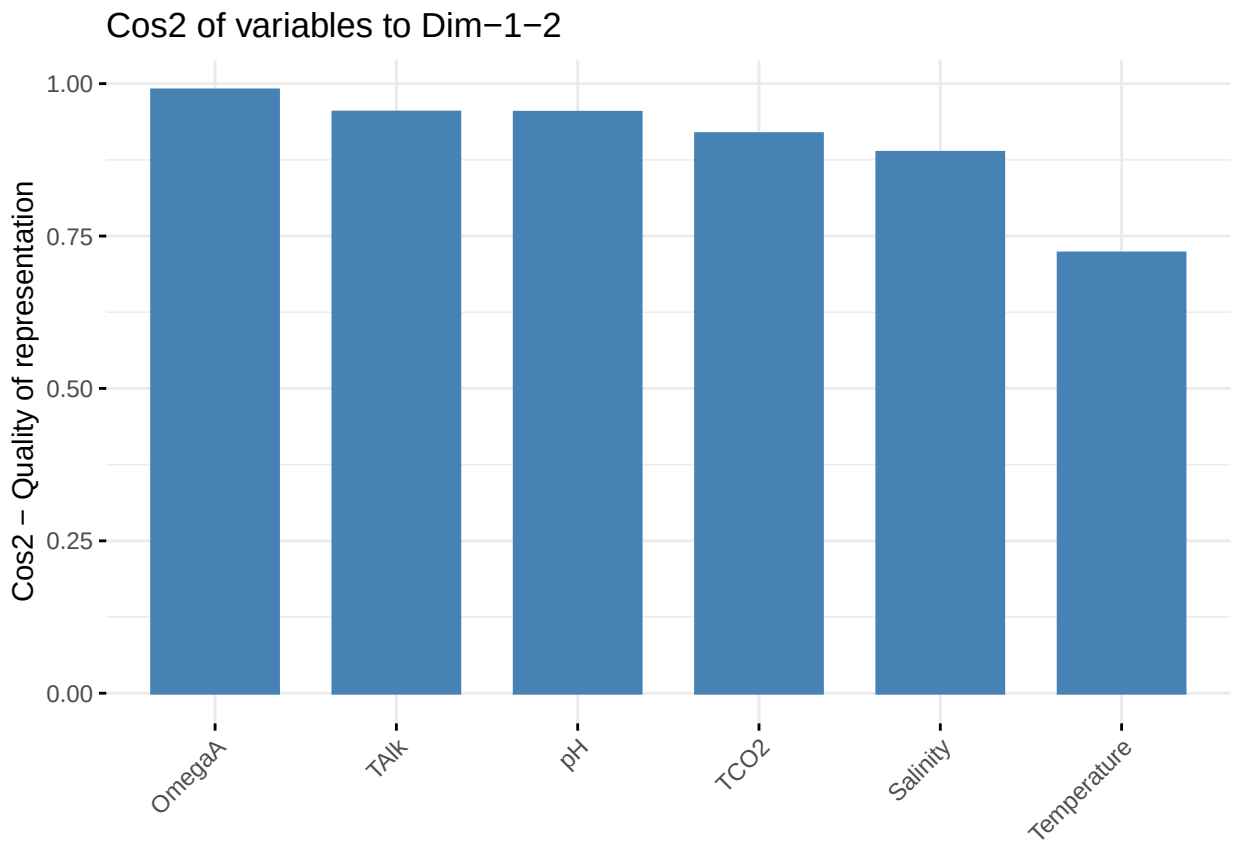
S. Eanes

11/25/2020

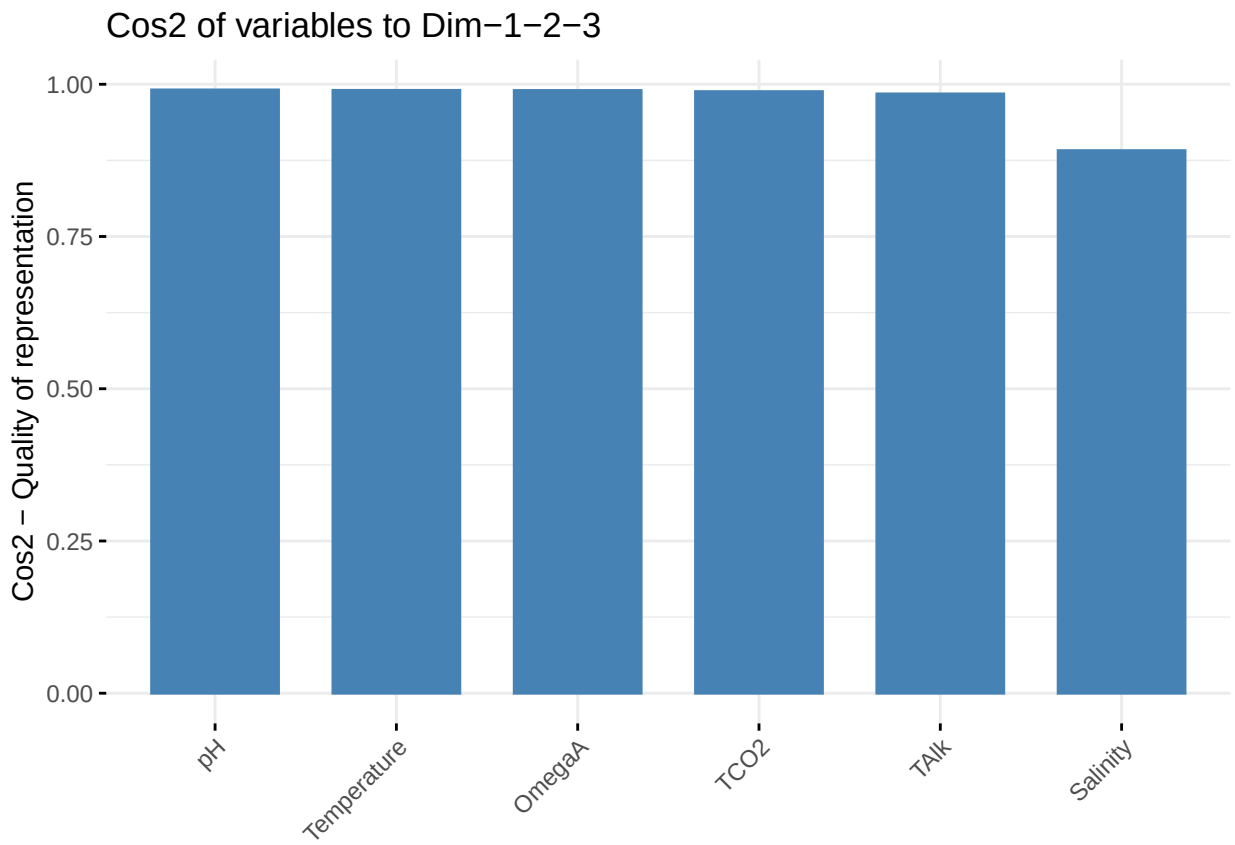
```
coral <- read.csv("C:/Users/shado/Documents/School/Fossil Coral/data/coral_3weighted.csv")

#omit data with no response
coral <- coral[!is.na(coral$U238),]

#omit data with no response
coral <- coral[!is.na(coral$U238),]

#select three largest species
genus_trim <- c("Acropora", "Porites")
#remove coral with age > 10
coral <- coral[which(coral$Genus %in% genus_trim),]  %>% dplyr::filter(Age < 10)
coral <- coral[(coral$Calcite <= 1 | is.na(coral$Calcite)),]

#clean up a nice dataframe
coral.df <- coral %>% mutate(Temperature = Temp, U238=U238*.421) %>%
  select(pH,TAlk,Salinity,Temperature,OmegaA,TCO2) %>%
  data.frame

nrow(coral.df)
```

```
## [1] 700
```

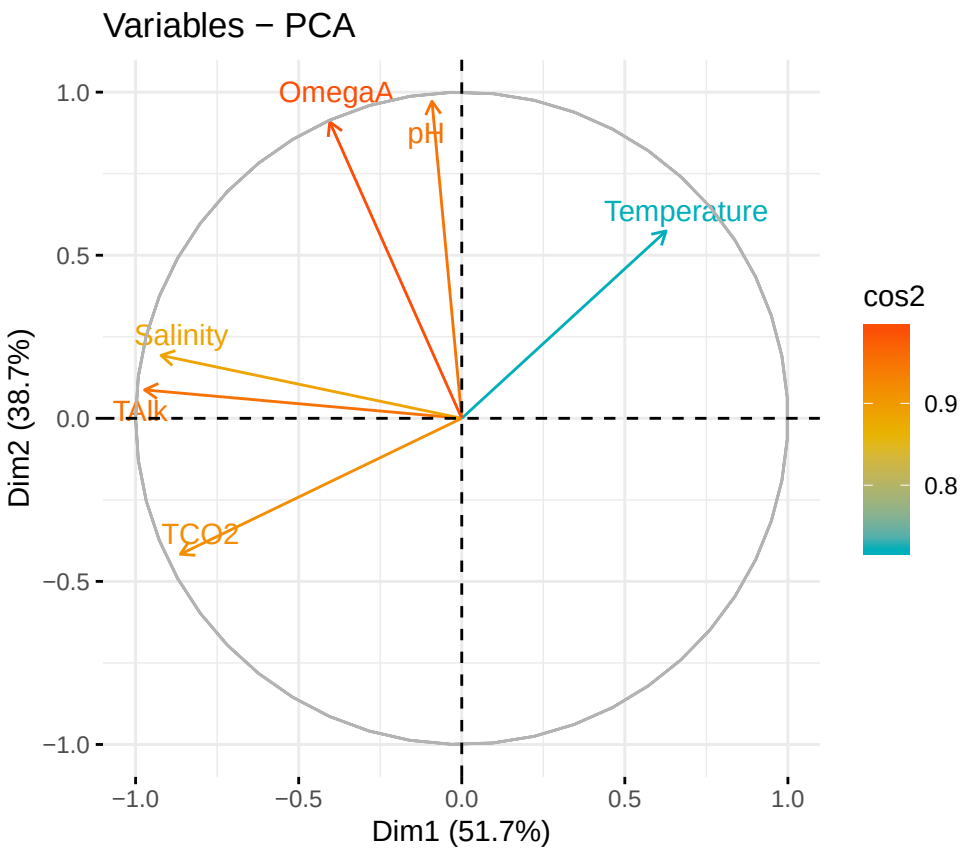
```
pca <- prcomp(coral.df,scale=T)
#in the first two dimensions temp has the lowest total cos2
fviz_cos2(pca, choice = "var", axes = 1:2)
```



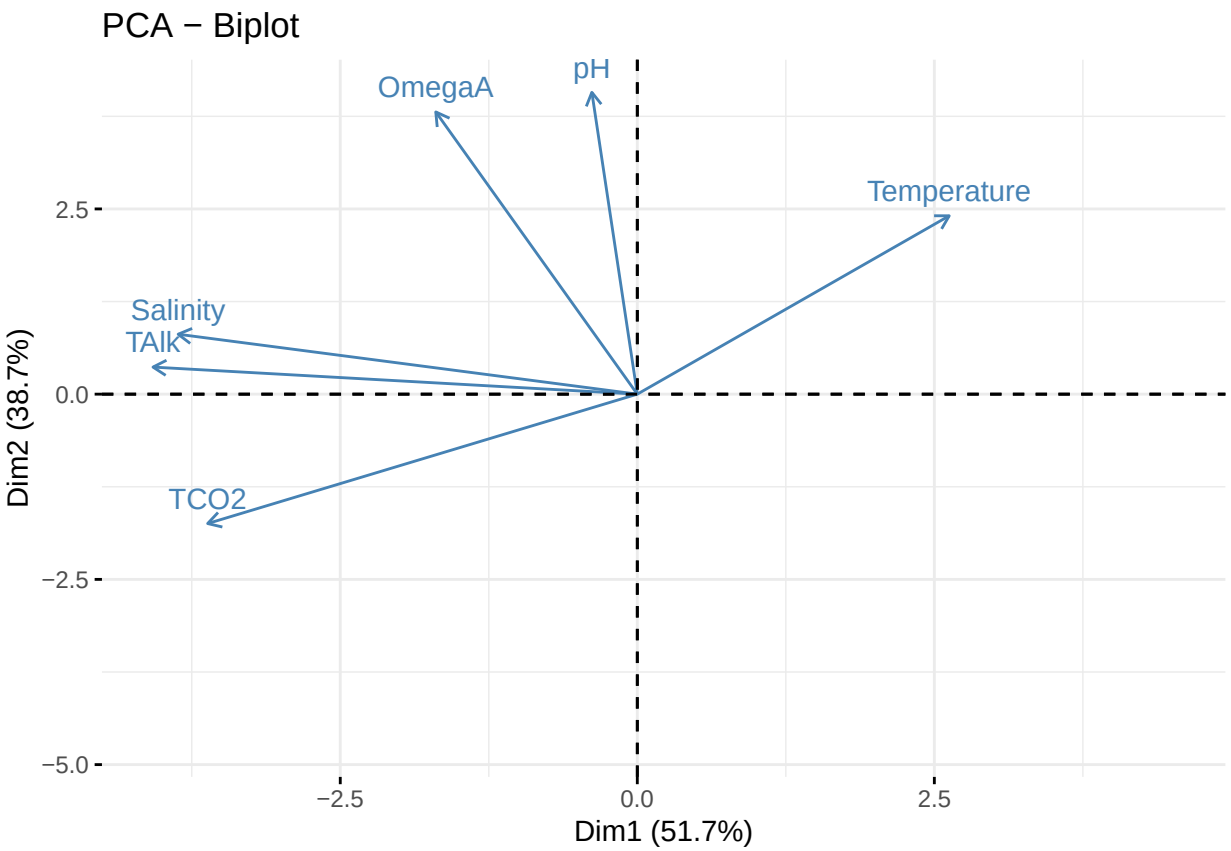

```
#it jumps up quite a bit when including the third pca dim
fviz_cos2(pca, choice = "var", axes = 1:3)
```




```
fviz_pca_var(pca, col.var="cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)
```




```
fviz_pca_biplot(pca, label ="var", invisible="ind")
```




```
pca$sdev
```

```
## [1] 1.76204659 1.52284098 0.64015376 0.39516319 0.08080561 0.06055450
```

```
pr.var <- pca$sdev^2
(PVE <- pr.var / sum(pr.var))
```

```
## [1] 0.5174680281 0.3865074420 0.0682994730 0.0260256580 0.0010882577
## [6] 0.0006111412
```

```
x <- 1:length(pr.var)
dat <- data.frame(x,PVE)
cumulative <- numeric(nrow(dat))
cumulative[1] <- dat$PVE[1]
for(i in 2:nrow(dat)){
  cumulative[i] <- dat$PVE[i]+cumulative[i-1]
}
dat$cumulative <- cumulative

coral.rot <- as.matrix(coral.df) %*% pca$rotation
coral.rot.df <- data.frame(coral.rot)
dim(coral.rot.df)
```

```
## [1] 700   6
```

```
scree <- ggplot(dat,aes(x,PVE))+
  geom_point()+
  geom_line()+
  xlab("Principal Component")+
  ggtitle("Scree Plot")+
  ylab("PVE")+ylim(0,1)

cumScree <- ggplot(dat,aes(x,cumulative))+
  geom_point()+
  geom_line()+
  ylab(NULL)+
  xlab("Principal Component")+
  ggtitle("Cumulative Scree Plot")+ylim(0,1)

grid.arrange(scree,cumScree,ncol=2)
```