Supplementary material

Contents

Supplementary Material 1

R script for PCA analysis of database using Hotelling, ggbiplot, cowplot, vegan, and factoextra packages, followed by PCA %contribution plot script.

N.B. text in green italics should be edited to your file names and location, "#" indicates instructions, not code.

Setup

Load packages and data files

```
library(Hotelling)
library(ggbiplot)
library(cowplot)
library(vegan)
library(factoextra)
```

```
setwd("C:/Users/abhre/OneDrive/Documents/Work/Active/To read/TephraNZ")
tephranz <- read.csv("TephraNZ_all_2.csv", fileEncoding="UTF-8-BOM")
str(tephranz)
```

```
## 'data.frame':    47 obs. of  47 variables:
##  $ Tephra.name: chr  "Eq F Ig" "Potaka" "Rewa" "Mangapipi" ...
##  $ Source     : chr  "KVC" "MgVC" "MgVC" "MgVC" ...
##  $ Si_ppm     : int  350450 342257 333346 329371 342687 350717 355250 362455 363111 357981 ...
##  $ Ti_ppm     : int  811 792 1072 1145 541 514 997 688 724 684 ...
##  $ Al_ppm     : int  63668 63117 66995 68731 63074 64580 66097 65916 66156 65071 ...
...
```

Perform PCA

Here, the character columns of `tephranz` have been removed, and the elemental data have been transformed using centered log-ratio. We ran a PCA on the centered and scaled transformed data.

```
tephranz.elements <- tephranz[, -c(1:2)]
tephranz.trans<- clr(tephranz.elements)

tephranz.pca <- prcomp(tephranz.trans, center = TRUE, scale. = TRUE)
```

Plot and export PCA

```
tephranz.plot <- ggbiplot(tephranz.pca, ellipse = TRUE,
                          labels = rownames(tephranz),
                          labels.size = 2, groups = tephranz$Source) +
  theme_bw() +
  guides(color = guide_legend("Tephra", override.aes = list(label = "")))
tephranz.plot
```
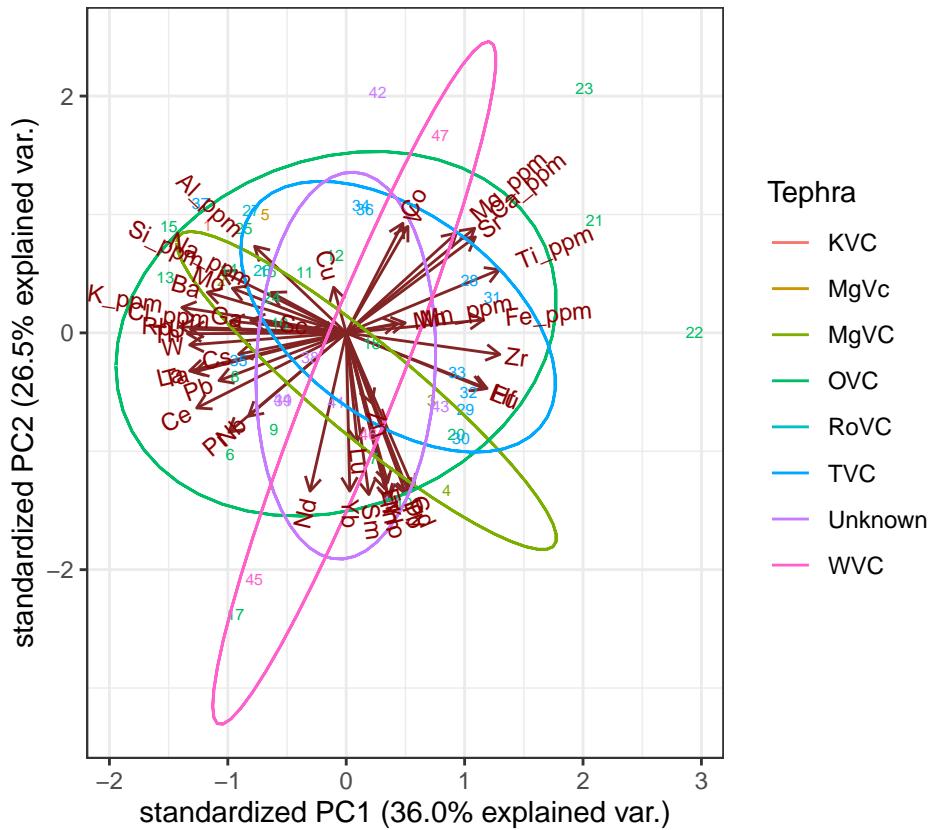
```
ggsave("TephraNZ_PCA.png", width = 25, height = 15, units = "cm", dpi = 600)
```

Contribution to PCA axes

```
contrib1 <- fviz_contrib(tephranz.pca, choice = "var", axes = 1, top = 10) +
  ggtitle("Contributions to PC1")

contrib2 <- fviz_contrib(tephranz.pca, choice = "var", axes = 2, top = 10) +
  ggtitle("Contributions to PC2") +
  ylab(NULL)

contributions <- plot_grid(
  contrib1,
  NULL,
  contrib2,
  align = 'vh',
  labels = c("A", "", "B"),
  hjust = -1,
  nrow = 1,
  rel_widths = c(1, .2, 1)
)
contributions
```
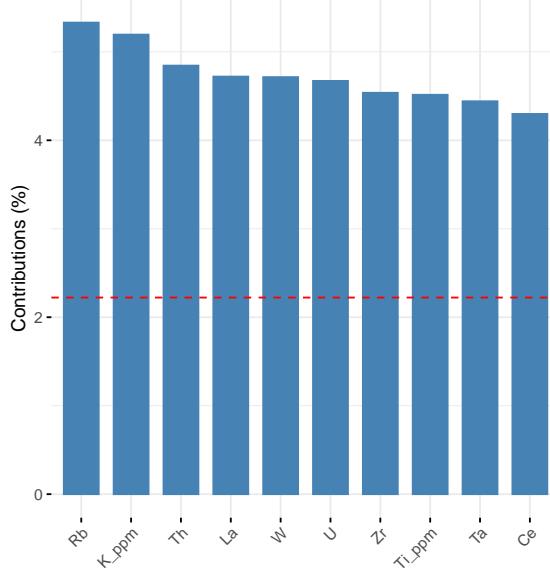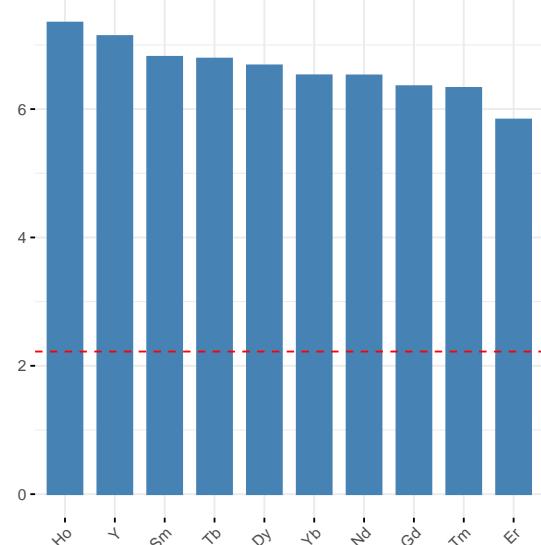


```
save_plot("Contributions plot.png", contributions, ncol = 3, base_asp = 1, dpi = 600)
```

Supplementary Material 2

R script for Euclidean similarity coefficients analysis

N.B. text in green italics should be edited to your file names and location, "#" indicates instructions, not code.

Setup

```
library(stats)
setwd("C:/Users/abhre/OneDrive/Documents/Work/Active/To read/TephraNZ")
tephranz.cor <- read.csv("TephraNZ_all_2.csv", fileEncoding="UTF-8-BOM")
```

Subset the data frame by selecting element columns of interest and sample rows of interest.

```
elements.keep <- c("Si_ppm", "Ti_ppm", "Al_ppm", "Fe_ppm", "Mn_ppm", "Na_ppm", "K_ppm")
tephranz.major <- (tephranz.cor[, elements.keep])
str(tephranz.major)
```

```
## 'data.frame':    47 obs. of  7 variables:
##  $ Si_ppm: int  350450 342257 333346 329371 342687 350717 355250 362455 363111 357981 ...
##  $ Ti_ppm: int  811 792 1072 1145 541 514 997 688 724 684 ...
##  $ Al_ppm: int  63668 63117 66995 68731 63074 64580 66097 65916 66156 65071 ...
##  $ Fe_ppm: int  6720 9078 14264 16688 9720 6390 11010 6752 6750 7325 ...
##  $ Mn_ppm: int  394 198 302 376 215 524 459 315 391 448 ...
##  $ Na_ppm: int  26226 25052 28161 29300 24725 26825 27469 26528 27518 28300 ...
##  $ K_ppm : int  26714 28593 25983 27288 26171 31911 26697 29546 28466 27006 ...
```

Run Euclidean Similarity Coefficient analysis and export results.

```
tephranz.major.dist <- as.matrix(dist(tephranz.major, method = "euclidean"))
write.csv(tephranz.major.dist, "TephraNZ Euclidean Similarity.csv")
```