# 08- PCA + Pearson, p450

## Setup

```
knitr::opts_chunk$set(
  echo = TRUE,          # Display code chunks
  eval = TRUE,          # Evaluate code chunks
  warning = FALSE,      # Hide warnings
  message = FALSE,      # Hide messages
  fig.width = 8,        # Set plot width in inches
  fig.height = 5,       # Set plot height in inches
  fig.align = "center" # Align plots to the center
)
```

## Load packages

```
library(tidyr)
library(tidyverse)
library(ggplot2)
library(vegan)
#library(tinytex)
```

## Load data

**Note:**

For *data* the units are listed below. Weight = g
Length, width, height = mm
p450, SOD = activity/ (mg/protein)
Condition factor, economic factor = unitless
For *pah*, *indv*, and *allana* the units are ng/g
For *metal* the units are mg/kg

```
getwd()
```

```
## [1] "/Users/cmantegna/Documents/WDFWmussels/code"
```

```
#data has all sites, coordinates, p450, sod, condition factor, economic factor data
data<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/biomarkerfull.csv")

#pah has complete site values and different summed pah analyte whole tissue values
pah<- read.csv("/Users/cmantegna/Documents/Biomarker Data Analysis/sum_analytes.csv")

#indv has complete site values and individual named pah analyte whole tissue values
indv<- read.csv("/Users/cmantegna/Documents/Biomarker Data Analysis/individual_analytes.csv")

metal<- read.csv("/Users/cmantegna/Documents/Biomarker Data Analysis/metal.csv")
```

```
allana<- read.csv("/Users/cmantegna/Documents/Biomarker Data Analysis/allana.csv")
```

```
# Review data frame structure

#str(metal)
#str(allana)
#str(indv)
```

```
# Review basic data types and stats

#summary(data)
#summary(pah)
#summary(indv)
```

```
head(data)
```

```
##   latitude longitude       site_name site_number sample    p450   SOD
## 1 48.67938 -122.6301 Aiston Preserve          77    239 5965780 0.000
## 2 48.67938 -122.6301 Aiston Preserve          77    240 1508156 4.877
## 3 48.67938 -122.6301 Aiston Preserve          77    241 4674882 8.871
## 4 48.67938 -122.6301 Aiston Preserve          77    242 2861653 0.010
## 5 47.50161 -122.3859    Arroyo Beach          13    281 3448794 7.084
## 6 47.50161 -122.3859    Arroyo Beach          13    282 6485447 0.635
##   weight_initial length width height weight_final weight_change
## 1        11.6884   53.9 22.73  18.59       3.2826          8.41
## 2         10.833  53.49 23.92  18.36       3.4809          7.35
## 3        14.7041  55.99 27.79  19.57       4.7251          9.98
## 4        14.6121  58.55 28.38  19.55       4.4461         10.17
## 5        15.4756  58.14 26.11  20.16       4.6221         10.85
## 6        17.9501  60.43 27.56   22.3       6.1066         11.84
##   condition_factor avg_thickness economic_index
## 1           0.1560         0.700         0.0018
## 2           0.1374         0.790          0.002
## 3           0.1782         0.825          0.002
## 4           0.1737         0.930         0.0021
## 5           0.1866         0.920         0.0022
## 6           0.1959         0.965         0.0022
```

```
head(metal)
```

```
##   Latitude Longitude LabSampleID             SiteName LabSampleID.1
## 1 47.50159 -122.3858    L79603-1         Arroyo Beach      L79603-1
## 2 47.68203 -122.5067    L79603-2       Brackenwood Ln      L79603-2
## 3 47.29469 -122.5305    L79603-3         Salmon Beach      L79603-3
## 4 48.04887 -122.7711    L79603-4 Chimacum Creek delta      L79603-4
## 5 47.66141 -122.4989    L79603-5          Skiff Point      L79603-5
## 6 48.02655 -122.7509    L79603-6    S of Skunk Island      L79603-6
##        Analyte Qualifier Units PctSolids   DryValue
## 1 mercuryTotal         D mg/Kg      17.0 0.03600000
## 2 mercuryTotal         D mg/Kg      16.9 0.03745562
## 3 mercuryTotal         D mg/Kg      17.9 0.02379888
## 4 mercuryTotal         D mg/Kg      17.0 0.03264706
## 5 mercuryTotal         D mg/Kg      17.8 0.03932584
## 6 mercuryTotal         D mg/Kg      17.5 0.02868571
```

```
head(allana)
```

```
##               SiteName Latitude Longitude      Analyte Qualifier Units
## 1         Arroyo Beach 47.50159 -122.3858 mercuryTotal         D mg/Kg
## 2       Brackenwood Ln 47.68203 -122.5067 mercuryTotal         D mg/Kg
## 3         Salmon Beach 47.29469 -122.5305 mercuryTotal         D mg/Kg
## 4 Chimacum Creek delta 48.04887 -122.7711 mercuryTotal         D mg/Kg
## 5          Skiff Point 47.66141 -122.4989 mercuryTotal         D mg/Kg
## 6    S of Skunk Island 48.02655 -122.7509 mercuryTotal         D mg/Kg
##   PctSolids   DryValue
## 1      17.0 0.03600000
## 2      16.9 0.03745562
## 3      17.9 0.02379888
## 4      17.0 0.03264706
## 5      17.8 0.03932584
## 6      17.5 0.02868571
```

# Data frame manipulations

### Adjusting biomarker values for accurate stats

```
# Data contains 0's and must be adjusted in this order to preserve all usable data.

#sod
#replace any SOD values at or below 0 with half of the lower detection limit of .005 (.005*.5). Lower d
data$SOD[data$SOD <= 0] <- 0.0025

#p450
#remove any p450 values that are 0 - those are true 0's not non-detectable. I am replacing with na so I
data$p450[data$p450 <= 0] <- NA
```

### Data adjustment for analysis- SOD & p450

```
#Average the

library(dplyr)

#simplifying the dataframe for joining with next steps
averaged_data <- data %>%
  group_by(site_number, latitude, longitude, site_name) %>%
  summarise(
    avg_p450 = mean(p450, na.rm = TRUE),
    avg_SOD = mean(SOD, na.rm = TRUE)
  ) %>%
  ungroup()  # Remove grouping for the new dataframe

head(averaged_data)
```

```
## # A tibble: 6 x 6
##   site_number latitude longitude site_name                avg_p450 avg_SOD
##         <int>    <dbl>     <dbl> <chr>                       <dbl>   <dbl>
## 1           1     48.1     -123. Port Angeles Yacht Club  5751355    7.39
## 2           2     48.0     -123. Jamestown                3263515   24.5
```

```
## 3           3     48.2     -123. Penn Cove Reference      2427656.  23.9
## 4           7     48.3     -123. North Camano            12290521    0.752
## 5           8     48.0     -123. Chimacum Creek delta     2641574.   2.19
## 6           9     48.0     -123. S of Skunk Island        3556923.  11.3
```

```
library(reshape2)
#merge data frames and reshape for input.
colnames(allana)[colnames(allana) == "SiteName"] <- "site_name"
merged_df <- merge(averaged_data, allana, by = c("site_name"), all.x = TRUE)

#reshape to get the analytes into their own columns with the DryValue as their values
reshaped_df2 <- dcast(merged_df, site_name + site_number +latitude + longitude + avg_p450 + avg_SOD ~ An

head(reshaped_df2)
```

```
##                       site_name site_number latitude longitude avg_p450
## 1                Aiston Preserve          77 48.67938 -122.6301  3752618
## 2                   Arroyo Beach          13 47.50161 -122.3859  4480860
## 3                  Blair Waterway          41 47.27568 -122.4173  4879642
## 4               Blair Waterway #2          42 47.26324 -122.3857  3714918
## 5                  Brackenwood Ln          23 47.68234 -122.5064  1857012
## 6 Broad Spit (Fisherman's Point)          30 47.78184 -122.8347  2311731
##     avg_SOD  arsenic  cadmium   copper      lead mercuryTotal Sum40CBs
## 1  3.440125 7.245509 1.652695 4.940120 0.1772455   0.03305389 23.13322
## 2  8.832583 9.647059 1.952941 5.623529 0.2423529   0.03600000 34.81931
## 3  6.517750 8.114286 1.622857 5.828571 0.2554286   0.03205714 37.40489
## 4 10.796000 8.373494 1.704819 8.132530 0.1849398  -0.02360000 42.15557
## 5  9.835125 8.698225 1.857988 6.213018 0.2201183   0.03745562 29.49750
## 6  7.116250       NA       NA       NA        NA           NA       NA
##     SumBDEs  SumCHLDs   SumDDTs    SumHCHs    SumPAHs SumPAHs16
## 1 -1.095784 -1.095784  1.826307 -0.9131536   97.40305  31.65599
## 2  5.643129  3.121731  2.401332 -1.1406325  408.22636 168.09321
## 3 15.952084  4.345568  9.901294  0.8801150  715.09344 247.53234
## 4  9.164255 -1.710661 17.717559  0.9775205 1038.61553 299.36565
## 5  2.005830 -1.592865  2.182815 -1.2978901  389.36704 176.98502
## 6        NA        NA        NA         NA         NA        NA
##   SumPAHs42_DMNcorrected SumPAHsHMW SumPAHsLMW SumPCBs2x17      Zinc NA
## 1               97.40305   19.48061   79.13998    28.00338  77.24551 NA
## 2              408.22636  186.10320  216.11984    49.22730 104.11765 NA
## 3              715.09344  280.53666  456.55966    50.60661  98.85714 NA
## 4             1038.61553  403.22721  610.95031    54.98553  85.54217 NA
## 5              389.36704  176.98502  212.38202    44.24625  90.53254 NA
## 6                     NA         NA         NA          NA        NA NA
```

### Data frame for p450 Pearson & PCA- summed PAH analytes

```
#create a table without the avg_SOD and NA column for p450 work
cols_to_keep2 <- colnames(reshaped_df2)[!colnames(reshaped_df2) %in% c("avg_SOD", "arsenic", "cadmium",

p450PAH <- reshaped_df2[, cols_to_keep2]


head(p450PAH)
```

```
##                       site_name site_number latitude longitude avg_p450
```

```
## 1                Aiston Preserve          77 48.67938 -122.6301  3752618
## 2                   Arroyo Beach          13 47.50161 -122.3859  4480860
## 3                  Blair Waterway          41 47.27568 -122.4173  4879642
## 4              Blair Waterway #2          42 47.26324 -122.3857  3714918
## 5                 Brackenwood Ln          23 47.68234 -122.5064  1857012
## 6 Broad Spit (Fisherman's Point)          30 47.78184 -122.8347  2311731
##      SumPAHs SumPAHs16 SumPAHs42_DMNcorrected SumPAHsHMW SumPAHsLMW
## 1   97.40305  31.65599               97.40305   19.48061   79.13998
## 2  408.22636 168.09321              408.22636  186.10320  216.11984
## 3  715.09344 247.53234              715.09344  280.53666  456.55966
## 4 1038.61553 299.36565             1038.61553  403.22721  610.95031
## 5  389.36704 176.98502              389.36704  176.98502  212.38202
## 6         NA        NA                     NA         NA         NA
```

## p450 Pearson- summed analytes

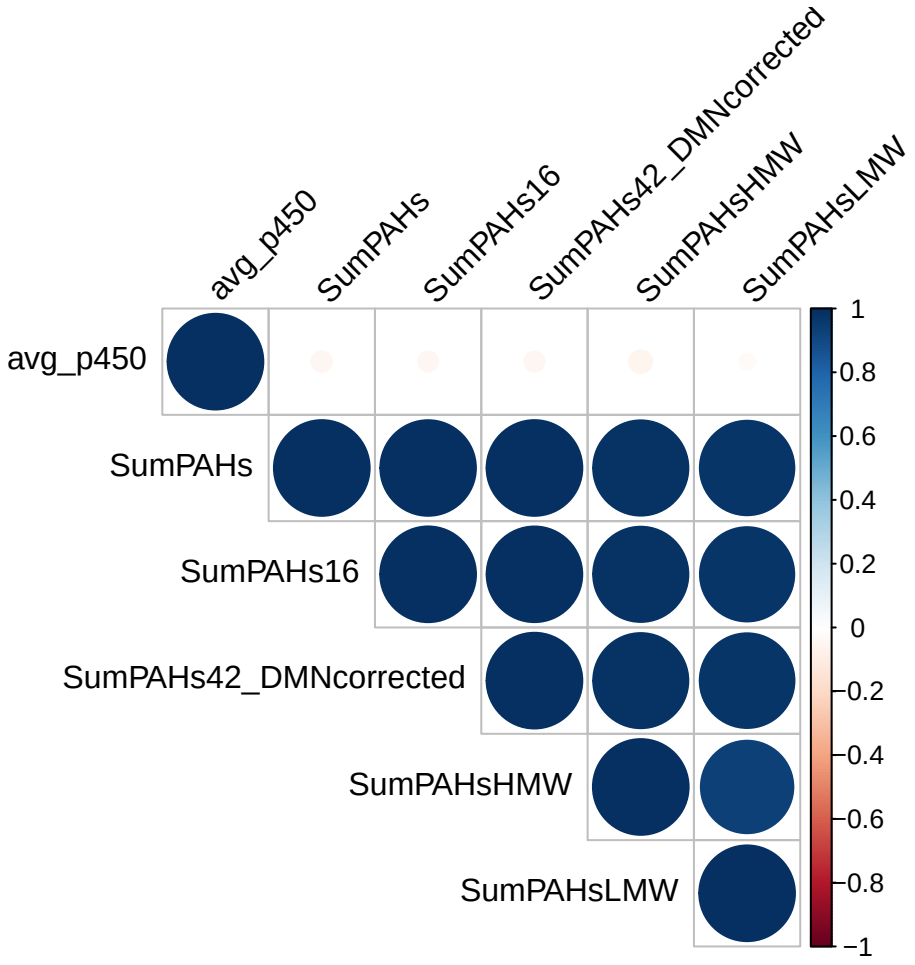
```
#get the column names from sod_all so I don't have to individually type each one

all_columns <- names(p450PAH)

# Remove the columns you don't want to include in the model
excluded_columns <- c('latitude', 'longitude', 'site_name', 'site_number')
independent_columns <- all_columns[!all_columns %in% excluded_columns]

# Enclose each column name in backticks to handle special characters
independent_columns <- sapply(independent_columns, function(x) paste0("`", x, "`"))

# Create a string representing the formula
formula_str <- paste("avg_p450 ~", paste(independent_columns, collapse = " + "))

# Convert the string to a formula object
formula <- as.formula(formula_str)
```

```
library(corrplot)
# Extract variable names from the formula
variables <- all.vars(formula)

# Subset the dataframe 'sod_all' using the extracted variables
subset_data <- p450PAH[, variables]

# Compute Pearson correlation for each pair of variables
correlation_results <- cor(subset_data, method = "pearson", use = "complete.obs")

# View the correlation matrix
print(correlation_results)
```

```
##                            avg_p450     SumPAHs   SumPAHs16
## avg_p450                 1.00000000 -0.04582778 -0.04345867
## SumPAHs                 -0.04582778  1.00000000  0.99718015
## SumPAHs16               -0.04345867  0.99718015  1.00000000
## SumPAHs42_DMNcorrected -0.04391634  0.99975446  0.99755481
## SumPAHsHMW              -0.05759445  0.98952982  0.98921868
## SumPAHsLMW              -0.02845483  0.97448144  0.97002607
```

```
##                          SumPAHs42_DMNcorrected  SumPAHsHMW  SumPAHsLMW
## avg_p450                             -0.04391634 -0.05759445 -0.02845483
## SumPAHs                               0.99975446  0.98952982  0.97448144
## SumPAHs16                             0.99755481  0.98921868  0.97002607
## SumPAHs42_DMNcorrected                1.00000000  0.98673513  0.97849736
## SumPAHsHMW                            0.98673513  1.00000000  0.93319226
## SumPAHsLMW                            0.97849736  0.93319226  1.00000000
```

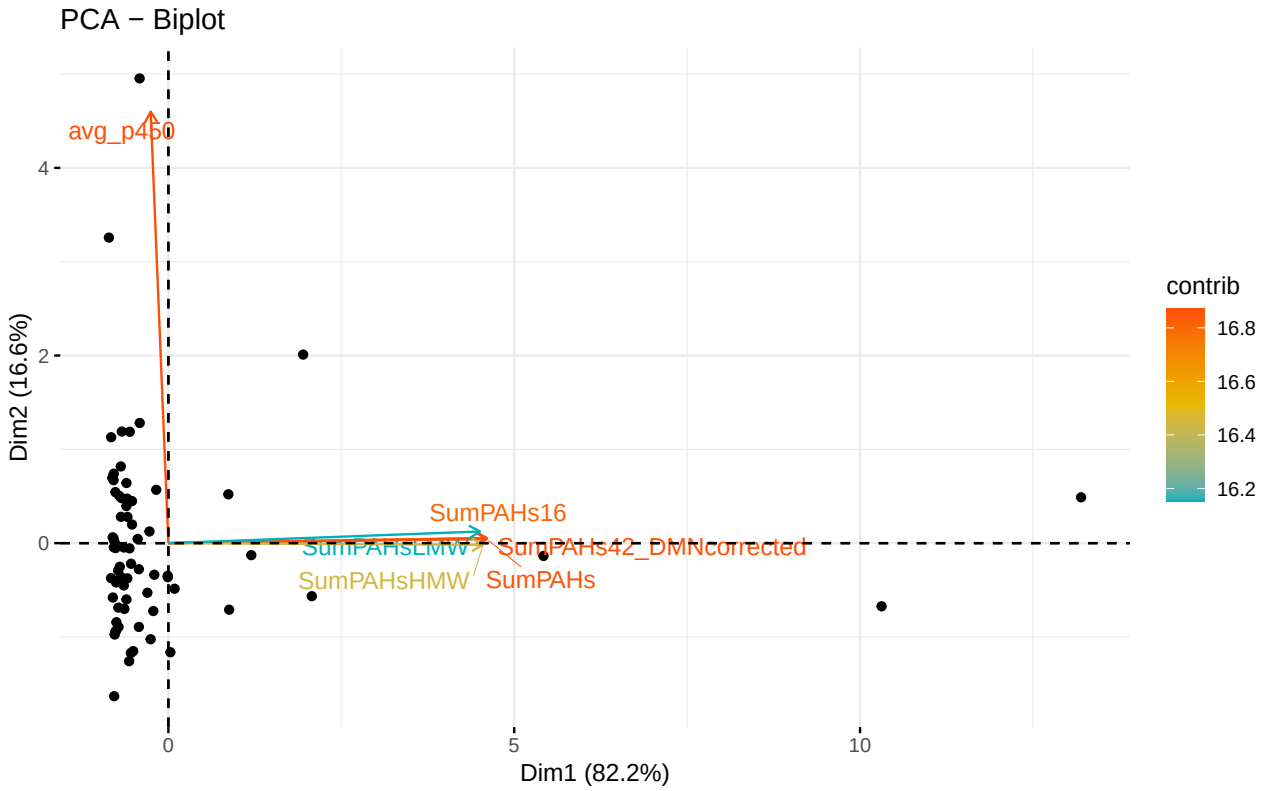
```
corrplot(correlation_results, method = "circle", type = "upper", tl.col = "black", tl.srt = 45)
```




## p450 PCA - summed analytes

```
# PCA Plot with biomarkers
#install.packages("FactoMineR")
#install.packages("factoextra")
library('FactoMineR')
library("factoextra")


# Remove NAs from the dataset
df_clean <- na.omit(p450PAH)

# Selecting the relevant variables for PCA
```

```
pca_data <- df_clean[, c("avg_p450", "SumPAHs", "SumPAHs16", "SumPAHs42_DMNcorrected", "SumPAHsHMW", "Su

# Performing PCA
pca_res <- PCA(pca_data, scale.unit = TRUE, graph = FALSE)

# Plotting the PCA
pcaplot<- fviz_pca_biplot(pca_res, label = "var", col.var = "contrib",
                gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                repel = TRUE)  # Avoid text overlapping (slow if many points)



print(pcaplot)
```




```
#ggsave(plot=pcaplot, filename="/Users/cmantegna/Documents/WDFWmussels/output/pca.png", width=15, heigh
```

**Data frame for p450 Pearson & PCA- individual PAH analytes**

```
library(reshape2)
#merge data frames and reshape for input.
colnames(indv)[colnames(indv) == "SiteName"] <- "site_name"
merged_df2 <- merge(averaged_data, indv, by = c("site_name"), all.x = TRUE)

#reshape to get the analytes into their own columns with the DryValue as their values
reshaped_df3 <- dcast(merged_df2, site_name + site_number +latitude + longitude + avg_p450 + avg_SOD ~ A

head(reshaped_df3)
```

```
##                       site_name site_number latitude longitude avg_p450
## 1                Aiston Preserve          77 48.67938 -122.6301  3752618
## 2                   Arroyo Beach          13 47.50161 -122.3859  4480860
## 3                  Blair Waterway          41 47.27568 -122.4173  4879642
## 4              Blair Waterway #2          42 47.26324 -122.3857  3714918
## 5                 Brackenwood Ln          23 47.68234 -122.5064  1857012
## 6 Broad Spit (Fisherman's Point)          30 47.78184 -122.8347  2311731
##     avg_SOD acenaphthene acenaphthylene anthracene benz[a]anthracene
## 1  3.440125    -3.165599      -3.165599  -3.165599         -3.104722
## 2  8.832583    -3.902164      -3.902164  -3.902164         13.207324
## 3  6.517750     5.500719      -2.915381   6.050791         17.052228
## 4 10.796000     7.942354      -4.948698   5.315268         17.717559
## 5  9.835125    -3.598695      -3.598695  -3.598695         12.978901
## 6  7.116250           NA             NA         NA                NA
##   benzo[a]pyrene benzo[b]fluoranthene benzo[e]pyrene benzo[ghi]perylene
## 1      -3.104722            -3.104722      -3.104722          -3.104722
## 2      -3.001664            11.406325       8.404660          -2.941631
## 3       2.915381             9.901294      12.101581           2.915381
## 4      -5.009793            17.106609      26.270863           6.720453
## 5      -2.713770             9.439201       6.489451          -2.713770
## 6             NA                   NA             NA                 NA
##   benzo[k]fluoranthene C1-benzanthracenes/chrysenes C1-dibenzothiophenes
## 1            -3.104722                    -2.252446            -3.104722
## 2            13.207324                     9.605326            -3.842130
## 3            10.451366                    12.651653             3.905510
## 4            16.495658                    28.103714            -4.887603
## 5            11.799001                     7.079401            -3.539700
## 6                   NA                           NA                   NA
##   C1-fluoranthenes/pyrenes C1-fluorenes C1-naphthalenes
## 1                -3.104722    -3.104722       -3.104722
## 2                15.008322     3.902164       -6.003329
## 3                22.002875     5.115668       -3.630474
## 4                28.714665     6.109503       -4.887603
## 5                15.338702    -3.539700       -5.309551
## 6                       NA           NA              NA
##   C1-phenanthrenes/anthracenes C2-benzanthracenes/chrysenes
## 1                     8.522767                    -2.252446
## 2                    25.814314                     6.603662
## 3                    45.105894                    11.001437
## 4                    45.210323                    29.325615
## 5                    27.137703                     3.775680
## 6                           NA                           NA
##   C2-dibenzothiophenes C2-fluoranthenes/pyrenes C2-fluorenes C2-naphthalenes
## 1            -3.104722                -3.104722    -3.104722        7.305229
## 2            -3.842130                 9.004993     4.742630       10.205659
## 3             9.351222                14.851941     9.901294       17.437278
## 4            13.440907                23.827062    17.717559       18.328509
## 5            -3.539700                 8.259301     4.719600       10.029151
## 6                   NA                       NA           NA              NA
##   C2-phenanthrenes/anthracenes C3-benzanthracenes/chrysenes
## 1                     9.740305                    -2.252446
## 2                    40.222304                     3.181764
## 3                    66.008625                     5.500719
## 4                   103.861553                    15.884708
```

```
## 5                    38.346754                    -2.182815
## 6                           NA                           NA
##   C3-dibenzothiophenes C3-fluoranthenes/pyrenes C3-fluorenes C3-naphthalenes
## 1            -3.104722                -3.104722   -3.104722        9.131536
## 2            -3.842130                 6.003329    5.883262       13.807656
## 3            10.451366                13.751797   17.602300       34.104456
## 4            18.939460                26.881814   40.933671       34.824168
## 5            -3.539700                 4.424625   10.619101       14.158801
## 6                   NA                       NA          NA              NA
##   C3-phenanthrenes/anthracenes C4-benzanthracenes/chrysenes
## 1                     5.600675                   -2.252446
## 2                    34.819308                   -2.401332
## 3                    71.509344                   -1.815237
## 4                   128.299566                   -3.543512
## 5                    30.087453                   -2.182815
## 6                           NA                           NA
##   C4-dibenzothiophenes C4-fluoranthenes/pyrenes C4-naphthalenes
## 1            -3.104722                -3.104722        3.530861
## 2            -3.842130                -3.842130        9.004993
## 3            11.551509                 5.005654       26.403450
## 4            15.884708                 8.553304       32.991317
## 5            -3.539700                -3.539700        9.439201
## 6                   NA                       NA              NA
##   C4-phenanthrenes/anthracenes chrysene dibenz[a,h]anthracene dibenzothiophene
## 1                     6.026814  4.13963             -3.104722        -3.104722
## 2                    30.616977 23.41298             -3.001664        -3.842130
## 3                    55.007187 24.75323             -2.200287        -2.915381
## 4                    85.533044 39.71177             -4.948698        -4.887603
## 5                    23.008052 21.82815             -2.713770        -3.539700
## 6                           NA       NA                    NA               NA
##   indeno[1,2,3-cd]pyrene phenanthrene NA
## 1              -3.104722     12.17538 NA
## 2              -3.001664     30.61698 NA
## 3              -2.200287     37.95496 NA
## 4              -5.009793     39.10082 NA
## 5              -2.713770     37.16685 NA
## 6                     NA           NA NA
```

```
#create a table without the avg_SOD and NA column for p450 work
cols_to_keep3 <- colnames(reshaped_df3)[!colnames(reshaped_df3) %in% c("avg_SOD", "NA")]

p450all <- reshaped_df3[, cols_to_keep3]


head(p450all)
```

```
##                       site_name site_number latitude longitude avg_p450
## 1                 Aiston Preserve          77 48.67938 -122.6301  3752618
## 2                    Arroyo Beach          13 47.50161 -122.3859  4480860
## 3                  Blair Waterway          41 47.27568 -122.4173  4879642
## 4               Blair Waterway #2          42 47.26324 -122.3857  3714918
## 5                  Brackenwood Ln          23 47.68234 -122.5064  1857012
## 6 Broad Spit (Fisherman's Point)          30 47.78184 -122.8347  2311731
##   acenaphthene acenaphthylene anthracene benz[a]anthracene benzo[a]pyrene
## 1    -3.165599      -3.165599 -3.165599         -3.104722      -3.104722
```

```
## 2    -3.902164      -3.902164  -3.902164         13.207324      -3.001664
## 3     5.500719      -2.915381   6.050791         17.052228       2.915381
## 4     7.942354      -4.948698   5.315268         17.717559      -5.009793
## 5    -3.598695      -3.598695  -3.598695         12.978901      -2.713770
## 6           NA             NA         NA                NA             NA
##   benzo[b]fluoranthene benzo[e]pyrene benzo[ghi]perylene benzo[k]fluoranthene
## 1            -3.104722      -3.104722          -3.104722            -3.104722
## 2            11.406325       8.404660          -2.941631            13.207324
## 3             9.901294      12.101581           2.915381            10.451366
## 4            17.106609      26.270863           6.720453            16.495658
## 5             9.439201       6.489451          -2.713770            11.799001
## 6                   NA             NA                 NA                   NA
##   C1-benzanthracenes/chrysenes C1-dibenzothiophenes C1-fluoranthenes/pyrenes
## 1                    -2.252446           -3.104722                -3.104722
## 2                     9.605326           -3.842130                15.008322
## 3                    12.651653            3.905510                22.002875
## 4                    28.103714           -4.887603                28.714665
## 5                     7.079401           -3.539700                15.338702
## 6                           NA                 NA                       NA
##   C1-fluorenes C1-naphthalenes C1-phenanthrenes/anthracenes
## 1    -3.104722       -3.104722                    8.522767
## 2     3.902164       -6.003329                   25.814314
## 3     5.115668       -3.630474                   45.105894
## 4     6.109503       -4.887603                   45.210323
## 5    -3.539700       -5.309551                   27.137703
## 6           NA             NA                          NA
##   C2-benzanthracenes/chrysenes C2-dibenzothiophenes C2-fluoranthenes/pyrenes
## 1                    -2.252446           -3.104722                -3.104722
## 2                     6.603662           -3.842130                 9.004993
## 3                    11.001437            9.351222                14.851941
## 4                    29.325615           13.440907                23.827062
## 5                     3.775680           -3.539700                 8.259301
## 6                           NA                 NA                       NA
##   C2-fluorenes C2-naphthalenes C2-phenanthrenes/anthracenes
## 1    -3.104722        7.305229                    9.740305
## 2     4.742630       10.205659                   40.222304
## 3     9.901294       17.437278                   66.008625
## 4    17.717559       18.328509                  103.861553
## 5     4.719600       10.029151                   38.346754
## 6           NA             NA                          NA
##   C3-benzanthracenes/chrysenes C3-dibenzothiophenes C3-fluoranthenes/pyrenes
## 1                    -2.252446           -3.104722                -3.104722
## 2                     3.181764           -3.842130                 6.003329
## 3                     5.500719           10.451366                13.751797
## 4                    15.884708           18.939460                26.881814
## 5                    -2.182815           -3.539700                 4.424625
## 6                           NA                 NA                       NA
##   C3-fluorenes C3-naphthalenes C3-phenanthrenes/anthracenes
## 1    -3.104722        9.131536                    5.600675
## 2     5.883262       13.807656                   34.819308
## 3    17.602300       34.104456                   71.509344
## 4    40.933671       34.824168                  128.299566
## 5    10.619101       14.158801                   30.087453
## 6           NA             NA                          NA
```

```
##   C4-benzanthracenes/chrysenes C4-dibenzothiophenes C4-fluoranthenes/pyrenes
## 1                   -2.252446            -3.104722                -3.104722
## 2                   -2.401332            -3.842130                -3.842130
## 3                   -1.815237            11.551509                 5.005654
## 4                   -3.543512            15.884708                 8.553304
## 5                   -2.182815            -3.539700                -3.539700
## 6                          NA                   NA                       NA
##   C4-naphthalenes C4-phenanthrenes/anthracenes chrysene dibenz[a,h]anthracene
## 1        3.530861                     6.026814  4.13963             -3.104722
## 2        9.004993                    30.616977 23.41298             -3.001664
## 3       26.403450                    55.007187 24.75323             -2.200287
## 4       32.991317                    85.533044 39.71177             -4.948698
## 5        9.439201                    23.008052 21.82815             -2.713770
## 6              NA                           NA       NA                    NA
##   dibenzothiophene indeno[1,2,3-cd]pyrene phenanthrene
## 1        -3.104722              -3.104722     12.17538
## 2        -3.842130              -3.001664     30.61698
## 3        -2.915381              -2.200287     37.95496
## 4        -4.887603              -5.009793     39.10082
## 5        -3.539700              -2.713770     37.16685
## 6               NA                     NA           NA
```

```
all_columns <- names(reshaped_df3)

# Remove the columns you don't want
excluded_columns <- c('site_name', 'site_number' ,'latitude', 'longitude', 'avg_SOD', 'NA')
independent_columns <- all_columns[!all_columns %in% excluded_columns]

# Enclose each column name in backticks to handle special characters
independent_columns <- sapply(independent_columns, function(x) paste0("`", x, "`"))

# Create a string representing the formula
formula_str <- paste("avg_p450 ~", paste(independent_columns, collapse = " + "))

# Convert the string to a formula object
formula1 <- as.formula(formula_str)
```

## Pearson - individual analytes

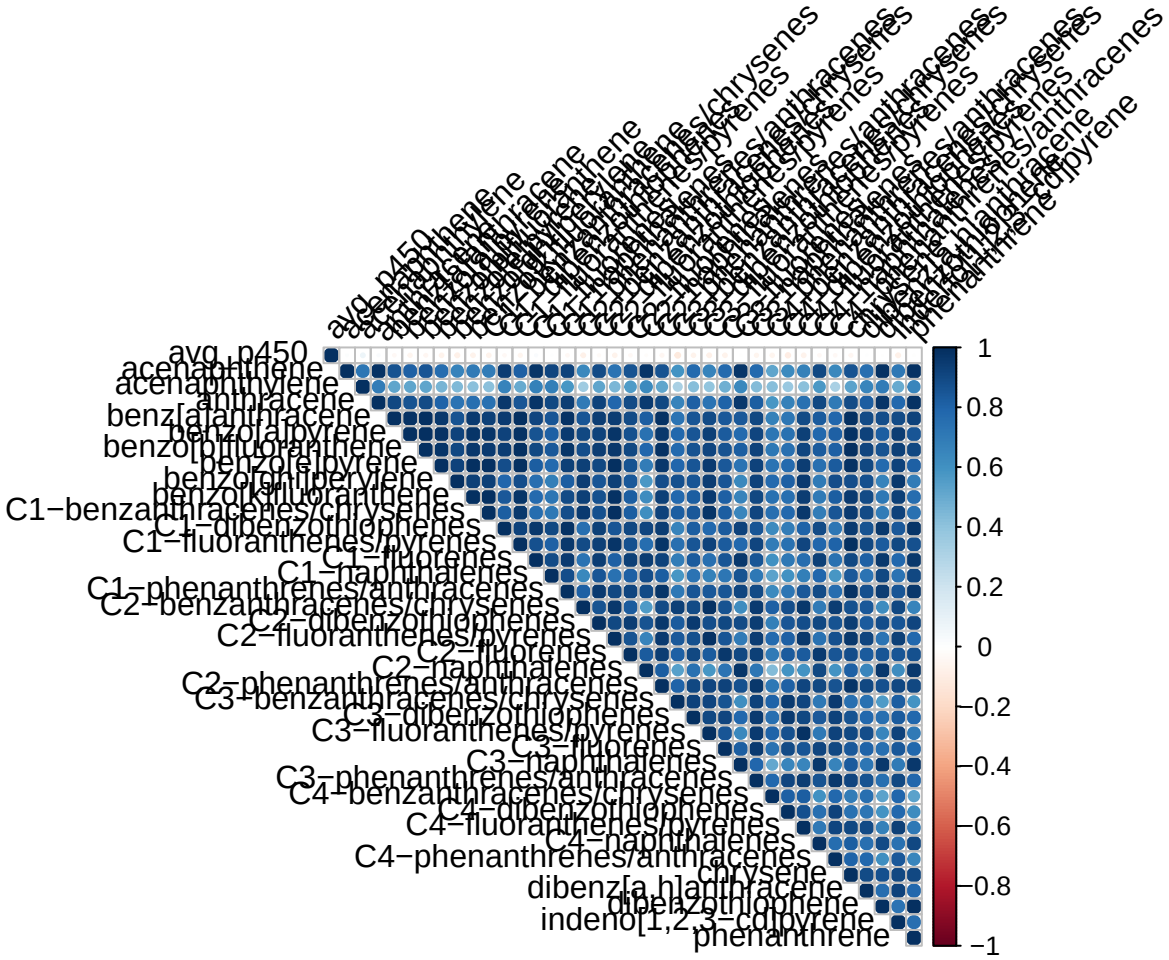
```
#picking up after excluding non-analysis columns
# Extract variable names from the formula
variables <- all.vars(formula1)

# Subset the dataframe 'sod_all' using the extracted variables
subset_data <- p450all[, variables]

# Compute Pearson correlation for each pair of variables
correlation_results2 <- cor(subset_data, method = "pearson", use = "complete.obs")

# View the correlation matrix
head(correlation_results2)
```

```
##                     avg_p450 acenaphthene acenaphthylene anthracene
## avg_p450          1.00000000   0.01503212     0.07417072 0.02232981
```

```
## acenaphthene       0.01503212   1.00000000     0.72330063 0.98271221
## acenaphthylene     0.07417072   0.72330063     1.00000000 0.69569843
## anthracene         0.02232981   0.98271221     0.69569843 1.00000000
## benz[a]anthracene -0.03262543   0.86864317     0.52067826 0.90454499
## benzo[a]pyrene    -0.05478740   0.81139408     0.52429936 0.86189209
##                   benz[a]anthracene benzo[a]pyrene benzo[b]fluoranthene
## avg_p450                -0.03262543     -0.0547874          -0.04106729
## acenaphthene             0.86864317      0.8113941           0.85325834
## acenaphthylene           0.52067826      0.5242994           0.52913995
## anthracene               0.90454499      0.8618921           0.89624457
## benz[a]anthracene        1.00000000      0.9821846           0.99463121
## benzo[a]pyrene           0.98218460      1.0000000           0.99065491
##                   benzo[e]pyrene benzo[ghi]perylene benzo[k]fluoranthene
## avg_p450             -0.06677933        -0.07927487           -0.0654972
## acenaphthene          0.76242077         0.66777082            0.7009633
## acenaphthylene        0.46722951         0.44346086            0.4102802
## anthracene            0.80938989         0.73631103            0.7597302
## benz[a]anthracene     0.96997390         0.86881496            0.9510819
## benzo[a]pyrene        0.98784570         0.92634271            0.9791003
##                   C1-benzanthracenes/chrysenes C1-dibenzothiophenes
## avg_p450                          -0.06378239          -0.03045705
## acenaphthene                       0.69161824           0.94716781
## acenaphthylene                     0.41523885           0.66705865
## anthracene                         0.74901770           0.96342956
## benz[a]anthracene                  0.93579090           0.95974877
## benzo[a]pyrene                     0.97003322           0.92338458
##                   C1-fluoranthenes/pyrenes C1-fluorenes C1-naphthalenes
## avg_p450                       -0.0515490   0.02252611     0.003964476
## acenaphthene                    0.8185509   0.97127004     0.913248300
## acenaphthylene                  0.5018771   0.67328052     0.682234611
## anthracene                      0.8572856   0.97784372     0.908026114
## benz[a]anthracene               0.9905630   0.91072555     0.853317337
## benzo[a]pyrene                  0.9893064   0.86439564     0.811582033
##                   C1-phenanthrenes/anthracenes C2-benzanthracenes/chrysenes
## avg_p450                          -0.03359885                  -0.07269662
## acenaphthene                       0.92676479                   0.64242550
## acenaphthylene                     0.58580308                   0.35605728
## anthracene                         0.94709453                   0.70948858
## benz[a]anthracene                  0.98696186                   0.85951598
## benzo[a]pyrene                     0.95364800                   0.90727258
##                   C2-dibenzothiophenes C2-fluoranthenes/pyrenes C2-fluorenes
## avg_p450                     -0.0247711              -0.06257841  -0.03300888
## acenaphthene                  0.8984597               0.73603496   0.86939981
## acenaphthylene                0.5287622               0.45240875   0.56660609
## anthracene                    0.9212452               0.79358909   0.89562156
## benz[a]anthracene             0.9472646               0.94988375   0.89206515
## benzo[a]pyrene                0.9222605               0.97791135   0.87133689
##                   C2-naphthalenes C2-phenanthrenes/anthracenes
## avg_p450               -0.01169963                  -0.04495845
## acenaphthene            0.97972654                   0.86601658
## acenaphthylene          0.63443539                   0.52200005
## anthracene              0.95487061                   0.90186664
## benz[a]anthracene       0.82263656                   0.98012427
## benzo[a]pyrene          0.74258393                   0.96529407
```

```
##                   C3-benzanthracenes/chrysenes C3-dibenzothiophenes
## avg_p450                           -0.1148864          -0.06107122
## acenaphthene                        0.5987241           0.77590021
## acenaphthylene                      0.3245099           0.42313209
## anthracene                          0.6610285           0.81369675
## benz[a]anthracene                   0.7378418           0.85534136
## benzo[a]pyrene                      0.7802671           0.85133278
##                   C3-fluoranthenes/pyrenes C3-fluorenes C3-naphthalenes
## avg_p450                        -0.07125938  -0.06414187    -0.006780428
## acenaphthene                     0.66864197   0.77184833     0.973619255
## acenaphthylene                   0.39583799   0.48636259     0.648971805
## anthracene                       0.73607146   0.80187148     0.961098887
## benz[a]anthracene                0.89193691   0.84077940     0.843681756
## benzo[a]pyrene                   0.93551465   0.83206172     0.776377539
##                   C3-phenanthrenes/anthracenes C4-benzanthracenes/chrysenes
## avg_p450                          -0.05232007                  -0.05864476
## acenaphthene                       0.77125857                   0.52473730
## acenaphthylene                     0.43416543                   0.42850582
## anthracene                         0.82252300                   0.58616179
## benz[a]anthracene                  0.90553217                   0.67848116
## benzo[a]pyrene                     0.91420786                   0.72923139
##                   C4-dibenzothiophenes C4-fluoranthenes/pyrenes C4-naphthalenes
## avg_p450                    -0.09567162              -0.06794471     -0.03454797
## acenaphthene                 0.63662649               0.67895833      0.90570016
## acenaphthylene               0.37760775               0.41955627      0.57802164
## anthracene                   0.67930129               0.74951749      0.90612006
## benz[a]anthracene            0.72972908               0.89224603      0.83466068
## benzo[a]pyrene               0.74894618               0.93742765      0.78419648
##                   C4-phenanthrenes/anthracenes    chrysene
## avg_p450                          -0.03398365 -0.04388444
## acenaphthene                       0.64642782  0.86376328
## acenaphthylene                     0.32649629  0.53597109
## anthracene                         0.70978226  0.89530697
## benz[a]anthracene                  0.79317325  0.99618528
## benzo[a]pyrene                     0.81616188  0.98597715
##                   dibenz[a,h]anthracene dibenzothiophene indeno[1,2,3-cd]pyrene
## avg_p450                     0.00375358    -0.0008213366            -0.07714225
## acenaphthene                 0.75039279     0.9832252506             0.71883784
## acenaphthylene               0.65429539     0.6939551298             0.50872215
## anthracene                   0.82225299     0.9778111371             0.78993132
## benz[a]anthracene            0.89074548     0.8893798619             0.89708154
## benzo[a]pyrene               0.92485860     0.8253457271             0.94230531
##                   phenanthrene
## avg_p450          -0.009030606
## acenaphthene       0.982931249
## acenaphthylene     0.654308354
## anthracene         0.981258242
## benz[a]anthracene  0.920052925
## benzo[a]pyrene     0.859014842
```

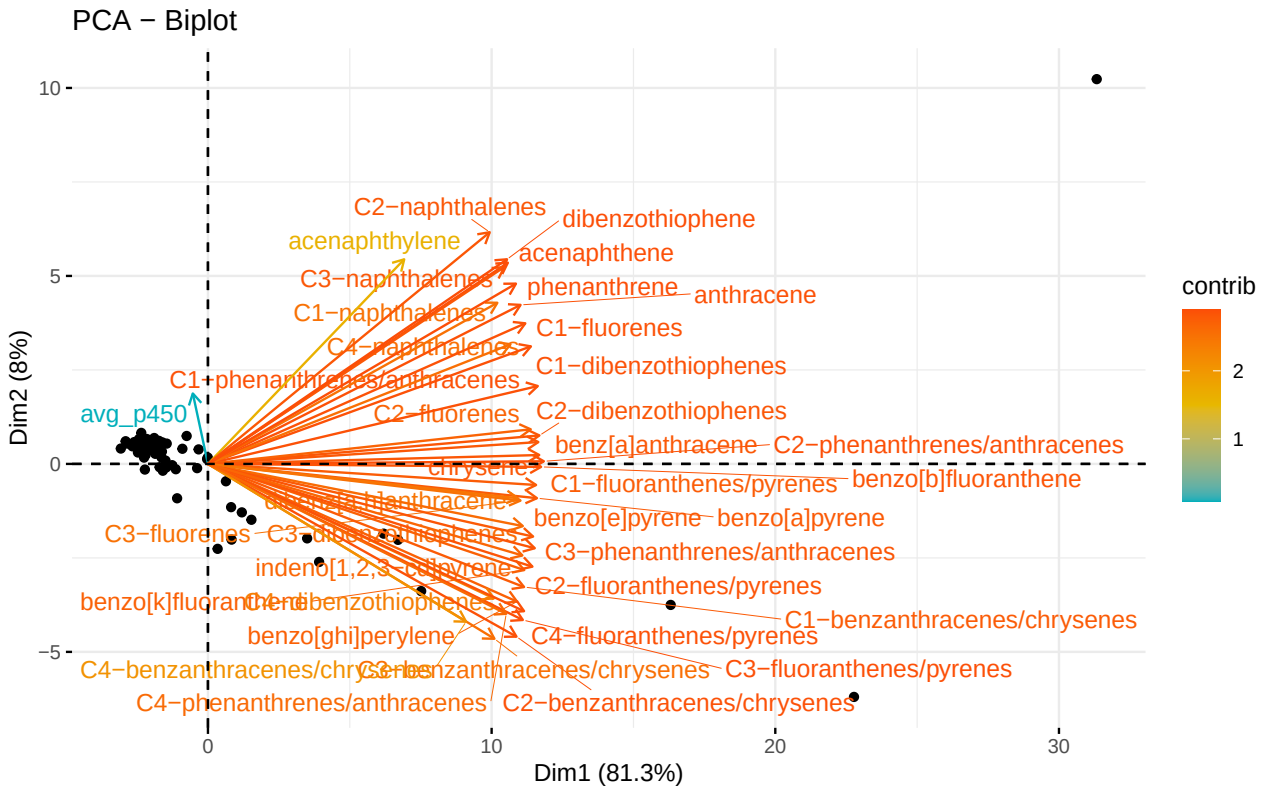
```
corrplot(correlation_results2, method = "circle", type = "upper", tl.col = "black", tl.srt = 45)
```




#PCA- individual analytes

```
# PCA Plot with biomarkers
#install.packages("FactoMineR")
#install.packages("factoextra")
library('FactoMineR')
library("factoextra")

# Remove NAs from the dataset
df_clean <- na.omit(p450all)

# Sample cols_to_keep3 vector (replace this with your actual vector of column names)
cols_to_keep3 <- colnames(df_clean)

# Remove the first four columns
selected_cols <- cols_to_keep3[-(1:4)]

# Select columns based on the selected_cols vector
pca_data <- df_clean[, selected_cols]

# Performing PCA
pca_res <- PCA(pca_data, scale.unit = TRUE, graph = FALSE)

# Plotting the PCA
```

```
pcaplot<- fviz_pca_biplot(pca_res, label = "var", col.var = "contrib",
                gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                repel = TRUE)  # Avoid text overlapping (slow if many points)



print(pcaplot)
```




```
#ggsave(plot=pcaplot, filename="/Users/cmantegna/Documents/WDFWmussels/output/pca.png", width=15, heigh
```