01- p450 Data Exploration

Directory and doc rules

```
knitr::opts_chunk$set(
  echo = TRUE,          # Display code chunks
  eval = TRUE,          # Evaluate code chunks
  warning = FALSE,      # Hide warnings
  message = FALSE,      # Hide messages
  fig.width = 20,        # Set plot width in inches
  fig.height = 9,       # Set plot height in inches
  fig.align = "center" # Align plots to the center
)
```

Load packages

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

Load data

```
getwd()
```

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

```
pdata<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/p450data.csv")
apdata<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/analytes_p450.csv")
```

Check data

```
#summary(apdata)
```

Plot data - boxplot, all p450 values by site name

```
#p450 is in activity/ mg protein
```

```
pplot<- ggplot(pdata, aes(x = site_name, y = p450)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```
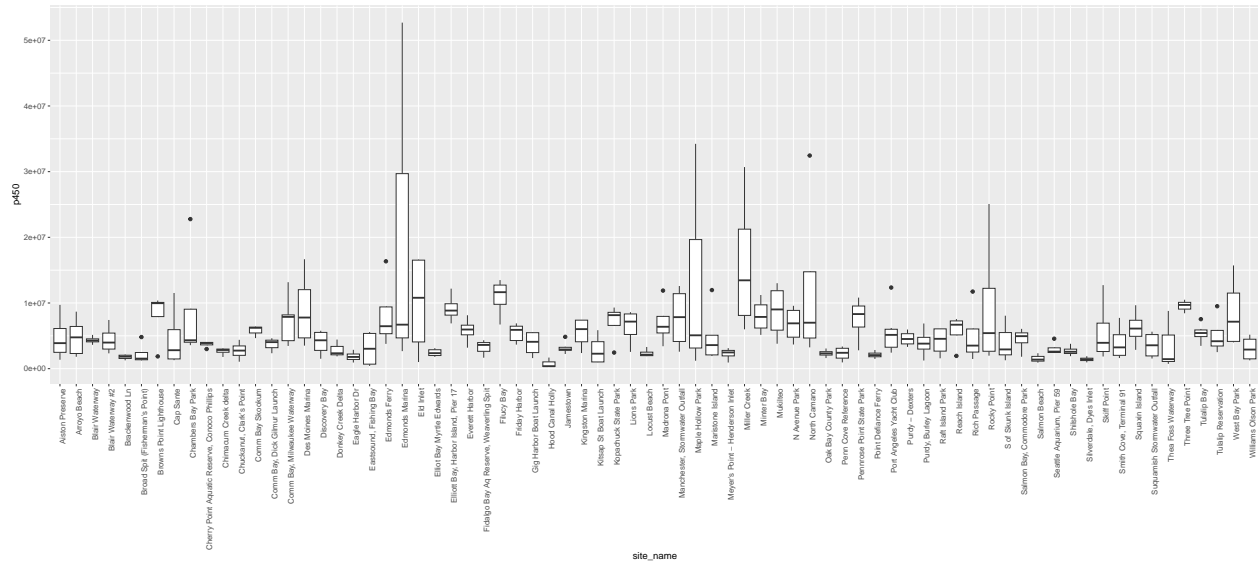
Plot data - ranked box plot, all p450 values by site name

```r
#order the sites by value

data_ordered <- pdata[order(pdata$p450),]

#create a factor with the ordered site names
data_ordered$site_name <- factor(data_ordered$site_name, levels = unique(data_ordered$site_name))

#plot with ordered site names
ranked<- ggplot(data_ordered, aes(x = site_name, y = p450)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) # Rotate x labels if needed

print(ranked)
```
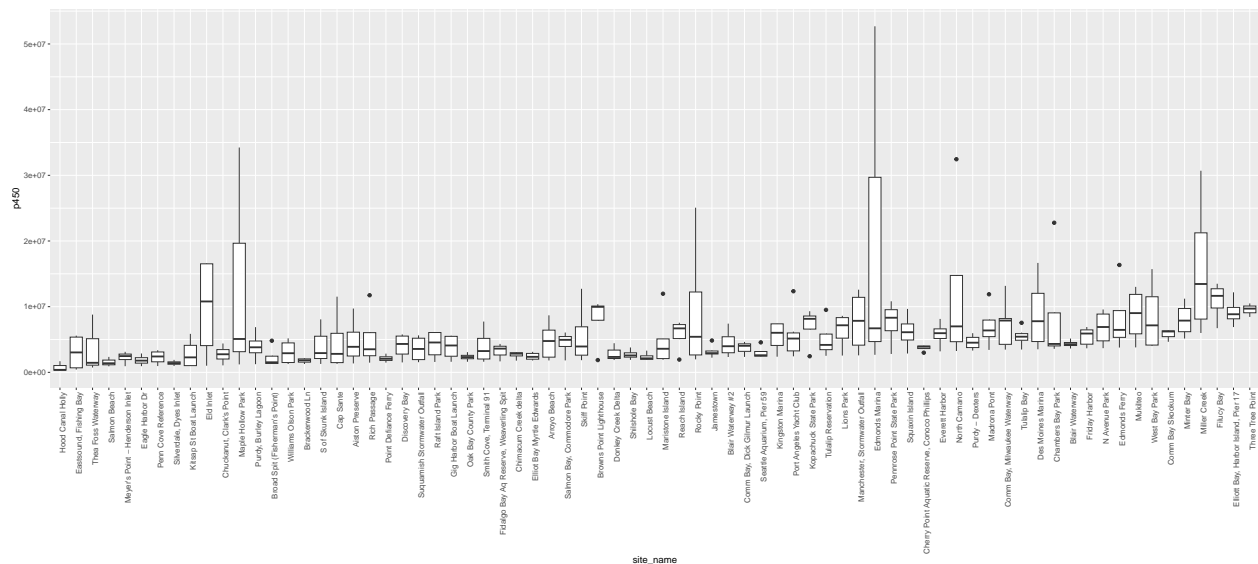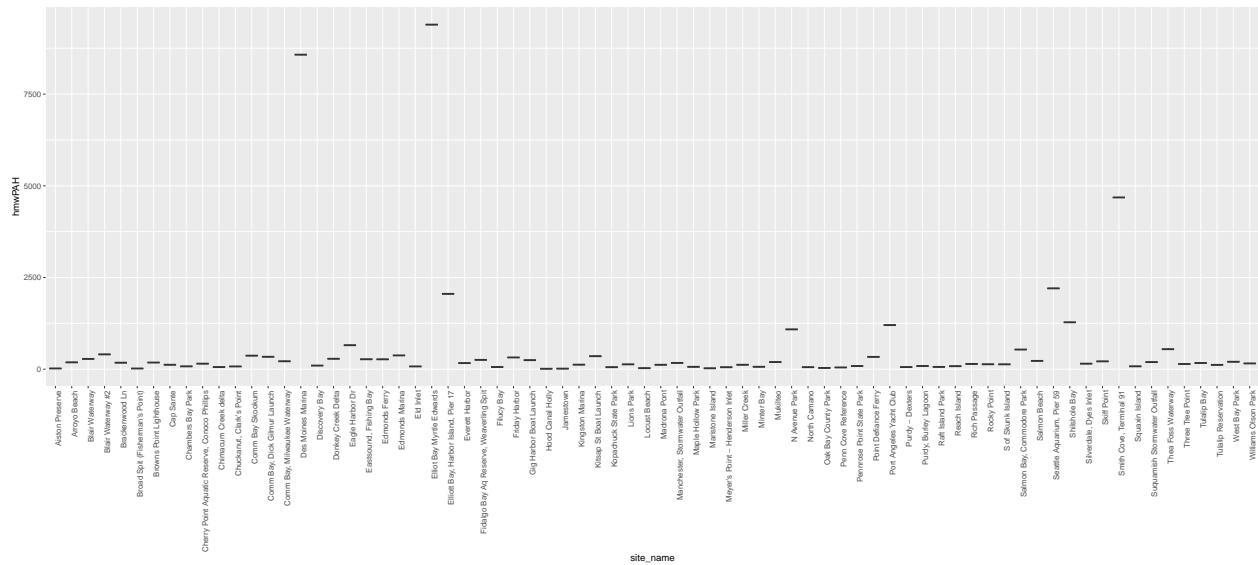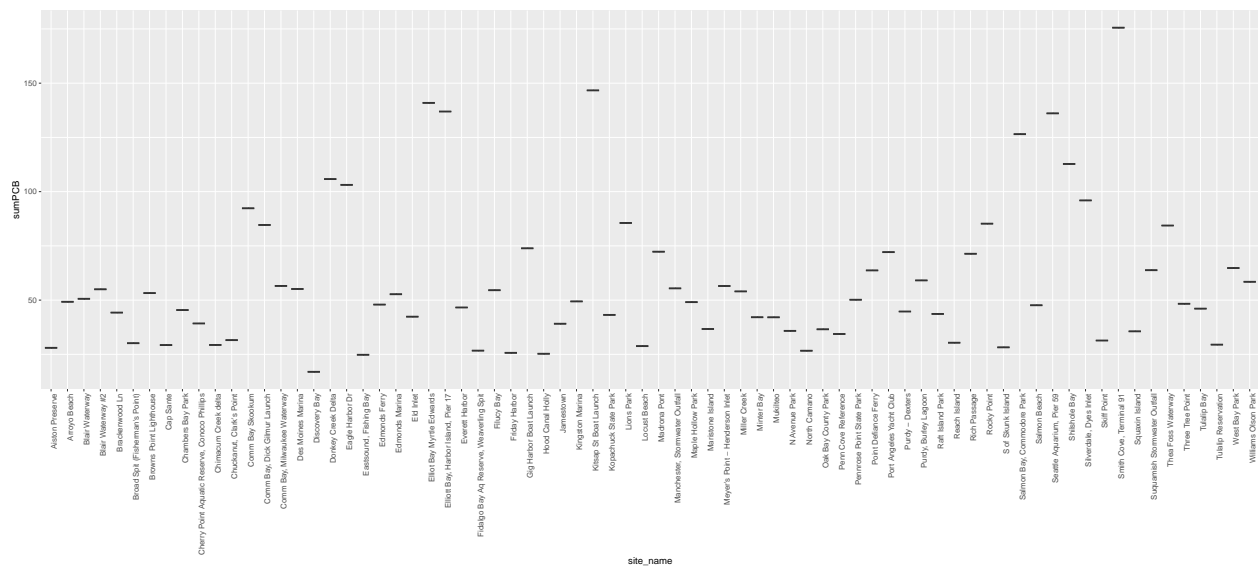
Plot data - boxplot, all SumPAH values by site name

```
#SumPAHs in ng/g tissue

pplot<- ggplot(apdata, aes(x = site_name, y = sumPAH)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```


Plot data - boxplot, all hmwPAH values by site name

```
#SumPAHs in ng/g tissue

pplot<- ggplot(apdata, aes(x = site_name, y = hmwPAH)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```

Plot data - boxplot, all sumPCB values by site name

```r
#SumPAHs in ng/g tissue

pplot<- ggplot(apdata, aes(x = site_name, y = sumPCB)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```
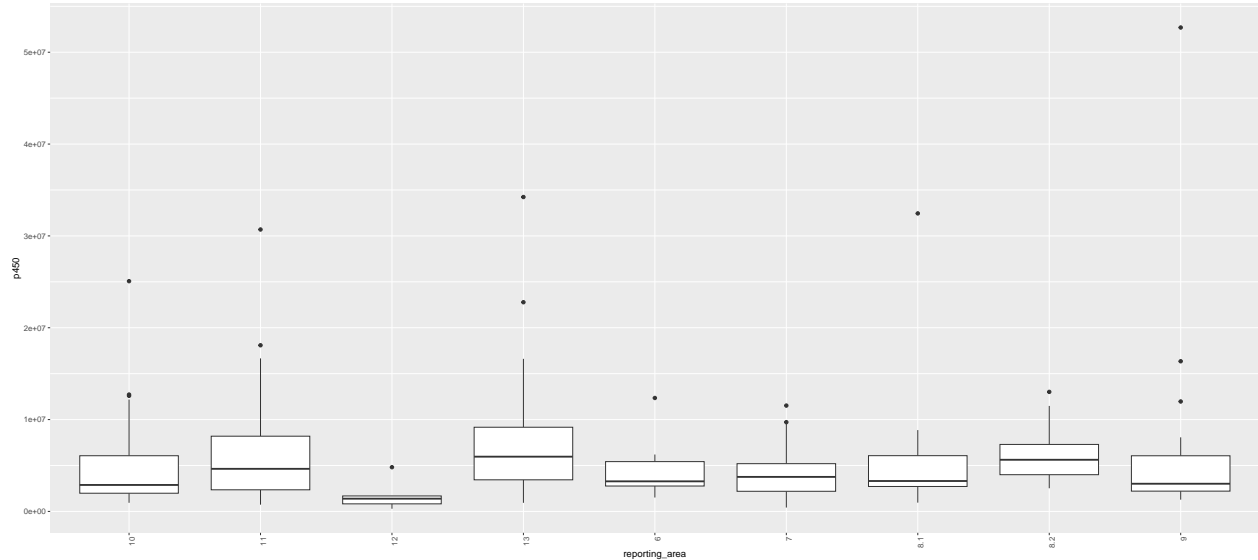

Plot data - boxplot, all p450 values by reporting area

```r
pdata$reporting_area <- as.character(pdata$reporting_area)

#p450
```

```
pplot<- ggplot(pdata, aes(x = reporting_area, y = p450)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```
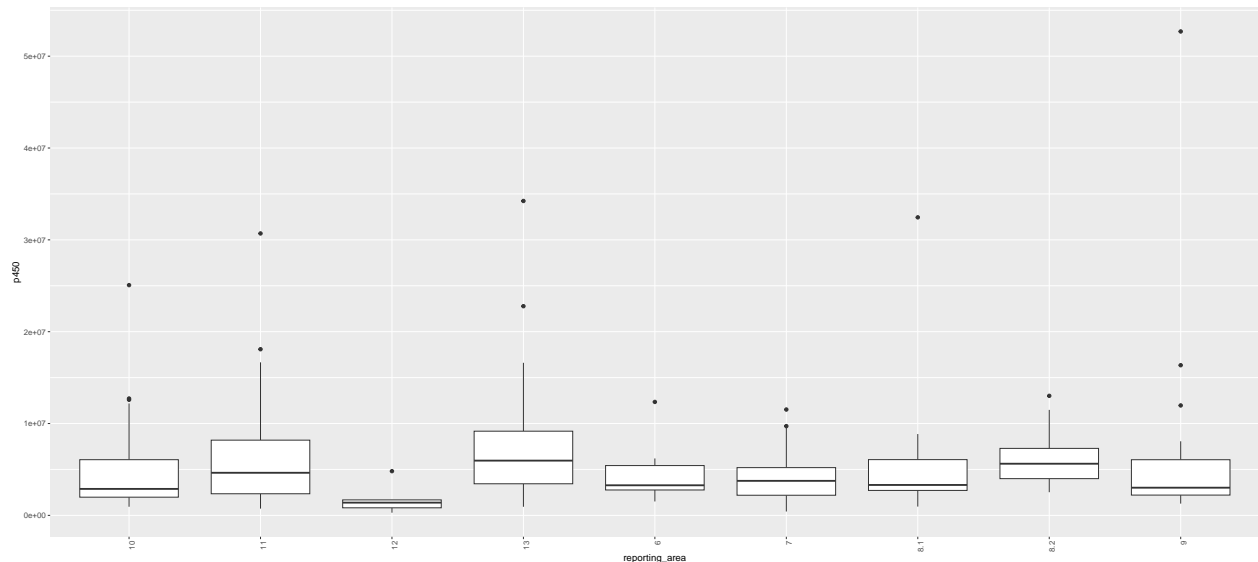

Plot data - boxplot, all SumPAH values by reporting area

```
apdata$reporting_area <- as.character(apdata$reporting_area)

#p450
pplot<- ggplot(apdata, aes(x = reporting_area, y = p450)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```

Plot data - boxplot, all hmwPAH values by reporting area

```
apdata$reporting_area <- as.character(apdata$reporting_area)

#p450
pplot<- ggplot(apdata, aes(x = reporting_area, y = hmwPAH)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```
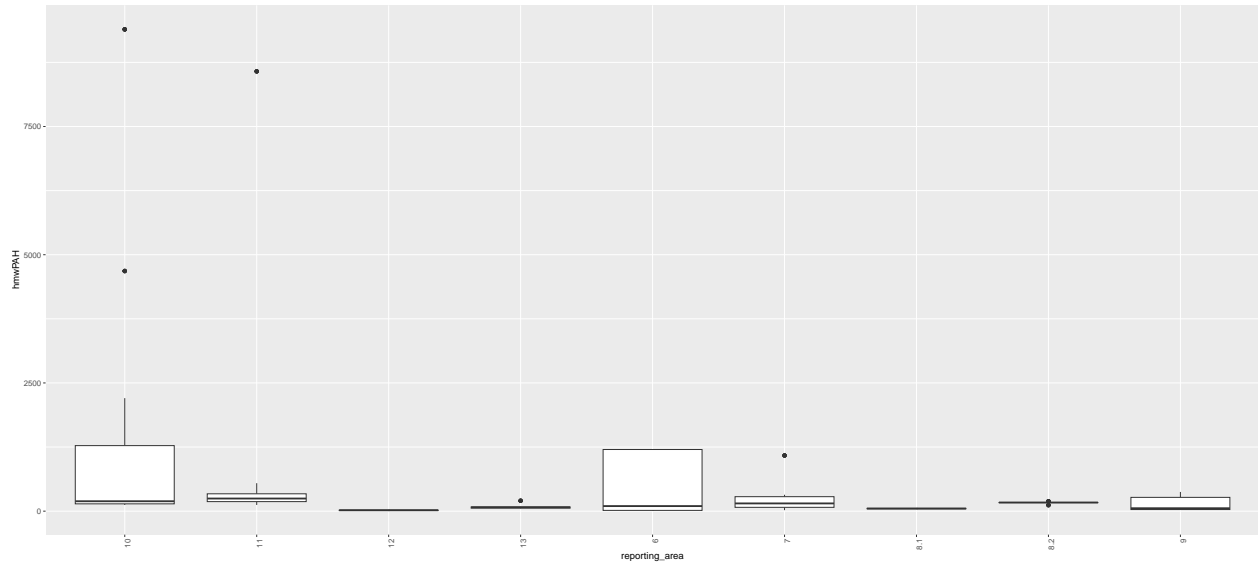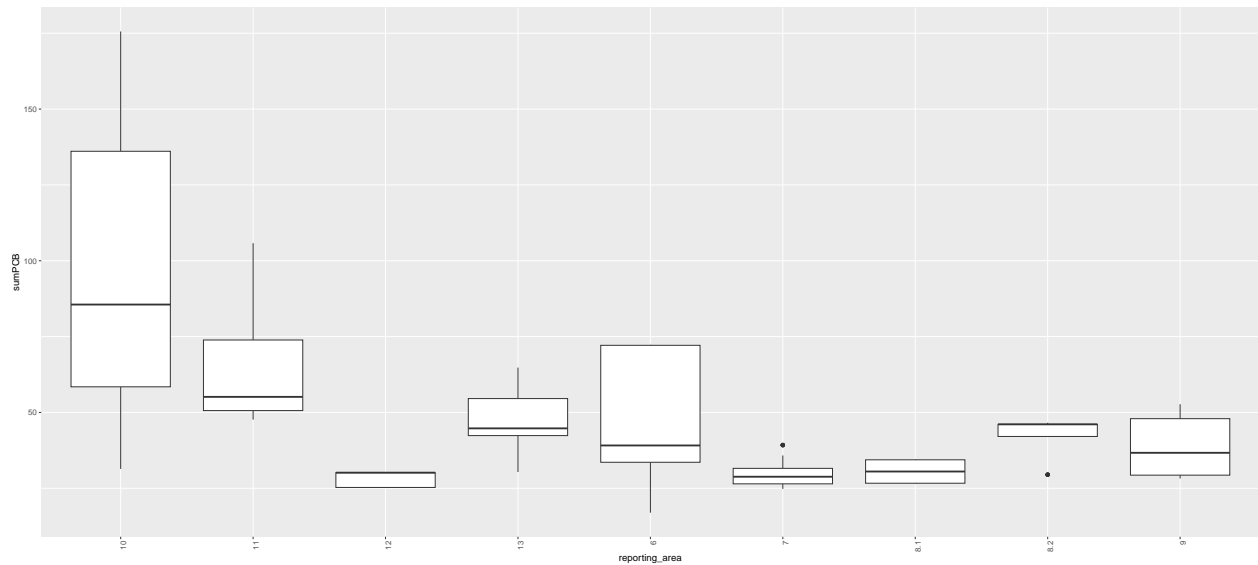

Plot data - boxplot, all sumPCB values by reporting area

```
apdata$reporting_area <- as.character(apdata$reporting_area)

#p450
pplot<- ggplot(apdata, aes(x = reporting_area, y = sumPCB)) +
  geom_boxplot() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

print(pplot)
```

6

Shapiro-Wilkes

```
#test for normality. No data is normally distributed.
```

```
shapiro.test(pdata$p450)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  pdata$p450
## W = 0.66202, p-value < 2.2e-16
```

```
shapiro.test(apdata$sumPAH)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  apdata$sumPAH
## W = 0.36508, p-value < 2.2e-16
```

```
shapiro.test(apdata$hmwPAH)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  apdata$hmwPAH
## W = 0.34187, p-value < 2.2e-16
```

```
shapiro.test(apdata$sumPCB)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  apdata$sumPCB
## W = 0.82832, p-value < 2.2e-16
```

Kruskal-Wallis

```
#test for significant interaction

kruskal.test(p450 ~ site_name, data = pdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  p450 by site_name
## Kruskal-Wallis chi-squared = 140.14, df = 73, p-value = 3.886e-06
```

```
kruskal.test(p450 ~ reporting_area, data = pdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  p450 by reporting_area
## Kruskal-Wallis chi-squared = 30.543, df = 8, p-value = 0.0001694
```

Kruskal-Wallac Multiple Comparisons (post hoc)

Reporting Areas Are: 6 - East Juan de Fuca Strait 7 - San Juan Islands 8.1 - Deception Pass, Hope Island, and Skagit Bay 8.2 - Port Susan and Port Gardner 9 - Admiralty Inlet 10 - Seattle-Bremerton 11 - Tacoma-Vashon 12 - Hood Canal 13 - South Puget Sound

```
library(pgirmess)

# no significance confirmed between sites
# significant differences across reporting area pairwise matches: 8.2-12, 10-13, 11-12, and 12-13

mc_site<- as.data.frame(kruskalmc(p450 ~ site_name, data = pdata, method = "bonferroni"))
mc_reporting<- as.data.frame(kruskalmc(p450 ~ reporting_area, data = pdata, method = "bonferroni"))

#head(mc_site)
#head(mc_reporting)
```

Correlation test - p450 and SumPAH

```
# no correlation
correlation_result <- cor.test(apdata$p450, apdata$sumPAH, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  apdata$p450 and apdata$sumPAH
## t = -0.14165, df = 296, p-value = 0.8875
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1217399  0.1054864
## sample estimates:
##         cor
## -0.008233005
```

Correlation test - p450 and hmwPAH

```r
# no correlation
correlation_result <- cor.test(apdata$p450, apdata$hmwPAH, method = "pearson")
print(correlation_result)
```

```
## 
##  Pearson's product-moment correlation
## 
## data:  apdata$p450 and apdata$hmwPAH
## t = -0.37535, df = 296, p-value = 0.7077
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.13509767  0.09203713
## sample estimates:
##         cor 
## -0.02181172
```

Correlation test - p450 and SumPCB

```r
# no correlation
correlation_result <- cor.test(apdata$p450, apdata$sumPCB, method = "pearson")
print(correlation_result)
```

```
## 
##  Pearson's product-moment correlation
## 
## data:  apdata$p450 and apdata$sumPCB
## t = -1.1016, df = 296, p-value = 0.2715
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.17623849  0.05008716
## sample estimates:
##         cor 
## -0.06389724
```