

M11 Group Work: Correlations for TAs

Pablo E. Gutiérrez-Fonseca

2024-04-11

On Brightspace you have a data set **marine_redox.xlsx** that quantifies the relationships among temperature, algal biomass, detritus and oxygen concentrations and redox potential (Eh) in various ocean water samples. Measurements of Eh are used to evaluate geochemical speciation models, and Eh data can provide insights on the evolution and status of water quality in aqueous systems. We want to better understand what factors influence redox potential in ocean ecosystems to inform how climate change might impact redox potential (temperature) and if there are other water quality conditions we could target for management actions to minimize redox potential.

Your task is to examine the relationships between the various measured biotic and abiotic variables in this data set with the goal of finding which has the strongest, unique relationship with redox potential (Eh).

#1. Import libraries and load packages

```
library(tidyverse)
library(rstatix)
library(readxl)
library(dplyr)
library(PMCMRplus)
library(lessR)
library(ppcor)
```

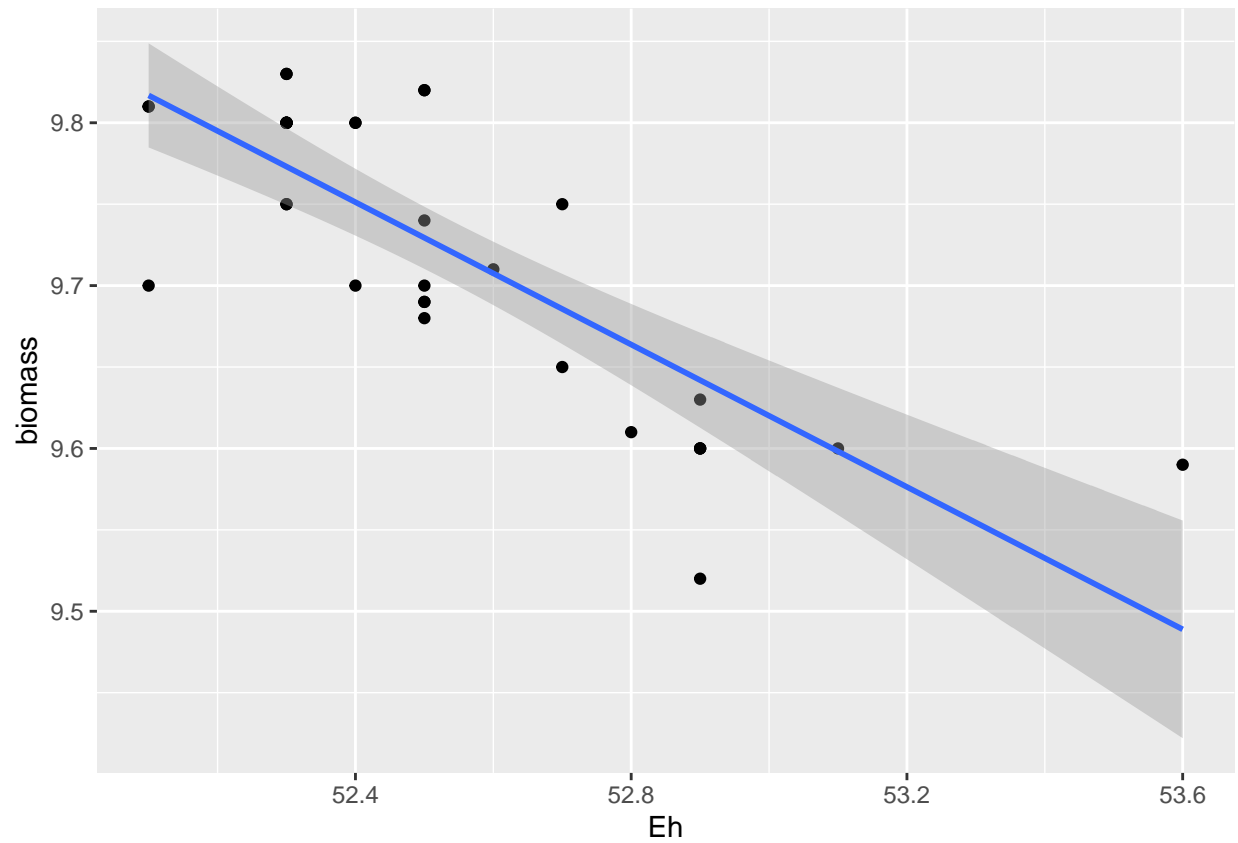
#2. Importing our data

```
redox <- read_excel("marine_redox.xlsx", sheet = 1, col_names = TRUE)
```

1. Your first task will be to visualize the relationships among variables. Visualize your data – examine scatterplots of each of your variables. Paste your figures here, and generally describe the expected type of correlation you see. **Eh and Biomass:** Include a scatterplot. Based on your visualization, describe if
 2. Linear (Y/N),
 3. Pos or Neg,
 4. Strength (weak, mild, moderate or strong).

```
ggplot(redox, aes(x=Eh, y=biomass)) +
  geom_point() +
  geom_smooth(method=lm)
```

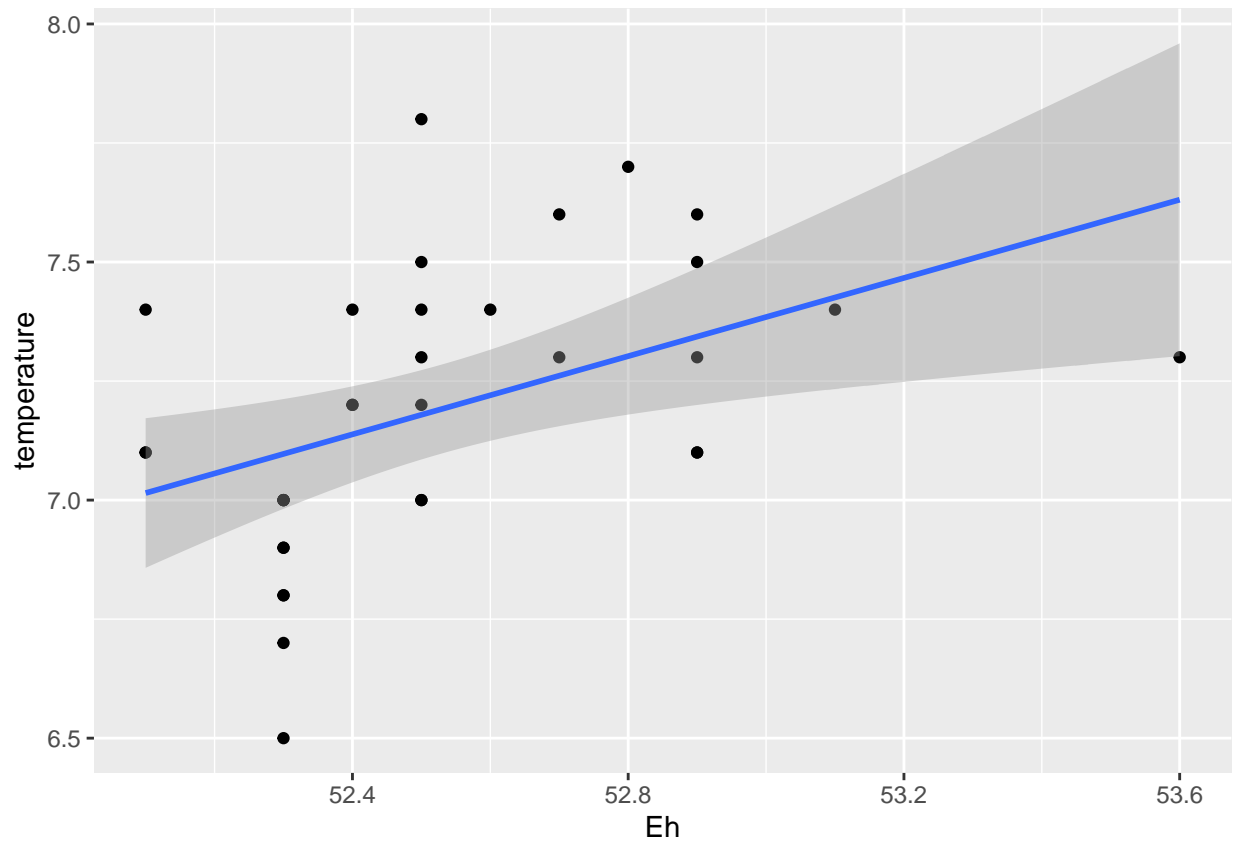
```
## `geom_smooth()` using formula = 'y ~ x'
```



Eh and Temperature: Include a scatterplot. Based on your visualization, describe if 1. Linear (Y/N), 2. Pos or Neg, 3. Strength (weak, mild, moderate or strong).

```
ggplot(redox, aes(x=Eh, y=temperature)) +  
  geom_point() +  
  geom_smooth(method=lm)
```

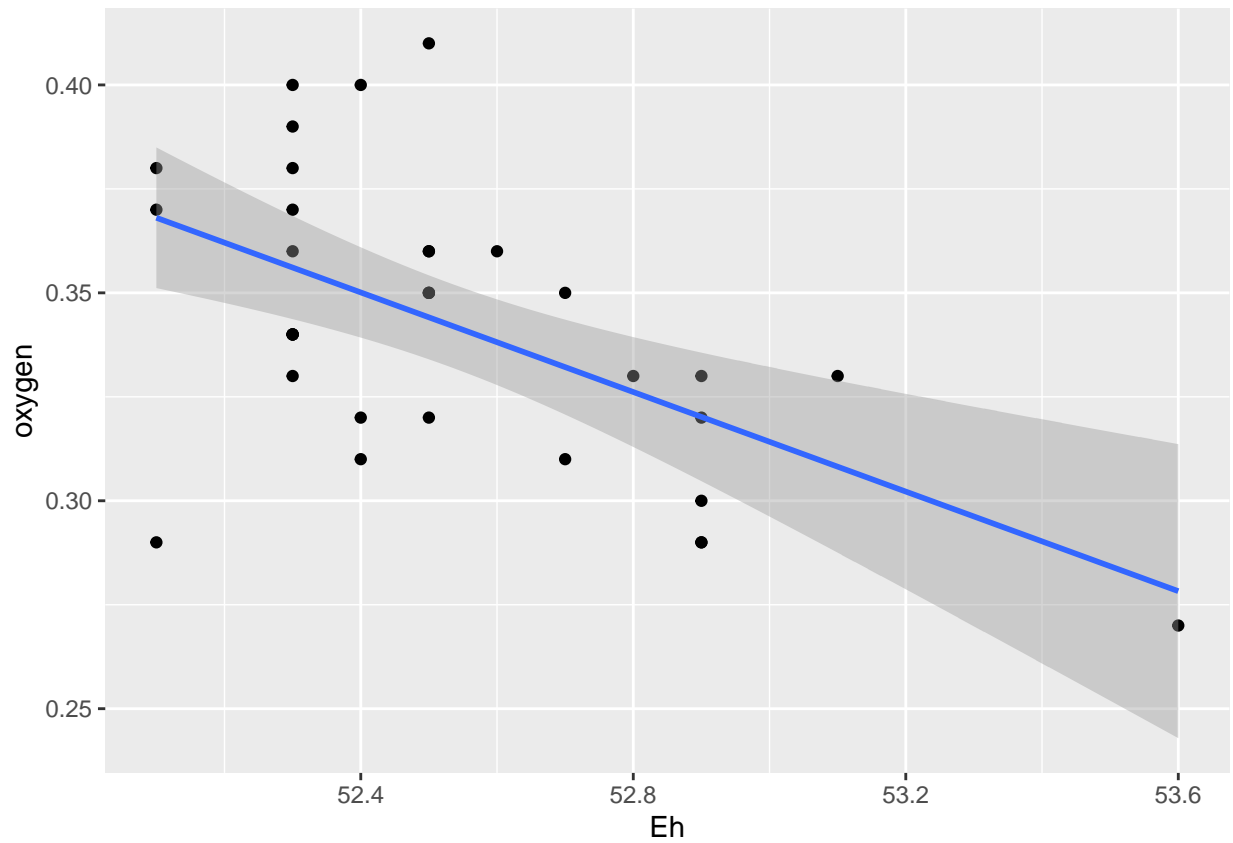
```
## `geom_smooth()` using formula = 'y ~ x'
```



Eh and Oxygen: Include a scatterplot. Based on your visualization, describe if 1. Linear (Y/N), 2. Pos or Neg, 3. Strength (weak, mild, moderate or strong).

```
ggplot(redox, aes(x=Eh, y=oxygen)) +  
  geom_point()+  
  geom_smooth(method=lm)
```

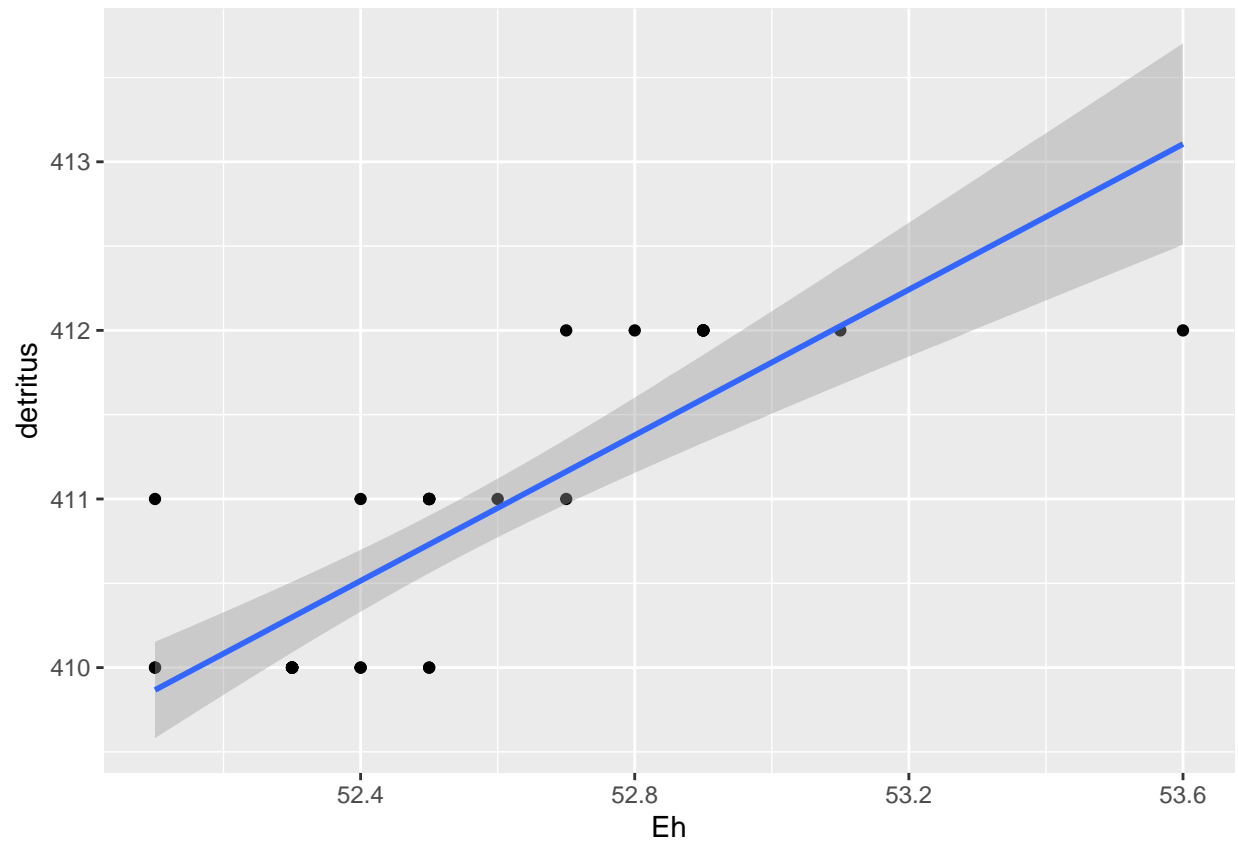
```
## `geom_smooth()` using formula = 'y ~ x'
```



Eh and Detritus: Include a scatterplot. Based on your visualization, describe if 1. Linear (Y/N), 2. Pos or Neg, 3. Strength (weak, mild, moderate or strong).

```
ggplot(redox, aes(x=Eh, y=detritus)) +  
  geom_point() +  
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



2. Now describe each of your variables and analyze your distributions. For clarity, work to aggregate your data into a table like this:

#pivot gives us more summary statistics.

```
pivot(redox, c(mean,median, sd, min, max), biomass)
```

```
##  n na  mean  mdn   sd min max
## 34  0 9.723 9.745 0.09 9.52 9.83
```

```
pivot(redox, c(mean,median, sd, min, max), temperature)
```

```
##  n na  mean mdn   sd min max
## 34  0 7.191 7.2 0.295 6.5 7.8
```

```
pivot(redox, c(mean,median, sd, min, max), oxygen)
```

```
##  n na  mean mdn   sd min max
## 34  0 0.342 0.34 0.03 0.27 0.41
```

```
pivot(redox, c(mean,median, sd, min, max), detritus)
```

```
##  n na  mean mdn   sd min max
## 34  0 410.794 411 0.845 410 412
```

```
pivot(redox, c(mean,median, sd, min, max), Eh)
```

```
##   n na   mean mdn    sd min max
##  34  0 52.529 52.5 0.322 52.1 53.6
```

```
#normality
shapiro.test(redox$biomass)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  redox$biomass
## W = 0.8962, p-value = 0.003645
```

```
shapiro.test(redox$temperature)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  redox$temperature
## W = 0.98575, p-value = 0.9272
```

```
shapiro.test(redox$oxygen)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  redox$oxygen
## W = 0.98373, p-value = 0.8805
```

```
shapiro.test(redox$detritus)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  redox$detritus
## W = 0.76059, p-value = 0.000004831
```

```
shapiro.test(redox$Eh)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  redox$Eh
## W = 0.88026, p-value = 0.001437
```

3. Time to run some correlations in R. Report the shorthand for each of your correlations with Eh. Be sure to clarify if you ran a Pearson's (r) or Spearman's (rho) correlation based on your normality. Eh and Biomass:

```
cor.test(redox$Eh,redox$biomass,  
         method='spearman')
```

```
##  
## Spearman's rank correlation rho  
##  
## data: redox$Eh and redox$biomass  
## S = 11605, p-value = 0.00000008368  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
##      rho  
## -0.773098
```

Eh and Temperature:

```
cor.test(redox$Eh,redox$temperature,  
         method='spearman')
```

```
##  
## Spearman's rank correlation rho  
##  
## data: redox$Eh and redox$temperature  
## S = 2667, p-value = 0.0002225  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
##      rho  
## 0.5925098
```

Eh and Biomass:

```
cor.test(redox$Eh,redox$oxygen,  
         method='spearman')
```

```
##  
## Spearman's rank correlation rho  
##  
## data: redox$Eh and redox$oxygen  
## S = 9753.9, p-value = 0.003244  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
##      rho  
## -0.490284
```

Eh and Biomass:

```
cor.test(redox$Eh,redox$detritus,  
         method='spearman')
```

```
##  
## Spearman's rank correlation rho
```

```
##
## data: redox$Eh and redox$detritus
## S = 1000.4, p-value = 0.0000000002665
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.8471446
```

4. Because we want to identify the metric with the strongest relationship on Eh, while controlling for any autocorrelation among variables, run partial correlations that includes all variables with a significant simple correlation with Eh. Report the results of your partial correlation and identify which variable has the strongest, unique relationship with Eh.

```
pcor(redox)
```

```
## $estimate
##           biomass temperature      oxygen  detritus      Eh
## biomass      1.00000000    0.1361763 -0.04392137 -0.8422946 -0.03603835
## temperature  0.13617633    1.0000000  0.14750239  0.4222405 -0.15077662
## oxygen       -0.04392137    0.1475024  1.00000000 -0.0621052 -0.35909890
## detritus     -0.84229457    0.4222405 -0.06210520  1.0000000  0.35556226
## Eh          -0.03603835   -0.1507766 -0.35909890  0.3555623  1.00000000
##
## $p.value
##           biomass temperature      oxygen      detritus
## biomass      0.0000000000000000  0.46511309 0.81451336 0.0000000002834301
## temperature  0.465113091840075   0.00000000 0.42844508 0.017972344786418
## oxygen       0.814513355142349   0.42844508 0.00000000 0.739966398297423
## detritus     0.0000000002834301  0.01797234 0.73996640 0.0000000000000000
## Eh          0.847374355359205   0.41815018 0.04726141 0.049644252443710
##           Eh
## biomass      0.84737436
## temperature  0.41815018
## oxygen       0.04726141
## detritus     0.04964425
## Eh          0.00000000
##
## $statistic
##           biomass temperature      oxygen  detritus      Eh
## biomass      0.0000000    0.7402275 -0.2367523 -8.4151166 -0.1941986
## temperature  0.7402275    0.0000000  0.8031094  2.5084116 -0.8213467
## oxygen       -0.2367523    0.8031094  0.0000000 -0.3350936 -2.0720109
## detritus     -8.4151166    2.5084116 -0.3350936  0.0000000  2.0486343
## Eh          -0.1941986   -0.8213467 -2.0720109  2.0486343  0.0000000
##
## $n
## [1] 34
##
## $gp
## [1] 3
##
## $method
## [1] "pearson"
```