Ben_Linear_Regression

TJ

10/14/2021

Libraries

First we need to load in the data

```
#Summer data
top100phyvSum <- read.csv("data/NewSumWinData/POWOW_RawCounts_EcotypePhyla_Summer.csv") %>%
  column_to_rownames(., var = "Sample") %>%
  subset(., select = eMED4:Verrucomicrobia)
top100phyvSum.hell <- decostand(top100phyvSum, "hellinger") #Preforming a hellinger transfermation

#Winter data
top100phyvWin <- read.csv("data/NewSumWinData/POWOW_RawCounts_EcotypePhyla_Winter.csv") %>%
  column_to_rownames(., var = "Sample") %>%
  subset(., select = eMED4:Verrucomicrobia)
top100phyvWin.hell <- decostand(top100phyvWin, "hellinger") #Preforming a hellinger transfermation
```

Summer analysis

Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eMIT9312lm_Sum_Intersept <- lm(eMIT9312 ~ 1, data = top100phyvSum.hell) #First we will define the inter
eMIT9312lm_Sum <- lm(eMIT9312 ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
  Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
  Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
  Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvSum.hell) #Now I will
```

```
vif(eMIT9312lm_Sum)
```

```
##                 Chloroflexi               Cyanobacteria
##                    9.439458                    3.903344
##         Deinococcus.Thermus               Euryarchaeota
##                    1.639400                    8.018580
##                  Firmicutes                Fusobacteria
##                    4.741412                    1.609413
##            Gemmatimonadetes             Gracilibacteria
```

```
##                    5.381990                    2.023682
##               Lentisphaerae Marinimicrobia_SAR406_clade
##                    2.010246                    8.027217
##                     PAUC34f             Planctomycetes
##                    3.398654                    2.102418
##             Proteobacteria                     SHA.109
##                    6.057502                    1.313680
##             Thaumarchaeota            Verrucomicrobia
##                   10.288416                    2.990001
```

```
eMIT9312lm_Sum_both <- step(eMIT9312lm_Sum_Intersept, direction='both', scope=formula(eMIT9312lm_Sum), 
```

```
eMIT9312lm_Sum_both$anova
```

```
##                     Step Df   Deviance Resid. Df Resid. Dev       AIC
## 1                       NA         NA        51   2.394479 -158.0601
## 2      + Proteobacteria -1 0.40757846        50   1.986901 -165.7627
## 3 + Deinococcus.Thermus -1 0.14579905        49   1.841102 -167.7257
## 4      + Planctomycetes -1 0.09743226        48   1.743669 -168.5531
```

```
eMIT9312lm_Sum_both$coefficients
```

```
##         (Intercept)      Proteobacteria Deinococcus.Thermus      Planctomycetes
##           0.3153202          -0.3806562          10.4842631          -0.6532889
```

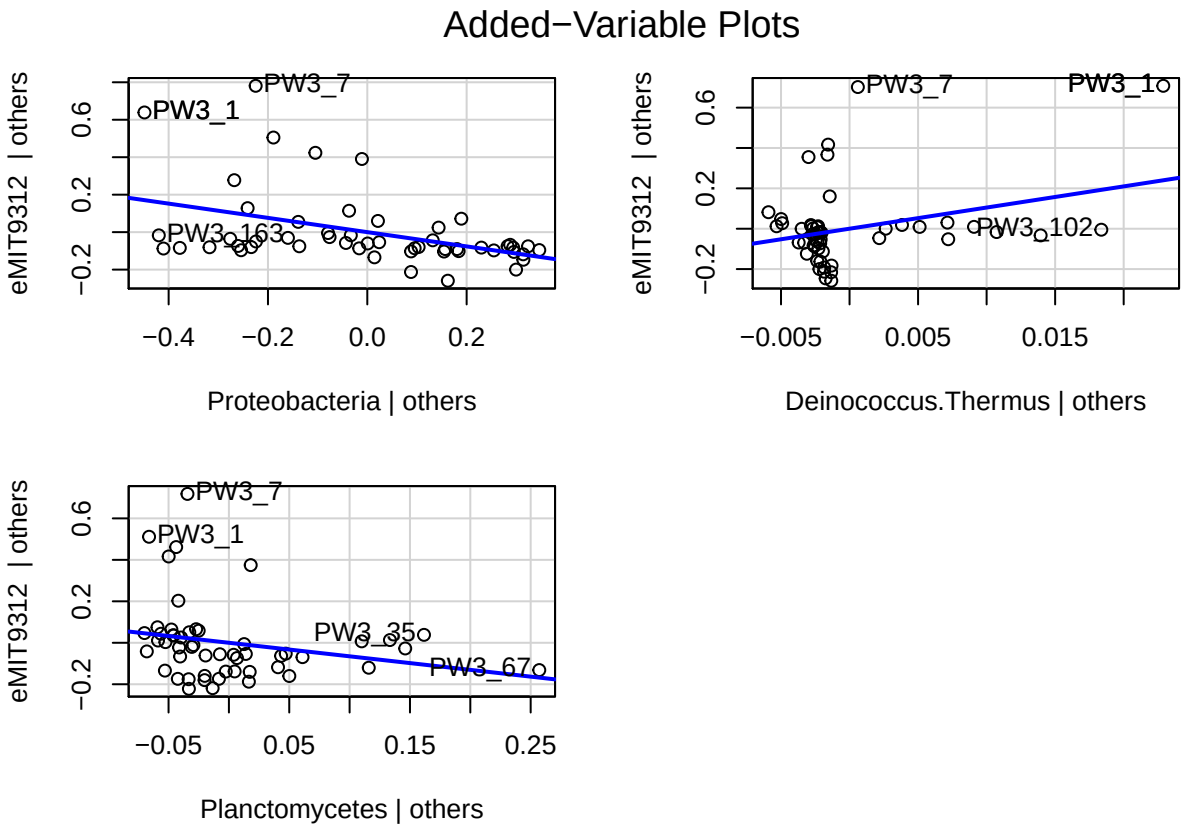
```
summary(eMIT9312lm_Sum_both)
```

```
##
## Call:
## lm(formula = eMIT9312 ~ Proteobacteria + Deinococcus.Thermus +
##     Planctomycetes, data = top100phyvSum.hell)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24368 -0.10140 -0.03093  0.03618  0.69555
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.31532    0.06936   4.546 3.71e-05 ***
## Proteobacteria      -0.38066    0.11760  -3.237  0.00219 **
## Deinococcus.Thermus 10.48426    4.59146   2.283  0.02687 *
## Planctomycetes      -0.65329    0.39890  -1.638  0.10802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1906 on 48 degrees of freedom
## Multiple R-squared:  0.2718, Adjusted R-squared:  0.2263
## F-statistic: 5.972 on 3 and 48 DF,  p-value: 0.001521
```

When comparing eMIT9312 against the Phylum in the summer samples, we found this ecotype has significiant relationships with Proteobacteria ($p < .005$) and Deinococcus-Thermus ($p < .05$). Specifically we found a

0.4% decrease (± 0.11) in eMIT9312 abundance for every 1% increase in Proteobacteria and a 10% increase (± 4.6) in eMIT9312 abundance for every 1% increase in Deinococcus-Thermus abundance. Though we saw a decrease of .66 in eMIT9312 for every 1% increase in Planctomycetes, this decrease was not statistically significant (p > .05).

```
avPlots(eMIT9312lm_Sum_both)
```



Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eNATL2Alm_Sum_Intersept <- lm(eNATL2A ~ 1, data = top100phyvSum.hell) #First we will define the intersep
eNATL2Alm_Sum <- lm(eNATL2A ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
  Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
  Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
  Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvSum.hell) #Now I will
```

#Preforming a stepwise in both directions to find the best model

```
eNATL2Alm_Sum_both <- step(eNATL2Alm_Sum_Intersept, direction='both', scope=formula(eNATL2Alm_Sum), tra
```

```
eNATL2Alm_Sum_both$anova
```

```
##   Step Df Deviance Resid. Df Resid. Dev       AIC
## 1      NA       NA        51   2.742193 -151.0093
```

```
eNATL2Alm_Sum_both$coefficients
```

```
## (Intercept)
##    0.152451
```

```
summary(eNATL2Alm_Sum_both)
```

```
##
## Call:
## lm(formula = eNATL2A ~ 1, data = top100phyvSum.hell)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1525 -0.1502 -0.1426  0.1080  0.6204
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.15245    0.03216   4.741 1.75e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2319 on 51 degrees of freedom
```

There is no sig correlation for eNATL2A

Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eMIT9313lm_Sum_Intersept <- lm(eMIT9313 ~ 1, data = top100phyvSum.hell) #First we will define the inter
eMIT9313lm_Sum <- lm(eMIT9313 ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
  Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
  Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
  Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvSum.hell) #Now I will
```

```
#Preforming a stepwise in both directions to find the best model
eMIT9313lm_Sum_both <- step(eMIT9313lm_Sum_Intersept, direction='both', scope=formula(eMIT9313lm_Sum),
```

```
eMIT9313lm_Sum_both$anova
```

```
##            Step Df   Deviance Resid. Df Resid. Dev       AIC
## 1               NA         NA        51 0.06073271 -349.1309
## 2 + Chloroflexi -1 0.00295798        50 0.05777473 -349.7273
```

```
eMIT9313lm_Sum_both$coefficients
```

```
## (Intercept) Chloroflexi
##  0.01764023 -0.22909016
```

```
summary(eMIT9313lm_Sum_both)
```

```
##
## Call:
## lm(formula = eMIT9313 ~ Chloroflexi, data = top100phyvSum.hell)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.017640 -0.016232 -0.009533 -0.001025  0.150683
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.017640   0.005875   3.003  0.00417 **
## Chloroflexi -0.229090   0.143183  -1.600  0.11590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03399 on 50 degrees of freedom
## Multiple R-squared:  0.0487, Adjusted R-squared:  0.02968
## F-statistic:  2.56 on 1 and 50 DF,  p-value: 0.1159
```

Only the intercept is sig here

Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eMED4lm_Sum_Intersept <- lm(eMED4 ~ 1, data = top100phyvSum.hell) #First we will define the intersept
eMED4lm_Sum <- lm(eMED4 ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
  Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
  Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
  Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvSum.hell) #Now I will
```

```
#Preforming a stepwise in both directions to find the best model
eMED4lm_Sum_both <- step(eMED4lm_Sum_Intersept, direction='both', scope=formula(eMED4lm_Sum), trace=0)
```

```
eMED4lm_Sum_both$anova
```

```
##                     Step Df  Deviance Resid. Df Resid. Dev       AIC
## 1                         NA        NA        51   6.705967 -104.5088
## 2      + Proteobacteria -1 3.6507605        50   3.055207 -143.3886
## 3 + Deinococcus.Thermus -1 0.3633326        49   2.691874 -147.9723
## 4       + Euryarchaeota -1 0.1121722        48   2.579702 -148.1856
```

```
eMED4lm_Sum_both$coefficients
```

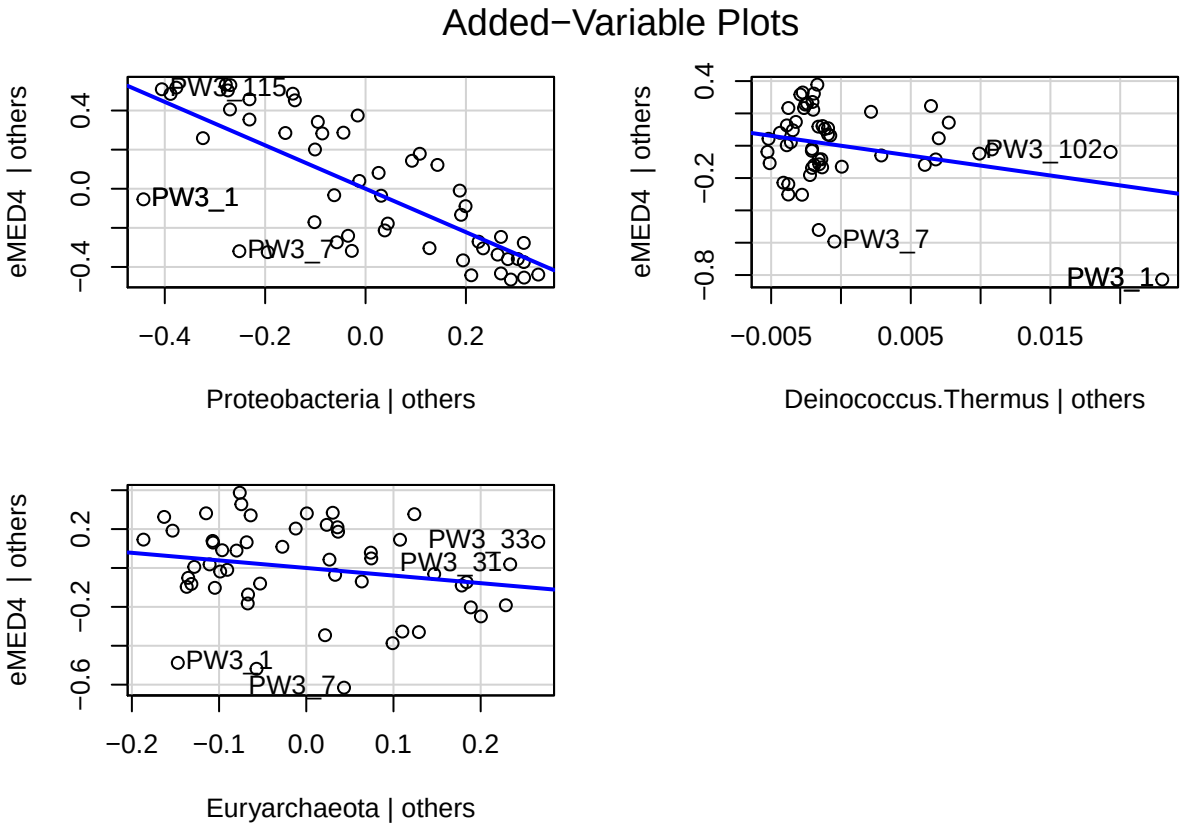
```
##         (Intercept)      Proteobacteria Deinococcus.Thermus       Euryarchaeota
##           1.0359735          -1.1104049         -12.2798426          -0.3903005
```

```
summary(eMED4lm_Sum_both)
```

```
##
## Call:
## lm(formula = eMED4 ~ Proteobacteria + Deinococcus.Thermus + Euryarchaeota,
##     data = top100phyvSum.hell)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59850 -0.13013  0.03972  0.18986  0.35708
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.03597    0.08509  12.176 2.75e-16 ***
## Proteobacteria       -1.11040    0.14299  -7.766 4.98e-10 ***
## Deinococcus.Thermus -12.27984    5.64451  -2.176   0.0345 *
## Euryarchaeota        -0.39030    0.27016  -1.445   0.1550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2318 on 48 degrees of freedom
## Multiple R-squared:  0.6153, Adjusted R-squared:  0.5913
## F-statistic: 25.59 on 3 and 48 DF,  p-value: 4.917e-10
```

When comparing eMED4 against the Phylum in the summary samples, we found this ecotype has significiant relationships with Proteobacteria ($p < .001$) and Deinococcus-Thermus ($p < .05$). Specifically we found a 1.1% decrease ($\pm$ 0.14) in eMED4 abundance for every 1% increase in Proteobacteria and a 12% decrease ($\pm$ 5.6) in eMED4 abundance for every 1% increase in Deinococcus-Thermus abundance. Though we saw a decrease of .4% in eMED4 for every 1% increase in Euryarchaeota, this decrease was not statistically significant ($p > .05$).

```
avPlots(eMED4lm_Sum_both)
```



Winter analysis

Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eMIT9312lm_Win_Intersept <- lm(eMIT9312 ~ 1, data = top100phyvWin.hell) #First we will define the inter
eMIT9312lm_Win <- lm(eMIT9312 ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
                       Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
                       Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
                       Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvW
```

```
vif(eMIT9312lm_Win)
```

```
##                 Chloroflexi               Cyanobacteria
##                    7.281146                    4.581259
##         Deinococcus.Thermus               Euryarchaeota
##                    3.314307                   15.599366
##                  Firmicutes               Fusobacteria
##                    9.626650                    1.884679
##            Gemmatimonadetes             Gracilibacteria
##                    3.660177                    4.701867
##               Lentisphaerae Marinimicrobia_SAR406_clade
##                    3.977338                   26.827255
```

```
##                     PAUC34f              Planctomycetes
##                    4.937170                    3.616936
##              Proteobacteria                     SHA.109
##                    5.094804                    3.066283
##              Thaumarchaeota             Verrucomicrobia
##                    4.332598                    8.003607
```

```
eMIT9312lm_Win_both <- step(eMIT9312lm_Win_Intersept, direction='both', scope=formula(eMIT9312lm_Win),
```

```
eMIT9312lm_Win_both$anova
```

```
##               Step Df Deviance Resid. Df Resid. Dev        AIC
## 1                   NA       NA        42   5.222196  -88.65613
## 2 + Proteobacteria -1 2.622288        41   2.599908 -116.64614
## 3    + Chloroflexi -1 0.438386        40   2.161522 -122.58667
```

```
eMIT9312lm_Win_both$coefficients
```

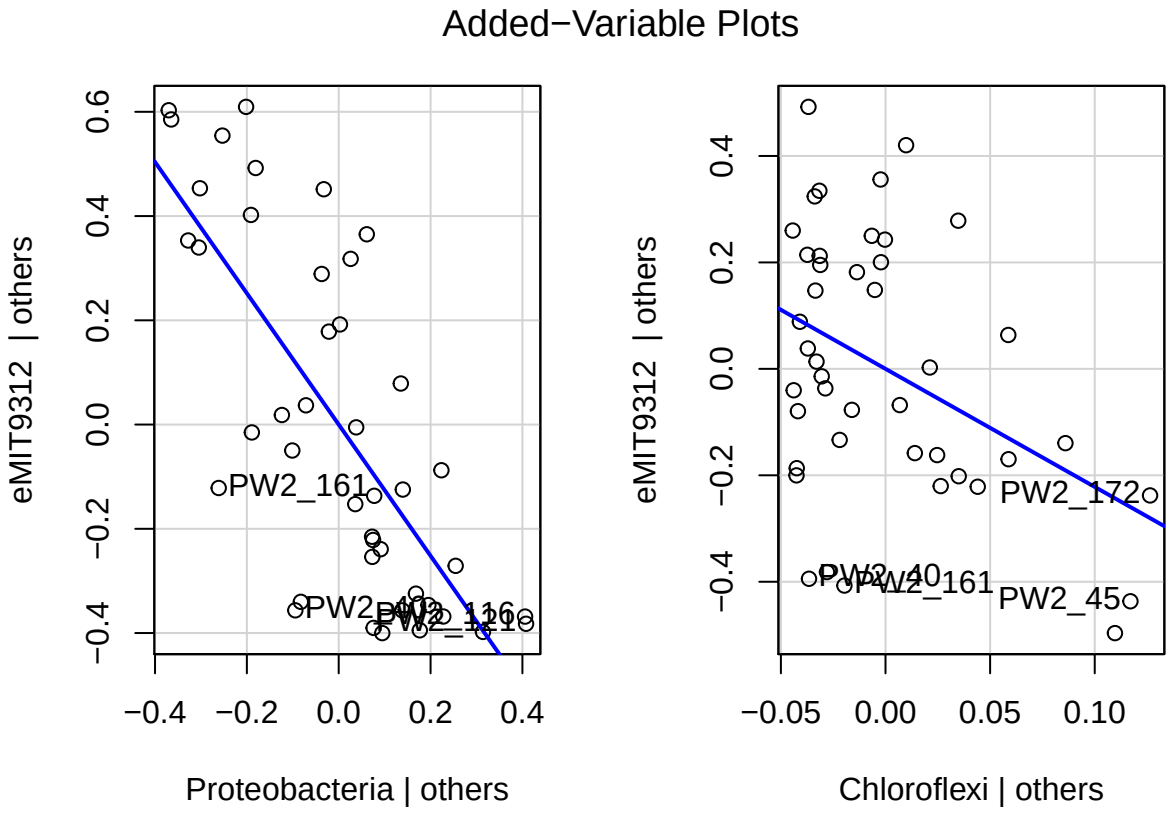
```
##    (Intercept) Proteobacteria    Chloroflexi
##      0.9885717     -1.2591248     -2.2176446
```

```
summary(eMIT9312lm_Win_both)
```

```
##
## Call:
## lm(formula = eMIT9312 ~ Proteobacteria + Chloroflexi, data = top100phyvWin.hell)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47548 -0.13240 -0.03934  0.15643  0.44203
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.98857    0.09644  10.251 9.40e-13 ***
## Proteobacteria -1.25912    0.17868  -7.047 1.61e-08 ***
## Chloroflexi    -2.21764    0.77860  -2.848  0.00691 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2325 on 40 degrees of freedom
## Multiple R-squared:  0.5861, Adjusted R-squared:  0.5654
## F-statistic: 28.32 on 2 and 40 DF,  p-value: 2.178e-08
```

When comparing eMIT9312 against the Phylum in the winter samples, we found this ecotype has significant relationships with Proteobacteria ($p < .001$) and Chloroflexi ($p < .001$). Specifically we found a 1% decrease ($\pm$ 0.2) in eMIT9312 abundance for every 1% increase in Proteobacteria and a 2% decrease ($\pm$ .8) in eMIT9312 abundance for every 1% increase in Chloroflexi abundance.

```
avPlots(eMIT9312lm_Win_both)
```



Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eNATL2Alm_Win_Intersept <- lm(eNATL2A ~ 1, data = top100phyvWin.hell) #First we will define the intersep
eNATL2Alm_Win <- lm(eNATL2A ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
                      Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
                      Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
                      Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvWi
```

```
#Preforming a stepwise in both directions to find the best model
eNATL2Alm_Win_both <- step(eNATL2Alm_Win_Intersept, direction='both', scope=formula(eNATL2Alm_Win), tra
```

```
eNATL2Alm_Win_both$anova
```

```
##                 Step Df   Deviance Resid. Df Resid. Dev       AIC
## 1                    NA         NA        42  1.0863957 -156.1684
## 2    + Lentisphaerae -1 0.20703655        41  0.8793592 -163.2598
## 3          + SHA.109 -1 0.08435733        40  0.7950018 -165.5963
## 4    + Euryarchaeota -1 0.14629648        39  0.6487054 -172.3410
## 5 + Gemmatimonadetes -1 0.02998499        38  0.6187204 -172.3760
## 6      + Chloroflexi -1 0.06308949        37  0.5556309 -175.0006
```

```
eNATL2Alm_Win_both$coefficients
```

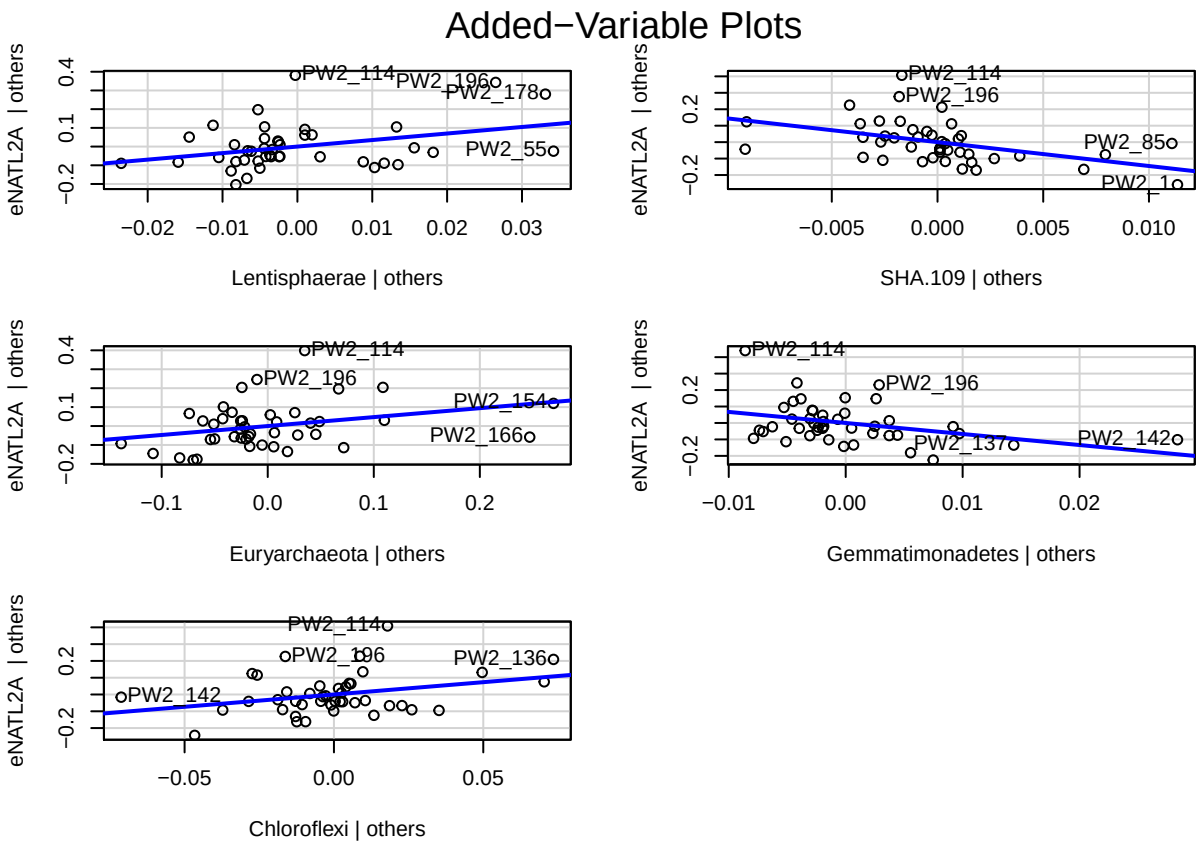
```
##      (Intercept)    Lentisphaerae          SHA.109    Euryarchaeota
##        0.1238786        3.4909464      -14.5247285        0.4720197
## Gemmatimonadetes      Chloroflexi
##       -6.7064609        1.4658046
```

```
summary(eNATL2Alm_Win_both)
```

```
##
## Call:
## lm(formula = eNATL2A ~ Lentisphaerae + SHA.109 + Euryarchaeota +
##     Gemmatimonadetes + Chloroflexi, data = top100phyvWin.hell)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17516 -0.06311 -0.02168  0.05683  0.38218
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.1239     0.0305   4.062 0.000243 ***
## Lentisphaerae      3.4910     1.5603   2.237 0.031372 *
## SHA.109          -14.5247     4.7546  -3.055 0.004161 **
## Euryarchaeota      0.4720     0.2445   1.931 0.061185 .
## Gemmatimonadetes  -6.7065     2.8527  -2.351 0.024165 *
## Chloroflexi        1.4658     0.7151   2.050 0.047534 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1225 on 37 degrees of freedom
## Multiple R-squared:  0.4886, Adjusted R-squared:  0.4194
## F-statistic: 7.069 on 5 and 37 DF,  p-value: 1e-04
```

When comparing eNATL2A against the Phylum in the winter samples, we found this ecotype has significant relationships with Lentisphaerae ($p < .05$), the uncultured phylum SHA.109 ($p < .005$), Gemmatimonadetes ($p < .05$), and Chloroflexi ($p < .05$). Specifically we found a 3.4% increase ($\pm$ 1.5) in eNATL2A abundance for every 1% increase in Lentisphaerae, a 15% decrease ($\pm$ 5) in eNATL2A abundance for every 1% increase in SHA.109 abundance, a 7% ($\pm$ 3) decrease in eNATL2A abundance for every 1% increase in Gemmatimonadetes, and a 2% increase in eNATL2A abundance for every 1% increase in Chloroflexi. While we found a .5% increase in eNATL2A for every 1% increase in Euryarchaeota, this was not statistically significant ($p > .05$)

```
avPlots(eNATL2Alm_Win_both)
```



Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eMIT9313lm_Win_Intersept <- lm(eMIT9313 ~ 1, data = top100phyvWin.hell) #First we will define the inter
eMIT9313lm_Win <- lm(eMIT9313 ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
                       Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
                       Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
                       Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvW
```

```
#Preforming a stepwise in both directions to find the best model
eMIT9313lm_Win_both <- step(eMIT9313lm_Win_Intersept, direction='both', scope=formula(eMIT9313lm_Win),
```

```
eMIT9313lm_Win_both$anova
```

```
##                             Step Df     Deviance Resid. Df Resid. Dev       AIC
## 1                                 NA           NA        42 0.09969080 -258.8759
## 2               + Lentisphaerae -1 0.015145467        41 0.08454533 -263.9617
## 3                 + Chloroflexi -1 0.006593508        40 0.07795182 -265.4532
## 4         + Deinococcus.Thermus -1 0.010578864        39 0.06737296 -269.7246
## 5 + Marinimicrobia_SAR406_clade -1 0.010032686        38 0.05734027 -274.6579
## 6               + Euryarchaeota -1 0.007122087        37 0.05021818 -278.3609
```

```
eMIT9313lm_Win_both$coefficients
```

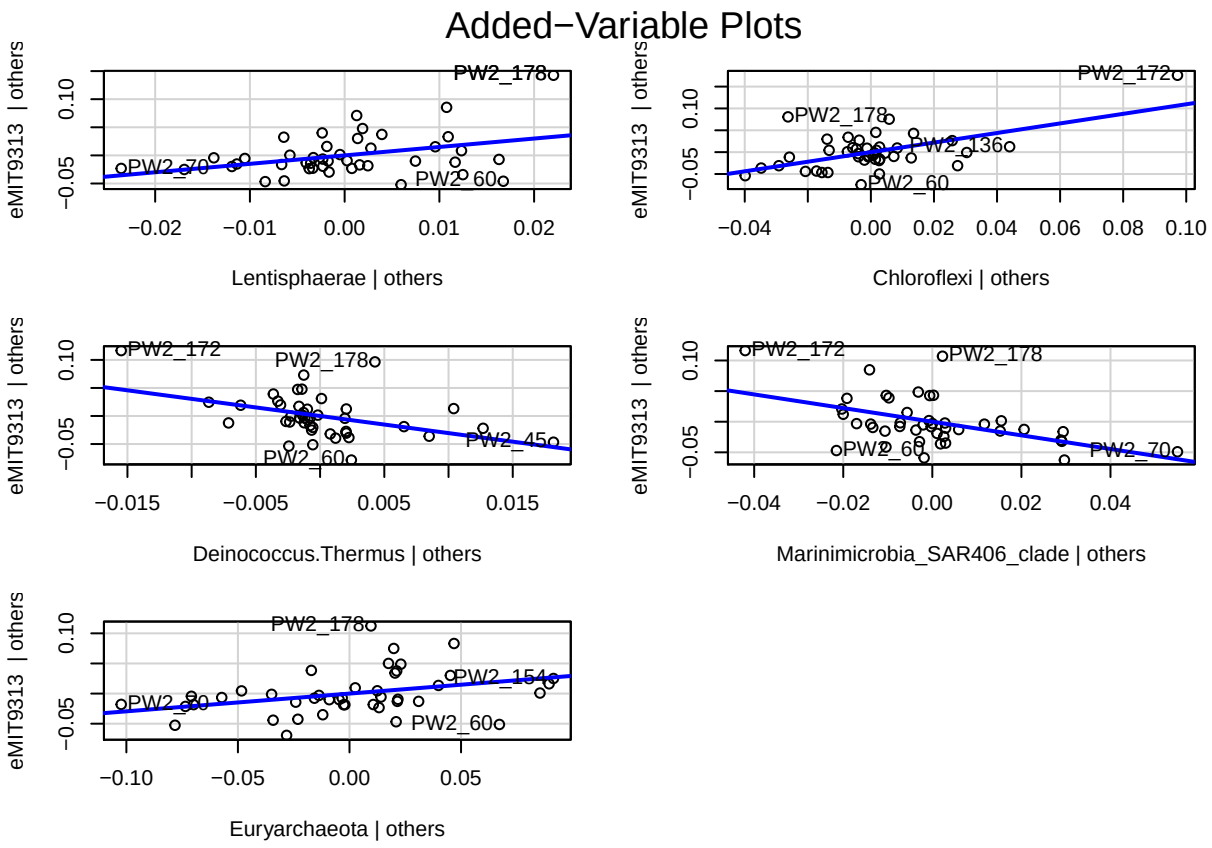
```
##                  (Intercept)               Lentisphaerae
##                    0.0101600                   1.4998094
##                  Chloroflexi         Deinococcus.Thermus
##                    1.0958169                  -3.0503470
## Marinimicrobia_SAR406_clade               Euryarchaeota
##                   -1.1138399                   0.2947599
```

```
summary(eMIT9313lm_Win_both)
```

```
##
## Call:
## lm(formula = eMIT9313 ~ Lentisphaerae + Chloroflexi + Deinococcus.Thermus +
##     Marinimicrobia_SAR406_clade + Euryarchaeota, data = top100phyvWin.hell)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.071085 -0.020302 -0.003054  0.014299  0.109504
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   0.010160   0.009119   1.114 0.272415
## Lentisphaerae                 1.499809   0.608108   2.466 0.018406 *
## Chloroflexi                   1.095817   0.253925   4.316 0.000114 ***
## Deinococcus.Thermus          -3.050347   1.043942  -2.922 0.005899 **
## Marinimicrobia_SAR406_clade -1.113840   0.330429  -3.371 0.001765 **
## Euryarchaeota                 0.294760   0.128675   2.291 0.027774 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03684 on 37 degrees of freedom
## Multiple R-squared:  0.4963, Adjusted R-squared:  0.4282
## F-statistic:  7.29 on 5 and 37 DF,  p-value: 7.716e-05
```

When comparing eMIT9313 against the Phylum in the winter samples, we found this ecotype has significant relationships with Lentisphaerae ($p < .05$), the uncultured phylum Chloroflexi ($p < .0005$), Deinococcus-Thermus ($p < .01$), Marinimicrobia_SAR406_clade ($p < .005$), and Euryarchaeota ($p < .05$). Specifically we found a 1.5% increase ($\pm$ .6) in eMIT9313 abundance for every 1% increase in Lentisphaerae, a 1.1% increase ($\pm$ .25) in eMIT9313 abundance for every 1% increase in Chloroflexi abundance, a 3% ($\pm$ 1) decrease in eMIT9313 abundance for every 1% increase in Deinococcus-Thermus, a 1.1% ($\pm$ .33) decrease in eMIT9313 abundance for every 1% increase in Marinimicrobia_SAR406_clade, and a .3% ($\pm$ .13) increase in eMIT9313 abundance for every 1% increase in Euryarchaeota.

```
avPlots(eMIT9313lm_Win_both)
```



Now I want to evaluate the correlations between eMIT9312 and the top100 phylum

```
eMED4lm_Win_Intersept <- lm(eMED4 ~ 1, data = top100phyvWin.hell) #First we will define the intersept
eMED4lm_Win <- lm(eMED4 ~  Chloroflexi + Cyanobacteria + Deinococcus.Thermus + Euryarchaeota +
                    Firmicutes + Fusobacteria +  Gemmatimonadetes + Gracilibacteria +
                    Lentisphaerae +  Marinimicrobia_SAR406_clade + PAUC34f +  Planctomycetes +
                    Proteobacteria + SHA.109 +  Thaumarchaeota + Verrucomicrobia, data = top100phyvWin.h
```

```
#Preforming a stepwise in both directions to find the best model
eMED4lm_Win_both <- step(eMED4lm_Win_Intersept, direction='both', scope=formula(eMED4lm_Win), trace=0)
```

```
eMED4lm_Win_both$anova
```

```
##               Step Df  Deviance Resid. Df Resid. Dev       AIC
## 1                   NA        NA        42   2.124159 -127.3364
## 2 + Proteobacteria -1 0.5683472        41   1.555811 -138.7257
## 3 + Planctomycetes -1 0.1313164        40   1.424495 -140.5175
```

```
eMED4lm_Win_both$coefficients
```

```
##    (Intercept) Proteobacteria Planctomycetes
##      0.4911182     -0.6401957      0.6483462
```

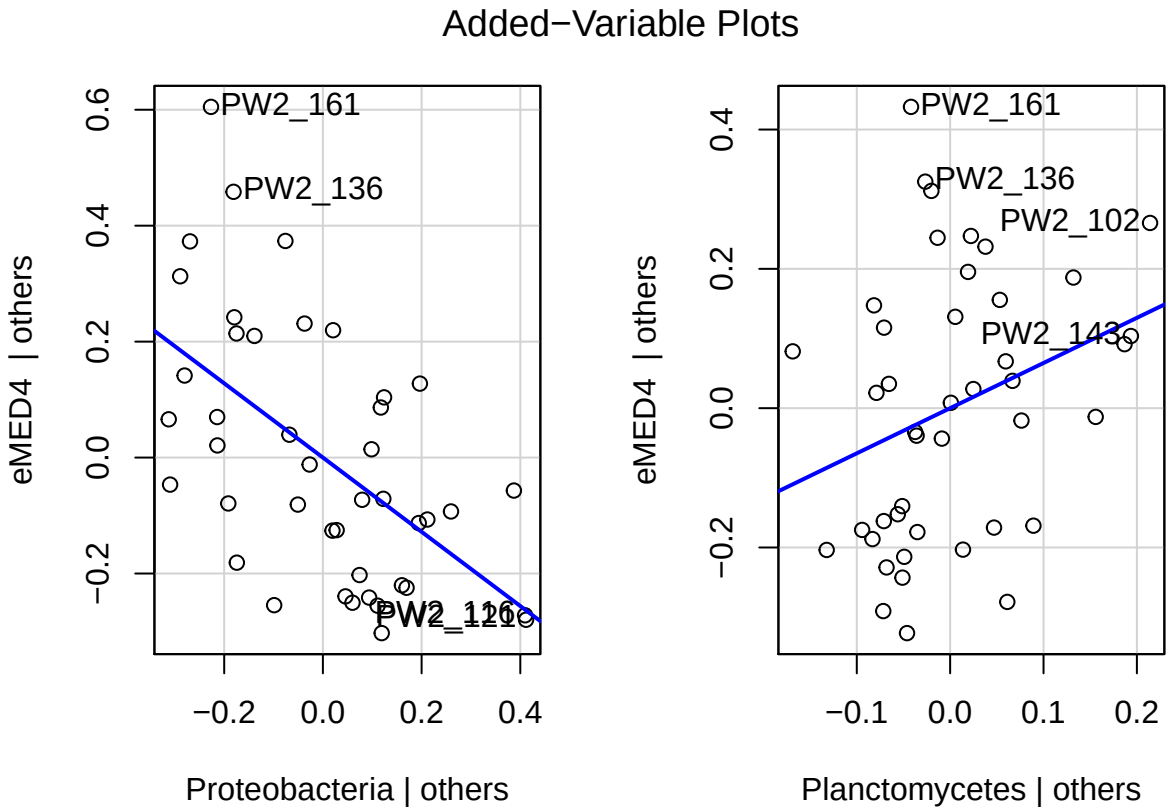
```
summary(eMED4lm_Win_both)
```

```
##
## Call:
## lm(formula = eMED4 ~ Proteobacteria + Planctomycetes, data = top100phyvWin.hell)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31781 -0.12590 -0.01619  0.12732  0.45972
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.49112    0.08053   6.099 3.44e-07 ***
## Proteobacteria -0.64020    0.14843  -4.313 0.000102 ***
## Planctomycetes  0.64835    0.33764   1.920 0.061976 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1887 on 40 degrees of freedom
## Multiple R-squared:  0.3294, Adjusted R-squared:  0.2959
## F-statistic: 9.823 on 2 and 40 DF,  p-value: 0.0003384
```

When comparing eMED4 against the Phylum in the winter samples, we found this ecotype has significant relationships with Proteobacteria ($p < .05$)

Specifically we found a .6% decrease ($\pm$ .1) in eMED4 abundance for every 1% increase in Proteobacteria. While we did find a .6% increase in eMED4 abundance for every 1% increase in Planctomycetes, this was not statistically significant ($p > .05$)

```
avPlots(eMED4lm_Win_both)
```


Added−Variable Plots
eMED4 | others
Proteobacteria | others
PW2_161
PW2_136
PW2_121
−0.2
0.0
0.2
0.4
0.6
Planctomycetes | others
PW2_102
PW2_143
−0.1
0.1