Data Exploration

C. Coleman

10/7/2021

# The following is an exploration of Stream Health Indicator correlation and comparison. The exploration will be structured as follows:

### Each dataframe (MIDO, NORO, and BICO) will be analyzed separately. Within each dataframe, there are data explaining biological indicator values, stream chemistry, and fecal indicators over time. The only group with multiple components is stream chemistry. I plan to do a series of correlative analysis showing if there is some sort of relationship between the three indicator types to help inform further analysis via principle component, multivariate analysis, etc. Therfore, the goal of this exploration is basically to summarize how these indicators correlate.

#### First, we need to load all the libraries we will need.

library(dplyr) #data manipulation

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(here) #setting pathways for saving files

## here() starts at C:/Users/socce/Downloads/BIOS8060/CARTERCOLEMAN\_MADA\_PROJECT

library(tidyverse) #Working with multiple Tidy packages

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.5 v stringr 1.4.0  
## v tidyr 1.1.4 v forcats 0.5.1  
## v readr 2.0.2

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(tidyr) #data cleaning  
library(readr) #read in files  
library(stringr) #working with strings  
library(gridExtra) #grouping figure outputs together

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

#### Loading the data from the processing script. Note that because you are loading a library at the beginning, it may either take a while or take several tries.

#Creates locations from file pathways that the required data frame .RDS files are stored.  
MIDO\_location <- here::here("data","processed\_data","MIDO.RDS")  
NORO\_location <- here::here("data","processed\_data","NORO.RDS")  
BICO\_location <- here::here("data","processed\_data","BICO.RDS")  
  
#Reads files called from defined locations  
MIDO <- readRDS(MIDO\_location)  
NORO <- readRDS(NORO\_location)  
BICO <- readRDS(BICO\_location)

#### take a look at the data frames to make sure everything worked correctly.

view(MIDO)  
view(NORO)  
view(BICO)

## MIDO

*This section of regression analysis and plotting deals exclusively with the MIDO data set.*

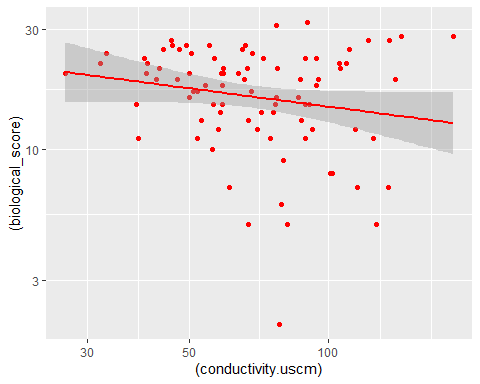
### *Biological Indicator:*

#### Biological Vs Conductivity:

*This plot shows the relationship between biological score as a function of conductivity.*

#Plot biological score vs. conductivity  
BS\_Con <- MIDO %>%   
 ggplot(aes(x=(conductivity.uscm),   
 y=(biological\_score))) +  
 geom\_point(color = "red") +  
 geom\_smooth(method=lm, color = "red") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
print(BS\_Con)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for BS\_Con.

#Linear regression analysis for biological score versus conductivity.  
BS\_Con\_lm <- lm(biological\_score ~ conductivity.uscm, data=MIDO)  
  
summary(BS\_Con\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ conductivity.uscm, data = MIDO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.6752 -5.0804 0.7546 5.0198 14.5213   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 19.00362 1.95783 9.706 1.8e-15 \*\*\*  
## conductivity.uscm -0.01694 0.02441 -0.694 0.489   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.721 on 86 degrees of freedom  
## Multiple R-squared: 0.005574, Adjusted R-squared: -0.005989   
## F-statistic: 0.482 on 1 and 86 DF, p-value: 0.4894

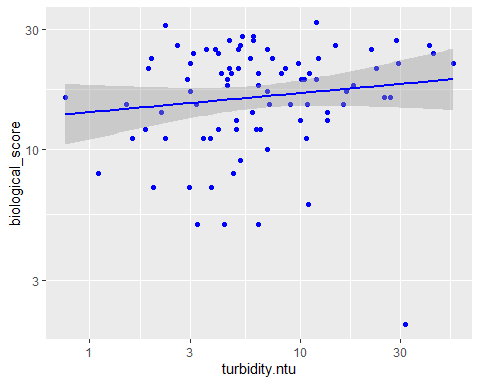
#### With *m = -0.017* and *p = 0.49*, there is a *negative*, *non-significant* correlation between *biological score* and *conductivity*.

#### Biological Vs Turbidity:

*This plot shows the relationship between biological score as a function of turbidity.*

#Plot biological score vs. turbidity  
BS\_Turb <- MIDO %>%  
 ggplot(aes(x=turbidity.ntu, y=biological\_score)) +  
 geom\_point(color = "blue") +  
 geom\_smooth(method=lm, color = "blue") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(BS\_Turb)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for BS\_Turb.

BS\_Turb\_lm <- lm(biological\_score ~ turbidity.ntu, data=MIDO)  
  
summary(BS\_Turb\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ turbidity.ntu, data = MIDO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.2293 -4.9714 0.8791 4.9388 14.0396   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 16.70297 0.98476 16.961 <2e-16 \*\*\*  
## turbidity.ntu 0.11195 0.07386 1.516 0.133   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.652 on 86 degrees of freedom  
## Multiple R-squared: 0.02602, Adjusted R-squared: 0.01469   
## F-statistic: 2.297 on 1 and 86 DF, p-value: 0.1333

#### With *m = 0.11* and *p = 0.13*, there is a *positive*, *non-significant* correlation between *biological score* and *turbidity*.

#### Biological Vs Nitrate Concentration

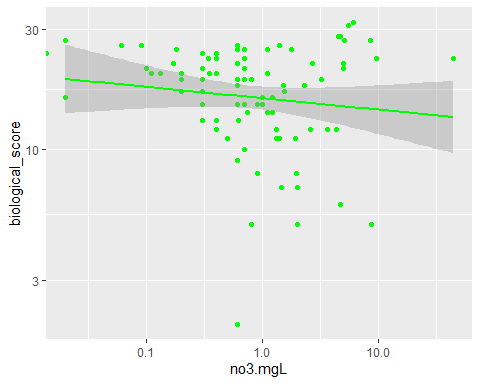
*This plot shows the relationship between biological score as a function of nitrate (no3) concentration.*

BS\_no3 <- MIDO %>%  
 ggplot(aes(x=no3.mgL, y=biological\_score)) +  
 geom\_point(color = "green") +  
 geom\_smooth(method=lm, color = "green") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(BS\_no3)

## Warning: Transformation introduced infinite values in continuous x-axis  
  
## Warning: Transformation introduced infinite values in continuous x-axis

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat\_smooth).



#### Now run linear regression for BS\_no3.

BS\_no3\_lm <- lm(biological\_score ~ no3.mgL, data=MIDO)  
  
summary(BS\_no3\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ no3.mgL, data = MIDO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.5016 -4.7460 0.7879 4.7965 13.6079   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.4045 0.7737 22.496 <2e-16 \*\*\*  
## no3.mgL 0.1619 0.1451 1.116 0.268   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.692 on 86 degrees of freedom  
## Multiple R-squared: 0.01427, Adjusted R-squared: 0.002806   
## F-statistic: 1.245 on 1 and 86 DF, p-value: 0.2677

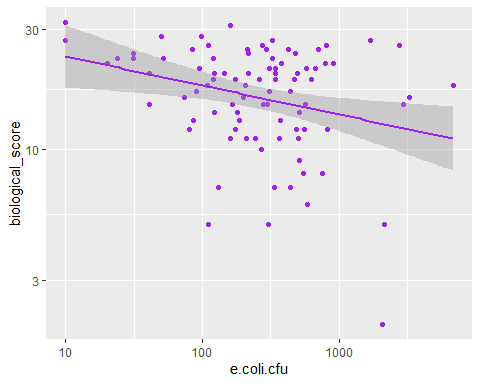
#### With *m = 0.11* and *p = 0.27*, there is a *positive*, *non-significant* correlation between *biological score* and *no3*.

#### Biological Vs E. Coli CFU

*This plot shows the relationship between biological score as a function of E. Coli (cfu).*

BS\_ECFU <- MIDO %>%  
 ggplot(aes(x=e.coli.cfu, y=biological\_score)) +  
 geom\_point(color = "purple") +  
 geom\_smooth(method=lm, color = "purple") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(BS\_ECFU)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for BS\_ECFU.

BS\_ECFU\_lm <- lm(biological\_score ~ e.coli.cfu, data=MIDO)  
  
summary(BS\_ECFU\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ e.coli.cfu, data = MIDO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -14.686 -5.160 1.055 4.932 13.914   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.0924039 0.8243184 21.948 <2e-16 \*\*\*  
## e.coli.cfu -0.0006778 0.0007846 -0.864 0.39   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.711 on 86 degrees of freedom  
## Multiple R-squared: 0.008603, Adjusted R-squared: -0.002924   
## F-statistic: 0.7463 on 1 and 86 DF, p-value: 0.39

#### With *m = -0.0007* and *p = 0.39*, there really isn’t a relationship at all between *biological score* and *ECFU*.

#### Now let’s use the gridExtra package to view all the MIDO Biological Indicator outcome relationship graphs together.

#Creates a plot with all the MIDO biological score relationships in it.  
MIDO\_bs\_lm <- grid.arrange(BS\_Con, BS\_ECFU, BS\_no3, BS\_Turb, ncol = 2)

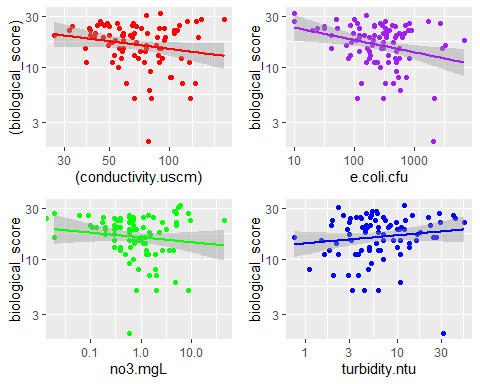
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'

## Warning: Transformation introduced infinite values in continuous x-axis  
  
## Warning: Transformation introduced infinite values in continuous x-axis

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat\_smooth).

## `geom\_smooth()` using formula 'y ~ x'

 *Save MIDO Biological Score figures*

#Stores the location I want to save the graphs to.  
MIDO\_bs\_lm\_location <- here::here("results","MIDO\_bs\_lm.png")  
#Saves graphs to stored location in ,my working directory.  
ggsave(MIDO\_bs\_lm, file = MIDO\_bs\_lm\_location)

## Saving 5 x 4 in image

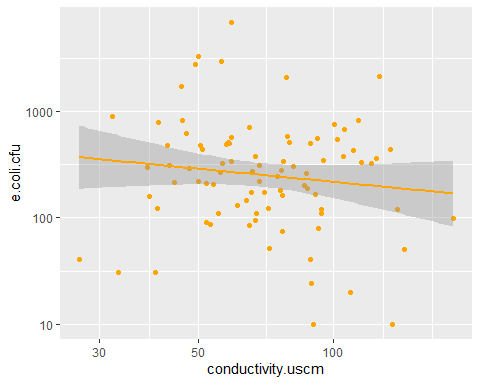
### *E. Coli*

#### E. Coli Vs Conductivity

*This plot shows the relationship between E. coli (cfu) as a function of Conductivity.*

ECFU\_Con <- MIDO %>%  
 ggplot(aes(x=conductivity.uscm, y=e.coli.cfu)) +  
 geom\_point(color = "orange") +  
 geom\_smooth(method=lm, color = "orange") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(ECFU\_Con)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for ECFU\_Con.

ECFU\_Con\_lm <- lm(e.coli.cfu ~ conductivity.uscm, data=MIDO)  
  
summary(ECFU\_Con\_lm)

##   
## Call:  
## lm(formula = e.coli.cfu ~ conductivity.uscm, data = MIDO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -654.2 -392.3 -236.8 3.7 6288.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 793.308 266.825 2.973 0.00382 \*\*  
## conductivity.uscm -3.635 3.326 -1.093 0.27750   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 916 on 86 degrees of freedom  
## Multiple R-squared: 0.0137, Adjusted R-squared: 0.002229   
## F-statistic: 1.194 on 1 and 86 DF, p-value: 0.2775

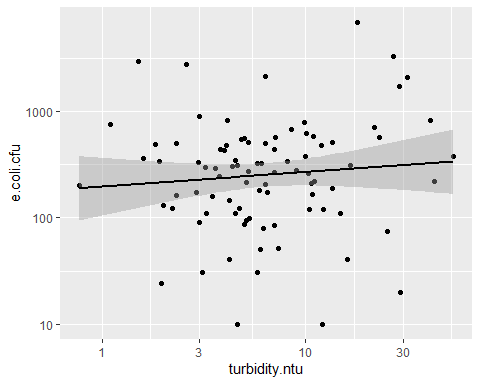
#### With *m = -3.635* and *p = 0.28*, there is a *negative*, *non-significant* correlation between *ECFU* and *Conductivity*.

#### E. Coli Vs Turbidity

*This plot shows the relationship between E. coli as a function of Turbidity.*

ECFU\_Turb <- MIDO %>%  
 ggplot(aes(x=turbidity.ntu, y=e.coli.cfu)) +  
 geom\_point(color = "black") +  
 geom\_smooth(method=lm, color = "black") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(ECFU\_Turb)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for ECFU\_Turb.

ECFU\_Turb\_lm <- lm(e.coli.cfu ~ turbidity.ntu, data=MIDO)  
  
summary(ECFU\_Turb\_lm)

##   
## Call:  
## lm(formula = e.coli.cfu ~ turbidity.ntu, data = MIDO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -961.2 -349.9 -206.9 22.6 6184.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 349.85 133.90 2.613 0.0106 \*  
## turbidity.ntu 18.60 10.04 1.852 0.0675 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 904.5 on 86 degrees of freedom  
## Multiple R-squared: 0.03835, Adjusted R-squared: 0.02717   
## F-statistic: 3.43 on 1 and 86 DF, p-value: 0.06746

#### With *m = 18.6* and *p = 0.07*, there is a *positive*, *non-significant* correlation between *ECFU* and *turbidity*.

#### E. Coli Vs Nitrate Concentration

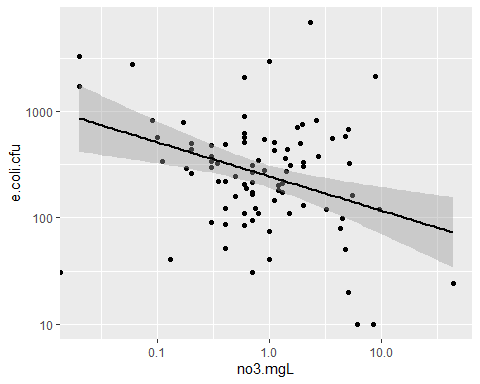
*This plot shows the relationship between E. coli (cfu) as a function of nitrate (no3) concentration.*

ECFU\_no3 <- MIDO %>%  
 ggplot(aes(x=no3.mgL, y=e.coli.cfu)) +  
 geom\_point(color = "black") +  
 geom\_smooth(method=lm, color = "black") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(ECFU\_no3)

## Warning: Transformation introduced infinite values in continuous x-axis  
  
## Warning: Transformation introduced infinite values in continuous x-axis

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat\_smooth).



#### Now run linear regression for ECFU\_no3.

ECFU\_no3\_lm <- lm(e.coli.cfu ~ no3.mgL, data=MIDO)  
  
summary(ECFU\_no3\_lm)

##   
## Call:  
## lm(formula = e.coli.cfu ~ no3.mgL, data = MIDO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -512.8 -390.1 -245.3 -26.1 6347.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 543.82 106.46 5.108 1.94e-06 \*\*\*  
## no3.mgL -10.61 19.97 -0.531 0.597   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 920.9 on 86 degrees of freedom  
## Multiple R-squared: 0.00327, Adjusted R-squared: -0.00832   
## F-statistic: 0.2821 on 1 and 86 DF, p-value: 0.5967

#### With *m = -10.61* and *p = 0.60*, there is a *negative*, *non-significant* correlation between *ECFU* and *no3*.

#### Now let’s use the gridExtra package to view all the MIDO E. coli CFU outcome relationship graphs together.

MIDO\_ECFU\_lm <- grid.arrange(ECFU\_Con, ECFU\_no3, ECFU\_Turb, ncol = 2)

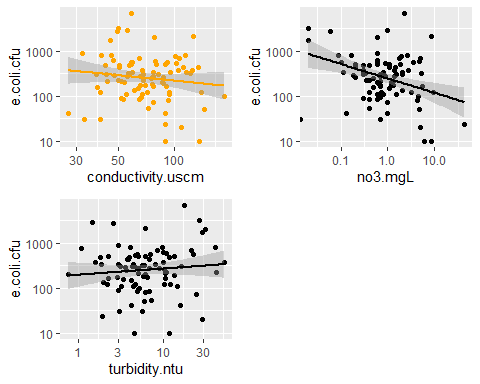
## `geom\_smooth()` using formula 'y ~ x'

## Warning: Transformation introduced infinite values in continuous x-axis  
  
## Warning: Transformation introduced infinite values in continuous x-axis

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat\_smooth).

## `geom\_smooth()` using formula 'y ~ x'

 *Save MIDO ECFU figures*

#Stores the location I want to save the graphs to.  
MIDO\_ECFU\_lm\_location <- here::here("results","MIDO\_ECFU\_lm.png")  
#Saves graphs to stored location in ,my working directory.  
ggsave(MIDO\_ECFU\_lm, file = MIDO\_ECFU\_lm\_location)

## Saving 5 x 4 in image

## NORO *the following linear analysis deals exclusively with the NORO dataset.*

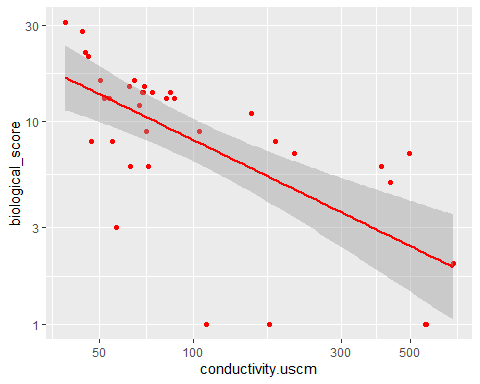
### *Biological Indicator:*

#### Biological Vs Conductivity

*This plot shows the relationship between Biological Score as a function of Conductivity.*

N\_BS\_Con <- NORO %>%  
 ggplot(aes(x=conductivity.uscm, y=biological\_score)) +  
 geom\_point(color = "red") +  
 geom\_smooth(method=lm, color = "red") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(N\_BS\_Con)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for N\_BS\_Con.

N\_BS\_Con\_lm <- lm(biological\_score ~ conductivity.uscm, data=NORO)  
  
summary(N\_BS\_Con\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ conductivity.uscm, data = NORO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.014 -2.907 0.296 1.978 17.353   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 14.541261 1.372711 10.593 3.75e-12 \*\*\*  
## conductivity.uscm -0.022931 0.005843 -3.924 0.000417 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.044 on 33 degrees of freedom  
## Multiple R-squared: 0.3182, Adjusted R-squared: 0.2975   
## F-statistic: 15.4 on 1 and 33 DF, p-value: 0.0004167

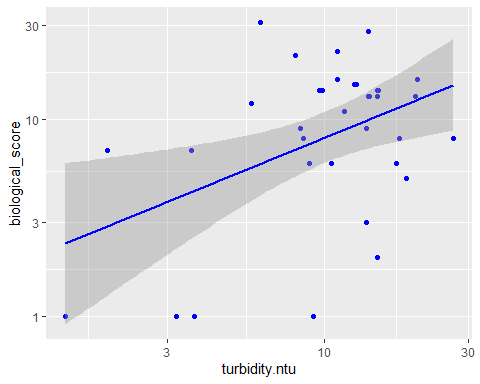
#### With *m = -0.023* and *p = 0.0004*, there is a *negative*, *significant* correlation between *biological score* and *conductivity*.

#### Biological Vs Turbidity

*This plot shows the relationship between Biological Score as a function of Turbidity.*

N\_BS\_Turb <- NORO %>%  
 ggplot(aes(x=turbidity.ntu, y=biological\_score)) +  
 geom\_point(color = "blue") +  
 geom\_smooth(method=lm, color = "blue") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(N\_BS\_Turb)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for N\_BS\_Turb.

N\_BS\_Turb\_lm <- lm(biological\_score ~ turbidity.ntu, data=NORO)  
  
summary(N\_BS\_Turb\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ turbidity.ntu, data = NORO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.5326 -5.1174 0.0636 3.3833 20.9890   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.9848 2.8434 3.160 0.00337 \*\*  
## turbidity.ntu 0.1682 0.2204 0.763 0.45071   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.256 on 33 degrees of freedom  
## Multiple R-squared: 0.01735, Adjusted R-squared: -0.01243   
## F-statistic: 0.5826 on 1 and 33 DF, p-value: 0.4507

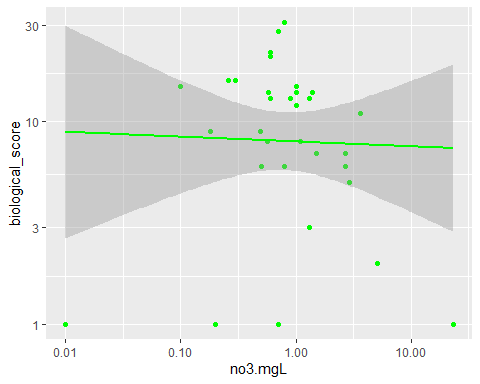
#### With *m = 0.17* and *p = 0.45*, there is a *positive*, *non-significant* correlation between *bilogical score* and *turbidity*.

#### Biological Vs Nitrate Concentration

*This plot shows the relationship between Biological Score as a function of nitrate (no3) Concentration.*

N\_BS\_no3 <- NORO %>%  
 ggplot(aes(x=no3.mgL, y=biological\_score)) +  
 geom\_point(color = "green") +  
 geom\_smooth(method=lm, color = "green") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(N\_BS\_no3)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for N\_BS\_no3.

N\_BS\_no3\_lm <- lm(biological\_score ~ no3.mgL, data=NORO)  
  
summary(N\_BS\_no3\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ no3.mgL, data = NORO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.931 -4.254 1.091 2.736 19.514   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.9371 1.2970 9.203 1.24e-10 \*\*\*  
## no3.mgL -0.5632 0.3069 -1.835 0.0755 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.972 on 33 degrees of freedom  
## Multiple R-squared: 0.09262, Adjusted R-squared: 0.06512   
## F-statistic: 3.368 on 1 and 33 DF, p-value: 0.07548

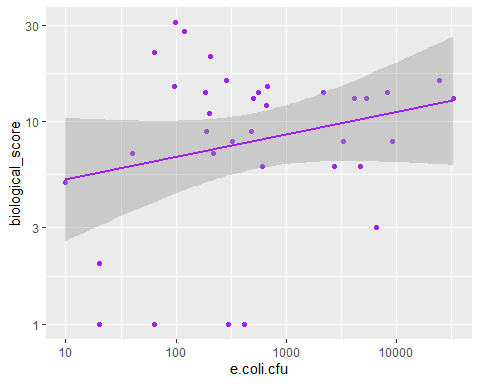
#### With *m = -0.56* and *p = 0.08*, there is a *negative*, *non-significant* correlation between *biological score* and *no3*.

#### Biological Vs E. Coli CFU

*This plot shows the relationship between Biological Score as a function of E. coli (cfu).*

N\_BS\_ECFU <- NORO %>%  
 ggplot(aes(x=e.coli.cfu, y=biological\_score)) +  
 geom\_point(color = "purple") +  
 geom\_smooth(method=lm, color = "purple") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(N\_BS\_ECFU)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for N\_BS\_ECFU.

N\_BS\_ECFU\_lm <- lm(biological\_score ~ e.coli.cfu, data=NORO)  
  
summary(N\_BS\_ECFU\_lm)

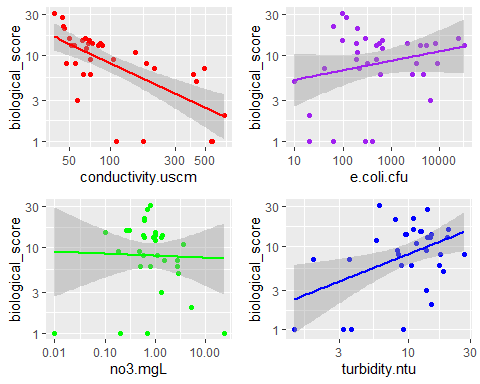
##   
## Call:  
## lm(formula = biological\_score ~ e.coli.cfu, data = NORO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.7460 -4.8366 -0.0854 3.2572 20.2770   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.072e+01 1.360e+00 7.880 4.39e-09 \*\*\*  
## e.coli.cfu 7.274e-05 1.830e-04 0.397 0.694   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.302 on 33 degrees of freedom  
## Multiple R-squared: 0.004765, Adjusted R-squared: -0.02539   
## F-statistic: 0.158 on 1 and 33 DF, p-value: 0.6936

#### With *m = 7.274e-05* and *p = 0.69*, there is not relationship between *Bbiological score* and *ECFU*.

#### Now let’s use the gridExtra package to view all the NORO Biological Indicator outcome relationship graphs together.

NORO\_bs\_lm <- grid.arrange(N\_BS\_Con, N\_BS\_ECFU, N\_BS\_no3, N\_BS\_Turb, ncol = 2)

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'

 *Save NORO Biological Score figures*

#Stores the location I want to save the graphs to.  
NORO\_bs\_lm\_location <- here::here("results","NORO\_bs\_lm.png")  
#Saves graphs to stored location in ,my working directory.  
ggsave(NORO\_bs\_lm, file = NORO\_bs\_lm\_location)

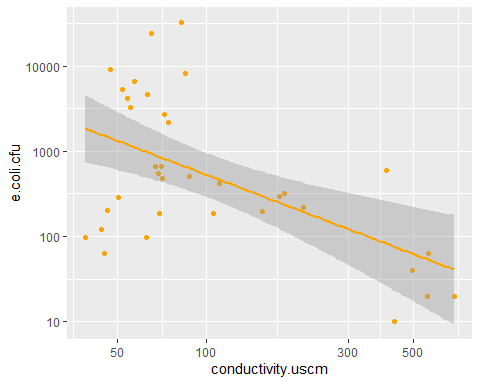
## Saving 5 x 4 in image

### *E. Coli*

#### E. Coli Vs Conductivity *This plot shows the relationship between E. Coli (cfu) as a function of Conductivity.*

N\_ECFU\_Con <- NORO %>%  
 ggplot(aes(x=conductivity.uscm, y=e.coli.cfu)) +  
 geom\_point(color = "orange") +  
 geom\_smooth(method=lm, color = "orange") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(N\_ECFU\_Con)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for N\_ECFU\_Con.

N\_ECFU\_Con\_lm <- lm(e.coli.cfu ~ conductivity.uscm, data=NORO)  
  
summary(N\_ECFU\_Con\_lm)

##   
## Call:  
## lm(formula = e.coli.cfu ~ conductivity.uscm, data = NORO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4040.6 -3353.7 -2425.3 382.9 28807.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4475.297 1537.659 2.910 0.00642 \*\*  
## conductivity.uscm -8.632 6.545 -1.319 0.19630   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6770 on 33 degrees of freedom  
## Multiple R-squared: 0.05007, Adjusted R-squared: 0.02128   
## F-statistic: 1.739 on 1 and 33 DF, p-value: 0.1963

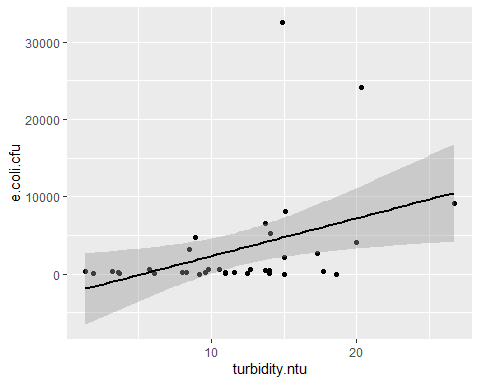
#### With *m = -8.632* and *p = 0.20*, there is a *negative*, *non-significant* correlation between *ECFU* and *conductivity*.

#### E. Coli Vs Turbidity

*This plot shows the relationship between E. Coli (cfu) as a function of Turbidity.*

N\_ECFU\_Turb <- NORO %>%  
 ggplot(aes(x=turbidity.ntu, y=e.coli.cfu)) +  
 geom\_point(color = "black") +  
 geom\_smooth(method=lm, color = "black")  
  
  
print(N\_ECFU\_Turb)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for N\_ECFU\_Turb.

N\_ECFU\_Turb\_lm <- lm(e.coli.cfu ~ turbidity.ntu, data=NORO)  
  
summary(N\_ECFU\_Turb\_lm)

##   
## Call:  
## lm(formula = e.coli.cfu ~ turbidity.ntu, data = NORO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6522 -3038 -1665 1171 27856   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2582 2490 -1.037 0.307   
## turbidity.ntu 490 193 2.539 0.016 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6353 on 33 degrees of freedom  
## Multiple R-squared: 0.1634, Adjusted R-squared: 0.138   
## F-statistic: 6.444 on 1 and 33 DF, p-value: 0.01603

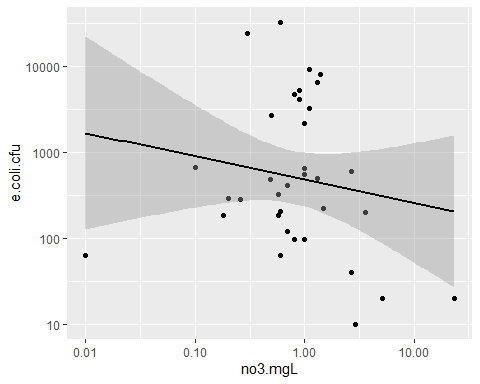
#### With *m = 490* and *p = 0.02*, there is a *positive*, *significant* correlation between *ECFU* and *Turbidity*.

#### E. Coli Vs Nitrate Concentration

*This plot shows the relationship between E. coli as a function of nitrate (no3) concentration.*

N\_ECFU\_no3 <- NORO %>%  
 ggplot(aes(x=no3.mgL, y=e.coli.cfu)) +  
 geom\_point(color = "black") +  
 geom\_smooth(method=lm, color = "black") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(N\_ECFU\_no3)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for N\_ECFU\_no3.

N\_ECFU\_no3\_lm <- lm(e.coli.cfu ~ no3.mgL, data=NORO)  
  
summary(N\_ECFU\_no3\_lm)

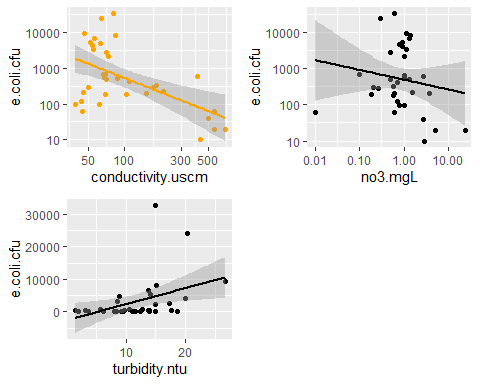
##   
## Call:  
## lm(formula = e.coli.cfu ~ no3.mgL, data = NORO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3431.7 -3157.8 -2728.8 399.7 29206.0   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3496.9 1282.6 2.726 0.0102 \*  
## no3.mgL -213.1 303.5 -0.702 0.4875   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6895 on 33 degrees of freedom  
## Multiple R-squared: 0.01472, Adjusted R-squared: -0.01514   
## F-statistic: 0.493 on 1 and 33 DF, p-value: 0.4875

#### With *m = -213.1* and *p = 0.49*, there is a *negative*, *non-significant* correlation between *ECFU* and *no3*.

#### Now let’s use the gridExtra package to view all the NORO E. coli CFU outcome relationship graphs together.

NORO\_ECFU\_lm <- grid.arrange(N\_ECFU\_Con, N\_ECFU\_no3, N\_ECFU\_Turb, ncol = 2)

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'

 *Save NORO ECFU figures*

#Stores the location I want to save the graphs to.  
NORO\_ECFU\_lm\_location <- here::here("results","NORO\_ECFU\_lm.png")  
#Saves graphs to stored location in ,my working directory.  
ggsave(NORO\_ECFU\_lm, file = NORO\_ECFU\_lm\_location)

## Saving 5 x 4 in image

## BICO:

*The following Linear Regression analysis deals exclusively with the BICO data set.*

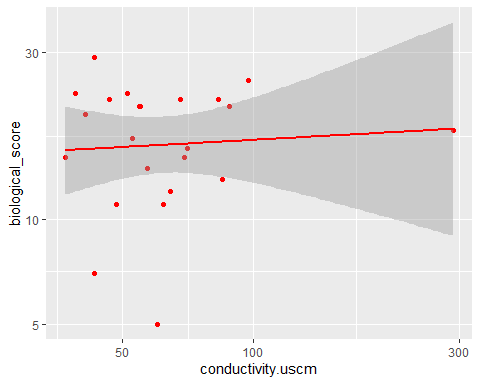
### *Biological Indicator:*

#### Biological Vs Conductivity

*This plot shows the relationship between biological score as a function of conductivity.*

B\_BS\_Con <- BICO %>%  
 ggplot(aes(x=conductivity.uscm, y=biological\_score)) +  
 geom\_point(color = "red") +  
 geom\_smooth(method=lm, color = "red") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(B\_BS\_Con)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for B\_BS\_Con.

B\_BS\_Con\_lm <- lm(biological\_score ~ conductivity.uscm, data=BICO)  
  
summary(B\_BS\_Con\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ conductivity.uscm, data = BICO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.4765 -4.0273 -0.4428 4.4520 11.6024   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.199651 2.192576 7.844 1.13e-07 \*\*\*  
## conductivity.uscm 0.004614 0.025592 0.180 0.859   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.097 on 21 degrees of freedom  
## Multiple R-squared: 0.001546, Adjusted R-squared: -0.046   
## F-statistic: 0.03251 on 1 and 21 DF, p-value: 0.8586

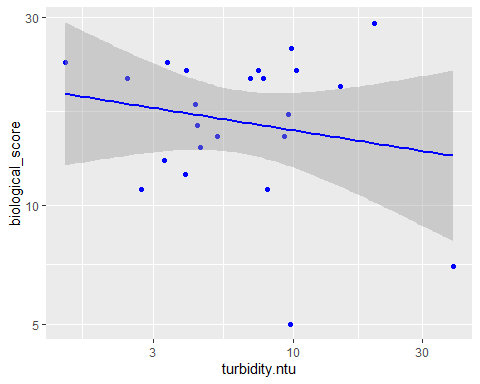
#### With *m = 0.0046* and *p = 0.86*, there is a *positive*, *non-significant* correlation between *biological score* and *conductivity*.

#### Biological Vs Turbidity

*This plot shows the relationship between biological score as a function of turbidity.*

B\_BS\_Turb <- BICO %>%  
 ggplot(aes(x=turbidity.ntu, y=biological\_score)) +  
 geom\_point(color = "blue") +  
 geom\_smooth(method=lm, color = "blue") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(B\_BS\_Turb)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for B\_BS\_Turb.

B\_BS\_Turb\_lm <- lm(biological\_score ~ turbidity.ntu, data=BICO)  
  
summary(B\_BS\_Turb\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ turbidity.ntu, data = BICO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.3528 -4.5901 -0.0402 4.1399 12.9672   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.5884 1.8365 10.122 1.57e-09 \*\*\*  
## turbidity.ntu -0.1278 0.1608 -0.795 0.436   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.012 on 21 degrees of freedom  
## Multiple R-squared: 0.0292, Adjusted R-squared: -0.01703   
## F-statistic: 0.6316 on 1 and 21 DF, p-value: 0.4357

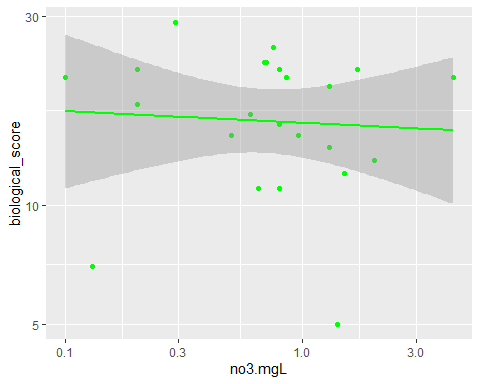
#### With *m = -0.13* and *p = 0.44*, there is a *negative*, *non-significant* correlation between *biological score* and *turbidity*.

#### Biological Vs Nitrate Concentration

*This plot shows the relationship between biological score as a function of nitrate (no3) concentration.*

B\_BS\_no3 <- BICO %>%  
 ggplot(aes(x=no3.mgL, y=biological\_score)) +  
 geom\_point(color = "green") +  
 geom\_smooth(method=lm, color = "green") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(B\_BS\_no3)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for B\_BS\_no3.

B\_BS\_no3\_lm <- lm(biological\_score ~ no3.mgL, data=BICO)  
  
summary(B\_BS\_no3\_lm)

##   
## Call:  
## lm(formula = biological\_score ~ no3.mgL, data = BICO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.3950 -3.8197 0.2433 4.4531 11.2704   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.8170 1.9172 9.293 6.88e-09 \*\*\*  
## no3.mgL -0.3014 1.4653 -0.206 0.839   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.095 on 21 degrees of freedom  
## Multiple R-squared: 0.00201, Adjusted R-squared: -0.04551   
## F-statistic: 0.0423 on 1 and 21 DF, p-value: 0.839

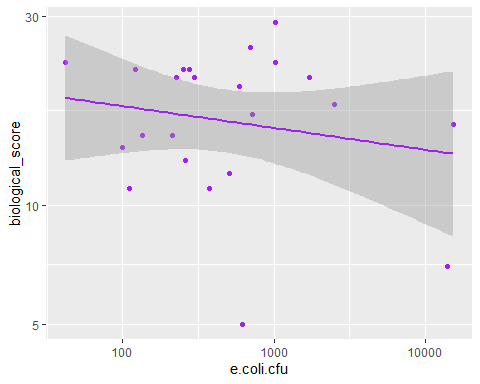
#### With *m = -0.30* and *p = 0.84*, there is a *negative*, *non-significant* correlation between *biological score* and *no3*.

#### Biological Vs E. Coli CFU

*This plot shows the relationship between biological score as a function of E. coli (cfu).*

B\_BS\_ECFU <- BICO %>%  
 ggplot(aes(x=e.coli.cfu, y=biological\_score)) +  
 geom\_point(color = "purple") +  
 geom\_smooth(method=lm, color = "purple") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(B\_BS\_ECFU)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for B\_BS\_ECFU.

B\_BS\_ECFU\_lm <- lm(biological\_score ~ e.coli.cfu, data=BICO)  
  
summary(B\_BS\_ECFU\_lm)

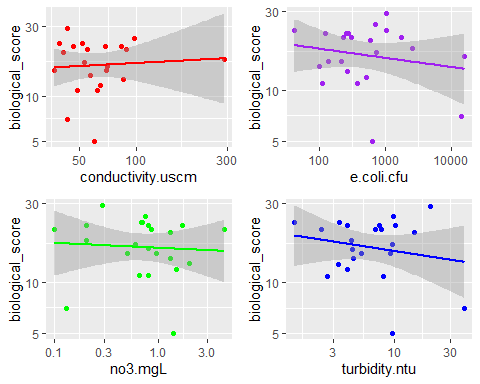
##   
## Call:  
## lm(formula = biological\_score ~ e.coli.cfu, data = BICO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.999 -4.678 1.989 3.855 11.162   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.2514270 1.3359266 13.662 6.43e-12 \*\*\*  
## e.coli.cfu -0.0004064 0.0003022 -1.345 0.193   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.854 on 21 degrees of freedom  
## Multiple R-squared: 0.07927, Adjusted R-squared: 0.03543   
## F-statistic: 1.808 on 1 and 21 DF, p-value: 0.1931

#### With *m = -0.00041* and *p = 0.19*, there is no relationship between *biological score* and *ECFU*.

#### Now let’s use the gridExtra package to view all the BICO Biological Indicator outcome relationship graphs together.

BICO\_bs\_lm <- grid.arrange(B\_BS\_Con, B\_BS\_ECFU, B\_BS\_no3, B\_BS\_Turb, ncol = 2)

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'

 *Save BICO biological score figures*

#Stores the location I want to save the graphs to.  
BICO\_bs\_lm\_location <- here::here("results","BICO\_bs\_lm.png")  
#Saves graphs to stored location in ,my working directory.  
ggsave(BICO\_bs\_lm, file = BICO\_bs\_lm\_location)

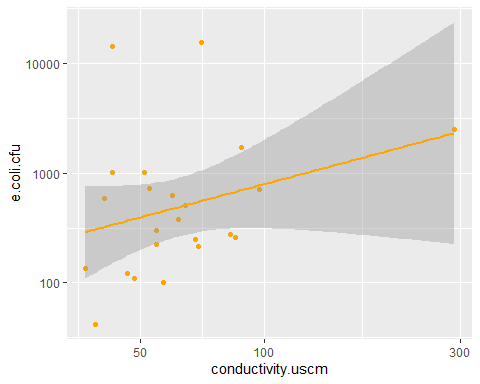
## Saving 5 x 4 in image

### *E. Coli*

#### E. Coli Vs Conductivity *This plot shows the relationship between E. coli (cfu) as a function of conductivity.*

B\_ECFU\_Con <- BICO %>%  
 ggplot(aes(x=conductivity.uscm, y=e.coli.cfu)) +  
 geom\_point(color = "orange") +  
 geom\_smooth(method=lm, color = "orange") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(B\_ECFU\_Con)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for B\_ECFU\_Con.

B\_ECFU\_Con\_lm <- lm(e.coli.cfu ~ conductivity.uscm, data=BICO)  
  
summary(B\_ECFU\_Con\_lm)

##   
## Call:  
## lm(formula = e.coli.cfu ~ conductivity.uscm, data = BICO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1683.1 -1575.6 -1402.7 -885.4 13553.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1636.516 1519.584 1.077 0.294  
## conductivity.uscm 2.278 17.737 0.128 0.899  
##   
## Residual standard error: 4225 on 21 degrees of freedom  
## Multiple R-squared: 0.0007846, Adjusted R-squared: -0.0468   
## F-statistic: 0.01649 on 1 and 21 DF, p-value: 0.899

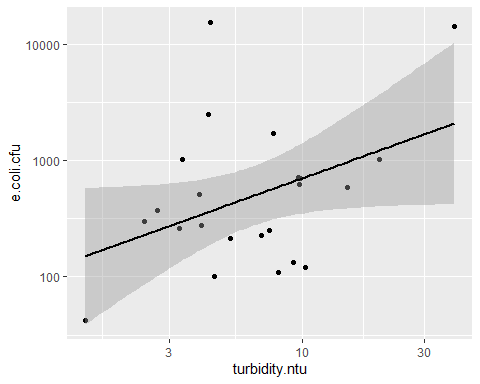
#### With *m = 2.28* and *p = 0.90*, there is a *positive*, *non-significant* correlation between *ECFU* and *conductivity*.

#### E. Coli Vs Turbidity

*This plot shows the relationship between E. coli (cfu) as a function of turbidity.*

B\_ECFU\_Turb <- BICO %>%  
 ggplot(aes(x=turbidity.ntu, y=e.coli.cfu)) +  
 geom\_point(color = "black") +  
 geom\_smooth(method=lm, color = "black") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(B\_ECFU\_Turb)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for B\_ECFU\_Turb.

B\_ECFU\_Turb\_lm <- lm(e.coli.cfu ~ turbidity.ntu, data=BICO)  
  
summary(B\_ECFU\_Turb\_lm)

##   
## Call:  
## lm(formula = e.coli.cfu ~ turbidity.ntu, data = BICO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3825.2 -1496.7 -689.4 58.8 14594.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -387.2 1114.8 -0.347 0.732   
## turbidity.ntu 261.5 97.6 2.679 0.014 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3649 on 21 degrees of freedom  
## Multiple R-squared: 0.2547, Adjusted R-squared: 0.2192   
## F-statistic: 7.177 on 1 and 21 DF, p-value: 0.01405

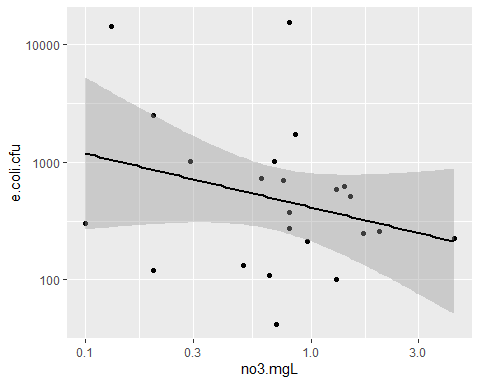
#### With *m = 261.5* and *p = 0.01*, there is a *positive*, *significant* correlation between *ECFU* and *turbidity*.

#### E. Coli Vs Nitrate Concentration

*This plot shows the relationship between E. coli (cfu) as a function of nitrate (no3) concentration.*

B\_ECFU\_no3 <- BICO %>%  
 ggplot(aes(x=no3.mgL, y=e.coli.cfu)) +  
 geom\_point(color = "black") +  
 geom\_smooth(method=lm, color = "black") +  
 scale\_x\_continuous(trans = 'log10') +  
 scale\_y\_continuous(trans = 'log10')  
  
  
print(B\_ECFU\_no3)

## `geom\_smooth()` using formula 'y ~ x'



#### Now run linear regression for B\_ECFU\_no3.

B\_ECFU\_no3\_lm <- lm(e.coli.cfu ~ no3.mgL, data=BICO)  
  
summary(B\_ECFU\_no3\_lm)

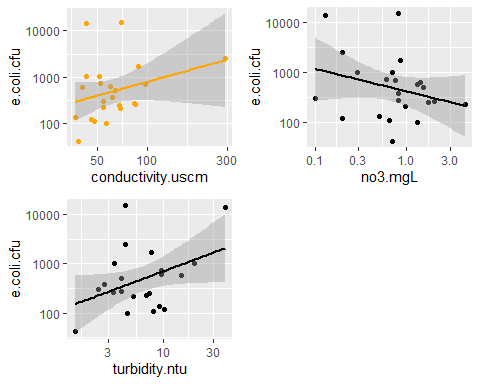
##   
## Call:  
## lm(formula = e.coli.cfu ~ no3.mgL, data = BICO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2441.3 -1649.4 -1318.3 -646.8 13377.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2759.0 1299.5 2.123 0.0458 \*  
## no3.mgL -983.6 993.2 -0.990 0.3333   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4132 on 21 degrees of freedom  
## Multiple R-squared: 0.04462, Adjusted R-squared: -0.0008751   
## F-statistic: 0.9808 on 1 and 21 DF, p-value: 0.3333

#### With *m = -983.6* and *p = 0.33*, there is a *negative*, *non-significant* correlation between *ECFU* and *no3*.

#### Now let’s use the gridExtra package to view all the BICO E. coli CFU outcome relationship graphs together.

BICO\_ECFU\_lm <- grid.arrange(B\_ECFU\_Con, B\_ECFU\_no3, B\_ECFU\_Turb, ncol = 2)

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'

 *Save BICO ECFU figures*

#Stores the location I want to save the graphs to.  
BICO\_ECFU\_lm\_location <- here::here("results","BICO\_ECFU\_lm.png")  
#Saves graphs to stored location in ,my working directory.  
ggsave(BICO\_ECFU\_lm, file = BICO\_ECFU\_lm\_location)

## Saving 5 x 4 in image

### Save the exploratory linear models #### MIDO *Biological Score*

BS\_Con\_location <- here::here("data","processed\_data","BS\_Con.rds")  
saveRDS(BS\_Con\_lm, file = BS\_Con\_location)  
  
BS\_ECFU\_location <- here::here("data","processed\_data","BS\_ECFU.rds")  
saveRDS(BS\_ECFU\_lm, file = BS\_ECFU\_location)  
  
BS\_no3\_location <- here::here("data","processed\_data","BS\_no3.rds")  
saveRDS(BS\_no3\_lm, file = BS\_no3\_location)  
  
BS\_Turb\_location <- here::here("data","processed\_data","BS\_Turb.rds")  
saveRDS(BS\_Turb\_lm, file = BS\_Turb\_location)

*ECFU*

ECFU\_Con\_location <- here::here("data","processed\_data","BS\_Con.rds")  
saveRDS(ECFU\_Con\_lm, file = ECFU\_Con\_location)  
  
ECFU\_no3\_location <- here::here("data","processed\_data","BS\_no3.rds")  
saveRDS(ECFU\_no3\_lm, file = ECFU\_no3\_location)  
  
ECFU\_Turb\_location <- here::here("data","processed\_data","BS\_Turb.rds")  
saveRDS(ECFU\_Turb\_lm, file = ECFU\_Turb\_location)

#### BICO

*Biological Score*

B\_BS\_Con\_location <- here::here("data","processed\_data","B\_BS\_Con.rds")  
saveRDS(B\_BS\_Con\_lm, file = B\_BS\_Con\_location)  
  
B\_BS\_ECFU\_location <- here::here("data","processed\_data","B\_BS\_ECFU.rds")  
saveRDS(B\_BS\_ECFU\_lm, file = B\_BS\_ECFU\_location)  
  
B\_BS\_no3\_location <- here::here("data","processed\_data","B\_BS\_no3.rds")  
saveRDS(B\_BS\_no3\_lm, file = B\_BS\_no3\_location)  
  
B\_BS\_Turb\_location <- here::here("data","processed\_data","B\_BS\_Turb.rds")  
saveRDS(B\_BS\_Turb\_lm, file = B\_BS\_Turb\_location)

*ECFU*

B\_ECFU\_Con\_location <- here::here("data","processed\_data","B\_BS\_Con.rds")  
saveRDS(B\_ECFU\_Con\_lm, file = B\_ECFU\_Con\_location)  
  
B\_ECFU\_no3\_location <- here::here("data","processed\_data","B\_BS\_no3.rds")  
saveRDS(B\_ECFU\_no3\_lm, file = B\_ECFU\_no3\_location)  
  
B\_ECFU\_Turb\_location <- here::here("data","processed\_data","B\_BS\_Turb.rds")  
saveRDS(B\_ECFU\_Turb\_lm, file = B\_ECFU\_Turb\_location)

#### NORO

*Biological Score*

N\_BS\_Con\_location <- here::here("data","processed\_data","N\_BS\_Con.rds")  
saveRDS(N\_BS\_Con\_lm, file = N\_BS\_Con\_location)  
  
N\_BS\_ECFU\_location <- here::here("data","processed\_data","N\_BS\_ECFU.rds")  
saveRDS(N\_BS\_ECFU\_lm, file = N\_BS\_ECFU\_location)  
  
N\_BS\_no3\_location <- here::here("data","processed\_data","N\_BS\_no3.rds")  
saveRDS(N\_BS\_no3\_lm, file = N\_BS\_no3\_location)  
  
N\_BS\_Turb\_location <- here::here("data","processed\_data","N\_BS\_Turb.rds")  
saveRDS(N\_BS\_Turb\_lm, file = N\_BS\_Turb\_location)

*ECFU*

N\_ECFU\_Con\_location <- here::here("data","processed\_data","N\_BS\_Con.rds")  
saveRDS(N\_ECFU\_Con\_lm, file = N\_ECFU\_Con\_location)  
  
N\_ECFU\_no3\_location <- here::here("data","processed\_data","N\_BS\_no3.rds")  
saveRDS(N\_ECFU\_no3\_lm, file = N\_ECFU\_no3\_location)  
  
N\_ECFU\_Turb\_location <- here::here("data","processed\_data","N\_BS\_Turb.rds")  
saveRDS(N\_ECFU\_Turb\_lm, file = N\_ECFU\_Turb\_location)

### In summary, exploration of correlation of stream health indicators showed that most interactions were non-significant. However, it is to be noted that the ECFU and turbidity interaction was significantly positive. In further analysis, this may serve as a check to show that correlative analyses are robust enough with data provided. In aquatic microbiology, the positive relationship between ECFU growth and turbidity is generally accepted because the more bacteria that grow in the water column, the cloudier it will be.

### Additionally, there was a weird significant correlation between biological score and conductivity at the NORO site only. The relationship is negative, meaning that biological score decreased as conductivity increases, which make sense. However, NORO was the only site to show this interaction as significant.