FinalProject_Group6

BillyMcLean

2024-10-18

Intro

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.0
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
to become errors
```

```
library(ggthemes)
library(dplyr)
library(stringr)
library(gridExtra)
```

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

```
#Read in ProfileData
G01<-read_csv("/Users/wamclean/github/LimnologyLabs/Lab4/Group01_09062024_profiledata.cs
v")
```

```
## Rows: 12 Columns: 9
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): group
## dbl (6): z_m, temp_c, ph, cond, do_percent, do_mgL
## lgl (2): sal, chla_ugL
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
G03<-read_csv("/Users/wamclean/github/LimnologyLabs/Lab4/Group03_09062024_profiledata.csv")
```

```
## Rows: 9 Columns: 9
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): group
## dbl (6): z_m, temp_c, ph, cond, do_percent, do_mgL
## lgl (2): sal, chla_ugL
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
combined_data<- rbind(G01,G03)
```

```
combined_data <- combined_data %>%
  select(where(~ any(!is.na(.)))) #remove empty columns from the dataset, in this case "sal" and "chla_ugL"

#use the select tool to chose which parameters you want in your new file
#use drop_na() last to remove any empty rows

temp<- combined_data %>%
  select(group, z_m, temp_c) %>%
  drop_na()

pH<- combined_data %>%
  select(group, z_m, ph)%>%
  drop_na()

conductivity<- combined_data %>%
  select(group, z_m, cond)%>%
  drop_na()

do_percent<- combined_data %>%
  select(group, z_m, do_percent) %>%
  drop_na()

do_mgL<- combined_data %>%
  select(group, z_m, do_mgL) %>%
  drop_na()
```

3. Results (figures and calculations) [32 points]. A presentation of the physical, chemical, and biological properties of the water column where we sampled.

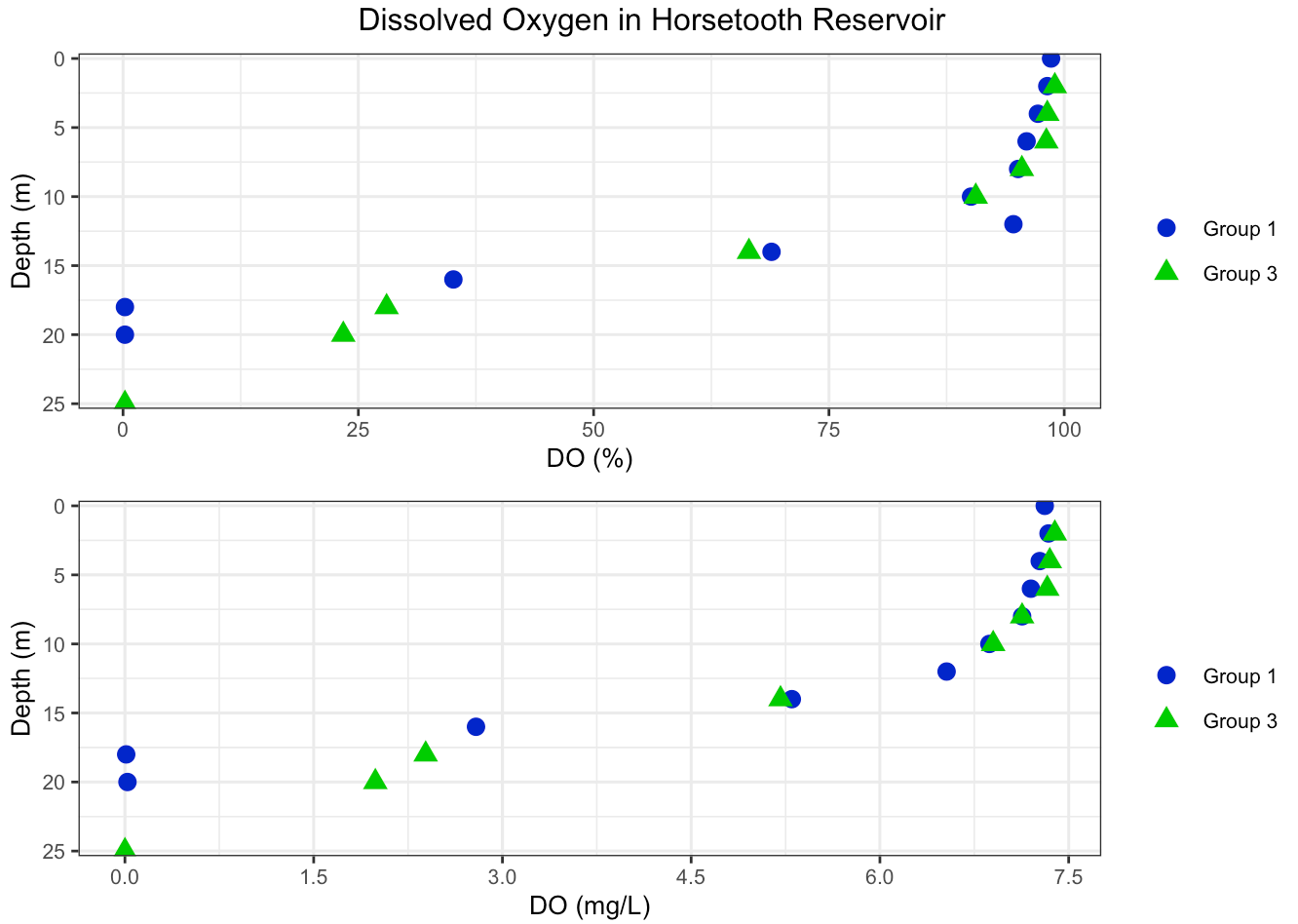
a. Each figure should have meaningful and axis labels (changed from the default label if appropriate, and with appropriate units), a legend, and a title.
b. Profiles should have the y axis inverted.
c. All figures should be indexed (i.e., Figure 1, Figure 2 etc.) and include a caption. In the remaining text of the report, when you want to refer to your figure to make a point, please use your figure numbers (like in a

scientific journal article)

d. For full credit you must include the following data. For items 1-3, use the data from whichever depth profile you uploaded to Canvas under the “Horsetooth Sampling Data Entry” assignment. For items 5-7, include class means and your group’s means, incorporated as you see fit into an appropriate figure.

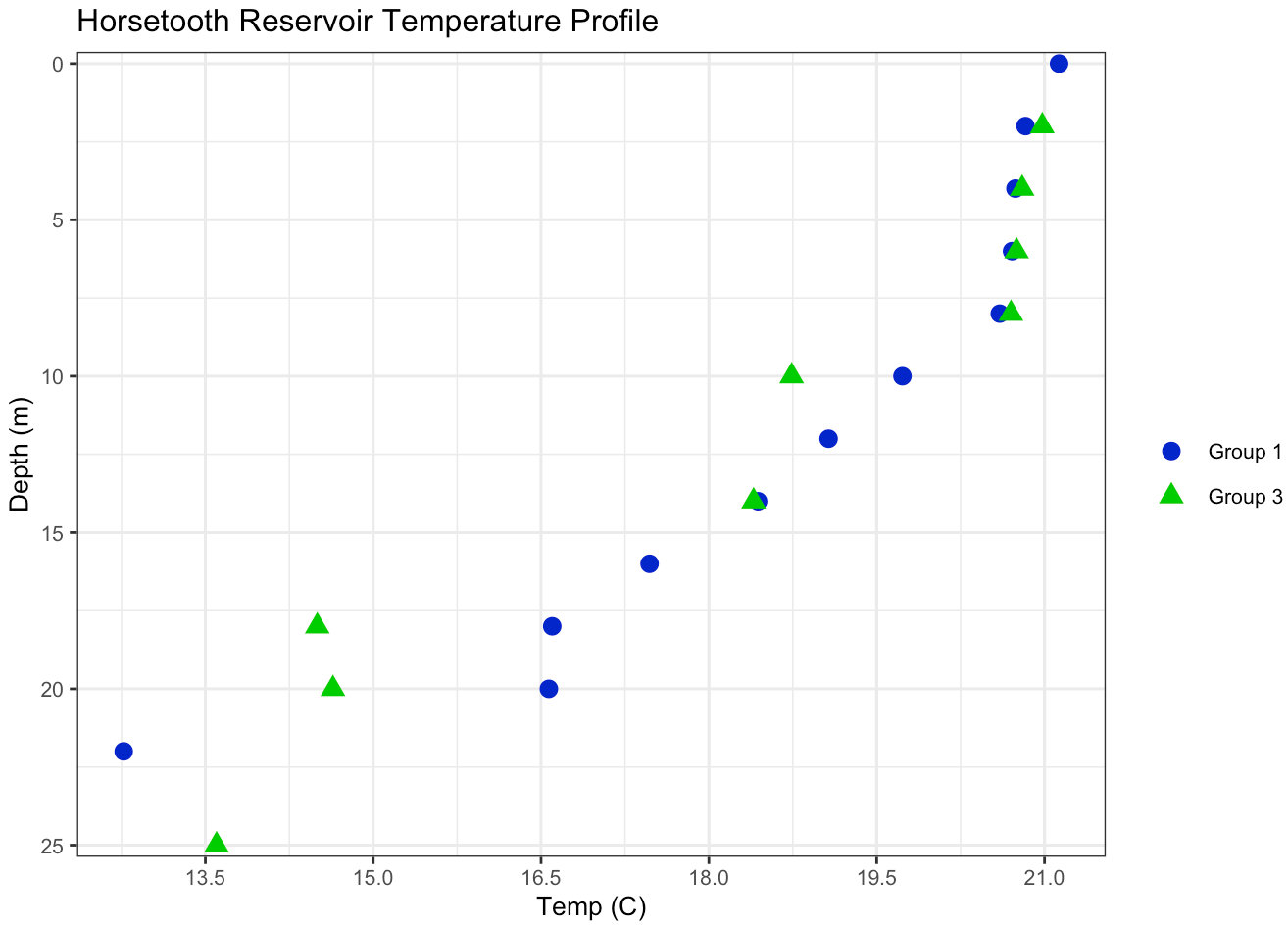
1. An oxygen profile, containing both the DO (percent) and DO (mg/L) series

```
do_percent_prof <- ggplot() +
  geom_point(data = do_percent, mapping = aes(x = do_percent,y = z_m, color = group, sha
pe = group), size = 3, show.legend = TRUE)+
  theme_bw() +
  labs( x='DO (%)',
        y= 'Depth (m)',
        color = "",
        shape = "")+
  scale_y_reverse(expand = c(0.015,0))+
  scale_color_manual(values = c("Group_1" = "blue3", "Group_3" = "green3"),  # Specify c
olors
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) + # Custo
m labels
  scale_shape_manual(values = c("Group_1" = 16, "Group_3" = 17),           # Specify sha
pes
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) +
  theme(text = element_text(size = 10))

do_mgL_prof <- ggplot() +
  geom_point(data = do_mgL, mapping = aes(x = do_mgL,y = z_m, color = group, shape = gro
up), size = 3, show.legend = TRUE)+
  theme_bw() +
  labs( x='DO (mg/L)',
        y= 'Depth (m)',
        color = "",
        shape = "")+
  scale_y_reverse(expand = c(0.015,0))+
  scale_color_manual(values = c("Group_1" = "blue3", "Group_3" = "green3"),  # Specify c
olors
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) + # Custo
m labels
  scale_shape_manual(values = c("Group_1" = 16, "Group_3" = 17),           # Specify sha
pes
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) +
  scale_x_continuous(breaks = seq(floor(min(do_mgL$do_mgL)), ceiling(max(do_mgL$do_mg
L)), by = 1.5)) +
  theme(text = element_text(size = 10))

grid.arrange(do_percent_prof, do_mgL_prof, top = "Dissolved Oxygen in Horsetooth Reservo
ir")
```



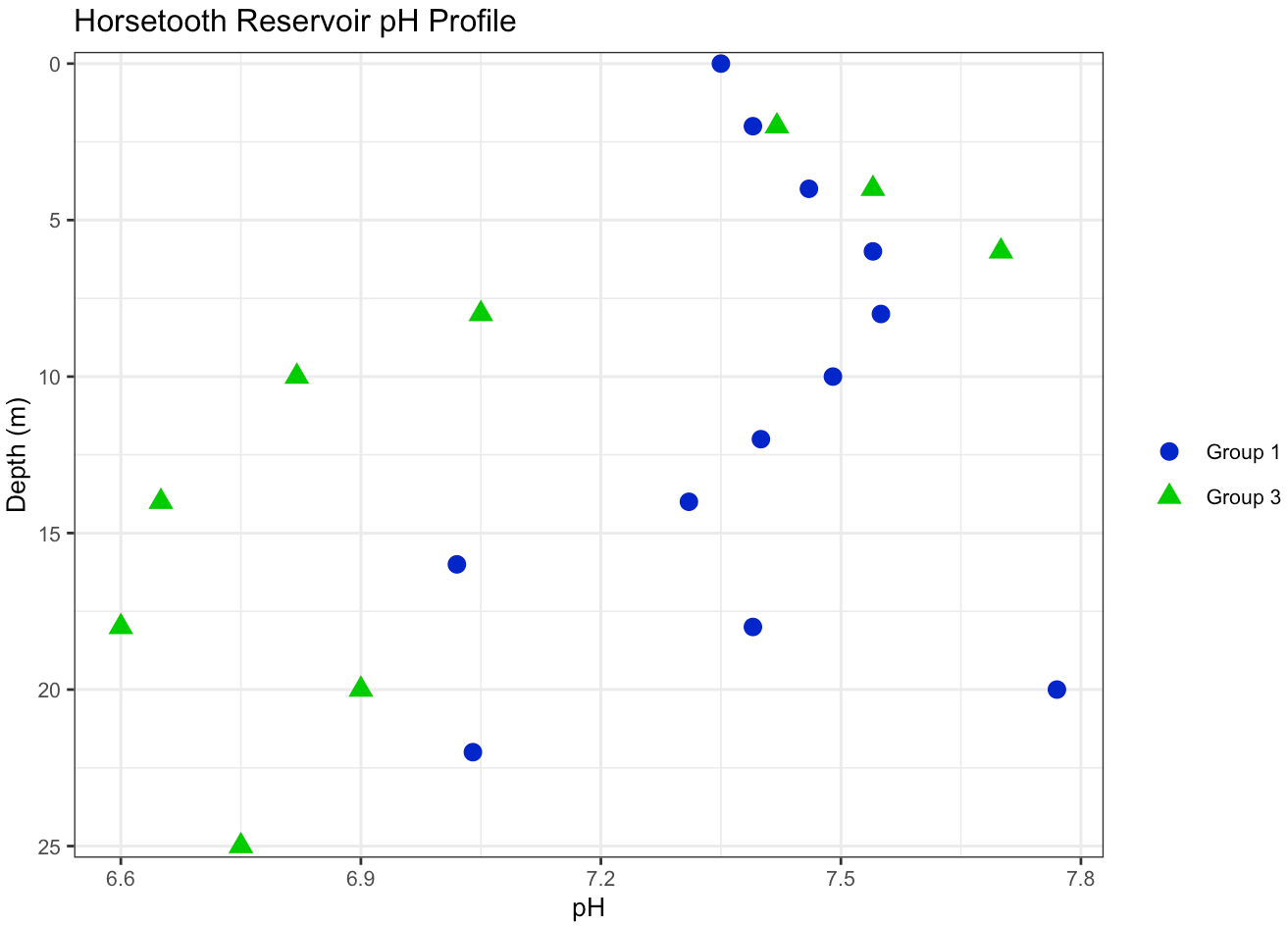
2. A temperature profile

```
temp_prof <- ggplot(data = temp, mapping = aes(x=temp_c, y=z_m, color=group, shape= grou
p)) +
  geom_point(size=3, show.legend = TRUE)+
  theme_bw() +
  labs( x='Temp (C)',
        y= 'Depth (m)',
        color = "",
        shape = "",
        title = "Horsetooth Reservoir Temperature Profile")+
  scale_y_reverse(expand = c(0.015,0))+
  scale_color_manual(values = c("Group_1" = "blue3", "Group_3" = "green3"),
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) +
  scale_shape_manual(values = c("Group_1" = 16, "Group_3" = 17),
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) +
  scale_x_continuous(breaks = seq(floor(min(temp$temp_c)), ceiling(max(temp$temp_c)), by
= 1.5)) +
  theme(text = element_text(size = 10))

temp_prof
```



3. A pH profile

```
pH_prof <- ggplot(pH,aes(x=ph,y=z_m, color=group, shape= group)) +
  geom_point(size=3, show.legend = TRUE)+
  theme_bw() +
  labs( x='pH',
        y= 'Depth (m)',
        color = "",
        shape = "",
        title = "Horsetooth Reservoir pH Profile")+
  scale_y_reverse(expand = c(0.015,0))+
  scale_color_manual(values = c("Group_1" = "blue3", "Group_3" = "green3"),
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) +
  scale_shape_manual(values = c("Group_1" = 16, "Group_3" = 17),
                     labels = c("Group_1" = "Group 1", "Group_3" = "Group 3")) +
  theme(text = element_text(size = 10))

pH_prof
```



4. The standard curve for TN from that prior assignment (but created in R). Add your group's samples' TN values on the curve in a conspicuous manner (e.g. different color or shape)

```
mgL <- read_csv("/Users/wamclean/github/LimnologyLabs/FinalProject/ESS474Fall2024stdcurvedata.csv")
```

```
## Rows: 26 Columns: 2
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## dbl (2): mg_L_TN, Area
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
mgL <- mgL %>%
  mutate(color = "1")

SampleData <- read_csv("/Users/wamclean/github/LimnologyLabs/Lab5/stdcurvedata.csv")
```

```
## Rows: 104 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (2): Sample_Name, Analysis_Inj
## dbl (4): sample_order, Inj_Number, Area, Excluded
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

What is the equation of the standard curve?

```
model_data <- mgL %>%
  filter(Area != 249.1000)

AreaModel <- lm(mg_L_TN ~ Area, model_data)
summary(AreaModel)
```

```
##
## Call:
## lm(formula = mg_L_TN ~ Area, data = model_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.047935 -0.021585 -0.007223  0.022731  0.053688
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.810e-04  7.365e-03   0.079    0.938
## Area        3.927e-02  7.792e-05 504.055   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0292 on 23 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 2.541e+05 on 1 and 23 DF,  p-value: < 2.2e-16
```

y = 0.03927x - 0.0005810

What is the R2 of your curve?

R-Squared = 0.9999

Did you remove any of the points from your curve? Why or why not?

Yes, I removed the point where the Area value was 249.1000 because it appeared to be an outlier. Doing so resulted in a change in my R-squared value from 0.9998 to 0.9999

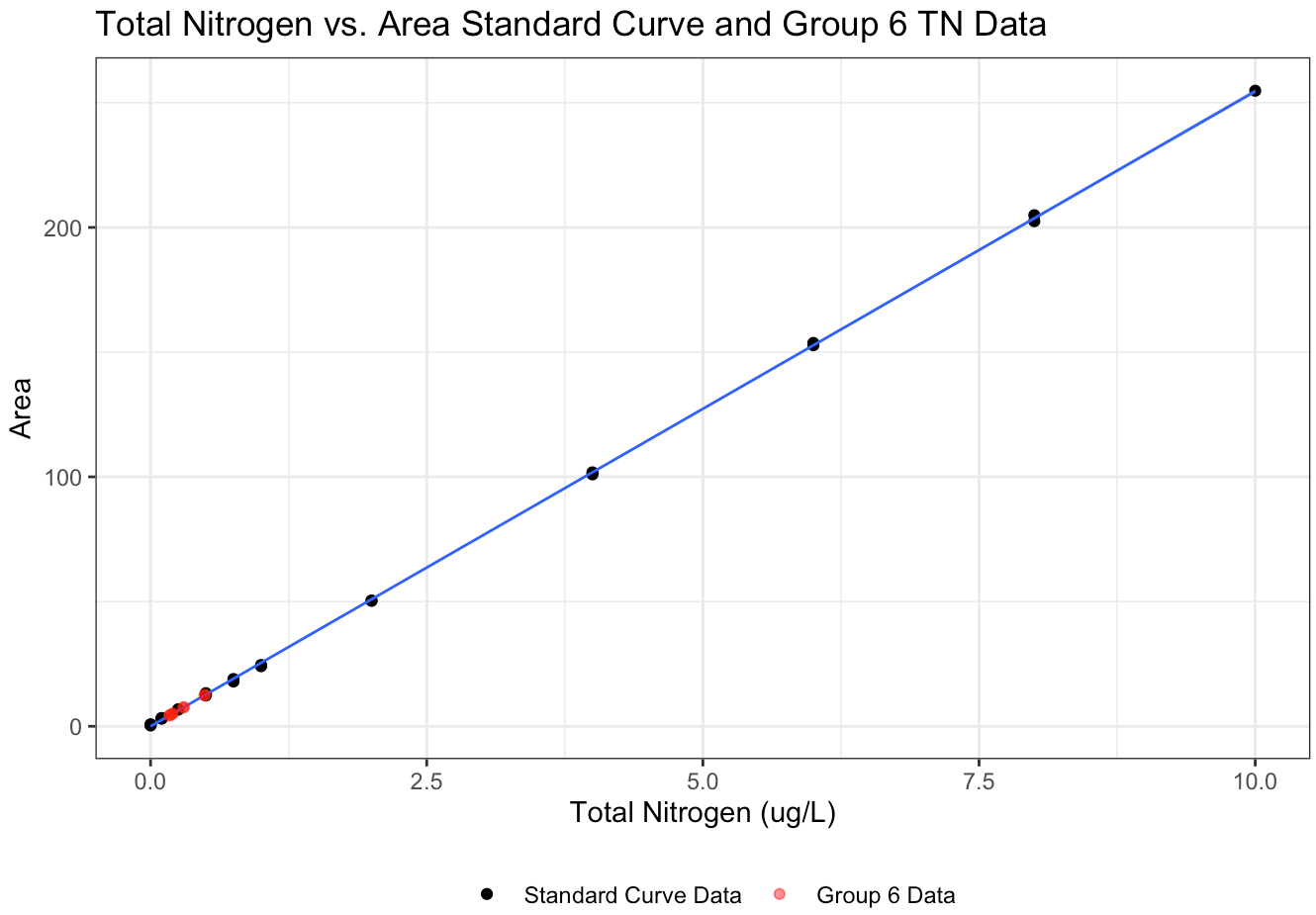
```
x <- SampleData$Area

SampleData <- SampleData %>%
  mutate(N = 0.03927*x - 0.0005810) %>%
  filter(Excluded != 1)

GroupTotals <- SampleData %>%
  separate(Sample_Name, into = c("Group_Number", "Sample_Number"), sep = "-", extra = "m
erge", fill = "right")

Group6_TN <- GroupTotals %>%
  filter(Group_Number == "G6")
```

```
#Plot Standard Curve, filtering out Area = 249.1000 as doing to brings the R2 from 0.999
8 to 0.9999

TN <- ggplot(data = filter(mgL, Area != 249.1000), mapping = aes(x = mg_L_TN, y = Area))
+
  theme_bw()+
  geom_point(aes(color = color))+
  geom_smooth(method="lm", size = .5)+
  geom_point(data = Group6_TN, mapping = aes(x = N, y = Area, color = Group_Number), alp
ha = .5)+
  labs(
    x = "Total Nitrogen (ug/L)",
    y = "Area",
    title = "Total Nitrogen vs. Area Standard Curve and Group 6 TN Data",
    color = "",
  )+
  scale_color_manual(values = c("1" = "black", "G6" = "red"),
                     labels = c("1" = "Standard Curve Data", "G6" = "Group 6 Data"))+
  theme(legend.position = "bottom")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

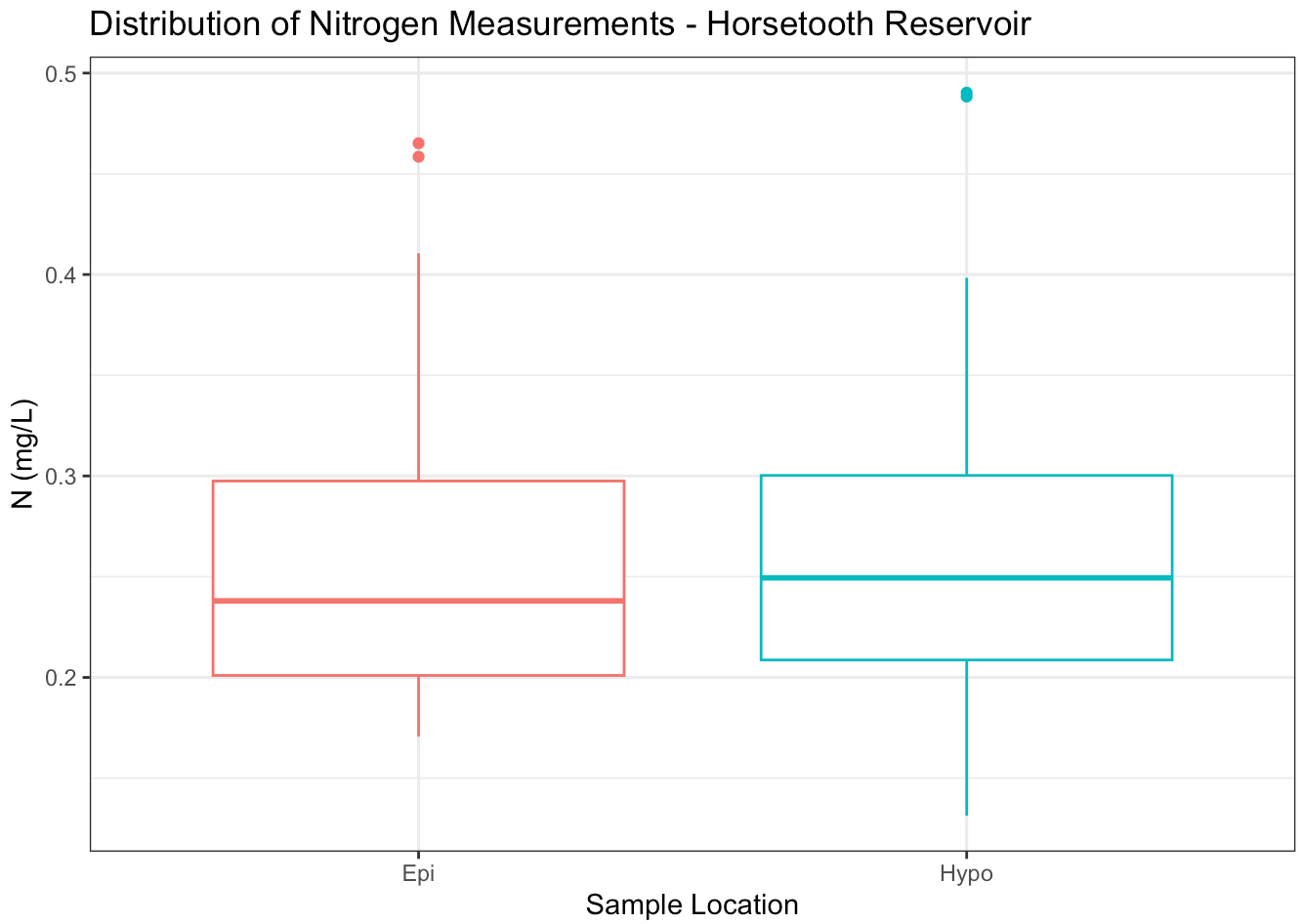
```
TN
```

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



5. A comparison of total nitrogen between the epilimnion and hypolimnion depths

```
#Filter to just epilimnion and hypolimnion samples, leaving out samples that do not desi
gnate location in the water column. Adding a SampleLocation Column for ease of plotting
Epi <- GroupTotals %>%
  filter(str_detect(Sample_Number, "E-")) %>%
  mutate(SampleLocation = "Epi")

Hypo <- GroupTotals %>%
  filter(str_detect(Sample_Number, "H-")) %>%
  mutate(SampleLocation = "Hypo")

#Bind them back together to compare
N_Comparison <- bind_rows(Epi, Hypo)

#Boxplot to show distribution of data
ggplot()+
  theme_bw()+
  geom_boxplot(data = N_Comparison, mapping = aes(x = SampleLocation, y = N, color = Sam
pleLocation))+
  labs(x = "Sample Location",
       y = "N (mg/L)",
       color = "",
       title = "Distribution of Nitrogen Measurements - Horsetooth Reservoir")+
  theme(legend.position = "none")
```




```
#Calculate average values
MeanN_Epi <- mean(Epi$N)
MeanN_Epi
```

```
## [1] 0.2637103
```

```
MeanN_Hypo <- mean(Hypo$N)
MeanN_Hypo
```

```
## [1] 0.2677313
```

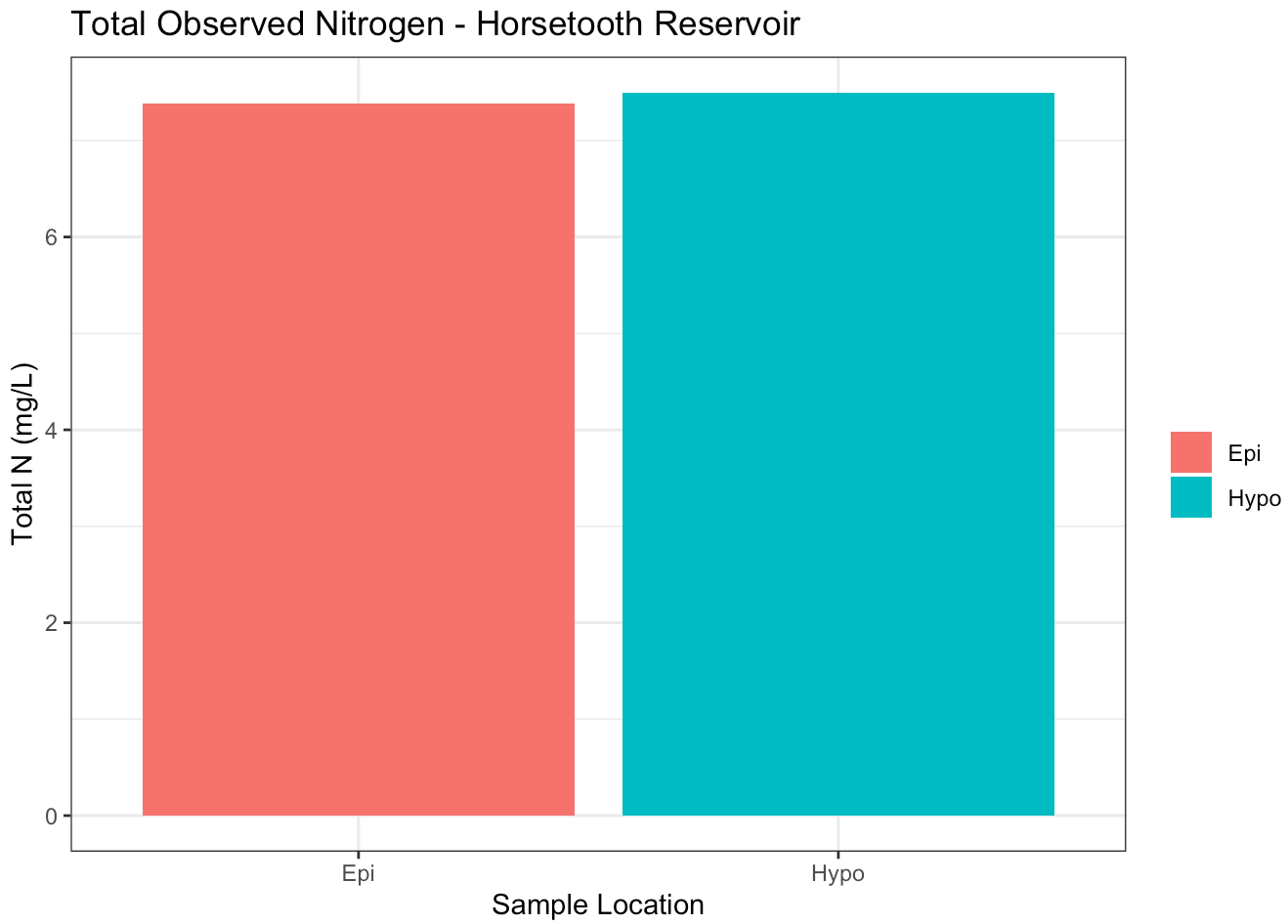
```
#Total N for each
Epi_TotalN <- sum(Epi$N)
Epi_TotalN
```

```
## [1] 7.383889
```

```
Hypo_TotalN <- sum(Hypo$N)
Hypo_TotalN
```

```
## [1] 7.496476
```

```
N_Totals <- N_Comparison %>%
  group_by(SampleLocation) %>%
  summarize(totalN = sum(N)) %>%
  ungroup()

ggplot()+
  theme_bw()+
  geom_col(data = N_Totals, mapping = aes(x = SampleLocation, y = totalN, fill = SampleL
ocation))+
  labs(x = "Sample Location",
       y = "Total N (mg/L)",
       fill = "",
       title = "Total Observed Nitrogen - Horsetooth Reservoir")
```

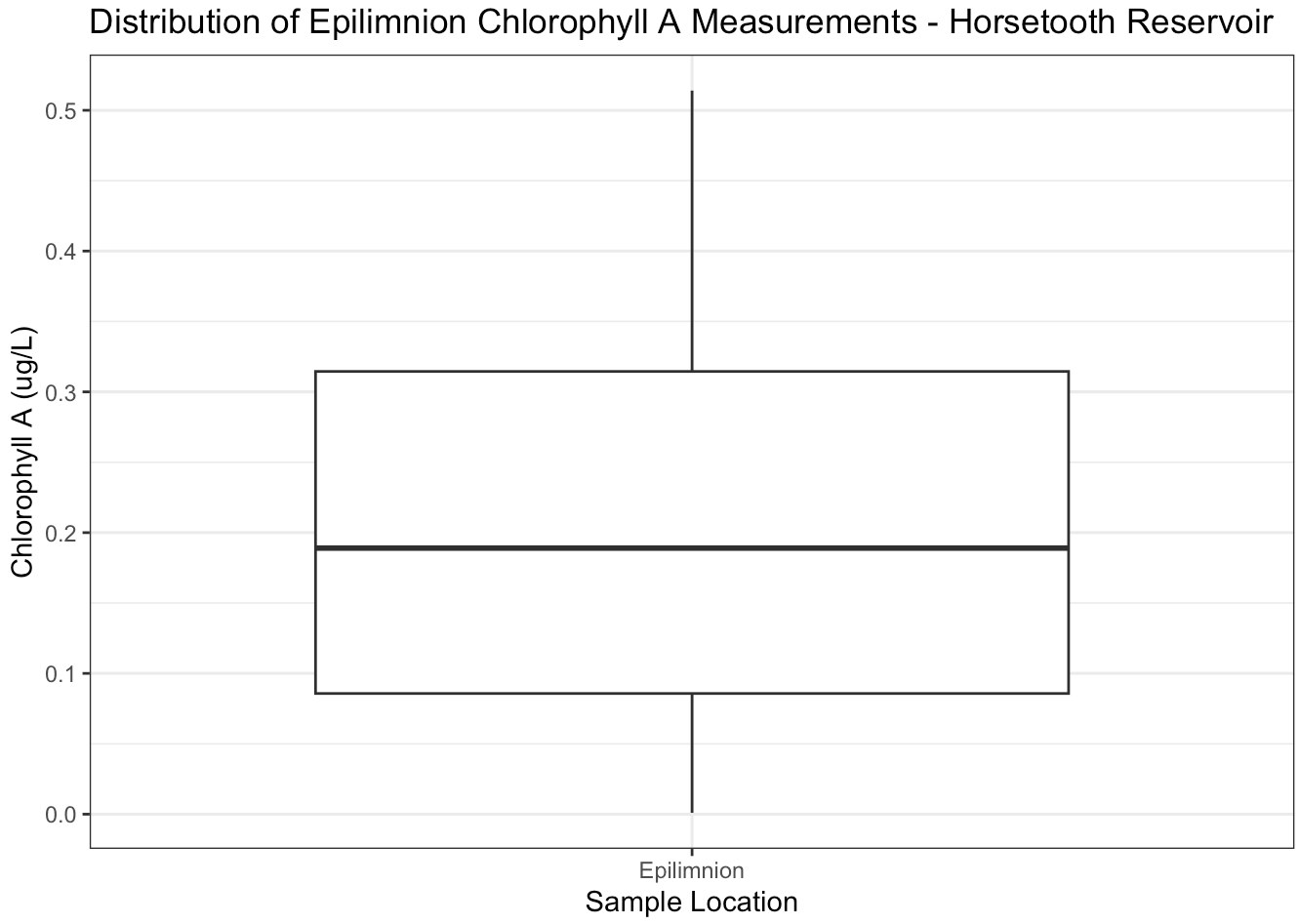


6. A figure illustrating the distribution of epilimnion chlorophyll a values taken by everyone, and the mean value.

```
ChlA <- read_csv("/Users/wamclean/github/LimnologyLabs/FinalProject/Horsetooth_Chla_data
_2024.csv") %>%
  mutate(SampleLocation = "Epilimnion")
```

```
## Rows: 14 Columns: 4
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (3): Group_Number, Sample_Label, FilterFolded
## dbl (1): Chla_ugL
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
ChlA_Plot <- ggplot()+
  theme_bw()+
  geom_boxplot(data = ChlA, mapping = aes(x = SampleLocation, y = Chla_ugL))+
  labs(x = "Sample Location",
       y = "Chlorophyll A (ug/L)",
       title = "Distribution of Epilimnion Chlorophyll A Measurements - Horsetooth Reser
voir")

ChlA_Plot
```




```
ChlA_Mean <- mean(ChlA$Chla_ugL)
ChlA_Mean
```

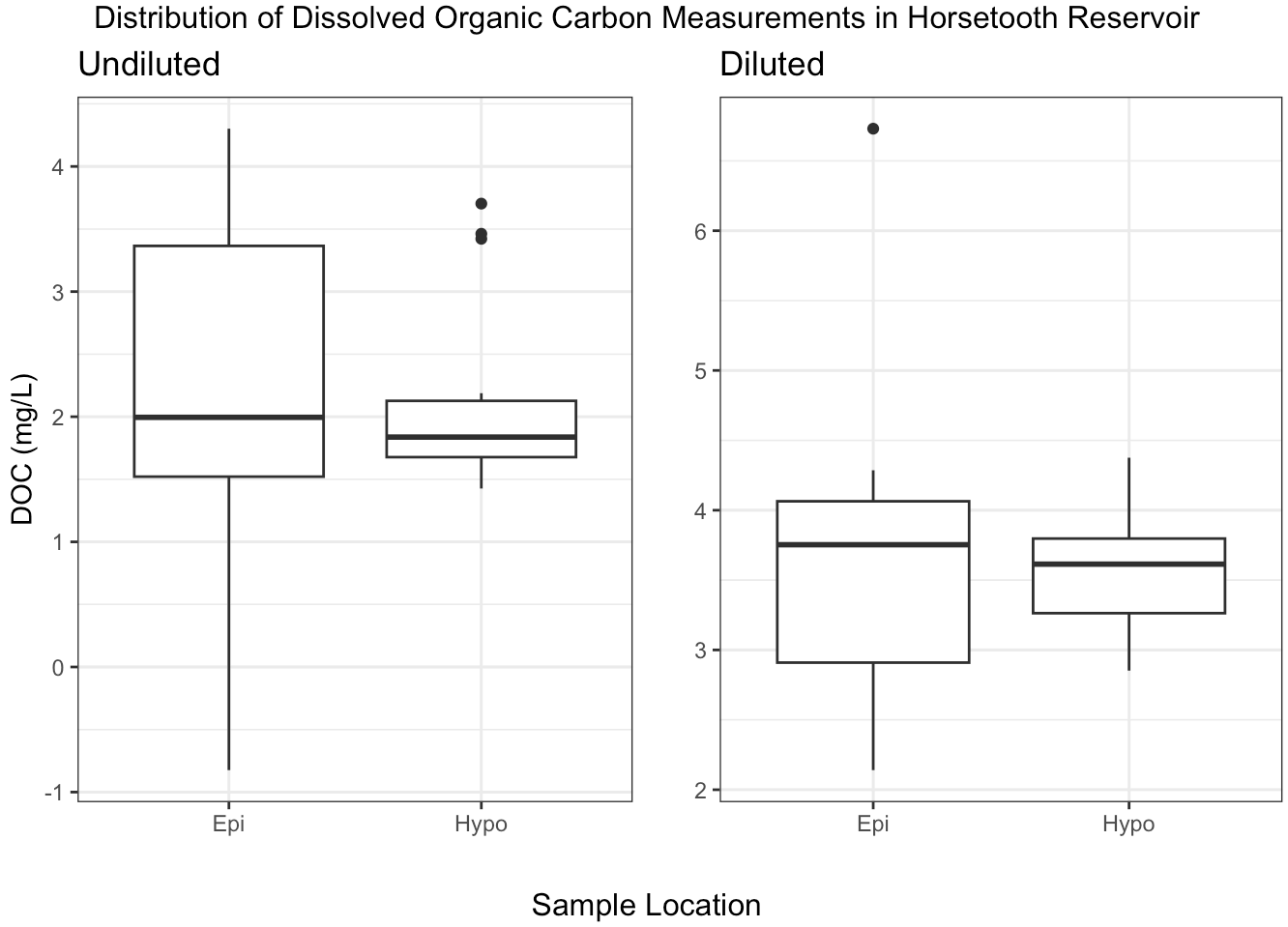
```
## [1] 0.2212857
```

7. A comparison of dissolved organic carbon values between the epilimnion and hypolimnion depths

```
#Read in DOC
DOC <- read_csv("/Users/wamclean/github/LimnologyLabs/FinalProject/ESS474_HorsetoothRese
rvoir_DOC_2024.csv")
```

```
## Rows: 51 Columns: 4
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (2): Sample_Name, Unit
## dbl (2): Result_DissolvedOrganicCarbon, Dilution_Result_DissolvedOrganicCarbon
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
DOC <- DOC %>%
  separate(Sample_Name, into = c("Group_Number", "Sample_Number"), sep = "-", extra = "m
erge", fill = "right")

DOC <- DOC %>%
  filter(str_detect(Sample_Number, "-")) %>%
  mutate(SampleLocation = if_else(str_detect(Sample_Number, "E-"), "Epi", "Hypo"))

DOC_Dil <- DOC %>%
  filter(str_detect(Sample_Number, "Dil"))

diluted <- ggplot()+
  theme_bw()+
  geom_boxplot(data = DOC_Dil, mapping = aes(x = SampleLocation, y = Dilution_Result_Dis
solvedOrganicCarbon))+
  labs(title = "Diluted",
       x = "",
       y = "")

undiluted <- ggplot()+
  theme_bw()+
  geom_boxplot(data = DOC, mapping = aes(x = SampleLocation, y = Result_DissolvedOrganic
Carbon))+
  labs(title = "Undiluted",
       x = "",
       y = "DOC (mg/L)")

grid.arrange(undiluted, diluted, ncol = 2, top = "Distribution of Dissolved Organic Carb
on Measurements in Horsetooth Reservoir", bottom = "Sample Location")
```



8. An estimate of your group's phytoplankton diversity (no figure required) #Shannon Index Value

```
Shannon <- 1.080317
Simpson <- 0.6127385

Shannon
```

```
## [1] 1.080317
```

```
Simpson
```

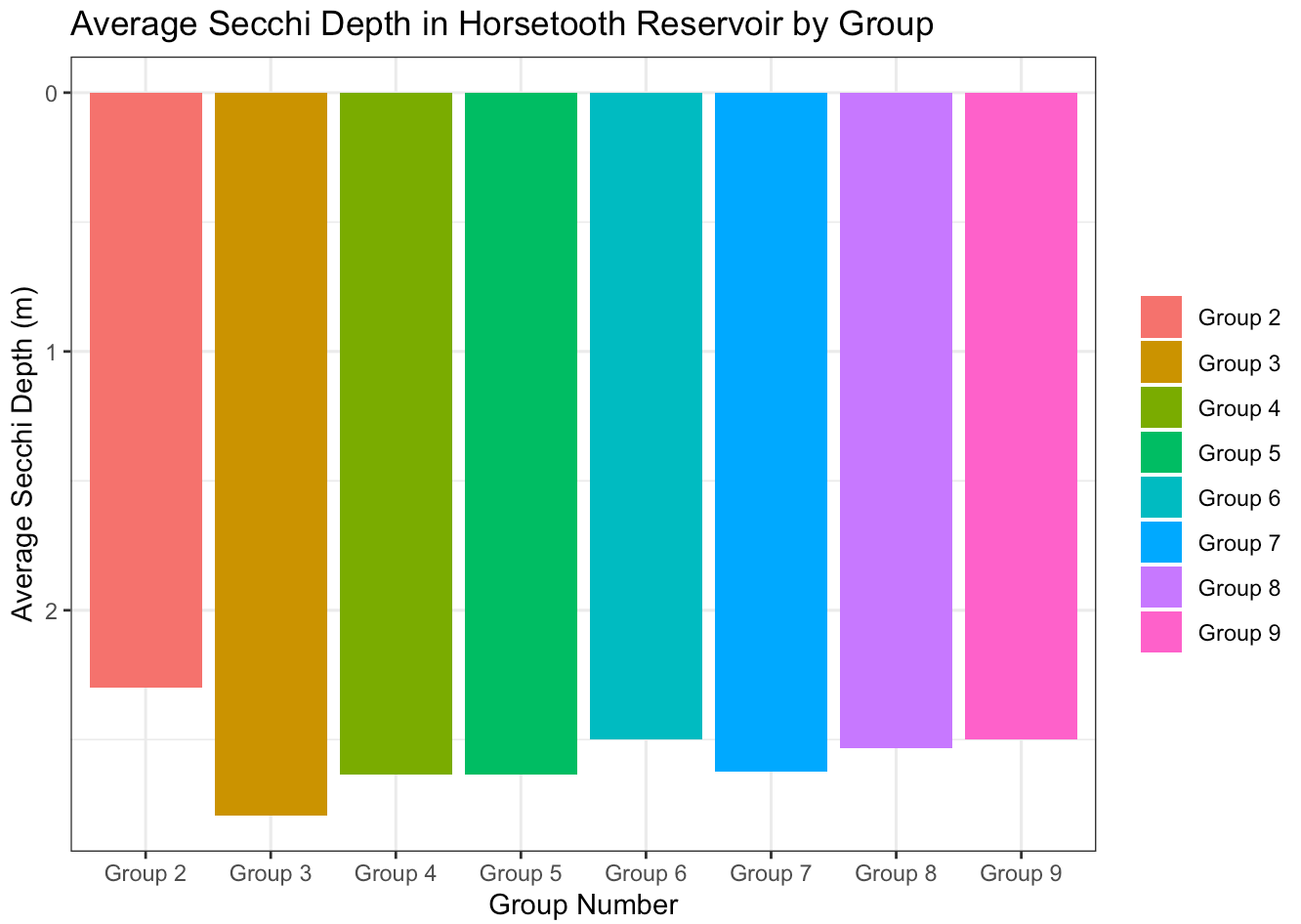
```
## [1] 0.6127385
```

9. A figure illustrating the distribution of Secchi depths taken by everyone, and the mean value.

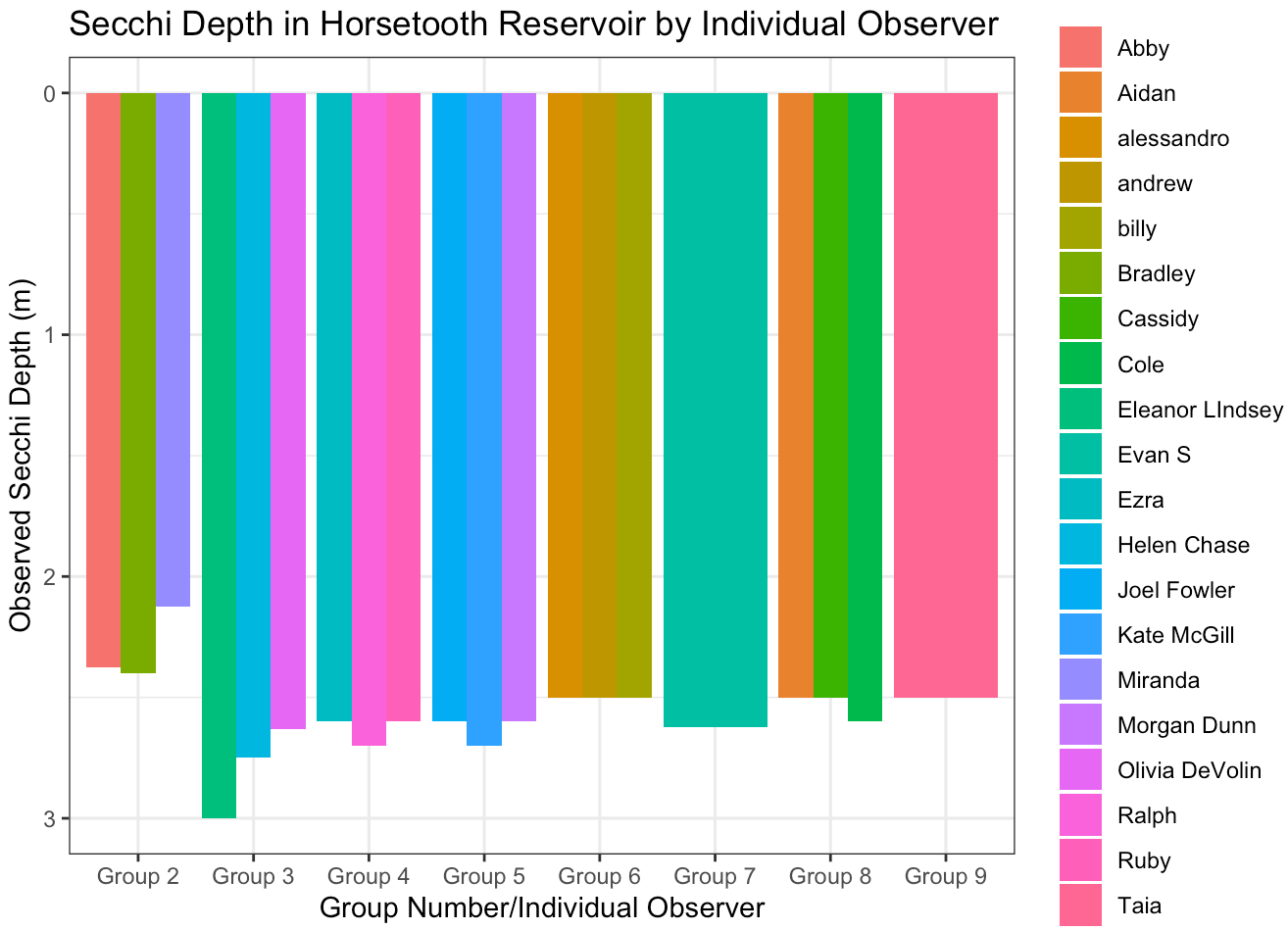
```
Secchi <- read_csv("/Users/wamclean/github/LimnologyLabs/FinalProject/HorsetoothSampling
_SecchiDepth_2024.csv")
```

```
## Rows: 23 Columns: 3
## ── Column specification ────────────────────────────────────────────────────
## Delimiter: ","
## chr (3): group_number, Sample_taker, depth_m
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Filtering out group 10 as their notation is unclear
Secchi <- Secchi %>%
  filter(group_number != "Group 10") %>%
  mutate(depth_m = as.numeric(depth_m))

SecchiMean <- Secchi %>%
  group_by(group_number) %>%
  summarize(MeanDepth = mean(depth_m)) %>%
  ungroup()


ggplot()+
  theme_bw()+
  geom_col(data = SecchiMean, mapping = aes(x = group_number, y = MeanDepth, fill = grou
p_number))+
  scale_y_reverse()+
  labs(x = "Group Number",
       y = "Average Secchi Depth (m)",
       fill = "",
       title = "Average Secchi Depth in Horsetooth Reservoir by Group")
```




```
ggplot()+
  theme_bw()+
  geom_col(data = Secchi, mapping = aes(x = group_number, y = depth_m, fill = Sample_tak
er), position = "dodge")+
  scale_y_reverse()+
  labs(x = "Group Number/Individual Observer",
       y = "Observed Secchi Depth (m)",
       fill = "",
       title = "Secchi Depth in Horsetooth Reservoir by Individual Observer")
```



e. In some cases, not every group took samples for each water quality property, or they took them with the incorrect procedure, or their data is not realistic. Thus, before calculating class means, you should filter out missing data and data from groups who you think is not reliable (include an explanation of this filtering out in the Results (interpretation) section below). If you think your own group's data is unreliable, still create a figure with the filtered class data and your group's data if it exists, but also include a note in the Results (interpretation) section below explaining the circumstances; also, rely on the filtered pooled class data values to answer any subsequent questions.