Dionex_COMPASS_TEMPEST

Stephanie J. Wilson

2023-06-06

Daily Set up

Read in and Format the raw data - change wd & file names (Make sure files are UTF-8 encoded)

```
# SULFATE DATA:

## Read in raw data file from Dionex - copied and saved as a txt
Sdat <- read.table("Raw Data/COMPASS_TEMPEST_20230606_SO4a.txt",sep='\t' , header=T, skip=3)
head(Sdat)
```

```
##    X                    X.1     X.2 IC.SO4 IC.SO4.1 IC.SO4.2 IC.SO4.3 IC.SO4.4
## 1 44  TEMPEST_C_C6_20230603 Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 2 45  TEMPEST_C_H6_20230603 Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 3 46  TEMPEST_C_I5_20230603 Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 4 47  TEMPEST_C_H6_20230605 Unknown 4.373   2.3894     0.47   0.0544     0.35
## 5 48 TEMPEST_FW_C3_20230603 Unknown 4.383   6.9627     0.69   0.1403     0.90
## 6 49 TEMPEST_FW_D5_20230603 Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
##   IC.SO4.5
## 1     n.a.
## 2     n.a.
## 3     n.a.
## 4       M
## 5       M
## 6     n.a.
```

```
## Only keep the columns that we need
Sdat <- Sdat[ ,c(2,5,7)] # dont need this here
head(Sdat)
```

```
##                     X.1 IC.SO4.1 IC.SO4.3
## 1  TEMPEST_C_C6_20230603     n.a.     n.a.
## 2  TEMPEST_C_H6_20230603     n.a.     n.a.
## 3  TEMPEST_C_I5_20230603     n.a.     n.a.
## 4  TEMPEST_C_H6_20230605   2.3894   0.0544
## 5 TEMPEST_FW_C3_20230603   6.9627   0.1403
## 6 TEMPEST_FW_D5_20230603     n.a.     n.a.
```

```
## Name the columns correctly
colnames(Sdat) <- c( "Sample_ID", "SO4_ppm", "SO4_Area")
Sdat$Sample_ID <- as.factor(Sdat$Sample_ID)
Sdat$SO4_ppm <- as.numeric(Sdat$SO4_ppm)
```

```
## Warning: NAs introduced by coercion
```

```
Sdat$SO4_Area <- as.numeric(Sdat$SO4_Area)
```

```
## Warning: NAs introduced by coercion
```

```
Sdat <- as.data.frame(Sdat)
head(Sdat)
```

```
##                Sample_ID SO4_ppm SO4_Area
## 1  TEMPEST_C_C6_20230603      NA       NA
## 2  TEMPEST_C_H6_20230603      NA       NA
## 3  TEMPEST_C_I5_20230603      NA       NA
## 4  TEMPEST_C_H6_20230605  2.3894   0.0544
## 5 TEMPEST_FW_C3_20230603  6.9627   0.1403
## 6 TEMPEST_FW_D5_20230603      NA       NA
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_TEMPEST_20230606_Cla.txt",sep='\t' , header=T, skip=3)
head(Cldat)
```

```
##    X                    X.1     X.2 IC.Cl IC.Cl.1 IC.Cl.2 IC.Cl.3 IC.Cl.4
## 1 44  TEMPEST_C_C6_20230603 Unknown  n.a.    n.a.    n.a.    n.a.    n.a.
## 2 45  TEMPEST_C_H6_20230603 Unknown  n.a.    n.a.    n.a.    n.a.    n.a.
## 3 46  TEMPEST_C_I5_20230603 Unknown  n.a.    n.a.    n.a.    n.a.    n.a.
## 4 47  TEMPEST_C_H6_20230605 Unknown  n.a.    n.a.    n.a.    n.a.    n.a.
## 5 48 TEMPEST_FW_C3_20230603 Unknown  n.a.    n.a.    n.a.    n.a.    n.a.
## 6 49 TEMPEST_FW_D5_20230603 Unknown  n.a.    n.a.    n.a.    n.a.    n.a.
##   IC.Cl.5
## 1    n.a.
## 2    n.a.
## 3    n.a.
## 4    n.a.
## 5    n.a.
## 6    n.a.
```

```
## Only keep the columns that we need
Cldat <- Cldat[ ,c(2,5,7)]
head(Cldat)
```

```
##                      X.1 IC.Cl.1 IC.Cl.3
## 1  TEMPEST_C_C6_20230603    n.a.    n.a.
## 2  TEMPEST_C_H6_20230603    n.a.    n.a.
## 3  TEMPEST_C_I5_20230603    n.a.    n.a.
## 4  TEMPEST_C_H6_20230605    n.a.    n.a.
## 5 TEMPEST_FW_C3_20230603    n.a.    n.a.
## 6 TEMPEST_FW_D5_20230603    n.a.    n.a.
```

```r
## Name the columns correctly
colnames(Cldat) <- c( "Sample_ID", "Cl_ppm", "Cl_Area")
Cldat$Sample_ID <- as.factor(Cldat$Sample_ID)
Cldat$Cl_ppm <- as.numeric(Cldat$Cl_ppm)
```

```
## Warning: NAs introduced by coercion
```

```r
Cldat$Cl_Area <- as.numeric(Cldat$Cl_Area)
```

```
## Warning: NAs introduced by coercion
```

```r
Cldat <- as.data.frame(Cldat)
head(Cldat)
```

```
##                Sample_ID Cl_ppm Cl_Area
## 1  TEMPEST_C_C6_20230603     NA      NA
## 2  TEMPEST_C_H6_20230603     NA      NA
## 3  TEMPEST_C_I5_20230603     NA      NA
## 4  TEMPEST_C_H6_20230605     NA      NA
## 5 TEMPEST_FW_C3_20230603     NA      NA
## 6 TEMPEST_FW_D5_20230603     NA      NA
```

```r
## Bring the data back together:
all_dat <- merge(Sdat, Cldat, by="Sample_ID")
head(all_dat)
```

```
##   Sample_ID  SO4_ppm SO4_Area   Cl_ppm  Cl_Area
## 1           168.9031  23.3576 203.7164 279.6752
## 2           168.9031  23.3576  29.8315  32.6201
## 3           168.9031  23.3576   0.0014   0.0001
## 4           168.9031  23.3576  40.9777  53.7128
## 5           168.9031  23.3576       NA       NA
## 6            13.6022   2.4364 203.7164 279.6752
```

```r
## Remove empty lines
all_dat <- all_dat[!(is.na(all_dat$Sample_ID) | all_dat$Sample_ID==""), ]
head(all_dat)
```

```
##    Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area
## 26 Lab Blank      NA       NA     NA      NA
## 27 Lab Blank      NA       NA 0.0079  0.0108
## 28 Lab Blank      NA       NA     NA      NA
## 29 Lab Blank      NA       NA     NA      NA
## 30 Lab Blank      NA       NA 0.0114  0.0156
## 31 Lab Blank      NA       NA     NA      NA
```

```r
all_dat <- droplevels(all_dat[!all_dat$Sample_ID == 'Lab Blank',])
head(all_dat)
```

```
##      Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area
## 107 Standard 1  0.4622   0.5264  5.2376  7.1905
## 108 Standard 1  0.4622   0.5264  5.1529  7.0742
## 109 Standard 1  0.4374   0.4982  5.2376  7.1905
## 110 Standard 1  0.4374   0.4982  5.1529  7.0742
## 111 Standard 2  1.0048   1.1446 11.4708 15.7479
## 112 Standard 2  1.0048   1.1446 11.3597 15.5954
```

```
all_dat[is.na(all_dat)] <- 0
```

Pull out standards - could do some checks here if we want

```
stds <- all_dat[grepl("Standard", all_dat$Sample_ID),]
head(stds)
```

```
##      Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area
## 107 Standard 1  0.4622   0.5264  5.2376  7.1905
## 108 Standard 1  0.4622   0.5264  5.1529  7.0742
## 109 Standard 1  0.4374   0.4982  5.2376  7.1905
## 110 Standard 1  0.4374   0.4982  5.1529  7.0742
## 111 Standard 2  1.0048   1.1446 11.4708 15.7479
## 112 Standard 2  1.0048   1.1446 11.3597 15.5954
```

```
stds_chk_S <- stds %>%
    group_by(Sample_ID) %>%
    summarise(mean = mean(SO4_ppm), sd = sd(SO4_ppm))
stds_chk_S$cv <- (stds_chk_S$sd/stds_chk_S$mean)*100
stds_chk_S$flag <-  ifelse(stds_chk_S$cv <2, 'YES', 'NO, rerun')
head(stds_chk_S)
```

```
## # A tibble: 5 x 5
##   Sample_ID    mean      sd    cv flag
##   <fct>       <dbl>   <dbl> <dbl> <chr>
## 1 Standard 1  0.450  0.0143  3.18 NO, rerun
## 2 Standard 2  0.993  0.0138  1.39 YES
## 3 Standard 3  2.02   0.0309  1.53 YES
## 4 Standard 4 10.5   NA      NA    <NA>
## 5 Standard 5 20.5   NA      NA    <NA>
```

```
stds_chk_Cl <- stds %>%
    group_by(Sample_ID) %>%
    summarise(mean = mean(Cl_ppm), sd = sd(Cl_ppm))
stds_chk_Cl$cv <- (stds_chk_Cl$sd/stds_chk_Cl$mean)*100
stds_chk_Cl$flag <-  ifelse(stds_chk_Cl$cv <2, 'YES', 'NO, rerun')
head(stds_chk_Cl)
```

```
## # A tibble: 5 x 5
##   Sample_ID    mean      sd     cv flag
##   <fct>       <dbl>   <dbl>  <dbl> <chr>
## 1 Standard 1   5.20  0.0489  0.941 YES
```

```
## 2 Standard 2  11.4   0.0641  0.562 YES
## 3 Standard 3  20.8   0.117   0.564 YES
## 4 Standard 4 103.   NA      NA     <NA>
## 5 Standard 5 204.   NA      NA     <NA>
```

```
lmS <- lm(stds$SO4_Area ~ stds$SO4_ppm)
SO4_sum <- summary(lmS)
SO4_Slope <- SO4_sum$coefficients[2, 1]
SO4_Int <- SO4_sum$coefficients[1, 1]


lmCl <- lm(stds$Cl_Area ~ stds$Cl_ppm)
Cl_sum <- summary(lmCl)
Cl_Slope <- Cl_sum$coefficients[2, 1]
Cl_Int <- Cl_sum$coefficients[1, 1]
```

Calculate mmol/L concentrations

```
#remove standards from sample dataframe
sampledat <- all_dat[!grepl("Standard", all_dat$Sample_ID),]
head(sampledat)
```

```
##                 Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area
## 126 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427
## 127 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000
## 128 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000
## 129 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000
## 130 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000
## 131 TEMPEST_C_F6_20230515  0.0000   0.0000 28.2531 35.2615
```

```
# Constants needed for calculations:
clmw <- 35.45     #molecular weight of Chloride: 35.45
smw <- 32.06      #molecular weight of sulfur: 32.06


# Convert ppm to mmol/L
sampledat$SO4_mM <- (sampledat$SO4_ppm / smw)
sampledat$Cl_mM <- (sampledat$Cl_ppm / clmw)


# Calculate Salinity
# calculated using the Knudsen equation
# Salinity = 0.03 + 1.8050 * Chlorinity
# Ref: A Practical Handbook of Seawater Analysis by Strickland & Parsons (P. 11)
# =((1.807*Cl_ppm)+0.026)/1000
sampledat$salinity <- ((1.8070 * sampledat$Cl_ppm) + 0.026) / 1000


head(sampledat)
```

```
##                 Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
```

```
## 126 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 127 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 128 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 129 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 130 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 131 TEMPEST_C_F6_20230515  0.0000   0.0000 28.2531 35.2615 0.00000000 0.7969845
##       salinity
## 126 0.05525081
## 127 0.00002600
## 128 0.00002600
## 129 0.00002600
## 130 0.00002600
## 131 0.05107935
```

Pull out dups and check with percent difference

```
#Show me the data that we have from the calculations
head(sampledat)
```

```
##                 Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 126 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 127 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 128 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 129 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 130 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 131 TEMPEST_C_F6_20230515  0.0000   0.0000 28.2531 35.2615 0.00000000 0.7969845
##       salinity
## 126 0.05525081
## 127 0.00002600
## 128 0.00002600
## 129 0.00002600
## 130 0.00002600
## 131 0.05107935
```

```
#pull out any rows that have "dup" in the SampleID column
dups <- sampledat %>%
  filter(str_detect(Sample_ID, "_dup"))      #have to change this to match data
head(dups)
```

```
##                    Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM
## 1  TEMPEST_C_I5_20230518_dup  4.1578   4.3055 25.5009 31.8266 0.12968808
## 2 TEMPEST_SW_H3_20230603_dup  0.0000   0.0000  0.0000  0.0000 0.00000000
## 3 TEMPEST_SW_H6_20230515_dup  0.4053   0.4197 36.6535 45.7458 0.01264192
##       Cl_mM   salinity
## 1 0.7193484 0.04610613
## 2 0.0000000 0.00002600
## 3 1.0339492 0.06625887
```

```
#remove these from sample dataframe in a new dataframe
sampledat2 <- sampledat %>%
  filter(!str_detect(Sample_ID, "_dup")) %>%
```

```
  filter(!str_detect(Sample_ID, "_spk"))
head(sampledat2)
```

```
##               Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 1 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 2 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 3 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 4 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 5 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 6 TEMPEST_C_F6_20230515  0.0000   0.0000 28.2531 35.2615 0.00000000 0.7969845
##     salinity
## 1 0.05525081
## 2 0.00002600
## 3 0.00002600
## 4 0.00002600
## 5 0.00002600
## 6 0.05107935
```

```
#remove the dup from these IDs so we will have duplicate sample names
dups$Sample_ID<-gsub("_dup","",as.character(dups$Sample_ID))
dups <- dups[ ,-c(2:5)]
colnames(dups) <- c('Sample_ID', 'SO4_mM_dup', "Cl_mM_dup", "salinity_dup")
head(dups)
```

```
##               Sample_ID SO4_mM_dup Cl_mM_dup salinity_dup
## 1  TEMPEST_C_I5_20230518 0.12968808 0.7193484   0.04610613
## 2 TEMPEST_SW_H3_20230603 0.00000000 0.0000000   0.00002600
## 3 TEMPEST_SW_H6_20230515 0.01264192 1.0339492   0.06625887
```

```
#put it back together with the old data set and look for duplicates
QAdups <- merge(sampledat2, dups)
head(QAdups)
```

```
##               Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 1  TEMPEST_C_I5_20230518  4.4207   4.5778 25.4606 31.7763 0.13788833 0.7182116
## 2 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 3 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 4 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 5 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 6 TEMPEST_SW_H6_20230515  0.3839   0.3975 28.1246 35.1012 0.01197442 0.7933597
##     salinity SO4_mM_dup Cl_mM_dup salinity_dup
## 1 0.04603330 0.12968808 0.7193484   0.04610613
## 2 0.00002600 0.00000000 0.0000000   0.00002600
## 3 0.00002600 0.00000000 0.0000000   0.00002600
## 4 0.00002600 0.00000000 0.0000000   0.00002600
## 5 0.00002600 0.00000000 0.0000000   0.00002600
## 6 0.05084715 0.01264192 1.0339492   0.06625887
```

```
QAdups$SO4_dups_chk <- ((abs(QAdups$SO4_mM-QAdups$SO4_mM_dup))/((QAdups$SO4_mM+QAdups$SO4_mM_dup)/2))*1
QAdups$SO4_dups_flag <-  ifelse(QAdups$SO4_dups_chk <10, 'YES', 'NO, rerun')
```

```
QAdups$Cl_dups_chk <- ((abs(QAdups$Cl_mM-QAdups$Cl_mM_dup))/((QAdups$Cl_mM+QAdups$Cl_mM_dup)/2))*100
QAdups$Cl_dups_flag <-  ifelse(QAdups$Cl_dups_chk <10, 'YES', 'NO, rerun')

head(QAdups)
```

```
##                Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 1  TEMPEST_C_I5_20230518  4.4207   4.5778 25.4606 31.7763 0.13788833 0.7182116
## 2 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 3 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 4 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 5 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 6 TEMPEST_SW_H6_20230515  0.3839   0.3975 28.1246 35.1012 0.01197442 0.7933597
##     salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 0.04603330 0.12968808 0.7193484   0.04610613     6.129277           YES
## 2 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 3 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 4 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 5 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 6 0.05084715 0.01264192 1.0339492   0.06625887     5.423213           YES
##   Cl_dups_chk Cl_dups_flag
## 1   0.1581586          YES
## 2         NaN         <NA>
## 3         NaN         <NA>
## 4         NaN         <NA>
## 5         NaN         <NA>
## 6  26.3326649    NO, rerun
```

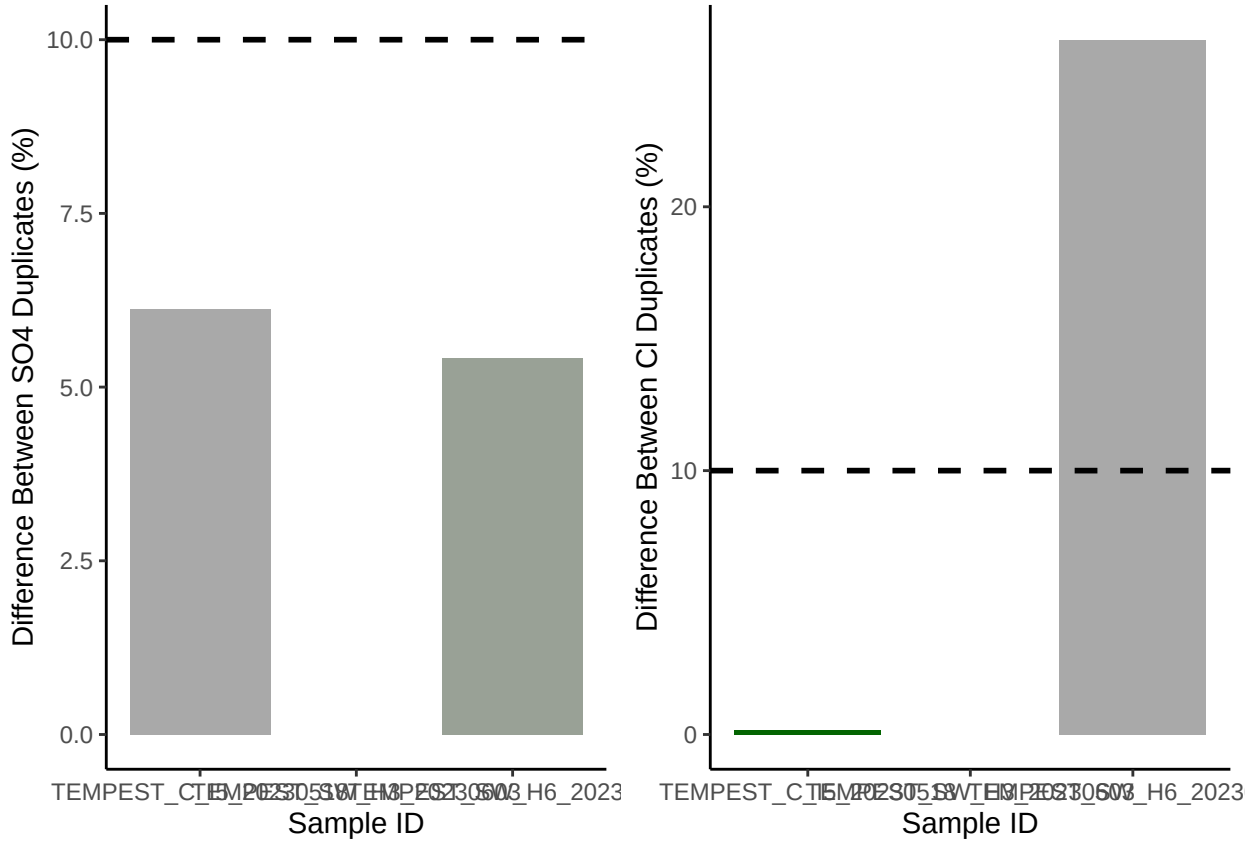
```
#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
Sdupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = SO4_dups_chk, fill=SO4_dups_chk)) +
       geom_bar(stat = 'identity') +
        scale_fill_gradient2(low='darkgreen', mid='darkgreen', high='darkgrey', space='Lab') +
        theme_classic() + labs(x= "Sample ID", y="Difference Between SO4 Duplicates (%)") +
        theme(legend.position="none") +  geom_hline(yintercept=10, linetype="dashed",
                color = "black", size=1)
```

```
## 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.
```

```
Cldupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = Cl_dups_chk, fill=Cl_dups_chk)) +
       geom_bar(stat = 'identity') +
        scale_fill_gradient2(low='darkgreen', mid='darkgreen', high='darkgrey', space='Lab') +
        theme_classic() + labs(x= "Sample ID", y="Difference Between Cl Duplicates (%)") +
        theme(legend.position="none") +  geom_hline(yintercept=10, linetype="dashed",
                color = "black", size=1)

ggarrange(Sdupsbar, Cldupsbar,ncol=2, nrow=1)
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (‘geom_bar()‘).
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_dups <- QAdups %>%
  group_by(SO4_dups_flag) %>%
  summarise(S_no_rows = length(SO4_dups_flag))
Perc_dups <- Perc_dups %>%
    filter(!is.na(SO4_dups_flag))
Perc_dups1 <- QAdups %>%
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(Cl_dups_flag))
Perc_dups1 <- Perc_dups1 %>%
    filter(!is.na(Cl_dups_flag))
colnames(Perc_dups) <- c("Flag", "S_no_rows")
colnames(Perc_dups1) <- c("Flag", "Cl_no_rows")
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)

Perc_dups2$Total <- length(QAdups$SO4_dups_flag)
Perc_dups2$S_Percent <- (Perc_dups2$S_no_rows / Perc_dups2$Total)*100
Perc_dups2$Cl_Percent <- (Perc_dups2$Cl_no_rows / Perc_dups2$Total)*100
head(Perc_dups2)
```

```
##   Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES         2 NO, rerun          1     6  33.33333   16.66667
## 2  YES         2       YES          1     6  33.33333   16.66667
```

Pull out dups and check with cv

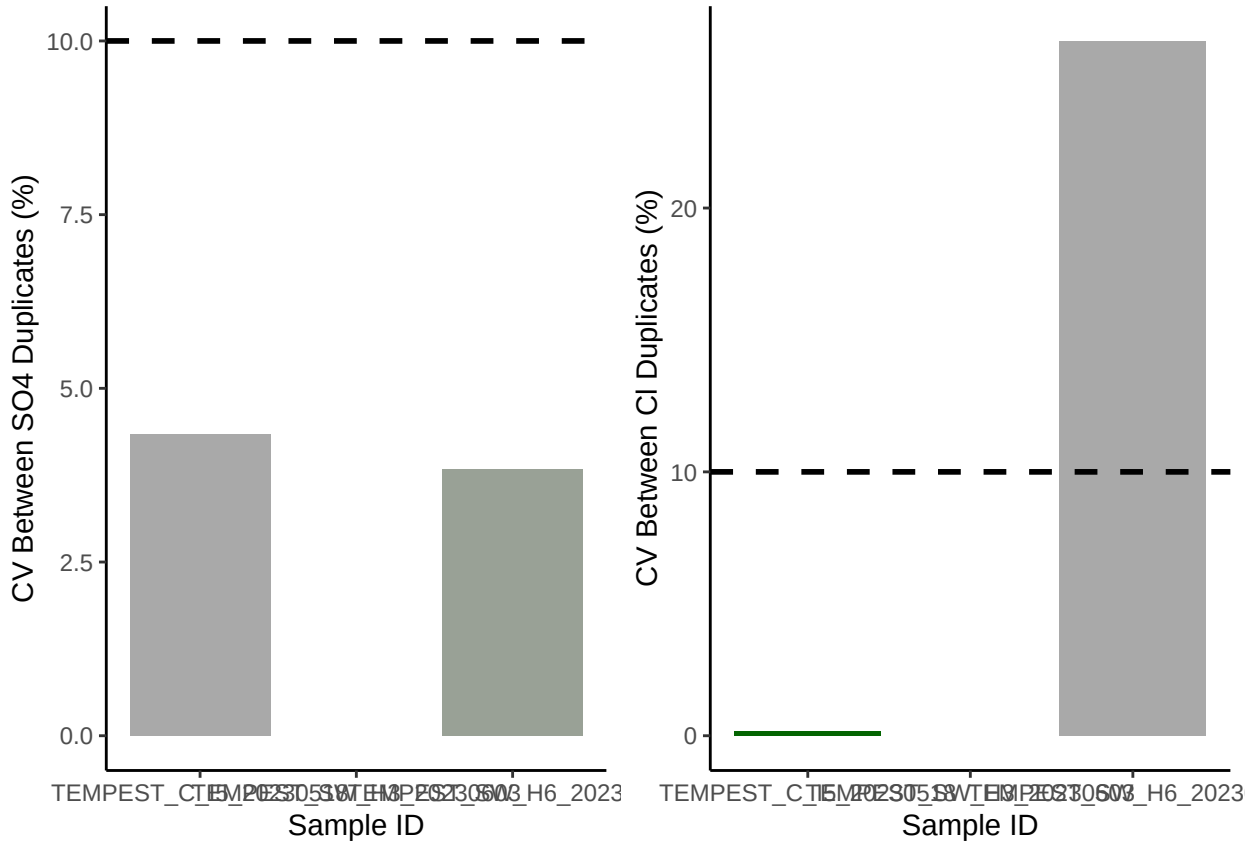
```
#the cv
# calculate the sd bewteen the two duplicates then divide by the mean and multiply by 100

df2 <- as.data.frame(QAdups$SO4_mM)
df2$dups <- QAdups$SO4_mM_dup

df2$sds <- apply(df2,1,sd)

QAdups$SO4_dups_cv <- (df2$sds)/((QAdups$SO4_mM+QAdups$SO4_mM_dup)/2) * 100
QAdups$SO4_dups_cv_flag <-  ifelse(QAdups$SO4_dups_cv <11, 'YES', 'NO, rerun')

head(QAdups)
```

```
##                Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 1  TEMPEST_C_I5_20230518  4.4207   4.5778 25.4606 31.7763 0.13788833 0.7182116
## 2 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 3 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 4 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 5 TEMPEST_SW_H3_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 6 TEMPEST_SW_H6_20230515  0.3839   0.3975 28.1246 35.1012 0.01197442 0.7933597
##     salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 0.04603330 0.12968808 0.7193484   0.04610613     6.129277           YES
## 2 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 3 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 4 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 5 0.00002600 0.00000000 0.0000000   0.00002600          NaN          <NA>
## 6 0.05084715 0.01264192 1.0339492   0.06625887     5.423213           YES
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1   0.1581586          YES    4.334053              YES
## 2         NaN         <NA>         NaN             <NA>
## 3         NaN         <NA>         NaN             <NA>
## 4         NaN         <NA>         NaN             <NA>
## 5         NaN         <NA>         NaN             <NA>
## 6  26.3326649    NO, rerun    3.834791              YES
```

```
#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
Sdupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = SO4_dups_cv, fill=SO4_dups_chk)) +
       geom_bar(stat = 'identity') +
        scale_fill_gradient2(low='darkgreen', mid='darkgreen', high='darkgrey', space='Lab') +
        theme_classic() + labs(x= "Sample ID", y="CV Between SO4 Duplicates (%)") +
        theme(legend.position="none") +  geom_hline(yintercept=10, linetype="dashed",
                color = "black", size=1)


Cldupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = Cl_dups_chk, fill=Cl_dups_chk)) +
       geom_bar(stat = 'identity') +
        scale_fill_gradient2(low='darkgreen', mid='darkgreen', high='darkgrey', space='Lab') +
        theme_classic() + labs(x= "Sample ID", y="CV Between Cl Duplicates (%)") +
        theme(legend.position="none") +  geom_hline(yintercept=10, linetype="dashed",
                color = "black", size=1)
```

```
ggarrange(Sdupsbar, Cldupsbar,ncol=2, nrow=1)
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (‘geom_bar()‘).
## Removed 4 rows containing missing values or values outside the scale range
## (‘geom_bar()‘).
```




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_dups <- QAdups %>%
  group_by(SO4_dups_flag) %>%
  summarise(S_no_rows = length(SO4_dups_flag))
Perc_dups <- Perc_dups %>%
    filter(!is.na(SO4_dups_flag))
Perc_dups1 <- QAdups %>%
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(Cl_dups_flag))
Perc_dups1 <- Perc_dups1 %>%
    filter(!is.na(Cl_dups_flag))
colnames(Perc_dups) <- c("Flag", "S_no_rows")
colnames(Perc_dups1) <- c("Flag", "Cl_no_rows")
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)

Perc_dups2$Total <- length(QAdups$SO4_dups_flag)
Perc_dups2$S_Percent <- (Perc_dups2$S_no_rows / Perc_dups2$Total)*100
```

```
Perc_dups2$Cl_Percent <- (Perc_dups2$Cl_no_rows / Perc_dups2$Total)*100
head(Perc_dups2)
```

```
##   Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES         2 NO, rerun          1     6  33.33333   16.66667
## 2  YES         2       YES          1     6  33.33333   16.66667
```

Pull out spikes and check with dionex output conc.

```
##Merge sample data with dilution data
Dilutions <- read.csv("Raw Data/TMP 2023 Sample Dilutions from Dionex.csv")
sampledat_dils <- merge(sampledat, Dilutions, by = "Sample_ID")
unique(sampledat_dils$Dilution)
```

```
## [1]  1.10 50.00  5.15 56.51  4.53
```

```
#Show me the data that we have from the calculations
head(sampledat_dils)
```

```
##               Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 1 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 2 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 3 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 4 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 5 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 6 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
##     salinity Dilution
## 1 0.05525081     1.10
## 2 0.05525081    50.00
## 3 0.00002600     5.15
## 4 0.00002600    56.51
## 5 0.00002600     5.15
## 6 0.00002600    56.51
```

```
#pull out any rows that have "spk" in the SampleID column
spks <- sampledat_dils %>%
  filter(str_detect(Sample_ID, "spk"))      #have to change this to match data
head(spks)
```

```
##                    Sample_ID  SO4_ppm SO4_Area Cl_ppm Cl_Area   SO4_mM Cl_mM
## 1 TEMPEST_SW_H3_20230603_spk 168.9031   3.4046      0       0 5.268344     0
##   salinity Dilution
## 1  2.6e-05    56.51
```

```
#remove the dup from these IDs so we will have duplicate sample names
spks$Sample_ID<-gsub("_spk","",as.character(spks$Sample_ID))
spks <- spks[ ,c(1,6)]
colnames(spks) <- c('Sample_ID', 'SO4_mM_spk')
head(spks)
```

```
##                Sample_ID SO4_mM_spk
## 1 TEMPEST_SW_H3_20230603   5.268344
```

```
#put it back together with the old data set and look for duplicates
QAspks <- merge(sampledat_dils, spks)
head(QAspks)
```

```
##                Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM salinity
## 1 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 2 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 3 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 4 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 5 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 6 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
##   Dilution SO4_mM_spk
## 1    56.51   5.268344
## 2     5.15   5.268344
## 3    56.51   5.268344
## 4    56.51   5.268344
## 5     5.15   5.268344
## 6    56.51   5.268344
```

```
#now we need to calculate the spike concentration and calculate the spike recovery
spkconc <- (250/smw)    # in mM
spkvol <- 10             # in uL
spkvol <- spkvol/1000000
#spike for these samples was 10uL of the 250mM standard
QAspks$SO4_spk_Conc <- (spkconc)*spkvol         # mmoles of SO4
head(QAspks)
```

```
##                Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM salinity
## 1 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 2 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 3 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 4 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 5 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 6 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
##   Dilution SO4_mM_spk SO4_spk_Conc
## 1    56.51   5.268344 7.797879e-05
## 2     5.15   5.268344 7.797879e-05
## 3    56.51   5.268344 7.797879e-05
## 4    56.51   5.268344 7.797879e-05
## 5     5.15   5.268344 7.797879e-05
## 6    56.51   5.268344 7.797879e-05
```

```
#need to determine dilution factors and initial amount of sample added
#if your samples are all the same dilution just use the first line
#for Steph / COMPASS this depends on the site so...
# QAspks$Dilution <- 1
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "TEMPEST"), 50, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 50, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 50, QAspks$Dilution)
```

```
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 100, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 50, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 50, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 100, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 100, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 100, QAspks$Dilution)
# QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 200, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$Dilution)

#Set Sample volumes in uL
unique(QAspks$Dilution)
```

```
## [1] 56.51  5.15
```

```
QAspks$SampleVol <- 1
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 50, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 56.85, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 51.16, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 49.27, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 56.51, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 5.15, 1500, QAspks$SampleVol)

# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "TEMPEST"), 1483, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 1462, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$SampleVol)

#change sample volume to L
QAspks$SampleVol <- QAspks$SampleVol/1000000
head(QAspks)
```

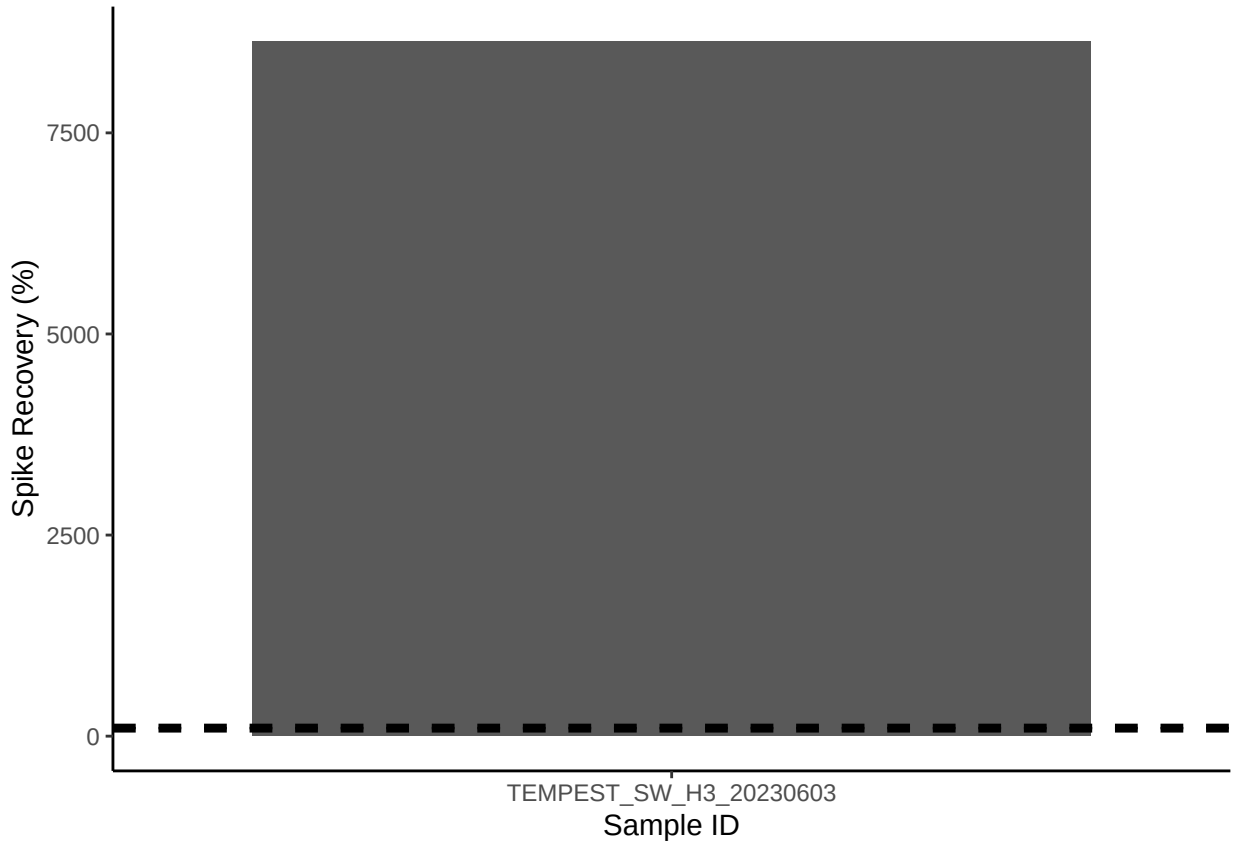
```
##                Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM salinity
## 1 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 2 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 3 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 4 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 5 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 6 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
##   Dilution SO4_mM_spk SO4_spk_Conc SampleVol
## 1    56.51   5.268344 7.797879e-05  0.001483
## 2     5.15   5.268344 7.797879e-05  0.001500
```

```
## 3    56.51   5.268344 7.797879e-05  0.001483
## 4    56.51   5.268344 7.797879e-05  0.001483
## 5     5.15   5.268344 7.797879e-05  0.001500
## 6    56.51   5.268344 7.797879e-05  0.001483
```

```
#gives us the total SO4 in the sample in mmoles
QAspks$SO4_Total_unspkd <- (QAspks$SO4_mM/QAspks$Dilution)*(QAspks$SampleVol)

##total SO4 in spiked sample in mmoles
QAspks$SO4_Total_spkd <- (QAspks$SO4_mM_spk/QAspks$Dilution)*(QAspks$SampleVol+spkvol)

QAspks$SO4_expctd_spkd <-  (QAspks$SO4_Total_unspkd + QAspks$SO4_spk_Conc)
QAspks$spk_recovery <-    (QAspks$SO4_Total_spkd/QAspks$SO4_expctd_spkd)*100
QAspks$SO4_spks_flag <-  ifelse(QAspks$spk_recovery <=120 & QAspks$spk_recovery >=80 , 'YES', 'NO, reru

head(QAspks)
```

```
##                Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM salinity
## 1 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 2 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 3 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 4 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 5 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
## 6 TEMPEST_SW_H3_20230603       0        0      0       0      0     0  2.6e-05
##   Dilution SO4_mM_spk SO4_spk_Conc SampleVol SO4_Total_unspkd SO4_Total_spkd
## 1    56.51   5.268344 7.797879e-05  0.001483                0   0.0001391902
## 2     5.15   5.268344 7.797879e-05  0.001500                0   0.0015446988
## 3    56.51   5.268344 7.797879e-05  0.001483                0   0.0001391902
## 4    56.51   5.268344 7.797879e-05  0.001483                0   0.0001391902
## 5     5.15   5.268344 7.797879e-05  0.001500                0   0.0015446988
## 6    56.51   5.268344 7.797879e-05  0.001483                0   0.0001391902
##   SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1    7.797879e-05     178.4975     NO, rerun
## 2    7.797879e-05    1980.9218     NO, rerun
## 3    7.797879e-05     178.4975     NO, rerun
## 4    7.797879e-05     178.4975     NO, rerun
## 5    7.797879e-05    1980.9218     NO, rerun
## 6    7.797879e-05     178.4975     NO, rerun
```

```
#plot spk recoveries output as a bar graph to easily check - want any over 10% to be red need to work o
spksbar <- ggplot(data = QAspks, aes(x = Sample_ID, y = spk_recovery)) +
        geom_bar(stat = 'identity') +
        scale_fill_manual(values=c("lightblue4")) +
        theme_classic() + labs(x= "Sample ID", y="Spike Recovery (%)") +
        theme(legend.position="none") +
        geom_hline(yintercept=80, linetype="dashed", color = "black", size=1) +
        geom_hline(yintercept=120, linetype="dashed", color = "black", size=1)

spksbar
```




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_spks <- QAspks %>%
  group_by(SO4_spks_flag) %>%
  summarise(no_rows = length(SO4_spks_flag))
Perc_spks$Total <- length(QAspks$SO4_spks_flag)
Perc_spks$Percent <- (Perc_spks$no_rows / Perc_spks$Total)*100
head(Perc_spks)
```

```
## # A tibble: 1 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 NO, rerun           8     8     100
```

Pull out spikes and check with area calc

```
##Merge sample data with dilution data
Dilutions <- read.csv("Raw Data/TMP 2023 Sample Dilutions from Dionex.csv")
sampledat_dils <- merge(sampledat, Dilutions, by = "Sample_ID")
unique(sampledat_dils$Dilution)
```

```
## [1]  1.10 50.00  5.15 56.51  4.53
```

```
#Show me the data that we have from the calculations
head(sampledat_dils)
```

```
##               Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 1 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 2 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 3 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 4 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 5 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 6 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
##     salinity Dilution
## 1 0.05525081     1.10
## 2 0.05525081    50.00
## 3 0.00002600     5.15
## 4 0.00002600    56.51
## 5 0.00002600     5.15
## 6 0.00002600    56.51
```

```
#now use area of all samples and calculate undilution corrected concentrations in ug/mL
sampledat_dils$SO4_ugmL <- ((sampledat_dils$SO4_Area)-SO4_Int)/SO4_Slope
sampledat_dils$Cl_ugmL <- (sampledat_dils$Cl_Area-Cl_Int)/Cl_Slope
head(sampledat_dils)
```

```
##               Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area     SO4_mM     Cl_mM
## 1 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 2 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616 38.1427 0.01005614 0.8621044
## 3 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 4 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 5 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
## 6 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000  0.0000 0.00000000 0.0000000
##     salinity Dilution    SO4_ugmL      Cl_ugmL
## 1 0.05525081     1.10 2.93060e-01 2.778329e+01
## 2 0.05525081    50.00 2.93060e-01 2.778329e+01
## 3 0.00002600     5.15 1.61704e-05 7.223654e-06
## 4 0.00002600    56.51 1.61704e-05 7.223654e-06
## 5 0.00002600     5.15 1.61704e-05 7.223654e-06
## 6 0.00002600    56.51 1.61704e-05 7.223654e-06
```

```
#pull out any rows that have "spk" in the SampleID column
spks <- sampledat_dils %>%
  filter(str_detect(Sample_ID, "spk"))      #have to change this to match data
head(spks)
```

```
##                    Sample_ID  SO4_ppm SO4_Area Cl_ppm Cl_Area   SO4_mM Cl_mM
## 1 TEMPEST_SW_H3_20230603_spk 168.9031   3.4046      0       0 5.268344     0
##   salinity Dilution SO4_ugmL      Cl_ugmL
## 1  2.6e-05    56.51 2.988922 7.223654e-06
```

```
## Only keep the columns that we need
spks <- spks[ ,c(1,9,10)]
head(spks)
```

```
##                    Sample_ID Dilution SO4_ugmL
## 1 TEMPEST_SW_H3_20230603_spk    56.51 2.988922
```

```
#remove the spk from these IDs so we will have duplicate sample names
spks$Sample_ID<-gsub("_spk","",as.character(spks$Sample_ID))
colnames(spks) <- c('Sample_ID', 'Dilution', 'SO4_ugmL_spk')
head(spks)
```

```
##                Sample_ID Dilution SO4_ugmL_spk
## 1 TEMPEST_SW_H3_20230603    56.51     2.988922
```

```
#put it back together with the old data set and look for duplicates
QAspks <- merge(sampledat_dils, spks)
head(QAspks)
```

```
##                Sample_ID Dilution SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM
## 1 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 2 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 3 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 4 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
##   salinity    SO4_ugmL      Cl_ugmL SO4_ugmL_spk
## 1  2.6e-05 1.61704e-05 7.223654e-06     2.988922
## 2  2.6e-05 1.61704e-05 7.223654e-06     2.988922
## 3  2.6e-05 1.61704e-05 7.223654e-06     2.988922
## 4  2.6e-05 1.61704e-05 7.223654e-06     2.988922
```

```
#now we need to calculate the spike concentration and calculate the spike recovery
spkconc <- (250)    # in ug
spkvol <- 10             # in uL
spkvol <- spkvol/1000
#spike for these samples was 10uL of the 250mM standard
QAspks$SO4_spk_Conc <- (spkconc)*spkvol         # mmoles of SO4
head(QAspks)
```

```
##                Sample_ID Dilution SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM
## 1 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 2 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 3 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 4 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
##   salinity    SO4_ugmL      Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5
## 2  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5
## 3  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5
## 4  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5
```

```
#need to determine dilution factors and initial amount of sample added
#if your samples are all the same dilution just use the first line
#for Steph / COMPASS this depends on the site so...
#QAspks$Dilution <- 1
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 50, QAspks$Dilution)
#$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 50, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 100, QAspks$Dilution)
```

```
#$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 50, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 50, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 100, QAspks$Dilution)
#$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 200, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$Dilution)

#Set Sample volumes in uL
QAspks$SampleVol <- 1
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 50, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.1, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 5.15, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 4.53, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 56.51, 1483, QAspks$SampleVol)

# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "TEMPEST"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 1462, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$SampleVol)

#change sample volume to L
QAspks$SampleVol <- QAspks$SampleVol/1000
head(QAspks)
```

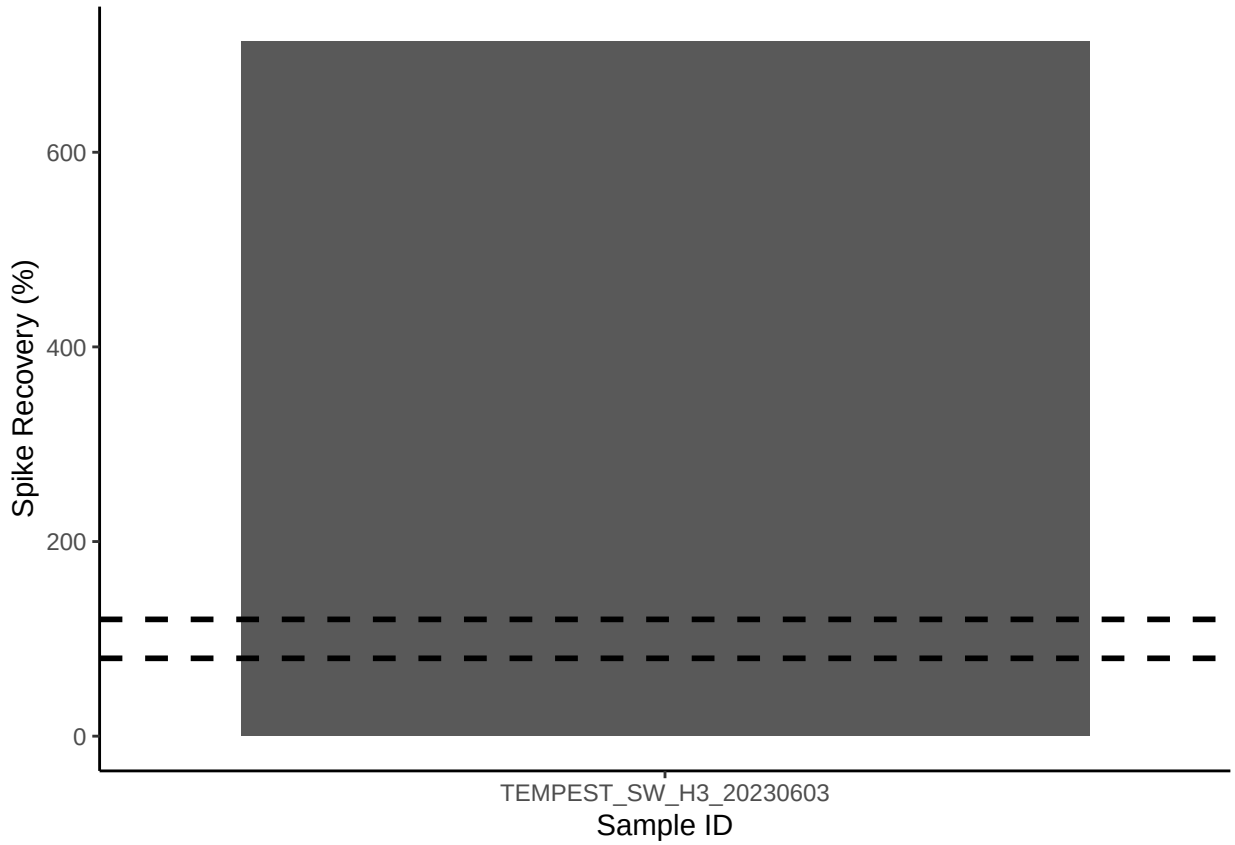
```
##               Sample_ID Dilution SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM
## 1 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 2 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 3 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 4 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
##   salinity    SO4_ugmL      Cl_ugmL SO4_ugmL_spk SO4_spk_Conc SampleVol
## 1  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
## 2  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
## 3  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
## 4  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
```

```
#gives us the total SO4 in the sample in mmoles
QAspks$SO4_Total_unspkd <- QAspks$SO4_ugmL*QAspks$SampleVol
```

```
##total SO4 in spiked sample in mmoles
QAspks$SO4_Total_spkd <- (QAspks$SO4_ugmL_spk)*(QAspks$SampleVol+spkvol)

QAspks$SO4_expctd_spkd <-  (QAspks$SO4_Total_unspkd + QAspks$SO4_spk_Conc)
QAspks$spk_recovery <-    (QAspks$SO4_Total_spkd/QAspks$SO4_expctd_spkd)*100
QAspks$SO4_spks_flag <-  ifelse(QAspks$spk_recovery <=120 & QAspks$spk_recovery >=80 , 'YES', 'NO, reru

head(QAspks)
```

```
##                Sample_ID Dilution SO4_ppm SO4_Area Cl_ppm Cl_Area SO4_mM Cl_mM
## 1 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 2 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 3 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
## 4 TEMPEST_SW_H3_20230603    56.51       0        0      0       0      0     0
##   salinity    SO4_ugmL      Cl_ugmL SO4_ugmL_spk SO4_spk_Conc SampleVol
## 1  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
## 2  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
## 3  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
## 4  2.6e-05 1.61704e-05 7.223654e-06     2.988922          2.5     1.483
##   SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1     2.398071e-05       4.462461        2.500024     178.4967     NO, rerun
## 2     2.398071e-05       4.462461        2.500024     178.4967     NO, rerun
## 3     2.398071e-05       4.462461        2.500024     178.4967     NO, rerun
## 4     2.398071e-05       4.462461        2.500024     178.4967     NO, rerun
```

```
#plot spk recoveries output as a bar graph to easily check - want any over 10% to be red need to work o
spksbar <- ggplot(data = QAspks, aes(x = Sample_ID, y = spk_recovery)) +
        geom_bar(stat = 'identity') +
        scale_fill_manual(values=c("lightblue4")) +
        theme_classic() + labs(x= "Sample ID", y="Spike Recovery (%)") +
        theme(legend.position="none") +
        geom_hline(yintercept=80, linetype="dashed", color = "black", size=1) +
        geom_hline(yintercept=120, linetype="dashed", color = "black", size=1)

spksbar
```




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_spks <- QAspks %>%
  group_by(SO4_spks_flag) %>%
  summarise(no_rows = length(SO4_spks_flag))
Perc_spks$Total <- length(QAspks$SO4_spks_flag)
Perc_spks$Percent <- (Perc_spks$no_rows / Perc_spks$Total)*100
head(Perc_spks)
```

```
## # A tibble: 1 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 NO, rerun           4     4     100
```

Make final dataframe with IDs

```
#for steph <- pull out identifiers of the sample names
#pull the sample ID and separate it by the underscores
IDs <- data.frame(do.call('rbind', strsplit(as.character(sampledat2$Sample_ID),'_',fixed=TRUE)))
colnames(IDs) <- c("Project" , "Plot","Grid", "Date")
head(IDs)
```

```
##   Project Plot Grid     Date
## 1 TEMPEST    C   C6 20230515
```

```
## 2 TEMPEST    C   C6 20230603
## 3 TEMPEST    C   C6 20230603
## 4 TEMPEST    C   C6 20230603
## 5 TEMPEST    C   C6 20230603
## 6 TEMPEST    C   F6 20230515
```

```
#rejoin them to the dataframe
alldat <- cbind(IDs, sampledat2)
head(alldat)
```

```
##   Project Plot Grid     Date             Sample_ID SO4_ppm SO4_Area  Cl_ppm
## 1 TEMPEST    C   C6 20230515 TEMPEST_C_C6_20230515  0.3224   0.3338 30.5616
## 2 TEMPEST    C   C6 20230603 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000
## 3 TEMPEST    C   C6 20230603 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000
## 4 TEMPEST    C   C6 20230603 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000
## 5 TEMPEST    C   C6 20230603 TEMPEST_C_C6_20230603  0.0000   0.0000  0.0000
## 6 TEMPEST    C   F6 20230515 TEMPEST_C_F6_20230515  0.0000   0.0000 28.2531
##   Cl_Area     SO4_mM     Cl_mM   salinity
## 1 38.1427 0.01005614 0.8621044 0.05525081
## 2  0.0000 0.00000000 0.0000000 0.00002600
## 3  0.0000 0.00000000 0.0000000 0.00002600
## 4  0.0000 0.00000000 0.0000000 0.00002600
## 5  0.0000 0.00000000 0.0000000 0.00002600
## 6 35.2615 0.00000000 0.7969845 0.05107935
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_TEMPEST_Dionex_ClSo4_Final_20230606.csv")          #Cha
```

END