Dionex_COMPASS_TEMPEST

Stephanie J. Wilson

2023-06-08

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_20230608_SO4.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 1  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 2 2  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 3 3  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 4 4 Standard 1 Calibration Standard 4.343   0.4565     6.17   0.5324     2.53
## 5 5 Standard 2 Calibration Standard 4.343   0.9668     6.58   1.1275     5.32
## 6 6 Standard 3 Calibration Standard 4.330   1.9262     7.25   2.2463    10.35
##   IC.SO4.5
## 1     n.a.
## 2     n.a.
## 3     n.a.
## 4       MB
## 5       MB
## 6       M
```

```
## 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  Lab Blank     n.a.     n.a.
## 2  Lab Blank     n.a.     n.a.
## 3  Lab Blank     n.a.     n.a.
## 4 Standard 1  0.4565   0.5324
## 5 Standard 2  0.9668   1.1275
## 6 Standard 3  1.9262   2.2463
```

```r
## 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
```

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

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

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

```
##    Sample_ID SO4_ppm SO4_Area
## 1  Lab Blank      NA       NA
## 2  Lab Blank      NA       NA
## 3  Lab Blank      NA       NA
## 4 Standard 1  0.4565   0.5324
## 5 Standard 2  0.9668   1.1275
## 6 Standard 3  1.9262   2.2463
```

```r
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_TEMPEST_20230608_Cl.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 1  Lab Blank             Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 2 2  Lab Blank             Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 3 3  Lab Blank             Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 4 4 Standard 1 Calibration Standard 3.417   5.7300   93.70   8.0833   45.41
## 5 5 Standard 2 Calibration Standard 3.417  11.3295   93.26  15.9825   88.56
## 6 6 Standard 3 Calibration Standard 3.413  20.3441   92.64  28.6992  159.19
##   IC.Cl.5
## 1    n.a.
## 2    n.a.
## 3    n.a.
## 4     BM
## 5      M
## 6     BM
```

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

```
##          X.1  IC.Cl.1  IC.Cl.3
## 1  Lab Blank     n.a.     n.a.
## 2  Lab Blank     n.a.     n.a.
## 3  Lab Blank     n.a.     n.a.
## 4 Standard 1  5.7300   8.0833
## 5 Standard 2 11.3295  15.9825
## 6 Standard 3 20.3441  28.6992
```

```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  Lab Blank      NA      NA
## 2  Lab Blank      NA      NA
## 3  Lab Blank      NA      NA
## 4 Standard 1  5.7300  8.0833
## 5 Standard 2 11.3295 15.9825
## 6 Standard 3 20.3441 28.6992
```

```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           581.2849  23.7212 4345.0138 286.6714
## 2           581.2849  23.7212 1400.2957  69.8002
## 3           581.2849  23.7212    0.0058   0.0082
## 4           581.2849  23.7212 1605.3399  61.2697
## 5           581.2849  23.7212        NA       NA
## 6           202.4819   8.0517 4345.0138 286.6714
```

```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     NA      NA
## 28 Lab Blank      NA       NA     NA      NA
## 29 Lab Blank      NA       NA     NA      NA
## 30 Lab Blank      NA       NA 0.0196  0.0277
## 31 Lab Blank      NA       NA 0.0100  0.0141
```

```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
## 170 Standard 1  0.4565   0.5324 5.7300  8.0833
## 171 Standard 1  0.4565   0.5324 5.8701  8.2810
## 172 Standard 1  0.4565   0.5324 5.7836  8.1589
## 173 Standard 1  0.4826   0.5628 5.7300  8.0833
## 174 Standard 1  0.4826   0.5628 5.8701  8.2810
## 175 Standard 1  0.4826   0.5628 5.7836  8.1589
```

```
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
## 170 Standard 1  0.4565   0.5324 5.7300  8.0833
## 171 Standard 1  0.4565   0.5324 5.8701  8.2810
## 172 Standard 1  0.4565   0.5324 5.7836  8.1589
## 173 Standard 1  0.4826   0.5628 5.7300  8.0833
## 174 Standard 1  0.4826   0.5628 5.8701  8.2810
## 175 Standard 1  0.4826   0.5628 5.7836  8.1589
```

```
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  0.472 0.0118  2.50  NO, rerun
## 2 Standard 2  0.985 0.0218  2.21  NO, rerun
## 3 Standard 3  1.92  0.00999 0.520 YES
## 4 Standard 4 10.1   0.121   1.21  YES
## 5 Standard 5 20.2   0.183   0.904 YES
```

```
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   5.79 0.0612 1.06  YES
```

```
## 2 Standard 2  11.4  0.0937 0.819 YES
## 3 Standard 3  20.5  0.159  0.777 YES
## 4 Standard 4 101.   1.14   1.12  YES
## 5 Standard 5 201.   1.98   0.982 YES
```

```r
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

```r
#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
## 205 TEMPEST_ESTUARY_SOURCE_20230607_0710 581.2849  13.5575 4335.7195 122.3274
## 206 TEMPEST_ESTUARY_SOURCE_20230607_1346 547.2306  12.7632 4145.0144 116.9468
## 207 TEMPEST_ESTUARY_SOURCE_20230607_1700 553.9434  12.9198 4153.6146 117.1895
## 208          TEMPEST_FW_B4_20230607_0800  11.1494  11.8200   27.9934  35.9001
## 209          TEMPEST_FW_B4_20230607_1600   2.4624   2.2973  206.4674 233.0096
## 210          TEMPEST_FW_C3_20230607_0800  10.7750  11.4231   23.9976  30.7756
```

```r
# 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
```

```
## 205 TEMPEST_ESTUARY_SOURCE_20230607_0710 581.2849  13.5575 4335.7195 122.3274
## 206 TEMPEST_ESTUARY_SOURCE_20230607_1346 547.2306  12.7632 4145.0144 116.9468
## 207 TEMPEST_ESTUARY_SOURCE_20230607_1700 553.9434  12.9198 4153.6146 117.1895
## 208          TEMPEST_FW_B4_20230607_0800  11.1494  11.8200   27.9934  35.9001
## 209          TEMPEST_FW_B4_20230607_1600   2.4624   2.2973  206.4674 233.0096
## 210          TEMPEST_FW_C3_20230607_0800  10.7750  11.4231   23.9976  30.7756
##          SO4_mM       Cl_mM   salinity
## 205 18.13115721 122.3052045 7.83467114
## 206 17.06895197 116.9256530 7.49006702
## 207 17.27833437 117.1682539 7.50560758
## 208  0.34776669   0.7896587 0.05061007
## 209  0.07680599   5.8241862 0.37311259
## 210  0.33608858   0.6769422 0.04338966
```

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
## 205 TEMPEST_ESTUARY_SOURCE_20230607_0710 581.2849  13.5575 4335.7195 122.3274
## 206 TEMPEST_ESTUARY_SOURCE_20230607_1346 547.2306  12.7632 4145.0144 116.9468
## 207 TEMPEST_ESTUARY_SOURCE_20230607_1700 553.9434  12.9198 4153.6146 117.1895
## 208          TEMPEST_FW_B4_20230607_0800  11.1494  11.8200   27.9934  35.9001
## 209          TEMPEST_FW_B4_20230607_1600   2.4624   2.2973  206.4674 233.0096
## 210          TEMPEST_FW_C3_20230607_0800  10.7750  11.4231   23.9976  30.7756
##          SO4_mM       Cl_mM   salinity
## 205 18.13115721 122.3052045 7.83467114
## 206 17.06895197 116.9256530 7.49006702
## 207 17.27833437 117.1682539 7.50560758
## 208  0.34776669   0.7896587 0.05061007
## 209  0.07680599   5.8241862 0.37311259
## 210  0.33608858   0.6769422 0.04338966
```

```
#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
## 1 TEMPEST_FW_SOURCE_20230607_1519_dup 574.2564  13.3935    0.000  0.0000
## 2     TEMPEST_SW_B4_20230607_0800_dup 403.8132   9.4182 3303.890 93.2155
## 3     TEMPEST_SW_B4_20230607_1600_dup 440.2724  10.2686 3510.899 99.0560
## 4     TEMPEST_SW_D5_20230607_0800_dup 292.2655   6.8166 2343.111 66.1082
## 5     TEMPEST_SW_I5_20230607_0800_dup 273.6440   6.3823 2744.173 77.4237
##      SO4_mM    Cl_mM salinity
## 1 17.911928  0.00000 0.000026
## 2 12.595546 93.19858 5.970155
## 3 13.732764 99.03805 6.344220
## 4  9.116204 66.09623 4.234028
## 5  8.535371 77.40967 4.958746
```

```r
#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
## 1 TEMPEST_ESTUARY_SOURCE_20230607_0710 581.2849  13.5575 4335.7195 122.3274
## 2 TEMPEST_ESTUARY_SOURCE_20230607_1346 547.2306  12.7632 4145.0144 116.9468
## 3 TEMPEST_ESTUARY_SOURCE_20230607_1700 553.9434  12.9198 4153.6146 117.1895
## 4          TEMPEST_FW_B4_20230607_0800  11.1494  11.8200   27.9934  35.9001
## 5          TEMPEST_FW_B4_20230607_1600   2.4624   2.2973  206.4674 233.0096
## 6          TEMPEST_FW_C3_20230607_0800  10.7750  11.4231   23.9976  30.7756
##        SO4_mM       Cl_mM   salinity
## 1 18.13115721 122.3052045 7.83467114
## 2 17.06895197 116.9256530 7.49006702
## 3 17.27833437 117.1682539 7.50560758
## 4  0.34776669   0.7896587 0.05061007
## 5  0.07680599   5.8241862 0.37311259
## 6  0.33608858   0.6769422 0.04338966
```

```r
#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_FW_SOURCE_20230607_1519  17.911928   0.00000     0.000026
## 2     TEMPEST_SW_B4_20230607_0800  12.595546  93.19858     5.970155
## 3     TEMPEST_SW_B4_20230607_1600  13.732764  99.03805     6.344220
## 4     TEMPEST_SW_D5_20230607_0800   9.116204  66.09623     4.234028
## 5     TEMPEST_SW_I5_20230607_0800   8.535371  77.40967     4.958746
```

```r
#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
## 1 TEMPEST_FW_SOURCE_20230607_1519 581.1936  13.5553    0.000  0.0000 18.128309
## 2     TEMPEST_SW_B4_20230607_0800 377.6524   8.8081 3087.801 87.1188 11.779551
## 3     TEMPEST_SW_B4_20230607_1600 440.3089  10.2694 3507.065 98.9478 13.733902
## 4     TEMPEST_SW_D5_20230607_0800 293.4522   6.8443 2353.532 66.4022  9.153219
## 5     TEMPEST_SW_I5_20230607_0800 284.5868   6.6375 2842.504 80.1980  8.876694
##      Cl_mM salinity SO4_mM_dup Cl_mM_dup salinity_dup
## 1  0.00000 0.000026  17.911928   0.00000     0.000026
## 2 87.10298 5.579682  12.595546  93.19858     5.970155
## 3 98.92990 6.337292  13.732764  99.03805     6.344220
## 4 66.39019 4.252859   9.116204  66.09623     4.234028
## 5 80.18346 5.136430   8.535371  77.40967     4.958746
```

```
QAdups$SO4_dups_chk <- ((abs(QAdups$SO4_mM-QAdups$SO4_mM_dup))/((QAdups$SO4_mM+QAdups$SO4_mM_dup)/2))*10
QAdups$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)
```

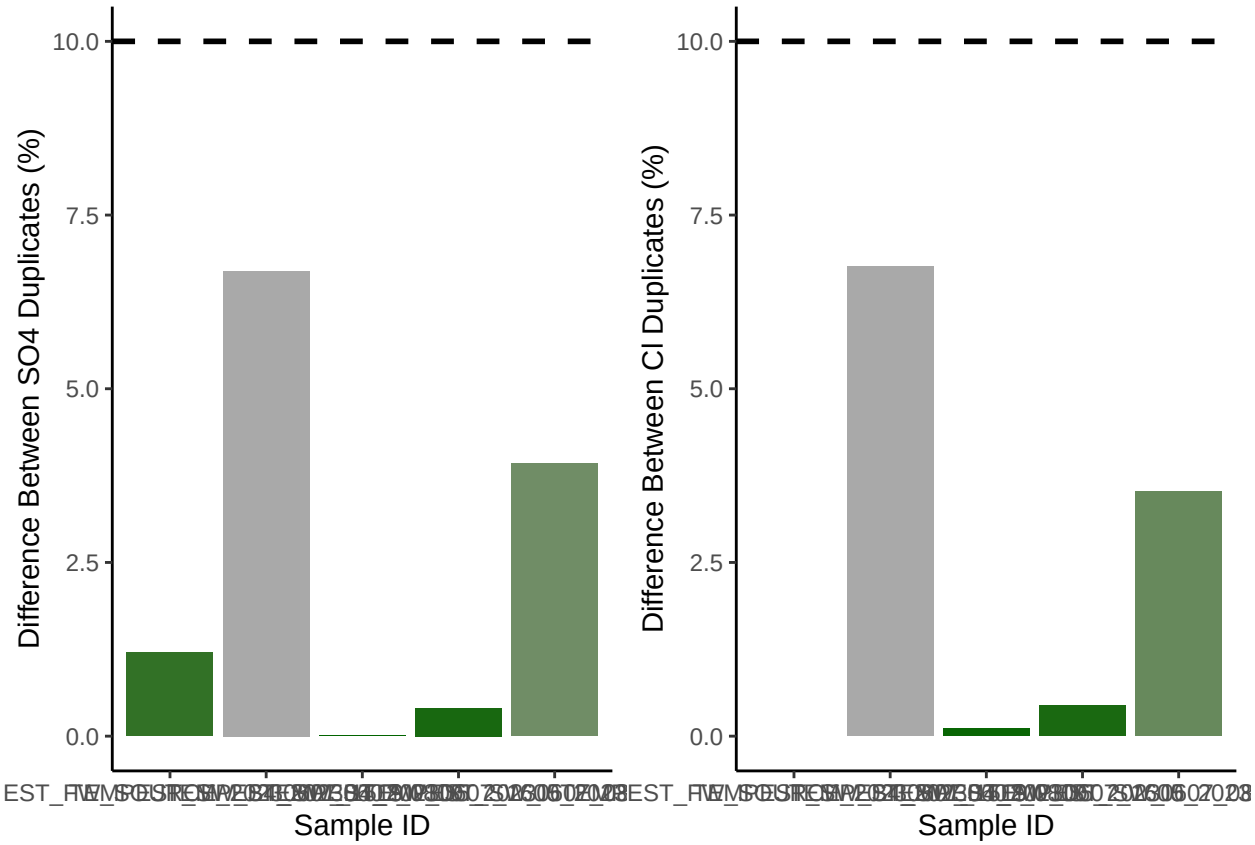
```
##                         Sample_ID  SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1 TEMPEST_FW_SOURCE_20230607_1519 581.1936  13.5553    0.000  0.0000 18.128309
## 2     TEMPEST_SW_B4_20230607_0800 377.6524   8.8081 3087.801 87.1188 11.779551
## 3     TEMPEST_SW_B4_20230607_1600 440.3089  10.2694 3507.065 98.9478 13.733902
## 4     TEMPEST_SW_D5_20230607_0800 293.4522   6.8443 2353.532 66.4022  9.153219
## 5     TEMPEST_SW_I5_20230607_0800 284.5868   6.6375 2842.504 80.1980  8.876694
##      Cl_mM salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk
## 1  0.00000 0.000026  17.911928   0.00000     0.000026  1.200778917
## 2 87.10298 5.579682  12.595546  93.19858     5.970155  6.695317107
## 3 98.92990 6.337292  13.732764  99.03805     6.344220  0.008289978
## 4 66.39019 4.252859   9.116204  66.09623     4.234028  0.405212272
## 5 80.18346 5.136430   8.535371  77.40967     4.958746  3.920528928
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1           YES         NaN         <NA>
## 2           YES   6.7615540          YES
## 3           YES   0.1092568          YES
## 4           YES   0.4437552          YES
## 5           YES   3.5201930          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_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 1 row 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_dups1 <- QAdups %>%
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(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         5  YES          4     5       100         80
## 2  YES         5 <NA>          1     5       100         20
```

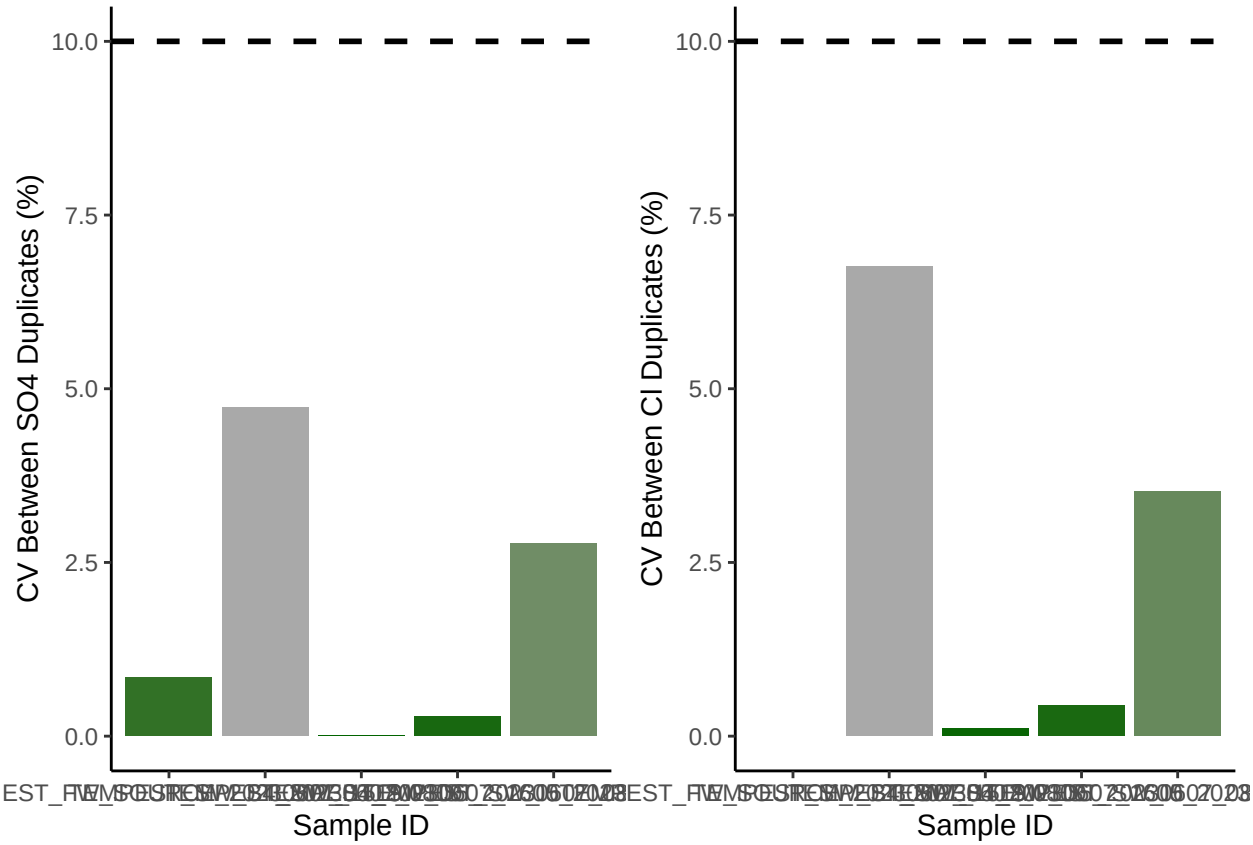
Pull out dups and check with cv

```r
#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
## 1 TEMPEST_FW_SOURCE_20230607_1519 581.1936  13.5553    0.000  0.0000 18.128309
## 2     TEMPEST_SW_B4_20230607_0800 377.6524   8.8081 3087.801 87.1188 11.779551
## 3     TEMPEST_SW_B4_20230607_1600 440.3089  10.2694 3507.065 98.9478 13.733902
## 4     TEMPEST_SW_D5_20230607_0800 293.4522   6.8443 2353.532 66.4022  9.153219
## 5     TEMPEST_SW_I5_20230607_0800 284.5868   6.6375 2842.504 80.1980  8.876694
##      Cl_mM salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk
## 1  0.00000 0.000026  17.911928   0.00000     0.000026  1.200778917
## 2 87.10298 5.579682  12.595546  93.19858     5.970155  6.695317107
## 3 98.92990 6.337292  13.732764  99.03805     6.344220  0.008289978
## 4 66.39019 4.252859   9.116204  66.09623     4.234028  0.405212272
## 5 80.18346 5.136430   8.535371  77.40967     4.958746  3.920528928
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1           YES         NaN         <NA>   0.8490789              YES
## 2           YES   6.7615540          YES   4.7343041              YES
## 3           YES   0.1092568          YES   0.0058619              YES
## 4           YES   0.4437552          YES   0.2865283              YES
## 5           YES   3.5201930          YES   2.7722326              YES
```

```r
#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 1 row 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_dups1 <- QAdups %>%
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(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         5  YES          4     5       100         80
## 2  YES         5 <NA>          1     5       100         20
```

Pull out spikes and check with dionex output conc.

```r
##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] 50.00  1.10  1.25 56.51 52.74
```

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

```
##                              Sample_ID  SO4_ppm SO4_Area    Cl_ppm  Cl_Area
## 1 TEMPEST_ESTUARY_SOURCE_20230607_0710 581.2849  13.5575 4335.7195 122.3274
## 2 TEMPEST_ESTUARY_SOURCE_20230607_1346 547.2306  12.7632 4145.0144 116.9468
## 3 TEMPEST_ESTUARY_SOURCE_20230607_1700 553.9434  12.9198 4153.6146 117.1895
## 4          TEMPEST_FW_B4_20230607_0800  11.1494  11.8200   27.9934  35.9001
## 5          TEMPEST_FW_B4_20230607_1600   2.4624   2.2973  206.4674 233.0096
## 6          TEMPEST_FW_C3_20230607_0800  10.7750  11.4231   23.9976  30.7756
##        SO4_mM       Cl_mM   salinity Dilution
## 1 18.13115721 122.3052045 7.83467114    50.00
## 2 17.06895197 116.9256530 7.49006702    50.00
## 3 17.27833437 117.1682539 7.50560758    50.00
## 4  0.34776669   0.7896587 0.05061007     1.10
## 5  0.07680599   5.8241862 0.37311259     1.25
## 6  0.33608858   0.6769422 0.04338966     1.10
```

```r
#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
## 1 TEMPEST_FW_SOURCE_20230607_0611_spk  16.0473  17.0125   23.1723  29.7173
## 2     TEMPEST_SW_E3_20230607_1600_spk 332.1500   7.7468 2355.1999  66.4493
## 3     TEMPEST_SW_I5_20230607_1600_spk 468.0219  10.9158 3179.1880  89.6971
## 4 TEMPEST_SW_SOURCE_20230607_0621_spk 563.1599  13.1347 4002.2094 112.9177
## 5  TEMPEST_SW_SOURCE_20230607_1515spk 535.8099  12.4968 4345.0138 122.5896
##       SO4_mM       Cl_mM   salinity Dilution
## 1  0.5005396   0.6536615 0.04189835      1.1
## 2 10.3602620  66.4372327 4.25587222     50.0
## 3 14.5983125  89.6809027 5.74481872     50.0
## 4 17.5658110 112.8973032 7.23201839     50.0
## 5 16.7127230 122.5673850 7.85146594     50.0
```

```r
#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_FW_SOURCE_20230607_0611  0.5005396
## 2        TEMPEST_SW_E3_20230607_1600 10.3602620
## 3        TEMPEST_SW_I5_20230607_1600 14.5983125
## 4    TEMPEST_SW_SOURCE_20230607_0621 17.5658110
## 5 TEMPEST_SW_SOURCE_20230607_1515spk 16.7127230
```

```
#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
## 1    TEMPEST_FW_SOURCE_20230607_0611  12.7760  13.5445   22.8178  29.2627
## 2        TEMPEST_SW_E3_20230607_1600 259.0745   6.0425 2327.3879  65.6646
## 3        TEMPEST_SW_I5_20230607_1600 336.2264   7.8419 3120.4635  88.0403
## 4    TEMPEST_SW_SOURCE_20230607_0621 571.0464  13.3187 4300.3781 121.3302
## 5 TEMPEST_SW_SOURCE_20230607_1515spk 535.8099  12.4968 4345.0138 122.5896
##       SO4_mM       Cl_mM   salinity Dilution SO4_mM_spk
## 1  0.3985028   0.6436615 0.04125776      1.1  0.5005396
## 2  8.0809264  65.6526911 4.20561594     50.0 10.3602620
## 3 10.4874111  88.0243583 5.63870354     50.0 14.5983125
## 4 17.8118029 121.3082680 7.77080923     50.0 17.5658110
## 5 16.7127230 122.5673850 7.85146594     50.0 16.7127230
```

```
#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
## 1    TEMPEST_FW_SOURCE_20230607_0611  12.7760  13.5445   22.8178  29.2627
## 2        TEMPEST_SW_E3_20230607_1600 259.0745   6.0425 2327.3879  65.6646
## 3        TEMPEST_SW_I5_20230607_1600 336.2264   7.8419 3120.4635  88.0403
## 4    TEMPEST_SW_SOURCE_20230607_0621 571.0464  13.3187 4300.3781 121.3302
## 5 TEMPEST_SW_SOURCE_20230607_1515spk 535.8099  12.4968 4345.0138 122.5896
##       SO4_mM       Cl_mM   salinity Dilution SO4_mM_spk SO4_spk_Conc
## 1  0.3985028   0.6436615 0.04125776      1.1  0.5005396 7.797879e-05
## 2  8.0809264  65.6526911 4.20561594     50.0 10.3602620 7.797879e-05
## 3 10.4874111  88.0243583 5.63870354     50.0 14.5983125 7.797879e-05
## 4 17.8118029 121.3082680 7.77080923     50.0 17.5658110 7.797879e-05
## 5 16.7127230 122.5673850 7.85146594     50.0 16.7127230 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]  1.1 50.0
```

```
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(QAspks$Dilution == 1.25, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.1, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 52.74, 1483, QAspks$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)
```

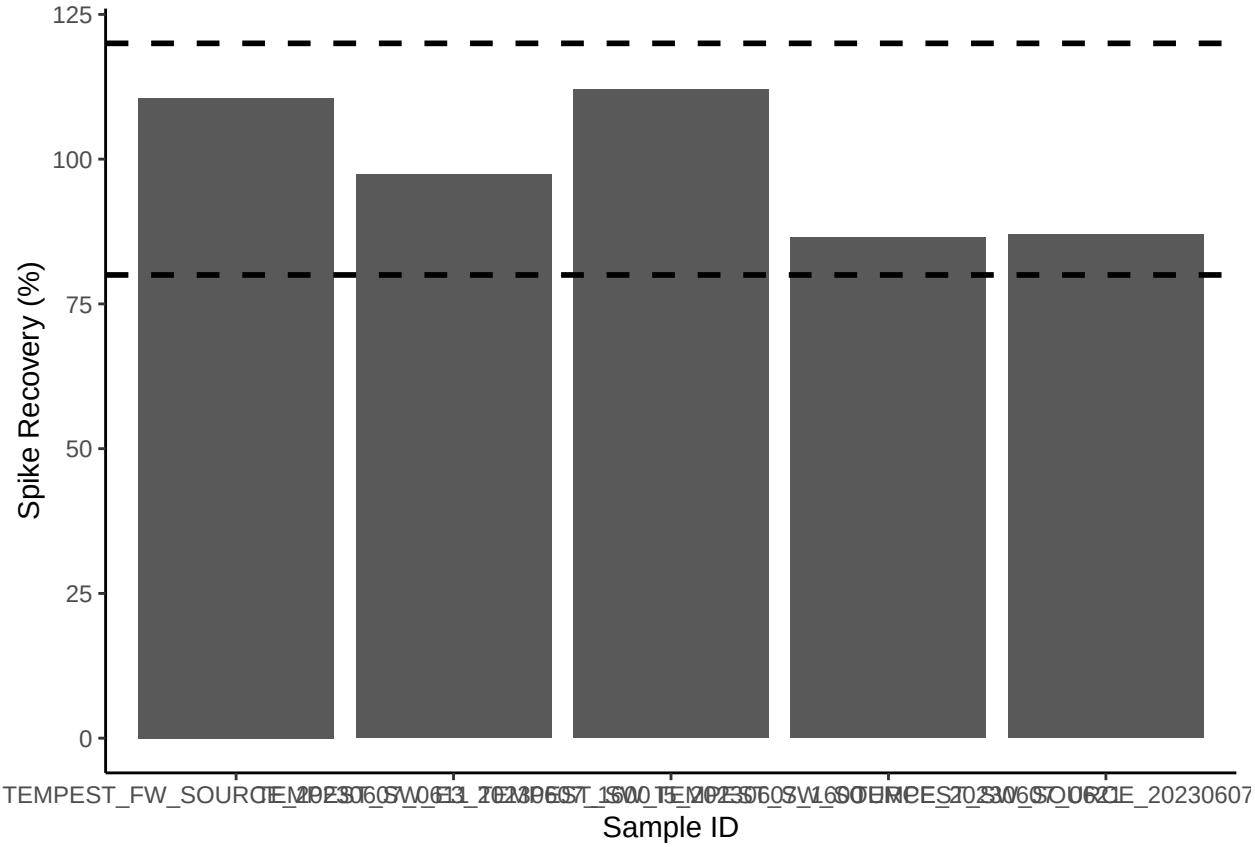
```
##                            Sample_ID  SO4_ppm SO4_Area    Cl_ppm  Cl_Area
## 1    TEMPEST_FW_SOURCE_20230607_0611  12.7760  13.5445   22.8178  29.2627
## 2        TEMPEST_SW_E3_20230607_1600 259.0745   6.0425 2327.3879  65.6646
## 3        TEMPEST_SW_I5_20230607_1600 336.2264   7.8419 3120.4635  88.0403
## 4    TEMPEST_SW_SOURCE_20230607_0621 571.0464  13.3187 4300.3781 121.3302
## 5 TEMPEST_SW_SOURCE_20230607_1515spk 535.8099  12.4968 4345.0138 122.5896
##       SO4_mM       Cl_mM   salinity Dilution SO4_mM_spk SO4_spk_Conc SampleVol
## 1  0.3985028   0.6436615 0.04125776      1.1  0.5005396 7.797879e-05  0.001500
```

```
## 2  8.0809264  65.6526911 4.20561594     50.0 10.3602620 7.797879e-05  0.001483
## 3 10.4874111  88.0243583 5.63870354     50.0 14.5983125 7.797879e-05  0.001483
## 4 17.8118029 121.3082680 7.77080923     50.0 17.5658110 7.797879e-05  0.001483
## 5 16.7127230 122.5673850 7.85146594     50.0 16.7127230 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, rerun

head(QAspks)
```

```
##                           Sample_ID  SO4_ppm SO4_Area    Cl_ppm  Cl_Area
## 1    TEMPEST_FW_SOURCE_20230607_0611  12.7760  13.5445   22.8178  29.2627
## 2        TEMPEST_SW_E3_20230607_1600 259.0745   6.0425 2327.3879  65.6646
## 3        TEMPEST_SW_I5_20230607_1600 336.2264   7.8419 3120.4635  88.0403
## 4    TEMPEST_SW_SOURCE_20230607_0621 571.0464  13.3187 4300.3781 121.3302
## 5 TEMPEST_SW_SOURCE_20230607_1515spk 535.8099  12.4968 4345.0138 122.5896
##       SO4_mM       Cl_mM   salinity Dilution SO4_mM_spk SO4_spk_Conc SampleVol
## 1  0.3985028   0.6436615 0.04125776      1.1  0.5005396 7.797879e-05  0.001500
## 2  8.0809264  65.6526911 4.20561594     50.0 10.3602620 7.797879e-05  0.001483
## 3 10.4874111  88.0243583 5.63870354     50.0 14.5983125 7.797879e-05  0.001483
## 4 17.8118029 121.3082680 7.77080923     50.0 17.5658110 7.797879e-05  0.001483
## 5 16.7127230 122.5673850 7.85146594     50.0 16.7127230 7.797879e-05  0.001483
##   SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1     0.0005434129   0.0006871044    0.0006213917    110.57508           YES
## 2     0.0002396803   0.0003093574    0.0003176591     97.38662           YES
## 3     0.0003110566   0.0004359056    0.0003890354    112.04780           YES
## 4     0.0005282981   0.0005245151    0.0006062769     86.51412           YES
## 5     0.0004956994   0.0004990419    0.0005736782     86.98987           YES
```

```
#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 YES                 5     5     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] 50.00  1.10  1.25 56.51 52.74
```

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

```
##                              Sample_ID  SO4_ppm SO4_Area    Cl_ppm  Cl_Area
## 1 TEMPEST_ESTUARY_SOURCE_20230607_0710 581.2849  13.5575 4335.7195 122.3274
## 2 TEMPEST_ESTUARY_SOURCE_20230607_1346 547.2306  12.7632 4145.0144 116.9468
## 3 TEMPEST_ESTUARY_SOURCE_20230607_1700 553.9434  12.9198 4153.6146 117.1895
## 4          TEMPEST_FW_B4_20230607_0800  11.1494  11.8200   27.9934  35.9001
## 5          TEMPEST_FW_B4_20230607_1600   2.4624   2.2973  206.4674 233.0096
## 6          TEMPEST_FW_C3_20230607_0800  10.7750  11.4231   23.9976  30.7756
##        SO4_mM       Cl_mM   salinity Dilution
## 1 18.13115721 122.3052045 7.83467114    50.00
## 2 17.06895197 116.9256530 7.49006702    50.00
## 3 17.27833437 117.1682539 7.50560758    50.00
## 4  0.34776669   0.7896587 0.05061007     1.10
## 5  0.07680599   5.8241862 0.37311259     1.25
## 6  0.33608858   0.6769422 0.04338966     1.10
```

```r
#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
## 1 TEMPEST_ESTUARY_SOURCE_20230607_0710 581.2849  13.5575 4335.7195 122.3274
## 2 TEMPEST_ESTUARY_SOURCE_20230607_1346 547.2306  12.7632 4145.0144 116.9468
## 3 TEMPEST_ESTUARY_SOURCE_20230607_1700 553.9434  12.9198 4153.6146 117.1895
## 4          TEMPEST_FW_B4_20230607_0800  11.1494  11.8200   27.9934  35.9001
## 5          TEMPEST_FW_B4_20230607_1600   2.4624   2.2973  206.4674 233.0096
## 6          TEMPEST_FW_C3_20230607_0800  10.7750  11.4231   23.9976  30.7756
##        SO4_mM       Cl_mM   salinity Dilution  SO4_ugmL   Cl_ugmL
## 1 18.13115721 122.3052045 7.83467114    50.00 11.625713  86.71442
## 2 17.06895197 116.9256530 7.49006702    50.00 10.944591  82.90027
## 3 17.27833437 117.1682539 7.50560758    50.00 11.078877  83.07231
## 4  0.34776669   0.7896587 0.05061007     1.10 10.135786  25.44856
## 5  0.07680599   5.8241862 0.37311259     1.25  1.969955 165.17390
## 6  0.33608858   0.6769422 0.04338966     1.10  9.795439  21.81594
```

```r
#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
## 1 TEMPEST_FW_SOURCE_20230607_0611_spk  16.0473  17.0125   23.1723  29.7173
## 2     TEMPEST_SW_E3_20230607_1600_spk 332.1500   7.7468 2355.1999  66.4493
## 3     TEMPEST_SW_I5_20230607_1600_spk 468.0219  10.9158 3179.1880  89.6971
## 4 TEMPEST_SW_SOURCE_20230607_0621_spk 563.1599  13.1347 4002.2094 112.9177
## 5  TEMPEST_SW_SOURCE_20230607_1515spk 535.8099  12.4968 4345.0138 122.5896
##       SO4_mM       Cl_mM   salinity Dilution  SO4_ugmL  Cl_ugmL
## 1  0.5005396   0.6536615 0.04189835      1.1 14.588417 21.06574
## 2 10.3602620  66.4372327 4.25587222     50.0  6.642967 47.10402
```

```
## 3 14.5983125  89.6809027 5.74481872     50.0  9.360423 63.58373
## 4 17.5658110 112.8973032 7.23201839     50.0 11.263157 80.04415
## 5 16.7127230 122.5673850 7.85146594     50.0 10.716150 86.90029
```

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

```
##                             Sample_ID Dilution  SO4_ugmL
## 1 TEMPEST_FW_SOURCE_20230607_0611_spk      1.1 14.588417
## 2     TEMPEST_SW_E3_20230607_1600_spk     50.0  6.642967
## 3     TEMPEST_SW_I5_20230607_1600_spk     50.0  9.360423
## 4 TEMPEST_SW_SOURCE_20230607_0621_spk     50.0 11.263157
## 5  TEMPEST_SW_SOURCE_20230607_1515spk     50.0 10.716150
```

```
#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_FW_SOURCE_20230607_0611      1.1    14.588417
## 2        TEMPEST_SW_E3_20230607_1600     50.0     6.642967
## 3        TEMPEST_SW_I5_20230607_1600     50.0     9.360423
## 4    TEMPEST_SW_SOURCE_20230607_0621     50.0    11.263157
## 5 TEMPEST_SW_SOURCE_20230607_1515spk     50.0    10.716150
```

```
#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
## 1    TEMPEST_FW_SOURCE_20230607_0611      1.1  12.7760  13.5445   22.8178
## 2        TEMPEST_SW_E3_20230607_1600     50.0 259.0745   6.0425 2327.3879
## 3        TEMPEST_SW_I5_20230607_1600     50.0 336.2264   7.8419 3120.4635
## 4    TEMPEST_SW_SOURCE_20230607_0621     50.0 571.0464  13.3187 4300.3781
## 5 TEMPEST_SW_SOURCE_20230607_1515spk     50.0 535.8099  12.4968 4345.0138
##    Cl_Area     SO4_mM       Cl_mM   salinity  SO4_ugmL  Cl_ugmL SO4_ugmL_spk
## 1  29.2627  0.3985028   0.6436615 0.04125776 11.614565 20.74349    14.588417
## 2  65.6646  8.0809264  65.6526911 4.20561594  5.181510 46.54777     6.642967
## 3  88.0403 10.4874111  88.0243583 5.63870354  6.724517 62.40927     9.360423
## 4 121.3302 17.8118029 121.3082680 7.77080923 11.420939 86.00754    11.263157
## 5 122.5896 16.7127230 122.5673850 7.85146594 10.716150 86.90029    10.716150
```

```
#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
## 1    TEMPEST_FW_SOURCE_20230607_0611      1.1  12.7760  13.5445   22.8178
## 2        TEMPEST_SW_E3_20230607_1600     50.0 259.0745   6.0425 2327.3879
## 3        TEMPEST_SW_I5_20230607_1600     50.0 336.2264   7.8419 3120.4635
## 4    TEMPEST_SW_SOURCE_20230607_0621     50.0 571.0464  13.3187 4300.3781
## 5 TEMPEST_SW_SOURCE_20230607_1515spk     50.0 535.8099  12.4968 4345.0138
##    Cl_Area     SO4_mM       Cl_mM   salinity  SO4_ugmL  Cl_ugmL SO4_ugmL_spk
## 1  29.2627  0.3985028   0.6436615 0.04125776 11.614565 20.74349    14.588417
## 2  65.6646  8.0809264  65.6526911 4.20561594  5.181510 46.54777     6.642967
## 3  88.0403 10.4874111  88.0243583 5.63870354  6.724517 62.40927     9.360423
## 4 121.3302 17.8118029 121.3082680 7.77080923 11.420939 86.00754    11.263157
## 5 122.5896 16.7127230 122.5673850 7.85146594 10.716150 86.90029    10.716150
##   SO4_spk_Conc
## 1          2.5
## 2          2.5
## 3          2.5
## 4          2.5
## 5          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$SampleVol <-  ifelse(QAspks$Dilution == 1.25, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.1, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 52.74, 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)
```

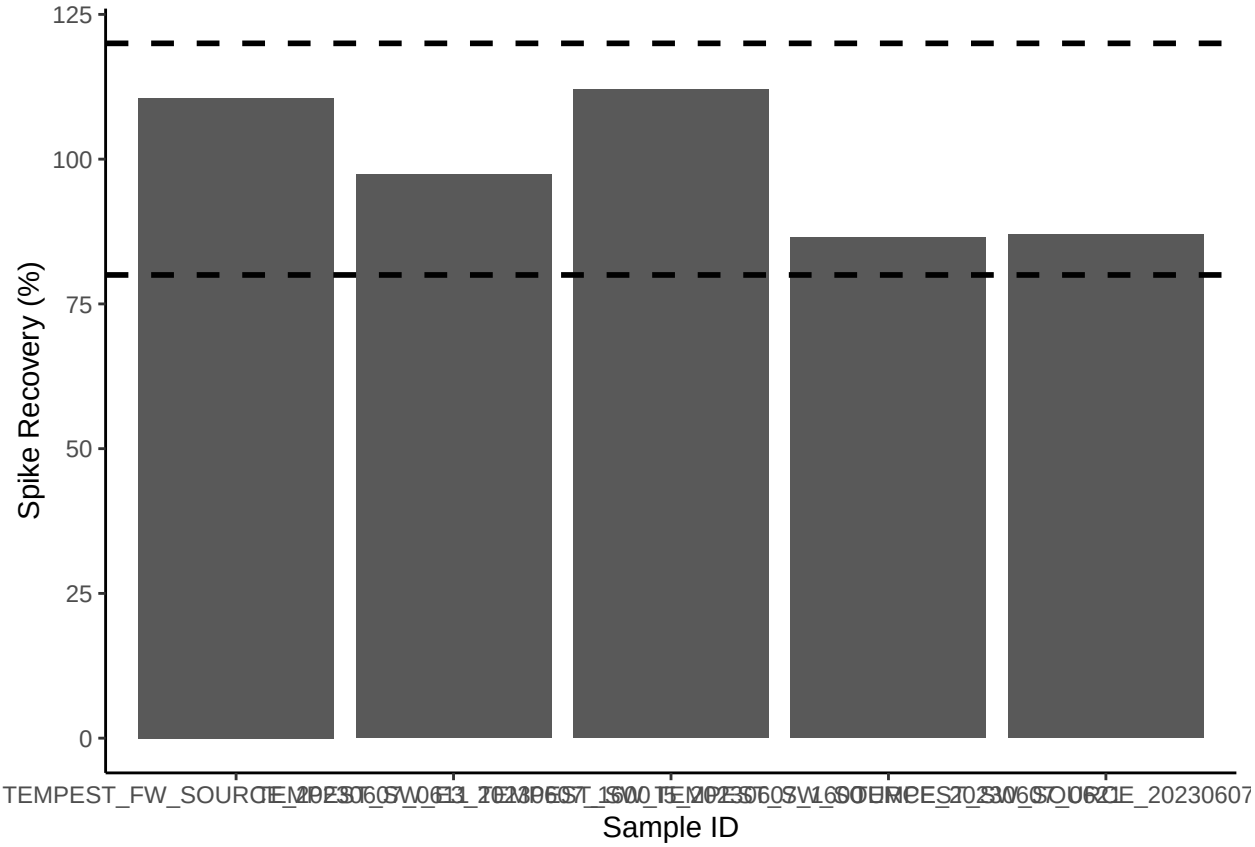
```
##                            Sample_ID Dilution  SO4_ppm SO4_Area    Cl_ppm
## 1    TEMPEST_FW_SOURCE_20230607_0611      1.1  12.7760  13.5445   22.8178
## 2        TEMPEST_SW_E3_20230607_1600     50.0 259.0745   6.0425 2327.3879
## 3        TEMPEST_SW_I5_20230607_1600     50.0 336.2264   7.8419 3120.4635
## 4    TEMPEST_SW_SOURCE_20230607_0621     50.0 571.0464  13.3187 4300.3781
## 5 TEMPEST_SW_SOURCE_20230607_1515spk     50.0 535.8099  12.4968 4345.0138
##    Cl_Area     SO4_mM       Cl_mM   salinity  SO4_ugmL  Cl_ugmL SO4_ugmL_spk
## 1  29.2627  0.3985028   0.6436615 0.04125776 11.614565 20.74349    14.588417
## 2  65.6646  8.0809264  65.6526911 4.20561594  5.181510 46.54777     6.642967
## 3  88.0403 10.4874111  88.0243583 5.63870354  6.724517 62.40927     9.360423
## 4 121.3302 17.8118029 121.3082680 7.77080923 11.420939 86.00754    11.263157
## 5 122.5896 16.7127230 122.5673850 7.85146594 10.716150 86.90029    10.716150
##   SO4_spk_Conc SampleVol
## 1          2.5     1.500
## 2          2.5     1.483
## 3          2.5     1.483
## 4          2.5     1.483
## 5          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, rerun

head(QAspks)
```

```
##                            Sample_ID Dilution  SO4_ppm SO4_Area    Cl_ppm
## 1    TEMPEST_FW_SOURCE_20230607_0611      1.1  12.7760  13.5445   22.8178
## 2        TEMPEST_SW_E3_20230607_1600     50.0 259.0745   6.0425 2327.3879
## 3        TEMPEST_SW_I5_20230607_1600     50.0 336.2264   7.8419 3120.4635
## 4    TEMPEST_SW_SOURCE_20230607_0621     50.0 571.0464  13.3187 4300.3781
## 5 TEMPEST_SW_SOURCE_20230607_1515spk     50.0 535.8099  12.4968 4345.0138
##    Cl_Area     SO4_mM       Cl_mM   salinity  SO4_ugmL  Cl_ugmL SO4_ugmL_spk
```

```
## 1  29.2627  0.3985028   0.6436615 0.04125776 11.614565 20.74349    14.588417
## 2  65.6646  8.0809264  65.6526911 4.20561594  5.181510 46.54777     6.642967
## 3  88.0403 10.4874111  88.0243583 5.63870354  6.724517 62.40927     9.360423
## 4 121.3302 17.8118029 121.3082680 7.77080923 11.420939 86.00754    11.263157
## 5 122.5896 16.7127230 122.5673850 7.85146594 10.716150 86.90029    10.716150
##   SO4_spk_Conc SampleVol SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd
## 1          2.5     1.500        17.421848       22.02851        19.92185
## 2          2.5     1.483         7.684179        9.91795        10.18418
## 3          2.5     1.483         9.972458       13.97511        12.47246
## 4          2.5     1.483        16.937252       16.81589        19.43725
## 5          2.5     1.483        15.892050       15.99921        18.39205
##   spk_recovery SO4_spks_flag
## 1    110.57463           YES
## 2     97.38586           YES
## 3    112.04777           YES
## 4     86.51373           YES
## 5     86.98982           YES
```

```
#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 YES                 5     5     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", "Time")
head(IDs)
```

```
##   Project    Plot   Grid     Date Time
## 1 TEMPEST ESTUARY SOURCE 20230607 0710
```

```
## 2 TEMPEST ESTUARY SOURCE 20230607 1346
## 3 TEMPEST ESTUARY SOURCE 20230607 1700
## 4 TEMPEST      FW     B4 20230607 0800
## 5 TEMPEST      FW     B4 20230607 1600
## 6 TEMPEST      FW     C3 20230607 0800
```

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

```
##   Project    Plot   Grid     Date Time                             Sample_ID
## 1 TEMPEST ESTUARY SOURCE 20230607 0710 TEMPEST_ESTUARY_SOURCE_20230607_0710
## 2 TEMPEST ESTUARY SOURCE 20230607 1346 TEMPEST_ESTUARY_SOURCE_20230607_1346
## 3 TEMPEST ESTUARY SOURCE 20230607 1700 TEMPEST_ESTUARY_SOURCE_20230607_1700
## 4 TEMPEST      FW     B4 20230607 0800          TEMPEST_FW_B4_20230607_0800
## 5 TEMPEST      FW     B4 20230607 1600          TEMPEST_FW_B4_20230607_1600
## 6 TEMPEST      FW     C3 20230607 0800          TEMPEST_FW_C3_20230607_0800
##    SO4_ppm SO4_Area    Cl_ppm  Cl_Area      SO4_mM       Cl_mM   salinity
## 1 581.2849  13.5575 4335.7195 122.3274 18.13115721 122.3052045 7.83467114
## 2 547.2306  12.7632 4145.0144 116.9468 17.06895197 116.9256530 7.49006702
## 3 553.9434  12.9198 4153.6146 117.1895 17.27833437 117.1682539 7.50560758
## 4  11.1494  11.8200   27.9934  35.9001  0.34776669   0.7896587 0.05061007
## 5   2.4624   2.2973  206.4674 233.0096  0.07680599   5.8241862 0.37311259
## 6  10.7750  11.4231   23.9976  30.7756  0.33608858   0.6769422 0.04338966
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_TEMPEST_Dionex_ClSo4_Final_20230608.csv")          #Cha
```

END