Dionex_COMPASS_TEMPEST

Stephanie J. Wilson

2023-06-09

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_2023_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 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  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 5 5 Standard 1 Calibration Standard 5.733   0.4858     6.82   0.6526     3.36
## 6 6 Standard 2 Calibration Standard 5.767   0.9999     6.94   1.3433     6.78
```

```
## 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  Lab Blank     n.a.     n.a.
## 5 Standard 1  0.4858   0.6526
## 6 Standard 2  0.9999   1.3433
```

```
## 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  Lab Blank      NA       NA
## 2  Lab Blank      NA       NA
## 3  Lab Blank      NA       NA
## 4  Lab Blank      NA       NA
## 5 Standard 1  0.4858   0.6526
## 6 Standard 2  0.9999   1.3433
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_TEMPEST_2023_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 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  Lab Blank             Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 5 5 Standard 1 Calibration Standard 4.523   4.9712   93.18   8.9150   54.84
## 6 6 Standard 2 Calibration Standard 4.497  10.0411   93.01  18.0069  109.84
```

```
## 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  Lab Blank     n.a.     n.a.
## 5 Standard 1  4.9712   8.9150
## 6 Standard 2 10.0411  18.0069
```

```
## 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
```

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

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

```
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  Lab Blank      NA      NA
## 5 Standard 1  4.9712  8.9150
## 6 Standard 2 10.0411 18.0069
```

```
## 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           154.3618  26.9769 1207.9926 359.0921
## 2           154.3618  26.9769   76.2651  22.4134
## 3           154.3618  26.9769    0.0009   0.0000
## 4           154.3618  26.9769  193.2321  69.7428
## 5           154.3618  26.9769        NA       NA
## 6            21.9166   2.1247 1207.9926 359.0921
```

```
## 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     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013
```

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

```
##                        Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area
## 26     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013
```

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

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

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

```
##      Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area
## 297 Standard 1  0.4858   0.6526 4.9712  8.9150
## 298 Standard 1  0.4858   0.6526 5.0469  9.0507
## 299 Standard 1  0.4858   0.6526 5.0406  9.0394
## 300 Standard 1  0.4924   0.6615 4.9712  8.9150
## 301 Standard 1  0.4924   0.6615 5.0469  9.0507
## 302 Standard 1  0.4924   0.6615 5.0406  9.0394
```

```r
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  0.490 0.00293 0.599 YES
## 2 Standard 2  1.00  0.00216 0.216 YES
## 3 Standard 3  2.01  0.00351 0.175 YES
## 4 Standard 4 10.1   0.0244  0.243 YES
## 5 Standard 5 20.0   0.0554  0.277 YES
```

```r
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   5.02 0.0364 0.725 YES
## 2 Standard 2  10.1  0.0540 0.534 YES
## 3 Standard 3  20.1  0.0868 0.431 YES
## 4 Standard 4 101.   0.226  0.225 YES
## 5 Standard 5 200.   0.220  0.110 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)
```

```r
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
## 26     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013
```

```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    SO4_mM
## 26     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025 0.1208734
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014 0.1311229
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349 0.1453088
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379 2.8460543
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003 0.1712695
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013 0.1508546
##           Cl_mM     salinity
## 26 0.0019887165 0.0001533935
## 27 0.0011170663 0.0000975572
## 28 0.0274104372 0.0017818619
## 29 0.0298053597 0.0019352762
## 30 0.0002454161 0.0000417209
## 31 0.0010465444 0.0000930397
```

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
## 26     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025 0.1208734
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014 0.1311229
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349 0.1453088
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379 2.8460543
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003 0.1712695
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013 0.1508546
##           Cl_mM     salinity
## 26 0.0019887165 0.0001533935
## 27 0.0011170663 0.0000975572
## 28 0.0274104372 0.0017818619
## 29 0.0298053597 0.0019352762
## 30 0.0002454161 0.0000417209
## 31 0.0010465444 0.0000930397
```

```
#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  100_TEMPEST_C_B4_20230810_dup   4.2038   0.1129   0.0396  0.0014 0.1311229
## 2 110_TEMPEST_SW_C3_20231002_dup 125.8279   3.3807 983.5062 35.2749 3.9247629
## 3   120_TEMPEST_SW_B4_202311_dup  57.1233   1.5348 351.4739 12.6061 1.7817623
## 4 130_TEMPEST_FW_F6_20231208_dup   5.1334   0.1379   0.5484  0.0197 0.1601185
## 5  140_TEMPEST_C_H3_20231211_dup   5.4889   0.1475   0.5758  0.0207 0.1712071
## 6  90_TEMPEST_FW_D5_20230807_dup   5.1056   0.1372   0.0475  0.0017 0.1592514
##          Cl_mM     salinity
## 1  0.001117066 0.0000975572
## 2 27.743475317 1.7772217034
## 3  9.914637518 0.6351393373
## 4  0.015469676 0.0010169588
## 5  0.016242595 0.0010664706
## 6  0.001339915 0.0001118325
```

```
#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
## 1  100_TEMPEST_C_B4_20230810   3.8752   0.1041   0.0705  0.0025 0.1208734
## 2  101_TEMPEST_C_H6_20230807   4.6586   0.1252   0.9717  0.0349 0.1453088
## 3  102_TEMPEST_C_H3_20230807   5.4909   0.1475   0.0087  0.0003 0.1712695
## 4  103_TEMPEST_C_F4_20230810   4.8364   0.1299   0.0371  0.0013 0.1508546
```

```
## 5  104_TEMPEST_C_C6_20230807   4.8123   0.1293   3.7102  0.1331 0.1501029
## 6 105_TEMPEST_SW_I5_20231002 114.6553   3.0805 281.4327 10.0940 3.5762726
##          Cl_mM     salinity
## 1 0.0019887165 0.0001533935
## 2 0.0274104372 0.0017818619
## 3 0.0002454161 0.0000417209
## 4 0.0010465444 0.0000930397
## 5 0.1046600846 0.0067303314
## 6 7.9388631876 0.5085748889
```

```
#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  100_TEMPEST_C_B4_20230810  0.1311229  0.001117066 0.0000975572
## 2 110_TEMPEST_SW_C3_20231002  3.9247629 27.743475317 1.7772217034
## 3   120_TEMPEST_SW_B4_202311  1.7817623  9.914637518 0.6351393373
## 4 130_TEMPEST_FW_F6_20231208  0.1601185  0.015469676 0.0010169588
## 5  140_TEMPEST_C_H3_20231211  0.1712071  0.016242595 0.0010664706
## 6  90_TEMPEST_FW_D5_20230807  0.1592514  0.001339915 0.0001118325
```

```
#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  100_TEMPEST_C_B4_20230810   3.8752   0.1041   0.0705  0.0025 0.1208734
## 2 110_TEMPEST_SW_C3_20231002 126.0340   3.3862 988.1749 35.4423 3.9311915
## 3   120_TEMPEST_SW_B4_202311  57.4551   1.5437 353.8215 12.6903 1.7921117
## 4 130_TEMPEST_FW_F6_20231208   5.2059   0.1399   0.5455  0.0196 0.1623799
## 5  140_TEMPEST_C_H3_20231211   5.5834   0.1500   0.3123  0.0112 0.1741547
## 6  90_TEMPEST_FW_D5_20230807   5.0450   0.1355   0.0494  0.0018 0.1573612
##          Cl_mM     salinity SO4_mM_dup    Cl_mM_dup salinity_dup
## 1  0.001988717 0.0001533935  0.1311229  0.001117066 0.0000975572
## 2 27.875173484 1.7856580443  3.9247629 27.743475317 1.7772217034
## 3  9.980860367 0.6393814505  1.7817623  9.914637518 0.6351393373
## 4  0.015387870 0.0010117185  0.1601185  0.015469676 0.0010169588
## 5  0.008809591 0.0005903261  0.1712071  0.016242595 0.0010664706
## 6  0.001393512 0.0001152658  0.1592514  0.001339915 0.0001118325
```

```
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)
```

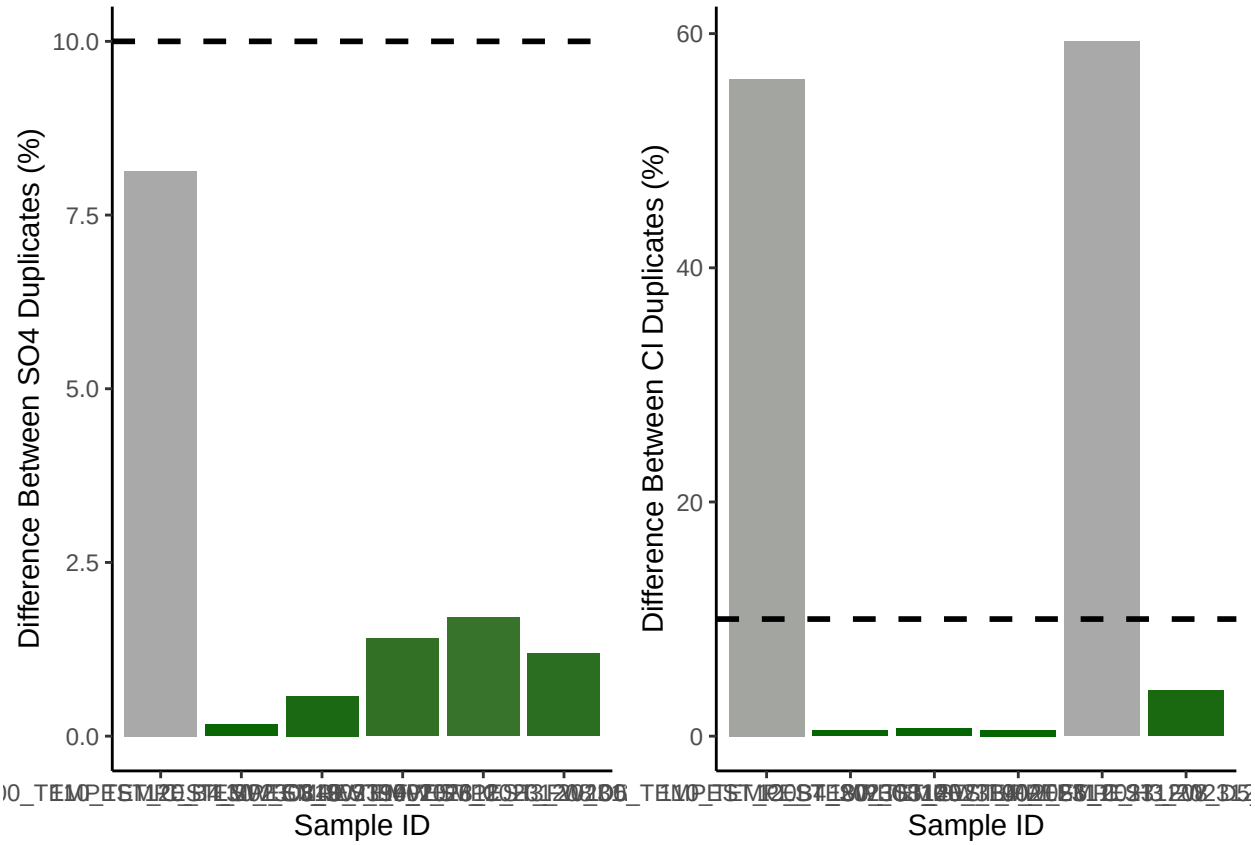
```
##                    Sample_ID  SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
```

```
## 1  100_TEMPEST_C_B4_20230810   3.8752   0.1041   0.0705  0.0025 0.1208734
## 2 110_TEMPEST_SW_C3_20231002 126.0340   3.3862 988.1749 35.4423 3.9311915
## 3   120_TEMPEST_SW_B4_202311  57.4551   1.5437 353.8215 12.6903 1.7921117
## 4 130_TEMPEST_FW_F6_20231208   5.2059   0.1399   0.5455  0.0196 0.1623799
## 5  140_TEMPEST_C_H3_20231211   5.5834   0.1500   0.3123  0.0112 0.1741547
## 6  90_TEMPEST_FW_D5_20230807   5.0450   0.1355   0.0494  0.0018 0.1573612
##          Cl_mM     salinity SO4_mM_dup    Cl_mM_dup salinity_dup SO4_dups_chk
## 1  0.001988717 0.0001533935  0.1311229  0.001117066 0.0000975572    8.1346701
## 2 27.875173484 1.7856580443  3.9247629 27.743475317 1.7772217034    0.1636611
## 3  9.980860367 0.6393814505  1.7817623  9.914637518 0.6351393373    0.5791668
## 4  0.015387870 0.0010117185  0.1601185  0.015469676 0.0010169588    1.4024160
## 5  0.008809591 0.0005903261  0.1712071  0.016242595 0.0010664706    1.7069624
## 6  0.001393512 0.0001152658  0.1592514  0.001339915 0.0001118325    1.1940181
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1           YES  56.1307902    NO, rerun
## 2           YES   0.4735756          YES
## 3           YES   0.6657069          YES
## 4           YES   0.5302130          YES
## 5           YES  59.3401644    NO, rerun
## 6           YES   3.9215686          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)
```




```
#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)
```

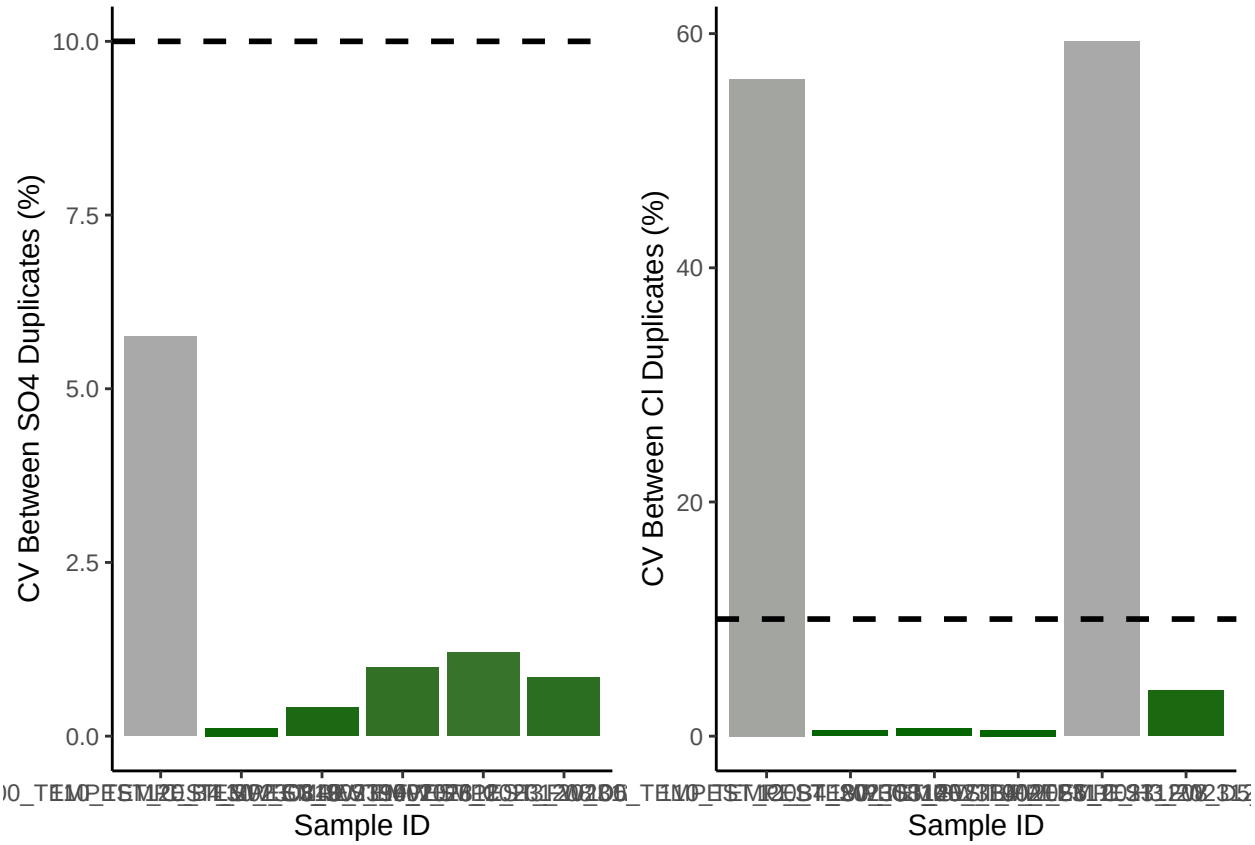
```
##   Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES         6 NO, rerun          2     6       100   33.33333
## 2  YES         6       YES          4     6       100   66.66667
```

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  100_TEMPEST_C_B4_20230810   3.8752   0.1041   0.0705  0.0025 0.1208734
## 2 110_TEMPEST_SW_C3_20231002 126.0340   3.3862 988.1749 35.4423 3.9311915
## 3   120_TEMPEST_SW_B4_202311  57.4551   1.5437 353.8215 12.6903 1.7921117
## 4 130_TEMPEST_FW_F6_20231208   5.2059   0.1399   0.5455  0.0196 0.1623799
## 5  140_TEMPEST_C_H3_20231211   5.5834   0.1500   0.3123  0.0112 0.1741547
## 6  90_TEMPEST_FW_D5_20230807   5.0450   0.1355   0.0494  0.0018 0.1573612
##          Cl_mM     salinity SO4_mM_dup    Cl_mM_dup salinity_dup SO4_dups_chk
## 1  0.001988717 0.0001533935  0.1311229  0.001117066 0.0000975572    8.1346701
## 2 27.875173484 1.7856580443  3.9247629 27.743475317 1.7772217034    0.1636611
## 3  9.980860367 0.6393814505  1.7817623  9.914637518 0.6351393373    0.5791668
## 4  0.015387870 0.0010117185  0.1601185  0.015469676 0.0010169588    1.4024160
## 5  0.008809591 0.0005903261  0.1712071  0.016242595 0.0010664706    1.7069624
## 6  0.001393512 0.0001152658  0.1592514  0.001339915 0.0001118325    1.1940181
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1           YES  56.1307902    NO, rerun   5.7520804              YES
## 2           YES   0.4735756          YES   0.1157259              YES
## 3           YES   0.6657069          YES   0.4095327              YES
## 4           YES   0.5302130          YES   0.9916579              YES
## 5           YES  59.3401644    NO, rerun   1.2070047              YES
## 6           YES   3.9215686          YES   0.8442983              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)
```




```
#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         6 NO, rerun          2     6       100   33.33333
## 2  YES         6       YES          4     6       100   66.66667
```

Pull out spikes and check with dionex output conc.

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

```
##                         Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area    SO4_mM
## 26     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025 0.1208734
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014 0.1311229
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349 0.1453088
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379 2.8460543
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003 0.1712695
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013 0.1508546
##           Cl_mM     salinity
## 26 0.0019887165 0.0001533935
## 27 0.0011170663 0.0000975572
## 28 0.0274104372 0.0017818619
## 29 0.0298053597 0.0019352762
## 30 0.0002454161 0.0000417209
## 31 0.0010465444 0.0000930397
```

```r
#pull out any rows that have "spk" in the SampleID column
spks <- sampledat %>%
  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
## 1  101_TEMPEST_C_H6_20230807_spk  91.2445   2.4515   1.0566  0.0379 2.846054
## 2 111_TEMPEST_FW_I5_20231002_spk  89.9420   2.4165   0.3888  0.0139 2.805427
## 3   121_TEMPEST_SW_H6_202311_spk 106.4642   2.8604 141.0457  5.0588 3.320780
## 4 131_TEMPEST_FW_E3_20231208_spk  90.8868   2.4419   0.8722  0.0313 2.834897
## 5  141_TEMPEST_C_C6_20231208_spk  88.8596   2.3874   1.1352  0.0407 2.771666
## 6  91_TEMPEST_FW_B4_20230807_spk  93.5863   2.5144   0.0059  0.0002 2.919099
##          Cl_mM     salinity
## 1 0.0298053597 0.0019352762
## 2 0.0109675599 0.0007285616
## 3 3.9787221439 0.2548955799
## 4 0.0246036671 0.0016020654
## 5 0.0320225670 0.0020773064
## 6 0.0001664316 0.0000366613
```

```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(2,3,4 ,5, 7,8)]
colnames(spks) <- c('Sample_ID', 'SO4_mM_spk')
head(spks)
```

```
##                    Sample_ID SO4_mM_spk
## 1  101_TEMPEST_C_H6_20230807   2.846054
## 2 111_TEMPEST_FW_I5_20231002   2.805427
## 3   121_TEMPEST_SW_H6_202311   3.320780
## 4 131_TEMPEST_FW_E3_20231208   2.834897
## 5  141_TEMPEST_C_C6_20231208   2.771666
## 6  91_TEMPEST_FW_B4_20230807   2.919099
```

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

```
##                    Sample_ID SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity SO4_mM_spk
## 1 0.027410437 0.0017818619   2.846054
## 2 0.009726375 0.0006490536   2.805427
## 3 4.006081805 0.2566481892   3.320780
## 4 0.022592384 0.0014732263   2.834897
## 5 0.024942172 0.0016237494   2.771666
## 6 0.925957687 0.0593411364   2.919099
```

```r
#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
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity SO4_mM_spk SO4_spk_Conc
## 1 0.027410437 0.0017818619   2.846054 7.797879e-05
## 2 0.009726375 0.0006490536   2.805427 7.797879e-05
## 3 4.006081805 0.2566481892   3.320780 7.797879e-05
## 4 0.022592384 0.0014732263   2.834897 7.797879e-05
## 5 0.024942172 0.0016237494   2.771666 7.797879e-05
## 6 0.925957687 0.0593411364   2.919099 7.797879e-05
```

```r
#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
QAspks$SampleVol <- 1
QAspks$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$Sample_ID, "GCrew_UP"), 1501, QAspks$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)
```

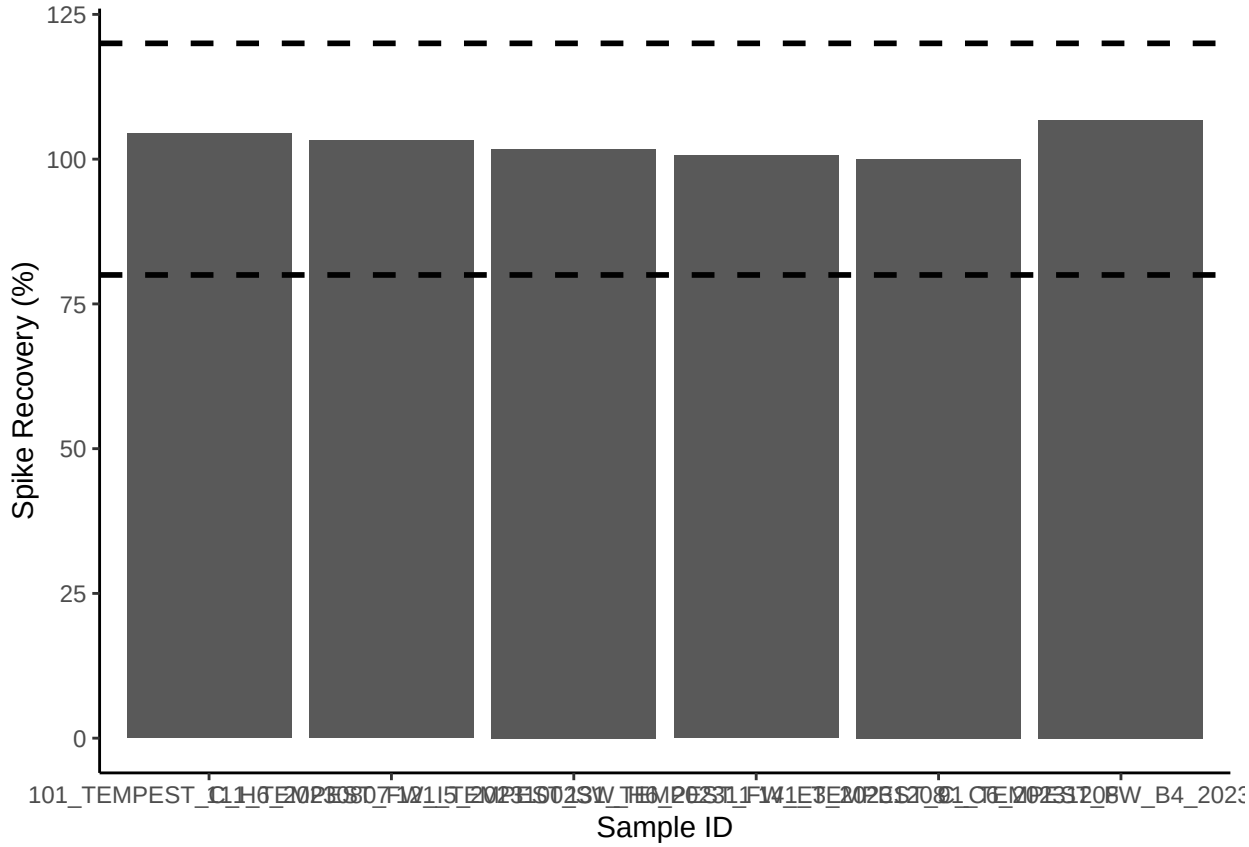
```
##                    Sample_ID SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 0.027410437 0.0017818619   2.846054 7.797879e-05       50  0.001501
## 2 0.009726375 0.0006490536   2.805427 7.797879e-05       50  0.001501
## 3 4.006081805 0.2566481892   3.320780 7.797879e-05       50  0.001501
## 4 0.022592384 0.0014732263   2.834897 7.797879e-05       50  0.001501
## 5 0.024942172 0.0016237494   2.771666 7.797879e-05       50  0.001501
## 6 0.925957687 0.0593411364   2.919099 7.797879e-05       50  0.001501
```

```
#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
```

```r
head(QAspks)
```

```
##                    Sample_ID SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 0.027410437 0.0017818619   2.846054 7.797879e-05       50  0.001501
## 2 0.009726375 0.0006490536   2.805427 7.797879e-05       50  0.001501
## 3 4.006081805 0.2566481892   3.320780 7.797879e-05       50  0.001501
## 4 0.022592384 0.0014732263   2.834897 7.797879e-05       50  0.001501
## 5 0.024942172 0.0016237494   2.771666 7.797879e-05       50  0.001501
## 6 0.925957687 0.0593411364   2.919099 7.797879e-05       50  0.001501
##   SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1     4.362170e-06   8.600776e-05    8.234096e-05     104.4532           YES
## 2     4.132947e-06   8.478001e-05    8.211174e-05     103.2496           YES
## 3     2.064217e-05   1.003540e-04    9.862096e-05     101.7572           YES
## 4     7.195155e-06   8.567059e-05    8.517395e-05     100.5831           YES
## 5     5.759233e-06   8.375974e-05    8.373802e-05     100.0259           YES
## 6     4.649167e-06   8.821516e-05    8.262796e-05     106.7619           YES
```

```r
#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                 6     6     100
```

Pull out spikes and check with area calc

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

```
##                         Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area    SO4_mM
## 26     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025 0.1208734
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014 0.1311229
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349 0.1453088
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379 2.8460543
```

```
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003 0.1712695
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013 0.1508546
##           Cl_mM     salinity
## 26 0.0019887165 0.0001533935
## 27 0.0011170663 0.0000975572
## 28 0.0274104372 0.0017818619
## 29 0.0298053597 0.0019352762
## 30 0.0002454161 0.0000417209
## 31 0.0010465444 0.0000930397
```

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

```
##                         Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area    SO4_mM
## 26     100_TEMPEST_C_B4_20230810  3.8752   0.1041 0.0705  0.0025 0.1208734
## 27 100_TEMPEST_C_B4_20230810_dup  4.2038   0.1129 0.0396  0.0014 0.1311229
## 28     101_TEMPEST_C_H6_20230807  4.6586   0.1252 0.9717  0.0349 0.1453088
## 29 101_TEMPEST_C_H6_20230807_spk 91.2445   2.4515 1.0566  0.0379 2.8460543
## 30     102_TEMPEST_C_H3_20230807  5.4909   0.1475 0.0087  0.0003 0.1712695
## 31     103_TEMPEST_C_F4_20230810  4.8364   0.1299 0.0371  0.0013 0.1508546
##           Cl_mM     salinity   SO4_ugmL      Cl_ugmL
## 26 0.0019887165 0.0001533935 0.07747505 0.0013990030
## 27 0.0011170663 0.0000975572 0.08402573 0.0007856160
## 28 0.0274104372 0.0017818619 0.09318179 0.0194660385
## 29 0.0298053597 0.0019352762 1.82486858 0.0211389122
## 30 0.0002454161 0.0000417209 0.10978180 0.0001722290
## 31 0.0010465444 0.0000930397 0.09668045 0.0007298536
```

```
#pull out any rows that have "spk" in the SampleID column
spks <- sampledat %>%
  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
## 1  101_TEMPEST_C_H6_20230807_spk  91.2445   2.4515   1.0566  0.0379 2.846054
## 2 111_TEMPEST_FW_I5_20231002_spk  89.9420   2.4165   0.3888  0.0139 2.805427
## 3   121_TEMPEST_SW_H6_202311_spk 106.4642   2.8604 141.0457  5.0588 3.320780
## 4 131_TEMPEST_FW_E3_20231208_spk  90.8868   2.4419   0.8722  0.0313 2.834897
## 5  141_TEMPEST_C_C6_20231208_spk  88.8596   2.3874   1.1352  0.0407 2.771666
## 6  91_TEMPEST_FW_B4_20230807_spk  93.5863   2.5144   0.0059  0.0002 2.919099
##          Cl_mM     salinity SO4_ugmL      Cl_ugmL
## 1 0.0298053597 0.0019352762 1.824869 0.0211389122
## 2 0.0109675599 0.0007285616 1.798815 0.0077559229
## 3 3.9787221439 0.2548955799 2.129252 2.8209160213
## 4 0.0246036671 0.0016020654 1.817722 0.0174585901
## 5 0.0320225670 0.0020773064 1.777153 0.0227002609
## 6 0.0001664316 0.0000366613 1.871691 0.0001164666
```

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

```
##                         Sample_ID SO4_ugmL      Cl_ugmL
## 1  101_TEMPEST_C_H6_20230807_spk 1.824869 0.0211389122
## 2 111_TEMPEST_FW_I5_20231002_spk 1.798815 0.0077559229
## 3   121_TEMPEST_SW_H6_202311_spk 2.129252 2.8209160213
## 4 131_TEMPEST_FW_E3_20231208_spk 1.817722 0.0174585901
## 5  141_TEMPEST_C_C6_20231208_spk 1.777153 0.0227002609
## 6  91_TEMPEST_FW_B4_20230807_spk 1.871691 0.0001164666
```

```
#remove the spk from these IDs so we will have duplicate sample names
spks$Sample_ID<-gsub("_spk","",as.character(spks$Sample_ID))
spks <- spks[ ,c(1,2)]
colnames(spks) <- c('Sample_ID', 'SO4_ugmL_spk')
head(spks)
```

```
##                     Sample_ID SO4_ugmL_spk
## 1  101_TEMPEST_C_H6_20230807     1.824869
## 2 111_TEMPEST_FW_I5_20231002     1.798815
## 3   121_TEMPEST_SW_H6_202311     2.129252
## 4 131_TEMPEST_FW_E3_20231208     1.817722
## 5  141_TEMPEST_C_C6_20231208     1.777153
## 6  91_TEMPEST_FW_B4_20230807     1.871691
```

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

```
##                     Sample_ID SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity   SO4_ugmL     Cl_ugmL SO4_ugmL_spk
## 1 0.027410437 0.0017818619 0.09318179 0.019466039     1.824869
## 2 0.009726375 0.0006490536 0.08826878 0.006919486     1.798815
## 3 4.006081805 0.2566481892 0.44088883 2.840321356     2.129252
## 4 0.022592384 0.0014732263 0.15370113 0.016008766     1.817722
## 5 0.024942172 0.0016237494 0.12303204 0.017681640     1.777153
## 6 0.925957687 0.0593411364 0.09928583 0.656496326     1.871691
```

```
#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 SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
```

```
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity   SO4_ugmL     Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 0.027410437 0.0017818619 0.09318179 0.019466039     1.824869          2.5
## 2 0.009726375 0.0006490536 0.08826878 0.006919486     1.798815          2.5
## 3 4.006081805 0.2566481892 0.44088883 2.840321356     2.129252          2.5
## 4 0.022592384 0.0014732263 0.15370113 0.016008766     1.817722          2.5
## 5 0.024942172 0.0016237494 0.12303204 0.017681640     1.777153          2.5
## 6 0.925957687 0.0593411364 0.09928583 0.656496326     1.871691          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(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 SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
```

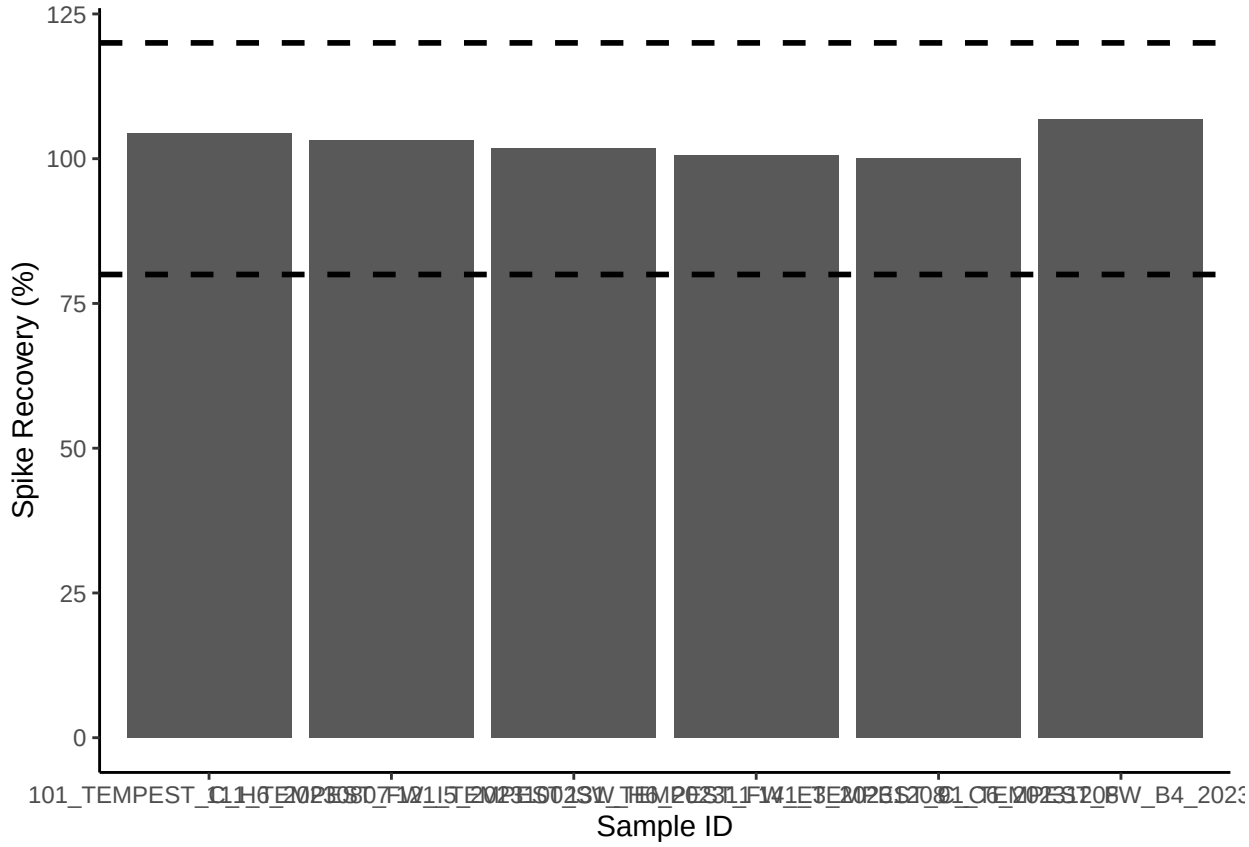
```
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity   SO4_ugmL     Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 0.027410437 0.0017818619 0.09318179 0.019466039     1.824869          2.5
## 2 0.009726375 0.0006490536 0.08826878 0.006919486     1.798815          2.5
## 3 4.006081805 0.2566481892 0.44088883 2.840321356     2.129252          2.5
## 4 0.022592384 0.0014732263 0.15370113 0.016008766     1.817722          2.5
## 5 0.024942172 0.0016237494 0.12303204 0.017681640     1.777153          2.5
## 6 0.925957687 0.0593411364 0.09928583 0.656496326     1.871691          2.5
##   SampleVol
## 1     1.501
## 2     1.501
## 3     1.501
## 4     1.501
## 5     1.501
## 6     1.501
```

```
#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 SO4_ppm SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1  101_TEMPEST_C_H6_20230807  4.6586   0.1252   0.9717  0.0349 0.1453088
## 2 111_TEMPEST_FW_I5_20231002  4.4138   0.1186   0.3448  0.0124 0.1376731
## 3   121_TEMPEST_SW_H6_202311 22.0449   0.5923 142.0156  5.0936 0.6876138
## 4 131_TEMPEST_FW_E3_20231208  7.6841   0.2065   0.8009  0.0287 0.2396787
## 5  141_TEMPEST_C_C6_20231208  6.1506   0.1653   0.8842  0.0317 0.1918465
## 6  91_TEMPEST_FW_B4_20230807  4.9651   0.1334  32.8252  1.1773 0.1548690
##         Cl_mM     salinity   SO4_ugmL     Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 0.027410437 0.0017818619 0.09318179 0.019466039     1.824869          2.5
## 2 0.009726375 0.0006490536 0.08826878 0.006919486     1.798815          2.5
## 3 4.006081805 0.2566481892 0.44088883 2.840321356     2.129252          2.5
## 4 0.022592384 0.0014732263 0.15370113 0.016008766     1.817722          2.5
## 5 0.024942172 0.0016237494 0.12303204 0.017681640     1.777153          2.5
## 6 0.925957687 0.0593411364 0.09928583 0.656496326     1.871691          2.5
##   SampleVol SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery
## 1     1.501        0.1398659       2.757376        2.639866     104.4514
## 2     1.501        0.1324914       2.718009        2.632491     103.2485
## 3     1.501        0.6617741       3.217299        3.161774     101.7561
## 4     1.501        0.2307054       2.746579        2.730705     100.5813
## 5     1.501        0.1846711       2.685278        2.684671     100.0226
## 6     1.501        0.1490280       2.828125        2.649028     106.7609
##   SO4_spks_flag
```

```
## 1           YES
## 2           YES
## 3           YES
## 4           YES
## 5           YES
## 6           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                 6     6     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       NA
## 1     100 TEMPEST    C   B4 20230810
## 2     101 TEMPEST    C   H6 20230807
## 3     102 TEMPEST    C   H3 20230807
## 4     103 TEMPEST    C   F4 20230810
## 5     104 TEMPEST    C   C6 20230807
## 6     105 TEMPEST   SW   I5 20231002
```

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

```
##   Project    Plot Grid Date       NA                  Sample_ID  SO4_ppm
## 1     100 TEMPEST    C   B4 20230810  100_TEMPEST_C_B4_20230810   3.8752
## 2     101 TEMPEST    C   H6 20230807  101_TEMPEST_C_H6_20230807   4.6586
## 3     102 TEMPEST    C   H3 20230807  102_TEMPEST_C_H3_20230807   5.4909
## 4     103 TEMPEST    C   F4 20230810  103_TEMPEST_C_F4_20230810   4.8364
## 5     104 TEMPEST    C   C6 20230807  104_TEMPEST_C_C6_20230807   4.8123
## 6     105 TEMPEST   SW   I5 20231002 105_TEMPEST_SW_I5_20231002 114.6553
##   SO4_Area   Cl_ppm Cl_Area    SO4_mM        Cl_mM     salinity
## 1   0.1041   0.0705  0.0025 0.1208734 0.0019887165 0.0001533935
## 2   0.1252   0.9717  0.0349 0.1453088 0.0274104372 0.0017818619
## 3   0.1475   0.0087  0.0003 0.1712695 0.0002454161 0.0000417209
## 4   0.1299   0.0371  0.0013 0.1508546 0.0010465444 0.0000930397
## 5   0.1293   3.7102  0.1331 0.1501029 0.1046600846 0.0067303314
## 6   3.0805 281.4327 10.0940 3.5762726 7.9388631876 0.5085748889
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_TEMPEST_Dionex_ClSo4_Final_2023a.csv")        #Change
```

END