Dionex_COMPASS_TEMPEST

Stephanie J. Wilson

2023-06-07

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_20230607_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 4.347   0.0005     1.32   0.0006     0.00
## 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 4.357   0.4391     6.60   0.5115     2.29
## 5 5 Standard 2 Calibration Standard 4.353   0.9724     6.67   1.1327     5.14
## 6 6 Standard 3 Calibration Standard 4.350   1.9875     7.45   2.3152    10.10
##   IC.SO4.5
## 1       Rd
## 2     n.a.
## 3     n.a.
## 4       M
## 5       M
## 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  0.0005   0.0006
## 2  Lab Blank     n.a.     n.a.
## 3  Lab Blank     n.a.     n.a.
## 4 Standard 1  0.4391   0.5115
## 5 Standard 2  0.9724   1.1327
## 6 Standard 3  1.9875   2.3152
```

```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  0.0005   0.0006
## 2  Lab Blank      NA       NA
## 3  Lab Blank      NA       NA
## 4 Standard 1  0.4391   0.5115
## 5 Standard 2  0.9724   1.1327
## 6 Standard 3  1.9875   2.3152
```

```r
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_TEMPEST_20230607_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 3.420   0.0186   54.19   0.0261    0.07
## 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 3.430   5.1187   92.59   7.1765   39.52
## 5 5 Standard 2 Calibration Standard 3.427  11.2205   92.58  15.7314   85.63
## 6 6 Standard 3 Calibration Standard 3.427  20.4308   92.19  28.6445  154.92
##   IC.Cl.5
## 1      MB
## 2    n.a.
## 3    n.a.
## 4      M
## 5      M
## 6      M
```

```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  0.0186   0.0261
## 2  Lab Blank     n.a.     n.a.
## 3  Lab Blank     n.a.     n.a.
## 4 Standard 1  5.1187   7.1765
## 5 Standard 2 11.2205  15.7314
## 6 Standard 3 20.4308  28.6445
```

```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  0.0186  0.0261
## 2  Lab Blank      NA      NA
## 3  Lab Blank      NA      NA
## 4 Standard 1  5.1187  7.1765
## 5 Standard 2 11.2205 15.7314
## 6 Standard 3 20.4308 28.6445
```

```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           557.6408  23.3103 4194.5484 280.6096
## 2           557.6408  23.3103  979.1875  68.8194
## 3           557.6408  23.3103    0.0014   0.0020
## 4           557.6408  23.3103 1429.0768  72.1471
## 5           557.6408  23.3103        NA       NA
## 6           140.4325   7.6988 4194.5484 280.6096
```

```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   5e-04    6e-04 0.0186  0.0261
## 27 Lab Blank   5e-04    6e-04     NA      NA
## 28 Lab Blank   5e-04    6e-04     NA      NA
## 29 Lab Blank   5e-04    6e-04 0.0014  0.0020
## 30 Lab Blank   5e-04    6e-04 0.0162  0.0227
## 31 Lab Blank   5e-04    6e-04 0.0093  0.0131
```

```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
## 147 Standard 1  0.4391   0.5115 5.1187  7.1765
## 148 Standard 1  0.4391   0.5115 5.1585  7.2323
## 149 Standard 1  0.4391   0.5115 5.1432  7.2109
## 150 Standard 1  0.4479   0.5218 5.1187  7.1765
## 151 Standard 1  0.4479   0.5218 5.1585  7.2323
## 152 Standard 1  0.4479   0.5218 5.1432  7.2109
```

```
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
## 147 Standard 1  0.4391   0.5115 5.1187  7.1765
## 148 Standard 1  0.4391   0.5115 5.1585  7.2323
## 149 Standard 1  0.4391   0.5115 5.1432  7.2109
## 150 Standard 1  0.4479   0.5218 5.1187  7.1765
## 151 Standard 1  0.4479   0.5218 5.1585  7.2323
## 152 Standard 1  0.4479   0.5218 5.1432  7.2109
```

```
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  0.452 0.0136   3.02   NO, rerun
## 2 Standard 2  0.984 0.0162   1.64   YES
## 3 Standard 3  1.94  0.0543   2.80   NO, rerun
## 4 Standard 4  8.41  1.83    21.8    NO, rerun
## 5 Standard 5 20.0   0.00254  0.0127 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   5.14 0.0174  0.338  YES
```

```
## 2 Standard 2  11.2  0.00524 0.0467 YES
## 3 Standard 3  20.4  0.00439 0.0215 YES
## 4 Standard 4 101.   0.0511  0.0506 YES
## 5 Standard 5 200.   0.407   0.204  YES
```

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


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

Calculate mmol/L concentrations

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

```
##                                    Sample_ID  SO4_ppm SO4_Area    Cl_ppm
## 177          TEMEPST_SW_SOURCE_20230606_1301 556.0527  12.9549 4182.4438
## 178     TEMPEST_ESTUARY_SOURCE_20230606_0632 557.6408  12.9919 4194.5484
## 179     TEMPEST_ESTUARY_SOURCE_20230606_1236 480.7045  11.1994 3636.5862
## 180     TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.0580
## 181 TEMPEST_ESTUARY_SOURCE_20230606_1524_dup 517.1687  12.0490 3902.2588
## 182                   TEMPEST_FW_B4_20230606   2.2280   2.0763  173.8469
##      Cl_Area
## 177 117.2776
## 178 117.6170
## 179 101.9715
## 180 110.1445
## 181 109.4211
## 182 194.9897
```

```
# 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
## 177          TEMEPST_SW_SOURCE_20230606_1301 556.0527  12.9549 4182.4438
## 178     TEMPEST_ESTUARY_SOURCE_20230606_0632 557.6408  12.9919 4194.5484
## 179     TEMPEST_ESTUARY_SOURCE_20230606_1236 480.7045  11.1994 3636.5862
## 180     TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.0580
## 181 TEMPEST_ESTUARY_SOURCE_20230606_1524_dup 517.1687  12.0490 3902.2588
## 182                   TEMPEST_FW_B4_20230606   2.2280   2.0763  173.8469
##      Cl_Area     SO4_mM      Cl_mM  salinity
## 177 117.2776 17.3441266 117.981489 7.5577019
## 178 117.6170 17.3936619 118.322945 7.5795750
## 179 101.9715 14.9939021 102.583532 6.5713373
## 180 110.1445 16.1900031 110.805585 7.0980268
## 181 109.4211 16.1312757 110.077822 7.0514077
## 182 194.9897  0.0694947   4.904003 0.3141673
```

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
## 177          TEMEPST_SW_SOURCE_20230606_1301 556.0527  12.9549 4182.4438
## 178     TEMPEST_ESTUARY_SOURCE_20230606_0632 557.6408  12.9919 4194.5484
## 179     TEMPEST_ESTUARY_SOURCE_20230606_1236 480.7045  11.1994 3636.5862
## 180     TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.0580
## 181 TEMPEST_ESTUARY_SOURCE_20230606_1524_dup 517.1687  12.0490 3902.2588
## 182                   TEMPEST_FW_B4_20230606   2.2280   2.0763  173.8469
##      Cl_Area     SO4_mM      Cl_mM  salinity
## 177 117.2776 17.3441266 117.981489 7.5577019
## 178 117.6170 17.3936619 118.322945 7.5795750
## 179 101.9715 14.9939021 102.583532 6.5713373
## 180 110.1445 16.1900031 110.805585 7.0980268
## 181 109.4211 16.1312757 110.077822 7.0514077
## 182 194.9897  0.0694947   4.904003 0.3141673
```

```
#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_ESTUARY_SOURCE_20230606_1524_dup 517.1687  12.0490 3902.259 109.4211
## 2              TEMPEST_SW_I5_20230606_dup 240.9080   5.6127 2335.150  65.4786
##      SO4_mM     Cl_mM salinity
```

```
## 1 16.131276 110.07782 7.051408
## 2  7.514286  65.87165 4.219642
```

```
#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      TEMEPST_SW_SOURCE_20230606_1301 556.0527  12.9549 4182.4438 117.2776
## 2 TEMPEST_ESTUARY_SOURCE_20230606_0632 557.6408  12.9919 4194.5484 117.6170
## 3 TEMPEST_ESTUARY_SOURCE_20230606_1236 480.7045  11.1994 3636.5862 101.9715
## 4 TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.0580 110.1445
## 5               TEMPEST_FW_B4_20230606   2.2280   2.0763  173.8469 194.9897
## 6               TEMPEST_FW_C3_20230606  12.2001  12.9198   27.3401  34.8467
##       SO4_mM       Cl_mM   salinity
## 1 17.3441266 117.9814894 7.55770195
## 2 17.3936619 118.3229450 7.57957496
## 3 14.9939021 102.5835317 6.57133726
## 4 16.1900031 110.8055853 7.09802681
## 5  0.0694947   4.9040028 0.31416735
## 6  0.3805396   0.7712299 0.04942956
```

```
#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_ESTUARY_SOURCE_20230606_1524  16.131276 110.07782     7.051408
## 2               TEMPEST_SW_I5_20230606   7.514286  65.87165     4.219642
```

```
#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
## 1 TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.058 110.1445
## 2               TEMPEST_SW_I5_20230606 236.2910   5.5051 2285.850  64.0963
##      SO4_mM     Cl_mM salinity SO4_mM_dup Cl_mM_dup salinity_dup
## 1 16.190003 110.80559 7.098027  16.131276 110.07782     7.051408
## 2  7.370274  64.48097 4.130558   7.514286  65.87165     4.219642
```

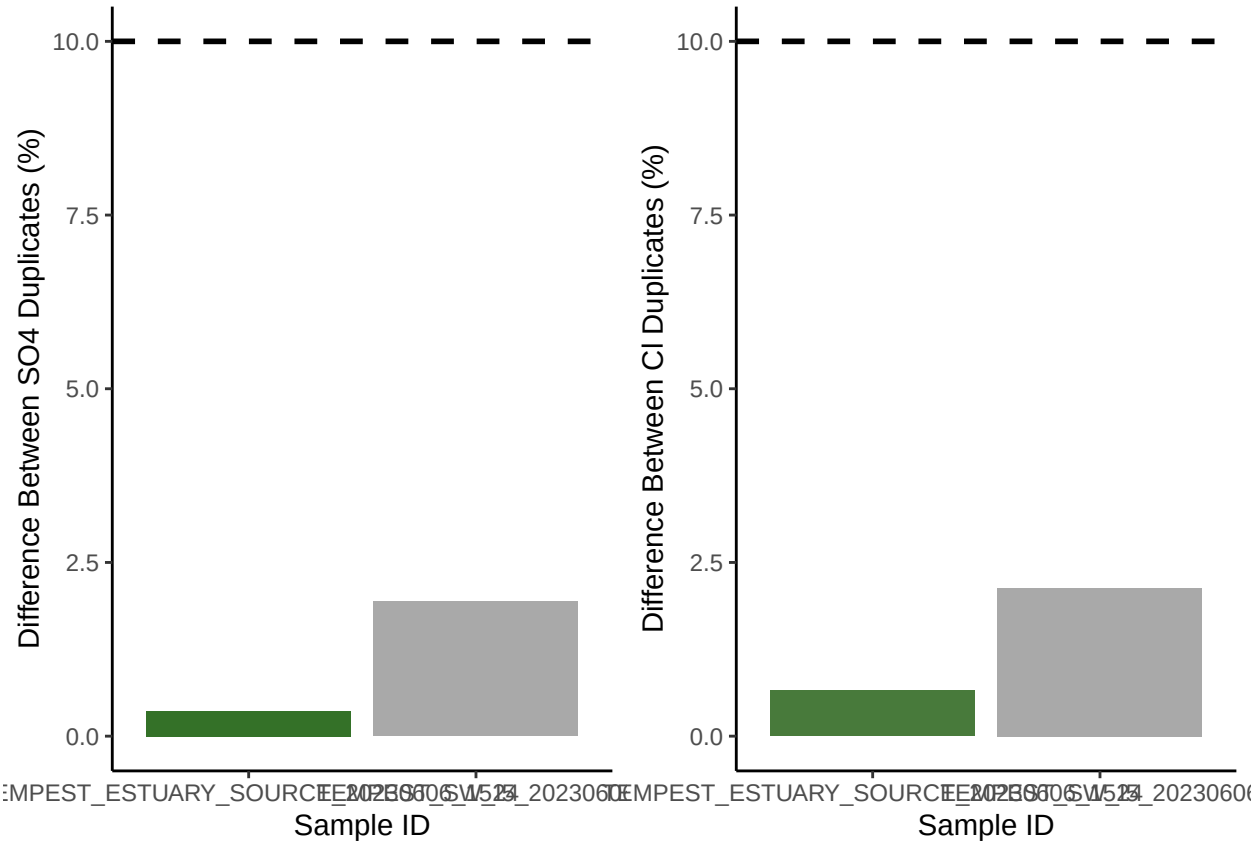
```
QAdups$SO4_dups_chk <- ((abs(QAdups$SO4_mM-QAdups$SO4_mM_dup))/((QAdups$SO4_mM+QAdups$SO4_mM_dup)/2))*10
QAdups$SO4_dups_flag <-  ifelse(QAdups$SO4_dups_chk <10, 'YES', 'NO, rerun')

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
## 1 TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.058 110.1445
## 2               TEMPEST_SW_I5_20230606 236.2910   5.5051 2285.850  64.0963
##      SO4_mM     Cl_mM salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk
## 1 16.190003 110.80559 7.098027  16.131276 110.07782     7.051408    0.3633977
## 2  7.370274  64.48097 4.130558   7.514286  65.87165     4.219642    1.9350418
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1           YES   0.6589567          YES
## 2           YES   2.1337241          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         2  YES          2     2       100        100
```

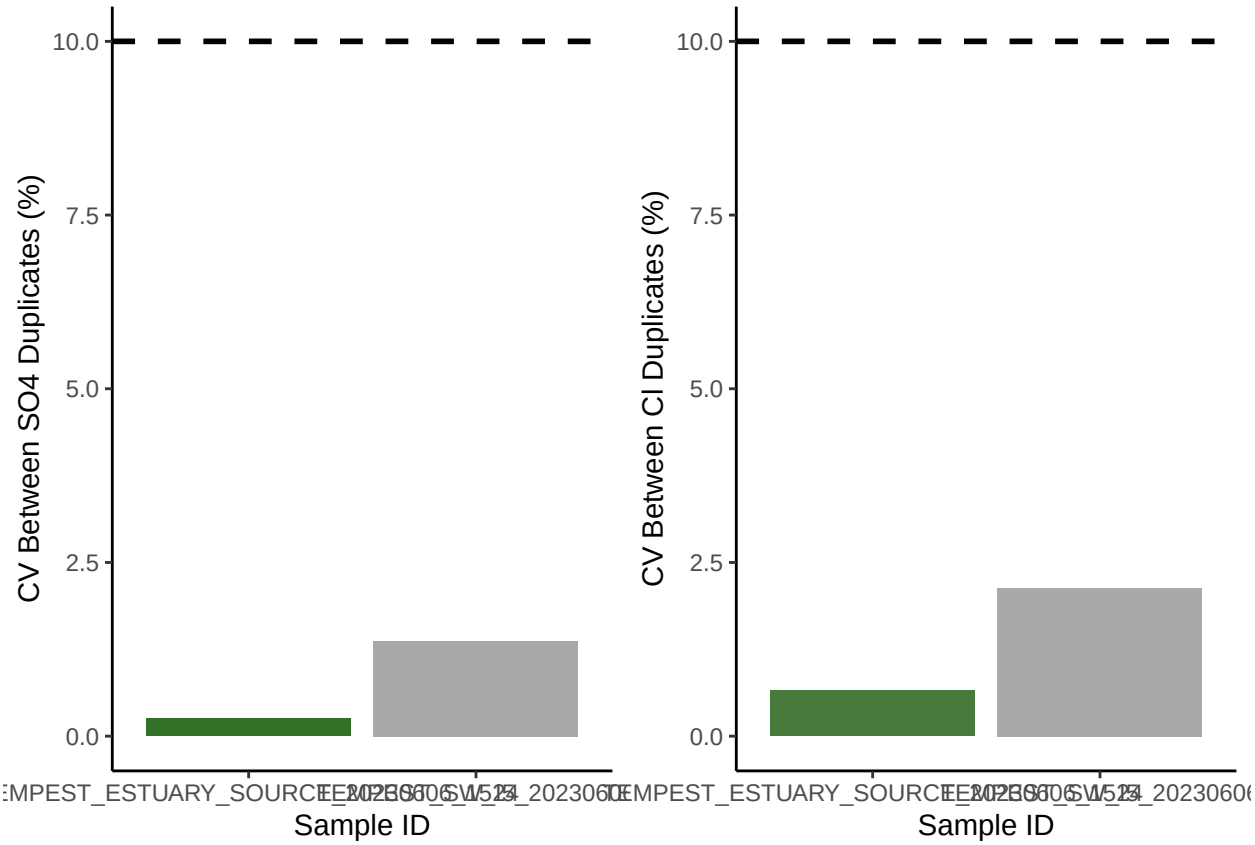
Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

```
##                             Sample_ID  SO4_ppm SO4_Area   Cl_ppm  Cl_Area
## 1 TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.058 110.1445
## 2               TEMPEST_SW_I5_20230606 236.2910   5.5051 2285.850  64.0963
##      SO4_mM     Cl_mM salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk
## 1 16.190003 110.80559 7.098027  16.131276 110.07782     7.051408    0.3633977
## 2  7.370274  64.48097 4.130558   7.514286  65.87165     4.219642    1.9350418
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1           YES   0.6589567          YES    0.256961              YES
## 2           YES   2.1337241          YES    1.368281              YES
```

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


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

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




```
#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         2  YES          2     2       100        100
```

Pull out spikes and check with dionex output conc.

```
##Merge sample data with dilution data
Dilutions <- read.csv("Raw Data/TMP 2023 Sample Dilutions from Dionex.csv")
```

```r
sampledat_dils <- merge(sampledat, Dilutions, by = "Sample_ID")
unique(sampledat_dils$Dilution)
```

```
## [1] 50.00  1.25  1.10
```

```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          TEMEPST_SW_SOURCE_20230606_1301 556.0527  12.9549 4182.4438 117.2776
## 2     TEMPEST_ESTUARY_SOURCE_20230606_0632 557.6408  12.9919 4194.5484 117.6170
## 3     TEMPEST_ESTUARY_SOURCE_20230606_1236 480.7045  11.1994 3636.5862 101.9715
## 4     TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.0580 110.1445
## 5 TEMPEST_ESTUARY_SOURCE_20230606_1524_dup 517.1687  12.0490 3902.2588 109.4211
## 6                   TEMPEST_FW_B4_20230606   2.2280   2.0763  173.8469 194.9897
##       SO4_mM      Cl_mM  salinity Dilution
## 1 17.3441266 117.981489 7.5577019    50.00
## 2 17.3936619 118.322945 7.5795750    50.00
## 3 14.9939021 102.583532 6.5713373    50.00
## 4 16.1900031 110.805585 7.0980268    50.00
## 5 16.1312757 110.077822 7.0514077    50.00
## 6  0.0694947   4.904003 0.3141673     1.25
```

```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_I5_20230606_spk 16.5901  15.4606 31.4901 35.3198
## 2 TEMPEST_FW_SOURCE_20230606_1511_spk 16.3183  17.2810 24.0314 30.6296
##      SO4_mM     Cl_mM   salinity Dilution
## 1 0.5174704 0.8882962 0.05692861     1.25
## 2 0.5089925 0.6778956 0.04345074     1.10
```

```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_I5_20230606  0.5174704
## 2 TEMPEST_FW_SOURCE_20230606_1511  0.5089925
```

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

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

```
## 1          TEMPEST_FW_I5_20230606 12.8534  11.9783 31.6279 35.4744 0.4009170
## 2 TEMPEST_FW_SOURCE_20230606_1511 13.9474  14.7702 24.5413 31.2795 0.4350405
##       Cl_mM   salinity Dilution SO4_mM_spk
## 1 0.8921834 0.05717762     1.25  0.5174704
## 2 0.6922793 0.04437213     1.10  0.5089925
```

```
#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          TEMPEST_FW_I5_20230606 12.8534  11.9783 31.6279 35.4744 0.4009170
## 2 TEMPEST_FW_SOURCE_20230606_1511 13.9474  14.7702 24.5413 31.2795 0.4350405
##       Cl_mM   salinity Dilution SO4_mM_spk SO4_spk_Conc
## 1 0.8921834 0.05717762     1.25  0.5174704 7.797879e-05
## 2 0.6922793 0.04437213     1.10  0.5089925 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] 1.25 1.10
```

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

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

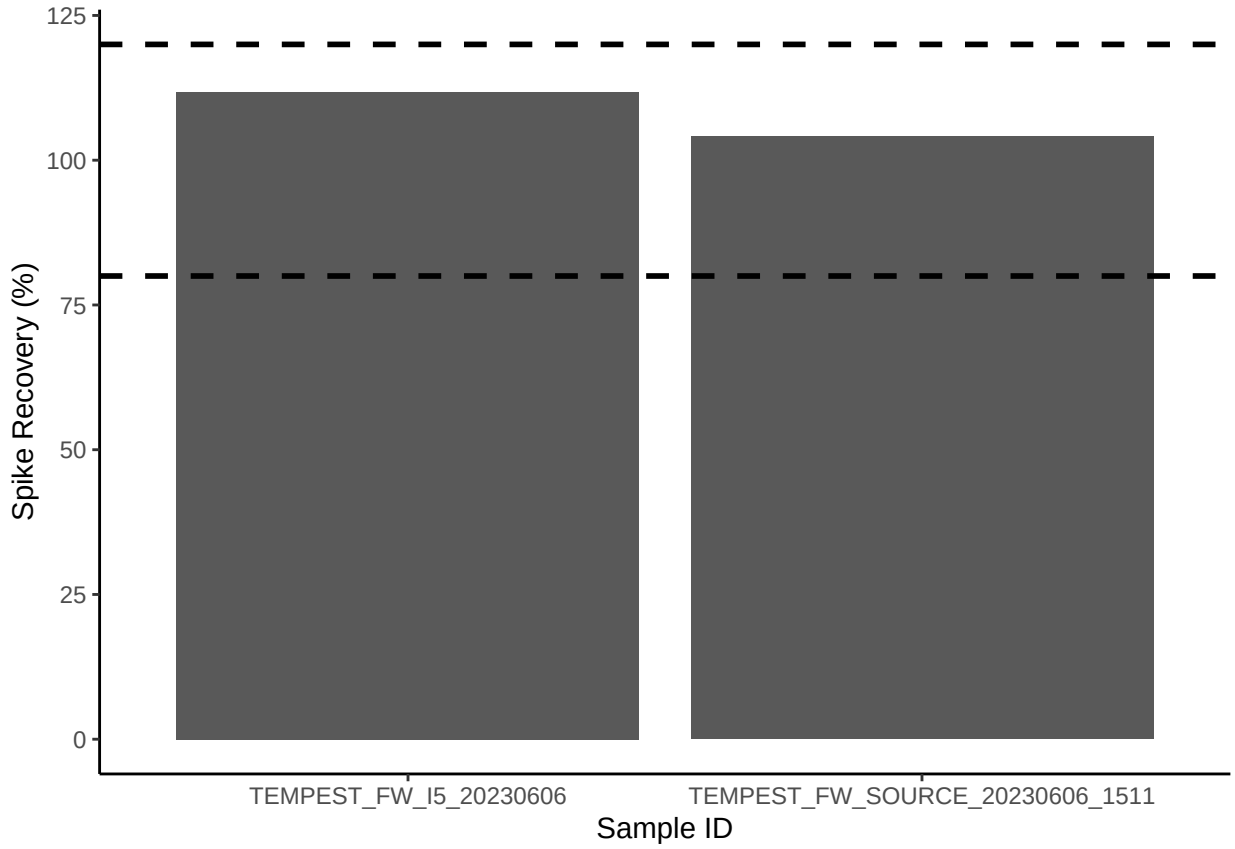
```
##                          Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area    SO4_mM
## 1          TEMPEST_FW_I5_20230606 12.8534  11.9783 31.6279 35.4744 0.4009170
## 2 TEMPEST_FW_SOURCE_20230606_1511 13.9474  14.7702 24.5413 31.2795 0.4350405
##       Cl_mM   salinity Dilution SO4_mM_spk SO4_spk_Conc SampleVol
## 1 0.8921834 0.05717762     1.25  0.5174704 7.797879e-05    0.0015
## 2 0.6922793 0.04437213     1.10  0.5089925 7.797879e-05    0.0015
```

```
#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    SO4_mM
## 1          TEMPEST_FW_I5_20230606 12.8534  11.9783 31.6279 35.4744 0.4009170
## 2 TEMPEST_FW_SOURCE_20230606_1511 13.9474  14.7702 24.5413 31.2795 0.4350405
##       Cl_mM   salinity Dilution SO4_mM_spk SO4_spk_Conc SampleVol
## 1 0.8921834 0.05717762     1.25  0.5174704 7.797879e-05    0.0015
## 2 0.6922793 0.04437213     1.10  0.5089925 7.797879e-05    0.0015
##   SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1     0.0004811004   0.0006251042    0.0005590792     111.8096           YES
## 2     0.0005932371   0.0006987079    0.0006712159     104.0959           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                 2     2     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  1.25  1.10
```

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

```
##                                 Sample_ID  SO4_ppm SO4_Area    Cl_ppm  Cl_Area
## 1          TEMEPST_SW_SOURCE_20230606_1301 556.0527  12.9549 4182.4438 117.2776
## 2     TEMPEST_ESTUARY_SOURCE_20230606_0632 557.6408  12.9919 4194.5484 117.6170
## 3     TEMPEST_ESTUARY_SOURCE_20230606_1236 480.7045  11.1994 3636.5862 101.9715
## 4     TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.0580 110.1445
## 5 TEMPEST_ESTUARY_SOURCE_20230606_1524_dup 517.1687  12.0490 3902.2588 109.4211
## 6                   TEMPEST_FW_B4_20230606   2.2280   2.0763  173.8469 194.9897
##       SO4_mM      Cl_mM  salinity Dilution
## 1 17.3441266 117.981489 7.5577019    50.00
## 2 17.3936619 118.322945 7.5795750    50.00
## 3 14.9939021 102.583532 6.5713373    50.00
## 4 16.1900031 110.805585 7.0980268    50.00
## 5 16.1312757 110.077822 7.0514077    50.00
## 6  0.0694947   4.904003 0.3141673     1.25
```

```
#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          TEMEPST_SW_SOURCE_20230606_1301 556.0527  12.9549 4182.4438 117.2776
## 2     TEMPEST_ESTUARY_SOURCE_20230606_0632 557.6408  12.9919 4194.5484 117.6170
## 3     TEMPEST_ESTUARY_SOURCE_20230606_1236 480.7045  11.1994 3636.5862 101.9715
## 4     TEMPEST_ESTUARY_SOURCE_20230606_1524 519.0515  12.0928 3928.0580 110.1445
## 5 TEMPEST_ESTUARY_SOURCE_20230606_1524_dup 517.1687  12.0490 3902.2588 109.4211
## 6                   TEMPEST_FW_B4_20230606   2.2280   2.0763  173.8469 194.9897
##       SO4_mM      Cl_mM  salinity Dilution  SO4_ugmL   Cl_ugmL
## 1 17.3441266 117.981489 7.5577019    50.00 11.121064  83.64890
## 2 17.3936619 118.322945 7.5795750    50.00 11.152826  83.89098
## 3 14.9939021 102.583532 6.5713373    50.00  9.614066  72.73174
## 4 16.1900031 110.805585 7.0980268    50.00 10.381000  78.56118
## 5 16.1312757 110.077822 7.0514077    50.00 10.343400  78.04521
## 6  0.0694947   4.904003 0.3141673     1.25  1.782401 139.07750
```

```
#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_I5_20230606_spk 16.5901  15.4606 31.4901 35.3198
## 2 TEMPEST_FW_SOURCE_20230606_1511_spk 16.3183  17.2810 24.0314 30.6296
##      SO4_mM     Cl_mM   salinity Dilution SO4_ugmL  Cl_ugmL
## 1 0.5174704 0.8882962 0.05692861     1.25 13.27207 25.19204
## 2 0.5089925 0.6778956 0.04345074     1.10 14.83478 21.84672
```

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

```
##                             Sample_ID Dilution SO4_ugmL
## 1          TEMPEST_FW_I5_20230606_spk     1.25 13.27207
## 2 TEMPEST_FW_SOURCE_20230606_1511_spk     1.10 14.83478
```

```
#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_I5_20230606     1.25     13.27207
## 2 TEMPEST_FW_SOURCE_20230606_1511     1.10     14.83478
```

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

```
##                         Sample_ID Dilution SO4_ppm SO4_Area  Cl_ppm Cl_Area
## 1          TEMPEST_FW_I5_20230606     1.25 12.8534  11.9783 31.6279 35.4744
## 2 TEMPEST_FW_SOURCE_20230606_1511     1.10 13.9474  14.7702 24.5413 31.2795
##      SO4_mM     Cl_mM   salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk
## 1 0.4009170 0.8921834 0.05717762 10.28271 25.30230     13.27207
## 2 0.4350405 0.6922793 0.04437213 12.67940 22.31027     14.83478
```

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

```
##                         Sample_ID Dilution SO4_ppm SO4_Area  Cl_ppm Cl_Area
## 1          TEMPEST_FW_I5_20230606     1.25 12.8534  11.9783 31.6279 35.4744
## 2 TEMPEST_FW_SOURCE_20230606_1511     1.10 13.9474  14.7702 24.5413 31.2795
##      SO4_mM     Cl_mM   salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 0.4009170 0.8921834 0.05717762 10.28271 25.30230     13.27207          2.5
## 2 0.4350405 0.6922793 0.04437213 12.67940 22.31027     14.83478          2.5
```

```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$Sample_ID, "MSM_UP"), 50, QAspks$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$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "TEMPEST"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 1501, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 1475, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 1462, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$SampleVol)
# QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$SampleVol)

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

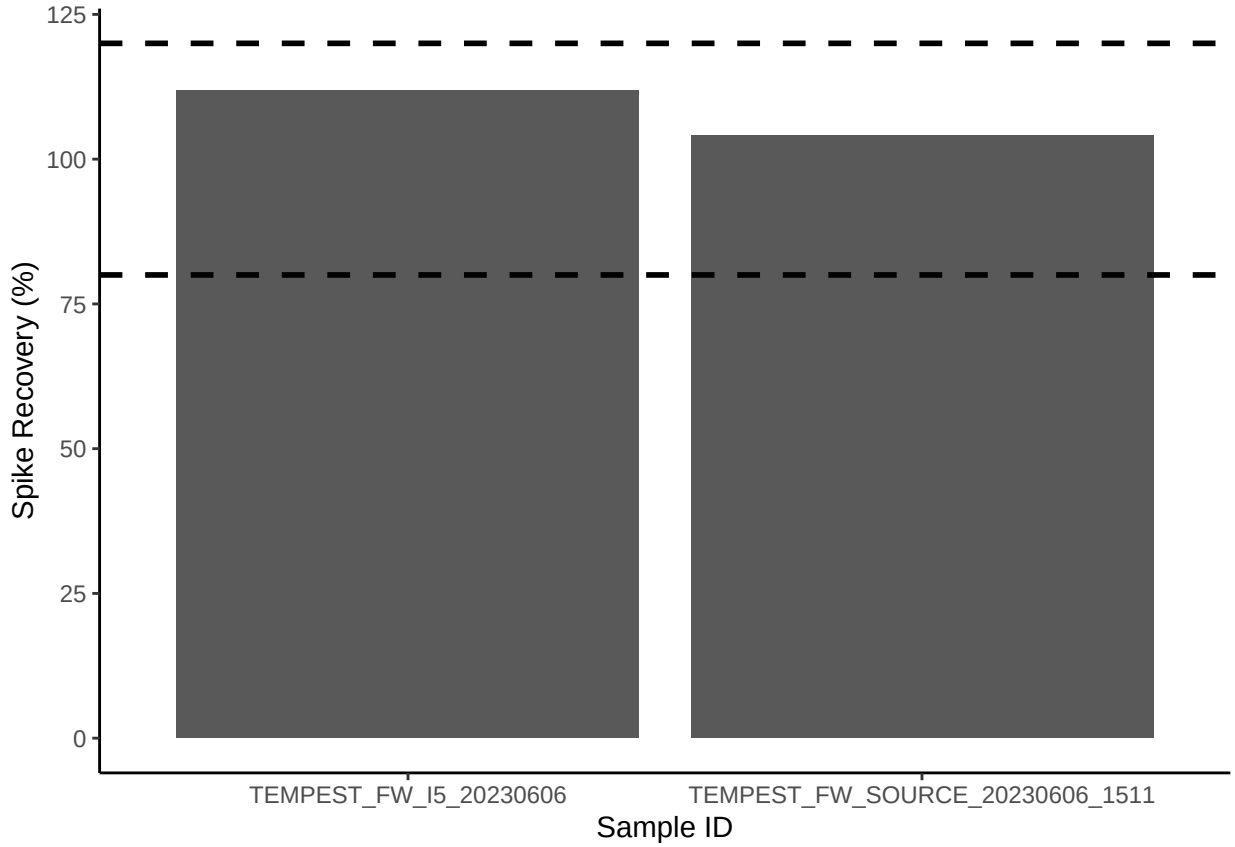
```
##                         Sample_ID Dilution SO4_ppm SO4_Area  Cl_ppm Cl_Area
## 1          TEMPEST_FW_I5_20230606     1.25 12.8534  11.9783 31.6279 35.4744
## 2 TEMPEST_FW_SOURCE_20230606_1511     1.10 13.9474  14.7702 24.5413 31.2795
##      SO4_mM     Cl_mM   salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 0.4009170 0.8921834 0.05717762 10.28271 25.30230     13.27207          2.5
```

```
## 2 0.4350405 0.6922793 0.04437213 12.67940 22.31027     14.83478          2.5
##   SampleVol
## 1       1.5
## 2       1.5
```

```
#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 Cl_Area
## 1          TEMPEST_FW_I5_20230606     1.25 12.8534  11.9783 31.6279 35.4744
## 2 TEMPEST_FW_SOURCE_20230606_1511     1.10 13.9474  14.7702 24.5413 31.2795
##      SO4_mM     Cl_mM   salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 0.4009170 0.8921834 0.05717762 10.28271 25.30230     13.27207          2.5
## 2 0.4350405 0.6922793 0.04437213 12.67940 22.31027     14.83478          2.5
##   SampleVol SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery
## 1       1.5         15.42406       20.04082        17.92406     111.8096
## 2       1.5         19.01909       22.40051        21.51909     104.0960
##   SO4_spks_flag
## 1           YES
## 2           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                 2     2     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)))
```

```
## Warning in rbind(c("TEMEPST", "SW", "SOURCE", "20230606", "1301"), c("TEMPEST",
## : number of columns of result is not a multiple of vector length (arg 5)
```

```r
colnames(IDs) <- c("Project" , "Plot","Grid", "Date", "Time")
head(IDs)
```

```
##   Project    Plot   Grid     Date    Time
## 1 TEMEPST      SW SOURCE 20230606    1301
## 2 TEMPEST ESTUARY SOURCE 20230606    0632
## 3 TEMPEST ESTUARY SOURCE 20230606    1236
## 4 TEMPEST ESTUARY SOURCE 20230606    1524
## 5 TEMPEST      FW     B4 20230606 TEMPEST
## 6 TEMPEST      FW     C3 20230606 TEMPEST
```

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

```
##   Project    Plot   Grid     Date    Time                             Sample_ID
## 1 TEMEPST      SW SOURCE 20230606    1301      TEMEPST_SW_SOURCE_20230606_1301
## 2 TEMPEST ESTUARY SOURCE 20230606    0632 TEMPEST_ESTUARY_SOURCE_20230606_0632
## 3 TEMPEST ESTUARY SOURCE 20230606    1236 TEMPEST_ESTUARY_SOURCE_20230606_1236
## 4 TEMPEST ESTUARY SOURCE 20230606    1524 TEMPEST_ESTUARY_SOURCE_20230606_1524
## 5 TEMPEST      FW     B4 20230606 TEMPEST               TEMPEST_FW_B4_20230606
## 6 TEMPEST      FW     C3 20230606 TEMPEST               TEMPEST_FW_C3_20230606
##    SO4_ppm SO4_Area    Cl_ppm  Cl_Area     SO4_mM       Cl_mM   salinity
## 1 556.0527  12.9549 4182.4438 117.2776 17.3441266 117.9814894 7.55770195
## 2 557.6408  12.9919 4194.5484 117.6170 17.3936619 118.3229450 7.57957496
## 3 480.7045  11.1994 3636.5862 101.9715 14.9939021 102.5835317 6.57133726
## 4 519.0515  12.0928 3928.0580 110.1445 16.1900031 110.8055853 7.09802681
## 5   2.2280   2.0763  173.8469 194.9897  0.0694947   4.9040028 0.31416735
## 6  12.2001  12.9198   27.3401  34.8467  0.3805396   0.7712299 0.04942956
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_TEMPEST_Dionex_ClSo4_Final_20230607.csv")          #Cha
```

END