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)

```r
# SULFATE DATA:

## Read in raw data file from Dionex - copied and saved as a txt
Sdat <- read.table("Raw Data/COMPASS_TEMPEST_20230609_SO4.txt",sep='\t' , header=T, skip=3)
head(Sdat)
```

```
##   X        X.1                X.2 IC.SO4 IC.SO4.1 IC.SO4.2 IC.SO4.3 IC.SO4.4
## 1 1  Lab Blank             Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 2 2  Lab Blank             Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 3 3  Lab Blank             Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 4 4 Standard 1 Calibration Standard  4.400   0.4517     6.08   0.5351     2.55
## 5 5 Standard 2 Calibration Standard  4.397   0.9638     6.53   1.1415     5.33
## 6 6 Standard 3 Calibration Standard  4.393   1.9094     7.18   2.2616    10.33
##   IC.SO4.5
## 1     n.a.
## 2     n.a.
## 3     n.a.
## 4        M
## 5        M
## 6        M
```

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

```
##          X.1 IC.SO4.1 IC.SO4.3
## 1  Lab Blank     n.a.     n.a.
## 2  Lab Blank     n.a.     n.a.
## 3  Lab Blank     n.a.     n.a.
## 4 Standard 1   0.4517   0.5351
## 5 Standard 2   0.9638   1.1415
## 6 Standard 3   1.9094   2.2616
```

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

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

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

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

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

```
##     Sample_ID SO4_ppm SO4_Area
## 1   Lab Blank      NA       NA
## 2   Lab Blank      NA       NA
## 3   Lab Blank      NA       NA
## 4  Standard 1  0.4517   0.5351
## 5  Standard 2  0.9638   1.1415
## 6  Standard 3  1.9094   2.2616
```

```r
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_TEMPEST_20230609_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  3.470   0.0141   22.51   0.0202    0.10
## 2 2  Lab Blank              Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 3 3  Lab Blank              Unknown  3.467   0.0076   17.96   0.0109    0.06
## 4 4 Standard 1 Calibration Standard  3.470   5.7586   93.79   8.2504   45.07
## 5 5 Standard 2 Calibration Standard  3.467  11.3682   93.12  16.2873   88.33
## 6 6 Standard 3 Calibration Standard  3.467  20.3417   92.57  29.1437  156.75
##   IC.Cl.5
## 1      MB
## 2    n.a.
## 3      MB
## 4      BM
## 5      BM
## 6      BM
```

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

```
##          X.1   IC.Cl.1  IC.Cl.3
## 1  Lab Blank   0.0141   0.0202
## 2  Lab Blank     n.a.     n.a.
## 3  Lab Blank   0.0076   0.0109
## 4 Standard 1   5.7586   8.2504
## 5 Standard 2  11.3682  16.2873
## 6 Standard 3  20.3417  29.1437
```

```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   0.0141  0.0202
## 2  Lab Blank       NA      NA
## 3  Lab Blank   0.0076  0.0109
## 4 Standard 1   5.7586  8.2504
## 5 Standard 2  11.3682 16.2873
## 6 Standard 3  20.3417 29.1437
```

```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            457.3600  28.7072 3706.9076 287.1185
## 2            457.3600  28.7072 1073.4812  62.2928
## 3            457.3600  28.7072    0.0076   0.0109
## 4            457.3600  28.7072 1469.0566  71.5412
## 5            457.3600  28.7072        NA       NA
## 6            142.4999   9.0068 3706.9076 287.1185
```

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

```
##     Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area
## 26 Lab Blank      NA       NA 0.0141  0.0202
## 27 Lab Blank      NA       NA     NA      NA
## 28 Lab Blank      NA       NA 0.0076  0.0109
## 29 Lab Blank      NA       NA 0.0091  0.0130
## 30 Lab Blank      NA       NA 0.0160  0.0229
## 31 Lab Blank      NA       NA 0.0224  0.0321
```

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

```
##         Sample_ID SO4_ppm SO4_Area Cl_ppm Cl_Area
## 107 Standard 1  0.4633   0.5488 5.7703  8.2671
## 108 Standard 1  0.4633   0.5488 5.7586  8.2504
## 109 Standard 1  0.4633   0.5488 5.7883  8.2930
## 110 Standard 1  0.4517   0.5351 5.7703  8.2671
## 111 Standard 1  0.4517   0.5351 5.7586  8.2504
## 112 Standard 1  0.4517   0.5351 5.7883  8.2930
```

```r
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
## 107 Standard 1  0.4633   0.5488 5.7703  8.2671
## 108 Standard 1  0.4633   0.5488 5.7586  8.2504
## 109 Standard 1  0.4633   0.5488 5.7883  8.2930
## 110 Standard 1  0.4517   0.5351 5.7703  8.2671
## 111 Standard 1  0.4517   0.5351 5.7586  8.2504
## 112 Standard 1  0.4517   0.5351 5.7883  8.2930
```

```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  0.456 0.00532    1.17   YES
## 2 Standard 2  0.963 0.000635  0.0659 YES
## 3 Standard 3  1.91  0.00318   0.167   YES
## 4 Standard 4  8.22  2.02      24.6    NO, rerun
## 5 Standard 5 20.0   0.0280    0.140   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   5.77 0.0130  0.224   YES
```

4

```
## 2 Standard 2  11.4  0.00803 0.0706 YES
## 3 Standard 3  20.4  0.0289  0.142  YES
## 4 Standard 4 100.   0.0318  0.0317 YES
## 5 Standard 5 200.   0.381   0.191  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   Cl_Area
## 132      TEMEPST_FW_H6_20230608   10.0574  10.8297    0.0000    0.0000
## 133      TEMEPST_SW_C3_20230608  436.0629  10.3300 3597.9994 103.0974
## 134      TEMPEST_FW_B4_20230608   28.2839  28.7072    0.0000    0.0000
## 135      TEMPEST_FW_C3_20230608   11.6026  12.4935   24.1648   31.4737
## 136 TEMPEST_FW_C3_20230608_dup   11.5730  12.4616   24.3234   31.6803
## 137      TEMPEST_FW_D5_20230608    3.1147   3.3538   27.7577   36.1532
```

```
# 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
```

```
## 132       TEMEPST_FW_H6_20230608  10.0574  10.8297     0.0000    0.0000  0.31370555
## 133       TEMEPST_SW_C3_20230608 436.0629  10.3300  3597.9994 103.0974 13.60146288
## 134       TEMPEST_FW_B4_20230608  28.2839  28.7072     0.0000    0.0000  0.88221772
## 135       TEMPEST_FW_C3_20230608  11.6026  12.4935    24.1648   31.4737  0.36190268
## 136 TEMPEST_FW_C3_20230608_dup  11.5730  12.4616    24.3234   31.6803  0.36097941
## 137       TEMPEST_FW_D5_20230608   3.1147   3.3538    27.7577   36.1532  0.09715221
##            Cl_mM   salinity
## 132   0.0000000 0.00002600
## 133 101.4950465 6.50161092
## 134   0.0000000 0.00002600
## 135   0.6816587 0.04369179
## 136   0.6861326 0.04397838
## 137   0.7830099 0.05018416
```

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
## 132       TEMEPST_FW_H6_20230608  10.0574  10.8297    0.0000    0.0000  0.31370555
## 133       TEMEPST_SW_C3_20230608 436.0629  10.3300 3597.9994 103.0974 13.60146288
## 134       TEMPEST_FW_B4_20230608  28.2839  28.7072    0.0000    0.0000  0.88221772
## 135       TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648   31.4737  0.36190268
## 136 TEMPEST_FW_C3_20230608_dup  11.5730  12.4616   24.3234   31.6803  0.36097941
## 137       TEMPEST_FW_D5_20230608   3.1147   3.3538   27.7577   36.1532  0.09715221
##            Cl_mM   salinity
## 132   0.0000000 0.00002600
## 133 101.4950465 6.50161092
## 134   0.0000000 0.00002600
## 135   0.6816587 0.04369179
## 136   0.6861326 0.04397838
## 137   0.7830099 0.05018416
```

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

```
##                          Sample_ID  SO4_ppm SO4_Area    Cl_ppm  Cl_Area
## 1       TEMPEST_FW_C3_20230608_dup  11.5730  12.4616   24.3234   31.6803
## 2 TEMPEST_SW_I5_20230607_0800_dup   0.0000   0.0000    0.0000    0.0000
## 3       TEMPEST_SW_I5_20230608_dup 420.3553   9.9579 3651.4021 104.6276
##        SO4_mM       Cl_mM   salinity
## 1   0.3609794   0.6861326 0.04397838
## 2   0.0000000   0.0000000 0.00002600
## 3 13.1115190 103.0014697 6.59810959
```

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

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

```
##                Sample_ID SO4_ppm SO4_Area    Cl_ppm   Cl_Area      SO4_mM
## 1 TEMEPST_FW_H6_20230608  10.0574  10.8297    0.0000    0.0000  0.31370555
## 2 TEMEPST_SW_C3_20230608 436.0629  10.3300 3597.9994 103.0974 13.60146288
## 3 TEMPEST_FW_B4_20230608  28.2839  28.7072    0.0000    0.0000  0.88221772
## 4 TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648  31.4737  0.36190268
## 5 TEMPEST_FW_D5_20230608   3.1147   3.3538   27.7577  36.1532  0.09715221
## 6 TEMPEST_FW_E3_20230608  23.9154  24.2733    0.0000    0.0000  0.74595758
##        Cl_mM   salinity
## 1   0.0000000 0.00002600
## 2 101.4950465 6.50161092
## 3   0.0000000 0.00002600
## 4   0.6816587 0.04369179
## 5   0.7830099 0.05018416
## 6   0.0000000 0.00002600
```

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

```
##                    Sample_ID SO4_mM_dup   Cl_mM_dup salinity_dup
## 1      TEMPEST_FW_C3_20230608  0.3609794   0.6861326   0.04397838
## 2 TEMPEST_SW_I5_20230607_0800  0.0000000   0.0000000   0.00002600
## 3      TEMPEST_SW_I5_20230608 13.1115190 103.0014697   6.59810959
```

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

```
##                Sample_ID SO4_ppm SO4_Area    Cl_ppm   Cl_Area     SO4_mM
## 1 TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648  31.4737  0.3619027
## 2 TEMPEST_SW_I5_20230608 413.5339   9.7963 3579.4092 102.5647 12.8987492
##        Cl_mM   salinity SO4_mM_dup   Cl_mM_dup salinity_dup
## 1   0.6816587 0.04369179  0.3609794   0.6861326   0.04397838
## 2 100.9706403 6.46801842 13.1115190 103.0014697   6.59810959
```

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

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

head(QAdups)
```

```
##                Sample_ID SO4_ppm SO4_Area    Cl_ppm   Cl_Area     SO4_mM
## 1 TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648  31.4737  0.3619027
## 2 TEMPEST_SW_I5_20230608 413.5339   9.7963 3579.4092 102.5647 12.8987492
```

```
##           Cl_mM    salinity SO4_mM_dup   Cl_mM_dup salinity_dup SO4_dups_chk
## 1   0.6816587 0.04369179  0.3609794   0.6861326   0.04397838    0.2554411
## 2 100.9706403 6.46801842 13.1115190 103.0014697   6.59810959    1.6360447
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1           YES   0.6541798          YES
## 2           YES   1.9912814          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_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)
```
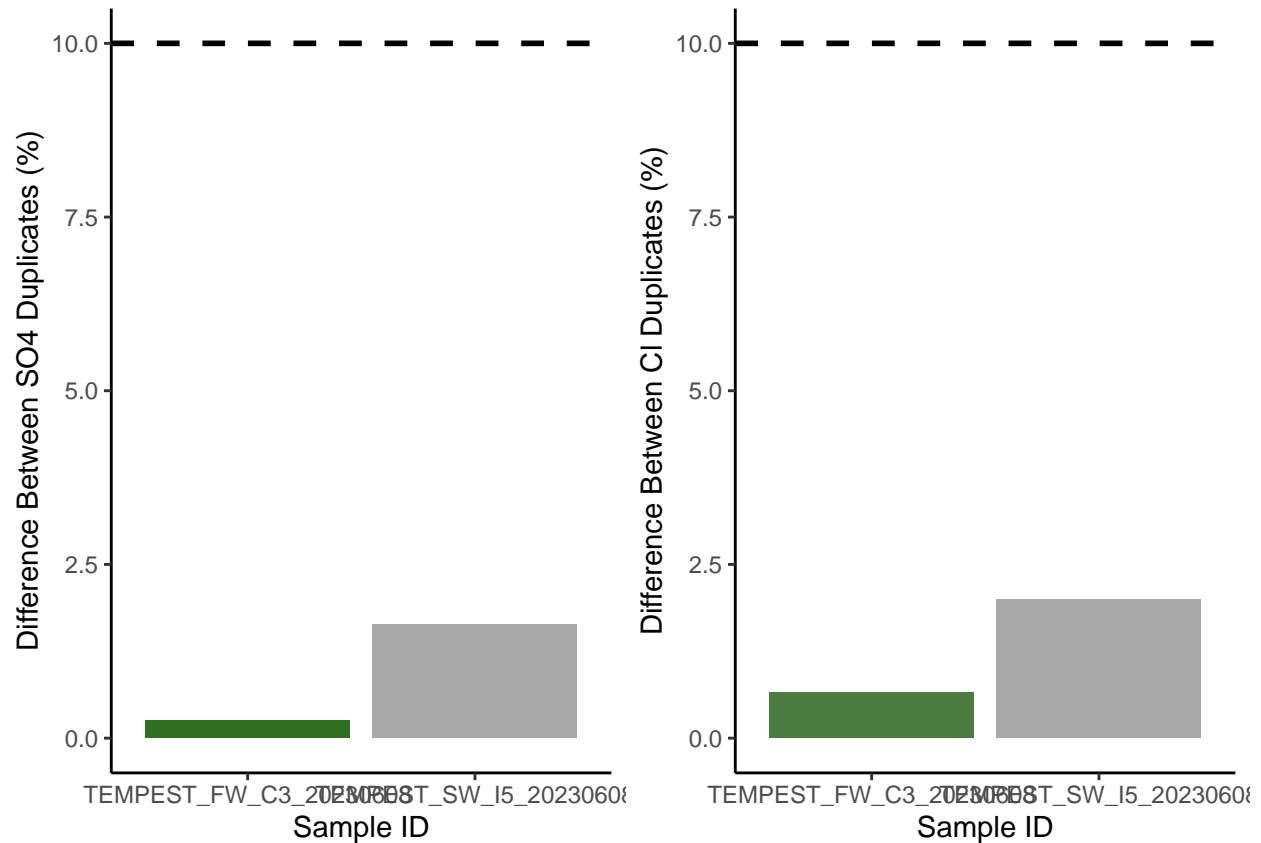
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
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)
```

```r
#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

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

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

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

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

head(QAdups)
```
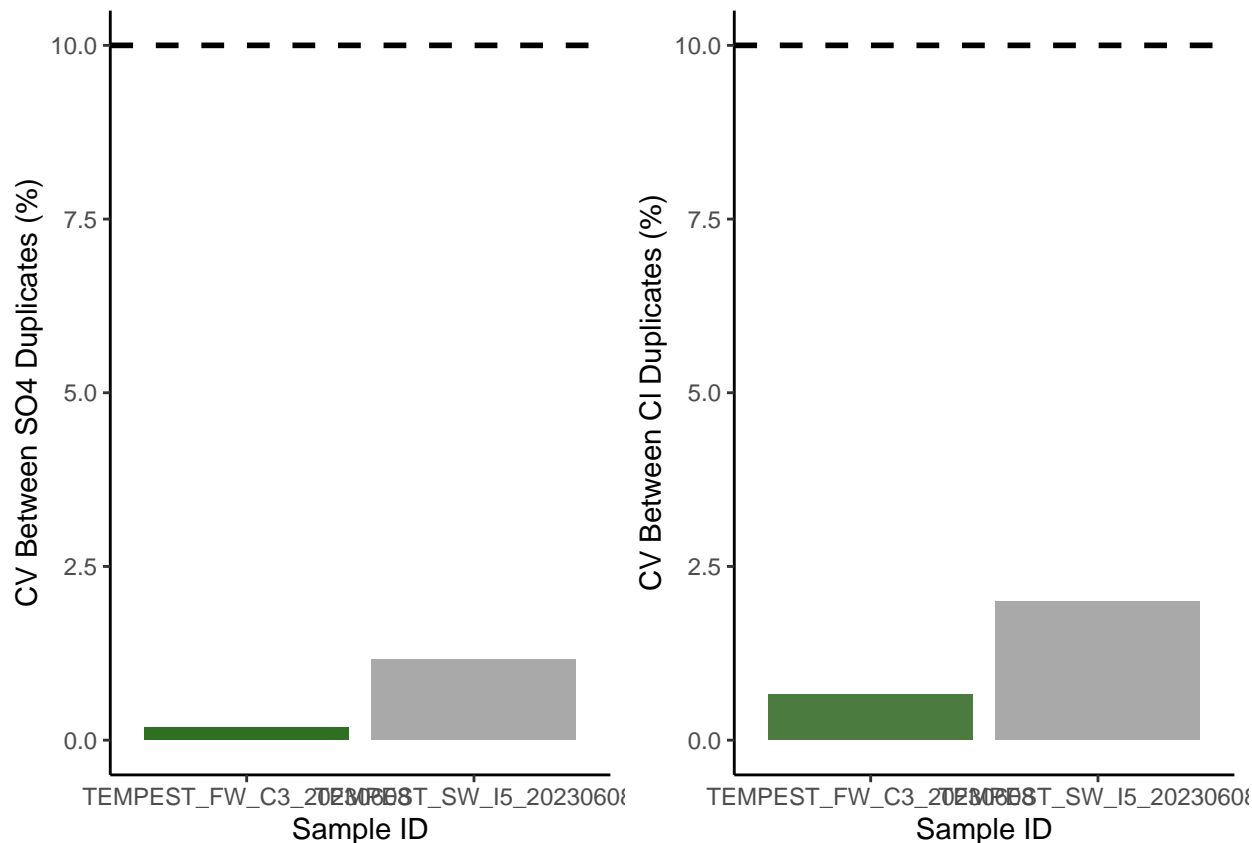
```
##               Sample_ID  SO4_ppm SO4_Area    Cl_ppm   Cl_Area      SO4_mM
## 1 TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648   31.4737   0.3619027
## 2 TEMPEST_SW_I5_20230608 413.5339   9.7963 3579.4092 102.5647  12.8987492
##        Cl_mM   salinity SO4_mM_dup    Cl_mM_dup salinity_dup SO4_dups_chk
## 1   0.6816587 0.04369179  0.3609794    0.6861326   0.04397838    0.2554411
## 2 100.9706403 6.46801842 13.1115190  103.0014697   6.59810959    1.6360447
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1           YES   0.6541798          YES   0.1806241              YES
## 2           YES   1.9912814          YES   1.1568583              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)
```

```r
#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.

```r
##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]  1.100 50.000  1.167 52.740
```

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

```
##                    Sample_ID SO4_ppm SO4_Area    Cl_ppm   Cl_Area       SO4_mM
## 1     TEMEPST_FW_H6_20230608  10.0574  10.8297    0.0000    0.0000   0.31370555
## 2     TEMEPST_SW_C3_20230608 436.0629  10.3300 3597.9994 103.0974  13.60146288
## 3     TEMPEST_FW_B4_20230608  28.2839  28.7072    0.0000    0.0000   0.88221772
## 4     TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648  31.4737   0.36190268
## 5 TEMPEST_FW_C3_20230608_dup  11.5730  12.4616   24.3234  31.6803   0.36097941
## 6     TEMPEST_FW_D5_20230608   3.1147   3.3538   27.7577  36.1532   0.09715221
##          Cl_mM    salinity Dilution
## 1    0.0000000 0.00002600    1.100
## 2  101.4950465 6.50161092   50.000
## 3    0.0000000 0.00002600    1.167
## 4    0.6816587 0.04369179    1.100
## 5    0.6861326 0.04397838    1.100
## 6    0.7830099 0.05018416    1.100
```

```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   SO4_mM
## 1 TEMPEST_SW_E3_20230608_spk 320.6901   7.5969 2356.323 67.5183 10.00281
##      Cl_mM salinity Dilution
## 1 66.46891 4.257901       50
```

```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_SW_E3_20230608   10.00281
```

```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    Cl_mM
## 1 TEMPEST_SW_E3_20230608 243.9557   5.7791 2299.434 65.8882 7.609348 64.86415
##   salinity Dilution SO4_mM_spk
## 1 4.155103       50   10.00281
```

```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    Cl_mM
## 1 TEMPEST_SW_E3_20230608 243.9557   5.7791 2299.434 65.8882 7.609348 64.86415
##   salinity Dilution SO4_mM_spk SO4_spk_Conc
## 1 4.155103       50    10.00281 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
unique(QAspks$Dilution)
```

```
## [1] 50
```

```r
QAspks$SampleVol <- 1
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 50, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 56.85, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 51.16, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 49.27, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 56.51, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 5.15, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.25, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.1, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 52.74, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.167, 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)
```

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

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

```
##            Sample_ID SO4_ppm SO4_Area  Cl_ppm Cl_Area   SO4_mM    Cl_mM
## 1 TEMPEST_SW_E3_20230608 243.9557   5.7791 2299.434 65.8882 7.609348 64.86415
##   salinity Dilution SO4_mM_spk SO4_spk_Conc SampleVol
## 1 4.155103       50   10.00281 7.797879e-05  0.001483
```

```r
#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    Cl_mM
## 1 TEMPEST_SW_E3_20230608 243.9557   5.7791 2299.434 65.8882 7.609348 64.86415
##   salinity Dilution SO4_mM_spk SO4_spk_Conc SampleVol SO4_Total_unspkd
## 1 4.155103       50   10.00281 7.797879e-05  0.001483      0.0002256933
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   0.0002986839    0.0003036721    98.35739          YES
```
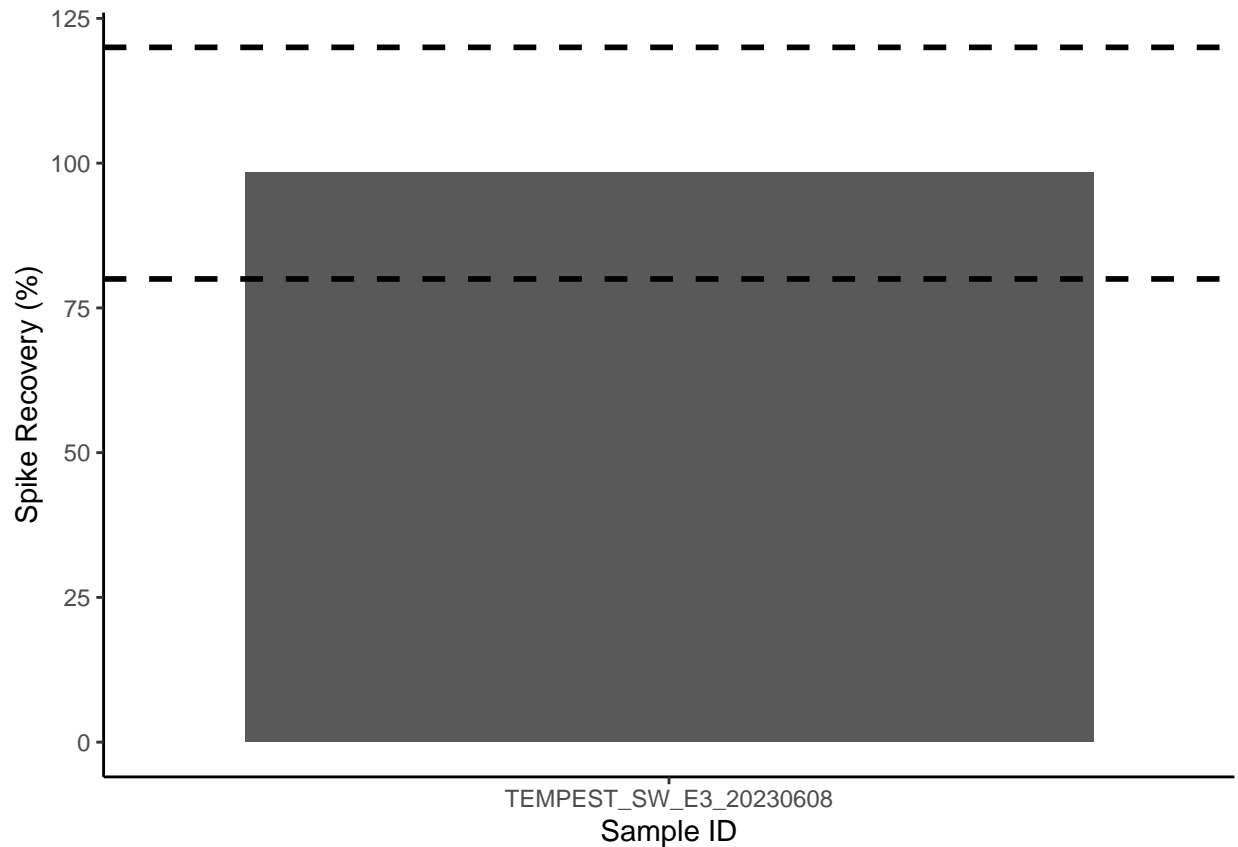
```r
#plot spk recoveries output as a bar graph to easily check - want any over 10% to be red need to work o
spksbar <- ggplot(data = QAspks, aes(x = Sample_ID, y = spk_recovery)) +
      geom_bar(stat = 'identity') +
      scale_fill_manual(values=c("lightblue4")) +
      theme_classic() + labs(x= "Sample 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
```

```r
#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                 1     1     100
```

Pull out spikes and check with area calc

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

```
## [1]  1.100 50.000  1.167 52.740
```

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

```
##                   Sample_ID SO4_ppm SO4_Area    Cl_ppm   Cl_Area      SO4_mM
## 1     TEMEPST_FW_H6_20230608  10.0574  10.8297    0.0000    0.0000  0.31370555
## 2     TEMEPST_SW_C3_20230608 436.0629  10.3300 3597.9994 103.0974 13.60146288
## 3     TEMPEST_FW_B4_20230608  28.2839  28.7072    0.0000    0.0000  0.88221772
## 4     TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648  31.4737  0.36190268
## 5 TEMPEST_FW_C3_20230608_dup  11.5730  12.4616   24.3234  31.6803  0.36097941
## 6     TEMPEST_FW_D5_20230608   3.1147   3.3538   27.7577  36.1532  0.09715221
##         Cl_mM    salinity Dilution
## 1   0.0000000 0.00002600    1.100
## 2 101.4950465 6.50161092   50.000
## 3   0.0000000 0.00002600    1.167
## 4   0.6816587 0.04369179    1.100
## 5   0.6861326 0.04397838    1.100
## 6   0.7830099 0.05018416    1.100
```

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

```
##                   Sample_ID SO4_ppm SO4_Area    Cl_ppm   Cl_Area      SO4_mM
## 1     TEMEPST_FW_H6_20230608  10.0574  10.8297    0.0000    0.0000  0.31370555
## 2     TEMEPST_SW_C3_20230608 436.0629  10.3300 3597.9994 103.0974 13.60146288
## 3     TEMPEST_FW_B4_20230608  28.2839  28.7072    0.0000    0.0000  0.88221772
## 4     TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648  31.4737  0.36190268
## 5 TEMPEST_FW_C3_20230608_dup  11.5730  12.4616   24.3234  31.6803  0.36097941
## 6     TEMPEST_FW_D5_20230608   3.1147   3.3538   27.7577  36.1532  0.09715221
##         Cl_mM    salinity Dilution  SO4_ugmL       Cl_ugmL
## 1   0.0000000 0.00002600    1.100  9.143118 -1.166337e-06
## 2 101.4950465 6.50161092   50.000  8.721239  7.195999e+01
## 3   0.0000000 0.00002600    1.167 24.236439 -1.166337e-06
## 4   0.6816587 0.04369179    1.100 10.547804  2.196803e+01
## 5   0.6861326 0.04397838    1.100 10.520872  2.211223e+01
## 6   0.7830099 0.05018416    1.100  2.831488  2.523423e+01
```

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

```
##                   Sample_ID SO4_ppm SO4_Area   Cl_ppm Cl_Area   SO4_mM
## 1 TEMPEST_SW_E3_20230608_spk 320.6901   7.5969 2356.323 67.5183 10.00281
##      Cl_mM salinity Dilution SO4_ugmL  Cl_ugmL
## 1 66.46891 4.257901       50 6.413783 47.12647
```

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

```
##                 Sample_ID Dilution SO4_ugmL
## 1 TEMPEST_SW_E3_20230608_spk       50 6.413783
```

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

```
##                 Sample_ID Dilution SO4_ugmL_spk
## 1 TEMPEST_SW_E3_20230608       50      6.413783
```

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

```
##                 Sample_ID Dilution  SO4_ppm SO4_Area   Cl_ppm  Cl_Area    SO4_mM
## 1 TEMPEST_SW_E3_20230608       50 243.9557   5.7791 2299.434 65.8882 7.609348
##      Cl_mM salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk
## 1 64.86415 4.155103 4.879081 45.98869     6.413783
```

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

```
##                 Sample_ID Dilution  SO4_ppm SO4_Area   Cl_ppm  Cl_Area    SO4_mM
## 1 TEMPEST_SW_E3_20230608       50 243.9557   5.7791 2299.434 65.8882 7.609348
##      Cl_mM salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 64.86415 4.155103 4.879081 45.98869     6.413783          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
```

```r
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 50, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.1, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 5.15, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 4.53, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 56.51, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.25, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.1, 1500, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 52.74, 1483, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(QAspks$Dilution == 1.167, 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   SO4_mM
## 1 TEMPEST_SW_E3_20230608       50 243.9557   5.7791 2299.434 65.8882 7.609348
##      Cl_mM salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk SO4_spk_Conc SampleVol
## 1 64.86415 4.155103 4.879081 45.98869     6.413783          2.5     1.483
```

```r
#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   SO4_mM
## 1 TEMPEST_SW_E3_20230608       50 243.9557   5.7791 2299.434 65.8882 7.609348
##      Cl_mM salinity SO4_ugmL  Cl_ugmL SO4_ugmL_spk SO4_spk_Conc SampleVol
## 1 64.86415 4.155103 4.879081 45.98869     6.413783          2.5     1.483
```
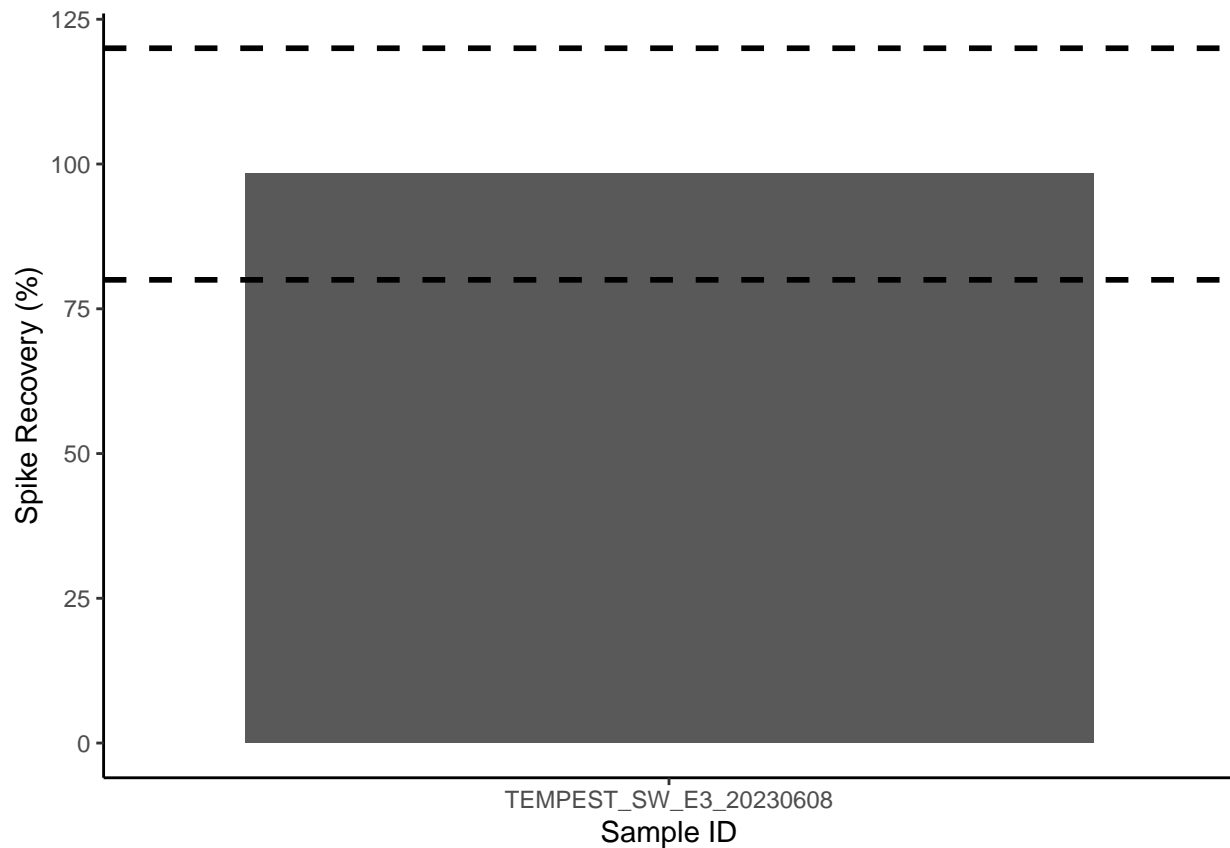
```
##    SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1          7.235676       9.575778        9.735676      98.3576           YES
```

```
#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                 1     1     100
```

Make final dataframe with IDs

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

```
##    Project Plot Grid     Date
## 1 TEMEPST   FW   H6 20230608
## 2 TEMEPST   SW   C3 20230608
## 3 TEMPEST   FW   B4 20230608
## 4 TEMPEST   FW   C3 20230608
## 5 TEMPEST   FW   D5 20230608
## 6 TEMPEST   FW   E3 20230608
```

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

```
##    Project Plot Grid     Date             Sample_ID SO4_ppm SO4_Area    Cl_ppm
## 1 TEMEPST   FW   H6 20230608 TEMEPST_FW_H6_20230608  10.0574  10.8297    0.0000
## 2 TEMEPST   SW   C3 20230608 TEMEPST_SW_C3_20230608 436.0629  10.3300 3597.9994
## 3 TEMPEST   FW   B4 20230608 TEMPEST_FW_B4_20230608  28.2839  28.7072    0.0000
## 4 TEMPEST   FW   C3 20230608 TEMPEST_FW_C3_20230608  11.6026  12.4935   24.1648
## 5 TEMPEST   FW   D5 20230608 TEMPEST_FW_D5_20230608   3.1147   3.3538   27.7577
## 6 TEMPEST   FW   E3 20230608 TEMPEST_FW_E3_20230608  23.9154  24.2733    0.0000
##    Cl_Area       SO4_mM        Cl_mM   salinity
## 1   0.0000   0.31370555    0.0000000 0.00002600
## 2 103.0974  13.60146288  101.4950465 6.50161092
## 3   0.0000   0.88221772    0.0000000 0.00002600
## 4  31.4737   0.36190268    0.6816587 0.04369179
## 5  36.1532   0.09715221    0.7830099 0.05018416
## 6   0.0000   0.74595758    0.0000000 0.00002600
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_TEMPEST_Dionex_ClSo4_Final_20230609.csv")                    #Cha
```

END