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_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  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 5 5 Standard 1 Calibration Standard 5.703   0.4755     6.72   0.6249     3.52
## 6 6 Standard 2 Calibration Standard 5.690   0.9860     6.88   1.2958     7.22
```

```
## 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  0.4755   0.6249
## 6 Standard 2  0.9860   1.2958
```

```
## 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  0.4755   0.6249
## 6 Standard 2  0.9860   1.2958
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_TEMPEST_2023_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 4.473   0.0115   42.99   0.0201    0.11
## 2 2  Lab Blank             Unknown 4.450   0.0112   56.48   0.0196    0.11
## 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 4.443   4.9570   93.28   8.6688   60.00
## 6 6 Standard 2 Calibration Standard 4.497  10.0293   93.06  17.5391  120.14
```

```
## 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  0.0115   0.0201
## 2  Lab Blank  0.0112   0.0196
## 3  Lab Blank     n.a.     n.a.
## 4  Lab Blank     n.a.     n.a.
## 5 Standard 1  4.9570   8.6688
## 6 Standard 2 10.0293  17.5391
```

```
## 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
```

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

```
##    Sample_ID  Cl_ppm Cl_Area
## 1  Lab Blank  0.0115  0.0201
## 2  Lab Blank  0.0112  0.0196
## 3  Lab Blank      NA      NA
## 4  Lab Blank      NA      NA
## 5 Standard 1  4.9570  8.6688
## 6 Standard 2 10.0293 17.5391
```

```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           423.1609  26.8194 3686.5172 356.9371
## 2           423.1609  26.8194  596.4769  34.8846
## 3           423.1609  26.8194    0.0096   0.0013
## 4           423.1609  26.8194  996.8567  65.0799
## 5           423.1609  26.8194        NA       NA
## 6            78.0880   3.1734 3686.5172 356.9371
```

```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      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
```

```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
## 26      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
```

```r
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
## 379 Standard 1  0.4887   0.6422 5.0614  8.8512
## 380 Standard 1  0.4887   0.6422 5.1257  8.9637
## 381 Standard 1  0.4887   0.6422 4.9570  8.6688
## 382 Standard 1  0.4887   0.6422 5.0846  8.8920
## 383 Standard 1  0.4990   0.6558 5.0614  8.8512
## 384 Standard 1  0.4990   0.6558 5.1257  8.9637
```

```
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  0.489 0.00896 1.83  YES
## 2 Standard 2  1.00  0.0122  1.21  YES
## 3 Standard 3  2.02  0.0197  0.975 YES
## 4 Standard 4 10.2   0.0874  0.860 YES
## 5 Standard 5 20.2   0.191   0.947 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   5.06 0.0643 1.27  YES
## 2 Standard 2  10.2  0.104  1.02  YES
## 3 Standard 3  20.3  0.239  1.18  YES
## 4 Standard 4 102.   1.10   1.09  YES
## 5 Standard 5 202.   1.84   0.912 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
## 26      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
```

```
# 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
## 26      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
##        SO4_mM       Cl_mM    salinity
## 26  0.3351716   0.0507165 0.003274805
## 27 10.8818465 101.1889929 6.482005689
## 28 10.7344136 100.6126544 6.445086510
## 29  8.8873487  69.7579464 4.468590994
## 30 11.5254211  69.4091283 4.446246355
## 31  5.0303961  48.3220423 3.095446635
```

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
## 26      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
##        SO4_mM       Cl_mM    salinity
## 26  0.3351716   0.0507165 0.003274805
## 27 10.8818465 101.1889929 6.482005689
## 28 10.7344136 100.6126544 6.445086510
## 29  8.8873487  69.7579464 4.468590994
## 30 11.5254211  69.4091283 4.446246355
## 31  5.0303961  48.3220423 3.095446635
```

```
#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 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486 10.7344136
## 2 20_TEMPEST_FW_C6_20230614_dup   6.5898   0.1732    2.7143   0.0949  0.2055459
## 3 30_TEMPEST_FW_D5_20230617_dup   5.8831   0.1546    2.8461   0.0995  0.1835028
## 4 40_TEMPEST_SW_H6_20230627_dup 134.6079   3.5380 1079.6321  37.7609  4.1986245
## 5 50_TEMPEST_FW_H6_20230627_dup   8.9357   0.2349    1.2131   0.0424  0.2787180
## 6  60_TEMPEST_C_H3_20230628_dup   6.0865   0.1600    0.6513   0.0228  0.1898472
##          Cl_mM    salinity
## 1 100.61265444 6.445086510
## 2   0.07656700 0.004930740
## 3   0.08028491 0.005168903
## 4  30.45506629 1.950921205
## 5   0.03422003 0.002218072
## 6   0.01837236 0.001202899
```

```
#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  1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629  0.3351716
## 2 10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632 10.8818465
## 3 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922  8.8873487
## 4 12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140  5.0303961
```

```
## 5 13_TEMPEST_SW_D5_20230614 194.2448   5.1055 2204.8413  77.1159  6.0587898
## 6 14_TEMPEST_SW_B4_20230614 389.9092  10.2483 3653.5580 127.7859 12.1618590
##         Cl_mM    salinity
## 1   0.0507165 0.003274805
## 2 101.1889929 6.482005689
## 3  69.7579464 4.468590994
## 4  48.3220423 3.095446635
## 5  62.1958054 3.984174229
## 6 103.0622849 6.602005306
```

```
#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 10_TEMPEST_SW_I5_20230614 10.7344136 100.61265444  6.445086510
## 2 20_TEMPEST_FW_C6_20230614  0.2055459   0.07656700  0.004930740
## 3 30_TEMPEST_FW_D5_20230617  0.1835028   0.08028491  0.005168903
## 4 40_TEMPEST_SW_H6_20230627  4.1986245  30.45506629  1.950921205
## 5 50_TEMPEST_FW_H6_20230627  0.2787180   0.03422003  0.002218072
## 6  60_TEMPEST_C_H3_20230628  0.1898472   0.01837236  0.001202899
```

```
#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 10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632 10.8818465
## 2 20_TEMPEST_FW_C6_20230614   6.9070   0.1815    2.6658   0.0932  0.2154398
## 3 30_TEMPEST_FW_D5_20230617   5.7945   0.1523    2.7877   0.0975  0.1807392
## 4 40_TEMPEST_SW_H6_20230627 135.5781   3.5635 1084.5743  37.9338  4.2288865
## 5 50_TEMPEST_FW_H6_20230627   8.9315   0.2348    1.3663   0.0478  0.2785870
## 6  60_TEMPEST_C_H3_20230628   6.2648   0.1647    0.6405   0.0224  0.1954086
##          Cl_mM    salinity SO4_mM_dup    Cl_mM_dup salinity_dup
## 1 101.18899295 6.482005689 10.7344136 100.61265444  6.445086510
## 2   0.07519887 0.004843101  0.2055459   0.07656700  0.004930740
## 3   0.07863752 0.005063374  0.1835028   0.08028491  0.005168903
## 4  30.59447955 1.959851760  4.1986245  30.45506629  1.950921205
## 5   0.03854161 0.002494904  0.2787180   0.03422003  0.002218072
## 6   0.01806770 0.001183383  0.1898472   0.01837236  0.001202899
```

```
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)
```

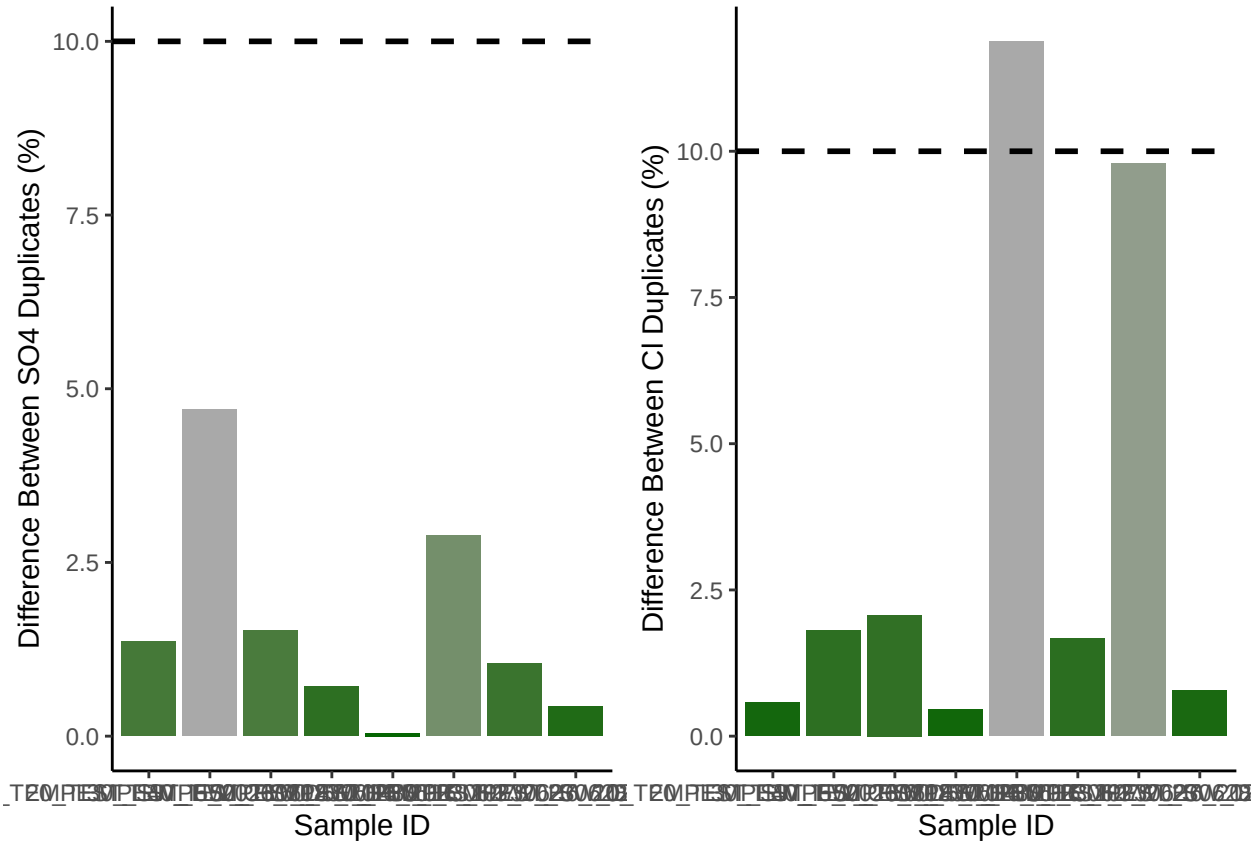
```
##                   Sample_ID  SO4_ppm SO4_Area    Cl_ppm  Cl_Area     SO4_mM
```

```
## 1 10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632 10.8818465
## 2 20_TEMPEST_FW_C6_20230614   6.9070   0.1815    2.6658   0.0932  0.2154398
## 3 30_TEMPEST_FW_D5_20230617   5.7945   0.1523    2.7877   0.0975  0.1807392
## 4 40_TEMPEST_SW_H6_20230627 135.5781   3.5635 1084.5743  37.9338  4.2288865
## 5 50_TEMPEST_FW_H6_20230627   8.9315   0.2348    1.3663   0.0478  0.2785870
## 6  60_TEMPEST_C_H3_20230628   6.2648   0.1647    0.6405   0.0224  0.1954086
##          Cl_mM    salinity SO4_mM_dup    Cl_mM_dup salinity_dup SO4_dups_chk
## 1 101.18899295 6.482005689 10.7344136 100.61265444  6.445086510   1.36409293
## 2   0.07519887 0.004843101  0.2055459   0.07656700  0.004930740   4.70037342
## 3   0.07863752 0.005063374  0.1835028   0.08028491  0.005168903   1.51743509
## 4  30.59447955 1.959851760  4.1986245  30.45506629  1.950921205   0.71817193
## 5   0.03854161 0.002494904  0.2787180   0.03422003  0.002218072   0.04701352
## 6   0.01806770 0.001183383  0.1898472   0.01837236  0.001202899   2.88714548
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1           YES   0.5711931          YES
## 2           YES   1.8029405          YES
## 3           YES   2.0732010          YES
## 4           YES   0.4567217          YES
## 5           YES  11.8787315    NO, rerun
## 6           YES   1.6720855          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)
```

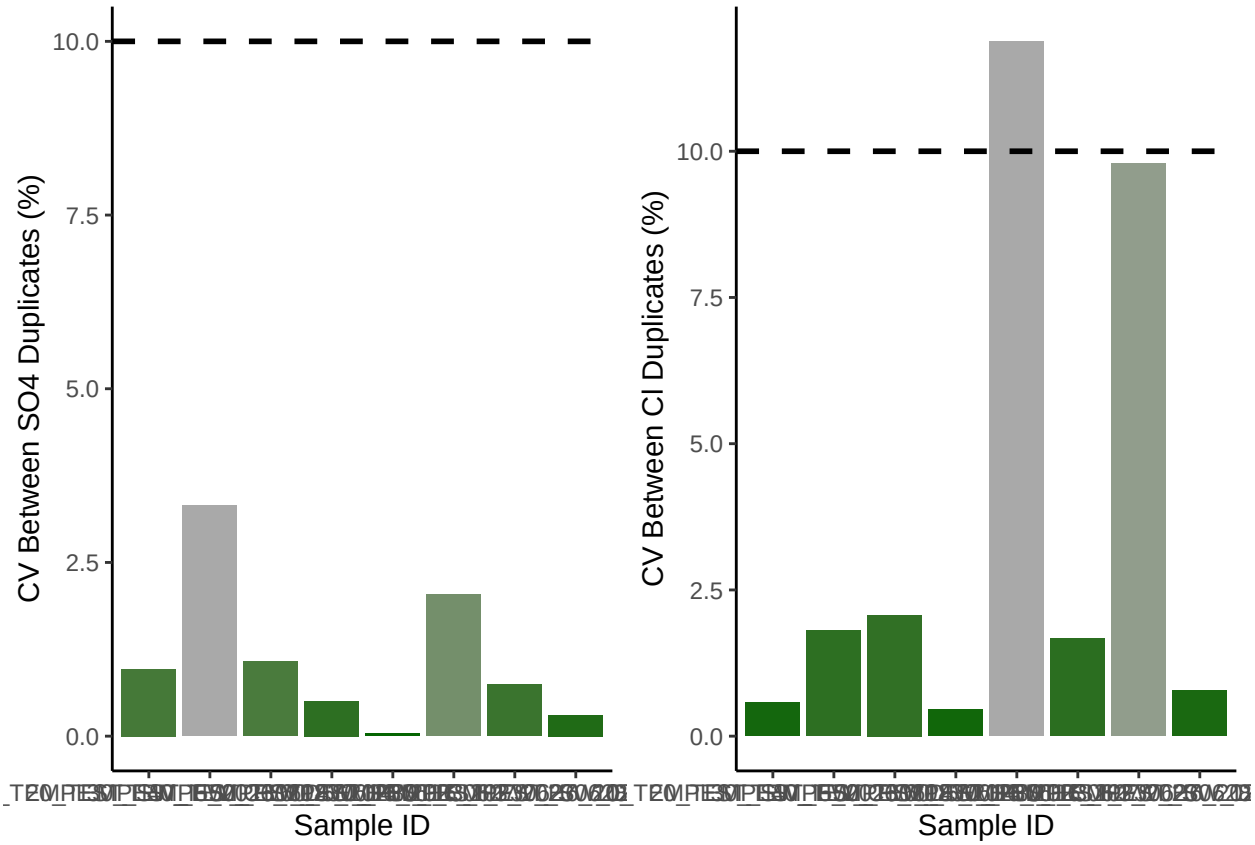
```
##   Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES         8 NO, rerun          1     8       100       12.5
## 2  YES         8       YES          7     8       100       87.5
```

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 10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632 10.8818465
## 2 20_TEMPEST_FW_C6_20230614   6.9070   0.1815    2.6658   0.0932  0.2154398
## 3 30_TEMPEST_FW_D5_20230617   5.7945   0.1523    2.7877   0.0975  0.1807392
## 4 40_TEMPEST_SW_H6_20230627 135.5781   3.5635 1084.5743  37.9338  4.2288865
## 5 50_TEMPEST_FW_H6_20230627   8.9315   0.2348    1.3663   0.0478  0.2785870
## 6  60_TEMPEST_C_H3_20230628   6.2648   0.1647    0.6405   0.0224  0.1954086
##          Cl_mM    salinity SO4_mM_dup    Cl_mM_dup salinity_dup SO4_dups_chk
## 1 101.18899295 6.482005689 10.7344136 100.61265444  6.445086510   1.36409293
## 2   0.07519887 0.004843101  0.2055459   0.07656700  0.004930740   4.70037342
## 3   0.07863752 0.005063374  0.1835028   0.08028491  0.005168903   1.51743509
## 4  30.59447955 1.959851760  4.1986245  30.45506629  1.950921205   0.71817193
## 5   0.03854161 0.002494904  0.2787180   0.03422003  0.002218072   0.04701352
## 6   0.01806770 0.001183383  0.1898472   0.01837236  0.001202899   2.88714548
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1           YES   0.5711931          YES  0.96455936              YES
## 2           YES   1.8029405          YES  3.32366592              YES
## 3           YES   2.0732010          YES  1.07298864              YES
## 4           YES   0.4567217          YES  0.50782424              YES
## 5           YES  11.8787315    NO, rerun  0.03324358              YES
## 6           YES   1.6720855          YES  2.04152015              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         8 NO, rerun          1     8       100       12.5
## 2  YES         8       YES          7     8       100       87.5
```

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
## 26      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
##        SO4_mM       Cl_mM    salinity
## 26  0.3351716   0.0507165 0.003274805
## 27 10.8818465 101.1889929 6.482005689
## 28 10.7344136 100.6126544 6.445086510
## 29  8.8873487  69.7579464 4.468590994
## 30 11.5254211  69.4091283 4.446246355
## 31  5.0303961  48.3220423 3.095446635
```

```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 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536 86.0597 11.525421
## 2 21_TEMPEST_FW_C3_20230614_spk  94.7296   2.4899    3.1081  0.1087  2.954760
## 3 31_TEMPEST_FW_H6_20230617_spk  95.1550   2.5010    3.3125  0.1159  2.968029
## 4 41_TEMPEST_SW_H3_20230627_spk 236.0189   6.2035 1271.7793 44.4814  7.361787
## 5 51_TEMPEST_FW_H3_20230627_spk  96.3773   2.5332    1.6148  0.0565  3.006154
## 6  61_TEMPEST_C_F4_20230628_spk  94.3948   2.4811    0.5600  0.0196  2.944317
##         Cl_mM    salinity
## 1 69.40912835 4.446246355
## 2  0.08767560 0.005642337
## 3  0.09344147 0.006011687
## 4 35.87529760 2.298131195
## 5  0.04555148 0.002943944
## 6  0.01579690 0.001037920
```

```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 11_TEMPEST_SW_F6_20230614  11.525421
## 2 21_TEMPEST_FW_C3_20230614   2.954760
## 3 31_TEMPEST_FW_H6_20230617   2.968029
## 4 41_TEMPEST_SW_H3_20230627   7.361787
## 5 51_TEMPEST_FW_H3_20230627   3.006154
## 6  61_TEMPEST_C_F4_20230628   2.944317
```

```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 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity SO4_mM_spk
## 1 69.757946403 4.4685909944  11.525421
## 2  0.083929478 0.0054023671   2.954760
## 3  0.082964739 0.0053405677   2.968029
## 4 36.525114245 2.3397572471   7.361787
## 5  0.043136812 0.0027892644   3.006154
## 6  0.006637518 0.0004511871   2.944317
```

```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 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity SO4_mM_spk SO4_spk_Conc
## 1 69.757946403 4.4685909944  11.525421 7.797879e-05
## 2  0.083929478 0.0054023671   2.954760 7.797879e-05
## 3  0.082964739 0.0053405677   2.968029 7.797879e-05
## 4 36.525114245 2.3397572471   7.361787 7.797879e-05
## 5  0.043136812 0.0027892644   3.006154 7.797879e-05
## 6  0.006637518 0.0004511871   2.944317 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)
```

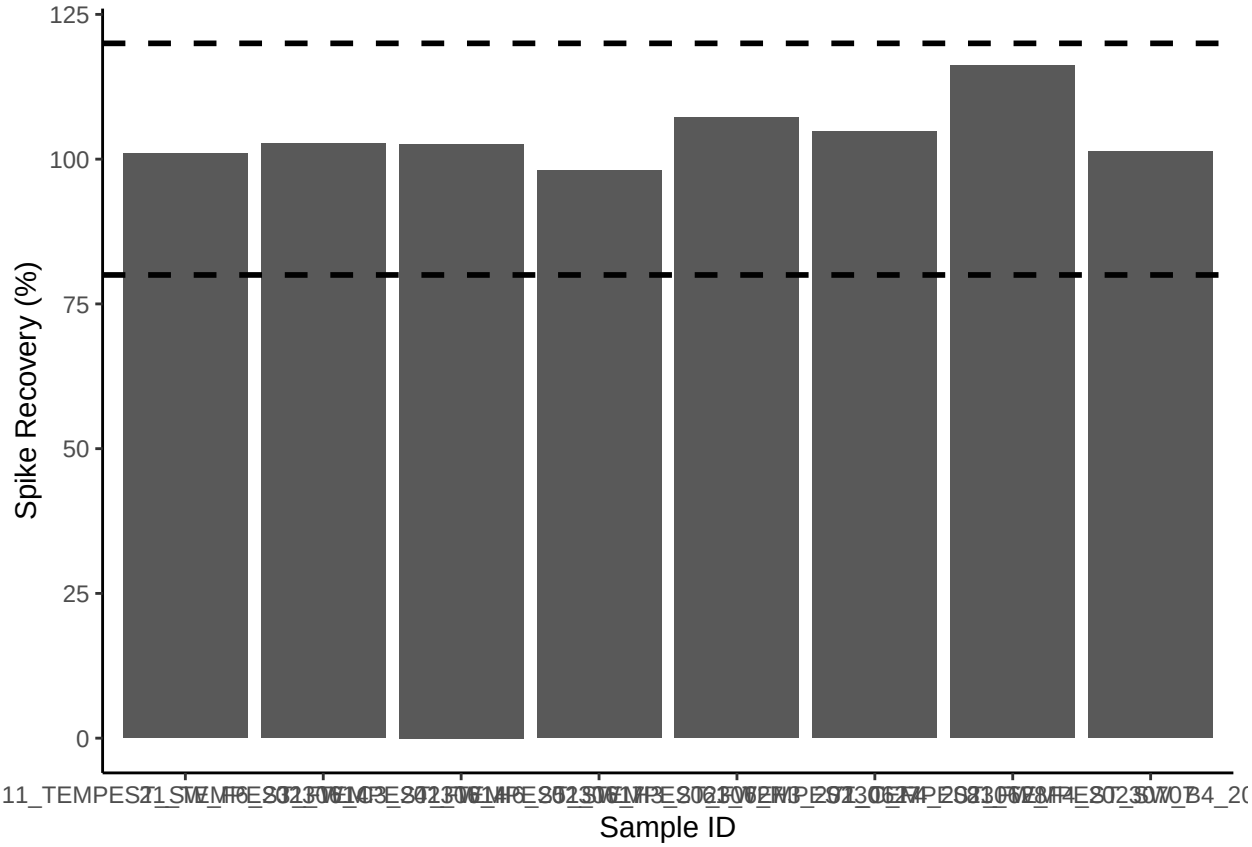
```
##                   Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area    SO4_mM
## 1 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 69.757946403 4.4685909944  11.525421 7.797879e-05       50  0.001501
## 2  0.083929478 0.0054023671   2.954760 7.797879e-05       50  0.001501
## 3  0.082964739 0.0053405677   2.968029 7.797879e-05       50  0.001501
## 4 36.525114245 2.3397572471   7.361787 7.797879e-05       50  0.001501
## 5  0.043136812 0.0027892644   3.006154 7.797879e-05       50  0.001501
## 6  0.006637518 0.0004511871   2.944317 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 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 69.757946403 4.4685909944  11.525421 7.797879e-05       50  0.001501
## 2  0.083929478 0.0054023671   2.954760 7.797879e-05       50  0.001501
## 3  0.082964739 0.0053405677   2.968029 7.797879e-05       50  0.001501
## 4 36.525114245 2.3397572471   7.361787 7.797879e-05       50  0.001501
## 5  0.043136812 0.0027892644   3.006154 7.797879e-05       50  0.001501
## 6  0.006637518 0.0004511871   2.944317 7.797879e-05       50  0.001501
##   SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1     2.667982e-04   3.482982e-04    3.447770e-04    101.02131           YES
## 2     8.926409e-06   8.929284e-05    8.690520e-05    102.74741           YES
## 3     9.431393e-06   8.969383e-05    8.741018e-05    102.61256           YES
## 4     1.489409e-04   2.224732e-04    2.269197e-04     98.04052           YES
## 5     6.798978e-06   9.084598e-05    8.477777e-05    107.15778           YES
## 6     6.927073e-06   8.897726e-05    8.490586e-05    104.79519           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                 8     8     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
## 26      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
```

```
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
##        SO4_mM       Cl_mM    salinity
## 26  0.3351716   0.0507165 0.003274805
## 27 10.8818465 101.1889929 6.482005689
## 28 10.7344136 100.6126544 6.445086510
## 29  8.8873487  69.7579464 4.468590994
## 30 11.5254211  69.4091283 4.446246355
## 31  5.0303961  48.3220423 3.095446635
```

```
#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
## 26      1_TEMPEST_FW_I5_20230608  10.7456   0.2824    1.7979   0.0629
## 27     10_TEMPEST_SW_I5_20230614 348.8720   9.1697 3587.1498 125.4632
## 28 10_TEMPEST_SW_I5_20230614_dup 344.1453   9.0455 3566.7186 124.7486
## 29     11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192  86.4922
## 30 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536  86.0597
## 31     12_TEMPEST_SW_E3_20230614 161.2745   4.2389 1713.0164  59.9140
##        SO4_mM       Cl_mM    salinity  SO4_ugmL     Cl_ugmL
## 26  0.3351716   0.0507165 0.003274805 0.2148854  0.03596502
## 27 10.8818465 101.1889929 6.482005689 6.9774197 71.74298929
## 28 10.7344136 100.6126544 6.445086510 6.8829133 71.33436316
## 29  8.8873487  69.7579464 4.468590994 5.6985396 49.45839792
## 30 11.5254211  69.4091283 4.446246355 7.3900672 49.21108362
## 31  5.0303961  48.3220423 3.095446635 3.2254698 34.26031919
```

```
#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 11_TEMPEST_SW_F6_20230614_spk 369.5050   9.7120 2460.5536 86.0597 11.525421
## 2 21_TEMPEST_FW_C3_20230614_spk  94.7296   2.4899    3.1081  0.1087  2.954760
## 3 31_TEMPEST_FW_H6_20230617_spk  95.1550   2.5010    3.3125  0.1159  2.968029
## 4 41_TEMPEST_SW_H3_20230627_spk 236.0189   6.2035 1271.7793 44.4814  7.361787
## 5 51_TEMPEST_FW_H3_20230627_spk  96.3773   2.5332    1.6148  0.0565  3.006154
## 6  61_TEMPEST_C_F4_20230628_spk  94.3948   2.4811    0.5600  0.0196  2.944317
##         Cl_mM    salinity SO4_ugmL     Cl_ugmL
## 1 69.40912835 4.446246355 7.390067 49.21108362
## 2  0.08767560 0.005642337 1.894619  0.06215460
## 3  0.09344147 0.006011687 1.903065  0.06627174
## 4 35.87529760 2.298131195 4.720375 25.43557297
## 5  0.04555148 0.002943944 1.927567  0.03230534
## 6  0.01579690 0.001037920 1.887923  0.01120500
```

```
## 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 11_TEMPEST_SW_F6_20230614_spk 7.390067 49.21108362
## 2 21_TEMPEST_FW_C3_20230614_spk 1.894619  0.06215460
## 3 31_TEMPEST_FW_H6_20230617_spk 1.903065  0.06627174
## 4 41_TEMPEST_SW_H3_20230627_spk 4.720375 25.43557297
## 5 51_TEMPEST_FW_H3_20230627_spk 1.927567  0.03230534
## 6  61_TEMPEST_C_F4_20230628_spk 1.887923  0.01120500
```

```
#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 11_TEMPEST_SW_F6_20230614     7.390067
## 2 21_TEMPEST_FW_C3_20230614     1.894619
## 3 31_TEMPEST_FW_H6_20230617     1.903065
## 4 41_TEMPEST_SW_H3_20230627     4.720375
## 5 51_TEMPEST_FW_H3_20230627     1.927567
## 6  61_TEMPEST_C_F4_20230628     1.887923
```

```
#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 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity  SO4_ugmL      Cl_ugmL SO4_ugmL_spk
## 1 69.757946403 4.4685909944 5.6985396 49.458397921     7.390067
## 2  0.083929478 0.0054023671 0.1906881  0.059524206     1.894619
## 3  0.082964739 0.0053405677 0.2014171  0.058838016     1.903065
## 4 36.525114245 2.3397572471 3.1812603 25.896292361     4.720375
## 5  0.043136812 0.0027892644 0.1451850  0.030589862     1.927567
## 6  0.006637518 0.0004511871 0.1479243  0.004686191     1.887923
```

```
#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 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
```

```
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity  SO4_ugmL      Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 69.757946403 4.4685909944 5.6985396 49.458397921     7.390067          2.5
## 2  0.083929478 0.0054023671 0.1906881  0.059524206     1.894619          2.5
## 3  0.082964739 0.0053405677 0.2014171  0.058838016     1.903065          2.5
## 4 36.525114245 2.3397572471 3.1812603 25.896292361     4.720375          2.5
## 5  0.043136812 0.0027892644 0.1451850  0.030589862     1.927567          2.5
## 6  0.006637518 0.0004511871 0.1479243  0.004686191     1.887923          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 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
```

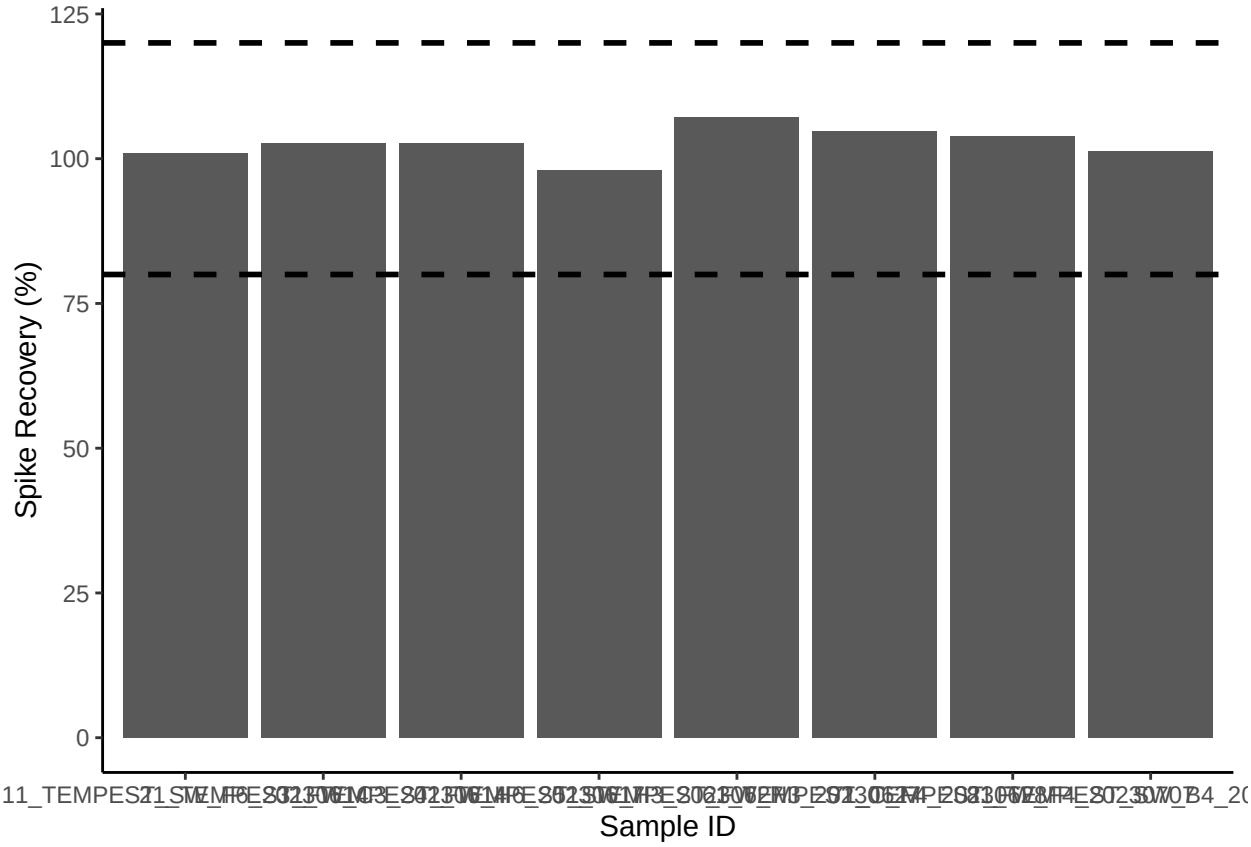
```
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity  SO4_ugmL      Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 69.757946403 4.4685909944 5.6985396 49.458397921     7.390067          2.5
## 2  0.083929478 0.0054023671 0.1906881  0.059524206     1.894619          2.5
## 3  0.082964739 0.0053405677 0.2014171  0.058838016     1.903065          2.5
## 4 36.525114245 2.3397572471 3.1812603 25.896292361     4.720375          2.5
## 5  0.043136812 0.0027892644 0.1451850  0.030589862     1.927567          2.5
## 6  0.006637518 0.0004511871 0.1479243  0.004686191     1.887923          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 11_TEMPEST_SW_F6_20230614 284.9284   7.4890 2472.9192 86.4922 8.8873487
## 2 21_TEMPEST_FW_C3_20230614   9.5330   0.2506    2.9753  0.1041 0.2973487
## 3 31_TEMPEST_FW_H6_20230617  10.0723   0.2647    2.9411  0.1029 0.3141703
## 4 41_TEMPEST_SW_H3_20230627 159.0621   4.1808 1294.8153 45.2871 4.9613880
## 5 51_TEMPEST_FW_H3_20230627   7.2610   0.1908    1.5292  0.0535 0.2264816
## 6  61_TEMPEST_C_F4_20230628   7.3978   0.1944    0.2353  0.0082 0.2307486
##          Cl_mM     salinity  SO4_ugmL      Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1 69.757946403 4.4685909944 5.6985396 49.458397921     7.390067          2.5
## 2  0.083929478 0.0054023671 0.1906881  0.059524206     1.894619          2.5
## 3  0.082964739 0.0053405677 0.2014171  0.058838016     1.903065          2.5
## 4 36.525114245 2.3397572471 3.1812603 25.896292361     4.720375          2.5
## 5  0.043136812 0.0027892644 0.1451850  0.030589862     1.927567          2.5
## 6  0.006637518 0.0004511871 0.1479243  0.004686191     1.887923          2.5
##   SampleVol SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery
## 1     1.501        8.5535079      11.166392       11.053508     101.0212
## 2     1.501        0.2862228       2.862769        2.786223     102.7473
## 3     1.501        0.3023270       2.875531        2.802327     102.6123
## 4     1.501        4.7750717       7.132487        7.275072      98.0401
## 5     1.501        0.2179227       2.912553        2.717923     107.1610
## 6     1.501        0.2220344       2.852651        2.722034     104.7985
##   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                 8     8     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       1 TEMPEST   FW   I5 20230608
## 2      10 TEMPEST   SW   I5 20230614
## 3      11 TEMPEST   SW   F6 20230614
## 4      12 TEMPEST   SW   E3 20230614
## 5      13 TEMPEST   SW   D5 20230614
## 6      14 TEMPEST   SW   B4 20230614
```

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

```
##   Project    Plot Grid Date       NA                 Sample_ID  SO4_ppm
## 1       1 TEMPEST   FW   I5 20230608  1_TEMPEST_FW_I5_20230608  10.7456
## 2      10 TEMPEST   SW   I5 20230614 10_TEMPEST_SW_I5_20230614 348.8720
## 3      11 TEMPEST   SW   F6 20230614 11_TEMPEST_SW_F6_20230614 284.9284
## 4      12 TEMPEST   SW   E3 20230614 12_TEMPEST_SW_E3_20230614 161.2745
## 5      13 TEMPEST   SW   D5 20230614 13_TEMPEST_SW_D5_20230614 194.2448
## 6      14 TEMPEST   SW   B4 20230614 14_TEMPEST_SW_B4_20230614 389.9092
##   SO4_Area    Cl_ppm  Cl_Area     SO4_mM       Cl_mM    salinity
## 1   0.2824    1.7979   0.0629  0.3351716   0.0507165 0.003274805
## 2   9.1697 3587.1498 125.4632 10.8818465 101.1889929 6.482005689
## 3   7.4890 2472.9192  86.4922  8.8873487  69.7579464 4.468590994
## 4   4.2389 1713.0164  59.9140  5.0303961  48.3220423 3.095446635
## 5   5.1055 2204.8413  77.1159  6.0587898  62.1958054 3.984174229
## 6  10.2483 3653.5580 127.7859 12.1618590 103.0622849 6.602005306
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_TEMPEST_Dionex_ClSo4_Final_2023.csv")        #Change
```

END