Dionex_COMPASS_June2023

Stephanie J. Wilson

2023-06-23

Daily Set up

Read in and Format the raw data - change wd & file names

```
# SULFATE DATA:

## Read in raw data file from Dionex - copied and saved as a txt
Sdat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202309_SO4a.txt",sep='\t' , header=T, skip=3)
head(Sdat)
```

```
##   X        X.1                  X.2 IC.SO4 IC.SO4.1 IC.SO4.2 IC.SO4.3 IC.SO4.4
## 1 1  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 2 2  Lab Blank              Unknown 5.000   0.0009     4.50   0.0011     0.01
## 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 4.993   0.4693     6.61   0.5731     3.70
## 6 6 Standard 2 Calibration Standard 5.020   0.9856     6.87   1.2036     7.51
```

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

```
##          X.1 IC.SO4.1
## 1  Lab Blank     n.a.
## 2  Lab Blank  0.0009
## 3  Lab Blank     n.a.
## 4  Lab Blank     n.a.
## 5 Standard 1  0.4693
## 6 Standard 2  0.9856
```

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

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

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

```
##    Sample_ID SO4_ppm
## 1  Lab Blank      NA
## 2  Lab Blank  0.0009
## 3  Lab Blank      NA
## 4  Lab Blank      NA
## 5 Standard 1  0.4693
## 6 Standard 2  0.9856
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202309_Cla.txt",sep='\t' , header=T, skip=3)
head(Cldat)
```

```
##   X        X.1                  X.2  IC.Cl IC.Cl.1 IC.Cl.2  IC.Cl.3 IC.Cl.4
## 1 1  Lab Blank              Unknown 3.910  0.0064  100.00   0.0105    0.07
## 2 2  Lab Blank              Unknown 3.900  0.0129   88.96   0.0211    0.07
## 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 3.950  4.9511   93.39   8.0938   63.84
## 6 6 Standard 2 Calibration Standard 3.920  9.9749   93.13  16.3066  126.89
```

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

```
##          X.1 IC.Cl.1
## 1  Lab Blank 0.0064
## 2  Lab Blank 0.0129
## 3  Lab Blank    n.a.
## 4  Lab Blank    n.a.
## 5 Standard 1 4.9511
## 6 Standard 2 9.9749
```

```
## Name the columns correctly
colnames(Cldat) <- c( "Sample_ID", "Cl_ppm")
Cldat$Sample_ID <- as.factor(Cldat$Sample_ID)
Cldat$Cl_ppm <- as.numeric(Cldat$Cl_ppm)
```

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

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

```
##    Sample_ID Cl_ppm
## 1  Lab Blank 0.0064
## 2  Lab Blank 0.0129
## 3  Lab Blank     NA
## 4  Lab Blank     NA
## 5 Standard 1 4.9511
## 6 Standard 2 9.9749
```

```r
## Bring the data back together:
all_dat <- merge(Sdat, Cldat, by="Sample_ID")
head(all_dat)
```

```
##   Sample_ID   SO4_ppm     Cl_ppm
## 1           2352.6008 16941.3424
## 2           2352.6008  5228.0332
## 3           2352.6008     0.0064
## 4           2352.6008  4945.1061
## 5           2352.6008         NA
## 6            662.6104 16941.3424
```

```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    Cl_ppm
## 26     100_GWI_202309_TR_LysC_10cm  788.8810  6095.389
## 27 100_GWI_202309_TR_LysC_10cm_dup  741.2978  5736.637
## 28     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.018
## 29 101_GWI_202309_TR_LysC_20cm_spk 1215.0450  8252.851
## 30     102_GWI_202309_TR_LysC_45cm 1187.1120  9569.575
## 31     103_GWI_202309_WC_LysA_10cm 1855.2192 13352.361
```

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

## If there is an n.a. in the dataframe that is because there was no peak, which
# would mean there was no Cl or SO4 and we want to know that so make all n.a.'s into zeros
all_dat[is.na(all_dat)] <- 0
head(all_dat)
```

```
##                          Sample_ID   SO4_ppm    Cl_ppm
## 26     100_GWI_202309_TR_LysC_10cm  788.8810  6095.389
## 27 100_GWI_202309_TR_LysC_10cm_dup  741.2978  5736.637
## 28     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.018
## 29 101_GWI_202309_TR_LysC_20cm_spk 1215.0450  8252.851
## 30     102_GWI_202309_TR_LysC_45cm 1187.1120  9569.575
## 31     103_GWI_202309_WC_LysA_10cm 1855.2192 13352.361
```

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

```r
stds <- all_dat[grepl("Standard", all_dat$Sample_ID),]
#stds <- stds[-c(17),]   #this is if you need to remove one for any reason
head(stds)
```

```
##      Sample_ID SO4_ppm Cl_ppm
## 266 Standard 1  0.4693 4.9511
## 267 Standard 1  0.4693 4.9500
## 268 Standard 1  0.4693 5.0158
```

```
## 269 Standard 1  0.4752 4.9511
## 270 Standard 1  0.4752 4.9500
## 271 Standard 1  0.4752 5.0158
```

```
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 <5, '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  0.477 0.00818 1.71  YES
## 2 Standard 2  0.995 0.0158  1.59  YES
## 3 Standard 3  1.96  0.00367 0.187 YES
## 4 Standard 4 10.1   0.0523  0.519 YES
## 5 Standard 5 19.9   0.0701  0.353 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 <5, '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   4.97 0.0326 0.656 YES
## 2 Standard 2  10.0  0.0971 0.970 YES
## 3 Standard 3  19.5  0.0669 0.343 YES
## 4 Standard 4  99.6  0.349  0.351 YES
## 5 Standard 5 199.   0.846  0.425 YES
```

Calculate mmol/L concentrations

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

```
##                          Sample_ID   SO4_ppm    Cl_ppm
## 26     100_GWI_202309_TR_LysC_10cm  788.8810  6095.389
## 27 100_GWI_202309_TR_LysC_10cm_dup  741.2978  5736.637
## 28     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.018
## 29 101_GWI_202309_TR_LysC_20cm_spk 1215.0450  8252.851
## 30     102_GWI_202309_TR_LysC_45cm 1187.1120  9569.575
## 31     103_GWI_202309_WC_LysA_10cm 1855.2192 13352.361
```

```r
# Constants needed for calculations:
clmw <- 35.45      #molecular weight of Chloride: 35.45
smw <- 32.06       #molecular weight of sulfur: 32.06


# Convert ppm to mmol/L
sampledat$SO4_mM <- (sampledat$SO4_ppm / smw)
sampledat$Cl_mM <- (sampledat$Cl_ppm / clmw)


# Calculate Salinity
# calculated using the Knudsen equation
# Salinity = 0.03 + 1.8050 * Chlorinity
# Ref: A Practical Handbook of Seawater Analysis by Strickland & Parsons (P. 11)
# =((1.807*Cl_ppm)+0.026)/1000
sampledat$salinity <- ((1.8070 * sampledat$Cl_ppm) + 0.026) / 1000


head(sampledat)
```

```
##                          Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM
## 26     100_GWI_202309_TR_LysC_10cm  788.8810  6095.389 24.60639 171.9433
## 27 100_GWI_202309_TR_LysC_10cm_dup  741.2978  5736.637 23.12220 161.8233
## 28     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.018 32.94783 237.3771
## 29 101_GWI_202309_TR_LysC_20cm_spk 1215.0450  8252.851 37.89910 232.8026
## 30     102_GWI_202309_TR_LysC_45cm 1187.1120  9569.575 37.02782 269.9457
## 31     103_GWI_202309_WC_LysA_10cm 1855.2192 13352.361 57.86710 376.6533
##    salinity
## 26 11.01439
## 27 10.36613
## 28 15.20596
## 29 14.91293
## 30 17.29225
## 31 24.12774
```

Pull out dups and check with percent difference

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

```
##                          Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM
## 26     100_GWI_202309_TR_LysC_10cm  788.8810  6095.389 24.60639 171.9433
## 27 100_GWI_202309_TR_LysC_10cm_dup  741.2978  5736.637 23.12220 161.8233
## 28     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.018 32.94783 237.3771
## 29 101_GWI_202309_TR_LysC_20cm_spk 1215.0450  8252.851 37.89910 232.8026
## 30     102_GWI_202309_TR_LysC_45cm 1187.1120  9569.575 37.02782 269.9457
## 31     103_GWI_202309_WC_LysA_10cm 1855.2192 13352.361 57.86710 376.6533
##    salinity
## 26 11.01439
## 27 10.36613
## 28 15.20596
```

```
## 29 14.91293
## 30 17.29225
## 31 24.12774
```

```
#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     Cl_ppm    SO4_mM      Cl_mM
## 1     100_GWI_202309_TR_LysC_10cm_dup  741.2978  5736.6367 23.122202 161.823320
## 2     110_GWI_202309_WC_LysC_20cm_dup 1129.5717 12989.4688 35.233054 366.416609
## 3 120_MSM_202309_TR_RHZ_SF_Tree_7_dup  945.2658  9369.8967 29.484273 264.313024
## 4      127_MSM_202309_WC_RHZ_LysA_dup 1109.8533  8970.6600 34.618007 253.051058
## 5   50_SWH_202309_UPCON_LysA_20cm_dup  115.4148    16.9133  3.599963   0.477103
## 6      80_SWH_202309_WC_LysB_20cm_dup    0.0000  1348.3927  0.000000  38.036465
##      salinity
## 1 10.36612852
## 2 23.47199612
## 3 16.93142934
## 4 16.21000862
## 5  0.03058833
## 6  2.43657161
```

```
#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    Cl_ppm   SO4_mM    Cl_mM salinity
## 1 100_GWI_202309_TR_LysC_10cm  788.881  6095.389 24.60639 171.9433 11.01439
## 2 101_GWI_202309_TR_LysC_20cm 1056.307  8415.018 32.94783 237.3771 15.20596
## 3 102_GWI_202309_TR_LysC_45cm 1187.112  9569.575 37.02782 269.9457 17.29225
## 4 103_GWI_202309_WC_LysA_10cm 1855.219 13352.361 57.86710 376.6533 24.12774
## 5 104_GWI_202309_WC_LysA_20cm 1376.219 13901.322 42.92634 392.1389 25.11972
## 6 105_GWI_202309_WC_LysA_45cm 1402.715 12239.430 43.75279 345.2590 22.11668
```

```
#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(2,3)]
colnames(dups) <- c('Sample_ID', 'SO4_mM_dup', "Cl_mM_dup", "salinity_dup")
head(dups)
```

```
##                         Sample_ID SO4_mM_dup  Cl_mM_dup salinity_dup
## 1     100_GWI_202309_TR_LysC_10cm  23.122202 161.823320  10.36612852
## 2     110_GWI_202309_WC_LysC_20cm  35.233054 366.416609  23.47199612
## 3 120_MSM_202309_TR_RHZ_SF_Tree_7  29.484273 264.313024  16.93142934
## 4      127_MSM_202309_WC_RHZ_LysA  34.618007 253.051058  16.21000862
## 5   50_SWH_202309_UPCON_LysA_20cm   3.599963   0.477103   0.03058833
## 6      80_SWH_202309_WC_LysB_20cm   0.000000  38.036465   2.43657161
```

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

```
##                          Sample_ID   SO4_ppm     Cl_ppm    SO4_mM       Cl_mM
## 1     100_GWI_202309_TR_LysC_10cm  788.8810  6095.3894 24.606394 171.9432835
## 2     110_GWI_202309_WC_LysC_20cm 1175.5635 13445.2265 36.667608 379.2729619
## 3 120_MSM_202309_TR_RHZ_SF_Tree_7  963.1908  9562.4377 30.043381 269.7443639
## 4      127_MSM_202309_WC_RHZ_LysA 1073.4254  8667.1153 33.481765 244.4884429
## 5   50_SWH_202309_UPCON_LysA_20cm  116.9987    18.4797  3.649367   0.5212891
## 6      80_SWH_202309_WC_LysB_20cm    0.0000  1415.1307  0.000000  39.9190606
##      salinity SO4_mM_dup  Cl_mM_dup salinity_dup
## 1 11.01439465  23.122202 161.823320  10.36612852
## 2 24.29555029  35.233054 366.416609  23.47199612
## 3 17.27935092  29.484273 264.313024  16.93142934
## 4 15.66150335  34.618007 253.051058  16.21000862
## 5  0.03341882   3.599963   0.477103   0.03058833
## 6  2.55716717   0.000000  38.036465   2.43657161
```

```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     Cl_ppm    SO4_mM       Cl_mM
## 1     100_GWI_202309_TR_LysC_10cm  788.8810  6095.3894 24.606394 171.9432835
## 2     110_GWI_202309_WC_LysC_20cm 1175.5635 13445.2265 36.667608 379.2729619
## 3 120_MSM_202309_TR_RHZ_SF_Tree_7  963.1908  9562.4377 30.043381 269.7443639
## 4      127_MSM_202309_WC_RHZ_LysA 1073.4254  8667.1153 33.481765 244.4884429
## 5   50_SWH_202309_UPCON_LysA_20cm  116.9987    18.4797  3.649367   0.5212891
## 6      80_SWH_202309_WC_LysB_20cm    0.0000  1415.1307  0.000000  39.9190606
##      salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 11.01439465  23.122202 161.823320  10.36612852     6.219299           YES
## 2 24.29555029  35.233054 366.416609  23.47199612     3.990378           YES
## 3 17.27935092  29.484273 264.313024  16.93142934     1.878481           YES
## 4 15.66150335  34.618007 253.051058  16.21000862     3.336990           YES
## 5  0.03341882   3.599963   0.477103   0.03058833     1.363002           YES
## 6  2.55716717   0.000000  38.036465   2.43657161          NaN          <NA>
##   Cl_dups_chk Cl_dups_flag
## 1    6.064096          YES
## 2    3.448178          YES
## 3    2.033991          YES
## 4    3.441984          YES
## 5    8.851468          YES
## 6    4.829921          YES
```

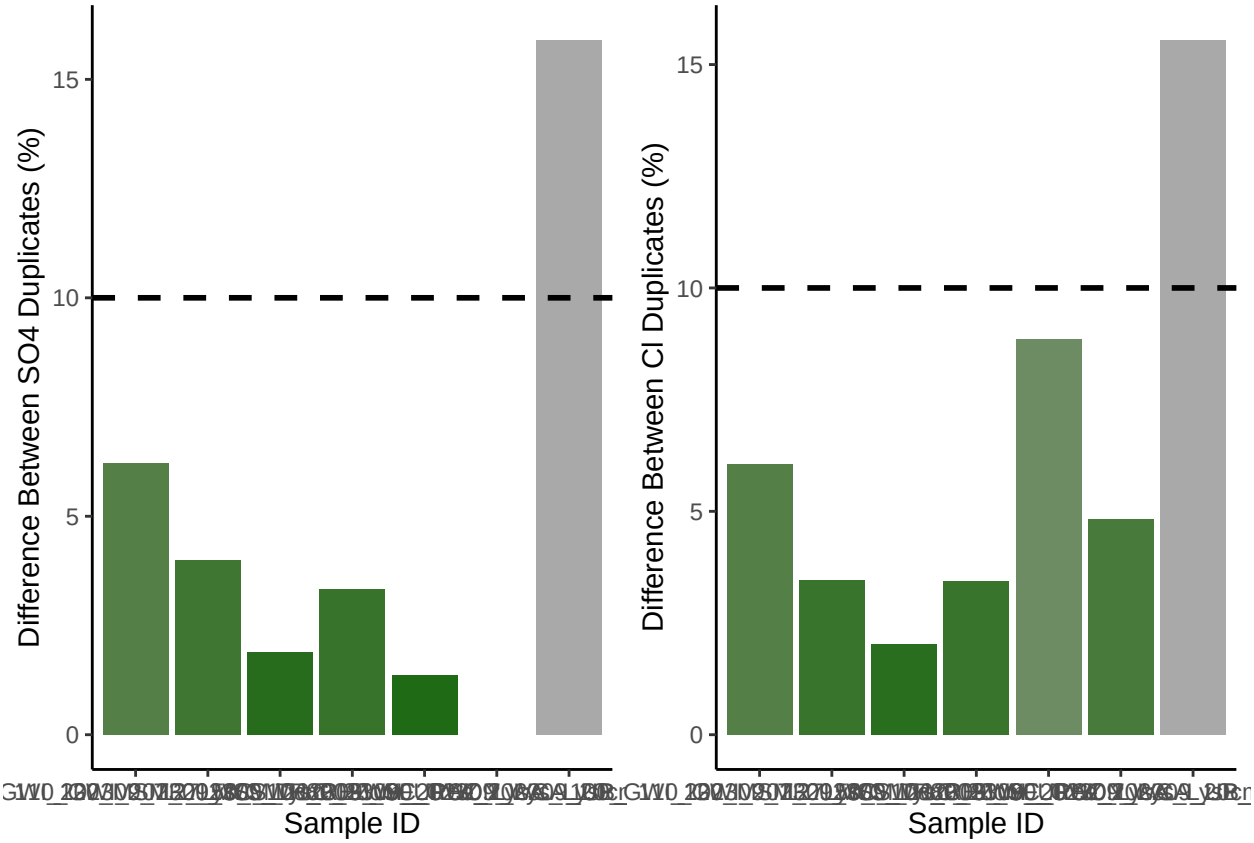
```r
#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
Sdupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = SO4_dups_chk, fill=SO4_dups_chk)) +
       geom_bar(stat = 'identity') +
```

```
        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.
```

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

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

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_bar()`).
```



```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")
Perc_dups <- Perc_dups[-c(3),]
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 NO, rerun         1 NO, rerun          1     7  14.28571   14.28571
## 2       YES         5       YES          6     7  71.42857   85.71429
```

Pull out dups and check with cv

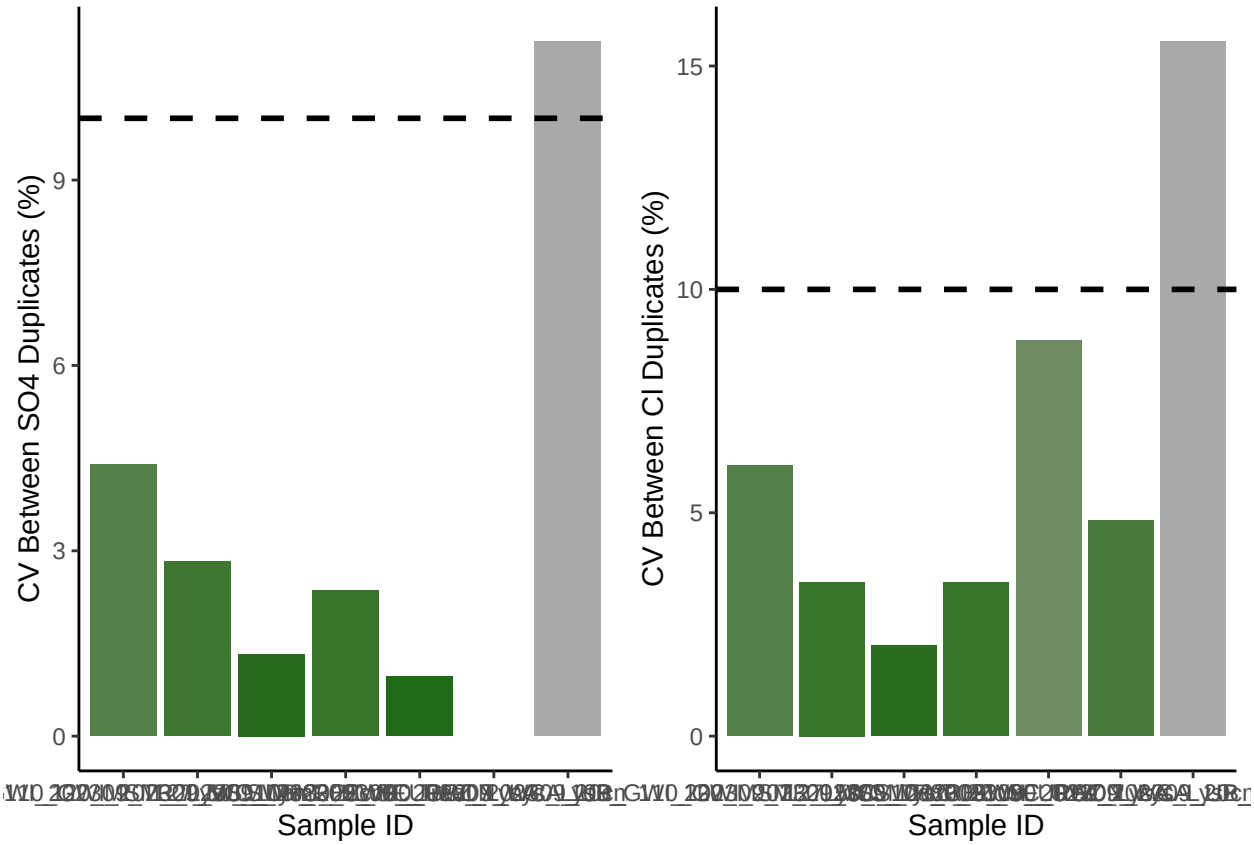
```r
#the cv
# calculate the sd bewteen the two duplicates then divide by the mean and multiply by 100

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     Cl_ppm    SO4_mM       Cl_mM
## 1     100_GWI_202309_TR_LysC_10cm  788.8810  6095.3894 24.606394 171.9432835
## 2     110_GWI_202309_WC_LysC_20cm 1175.5635 13445.2265 36.667608 379.2729619
## 3 120_MSM_202309_TR_RHZ_SF_Tree_7  963.1908  9562.4377 30.043381 269.7443639
## 4      127_MSM_202309_WC_RHZ_LysA 1073.4254  8667.1153 33.481765 244.4884429
## 5   50_SWH_202309_UPCON_LysA_20cm  116.9987    18.4797  3.649367   0.5212891
## 6      80_SWH_202309_WC_LysB_20cm    0.0000  1415.1307  0.000000  39.9190606
##      salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 11.01439465  23.122202 161.823320  10.36612852     6.219299           YES
## 2 24.29555029  35.233054 366.416609  23.47199612     3.990378           YES
## 3 17.27935092  29.484273 264.313024  16.93142934     1.878481           YES
## 4 15.66150335  34.618007 253.051058  16.21000862     3.336990           YES
## 5  0.03341882   3.599963   0.477103   0.03058833     1.363002           YES
## 6  2.55716717   0.000000  38.036465   2.43657161          NaN          <NA>
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
```

```
## 1    6.064096          YES   4.3977087             YES
## 2    3.448178          YES   2.8216231             YES
## 3    2.033991          YES   1.3282869             YES
## 4    3.441984          YES   2.3596085             YES
## 5    8.851468          YES   0.9637878             YES
## 6    4.829921          YES         NaN            <NA>
```

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


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

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

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_bar()`).
```




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_dups <- QAdups %>%
  group_by(SO4_dups_flag) %>%
  summarise(S_no_rows = length(SO4_dups_flag))
Perc_dups1 <- QAdups %>%
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(Cl_dups_flag))
colnames(Perc_dups) <- c("Flag", "S_no_rows")
Perc_dups <- Perc_dups[-c(3),]
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 NO, rerun         1 NO, rerun          1     7  14.28571   14.28571
## 2       YES         5       YES          6     7  71.42857   85.71429
```

Pull out spikes and check

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

```
##                          Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM
## 26     100_GWI_202309_TR_LysC_10cm  788.8810  6095.389 24.60639 171.9433
## 27 100_GWI_202309_TR_LysC_10cm_dup  741.2978  5736.637 23.12220 161.8233
## 28     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.018 32.94783 237.3771
## 29 101_GWI_202309_TR_LysC_20cm_spk 1215.0450  8252.851 37.89910 232.8026
## 30     102_GWI_202309_TR_LysC_45cm 1187.1120  9569.575 37.02782 269.9457
## 31     103_GWI_202309_WC_LysA_10cm 1855.2192 13352.361 57.86710 376.6533
##    salinity
## 26 11.01439
## 27 10.36613
## 28 15.20596
## 29 14.91293
## 30 17.29225
## 31 24.12774
```

```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     Cl_ppm    SO4_mM     Cl_mM
## 1     101_GWI_202309_TR_LysC_20cm_spk 1215.0450  8252.8505 37.899095 232.80255
## 2             111_GWI_202309_SW_A_spk 2144.7441 12786.3210 66.897820 360.68606
## 3 121_MSM_202309_TR_RHZ_SF_Tree_8_spk  790.9077  9044.4139 24.669610 255.13156
## 4      128_MSM_202309_WC_RHZ_LysC_spk  988.9780  9191.8894 30.847723 259.29166
## 5   51_SWH_202309_UPCON_LysA_45cm_spk  273.4441   211.4954  8.529136   5.96602
## 6      61_SWH_202309_UP_LysB_10cm_spk  271.7634   960.8437  8.476712  27.10419
##     salinity
## 1 14.9129269
## 2 23.1049080
## 3 16.3432819
## 4 16.6097701
## 5  0.3821982
## 6  1.7362706
```

```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, 5,6)]
colnames(spks) <- c('Sample_ID', 'SO4_mM_spk')
head(spks)
```

```
##                        Sample_ID SO4_mM_spk
## 1     101_GWI_202309_TR_LysC_20cm  37.899095
## 2             111_GWI_202309_SW_A  66.897820
## 3 121_MSM_202309_TR_RHZ_SF_Tree_8  24.669610
## 4      128_MSM_202309_WC_RHZ_LysC  30.847723
## 5   51_SWH_202309_UPCON_LysA_45cm   8.529136
## 6      61_SWH_202309_UP_LysB_10cm   8.476712
```

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

```
##                          Sample_ID   SO4_ppm     Cl_ppm    SO4_mM      Cl_mM
## 1     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.0182 32.947826 237.377100
## 2             111_GWI_202309_SW_A 1931.3778 13871.3668 60.242601 391.293845
## 3 121_MSM_202309_TR_RHZ_SF_Tree_8  672.8113  8669.9839 20.986004 244.569362
## 4      128_MSM_202309_WC_RHZ_LysC  807.4715  9248.3920 25.186260 260.885529
## 5   51_SWH_202309_UPCON_LysA_45cm   56.7756   216.1650  1.770917   6.097743
## 6      61_SWH_202309_UP_LysB_10cm   90.8451   932.8824  2.833596  26.315441
##     salinity SO4_mM_spk
## 1 15.2059639  37.899095
## 2 25.0655858  66.897820
## 3 15.6666869  24.669610
## 4 16.7118703  30.847723
## 5  0.3906362   8.529136
## 6  1.6857445   8.476712
```

```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     Cl_ppm    SO4_mM      Cl_mM
## 1     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.0182 32.947826 237.377100
## 2             111_GWI_202309_SW_A 1931.3778 13871.3668 60.242601 391.293845
## 3 121_MSM_202309_TR_RHZ_SF_Tree_8  672.8113  8669.9839 20.986004 244.569362
## 4      128_MSM_202309_WC_RHZ_LysC  807.4715  9248.3920 25.186260 260.885529
## 5   51_SWH_202309_UPCON_LysA_45cm   56.7756   216.1650  1.770917   6.097743
## 6      61_SWH_202309_UP_LysB_10cm   90.8451   932.8824  2.833596  26.315441
##     salinity SO4_mM_spk SO4_spk_Conc
## 1 15.2059639  37.899095 7.797879e-05
## 2 25.0655858  66.897820 7.797879e-05
## 3 15.6666869  24.669610 7.797879e-05
## 4 16.7118703  30.847723 7.797879e-05
## 5  0.3906362   8.529136 7.797879e-05
## 6  1.6857445   8.476712 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$Sample_ID, "MSM_202306_UP"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_202306_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_202306_WC"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_202306_UP"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_202306_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_202306_WC"), 100, QAspks$Dilution)
```

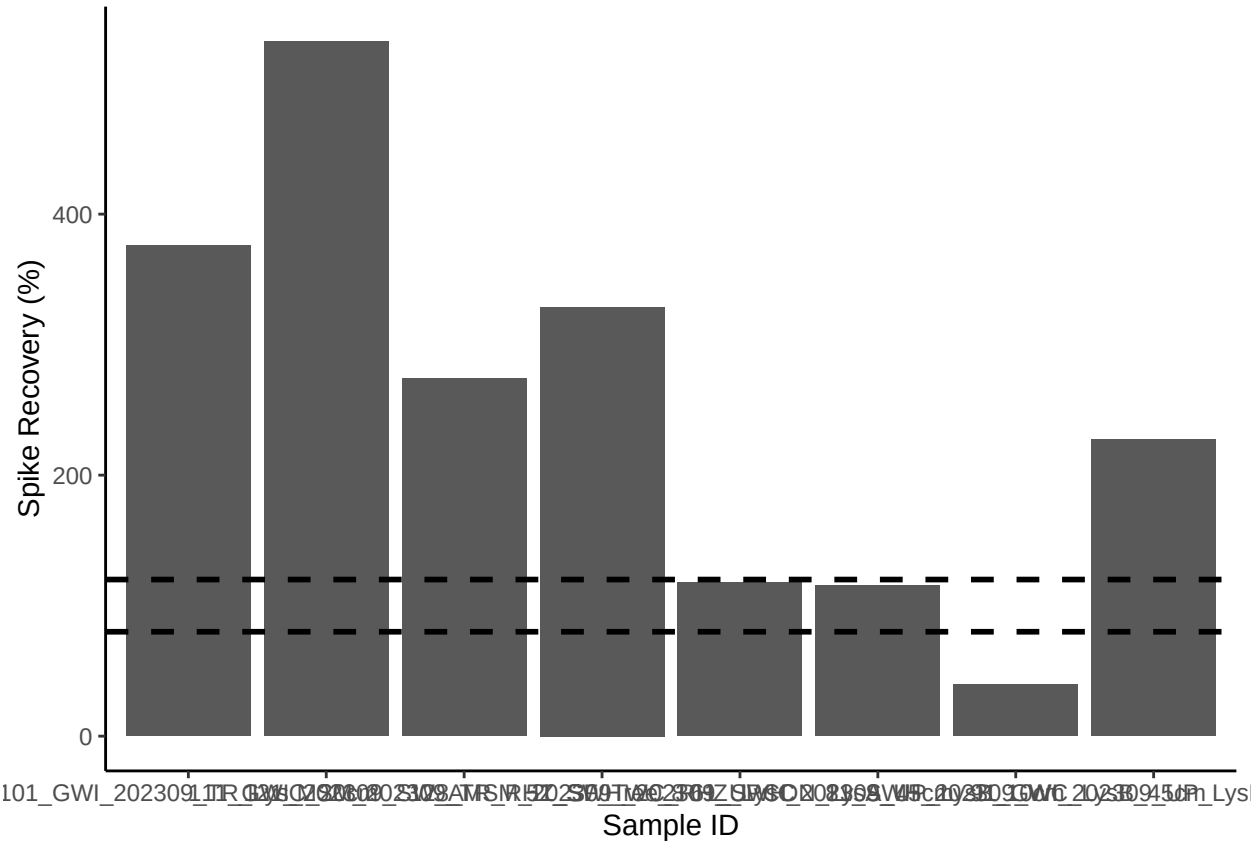
```r
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_202306_UP"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_202306_TR"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_202306_WC"), 200, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_202306_UP"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_202306_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_202306_WC"), 50, QAspks$Dilution)

#Set Sample volumes in uL
QAspks$SampleVol <- 1
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_202306_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_202306_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_202306_WC"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_202306_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_202306_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_202306_WC"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_202306_UP"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_202306_TR"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_202306_WC"), 1462, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_202306_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_202306_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_202306_WC"), 1501, QAspks$SampleVol)

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

```
##                          Sample_ID   SO4_ppm     Cl_ppm    SO4_mM      Cl_mM
## 1     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.0182 32.947826 237.377100
## 2             111_GWI_202309_SW_A 1931.3778 13871.3668 60.242601 391.293845
## 3 121_MSM_202309_TR_RHZ_SF_Tree_8  672.8113  8669.9839 20.986004 244.569362
## 4      128_MSM_202309_WC_RHZ_LysC  807.4715  9248.3920 25.186260 260.885529
## 5   51_SWH_202309_UPCON_LysA_45cm   56.7756   216.1650  1.770917   6.097743
## 6      61_SWH_202309_UP_LysB_10cm   90.8451   932.8824  2.833596  26.315441
##     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 15.2059639  37.899095 7.797879e-05        1     1e-06
## 2 25.0655858  66.897820 7.797879e-05        1     1e-06
## 3 15.6666869  24.669610 7.797879e-05        1     1e-06
## 4 16.7118703  30.847723 7.797879e-05        1     1e-06
## 5  0.3906362   8.529136 7.797879e-05        1     1e-06
## 6  1.6857445   8.476712 7.797879e-05        1     1e-06
```

```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     Cl_ppm    SO4_mM      Cl_mM
## 1     101_GWI_202309_TR_LysC_20cm 1056.3073  8415.0182 32.947826 237.377100
## 2             111_GWI_202309_SW_A 1931.3778 13871.3668 60.242601 391.293845
## 3 121_MSM_202309_TR_RHZ_SF_Tree_8  672.8113  8669.9839 20.986004 244.569362
## 4      128_MSM_202309_WC_RHZ_LysC  807.4715  9248.3920 25.186260 260.885529
## 5   51_SWH_202309_UPCON_LysA_45cm   56.7756   216.1650  1.770917   6.097743
## 6      61_SWH_202309_UP_LysB_10cm   90.8451   932.8824  2.833596  26.315441
##     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1 15.2059639  37.899095 7.797879e-05        1     1e-06     3.294783e-05
## 2 25.0655858  66.897820 7.797879e-05        1     1e-06     6.024260e-05
## 3 15.6666869  24.669610 7.797879e-05        1     1e-06     2.098600e-05
## 4 16.7118703  30.847723 7.797879e-05        1     1e-06     2.518626e-05
## 5  0.3906362   8.529136 7.797879e-05        1     1e-06     1.770917e-06
## 6  1.6857445   8.476712 7.797879e-05        1     1e-06     2.833596e-06
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   4.168900e-04    1.109266e-04     375.8251     NO, rerun
## 2   7.358760e-04    1.382214e-04     532.3894     NO, rerun
## 3   2.713657e-04    9.896479e-05     274.2043     NO, rerun
## 4   3.393250e-04    1.031650e-04     328.9146     NO, rerun
## 5   9.382050e-05    7.974971e-05     117.6437           YES
## 6   9.324384e-05    8.081239e-05     115.3831           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: 2 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 NO, rerun           6     8      75
## 2 YES                 2     8      25
```

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("100", "GWI", "202309", "TR", "LysC", "10cm"), c("101", :
## number of columns of result is not a multiple of vector length (arg 1)
```

```r
colnames(IDs) <- c("Analysis_No" ,"Site", "Date", "Zone", "Replicate", "Depth", "RHZ", "RHZ_Rep")
head(IDs)
```

```
##   Analysis_No Site   Date Zone Replicate Depth RHZ RHZ_Rep
## 1         100  GWI 202309   TR      LysC  10cm 100     GWI
## 2         101  GWI 202309   TR      LysC  20cm 101     GWI
## 3         102  GWI 202309   TR      LysC  45cm 102     GWI
## 4         103  GWI 202309   WC      LysA  10cm 103     GWI
## 5         104  GWI 202309   WC      LysA  20cm 104     GWI
## 6         105  GWI 202309   WC      LysA  45cm 105     GWI
```

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

```
##   Analysis_No Site   Date Zone Replicate Depth RHZ RHZ_Rep
## 1         100  GWI 202309   TR      LysC  10cm 100     GWI
## 2         101  GWI 202309   TR      LysC  20cm 101     GWI
## 3         102  GWI 202309   TR      LysC  45cm 102     GWI
## 4         103  GWI 202309   WC      LysA  10cm 103     GWI
## 5         104  GWI 202309   WC      LysA  20cm 104     GWI
## 6         105  GWI 202309   WC      LysA  45cm 105     GWI
##                     Sample_ID  SO4_ppm    Cl_ppm   SO4_mM    Cl_mM salinity
## 1 100_GWI_202309_TR_LysC_10cm  788.881  6095.389 24.60639 171.9433 11.01439
## 2 101_GWI_202309_TR_LysC_20cm 1056.307  8415.018 32.94783 237.3771 15.20596
## 3 102_GWI_202309_TR_LysC_45cm 1187.112  9569.575 37.02782 269.9457 17.29225
## 4 103_GWI_202309_WC_LysA_10cm 1855.219 13352.361 57.86710 376.6533 24.12774
## 5 104_GWI_202309_WC_LysA_20cm 1376.219 13901.322 42.92634 392.1389 25.11972
## 6 105_GWI_202309_WC_LysA_45cm 1402.715 12239.430 43.75279 345.2590 22.11668
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202309a.csv")          #
```

END