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_202310_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 5.097   0.4573     6.48   0.5739     3.63
## 6 6 Standard 2 Calibration Standard 5.150   0.9857     6.85   1.2371     7.50
```

```
## 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     n.a.
## 3  Lab Blank     n.a.
## 4  Lab Blank     n.a.
## 5 Standard 1  0.4573
## 6 Standard 2  0.9857
```

```
## 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
```

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

```
##    Sample_ID SO4_ppm
## 1  Lab Blank      NA
## 2  Lab Blank      NA
## 3  Lab Blank      NA
## 4  Lab Blank      NA
## 5 Standard 1  0.4573
## 6 Standard 2  0.9857
```

```r
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("raw Data/COMPASS_Synoptic_CB_MonMon_202310_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   n.a.     n.a.    n.a.     n.a.    n.a.
## 2 2  Lab Blank             Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 3 3  Lab Blank             Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 4 4  Lab Blank             Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 5 5 Standard 1 Calibration Standard 4.040   4.9368   93.52   8.2828   64.04
## 6 6 Standard 2 Calibration Standard 4.030  10.0322   93.15  16.8317  127.41
```

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

```
##          X.1  IC.Cl.1
## 1  Lab Blank     n.a.
## 2  Lab Blank     n.a.
## 3  Lab Blank     n.a.
## 4  Lab Blank     n.a.
## 5 Standard 1  4.9368
## 6 Standard 2 10.0322
```

```r
## 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
```

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

```
##    Sample_ID  Cl_ppm
## 1  Lab Blank      NA
## 2  Lab Blank      NA
## 3  Lab Blank      NA
## 4  Lab Blank      NA
## 5 Standard 1  4.9368
## 6 Standard 2 10.0322
```

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

```
##   Sample_ID   SO4_ppm     Cl_ppm
## 1           2322.4204 33352.2387
## 2           2322.4204  5714.1618
## 3           2322.4204     0.0061
## 4           2322.4204  5966.0950
## 5           2322.4204         NA
## 6            592.7974 33352.2387
```

```
## 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       1_MSM_202310_UP_LysA_20cm  156.4865  2514.586
## 27      10_MSM_202310_TR_LysA_45cm  543.5656  4326.371
## 28  10_MSM_202310_TR_LysA_45cm_dup  533.3889  4242.405
## 29     100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979
## 30 100_GWI_202310_TR_LysC_45cm_dup 1151.9205  9252.682
## 31     101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559
```

```
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       1_MSM_202310_UP_LysA_20cm  156.4865  2514.586
## 27      10_MSM_202310_TR_LysA_45cm  543.5656  4326.371
## 28  10_MSM_202310_TR_LysA_45cm_dup  533.3889  4242.405
## 29     100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979
## 30 100_GWI_202310_TR_LysC_45cm_dup 1151.9205  9252.682
## 31     101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559
```

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

```
stds <- all_dat[grepl("Standard", all_dat$Sample_ID),]
#stds <- stds[-c(80),]   #this is if you need to remove one for any reason
head(stds)
```

```
##      Sample_ID SO4_ppm Cl_ppm
## 691 Standard 1  0.5460 5.6530
## 692 Standard 1  0.5460 4.9368
## 693 Standard 1  0.5460 5.1270
```

```
## 694 Standard 1  0.5460 5.6753
## 695 Standard 1  0.5460 5.5434
## 696 Standard 1  0.4573 5.6530
```

```
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  0.516 0.0355  6.87 NO, rerun
## 2 Standard 2  1.09  0.0604  5.55 NO, rerun
## 3 Standard 3  2.14  0.107   5.03 NO, rerun
## 4 Standard 4 11.0   0.504   4.57 YES
## 5 Standard 5 21.4   0.977   4.55 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   5.39  0.306  5.67 NO, rerun
## 2 Standard 2  11.0   0.538  4.91 YES
## 3 Standard 3  21.4   1.05   4.92 YES
## 4 Standard 4  89.4  45.6   51.1  NO, rerun
## 5 Standard 5 216.   10.1    4.70 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       1_MSM_202310_UP_LysA_20cm  156.4865  2514.586
## 27      10_MSM_202310_TR_LysA_45cm  543.5656  4326.371
## 28  10_MSM_202310_TR_LysA_45cm_dup  533.3889  4242.405
## 29     100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979
## 30 100_GWI_202310_TR_LysC_45cm_dup 1151.9205  9252.682
## 31     101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559
```

```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       1_MSM_202310_UP_LysA_20cm  156.4865  2514.586  4.881051  70.93331
## 27      10_MSM_202310_TR_LysA_45cm  543.5656  4326.371 16.954635 122.04149
## 28  10_MSM_202310_TR_LysA_45cm_dup  533.3889  4242.405 16.637208 119.67292
## 29     100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979 36.468936 266.14891
## 30 100_GWI_202310_TR_LysC_45cm_dup 1151.9205  9252.682 35.930147 261.00655
## 31     101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559 56.052024 411.01717
##     salinity
## 26  4.543882
## 27  7.817778
## 28  7.666052
## 29 17.049033
## 30 16.719623
## 31 26.329026
```

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       1_MSM_202310_UP_LysA_20cm  156.4865  2514.586  4.881051  70.93331
## 27      10_MSM_202310_TR_LysA_45cm  543.5656  4326.371 16.954635 122.04149
## 28  10_MSM_202310_TR_LysA_45cm_dup  533.3889  4242.405 16.637208 119.67292
## 29     100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979 36.468936 266.14891
## 30 100_GWI_202310_TR_LysC_45cm_dup 1151.9205  9252.682 35.930147 261.00655
## 31     101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559 56.052024 411.01717
##     salinity
## 26  4.543882
## 27  7.817778
## 28  7.666052
```

```
## 29 17.049033
## 30 16.719623
## 31 26.329026
```

```
#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        10_MSM_202310_TR_LysA_45cm_dup  533.3889  4242.405 16.63721 119.6729
## 2       100_GWI_202310_TR_LysC_45cm_dup 1151.9205  9252.682 35.93015 261.0065
## 3               110_GWI_202310_SW_B_dup 1924.4210 13794.211 60.02561 389.1174
## 4   120_MSM_202310_TR_RHZ_SF_Tree_3_dup  938.2737  8470.767 29.26618 238.9497
## 5 130_MSM_202310_WC_RHZ_SF_Collar_5_dup 1047.0756  9863.028 32.65988 278.2236
## 6        20_MSM_202310_WC_LysB_10cm_dup  872.9288  9105.025 27.22797 256.8413
##    salinity
## 1  7.666052
## 2 16.719623
## 3 24.926166
## 4 15.306701
## 5 17.822518
## 6 16.452806
```

```
#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   1_MSM_202310_UP_LysA_20cm  156.4865  2514.586  4.881051  70.93331  4.543882
## 2  10_MSM_202310_TR_LysA_45cm  543.5656  4326.371 16.954635 122.04149  7.817778
## 3 100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979 36.468936 266.14891 17.049033
## 4 101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559 56.052024 411.01717 26.329026
## 5 102_GWI_202310_WC_LysA_20cm 1448.6028 13511.788 45.184117 381.15059 24.415828
## 6 103_GWI_202310_WC_LysA_45cm 1413.1583 12517.120 44.078550 353.09223 22.618461
```

```
#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        10_MSM_202310_TR_LysA_45cm   16.63721  119.6729     7.666052
## 2       100_GWI_202310_TR_LysC_45cm   35.93015  261.0065    16.719623
## 3               110_GWI_202310_SW_B   60.02561  389.1174    24.926166
## 4   120_MSM_202310_TR_RHZ_SF_Tree_3   29.26618  238.9497    15.306701
## 5 130_MSM_202310_WC_RHZ_SF_Collar_5   32.65988  278.2236    17.822518
## 6        20_MSM_202310_WC_LysB_10cm   27.22797  256.8413    16.452806
```

```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        10_MSM_202310_TR_LysA_45cm  543.5656 4326.3707 16.95464 122.041487
## 2       100_GWI_202310_TR_LysC_45cm 1169.1941 9434.9790 36.46894 266.148914
## 3               110_GWI_202310_SW_B 1925.0477  108.1539 60.04516   3.050886
## 4   120_MSM_202310_TR_RHZ_SF_Tree_3  904.1450 8163.1000 28.20165 230.270804
## 5 130_MSM_202310_WC_RHZ_SF_Collar_5 1056.5185 9928.2328 32.95441 280.062984
## 6        20_MSM_202310_WC_LysB_10cm  888.9918 9296.0155 27.72900 262.228928
##     salinity SO4_mM_dup Cl_mM_dup salinity_dup
## 1  7.8177779   16.63721  119.6729     7.666052
## 2 17.0490331   35.93015  261.0065    16.719623
## 3  0.1954601   60.02561  389.1174    24.926166
## 4 14.7507477   29.26618  238.9497    15.306701
## 5 17.9403427   32.65988  278.2236    17.822518
## 6 16.7979260   27.22797  256.8413    16.452806
```

```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)
```

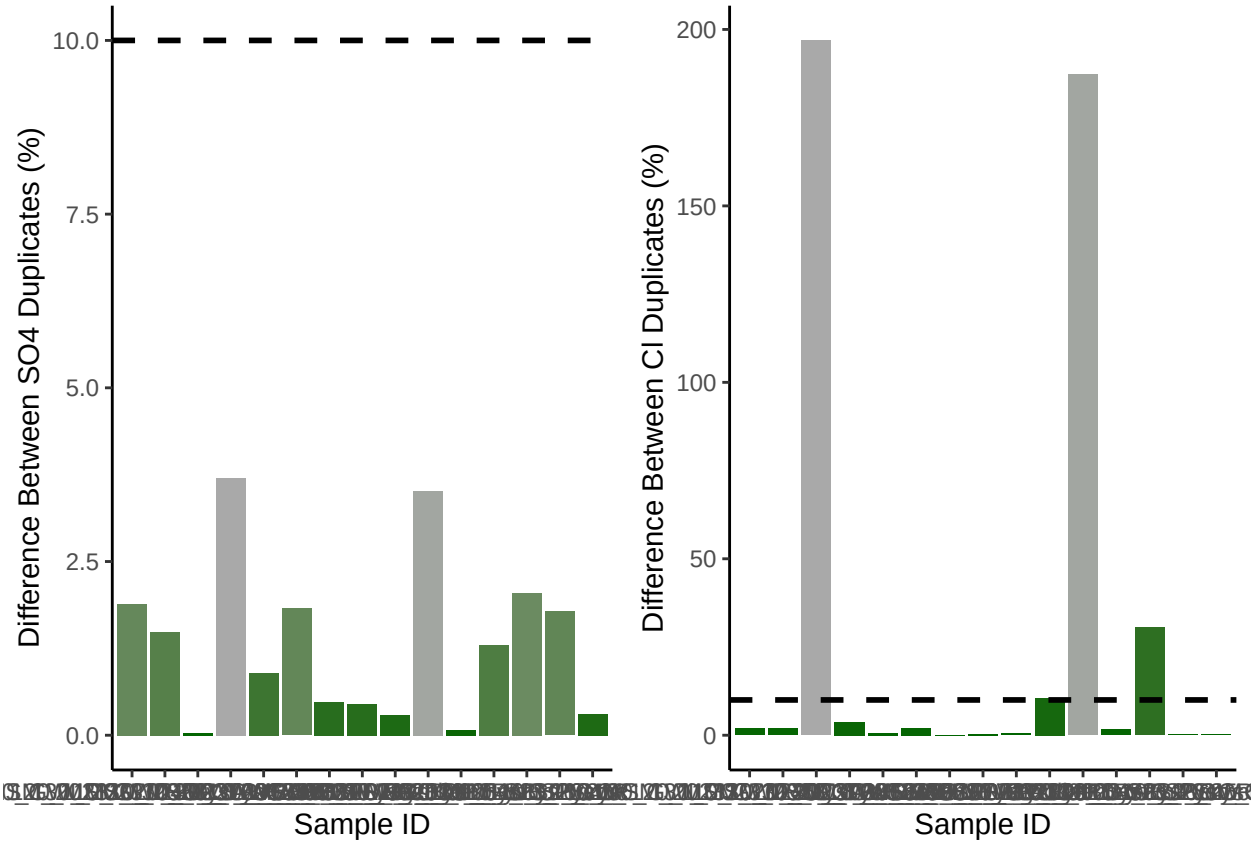
```
##                            Sample_ID   SO4_ppm    Cl_ppm   SO4_mM      Cl_mM
## 1        10_MSM_202310_TR_LysA_45cm  543.5656 4326.3707 16.95464 122.041487
## 2       100_GWI_202310_TR_LysC_45cm 1169.1941 9434.9790 36.46894 266.148914
## 3               110_GWI_202310_SW_B 1925.0477  108.1539 60.04516   3.050886
## 4   120_MSM_202310_TR_RHZ_SF_Tree_3  904.1450 8163.1000 28.20165 230.270804
## 5 130_MSM_202310_WC_RHZ_SF_Collar_5 1056.5185 9928.2328 32.95441 280.062984
## 6        20_MSM_202310_WC_LysB_10cm  888.9918 9296.0155 27.72900 262.228928
##     salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  7.8177779   16.63721  119.6729     7.666052   1.88990343           YES
## 2 17.0490331   35.93015  261.0065    16.719623   1.48838838           YES
## 3  0.1954601   60.02561  389.1174    24.926166   0.03256034           YES
## 4 14.7507477   29.26618  238.9497    15.306701   3.70477134           YES
## 5 17.9403427   32.65988  278.2236    17.822518   0.89778727           YES
## 6 16.7979260   27.22797  256.8413    16.452806   1.82335118           YES
##   Cl_dups_chk Cl_dups_flag
## 1   1.9598039          YES
## 2   1.9509868          YES
## 3 196.8881871    NO, rerun
## 4   3.6992794          YES
## 5   0.6589252          YES
## 6   2.0758663          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)
```




```
#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        15 NO, rerun          4    15       100   26.66667
## 2  YES        15       YES         11    15       100   73.33333
```

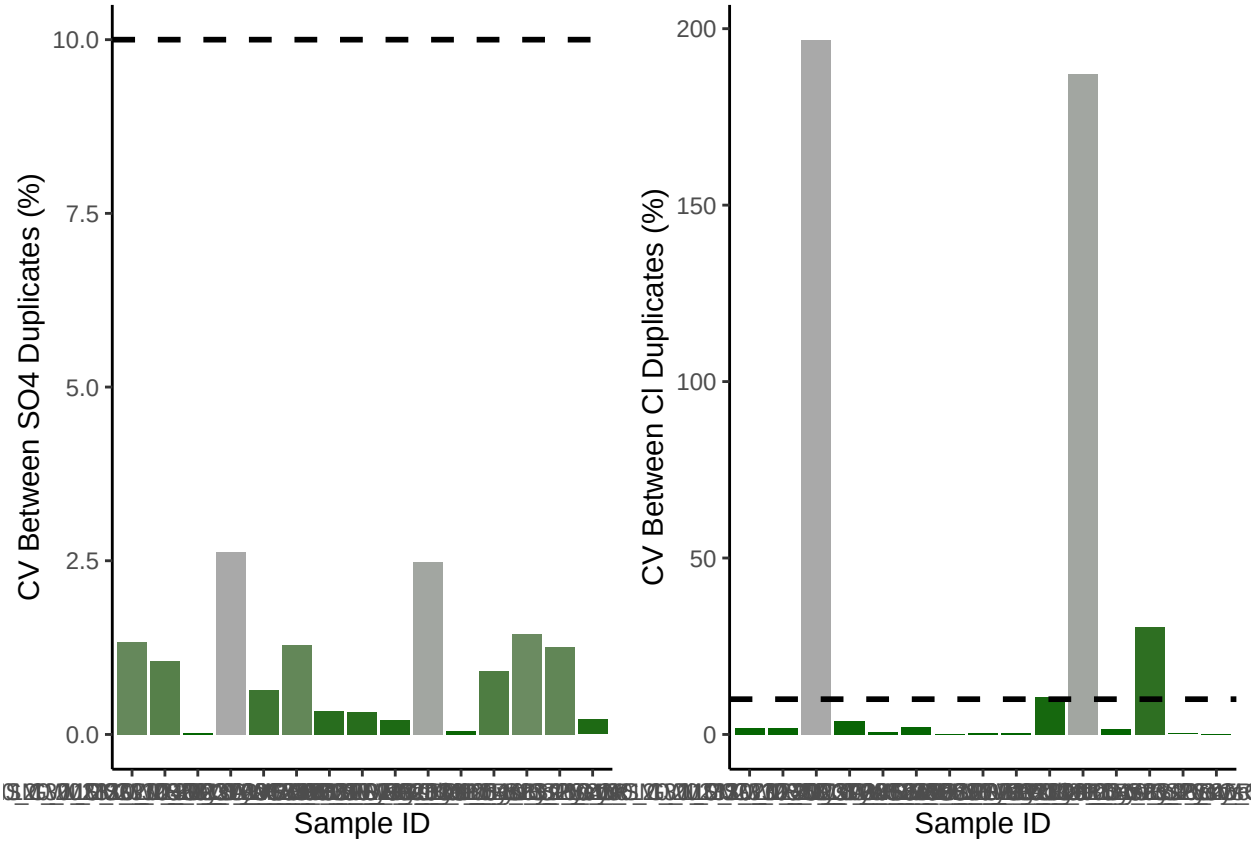
Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

```
##                          Sample_ID   SO4_ppm    Cl_ppm   SO4_mM      Cl_mM
## 1       10_MSM_202310_TR_LysA_45cm  543.5656 4326.3707 16.95464 122.041487
## 2      100_GWI_202310_TR_LysC_45cm 1169.1941 9434.9790 36.46894 266.148914
## 3              110_GWI_202310_SW_B 1925.0477  108.1539 60.04516   3.050886
## 4   120_MSM_202310_TR_RHZ_SF_Tree_3  904.1450 8163.1000 28.20165 230.270804
## 5 130_MSM_202310_WC_RHZ_SF_Collar_5 1056.5185 9928.2328 32.95441 280.062984
## 6       20_MSM_202310_WC_LysB_10cm  888.9918 9296.0155 27.72900 262.228928
##     salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  7.8177779   16.63721  119.6729     7.666052   1.88990343           YES
## 2 17.0490331   35.93015  261.0065    16.719623   1.48838838           YES
## 3  0.1954601   60.02561  389.1174    24.926166   0.03256034           YES
## 4 14.7507477   29.26618  238.9497    15.306701   3.70477134           YES
## 5 17.9403427   32.65988  278.2236    17.822518   0.89778727           YES
## 6 16.7979260   27.22797  256.8413    16.452806   1.82335118           YES
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1   1.9598039          YES  1.33636353              YES
## 2   1.9509868          YES  1.05244952              YES
## 3 196.8881871    NO, rerun  0.02302364              YES
## 4   3.6992794          YES  2.61966894              YES
## 5   0.6589252          YES  0.63483147              YES
```

```
## 6   2.0758663          YES  1.28930398              YES
```

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


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

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




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_dups <- QAdups %>%
  group_by(SO4_dups_flag) %>%
  summarise(S_no_rows = length(SO4_dups_flag))
Perc_dups1 <- QAdups %>%
```

```
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(Cl_dups_flag))
colnames(Perc_dups) <- c("Flag", "S_no_rows")
colnames(Perc_dups1) <- c("Flag", "Cl_no_rows")
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)

Perc_dups2$Total <- length(QAdups$SO4_dups_flag)
Perc_dups2$S_Percent <- (Perc_dups2$S_no_rows / Perc_dups2$Total)*100
Perc_dups2$Cl_Percent <- (Perc_dups2$Cl_no_rows / Perc_dups2$Total)*100
head(Perc_dups2)
```

```
##   Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES        15 NO, rerun          4    15       100   26.66667
## 2  YES        15       YES         11    15       100   73.33333
```

Pull out spikes and check

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

```
##                          Sample_ID   SO4_ppm    Cl_ppm    SO4_mM     Cl_mM
## 26       1_MSM_202310_UP_LysA_20cm  156.4865  2514.586  4.881051  70.93331
## 27      10_MSM_202310_TR_LysA_45cm  543.5656  4326.371 16.954635 122.04149
## 28  10_MSM_202310_TR_LysA_45cm_dup  533.3889  4242.405 16.637208 119.67292
## 29     100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979 36.468936 266.14891
## 30 100_GWI_202310_TR_LysC_45cm_dup 1151.9205  9252.682 35.930147 261.00655
## 31     101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559 56.052024 411.01717
##     salinity
## 26  4.543882
## 27  7.817778
## 28  7.666052
## 29 17.049033
## 30 16.719623
## 31 26.329026
```

```
#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_202310_WC_LysA_10cm_spk 2123.4532 13985.840 66.23372 394.5230
## 2      11_MSM_202310_TR_LysB_10cm_spk  824.2115  6953.347 25.70841 196.1452
## 3             111_GWI_202310_SW_C_spk 2322.4204 13800.892 72.43981 389.3058
## 4 121_MSM_202310_TR_RHZ_SF_Tree_4_spk 1015.5222  8245.100 31.67568 232.5839
## 5      131_MSM_202307_WC_RHZ_LysA_spk 1420.6403  9776.295 44.31192 275.7770
## 6      21_MSM_202310_WC_LysB_20cm_spk  931.4789  9421.024 29.05424 265.7553
##   salinity
## 1 25.27244
## 2 12.56472
```

```
## 3 24.93824
## 4 14.89892
## 5 17.66579
## 6 17.02382
```

```
#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_202310_WC_LysA_10cm   66.23372
## 2      11_MSM_202310_TR_LysB_10cm   25.70841
## 3             111_GWI_202310_SW_C   72.43981
## 4 121_MSM_202310_TR_RHZ_SF_Tree_4   31.67568
## 5      131_MSM_202307_WC_RHZ_LysA   44.31192
## 6      21_MSM_202310_WC_LysB_20cm   29.05424
```

```
#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_202310_WC_LysA_10cm 1797.0279 14570.5587 56.052024 411.01717
## 2      11_MSM_202310_TR_LysB_10cm  744.7543  7094.4554 23.230016 200.12568
## 3             111_GWI_202310_SW_C 1946.6882 13991.2256 60.720156 394.67491
## 4 121_MSM_202310_TR_RHZ_SF_Tree_4  894.6568  8047.9702 27.905702 227.02314
## 5      21_MSM_202310_WC_LysB_20cm  747.1359  9510.8725 23.304301 268.28977
## 6      31_GCW_202310_TR_LysB_20cm   50.6452   776.0658  1.579701  21.89184
##    salinity SO4_mM_spk
## 1 26.329026  66.233724
## 2 12.819707  25.708406
## 3 25.282171  72.439813
## 4 14.542708  31.675677
## 5 17.186173  29.054239
## 6  1.402377   4.434775
```

```
#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_202310_WC_LysA_10cm 1797.0279 14570.5587 56.052024 411.01717
## 2      11_MSM_202310_TR_LysB_10cm  744.7543  7094.4554 23.230016 200.12568
## 3             111_GWI_202310_SW_C 1946.6882 13991.2256 60.720156 394.67491
## 4 121_MSM_202310_TR_RHZ_SF_Tree_4  894.6568  8047.9702 27.905702 227.02314
## 5      21_MSM_202310_WC_LysB_20cm  747.1359  9510.8725 23.304301 268.28977
```

```
## 6      31_GCW_202310_TR_LysB_20cm   50.6452   776.0658  1.579701  21.89184
##    salinity SO4_mM_spk SO4_spk_Conc
## 1 26.329026  66.233724 7.797879e-05
## 2 12.819707  25.708406 7.797879e-05
## 3 25.282171  72.439813 7.797879e-05
## 4 14.542708  31.675677 7.797879e-05
## 5 17.186173  29.054239 7.797879e-05
## 6  1.402377   4.434775 7.797879e-05
```

```
#need to determine dilution factors and initial amount of sample added
#if your samples are all the same dilution just use the first line
#for Steph / COMPASS this depends on the site so...
QAspks$Dilution <- 1
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_202306_UP"), 50, QAspks$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$Dilution)
QAspks$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)
```

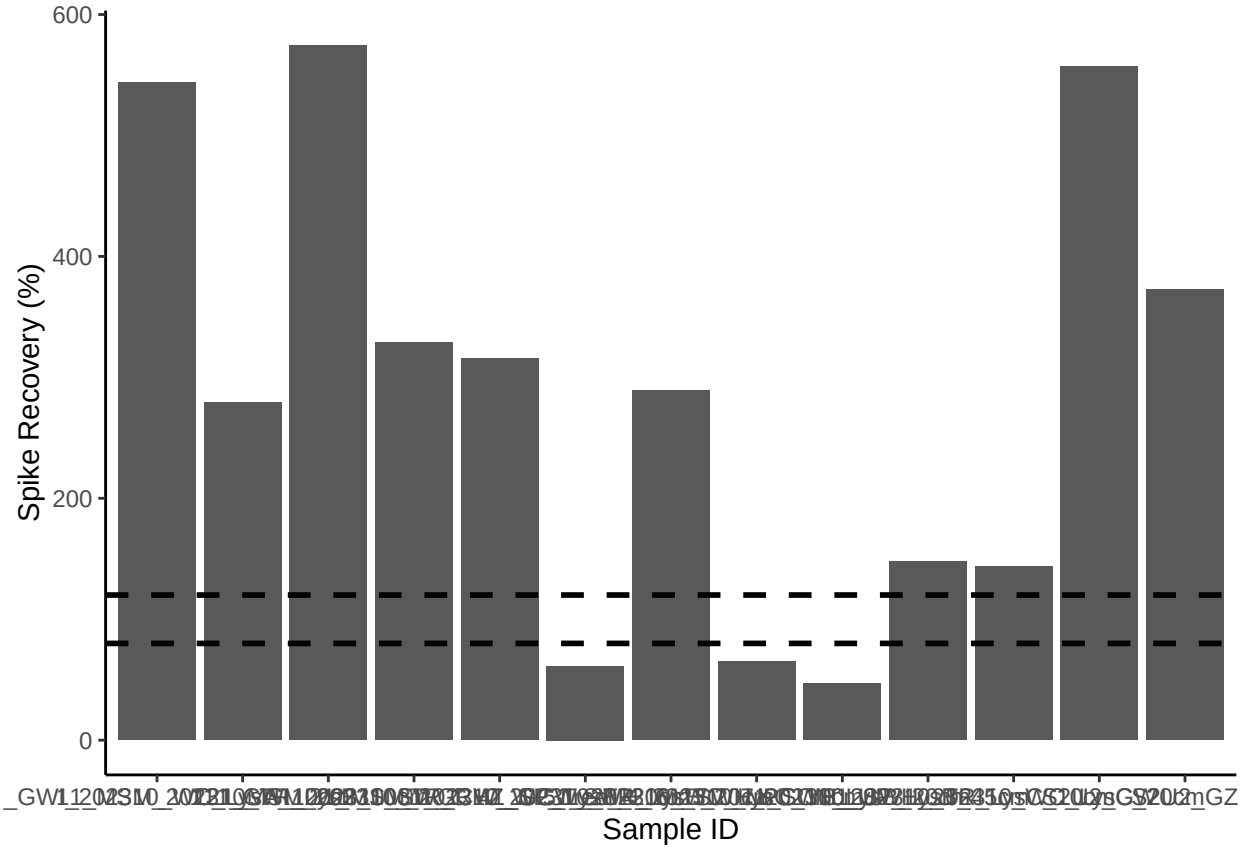
```
##                          Sample_ID   SO4_ppm     Cl_ppm    SO4_mM     Cl_mM
## 1     101_GWI_202310_WC_LysA_10cm 1797.0279 14570.5587 56.052024 411.01717
## 2      11_MSM_202310_TR_LysB_10cm  744.7543  7094.4554 23.230016 200.12568
## 3             111_GWI_202310_SW_C 1946.6882 13991.2256 60.720156 394.67491
## 4 121_MSM_202310_TR_RHZ_SF_Tree_4  894.6568  8047.9702 27.905702 227.02314
## 5      21_MSM_202310_WC_LysB_20cm  747.1359  9510.8725 23.304301 268.28977
## 6      31_GCW_202310_TR_LysB_20cm   50.6452   776.0658  1.579701  21.89184
##    salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 26.329026  66.233724 7.797879e-05        1     1e-06
```

```
## 2 12.819707  25.708406 7.797879e-05        1     1e-06
## 3 25.282171  72.439813 7.797879e-05        1     1e-06
## 4 14.542708  31.675677 7.797879e-05        1     1e-06
## 5 17.186173  29.054239 7.797879e-05        1     1e-06
## 6  1.402377   4.434775 7.797879e-05        1     1e-06
```

```
#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_202310_WC_LysA_10cm 1797.0279 14570.5587 56.052024 411.01717
## 2      11_MSM_202310_TR_LysB_10cm  744.7543  7094.4554 23.230016 200.12568
## 3             111_GWI_202310_SW_C 1946.6882 13991.2256 60.720156 394.67491
## 4 121_MSM_202310_TR_RHZ_SF_Tree_4  894.6568  8047.9702 27.905702 227.02314
## 5      21_MSM_202310_WC_LysB_20cm  747.1359  9510.8725 23.304301 268.28977
## 6      31_GCW_202310_TR_LysB_20cm   50.6452   776.0658  1.579701  21.89184
##    salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1 26.329026  66.233724 7.797879e-05        1     1e-06     5.605202e-05
## 2 12.819707  25.708406 7.797879e-05        1     1e-06     2.323002e-05
## 3 25.282171  72.439813 7.797879e-05        1     1e-06     6.072016e-05
## 4 14.542708  31.675677 7.797879e-05        1     1e-06     2.790570e-05
## 5 17.186173  29.054239 7.797879e-05        1     1e-06     2.330430e-05
## 6  1.402377   4.434775 7.797879e-05        1     1e-06     1.579701e-06
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   7.285710e-04    1.340308e-04    543.58468     NO, rerun
## 2   2.827925e-04    1.012088e-04    279.41489     NO, rerun
## 3   7.968379e-04    1.386989e-04    574.50901     NO, rerun
## 4   3.484324e-04    1.058845e-04    329.06844     NO, rerun
## 5   3.195966e-04    1.012831e-04    315.54786     NO, rerun
## 6   4.878253e-05    7.955849e-05     61.31656     NO, rerun
```

```
#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 NO, rerun          13    13     100
```

Make final dataframe with IDs

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

```
## Warning in rbind(c("1", "MSM", "202310", "UP", "LysA", "20cm"), c("10", :
## 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           1  MSM 202310   UP      LysA  20cm   1     MSM
## 2          10  MSM 202310   TR      LysA  45cm  10     MSM
## 3         100  GWI 202310   TR      LysC  45cm 100     GWI
## 4         101  GWI 202310   WC      LysA  10cm 101     GWI
## 5         102  GWI 202310   WC      LysA  20cm 102     GWI
## 6         103  GWI 202310   WC      LysA  45cm 103     GWI
```

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

```
##   Analysis_No Site   Date Zone Replicate Depth RHZ RHZ_Rep
## 1           1  MSM 202310   UP      LysA  20cm   1     MSM
## 2          10  MSM 202310   TR      LysA  45cm  10     MSM
## 3         100  GWI 202310   TR      LysC  45cm 100     GWI
## 4         101  GWI 202310   WC      LysA  10cm 101     GWI
## 5         102  GWI 202310   WC      LysA  20cm 102     GWI
## 6         103  GWI 202310   WC      LysA  45cm 103     GWI
##                     Sample_ID   SO4_ppm    Cl_ppm    SO4_mM     Cl_mM  salinity
## 1   1_MSM_202310_UP_LysA_20cm  156.4865  2514.586  4.881051  70.93331  4.543882
## 2  10_MSM_202310_TR_LysA_45cm  543.5656  4326.371 16.954635 122.04149  7.817778
## 3 100_GWI_202310_TR_LysC_45cm 1169.1941  9434.979 36.468936 266.14891 17.049033
## 4 101_GWI_202310_WC_LysA_10cm 1797.0279 14570.559 56.052024 411.01717 26.329026
## 5 102_GWI_202310_WC_LysA_20cm 1448.6028 13511.788 45.184117 381.15059 24.415828
## 6 103_GWI_202310_WC_LysA_45cm 1413.1583 12517.120 44.078550 353.09223 22.618461
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_Synoptic_CB_Dionex_ClSo4_Final_202310.csv")          #C
```

END