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_202311_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  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 6 6 Standard 1 Calibration Standard 5.797   0.4644     6.54   0.5648     3.18
```

```
## 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  Lab Blank     n.a.
## 6 Standard 1  0.4644
```

```
## 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      NA
## 3  Lab Blank      NA
## 4  Lab Blank      NA
## 5  Lab Blank      NA
## 6 Standard 1  0.4644
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202311_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  Lab Blank              Unknown   n.a.    n.a.    n.a.    n.a.    n.a.
## 6 6 Standard 1 Calibration Standard 4.567  5.0079   93.30  8.0534   55.78
```

```
## 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  Lab Blank    n.a.
## 6 Standard 1 5.0079
```

```
## 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     NA
## 2  Lab Blank     NA
## 3  Lab Blank     NA
## 4  Lab Blank     NA
## 5  Lab Blank     NA
## 6 Standard 1 5.0079
```

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

```
##   Sample_ID   SO4_ppm     Cl_ppm
## 1           1821.9714 14254.5661
## 2           1821.9714  4583.2258
## 3           1821.9714     0.0054
## 4           1821.9714  4220.2294
## 5           1821.9714         NA
## 6            503.9588 14254.5661
```

```
## 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_202311_UP_LysA_20cm  200.7171  2967.900
## 27      10_MSM_202311_TR_LysA_45cm  532.8335  4351.421
## 28  10_MSM_202311_TR_LysA_45cm_dup  534.2280  4378.928
## 29     100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329
## 30 100_GWI_202311_WC_LysA_10cm_dup 1544.0406 12622.756
## 31     101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407
```

```
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_202311_UP_LysA_20cm  200.7171  2967.900
## 27      10_MSM_202311_TR_LysA_45cm  532.8335  4351.421
## 28  10_MSM_202311_TR_LysA_45cm_dup  534.2280  4378.928
## 29     100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329
## 30 100_GWI_202311_WC_LysA_10cm_dup 1544.0406 12622.756
## 31     101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407
```

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
## 625 Standard 1  0.5148 5.1969
## 626 Standard 1  0.5148 5.2412
## 627 Standard 1  0.5148 5.2076
```

```
## 628 Standard 1  0.5148 5.1488
## 629 Standard 1  0.5148 5.0079
## 630 Standard 1  0.5053 5.1969
```

```
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  0.496 0.0184 3.72  YES
## 2 Standard 2  1.02  0.0212 2.08  YES
## 3 Standard 3  2.04  0.0263 1.29  YES
## 4 Standard 4 10.2   0.0990 0.969 YES
## 5 Standard 5 20.3   0.225  1.11  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   5.16 0.0835  1.62 YES
## 2 Standard 2  10.4  0.134   1.28 YES
## 3 Standard 3  20.6  0.248   1.20 YES
## 4 Standard 4 102.   1.17    1.14 YES
## 5 Standard 5 203.   2.13    1.05 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_202311_UP_LysA_20cm  200.7171  2967.900
## 27      10_MSM_202311_TR_LysA_45cm  532.8335  4351.421
## 28  10_MSM_202311_TR_LysA_45cm_dup  534.2280  4378.928
## 29     100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329
## 30 100_GWI_202311_WC_LysA_10cm_dup 1544.0406 12622.756
## 31     101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407
```

```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_202311_UP_LysA_20cm  200.7171  2967.900  6.260671  83.72074
## 27      10_MSM_202311_TR_LysA_45cm  532.8335  4351.421 16.619885 122.74811
## 28  10_MSM_202311_TR_LysA_45cm_dup  534.2280  4378.928 16.663381 123.52407
## 29     100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.866254 361.64539
## 30 100_GWI_202311_WC_LysA_10cm_dup 1544.0406 12622.756 48.160967 356.07211
## 31     101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623
##     salinity
## 26  5.363022
## 27  7.863043
## 28  7.912749
## 29 23.166360
## 30 22.809347
## 31 21.977495
```

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_202311_UP_LysA_20cm  200.7171  2967.900  6.260671  83.72074
## 27      10_MSM_202311_TR_LysA_45cm  532.8335  4351.421 16.619885 122.74811
## 28  10_MSM_202311_TR_LysA_45cm_dup  534.2280  4378.928 16.663381 123.52407
## 29     100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.866254 361.64539
## 30 100_GWI_202311_WC_LysA_10cm_dup 1544.0406 12622.756 48.160967 356.07211
## 31     101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623
##     salinity
## 26  5.363022
## 27  7.863043
## 28  7.912749
```

```
## 29 23.166360
## 30 22.809347
## 31 21.977495
```

```
#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_202311_TR_LysA_45cm_dup  534.2280  4378.928 16.66338 123.5241
## 2       100_GWI_202311_WC_LysA_10cm_dup 1544.0406 12622.756 48.16097 356.0721
## 3               110_GWI_202311_SW_C_dup 1736.4688 12657.213 54.16309 357.0441
## 4   120_MSM_202311_TR_RHZ_SF_Tree_3_dup  877.2742  8017.520 27.36351 226.1642
## 5 130_MSM_202311_WC_RHZ_SF_Collar_5_dup 1036.8625  9369.887 32.34131 264.3127
## 6        20_MSM_202311_WC_LysB_20cm_dup  821.2223  9048.720 25.61517 255.2530
##    salinity
## 1  7.912749
## 2 22.809347
## 3 22.871611
## 4 14.487685
## 5 16.931411
## 6 16.351064
```

```
#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_202311_UP_LysA_20cm  200.7171  2967.900  6.260671  83.72074  5.363022
## 2  10_MSM_202311_TR_LysA_45cm  532.8335  4351.421 16.619885 122.74811  7.863043
## 3 100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.866254 361.64539 23.166360
## 4 101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623 21.977495
## 5 102_GWI_202311_WC_LysA_45cm 1199.1944 10824.812 37.404691 305.35435 19.560461
## 6 103_GWI_202311_WC_LysB_10cm 1592.8158 13596.121 49.682339 383.52951 24.568217
```

```
#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_202311_TR_LysA_45cm   16.66338  123.5241     7.912749
## 2       100_GWI_202311_WC_LysA_10cm   48.16097  356.0721    22.809347
## 3               110_GWI_202311_SW_C   54.16309  357.0441    22.871611
## 4   120_MSM_202311_TR_RHZ_SF_Tree_3   27.36351  226.1642    14.487685
## 5 130_MSM_202311_WC_RHZ_SF_Collar_5   32.34131  264.3127    16.931411
## 6        20_MSM_202311_WC_LysB_20cm   25.61517  255.2530    16.351064
```

```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_202311_TR_LysA_45cm  532.8335  4351.421 16.61988 122.7481
## 2       100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.86625 361.6454
## 3               110_GWI_202311_SW_C 1766.9177 12862.314 55.11284 362.8297
## 4   120_MSM_202311_TR_RHZ_SF_Tree_3  858.8025  7845.456 26.78735 221.3105
## 5 130_MSM_202311_WC_RHZ_SF_Collar_5 1046.7724  9517.950 32.65042 268.4894
## 6        20_MSM_202311_WC_LysB_20cm  842.8573  9269.786 26.29000 261.4890
##    salinity SO4_mM_dup Cl_mM_dup salinity_dup
## 1  7.863043   16.66338  123.5241     7.912749
## 2 23.166360   48.16097  356.0721    22.809347
## 3 23.242227   54.16309  357.0441    22.871611
## 4 14.176765   27.36351  226.1642    14.487685
## 5 17.198962   32.34131  264.3127    16.931411
## 6 16.750529   25.61517  255.2530    16.351064
```

```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)
```

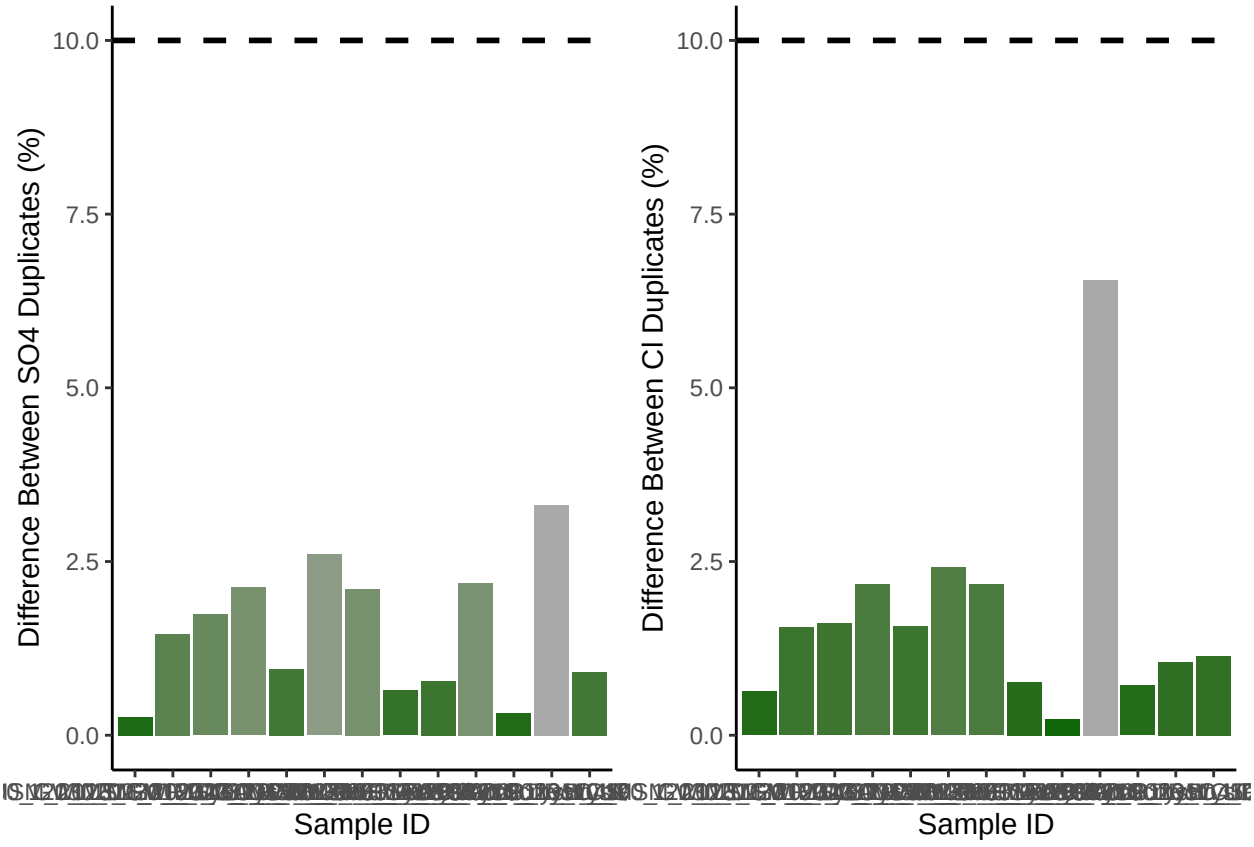
```
##                            Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM
## 1        10_MSM_202311_TR_LysA_45cm  532.8335  4351.421 16.61988 122.7481
## 2       100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.86625 361.6454
## 3               110_GWI_202311_SW_C 1766.9177 12862.314 55.11284 362.8297
## 4   120_MSM_202311_TR_RHZ_SF_Tree_3  858.8025  7845.456 26.78735 221.3105
## 5 130_MSM_202311_WC_RHZ_SF_Collar_5 1046.7724  9517.950 32.65042 268.4894
## 6        20_MSM_202311_WC_LysB_20cm  842.8573  9269.786 26.29000 261.4890
##    salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  7.863043   16.66338  123.5241     7.912749    0.2613720           YES
## 2 23.166360   48.16097  356.0721    22.809347    1.4537919           YES
## 3 23.242227   54.16309  357.0441    22.871611    1.7382553           YES
## 4 14.176765   27.36351  226.1642    14.487685    2.1279820           YES
## 5 17.198962   32.34131  264.3127    16.931411    0.9512127           YES
## 6 16.750529   25.61517  255.2530    16.351064    2.6002362           YES
##   Cl_dups_chk Cl_dups_flag
## 1   0.6301604          YES
## 2   1.5530546          YES
## 3   1.6073981          YES
## 4   2.1693861          YES
## 5   1.5678185          YES
## 6   2.4135756          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)
```

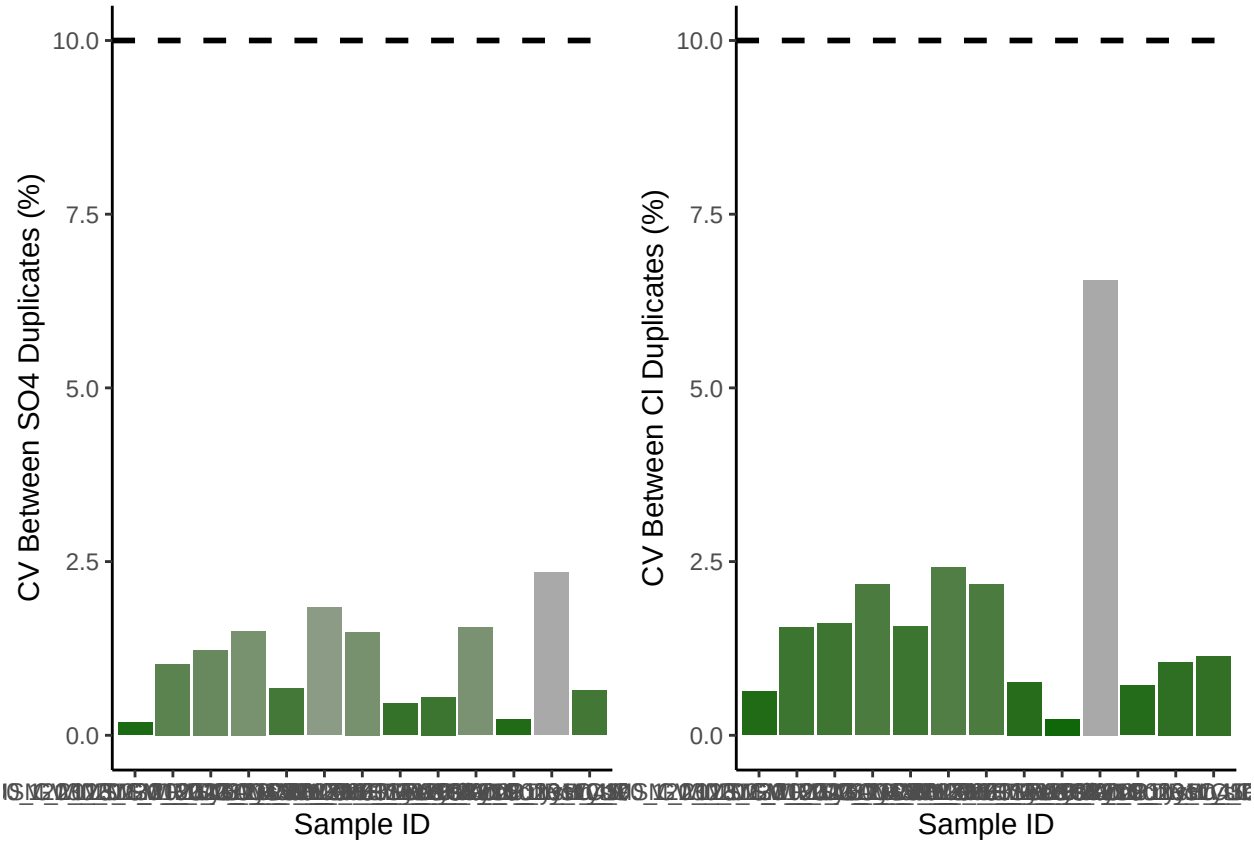
```
##   Flag S_no_rows Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES        13  YES         13    13       100        100
```

Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

```
##                           Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM
## 1        10_MSM_202311_TR_LysA_45cm  532.8335  4351.421 16.61988 122.7481
## 2       100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.86625 361.6454
## 3               110_GWI_202311_SW_C 1766.9177 12862.314 55.11284 362.8297
## 4   120_MSM_202311_TR_RHZ_SF_Tree_3  858.8025  7845.456 26.78735 221.3105
## 5 130_MSM_202311_WC_RHZ_SF_Collar_5 1046.7724  9517.950 32.65042 268.4894
## 6        20_MSM_202311_WC_LysB_20cm  842.8573  9269.786 26.29000 261.4890
##    salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  7.863043   16.66338  123.5241     7.912749    0.2613720           YES
## 2 23.166360   48.16097  356.0721    22.809347    1.4537919           YES
## 3 23.242227   54.16309  357.0441    22.871611    1.7382553           YES
## 4 14.176765   27.36351  226.1642    14.487685    2.1279820           YES
## 5 17.198962   32.34131  264.3127    16.931411    0.9512127           YES
## 6 16.750529   25.61517  255.2530    16.351064    2.6002362           YES
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1   0.6301604          YES   0.1848179              YES
## 2   1.5530546          YES   1.0279861              YES
## 3   1.6073981          YES   1.2291321              YES
## 4   2.1693861          YES   1.5047105              YES
## 5   1.5678185          YES   0.6726090              YES
## 6   2.4135756          YES   1.8386446              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)
```




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

```
colnames(Perc_dups) <- c("Flag", "S_no_rows")
colnames(Perc_dups1) <- c("Flag", "Cl_no_rows")
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)

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

```
##   Flag S_no_rows Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES        13  YES         13    13       100        100
```

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_202311_UP_LysA_20cm  200.7171  2967.900  6.260671  83.72074
## 27      10_MSM_202311_TR_LysA_45cm  532.8335  4351.421 16.619885 122.74811
## 28  10_MSM_202311_TR_LysA_45cm_dup  534.2280  4378.928 16.663381 123.52407
## 29     100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.866254 361.64539
## 30 100_GWI_202311_WC_LysA_10cm_dup 1544.0406 12622.756 48.160967 356.07211
## 31     101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623
##     salinity
## 26  5.363022
## 27  7.863043
## 28  7.912749
## 29 23.166360
## 30 22.809347
## 31 21.977495
```

```
#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_202311_WC_LysA_20cm_spk 1651.5201 11892.932 51.51342 335.48469
## 2      11_MSM_202311_TR_LysB_10cm_spk  803.5885  7175.359 25.06514 202.40786
## 3 111_MSM_202311_UP_RHZ_SF_Tree_1_spk  398.1390  2820.719 12.41856  79.56895
## 4 121_MSM_202311_TR_RHZ_SF_Tree_4_spk  999.5127  7901.435 31.17632 222.88955
## 5      131_MSM_202311_WC_RHZ_LysA_spk 1481.2258  9900.823 46.20168 279.28978
## 6      21_MSM_202311_WC_LysB_45cm_spk  328.0532  8615.336 10.23248 243.02780
##    salinity
## 1 21.490554
## 2 12.965899
## 3  5.097066
## 4 14.277918
## 5 17.890813
## 6 15.567937
```

```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_202311_WC_LysA_20cm   51.51342
## 2      11_MSM_202311_TR_LysB_10cm   25.06514
## 3 111_MSM_202311_UP_RHZ_SF_Tree_1   12.41856
## 4 121_MSM_202311_TR_RHZ_SF_Tree_4   31.17632
## 5      131_MSM_202311_WC_RHZ_LysA   46.20168
## 6      21_MSM_202311_WC_LysB_45cm   10.23248
```

```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_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623
## 2      11_MSM_202311_TR_LysB_10cm  709.2730  7178.410 22.123300 202.49394
## 3 111_MSM_202311_UP_RHZ_SF_Tree_1  310.6557  2855.731  9.689822  80.55658
## 4 121_MSM_202311_TR_RHZ_SF_Tree_4  961.2419  8335.104 29.982592 235.12283
## 5      131_MSM_202311_WC_RHZ_LysA 1320.9987 10110.144 41.203952 285.19447
## 6      21_MSM_202311_WC_LysB_45cm  145.5572  8790.947  4.540150 247.98159
##    salinity SO4_mM_spk
## 1 21.977495   51.51342
## 2 12.971413   25.06514
## 3  5.160331   12.41856
## 4 15.061560   31.17632
## 5 18.269056   46.20168
## 6 15.885268   10.23248
```

```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_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623
## 2      11_MSM_202311_TR_LysB_10cm  709.2730  7178.410 22.123300 202.49394
## 3 111_MSM_202311_UP_RHZ_SF_Tree_1  310.6557  2855.731  9.689822  80.55658
## 4 121_MSM_202311_TR_RHZ_SF_Tree_4  961.2419  8335.104 29.982592 235.12283
## 5      131_MSM_202311_WC_RHZ_LysA 1320.9987 10110.144 41.203952 285.19447
## 6      21_MSM_202311_WC_LysB_45cm  145.5572  8790.947  4.540150 247.98159
##    salinity SO4_mM_spk SO4_spk_Conc
## 1 21.977495   51.51342 7.797879e-05
## 2 12.971413   25.06514 7.797879e-05
## 3  5.160331   12.41856 7.797879e-05
```

```
## 4 15.061560   31.17632 7.797879e-05
## 5 18.269056   46.20168 7.797879e-05
## 6 15.885268   10.23248 7.797879e-05
```

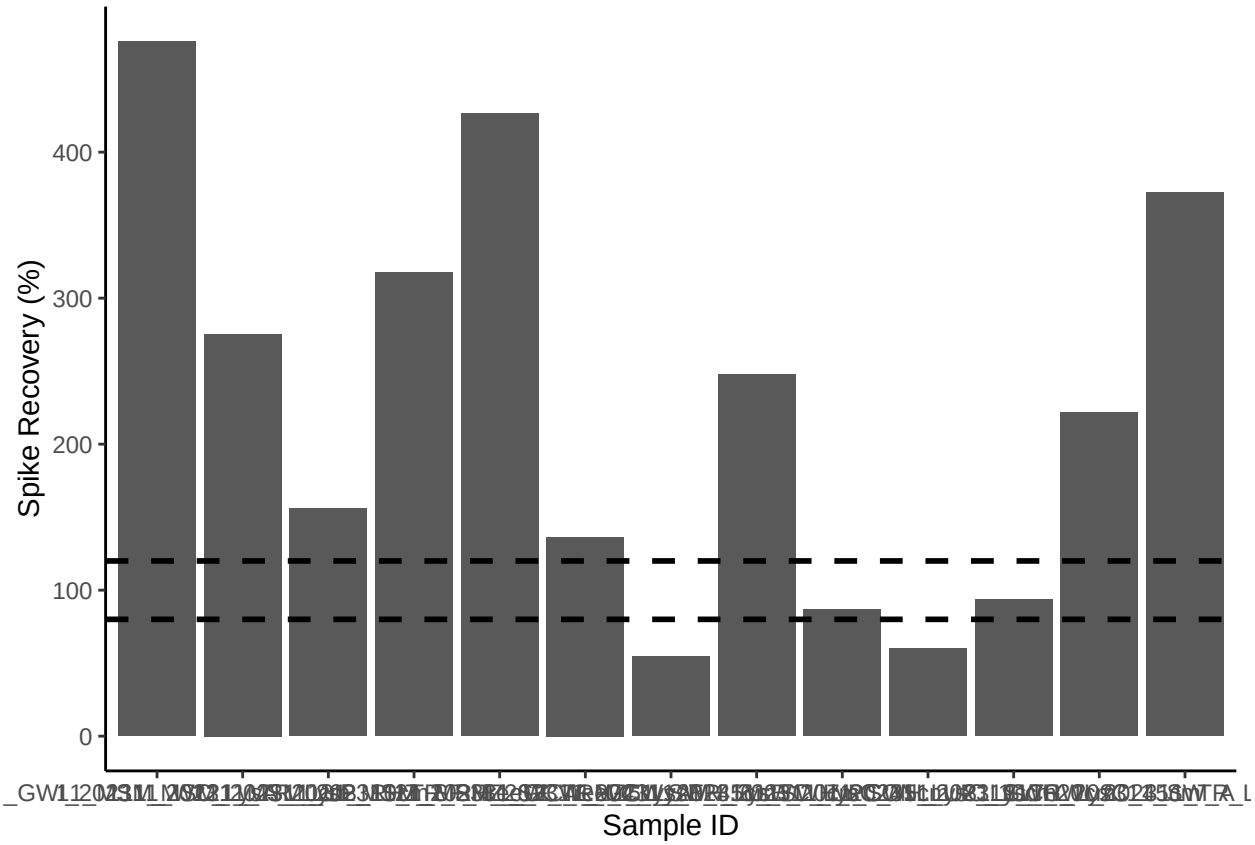
```
#need to determine dilution factors and initial amount of sample added
#if your samples are all the same dilution just use the first line
#for Steph / COMPASS this depends on the site so...
QAspks$Dilution <- 1
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_202306_UP"), 50, QAspks$Dilution)
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_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623
## 2      11_MSM_202311_TR_LysB_10cm  709.2730  7178.410 22.123300 202.49394
## 3 111_MSM_202311_UP_RHZ_SF_Tree_1  310.6557  2855.731  9.689822  80.55658
## 4 121_MSM_202311_TR_RHZ_SF_Tree_4  961.2419  8335.104 29.982592 235.12283
## 5      131_MSM_202311_WC_RHZ_LysA 1320.9987 10110.144 41.203952 285.19447
## 6      21_MSM_202311_WC_LysB_45cm  145.5572  8790.947  4.540150 247.98159
##    salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 21.977495   51.51342 7.797879e-05        1     1e-06
## 2 12.971413   25.06514 7.797879e-05        1     1e-06
## 3  5.160331   12.41856 7.797879e-05        1     1e-06
## 4 15.061560   31.17632 7.797879e-05        1     1e-06
## 5 18.269056   46.20168 7.797879e-05        1     1e-06
## 6 15.885268   10.23248 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_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623
## 2      11_MSM_202311_TR_LysB_10cm  709.2730  7178.410 22.123300 202.49394
## 3 111_MSM_202311_UP_RHZ_SF_Tree_1  310.6557  2855.731  9.689822  80.55658
## 4 121_MSM_202311_TR_RHZ_SF_Tree_4  961.2419  8335.104 29.982592 235.12283
## 5      131_MSM_202311_WC_RHZ_LysA 1320.9987 10110.144 41.203952 285.19447
## 6      21_MSM_202311_WC_LysB_45cm  145.5572  8790.947  4.540150 247.98159
##    salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1 21.977495   51.51342 7.797879e-05        1     1e-06     4.109203e-05
## 2 12.971413   25.06514 7.797879e-05        1     1e-06     2.212330e-05
## 3  5.160331   12.41856 7.797879e-05        1     1e-06     9.689822e-06
## 4 15.061560   31.17632 7.797879e-05        1     1e-06     2.998259e-05
## 5 18.269056   46.20168 7.797879e-05        1     1e-06     4.120395e-05
## 6 15.885268   10.23248 7.797879e-05        1     1e-06     4.540150e-06
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   0.0005666476    1.190708e-04     475.8912     NO, rerun
## 2   0.0002757166    1.001021e-04     275.4354     NO, rerun
## 3   0.0001366041    8.766861e-05     155.8188     NO, rerun
## 4   0.0003429395    1.079614e-04     317.6501     NO, rerun
## 5   0.0005082185    1.191827e-04     426.4195     NO, rerun
## 6   0.0001125572    8.251894e-05     136.4017     NO, rerun
```

```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: 2 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 NO, rerun          11    13    84.6
## 2 YES                 2    13    15.4
```

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", "202311", "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 202311   UP      LysA  20cm   1     MSM
## 2          10  MSM 202311   TR      LysA  45cm  10     MSM
## 3         100  GWI 202311   WC      LysA  10cm 100     GWI
## 4         101  GWI 202311   WC      LysA  20cm 101     GWI
## 5         102  GWI 202311   WC      LysA  45cm 102     GWI
## 6         103  GWI 202311   WC      LysB  10cm 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 202311   UP      LysA  20cm   1     MSM
## 2          10  MSM 202311   TR      LysA  45cm  10     MSM
## 3         100  GWI 202311   WC      LysA  10cm 100     GWI
## 4         101  GWI 202311   WC      LysA  20cm 101     GWI
## 5         102  GWI 202311   WC      LysA  45cm 102     GWI
## 6         103  GWI 202311   WC      LysB  10cm 103     GWI
##                     Sample_ID   SO4_ppm    Cl_ppm    SO4_mM     Cl_mM  salinity
## 1   1_MSM_202311_UP_LysA_20cm  200.7171  2967.900  6.260671  83.72074  5.363022
## 2  10_MSM_202311_TR_LysA_45cm  532.8335  4351.421 16.619885 122.74811  7.863043
## 3 100_GWI_202311_WC_LysA_10cm 1566.6521 12820.329 48.866254 361.64539 23.166360
## 4 101_GWI_202311_WC_LysA_20cm 1317.4105 12162.407 41.092031 343.08623 21.977495
## 5 102_GWI_202311_WC_LysA_45cm 1199.1944 10824.812 37.404691 305.35435 19.560461
## 6 103_GWI_202311_WC_LysB_10cm 1592.8158 13596.121 49.682339 383.52951 24.568217
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202311.csv")          #C
```

END