# Dionex_COMPASS_August2022

Stephanie J. Wilson

2023-01-23

## 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_202208_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  4.297   0.0001     0.09   0.0001        0
## 5 5 Lab Blank Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 6 6 Lab Blank Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
```

```
## 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   0.0001
## 5 Lab Blank     n.a.
## 6 Lab Blank     n.a.
```

```
## 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   1e-04
## 5 Lab Blank      NA
## 6 Lab Blank      NA
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202208_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  3.22  0.0215   31.11  0.0287     0.1
## 5 5 Lab Blank Unknown  n.a.    n.a.    n.a.    n.a.    n.a.
## 6 6 Lab Blank Unknown 3.217   0.013   79.27  0.0174     0.1
```

```
## 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  0.0215
## 5 Lab Blank    n.a.
## 6 Lab Blank   0.013
```

```
## 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 0.0215
## 5 Lab Blank     NA
## 6 Lab Blank 0.0130
```

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

```
##   Sample_ID   SO4_ppm     Cl_ppm
## 1           2659.6013 14576.8593
## 2           2659.6013  3622.8327
## 3           2659.6013     0.0084
## 4           2659.6013  3759.8455
## 5           2659.6013         NA
## 6            556.7597 14576.8593
```

```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      1_GCrew_202208_UP_LysA_20cm   2.4251 10.7228
## 27     10_GCrew_202208_TR_LysB_20cm   1.9019  6.6606
## 28 10_GCrew_202208_TR_LysB_20cm_dup   2.0380  6.5797
## 29 10_GCrew_202208_TR_LysB_20cm_spk 108.4972 24.4734
## 30     11_GCrew_202208_TR_LysB_45cm   2.7212  3.1153
## 31     12_GCrew_202208_TR_LysC_10cm   5.3369 27.0484
```

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

```
##                           Sample_ID  SO4_ppm  Cl_ppm
## 26      1_GCrew_202208_UP_LysA_20cm   2.4251 10.7228
## 27     10_GCrew_202208_TR_LysB_20cm   1.9019  6.6606
## 28 10_GCrew_202208_TR_LysB_20cm_dup   2.0380  6.5797
## 29 10_GCrew_202208_TR_LysB_20cm_spk 108.4972 24.4734
## 30     11_GCrew_202208_TR_LysB_45cm   2.7212  3.1153
## 31     12_GCrew_202208_TR_LysC_10cm   5.3369 27.0484
```

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

```r
stds <- all_dat[grepl("Standard", all_dat$Sample_ID),]
head(stds)
```

```
##      Sample_ID SO4_ppm Cl_ppm
## 695 Standard 1  0.4842 5.0438
## 696 Standard 1  0.4842 5.0704
## 697 Standard 1  0.4842 5.0358
## 698 Standard 1  0.4842 5.0613
## 699 Standard 1  0.4842 5.0642
## 700 Standard 1  0.4875 5.0438
```

```
stds_chk_S <- stds %>%
    group_by(Sample_ID) %>%
    summarise(mean = mean(SO4_ppm), sd = sd(SO4_ppm))
stds_chk_S$cv <- (stds_chk_S$sd/stds_chk_S$mean)*100
stds_chk_S$flag <-  ifelse(stds_chk_S$cv <2, 'YES', 'NO, rerun')
head(stds_chk_S)
```

```
## # A tibble: 5 x 5
##   Sample_ID    mean      sd    cv flag
##   <fct>       <dbl>   <dbl> <dbl> <chr>
## 1 Standard 1  0.486 0.00183 0.378 YES
## 2 Standard 2  0.940 0.00549 0.584 YES
## 3 Standard 3  1.94  0.00444 0.228 YES
## 4 Standard 4  9.94  0.0232  0.233 YES
## 5 Standard 5 20.0   0.0279  0.139 YES
```

```
stds_chk_Cl <- stds %>%
    group_by(Sample_ID) %>%
    summarise(mean = mean(Cl_ppm), sd = sd(Cl_ppm))
stds_chk_Cl$cv <- (stds_chk_Cl$sd/stds_chk_Cl$mean)*100
stds_chk_Cl$flag <-  ifelse(stds_chk_Cl$cv <2, 'YES', 'NO, rerun')
head(stds_chk_Cl)
```

```
## # A tibble: 5 x 5
##   Sample_ID    mean     sd    cv flag
##   <fct>       <dbl>  <dbl> <dbl> <chr>
## 1 Standard 1   5.06 0.0134 0.264 YES
## 2 Standard 2  10.3  0.0482 0.470 YES
## 3 Standard 3  20.2  0.0366 0.182 YES
## 4 Standard 4  99.3  0.141  0.142 YES
## 5 Standard 5 200.   0.247  0.123 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_GCrew_202208_UP_LysA_20cm   2.4251 10.7228
## 27     10_GCrew_202208_TR_LysB_20cm   1.9019  6.6606
## 28 10_GCrew_202208_TR_LysB_20cm_dup   2.0380  6.5797
## 29 10_GCrew_202208_TR_LysB_20cm_spk 108.4972 24.4734
## 30     11_GCrew_202208_TR_LysB_45cm   2.7212  3.1153
## 31     12_GCrew_202208_TR_LysC_10cm   5.3369 27.0484
```

```
# 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_GCrew_202208_UP_LysA_20cm   2.4251 10.7228 0.07564255 0.3024767
## 27     10_GCrew_202208_TR_LysB_20cm   1.9019  6.6606 0.05932314 0.1878872
## 28 10_GCrew_202208_TR_LysB_20cm_dup   2.0380  6.5797 0.06356831 0.1856051
## 29 10_GCrew_202208_TR_LysB_20cm_spk 108.4972 24.4734 3.38419214 0.6903639
## 30     11_GCrew_202208_TR_LysB_45cm   2.7212  3.1153 0.08487835 0.0878787
## 31     12_GCrew_202208_TR_LysC_10cm   5.3369 27.0484 0.16646600 0.7630014
##       salinity
## 26 0.019402100
## 27 0.012061704
## 28 0.011915518
## 29 0.044249434
## 30 0.005655347
## 31 0.048902459
```

## Pull out dups and check with percent difference

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

```
##                            Sample_ID  SO4_ppm  Cl_ppm     SO4_mM     Cl_mM
## 26      1_GCrew_202208_UP_LysA_20cm   2.4251 10.7228 0.07564255 0.3024767
## 27     10_GCrew_202208_TR_LysB_20cm   1.9019  6.6606 0.05932314 0.1878872
## 28 10_GCrew_202208_TR_LysB_20cm_dup   2.0380  6.5797 0.06356831 0.1856051
## 29 10_GCrew_202208_TR_LysB_20cm_spk 108.4972 24.4734 3.38419214 0.6903639
## 30     11_GCrew_202208_TR_LysB_45cm   2.7212  3.1153 0.08487835 0.0878787
## 31     12_GCrew_202208_TR_LysC_10cm   5.3369 27.0484 0.16646600 0.7630014
##       salinity
## 26 0.019402100
## 27 0.012061704
## 28 0.011915518
## 29 0.044249434
## 30 0.005655347
## 31 0.048902459
```

```r
#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_GCrew_202208_TR_LysB_20cm_dup    2.0380    6.5797  0.06356831   0.1856051
## 2 20_GCrew_202208_WC_SipB_45cm_dup  405.2386 3967.4580 12.64000624 111.9170099
## 3   30_GWI_202208_UP_LysA_10cm_dup  508.7221 3659.4661 15.86781347 103.2289450
## 4   40_GWI_202208_TR_LysA_45cm_dup  735.8219 8149.1088 22.95140050 229.8761298
## 5   50_GWI_202208_WC_SipB_10cm_dup 1310.7931 9277.3775 40.88562383 261.7031735
## 6    60_GWI_202208_RHZ_TR_SF_2_dup  613.3640 4550.2601 19.13175296 128.3571255
##      salinity
## 1  0.01191552
## 2  7.16922261
## 3  6.61268124
## 4 14.72546560
## 5 16.76424714
## 6  8.22234600
```

```r
#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_GCrew_202208_UP_LysA_20cm  2.4251 10.7228 0.07564255 0.3024767 0.019402100
## 2 10_GCrew_202208_TR_LysB_20cm  1.9019  6.6606 0.05932314 0.1878872 0.012061704
## 3 11_GCrew_202208_TR_LysB_45cm  2.7212  3.1153 0.08487835 0.0878787 0.005655347
## 4 12_GCrew_202208_TR_LysC_10cm  5.3369 27.0484 0.16646600 0.7630014 0.048902459
## 5 13_GCrew_202208_TR_LysC_20cm  2.5114 35.9906 0.07833437 1.0152496 0.065061014
## 6 14_GCrew_202208_TR_LysC_45cm 20.8843  4.8610 0.65141298 0.1371227 0.008809827
```

```r
#remove the dup from these IDs so we will have duplicate sample names
dups$Sample_ID<-gsub("_dup","",as.character(dups$Sample_ID))
dups <- dups[ ,-c(2,3)]
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_GCrew_202208_TR_LysB_20cm  0.06356831   0.1856051   0.01191552
## 2 20_GCrew_202208_WC_SipB_45cm 12.64000624 111.9170099   7.16922261
## 3   30_GWI_202208_UP_LysA_10cm 15.86781347 103.2289450   6.61268124
## 4   40_GWI_202208_TR_LysA_45cm 22.95140050 229.8761298  14.72546560
## 5   50_GWI_202208_WC_SipB_10cm 40.88562383 261.7031735  16.76424714
## 6    60_GWI_202208_RHZ_TR_SF_2 19.13175296 128.3571255   8.22234600
```

```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_GCrew_202208_TR_LysB_20cm    1.9019    6.6606  0.05932314   0.1878872
## 2 20_GCrew_202208_WC_SipB_45cm  412.4507 4032.5780 12.86496257 113.7539633
## 3   30_GWI_202208_UP_LysA_10cm  486.9415 3506.9076 15.18844354  98.9254612
## 4   40_GWI_202208_TR_LysA_45cm  658.9266 7268.5465 20.55291953 205.0365726
## 5   50_GWI_202208_WC_SipB_10cm 1397.0580 9892.3338 43.57635683 279.0503188
## 6    60_GWI_202208_RHZ_TR_SF_2  600.6481 4449.2865 18.73512477 125.5087870
##     salinity  SO4_mM_dup   Cl_mM_dup salinity_dup
## 1  0.0120617  0.06356831   0.1856051   0.01191552
## 2  7.2868944 12.64000624 111.9170099   7.16922261
## 3  6.3370080 15.86781347 103.2289450   6.61268124
## 4 13.1342895 22.95140050 229.8761298  14.72546560
## 5 17.8754732 40.88562383 261.7031735  16.76424714
## 6  8.0398867 19.13175296 128.3571255   8.22234600
```

```
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)
```

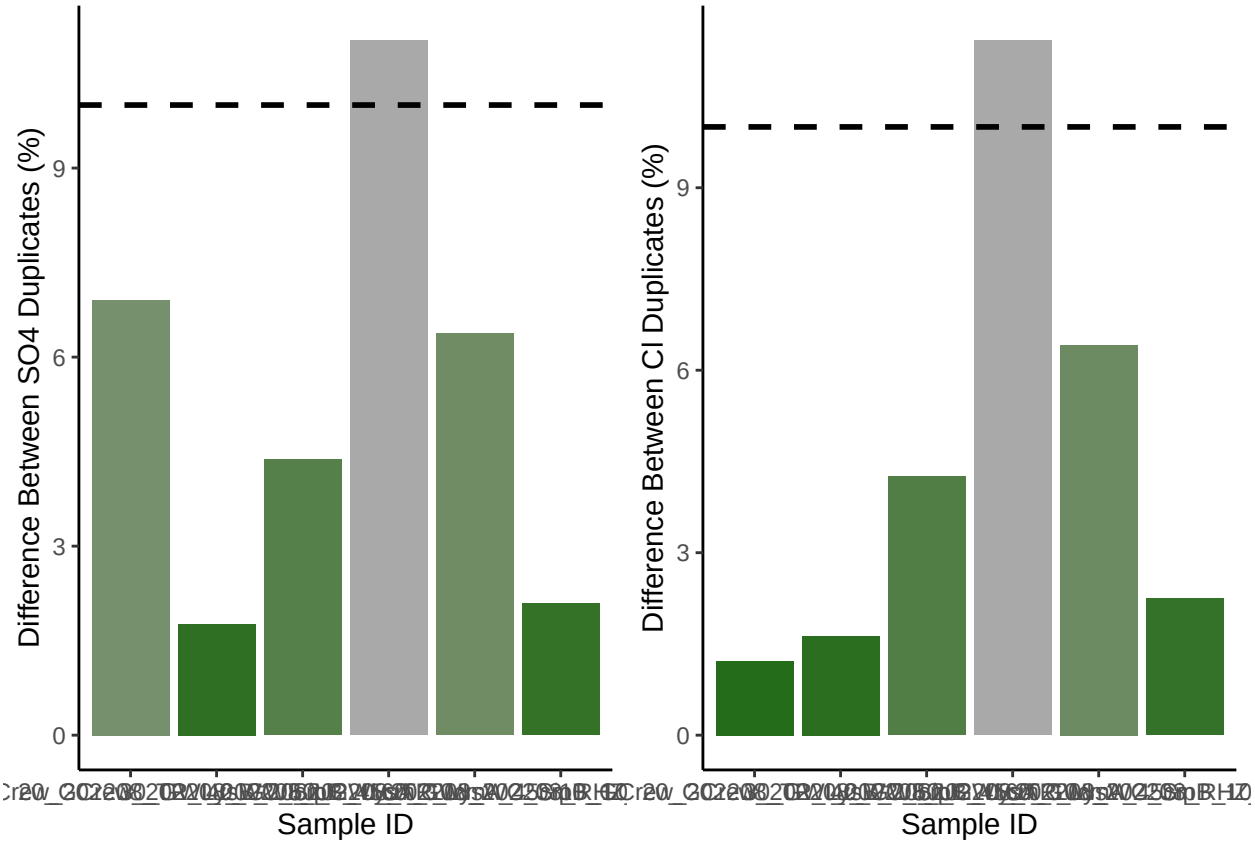
```
##                      Sample_ID   SO4_ppm    Cl_ppm      SO4_mM       Cl_mM
## 1 10_GCrew_202208_TR_LysB_20cm    1.9019    6.6606  0.05932314   0.1878872
## 2 20_GCrew_202208_WC_SipB_45cm  412.4507 4032.5780 12.86496257 113.7539633
## 3   30_GWI_202208_UP_LysA_10cm  486.9415 3506.9076 15.18844354  98.9254612
## 4   40_GWI_202208_TR_LysA_45cm  658.9266 7268.5465 20.55291953 205.0365726
## 5   50_GWI_202208_WC_SipB_10cm 1397.0580 9892.3338 43.57635683 279.0503188
## 6    60_GWI_202208_RHZ_TR_SF_2  600.6481 4449.2865 18.73512477 125.5087870
##     salinity  SO4_mM_dup   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.0120617  0.06356831   0.1856051   0.01191552     6.908805           YES
## 2  7.2868944 12.64000624 111.9170099   7.16922261     1.764020           YES
## 3  6.3370080 15.86781347 103.2289450   6.61268124     4.375092           YES
## 4 13.1342895 22.95140050 229.8761298  14.72546560    11.026404     NO, rerun
## 5 17.8754732 40.88562383 261.7031735  16.76424714     6.371466           YES
## 6  8.0398867 19.13175296 128.3571255   8.22234600     2.094856           YES
##   Cl_dups_chk Cl_dups_flag
## 1    1.222027          YES
## 2    1.627993          YES
## 3    4.257621          YES
## 4   11.422778    NO, rerun
## 5    6.415916          YES
## 6    2.243971          YES
```

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

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
```

```
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

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

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




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_dups <- QAdups %>%
  group_by(SO4_dups_flag) %>%
  summarise(S_no_rows = length(SO4_dups_flag))
Perc_dups1 <- QAdups %>%
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(Cl_dups_flag))
colnames(Perc_dups) <- c("Flag", "S_no_rows")
colnames(Perc_dups1) <- c("Flag", "Cl_no_rows")
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)
```

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

```
##        Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1 NO, rerun         1 NO, rerun          1     6  16.66667   16.66667
## 2       YES         5       YES          5     6  83.33333   83.33333
```

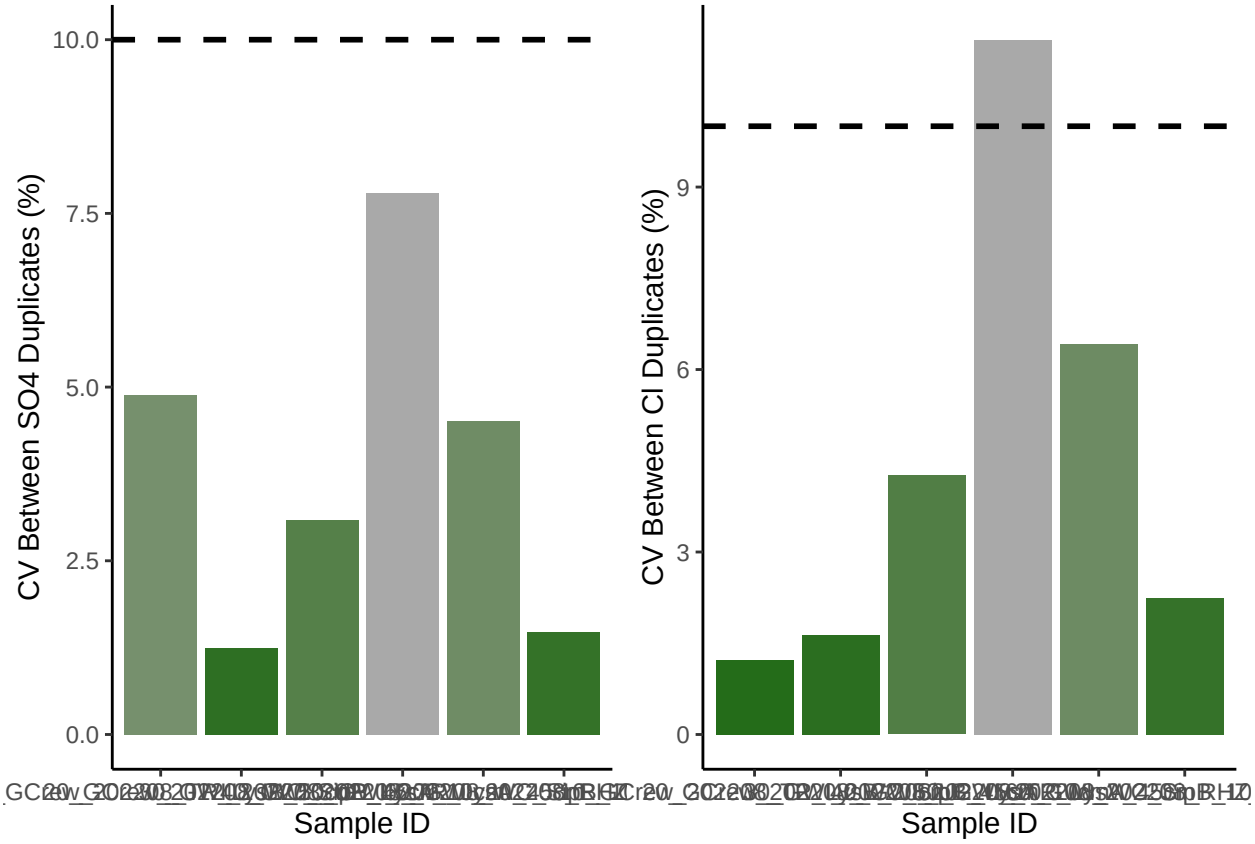
### Pull out dups and check with cv

```
#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_GCrew_202208_TR_LysB_20cm    1.9019    6.6606  0.05932314   0.1878872
## 2 20_GCrew_202208_WC_SipB_45cm  412.4507 4032.5780 12.86496257 113.7539633
## 3   30_GWI_202208_UP_LysA_10cm  486.9415 3506.9076 15.18844354  98.9254612
## 4   40_GWI_202208_TR_LysA_45cm  658.9266 7268.5465 20.55291953 205.0365726
## 5   50_GWI_202208_WC_SipB_10cm 1397.0580 9892.3338 43.57635683 279.0503188
## 6    60_GWI_202208_RHZ_TR_SF_2  600.6481 4449.2865 18.73512477 125.5087870
##     salinity  SO4_mM_dup   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.0120617  0.06356831   0.1856051   0.01191552     6.908805           YES
## 2  7.2868944 12.64000624 111.9170099   7.16922261     1.764020           YES
## 3  6.3370080 15.86781347 103.2289450   6.61268124     4.375092           YES
## 4 13.1342895 22.95140050 229.8761298  14.72546560    11.026404     NO, rerun
## 5 17.8754732 40.88562383 261.7031735  16.76424714     6.371466           YES
## 6  8.0398867 19.13175296 128.3571255   8.22234600     2.094856           YES
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1    1.222027          YES    4.885263              YES
## 2    1.627993          YES    1.247350              YES
## 3    4.257621          YES    3.093657              YES
## 4   11.422778    NO, rerun    7.796845              YES
## 5    6.415916          YES    4.505306              YES
## 6    2.243971          YES    1.481287              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 NO, rerun         1 NO, rerun          1     6  16.66667   16.66667
## 2       YES         5       YES          5     6  83.33333   83.33333
```

## 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_GCrew_202208_UP_LysA_20cm   2.4251 10.7228 0.07564255 0.3024767
## 27     10_GCrew_202208_TR_LysB_20cm   1.9019  6.6606 0.05932314 0.1878872
## 28 10_GCrew_202208_TR_LysB_20cm_dup   2.0380  6.5797 0.06356831 0.1856051
## 29 10_GCrew_202208_TR_LysB_20cm_spk 108.4972 24.4734 3.38419214 0.6903639
## 30     11_GCrew_202208_TR_LysB_45cm   2.7212  3.1153 0.08487835 0.0878787
## 31     12_GCrew_202208_TR_LysC_10cm   5.3369 27.0484 0.16646600 0.7630014
##       salinity
## 26 0.019402100
## 27 0.012061704
## 28 0.011915518
## 29 0.044249434
## 30 0.005655347
## 31 0.048902459
```

```
#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 10_GCrew_202208_TR_LysB_20cm_spk  108.4972   24.4734  3.384192   0.6903639
## 2 20_GCrew_202208_WC_SipB_45cm_spk  620.2658 3954.0032 19.347031 111.5374669
## 3   30_GWI_202208_UP_LysA_10cm_spk  702.4759 3526.1084 21.911288  99.4670917
## 4   40_GWI_202208_TR_LysA_45cm_spk  855.0443 7143.8038 26.670128 201.5177377
## 5   50_GWI_202208_WC_SipB_10cm_spk 1854.8338 9964.9352 57.855078 281.0983131
## 6    60_GWI_202208_RHZ_TR_SF_2_spk  827.0714 4549.3304 25.797611 128.3308999
##      salinity
## 1  0.04424943
## 2  7.14490978
## 3  6.37170388
## 4 12.90887947
## 5 18.00666391
## 6  8.22066603
```

```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 10_GCrew_202208_TR_LysB_20cm   3.384192
## 2 20_GCrew_202208_WC_SipB_45cm  19.347031
## 3   30_GWI_202208_UP_LysA_10cm  21.911288
## 4   40_GWI_202208_TR_LysA_45cm  26.670128
## 5   50_GWI_202208_WC_SipB_10cm  57.855078
## 6    60_GWI_202208_RHZ_TR_SF_2  25.797611
```

```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 10_GCrew_202208_TR_LysB_20cm    1.9019    6.6606  0.05932314   0.1878872
## 2 20_GCrew_202208_WC_SipB_45cm  412.4507 4032.5780 12.86496257 113.7539633
## 3   30_GWI_202208_UP_LysA_10cm  486.9415 3506.9076 15.18844354  98.9254612
## 4   40_GWI_202208_TR_LysA_45cm  658.9266 7268.5465 20.55291953 205.0365726
## 5   50_GWI_202208_WC_SipB_10cm 1397.0580 9892.3338 43.57635683 279.0503188
## 6    60_GWI_202208_RHZ_TR_SF_2  600.6481 4449.2865 18.73512477 125.5087870
##     salinity SO4_mM_spk
## 1  0.0120617   3.384192
## 2  7.2868944  19.347031
## 3  6.3370080  21.911288
## 4 13.1342895  26.670128
## 5 17.8754732  57.855078
## 6  8.0398867  25.797611
```

```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 10_GCrew_202208_TR_LysB_20cm    1.9019    6.6606  0.05932314   0.1878872
## 2 20_GCrew_202208_WC_SipB_45cm  412.4507 4032.5780 12.86496257 113.7539633
## 3   30_GWI_202208_UP_LysA_10cm  486.9415 3506.9076 15.18844354  98.9254612
## 4   40_GWI_202208_TR_LysA_45cm  658.9266 7268.5465 20.55291953 205.0365726
## 5   50_GWI_202208_WC_SipB_10cm 1397.0580 9892.3338 43.57635683 279.0503188
## 6    60_GWI_202208_RHZ_TR_SF_2  600.6481 4449.2865 18.73512477 125.5087870
##     salinity SO4_mM_spk SO4_spk_Conc
## 1  0.0120617   3.384192 7.797879e-05
## 2  7.2868944  19.347031 7.797879e-05
## 3  6.3370080  21.911288 7.797879e-05
```

```
## 4 13.1342895  26.670128 7.797879e-05
## 5 17.8754732  57.855078 7.797879e-05
## 6  8.0398867  25.797611 7.797879e-05
```

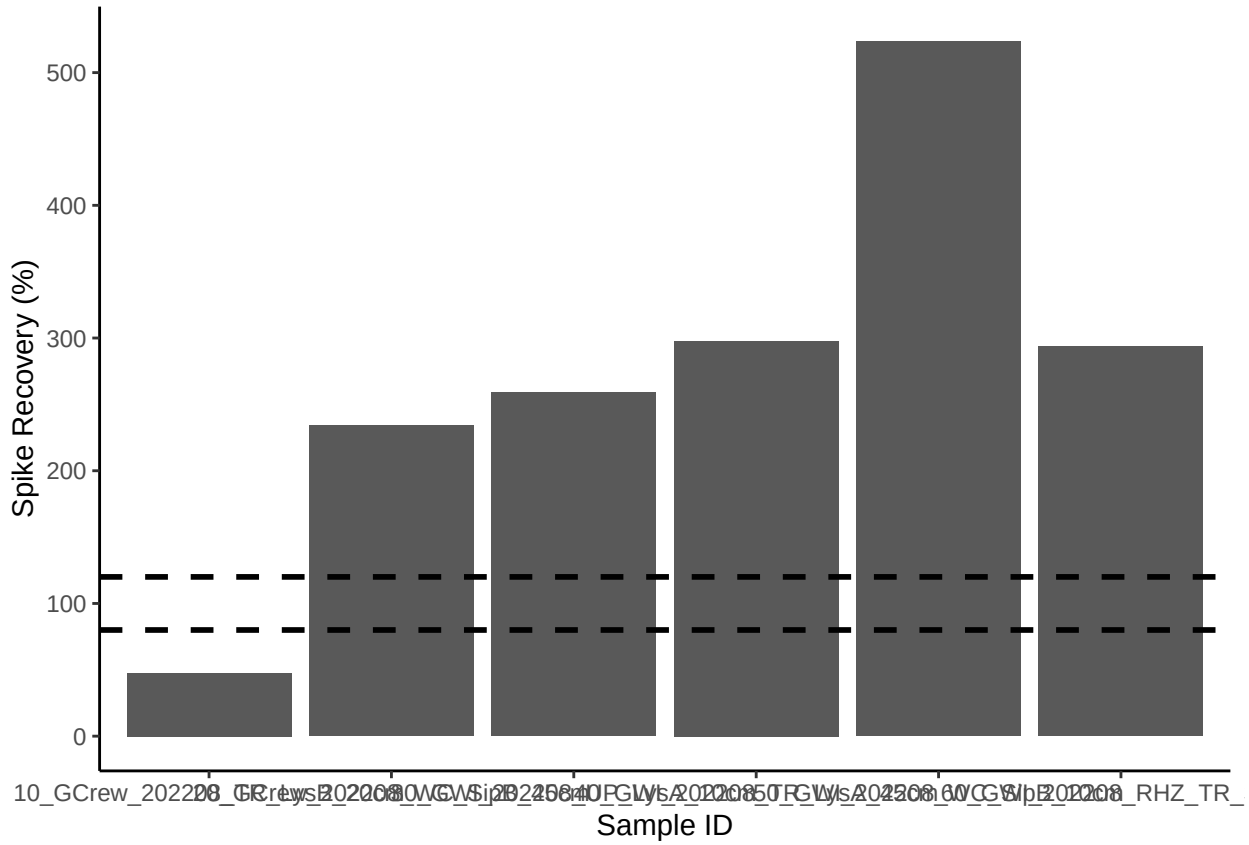
```
#need to determine dilution factors and initial amount of sample added
#if your samples are all the same dilution just use the first line
#for Steph / COMPASS this depends on the site so...
QAspks$Dilution <- 1
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 200, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$Dilution)

#Set Sample volumes in uL
QAspks$SampleVol <- 1
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 1462, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$SampleVol)

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

```
##                     Sample_ID   SO4_ppm    Cl_ppm      SO4_mM       Cl_mM
## 1 10_GCrew_202208_TR_LysB_20cm    1.9019    6.6606  0.05932314   0.1878872
## 2 20_GCrew_202208_WC_SipB_45cm  412.4507 4032.5780 12.86496257 113.7539633
## 3   30_GWI_202208_UP_LysA_10cm  486.9415 3506.9076 15.18844354  98.9254612
## 4   40_GWI_202208_TR_LysA_45cm  658.9266 7268.5465 20.55291953 205.0365726
## 5   50_GWI_202208_WC_SipB_10cm 1397.0580 9892.3338 43.57635683 279.0503188
## 6    60_GWI_202208_RHZ_TR_SF_2  600.6481 4449.2865 18.73512477 125.5087870
##     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1  0.0120617   3.384192 7.797879e-05        1     1e-06
## 2  7.2868944  19.347031 7.797879e-05        1     1e-06
## 3  6.3370080  21.911288 7.797879e-05        1     1e-06
## 4 13.1342895  26.670128 7.797879e-05        1     1e-06
## 5 17.8754732  57.855078 7.797879e-05        1     1e-06
## 6  8.0398867  25.797611 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 10_GCrew_202208_TR_LysB_20cm    1.9019    6.6606  0.05932314   0.1878872
## 2 20_GCrew_202208_WC_SipB_45cm  412.4507 4032.5780 12.86496257 113.7539633
## 3   30_GWI_202208_UP_LysA_10cm  486.9415 3506.9076 15.18844354  98.9254612
## 4   40_GWI_202208_TR_LysA_45cm  658.9266 7268.5465 20.55291953 205.0365726
## 5   50_GWI_202208_WC_SipB_10cm 1397.0580 9892.3338 43.57635683 279.0503188
## 6    60_GWI_202208_RHZ_TR_SF_2  600.6481 4449.2865 18.73512477 125.5087870
##     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1  0.0120617   3.384192 7.797879e-05        1     1e-06     5.932314e-08
## 2  7.2868944  19.347031 7.797879e-05        1     1e-06     1.286496e-05
## 3  6.3370080  21.911288 7.797879e-05        1     1e-06     1.518844e-05
## 4 13.1342895  26.670128 7.797879e-05        1     1e-06     2.055292e-05
## 5 17.8754732  57.855078 7.797879e-05        1     1e-06     4.357636e-05
## 6  8.0398867  25.797611 7.797879e-05        1     1e-06     1.873512e-05
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   3.722611e-05    7.803811e-05     47.70248     NO, rerun
## 2   2.128173e-04    9.084375e-05    234.26744     NO, rerun
## 3   2.410242e-04    9.316723e-05    258.70058     NO, rerun
## 4   2.933714e-04    9.853171e-05    297.74314     NO, rerun
## 5   6.364059e-04    1.215551e-04    523.55320     NO, rerun
## 6   2.837737e-04    9.671391e-05    293.41561     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: 1 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 NO, rerun           6     6     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", "GCrew", "202208", "UP", "LysA", "20cm"), c("10", :
## number of columns of result is not a multiple of vector length (arg 1)
```

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

```
##   Analysis_No  Date   Site Zone Replicate Depth Tree
## 1           1 GCrew 202208   UP      LysA  20cm    1
## 2          10 GCrew 202208   TR      LysB  20cm   10
## 3          11 GCrew 202208   TR      LysB  45cm   11
## 4          12 GCrew 202208   TR      LysC  10cm   12
## 5          13 GCrew 202208   TR      LysC  20cm   13
## 6          14 GCrew 202208   TR      LysC  45cm   14
```

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

```
##   Analysis_No  Date   Site Zone Replicate Depth Tree
## 1           1 GCrew 202208   UP      LysA  20cm    1
## 2          10 GCrew 202208   TR      LysB  20cm   10
## 3          11 GCrew 202208   TR      LysB  45cm   11
## 4          12 GCrew 202208   TR      LysC  10cm   12
## 5          13 GCrew 202208   TR      LysC  20cm   13
## 6          14 GCrew 202208   TR      LysC  45cm   14
##                      Sample_ID SO4_ppm  Cl_ppm     SO4_mM     Cl_mM    salinity
## 1  1_GCrew_202208_UP_LysA_20cm  2.4251 10.7228 0.07564255 0.3024767 0.019402100
## 2 10_GCrew_202208_TR_LysB_20cm  1.9019  6.6606 0.05932314 0.1878872 0.012061704
## 3 11_GCrew_202208_TR_LysB_45cm  2.7212  3.1153 0.08487835 0.0878787 0.005655347
## 4 12_GCrew_202208_TR_LysC_10cm  5.3369 27.0484 0.16646600 0.7630014 0.048902459
## 5 13_GCrew_202208_TR_LysC_20cm  2.5114 35.9906 0.07833437 1.0152496 0.065061014
## 6 14_GCrew_202208_TR_LysC_45cm 20.8843  4.8610 0.65141298 0.1371227 0.008809827
```

### Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_So4_202208.csv")    #Change f
```

### END