Dionex_COMPASS_June2023

Stephanie J. Wilson

2023-06-23

Daily Set up

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

```
#setwd("S:/Biogeochemistry/People/Wilson (Steph)/Data/Dionex/Raw Data Files")

# SULFATE DATA:

## Read in raw data file from Dionex - copied and saved as a txt
Sdat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202304_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 4.680   0.4466     6.56   0.6608     2.79
## 6 6 Standard 2 Calibration Standard 4.677   0.9155     7.16   1.3545     5.57
##   IC.SO4.5
## 1     n.a.
## 2     n.a.
## 3     n.a.
## 4     n.a.
## 5       M
## 6       M
```

```
## 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  0.4466
## 6 Standard 2  0.9155
```

```
## 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 Standard 1  0.4466
## 6 Standard 2  0.9155
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202304_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 3.710   6.0596   92.44   9.3112   42.63
## 6 6 Standard 2 Calibration Standard 3.710  11.3033   91.83  17.3688   78.65
##   IC.Cl.5
## 1    n.a.
## 2    n.a.
## 3    n.a.
## 4    n.a.
## 5      M
## 6      M
```

```
## 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  6.0596
## 6 Standard 2 11.3033
```

```
## 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 Standard 1  6.0596
## 6 Standard 2 11.3033
```

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

```
##   Sample_ID   SO4_ppm     Cl_ppm
## 1           2556.7130 21806.2530
## 2           2556.7130  6022.9406
## 3           2556.7130     0.0017
## 4           2556.7130  5600.4666
## 5           2556.7130         NA
## 6            597.3037 21806.2530
```

```
## 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_GCW_202304_UP_LysA_20cm   8.2441    0.0017
## 27     10_GCW_202304_TR_LysB_20cm  24.9499   73.6636
## 28 10_GCW_202304_TR_LysB_20cm_dup  25.0390   73.6625
## 29        100_MSM_202304_PPR_TR_4 741.9526 5993.3481
## 30        100_MSM_202304_PPR_TR_4 741.9526        NA
## 31        100_MSM_202304_PPR_TR_4 741.9526        NA
```

```
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_GCW_202304_UP_LysA_20cm   8.2441    0.0017
## 27     10_GCW_202304_TR_LysB_20cm  24.9499   73.6636
## 28 10_GCW_202304_TR_LysB_20cm_dup  25.0390   73.6625
## 29        100_MSM_202304_PPR_TR_4 741.9526 5993.3481
## 30        100_MSM_202304_PPR_TR_4 741.9526    0.0000
## 31        100_MSM_202304_PPR_TR_4 741.9526    0.0000
```

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

```
stds <- all_dat[grepl("Standard", all_dat$Sample_ID),]
stds <- stds[-c(3, 8, 13, 18, 23),]   #this is if you need to remove one for any reason
head(stds)
```

```
##      Sample_ID SO4_ppm Cl_ppm
## 638 Standard 1  0.6281 6.4219
## 639 Standard 1  0.6281 6.0596
## 641 Standard 1  0.6281 5.9994
## 642 Standard 1  0.6281 6.2815
## 643 Standard 1  0.4466 6.4219
## 644 Standard 1  0.4466 6.0596
```

```
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  5.40 10.1   186.   NO, rerun
## 2 Standard 2  1.01  0.137  13.6  NO, rerun
## 3 Standard 3  2.20  0.284  12.9  NO, rerun
## 4 Standard 4 10.7   1.25   11.7  NO, rerun
## 5 Standard 5 21.7   1.91    8.80 NO, rerun
```

```
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   6.19 0.174  2.82 YES
## 2 Standard 2  11.7  0.231  1.99 YES
## 3 Standard 3  22.9  0.533  2.33 YES
## 4 Standard 4 103.   2.59   2.51 YES
## 5 Standard 5 206.   4.47   2.17 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_GCW_202304_UP_LysA_20cm   8.2441    0.0017
## 27     10_GCW_202304_TR_LysB_20cm  24.9499   73.6636
## 28 10_GCW_202304_TR_LysB_20cm_dup  25.0390   73.6625
## 29        100_MSM_202304_PPR_TR_4 741.9526 5993.3481
## 30        100_MSM_202304_PPR_TR_4 741.9526    0.0000
## 31        100_MSM_202304_PPR_TR_4 741.9526    0.0000
```

```
# 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_GCW_202304_UP_LysA_20cm   8.2441    0.0017  0.2571460 4.795487e-05
## 27     10_GCW_202304_TR_LysB_20cm  24.9499   73.6636  0.7782252 2.077958e+00
## 28 10_GCW_202304_TR_LysB_20cm_dup  25.0390   73.6625  0.7810044 2.077927e+00
## 29        100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 1.690648e+02
## 30        100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
## 31        100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
##        salinity
## 26 2.907190e-05
## 27 1.331361e-01
## 28 1.331341e-01
## 29 1.083001e+01
## 30 2.600000e-05
## 31 2.600000e-05
```

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_GCW_202304_UP_LysA_20cm   8.2441    0.0017  0.2571460 4.795487e-05
## 27     10_GCW_202304_TR_LysB_20cm  24.9499   73.6636  0.7782252 2.077958e+00
## 28 10_GCW_202304_TR_LysB_20cm_dup  25.0390   73.6625  0.7810044 2.077927e+00
## 29        100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 1.690648e+02
## 30        100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
## 31        100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
##        salinity
## 26 2.907190e-05
## 27 1.331361e-01
## 28 1.331341e-01
## 29 1.083001e+01
## 30 2.600000e-05
## 31 2.600000e-05
```

```
#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_GCW_202304_TR_LysB_20cm_dup  25.0390   73.6625  0.7810044   2.077927
## 2    100_MSM_202304_PPR_TR_4_dup 739.1719 6000.1002 23.0558921 169.255295
## 3    100_MSM_202304_PPR_TR_4_dup 739.1719    0.0000 23.0558921   0.000000
## 4    100_MSM_202304_PPR_TR_4_dup 739.1719    0.0000 23.0558921   0.000000
## 5    100_MSM_202304_PPR_TR_4_dup  11.4443 6000.1002  0.3569651 169.255295
## 6    100_MSM_202304_PPR_TR_4_dup  11.4443    0.0000  0.3569651   0.000000
##     salinity
## 1  0.1331341
## 2 10.8422071
## 3  0.0000260
## 4  0.0000260
## 5 10.8422071
## 6  0.0000260
```

```
#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
## 1  1_GCW_202304_UP_LysA_20cm   8.2441    0.0017  0.2571460 4.795487e-05
## 2 10_GCW_202304_TR_LysB_20cm  24.9499   73.6636  0.7782252 2.077958e+00
## 3    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 1.690648e+02
## 4    100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
```

```
## 5    100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
## 6    100_MSM_202304_PPR_TR_4   9.2977 5993.3481  0.2900094 1.690648e+02
##       salinity
## 1 2.907190e-05
## 2 1.331361e-01
## 3 1.083001e+01
## 4 2.600000e-05
## 5 2.600000e-05
## 6 1.083001e+01
```

```
#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_GCW_202304_TR_LysB_20cm  0.7810044   2.077927    0.1331341
## 2    100_MSM_202304_PPR_TR_4 23.0558921 169.255295   10.8422071
## 3    100_MSM_202304_PPR_TR_4 23.0558921   0.000000    0.0000260
## 4    100_MSM_202304_PPR_TR_4 23.0558921   0.000000    0.0000260
## 5    100_MSM_202304_PPR_TR_4  0.3569651 169.255295   10.8422071
## 6    100_MSM_202304_PPR_TR_4  0.3569651   0.000000    0.0000260
```

```
#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_GCW_202304_TR_LysB_20cm  24.9499   73.6636  0.7782252   2.077958
## 2    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 3    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 4    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 5    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 6    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
##     salinity SO4_mM_dup  Cl_mM_dup salinity_dup
## 1  0.1331361  0.7810044   2.077927    0.1331341
## 2 10.8300060 23.0558921 169.255295   10.8422071
## 3 10.8300060 23.0558921   0.000000    0.0000260
## 4 10.8300060 23.0558921   0.000000    0.0000260
## 5 10.8300060  0.3569651 169.255295   10.8422071
## 6 10.8300060  0.3569651   0.000000    0.0000260
```

```
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)
```

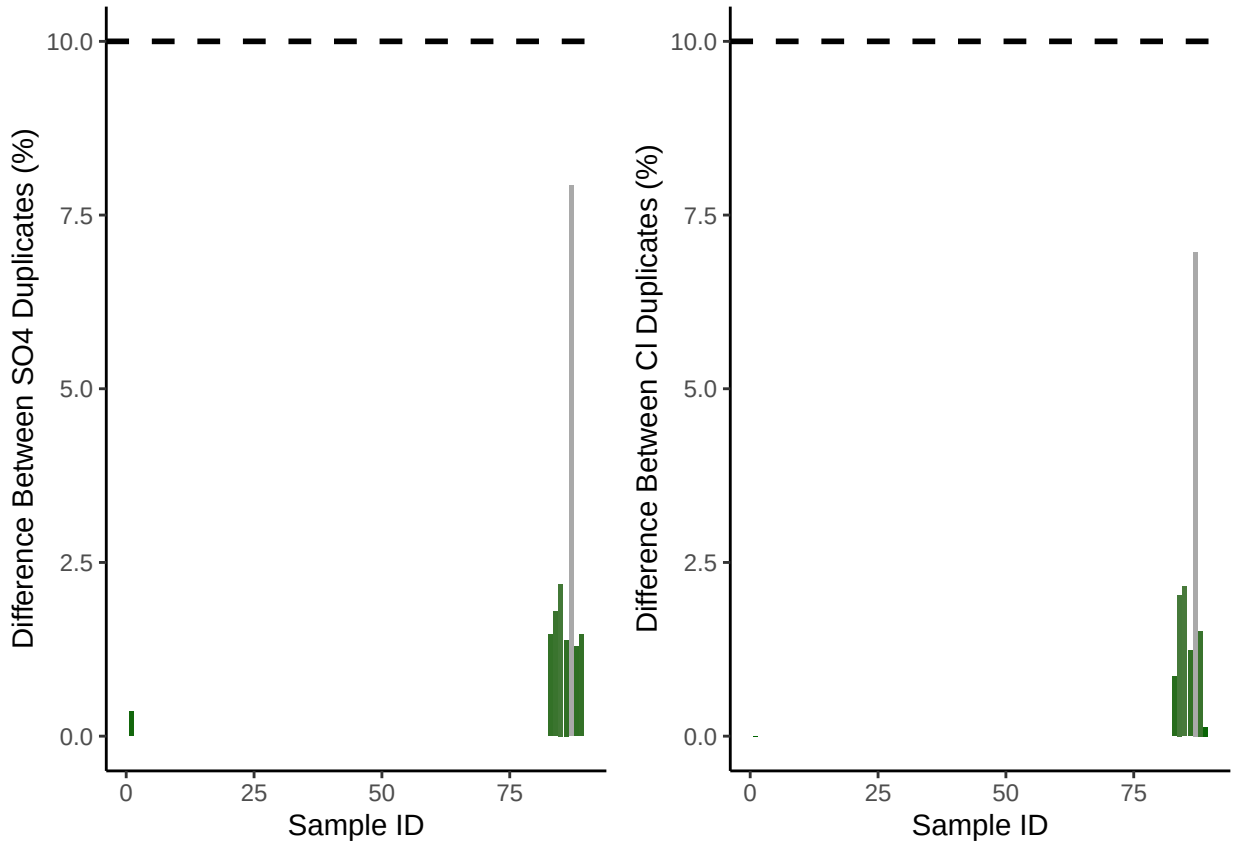
```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM      Cl_mM
```

```
## 1 10_GCW_202304_TR_LysB_20cm  24.9499   73.6636  0.7782252   2.077958
## 2    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 3    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 4    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 5    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
## 6    100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 169.064827
##     salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.1331361  0.7810044   2.077927    0.1331341    0.3564791           YES
## 2 10.8300060 23.0558921 169.255295   10.8422071    0.3754850           YES
## 3 10.8300060 23.0558921   0.000000    0.0000260    0.3754850           YES
## 4 10.8300060 23.0558921   0.000000    0.0000260    0.3754850           YES
## 5 10.8300060  0.3569651 169.255295   10.8422071  193.9238932     NO, rerun
## 6 10.8300060  0.3569651   0.000000    0.0000260  193.9238932     NO, rerun
##    Cl_dups_chk Cl_dups_flag
## 1 1.493286e-03          YES
## 2 1.125965e-01          YES
## 3 2.000000e+02    NO, rerun
## 4 2.000000e+02    NO, rerun
## 5 1.125965e-01          YES
## 6 2.000000e+02    NO, rerun
```

```
QAdups <- QAdups %>%
  mutate(row_number = row_number())  %>%
  filter(!Sample_ID == "100_MSM_202304_PPR_TR_4")

#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 = row_number, 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 = row_number, 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_dups1 <- Perc_dups1[-c(3),]
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)

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

```
##   Flag S_no_rows Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES         8  YES          8     8       100        100
```

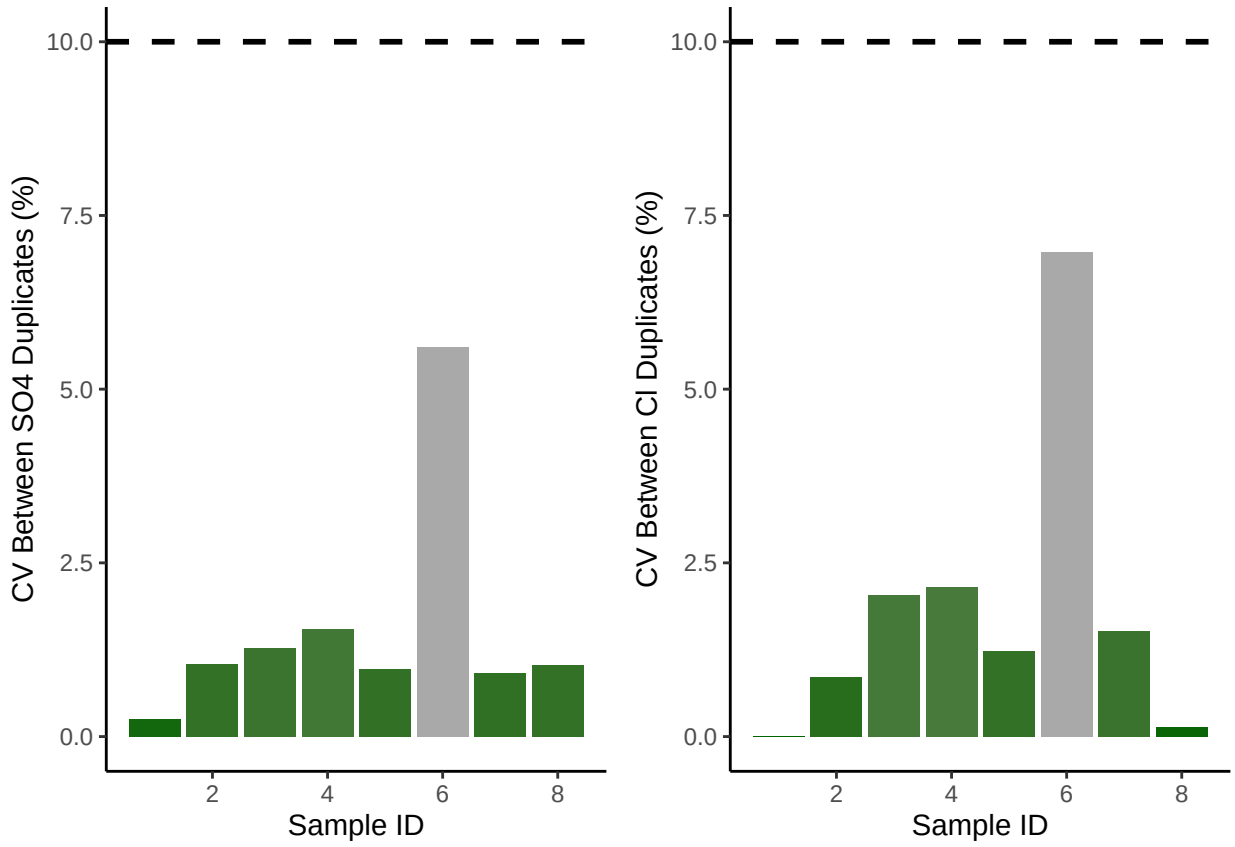
Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

```
##                    Sample_ID   SO4_ppm     Cl_ppm     SO4_mM      Cl_mM
## 1 10_GCW_202304_TR_LysB_20cm   24.9499    73.6636  0.7782252   2.077958
## 2    110_MSM_202304_PPR_WC_2  728.5031 13337.4514 22.7231160 376.232762
## 3 20_GCW_202304_WC_LysB_45cm  440.9736  7417.9642 13.7546351 209.251458
## 4 30_MSM_202304_UP_LysB_45cm  222.6503  2656.9852  6.9448004  74.950217
## 5 40_MSM_202304_TR_LysC_10cm  579.3436  5440.5188 18.0706051 153.470206
## 6 60_GWI_202304_WC_LysB_20cm 2050.6464 17258.5113 63.9627698 486.840939
##     salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.1331361  0.7810044   2.077927    0.1331341    0.3564791           YES
## 2 24.1008007 22.3924579 373.019901   23.8949908    1.4658270           YES
## 3 13.4042873 13.5096163 205.038812   13.1344330    1.7973625           YES
## 4  4.8011983  6.7942452  73.354116    4.6989549    2.1916400           YES
## 5  9.8310435 18.3222333 151.588496    9.7105046    1.3828446           YES
## 6 31.1861559 59.0870056 454.059289   29.0862241    7.9248648           YES
##   Cl_dups_chk Cl_dups_flag row_number SO4_dups_cv SO4_dups_cv_flag
## 1 0.001493286          YES          1   0.2520688              YES
## 2 0.857617337          YES         83   1.0364962              YES
## 3 2.033668796          YES         84   1.2709272              YES
## 4 2.152467862          YES         85   1.5497235              YES
## 5 1.233670395          YES         86   0.9778188              YES
## 6 6.968145871          YES         87   5.6037257              YES
```

```r
QAdups <- QAdups %>%
  mutate(row_number = row_number()) %>%
  filter(!Sample_ID == "100_MSM_202304_PPR_TR_4")

#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 = row_number, 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 = row_number, 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_dups1 <- Perc_dups1[-c(3),]
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)

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

```
##   Flag S_no_rows Flag Cl_no_rows Total S_Percent Cl_Percent
## 1  YES         8  YES          8     8       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_GCW_202304_UP_LysA_20cm   8.2441    0.0017  0.2571460 4.795487e-05
## 27     10_GCW_202304_TR_LysB_20cm  24.9499   73.6636  0.7782252 2.077958e+00
## 28 10_GCW_202304_TR_LysB_20cm_dup  25.0390   73.6625  0.7810044 2.077927e+00
## 29        100_MSM_202304_PPR_TR_4 741.9526 5993.3481 23.1426263 1.690648e+02
## 30        100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
## 31        100_MSM_202304_PPR_TR_4 741.9526    0.0000 23.1426263 0.000000e+00
##        salinity
## 26 2.907190e-05
## 27 1.331361e-01
## 28 1.331341e-01
## 29 1.083001e+01
## 30 2.600000e-05
## 31 2.600000e-05
```

```
#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_MSM_202304_PPR_TR_5_spk    5.7280     0.0000  0.178665   0.000000
## 2    101_MSM_202304_PPR_TR_5_spk    5.7280  6351.8704  0.178665 179.178291
## 3    101_MSM_202304_PPR_TR_5_spk  428.2802     0.0000 13.358709   0.000000
## 4    101_MSM_202304_PPR_TR_5_spk  428.2802  6351.8704 13.358709 179.178291
## 5 11_GCW_202304_TR_LysB_45cm_spk   80.9716    98.3241  2.525627   2.773599
## 6    111_MSM_202304_PPR_WC_3_spk 1822.7867 12621.9929 56.855480 356.050575
##     salinity
## 1  0.0000260
## 2 11.4778558
## 3  0.0000260
## 4 11.4778558
## 5  0.1776976
## 6 22.8079672
```

```
#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_MSM_202304_PPR_TR_5   0.178665
## 2    101_MSM_202304_PPR_TR_5   0.178665
## 3    101_MSM_202304_PPR_TR_5  13.358709
## 4    101_MSM_202304_PPR_TR_5  13.358709
## 5 11_GCW_202304_TR_LysB_45cm   2.525627
## 6    111_MSM_202304_PPR_WC_3  56.855480
```

```
#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 salinity SO4_mM_spk
## 1 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 2 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 3 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 4 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 5 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
## 6 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
```

```
#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 salinity SO4_mM_spk
## 1 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 2 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 3 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 4 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 5 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
## 6 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
##   SO4_spk_Conc
## 1 7.797879e-05
## 2 7.797879e-05
## 3 7.797879e-05
## 4 7.797879e-05
## 5 7.797879e-05
## 6 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)
```

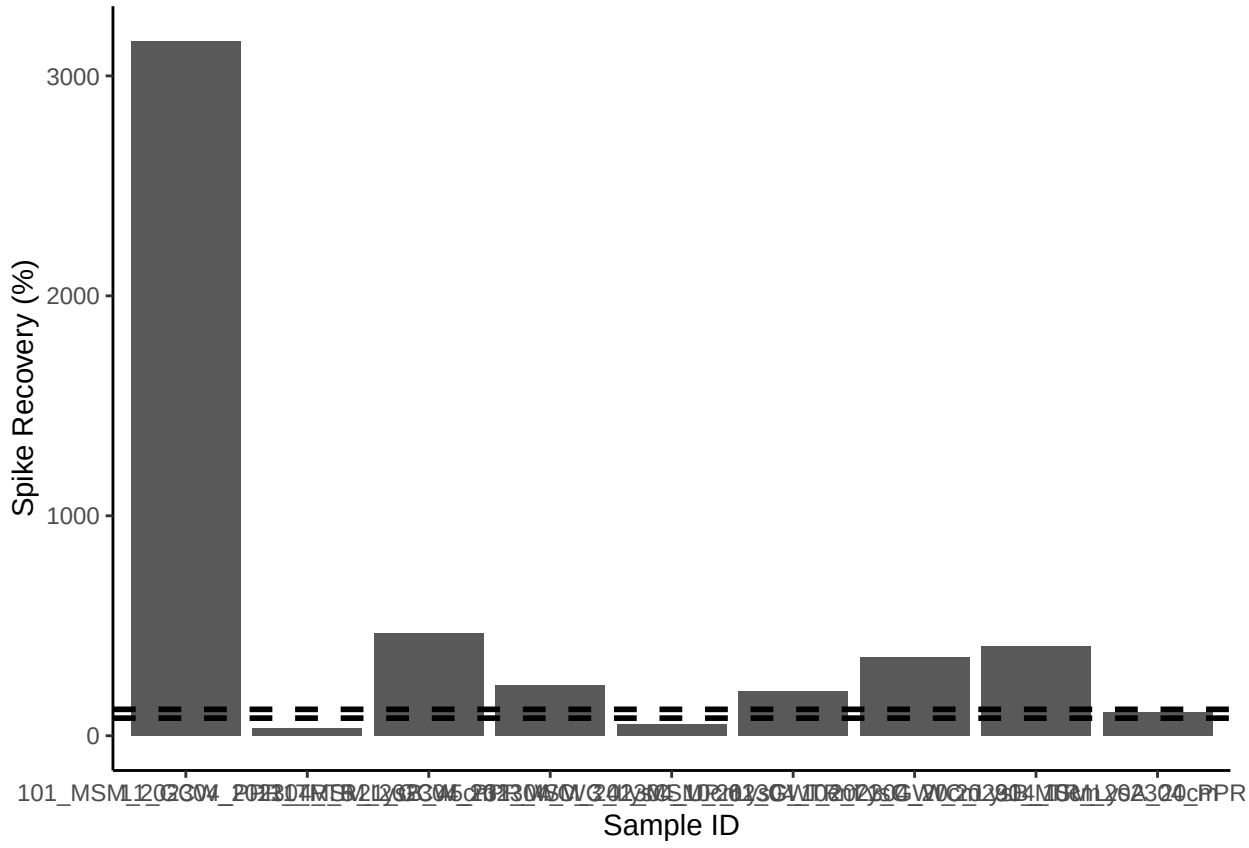
```
##                 Sample_ID  SO4_ppm Cl_ppm   SO4_mM Cl_mM salinity SO4_mM_spk
## 1 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 2 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 3 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 4 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 5 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
## 6 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
##   SO4_spk_Conc Dilution SampleVol
## 1 7.797879e-05        1     1e-06
## 2 7.797879e-05        1     1e-06
## 3 7.797879e-05        1     1e-06
## 4 7.797879e-05        1     1e-06
## 5 7.797879e-05        1     1e-06
## 6 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 salinity SO4_mM_spk
## 1 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 2 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05   0.178665
## 3 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 4 101_MSM_202304_PPR_TR_5   0.0000      0  0.00000     0  2.6e-05  13.358709
## 5 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
## 6 101_MSM_202304_PPR_TR_5 783.8743      0 24.45023     0  2.6e-05   0.178665
```

```
##   SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd SO4_Total_spkd
## 1 7.797879e-05        1     1e-06     0.000000e+00   1.965315e-06
## 2 7.797879e-05        1     1e-06     0.000000e+00   1.965315e-06
## 3 7.797879e-05        1     1e-06     0.000000e+00   1.469458e-04
## 4 7.797879e-05        1     1e-06     0.000000e+00   1.469458e-04
## 5 7.797879e-05        1     1e-06     2.445023e-05   1.965315e-06
## 6 7.797879e-05        1     1e-06     2.445023e-05   1.965315e-06
##   SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1    7.797879e-05     2.520320     NO, rerun
## 2    7.797879e-05     2.520320     NO, rerun
## 3    7.797879e-05   188.443288     NO, rerun
## 4    7.797879e-05   188.443288     NO, rerun
## 5    1.024290e-04     1.918709     NO, rerun
## 6    1.024290e-04     1.918709     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: 2 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 NO, rerun          43    44   97.7
## 2 YES                 1    44    2.27
```

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", "GCW", "202304", "UP", "LysA", "20cm"), c("10", :
## number of columns of result is not a multiple of vector length (arg 1)
```

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

```
##   Analysis_No Site   Date Zone Replicate Depth  NA
## 1           1  GCW 202304   UP      LysA  20cm   1
## 2          10  GCW 202304   TR      LysB  20cm  10
## 3         100  MSM 202304  PPR        TR     4 100
## 4         100  MSM 202304  PPR        TR     4 100
## 5         100  MSM 202304  PPR        TR     4 100
## 6         100  MSM 202304  PPR        TR     4 100
```

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

```
##   Analysis_No Site   Date Zone Replicate Depth  NA                  Sample_ID
## 1           1  GCW 202304   UP      LysA  20cm   1  1_GCW_202304_UP_LysA_20cm
## 2          10  GCW 202304   TR      LysB  20cm  10 10_GCW_202304_TR_LysB_20cm
## 3         100  MSM 202304  PPR        TR     4 100    100_MSM_202304_PPR_TR_4
## 4         100  MSM 202304  PPR        TR     4 100    100_MSM_202304_PPR_TR_4
## 5         100  MSM 202304  PPR        TR     4 100    100_MSM_202304_PPR_TR_4
## 6         100  MSM 202304  PPR        TR     4 100    100_MSM_202304_PPR_TR_4
##    SO4_ppm    Cl_ppm     SO4_mM        Cl_mM     salinity
## 1   8.2441    0.0017  0.2571460 4.795487e-05 2.907190e-05
## 2  24.9499   73.6636  0.7782252 2.077958e+00 1.331361e-01
## 3 741.9526 5993.3481 23.1426263 1.690648e+02 1.083001e+01
## 4 741.9526    0.0000 23.1426263 0.000000e+00 2.600000e-05
## 5 741.9526    0.0000 23.1426263 0.000000e+00 2.600000e-05
## 6   9.2977 5993.3481  0.2900094 1.690648e+02 1.083001e+01
```

Make final dataframe with IDs

```
#setwd("S:/Biogeochemistry/People/Wilson (Steph)/Data/Dionex/Final Data Files")        #Change wd

write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202304.csv")           #C
```

END