Dionex_COMPASS_September2022

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_202305_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 4.887   0.0084    22.74   0.0132     0.06
## 2 2  Lab Blank             Unknown 4.907   0.0091    23.29   0.0143     0.06
## 3 3  Lab Blank             Unknown 4.907   0.0094    32.59   0.0148     0.06
## 4 4  Lab Blank             Unknown 4.890   0.0090    32.12   0.0142     0.06
## 5 5 Standard 1 Calibration Standard 4.893   0.4337     7.79   0.6859     3.45
## 6 6 Standard 2 Calibration Standard 4.900   1.0909    10.03   1.7251     6.45
##   IC.SO4.5
## 1      BMB
## 2      BMB
## 3      BMB
## 4      BMB
## 5       Rd
## 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  0.0084
## 2  Lab Blank  0.0091
## 3  Lab Blank  0.0094
## 4  Lab Blank  0.0090
## 5 Standard 1  0.4337
## 6 Standard 2  1.0909
```

```
## 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  0.0084
## 2  Lab Blank  0.0091
## 3  Lab Blank  0.0094
## 4  Lab Blank  0.0090
## 5 Standard 1  0.4337
## 6 Standard 2  1.0909
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202305_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 3.890  0.0086   22.84   0.0133    0.05
## 2 2  Lab Blank             Unknown 3.857  0.0093   23.54   0.0145    0.05
## 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 3.900  5.1547   90.74   7.9918   33.37
## 6 6 Standard 2 Calibration Standard 3.860  9.9035   89.27  15.3544   65.84
##   IC.Cl.5
## 1      MB
## 2      MB
## 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 0.0086
## 2  Lab Blank 0.0093
## 3  Lab Blank    n.a.
## 4  Lab Blank    n.a.
## 5 Standard 1 5.1547
## 6 Standard 2 9.9035
```

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

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

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

```
##    Sample_ID Cl_ppm
## 1  Lab Blank 0.0086
## 2  Lab Blank 0.0093
## 3  Lab Blank     NA
## 4  Lab Blank     NA
## 5 Standard 1 5.1547
## 6 Standard 2 9.9035
```

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

```
##   Sample_ID  SO4_ppm    Cl_ppm
## 1           911.5208 8679.8805
## 2           911.5208 2977.5413
## 3           911.5208    0.0086
## 4           911.5208 2622.8924
## 5           911.5208        NA
## 6           265.1608 8679.8805
```

```
## 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_202305_UP_lysA_20cm  6.9968  66.1749
## 27      10_GCW_202305_TR_LysB_45cm 10.4414  75.9797
## 28  10_GCW_202305_TR_LysB_45cm_dup 10.4558  75.9221
## 29     100_GWI_202305_WC_LysB_20cm 64.0087 276.9108
## 30 100_GWI_202305_WC_LysB_20cm_dup 33.9433 300.0573
## 31     101_GWI_202305_WC_LysB_45cm 25.5771 260.4622
```

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

```
##                          Sample_ID SO4_ppm   Cl_ppm
## 26       1_GCW_202305_UP_lysA_20cm  6.9968  66.1749
## 27      10_GCW_202305_TR_LysB_45cm 10.4414  75.9797
## 28  10_GCW_202305_TR_LysB_45cm_dup 10.4558  75.9221
## 29     100_GWI_202305_WC_LysB_20cm 64.0087 276.9108
## 30 100_GWI_202305_WC_LysB_20cm_dup 33.9433 300.0573
## 31     101_GWI_202305_WC_LysB_45cm 25.5771 260.4622
```

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

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

```
##      Sample_ID SO4_ppm Cl_ppm
## 459 Standard 1  0.4622 5.2745
## 460 Standard 1  0.4622 5.1604
## 461 Standard 1  0.4622 5.1547
## 462 Standard 1  0.4622 5.1937
## 463 Standard 1  0.4622 5.2025
## 464 Standard 1  0.4388 5.2745
```

```
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  0.444 0.00987  2.22  NO, rerun
## 2 Standard 2  1.03  0.150   14.5   NO, rerun
## 3 Standard 3  2.06  0.0194   0.940 YES
## 4 Standard 4  9.97  0.165    1.66  YES
## 5 Standard 5 20.0   0.219    1.10  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   5.20 0.0437 0.841 YES
## 2 Standard 2  10.1  0.133  1.32  YES
## 3 Standard 3  20.1  0.188  0.934 YES
## 4 Standard 4 102.   0.860  0.845 YES
## 5 Standard 5 202.   1.58   0.785 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_202305_UP_lysA_20cm  6.9968  66.1749
## 27      10_GCW_202305_TR_LysB_45cm 10.4414  75.9797
## 28  10_GCW_202305_TR_LysB_45cm_dup 10.4558  75.9221
## 29     100_GWI_202305_WC_LysB_20cm 64.0087 276.9108
## 30 100_GWI_202305_WC_LysB_20cm_dup 33.9433 300.0573
## 31     101_GWI_202305_WC_LysB_45cm 25.5771 260.4622
```

```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_GCW_202305_UP_lysA_20cm  6.9968  66.1749 0.2182408 1.866711
## 27      10_GCW_202305_TR_LysB_45cm 10.4414  75.9797 0.3256831 2.143292
## 28  10_GCW_202305_TR_LysB_45cm_dup 10.4558  75.9221 0.3261323 2.141667
## 29     100_GWI_202305_WC_LysB_20cm 64.0087 276.9108 1.9965284 7.811306
## 30 100_GWI_202305_WC_LysB_20cm_dup 33.9433 300.0573 1.0587430 8.464240
## 31     101_GWI_202305_WC_LysB_45cm 25.5771 260.4622 0.7977885 7.347312
##     salinity
## 26 0.1196040
## 27 0.1373213
## 28 0.1372172
## 29 0.5004038
## 30 0.5422295
## 31 0.4706812
```

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_GCW_202305_UP_lysA_20cm  6.9968  66.1749 0.2182408 1.866711
## 27      10_GCW_202305_TR_LysB_45cm 10.4414  75.9797 0.3256831 2.143292
```

```
## 28  10_GCW_202305_TR_LysB_45cm_dup 10.4558  75.9221 0.3261323 2.141667
## 29     100_GWI_202305_WC_LysB_20cm 64.0087 276.9108 1.9965284 7.811306
## 30 100_GWI_202305_WC_LysB_20cm_dup 33.9433 300.0573 1.0587430 8.464240
## 31     101_GWI_202305_WC_LysB_45cm 25.5771 260.4622 0.7977885 7.347312
##     salinity
## 26 0.1196040
## 27 0.1373213
## 28 0.1372172
## 29 0.5004038
## 30 0.5422295
## 31 0.4706812
```

```
#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_202305_TR_LysB_45cm_dup  10.4558   75.9221  0.3261323   2.141667
## 2    100_GWI_202305_WC_LysB_20cm_dup  33.9433  300.0573  1.0587430   8.464240
## 3    110_SWH_202305_UP_LysB_10cm_dup 147.8545 2731.8051  4.6118060  77.060793
## 4 120_SWH_202305_UPCON_LysB_20cm_dup 133.1384 1467.6991  4.1527885  41.401949
## 5    130_SWH_202305_TR_LysB_45cm_dup 462.7866 3997.2233 14.4350156 112.756652
## 6    140_SWH_202305_WC_LysC_10cm_dup 409.7469 4046.0590 12.7806269 114.134245
##    salinity
## 1 0.1372172
## 2 0.5422295
## 3 4.9363978
## 4 2.6521583
## 5 7.2230085
## 6 7.3112546
```

```
#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_GCW_202305_UP_lysA_20cm  6.9968  66.1749 0.2182408 1.866711 0.1196040
## 2  10_GCW_202305_TR_LysB_45cm 10.4414  75.9797 0.3256831 2.143292 0.1373213
## 3 100_GWI_202305_WC_LysB_20cm 64.0087 276.9108 1.9965284 7.811306 0.5004038
## 4 101_GWI_202305_WC_LysB_45cm 25.5771 260.4622 0.7977885 7.347312 0.4706812
## 5 102_GWI_202305_WC_LysC_10cm 34.8612 272.9539 1.0873737 7.699687 0.4932537
## 6 103_GWI_202305_WC_LysC_20cm 39.7708 302.7670 1.2405115 8.540677 0.5471260
```

```
#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_202305_TR_LysB_45cm  0.3261323   2.141667    0.1372172
## 2    100_GWI_202305_WC_LysB_20cm  1.0587430   8.464240    0.5422295
## 3    110_SWH_202305_UP_LysB_10cm  4.6118060  77.060793    4.9363978
## 4 120_SWH_202305_UPCON_LysB_20cm  4.1527885  41.401949    2.6521583
## 5    130_SWH_202305_TR_LysB_45cm 14.4350156 112.756652    7.2230085
## 6    140_SWH_202305_WC_LysC_10cm 12.7806269 114.134245    7.3112546
```

```
#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_202305_TR_LysB_45cm  10.4414   75.9797 0.3256831   2.143292
## 2    100_GWI_202305_WC_LysB_20cm  64.0087  276.9108 1.9965284   7.811306
## 3    110_SWH_202305_UP_LysB_10cm 215.0081 2446.1470 6.7064286  69.002736
## 4 120_SWH_202305_UPCON_LysB_20cm 162.4131 2237.4438 5.0659108  63.115481
## 5    130_SWH_202305_TR_LysB_45cm 271.3318 3655.9974 8.4632502 103.131097
## 6    140_SWH_202305_WC_LysC_10cm  92.1882 4023.6572 2.8754897 113.502319
##    salinity SO4_mM_dup  Cl_mM_dup salinity_dup
## 1 0.1373213  0.3261323   2.141667    0.1372172
## 2 0.5004038  1.0587430   8.464240    0.5422295
## 3 4.4202136  4.6118060  77.060793    4.9363978
## 4 4.0430869  4.1527885  41.401949    2.6521583
## 5 6.6064133 14.4350156 112.756652    7.2230085
## 6 7.2707746 12.7806269 114.134245    7.3112546
```

```
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)
```

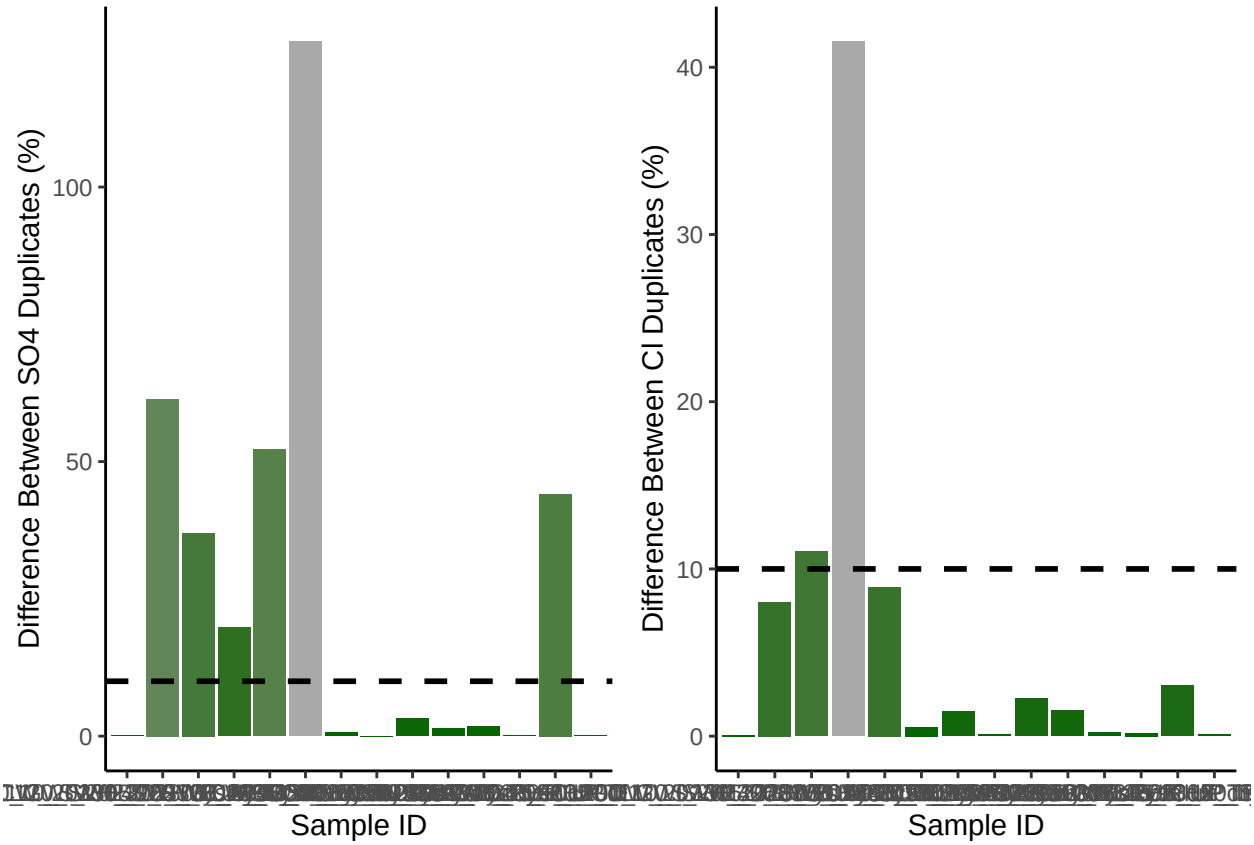
```
##                         Sample_ID  SO4_ppm    Cl_ppm    SO4_mM      Cl_mM
## 1     10_GCW_202305_TR_LysB_45cm  10.4414   75.9797 0.3256831   2.143292
## 2    100_GWI_202305_WC_LysB_20cm  64.0087  276.9108 1.9965284   7.811306
## 3    110_SWH_202305_UP_LysB_10cm 215.0081 2446.1470 6.7064286  69.002736
## 4 120_SWH_202305_UPCON_LysB_20cm 162.4131 2237.4438 5.0659108  63.115481
## 5    130_SWH_202305_TR_LysB_45cm 271.3318 3655.9974 8.4632502 103.131097
## 6    140_SWH_202305_WC_LysC_10cm  92.1882 4023.6572 2.8754897 113.502319
##    salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 0.1373213  0.3261323   2.141667    0.1372172    0.1378175           YES
## 2 0.5004038  1.0587430   8.464240    0.5422295   61.3880268     NO, rerun
## 3 4.4202136  4.6118060  77.060793    4.9363978   37.0132386     NO, rerun
## 4 4.0430869  4.1527885  41.401949    2.6521583   19.8102192     NO, rerun
## 5 6.6064133 14.4350156 112.756652    7.2230085   52.1591068     NO, rerun
## 6 7.2707746 12.7806269 114.134245    7.3112546  126.5337690     NO, rerun
##   Cl_dups_chk Cl_dups_flag
## 1  0.07583847          YES
## 2  8.02349385          YES
## 3 11.03363239    NO, rerun
```

```
## 4 41.55006815    NO, rerun
## 5  8.91718437          YES
## 6  0.55520664          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         6 NO, rerun          2    14  42.85714   14.28571
## 2       YES         8       YES         12    14  57.14286   85.71429
```

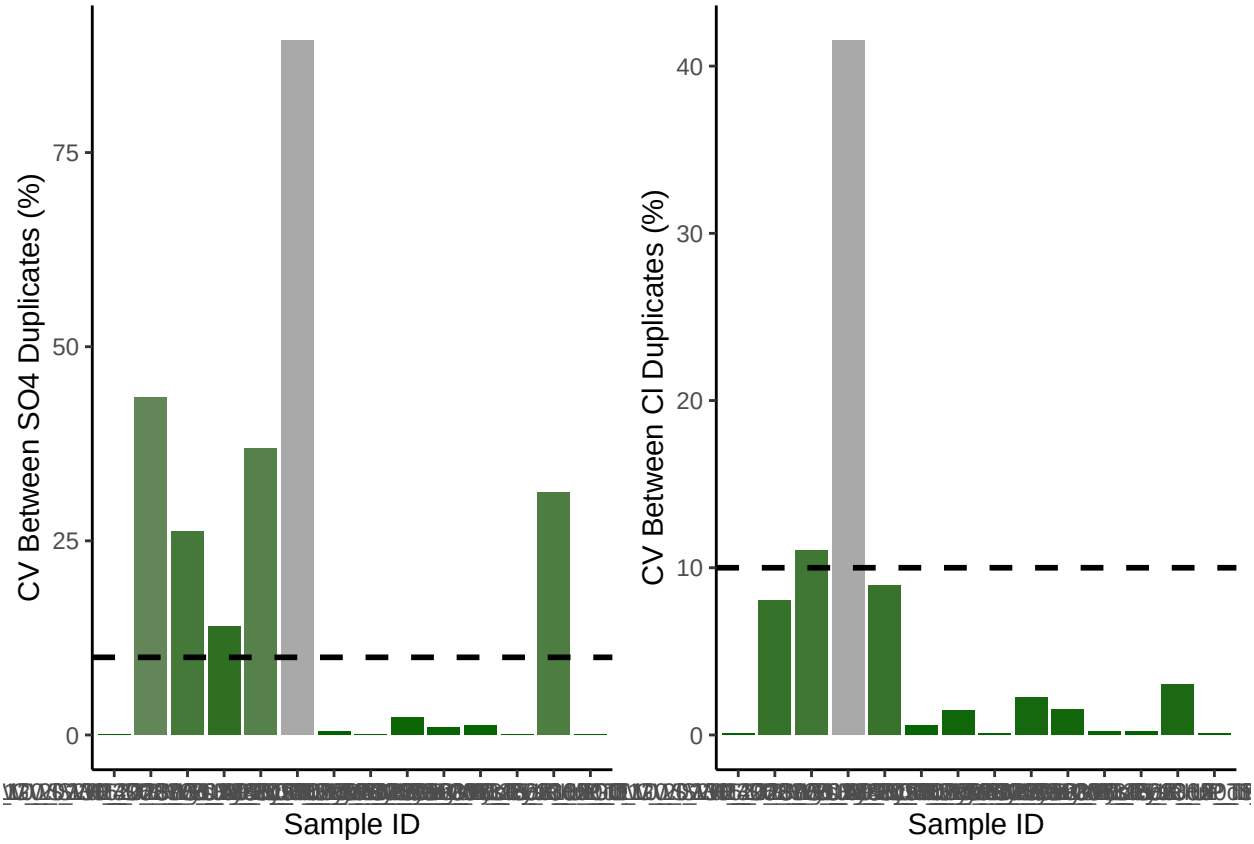
Pull out dups and check with cv

```
#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_GCW_202305_TR_LysB_45cm  10.4414   75.9797 0.3256831   2.143292
## 2    100_GWI_202305_WC_LysB_20cm  64.0087  276.9108 1.9965284   7.811306
## 3    110_SWH_202305_UP_LysB_10cm 215.0081 2446.1470 6.7064286  69.002736
## 4 120_SWH_202305_UPCON_LysB_20cm 162.4131 2237.4438 5.0659108  63.115481
## 5    130_SWH_202305_TR_LysB_45cm 271.3318 3655.9974 8.4632502 103.131097
## 6    140_SWH_202305_WC_LysC_10cm  92.1882 4023.6572 2.8754897 113.502319
##    salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 0.1373213  0.3261323   2.141667    0.1372172    0.1378175           YES
## 2 0.5004038  1.0587430   8.464240    0.5422295   61.3880268     NO, rerun
## 3 4.4202136  4.6118060  77.060793    4.9363978   37.0132386     NO, rerun
## 4 4.0430869  4.1527885  41.401949    2.6521583   19.8102192     NO, rerun
## 5 6.6064133 14.4350156 112.756652    7.2230085   52.1591068     NO, rerun
## 6 7.2707746 12.7806269 114.134245    7.3112546  126.5337690     NO, rerun
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1  0.07583847          YES  0.09745169              YES
```

```
## 2  8.02349385          YES 43.40789003        NO, rerun
## 3 11.03363239    NO, rerun 26.17231202        NO, rerun
## 4 41.55006815    NO, rerun 14.00794033        NO, rerun
## 5  8.91718437          YES 36.88205809        NO, rerun
## 6  0.55520664          YES 89.47288611        NO, rerun
```

```
#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 NO, rerun         6 NO, rerun          2    14  42.85714   14.28571
## 2       YES         8       YES         12    14  57.14286   85.71429
```

Pull out spikes and check

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

```
##                          Sample_ID SO4_ppm   Cl_ppm    SO4_mM    Cl_mM
## 26       1_GCW_202305_UP_lysA_20cm  6.9968  66.1749 0.2182408 1.866711
## 27      10_GCW_202305_TR_LysB_45cm 10.4414  75.9797 0.3256831 2.143292
## 28  10_GCW_202305_TR_LysB_45cm_dup 10.4558  75.9221 0.3261323 2.141667
## 29     100_GWI_202305_WC_LysB_20cm 64.0087 276.9108 1.9965284 7.811306
## 30 100_GWI_202305_WC_LysB_20cm_dup 33.9433 300.0573 1.0587430 8.464240
## 31     101_GWI_202305_WC_LysB_45cm 25.5771 260.4622 0.7977885 7.347312
##     salinity
## 26 0.1196040
## 27 0.1373213
## 28 0.1372172
## 29 0.5004038
## 30 0.5422295
## 31 0.4706812
```

```r
#pull out any rows that have "spk" in the SampleID column
spks <- sampledat %>%
  filter(str_detect(Sample_ID, "spk"))      #have to change this to match data
head(spks)
```

```
##                            Sample_ID  SO4_ppm    Cl_ppm    SO4_mM      Cl_mM
## 1    101_GWI_202305_WC_LysB_45cm_spk  71.1032  289.7112 2.2178166   8.172389
## 2     11_GCW_202305_TR_LysC_10cm_spk   8.4098   78.3821 0.2623144   2.211061
## 3    111_SWH_202305_UP_LysB_20cm_spk 201.8888 2235.9854 6.2972177  63.074341
## 4 121_SWH_202305_UPCON_LysB_45cm_spk  28.7040  473.5673 0.8953213  13.358739
```

```
## 5    131_SWH_202305_TR_LysC_10cm_spk 318.4301 3501.4783 9.9323175  98.772307
## 6    141_SWH_202305_WC_LysC_20cm_spk 109.9728 3997.5809 3.4302183 112.766739
##    salinity
## 1 0.5235341
## 2 0.1416625
## 3 4.0404516
## 4 0.8557621
## 5 6.3271973
## 6 7.2236547
```

```
#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_202305_WC_LysB_45cm  2.2178166
## 2     11_GCW_202305_TR_LysC_10cm  0.2623144
## 3    111_SWH_202305_UP_LysB_20cm  6.2972177
## 4 121_SWH_202305_UPCON_LysB_45cm  0.8953213
## 5    131_SWH_202305_TR_LysC_10cm  9.9323175
## 6    141_SWH_202305_WC_LysC_20cm  3.4302183
```

```
#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_202305_WC_LysB_45cm  25.5771  260.4622  0.7977885   7.347312
## 2     11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 3    111_SWH_202305_UP_LysB_20cm 190.0090 2248.7667  5.9266687  63.434886
## 4 121_SWH_202305_UPCON_LysB_45cm 204.1464 2663.3140  6.3676357  75.128745
## 5    131_SWH_202305_TR_LysC_10cm 353.5336 3529.3181 11.0272489  99.557633
## 6    141_SWH_202305_WC_LysC_20cm 388.4101 3977.8297 12.1150998 112.209583
##    salinity SO4_mM_spk
## 1 0.4706812  2.2178166
## 2 0.1400748  0.2623144
## 3 4.0635474  6.2972177
## 4 4.8126344  0.8953213
## 5 6.3775038  9.9323175
## 6 7.1879643  3.4302183
```

```
#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_202305_WC_LysB_45cm  25.5771  260.4622  0.7977885   7.347312
## 2     11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 3    111_SWH_202305_UP_LysB_20cm 190.0090 2248.7667  5.9266687  63.434886
## 4 121_SWH_202305_UPCON_LysB_45cm 204.1464 2663.3140  6.3676357  75.128745
## 5    131_SWH_202305_TR_LysC_10cm 353.5336 3529.3181 11.0272489  99.557633
## 6    141_SWH_202305_WC_LysC_20cm 388.4101 3977.8297 12.1150998 112.209583
##    salinity SO4_mM_spk SO4_spk_Conc
## 1 0.4706812  2.2178166 7.797879e-05
## 2 0.1400748  0.2623144 7.797879e-05
## 3 4.0635474  6.2972177 7.797879e-05
## 4 4.8126344  0.8953213 7.797879e-05
## 5 6.3775038  9.9323175 7.797879e-05
## 6 7.1879643  3.4302183 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"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 50, 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"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_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_202305_WC_LysB_45cm  25.5771  260.4622  0.7977885   7.347312
## 2     11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 3    111_SWH_202305_UP_LysB_20cm 190.0090 2248.7667  5.9266687  63.434886
```

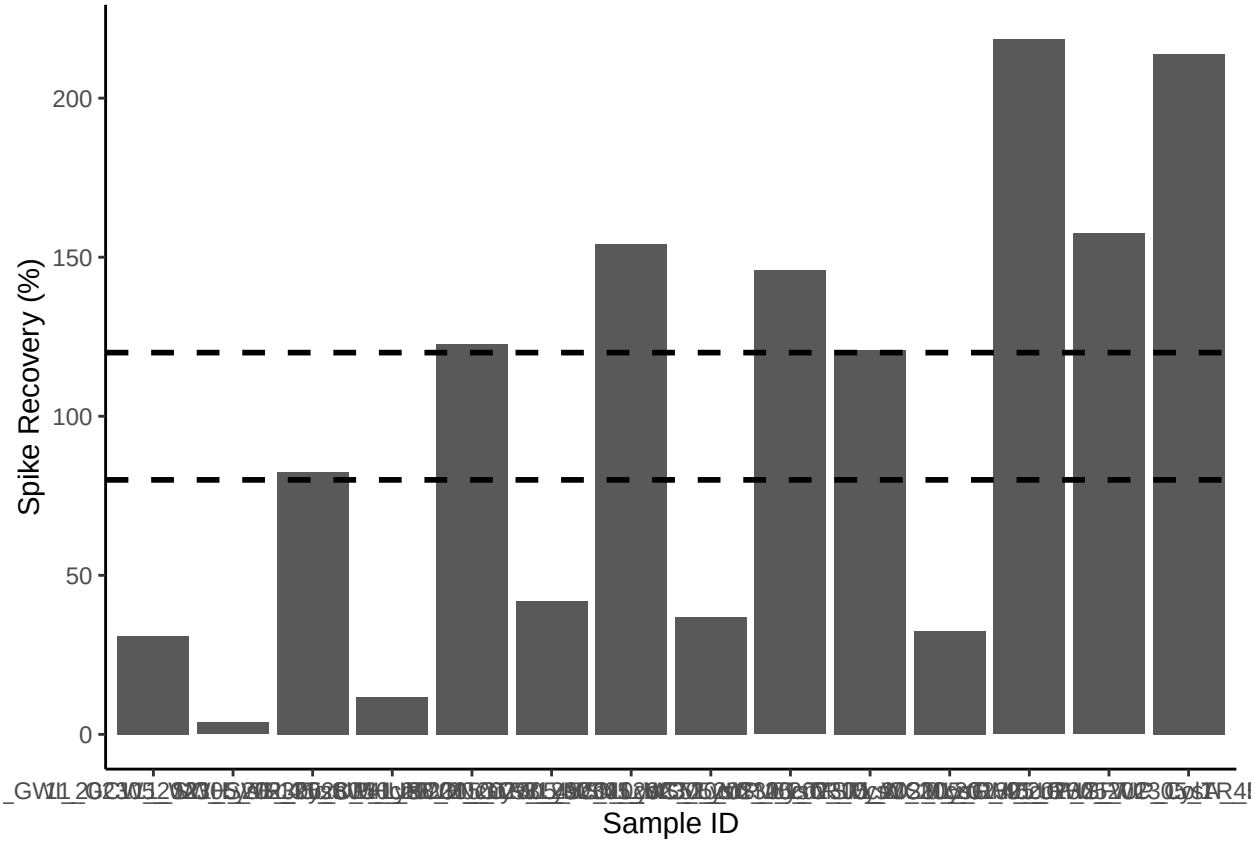
```
## 4 121_SWH_202305_UPCON_LysB_45cm 204.1464 2663.3140  6.3676357  75.128745
## 5    131_SWH_202305_TR_LysC_10cm 353.5336 3529.3181 11.0272489  99.557633
## 6    141_SWH_202305_WC_LysC_20cm 388.4101 3977.8297 12.1150998 112.209583
##    salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 0.4706812  2.2178166 7.797879e-05        1     1e-06
## 2 0.1400748  0.2623144 7.797879e-05        1     1e-06
## 3 4.0635474  6.2972177 7.797879e-05        1     1e-06
## 4 4.8126344  0.8953213 7.797879e-05        1     1e-06
## 5 6.3775038  9.9323175 7.797879e-05        1     1e-06
## 6 7.1879643  3.4302183 7.797879e-05        1     1e-06
```

```
#gives us the total SO4 in the sample in mmoles
QAspks$SO4_Total_unspkd <- (QAspks$SO4_mM/QAspks$Dilution)*(QAspks$SampleVol)

##total SO4 in spiked sample in mmoles
QAspks$SO4_Total_spkd <- (QAspks$SO4_mM_spk/QAspks$Dilution)*(QAspks$SampleVol+spkvol)

QAspks$SO4_expctd_spkd <-  (QAspks$SO4_Total_unspkd + QAspks$SO4_spk_Conc)
QAspks$spk_recovery <-    (QAspks$SO4_Total_spkd/QAspks$SO4_expctd_spkd)*100
QAspks$SO4_spks_flag <-  ifelse(QAspks$spk_recovery <=120 & QAspks$spk_recovery >=80 , 'YES', 'NO, rerun

head(QAspks)
```

```
##                        Sample_ID  SO4_ppm    Cl_ppm     SO4_mM      Cl_mM
## 1    101_GWI_202305_WC_LysB_45cm  25.5771  260.4622  0.7977885   7.347312
## 2     11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 3    111_SWH_202305_UP_LysB_20cm 190.0090 2248.7667  5.9266687  63.434886
## 4 121_SWH_202305_UPCON_LysB_45cm 204.1464 2663.3140  6.3676357  75.128745
## 5    131_SWH_202305_TR_LysC_10cm 353.5336 3529.3181 11.0272489  99.557633
## 6    141_SWH_202305_WC_LysC_20cm 388.4101 3977.8297 12.1150998 112.209583
##    salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1 0.4706812  2.2178166 7.797879e-05        1     1e-06     7.977885e-07
## 2 0.1400748  0.2623144 7.797879e-05        1     1e-06     2.713506e-07
## 3 4.0635474  6.2972177 7.797879e-05        1     1e-06     5.926669e-06
## 4 4.8126344  0.8953213 7.797879e-05        1     1e-06     6.367636e-06
## 5 6.3775038  9.9323175 7.797879e-05        1     1e-06     1.102725e-05
## 6 7.1879643  3.4302183 7.797879e-05        1     1e-06     1.211510e-05
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   2.439598e-05    7.877658e-05     30.96857     NO, rerun
## 2   2.885459e-06    7.825014e-05      3.68748     NO, rerun
## 3   6.926939e-05    8.390546e-05     82.55648           YES
## 4   9.848534e-06    8.434643e-05     11.67629     NO, rerun
## 5   1.092555e-04    8.900604e-05    122.75065     NO, rerun
## 6   3.773240e-05    9.009389e-05     41.88120     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          13    14   92.9
## 2 YES                 1    14    7.14
```

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", "202305", "UP", "lysA", "20cm"), c("10", :
## number of columns of result is not a multiple of vector length (arg 7)
```

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

```
##   Analysis_No Date   Site Zone Replicate Depth
## 1           1  GCW 202305   UP      lysA  20cm
## 2          10  GCW 202305   TR      LysB  45cm
## 3         100  GWI 202305   WC      LysB  20cm
## 4         101  GWI 202305   WC      LysB  45cm
## 5         102  GWI 202305   WC      LysC  10cm
## 6         103  GWI 202305   WC      LysC  20cm
```

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

```
##   Analysis_No Date   Site Zone Replicate Depth                    Sample_ID
## 1           1  GCW 202305   UP      lysA  20cm   1_GCW_202305_UP_lysA_20cm
## 2          10  GCW 202305   TR      LysB  45cm  10_GCW_202305_TR_LysB_45cm
## 3         100  GWI 202305   WC      LysB  20cm 100_GWI_202305_WC_LysB_20cm
## 4         101  GWI 202305   WC      LysB  45cm 101_GWI_202305_WC_LysB_45cm
## 5         102  GWI 202305   WC      LysC  10cm 102_GWI_202305_WC_LysC_10cm
## 6         103  GWI 202305   WC      LysC  20cm 103_GWI_202305_WC_LysC_20cm
##   SO4_ppm   Cl_ppm    SO4_mM    Cl_mM  salinity
## 1  6.9968  66.1749 0.2182408 1.866711 0.1196040
## 2 10.4414  75.9797 0.3256831 2.143292 0.1373213
## 3 64.0087 276.9108 1.9965284 7.811306 0.5004038
## 4 25.5771 260.4622 0.7977885 7.347312 0.4706812
## 5 34.8612 272.9539 1.0873737 7.699687 0.4932537
## 6 39.7708 302.7670 1.2405115 8.540677 0.5471260
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202305.csv")    #Change f
```

END