Dionex_COMPASS_November2022

Stephanie J. Wilson

2023-01-23

Daily Set up

Read in and Format the raw data - change wd & file names (Make sure files are UTF-8 encoded)

```
# SULFATE DATA:

## Read in raw data file from Dionex - copied and saved as a txt
Sdat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202211_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 Standard 1             Unknown 4.397   0.5017     5.85   0.4070     2.53
## 4 4      Blank             Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 5 5 Standard 1 Calibration Standard 4.393   0.5034     5.87   0.4084     2.54
## 6 6 Standard 2 Calibration Standard 4.400   1.0993     6.12   0.8917     5.39
```

```
## 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 Standard 1  0.5017
## 4      Blank     n.a.
## 5 Standard 1  0.5034
## 6 Standard 2  1.0993
```

```
## 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 Standard 1  0.5017
## 4      Blank      NA
## 5 Standard 1  0.5034
## 6 Standard 2  1.0993
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202211_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 3.270   0.0241  100.00   0.0271    0.19
## 2 2  Lab Blank             Unknown 3.270   0.0251  100.00   0.0282    0.20
## 3 3 Standard 1             Unknown 3.273   5.7878   93.58   6.5054   43.94
## 4 4      Blank             Unknown 3.270   0.0260  100.00   0.0292    0.21
## 5 5 Standard 1 Calibration Standard 3.273   5.7959   93.56   6.5145   44.12
## 6 6 Standard 2 Calibration Standard 3.273  12.0880   93.22  13.5866   90.14
```

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

```
##          X.1  IC.Cl.1
## 1  Lab Blank  0.0241
## 2  Lab Blank  0.0251
## 3 Standard 1  5.7878
## 4      Blank  0.0260
## 5 Standard 1  5.7959
## 6 Standard 2 12.0880
```

```
## 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  0.0241
## 2  Lab Blank  0.0251
## 3 Standard 1  5.7878
## 4      Blank  0.0260
## 5 Standard 1  5.7959
## 6 Standard 2 12.0880
```

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

```
##   Sample_ID   SO4_ppm     Cl_ppm
## 1           2544.4789 19269.7855
## 2           2544.4789  5007.1144
## 3           2544.4789     0.0179
## 4           2544.4789  5111.7920
## 5           2544.4789         NA
## 6            784.5324 19269.7855
```

```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_202211_MSM_UP_LysA_10cm  264.6481 3270.8644
## 27     10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636
## 28 10_202211_MSM_TR_LysA_20cm_spk 1248.1115 6395.9330
## 29  100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034
## 30  100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425
## 31  100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034
```

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

```
##                          Sample_ID   SO4_ppm    Cl_ppm
## 26      1_202211_MSM_UP_LysA_10cm  264.6481 3270.8644
## 27     10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636
## 28 10_202211_MSM_TR_LysA_20cm_spk 1248.1115 6395.9330
## 29  100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034
## 30  100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425
## 31  100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034
```

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
## 1730 Standard 1  0.5034 5.7959
## 1731 Standard 1  0.5034 5.8500
## 1732 Standard 1  0.5034 5.8551
## 1733 Standard 1  0.5034 5.9283
## 1734 Standard 1  0.5034 5.9000
## 1735 Standard 1  0.5034 5.7635
```

```
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  0.506 0.0109 2.16  NO, rerun
## 2 Standard 2  1.11  0.0162 1.46  YES
## 3 Standard 3  2.22  0.0416 1.88  YES
## 4 Standard 4  8.80  0.100  1.14  YES
## 5 Standard 5 20.4   0.190  0.935 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   5.84 0.0570 0.975 YES
## 2 Standard 2  12.2  0.0665 0.546 YES
## 3 Standard 3  23.4  0.186  0.796 YES
## 4 Standard 4  89.3  0.628  0.704 YES
## 5 Standard 5 203.   1.53   0.753 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_202211_MSM_UP_LysA_10cm  264.6481 3270.8644
## 27     10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636
## 28 10_202211_MSM_TR_LysA_20cm_spk 1248.1115 6395.9330
## 29  100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034
## 30  100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425
## 31  100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034
```

```
# 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_202211_MSM_UP_LysA_10cm  264.6481 3270.8644  8.254775  92.266979
## 27     10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636 41.341089 193.951018
## 28 10_202211_MSM_TR_LysA_20cm_spk 1248.1115 6395.9330 38.930490 180.421241
## 29  100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034  4.951578   3.780068
## 30  100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425  4.951578   3.758604
## 31  100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034  0.604975   3.780068
##      salinity
## 26  5.9104780
## 27 12.4241694
## 28 11.5574769
## 29  0.2421701
## 30  0.2407952
## 31  0.2421701
```

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_202211_MSM_UP_LysA_10cm  264.6481 3270.8644  8.254775  92.266979
## 27     10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636 41.341089 193.951018
## 28 10_202211_MSM_TR_LysA_20cm_spk 1248.1115 6395.9330 38.930490 180.421241
## 29  100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034  4.951578   3.780068
## 30  100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425  4.951578   3.758604
## 31  100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034  0.604975   3.780068
##      salinity
## 26  5.9104780
## 27 12.4241694
## 28 11.5574769
## 29  0.2421701
## 30  0.2407952
## 31  0.2421701
```

```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
## 1 109_202211_GCrew_WC_SipC_45cm_dup  490.3857  5847.447 15.29587 164.9491
## 2   129_202211_GWI_WC_SIpA_10cm_dup 2192.9807 15519.800 68.40239 437.7941
## 3           139_202211_GWI_SW_C_dup 1972.7418 14364.361 61.53281 405.2006
## 4    19_202211_MSM_WC_SipA_20cm_dup 2527.3082 14813.024 78.83057 417.8568
## 5            29_202211_MSM_SW_C_dup 2245.2770 11988.478 70.03359 338.1799
## 6     39_202211_MSM_RHZ_TR_SF_2_dup  658.3043  4849.288 20.53351 136.7923
##   salinity
## 1 10.56636
## 2 28.04430
## 3 25.95643
## 4 26.76716
## 5 21.66321
## 6  8.76269
```

```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
## 1     1_202211_MSM_UP_LysA_10cm  264.6481 3270.8644  8.254775  92.266979
## 2    10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636 41.341089 193.951018
## 3 100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034  4.951578   3.780068
## 4 100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425  4.951578   3.758604
## 5 100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034  0.604975   3.780068
## 6 100_202211_GCrew_TR_LysC_45cm   19.3955  133.2425  0.604975   3.758604
##     salinity
## 1  5.9104780
## 2 12.4241694
## 3  0.2421701
## 4  0.2407952
## 5  0.2421701
## 6  0.2407952
```

```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
## 1 109_202211_GCrew_WC_SipC_45cm   15.29587  164.9491     10.56636
## 2   129_202211_GWI_WC_SIpA_10cm   68.40239  437.7941     28.04430
## 3           139_202211_GWI_SW_C   61.53281  405.2006     25.95643
## 4    19_202211_MSM_WC_SipA_20cm   78.83057  417.8568     26.76716
```

```
## 5            29_202211_MSM_SW_C   70.03359  338.1799     21.66321
## 6     39_202211_MSM_RHZ_TR_SF_2   20.53351  136.7923      8.76269
```

```
#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  salinity
## 1 109_202211_GCrew_WC_SipC_45cm  467.7261  5581.663 14.58909 157.4517 10.086092
## 2           139_202211_GWI_SW_C 2131.0356 15474.590 66.47023 436.5188 27.962610
## 3    19_202211_MSM_WC_SipA_20cm 2365.1949 14059.697 73.77401 396.6064 25.405898
## 4            29_202211_MSM_SW_C 2098.9347 12472.348 65.46896 351.8293 22.537559
## 5     39_202211_MSM_RHZ_TR_SF_2  713.7607  5163.679 22.26328 145.6609  9.330794
## 6     49_202211_MSM_RHZ_WC_Col4 1787.4021 11210.941 55.75178 316.2466 20.258196
##   SO4_mM_dup Cl_mM_dup salinity_dup
## 1   15.29587  164.9491     10.56636
## 2   61.53281  405.2006     25.95643
## 3   78.83057  417.8568     26.76716
## 4   70.03359  338.1799     21.66321
## 5   20.53351  136.7923      8.76269
## 6   58.95680  324.4860     20.78600
```

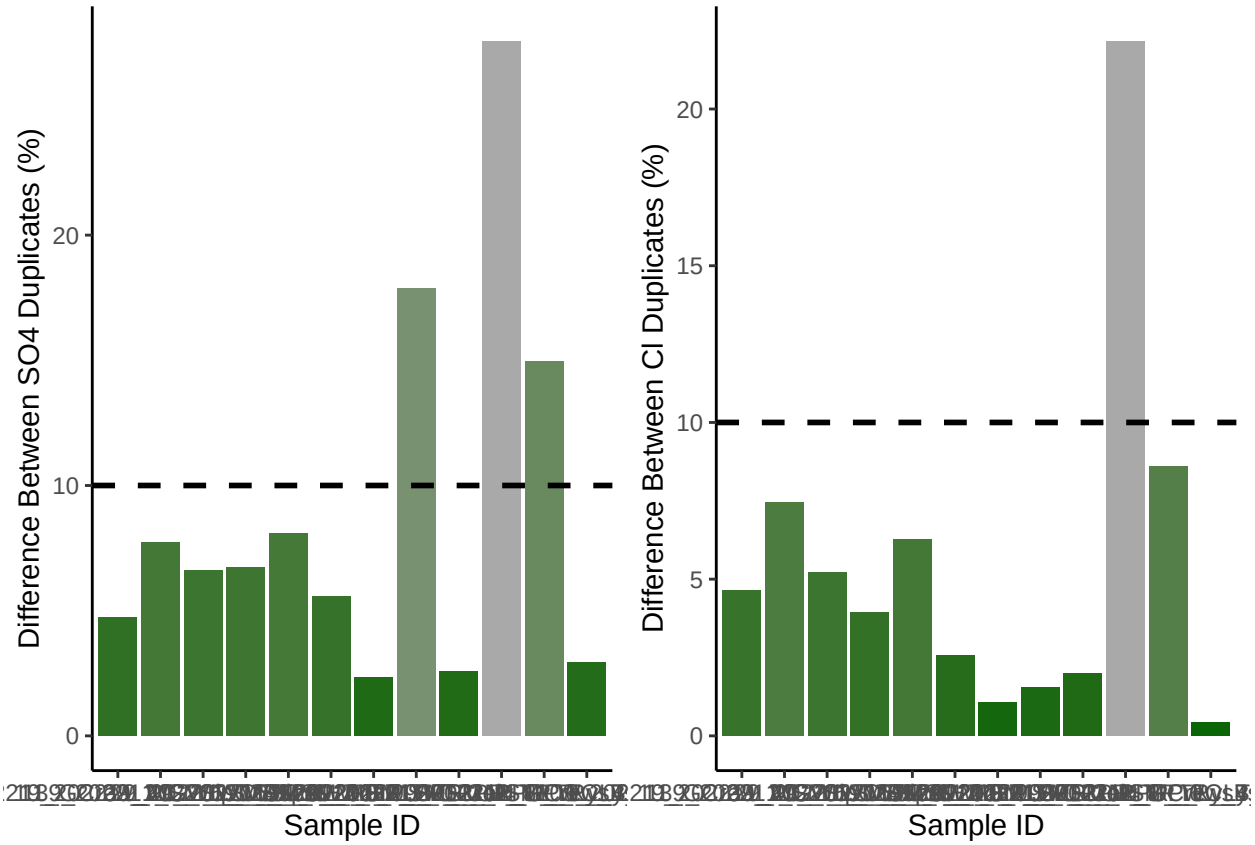
```
QAdups$SO4_dups_chk <- ((abs(QAdups$SO4_mM-QAdups$SO4_mM_dup))/((QAdups$SO4_mM+QAdups$SO4_mM_dup)/2))*1
QAdups$SO4_dups_flag <-  ifelse(QAdups$SO4_dups_chk <10, 'YES', 'NO, rerun')

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  salinity
## 1 109_202211_GCrew_WC_SipC_45cm  467.7261  5581.663 14.58909 157.4517 10.086092
## 2           139_202211_GWI_SW_C 2131.0356 15474.590 66.47023 436.5188 27.962610
## 3    19_202211_MSM_WC_SipA_20cm 2365.1949 14059.697 73.77401 396.6064 25.405898
## 4            29_202211_MSM_SW_C 2098.9347 12472.348 65.46896 351.8293 22.537559
## 5     39_202211_MSM_RHZ_TR_SF_2  713.7607  5163.679 22.26328 145.6609  9.330794
## 6     49_202211_MSM_RHZ_WC_Col4 1787.4021 11210.941 55.75178 316.2466 20.258196
##   SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag Cl_dups_chk
## 1   15.29587  164.9491     10.56636     4.730053           YES    4.650997
## 2   61.53281  405.2006     25.95643     7.714541           YES    7.441473
## 3   78.83057  417.8568     26.76716     6.627009           YES    5.218266
## 4   70.03359  338.1799     21.66321     6.737347           YES    3.956287
## 5   20.53351  136.7923      8.76269     8.083640           YES    6.279669
## 6   58.95680  324.4860     20.78600     5.588112           YES    2.571869
##   Cl_dups_flag
## 1          YES
## 2          YES
## 3          YES
## 4          YES
## 5          YES
## 6          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)
```



```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         3 NO, rerun          1    12        25   8.333333
## 2       YES         9       YES         11    12        75  91.666667
```

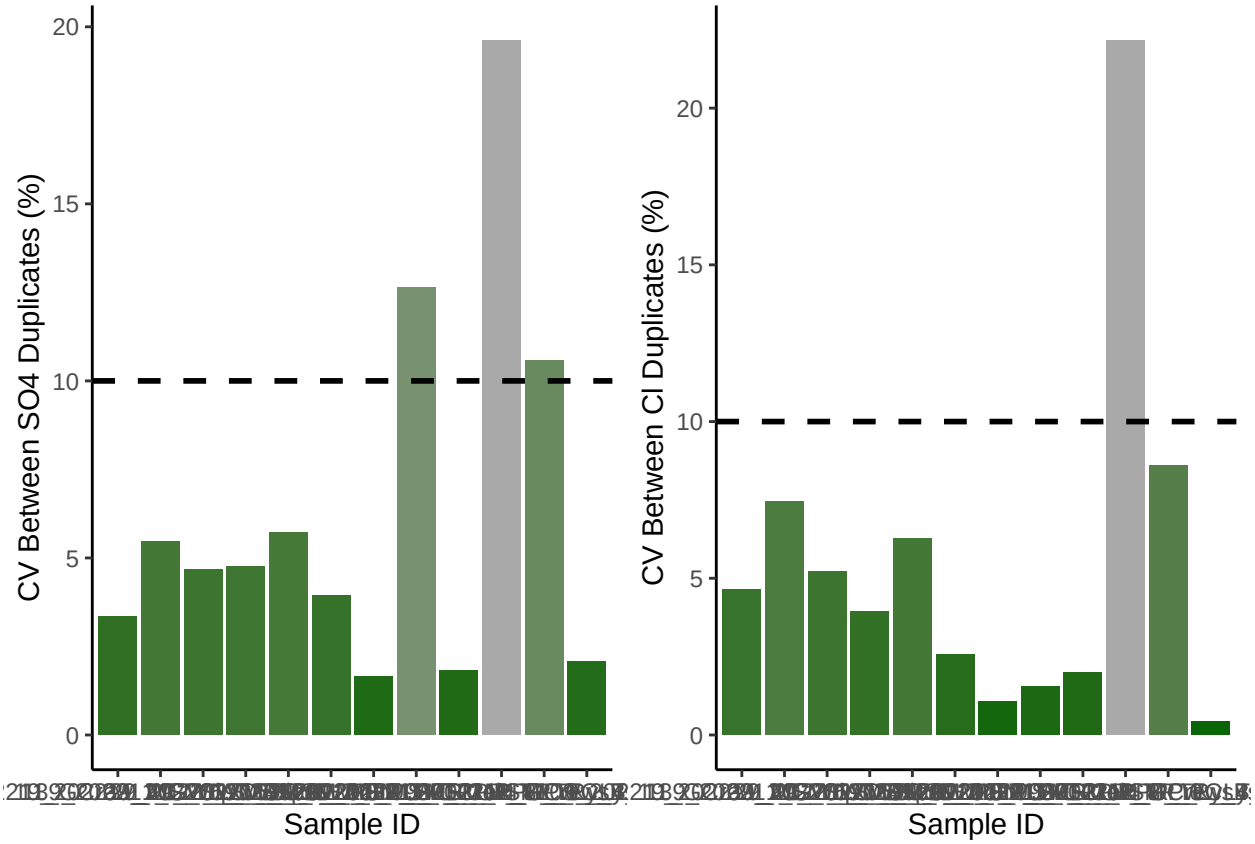
Pull out dups and check with cv

```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  salinity
## 1 109_202211_GCrew_WC_SipC_45cm  467.7261  5581.663 14.58909 157.4517 10.086092
## 2           139_202211_GWI_SW_C 2131.0356 15474.590 66.47023 436.5188 27.962610
## 3    19_202211_MSM_WC_SipA_20cm 2365.1949 14059.697 73.77401 396.6064 25.405898
## 4            29_202211_MSM_SW_C 2098.9347 12472.348 65.46896 351.8293 22.537559
## 5     39_202211_MSM_RHZ_TR_SF_2  713.7607  5163.679 22.26328 145.6609  9.330794
## 6     49_202211_MSM_RHZ_WC_Col4 1787.4021 11210.941 55.75178 316.2466 20.258196
##   SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag Cl_dups_chk
## 1   15.29587  164.9491     10.56636     4.730053           YES    4.650997
## 2   61.53281  405.2006     25.95643     7.714541           YES    7.441473
## 3   78.83057  417.8568     26.76716     6.627009           YES    5.218266
## 4   70.03359  338.1799     21.66321     6.737347           YES    3.956287
## 5   20.53351  136.7923      8.76269     8.083640           YES    6.279669
## 6   58.95680  324.4860     20.78600     5.588112           YES    2.571869
##   Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1          YES    3.344653              YES
```

```
## 2          YES    5.455004             YES
## 3          YES    4.686003             YES
## 4          YES    4.764023             YES
## 5          YES    5.715997             YES
## 6          YES    3.951392             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)
```



```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         3 NO, rerun          1    12        25   8.333333
## 2       YES         9       YES         11    12        75  91.666667
```

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_202211_MSM_UP_LysA_10cm  264.6481 3270.8644  8.254775  92.266979
## 27     10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636 41.341089 193.951018
## 28 10_202211_MSM_TR_LysA_20cm_spk 1248.1115 6395.9330 38.930490 180.421241
## 29  100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034  4.951578   3.780068
## 30  100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425  4.951578   3.758604
## 31  100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034  0.604975   3.780068
##      salinity
## 26  5.9104780
## 27 12.4241694
## 28 11.5574769
## 29  0.2421701
## 30  0.2407952
## 31  0.2421701
```

```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  10_202211_MSM_TR_LysA_20cm_spk 1248.1115  6395.933 38.93049 180.4212
## 2       110_202211_GCrew_SW_A_spk  956.5879  5163.239 29.83743 145.6485
## 3 120_202211_GWI_TR_LysA_10cm_spk 1664.3307 12466.940 51.91300 351.6767
## 4 130_202211_GWI_WC_SipA_20cm_spk 1900.5139 13145.045 59.27991 370.8052
```

```
## 5  140_202211_GWI_RHZ_TR_SF_1_spk 1606.2517 11339.052 50.10143 319.8604
## 6  20_202211_MSM_WC_SipA_45cm_spk  326.3741 12616.349 10.18010 355.8914
##    salinity
## 1 11.557477
## 2  9.329999
## 3 22.527787
## 4 23.753122
## 5 20.489693
## 6 22.797768
```

```
#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_202211_MSM_TR_LysA_20cm   38.93049
## 2       110_202211_GCrew_SW_A   29.83743
## 3 120_202211_GWI_TR_LysA_10cm   51.91300
## 4 130_202211_GWI_WC_SipA_20cm   59.27991
## 5  140_202211_GWI_RHZ_TR_SF_1   50.10143
## 6  20_202211_MSM_WC_SipA_45cm   10.18010
```

```
#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
## 1  10_202211_MSM_TR_LysA_20cm 1325.3953  6875.564 41.3410886 193.9510 12.424169
## 2       110_202211_GCrew_SW_A  663.6499  5151.559 20.7002464 145.3190  9.308893
## 3 120_202211_GWI_TR_LysA_10cm 1862.6278 14578.719 58.0981847 411.2474 26.343772
## 4 130_202211_GWI_WC_SipA_20cm 1879.0113 13943.574 58.6092109 393.3307 25.196064
## 5  140_202211_GWI_RHZ_TR_SF_1 1641.0884 12090.940 51.1880349 341.0702 21.848355
## 6  20_202211_MSM_WC_SipA_45cm   29.5728 12145.786  0.9224205 342.6174 21.947462
##   SO4_mM_spk
## 1   38.93049
## 2   29.83743
## 3   51.91300
## 4   59.27991
## 5   50.10143
## 6   10.18010
```

```
#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
```

```
## 1  10_202211_MSM_TR_LysA_20cm 1325.3953  6875.564 41.3410886 193.9510 12.424169
## 2       110_202211_GCrew_SW_A  663.6499  5151.559 20.7002464 145.3190  9.308893
## 3 120_202211_GWI_TR_LysA_10cm 1862.6278 14578.719 58.0981847 411.2474 26.343772
## 4 130_202211_GWI_WC_SipA_20cm 1879.0113 13943.574 58.6092109 393.3307 25.196064
## 5  140_202211_GWI_RHZ_TR_SF_1 1641.0884 12090.940 51.1880349 341.0702 21.848355
## 6  20_202211_MSM_WC_SipA_45cm   29.5728 12145.786  0.9224205 342.6174 21.947462
##   SO4_mM_spk SO4_spk_Conc
## 1   38.93049 7.797879e-05
## 2   29.83743 7.797879e-05
## 3   51.91300 7.797879e-05
## 4   59.27991 7.797879e-05
## 5   50.10143 7.797879e-05
## 6   10.18010 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  salinity
## 1  10_202211_MSM_TR_LysA_20cm 1325.3953  6875.564 41.3410886 193.9510 12.424169
## 2       110_202211_GCrew_SW_A  663.6499  5151.559 20.7002464 145.3190  9.308893
## 3 120_202211_GWI_TR_LysA_10cm 1862.6278 14578.719 58.0981847 411.2474 26.343772
```

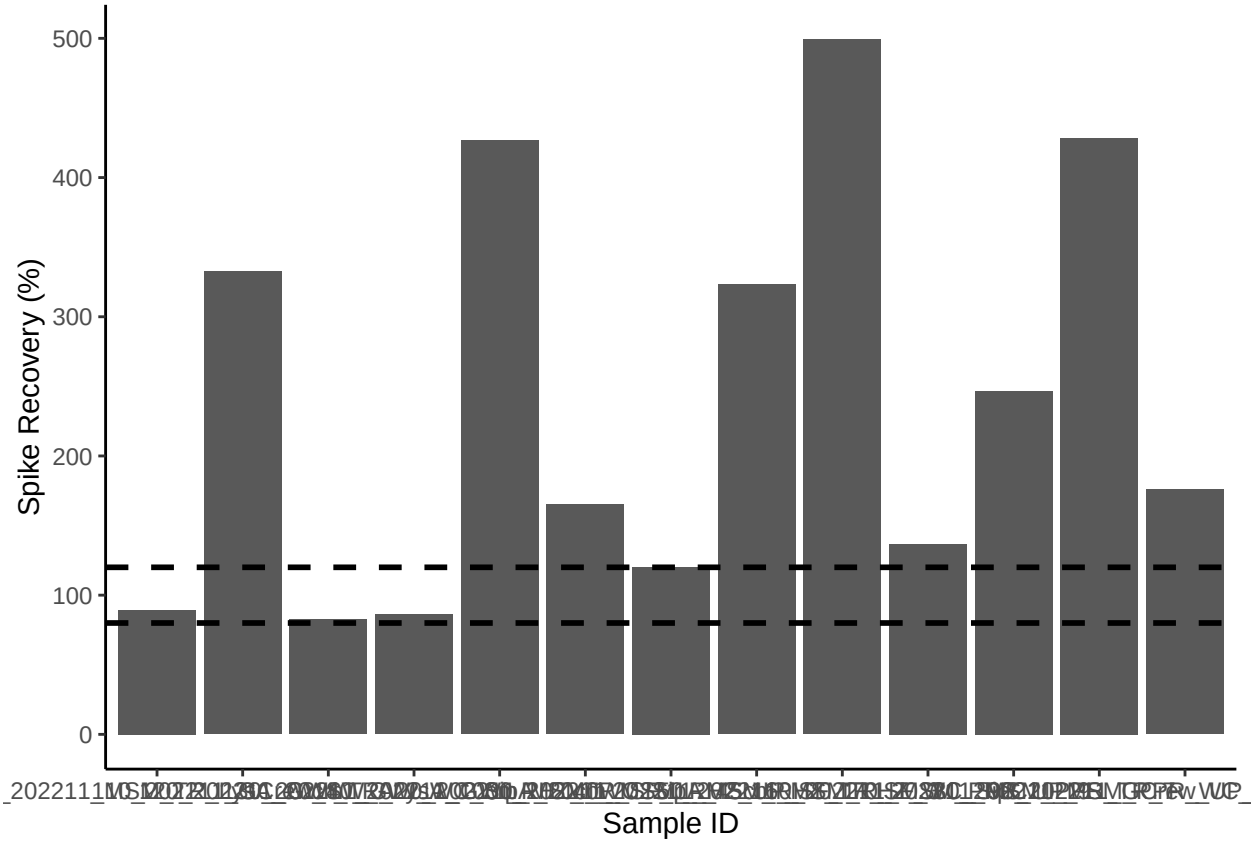
```
## 4 130_202211_GWI_WC_SipA_20cm 1879.0113 13943.574 58.6092109 393.3307 25.196064
## 5  140_202211_GWI_RHZ_TR_SF_1 1641.0884 12090.940 51.1880349 341.0702 21.848355
## 6  20_202211_MSM_WC_SipA_45cm   29.5728 12145.786  0.9224205 342.6174 21.947462
##   SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1   38.93049 7.797879e-05       50  0.001501
## 2   29.83743 7.797879e-05        1  0.000001
## 3   51.91300 7.797879e-05      100  0.001475
## 4   59.27991 7.797879e-05      200  0.001462
## 5   50.10143 7.797879e-05        1  0.000001
## 6   10.18010 7.797879e-05      100  0.001475
```

```
#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
## 1  10_202211_MSM_TR_LysA_20cm 1325.3953  6875.564 41.3410886 193.9510 12.424169
## 2       110_202211_GCrew_SW_A  663.6499  5151.559 20.7002464 145.3190  9.308893
## 3 120_202211_GWI_TR_LysA_10cm 1862.6278 14578.719 58.0981847 411.2474 26.343772
## 4 130_202211_GWI_WC_SipA_20cm 1879.0113 13943.574 58.6092109 393.3307 25.196064
## 5  140_202211_GWI_RHZ_TR_SF_1 1641.0884 12090.940 51.1880349 341.0702 21.848355
## 6  20_202211_MSM_WC_SipA_45cm   29.5728 12145.786  0.9224205 342.6174 21.947462
##   SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd SO4_Total_spkd
## 1   38.93049 7.797879e-05       50  0.001501     1.241059e-03   0.0011764794
## 2   29.83743 7.797879e-05        1  0.000001     2.070025e-05   0.0003282117
## 3   51.91300 7.797879e-05      100  0.001475     8.569482e-04   0.0007709080
## 4   59.27991 7.797879e-05      200  0.001462     4.284333e-04   0.0004363001
## 5   50.10143 7.797879e-05        1  0.000001     5.118803e-05   0.0005511157
## 6   10.18010 7.797879e-05      100  0.001475     1.360570e-05   0.0001511745
##   SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1    1.319038e-03     89.19221           YES
## 2    9.867904e-05    332.60529     NO, rerun
## 3    9.349270e-04     82.45649           YES
## 4    5.064121e-04     86.15515           YES
## 5    1.291668e-04    426.66968     NO, rerun
## 6    9.158449e-05    165.06564     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          10    13    76.9
## 2 YES                 3    13    23.1
```

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", "202211", "MSM", "UP", "LysA", "10cm"), 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 202211   MSM   UP      LysA  10cm    1
## 2          10 202211   MSM   TR      LysA  20cm   10
## 3         100 202211 GCrew   TR      LysC  45cm  100
## 4         100 202211 GCrew   TR      LysC  45cm  100
## 5         100 202211 GCrew   TR      LysC  45cm  100
## 6         100 202211 GCrew   TR      LysC  45cm  100
```

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

```
##   Analysis_No   Date  Site Zone Replicate Depth Tree
## 1           1 202211   MSM   UP      LysA  10cm    1
## 2          10 202211   MSM   TR      LysA  20cm   10
## 3         100 202211 GCrew   TR      LysC  45cm  100
## 4         100 202211 GCrew   TR      LysC  45cm  100
## 5         100 202211 GCrew   TR      LysC  45cm  100
## 6         100 202211 GCrew   TR      LysC  45cm  100
##                       Sample_ID   SO4_ppm    Cl_ppm    SO4_mM      Cl_mM
## 1     1_202211_MSM_UP_LysA_10cm  264.6481 3270.8644  8.254775  92.266979
## 2    10_202211_MSM_TR_LysA_20cm 1325.3953 6875.5636 41.341089 193.951018
## 3 100_202211_GCrew_TR_LysC_45cm  158.7476  134.0034  4.951578   3.780068
## 4 100_202211_GCrew_TR_LysC_45cm  158.7476  133.2425  4.951578   3.758604
## 5 100_202211_GCrew_TR_LysC_45cm   19.3955  134.0034  0.604975   3.780068
## 6 100_202211_GCrew_TR_LysC_45cm   19.3955  133.2425  0.604975   3.758604
##     salinity
## 1  5.9104780
## 2 12.4241694
## 3  0.2421701
## 4  0.2407952
## 5  0.2421701
## 6  0.2407952
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202211.csv")  #Change fil
```

END