Dionex_COMPASS_October2022

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_202210_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 4.400   0.5687     6.39   0.4728     2.96
## 4 4      Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 5 5 Standard 1 Calibration Standard 4.397   0.5680     6.40   0.4722     2.98
## 6 6 Standard 2 Calibration Standard 4.400   1.1543     6.46   0.9595     5.88
```

```
## Only keep the columns that we need
Sdat <- Sdat[ ,c(2,5,7)] # dont need this here
head(Sdat)
```

```
##          X.1 IC.SO4.1 IC.SO4.3
## 1  Lab Blank     n.a.     n.a.
## 2  Lab Blank     n.a.     n.a.
## 3 Standard 1  0.5687   0.4728
## 4      Blank     n.a.     n.a.
## 5 Standard 1  0.5680   0.4722
## 6 Standard 2  1.1543   0.9595
```

```
## Name the columns correctly
colnames(Sdat) <- c( "Sample_ID", "SO4_ppm", "SO4_Area")
Sdat$Sample_ID <- as.factor(Sdat$Sample_ID)
Sdat$SO4_ppm <- as.numeric(Sdat$SO4_ppm)
```

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

```
Sdat$SO4_Area <- as.numeric(Sdat$SO4_Area)
```

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

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

```
##    Sample_ID SO4_ppm SO4_Area
## 1  Lab Blank      NA       NA
## 2  Lab Blank      NA       NA
## 3 Standard 1  0.5687   0.4728
## 4      Blank      NA       NA
## 5 Standard 1  0.5680   0.4722
## 6 Standard 2  1.1543   0.9595
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202210_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 3.273   0.0471  100.00   0.0531    0.37
## 2 2  Lab Blank             Unknown 3.280   0.0472  100.00   0.0531    0.38
## 3 3 Standard 1             Unknown 3.280   6.0990   92.87   6.8696   47.28
## 4 4      Blank             Unknown 3.273   0.0480  100.00   0.0541    0.39
## 5 5 Standard 1 Calibration Standard 3.280   6.0838   92.90   6.8525   47.38
## 6 6 Standard 2 Calibration Standard 3.280  12.2230   92.75  13.7674   94.28
```

```
## Only keep the columns that we need
Cldat <- Cldat[ ,c(2,5,7)]
head(Cldat)
```

```
##          X.1  IC.Cl.1  IC.Cl.3
## 1  Lab Blank  0.0471   0.0531
## 2  Lab Blank  0.0472   0.0531
## 3 Standard 1  6.0990   6.8696
## 4      Blank  0.0480   0.0541
## 5 Standard 1  6.0838   6.8525
## 6 Standard 2 12.2230  13.7674
```

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

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

```
Cldat$Cl_Area <- as.numeric(Cldat$Cl_Area)
```

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

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

```
##    Sample_ID  Cl_ppm Cl_Area
## 1  Lab Blank  0.0471  0.0531
## 2  Lab Blank  0.0472  0.0531
## 3 Standard 1  6.0990  6.8696
## 4      Blank  0.0480  0.0541
## 5 Standard 1  6.0838  6.8525
## 6 Standard 2 12.2230 13.7674
```

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

```
##   Sample_ID   SO4_ppm SO4_Area     Cl_ppm  Cl_Area
## 1           2418.1928  17.1467 15120.9494 224.2274
## 2           2418.1928  17.1467  4224.4183  60.4754
## 3           2418.1928  17.1467     0.0212   0.0239
## 4           2418.1928  17.1467  4473.1255  50.9915
## 5           2418.1928  17.1467         NA       NA
## 6            586.9775   5.9183 15120.9494 224.2274
```

```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 SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
```

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

```
##                          Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
```

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 SO4_Area Cl_ppm Cl_Area
## 1424 Standard 1  0.5615   0.4668 6.1419  6.9180
## 1425 Standard 1  0.5615   0.4668 6.1360  6.9113
## 1426 Standard 1  0.5615   0.4668 6.0990  6.8696
## 1427 Standard 1  0.5615   0.4668 6.1169  6.8898
## 1428 Standard 1  0.5615   0.4668 6.0921  6.8619
## 1429 Standard 1  0.5615   0.4668 6.0838  6.8525
```

```
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  0.568 0.00410 0.722 YES
## 2 Standard 2  1.16  0.00688 0.595 YES
## 3 Standard 3  2.30  0.00792 0.345 YES
## 4 Standard 4 10.4   0.0278  0.268 YES
## 5 Standard 5 19.9   0.0475  0.239 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   6.11 0.0210 0.344 YES
## 2 Standard 2  12.2  0.0218 0.178 YES
## 3 Standard 3  23.6  0.0444 0.189 YES
## 4 Standard 4 104.   0.233  0.224 YES
## 5 Standard 5 199.   0.313  0.157 YES
```

```
lmS <- lm(stds$SO4_Area ~ stds$SO4_ppm)
SO4_sum <- summary(lmS)
SO4_Slope <- SO4_sum$coefficients[2, 1]
SO4_Int <- SO4_sum$coefficients[1, 1]


lmCl <- lm(stds$Cl_Area ~ stds$Cl_ppm)
Cl_sum <- summary(lmCl)
```

```
Cl_Slope <- Cl_sum$coefficients[2, 1]
Cl_Int <- Cl_sum$coefficients[1, 1]
```

Calculate mmol/L concentrations

```
#remove standards from sample dataframe
sampledat <- all_dat[!grepl("Standard", all_dat$Sample_ID),]
head(sampledat)
```

```
##                           Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
```

```
# 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 SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
##       SO4_mM      Cl_mM   salinity
## 26 0.1294167  0.4072976 0.02611673
## 27 0.8915721  1.1692750 0.07492760
## 28 5.1679538  1.0954866 0.07020084
## 29 0.9785652 18.9505134 1.21396083
## 30 6.3468715 37.8097320 2.42204748
## 31 2.0451497 26.3340367 1.68693567
```

Pull out dups and check with percent difference

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

```
##                           Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
##       SO4_mM      Cl_mM   salinity
## 26 0.1294167  0.4072976 0.02611673
## 27 0.8915721  1.1692750 0.07492760
## 28 5.1679538  1.0954866 0.07020084
## 29 0.9785652 18.9505134 1.21396083
## 30 6.3468715 37.8097320 2.42204748
## 31 2.0451497 26.3340367 1.68693567
```

```
#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 SO4_Area   Cl_ppm  Cl_Area
## 1      109_202210_MSM_PPR_UP_2_dup   57.3788   0.9540 1055.464  23.7765
## 2     119_202210_MSM_PPR_UP_12_dup  140.2629   2.3320 2355.659  53.0661
## 3     129_202210_MSM_PPR_TR_10_dup  466.9523   7.7634 3585.616  80.7734
## 4 19_202210_GCrew_WC_SipB_20cm_dup  436.3971   3.6273 5767.606  64.9571
## 5   29_202210_GWI_UP_LysA_45cm_dup  651.2403   5.4131 5494.783  61.8845
## 6   39_202210_GWI_TR_LysB_45cm_dup 1084.1925   9.0118 9592.738 108.0373
##       SO4_mM     Cl_mM  salinity
## 1  1.789732  29.77331  1.907249
## 2  4.375012  66.45018  4.256702
## 3 14.564950 101.14572  6.479233
## 4 13.611887 162.69693 10.422090
## 5 20.313172 155.00093  9.929099
## 6 33.817608 270.59911 17.334104
```

```
#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 SO4_Area    Cl_ppm Cl_Area    SO4_mM
## 1  1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253 0.1294167
## 2 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338 0.8915721
## 3   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336 0.9785652
## 4   101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300 2.0451497
```

```
## 5   102_202210_MSM_UP_RHZ_SF_3 105.7117   1.7575 1230.2357 27.7136 3.2973082
## 6   103_202210_MSM_UP_RHZ_SF_4  54.7246   0.9098 1060.9960 23.9011 1.7069432
##        Cl_mM   salinity
## 1  0.4072976 0.02611673
## 2  1.1692750 0.07492760
## 3 18.9505134 1.21396083
## 4 26.3340367 1.68693567
## 5 34.7034048 2.22306191
## 6 29.9293653 1.91724577
```

```
#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(2:5)]
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_202210_MSM_PPR_UP_2   1.789732  29.77331     1.907249
## 2     119_202210_MSM_PPR_UP_12   4.375012  66.45018     4.256702
## 3     129_202210_MSM_PPR_TR_10  14.564950 101.14572     6.479233
## 4 19_202210_GCrew_WC_SipB_20cm  13.611887 162.69693    10.422090
## 5   29_202210_GWI_UP_LysA_45cm  20.313172 155.00093     9.929099
## 6   39_202210_GWI_TR_LysB_45cm  33.817608 270.59911    17.334104
```

```
#put it back together with the old data set and look for duplicates
QAdups <- merge(sampledat2, dups)
head(QAdups)
```

```
##                      Sample_ID   SO4_ppm SO4_Area   Cl_ppm  Cl_Area    SO4_mM
## 1      109_202210_MSM_PPR_UP_2   56.9315   0.9465 1042.804  23.4913  1.775780
## 2     119_202210_MSM_PPR_UP_12  141.9209   2.3595 2376.423  53.5338  4.426728
## 3     129_202210_MSM_PPR_TR_10  462.9205   7.6964 3552.531  80.0281 14.439192
## 4 19_202210_GCrew_WC_SipB_20cm  441.4171   3.6691 5832.465  65.6876 13.768468
## 5   29_202210_GWI_UP_LysA_45cm  655.8744   5.4516 5534.610  62.3330 20.457717
## 6   39_202210_GWI_TR_LysB_45cm 1109.8531   9.2251 9834.080 110.7554 34.618001
##       Cl_mM  salinity SO4_mM_dup Cl_mM_dup salinity_dup
## 1  29.41620  1.884373   1.789732  29.77331     1.907249
## 2  67.03591  4.294222   4.375012  66.45018     4.256702
## 3 100.21243  6.419449  14.564950 101.14572     6.479233
## 4 164.52650 10.539290  13.611887 162.69693    10.422090
## 5 156.12440 10.001066  20.313172 155.00093     9.929099
## 6 277.40705 17.770208  33.817608 270.59911    17.334104
```

```
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)
```

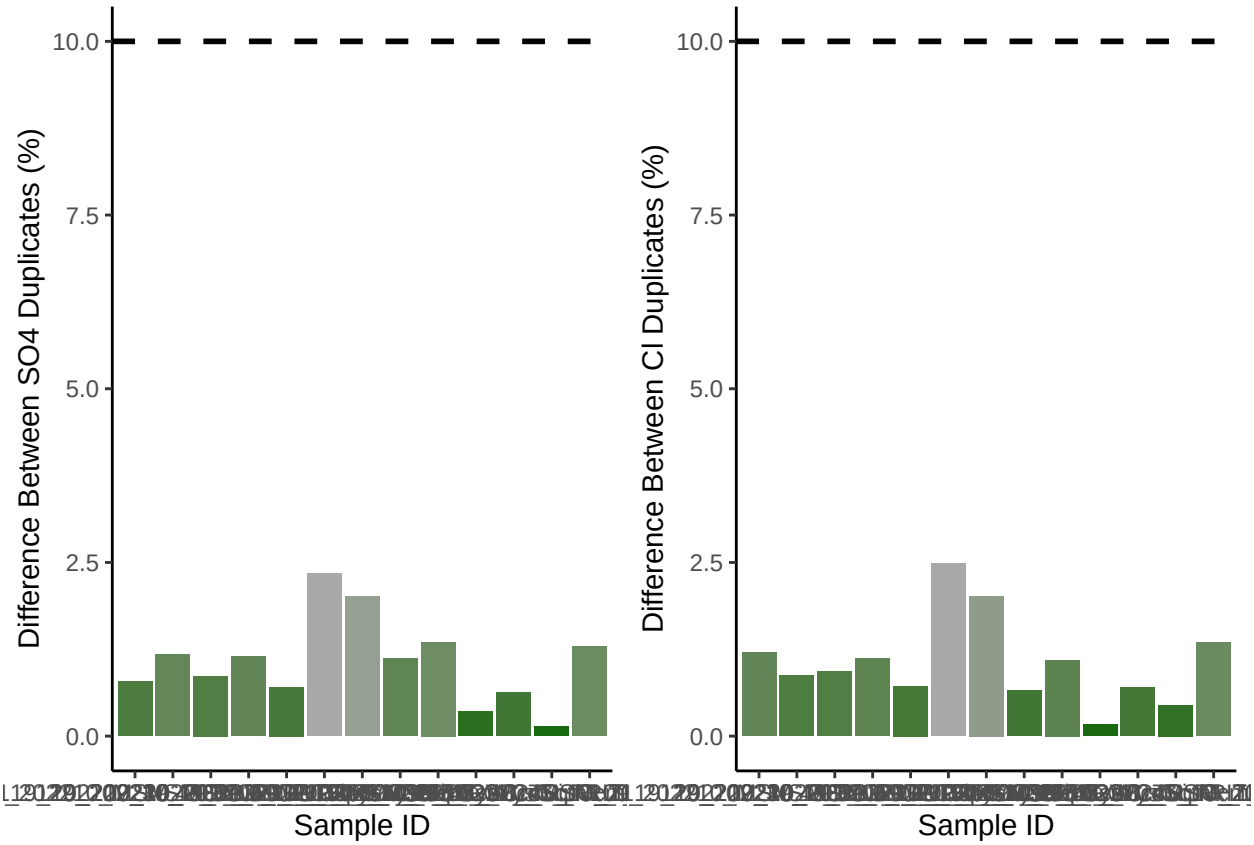
```
##                      Sample_ID   SO4_ppm SO4_Area   Cl_ppm  Cl_Area    SO4_mM
```

```
## 1      109_202210_MSM_PPR_UP_2   56.9315   0.9465 1042.804  23.4913  1.775780
## 2     119_202210_MSM_PPR_UP_12  141.9209   2.3595 2376.423  53.5338  4.426728
## 3     129_202210_MSM_PPR_TR_10  462.9205   7.6964 3552.531  80.0281 14.439192
## 4 19_202210_GCrew_WC_SipB_20cm  441.4171   3.6691 5832.465  65.6876 13.768468
## 5   29_202210_GWI_UP_LysA_45cm  655.8744   5.4516 5534.610  62.3330 20.457717
## 6   39_202210_GWI_TR_LysB_45cm 1109.8531   9.2251 9834.080 110.7554 34.618001
##       Cl_mM  salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk
## 1  29.41620  1.884373   1.789732  29.77331     1.907249    0.7826066
## 2  67.03591  4.294222   4.375012  66.45018     4.256702    1.1751206
## 3 100.21243  6.419449  14.564950 101.14572     6.479233    0.8671724
## 4 164.52650 10.539290  13.611887 162.69693    10.422090    1.1437500
## 5 156.12440 10.001066  20.313172 155.00093     9.929099    0.7090579
## 6 277.40705 17.770208  33.817608 270.59911    17.334104    2.3391127
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1           YES   1.2066713          YES
## 2           YES   0.8775842          YES
## 3           YES   0.9269914          YES
## 4           YES   1.1182432          YES
## 5           YES   0.7222011          YES
## 6           YES   2.4846220          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  YES        13  YES         13    13       100        100
```

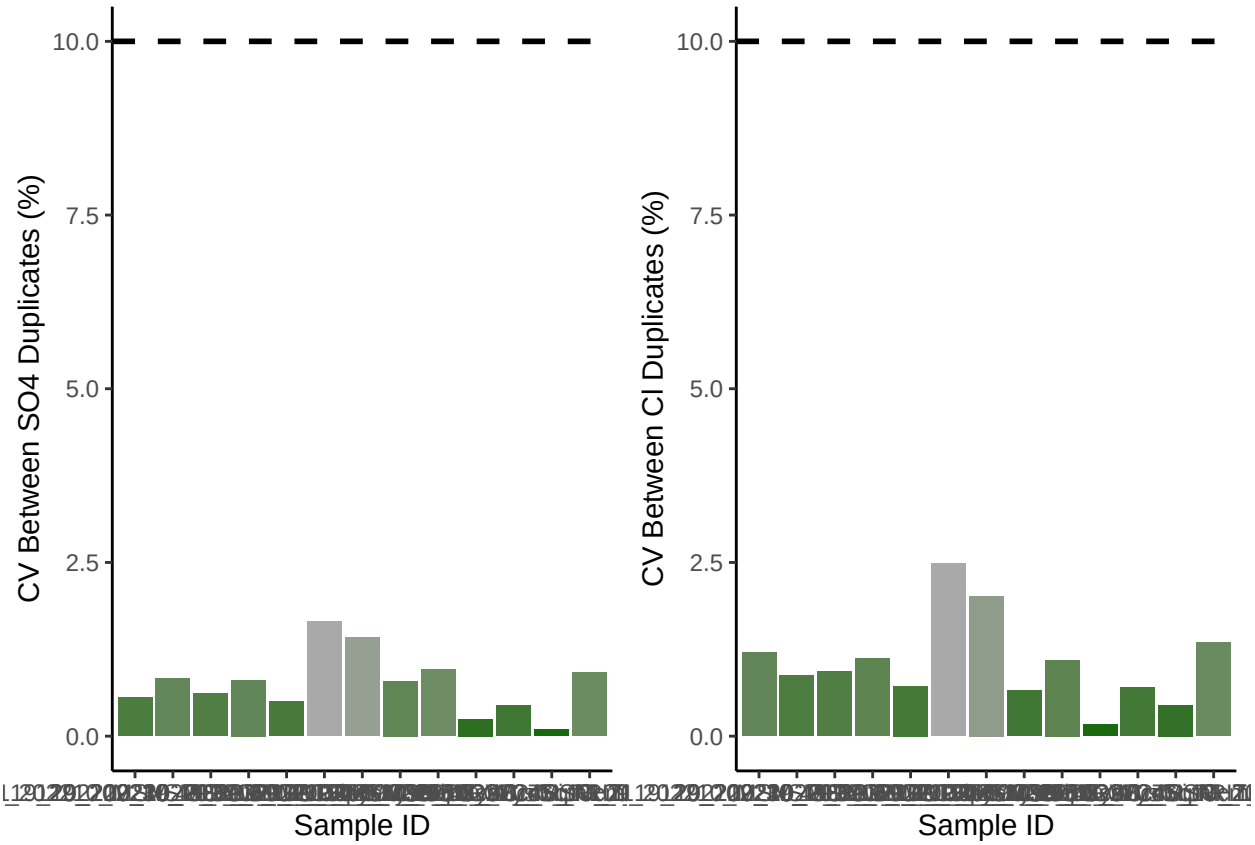
Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

```
##                      Sample_ID   SO4_ppm SO4_Area   Cl_ppm  Cl_Area    SO4_mM
## 1      109_202210_MSM_PPR_UP_2   56.9315   0.9465 1042.804  23.4913  1.775780
## 2     119_202210_MSM_PPR_UP_12  141.9209   2.3595 2376.423  53.5338  4.426728
## 3     129_202210_MSM_PPR_TR_10  462.9205   7.6964 3552.531  80.0281 14.439192
## 4 19_202210_GCrew_WC_SipB_20cm  441.4171   3.6691 5832.465  65.6876 13.768468
## 5   29_202210_GWI_UP_LysA_45cm  655.8744   5.4516 5534.610  62.3330 20.457717
## 6   39_202210_GWI_TR_LysB_45cm 1109.8531   9.2251 9834.080 110.7554 34.618001
##       Cl_mM  salinity SO4_mM_dup Cl_mM_dup salinity_dup SO4_dups_chk
## 1  29.41620  1.884373   1.789732  29.77331     1.907249    0.7826066
## 2  67.03591  4.294222   4.375012  66.45018     4.256702    1.1751206
## 3 100.21243  6.419449  14.564950 101.14572     6.479233    0.8671724
## 4 164.52650 10.539290  13.611887 162.69693    10.422090    1.1437500
## 5 156.12440 10.001066  20.313172 155.00093     9.929099    0.7090579
## 6 277.40705 17.770208  33.817608 270.59911    17.334104    2.3391127
##   SO4_dups_flag Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1           YES   1.2066713          YES   0.5533865              YES
## 2           YES   0.8775842          YES   0.8309358              YES
## 3           YES   0.9269914          YES   0.6131835              YES
## 4           YES   1.1182432          YES   0.8087534              YES
## 5           YES   0.7222011          YES   0.5013796              YES
## 6           YES   2.4846220          YES   1.6540025              YES
```

```
#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
Sdupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = SO4_dups_cv, fill=SO4_dups_chk)) +
       geom_bar(stat = 'identity') +
        scale_fill_gradient2(low='darkgreen', mid='darkgreen', high='darkgrey', space='Lab') +
        theme_classic() + labs(x= "Sample ID", y="CV Between SO4 Duplicates (%)") +
        theme(legend.position="none") +  geom_hline(yintercept=10, linetype="dashed",
                color = "black", size=1)


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

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




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

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

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

Pull out spikes and check with dionex output conc.

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

```
##                           Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
##       SO4_mM      Cl_mM   salinity
## 26 0.1294167  0.4072976 0.02611673
## 27 0.8915721  1.1692750 0.07492760
## 28 5.1679538  1.0954866 0.07020084
## 29 0.9785652 18.9505134 1.21396083
## 30 6.3468715 37.8097320 2.42204748
## 31 2.0451497 26.3340367 1.68693567
```

```
#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 SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.835  0.8748  5.167954
## 2   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.355 30.1943  6.346871
## 3      110_202210_MSM_PPR_UP_3_spk 206.8264   3.4386 1094.155 24.6481  6.451229
## 4      120_202210_MSM_PPR_TR_1_spk 374.0176   6.2183 2572.971 57.9615 11.666176
## 5     130_202210_MSM_PPR_TR_11_spk 577.9479   9.6088 3459.534 77.9332 18.027071
## 6 20_202210_GCrew_WC_SipB_45cm_spk 698.3096   5.8043 5526.222 62.2386 21.781335
##       Cl_mM   salinity
## 1  1.095487 0.07020084
## 2 37.809732 2.42204748
## 3 30.864725 1.97716318
## 4 72.580279 4.64938442
## 5 97.589109 6.25140376
## 6 155.887797 9.98590988
```

```
#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(2:5,7,8)]
colnames(spks) <- c('Sample_ID', 'SO4_mM_spk')
head(spks)
```

```
##                       Sample_ID SO4_mM_spk
## 1 10_202210_GCrew_TR_LysB_20cm   5.167954
## 2   100_202210_MSM_UP_RHZ_SF_1   6.346871
## 3      110_202210_MSM_PPR_UP_3   6.451229
## 4      120_202210_MSM_PPR_TR_1  11.666176
## 5     130_202210_MSM_PPR_TR_11  18.027071
## 6 20_202210_GCrew_WC_SipB_45cm  21.781335
```

```
#put it back together with the old data set and look for duplicates
QAspks <- merge(sampledat, spks)
head(QAspks)
```

```
##                      Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity SO4_mM_spk
## 1   1.169275  0.0749276   5.167954
## 2  18.950513  1.2139608   6.346871
## 3  31.166155  1.9964722   6.451229
## 4  71.607430  4.5870655  11.666176
## 5  97.454429  6.2427764  18.027071
## 6 158.721315 10.1674198  21.781335
```

```
#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 SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity SO4_mM_spk SO4_spk_Conc
## 1   1.169275  0.0749276   5.167954 7.797879e-05
## 2  18.950513  1.2139608   6.346871 7.797879e-05
## 3  31.166155  1.9964722   6.451229 7.797879e-05
## 4  71.607430  4.5870655  11.666176 7.797879e-05
## 5  97.454429  6.2427764  18.027071 7.797879e-05
## 6 158.721315 10.1674198  21.781335 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)
```

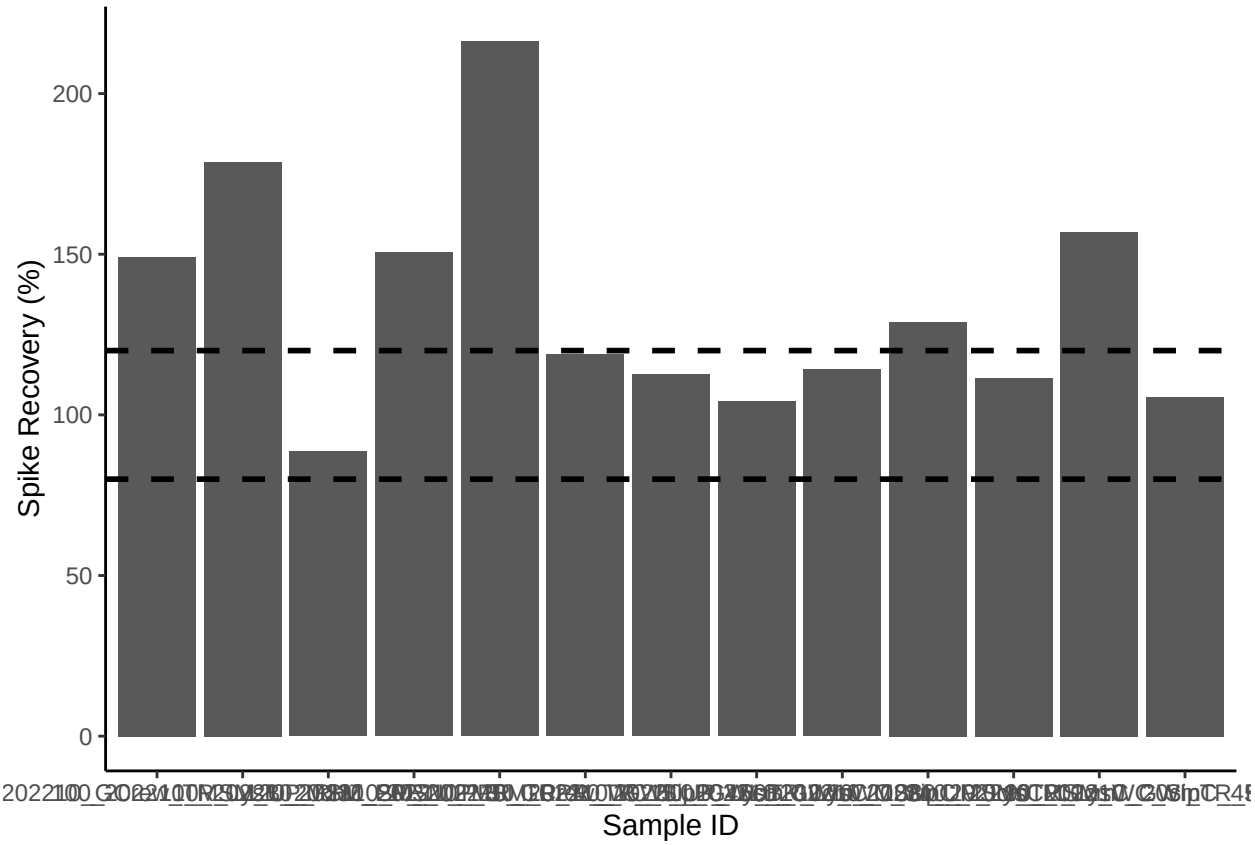
```
##                      Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1   1.169275  0.0749276   5.167954 7.797879e-05       50  0.001501
## 2  18.950513  1.2139608   6.346871 7.797879e-05       50  0.001501
## 3  31.166155  1.9964722   6.451229 7.797879e-05        1  0.000001
## 4  71.607430  4.5870655  11.666176 7.797879e-05        1  0.000001
## 5  97.454429  6.2427764  18.027071 7.797879e-05        1  0.000001
## 6 158.721315 10.1674198  21.781335 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 SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
```

```
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1   1.169275  0.0749276   5.167954 7.797879e-05       50  0.001501
## 2  18.950513  1.2139608   6.346871 7.797879e-05       50  0.001501
## 3  31.166155  1.9964722   6.451229 7.797879e-05        1  0.000001
## 4  71.607430  4.5870655  11.666176 7.797879e-05        1  0.000001
## 5  97.454429  6.2427764  18.027071 7.797879e-05        1  0.000001
## 6 158.721315 10.1674198  21.781335 7.797879e-05      100  0.001475
##   SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1     2.676499e-05   1.561756e-04    1.047438e-04    149.10247     NO, rerun
## 2     2.937653e-05   1.918025e-04    1.073553e-04    178.66135     NO, rerun
## 3     2.188391e-06   7.096352e-05    8.016718e-05     88.51941           YES
## 4     7.310031e-06   1.283279e-04    8.528882e-05    150.46278     NO, rerun
## 5     1.371568e-05   1.982978e-04    9.169447e-05    216.25927     NO, rerun
## 6     1.940789e-04   3.234528e-04    2.720577e-04    118.89126           YES
```

```
#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           6    13    46.2
## 2 YES                 7    13    53.8
```

Pull out spikes and check with area calc

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

```
##                            Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
```

```
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
##       SO4_mM      Cl_mM   salinity
## 26 0.1294167  0.4072976 0.02611673
## 27 0.8915721  1.1692750 0.07492760
## 28 5.1679538  1.0954866 0.07020084
## 29 0.9785652 18.9505134 1.21396083
## 30 6.3468715 37.8097320 2.42204748
## 31 2.0451497 26.3340367 1.68693567
```

```
#now use area of all samples and calculate undilution corrected concentrations in ug/mL
sampledat$SO4_ugmL <- ((sampledat$SO4_Area)-SO4_Int)/SO4_Slope
sampledat$Cl_ugmL <- (sampledat$Cl_Area-Cl_Int)/Cl_Slope
head(sampledat)
```

```
##                           Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area
## 26      1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253
## 27     10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338
## 28 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.8350  0.8748
## 29       100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336
## 30   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.3550 30.1943
## 31       101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300
##       SO4_mM      Cl_mM   salinity   SO4_ugmL    Cl_ugmL
## 26 0.1294167  0.4072976 0.02611673 0.08300035  0.2888063
## 27 0.8915721  1.1692750 0.07492760 0.57164426  0.8290452
## 28 5.1679538  1.0954866 0.07020084 3.31368003  0.7766638
## 29 0.9785652 18.9505134 1.21396083 0.62746178 13.4359223
## 30 6.3468715 37.8097320 2.42204748 4.06962251 26.8071243
## 31 2.0451497 26.3340367 1.68693567 1.31134672 18.6708686
```

```
#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 SO4_Area   Cl_ppm Cl_Area    SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm_spk 165.6846   2.7546   38.835  0.8748  5.167954
## 2   100_202210_MSM_UP_RHZ_SF_1_spk 203.4807   3.3830 1340.355 30.1943  6.346871
## 3      110_202210_MSM_PPR_UP_3_spk 206.8264   3.4386 1094.155 24.6481  6.451229
## 4      120_202210_MSM_PPR_TR_1_spk 374.0176   6.2183 2572.971 57.9615 11.666176
## 5     130_202210_MSM_PPR_TR_11_spk 577.9479   9.6088 3459.534 77.9332 18.027071
## 6 20_202210_GCrew_WC_SipB_45cm_spk 698.3096   5.8043 5526.222 62.2386 21.781335
##        Cl_mM   salinity  SO4_ugmL    Cl_ugmL
## 1   1.095487 0.07020084  3.313680  0.7766638
## 2  37.809732 2.42204748  4.069623 26.8071243
## 3  30.864725 1.97716318  4.136507 21.8830928
## 4  72.580279 4.64938442  7.480386 51.4594206
## 5  97.589109 6.25140376 11.559035 69.1907103
## 6 155.887797 9.98590988  6.982359 55.2567187
```

```
## Only keep the columns that we need
spks <- spks[ ,c(1,9,10)] # dont need this here
head(spks)
```

```
##                          Sample_ID  SO4_ugmL    Cl_ugmL
## 1 10_202210_GCrew_TR_LysB_20cm_spk  3.313680  0.7766638
## 2   100_202210_MSM_UP_RHZ_SF_1_spk  4.069623 26.8071243
## 3      110_202210_MSM_PPR_UP_3_spk  4.136507 21.8830928
## 4      120_202210_MSM_PPR_TR_1_spk  7.480386 51.4594206
## 5     130_202210_MSM_PPR_TR_11_spk 11.559035 69.1907103
## 6 20_202210_GCrew_WC_SipB_45cm_spk  6.982359 55.2567187
```

```
#remove the spk from these IDs so we will have duplicate sample names
spks$Sample_ID<-gsub("_spk","",as.character(spks$Sample_ID))
spks <- spks[ ,c(1,2)]
colnames(spks) <- c('Sample_ID', 'SO4_ugmL_spk')
head(spks)
```

```
##                      Sample_ID SO4_ugmL_spk
## 1 10_202210_GCrew_TR_LysB_20cm     3.313680
## 2   100_202210_MSM_UP_RHZ_SF_1     4.069623
## 3      110_202210_MSM_PPR_UP_3     4.136507
## 4      120_202210_MSM_PPR_TR_1     7.480386
## 5     130_202210_MSM_PPR_TR_11    11.559035
## 6 20_202210_GCrew_WC_SipB_45cm     6.982359
```

```
#put it back together with the old data set and look for duplicates
QAspks <- merge(sampledat, spks)
head(QAspks)
```

```
##                      Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity  SO4_ugmL    Cl_ugmL SO4_ugmL_spk
## 1   1.169275  0.0749276 0.5716443  0.8290452     3.313680
## 2  18.950513  1.2139608 0.6274618 13.4359223     4.069623
## 3  31.166155  1.9964722 1.4032532 22.0967913     4.136507
## 4  71.607430  4.5870655 4.6872241 50.7696727     7.480386
## 5  97.454429  6.2427764 8.7945036 69.0951807    11.559035
## 6 158.721315 10.1674198 4.2179479 56.2611103     6.982359
```

```
#now we need to calculate the spike concentration and calculate the spike recovery
spkconc <- (250)    # in ug
spkvol <- 10             # in uL
spkvol <- spkvol/1000
#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 SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity  SO4_ugmL    Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1   1.169275  0.0749276 0.5716443  0.8290452     3.313680          2.5
## 2  18.950513  1.2139608 0.6274618 13.4359223     4.069623          2.5
## 3  31.166155  1.9964722 1.4032532 22.0967913     4.136507          2.5
## 4  71.607430  4.5870655 4.6872241 50.7696727     7.480386          2.5
## 5  97.454429  6.2427764 8.7945036 69.0951807    11.559035          2.5
## 6 158.721315 10.1674198 4.2179479 56.2611103     6.982359          2.5
```

```
#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)
#$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 50, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 100, QAspks$Dilution)
#$Dilution <-  ifelse(str_detect(QAspks$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)
#$Dilution <-  ifelse(str_detect(QAspks$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, "MSM_PPR_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_PPR_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/1000
head(QAspks)
```

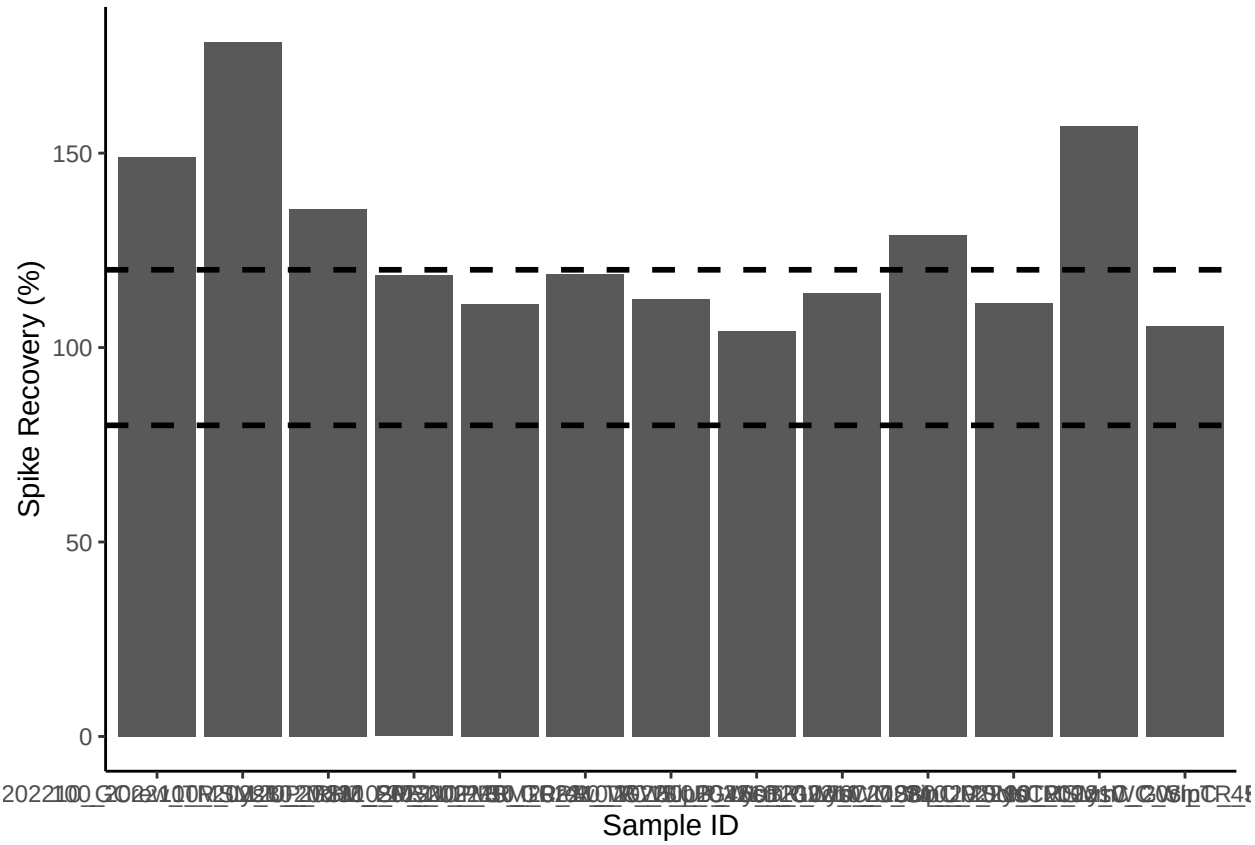
```
##                      Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity  SO4_ugmL    Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1   1.169275  0.0749276 0.5716443  0.8290452     3.313680          2.5
## 2  18.950513  1.2139608 0.6274618 13.4359223     4.069623          2.5
## 3  31.166155  1.9964722 1.4032532 22.0967913     4.136507          2.5
## 4  71.607430  4.5870655 4.6872241 50.7696727     7.480386          2.5
## 5  97.454429  6.2427764 8.7945036 69.0951807    11.559035          2.5
## 6 158.721315 10.1674198 4.2179479 56.2611103     6.982359          2.5
##   SampleVol
## 1     1.501
## 2     1.501
## 3     1.501
## 4     1.501
## 5     1.501
## 6     1.475
```

```
#gives us the total SO4 in the sample in mmoles
QAspks$SO4_Total_unspkd <- QAspks$SO4_ugmL*QAspks$SampleVol

##total SO4 in spiked sample in mmoles
QAspks$SO4_Total_spkd <- (QAspks$SO4_ugmL_spk)*(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 SO4_Area    Cl_ppm Cl_Area     SO4_mM
## 1 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338  0.8915721
## 2   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336  0.9785652
## 3      110_202210_MSM_PPR_UP_3  70.1598   1.1665 1104.8402 24.8888  2.1883905
## 4      120_202210_MSM_PPR_TR_1 234.3596   3.8964 2538.4834 57.1846  7.3100312
## 5     130_202210_MSM_PPR_TR_11 439.7246   7.3107 3454.7595 77.8256 13.7156769
## 6 20_202210_GCrew_WC_SipB_45cm 421.8420   3.5063 5626.6706 63.3699 13.1578915
##        Cl_mM   salinity  SO4_ugmL    Cl_ugmL SO4_ugmL_spk SO4_spk_Conc
## 1   1.169275  0.0749276 0.5716443  0.8290452     3.313680          2.5
## 2  18.950513  1.2139608 0.6274618 13.4359223     4.069623          2.5
## 3  31.166155  1.9964722 1.4032532 22.0967913     4.136507          2.5
## 4  71.607430  4.5870655 4.6872241 50.7696727     7.480386          2.5
## 5  97.454429  6.2427764 8.7945036 69.0951807    11.559035          2.5
## 6 158.721315 10.1674198 4.2179479 56.2611103     6.982359          2.5
##   SampleVol SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery
## 1     1.501        0.8580380       5.006971        3.358038     149.1040
## 2     1.501        0.9418201       6.149200        3.441820     178.6613
## 3     1.501        2.1062830       6.250263        4.606283     135.6899
## 4     1.501        7.0355234      11.302863        9.535523     118.5343
## 5     1.501       13.2005499      17.465702       15.700550     111.2426
```

```
## 6     1.475        6.2214732      10.368803        8.721473     118.8882
##   SO4_spks_flag
## 1     NO, rerun
## 2     NO, rerun
## 3     NO, rerun
## 4           YES
## 5           YES
## 6           YES
```

```
#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           5    13    38.5
## 2 YES                 8    13    61.5
```

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", "202210", "Gcrew", "UP", "LysA", "20cm"), c("10", :
## number of columns of result is not a multiple of vector length (arg 1)
```

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

```
##   Analysis_No   Date  Site Zone Replicate Depth Tree
## 1           1 202210 Gcrew   UP      LysA  20cm    1
## 2          10 202210 GCrew   TR      LysB  20cm   10
## 3         100 202210   MSM   UP       RHZ    SF    1
## 4         101 202210   MSM   UP       RHZ    SF    2
## 5         102 202210   MSM   UP       RHZ    SF    3
## 6         103 202210   MSM   UP       RHZ    SF    4
```

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

```
##   Analysis_No   Date  Site Zone Replicate Depth Tree
## 1           1 202210 Gcrew   UP      LysA  20cm    1
## 2          10 202210 GCrew   TR      LysB  20cm   10
## 3         100 202210   MSM   UP       RHZ    SF    1
## 4         101 202210   MSM   UP       RHZ    SF    2
## 5         102 202210   MSM   UP       RHZ    SF    3
## 6         103 202210   MSM   UP       RHZ    SF    4
##                      Sample_ID  SO4_ppm SO4_Area    Cl_ppm Cl_Area    SO4_mM
## 1  1_202210_Gcrew_UP_LysA_20cm   4.1491   0.0690   14.4387  0.3253 0.1294167
## 2 10_202210_GCrew_TR_LysB_20cm  28.5838   0.4752   41.4508  0.9338 0.8915721
## 3   100_202210_MSM_UP_RHZ_SF_1  31.3728   0.5216  671.7957 15.1336 0.9785652
## 4   101_202210_MSM_UP_RHZ_SF_2  65.5675   1.0901  933.5416 21.0300 2.0451497
## 5   102_202210_MSM_UP_RHZ_SF_3 105.7117   1.7575 1230.2357 27.7136 3.2973082
## 6   103_202210_MSM_UP_RHZ_SF_4  54.7246   0.9098 1060.9960 23.9011 1.7069432
##        Cl_mM   salinity
## 1  0.4072976 0.02611673
```

```
## 2  1.1692750 0.07492760
## 3 18.9505134 1.21396083
## 4 26.3340367 1.68693567
## 5 34.7034048 2.22306191
## 6 29.9293653 1.91724577
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202210.csv") #Change file
```

END