Dionex_COMPASS_June2022

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_202206_SO4.txt",sep='\t' , header=T, skip=3)
head(Sdat)
```

```
##   X        X.1                  X.2 IC.SO4 IC.SO4.1 IC.SO4.2 IC.SO4.3 IC.SO4.4
## 1 1  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 2 2  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 3 3  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 4 4  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 5 5  Lab Blank              Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 6 6 Standard 1 Calibration Standard 4.287   0.0820     9.52   0.0734     0.91
```

```
## 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  Lab Blank     n.a.     n.a.
## 4  Lab Blank     n.a.     n.a.
## 5  Lab Blank     n.a.     n.a.
## 6 Standard 1  0.0820   0.0734
```

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

```
##    Sample_ID SO4_ppm SO4_Area
## 1  Lab Blank      NA     n.a.
## 2  Lab Blank      NA     n.a.
## 3  Lab Blank      NA     n.a.
## 4  Lab Blank      NA     n.a.
## 5  Lab Blank      NA     n.a.
## 6 Standard 1   0.082  0.0734
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202206_Cl.txt",sep='\t' , header=T, skip=3)
head(Cldat)
```

```
##   X        X.1                 X.2  IC.Cl IC.Cl.1 IC.Cl.2 IC.Cl.3 IC.Cl.4
## 1 1  Lab Blank             Unknown   n.a.    n.a.    n.a.    n.a.    n.a.
## 2 2  Lab Blank             Unknown 3.237  0.0092   25.58  0.0117    0.03
## 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  Lab Blank             Unknown   n.a.    n.a.    n.a.    n.a.    n.a.
## 6 6 Standard 1 Calibration Standard   n.a.    n.a.    n.a.    n.a.    n.a.
```

```
## 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    n.a.    n.a.
## 2  Lab Blank 0.0092  0.0117
## 3  Lab Blank    n.a.    n.a.
## 4  Lab Blank    n.a.    n.a.
## 5  Lab Blank    n.a.    n.a.
## 6 Standard 1    n.a.    n.a.
```

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

```
##    Sample_ID Cl_ppm Cl_Area
## 1  Lab Blank     NA    n.a.
## 2  Lab Blank 0.0092 0.0117
## 3  Lab Blank     NA    n.a.
## 4  Lab Blank     NA    n.a.
## 5  Lab Blank     NA    n.a.
## 6 Standard 1     NA    n.a.
```

```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           1858.5917 20.0924  14717.2535 262.5010
## 2           1858.5917 20.0924   3732.9877  41.0622
## 3           1858.5917 20.0924      0.0079   0.0101
## 4           1858.5917 20.0924   4029.0080  45.6129
## 5           1858.5917 20.0924          NA   111.08%
## 6            447.9436  3.3915  14717.2535 262.5010
```

```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_MSM_UP_LysA_10cm 126.0804  1.3277  1568.675 23.5520
## 27     10_MSM_TR_LysA_10cm 331.4201  3.4901  2755.854 41.3762
## 28     11_MSM_TR_LysB_10cm 367.1323  3.8662  3027.198 45.4502
## 29  12_MSM_TR_LysC_10c_spk 484.6522  5.1038  2789.309 41.8785
## 30     12_MSM_TR_LysC_10cm 314.1841  3.3086  2868.593 43.0689
## 31 12_MSM_TR_LysC_10cm_dup 305.1959  3.2139  2796.526 41.9869
```

```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_MSM_UP_LysA_10cm 126.0804  1.3277  1568.675 23.5520
## 27     10_MSM_TR_LysA_10cm 331.4201  3.4901  2755.854 41.3762
## 28     11_MSM_TR_LysB_10cm 367.1323  3.8662  3027.198 45.4502
## 29  12_MSM_TR_LysC_10c_spk 484.6522  5.1038  2789.309 41.8785
## 30     12_MSM_TR_LysC_10cm 314.1841  3.3086  2868.593 43.0689
## 31 12_MSM_TR_LysC_10cm_dup 305.1959  3.2139  2796.526 41.9869
```

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 SO4_Area Cl_ppm Cl_Area
## 981 Standard 1  0.4149  0.3716  5.1137 6.5287
## 982 Standard 1  0.4149  0.3716      NA    n.a.
## 983 Standard 1  0.4149  0.3716      NA    n.a.
## 984 Standard 1  0.4149  0.3716      NA    n.a.
## 985 Standard 1  0.4149  0.3716  4.9890 6.3695
## 986 Standard 1  0.0679  0.0608  5.1137 6.5287
```

```
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  2.72   5.09   187.   NO, rerun
## 2 Standard 2  0.759  0.0925  12.2  NO, rerun
## 3 Standard 3  1.86   0.105    5.65 NO, rerun
## 4 Standard 4 NA     NA       NA    <NA>
## 5 Standard 5 NA     NA       NA    <NA>
```

```
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  NA   NA     NA    <NA>
## 2 Standard 2  10.1  0.252  2.51 NO, rerun
## 3 Standard 3  20.3  0.350  1.72 YES
## 4 Standard 4  NA   NA     NA    <NA>
## 5 Standard 5  NA   NA     NA    <NA>
```

```
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_MSM_UP_LysA_10cm 126.0804  1.3277  1568.675 23.5520
## 27     10_MSM_TR_LysA_10cm 331.4201  3.4901  2755.854 41.3762
## 28     11_MSM_TR_LysB_10cm 367.1323  3.8662  3027.198 45.4502
## 29  12_MSM_TR_LysC_10c_spk 484.6522  5.1038  2789.309 41.8785
## 30     12_MSM_TR_LysC_10cm 314.1841  3.3086  2868.593 43.0689
## 31 12_MSM_TR_LysC_10cm_dup 305.1959  3.2139  2796.526 41.9869
```

```
# 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    SO4_mM
## 26      1_MSM_UP_LysA_10cm 126.0804  1.3277  1568.675 23.5520   3.932639
## 27     10_MSM_TR_LysA_10cm 331.4201  3.4901  2755.854 41.3762  10.337495
## 28     11_MSM_TR_LysB_10cm 367.1323  3.8662  3027.198 45.4502  11.451413
## 29  12_MSM_TR_LysC_10c_spk 484.6522  5.1038  2789.309 41.8785  15.117037
## 30     12_MSM_TR_LysC_10cm 314.1841  3.3086  2868.593 43.0689   9.799878
## 31 12_MSM_TR_LysC_10cm_dup 305.1959  3.2139  2796.526 41.9869   9.519523
##       Cl_mM salinity
## 26 44.25036 2.834622
## 27 77.73917 4.979853
## 28 85.39347 5.470174
## 29 78.68290 5.040307
## 30 80.91941 5.183574
## 31 78.88649 5.053348
```

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    SO4_mM
## 26      1_MSM_UP_LysA_10cm 126.0804  1.3277  1568.675 23.5520   3.932639
## 27     10_MSM_TR_LysA_10cm 331.4201  3.4901  2755.854 41.3762  10.337495
```

```
## 28     11_MSM_TR_LysB_10cm 367.1323  3.8662  3027.198 45.4502  11.451413
## 29  12_MSM_TR_LysC_10c_spk 484.6522  5.1038  2789.309 41.8785  15.117037
## 30     12_MSM_TR_LysC_10cm 314.1841  3.3086  2868.593 43.0689   9.799878
## 31 12_MSM_TR_LysC_10cm_dup 305.1959  3.2139  2796.526 41.9869   9.519523
##       Cl_mM salinity
## 26 44.25036 2.834622
## 27 77.73917 4.979853
## 28 85.39347 5.470174
## 29 78.68290 5.040307
## 30 80.91941 5.183574
## 31 78.88649 5.053348
```

```
#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    SO4_mM
## 1 12_MSM_TR_LysC_10cm_dup  305.1959  3.2139   2796.526  41.9869   9.519523
## 2 15_MSM_TR_LysC_20cm_dup  304.1713  3.2032   2874.954  43.1644   9.487564
## 3 18_MSM_TR_LysC_45cm_dup  462.4204  4.8696   3100.760  46.5546  14.423593
## 4 21_MSM_WC_SipC_10cm_dup 1600.7862  8.4288  13348.587 100.2075  49.930948
## 5 24_MSM_WC_SipC_20cm_dup 1398.3514  7.3629  11907.618  89.3902  43.616700
## 6 27_MSM_WC_SipC_45cm_dup  453.3821  2.3872  14650.695 109.9824  14.141675
##       Cl_mM  salinity
## 1  78.88649  5.053348
## 2  81.09884  5.195068
## 3  87.46854  5.603099
## 4 376.54689 24.120923
## 5 335.89895 21.517092
## 6 413.27771 26.473832
```

```
#remove these from sample dataframe in a new dataframe
sampledat2 <- sampledat %>%
  filter(!str_detect(Sample_ID, "dup")) %>%
  filter(!str_detect(Sample_ID, "spk"))
sampledat2 <- sampledat2[ ,-c(2:4)]
head(sampledat2)
```

```
##             Sample_ID  Cl_Area    SO4_mM    Cl_mM salinity
## 1  1_MSM_UP_LysA_10cm 23.5520   3.932639 44.25036 2.834622
## 2 10_MSM_TR_LysA_10cm 41.3762  10.337495 77.73917 4.979853
## 3 11_MSM_TR_LysB_10cm 45.4502  11.451413 85.39347 5.470174
## 4 12_MSM_TR_LysC_10cm 43.0689   9.799878 80.91941 5.183574
## 5 13_MSM_TR_LysA_20cm 42.3007  13.084039 79.47600 5.091112
## 6 14_MSM_TR_LysB_20cm 44.3281  10.877742 83.28524 5.335124
```

```
#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 12_MSM_TR_LysC_10cm   9.519523  78.88649     5.053348
## 2 15_MSM_TR_LysC_20cm   9.487564  81.09884     5.195068
## 3 18_MSM_TR_LysC_45cm  14.423593  87.46854     5.603099
## 4 21_MSM_WC_SipC_10cm  49.930948 376.54689    24.120923
## 5 24_MSM_WC_SipC_20cm  43.616700 335.89895    21.517092
## 6 27_MSM_WC_SipC_45cm  14.141675 413.27771    26.473832
```

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

```
##             Sample_ID   Cl_Area    SO4_mM     Cl_mM  salinity SO4_mM_dup
## 1 12_MSM_TR_LysC_10cm  43.0689   9.799878  80.91941  5.183574   9.519523
## 2 15_MSM_TR_LysC_20cm  43.8109   9.625306  82.31358  5.272881   9.487564
## 3 18_MSM_TR_LysC_45cm  45.4714  14.067405  85.43324  5.472722  14.423593
## 4 21_MSM_WC_SipC_10cm  98.0269  48.735540 368.35315 23.596047  49.930948
## 5 24_MSM_WC_SipC_20cm  87.4088  42.449308 328.45362 21.040157  43.616700
## 6 27_MSM_WC_SipC_45cm 110.4820  14.149211 415.15525 26.594103  14.141675
##   Cl_mM_dup salinity_dup
## 1  78.88649     5.053348
## 2  81.09884     5.195068
## 3  87.46854     5.603099
## 4 376.54689    24.120923
## 5 335.89895    21.517092
## 6 413.27771    26.473832
```

```
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)
```

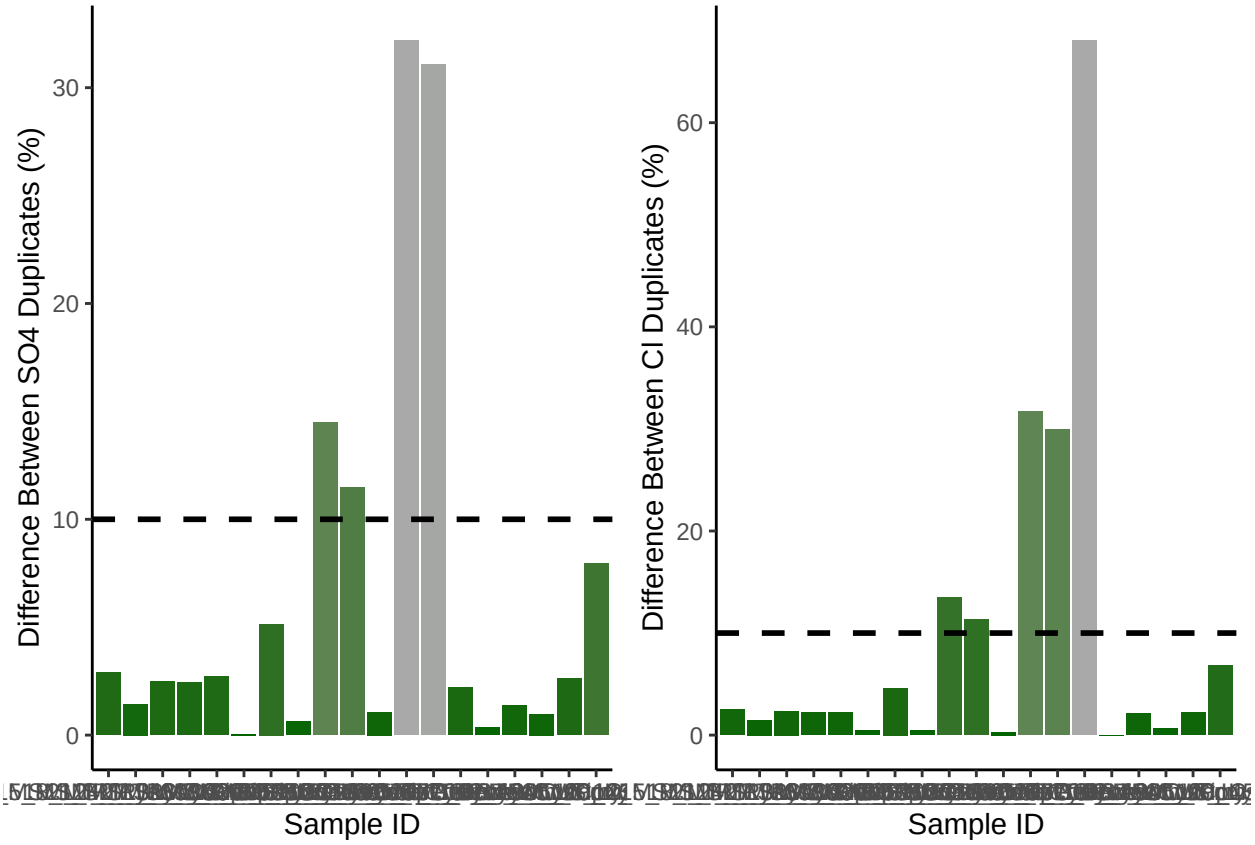
```
##             Sample_ID   Cl_Area    SO4_mM     Cl_mM  salinity SO4_mM_dup
## 1 12_MSM_TR_LysC_10cm  43.0689   9.799878  80.91941  5.183574   9.519523
## 2 15_MSM_TR_LysC_20cm  43.8109   9.625306  82.31358  5.272881   9.487564
## 3 18_MSM_TR_LysC_45cm  45.4714  14.067405  85.43324  5.472722  14.423593
## 4 21_MSM_WC_SipC_10cm  98.0269  48.735540 368.35315 23.596047  49.930948
## 5 24_MSM_WC_SipC_20cm  87.4088  42.449308 328.45362 21.040157  43.616700
## 6 27_MSM_WC_SipC_45cm 110.4820  14.149211 415.15525 26.594103  14.141675
##   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1  78.88649     5.053348   2.90232168           YES   2.5442466          YES
## 2  81.09884     5.195068   1.44135064           YES   1.4867088          YES
## 3  87.46854     5.603099   2.50035745           YES   2.3542785          YES
## 4 376.54689    24.120923   2.42312995           YES   2.1999586          YES
## 5 335.89895    21.517092   2.71278388           YES   2.2413796          YES
## 6 413.27771    26.473832   0.05327419           YES   0.4532744          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         4 NO, rerun          5    19  21.05263   26.31579
## 2       YES        15       YES         14    19  78.94737   73.68421
```

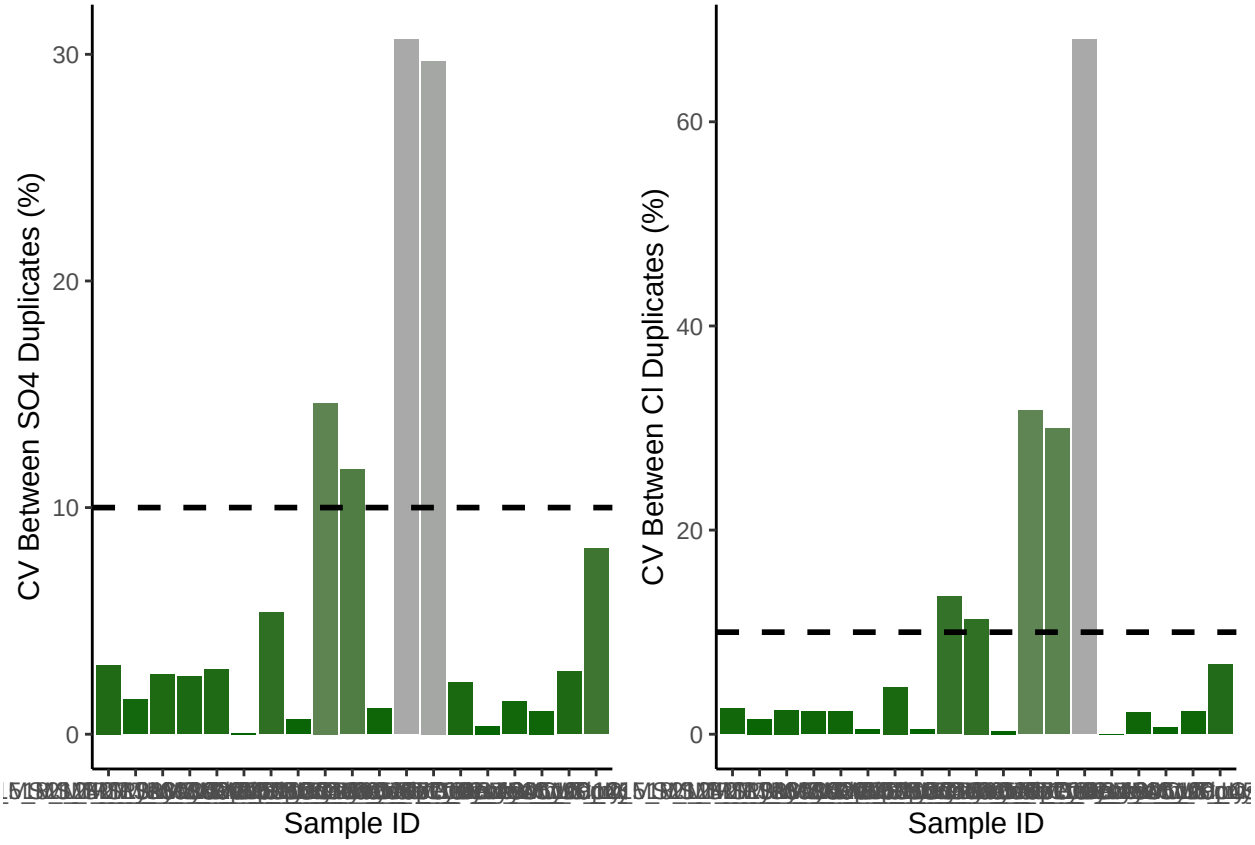
Pull out dups and check with cv

```
#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$sds <- apply(df2,1,sd)
df2$mean <- apply(df2, 1, mean)

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

head(QAdups)
```

```
##             Sample_ID   Cl_Area    SO4_mM     Cl_mM  salinity SO4_mM_dup
## 1 12_MSM_TR_LysC_10cm  43.0689   9.799878  80.91941  5.183574   9.519523
## 2 15_MSM_TR_LysC_20cm  43.8109   9.625306  82.31358  5.272881   9.487564
## 3 18_MSM_TR_LysC_45cm  45.4714  14.067405  85.43324  5.472722  14.423593
## 4 21_MSM_WC_SipC_10cm  98.0269  48.735540 368.35315 23.596047  49.930948
## 5 24_MSM_WC_SipC_20cm  87.4088  42.449308 328.45362 21.040157  43.616700
## 6 27_MSM_WC_SipC_45cm 110.4820  14.149211 415.15525 26.594103  14.141675
##   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag Cl_dups_chk Cl_dups_flag
## 1  78.88649     5.053348   2.90232168           YES   2.5442466          YES
## 2  81.09884     5.195068   1.44135064           YES   1.4867088          YES
## 3  87.46854     5.603099   2.50035745           YES   2.3542785          YES
## 4 376.54689    24.120923   2.42312995           YES   2.1999586          YES
## 5 335.89895    21.517092   2.71278388           YES   2.2413796          YES
## 6 413.27771    26.473832   0.05327419           YES   0.4532744          YES
##   SO4_dups_cv SO4_dups_cv_flag
## 1  3.04710983              YES
## 2  1.52103212              YES
## 3  2.62879079              YES
```

```
## 4  2.54828611              YES
## 5  2.85000700              YES
## 6  0.05649517              YES
```

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


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

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




```
#check for percent of no, reruns to see if it  would warrant reruns
Perc_dups <- QAdups %>%
  group_by(SO4_dups_flag) %>%
```

```
  summarise(S_no_rows = length(SO4_dups_flag))
Perc_dups1 <- QAdups %>%
  group_by(Cl_dups_flag) %>%
  summarise(Cl_no_rows = length(Cl_dups_flag))
colnames(Perc_dups) <- c("Flag", "S_no_rows")
colnames(Perc_dups1) <- c("Flag", "Cl_no_rows")
Perc_dups2 <- cbind(Perc_dups, Perc_dups1)

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

```
##        Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1 NO, rerun         4 NO, rerun          5    19  21.05263   26.31579
## 2       YES        15       YES         14    19  78.94737   73.68421
```

Pull out spikes and check - with dionex output

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

```
##                  Sample_ID  SO4_ppm SO4_Area   Cl_ppm  Cl_Area    SO4_mM
## 26      1_MSM_UP_LysA_10cm 126.0804  1.3277  1568.675 23.5520   3.932639
## 27     10_MSM_TR_LysA_10cm 331.4201  3.4901  2755.854 41.3762  10.337495
## 28     11_MSM_TR_LysB_10cm 367.1323  3.8662  3027.198 45.4502  11.451413
## 29  12_MSM_TR_LysC_10c_spk 484.6522  5.1038  2789.309 41.8785  15.117037
## 30     12_MSM_TR_LysC_10cm 314.1841  3.3086  2868.593 43.0689   9.799878
## 31 12_MSM_TR_LysC_10cm_dup 305.1959  3.2139  2796.526 41.9869   9.519523
##       Cl_mM salinity
## 26 44.25036 2.834622
## 27 77.73917 4.979853
## 28 85.39347 5.470174
## 29 78.68290 5.040307
## 30 80.91941 5.183574
## 31 78.88649 5.053348
```

```
#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  12_MSM_TR_LysC_10c_spk  484.6522  5.1038   2789.309  41.8785  15.117037
## 2 15_MSM_TR_LysC_20cm_spk  484.8206  5.1055   2900.910  43.5541  15.122289
## 3 18_MSM_TR_LysC_45cm_spk  636.6021  6.7039   3085.539  46.3261  19.856585
## 4 24_MSM_WC_SipC_20cm_spk 1395.8815  7.3499  11928.089  89.5438  43.539660
## 5 27_MSM_WC_SipC_45cm_spk  793.6321  4.1788  14587.499 109.5080  24.754588
## 6  3_MSM_UP_LysC_10cm_spk  306.0948  3.2234   1440.092  21.6215   9.547561
##       Cl_mM  salinity
```

```
## 1  78.68290  5.040307
## 2  81.83103  5.241970
## 3  87.03919  5.575596
## 4 336.47641 21.554082
## 5 411.49505 26.359638
## 6  40.62320  2.602273
```

```
#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  12_MSM_TR_LysC_10c  15.117037
## 2 15_MSM_TR_LysC_20cm  15.122289
## 3 18_MSM_TR_LysC_45cm  19.856585
## 4 24_MSM_WC_SipC_20cm  43.539660
## 5 27_MSM_WC_SipC_45cm  24.754588
## 6  3_MSM_UP_LysC_10cm   9.547561
```

```
#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 15_MSM_TR_LysC_20cm  308.5873  3.2497   2918.016  43.8109   9.625306
## 2 18_MSM_TR_LysC_45cm  451.0010  4.7494   3028.608  45.4714  14.067405
## 3 24_MSM_WC_SipC_20cm 1360.9248  7.1658  11643.681  87.4088  42.449308
## 4 27_MSM_WC_SipC_45cm  453.6237  2.3885  14717.254 110.4820  14.149211
## 5  3_MSM_UP_LysC_10cm  139.2101  1.4660   1527.373  22.9319   4.342174
## 6 30_GWI_UP_LysC_10cm 1136.6741  5.9850   7773.147  58.3528  35.454588
##       Cl_mM  salinity SO4_mM_spk
## 1  82.31358  5.272881  15.122289
## 2  85.43324  5.472722  19.856585
## 3 328.45362 21.040157  43.539660
## 4 415.15525 26.594103  24.754588
## 5  43.08528  2.759989   9.547561
## 6 219.27072 14.046103  45.109582
```

```
#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 15_MSM_TR_LysC_20cm  308.5873  3.2497   2918.016  43.8109   9.625306
## 2 18_MSM_TR_LysC_45cm  451.0010  4.7494   3028.608  45.4714  14.067405
## 3 24_MSM_WC_SipC_20cm 1360.9248  7.1658  11643.681  87.4088  42.449308
```

```
## 4 27_MSM_WC_SipC_45cm  453.6237  2.3885  14717.254 110.4820  14.149211
## 5  3_MSM_UP_LysC_10cm  139.2101  1.4660   1527.373  22.9319   4.342174
## 6 30_GWI_UP_LysC_10cm 1136.6741  5.9850   7773.147  58.3528  35.454588
##       Cl_mM  salinity SO4_mM_spk SO4_spk_Conc
## 1  82.31358  5.272881  15.122289 7.797879e-05
## 2  85.43324  5.472722  19.856585 7.797879e-05
## 3 328.45362 21.040157  43.539660 7.797879e-05
## 4 415.15525 26.594103  24.754588 7.797879e-05
## 5  43.08528  2.759989   9.547561 7.797879e-05
## 6 219.27072 14.046103  45.109582 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)
```

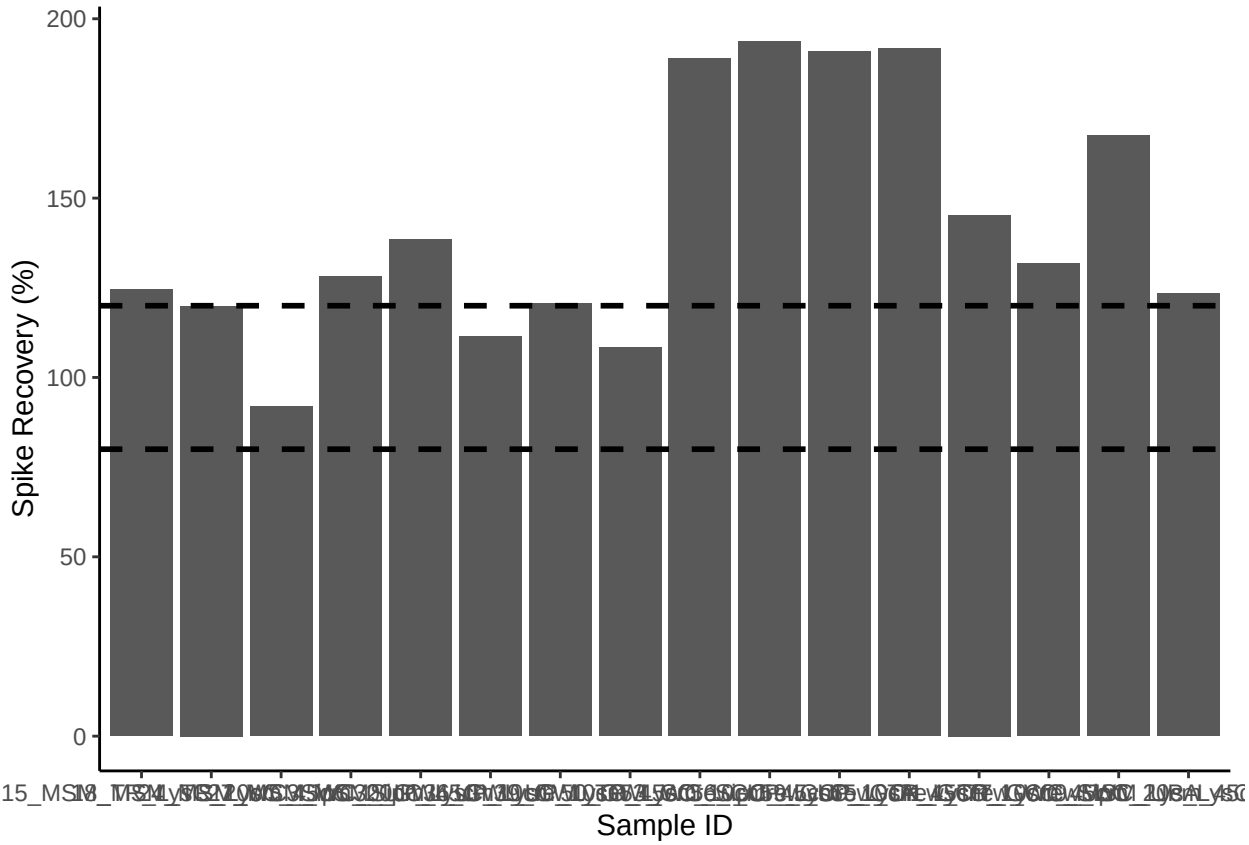
```
##             Sample_ID   SO4_ppm SO4_Area    Cl_ppm   Cl_Area    SO4_mM
## 1 15_MSM_TR_LysC_20cm  308.5873  3.2497   2918.016  43.8109   9.625306
## 2 18_MSM_TR_LysC_45cm  451.0010  4.7494   3028.608  45.4714  14.067405
## 3 24_MSM_WC_SipC_20cm 1360.9248  7.1658  11643.681  87.4088  42.449308
## 4 27_MSM_WC_SipC_45cm  453.6237  2.3885  14717.254 110.4820  14.149211
## 5  3_MSM_UP_LysC_10cm  139.2101  1.4660   1527.373  22.9319   4.342174
## 6 30_GWI_UP_LysC_10cm 1136.6741  5.9850   7773.147  58.3528  35.454588
```

```
##       Cl_mM  salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1  82.31358  5.272881  15.122289 7.797879e-05       50  0.001501
## 2  85.43324  5.472722  19.856585 7.797879e-05       50  0.001501
## 3 328.45362 21.040157  43.539660 7.797879e-05      100  0.001475
## 4 415.15525 26.594103  24.754588 7.797879e-05      100  0.001475
## 5  43.08528  2.759989   9.547561 7.797879e-05       50  0.001501
## 6 219.27072 14.046103  45.109582 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 15_MSM_TR_LysC_20cm  308.5873  3.2497   2918.016  43.8109   9.625306
## 2 18_MSM_TR_LysC_45cm  451.0010  4.7494   3028.608  45.4714  14.067405
## 3 24_MSM_WC_SipC_20cm 1360.9248  7.1658  11643.681  87.4088  42.449308
## 4 27_MSM_WC_SipC_45cm  453.6237  2.3885  14717.254 110.4820  14.149211
## 5  3_MSM_UP_LysC_10cm  139.2101  1.4660   1527.373  22.9319   4.342174
## 6 30_GWI_UP_LysC_10cm 1136.6741  5.9850   7773.147  58.3528  35.454588
##       Cl_mM  salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1  82.31358  5.272881  15.122289 7.797879e-05       50  0.001501
## 2  85.43324  5.472722  19.856585 7.797879e-05       50  0.001501
## 3 328.45362 21.040157  43.539660 7.797879e-05      100  0.001475
## 4 415.15525 26.594103  24.754588 7.797879e-05      100  0.001475
## 5  43.08528  2.759989   9.547561 7.797879e-05       50  0.001501
## 6 219.27072 14.046103  45.109582 7.797879e-05      100  0.001475
##   SO4_Total_unspkd SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1     0.0002889517   0.0004569956    0.0003669305    124.54556     NO, rerun
## 2     0.0004223035   0.0006000660    0.0005002823    119.94548           YES
## 3     0.0006261273   0.0006465640    0.0007041061     91.82763           YES
## 4     0.0002087009   0.0003676056    0.0002866796    128.22872     NO, rerun
## 5     0.0001303521   0.0002885273    0.0002083309    138.49475     NO, rerun
## 6     0.0005229552   0.0006698773    0.0006009340    111.47270           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          12    16      75
## 2 YES                 4    16      25
```

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", "MSM", "UP", "LysA", "10cm"), c("10", "MSM", "TR", :
## number of columns of result is not a multiple of vector length (arg 77)
```

```r
colnames(IDs) <- c("Analysis_No" , "Site", "Zone", "Replicate", "Depth") #, #"Tree")
IDs$Date <- 202206
head(IDs)
```

```
##   Analysis_No Site Zone Replicate Depth   Date
## 1           1  MSM   UP      LysA  10cm 202206
## 2          10  MSM   TR      LysA  10cm 202206
## 3          11  MSM   TR      LysB  10cm 202206
## 4          12  MSM   TR      LysC  10cm 202206
## 5          13  MSM   TR      LysA  20cm 202206
## 6          14  MSM   TR      LysB  20cm 202206
```

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

```
##   Analysis_No Site Zone Replicate Depth   Date           Sample_ID  Cl_Area
## 1           1  MSM   UP      LysA  10cm 202206  1_MSM_UP_LysA_10cm 23.5520
## 2          10  MSM   TR      LysA  10cm 202206 10_MSM_TR_LysA_10cm 41.3762
## 3          11  MSM   TR      LysB  10cm 202206 11_MSM_TR_LysB_10cm 45.4502
## 4          12  MSM   TR      LysC  10cm 202206 12_MSM_TR_LysC_10cm 43.0689
## 5          13  MSM   TR      LysA  20cm 202206 13_MSM_TR_LysA_20cm 42.3007
## 6          14  MSM   TR      LysB  20cm 202206 14_MSM_TR_LysB_20cm 44.3281
##      SO4_mM    Cl_mM salinity
## 1  3.932639 44.25036 2.834622
## 2 10.337495 77.73917 4.979853
## 3 11.451413 85.39347 5.470174
## 4  9.799878 80.91941 5.183574
## 5 13.084039 79.47600 5.091112
## 6 10.877742 83.28524 5.335124
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_Processed_Cl_So4_202206.csv")          #Chan
```

END