Dionex_COMPASS_August2022

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_202208_SO4a.txt", sep='\t' , header=T, skip=3)
head(Sdat)
```

```
##   X       X.1     X.2 IC.SO4 IC.SO4.1 IC.SO4.2 IC.SO4.3 IC.SO4.4
## 1 1 Lab Blank Unknown 4.453   0.0337     5.56   0.0343     0.09
## 2 2 Lab Blank Unknown 4.100   0.0238     3.82   0.0242     0.07
## 3 3 Lab Blank Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 4 4 Lab Blank Unknown 4.450   0.0349     5.48   0.0356     0.09
## 5 5 Lab Blank Unknown   n.a.     n.a.     n.a.     n.a.     n.a.
## 6 6 Lab Blank Unknown 4.450   0.0117     2.36   0.0119     0.04
```

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

```
##         X.1 IC.SO4.1
## 1 Lab Blank  0.0337
## 2 Lab Blank  0.0238
## 3 Lab Blank     n.a.
## 4 Lab Blank  0.0349
## 5 Lab Blank     n.a.
## 6 Lab Blank  0.0117
```

```
## 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
```

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

```
##   Sample_ID SO4_ppm
## 1 Lab Blank  0.0337
## 2 Lab Blank  0.0238
## 3 Lab Blank      NA
## 4 Lab Blank  0.0349
## 5 Lab Blank      NA
## 6 Lab Blank  0.0117
```

```r
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202208_Cla.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 3.287  0.0017    0.37  0.0023    0.03
## 2 2 Lab Blank Unknown 3.287  0.0068    1.46  0.0092    0.09
## 3 3 Lab Blank Unknown 3.290  0.3186  100.00  0.4348    3.14
## 4 4 Lab Blank Unknown 3.287  0.0814   17.08  0.1111    0.37
## 5 5 Lab Blank Unknown 3.290  0.3173  100.00  0.4331    3.13
## 6 6 Lab Blank Unknown 3.287  0.0079    2.14  0.0108    0.11
```

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

```
##         X.1 IC.Cl.1
## 1 Lab Blank 0.0017
## 2 Lab Blank 0.0068
## 3 Lab Blank 0.3186
## 4 Lab Blank 0.0814
## 5 Lab Blank 0.3173
## 6 Lab Blank 0.0079
```

```r
## 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
```

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

```
##   Sample_ID Cl_ppm
## 1 Lab Blank 0.0017
## 2 Lab Blank 0.0068
## 3 Lab Blank 0.3186
## 4 Lab Blank 0.0814
## 5 Lab Blank 0.3173
## 6 Lab Blank 0.0079
```

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

```
##   Sample_ID   SO4_ppm    Cl_ppm
## 1           1036.0978 7696.6153
## 2           1036.0978 2087.0619
## 3           1036.0978    0.0017
## 4           1036.0978 2460.2340
## 5           1036.0978        NA
## 6            226.5618 7696.6153
```

```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_MSM_202208_UP_LysA_20cm 157.5194 1860.322
## 27 10_MSM_202208_TR_LysA_20cm 406.2132 2697.053
## 28 11_MSM_202208_TR_LysA_45cm 469.2068 3444.071
## 29 12_MSM_202208_TR_LysB_10cm 378.0426 2925.598
## 30 13_MSM_202208_TR_LysB_20cm 169.5410 2694.274
## 31 14_MSM_202208_TR_LysB_45cm 285.5842 3085.438
```

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

```
##                     Sample_ID  SO4_ppm   Cl_ppm
## 26  1_MSM_202208_UP_LysA_20cm 157.5194 1860.322
## 27 10_MSM_202208_TR_LysA_20cm 406.2132 2697.053
## 28 11_MSM_202208_TR_LysA_45cm 469.2068 3444.071
## 29 12_MSM_202208_TR_LysB_10cm 378.0426 2925.598
## 30 13_MSM_202208_TR_LysB_20cm 169.5410 2694.274
## 31 14_MSM_202208_TR_LysB_45cm 285.5842 3085.438
```

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
## 520 Standard 1  0.5088 5.0017
## 521 Standard 1  0.5088 5.0358
## 522 Standard 1  0.5088 5.2399
## 523 Standard 1  0.5088 5.1696
## 524 Standard 1  0.5106 5.0017
## 525 Standard 1  0.5106 5.0358
```

```r
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  0.522 0.0132  2.53   NO, rerun
## 2 Standard 2  0.987 0.00391 0.396  YES
## 3 Standard 3  2.05  0.00517 0.252  YES
## 4 Standard 4  9.87  0.00692 0.0702 YES
## 5 Standard 5 20.1   0.0311  0.155  YES
```

```r
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   5.11 0.100  1.96  YES
## 2 Standard 2  10.3  0.0303 0.294 YES
## 3 Standard 3  20.7  0.0363 0.175 YES
## 4 Standard 4  98.4  0.173  0.176 YES
## 5 Standard 5 201.   0.241  0.120 YES
```

Calculate mmol/L concentrations

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

```
##                     Sample_ID  SO4_ppm   Cl_ppm
## 26  1_MSM_202208_UP_LysA_20cm 157.5194 1860.322
## 27 10_MSM_202208_TR_LysA_20cm 406.2132 2697.053
## 28 11_MSM_202208_TR_LysA_45cm 469.2068 3444.071
## 29 12_MSM_202208_TR_LysB_10cm 378.0426 2925.598
## 30 13_MSM_202208_TR_LysB_20cm 169.5410 2694.274
## 31 14_MSM_202208_TR_LysB_45cm 285.5842 3085.438
```

```r
# Constants needed for calculations:
clmw <- 35.45     #molecular weight of Chloride: 35.45
smw <- 32.06      #molecular weight of sulfur: 32.06
```

```
# Convert ppm to mmol/L
sampledat$SO4_mM <- (sampledat$SO4_ppm / smw)
sampledat$Cl_mM <- (sampledat$Cl_ppm / clmw)


# Calculate Salinity
# calculated using the Knudsen equation
# Salinity = 0.03 + 1.8050 * Chlorinity
# Ref: A Practical Handbook of Seawater Analysis by Strickland & Parsons (P. 11)
# =((1.807*Cl_ppm)+0.026)/1000
sampledat$salinity <- ((1.8070 * sampledat$Cl_ppm) + 0.026) / 1000


head(sampledat)
```

```
##                     Sample_ID  SO4_ppm   Cl_ppm    SO4_mM    Cl_mM salinity
## 26  1_MSM_202208_UP_LysA_20cm 157.5194 1860.322  4.913269 52.47736 3.361629
## 27 10_MSM_202208_TR_LysA_20cm 406.2132 2697.053 12.670405 76.08049 4.873601
## 28 11_MSM_202208_TR_LysA_45cm 469.2068 3444.071 14.635271 97.15292 6.223462
## 29 12_MSM_202208_TR_LysB_10cm 378.0426 2925.598 11.791722 82.52745 5.286582
## 30 13_MSM_202208_TR_LysB_20cm 169.5410 2694.274  5.288241 76.00210 4.868580
## 31 14_MSM_202208_TR_LysB_45cm 285.5842 3085.438  8.907804 87.03632 5.575412
```

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 salinity
## 26  1_MSM_202208_UP_LysA_20cm 157.5194 1860.322  4.913269 52.47736 3.361629
## 27 10_MSM_202208_TR_LysA_20cm 406.2132 2697.053 12.670405 76.08049 4.873601
## 28 11_MSM_202208_TR_LysA_45cm 469.2068 3444.071 14.635271 97.15292 6.223462
## 29 12_MSM_202208_TR_LysB_10cm 378.0426 2925.598 11.791722 82.52745 5.286582
## 30 13_MSM_202208_TR_LysB_20cm 169.5410 2694.274  5.288241 76.00210 4.868580
## 31 14_MSM_202208_TR_LysB_45cm 285.5842 3085.438  8.907804 87.03632 5.575412
```

```
#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 26_MSM_202208_WC_SipC_45cm_dup  52.7837 6322.2044  1.646404 1.783414e+02
## 2  37_MSM_202208_RHZ_UP_SF_8_dup  61.1325 1659.3508  1.906815 4.680820e+01
## 3  45_MSM_202208_RHZ_TR_SF_8_dup 534.3976 4088.4286 16.668671 1.153294e+02
## 4  8_MSM_202208_UP_LysC_45cm_dup  44.5578    0.3469  1.389825 9.785614e-03
##       salinity
## 1 1.142425e+01
## 2 2.998473e+00
## 3 7.387816e+00
## 4 6.528483e-04
```

```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 salinity
## 1  1_MSM_202208_UP_LysA_20cm 157.5194 1860.322  4.913269 52.47736 3.361629
## 2 10_MSM_202208_TR_LysA_20cm 406.2132 2697.053 12.670405 76.08049 4.873601
## 3 11_MSM_202208_TR_LysA_45cm 469.2068 3444.071 14.635271 97.15292 6.223462
## 4 12_MSM_202208_TR_LysB_10cm 378.0426 2925.598 11.791722 82.52745 5.286582
## 5 13_MSM_202208_TR_LysB_20cm 169.5410 2694.274  5.288241 76.00210 4.868580
## 6 14_MSM_202208_TR_LysB_45cm 285.5842 3085.438  8.907804 87.03632 5.575412
```

```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 26_MSM_202208_WC_SipC_45cm   1.646404 1.783414e+02 1.142425e+01
## 2  37_MSM_202208_RHZ_UP_SF_8   1.906815 4.680820e+01 2.998473e+00
## 3  45_MSM_202208_RHZ_TR_SF_8  16.668671 1.153294e+02 7.387816e+00
## 4  8_MSM_202208_UP_LysC_45cm   1.389825 9.785614e-03 6.528483e-04
```

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

```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM        Cl_mM
## 1 26_MSM_202208_WC_SipC_45cm  52.4325 6240.8845  1.6354492 176.04751763
## 2  37_MSM_202208_RHZ_UP_SF_8  60.4674 1640.4626  1.8860699  46.27539069
## 3  45_MSM_202208_RHZ_TR_SF_8 520.7936 3992.5401 16.2443419 112.62454443
## 4  8_MSM_202208_UP_LysC_45cm   9.8873    2.0883  0.3083999   0.05890832
##       salinity SO4_mM_dup    Cl_mM_dup salinity_dup
## 1 11.277304292   1.646404 1.783414e+02 1.142425e+01
## 2  2.964341918   1.906815 4.680820e+01 2.998473e+00
## 3  7.214545961  16.668671 1.153294e+02 7.387816e+00
## 4  0.003799558   1.389825 9.785614e-03 6.528483e-04
```

```r
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)
```

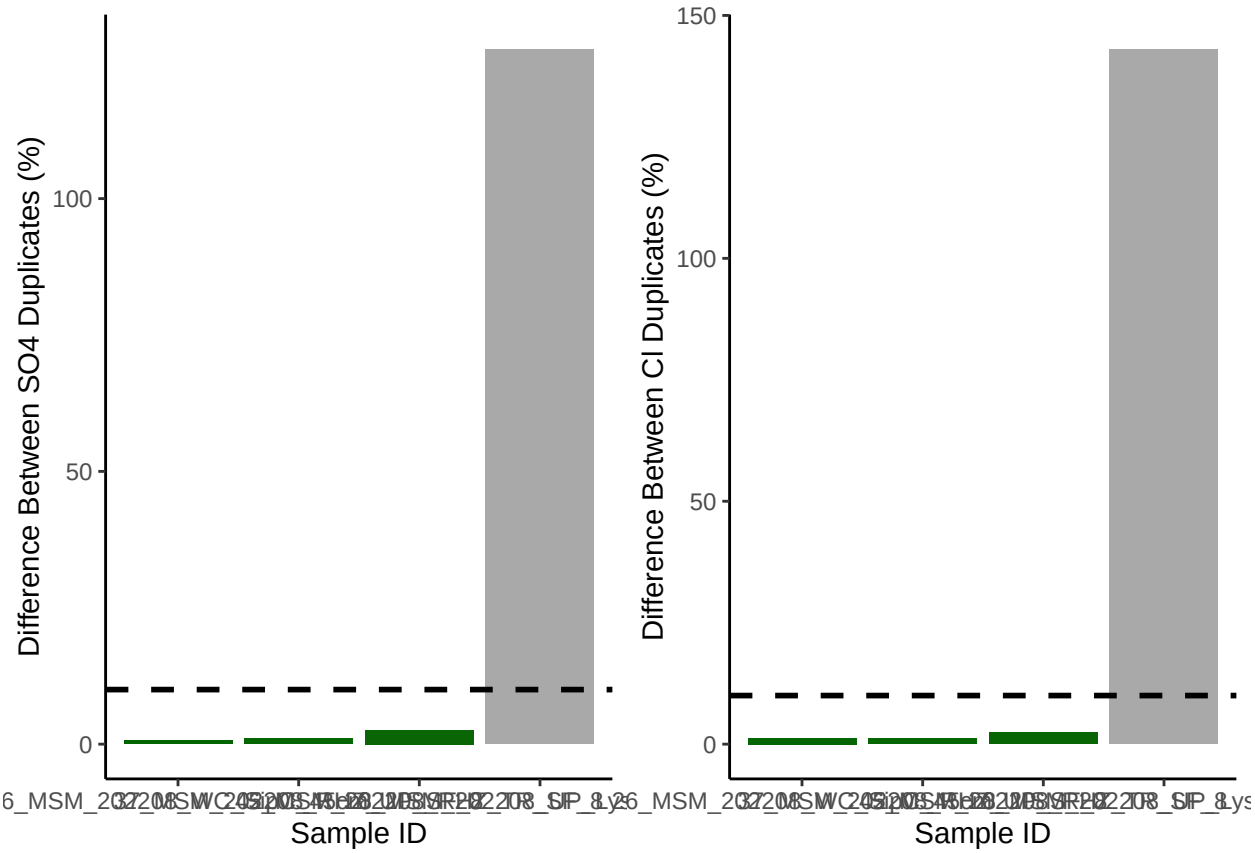
```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM        Cl_mM
## 1 26_MSM_202208_WC_SipC_45cm  52.4325 6240.8845  1.6354492 176.04751763
```

```
## 2  37_MSM_202208_RHZ_UP_SF_8  60.4674 1640.4626  1.8860699  46.27539069
## 3  45_MSM_202208_RHZ_TR_SF_8 520.7936 3992.5401 16.2443419 112.62454443
## 4  8_MSM_202208_UP_LysC_45cm   9.8873    2.0883  0.3083999   0.05890832
##       salinity SO4_mM_dup    Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 11.277304292   1.646404 1.783414e+02 1.142425e+01    0.6675778           YES
## 2  2.964341918   1.906815 4.680820e+01 2.998473e+00    1.0939154           YES
## 3  7.214545961  16.668671 1.153294e+02 7.387816e+00    2.5784900           YES
## 4  0.003799558   1.389825 9.785614e-03 6.528483e-04  127.3594869     NO, rerun
##   Cl_dups_chk Cl_dups_flag
## 1    1.294584          YES
## 2    1.144804          YES
## 3    2.373193          YES
## 4  143.019054    NO, rerun
```

```
#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
Sdupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = SO4_dups_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)
```

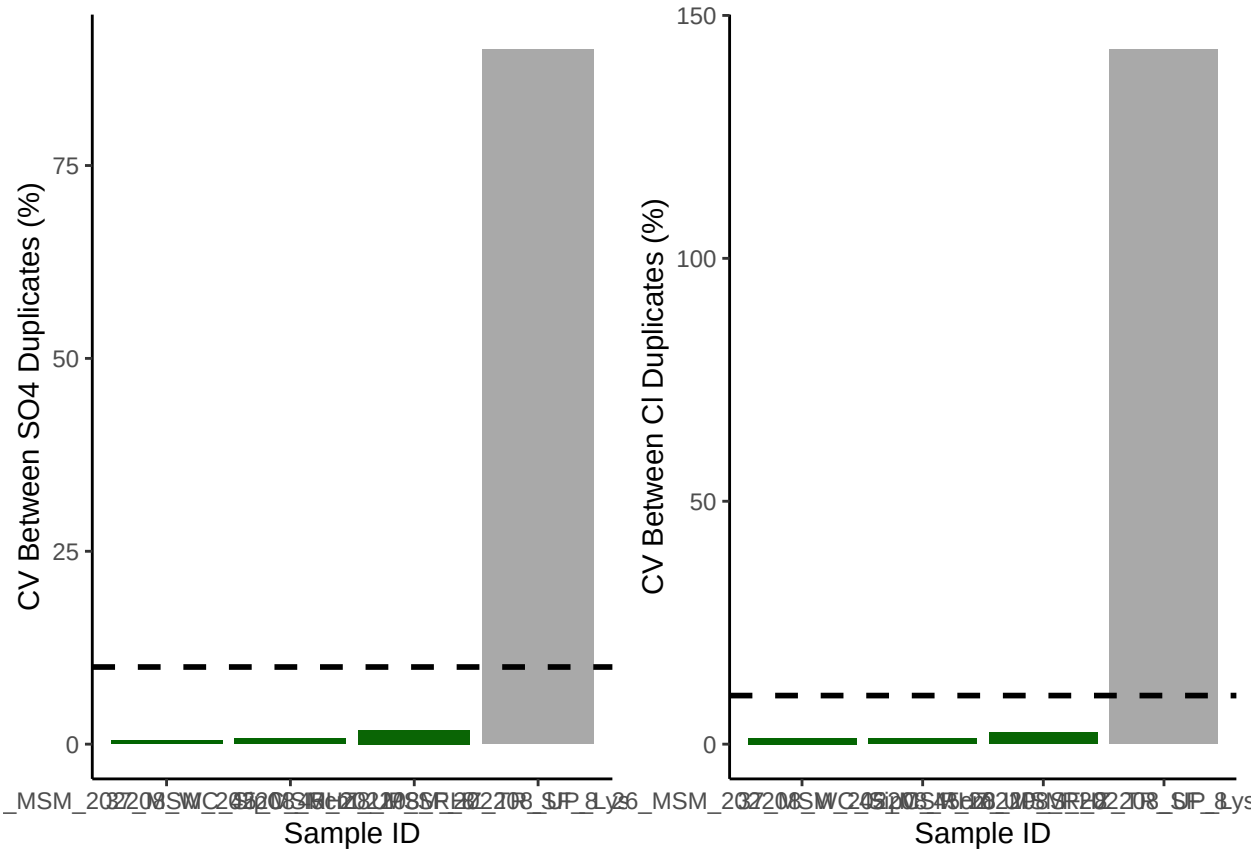
```
##        Flag S_no_rows      Flag Cl_no_rows Total S_Percent Cl_Percent
## 1 NO, rerun         1 NO, rerun          1     4        25         25
## 2       YES         3       YES          3     4        75         75
```

Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM        Cl_mM
## 1 26_MSM_202208_WC_SipC_45cm  52.4325 6240.8845  1.6354492 176.04751763
## 2  37_MSM_202208_RHZ_UP_SF_8  60.4674 1640.4626  1.8860699  46.27539069
## 3  45_MSM_202208_RHZ_TR_SF_8 520.7936 3992.5401 16.2443419 112.62454443
## 4  8_MSM_202208_UP_LysC_45cm   9.8873    2.0883  0.3083999   0.05890832
##       salinity SO4_mM_dup    Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 11.277304292   1.646404 1.783414e+02 1.142425e+01    0.6675778           YES
## 2  2.964341918   1.906815 4.680820e+01 2.998473e+00    1.0939154           YES
## 3  7.214545961  16.668671 1.153294e+02 7.387816e+00    2.5784900           YES
## 4  0.003799558   1.389825 9.785614e-03 6.528483e-04  127.3594869     NO, rerun
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1    1.294584          YES   0.4720488              YES
## 2    1.144804          YES   0.7735150              YES
## 3    2.373193          YES   1.8232678              YES
## 4  143.019054    NO, rerun  90.0567568        NO, rerun
```

```r
#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         1 NO, rerun          1     4        25         25
## 2       YES         3       YES          3     4        75         75
```

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 salinity
## 26  1_MSM_202208_UP_LysA_20cm 157.5194 1860.322  4.913269 52.47736 3.361629
## 27 10_MSM_202208_TR_LysA_20cm 406.2132 2697.053 12.670405 76.08049 4.873601
## 28 11_MSM_202208_TR_LysA_45cm 469.2068 3444.071 14.635271 97.15292 6.223462
## 29 12_MSM_202208_TR_LysB_10cm 378.0426 2925.598 11.791722 82.52745 5.286582
## 30 13_MSM_202208_TR_LysB_20cm 169.5410 2694.274  5.288241 76.00210 4.868580
## 31 14_MSM_202208_TR_LysB_45cm 285.5842 3085.438  8.907804 87.03632 5.575412
```

```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 26_MSM_202208_WC_SipC_45cm_spk 290.4766 6810.1389  9.060405 1.921055e+02
## 2  37_MSM_202208_RHZ_UP_SF_8_spk 178.5871 1666.1291  5.570402 4.699941e+01
## 3  45_MSM_202208_RHZ_TR_SF_8_spk 649.8361 4082.0336 20.269373 1.151490e+02
## 4  8_MSM_202208_UP_LysC_45cm_spk 156.6900    0.0336  4.887399 9.478138e-04
##       salinity
## 1 1.230595e+01
## 2 3.010721e+00
## 3 7.376261e+00
## 4 8.671520e-05
```

```r
#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 26_MSM_202208_WC_SipC_45cm   9.060405
## 2  37_MSM_202208_RHZ_UP_SF_8   5.570402
## 3  45_MSM_202208_RHZ_TR_SF_8  20.269373
## 4  8_MSM_202208_UP_LysC_45cm   4.887399
```

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

```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM        Cl_mM
## 1 26_MSM_202208_WC_SipC_45cm  52.4325 6240.8845  1.6354492 176.04751763
## 2  37_MSM_202208_RHZ_UP_SF_8  60.4674 1640.4626  1.8860699  46.27539069
## 3  45_MSM_202208_RHZ_TR_SF_8 520.7936 3992.5401 16.2443419 112.62454443
## 4  8_MSM_202208_UP_LysC_45cm   9.8873    2.0883  0.3083999   0.05890832
##       salinity SO4_mM_spk
## 1 11.277304292   9.060405
```

```
## 2  2.964341918   5.570402
## 3  7.214545961  20.269373
## 4  0.003799558   4.887399
```

```
#now we need to calculate the spike concentration and calculate the spike recovery
spkconc <- (250/smw)    # in mM
spkvol <- 10             # in uL
spkvol <- spkvol/1000000
#spike for these samples was 10uL of the 250mM standard
QAspks$SO4_spk_Conc <- (spkconc)*spkvol         # mmoles of SO4
head(QAspks)
```

```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM        Cl_mM
## 1 26_MSM_202208_WC_SipC_45cm  52.4325 6240.8845  1.6354492 176.04751763
## 2  37_MSM_202208_RHZ_UP_SF_8  60.4674 1640.4626  1.8860699  46.27539069
## 3  45_MSM_202208_RHZ_TR_SF_8 520.7936 3992.5401 16.2443419 112.62454443
## 4  8_MSM_202208_UP_LysC_45cm   9.8873    2.0883  0.3083999   0.05890832
##       salinity SO4_mM_spk SO4_spk_Conc
## 1 11.277304292   9.060405 7.797879e-05
## 2  2.964341918   5.570402 7.797879e-05
## 3  7.214545961  20.269373 7.797879e-05
## 4  0.003799558   4.887399 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)
```

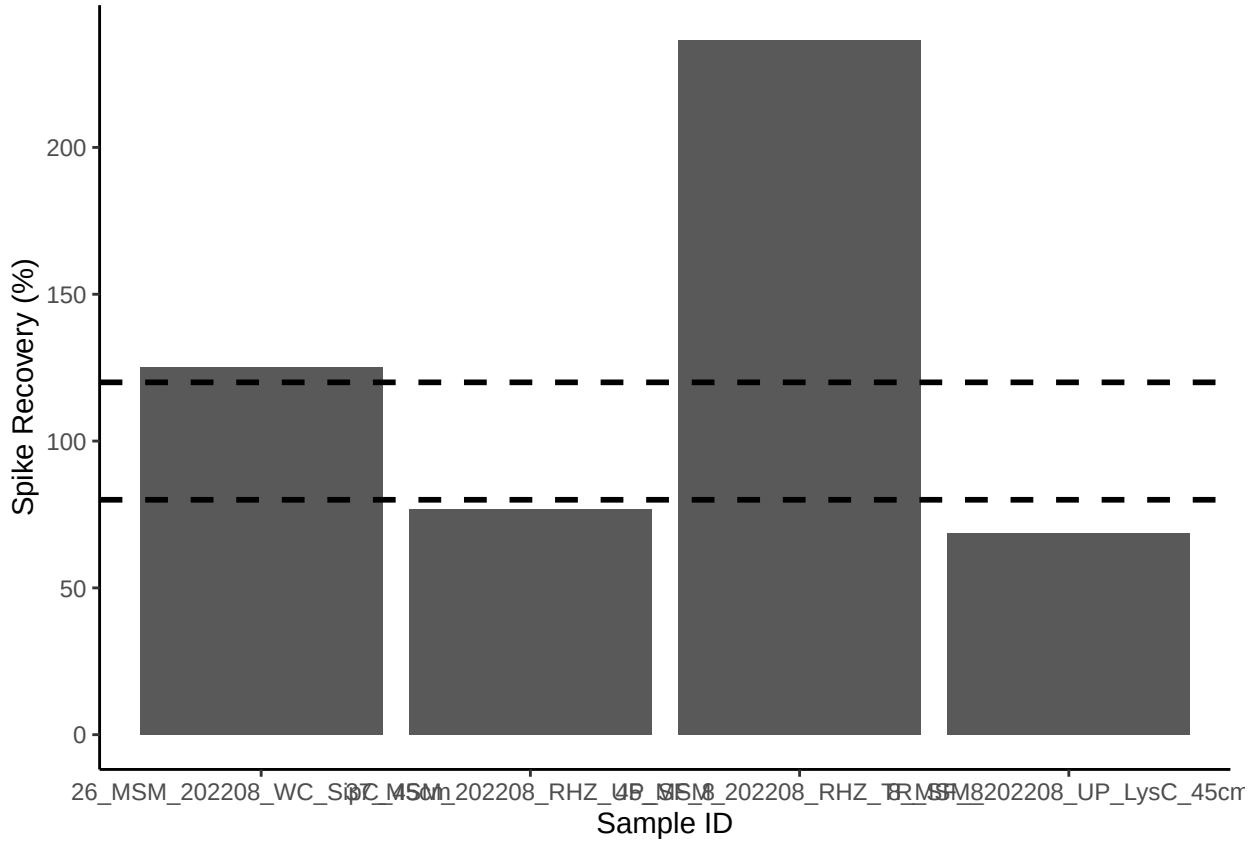
```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM        Cl_mM
## 1 26_MSM_202208_WC_SipC_45cm  52.4325 6240.8845  1.6354492 176.04751763
## 2  37_MSM_202208_RHZ_UP_SF_8  60.4674 1640.4626  1.8860699  46.27539069
## 3  45_MSM_202208_RHZ_TR_SF_8 520.7936 3992.5401 16.2443419 112.62454443
## 4  8_MSM_202208_UP_LysC_45cm   9.8873    2.0883  0.3083999   0.05890832
##       salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1 11.277304292   9.060405 7.797879e-05        1     1e-06
## 2  2.964341918   5.570402 7.797879e-05        1     1e-06
## 3  7.214545961  20.269373 7.797879e-05        1     1e-06
## 4  0.003799558   4.887399 7.797879e-05        1     1e-06
```

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

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

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

head(QAspks)
```

```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM        Cl_mM
## 1 26_MSM_202208_WC_SipC_45cm  52.4325 6240.8845  1.6354492 176.04751763
## 2  37_MSM_202208_RHZ_UP_SF_8  60.4674 1640.4626  1.8860699  46.27539069
## 3  45_MSM_202208_RHZ_TR_SF_8 520.7936 3992.5401 16.2443419 112.62454443
## 4  8_MSM_202208_UP_LysC_45cm   9.8873    2.0883  0.3083999   0.05890832
##       salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1 11.277304292   9.060405 7.797879e-05        1     1e-06     1.635449e-06
## 2  2.964341918   5.570402 7.797879e-05        1     1e-06     1.886070e-06
## 3  7.214545961  20.269373 7.797879e-05        1     1e-06     1.624434e-05
## 4  0.003799558   4.887399 7.797879e-05        1     1e-06     3.083999e-07
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   9.966446e-05    7.961424e-05    125.18422     NO, rerun
## 2   6.127443e-05    7.986486e-05     76.72264     NO, rerun
## 3   2.229631e-04    9.422313e-05    236.63309     NO, rerun
## 4   5.376138e-05    7.828719e-05     68.67201     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: 1 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 NO, rerun           4     4     100
```

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", "202208", "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  MSM 202208   UP      LysA  20cm    1
## 2          10  MSM 202208   TR      LysA  20cm   10
## 3          11  MSM 202208   TR      LysA  45cm   11
## 4          12  MSM 202208   TR      LysB  10cm   12
## 5          13  MSM 202208   TR      LysB  20cm   13
## 6          14  MSM 202208   TR      LysB  45cm   14
```

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

```
##   Analysis_No Date   Site Zone Replicate Depth Tree                  Sample_ID
## 1           1  MSM 202208   UP      LysA  20cm    1  1_MSM_202208_UP_LysA_20cm
## 2          10  MSM 202208   TR      LysA  20cm   10 10_MSM_202208_TR_LysA_20cm
## 3          11  MSM 202208   TR      LysA  45cm   11 11_MSM_202208_TR_LysA_45cm
## 4          12  MSM 202208   TR      LysB  10cm   12 12_MSM_202208_TR_LysB_10cm
## 5          13  MSM 202208   TR      LysB  20cm   13 13_MSM_202208_TR_LysB_20cm
## 6          14  MSM 202208   TR      LysB  45cm   14 14_MSM_202208_TR_LysB_45cm
##    SO4_ppm   Cl_ppm    SO4_mM    Cl_mM salinity
## 1 157.5194 1860.322  4.913269 52.47736 3.361629
## 2 406.2132 2697.053 12.670405 76.08049 4.873601
## 3 469.2068 3444.071 14.635271 97.15292 6.223462
## 4 378.0426 2925.598 11.791722 82.52745 5.286582
## 5 169.5410 2694.274  5.288241 76.00210 4.868580
## 6 285.5842 3085.438  8.907804 87.03632 5.575412
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202208a.csv")  #Change fi
```

END