Dionex_COMPASS_September2022

Stephanie J. Wilson

2023-01-23

Daily Set up

Read in and Format the raw data - change wd & file names

```
# SULFATE DATA:

## Read in raw data file from Dionex - copied and saved as a txt
Sdat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202209_SO4.txt",sep='\t' , header=T, skip=3)
head(Sdat)
```

```
##   X        X.1                  X.2 IC.SO4 IC.SO4.1 IC.SO4.2 IC.SO4.3 IC.SO4.4
## 1 1  Lab Blank              Unknown 4.413   0.5525     6.21   0.4551     2.87
## 2 2  Lab Blank              Unknown 4.417   0.5520     6.21   0.4546     2.87
## 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 4.423   0.5525     6.20   0.4550     2.87
## 6 6 Standard 1 Calibration Standard 4.427   1.1273     6.27   0.9285     5.72
```

```
## 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  0.5525
## 2  Lab Blank  0.5520
## 3  Lab Blank     n.a.
## 4  Lab Blank     n.a.
## 5  Lab Blank  0.5525
## 6 Standard 1  1.1273
```

```
## 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  0.5525
## 2  Lab Blank  0.5520
## 3  Lab Blank      NA
## 4  Lab Blank      NA
## 5  Lab Blank  0.5525
## 6 Standard 1  1.1273
```

```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_202209_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 3.293   6.0694   93.08   6.8202   47.25
## 2 2  Lab Blank              Unknown 3.297   6.0646   93.06   6.8149   47.26
## 3 3  Lab Blank              Unknown   n.a.     n.a.    n.a.     n.a.    n.a.
## 4 4  Lab Blank              Unknown 3.290   0.0090  100.00   0.0101    0.07
## 5 5  Lab Blank              Unknown 3.300   6.0747   93.07   6.8262   47.33
## 6 6 Standard 1 Calibration Standard 3.300  12.2365   92.93  13.7503   94.39
```

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

```
##          X.1  IC.Cl.1
## 1  Lab Blank  6.0694
## 2  Lab Blank  6.0646
## 3  Lab Blank     n.a.
## 4  Lab Blank  0.0090
## 5  Lab Blank  6.0747
## 6 Standard 1 12.2365
```

```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  6.0694
## 2  Lab Blank  6.0646
## 3  Lab Blank      NA
## 4  Lab Blank  0.0090
## 5  Lab Blank  6.0747
## 6 Standard 1 12.2365
```

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

```
##   Sample_ID   SO4_ppm    Cl_ppm
## 1           2568.7298 19169.228
## 2           2568.7298  4914.293
## 3           2568.7298     0.009
## 4           2568.7298  4649.136
## 5           2568.7298        NA
## 6            607.8371 19169.228
```

```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_202209_MSM_UP_LysA_20cm  185.4786 2155.484
## 27     10_202209_MSM_TR_LysA_20cm  505.3590 3550.881
## 28 10_202209_MSM_TR_LysA_20cm_spk  627.6741 3531.850
## 29        100_202209_MSM_WC_RHZ_6  794.4041 8421.266
## 30    100_202209_MSM_WC_RHZ_6_spk 1072.7880 8331.226
## 31        101_202209_MSM_WC_RHZ_7  815.7621 8675.486
```

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

```
##                          Sample_ID   SO4_ppm   Cl_ppm
## 26      1_202209_MSM_UP_LysA_20cm  185.4786 2155.484
## 27     10_202209_MSM_TR_LysA_20cm  505.3590 3550.881
## 28 10_202209_MSM_TR_LysA_20cm_spk  627.6741 3531.850
## 29        100_202209_MSM_WC_RHZ_6  794.4041 8421.266
## 30    100_202209_MSM_WC_RHZ_6_spk 1072.7880 8331.226
## 31        101_202209_MSM_WC_RHZ_7  815.7621 8675.486
```

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
## 1128 Standard 1  1.1273 12.2365
## 1129 Standard 1  1.1273 12.3306
## 1130 Standard 1  1.1273 12.2419
## 1131 Standard 1  1.1273  6.0938
## 1132 Standard 1  1.1501 12.2365
## 1133 Standard 1  1.1501 12.3306
```

```
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  0.991  0.258  26.1 NO, rerun
## 2 Standard 2  1.89   0.562  29.8 NO, rerun
## 3 Standard 3  7.70   4.08   53.0 NO, rerun
## 4 Standard 4 16.8    4.79   28.6 NO, rerun
## 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  10.7  2.76  25.8 NO, rerun
## 2 Standard 2  19.8  5.62  28.4 NO, rerun
## 3 Standard 3  77.8 40.7   52.3 NO, rerun
## 4 Standard 4 168.  47.4   28.3 NO, rerun
## 5 Standard 5  NA   NA     NA   <NA>
```

Calculate mmol/L concentrations

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

```
##                         Sample_ID   SO4_ppm   Cl_ppm
## 26      1_202209_MSM_UP_LysA_20cm  185.4786 2155.484
## 27     10_202209_MSM_TR_LysA_20cm  505.3590 3550.881
## 28 10_202209_MSM_TR_LysA_20cm_spk  627.6741 3531.850
## 29        100_202209_MSM_WC_RHZ_6  794.4041 8421.266
## 30    100_202209_MSM_WC_RHZ_6_spk 1072.7880 8331.226
## 31        101_202209_MSM_WC_RHZ_7  815.7621 8675.486
```

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

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


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


head(sampledat)
```

```
##                          Sample_ID   SO4_ppm   Cl_ppm    SO4_mM     Cl_mM
## 26      1_202209_MSM_UP_LysA_20cm  185.4786 2155.484  5.785359  60.80349
## 27     10_202209_MSM_TR_LysA_20cm  505.3590 3550.881 15.762913 100.16590
## 28 10_202209_MSM_TR_LysA_20cm_spk  627.6741 3531.850 19.578107  99.62906
## 29        100_202209_MSM_WC_RHZ_6  794.4041 8421.266 24.778668 237.55333
## 30    100_202209_MSM_WC_RHZ_6_spk 1072.7880 8331.226 33.461884 235.01341
## 31        101_202209_MSM_WC_RHZ_7  815.7621 8675.486 25.444857 244.72456
##     salinity
## 26  3.894985
## 27  6.416469
## 28  6.382079
## 29 15.217253
## 30 15.054550
## 31 15.676629
```

Pull out dups and check with percent difference

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

```
##                          Sample_ID   SO4_ppm   Cl_ppm    SO4_mM     Cl_mM
## 26      1_202209_MSM_UP_LysA_20cm  185.4786 2155.484  5.785359  60.80349
## 27     10_202209_MSM_TR_LysA_20cm  505.3590 3550.881 15.762913 100.16590
## 28 10_202209_MSM_TR_LysA_20cm_spk  627.6741 3531.850 19.578107  99.62906
## 29        100_202209_MSM_WC_RHZ_6  794.4041 8421.266 24.778668 237.55333
## 30    100_202209_MSM_WC_RHZ_6_spk 1072.7880 8331.226 33.461884 235.01341
## 31        101_202209_MSM_WC_RHZ_7  815.7621 8675.486 25.444857 244.72456
##     salinity
## 26  3.894985
## 27  6.416469
## 28  6.382079
## 29 15.217253
## 30 15.054550
## 31 15.676629
```

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

```
##                          Sample_ID   SO4_ppm    Cl_ppm     SO4_mM       Cl_mM
## 1   109_202209_GWI_TR_RHZ_SF_7_dup  933.3051  6544.394 29.1112009 184.6091312
## 2   19_202209_MSM_WC_SipA_45cm_dup   51.6200  8507.394  1.6101061 239.9828999
## 3   29_202209_GWI_UP_LysA_20cm_dup  772.0785  6490.409 24.0822988 183.0862934
## 4   39_202209_GWI_TR_LysA_45cm_dup  701.8766  8857.495 21.8925951 249.8588124
## 5   49_202209_GWI_WC_SipB_10cm_dup 1907.2701 13781.568 59.4906457 388.7607419
## 6 59_202209_GCrew_TR_LysA_20cm_dup   10.1312     3.942  0.3160075   0.1111989
##       salinity
## 1 11.825745416
## 2 15.372886597
## 3 11.728195244
## 4 16.005519284
## 5 24.903319918
## 6  0.007149194
```

```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_202209_MSM_UP_LysA_20cm 185.4786 2155.484  5.785359  60.80349  3.894985
## 2 10_202209_MSM_TR_LysA_20cm 505.3590 3550.881 15.762913 100.16590  6.416469
## 3    100_202209_MSM_WC_RHZ_6 794.4041 8421.266 24.778668 237.55333 15.217253
## 4    101_202209_MSM_WC_RHZ_7 815.7621 8675.486 25.444857 244.72456 15.676629
## 5    102_202209_MSM_WC_RHZ_8 872.0692 8465.177 27.201160 238.79201 15.296600
## 6 103_202209_GWI_TR_RHZ_SF_1 951.8538 6184.023 29.689763 174.44353 11.174556
```

```r
#remove the dup from these IDs so we will have duplicate sample names
dups$Sample_ID<-gsub("_dup","",as.character(dups$Sample_ID))
dups <- dups[ ,-c(2,3)]
colnames(dups) <- c('Sample_ID', 'SO4_mM_dup', "Cl_mM_dup", "salinity_dup")
head(dups)
```

```
##                      Sample_ID SO4_mM_dup   Cl_mM_dup salinity_dup
## 1   109_202209_GWI_TR_RHZ_SF_7 29.1112009 184.6091312 11.825745416
## 2   19_202209_MSM_WC_SipA_45cm  1.6101061 239.9828999 15.372886597
## 3   29_202209_GWI_UP_LysA_20cm 24.0822988 183.0862934 11.728195244
## 4   39_202209_GWI_TR_LysA_45cm 21.8925951 249.8588124 16.005519284
## 5   49_202209_GWI_WC_SipB_10cm 59.4906457 388.7607419 24.903319918
## 6 59_202209_GCrew_TR_LysA_20cm  0.3160075   0.1111989  0.007149194
```

```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   109_202209_GWI_TR_RHZ_SF_7  927.5268  6420.4224 28.9309669 181.1120564
## 2   19_202209_MSM_WC_SipA_45cm   52.8767  8548.8314  1.6493044 241.1518025
## 3   29_202209_GWI_UP_LysA_20cm  768.0581  6435.7454 23.9568964 181.5442990
## 4   39_202209_GWI_TR_LysA_45cm  689.8972  8687.1963 21.5189395 245.0549027
## 5   49_202209_GWI_WC_SipB_10cm 1925.9192 13905.9219 60.0723394 392.2686008
## 6 59_202209_GCrew_TR_LysA_20cm   10.4618     4.9391  0.3263194   0.1393258
##       salinity SO4_mM_dup   Cl_mM_dup salinity_dup
## 1 11.601729277 29.1112009 184.6091312 11.825745416
## 2 15.447764340  1.6101061 239.9828999 15.372886597
## 3 11.629417938 24.0822988 183.0862934 11.728195244
## 4 15.697789714 21.8925951 249.8588124 16.005519284
## 5 25.128026873 59.4906457 388.7607419 24.903319918
## 6  0.008950954  0.3160075   0.1111989  0.007149194
```

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

QAdups$Cl_dups_chk <- ((abs(QAdups$Cl_mM-QAdups$Cl_mM_dup))/((QAdups$Cl_mM+QAdups$Cl_mM_dup)/2))*100
QAdups$Cl_dups_flag <-  ifelse(QAdups$Cl_dups_chk <10, 'YES', 'NO, rerun')

head(QAdups)
```

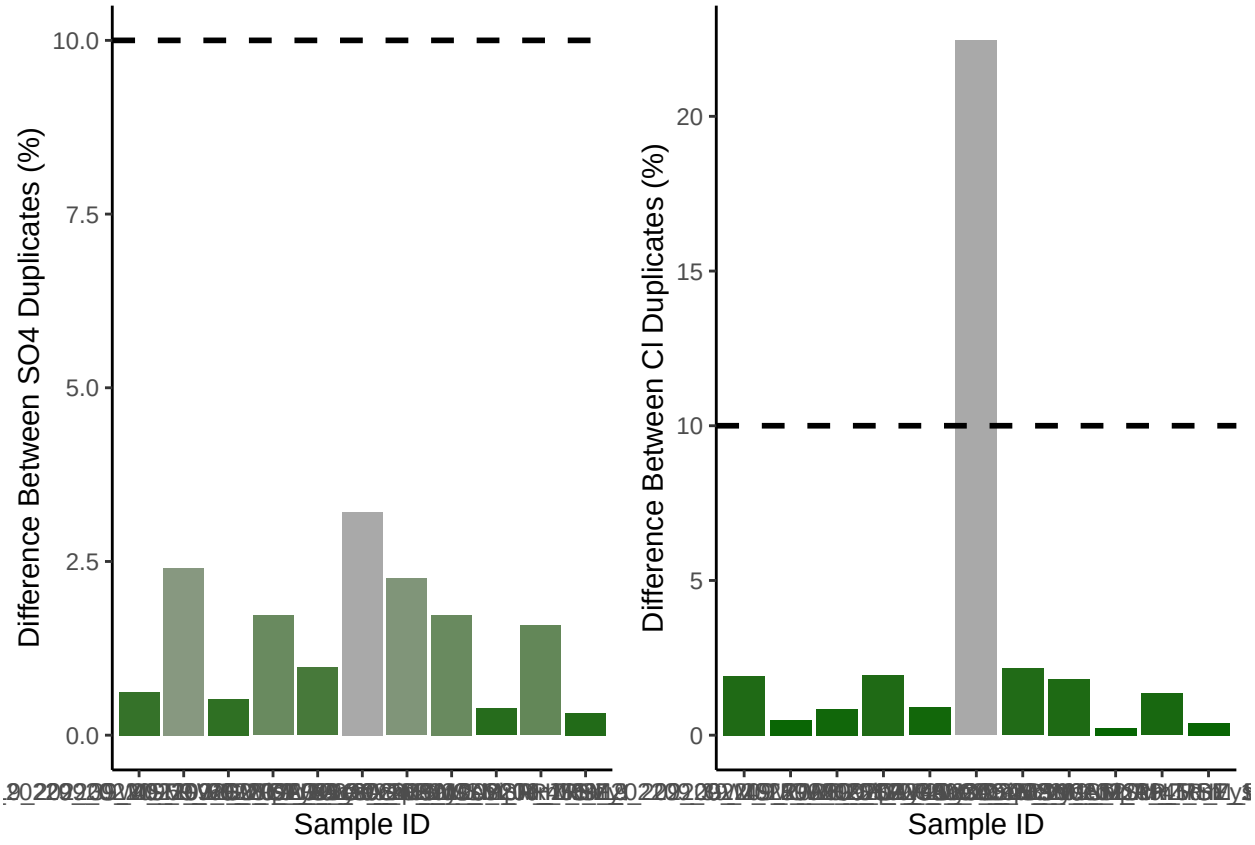
```
##                       Sample_ID   SO4_ppm     Cl_ppm     SO4_mM       Cl_mM
## 1   109_202209_GWI_TR_RHZ_SF_7  927.5268  6420.4224 28.9309669 181.1120564
## 2   19_202209_MSM_WC_SipA_45cm   52.8767  8548.8314  1.6493044 241.1518025
## 3   29_202209_GWI_UP_LysA_20cm  768.0581  6435.7454 23.9568964 181.5442990
## 4   39_202209_GWI_TR_LysA_45cm  689.8972  8687.1963 21.5189395 245.0549027
## 5   49_202209_GWI_WC_SipB_10cm 1925.9192 13905.9219 60.0723394 392.2686008
## 6 59_202209_GCrew_TR_LysA_20cm   10.4618     4.9391  0.3263194   0.1393258
##       salinity SO4_mM_dup   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 11.601729277 29.1112009 184.6091312 11.825745416    0.6210448           YES
## 2 15.447764340  1.6101061 239.9828999 15.372886597    2.4052434           YES
## 3 11.629417938 24.0822988 183.0862934 11.728195244    0.5220836           YES
## 4 15.697789714 21.8925951 249.8588124 16.005519284    1.7214579           YES
## 5 25.128026873 59.4906457 388.7607419 24.903319918    0.9730331           YES
## 6  0.008950954  0.3160075   0.1111989  0.007149194    3.2107998           YES
##   Cl_dups_chk Cl_dups_flag
## 1   1.9124267          YES
## 2   0.4858941          YES
## 3   0.8457844          YES
## 4   1.9413120          YES
## 5   0.8982656          YES
## 6  22.4544257    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  YES        11 NO, rerun          1    11       100   9.090909
## 2  YES        11       YES         10    11       100  90.909091
```

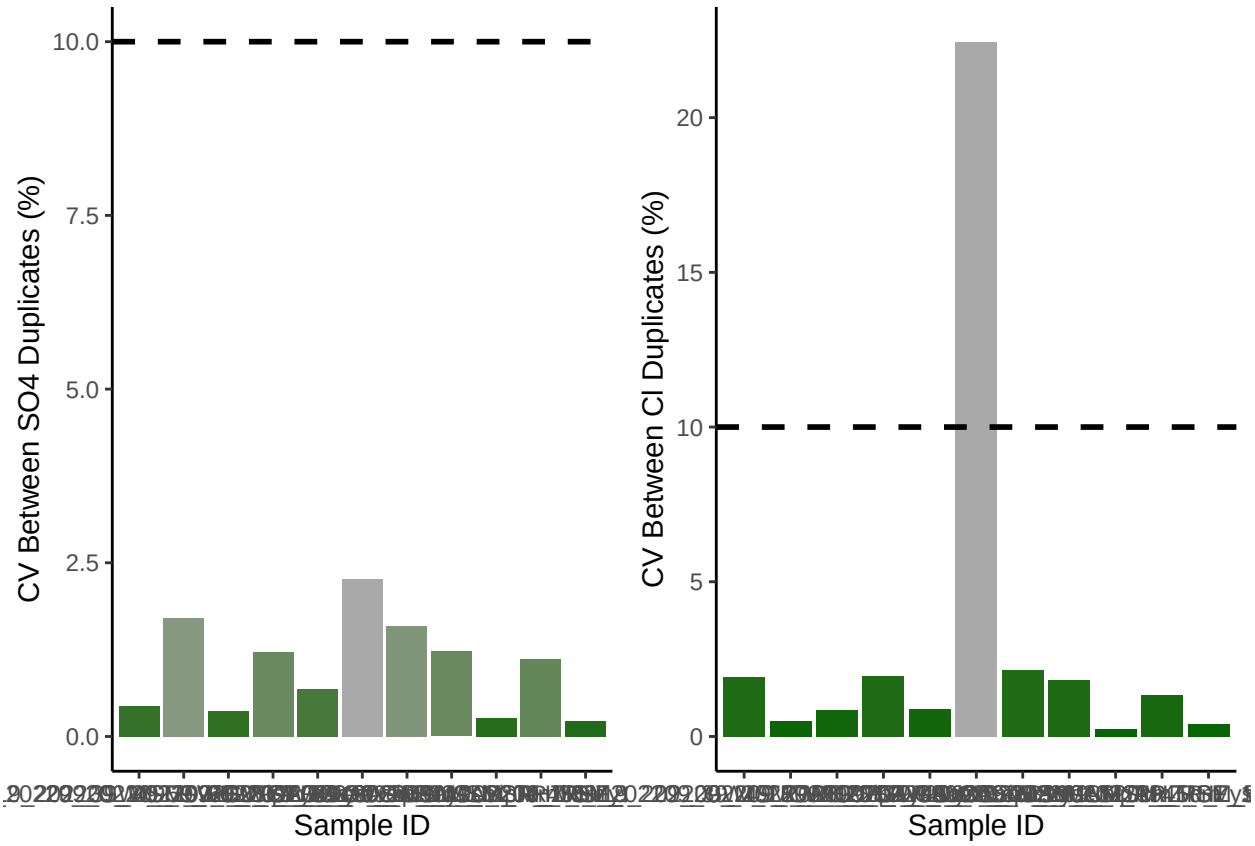
Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

```
##                      Sample_ID   SO4_ppm     Cl_ppm     SO4_mM       Cl_mM
## 1   109_202209_GWI_TR_RHZ_SF_7  927.5268  6420.4224 28.9309669 181.1120564
## 2   19_202209_MSM_WC_SipA_45cm   52.8767  8548.8314  1.6493044 241.1518025
## 3   29_202209_GWI_UP_LysA_20cm  768.0581  6435.7454 23.9568964 181.5442990
## 4   39_202209_GWI_TR_LysA_45cm  689.8972  8687.1963 21.5189395 245.0549027
## 5   49_202209_GWI_WC_SipB_10cm 1925.9192 13905.9219 60.0723394 392.2686008
## 6 59_202209_GCrew_TR_LysA_20cm   10.4618     4.9391  0.3263194   0.1393258
##       salinity SO4_mM_dup   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1 11.601729277 29.1112009 184.6091312 11.825745416    0.6210448           YES
## 2 15.447764340  1.6101061 239.9828999 15.372886597    2.4052434           YES
## 3 11.629417938 24.0822988 183.0862934 11.728195244    0.5220836           YES
## 4 15.697789714 21.8925951 249.8588124 16.005519284    1.7214579           YES
## 5 25.128026873 59.4906457 388.7607419 24.903319918    0.9730331           YES
## 6  0.008950954  0.3160075   0.1111989  0.007149194    3.2107998           YES
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1   1.9124267          YES   0.4391450              YES
## 2   0.4858941          YES   1.7007639              YES
## 3   0.8457844          YES   0.3691688              YES
## 4   1.9413120          YES   1.2172546              YES
## 5   0.8982656          YES   0.6880383              YES
## 6  22.4544257    NO, rerun   2.2703783              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        11 NO, rerun          1    11       100   9.090909
## 2  YES        11       YES         10    11       100  90.909091
```

Pull out spikes and check

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

```
##                         Sample_ID   SO4_ppm   Cl_ppm    SO4_mM     Cl_mM
## 26      1_202209_MSM_UP_LysA_20cm  185.4786 2155.484  5.785359  60.80349
## 27     10_202209_MSM_TR_LysA_20cm  505.3590 3550.881 15.762913 100.16590
## 28 10_202209_MSM_TR_LysA_20cm_spk  627.6741 3531.850 19.578107  99.62906
## 29        100_202209_MSM_WC_RHZ_6  794.4041 8421.266 24.778668 237.55333
## 30    100_202209_MSM_WC_RHZ_6_spk 1072.7880 8331.226 33.461884 235.01341
## 31        101_202209_MSM_WC_RHZ_7  815.7621 8675.486 25.444857 244.72456
##     salinity
## 26  3.894985
## 27  6.416469
## 28  6.382079
## 29 15.217253
## 30 15.054550
## 31 15.676629
```

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

```
##                        Sample_ID   SO4_ppm    Cl_ppm   SO4_mM     Cl_mM
## 1 10_202209_MSM_TR_LysA_20cm_spk  627.6741  3531.850 19.57811  99.62906
## 2    100_202209_MSM_WC_RHZ_6_spk 1072.7880  8331.226 33.46188 235.01341
## 3 110_202209_GWI_TR_RHZ_SF_8_spk  793.6349  4702.940 24.75468 132.66403
## 4 20_202209_MSM_WC_SipB_10cm_spk 1122.1770  9663.288 35.00240 272.58922
## 5 30_202209_GWI_UP_LysA_20cm_spk  945.0486  5694.335 29.47750 160.63003
## 6 40_202209_GWI_TR_LysB_10cm_spk 1990.9739 12742.654 62.10149 359.45428
##    salinity
## 1  6.382079
## 2 15.054550
## 3  8.498238
## 4 17.461587
## 5 10.289689
## 6 23.026002
```

```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 10_202209_MSM_TR_LysA_20cm   19.57811
## 2    100_202209_MSM_WC_RHZ_6   33.46188
## 3 110_202209_GWI_TR_RHZ_SF_8   24.75468
## 4 20_202209_MSM_WC_SipB_10cm   35.00240
## 5 30_202209_GWI_UP_LysA_20cm   29.47750
## 6 40_202209_GWI_TR_LysB_10cm   62.10149
```

```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  salinity
## 1 10_202209_MSM_TR_LysA_20cm  505.3590  3550.881 15.76291 100.1659  6.416469
## 2    100_202209_MSM_WC_RHZ_6  794.4041  8421.266 24.77867 237.5533 15.217253
## 3 110_202209_GWI_TR_RHZ_SF_8  962.0158  6500.256 30.00673 183.3641 11.745989
## 4 20_202209_MSM_WC_SipB_10cm  860.1581  9655.185 26.82964 272.3606 17.446945
## 5 30_202209_GWI_UP_LysA_20cm  658.0800  5681.553 20.52651 160.2695 10.266592
## 6 40_202209_GWI_TR_LysB_10cm 1698.7295 12719.267 52.98595 358.7946 22.983742
##   SO4_mM_spk
## 1   19.57811
## 2   33.46188
## 3   24.75468
## 4   35.00240
## 5   29.47750
## 6   62.10149
```

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

```
##                   Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM  salinity
## 1 10_202209_MSM_TR_LysA_20cm  505.3590  3550.881 15.76291 100.1659  6.416469
## 2    100_202209_MSM_WC_RHZ_6  794.4041  8421.266 24.77867 237.5533 15.217253
## 3 110_202209_GWI_TR_RHZ_SF_8  962.0158  6500.256 30.00673 183.3641 11.745989
## 4 20_202209_MSM_WC_SipB_10cm  860.1581  9655.185 26.82964 272.3606 17.446945
## 5 30_202209_GWI_UP_LysA_20cm  658.0800  5681.553 20.52651 160.2695 10.266592
## 6 40_202209_GWI_TR_LysB_10cm 1698.7295 12719.267 52.98595 358.7946 22.983742
##   SO4_mM_spk SO4_spk_Conc
## 1   19.57811 7.797879e-05
## 2   33.46188 7.797879e-05
## 3   24.75468 7.797879e-05
```

```
## 4   35.00240 7.797879e-05
## 5   29.47750 7.797879e-05
## 6   62.10149 7.797879e-05
```

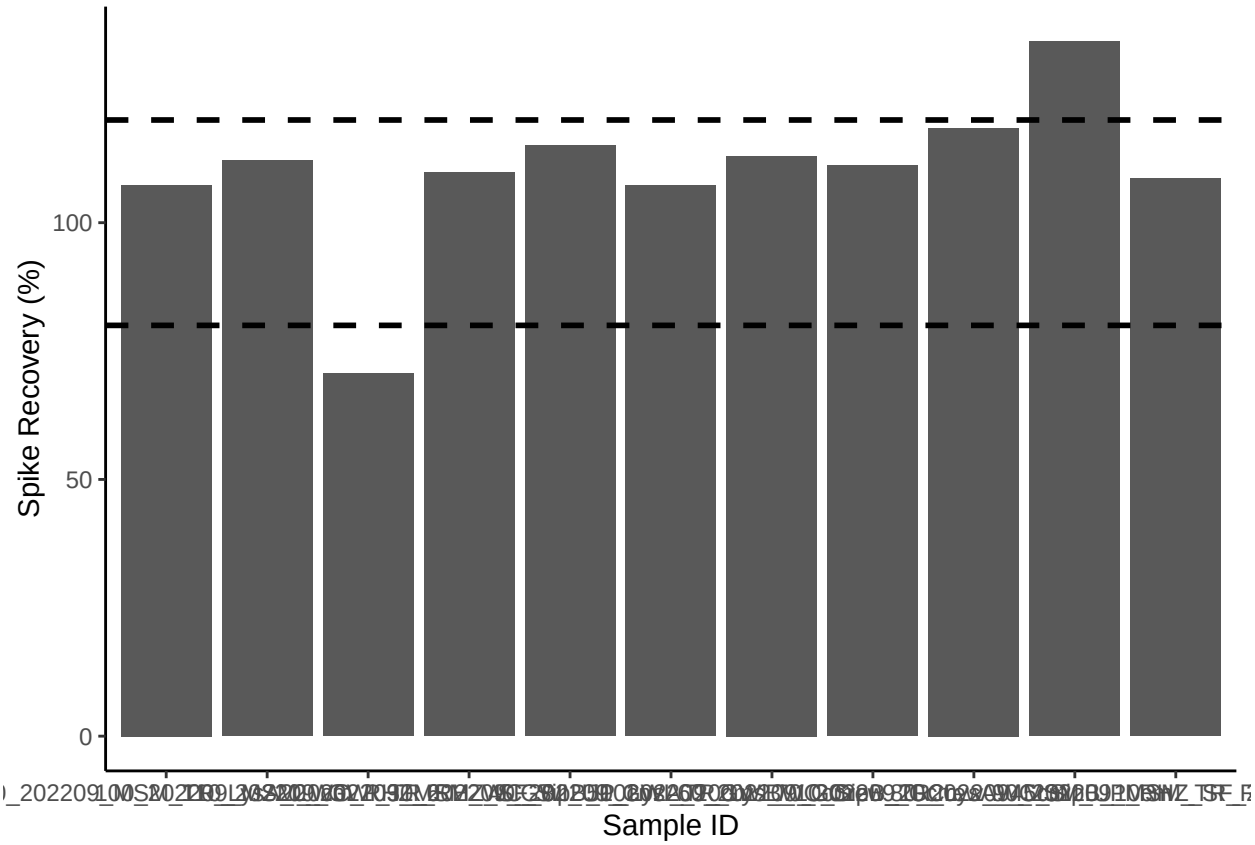
```
#need to determine dilution factors and initial amount of sample added
#if your samples are all the same dilution just use the first line
#for Steph / COMPASS this depends on the site so...
QAspks$Dilution <- 1
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 200, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$Dilution)

#Set Sample volumes in uL
QAspks$SampleVol <- 1
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "MSM_WC"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_UP"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GCrew_WC"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 1475, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 1462, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 100, QAspks$SampleVol)
#QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 100, QAspks$SampleVol)

#change sample volume to L
QAspks$SampleVol <- QAspks$SampleVol/1000000
head(QAspks)
```

```
##                   Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM  salinity
## 1 10_202209_MSM_TR_LysA_20cm  505.3590  3550.881 15.76291 100.1659  6.416469
## 2    100_202209_MSM_WC_RHZ_6  794.4041  8421.266 24.77867 237.5533 15.217253
## 3 110_202209_GWI_TR_RHZ_SF_8  962.0158  6500.256 30.00673 183.3641 11.745989
## 4 20_202209_MSM_WC_SipB_10cm  860.1581  9655.185 26.82964 272.3606 17.446945
## 5 30_202209_GWI_UP_LysA_20cm  658.0800  5681.553 20.52651 160.2695 10.266592
## 6 40_202209_GWI_TR_LysB_10cm 1698.7295 12719.267 52.98595 358.7946 22.983742
##   SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1   19.57811 7.797879e-05       50  0.001501
## 2   33.46188 7.797879e-05      100  0.001475
## 3   24.75468 7.797879e-05      100  0.001475
## 4   35.00240 7.797879e-05      100  0.001475
## 5   29.47750 7.797879e-05      100  0.001475
## 6   62.10149 7.797879e-05      100  0.001475
```

```r
#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, reru

head(QAspks)
```

```
##                    Sample_ID   SO4_ppm    Cl_ppm   SO4_mM    Cl_mM  salinity
## 1 10_202209_MSM_TR_LysA_20cm  505.3590  3550.881 15.76291 100.1659  6.416469
## 2    100_202209_MSM_WC_RHZ_6  794.4041  8421.266 24.77867 237.5533 15.217253
## 3 110_202209_GWI_TR_RHZ_SF_8  962.0158  6500.256 30.00673 183.3641 11.745989
## 4 20_202209_MSM_WC_SipB_10cm  860.1581  9655.185 26.82964 272.3606 17.446945
## 5 30_202209_GWI_UP_LysA_20cm  658.0800  5681.553 20.52651 160.2695 10.266592
## 6 40_202209_GWI_TR_LysB_10cm 1698.7295 12719.267 52.98595 358.7946 22.983742
##   SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd SO4_Total_spkd
## 1   19.57811 7.797879e-05       50  0.001501     0.0004732027   0.0005916504
## 2   33.46188 7.797879e-05      100  0.001475     0.0003654854   0.0004969090
## 3   24.75468 7.797879e-05      100  0.001475     0.0004425993   0.0003676069
## 4   35.00240 7.797879e-05      100  0.001475     0.0003957371   0.0005197857
## 5   29.47750 7.797879e-05      100  0.001475     0.0003027661   0.0004377409
## 6   62.10149 7.797879e-05      100  0.001475     0.0007815427   0.0009222072
##   SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1    0.0005511814    107.34222           YES
## 2    0.0004434641    112.05167           YES
## 3    0.0005205781     70.61514     NO, rerun
## 4    0.0004737159    109.72519           YES
## 5    0.0003807449    114.96960           YES
## 6    0.0008595215    107.29309           YES
```

```r
#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           2    11    18.2
## 2 YES                 9    11    81.8
```

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", "202209", "MSM", "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 202209  MSM   UP      LysA  20cm    1
## 2          10 202209  MSM   TR      LysA  20cm   10
## 3         100 202209  MSM   WC       RHZ     6  100
## 4         101 202209  MSM   WC       RHZ     7  101
## 5         102 202209  MSM   WC       RHZ     8  102
## 6         103 202209  GWI   TR       RHZ    SF    1
```

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

```
##   Analysis_No   Date Site Zone Replicate Depth Tree                  Sample_ID
## 1           1 202209  MSM   UP      LysA  20cm    1  1_202209_MSM_UP_LysA_20cm
## 2          10 202209  MSM   TR      LysA  20cm   10 10_202209_MSM_TR_LysA_20cm
## 3         100 202209  MSM   WC       RHZ     6  100    100_202209_MSM_WC_RHZ_6
## 4         101 202209  MSM   WC       RHZ     7  101    101_202209_MSM_WC_RHZ_7
## 5         102 202209  MSM   WC       RHZ     8  102    102_202209_MSM_WC_RHZ_8
## 6         103 202209  GWI   TR       RHZ    SF    1 103_202209_GWI_TR_RHZ_SF_1
##    SO4_ppm   Cl_ppm    SO4_mM     Cl_mM  salinity
## 1 185.4786 2155.484  5.785359  60.80349  3.894985
## 2 505.3590 3550.881 15.762913 100.16590  6.416469
## 3 794.4041 8421.266 24.778668 237.55333 15.217253
## 4 815.7621 8675.486 25.444857 244.72456 15.676629
## 5 872.0692 8465.177 27.201160 238.79201 15.296600
## 6 951.8538 6184.023 29.689763 174.44353 11.174556
```

Make final dataframe with IDs

```
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_Final_202209.csv")  #Chan
```

END