Dionex_COMPASS_July2022

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_202207_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 4.303   0.5203     6.97   0.5091     3.17
```

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

```
##          X.1 IC.SO4.1
## 1  Lab Blank     n.a.
## 2  Lab Blank     n.a.
## 3  Lab Blank     n.a.
## 4  Lab Blank     n.a.
## 5  Lab Blank     n.a.
## 6 Standard 1  0.5203
```

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

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

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

```
##    Sample_ID SO4_ppm
## 1  Lab Blank      NA
## 2  Lab Blank      NA
## 3  Lab Blank      NA
## 4  Lab Blank      NA
## 5  Lab Blank      NA
## 6 Standard 1  0.5203
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202207_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 3.210  0.0084    6.93  0.0111    0.08
## 3 3  Lab Blank             Unknown 3.207  0.0154    6.85  0.0204    0.12
## 4 4  Lab Blank             Unknown 3.207  0.0162    7.07  0.0214    0.12
## 5 5  Lab Blank             Unknown 3.207  0.0192   14.60  0.0253    0.16
## 6 6 Standard 1 Calibration Standard 3.210  5.1132   92.48  6.7527   46.95
```

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

```
##          X.1 IC.Cl.1
## 1  Lab Blank    n.a.
## 2  Lab Blank  0.0084
## 3  Lab Blank  0.0154
## 4  Lab Blank  0.0162
## 5  Lab Blank  0.0192
## 6 Standard 1 5.1132
```

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

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

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

```
##    Sample_ID Cl_ppm
## 1  Lab Blank     NA
## 2  Lab Blank 0.0084
## 3  Lab Blank 0.0154
## 4  Lab Blank 0.0162
## 5  Lab Blank 0.0192
## 6 Standard 1 5.1132
```

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

```
##   Sample_ID   SO4_ppm     Cl_ppm
## 1           1635.7937 11154.3742
## 2           1635.7937  2859.5887
## 3           1635.7937     0.0084
## 4           1635.7937  2990.3171
## 5           1635.7937         NA
## 6            410.8412 11154.3742
```

```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_202207_GCrew_UP_LysB_10cm   3.3588   12.7223
## 27     10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462
## 28 10_202207_GCrew_TR_LysA_20cm_dup  22.7271    4.0763
## 29 10_202207_GCrew_TR_LysA_20cm_spk 117.7331    5.1920
## 30     100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358
## 31     101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360
```

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

```
##                           Sample_ID  SO4_ppm    Cl_ppm
## 26      1_202207_GCrew_UP_LysB_10cm   3.3588   12.7223
## 27     10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462
## 28 10_202207_GCrew_TR_LysA_20cm_dup  22.7271    4.0763
## 29 10_202207_GCrew_TR_LysA_20cm_spk 117.7331    5.1920
## 30     100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358
## 31     101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360
```

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
## 987 Standard 1  0.5203   5.1132
## 988 Standard 1  0.5203   5.1051
## 989 Standard 1  0.5203 255.5158
## 990 Standard 1  0.5203   5.2549
## 991 Standard 1  0.5203   5.1533
## 992 Standard 1  0.5198   5.1132
```

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_202207_GCrew_UP_LysB_10cm   3.3588   12.7223
## 27     10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462
## 28 10_202207_GCrew_TR_LysA_20cm_dup  22.7271    4.0763
## 29 10_202207_GCrew_TR_LysA_20cm_spk 117.7331    5.1920
## 30     100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358
## 31     101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360
```

```
# 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_202207_GCrew_UP_LysB_10cm   3.3588   12.7223  0.1047661  0.3588801
## 27     10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462  0.7033874  0.1197800
## 28 10_202207_GCrew_TR_LysA_20cm_dup  22.7271    4.0763  0.7088927  0.1149873
## 29 10_202207_GCrew_TR_LysA_20cm_spk 117.7331    5.1920  3.6722739  0.1464598
## 30     100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358 13.9548191 99.2393738
## 31     101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360 13.7340954 85.3973484
##       salinity
## 26 0.023015196
## 27 0.007698883
## 28 0.007391874
## 29 0.009407944
## 30 6.357116691
## 31 5.470422152
```

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_202207_GCrew_UP_LysB_10cm   3.3588   12.7223  0.1047661  0.3588801
## 27     10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462  0.7033874  0.1197800
## 28 10_202207_GCrew_TR_LysA_20cm_dup  22.7271    4.0763  0.7088927  0.1149873
## 29 10_202207_GCrew_TR_LysA_20cm_spk 117.7331    5.1920  3.6722739  0.1464598
## 30     100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358 13.9548191 99.2393738
## 31     101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360 13.7340954 85.3973484
##       salinity
## 26 0.023015196
## 27 0.007698883
## 28 0.007391874
## 29 0.009407944
## 30 6.357116691
## 31 5.470422152
```

```
#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 10_202207_GCrew_TR_LysA_20cm_dup   22.7271    4.0763  0.7088927   0.1149873
## 2 102_202207_MSM_TR_RHZ_Tree_8_dup  503.8052 3693.2665 15.7144479 104.1824118
## 3 20_202207_GCrew_WC_SipB_20cm_dup  532.6823 3891.5436 16.6151684 109.7755599
## 4   30_202207_GWI_UP_LysC_10cm_dup  691.7529 5065.9526 21.5768216 142.9041636
## 5   40_202207_GWI_TR_LysB_20cm_dup  403.8785 8042.8357 12.5975827 226.8782990
## 6   50_202207_GWI_WC_SipC_20cm_dup 1107.7006 7977.1769 34.5508609 225.0261467
##       salinity
## 1  0.007391874
## 2  6.673758566
## 3  7.032045285
## 4  9.154202348
## 5 14.533430110
## 6 14.414784658
```

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

```
##                       Sample_ID  SO4_ppm    Cl_ppm     SO4_mM       Cl_mM
## 1  1_202207_GCrew_UP_LysB_10cm   3.3588   12.7223  0.1047661   0.3588801
## 2 10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462  0.7033874   0.1197800
## 3 100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358 13.9548191  99.2393738
## 4 101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360 13.7340954  85.3973484
```

```
## 5 102_202207_MSM_TR_RHZ_Tree_8 499.3243 3655.9609 15.5746818 103.1300677
## 6 11_202207_GCrew_TR_LysB_20cm   4.1594   24.0380  0.1297380   0.6780818
##      salinity
## 1 0.023015196
## 2 0.007698883
## 3 6.357116691
## 4 5.470422152
## 5 6.606347346
## 6 0.043462666
```

```
#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 10_202207_GCrew_TR_LysA_20cm  0.7088927   0.1149873  0.007391874
## 2 102_202207_MSM_TR_RHZ_Tree_8 15.7144479 104.1824118  6.673758566
## 3 20_202207_GCrew_WC_SipB_20cm 16.6151684 109.7755599  7.032045285
## 4   30_202207_GWI_UP_LysC_10cm 21.5768216 142.9041636  9.154202348
## 5   40_202207_GWI_TR_LysB_20cm 12.5975827 226.8782990 14.533430110
## 6   50_202207_GWI_WC_SipC_20cm 34.5508609 225.0261467 14.414784658
```

```
#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 10_202207_GCrew_TR_LysA_20cm   22.5506    4.2462  0.7033874   0.11978
## 2 102_202207_MSM_TR_RHZ_Tree_8  499.3243 3655.9609 15.5746818 103.13007
## 3 20_202207_GCrew_WC_SipB_20cm  554.4386 4062.0082 17.2937804 114.58415
## 4   30_202207_GWI_UP_LysC_10cm  702.7388 5105.3620 21.9194885 144.01585
## 5   40_202207_GWI_TR_LysB_20cm  386.4381 7630.9817 12.0535901 215.26041
## 6   50_202207_GWI_WC_SipC_20cm 1063.0874 7666.9799 33.1593075 216.27588
##       salinity SO4_mM_dup   Cl_mM_dup salinity_dup
## 1  0.007698883  0.7088927   0.1149873  0.007391874
## 2  6.606347346 15.7144479 104.1824118  6.673758566
## 3  7.340074817 16.6151684 109.7755599  7.032045285
## 4  9.225415134 21.5768216 142.9041636  9.154202348
## 5 13.789209932 12.5975827 226.8782990 14.533430110
## 6 13.854258679 34.5508609 225.0261467 14.414784658
```

```
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)
```

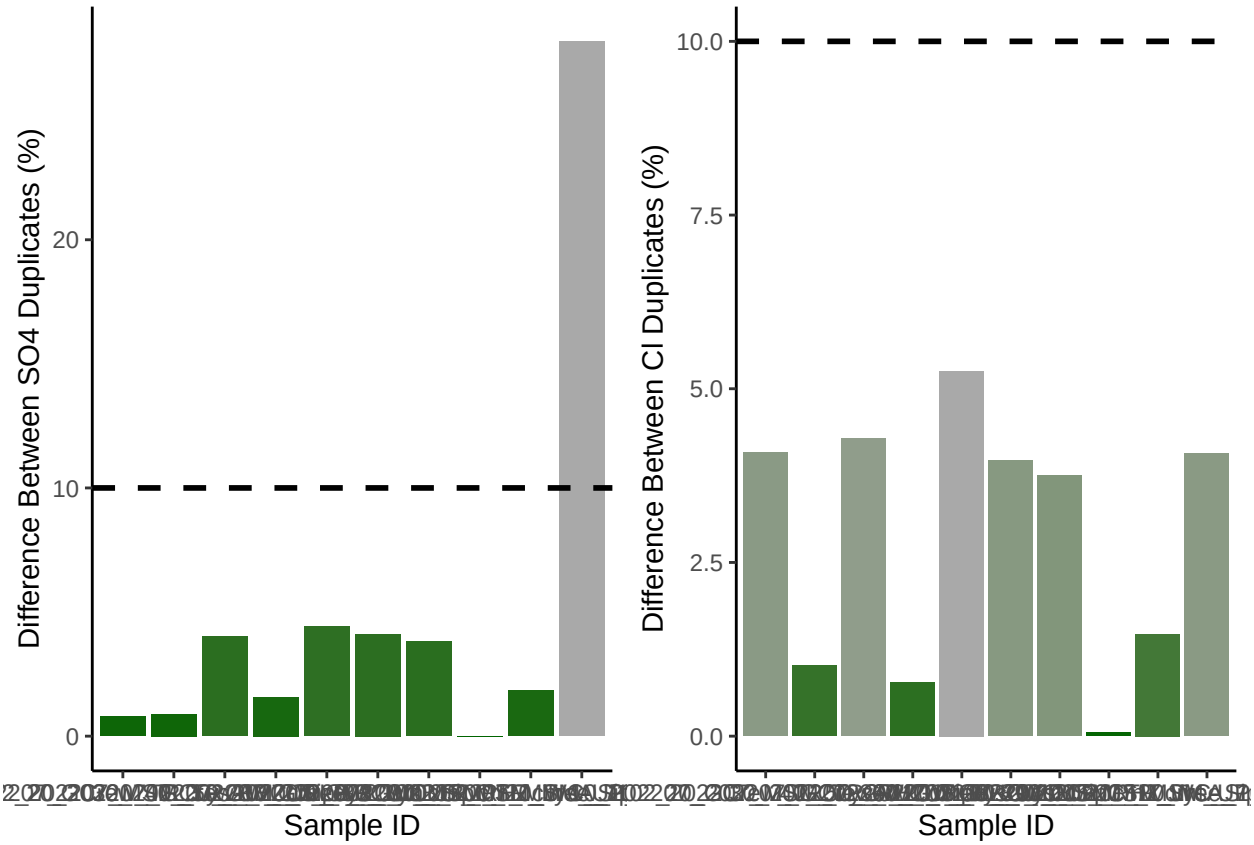
```
##                      Sample_ID   SO4_ppm    Cl_ppm     SO4_mM     Cl_mM
```

```
## 1 10_202207_GCrew_TR_LysA_20cm   22.5506    4.2462  0.7033874   0.11978
## 2 102_202207_MSM_TR_RHZ_Tree_8  499.3243 3655.9609 15.5746818 103.13007
## 3 20_202207_GCrew_WC_SipB_20cm  554.4386 4062.0082 17.2937804 114.58415
## 4   30_202207_GWI_UP_LysC_10cm  702.7388 5105.3620 21.9194885 144.01585
## 5   40_202207_GWI_TR_LysB_20cm  386.4381 7630.9817 12.0535901 215.26041
## 6   50_202207_GWI_WC_SipC_20cm 1063.0874 7666.9799 33.1593075 216.27588
##       salinity SO4_mM_dup   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.007698883  0.7088927   0.1149873  0.007391874    0.7796332           YES
## 2  6.606347346 15.7144479 104.1824118  6.673758566    0.8933842           YES
## 3  7.340074817 16.6151684 109.7755599  7.032045285    4.0025539           YES
## 4  9.225415134 21.5768216 142.9041636  9.154202348    1.5756135           YES
## 5 13.789209932 12.5975827 226.8782990 14.533430110    4.4135224           YES
## 6 13.854258679 34.5508609 225.0261467 14.414784658    4.1103231           YES
##   Cl_dups_chk Cl_dups_flag
## 1   4.0829078          YES
## 2   1.0152251          YES
## 3   4.2865025          YES
## 4   0.7749126          YES
## 5   5.2553120          YES
## 6   3.9656596          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)
```

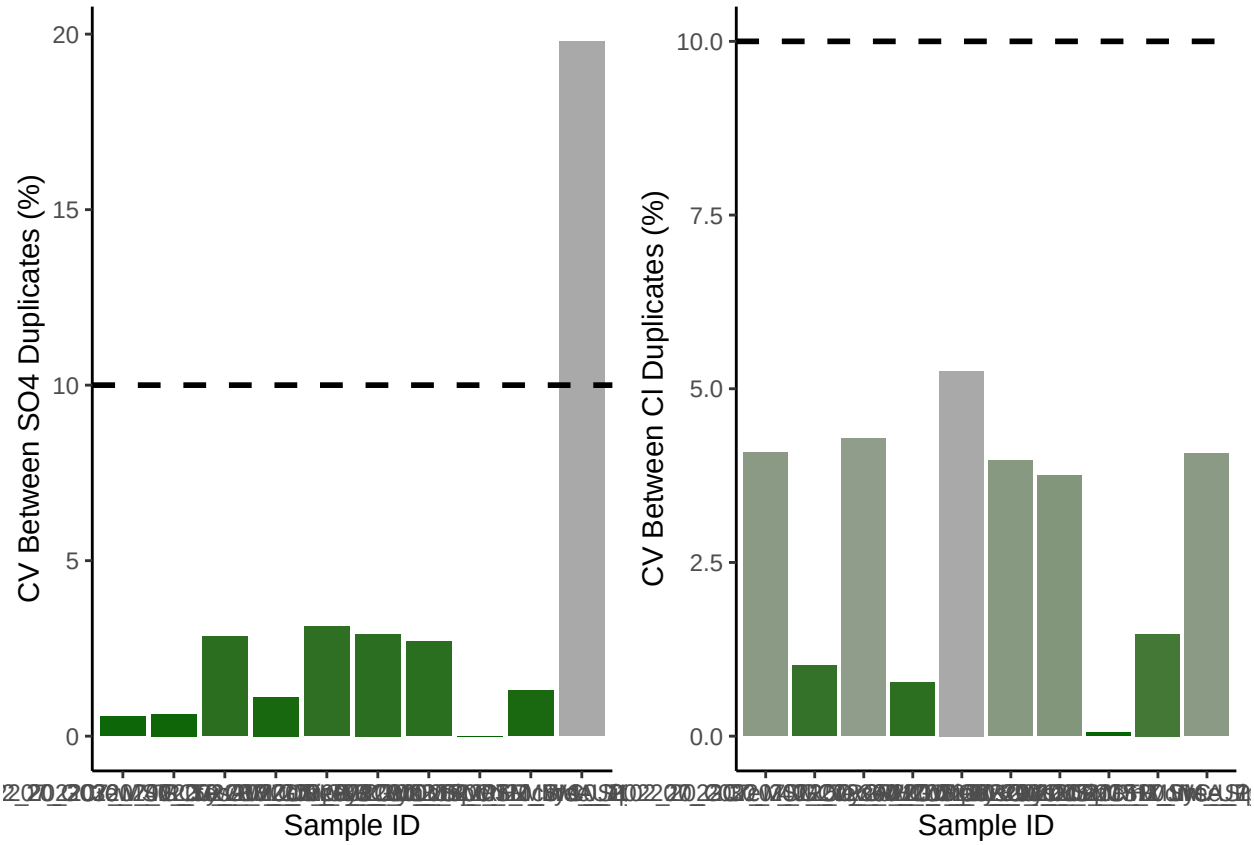
```
##        Flag S_no_rows Flag Cl_no_rows Total S_Percent Cl_Percent
## 1 NO, rerun         1  YES         10    10        10         100
## 2       YES         9  YES         10    10        90         100
```

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 10_202207_GCrew_TR_LysA_20cm   22.5506    4.2462  0.7033874   0.11978
## 2 102_202207_MSM_TR_RHZ_Tree_8  499.3243 3655.9609 15.5746818 103.13007
## 3 20_202207_GCrew_WC_SipB_20cm  554.4386 4062.0082 17.2937804 114.58415
## 4   30_202207_GWI_UP_LysC_10cm  702.7388 5105.3620 21.9194885 144.01585
## 5   40_202207_GWI_TR_LysB_20cm  386.4381 7630.9817 12.0535901 215.26041
## 6   50_202207_GWI_WC_SipC_20cm 1063.0874 7666.9799 33.1593075 216.27588
##       salinity SO4_mM_dup   Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.007698883  0.7088927   0.1149873  0.007391874    0.7796332           YES
## 2  6.606347346 15.7144479 104.1824118  6.673758566    0.8933842           YES
## 3  7.340074817 16.6151684 109.7755599  7.032045285    4.0025539           YES
## 4  9.225415134 21.5768216 142.9041636  9.154202348    1.5756135           YES
## 5 13.789209932 12.5975827 226.8782990 14.533430110    4.4135224           YES
## 6 13.854258679 34.5508609 225.0261467 14.414784658    4.1103231           YES
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1   4.0829078          YES    0.551284              YES
## 2   1.0152251          YES    0.631718              YES
## 3   4.2865025          YES    2.830233              YES
## 4   0.7749126          YES    1.114127              YES
## 5   5.2553120          YES    3.120832              YES
## 6   3.9656596          YES    2.906437              YES
```

```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  YES         10    10        10         100
## 2       YES         9  YES         10    10        90         100
```

Pull out spikes and check

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

```
##                            Sample_ID  SO4_ppm    Cl_ppm     SO4_mM      Cl_mM
## 26      1_202207_GCrew_UP_LysB_10cm   3.3588   12.7223  0.1047661  0.3588801
## 27     10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462  0.7033874  0.1197800
## 28 10_202207_GCrew_TR_LysA_20cm_dup  22.7271    4.0763  0.7088927  0.1149873
## 29 10_202207_GCrew_TR_LysA_20cm_spk 117.7331    5.1920  3.6722739  0.1464598
## 30     100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358 13.9548191 99.2393738
## 31     101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360 13.7340954 85.3973484
##       salinity
## 26 0.023015196
## 27 0.007698883
## 28 0.007391874
## 29 0.009407944
## 30 6.357116691
## 31 5.470422152
```

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

```
##                           Sample_ID   SO4_ppm   Cl_ppm    SO4_mM       Cl_mM
## 1 10_202207_GCrew_TR_LysA_20cm_spk  117.7331    5.192  3.672274   0.1464598
## 2 20_202207_GCrew_WC_SipB_20cm_spk  725.9881 3905.243 22.644669 110.1620028
## 3   30_202207_GWI_UP_LysC_10cm_spk  863.2929 5146.109 26.927414 145.1652863
## 4   40_202207_GWI_TR_LysB_20cm_spk  535.5176 7186.561 16.703606 202.7238561
## 5   50_202207_GWI_WC_SipC_20cm_spk 1505.9961 8057.812 46.974301 227.3007588
## 6  60_202207_GWI_TR_RHZ_Tree_7_spk  787.8655 4674.293 24.574719 131.8559436
##       salinity
## 1  0.009407944
## 2  7.056800101
## 3  9.299045686
## 4 12.986141185
## 5 14.560492103
## 6  8.446473812
```

```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_202207_GCrew_TR_LysA_20cm   3.672274
## 2 20_202207_GCrew_WC_SipB_20cm  22.644669
## 3   30_202207_GWI_UP_LysC_10cm  26.927414
## 4   40_202207_GWI_TR_LysB_20cm  16.703606
## 5   50_202207_GWI_WC_SipC_20cm  46.974301
## 6  60_202207_GWI_TR_RHZ_Tree_7  24.574719
```

```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 10_202207_GCrew_TR_LysA_20cm   22.5506    4.2462  0.7033874   0.11978
## 2 20_202207_GCrew_WC_SipB_20cm  554.4386 4062.0082 17.2937804 114.58415
## 3   30_202207_GWI_UP_LysC_10cm  702.7388 5105.3620 21.9194885 144.01585
## 4   40_202207_GWI_TR_LysB_20cm  386.4381 7630.9817 12.0535901 215.26041
## 5   50_202207_GWI_WC_SipC_20cm 1063.0874 7666.9799 33.1593075 216.27588
## 6  60_202207_GWI_TR_RHZ_Tree_7  659.9456 5156.1324 20.5847037 145.44802
##       salinity SO4_mM_spk
## 1  0.007698883   3.672274
## 2  7.340074817  22.644669
## 3  9.225415134  26.927414
## 4 13.789209932  16.703606
## 5 13.854258679  46.974301
## 6  9.317157247  24.574719
```

```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
## 1 10_202207_GCrew_TR_LysA_20cm   22.5506    4.2462  0.7033874   0.11978
## 2 20_202207_GCrew_WC_SipB_20cm  554.4386 4062.0082 17.2937804 114.58415
## 3   30_202207_GWI_UP_LysC_10cm  702.7388 5105.3620 21.9194885 144.01585
## 4   40_202207_GWI_TR_LysB_20cm  386.4381 7630.9817 12.0535901 215.26041
## 5   50_202207_GWI_WC_SipC_20cm 1063.0874 7666.9799 33.1593075 216.27588
## 6  60_202207_GWI_TR_RHZ_Tree_7  659.9456 5156.1324 20.5847037 145.44802
##       salinity SO4_mM_spk SO4_spk_Conc
## 1  0.007698883   3.672274 7.797879e-05
## 2  7.340074817  22.644669 7.797879e-05
## 3  9.225415134  26.927414 7.797879e-05
## 4 13.789209932  16.703606 7.797879e-05
## 5 13.854258679  46.974301 7.797879e-05
## 6  9.317157247  24.574719 7.797879e-05
```

```r
#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)
```

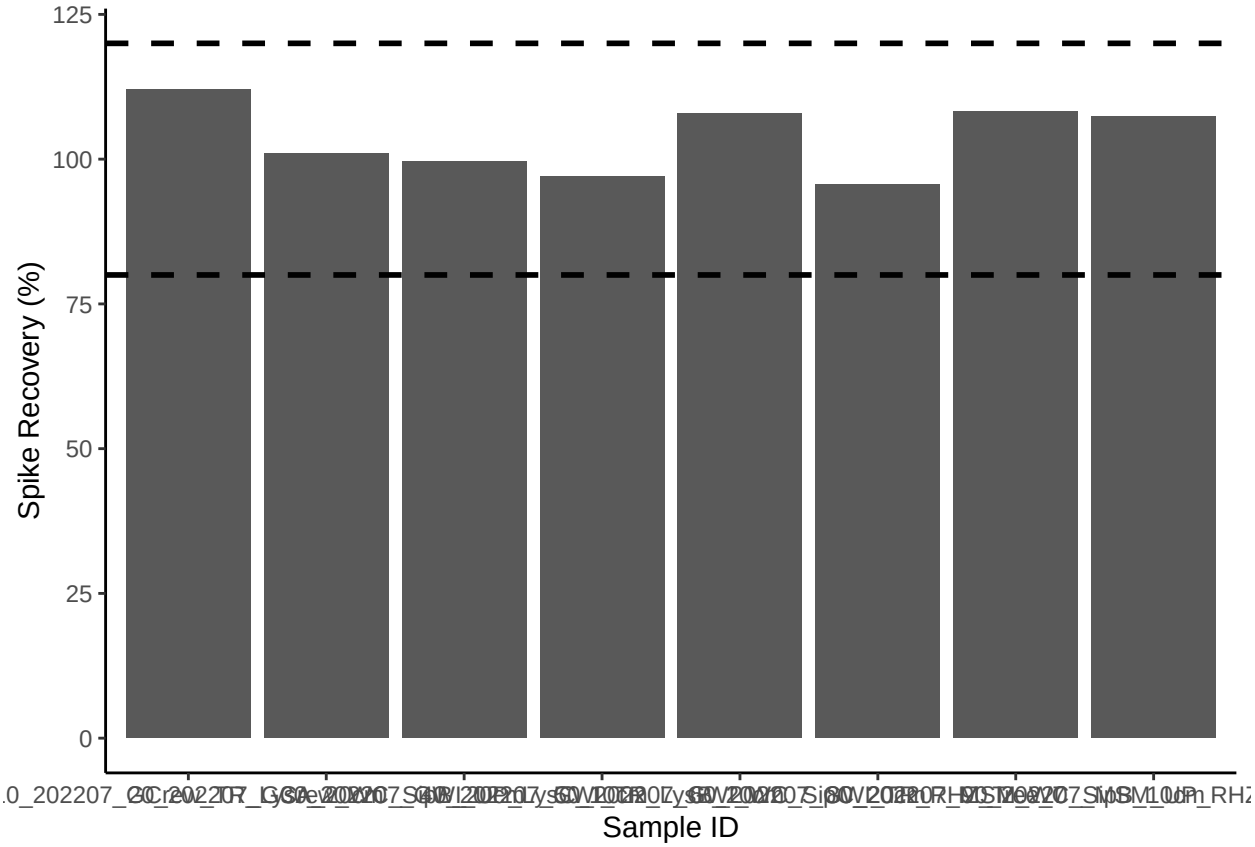
```
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_UP"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_TR"), 100, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "GWI_WC"), 200, QAspks$Dilution)
#QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_UP"), 100, QAspks$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 10_202207_GCrew_TR_LysA_20cm   22.5506    4.2462  0.7033874   0.11978
## 2 20_202207_GCrew_WC_SipB_20cm  554.4386 4062.0082 17.2937804 114.58415
## 3   30_202207_GWI_UP_LysC_10cm  702.7388 5105.3620 21.9194885 144.01585
## 4   40_202207_GWI_TR_LysB_20cm  386.4381 7630.9817 12.0535901 215.26041
## 5   50_202207_GWI_WC_SipC_20cm 1063.0874 7666.9799 33.1593075 216.27588
## 6  60_202207_GWI_TR_RHZ_Tree_7  659.9456 5156.1324 20.5847037 145.44802
##       salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1  0.007698883   3.672274 7.797879e-05       50  0.001501
## 2  7.340074817  22.644669 7.797879e-05      100  0.001475
## 3  9.225415134  26.927414 7.797879e-05      100  0.001475
## 4 13.789209932  16.703606 7.797879e-05      100  0.001475
## 5 13.854258679  46.974301 7.797879e-05      200  0.001462
## 6  9.317157247  24.574719 7.797879e-05      100  0.001475
```

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

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

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

head(QAspks)
```

```
##                     Sample_ID   SO4_ppm    Cl_ppm     SO4_mM     Cl_mM
## 1 10_202207_GCrew_TR_LysA_20cm   22.5506    4.2462  0.7033874   0.11978
## 2 20_202207_GCrew_WC_SipB_20cm  554.4386 4062.0082 17.2937804 114.58415
## 3   30_202207_GWI_UP_LysC_10cm  702.7388 5105.3620 21.9194885 144.01585
## 4   40_202207_GWI_TR_LysB_20cm  386.4381 7630.9817 12.0535901 215.26041
## 5   50_202207_GWI_WC_SipC_20cm 1063.0874 7666.9799 33.1593075 216.27588
## 6  60_202207_GWI_TR_RHZ_Tree_7  659.9456 5156.1324 20.5847037 145.44802
##       salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1  0.007698883   3.672274 7.797879e-05       50  0.001501     2.111569e-05
## 2  7.340074817  22.644669 7.797879e-05      100  0.001475     2.550833e-04
## 3  9.225415134  26.927414 7.797879e-05      100  0.001475     3.233125e-04
## 4 13.789209932  16.703606 7.797879e-05      100  0.001475     1.777905e-04
## 5 13.854258679  46.974301 7.797879e-05      200  0.001462     2.423945e-04
## 6  9.317157247  24.574719 7.797879e-05      100  0.001475     3.036244e-04
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   0.0001109761    9.909448e-05    111.99021           YES
## 2   0.0003362733    3.330621e-04    100.96417           YES
## 3   0.0003998721    4.012912e-04     99.64636           YES
## 4   0.0002480485    2.557692e-04     96.98138           YES
## 5   0.0003457309    3.203733e-04    107.91499           YES
## 6   0.0003649346    3.816032e-04     95.63196           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: 1 x 4
##   SO4_spks_flag no_rows Total Percent
##   <chr>           <int> <int>   <dbl>
## 1 YES                 8     8     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", "202207", "GCrew", "UP", "LysB", "10cm"), c("10", :
## number of columns of result is not a multiple of vector length (arg 1)
```

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

```
##   Analysis_No   Date  Site Zone Replicate Depth Tree_ID
## 1           1 202207 GCrew   UP      LysB  10cm       1
## 2          10 202207 GCrew   TR      LysA  20cm      10
## 3         100 202207   MSM   TR       RHZ  Tree       6
## 4         101 202207   MSM   TR       RHZ  Tree       7
## 5         102 202207   MSM   TR       RHZ  Tree       8
## 6          11 202207 GCrew   TR      LysB  20cm      11
```

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

```
##   Analysis_No   Date  Site Zone Replicate Depth Tree_ID
## 1           1 202207 GCrew   UP      LysB  10cm       1
## 2          10 202207 GCrew   TR      LysA  20cm      10
## 3         100 202207   MSM   TR       RHZ  Tree       6
## 4         101 202207   MSM   TR       RHZ  Tree       7
## 5         102 202207   MSM   TR       RHZ  Tree       8
## 6          11 202207 GCrew   TR      LysB  20cm      11
##                      Sample_ID  SO4_ppm    Cl_ppm     SO4_mM       Cl_mM
## 1  1_202207_GCrew_UP_LysB_10cm   3.3588   12.7223  0.1047661   0.3588801
## 2 10_202207_GCrew_TR_LysA_20cm  22.5506    4.2462  0.7033874   0.1197800
## 3 100_202207_MSM_TR_RHZ_Tree_6 447.3915 3518.0358 13.9548191  99.2393738
## 4 101_202207_MSM_TR_RHZ_Tree_7 440.3151 3027.3360 13.7340954  85.3973484
## 5 102_202207_MSM_TR_RHZ_Tree_8 499.3243 3655.9609 15.5746818 103.1300677
## 6 11_202207_GCrew_TR_LysB_20cm   4.1594   24.0380  0.1297380   0.6780818
##      salinity
## 1 0.023015196
## 2 0.007698883
## 3 6.357116691
## 4 5.470422152
## 5 6.606347346
## 6 0.043462666
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202207.csv")      #Change
```

END