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_202305_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 4.887   0.0084    22.74   0.0132     0.06
## 2 2  Lab Blank             Unknown 4.907   0.0091    23.29   0.0143     0.06
## 3 3  Lab Blank             Unknown 4.907   0.0094    32.59   0.0148     0.06
## 4 4  Lab Blank             Unknown 4.890   0.0090    32.12   0.0142     0.06
## 5 5 Standard 1 Calibration Standard 4.893   0.4337     7.79   0.6859     3.45
## 6 6 Standard 2 Calibration Standard 4.900   1.0909    10.03   1.7251     6.45
##   IC.SO4.5
## 1      BMB
## 2      BMB
## 3      BMB
## 4      BMB
## 5       Rd
## 6       M
```

```
## 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  0.0084
## 2  Lab Blank  0.0091
## 3  Lab Blank  0.0094
## 4  Lab Blank  0.0090
## 5 Standard 1  0.4337
## 6 Standard 2  1.0909
```

```
## 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  0.0084
## 2  Lab Blank  0.0091
## 3  Lab Blank  0.0094
## 4  Lab Blank  0.0090
## 5 Standard 1  0.4337
## 6 Standard 2  1.0909
```

```
#Chloride data
## Read in raw data file from Dionex - copied and saved as a txt
Cldat <- read.table("Raw Data/COMPASS_Synoptic_CB_MonMon_202305_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 3.890  0.0086   22.84   0.0133    0.05
## 2 2  Lab Blank             Unknown 3.857  0.0093   23.54   0.0145    0.05
## 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 Standard 1 Calibration Standard 3.900  5.1547   90.74   7.9918   33.37
## 6 6 Standard 2 Calibration Standard 3.860  9.9035   89.27  15.3544   65.84
##   IC.Cl.5
## 1      MB
## 2      MB
## 3    n.a.
## 4    n.a.
## 5       M
## 6       M
```

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

```
##          X.1 IC.Cl.1
## 1  Lab Blank 0.0086
## 2  Lab Blank 0.0093
## 3  Lab Blank    n.a.
## 4  Lab Blank    n.a.
## 5 Standard 1 5.1547
## 6 Standard 2 9.9035
```

```
## 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 0.0086
## 2  Lab Blank 0.0093
## 3  Lab Blank     NA
## 4  Lab Blank     NA
## 5 Standard 1 5.1547
## 6 Standard 2 9.9035
```

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

```
##   Sample_ID  SO4_ppm    Cl_ppm
## 1           911.5208 8679.8805
## 2           911.5208 3359.8059
## 3           911.5208    0.0086
## 4           911.5208 2923.7016
## 5           911.5208        NA
## 6           306.9453 8679.8805
```

```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_GCW_202305_UP_lysA_20cm  6.9968 66.1749
## 27     10_GCW_202305_TR_LysB_45cm 10.4414 75.9797
## 28 10_GCW_202305_TR_LysB_45cm_dup 10.4558 75.9221
## 29     11_GCW_202305_TR_LysC_10cm  8.6995 77.5035
## 30 11_GCW_202305_TR_LysC_10cm_spk  8.4098 78.3821
## 31     12_GCW_202305_TR_LysC_20cm 10.0383 80.5907
```

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

```
##                          Sample_ID SO4_ppm  Cl_ppm
## 26      1_GCW_202305_UP_lysA_20cm  6.9968 66.1749
## 27     10_GCW_202305_TR_LysB_45cm 10.4414 75.9797
## 28 10_GCW_202305_TR_LysB_45cm_dup 10.4558 75.9221
## 29     11_GCW_202305_TR_LysC_10cm  8.6995 77.5035
## 30 11_GCW_202305_TR_LysC_10cm_spk  8.4098 78.3821
## 31     12_GCW_202305_TR_LysC_20cm 10.0383 80.5907
```

Pull out standards - could do some checks here if we want

```
stds <- all_dat[grepl("Standard", all_dat$Sample_ID),]
head(stds)
```

```
##      Sample_ID SO4_ppm Cl_ppm
## 311 Standard 1  0.4337 5.1547
## 312 Standard 1  0.4337 5.2745
## 313 Standard 1  0.4337 5.2025
## 314 Standard 1  0.4337 5.1604
## 315 Standard 1  0.4622 5.1547
## 316 Standard 1  0.4622 5.2745
```

```
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  0.444 0.0111  2.51 NO, rerun
## 2 Standard 2  1.02  0.165  16.2  NO, rerun
## 3 Standard 3  2.06  0.0230  1.11 YES
## 4 Standard 4  9.97  0.165   1.66 YES
## 5 Standard 5 20.0   0.219   1.10 YES
```

```
stds_chk_Cl <- stds %>%
    group_by(Sample_ID) %>%
    summarise(mean = mean(Cl_ppm), sd = sd(Cl_ppm))
stds_chk_Cl$cv <- (stds_chk_Cl$sd/stds_chk_Cl$mean)*100
stds_chk_Cl$flag <-  ifelse(stds_chk_Cl$cv <2, 'YES', 'NO, rerun')
head(stds_chk_Cl)
```

```
## # A tibble: 5 x 5
##   Sample_ID    mean     sd    cv flag
##   <fct>       <dbl>  <dbl> <dbl> <chr>
## 1 Standard 1   5.20 0.0494 0.951 YES
## 2 Standard 2  10.1  0.150  1.49  YES
## 3 Standard 3  20.0  0.192  0.956 YES
## 4 Standard 4 102.   0.860  0.845 YES
## 5 Standard 5 202.   1.58   0.785 YES
```

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_GCW_202305_UP_lysA_20cm  6.9968 66.1749
## 27     10_GCW_202305_TR_LysB_45cm 10.4414 75.9797
## 28 10_GCW_202305_TR_LysB_45cm_dup 10.4558 75.9221
## 29     11_GCW_202305_TR_LysC_10cm  8.6995 77.5035
## 30 11_GCW_202305_TR_LysC_10cm_spk  8.4098 78.3821
## 31     12_GCW_202305_TR_LysC_20cm 10.0383 80.5907
```

```
# 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_GCW_202305_UP_lysA_20cm  6.9968 66.1749 0.2182408 1.866711 0.1196040
## 27     10_GCW_202305_TR_LysB_45cm 10.4414 75.9797 0.3256831 2.143292 0.1373213
## 28 10_GCW_202305_TR_LysB_45cm_dup 10.4558 75.9221 0.3261323 2.141667 0.1372172
## 29     11_GCW_202305_TR_LysC_10cm  8.6995 77.5035 0.2713506 2.186276 0.1400748
## 30 11_GCW_202305_TR_LysC_10cm_spk  8.4098 78.3821 0.2623144 2.211061 0.1416625
## 31     12_GCW_202305_TR_LysC_20cm 10.0383 80.5907 0.3131098 2.273362 0.1456534
```

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_GCW_202305_UP_lysA_20cm  6.9968 66.1749 0.2182408 1.866711 0.1196040
## 27     10_GCW_202305_TR_LysB_45cm 10.4414 75.9797 0.3256831 2.143292 0.1373213
## 28 10_GCW_202305_TR_LysB_45cm_dup 10.4558 75.9221 0.3261323 2.141667 0.1372172
## 29     11_GCW_202305_TR_LysC_10cm  8.6995 77.5035 0.2713506 2.186276 0.1400748
## 30 11_GCW_202305_TR_LysC_10cm_spk  8.4098 78.3821 0.2623144 2.211061 0.1416625
## 31     12_GCW_202305_TR_LysC_20cm 10.0383 80.5907 0.3131098 2.273362 0.1456534
```

```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 10_GCW_202305_TR_LysB_45cm_dup  10.4558   75.9221  0.3261323   2.141667
## 2 20_GCW_202305_WC_LysC_10cm_dup 389.4078 4084.1041 12.1462196 115.207450
## 3 30_MSM_202305_UP_LysB_45cm_dup 125.9109 1726.2511  3.9273518  48.695377
## 4 40_MSM_202305_TR_LysC_10cm_dup 396.7954 3822.0579 12.3766500 107.815456
## 5 50_MSM_202305_WC_LysC_20cm_dup 680.4251 7416.7719 21.2234903 209.217825
## 6   60_MSM_202305_RHZ_UP_SF6_dup  46.5377  763.0388  1.4515814  21.524367
##     salinity
## 1  0.1372172
## 2  7.3800021
## 3  3.1193617
## 4  6.9064846
## 5 13.4021328
## 6  1.3788371
```

```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_GCW_202305_UP_lysA_20cm   6.9968   66.1749  0.2182408   1.866711 0.1196040
## 2 10_GCW_202305_TR_LysB_45cm  10.4414   75.9797  0.3256831   2.143292 0.1373213
## 3 11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276 0.1400748
## 4 12_GCW_202305_TR_LysC_20cm  10.0383   80.5907  0.3131098   2.273362 0.1456534
## 5 13_GCW_202305_TR_LysC_45cm  22.4937   81.0706  0.7016126   2.286900 0.1465206
## 6 14_GCW_202305_WC_LysA_10cm 437.5432 4637.3172 13.6476357 130.812897 8.3796582
```

```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 10_GCW_202305_TR_LysB_45cm  0.3261323   2.141667    0.1372172
## 2 20_GCW_202305_WC_LysC_10cm 12.1462196 115.207450    7.3800021
## 3 30_MSM_202305_UP_LysB_45cm  3.9273518  48.695377    3.1193617
## 4 40_MSM_202305_TR_LysC_10cm 12.3766500 107.815456    6.9064846
## 5 50_MSM_202305_WC_LysC_20cm 21.2234903 209.217825   13.4021328
## 6   60_MSM_202305_RHZ_UP_SF6  1.4515814  21.524367    1.3788371
```

```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 10_GCW_202305_TR_LysB_45cm  10.4414   75.9797  0.3256831   2.143292
## 2 20_GCW_202305_WC_LysC_10cm 386.8014 4143.9913 12.0649220 116.896793
## 3 30_MSM_202305_UP_LysB_45cm 125.9329 1724.7658  3.9280381  48.653478
## 4 40_MSM_202305_TR_LysC_10cm 409.6475 3908.6209 12.7775265 110.257289
## 5 50_MSM_202305_WC_LysC_20cm 671.1855 7304.4209 20.9352932 206.048544
## 6   60_MSM_202305_RHZ_UP_SF6  45.7372  761.3144  1.4266126  21.475724
##     salinity SO4_mM_dup  Cl_mM_dup salinity_dup
## 1  0.1373213  0.3261323   2.141667    0.1372172
## 2  7.4882183 12.1462196 115.207450    7.3800021
## 3  3.1166778  3.9273518  48.695377    3.1193617
## 4  7.0629040 12.3766500 107.815456    6.9064846
## 5 13.1991146 21.2234903 209.217825   13.4021328
## 6  1.3757211  1.4515814  21.524367    1.3788371
```

```
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_GCW_202305_TR_LysB_45cm  10.4414   75.9797  0.3256831   2.143292
## 2 20_GCW_202305_WC_LysC_10cm 386.8014 4143.9913 12.0649220 116.896793
## 3 30_MSM_202305_UP_LysB_45cm 125.9329 1724.7658  3.9280381  48.653478
## 4 40_MSM_202305_TR_LysC_10cm 409.6475 3908.6209 12.7775265 110.257289
## 5 50_MSM_202305_WC_LysC_20cm 671.1855 7304.4209 20.9352932 206.048544
## 6   60_MSM_202305_RHZ_UP_SF6  45.7372  761.3144  1.4266126  21.475724
##     salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.1373213  0.3261323   2.141667    0.1372172   0.13781751           YES
## 2  7.4882183 12.1462196 115.207450    7.3800021   0.67157153           YES
## 3  3.1166778  3.9273518  48.695377    3.1193617   0.01747115           YES
## 4  7.0629040 12.3766500 107.815456    6.9064846   3.18735524           YES
## 5 13.1991146 21.2234903 209.217825   13.4021328   1.36719851           YES
## 6  1.3757211  1.4515814  21.524367    1.3788371   1.73503304           YES
##   Cl_dups_chk Cl_dups_flag
## 1  0.07583847          YES
## 2  1.45567588          YES
## 3  0.08607898          YES
## 4  2.23946699          YES
## 5  1.52638447          YES
## 6  0.22624678          YES
```

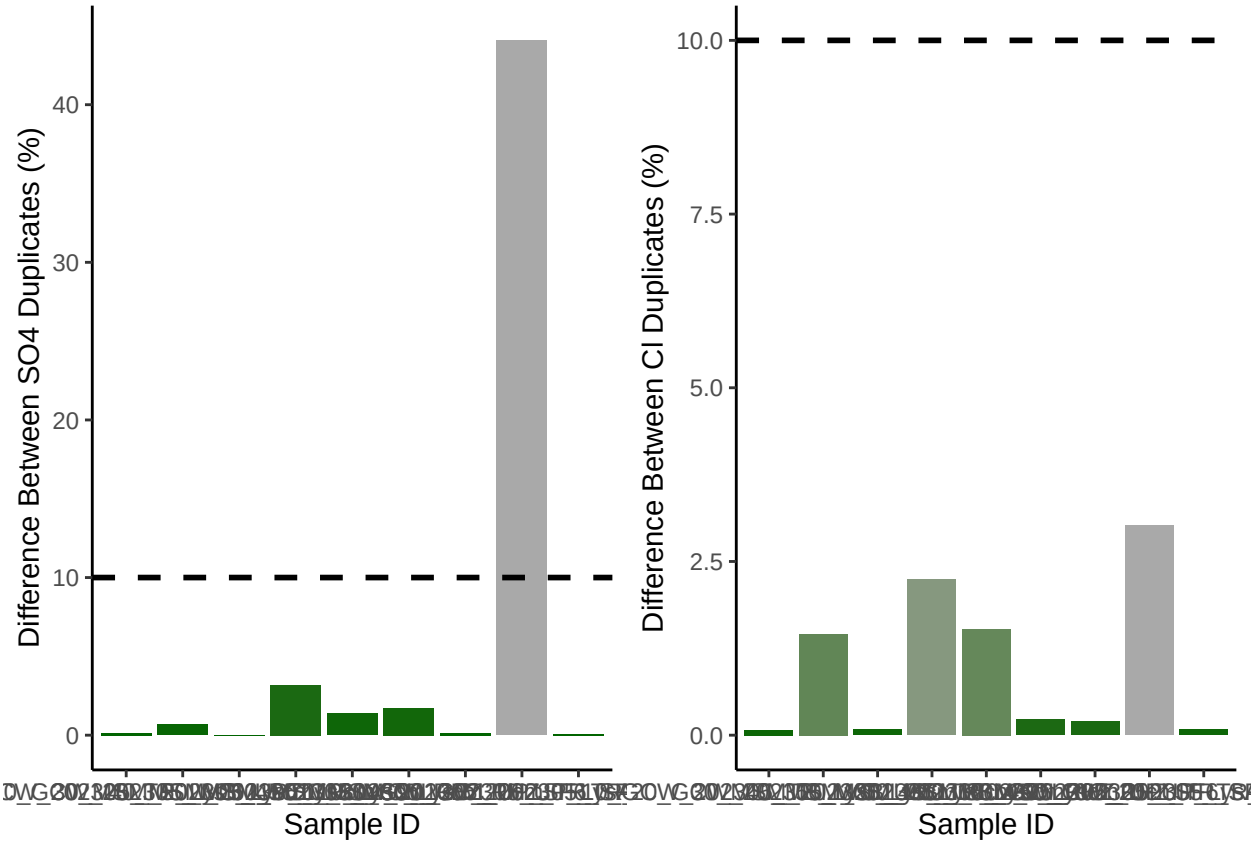
```
#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
Sdupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = SO4_dups_chk, fill=SO4_dups_chk)) +
       geom_bar(stat = 'identity') +
        scale_fill_gradient2(low='darkgreen', mid='darkgreen', high='darkgrey', 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          9     9  11.11111        100
## 2       YES         8  YES          9     9  88.88889        100
```

Pull out dups and check with cv

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

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

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

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

head(QAdups)
```

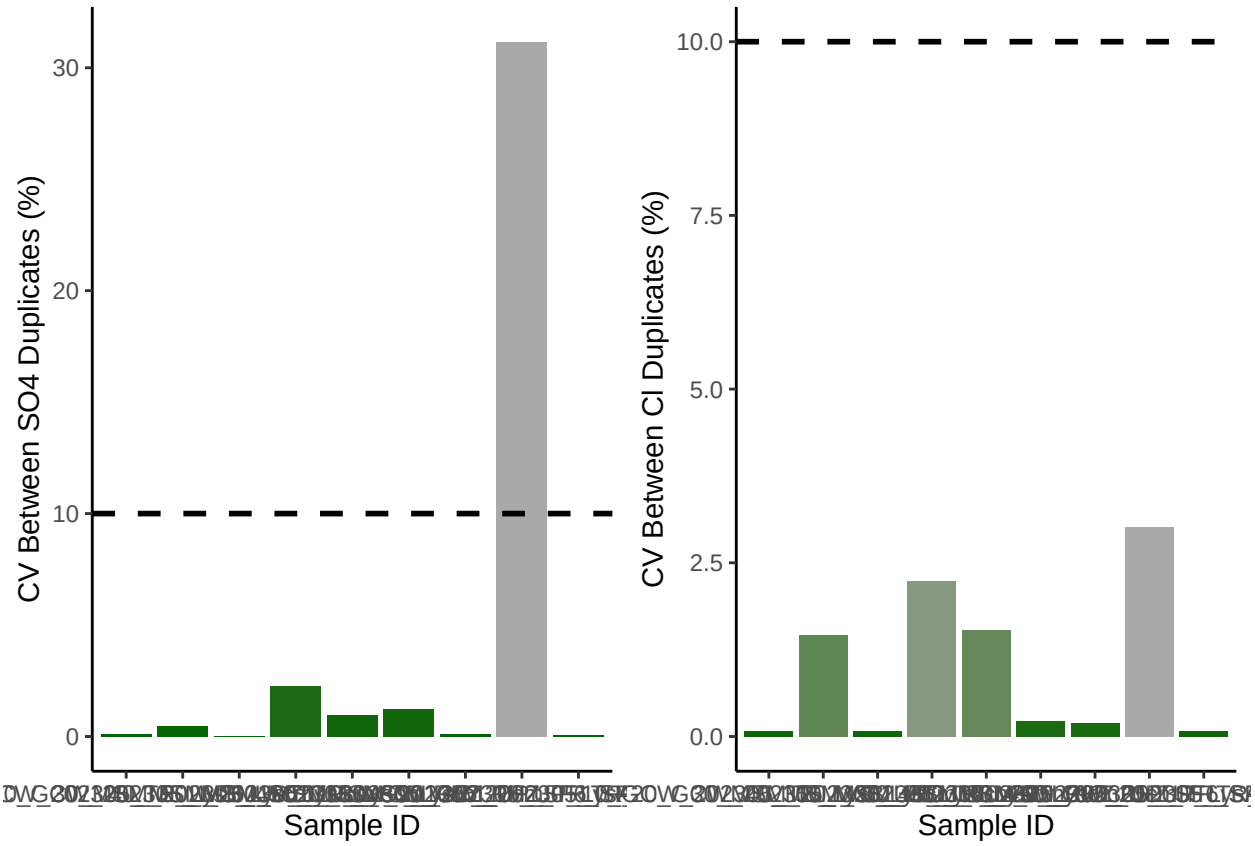
```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM      Cl_mM
## 1 10_GCW_202305_TR_LysB_45cm  10.4414   75.9797  0.3256831   2.143292
## 2 20_GCW_202305_WC_LysC_10cm 386.8014 4143.9913 12.0649220 116.896793
## 3 30_MSM_202305_UP_LysB_45cm 125.9329 1724.7658  3.9280381  48.653478
## 4 40_MSM_202305_TR_LysC_10cm 409.6475 3908.6209 12.7775265 110.257289
## 5 50_MSM_202305_WC_LysC_20cm 671.1855 7304.4209 20.9352932 206.048544
## 6   60_MSM_202305_RHZ_UP_SF6  45.7372  761.3144  1.4266126  21.475724
##     salinity SO4_mM_dup  Cl_mM_dup salinity_dup SO4_dups_chk SO4_dups_flag
## 1  0.1373213  0.3261323   2.141667    0.1372172   0.13781751           YES
## 2  7.4882183 12.1462196 115.207450    7.3800021   0.67157153           YES
## 3  3.1166778  3.9273518  48.695377    3.1193617   0.01747115           YES
## 4  7.0629040 12.3766500 107.815456    6.9064846   3.18735524           YES
## 5 13.1991146 21.2234903 209.217825   13.4021328   1.36719851           YES
## 6  1.3757211  1.4515814  21.524367    1.3788371   1.73503304           YES
##   Cl_dups_chk Cl_dups_flag SO4_dups_cv SO4_dups_cv_flag
## 1  0.07583847          YES  0.09745169              YES
## 2  1.45567588          YES  0.47487278              YES
## 3  0.08607898          YES  0.01235397              YES
## 4  2.23946699          YES  2.25380050              YES
## 5  1.52638447          YES  0.96675534              YES
## 6  0.22624678          YES  1.22685363              YES
```

```
#plot dups output as a bar graph to easily check - want any over 10% to be red need to work on this
Sdupsbar <- ggplot(data = QAdups, aes(x = Sample_ID, y = SO4_dups_cv, fill=SO4_dups_chk)) +
       geom_bar(stat = 'identity') +
        scale_fill_gradient2(low='darkgreen', mid='darkgreen', high='darkgrey', space='Lab') +
```

```
        theme_classic() + labs(x= "Sample ID", y="CV Between SO4 Duplicates (%)") +
        theme(legend.position="none") +  geom_hline(yintercept=10, linetype="dashed",
                color = "black", size=1)


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

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




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

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

```
##        Flag S_no_rows Flag Cl_no_rows Total S_Percent Cl_Percent
## 1 NO, rerun         1  YES          9     9  11.11111        100
## 2       YES         8  YES          9     9  88.88889        100
```

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  salinity
## 26      1_GCW_202305_UP_lysA_20cm  6.9968 66.1749 0.2182408 1.866711 0.1196040
## 27     10_GCW_202305_TR_LysB_45cm 10.4414 75.9797 0.3256831 2.143292 0.1373213
## 28 10_GCW_202305_TR_LysB_45cm_dup 10.4558 75.9221 0.3261323 2.141667 0.1372172
## 29     11_GCW_202305_TR_LysC_10cm  8.6995 77.5035 0.2713506 2.186276 0.1400748
## 30 11_GCW_202305_TR_LysC_10cm_spk  8.4098 78.3821 0.2623144 2.211061 0.1416625
## 31     12_GCW_202305_TR_LysC_20cm 10.0383 80.5907 0.3131098 2.273362 0.1456534
```

```
#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 11_GCW_202305_TR_LysC_10cm_spk   8.4098   78.3821  0.2623144   2.211061
## 2 21_GCW_202305_WC_LysC_20cm_spk 418.9831 4870.7450 13.0687180 137.397602
## 3 31_MSM_202305_UP_LysC_10cm_spk  85.7760  747.6468  2.6754835  21.090178
## 4 41_MSM_202305_TR_LysC_20cm_spk 371.5310 3464.3339 11.5886151  97.724511
## 5 51_MSM_202305_WC_LysC_45cm_spk 291.3934 7870.2669  9.0890019 222.010350
## 6   61_MSM_202305_RHZ_UP_SF7_spk  75.5919 1252.5837  2.3578260  35.333814
##     salinity
## 1  0.1416625
## 2  8.8014622
## 3  1.3510238
## 4  6.2600774
## 5 14.2215983
## 6  2.2634447
```

```
#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 11_GCW_202305_TR_LysC_10cm  0.2623144
## 2 21_GCW_202305_WC_LysC_20cm 13.0687180
## 3 31_MSM_202305_UP_LysC_10cm  2.6754835
## 4 41_MSM_202305_TR_LysC_20cm 11.5886151
## 5 51_MSM_202305_WC_LysC_45cm  9.0890019
## 6   61_MSM_202305_RHZ_UP_SF7  2.3578260
```

```
#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 11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 2 21_GCW_202305_WC_LysC_20cm 488.8230 4727.0862 15.2471304 133.345168
## 3 31_MSM_202305_UP_LysC_10cm  62.9016  707.2833  1.9619963  19.951574
## 4 41_MSM_202305_TR_LysC_20cm 303.1789 3523.8549  9.4566095  99.403523
## 5 51_MSM_202305_WC_LysC_45cm 154.8114 8488.3719  4.8288022 239.446316
## 6   61_MSM_202305_RHZ_UP_SF7  62.2802 1262.2168  1.9426138  35.605551
##     salinity SO4_mM_spk
## 1  0.1400748  0.2623144
## 2  8.5418708 13.0687180
## 3  1.2780869  2.6754835
## 4  6.3676318 11.5886151
## 5 15.3385140  9.0890019
## 6  2.2808518  2.3578260
```

```
#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 11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 2 21_GCW_202305_WC_LysC_20cm 488.8230 4727.0862 15.2471304 133.345168
## 3 31_MSM_202305_UP_LysC_10cm  62.9016  707.2833  1.9619963  19.951574
## 4 41_MSM_202305_TR_LysC_20cm 303.1789 3523.8549  9.4566095  99.403523
## 5 51_MSM_202305_WC_LysC_45cm 154.8114 8488.3719  4.8288022 239.446316
## 6   61_MSM_202305_RHZ_UP_SF7  62.2802 1262.2168  1.9426138  35.605551
##     salinity SO4_mM_spk SO4_spk_Conc
## 1  0.1400748  0.2623144 7.797879e-05
## 2  8.5418708 13.0687180 7.797879e-05
## 3  1.2780869  2.6754835 7.797879e-05
## 4  6.3676318 11.5886151 7.797879e-05
## 5 15.3385140  9.0890019 7.797879e-05
## 6  2.2808518  2.3578260 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"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 50, QAspks$Dilution)
QAspks$Dilution <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 50, 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"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_TR"), 1501, QAspks$SampleVol)
QAspks$SampleVol <-  ifelse(str_detect(QAspks$Sample_ID, "SWH_WC"), 1501, 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 11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 2 21_GCW_202305_WC_LysC_20cm 488.8230 4727.0862 15.2471304 133.345168
## 3 31_MSM_202305_UP_LysC_10cm  62.9016  707.2833  1.9619963  19.951574
## 4 41_MSM_202305_TR_LysC_20cm 303.1789 3523.8549  9.4566095  99.403523
## 5 51_MSM_202305_WC_LysC_45cm 154.8114 8488.3719  4.8288022 239.446316
## 6   61_MSM_202305_RHZ_UP_SF7  62.2802 1262.2168  1.9426138  35.605551
##     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol
## 1  0.1400748  0.2623144 7.797879e-05        1     1e-06
## 2  8.5418708 13.0687180 7.797879e-05        1     1e-06
## 3  1.2780869  2.6754835 7.797879e-05        1     1e-06
## 4  6.3676318 11.5886151 7.797879e-05        1     1e-06
## 5 15.3385140  9.0890019 7.797879e-05        1     1e-06
## 6  2.2808518  2.3578260 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)
```

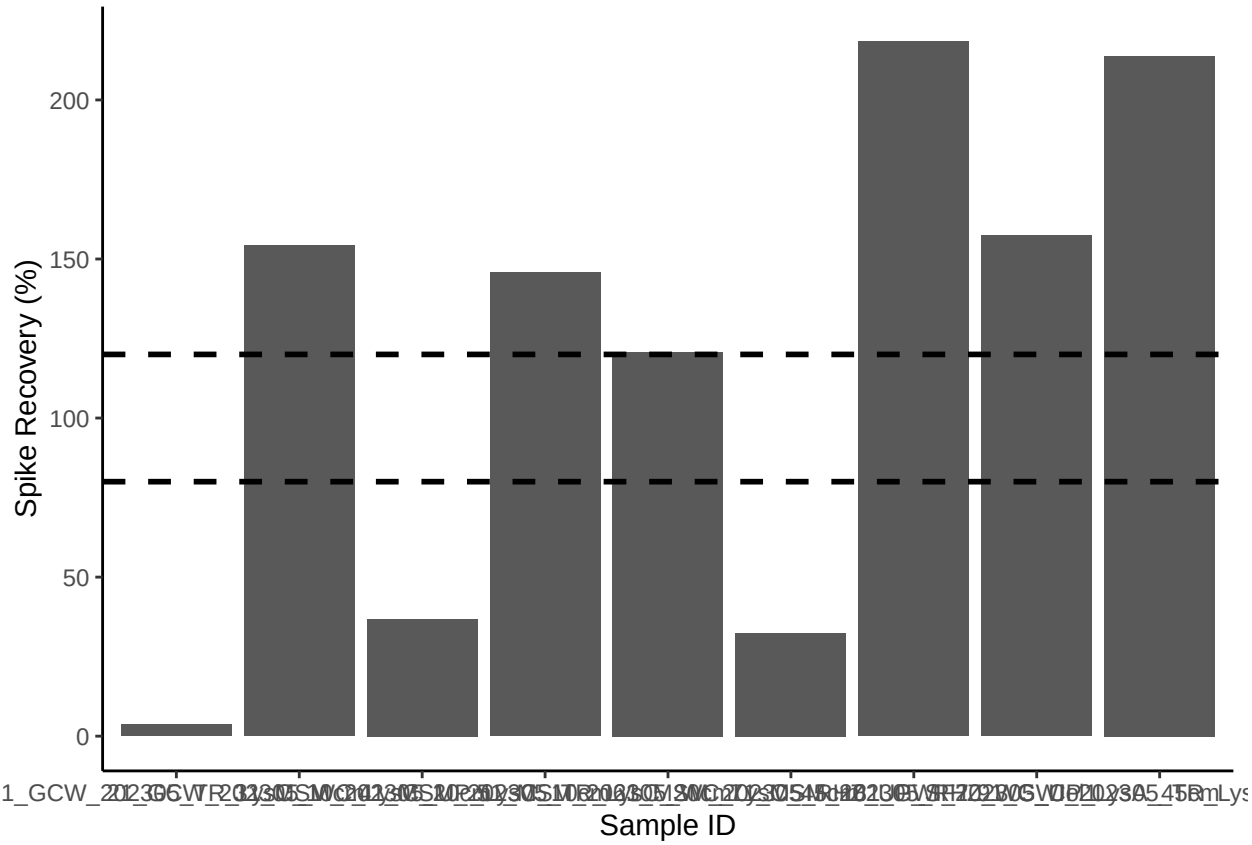
```
QAspks$SO4_expctd_spkd <-  (QAspks$SO4_Total_unspkd + QAspks$SO4_spk_Conc)
QAspks$spk_recovery <-    (QAspks$SO4_Total_spkd/QAspks$SO4_expctd_spkd)*100
QAspks$SO4_spks_flag <-  ifelse(QAspks$spk_recovery <=120 & QAspks$spk_recovery >=80 , 'YES', 'NO, rerun

head(QAspks)
```

```
##                    Sample_ID  SO4_ppm    Cl_ppm     SO4_mM      Cl_mM
## 1 11_GCW_202305_TR_LysC_10cm   8.6995   77.5035  0.2713506   2.186276
## 2 21_GCW_202305_WC_LysC_20cm 488.8230 4727.0862 15.2471304 133.345168
## 3 31_MSM_202305_UP_LysC_10cm  62.9016  707.2833  1.9619963  19.951574
## 4 41_MSM_202305_TR_LysC_20cm 303.1789 3523.8549  9.4566095  99.403523
## 5 51_MSM_202305_WC_LysC_45cm 154.8114 8488.3719  4.8288022 239.446316
## 6   61_MSM_202305_RHZ_UP_SF7  62.2802 1262.2168  1.9426138  35.605551
##     salinity SO4_mM_spk SO4_spk_Conc Dilution SampleVol SO4_Total_unspkd
## 1  0.1400748  0.2623144 7.797879e-05        1     1e-06     2.713506e-07
## 2  8.5418708 13.0687180 7.797879e-05        1     1e-06     1.524713e-05
## 3  1.2780869  2.6754835 7.797879e-05        1     1e-06     1.961996e-06
## 4  6.3676318 11.5886151 7.797879e-05        1     1e-06     9.456609e-06
## 5 15.3385140  9.0890019 7.797879e-05        1     1e-06     4.828802e-06
## 6  2.2808518  2.3578260 7.797879e-05        1     1e-06     1.942614e-06
##   SO4_Total_spkd SO4_expctd_spkd spk_recovery SO4_spks_flag
## 1   2.885459e-06    7.825014e-05      3.68748     NO, rerun
## 2   1.437559e-04    9.322592e-05    154.20164     NO, rerun
## 3   2.943032e-05    7.994079e-05     36.81515     NO, rerun
## 4   1.274748e-04    8.743540e-05    145.79309     NO, rerun
## 5   9.997902e-05    8.280759e-05    120.73654     NO, rerun
## 6   2.593609e-05    7.992140e-05     32.45199     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           9     9     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", "GCW", "202305", "UP", "lysA", "20cm"), c("10", :
## number of columns of result is not a multiple of vector length (arg 16)
```

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

```
##   Analysis_No Site   Date Zone Replicate Depth
## 1           1  GCW 202305   UP      lysA  20cm
## 2          10  GCW 202305   TR      LysB  45cm
## 3          11  GCW 202305   TR      LysC  10cm
## 4          12  GCW 202305   TR      LysC  20cm
## 5          13  GCW 202305   TR      LysC  45cm
## 6          14  GCW 202305   WC      LysA  10cm
```

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

```
##   Analysis_No Site   Date Zone Replicate Depth                  Sample_ID
## 1           1  GCW 202305   UP      lysA  20cm  1_GCW_202305_UP_lysA_20cm
## 2          10  GCW 202305   TR      LysB  45cm 10_GCW_202305_TR_LysB_45cm
## 3          11  GCW 202305   TR      LysC  10cm 11_GCW_202305_TR_LysC_10cm
## 4          12  GCW 202305   TR      LysC  20cm 12_GCW_202305_TR_LysC_20cm
## 5          13  GCW 202305   TR      LysC  45cm 13_GCW_202305_TR_LysC_45cm
## 6          14  GCW 202305   WC      LysA  10cm 14_GCW_202305_WC_LysA_10cm
##    SO4_ppm    Cl_ppm     SO4_mM      Cl_mM  salinity
## 1   6.9968   66.1749  0.2182408   1.866711 0.1196040
## 2  10.4414   75.9797  0.3256831   2.143292 0.1373213
## 3   8.6995   77.5035  0.2713506   2.186276 0.1400748
## 4  10.0383   80.5907  0.3131098   2.273362 0.1456534
## 5  22.4937   81.0706  0.7016126   2.286900 0.1465206
## 6 437.5432 4637.3172 13.6476357 130.812897 8.3796582
```

Make final dataframe with IDs

```r
write.csv(alldat, file="Processed Data/COMPASS_SynopticCB_PW_Processed_Cl_SO4_202305a.csv")          #
```

END