Check\_data\_integrity

Tobias Opialla

9/27/2019

***Nothing conclusive that would explain the huge shifts in CA-intensity by treatment or handling bias from experimental design or sample clean up.***

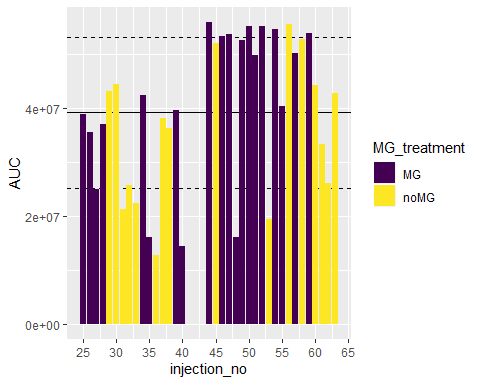
## -- Attaching packages -------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.2  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 0.8.3 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

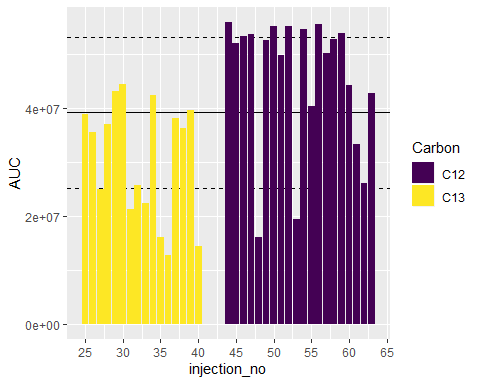
## -- Conflicts ----------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# cinnamic acid

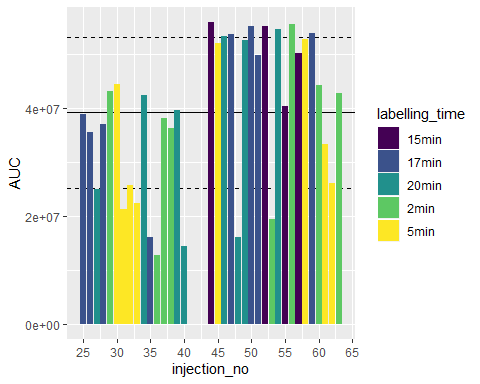
df\_cinnamic=read.csv('./outputfiles\_20190914/cinnamic\_acid\_09-14-2019\_19-54-24/matrices/quantMassAreas\_cinAcidNormMatrix.csv',stringsAsFactors = F)  
df\_sample\_metadata=read.xlsx("../Sample\_overview.xlsx",sheet = "Sample\_overview",rows=1:37) #row 38:43=Testsamples  
  
df\_cinnamic=gather(df\_cinnamic,key = "file",value = "AUC",-Metabolite,-QuantMasses)  
df\_cinnamic=df\_cinnamic %>% mutate(  
 injection\_no=str\_replace\_all(file,".\*\_to\_(\\d{3})\_.\*","\\1") %>%   
 as.numeric(),  
 sample\_no=str\_replace\_all(file,".\*\_to\_\\d{3}\_(\\d{2})\_.\*","\\1") %>%   
 as.numeric(),  
 MG\_treatment=ifelse(str\_detect(file,"\_MG\_"),"MG","noMG"),  
 Carbon=str\_replace\_all(file,".\*\_(C1[23])\_.\*","\\1"),  
 labelling\_time=str\_replace\_all(file,".\*\_(\\d+min)\_.\*","\\1"),  
 labelling\_time\_num=str\_remove\_all(labelling\_time,"min") %>%  
 as.numeric(),  
 Replicate=str\_replace\_all(file,".\*\_(R\\d)..\*","\\1"),  
 Replicate\_num=str\_remove\_all(Replicate,"R") %>% as.numeric()  
)  
  
df\_cinnamic=left\_join(df\_cinnamic,df\_sample\_metadata %>%  
 select(corrected.SampleNo.Fabiola,  
 cellnumber.according.to.conditions,  
 derivatization.same.MeOX.as.Ident.and.Quant),  
 by=c("sample\_no"="corrected.SampleNo.Fabiola"))  
  
p=ggplot(df\_cinnamic,aes(x=injection\_no,y=AUC))+  
 scale\_fill\_viridis\_d()+  
 scale\_x\_continuous(breaks = seq(0,65,5))+  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC,na.rm = T)+  
 sd(df\_cinnamic$AUC,na.rm = T),  
 color="black",linetype="dashed") +  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC),color="black")+  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC,na.rm = T) -  
 sd(df\_cinnamic$AUC,na.rm = T),  
 color="black",linetype="dashed")  
p+geom\_col(aes(fill=MG\_treatment))



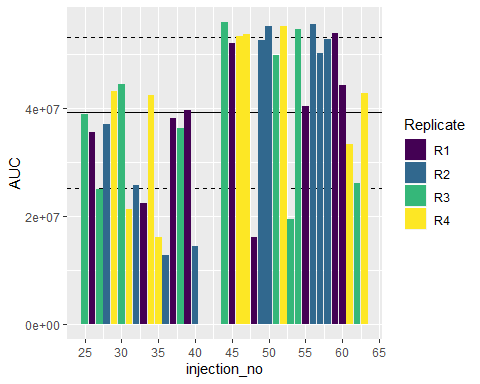
p+geom\_col(aes(fill=Carbon))



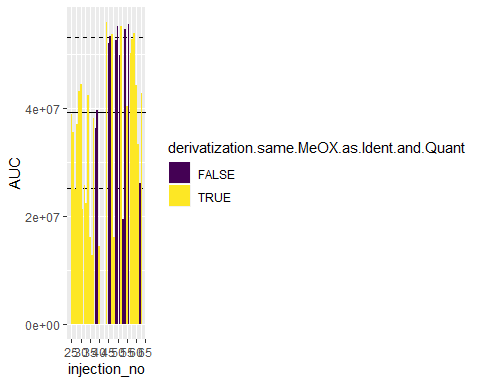
p+geom\_col(aes(fill=labelling\_time))



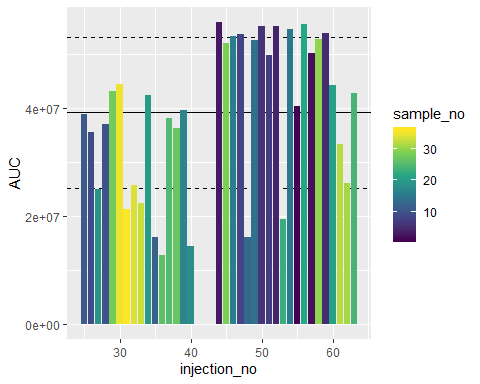
p+geom\_col(aes(fill=Replicate))



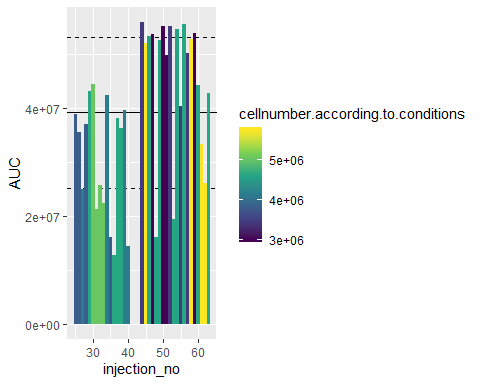
p+geom\_col(aes(fill=derivatization.same.MeOX.as.Ident.and.Quant))



ggplot(df\_cinnamic,aes(x=injection\_no,y=AUC))+  
 scale\_fill\_viridis\_c()+  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC,na.rm = T)+  
 sd(df\_cinnamic$AUC,na.rm = T),  
 color="black",linetype="dashed") +  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC),color="black")+  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC,na.rm = T) -  
 sd(df\_cinnamic$AUC,na.rm = T),  
 color="black",linetype="dashed")+  
 geom\_col(aes(fill=sample\_no))



ggplot(df\_cinnamic,aes(x=injection\_no,y=AUC))+  
 scale\_fill\_viridis\_c()+  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC,na.rm = T)+  
 sd(df\_cinnamic$AUC,na.rm = T),  
 color="black",linetype="dashed") +  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC),color="black")+  
 geom\_hline(yintercept = mean(df\_cinnamic$AUC,na.rm = T) -  
 sd(df\_cinnamic$AUC,na.rm = T),  
 color="black",linetype="dashed")+  
 geom\_col(aes(fill=cellnumber.according.to.conditions))



***Nothing conclusive that would explain the huge shifts in CA-intensity by treatment or handling bias from experimental design or sample clean up.***