Main Finding of outliers through Interceptor

Marlin

3/15/2022

"Files Used:"

Z:/COVID-19_WastewaterAnalysis/data/processed/MMSD_Interceptor_Cases_2_7_22.csv

Z:/COVID-19_WastewaterAnalysis/data/processed/LIMSWasteData_02-09-22.csv

```r
SizeUsed = 1
alphaUsed = .9
N1ShapeUnit = 4
N2ShapeUnit = 5
HeightJit <- 0
WidthJit <- 1



MadDF <- filter(FullDF,Site=="Madison")%>%
    mutate(FlagedOutliersN1 = IdentifyOutliers(N1,Bin = 21, Action = "Flag"),
           FlagedOutliersN2 = IdentifyOutliers(N2,Bin = 28, Action = "Flag"),
           NoOutlierVarN1 = ifelse(FlagedOutliersN1, NA, N1),
           NoOutlierVarN2 = ifelse(FlagedOutliersN2, NA, N2),)


NonOuliersDF <- MadDF%>%
  mutate(Outlier = ifelse(FlagedOutliersN1,N1,NA))%>%
           mutate(N1 = NoOutlierVarN1,
                  N2 = NoOutlierVarN2)

OutLierPlotDF <- MadDF%>%
  mutate(OutlierN1 = ifelse(FlagedOutliersN1,N1,NA),
         OutlierN2 = ifelse(FlagedOutliersN2,N2,NA))%>%
           mutate(N1 = NoOutlierVarN1,
                  N2 = NoOutlierVarN2)%>%
  #filter(!(is.na(N1)&is.na(Outlier)))%>%
  ggplot(aes(x=Date))+#Data depends on time
  geom_jitter(aes(y=N1, color="inlier", info = N1),
             data=NonOuliersDF, size = .5, shape = N1ShapeUnit, stroke = .2,
             width = WidthJit ,height = HeightJit )+
  geom_jitter(aes(y=N2, color="inlier",  info = N2),
             data=NonOuliersDF, size = .5, shape = N2ShapeUnit, stroke = .2,
             width = WidthJit ,height = HeightJit )+
  geom_jitter(aes(y=OutlierN1, color="Outlier", info = OutlierN1),
             shape=N1ShapeUnit, width = WidthJit ,height = HeightJit )+
  geom_jitter(aes(y=OutlierN2, color="Outlier", info = OutlierN2),
             shape=N2ShapeUnit, width = WidthJit ,height = HeightJit )+
```

```r
  theme_light() +
  #scale_y_log10()+
  scale_color_manual(values=c("#F8766D","#999999"))+
  ylab("Covid partical conventration (GC/L)")

ggplotly(OutLierPlotDF, tooltip = c("x","y","colour"))%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"))%>%
  config(displaylogo = FALSE)
```

```r
ggplotly(OutLierPlotDF+
           scale_y_continuous(limits = c(0, 5800000))


         , tooltip = c("x","y","colour"))%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"))%>%
  config(displaylogo = FALSE)
```

```r
#"2021-06-08","2021-10-17","2021-05-02","2021-01-26"
```

```r
IntercepterDF <- FullDF %>%
  group_by(Site)%>%
  mutate(FudgeFactor = mean(N1))%>%#Mean of sites to see if it works as normalizer
  filter(Site != "Madison")



RotSpacing <- seq(15, 375, length = 6)#Getting equal spaced colors
HueSpace <- hcl(h = RotSpacing, l = 65, c = 100)[1:5]#converting degrees to color
HueSpace <- c(tail(HueSpace, -3), head(HueSpace, 3))#Rotating so P18 has best color
HueSpace[2] <- "#000000" #Changing P18 to blac k


IntercepterOverLay <- IntercepterDF%>%
  filter(Date>mdy("1/1/2021"))%>%
  ggplot(aes(x=Date))+
  geom_point(aes(y=N1,color = Site),size = SizeUsed, alpha= alphaUsed,shape=N1ShapeUnit)+
  geom_point(aes(y=N2,color = Site),size = SizeUsed, alpha= alphaUsed,shape=N2ShapeUnit)+
  theme_light()+
  scale_colour_manual(values = HueSpace)+
  scale_y_log10()+
  ylab("Covid partical conventration (GC/L)")

ggplotly(IntercepterOverLay)%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"))%>%
  config(displaylogo = FALSE)
```

```r
#Get P18 to pop out
```

```r
IntercepterChangeDF <- IntercepterDF%>%
  filter(!is.na(N1))%>%
  mutate(ChangeN1 = lead(N1) - N1,
         ChangeN2 = lead(N2) - N2,
         PerChangeN1 = log(lead(N1) - N1),
         PerChangeN2 = log(lead(N2) - N2))


IntercepterChangeOverLay <- IntercepterChangeDF%>%
  ggplot(aes(x=Date))+
  geom_point(aes(y = ChangeN1, color = Site), size = SizeUsed,
             alpha = alphaUsed, shape=N1ShapeUnit)+
  geom_point(aes(y = ChangeN2,color = Site), size = SizeUsed,
             alpha = alphaUsed, shape = N2ShapeUnit)+
  scale_colour_manual(values = HueSpace)+
  theme_light()+
  ylab("gene concentration change (GC/L)")

ggplotly(IntercepterChangeOverLay)%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"))%>%
  config(displaylogo = FALSE)


IntercepterPerChangeOverLay <- IntercepterChangeDF%>%
  ggplot(aes(x = Date))+
  geom_point(aes(y = PerChangeN1, color = Site), size = SizeUsed,
             alpha = alphaUsed, shape = N1ShapeUnit)+
  geom_point(aes(y = PerChangeN2, color = Site), size = SizeUsed,
             alpha= alphaUsed,shape = N2ShapeUnit)+
  theme_light()#+
  #scale_y_log10()

#ggplotly(IntercepterPerChangeOverLay)


MadisonChangeDF <- FullDF%>%
  filter(Site=="Madison",!is.na(N1))%>%
  mutate(ChangeN1 = lead(N1) - N1,
         ChangeN2 = lead(N2) - N2,
         PerChangeN1 = log(lead(N1) - N1),
         PerChangeN2 = log(lead(N2) - N2))


IntercepterChangeOverLay <- MadisonChangeDF%>%
  ggplot(aes(x=Date))+
  geom_point(aes(y = ChangeN1, color = Site), size = SizeUsed,
             alpha = alphaUsed, shape=N1ShapeUnit)+
  geom_point(aes(y = ChangeN2,color = Site), size = SizeUsed,
             alpha = alphaUsed, shape = N2ShapeUnit)+
  scale_colour_manual(values = HueSpace)+
  theme_light()+
  ylab("gene concentration change (GC/L)")

ggplotly(IntercepterChangeOverLay)%>%
```

```r
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"))%>%
  config(displaylogo = FALSE)


IntercepterPerChangeOverLay <- MadisonChangeDF%>%
  ggplot(aes(x = Date))+
  geom_point(aes(y = PerChangeN1, color = Site), size = SizeUsed,
             alpha = alphaUsed, shape = N1ShapeUnit)+
  geom_point(aes(y = PerChangeN2, color = Site), size = SizeUsed,
             alpha= alphaUsed,shape = N2ShapeUnit)+
  theme_light()#+
  #scale_y_log10()

#ggplotly(IntercepterPerChangeOverLay)

#Flag outliers in graphic


LoessFunc <- function(SiteFilter,DF,SpanConstant = .163,Var){
  MainDF <- DF%>%
    filter(Site==SiteFilter)
  MainDF[[paste0("loess",Var)]] <- loessFit(y=(MainDF[[Var]]),
                      x=MainDF$Date, #create loess fit of the data
                      span=SpanConstant, #span of .2 seems to give the best result,
                                          #not rigorously chosen
                      iterations=5)$fitted#2 iterations remove some bad patterns
  #Same as above but for N2
  N2Name <- gsub("1","2",Var)
  MainDF[[paste0("loess", N2Name)]] <- loessFit(y=(MainDF[[N2Name]]),
                      x=MainDF$Date,
                      span=SpanConstant,
                      iterations=5)$fitted#
return(MainDF)
}
#Temp clone to test 7 MA
# LoessFunc <- function(SiteFilter,DF,SpanConstant = .163,Var){
#   N2Name <- gsub("1","2",Var)
#   MainDF <- DF%>%
#     filter(Site==SiteFilter)%>%
#     mutate(!!paste0("loess",Var) := rollapply(!!sym(Var), width = 51,
#                                     FUN = mean, fill = NA, na.rm = TRUE ),
#            !!paste0("loess",N2Name) := rollapply(!!sym(N2Name), width = 51,
#                                     FUN = mean, fill = NA, na.rm = TRUE ))
#   return(MainDF)
# }

SiteLoessDF <- IntercepterDF%>%
  mutate(FlowN1 = FlowRate*N1,
         PopN1 = N1/Pop,
         FlowN2 = FlowRate*N2,
         PopN2 = N1/Pop)
```

```r
BaseColDFN1 <- lapply(c("MMSD-P11","MMSD-P18","MMSD-P2","MMSD-P7","MMSD-P8"),
                      LoessFunc,SiteLoessDF,SpanConstant = .15,
        Var = "N1")%>%
         bind_rows()%>%
       mutate(Norm = "None",
        CovConcNorm = loessN1,
        Messure = "N1")

FlowColDFN1 <- lapply(c("MMSD-P11","MMSD-P18","MMSD-P2","MMSD-P7","MMSD-P8"),
                      LoessFunc,SiteLoessDF,SpanConstant = .15,
        Var = "FlowN1")%>%
              bind_rows()%>%
  mutate(Norm = "Flow",
        CovConcNorm = loessFlowN1,
        Messure = "N1")

PopColDFN1 <- lapply(c("MMSD-P11","MMSD-P18","MMSD-P2","MMSD-P7","MMSD-P8"),
                     LoessFunc,SiteLoessDF,SpanConstant = .15,
        Var = "PopN1")%>%
  bind_rows()%>%
  mutate(Norm = "Pop",
        CovConcNorm = loessPopN1,
        Messure = "N1")

BaseColDFN2 <- BaseColDFN1%>%
  mutate(Norm = "None",
        CovConcNorm = loessN2,
        Messure = "N2")

FlowColDFN2 <- FlowColDFN1%>%
  mutate(Norm = "Flow",
        CovConcNorm = loessFlowN2,
        Messure = "N2")

PopColDFN2 <- PopColDFN1%>%
  mutate(Norm = "Pop",
        CovConcNorm = loessPopN2,
        Messure = "N2")


MergedToBeFacetedDF <- rbind(BaseColDFN1, FlowColDFN1, PopColDFN1,
                             BaseColDFN2, FlowColDFN2, PopColDFN2)%>%
  mutate(Norm = factor(Norm,c("None","Flow","Pop")))%>%
  select(Norm,CovConcNorm,Messure,Date)


BaseGridPlot <- MergedToBeFacetedDF%>%
            filter(!is.na(CovConcNorm))%>%
            ggplot(aes(x=Date))+
            geom_line(aes(y=CovConcNorm, color = Site))+
            scale_colour_manual(values = HueSpace)+
            theme_light() +
            facet_grid(Norm~Messure, scales = "free_y") +
```

```r
                 theme(panel.spacing = unit(2, "lines"))+
  ylab("gene concentration with Normalization (GC/L)")


NonLogPlotlyPlot <- ggplotly( BaseGridPlot
  )%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"),
        displaylogo = FALSE)

LogPlotlyPlot <- ggplotly( BaseGridPlot+
                          scale_y_log10())%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"),
        displaylogo = FALSE)

NonLogPlotlyPlot


LogPlotlyPlot


#BaseGridPlot
#BaseGridPlot+
#scale_y_log10()


Mean <- 11.73
StandardDeviation <- 7.68
Scale = StandardDeviation^2/Mean
Shape = Mean/Scale
SLDWidth <- 21
weights <- dgamma(1:SLDWidth, scale = Scale, shape = Shape)

SiteLoessDF <- FullDF%>%
  filter(Site!="Madison")

A <- SiteLoessDF%>%
  filter(!is.na(SevenDayMACases),
         SevenDayMACases != 0)%>%
  ggplot(aes(x=Date))+
  geom_point(aes(y = FirstConfirmed, color = Site),
             data = IntercepterDF, size = .5, alpha = .5)+
  #geom_line(aes(y = SevenDayMACases, color = Site))+
  theme_light() +
  scale_colour_manual(values = HueSpace)+
  geom_line(aes(y = SLDCases, color = Site))+
  scale_y_log10()
ggplotly(A)%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"))%>%
  config(displaylogo = FALSE)


B <- SiteLoessDF%>%
  filter(!is.na(SevenDayMACases),
```

```r
        SevenDayMACases != 0)%>%
  ggplot(aes(x=Date))+
  geom_point(aes(y = 10000 * FirstConfirmed / Pop, color = Site),
             data = IntercepterDF, size = .5, alpha = .5)+
  geom_line(aes(y = 10000 * SLDCases / Pop, color = Site))+
  #geom_line(aes(y = SevenDayMACases / Pop, color = Site))+
  theme_light() +
  scale_colour_manual(values = HueSpace)+
  scale_y_log10()+
  ylab("Cases per 10,000 people")
ggplotly(B)%>%
  config(modeBarButtonsToRemove = c("zoomIn2d", "zoomOut2d","lasso2d",
                                    "select2d", "autoScale2d"),
         displaylogo = FALSE)
```