

# Final.r

muj\_m

2022-04-06

```
#importing Libraries
library(plyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarise
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v tibble  3.1.5      v purrr   0.3.4
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      vforcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::arrange()    masks plyr::arrange()
## x purrr::compact()   masks plyr::compact()
## x dplyr::count()     masks plyr::count()
## x dplyr::failwith()  masks plyr::failwith()
## x dplyr::filter()    masks stats::filter()
## x dplyr::id()        masks plyr::id()
## x dplyr::lag()       masks stats::lag()
## x dplyr::mutate()    masks plyr::mutate()
## x dplyr::rename()    masks plyr::rename()
## x dplyr::summarise() masks plyr::summarise()
## x dplyr::summarize() masks plyr::summarize()
```

```
library(plotly)
```

```
##  
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':  
##  
##     last_plot
```

```
## The following objects are masked from 'package:plyr':  
##  
##     arrange, mutate, rename, summarise
```

```
## The following object is masked from 'package:stats':  
##  
##     filter
```

```
## The following object is masked from 'package:graphics':  
##  
##     layout
```

```
library(hrbrthemes)
```

```
## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.
```

```
## Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and
```

```
## if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow
```

```
#Loading data  
setwd("C:\\\\Users\\\\muj_m\\\\Desktop\\\\Aly_6010")  
stats <- read.csv("All_Lakes_GLIP.csv")  
stat <- stats  
str(stat)
```

```

## 'data.frame': 425011 obs. of 14 variables:
## $ LAKE : chr "Lake Superior" "Lake Superior" "Lake Superior" "Lake Superior" ...
## $ FACILITY_NAME : chr "TERRACE BAY WATER TREATMENT PLANT" ...
## $ STATION : int 20110013 20110013 20110013 20110013 20110013 20110013 20110013 20110013 20110013 20110013 ...
## $ DATE_YYYYMMDD : chr "10/7/1985" "10/7/1985" "10/7/1985" "10/7/1985" ...
## $ YEAR : int 1985 1985 1985 1985 1985 1985 1985 1985 1985 ...
## $ MONTH : int 10 10 10 10 10 10 10 10 10 ...
## $ WEEK : int 41 41 41 41 41 41 41 41 41 ...
## $ ANALYTIC_METHOD : chr "" "" "" ...
## $ TEST_CODE : chr "CLIDUR" "COND25" "NNHTFR" "NNO2FR" ...
## $ PARAMETER : chr "Chloride" "Conductivity" "Nitrogen; ammonia+ammonium" "Nitrogen; nitrite" ...
## $ VALUE : num 2.6 111 0.02 0.001 0.27 0.003 0.003 1.1 0.1 0.28 ...
## $ UNIT : chr "mg/L" "uS/cm" "mg/L" "mg/L" ...
## $ VALUE_QUALIFIER : chr "" "" ...
## $ VALQUAL_DESCRIPTION: chr "" "" ...

```

```

#get all the variables
colnames(stats)

```

```

## [1] "LAKE"           "FACILITY_NAME"   "STATION"
## [4] "DATE_YYYYMMDD" "YEAR"          "MONTH"
## [7] "WEEK"           "ANALYTIC_METHOD" "TEST_CODE"
## [10] "PARAMETER"      "VALUE"          "UNIT"
## [13] "VALUE_QUALIFIER" "VALQUAL_DESCRIPTION"

```

```

#renaming column
colnames(stats)[4]="DATE"
colnames(stats)[2]="FACILITY"
colnames(stats)[13]="QUALIFIER"
colnames(stats)[14]="DESCRIPTION"
colnames(stats)

```

```

## [1] "LAKE"           "FACILITY"       "STATION"        "DATE"
## [5] "YEAR"           "MONTH"          "WEEK"          "ANALYTIC_METHOD"
## [9] "TEST_CODE"       "PARAMETER"      "VALUE"         "UNIT"
## [13] "QUALIFIER"      "DESCRIPTION"

```

```

#Converting date from factor to DATE format
stats$DATE <- as.Date(stats$DATE, "%m/%d/%Y")
str(stats)

```

```
## 'data.frame': 425011 obs. of 14 variables:
## $ LAKE : chr "Lake Superior" "Lake Superior" "Lake Superior" "Lake Superior"
...
## $ FACILITY : chr "TERRACE BAY WATER TREATMENT PLANT" ...
## $ STATION : int 20110013 20110013 20110013 20110013 20110013 20110013 20110013 20110013 20110013 ...
## $ DATE : Date, format: "1985-10-07" "1985-10-07" ...
## $ YEAR : int 1985 1985 1985 1985 1985 1985 1985 1985 1985 ...
## $ MONTH : int 10 10 10 10 10 10 10 10 10 ...
## $ WEEK : int 41 41 41 41 41 41 41 41 41 ...
## $ ANALYTIC_METHOD: chr "" "" "" ...
## $ TEST_CODE : chr "CLIDUR" "COND25" "NNHTFR" "NNO2FR" ...
## $ PARAMETER : chr "Chloride" "Conductivity" "Nitrogen; ammonia+ammonium" "Nitrogen; nitrite" ...
## $ VALUE : num 2.6 111 0.02 0.001 0.27 0.003 0.003 1.1 0.1 0.28 ...
## $ UNIT : chr "mg/L" "uS/cm" "mg/L" "mg/L" ...
## $ QUALIFIER : chr "" "" ...
## $ DESCRIPTION : chr "" "" ...
```

```
#Unique values in the data
unique(stats$LAKE)
```

```
## [1] "Lake Superior" "Lake Huron" "Lake Erie" "Lake Ontario"
```

```
unique(stats$FACILITY)
```

```
## [1] "TERRACE BAY WATER TREATMENT PLANT"
## [2] "BARE POINT WATER TREATMENT PLANT"
## [3] "GODERICH WATER TREATMENT PLANT"
## [4] "GRAND BEND - LAKE HURON PRIMARY WATER SUPPLY SYSTEM"
## [5] "LAMBTON AREA WATER SUPPLY SYSTEM"
## [6] "AMHERSTBURG WATER TREATMENT PLANT"
## [7] "UNION WATER SUPPLY SYSTEM"
## [8] "BLENHEIM WATER TREATMENT PLANT"
## [9] "ELGIN AREA PRIMARY WATER SUPPLY SYSTEM"
## [10] "DUNNVILLE DRINKING WATER SYSTEM"
## [11] "ROSEHILL WATER TREATMENT PLANT"
## [12] "SOUTH PEEL - LAKEVIEW WATER TREATMENT PLANT"
## [13] "KINGSTON - KING STREET WATER TREATMENT PLANT"
## [14] "TORONTO - R. L. CLARK WATER TREATMENT PLANT"
## [15] "GRIMSBY WATER TREATMENT PLANT"
## [16] "COBOURG WATER TREATMENT PLANT"
## [17] "BELLEVILLE - GERRY O'CONNOR WATER TREATMENT PLANT"
## [18] "BROCKVILLE WATER TREATMENT PLANT"
```

```
unique(stats$YEAR)
```

```
## [1] 1985 1986 1987 1995 1990 1991 1992 1993 1994 1996 1997 1979 1980 1981 1982
## [16] 1983 1984 1988 1989 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
## [31] 2016 2017 2009 2010 2011 2012 2013 2014 2015 2018 2019 1976 1977 1978
```

```
unique(stats$TEST_CODE)
```

```
## [1] "CLIDUR" "COND25" "NNHTFR" "NNO2FR" "NNO3FR" "PPO4FR" "PPUT"    "SIO3UR"  
## [9] "NNTKUR" "NNOTFR" "CHLRAT" "CHLRBT" "NNHTUR" "NNO2UR" "NNOTUR" "PH"  
## [17] "DIC"     "DOC"      "NTOT"     "TURB"
```

```
unique(stats$PARAMETER)
```

```
## [1] "Chloride"                  "Conductivity"  
## [3] "Nitrogen; ammonia+ammonium" "Nitrogen; nitrite"  
## [5] "Nitrogen; nitrate"          "Phosphorus; phosphate"  
## [7] "Phosphorus; total"          "Silicon; reactive silicate"  
## [9] "Nitrogen; total Kjeldahl"   "Nitrogen; nitrate+nitrite"  
## [11] "Chlorophyll - a; total"    "Chlorophyll - b; total"  
## [13] "pH"                        "Carbon; dissolved inorganic"  
## [15] "Carbon; dissolved organic" "Nitrogen; total"  
## [17] "Turbidity"
```

```
unique(stats$UNIT)
```

```
## [1] "mg/L"   "uS/cm"  "ug/L"   "none"   "NTU"    ""
```

```
#substituting empty rows with "empty" string with gsub  
stats$DESCRIPTION <- gsub("^$|^ $", "Empty", stats$DESCRIPTION)
```

```
#dropping column  
stats$ANALYTIC_METHOD<-NULL
```

```
#Descriptive Statistics  
class(stats)
```

```
## [1] "data.frame"
```

```
nrow(stats)
```

```
## [1] 425011
```

```
ncol(stats)
```

```
## [1] 13
```

```
head(stats)
```

```

##          LAKE             FACILITY   STATION      DATE YEAR
## 1 Lake Superior TERRACE BAY WATER TREATMENT PLANT 20110013 1985-10-07 1985
## 2 Lake Superior TERRACE BAY WATER TREATMENT PLANT 20110013 1985-10-07 1985
## 3 Lake Superior TERRACE BAY WATER TREATMENT PLANT 20110013 1985-10-07 1985
## 4 Lake Superior TERRACE BAY WATER TREATMENT PLANT 20110013 1985-10-07 1985
## 5 Lake Superior TERRACE BAY WATER TREATMENT PLANT 20110013 1985-10-07 1985
## 6 Lake Superior TERRACE BAY WATER TREATMENT PLANT 20110013 1985-10-07 1985
##    MONTH WEEK TEST_CODE           PARAMETER     VALUE UNIT QUALIFIER
## 1    10   41    CLIDUR        Chloride    2.600 mg/L
## 2    10   41    COND25       Conductivity 111.000 uS/cm
## 3    10   41    NNHTFR Nitrogen; ammonia+ammonium 0.020 mg/L
## 4    10   41    NNO2FR      Nitrogen; nitrite  0.001 mg/L
## 5    10   41    NNO3FR      Nitrogen; nitrate 0.270 mg/L
## 6    10   41    PPO4FR      Phosphorus; phosphate 0.003 mg/L
##    DESCRIPTION
## 1    Empty
## 2    Empty
## 3    Empty
## 4    Empty
## 5    Empty
## 6    Empty

```

```
tail(stats)
```

```

##          LAKE             FACILITY   STATION      DATE YEAR
## 425006 Lake Ontario BROCKVILLE WATER TREATMENT PLANT 20180011 2019-12-31 2019
## 425007 Lake Ontario BROCKVILLE WATER TREATMENT PLANT 20180011 2019-12-31 2019
## 425008 Lake Ontario BROCKVILLE WATER TREATMENT PLANT 20180011 2019-12-31 2019
## 425009 Lake Ontario BROCKVILLE WATER TREATMENT PLANT 20180011 2019-12-31 2019
## 425010 Lake Ontario BROCKVILLE WATER TREATMENT PLANT 20180011 2019-12-31 2019
## 425011 Lake Ontario BROCKVILLE WATER TREATMENT PLANT 20180011 2019-12-31 2019
##    MONTH WEEK TEST_CODE           PARAMETER     VALUE UNIT QUALIFIER
## 425006    12   53    NNOTUR Nitrogen; nitrate+nitrite 0.320 mg/L
## 425007    12   53    NNO2UR Nitrogen; nitrite  0.003 mg/L
## 425008    12   53    PPO4FR Phosphorus; phosphate 0.003 mg/L <MDL
## 425009    12   53    DIC     Carbon; dissolved inorganic 23.100 mg/L
## 425010    12   53    DOC     Carbon; dissolved organic 1.800 mg/L
## 425011    12   53    SIO3UR Silicon; reactive silicate 0.380 mg/L
##    DESCRIPTION
## 425006          Empty
## 425007          Empty
## 425008 LESS THAN METHOD DETECTION LIMIT
## 425009          Empty
## 425010          Empty
## 425011          Empty

```

```
str(stats)
```

```

## 'data.frame': 425011 obs. of 13 variables:
## $ LAKE      : chr "Lake Superior" "Lake Superior" "Lake Superior" "Lake Superior" ...
## $ FACILITY   : chr "TERRACE BAY WATER TREATMENT PLANT" "TERRACE BAY WATER TREATMENT PLAN T" "TERRACE BAY WATER TREATMENT PLANT" "TERRACE BAY WATER TREATMENT PLANT" ...
## $ STATION    : int 20110013 20110013 20110013 20110013 20110013 20110013 20110013 20110013 20110013 ...
## $ DATE       : Date, format: "1985-10-07" "1985-10-07" ...
## $ YEAR       : int 1985 1985 1985 1985 1985 1985 1985 1985 1985 ...
## $ MONTH      : int 10 10 10 10 10 10 10 10 10 ...
## $ WEEK       : int 41 41 41 41 41 41 41 41 41 ...
## $ TEST_CODE   : chr "CLIDUR" "COND25" "NNHTFR" "NNO2FR" ...
## $ PARAMETER   : chr "Chloride" "Conductivity" "Nitrogen; ammonia+ammonium" "Nitrogen; nitrite" ...
## $ VALUE       : num 2.6 111 0.02 0.001 0.27 0.003 0.003 1.1 0.1 0.28 ...
## $ UNIT        : chr "mg/L" "uS/cm" "mg/L" "mg/L" ...
## $ QUALIFIER   : chr "" "" "" ...
## $ DESCRIPTION: chr "Empty" "Empty" "Empty" "Empty" ...

```

```
summary(stats)
```

	LAKE	FACILITY	STATION	DATE
##	Length:425011	Length:425011	Min. :20110013	Min. :1976-01-26
##	Class :character	Class :character	1st Qu.:20134003	1st Qu.:1988-01-25
##	Mode :character	Mode :character	Median :20150008	Median :1998-02-23
##			Mean :20151135	Mean :1998-08-11
##			3rd Qu.:20170015	3rd Qu.:2009-04-20
##			Max. :20180011	Max. :2019-12-31
##				
	YEAR	MONTH	WEEK	TEST_CODE
##	Min. :1976	Min. : 1.000	Min. : 1.00	Length:425011
##	1st Qu.:1988	1st Qu.: 4.000	1st Qu.:14.00	Class :character
##	Median :1998	Median : 7.000	Median :27.00	Mode :character
##	Mean :1998	Mean : 6.537	Mean :27.28	
##	3rd Qu.:2009	3rd Qu.:10.000	3rd Qu.:40.00	
##	Max. :2019	Max. :12.000	Max. :53.00	
##				
	PARAMETER	VALUE	UNIT	QUALIFIER
##	Length:425011	Min. : 0.00	Length:425011	Length:425011
##	Class :character	1st Qu.: 0.01	Class :character	Class :character
##	Mode :character	Median : 0.28	Mode :character	Mode :character
##		Mean : 13.73		
##		3rd Qu.: 1.60		
##		Max. :2060.00		
##		NA's :153		
##	DESCRIPTION			
##	Length:425011			
##	Class :character			
##	Mode :character			
##				
##				
##				
##				

```
#total rows and columns  
dim(stats)
```

```
## [1] 425011      13
```

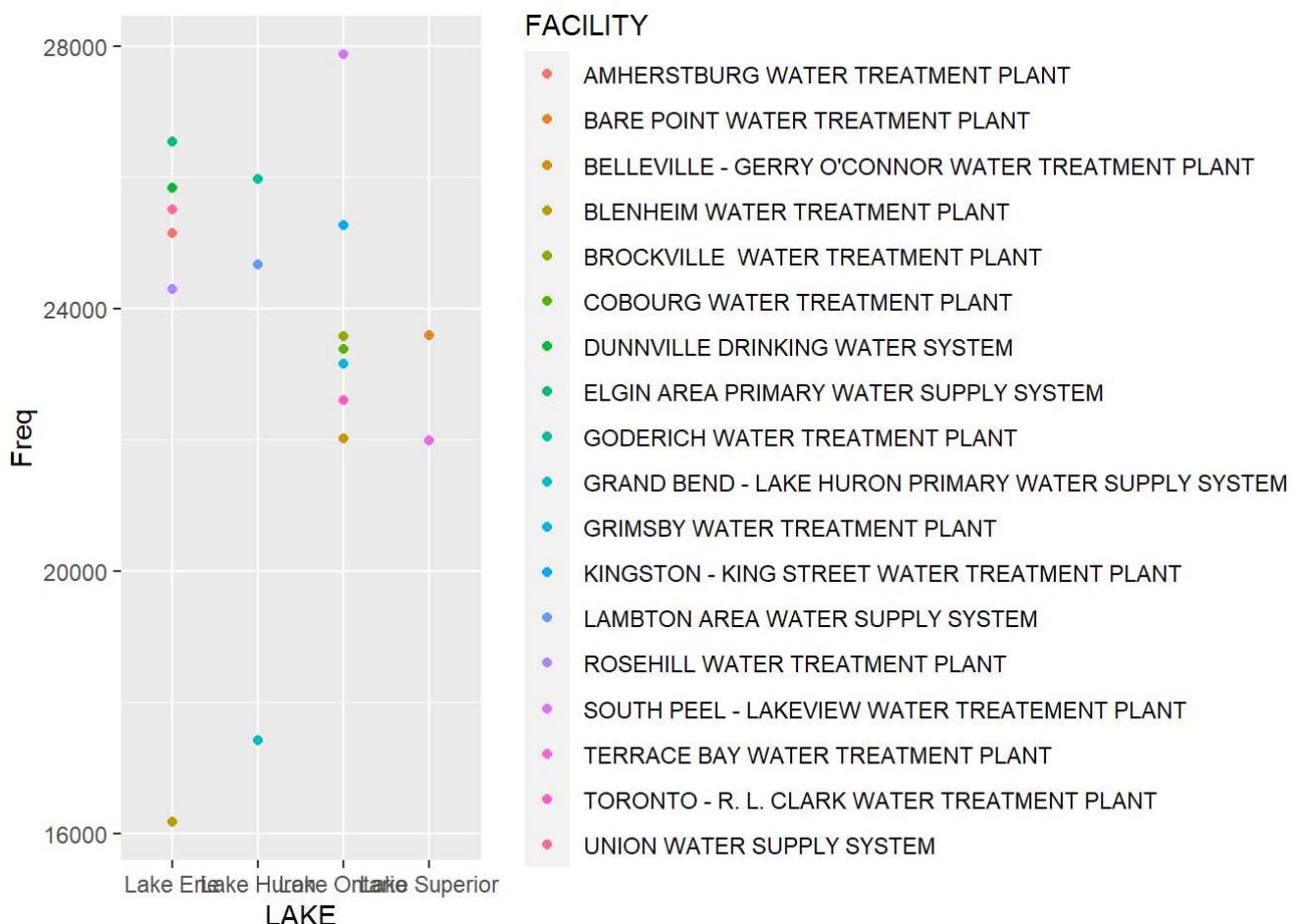
```
#total empty rows  
sum(is.na(stats))
```

```
## [1] 153
```

```
#Lakes and facility freq  
tt <- rename(count(stats, LAKE, FACILITY), Freq = n)  
view(tt)
```

```
#Cross tabulation with Lake and facility  
crossy <- xtabs(~ LAKE+FACILITY,data=stats)  
View(crossy)
```

```
ggplot(data = tt)+geom_point(mapping = aes(x=LAKE, y=Freq,color = FACILITY))
```



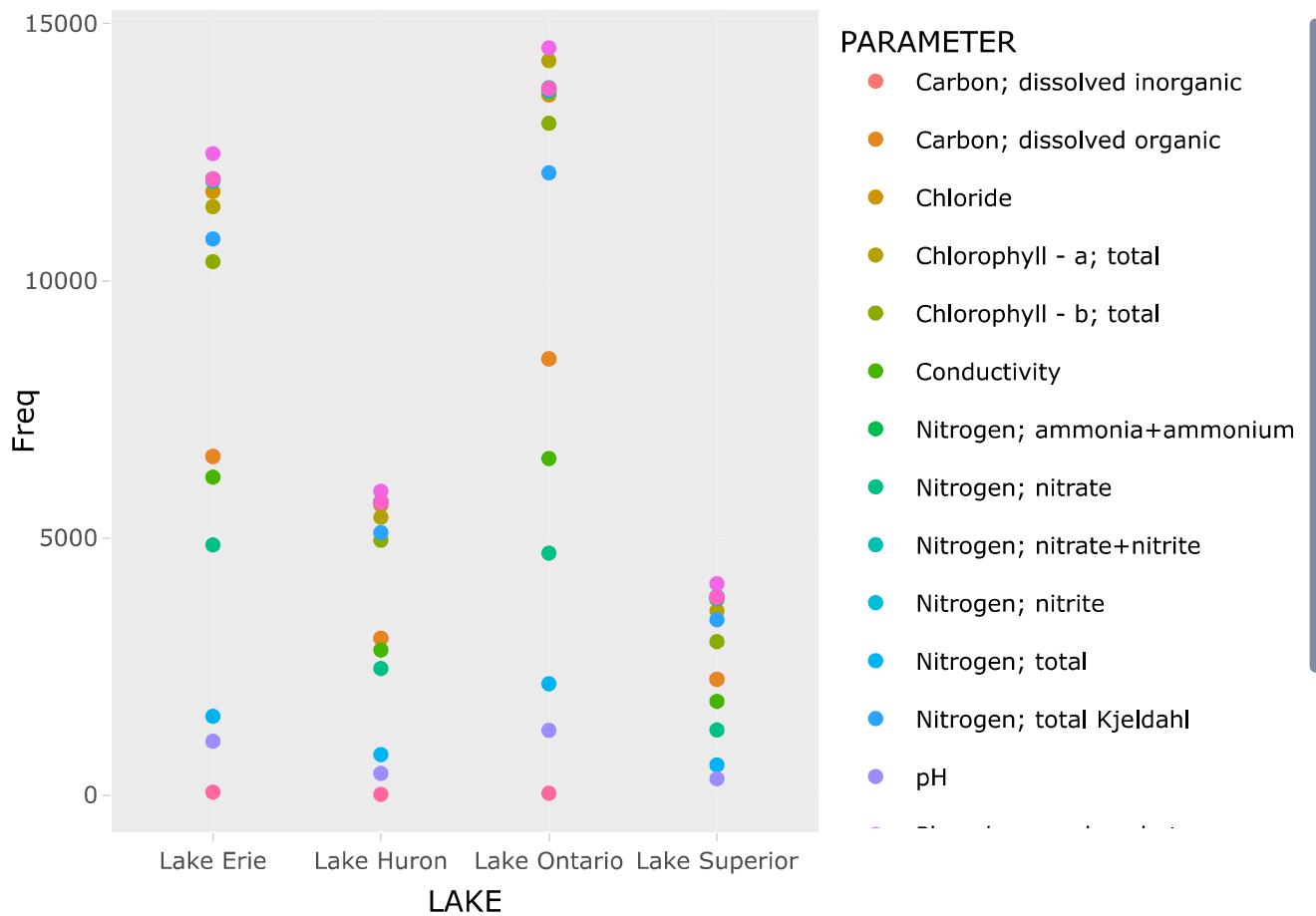
```

#Lakes and parameter freq
lp <- rename(count(stats, LAKE, PARAMETER), Freq = n)
View(lp)

#chemicals in Lake
content<-xtabs(~LAKE+PARAMETER,data=stats)
View(content)

#ggplotly for Lakes and parameter with freq
pip <- ggplot(data = lp)+geom_point(mapping = aes(x=LAKE, y=Freq,color = PARAMETER))
ggplotly(pip)

```



```

#ggPlot with Lakes year content and freq
yp <- rename(count(stats, LAKE, PARAMETER, YEAR), Freq = n)
View(yp)

lyc <- ggplot(data = yp)+geom_point(mapping = aes(x=LAKE, y=Freq, text=YEAR,color = PARAMETER))

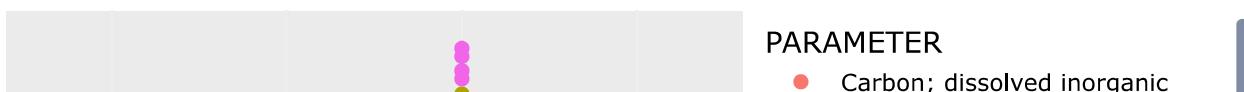
```

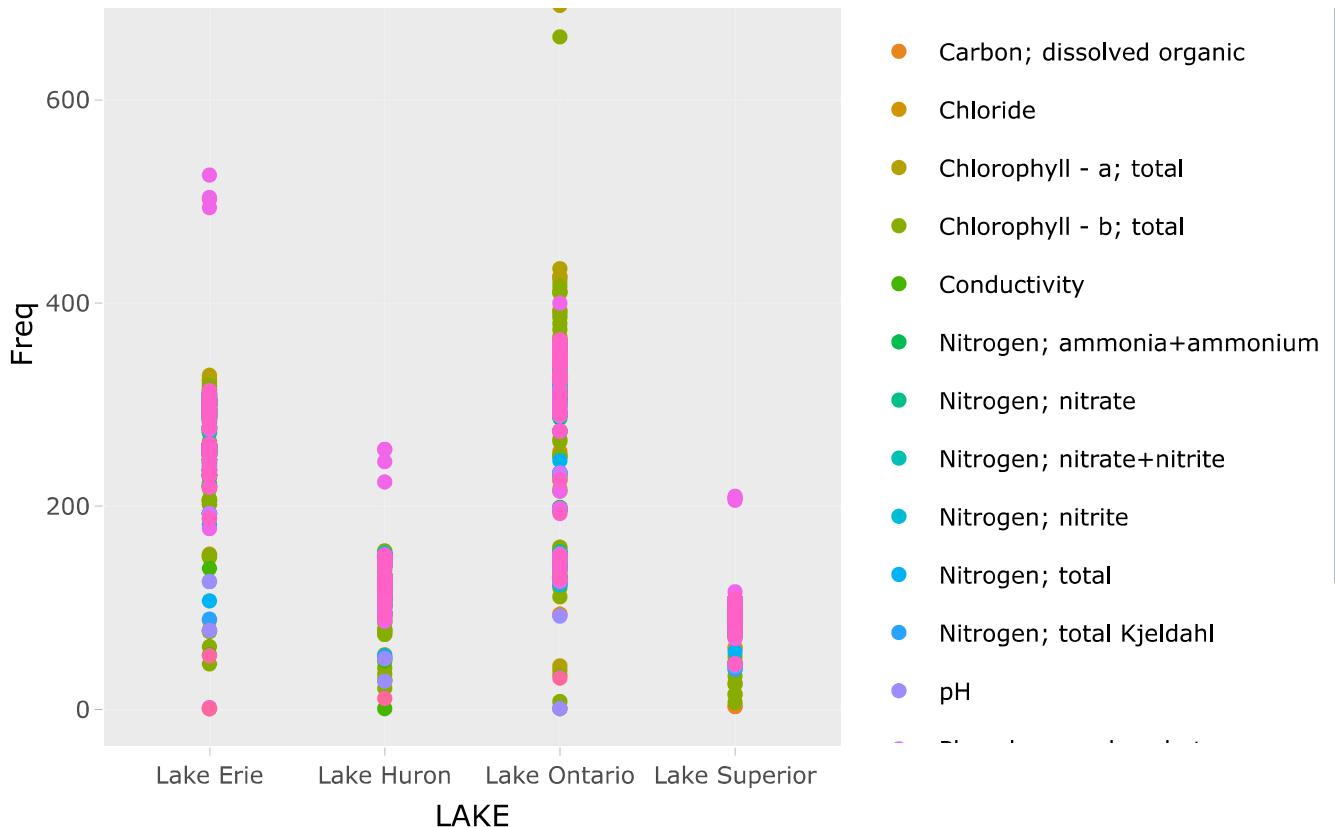
## Warning: Ignoring unknown aesthetics: text

```

ggplotly(lyc, tooltip = "text")

```





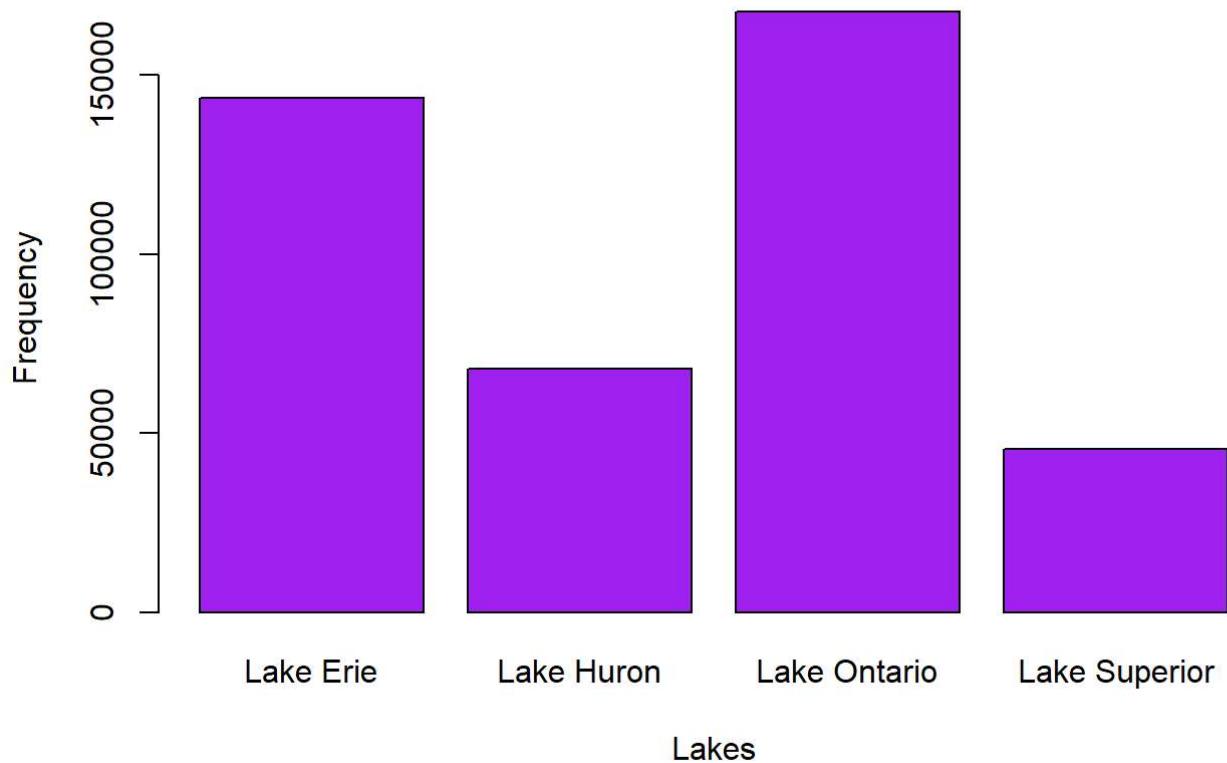
```
#no of records for each Lake
lakecount <- table(stats$LAKE)
lakecount
```

```
##
##      Lake Erie     Lake Huron   Lake Ontario  Lake Superior
##        143517       68049       167869       45576
```

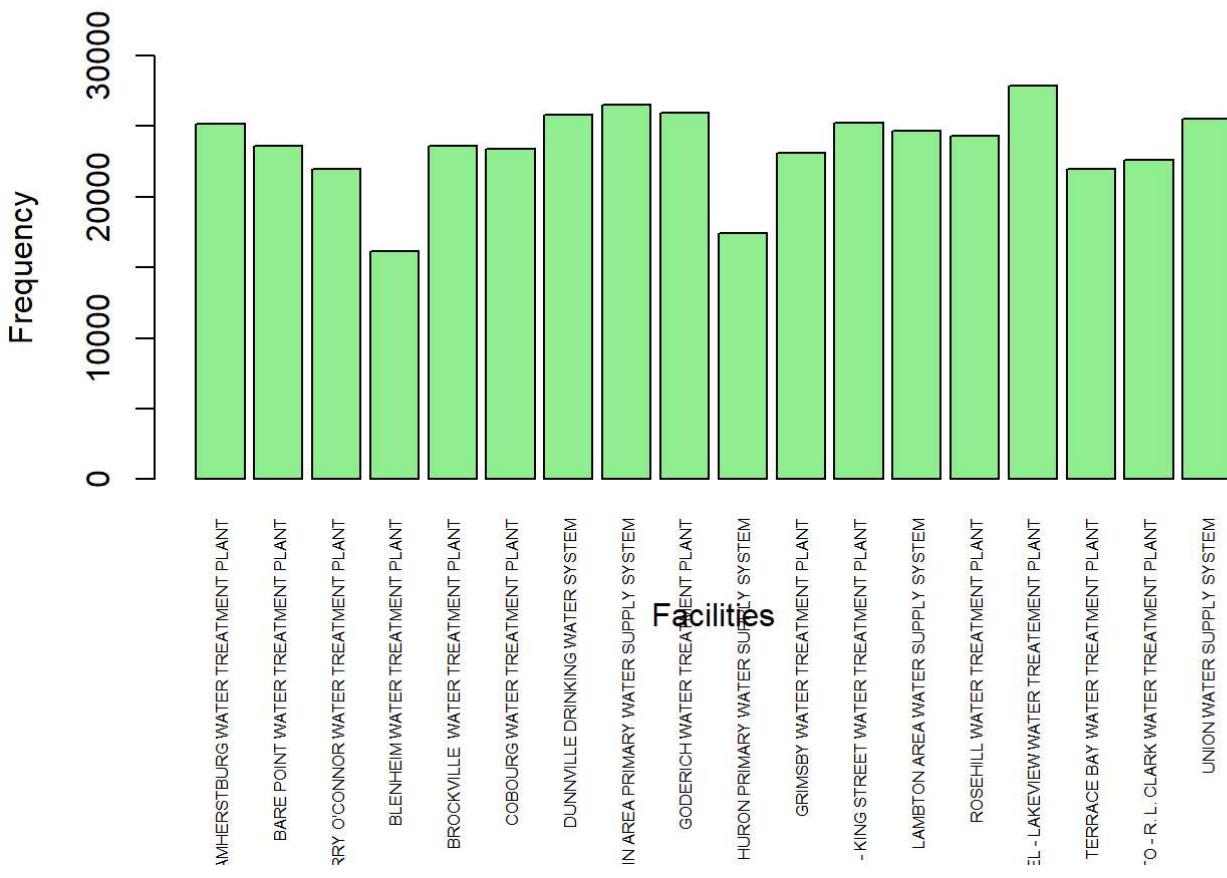
```
#no of record for each facility
fcc <- table(stats$FACILITY)
View(fcc)

#no of records for parameters
paracount <- table(stats$PARAMETER)
View(paracount)

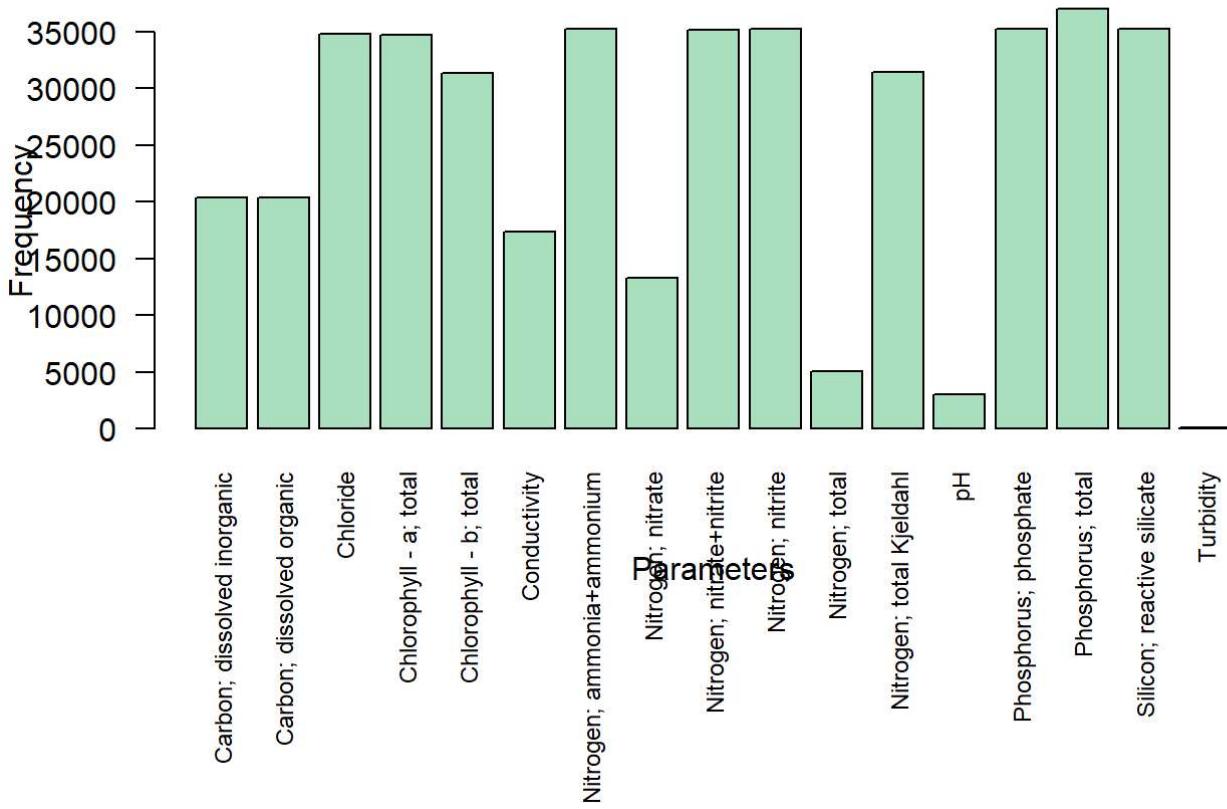
#plot
barplot(lakecount,col="purple", ylim=c(0, 170000), xlab = "Lakes", ylab="Frequency")
```



```
#facility freq plot
par(mar=c(10,4,4,2))
barplot(fcc,col="lightgreen", ylim=c(0, 30000), xlab = "Facilities", ylab="Frequency", las=3, cex.names=.5)
```



```
#parameter pPlot
barplot(paracount,las=2, xlab = "Parameters", ylab="Frequency", cex.names=.75, col="#a9dfbf")
```



```

#cross tab with Lake facility and their parameters
pol <- xtabs(~ LAKE+FACILITY+PARAMETER,data=stats)
View(pol)

#Getting rows based on chloride
cl <- filter(stats, PARAMETER=="Chloride")
cl <- cl[order(cl$VALUE,decreasing = TRUE),]

summary(cl)

```

```

##      LAKE          FACILITY        STATION        DATE
##  Length:34782    Length:34782    Min.   :20110013  Min.   :1976-01-26
##  Class :character  Class :character  1st Qu.:20134003  1st Qu.:1987-09-22
##  Mode  :character  Mode  :character  Median :20150008  Median :1997-11-03
##                               Mean   :20150953  Mean   :1998-04-09
##                               3rd Qu.:20170015  3rd Qu.:2008-10-14
##                               Max.   :20180011  Max.   :2019-12-31
##
##      YEAR          MONTH        WEEK        TEST_CODE
##  Min.   :1976   Min.   : 1.00   Min.   : 1.00  Length:34782
##  1st Qu.:1987   1st Qu.: 4.00   1st Qu.:14.00  Class :character
##  Median :1997   Median : 7.00   Median :27.00  Mode  :character
##  Mean   :1998   Mean   : 6.53   Mean   :27.25
##  3rd Qu.:2008   3rd Qu.: 9.00   3rd Qu.:40.00
##  Max.   :2019   Max.   :12.00   Max.   :53.00
##
##      PARAMETER        VALUE        UNIT        QUALIFIER
##  Length:34782    Min.   : 0.000  Length:34782  Length:34782
##  Class :character  1st Qu.: 8.885  Class :character  Class :character
##  Mode  :character  Median : 16.100  Mode  :character  Mode  :character
##                               Mean   : 15.955
##                               3rd Qu.: 22.500
##                               Max.   :607.000
##                               NA's   :3
##
##      DESCRIPTION
##  Length:34782
##  Class :character
##  Mode  :character
##
##      
```

```
#overall high chloride range
cl %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  geom_line(color="#69b3a2") +
  ylim(0,700) +
  annotate(geom="text", x=as.Date("1999-01-17"), y=650,
           label=" highest Chloride level") +
  annotate(geom="point", x=as.Date("1999-08-17"), y=607, size=10, shape=21, fill="transparent") +
  geom_hline(yintercept=200, color="orange", size=.5) +
  theme_ipsum()
```

```
## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font family not
## found in Windows font database
```

```
## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font family not
## found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

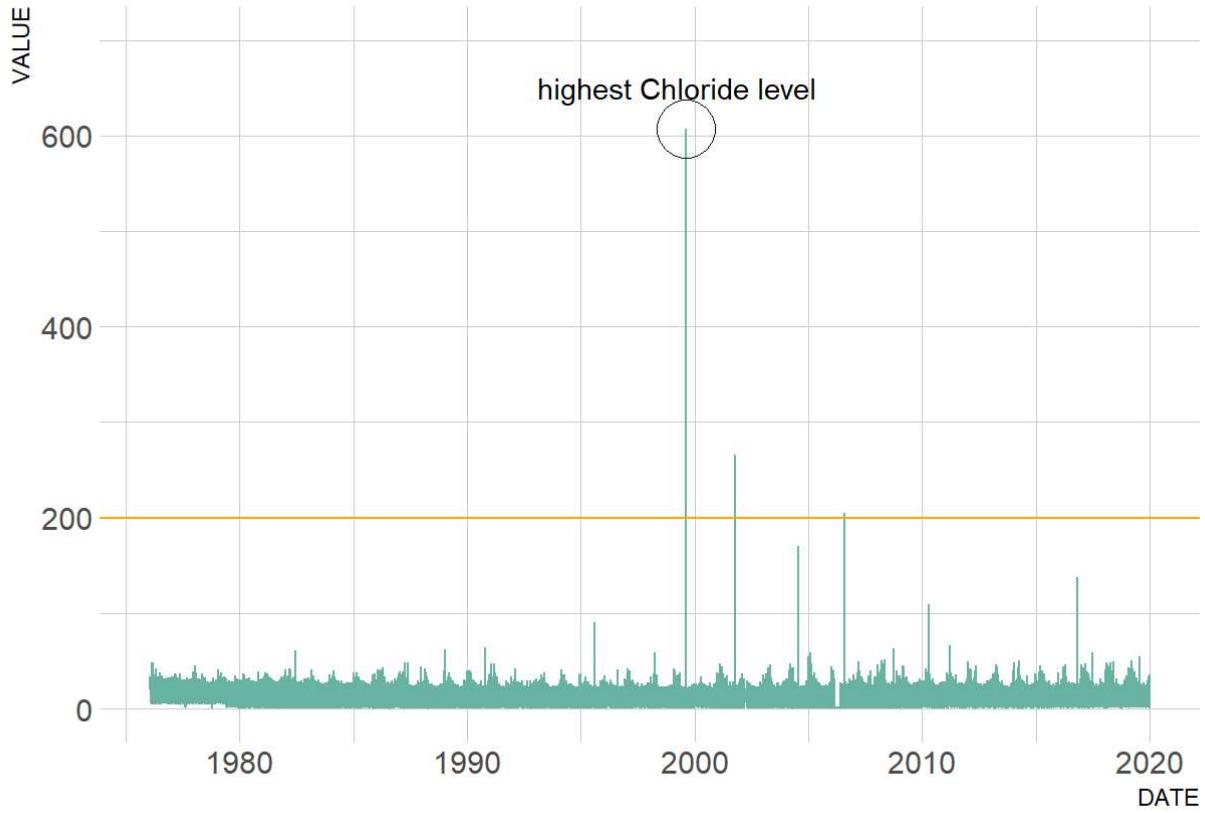
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

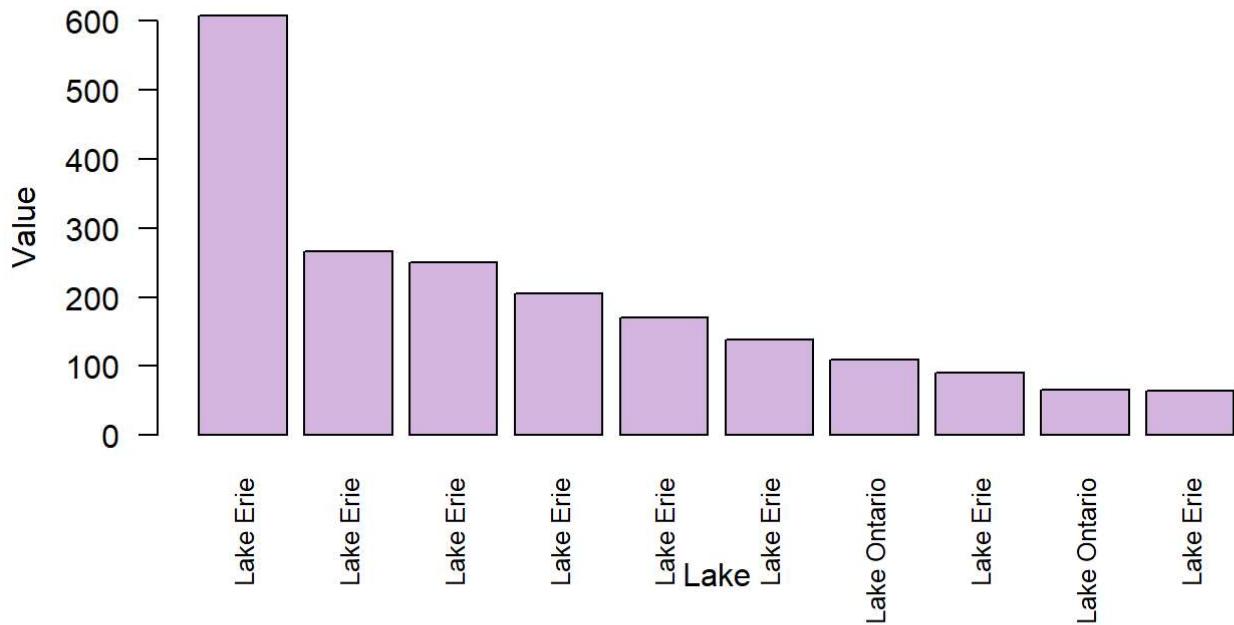
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database
```

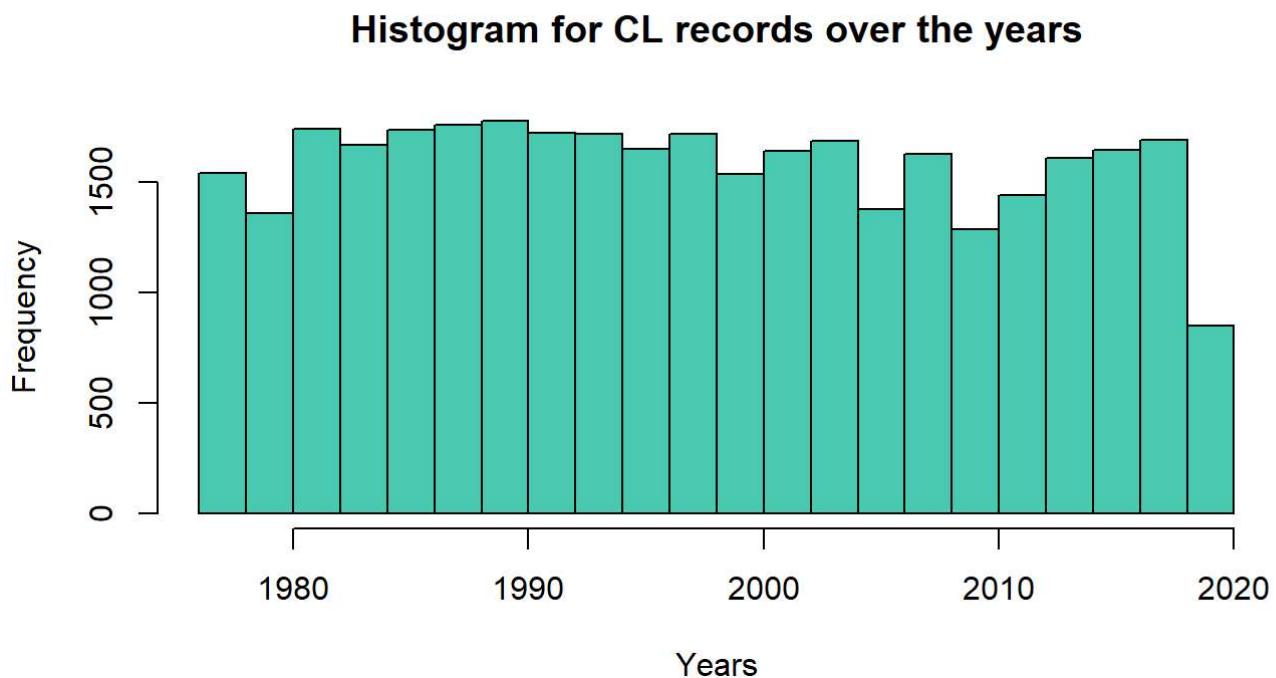


```
#top 10 records with most chloride  
barplot(head(cl$VALUE,n=10),names.arg = head(cl$LAKE,n=10),las=2,cex.names = 0.8, col="#d2b4d  
e", xlab = "Lake",ylab = "Value",main ="Records with most chloride content")
```

## Records with most chloride content



```
hist(cl$YEAR, col = "#48c9b0", main = "Histogram for CL records over the years", xlab="Years")
```



```

max(cl$VALUE)

## [1] NA

min(cl$VALUE)

## [1] NA

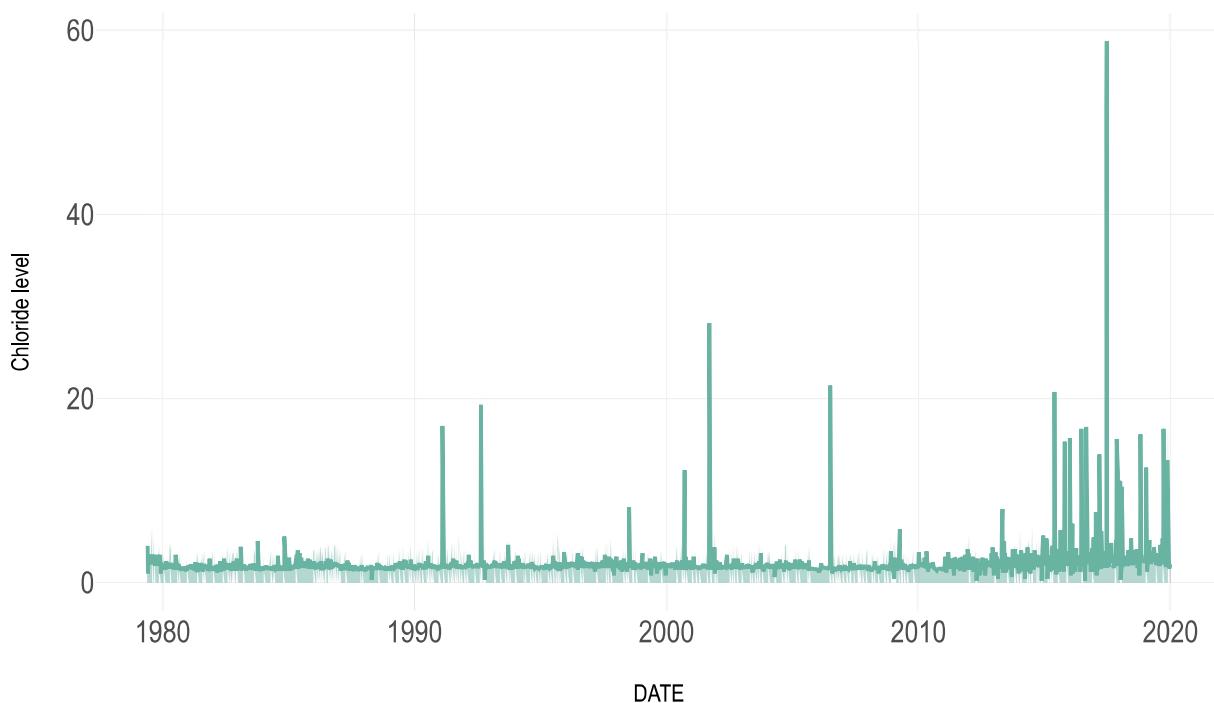
#chloride for each Lake
ls = cl[cl$LAKE=="Lake Superior",]
le = cl[cl$LAKE=="Lake Erie",]
lh = cl[cl$LAKE=="Lake Huron",]
lo = cl[cl$LAKE=="Lake Ontario",]

#Chloride Level in Lake superior
th<- ls %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggtitle("Chloride level in Lake Superior")+
  geom_area(fill="#69b3a2", alpha=0.5) +
  geom_line(color="#69b3a2") +
  ylab("Chloride level") +
  theme_ipsum()
ggplotly(th)

```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```

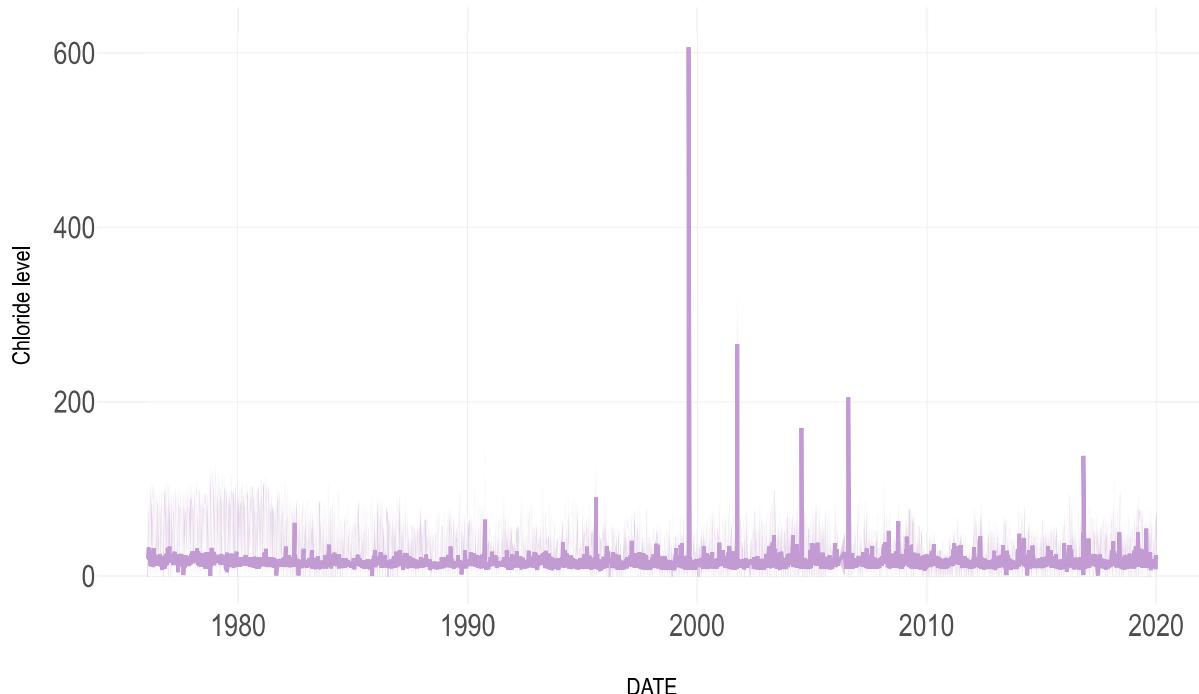
## Chloride level in Lake Superior



```
#chloride Level in Lake erie
te<- le %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggtitle("Chloride level in Lake Erie")+
  geom_area(fill="#c39bd3", alpha=0.5) +
  geom_line(color="#c39bd3") +
  ylab("Chloride level") +
  theme_ipsum()
ggplotly(te)
```

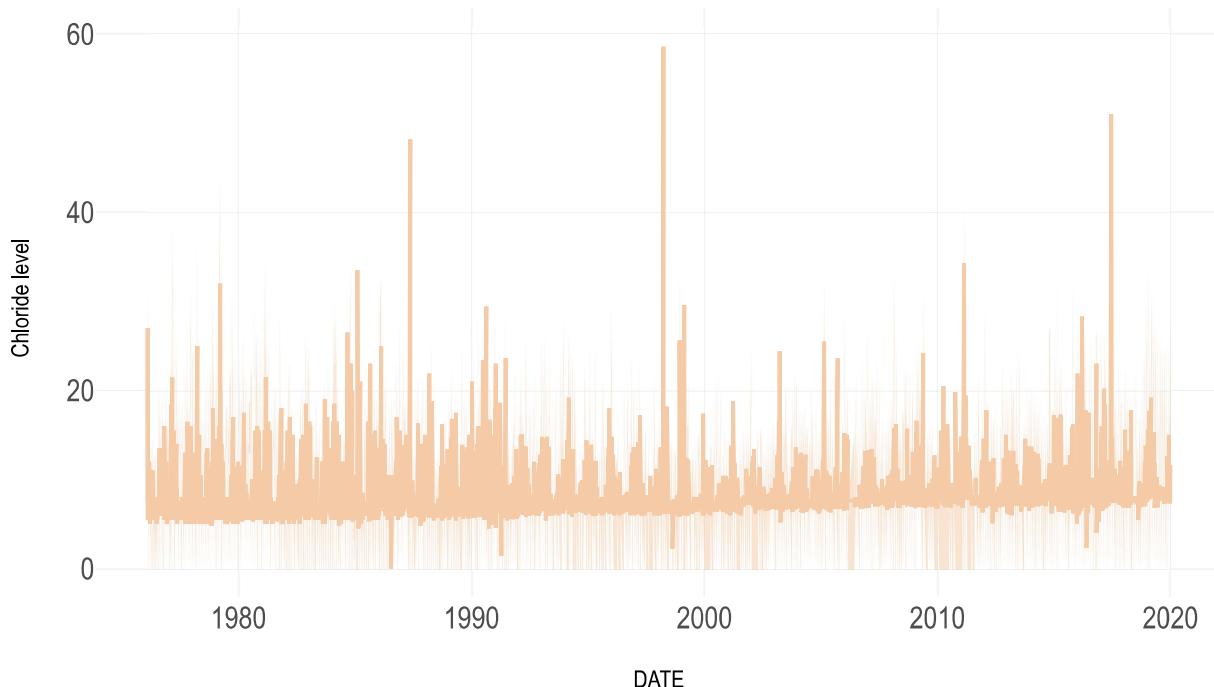
## Warning: Removed 2 rows containing missing values (position\_stack).

## Chloride level in Lake Erie



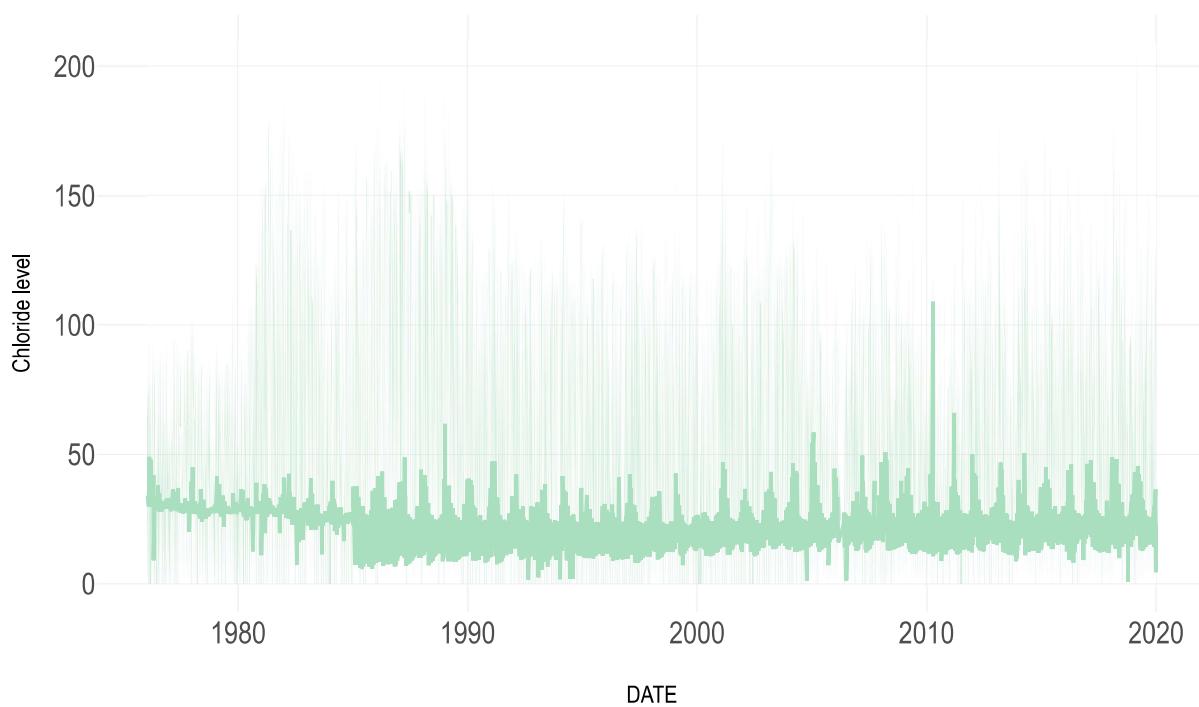
```
#Lake huron chloride
th<- lh %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggtitle("Chloride level in Lake Huron")+
  geom_area(fill="#f5cba7", alpha=0.5) +
  geom_line(color="#f5cba7") +
  ylab("Chloride level") +
  theme_ipsum()
ggplotly(th)
```

## Chloride level in Lake Huron



```
#Lake ontario chloride
to<- lo %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggtitle("Chloride level in Lake Ontario")+
  geom_area(fill="#a9dfbf", alpha=0.5) +
  geom_line(color="#a9dfbf") +
  ylab("Chloride level") +
  theme_ipsum()
ggplotly(to)
```

## Chloride level in Lake Ontario



```
#nitrogen for each lake

#nitrogen nitrate
nitra <- filter(stats, PARAMETER=="Nitrogen; nitrate")

nit <- nitra[order(nitra$VALUE,decreasing = TRUE),]

nit %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  geom_line(color="#f5b7b1") +
  ylim(0,6) +
  annotate(geom="text", x=as.Date("1981-02-23"), y=6,
           label=" highest nitrate level") +
  annotate(geom="point", x=as.Date("1981-02-23"), y=5.50, size=10, shape=21, fill="transparent") +
  geom_hline(yintercept=2, color="orange", size=.5) +
  theme_ipsum()
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

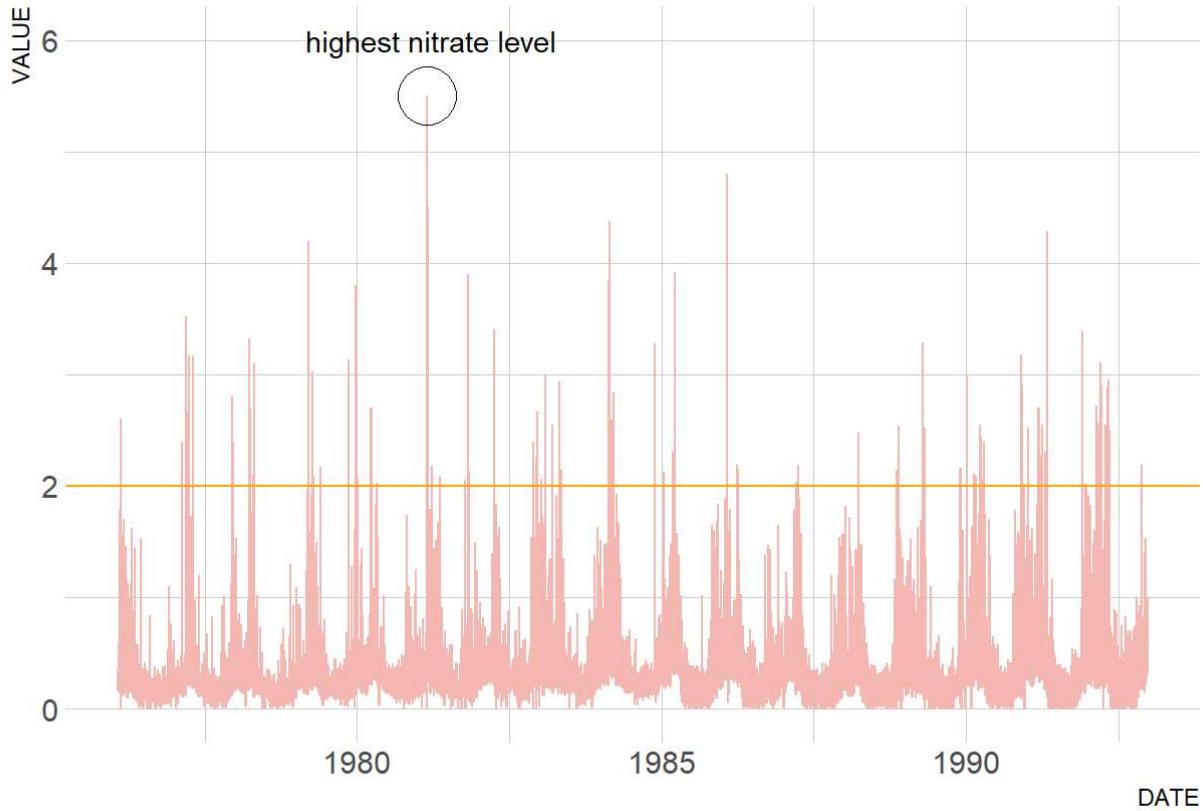
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

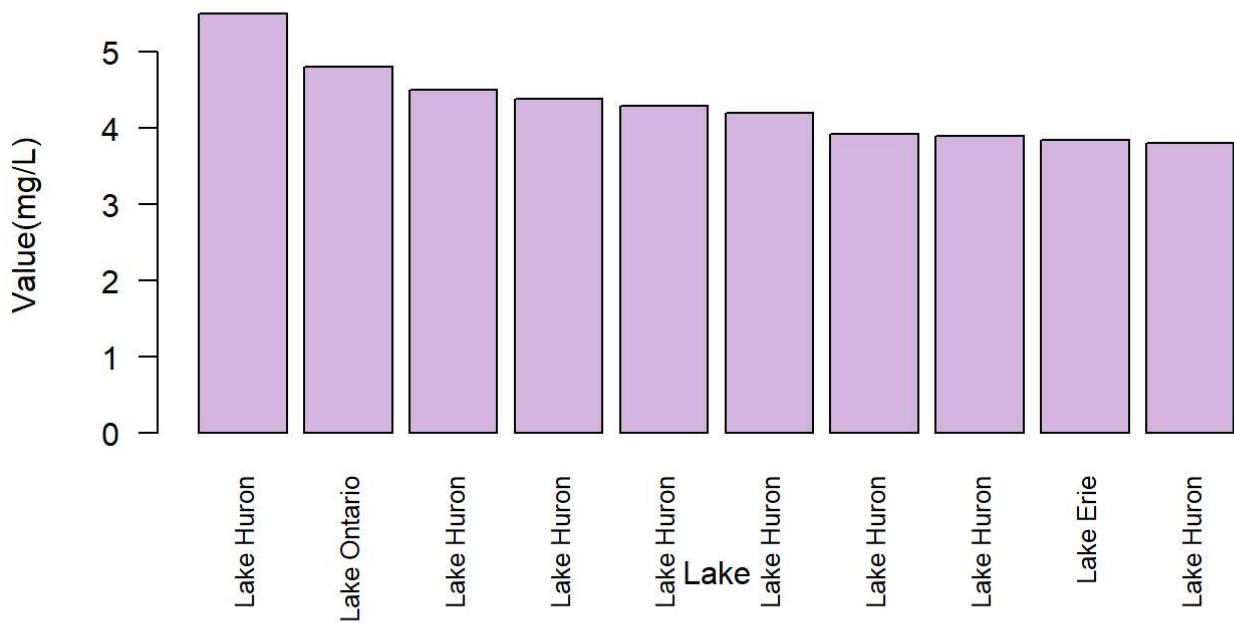
```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database
```

```
## Warning in grid.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database
```



```
#top 10 records with most nitrate
barplot(head(hit$VALUE,n=10),names.arg = head(hit$LAKE,n=10),las=2,cex.names = 0.8, col="#d2b4de", xlab = "Lake",ylab = "Value(mg/L)",main ="Records with most Nitrate content")
```

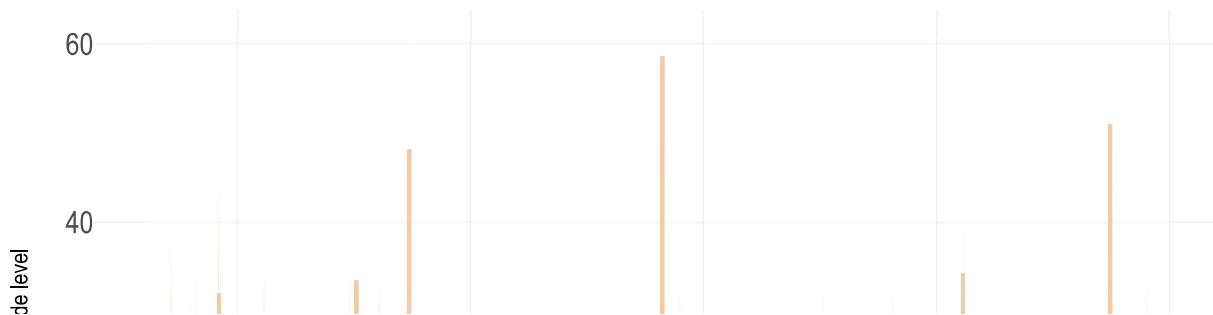
## Records with most Nitrate content

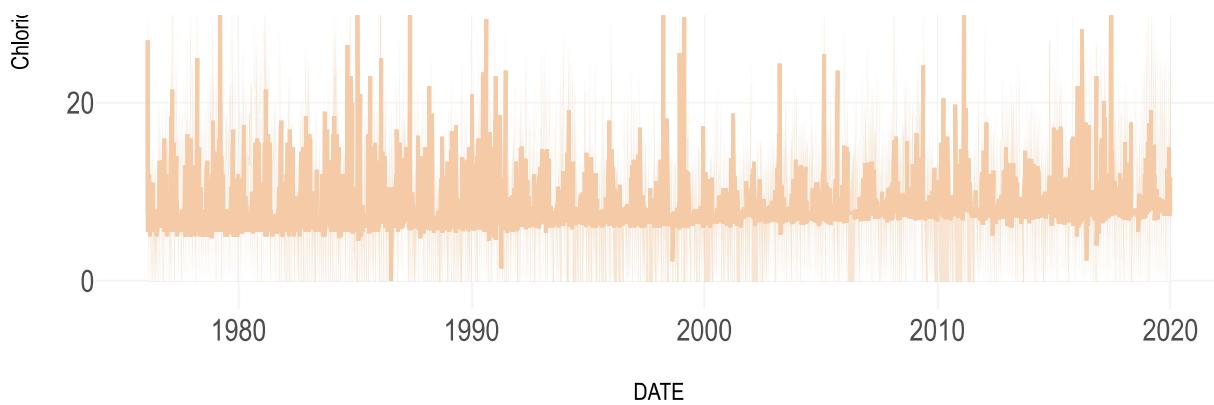


```
#nitrate for each Lake
ns = nit[nit$LAKE=="Lake Superior",]
ne = nit[nit$LAKE=="Lake Erie",]
nh = nit[nit$LAKE=="Lake Huron",]
no = nit[nit$LAKE=="Lake Ontario",]

#nitrate Level in Lake superior
tn<- ns %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggtitle("nitrate level in Lake Superior")+
  geom_area(fill="#69b3a2", alpha=0.5) +
  geom_line(color="#69b3a2") +
  ylab("nitrate level") +
  theme_ipsum()
ggplotly(th)
```

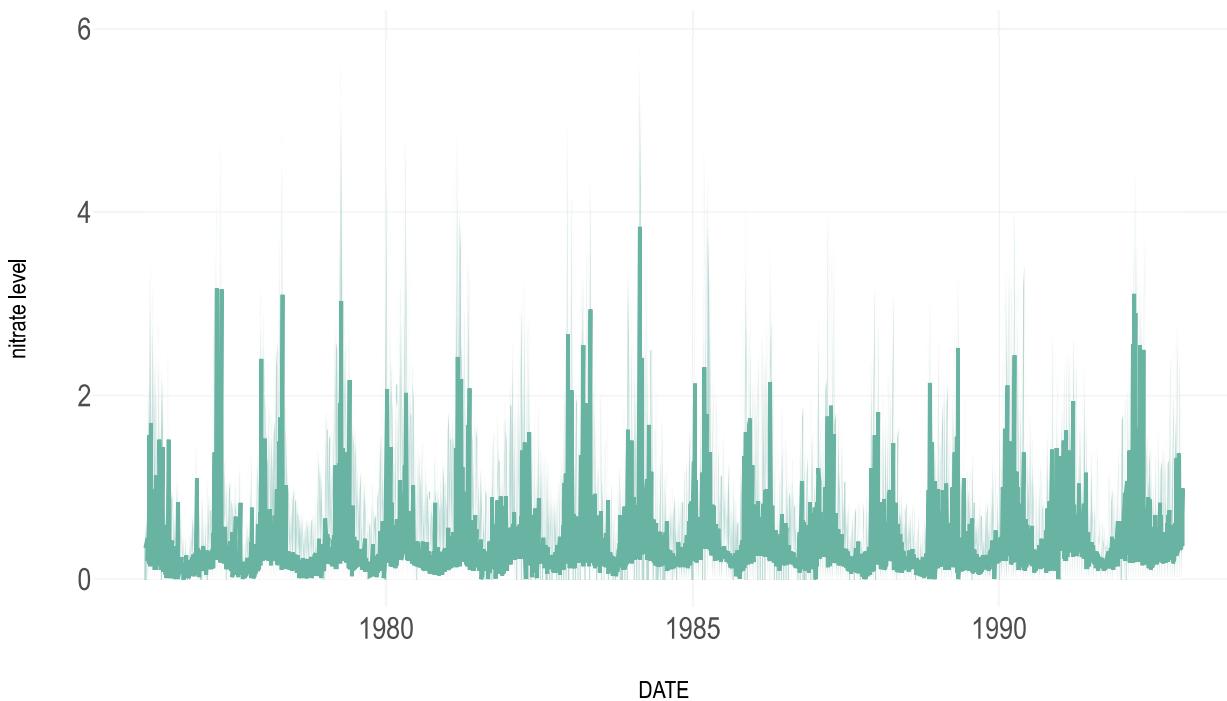
## Chloride level in Lake Huron





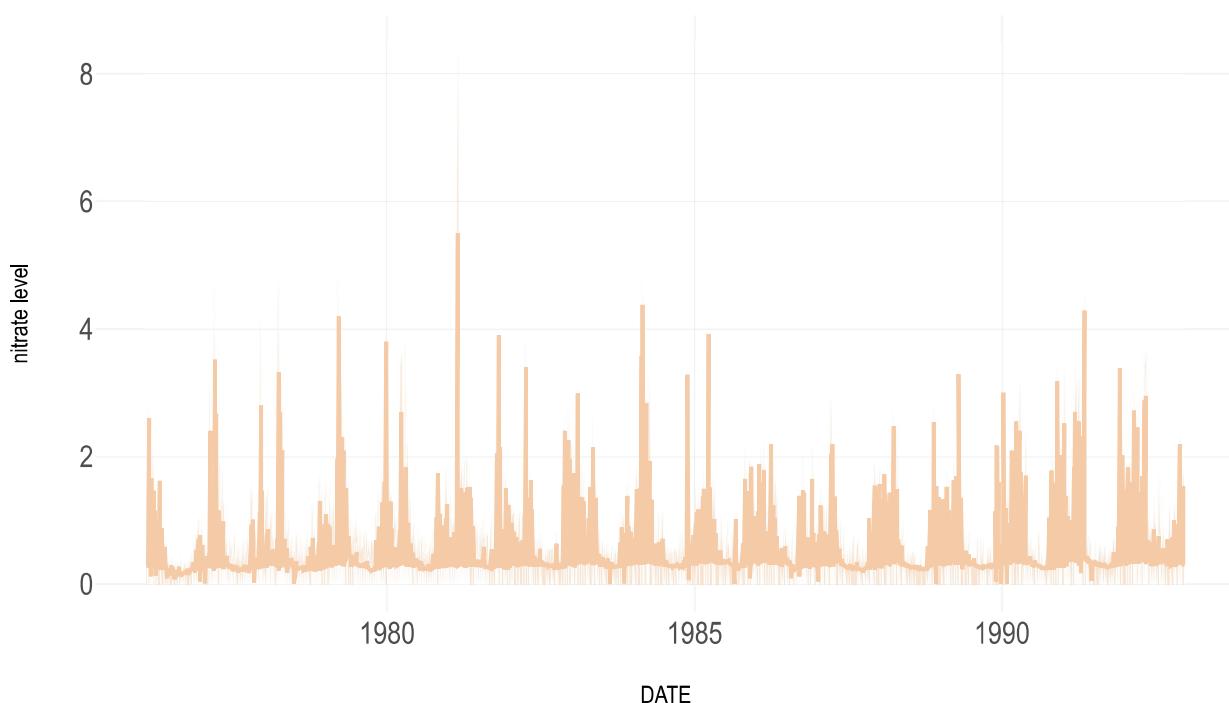
```
#nitrate Level in Lake erie
tne<- ne %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggtitle("nitrate level in Lake Erie")+
  geom_area(fill="#69b3a2", alpha=0.5) +
  geom_line(color="#69b3a2") +
  ylab("nitrate level") +
  theme_ipsum()
ggplotly(tne)
```

## nitrate level in Lake Erie



```
#Lake huron nitrate
tnh<- nh %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggttitle("nitrate level in Lake Huron")+
  geom_area(fill="#f5caba7", alpha=0.5) +
  geom_line(color="#f5caba7") +
  ylab("nitrate level") +
  theme_ipsum()
ggplotly(tnh)
```

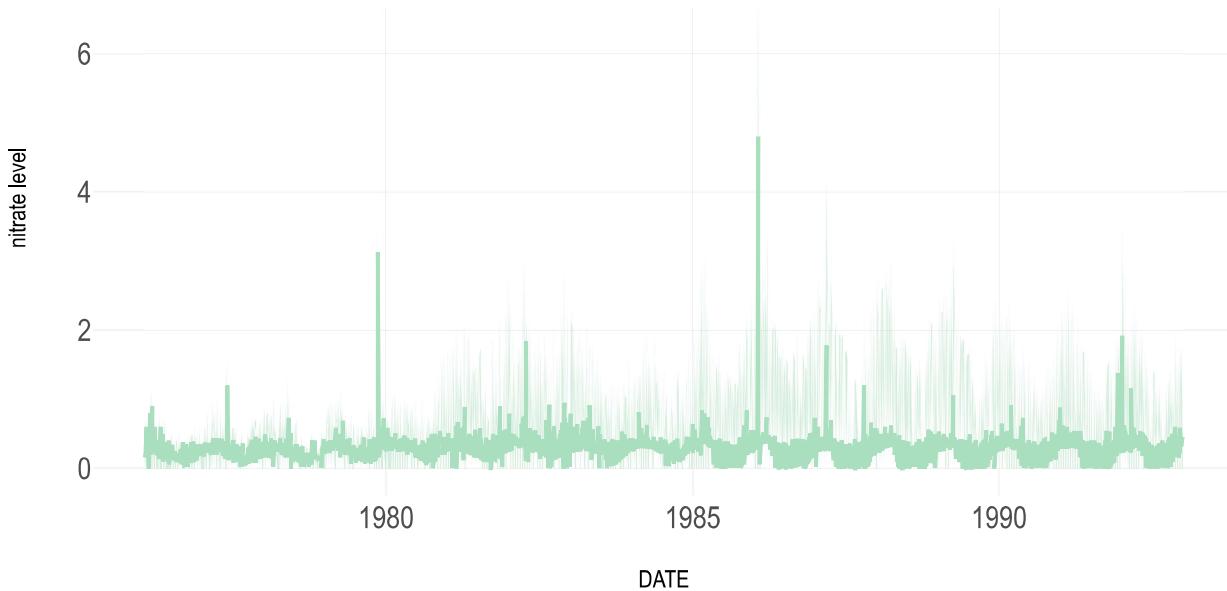
## nitrate level in Lake Huron



```
#Lake ontario nitrate
tno<- no %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  ggttitle("nitrate level in Lake Ontario")+
  geom_area(fill="#a9dfbf", alpha=0.5) +
  geom_line(color="#a9dfbf") +
  ylab("nitrate level") +
  theme_ipsum()
ggplotly(tno)
```

## nitrate level in Lake Ontario





```
#nitrogen nitrite
nitri <- filter(stats, PARAMETER=="Nitrogen; nitrite")

#nitrite
nitr <- nitri[order(nitri$VALUE,decreasing = TRUE),]
nitr %>%
  ggplot( aes(x=DATE, y=VALUE)) +
  geom_line(color="#f5b7b1") +
  ylim(0,1) +
  annotate(geom="text", x=as.Date("1991-10-01"), y=0.62,
           label=" highest nitrite level") +
  annotate(geom="point", x=as.Date("1991-10-01"), y=0.62, size=10, shape=21, fill="transparent") +
  geom_hline(yintercept=2, color="orange", size=.5) +
  theme_ipsum()

## Warning: Removed 1 rows containing missing values (geom_hline).
```

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

```

```

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font family not found in Windows font database

```

