

Project_Final

Joo Kim

2022-05-19

```
library(ggplot2)
library(lubridate)

##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
library(tidyr)
library(gridExtra)
library(reshape2)

##
## Attaching package: 'reshape2'
## The following object is masked from 'package:tidyr':
##
##   smiths
library(tseries)

## Registered S3 method overwritten by 'quantmod':
##   method      from
##   as.zoo.data.frame zoo
library(DescTools)
library("TTR")
library(Kendall)

## Warning: package 'Kendall' was built under R version 4.0.5
library(patchwork)
library(ggpubr)
library(corrplot)

## corrplot 0.84 loaded
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v tibble  3.0.4    v dplyr   1.0.7
## v readr   1.4.0    v stringr 1.4.0
## v purrr   0.3.4    v forcats 0.5.0
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x lubridate::as.difftime() masks base::as.difftime()
## x dplyr::combine() masks gridExtra::combine()
## x lubridate::date() masks base::date()
## x dplyr::filter() masks stats::filter()
## x lubridate::intersect() masks base::intersect()
## x dplyr::lag() masks stats::lag()
## x lubridate::setdiff() masks base::setdiff()
## x lubridate::union() masks base::union()
```

Data Dictionary

Field Name	Data Type	Description	Example
Date.UTC	character	Date (Universal Time Correlated)	2000-01-01T00:00:00
dbn_nez	float	Magnetic field north component, NEZ coordinates [nT]	-51.6
dbe_nez	float	Magnetic field east component, NEZ coordinates [nT]	-5.7
dbz_nez	float	Magnetic field vertical component, NEZ coordinates [nT]	15.0
dbn_geo	float	Magnetic field north component, geographic coordinates [nT]	-51.8
dbe_geo	float	Magnetic field east component, geographic coordinates [nT]	3.5
dbz_geo	float	Magnetic field vertical component, geographic coordinates [nT]	15.0
Extent	double	Extent of Record [seconds]	60
IAGA	character	Station Identifier	FRD
GEOLON	float	Geographic Longitude [degrees]	282.63
GEOLAT	float	Geographic Latitude [degrees]	38.2
MAGON	float	AACGM Longitude [degrees]	-2.13
MAGLAT	float	AACGM Latitude [degrees]	49.08
MLT	float	AACGM Magnetic Local Time [h]	18.70
MCOLAT	float	AACGM Colatitude [degrees]	40.92
IGRF_DECL	float	Magnetic Declination [degrees]	-10.18
SZA	float	Solar Zenith Angle [degrees]	113.02

- NEZ: North, East, Vertical directions
- AACGM: Altitude adjusted corrected geomagnetic
- nT: nano tesla (a unit of magnetic flux)

Visualizations

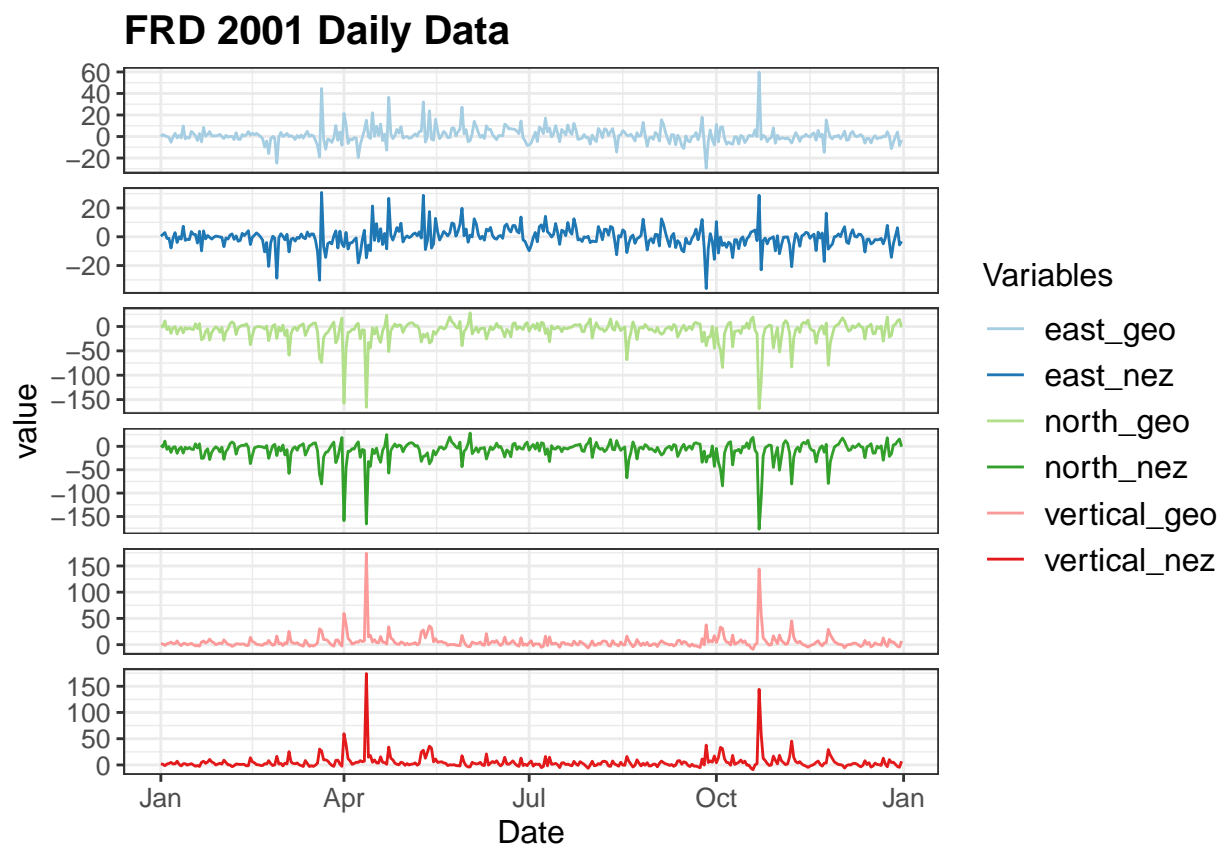
Daily Data

```
df = read.table("FRD2001.csv", sep = ',', header = T)
df_sample = df[(1:nrow(df) - 1) %% 24 == 0 ,]
df_sample$Date = ymd(substr(df_sample$Date.UTC, 1, 10))

df_sample = gather(df_sample, "Variables", "value", c(2:7))
```

```
p <- ggplot(aes(x = Date, y = value, colour = Variables), data = df_sample) +
  geom_line() + facet_grid(rows = vars(Variables), scales = "free") +
  scale_x_date(date_labels = "%b") +
  ggtitle("FRD 2001 Daily Data") +
  theme_bw() +
  theme(plot.title = element_text(face="bold", size=15), text = element_text(size = 12),
        legend.background = element_rect(fill = "white", size = 6, colour = "white"),
        strip.text.y = element_blank(),
        legend.text=element_text(size=12)) +
  scale_color_brewer(name="Variables",
                    labels=c("east_geo", "east_nez", "north_geo", "north_nez", "vertical_geo", "vertical_nez"),
                    palette="Paired")
```

p



Correlation matrix

```
clk <- read.csv("CLK2001_2.csv")
dso <- read.csv("DSO2001.csv")
gtf <- read.csv("GTF2001.csv")
frd <- read.csv("FRD2001.csv")
msh <- read.csv("MSH2001.csv")
```

keep only relevant variables

```
clk <- clk[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]
dso <- dso[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]
gtf <- gtf[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]
frd <- frd[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]
```

```

msh <- msh[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]

# number of rows of the data
print(nrow(clk))

## [1] 525600

print(nrow(frd))

## [1] 8760

# hourly data
clk2 = clk[(1:nrow(clk)-1) %% 60 == 0 ,]
frd2 = frd

# check the n of rows
print(nrow(clk2))

## [1] 8760

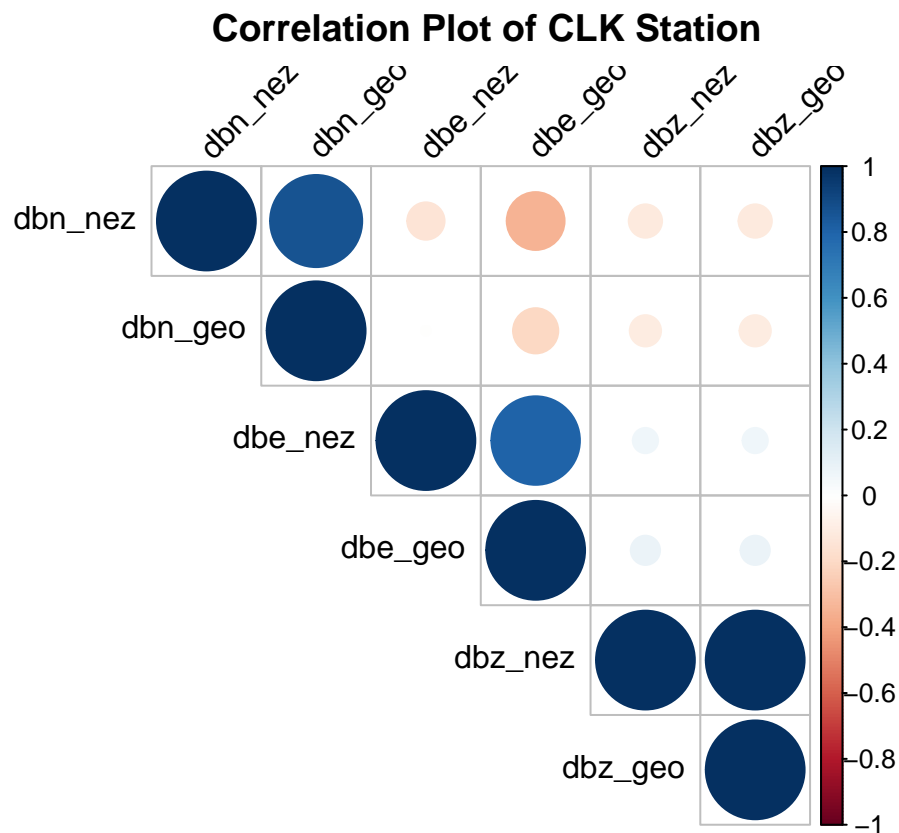
print(nrow(frd2))

## [1] 8760

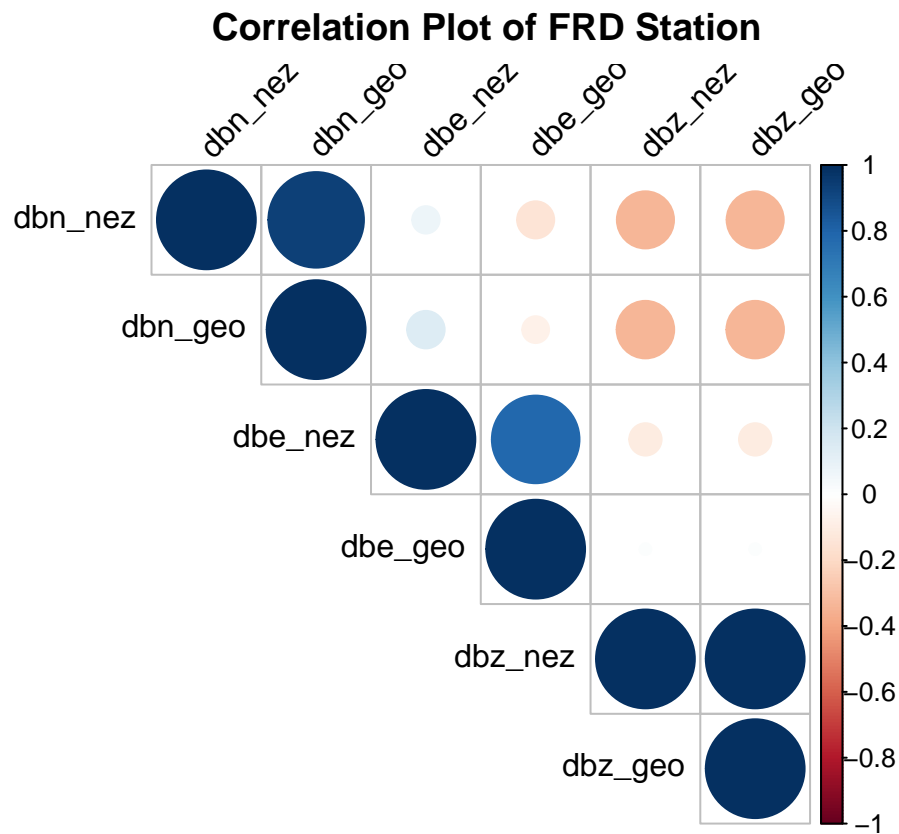
# getting correlation matrix
clk_cor = cor(clk2[,2:7], use = "complete.obs", method = c("kendall"))
frd_cor = cor(frd2[,2:7], use = "complete.obs", method = "kendall")

corrplot(clk_cor, type = "upper", order = "original",
          tl.col = "black", tl.srt = 45, title = "Correlation Plot of CLK Station",
          mar=c(0,0,2,0))

```



```
corrplot(frd_cor, type = "upper", order = "original",
         tl.col = "black", tl.srt = 45, title = "Correlation Plot of FRD Station",
         mar=c(0,0,2,0))
```

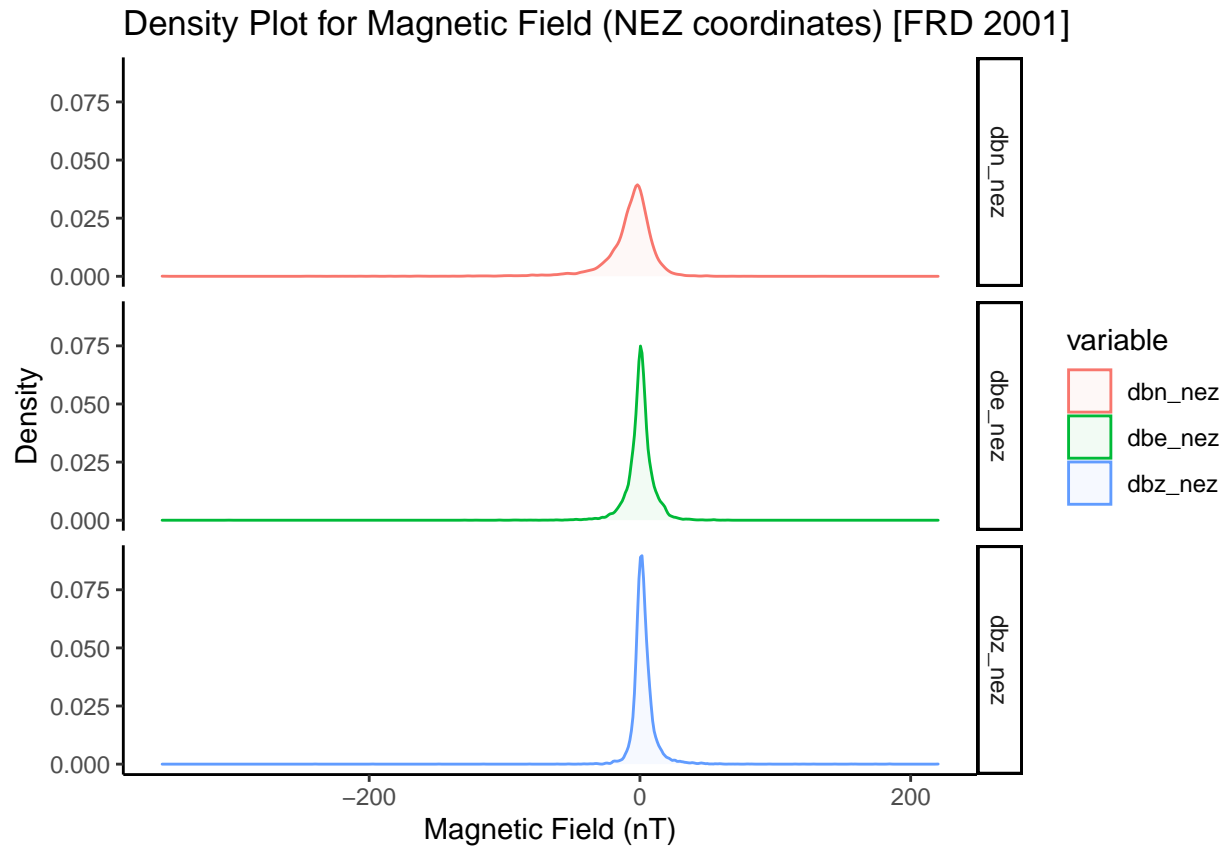


Density plots

```
frd2001 <- read.table("frd2001_samples.csv", sep = ',', header = T)
mfrd2001 <- melt(frd2001[,c("Date.UTC", "dbn_nez", "dbe_nez", "dbz_nez")], id = "Date.UTC")
mfrd2001_g <- melt(frd2001[,c("Date.UTC", "dbn_geo", "dbe_geo", "dbz_geo")], id = "Date.UTC")
```

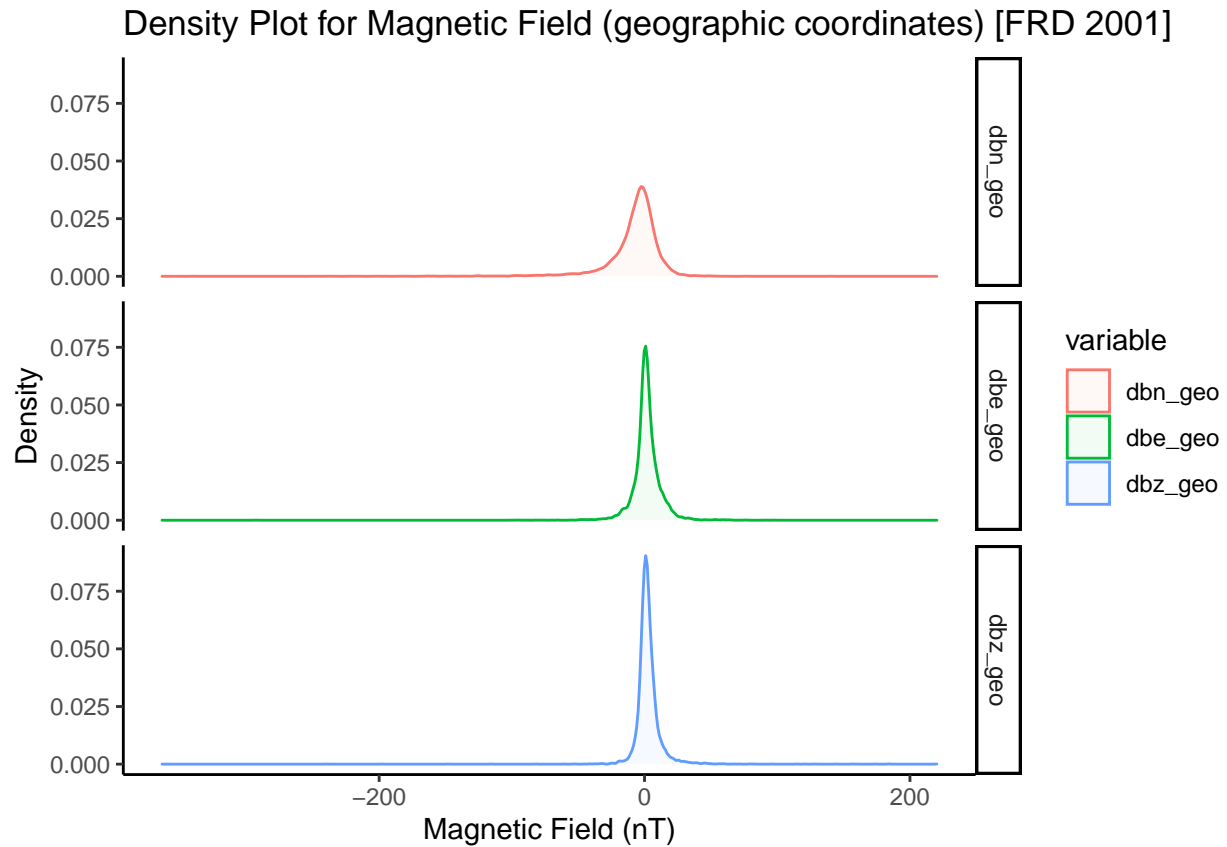
```
# NEZ coordinates
ggplot(mfrd2001, aes(x=value, color=variable, fill=variable)) +
  geom_density(alpha=0.05) +
  facet_grid(rows = vars(variable)) +
  ggtitle("Density Plot for Magnetic Field (NEZ coordinates) [FRD 2001]") +
  xlab("Magnetic Field (nT)") +
  ylab("Density") +
  theme_classic()
```

```
## Warning: Removed 96 rows containing non-finite values (stat_density).
```



```
# geo coordinates
ggplot(mfrd2001_g, aes(x=value, color=variable, fill=variable)) +
  geom_density(alpha=0.05) +
  facet_grid(rows = vars(variable)) +
  ggtitle("Density Plot for Magnetic Field (geographic coordinates) [FRD 2001]") +
  xlab("Magnetic Field (nT)") +
  ylab("Density") +
  theme_classic()
```

Warning: Removed 96 rows containing non-finite values (stat_density).



acf of hourly and daily data

```
frd2001 <- read.table("frd2001_samples.csv", sep = ',', header = T)
nrow(frd2001)
```

```
## [1] 8760
```

```
# finding the max magnitude value
```

```
max_abs = function(x){
  x = na.omit(x)
  if(length(x) == 0) return (0)
  else{
    return(x[which.max(abs(x))])
  }
}
```

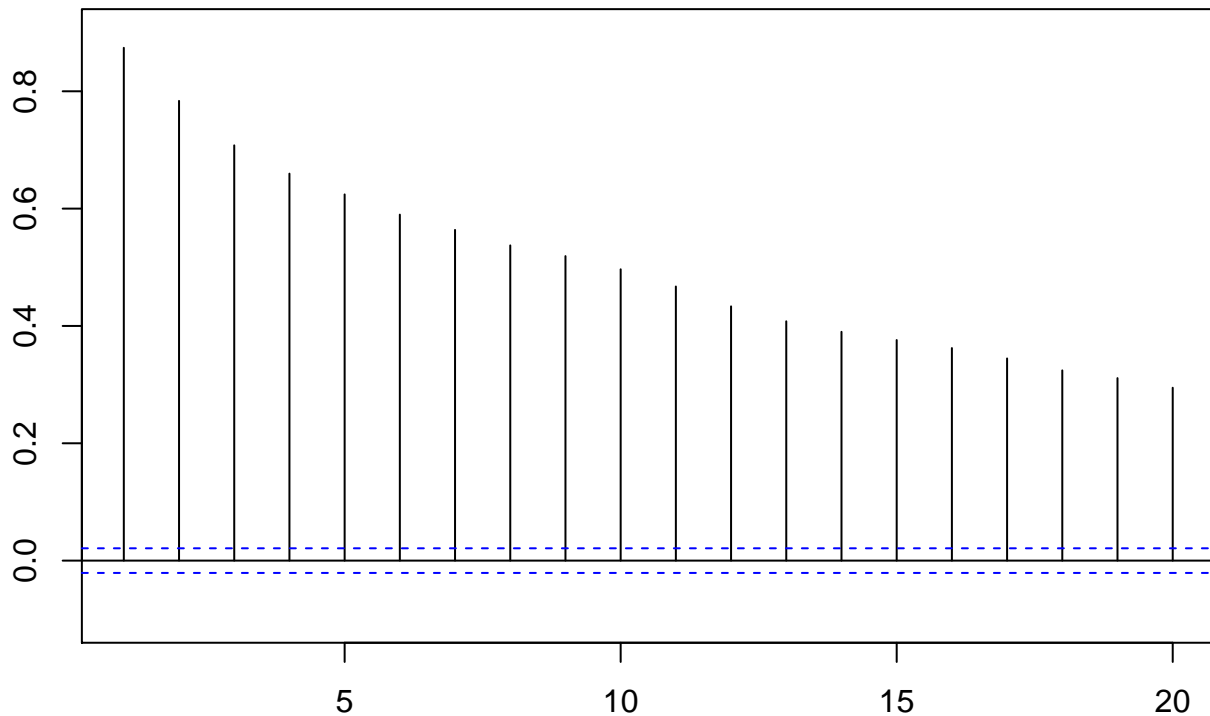
```
# obtaining daily data (using max_abs)
```

```
daily_pooling = function(df, varname, func)
{
  y = as.vector(df[, varname])
  y = matrix(y, ncol = 24, byrow = T)
  y = apply(y, 1, func) # max pooling
  return(y)
}
```

```
# summarizing into daily data
daily_dbn <- daily_pooling(frd2001, "dbn_nez", max_abs)
daily_dbe <- daily_pooling(frd2001, "dbe_nez", max_abs)
daily_dbz <- daily_pooling(frd2001, "dbz_nez", max_abs)

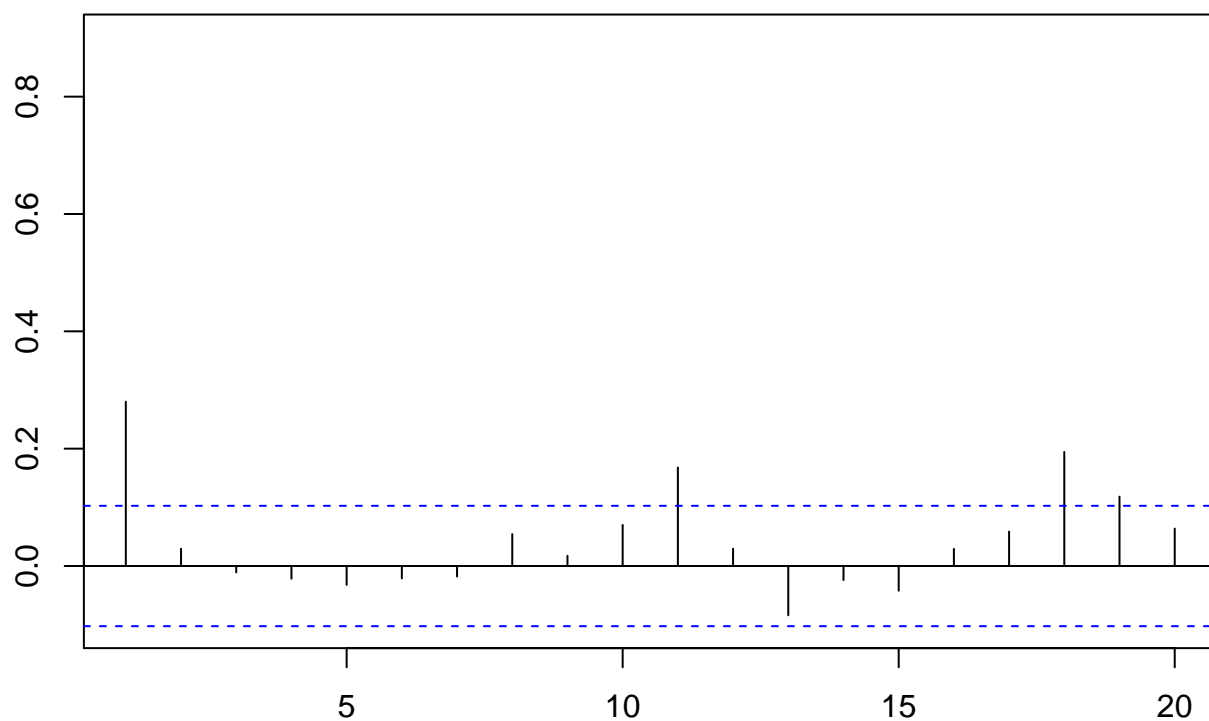
par(mar=c(3,3,3,0))
acf_dbn_h <- acf(frd2001$dbn_nez[!is.na(frd2001$dbn_nez)], plot = FALSE)
plot(acf_dbn_h[1:20], main="Autocorrelation for Hourly Data (north_nez)", ylim=c(-0.1,0.9))
```

Autocorrelation for Hourly Data (north_nez)



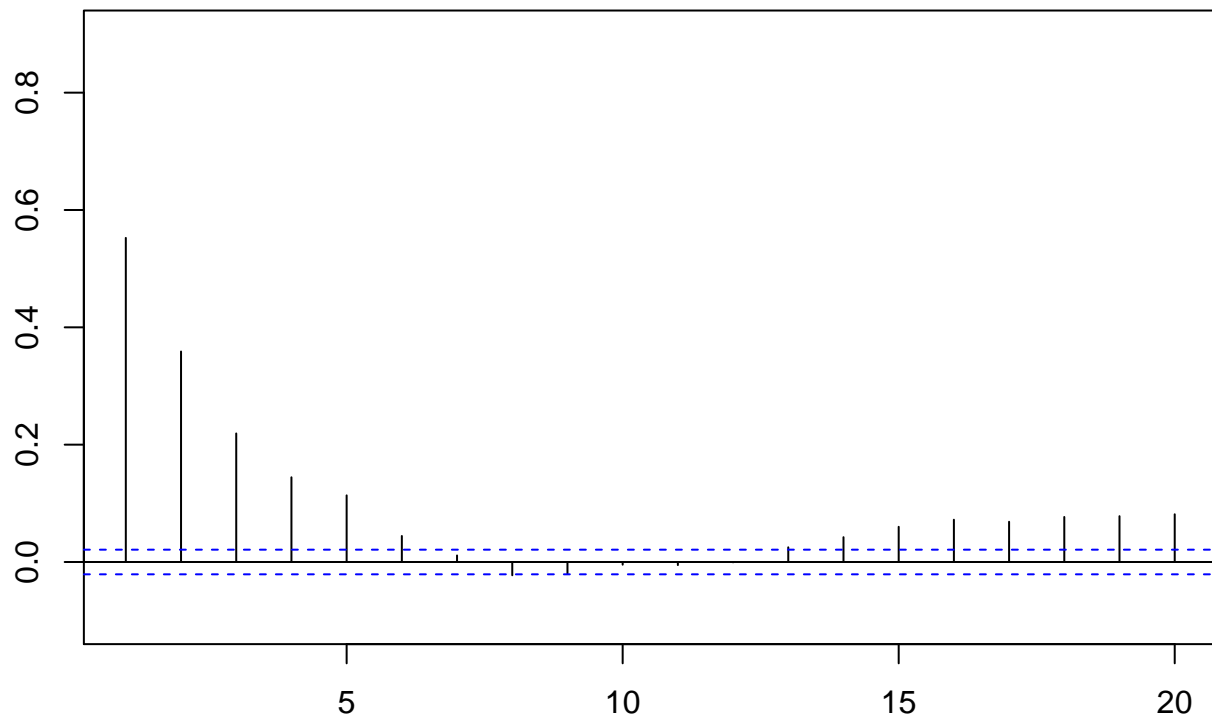
```
par(mar=c(3,3,3,0))
acf_dbn_d <- acf(daily_dbn, plot = FALSE)
plot(acf_dbn_d[1:20], main = "Autocorrelation for Daily Data (north_nez)", ylim=c(-0.1,0.9))
```


Autocorrelation for Daily Data (north_nez)



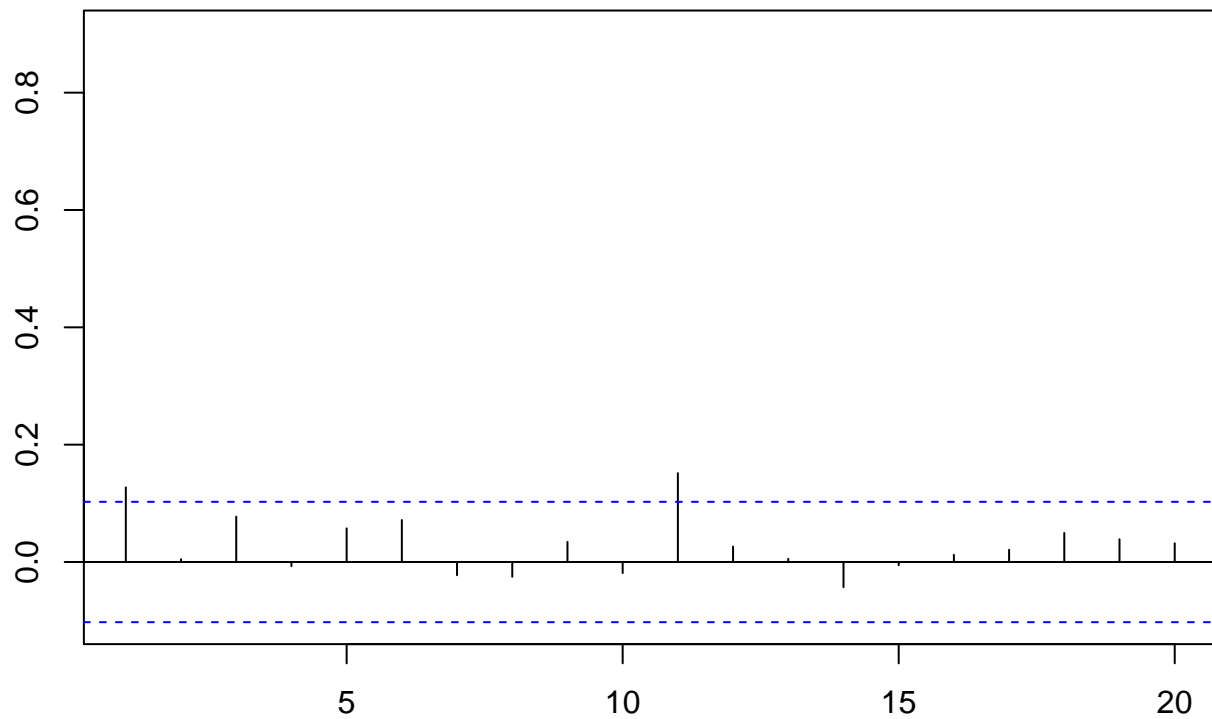
```
par(mar=c(3,3,3,0))
acf_dbe_h <- acf(frd2001$dbe_nez[!is.na(frd2001$dbe_nez)], plot = FALSE)
plot(acf_dbe_h[1:20], main="Autocorrelation for Hourly Data (east_nez)", ylim=c(-0.1,0.9))
```

Autocorrelation for Hourly Data (east_nez)



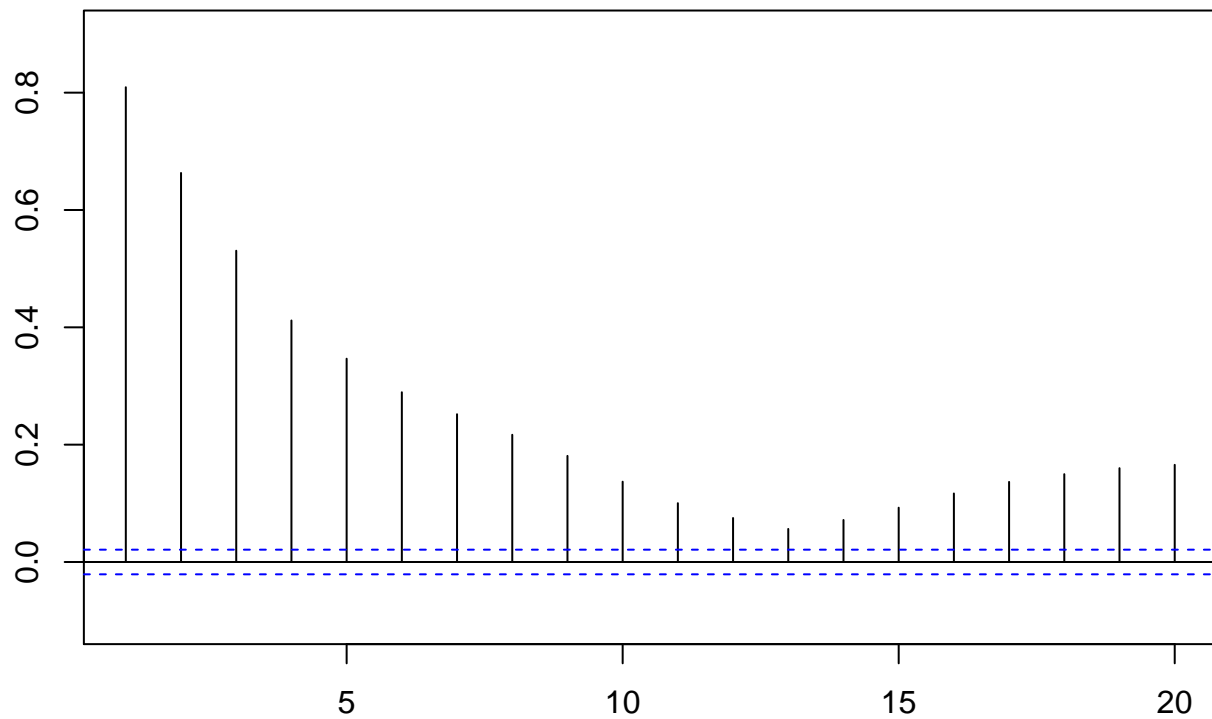
```
par(mar=c(3,3,3,0))
acf_dbe_d <- acf(daily_dbe, plot = FALSE)
plot(acf_dbe_d[1:20], main = "Autocorrelation for Daily Data (east_nez)", ylim=c(-0.1,0.9))
```

Autocorrelation for Daily Data (east_nez)



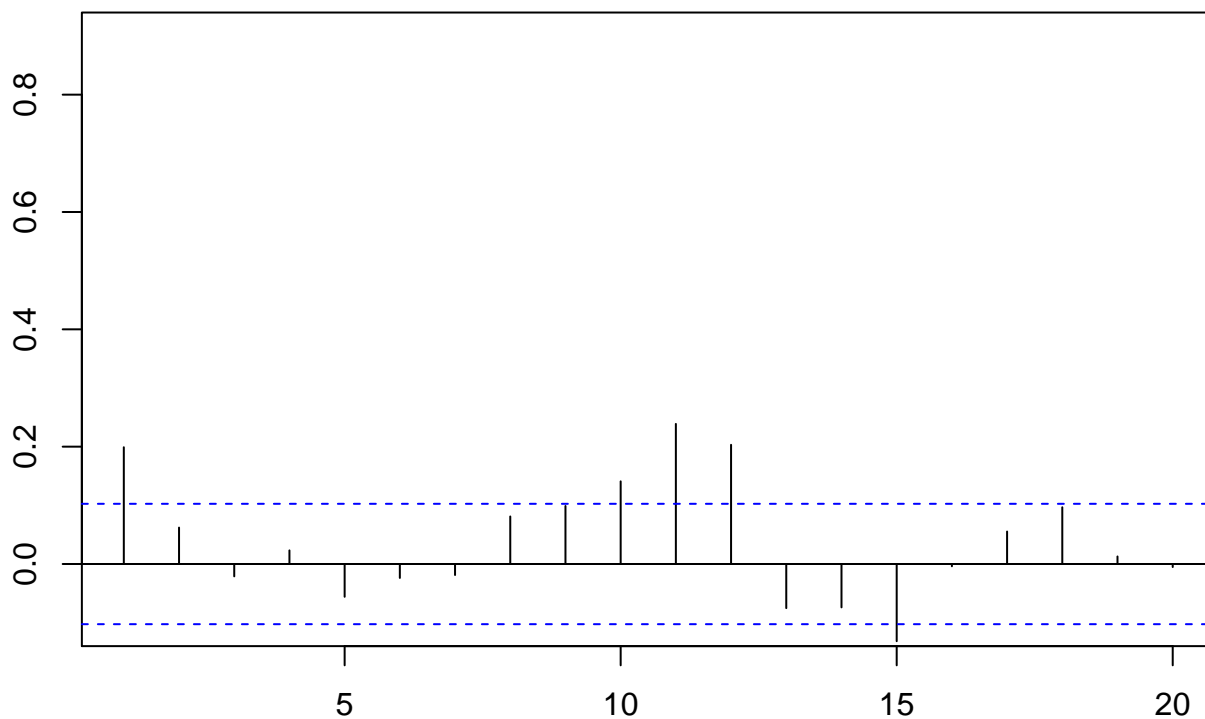
```
par(mar=c(3,3,3,0))
acf_dbz_h <- acf(frd2001$dbz_nez[!is.na(frd2001$dbz_nez)], plot = FALSE)
plot(acf_dbz_h[1:20], main="Autocorrelation for Hourly Data (east_nez)", ylim=c(-0.1,0.9))
```

Autocorrelation for Hourly Data (east_nez)



```
par(mar=c(3,3,3,0))
acf_dbz_d <- acf(daily_dbz, plot = FALSE)
plot(acf_dbz_d[1:20], main = "Autocorrelation for Daily Data (vertical_nez)", ylim=c(-0.1,0.9))
```

Autocorrelation for Daily Data (vertical_nez)



Comparing Stations

```
clk <- read.csv("CLK2001_2.csv")
msh <- read.csv("MSH2001.csv")

clk <- clk[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]
frd <- frd[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]
msh <- msh[,c("Date.UTC", "dbn_nez", "dbn_geo", "dbe_nez", "dbe_geo", "dbz_nez", "dbz_geo")]

clk2 = clk[(1:nrow(clk)-1) %% 60 == 0 ,]
frd2 = frd
msh2 = msh[(1:nrow(clk)-1) %% 60 == 0 ,]

# check the number of rows for each data is consistent
print(nrow(clk2))

## [1] 8760
print(nrow(frd2))

## [1] 8760
print(nrow(msh2))

## [1] 8760

# Date
clk2$Date = ymd(substr(clk2$Date.UTC, 1, 10))
frd2$Date = ymd(substr(frd2$Date.UTC, 1, 10))
msh2$Date = ymd(substr(msh2$Date.UTC, 1, 10))
```

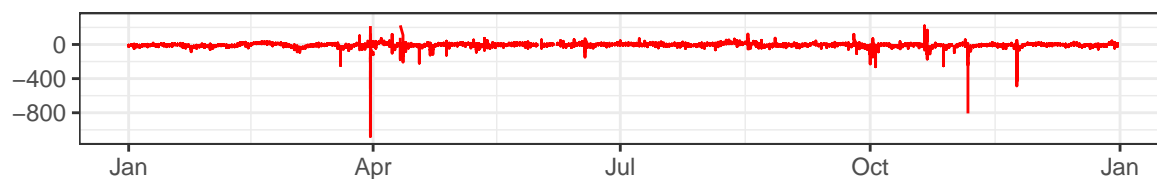
```

# For the presentation
p1 <- ggplot() +
  geom_line(data = clk2, aes(x = Date, y = dbn_nez), color = "red") +
  scale_x_date(date_labels = "%b") +
  ggtitle("CLK 2001 Hourly Data (North)") +
  theme_bw() +
  ylim(-1100, 300) +
  xlab("") + ylab("")
p2 <- ggplot() +
  geom_line(data = frd2, aes(x = Date, y = dbn_nez), color = "orange") +
  scale_x_date(date_labels = "%b") +
  ggtitle("FRD 2001 Hourly Data (North)") +
  theme_bw() +
  ylim(-1100, 300) +
  xlab("") + ylab("")
p3 <- ggplot() +
  geom_line(data = msh2, aes(x = Date, y = dbn_nez), color = "darkblue") +
  scale_x_date(date_labels = "%b") +
  ggtitle("MSH 2001 Hourly Data (North)") +
  theme_bw() +
  ylim(-1100, 300) +
  ylab("")

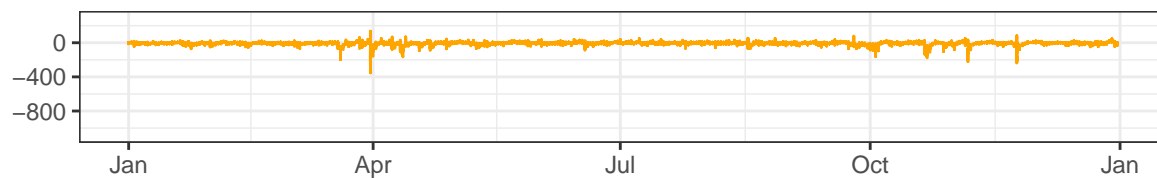
p1 / p2 / p3

```

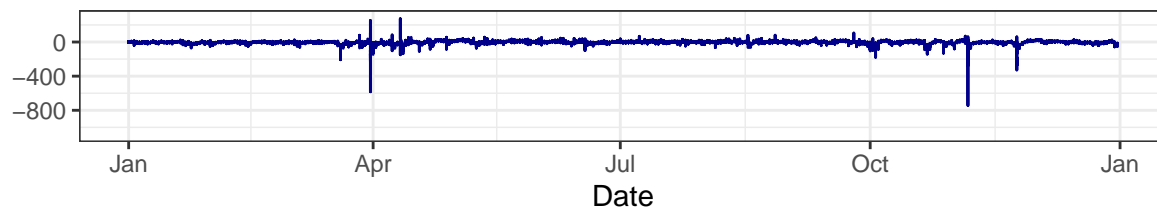
CLK 2001 Hourly Data (North)



FRD 2001 Hourly Data (North)



MSH 2001 Hourly Data (North)



```

p1 <- ggplot() +
  geom_line(data = clk2, aes(x = Date, y = dbn_nez), color = "red") +
  scale_x_date(date_labels = "%b") +

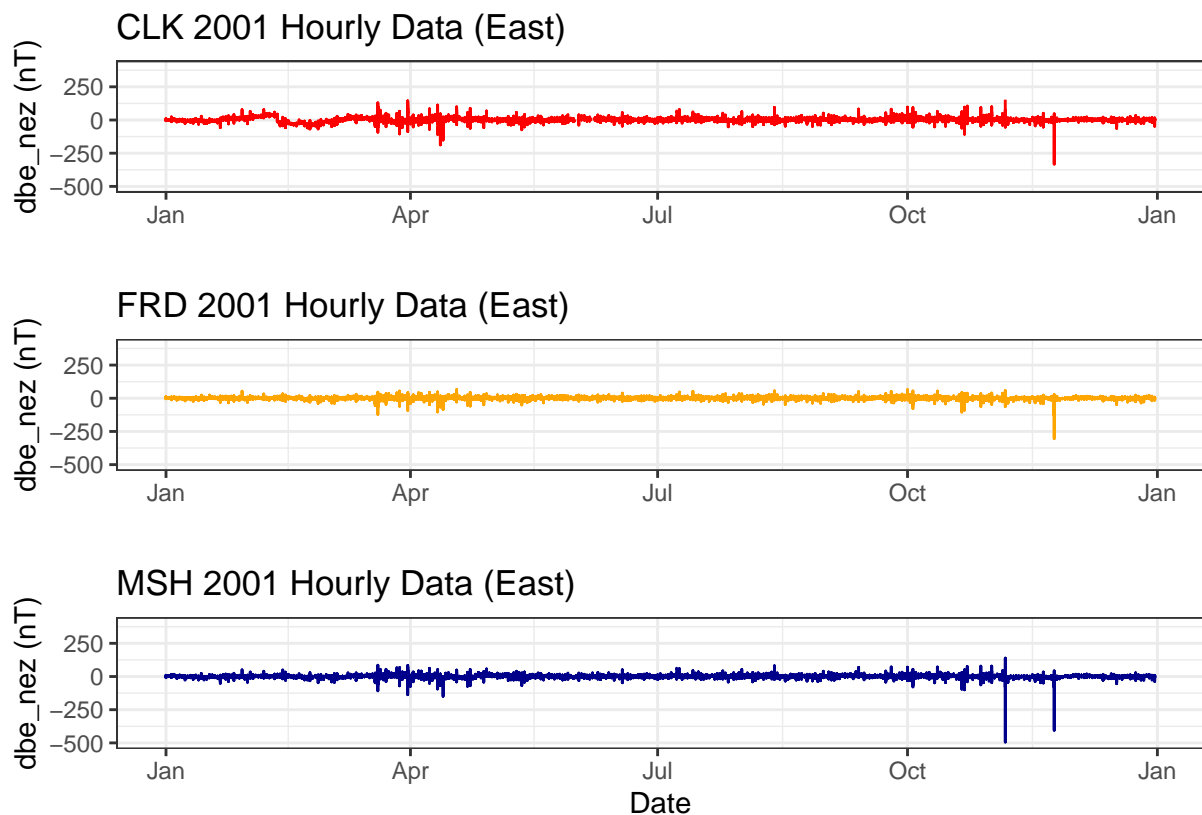
```

```

ggtitle("CLK 2001 Hourly Data (East)") +
theme_bw() +
ylim(-500, 400) +
xlab("") + ylab("dbe_nez (nT)")
p2 <- ggplot() +
geom_line(data = frd2, aes(x = Date, y = dbe_nez), color = "orange") +
scale_x_date(date_labels = "%b") +
ggtitle("FRD 2001 Hourly Data (East)") +
theme_bw() +
ylim(-500, 400) +
xlab("") + ylab("dbe_nez (nT)")
p3 <- ggplot() +
geom_line(data = msh2, aes(x = Date, y = dbe_nez), color = "darkblue") +
scale_x_date(date_labels = "%b") +
ggtitle("MSH 2001 Hourly Data (East)") +
theme_bw() +
ylim(-500, 400) +
ylab("dbe_nez (nT)")

```

p1 / p2 / p3



```

p1 <- ggplot() +
geom_line(data = clk2, aes(x = Date, y = dbz_nez), color = "red") +
scale_x_date(date_labels = "%b") +
ggtitle("CLK 2001 Hourly Data (Vertical)") +
theme_bw() +
ylim(-450, 300) +
xlab("") + ylab("dbz_nez (nT)")

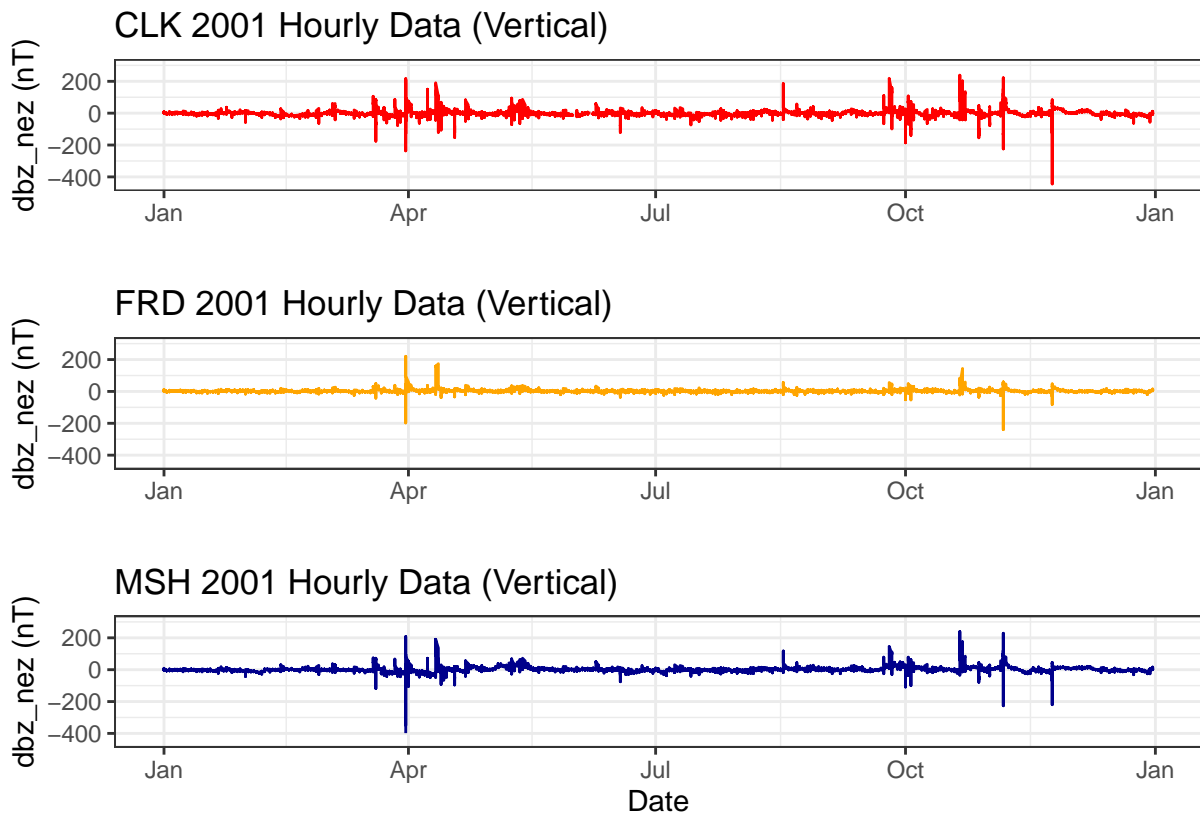
```

```

p2 <- ggplot() +
  geom_line(data = frd2, aes(x = Date, y = dbz_nez), color = "orange") +
  scale_x_date(date_labels = "%b") +
  ggtitle("FRD 2001 Hourly Data (Vertical)") +
  theme_bw() +
  ylim(-450, 300) +
  xlab("") + ylab("dbz_nez (nT)")
p3 <- ggplot() +
  geom_line(data = msh2, aes(x = Date, y = dbz_nez), color = "darkblue") +
  scale_x_date(date_labels = "%b") +
  ggtitle("MSH 2001 Hourly Data (Vertical)") +
  theme_bw() +
  ylim(-450, 300) +
  ylab("dbz_nez (nT)")

```

p1 / p2 / p3



Correlation of same variables from different stations

```

# for presentation
par(mar=c(3,3,3,0))
ggplot(data = data.frame(x = clk2$dbn_nez, y = frd2$dbn_nez), aes(x=x, y=y)) +
  geom_point() +
  stat_cor(method = "pearson", label.x = -1000, label.y = 150, size=7) +
  theme_bw() +
  ggtitle("The Correlation between north_nez Variables from Different Stations (CLK, FRD)") +

```

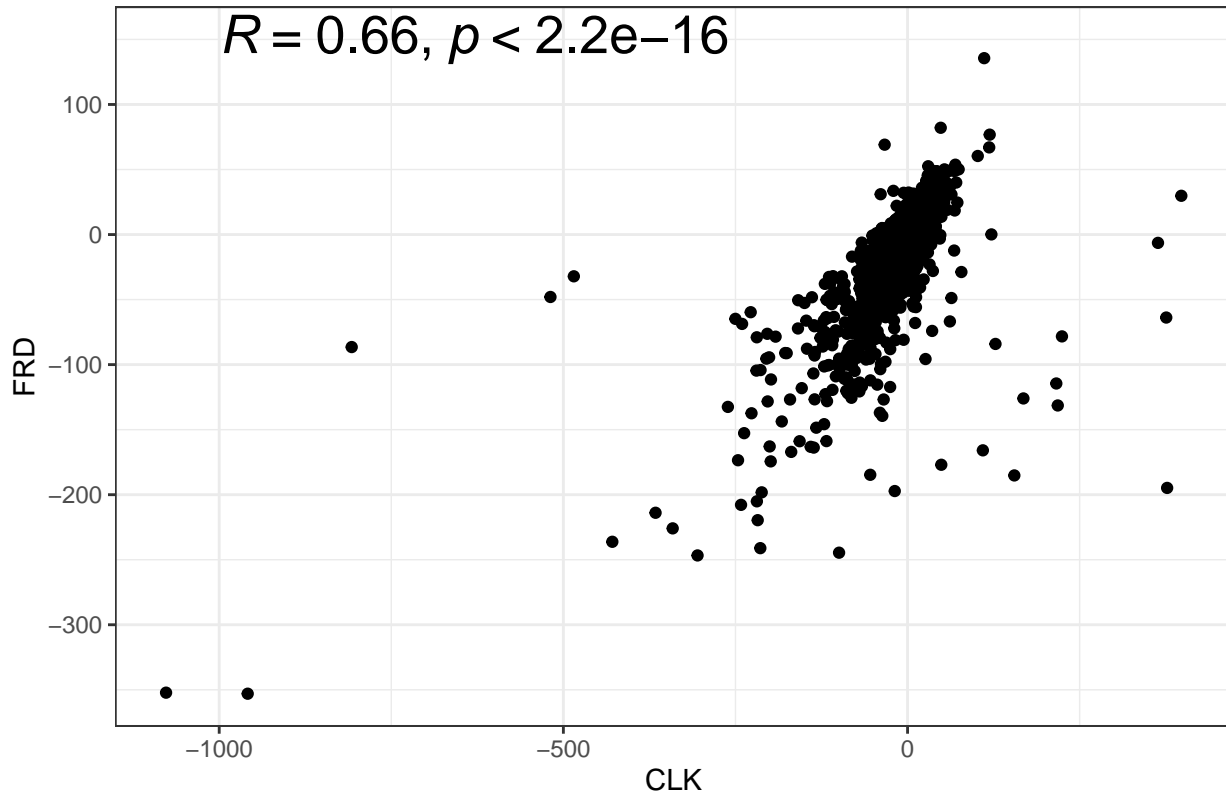


```
xlab("CLK") +
ylab("FRD") +
theme(plot.title = element_text(face="bold", size=12))
```

Warning: Removed 140 rows containing non-finite values (stat_cor).

Warning: Removed 140 rows containing missing values (geom_point).

The Correlation between north_nez Variables from Different Stations (CLK,

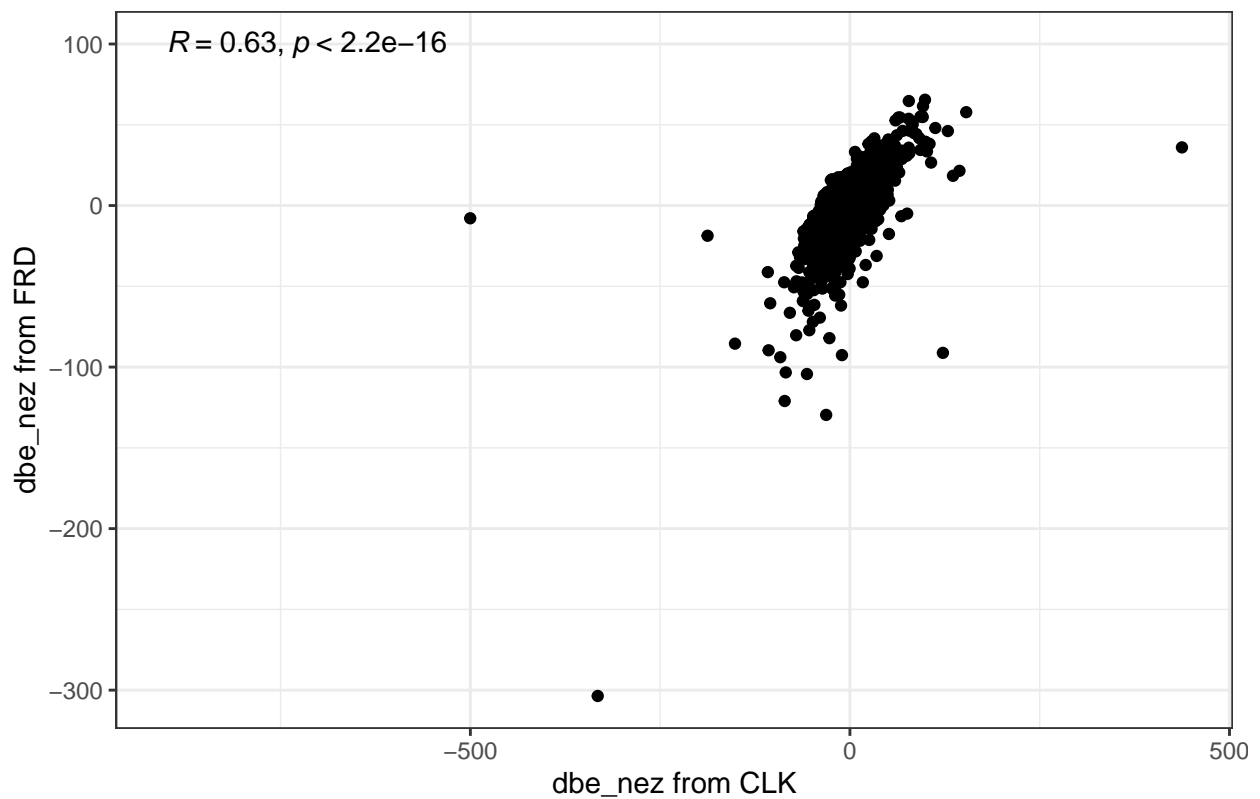


```
ggplot(data = data.frame(x = clk2$db_e_nez, y = frd2$db_e_nez), aes(x=x, y=y)) +
  geom_point() +
  stat_cor(method = "pearson", label.x = -900, label.y = 100) +
  theme_bw() +
  ggtitle("The Correlation between db_e_nez Variables from Different Stations (CLK, FRD)") +
  xlab("db_e_nez from CLK") +
  ylab("db_e_nez from FRD")
```

Warning: Removed 140 rows containing non-finite values (stat_cor).

Warning: Removed 140 rows containing missing values (geom_point).

The Correlation between dbe_nez Variables from Different Stations (CLK,



```
ggplot(data = data.frame(x = clk$dbz_nez, y = frd2$dbz_nez), aes(x=x, y=y)) +
  geom_point() +
  stat_cor(method = "pearson", label.x = -900, label.y = 150) +
  theme_bw() +
  ggtitle("The Correlation between dbz_nez Variables from Different Stations (CLK, FRD)") +
  xlab("dbz_nez from CLK") +
  ylab("dbz_nez from FRD")
```

```
## Warning: Removed 140 rows containing non-finite values (stat_cor).
```

```
## Warning: Removed 140 rows containing missing values (geom_point).
```

The Correlation between dbz_nez Variables from Different Stations (CLK,

