ENV872 Final Project

https://github.com/Jiaqi-Li-Duke/ENV872_Project_jl769

Jiaqi Li

Abstract

High concentrations of chloroprene have been measured in the vicinity of the Denka Performance Elastomer facility in LaPlace, LA. Chloroprene concentrations at six monitoring sites and meteorology data are available since May 2016. New emission reduction projects were implemented by the company to reduce chloroprene emissions in 2018. This project aims to investigate the relationship between wind speed and chloroprene concentrations and the effects of emission reduction projects using multiple statistical approaches. The results show that there is a statistically significant negative correlation between chloroprene concentration and wind speed at the three monitoring sites within 1 km to the Denka facility. Chloroprene concentrations decline significantly from 2016 to 2018, and the changing points occurred around January 2018. However, the current concentrations still far exceed the recommended level of 0.2 µg/m^3 without increasing risk of cancer. More efforts need to be made to protect public health in the LaPlace community.

Contents

List of Tables

List of Figures

1 Research Question

High concentrations of chloroprene, which is a monomer used to produce synthetic rubber and is classified as likely to be carcinogenic to humans, have been measured in the vicinity of the Denka Performance Elastomer facility in LaPlace, LA. New emission reduction projects were implemented by the company to reduce chloroprene emissions in 2018.

This project aims to answer two questions. First, is there any correlation between chloroprene concentrations and wind speed? Second, is there is a decline observed in the chloroprene concentrations over time and if so, at what point does the change point occur?

2 Dataset Information

There are two parts of data in this analysis, chloroprene concentrations at the six monitoring sites and meteorology data collected at the meteorological station. Chloroprene concentrations are available from May 2016 to December 2018 and are measured from noon to noon for 24 hours continuously every three days by U.S. EPA. For instance, data for 5/31/2016 is the mean chloroprene concentration from 5/30/2016 noon to 5/31/2016 noon.

EPA also collects local scale minute-by-minute level meteorology data, including air pressure, dewpoint, precipitation, relative humidity, temperature, wind direction, and wind speed since May 2016. To match with chloroprene data, the daily averages of meteorology data are computed from noon to noon.

```r
# Import 2016 Meteorology Data
wind2016 <- read.csv("../Data/Raw/2016Weather.csv", header = T)
colnames(wind2016) <- c('Time', 'Air.Pressure', 'Dewpoint', 'Precipitation',
                        'Precipitation.Intensity','Relative.Humidity',
                        'Temperature', 'Total.Precipitation',
                        'Wind.Chill.Temperature', 'Wind.Direction',
                        'Wind.Direction.vct', 'Wind.Speed', 'Wind.Speed.avg',
                        'Wind.Speed.max')
# Create wind components
wind2016$u.wind <- - wind2016$Wind.Speed * sin(2*pi*wind2016$Wind.Direction/360)
wind2016$v.wind <- - wind2016$Wind.Speed * cos(2*pi*wind2016$Wind.Direction/360)

# Convert time to date
wind2016$Time <- as.POSIXct(wind2016$Time)
wind2016$Date.noon <- as.Date(wind2016$Time, format = "%Y/%m/%d %H:%M:%S", tz =
                                "Antarctica/Davis")
wind2016$Date <- as.Date(wind2016$Time, format = "%Y/%m/%d %H:%M:%S", tz =
                           "America/Toronto")

# Compute mean wind speed and direction
mean2016 <- as.data.frame(aggregate(cbind(Temperature, Wind.Speed, u.wind,
                                          v.wind)~Date.noon, wind2016, mean))
mean2016$Wind.direction.avg <- (atan2(mean2016$u.wind, mean2016$v.wind)
                                * 360/2/pi) + 180
mean2016 <- select(mean2016, 'Date.noon', 'Temperature', 'Wind.Speed',
                   'Wind.direction.avg')

# Import 2017 Meteorology Data
wind2017 <- read.csv("../Data/Raw/2017Weather.csv", header = T)
colnames(wind2017) <- c('Time', 'Air.Pressure', 'Dewpoint', 'Precipitation',
                        'Precipitation.Intensity','Relative.Humidity',
                        'Temperature', 'Total.Precipitation',
```

```
                         'Wind.Chill.Temperature', 'Wind.Direction',
                         'Wind.Direction.vct', 'Wind.Speed', 'Wind.Speed.avg',
                         'Wind.Speed.max')
# Create wind components
wind2017$u.wind <- - wind2017$Wind.Speed * sin(2*pi*wind2017$Wind.Direction/360)
wind2017$v.wind <- - wind2017$Wind.Speed * cos(2*pi*wind2017$Wind.Direction/360)

# Convert time to date
wind2017$Time <- strptime(wind2017$Time, format= "%m/%d/%Y %H:%M", tz =
                             "America/Toronto")
wind2017$Time <- as.POSIXct(wind2017$Time)
wind2017$Date.noon <- as.Date(wind2017$Time, format = "%Y/%m/%d %H:%M:%S", tz =
                                 "Antarctica/Davis")
wind2017$Date <- as.Date(wind2017$Time, format = "%Y/%m/%d %H:%M:%S", tz =
                            "America/Toronto")

# Compute mean
mean2017 <- as.data.frame(aggregate(cbind(Temperature, Wind.Speed, u.wind,
                                          v.wind)~Date.noon, wind2017, mean))
mean2017$Wind.direction.avg <- (atan2(mean2017$u.wind, mean2017$v.wind)
                                 *360/2/pi) + 180
mean2017 <- select(mean2017, 'Date.noon', 'Temperature', 'Wind.Speed',
                   'Wind.direction.avg')

# Import 2018 Meteorology Data
wind2018 <- read.csv("../Data/Raw/2018Weather.csv", header = T)
colnames(wind2018) <- c('Time', 'Air.Pressure', 'Dewpoint', 'Precipitation',
                         'Precipitation.Intensity','Relative.Humidity',
                         'Temperature', 'Total.Precipitation',
                         'Wind.Chill.Temperature', 'Wind.Direction',
                         'Wind.Direction.vct', 'Wind.Speed', 'Wind.Speed.avg',
                         'Wind.Speed.max')
# Create wind components
wind2018$u.wind <- - wind2018$Wind.Speed * sin(2*pi*wind2018$Wind.Direction/360)
wind2018$v.wind <- - wind2018$Wind.Speed * (2*pi*wind2018$Wind.Direction/360)

# Convert time to date
wind2018$Time <- strptime(wind2018$Time, format= "%Y-%m-%d %H:%M:%S", tz =
                             "America/Toronto")
wind2018$Time <- as.POSIXct(wind2018$Time)
wind2018$Date.noon <- as.Date(wind2018$Time, format = "%Y-%m-%d %H:%M:%S", tz =
                                 "Antarctica/Davis")
wind2018$Date <- as.Date(wind2018$Time, format = "%Y-%m-%d %H:%M:%S", tz =
                            "America/Toronto")
```

```r
# Compute mean
mean2018 <- as.data.frame(aggregate(cbind(Temperature, Wind.Speed, u.wind,
                                          v.wind)~Date.noon, wind2018, mean))
mean2018$Wind.direction.avg <- (atan2(mean2018$u.wind, mean2018$v.wind)
                                *360/2/pi) + 180
mean2018 <- select(mean2018, 'Date.noon', 'Temperature', 'Wind.Speed',
                   'Wind.direction.avg')

#Merge with chloroprene data

air <- read.csv("../Data/Raw/Air.csv", header = T)
colnames(air) <- c("Date.noon", "Chad.Baker", "Hwy44", "Highschool",
                   "Elementary.School", "Levee", "Ochsner.Hospital")
air$Date.noon <- as.Date(air$Date.noon, format= "%m/%d/%Y")
data2016 <- merge(mean2016, air, by = "Date.noon")
data2016 <- mutate(data2016, Year= year(Date.noon), Month = month(Date.noon),
                   Week = week(Date.noon))
data2017 <- merge(mean2017, air, by = "Date.noon")
data2017 <- mutate(data2017, Year= year(Date.noon), Month = month(Date.noon),
                   Week = week(Date.noon))
data2018 <- merge(mean2018, air, by = "Date.noon")
data2018 <- mutate(data2018, Year= year(Date.noon), Month = month(Date.noon),
                   Week = week(Date.noon))
data.all <- rbind(data2016, data2017, data2018)

# Gather the chloroprene concentrations
data.gather <- gather(data.all, "Site.Name", "Concentration",
                      Chad.Baker:Ochsner.Hospital) %>%
  na.exclude() %>%
  filter(Concentration > 0.05)

# Save the processed data
#write.csv(data.all, row.names = FALSE,
#file ="./Data/Processed/Chloroprene_Meteorology_all.csv")
```

A summary of chloroprene concentrations (µg/m^3) at the six monitoring sites is shown in Table 1.

Table 1: Summary of chloroprene concentration

Site.Name	Mean	Minimum	Maximum	SD
Chad.Baker	8.978457	0.051	70.002	12.122639
Elementary.School	7.886266	0.052	149.616	15.390742
Highschool	2.382548	0.051	39.535	4.927251
Hwy44	5.826461	0.054	153.424	17.247675
Levee	6.652409	0.070	146.895	15.294339
Ochsner.Hospital	4.859043	0.053	89.225	11.789437

3 Exploratory Data Analysis and Wrangling

For the relationship between chloroprene concentration and wind speed, we selected three monitoring sites, Mississippi River Levee, Chad Baker Street, and Fifth Ward Elementary School, which are within 1 km to the facility. Concentrations which are not availabe or below detective level (0.05 µg/m^3) are removed. Some of the summary information for the full dataset, the gathered dataset, and dataset containing sites within 1 km is listed below.

```
# Select the three sites within 1 km to the facility
data.near <- select(data.all, 'Date.noon', 'Wind.Speed', 'Chad.Baker',
                    'Elementary.School', 'Levee')
data.near.gather <- gather(data.near, "Site.Name", "Concentration",
                     Chad.Baker:Levee) %>%
  na.exclude() %>%
  filter(Concentration > 0.05)

data.near.gather <- mutate(data.near.gather, Log.concentration
                           = log(Concentration))
# Data summary
dim(data.all)
```

```
## [1] 311  13
```

```
head(data.all)
```

```
##    Date.noon Temperature Wind.Speed Wind.direction.avg Chad.Baker  Hwy44
## 1 2016-05-31    77.88186   1.634399           198.1150      7.581 30.322
## 2 2016-06-02    77.94377   2.342937           171.8351      7.145  0.073
## 3 2016-06-05    75.08600   2.958148           126.6919     11.099  0.018
## 4 2016-06-09    83.46604   2.814583           119.2105      5.477  0.624
## 5 2016-06-12    80.20563   2.186588           209.3092      5.368  0.983
## 6 2016-06-15    80.85313   2.749930           230.4162      1.211  0.225
##   Highschool Elementary.School  Levee Ochsner.Hospital Year Month Week
## 1      2.017             3.072  6.130           17.482 2016     5   22
## 2      2.666             1.882  2.637            0.065 2016     6   22
## 3      0.341             4.969 20.493            0.809 2016     6   23
## 4      1.251             3.409  4.824            4.679 2016     6   23
## 5      5.441             0.573  0.272            1.277 2016     6   24
## 6      1.030             1.745  0.366           10.809 2016     6   24
```

```
colnames(data.all)
```

```
##  [1] "Date.noon"          "Temperature"        "Wind.Speed"
##  [4] "Wind.direction.avg" "Chad.Baker"         "Hwy44"
##  [7] "Highschool"         "Elementary.School"  "Levee"
## [10] "Ochsner.Hospital"   "Year"               "Month"
## [13] "Week"
```

```r
summary(data.all)
```

```
##    Date.noon           Temperature     Wind.Speed     Wind.direction.avg
##  Min.   :2016-05-31   Min.   :31.38   Min.   :0.5035   Min.   :  0.225
##  1st Qu.:2017-01-30   1st Qu.:64.62   1st Qu.:2.2679   1st Qu.: 29.944
##  Median :2017-09-20   Median :73.96   Median :2.8115   Median :117.248
##  Mean   :2017-09-18   Mean   :70.73   Mean   :3.1294   Mean   :143.232
##  3rd Qu.:2018-05-11   3rd Qu.:80.10   3rd Qu.:3.7453   3rd Qu.:228.687
##  Max.   :2018-12-30   Max.   :86.21   Max.   :9.8444   Max.   :359.547
##
##    Chad.Baker          Hwy44            Highschool        Elementary.School
##  Min.   : 0.000   Min.   :  0.018   Min.   : 0.0180   Min.   :  0.018
##  1st Qu.: 0.037   1st Qu.:  0.023   1st Qu.: 0.0370   1st Qu.:  0.037
##  Median : 0.925   Median :  0.076   Median : 0.1581   Median :  0.823
##  Mean   : 6.392   Mean   :  3.012   Mean   : 1.4073   Mean   :  5.335
##  3rd Qu.: 6.918   3rd Qu.:  1.117   3rd Qu.: 1.0410   3rd Qu.:  4.135
##  Max.   :70.002   Max.   :153.424   Max.   :39.5350   Max.   :149.616
##  NA's   :3        NA's   :4         NA's   :2         NA's   :6
##      Levee          Ochsner.Hospital       Year          Month
##  Min.   :  0.0180   Min.   : 0.0180   Min.   :2016   Min.   : 1.000
##  1st Qu.:  0.0370   1st Qu.: 0.0234   1st Qu.:2017   1st Qu.: 4.000
##  Median :  0.7325   Median : 0.1270   Median :2017   Median : 7.000
##  Mean   :  4.7568   Mean   : 2.8041   Mean   :2017   Mean   : 7.064
##  3rd Qu.:  3.4522   3rd Qu.: 1.1690   3rd Qu.:2018   3rd Qu.:10.000
##  Max.   :146.8950   Max.   :89.2250   Max.   :2018   Max.   :12.000
##  NA's   :7          NA's   :3
##       Week
##  Min.   : 1.00
##  1st Qu.:17.00
##  Median :30.00
##  Mean   :28.94
##  3rd Qu.:41.00
##  Max.   :53.00
##
```

```r
dim(data.gather)
```

```
## [1] 1158    9
```

```r
colnames(data.gather)
```

```
## [1] "Date.noon"          "Temperature"        "Wind.Speed"
## [4] "Wind.direction.avg" "Year"               "Month"
## [7] "Week"               "Site.Name"          "Concentration"
```

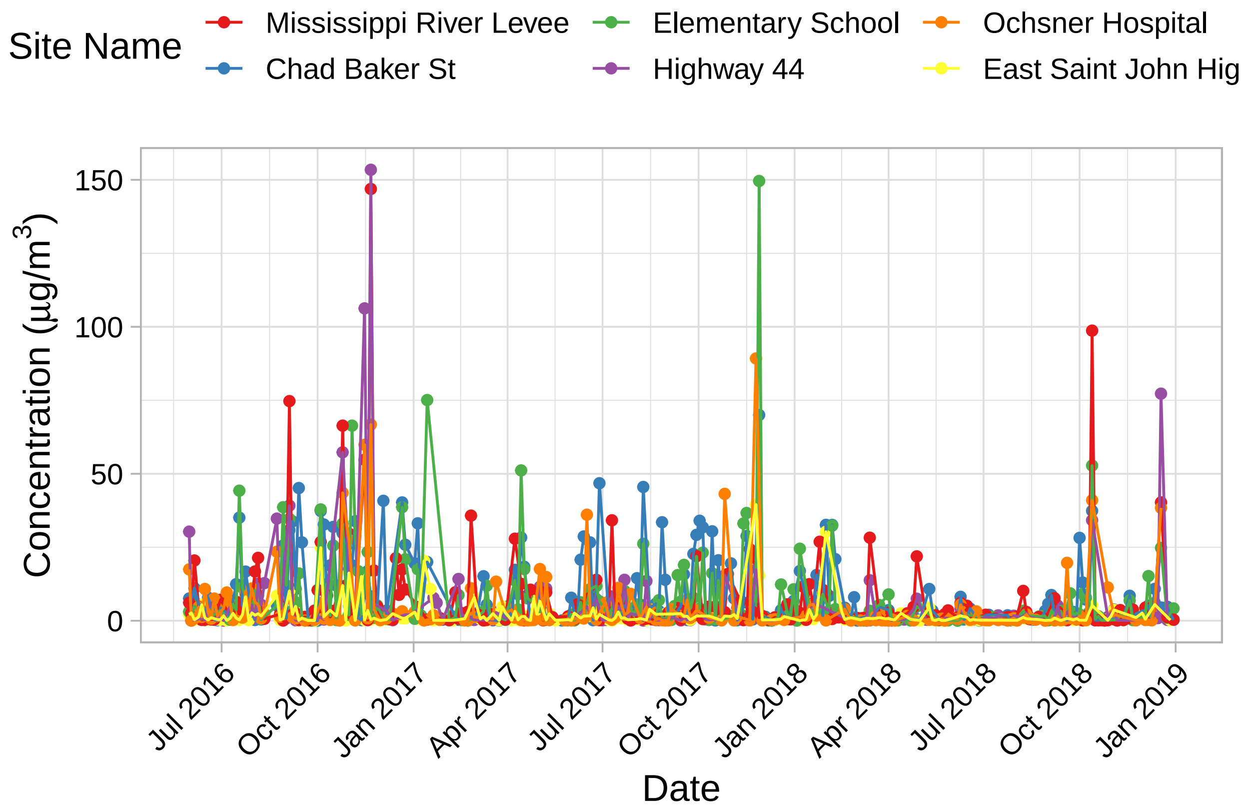

Figure 1: Chloroprene concentrations at six monitoring sites over time

```r
dim(data.near.gather)
```

```
## [1] 642   5
```

```r
colnames(data.near.gather)
```

```
## [1] "Date.noon"          "Wind.Speed"         "Site.Name"
## [4] "Concentration"      "Log.concentration"
```

Figure 1 shows the chloroprene concentrations at the six monitoring sites from May 2016 to December 2018. The distributions of original and log-transformed chloroprene concentrations at the three closer monitoring sites are shown in Figure 2 and 3. As we can see from the figures, the log-transformed concentrations are more normally distributed. Therefore, we use the log-transformed concentration in the following analysis.

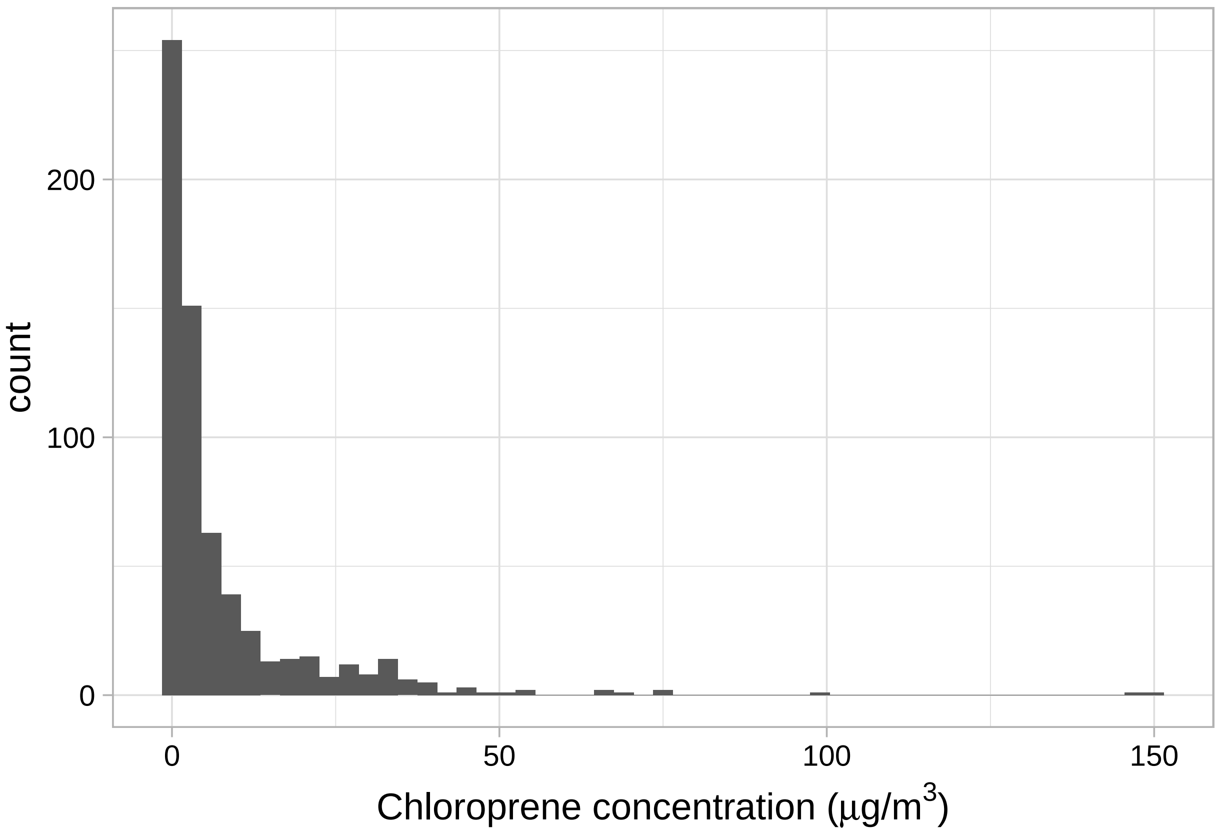

Figure 2: Distribution of chloroprene concentration

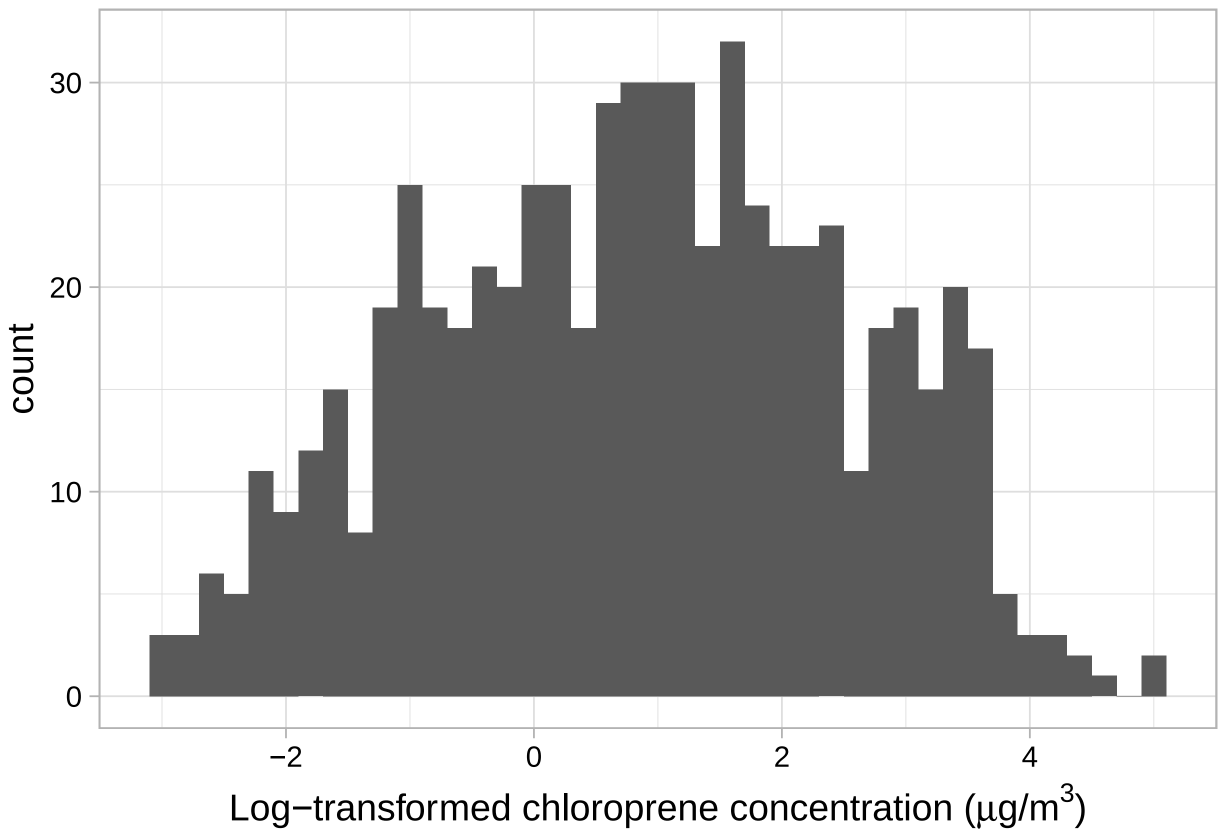

Figure 3: Distribution of log-transformed chloroprene concentration

4 Analysis

We begin with testing the normality of the chloroprene concentrations.

```
# Distribution test
# Normal distribution
shapiro.test(data.near.gather$Concentration)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data.near.gather$Concentration
## W = 0.54291, p-value < 2.2e-16
```

```
shapiro.test(data.near.gather$Log.concentration)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data.near.gather$Log.concentration
## W = 0.98788, p-value = 3.678e-05
```

Even though the results of Shapiro test show that neither of the original or log-transformed chloroprene concentrations is normally distributed, considering the nature of this dataset and the distribution figures above, we performed the generalized linear model using the log-transformed data. And the result shows that there is a statistically significant correlation between wind speed and chloroprene concentrtion (Generalized linear model; coefficient = -0.303, t-value = -4.940, $p < 0.0001$). The correlation is shown in Figure 4.

```
speed.glm <- glm(data = data.near.gather, Log.concentration ~ Wind.Speed)
summary(speed.glm)
```

```
##
## Call:
## glm(formula = Log.concentration ~ Wind.Speed, data = data.near.gather)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.0957  -1.1894  -0.0421   1.2365   4.0764
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.71967    0.18806   9.144  < 2e-16 ***
## Wind.Speed  -0.30294    0.06132  -4.940 9.96e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

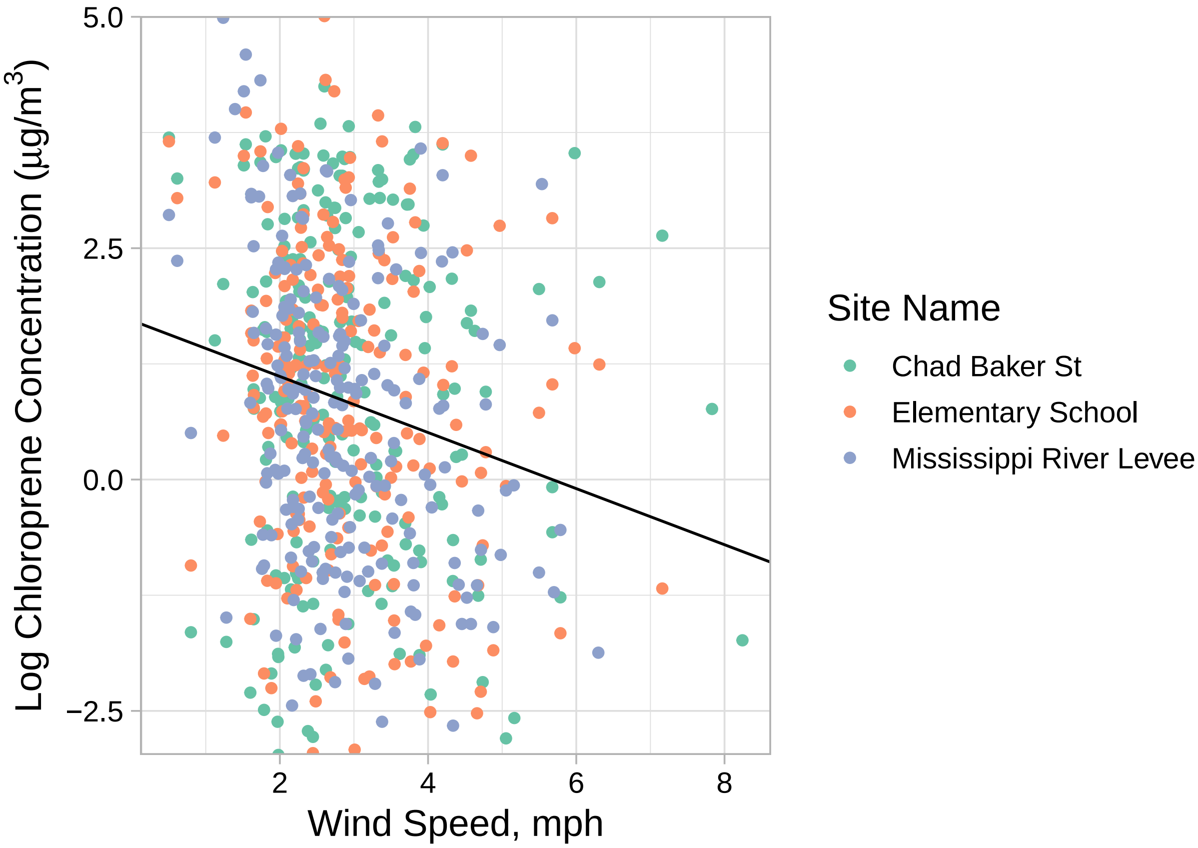

Figure 4: Chloroprene concentration and wind speed

```
## (Dispersion parameter for gaussian family taken to be 2.749833)
##
##     Null deviance: 1827.0  on 641  degrees of freedom
## Residual deviance: 1759.9  on 640  degrees of freedom
## AIC: 2475.3
##
## Number of Fisher Scoring iterations: 2
```

Another objective of this project is to investigate the change of chloroprene concentrations over time. As shown in Table 2, there is an obvious drop in chloroprene concentrations from 2016 to 2018. The result of ANOVA test also suggests that the concentrations are statistically significant different from each other in 2016, 2017, and 2018 (ANOVA test; F-statistic = 19.76, df = 1155, p-value < 0.0001).

Pettitt's test allows us to find the changing point in our data. According to Denka, the emission reduction projects were implemented by the company to reduce chloroprene emissions around the beginning in 2018. As the results shown below, the changing points occurred on 2017-11-13 at Chad Baker St, on 2018-01-15 at Fifth Ward Elementary School, and on 2018-01-09 at Missipi River Levee, which agree with the statement of the company.

```r
# Test for change over time
Year.mean <- data.gather %>%
  filter(!is.na(Concentration)) %>%
  group_by(Year) %>%
  summarize("Mean" = mean(Concentration))
kable(Year.mean, caption = "Annual average of chloroprene concentration")
```

Table 2: Annual average of chloroprene concentration

Year	Mean
2016	9.986195
2017	6.154984
2018	3.659313

```r
air.lm <- aov(data = data.gather,  Concentration ~ as.factor(Year))
summary(air.lm)
```

```
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## as.factor(Year)    2   7012    3506   19.76 3.65e-09 ***
## Residuals       1155 204943     177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Test for change point
# Remove NAs
data.clean <- na.exclude(data.all)
# Pettitt test
pettitt.test(data.clean$Chad.Baker)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  data.clean$Chad.Baker
## U* = 6044, p-value = 0.0004784
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                              174
```

```r
pettitt.test(data.clean$Elementary.School)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  data.clean$Elementary.School
```

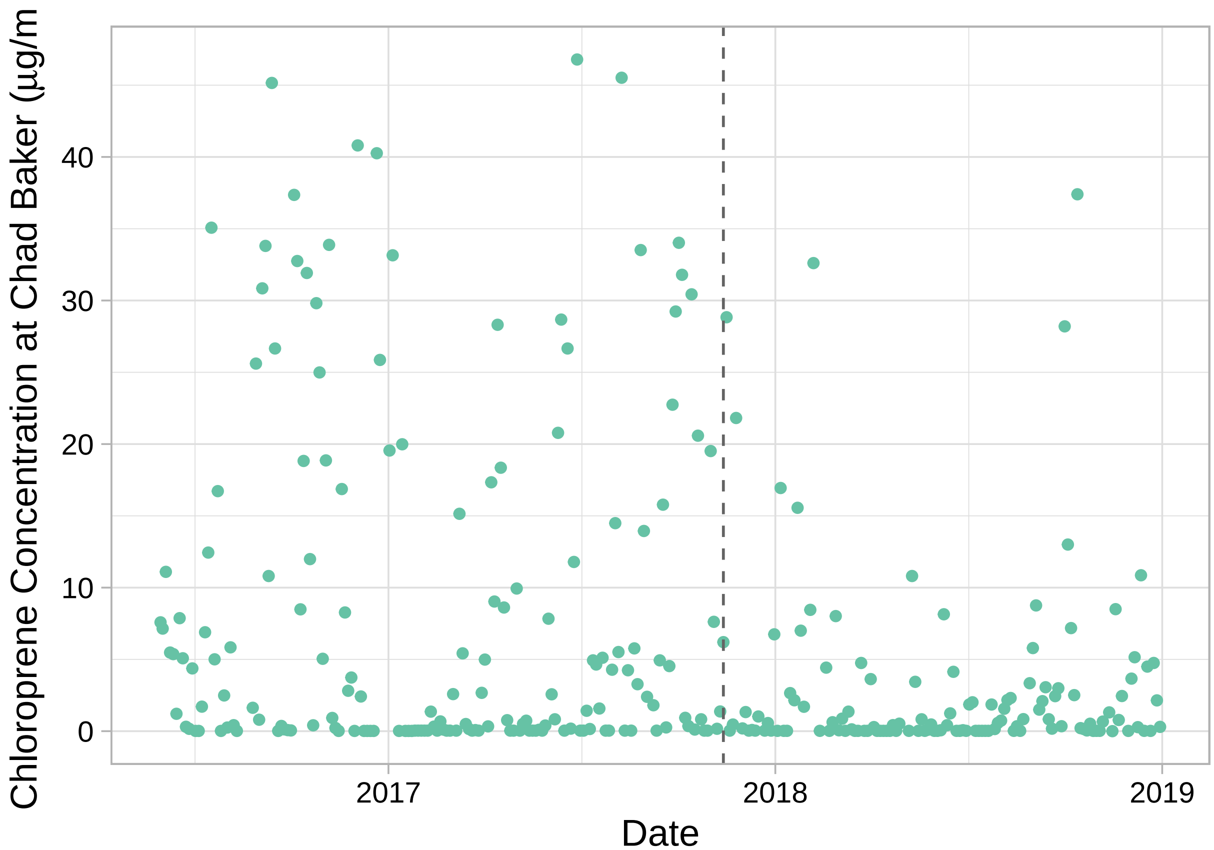

Figure 5: Chloroprene change point at Chade Baker St

```
## U* = 5878, p-value = 0.0007516
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                               195
```

```
pettitt.test(data.clean$Levee)
```

```
##
##  Pettitt's test for single change-point detection
##
## data:  data.clean$Levee
## U* = 4782, p-value = 0.01082
## alternative hypothesis: two.sided
## sample estimates:
## probable change point at time K
##                               193
```

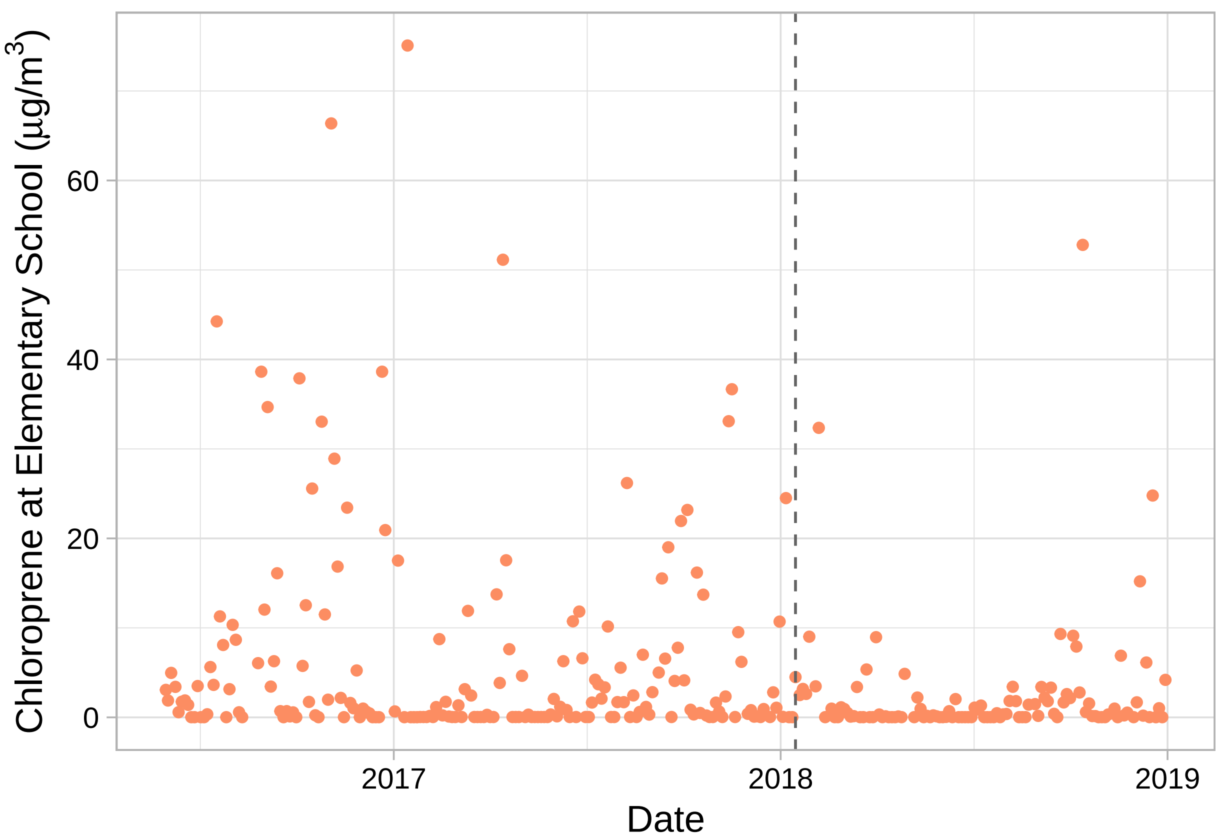

Figure 6: Chloroprene change point at Elementary School

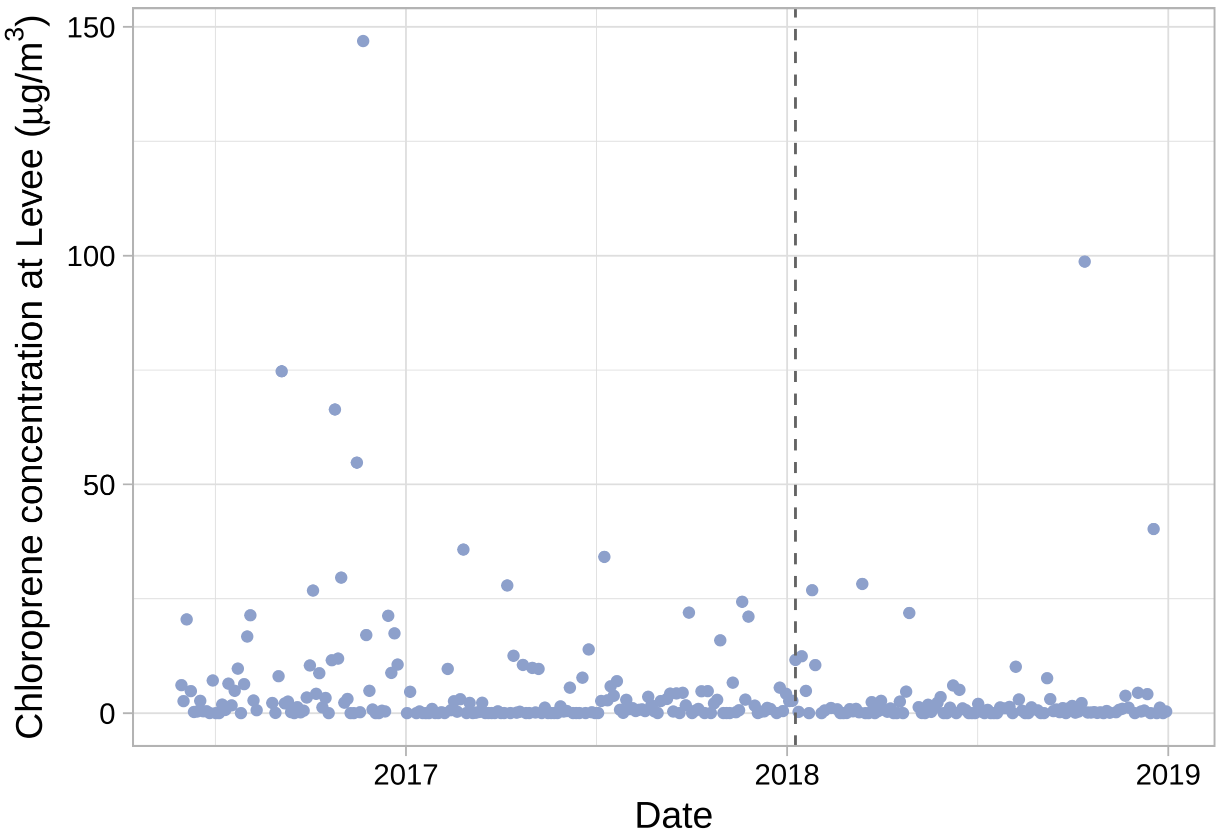

Figure 7: Chloroprene change point at Mississippi River Levee

5 Summary and Conclusions

The results show that there is a statistically significant negative correlation between chloroprene concentration and wind speed at the three monitoring sites within 1 km to the Denka facility. Residents living close to the Denka facility are facing high risk of developing cancer because of potential explosure of chloroprene. Meteorology factors may play an essential part in the distribution of chloroprene. The results of the project suggests that wind speed affacts the concentrations of chloroprene close to the facility.

Chloroprene concentrations decline significantly from 2016 to 2018, and the changing points at the three monitoring sites within 1 km to the facility all occurred around January 2018, which is accordant with the implementation time of the emission reduction projects announced by Denka. Substantial decreases in chloropren concentraitons have been seen from 2016 to 2018. However, the current concentrations still far exceed the recommended level without increasing risk of cancer (0.2 $\mu g/m^3$). More efforts need to be made to protect public health in the LaPlace community.