

# Merge LA refinery data ses

Jerry Wu

5/1/2023

```
# load data sets
full_data <- tibble()
for (file in list.files('data')){
  full_data <- bind_rows(full_data, read.csv(paste0('data/', file)))
}

glimpse(full_data)

## Rows: 3,654,646
## Columns: 23
## $ DateTime          <chr> "10/22/2021 15:15", "10/22/2021 15:20", "10/22/20~
## $ Date.Time         <chr> "10/22/2021 15:15", "10/22/2021 15:20", "10/22/20~
## $ H2S               <dbl> 138.79, 183.26, 149.24, 263.18, 308.97, 169.06, 2~
## $ Averaging.Hour    <chr> "5 Min", "5 Min", "5 Min", "5 Min", "5 Min", "5 M~
## $ ws               <dbl> 7.01, 6.32, 6.73, 6.54, 8.02, 7.82, 6.99, 6.71, 6~
## $ wd               <dbl> 276.87, 250.15, 263.97, 280.36, 269.26, 256.21, 2~
## $ latitude          <dbl> 33.83678, 33.83678, 33.83678, 33.83678, 33.83678,~
## $ longitude         <dbl> -118.2586, -118.2586, -118.2586, -118.2586, -118.~
## $ Monitor           <chr> "213th&Chico", "213th&Chico", "213th&Chico", "213~
## $ Join_Count        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ MinDist           <dbl> 2873.053, 2873.053, 2873.053, 2873.053, 2873.053,~
## $ Converted_Angle   <int> 306, 306, 306, 306, 306, 306, 306, 306, 306, 306,~
## $ Refinery          <chr> "Marathon (Carson)", "Marathon (Carson)", "Marath~
## $ Value             <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Unit              <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Date.Time.First   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Averaging.Hour.First <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Monitor.1         <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ siteName          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ windSpeed         <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ windDirection     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Wind_latitude     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Wind_longitude    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~

summary(full_data)
```

```
##   DateTime          Date.Time          H2S          Averaging.Hour
## Length:3654646      Length:3654646      Min.    : 0.0      Length:3654646
## Class :character    Class :character    1st Qu.: 0.3      Class :character
## Mode  :character    Mode  :character    Median : 0.6      Mode  :character
##                               Mean    : 1.1
##                               3rd Qu.: 1.1
##                               Max.    :3817.2
##                               NA's    :544521
```

```
##           ws           wd           latitude           longitude
## Min.      : 0.0       Min.      : 0.0       Min.      :33.78       Min.      : -118.4
## 1st Qu.: 1.4       1st Qu.:136.8       1st Qu.:33.79       1st Qu.: -118.4
## Median : 2.8       Median :246.3       Median :33.82       Median : -118.3
## Mean    : 3.4       Mean    :212.0       Mean    :33.83       Mean    : -118.3
## 3rd Qu.: 4.9       3rd Qu.:298.0       3rd Qu.:33.87       3rd Qu.: -118.3
## Max.    :57.8       Max.    :360.0       Max.    :33.92       Max.    : -118.2
## NA's    :1387384   NA's    :1709426   NA's    :211113     NA's    :211113
## Monitor      Join_Count      MinDist      Converted_Angle
## Length:3654646   Min.      :1.000   Min.      : 758.7   Min.      : 11.0
## Class :character 1st Qu.:1.000   1st Qu.:1308.3   1st Qu.:189.0
## Mode  :character Median :1.000   Median :1988.2   Median :209.0
##                               Mean  :1.949   Mean  :1909.0   Mean  :198.5
##                               3rd Qu.:3.000   3rd Qu.:2541.9   3rd Qu.:263.0
##                               Max.   :5.000   Max.   :3592.8   Max.   :338.0
##
## Refinery      Value      Unit      Date.Time.First
## Length:3654646   Min.      : 0      Length:3654646   Length:3654646
## Class :character 1st Qu.: 0      Class :character  Class :character
## Mode  :character Median : 0      Mode  :character  Mode  :character
##                               Mean   : 1
##                               3rd Qu.: 1
##                               Max.   :36
##                               NA's   :3321238
## Averaging.Hour.First Monitor.1      siteName      windSpeed
## Length:3654646   Length:3654646   Length:3654646   Min.      : 0
## Class :character  Class :character  Class :character  1st Qu.: 2
## Mode  :character  Mode  :character  Mode  :character  Median : 4
##                               Mean   : 4
##                               3rd Qu.: 6
##                               Max.   :24
##                               NA's   :3420724
## windDirection   Wind_latitude   Wind_longitude
## Min.      : 0      Min.      :34      Min.      : -118
## 1st Qu.:146      1st Qu.:34      1st Qu.: -118
## Median :251      Median :34      Median : -118
## Mean    :211      Mean    :34      Mean    : -118
## 3rd Qu.:268      3rd Qu.:34      3rd Qu.: -118
## Max.    :360      Max.    :34      Max.    : -118
## NA's    :3420724   NA's    :3420724   NA's    :3420724
```

```
# full_data %>%
#   ggplot(aes(x=DateTime, y=H2S, group=Monitor, color=Monitor)) +
#     geom_line() +
#     labs(title = "H2S concentration by monitor", y = 'H2S', x = 'time') +
#     theme_minimal()
```