Air Quality Data Processing

Shihab Sarker

2025-03-16

Air Quality Data

A data of ambient air quality PM2.5 measured in every hours in major cities in Bangladesh covering 2012 to 2021. Your task is to create a tidy data so that subsequent analysis can be done easily. The data is in an Excel file. One worksheet contains one city. The following are the variables in the data:

- `Date:` Date in Datetime format
- `Time:` Time of the day in 24 hours format
- `PM2.5:` Concentration of PM2.5 measured in $\mu g/m^3$
- `Temperature:` in Degree Celsius

Expected Output from Air Quality Data

The following is the expected output data after doing cleaning and necessary processing:

- `Date:` Only date part from the date time variable
- `Time:` Time of the day in 24 hours format
- `CityName:` Name of the city from the raw data (Worksheet name of the Excel file)
- `PM2.5:` Concentration of PM2.5 measured in $\mu g/m^3$
- `Temperature:` in Degree Celsius

Data Processing and Cleaning

```
# Load Packages
library(readxl)
library(dplyr)
library(tidyr)
library(lubridate)
library(hms)
library(purrr)

# Import data
city_names <- excel_sheets("data/Air Quality Data.xlsx")

dfAirQuality <- map_df(city_names, ~ {
  read_excel("data/Air Quality Data.xlsx",
             sheet = .x,
             col_types = c("date", "text", "numeric", "numeric")) %>%
    mutate(CityName = .x)
})

# Clean and process the data
dfAirQuality_cleaned <- dfAirQuality %>%
  transmute(
    Date = as.Date(Date),
    Time = parse_hm(Time),
    CityName = as.factor(CityName),
    PM2.5,
    Temperature
  ) %>%
  drop_na()

# View the cleaned data
show(dfAirQuality_cleaned)
```

```
## # A tibble: 220,206 × 5
##    Date       Time   CityName PM2.5 Temperature
##    <date>     <time> <fct>    <dbl>       <dbl>
##  1 2012-11-01 01:00  Dhaka     93.4        27.5
##  2 2012-11-01 02:00  Dhaka     73.6        27.4
##  3 2012-11-01 03:00  Dhaka     60.3        26.6
##  4 2012-11-01 04:00  Dhaka     70.6        26.1
##  5 2012-11-01 05:00  Dhaka     90.4        25.8
##  6 2012-11-01 06:00  Dhaka     69.9        25.3
##  7 2012-11-01 07:00  Dhaka     64.8        25.8
##  8 2012-11-01 08:00  Dhaka     60.1        27.6
##  9 2012-11-01 09:00  Dhaka     99.3        29.5
## 10 2012-11-01 10:00  Dhaka     72.8        22.0
## # ℹ 220,196 more rows
```