Assignment 4: Data Wrangling

Sashoy Milton

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on Data Wrangling

Set up your session

1. Check your working directory, load the `tidyverse` and `lubridate` packages, and upload all four raw data files associated with the EPA Air dataset, being sure to set string columns to be read in a factors. See the README file for the EPA air datasets for more information (especially if you have not worked with air quality data previously).

```
#Load packages

library(tidyverse)
library(lubridate)
library (skimr)

#Load data

EPA.Air.NC18 <- read.csv("./EPAair_O3_NC2018_raw.csv", stringsAsFactors = TRUE)
EPA.Air.NC19 <- read.csv("./EPAair_O3_NC2019_raw.csv", stringsAsFactors = TRUE)
EPA.Air.PM25_NC18 <- read.csv("./EPAair_PM25_NC2018_raw.csv", stringsAsFactors = TRUE)
EPA.Air.PM25_NC19 <- read.csv("./EPAair_PM25_NC2019_raw.csv", stringsAsFactors = TRUE)
```

2. Explore the dimensions, column names, and structure of the datasets.

```
#1. Air Data

############# Explore dimensions ####################

dim(EPA.Air.NC18)
```

```
## [1] 9737   20
```

```
dim (EPA.Air.NC19)
```

```
## [1] 10592    20
```

```
############# Explore column names and structure of databases ###########

summary(EPA.Air.NC18)
```

```
##          Date      Source        Site.ID              POC
##  04/01/2018:  40   AQS:9737   Min.   :370030005   Min.   :1
##  04/12/2018:  40              1st Qu.:370650099   1st Qu.:1
##  04/13/2018:  40              Median :371010002   Median :1
##  04/14/2018:  40              Mean   :370969118   Mean   :1
##  04/15/2018:  40              3rd Qu.:371290002   3rd Qu.:1
##  04/18/2018:  40              Max.   :371990004   Max.   :1
##  (Other)   :9497
##  Daily.Max.8.hour.Ozone.Concentration UNITS      DAILY_AQI_VALUE
##  Min.   :0.00200                      ppm:9737   Min.   :  2.00
##  1st Qu.:0.03400                                 1st Qu.: 31.00
##  Median :0.04200                                 Median : 39.00
##  Mean   :0.04194                                 Mean   : 40.22
##  3rd Qu.:0.04900                                 3rd Qu.: 45.00
##  Max.   :0.07700                                 Max.   :122.00
##
##                Site.Name    DAILY_OBS_COUNT PERCENT_COMPLETE
##  Coweeta             : 355   Min.   :12.00   Min.   : 71.00
##  Garinger High School: 354   1st Qu.:17.00   1st Qu.:100.00
##  Millbrook School    : 352   Median :17.00   Median :100.00
##  Candor              : 335   Mean   :16.94   Mean   : 99.65
##  Rockwell            : 335   3rd Qu.:17.00   3rd Qu.:100.00
##  Cranberry           : 323   Max.   :17.00   Max.   :100.00
##  (Other)             :7683
##  AQS_PARAMETER_CODE AQS_PARAMETER_DESC   CBSA_CODE
##  Min.   :44201      Ozone:9737         Min.   :11700
##  1st Qu.:44201                         1st Qu.:16740
##  Median :44201                         Median :24660
##  Mean   :44201                         Mean   :27247
##  3rd Qu.:44201                         3rd Qu.:39580
##  Max.   :44201                         Max.   :49180
##                                        NA's   :2609
##                               CBSA_NAME      STATE_CODE            STATE
##                                   :2609   Min.   :37   North Carolina:9737
##  Charlotte-Concord-Gastonia, NC-SC:1338   1st Qu.:37
##  Asheville, NC                    : 927   Median :37
##  Winston-Salem, NC                : 725   Mean   :37
##  Raleigh, NC                      : 585   3rd Qu.:37
##  Hickory-Lenoir-Morganton, NC     : 477   Max.   :37
##  (Other)                          :3076
##   COUNTY_CODE             COUNTY     SITE_LATITUDE   SITE_LONGITUDE
##  Min.   :  3.00   Forsyth    : 725   Min.   :34.36   Min.   :-83.80
##  1st Qu.: 65.00   Haywood    : 683   1st Qu.:35.26   1st Qu.:-82.05
##  Median :101.00   Mecklenburg: 592   Median :35.55   Median :-80.34
##  Mean   : 96.78   Avery      : 558   Mean   :35.62   Mean   :-80.42
##  3rd Qu.:129.00   Swain      : 483   3rd Qu.:36.03   3rd Qu.:-78.90
##  Max.   :199.00   Cumberland : 444   Max.   :36.31   Max.   :-76.62
##                   (Other)    :6252
```

```
# skim(EPA.Air.NC18)
head(EPA.Air.NC18)
```

```
##         Date Source   Site.ID POC Daily.Max.8.hour.Ozone.Concentration UNITS
## 1 03/01/2018    AQS 370030005   1                                 0.043   ppm
```

```
## 2 03/02/2018    AQS 370030005   1                                0.046   ppm
## 3 03/03/2018    AQS 370030005   1                                0.047   ppm
## 4 03/04/2018    AQS 370030005   1                                0.049   ppm
## 5 03/05/2018    AQS 370030005   1                                0.047   ppm
## 6 03/06/2018    AQS 370030005   1                                0.030   ppm
##   DAILY_AQI_VALUE            Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 1              40 Taylorsville Liledoun              17              100
## 2              43 Taylorsville Liledoun              17              100
## 3              44 Taylorsville Liledoun              17              100
## 4              45 Taylorsville Liledoun              17              100
## 5              44 Taylorsville Liledoun              17              100
## 6              28 Taylorsville Liledoun              17              100
##   AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE                  CBSA_NAME
## 1              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 2              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 3              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 4              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 5              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 6              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
##   STATE_CODE          STATE COUNTY_CODE    COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1         37 North Carolina           3 Alexander       35.9138        -81.191
## 2         37 North Carolina           3 Alexander       35.9138        -81.191
## 3         37 North Carolina           3 Alexander       35.9138        -81.191
## 4         37 North Carolina           3 Alexander       35.9138        -81.191
## 5         37 North Carolina           3 Alexander       35.9138        -81.191
## 6         37 North Carolina           3 Alexander       35.9138        -81.191
```

```
tail(EPA.Air.NC18)
```

```
##            Date Source   Site.ID POC Daily.Max.8.hour.Ozone.Concentration UNITS
## 9732 10/26/2018    AQS 371990004   1                                0.043   ppm
## 9733 10/27/2018    AQS 371990004   1                                0.041   ppm
## 9734 10/28/2018    AQS 371990004   1                                0.045   ppm
## 9735 10/29/2018    AQS 371990004   1                                0.050   ppm
## 9736 10/30/2018    AQS 371990004   1                                0.053   ppm
## 9737 10/31/2018    AQS 371990004   1                                0.048   ppm
##      DAILY_AQI_VALUE    Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 9732              40 Mt. Mitchell              17              100
## 9733              38 Mt. Mitchell              17              100
## 9734              42 Mt. Mitchell              17              100
## 9735              46 Mt. Mitchell              17              100
## 9736              49 Mt. Mitchell              17              100
## 9737              44 Mt. Mitchell              17              100
##      AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME STATE_CODE
## 9732              44201              Ozone        NA                  37
## 9733              44201              Ozone        NA                  37
## 9734              44201              Ozone        NA                  37
## 9735              44201              Ozone        NA                  37
## 9736              44201              Ozone        NA                  37
## 9737              44201              Ozone        NA                  37
##               STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 9732 North Carolina         199 Yancey      35.76541      -82.26494
## 9733 North Carolina         199 Yancey      35.76541      -82.26494
## 9734 North Carolina         199 Yancey      35.76541      -82.26494
```

```
## 9735 North Carolina         199 Yancey      35.76541      -82.26494
## 9736 North Carolina         199 Yancey      35.76541      -82.26494
## 9737 North Carolina         199 Yancey      35.76541      -82.26494
```

```
########################

summary(EPA.Air.NC19)
```

```
##          Date          Source        Site.ID               POC
##  03/18/2019:   38   AirNow:2126   Min.   :370030005   Min.   :1
##  03/19/2019:   38   AQS   :8466   1st Qu.:370630015   1st Qu.:1
##  03/20/2019:   38                 Median :370870036   Median :1
##  03/23/2019:   38                 Mean   :370960317   Mean   :1
##  03/24/2019:   38                 3rd Qu.:371290002   3rd Qu.:1
##  03/25/2019:   38                 Max.   :371990004   Max.   :1
##  (Other)   :10364
##  Daily.Max.8.hour.Ozone.Concentration UNITS       DAILY_AQI_VALUE
##  Min.   :0.00000                      ppm:10592   Min.   :  0.0
##  1st Qu.:0.03600                                  1st Qu.: 33.0
##  Median :0.04400                                  Median : 41.0
##  Mean   :0.04331                                  Mean   : 41.2
##  3rd Qu.:0.05000                                  3rd Qu.: 46.0
##  Max.   :0.08100                                  Max.   :136.0
##
##                 Site.Name    DAILY_OBS_COUNT PERCENT_COMPLETE
##  Garinger High School: 363   Min.   :13.00   Min.   : 75.00
##  Millbrook School    : 362   1st Qu.:17.00   1st Qu.:100.00
##  Coweeta             : 361   Median :17.00   Median :100.00
##  Rockwell            : 361   Mean   :18.34   Mean   : 99.69
##  Candor              : 358   3rd Qu.:17.00   3rd Qu.:100.00
##  Cranberry           : 351   Max.   :24.00   Max.   :100.00
##  (Other)             :8436
##  AQS_PARAMETER_CODE AQS_PARAMETER_DESC   CBSA_CODE
##  Min.   :44201      Ozone:10592        Min.   :11700
##  1st Qu.:44201                         1st Qu.:16740
##  Median :44201                         Median :24660
##  Mean   :44201                         Mean   :26617
##  3rd Qu.:44201                         3rd Qu.:37080
##  Max.   :44201                         Max.   :49180
##                                        NA's   :2852
##                              CBSA_NAME      STATE_CODE            STATE
##                                   :2852   Min.   :37   North Carolina:10592
##  Charlotte-Concord-Gastonia, NC-SC:1590   1st Qu.:37
##  Asheville, NC                    :1114   Median :37
##  Winston-Salem, NC                : 735   Mean   :37
##  Raleigh, NC                      : 646   3rd Qu.:37
##  Hickory-Lenoir-Morganton, NC     : 567   Max.   :37
##  (Other)                          :3088
##   COUNTY_CODE            COUNTY     SITE_LATITUDE   SITE_LONGITUDE
##  Min.   :  3.0   Haywood    : 864   Min.   :34.36   Min.   :-83.80
##  1st Qu.: 63.0   Forsyth    : 735   1st Qu.:35.26   1st Qu.:-82.05
##  Median : 87.0   Mecklenburg: 657   Median :35.59   Median :-80.34
##  Mean   : 95.9   Avery      : 607   Mean   :35.61   Mean   :-80.41
##  3rd Qu.:129.0   Cumberland : 498   3rd Qu.:36.03   3rd Qu.:-78.77
```

```
##  Max.   :199.0   Swain      : 476   Max.   :36.31   Max.   :-76.62
##                  (Other)    :6755
```

```
# skim(EPA.Air.NC19)
head(EPA.Air.NC19)
```

```
##         Date Source   Site.ID POC Daily.Max.8.hour.Ozone.Concentration UNITS
## 1 01/01/2019 AirNow 370030005   1                                 0.029   ppm
## 2 01/02/2019 AirNow 370030005   1                                 0.018   ppm
## 3 01/03/2019 AirNow 370030005   1                                 0.016   ppm
## 4 01/04/2019 AirNow 370030005   1                                 0.022   ppm
## 5 01/05/2019 AirNow 370030005   1                                 0.037   ppm
## 6 01/06/2019 AirNow 370030005   1                                 0.037   ppm
##   DAILY_AQI_VALUE            Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 1              27 Taylorsville Liledoun              24              100
## 2              17 Taylorsville Liledoun              24              100
## 3              15 Taylorsville Liledoun              24              100
## 4              20 Taylorsville Liledoun              24              100
## 5              34 Taylorsville Liledoun              24              100
## 6              34 Taylorsville Liledoun              24              100
##   AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE                    CBSA_NAME
## 1              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 2              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 3              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 4              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 5              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
## 6              44201              Ozone     25860 Hickory-Lenoir-Morganton, NC
##   STATE_CODE          STATE COUNTY_CODE    COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1         37 North Carolina           3 Alexander       35.9138        -81.191
## 2         37 North Carolina           3 Alexander       35.9138        -81.191
## 3         37 North Carolina           3 Alexander       35.9138        -81.191
## 4         37 North Carolina           3 Alexander       35.9138        -81.191
## 5         37 North Carolina           3 Alexander       35.9138        -81.191
## 6         37 North Carolina           3 Alexander       35.9138        -81.191
```

```
tail(EPA.Air.NC19)
```

```
##             Date Source   Site.ID POC Daily.Max.8.hour.Ozone.Concentration
## 10587 10/26/2019 AirNow 371990004   1                                 0.042
## 10588 10/27/2019 AirNow 371990004   1                                 0.059
## 10589 10/28/2019 AirNow 371990004   1                                 0.060
## 10590 10/29/2019 AirNow 371990004   1                                 0.065
## 10591 10/30/2019 AirNow 371990004   1                                 0.033
## 10592 10/31/2019 AirNow 371990004   1                                 0.032
##       UNITS DAILY_AQI_VALUE    Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 10587   ppm              39 Mt. Mitchell              24               100
## 10588   ppm              64 Mt. Mitchell              24               100
## 10589   ppm              67 Mt. Mitchell              24               100
## 10590   ppm              84 Mt. Mitchell              24               100
## 10591   ppm              31 Mt. Mitchell              24               100
## 10592   ppm              30 Mt. Mitchell              22                92
##       AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME STATE_CODE
## 10587              44201              Ozone        NA                   37
```

```
## 10588              44201              Ozone        NA                   37
## 10589              44201              Ozone        NA                   37
## 10590              44201              Ozone        NA                   37
## 10591              44201              Ozone        NA                   37
## 10592              44201              Ozone        NA                   37
##                STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 10587 North Carolina         199 Yancey      35.76541      -82.26494
## 10588 North Carolina         199 Yancey      35.76541      -82.26494
## 10589 North Carolina         199 Yancey      35.76541      -82.26494
## 10590 North Carolina         199 Yancey      35.76541      -82.26494
## 10591 North Carolina         199 Yancey      35.76541      -82.26494
## 10592 North Carolina         199 Yancey      35.76541      -82.26494
```

```
#2. PM 2.5 Data

############## Explore dimensions ######################

dim(EPA.Air.PM25_NC18)
```

```
## [1] 8983   20
```

```
dim (EPA.Air.PM25_NC19)
```

```
## [1] 8581   20
```

```
############## Explore column names and structure of databases ############

summary(EPA.Air.PM25_NC18)
```

```
##          Date      Source        Site.ID               POC
##  01/26/2018:  40   AQS:8983   Min.   :370110002   Min.   :1.000
##  02/01/2018:  40              1st Qu.:370630015   1st Qu.:3.000
##  02/19/2018:  40              Median :371010002   Median :3.000
##  03/21/2018:  40              Mean   :371002405   Mean   :2.812
##  04/02/2018:  40              3rd Qu.:371230001   3rd Qu.:3.000
##  04/08/2018:  40              Max.   :371830021   Max.   :5.000
##  (Other)   :8743
##  Daily.Mean.PM2.5.Concentration      UNITS      DAILY_AQI_VALUE
##  Min.   :-2.300                 ug/m3 LC:8983   Min.   : 0.00
##  1st Qu.: 4.900                                 1st Qu.:20.00
##  Median : 7.000                                 Median :29.00
##  Mean   : 7.491                                 Mean   :30.73
##  3rd Qu.: 9.700                                 3rd Qu.:40.00
##  Max.   :34.200                                 Max.   :97.00
##
##                 Site.Name    DAILY_OBS_COUNT PERCENT_COMPLETE
##  Millbrook School    : 717   Min.   :1       Min.   :100
##  Hattie Avenue       : 510   1st Qu.:1       1st Qu.:100
##  Board Of Ed. Bldg.  : 477   Median :1       Median :100
##  Garinger High School: 472   Mean   :1       Mean   :100
##  Durham Armory       : 466   3rd Qu.:1       3rd Qu.:100
##  Pitt Agri. Center   : 460   Max.   :1       Max.   :100
```

```
##  (Other)             :5881
##  AQS_PARAMETER_CODE                              AQS_PARAMETER_DESC
##  Min.   :88101      Acceptable PM2.5 AQI & Speciation Mass:1403
##  1st Qu.:88101      PM2.5 - Local Conditions              :7580
##  Median :88101
##  Mean   :88164
##  3rd Qu.:88101
##  Max.   :88502
##
##    CBSA_CODE                                  CBSA_NAME      STATE_CODE
##  Min.   :11700   Raleigh, NC                       :1396   Min.   :37
##  1st Qu.:19000   Winston-Salem, NC                 :1316   1st Qu.:37
##  Median :25860   Charlotte-Concord-Gastonia, NC-SC:1275   Median :37
##  Mean   :30946                                     :1263   Mean   :37
##  3rd Qu.:40580   Asheville, NC                     : 586   3rd Qu.:37
##  Max.   :49180   Durham-Chapel Hill, NC            : 466   Max.   :37
##  NA's   :1263    (Other)                           :2681
##             STATE       COUNTY_CODE            COUNTY     SITE_LATITUDE
##  North Carolina:8983   Min.   : 11.0   Mecklenburg:1275   Min.   :34.36
##                        1st Qu.: 63.0   Wake       :1049   1st Qu.:35.26
##                        Median :101.0   Forsyth    : 876   Median :35.64
##                        Mean   :100.2   Buncombe   : 477   Mean   :35.61
##                        3rd Qu.:123.0   Durham     : 466   3rd Qu.:35.91
##                        Max.   :183.0   Pitt       : 460   Max.   :36.11
##                                        (Other)    :4380
##  SITE_LONGITUDE
##  Min.   :-83.44
##  1st Qu.:-80.87
##  Median :-80.23
##  Mean   :-79.99
##  3rd Qu.:-78.57
##  Max.   :-76.21
##
```

```
# skim(EPA.Air.PM25_NC18)
head(EPA.Air.PM25_NC18)
```

```
##         Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    UNITS
## 1 01/02/2018    AQS 370110002   1                             2.9 ug/m3 LC
## 2 01/05/2018    AQS 370110002   1                             3.7 ug/m3 LC
## 3 01/08/2018    AQS 370110002   1                             5.3 ug/m3 LC
## 4 01/11/2018    AQS 370110002   1                             0.8 ug/m3 LC
## 5 01/14/2018    AQS 370110002   1                             2.5 ug/m3 LC
## 6 01/17/2018    AQS 370110002   1                             4.5 ug/m3 LC
##   DAILY_AQI_VALUE      Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 1              12 Linville Falls               1              100
## 2              15 Linville Falls               1              100
## 3              22 Linville Falls               1              100
## 4               3 Linville Falls               1              100
## 5              10 Linville Falls               1              100
## 6              19 Linville Falls               1              100
##   AQS_PARAMETER_CODE                     AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME
## 1              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 2              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
```

```
## 3              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 4              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 5              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 6              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
##   STATE_CODE          STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1         37 North Carolina          11  Avery      35.97235      -81.93307
## 2         37 North Carolina          11  Avery      35.97235      -81.93307
## 3         37 North Carolina          11  Avery      35.97235      -81.93307
## 4         37 North Carolina          11  Avery      35.97235      -81.93307
## 5         37 North Carolina          11  Avery      35.97235      -81.93307
## 6         37 North Carolina          11  Avery      35.97235      -81.93307
```

```
tail(EPA.Air.PM25_NC18)
```

```
##            Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    UNITS
## 8978 12/26/2018    AQS 371830021   3                            8.7 ug/m3 LC
## 8979 12/27/2018    AQS 371830021   3                            7.4 ug/m3 LC
## 8980 12/28/2018    AQS 371830021   3                            3.5 ug/m3 LC
## 8981 12/29/2018    AQS 371830021   3                            3.0 ug/m3 LC
## 8982 12/30/2018    AQS 371830021   3                            6.5 ug/m3 LC
## 8983 12/31/2018    AQS 371830021   3                            8.9 ug/m3 LC
##      DAILY_AQI_VALUE  Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 8978              36 Triple Oak               1              100
## 8979              31 Triple Oak               1              100
## 8980              15 Triple Oak               1              100
## 8981              13 Triple Oak               1              100
## 8982              27 Triple Oak               1              100
## 8983              37 Triple Oak               1              100
##      AQS_PARAMETER_CODE       AQS_PARAMETER_DESC CBSA_CODE   CBSA_NAME
## 8978              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8979              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8980              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8981              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8982              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8983              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
##      STATE_CODE          STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 8978         37 North Carolina         183   Wake       35.8652       -78.8197
## 8979         37 North Carolina         183   Wake       35.8652       -78.8197
## 8980         37 North Carolina         183   Wake       35.8652       -78.8197
## 8981         37 North Carolina         183   Wake       35.8652       -78.8197
## 8982         37 North Carolina         183   Wake       35.8652       -78.8197
## 8983         37 North Carolina         183   Wake       35.8652       -78.8197
```

```
#########################

summary(EPA.Air.PM25_NC19)
```

```
##          Date         Source        Site.ID              POC
##  02/26/2019:  41   AirNow:1670   Min.   :370110002   Min.   :1.000
##  01/21/2019:  40   AQS   :6911   1st Qu.:370630015   1st Qu.:3.000
##  02/14/2019:  40                 Median :371190041   Median :3.000
##  01/09/2019:  39                 Mean   :371023743   Mean   :3.032
##  01/27/2019:  39                 3rd Qu.:371290002   3rd Qu.:3.000
```

```
##  02/02/2019:  39                  Max.   :371830021   Max.   :5.000
##  (Other)   :8343
##  Daily.Mean.PM2.5.Concentration      UNITS      DAILY_AQI_VALUE
##  Min.   :-3.100                 ug/m3 LC:8581   Min.   : 0.00
##  1st Qu.: 4.900                                 1st Qu.:20.00
##  Median : 7.400                                 Median :31.00
##  Mean   : 7.684                                 Mean   :31.51
##  3rd Qu.:10.100                                 3rd Qu.:42.00
##  Max.   :31.200                                 Max.   :91.00
##
##                 Site.Name    DAILY_OBS_COUNT PERCENT_COMPLETE
##  Millbrook School    : 738   Min.   :1       Min.   :100
##  Garinger High School: 629   1st Qu.:1       1st Qu.:100
##  Remount             : 573   Median :1       Median :100
##  Hickory Water Tower : 518   Mean   :1       Mean   :100
##  Hattie Avenue       : 436   3rd Qu.:1       3rd Qu.:100
##  Durham Armory       : 431   Max.   :1       Max.   :100
##  (Other)             :5256
##  AQS_PARAMETER_CODE                              AQS_PARAMETER_DESC
##  Min.   :88101      Acceptable PM2.5 AQI & Speciation Mass:1029
##  1st Qu.:88101      PM2.5 - Local Conditions              :7552
##  Median :88101
##  Mean   :88149
##  3rd Qu.:88101
##  Max.   :88502
##
##    CBSA_CODE                                  CBSA_NAME      STATE_CODE
##  Min.   :11700   Raleigh, NC                       :1441   Min.   :37
##  1st Qu.:19000   Charlotte-Concord-Gastonia, NC-SC:1379   1st Qu.:37
##  Median :25860   Winston-Salem, NC                 :1235   Median :37
##  Mean   :31099                                     :1058   Mean   :37
##  3rd Qu.:40580   Hickory-Lenoir-Morganton, NC      : 518   3rd Qu.:37
##  Max.   :49180   Durham-Chapel Hill, NC            : 431   Max.   :37
##  NA's   :1058    (Other)                           :2519
##              STATE       COUNTY_CODE             COUNTY     SITE_LATITUDE
##  North Carolina:8581   Min.   : 11.0   Mecklenburg:1379   Min.   :34.36
##                        1st Qu.: 63.0   Wake       :1083   1st Qu.:35.26
##                        Median :119.0   Forsyth    : 839   Median :35.73
##                        Mean   :102.4   Catawba    : 518   Mean   :35.63
##                        3rd Qu.:129.0   Durham     : 431   3rd Qu.:35.91
##                        Max.   :183.0   Cumberland : 427   Max.   :36.51
##                                        (Other)    :3904
##  SITE_LONGITUDE
##  Min.   :-83.44
##  1st Qu.:-80.87
##  Median :-80.23
##  Mean   :-79.95
##  3rd Qu.:-78.57
##  Max.   :-76.21
##
```

```
# skim(EPA.Air.PM25_NC19)
head(EPA.Air.PM25_NC19)
```

```
##         Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    UNITS
## 1 01/03/2019    AQS 370110002   1                             1.6 ug/m3 LC
## 2 01/06/2019    AQS 370110002   1                             1.0 ug/m3 LC
## 3 01/09/2019    AQS 370110002   1                             1.3 ug/m3 LC
## 4 01/12/2019    AQS 370110002   1                             6.3 ug/m3 LC
## 5 01/15/2019    AQS 370110002   1                             2.6 ug/m3 LC
## 6 01/18/2019    AQS 370110002   1                             1.2 ug/m3 LC
##   DAILY_AQI_VALUE      Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 1               7 Linville Falls               1              100
## 2               4 Linville Falls               1              100
## 3               5 Linville Falls               1              100
## 4              26 Linville Falls               1              100
## 5              11 Linville Falls               1              100
## 6               5 Linville Falls               1              100
##   AQS_PARAMETER_CODE                     AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME
## 1              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 2              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 3              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 4              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 5              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
## 6              88502 Acceptable PM2.5 AQI & Speciation Mass        NA
##   STATE_CODE          STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1         37 North Carolina          11  Avery      35.97235      -81.93307
## 2         37 North Carolina          11  Avery      35.97235      -81.93307
## 3         37 North Carolina          11  Avery      35.97235      -81.93307
## 4         37 North Carolina          11  Avery      35.97235      -81.93307
## 5         37 North Carolina          11  Avery      35.97235      -81.93307
## 6         37 North Carolina          11  Avery      35.97235      -81.93307
```

```
tail(EPA.Air.PM25_NC19)
```

```
##            Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration    UNITS
## 8576 12/26/2019 AirNow 371830021   3                             9.2 ug/m3 LC
## 8577 12/27/2019 AirNow 371830021   3                            11.5 ug/m3 LC
## 8578 12/28/2019 AirNow 371830021   3                             9.9 ug/m3 LC
## 8579 12/29/2019 AirNow 371830021   3                             6.5 ug/m3 LC
## 8580 12/30/2019 AirNow 371830021   3                             3.6 ug/m3 LC
## 8581 12/31/2019 AirNow 371830021   3                             4.3 ug/m3 LC
##      DAILY_AQI_VALUE  Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 8576              38 Triple Oak               1              100
## 8577              48 Triple Oak               1              100
## 8578              41 Triple Oak               1              100
## 8579              27 Triple Oak               1              100
## 8580              15 Triple Oak               1              100
## 8581              18 Triple Oak               1              100
##      AQS_PARAMETER_CODE       AQS_PARAMETER_DESC CBSA_CODE   CBSA_NAME
## 8576              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8577              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8578              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8579              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8580              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
## 8581              88101 PM2.5 - Local Conditions     39580 Raleigh, NC
##      STATE_CODE          STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 8576         37 North Carolina         183   Wake       35.8652       -78.8197
```

```
## 8577         37 North Carolina         183   Wake       35.8652       -78.8197
## 8578         37 North Carolina         183   Wake       35.8652       -78.8197
## 8579         37 North Carolina         183   Wake       35.8652       -78.8197
## 8580         37 North Carolina         183   Wake       35.8652       -78.8197
## 8581         37 North Carolina         183   Wake       35.8652       -78.8197
```

Wrangle individual datasets to create processed files.

3. Change date to date
4. Select the following columns: Date, DAILY_AQI_VALUE, Site.Name, AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE
5. For the PM2.5 datasets, fill all cells in AQS_PARAMETER_DESC with "PM2.5" (all cells in this column should be identical).
6. Save all four processed datasets in the Processed folder. Use the same file names as the raw files but replace "raw" with "processed".

```
#3. Change date to date

############################ Air Data ##########################

class(EPA.Air.NC18$Date)  #Check class of variable
```

```
## [1] "factor"
```

```
EPA.Air.NC18$Date <- mdy(EPA.Air.NC18$Date)
class(EPA.Air.NC18$Date) #Check change of class
```

```
## [1] "Date"
```

```
View(EPA.Air.NC18)

#########################

class(EPA.Air.NC19$Date) #Check class of variable
```

```
## [1] "factor"
```

```
EPA.Air.NC19$Date <- mdy(EPA.Air.NC19$Date)
class(EPA.Air.NC19$Date) #Check change of class
```

```
## [1] "Date"
```

```
View(EPA.Air.NC19)

############################ PM 2.5 ############################

class(EPA.Air.PM25_NC18$Date) #Check class of variable
```

```
## [1] "factor"
```

```
EPA.Air.PM25_NC18$Date <- mdy(EPA.Air.PM25_NC18$Date)
class(EPA.Air.PM25_NC18$Date) #Check change of class
```

```
## [1] "Date"
```

```
View(EPA.Air.PM25_NC18)
#########################

class(EPA.Air.PM25_NC19$Date) #Check class of variable
```

```
## [1] "factor"
```

```
EPA.Air.PM25_NC19$Date <- mdy(EPA.Air.PM25_NC19$Date)
class(EPA.Air.PM25_NC19$Date)  #Check change of class
```

```
## [1] "Date"
```

```
View(EPA.Air.PM25_NC19)

#4. Select Columns

#Select the following columns: Date, DAILY_AQI_VALUE, Site.Name,
#AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE

########################### Air Data ##########################

processed_EPA.Air.NC18 <- EPA.Air.NC18[c("Date", "DAILY_AQI_VALUE", "Site.Name",
            "AQS_PARAMETER_DESC", "COUNTY","SITE_LATITUDE","SITE_LONGITUDE" )]


##############

processed_EPA.Air.NC19 <- EPA.Air.NC19[c("Date", "DAILY_AQI_VALUE", "Site.Name",
          "AQS_PARAMETER_DESC", "COUNTY","SITE_LATITUDE","SITE_LONGITUDE" )]

########################## PM 2.5 ###################################

processed_EPA.Air.PM25_NC18 <- EPA.Air.PM25_NC18 [c("Date", "DAILY_AQI_VALUE",
"Site.Name", "AQS_PARAMETER_DESC", "COUNTY","SITE_LATITUDE","SITE_LONGITUDE" )]

#################

processed_EPA.Air.PM25_NC19 <- EPA.Air.PM25_NC19 [c("Date", "DAILY_AQI_VALUE",
          "Site.Name", "AQS_PARAMETER_DESC", "COUNTY","SITE_LATITUDE",
          "SITE_LONGITUDE" )]

#5. Filling all cells in AQS_PARAMETER_DESC with "PM2.5"

processed_EPA.Air.PM25_NC18$AQS_PARAMETER_DESC <- "PM2.5"

processed_EPA.Air.PM25_NC19$AQS_PARAMETER_DESC <- "PM2.5"
```

```
#6. Save four processed data-sets in the Processed folder

write.csv(processed_EPA.Air.NC18,
      file = "C:/Users/sasho/Desktop/Environ Data Analytics/Env872 Workspace/EDA-Fall2022_SM/Data/Proces
      row.names = FALSE)

write.csv(processed_EPA.Air.NC18,
          file = "C:/Users/sasho/Desktop/Environ Data Analytics/Env872 Workspace/EDA-Fall2022_SM/Data/Pr
          row.names = FALSE)

write.csv(processed_EPA.Air.NC18,
          file = "C:/Users/sasho/Desktop/Environ Data Analytics/Env872 Workspace/EDA-Fall2022_SM/Data/Pr
          row.names = FALSE)

write.csv(processed_EPA.Air.NC18,
          file = "C:/Users/sasho/Desktop/Environ Data Analytics/Env872 Workspace/EDA-Fall2022_SM/Data/Pr
          row.names = FALSE)
```

Combine datasets

7. Combine the four datasets with `rbind`. Make sure your column names are identical prior to running this code.
8. Wrangle your new dataset with a pipe function (%>%) so that it fills the following conditions:

- Include all sites that the four data frames have in common: "Linville Falls", "Durham Armory", "Leggett", "Hattie Avenue", "Clemmons Middle", "Mendenhall School", "Frying Pan Mountain", "West Johnston Co.", "Garinger High School", "Castle Hayne", "Pitt Agri. Center", "Bryson City", "Millbrook School" (the function `intersect` can figure out common factor levels)
- Some sites have multiple measurements per day. Use the split-apply-combine strategy to generate daily means: group by date, site, aqs parameter, and county. Take the mean of the AQI value, latitude, and longitude.
- Add columns for "Month" and "Year" by parsing your "Date" column (hint: `lubridate` package)
- Hint: the dimensions of this dataset should be 14,752 x 9.

9. Spread your datasets such that AQI values for ozone and PM2.5 are in separate columns. Each location on a specific date should now occupy only one row.
10. Call up the dimensions of your new tidy dataset.
11. Save your processed dataset with the following file name: "EPAair_O3_PM25_NC1718_Processed.csv"

```
#7. Check that column names are the same

colnames(processed_EPA.Air.NC18)
```

```
## [1] "Date"               "DAILY_AQI_VALUE"    "Site.Name"
## [4] "AQS_PARAMETER_DESC" "COUNTY"             "SITE_LATITUDE"
## [7] "SITE_LONGITUDE"
```

```
colnames(processed_EPA.Air.NC19)
```

```
## [1] "Date"               "DAILY_AQI_VALUE"    "Site.Name"
## [4] "AQS_PARAMETER_DESC" "COUNTY"             "SITE_LATITUDE"
## [7] "SITE_LONGITUDE"
```

```
colnames(processed_EPA.Air.PM25_NC18)
```

```
## [1] "Date"               "DAILY_AQI_VALUE"    "Site.Name"
## [4] "AQS_PARAMETER_DESC" "COUNTY"             "SITE_LATITUDE"
## [7] "SITE_LONGITUDE"
```

```
colnames(processed_EPA.Air.PM25_NC19)
```

```
## [1] "Date"               "DAILY_AQI_VALUE"    "Site.Name"
## [4] "AQS_PARAMETER_DESC" "COUNTY"             "SITE_LATITUDE"
## [7] "SITE_LONGITUDE"
```

```
#Combine with rbind

combined_EPA.data <- rbind(processed_EPA.Air.NC18, processed_EPA.Air.NC19,
processed_EPA.Air.PM25_NC18, processed_EPA.Air.PM25_NC19)

#8. Using the pipe function

filtered_combined_EPA.data <- combined_EPA.data %>%
  filter(Site.Name %in% c("Linville Falls", "Durham Armory", "Leggett",
 "Hattie Avenue", "Clemmons Middle", "Mendenhall School", "Frying Pan Mountain",
                   "West Johnston Co.", "Garinger High School", "Castle Hayne",
                "Pitt Agri. Center", "Bryson City", "Millbrook School")) %>%
  group_by(AQS_PARAMETER_DESC,Date,Site.Name,COUNTY) %>%
  summarise(DAILY_AQI_VALUE = mean(DAILY_AQI_VALUE),
            SITE_LATITUDE = mean(SITE_LATITUDE),
            SITE_LONGITUDE = mean(SITE_LONGITUDE)) %>%
             mutate(Month = month(Date),
              Year = year(Date))
```

```
## `summarise()` has grouped output by 'AQS_PARAMETER_DESC', 'Date', 'Site.Name'.
## You can override using the `.groups` argument.
```

```
#9. Spreading Data sets

filtered_combined_EPA.data2 <- filtered_combined_EPA.data %>%
  pivot_wider(names_from = AQS_PARAMETER_DESC, values_from = DAILY_AQI_VALUE)

filtered_combined_EPA.data2  #Check spreading
```

```
## # A tibble: 8,976 x 9
## # Groups:   Date, Site.Name [8,976]
##    Date       Site.Name           COUNTY SITE_~1 SITE_~2 Month  Year Ozone PM2.5
```

```
##    <date>     <fct>                <fct>    <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 2018-01-01 Garinger High Scho~ Meckl~    35.2   -80.8     1  2018    32  20
##  2 2018-01-01 Millbrook School    Wake      35.9   -78.6     1  2018    34  28
##  3 2018-01-02 Garinger High Scho~ Meckl~    35.2   -80.8     1  2018    31  38
##  4 2018-01-02 Millbrook School    Wake      35.9   -78.6     1  2018    31  38.5
##  5 2018-01-03 Garinger High Scho~ Meckl~    35.2   -80.8     1  2018    21  59
##  6 2018-01-03 Millbrook School    Wake      35.9   -78.6     1  2018    30  61
##  7 2018-01-04 Garinger High Scho~ Meckl~    35.2   -80.8     1  2018    30  39
##  8 2018-01-04 Millbrook School    Wake      35.9   -78.6     1  2018    32  39
##  9 2018-01-05 Garinger High Scho~ Meckl~    35.2   -80.8     1  2018    30  32
## 10 2018-01-05 Millbrook School    Wake      35.9   -78.6     1  2018    33  32
## # ... with 8,966 more rows, and abbreviated variable names 1: SITE_LATITUDE,
## #   2: SITE_LONGITUDE
```

```
#10. Call up the dimensions of tidy dataset

dim(filtered_combined_EPA.data2)
```

```
## [1] 8976    9
```

```
#11. Save your dataset

write.csv(filtered_combined_EPA.data2,
          file = "C:/Users/sasho/Desktop/Environ Data Analytics/Env872 Workspace/EDA-Fall2022_SM/Data/P
          row.names = FALSE)
```

Generate summary tables

12. Use the split-apply-combine strategy to generate a summary data frame. Data should be grouped by site, month, and year. Generate the mean AQI values for ozone and PM2.5 for each group. Then, add a pipe to remove instances where a month and year are not available (use the function `drop_na` in your pipe).

13. Call up the dimensions of the summary dataset.

```
#12a + b

filtered_combined_EPA.data3 <- filtered_combined_EPA.data2 %>%
  group_by(Site.Name, Month, Year) %>%
  summarize (Ozone = mean(Ozone),
             PM2.5 = mean (PM2.5)) %>%
  drop_na(Ozone) %>%
  drop_na(PM2.5)
```

```
## `summarise()` has grouped output by 'Site.Name', 'Month'. You can override
## using the `.groups` argument.
```

```
filtered_combined_EPA.data3 # View data
```

```
## # A tibble: 101 x 5
## # Groups:   Site.Name, Month [74]
```

```
##    Site.Name    Month  Year Ozone PM2.5
##    <fct>        <dbl> <dbl> <dbl> <dbl>
##  1 Bryson City      3  2018  41.6  34.7
##  2 Bryson City      4  2018  44.5  28.2
##  3 Bryson City      4  2019  45.4  26.7
##  4 Bryson City      7  2019  30.4  33.6
##  5 Bryson City      9  2018  25.4  25.1
##  6 Bryson City     10  2018  31    31.3
##  7 Castle Hayne     4  2018  48.7  14.9
##  8 Castle Hayne     4  2019  45.1  14.3
##  9 Castle Hayne     5  2019  42.8  16.5
## 10 Castle Hayne     7  2018  36.5  15.5
## # ... with 91 more rows
```

```
#13.

dim (filtered_combined_EPA.data3)
```

```
## [1] 101   5
```

14. Why did we use the function `drop_na` rather than `na.omit`?

    Answer: In this case, we wanted to remove all the NA values from the dataset (which also deletes the record within which the NA was found) instead of keeping the NA values but not incorporating them into our calculations.