

Assignment 4: Data Wrangling

Sarah Ko

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics (ENV872L) on data wrangling.

Directions

1. Change “Student Name” on line 3 (above) with your name.
2. Use the lesson as a guide. It contains code that can be modified to complete the assignment.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure to **answer the questions** in this assignment document. Space for your answers is provided in this document and is indicated by the “>” character. If you need a second paragraph be sure to start the first line with “>”. You should notice that the answer is highlighted in green by RStudio.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file. You will need to have the correct software installed to do this (see Software Installation Guide) Press the **Knit** button in the RStudio scripting panel. This will save the PDF output in your Assignments folder.
6. After Knitting, please submit the completed exercise (PDF file) to the dropbox in Sakai. Please add your last name into the file name (e.g., “Salk_A04_DataWrangling.pdf”) prior to submission.

The completed exercise is due on Thursday, 7 February, 2019 before class begins.

Set up your session

1. Check your working directory, load the **tidyverse** package, and upload all four raw data files associated with the EPA Air dataset. See the README file for the EPA air datasets for more information (especially if you have not worked with air quality data previously).
2. Generate a few lines of code to get to know your datasets (basic data summaries, etc.).

```
#1

# get working directory
getwd()

## [1] "C:/Users/Sarah/Documents/Duke/Year 2/Spring 2019/Data Analytics/Environmental_Data_Analytics"
# set wd to the filepath of Environmental_Data_Analytics to use relative filepath

# Load necessary package 'tidyverse'
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.5.2

## -- Attaching packages ----- tidyverse 1.2.1 --

## v ggplot2 3.1.0      v purrr   0.3.0
## v tibble  2.0.1      v dplyr   0.7.8
## v tidyr   0.8.2      v stringr 1.3.1
## v readr   1.3.1      v forcats 0.3.0

## Warning: package 'ggplot2' was built under R version 3.5.2
## Warning: package 'tibble' was built under R version 3.5.2
## Warning: package 'tidyr' was built under R version 3.5.2
```

```
## Warning: package 'readr' was built under R version 3.5.2
## Warning: package 'purrr' was built under R version 3.5.2
## Warning: package 'dplyr' was built under R version 3.5.2
## -- Conflicts ----- tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
# upload 4 raw data files associated w the EPA Air dataset

O3_NC2017 <- read.csv("./Data/Raw/EPAair_O3_NC2017_raw.csv")
O3_NC2018 <- read.csv("./Data/Raw/EPAair_O3_NC2018_raw.csv")
PM25_NC2017 <- read.csv("./Data/Raw/EPAair_PM25_NC2017_raw.csv")
PM25_NC2018 <- read.csv("./Data/Raw/EPAair_PM25_NC2018_raw.csv")

#2

# explore O3_NC2017
dim(O3_NC2017)

## [1] 10219    20
colnames(O3_NC2017)

## [1] "Date"
## [2] "Source"
## [3] "Site.ID"
## [4] "POC"
## [5] "Daily.Max.8.hour.Ozone.Concentration"
## [6] "UNITS"
## [7] "DAILY_AQI_VALUE"
## [8] "Site.Name"
## [9] "DAILY_OBS_COUNT"
## [10] "PERCENT_COMPLETE"
## [11] "AQ5_PARAMETER_CODE"
## [12] "AQ5_PARAMETER_DESC"
## [13] "CBSA_CODE"
## [14] "CBSA_NAME"
## [15] "STATE_CODE"
## [16] "STATE"
## [17] "COUNTY_CODE"
## [18] "COUNTY"
## [19] "SITE_LATITUDE"
## [20] "SITE_LONGITUDE"
class(O3_NC2017)

## [1] "data.frame"
head(O3_NC2017)

##      Date Source   Site.ID POC Daily.Max.8.hour.Ozone.Concentration UNITS
## 1 3/1/17   AQS 370030005    1                0.041      ppm
## 2 3/2/17   AQS 370030005    1                0.046      ppm
## 3 3/3/17   AQS 370030005    1                0.046      ppm
## 4 3/4/17   AQS 370030005    1                0.046      ppm
## 5 3/5/17   AQS 370030005    1                0.046      ppm
```

```

## 6 3/6/17    AQS 370030005    1                                0.048    ppm
##    DAILY_AQI_VALUE                Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 1              38 Taylorsville Liledoun                17                100
## 2              43 Taylorsville Liledoun                17                100
## 3              43 Taylorsville Liledoun                17                100
## 4              43 Taylorsville Liledoun                17                100
## 5              43 Taylorsville Liledoun                17                100
## 6              44 Taylorsville Liledoun                17                100
##    AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE
## 1              44201                Ozone    25860
## 2              44201                Ozone    25860
## 3              44201                Ozone    25860
## 4              44201                Ozone    25860
## 5              44201                Ozone    25860
## 6              44201                Ozone    25860
##                CBSA_NAME STATE_CODE                STATE COUNTY_CODE
## 1 Hickory-Lenoir-Morganton, NC    37 North Carolina    3
## 2 Hickory-Lenoir-Morganton, NC    37 North Carolina    3
## 3 Hickory-Lenoir-Morganton, NC    37 North Carolina    3
## 4 Hickory-Lenoir-Morganton, NC    37 North Carolina    3
## 5 Hickory-Lenoir-Morganton, NC    37 North Carolina    3
## 6 Hickory-Lenoir-Morganton, NC    37 North Carolina    3
##    COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1 Alexander    35.9138    -81.191
## 2 Alexander    35.9138    -81.191
## 3 Alexander    35.9138    -81.191
## 4 Alexander    35.9138    -81.191
## 5 Alexander    35.9138    -81.191
## 6 Alexander    35.9138    -81.191

```

```
tail(03_NC2017)
```

```

##          Date Source    Site.ID POC Daily.Max.8.hour.Ozone.Concentration
## 10214 10/25/17    AQS 371990004    1                                0.038
## 10215 10/26/17    AQS 371990004    1                                0.044
## 10216 10/27/17    AQS 371990004    1                                0.044
## 10217 10/28/17    AQS 371990004    1                                0.042
## 10218 10/30/17    AQS 371990004    1                                0.047
## 10219 10/31/17    AQS 371990004    1                                0.047
##    UNITS DAILY_AQI_VALUE    Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 10214    ppm              35 Mt. Mitchell                17                100
## 10215    ppm              41 Mt. Mitchell                17                100
## 10216    ppm              41 Mt. Mitchell                17                100
## 10217    ppm              39 Mt. Mitchell                17                100
## 10218    ppm              44 Mt. Mitchell                13                 76
## 10219    ppm              44 Mt. Mitchell                17                100
##    AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME STATE_CODE
## 10214              44201                Ozone    NA                37
## 10215              44201                Ozone    NA                37
## 10216              44201                Ozone    NA                37
## 10217              44201                Ozone    NA                37
## 10218              44201                Ozone    NA                37
## 10219              44201                Ozone    NA                37
##    STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 10214 North Carolina    199 Yancey    35.76541    -82.26494

```

```
## 10215 North Carolina      199 Yancey      35.76541      -82.26494
## 10216 North Carolina      199 Yancey      35.76541      -82.26494
## 10217 North Carolina      199 Yancey      35.76541      -82.26494
## 10218 North Carolina      199 Yancey      35.76541      -82.26494
## 10219 North Carolina      199 Yancey      35.76541      -82.26494
```

```
summary(O3_NC2017$DAILY_AQI_VALUE)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      5.00   32.00   40.00   39.87   45.00  115.00
```

```
class(O3_NC2017$Date)
```

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

```
# explore O3_NC2018
```

```
dim(O3_NC2018)
```

```
## [1] 10781      20
```

```
colnames(O3_NC2018)
```

```
## [1] "Date"
## [2] "Source"
## [3] "Site.ID"
## [4] "POC"
## [5] "Daily.Max.8.hour.Ozone.Concentration"
## [6] "UNITS"
## [7] "DAILY_AQI_VALUE"
## [8] "Site.Name"
## [9] "DAILY_OBS_COUNT"
## [10] "PERCENT_COMPLETE"
## [11] "AQS_PARAMETER_CODE"
## [12] "AQS_PARAMETER_DESC"
## [13] "CBSA_CODE"
## [14] "CBSA_NAME"
## [15] "STATE_CODE"
## [16] "STATE"
## [17] "COUNTY_CODE"
## [18] "COUNTY"
## [19] "SITE_LATITUDE"
## [20] "SITE_LONGITUDE"
```

```
class(O3_NC2018)
```

```
## [1] "data.frame"
```

```
head(O3_NC2018)
```

```
##      Date Source  Site.ID POC Daily.Max.8.hour.Ozone.Concentration UNITS
## 1 2/16/18 AirNow 370030005    1                0.038      ppm
## 2 2/17/18 AirNow 370030005    1                0.033      ppm
## 3 2/18/18 AirNow 370030005    1                0.040      ppm
## 4 2/19/18 AirNow 370030005    1                0.020      ppm
## 5 2/20/18 AirNow 370030005    1                0.019      ppm
## 6 2/21/18 AirNow 370030005    1                0.021      ppm
##      DAILY_AQI_VALUE      Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 1                35 Taylorsville Liledoun                24                100
## 2                31 Taylorsville Liledoun                24                100
```

```

## 3          37 Taylorsville Liledoun          24          100
## 4          19 Taylorsville Liledoun          24          100
## 5          18 Taylorsville Liledoun          24          100
## 6          19 Taylorsville Liledoun          24          100
##  AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE
## 1          44201          Ozone          25860
## 2          44201          Ozone          25860
## 3          44201          Ozone          25860
## 4          44201          Ozone          25860
## 5          44201          Ozone          25860
## 6          44201          Ozone          25860
##          CBSA_NAME STATE_CODE          STATE COUNTY_CODE
## 1 Hickory-Lenoir-Morganton, NC          37 North Carolina          3
## 2 Hickory-Lenoir-Morganton, NC          37 North Carolina          3
## 3 Hickory-Lenoir-Morganton, NC          37 North Carolina          3
## 4 Hickory-Lenoir-Morganton, NC          37 North Carolina          3
## 5 Hickory-Lenoir-Morganton, NC          37 North Carolina          3
## 6 Hickory-Lenoir-Morganton, NC          37 North Carolina          3
##          COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1 Alexander          35.9138          -81.191
## 2 Alexander          35.9138          -81.191
## 3 Alexander          35.9138          -81.191
## 4 Alexander          35.9138          -81.191
## 5 Alexander          35.9138          -81.191
## 6 Alexander          35.9138          -81.191

```

```
tail(03_NC2018)
```

```

##          Date Source   Site.ID POC Daily.Max.8.hour.Ozone.Concentration
## 10776  11/4/18 AirNow 371990004    1                                0.043
## 10777  11/5/18 AirNow 371990004    1                                0.044
## 10778  11/6/18 AirNow 371990004    1                                0.053
## 10779  11/7/18 AirNow 371990004    1                                0.053
## 10780  11/8/18 AirNow 371990004    1                                0.039
## 10781  11/11/18 AirNow 371990004    1                                0.059
##          UNITS DAILY_AQI_VALUE   Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 10776   ppm          40 Mt. Mitchell          24          100
## 10777   ppm          41 Mt. Mitchell          24          100
## 10778   ppm          49 Mt. Mitchell          24          100
## 10779   ppm          49 Mt. Mitchell          24          100
## 10780   ppm          36 Mt. Mitchell          24          100
## 10781   ppm          64 Mt. Mitchell          24          100
##          AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME STATE_CODE
## 10776          44201          Ozone          NA          37
## 10777          44201          Ozone          NA          37
## 10778          44201          Ozone          NA          37
## 10779          44201          Ozone          NA          37
## 10780          44201          Ozone          NA          37
## 10781          44201          Ozone          NA          37
##          STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 10776 North Carolina          199 Yancey          35.76541          -82.26494
## 10777 North Carolina          199 Yancey          35.76541          -82.26494
## 10778 North Carolina          199 Yancey          35.76541          -82.26494
## 10779 North Carolina          199 Yancey          35.76541          -82.26494
## 10780 North Carolina          199 Yancey          35.76541          -82.26494

```

```
## 10781 North Carolina      199 Yancey      35.76541      -82.26494
```

```
summary(O3_NC2018$DAILY_AQI_VALUE)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.00   31.00   38.00   39.46   45.00  122.00
```

```
class(O3_NC2018$Date)
```

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

```
# explore PM25_NC2017
```

```
dim(PM25_NC2017)
```

```
## [1] 9494    20
```

```
colnames(PM25_NC2017)
```

```
## [1] "Date"                "Source"
## [3] "Site.ID"             "POC"
## [5] "Daily.Mean.PM2.5.Concentration" "UNITS"
## [7] "DAILY_AQI_VALUE"     "Site.Name"
## [9] "DAILY_OBS_COUNT"     "PERCENT_COMPLETE"
## [11] "AQS_PARAMETER_CODE"  "AQS_PARAMETER_DESC"
## [13] "CBSA_CODE"           "CBSA_NAME"
## [15] "STATE_CODE"          "STATE"
## [17] "COUNTY_CODE"        "COUNTY"
## [19] "SITE_LATITUDE"       "SITE_LONGITUDE"
```

```
class(PM25_NC2017)
```

```
## [1] "data.frame"
```

```
head(PM25_NC2017)
```

```
##      Date Source  Site.ID POC Daily.Mean.PM2.5.Concentration  UNITS
## 1  1/1/17   AQS 370110002   1                2.9 ug/m3 LC
## 2  1/4/17   AQS 370110002   1                1.2 ug/m3 LC
## 3  1/7/17   AQS 370110002   1                3.2 ug/m3 LC
## 4 1/10/17   AQS 370110002   1                6.4 ug/m3 LC
## 5 1/13/17   AQS 370110002   1                3.6 ug/m3 LC
## 6 1/16/17   AQS 370110002   1                5.8 ug/m3 LC
##  DAILY_AQI_VALUE      Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 1                12 Linville Falls                1            100
## 2                 5 Linville Falls                1            100
## 3                13 Linville Falls                1            100
## 4                27 Linville Falls                1            100
## 5                15 Linville Falls                1            100
## 6                24 Linville Falls                1            100
##  AQS_PARAMETER_CODE      AQS_PARAMETER_DESC CBSA_CODE
## 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
##  CBSA_NAME STATE_CODE      STATE COUNTY_CODE COUNTY SITE_LATITUDE
## 1                37 North Carolina            11 Avery      35.97235
## 2                37 North Carolina            11 Avery      35.97235
```

```
## 3          37 North Carolina          11 Avery          35.97235
## 4          37 North Carolina          11 Avery          35.97235
## 5          37 North Carolina          11 Avery          35.97235
## 6          37 North Carolina          11 Avery          35.97235
## SITE_LONGITUDE
## 1          -81.93307
## 2          -81.93307
## 3          -81.93307
## 4          -81.93307
## 5          -81.93307
## 6          -81.93307
```

```
tail(PM25_NC2017)
```

```
##          Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration   UNITS
## 9489 12/26/17   AQS 371830021   3                               4.1 ug/m3 LC
## 9490 12/27/17   AQS 371830021   3                               7.2 ug/m3 LC
## 9491 12/28/17   AQS 371830021   3                               7.1 ug/m3 LC
## 9492 12/29/17   AQS 371830021   3                              11.6 ug/m3 LC
## 9493 12/30/17   AQS 371830021   3                              15.3 ug/m3 LC
## 9494 12/31/17   AQS 371830021   3                               2.9 ug/m3 LC
## DAILY_AQI_VALUE Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 9489          17 Triple Oak           1           100
## 9490          30 Triple Oak           1           100
## 9491          30 Triple Oak           1           100
## 9492          48 Triple Oak           1           100
## 9493          58 Triple Oak           1           100
## 9494          12 Triple Oak           1           100
## AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME
## 9489          88101 PM2.5 - Local Conditions   39580 Raleigh, NC
## 9490          88101 PM2.5 - Local Conditions   39580 Raleigh, NC
## 9491          88101 PM2.5 - Local Conditions   39580 Raleigh, NC
## 9492          88101 PM2.5 - Local Conditions   39580 Raleigh, NC
## 9493          88101 PM2.5 - Local Conditions   39580 Raleigh, NC
## 9494          88101 PM2.5 - Local Conditions   39580 Raleigh, NC
## STATE_CODE STATE COUNTY_CODE COUNTY SITE_LATITUDE
## 9489          37 North Carolina          183 Wake          35.8652
## 9490          37 North Carolina          183 Wake          35.8652
## 9491          37 North Carolina          183 Wake          35.8652
## 9492          37 North Carolina          183 Wake          35.8652
## 9493          37 North Carolina          183 Wake          35.8652
## 9494          37 North Carolina          183 Wake          35.8652
## SITE_LONGITUDE
## 9489          -78.8197
## 9490          -78.8197
## 9491          -78.8197
## 9492          -78.8197
## 9493          -78.8197
## 9494          -78.8197
```

```
summary(PM25_NC2017$DAILY_AQI_VALUE)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.00   21.00   30.00   31.72   42.00   93.00
```

```
class(PM25_NC2017$Date)
```

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

```
# explore PM25_NC2018
```

```
dim(PM25_NC2018)
```

```
## [1] 7611 20
```

```
colnames(PM25_NC2018)
```

```
## [1] "Date" "Source"
## [3] "Site.ID" "POC"
## [5] "Daily.Mean.PM2.5.Concentration" "UNITS"
## [7] "DAILY_AQI_VALUE" "Site.Name"
## [9] "DAILY_OBS_COUNT" "PERCENT_COMPLETE"
## [11] "AQS_PARAMETER_CODE" "AQS_PARAMETER_DESC"
## [13] "CBSA_CODE" "CBSA_NAME"
## [15] "STATE_CODE" "STATE"
## [17] "COUNTY_CODE" "COUNTY"
## [19] "SITE_LATITUDE" "SITE_LONGITUDE"
```

```
class(PM25_NC2018)
```

```
## [1] "data.frame"
```

```
head(PM25_NC2018)
```

```
##      Date Source   Site.ID POC Daily.Mean.PM2.5.Concentration UNITS
## 1  1/2/18   AQS 370110002   1                2.9 ug/m3 LC
## 2  1/5/18   AQS 370110002   1                3.7 ug/m3 LC
## 3  1/8/18   AQS 370110002   1                5.3 ug/m3 LC
## 4 1/11/18   AQS 370110002   1                0.8 ug/m3 LC
## 5 1/14/18   AQS 370110002   1                2.5 ug/m3 LC
## 6 1/17/18   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
## 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
##  CBSA_NAME STATE_CODE   STATE COUNTY_CODE COUNTY SITE_LATITUDE
## 1                37 North Carolina                11 Avery      35.97235
## 2                37 North Carolina                11 Avery      35.97235
## 3                37 North Carolina                11 Avery      35.97235
## 4                37 North Carolina                11 Avery      35.97235
## 5                37 North Carolina                11 Avery      35.97235
## 6                37 North Carolina                11 Avery      35.97235
##  SITE_LONGITUDE
```



```
## 1      -81.93307
## 2      -81.93307
## 3      -81.93307
## 4      -81.93307
## 5      -81.93307
## 6      -81.93307
```

```
tail(03_NC2018)
```

```
##          Date Source   Site.ID POC Daily.Max.8.hour.Ozone.Concentration
## 10776 11/4/18 AirNow 371990004 1                                0.043
## 10777 11/5/18 AirNow 371990004 1                                0.044
## 10778 11/6/18 AirNow 371990004 1                                0.053
## 10779 11/7/18 AirNow 371990004 1                                0.053
## 10780 11/8/18 AirNow 371990004 1                                0.039
## 10781 11/11/18 AirNow 371990004 1                               0.059
##          UNITS DAILY_AQI_VALUE   Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## 10776   ppm                40 Mt. Mitchell           24           100
## 10777   ppm                41 Mt. Mitchell           24           100
## 10778   ppm                49 Mt. Mitchell           24           100
## 10779   ppm                49 Mt. Mitchell           24           100
## 10780   ppm                36 Mt. Mitchell           24           100
## 10781   ppm                64 Mt. Mitchell           24           100
##          AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE CBSA_NAME STATE_CODE
## 10776                44201                Ozone      NA           37
## 10777                44201                Ozone      NA           37
## 10778                44201                Ozone      NA           37
## 10779                44201                Ozone      NA           37
## 10780                44201                Ozone      NA           37
## 10781                44201                Ozone      NA           37
##          STATE COUNTY_CODE COUNTY SITE_LATITUDE SITE_LONGITUDE
## 10776 North Carolina      199 Yancey      35.76541      -82.26494
## 10777 North Carolina      199 Yancey      35.76541      -82.26494
## 10778 North Carolina      199 Yancey      35.76541      -82.26494
## 10779 North Carolina      199 Yancey      35.76541      -82.26494
## 10780 North Carolina      199 Yancey      35.76541      -82.26494
## 10781 North Carolina      199 Yancey      35.76541      -82.26494
```

```
summary(PM25_NC2018$DAILY_AQI_VALUE)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   21.00   30.00   31.03   41.00   97.00
```

```
class(PM25_NC2018$Date)
```

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

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.

```

#3

# check class of column date
class(O3_NC2017$Date)

## [1] "factor"

class(O3_NC2018$Date)

## [1] "factor"

class(PM25_NC2017$Date)

## [1] "factor"

class(PM25_NC2018$Date)

## [1] "factor"

# change the date columns from class factor to date
O3_NC2017$Date <- as.Date(O3_NC2017$Date, format = "%m/%d/%y")
O3_NC2018$Date <- as.Date(O3_NC2018$Date, format = "%m/%d/%y")
PM25_NC2017$Date <- as.Date(PM25_NC2017$Date, format = "%m/%d/%y")
PM25_NC2018$Date <- as.Date(PM25_NC2018$Date, format = "%m/%d/%y")

# confirm date columns are class date
class(O3_NC2017$Date)

## [1] "Date"

class(O3_NC2018$Date)

## [1] "Date"

class(PM25_NC2017$Date)

## [1] "Date"

class(PM25_NC2018$Date)

## [1] "Date"

#4

# create processed datasets
O3_NC2017_processed <- select(O3_NC2017, Date, DAILY_AQI_VALUE, Site.Name,
                              AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE)
O3_NC2018_processed <- select(O3_NC2018, Date, DAILY_AQI_VALUE, Site.Name,
                              AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE)
PM25_NC2017_processed <- select(PM25_NC2017, Date, DAILY_AQI_VALUE, Site.Name,
                                AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE)
PM25_NC2018_processed <- select(PM25_NC2018, 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"

PM25_NC2017_processed <- mutate(PM25_NC2017_processed, AQS_PARAMETER_DESC = "PM2.5")

```

```
PM25_NC2018_processed <- mutate(PM25_NC2018_processed, AQS_PARAMETER_DESC = "PM2.5")

class(PM25_NC2017_processed$AQS_PARAMETER_DESC)
```

```
## [1] "character"
```

```
#note: the column AQS_PARAMETER_DESC has changed from class factor to character
```

```
#6
```

```
# save processed datasets in processed folder
```

```
write.csv(O3_NC2017_processed, row.names = FALSE,
          file = "./Data/Processed/EPAair_O3_NC2017_processed.csv")
write.csv(O3_NC2018_processed, row.names = FALSE,
          file = "./Data/Processed/EPAair_O3_NC2018_processed.csv")
write.csv(PM25_NC2017_processed, row.names = FALSE,
          file = "./Data/Processed/EPAair_PM25_NC2017_processed.csv")
write.csv(PM25_NC2018_processed, row.names = FALSE,
          file = "./Data/Processed/EPAair_PM25_NC2018_processed.csv")
```

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:
 - Sites: Blackstone, Bryson City, Triple Oak
 - Add columns for “Month” and “Year” by parsing your “Date” column (hint: `separate` function or `lubridate` package)
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 for identical column names across the 4 datasets
```

```
all(colnames(O3_NC2017_processed) == colnames(O3_NC2018_processed))
```

```
## [1] TRUE
```

```
all(colnames(PM25_NC2017_processed) == colnames(PM25_NC2018_processed))
```

```
## [1] TRUE
```

```
all(colnames(O3_NC2017_processed) == colnames(PM25_NC2017_processed))
```

```
## [1] TRUE
```

```
# combine datasets vertically
```

```
EPAair_O3_PM25 <- rbind(O3_NC2017_processed, O3_NC2018_processed,
                       PM25_NC2017_processed, PM25_NC2018_processed)
```

```
#8
```

```
# wrangle based on conditions
```

```

EPAair_03_PM25_processed <-
EPAair_03_PM25 %>%
filter(Site.Name == "Blackstone" | Site.Name == "Bryson City" | Site.Name == "Triple Oak") %>%
separate(Date, c("Y", "m", "d")) %>%
rename(Year = "Y") %>%
rename(Month = "m") %>%
rename(Day = "d")

#9

# spread dataset so ozone and PM2.5 are in separate columns
EPAair_03_PM25_spread <- spread(EPAair_03_PM25_processed, AQS_PARAMETER_DESC, DAILY_AQI_VALUE)

#10

# check dimensions of spread dataset
dim(EPAair_03_PM25_spread)

## [1] 1953    9

#11

# save spread dataset
write.csv(EPAair_03_PM25_spread, row.names = FALSE,
          file = "./Data/Processed/EPAair_03_PM25_NC1718_Processed.csv")

```

Generate summary tables

12. Use the split-apply-combine strategy to generate two new data frames:

- A summary table of mean AQI values for O3 and PM2.5 by month
- A summary table of the mean, minimum, and maximum AQI values of O3 and PM2.5 for each site

13. Display the data frames.

```

#12a

# create summary of mean O3 and PM2.5 by month
mean_03_PM25_bymonth <-
  EPAair_03_PM25_spread %>%
  group_by(Month) %>%
  summarise(meanO3 = mean(Ozone, na.rm=TRUE),
            meanPM25 = mean(PM2.5, na.rm=TRUE))

#12b

# create a summary of mean/min/max O3 and PM2.5 by site
stats_03_PM25_bysite <-
  EPAair_03_PM25_spread %>%
  group_by(Site.Name) %>%
  summarise(meanO3 = mean(Ozone, na.rm=TRUE),
            meanPM25 = mean(PM2.5, na.rm=TRUE),
            minO3 = min(Ozone, na.rm=TRUE),
            minPM25 = min(PM2.5, na.rm=TRUE),
            maxO3 = max(Ozone, na.rm=TRUE),
            maxPM25 = max(PM2.5, na.rm=TRUE))

```

#13

display dataframes in console

```
print(mean_O3_PM25_bymonth)
```

```
## # A tibble: 12 x 3
##   Month meanO3 meanPM25
##   <chr>   <dbl>   <dbl>
## 1 01      31.5     34.6
## 2 02      35.5     36.7
## 3 03      42.4     35.1
## 4 04      44.3     32.5
## 5 05      38.9     31.7
## 6 06      38.7     33.3
## 7 07      38.2     33.1
## 8 08      34.0     33.7
## 9 09      32.6     31.9
## 10 10      32.1     29.3
## 11 11      30.1     36.8
## 12 12      29.8     41.1
```

```
print(stats_O3_PM25_bysite)
```

```
## # A tibble: 3 x 7
##   Site.Name meanO3 meanPM25 minO3 minPM25 maxO3 maxPM25
##   <fct>      <dbl>   <dbl> <dbl>   <dbl> <dbl>   <dbl>
## 1 Blackstone  38.5     36.7    8      0     97     83
## 2 Bryson City  35.2     32.3    5      3     71     78
## 3 Triple Oak   NaN      33.5   Inf     0    -Inf     74
```

note that the column Month is now a character