

# Assignment 4: Data Wrangling

*Yifei Zhang*

## 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
getwd()

## [1] "/Users/yifeizhang/R/Environmental Data Analytics"

library(tidyverse)

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

## v ggplot2 3.1.0      v purrr   0.2.5
## 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

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

EPAair_03_NC2017 <- read.csv("./Data/Raw/EPAair_03_NC2017_raw.csv")
EPAair_03_NC2018 <- read.csv("./Data/Raw/EPAair_03_NC2018_raw.csv")
EPAair_PM25_NC2018 <- read.csv("./Data/Raw/EPAair_PM25_NC2018_raw.csv")
EPAair_PM25_NC2017 <- read.csv("./Data/Raw/EPAair_PM25_NC2017_raw.csv")

#2
head(EPAair_03_NC2017, 3)
```

```
##      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
##      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
##      AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE
## 1              44201              Ozone      25860
## 2              44201              Ozone      25860
## 3              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
##      COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1 Alexander      35.9138      -81.191
## 2 Alexander      35.9138      -81.191
## 3 Alexander      35.9138      -81.191
```

```
head(EPAair_03_NC2018, 3)
```

```
##      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
##      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
##      AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE
## 1              44201              Ozone      25860
## 2              44201              Ozone      25860
## 3              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
##      COUNTY SITE_LATITUDE SITE_LONGITUDE
## 1 Alexander      35.9138      -81.191
## 2 Alexander      35.9138      -81.191
## 3 Alexander      35.9138      -81.191
```

```
head(EPAair_PM25_NC2017, 3)
```

```
##      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
##      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
##      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
##   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
##   SITE_LONGITUDE
## 1      -81.93307
## 2      -81.93307
## 3      -81.93307
```

```
head(EPAair_PM25_NC2018, 3)
```

```
##   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
##   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
##   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
##   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
##   SITE_LONGITUDE
## 1      -81.93307
## 2      -81.93307
## 3      -81.93307
```

```
summary(EPAair_03_NC2017)
```

```
##      Date      Source      Site.ID      POC
## 4/13/17: 40   AQS:10219   Min.    :370030005   Min.    :1
## 4/15/17: 40           1st Qu.:370650099   1st Qu.:1
## 4/18/17: 40           Median :371010002   Median :1
## 4/3/17 : 40           Mean   :370962005   Mean   :1
## 4/5/17 : 40           3rd Qu.:371239991   3rd Qu.:1
## 4/8/17 : 40           Max.    :371990004   Max.    :1
## (Other):9979
##   Daily.Max.8.hour.Ozone.Concentration UNITS      DAILY_AQI_VALUE
##   Min.    :0.00500                      ppm:10219   Min.    : 5.00
##   1st Qu.:0.03500                      1st Qu.: 32.00
##   Median :0.04300                      Median : 40.00
##   Mean   :0.04211                      Mean   : 39.87
##   3rd Qu.:0.04900                      3rd Qu.: 45.00
##   Max.    :0.07500                      Max.    :115.00
##
##           Site.Name      DAILY_OBS_COUNT PERCENT_COMPLETE
##   Garinger High School: 358   Min.    :13.00   Min.    : 76.00
```

```

## Blackstone      : 355  1st Qu.:17.00  1st Qu.:100.00
## Rockwell       : 354  Median :17.00  Median :100.00
## Coweeta        : 344  Mean   :16.94  Mean   : 99.63
## Millbrook School : 339  3rd Qu.:17.00  3rd Qu.:100.00
## Beaufort       : 338  Max.    :17.00  Max.    :100.00
## (Other)        :8131
## AQS_PARAMETER_CODE AQS_PARAMETER_DESC CBSA_CODE
## Min. :44201 Ozone:10219 Min. :11700
## 1st Qu.:44201 1st Qu.:16740
## Median :44201 Median :24660
## Mean :44201 Mean :27541
## 3rd Qu.:44201 3rd Qu.:39580
## Max. :44201 Max. :49180
## NA's :2541
## CBSA_NAME STATE_CODE
## :2541 Min. :37
## Charlotte-Concord-Gastonia, NC-SC:1428 1st Qu.:37
## Asheville, NC : 940 Median :37
## Winston-Salem, NC : 725 Mean :37
## Raleigh, NC : 584 3rd Qu.:37
## Durham-Chapel Hill, NC : 486 Max. :37
## (Other) :3515
## STATE COUNTY_CODE COUNTY
## North Carolina:10219 Min. : 3.00 Forsyth : 725
## 1st Qu.: 65.00 Haywood : 700
## Median :101.00 Mecklenburg: 601
## Mean : 96.07 Avery : 541
## 3rd Qu.:123.00 Cumberland : 464
## Max. :199.00 Swain : 429
## (Other) :6759
## SITE_LATITUDE SITE_LONGITUDE
## Min. :34.36 Min. : -83.80
## 1st Qu.:35.26 1st Qu.: -82.05
## Median :35.55 Median : -80.23
## Mean :35.60 Mean : -80.32
## 3rd Qu.:35.99 3rd Qu.: -78.77
## Max. :36.31 Max. : -76.62
##

```

#### summary(EPAair\_03\_NC2018)

```

## Date Source Site.ID POC
## 3/10/18: 39 AirNow:2718 Min. :370030005 Min. :1
## 3/11/18: 39 AQS :8063 1st Qu.:370630015 1st Qu.:1
## 3/13/18: 39 Median :370870036 Median :1
## 3/14/18: 39 Mean :370959550 Mean :1
## 3/15/18: 39 3rd Qu.:371290002 3rd Qu.:1
## 3/16/18: 39 Max. :371990004 Max. :1
## (Other):10547
## Daily.Max.8.hour.Ozone.Concentration UNITS DAILY_AQI_VALUE
## Min. :0.00000 ppm:10781 Min. : 0.00
## 1st Qu.:0.03400 1st Qu.: 31.00
## Median :0.04100 Median : 38.00
## Mean :0.04124 Mean : 39.46
## 3rd Qu.:0.04900 3rd Qu.: 45.00

```

```

## Max.      :0.07700                               Max.      :122.00
##
##           Site.Name   DAILY_OBS_COUNT PERCENT_COMPLETE
## Coweeta      : 340   Min.      :12.00   Min.      : 71.00
## Millbrook School : 338   1st Qu.:17.00   1st Qu.:100.00
## Candor       : 337   Median    :17.00   Median    :100.00
## Garinger High School: 333   Mean      :18.69   Mean      : 99.62
## Bethany sch.   : 332   3rd Qu.:18.00   3rd Qu.:100.00
## Cranberry     : 319   Max.      :24.00   Max.      :100.00
## (Other)       :8782
## AQS_PARAMETER_CODE AQS_PARAMETER_DESC   CBSA_CODE
## Min.      :44201   Ozone:10781   Min.      :11700
## 1st Qu.:44201           1st Qu.:16740
## Median :44201           Median    :24660
## Mean      :44201           Mean      :27015
## 3rd Qu.:44201           3rd Qu.:39580
## Max.      :44201           Max.      :49180
##                               NA's      :2802
##                               CBSA_NAME   STATE_CODE
##                               :2802   Min.      :37
## Charlotte-Concord-Gastonia, NC-SC:1469   1st Qu.:37
## Asheville, NC                               :1159   Median    :37
## Winston-Salem, NC                           : 754   Mean      :37
## Raleigh, NC                                 : 636   3rd Qu.:37
## Greensboro-High Point, NC                   : 595   Max.      :37
## (Other)                                     :3366
##           STATE      COUNTY_CODE      COUNTY
## North Carolina:10781   Min.      : 3.00   Haywood    : 879
##                               1st Qu.: 63.00   Forsyth    : 754
##                               Median    : 87.00   Mecklenburg: 632
##                               Mean      : 95.84   Avery      : 613
##                               3rd Qu.:129.00   Cumberland : 467
##                               Max.      :199.00   Swain      : 447
##                               (Other)    :6989
## SITE_LATITUDE   SITE_LONGITUDE
## Min.      :34.36   Min.      :-83.80
## 1st Qu.:35.26   1st Qu.: -82.05
## Median :35.59   Median    :-80.34
## Mean      :35.63   Mean      :-80.39
## 3rd Qu.:36.03   3rd Qu.: -78.90
## Max.      :36.31   Max.      :-76.62
##

```

#### summary(EPAair\_PM25\_NC2017)

```

##           Date      Source      Site.ID      POC
## 1/31/17: 45   AQS:9494   Min.      :370110002   Min.      :1.000
## 1/19/17: 44           1st Qu.:370630015   1st Qu.:3.000
## 11/3/17: 44           Median    :371010002   Median    :3.000
## 2/12/17: 44           Mean      :370980114   Mean      :2.734
## 4/1/17 : 44           3rd Qu.:371210004   3rd Qu.:3.000
## 5/31/17: 44           Max.      :371830021   Max.      :4.000
## (Other):9229
## Daily.Mean.PM2.5.Concentration      UNITS      DAILY_AQI_VALUE
## Min.      :-3.900      ug/m3 LC:9494   Min.      : 0.00

```

```

## 1st Qu.: 5.000                      1st Qu.:21.00
## Median : 7.300                      Median :30.00
## Mean   : 7.742                      Mean   :31.72
## 3rd Qu.:10.000                     3rd Qu.:42.00
## Max.    :31.900                     Max.    :93.00
##
##
##      Site.Name      DAILY_OBS_COUNT PERCENT_COMPLETE
## Board Of Ed. Bldg.   : 542   Min.      :1      Min.      :100
## Hattie Avenue       : 505   1st Qu.:1      1st Qu.:100
## Lexington water tower : 501   Median :1      Median :100
## Montclair Elementary School: 489   Mean    :1      Mean    :100
## Pitt Agri. Center   : 483   3rd Qu.:1      3rd Qu.:100
## West Johnston Co.    : 478   Max.     :1      Max.     :100
## (Other)              :6496
## AQS_PARAMETER_CODE      AQS_PARAMETER_DESC
## Min.      :88101      Acceptable PM2.5 AQI & Speciation Mass:2842
## 1st Qu.:88101      PM2.5 - Local Conditions      :6652
## Median :88101
## Mean    :88221
## 3rd Qu.:88502
## Max.    :88502
##
##      CBSA_CODE      CBSA_NAME      STATE_CODE
## Min.      :11700      Charlotte-Concord-Gastonia, NC-SC:1411   Min.      :37
## 1st Qu.:16740      Winston-Salem, NC      :1366   1st Qu.:37
## Median :25860      :1353   Median :37
## Mean    :30793      Raleigh, NC      :1285   Mean    :37
## 3rd Qu.:41820      Asheville, NC      : 657   3rd Qu.:37
## Max.    :49180      Greenville, NC      : 483   Max.    :37
## NA's    :1353      (Other)      :2939
##      STATE      COUNTY_CODE      COUNTY      SITE_LATITUDE
## North Carolina:9494   Min.      : 11   Mecklenburg:1411   Min.      :34.36
##      1st Qu.: 63   Forsyth      : 865   1st Qu.:35.26
##      Median :101   Wake        : 807   Median :35.64
##      Mean    : 98   Buncombe    : 542   Mean    :35.60
##      3rd Qu.:121   Davidson    : 501   3rd Qu.:35.91
##      Max.    :183   Pitt        : 483   Max.    :36.11
##      (Other)      :4885
## SITE_LONGITUDE
## Min.      :-83.44
## 1st Qu.: -80.87
## Median : -80.23
## Mean     : -80.03
## 3rd Qu.: -78.82
## Max.     : -76.21
##

```

[summary\(EPAair\\_PM25\\_NC2018\)](#)

```

##      Date      Source      Site.ID      POC
## 1/26/18: 39   AirNow: 783   Min.      :370110002   Min.      :1.000
## 2/1/18 : 39   AQS      :6828   1st Qu.:370630015   1st Qu.:3.000
## 2/19/18: 39      Median :371190041   Median :3.000
## 1/14/18: 38      Mean    :371031969   Mean     :3.011
## 1/8/18 : 38      3rd Qu.:371290002   3rd Qu.:3.000

```

```

## 2/7/18 : 38 Max. :371830021 Max. :5.000
## (Other):7380
## Daily.Mean.PM2.5.Concentration UNITS DAILY_AQI_VALUE
## Min. :-2.800 ug/m3 LC:7611 Min. : 0.00
## 1st Qu.: 5.000 1st Qu.:21.00
## Median : 7.200 Median :30.00
## Mean : 7.554 Mean :31.03
## 3rd Qu.: 9.800 3rd Qu.:41.00
## Max. :34.200 Max. :97.00
##
## Site.Name DAILY_OBS_COUNT PERCENT_COMPLETE
## Millbrook School : 621 Min. :1 Min. :100
## Board Of Ed. Bldg. : 428 1st Qu.:1 1st Qu.:100
## Garinger High School : 421 Median :1 Median :100
## Durham Armory : 415 Mean :1 Mean :100
## Lexington water tower: 411 3rd Qu.:1 3rd Qu.:100
## Pitt Agri. Center : 409 Max. :1 Max. :100
## (Other) :4906
## AQS_PARAMETER_CODE AQS_PARAMETER_DESC
## Min. :88101 Acceptable PM2.5 AQI & Speciation Mass:1246
## 1st Qu.:88101 PM2.5 - Local Conditions :6365
## Median :88101
## Mean :88167
## 3rd Qu.:88101
## Max. :88502
##
## CBSA_CODE CBSA_NAME STATE_CODE
## Min. :11700 Raleigh, NC :1274 Min. :37
## 1st Qu.:19000 Charlotte-Concord-Gastonia, NC-SC:1171 1st Qu.:37
## Median :25860 :1025 Median :37
## Mean :30249 Winston-Salem, NC : 803 Mean :37
## 3rd Qu.:39580 Asheville, NC : 447 3rd Qu.:37
## Max. :49180 Durham-Chapel Hill, NC : 415 Max. :37
## NA's :1025 (Other) :2476
## STATE COUNTY_CODE COUNTY SITE_LATITUDE
## North Carolina:7611 Min. : 11.0 Mecklenburg:1171 Min. :34.36
## 1st Qu.: 63.0 Wake : 947 1st Qu.:35.26
## Median :119.0 Buncombe : 428 Median :35.64
## Mean :103.2 Durham : 415 Mean :35.59
## 3rd Qu.:129.0 Davidson : 411 3rd Qu.:35.87
## Max. :183.0 Pitt : 409 Max. :36.11
## (Other) :3830
## SITE_LONGITUDE
## Min. :-83.44
## 1st Qu.: -80.87
## Median : -79.84
## Mean : -79.95
## 3rd Qu.: -78.57
## Max. : -76.21
##

```

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
class(EPAair_03_NC2017$Date)

## [1] "factor"
class(EPAair_03_NC2018$Date)

## [1] "factor"
class(EPAair_PM25_NC2017$Date)

## [1] "factor"
class(EPAair_PM25_NC2018$Date)

## [1] "factor"
EPAair_03_NC2017$Date <- as.Date(EPAair_03_NC2017$Date,format="%m/%d/%y")
EPAair_03_NC2018$Date <- as.Date(EPAair_03_NC2018$Date,format="%m/%d/%y")
EPAair_PM25_NC2017$Date <- as.Date(EPAair_PM25_NC2017$Date,format="%m/%d/%y")
EPAair_PM25_NC2018$Date <- as.Date(EPAair_PM25_NC2018$Date,format="%m/%d/%y")

#4
EPAair_03_NC2017_processed <- select(EPAair_03_NC2017, Date, DAILY_AQI_VALUE, Site.Name, AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE)
EPAair_03_NC2018_processed <- select(EPAair_03_NC2018, Date, DAILY_AQI_VALUE, Site.Name, AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE)
EPAair_PM25_NC2017_processed <- select(EPAair_PM25_NC2017, Date, DAILY_AQI_VALUE, Site.Name, AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE)
EPAair_PM25_NC2018_processed <- select(EPAair_PM25_NC2018, Date, DAILY_AQI_VALUE, Site.Name, AQS_PARAMETER_DESC, COUNTY, SITE_LATITUDE, SITE_LONGITUDE)

#5
EPAair_PM25_NC2017_processed$AQS_PARAMETER_DESC <- "PM2.5"
EPAair_PM25_NC2018_processed$AQS_PARAMETER_DESC <- "PM2.5"

#6
write.csv(EPAair_03_NC2017_processed, row.names = FALSE, file = "./Data/Processed/EPAair_03_NC2017_processed.csv")
write.csv(EPAair_03_NC2018_processed, row.names = FALSE, file = "./Data/Processed/EPAair_03_NC2018_processed.csv")
write.csv(EPAair_PM25_NC2017_processed, row.names = FALSE, file = "./Data/Processed/EPAair_PM25_NC2017_processed.csv")
write.csv(EPAair_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
EPAair_PM25_O3_NC2017_NC2018 <- rbind(EPAair_O3_NC2017_processed, EPAair_O3_NC2018_processed, EPAair_PM25_O3_NC2017_processed, EPAair_PM25_O3_NC2018_processed)

#8
EPAair_PM25_O3_NC2017_NC2018_processed <- EPAair_PM25_O3_NC2017_NC2018_processed %>%
  filter(Site.Name == "Blackstone" | Site.Name == "Bryson City" | Site.Name == "Triple Oak") %>%
  separate(Date, c("Y", "m", "d"))
  colnames(EPAair_PM25_O3_NC2017_NC2018_processed)[1:3] <- c("Year", "Month", "Day")

#9
EPAair_PM25_O3_NC2017_NC2018_processed <- spread(EPAair_PM25_O3_NC2017_NC2018_processed, AQS_PARAMETER_ID, value)

#10
dim(EPAair_PM25_O3_NC2017_NC2018_processed)

## [1] 1953    9

#11
write.csv(EPAair_PM25_O3_NC2017_NC2018_processed, row.names = FALSE, file = "./Data/Processed/EPAair_O3_NC2017_NC2018.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
EPAair_PM25_O3_NC1718_Summaries_by_Month <-
  EPAair_PM25_O3_NC2017_NC2018_processed %>%
  group_by(Month) %>%
  summarise(meanAQI_O3 = mean(na.omit(Ozone)),
            meanAQI_PM2.5 = mean(na.omit(PM2.5)))

#12b
EPAair_PM25_O3_NC1718_Summaries_by_Site <-
  EPAair_PM25_O3_NC2017_NC2018_processed %>%
  group_by(Site.Name) %>%
  summarise(meanAQI_O3 = mean(na.omit(Ozone)),
            meanAQI_PM2.5 = mean(na.omit(PM2.5)),
            minAQI_O3 = min(na.omit(Ozone)),
            minAQI_PM2.5 = min(na.omit(PM2.5)),
            maxAQI_O3 = max(na.omit(Ozone)),
            maxAQI_PM2.5 = max(na.omit(PM2.5)))

## Warning in min(na.omit(Ozone)): no non-missing arguments to min; returning
## Inf

## Warning in max(na.omit(Ozone)): no non-missing arguments to max; returning
## -Inf

#13
EPAair_PM25_O3_NC1718_Summaries_by_Month

```

```
## # A tibble: 12 x 3
##   Month meanAQI_O3 meanAQI_PM2.5
##   <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
```

EPAair\_PM25\_O3\_NC1718\_Summaries\_by\_Site

```
## # A tibble: 3 x 7
##   Site.Name meanAQI_O3 meanAQI_PM2.5 minAQI_O3 minAQI_PM2.5 maxAQI_O3
##   <fct>      <dbl>      <dbl>      <dbl>      <int>      <dbl>
## 1 Blacksto~    38.5        36.7         8         0         97
## 2 Bryson C~    35.2        32.3         5         3         71
## 3 Triple O~    NaN         33.5      Inf         0        -Inf
## # ... with 1 more variable: maxAQI_PM2.5 <int>
```