

# Assignment 10: Data Scraping

Hannah Nelson

## OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics on data scraping.

## Directions

1. Rename this file `<FirstLast>_A10_DataScraping.Rmd` (replacing `<FirstLast>` with your first and last name).
2. Change “Student Name” on line 3 (above) with your name.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure your code is tidy; use line breaks to ensure your code fits in the knitted output.
5. Be sure to **answer the questions** in this assignment document.
6. When you have completed the assignment, **Knit** the text and code into a single PDF file.

## Set up

1. Set up your session:
  - Load the packages `tidyverse`, `rvest`, and any others you end up using.
  - Check your working directory

```
#1
library(here)
library(rvest)
library(tidyverse)
library(tidycensus)
library(dataRetrieval)
library(lubridate)

here()
```

```
## [1] "/Users/hannahnelson/Desktop/env872/EDA-Fall2023"
```

2. We will be scraping data from the NC DEQs Local Water Supply Planning website, specifically the Durham’s 2022 Municipal Local Water Supply Plan (LWSP):
  - Navigate to <https://www.ncwater.org/WUDC/app/LWSP/search.php>
  - Scroll down and select the LWSP link next to Durham Municipality.
  - Note the web address: <https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=03-32-010&year=2022>

Indicate this website as the as the URL to be scraped. (In other words, read the contents into an `rvest` webpage object.)

```
#2
site <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?psid=03-32-010&year=2022")
```

3. The data we want to collect are listed below:

- From the “1. System Information” section:
- Water system name
- PWSID
- Ownership
- From the “3. Water Supply Sources” section:
- Maximum Day Use (MGD) - for each month

In the code chunk below scrape these values, assigning them to four separate variables.

HINT: The first value should be “Durham”, the second “03-32-010”, the third “Municipality”, and the last should be a vector of 12 numeric values (represented as strings)“.

```
#3

wsysname <- site %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

wsysname
```

```
## [1] "Durham"
```

```
PWSID
```

```
## [1] "03-32-010"
```

```
ownership
```

```
## [1] "Municipality"
```

```
MGD
```

```
## [1] "36.1000" "43.4200" "52.4900" "30.5000" "42.5900" "34.8800" "39.9100"  
## [8] "43.3200" "32.5300" "34.6600" "41.8000" "37.5300"
```

```
MGD <- c(36.1000, 30.5000, 39.9100, 34.6600, 43.4200, 42.5900, 43.3200,  
         41.8000, 52.4900, 34.8800, 32.5300, 37.5300)
```

4. Convert your scraped data into a dataframe. This dataframe should have a column for each of the 4 variables scraped and a row for the month corresponding to the withdrawal data. Also add a Date column that includes your month and year in data format. (Feel free to add a Year column too, if you wish.)

TIP: Use `rep()` to repeat a value when creating a dataframe.

NOTE: It's likely you won't be able to scrape the monthly withdrawal data in chronological order. You can overcome this by creating a month column manually assigning values in the order the data are scraped: "Jan", "May", "Sept", "Feb", etc... Or, you could scrape month values from the web page...

5. Create a line plot of the maximum daily withdrawals across the months for 2022

```
#4
```

```
durham_2022 <- data.frame(stringsAsFactors = T,  
                          "Water System Name" = wsysname,  
                          "Ownership" = ownership,  
                          "PWSID" = PWSID,  
                          "Month" = month.abb,  
                          "MGD" = MGD)  
  
durham_2022$Month <- factor(durham_2022$Month,  
                           levels=c("Jan", "Feb", "Mar", "Apr", "May",  
                                    "Jun", "Jul", "Aug", "Sep", "Oct",  
                                    "Nov", "Dec"))  
  
durham_2022
```

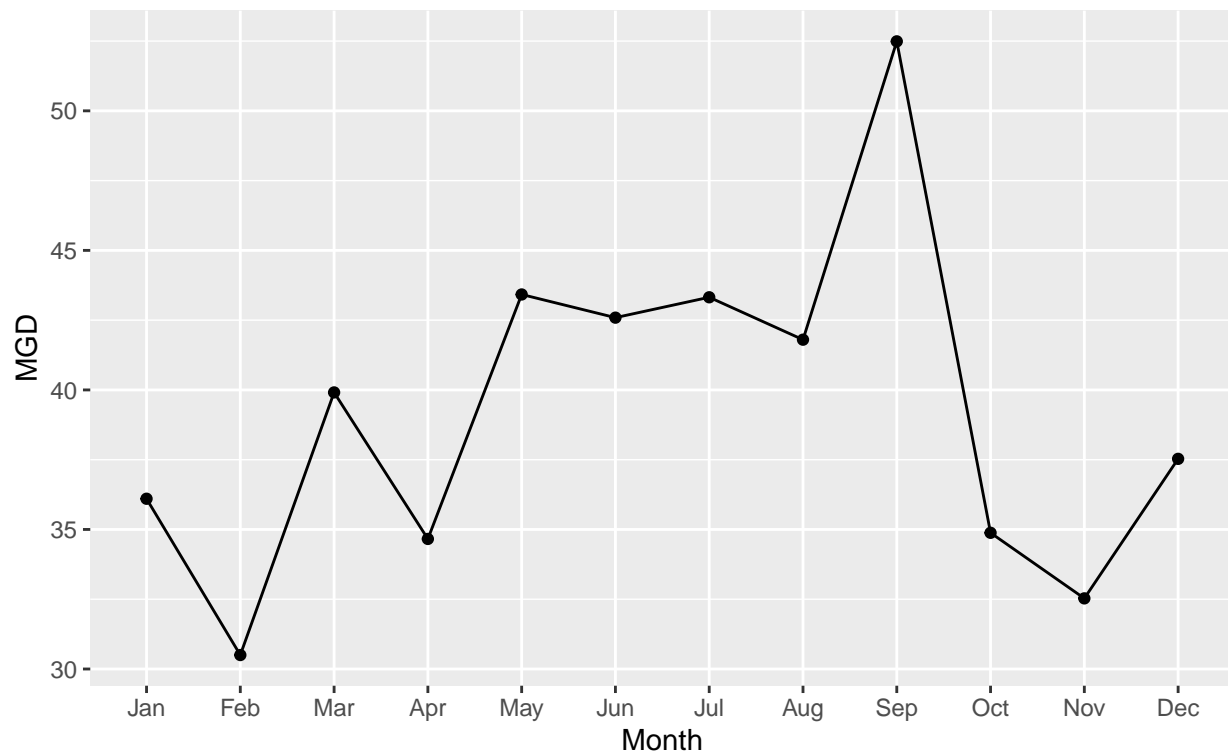
```
##   Water.System.Name Ownership   PWSID Month  MGD  
## 1      Durham Municipality 03-32-010   Jan 36.10  
## 2      Durham Municipality 03-32-010   Feb 30.50  
## 3      Durham Municipality 03-32-010   Mar 39.91  
## 4      Durham Municipality 03-32-010   Apr 34.66  
## 5      Durham Municipality 03-32-010   May 43.42  
## 6      Durham Municipality 03-32-010   Jun 42.59  
## 7      Durham Municipality 03-32-010   Jul 43.32
```

```
## 8      Durham Municipality 03-32-010    Aug 41.80
## 9      Durham Municipality 03-32-010    Sep 52.49
## 10     Durham Municipality 03-32-010    Oct 34.88
## 11     Durham Municipality 03-32-010    Nov 32.53
## 12     Durham Municipality 03-32-010    Dec 37.53
```

#5

```
ggplot(durham_2022, aes(y = MGD, x = Month, group = 1)) +
  geom_point() +
  geom_line() +
  labs(title = "Maximum Daily Withdrawals for Each Month of 2022",
        subtitle = "Durham, NC")
```

Maximum Daily Withdrawals for Each Month of 2022  
Durham, NC



- Note that the PWSID and the year appear in the web address for the page we scraped. Construct a function using your code above that can scrape data for any PWSID and year for which the NC DEQ has data. **Be sure to modify the code to reflect the year and site (pwsid) scraped.**

#6.

```
scrape <- function(PWSID, year) {
  url <- paste0("https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=",
                PWSID, "&year=", year)

  page <- read_html(url)
```

```

wsysname <- site %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text

PWSID <- site %>%
html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
html_text

ownership <- site %>%
html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
html_text

MGD <- site %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text

year <- 2022

months <- month.abb

}

```

7. Use the function above to extract and plot max daily withdrawals for Durham (PWSID='03-32-010') for each month in 2015

```

#7

site2 <-read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=03-32-010&year=2015")

wsysname <- site2 %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text

PWSID <- site2 %>%
html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
html_text

ownership <- site2 %>%
html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
html_text

MGD <- site2 %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text

wsysname

## [1] "Durham"

```

```
PWSID
```

```
## [1] "03-32-010"
```

```
ownership
```

```
## [1] "Municipality"
```

```
MGD
```

```
## [1] "40.2500" "53.1700" "40.0300" "43.5000" "57.0200" "38.7200" "43.1000"  
## [8] "41.6500" "43.5500" "49.6800" "44.7000" "48.7500"
```

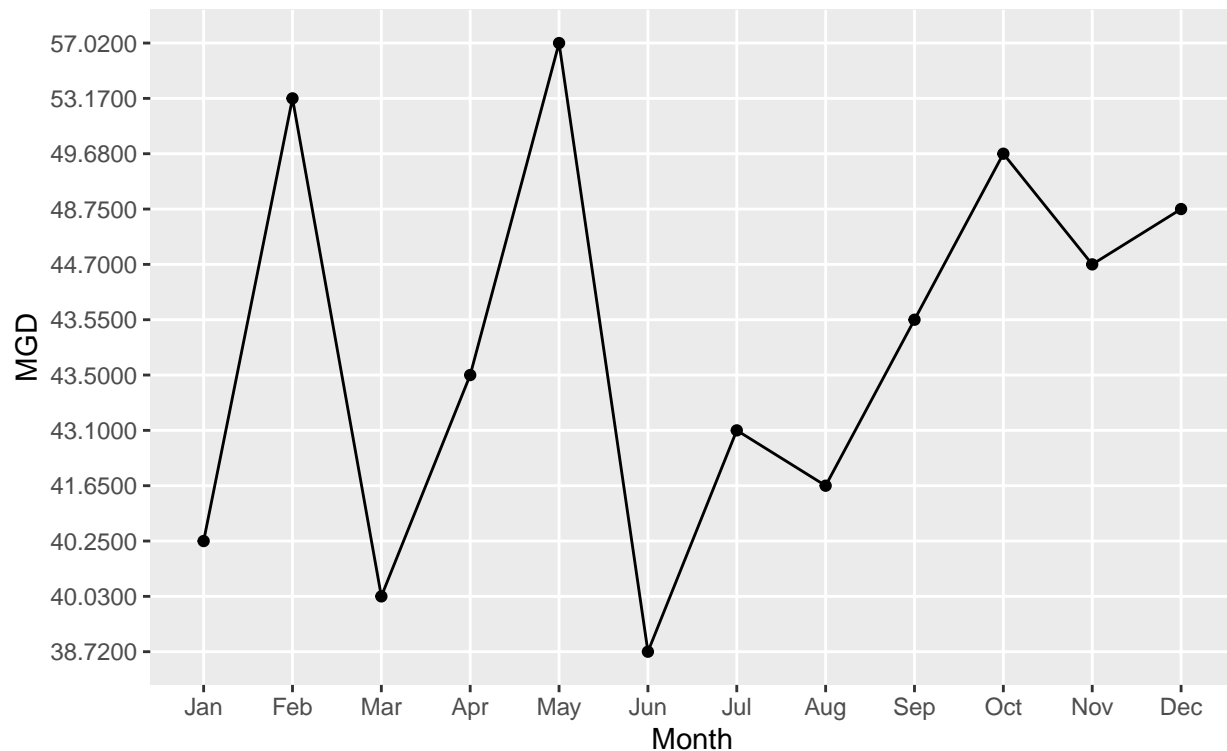
```
durham_2015 <- data.frame(stringsAsFactors = T,  
  "Water System Name" = wsysname,  
  "Ownership" = ownership,  
  "PWSID" = PWSID,  
  "Month" = month.abb,  
  "MGD" = MGD)  
  
durham_2015$Month <- factor(durham_2015$Month,  
  levels = c("Jan", "Feb", "Mar", "Apr", "May",  
    "Jun", "Jul", "Aug", "Sep", "Oct",  
    "Nov", "Dec"))
```

```
durham_2015
```

```
##   Water.System.Name Ownership   PWSID Month   MGD  
## 1      Durham Municipality 03-32-010   Jan 40.2500  
## 2      Durham Municipality 03-32-010   Feb 53.1700  
## 3      Durham Municipality 03-32-010   Mar 40.0300  
## 4      Durham Municipality 03-32-010   Apr 43.5000  
## 5      Durham Municipality 03-32-010   May 57.0200  
## 6      Durham Municipality 03-32-010   Jun 38.7200  
## 7      Durham Municipality 03-32-010   Jul 43.1000  
## 8      Durham Municipality 03-32-010   Aug 41.6500  
## 9      Durham Municipality 03-32-010   Sep 43.5500  
## 10     Durham Municipality 03-32-010   Oct 49.6800  
## 11     Durham Municipality 03-32-010   Nov 44.7000  
## 12     Durham Municipality 03-32-010   Dec 48.7500
```

```
ggplot(durham_2015, aes(y = MGD, x = Month, group = 1)) +  
  geom_point() +  
  geom_line() +  
  labs(title = "Maximum Daily Withdrawals for Each Month of 2015",  
    subtitle = "Durham, NC")
```

## Maximum Daily Withdrawals for Each Month of 2015 Durham, NC



- Use the function above to extract data for Asheville (PWSID = 01-11-010) in 2015. Combine this data with the Durham data collected above and create a plot that compares Asheville's to Durham's water withdrawals.

```
#8
site3 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2015")

wsysname <- site3 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site3 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site3 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site3 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

wsysname
```

```
## [1] "Asheville"
```

```
PWSID
```

```
## [1] "01-11-010"
```

```
ownership
```

```
## [1] "Municipality"
```

```
MGD
```

```
## [1] "20.8100" "23.9500" "22.9700" "24.5400" "23.5300" "21.3200" "21.4200"  
## [8] "23.6800" "20.4500" "21.6000" "24.1100" "19.8800"
```

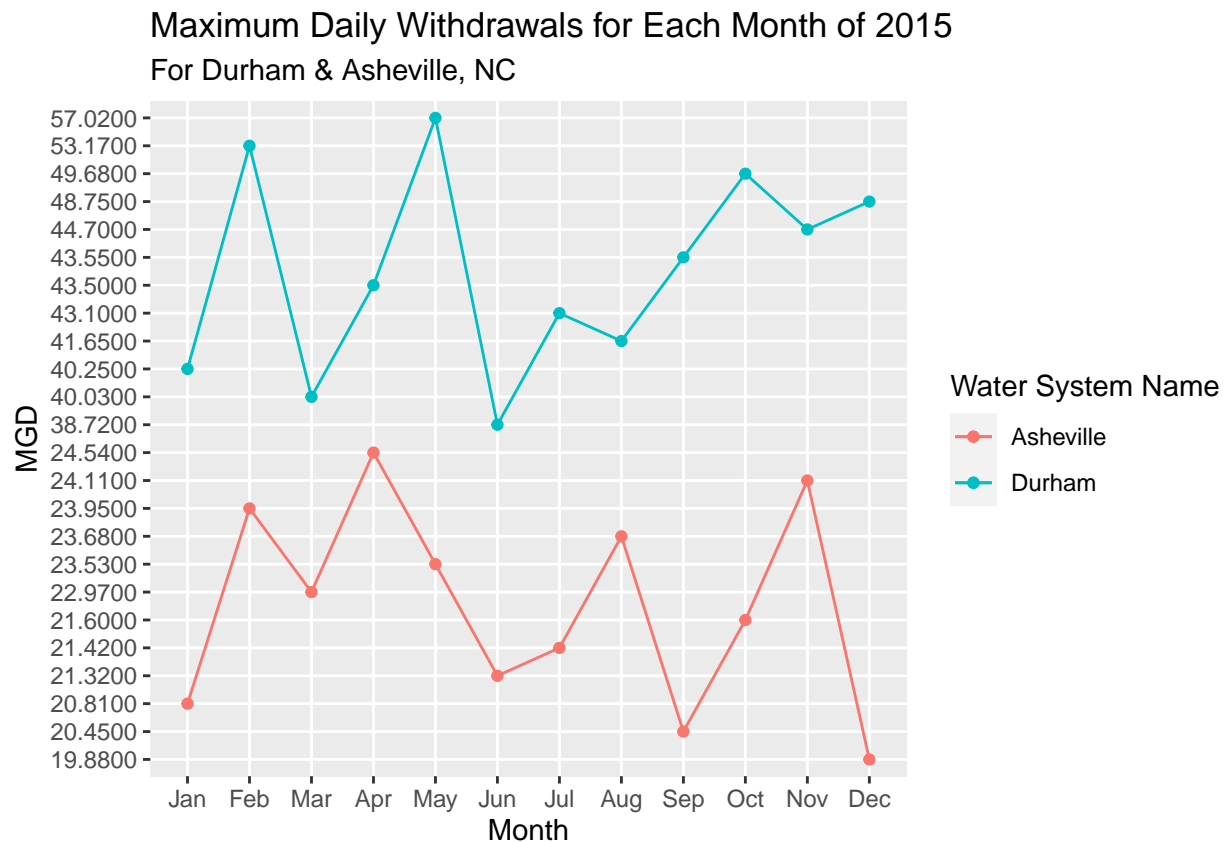
```
asheville_2015 <- data.frame(stringsAsFactors = T,  
  "Water System Name" = wsysname,  
  "Ownership" = ownership,  
  "PWSID" = PWSID,  
  "Month" = month.abb,  
  "MGD" = MGD)  
  
asheville_2015$Month <- factor(asheville_2015$Month,  
  levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",  
    "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))  
  
asheville_2015
```

```
##   Water.System.Name  Ownership   PWSID Month   MGD  
## 1      Asheville Municipality 01-11-010   Jan 20.8100  
## 2      Asheville Municipality 01-11-010   Feb 23.9500  
## 3      Asheville Municipality 01-11-010   Mar 22.9700  
## 4      Asheville Municipality 01-11-010   Apr 24.5400  
## 5      Asheville Municipality 01-11-010   May 23.5300  
## 6      Asheville Municipality 01-11-010   Jun 21.3200  
## 7      Asheville Municipality 01-11-010   Jul 21.4200  
## 8      Asheville Municipality 01-11-010   Aug 23.6800  
## 9      Asheville Municipality 01-11-010   Sep 20.4500  
## 10     Asheville Municipality 01-11-010   Oct 21.6000  
## 11     Asheville Municipality 01-11-010   Nov 24.1100  
## 12     Asheville Municipality 01-11-010   Dec 19.8800
```

```
ggplot() +  
  geom_line(data = asheville_2015,  
    aes(x = Month, y = MGD, group = 1, color = Water.System.Name)) +  
  geom_point(data = asheville_2015,  
    aes(x = Month, y = MGD, group = 1, color = Water.System.Name)) +  
  geom_line(data = durham_2015,  
    aes(x = Month, y = MGD, group = 1, color = Water.System.Name)) +
```



```
geom_point(data = durham_2015,
           aes(x = Month, y = MGD, group = 1, color = Water.System.Name)) +
labs(title = "Maximum Daily Withdrawals for Each Month of 2015",
     subtitle = "For Durham & Asheville, NC",
     color = "Water System Name")
```



- Use the code & function you created above to plot Asheville's max daily withdrawal by months for the years 2010 thru 2021. Add a smoothed line to the plot (method = 'loess').

TIP: See Section 3.2 in the "10\_Data\_Scraping.Rmd" where we apply "map2()" to iteratively run a function over two inputs. Pipe the output of the map2() function to bindrows() to combine the dataframes into a single one.

#9

```
site2010 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2010")

wsysname <- site2010 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2010 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
```

```

html_text

ownership <- site2010 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2010 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2010 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2010$Month <- factor(asheville_2010$Month,
  levels=c("Jan","Feb","Mar","Apr","May","Jun",
    "Jul","Aug","Sep","Oct","Nov","Dec"))

site2011 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2011")

wsysname <- site2011 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2011 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2011 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2011 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2011 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2011$Month <- factor(asheville_2011$Month,

```

```

                                levels=c("Jan","Feb","Mar","Apr","May","Jun",
                                           "Jul","Aug","Sep","Oct","Nov","Dec"))

site2012 <-read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?psid=01-11-010&year=2012")

wsysname <- site2012 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2012 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2012 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2012 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2012 <- data.frame(stringsAsFactors = T,
                             "Water System Name" = wsysname,
                             "Ownership" = ownership,
                             "PWSID" = PWSID,
                             "Month" = month.abb,
                             "MGD" = MGD)

asheville_2012$Month <- factor(asheville_2012$Month,
                              levels=c("Jan","Feb","Mar","Apr","May","Jun",
                                         "Jul","Aug","Sep","Oct","Nov","Dec"))

site2013 <-read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?psid=01-11-010&year=2013")

wsysname <- site2013 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2013 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2013 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

```

```

MGD <- site2013 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2013 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2013$Month <- factor(asheville_2013$Month,
  levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
    "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

site2014 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2014")

wsysname <- site2014 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2014 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2014 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2014 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2014 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2014$Month <- factor(asheville_2014$Month,
  levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
    "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

site2015 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2015")

wsysname <- site2015 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

```

```

PWSID <- site2015 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2015 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2015 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2015 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2015$Month <- factor(asheville_2015$Month,
  levels=c("Jan","Feb","Mar","Apr","May","Jun",
    "Jul","Aug","Sep","Oct","Nov","Dec"))

site2016 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2016")

wsysname <- site2016 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2016 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2016 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2016 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2016 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

```

```

asheville_2016$Month <- factor(asheville_2016$Month,
                              levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
                                         "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

site2017 <-read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2017")

wsysname <- site2017 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2017 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2017 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2017 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2017 <- data.frame(stringsAsFactors = T,
                             "Water System Name" = wsysname,
                             "Ownership" = ownership,
                             "PWSID" = PWSID,
                             "Month" = month.abb,
                             "MGD" = MGD)

asheville_2017$Month <- factor(asheville_2017$Month,
                              levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
                                         "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

site2018 <-read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2018")

wsysname <- site2018 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2018 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2018 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

```

```

MGD <- site2018 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2018 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2018$Month <- factor(asheville_2018$Month,
  levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
    "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

site2019 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2019")

wsysname <- site2019 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2019 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2019 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2019 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2019 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2019$Month <- factor(asheville_2019$Month,
  levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
    "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

site2020 <- read_html(

```

```

"https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2020")

wsysname <- site2020 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2020 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2020 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2020 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

asheville_2020 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2020$Month <- factor(asheville_2020$Month,
  levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
    "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

site2021 <- read_html(
  "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=01-11-010&year=2021")

wsysname <- site2021 %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text

PWSID <- site2021 %>%
  html_nodes("td tr:nth-child(1) td:nth-child(5)") %>%
  html_text

ownership <- site2021 %>%
  html_nodes("div+ table tr:nth-child(2) td:nth-child(4)") %>%
  html_text

MGD <- site2021 %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text

```



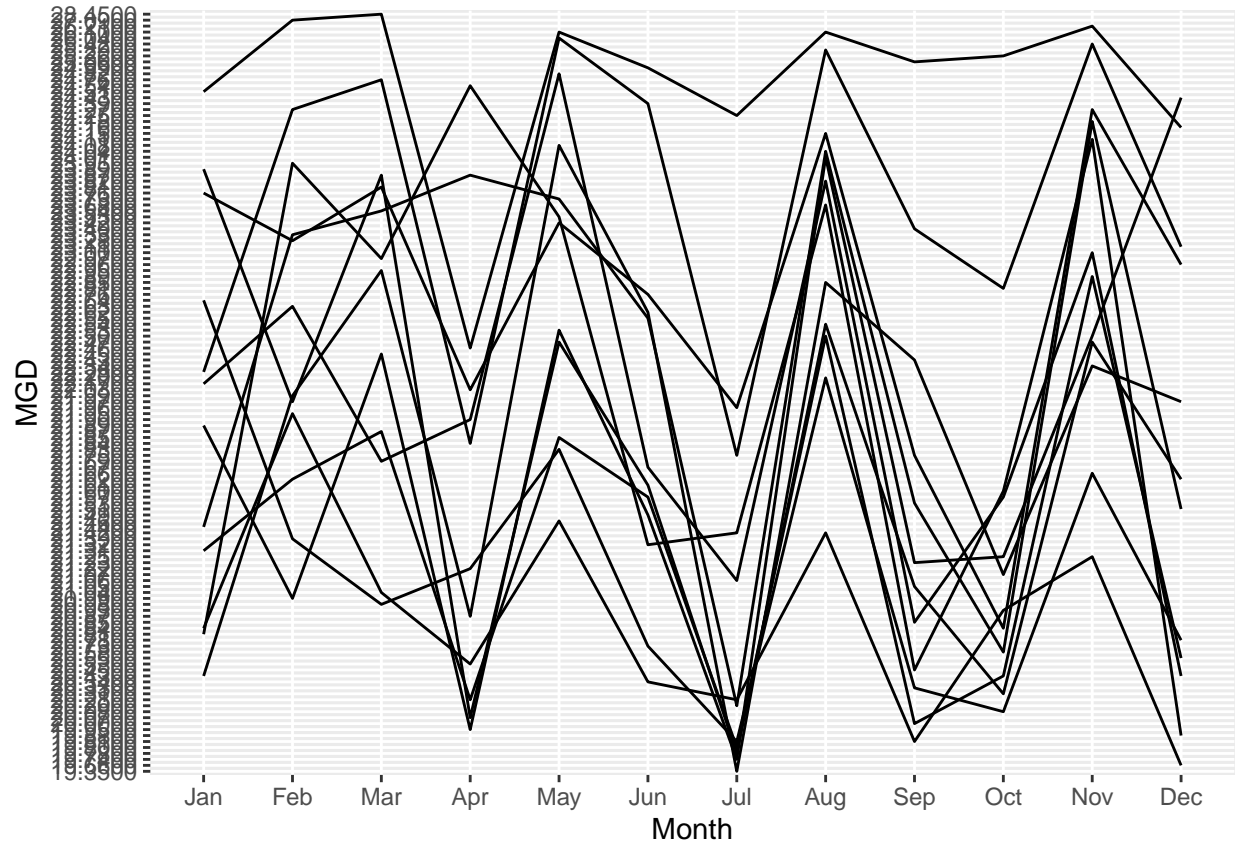
```

asheville_2021 <- data.frame(stringsAsFactors = T,
  "Water System Name" = wsysname,
  "Ownership" = ownership,
  "PWSID" = PWSID,
  "Month" = month.abb,
  "MGD" = MGD)

asheville_2021$Month <- factor(asheville_2021$Month,
  levels=c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
    "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

ggplot() +
  geom_line(data = asheville_2010,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2011,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2012,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2013,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2014,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2015,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2016,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2017,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2018,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2019,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2020,
    aes(x = Month, y = MGD, group = 1)) +
  geom_line(data = asheville_2021,
    aes(x = Month, y = MGD, group = 1))

```



Question: Just by looking at the plot (i.e. not running statistics), does Asheville have a trend in water usage over time? > Answer: >