

# Assignment 10: Data Scraping

Jaleesia Amos

## 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
#-----Check working directory-----#
getwd()

## [1] "/Users/jaleesiad.amos/Documents/EDA-Spring2023"

#-----Loading packages-----#
pacman::p_load(tidyverse, lubridate, here, rvest)

#-----Set ggplot theme-----#
jatheme <- theme_bw(base_size = 12) +
  theme(plot.title = element_text(face = "bold", hjust = 0.5),
        axis.title = element_text(face = "bold"),
        legend.position = "right")
theme_set(jatheme)
```

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
webpage.LWSP <- read_html('https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=03-32-010&year=2022')
webpage.LWSP

## {html_document}
## <html xmlns="http://www.w3.org/1999/xhtml" lang="en" xml:lang="en">
## [1] <head>\n<title>DWR :: Local Water Supply Planning</title>\n<meta http-equiv= ...
## [2] <body id="plan">\r\n<!--<div id="division-header">\r\n<a name="top" href= ...
```

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), with the first value being “27.6400”.

```
#3
#-----Scraped data values-----#
#_____System Information_____#
water_system_name <- webpage.LWSP %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text()
water_system_name

## [1] "Durham"

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

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

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

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

```
#-----Water Supply Sources-----#
max.withdrawals.mgd <- webpage.LWSP %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text()
max.withdrawals.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"
```

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...

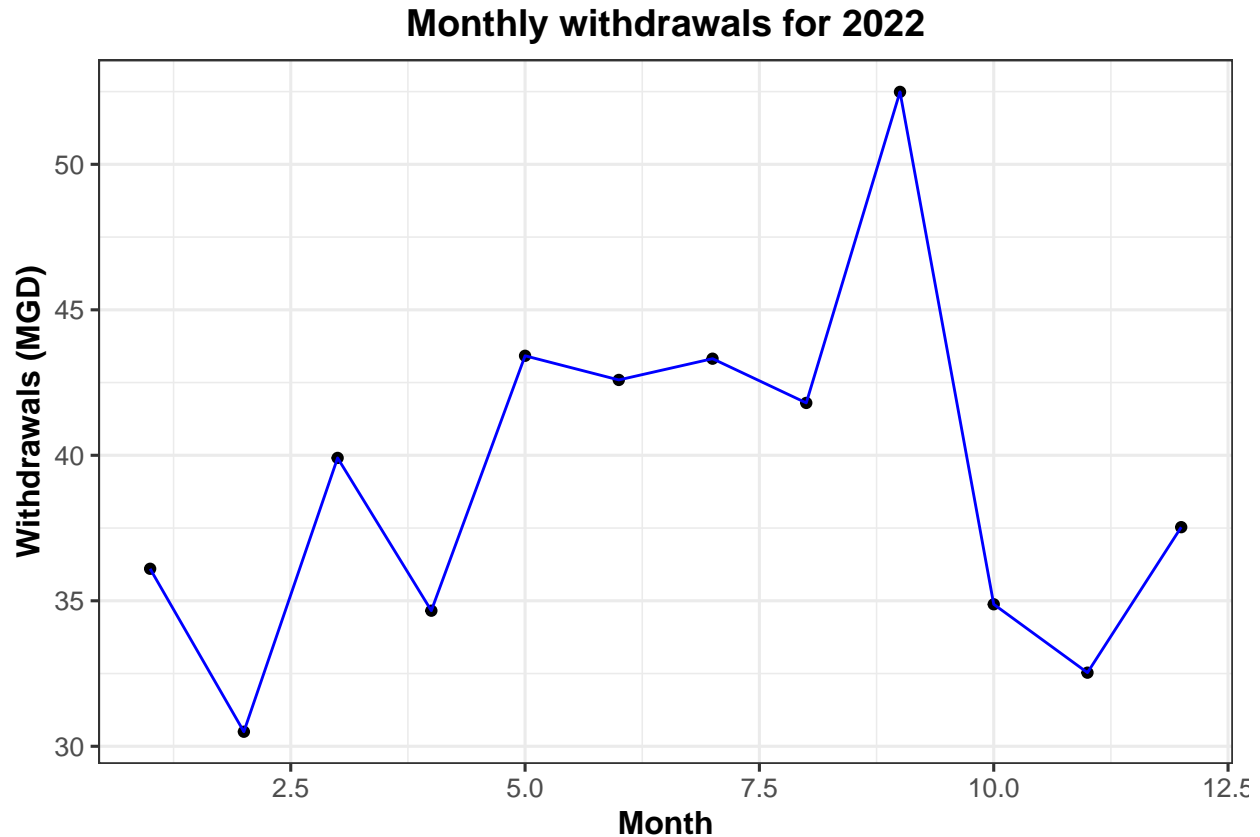
5. Create a line plot of the average daily withdrawals across the months for 2022

```
#4
#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2022,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.df <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#5
ggplot(monthly.withdrawals.df, aes(y = max.withdrawals.mgd, x = Month)) +
  geom_point() +
```

```
geom_line(color = "blue") +
  xlab("Month") +
  ylab("Withdrawals (MGD)") +
  ggtitle("Monthly withdrawals for 2022")
```



6. 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.
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "03-32-010"
the_year <- "2022"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

#-----Create function-----#
scrape.it.fun <- function(base_url, the_PWSID, the_year){

  #-----Read scraped website-----#
  scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)
  new_website <- read_html(scraped_url)

  #-----Scraped data elements from website-----#
  #-----System Information-----#
  water_system_name <- webpage.LWSP %>%
```

```

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

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

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

#-----Water Supply Sources-----#
max.withdrawals.mgd <- webpage.LWSP %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2022,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.dur <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#_____Return data frame_____#
return(monthly.withdrawals.dur)
}

```

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

```

#7
#-----Scraping Components-----#

the_PWSID <- "03-32-010"
the_year <- "2015"
scraped_url <- paste0(base_url, "=", the_PWSID, "&year=", the_year)

#---function for some reason wouldn't work so I had to solve the problem going through individually---#

scraped_url <- paste0("https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid=", the_PWSID, "&year=",

```

```

new_website <- read_html(scraped_url)

#-----Scraped data elements from website-----#
#-----System Information-----#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name

```

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

```

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

```

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

```

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

```

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

```

#-----Water Supply Sources-----#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.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"

```

```

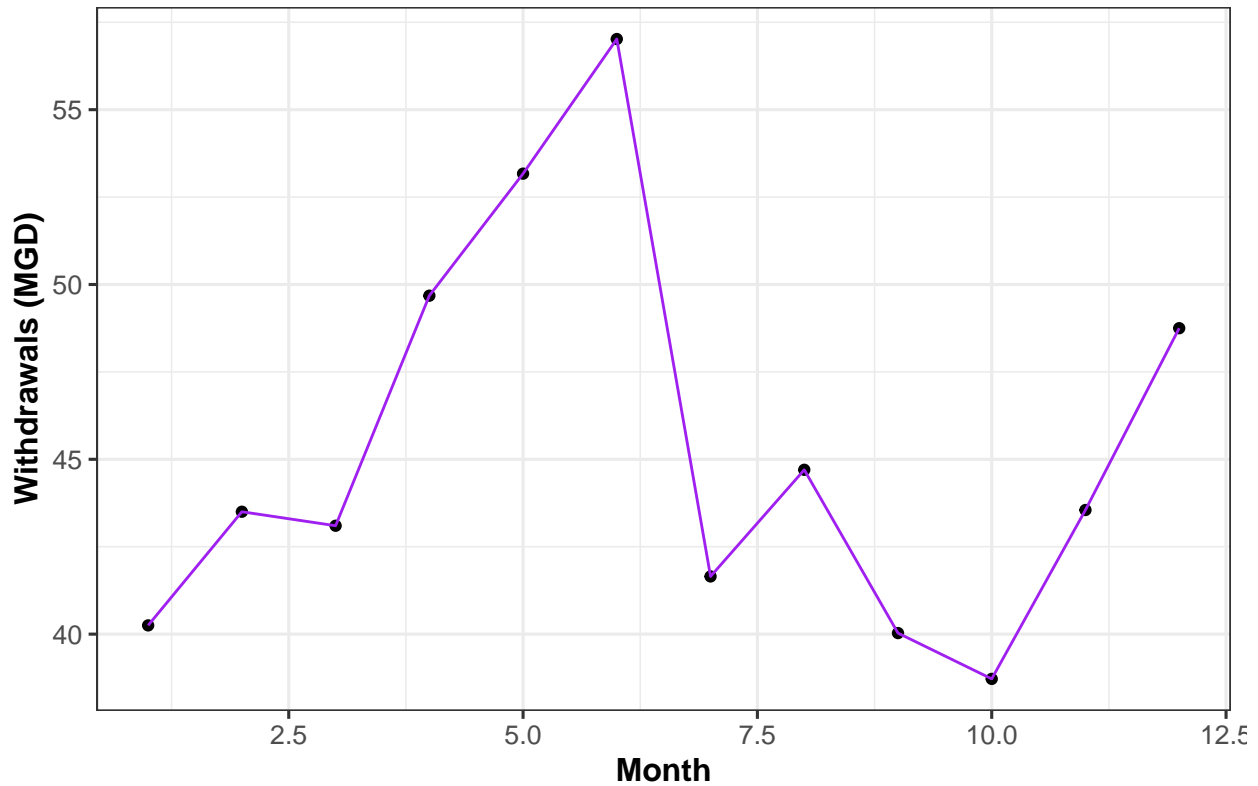
#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2015,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.dur <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

```

```
#-----Plot-----#
ggplot(monthly.withdrawals.dur, aes(y = max.withdrawals.mgd, x = Month)) +
  geom_point() +
  geom_line(color = "purple") +
  xlab("Month") +
  ylab("Withdrawals (MGD)") +
  ggtitle(paste("Monthly withdrawals for", water_system_name, "in", the_year))
```

## Monthly withdrawals for Durham in 2015



- 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
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2015"
scraped_url <- paste0(base_url, "=", the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#-----System Information-----#
water_system_name <- new_website %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
```

```
html_text()
water_system_name
```

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

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

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

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

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

```
#-----Water Supply Sources-----#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.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"
```

```
#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2015,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.ash <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#-----Combine datasets-----#
Durham.Asheville <- rbind(monthly.withdrawals.ash, monthly.withdrawals.dur)

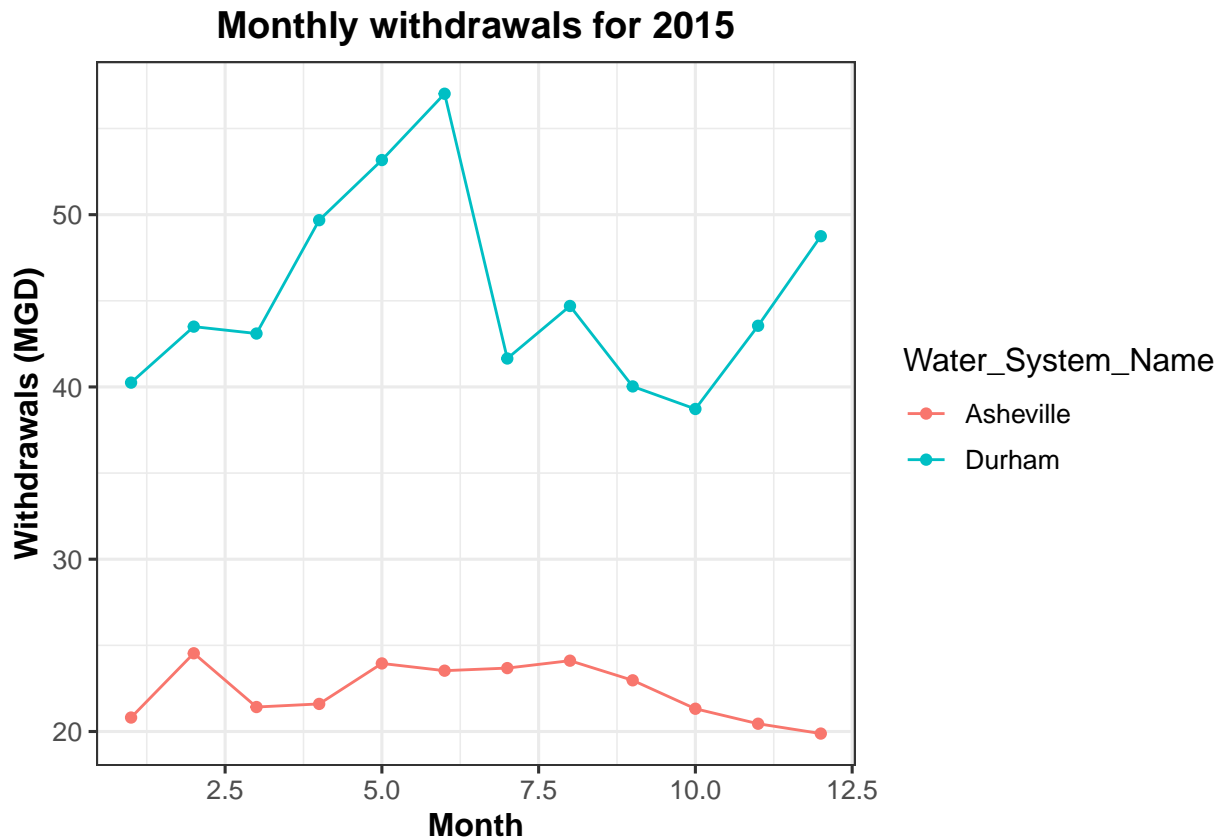
ggplot(Durham.Asheville, aes(y = max.withdrawals.mgd, x = Month,
```



```

                                color = Water_System_Name)) +
  geom_point() +
  geom_line() +
  xlab("Month") +
  ylab("Withdrawals (MGD)") +
  ggtitle(paste("Monthly withdrawals for 2015"))

```



- 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 "09\_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
#-----2010-----#

#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2010"
scraped_url <- paste0(base_url, "=", the_PWSID, "&year=", the_year)

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

```

```

#-----Scraped data elements from website-----#
#-----System Information-----#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name

## [1] "Asheville"

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

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

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

## [1] "Municipality"

#-----Water Supply Sources-----#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd

## [1] "21.8900" "20.9900" "22.4500" "19.9500" "22.5300" "21.4900" "19.7400"
## [8] "24.0100" "21.2300" "21.2500" "22.5000" "24.4300"

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2010,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.2010 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#----2011-----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"

```

```

the_PWSID <- "01-11-01"
the_year <- "2011"
scraped_url <- paste0(base_url, "=", the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name

```

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

```

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

```

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

```

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

```

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

```

#_____Water Supply Sources_____#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd

```

```

## [1] "21.4400" "23.3300" "23.5400" "23.8700" "23.7300" "22.5500" "20.2000"
## [8] "24.0400" "21.5300" "20.5800" "24.1800" "21.5100"

```

```

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2011, 12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.2011 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,

```

```

Ownership = !!ownership,
Date = my(paste(Month, "-", Year)))

#----2012----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?psid"
the_PWSID <- "01-11-01"
the_year <- "2012"
scraped_url <- paste0(base_url, "=", the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name

```

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

```

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

```

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

```

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

```

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

```

#_____Water Supply Sources_____#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd

```

```

## [1] "22.1700" "22.6300" "21.6900" "21.9000" "24.8200" "21.6700" "21.0600"
## [8] "23.8200" "20.8500" "21.5700" "23.0000" "20.4300"

```

```

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2012, 12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

```

```

#_2_____Modify dataframe_____#
monthly.withdrawals.2012 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#----2013----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?psid"
the_PWSID <- "01-11-01"
the_year <- "2013"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text()
water_system_name

```

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

```

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

```

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

```

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

```

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

```

#_____Water Supply Sources_____#
max.withdrawals.mgd <- new_website %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text()
max.withdrawals.mgd

```

```

## [1] "20.8400" "21.9500" "21.0400" "20.5300" "21.4600" "20.3400" "20.2800"
## [8] "21.4200" "19.8100" "20.9300" "21.2500" "19.6600"

```

```

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2013,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.2013 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#----2014----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2014"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text()
water_system_name

## [1] "Asheville"

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

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

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

## [1] "Municipality"

#_____Water Supply Sources_____#
max.withdrawals.mgd <- new_website %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text()
max.withdrawals.mgd

```

```
## [1] "22.6400" "21.3900" "20.9800" "21.2200" "21.8300" "20.7300" "19.8100"
## [8] "22.2000" "20.3300" "20.0800" "21.6600" "20.7800"
```

```
#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2014,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.2014 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#----2015----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2015"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text()
water_system_name
```

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

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

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

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

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

```
#-----Water Supply Sources-----#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.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"
```

```
#-----Withdrawals: Convert values into dataframe-----#
#_1-----match month to MGD-----#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2015,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))
```

```
#_2-----Modify dataframe-----#
monthly.withdrawals.2015 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))
```

```
#----2016----#
```

```
#-----Scraping Components-----#
```

```
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2016"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)
```

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

```
#-----Scraped data elements from website-----#
#-----System Information-----#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name
```

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

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

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



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

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

```
#-----Water Supply Sources-----#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd
```

```
## [1] "20.4300" "21.9900" "22.9500" "20.8700" "24.0800" "22.6200" "19.3500"
## [8] "22.8500" "22.4300" "21.0700" "22.3400" "21.9700"
```

```
#-----Withdrawals: Convert values into dataframe-----#
#_1-----match month to MGD-----#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2016,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))
```

```
#_2-----Modify dataframe-----#
monthly.withdrawals.2016 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))
```

```
#----2017----#
```

```
#-----Scraping Components-----#
```

```
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2017"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)
```

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

```
#-----Scraped data elements from website-----#
#-----System Information-----#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name
```

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

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

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

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

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

```
#-----Water Supply Sources-----#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd
```

```
## [1] "21.3100" "21.6200" "21.8700" "20.2800" "21.8500" "21.5700" "19.8000"
## [8] "22.5000" "20.0000" "20.4300" "22.8900" "20.5500"
```

```
#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2017,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.2017 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#----2018----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2018"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#-----System Information-----#
water_system_name <- new_website %>%
```

```
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name
```

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

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

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

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

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

```
#-----Water Supply Sources-----#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd
```

```
## [1] "23.8900" "21.9700" "23.8700" "20.0700" "22.4700" "21.6100" "19.7800"
## [8] "22.5400" "21.0500" "20.3100" "22.4700" "21.6200"
```

```
#-----Withdrawals: Convert values into dataframe-----#
#_1-----match month to MGD-----#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2018,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))
```

```
#_2-----Modify dataframe-----#
monthly.withdrawals.2018 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))
```

```
#----2019----
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
```

```

the_year <- "2019"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name

```

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

```

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

```

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

```

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

```

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

```

#_____Water Supply Sources_____#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd

```

```

## [1] "24.5100" "27.0900" "28.4500" "22.4600" "26.1000" "24.9900" "24.2500"
## [8] "26.1000" "25.0600" "25.2600" "26.2100" "24.1600"

```

```

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2019,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

#_2_____Modify dataframe_____#
monthly.withdrawals.2019 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,

```

```

    Date = my(paste(Month, "-", Year)))

#----2020----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?pwsid"
the_PWSID <- "01-11-01"
the_year <- "2020"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
html_text()
water_system_name

```

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

```

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

```

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

```

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

```

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

```

#_____Water Supply Sources_____#
max.withdrawals.mgd <- new_website %>%
html_nodes("th~ td+ td , th~ td+ td") %>%
html_text()
max.withdrawals.mgd

```

```

## [1] "23.7600" "23.2800" "23.8100" "22.0300" "23.4200" "22.7600" "21.9600"
## [8] "24.1500" "21.7500" "20.8400" "24.2700" "22.9600"

```

```

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                "Year" = rep(2020,12),
                                "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

```

```

#_2_____Modify dataframe_____#
monthly.withdrawals.2020 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#----2021----#
#-----Scraping Components-----#
base_url <- "https://www.ncwater.org/WUDC/app/LWSP/report.php?psid"
the_PWSID <- "01-11-01"
the_year <- "2021"
scraped_url <- paste0(base_url,"=",the_PWSID, "&year=", the_year)

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

#-----Scraped data elements from website-----#
#_____System Information_____#
water_system_name <- new_website %>%
  html_nodes("div+ table tr:nth-child(1) td:nth-child(2)") %>%
  html_text()
water_system_name

```

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

```

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

```

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

```

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

```

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

```

#_____Water Supply Sources_____#
max.withdrawals.mgd <- new_website %>%
  html_nodes("th~ td+ td , th~ td+ td") %>%
  html_text()
max.withdrawals.mgd

```

```

## [1] "22.2900" "24.2700" "24.7600" "21.8400" "26.0400" "24.3900" "21.7500"
## [8] "25.2900" "23.4000" "22.8100" "25.4200" "23.1100"

```

```

#-----Withdrawals: Convert values into dataframe-----#
#_1_____match month to MGD_____#
monthly.withdrawals <- data.frame("Month" = c(01, 05, 09, 02, 06, 10,
                                              03, 07, 11, 04, 08, 12),
                                  "Year" = rep(2021,12),
                                  "max.withdrawals.mgd" = as.numeric(max.withdrawals.mgd))

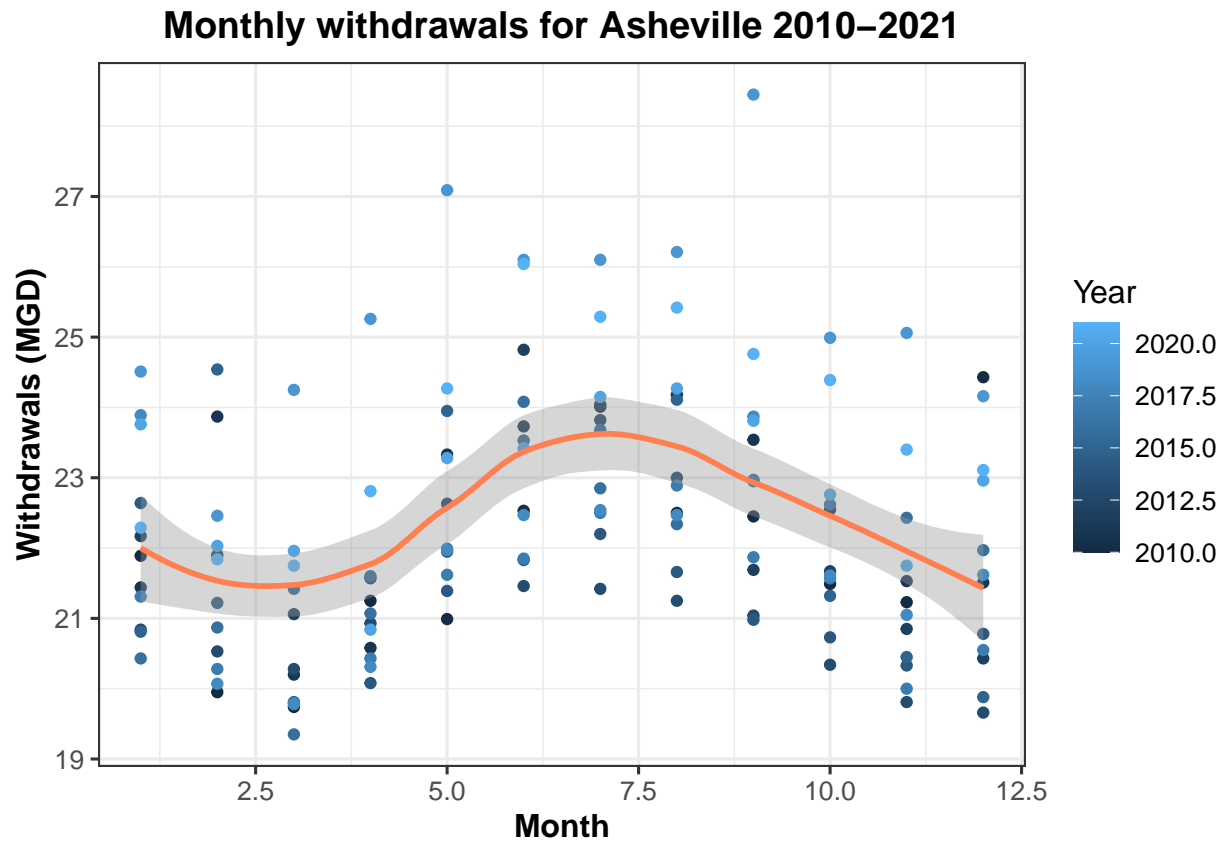
#_2_____Modify dataframe_____#
monthly.withdrawals.2021 <- monthly.withdrawals %>%
  arrange(Month) %>%
  mutate(Water_System_Name = !!water_system_name,
         PWSID = !!PWSID,
         Ownership = !!ownership,
         Date = my(paste(Month, "-", Year)))

#-----Combine datasets-----#
Asheville.2010.21 <- rbind(monthly.withdrawals.2010,
                           monthly.withdrawals.2011,
                           monthly.withdrawals.2012,
                           monthly.withdrawals.2013,
                           monthly.withdrawals.2014,
                           monthly.withdrawals.2015,
                           monthly.withdrawals.2016,
                           monthly.withdrawals.2017,
                           monthly.withdrawals.2018,
                           monthly.withdrawals.2019,
                           monthly.withdrawals.2020,
                           monthly.withdrawals.2021)

#-----plot asheville 2010-2021-----#
ggplot(Asheville.2010.21, aes(y = max.withdrawals.mgd, x = Month)) +
  geom_point(aes(color = Year)) +
  geom_smooth(method = "loess", color = "coral") +
  xlab("Month") +
  ylab("Withdrawals (MGD)") +
  ggtitle(paste("Monthly withdrawals for Asheville 2010-2021"))

```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



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

Yes, it appears as though the water usage for Asheville tends to peak during the summer months, and then gradually decrease into fall and winter. It also appears as though water demand is increasing over time.