HW 06: APIs and Web Scraping

David Gerard

2023-10-12

© 2023 David Gerard, not to be posted online.

Instructions

- Write your solutions in this starter file. You should modify the "author" field in the YAML header.
- Only commit R Markdown and HTML files (no PDF files). Make sure you have knitted to HTML for your final submission.
- **Make sure to commit each time you answer a question.**
- Only include the necessary code, not any extraneous code, to answer the questions.
- Learning objectives:
  - Obtain data from an API.
  - Scrape data from the web.

Open Brewery Database

Consider the Open Brewery Database API: https://www.openbrewerydb.org/

Only use `{httr2}`, not `{httr}`.

1. (1 pt) Get the list of all micro breweries in Ohio (163 of them) where we have longitude and latitude information (124 of them). These should be the microbreweries you obtained:

```
## 13 Below Brewery
## 2 Tones Brewing Co.
## Actual Brewing Company, LLC
## Akronym Brewing LLC
## Antiques on High
## Aqueduct Brewing
## Bad Tom Smith Brewing
## Bardwell Winery and Brewery
## Bascule Brewery And Public House
## BirdFish Brewing Co
## Black Frog Brewing Co
## BottleHouse Brewery
## Branch & Bone Artisan Ales
## Brausch Brewery
## Brew Kettle - Production Works
## BrewDog Brewing Company, LLC
## Brewery 33 Hocking Hills, LLC
## Brink Brewing Company
## Brick and Barrel
## Buck's Brewing Co
## Buckeye Lake Brewery
```

```
## CLE Brewing
## Clubhouse Brewing Company Ltd
## Commonhouse Ales
## Crooked Handle Brewing Co.
## DankHouse Brewing Company
## Dayton Beer Co Production Brewery & Bierhall
## Derive Brewing company
## Devil Wind Brewing LLC
## Devil's Kettle Brewing
## Double Edge Brewing Co
## Earnest Brew Works
## Elevator Brewing Co - Production facility
## Endeavor Brewing Company
## Eudora Brewing Co
## Fibonacci Brewing Company
## Fifty West Brewing Co
## FigLeaf Brewing Company
## Findlay Brewing Co.
## Flatrock Brewery
## Four String Brewing Company
## Franklin Brewing Co
## Fretboard Brewing Company
## Granville Brewing Company
## Great Black Swamp Brewing Co
## Gypsy Brewery
## Hasseman Brewing
## Headtrip Brewery
## Heavier Than Air Brewing Co
## Homestead Beer Co.
## Ignite Brewing Company, Ltd.
## Ill Mannered Brewing Company
## Jackie O's Brewery
## JAFB Brewery
## Lager Heads Brewing Co.
## Land-Grant Brewing Company
## Listermann Brewing Company
## Little Fish Brewing Company
## Loose Rail Brewing
## Madcap Brew Co.
## Magic City Brewing Company
## Maple Lawn Brewery
## Market Garden Brewery
## Market Garden Brewery
## McArthur's Brew House
## Moeller Brew Barn
## Mother Stewart's Brewing Co
## Multiple Brewing Company
## Municipal Brew Works
## Muskellunge Brewing Company
## Nocterra Brewing Co
## North High Brewing
## Northern Row Brewery & Distillery
## Off Track Brewing Company
## Ohio Brewing Company
```

```
## Old Dog Alehouse & Brewery
## Old Firehouse Brewery
## Olentangy River Brewing Company
## Paladin Brewing
## Paradigm Shift Brewing
## Patron Saints Brewery
## Pinups & Pints
## Platform Beer Co - Production Facility
## Platform Beer Co Taproom
## Platform Cincinnati
## Pretentious Barrel House
## R.Shea Brewing
## Railroad Brewing Company
## Random Precision Brewing Company
## Rivertown Brewing Co - Monroe Barrel House
## Rolling Mill Brewing Company
## Roundhouse Depot Brewing
## Royal Docks Brewhouse and Cannery
## Seventh Son Brewing Co
## Shale Brewing Co
## Sibling Revelry Brewing
## Sideswipe Brewing
## Sixth Sense Brewing Company
## Spires Social Brewing Co.
## Staas Brewing Company
## Star City Brewing Company
## Streetside Brewery
## Swine City Brewing Company
## Taft's Ale House
## Tailspin Brewing Co
## Terrestrial Brewing Company
## The BottleHouse Brewery And Meadery
## The Brew Mentor
## The Cleveland Brewery
## The Phoenix Brewing Company
## The Woodburn Brewery
## Thirsty Dog Brewing Company
## Three Points Urban Brewery
## Trek Brewing
## Two Monks Brewing Company
## Urban Artifact
## West Side Brewing
## Wiedemann Brewing Co, LLC
## Wild Side Brewing Company
## Wolf's Ridge Brewing
## Wooly Pig Farm Brewery
## Working Class Brewery
## Y Bridge Brewing Company
## Zaftig Brewing Co.
```

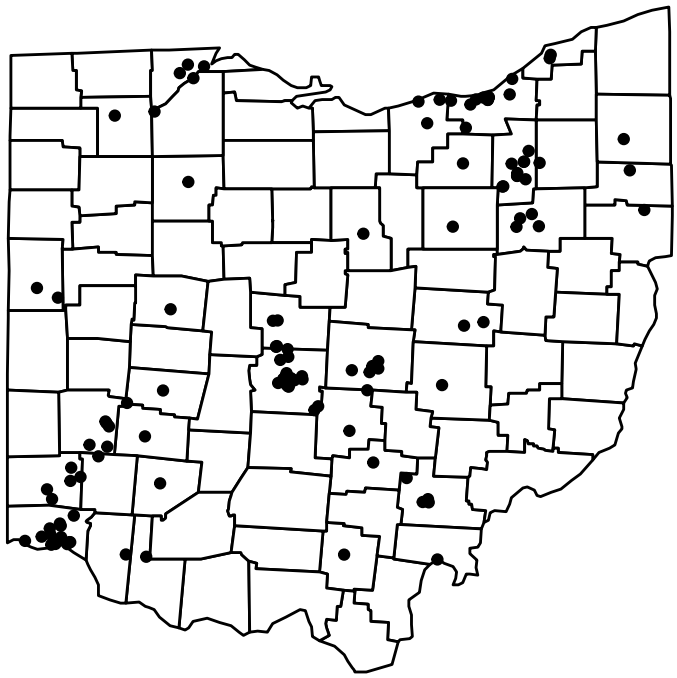
2. (1 pt) Clean up the data from part 1 to get the following data frame:

```
## # A tibble: 124 x 11
##    id        name  address_1 city  state_province postal_code longitude latitude
##    <chr>     <chr> <chr>     <chr> <chr>          <chr>           <dbl>    <dbl>
##  1 950180bd~ 13 B~ 7391 For~ Cinc~ Ohio           45233-1013      -84.7     39.1
##  2 836cb05e~ 2 To~ 4539 E B~ Whit~ Ohio           43213-1308      -82.9     40.0
##  3 ea7f7091~ Actu~ 655 N Ja~ Colu~ Ohio           43219-1837      -82.9     40.0
##  4 d4f4e76b~ Akro~ 58 E Mar~ Akron Ohio           44308           -81.5     41.1
##  5 1c215e2f~ Anti~ 714 S Hi~ Colu~ Ohio           43206           -83.0     39.9
##  6 fc2ce5e0~ Aque~ 529 Gran~ Akron Ohio           44311-1184      -81.5     41.1
##  7 ccb3350f~ Bad ~ 4720 Eas~ Cinc~ Ohio           45226-1893      -84.4     39.1
##  8 7574d2fa~ Bard~ 716 N Hi~ Moun~ Ohio           45154           -83.9     39.0
##  9 d73def5b~ Basc~ 1397 Col~ Lora~ Ohio           44052-3377      -82.2     41.5
## 10 55cf083b~ Bird~ 16 S Mai~ Colu~ Ohio           44408-1348      -80.7     40.9
## # i 114 more rows
## # i 3 more variables: phone <chr>, website_url <chr>, street <chr>
```

3. (1 pt) Edit the following ggplot code to obtain the following plot:

```
library(maps)
countymap <- map_data("county")

countymap |>
  filter(region == "ohio") |>
  ggplot(aes(x = long, y = lat, group = subregion)) +
  geom_polygon(fill = "white", color = "black")
```



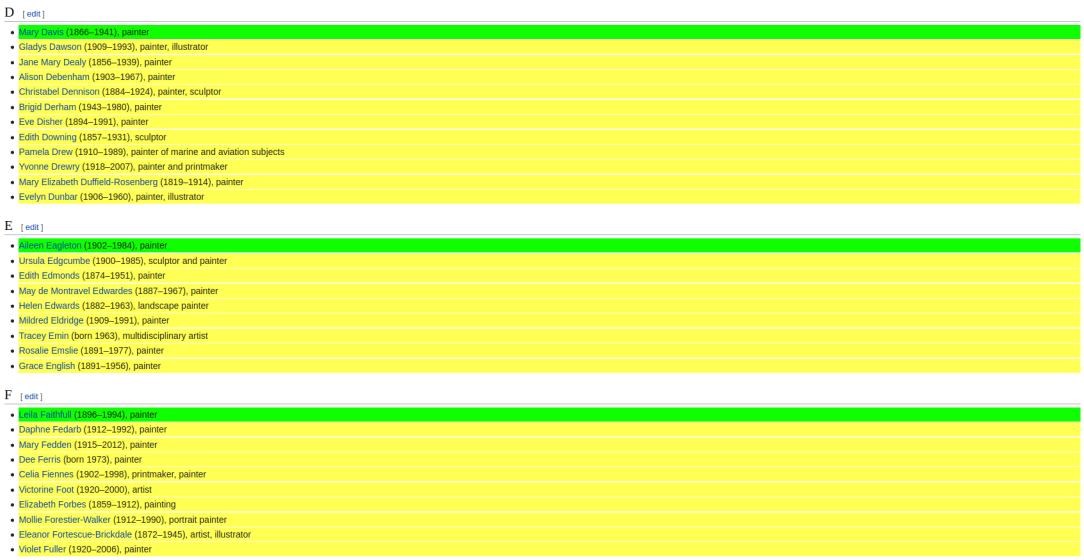
English women artists

Consider the copied Wikipedia page on English women artists: https://data-science-master.github.io/lectures/data/engart.html

We'll use the copied version on GitHub rather than the original version to make sure solutions are consistent. But the original version can be found here (but don't use it): https://en.wikipedia.org/wiki/List_of_English_women_artists

Use rvest to answer the following questions.

1. (1 pt) Download the html file and save the output to a variable.
2. (1 pt) Use SelectorGadget to extract each woman, years of life, and artistic medium. For example, my selections looked like this:



D [edit]

- Mary Davis (1866–1941), painter
- Gladys Dawson (1909–1993), painter, illustrator
- Jane Mary Dealy (1856–1939), painter
- Alison Debenham (1903–1967), painter
- Christabel Dennison (1884–1924), painter, sculptor
- Brigid Derham (1943–1980), painter
- Eve Disher (1894–1991), painter
- Edith Downing (1857–1931), sculptor
- Pamela Drew (1910–1989), painter of marine and aviation subjects
- Yvonne Drewry (1918–2007), painter and printmaker
- Mary Elizabeth Duffield-Rosenberg (1819–1914), painter
- Evelyn Dunbar (1906–1960), painter, illustrator

E [edit]

- Aileen Eagleton (1902–1984), painter
- Ursula Edgcumbe (1900–1985), sculptor and painter
- Edith Edmonds (1874–1951), painter
- May de Montravel Edwardes (1887–1967), painter
- Helen Edwards (1882–1963), landscape painter
- Mildred Eldridge (1909–1991), painter
- Tracey Emin (born 1963), multidisciplinary artist
- Rosalie Emslie (1891–1977), painter
- Grace English (1891–1956), painter

F [edit]

- Leila Faithfull (1896–1994), painter
- Daphne Fedarb (1912–1992), painter
- Mary Fedden (1915–2012), painter
- Dee Ferris (born 1973), painter
- Celia Fiennes (1902–1998), printmaker, painter
- Victorine Foot (1920–2000), artist
- Elizabeth Forbes (1859–1912), painting
- Mollie Forestier-Walker (1912–1990), portrait painter
- Eleanor Fortescue-Brickdale (1872–1945), artist, illustrator
- Violet Fuller (1920–2006), painter

3. (2 pts) Clean the data.

   Hints:

   1. Not all year ranges are of the form (Birth-Death). You should place NA's in the appropriate locations. No need to extract third-party datasets to obtain the true values.
   2. Be careful of parsing numbers like `"c.1888"`.
   3. Parentheses are used more than just to delimit years.
   4. Painters, sculptors, illustrators, and printmakers are the most common types of artists, so I included indicators for those mediums. Note that not all printmakers are called "printmakers".

   Your final data frame should look like this:

```
## # A tibble: 294 x 8
##    artist            birth death mediums painter sculptor illustrator printmaker
##    <chr>             <dbl> <dbl> <chr>   <lgl>   <lgl>    <lgl>       <lgl>
##  1 Evelyn Abelson     1886  1967 painter TRUE    FALSE    FALSE       FALSE
##  2 Ruth Abrahams      1931    NA painte~ TRUE    FALSE    TRUE        FALSE
##  3 Judith Ackland     1892  1971 landsc~ TRUE    FALSE    FALSE       FALSE
##  4 Elinor Proby Ada~  1885  1945 painter TRUE    FALSE    FALSE       FALSE
##  5 Sarah Gough Adam~  1888  1963 painter TRUE    FALSE    FALSE       FALSE
##  6 Marion Adnams      1898  1995 painte~ TRUE    FALSE    FALSE       TRUE
##  7 Mary Adshead       1904  1995 painte~ TRUE    FALSE    TRUE        FALSE
##  8 Eileen Agar        1899  1991 painte~ TRUE    FALSE    FALSE       FALSE
##  9 Sam Ainsley        1950    NA painte~ TRUE    FALSE    FALSE       FALSE
## 10 Eileen Aldridge    1916  1990 painter TRUE    FALSE    FALSE       FALSE
## # i 284 more rows
```

4. (1 pt) Tabulate the number of painters, sculptors, illustrators, and printmakers. You should get these numbers:

```
## # A tibble: 1 x 4
##   painters_n sculptor_n illustrator_n printmaker_n
##        <int>      <int>         <int>        <int>
## 1        234         26            22           15
```

5. (1 pt) Plot the lifespans of printmakers. Your plot should look like this: