HW 12

David Gerard

2020-04-14

Instructions

- Write your solutions in this starter file. You should modify the "author" field in the YAML header.
- Do not modify the paths of any files.
- Only commit R Markdown and HTML files (no PDF or Word files). Make sure you have knitted to HTML for your final submission.
- **Make sure to commit each time you answer a question.** Lack of informative and frequent commits will result in point deductions.
- 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/

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

```
##  [1] "Brick and Barrel"
##  [2] "Dayton Beer Co Production Brewery & Bierhall"
##  [3] "Fibonacci Brewing Company"
##  [4] "Granville Brewing Company"
##  [5] "Magic City Brewing Company"
##  [6] "Muskellunge Brewing Company"
##  [7] "Paradigm Shift Brewing"
##  [8] "Random Precision Brewing Company"
##  [9] "Swine City Brewing Company"
## [10] "The Woodburn Brewery"
## [11] "Two Monks Brewing Company"
## [12] "Wooly Pig Farm Brewery"
## [13] "Black Frog Brewing Co"
## [14] "Double Edge Brewing Co"
## [15] "Mother Stewart's Brewing Co"
## [16] "The Phoenix Brewing Company"
```

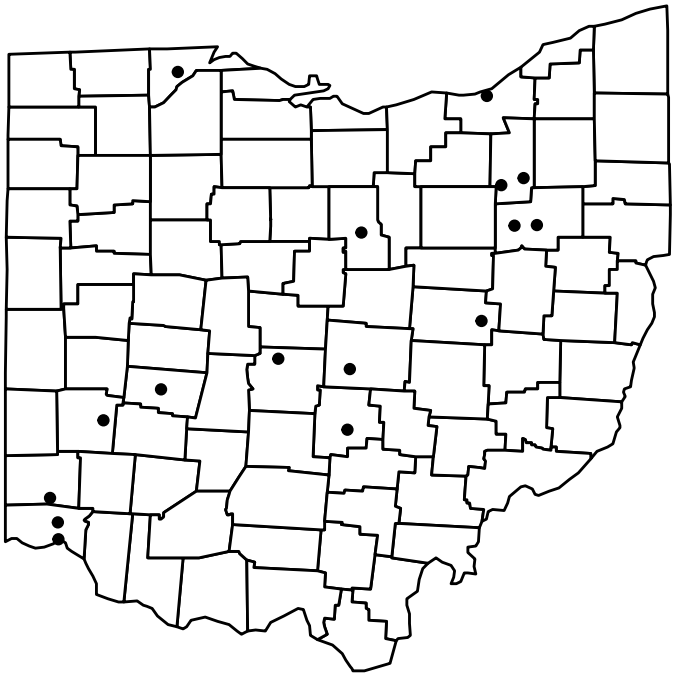
2. Clean up the data from part 1 to get the following data frame:

```
## # A tibble: 16 x 10
##    id    name  street city  postal_code longitude latitude phone website_url
##    <chr> <chr> <chr>  <chr> <chr>           <dbl>    <dbl> <chr> <chr>
##  1 5383  Bric~ 1844 ~ Clev~ 44113-2412      -81.7     41.5 "503~ "http://ww~
##  2 5408  Dayt~ 41 Ma~ Dayt~ 45402-2105      -84.2     39.8 "937~ "http://ww~
##  3 5427  Fibo~ 1445 ~ Cinc~ 45231-3559      -84.5     39.2 "513~ "http://fi~
##  4 5445  Gran~ 5371 ~ Gran~ 43023-9192      -82.6     40.0 ""    "http://ww~
##  5 5496  Magi~ 161 2~ Barb~ 44203           -81.6     41.0 "330~ "http://Ww~
##  6 5521  Musk~ 425 5~ Cant~ 44702           -81.4     40.8 ""    "http://ww~
##  7 5540  Para~ 128 N~ Mass~ 44646-5526      -81.5     40.8 "330~ "http://ww~
##  8 5562  Rand~ 2365 ~ Colu~ 43235-2700      -83.1     40.1 "614~ "http://ww~
##  9 5597  Swin~ 4614 ~ Fair~ 45014-1923      -84.5     39.3 "704~ "http://ww~
## 10 5617  The ~ 2800 ~ Cinc~ 45206-1793      -84.5     39.1 "513~ "http://ww~
## 11 5627  Two ~ 352 M~ Akron 44312-2021      -81.5     41.1 "234~ "http://ww~
## 12 5645  Wool~ 23631~ Fres~ 43824-9704      -81.7     40.3 "740~ "http://ww~
## 13 5357  Blac~ 831 S~ Holl~ 43528-8746      -83.7     41.6 "419~ "http://ww~
## 14 5413  Doub~ 158 W~ Lanc~ 43130-4308      -82.6     39.7 "740~ "http://ww~
## 15 5516  Moth~ 109 W~ Spri~ 45504-2546      -83.8     39.9 ""    ""
## 16 5615  The ~ 131 N~ Mans~ 44902-1331      -82.5     40.8 "419~ "http://ww~
## # ... with 1 more variable: updated_at <chr>
```

3. 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")
```



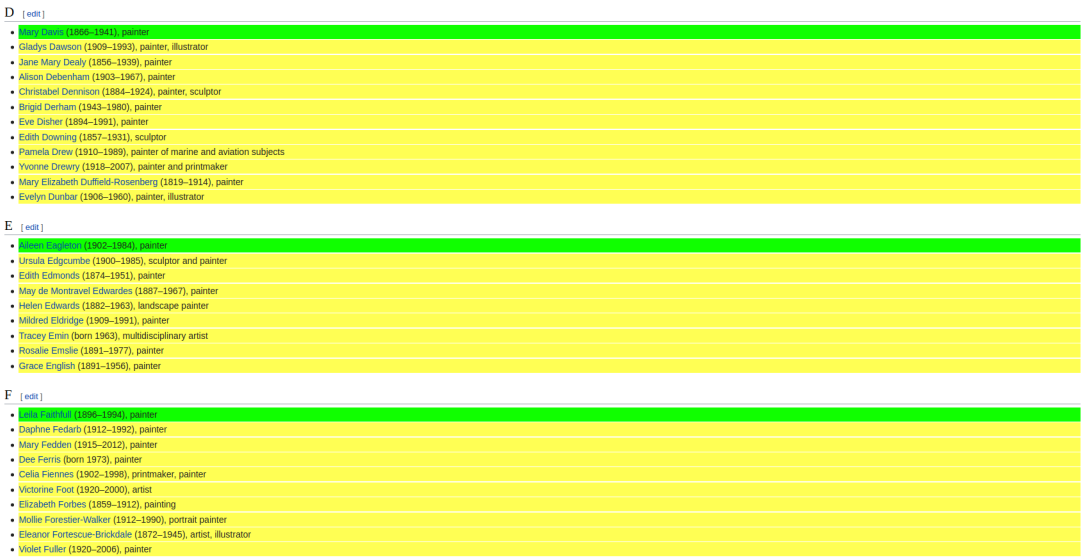
English women artists

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. Download the html file and save the output to a variable.
2. 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. 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 Ab~  1886  1967 painter        TRUE    FALSE    FALSE       FALSE
##  2 Ruth Abra~  1931    NA painter illus~ TRUE    FALSE    TRUE        FALSE
##  3 Judith Ac~  1892  1971 landscape pai~ TRUE    FALSE    FALSE       FALSE
##  4 Elinor Pr~  1885  1945 painter        TRUE    FALSE    FALSE       FALSE
##  5 Sarah Gou~  1888  1963 painter        TRUE    FALSE    FALSE       FALSE
##  6 Marion Ad~  1898  1995 painter print~ TRUE    FALSE    FALSE       TRUE
##  7 Mary Adsh~  1904  1995 painter illus~ TRUE    FALSE    TRUE        FALSE
##  8 Eileen Ag~  1899  1991 painter and p~ TRUE    FALSE    FALSE       FALSE
##  9 Sam Ainsl~  1950    NA painter and t~ TRUE    FALSE    FALSE       FALSE
## 10 Eileen Al~  1916  1990 painter        TRUE    FALSE    FALSE       FALSE
## # ... with 284 more rows
```

4. 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. Plot the lifespans of printmakers. Your plot should look like this: