Natl Gallery Art Web Scrape II

November 2, 2023

0.1 National Gallery of Art Web Scrape

```
[9]:  # import libraries
      from bs4 import BeautifulSoup
      import requests
      import smtplib
      import time
      import datetime
```

0.1.1 Read web page

```
[57]: page = requests.get('https://web.archive.org/web/20161025170331/http://www.nga.
      ↪gov/collection/anA1.htm')

      soup = BeautifulSoup(page.text, 'html.parser')
```

```
[ ]:  ## Parse html to obtain values
```

```
[19]: # Pull all text from the BodyText div
      artist_name_list = soup.find(class_='BodyText')

      # Pull text from all instances of <a> tag within BodyText div
      artist_name_list_items = artist_name_list.find_all('a')

      # Create for loop to print out all artists' names
      for artist_name in artist_name_list_items:
          print(artist_name.prettify())
```

```
<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=17840">
 A. B.
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=5566">
 Aachen, Hans von
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=10278">
 Abakanowicz, Magdalena
```

```
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=3474">
 Abate, Niccolò dell'
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=3474">
 Abbate, Niccolò dell'
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=21155">
 Abbati, Pietro Giovanni
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=19745">
 Abbatini, Guido Ubaldo
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2855">
 Abbe, Elfriede Martha
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2856">
 Abbé, Salomon van
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33127">
 Abbe, William Parker
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2337">
 Abbey, Edwin Austin
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=14725">
 Abbiati, Giuseppe
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=40407">
 Abbot, John
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=13553">
 Abbott, Berenice
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=10979">
 Abbott, John White
```

```
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=22">
 Abbott, Lemuel Francis
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2857">
 Abe, Nobuya
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34210">
 Abel, Mike
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34210">
 Abel, Myer
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33106">
 Abelman, Ida York
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2338">
 Abels, Jacobus Theodorus
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=17830">
 Abel-Truchet
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=7194">
 Aberdeen, Harry G.
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2006">
 Abondio, Antonio
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2858">
 Abraham, Tancrède
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=12430">
 Abrahami, Elie
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33107">
 Abramovitz, Albert
```

```
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34328">
 Abramowitz, Ben
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34328">
 Abramowitz, Benjamin
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA2.htm">
 Ami-And
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA2.htm">
 Abr-Ada
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA3.htm">
 And-And
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA3.htm">
 Ada-Aft
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA4.htm">
 And-Ann
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA4.htm">
 Aga-Alb
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA5.htm">
 Alb-Alb
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA5.htm">
 Ann-Ano
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA6.htm">
 Ano-Ano
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA6.htm">
 Alb-All
```

```
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA7.htm">
 Ano-Ano
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA7.htm">
 All-Alm
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA8.htm">
 Ano-Ano
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA8.htm">
 Alm-Alt
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA9.htm">
 Ano-Ano
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA10.htm">
 Ano-Ano
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA11.htm">
 Ano-Ant
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA12.htm">
 Ant-App
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA13.htm">
 App-Arn
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA14.htm">
 Arn-Art
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA15.htm">
 Asa-Atk
</a>

<a href="/web/20161025170331/http://www.nga.gov/collection/anA16.htm">
 Atk-Aud
```

```
 </a>

 <a href="/web/20161025170331/http://www.nga.gov/collection/anA17.htm">
  Aud-Aus
 </a>

 <a href="/web/20161025170331/http://www.nga.gov/collection/anA18.htm">
  Aus-Ayl
 </a>

 <a href="/web/20161025170331/http://www.nga.gov/collection/anA19.htm">
  Ayr-Azz
 </a>

 <a href="/web/20161025170331/http://www.nga.gov/collection/anA26.htm">
  <strong>
   next
   <br/>
   page
  </strong>
 </a>
```

```python
[22]:  # Remove bottom links
       last_links = soup.find(class_='AlphaNav')
       last_links.decompose()

       artist_name_list = soup.find(class_='BodyText')
       artist_name_list_items = artist_name_list.find_all('a')

       for artist_name in artist_name_list_items:
           print(artist_name.prettify())
```

```
<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=17840">
 A. B.
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=5566">
 Aachen, Hans von
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=10278">
 Abakanowicz, Magdalena
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=3474">
 Abate, Niccolò dell'
</a>
```

```
<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=3474">
 Abbate, Niccolò dell'
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=21155">
 Abbati, Pietro Giovanni
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=19745">
 Abbatini, Guido Ubaldo
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2855">
 Abbe, Elfriede Martha
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2856">
 Abbé, Salomon van
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33127">
 Abbe, William Parker
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2337">
 Abbey, Edwin Austin
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=14725">
 Abbiati, Giuseppe
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=40407">
 Abbot, John
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=13553">
 Abbott, Berenice
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=10979">
 Abbott, John White
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=22">
 Abbott, Lemuel Francis
</a>
```

```
<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2857">
 Abe, Nobuya
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34210">
 Abel, Mike
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34210">
 Abel, Myer
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33106">
 Abelman, Ida York
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2338">
 Abels, Jacobus Theodorus
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=17830">
 Abel-Truchet
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=7194">
 Aberdeen, Harry G.
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2006">
 Abondio, Antonio
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2858">
 Abraham, Tancrède
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=12430">
 Abrahami, Elie
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33107">
 Abramovitz, Albert
</a>

<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34328">
 Abramowitz, Ben
</a>
```

```
<a href="/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34328">
 Abramowitz, Benjamin
</a>
```

[59]:
```python
# Initialize
Artist_nm = []
Artist_lnk = []

# Read artist name and get link reference
# Append names and links to Series
# Print names and links
for artist_name in artist_name_list_items:
    names = artist_name.contents[0]
    links = 'https://web.archive.org' + artist_name.get('href')
    Artist_nm.append(names)
    Artist_lnk.append(links)
    print(names)
    print(links)
```

```
A. B.
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=17840
Aachen, Hans von
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=5566
Abakanowicz, Magdalena
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=10278
Abate, Niccolò dell'
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=3474
Abbate, Niccolò dell'
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=3474
Abbati, Pietro Giovanni
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=21155
Abbatini, Guido Ubaldo
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=19745
Abbe, Elfriede Martha
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=2855
Abbé, Salomon van
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=2856
```

Abbe, William Parker
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33127
Abbey, Edwin Austin
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2337
Abbiati, Giuseppe
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=14725
Abbot, John
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=40407
Abbott, Berenice
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=13553
Abbott, John White
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=10979
Abbott, Lemuel Francis
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=22
Abe, Nobuya
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2857
Abel, Mike
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34210
Abel, Myer
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=34210
Abelman, Ida York
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=33106
Abels, Jacobus Theodorus
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2338
Abel-Truchet
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=17830
Aberdeen, Harry G.
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=7194
Abondio, Antonio
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2006
Abraham, Tancrède
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-bin/tsearch?artistid=2858

Abrahami, Elie
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=12430
Abramovitz, Albert
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=33107
Abramowitz, Ben
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=34328
Abramowitz, Benjamin
https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=34328

0.1.2 Create Dataframe

```
[44]: import pandas as pd
      # set max column width
      pd.set_option('display.max_colwidth', None)

      # Create dataframe
      Gallery_artists = pd.DataFrame({'Name':Artist_nm, 'Link':Artist_lnk})
      Gallery_artists
```

```
[44]:                         Name  \
      0                        A. B.
      1           Aachen, Hans von
      2       Abakanowicz, Magdalena
      3           Abate, Niccolò dell'
      4          Abbate, Niccolò dell'
      5      Abbati, Pietro Giovanni
      6      Abbatini, Guido Ubaldo
      7       Abbe, Elfriede Martha
      8           Abbé, Salomon van
      9        Abbe, William Parker
      10        Abbey, Edwin Austin
      11         Abbiati, Giuseppe
      12                Abbot, John
      13           Abbott, Berenice
      14         Abbott, John White
      15     Abbott, Lemuel Francis
      16               Abe, Nobuya
      17                Abel, Mike
      18                Abel, Myer
      19          Abelman, Ida York
      20   Abels, Jacobus Theodorus
      21              Abel-Truchet
      22          Aberdeen, Harry G.
```

```
23           Abondio, Antonio
24         Abraham, Tancrède
25            Abrahami, Elie
26        Abramovitz, Albert
27             Abramowitz, Ben
28       Abramowitz, Benjamin

          Link
0   https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=17840
1     https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=5566
2    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=10278
3     https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=3474
4     https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=3474
5    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=21155
6    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=19745
7     https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=2855
8     https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=2856
9    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=33127
10    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=2337
11  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=14725
12  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=40407
13  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=13553
14  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=10979
15     https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=22
16    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=2857
17  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=34210
18  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
bin/tsearch?artistid=34210
19  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
```

```
    bin/tsearch?artistid=33106
20    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=2338
21  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=17830
22    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=7194
23    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=2006
24    https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=2858
25  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=12430
26  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=33107
27  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=34328
28  https://web.archive.org/web/20161025170331/http://www.nga.gov/cgi-
    bin/tsearch?artistid=34328
```

[38]:
```python
Gallery_artists.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29 entries, 0 to 28
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Name    29 non-null     object
 1   Link    29 non-null     object
dtypes: object(2)
memory usage: 596.0+ bytes
```

0.1.3 Create CSV File

[46]:
```python
# Create csv file
Gallery_artists.to_csv('Gallery_Artists.csv',index=False, encoding='UTF-8')
```

[ ]: