

Introduction to Data Science – Week 5

NETA LIVNEH

COURSE 55793

2019-2020

Last Week Recap

Data types:

- Sets
- Dictionaries

Looping through containers

Strings Formats

We encountered the functions: `format()`, `eval()`, `zip()`, `enumerate()`



This Week

Sorting by different keys

Importing libraries

Using built-in and third-party libraries:

- Random
- Requests
- Pathlib

A few words about web pages



Sorted Function

The built-in function `sorted(iterable, key, reverse)` builds a new sorted list from an iterable.

It has two optional parameters: `key`, receives a function that determines the comparison key, and `reverse`, which sorts the list in a reversed order.

```
In [140]: sorted([5, 2, 3, 1, 4])
```

```
Out[140]: [1, 2, 3, 4, 5]
```

```
In [141]: sorted({1: 'D', 2: 'B', 3: 'B', 4: 'E', 5: 'A'})
```

```
Out[141]: [1, 2, 3, 4, 5]
```

```
In [142]: sorted("This is a test string from Andrew".split(), key=str.lower)
```

```
['a', 'Andrew', 'from', 'is', 'string', 'test', 'This']
```

```
In [143]: student_tuples = [('john', 'A', 15), ('jane', 'B', 12), ('dave', 'B', 10), ]
```

```
In [144]: sorted(student_tuples, key=lambda student: student[2]) # sort by age
```

```
Out[144]: [('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```



Import Non Built-in Libraries

Python contains many functions, but not all of them are immediately available as built-in functions. Instead of being available as built-ins, some functions are saved in different modules. A *module* is a file containing function definitions and other statements.

To use a non build-in function or a module we use *import*.

For example:

```
import random  
  
dir(random) # Print a list of the module's attributes and methods.  
  
import numpy as np  
  
from math import cos, sin, pi  
  
From time import *
```



How to install a new package (or update)

Use the Anaconda Navigator or Pycharm

From the command line:

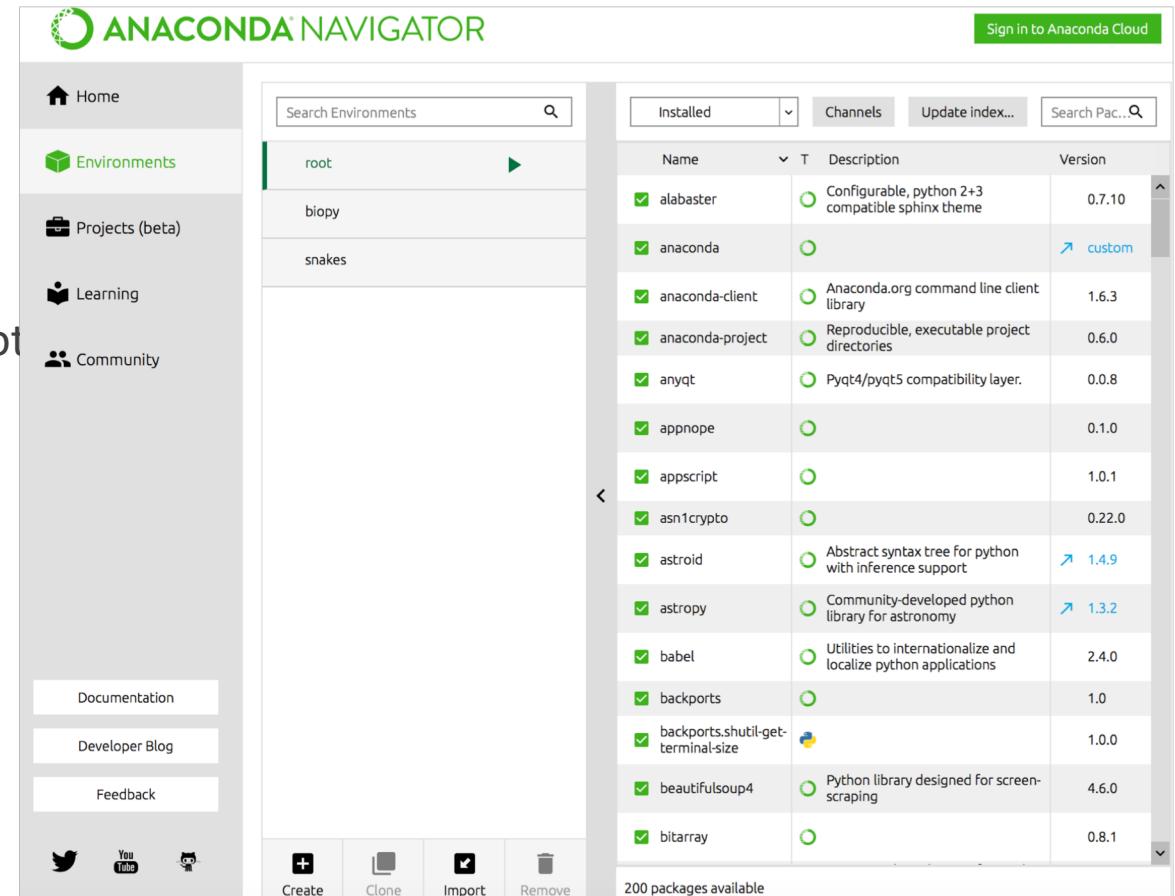
- `conda search packagename`
- Example: `conda install scipy`

If you are not using a naconda or the package is not supported:

- `pip install scipy`

To update:

- `conda update conda`
- `conda update python`
- `pip install scipy -u` (depends on the platform)



(pseudo-)Random numbers

Why?

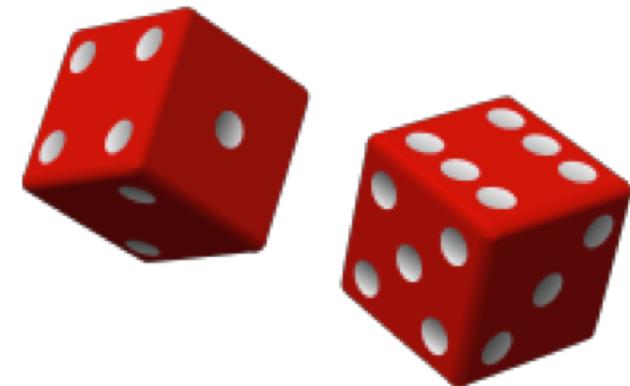
- Shuffle data
- Generate random sequences, stochastic noise
- Emulate unpredictable behavior (e.g. simulation of agents)
- Cryptography

Computers nearly never generate random numbers

Sequences of random number can be predictable (determined by seed) – can be good or bad

Random numbers are relatively expensive

Out of the box: uniform, normal (Gaussian), lognormal, negative exponential, gamma, and beta distributions



(pseudo-)Random numbers

Random number generators reside in `random.py` library (comes with python). It is the main library used when trying to generate numbers.

```
In [155]: import random
```

```
In [156]: random.randint(1, 20)  
Out[156]: 8
```

```
In [157]: random.randint(1, 20)  
Out[157]: 10
```

```
In [158]: random.randint(10,20)  
Out[158]: 11
```

```
In [159]: random.randint(10,20)  
Out[159]: 19
```

```
In [160]: random.choice(['to be', 'not to be'])  
Out[160]: 'not to be'
```

```
In [161]: random.choice('abcdefg')  
Out[161]: 'c'
```



(pseudo-)Random numbers

Method	Description
random.seed(a)	Initialize the random number generator.
random.getrandbits(<i>k</i>)	Returns a Python integer with <i>k</i> random bits.
random.range(start, stop[, step])	Return a randomly selected element from range.
random.randint(a, b)	Return a random integer <i>N</i> such that <i>a</i> <= <i>N</i> <= <i>b</i> .
random.choice(seq)	Return a random element from the non-empty sequence <i>seq</i> .
random.shuffle(seq)	Shuffle the sequence <i>seq</i> in place.
random.sample(seq, <i>k</i>)	Return a <i>k</i> length list of unique elements chosen from the population sequence or set. Used for random sampling without replacement.
random.random()	Return the next random floating point number in the range [0.0, 1.0).
random.uniform(a, b)	Return a random floating point number <i>N</i> such that <i>a</i> <= <i>N</i> <= <i>b</i> .
random.normalvariate(mu, sigma)	Normal distribution. <i>mu</i> is the mean, and <i>sigma</i> is the standard deviation.



An introduction to web pages

The screenshot shows a web browser window with the title "Madeleine BRES" in the tab bar. The address bar displays the URL "scientificwomen.net/women/bres-madeleine-145". The page content is titled "HISTORY OF SCIENTIFIC WOMEN". On the left, there is a sidebar with a red hexagonal background and white text links: "MAINPAGE" (with a red atom icon), "A-Z LIST", "NOBEL PRIZE and Co.", "BY PERIOD" (with sub-links "Antiquity", "Middle ages", "16th century", "17th century", "18th century", "19th century", "20th century"). To the right of the sidebar is a portrait of a woman, identified as Madeleine BRES. The main text area contains her name, "19th century", "Fields: Medicine", "Born: 1842", "Death: 1921", and "Main achievements: First French woman to obtain a PhD in medicine". A search bar with a red "SEARCH" button is located at the top right. The bottom of the page has a blue footer with the text "Born Madeleine Gebelin, she told in the Medical Chronicle on 1 April 1895 how her medical career was born - "I was barely eight years old when my".

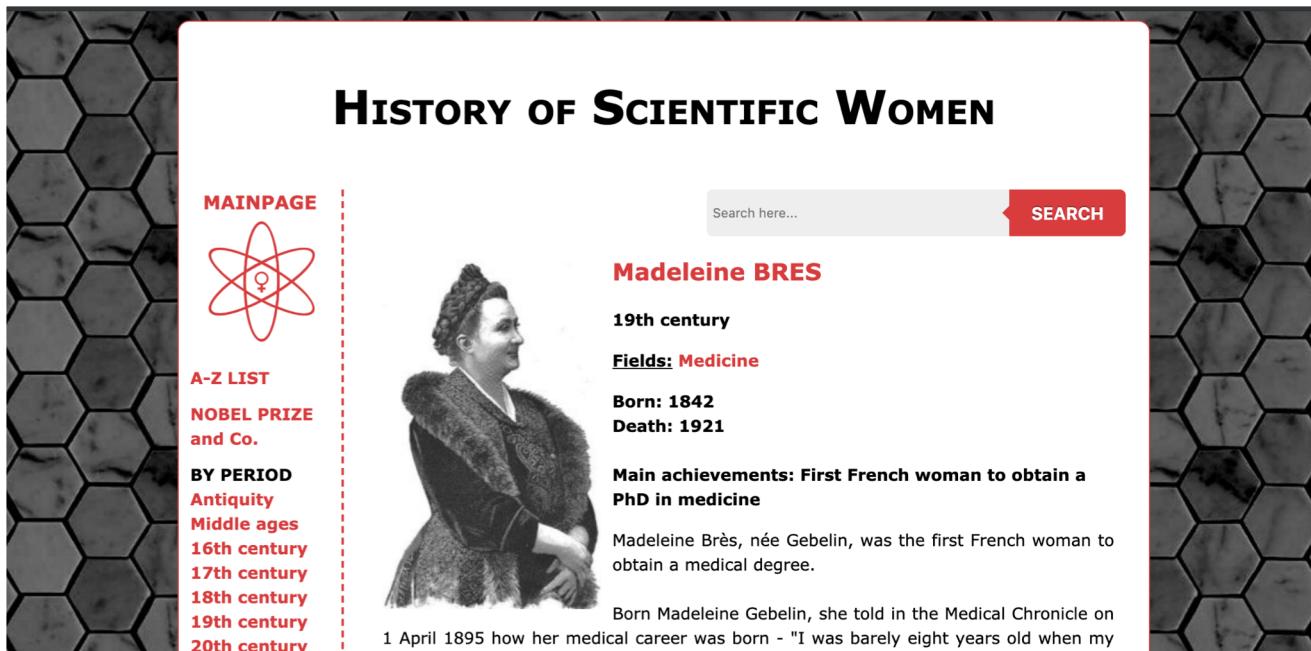


An introduction to web pages

Web address (url):

 scientificwomen.net/women/bres-madeleine-145

Page content: including anything on the page (even background)



The screenshot shows a website with a dark hexagonal background pattern. The main content area has a white background. At the top center, the title "HISTORY OF SCIENTIFIC WOMEN" is displayed in bold black capital letters. To the left, there is a sidebar with red text links: "MAINPAGE" (with a red atom icon), "A-Z LIST", "NOBEL PRIZE and Co.", "BY PERIOD" (with sub-links: "Antiquity", "Middle ages", "16th century", "17th century", "18th century", "19th century", "20th century"). In the center, there is a search bar with a placeholder "Search here..." and a red "SEARCH" button. Below the search bar is a portrait of a woman, identified as Madeleine BRES. To the right of the portrait, her name is in red, followed by her era ("19th century"), fields ("Fields: Medicine"), birth and death dates ("Born: 1842", "Death: 1921"), and main achievements ("Main achievements: First French woman to obtain a PhD in medicine"). A short biography notes she was the first French woman to obtain a medical degree. At the bottom, a caption reads: "Born Madeleine Gebelin, she told in the Medical Chronicle on 1 April 1895 how her medical career was born - 'I was barely eight years old when my".



```

1<!DOCTYPE html>
2<html lang="en">
3<head>
4    <meta charset="UTF-8">
5    <title>Madeleine&nbsp;BRES</title>
6    <base href="/">
7    <link href="css/style.css" rel="stylesheet" type="text/css" />
8    <link rel="icon" type="image/ico" href="img/favicon.ico" />
9    <meta name="viewport" content="width=device-width, initial-scale=1.0, maximum-scale=1.0" />
10</head>
11<body>
12<div id="main">
13    <div id="title"><h1>History of Scientific Women</h1></div>
14
15    <form class="search" method="post">
16        <input type="text" name="search" placeholder="Search here..." required>
17        <button type="submit" name="sub_search">Search</button>
18    </form>
19
20    <div id="menu">
21        <div id="logo">
22            <a href="mainpage">MAINPAGE
23            </a>
24        </div>
25        <ul>
26            <li><a href="list">A-Z LIST</a></li>
27            <li><a href="nobel">NOBEL PRIZE and Co.</a></li>
28            <li>BY PERIOD<br>
29                <a href="period/antiquity-1">Antiquity</a><br><a href='period/middle_ages-2'>Middle ages</a><br><a href='period/16th_century-3'>16th century</a><br><a href='period/17th_century-4'>17th century</a><br><a href='period/18th_century-5'>18th century</a><br><a href='period/19th_century-6'>19th century</a><br><a href='period/20th_century-7'>20th century</a><br><a href='period/21st_century-8'>21st century</a><br>           </li>
30                <li>BY FIELD<br>
31                    <a href="field/astronomy-1">Astronomy</a><br><a href='field/biology-2'>Biology</a><br><a href='field/chemistry-3'>Chemistry</a><br><a href='field/computing-4'>Computing</a><br><a href='field/mathematics-5'>Mathematics</a>
32                    <a href='field/medicine-6'>Medicine</a><br><a href='field/physics-7'>Physics</a><br><a href='field/astronauts-8'>Astronauts</a><br>           </li>
33                    <li><a href="resources">RESOURCES</a></li>
34                    <li><a href="contact">CONTACT</a></li>
35
36            </ul>
37        </div>
38        <div id="content">
39            <div id='profil'>
40                <div id='head'>
41                    
42                    <div id='right_head'>
43                        <h3>Madeleine&nbsp;BRES
44                    </h3>
45                    <p>19th century</p>
46                    <p><u>Fields:</u> <span style='color: #d83c3c;'><a href='field/Medicine-6'>Medicine</a><a href='field/-0'></a></span></p>
47                    <p>Born: 1842<br />
48 Death: 1921<br />
49 Main achievements: First French woman to obtain a PhD in medicine</p>
50
51
52                    <div class='biography'>Madeleine Br&egrave;s, n&eacute;e Gebelin, was the first French woman to obtain a medical degree. <br />
53<br />
54 Born Madeleine Gebelin, she told in the Medical Chronicle on 1 April 1895 how her medical career was born - "I was barely eight years old when my father, who was a wheelwright by trade - it is not a silly job - drove me to the sisters' home where he carried out his work..." In the hospital of N&icirc;mes, one nun took an affection to her and taught her some small procedures, such as the preparation of herbal teas and poultices.<br />
55<br />
56 She was twelve when the Gebelin family left for Paris, and only fifteen and one month when she married Adrien-St&eacute;phane Br&egrave;s, a tram conductor. Thanks to the efforts of Julie-Victoire Daubi&eacute;, French degrees became open to women since 1861. She first had to obtain the consent of her husband, as at the time French law judged married women to be the legal responsibility of their husbands. In 1866, she presented herself to the Dean of the Faculty of Medicine in Paris, Charles Adolphe Wurtz, and asked him for permission to enroll to study medicine. The latter announced that he would do it, but on the condition that she first obtained a degree in Arts and Sciences---a task which she accomplished in three years. She presented herself again to the Dean of the Faculty of Medicine at the University of Paris and remarked to him that nothing now opposed her enrollment in the medical course, and that there were three female foreigners - the American Mary Putnam, the Russian Catherine Gontcharoff and the English Elizabeth Garrett Anderson - were holders of nationally known equivalent degrees. She finally obtained the right to study the medicine course at the faculty.<br />
57<br />
58 The dean's decision was met with opposition from the university and medical community. The doctor Henri Montanier wrote in 1868 in the Hospitals Gazette "to make a woman a doctor, it is necessary to make her lose her sensitivity, her timidity, her modesty, harden her to the sight of the most horrible and frightening things. When the woman arrives at that, I ask myself, what remains of the woman? A being who is no longer either a young girl nor a woman; neither a wife, nor a mother." It is difficult to be more eloquent in closing, with the help of misogynistic arguments, the doors of medicine to women.<br />
59<br />
60 Despite his personal endorsement, the Dean W&uuml;rtz took this application to the Minister of Education, Victor Duruy, who approved the acceptance of Br&egrave;s, but pre-emptively brought the matter to the Council of Ministers. The

```



Using file paths on your computer

The problems:

1. Handling file paths across platforms
2. Reading multiple files
3. Manipulating files and directories

A few python libraries offer solutions. These are (as far as I know) OS, glob, shutil, and pathlib.



The pathlib library

```
from pathlib import path

p = Path('.')
Path.cmd()
list(Path.cwd().glob('*.*py'))
p.joinpath('python', 'scripts', 'test.py')
```



Pathlib

Method	Description
Path.exists()	Whether the path points to an existing file or directory.
Path.is_dir()	Return True if the path points to a directory, False if it points to another kind of file or the path doesn't exist.
Path.is_file()	Return True if the path points to a regular file, False if it points to another kind of file or the path doesn't exist.
Path.cwd()	Return a new path object representing the current directory.
Path.home()	Return a new path object representing the user's home directory.
Path.mkdir()	Create a new directory at this given path.
Path.rename(<i>target</i>)	Rename this file or directory to the given <i>target</i> , and return a new Path instance pointing to <i>target</i> .
Path.replace(<i>target</i>)	Rename this file or directory to the given <i>target</i> , and return a new Path instance pointing to <i>target</i> . If <i>target</i> points to an existing file or directory, it will be unconditionally replaced.



Pathlib

Method	Description
<code>Path.joinpath(*other)</code>	Combining the path with each of the other arguments in turn.
<code>Path.resolve()</code>	Make the path absolute, resolving any symlinks.
<code>Path.glob(pattern)</code>	Glob the given relative <i>pattern</i> in the directory represented by this path, yielding all matching files
<code>Path.parent</code>	Returns a string representing the parent directory
<code>Path.name</code>	Returns string representing the final path component, excluding the drive and root, if any:



Dealing with Windows paths

On Windows, the path separator is a backslash, \.

Backslash is also used as an *escape character* in order to represent non-printable characters.

Use *raw string literals* to represent Windows paths. These are string literals that have an r prepended to them.

For example: r'C:\Users'



Class Exercise – Libraries

Use the library math to solve the following questions:

1. Read about the library here: <https://docs.python.org/3/library/math.html>
2. Return the absolute value of a given number.
 - Check your answer for 3.45, -13, 0
3. Calculate the natural logarithm of 25. Then Calculate the base-4 logarithm of 25.
4. Generate pi.
5. Use the library requests <http://www.python-requests.org/en/master/user/quickstart> (you might need to install it).
 1. Find what is the function to get a request (read the section “Make a request”)
 2. Find how you can read the response (read the section “Response content”).
 3. Write a small function that receives as input a url (should be a string), and a word, and return two numbers: the page length and the number of times the word is found in the page.
 - For example, for the url: <http://www.gutenberg.org/cache/epub/844/pg844.txt> and word ‘Ernest’, the function should return (82, 142382)

