VIDL Introduction to Python

October 26, 2017

```
In [1]: x = 4
```

This is now text

1 This is going big

1.1 Less big

1.1.1 Even Smaller

$$e = mc^2 \times \sqrt{4\pi}$$

This is *italics* and this is **bold**

---

1. Item
2. Item

    - Subitem
    - Subitem

3. Item

    1. Subitem
    2. Subitem

4. Item.

```
In [2]: %lsmagic
```

```
Out[2]: Available line magics:
        %alias  %alias_magic  %autocall  %automagic  %autosave  %bookmark  %cat  %cd  %clear  %c

        Available cell magics:
        %%!  %%HTML  %%SVG  %%bash  %%capture  %%debug  %%file  %%html  %%javascript  %%js  %%la

        Automagic is ON, % prefix IS NOT needed for line magics.
```

```
In [3]: %pwd
```

```
Out[3]: '/Users/bobbyb/Documents/classes/sc250/Notebooks'

In [4]: %ls

100m-WR-back-M.csv
100m-WR-back-M.tsv
100m-WR-back-W.csv
100m-WR-back-W.tsv
100m-WR-breast-M.tsv
100m-WR-breast-W.tsv
100m-WR-fly-M.tsv
100m-WR-fly-W.tsv
100m-WR-free-M.csv
100m-WR-free-W.csv
200m-IM-M.tsv
200m-IM-W.tsv
200m-WR-IM-M.csv
Debugging - Quick Intro.ipynb
FastRecalibration.csv
Floating Point and Pandas.ipynb
Fourier/
FourierTalkOSCON-master/
FourierTransorm.zip
Graphs for IEEE VR Paper.ipynb
HW3-start-AT.ipynb
Homework 2 Solutions.ipynb
Homework 3 Solutions.ipynb
Homework 4 Pre-Solutions.ipynb
Homework 4 Solutions.ipynb
JupyterWorkflow.ipynb
LIGO/
Lecture 1 Notebook Example.ipynb
Lecture 2 - Matplotlib, Numpy, Scipy, and Modules.ipynb
Moving Average vs Convolution for Oceanographic Data.ipynb
PDF/
PrismData-Teens.csv
PrismData.csv
PrismDataAll.csv
Python Recitation Notebook.ipynb
Recalibration.csv
SN_m_tot_V2.0.csv
Swimming Scraping and Visualization.ipynb
Swimming-100m-free-M.csv
Tutorial for SC 3250.ipynb
Untitled.ipynb
Untitled1.ipynb
Untitled2.ipynb
Untitled3.ipynb
```

```
Untitled4.ipynb
VIDL Introduction to Python.ipynb
VIDL-Prep.ipynb
__pycache__/
ch4.tsv
ch4nat.txt
co2.tsv
co2nat.txt
data_2015-16_suppressed_district_level_act.csv
firstminuspre.pdf
freewr/
icedata.tsv
listings.csv
pickle_example
postminuspre.pdf
snippets.txt
swimmer/
swimmer.zip
swims/
swims2/
testmod/
tmp/
untitled.txt
wikitablescrape.py
```

```python
In [5]: %matplotlib inline

In [6]: import numpy as np
        import matplotlib.pylab as plt
        x = np.arange(0,1,0.001)*12*np.pi
        y = np.sin(x)
        plt.plot(x,y)
        plt.show()
```
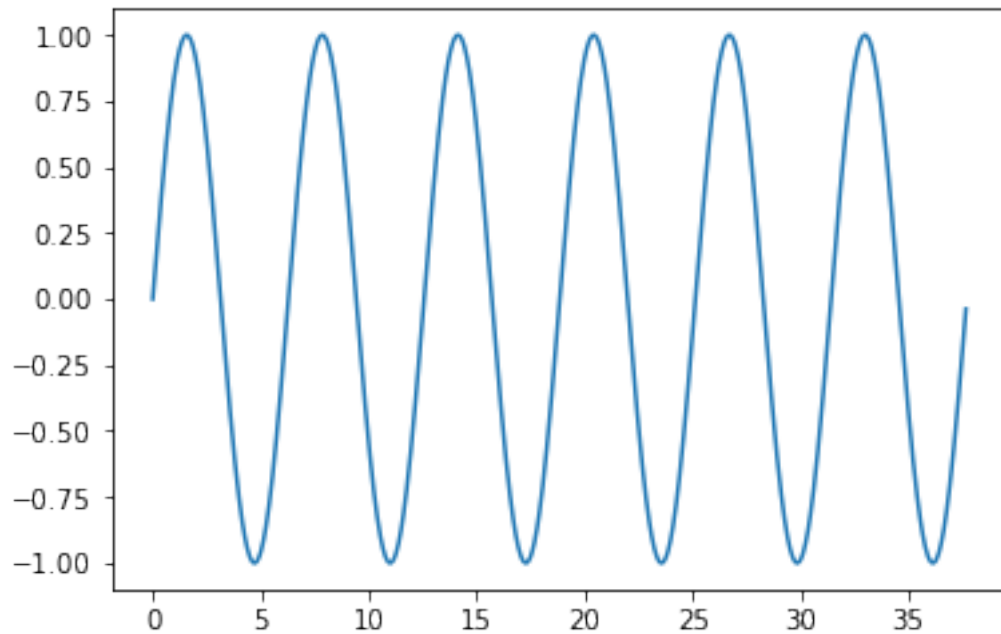
```
In [7]: y = 4.5

In [8]: print(y)

4.5


In [9]: print(y)

4.5
```

In the workshop, I couldn't get this to work because I was using a url for the youtube video of the form www.youtube.com/watch?v=whatever. The solution is to change the url to www.youtube.com/embed/whatever. There may be some urls for which this doesn't work.

```
In [30]: %%HTML
         <iframe width="400" height="300" src="https://www.youtube.com/embed/7i9rm6XOqs0"></ifra

<IPython.core.display.HTML object>


In [11]: %%timeit
         for i in range(10000):
             i*i*i

1000 loops, best of 3: 1.07 ms per loop
```

```
In [12]: import pandas as pd
         import wikitablescrape

In [13]: page = "https://en.wikipedia.org/wiki/World_record_progression_100_metres_freestyle"

In [14]: wikitablescrape.scrape(url=page,output_name='freewr')

In [15]: mf = pd.read_csv('freewr/freewr.csv',header=0)
         wf = pd.read_csv('freewr/freewr_2.csv',header=0)

In [16]: mf.head(10)

Out[16]:     #    Time Unnamed: 2               Name      Nationality         Date  \
         0   1  1:05.8        NaN      Zoltán Halmay          Hungary   3 Dec 1905
         1   2  1:05.6        NaN     Charles Daniels    United States  20 Jul 1908
         2   3  1:02.8       (tt)     Charles Daniels    United States  15 Apr 1910
         3   4  1:02.4        NaN       Kurt Bretting          Germany   6 Apr 1912
         4   5  1:01.6        NaN     Duke Kahanamoku    United States  20 Jul 1912
         5   6  1:01.4        NaN     Duke Kahanamoku    United States   9 Aug 1918
         6   7  1:00.4        NaN     Duke Kahanamoku    United States  24 Aug 1920
         7   8    58.6        NaN  Johnny Weissmuller    United States  19 Jul 1922
         8   9    57.4        NaN  Johnny Weissmuller    United States  17 Feb 1924
         9  10    56.8        NaN          Peter Fick    United States   2 Mar 1934

                                                 Meet  \
         0
         1                                Olympic Games
         2   Special record attempt by the New York AC
         3
         4
         5                      International exhibition
         6                                Olympic Games
         7
         8
         9           Yale University Swimming Carnival

                                    Location  Ref
         0                      Vienna, Austria  NaN
         1              London, United Kingdom  NaN
         2       New York City, United States  NaN
         3                    Brussels, Belgium  NaN
         4                     Hamburg, Germany  NaN
         5       New York City, United States  NaN
         6                    Antwerp, Belgium  NaN
         7  Alameda, California, United States  NaN
         8               Miami, United States  NaN
         9          New Haven, United States  NaN

In [17]: mf['Date']
```

```
0      3 Dec 1905
1     20 Jul 1908
2     15 Apr 1910
3      6 Apr 1912
4     20 Jul 1912
5      9 Aug 1918
6     24 Aug 1920
7     19 Jul 1922
8     17 Feb 1924
9      2 Mar 1934
10     5 Mar 1935
11    11 Feb 1936
12    13 Apr 1944
13    15 Sep 1947
14    29 Jun 1948
15     1 Apr 1954
16    30 Nov 1956
17    19 Jan 1957
18    28 Jan 1957
19    18 Aug 1961
20    20 Sep 1961
21    13 Sep 1964
22    14 Oct 1964
23    27 Jul 1967
24     2 Sep 1968
25    19 Oct 1968
26    23 Aug 1970
27     5 Aug 1972
28     3 Sep 1972
29    21 Jun 1975
30     3 Aug 1975
31    23 Aug 1975
32    24 Jul 1976
33    25 Jul 1976
34    14 Aug 1976
35     3 Apr 1981
36     6 Aug 1985
37     6 Aug 1985
38    24 Jun 1986
39    10 Aug 1988
40    18 Jun 1994
41    16 Sep 2000
42    19 Sep 2000
43    21 Mar 2008
44    22 Mar 2008
45    11 Aug 2008
46    13 Aug 2008
47    13 Aug 2008
```

```
        48     23 Apr 2009
        49     30 Jul 2009
        Name: Date, dtype: object

In [18]: mf['Date']=pd.to_datetime(mf['Date'])
         mf['Date']

Out[18]: 0      1905-12-03
         1      1908-07-20
         2      1910-04-15
         3      1912-04-06
         4      1912-07-20
         5      1918-08-09
         6      1920-08-24
         7      1922-07-19
         8      1924-02-17
         9      1934-03-02
         10     1935-03-05
         11     1936-02-11
         12     1944-04-13
         13     1947-09-15
         14     1948-06-29
         15     1954-04-01
         16     1956-11-30
         17     1957-01-19
         18     1957-01-28
         19     1961-08-18
         20     1961-09-20
         21     1964-09-13
         22     1964-10-14
         23     1967-07-27
         24     1968-09-02
         25     1968-10-19
         26     1970-08-23
         27     1972-08-05
         28     1972-09-03
         29     1975-06-21
         30     1975-08-03
         31     1975-08-23
         32     1976-07-24
         33     1976-07-25
         34     1976-08-14
         35     1981-04-03
         36     1985-08-06
         37     1985-08-06
         38     1986-06-24
         39     1988-08-10
         40     1994-06-18
```

```
41    2000-09-16
42    2000-09-19
43    2008-03-21
44    2008-03-22
45    2008-08-11
46    2008-08-13
47    2008-08-13
48    2009-04-23
49    2009-07-30
Name: Date, dtype: datetime64[ns]
```

In [25]:
```python
def process_time(x):
    x = str(x)  # At Kate's suggestion, always convert x to be a string
    if ":" in x:
        time = x.split(":")
        return 60.0*float(time[0])+float(time[1])
    else:
        return float(x)
```

In [26]:
```python
mf['Time']=mf['Time'].apply(process_time)
wf['Time']=wf['Time'].apply(process_time)
wf['Date']=pd.to_datetime(wf['Date'])
```

In [27]:
```python
wf.head(10)
```

Out[27]:

	#	Time	Unnamed: 2	Name	Nationality	Date
0	1	95.0	NaN	Martha Gerstung	Germany	1908-10-18
1	2	86.6	NaN	Claire Guttenstein	Belgium	1910-10-02
2	3	84.6	NaN	Daisy Curwen	Great Britain	1911-09-29
3	4	80.6	NaN	Daisy Curwen	Great Britain	1912-06-10
4	5	79.8	NaN	Fanny Durack	Australia	1912-07-09
5	6	78.8	NaN	Fanny Durack	Australia	1912-07-21
6	7	76.2	NaN	Fanny Durack	Australia	1915-02-06
7	8	74.4	sf	Ethelda Bleibtrey	United States	1920-08-23
8	9	73.6	NaN	Ethelda Bleibtrey	United States	1920-08-25
9	10	72.8	NaN	Gertrude Ederle	United States	1923-06-30

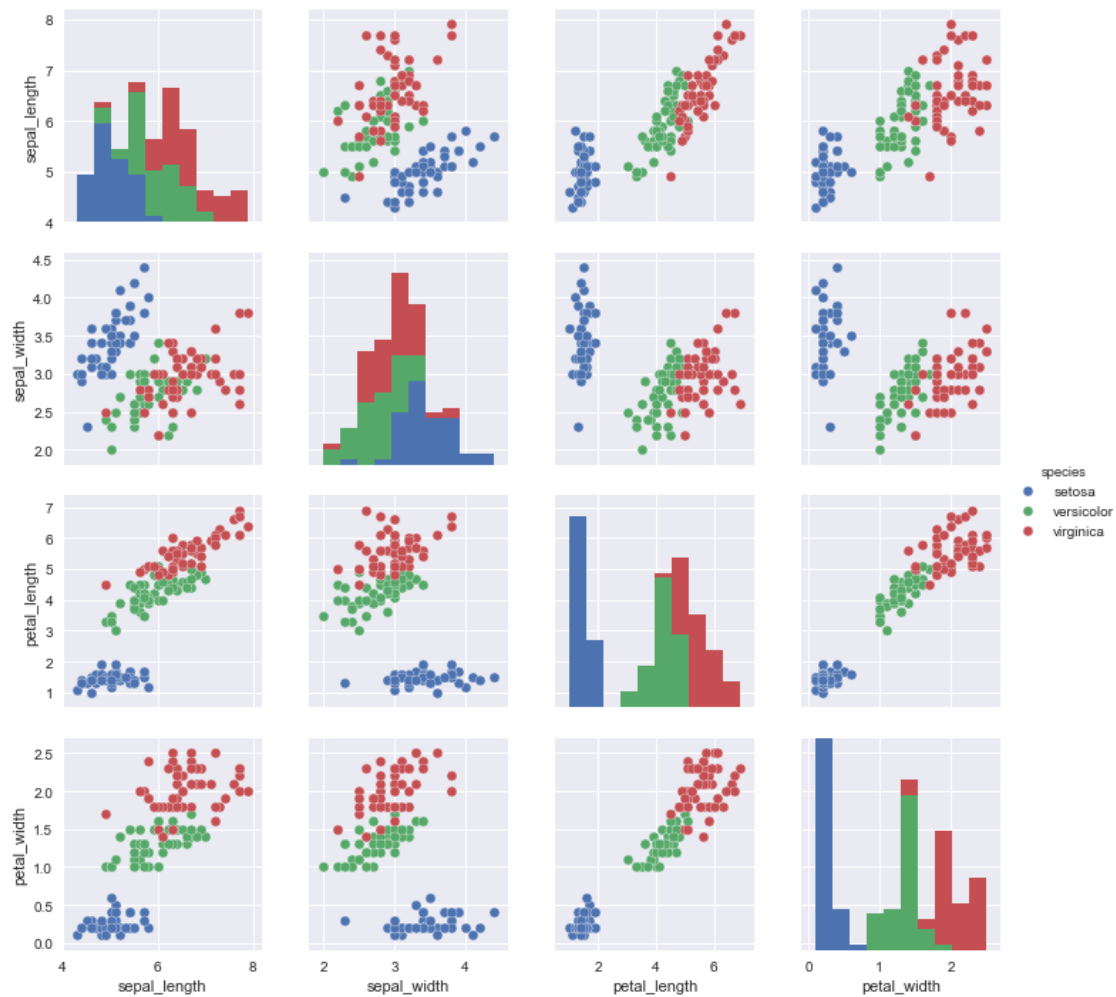|	Meet	Location	Ref
0		Magdeburg, Germany	NaN
1		Schaerbeek, Belgium	NaN
2		Liverpool, United Kingdom	NaN
3		Birkenhead, United Kingdom	NaN
4	Olympic Games	Stockholm, Sweden	NaN
5		Hamburg, Germany	NaN
6	NSW Ladies' Amateur Championships	Sydney, Australia	NaN
7	Olympic Games	Antwerp, Belgium	NaN
8	Olympic Games	Antwerp, Belgium	NaN
9		Newark, New Jersey, United States	NaN

```
In [28]: fig,ax=plt.subplots(figsize=(10,10))
         mf.plot(x='Date',y='Time',ax=ax,label='Men')
         wf.plot(x='Date',y='Time',ax=ax,label='Women')
```

Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x11faec8d0>
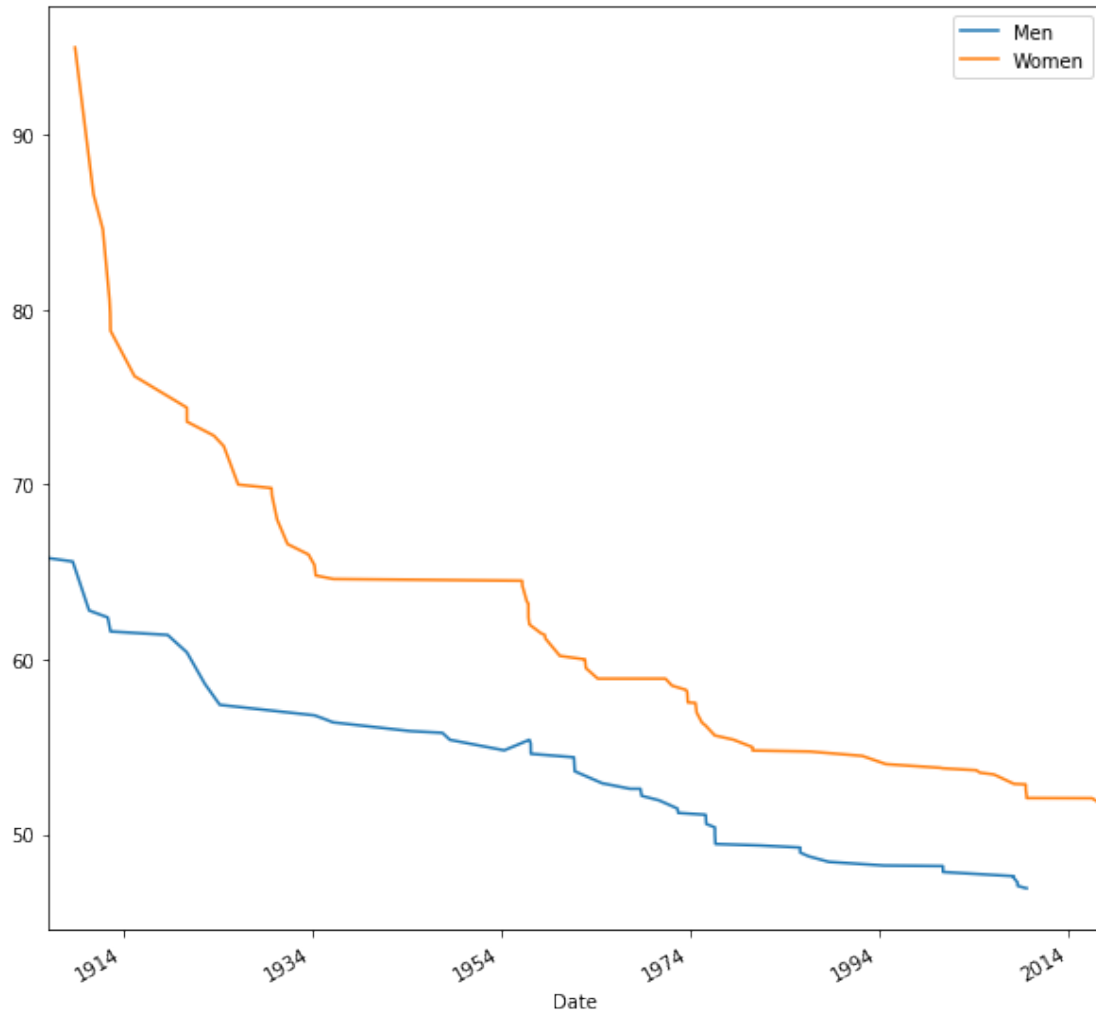

```
In [29]: import seaborn as sns
         sns.set()
         data = sns.load_dataset('iris')
         sns.pairplot(data,hue='species')
```

Out[29]: <seaborn.axisgrid.PairGrid at 0x11ff8e390>

In [ ]: