Python Data Analysis

Johnny Pribyl

***Abstract*—This lab intends to provide an introduction to using Python as a tool for data analysis. Learning how to do basic statistics and plotting in Python makes it possible to leverage very powerful libraries. This week, I will be investigating SciPy, NumPy, and Matplotlib.**

I. Introduction

THIS lab does not have an in-class portion. All you would need to reproduce the results is a computer with Python. I did all of my development in Vim. If you would like to emulate my environment, you can find my .vimrc on github at:

```
jpribyl/legendary-waddle
```

Additionally, you can find all of my code at:

```
jpribyl/cautious-palm-tree
```

II. Starting Python

I had a much easier time loading python than some of my peers. This is probably a result, at least in part, to this line in my .bashrc:

```
alias accio='sudo apt-get install'
```

While they had to mess around with the PATH and environment variables, I got to sail away into the sunset with:

```
accio python && python
```

It also probably doesn't hurt that I have used Python before.

III. Exercises: Python and Plotting

A. Numbers & Arrays

This section provides an extremely basic introduction to python. I think that it's main purpose was to check whether my version of python supports implicit conversion of integers to floats.

Although I occaisionally set up a virtual environment in python 2 by accident, I prefer usin python 3. Among other things, it does support this implicit conversion.

The code in this section was simple enough that I opted to do input it directly to the terminal. Here are the results. The characters >>> occur before inputs but not before outputs.

```
$ python3
Python 3.5.2
[GCC 5.4.0 20160609] on linux
>>> 5.0/3
1.6666666666666667
>>> 5/3
1.6666666666666667
>>> 5**2
25
>>> import numpy as np
>>> np.sin(np.pi/6)
0.49999999999999994
>>> t=10
>>> t2=2*t
>>> g=-9.8
>>> y=g*t**2/2
>>> print(y)
-490.00000000000006
```

Had I tried using NumPy before importing it, I would have gotten an error:

```
>>> np.pi
Traceback (most recent call last):
 File "<stdin>", line 1, in <module>
NameError: name 'np' is not defined
>>>
```

Which could be good to keep in mind going forward.

B. Plotting Data and Curves

This section is a slight step up from the previous section. Although I do love jupyter, I'm not quite ready to let go of vim, so I got familiar with

```
plt.show()
```

The code for this section is detailed almost word for word in *lab_descrip/data_analysis.pdf* so I'm not going to recopy it here. You can find it at *cautious − palm − tree/simple_plot.py*. The plot is in Figure 1.

If we alter the code a little bit, then we can easily include error bars, label axes, add a legend, change the color, symbol, size, and line type.

Brian D'Urso

Listing 1. Altering the Graphs

```
# changing color, size, and symbol
plt.plot(
                    t_theory,
                    y_theory,
                    'r.',
                    MarkerSize=1
)

# labelling axes
plt.xlabel('time (s)')
plt.ylabel('height (m)')

# and, adding a legend
plt.legend(['Theory', 'Actual'])
plt.show()
plt.savefig('figures/figure3.png')
```

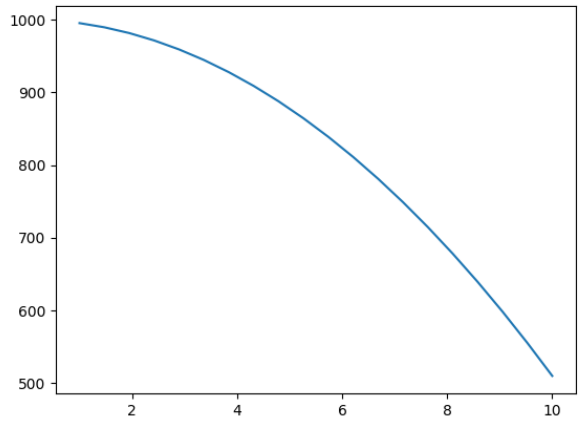

Fig. 1. Plot generated by a simple plot of kinematics

Which produces this plot:

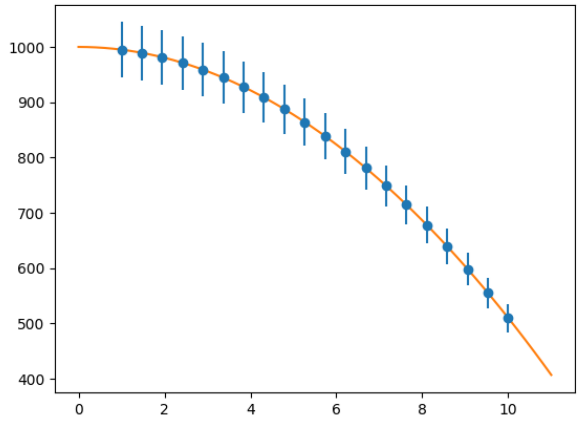

Fig. 2. Including mock error bars

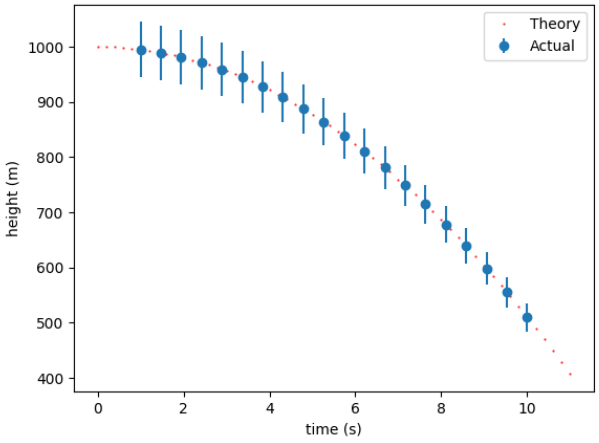

Fig. 3. Changing size and introducing a legend

IV. Format

The report can be written in LaTeX or Microsoft Word, but LaTeX is definitely preferred. Its appearance should be as close to this document as possible to achieve consistency in the proceedings.

References should be cited as numbers, and should be ordered by their appearance (example: "... as shown in [1], ..."). Only references that are actually cited can be listed in the references section. The references' format should be evident from the examples in this text.

References should be of academic character and should be published and accessible. Your advisor can answer your questions regarding literature research. You must cite all used sources. Examples of good references include text books and scientific journals or conference proceedings. If possible, citing internet pages should be avoided. In particular, Wikipedia is *not* an appropriate reference in academic reports. Avoiding references in languages other than English is recommended.

TABLE I
Simulation Parameters

Information message length	$k = 16000$ bit
Radio segment size	$b = 160$ bit
Rate of component codes	$R_{cc} = 1/3$
Polynomial of component encoders	$[1, 33/37, 25/37]_8$

V. Filling this page

Gallia est omnis divisa in partes tres, quarum unam incolunt Belgae, aliam Aquitani, tertiam qui ipsorum lingua Celtae, nostra Galli appellantur. Gallos ab Aquitanis Garumna flumen, a Belgis Matrona et Sequana dividit. Horum omnium fortissimi sunt Belgae, propterea quod a cultu atque humanitate provinciae longissime absunt, minimeque ad eos mercatores saepe commeant atque ea quae ad effeminandos animos pertinent important, proximique sunt Germanis, qui

trans Rhenum incolunt, quibuscum continenter bellum gerunt. Qua de causa Helvetii quoque reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis cum Germanis contendunt, cum aut suis finibus eos prohibent aut ipsi in eorum finibus bellum gerunt. Eorum una, pars, quam Gallos obtinere dictum est, initium capit a flumine Rhodano, continetur Garumna flumine, Oceano, finibus Belgarum, attingit etiam ab Sequanis et Helvetiis flumen Rhenum, vergit ad septentriones. Belgae ab extremis Galliae finibus oriuntur, pertinent ad inferiorem partem fluminis Rheni, spectant in septentrionem et orientem solem.

VI. Conclusion

This section summarizes the paper.

References

[1] J. Hagenauer, E. Offer, and L. Papke. Iterative decoding of binary block and convolutional codes. *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 429-445, Mar. 1996.

[2] T. Mayer, H. Jenkac, and J. Hagenauer. Turbo base-station cooperation for intercell interference cancellation. *IEEE Int. Conf. Commun. (ICC)*, Istanbul, Turkey, pp. 356–361, June 2006.

[3] J. G. Proakis. *Digital Communications*. McGraw-Hill Book Co., New York, USA, 3rd edition, 1995.

[4] F. R. Kschischang. Giving a talk: Guidelines for the Preparation and Presentation of Technical Seminars. http://www.comm.toronto.edu/frank/guide/guide.pdf.

[5] IEEE Transactions LaTeXand Microsoft Word Style Files. http://www.ieee.org/web/publications/authors/transjnl/index.html