

Note: For this homework you will need to make an account on github. You will share the repo with me. My github username is "**Yunhua468**". The repo should be called CSc47400, which contains several folders, and put each problem in a **separate folder**, which is "prob1", "prob2", "prob3".

Problem 1: Make line plots in Python

Data Source

We will make some simple line plots in matplotlib.

We will use the U.S. Food Commodity Availability by Food Source, 1994-2008

<https://www.ers.usda.gov/publications/pub-details/?pubid=81817>

You will use the **Appendix B: Consumption share tables**

Appendix B: Consumption share tables

[Download XLS](#)

Table structure:

First of all there are 4 time periods considered:

- 1994-1998,
- 2003-2004,
- 2005-2006,
- 2007-2008.

For each time period there are two worksheets,

- a Food At Home Consumption (FAH),
- a Food Away From Home (FAFH).

Lets focus only on Food At Home.

Each worksheet is grouped into groups of rows

- the first being "LAFA away-fromhome consumption amounts: 2007-08 means and confidence intervals for U.S., children, and adults",
- the second being "Appendix Table C-8 (cont'd). LAFA away-fromhome consumption amounts: 2007-08 means and confidence intervals for boys, girls, men, and women," and so forth.

Task

Note: Don't reformat with excel or open office. Use the "**pandas.read_excel**" function to load the data.

You will need to use the sheetname properly to select the sheets we need (**FAH** for each time).

You will extract the data you need from the loaded dataframes and create new data frames with just the information you need, and the proper headers.

Lets concentrate on **Men and Women** and the at home consumption of a few kinds of fruit and dairy products to see if there are any trends.

Make 4 line plots using python:

1. Fruit types over time (Men)
2. Fruit types over time (Women)
3. Dairy products over time (Men)
4. Dairy products over time (Women)

The fruit types you should consider will be

- "Apples as fruit",
- "Bananas",
- "Berries",
- "Grapes",
- "Melons",
- "Oranges, Total",
- "Other citrus fruit",
- "Stone fruit",
- "Tropical fruit".

For dairy products you should look at

- "Fluid milk total",
- "Butter",
- "Cheese",
- "Yogurt",
- "Dairy, Other".

Requirement:

- Each product should have **a line with its own color, line style**.
- The graphs should have **each axis labeled with the variable measured and units**.

- There should be a **legend** to understand which product is which line.
- The **data source and source url** should appear below the graph in the image. This should be part of the same ipython notebook with the code extracting the data.
- We are going to just use **mean pounds** here, which, for Men, appears in the 7th column, after the 77th row. You will figure the details out through inspecting the sheets in a spreadsheet program and checking that you get the same numbers in the ipython notebook.
- Below each plot image, use a markdown cell to write a **caption**. The caption should quickly summarize what is in the fig (2 sentences). Those summaries should fill in some details in the attributes not obvious from reading the titles, legends and axis. You might get this information from the report summary. Each caption should say something interesting thing we conclude from the figure. Like "in the figure we see an increase in the consumption of rabbit eyeballs", or "the overall consumption of bird droppings remained flat but rained higher than red grapes." Just to be clear, **do not use these quotes as captions. The analysis should be your own.**

Problem 2: Make bar graphs in Python

Task

Use the **data** from Problem 1 task. Compute the **percent increase or decrease** from 1994-1998 to 2007-2008 for each product.

Again make **2 charts**: one for men and one for women.

Requirement

- Here **group** the bars representing fruit products, and dairy products separately with a space between them.
- Within each category of fruit or dairy, **sort** from largest decrease to largest increasing.
- Rather than legend, here **use x axis to label by product type** and **use a 45 degree slant** so that the labels all fit.

Problem 3: Data Set Types/Ipython Slides

Task

In this problem I want you to think about different **data set types**.

Find an example of a visualization (image) of each of these data types

1. 1D Time Series
2. 2D Scaler Field
3. 3D Scaler Field
4. 2D Vector Field
5. 3D Shape
6. Graph (not tree)

Requirement

- First **collect (download)** the images you find.
- Just to make this a double challenge, you will make this as an **ipython slideshow**.
- You will use the commands to display in images within your ipython notebook as "results".
 - `from IPython.display import display, Image`
 - `display(Image(filename='misc-you-dont-say-l.png'))`
- Next watch this: <https://mljar.com/blog/jupyter-notebook-presentation/> and figure out how to **make slides** with your ipython notebook and export them to **html/js** slideshow in your repo.
- With each of the 6 examples you should have a short **explanation** of how this is a visualization of the data (what is the data and how it is this data type).