YOU CAN GOOGLE ALMOST ANY QUESTION – Stackoverflow is a website that has lots of answers

Web scraping (collecting data)

* You get data (scrape)
* Store the data
* Visualize the data

Install packages for today

* Conda install -c anaconda requests
* Conda install -c anaconda beautifulsoup4
* Conda install pandas

What are packages?

* Packages are packages of code that do things. You import packages to do a function.
* Anaconda has a large library that preinstalls many packages on your computer

Matplotlib

* Big library for visualizations
* Sometimes libraries are too big: you can import sections of them by adding a dot and writing import **matplotlib.pyplot** as plt.
  + THAT DOES: import [main program].[section] as [the name that you want to call it]
* Import requests.
  + **Requests** is a program that collects data (from websites)
* Import beautifulsoup.
  + **Beautifulsoup** parses the website data (html) and turn it into useful data
* Import pandas.
  + **Pandas** visualizes data
* # signifies a personal comment is coming (all information that follows is ignored by the program)
* ´´´ creates a comment too, that can enclose multiple lines ´´´

The more ### you used the smaller the comment font

Variables

* Variables store information
* Here is a variable: URL = ’https://coinmarketcap.com/’
  + URL (variable name)
  + = (computer knows we’re creating a variable)
  + ’https://coinmarketcap.com/’ (the actual variable)
  + ’’ apostrophes tell the computer that the information inside them is text to be read as a variable. If your variable is numerical you don’t need them. Example:
    - Print (1+1) = 2
    - Print (‘py’+’thon’) = pyton

**Finding information in the website:**

res = requests.get(url)

* Runs a command to request the HTLM code from the url

soup = BeautifulSoup(res.content, 'lxml') #The second argument 'lxml' is the html parser

* Interprets the information as text lines we can read

Table = soup.find\_all(´table´)[0]

* Defines variable “table” as what “soup” package’s function “find\_all” finds??? (this is probably wrong)