Day 1

* Data portals (public)
  + https://data.gov.uk/
  + http://www.opendatanetwork.com/
  + https://www.data.gov/
  + Compare the above 3 sites – how they package the data and make it easier or harder for a user. Data.gov is probably the crappiest
  + Also Socrata
* GreatFire.org
  + Check for data sites that China is censoring
* Open a web browser
  + from selenium import webdriver
  + driver = webdriver.Firefox()
  + websitename = ‘www.youtube.com’
  + driver.get(‘http.//’ + websitename)
* Screenshot
  + driver.save\_screenshot
* curl – to get lots of links from a page
* Go on github web interface to create new folder
  + week-01/ - creates folder
  + Within folder, create DATAREPORT.md (HTML file – markdown)
  + #data.gov, #data.gov.uk, #opendatanetwork.com, #socrata
* Gentrification? Interesting data – effects on crime, housing prices
* Homework: DATAREPORT file

Day 2

* import requests
  + requests.get(downloadable file)
* str.split(“,”) splits based on commas
  + data.split(‘\r\n’)
  + mytext.splitlines()
* whoami – tells you your username
* pip install \_\_whatever\_package\_\_
* list(range[2003, 2016]) – creates a list from 2003 to 2015

Day 3

* Code is in github to take pdf and turn it into csv
* C = csv.writer(file)
* pip install pdfplumber
* from string import Template
  + t = Template(“What’s up ${name}?”)
  + t.substitute(name = “Alan”)
* {whatever} allows user to input whatever they want and that becomes whatever
* To find where link in HTML goes: right click on link and type “inspect”. Don’t forget that that’s the string appended on the homepage URL string
* “Beautiful Soup” library
  + from bs4 import BeautifulSoup
  + soup = BeautifulSoup(txt)
  + soup.find\_all(‘h3’) – generates list of all strings in txt that have ‘h3’