File Creation
IS590OM Final Project

Ke-Rou Wang (kerourw2)

There are three intermediate datasets that I used for creating final dataset:

• Population dataset: Resident Population by State, Annual

• UFO dataset: ufo-scrubbed-geocoded-time-standardized-00

• Climate dataset: NOAA National Centers for Environmental information(IS590OM)

The UFO dataset is a dataset that has already been formed, whereas the other two, population dataset and climate dataset, are not. Based on this, I search online and found appropriate data from websites FRED Economic Data, and NOAA National Centers for Environmental Information. After searching the information, I do web-scrapping and then create the date file manually.

Population dataset:

I obtained the data from FRED Economic Data, here's the link: https://fred.stlouisfed.org/release/tables?rid=118&eid=259194&od=1900-01-01#, and the sample of the webpage that I used to gather data.

Release Tables:
Resident Population by State, Annual

| Add to Data List | Add to Graph | | Expand All | Collapse All | | | 1900 | 2014 | **2015** | 2016 | 2018 |

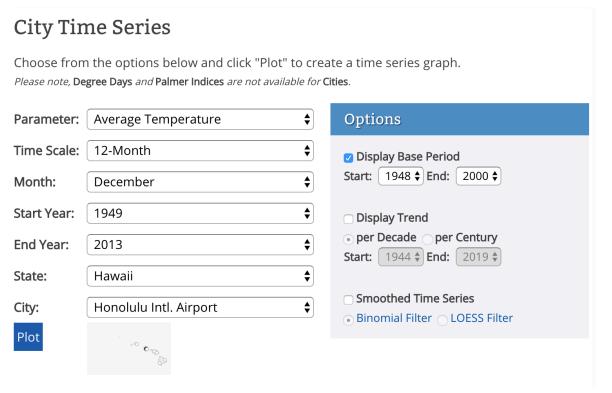|	Name	Thousands of Persons		
		2015	Preceding Period	Year Ago from Period
	Alabama	4,853.160	4,842.481	4,842.481
	Alaska	737.547	736.307	736.307
	Arizona	6,833.596	6,733.840	6,733.840
	Arkansas	2,978.407	2,967.726	2,967.726
	California	38,953.142	38,625.139	38,625.139
	Colorado	5,452.107	5,351.218	5,351.218
	Connecticut	3,587.509	3,594.783	3,594.783
	Delaware	941.413	932.596	932.596
	District of Columbia	675.254	662.513	662.513

The website is easy for users to retrieve the information by year, whereas its format of the data is not suitable for pasting the data to excel directly, so I have to edit each row/value one by one; after this, I will be able to get a table like this:

States	1949	1950	1951	1952	1953	1954	1955
Alabama	3000.00	3058.00	3059.00	3068.00	3053.00	3014.00	3050.00
Alaska	0.00	135.00	158.00	189.00	205.00	215.00	222.00
Arizona	714.00	756.00	785.00	842.00	894.00	933.00	987.00
Arkansas	1844.00	1908.00	1901.00	1838.00	1780.00	1734.00	1725.00
California	10337.00	10677.00	11134.00	11635.00	12251.00	12746.00	13133.00
Colorado	1295.00	1325.00	1326.00	1365.00	1431.00	1493.00	1546.00
Connecticut	2032.00	2016.00	2028.00	2081.00	2168.00	2249.00	2300.00
Delaware	316.00	321.00	331.00	341.00	351.00	368.00	389.00
Florida	2668.00	2810.00	2980.00	3157.00	3310.00	3505.00	3747.00
Georgia	3325.00	3458.00	3531.00	3584.00	3558.00	3602.00	3636.00
Hawaii	0.00	498.00	514.00	517.00	510.00	505.00	539.00
Idaho	570.00	590.00	589.00	587.00	596.00	600.00	618.00
Illinois	8670.00	8738.00	8790.00	8956.00	9065.00	9252.00	9435.00
Indiana	3958.00	3967.00	4096.00	4148.00	4182.00	4264.00	4363.00
Iowa	2578.00	2625.00	2617.00	2626.00	2629.00	2626.00	2679.00

Climate dataset:

I obtained the data from NOAA National Centers for Environmental Information, here's the link: https://www.ncdc.noaa.gov/%20cag/city/time-series/USW00022521/tavg/12/12/1949-2013?base_prd=true&begbaseyear=1948&endbaseyear=2000, and the sample of the webpage that I used for retrieving data.

City Time Series

Choose from the options below and click "Plot" to create a time series graph.
Please note, **Degree Days** *and* **Palmer Indices** *are not available for* **Cities**.

		Options
Parameter:	Average Temperature ⬍	
Time Scale:	12-Month ⬍	☑ Display Base Period
Month:	December ⬍	Start: 1948 ⬍ End: 2000 ⬍
Start Year:	1949 ⬍	
End Year:	2013 ⬍	☐ Display Trend
		◉ per Decade ○ per Century
State:	Hawaii ⬍	Start: 1944 ⬍ End: 2019 ⬍
City:	Honolulu Intl. Airport ⬍	
Plot		☐ Smoothed Time Series
		◉ Binomial Filter ○ LOESS Filter

DATES	VALUE	RANK	DEPARTURE FROM MEAN (78.0°F) 1950-2000 BASE PERIOD
195001 - 195012	77.4°F	8	-0.6°F
195101 - 195112	77.9°F	22	-0.1°F
195201 - 195212	77.1°F	4	-0.9°F
195301 - 195312	77.6°F	15	-0.4°F
195401 - 195412	77.4°F	8	-0.6°F
195501 - 195512	76.1°F	1	-1.9°F

Compare to FRED website, NOAA's website is more user-friendly since the data format can remain the same while I would like to paste it directly to excel. After this, I will be able to get a table like this:

States	States Abbr.	Record Dates	Temperature Value	Temperature Rank
Alabama	AL	194901 - 194912	64.2°F	61
Alabama	AL	195001 - 195012	63.1°F	43
Alabama	AL	195101 - 195112	63.8°F	54
Alabama	AL	195201 - 195212	63.3°F	46
Alabama	AL	195301 - 195312	63.8°F	54
Alabama	AL	195401 - 195412	64.2°F	61
Alabama	AL	195501 - 195512	63.4°F	48
Alabama	AL	195601 - 195612	63.6°F	51
Alabama	AL	195701 - 195712	64.0°F	56
Alabama	AL	195801 - 195812	61.2°F	3
Alabama	AL	195901 - 195912	63.0°F	40

References:

Population dataset: https://fred.stlouisfed.org/release/tables?rid=118&eid=259194&od=1900-01-01#

Climate dataset: https://www.ncdc.noaa.gov/cag/city/time-series/USW00022521/tavg/12/12/1949-2013?base_prd=true&begbaseyear=1948&endbaseyear=2000