**Government macroeconomic data**

The data from 1959 to 2009 is collected from the government of a country. Below is the description of various aspects of the data.

The data is in long format.

year - 1959q1 - 2009q3

quarter - 1-4

realgdp - Real gross domestic product (Bil. of chained 2005 US$,

seasonally adjusted annual rate)

realcons - Real personal consumption expenditures (Bil. of chained

2005 US$, seasonally adjusted annual rate)

realinv - Real gross private domestic investment (Bil. of chained

2005 US$, seasonally adjusted annual rate)

realgovt - Real federal consumption expenditures & gross investment

(Bil. of chained 2005 US$, seasonally adjusted annual rate)

realdpi - Real private disposable income (Bil. of chained 2005

US$, seasonally adjusted annual rate)

cpi - End of the quarter consumer price index for all urban

consumers: all items (1982-84 = 100, seasonally adjusted).

m1 - End of the quarter M1 nominal money stock (Seasonally

adjusted)

tbilrate - Quarterly monthly average of the monthly 3-month

treasury bill: secondary market rate

unemp - Seasonally adjusted unemployment rate (%)

pop - End of the quarter total population: all ages incl. armed

forces over seas

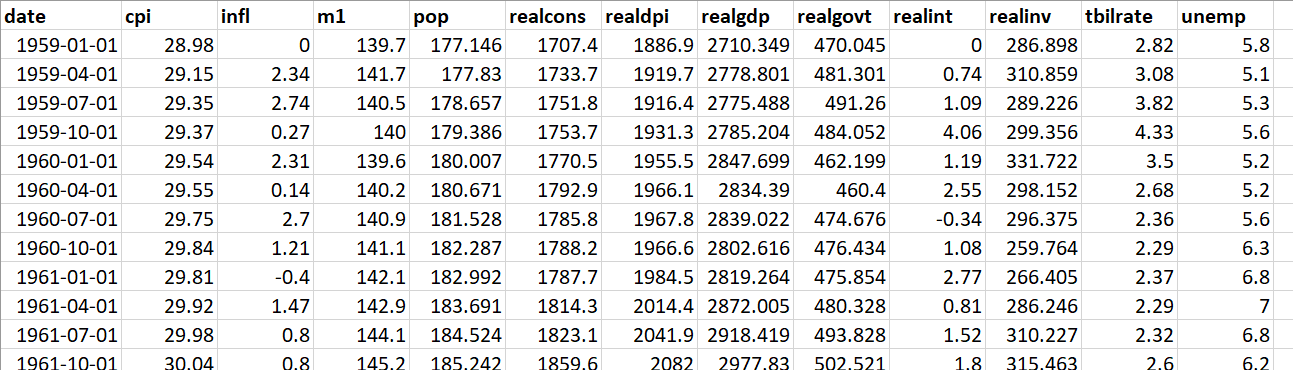
infl - Inflation rate (ln(cpi\_{t}/cpi\_{t-1}) \* 400)

realint - Real interest rate (tbilrate - infl)

Long format of the data:

|  |  |  |
| --- | --- | --- |
| **date** | **item** | **value** |
| 1959-01-01 | realgdp | 2710.349 |
| 1959-01-01 | realcons | 1707.4 |
| 1959-01-01 | realinv | 286.898 |
| 1959-01-01 | realgovt | 470.045 |
| 1959-01-01 | realdpi | 1886.9 |
| 1959-01-01 | cpi | 28.98 |
| 1959-01-01 | m1 | 139.7 |
| 1959-01-01 | tbilrate | 2.82 |
| 1959-01-01 | unemp | 5.8 |
| 1959-01-01 | pop | 177.146 |
| 1959-01-01 | infl | 0 |
| 1959-01-01 | realint | 0 |
| 1959-04-01 | realgdp | 2778.801 |
| 1959-04-01 | realcons | 1733.7 |
| 1959-04-01 | realinv | 310.859 |
| 1959-04-01 | realgovt | 481.301 |
| 1959-04-01 | realdpi | 1919.7 |

The data must be pivoted to make the data in the following format:



Once data is transformed, please find the correlation matrix. You should be able to answer the following questions:

1. Which field is an indicator of unemployment.
2. Make scatter plots of the data (only real valued)
3. Make a correlational matrix showing which columns are most correlated with unemployment

**Demo script for making a correlation matrix:**

**Python script**

import pandas as pd

import numpy as np

import os

os.chdir('E:\\0\_ulab\\big-data\\lectures\\experiment3\\macrodata')

p = pd.read\_csv('all\_long.csv')

p = p.drop(columns=['Unnamed: 0'])

# q = p.pivot(index="date", columns="item", values="value")

# q['year'] = q.index

p[['year','month', 'day']] = p['date'].str.split('-',expand=True)r = q.drop(columns=['month', 'day'])

p = p.drop(columns=['month', 'day'])

# make a cross tab

q = pd.crosstab(index=p['year'], columns=p['item'], values=p['value'], aggfunc="sum")

q.select\_dtypes(np.number).corr()