Economic Activity or Measures, Michigan, United States

Mini-project

Before working on this assignment please read these instructions fully. In the submission area, you will notice that you can click the link to **Preview the Grading** for each step of the assignment. This is the criteria that will be used for peer grading. Please familiarize yourself with the criteria before beginning the assignment.

This assignment requires that you to find **at least** two datasets on the web which are related, and that you visualize these datasets to answer a question with the broad topic of **economic activity or measures** (see below) for the region of **Ann Arbor, Michigan, United States**, or **United States** more broadly.

You can merge these datasets with data from different regions if you like! For instance, you might want to compare **Ann Arbor, Michigan, United States** to Ann Arbor, USA. In that case at least one source file must be about **Ann Arbor, Michigan, United States**.

You are welcome to choose datasets at your discretion, but keep in mind **they will be shared with your peers**, so choose appropriate datasets. Sensitive, confidential, illicit, and proprietary materials are not good choices for datasets for this assignment. You are welcome to upload datasets of your own as well, and link to them using a third party repository such as github, bitbucket, pastebin, etc. Please be aware of the Coursera terms of service with respect to intellectual property.

Also, you are welcome to preserve data in its original language, but for the purposes of grading you should provide english translations. You are welcome to provide multiple visuals in different languages if you would like!

As this assignment is for the whole course, you must incorporate principles discussed in the first week, such as having as high data-ink ratio (Tufte) and aligning with Cairo's principles of truth, beauty, function, and insight.

Here are the assignment instructions:

- State the region and the domain category that your data sets are about (e.g., **Ann Arbor, Michigan, United States** and **economic activity or measures**).
- You must state a question about the domain category and region that you identified as being interesting.
- You must provide at least two links to available datasets. These could be links to files such as CSV or Excel files, or links to websites which might have data in tabular form, such as Wikipedia pages.
- You must upload an image which addresses the research question you stated. In addition to addressing the question, this visual should follow Cairo's principles of truthfulness, functionality, beauty, and insightfulness.
- You must contribute a short (1-2 paragraph) written justification of how your visualization addresses your stated research question.

What do we mean by **economic activity or measures**? For this category you might look at the inputs or outputs to the given economy, or major changes in the economy compared to other regions.

Tips

- Wikipedia is an excellent source of data, and I strongly encourage you to explore it for new data sources.

- Many governments run open data initiatives at the city, region, and country levels, and these are wonderful resources for localized data sources.
- Several international agencies, such as the United Nations (http://data.un.org/), the World Bank (http://data.worldbank.org/), the Global Open Data Index (http://index.okfn.org/place/) are other great places to look for data.
- This assignment requires you to convert and clean datafiles. Check out the discussion forums for tips on how to do this from various sources, and share your successes with your fellow students!

Example

Looking for an example? Here's what our course assistant put together for the **Ann Arbor, MI, USA** area using **sports and athletics** as the topic. Example Solution File (./readonly/Assignment4_example.pdf)

Solution

This project intends to explore the following questions from the publically available datasets:

- The safety situation in the State of Michigan, are the crimes increasing or decreasing over the years?
- How the crime rates are associated to the Unemployment in youth and overall unemployment in the State?
- Is there any relation between the number of crimes and the average per-capita income of the people of Michigan?

This project uses the datasets from the State of Michigan (Year 2009 to 2016) obtained from the following sources:

1. Michigan Crime Incidents: https://data.world/detroit/dpd-crime-incidents-2009-2016/workspace/query?queryid=sample-0 (https://data.world/detroit/dpd-crime-incidents-2009-2016/workspace/query?queryid=sample-0)
2. Michigan Youth Unemployment Rate: https://data.michigan.gov/Economy/Youth-Unemployment-Rate-16-24-years-old-/tdrw-5jnz (https://data.michigan.gov/Economy/Youth-Unemployment-Rate-16-24-years-old-/tdrw-5jnz)
3. Michigan Unemployment Rate: https://data.michigan.gov/Economy/Unemployment-Rate/kamp-ngb9 (https://data.michigan.gov/Economy/Unemployment-Rate/kamp-ngb9)
4. Michigan Per-Capita Personal Income: https://data.michigan.gov/Economy/Michigan-Per-Capita-Personal-Income/7frj-2rv6 (https://data.michigan.gov/Economy/Michigan-Per-Capita-Personal-Income/7frj-2rv6)

In [215]:

```python
# Load Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Preparing Datasets

In [216]:

```python
# Dataset 1
# Michigan Crime Incidents Dataset (Normalized)

df1 = pd.read_csv('dpdallcrimeincidents-dpd-crime-incidents-2009-2016-QueryResult.csv',
usecols=[2,8], index_col=1,
                names=['Crime Rate', 'Date'], skiprows=1)
df1 = df1.groupby('Date').count()
df1.index = pd.to_datetime(df1.index, format='%Y-%m-%d')
df1 = df1.resample('MS').sum()
df1 = df1 / df1.max()
df1.head()
```

Out[216]:

| | Crime Rate |
Date	
2009-01-01	0.785349
2009-02-01	0.729410
2009-03-01	0.892883
2009-04-01	0.888836
2009-05-01	0.928529

In [217]:

```python
# Dataset 2
# Michigan Youth Unemployment Rate Dataset (Normalized)

df2 = pd.read_csv('Youth_Unemployment_Rate__16_-_24_years_old_.csv', usecols=[0,1], index_col=0,
                names=['Date', 'Youth Unemployment Rate'], skiprows=1)
df2.index = pd.to_datetime(df2.index,format='%b-%y')
df2 = df2[(df2.index.year>2008)&(df2.index.year<2017)][:-6]
df2 = df2 / df2.max()
df2.head()
```

Out[217]:

| | Youth Unemployment Rate |
Date	
2009-01-01	0.834372
2009-02-01	0.839947
2009-03-01	0.847063
2009-04-01	0.895059
2009-05-01	0.907350

In [218]:

```python
# Dataset 3
# Michigan Unemployment Rate Dataset (Normalized)

df3 = pd.read_csv('Unemployment_Rate.csv', usecols=[0,2], index_col=0,
                  names=['Date', 'Overall Unemployment Rate'], skiprows=1)
df3.index = pd.to_datetime(df3.index)
df3 = df3[(df3.index.year>2008)&(df3.index.year<2017)][:-6]
df3 = df3 / df3.max()
df3.head()
```

Out[218]:

Date	Overall Unemployment Rate
2009-01-01	0.746575
2009-02-01	0.821918
2009-03-01	0.883562
2009-04-01	0.931507
2009-05-01	0.972603

In [219]:

```python
df4 = pd.read_csv('Michigan_Per_Capita_Personal_Income.csv', usecols=[0,2], index_col=0
,
                  names=['Date', 'Personal Income'], skiprows=1)
df4.index = pd.to_datetime(df4.index)
df4 = df4 / df4.max()
df4 = df4[(df4.index.year>2008)&(df4.index.year<2018)]
df4 = df4.resample('MS').bfill()[:-7]
df4.head()
```

Out[219]:

Date	Personal Income
2009-01-01	0.702765
2009-02-01	0.730872
2009-03-01	0.730872
2009-04-01	0.730872
2009-05-01	0.730872

The data to be plotted is normalized (values between 0 and 1), so that it can be mapped on the same plane.

```
df = pd.concat([df1, df2, df3, df4], axis=1)
df
```

Out[220]:

| Date	Crime Rate	Youth Unemployment Rate	Overall Unemployment Rate	Personal Income
2009-01-01	0.785349	0.834372	0.746575	0.702765
2009-02-01	0.729410	0.839947	0.821918	0.730872
2009-03-01	0.892883	0.847063	0.883562	0.730872
2009-04-01	0.888836	0.895059	0.931507	0.730872
2009-05-01	0.928529	0.907350	0.972603	0.730872
...	...	...	...	...
2016-02-01	0.581826	0.570417	0.335616	0.955290
2016-03-01	0.692395	0.564508	0.335616	0.955290
2016-04-01	0.681385	0.553089	0.342466	0.955290
2016-05-01	0.712152	0.535791	0.342466	0.955290
2016-06-01	0.560224	0.521609	0.342466	0.955290

90 rows × 4 columns

Plotting

Let us first analyze the Crime Rate data of Michigan and its key time series patterns.

```
ax1 = df1.plot(figsize=(15,8))
df1['Crime Rate'].rolling(12).mean().plot();
ax1.legend(['Crime Rate', 'Crime Rate Trend']);
plt.ylabel('Normalized Parameter Values');
```
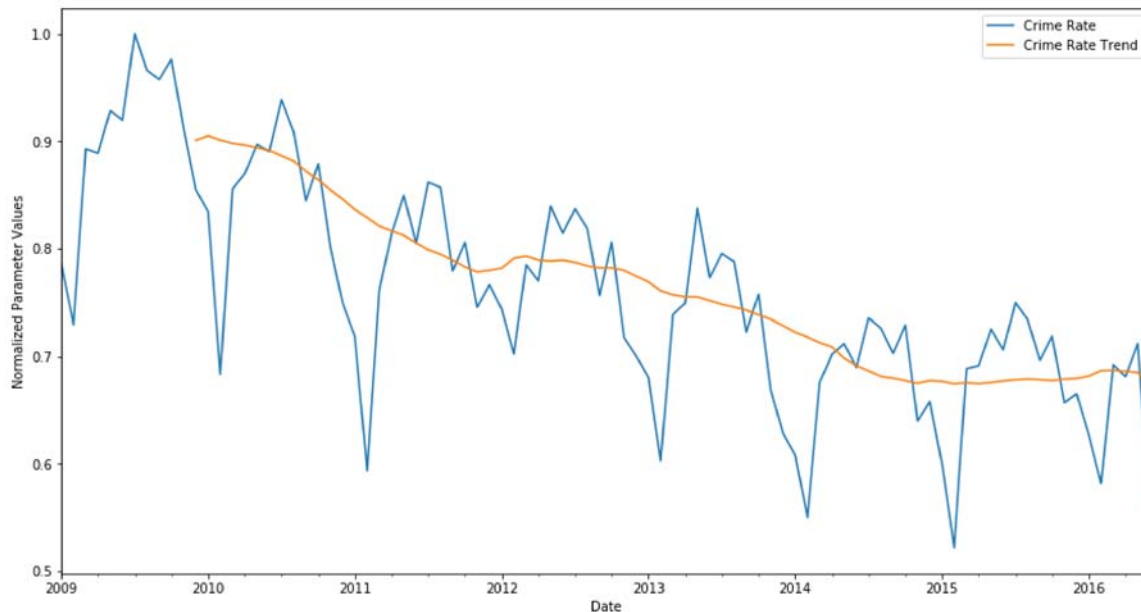

The Crime Rate data shows us a typical time series 'seasonality' and 'trend' patterns. Now lets plot it against the other key 'Economic Factors' of the region.

In [222]:

```
df.plot(style=['-',':','--','-.'], figsize=(15,8), fontsize=12);
plt.ylabel('Normalized Parameter Values');
```
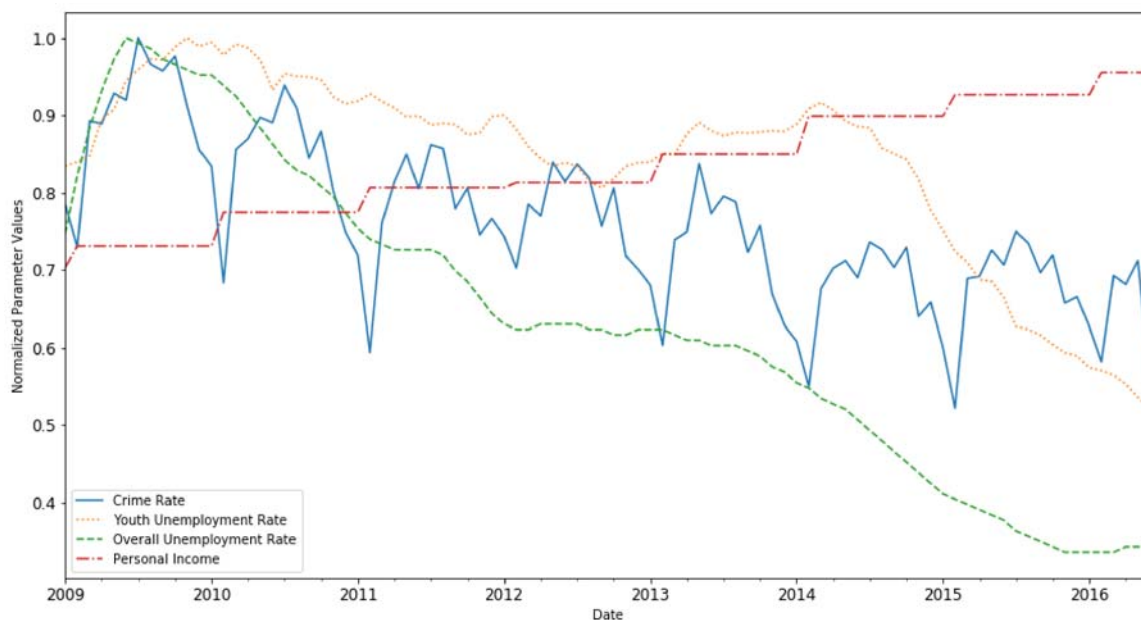
Results

Following observations can be deduced from the above two graphics:

- Crime rates are continuously **on the fall** for the past seven years 2009 to 2016.
- The crime rate trend has **a strong correlation** with the overall unemployment in the state of Michigan. Hence, more the unemployment in the society, it leads to more criminal activities.
- The crime rate trend maintains **positive correlation** with the youth unemployment with the exception from years 2013 to 2015.
- Crime rate is **negatively correlated** to the per-capita personal income of the society. Stronger financial status of the masses has in turn resulted in lesser crimes over the years.

In [ ]:

In [ ]:

In [ ]: