Econometrics 322 Lab #2

Introduction to Data Visualization in Econometrics

Prof. Paczkowski

Enter your Name in the Next Cell

Maanya Tandon

Grading Rubric

Score:___: Max(0, 20 - Total Deductions)

Content Area	Deduction	Times Deducted	Check	Comments
Abstract				
Missing	5		[ ]	
Insufficient/Wrong Focus	1		[ ]	
Data Dictionary (Metadata)				
Missing	5		[ ]	
Insufficient/Wrong Form or Wording	1		[ ]	
Graphs				
Missing	5		[ ]	
Missing Title	1		[ ]	
Missing/Wrong Labels	1		[ ]	
Pre-Lab				
Missing	5		[ ]	
Insufficient/Wrong Answer	2 Each		[ ]	
No/Incorrect/Insufficient Model Specification	2		[ ]	
No/Incorrect Statistical Hypothesis Statement	2 Each		[ ]	
Post-Lab				
Missing	5		[ ]	
Insufficient/Wrong Answer	2 Each		[ ]	
Correlations				
Missing	5		[ ]	
Insufficient/Wrong Analysis	2		[ ]	
Missing Graph	2		[ ]	
Estimations				
Missing	5		[ ]	
No or incorrect discussion/interpretation of...				
Hypothesis tests and p-values	2 Each		[ ]	
R^2	2		[ ]	
F-Statistic	2		[ ]	
Multicollinearity/VIF	2		[ ]	
Heteroskedasticity/Test	2		[ ]	
Autocorrelation/Test	2		[ ]	
No/insufficient model selection	2		[ ]	
Elasticities				
Missing	5		[ ]	
Incorrect Interpretation	2		[ ]	
Missing Summary Table	2		[ ]	
Model Portfolio			[ ]	
Missing	5		[ ]	
General Comments:				

Contents

Collaboration Policy

Back to Contents

```
1. Study groups are allowed but I expect students to understand and complete their own assignmen
ts and to hand in one
assignment per student.
2. If you worked in a group, please put the names of your study group in the following table.
3. Just like all other classes at Rutgers, the student Honor Code is taken seriously.

The submitted assignment must be your work.
```

Collaborator(s) Name(s)
name(s) here

Introduction

Back to Contents

A powerful, but often overlooked tool of fundamental data analysis is the graph. Economists tend to use simple graphs to portray their data, but spend very little time analyzing the graphs for messages about the *Data Generating Process* (*DGP*).

We can group graph types into two categories consistent with the time/spatial domain of the data:

1. time series plots; and
2. scatter plots.

A third type is purely descriptive and cuts across the time/spatial domains. Examples of these graphs are:

1. boxplots; and
2. histograms.

Boxplots are extremely powerful graphs that robustly summarize the distribution of the data. I often use these in my classes. Histograms are also very useful, but controversial.

Time series plots are self-evident. The horizontal axis is the time span. Since the nature of this axis is clear, you do not have to label it "Time". The vertical axis, on the other hand, must always be clearly labeled with the units.

Scatter plots are a plot of two series.

Always put a title on a graph to clearly label it. Do not use the variable mnemonics in a title or axis label since these are not always clear. Be clear, precise, and succinct.

Purpose

Back to Contents

This lab will introduce you to the use of graphs and their analysis. These graphs will be required throughout the course.

At the end of this lab, you will be able to:

- build appropriate graphs for your data;
- look for:
  - relationships;
  - trends;
  - patterns; and
  - anomalies;
- interpret graphs.

Problem

Back to Contents

The stock market is always in the news. Everyone has been affected by the almost overnight collapse of stock portfolios in 1997. As economists, we would like to understand the relationship between key economic measures of the **health of the economy** and the **stock market**. The focus is the stock market. In this lab, you will explore, graphically, these relationships.

It is very important that you clearly label all axes and put a title on the graphs. Points will be deducted for anything that is missing. In addition, points will be deducted if your title or labels are just the variable name; the variable names are unacceptable.

Assignment

Back to Contents

Collect monthly data on the following:

- Industrial Production Index
- Unemployment Rate
- Moody's Seasoned Aaa Corporate Bond Yield

The S&P 500 monthly index is available on Sakai. Use the Adjusted Close variable for this lab. You can collect the other three series at FRED. Use the *Search* feature at the top right of the FRED home page. Download monthly data and build a data set of the four series. Make sure the months line-up. Use as much data as you can collect that have common months and years.

Documentation

Back to Contents

Abstract

Back to Contents

This lab helped me understand how the S&P monthly 500 index, the unemployment rate, corporate bond yield, and the industrial production rate all affect the health of the economy. Plotting the graphs was essential to see cycles of economic stress and economic growth and how the change in these 4 variables affect the general health of the economy.

Data Dictionary

Back to Contents

Variable	Values	Source	Mnemonic
Unemployment Rate (monthly)	Percent	FRED	UNRATE
Industrial Production Total Index (monthly)	Index, Year2012=100	FRED	INDPRO
Moody's Seasoned Aaa Corporate Bond Yield (monthly)	Percent	FRED	AAA
S&P 500 Index (monthly)	Index	SAKAI/S&P	SPX

Pre-lab Questions

Back to Contents

Before you do any work, please think about the relationship among these macro variables. In particular, think how you would answer the following if called on in class.

What type of data is this and why (i.e., source and domain)?

Since I would like to understand the relationship between key economic measures of the health of the economy and the **stock market**, we have the following data. The data is all secondary data source collected from FRED and Sakai. we have following data:

- Industrial Production Index (INDPRO): This data from FRED represnts the output of the industry. Higher the index, better the industry is doing in terms of production.
- Unemployment Rate (UNRATE): This data from FRED reprsents percentage of people unemployed in the economy.
- Moody's Seasoned Aaa Corporate Bond Yield (AAA): This data from FRED reprsents average yield of the highest rated (least risky) AAA corporate bonds - and it is published by Moody's. In the times of economic distress the corporate bond yield goes higher as the credit-worthyness of the issuers deteriorates and the their bond prices go down.
- S&P 500 Index (SPX): The data is published by Satndard & Poor, and it reprsnts the market-capitalization weighted index of the share prices for the top 500 public companies in the US. The higher is SPX, the better the **stock market** is performing.

What pattern do you expect to see for stock prices? Explain your answer.

I expect S&P 500 Index to tank during the times of economic distress such as in October 1987, burst of dot-com bubble in 2001-02, and financial crisis of 2008-09. When business cycle is not in a recession, or trough, I expect S&P 500 to be going up.

What relationship would you expect to see between stock prices and unemployment and industrial production? Explain your answer.

I expect stock prices to be positively related to industrial production and negatively corrleated with unemployment rate.

When economy is growing and stock prices goes up, the industrial production goes up because of high demand for products. In recessions, however, I expect stock prices and industrial production - both to go down.

When economy is growing, stock prices go up, and unemployment goes down becasue a plenty of jobs become available. However, stock prices can also go up when economy becomes more efficient -- i.e. labor productivity goes up -- which means that employees may be getting laid off. Hence the relationship between stock prices and unemployment rate is not very strong.

How do you think these variables will change over time? Do you think they will grow or decline or stay constant? Explain your answer.

- Industrial Production Index (INDPRO): Since I expect the industry to keep growing in size over long-term, I expect Industrial Production Index to display a upward trend over long-term. But it should temporarily go down when economy is under stress.
- Unemployment Rate (UNRATE): I expect unemployement to go up when economy is under stress, and go down during normal periods. But there should not be a long-term trend in it.
- Moody's Seasoned Aaa Corporate Bond Yield (AAA): I expect bond yields to go up when economy is under stress, and go down during normal periods. But there should not be a long-term trend in it.
- S&P 500 Index (SPX): Since I expect the industry and businesses to keep growing in size over long-term, I expect S&P 500 Index to display a upward trend over long-term. But it should temporarily go down when economy is under stress.

What are good testable hypotheses about what drives or determines the stock market?

Since S&P 500 Index and Indusrial Production Index are upward trending over the long-term, but other variables are not, I should look at the %-change in these two variables. However, the issue is, what lag I should use to calculate %-change? Monthly, quarterly, or annual? I will limit ourseleves to monthly %-changes.

Correlation between S&P 500 Index %-change (SPXpc) and AAA bond yields %-change (AAApc) should be highly negative:

$$Hypothesis 1. H_0 : correl(SPXpc, AAApc) < -0.60.$$

Correlation between S&P 500 Index %-change (SPXpc) and Industrial Production Index %-change I(NDPROpc) should be highly positive:

$$Hypothesis 2. H_0 : correl(SPXpc, INDPROpc) > 0.60.$$

Correlation between S&P 500 Index (SPX) and Industrial Production Index (INDPRO) should be reasonbaly positive:

$$Hypothesis 3. H_0 : correl(SPX, INDPRO) > 0.30$$

Correlation between S&P 500 Index %-change (SPXpc) and Unemployment Rate (UNRATE) should be resonably negative:

$$Hypothesis 4. H_0 : correl(SPXpc, UNRATE) < -0.30.$$

Correlation between S&P 500 Index (SPX) and Unemployment Rate (UNRATE) should be low negative correlation:

$$Hypothesis 5. H_0 : correl(SPX, UNRATE) < -0.30;$$

Write a tentative specific model for the stock market. Explain why you wrote the model this way.

Stock market may be leading or lagging the other measures of the economy - and if so what would be the length of the lag? I will ignore the lead/lag issue while writing the model.

Since SPX & INDPRO are long-term trending variable - hence I take monthly %-change in these two ($SPXpc$, $INDPROpc$). Since UNRATE would have no long-term trend, I do not include %-change in it in the following regression relationship for S&P 500 Index.

$$SPXpc = \beta_0 + \beta_1 * INDPROpc + \beta_2 * UNRATE.$$

or, if I only relate the variables that have a long-term trend:

$$SPX = \beta_0 + \beta_1 * INDPRO.$$

Since I do not know if the above relationship would be linear, I can conside the following alternative,

$$SPX = \beta_0 + \beta_1 * f(INDPRO).$$

The function f would take care of the non-linearity in the relationship, and it would emerge after I look at the plots of the two variables and their scatter plots.

Write the statistical hypotheses to go along with your testable hypotheses. Define the terms.

Since I do not have the probability distribution for correlations - I cannot test hypothesis on correlations statistically. However I can test hypothesis on stock market monthly returns (S&P 500 Index monthly %-change) - since I can calculate the time-seirs for it. So I plan to invetigate if on average of the S&P 500 Index goes up by 0.5% on a monthly basis, or

$$H_0 : E[SPXpc] = 0.5\%$$

with confidence level $\alpha = 0.05$.

Tasks and Questions

Back to Contents

Load the necessary packages.

In [2]:

```
##
## Analysis
import pandas as pd
import numpy as np
import statsmodels.api as sm
import statsmodels.stats.weightstats as ztest

## Visualization
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
## sns.set()
sns.set(rc={'figure.figsize':(11.7,8.27)})
##
```

Import the data.

Be sure to include the S&P 500. Insert more code cells if necessary.

In [4]:

```
## Enter the code here for the S&P 500
data_path = '/Users/maanyatandon/Documents/Maanya/Econometrics/Lab2/'
df_SPX    = pd.read_csv( data_path+'SP500.csv', parse_dates = ['Date'])
# rename column
df_SPX.rename(columns={"Date": "DATE"})
## Set 'Date' as index
df_SPX.set_index( 'Date', inplace = True)
display( df_SPX.head() )
## display( df_SPX.tail() )
## print( 'Shape: {}'.format(df_SPX.shape) )
```

Date	Open	High	Low	Close	Adj Close	Volume
1960-01-01	59.910000	60.389999	55.610001	55.610001	55.610001	63920000.0
1960-02-01	55.959999	56.820000	54.730000	56.119999	56.119999	60500000.0
1960-03-01	56.009998	56.009998	53.470001	55.340000	55.340000	65710000.0
1960-04-01	55.430000	56.590000	54.369999	54.369999	54.369999	57300000.0
1960-05-01	54.130001	55.830002	54.130001	55.830002	55.830002	68790000.0

In [6]:

```
## Enter the code here for the other series
## In 'AAA_mod.csv', data before 1948-01-01 was deleted
df_AAA    = pd.read_csv( data_path+'AAA_mod.csv', parse_dates = ['DATE'])
df_AAA.set_index( 'DATE', inplace = True)   # Set 'DATE' as index
## df_AAA.index = df_AAA.index.DATE
display( df_AAA.head() )
## display( df_AAA.tail() )
## print( 'Shape: {}'.format(df_AAA.shape) )
## a_dataframe.drop( a_dataframe[a_dataframe.B > 3].index, inplace=True)
##  df_AAA.drop( df_AAA[ 'DATE' < '1948-01-01'], inplace=True)
## df_AAA.head()
```

DATE	AAA
1948-01-01	2.86
1948-02-01	2.85
1948-03-01	2.83
1948-04-01	2.78
1948-05-01	2.76

In [7]:

```
df_INDPRO    = pd.read_csv( data_path+'INDPRO.csv', parse_dates = ['DATE'])
df_INDPRO.set_index( 'DATE', inplace = True)   # Set 'DATE' as index
display( df_INDPRO.head() )
## display( df_INDPRO.tail() )
## display( df_INDPRO.shape )
```

DATE	INDPRO
1919-01-01	5.0124
1919-02-01	4.7908
1919-03-01	4.6524
1919-04-01	4.7355
1919-05-01	4.7632

In [8]:

```
df_UNRATE    = pd.read_csv( data_path+'UNRATE.csv', parse_dates = ['DATE'])
df_UNRATE.set_index( 'DATE', inplace = True)   # Set 'DATE' as index
display( df_UNRATE.head() )
## display( df_UNRATE.tail() )
## print( 'Shape: {}'.format(df_UNRATE.shape) )
```

DATE	UNRATE
1948-01-01	3.4
1948-02-01	3.8
1948-03-01	4.0
1948-04-01	3.9
1948-05-01	3.5

In [9]:

```
## merge the two data sets.
## Call the final DataFrame df
##
## df_all = pd.merge( df_SPX, df_AAA,    how = 'inner', left_index = True, right_index = True)
df_all = pd.merge( df_SPX, df_AAA,    how = 'inner', left_index = True, left_on = 'Date', right_index=T
rue, right_on = 'Date')
## df_all.set_index('DATE', inplace=True)
## df_all.index
## df_all[ 'Date'] = df_all.index
df_all = pd.merge( df_all, df_UNRATE, how = 'inner', left_index = True, right_index = True)
df_all = pd.merge( df_all, df_INDPRO, how = 'inner', left_index = True, right_index = True)
```

Print the first five (5) records of *df*.

In [10]:

```
display( df_all.head() )
## display( df_all.tail() )
## print( 'Shape: {}'.format(df_all.shape) )
```

	Date	Open	High	Low	Close	Adj Close	Volume	AAA	UNRATE	INDPRO
1960-01-01	1960-01-01	59.910000	60.389999	55.610001	55.610001	55.610001	63920000.0	4.61	5.2	24.8958
1960-02-01	1960-02-01	55.959999	56.820000	54.730000	56.119999	56.119999	60500000.0	4.56	4.8	24.6743
1960-03-01	1960-03-01	56.009998	56.009998	53.470001	55.340000	55.340000	65710000.0	4.49	5.4	24.4528
1960-04-01	1960-04-01	55.430000	56.590000	54.369999	54.369999	54.369999	57300000.0	4.45	5.2	24.2589
1960-05-01	1960-05-01	54.130001	55.830002	54.130001	55.830002	55.830002	68790000.0	4.46	5.1	24.2312

Plot the stock prices with time on the x-axis.

In [49]:

```
## Plot 1: SPX vs time
ax = sns.lineplot( y='Adj Close', x= 'Date', data=df_all )
ax.set( title='Plot 1. S&P 500 Index Time Series', xlabel= '', ylabel= 'Adj Close');
```

Plot the other data series using your own judgement. Use your imagination and creativity.

Insert more code cells if necessary.

In [67]:

```
##  Plot 2: Other variables  vs time
ax_AAA = sns.lineplot( y='AAA', x= 'Date', data=df_all, legend='full', label='AAA Bond Yield', color='b
lue' )
#ax_AAA.set( title='AAA Bond Yield, Unemployment Rate & Ind. Production Indx', xlabel= '', ylabel= 'Yie
ld');
ax_UNRATE = sns.lineplot( y='UNRATE', x= 'Date', data=df_all, legend='full', label='Unemployment Rate',
color='orange')
#ax_UNRATE.set( title='', xlabel= '', ylabel= 'Unemp Rate & Bond Yield');
ax2 = ax_AAA.twinx()
ax_INDPRO = sns.lineplot( y='INDPRO', x= 'Date', data=df_all, legend='full', label='Ind. Production Ind
x', color='green',ax=ax2)
ax_INDPRO.set( title='Plot 2. AAA Bond Yield, Unemployment Rate & Ind. Production Index Time Series', x
label= '', ylabel= 'Ind. Production Indx ');
```

In [66]:

```
## Plot 3: Plot ALL in one graph with legend
Date = df_all['Date']
df_preproc = pd.DataFrame({
    'Date': Date,
    'Unemployment Rate x10':   df_all['UNRATE']*10,
    'AAA Bond Yield x10':      df_all['AAA']*10,
    'Ind. Production Index':   df_all['INDPRO'],
    'S&P 500 Index / 20':      df_all['Adj Close'] / 20 })
## df_preproc.head()
df_preproc2 = pd.melt( df_preproc, ['Date'])
ax_preproc2 = sns.lineplot( x='Date', y='value',hue='variable', data=df_preproc2)
ax_preproc2.set( title='Plot 3. S&P 500 Index, AAA Bond Yields, Unemployment Rate, Industrial Productio
n Index Time Series');
```

In [36]:

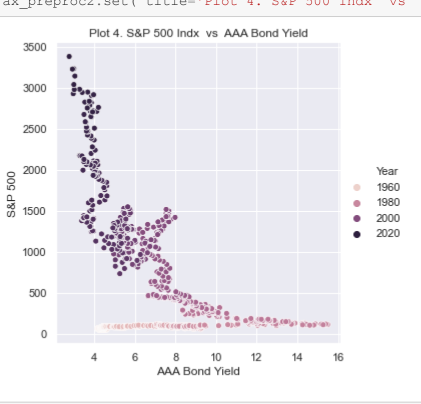
```
## Create df with only the variables that we will use.
df_preproc3 = pd.DataFrame({
    'Date':                    df_all.index,
    'Year':                    df_all.index.year,
    'Unemployment Rate':       df_all['UNRATE'],
    'AAA Bond Yield':          df_all['AAA'],
    'Ind. Production Indx':    df_all['INDPRO'],
    'S&P 500':                 df_all['Adj Close'],
    'Unemployment Rate Pct Chg':      df_all['UNRATE'].pct_change(),
    'AAA Bond Yield Pct Chg':         df_all['AAA'].pct_change(),
    'Ind. Production Indx Pct Chg':   df_all['INDPRO'].pct_change(),
    'S&P 500 Pct Chg':                df_all['Adj Close'].pct_change() })
df_preproc3.head()
```

Out[36]:

	Date	Year	Unemployment Rate	AAA Bond Yield	Ind. Production Indx	S&P 500	Unemployment Rate Pct Chg	AAA Bond Yield Pct Chg	Ind. Production Indx Pct Chg	S&P 500 Pct Chg
1960-01-01	1960-01-01	1960	5.2	4.61	24.8958	55.610001	NaN	NaN	NaN	NaN
1960-02-01	1960-02-01	1960	4.8	4.56	24.6743	56.119999	-0.076923	-0.010846	-0.008897	0.009171
1960-03-01	1960-03-01	1960	5.4	4.49	24.4528	55.340000	0.125000	-0.015351	-0.008977	-0.013899
1960-04-01	1960-04-01	1960	5.2	4.45	24.2589	54.369999	-0.037037	-0.008909	-0.007930	-0.017528
1960-05-01	1960-05-01	1960	5.1	4.46	24.2312	55.830002	-0.019231	0.002247	-0.001142	0.026853

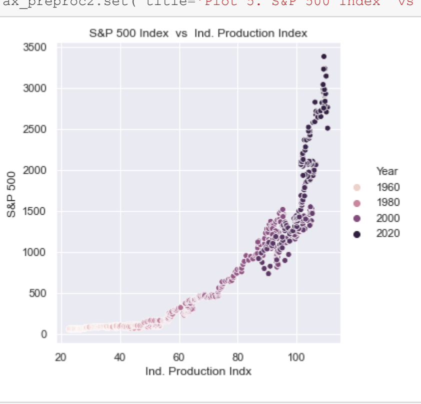
In [52]:

```
# Plot 4: S&P 500 vs AAA yield
ax_preproc2 = sns.relplot( y='S&P 500', x='AAA Bond Yield', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 4. S&P 500 Indx  vs  AAA Bond Yield');
```



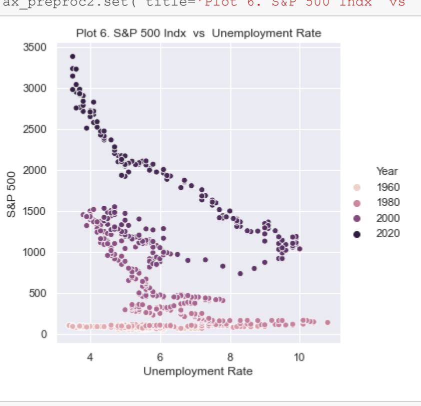
In [38]:

```
# Plot 5: S&P 500 vs 'Ind. Production Index'
ax_preproc2 = sns.relplot( y='S&P 500', x='Ind. Production Indx', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 5. S&P 500 Index  vs  Ind. Production Index');
```



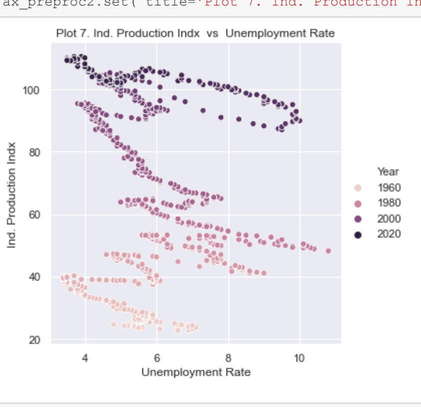
In [53]:

```
## Plot 6: S&P 500 vs 'Unemployment Rate'
ax_preproc2 = sns.relplot( y='S&P 500', x='Unemployment Rate', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 6. S&P 500 Indx  vs  Unemployment Rate');
```



In [54]:

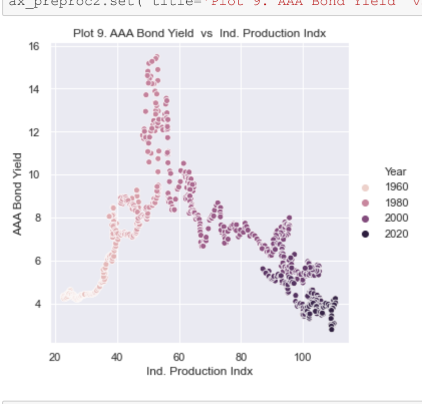
```
## Plot 7:  Industrial Production  vs Unemployment Rate
ax_preproc2 = sns.relplot( y='Ind. Production Indx', x='Unemployment Rate', hue='Year',data=df_preproc3
)
ax_preproc2.set( title='Plot 7. Ind. Production Indx  vs  Unemployment Rate');
```



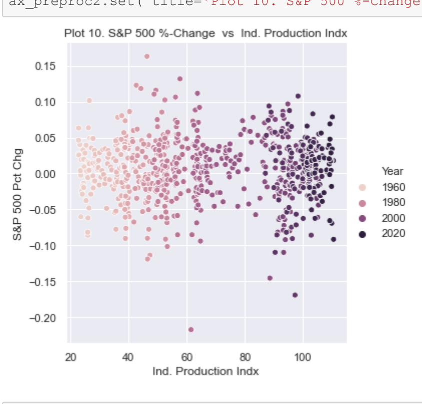
In [55]:

```
## Plot 8:  AAA Bond Yield  vs Unemployment Rate
ax_preproc2 = sns.relplot( y='AAA Bond Yield', x='Unemployment Rate', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 8. AAA Bond Yield  vs  Unemployment Rate');
```

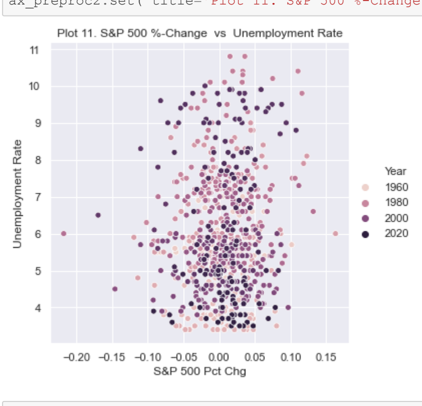
In [56]:
```python
## Plot 9:  AAA Bond Yield  vs Ind. Production Indx
ax_preproc2 = sns.relplot( y='AAA Bond Yield', x='Ind. Production Indx', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 9. AAA Bond Yield  vs  Ind. Production Indx');
```



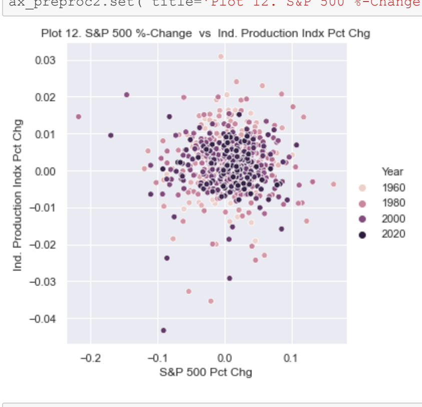
In [60]:
```python
## Plot 10: Relation between Stock Prices PERCENT CHANGE and Ind. Production Indx
ax_preproc2 = sns.relplot( y='S&P 500 Pct Chg', x='Ind. Production Indx', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 10. S&P 500 %-Change  vs  Ind. Production Indx');
```



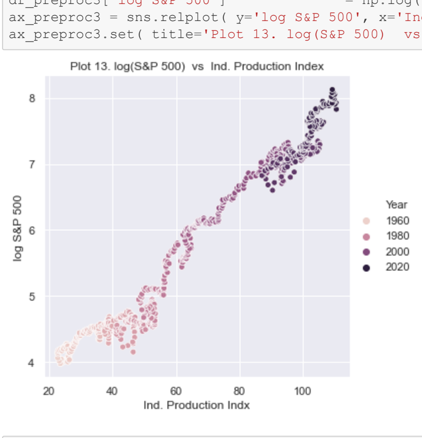
In [58]:
```python
## Plot 11: Relation between Stock Prices PERCENT CHANGE and Unemployment Rate
ax_preproc2 = sns.relplot( x='S&P 500 Pct Chg', y='Unemployment Rate', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 11. S&P 500 %-Change  vs  Unemployment Rate');
```



In [59]:
```python
## Plot 12: Relation between Stock Prices PERCENT CHANGE and Ind. Production Indx PERCENT CHANGE
ax_preproc2 = sns.relplot( x='S&P 500 Pct Chg', y='Ind. Production Indx Pct Chg', hue='Year',data=df_preproc3)
ax_preproc2.set( title='Plot 12. S&P 500 %-Change  vs  Ind. Production Indx Pct Chg');
```

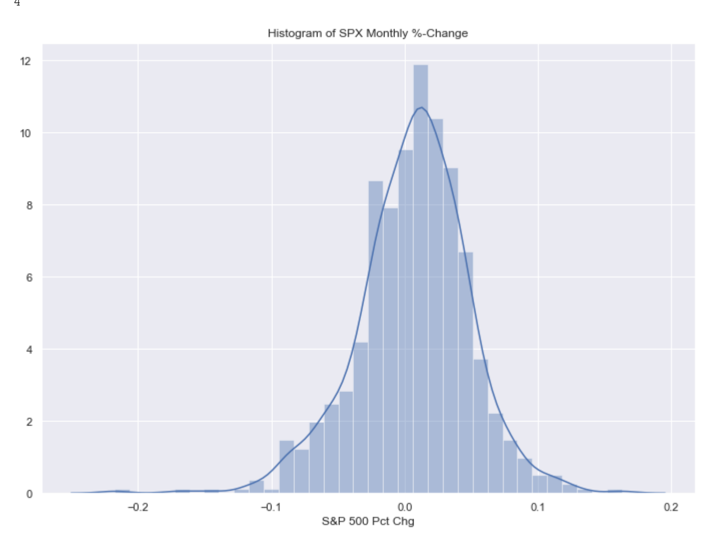


In [62]:
```python
## Plot 13. LOG S&P 500  vs  Industrial Production Index
## df_preproc3['log Ind. Production Indx'] = np.log( df_preproc3['Ind. Production Indx'] )
df_preproc3['log S&P 500']              = np.log( df_preproc3['S&P 500'] )
ax_preproc3 = sns.relplot( y='log S&P 500', x='Ind. Production Indx', hue='Year', data=df_preproc3)
ax_preproc3.set( title='Plot 13. log(S&P 500)  vs  Ind. Production Index');
```



In [78]:
```python
## Plot 14: Histogram of SPX %-change
ax = sns.distplot( df_preproc3['S&P 500 Pct Chg'] ).set_title('Histogram of SPX Monthly %-Change')
SPXpc_mean   = df_preproc3['S&P 500 Pct Chg'].mean()
SPXpc_stdev  = df_preproc3['S&P 500 Pct Chg'].std()
n            = df_preproc3['S&P 500 Pct Chg'].count()
Z_score      = (SPXpc_mean - 0.005) / (SPXpc_stdev / np.sqrt(n))
print('SPXpc_mean ' + repr(SPXpc_mean) + ', SPXpc_stdev ', repr(SPXpc_stdev), ', n', repr(n), ', Z_score ', repr(Z_score))
```

```
SPXpc_mean 0.006600524840935346, SPXpc_stdev  0.04196180452040785 , n 721 , Z_score  1.024179066902534
```



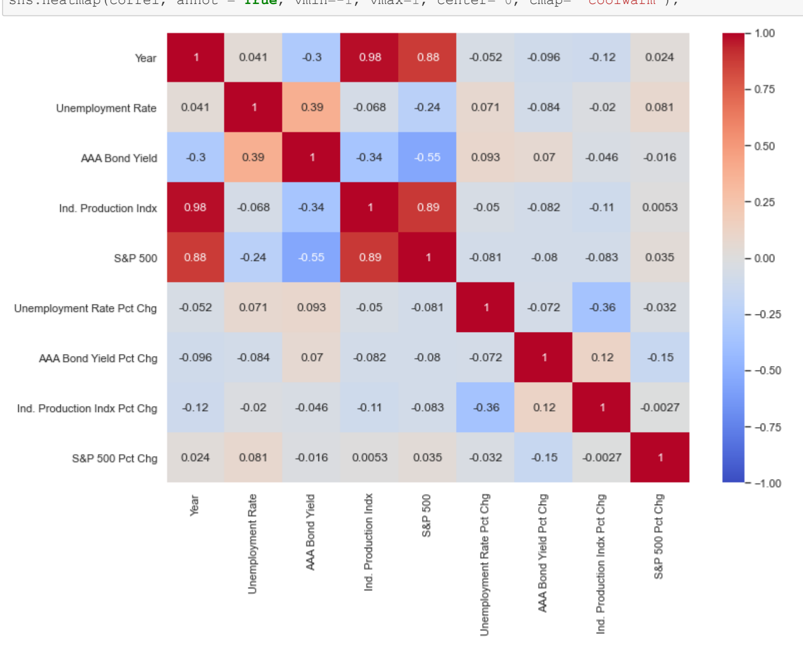
Print and graph a correlation matrix.

In [48]:
```python
correl = df_preproc3.corr()
correl
```

Out[48]:

	Year	Unemployment Rate	AAA Bond Yield	Ind. Production Indx	S&P 500	Unemployment Rate Pct Chg	AAA Bond Yield Pct Chg	Ind. Production Indx Pct Chg	S&P 500 Pct Chg
Year	1.000000	0.040900	-0.304497	0.983368	0.882419	-0.051745	-0.096416	-0.115389	0.024267
Unemployment Rate	0.040900	1.000000	0.390361	-0.068011	-0.241949	0.070741	-0.083642	-0.020170	0.080893
AAA Bond Yield	-0.304497	0.390361	1.000000	-0.344313	-0.551204	0.093165	0.070047	-0.045796	-0.015885
Ind. Production Indx	0.983368	-0.068011	-0.344313	1.000000	0.888460	-0.050061	-0.082278	-0.105406	0.005324
S&P 500	0.882419	-0.241949	-0.551204	0.888460	1.000000	-0.081079	-0.079910	-0.083400	0.034790
Unemployment Rate Pct Chg	-0.051745	0.070741	0.093165	-0.050061	-0.081079	1.000000	-0.072358	-0.362041	-0.032224
AAA Bond Yield Pct Chg	-0.096416	-0.083642	0.070047	-0.082278	-0.079910	-0.072358	1.000000	0.120795	-0.151324
Ind. Production Indx Pct Chg	-0.115389	-0.020170	-0.045796	-0.105406	-0.083400	-0.362041	0.120795	1.000000	-0.002719
S&P 500 Pct Chg	0.024267	0.080893	-0.015885	0.005324	0.034790	-0.032224	-0.151324	-0.002719	1.000000

In [61]:
```python
sns.heatmap(correl, annot = True, vmin=-1, vmax=1, center= 0, cmap= 'coolwarm');
```



Post-lab Questions

Back to Contents

What patterns do you see in stock prices?

From year 1960 to 2000, stocks were going up exponentially. Then in 2001, S&P 500 crashed from around 1500 to 800 level because of the dot-com bust. Then again stocks went up exponentially from year to year 2007. In year 2008 financial crisis, S&P 500 crashed agin from 1500 to 800. After this, S&P 500 has been going up exponentially till February 2020 - with minor blips in between.

What patterns or relationships do you see in your graphs?

From Plot 3, S&P 500 Index goes down - when Industrial Production Index goes down and Employment Rate goes down. I specially notice it for 2001-02 and 2008-09. Also from Plot 3, AAA Bond yield goes up - when Industrial Production Index goes down and Employment Rate goes down. I specially notice it for 19282-83, and 2001-02.

From Plot 4, AAA bond yields is highly correlated with unemployment rates; also from Plot 3, I see that the peaks of Unemployment Rate are lagging peaks of AAA bond yields by a year or two. However this relationship did not seem to hold in 1960's.

From scatter Plot 4, S&P500 Index is higly negatively related with AAA Bond Yields - with Bond yields going up when S&P 500 index is low.

From scatter Plot 5, S&P500 Index is higly positively related with Industrial Production Index - with both the indexes going up together.

From scatter Plot 6, S&P500 Index is higly negatively related with Unemployment Rate - with S&P 500 Index rising with decreasing Unmployment Rate.

From scatter Plot 8, AAA Bond Yield is higly positively related with Unemployment Rate - with AAA Bond Yield rising with increasing Unemployment Rate.

From scatter Plot 9, AAA Bond Yield was higly positively related with Industrial Production Index in 1960's - while after 1980's it reversed - and I see that a rise in AAA Bond Yield is associated with decreasing Unemployment Rate in later years.

In Plots 10, 11 and 12, I do scatter plots of monthly %-Change in S&P 500 Index against:

1. Ind. Production Index,
2. Unemployment Rate,
3. Monthly %-Change in Ind. Production Index

From these three plots, I see that the relationshsip, if any, gets destroyed when I use monthly %-changes in the economic variables!

Plots also show that when industrial production index goes down, unemployment rates goes up - as in years 1974-75, 1982-83, 2001-03 and 2008-09.

From Plot 13, it seems that log(S&P 500 Index) is linearly related with Industrial Production Index, i.e.,

$$log(SPX) = \alpha + \beta * INDPRO$$

or, I possibly have the realtionship:

$$SPX = \alpha_1 * exp(\beta * INDPRO)$$

What can you observe about the correlation matrix? Explain.

I see a resonable postive correlation of 0.39 between *AAA bond yields* and *unemployment rates*.
I see a resonable negative correlation of -0.34 between *AAA bond yields* and *industrial production index*.
Also as expected, I have a negative correlation of -0.24 between *S&P 500* and *unemployment rates*.
As is expected, between *S&P 500* and *industrial production index* - I see a high positive correlation of 0.89.
Also as expected, I see a negative correlation of -0.55 between *S&P 500* and *AAA yields*.
Industrial production index seems to lag *unemployment rate* by a year or so - that is why I see a low negative correlation of -0.06 between the two variables. However when I switch to %-changes, I see a decent negative correlation of -0.36 between *%-change in Industrial production* and *%-change in unemployment rate*.

Do you see anything to support your testable hypothesis? Explain.

Correlation between S&P 500 Index %-change (SPXpc) and AAA bond yields %-change (AAApc):

$$Hypothesis 1. H_0 : correl(SPXpc, AAA) < -0.60;$$

$$Actual = -0.15$$

Correlation between S&P 500 Index %-change (SPXpc) and Industrial Production Index %-change:

$$Hypothesis 2. H_0 : correl(SPXpc, INDPROpc) > 0.60;$$

$$Actual = -0.0027$$

Correlation between S&P 500 Index (SPX) and Industrial Production Index (INDPRO):

$$Hypothesis 3. H_0 : correl(SPX, INDPRO) > 0.30;$$

$$Actual = 0.89$$

Correlation between S&P 500 Index %-change (SPXpc) and Unemployment Rate (UNRATE):

$$Hypothesis 4. H_0 : correl(SPXpc, UNRATE) < -0.30;$$

$$Actual = 0.08$$

Correlation between S&P 500 Index (SPX) and Unemployment Rate (UNRATE):

$$Hypothesis 5. H_0 : correl(SPX, UNRATE) < -0.30;$$

$$Actual = -0.36$$

Test of Statistical Hypothesis
In Plot 14 cell, monthly returns on S&P 500 index seem to be normally distributed; so for this variable, I calculate a mean (SPXpc_mean) = 0.66%, and standard deviation (SPXpc_stdev) = 4.196%, using n=721 sample points. Thus to test the null hypothesis, that the mean monthly return on S&P 500 is 0.5%, I calculate Zscore as:

Zscore = (SPXpc_mean - 0.005) / (SPXpc_stdev / sqrt(n)) = 1.0241.

Since Zscore = 1.0241 is outside the rejection region of z > 1.645 for $\alpha = 0.05$ level of confidence, I cannot reject the null hypothesis that mean of long-term monthly returns on S&P 500 Index is 0.5%.

Well done!

Make sure your name is on this notebook at the top and on the file.
Please submit this notebook as a PDF file. Nothing else will be accepted.

In [ ]: