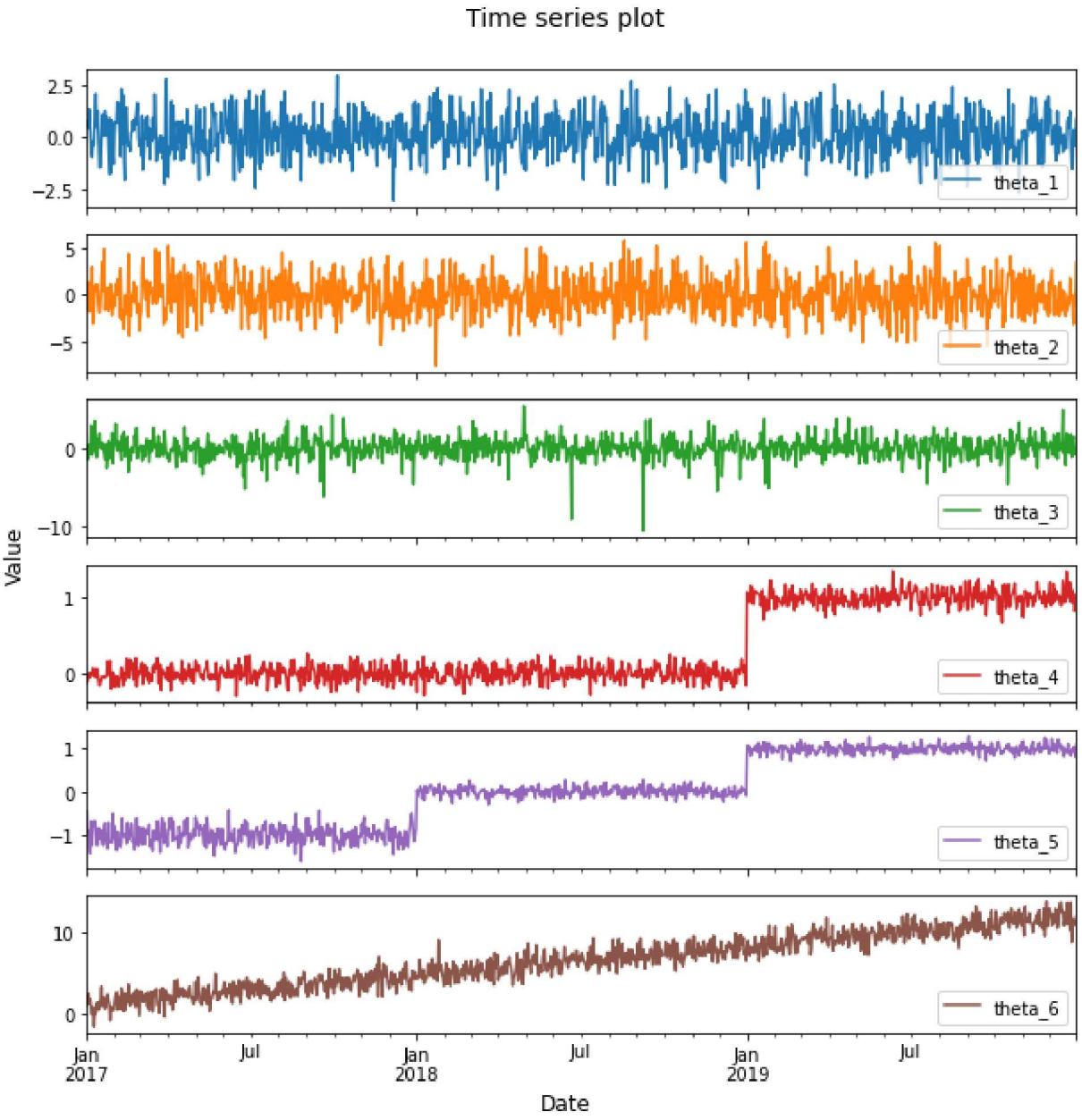
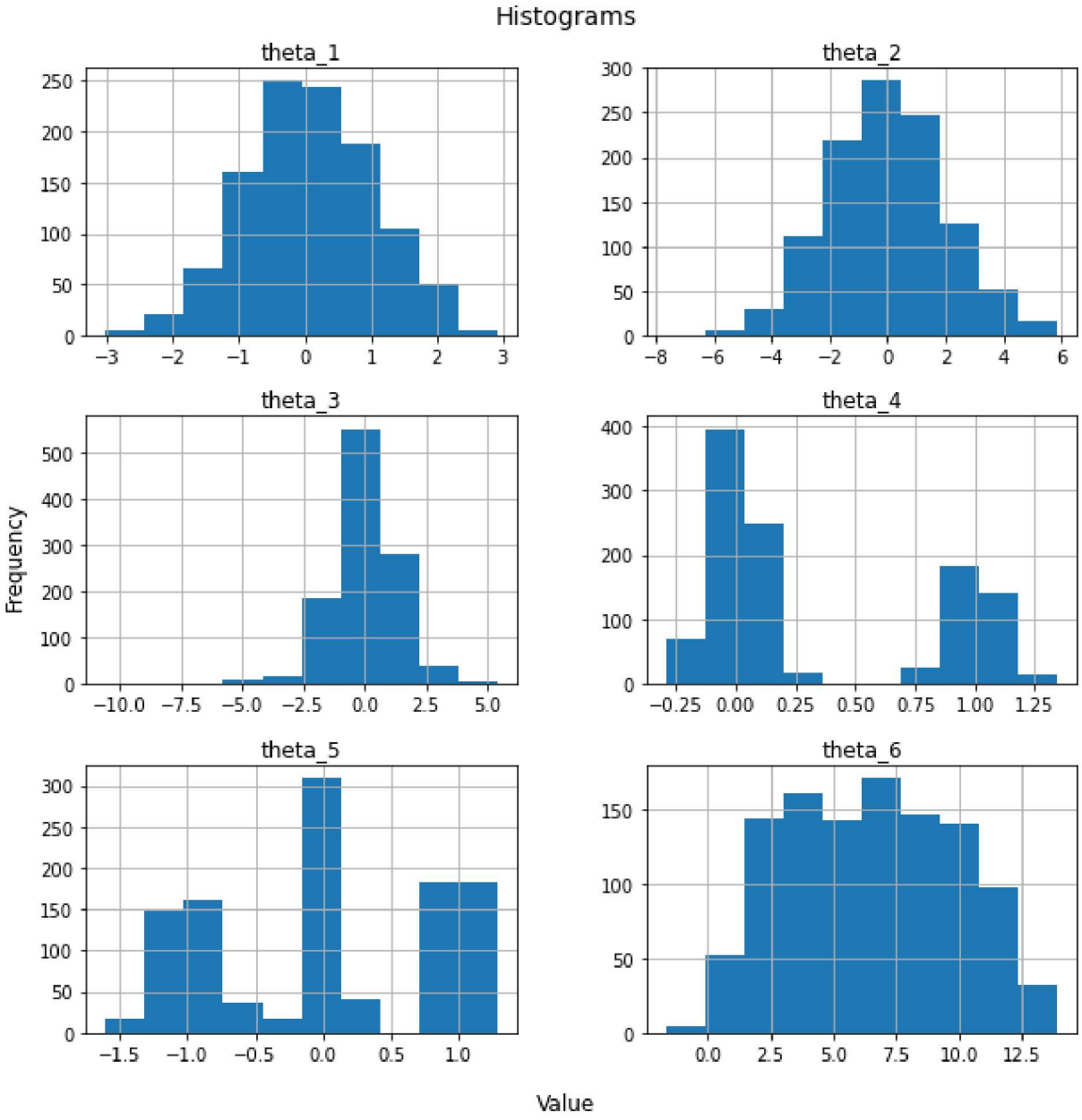
Exercise 1

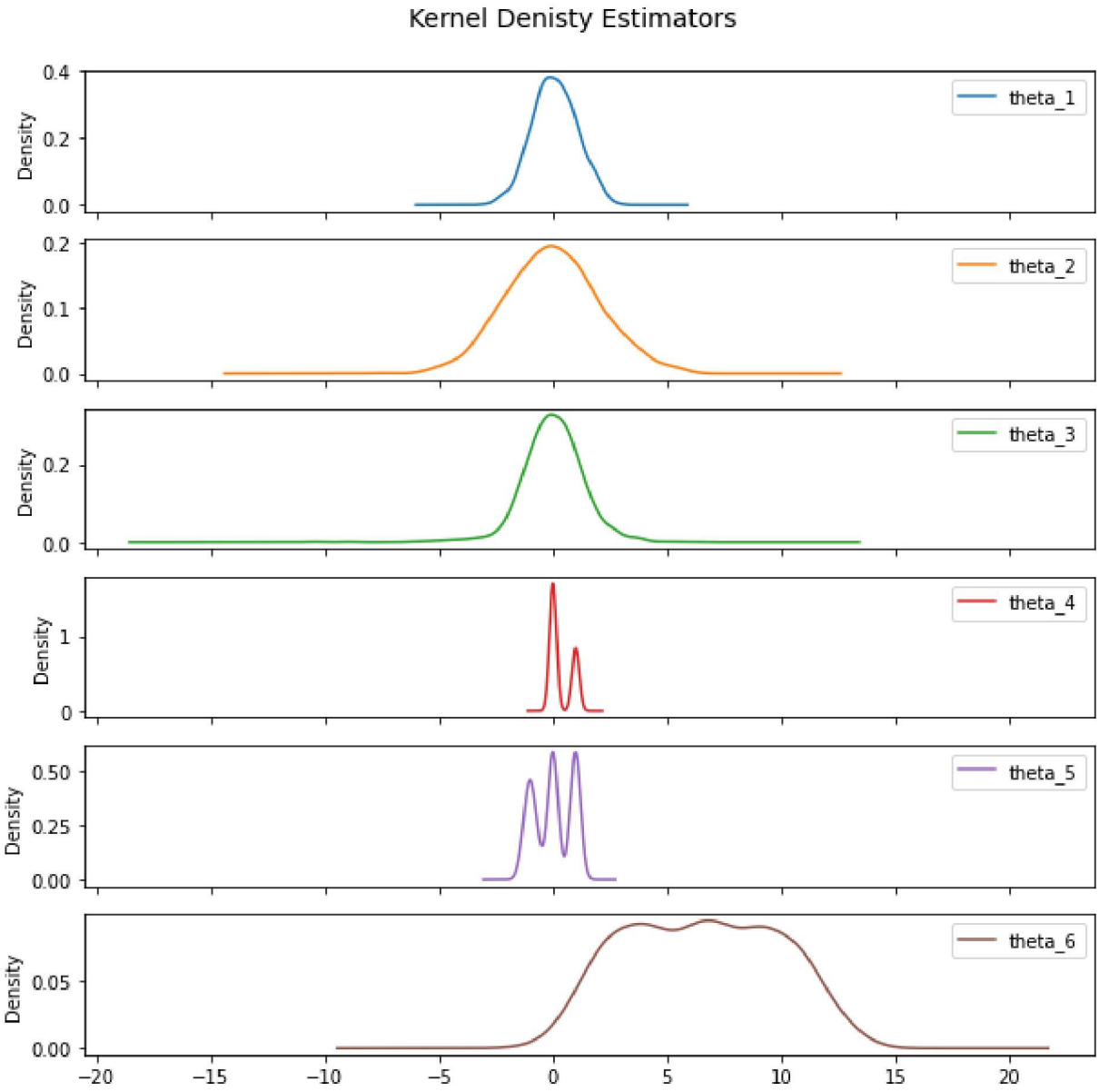
Review of data wrangling and visualization in Python

In [ ]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv('Data1.csv', index_col=0, parse_dates=True)

axes = df.plot(subplots=True, figsize=(10,10))
for ax in axes:
    ax.legend(loc='lower right')
fig=axes[0].figure
fig.text(0.5,0.92, 'Time series plot', ha='center', va='center', size=14)
fig.text(0.5,0.07, 'Date', ha='center', va='center', size=12)
fig.text(0.07,0.5, 'Value', ha='center', va='center', rotation=90, size=12)
plt.show()

axes = df.hist(figsize=(10,10), bins=10)
fig=axes[0][0].figure
fig.text(0.5,0.92, 'Histograms', ha='center', va='center', size=14)
fig.text(0.5,0.07, 'Value', ha='center', va='center', size=12)
fig.text(0.07,0.5, 'Frequency', ha='center', va='center', rotation=90, size=12)
plt.show()

axes = df.plot.density(subplots=True, figsize=(10,10))
fig=axes[0].figure
fig.text(0.5,0.92, 'Kernel Denisty Estimators', ha='center', va='center', size=14)
plt.show()
```

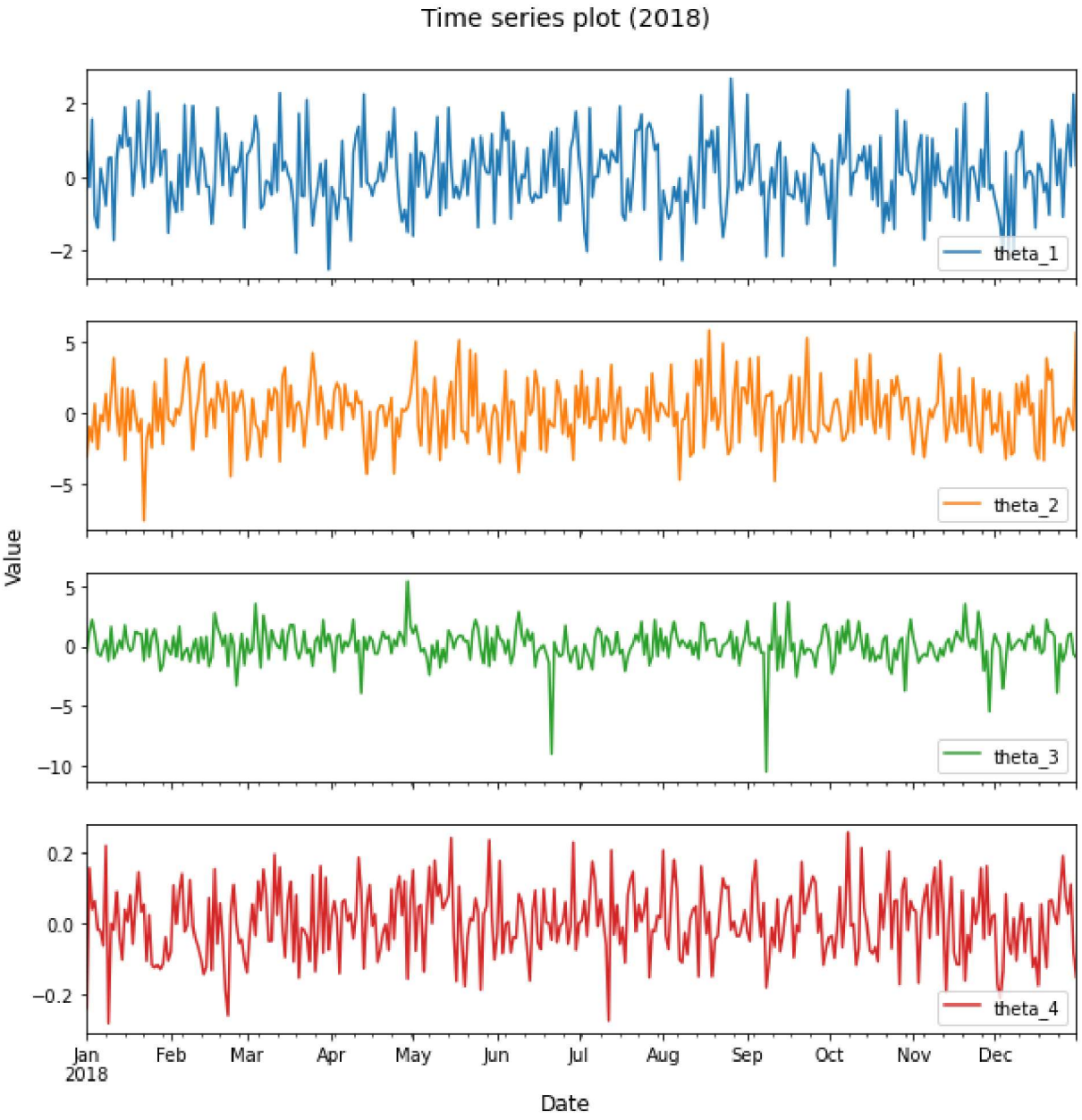

Time series plot
theta_1
theta_2
theta_3
theta_4
theta_5
theta_6
Value
Jan 2017
Jul
Jan 2018
Jul
Jan 2019
Jul
Date

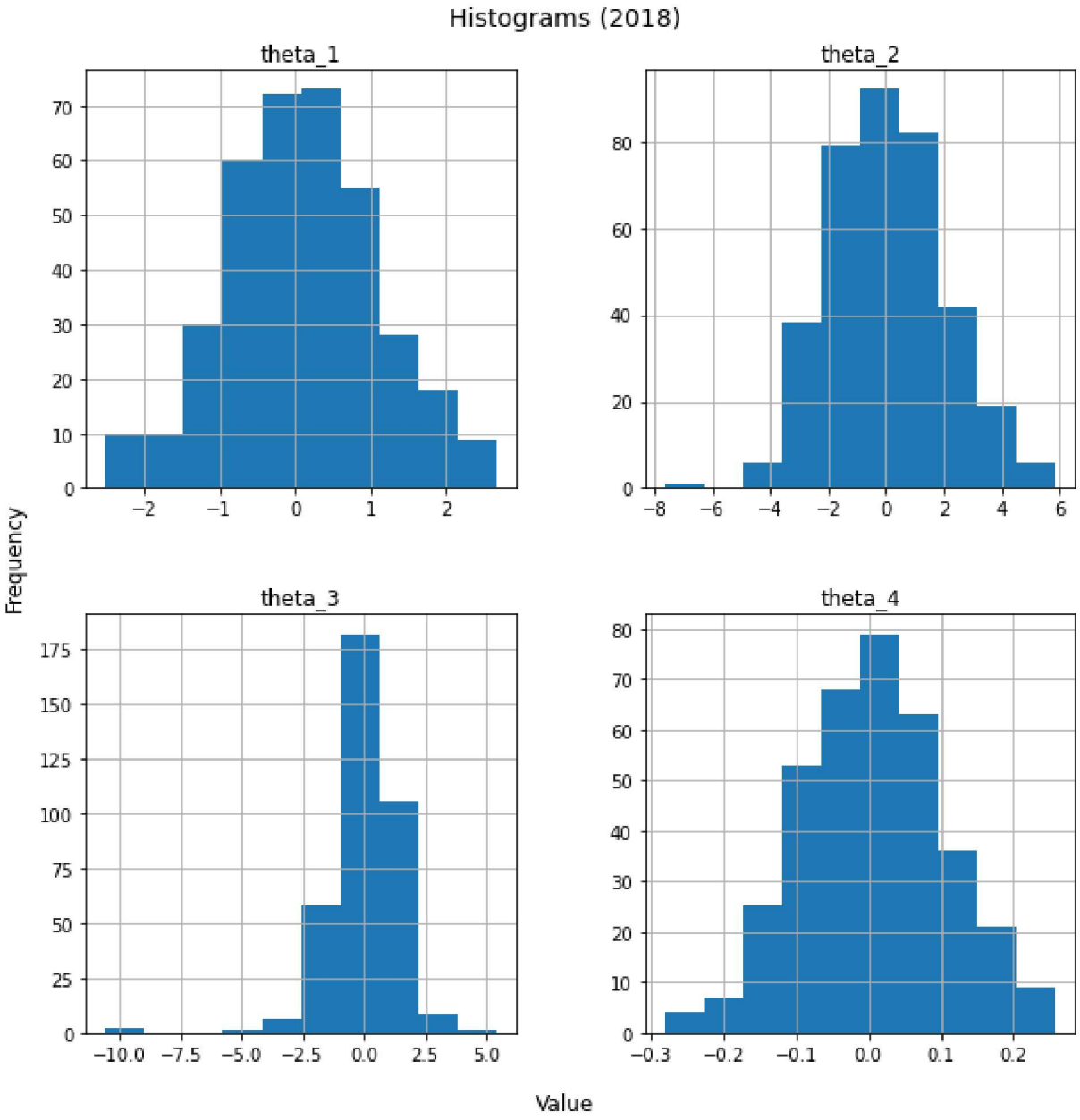

Histograms
theta_1
theta_2
theta_3
theta_4
theta_5
theta_6
Frequency
Value

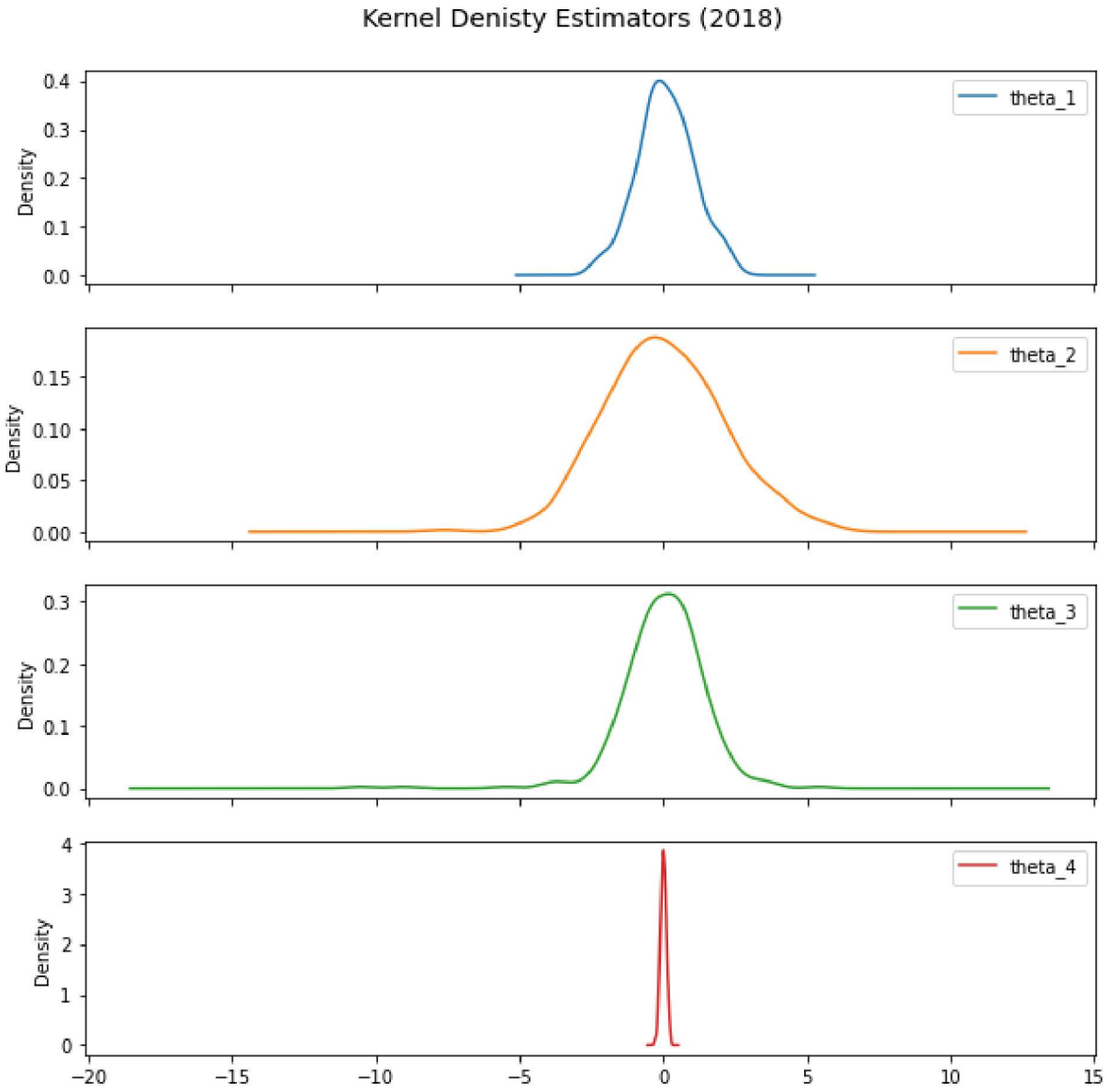


In [ ]:

```
df_2 = df.loc['2018', 'theta_1':'theta_4']

axes = df_2.plot(subplots=True, figsize=(10,10))
for ax in axes:
    ax.legend(loc='lower right')
fig=axes[0].figure
fig.text(0.5,0.92, 'Time series plot (2018)', ha='center', va='center', size=14)
fig.text(0.5,0.07, 'Date', ha='center', va='center', size=12)
fig.text(0.07,0.5, 'Value', ha='center', va='center', rotation=90, size=12)
plt.show()

axes = df_2.hist(figsize=(10,10), bins=10)
fig=axes[0][0].figure
fig.text(0.5,0.92, 'Histograms (2018)', ha='center', va='center', size=14)
fig.text(0.5,0.07, 'Value', ha='center', va='center', size=12)
fig.text(0.07,0.5, 'Frequency', ha='center', va='center', rotation=90, size=12)
plt.show()

axes = df_2.plot.density(subplots=True, figsize=(10,10))
fig=axes[0].figure
fig.text(0.5,0.92, 'Kernel Denisty Estimators (2018)', ha='center', va='center', siz
plt.show()
```


Time series plot (2018)
theta_1
theta_2
theta_3
theta_4
Value
Date
Jan
2018
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Histograms (2018)
theta_1
theta_2
theta_3
theta_4
Frequency
Value



Exercise 2

In [ ]:

```
from cmdstanpy import CmdStanModel

dataset = {"N":16, "y":[1,0,0,1,1,1,0,1,1,0,1,1,1,0,1,0]}
model = CmdStanModel(stan_file="bern_1.stan")
print(model)

model_sample = model.sample(data=dataset, output_dir='output')
theta = model_sample.stan_variable(name='theta')

df=model_sample.summary()
df_theta = df.loc['theta']
quan_50=df_theta['50%']
quan_5=df_theta['5%']
quan_95=df_theta['95%']
print('\n\n',df,'\n\n')

axes = plt.hist(theta, bins=50, density=True)
plt.axvline(theta.mean(), color='indigo')
plt.axvline(quan_5, color='yellow')
plt.axvline(quan_50, color='orange')
plt.axvline(quan_95, color='red')
plt.title('Histogram')
plt.show()
```

```
INFO:cmdstanpy:found newer exe file, not recompiling
INFO:cmdstanpy:compiled model file: E:/Programowanie/Microsoft VS Code Projects/Lab_
1/bern_1.exe
INFO:cmdstanpy:start chain 1
INFO:cmdstanpy:start chain 2
INFO:cmdstanpy:start chain 3
INFO:cmdstanpy:start chain 4
INFO:cmdstanpy:finish chain 1
INFO:cmdstanpy:finish chain 2
INFO:cmdstanpy:finish chain 4
INFO:cmdstanpy:finish chain 3
CmdStanModel: name=bern_1
        stan_file=E:\Programowanie\Microsoft VS Code Projects\Lab_1\bern_1.stan
        exe_file=E:/Programowanie/Microsoft VS Code Projects/Lab_1/bern_1.exe
        compiler_options=stanc_options=None, cpp_options=None


         Mean   MCSE  StdDev     5%    50%    95%   N_Eff  N_Eff/s  R_hat
name
lp__   -13.00  0.016    0.69 -14.00 -12.00 -12.00  1800.0  13000.0    1.0
theta    0.61  0.003    0.11   0.42   0.61   0.78  1300.0   9100.0    1.0
```