Data Visualisation - Lab assignment 5

September 19, 2022

[1]:
```
import pandas as pd
import seaborn as sns
import plotly.express as px

import matplotlib.pyplot as plt
```

```
C:\Users\MaxSc\Anaconda\envs\TIL6022\lib\site-packages\scipy\__init__.py:146:
UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this version
of SciPy (detected version 1.23.1
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

[2]:
```
import plotly.io as pio
pio.renderers.default = "plotly_mimetype+notebook"
```

1 Matplotlib

For this excercise, we have written the following code to load the stock dataset built into plotly express.

[3]:
```
stocks = px.data.stocks()
stocks.head()
```

[3]:

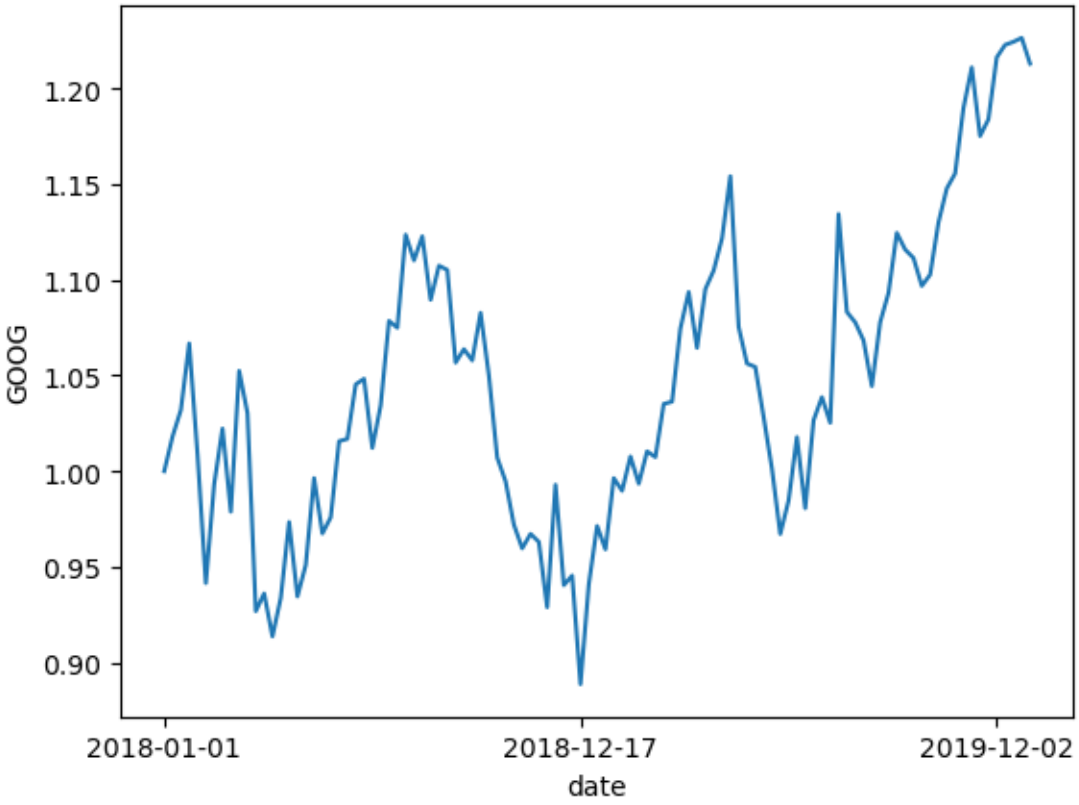
	date	GOOG	AAPL	AMZN	FB	NFLX	MSFT
0	2018-01-01	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	2018-01-08	1.018172	1.011943	1.061881	0.959968	1.053526	1.015988
2	2018-01-15	1.032008	1.019771	1.053240	0.970243	1.049860	1.020524
3	2018-01-22	1.066783	0.980057	1.140676	1.016858	1.307681	1.066561
4	2018-01-29	1.008773	0.917143	1.163374	1.018357	1.273537	1.040708

1.1 Question 1:

Select a stock and create a suitable plot for it. Make sure the plot is readable with relevant information, such as date, values.

[4]:
```
sns.lineplot(data=stocks, x='date', y='GOOG').set_xticks([0,50,100])
#plt.xticks([0,50,100])
```

[4]:
```
[<matplotlib.axis.XTick at 0x15e6df6ddf0>,
 <matplotlib.axis.XTick at 0x15e6df6ddc0>,
 <matplotlib.axis.XTick at 0x15e6dfb20d0>]
```



1.2 Question 2:

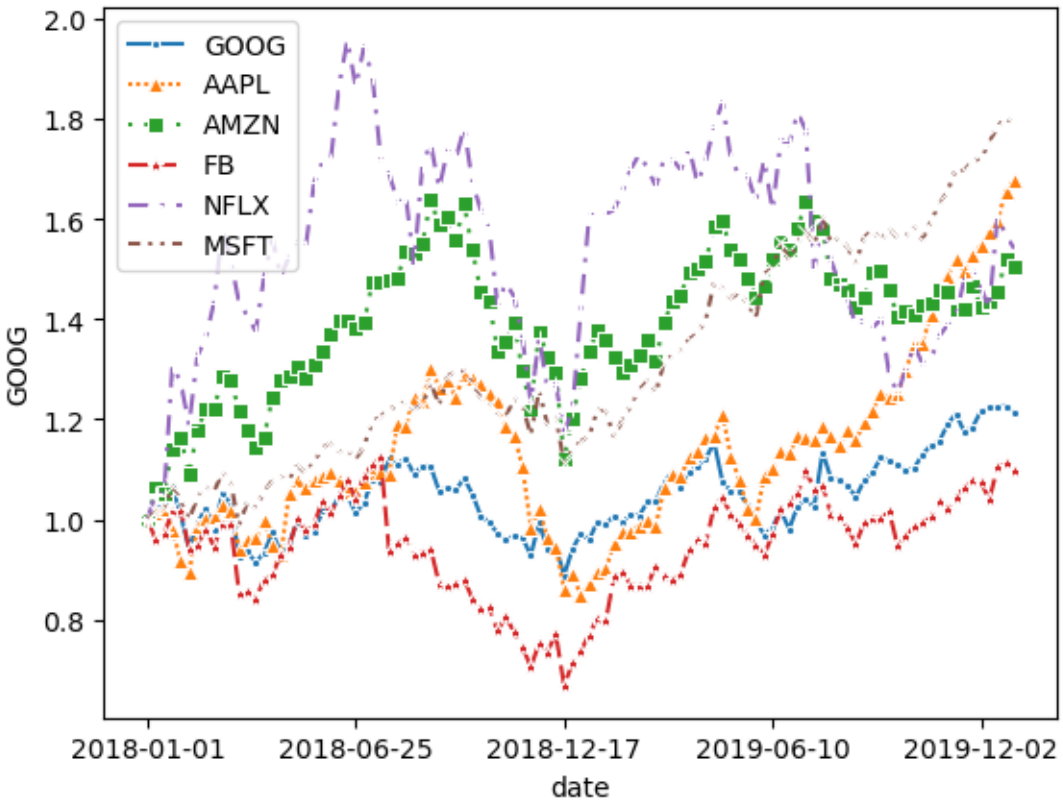
You've already plot data from one stock. It is possible to plot multiples of them to support comparison.
To highlight different lines, customise line styles, markers, colors and include a legend to the plot.

[5]:
```
companies = ['GOOG','AAPL','AMZN','FB','NFLX','MSFT']
linestyles = ['-', (0,(1,1)), (0,(1,3)), (0,(5,2)), (0,(5,3,1,3)),␣
 ↪(0,(3,2,1,2,1,2))]
marker = ['.','^','s','*','+','x']

for i in range(len(companies)):
    sns.lineplot(data=stocks, x='date', y=companies[i], marker=marker[i]).
 ↪lines[i].set_linestyle(linestyles[i])

plt.legend(companies)
plt.xticks([0,25,50,75,100])
```

[5]:
```
([<matplotlib.axis.XTick at 0x15e6e7a9ca0>,
  <matplotlib.axis.XTick at 0x15e6e7a9c70>,
  <matplotlib.axis.XTick at 0x15e6e7a9970>,
  <matplotlib.axis.XTick at 0x15e6e7ed430>,
  <matplotlib.axis.XTick at 0x15e6e7edcd0>],
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')])
```



2 Seaborn

First, load the tips dataset

[6]:
```
tips = sns.load_dataset('tips')
tips.head()
```

[6]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3

```
2       21.01  3.50    Male     No  Sun  Dinner     3
3       23.68  3.31    Male     No  Sun  Dinner     2
4       24.59  3.61  Female     No  Sun  Dinner     4
```

2.1 Question 3:

Let's explore this dataset. Pose a question and create a plot that support drawing answers for your question.

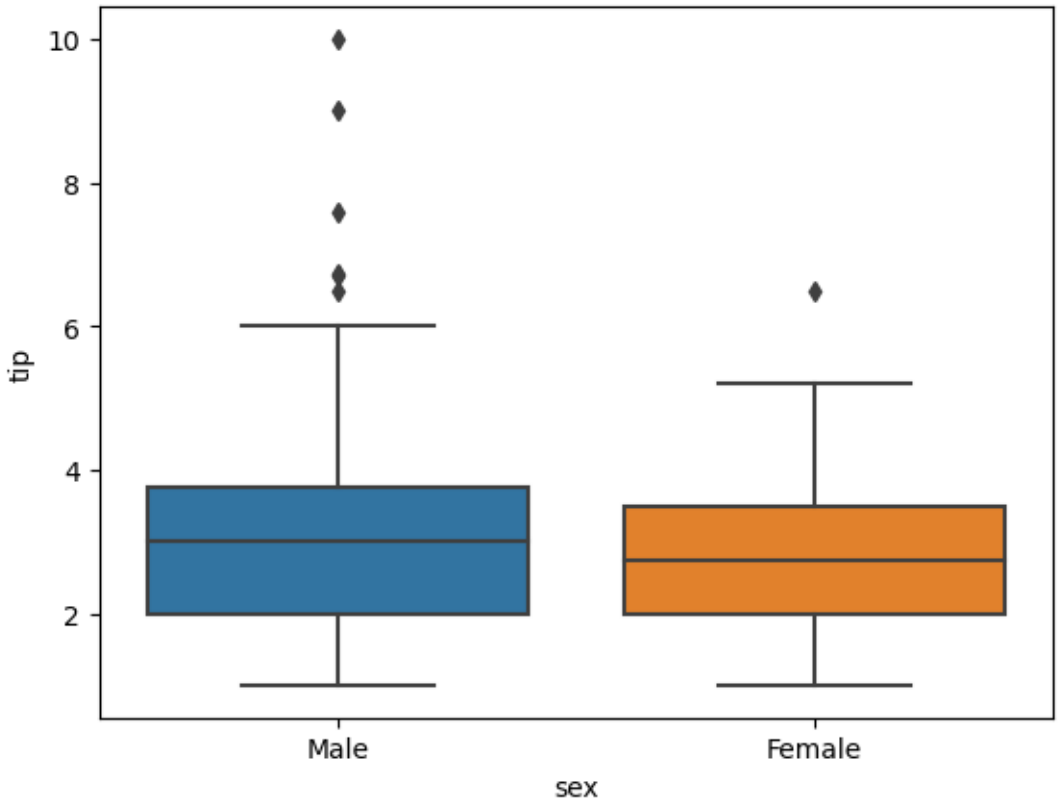
Some possible questions: - Are there differences between male and female when it comes to giving tips? - What attribute correlate the most with tip?

[7]:
```
#Are there differences between male and female when it comes to giving tips?
print('Are there differences between male and female when it comes to giving␣
 ↪tips?')

sns.boxplot(x='sex', y='tip', data=tips)

#There is no significant difference between male and female when it comes to␣
 ↪giving tips.
print('There is no significant difference between male and female when it comes␣
 ↪to giving tips.')
```

```
Are there differences between male and female when it comes to giving tips?
There is no significant difference between male and female when it comes to
giving tips.
```



3 Plotly Express

3.1 Question 4:

Redo the above exercises (challenges 2 & 3) with plotly express. Create diagrams which you can interact with.

[8]:
```
stocks = px.data.stocks()
companies = ['GOOG','AAPL','AMZN','FB','NFLX','MSFT']
```

3.1.1 The stocks dataset

Hints: - Turn stocks dataframe into a structure that can be picked up easily with plotly express

[9]:
```
fig = px.line(stocks, x = 'date', y = companies, markers=True)
fig.show()
```

3.1.2 The tips dataset

[10]:
```
tips = sns.load_dataset('tips')
boxplot = px.box(tips, x="sex", y="tip", color="sex")
boxplot.show()
```

3.2 Question 5:

Recreate the barplot below that shows the population of different continents for the year 2007.

Hints:

- Extract the 2007 year data from the dataframe. You have to process the data accordingly
- use plotly bar
- Add different colors for different continents
- Sort the order of the continent for the visualisation. Use axis layout setting
- Add text to each bar that represents the population

[11]:
```
#load data
df = px.data.gapminder()
df.head()
```

[11]:

	country	continent	year	lifeExp	pop	gdpPercap	iso_alpha	iso_num
0	Afghanistan	Asia	1952	28.801	8425333	779.445314	AFG	4
1	Afghanistan	Asia	1957	30.332	9240934	820.853030	AFG	4
2	Afghanistan	Asia	1962	31.997	10267083	853.100710	AFG	4
3	Afghanistan	Asia	1967	34.020	11537966	836.197138	AFG	4
4	Afghanistan	Asia	1972	36.088	13079460	739.981106	AFG	4

[12]:
```
# YOUR CODE HERE
data_2007 = df.query('year==2007')
data_2007_continent = data_2007.groupby('continent').sum()
fig = px.bar(data_2007_continent, x = 'pop', y = data_2007_continent.index,␣
 ↪color = data_2007_continent.index, text_auto = '.2s')
fig.update_yaxes(categoryorder = 'total ascending')
fig.update_traces(textposition="outside")
fig.update_layout(showlegend=False)
fig.show()
```