In [1]:

```
import pandas as pd
import seaborn as sns
import plotly.express as px

import matplotlib.pyplot as plt
```

In [2]:

```
import plotly.io as pio
pio.renderers.default = "plotly_mimetype+notebook"
```

Matplotlib

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

In [3]:

```
stocks = px.data.stocks()
stocks.head()
```

Out[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

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.

In [4]:

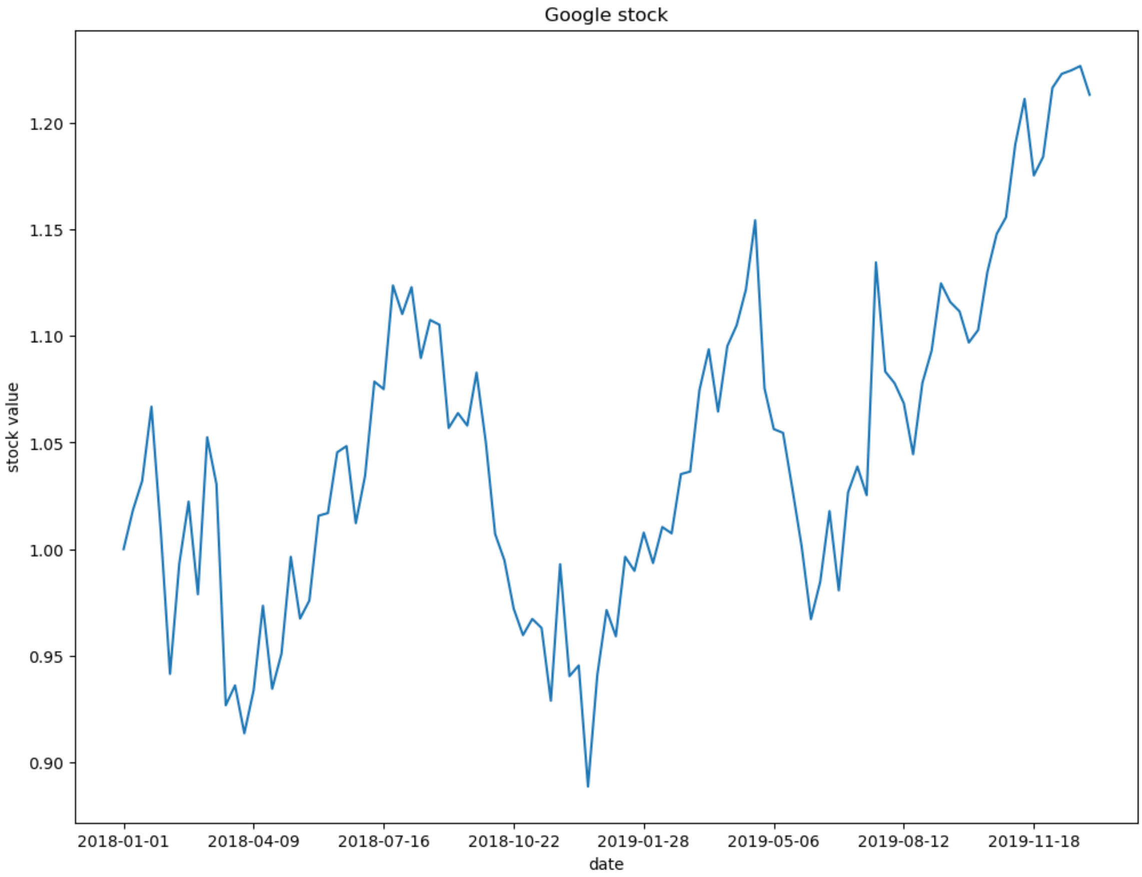
```
import numpy as np

fig, ax = plt.subplots(figsize=(12,9))

x = stocks['date']
y = stocks['GOOG']
plt.plot(x,y)

ax.set_xticks(np.arange(0, len(x)+1, 14)) #Length of file = 105, steps of 14 makes 8 da
tes show in the x-axis.

ax.set_xlabel('date')
ax.set_ylabel('stock value')
ax.set_title('Google stock');
```

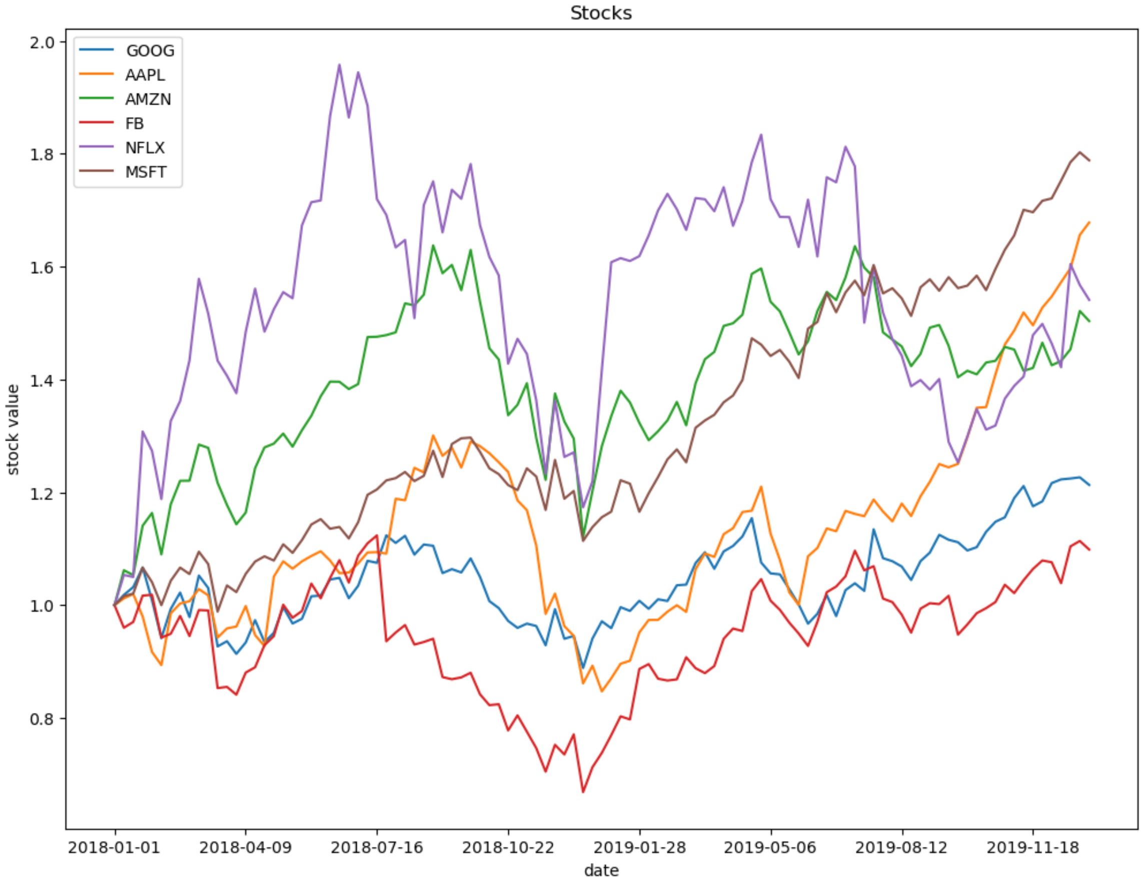


Question 2:

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.

In [5]:

```
stocks.plot(x = 'date', figsize=(12,9), xticks=np.arange(0, len(x)+1, 14), xlabel = 'da
te', ylabel = 'stock value', title = 'Stocks');
```



Seaborn

First, load the tips (https://github.com/mwaskom/seaborn-data/blob/master/tips.csv) dataset

In [6]:

```
tips = sns.load_dataset('tips')
tips.head()
```

Out[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

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?

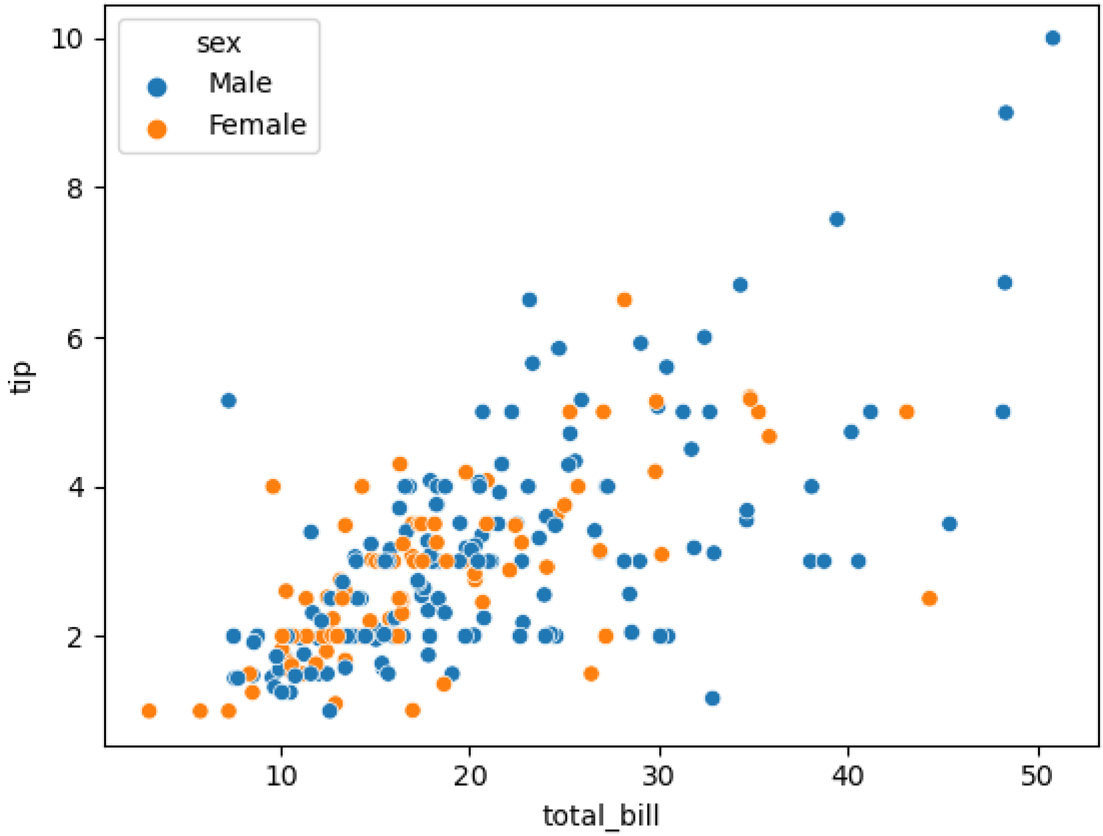
In [7]:

```
print("Question: Are there differences between male and female when it comes to giving
 tips?")

sns.scatterplot(x='total_bill', y='tip', data=tips, hue = 'sex')
plt.show()

print("Answer: The scatterplot shows that give a little more tip than women.")
print("The average tip men give is",
      round(tips[tips.sex == 'Male'].tip.mean(),2), ", while women on average give a ti
p of",
      round(tips[tips.sex == 'Female'].tip.mean(),2),'.');
```

```
Question: Are there differences between male and female when it comes to g
iving tips?
```



```
Answer: The scatterplot shows that give a little more tip than women.
The average tip men give is 3.09 , while women on average give a tip of 2.
83 .
```

Plotly Express

Question 4:

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

The stocks dataset

Hints:

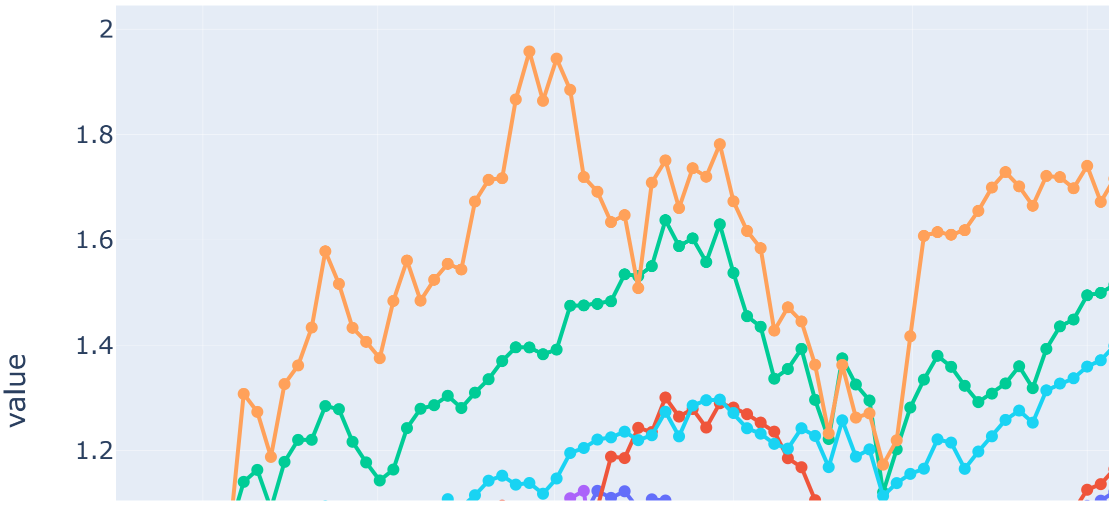
- Turn stocks dataframe into a structure that can be picked up easily with plotly express

In [8]:

```
stock_names = stocks.columns[1:]
print(stock_names)

fig1 = px.line(stocks, x = 'date', y = stock_names, markers = True)
fig1.show();
```

```
Index(['GOOG', 'AAPL', 'AMZN', 'FB', 'NFLX', 'MSFT'], dtype='object')
```

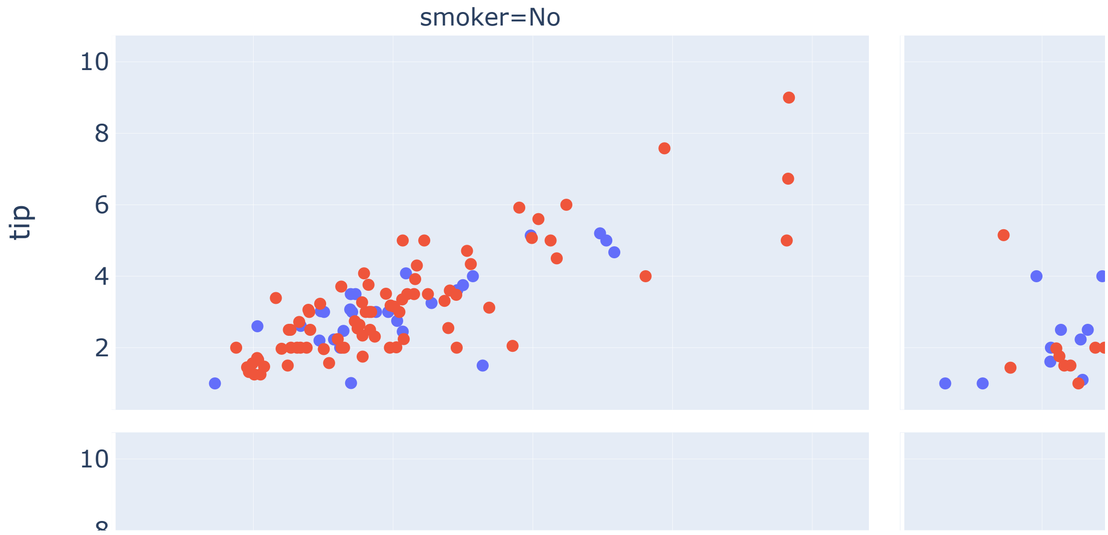


The tips dataset

In [9]:

```
#fig2 = px.scatter(tips, x='total_bill', y='tip', color='sex')
#fig2.show()

fig1 = px.scatter(tips, x="total_bill", y="tip", color="sex", facet_col="smoker", facet
_row="time")
fig1.show(figsize = (12,9));
```



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 (https://plotly.com/python-api-reference/generated/plotly.express.bar)
- Add different colors for different continents
- Sort the order of the continent for the visualisation. Use axis layout setting (https://plotly.com/python/reference/layout/xaxis/)
- Add text to each bar that represents the population

In [10]:

```
#Load data
df = px.data.gapminder()
df.head()
```

Out[10]:

	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

In [11]:

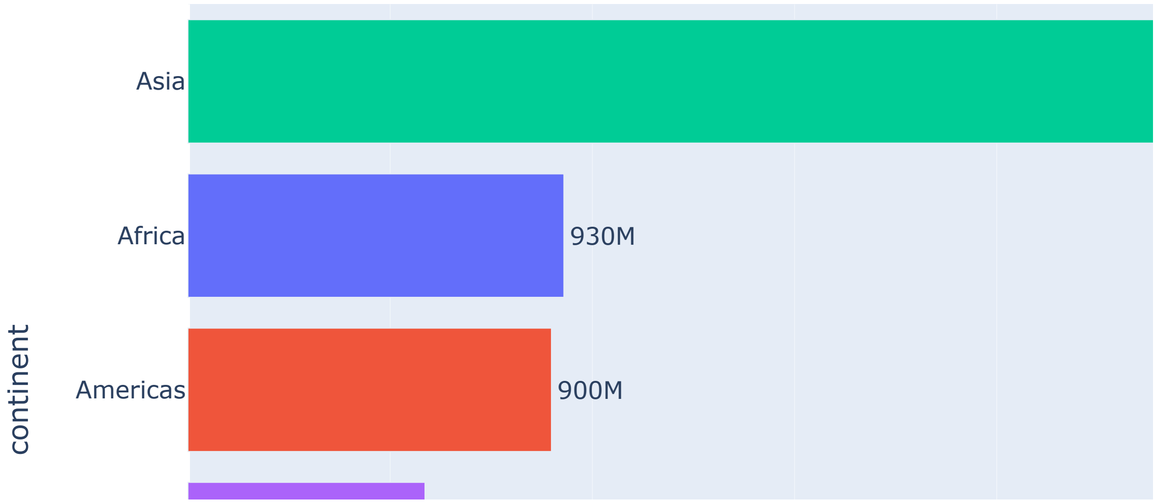
```
df_2007 = df.query('year==2007')
df_2007_new = df_2007.groupby('continent').sum()

fig = px.bar(df_2007_new, x="pop", y=df_2007_new.index, orientation='h',
             color = df_2007_new["pop"].astype(str), text='pop')

fig.update_yaxes(categoryorder="total ascending")

fig.update_traces(textposition='outside', showlegend=False,texttemplate = "%{text:.2s}"
)
fig.update_layout(uniformtext_minsize=12)

fig.show()
```


Asia
Africa
930M
Americas
900M
continent