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

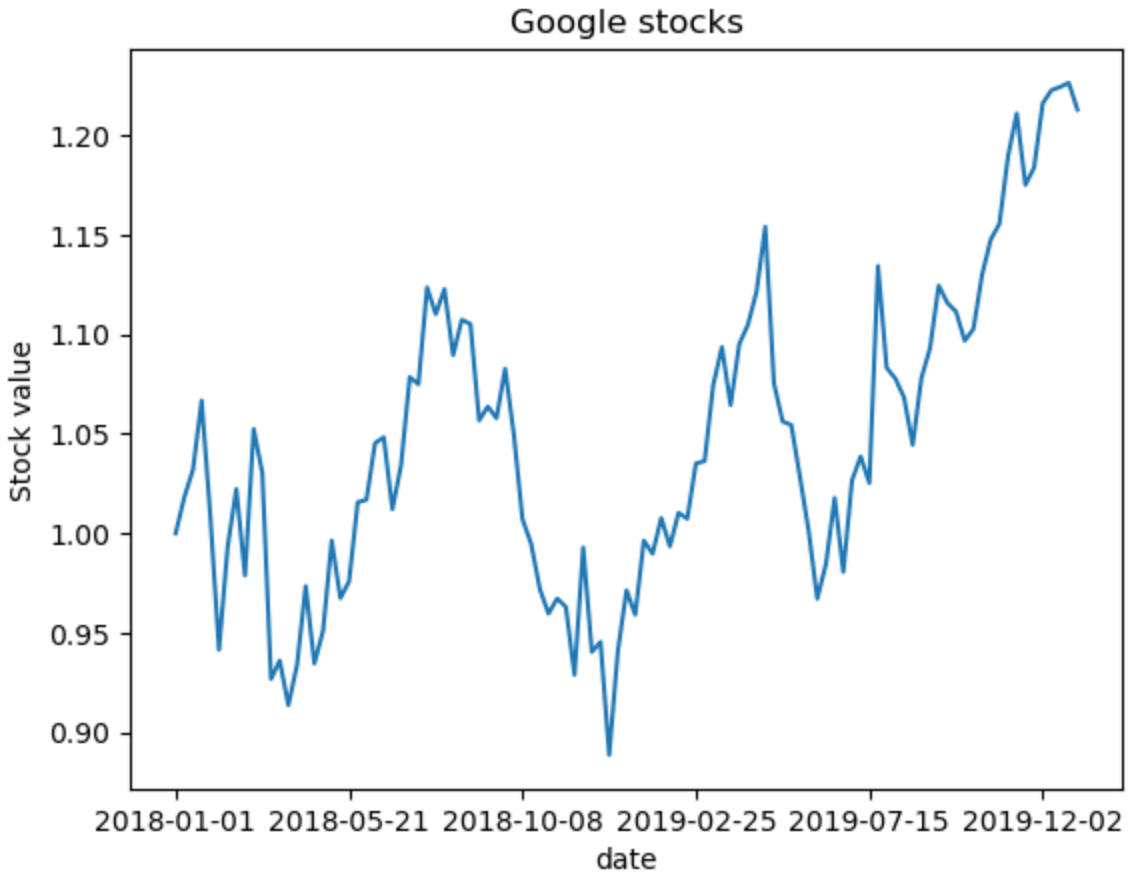
For this excercise, we have written the following code to load the stock dataset built into plotly express.

In [3]:
```
stocks = px.data.stocks()
stocks.head()
stocks.set_index('date', inplace = True)
```

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.

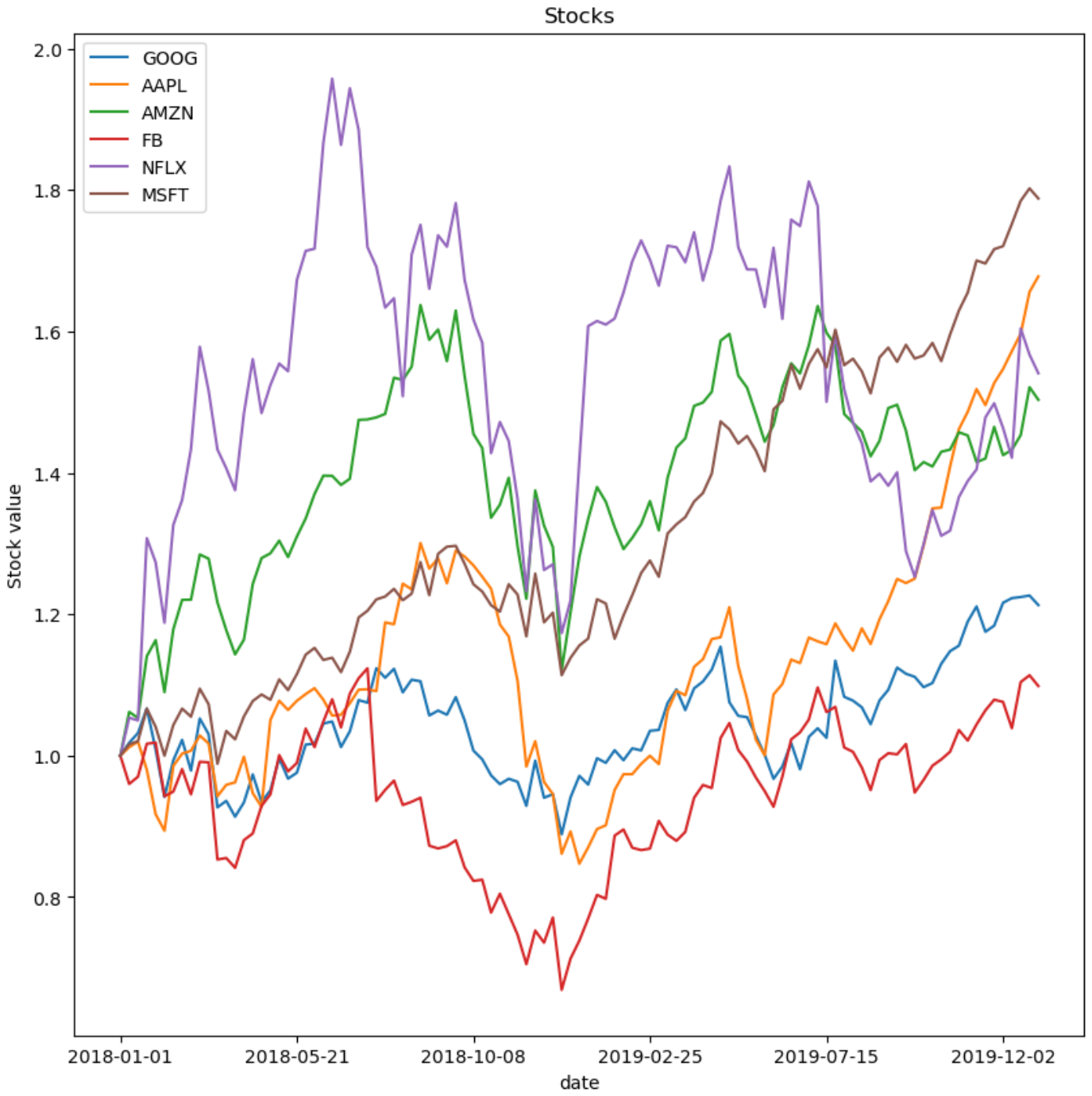
In [4]:
```
# YOUR CODE HERE
stocks.plot(y='GOOG', legend = False);
plt.title('Google stocks');
plt.ylabel('Stock value');
plt.show()
```



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]:
```
# YOUR CODE HERE
stocks.plot(figsize=(10,10));
plt.title('Stocks');
plt.ylabel('Stock value');
plt.show()
```



Seaborn

First, load the tips 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?

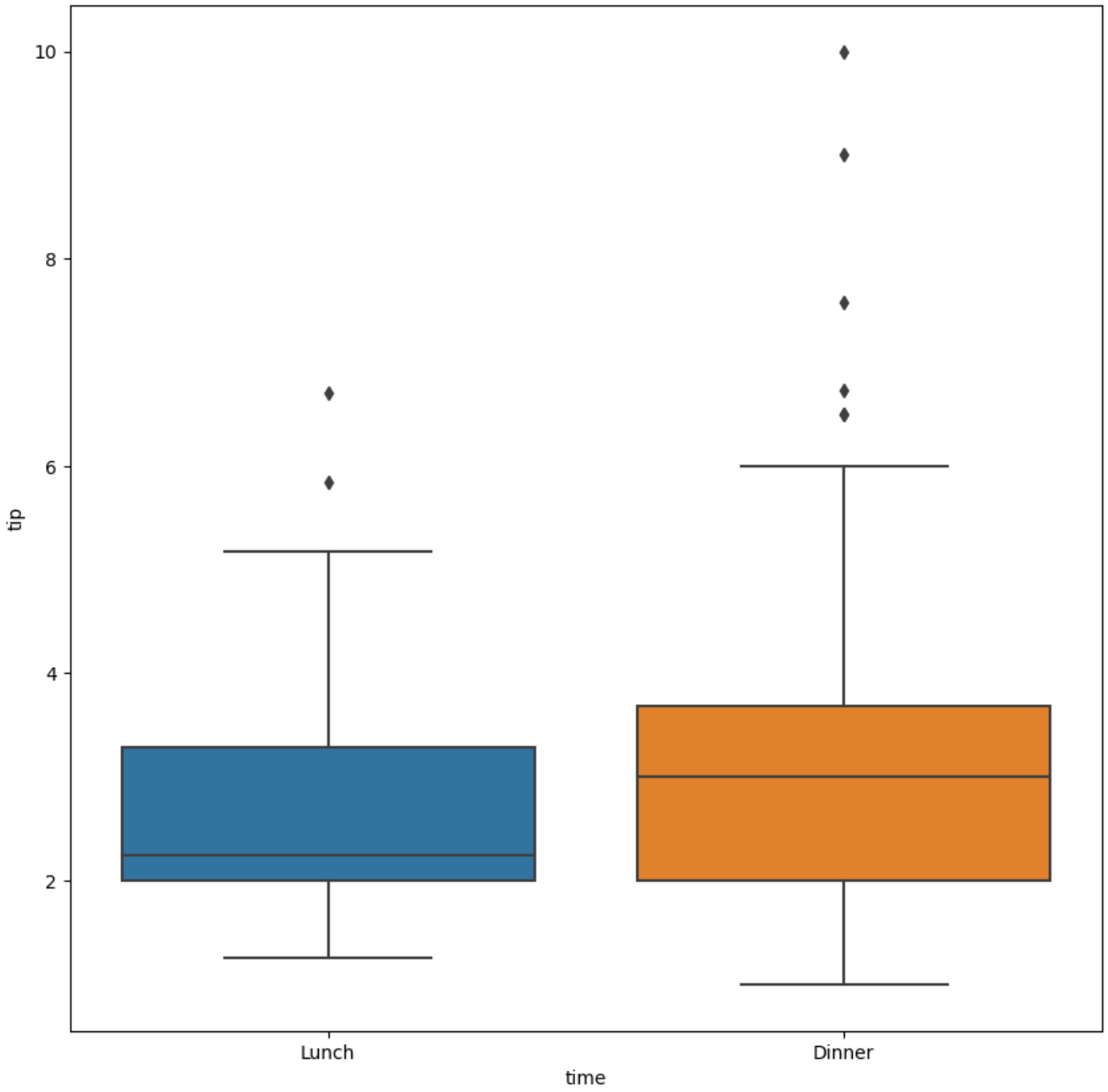
In [7]:

```
# YOUR CODE HERE
print('The question asked is: \nWhen are the tips the highest?')

fig,ax = plt.subplots(figsize=(10,10))
sns.boxplot(x='time', y='tip', data=tips)
plt.show()

print('From the plots we can conclude that on average, tips are higher at dinner than at
```

```
The question asked is:
When are the tips the highest?
```




```
From the plots we can conclude that on average, tips are higher at dinner than at lunch
```

Plotly Express

Question 4:

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

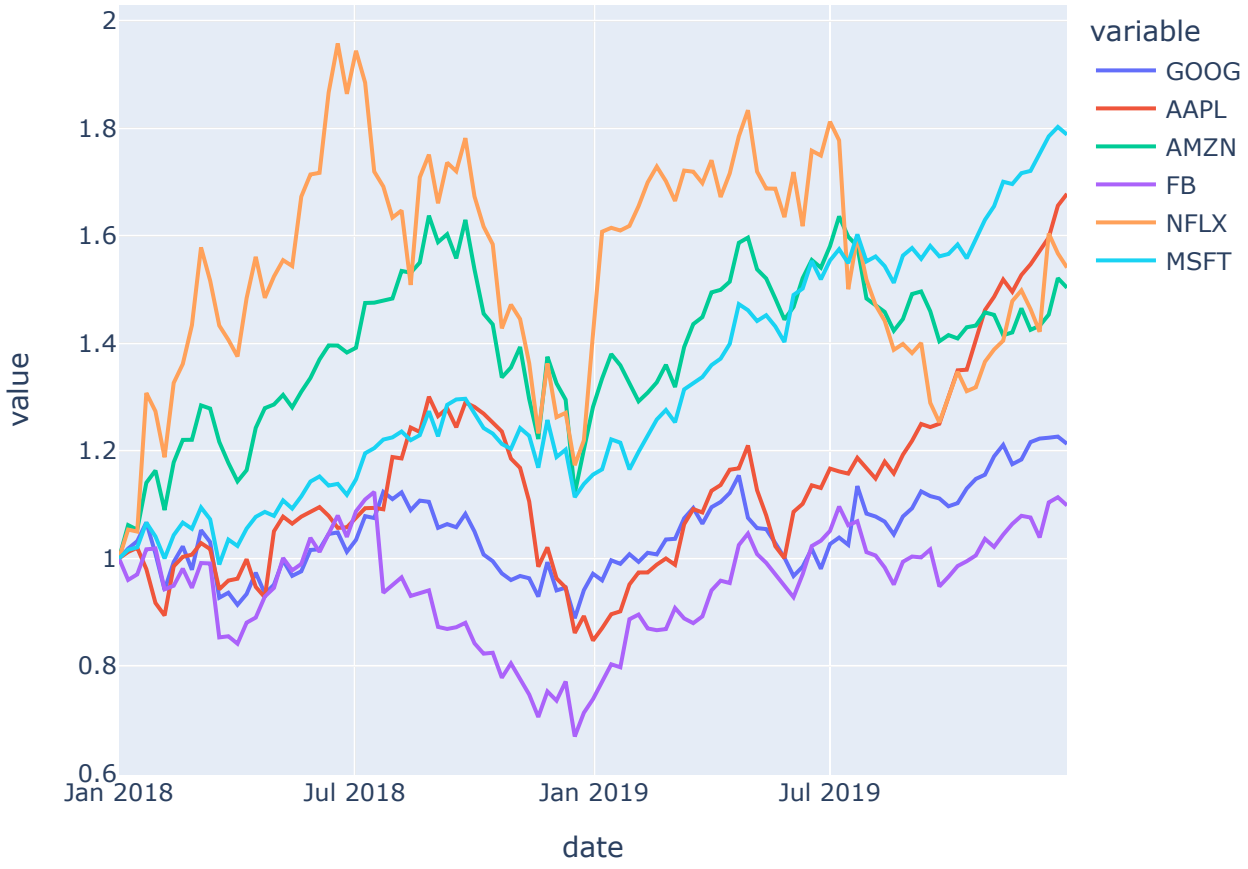
The stocks dataset

Hints:

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

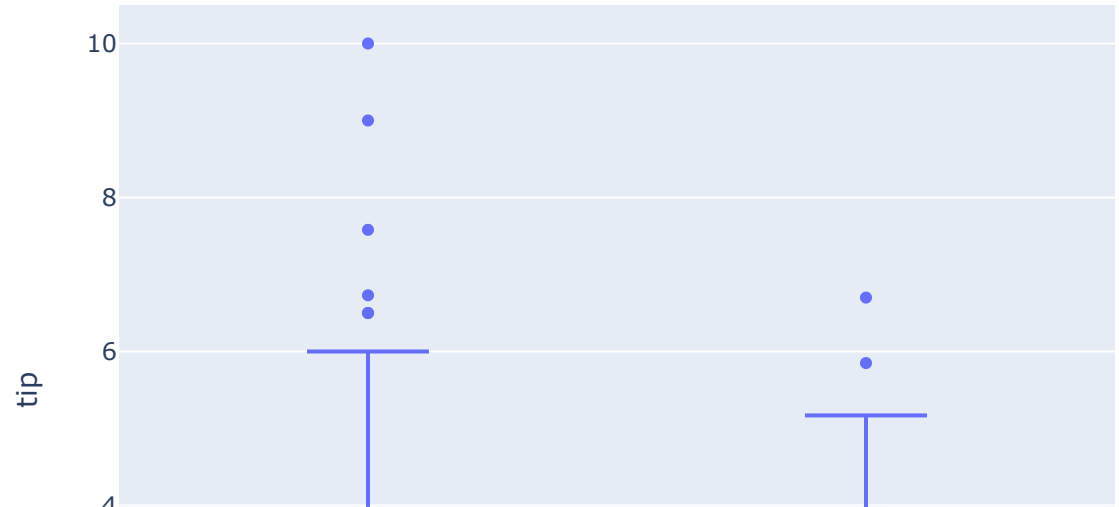
In [8]:
```
# YOUR CODE HERE
```

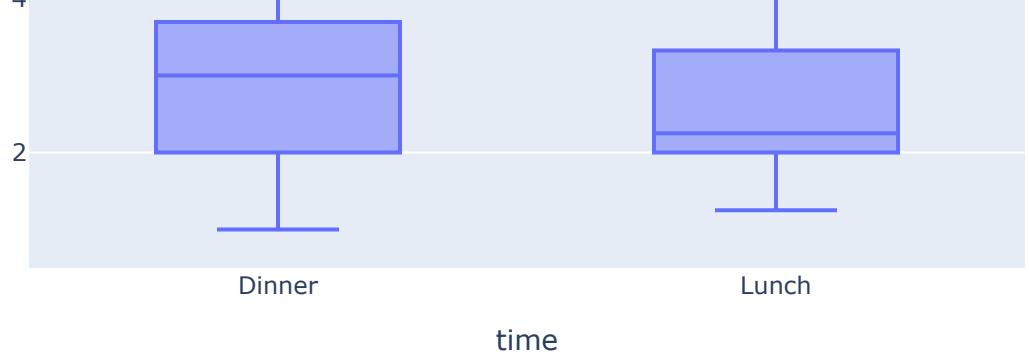
```
df = px.data.stocks()
fig = px.line(df, x= 'date', y=['GOOG','AAPL','AMZN','FB','NFLX', 'MSFT'])
fig.show()
```



The tips dataset

In [9]:
```
# YOUR CODE HERE
df = px.data.tips()
fig = px.box(df,x = 'time', y = 'tip')
fig.show()
```





Question 5:

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

Hints:

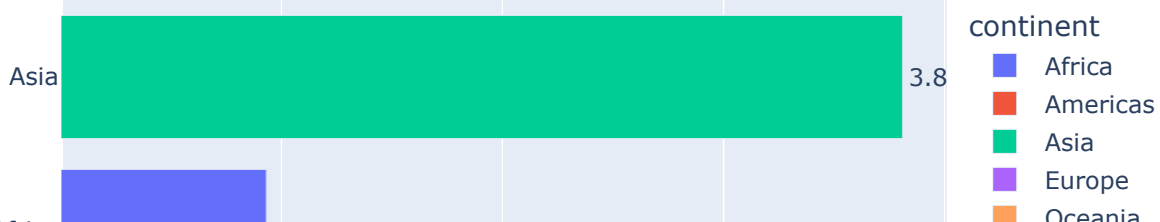
- Extract the 2007 year data from the dataframe. You have to process the data accordingly
- use plotly bar
- Add different colors for different continents
- Sort the order of the continent for the visualisation. Use axis layout setting
- Add text to each bar that represents the population

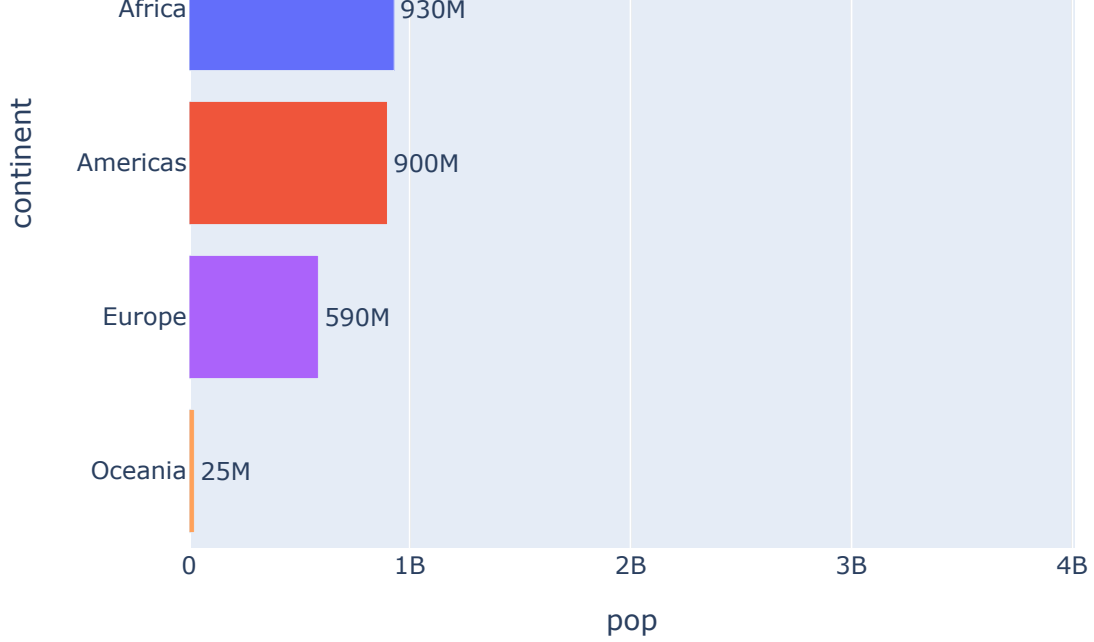
In [10]:
```
#load data
df = px.data.gapminder().query("year == 2007")
df = df.groupby(['continent']).sum()
df.head()
```

Out[10]:

continent	year	lifeExp	pop	gdpPercap	iso_num
Africa	104364	2849.914	929539692	160629.695446	23859
Americas	50175	1840.203	898871184	275075.790634	9843
Asia	66231	2334.040	3811953827	411609.886714	13354
Europe	60210	2329.458	586098529	751634.449078	12829
Oceania	4014	161.439	24549947	59620.376550	590

In [11]:
```
# YOUR CODE HERE
fig = px.bar(df, x = 'pop', y=df.index, color = df.index, text_auto = '.2s')
fig.update_traces(textposition="outside")
fig.update_layout(yaxis={'categoryorder':'total ascending'})
fig.show()
```




Africa
930M
Americas
900M
Europe
590M
Oceania
25M
continent
0
1B
2B
3B
4B
pop

In [ ]: