M4Lab-Seaborn

June 24, 2023

#Welcome to our Lab practice!

This lab is all about the package, Seaborn. Are you ready? Let's go!

You will find some small tasks in sections below.

> Try to figure out by yourself, or search for references. Being able to search and find information needed is an important skill that benefits you and your career for a long time.
>
> Please note this lab is exploratory, and there is no correct solution – Just try your best!

0.0.1 Choose One dataset from the list below, and play with it.

For practice purpose, you can load and play with one or more toy datasets in seaborn package (as we did for *tips* in lectures). You can get them by:

```
import seaborn as sns
data = sns.load_dataset('NAME')
```

The list of names can be found at Seaborn Datasets:

```
['anagrams', 'anscombe', 'attention', 'brain_networks', 'car_crashes',
 'diamonds', 'dots', 'exercise', 'flights', 'fmri', 'gammas', 'geyser',
'iris', 'mpg', 'penguins', 'planets', 'taxis', 'tips', 'titanic']
```

For example, to play with the dataset *iris*, you can do:

```
import seaborn as sns
data = sns.load_dataset('iris')
```

0.1 Set up the environment

0.1.1 Task: import packages and rename them accordingly

[1]:
```
# your code is here
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

0.2 Load the dataset

[2]:
```
# your code is here
data = sns.load_dataset('flights')
data
```

[2]:

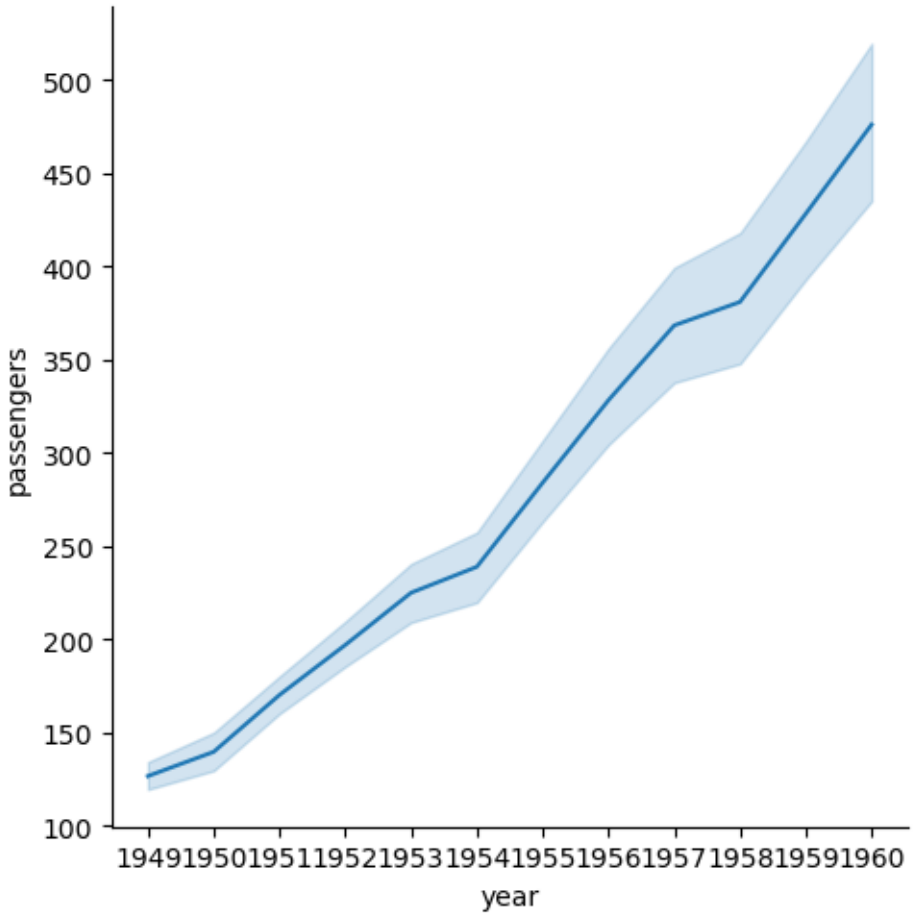
	year	month	passengers
0	1949	Jan	112
1	1949	Feb	118
2	1949	Mar	132
3	1949	Apr	129
4	1949	May	121
..	…	…	…
139	1960	Aug	606
140	1960	Sep	508
141	1960	Oct	461
142	1960	Nov	390
143	1960	Dec	432

[144 rows x 3 columns]

0.3 Play with relational plots

[40]:
```
# your code is here
sns.relplot(data['year'], data['passengers'], kind= 'line')
```

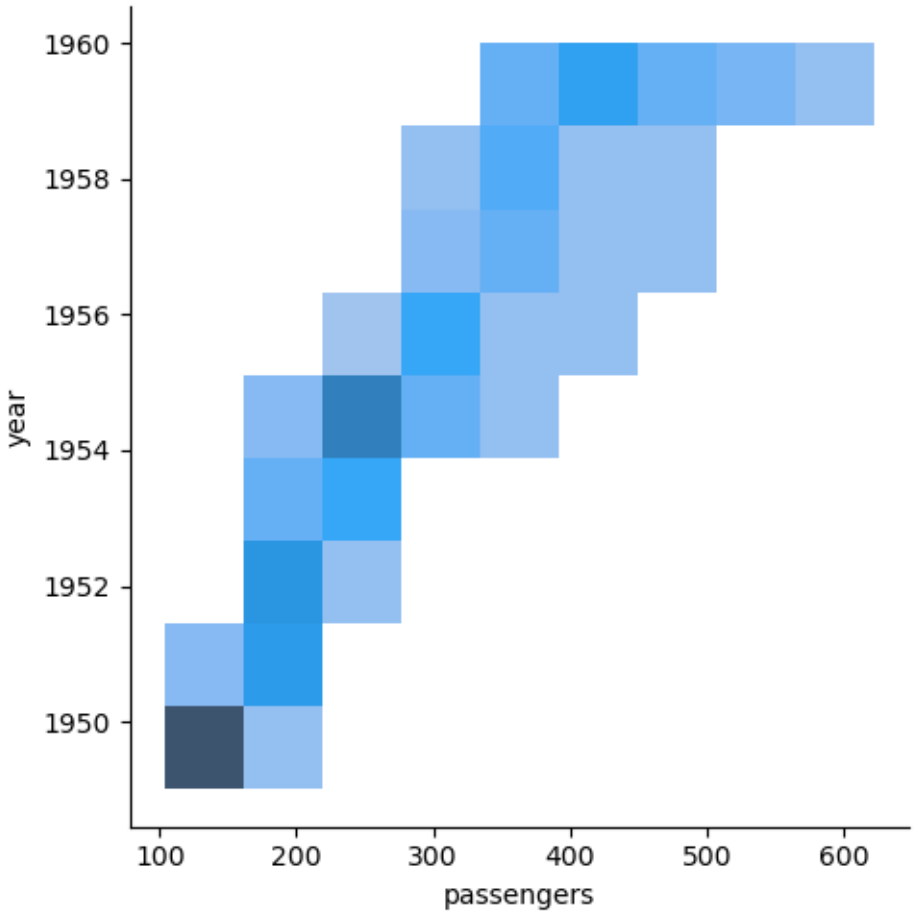
[40]:
```
<seaborn.axisgrid.FacetGrid at 0x22e12d55190>
```



0.4 Play with distribution plots

[11]:
```
# your code is here
sns.displot(x =data['passengers'], y = data['year'])
```

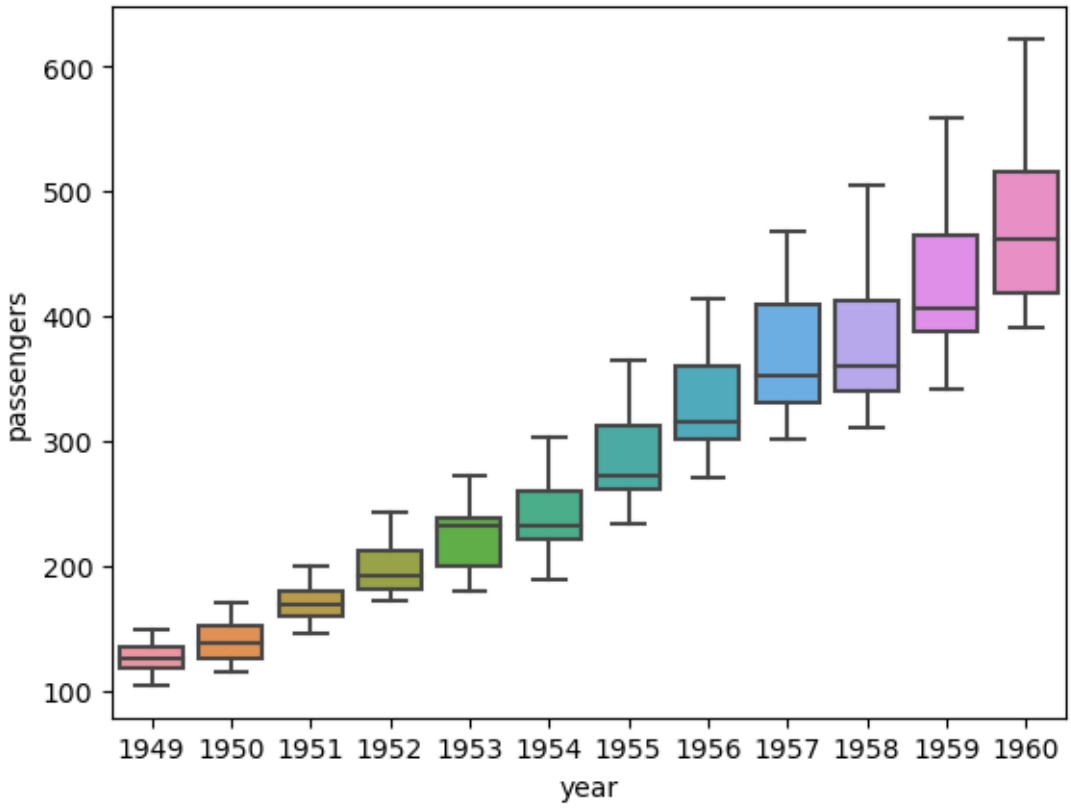
[11]: <seaborn.axisgrid.FacetGrid at 0x22e10ffcd30>



0.5 Play with other plots

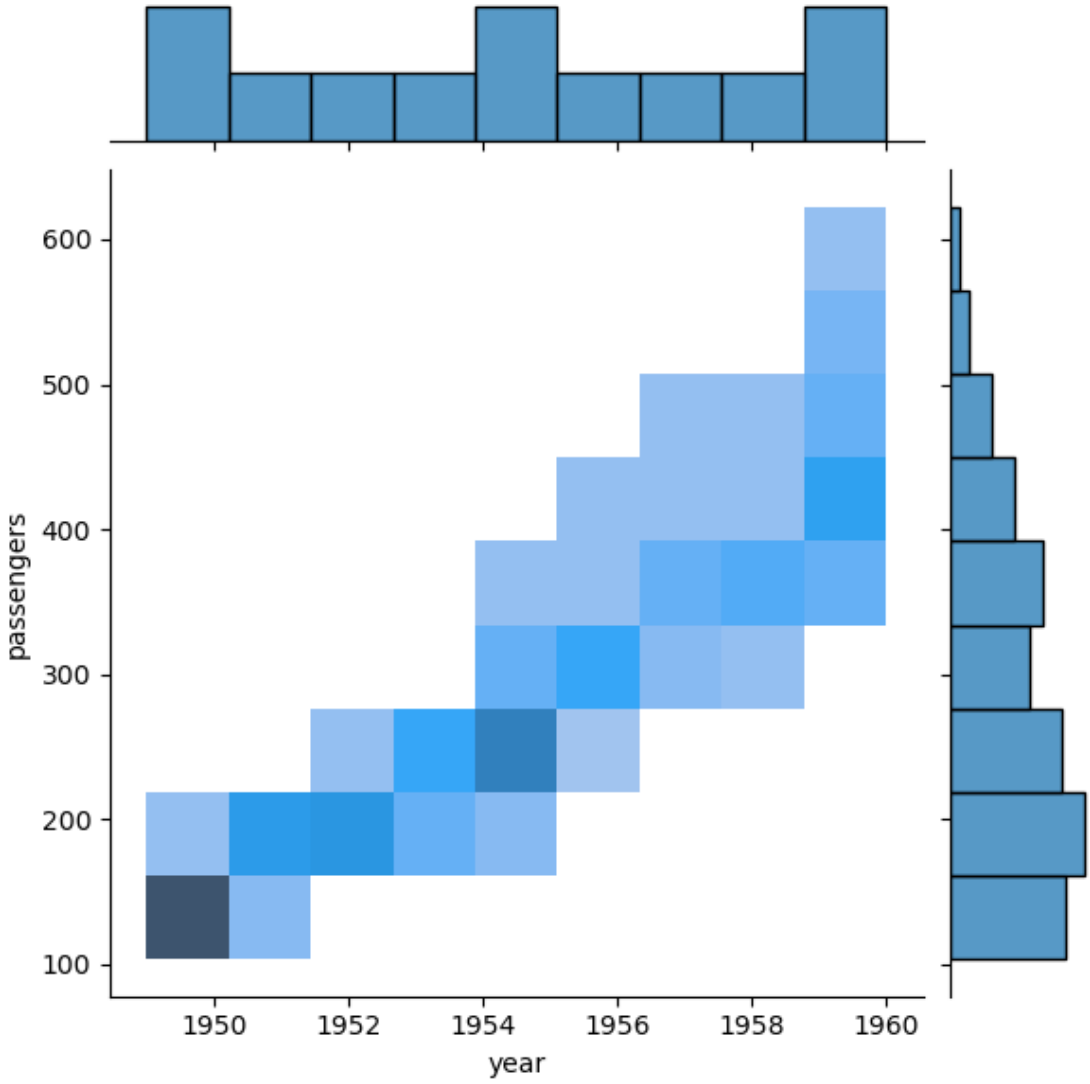
[13]:
```
# your code is here
sns.boxplot(x=data['year'], y=data['passengers'])
```

[13]:
```
<AxesSubplot:xlabel='year', ylabel='passengers'>
```



[14]:
```
sns.jointplot(data=data, x="year", y="passengers", kind="hist")
```

[14]: <seaborn.axisgrid.JointGrid at 0x22e1216ae80>



0.6 Additional Challenge

You may want to create new columns based on existing data using Pandas. You can play with the new columns as well.

[35]:
```
# your code is here
new_data = data
new_data['year']= new_data['year'].astype(str)
new_column = new_data[['year','month']].agg('-'.join, axis=1)
new_data['date']= new_column
new_data
```

[35]:

	year	month	passengers	date
0	1949	Jan	112	1949-Jan
1	1949	Feb	118	1949-Feb
2	1949	Mar	132	1949-Mar
3	1949	Apr	129	1949-Apr
4	1949	May	121	1949-May
..	…	…	…	…
139	1960	Aug	606	1960-Aug
140	1960	Sep	508	1960-Sep
141	1960	Oct	461	1960-Oct
142	1960	Nov	390	1960-Nov
143	1960	Dec	432	1960-Dec

[144 rows x 4 columns]

1 Answering the Questions

1.0.1 A1: The plots created with seaborn are of a much higher quality and also the best ones to understand the 'flights' data. A simple line chart is more than sufficient for showing the progressive increase in the amount of passengers through the years.

1.0.2 A3: The most difficult part is figuring out the parameters for a good representation. Not only it's hard to know which parameters are available, but also the values for those same parameters.

1.0.3 A3: The most important takeaway is that it can actually be fun to explore the graphs, even when they don't add much information to what is in question.The data used is a very linear one and at best the graph helps us by showing that in the latter years, the increase in passengers has been much higher than in former years.

1.0.4 A4: My learning of visualization has currently changed very little. I wish we had more exercises here with more specifics like we've had so far. I think this was a fall in quality.

1.0.5 A5: I don't know why this question persists.

[ ]: