Serghei Mangul

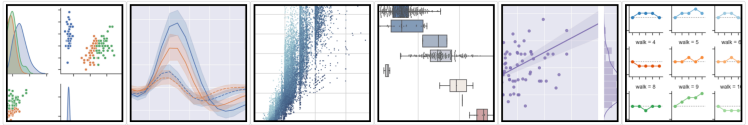
Data visualization using python

Visualization via python

- Pandas
- Seaborn

Seaborn

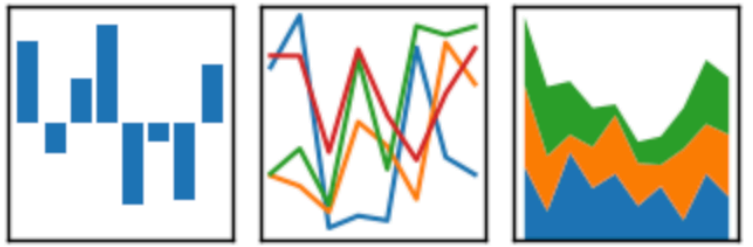
seaborn: statistical data visualization



- **Seaborn** is a Python data visualization library. It provides a high-level interface for drawing attractive and informative statistical graphics
- Allow to generate publication quality figures using few commands

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



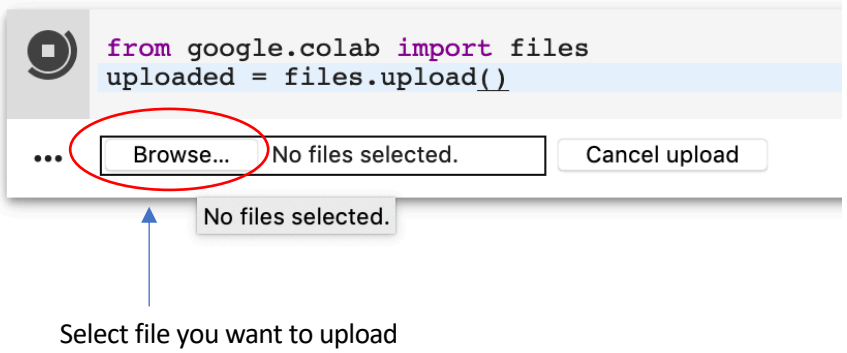
Pandas

- **Pandas** is a software library written for the Python programming language for data manipulation and analysis
- Particularly useful for .csv files
  - Any .xls file can be converted to .csv by using ‘Save as’ option

Upload files

from google.colab import files

uploaded = files.upload()

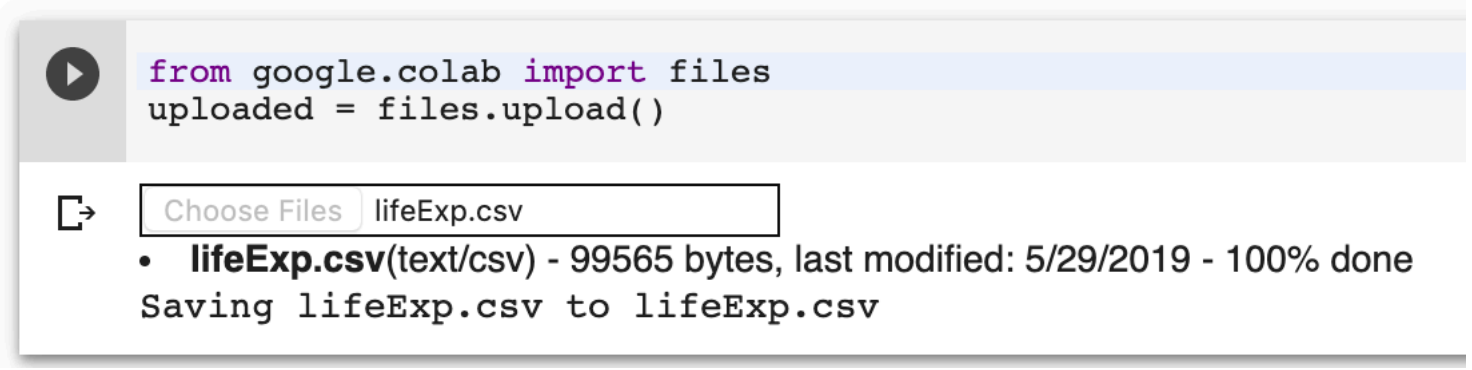


Select file you want to upload

Confirmation that the file was successfully uploaded

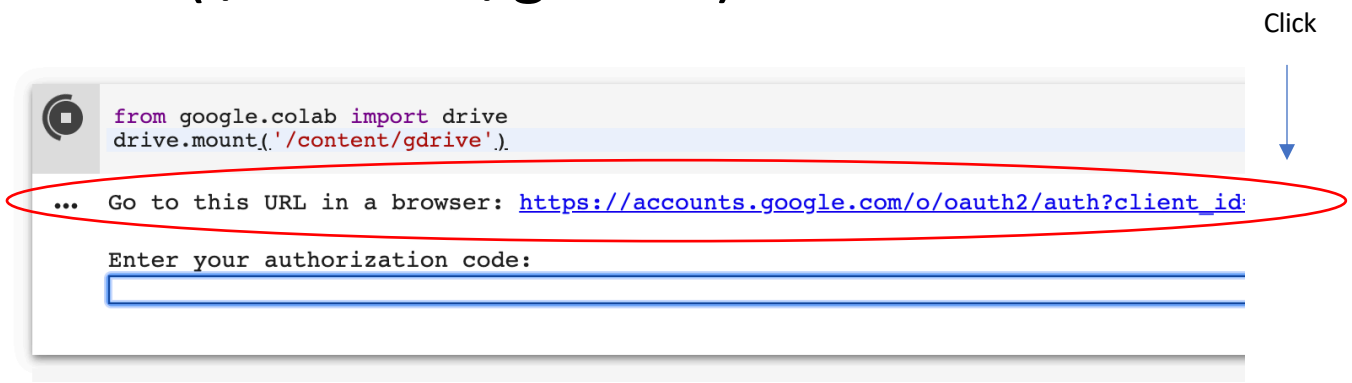
Upload file to Colaboratory - Google

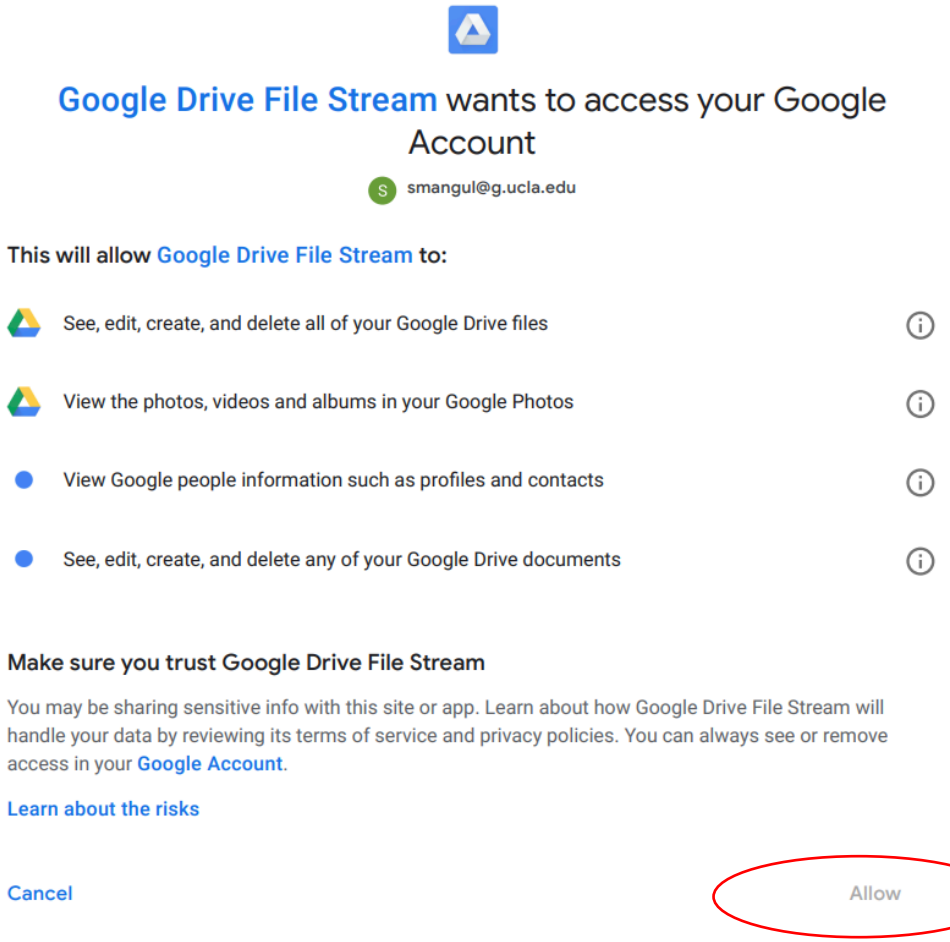
```
from google.colab import files
uploaded = files.upload()
```

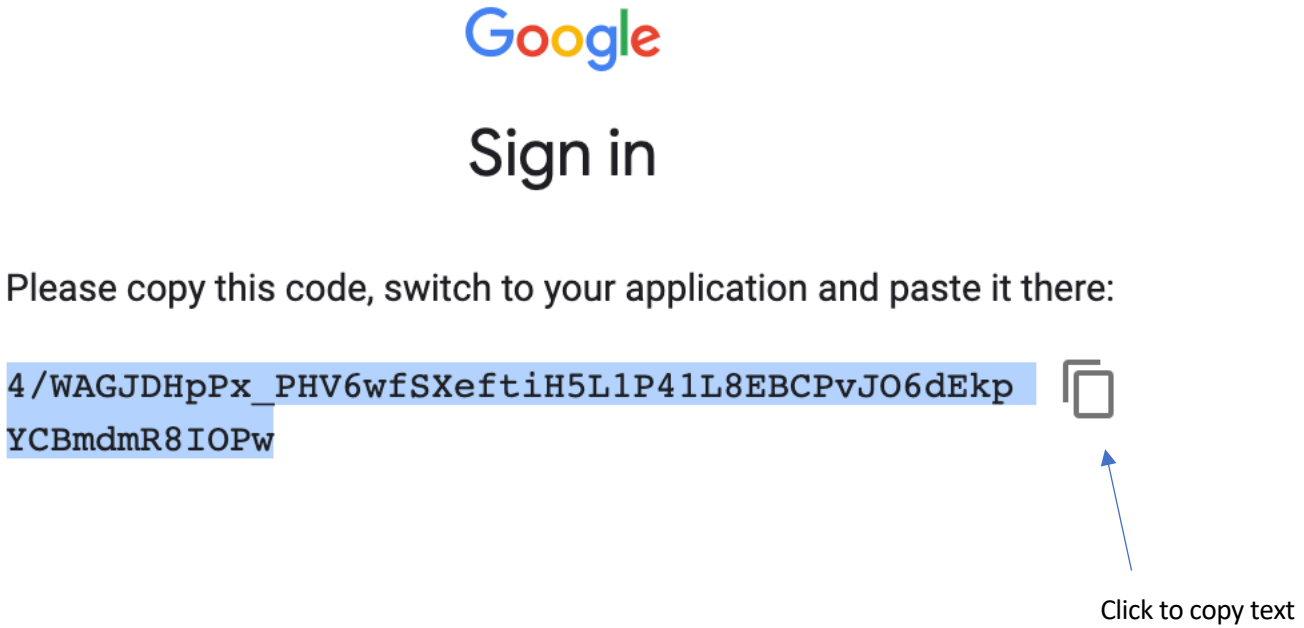


Download files

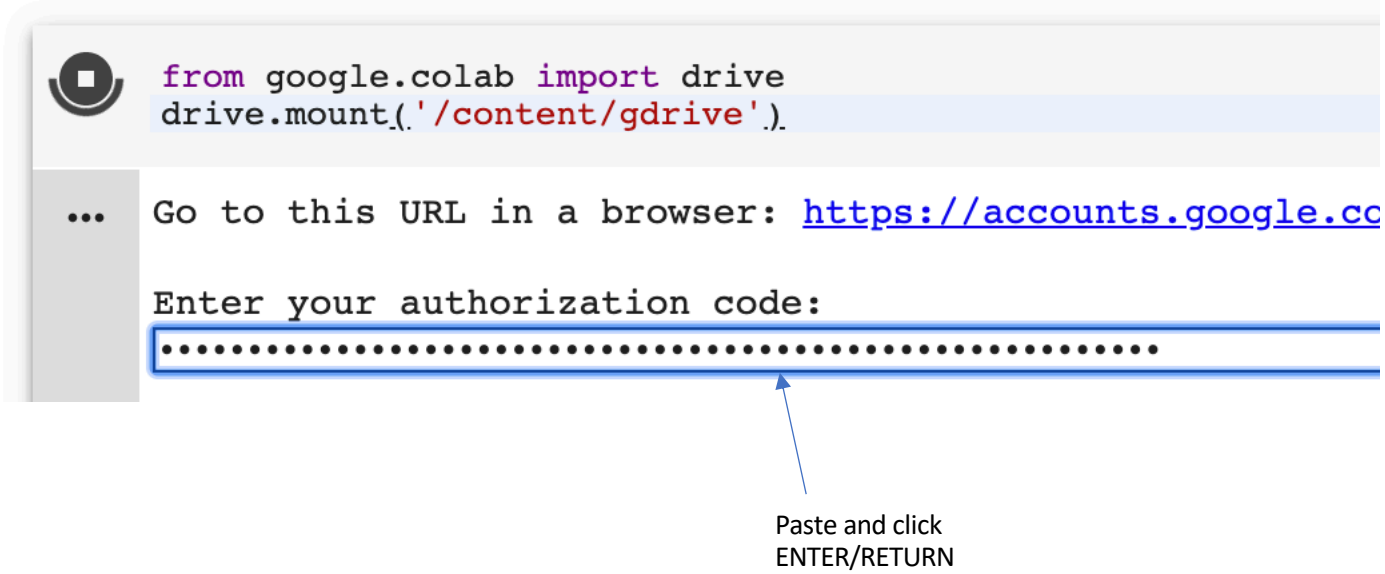
```
from google.colab import drive
drive.mount('/content/gdrive')
```




Google Drive File Stream wants to access your Google Account
smangul@g.ucla.edu
This will allow Google Drive File Stream to:
See, edit, create, and delete all of your Google Drive files
View the photos, videos and albums in your Google Photos
View Google people information such as profiles and contacts
See, edit, create, and delete any of your Google Drive documents
Make sure you trust Google Drive File Stream
You may be sharing sensitive info with this site or app. Learn about how Google Drive File Stream will handle your data by reviewing its terms of service and privacy policies. You can always see or remove access in your Google Account.
Learn about the risks
Cancel
Allow
Click


Google
Sign in
Please copy this code, switch to your application and paste it there:
4/WAGJDHpPx_PHV6wfSXeftiH5L1P41L8EBCPvJO6dEkp
YCBmdmR8IOPw
Click to copy text

Connect notebook with Google Drive



```
%%bash
ls
cd gdrive/My\ Drive
ls
```

```
gdrive
sample_data
7L5A9899.jpg
7L5A9904.jpg
7L5A9911.jpg
7L5A9917.jpg
7L5A9931.jpg
7L5A9938.jpg
```

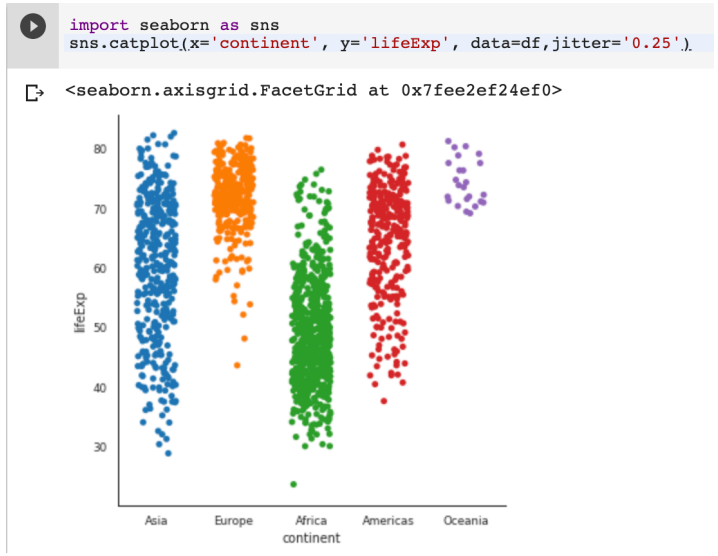
Read csv file using pandas

```
import pandas as pd
df = pd.read_csv('lifeExp.csv')
df.head()
```

```
import pandas as pd
df = pd.read_csv('lifeExp.csv')
df.head()
```

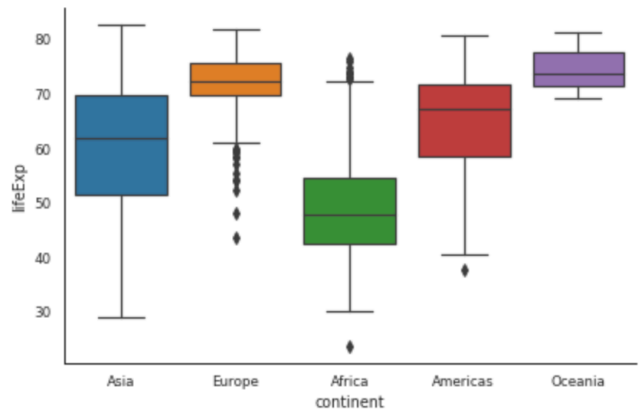
	Unnamed: 0	country	year	pop	continent	lifeExp	gdpPercap
0	0	Afghanistan	1952	8425333.0	Asia	28.801	779.445314
1	1	Afghanistan	1957	9240934.0	Asia	30.332	820.853030
2	2	Afghanistan	1962	10267083.0	Asia	31.997	853.100710
3	3	Afghanistan	1967	11537966.0	Asia	34.020	836.197138
4	4	Afghanistan	1972	13079460.0	Asia	36.088	739.981106

Visualize data as **Stripplot**



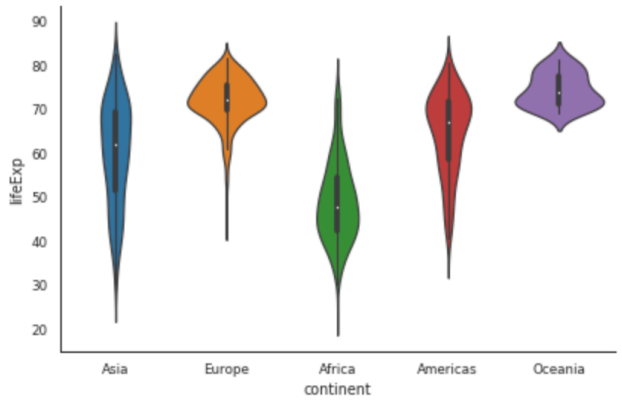

```
import seaborn as sns

sns.catplot(x='continent',
y='lifeExp',
data=df,jitter='0.25')
```

Visualize data as **Boxplot**



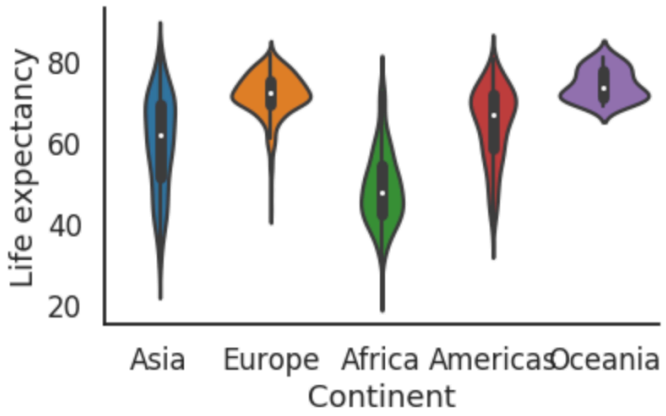

```
import seaborn as sns

sns.catplot(x='continent',
y='lifeExp',
data=df,kind='box',height=4,
aspect=1.5)
```

Visualize data as **violin plot**



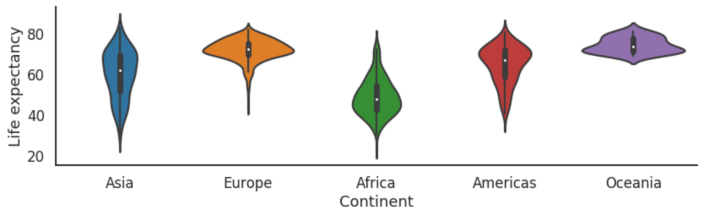

```
import seaborn as sns

sns.catplot(x='continent',
y='lifeExp',
data=df,height=4,aspect=1.5,kind=
'violin')
```

Save as .png or .pdf file




```
import matplotlib.pyplot as plt

import seaborn as sns

g=sns.set_style("white")

g=sns.set_context("paper")

g=sns.catplot(x='continent', y='lifeExp',
data=df,height=4,aspect=1.5,kind='violin')

g.set(xlabel='Continent', ylabel='Life expectancy')

g=sns.despine()

plt.savefig("./gdrive/My Drive/lifeExp.png ")
```

Adjust size of the figure




```
import matplotlib.pyplot as plt

import seaborn as sns

g=sns.set_style("white")

g=sns.set_context("talk")

g=sns.catplot(x='continent', y='lifeExp',
data=df,height=4,aspect=3,kind='violin')

g.set(xlabel='Continent', ylabel='Life expectancy')

g=sns.despine()

plt.savefig("./gdrive/My Drive/lifeExp.png")
```

- I have used this great blog
- https://cmdlinetips.com/2019/03/catplot-in-seaborn-python/