

AnalysisQ4(e)

August 28, 2019

1 DATA420-19S1 Assignment 1 Peng Shen(57408055)

2 Analysis Q4(e)

plot the cumulative rainfall for each country

```
In [1]: import numpy as np
import pandas as pd
import plotly.plotly as plt
import plotly.tools as tls
```

```
In [2]: tls.set_credentials_file(username='dylansp', api_key='I3h0HdVaQKa1gbSLegzU')
```

```
In [3]: # Load the parquet file as pandas dataframe
df = pd.read_parquet('./cum_rainfall_by_country.parquet', engine='pyarrow')
df.head()
```

```
Out[3]:
```

	COUNTRY_CODE	COUNTRY_NAME	CUM_RAINFALL
0	EK	Equatorial Guinea	623.607759
1	BH	Belize	192.375666
2	CS	Costa Rica	182.552504
3	TL	Tokelau [New Zealand]	179.442811
4	NC	New Caledonia [France]	172.842959

```
In [4]: df.describe()
```

```
Out[4]:
```

	CUM_RAINFALL
count	217.000000
mean	53.253083
std	54.120876
min	5.224139
25%	23.003007
50%	39.370455
75%	65.890637
max	623.607759

```
In [5]: # load the dataset containing 3-letter country code with corresponding 2-letter code
country = pd.read_csv("country-codes.csv")
country.head()
```

```
Out[5]: English short name lower case Alpha-2 code Alpha-3 code Numeric code \
0          Afghanistan                AF          AFG          4
1          Åland Islands                AX          ALA         248
2          Albania                    AL          ALB          8
3          Algeria                    DZ          DZA         12
4          American Samoa              AS          ASM         16
```

```
ISO 3166-2
0 ISO 3166-2:AF
1 ISO 3166-2:AX
2 ISO 3166-2:AL
3 ISO 3166-2:DZ
4 ISO 3166-2:AS
```

```
In [6]: # Create a subset of country dataframe above
country = country.iloc[:, [0,1,2]]
country.columns = ['COUNTRY_NAME', 'ALPHA2_CODE', 'ALPHA3_CODE' ]
country.head()
```

```
Out[6]: COUNTRY_NAME ALPHA2_CODE ALPHA3_CODE
0      Afghanistan          AF          AFG
1      Åland Islands          AX          ALA
2          Albania          AL          ALB
3          Algeria          DZ          DZA
4    American Samoa          AS          ASM
```

```
In [7]: # Add alpha-3 code to our data
df = pd.merge(df, country, how='left', on=['COUNTRY_NAME'])
df.head()
```

```
Out[7]: COUNTRY_CODE COUNTRY_NAME CUM_RAINFALL ALPHA2_CODE ALPHA3_CODE
0      EK      Equatorial Guinea    623.607759          GQ          GNQ
1      BH              Belize    192.375666          BZ          BLZ
2      CS      Costa Rica    182.552504          CR          CRI
3      TL  Tokelau [New Zealand]    179.442811         NaN          NaN
4      NC  New Caledonia [France]    172.842959         NaN          NaN
```

```
In [8]: # We can find that some inconsistency between two tables in country name, then we will
df = pd.merge(df, country, how='left', left_on='COUNTRY_CODE', right_on='ALPHA2_CODE' )
df.head()
```

```
Out[8]: COUNTRY_CODE COUNTRY_NAME_x CUM_RAINFALL ALPHA2_CODE_x \
0      EK      Equatorial Guinea    623.607759          GQ
1      BH              Belize    192.375666          BZ
2      CS      Costa Rica    182.552504          CR
3      TL  Tokelau [New Zealand]    179.442811         NaN
4      NC  New Caledonia [France]    172.842959         NaN
```

```
ALPHA3_CODE_x COUNTRY_NAME_y ALPHA2_CODE_y ALPHA3_CODE_y
```

0	GNQ	NaN	NaN	NaN
1	BLZ	Bahrain	BH	BHR
2	CRI	NaN	NaN	NaN
3	NaN	Timor-Leste	TL	TLS
4	NaN	New Caledonia	NC	NCL

In [9]: # Get the final table with Alpha-3 code

```
mask = df.ALPHA3_CODE_x.isnull()
df.loc[mask, 'ALPHA3_CODE_x'] = df[mask] ['ALPHA3_CODE_y']
df.head()
```

Out[9]:

	COUNTRY_CODE	COUNTRY_NAME_x	CUM_RAINFALL	ALPHA2_CODE_x	\
0	EK	Equatorial Guinea	623.607759	GQ	
1	BH	Belize	192.375666	BZ	
2	CS	Costa Rica	182.552504	CR	
3	TL	Tokelau [New Zealand]	179.442811	NaN	
4	NC	New Caledonia [France]	172.842959	NaN	

	ALPHA3_CODE_x	COUNTRY_NAME_y	ALPHA2_CODE_y	ALPHA3_CODE_y
0	GNQ	NaN	NaN	NaN
1	BLZ	Bahrain	BH	BHR
2	CRI	NaN	NaN	NaN
3	TLS	Timor-Leste	TL	TLS
4	NCL	New Caledonia	NC	NCL

In [10]: # Define elements for

```
data = [dict(type='choropleth',
             autocolorscale=True,
             locations=df.ALPHA3_CODE_x,
             z=df.CUM_RAINFALL,
             locationmode='ISO-3',
             colorbar=dict(title='cumulative rainfall')
        )
    ]
data
```

Out[10]:

0	GNQ
1	BLZ
2	CRI
3	TLS
4	NCL
5	DOM
6	NaN
7	TON
8	DMA
9	GUY
10	QAT
11	LAO
12	GNB

13	SLV
14	BHR
15	NGA
16	BMU
17	TUV
18	SLE
19	NaN
20	SUR
21	ATA
22	HND
23	JAM
24	SLB
25	FSM
26	NaN
27	PLW
28	CMR
29	NaN
	...
187	BLR
188	ZAF
189	GRC
190	ARM
191	NaN
192	ROU
193	UKR
194	SWE
195	SYR
196	GRL
197	FLK
198	TJK
199	ESP
200	OMN
201	FIN
202	MAR
203	CZE
204	RUS
205	MNG
206	NAM
207	CYP
208	ISR
209	KGZ
210	DZA
211	SLV
212	UZB
213	ERI
214	TKM
215	EGY
216	ESH

Name: ALPHA3_CODE_x, Length: 217, dtype: object, 'z': 0 623.607759

1	192.375666
2	182.552504
3	179.442811
4	172.842959
5	167.398318
6	152.721431
7	145.849839
8	144.105246
9	143.642269
10	136.923599
11	129.357080
12	125.598585
13	123.150394
14	117.410242
15	115.435700
16	114.764076
17	114.656136
18	113.781187
19	112.412988
20	112.181660
21	111.655264
22	111.239524
23	107.402789
24	103.840805
25	102.967195
26	100.861245
27	99.548792
28	98.130897
29	95.450890
	...
187	17.870352
188	17.859072
189	17.856126
190	17.403298
191	17.239434
192	17.131301
193	16.888682
194	16.747105
195	16.503423
196	16.201742
197	16.107007
198	16.053343
199	15.855416
200	15.733303
201	15.700481
202	15.612557
203	15.425811

```

204      15.368403
205      14.659854
206      13.970321
207      13.406229
208      12.814487
209      12.687286
210      12.507393
211      12.022520
212      11.736697
213      11.282357
214       9.764121
215       6.509744
216       5.224139

```

Name: CUM_RAINFALL, Length: 217, dtype: float64, 'locationmode': 'ISO-3', 'colorbar'

```
In [11]: # Define layout
```

```

layout = dict(title='Cumulative Rainfall for Countries',
              geo = dict(scope='world',
                        projection=dict(type='natural earth')
                        )
              )
layout

```

```
Out[11]: {'title': 'Cumulative Rainfall for Countries',
          'geo': {'scope': 'world', 'projection': {'type': 'natural earth'}}}
```

```
In [12]: fig = dict(data=data, layout=layout)
plt.iplot(fig, filename='Cumulative Rainfall')
```

High five! You successfully sent some data to your account on plotly. View your plot in your browser.

/Users/dylan/anaconda3/lib/python3.7/site-packages/IPython/core/display.py:689: UserWarning:

Consider using IPython.display.IFrame instead

```
Out[12]: <chart_studio.tools.PlotlyDisplay object>
```

```
In [ ]:
```