data_cleaning_script

April 17, 2018

1 Data Cleaning code in Python

In [1]:
```
import pandas as pd
import numpy as np
from numpy import nan as Nan
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
terror=pd.read_csv('./input/globalterrorismdb_0617dist.csv',encoding='ISO-8859-1')
terror.rename(columns={'iyear':'Year','imonth':'Month','iday':'Day','country_txt':'Cou
terror=terror[['Year','Month','Day','Country','Region','city','latitude','longitude','
terror['casualities']=terror['Killed']+terror['Wounded']

#This is a basic table
terror_og = terror.reset_index()
terror = terror[["Year", "Month", "Killed", "Wounded", "casualities"]]

#condition
terror["kill thresh"] = terror["casualities"] >= 10

#This is the table that we can mess with
terror1 = terror.groupby(['Year', 'Month']).agg(['sum']).reset_index()
terror1.columns = [' '.join(col).strip() for col in terror1.columns.values]
#terror1 = terror1.drop(["Wounded size", "casualities size"], axis=1)
terror1 = terror1.rename(index=str, columns={"kill thresh sum": "num attacks with kill
terror1 = terror1[terror1["Month"] != 0]

#Add 1993 Nan values
sub_terror_1993 = pd.DataFrame(np.array([[1993, i, Nan,Nan, Nan, Nan] for i in range(1
sub_terror_1993.columns = terror1.columns
sub_terror_1993["Year"] = sub_terror_1993["Year"].astype(int)
sub_terror_1993["Month"] = sub_terror_1993["Month"].astype(int)

terror2 = pd.concat([terror1[:276], sub_terror_1993, terror1[276:]]).reset_index().dro
#terror2.to_csv("./input/casualities.csv", index=False)
```

```
/Users/cheng/miniconda3/envs/tensorflow/lib/python3.5/site-packages/IPython/core/interactivesh
```

```
interactivity=interactivity, compiler=compiler, result=result)
```