sample_convert_phi

March 15, 2022

The Coastal Grain Size Portal (C-GRASP) dataset Will Speiser, Daniel Buscombe, Evan Goldstein > Convert Sample Measurements to Phi

The purpose of this notebook
This notebook will output a CSV containing all of the data from a chosen C-GRASP dataset with all measurements in mm units converted to Phi units. As C-Grasp file sizes vary completion of this task will vary with internet connectivity. This notebook provides simple code in order to convert measurements in mm to phi for each sample within a dataset.
To do so, a user can choose a dataset of choice. The notebook then runs a simple conversion on each column containing mm untis and converts them to Phi.

[1]:
```
import pandas as pd
import numpy as np
import ipywidgets
```

Select a dataset

[2]:
```
#Dataset collection widget
zen=ipywidgets.Select(
    options=['Entire Dataset', 'Estimated Onshore Data', 'Verified Onshore␣
 ↪Data', 'Verified Onshore Post 2012 Data'],
    value='Entire Dataset',
    # rows=10,
    description='Dataset:',
    disabled=False
)

display(zen)
```

```
Select(description='Dataset:', options=('Entire Dataset', 'Estimated Onshore␣
 ↪Data', 'Verified Onshore Data', '…
```

Download that dataset

[3]:
```
url = 'https://zenodo.org/record/6099266/files/'
if zen.value=='Entire Dataset':
    filename='dataset_10kmcoast.csv'
if zen.value=='Estimated Onshore Data':
    filename='Data_EstimatedOnshore.csv'
```

```
if zen.value=='Verified Onshore Data':
    filename='Data_VerifiedOnshore.csv'
if zen.value=='Verified Onshore Post 2012 Data':
    filename='Data_Post2012_VerifiedOnshore.csv'

print("Downloading {}".format(url+filename))
```

```
Downloading https://zenodo.org/record/6099266/files/dataset_10kmcoast.csv
```

The next cell will download the CGRASP dataset and read it in as a pandas dataframe with variable name `df`

[4]:
```
url=(url+filename)
print('Retrieving Data, Please Wait')
#retrieve data
df=pd.read_csv(url)
print('Sediment Data Retrieved!')
```

```
Retrieving Data, Please Wait
Sediment Data Retrieved!
```

```
/tmp/ipykernel_889149/2390288443.py:4: DtypeWarning: Columns (6,11) have mixed
types. Specify dtype option on import or set low_memory=False.
  df=pd.read_csv(url)
```

Let's take a quick look at the top of the file

[5]:
```
df.head()
```

[5]:

	ID	Sample_ID	Sample_Type_Code	Project	dataset	Date	\
0	81	NaN	NaN	ussb_project_259	US_SeaBed	NaN	
1	80	NaN	NaN	ussb_project_259	US_SeaBed	NaN	
2	85	NaN	NaN	ussb_project_115	US_SeaBed	NaN	
3	86	NaN	NaN	ussb_project_115	US_SeaBed	NaN	
4	88	NaN	NaN	ussb_project_115	US_SeaBed	NaN	

	Location_Type	latitude	longitude	Contact	…	d16	d25	d30	d50	d65	d75	\
0	NaN	25.96090	-97.12251	NaN	…	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	25.96090	-97.12251	NaN	…	NaN	NaN	NaN	NaN	NaN	NaN	
2	NaN	25.96667	-97.08334	NaN	…	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	25.96667	-97.08334	NaN	…	NaN	NaN	NaN	NaN	NaN	NaN	
4	NaN	25.96667	-97.09972	NaN	…	NaN	NaN	NaN	NaN	NaN	NaN	

	d84	d90	d95	Notes
0	NaN	NaN	NaN	LocnName: Chart_11301_2001_3;ObsvnKey: 302126;…
1	NaN	NaN	NaN	LocnName: Chart_11301_2001_3;ObsvnKey: 302126;…
2	NaN	NaN	NaN	LocnName: HE-20-2-92: H-10429: (CatNo: 48734);…
3	NaN	NaN	NaN	LocnName: HE-20-2-92: H-10429: (CatNo: 48734);…
4	NaN	NaN	NaN	LocnName: HE-20-2-92: H-10429: (CatNo: 48735);…

```
[5 rows x 34 columns]
```

0.0.1 Convert from mm to Phi

This cell converts a column in mm to Phi units using numpy by looping through each column listed as containing mm values

[6]:
```
measurement_columns=('Grainsize','Mean','Median','Kurtosis','Skewness','Std',␣
↪'d5',
                     ␣
↪'d10','d16','d25','d30','d50','d65','d75','d84','d90','d95') #All columns␣
↪with mm measurements


i=0
for i in range(0, len(measurement_columns)): #loop through each field in mm
    df[measurement_columns[i]]= -1*(np.log2((df[measurement_columns[i]].
↪astype(float)))) #convert that field to mm
    i=i+1

print('Done!')
```

```
/home/marda/anaconda3/envs/cgrasp/lib/python3.10/site-
packages/pandas/core/arraylike.py:397: RuntimeWarning: divide by zero
encountered in log2
  result = getattr(ufunc, method)(*inputs, **kwargs)
/home/marda/anaconda3/envs/cgrasp/lib/python3.10/site-
packages/pandas/core/arraylike.py:397: RuntimeWarning: invalid value encountered
in log2
  result = getattr(ufunc, method)(*inputs, **kwargs)
```

```
Done!
```

Let's view the file again, this time converted to phi units

[7]:
```
df
```

[7]:
```
             ID Sample_ID  Sample_Type_Code           Project  \
0            81       NaN               NaN  ussb_project_259
1            80       NaN               NaN  ussb_project_259
2            85       NaN               NaN  ussb_project_115
3            86       NaN               NaN  ussb_project_115
4            88       NaN               NaN  ussb_project_115
…         …       …                 …                 …
435196   636550       NaN               NaN  ussb_project_251
435197   636549       NaN               NaN  ussb_project_251
435198   636554      H152               NaN         USGS/WHOI
```

```
435199  636553      H151               NaN         USGS/WHOI
435200  636555      H153               NaN         USGS/WHOI

                                          dataset        Date Location_Type  \
0                                        US_SeaBed         NaN           NaN
1                                        US_SeaBed         NaN           NaN
2                                        US_SeaBed         NaN           NaN
3                                        US_SeaBed         NaN           NaN
4                                        US_SeaBed         NaN           NaN
…                                           …          …           …
435196                                   US_SeaBed         NaN           NaN
435197                                   US_SeaBed         NaN           NaN
435198  USGS East Coast Sediment Texture Database  1964-11-06           NaN
435199  USGS East Coast Sediment Texture Database  1964-11-06           NaN
435200  USGS East Coast Sediment Texture Database  1964-11-06           NaN

        latitude  longitude Contact  …  d16 d25  d30  d50  d65 d75  d84 d90  \
0       25.96090  -97.12251     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
1       25.96090  -97.12251     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
2       25.96667  -97.08334     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
3       25.96667  -97.08334     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
4       25.96667  -97.09972     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
…          …         …      …   …  …  ..  …  …  …  ..  …  ..
435196  45.10000  -67.63167     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
435197  45.10000  -67.63167     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
435198  46.08667  -67.55000     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
435199  46.08667  -67.55000     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN
435200  46.08667  -67.55000     NaN  …  NaN NaN  NaN  NaN  NaN NaN  NaN NaN

        d95                                                 Notes
0       NaN  LocnName: Chart_11301_2001_3;ObsvnKey: 302126;…
1       NaN  LocnName: Chart_11301_2001_3;ObsvnKey: 302126;…
2       NaN  LocnName: HE-20-2-92: H-10429: (CatNo: 48734);…
3       NaN  LocnName: HE-20-2-92: H-10429: (CatNo: 48734);…
4       NaN  LocnName: HE-20-2-92: H-10429: (CatNo: 48735);…
…     …                                                   …
435196  NaN  LocnName: H010;ObsvnKey: 214472;Device: UnidDe…
435197  NaN  LocnName: H010;ObsvnKey: 214472;Device: UnidDe…
435198  NaN  Listed Location: WOODSTOCK,N.BRUNSWICK Litholo…
435199  NaN  Listed Location: WOODSTOCK,N.BRUNSWICK Litholo…
435200  NaN  Listed Location: WOODSTOCK,N.BRUNSWICK Litholo…

[435201 rows x 34 columns]
```

0.0.2 Write to file

Finally, define a csv file name for the output dataframe

[ ]:
```
output_csvfile='../data_phi.csv'
```

Write the data to that csv file

[ ]:
```
df.to_csv(output_csvfile)
```