Natural Resource Protection & Child Health Indicators

Socioeconomic Data and Applications Center (SEDAC)
A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIESIN at Columbia University

CONTENTS

SEDAC DATASET

1. **NRPI_v2019_xx:** The Natural Resource Protection Indicator is the proximity to target of 17% protection across all biomes on a 0-100 scale (2019 release).100 corresponds to 17% protection across all biomes, and 0 corresponds to zero protection.
2. **cmr_xx:** Deaths of children age 1 to exact age 5 per 1,000 live births in that year (4q1)
3. **wat_xx:** water (raw and proximity to target), 0-100 scale
4. **san_xx:** sanitation (raw and proximity to target), 0-100 scale
5. **chmort_pt_xx:** Probability of dying between age 1 and 5 (4q1) (expressed in chances out of 1,000)
6. **CHI_v2019_xx:** Child Health Indicator, unweighted average of three proximity to target indicators - water, sanitation, and mortality (2019 release)

United Nations Department of Economic and Social Affairs Statistics Grouping by ISO3:

1. **Region Name**
2. **Sub-region Name**

Millennium Challenge Corporation Initiatives: https://www.mcc.gov/initiatives

> **Abstract:**
>
> "The Natural Resource Protection and Child Health Indicators, 2019 Release, is produced in support of the U.S. Millennium Challenge Corporation (MCC) as selection criteria for funding eligibility. The Natural Resource Protection Indicator (NRPI) and Child Health Indicator (CHI) are based on proximity-to-target scores ranging from 0 to 100 (at target). The NRPI covers 234 countries and is calculated based on the weighted average percentage of biomes under protected status. The CHI is a composite index for 195 countries derived from the average of three proximity-to-target scores for access to at least basic water and sanitation, along with child mortality. The 2019 release includes a consistent time series of NRPI scores for 2015 to 2019 and CHI scores for 2010 to 2018."

> **Initiatives in Analysis**
>
> 1. **Aid Effectiveness:** Region & Sub-Region, where do Child Health Indicators have the lowest 3 year mean?
> 2. **Climate-Resilient Development:** Where can relationships be identified between Natural Resource Protection & Child Health?
> 3. **Country-Led Poverty Reduction:** Which countries show significant relationships for inferces on future funding?

EDA

In [1]:
```python
# import libraries
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.stats as stats
import statsmodels.api as sm
import statsmodels.formula.api as smf
import requests
from requests.auth import HTTPBasicAuth
import zipfile
import urllib.request
import shutil
```

In [2]:
```python
# bring in Natural Resource Protection Indicator (NRPI) data
resources_df = pd.read_excel('http://atxmoon.com/cv/assets/custom/data/resource-health-indicators.xlsx'
, sheet_name='NRPI_2019')
```

In [3]:
```python
# bring in Child Health Indicator data
child_health_df = pd.read_excel('http://atxmoon.com/cv/assets/custom/data/resource-health-indicators.xl
sx', sheet_name='CHI_2019')
```

In [4]:
```python
# keep only 2016 forward (observe current state of indicators)
child_health_df = child_health_df.drop(columns=['cmr_10', 'cmr_11', 'cmr_12', 'cmr_13', 'cmr_14', 'cmr_
15', 'wat_10', 'wat_11', 'wat_12', 'wat_13', 'wat_14', 'wat_15', 'san_10', 'san_11', 'san_12', 'san_13'
, 'san_14', 'san_15', 'chmort_pt_10', 'chmort_pt_11', 'chmort_pt_12', 'chmort_pt_13', 'chmort_pt_14',
'chmort_pt_15', 'CHI_v2019_10', 'CHI_v2019_11', 'CHI_v2019_12', 'CHI_v2019_13', 'CHI_v2019_14', 'CHI_v2
019_15'])
```

In [5]:
```python
# Join Natural Resource & Child Health data on ISO3
df = resources_df.merge(child_health_df, left_on='ISO3', right_on='ISO3')
df = df.drop(columns=['NRPI_v2019_15'])
df.head(2)
```

Out[5]:

	ISO3	CountryName_x	NRPI_v2019_16	NRPI_v2019_17	NRPI_v2019_18	NRPI_v2019_19	CountryName_y	cmr_16	cmr_17	cmr_
0	ABW	Aruba	58.964862	58.964862	58.964862	58.964862	Aruba	NaN	NaN	N
1	AFG	Afghanistan	0.613273	0.613273	0.613273	0.613273	Afghanistan	17.086026	16.050861	15.1359

2 rows × 22 columns

In [6]:
```python
# add column for 3 year CMR mean, WAT mean, SAN mean, CHMORT mean, CHI mean, NRPI mean
# approximation of indicators for 2016, 2017, 2018
col = df.loc[: , "NRPI_v2019_16":"NRPI_v2019_18"]
df['nrpi_mean'] = col.mean(axis=1)
col = df.loc[: , "cmr_16":"cmr_18"]
df['cmr_mean'] = col.mean(axis=1)
col = df.loc[: , "wat_16":"wat_18"]
df['wat_mean'] = col.mean(axis=1)
col = df.loc[: , "san_16":"san_18"]
df['san_mean'] = col.mean(axis=1)
col = df.loc[: , "chmort_pt_16":"chmort_pt_18"]
df['chmort_pt_mean'] = col.mean(axis=1)
col = df.loc[: , "CHI_v2019_16":"CHI_v2019_18"]
df['CHI_mean'] = col.mean(axis=1)
df.head(2)
```

Out[6]:

	ISO3	CountryName_x	NRPI_v2019_16	NRPI_v2019_17	NRPI_v2019_18	NRPI_v2019_19	CountryName_y	cmr_16	cmr_17	cmr_
0	ABW	Aruba	58.964862	58.964862	58.964862	58.964862	Aruba	NaN	NaN	N
1	AFG	Afghanistan	0.613273	0.613273	0.613273	0.613273	Afghanistan	17.086026	16.050861	15.1359

2 rows × 28 columns

In [7]:
```python
# bring in regional/geographic grouping by ISO3
# (observe current state of indicators by geographic region)
temp_df = pd.read_csv('https://datahub.io/core/country-codes/r/country-codes.csv')
```

In [8]:
```python
# keep only regional/geographic grouping
temp_df = temp_df[['ISO3166-1-Alpha-3', 'Region Name', 'Sub-region Name']]
temp_df.head(2)
```

Out[8]:

	ISO3166-1-Alpha-3	Region Name	Sub-region Name
0	TWN	NaN	NaN
1	AFG	Asia	Southern Asia

In [9]:
```python
# Join Natural Resource & Child Health data with Country/Regional Grouping on ISO3
df = df.merge(temp_df, left_on='ISO3', right_on='ISO3166-1-Alpha-3')
df.head(2)
```

Out[9]:

	ISO3	CountryName_x	NRPI_v2019_16	NRPI_v2019_17	NRPI_v2019_18	NRPI_v2019_19	CountryName_y	cmr_16	cmr_17	cmr_
0	ABW	Aruba	58.964862	58.964862	58.964862	58.964862	Aruba	NaN	NaN	N
1	AFG	Afghanistan	0.613273	0.613273	0.613273	0.613273	Afghanistan	17.086026	16.050861	15.1359

2 rows × 31 columns

In [10]:
```python
df = df.drop(columns=['CountryName_y', 'ISO3166-1-Alpha-3'])
```

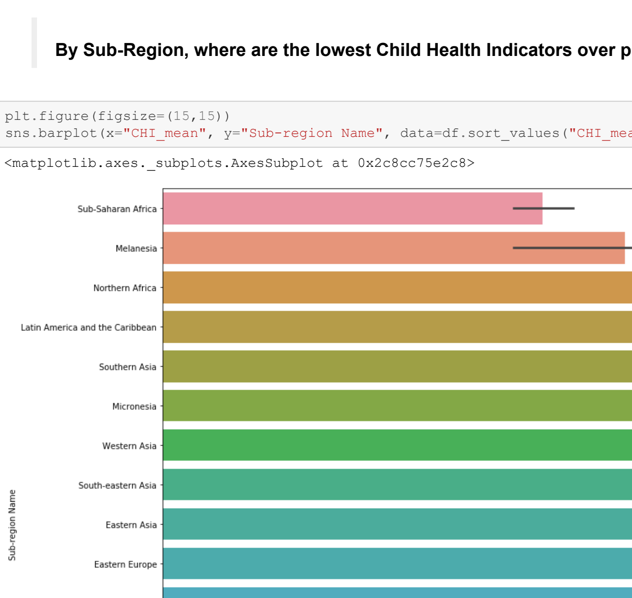
In [11]:
```python
df.to_excel('output.xlsx')
```

Data Visualization

> **By Global Region, where are the lowest Child Health Indicators over preceding 3 years?**

In [12]:
```python
plt.figure(figsize=(10,10))
sns.barplot(x="CHI_mean", y="Region Name", data=df.sort_values("CHI_mean"))
```

Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x2c8cc73ef08>

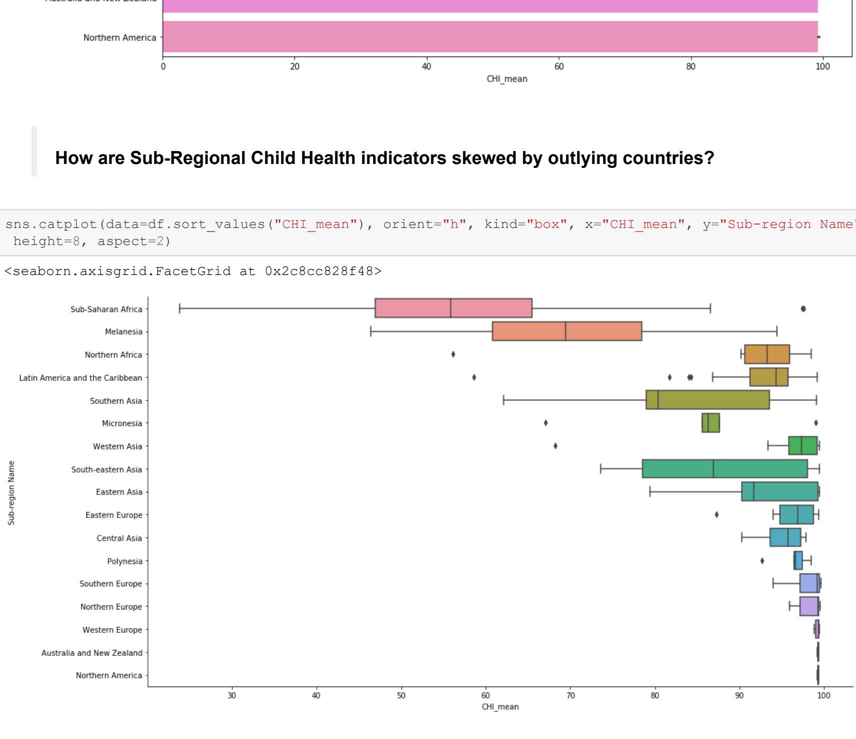


Error bars are graphical representations of the variability of data. We see the most variation in Oceana, followed by Africa.

> **By Sub-Region, where are the lowest Child Health Indicators over preceding 3 years?**

In [13]:
```python
plt.figure(figsize=(15,15))
sns.barplot(x="CHI_mean", y="Sub-region Name", data=df.sort_values("CHI_mean"))
```

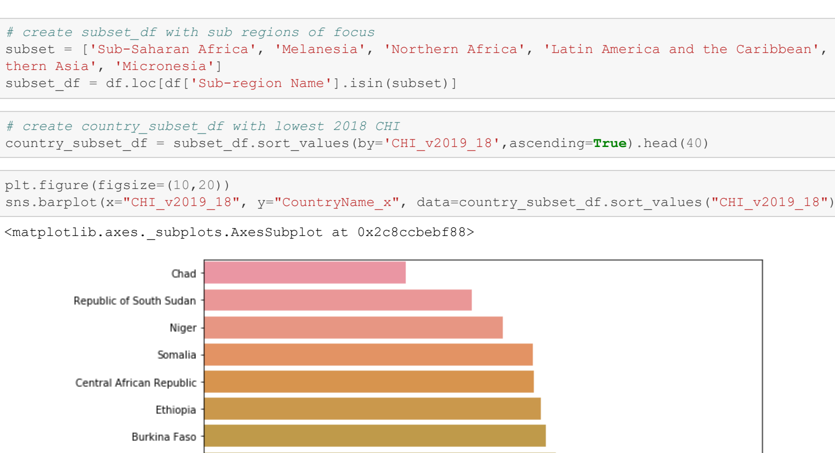
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x2c8cc75e2c8>



> **How are Sub-Regional Child Health indicators skewed by outlying countries?**

In [14]:
```python
sns.catplot(data=df.sort_values("CHI_mean"), orient="h", kind="box", x="CHI_mean", y="Sub-region Name",
 height=8, aspect=2)
```

Out[14]: <seaborn.axisgrid.FacetGrid at 0x2c8cc828f48>



Within Subset of Low CHI Sub-Regions, which countries had the lowest Child Health Indicators in 2018? (most recent year in CHI data)

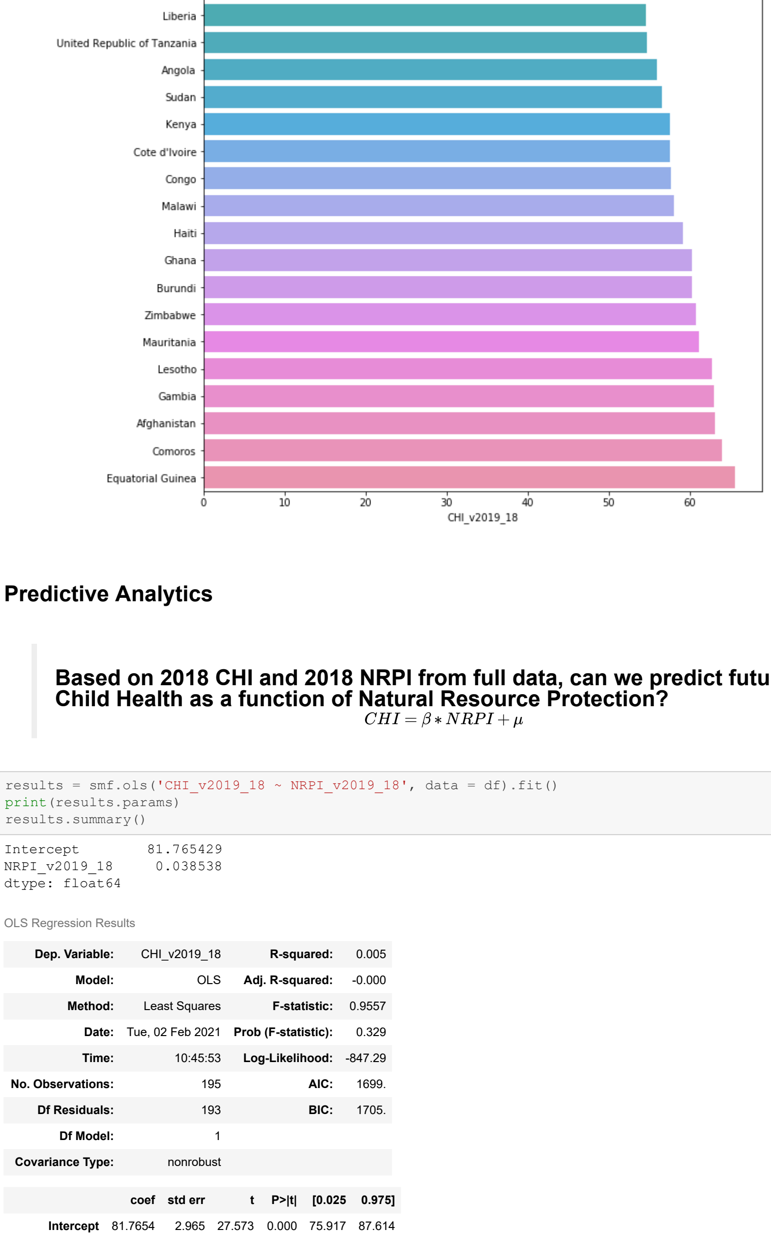
> Sub-Saharan Africa, Melanesia, Northern Africa, LATAM, Southern Asia, Micronesia (keeps 6 of 17 Sub-Regions, 55% of full data set)

In [15]:
```python
# create subset_df with sub regions of focus
subset = ['Sub-Saharan Africa', 'Melanesia', 'Northern Africa', 'Latin America and the Caribbean', 'Sou
thern Asia', 'Micronesia']
subset_df = df.loc[df['Sub-region Name'].isin(subset)]
```

In [16]:
```python
# create country_subset_df with lowest 2018 CHI
country_subset_df = subset_df.sort_values(by='CHI_v2019_18',ascending=True).head(40)
```

In [17]:
```python
plt.figure(figsize=(10,20))
sns.barplot(x="CHI_v2019_18", y="CountryName_x", data=country_subset_df.sort_values("CHI_v2019_18"))
```

Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x2c8cbebf88>



Predictive Analytics

> **Based on 2018 CHI and 2018 NRPI from full data, can we predict future Child Health as a function of Natural Resource Protection?**
> $$CHI = \beta * NRPI + \mu$$

In [18]:
```python
results = smf.ols('CHI_v2019_18 ~ NRPI_v2019_18', data = df).fit()
print(results.params)
results.summary()
```

```
Intercept        81.765429
NRPI_v2019_18     0.038538
dtype: float64
```

Out[18]:

OLS Regression Results

Dep. Variable:	CHI_v2019_18	R-squared:	0.005
Model:	OLS	Adj. R-squared:	-0.000
Method:	Least Squares	F-statistic:	0.9557
Date:	Tue, 02 Feb 2021	Prob (F-statistic):	0.329
Time:	10:45:53	Log-Likelihood:	-847.29
No. Observations:	195	AIC:	1699.
Df Residuals:	193	BIC:	1705.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	81.7654	2.965	27.573	0.000	75.917	87.614
NRPI_v2019_18	0.0385	0.039	0.978	0.329	-0.039	0.116

Omnibus:	34.929	Durbin-Watson:	1.675
Prob(Omnibus):	0.000	Jarque-Bera (JB):	48.990
Skew:	-1.222	Prob(JB):	2.30e-11
Kurtosis:	3.236	Cond. No.	166.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

CITATIONS

> ### Natural Resource & Child Health Indicators
>
> Center for International Earth Science Information Network (CIESIN), Columbia University. 2019. Natural Resource Protection and Child Health Indicators, 2019 Release. Palisades, NY: NASA Socioeconomic Data and Applications Center. https://doi.org/10.7927/r6mv-sv82. Accessed 16 OCT 2020.
>
> ### Regional Country Grouping Data
>
> Datahub. 2020. Comprehensive Country Codes, https://datahub.io/core/country-codes. Accessed 16 OCT 2020.

In [ ]: