Data_Coverage

August 15, 2017

In [1]:
```
from IPython.display import HTML
HTML('''<script>
code_show=true;
function code_toggle() {
 if (code_show){
 $('div.input').hide();
 } else {
 $('div.input').show();
 }
 code_show = !code_show
}
$( document ).ready(code_toggle);
</script>''')
```

Out[1]:
```
<IPython.core.display.HTML object>
```

In [2]:
```
from IPython.display import HTML
HTML('''
<style>
    .yourDiv {position: fixed;top: 100px; right: 0px;
              background: white;
              height: 100%;
              width: 175px;
              padding: 10px;
              z-index: 10000}
</style>
<script>
function showthis(url) {
        window.open(url, "pres",
                "toolbar=yes,scrollbars=yes,resizable=yes,top=10,left=400,width=500,heig
        return(false);
}
</script>

<div class=yourDiv>
    <h4>MENU</h4><br>
    <a href=#Data>1. Data</a><br>
```

```
            <a href=#SpatialCoverage>2. Spatial Coverage</a><br>
            <a href=#TemporalCoverage>3. Temporal Coverage</a><br>
            <a href=#ClassOverlaps>4. Farm size class overlaps</a><br>
            <a href=#YieldLookUpTable>5. Yield look-up table</a><br><br>

            <a href=#Top>Top</a><br>
            <a href="javascript:code_toggle()">Toggle Code On/Off</a><br>
            <a href=#LeftOff>Left Off Here</a><br>
            <a href='https://vinnyricciardi.github.io/farmsize_site/'>Site Index</a><br>
        </div>
        ''')
```

Out[2]: <IPython.core.display.HTML object>

Data Coverage Overview

In [3]:
```
# Import dependencies
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import geopandas as gpd
import seaborn as sns
from matplotlib import pyplot as plt
import matplotlib.pyplot as plt
from matplotlib.path import Path
import matplotlib.patches as patches
from matplotlib.pyplot import cm
import matplotlib as mpl
import numpy as np
import re
import geopy
import mpld3
import plotly.plotly as py
import cmocean

pd.set_option('display.max_columns', 500)
%matplotlib inline
```

In [4]:
```
# Set all plotting params:
title_sz = 20
x_lab_tick_sz = 18
y_lab_tick_sz = 18
x_lab_label_sz = 18
y_lab_label_sz = 18
lengend_sz = 16
```

In [5]:
```
PATH = '/Users/Vinny_Ricciardi/Documents/Data_Library_Big/Survey/Global/Farm_Size/Data/F
df = pd.read_csv(PATH, low_memory=False)
```

Data

```
In [6]: num_countries = len(df.NAME_0.unique())
        num_crops = len(df.Crop.unique())
        num_crops_fao = len(df.query("production_Food_kcal == production_Food_kcal").Crop.unique
        num_admin = len(df.shpID.unique())
        num_obs = len(df)
        num_micro = len(df.query("microdata == 1").NAME_0.unique())
        num_tab = len(df.query("microdata == 0").NAME_0.unique())
        num_sur = len(df.query("cen_sur == 'sur'").NAME_0.unique())
        num_cen = len(df.query("cen_sur == 'cen'").NAME_0.unique())
        avg_year = int(round(df.year.mean(), 0))
        min_year = df.year.min().astype(int)
        max_year = df.year.max().astype(int)
```

General - Our dataset caputres the amount of crops produced by farms of different sizes. - We used the World Census of Agriculture's (WCA) farm size categories to be consistent with other studies. - Our dataset consists of 564134 observations. - 58 countries are represented at either the national or subnational level. - In total, there are 2804 national or subnational units. - There are 151 crops, of which we were able to match 127 with the FAO's database to calculate the amount of crops produced by farm size class for food, feed, waste, seed, proccessing, and other in terms of kcal. - We used 37 tabulated datasets, and 21 microdatasets (i.e., data at the household record level) - 41 agricultural censuses were used. Where census data was not used, nationally or subnationally representative household surveys were used (17 in total). - On average the data was from 2011, with the oldest datasets from 2001 and the newest from 2013

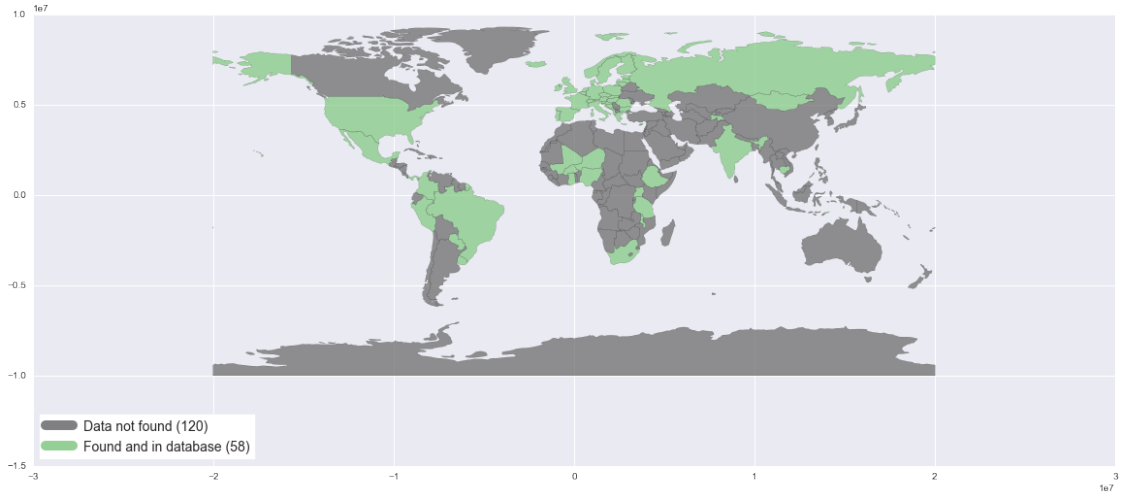
Spatial Coverage

```
In [7]: df['NAME_0'].replace(['United States of America'], ['United States'], inplace=True)
        df['NAME_0'].replace(['Bosnia and Herzegovina'], ['Bosnia and Herz.'], inplace=True)
        df['NAME_0'].replace(['United Republic of Tanzania'], ['Tanzania'], inplace=True)
        df['NAME_0'].replace(['Russian Federation'], ['Russia'], inplace=True)
        df['NAME_0'].replace(['Czech Republic'], ['Czech Rep.'], inplace=True)
        df['NAME_0'].replace(['Czech Republic'], ['Czech Rep.'], inplace=True)
        df['NAME_0'].replace(['Czech Republic'], ['Czech Rep.'], inplace=True)
```

To do: - Map to be replaced with map of sub-national units (and in a better projection!) after we spatially match all admin units

```
In [8]: pivoted = pd.pivot_table(df,
                                  index='NAME_0',
                                  values='Crop',
                                  aggfunc=lambda x: len(x.unique()))
        pivoted = pivoted.reset_index()
        pivoted = pivoted.sort_values('Crop', ascending=False)
        pivoted['Data_Available'] = pivoted['Crop'].astype(int)

        world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))

        world = world.to_crs(epsg=3786)
```

```
world = pd.merge(world, pivoted,
                 how='outer',
                 left_on='name',
                 right_on='NAME_0')

world['Orig_crop'] = world['Crop'].fillna(0)
world['coverage'] = np.where(world['Crop'] > 0,
                             'Found and downloaded',
                             np.where(world['Crop'] == -1,
                                      'Found not downloaded',
                                      'No data found'))

warnings.filterwarnings('ignore')

x = len(pivoted.NAME_0.unique())

try:
    fig, ax = plt.subplots(figsize=(20, 10))
    ax.set_aspect('equal')
    world.plot(column='coverage', cmap='Accent', ax=ax, alpha=0.7, linewidth=0.1)  #cmoc
except:
    pass

ndf, fad = world.coverage.value_counts()
cmap_ = cmocean.tools.get_dict(cmocean.cm.deep, N=4)

p1 = mpl.lines.Line2D([], [],
                          color=[x / 255. for x in [128, 128, 130]],
                          linewidth=10,
                          label='Data not found ({})'.format(ndf))
p2 = mpl.lines.Line2D([], [],
                          color=[x / 255. for x in [148, 207, 150]],
                          linewidth=10,
                          label='Found and in database ({})'.format(fad))

handles = [p1, p2]
labels = [h.get_label() for h in handles]

legend = ax.legend(handles=handles, labels=labels, frameon=True,
                   fontsize=14, loc='lower left')

legend.get_frame().set_facecolor('#ffffff')

plt.show()
```



Left Off
To Do: - Need to put into the global context

```
In [9]: df.head()
```

Out[9]:

	Unnamed: 0	Crop	Item_Code	NAME_0	NAME_1	NAME_2	NAME_3	es1	shpID	data_unit	fs_class_min	fs_class_max	cen_sur	microdata	year	Crop_area	Cultivated_area	Harvested_area	Planted_area	Production	Production_fix	Production_fix_dummy
0	0	Agave fibres nes	800.0	Mexico	1	1	1	MEX	MEX	t	1.0	2.0	sur	1	2007.0	NaN	NaN	NaN	NaN	37.702929	37.702929	0
1	1	Agave fibres nes	800.0	Mexico	1	1	1	MEX	MEX	t	20.0	50.0	sur	1	2007.0	NaN	NaN	NaN	NaN	37.702929	37.702929	0
2	2	Almonds, with shell	221.0	Colombia	Amazonas	La Chorrera	1	COL	COL001002	ha	20.0	50.0	cen	1	2013.0	NaN	NaN	0.0	NaN	NaN	0.000000	1
3	3	Almonds, with shell	221.0	Colombia	Amazonas	La Chorrera	1	COL	COL001002	ha	50.0	100.0	cen	1	2013.0	NaN	NaN	0.0	NaN	NaN	0.000000	1
4	4	Almonds, with shell	221.0	Colombia	Amazonas	La Chorrera	1	COL	COL001002	t	20.0	50.0	cen	1	2013.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
4           NaN    0.000000        0.000000                     0

   Production_constant  perc_Feed  perc_Food  perc_Seed  perc_Waste  \
0                  NaN   0.255413   0.927016   0.000674    0.087989
1                  NaN   0.255413   0.927016   0.000674    0.087989
2                  0.0   0.905842   1.000000   0.135892    0.002301
3                  0.0   0.905842   1.000000   0.135892    0.002301
4                  NaN   0.905842   1.000000   0.135892    0.002301

   perc_Processing  perc_Other  production_Feed  production_Feed_k  \
0         0.877703    1.000000          9.629819                NaN
1         0.877703    1.000000          9.629819                NaN
2         0.782199    0.089903          0.000000                0.0
3         0.782199    0.089903          0.000000                0.0
4         0.782199    0.089903          0.000000                NaN

   production_Food  production_Food_k  production_Other  production_Other_k  \
0        34.951226                NaN         37.702929                 NaN
1        34.951226                NaN         37.702929                 NaN
2         0.000000                0.0          0.000000                 0.0
3         0.000000                0.0          0.000000                 0.0
4         0.000000                NaN          0.000000                 NaN

   production_Seed  production_Seed_k  production_Waste  production_Waste_k  \
0         0.025416                NaN          3.317445                 NaN
1         0.025416                NaN          3.317445                 NaN
2         0.000000                0.0          0.000000                 0.0
3         0.000000                0.0          0.000000                 0.0
4         0.000000                NaN          0.000000                 NaN

   production_Processing  production_Processing_k  kcal   fat  protein  \
0              33.091975                      NaN   NaN   NaN      NaN
1              33.091975                      NaN   NaN   NaN      NaN
2               0.000000                      0.0   1.0  0.11     0.03
3               0.000000                      0.0   1.0  0.11     0.03
4               0.000000                      NaN   1.0  0.11     0.03

   production_Feed_kcal  production_Feed_k_kcal  production_Food_kcal  \
0                   NaN                     NaN                   NaN
1                   NaN                     NaN                   NaN
2                   0.0                     0.0                   0.0
3                   0.0                     0.0                   0.0
4                   0.0                     NaN                   0.0

   production_Food_k_kcal  production_Other_kcal  production_Other_k_kcal  \
0                     NaN                    NaN                      NaN
1                     NaN                    NaN                      NaN
2                     0.0                    0.0                      0.0
```

```
3                    0.0                    0.0                    0.0
4                    NaN                    0.0                    NaN

   production_Seed_kcal  production_Seed_k_kcal  production_Waste_kcal  \
0                   NaN                     NaN                    NaN
1                   NaN                     NaN                    NaN
2                   0.0                     0.0                    0.0
3                   0.0                     0.0                    0.0
4                   0.0                     NaN                    0.0

   production_Waste_k_kcal  production_Processing_kcal  \
0                      NaN                         NaN
1                      NaN                         NaN
2                      0.0                         0.0
3                      0.0                         0.0
4                      NaN                         0.0

   production_Processing_k_kcal  production_Feed_fat  production_Feed_k_fat  \
0                           NaN                 NaN                    NaN
1                           NaN                 NaN                    NaN
2                           0.0                 0.0                    0.0
3                           0.0                 0.0                    0.0
4                           NaN                 0.0                    NaN

   production_Food_fat  production_Food_k_fat  production_Other_fat  \
0                  NaN                    NaN                   NaN
1                  NaN                    NaN                   NaN
2                  0.0                    0.0                   0.0
3                  0.0                    0.0                   0.0
4                  0.0                    NaN                   0.0

   production_Other_k_fat  production_Seed_fat  production_Seed_k_fat  \
0                     NaN                  NaN                    NaN
1                     NaN                  NaN                    NaN
2                     0.0                  0.0                    0.0
3                     0.0                  0.0                    0.0
4                     NaN                  0.0                    NaN

   production_Waste_fat  production_Waste_k_fat  production_Processing_fat  \
0                   NaN                     NaN                        NaN
1                   NaN                     NaN                        NaN
2                   0.0                     0.0                        0.0
3                   0.0                     0.0                        0.0
4                   0.0                     NaN                        0.0

   production_Processing_k_fat  production_Feed_protein  \
0                          NaN                      NaN
1                          NaN                      NaN
```

```
2                          0.0                        0.0
3                          0.0                        0.0
4                          NaN                        0.0

   production_Feed_k_protein  production_Food_protein  \
0                        NaN                      NaN
1                        NaN                      NaN
2                        0.0                      0.0
3                        0.0                      0.0
4                        NaN                      0.0

   production_Food_k_protein  production_Other_protein  \
0                        NaN                       NaN
1                        NaN                       NaN
2                        0.0                       0.0
3                        0.0                       0.0
4                        NaN                       0.0

   production_Other_k_protein  production_Seed_protein  \
0                         NaN                      NaN
1                         NaN                      NaN
2                         0.0                      0.0
3                         0.0                      0.0
4                         NaN                      0.0

   production_Seed_k_protein  production_Waste_protein  \
0                        NaN                       NaN
1                        NaN                       NaN
2                        0.0                       0.0
3                        0.0                       0.0
4                        NaN                       0.0

   production_Waste_k_protein  production_Processing_protein  \
0                         NaN                            NaN
1                         NaN                            NaN
2                         0.0                            0.0
3                         0.0                            0.0
4                         NaN                            0.0

   production_Processing_k_protein
0                              NaN
1                              NaN
2                              0.0
3                              0.0
4                              NaN
```

In [ ]:
```
fao = pd.read_csv('/Users/Vinny_Ricciardi/Documents/Data_Library_Big/Survey/Global/FaoSt
```

In [ ]:
```
fao.head()
```

```
In [ ]: tmp = fao.copy()
        tmp2 = df.copy()

In [ ]: tmp1 = tmp[tmp['Element'] == 'Area harvested']

In [ ]: tmp1 = pd.merge(tmp1, tmp2, how='inner', left_on='Country', right_on='NAME_0', indicator
        tmp1.head()

In [40]:

In [ ]:
```