

Data visualization of world contraceptive use

```
import geopandas
import pandas as pd
import seaborn as sns
import numpy as np
import fiona
import pycountry
from fuzzywuzzy import fuzz
from fuzzywuzzy import process
import matplotlib.pyplot as plt
from mpl_toolkits.axes_grid1 import make_axes_locatable
```

```
# The data is used for map canvas
world = geopandas.read_file(geopandas.datasets.get_path('naturalearth_lowres'))
```

```
# loading contraceptive data
with pd.ExcelFile('Contraceptive_2019.xls') as xls:
    country2019 = pd.read_excel(xls, 'Sheet1', na_values= ["."])
    area2019 = pd.read_excel(xls, 'Sheet2', na_values= ["."])
    country_trend= pd.read_excel(xls, "Sheet3", na_values= ["."])
```

Merge data

```
# The contraceptive data do not contain a col of unique identifiers.
# Use fuzzy merge to join two datasets.
def fuzzy_merge(data1, data2, key1, key2, threshold=95, limit=1):

    s = data2[key2].tolist()

    m = data1[key1].apply(lambda x: process.extract(x, s, limit=limit))
    data1['matches'] = m

    m2 = data1['matches'].apply(lambda x: ', '.join([i[0] for i in x if i[1] >= threshold]))
    data1['matches'] = m2

    return data1
```

```
match=fuzzy_merge(world, country2019, 'name', 'area')
```

```
#match.loc[match['matches'] == ""] ## display countries that have no matches from contraceptive data
```

```
World_con = pd.merge(match, country2019,
                      left_on='matches',
                      right_on='area',
                      how = 'left'
                      )
```

```
# Clean the data
World_con=World_con[World_con.columns.drop('matches')];
World_con=World_con[World_con.columns.drop('area')]
World_con=World_con[World_con.continent!='Antarctica']
```

```
# a quick inspect of the data
round(World_con.describe(), 2)
```

	pop_est	gdp_md_est	Any method	Female sterilisation	Male sterilisation	Pill	Injectable	Implant	IUD	Male condom	Rhythm	Withdrawal	Other methods	women(15- 49 in thousands)	
count	1.760000e+02	176.00	148.00	136.00	136.00	136.00	136.00	136.00	136.00	136.00	136.00	136.00	136.00	148.00	1
mean	4.194935e+07	670681.31	41.33	4.60	0.61	9.80	5.35	1.49	5.16	8.30	1.71	2.64	1.08	11764.28	0
std	1.463514e+08	2295416.80	16.16	6.69	1.55	8.88	6.20	2.44	7.32	8.15	1.73	4.06	1.14	41438.57	0
min	1.400000e+02	16.00	6.50	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	72.00	0
25%	3.408326e+06	25945.00	28.98	0.50	0.00	3.08	0.40	0.10	0.70	1.70	0.50	0.40	0.38	1133.00	0

	pop_est	gdp_md_est	Any method	Female sterilisation	Male sterilisation	Pill	Injectable	Implant	IUD	Male condom	Rhythm	Withdrawal	Other methods	women(15- 49 in thousands)	
50%	1.010473e+07	85045.00	42.40	1.85	0.10	6.25	2.90	0.35	2.05	5.25	1.15	1.20	0.80	2611.00	0
75%	2.947544e+07	411950.00	53.92	4.93	0.20	13.55	7.98	1.83	7.30	12.12	2.30	2.95	1.30	8487.00	0
max	1.379303e+09	21140000.00	78.00	30.60	9.50	34.40	23.40	14.90	47.00	34.90	9.50	24.50	8.80	354103.00	0

Contraceptive prevalence in the world

General trend

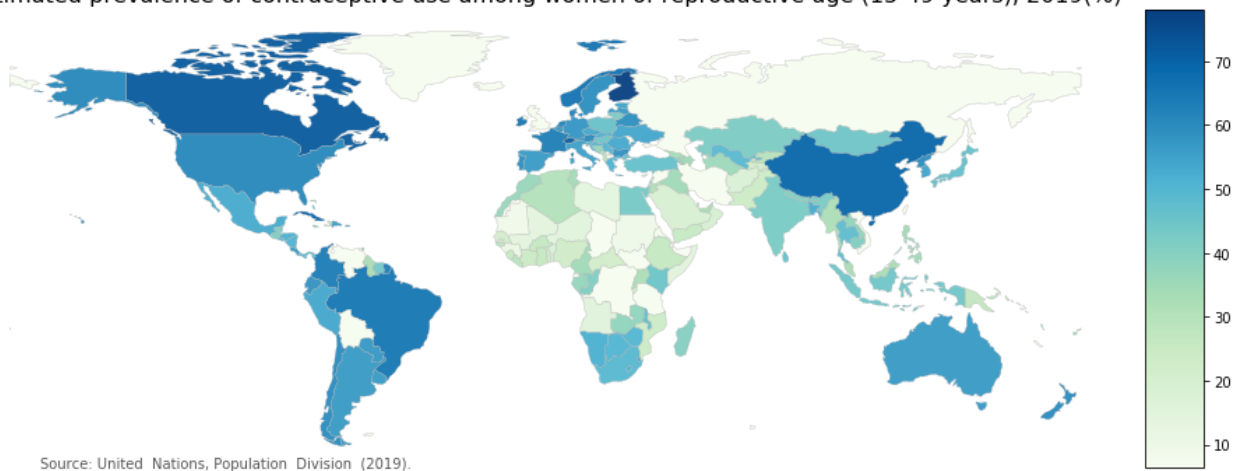
A world map to show the estimated prevalence of contraceptive use among women of reproductive age

```
fig, ax = plt.subplots(1, figsize=(16, 17))
divider = make_axes_locatable(ax) # align the legend to the plot
cax = divider.append_axes("right", size="5%", pad=0.1)
World_con.plot(column='Any method', linewidth=0.3, ax=ax, edgecolor='0.7',alpha=0.95,\
               cax=cax, cmap='GnBu',legend= True)
ax.axis('off')# remove the axis
# add plot title and annotation
ax.set_title('Estimated prevalence of contraceptive use among women of reproductive age (15-49 years), 2019(%)',
            fontdict={'fontsize': '16', 'fontweight': '5','horizontalalignment': 'center'})
ax.annotate('Source: United Nations, Population Division (2019).',xy=(0.1, 0.28),
            xycoords='figure fraction', horizontalalignment='left', verticalalignment='top',
            fontsize=10, color='#555555')
```

/anaconda3/lib/python3.7/site-packages/matplotlib/colors.py:512: RuntimeWarning: invalid value encountered in less xa[xa < 0] = -1

Text(0.1, 0.28, 'Source: United Nations, Population Division (2019).')

Estimated prevalence of contraceptive use among women of reproductive age (15-49 years), 2019(%)



png

Contraceptive use prevalence among women in geographic regions

```
g_contra=area2019.iloc[1:9,:2]
g_contra=g_contra.sort_values("Any method", ascending=False)
```

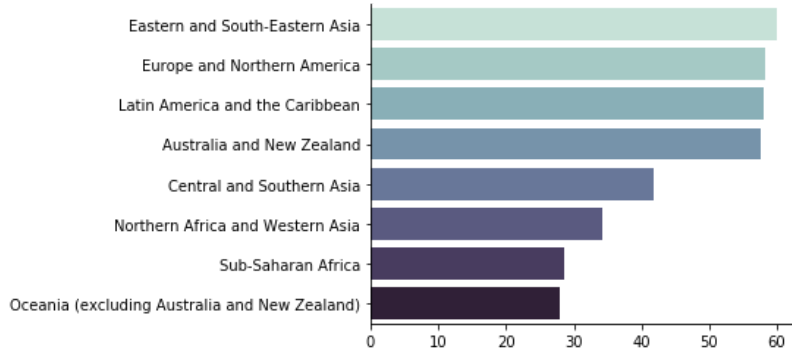
```

g=sns.catplot(y="area", x="Any method",
              palette=(sns.cubehelix_palette(8, start=.5, rot=-.5)),
              height=4,aspect=2, kind="bar", data=g_contra)
#plt.suptitle('Contraceptive use in geographic regions',size=16)
#plt.show()
(g.set_ylabels(""))
.set_xlabels(""))
g.fig.subplots_adjust(top=0.9)
g.fig.suptitle('Contraceptive use prevalence in geographic regions: all method (%)',
              fontsize=18)

```

Text(0.5, 0.98, 'Contraceptive use prevalence in geographic regions: all method (%)')

Contraceptive use prevalence in geographic regions: all method (%)



png

Contraceptive use prevalence among women by income levels

```
inc_contra=area2019.iloc[16:21,[0,1,2,3,4,7,8]]
```

```
inc_contra
```

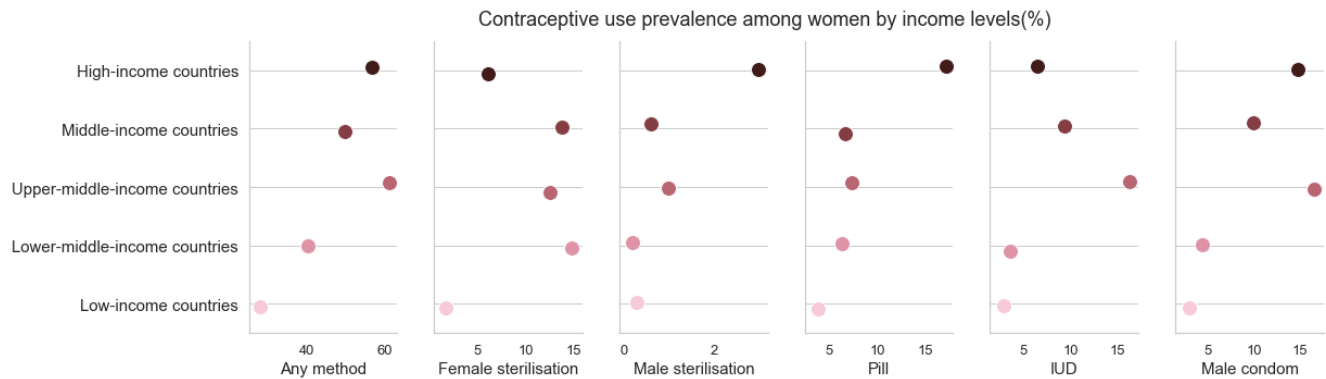
	area	Any method	Female sterilisation	Male sterilisation	Pill	IUD	Male condom
16	High-income countries	56.6	6.1	3.0	17.2	6.5	14.8
17	Middle-income countries	49.6	13.8	0.6	6.7	9.4	9.9
18	Upper-middle-income countries	61.0	12.6	1.0	7.3	16.3	16.6
19	Lower-middle-income countries	40.1	14.8	0.2	6.3	3.6	4.4
20	Low-income countries	28.0	1.7	0.3	3.8	3.0	2.9

```

sns.set(style="whitegrid",rc={'font.size': 15, 'axes.labelsize': 15, 'legend.fontsize': 15,
    'axes.titlesize': 15, 'xtick.labelsize': 13, 'ytick.labelsize': 15})
inc= sns.PairGrid(inc_contra,
                  x_vars=inc_contra.columns[1:7], y_vars=["area"],
                  height=5, aspect=.5)
inc.map(sns.stripplot, size=14, orient="h",
        palette="ch:s=1,r=-.1,h=1_r", linewidth=1, edgecolor="w")
inc.set(ylabel="")
for ax in inc.axes.flat:
    ax.xaxis.grid(False)
    ax.yaxis.grid(True)
inc.fig.subplots_adjust(top=0.9)
inc.fig.suptitle('Contraceptive use prevalence among women by income levels(%)', fontsize=18)

```

Text(0.5, 0.98, 'Contraceptive use prevalence among women by income levels(%)')



png

Country specific trend

```
## Select 12 countries to see specific trend
countryname=[]
for i in ['USA', 'CAN', 'GBR', 'FRA', 'DEU', 'MEX', 'JPN', 'KOR', 'CHN', 'IND', 'AUS', 'TUR']:
    name=pycountry.countries.get(alpha_3=i).numeric
    countryname.append(int(name))
subcountry=country_trend.loc[country_trend['ISO code'].isin(countryname),\
                             ['ISO code', 'area', 'Survey\ntend year', 'Any method', 'Female\nsterilization',\
                              'Pill', 'Male condom' ] ]
```

```
subcountry = subcountry.rename(columns = {'ISO code':'iso', 'Survey\ntend year':'year', 'Any method':'Any method' ,\
                                          'Female\nsterilization':'Female sterilization', \
                                          'Pill':'Pill', 'Male condom': 'Male condom'})
```

```
round(pd.pivot_table(subcountry, values=['Any method', 'Female sterilization', 'Pill', 'Male condom'],
                     index=['area'],aggfunc=np.mean),2)
```

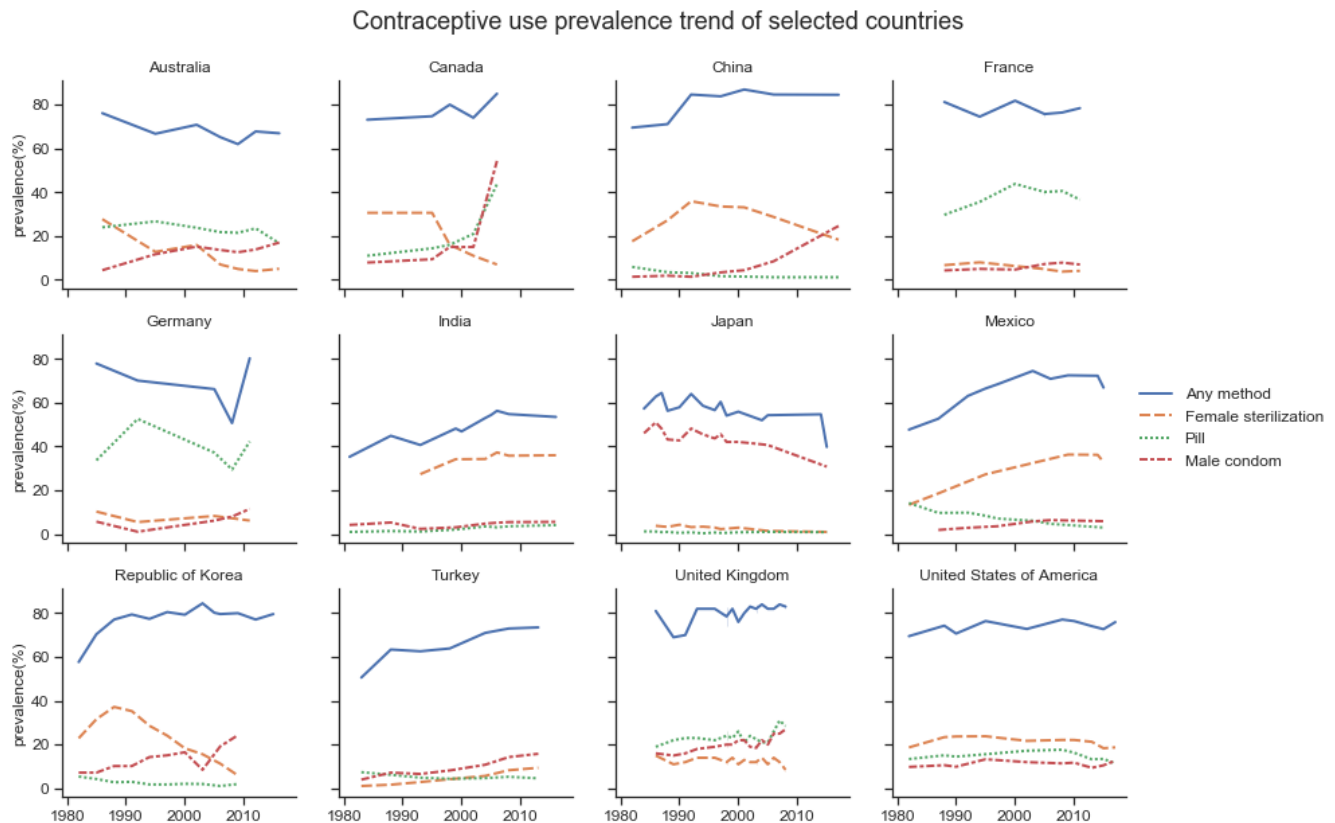
	Any method	Female sterilization	Male condom	Pill
area				
Australia	67.91	11.07	12.63	22.58
Canada	77.36	19.04	20.32	21.22
China	80.71	27.76	6.51	2.77
France	77.67	5.49	6.19	38.07
Germany	69.04	7.57	6.49	39.06
India	48.80	34.17	4.51	2.50
Japan	56.61	2.90	43.63	0.83
Mexico	65.55	27.45	4.42	7.88
Republic of Korea	77.14	23.08	13.27	2.63
Turkey	65.43	4.76	9.56	5.36
United Kingdom	80.21	12.21	20.53	23.79
United States of America	73.99	21.39	11.14	14.79

```
## Only examine the trend since 1980
subcountry = subcountry.query('year>1979')
long_subc = pd.melt(subcountry, id_vars=['area', 'iso', 'year'],\
                    value_vars=['Any method', 'Female sterilization', 'Pill', 'Male condom'])
```

```
sns.set(style="ticks",rc={"lines.linewidth": 2,'xtick.labelsize': 12, 'ytick.labelsize': 12,\
                        'font.size': 15, 'axes.labelsize': 12, 'legend.fontsize': 12})

g=sns.relplot(x="year", y="value",
              hue="variable",
              kind="line",col="area",style="variable", col_wrap=4, height=3,data=long_subc)
(g.set_ylabels("prevalence(%)")
 .set_xlabels(""))
g._legend.texts[0].set_text("") ## remove legend title
g.fig.subplots_adjust(top=0.9)
g.fig.suptitle('Contraceptive use prevalence trend of selected countries', fontsize=18)
# change column titles for each plot.
name=list(set(subcountry['area']))
titles=sorted(name)
for ax, title in zip(g.axes.flat, titles):
    ax.set_title(title)
```

/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result. return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval



png