COVID-19 Data Exploration

February 26, 2021

1 COVID-19 Data Exploration

This notebook will load in a dataset of COVID-19 Community Mobility report data from Google and explore it using Python and popular python data science packages.

1.0.1 Imports

First let's load in some packages.

[1]:
```
from __future__ import print_function, division
import pandas as pd
from pandas.plotting import scatter_matrix
import numpy as np
import scipy as sp
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
from IPython.core.display import HTML
```

[2]:
```
%matplotlib inline
```

1.0.2 Data ingest

Next let's load in the dataset into a panda's dataframe.

[3]:
```
df = pd.read_csv('../datasources/Global_Mobility_Report.csv', header=0,␣
↪na_values = [''])
countries = pd.read_csv('../datasources/country-and-continent-codes-list-csv.
↪csv')
```

```
/opt/conda/lib/python3.8/site-packages/IPython/core/interactiveshell.py:3155:
DtypeWarning: Columns (4,5) have mixed types.Specify dtype option on import or
set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

[4]:
```
datetimes = ['date']
df = df.set_index('country_region_code').join(countries.
↪set_index('Two_Letter_Country_Code'))
```

```
df[datetimes] = df[datetimes].apply(pd.to_datetime)
```

[5]:
```
sdf = df.sample(n = 50000)
```

1.0.3 Let's explore!

[6]:
```
df
```

[6]:
```
           country_region         sub_region_1 sub_region_2 metro_area  \
AE   United Arab Emirates                  NaN          NaN        NaN
AE   United Arab Emirates                  NaN          NaN        NaN
AE   United Arab Emirates                  NaN          NaN        NaN
AE   United Arab Emirates                  NaN          NaN        NaN
AE   United Arab Emirates                  NaN          NaN        NaN
..                    …                    …            …        …
NaN              Namibia  Otjozondjupa Region          NaN        NaN
NaN              Namibia  Otjozondjupa Region          NaN        NaN
NaN              Namibia       Zambezi Region          NaN        NaN
NaN              Namibia       Zambezi Region          NaN        NaN
NaN              Namibia       Zambezi Region          NaN        NaN

    iso_3166_2_code  census_fips_code                    place_id       date  \
AE              NaN               NaN  ChIJvRKrsd9IXj4RpwoIwFYv0zM 2020-02-15
AE              NaN               NaN  ChIJvRKrsd9IXj4RpwoIwFYv0zM 2020-02-16
AE              NaN               NaN  ChIJvRKrsd9IXj4RpwoIwFYv0zM 2020-02-17
AE              NaN               NaN  ChIJvRKrsd9IXj4RpwoIwFYv0zM 2020-02-18
AE              NaN               NaN  ChIJvRKrsd9IXj4RpwoIwFYv0zM 2020-02-19
..              …                 …                          …        …
NaN           NA-OD               NaN  ChIJ_7SqpVK99hsRJoAG7f5HMeY 2021-02-19
NaN           NA-OD               NaN  ChIJ_7SqpVK99hsRJoAG7f5HMeY 2021-02-20
NaN           NA-CA               NaN  ChIJ6ZWCJLviWRkR7vczZyMxxu4 2020-04-13
NaN           NA-CA               NaN  ChIJ6ZWCJLviWRkR7vczZyMxxu4 2020-12-25
NaN           NA-CA               NaN  ChIJ6ZWCJLviWRkR7vczZyMxxu4 2021-01-01

     retail_and_recreation_percent_change_from_baseline  \
AE                                                  0.0
AE                                                  1.0
AE                                                 -1.0
AE                                                 -2.0
AE                                                 -2.0
..                                                  …
NaN                                                 NaN
NaN                                                 NaN
NaN                                                 NaN
NaN                                                 NaN
NaN                                                 NaN
```

```
     grocery_and_pharmacy_percent_change_from_baseline  \
AE                                                 4.0
AE                                                 4.0
AE                                                 1.0
AE                                                 1.0
AE                                                 0.0
..                                                   …
NaN                                                NaN
NaN                                                NaN
NaN                                                NaN
NaN                                                NaN
NaN                                                NaN

     parks_percent_change_from_baseline  \
AE                                 5.0
AE                                 4.0
AE                                 5.0
AE                                 5.0
AE                                 4.0
..                                   …
NaN                                NaN
NaN                                NaN
NaN                                NaN
NaN                                NaN
NaN                                NaN

     transit_stations_percent_change_from_baseline  \
AE                                             0.0
AE                                             1.0
AE                                             1.0
AE                                             0.0
AE                                            -1.0
..                                             …
NaN                                            NaN
NaN                                            NaN
NaN                                            NaN
NaN                                            NaN
NaN                                            NaN

     workplaces_percent_change_from_baseline  \
AE                                       2.0
AE                                       2.0
AE                                       2.0
AE                                       2.0
AE                                       2.0
..                                       …
NaN                                    -14.0
```

```
NaN                                    -17.0
NaN                                    -59.0
NaN                                    -75.0
NaN                                    -68.0

     residential_percent_change_from_baseline Continent_Name Continent_Code  \
AE                                        1.0           Asia             AS
AE                                        1.0           Asia             AS
AE                                        1.0           Asia             AS
AE                                        1.0           Asia             AS
AE                                        1.0           Asia             AS
..                                        …             …              …
NaN                                       NaN         Africa             AF
NaN                                       NaN         Africa             AF
NaN                                       NaN         Africa             AF
NaN                                       NaN         Africa             AF
NaN                                       NaN         Africa             AF

             Country_Name Three_Letter_Country_Code  Country_Number
AE   United Arab Emirates                       ARE           784.0
AE   United Arab Emirates                       ARE           784.0
AE   United Arab Emirates                       ARE           784.0
AE   United Arab Emirates                       ARE           784.0
AE   United Arab Emirates                       ARE           784.0
..                    …                        …             …
NaN  Namibia, Republic of                       NAM           516.0
NaN  Namibia, Republic of                       NAM           516.0
NaN  Namibia, Republic of                       NAM           516.0
NaN  Namibia, Republic of                       NAM           516.0
NaN  Namibia, Republic of                       NAM           516.0

[4567966 rows x 19 columns]
```

[7]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 4567966 entries, AE to nan
Data columns (total 19 columns):
 #   Column                                  Dtype
---  ------                                  -----
 0   country_region                          object
 1   sub_region_1                            object
 2   sub_region_2                            object
 3   metro_area                              object
 4   iso_3166_2_code                         object
 5   census_fips_code                        float64
 6   place_id                                object
```

```
 7   date                                                  datetime64[ns]
 8   retail_and_recreation_percent_change_from_baseline    float64
 9   grocery_and_pharmacy_percent_change_from_baseline     float64
 10  parks_percent_change_from_baseline                    float64
 11  transit_stations_percent_change_from_baseline         float64
 12  workplaces_percent_change_from_baseline               float64
 13  residential_percent_change_from_baseline              float64
 14  Continent_Name                                        object
 15  Continent_Code                                        object
 16  Country_Name                                          object
 17  Three_Letter_Country_Code                             object
 18  Country_Number                                        float64
dtypes: datetime64[ns](1), float64(8), object(10)
memory usage: 697.0+ MB
```

[8]:
```
df.describe()
```

[8]:
```
       census_fips_code  retail_and_recreation_percent_change_from_baseline  \
count     928217.000000                                       2.868086e+06
mean       30357.407207                                      -2.389985e+01
std        15299.174033                                       2.769400e+01
min         1001.000000                                      -1.000000e+02
25%        18105.000000                                      -4.100000e+01
50%        29115.000000                                      -2.000000e+01
75%        45051.000000                                      -5.000000e+00
max        56045.000000                                       5.450000e+02

       grocery_and_pharmacy_percent_change_from_baseline  \
count                                       2.774222e+06
mean                                       -2.756646e+00
std                                         2.500924e+01
min                                        -1.000000e+02
25%                                        -1.400000e+01
50%                                        -2.000000e+00
75%                                         9.000000e+00
max                                         6.150000e+02

       parks_percent_change_from_baseline  \
count                        2.218548e+06
mean                        -9.899935e+00
std                          5.352491e+01
min                         -1.000000e+02
25%                         -4.300000e+01
50%                         -1.700000e+01
75%                          1.100000e+01
max                          1.206000e+03
```

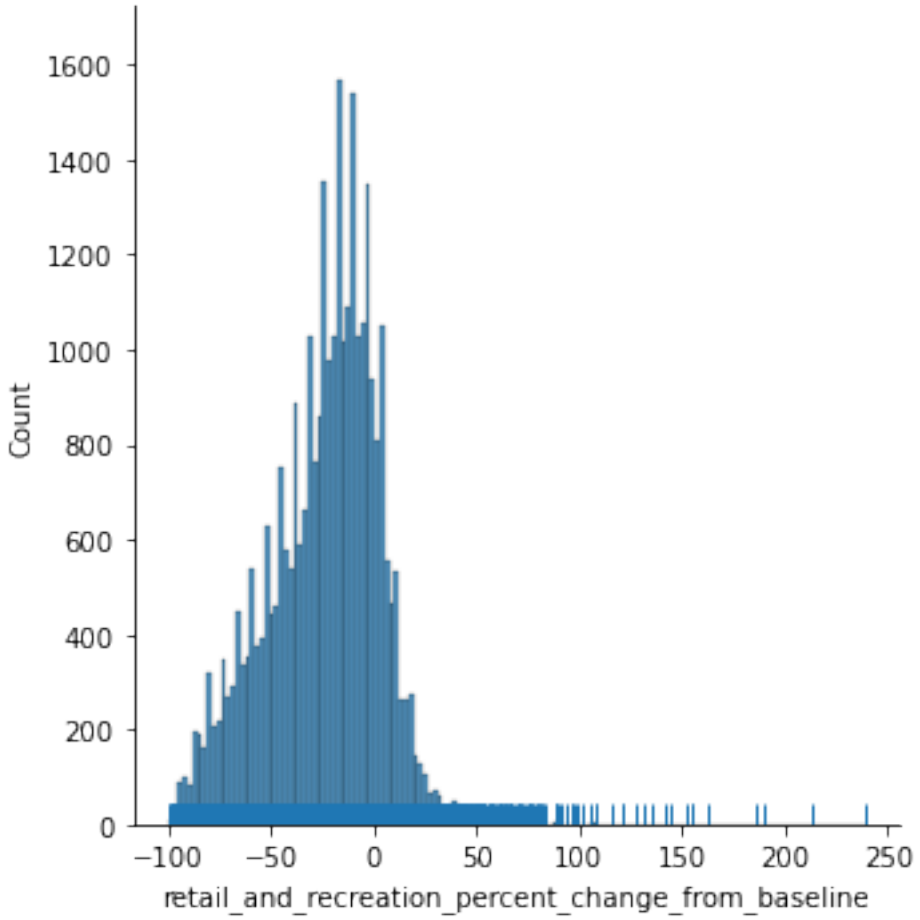
```
       transit_stations_percent_change_from_baseline  \
count                                   2.315340e+06
mean                                   -2.736464e+01
std                                     3.013115e+01
min                                    -1.000000e+02
25%                                    -4.800000e+01
50%                                    -2.800000e+01
75%                                    -8.000000e+00
max                                     5.540000e+02

       workplaces_percent_change_from_baseline  \
count                              4.358439e+06
mean                              -2.003840e+01
std                                2.014508e+01
min                               -1.000000e+02
25%                               -3.200000e+01
50%                               -1.900000e+01
75%                               -6.000000e+00
max                                2.600000e+02

       residential_percent_change_from_baseline  Country_Number
count                              2.664235e+06    4.567966e+06
mean                               9.235258e+00    4.858650e+02
std                                7.832266e+00    3.122592e+02
min                               -4.600000e+01    4.000000e+00
25%                                4.000000e+00    1.240000e+02
50%                                8.000000e+00    5.660000e+02
75%                                1.300000e+01    8.180000e+02
max                                6.500000e+01    8.940000e+02
```

Single column exploration

[9]:
```
single_df = sdf['retail_and_recreation_percent_change_from_baseline']
```

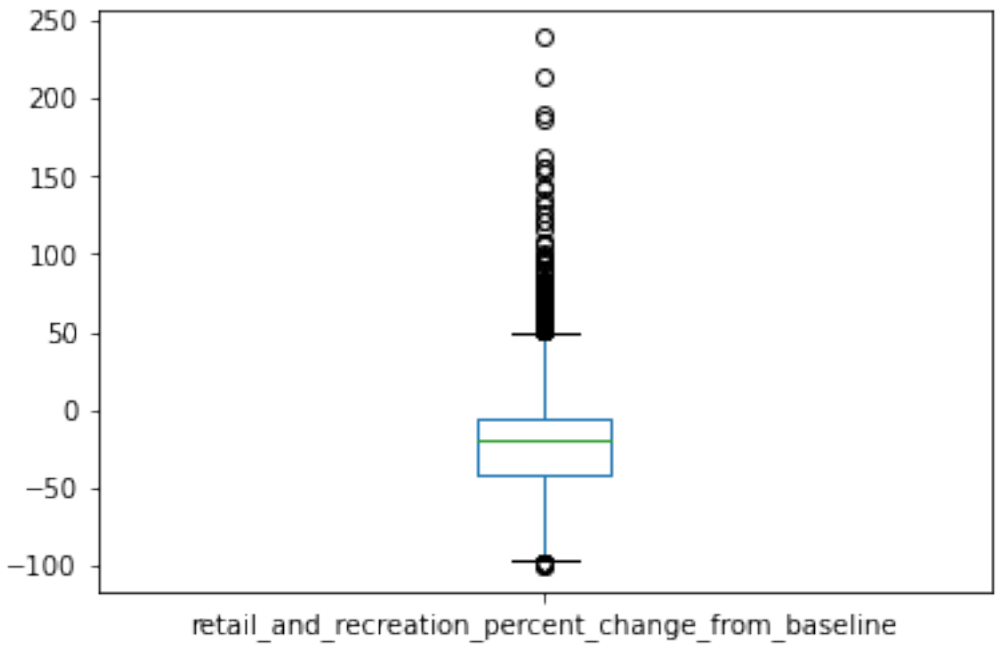
[10]:
```
sns.displot(single_df, rug=True)
```

[10]:
```
<seaborn.axisgrid.FacetGrid at 0x7f43cc592f10>
```



[11]:
```
single_df.plot.box()
```

[11]: <AxesSubplot:>



1.0.4 Multi column data exploration

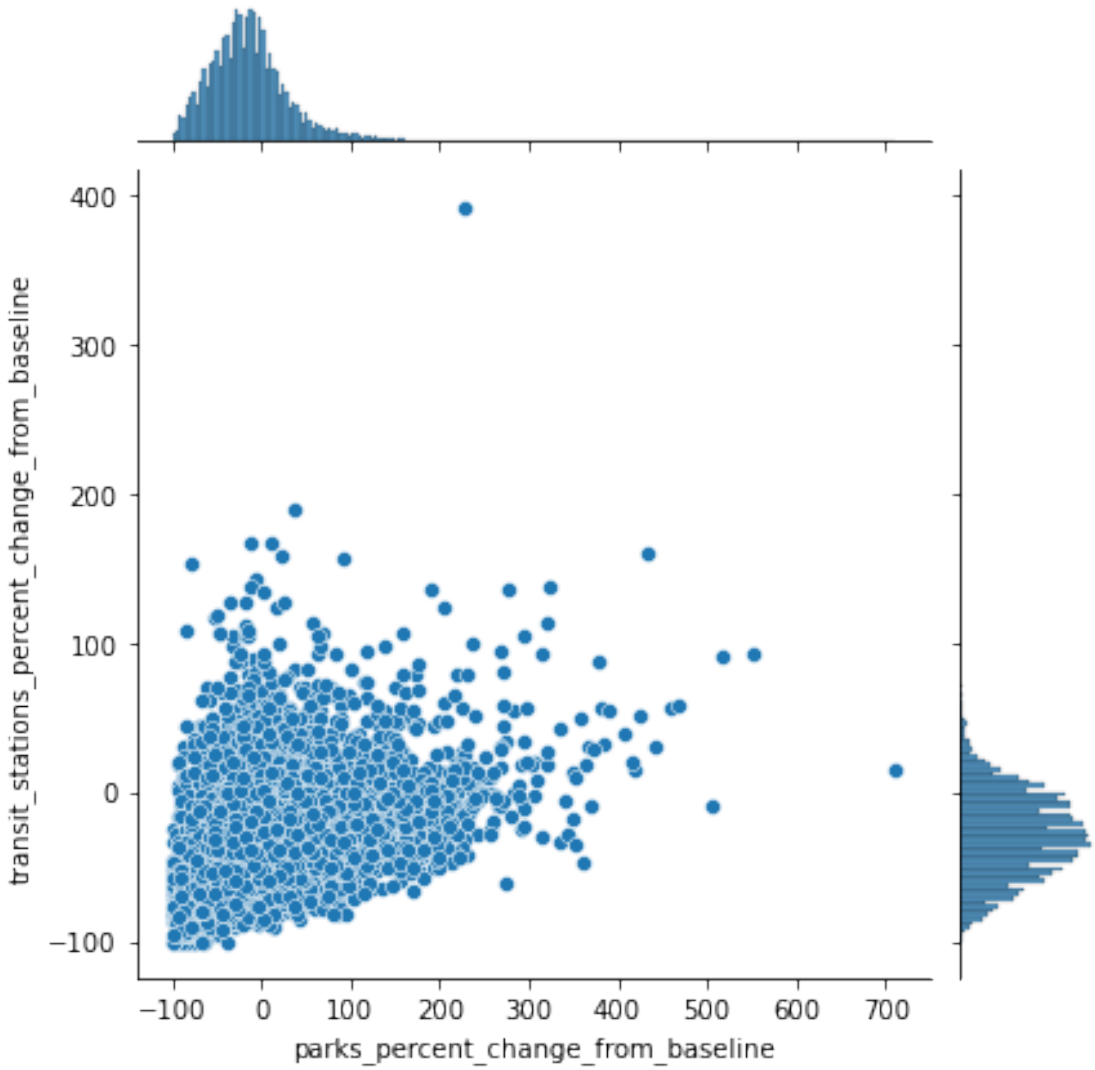
[12]:
```
multi_1_df = sdf['parks_percent_change_from_baseline']
multi_2_df = sdf['transit_stations_percent_change_from_baseline']
multi_df = sdf[['parks_percent_change_from_baseline',␣
 ↪'transit_stations_percent_change_from_baseline']].copy()
```

[13]:
```
multi_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 50000 entries, TR to BR
Data columns (total 2 columns):
 #   Column                                        Non-Null Count  Dtype
---  ------                                        --------------  -----
 0   parks_percent_change_from_baseline            24232 non-null  float64
 1   transit_stations_percent_change_from_baseline  25245 non-null  float64
dtypes: float64(2)
memory usage: 1.1+ MB
```

[14]:
```
sns.jointplot(x=multi_1_df, y=multi_2_df, data=multi_df)
```

[14]:
```
<seaborn.axisgrid.JointGrid at 0x7f436d7dbc40>
```



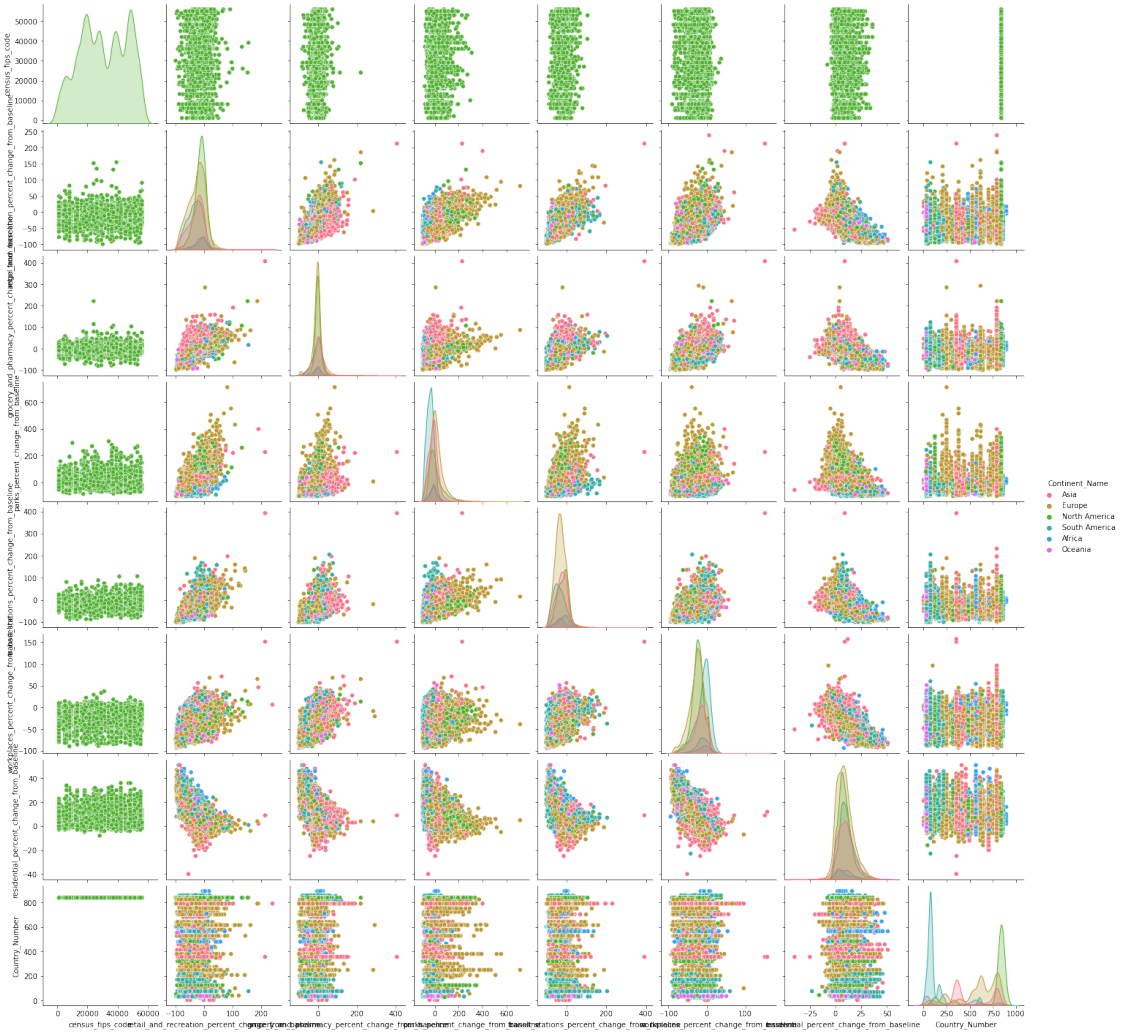
1.0.5 And now for the entire dataset...

[15]:
```
sns.pairplot(sdf, hue='Continent_Name', palette="husl")
```

```
/opt/conda/lib/python3.8/site-packages/seaborn/distributions.py:306:
UserWarning: Dataset has 0 variance; skipping density estimate.
  warnings.warn(msg, UserWarning)
/opt/conda/lib/python3.8/site-packages/seaborn/distributions.py:306:
UserWarning: Dataset has 0 variance; skipping density estimate.
  warnings.warn(msg, UserWarning)
/opt/conda/lib/python3.8/site-packages/seaborn/distributions.py:306:
UserWarning: Dataset has 0 variance; skipping density estimate.
  warnings.warn(msg, UserWarning)
/opt/conda/lib/python3.8/site-packages/seaborn/distributions.py:306:
```

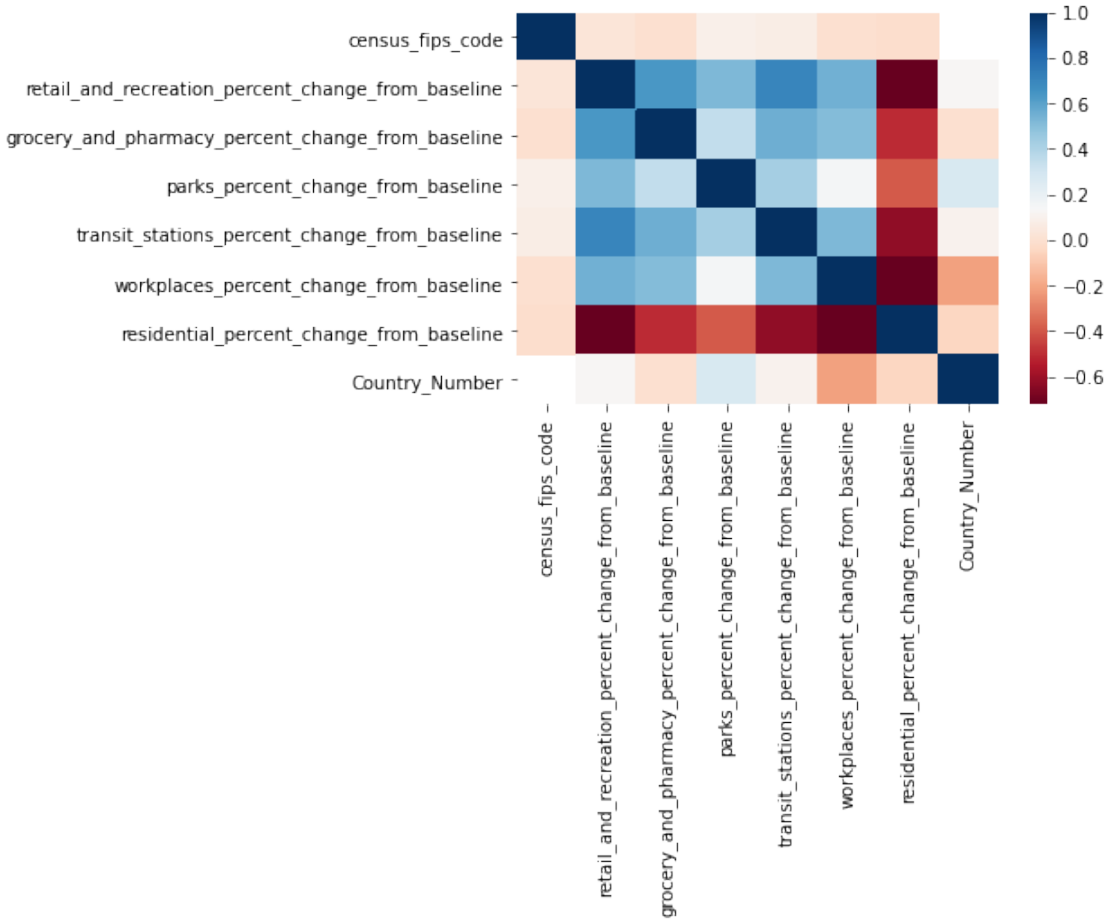
```
UserWarning: Dataset has 0 variance; skipping density estimate.
  warnings.warn(msg, UserWarning)
/opt/conda/lib/python3.8/site-packages/seaborn/distributions.py:306:
UserWarning: Dataset has 0 variance; skipping density estimate.
  warnings.warn(msg, UserWarning)
```

[15]: <seaborn.axisgrid.PairGrid at 0x7f436d28bd00>



[16]:
```
corr = sdf.corr()
sns.heatmap(corr, xticklabels=True, yticklabels=True, cmap='RdBu')
```

[16]: <AxesSubplot:>


census_fips_code
retail_and_recreation_percent_change_from_baseline
grocery_and_pharmacy_percent_change_from_baseline
parks_percent_change_from_baseline
transit_stations_percent_change_from_baseline
workplaces_percent_change_from_baseline
residential_percent_change_from_baseline
Country_Number
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6