itproject-2

April 18, 2024

[2]:
```
import pandas as pd
import numpy as np
import importlib
import geopandas
import shapely
```

[3]:
```
import plotly
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
plotly.__version__
```

[3]:
```
'5.20.0'
```

[4]:
```
#use plotly offline notebook mode
init_notebook_mode(connected=True)
```

[5]:
```
#set parameters for notebook display
#pd.set_option('display.height', 1000)
pd.set_option('display.max_rows', 4000)
pd.set_option('display.max_columns', 500)
#pd.set_option('display.width', 2000)
```

[64]:
```
HDLo = pd.read_csv('C:/Users/shett/Downloads/data/Home_Depot_Lowes_Data.csv')
region = pd.read_csv('C:/Users/shett/Downloads/data/state_region.csv')
Property_tax = pd.read_csv('C:/Users/shett/Downloads/data/Property_Tax.csv')
highways = pd.read_csv('C:/Users/shett/Downloads/data/highways.csv')
```

[65]:
```
#merge state codes
merge = pd.merge(HDLo,region, left_on='state', right_on='State Code',␣
 ↪how='inner')
HDLo_NE = merge[merge.Division == 'New England']
#HDLo_NE with Highways
HDLo_NE = pd.merge(HDLo_NE, highways, how='left', left_on=['areaname','state'],␣
 ↪right_on = ['County','state'])
#merge of property tax
HDLo_NE = pd.merge(HDLo_NE, Property_tax, how='left',␣
 ↪left_on=['areaname','state'], right_on = ['County','state'])
```

```
[66]: HDLo_NE.columns
```

```
[66]: Index(['areaname', 'county', 'state', 'r1', 'r2', 'Lcount', 'HDcount',
             'pop_2000', 'pop_2010', 'income_2000', 'income_2010', 'pct_U18_2000',
             'pct_U18_2010', 'pctcollege_2000', 'pctcollege_2010', 'ownhome_2000',
             'ownhome_2010', 'density_2000', 'density_2010', 'pctwhite_2000',
             'pctwhite_2010', 'pctblack_2000', 'pctblack_2010', 'State',
             'State Code', 'Region', 'Division', 'County_x', 'Interstate Highways',
             'U.S Highways', 'Toll Roads', 'County_y', 'Median Home Value',
             'Median Annual Property Tax Payment',
             'Average Effective Property Tax Rate'],
            dtype='object')
```

```
[67]: HDLo_NE['highway_count'] = HDLo_NE['U.S Highways'] + HDLo_NE['Toll Roads'] +␣
        ↪HDLo_NE['Interstate Highways'];
      HDLo_NE[:10]
```

```
[67]:        areaname  county  state  r1  r2  Lcount  HDcount  pop_2000  pop_2010  \
      0     Fairfield    9001     CT   1   1       1        6  882567.0  916829.0
      1      Hartford    9003     CT   1   1       5        9  857183.0  894014.0
      2    Litchfield    9005     CT   1   1       1        2  182193.0  189927.0
      3     Middlesex    9007     CT   1   1       1        1  155071.0  165676.0
      4     New Haven    9009     CT   1   1       5        7  824008.0  862477.0
      5    New London    9011     CT   1   1       2        3  259088.0  274055.0
      6       Tolland    9013     CT   1   1       0        0  136364.0  152691.0
      7       Windham    9015     CT   1   1       1        1  109091.0  118428.0
      8  Androscoggin   23001     ME   1   1       1        1  103793.0  107702.0
      9     Aroostook   23003     ME   1   1       1        0   73938.0   71870.0

         income_2000  income_2010  pct_U18_2000  pct_U18_2010  pctcollege_2000  \
      0        77690       100179          25.6          24.8             39.9
      1        62144        78826          24.6          22.8             29.6
      2        66445        84422          24.6          21.6             27.5
      3        71319        90666          23.2          21.2             33.8
      4        60549        77451          24.5          22.4             27.6
      5        59857        79236          24.4          21.7             26.2
      6        70856        91048          23.1          20.2             32.8
      7        52490        67520          25.1          22.3             19.0
      8        44082        54081          23.9          22.6             14.4
      9        36044        45592          22.6          20.0             14.6

         pctcollege_2010  ownhome_2000  ownhome_2010  density_2000  density_2010  \
      0             43.4          69.2          68.6        1409.9        1467.2
      1             32.8          64.2          65.5        1166.2        1216.2
      2             32.4          75.2          76.3         198.0         206.3
      3             36.8          72.1          74.4         420.2         448.6
      4             32.0          63.1          63.4        1359.7        1426.7
```

```
5             30.7          66.7          67.7         389.0         412.2
6             36.0          73.5          75.3         332.6         372.2
7             21.6          67.4          69.3         212.7         230.9
8             17.7          63.4          64.4         220.8         230.2
9             15.7          73.0          71.2          11.1          10.8

   pctwhite_2000  pctwhite_2010  pctblack_2000  pctblack_2010        State  \
0           79.3           74.8           11.3           10.8  Connecticut   
1           76.9           72.4           13.9           13.3  Connecticut   
2           95.8           93.9            1.4            1.3  Connecticut   
3           91.3           89.2            5.0            4.7  Connecticut   
4           79.4           74.8           13.3           12.7  Connecticut   
5           87.0           82.2            6.2            5.8  Connecticut   
6           92.3           89.8            3.7            3.3  Connecticut   
7           91.3           89.6            2.4            2.2  Connecticut   
8           97.0           92.8            3.8            3.6        Maine   
9           96.8           95.7            0.6            0.6        Maine   

  State Code     Region     Division      County_x  Interstate Highways  \
0         CT  Northeast  New England     Fairfield                    1   
1         CT  Northeast  New England      Hartford                    1   
2         CT  Northeast  New England    Litchfield                    0   
3         CT  Northeast  New England     Middlesex                    1   
4         CT  Northeast  New England     New Haven                    2   
5         CT  Northeast  New England    New London                    1   
6         CT  Northeast  New England       Tolland                    0   
7         CT  Northeast  New England       Windham                    2   
8         ME  Northeast  New England  Androscoggin                    1   
9         ME  Northeast  New England     Aroostook                    2   

   U.S Highways  Toll Roads      County_y  Median Home Value  \
0             2           0     Fairfield             413400   
1             5           0      Hartford             234900   
2             4           0    Litchfield             249500   
3             2           0     Middlesex             283800   
4             2           0     New Haven             244000   
5             2           0    New London             241500   
6             1           0       Tolland             247800   
7             2           0       Windham             196900   
8             2           0  Androscoggin             152100   
9             2           1     Aroostook              95800   

   Median Annual Property Tax Payment  Average Effective Property Tax Rate  \
0                                7057                               0.0171   
1                                5035                               0.0214   
2                                4639                               0.0186   
3                                5298                               0.0187   
```

```
4                                5486                                0.0225
5                                4227                                0.0175
6                                5133                                0.0207
7                                3557                                0.0181
8                                2460                                0.0162
9                                1364                                0.0142

   highway_count
0              3
1              6
2              4
3              3
4              4
5              3
6              1
7              4
8              3
9              5
```

```
[68]: import plotly.graph_objects as go
      import plotly.express as px
```

```
[69]: fips = HDLo_NE['county'].tolist()
      fips = ['0500000US' + str(value).zfill(5) if len(str(value)) < 5 else␣
       ↪'0500000US' + str(value) for value in fips]
      values = HDLo_NE['Lcount'].tolist()
      print(fips[:10])
      print(values[:10])
```

```
['0500000US09001', '0500000US09003', '0500000US09005', '0500000US09007',
'0500000US09009', '0500000US09011', '0500000US09013', '0500000US09015',
'0500000US23001', '0500000US23003']
[1, 5, 1, 1, 5, 2, 0, 1, 1, 1]
```

```
[70]: import geopandas as gpd

      geo_data = gpd.read_file('C:/Users/shett/Downloads/geojson-counties-fips.json')

      geo_data['FIPS'] = geo_data['GEO_ID']
      geo_data2 = geo_data[['FIPS', "geometry"]]
      geo_data2.to_file("main_map.geojson",driver='GeoJSON')
      print(geo_data2[:3])
```

```
             FIPS                                           geometry
0  0500000US01001  POLYGON ((-86.49677 32.34444, -86.71790 32.402…
1  0500000US01009  POLYGON ((-86.57780 33.76532, -86.75914 33.840…
2  0500000US01017  POLYGON ((-85.18413 32.87053, -85.12342 32.772…
```

[71]:

```
data = { 'FIPS': fips, 'Value': values }
df = pd.DataFrame(data)
print(df[:3])
```

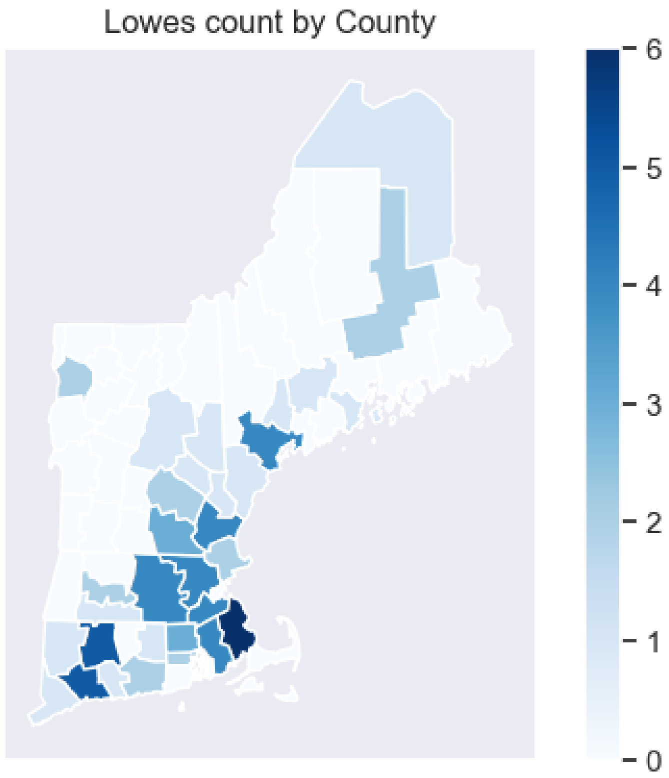
```
               FIPS  Value
0  0500000US09001      1
1  0500000US09003      5
2  0500000US09005      1
```

[72]:

```
ax = pd.merge(geo_data2, df).plot(column="Value", cmap='Blues', legend=True)
ax.set_title('Lowes count by County')
ax.set_xticks([])
ax.set_yticks([])
```

[72]:

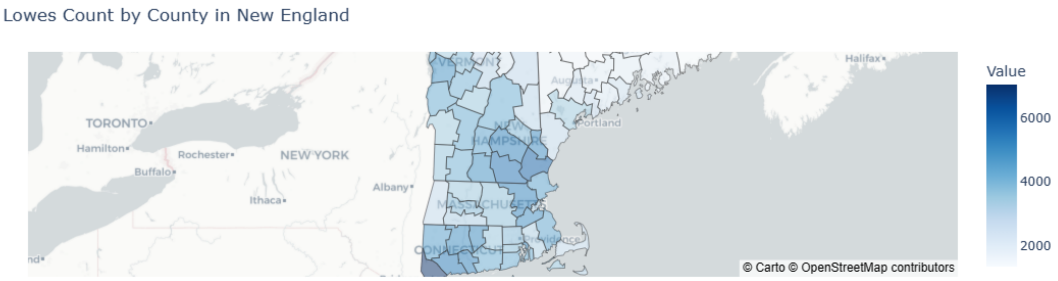
```
[]
```



[47]:

```
import json

with open("main_map.geojson", 'r') as infile:
    pacitiesjson = json.load(infile)
```

```
colorscale =␣
 ↪["#f7fbff","#ebf3fb","#deebf7","#d2e3f3","#c6dbef","#b3d2e9","#9ecae1",
 "#85bcdb","#6baed6","#57a0ce","#4292c6","#3082be","#2171b5","#1361a9",
 "#08519c","#0b4083","#08306b"]


fig = px.choropleth_mapbox(df, geojson=pacitiesjson,
                          locations='FIPS', featureidkey='properties.FIPS',
                          color='Value',
                              color_continuous_scale=colorscale,
                              range_color=(min(values), max(values)),
                              mapbox_style="carto-positron",  # Map style
                              center={"lat": 43.0, "lon": -72.0},  # Center map␣
 ↪on New England
                              zoom=5,  # Adjust zoom level
                              opacity=0.5,  # Adjust opacity
                              labels={'color': 'Lowes Count'},  # Label for␣
 ↪colorbar
                              )

fig.update_layout(
    title_text='Lowes Count by County in New England',
    geo=dict(
        projection=go.layout.geo.Projection(type='albers usa'),
        scope='usa',
        showland=True,
        landcolor='rgb(229, 229, 229)',
        showsubunits=True,
        subunitcolor='rgb(255, 255, 255)',
        showframe=False,
        resolution=50,
        center={'lat': 43.0, 'lon': -72.0},  # Center map on New England
        lonaxis=dict(range=[-75, -68]),  # Adjust longitude range
        lataxis=dict(range=[41, 47]),  # Adjust latitude range
    ),
)

fig.show()
```

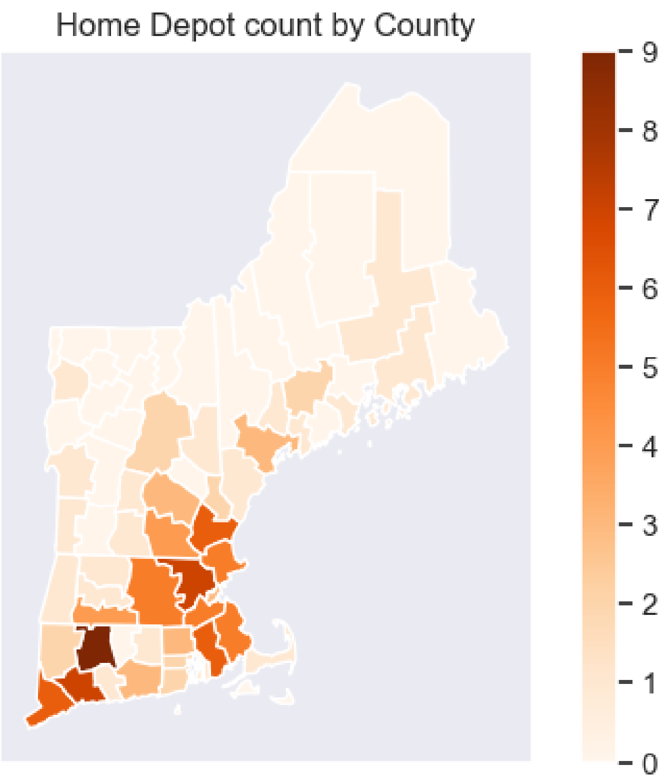


```
[73]: values = HDLo_NE['HDcount'].tolist()
      data = { 'FIPS': fips, 'Value': values }
      df = pd.DataFrame(data)
      print(df[:3])
```

```
                FIPS  Value
0   0500000US09001      6
1   0500000US09003      9
2   0500000US09005      2
```

```
[74]: ax = pd.merge(geo_data2, df).plot(column="Value", cmap='Oranges', legend=True)
      ax.set_title('Home Depot count by County')
      ax.set_xticks([])
      ax.set_yticks([])
```

```
[74]: []
```

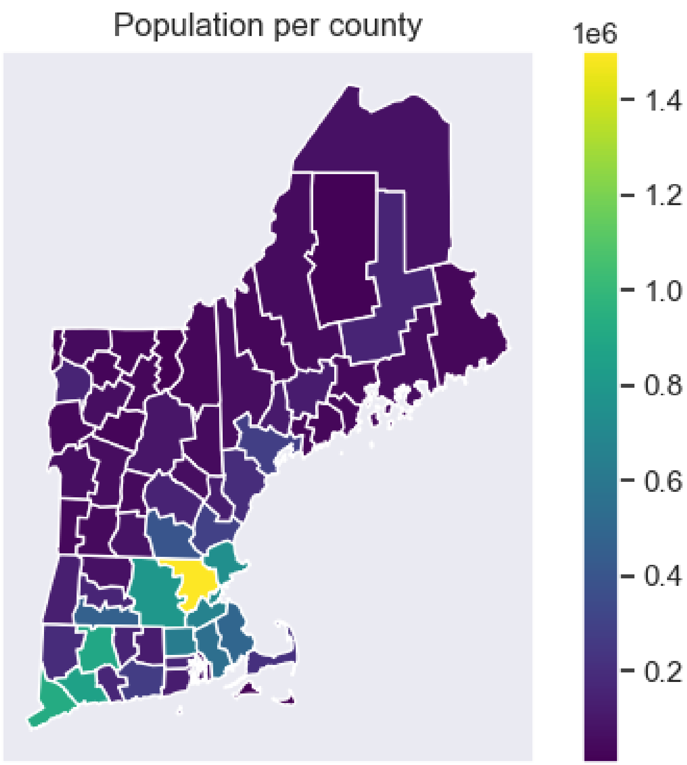


[75]:
```
values = HDLo_NE['pop_2010'].tolist()
data = { 'FIPS': fips, 'Value': values }
df = pd.DataFrame(data)
print(df[:3])
```

```
              FIPS     Value
0  0500000US09001  916829.0
1  0500000US09003  894014.0
2  0500000US09005  189927.0
```

[80]:
```
ax = pd.merge(geo_data2, df).plot(column="Value", cmap='viridis', legend=True)
ax.set_title('Population per county')
ax.set_xticks([])
ax.set_yticks([])
```

[80]: []

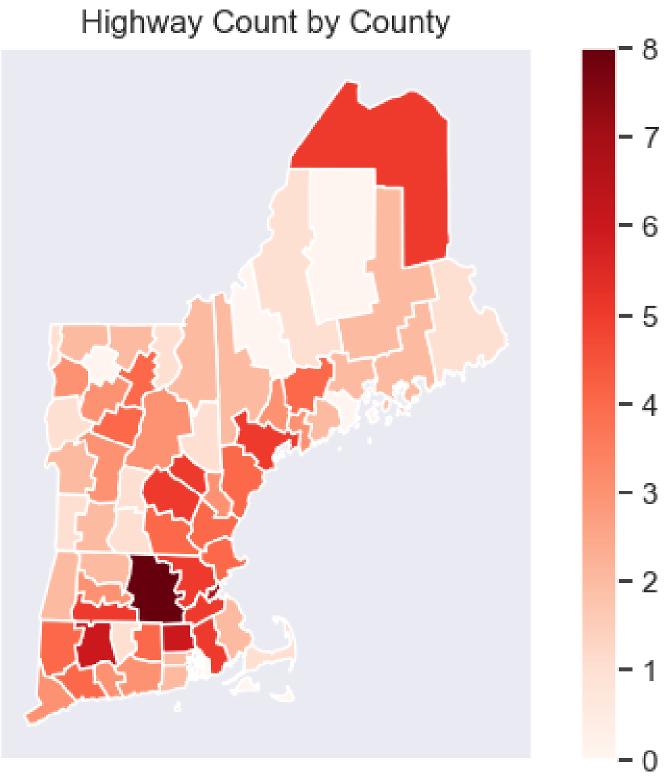


[81]:
```
values = HDLo_NE['highway_count'].tolist()
data = { 'FIPS': fips, 'Value': values }
df = pd.DataFrame(data)
print(df[:3])
```

```
             FIPS  Value
0  0500000US09001      3
1  0500000US09003      6
2  0500000US09005      4
```

[82]:
```
ax = pd.merge(geo_data2, df).plot(column="Value", cmap='Reds', legend=True)
ax.set_title('Highway Count by County')
ax.set_xticks([])
ax.set_yticks([])
```

[82]: []

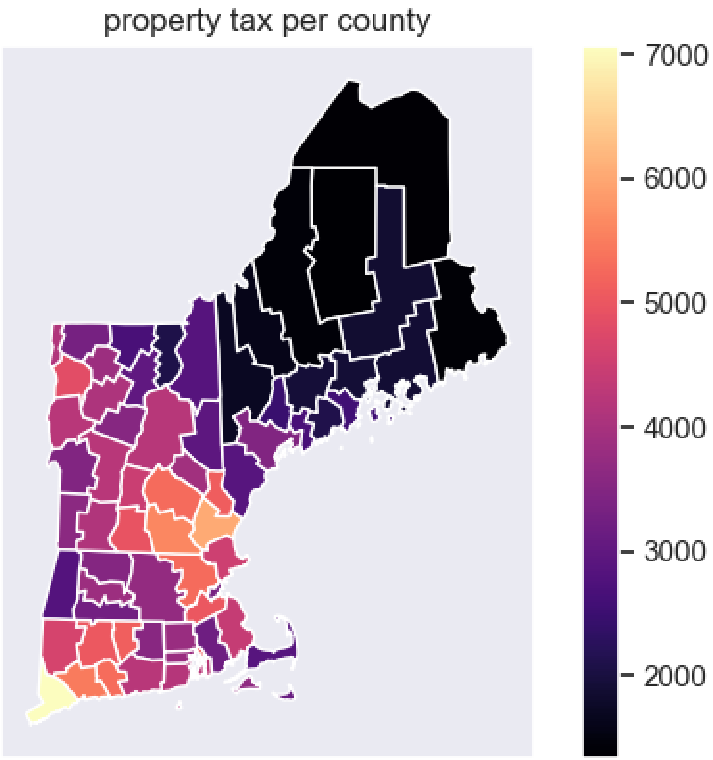


```
[83]: values = HDLo_NE['Median Annual Property Tax Payment'].tolist()
      data = { 'FIPS': fips, 'Value': values }
      df = pd.DataFrame(data)
      print(df[:3])
```

```
                FIPS  Value
0  0500000US09001   7057
1  0500000US09003   5035
2  0500000US09005   4639
```

```
[84]: ax = pd.merge(geo_data2, df).plot(column="Value", cmap='magma', legend=True)
      ax.set_title('property tax per county')
      ax.set_xticks([])
      ax.set_yticks([])
```

```
[84]: []
```



[38]:
```
%matplotlib inline
```

[39]:
```
#normalize census data
#remove collinearity
HDLo_NE['population'] = abs(HDLo_NE['pop_2010'] - HDLo_NE['pop_2000'])
HDLo_NE['income'] = abs(HDLo_NE['income_2010'] - HDLo_NE['income_2000'])
HDLo_NE['pct_U18'] = abs(HDLo_NE['pct_U18_2010'] - HDLo_NE['pct_U18_2000'])
HDLo_NE['pctcollege'] = abs(HDLo_NE['pctcollege_2010'] -␣
  ↪HDLo_NE['pctcollege_2000'])
HDLo_NE['ownhome'] = abs(HDLo_NE['ownhome_2010'] - HDLo_NE['ownhome_2000'])
HDLo_NE['density'] = abs(HDLo_NE['density_2010'] - HDLo_NE['density_2000'])
HDLo_NE['pctwhite'] = abs(HDLo_NE['pctwhite_2010'] - HDLo_NE['pctwhite_2000'])
HDLo_NE['pctblack'] = abs(HDLo_NE['pctblack_2010'] - HDLo_NE['pctblack_2000'])
#remove *_2010 and *_2000 fields
HDLo_NE = HDLo_NE[HDLo_NE.columns.drop(list(HDLo_NE.filter(regex=('(.*_2010|.
  ↪*_2000)'))))]
HDLo_NE.head()
```

[39]:
```
     areaname  county state  r1  r2  Lcount  HDcount        State State Code  \
0   Fairfield    9001    CT   1   1       1        6  Connecticut         CT
1    Hartford    9003    CT   1   1       5        9  Connecticut         CT
2  Litchfield    9005    CT   1   1       1        2  Connecticut         CT
```

```
3   Middlesex    9007    CT    1    1        1        1  Connecticut         CT
4   New Haven    9009    CT    1    1        5        7  Connecticut         CT

      Region     Division     County_x  Interstate Highways  U.S Highways  \
0  Northeast  New England    Fairfield                    1             2
1  Northeast  New England     Hartford                    1             5
2  Northeast  New England   Litchfield                    0             4
3  Northeast  New England    Middlesex                    1             2
4  Northeast  New England    New Haven                    2             2

   Toll Roads     County_y  Median Home Value  \
0           0    Fairfield             413400
1           0     Hartford             234900
2           0   Litchfield             249500
3           0    Middlesex             283800
4           0    New Haven             244000

   Median Annual Property Tax Payment  Average Effective Property Tax Rate  \
0                                7057                               0.0171
1                                5035                               0.0214
2                                4639                               0.0186
3                                5298                               0.0187
4                                5486                               0.0225

   highway_count  population  income  pct_U18  pctcollege  ownhome  density  \
0              3     34262.0   22489      0.8         3.5      0.6     57.3
1              6     36831.0   16682      1.8         3.2      1.3     50.0
2              4      7734.0   17977      3.0         4.9      1.1      8.3
3              3     10605.0   19347      2.0         3.0      2.3     28.4
4              4     38469.0   16902      2.1         4.4      0.3     67.0

   pctwhite  pctblack
0       4.5       0.5
1       4.5       0.6
2       1.9       0.1
3       2.1       0.3
4       4.6       0.6
```

[40]:
```
#add store presence
HDLo_NE['Store_present'] = np.where(HDLo_NE['Lcount'] > 0,1, np.
 ↪where(HDLo_NE['HDcount'] > 0,1,0))
HDLo_NE.head()
```

[40]:
```
     areaname  county state  r1  r2  Lcount  HDcount        State State Code  \
0   Fairfield    9001    CT   1   1       1        6  Connecticut         CT
1    Hartford    9003    CT   1   1       5        9  Connecticut         CT
2  Litchfield    9005    CT   1   1       1        2  Connecticut         CT
```

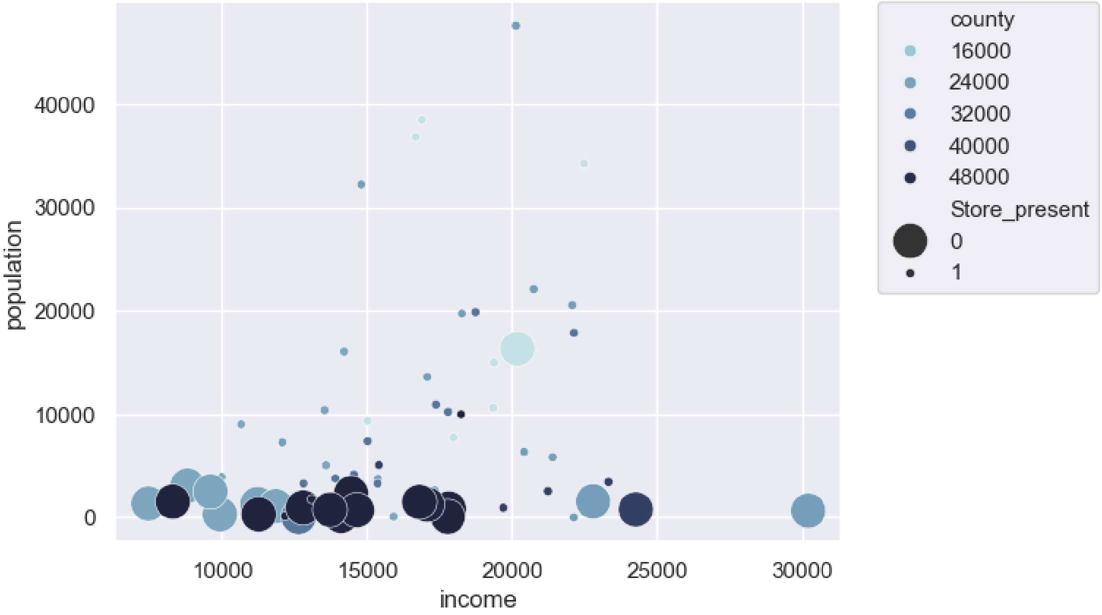
```
3   Middlesex    9007     CT    1   1       1        1   Connecticut         CT
4   New Haven    9009     CT    1   1       5        7   Connecticut         CT

      Region     Division    County_x  Interstate Highways  U.S Highways  \
0  Northeast  New England   Fairfield                    1             2
1  Northeast  New England    Hartford                    1             5
2  Northeast  New England  Litchfield                    0             4
3  Northeast  New England   Middlesex                    1             2
4  Northeast  New England   New Haven                    2             2

   Toll Roads    County_y  Median Home Value  \
0           0   Fairfield             413400
1           0    Hartford             234900
2           0  Litchfield             249500
3           0   Middlesex             283800
4           0   New Haven             244000

   Median Annual Property Tax Payment  Average Effective Property Tax Rate  \
0                                7057                               0.0171
1                                5035                               0.0214
2                                4639                               0.0186
3                                5298                               0.0187
4                                5486                               0.0225

   highway_count  population  income  pct_U18  pctcollege  ownhome  density  \
0              3     34262.0   22489      0.8         3.5      0.6     57.3
1              6     36831.0   16682      1.8         3.2      1.3     50.0
2              4      7734.0   17977      3.0         4.9      1.1      8.3
3              3     10605.0   19347      2.0         3.0      2.3     28.4
4              4     38469.0   16902      2.1         4.4      0.3     67.0

   pctwhite  pctblack  Store_present
0       4.5       0.5              1
1       4.5       0.6              1
2       1.9       0.1              1
3       2.1       0.3              1
4       4.6       0.6              1
```

[44]:

```
import seaborn as sns

sns.set()
cmap = sns.cubehelix_palette(rot=-.2, as_cmap=True)
ax = sns.scatterplot(x="income", y="population",
 hue="county", size="Store_present",
 palette=cmap, sizes=(20, 300),
 data=HDLo_NE)
# Put the legend out of the figure
```

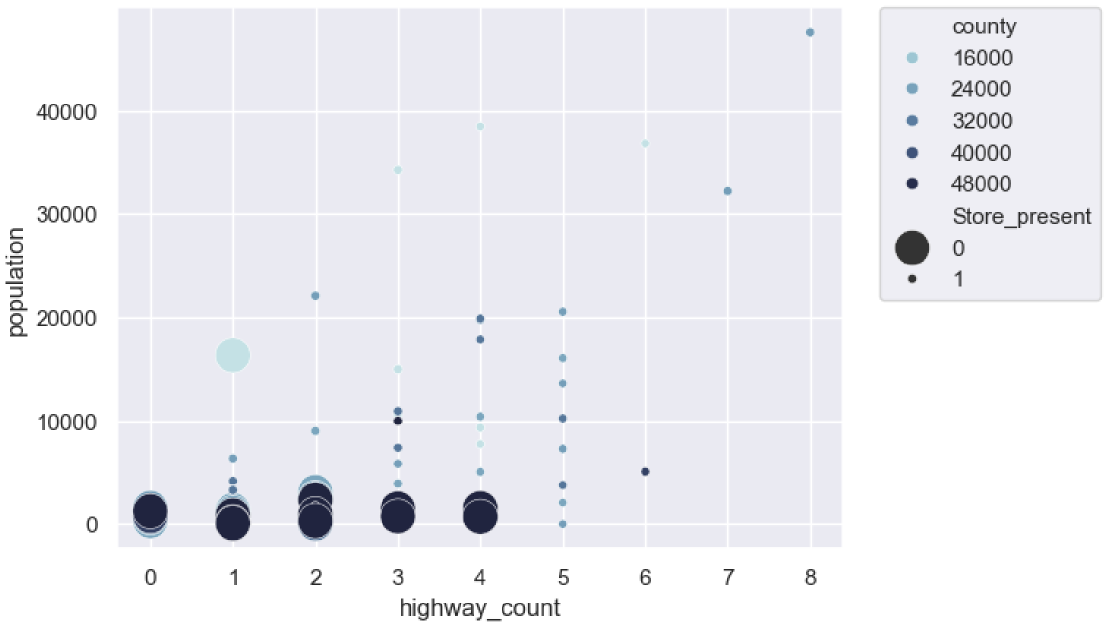
```
ax.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.)
```

[44]: <matplotlib.legend.Legend at 0x25876093b50>



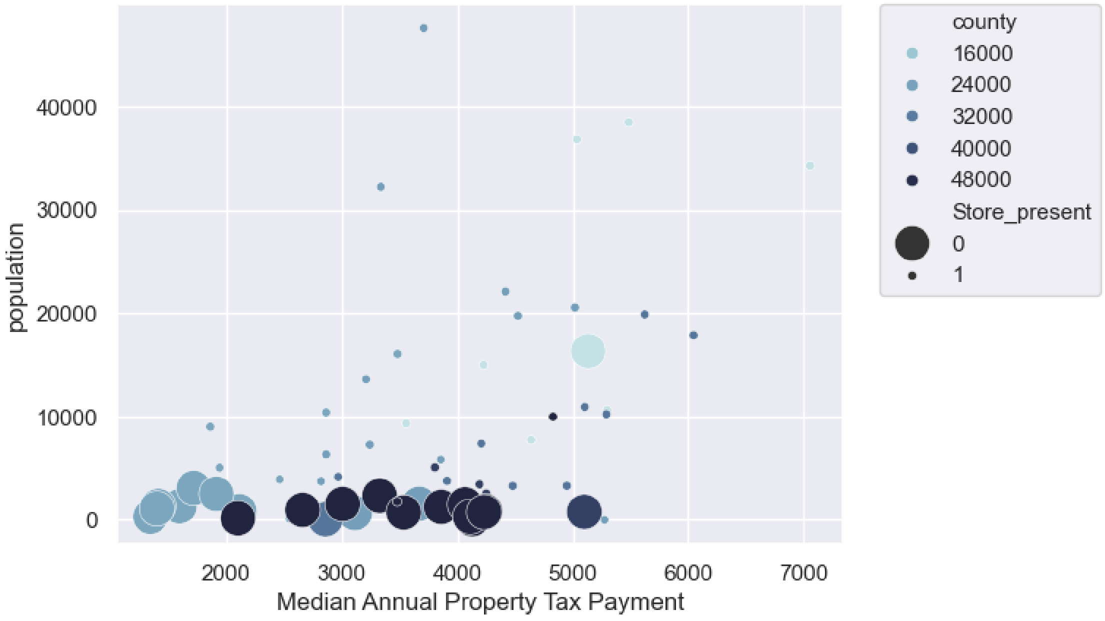
[45]:
```
ax = sns.scatterplot(x="highway_count", y="population",
 hue="county", size="Store_present",
 palette=cmap, sizes=(20, 300),
 data=HDLo_NE)
# Put the legend out of the figure
ax.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.)
```

[45]: <matplotlib.legend.Legend at 0x25876301090>



```
[46]: ax = sns.scatterplot(x="Median Annual Property Tax Payment", y="population",
       hue="county", size="Store_present",
       palette=cmap, sizes=(20, 300),
       data=HDLo_NE)
      # Put the legend out of the figure
      ax.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.)
```

```
[46]: <matplotlib.legend.Legend at 0x25876305ad0>
```



[ ]: