hw_3

November 26, 2022

1 Homework 3

1.1 In-class exercise 1

1.1.1 Step 1

Load the request per capita dataset from https://storage.googleapis.com/python-public-policy/data/311_community_districts.csv.zip as `requests_by_cd` and display it.

[1]:
```
# your code here
import pandas as pd
import plotly.express as px

# boilerplate for allowing PDF export
import plotly.io as pio

pio.renderers.default = "notebook_connected+pdf"

requests_by_cd = pd.read_csv(
    "https://storage.googleapis.com/python-public-policy/data/
 ↪311_community_districts.csv.zip"
)
requests_by_cd.head()
```

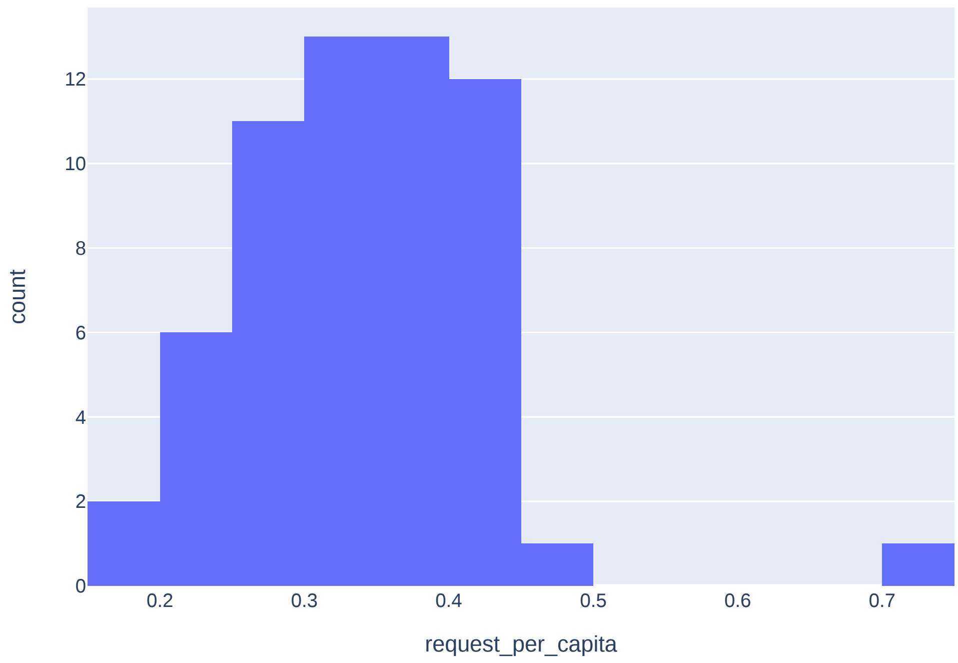
[1]:
```
   borocd    Borough                        CD Name  2010 Population  \
0     112  Manhattan    Washington Heights, Inwood           190020
1     405     Queens  Ridgewood, Glendale, Maspeth           169190
2     412     Queens   Jamaica, St. Albans, Hollis           225919
3     301   Brooklyn      Williamsburg, Greenpoint           173083
4     303   Brooklyn            Bedford Stuyvesant           152985

   count_of_311_requests  request_per_capita
0                  81403            0.428392
1                  71506            0.422637
2                  70362            0.311448
3                  68104            0.393476
4                  66360            0.433768
```

1.1.2 Step 2

Make a histogram of the requests per capita.

```
[2]: # your code here (dataframe)
     fig = px.histogram(requests_by_cd, x="request_per_capita")
     fig.show()
```

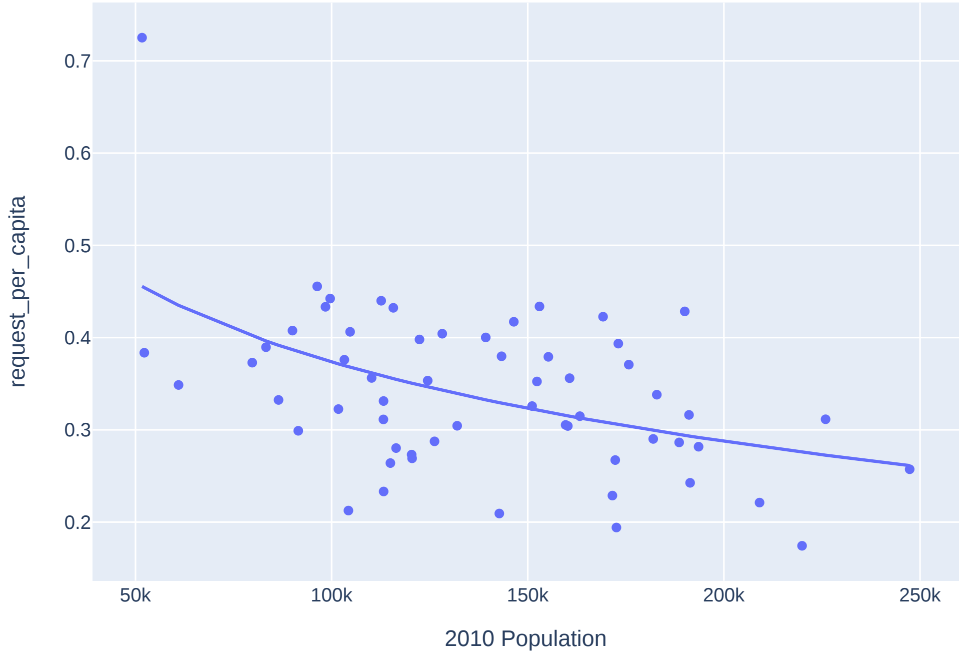


1.2 In-class exercise 2

Take the scatterplot example from the lecture and add a trendline.

```
[15]: # your code here
      fig = px.scatter(
          requests_by_cd,
          x="2010 Population",
          y="request_per_capita",
          trendline='ols',
          trendline_options=dict(log_x=True),
          hover_data=["borocd", "CD Name"],
      )

      fig.show()
```



1.3 Coding

We are going to look at the population count of different community districts over time.

[4]:

```python
import plotly.express as px

# boilerplate for allowing PDF export
import plotly.io as pio

pio.renderers.default = "notebook_connected+pdf"
```

1.3.1 Step 1

Read the data from the New York City Population By Community Districts data set into a DataFrame called `pop_by_cd`. To get the URL:

1. Visit the page linked above.
2. Click `Export`.
3. Right-click `CSV`.
4. Click `Copy Link Address` (or `Location`, depending on your browser).

[5]:

```
# your code here
import pandas as pd
pop_by_cd = pd.read_csv(
    "https://data.cityofnewyork.us/api/views/xi7c-iiu2/rows.csv?
 ↪accessType=DOWNLOAD"
)
```

1.3.2 Step 2

Prepare the data. Use the following code to reshape the DataFrame to have one row per community district per Census year.

[6]:

```
# turn the population columns into rows
populations = pd.melt(
    pop_by_cd,
    id_vars=["Borough", "CD Number", "CD Name"],
    var_name="year",
    value_name="population",
)

# turn the years into numbers
populations.year = populations.year.str.replace(" Population", "").astype(int)

populations
```

[6]:

```
           Borough  CD Number                           CD Name  year  \
0            Bronx          1   Melrose, Mott Haven, Port Morris  1970
1            Bronx          2             Hunts Point, Longwood  1970
2            Bronx          3       Morrisania, Crotona Park East  1970
3            Bronx          4       Highbridge, Concourse Village  1970
4            Bronx          5  University Hts., Fordham, Mt. Hope  1970
..             …        …                                 …   …
290         Queens         13           Queens Village, Rosedale  2010
291         Queens         14         The Rockaways, Broad Channel  2010
292  Staten Island          1           Stapleton, Port Richmond  2010
293  Staten Island          2         New Springville, South Beach  2010
294  Staten Island          3   Tottenville, Woodrow, Great Kills  2010

     population
0        138557
1         99493
2        150636
3        144207
4        121807
..          …
290      188593
291      114978
```

```
292      175756
293      132003
294      160209

[295 rows x 5 columns]
```

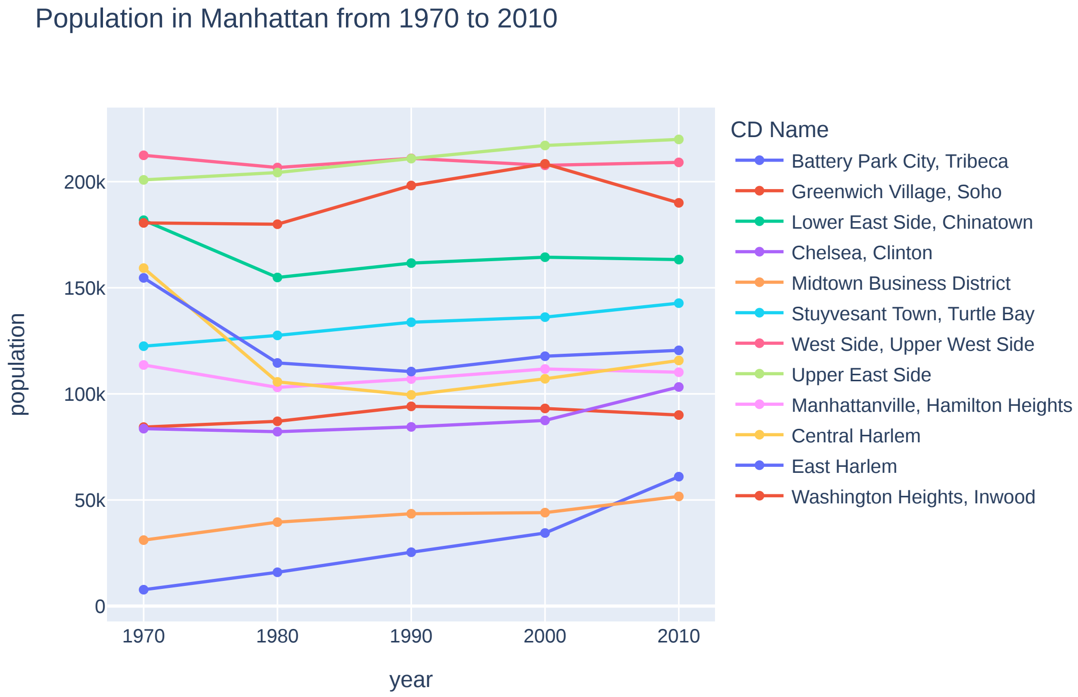
1.3.3 Step 3

Create a line chart of the population over time for each community district in Manhattan. There should be one line for each.

See the Plotly Line Plot with column encoding color examples.

[7]:

```
# your code here
import plotly.express as px
df = populations.query("Borough =='Manhattan'")
fig = px.line(df, x="year", y="population",color='CD␣
 ↪Name',markers=True,title='Population in Manhattan from 1970 to 2010')
fig.show()
```



1.3.4 Step 4

We are going to do some mapping using the `pop_by_cd` DataFrame from before. To do so, we need `borocds`. Create that column with the values filled in. (See Lecture 2.)

[8]:

```
# your code here
def recode_borocd_pop(row):
    if row.Borough == "Manhattan":
        return str(100 + row["CD Number"])
    elif row.Borough == "Bronx":
        return str(200 + row["CD Number"])
    elif row.Borough == "Brooklyn":
        return str(300 + row["CD Number"])
    elif row.Borough == "Queens":
        return str(400 + row["CD Number"])
    elif row.Borough == "Staten Island":
        return str(500 + row["CD Number"])
    else:
        return "Invalid BoroCD"
```

[9]:

```
pop_by_cd["borocd"] = pop_by_cd.apply(recode_borocd_pop, axis=1)
pop_by_cd.head()
```

[9]:

```
  Borough  CD Number                           CD Name  1970 Population  \
0   Bronx          1    Melrose, Mott Haven, Port Morris           138557
1   Bronx          2               Hunts Point, Longwood            99493
2   Bronx          3        Morrisania, Crotona Park East           150636
3   Bronx          4        Highbridge, Concourse Village           144207
4   Bronx          5  University Hts., Fordham, Mt. Hope           121807

   1980 Population  1990 Population  2000 Population  2010 Population borocd
0            78441            77214            82159            91497    201
1            34399            39443            46824            52246    202
2            53635            57162            68574            79762    203
3           114312           119962           139563           146441    204
4           107995           118435           128313           128200    205
```

1.3.5 Step 5

Let make a choropleth map showing the population change from 2000 to 2010 for each community district. Adapt the `.choropleth_mapbox()` example in Lecture 3.

[10]:

```
pop_by_cd['pop_change'] = pop_by_cd['2010 Population'] - pop_by_cd['2000␣
 ↪Population']
pop_by_cd.head()
```

[10]:

```
  Borough  CD Number                           CD Name  1970 Population  \
0   Bronx          1    Melrose, Mott Haven, Port Morris           138557
```

```
1   Bronx          2             Hunts Point, Longwood           99493
2   Bronx          3       Morrisania, Crotona Park East         150636
3   Bronx          4       Highbridge, Concourse Village         144207
4   Bronx          5  University Hts., Fordham, Mt. Hope         121807

  1980 Population  1990 Population  2000 Population  2010 Population borocd  \
0           78441            77214            82159            91497    201
1           34399            39443            46824            52246    202
2           53635            57162            68574            79762    203
3          114312           119962           139563           146441    204
4          107995           118435           128313           128200    205

  pop_change
0       9338
1       5422
2      11188
3       6878
4       -113
```

[11]:

```
import requests

response = requests.get("https://data.cityofnewyork.us/resource/jp9i-3b7y.
 ↪geojson")
shapes = response.json()
print("loaded")
```

```
loaded
```

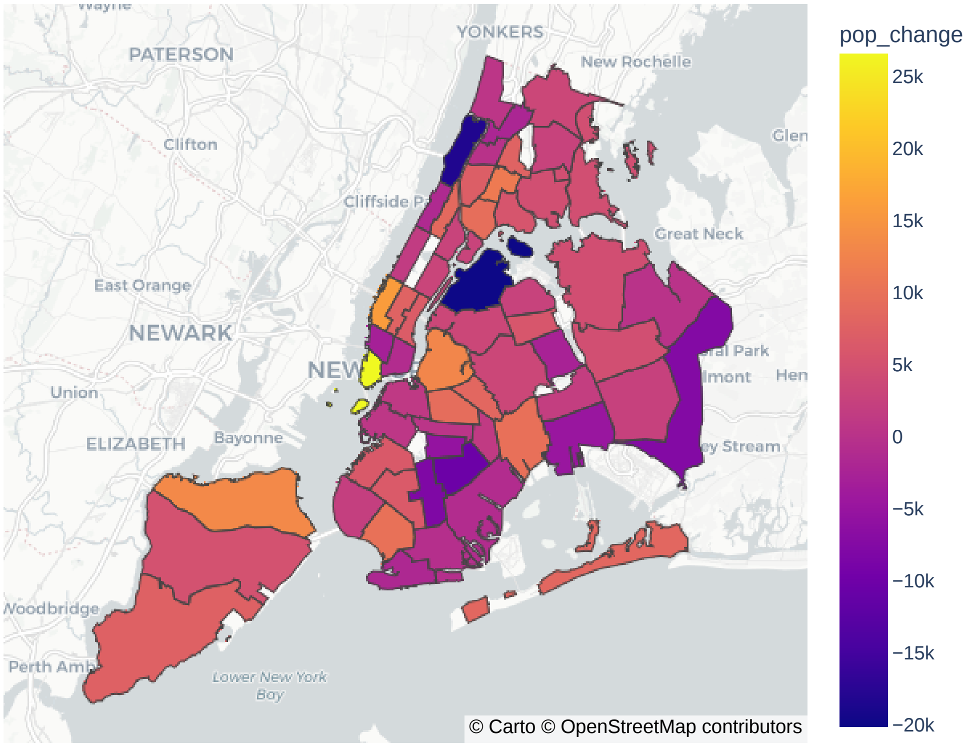
[12]:

```
# your code here
def plot_nyc(pop_by_cd):
    # putting this in a function to save space on subsequent slides

    fig = px.choropleth_mapbox(
        pop_by_cd,
        geojson=shapes,
        locations="borocd",
        featureidkey="properties.boro_cd",
        color="pop_change",
        hover_data=["CD Name"],
        center={"lat": 40.71, "lon": -73.98},
        zoom=9,
        mapbox_style="carto-positron",
        height=600,
    )

    fig.show()
```

```
[13]: plot_nyc(pop_by_cd)
```



© Carto © OpenStreetMap contributors

1.3.6 Step 6

Analysis: Washington Heights and Inwood (the tall skinny community district at the top of Manhattan) are "up and coming" neighborhoods. In a few sentences: Why might might the population have decreased?

The population has decreased might because the residents cannot afford the housing rent.Local residents in up & coming neighborhoods earned relatively lower incomes but the housing rent was relatively high.

Then, read the first three paragraphs of the `Demographics` section of An Economic Snapshot of Washington Heights and Inwood from June 2015.

Now turn in the assignment.

1.4 Tutorials

1. Go through the first third of Time Series Analysis with Pandas, up until the "Visualizing time series data" section.
2. Read how to handle time series data in pandas.
3. Read the Data Design Standards.
4. Watch this talk on audification/sonification. We won't be doing so in this class, but hopefully will provide some inspiration about different ways that data can be represented.
5. Optional: Read about other tools and techniques for visualization in Python.
   - PyViz
   - Python Graph Gallery

1.5 Participation

Reminder about the between-class participation requirement.