CMPINF 0010 Group 24

Leo Liang and Peter Ju

Agenda

1. Introduction
2. Metrics and Datasets
3. Data Analyses
4. Result Explanation
5. Discussion
6. Conclusion

1. Introduction

- The environment we live in is essential.

- We would like to know which Pittsburgh neighborhood is the best on an environmental and individual wellbeing basis.

2. Metrics and Datasets

2.1 Metrics

Our metric in determining which neighborhood is the best is based on three dimensions/datasets (WPRDC)

- Environment (Air Quality PM 2.5)
- Environment (Illegal Dump Sites)
- Individual Health (Hypertension + Hyperlipidemia)

2.2 Datasets

Our metric in determining which neighborhood is the best is based on three dimensions/datasets (WPRDC)

- Allegheny County Particulate Matter 2.5

  https://data.wprdc.org/dataset/particulate-matter-2-5

- Allegheny County Illegal Dump Site

  https://data.wprdc.org/dataset/allegheny-county-illegal-dump-sites

- Hypertension + Hyperlipidemia Comorbidity

  https://data.wprdc.org/dataset/hypertension-hyperlipidemia-comorbidity

2.2 Datasets

Besides these three datasets, we also incorporated the dataset (shapefile) of neighborhoods.

https://data.wprdc.org/dataset/neighborhoods2

3. Data Analyses

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

- Import modules

```
import geopandas as geopd
import pandas as pd
```

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

• Read the dataset

```
pm25 = pd.read_csv("Data/pm25.csv")
pm25.head()
```

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

- Read the dataset

Used for location reference

	CensusTract	PM25
0	10300	12.264655
1	20100	12.283052
2	20300	12.241216
3	30500	12.234197
4	40200	12.315312

The critical values that we are looking for

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

• Read the shapefile

```
neighborhoods = geopd.read_file("Data/pittsburghpaneighborhoods-/Neighborhoods_.shp")
neighborhoods.head()
```

	OBJECTID	fid_blockg	statefp10	countyfp10	tractce10	blkgrpce10	geoid10	namelsad10	mtfcc10	funcstat10	...	created_us	created_da	last_edite	last_edi_1	temp	pghdb_sde_	GlobalID	Shape__Are	Shape__Len	geometry
0	1	233	42	003	140500	3	420031405003	Block Group 3	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	7968760.0	a642a6b6-a2a9-48d5-acf2-3e2ce87d4bc0	8.153647e+06	14047.065024	POLYGON ((-79.90560 40.45522, -79.90553 40.455...
1	2	203	42	003	140100	3	420031401003	Block Group 3	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	34078712.0	20d4605b-258e-441a-a57e-a28fdabb6d30	3.407739e+07	30347.886823	POLYGON ((-79.91309 40.43804, -79.91321 40.438...
2	3	86	42	003	101700	2	420031017002	Block Group 2	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	12751163.0	bf4065e7-c29c-48bd-b529-e7fd99077238	1.275116e+07	14577.441571	POLYGON ((-79.93457 40.47217, -79.93455 40.472...
3	4	214	42	003	050900	1	420030509001	Block Group 1	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	4997335.5	fed3fb32-0cd3-4556-aa41-17ee03b4fe41	4.997845e+06	14216.959213	POLYGON ((-79.96601 40.45520, -79.96596 40.454...
4	5	173	42	003	300100	5	420033001005	Block Group 5	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	8411048.0	8669d2f0-b886-4519-99d0-bb1015b592b0	8.410563e+06	12003.269761	POLYGON ((-79.98845 40.41851, -79.98821 40.418...

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

- Create our own pandas.DataFrame

```python
neighborhood_pm25 = pd.DataFrame(columns = ["Neighborhood", "tractce10", "PM25"])
neighborhood_pm25["Neighborhood"] = neighborhoods["hood"]
neighborhood_pm25["tractce10"] = neighborhoods["tractce10"]

neighborhood_pm25[["PM25"]] = neighborhood_pm25[["PM25"]].astype(float)
neighborhood_pm25
```

	Neighborhood	tractce10	PM25
0	Point Breeze North	140500	NaN
1	Squirrel Hill North	140100	NaN
2	Garfield	101700	NaN
3	Bedford Dwellings	050900	NaN
4	Knoxville	300100	NaN
...	...	...	...

Name of the neighborhood

Critical values: How are they critical if they are all NaNs?

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

• Iterate through the dataset containing PM 2.5 values

```python
for ignored, row in pm25.iterrows():
    ct = row["CensusTract"]
    for index, each_neighborhood in neighborhood_pm25.iterrows():
        if str(each_neighborhood["tractce10"]) == "None":
            continue
        else:
            if float(str(each_neighborhood["tractce10"])) == float((str(ct))):
                if pd.isna(neighborhood_pm25.at[index, "PM25"]):
                    neighborhood_pm25.at[index, "PM25"] = row["PM25"]

neighborhood_pm25
```

	Neighborhood	tractce10	PM25
0	Point Breeze North	140500	12.280433
1	Squirrel Hill North	140100	12.285337
2	Garfield	101700	12.199422
3	Bedford Dwellings	050900	12.243351
4	Knoxville	300100	12.317734
...	...	...	...

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

• Sort the data ascendingly

```
neighborhood_pm25 = neighborhood_pm25.sort_values("PM25", ascending=True)
neighborhood_pm25.head(10)
```

|	Neighborhood	tractce10	PM25
78	Windgap	563000	12.157559
23	Fairywood	563000	12.157559
50	Chartiers City	563000	12.157559
84	Upper Lawrenceville	101100	12.157773
21	Lincoln-Lemington-Belmar	980000	12.158553
71	Stanton Heights	101800	12.159449
76	Summer Hill	261200	12.160112
13	Morningside	101400	12.162134
28	East Carnegie	562800	12.169079
7	Oakwood	562800	12.169079

3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

• Draw a bar graph

```
plot = neighborhood_pm25.plot(kind='bar', x="Neighborhood", y="PM25", title="PM 2.5 by Neighborhood", figsize=(20,10)).set_ylim([12,12.5])
```
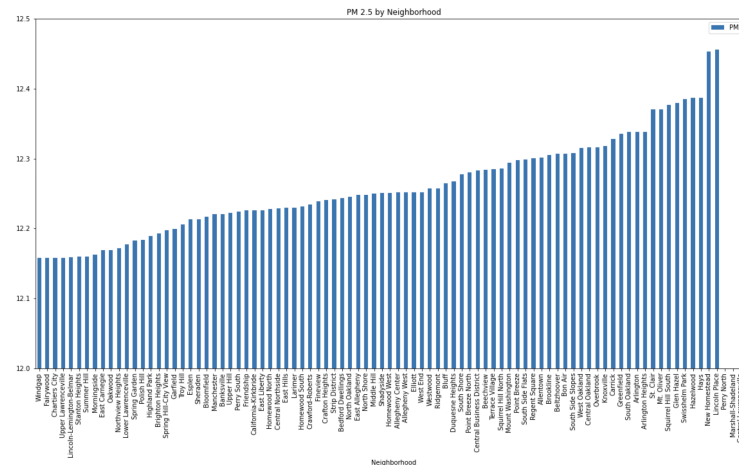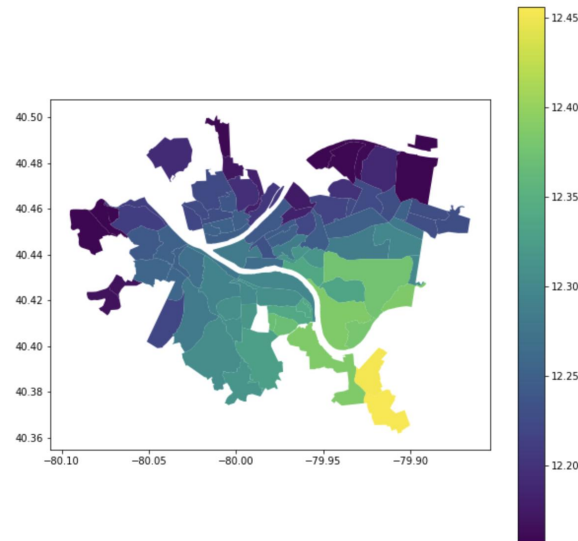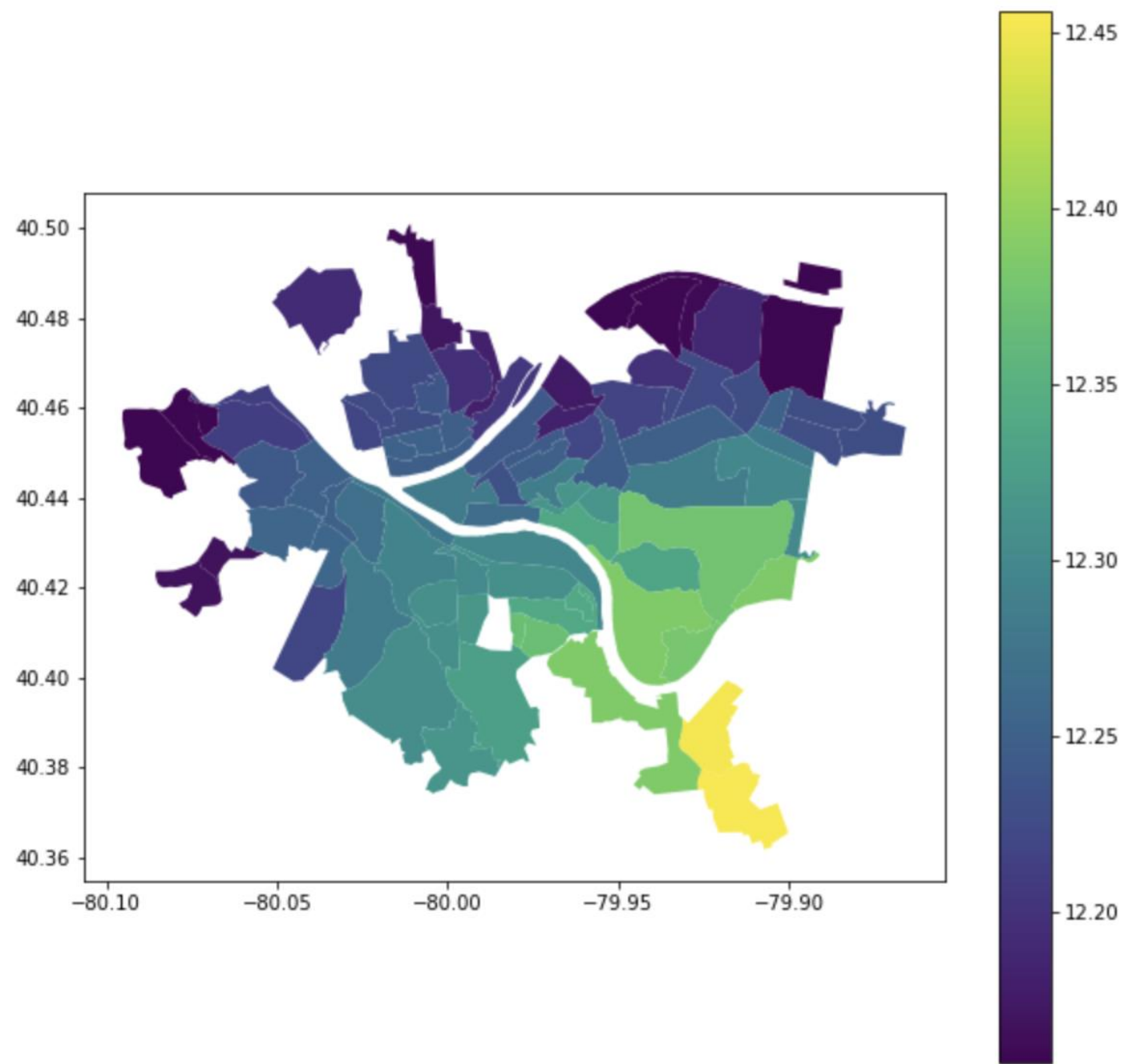
PM 2.5 by Neighborhood
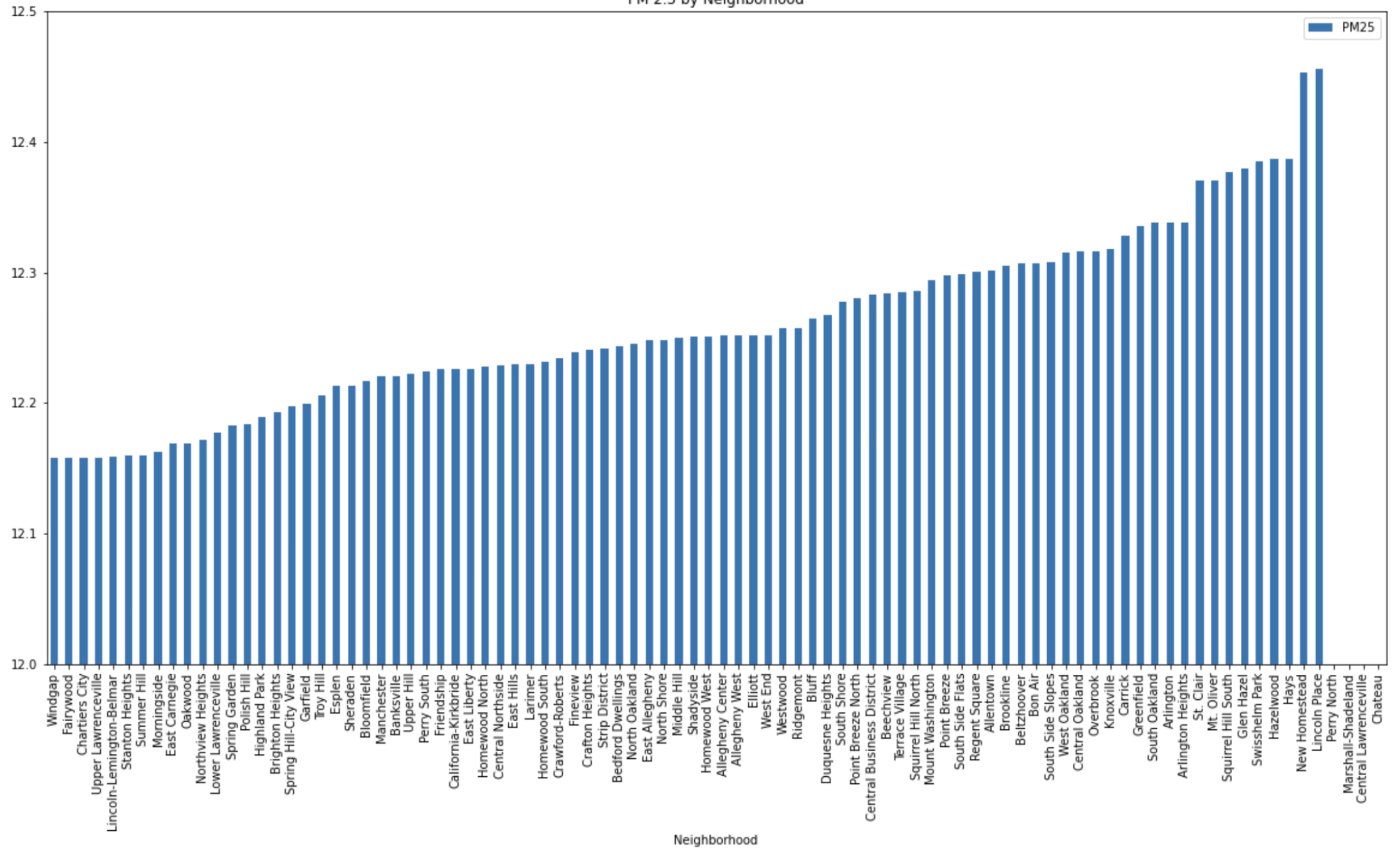
3.1 Data Analyses

Allegheny County Particulate Matter 2.5 (Peter)

- Put this on a map

```python
neighborhoods["PM25"] = neighborhood_pm25["PM25"]

neighborhoods.plot(column="PM25", legend=True, figsize=(10,10))
```

3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

- Read the dataset

```
dump = pd.read_csv("Data/dump-site.csv")
dump.head()
```

	site_name	Status	City	Neighborhood	estimated_tons	location_description	latitude	longitude
0	St. Martin Street	Surveyed	Pittsburgh	Allentown	0.5	NaN	40.42221971	-79.99022525
1	Brosville Street	Surveyed	Pittsburgh	Allentown	3	NaN	40.42370101	-79.98657393
2	McCain Street	Surveyed	Pittsburgh	Allentown	1	NaN	40.42427063	-79.99022675
3	Ceres Way	Surveyed	Pittsburgh	Allentown	0.5	NaN	40.42327242	-79.99248958
4	Eureka Street	Completed	Pittsburgh	Allentown	0.1	NaN	40.42372131	-79.99478912

Used for location reference   Critical values

3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

• Read the shapefile

```
neighborhoods = geopd.read_file("Data/pittsburghpaneighborhoods-/Neighborhoods_.shp")
neighborhoods.head()
```

	OBJECTID	fid_blockg	statefp10	countyfp10	tractce10	blkgrpce10	geoid10	namelsad10	mtfcc10	funcstat10	...	created_us	created_da	last_edite	last_edi_1	temp	pghdb_sde_	GlobalID	Shape__Are	Shape__Len	geometry
0	1	233	42	003	140500	3	420031405003	Block Group 3	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	7968760.0	a642a6b6-a2a9-48d5-acf2-3e2ce87d4bc0	8.153647e+06	14047.065024	POLYGON ((-79.90560 40.45522, -79.90553 40.455...
1	2	203	42	003	140100	3	420031401003	Block Group 3	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	34078712.0	20d4605b-258e-441a-a57e-a28fdabb6d30	3.407739e+07	30347.886823	POLYGON ((-79.91309 40.43804, -79.91321 40.438...
2	3	86	42	003	101700	2	420031017002	Block Group 2	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	12751163.0	bf4065e7-c29c-48bd-b529-e7fd99077238	1.275116e+07	14577.441571	POLYGON ((-79.93457 40.47217, -79.93455 40.472...
3	4	214	42	003	050900	1	420030509001	Block Group 1	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	4997335.5	fed3fb32-0cd3-4556-aa41-17ee03b4fe41	4.997845e+06	14216.959213	POLYGON ((-79.96601 40.45520, -79.96596 40.454...
4	5	173	42	003	300100	5	420033001005	Block Group 5	G5030	S	...	SDE	2020-08-14	SDE	2020-08-14	None	8411048.0	8669d2f0-b886-4519-99d0-bb1015b592b0	8.410563e+06	12003.269761	POLYGON ((-79.98845 40.41851, -79.98821 40.418...

3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

- Create our own pandas.DataFrame

```python
neighborhood_dump = pd.DataFrame(columns = ["Neighborhood", "Amount"])
neighborhood_dump["Neighborhood"] = neighborhoods["hood"]

neighborhood_dump[["Amount"]] = 0
neighborhood_dump
```

	Neighborhood	Amount
0	Point Breeze North	0
1	Squirrel Hill North	0
2	Garfield	0
3	Bedford Dwellings	0
4	Knoxville	0
...	...	...

3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

• Iterate through the dataset containing the amounts

```python
for ignored, row in dump.iterrows():

    for index, neighborhood in neighborhood_dump.iterrows():
        if neighborhood["Neighborhood"] == row["Neighborhood"]:
            if neighborhood_dump.at[index, "Amount"] == 0:
                neighborhood_dump.at[index, "Amount"] = row["estimated_tons"]
            else:
                neighborhood_dump.at[index, "Amount"] = neighborhood_dump.at[index, "Amount"] + row["estimated_tons"]

neighborhood_dump
```

	Neighborhood	Amount
0	Point Breeze North	0.3
1	Squirrel Hill North	0.0
2	Garfield	19.4
3	Bedford Dwellings	1.0
4	Knoxville	9.5
...	...	...

3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

- Sort the data ascendingly

```
neighborhood_dump = neighborhood_dump.sort_values("Amount", ascending=True)
neighborhood_dump.head(20)
```

	Neighborhood	Amount
44	East Allegheny	0.0
34	Manchester	0.0
35	Marshall-Shadeland	0.0
41	Friendship	0.0
43	Central Business District	0.0
88	South Side Flats	0.0
45	Allegheny Center	0.0
48	South Side Slopes	0.0
19	Lower Lawrenceville	0.0
18	Arlington Heights	0.0
54	Mount Washington	0.0

	Neighborhood	Amount
56	Central Lawrenceville	0.0
59	Point Breeze	0.0
32	Allegheny West	0.0
14	Lincoln Place	0.0
12	Westwood	0.0
11	East Liberty	0.0
64	Ridgemont	0.0
66	South Shore	0.0
69	Spring Hill-City View	0.0

3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

- Look at the tails

```
neighborhood_dump.tail(20)
```

	Neighborhood	Amount
67	Mt. Oliver	11.3
5	Homewood South	12.9
58	Esplen	15.0
20	Larimer	16.9
77	West Oakland	17.3
49	Arlington	17.5
55	East Hills	17.8
9	Duquesne Heights	19.0
2	Garfield	19.4
37	New Homestead	20.0
71	Stanton Heights	21.7

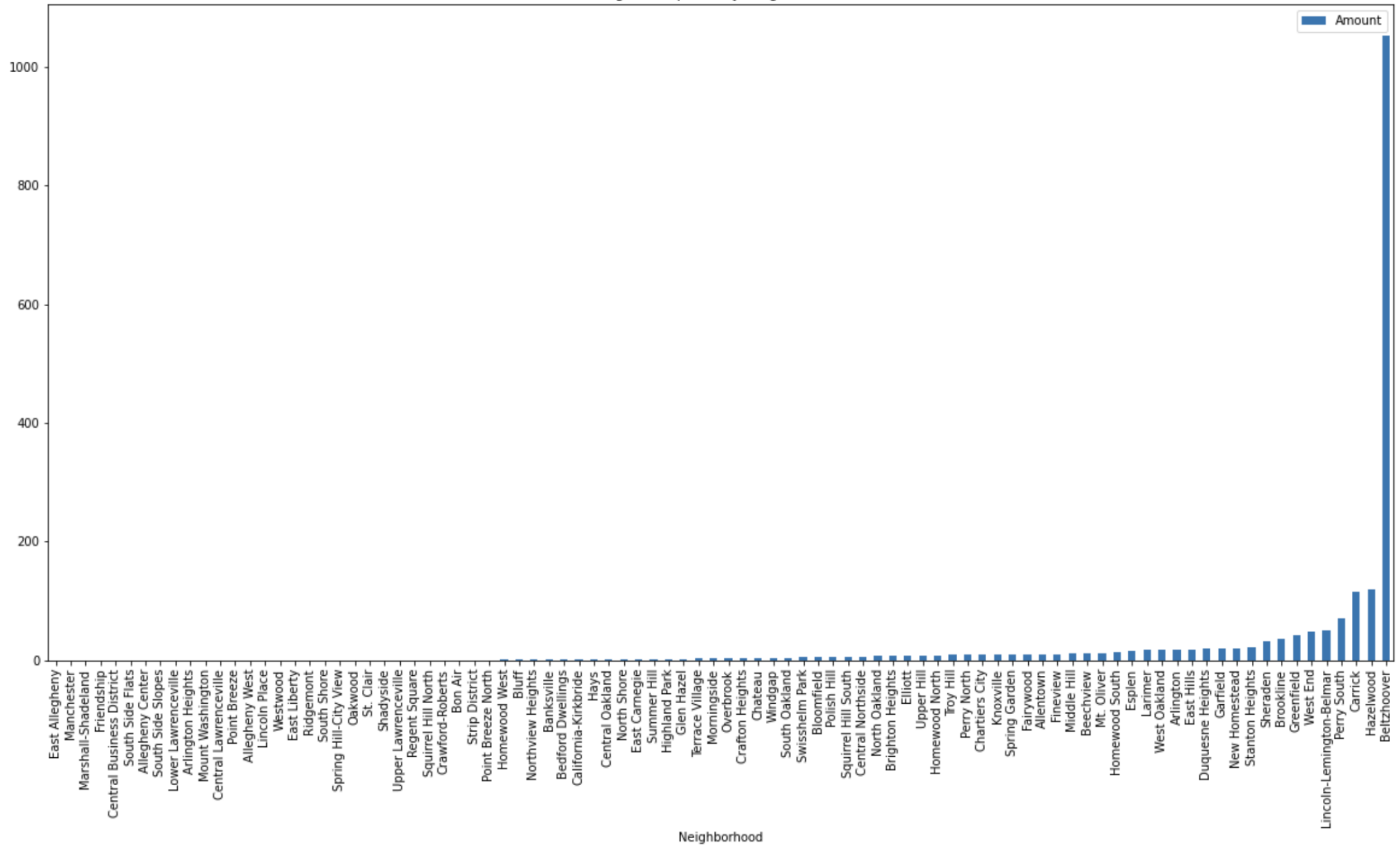|	Neighborhood	Amount
70	Sheraden	31.8
24	Brookline	35.7
27	Greenfield	41.8
57	West End	47.2
21	Lincoln-Lemington-Belmar	49.6
83	Perry South	70.7
51	Carrick	115.3
60	Hazelwood	118.6
26	Beltzhoover	1053.0

3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

- Draw a bar graph

```
plot = neighborhood_dump.plot(kind='bar', x="Neighborhood", y="Amount", title="Illegal Dump Site by Neighborhood", figsize=(20,10))
```
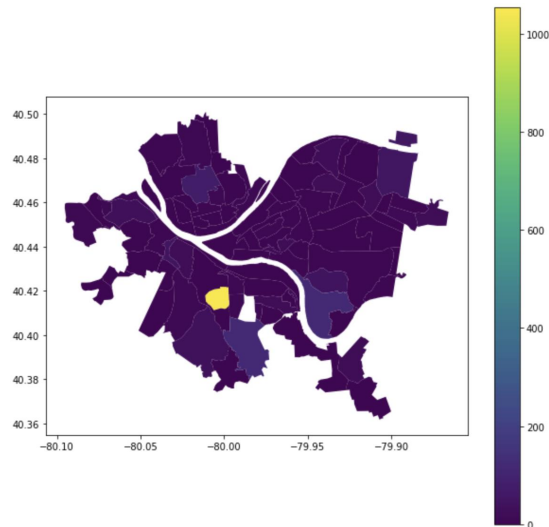
Illegal Dump Site by Neighborhood
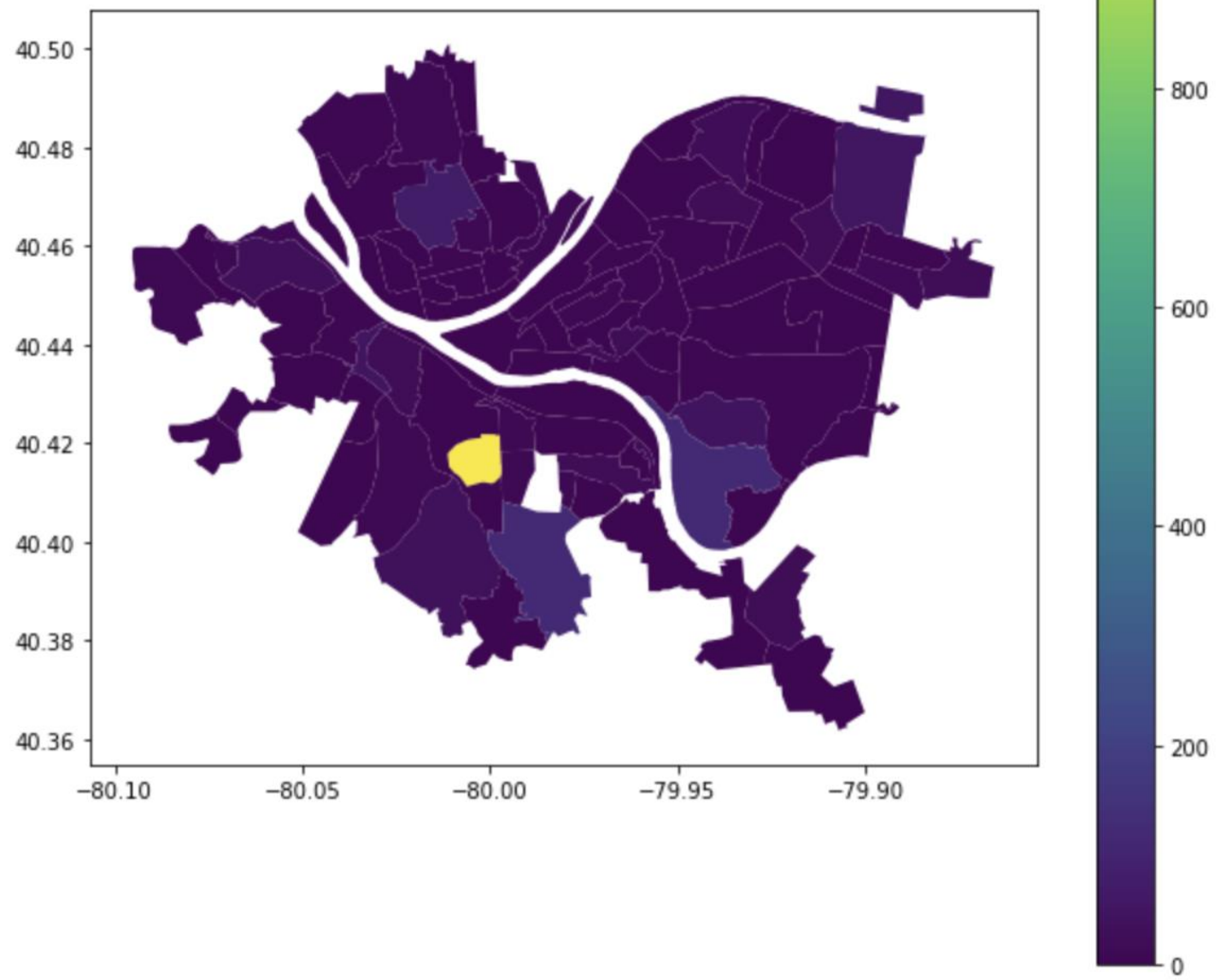
3.2 Data Analyses

Allegheny County Illegal Dump Site (Peter)

- Put this on a map

```
neighborhoods["Amount"] = neighborhood_dump["Amount"]

neighborhoods.plot(column="Amount", legend=True, figsize=(10,10))
```

Hypertension/Hyperlipidemia

3.3 Data Analysis

Freshly imported dataset, though labels are a bit cryptic

	CT	LTPAD	LTPAN	LTWAD	LTWAN	LTMAD	LTMAN	geometry
0	10300	422	28	197	13	225	15	None
1	20100	2610	175	1242	73	1368	102	None
2	20300	354	38	175	12	179	26	None
3	30500	859	147	553	103	306	44	None
4	40200	388	44	233	29	155	15	None
...	...	...	...	...	...	...	...	...
385	564400	2797	630	1529	333	1268	297	None
386	564500	806	146	440	78	366	68	None
387	980000	4	1	3	1	1	0	None
388	980100	7	0	6	0	1	0	None
389	980700	9	0	5	0	4	0	None

But luckily...

3.3 Data Analysis

There's a dictionary

	column	type	label	description
1	CT	text		Census Tract
2	LTPAD	text		dar year (Jan 1 - Dec 31).
3	LTPAN	text		osed with Hyperlipidemia.
4	LTWAD	text		dar year (Jan 1 - Dec 31).
5	LTWAN	text		osed with Hyperlipidemia.
6	LTMAD	text		dar year (Jan 1 - Dec 31).
7	LTMAN	text		osed with Hyperlipidemia.

CT: Census Tract

LTPAD: Total number of members, who are Allegheny County residents and who were effectively enrolled for at least 90 continuous days during the 2015 calendar year (Jan 1 - Dec 31).

LTPAN: Total number of members who met the denominator criteria, who are diagnosed with Hypertension & diagnosed with Hyperlipidemia.

LTWAD: Total number of women members, who are Allegheny County residents and who were effectively enrolled for at least 90 continuous days during the 2015 calendar year (Jan 1 - Dec 31).

LTWAN: Total number of women members who met the denominator criteria, who are diagnosed with Hypertension & diagnosed with Hyperlipidemia.

LTMAD: Total number of male members, who are Allegheny County residents and who were effectively enrolled for at least 90 continuous days during the 2015 calendar year (Jan 1 - Dec 31).

LTMAN: Total number of male members who met the denominator criteria, who are diagnosed with Hypertension & diagnosed with Hyperlipidemia.
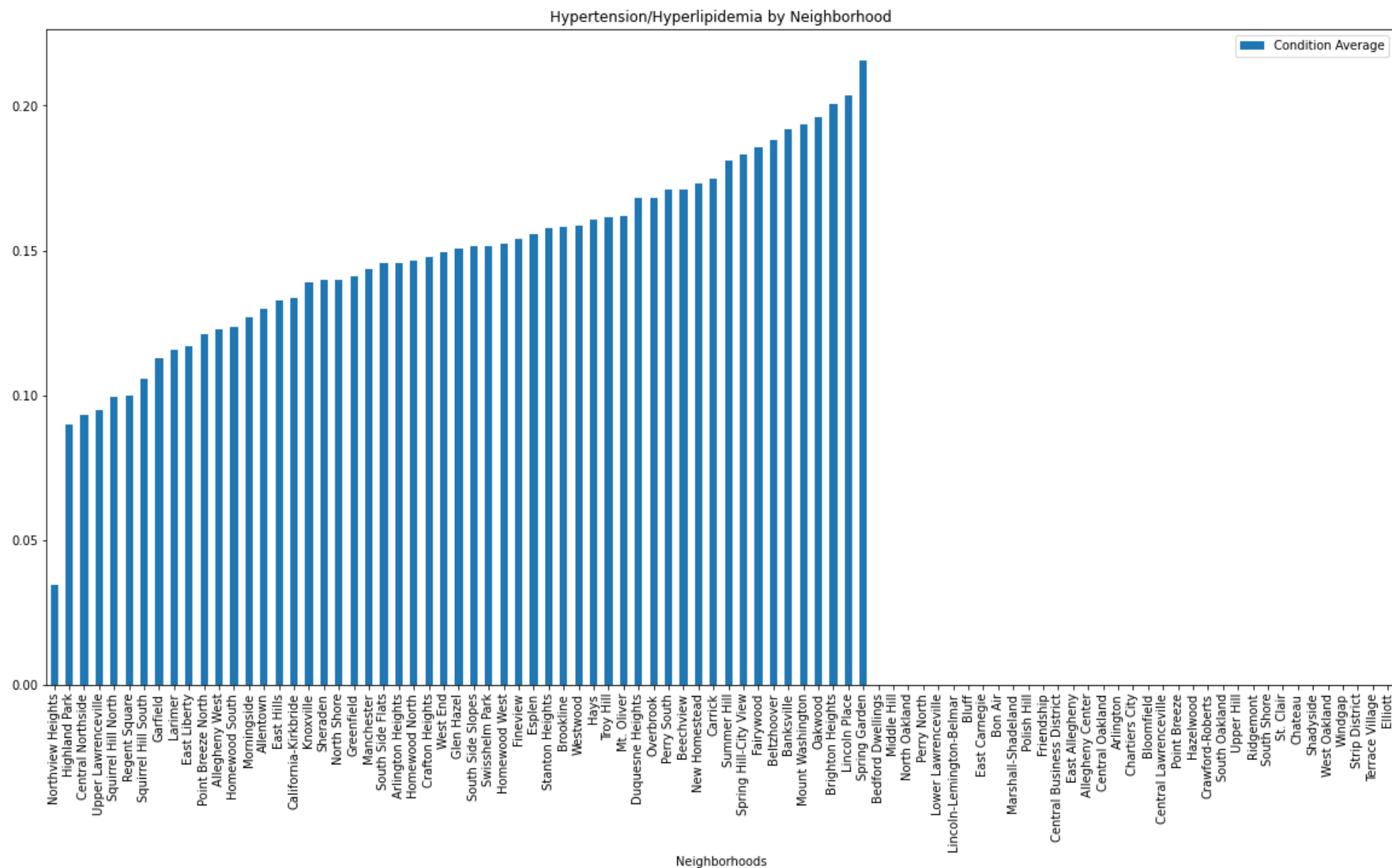
3.3 Data Analysis

Averaging the data...

Some may be missing :O

CT	Neighborhoods	Condition Average
140500	Point Breeze North	0.121212
140100	Squirrel Hill North	0.099307
101700	Garfield	0.112782
050900	Bedford Dwellings	NaN
300100	Knoxville	0.13883
...	...	...
141000	Regent Square	0.1
051100	Terrace Village	NaN
562600	Elliott	NaN
170200	South Side Flats	0.145564
250900	Fineview	0.154034

Putting it all together

Hypertension/Hyperlipidemia by Neighborhood

3.3 Data Analysis

Per the data, the top 5 neighborhoods with the lowest rates of hypertension and hyperlipidemia are:

1. Northview Heights
2. Highland Park
3. Central Northside
4. Upper Lawrenceville
5. Squirrel Hill North
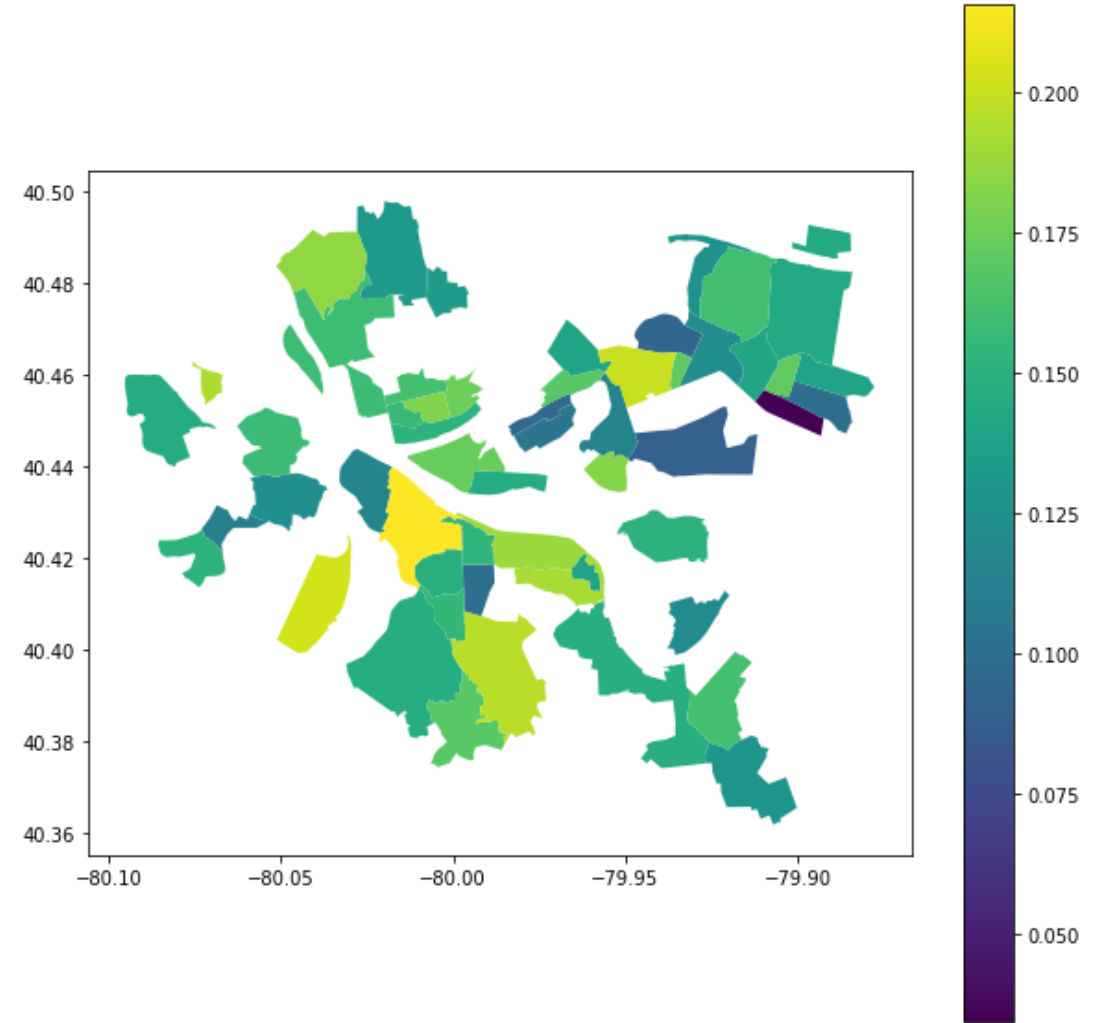
4. Result Explanation

	Neighborhood	Rank_PM25	Rank_Dump	Rank_Hype	Total
16	Northview Heights	11	32	1	0
36	Highland Park	15	42	2	0
38	Central Northside	31	53	3	0
84	Upper Lawrenceville	4	1	4	0
1	Squirrel Hill North	59	1	5	0
...	...	...	...	...	...
77	West Oakland	69	75	86	0
78	Windgap	1	48	87	0
81	Strip District	38	29	88	0
86	Terrace Village	58	44	89	0
87	Elliott	48	58	90	0

We ranked the neighborhoods based on the 3 categories

1. Lowest Pollution

2. Lowest Dump Sites

3. Lowest Hypertension and Hyperlipidemia average

Then added them together to obtain their final score

And the best Neighborhood
(by our metric) is…

Upper Lawrenceville!

	Neighborhood	Rank_PM25	Rank_Dump	Rank_Hype	Total
84	Upper Lawrenceville	4	1	4	9
11	East Liberty	29	1	10	40
16	Northview Heights	11	32	1	44
34	Manchester	23	1	22	46
36	Highland Park	15	42	2	59
32	Allegheny West	46	1	12	59
7	Oakwood	9	1	52	62
69	Spring Hill-City View	17	1	47	65
1	Squirrel Hill North	59	1	5	65
13	Morningside	8	45	14	67

5. Discussion

- Counterclaims:
  - More metrics?
  - Time span
  - Ranking:
    - Weight?

6. Conclusion

References

https://data.wprdc.org/dataset/particulate-matter-2-5

https://data.wprdc.org/dataset/allegheny-county-illegal-dump-sites

https://data.wprdc.org/dataset/hypertension-hyperlipidemia-comorbidity

https://data.wprdc.org/dataset/neighborhoods2

Thanks

CMPINF 0010 Group 24