FinalProject_stephanielzhou

December 16, 2020

1 COGS 108 - Final Project

2 Overview

Sorry to my graders <3 I am trying to submit an incomplete for the class. This is my best attempt now.

The advent of the COVID-19 pandemic means that gathering in indoor spaces is no longer as safe [1], as a result, this "Quarantine Fatigue" has more people going outside [3]. Studies continue to show that neighborhoods with higher access to outdoor spaces have a number of health benefits such as lower obesity rates[2]. However, like majority of public resources, there is often inequity disproportionately affecting people of lower incomes, BIPOC's, or other marginalized groups. In this specific exploration, I will looking further into the distribution of park resources compared to the wealth distribution of San Diego county in the hopes of informing future park creation by the City of San Diego's Parks and Recreation Council.

3 Name & GitHub ID

- Name: Lan (stephanie) Zhou
- GitHub Username: stephanielzhou

4 Research Question

How does income affect the population of San Diego County's access to outdoor parks and spaces? Is the distribution equitable per person in different neighborhoods?

4.1 Background and Prior Work

Why is this important? Especially with the onset of COVID-19, outdoor spaces are a healthy and safe way to social distance and relieve stress. Mental health is

What has already been done on this topic? What is already known? There is a study in Los Angeles county[4] that concluded Latinx, African American, and lower-income folx are more likely to live in areas with fewer park acres per person.The equitable distribution of parks and recreation resources is important because their study indicates a correlation between children's distances from parks directly to their health. Noticeably, the study emphasized that the issue is centered around race because of the disproportionate amount of access between White communities and the others, which intersects a multitude of other social issues.

San Diego has similar disparities in wealth with large Hispanic and Latinx community, leading me to believe that there may be similar circumstances in the county. Addressing this disparity and analyzing public data can further the city's decisions of fund distribution for further park and recreation additions.

References (include links): - 1) CDC social distancing guidelines: https://www.cdc.gov/pcd/issues/2014/13_0407.htm - 2) Guide to Designing Healthier Spaces: https://healthyplacesbydesign.org/parks-and-community-health/ - 3) "quarantine fever": https://www.nytimes.com/2020/04/27/us/coronavirus-social-distancing-lockdown.html - 4) LA study: https://www.activelivingresearch.org/disparities-park-space-race-and-income - 5) San Diego Wealth Gap: https://www.sandiegouniontribune.com/business/economy/sd-fi-wage-gap-20180511-story.html

5 Hypothesis

I predict that there is an unequitable distribution of parks and recreation spaces readily available lower-income families. There is less acreage per person in a lower income area than a higher income neighborhood.

6 Dataset(s)

1. Census_ZIP: Link: https://sdgis-sandag.opendata.arcgis.com/datasets/census-zip?layer=5 Number of Observations:

2. Population_estimate: Link: https://datasurfer.sandag.org/dataoverview Number of Observations:

3. Park Distribution for Population: Link: https://data.ca.gov/dataset/park-beach-open-space-or-coastline-access Number of Observations:

4. Income Distribution: Link: https://datasurfer.sandag.org/dataoverview Number of Observations:

5. Park Locations (for mapping purposes): Link: https://data.sandiego.gov/datasets/park-locations/ Number of Observations:

7 Setup

```python
# Imports
%matplotlib inline
import pandas as pd
import geopandas as gpd
import numpy as np
import fiona

import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = (17, 5)
plt.rcParams.update({'font.size': 16})
from mpl_toolkits.axes_grid1 import make_axes_locatable
```

```
import seaborn as sns

import shapely.geometry as shp

import sklearn.neighbors as skn
import sklearn.metrics as skm
```

8 Data Cleaning

Describe your data cleaning steps here.

```
[39]: file = "Census_ZIP.csv"
      tempmap = gpd.read_file(file)

      pop = pd.read_excel("Population_Estimate_2019.xlsx").
       ↪groupby('ZIP')['POPULATION'].sum().to_frame()
```

```
[ ]: fig, ax = plt.subplots(1, 1, figsize=(17, 7))
     divider = make_axes_locatable(ax)
     sdmap.plot(column='POPULATION', ax=ax, cmap='coolwarm', legend=True);
```

9 Data Analysis & Results

Include cells that describe the steps in your data analysis.

visualization: population map of san diego

```
[ ]: visualization:
```

```
[ ]:
```

```
[ ]: looking at
```

```
[ ]:
```

```
[ ]:
```

10 Ethics & Privacy

There were no data sets that included personal identifiers, so individual data privacy is not a concern. It is important to recognize in the Park Proximity dataset that proximity does not necessarily imply usage. There would need to be further analysis.

11 Conclusion & Discussion

consider building parks in better areas

[ ]: