My Final Project

Jessica De Anda

February 2020

Abstract

Contents

List of Figures

1 Introduction

OUTLINE:

- Present the issue: access to supermarkets in the US is not uniform across the country. Poor access to supermarkets in this report is defined as living far from a supermarket and lacking vehicle access.
  - Why we should care: poor access to food = lower quality of life
  - Biological impact of food deserts: health; research demonstrates that obesity is prominent in communities living in food deserts. The lack of access to supermarkets with fresh produce perpetuates community dependence on fast food, which is typically unhealthy (high fat, low nutrient).
  - REFERENCES: Gordon-Larsen, 2014 ; National Research Council, 2007 ; White, 2007
- Factors that influence access to food: socioeconomic status
  - Income/economic standing: low income communities are less likely to have access to fresh produce
  - Race: racial profile often influences economic standing. Lower income communities tend to be comprised of racial minorities (largely due to social inequity), therefore there may be a correlation between access to food and racial identity (REFERENCE: Tobin and Weaver, 2017)
- Overview of this report
  - Purpose: to assess trends betwwen socioeconomic status and access to food in the US in order to identify the groups that may need additional services from the government to improve their quality of life
  - Methods: obtained data on 2010 food access in the US from USDA. Calculated the percentage of each racial group (white, black, asian, latino, islander, native american, and other) living at 0.5, 1, 10, and 20 miles from a supermarket.
  - Hypothesis: based on past research, it is likely that all racial minorities will have less access to supermarkets than the white population in the US
- Future directions
  - Use GIS to construct a map to better visualize the data. Report randomly selected case studies in order to get a better idea of how living at each of these distances from the supermarket influences different people
  - also look at price –¿ doesn't matter that a supermarket is close by if the food is too expensive to buy!
  - Research ways to address food insecurity

2 Methods

2.1 Importing the data file

Public data on food access in the United States in 2010 was downloaded from the United States Department of Agriculture (USDA) as a csv file. A function was defined within a python script to open the csv file, import the data using the pandas library, and convert the data to a numpy file for future numerical analysis.

```python
import pandas as pd
import numpy as np

def np_from_csv(csv_file):
    temp_data = pd.read_csv(csv_file, header = None)
    data = temp_data.to_numpy()
    return data

data = np_from_csv('food_access_data.csv')
```

2.2 Analysis of access to supermarkets across racial categories

A function was defined to calculate the percentage of a race living at a particular distance from the supermarket. Given an input of race and distance, the defined function first used the numpy.sum operation to calculate the total number of individuals of the race living at the specified distance from the supermarket. This value was then divided by total number of individuals of the race, which was also generated using the numpy.sum operation. The function returned the calculation as a percentage by multiplying the value by 100.

```python
def percent_race_at_distance(input_data, race_distance, race_total, dec = 2):
    # inside the parentheses, the format is: dataset_imported[rows, columns]
    race_distance_sum = np.sum(input_data[:, race_distance])
    race_total_sum = np.sum(input_data[:, race_total])
    percentage = race_distance_sum/race_total_sum * 100
    return round(percentage, dec)
```

A second function took an input of race name and the previously calculated percentage of that race living at a given distance from the supermarket to plot the percentages against the corresponding distances. The plot function of the matplotlib.pyploy library was used to graph this relationship and additional functionalities of this library were used to format the plot.

```python
import matplotlib.pyplot as plt

def plot_race_percentage_vs_distance(race, perc):
    [half,one,ten,twenty] = perc
    plt.plot([0.5, 1, 10, 20],[half, one, ten, twenty])
    # the following code formats the plot
    plt.axis([0, 20, 0, 100])
    plt.xlabel('Distance from Supermarket (Miles)')
    plt.ylabel('Percentage')
    plt.title('Percentage of %s Population Living at a Distance from a Supermarket' % race)
    return
```

All of the data in the file was organized into arrays. The data would later be extracted sequentially and inputted into a for loop that combined the two defined functions.

```python
total = [white_total, black_total, asian_total, islander_total, native_total, other_total,
    latino_total]
# total is an array of the data columns corresponding to the total population of each race

distance = [[white_half, white_1, white_10, white_20],
            [black_half, black_1, black_10, black_20],
            [asian_half, asian_1, asian_10, asian_20],
            [islander_half, islander_1, islander_10, islander_20],
            [native_half, native_1, native_10, native_20],
            [other_half, other_1, other_10, other_20],
            [latino_half, latino_1, latino_10, latino_20]]
# distance is a 7X4 array given that it holds data for 7 racial categories and 4 distances
# each row corresponds to a race and each column corresponds to a distance

name = ['White', 'Black', 'Asian', 'Islander', 'Native', 'Other', 'Latino']
# name is a list of the racial categories
```

The for loop containing both functions was designed to automatically input and analyze all of the data without having to manually enter each race and distance.

```python
perc = np.zeros(7,4)
# create an empty 7X4 array corresponding to the number of races by the number of distances
# the percentages of each race living at a given distance will be organized in this array

```

```
5  for race in range(7):
6  # the following for loop will run 7 times (1X per race)
7      for dist in range(4):
8      # nested within the for loop is a second for loop that will run 4 times (1X per distance)
9      # this set of code will run a total of 28 times (1X per race and distance)
10         perc[race][dist] = percent_race_at_distance(data, distance[race][dist], total[race], 2)
11         # each output of the function "percent_race_at_distance" will be added to array "perc"
12      plot_race_percentage_vs_distance(name[race], perc[race])
13      # the percentages in "perc" will be inputted in "plot_race_percentage_vs_distance"
14      # it will also pull each race from the list "name"
15 plt.legend(name)
16 plt.show()
```

3   Results

4   Discussion

5 References

*** This section will be edited

6 Figures

Figure 1: Food access across racial categories