Proposal: Food Accessibility at the County Level

DATA 450 Capstone

Tenzin Sherpa

2/5/23

1 Introduction

[This section should contain background and introduction to your general topic.]

2 Dataset

The project will use the Food Environment Atlas Data from 9/10/2020 found here. The dataset has been compiled by the US Department of Agriculture (USDA) Economic Research Service (ERS) for purposes of studying factors that affect food choices and the accessibility to healthy foods in communities. The dataset contains county level information on food environment factors such as access to grocery stores/supermarkets/restaurants, local food sales, food prices, food assistance programs like SNAP (Supplemental Nutrition Assistance Program), National School Lunch Program, etc., socioeconomic characteristics and some health/physical activities. State level information on food taxes, and household food insecurity are also available. The data have been collected through various reports from the Census, USDA for years ranging from 2010 to 2019. For further interest on the variables, a detailed documentation can be found here.

Describe how the data was obtained by the data owner/curator, as best as you can. List the variables that you plan to use in your analysis, for example:

- weight: The patient's weight (kg)
- sex: The patient's sex, male or female
- age: The patient's age (months)

3 Data Acquisition and Processing

[In this section, if applicable, describe how you will obtain the data (if it's anything more complicated than a simple download). Discuss what data processing steps will be needed, such as recoding variables, data cleaning, data tidying, imputing missing values, etc. See sections 1c, 1d, 1e in the "Good Enough Practices" paper.]

4 Research Questions and Methodology

[In this section, list each of the questions you will explore. Following each question, provide a detailed and specific plan for how you plan to answer the question. Include the specific steps you will take, what form the answer will take (a number? table? visualization? model? Give all the specifics), and estimate how many hours each question will take to complete.]

1. Is smoking correlated with diabetes? To answer this, I will create a filled bar plot, with the left bar representing non-smokers, the middle bar representing people who smoke moderately, and the right bar representing heavy smokers. The bars will be the same height, and each bar will be colored two colors based on the proportion of patients in the group who do or do not have diabetes.
2. Question 2? Plan for question 2.
3. Question 3? Plan for question 3.
4. etc.

5 Work plan

[Fill in the list below with a plan for what you will do each week. You should have around 7 hours worth of work each week. Writing work counts. Several tasks have already been filled in for you.]

Week 4 (2/6 - 2/12): [Just an example:

- Data tidying and recoding (4 hours)
- Question 2 (4 hours).]

Week 5 (2/13 - 2/19):

Week 6 (2/20 - 2/26):

Week 7 (2/27 - 3/5):

- Presentation prep and practice (4 hours)

Week 8 (3/6 - 3/12): *Presentations in class on Thurs 3/9.*

- Presentation peer review (1.5 hours)

Week 9 (3/20 - 3/26):

- Poster prep (4 hours)

Week 10 (3/27 - 4/2): *Poster Draft 1 due Monday 3/27. Peer feedback due Thursday 3/30.*

- Peer feedback (2.5 hours)
- Poster revisions (2 hours)

Week 11 (4/3 - 4/9): *Poster Draft 2 due Monday 4/3. Final Poster due Saturday 4/8.*

- Poster revisions (1 hour).

Week 12 (4/10 - 4/16):

Week 13 (4/17 - 4/23): [All project work should be done by the end of this week. The remaining time will be used for writing up and presenting your results.]

- Draft blog post (5 hours).

Week 14 (4/24 - 4/30): *Blog post draft 1 due Monday 4/24. Peer feedback due Thursday 4/27. Blog post draft 2 due Sunday 4/30.*

- Peer feedback (2.5 hours)
- Blog post revisions (2 hours)
- [Do not schedule any other tasks for this week.]

Week 15 (5/1 - 5/7): *Final blog post due Tuesday 5/2.*

- Final presentation prep and practice.
- [Do not schedule any other tasks for this week.]

Final Exam Week (5/8): *Final Presentations during final exam slot, Monday May 9th 3:20-6:40pm.* [Do not schedule any other tasks for this week.]

5.1 Some cool Quarto stuff

[You can delete this section from your proposal.]

For your reference, here's an example of a Python code cell in Quarto, along with a figure that gets generated, along with a caption and a label so that it can be referred to automatically as "Figure 1" (or whatever) in the writeup.

For a demonstration of a line plot on a polar axis, see Figure 1.

```
import numpy as np
import matplotlib.pyplot as plt

r = np.arange(0, 2, 0.01)
theta = 2 * np.pi * r
fig, ax = plt.subplots(
  subplot_kw = {'projection': 'polar'}
)
ax.plot(theta, r)
ax.set_rticks([0.5, 1, 1.5, 2])
ax.grid(True)
plt.show()
```

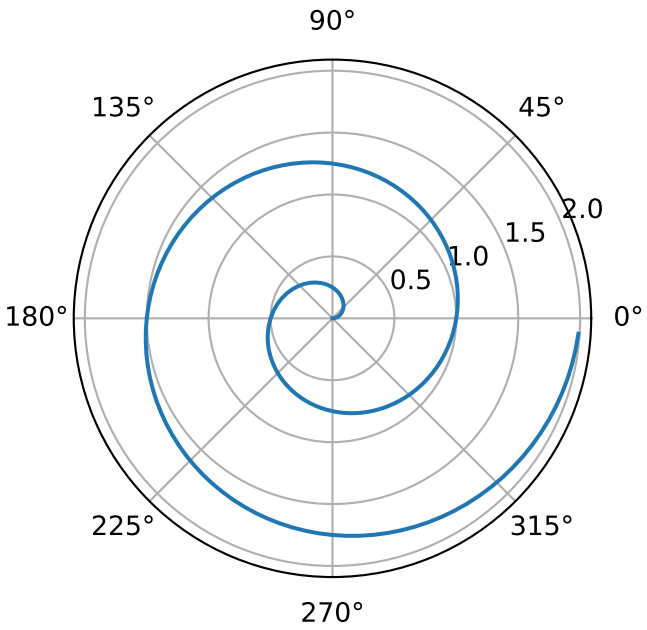

Figure 1: A line plot on a polar axis

Here's an example of citing a source (see Phillips 1999, 33–35). Be sure the source information is entered in "BibTeX" form in the `references.bib` file.

6 References

[The bibliography will automatically get generated. Any sources you cite in the document will be included. Other entries in the `.bib` file will not be included.]

Phillips, T. P. 1999. "Possible Influence of the Magnetosphere on American History." *J. Oddball Res.* 98: 1000–1003.