What Parts of the United States are the Most Vegetarian- and Vegan-Friendly?

Noah O'Neill

October 2022

Introduction

In 2022 I made the decision to become a vegetarian, while also putting in extra effort to cook almost exclusively vegan meals at home. A common topic of debate in the vegan community is whether or not it is easier for certain communities to adopt healthy and plant-based diets than it is for others. Personally, I think the answer to this question is a no-brainer, but nonetheless would like to explore it with data.

While a vegetarian or vegan diet isn't for everyone, the reduction of human consumption of meat for nutritionally equal or healthier plant-based alternatives has been shown to be beneficial on multiple levels. Farming for animal products is one of the world's biggest contributors to climate change, with 70% of freshwater use and 77% of land use being used for maintaining and feeding livestock. On an individual level, there is some evidence that reducing the amount of meat one eats (while, again, still taking care to keep a well-balanced diet) can lower the risk of certain health conditions like heart disease.

There are many factors to consider when it comes to transitioning to a new diet, but for this study I will be judging how easy it is to be a vegetarian or vegan based on how many restaurants with vegetarian or vegan options are available, as this data is easy to access and eating out at restaurants that often will not accommodate your dietary restrictions is one of the most difficult parts of having a plant-based diet.

Hypothesis

There are more meat-free and plant-based dining options in areas with high levels of income, as well as those with larger populations.

Preparation

The primary dataset I'll be using is a table containing the amount of vegetarian- and vegan-friendly restaurants per state. By vegetarian/vegan-friendly, I refer to restaurants that provide food that is meat- or animal-product-free on their menus. This data was scraped from the website HappyCow, an online directory of such restaurants, and was the most accurate I could find. Most of the sets containing the demographic data I'll be using are sourced from the U.S. Census Bureau; for consistency, all of this government data was collected in 2018. The additional data on political party preference was collected in 2018 by Gallup.

The tools I will be using are:

- Data Scraping: Python with Selenium
- Cleaning: PostgreSQL/PGAd4min4, Microsoft Excel
- Exploration/Analysis: Python with numpy, pandas, and scipy
- Visualization: Python with Seaborn, Tableau
- Final report: rmarkdown

Processing and Cleaning Data

The collected Happy Cow data was simple and straightforward, and thus did not require any major cleaning. I also originally used SQL to process a dataset of fast food restaurants, but tossed this set after realizing it doesn't contribute to answering the question at hand.

Analysis

From here I began my exploratory data analysis.

```
# Loading the necessary packages
import numpy as np
import pandas as pd
import openpyxl
import seaborn as sns
import matplotlib.pyplot as plt
sns.set_theme(style="whitegrid")
from scipy.stats import pearsonr

# Loading the vegetarian restaurant data
veggiefood = pd.read_excel("C:/Users/Noah/Coding/food_study/data/clean_data/veggie_res

# Loading in state data
statepop = pd.read_excel("C:/Users/Noah/Coding/food_study/data/clean_data/state_popula
statepop.columns = ['name', 'population']
# Creating a Series of each state and its population
populations = statepop['population']
populations.index = statepop['name']

# Our first look at the vegetarian restaurant data, sorted by the number of restaurant
veggiefood.sort_values(['num_of_restaurants'], ascending=False).head()
```

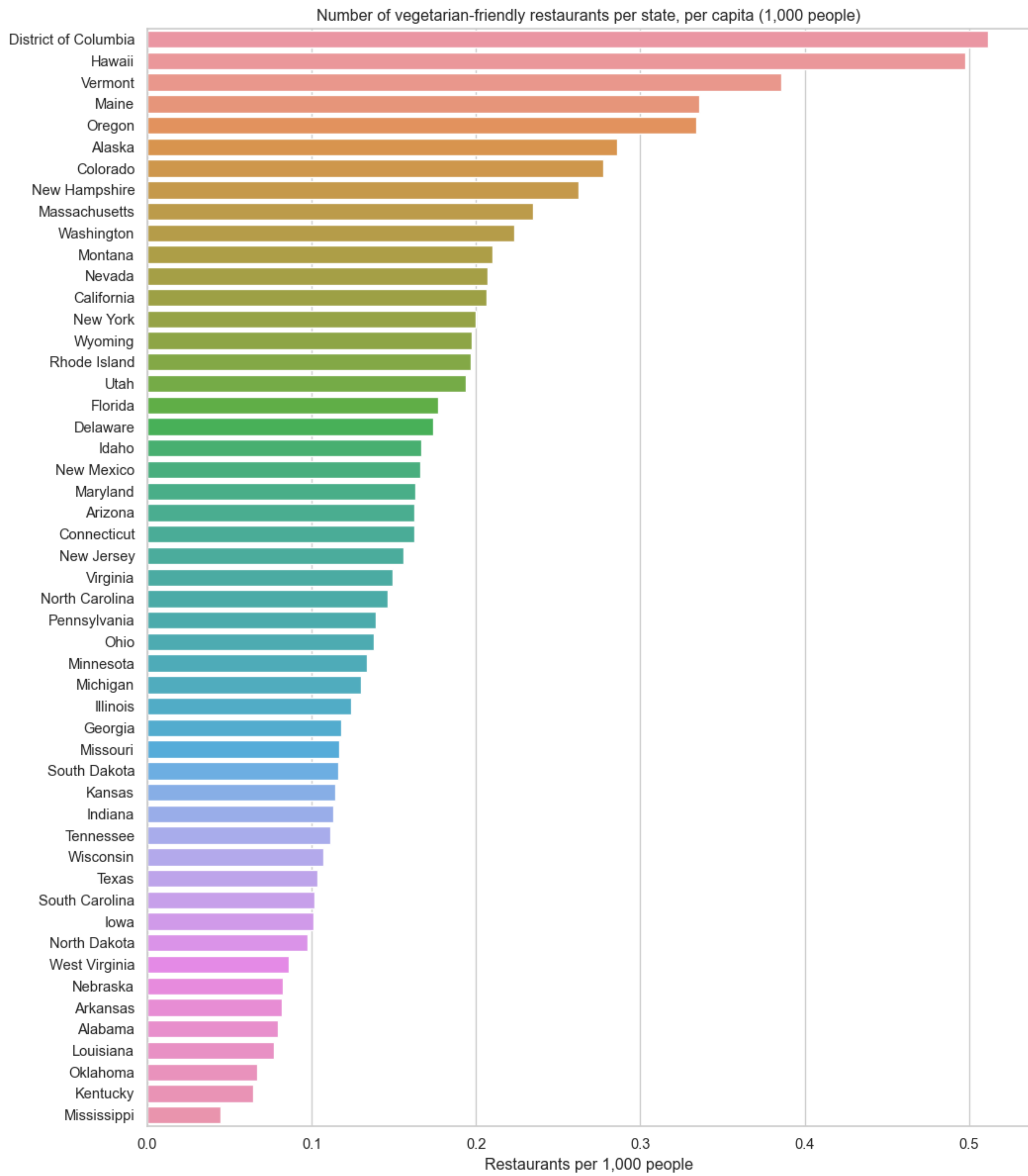
```
##            state  num_of_restaurants
## 50    California                8137
## 49      New York                3904
## 48       Florida                3764
## 47         Texas                2965
## 46  Pennsylvania                1776
```

Once again, many of the highest-ranking states, such as New York, California, Florida, and Texas, are those with the highest populations. Per-capita calculations follow.

```
# Convert vegetarian data to a series so we can perform arithmetic with the populatio
veggie_sorted = veggiefood.sort_values(['state'])
veggie_sorted = pd.Series(data=list(veggie_sorted['num_of_restaurants']), index=veggie

veg_per_capita = veggie_sorted.divide(populations) * 1000
veg_per_capita = veg_per_capita.sort_index(0, False)
```

```
veg_per_capita = veg_per_capita.sort_values(0, False)
```

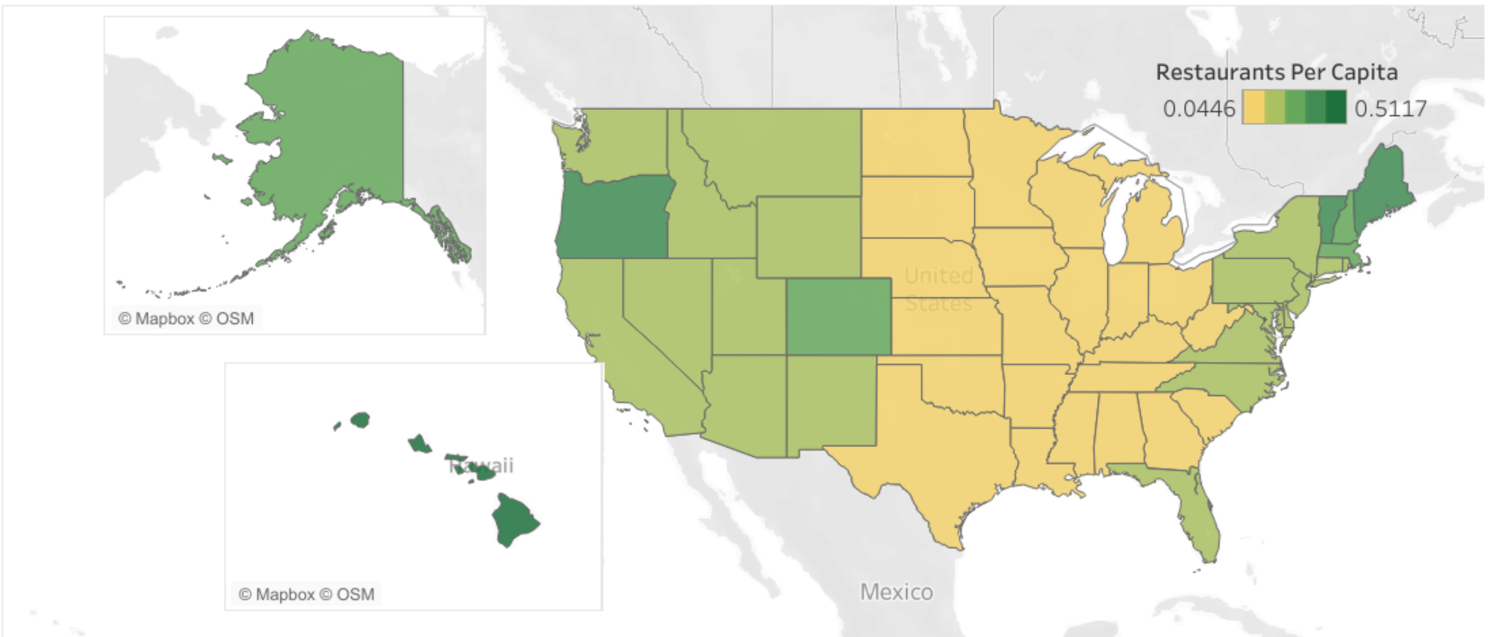
```
fig, ax = plt.subplots(figsize=(12, 15))
graph = sns.barplot(x=veg_per_capita.values, y=veg_per_capita.index, orient='h', ax=a
graph.set(xlabel = 'Restaurants per 1,000 people',
          ylabel= 'State',
          title='Number of vegetarian-friendly restaurants per state, per capita (1,0
```



Disregarding Washington DC, the state with the highest per-capita amount of vegetarian-friendly restaurants is Hawaii, with one such business per 2,000 people. Mississippi, in last place, has a dearth of vegetarian options with only one to every 22,200 people. However, it's important to note that HappyCow's data is user-provided; it's entirely possible that states such as OK, KY, and MS simply may not have many people who know about their website, and thus few submissions.

Before testing these figures against several demographic variables, we can also see a pattern in an initial geographic map, as shown below.

Number of vegetarian-friendly restaurants per 1,000 people



It's now clear that eating out as a vegetarian or vegan is not as easy in the American Midwest than as it is on either coast. However, it's worth exploring just what characteristics of different states, such as wealth and political beliefs, may be contributing the most to this contrast.

Exploring the relationship between the number of vegetarian restaurants and demographic variables

Variable 1: Political Party

```
# Load in the data containing political party identification figures per state
party = pd.read_excel("C:/Users/Noah/Coding/food_study/data/clean_data/state_party_and

party.head()
```

```
##          State  Democrat  Republican      Classification
## 0      Alabama        35          52  Strong Republican
## 1       Alaska        33          51  Strong Republican
## 2      Arizona        41          41         Competitive
```

```
## 3    Arkansas        35          48  Strong Republican
## 4  California        51          31  Strong Democratic
```

```
# Calculate the percentage of democratic identification per state
partyTotal = pd.Series(party['Democrat'] + party['Republican'])
partyTotal.index = party['State']
dem = party['Democrat']
dem.index = party['State']
```

```
demPerc = dem.divide(partyTotal)
demPerc.head()
```

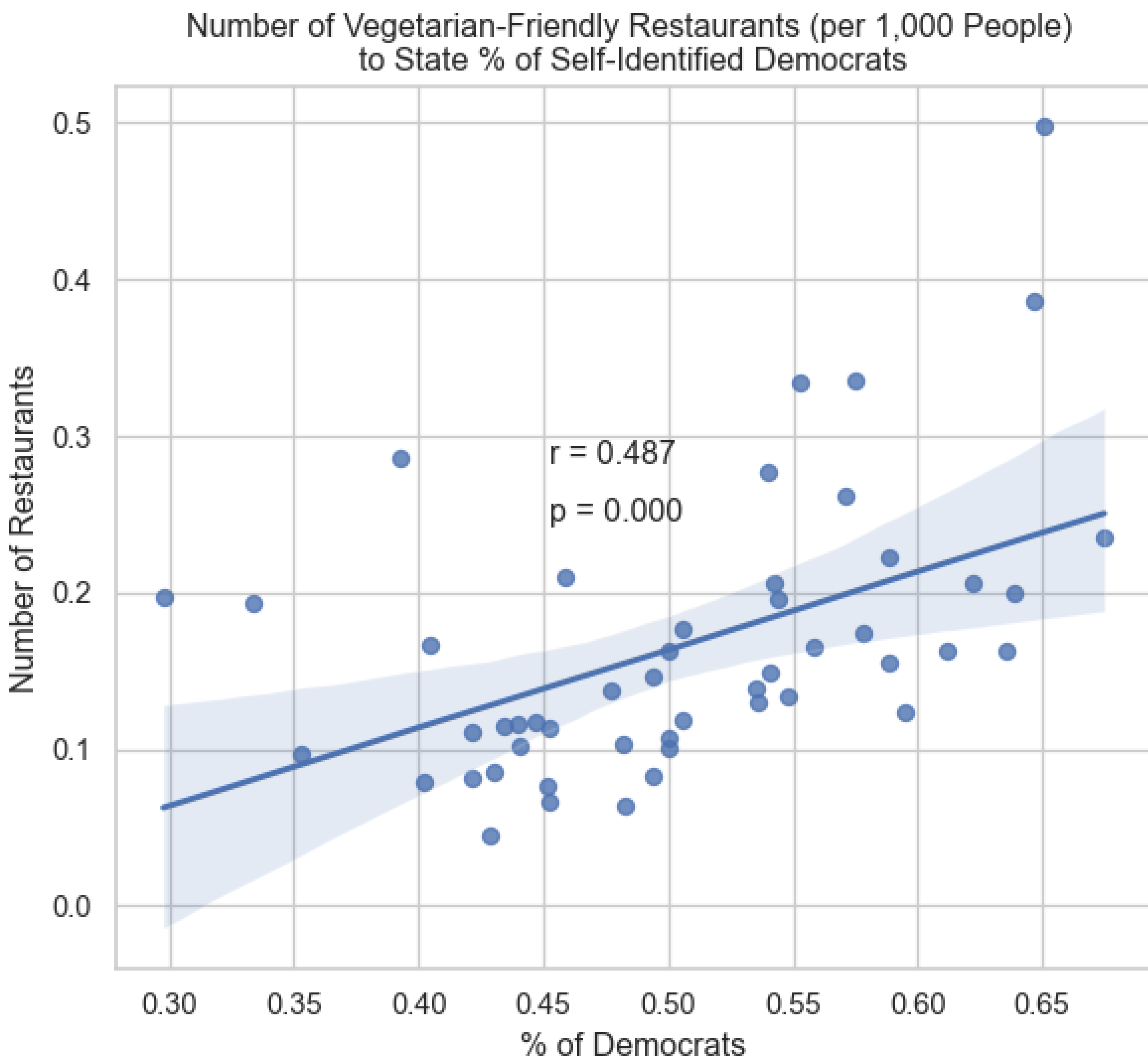
```
## State
## Alabama       0.402299
## Alaska        0.392857
## Arizona       0.500000
## Arkansas      0.421687
## California    0.621951
## dtype: float64
```

```
# Removing DC from vegetarian data, since it's not in political data
veg2 = veg_per_capita.drop(labels='District of Columbia',
        axis=0, inplace=False)
veg2 = veg2.sort_index(0,False)

# Merging the data to later plot it
```

```
partyFull = pd.DataFrame({'restaurants_per_capita': veg2.values, 'percent_democrat': d
        index=veg2.index)
partyFull.head()
```

```
# create scatterplot with the two variables of interest
fig, ax = plt.subplots(figsize=(7, 6))
sns.regplot(data=partyFull,
    x='percent_democrat',
    y='restaurants_per_capita').set(
        xlabel='% of Democrats',
        ylabel='Number of Restaurants',
        title='Number of Vegetarian-Friendly Restaurants (per 1,000 People) \nto State
```

```
# calculate pearson correlation coefficient
r, p = pearsonr(x=partyFull['percent_democrat'], y=partyFull['restaurants_per_capita']
# include values as annotations
ax.annotate('r = {:.3f}'.format(r), xy=(0.45, 0.55), xycoords='figure fraction')
ax.annotate('p = {:.3f}'.format(p), xy=(0.45, 0.50), xycoords='figure fraction')

plt.show()
```



The relationship between political party and the number of vegetarian-friendly restaurants has a low-to-moderate *r* of 0.487. With its nearly-zero p-value indicating the statistical significance of this result, we can confidently conclude that there is a weak relationship between these two variables, and that we can disregard political party as a factor.

Variable 2: Race

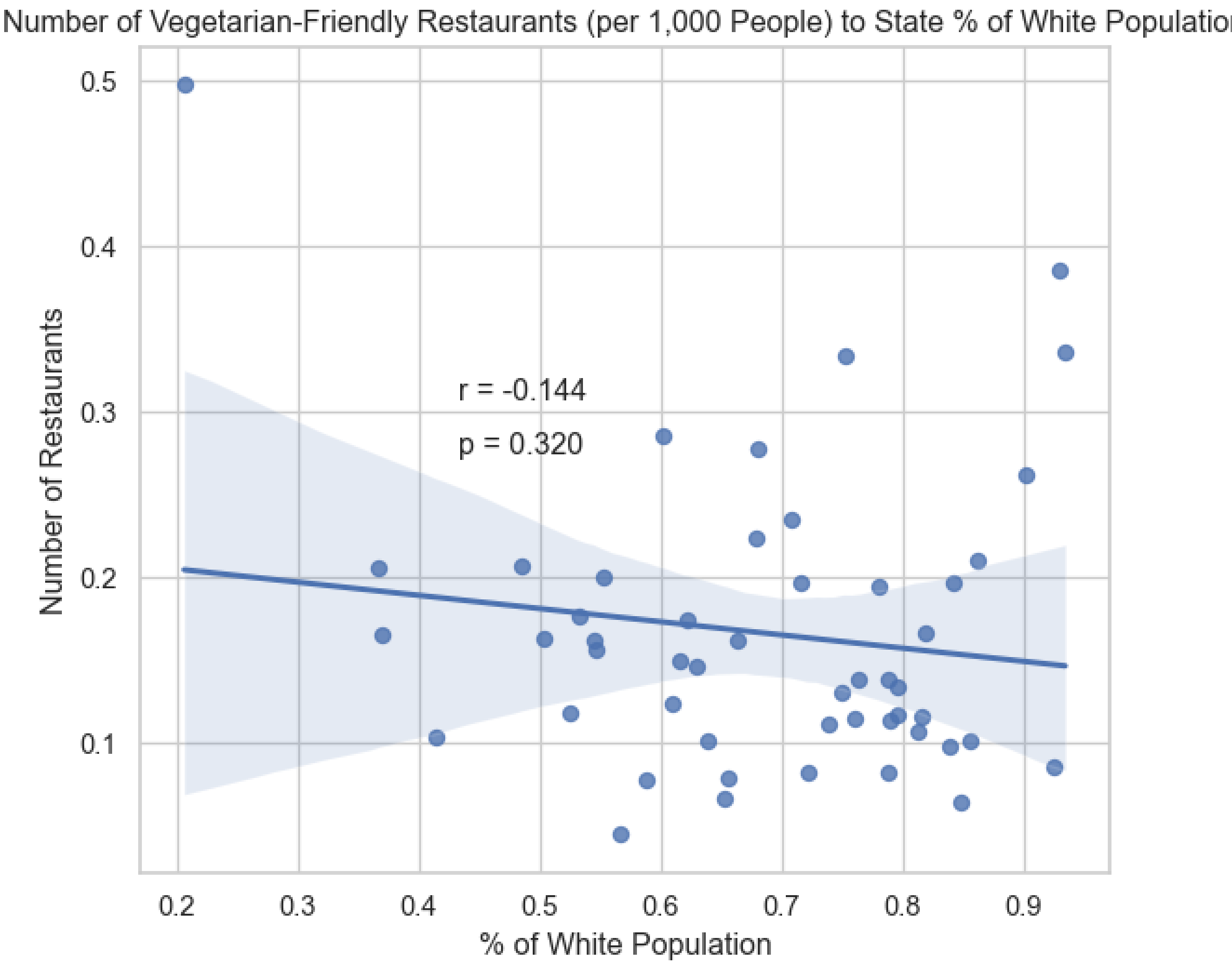
```python
race = pd.read_csv("C:/Users/Noah/Coding/food_study/data/clean_data/race_by_state_2018
race = race.rename(columns={'Location': 'state'})
race.head()
```

```
##         state  White  ... Multiple Races  Non-white
## 0     Alabama  0.656  ...          0.019      0.340
## 1      Alaska  0.601  ...          0.072      0.400
## 2     Arizona  0.545  ...          0.024      0.453
## 3    Arkansas  0.722  ...          0.027      0.279
## 4  California  0.367  ...          0.033      0.627
##
## [5 rows x 9 columns]
```

```python
# combine restaurant and race data
raceFull = pd.DataFrame({'restaurants_per_capita': veg2.values, 'percent_white': race
raceFull.index = veg2.index
raceFull.head()
```

```
##               restaurants_per_capita  percent_white
## state
## Alabama                     0.079383          0.656
## Alaska                      0.285660          0.601
## Arizona                     0.162475          0.545
## Arkansas                    0.081735          0.722
## California                  0.206201          0.367
```

```python
fig, ax = plt.subplots(figsize=(7, 6))
sns.regplot(data=raceFull,
    x='percent_white',
    y='restaurants_per_capita').set(
        xlabel='% of White Population',
        ylabel='Number of Restaurants',
        title='Number of Vegetarian-Friendly Restaurants (per 1,000 People) to State %

# calculate pearson correlation coefficient
r, p = pearsonr(x=raceFull['percent_white'], y=raceFull['restaurants_per_capita'])
# include values as annotations
ax.annotate('r = {:.3f}'.format(r), xy=(0.38, 0.55), xycoords='figure fraction')
ax.annotate('p = {:.3f}'.format(p), xy=(0.38, 0.50), xycoords='figure fraction')

plt.show()
```



There is a negative, albeit nearly non-existent, relationship between a state's white population percentage and its number of vegetarian-friendly restaurants. However, the rather high p-value suggests that there is not much of a linear relationship here at all. Much of the negative nature of *r* is due to data from Hawaii, a state which is mostly composed of people of color and is also the leading state in terms of accommodating vegetarians and vegans. Compared to traditional American food, vegetarian dishes are much easier to find in the cuisine of various Asian cultures; with such a high Asian population (38%), this may be what's at play in Hawaii. Aside from that, race seems to be insignificant.

Variable 3: Income

```
# load data for each state's average yearly income
income = pd.read_excel("C:/Users/Noah/Coding/food_study/data/clean_data/state_income_2
income.columns = ['state', 'income']
income.head()
```

```
##         state  income
## 0     Alabama   49936
## 1      Alaska   68734
## 2     Arizona   62283
## 3    Arkansas   49781
## 4  California   70489
```

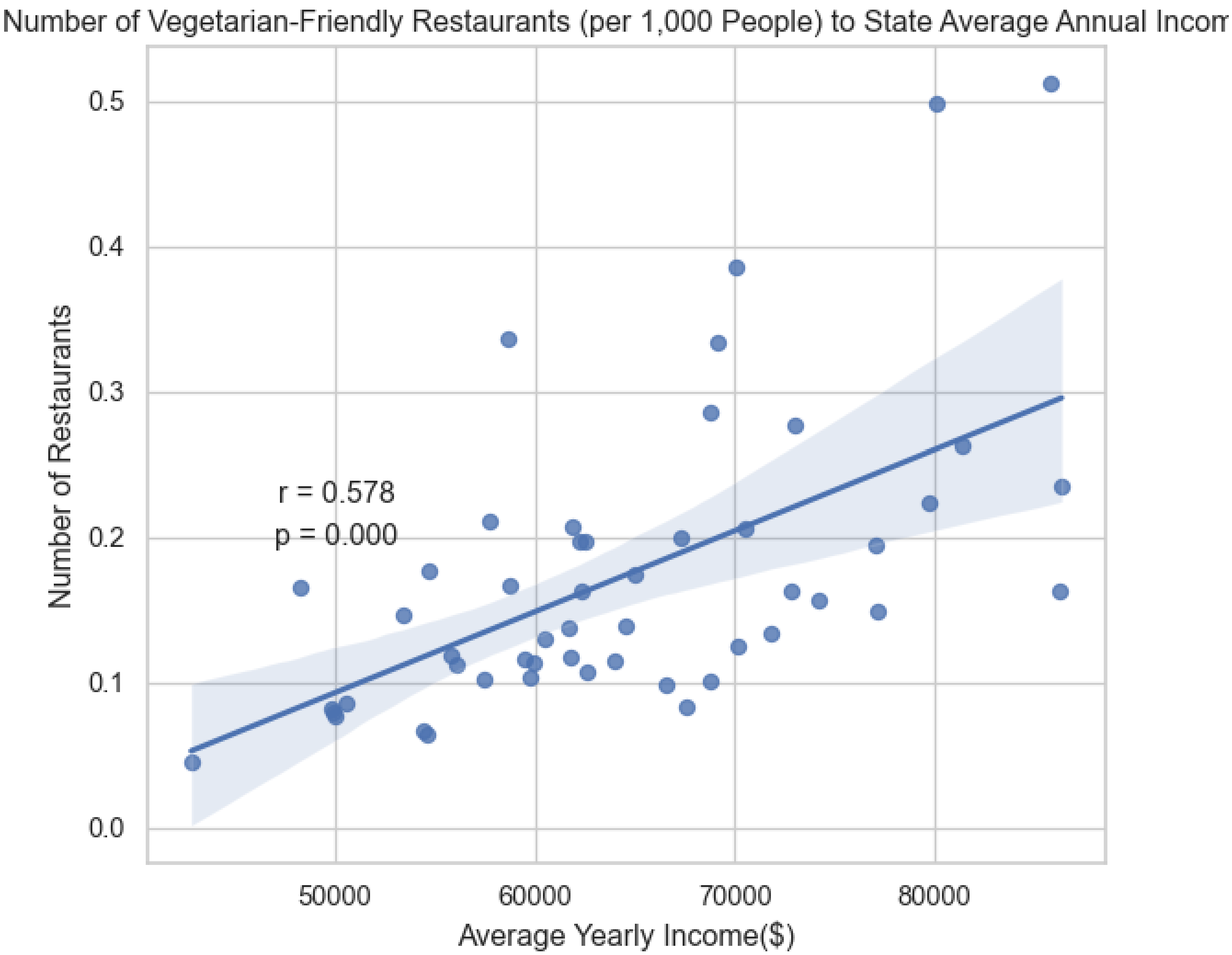
```
# combine data
veg_per_capita = veg_per_capita.sort_index(0, False)
    # use the data that has D.C. in it, since the income data does too
```

```
## <string>:1: FutureWarning: In a future version of pandas all arguments of Series.so
```

```
incomeFull = pd.DataFrame({'num_of_restaurants': veg_per_capita.values, 'income': inco
incomeFull.index = veg_per_capita.index
incomeFull.head()
```

```
##             num_of_restaurants  income
## state
## Alabama               0.079383   49936
## Alaska                0.285660   68734
## Arizona               0.162475   62283
## Arkansas              0.081735   49781
## California            0.206201   70489
```

```
fig, ax = plt.subplots(figsize=(7, 6))
sns.regplot(data=incomeFull,
    x='income',
    y='num_of_restaurants').set(
        xlabel='Average Yearly Income($)',
        ylabel='Number of Restaurants',
        title='Number of Vegetarian-Friendly Restaurants (per 1,000 People) to State A

# calculate pearson correlation coefficient
r, p = pearsonr(x=incomeFull['income'], y=incomeFull['num_of_restaurants'])
# include value as annotations
ax.annotate('r = {:.3f}'.format(r), xy=(0.23, 0.45), xycoords='figure fraction')
ax.annotate('p = {:.3f}'.format(p), xy=(0.227, 0.41), xycoords='figure fraction')

plt.show()
```



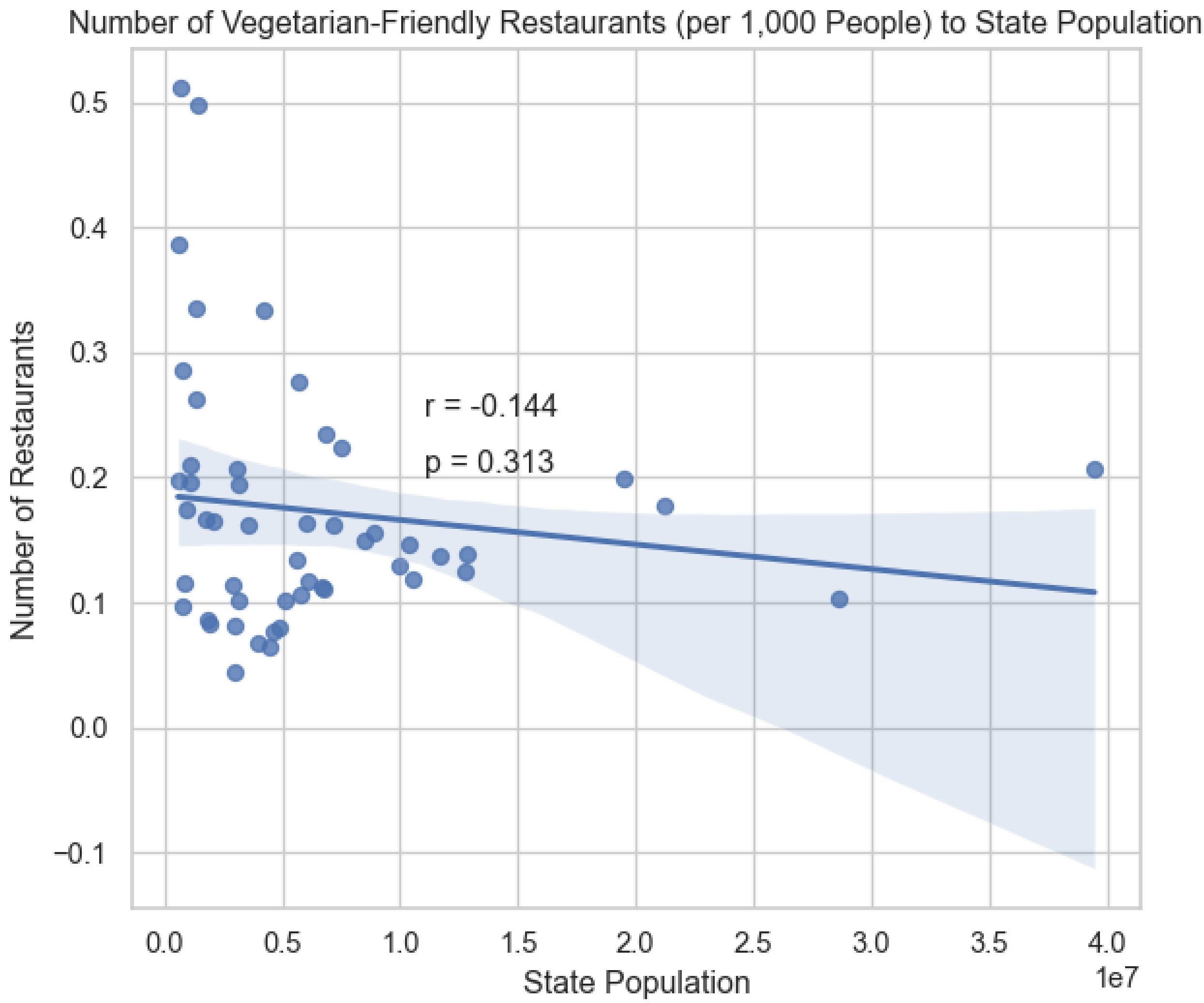
(Note: This 2018 income data has not been adjusted for inflation.)
Here we see the strongest relationship of all: an *r* of 0.578, meaning there is a moderate relationship between the average yearly income and number of vegetarian-friendly restaurants of a given state. The very low p-value of nearly 0 indicates that these results are indeed significant. There are many sub-variables resulting from having a higher income that must be contributing to this. For example, a guess as to why fewer people with a lower income are likely to become vegetarians in the first place would be that working class people have less time to cook their own food; a single, working mother may find it much more difficult than I did to take the time to learn new, vegetarian recipes for her picky children. Additionally, those with less income may have less access to pricier grocery stores that provide more vegetarian ingredients and alternatives, such as Sprouts and Whole Foods. If an area has fewer vegetarians, it would make sense that the lower demand would result in fewer vegetarian restaurants popping up.

Variable 4: Population

```
popFull = pd.DataFrame({'restaurants_per_capita': veg_per_capita.values, 'population'
# incomeFull.index = veg_per_capita.index
popFull.head()
```

```
##              restaurants_per_capita  population
## name
## Alabama                    0.079383     4887681
## Alaska                     0.285660      735139
## Arizona                    0.162475     7158024
## Arkansas                   0.081735     3009733
## California                 0.206201    39461588
```

```
fig, ax = plt.subplots(figsize=(7, 6))
sns.regplot(data=popFull,
    x='population',
    y='restaurants_per_capita').set(
        xlabel='State Population',
        ylabel='Number of Restaurants',
        title='Number of Vegetarian-Friendly Restaurants (per 1,000 People) to State P

# calculate pearson correlation coefficient
r, p = pearsonr(x=popFull['population'], y=popFull['restaurants_per_capita'])
# include value as annotations
ax.annotate('r = {:.3f}'.format(r), xy=(0.35, 0.55), xycoords='figure fraction')
ax.annotate('p = {:.3f}'.format(p), xy=(0.35, 0.50), xycoords='figure fraction')

plt.show()
```



The correlation between a state's population and its per-capita number of vegetarian-friendly restaurants is nearly nonexistent. This, combined with the p-value of 0.313, indicates that this variable can be disregarded.

Conclusions

The only one of these tests that suggested a relationship between a demographic variable and the number of restaurants with vegetarian/vegan restaurants in a state was that which was run on state average income, confirming one of my hypotheses. If one wants to live in an area that provides a relative abundance of plant-based and meat-free options, they'd be best off relocating to a more financially affluent state. The problematic nature of this conclusion is obvious – maintaining a vegetarian or vegan diet will be easier for those who can actually afford to live in these expensive areas.

My other hypothesis, however – that states with higher populations would have more restaurants with vegetarian/vegan options – was proven incorrect. That said, I think this ties into a major weakness in this study, which is just how much variables like income and population can vary in one state from city to city. Thus I believe a similar project done exactly the same but with data from counties or cities instead of states could have provided more illuminating insights.

Another major caveat is the fact that the webscraped HappyCow data wasn't collected in the same year as the other datasets – This data includes, for example, many fast food restaurants which have only introduced vegetarian/vegan options in recent years since the other data was collected. This is a trend which will make it much easier for vegetarians and vegans to eat out across the country; but, with these restaurants only providing often unhealthy fast food, there are still improvements to be made.

It's difficult to say exactly what the solution to these differences in access based on class is. Whatever actions one feels should be taken, such as education and resources on how to cook vegetarian dishes (especially for owners of restaurants), any move towards accessibility of these cuisines will undoubtedly be a net positive – both for the health of our country's people and our environment.