The Heart Disease Pandemic and its correlation to Convenience Restaurant Density

Jeffrey Passaro

**Introduction:**

In the United States in 2020, 375,000 people died of COVID19 complications. Some will say that our government has not done enough to fight the disease, and some will say that our government has been too overreaching in its fight against COVID19. From my perspective, I believe our collective reaction to fighting COVID19 has been astounding, and when called to limit our mobility and protect our neighbors we responded. Often, we did not agree with the requirements, but we followed guidance. Although we have not overcome COVID19, in July 2021 the US is in a much better place, relative to COVID19, then we were a year ago.

In the United States in 2020, 690,000 people died of heart disease. Between 2010 and 2019, 6.2 million people in the United States died of heart disease. That is over 1.5 times the number of COVID19 deaths in 2020 on average for the past 10 years. The number of people dying from heart disease is staggering, and as most of us emerge from the only pandemic of our lifetime, I encourage us to consider that we have been living in a heart disease pandemic and a cancer pandemic for the past century but simply have not defined it that way.

Between 2000 and 2012, the death rate from heart disease decreased 24%. However, since 2012, the rate of heart disease deaths has increased 5%. Large metropolitan areas fared better with an increase of 1%, while the rest of the US has seen an increase of 7%.

While there are likely many reasons for an increase in heart disease deaths, this analysis explores convenience restaurants and their potential correlation to heart disease deaths. Healthy eating is a message that we have understood for some time. Often it is more convenient to stop for a meal, then to cook healthy at home. This Analysis explores the correlation between the density of convenience restaurants and the death rate from heart disease.

**Data Sources:**

To support the final correlation analysis, access to the Center for Disease Control and Prevention cause of death databased, Wonder, was obtained and queried for deaths related to heart disease. This data was evaluated for the period 2010 to 2019 and was aggregated for stability in areas with small populations. The CDC provides this information at the county FIPS level in order to understand how death rates are distributed in the United States.

The count of convenience restaurants was obtained by passing the county centroid latitude and longitude coordinates to the Foursquare API to search specific venue categories. This was completed for each county centroid in the United States and the results were compiled in a Pandas dataframe. County centroid latitude and longitude coordinates were obtained from a Wikipedia page titled “Table of United States counties.” The python package Beautiful Soup was used to parse the Wikipedia table and create a dataframe from which to pass Foursquare the required coordinates. The approach looped through Foursquare to obtain a list of venues that are within twenty miles of each county centroid. This dataframe was then used to create a count of convenience restaurants in each county. That count is referred to as the convenience food density for each county.

To better familiarize the audience with the concept of convenience food restaurant venues I have provided a sample of the top ten categories and venues. Most will gain a greater understanding of the analysis by reviewing the following lists.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Top 10 Venue Categories | |  | Top 10 Venues | |
| Venue Category | Count |  | Venue Name | Count |
| FAST FOOD RESTAURANT | 65,293 |  | SUBWAY | 16,966 |
| PIZZA PLACE | 49,115 |  | MCDONALD'S | 6,533 |
| BURGER JOINT | 12,699 |  | TACO BELL | 4,434 |
| DINER | 12,318 |  | DAIRY QUEEN | 4,372 |
| FRIED CHICKEN JOINT | 8,827 |  | PIZZA HUT | 4,260 |
| DONUT SHOP | 7,234 |  | BURGER KING | 3,664 |
| FOOD TRUCK | 2,673 |  | DOMINO'S PIZZA | 3,660 |
| BAGEL SHOP | 2,174 |  | DUNKIN' | 3,642 |
| FOOD COURT | 1,776 |  | WENDY’S | 3,349 |
| HOT DOG JOINT | 1,664 |  | ARBY'S | 3,122 |

These lists contain many of my favorite food choices, providing even more reason to explore the relationship between these kinds of restaurants and heart disease deaths.

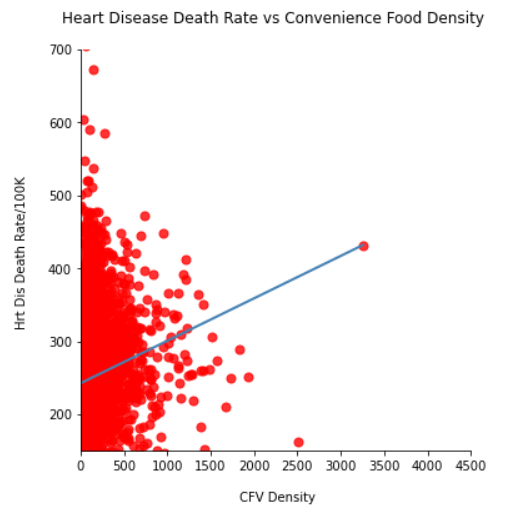
**Methodology:**

After accessing the required data sources, the heart disease data was explored for the depth of variation between counties. The heart disease death rate varied from 35 to 705 heart disease deaths per 100,000 people in the county. There are over 3,000 counties in the US with data, and the variation is wide, so the data provides a robust set of information to work with.

The Foursquare venue data was then explored, and it showed that there were several counties where the venue count stopped at one hundred. The max limit on venues to bring back using the Foursquare API is one hundred and while that suppresses the top end of the analysis, the correlation concept can still be evaluated using all the data in the data set. Convenience food density ranges from zero to 3,256.

After joining the heart disease death rate data to the venue data, an evaluation of the variation between the two variables occurred. The initial scatter plot using convenience food density on the x axis and heart disease death rate on the y axis showed data concentrated at the lower range of the convenience food density axis, while becoming much more dispersed with larger density.

*Figure 1*



Further exploration of the data showed there was a significant outlier. The outlier in in this case is Robertson County Kentucky with 9 Subways, 5 McDonalds and 3 Taco Bells among the seventy-one convenience food venues for a population of 2,100 people. Robertson County Kentucky has a heart disease death rate of 431.2 per 100,000 people, 120% higher than the heart disease death rate in the United States. Robertson County also has a poverty rate 73% higher than the US average and 22% of its residents did not graduate from high school compared to a US average of 13%. This suggests further research into the correlation of these measures of social determinants of health.

The visually low correlation between convenience restaurant density and heart disease death rate from the scatter plot above is inconsistent with conventional wisdom. Therefore, the data was supplemented by calculating twenty equal bins for convenience food venue density based on rank. The data was then grouped into larger bin segments to stabilize the variability between counties.

Figure 2 summarizes the new set of data by bins.

*Figure 2*

Table

Description automatically generated

After grouping the data by bin, the data was explored to observe how the heart disease death rate varied by convenience food density. This observation was simply to examine the counties to assess if anything seemed initially inconsistent. Figure 3 shows the Top 10 Counties in the top bin of convenience food density. For comparison, the heart disease death rate per 100,000 people in the US is 195 and the convenience food density in the US is 18 per 100,000 people. You will find that the convenience food density of this bin is extremely high as is the heart disease death rate.

Figure

Table

Description automatically generated

In contrast, Figure 4 shows bin one of convenience food restaurant density sorted by heart disease death rate. There is also consistency in this cohort, and it is quite different than then top bin cohort.

Figure

Table

Description automatically generated with medium confidence

**Results:**

Once the heart disease death rate and convenience food density data were summarized by bin a scatter plot was produced to visualize the correlation between the two measures. Figure 5 shows the visualization of the correlation between the variables and suggests a moderately strong correlation.

*Figure 5*

Chart, scatter chart

Description automatically generated

The plot of the convenience food density bin compared to the heart disease death rate clearly shows a visual relationship. This should be interpreted, in general, as the higher the density of convenience food restaurants, the higher the heart disease death rate. Since we grouped counties by bin to normalize the data, this relationship will not hold precisely by county, however it does shows a general correlation between the two metrics.

The Pearson correlation coefficient was calculated using the SciPy package and showed a correlation of .65 with a p-value of .002, showing a statistically significant, moderately strong correlation between convenience food restaurant density and heart disease death rate.

Pearson Correlation Coefficient: .65

p-value: .002

While the correlation is clear based on the scatter plot above and the Pearson correlation coefficient, further evidence of the relationship is depicted in Figure 6 and Figure 7, a mapping of convenience food venue density and heart disease death rates.

Figure

Map

Description automatically generated

Figure

Map

Description automatically generated

This map comparison shows that heart disease deaths are in fact higher in areas with a higher density of convenience food restaurants, validating the statistically significant Pearson correlation of .65.

**Discussion:**

This analysis confirms that access to convenience, typically less healthy, restaurants is correlated with heart disease death. Correlation is not causation, and these results should be considered with that in mind.

As an individual one may choose to live in an area with a lower density convenience food rate. The notion is that if there is less access to convenience food restaurants then the likelihood of dying of heart disease is lower. There are many reasons to select a place to live. However, if you have a family and are concerned about their health, one of the many things to consider is the density of convenience food restaurants. I enjoy convenience food, especially pizza, but I recently moved to a location where I cannot find a pizza restaurant (not for lack of trying), and I have not had pizza since my move. There is a practicality to what I am suggesting that some will oppose and that is expected. This proposal is a rational way to consider data driven results in your choices.

Government regulation on these types of venues is also something to consider. While government most often regulates things and not establishments, there is a precedent, ironically, in the health care certificate of need. Certificate of need laws are present in most states that require health care organization to demonstrate community need before building a new health care facility. Perhaps regulating restaurants along similar (different in content) lines could lead to a sustained overall lower heart disease death rate.

**Conclusion:**

Understanding the correlation between convenience food restaurant density and heart disease deaths enables both individuals and policymakers to make analytic informed decisions. As more health care organizations begin to understand how social determinants of health impact their served population, it is critical to apply structured analytics to make informed decisions. This analysis provides a practical example of how someone’s environment impacts their health. Each applicable social determinant measure and health challenge require a unique approach to inform a rationale decision. This is a worthwhile endeavor as these are genuine issues with significant health impact for our neighbors.

Simplifying robust analysis enables sound decisions that many will respond to. We learned this with the COVID19 pandemic. Our leaders provided us with simple Covid statistics that made an impact. In 2020, most of us wore masks, washed our hands frequently and socially distanced. In 2021, it is time to use analytics to understand the correlation between our environment and ongoing pandemics like heart disease.

**References**:

2020 CDC Death rates: https://www.cdc.gov/mmwr/volumes/70/wr/mm7014e1.htm

CDC Heart Disease Death Rate: https://wonder.cdc.gov/controller/datarequest/D76;jsessionid=83F854F8AA809E9EA32BD53A633B

County Coordinates Wikipedia Page: https://en.wikipedia.org/wiki/User:Michael\_J/County\_table

CDC Social Vulnerability Index: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

Foursquare Venue Collection: https://foursquare.com/