Data Visualization and Communication 22103 Group 5

University of Arkansas – Data Science

December 10, 2024

Executive Summary

Introduction:

My group has spent so much time working on our game changing dataset analysis, where we have spent hours working on technical data, analyzations through calculated fields, using Tableau's TabPy coding function to use python scripts, and more. Our dataset covers the basis of the leading causes of death over the course from 1999-2017 in the U.S. Our purpose was to try and find any leading correlations with different causes of death; such as smoking, obesity levels, heart disease, strokes, and many other causes of leading deaths. To try and see if we could really affect the number of deaths by seeing if they have a high correlation with certain causes. For instance, if we found out that having high cholesterol levels leads to more causes of death, we would focus more on healthcare and spending for that specific cause, and so forth. Our final vision was to see if we could show visual analysis to see how we could change the way causes of deaths are being set up today. Our datasets include information about every leading cause of death in the U.S. and breaking off into that we found datasets covering obesity and smoking, then to take a turn from causes of death, we also took in datasets revolving around healthcare analytics, Medicaid and Medicare to be specific. All datasets used were NOT from Kaggle. We feel as though this kind of data is very crucial to the way everybody lives their lives. Obviously, everyone wants to live for as long as possible, so we felt this would be a topic anybody could be interested in.

Objectives:

We had a set variety of goals in the beginning of this group project. But once we learned exactly what our data was saying after analyzing it. We decided we wanted to express if we could see the correlation between lifestyle habits and cause of death. Our first bit was to see if smoking has any effect on cancer because that was a leading cause of death, same as heart disease caused from obesity. We then decided instead of computing 4 different causes of death, we decided to spread out our data and use healthcare data to see if that has any leading correlation between deaths. All this to better prepare for the future with our analyzation of a longer living life! We aimed to achieve finding at least 2 good correlations between cause of death and healthcare.

Methodology:

To start, with any project we needed to find some questions about what we wanted to do or where to start. We ended up picking our topic due to the seriousness and importance of the concept. Once we found the topic, we each split up to do different takeaways from the analysis. After having spent some time looking for the correct data, we had to aggregate many of them into 2 and had to cross reference similar statistics to show accurate data.

We all split up into parts for this part to find datasets we thought would back our stories up. Once we aggregated our data and we have a beautiful looking dataset, we then got to work. Everybody picked out their topic in mind, one was obesity rates leading to heart disease, one was smoking statistics leading to lung cancer, one death with Medicaid data (which is healthcare benefits for people in poverty), while another was dealing with Medicare (which is the healthcare benefits for elderly people).

To fully analyze specific subjects, such as calculated fields or parameters, we had to spend some time in the terminal and python using the special Tableau function TabPy, which helped us create more dynamic looking variables to stand out in our dashboards. I know for the obesity and smoking datasets we had to go deep into the analyzation to find specific things we were experimenting for. That took quite a bit of our time, as well as creating our visuals to define a good and well thought out story.

Key Findings:

After having spent much time looking at our datasets and creating visual analyzations, each of the 4 parts of this project had a takeaway that they are proud to share. We each had found correlations between our topics, as well as some interesting finds along the way!

Smoking:

Cigarette smoking rates have decreased significantly in the majority of states and Washington DC (from 2011 to 2012). Although, even on the decline, the smoking rates for certain states like Kentuckey and West Virginia remain high. The smoking rate of all states correlate with the age adjusted death rate of lung disease for each state the most, with an r value of 0.8379. The age adjusted death rate for cancer also correlates well with the smoking rate, with an r value of 0.7575. This means that states with a higher smoking rate are much more likely to have lung disease and/or cancer related deaths compared to lower smoking rate states. In order to prevent smoking rates from increasing (and therefore reduce lung disease and cancer rates), campaigns should target how lung disease and cancer fatalities are more likely to be found in states which have more cigarette smokers.

Obesity:

Obesity is correlated very highly with many causes of death, the main ones being cancer, heart disease, stroke, and diabetes. These causes account for a very large portion of deaths in the United States. Heart disease, cancer, and stroke have been the three leading causes of deaths. The combination of the clear link with these causes, the proportion of deaths that come from these causes, and the rapid rise in obesity is extremely alarming.

Obesity has been on the rapid rise across every state in the US. The South and Midwest are the two regions hit the hardest, but every state has seen an increase. These regions are some of the least wealthy areas of the US, which leaves them prone to food deserts. Food deserts are areas with no close source of healthy, reliable, and cheap food. This leaves many people to get their food from convenience stores and fast food establishments. This lack of access to healthy food is key in understanding the rise in obesity. In order to address this, policymakers should consider addressing policy to target the growing number of food deserts in impoverished areas.

Medicaid:

To start off with clear clarification, Medicaid is a governmental ran program that helps Americans in lower poverty income ranges. Anywhere typically under the age of 65 is when people of lower income get indicted into Medicaid if they need it. A few important things we found looking through this dataset was that Medicaid in fact does not have a strong correlation to causes of death. Causes such as heart disease or cancer, there isn't necessarily a way to prevent those kinds of things rather than the flu. Not only does Medicaid have a low correlation to causes of death, but it also has a low correlation with decreasing deaths. Meaning, there isn't a proven statistic that shows Medicaid helps save lives yet. Biggest takeaway we found for this concept was that although low correlation linkage between Medicaid, it still is critical when focusing on preventable deaths rather than non-affected causes of death. A beneficiary takeaway we found for this topic was Medicaid does have a very strong correlation to decreasing stroke deaths.

Medicare:

On the flip side of Medicaid, there's always Medicare. This is the government ran program that does the exact same thing as Medicaid, except it's for Americans tending the age of 65 or older. After looking at different data from Medicaid, Medicare shares somewhat the same descriptions, although there is a big difference as well. The growth rate of Americans enrolling into Medicare have presumably no set pattern. It seems to increase at specific times on our interval, and other times it drops tremendously. After having analyzed why this was, it seems that along with getting old and signing up for healthcare, nothing really has an everlasting effect on the growth rate of enrollees using Medicare. Some weird but interesting takeaways from this topic was that any kind of policy or government act that deals with healthcare affects the growth rate of enrollment for that year. Although growth rate increases and decreases, its annual spending per user is still expensive overall for healthcare. We also took a turn and wanted to see geographical data to see where and maybe why causes of deaths have bigger rates in different areas. We then noticed that on the West coast there's more heart disease rates and obesity levels, compared to the East coast having more of a cancerous environment. These possible outcomes could be due to environmental change, healthcare access, or even personal lifestyle. Overall, Medicare has no correlation between causes of death due to the fact Medicare users are elder or old. Meaning they don't have much to live for, meaning much easier to pass away from old age than anything. Medicare does have a high correlation of populated areas, which is common knowledge almost. But the goal for reaching high correlation of Medicare to cause of death is very poor or low.

Recommendations/Conclusion:

We feel as though there are several things that could be done to help prevent causes of death in the U.S. For smoking, we can clearly say that the less people smoke or vape, the longer happy lasting life they will have. This isn't always true, but from our analyzation that is a correct statement. Also to try and decrease smoking levels, to try and push for more anti-smoking campaigns and startups to take over and ultimately decrease smoking rates.

For obesity, rates are getting more and more out of hand when it comes to the United States. Studies have shown that the U.S. is one of the leading countries in overweight/obese populations. An easy recommendation from us, is to not necessarily "watch" your habits but seek help if it becomes a serious problem. Obviously, obesity isn't always a personal factor, due to that, we say that there's always help there to seek weight loss or weight sustainment.

For enrollees that use the Medicaid system, it isn’t “useless” when it comes to cause of death. Only for underlying causes that have no prevention, for instance cancer. Being enrolled and getting benefits does help tremendously with less severe injuries or diseases. Ultimately, we recommend continuing to use Medicaid. Our analysis doesn’t show a bad outcome for any user who enrolls in Medicaid, just a slower outcome reached.

For Medicare enrollees, keep doing what you are doing. Although there is no correlation between causes of death and Medicare, that doesn’t mean it's bad for you. It just means that when elder people are on Medicare, there isn’t a link between the death of that person and their healthcare. Medicare is more of a personal use protection plan, you don’t really need it, it really does ultimately lead to longer lasting life.