Darling Judah Hsu

Measurement of Income's effects on Heart Disease and Attacks

Heart disease and attacks are often the result of reduced blood flow for various reasons. Among the reasons, In this inquiry, we'll be focusing more specifically on the (primarily indirect) effects of income on Heart disease and attacks through its impact on mental health and physical health (indicated by high cholesterol).

My research has shown, higher income has a tendency to be correlated to higher risk of "heart disease and attacks" (HD/A). Although it's not as clear in the histograms (figure 1.1) due to the scale, it's much easier to see in the contingency table (table 1.1), that income does seem to have some relationship with HD/A.

Generally, if a factor doesn't have an effect on 2 samples of people you'd expect to see the same ratio of some indicator variable happening for each sample. In this situation, for example, you have 8 income levels (or 8 samples of people, each at a different income) and you wanted to see whether or not income has an effect on heart disease. If income doesn't affect heart disease, then regardless of the amount of people in each sample, you'd see approximately the same "odds," or proportion of people who have heart disease vs people who don't have heart disease at each level of income- that is to say- the ratio of having HD/A is the same regardless of your income. Another way to determine this is to look at the ratio of the number of people in total who have HD/A to those who don't. That proportion is essentially a base ratio; if other variables don't have any influence over having HD/A, then the probability of having HD/A at each income level should be approximately the same as in the total sample. So if you have a sample of 2000 people who had HD/A and 5000 people who didn't then the ratio is 2/5. If income has no effect, you'd expect that if you divided everyone into groups by their income, the

ratio of people who had HD/A in income 1 should be close to the same ratio 2/5 (the base ratio). When looking at contingency tables, this will be our primary method of discernment as to whether a variable is having an effect or not. Furthermore an "odds ratio" is defined as the ratio between the odds of a variable in one sample, over the base ratio. If the odds ratio is bigger, then the ratio in said-sample is larger than the base ratio and vice versa.

Going back to income, the odds ratios, from income levels of 1 (income<$10,000/yr), 2 ($10,000/yr < income < $20,000/yr), [...], 8 ($70,000/yr +) are 1.809, 2.204, 1.798, 1.571, 1.338, 1.068, 0.822, 0.514 respectively (where the odds are HD/No HD per income). As seen in the ratios, although the odds of HD to no HD seem to peak around an income level of 2, the odds of having a HD/A becomes increasingly lower as income levels increase. The peak at income 2 could be attributed to pure chance or other reasons but the focus for now will be on the decreasing trend of the odds ratio as income increases.

Additionally, in figure 1.7, although it's not perfectly obvious, you can see that the *ratio* for people with heart disease to no heart disease in income lv. 1 is greater than in income lv. 8 (due to the large number of values in income level 8). So, the evidence points to the notion that higher levels of income correlate to lower risk of HD/A. In this discourse, we will be inquiring into the various intermediaries from which income correlates to heart disease and attacks.

As for why income seems to correlate to HD/A, my research has shown that mental health, in particular, seems to relate back to HD/A, as the result of income. As can be seen in the histogram (figure 1.2b), although the least amount of people recorded are found in income level 1 (table 1.2), it holds equal to (if not more) counts of people who took 11+ mental health days in a month. Furthermore, in the table (table 1.2) comparing mental health days for people of income levels 1 & 8, the odds of being from income level 1 increases as the amount of mental health

days increases;for people who took 26-30 mental health days, the odds of a person being from income level 1 becomes 0.830 (compared to 0.109, the odds of a person being from income level 1 solely based on the amount of people). The odds ratio is then 7.644, indicating that income does have a huge effect on mental health. Afterall, the lower your income is, the more likely you are to be stressed, worried, and uncertain about meeting your needs. But why does it matter that income leads to mental health? What effect does mental health have on HD/A?

The primary reason why we're concerned about mental health is because mental health actually has a substantial effect on HD/A and whether someone smokes or not. In table 1.7, the odds of having heart disease increases as someone takes more mental health days. Focusing more on the effects of mental health on smoking: in figure 1.3, although there's a greater number of people who don't smoke, most of the people who reported 10+ mental health days happened to smoke. Generally, most people in the data only take 0-10 mental health days but for those who take more, the bars look primarily black (indicating large amounts of smokers). This can also be seen in the contingency table (table 1.3) where the odds of smokers to non-smokers steadily increases with more mental health days, until the odds ratio is approximately 2- the odds of smokers to non-smokers for people who took 25-30 mental health days is 2 times that of smokers to non-smokers in general. Since nicotine in cigarettes calms one's mind, stress (indicated by the number of mental health days taken) often relates to the amount of cigarettes you may smoke. Since smoking increases plaque among other impurities inside the blood vessels, smoking also has great potential for causing heart disease.

When looking at the effect of smoking on HD/A, the number of people with HD/A is *greater* in smokers than non smokers (figure 1.4), despite there being a larger quantity of non-smokers in general. In table 1.4, it's visible that the odds of having heart disease for a

smoker vs not having heart disease is actually 1.5 times the odds of having heart disease vs not having heart disease in general. So by this line of reasoning, we see that one of the ways income causes HD/A can be found in how income affects stress (quantified by number of mental health days taken), which affects one's possibility of smoking and, in turn, their risk of HD/A.

Another way in which income affects HD/A is in the food consumption of people at different levels. Although there isn't sufficient reasoning to say that, if you have more income, you'll buy healthier foods; it is necessary to acknowledge that with less income, you try to save more money. In saving more money, people often turn to eating cost-effective fast foods: some of the most addictive, cheap, and unhealthy foods in existence.

In tables 1.5 and 1.6, there does seem to be some relationship between income to BMI, and then BMI to HD/A. However, the odds ratio for BMI to HD/A is not particularly significant, nor does the histogram in figure 1.6 show a significant ratio between people with HD/A and people who don't, for the different BMIs. So although income does provide sufficient evidence that it relates to BMI, the relationship will not be considered in depth due to BMI's lack of significance in HD/A. There is, however, a noticeable correlation between income and whether or not someone has high cholesterol.

In table 1.8, the odds for people with high cholesterol as opposed to low cholesterol are rather high for people of lower incomes. In fact, aside from the peak at income level of 2, the odds of high cholesterol to low cholesterol only decrease as income level increases. When looking at figure 1.8, the counts for high cholesterol vs low cholesterol are approximately equal for incomes at levels 1-6, despite the proportion of total high cholesterol to low cholesterol being 107,600 to 146,100.

For the final variable, research has shown that there does seem to be an interaction between smoking and high cholesterol, on HD/A. The bar chart (1.10), the amount of people who smoke and have high cholesterol is the least. However, it also has the highest number of people who had HD/A. The odds of people who had HD/A vs. who didn't have HD/A is also 2.5 times that of the amount of people who had HD/A vs. those who didn't in total.

In the research conducted here, the evidence supports the notion that income has a substantial effect on mental health, as well as high cholesterol. In turn, mental health affects how often you might smoke- increasing the risk of heart disease and/or attacks. Additionally, since income seems to have an effect on food consumption, it also increases whether or not someone has high cholesterol. High cholesterol and smoking also seem to interact together to greatly add to one's risk of heart disease and attacks. These are the ways we've found why lower income might lead to higher risk of HD/A, based on the data we've been provided.

Tables

HD refers to whether or not an individual had a HD/A
MH refers to the number of days an individual reported poor mental health, out of a month
HC refers to whether or not a sampled individual has high cholesterol
SMHC refers to the interaction between smoking and high cholesterol: whether or not an individual smokes <u>and</u> has high cholesterol

Table 1.1

HD/ Income	1	2	3	4	5	6	7	8	Row-sum
HD	1553	2197	2519	2828	3161	3646	3404	4584	23893
NO HD	8258	9586	13475	17307	22722	32824	39815	85800	229787
Col-sum	9811	11783	15994	20135	25883	36470	43219	90385	253680
odds	0.18806	0.22919	0.18694	0.1634	0.13912	0.11108	0.0855	0.05343	0.10398

Table 1.2

For the sake of simplicity, only people of income levels 1 & 8 were compared

Income/MH	0-5	6-10	11-15	16-20	21-25	26-30	Row Sum
1	6408	659	614	311	172	1647	9811
8	82567	3191	1567	774	301	1985	90385
Col-sum	88975	3850	2181	1085	473	3632	100196
odds	0.07760969879	0.2065183328	0.3918315252	0.4018087855	0.5714285714	0.8297229219	0.1085467721

Table 1.3

Smoker/MH	0-5	6-10	11-15	16-20	21-25	26-30	Row Sum
Smoker	92677	5376	3784	1987	879	7720	112423
Non smoker	124795	5815	3368	1632	670	4977	141257
col sum	217472	11191	7152	3619	1549	12697	253680
odds	0.7426339196	0.924505589	1.123515439	1.21752451	1.311940299	1.551135222	0.7958756026

Table 1.4

	Smoker	Non Smoker	Row Sum
HD	14801	9092	23893
NO HD	97622	132165	229787
Col-sum	112423	141257	253680
odds	0.1516154146	0.06879279688	0.1039789022

Table 1.5

Income/BMI	0-10	11-20	21-30	30-40	41-50	51-60	61-70	71-80	81-90	91-100	Row Sum
1	0	687	5473	2778	665	149	28	17	9	5	9811
8	0	4755	64384	18882	1937	188	19	128	74	18	90385
Col Sum	0	5442	69857	21660	2602	337	47	145	83	23	100196
Odds	#DIV/0!	0.1444794953	0.08500559145	0.1471242453	0.3433144037	0.7925531915	1.473684211	0.1328125	0.1216216216	0.2777777778	0.1085467721

Table 1.6

HD/BMI	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	Row Sum
HD	0	983	14312	7181	1166	179	31	21	14	6	23893
NO HD	0	12439	152668	54618	8138	1191	190	284	212	47	229787
Col-sum	0	13422	166980	61799	9304	1370	221	305	226	53	253680
odds	#DIV/0!	0.07902564515	0.09374590615	0.1314768025	0.1432784468	0.1502938707	0.1631578947	0.07394366197	0.06603773585	0.1276595745	0.1039789022

Table 1.7

HD/MH	0-5	6-10	11-15	16-20	21-25	26-30	Row Sum
HD	18988	1153	909	507	238	2098	23893
NO HD	198484	10038	6243	3112	1311	10599	229787
Col-sum	217472	11191	7152	3619	1549	12697	253680
odds	0.09566514177	0.1148635186	0.1456030754	0.1629177378	0.1815408085	0.1979432022	0.1039789022

Table 1.8

HC/Income	1	2	3	4	5	6	7	8	Row Sum
HC	4817	6207	7618	9335	11751	15860	18212	33791	107591
NO HC	4994	5576	8376	10800	14132	20610	25007	56594	146089
Col-sum	9811	11783	15994	20135	25883	36470	43219	90385	253680
odds	0.964557469	1.113163558	0.9095033429	0.8643518519	0.8315171243	0.7695293547	0.7282760827	0.5970774287	0.73647571

Table 1.9

HD/SMHC	SMHC11	SMHC10	SMHC01	SMHC00	Row Sum
HD	10608	4193	6145	2947	23893
NO HD	42759	54863	48079	84086	229787
Col-sum	53367	59056	54224	87033	253680
odds	0.2480881218	0.07642673569	0.1278104786	0.03504745142	0.1039789022

Figures

Note that all colors used in charts are overlapping- not segmented

Figure 1.1

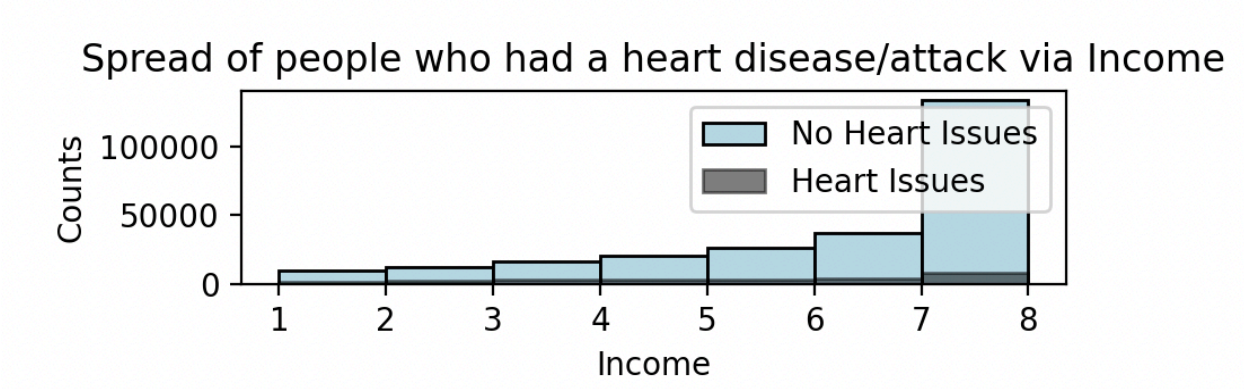


Figure 1.2a

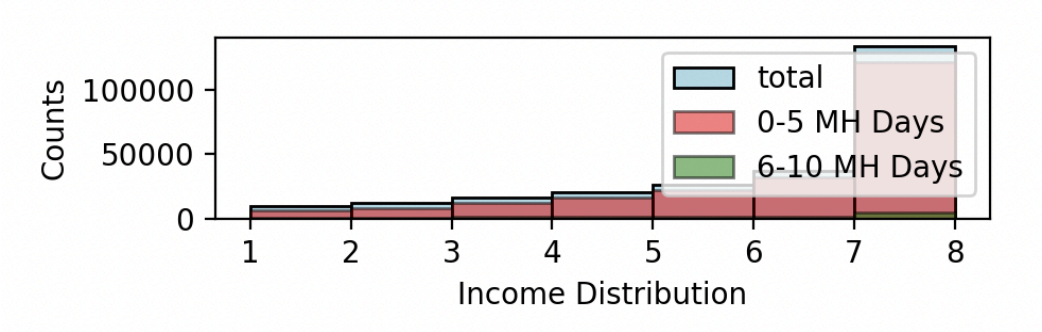


Figure 1.2b

(*Note that the Y-scaling is different due to the disproportionate amount of people who indicated 11+ mental health days in a month*)

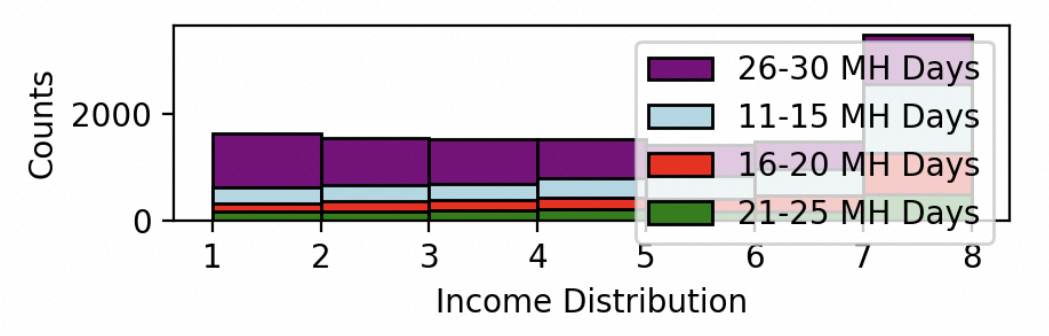


Figure 1.3

Note that although whether or not someone is alcoholic is measured here, it is not considered due to its lack of significance as can be seen in the histogram

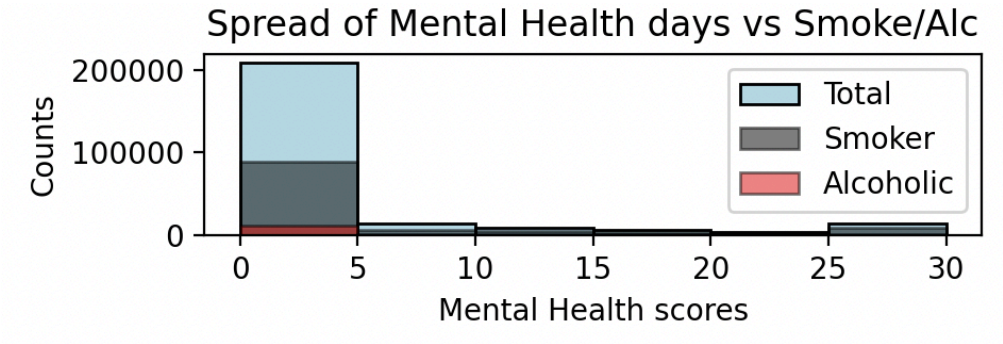


Figure 1.4

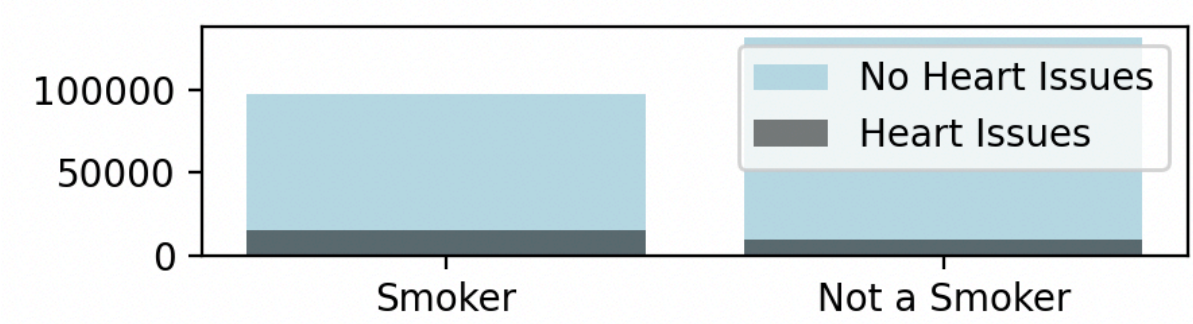


Figure 1.5a

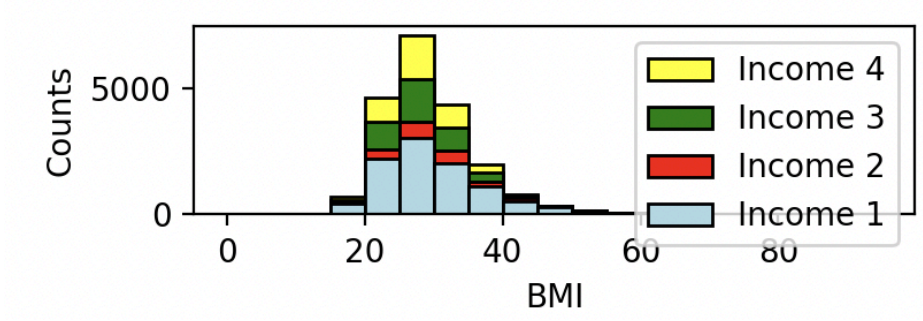


Figure 1.5b

Note again that the scaling is much larger in figure 1.5b due to the significantly more amount of people sampled at income levels 5-8

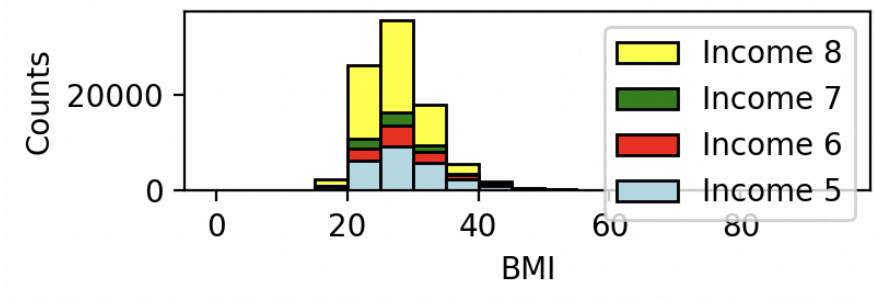


Figure 1.6

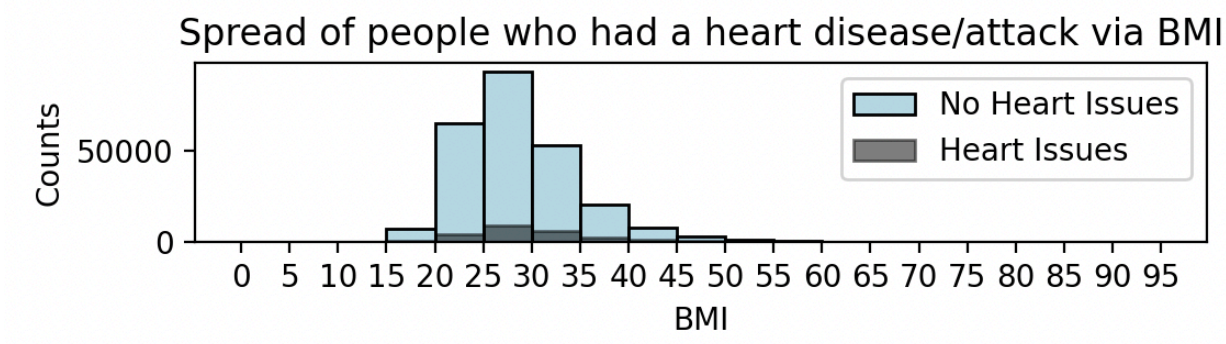


Figure 1.7

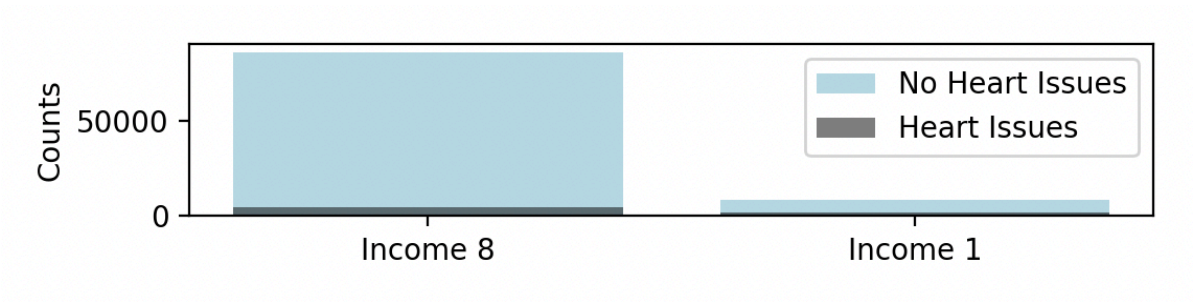


Figure 1.8

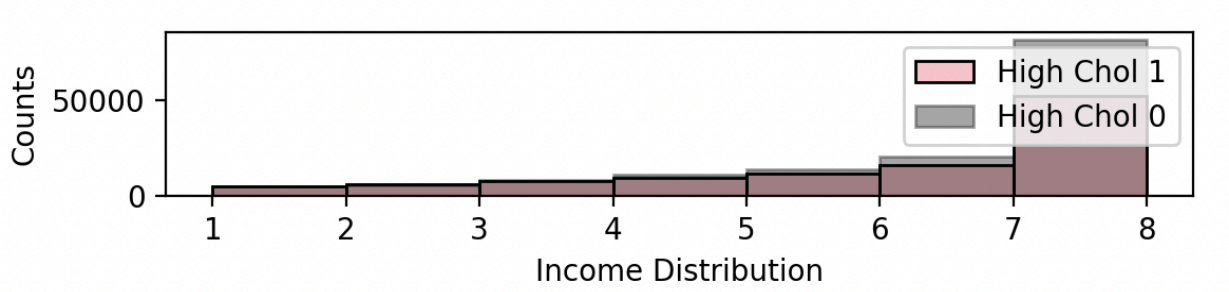


Figure 1.9

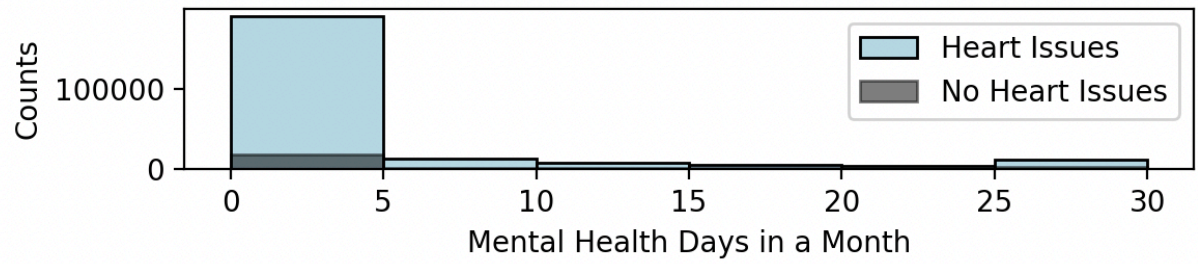


Figure 1.10

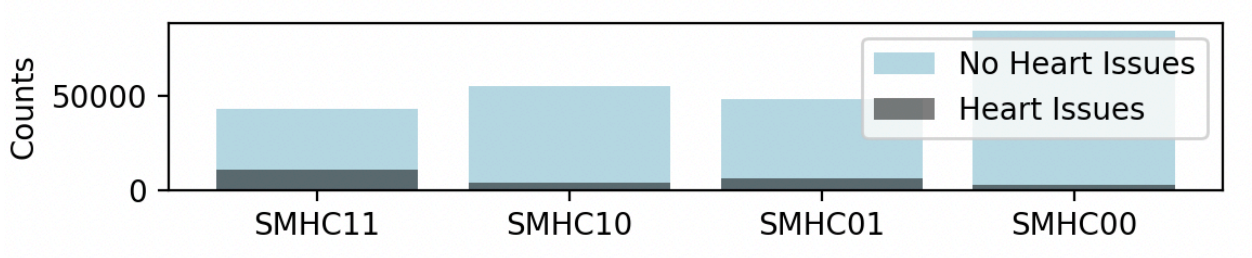
Note that the format of the variables is the following:

SM=Smoker

HC=High Cholesterol

Where the numbers are associated with the respective variable

Eg. SMHC11= count of people who smoke and have high cholesterol



Raw Dataset provided by CDC on Kaggle

Cleaned dataset provided by Alex Teboul on Kaggle

Tables and figures provided by Darling Judah Hsu