Tyler Benjamin

CPSC 4530 Final Project Paper
Effects of Alcohol Consumption in the Domain of Wellness and Health

The application domain I've chosen for this project is wellness and health. In this chosen domain, I will be exploring relationships between alcohol consumption and health using three different but related datasets. I will be analyzing the relationship between life expectancy and alcohol consumption, the relationship between a country's happiness scores and alcohol consumption, and finally (and most grimly), the relationship between car accident fatalities and blood alcohol concentration (BAC) level. My hope is that this project will explore the effects alcohol consumption can have on a person's life through interactive data visualization.

I chose health as a domain because I believe it is more relatable than a domain such as transportation or energy, especially when my chosen topic within that domain is alcohol consumption. Data visualization can be a powerful tool for illustrating the importance of a given issue, and substance abuse is no exception. The interactive visualizations I produced should emphasize any positive or negative effects of alcohol consumption.

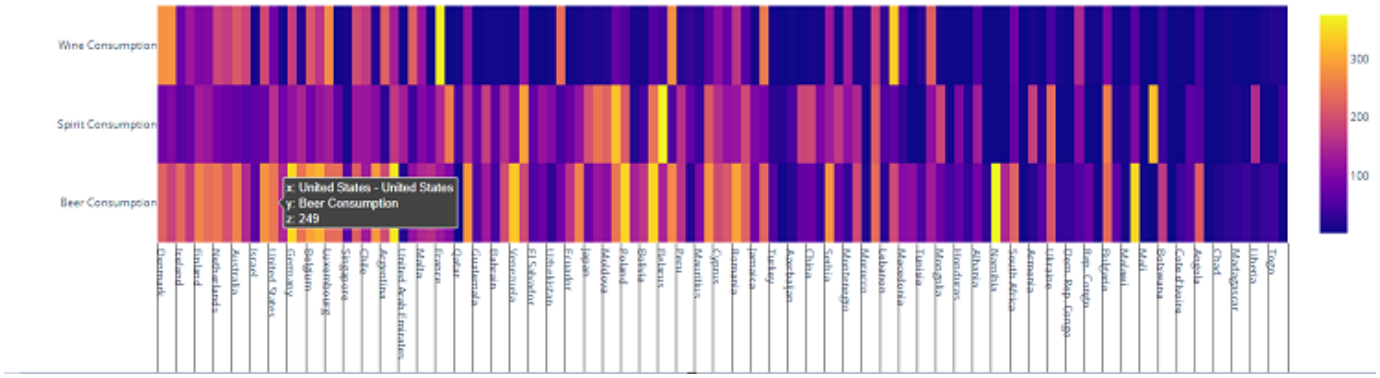
The first dataset I will be exploring through visualization consists of every country's happiness ratings amongst its population and various other statistics, including HDI (Human Development Index), GDP (Gross Domestic Product) per capita, and the amount of wine, spirits, and beer consumed per capita. The reason I selected this dataset and why I decided to visualize it first is to identify one possible reason why people consume alcohol, i.e. if happiness generally correlates with higher alcohol consumption.

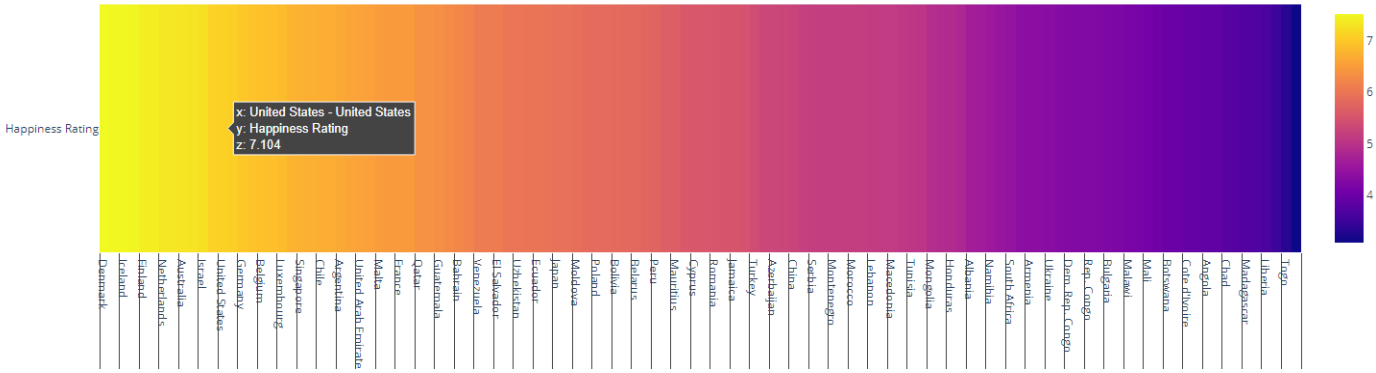
This dataset is a static, flat table with nine attributes. The country name serves as the key as each country is a unique value in the table with eight corresponding attributes. In addition to the previously mentioned attributes, the dataset also includes a region attribute and a hemisphere attribute. These attributes locate the country in a certain section of the world, in case regional analysis of alcohol consumption would be performed. I will only be analyzing alcohol consumption on a country-by-country basis, so these two attributes will not be used.

The happiness rating attribute is a decimal number ranging from 0 to 10, with lower scores indicating low levels of happiness and vice versa. The HDI attribute is a whole number and the GDP attribute is a non-zero decimal number. The consumption rates per capita for beer, wine, and spirits are all non-zero whole numbers.

Country	Region	Hemisphe	Happiness	HDI	GDP_PerC	Beer_Per(	Spirit_Per	Wine_PerCapita
Denmark	Western E	north	7.526	928	53.579	224	81	278
Switzerlar	Western E	north	7.509	943	79.866	185	100	280
Iceland	Western E	north	7.501	933	60.53	233	61	78
Norway	Western E	north	7.498	951	70.89	169	71	129
Finland	Western E	north	7.413	918	43.433	263	133	97
Canada	North Am	north	7.404	922	42.349	240	122	100
Netherlar	Western E	north	7.339	928	45.638	251	88	190
New Zeala	Australia a	south	7.334	915	40.332	203	79	175
Australia	Australia a	south	7.313	938	49.897	261	72	212
Sweden	Western E	north	7.291	932	51.845	152	60	186
Israel	Middle Ea	north	7.267	902	37.181	63	69	9
Austria	Western E	north	7.119	906	44.731	279	75	191
United Sta	North Am	north	7.104	922	57.589	249	158	84
Costa Rica	Latin Ame	north	7.087	791	11.733	149	87	11
Germany	Western E	north	6.994	934	42.233	346	117	175
Brazil	Latin Ame	both	6.952	758	8.639	245	145	16
Belgium	Western E	north	6.929	915	41.261	295	84	212
Ireland	Western E	north	6.907	934	64.1	313	118	165
Luxembou	Western E	north	6.871	904	100.739	236	133	271
Mexico	Latin Ame	north	6.778	772	8.444	238	68	5
Singapore	Southeast	north	6.739	930	55.243	60	12	11
United Kir	Western E	north	6.725	920	40.412	219	126	195
Chile	Latin Ame	south	6.705	842	13.961	130	124	172
Panama	Latin Ame	north	6.701	785	14.333	285	104	18
Argentina	Latin Ame	south	6.65	822	12.654	193	25	221
Czech Rep	Central ar	north	6.596	885	18.484	361	170	134
United Ar	Middle Ea	north	6.573	862	38.518	16	135	5

I decided to represent this dataset using two separate interactive heatmaps. One heatmap shows the consumption rates per capita per country of beer, wine, and spirits, and the other shows the happiness ratings per country. For this visualization, I am using Plotly, a Python graphing library.





Both heatmaps are organized by happiness rating; countries with higher happiness ratings are more left on the scale and vice versa. Based on the color numerical scale on the right, we can compare the two heatmaps to see if alcohol consumption correlates with higher happiness.

Looking at the map, if there is any correlation between happiness and alcohol consumption, it is rather weak. For wine consumption, there seems to be some correlation between it and happiness rating as the countries with higher wine consumption appear mostly at the left side of the heatmap. For spirit consumption, I would say there is no correlation between it and happiness rating, as countries with higher spirit consumption can be found all throughout the heatmap. Finally, for beer consumption, there seems to be some correlation, but very weak. There are a lot of countries with higher beer consumption towards the left side of the heatmap, but quite a few countries with high beer consumption can be found all over the heatmap. There appears to be a few outliers from at least a cursory look at the heatmap, as we can see countries with rather abysmal happiness ratings in comparison to other countries that also have high

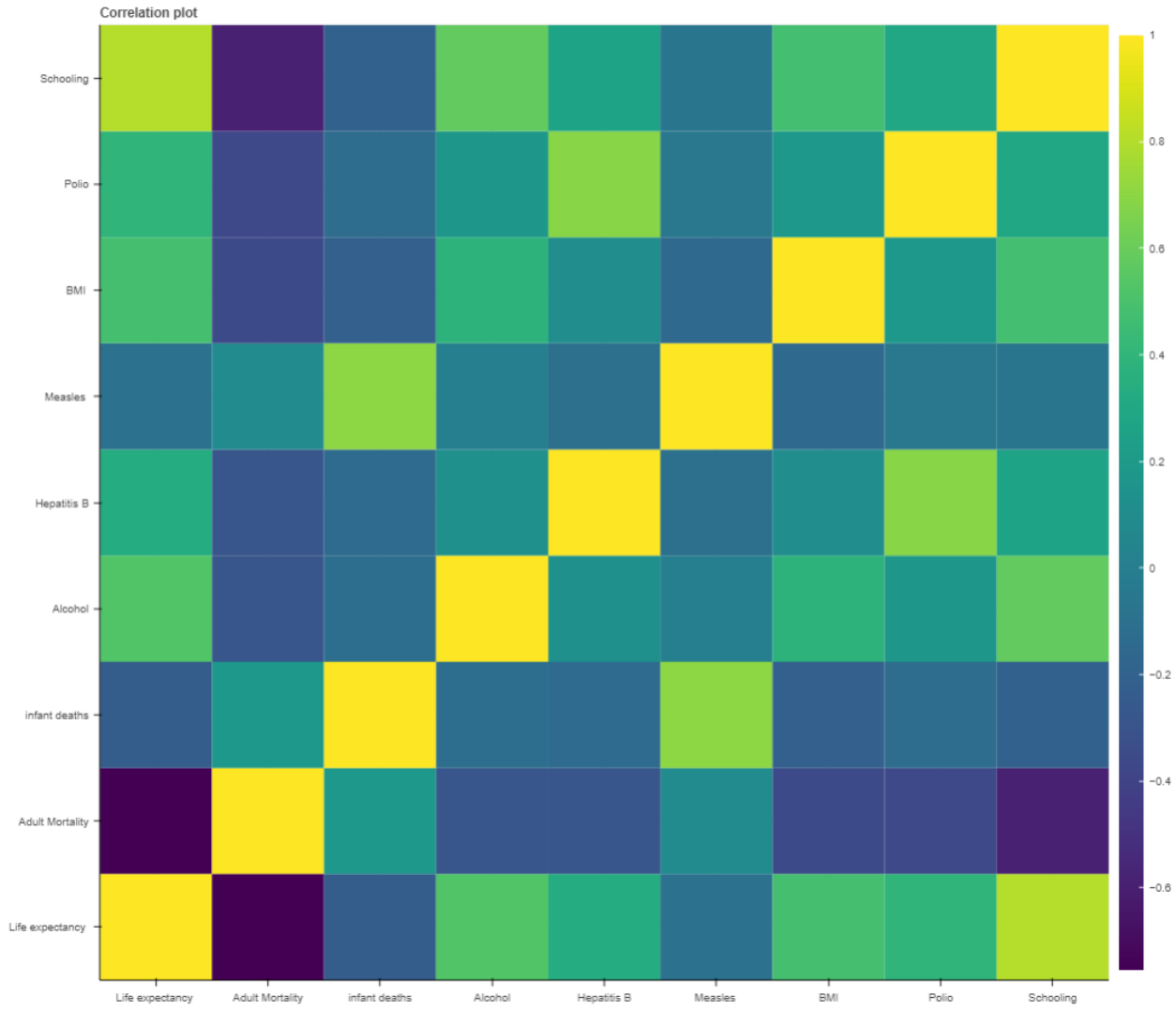
alcohol consumption rates. Some high happiness countries also have abysmal alcohol consumption rates, like Israel.

Based on these observations, there appears to be a weak correlation between happiness and alcohol consumption, which explains its popularity, but happiness is not necessarily an indicator of health as much as it can be of wellness. To look more closely at health, I will need to look at one of the biggest indicators of healthiness in a population: life expectancy.

The second dataset I will be exploring contains information about life expectancy rates per country from 2000-2015. It also contains information about various other measures of health, including but not limited to alcohol consumption per capita, Hepatitis B immunization rates, Measles case numbers, Polio immunization rates, etc. After exploring the relationship between happiness and alcohol consumption, I wanted to explore the relationship between life expectancy and alcohol consumption to examine alcohol's more meaningful consequences.

This dataset is a static, flat table and is item-based. There is no single attribute that serves as the key, as each country is repeated 15 times for every year between the range of 2000 to 2015. There are twenty-two attributes in total. All attributes are either a positive non-zero decimal number or a positive whole number. I only want to focus on life expectancy rates for 2014 as it is the most recent year with complete alcohol consumption rates in the dataset, so to preprocess the data I removed all other years' data from the dataset. I did this by cutting and pasting only 2014 data below the other entries in Excel, then deleting the other year data and shifting the cells up.

For this visualization, I used the Bokeh library that is available in both Python and Javascript, but I coded the visualization in Python.



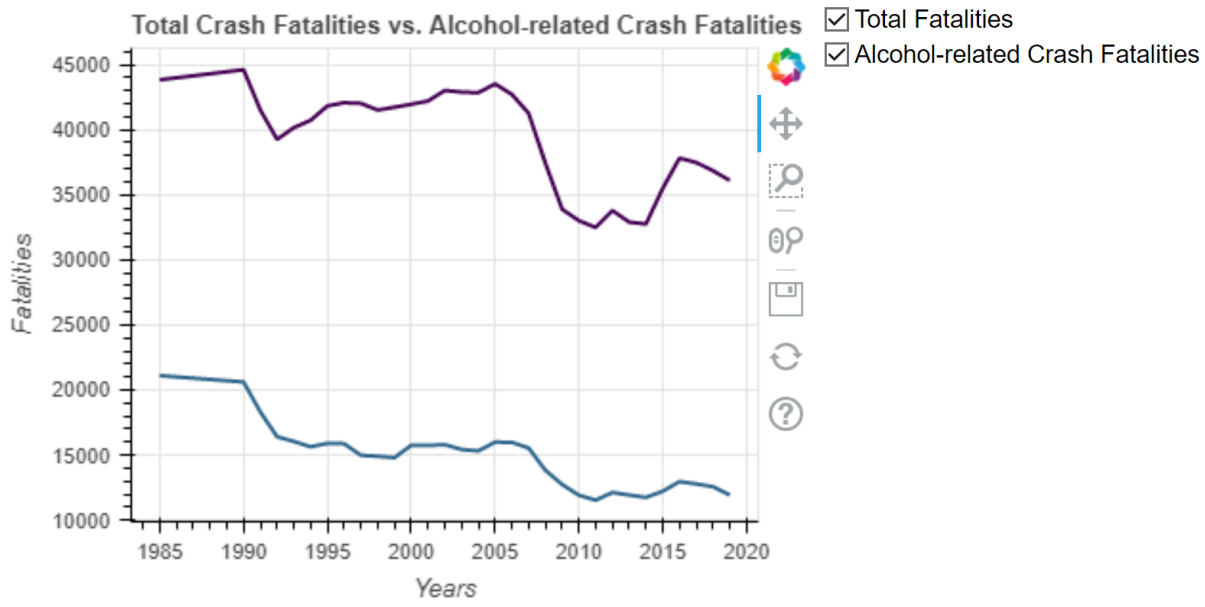
This visualization, in its interactive format, allows for the free movement of the correlation matrix to easily identify the correlation between two attributes. For example, since we're looking at the relationship between alcohol consumption and life expectancy, we can move the matrix so that the square corresponding to that relationship is right next to the color scale on the right. Doing this, we can observe that the correlation is between the range of 0.5-0.6. This indicates a positive correlation between alcohol consumption and life expectancy. However, the measures of happiness and life expectancy are merely averages that can vary wildly depending on the person. The previous two visualizations painted too rosy a picture of the effects of alcohol. I wanted to at least provide some semblance of balance between the potential benefits and consequences associated with alcohol consumption. I decided to choose alcohol-related car accident fatalities.

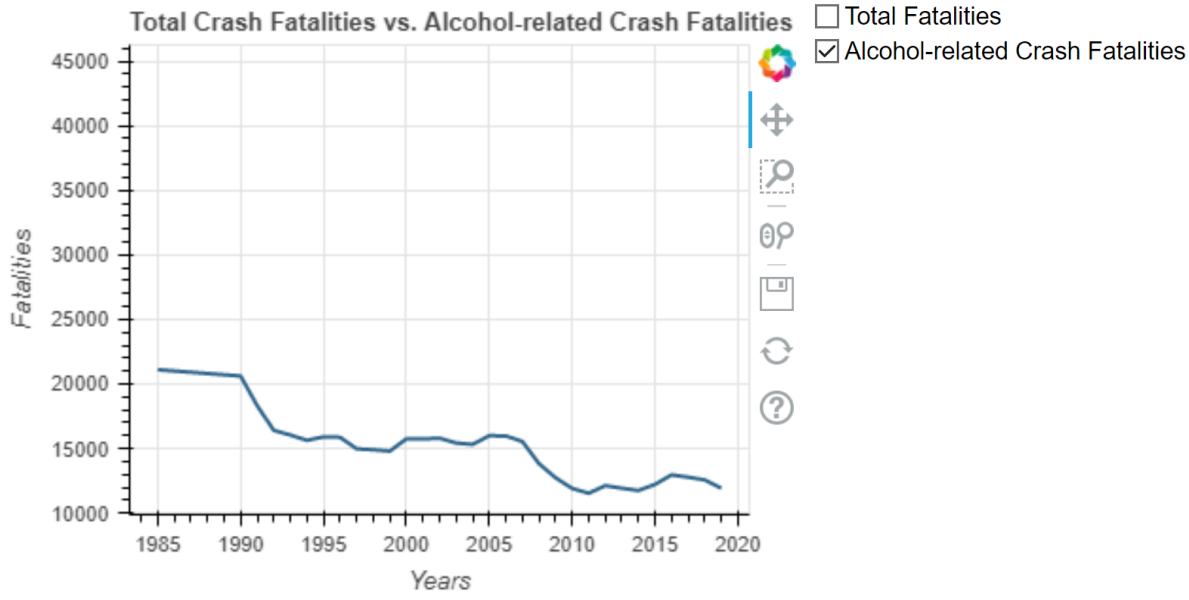
The third and final dataset I will be using for visualization contains data on highway car accident fatalities and the BAC of the driver. This dataset is formatted a bit differently than the other two datasets, so I needed to perform some preprocessing to make the data more accessible. I'm only concerned with the percentage of fatalities with a BAC = 0.01+, so I removed all other data or notes included with the data in the .csv file. I then transposed the data so that the attribute names were at the top of the data's respective columns and added an attribute name to the column containing the years.

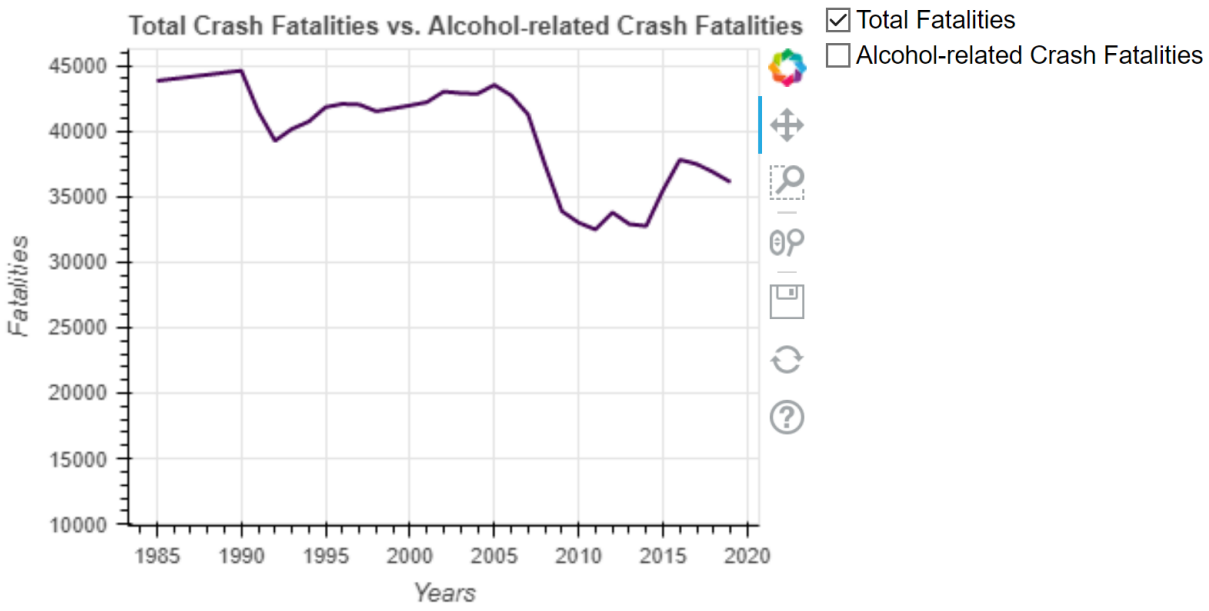
For this visualization, I used the Bokeh library that is available in both Python and Javascript, but I coded the visualization in Python.

Year	Total fatalities	Fatalities in alcohol-related crashes	Percent
1985	43825	21098	48.1
1990	44599	20607	46.2
1991	41508	18307	44.1
1992	39250	16401	41.8
1993	40150	16039	39.9
1994	40716	15626	38.4
1995	41817	15893	38.0
1996	42065	15866	37.7
1997	42013	14973	35.6
1998	41501	14899	35.9
1999	41717	14790	35.5
2000	41945	15746	37.5
2001	42196	15731	37.3
2002	43005	15793	36.7
2003	42884	15423	36.0
2004	42836	15311	35.7
2005	43510	15985	36.7
2006	42708	15970	37.4
2007	41259	15534	37.6
2008	37423	13826	36.9
2009	33883	12731	37.6
2010	32999	11906	36.1
2011	32479	11527	35.5
2012	33782	12118	35.9

This screenshot is an excerpt of data from the dataset, and the attributes I used in my visualization are in bold. My interactive visualization using this data is a line graph that shows both the total number of fatalities and the total number of fatalities in alcohol-related crashes. Whether or not a crash is alcohol-related is determined by the BAC of the driver being equal to or greater than 0.01. Users can toggle on or off both graphs in the visualization to compare the two fatality amounts.







Based on the visualization, we can observe that a frighteningly high number of fatalities are alcohol-related; this number varies anywhere from 33% to 48%. Though there is some observable variation over the years, the overall number of fatalities (both alcohol-related and not) is decreasing. However, it is clear that driving while under the influence is an extremely hazardous activity and a very common cause of crash fatalities across the years included in the dataset. I did not notice any outliers in the data.

According to the visualizations I produced, there seems to be some positive correlation between alcohol consumption and health/wellness, with the caveat of moderation. If I were to continue this project, I would examine more effects of alcohol consumption on factors like other causes of death besides crash fatalities, such as liver failure. That would likely indicate more serious consequences for consuming alcohol in excess.