title: "Alcohol in Texas"

subtitle: "DA 6233"

author: "Connor Allison, Joel Fry, Nicholas Mata, Royalpriesthood Olala"

date: "December 2024"

First off, let us apologies for the lack of organization apparent in this folder. We have opted to keep all of the Rmd and CSV files that were the result of data cleaning, exploration and educationally motivated trial and error. **Descriptions of all files**:

“Mixed\_Beverage\_Gross\_Receipts”: original unedited data

“Practice\_FinalProject.Rmd”: used to explore the original data, create and format the year column, then ultimently create the “Alc\_per\_year.csv” which was then copied and added to in Excel to create what became the “alc\_per\_person.csv”

“Presentation\_Final.qmd”: contains the three graphs used in the presentation

“Total Alcohol Receipts Across San Antonio.twbx”: tableau map used in presentation

“alc\_per\_person.csv”: used line plot 2 and 3. Includes alcohol by type adjusted and not adjusted for population and inflation. CPI was used to adjust data for inflation. Our source for this CPI data was <https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913->

Population data was retrieved from <https://www.macrotrends.net/global-metrics/cities/23128/san-antonio/population>

‘Alc\_VS.Rmd’: contains the second graph in the final presentation

‘AlcTypesComp.Rmd’: contains the third graph in the final presentation

“Joels\_Part.Rmd”: here is were the original data was manipulated to form “Total\_Receipts\_Big4.csv” which was used to make the first graph in “Presentation\_Final.qmd”. Joels\_part also led to the creation of Gross\_Receipts\_AUS.csv,

Gross\_Receipts\_Big\_Four.csv, Gross\_Receipts\_DAL.csv, Gross\_Receipts\_HOU.csv, Gross\_Receipts\_SA.csv

Our Project was to tell the story of alcohol sales in Texas and convey our findings. Our first step was searching for a data set, and we came across

“Mixed\_Beverage\_Gross\_Receipts \_20240922.csv”

which contained all the alcohol sales for every bar and restaurant in Texas from 2007 to the present. The CSV file was 661 megabytes of simple text, so we quickly recognized that we needed to figure out a way to pare it down. Beer sales in Devine are not the same as liquor sales in Austin, so we made the subjective decision to use data from the four largest cities in Texas. When we went to familiarize ourselves with the data set, we found there were four different dates, and did not all have a value, and the first address we saw was not always the address of the venue. Once we identified the usable columns, we started reading out data to smaller CSV’s to cut down on computer resources. The primary foundational filters and mutations we made were filtering location by “Location City” for Houston, Dallas, San Antonio, and Austin”. Then there were four columns that contributed to “Total Sales”, which were “Liquor Receipts”, “Wine Receipts”, “Beer Receipts”, and “Cover Charge.” Because “Cover Charge” was sparsely populated, and was not an alcohol type, we made a new “Total Sales” column by adding the three other receipt columns together. We also removed any locations that had all missing or all zero data for receipt columns.

The next stage of cleaning the data found us filtering into five data frames for each city and one with all four cities together. Additionally, 2024 data was incomplete, so we took data from 2007 to 2023, and summarized them across the four sales columns to get totals for each year. To create the ‘Years’ column we had to take the “Obligation End Date” column (which was in str format) and convert it to MM/DD/YYYY format with the MDY function. From there, we isolated and grouped by year. After a pivot\_longer mutation on the data we were able to throw them in several charts, to include totals by each city, and sales for San Antonio. This led to our first graph showing the increase in sales and the covid dip as found in the

“Total\_Receipts\_Big4.csv”

The next big data manipulation was to adjust the sales numbers we now had for population and inflation. Because this was a complicated process of applying changes to the data, we found it was easier to pull data out to other CSV files to do the calculations and read them back in to chart the data as found in the

“alc\_per\_person.csv” file.

These adjusted data calculations were applied to one graph for the total sales to show how sales have changed over time, and how sales changed by type. Because we pulled the data out of R, it was easier to add a column for the consumer price index to adjust for inflation and another column for population, which we could only find total numbers, not just the legal drinking age subset. Additionally, the sales would have included tourists coming here and buying alcohol, but they would not be counted in the population, so we felt it was a reasonable compromise to use the population data.

Lastly, we took the location data to Tableau to show where the alcohol was consumed across the city. As this could be used for future restaurants and bars to find a location that is either underserved or they could find a location where all the money is already going. The resulting file is

“Total Alcohol Receipts Across San Antonio.twbx”