

Week 2

Advay Vyas

4/15/25

Contents

Introduction	1
Improving statistical analysis and data cleaning	2
Weekly time-series data	2
Comparing state-wide	3
Social Vulnerability Index	5
How it is calculated	5
Why it's relevant + trends	6
Fort Worth vs. Dallas	7
SVI Comparison	7
Demographics	7
Healthcare access	7
Conclusion	8
Last week's hypothesis and relevance	8
Conclusion	8

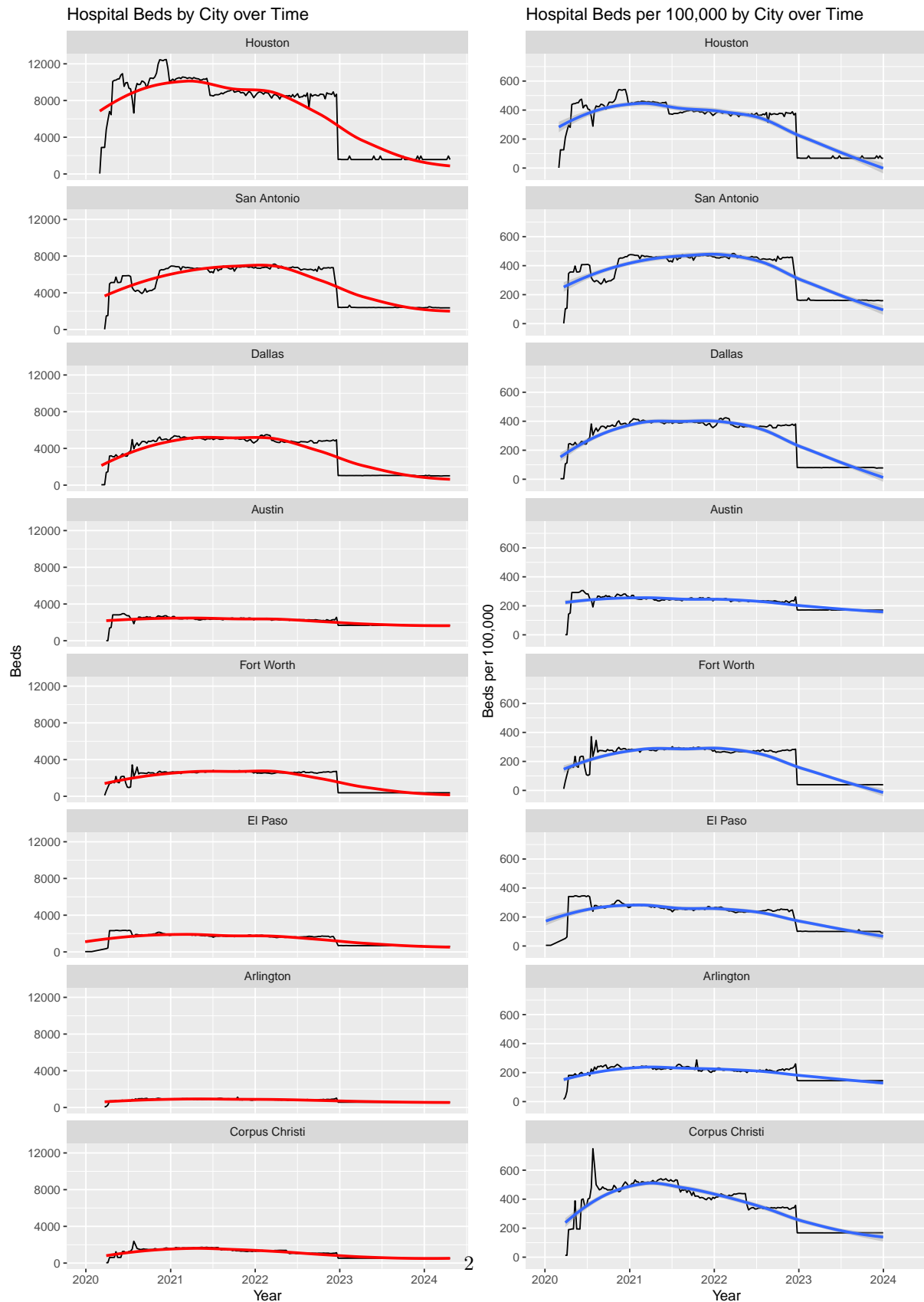
Introduction

This week, we need to target why specific pairs of cities have disparities so that we can generalize in the following weeks. In particular, why do Fort Worth and Dallas (nearby cities) have differences in their flu patients over time? We could use the Social Vulnerability Index (SVI) as a key feature in our analysis, provided we understand how it is calculated and extract the relevant information.

As far as changing statistical techniques, we'd also like to focus more on weekly time-series data, normalizing statistics per capita, and trying to visualize this data with maps. Time permitting, I'd also like to take a look at last week's hypothesis of faster "recovery time" with more hospitals (Houston, Dallas).

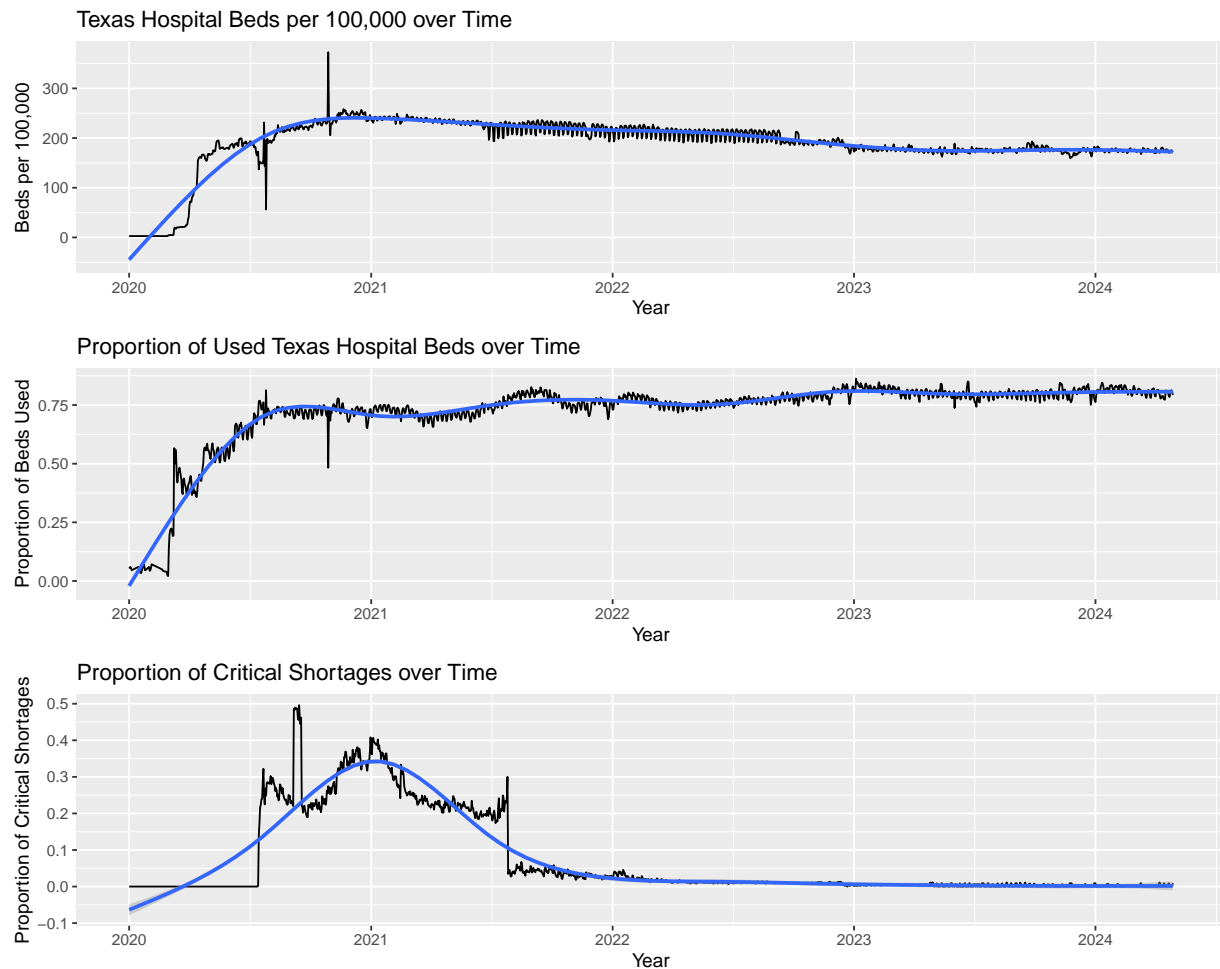
Improving statistical analysis and data cleaning

Weekly time-series data



What does this graph show us? Let's unpack our findings real quick. Clearly, we can see that the trends are much more visible on the right hand side. Corpus Christi, Houston, and San Antonio have the most beds per 100,000 while Austin and El Paso have the fewest. We can also see the Ft. Worth vs Dallas comparison here (a little spoiler for the analysis that's later in the report).

Comparing state-wide



A quick aside for the data wrangling steps: I removed about 10 entries that had a bed count under 100, which seemed completely unreasonable for an entire state.

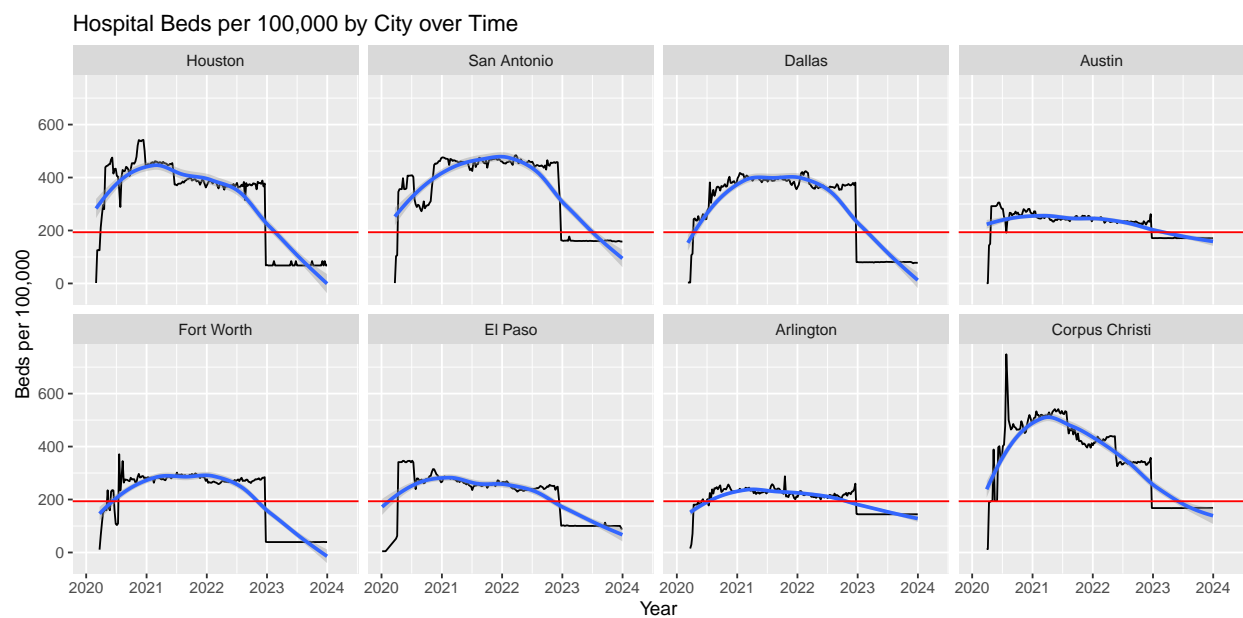
First off, we can ignore the critical shortages graph because it is mainly just representing the onset of COVID-19 and accompanying pandemic. Next, the proportion of used hospital beds looks to stay stagnant over time, and it is likely that the early low values are due to a lack of reporting in that area until about mid-2020.

The most useful graph in this set is the last one, which displays the beds per 100,000. We can say that after the stabilization, Texas' beds per 100,000 hovers roughly around 200 per 100,000, a figure that is supported by the below table.

Table 1: Distribution statistics for Texas hospital beds per 100,000

Statistic	Value
Min.	2.62
1st Qu.	176.20
Median	194.12
Mean	193.08
3rd Qu.	223.83
Max.	372.67

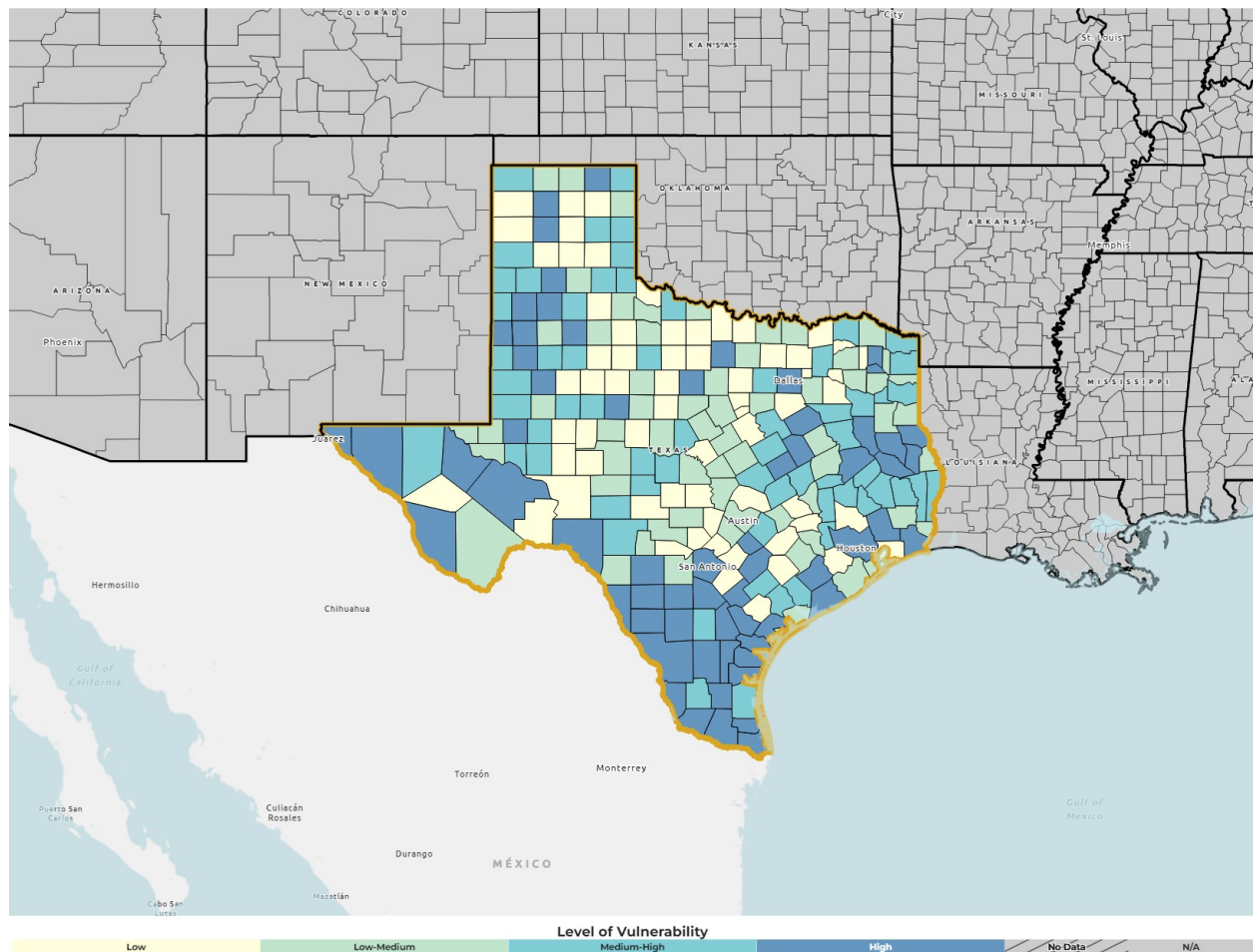
Now, let's take this average state-wide data and compare it to each individual city.



Right away, we can say that Corpus Christi has a lot more beds per 100,000 relative to its population size, while cities like Austin and Arlington barely pass the statewide mark. Therefore, we could investigate further and might find a trend that Austin and Arlington take longer to “recover” due to possible shortages or limited equipment (another foreshadowing).

Social Vulnerability Index

Overall SVI Texas: Statewide Comparison By County | 2022



The above graphic is a screenshot displaying the SVI map for the state of Texas.

How it is calculated

The following information comes directly from the attached PDF that comes with the SVI data.

Every community must prepare for and respond to hazardous events, whether a natural disaster like a tornado or a disease outbreak, or an anthropogenic event such as a harmful chemical spill. The degree to which a community exhibits certain social conditions, including high poverty, low percentage of vehicle access, or crowded households, among others, may affect that community's ability to prevent human suffering and financial loss in the event of a disaster. These factors describe a community's social vulnerability.

SVI indicates the relative vulnerability of every U.S. census tract. Census tracts are subdivisions of counties for which the Census collects statistical data. SVI ranks the tracts on 16 social factors, such as unemployment, racial and ethnic minority status, and disability status. Then, SVI further groups the factors into four related themes. Thus, each tract receives a ranking for each Census variable and for each of the four themes as well as an overall ranking.

We will be using the overall ranking in our calculations since there is no healthcare metric - below is a list of the 16 factors.

- **Socioeconomic Status**
 - Below 150% Poverty
 - Unemployed
 - Housing Cost Burden
 - No High School Diploma
 - No Health Insurance
- **Household Characteristics**
 - Aged 65 & Older
 - Aged 17 & Younger
 - Civilian with a Disability
 - Single-Parent Households
 - English Language Proficiency
- **Racial & Ethnic Minority Status***
 - Hispanic or Latino (of any race)
 - Black and African American, Not Hispanic or Latino
 - American Indian and Alaska Native, Not Hispanic or Latino
 - Asian, Not Hispanic or Latino
 - Native Hawaiian and Other Pacific Islander, Not Hispanic or Latino
 - Two or More Races, Not Hispanic or Latino
 - Other Races, Not Hispanic or Latino
- **Housing Type & Transportation**
 - Multi-Unit Structures
 - Mobile Homes
 - Crowding
 - No Vehicle
 - Group Quarters

Why it's relevant + trends

A higher social vulnerability corresponds to a higher SVI percentile and the below table represents the major cities and their SVI values, sorted by the least socially vulnerable county.

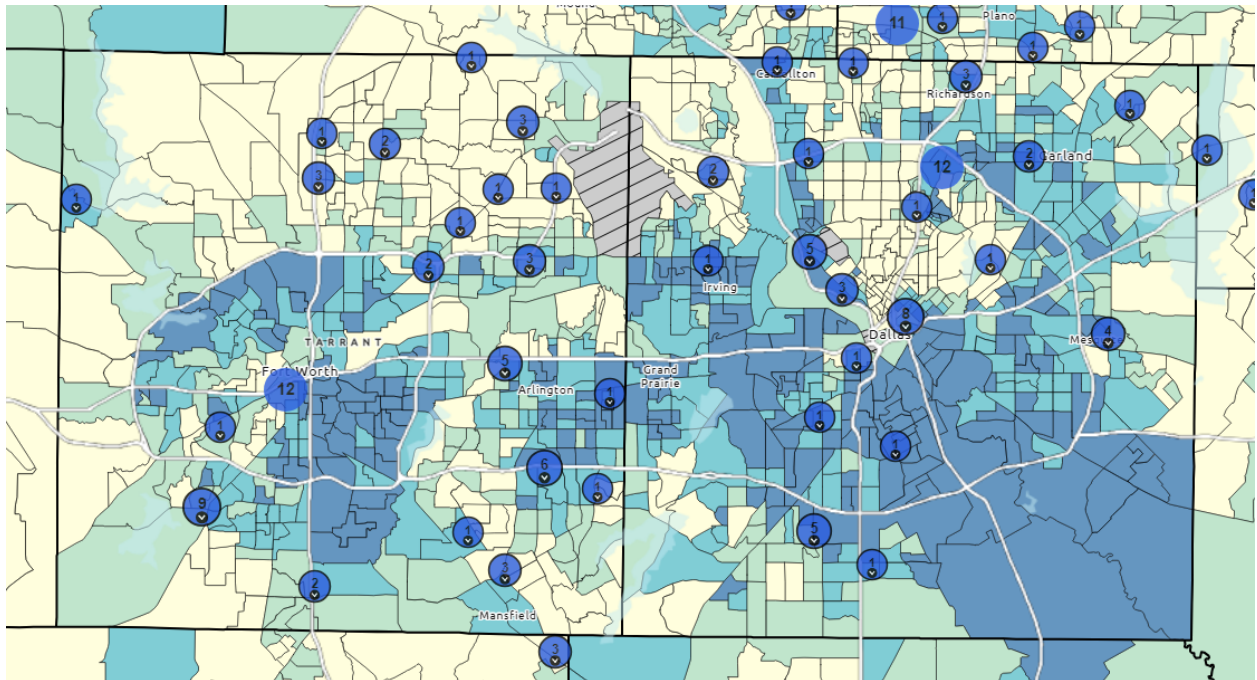
Table 2: Texas County SVI Percentile with corresponding Major Texas Cities

County	Corresponding Cities	SVI Percentile (0-1)
Travis County	Austin	0.2688
Tarrant County	Fort Worth & Arlington	0.5178
Bexar County	San Antonio	0.7589
Nueces County	Corpus Christi	0.7708
Dallas County	Dallas	0.8142
Harris County	Houston	0.8340
El Paso County	El Paso	0.9565

What's interesting is that even though Austin and Arlington barely had more beds than the state-wide mark, their counties seem to have less social vulnerability. However, there are lots of confounders because of the other factors in SVI and an improvement would be to only select healthcare factors (none of the SVI themes directly match up so I decided to use the entire metric). Disregarding hospital beds per 100,000, this trend makes sense because El Paso is a border city and likely has conflict while Dallas and Houston bring social vulnerability by virtue of having massive populations.

Fort Worth vs. Dallas

SVI Comparison



If we examine this map above, we can see that Fort Worth and Dallas have slightly different distributions of the SVI index (the darker, the more vulnerable just like before). It is important to understand that these areas are census tracts, which are created outside of political boundaries and meant to average a certain population per tract, which leads to major variation in the land size. However, upon a closer look, those more vulnerable areas simply have more land area and falsely portray greater vulnerability in Dallas. Therefore, I think we can conclude that Fort Worth and Dallas likely have similar SVI distributions.

Demographics

Table 3: Population of Dallas and Fort Worth (2020-2023)

City	2020	2021	2022	2023
Dallas	1303212	1289705	1297358	1302868
Fort Worth	923602	937590	957103	978468

Dallas has a larger population than Fort Worth yet more beds per 100,000 by 100 beds.

Healthcare access

Table 4: Hospitals of Dallas and Fort Worth

City	Hospitals
Dallas	19

City	Hospitals
Fort Worth	11

Makes sense, the reason Dallas has more beds/100k is because they simply have more hospitals. This may be due to Dallas being more dense and creating more hospitals per population because it is a major city built to handle crises. Anyway, the hospitals explain the beds.

Conclusion

While the flu patients were almost identical, the beds looked different when normalized but this simply boils down to more hospitals in the city of Dallas as compared to the city of Fort Worth.

Last week's hypothesis and relevance

While it is wishy-washy in a sense, I think another research rabbit hole to dive down is the recovery rate of cities after pandemics or seasonal flu outbreaks and how that is affected by hospitals, population density, the SVI metric, and other factors. We saw conflicting values today, particularly Austin/Arlington, which motivate me to shelve this idea for now.

Conclusion

This week, I investigated state-wide and city-wide time-series data and tried to discern trends with the help of external population and SVI data. I did find some interesting insights into SVI and normalizing the population revealed interesting outliers like Corpus Christi (which may due to its presence as a vacation spot) and underachievers like Austin/Arlington. Good progress this week, I think I can write more about actual results next week or very soon.