

# Week 4

Advay Vyas

4/29/25

## Contents

<b>Introduction</b>	<b>1</b>
<b>Flu spikes</b>	<b>2</b>
Isolating the seasonal spikes . . . . .	2
Comparing each spike . . . . .	3
<b>Correlation with other variables</b>	<b>5</b>
Correlation matrix . . . . .	5
Vaccination coverage rate . . . . .	6
Routine checkups . . . . .	7
Insurance . . . . .	8
Analysis . . . . .	9
<b>City comparisons</b>	<b>10</b>
Dallas vs. Austin . . . . .	10
Cycle by cycle analysis . . . . .	10
Important changes in major Texas cities (2022-2024) . . . . .	10
<b>Conclusion</b>	<b>10</b>

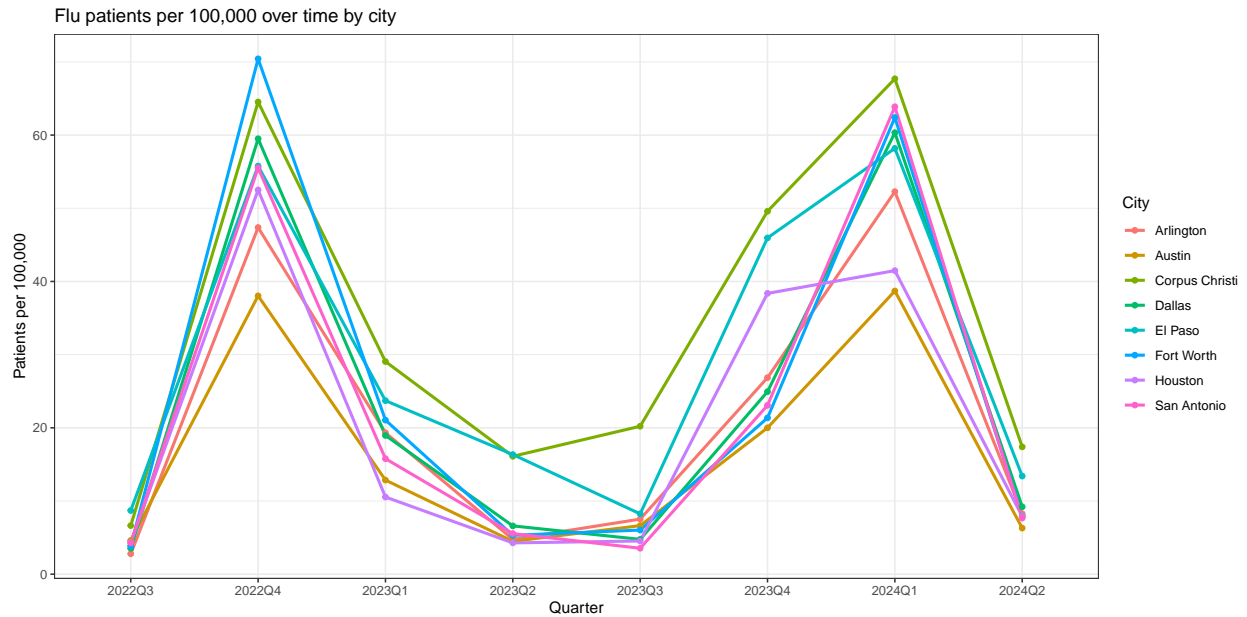
---

## Introduction

This week, we are going to look into vaccination and quantify the correlation between vaccination and flu patients. We will also attempt to look at the new datasets for specifically flu patients and maybe make a correlation matrix to find any variables that match up well. Also, we want to look into why cities like Dallas have much higher flu patients per capita compared to ciites like Austin. We can also look at why cities like Houston and San Antonio have different numbers every cycle.

# Flu spikes

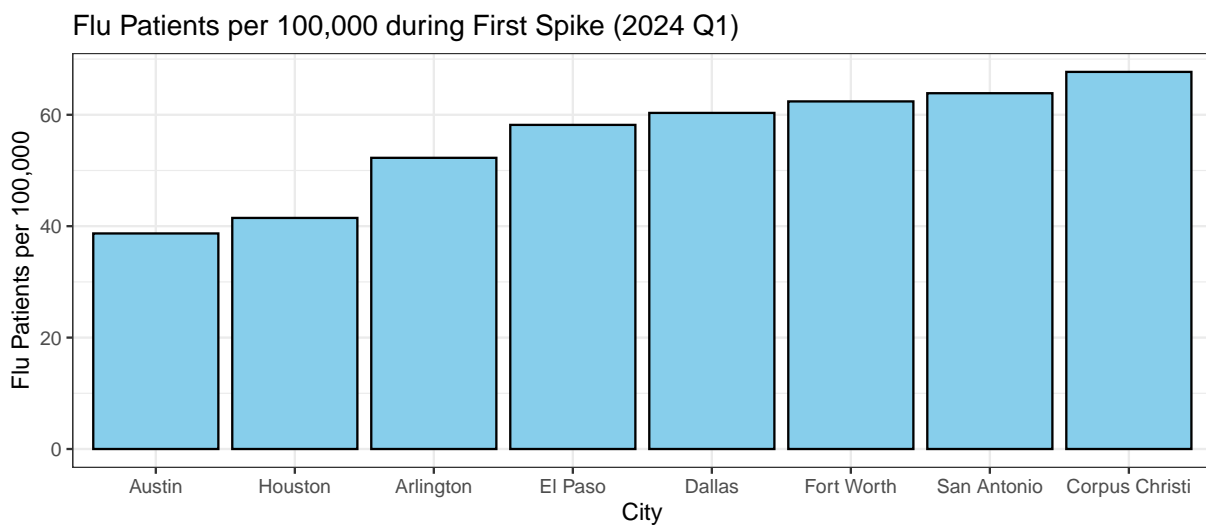
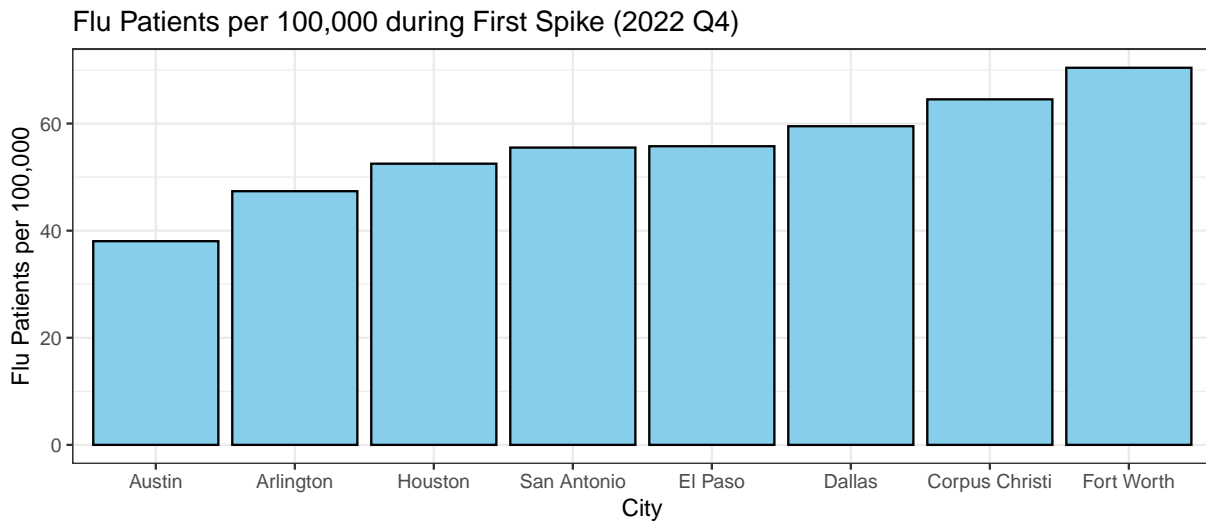
## Isolating the seasonal spikes



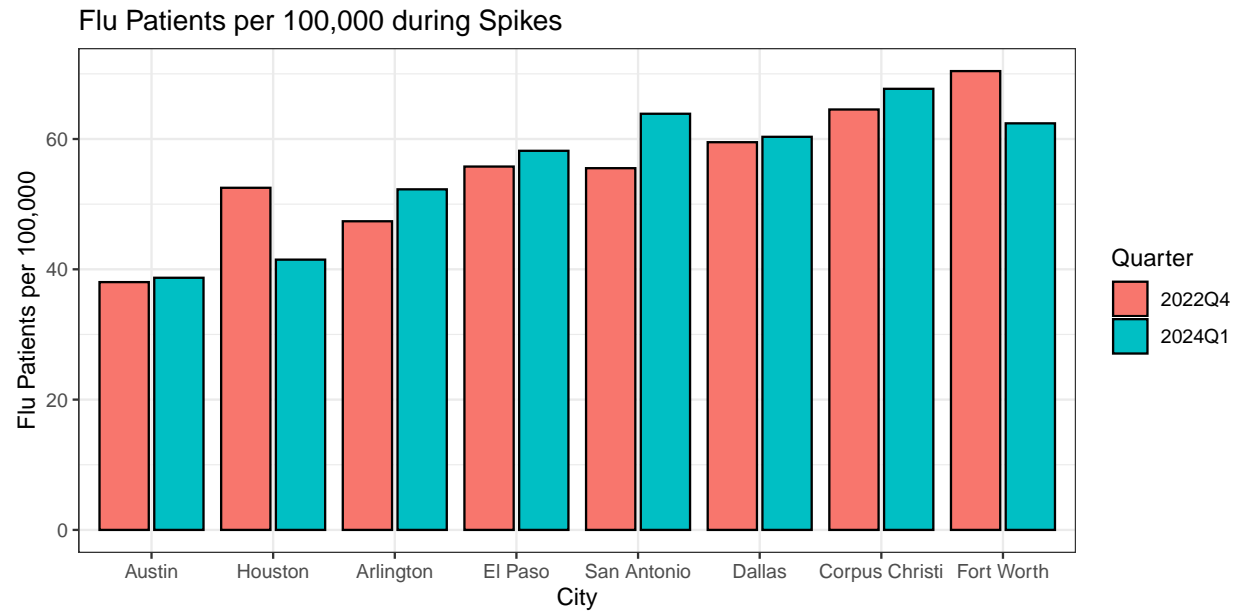
We can clearly see that the peaks of each season (“spikes”) are at 2022Q4 and 2024Q1. Now, let’s grab those specific values from the first and second spike to run statistical analysis on. I’d like to preface by saying that the correlation matrix will be much more accurate for 2022 since I only have 2022 data for city health and flu vaccination coverage.

## Comparing each spike

Let's get some bar graph representations of our data so we can interpret it a little better.



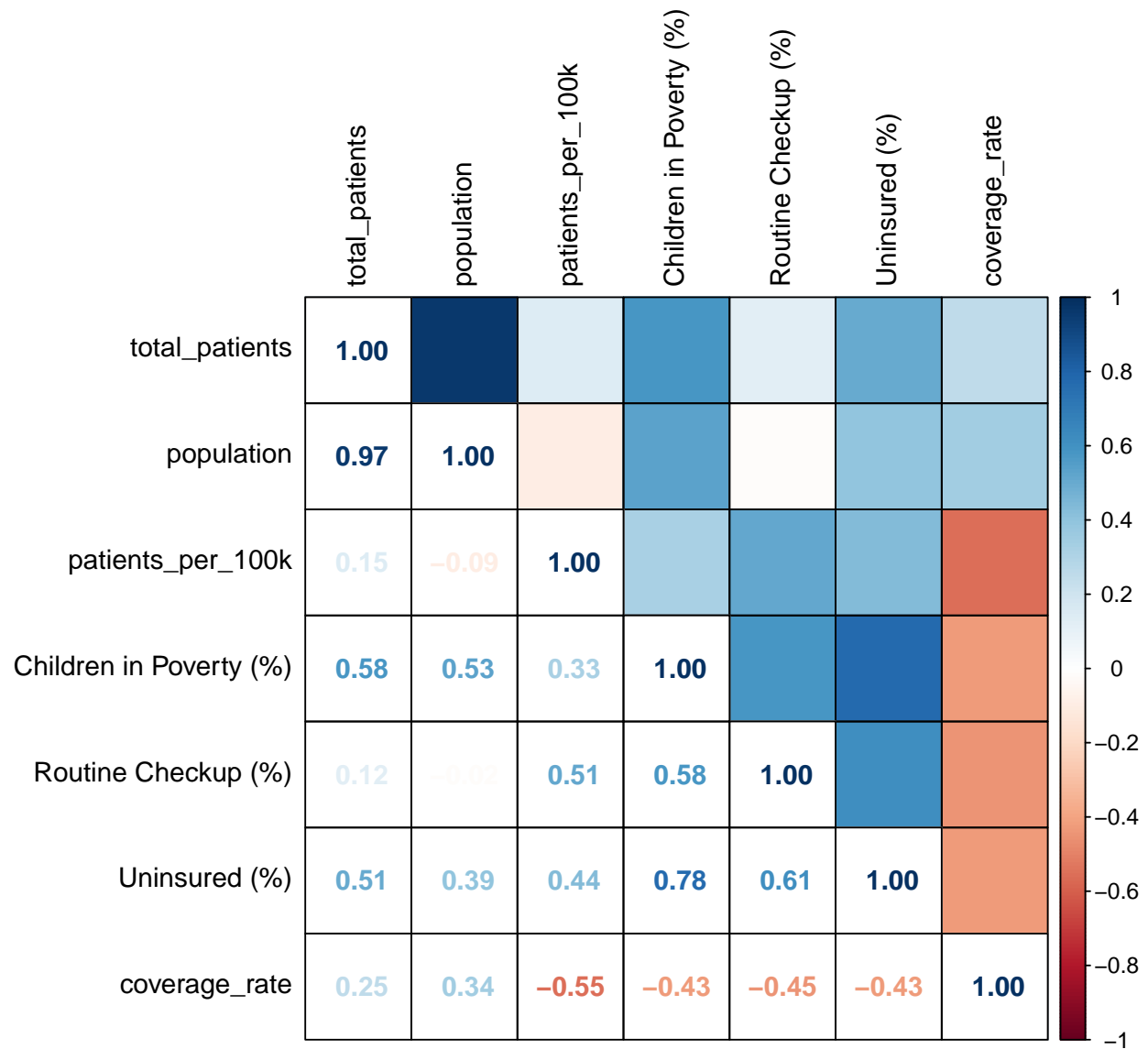
Looks like small changes in each spike, but Austin has the least both times (this aligns with what we saw last week). Now, let's compare the spikes side by side to get a better feel.



I don't see any specific trend with 2022 vs. 2024 (technically end of 2023, maybe it was just a late flu season). Austin and Dallas look about even both times, El Paso, San Antonio, Corpus Christi, and Arlington all increased while Fort Worth and Houston decreased - could look into why that is.

## Correlation with other variables

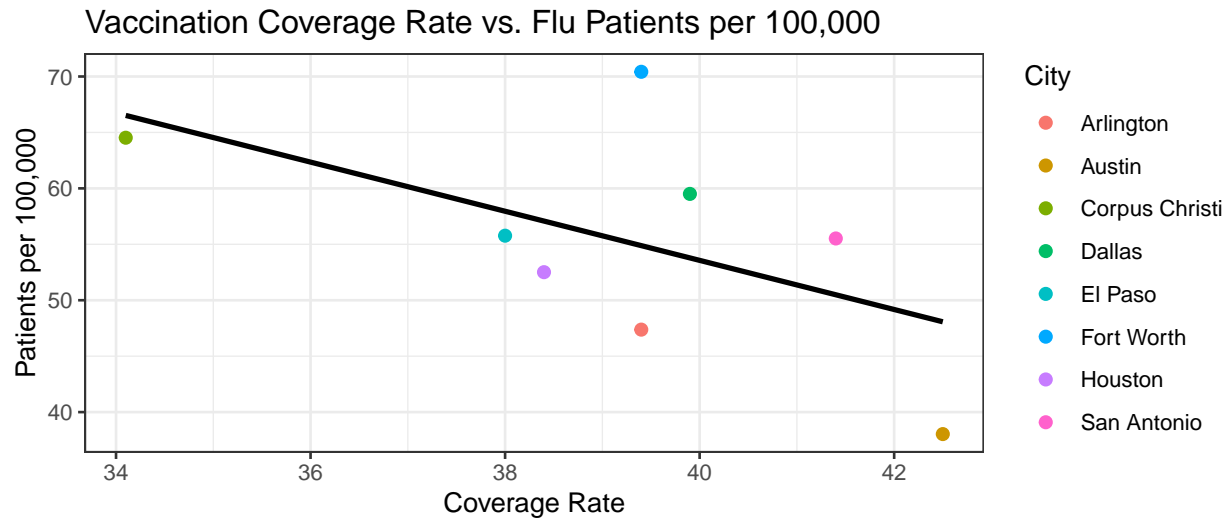
### Correlation matrix



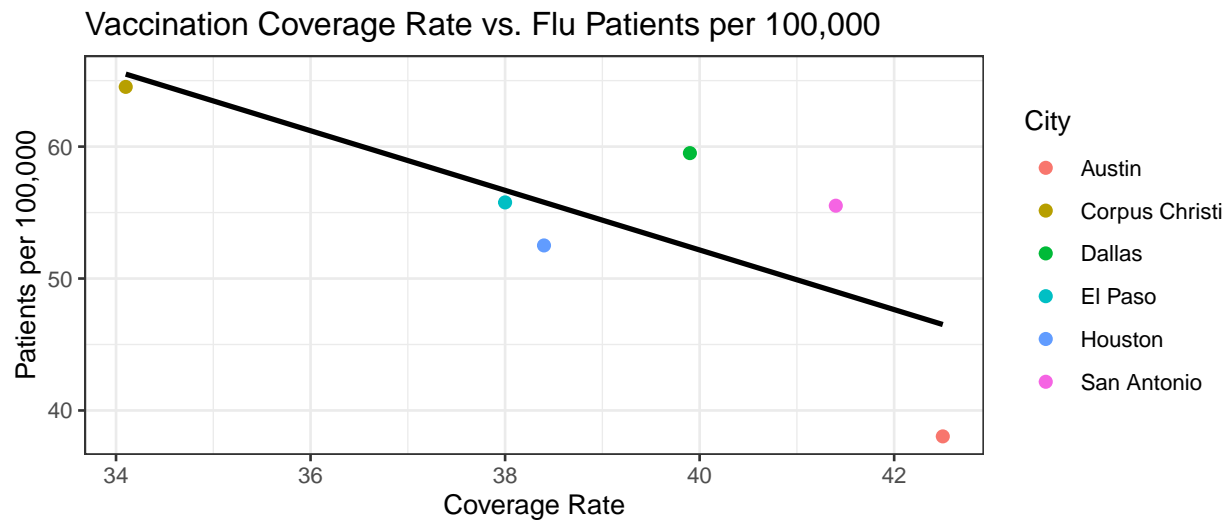
Above is the correlation matrix for City Health metrics, flu vaccination data, and patients per 100,000 for the first spike (2022Q4). We're looking for high correlation values with patients per 100k, which we see from Routine Checkups (%) at 0.51, Uninsured (%) at 0.44, and coverage\_rate at -0.55.

## Vaccination coverage rate

Note that for the coverage rate graph, the trend would be even stronger if Fort Worth and Arlington didn't share the same coverage rate by virtue of being in the same county - this first graph shows a very strong trend.

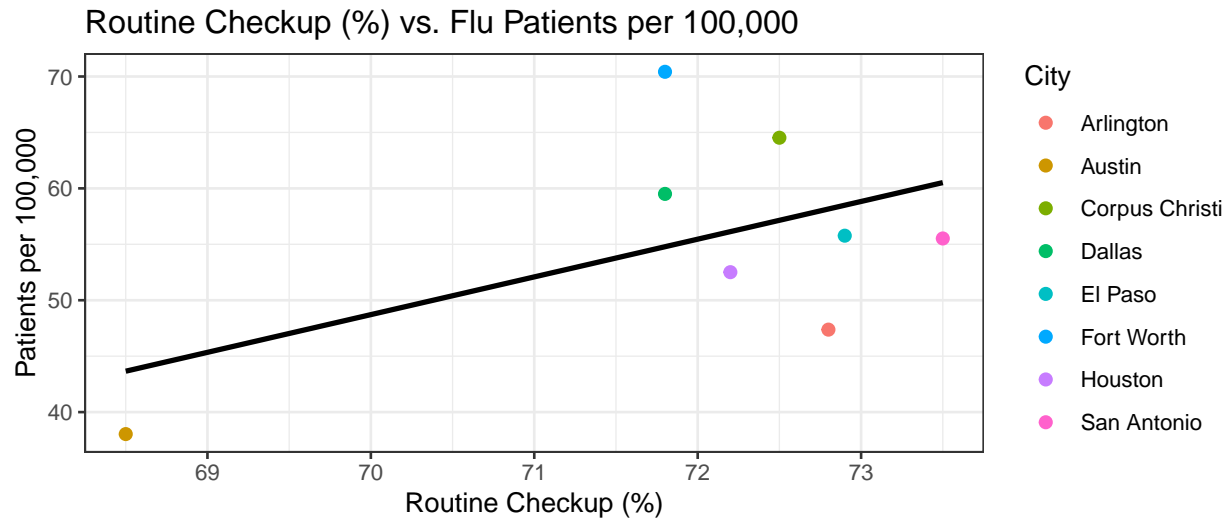


In fact, removing those two cities results in a correlation of -0.75.

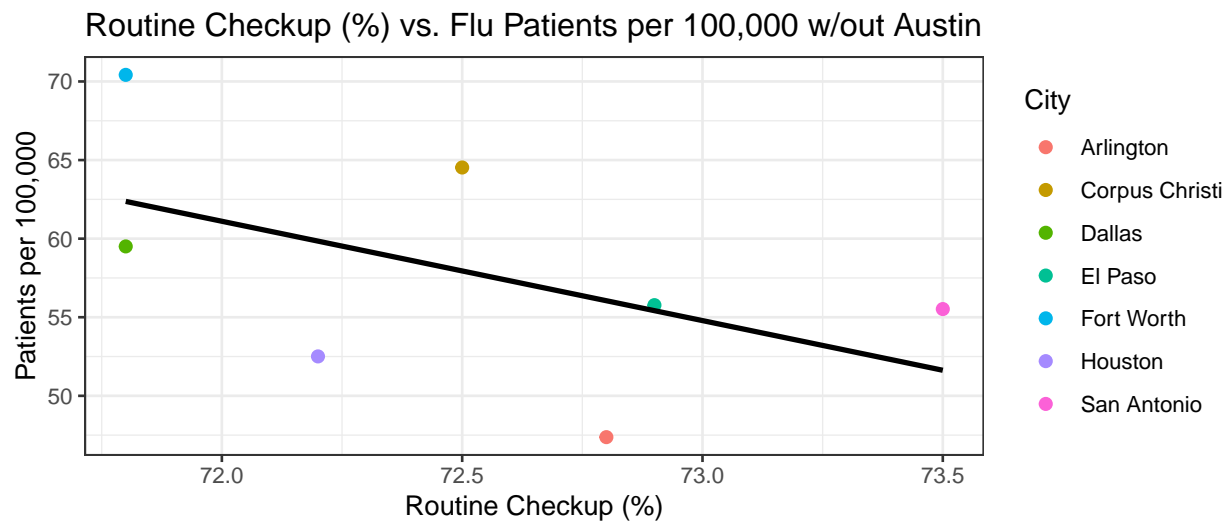


## Routine checkups

This graph kind of destroys any correlation this has because Austin “anchors” this line to have a positive slope. Furthermore, Austin actually has the least amount of Routine Checkup (%) so we should expect this graph to have a negative slope (more checkups means less patients).

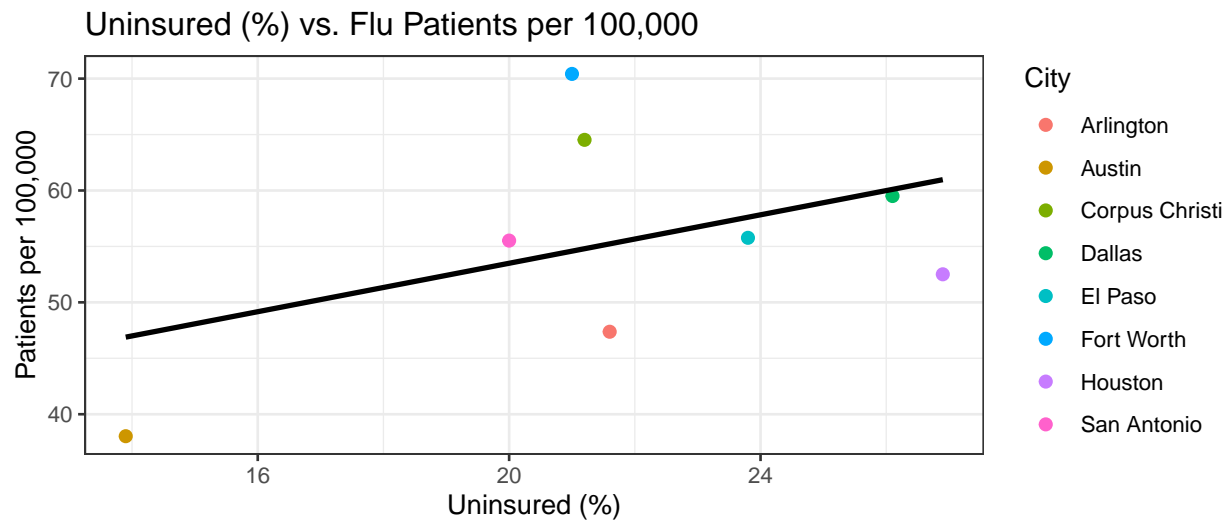


Let's try this graph without Austin and see what we get. This graph makes a lot more sense, not sure how accurate it is - we get a correlation of -0.51 after removing Austin (an upgrade over the previous 0.51, ironically just a flipped sign).

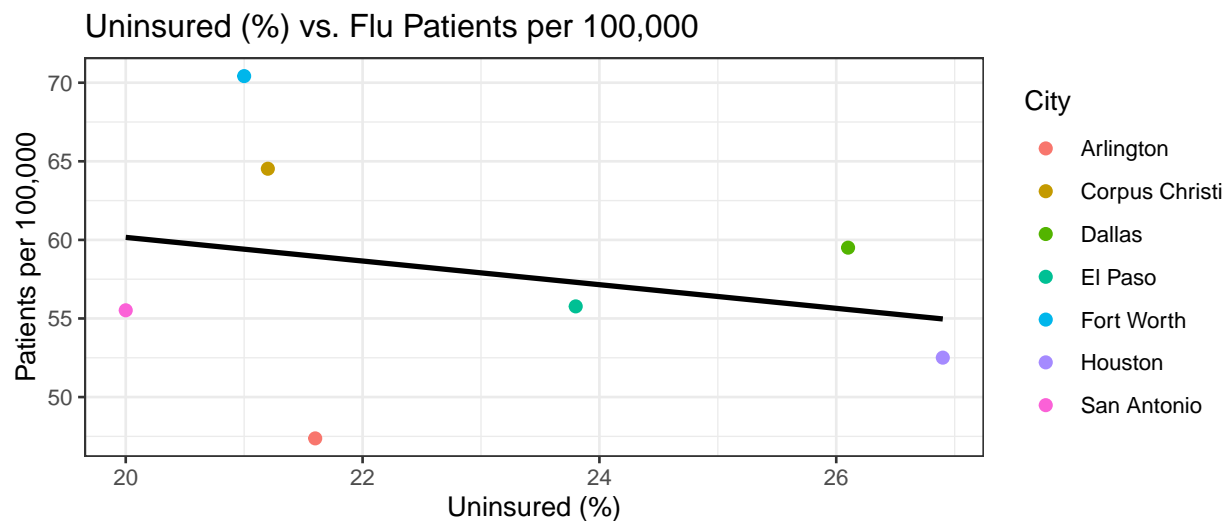


## Insurance

This variable looks like a general trend again anchored by Austin being very low. This time, the trend appears to make sense since less uninsured people should lead to less flu patients,



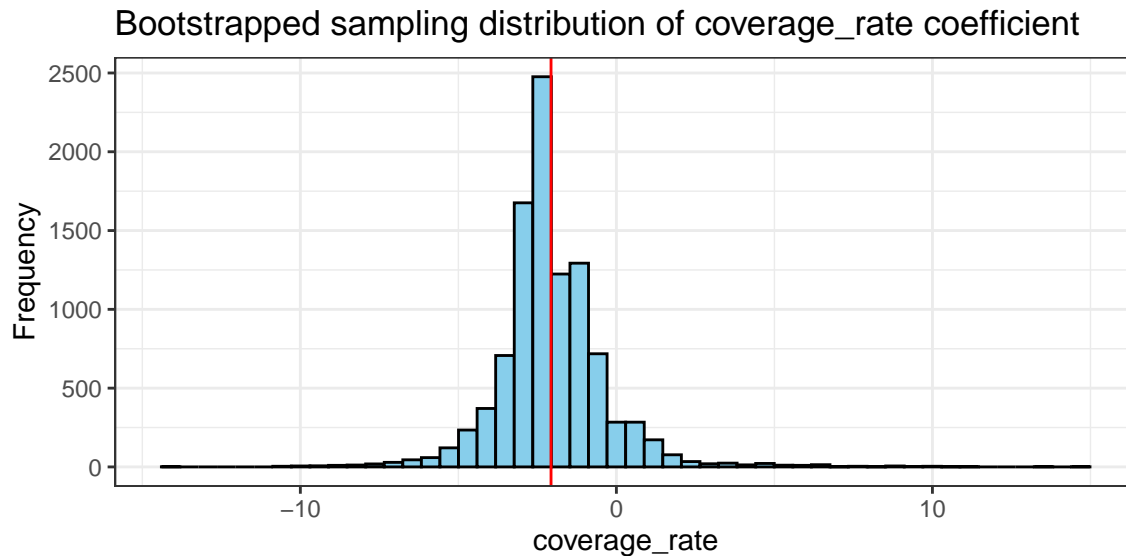
Aha! Without Austin, the correlation vanishes and is back to wishy-washy stuff. I definitely don't see a trend here, for what it's worth.



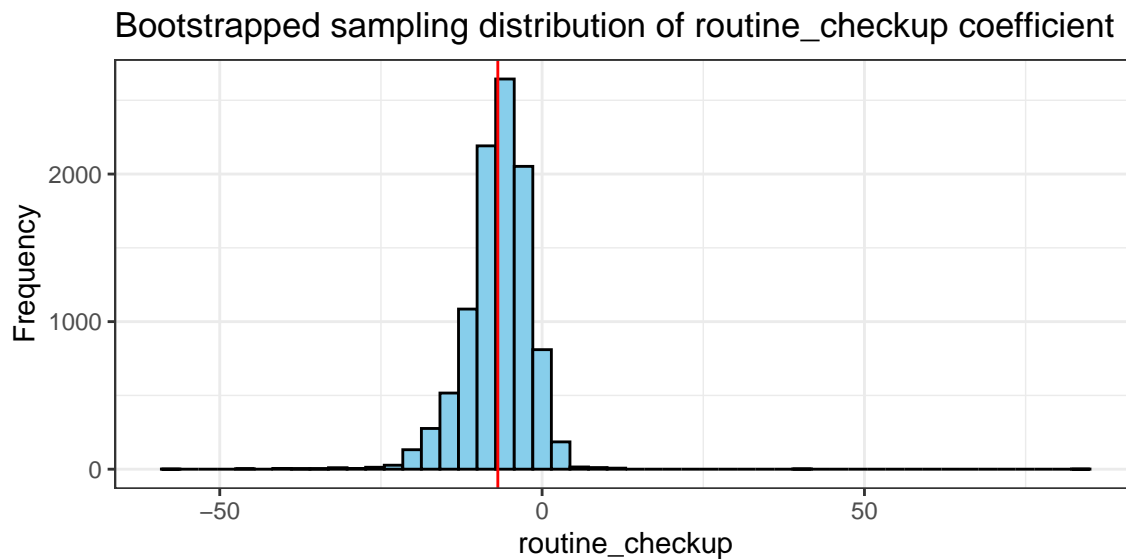


## Analysis

From this correlation matrix, I think we have sufficiently proved that vaccination coverage rate is absolutely crucial to understanding how many flu patients appear (at least in the 2022 spike). Furthermore, we can train a linear regression model and bootstrap the coefficient to get -2.07 as implying that with every 1% increase in flu vaccination coverage rate, the number of flu patients per 100,000 reduces by around 2.



For cities that are not Austin, the routine checkup % seems to make a difference, with more routine checkups implying less flu patients. We'll now do the same bootstrapping as before for this variable, except this time around we only resample from cities that are not Austin. Now, we get the value -6.87 as implying that with every 1% increase in the percent of the population that get routine checkups, the number of flu patients per 100,000 reduces by around 6 to 7.



Lastly, uninsured % seems to have no substantial difference, so we ignore this.

## City comparisons

Dallas vs. Austin

Vaccination coverage

Demographics

Politics and other qualitative aspects

Cycle by cycle analysis

Important changes in major Texas cities (2022-2024)

imma make a table of the pre and post and then the delta and then talk about major changes ## Key statistics by city

## Conclusion