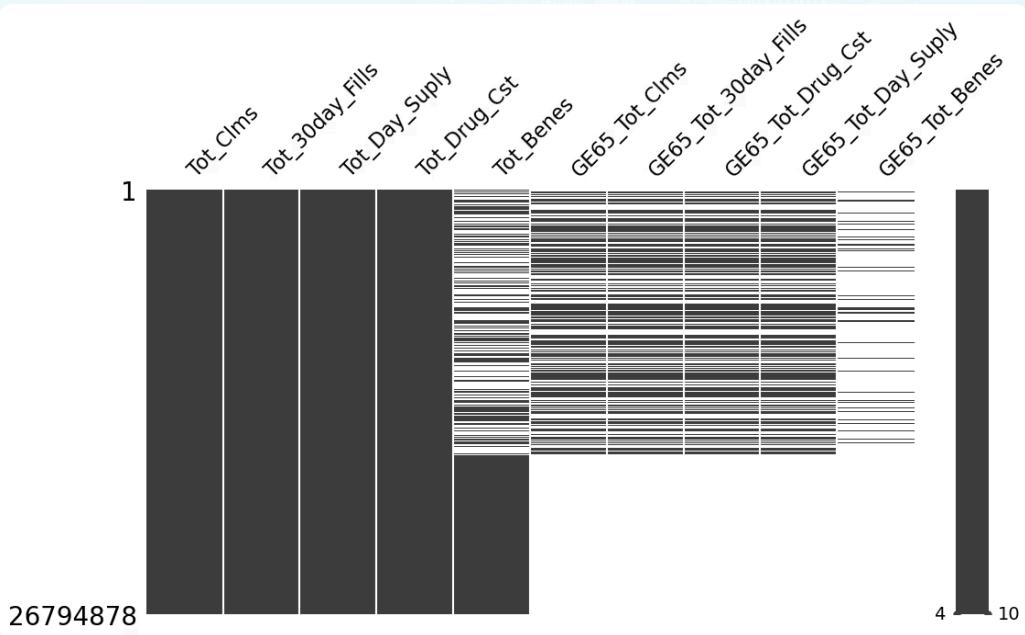


# Healthcare and Medicare Data

From Claims to Clarity: Visualizing Medicare Patterns & Provider Impact

 by Susan Schnitzel

## Evaluation and cleaning the data

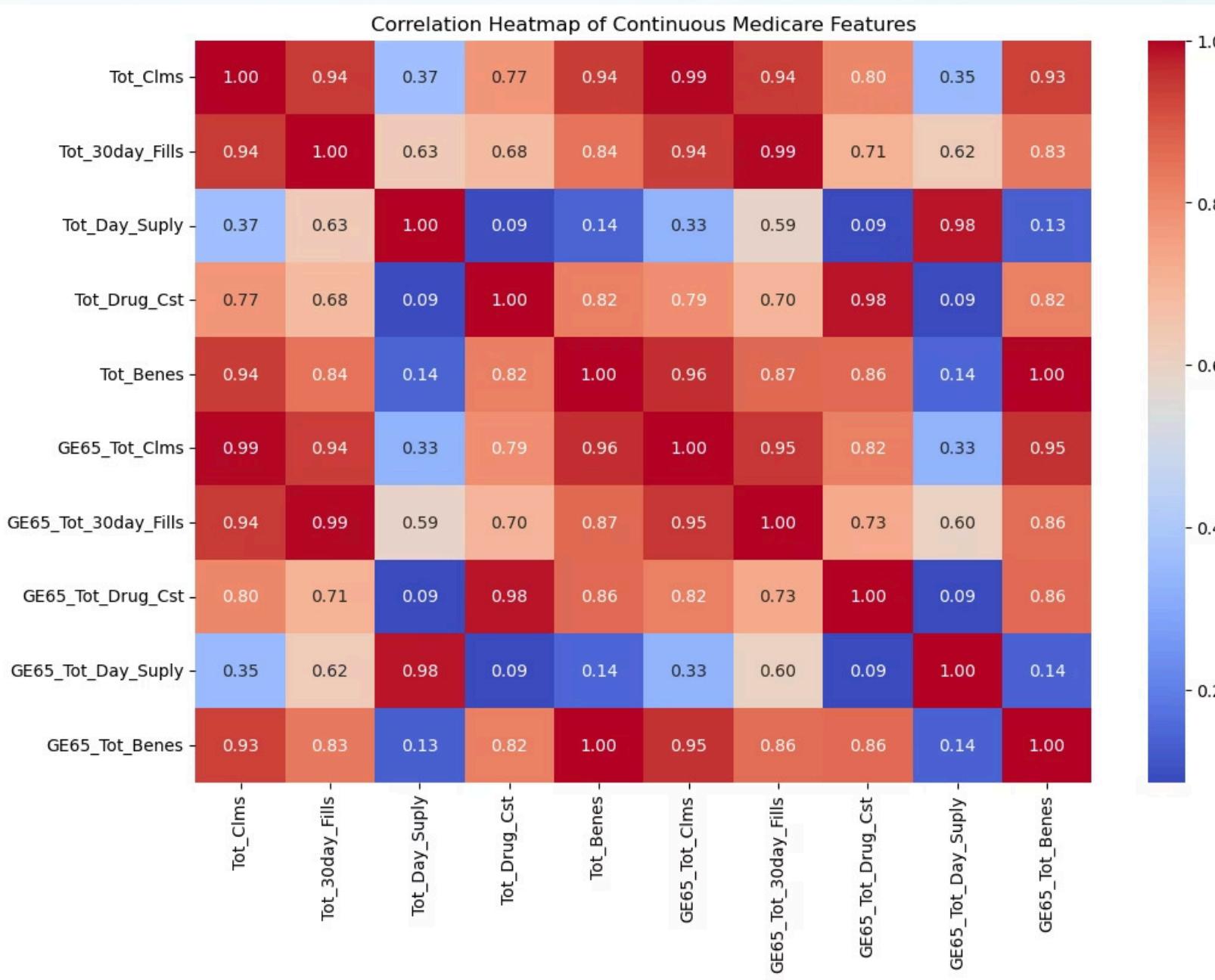


```
!pip install missingno import missingno as msno  
msno.matrix(df[continuous_cols], figsize=(12, 6))
```

*I will focus on the First Four Columns but will cleanup the "Total Benefits" to use for some visuals.*

# Patterns at a Glance:

## A Heatmap Exploration of Medicare Data

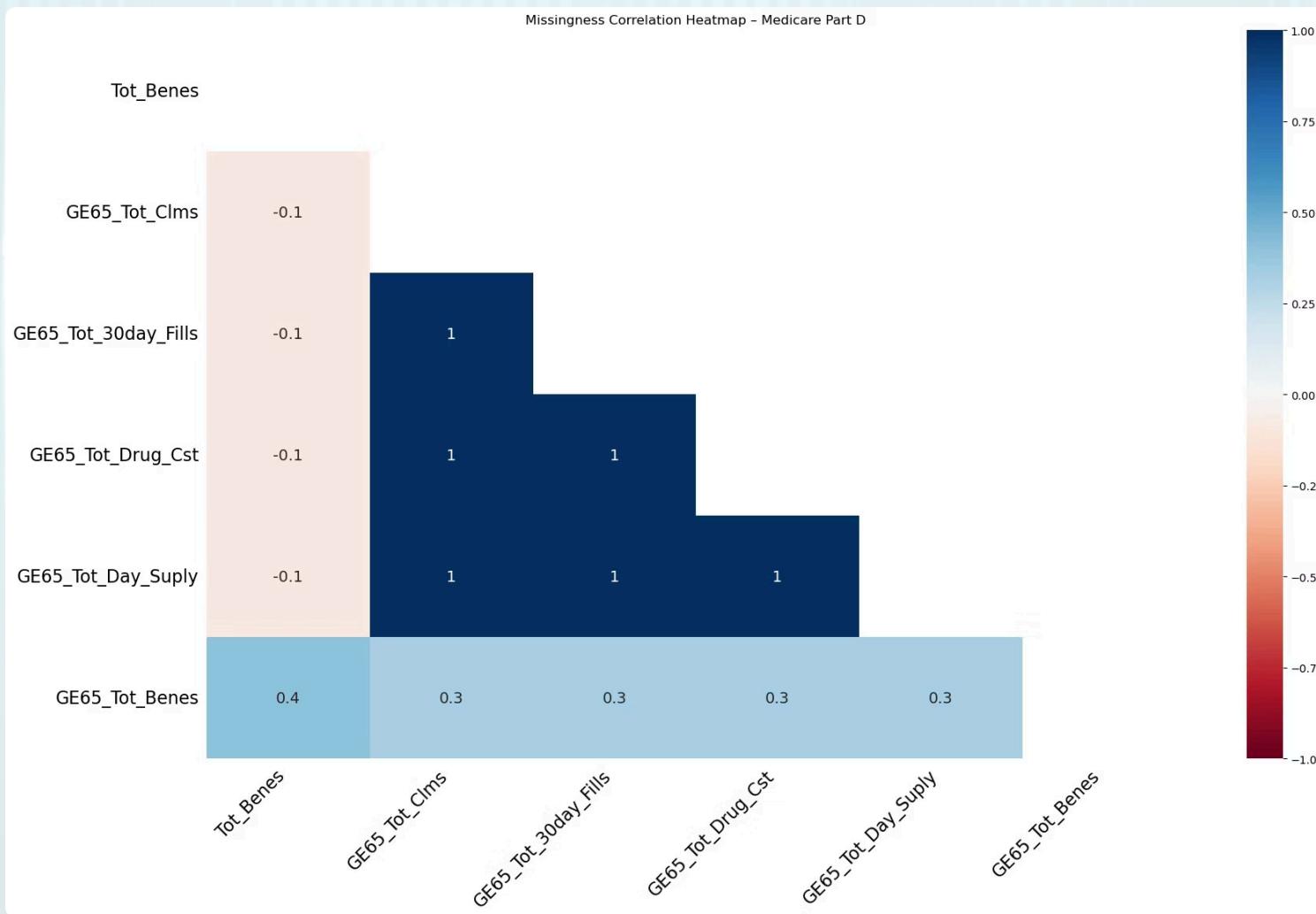


**Several trends emerged with some expected, others more revealing.**

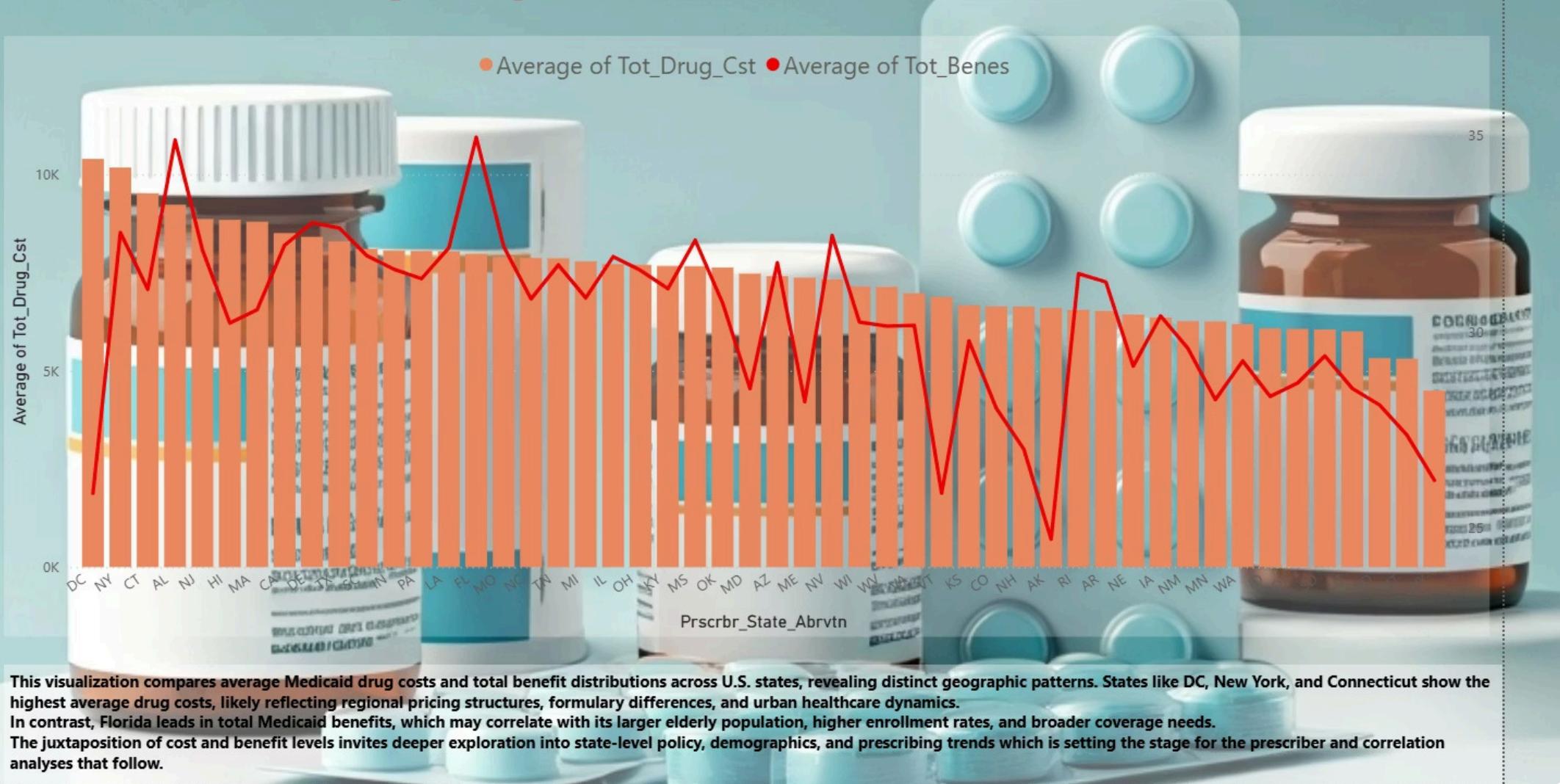
The GE65-titled columns show a strong correlation with the four key variables I initially planned to focus on. Across those four, the relationships are consistently strong, including with the Benefits column, which demonstrates notable alignment.

# Visualizing Data Gaps: Heatmap of Medicare Part D Variables

GE65 variables show perfect alignment, and potentially redundant or policy-driven. GE65 Benefits diverges slightly, suggesting a distinct cost pattern.



## Average Drug Cost and Benefits at US State Levels



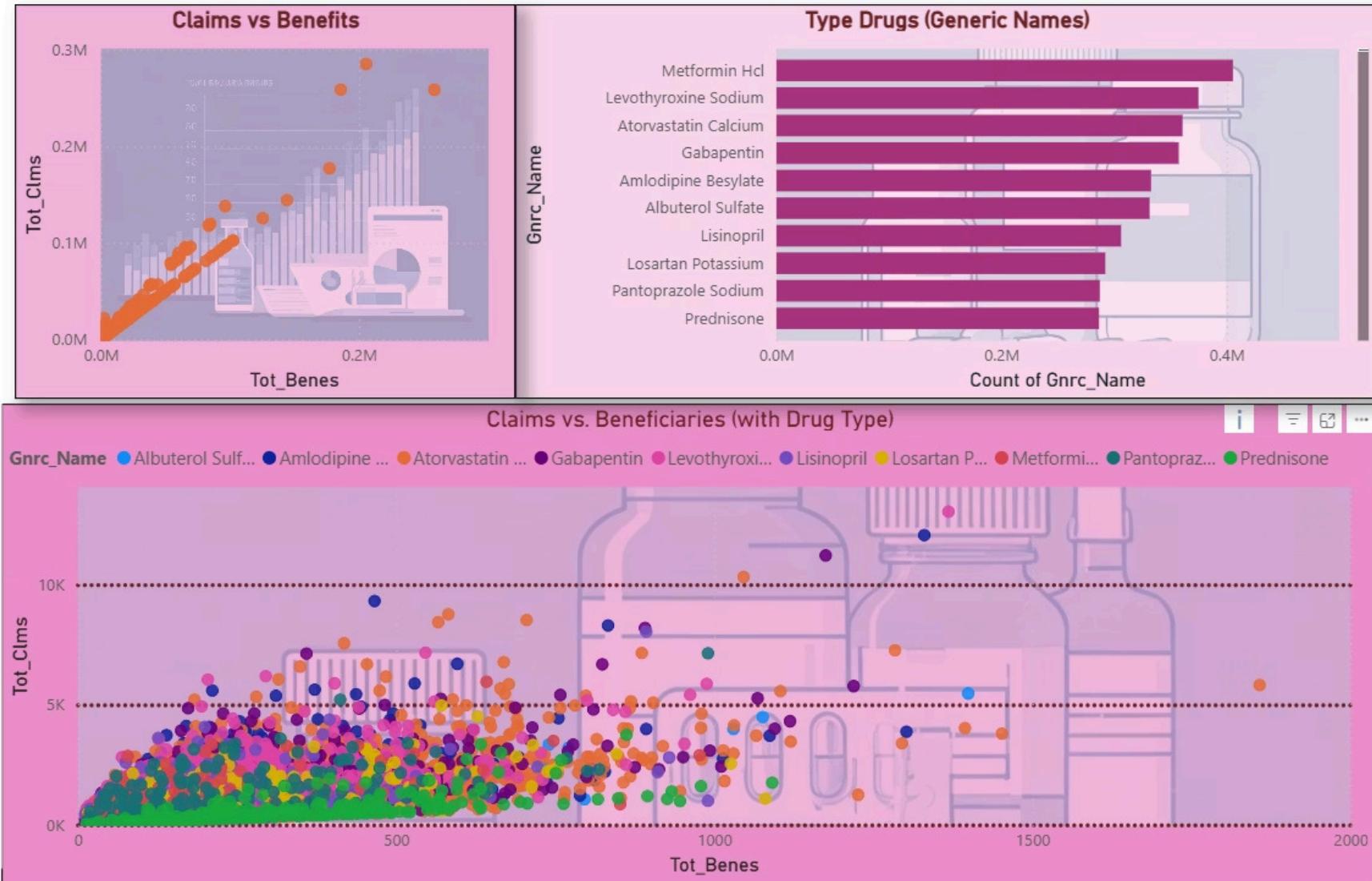


## Quick Look of "Claims vs. Benefits" and Top Drugs

Initial findings show a strong, direct correlation between total benefits and total cost.

To explore this further, I filtered out the top 10 drugs; see chart on top right.

After rerunning the correlation analysis, the relationship becomes more nuanced and slightly less linear. However, a meaningful correlation still persists, suggesting consistent cost dynamics even beyond high-volume prescriptions.



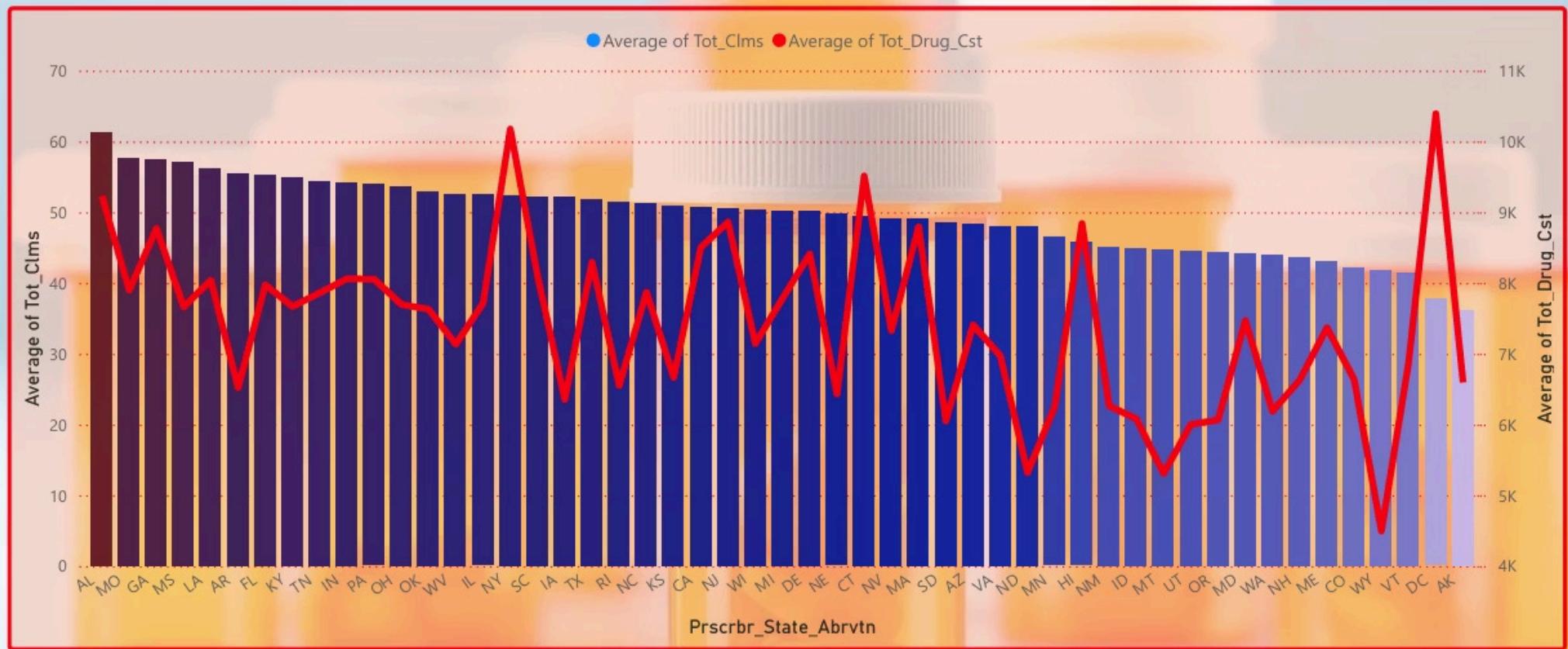
## A Snapshot of the Prescribers

*Some of the Prescribers will be combined or removed in future Analysis/Visuals*

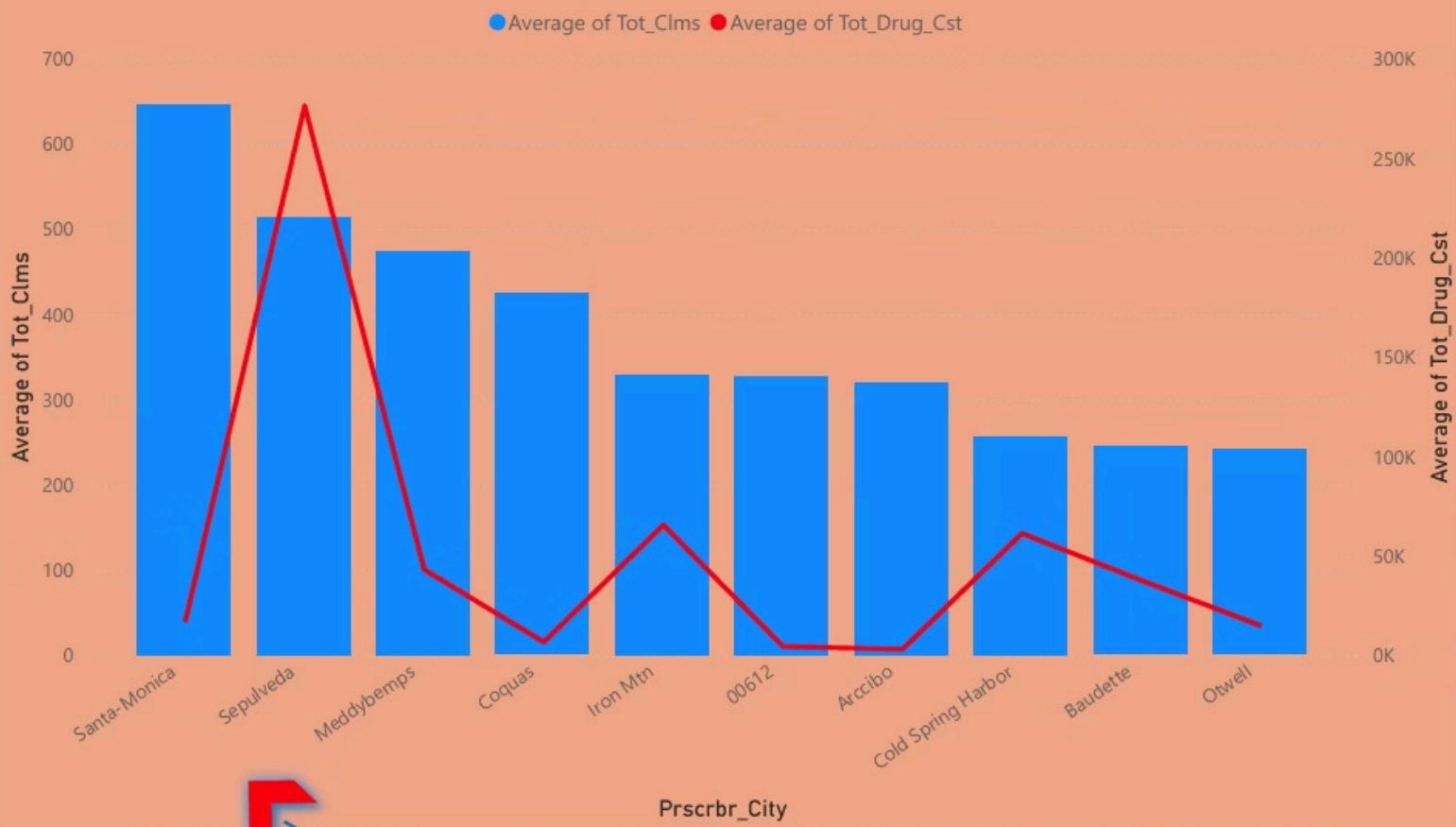


# Total Claims and Drug Cost vs. States

This visualization shifts focus to average claims (bars) against drug prices (lines). DC continues to spike in drug costs (as seen in previous visual), but Alabama emerges as the leader in total claims. This reframes the analysis toward utilization patterns and prescribing behavior. Slicers added on the next visuals to dig a little deeper.



## Average of Total Claims by CITY (US Territory)



North Hills, formerly known as Sepulveda, is a neighborhood in the San Fernando Valley region of Los Angeles, California. It was originally part of an agricultural community called Mission, and today it is also home to the Sepulveda VA Medical Center.

I'm a little concerned that the Drug Cost numbers here seem unusually high compared to other cities. This could point to a clerical error (perhaps repeated entries inflating the totals) or, on the more dramatic side, we might have stumbled onto something suspicious, maybe even uncovered a hidden ring. Personally, I lean toward the former, but I'll be taking a closer look at the data just to be sure.

Santa Monica is the leader in total claims. The other cities included ones that are all of the US territory, but also some with zip codes too. Messy Data still.

SLICER USED - ALL STATES  
ALL CITIES SHOWN

### STATE ABV

- Deselect...
- AK
- AL
- AR
- AZ
- CA
- CO
- CT

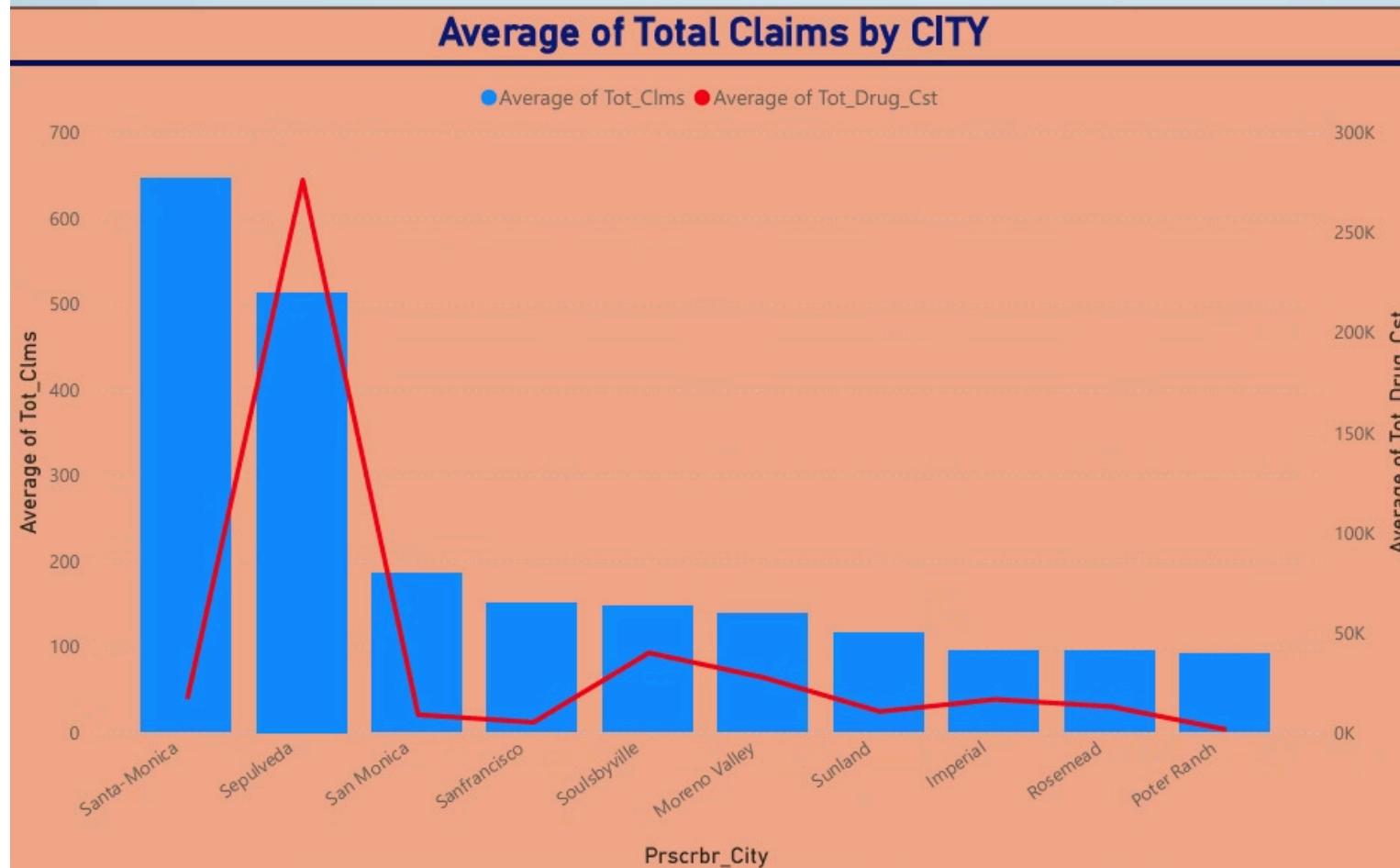


## Sliced on California

### Top Ten Cities based on Total Claims

As expected based on the previous slide, Santa Monica and Sepulveda are significantly higher in claims and I still believe there might be a clerical error with the drug costs for Sepulveda.

Selected CA only



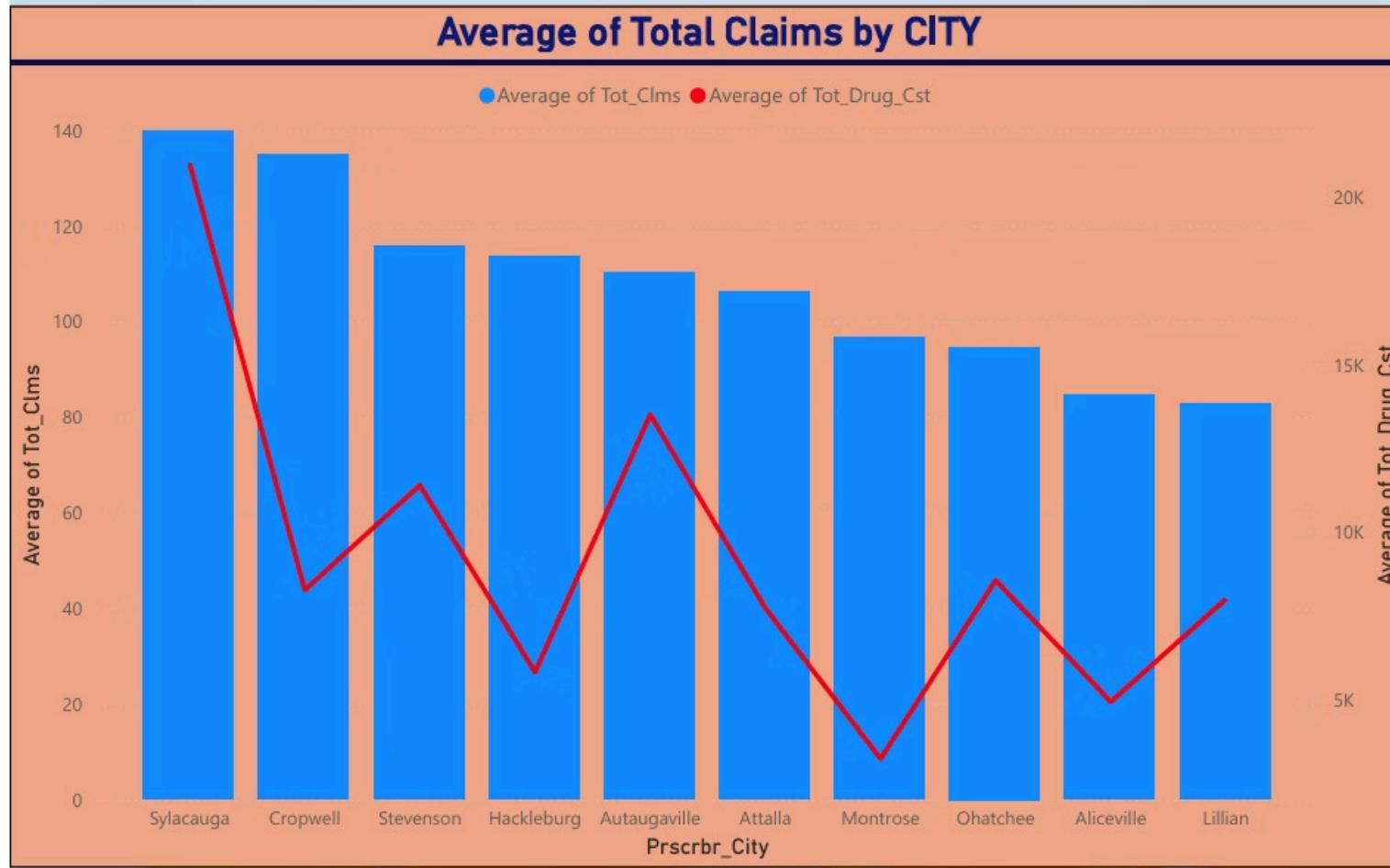


## Sliced on ALABAMA

### Top Ten Cities based on Total Claims

Alabama, in a previous visual, was the leading state for Claims, so wanted to dig deeper into the cities.

Roll Tide



Selected AL  
only

#### STATE ABV

- Select all
- AK
- AL
- AR
- AZ
- CA
- CO
- CT

#### Unusual Claim Patterns

Sylacauga, AL vs.

Sepulveda (North Hills), CA

Alabama leads the nation in total healthcare claims, and when I drill down into the city-level data, Sylacauga rises to the top. That's surprising on its own, but even more so when you consider that drug costs in Sylacauga are also notably high, not just the volume of claims.

For context, North Hills, CA (formerly Sepulveda) also shows up as an outlier, particularly in terms of average drug costs. (This can be found in a previous visual.)

Cropwell, AL isn't far behind Sylacauga in total claims, but its cost profile seems more in line with expectations. That contrast makes Sylacauga's numbers stand out even more.

Could this be a case of repeated entries or a data aggregation quirk? Possibly. But when two cities on opposite coasts both show unusually high drug-related claims, it's hard not to wonder if there's a deeper pattern ....or at least a shared reporting anomaly. Personally, I lean toward a clerical or structural explanation, but I'll be taking a closer look at the data.

Still, it's tempting to imagine we've just stumbled onto the opening scene of a healthcare noir.

**MORE TO COME.....**

