State-Level Analysis of Health Issues, Hospital Availability, and Healthcare Quality in the U.S.

BA780: Introduction to Data Analytics

Team 10: Suji Kim, Anna Baldzikowski, Linh Le, Shi Qiu, Tianqi Sun, Xiaoyan Wang

Goal & Data Preprocessing

Goal:

- Identify the most relevant health issues
- Assess elderly hospital accessibility
- Evaluate average nursing home quality ratings

Data Preprocessing:

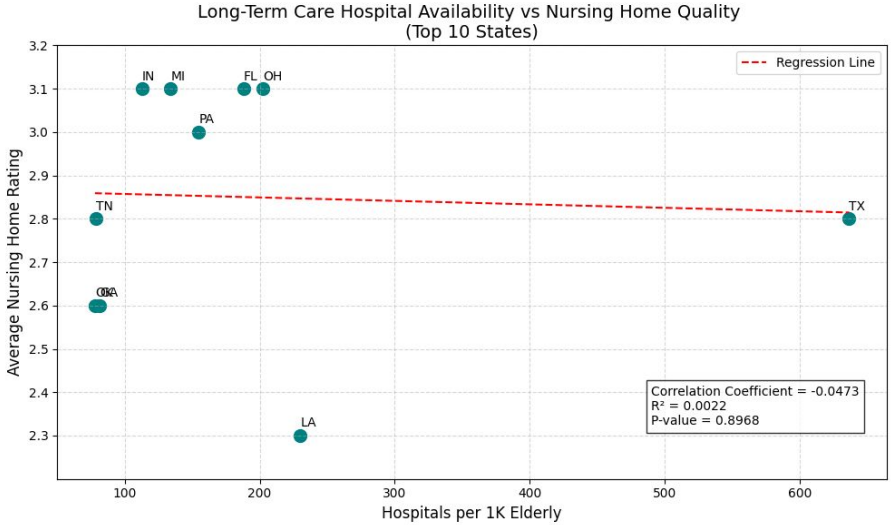
- Removed unnecessary columns
- Renamed columns
- Grouped by states
- Created dictionary and function to get region

Data Source

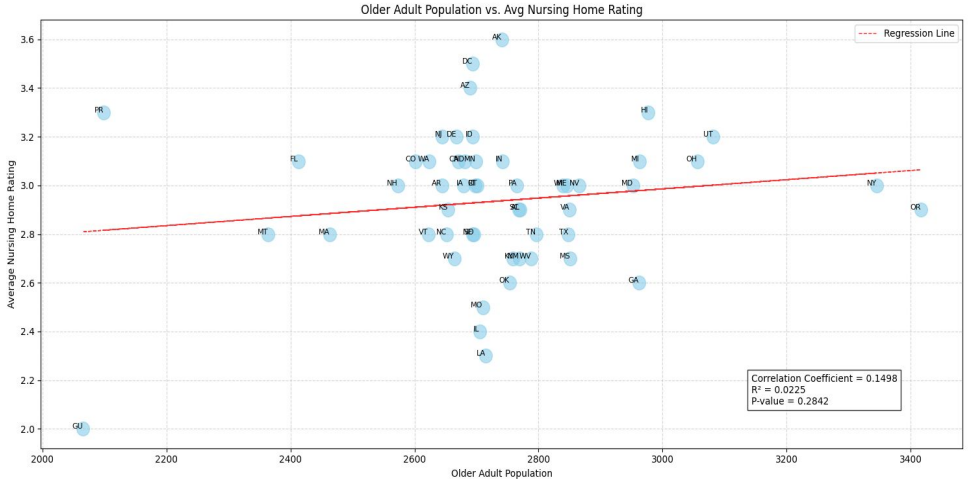
- **Alzheimer's Disease and Healthy Aging Data**
  - **Description**: Data including health issues, residence, age (adults over middle age), etc. of elderly living in the United States
  - **License**: CDC (BRFSS-based, Public Domain)
  - **Access**: Link (CSV, 138.4 MB)
- **USA Hospitals**
  - **Description**: Data including the names, types, locations, etc. of hospitals in the United States
  - **License:** U.S. Government Works (Public Domain)
  - **Access:** Link (CSV, 3.7 MB)
- **State US Averages**
  - **Description**: Data including various type of ratings of nursing homes in the United States by state
  - **License:** CMS (Public domain U.S. government data)
  - **Access:** Link (CSV, 23 KB)

Healthcare Accessibility and Quality for the Elderly

Does Higher *Hospital Accessibility* Mean Higher Rating?



Does *Higher Population Size* Mean Higher Rating?

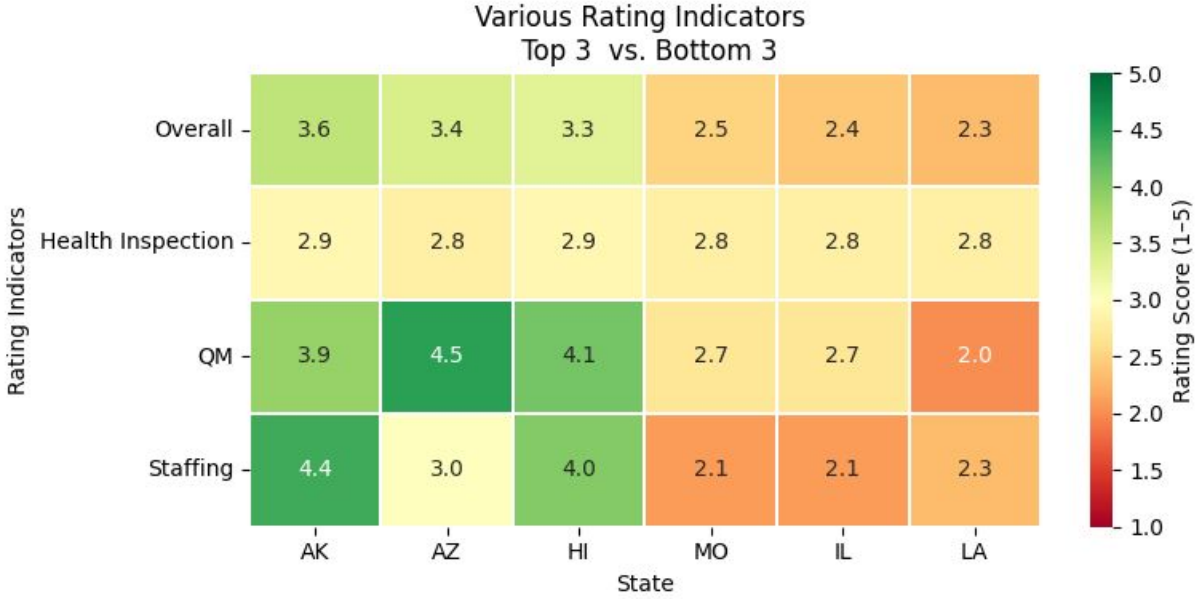


Almost No Correlation

(Quality Beats Quantity)

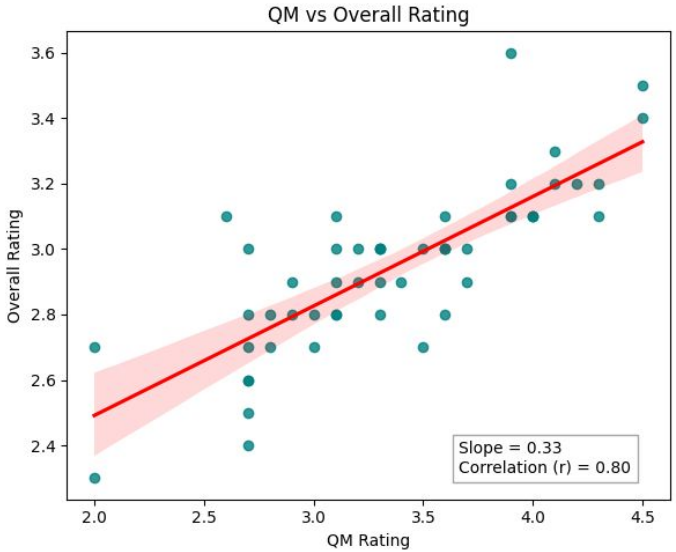
We calculated population by how many people participated in health surveys per state

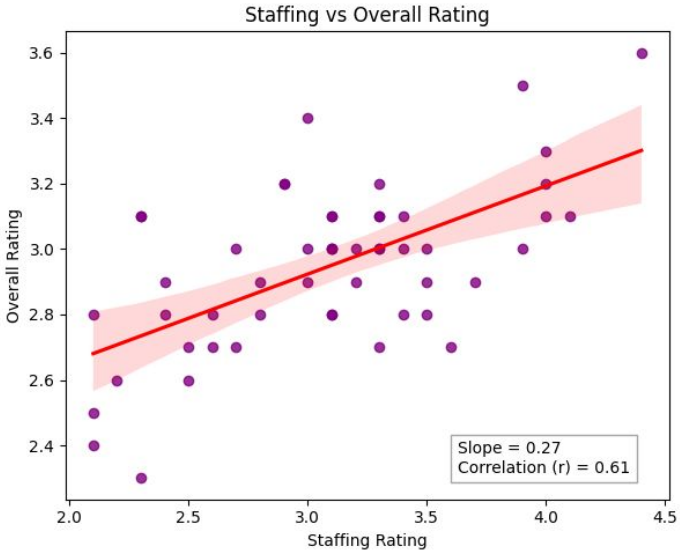
Which Rating Factor Has The Greatest Impact On Overall Rating?



QM vs. Staffing?

How Well Does This Model Fit The Dataset?





– Slope –
QM > Staffing
“QM has a greater impact”

– Correlation –
QM > Staffing
“QM has a stronger linear relationship (strength & direction)”

How Well Does This Model Fit The Dataset?

OLS Regression Results

Dep. Variable:	Overall	R-squared:	0.909
Model:	OLS	Adj. R-squared:	0.906
Method:	Least Squares	F-statistic:	245.6
Date:	Wed, 15 Oct 2025	Prob (F-statistic):	2.90e-26
Time:	14:47:12	Log-Likelihood:	61.033
No. Observations:	52	AIC:	-116.1
Df Residuals:	49	BIC:	-110.2
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.1897	0.081	14.766	0.000	1.028	1.352
QM	0.3091	0.018	16.977	0.000	0.272	0.346
Staffing	0.2328	0.019	12.206	0.000	0.194	0.271

Omnibus:	2.133	Durbin-Watson:	2.030
Prob(Omnibus):	0.344	Jarque-Bera (JB):	1.257
Skew:	-0.255	Prob(JB):	0.533
Kurtosis:	3.567	Cond. No.	36.3

Ordinary Least Squares Regression

R^2: 0.909

Adj. R^2: 0.906

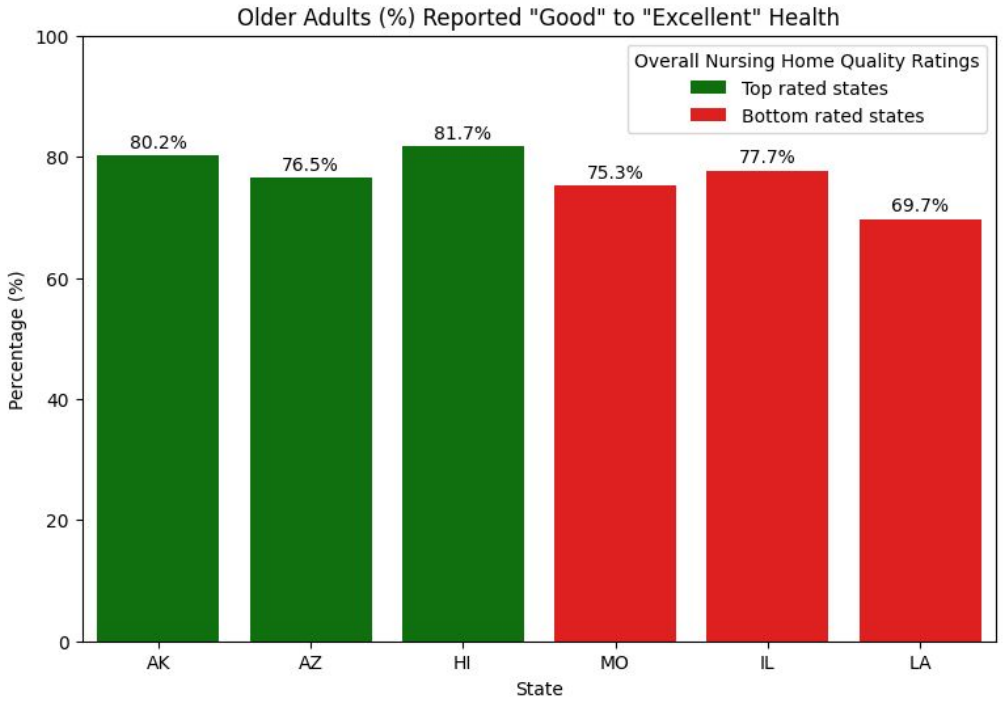
→ Linear model can explain approximately 90% of the data variance

F-statistic: 245.6

Prob (F-statistic): 2.90e-26

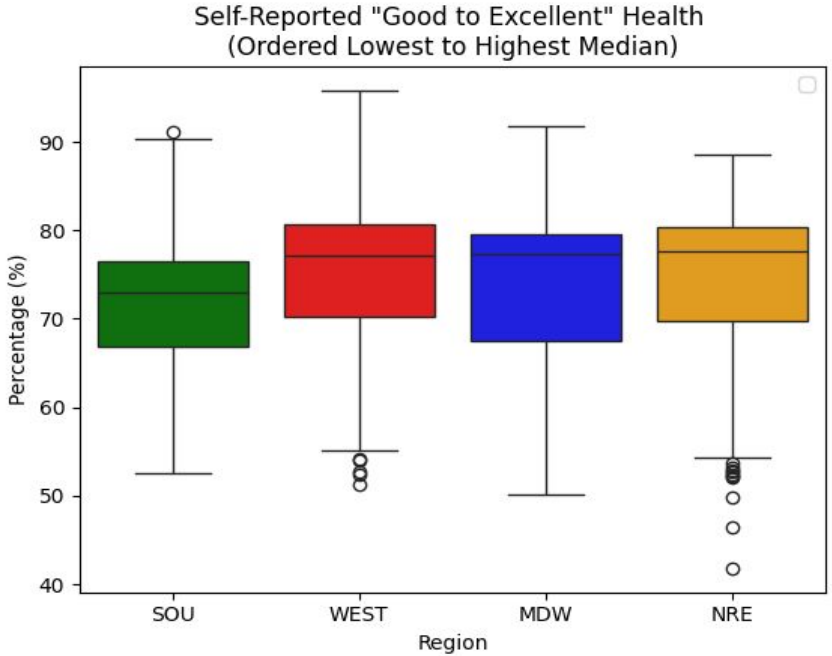
→ Confirm the statistical significance of this model

Do Higher Ratings Reflect Better Self-Reported Health?



Self-Reported “Good” to “Excellent” Health

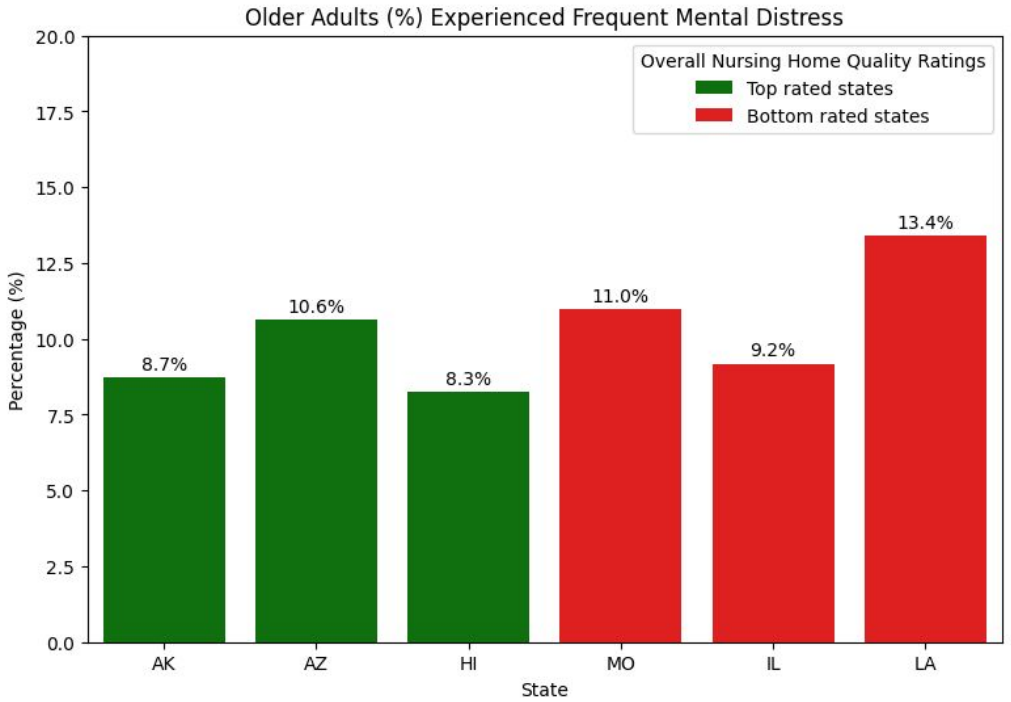
Visualizing Self-Reported Health Regionally



Range from **73.0% to 77.6%**

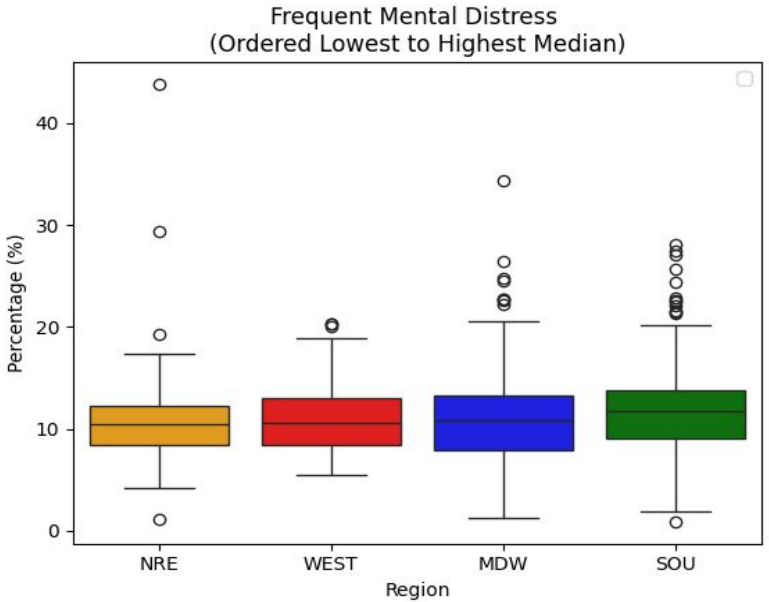
Northeast performs best,
South performs worst

Do Higher Ratings Reflect Lower Mental Distress?



Experienced Frequent Mental Distress

Visualizing Mental Distress Regionally



Not much variation (10.5%-11.7%), but Northeast performs best, South performs worst (we want percentage to be low)

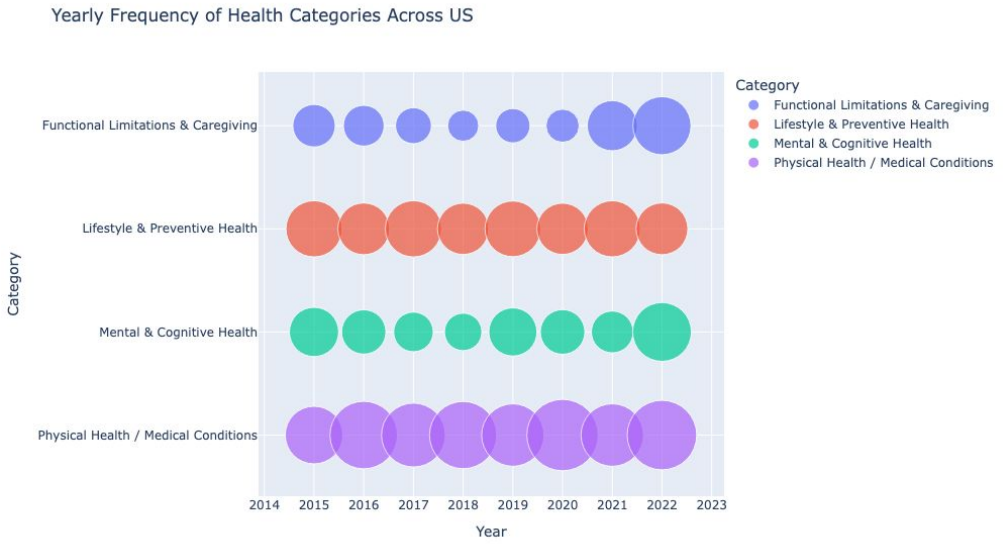
Takeaways:
Improving nursing home rating scores may have a positive impact on older population's perceptions of their health, and may lower their levels of mental distress.

Overall, the South needs to be prioritized for these indicators.

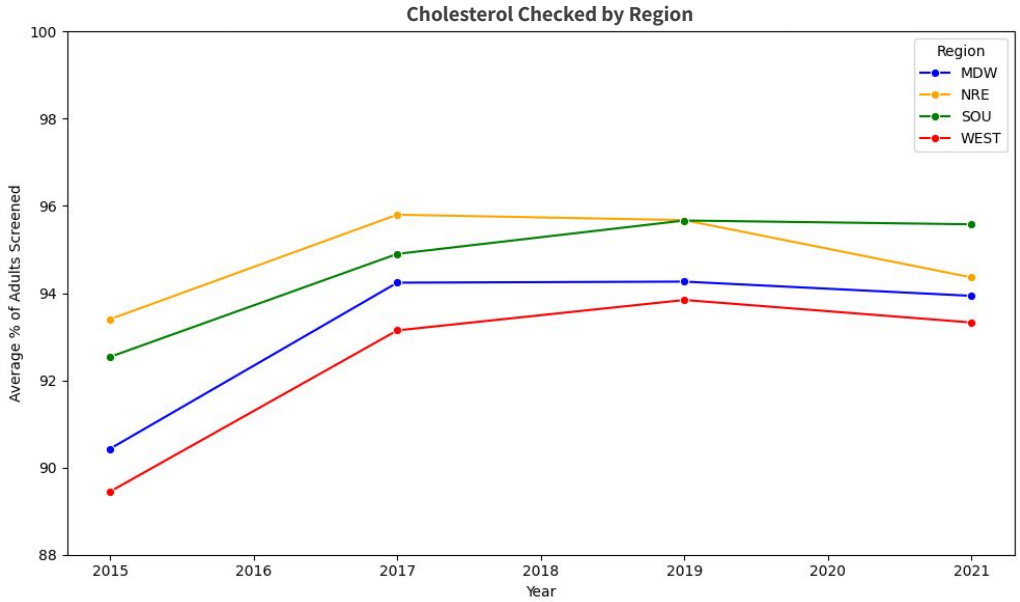
Most Prevalent Health Issues

The Most Prevalent Health Category & Issue

- **The Most Prevalent Health Category:** Physical Health/Medical Conditions
- **The Most Prevalent Health Issue:** Cholesterol Screening

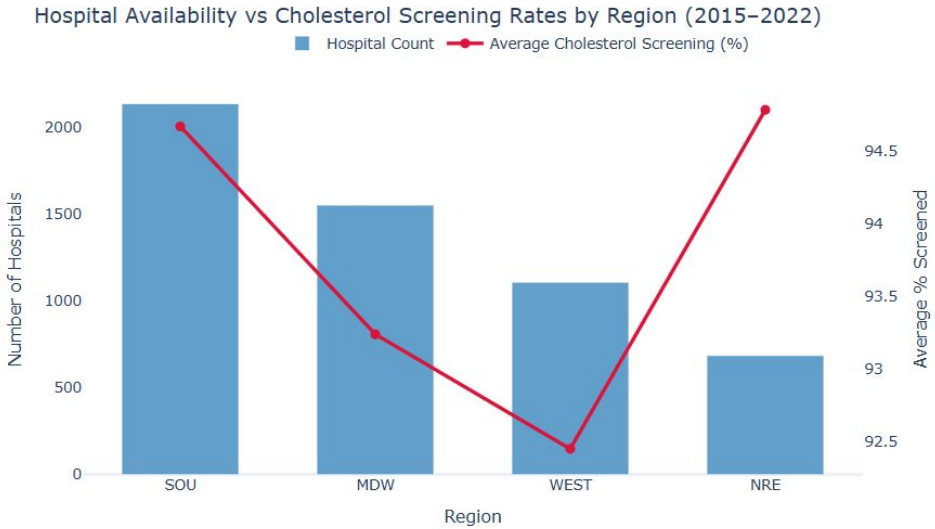


Cholesterol Screening by Region



Overall, very high screening rates (over 90%) remain across all regions, suggesting that preventive health screenings are widespread across the region.

Cholesterol Screening by Region



Filtered for General Acute Care, Critical Access, and Women's Hospitals from hospital types, likely to perform cholesterol screenings.

Quantity of Healthcare Infrastructure

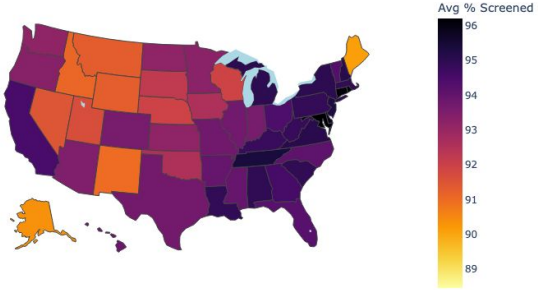
≠

Increase Participation Health Screening

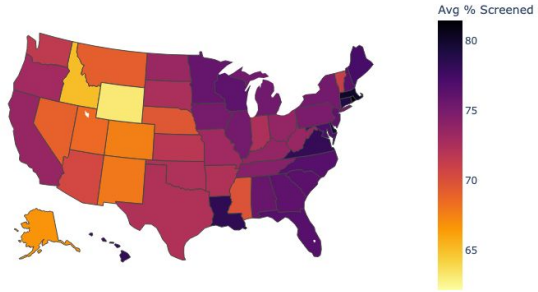
The U.S. government should prioritize support in areas with low screening efficiency by integrating and analyzing **regional healthcare infrastructure** and **health behavior data**.

Similar Trend of Cholesterol and Mammogram Screening

Average % of Cholesterol Screening by State (2015–2022)



Average % of Mammogram Screening by State (2015–2022)



- **Physical Health/Medical Conditions:**
  - Cholesterol and Mammogram Screening identified as key preventive indicators within this category
- **Screenings Patterns:**
  - Higher in the Eastern & Southern regions
  - Lower in the rural Western regions
  - Shared structural or behavioral factors influencing preventive health behavior.
- **Southern Region:**
  - Greatest number of hospitals ≠ highest cholesterol screening rates
- **Healthcare quantity alone does not ensure effective preventive care.**

Policy / Management Ideas

The U.S. Government

Focus on quality of Hospitals, not just quantity

Improve Quality Measures at nursing homes

Invest in regional equity

(South for health surveys, rural West for screenings, etc)

Promote preventive health & education

Thank You