State-Level Longevity Trends with the US Mortality Database

Longevity 17 2022 Waterloo

Doris Padilla and Mike Ludkovski

September 13th, 2022

University of California, Santa Barbara

Overview

Big Picture

The goal of this project is to investigate and effectively understand **state-level, age-specific** mortality trends throughout the US. Exploratory analysis shows that the mortality experience across different states is highly nonhomogeneous, offering opportunities for rich insights about drivers of mortality.

Steps

To achieve this objective, we make use of:

i) *Machine Learning Tool*: Multi-Output Gaussian Processes (MOGP).

ii) *State-Grouping Selection Procedure*: an algorithm which minimizes the 'distance' between neighboring states.

Table of contents

1. Data and Statistical Model

United States Mortality Database (USMDB)

Data Collection:

- The United States Mortality Database (USMDB) [1] offers a complete set of state-level life tables.
- Mortality data is organized by state, age, year, and gender.

Subsets of Interest:

Most applicable to actuarial research:

- *Ages:* 64-84 (retirement).
- *Calendar Years*: 1990-2018.
- *Population:* Males and Females.

Our goal is to model state-level mortality rates as a function of age and time (surface) $f_{\text{STATE, GNDR}}(x^n_{ag}, x^n_{yr})$.

[1] https://usa.mortality.org

Motivation for Data Pooling

Motivation for Joint Mortality Modeling:

1) Mortality data is **noisy** (especially for smaller states).

- Careful data-fusion can reduce random fluctuations and output a smoothed (accurate) surface.

2) Mortality experiences are **highly correlated**.

- Accepted hypothesis that states which have 'similar characteristics' also share an equivalent mortality outlook.

3) **Other Options:** *Single models* are not stable in estimating improvement factors (not consistent) and inputting *all states* into one model is computationally intractable and unnecessary.

Data Assumptions to use MOGP Modeling

Summary

Implement a spatial statistical framework of *Multi-Output Gaussian Processes (MOGP)* regression as a machine learning method for multi-population modeling.

Observation Likelihood

• Relationship between y_l^n (observed mortality rate for state l) and x^n (age and year) is described with a *latent function*:

$$y_l^n = f_l(x^n) + \epsilon^n$$

where the observation noise is Gaussian.

• Specify a prior distribution; $f_l \sim \mathcal{GP}(m_l, C_l)$. Use the data to compute the *posterior distribution* via MVN conditioning. Analytic results allow us to impose uncertainty quantification around future trajectories.

Multi-Output Gaussian Processes (i)

Covariance Function

The kernel captures the *dependence* of the response surface f_l on the varying Age and Year dimensions; **distance-based** in order to capture the expectation that mortality should be similar to its neighboring points.

$$C_l(x^n, x^m) = \eta^2 \exp\left\{ -\frac{(x^i_{\text{ag}} - x^j_{\text{ag}})^2}{2\theta^2_{\text{ag}}} - \frac{(x^i_{\text{yr}} - x^j_{\text{yr}})^2}{2\theta^2_{\text{yr}}} \right\}$$

Separable between Age and Year factors. It's fitted through MLE.

Multi-Output Modeling

Here, we consider data from *multiple populations* simultaneously $(\#L)$.

- Each population is a covariate. Jointly modeling surfaces on top of each other (stacked).

Multi-Output Gaussian Processes (ii)

- The fused covariance matrix becomes:

$$\mathcal{C}(x^n, x^m) = \underbrace{C_l(x^n, x^m)}_{\text{Covariance over Age \& Year}} \times \overbrace{\Gamma_{l_1,l_2}}^{\text{Cross-State Covariance}}$$

Requires estimation of $\binom{L}{2}$ parameters θ_{l_1,l_2}.

Kernel Dimension Reduction: ICM approach [2]

• Assume each surface is a linear combination of Q latent independent GPs $u_q(x), 1 \leq q \leq Q$ with shared covariance $C^{(u)}$. That is, $f_l(x) = \sum_{q=1}^{Q} a_{l,q} u_q(x)$.

• Number of parameters estimated in cross-population covariance is $Q \times L$ by:

$$\text{Cov}\big(f(x), f(x')\big) = B \otimes C^u(x, x')$$

$B \equiv \sum_{q=1}^{Q} a_q a_q^T$ has rank Q.

[2] (Alvarez et al 2011)

Optimal groupings

Grouping States

1) Targeted group sizes 3-4 (large enough population sizes).
2) ICM Rank $Q = 2$ (assumption).
3) Will create *over-lapping groupings*: different one for each state - no partitioning.

For a fixed state *s*, we calculate an 'optimal group' by using (a) a distance metric and (b) enforcing geographical contiguity.

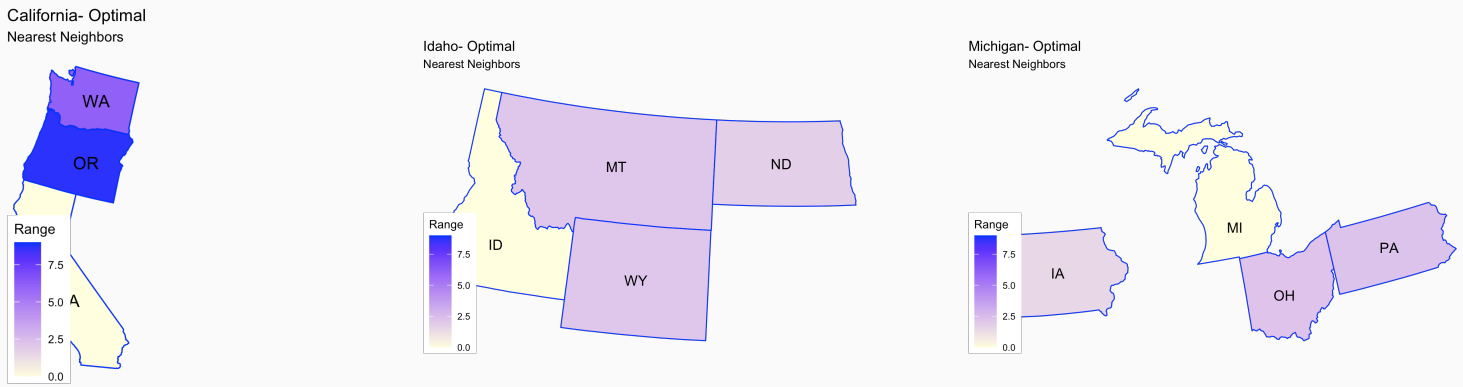


Figure 1: Optimal Groupings

Note: gradient = PC factors.

2. (Optimal) Model Grouping

Covariates

Overview:

We collect data on several state-level variables (ones which possibly have explanatory power in **observable mortality discrepancies**), then we run PCA analysis on this collection of covariates, and use the PCA factors to customize a similarity distance metric between neighboring states.

Assumption:

We assume that locations which are 'similar' in several measurable (*economic, demographic, biomedical* and to some extent *demographic*) characteristics, share akin mortality experiences.

Covariates

Collection of covariates are, (1) **Economic variables:** median income, regional price parities. (2) **Demographic variables:** % without insurance, % elderly (3) **Geographical variables:** average temperature, land in farms. Total amount = 18 (6 each category).

PCA Analysis

PCA Factor Loadings

Results show a strong relationship between economic, geographic, and possibly demographic variables with state-level variability.

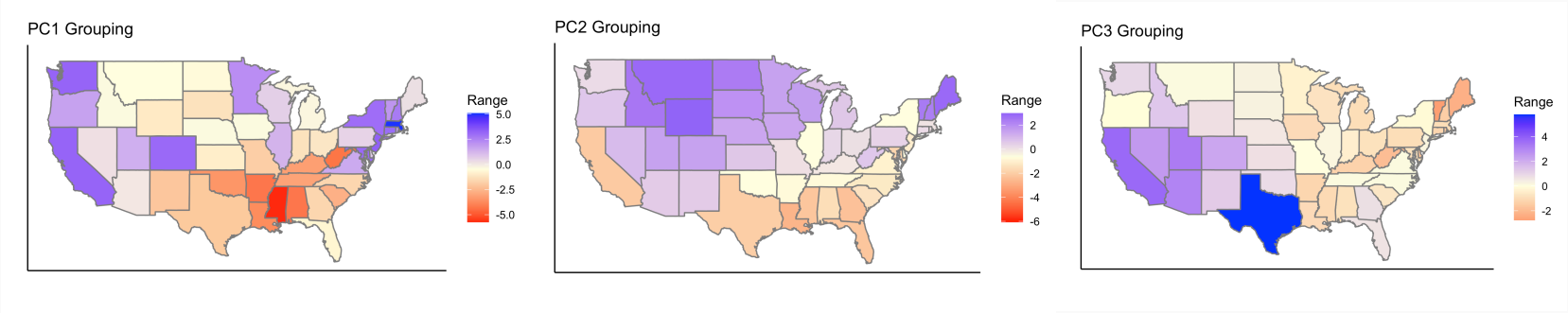


Figure 2: State-wise PCA factor loadings PC_k, $k = 1, 2, 3$.

- *PC1 Score:* Economic Variable; (DC, NY, CA) vs (MS,LA).
- *PC2 Score:* Climate Variable; North vs South.
- *PC3 Score:* Sun-Belt states; SW, Texas, Florida (pop. growth).

3. Results and Interpretation

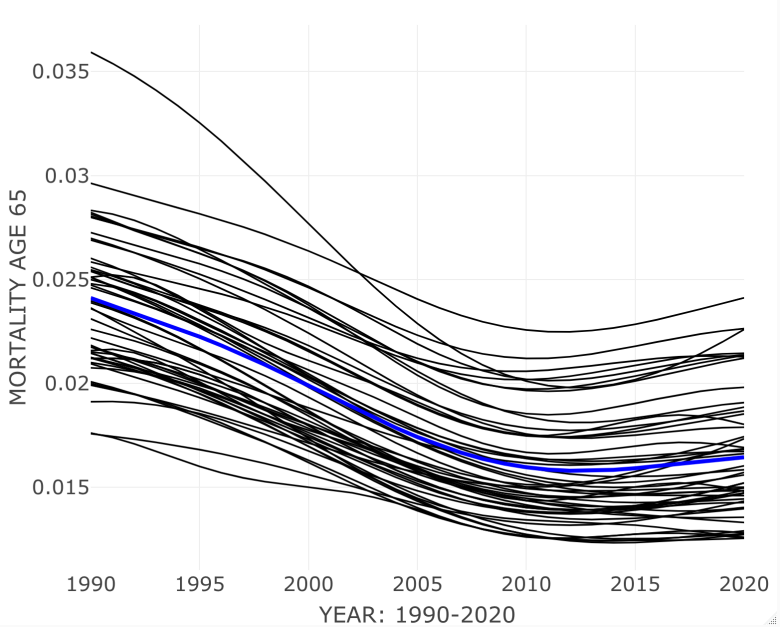
Exploratory Analysis

Next, we analyze the MOGP outputs and investigate:

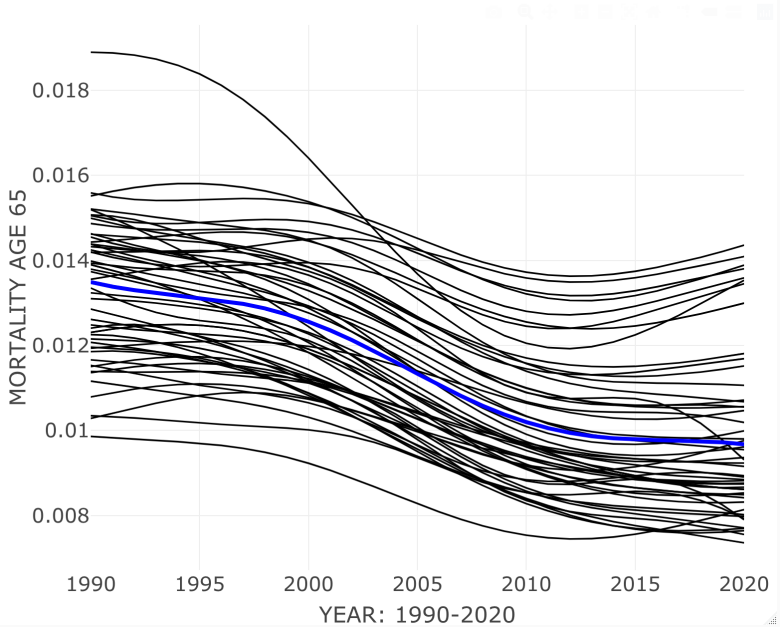
1) State-Level Mortality Rates and Aggregate Behavior.
    - Both across Ages and Time.
2) Improvement Structure (MI).
    - Understand the year over year (YOY) changes in mortality; i.e. extrapolate the recent and relative experience across all 50 U.S states.
    - *Note:* impossible to estimate well without a statistical model.

Mortality Predictions (Time: 1990-2020)

Smoothed mortality rates as a function of time.



Males Age 65



Females Age 65

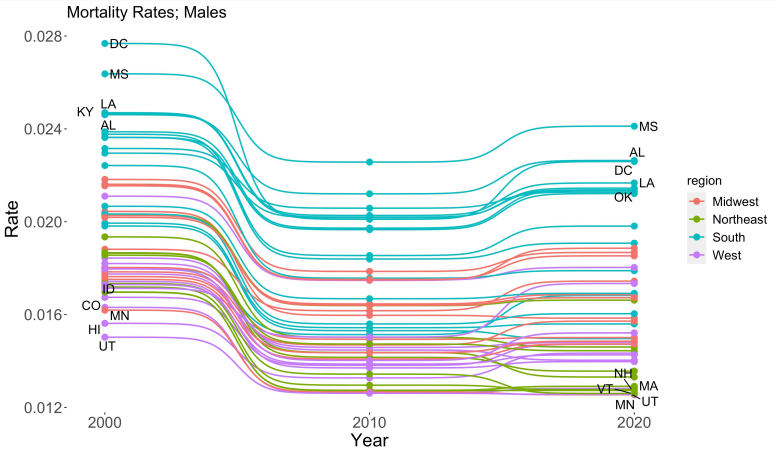
Figure 3: Multi-Output GP Mortality Rates. US national average in blue.

- Mortality evolution is variable across states (similar but not the same). Many 'cross-overs'; relative ranks change throughout time.

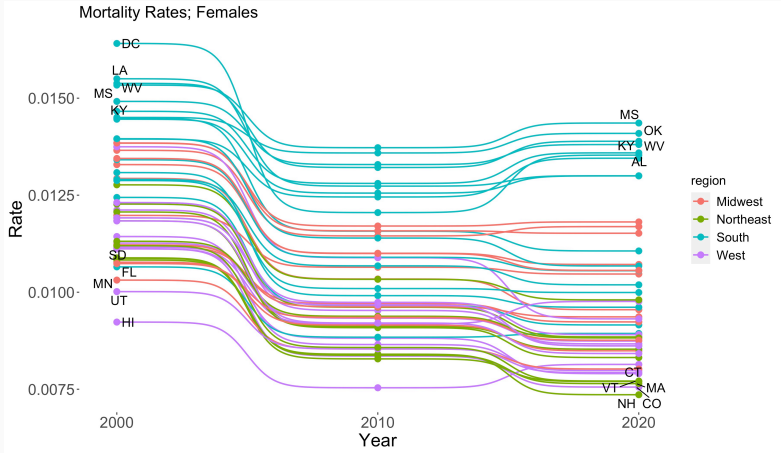
Bump Charts Comparing Mortality

Ranking Mortality across U.S. States between 2000, 2010, 2020

To better understand the heterogeneity in our mortality predictions, we compare state ranking across decades.



Males Age 65



Females Age 65

Figure 4: Bump Chart Comparing Mortality Rates in 2000, 2010, 2020.

- Consistent improvement between 2000 and 2010; however, post-2010 analysis is not as straightforward.

Summary of Aggregate Mortality Behavior

(1) Rapid improvement until 2010. In the past decade, mortality rates have been consistently worsening (or flattening) in all locations except these:

Alaska	Delaware	Illinois
Maine	Minnesota	New Hampshire
New Jersey	New York	Rhode Island
Utah	Vermont	

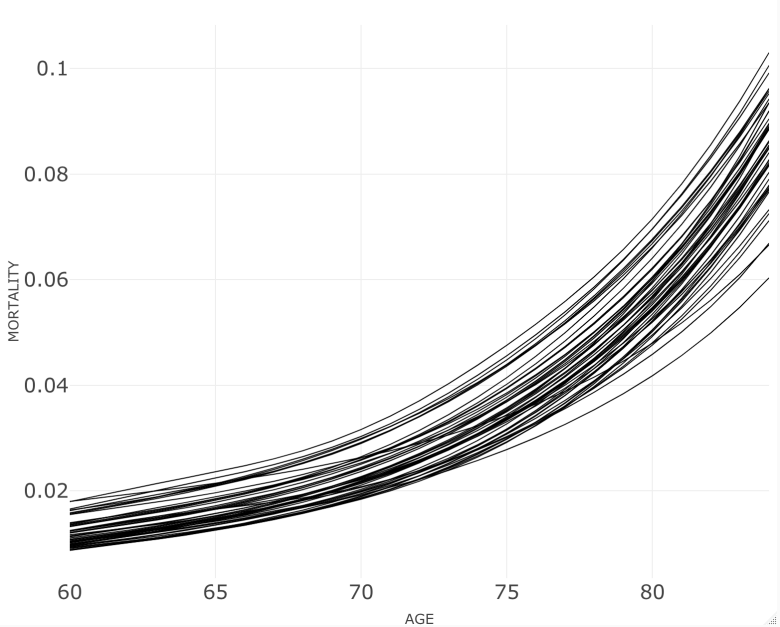
Table 1: Improved Mortality Predictions in 2020 (vs 2010)

(2) Stable behavior among states (no convergence).

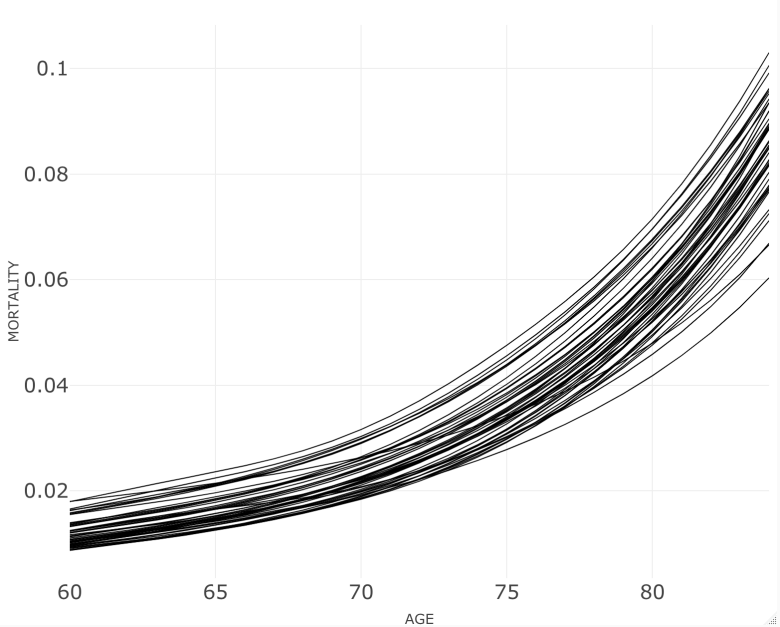
(3) States which had the best and worst mortality predictions pre-COVID (2020) were: MS, AL, OK and MA, NH, VT for both males and females.

Mortality Age Structure (Ages: 60-84)

Next, we analyze the *extrapolated mortality rates* in 2020 across the Age dimension.



Males Year 2020

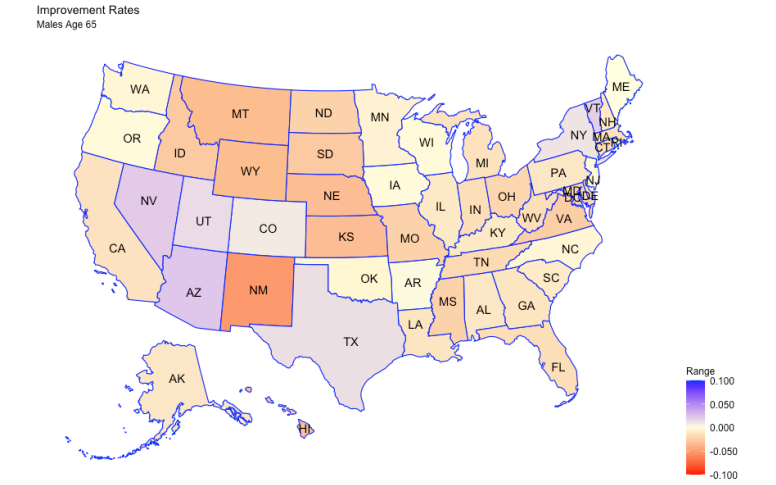


Females Year 2020

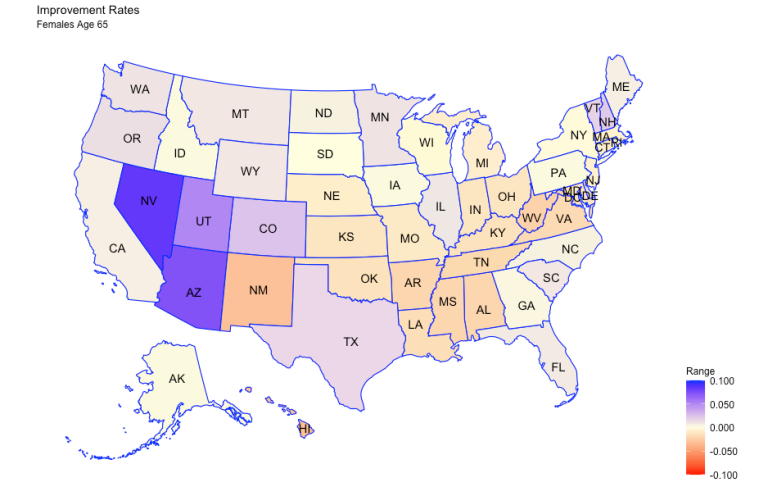
Figure 5: Multi-Output GP Mortality Rates for Ages 60-84.

• Mortality increases exponentially in Age with very few 'cross-overs'. Implies a high correlation between relative mortality experience at different ages.

State-Level MI Maps: Non-parametric ‘nowcasts’



Males Age 65



Females Age 65

Figure 6: MOGP-based annualized improvement factors in 2020; Age 65

Observations

a) **Heterogeneity.** MI for Females is higher and less variable.

b) **Geography Patterns.** Supports premise that mortality experience is similar for ‘alike’ states.

c) **Outlier states.** (NV, AZ, NM) and (AL, MS, TN), which have strong a upward or downward MI across both genders.

(i) Improvement Factors Across Ages 60-84

Age structure of improvement rates for the Year 2020 (heat maps).

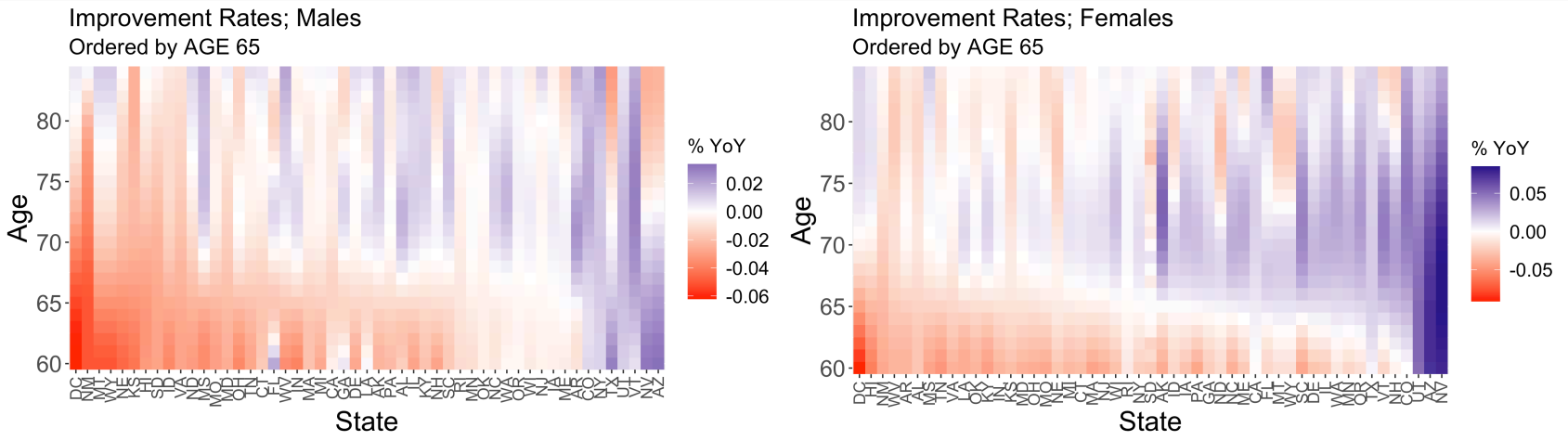


Figure 7: MOGP Improvement Rates. Ordered by Age 65 in 2020.

Color Scheme:

Negative MI = red, positive MI = purple, and neutral MI = white.

(ii) Improvement Factors Across Ages 60-84

a) **Most common pattern:** negative improvement for 'younger' population and a positive improvement for 'elderly' population.

b) **Outlier States:** negative improvement for 'elderly' population and a positive improvement for 'younger' population.
   - Arizona, Nevada, and Texas (for Males).

c) States with a High MI across **all** ages.
   - Utah, Arizona, Nevada (for Females).
   - Then, Vermont, Utah, Colorado, New York (for Males).

d) States with a Low MI across **all** ages.
   - Kansas, New Mexico, and Nebraska.

4. Conclusion

Overview

- By employing a two-step statistical procedure, we smoothed the raw data from the USMDB in order to, (a) extrapolate future mortality predictions and (b) analyze aggregate mortality trends in both Age- and Time- dimensions.

- This procedure consisted of:
  a) Combining states into *'optimal' groupings*; which advocated for data pooling and information fusion.
  b) Using a machine learning framework to analyze these **jointly correlated datasets**; in order to maximize the smoothing process and predictive power.

- The analysis highlighted some take-aways that appear new in terms of the aggregate behavior of the 50 states, some notable "outliers" and the outlook just prior to the Covid-19 pandemic.

Selected References i

N. Huynh and M. Ludkovski.
Multi-Output Gaussian Processes for Multi-Population Longevity Modeling.
Annals of Actuarial Science, 15(2):318–345, 2021.

N. Huynh, M. Ludkovski, and H. Zail.
Multi-population Longevity Models: A Spatial Random Field Approach.
Living to 100, 2020.

M. Ludkovski, J. Risk, and H. Zail.
Gaussian Process Models for Mortality Rates and Improvement Factors.
ASTIN Bulletin, 48(3):1307–1347, 2018.