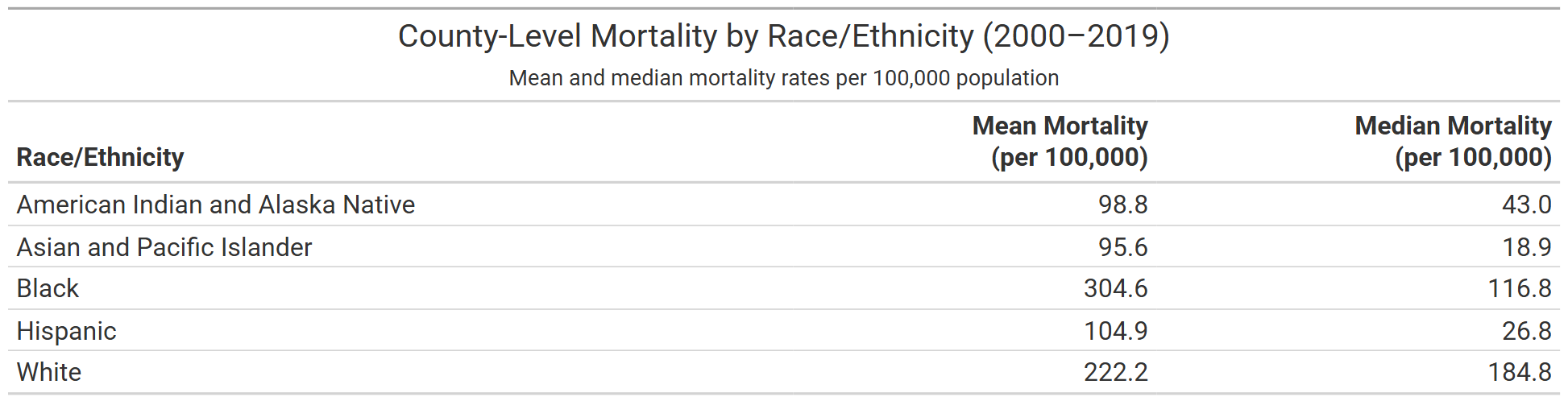
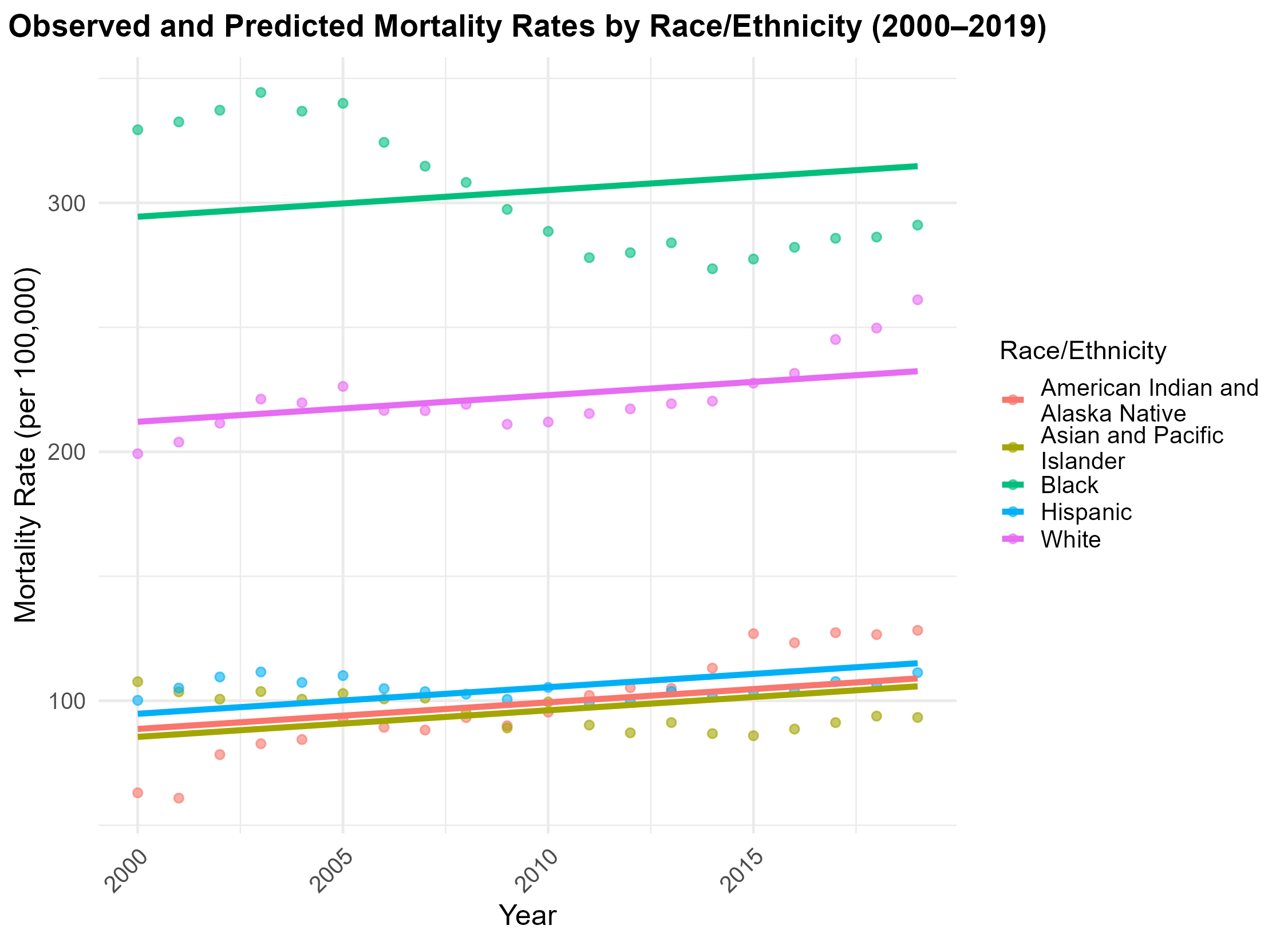
Supplement to Investigating the incidence and potential risk factors for hypertension-related mortalities in adults in the United States.

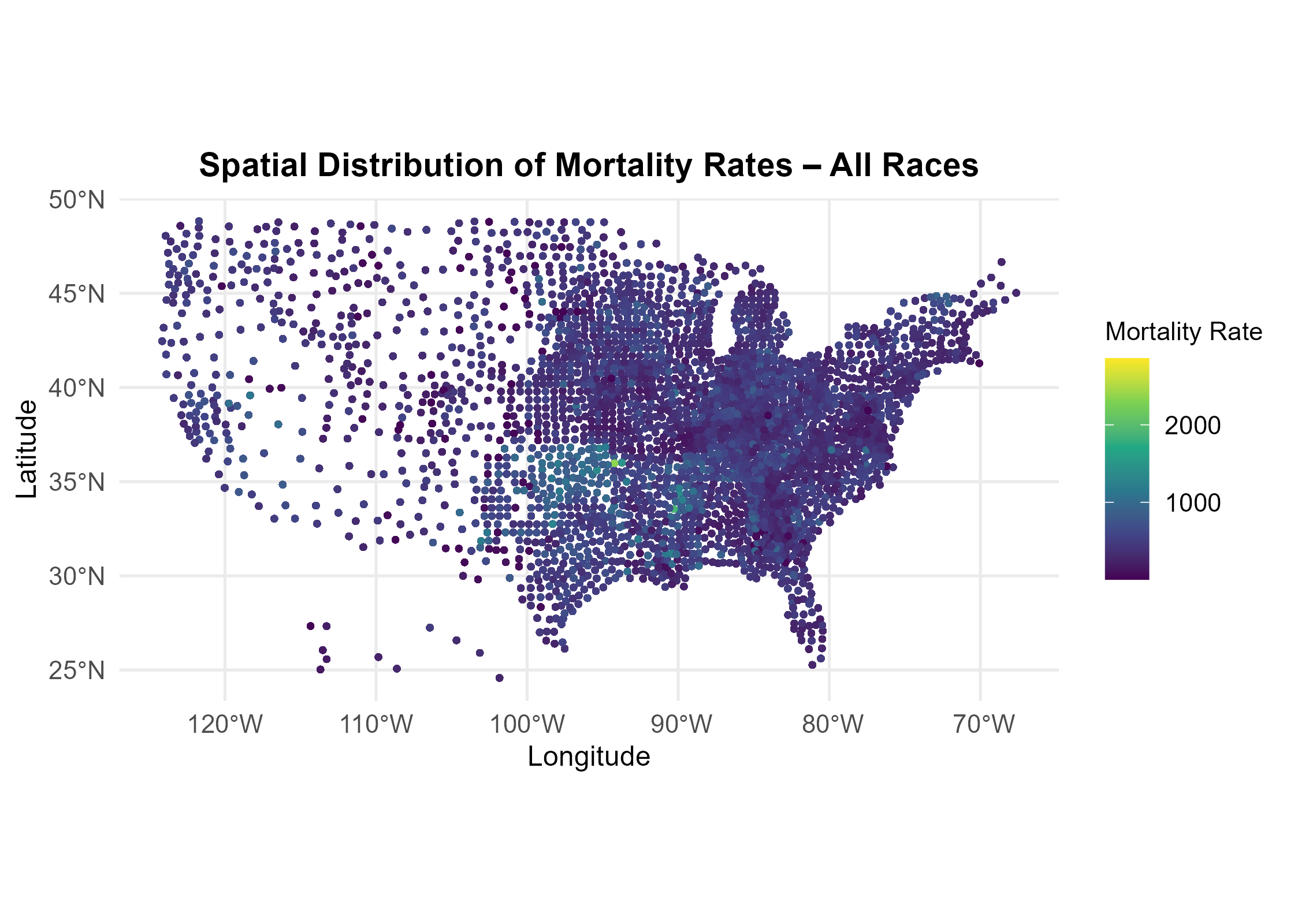
## Supplementary figures

**Supplementary Table S1.** Mean and median county-level mortality rates per 100,000 by race/ethnicity (2000–2019).

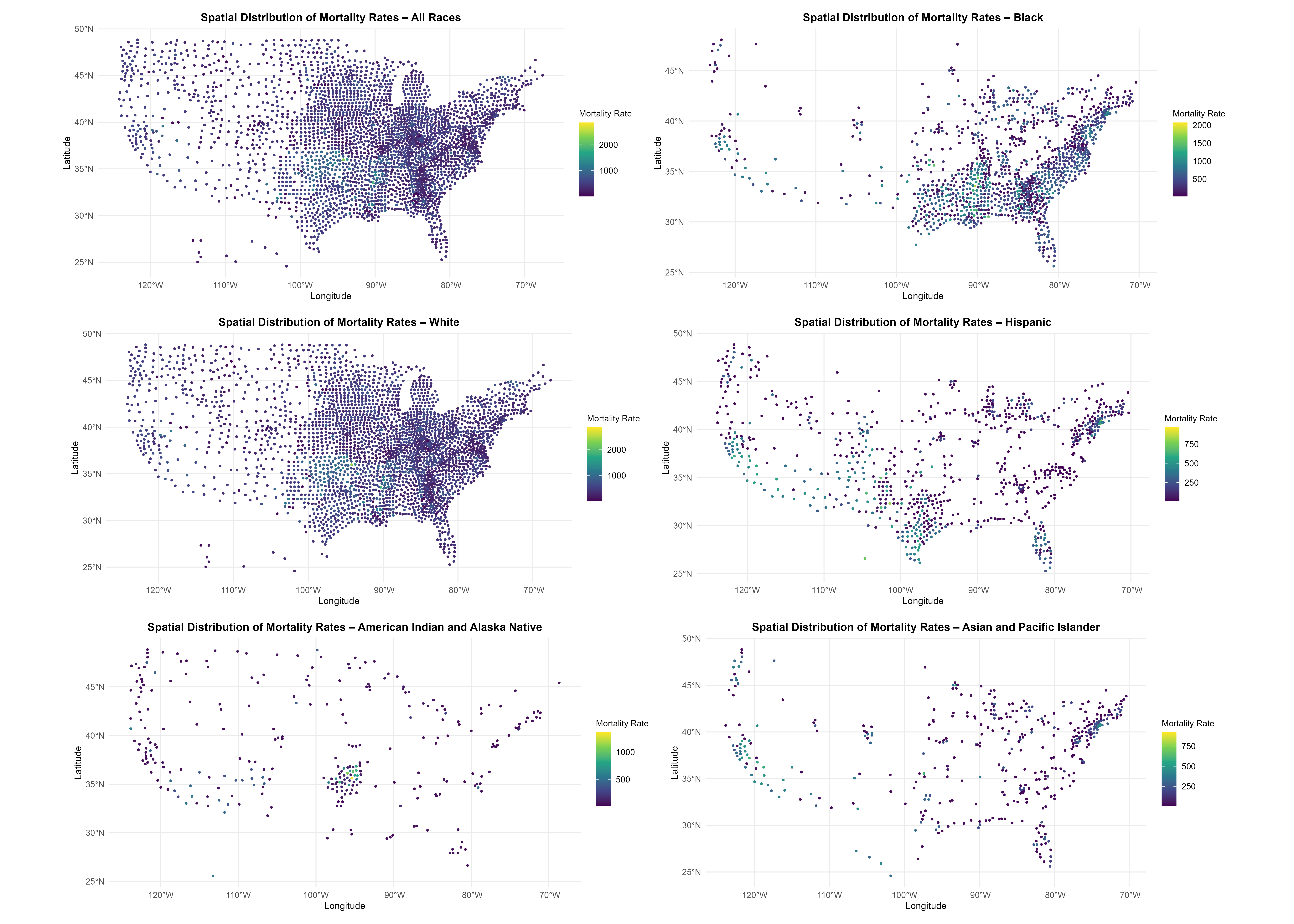
```

**Supplementary Figure S1.** Observed and predicted mortality rates by race/ethnicity (2000–2019), based on a generalized linear model without interaction terms. Assumes a shared temporal trend across all groups.

```

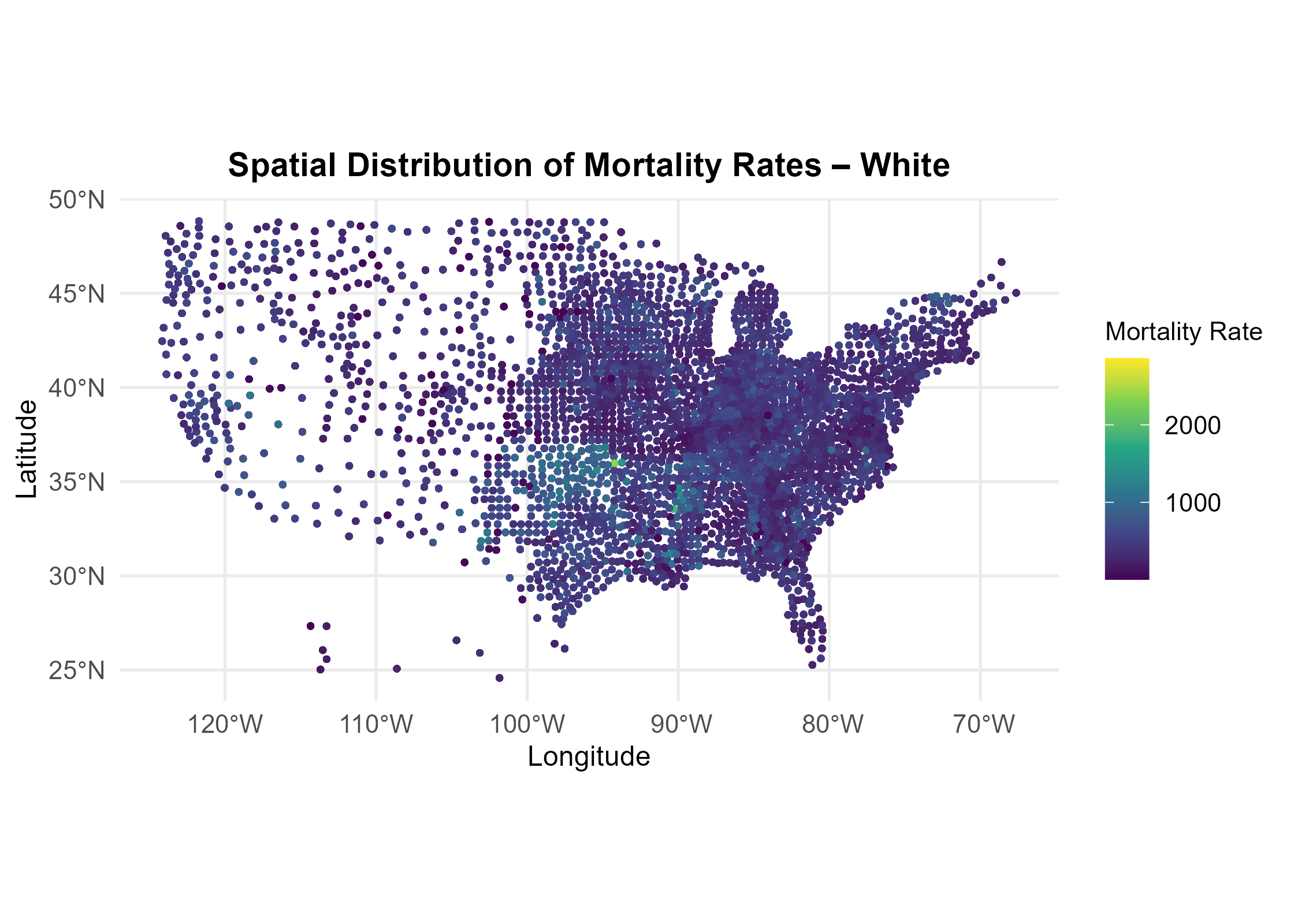
**Supplementary Figure S2.** Spatial distribution of hypertension-related mortality – All Races.

```

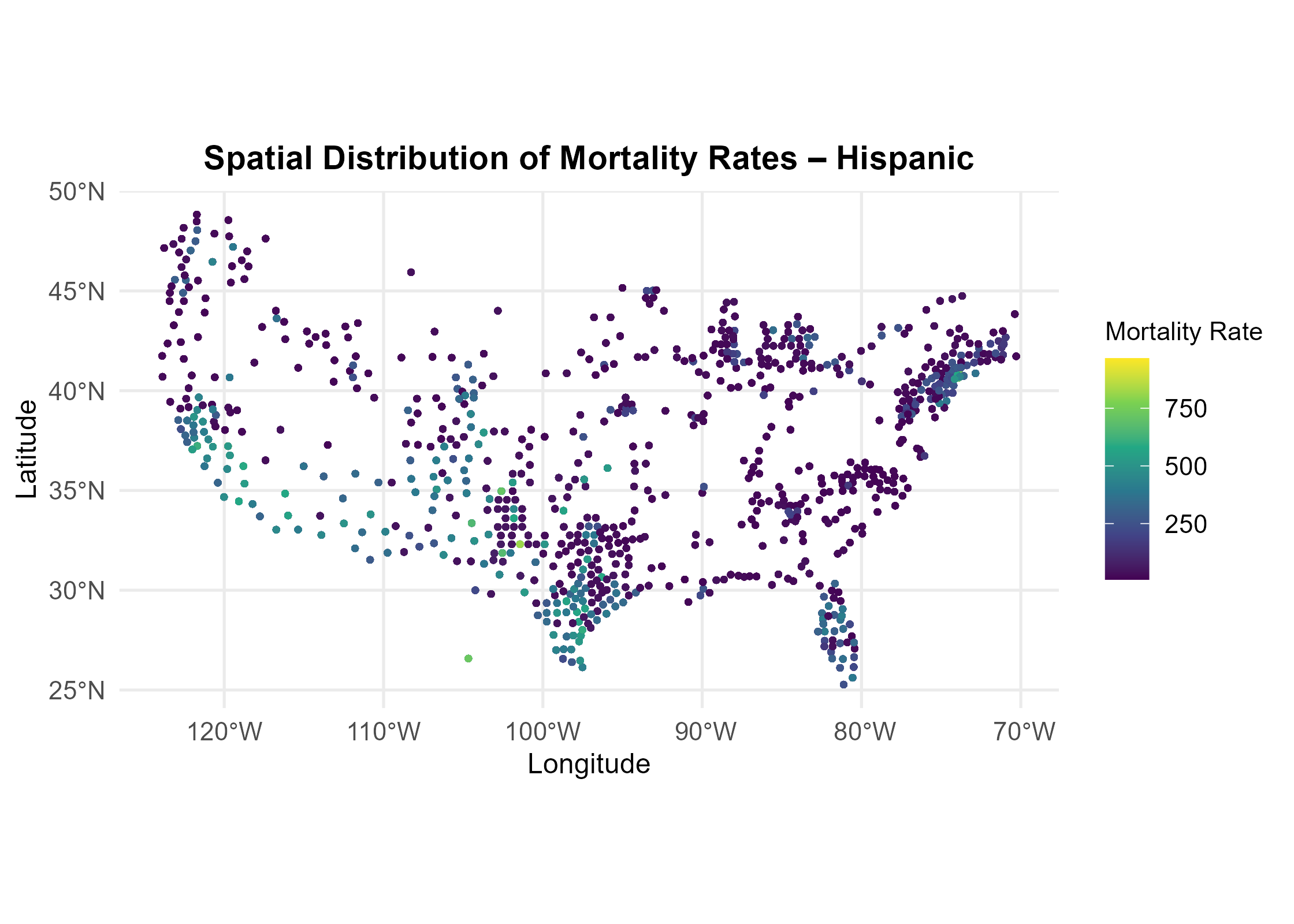
**Supplementary Figure S3.** Spatial distribution of hypertension-related mortality – Combined plot of all race/ethnicity groups.

The combined plot (Supplement Figure 3) provides an overview of mortality patterns across groups. While harder to read at the small scale once rendered in Word, it was still included as it offers helpful visual context when reviewed alongside the full-size individual maps.

```

**Supplementary Figure S4.** Spatial distribution of hypertension-related mortality – White population.

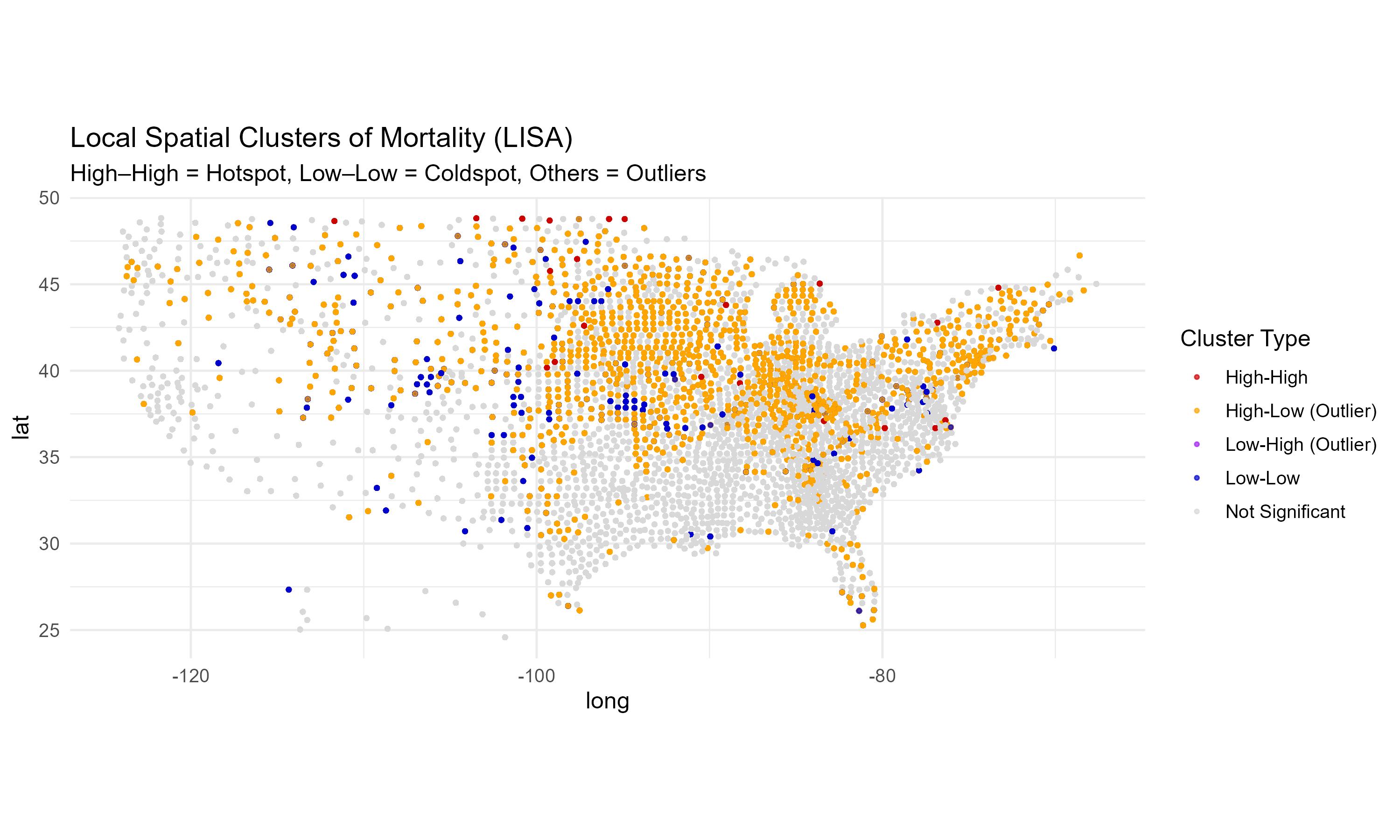
```

**Supplementary Figure S5.** Spatial distribution of hypertension-related mortality – Hispanic population.

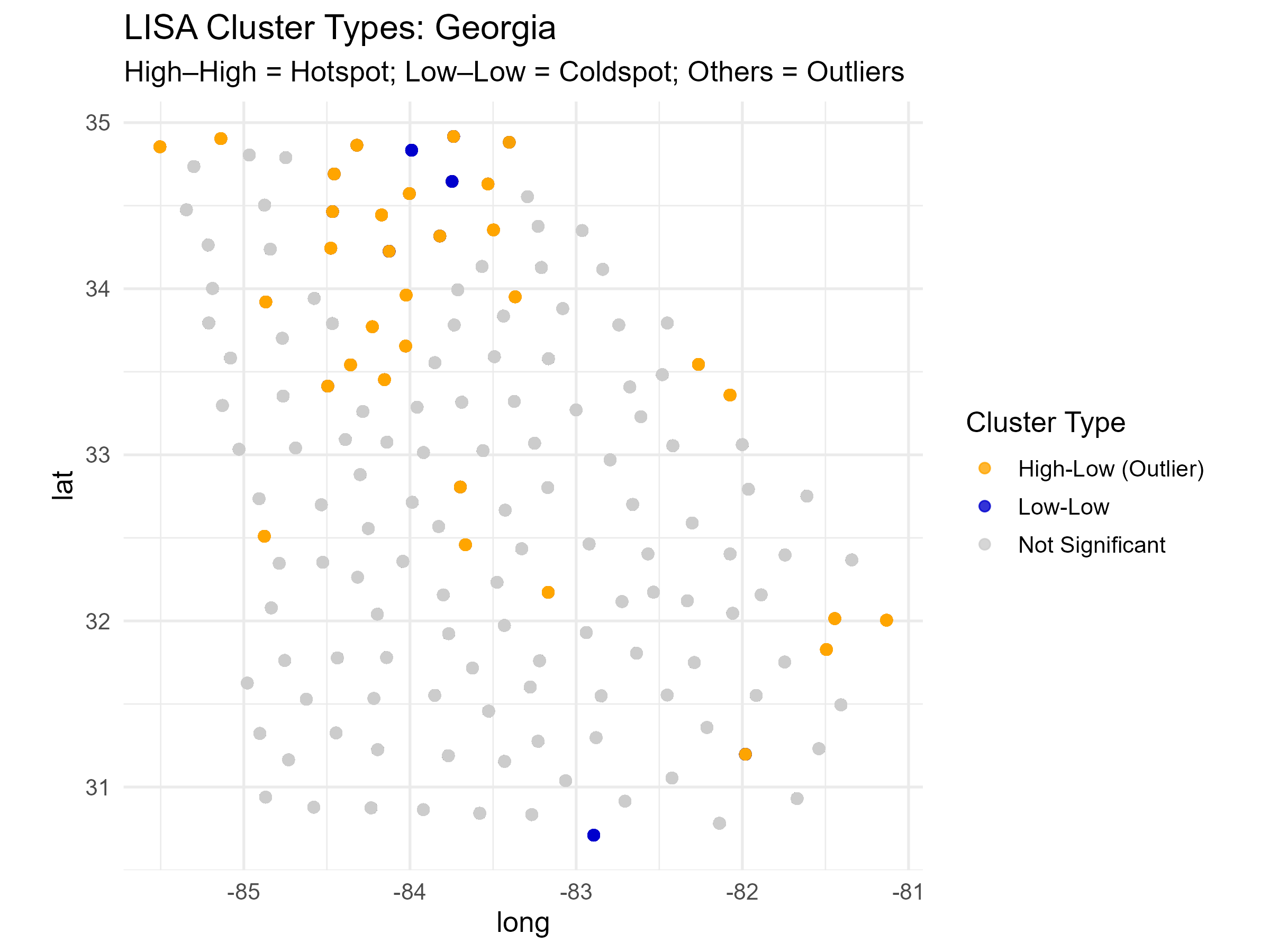
```

**Supplementary Figure S6.** Spatial distribution of hypertension-related mortality – American Indian and Alaska Native population.

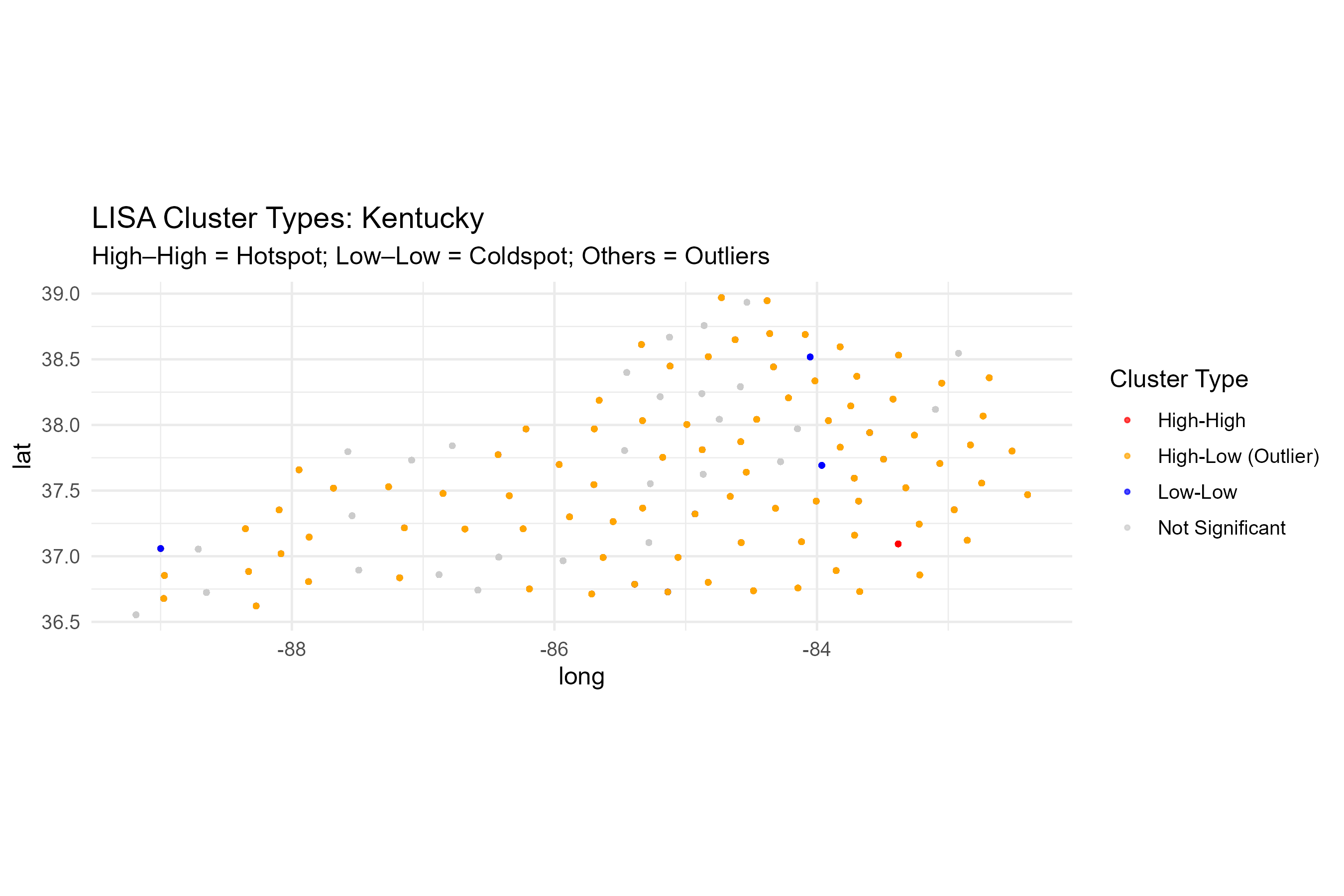
```

**Supplementary Figure S7.** Local Indicators of Spatial Association (LISA) cluster map for the contiguous United States. The map identifies county-level clusters of high or low mortality rates relative to their neighbors. Statistically significant clusters include high–high (mortality hotspots), low–low (coldspots), and spatial outliers (high–low or low–high), with color coding based on the spatial relationship and statistical significance.

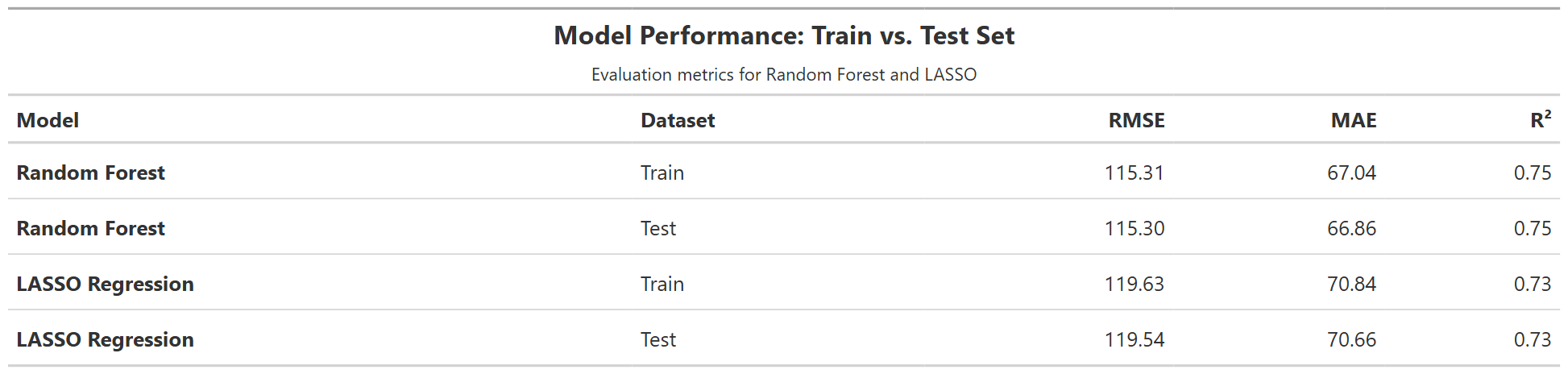
```

**Supplementary Figure S8.** Local Indicators of Spatial Association (LISA) cluster map for Georgia. This map visualizes statistically significant local spatial autocorrelation in hypertension-related mortality rates across counties in Georgia. High–high clusters indicate counties with high mortality rates surrounded by similarly high neighbors, while low–low clusters represent the opposite. Other categories denote spatial outliers and areas with no significant local clustering.

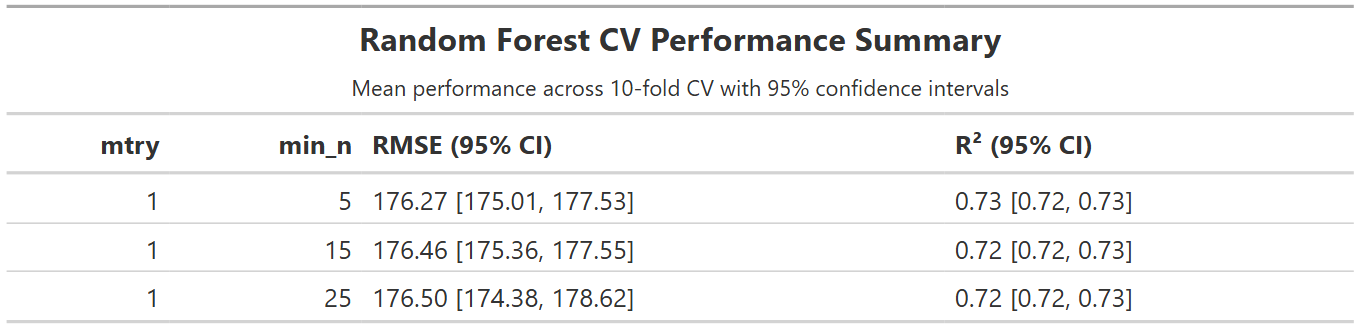
```

**Supplementary Figure S9.** Local Indicators of Spatial Association (LISA) cluster map for Kentucky. This figure shows localized spatial clusters of hypertension-related mortality, highlighting counties with significantly elevated or reduced mortality compared to their immediate neighbors. High–high clusters represent potential hotspots for targeted public health intervention.

```

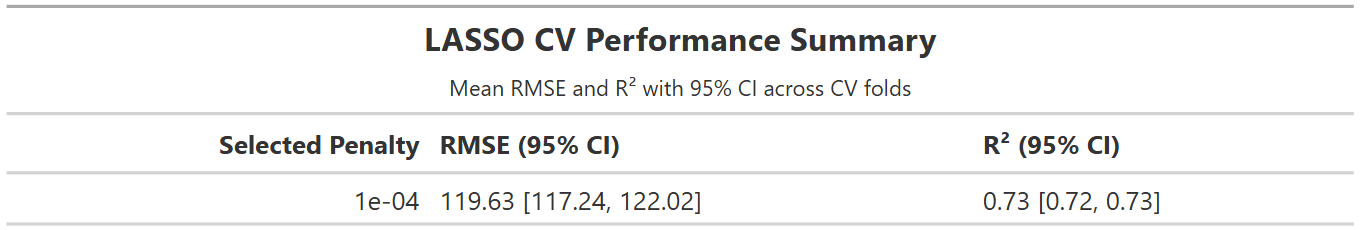
**Supplementary Table S2.** Train and test set performance metrics for Random Forest and LASSO models. Metrics include RMSE, MAE, and R². Similar values across splits indicate strong generalization and minimal overfitting.

```

**Supplementary Table S3.** Ten-fold cross-validation performance metrics for the Random Forest model. Values represent mean RMSE and R² across folds with 95% confidence intervals calculated using the standard error for each hyperparameter combination.

Cross-validation performance metrics for Random Forest model. Mean RMSE and R² values are shown across 10-fold cross-validation for selected hyperparameter combinations (mtry and min\_n). Values reflect consistent predictive performance across folds, indicating high model stability.

```

**Supplementary Table S4.** Ten-fold cross-validation performance metrics for the LASSO regression model. Values represent mean RMSE and R² with 95% confidence intervals based on variability across folds using the selected penalty parameter.

Cross-validation performance metrics for LASSO regression model. Mean RMSE and R² values with 95% confidence intervals are reported across 10-fold cross-validation folds using the optimal penalty value. The narrow confidence intervals reflect low variability and high consistency in model performance.

```

## Reproducibility and Software

### Steps for reproducing the project

To fully reproduce all steps of the project, the code should be run in the following order:

1. processing-code/processing-code.qmd^^ (Not required to reproduce results - This file will not run - see below)
2. processing-code/processingfile-v1.qmd
3. eda-code/eda.qmd
4. analysis-code/statistical-analysis-v2-slim.qmd\*
5. products/manuscript/Manuscript.qmd
6. products/manuscript/supplement/Supplement-CVD.qmd

\*Note: This is a large script and will take some time and processing power to run. The script uses loops to load saved data to reduce computation but expect it to take about 15 minutes to run the full script even with those steps.

Note: Because all data, figures, and tables are saved in the relevant subfolders, the manuscript can be produced by simply running the Manuscript.qmd file.

^^Note: processing-code/processing-code.qmd - This file will not run as the Excel CSV file is not in the repository due to size limitations. It was included purely to demonstrate processing steps. You can contact the author to get it or download the file directly from the CDC website if you wish to run this yourself.

### Software Packages and Environment

All analyses presented in this project were performed using R within a version-controlled environment managed by the renv package. This ensures that all package versions and dependencies used to generate figures and tables are locked and reproducible.

To reproduce this analysis:

1. Clone the repository.
2. Open the project in R or RStudio.
3. Restore the environment by running:

#renv::restore()

The analysis was conducted using R version 4.4.0 (2024-04-24 ucrt) on a Windows 11 x64 system via RStudio 2024.12.0+467. The Quarto engine, R packages, and their versions used in this analysis are listed below to support full reproducibility.

**R version:**  
R version 4.4.0 (2024-04-24 ucrt)

**Operating system:**  
Windows 11 x64 (build 26100)

**RStudio version:**  
2024.12.0+467 Kousa Dogwood

**Key R Packages Used (Attached or namespace-loaded)**

* **tidymodels** 1.3.0
* **tidyverse** 2.0.0
* **gt** 0.11.1
* **yardstick** 1.3.2
* **rsample** 1.3.0
* **parsnip** 1.3.0
* **tune** 1.3.0
* **recipes** 1.1.1
* **workflows** 1.2.0
* **workflowsets** 1.1.0
* **vip** 0.4.1
* **spdep** 1.3-10
* **spatialreg** 1.3-6
* **sf** 1.0-19
* **spData** 2.3.4
* **skimr** 2.1.5
* **dplyr** 1.1.4
* **ggplot2** 3.5.1
* **forcats** 1.0.0
* **purrr** 1.0.4
* **readr** 2.1.5
* **stringr** 1.5.1
* **tibble** 3.2.1
* **lubridate** 1.9.3
* **gtsummary** (if used)
* **webshot2** 0.1.1
* **here** 1.0.1
* **magick** 2.8.5
* **sessioninfo** 1.2.3
* Other dependencies (e.g., cli, glue, broom, etc.) were loaded automatically via dependencies of the core packages listed above.

**Note:** A full session output is available upon request. All analyses were conducted using a reproducible pipeline, and Quarto was used for document generation and reporting.