

Geographic Variation in the Quality of US Mortality Data, 1999–2022



Monica Alexander, PhD
University of Toronto
monica.alexander@utoronto.ca



Amy Mann, BS
Oxford University
amy.mann@chch.ox.ac.uk



Mathew Kiang, ScD
Stanford University
mkiang@stanford.edu

Provisional Mortality Statistics, X

wonder.cdc.gov/controller/datarequest/D176

120%



Search

CDC WONDER

FAQs

Help

Contact Us

WONDER Search

CDC's website is being modified to comply with President Trump's Executive Orders.

Provisional Mortality Statistics, 2018 through Last Week Request

Deaths occurring through April 11, 2026 as of April 19, 2026

Request Form

Results

Map

Chart

About

[Multiple Cause of Death Data](#)

[Dataset Documentation](#)

[Other Data Access](#)

[Data Use Restrictions](#)

[How to Use WONDER](#)

Save

Reset

Make all desired selections and then click any **Send** button one time to send your request.

1. Organize table layout:

Send

Help

Group Results By

Residence Census Region

And By

None

And By

None

And By

None

And By

None

Notes:

- Group Results By "15 Leading Causes" to see the top 15 rankable causes selected from the corresponding 113 or 130 Cause List.

[More information.](#)

71 y/o male presents with chest pain

WHO PRONOUNCES OR CERTIFIES DEATH		
26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable)		27. LICENSE NUMBER
		28. DATE SIGNED (Mo/Day/Yr)
29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month)	30. ACTUAL OR PRESUMED TIME OF DEATH	31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? ☐ Yes ☐ No
CAUSE OF DEATH (See instructions and examples) 32. PART I. Enter the <u>chain of events</u>--diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary. IMMEDIATE CAUSE (Final disease or condition -----> resulting in death) a. _____ Due to (or as a consequence of): Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST b. _____ Due to (or as a consequence of): c. _____ Due to (or as a consequence of): d. _____ Approximate interval: Onset to death _____ _____ _____ _____		
32. PART II. Enter other <u>significant conditions contributing to death</u> but not resulting in the underlying cause given in PART I		33. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No
		34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ☐ Yes ☐ No
35. DID TOBACCO USE CONTRIBUTE TO DEATH? ☐ Yes ☐ Probably	36. IF FEMALE: ☐ Not pregnant within past year ☐ Pregnant at time of death	37. MANNER OF DEATH ☐ Natural ☐ Homicide ☐ Accident ☐ Pending Investigation

Starting with the most immediate cause

WHO PRONOUNCES OR CERTIFIES DEATH		
26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable)		27. LICENSE NUMBER
		28. DATE SIGNED (Mo/Day/Yr)
29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month)	30. ACTUAL OR PRESUMED TIME OF DEATH	31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? ☐ Yes ☐ No
32. PART I. Enter the <u>chain of events</u> --diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary. IMMEDIATE CAUSE (Final disease or condition -----> resulting in death) a. <u>Acute MI</u> Due to (or as a consequence of): Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST b. _____ Due to (or as a consequence of): c. _____ Due to (or as a consequence of): d. _____		Approximate interval: Onset to death <u>4 h</u> _____ _____ _____ _____
33. PART II. Enter other <u>significant conditions contributing to death</u> but not resulting in the underlying cause given in PART I		33. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No
		34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ☐ Yes ☐ No
35. DID TOBACCO USE CONTRIBUTE TO DEATH? ☐ Yes ☐ Probably	36. IF FEMALE: ☐ Not pregnant within past year ☐ Pregnant at time of death	37. MANNER OF DEATH ☐ Natural ☐ Homicide ☐ Accident ☐ Pending Investigation

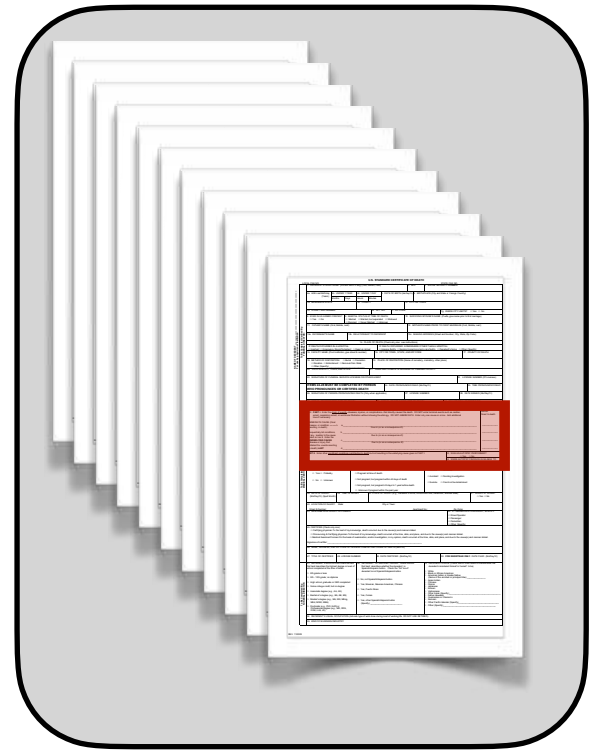
Working backwards to the underlying cause

WHO PRONOUNCES OR CERTIFIES DEATH			
26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable)		27. LICENSE NUMBER	28. DATE SIGNED (Mo/Day/Yr)
29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month)		30. ACTUAL OR PRESUMED TIME OF DEATH	31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? ☐ Yes ☐ No
CAUSE OF DEATH (See instructions and examples)			Approximate interval: Onset to death
32. PART I. Enter the <u>chain of events</u> --diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.			
IMMEDIATE CAUSE (Final disease or condition -----> resulting in death)	a. <u>Acute MI</u>	Due to (or as a consequence of):	<u>4 h</u>
Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST	b. <u>Coronary artery disease</u>	Due to (or as a consequence of):	<u>15 y</u>
	c.	Due to (or as a consequence of):	
	d.	Due to (or as a consequence of):	
33. PART II. Enter other <u>significant conditions contributing to death</u> but not resulting in the underlying cause given in PART I			33. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No
			34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ☐ Yes ☐ No
35. DID TOBACCO USE CONTRIBUTE TO DEATH? ☐ Yes ☐ Probably	36. IF FEMALE: ☐ Not pregnant within past year ☐ Pregnant at time of death	37. MANNER OF DEATH ☐ Natural ☐ Homicide ☐ Accident ☐ Pending Investigation	

Adding contributing factors

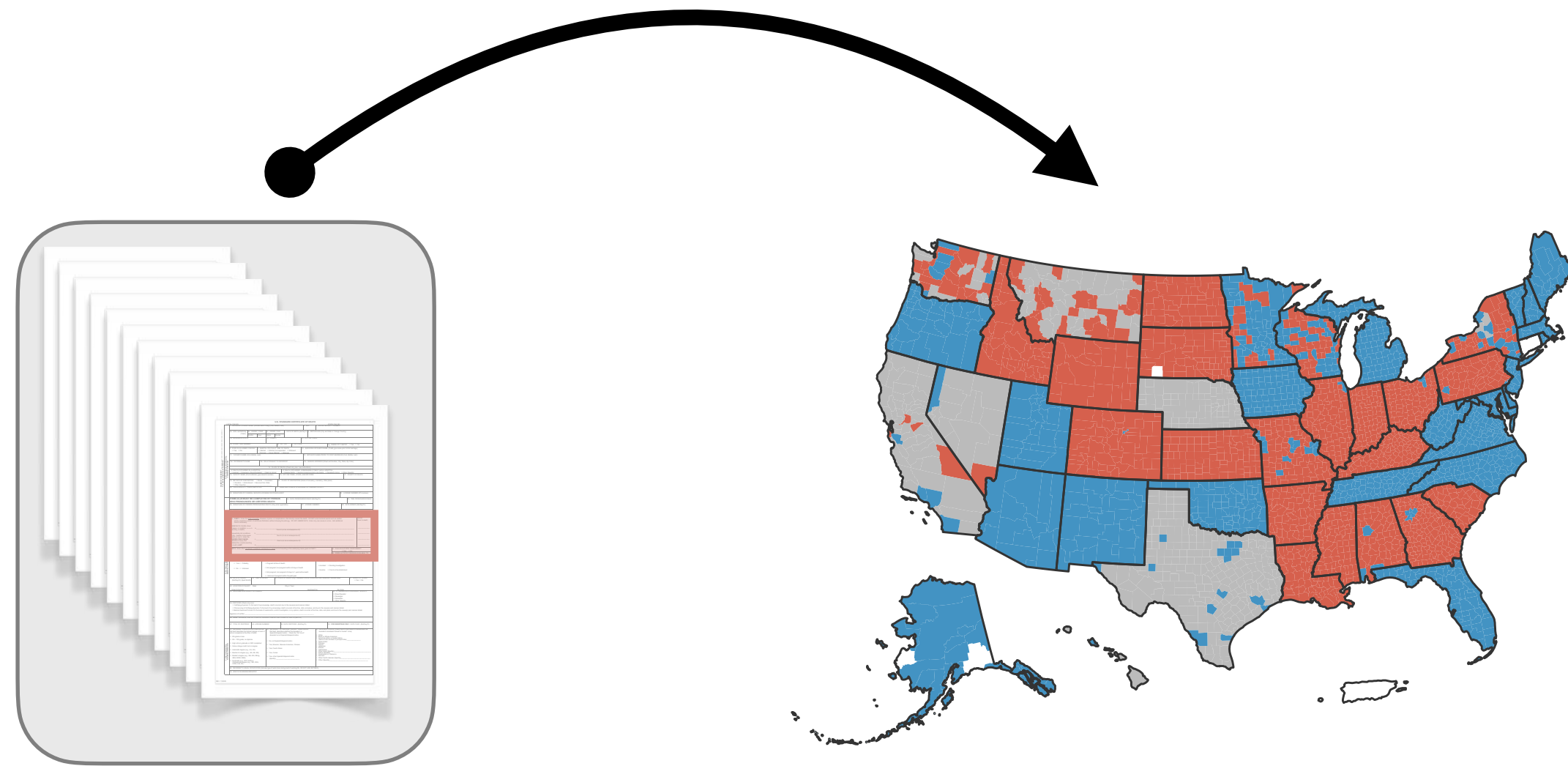
WHO PRONOUNCES OR CERTIFIES DEATH			
26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable)		27. LICENSE NUMBER	28. DATE SIGNED (Mo/Day/Yr)
29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month)		30. ACTUAL OR PRESUMED TIME OF DEATH	31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? ☐ Yes ☐ No
CAUSE OF DEATH (See instructions and examples)			Approximate interval: Onset to death
32. PART I. Enter the <u>chain of events</u> --diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.			
IMMEDIATE CAUSE (Final disease or condition -----> resulting in death)			
a. <u>Acute MI</u>			<u>4 h</u>
Due to (or as a consequence of):			
b. <u>Coronary artery disease</u>			<u>15 y</u>
Due to (or as a consequence of):			
c. _____			_____
Due to (or as a consequence of):			
d. _____			_____
PART II. Enter other <u>significant conditions contributing to death</u> but not resulting in the underlying cause given in PART I			33. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No
<u>Type-2 diabetes, hypertension, CKD stage 3</u>			34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ☐ Yes ☐ No
35. DID TOBACCO USE CONTRIBUTE TO DEATH? ☐ Yes ☐ Probably	36. IF FEMALE: ☐ Not pregnant within past year ☐ Pregnant at time of death	37. MANNER OF DEATH ☐ Natural ☐ Homicide ☐ Accident ☐ Pending Investigation	

Death certificates are sent to the state



Death certificates
are (electronically)
reported to the state.

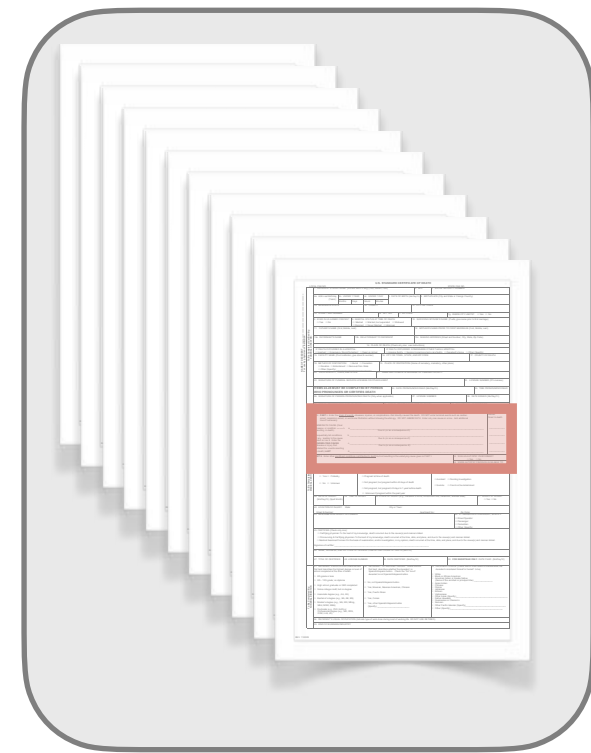
The state finalizes them and sends to NCHS



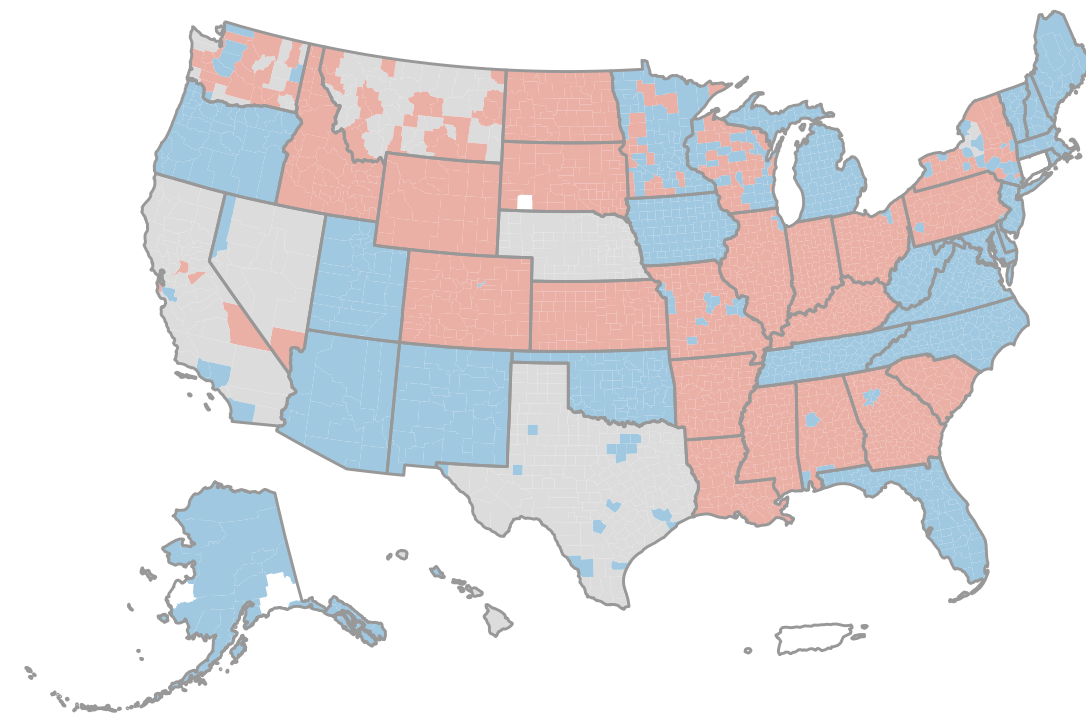
Death certificates
are (electronically)
reported to the state.

State finalizes death
certificates and transmits
to CDC NCHS.

NCHS codes the deaths



Death certificates are (electronically) **reported** to the state.

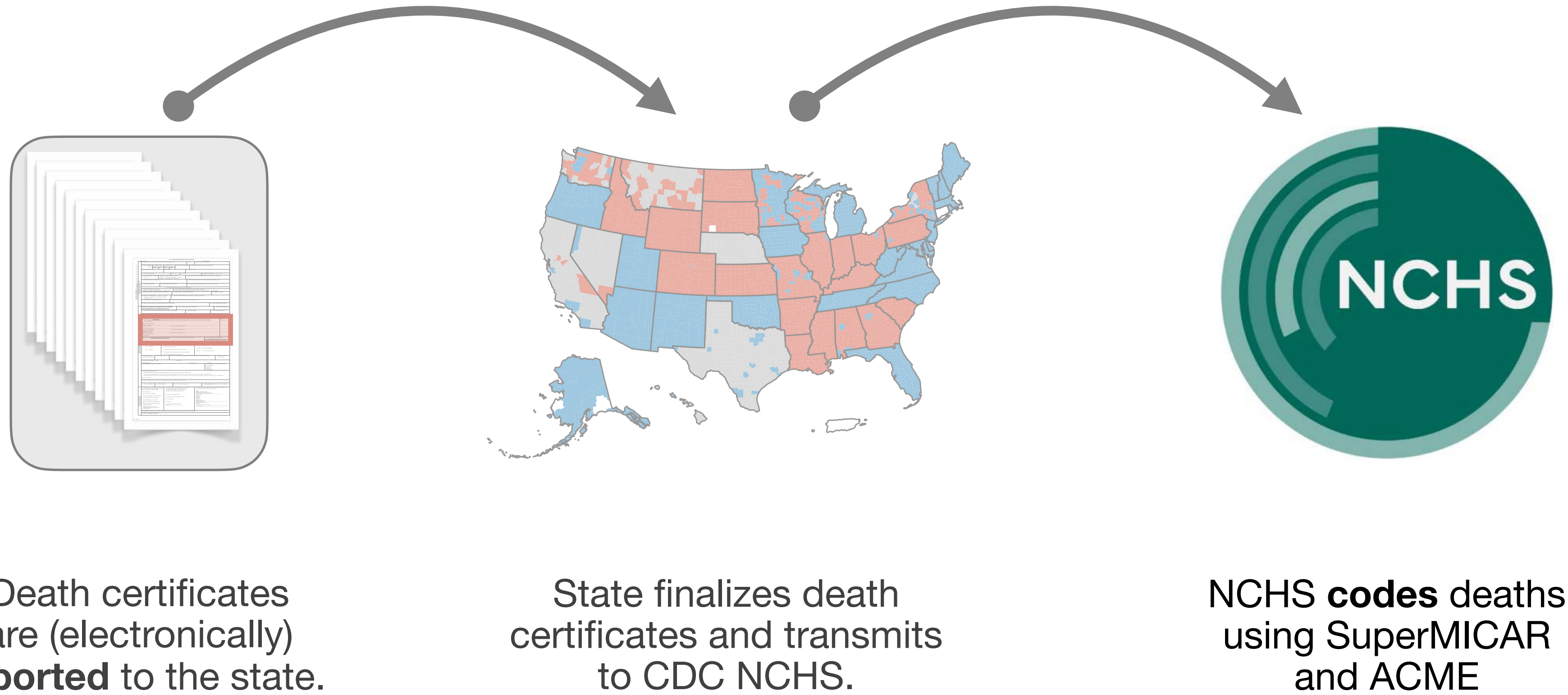


State finalizes death certificates and transmits to CDC NCHS.



NCHS **codes** deaths using SuperMICAR and ACME

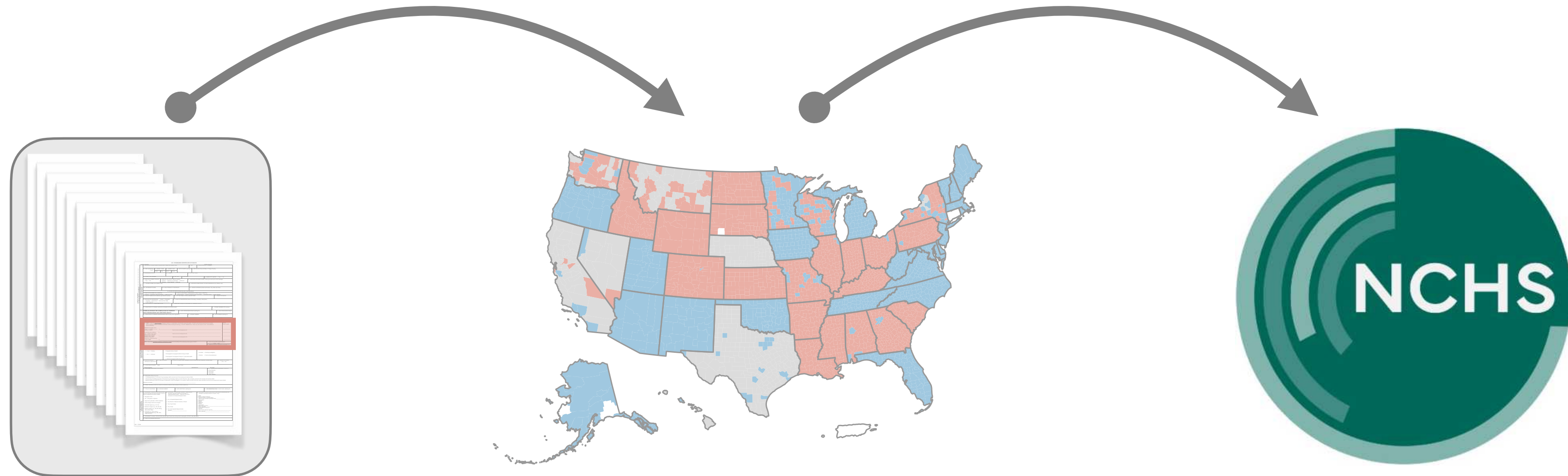
SuperMICAR converts free-text to ICD-10



SuperMICAR

- Acute MI → I21.9
- Coronary artery disease → I25.1
- Type 2 diabetes → E11.9
- Hypertension → I10
- CKD stage 3 → N18.3

ACME selects UCOD from ICD-10 codes



Death certificates
are (electronically)
reported to the state.

State finalizes death
certificates and transmits
to CDC NCHS.

NCHS **codes** deaths
using SuperMICAR
and ACME

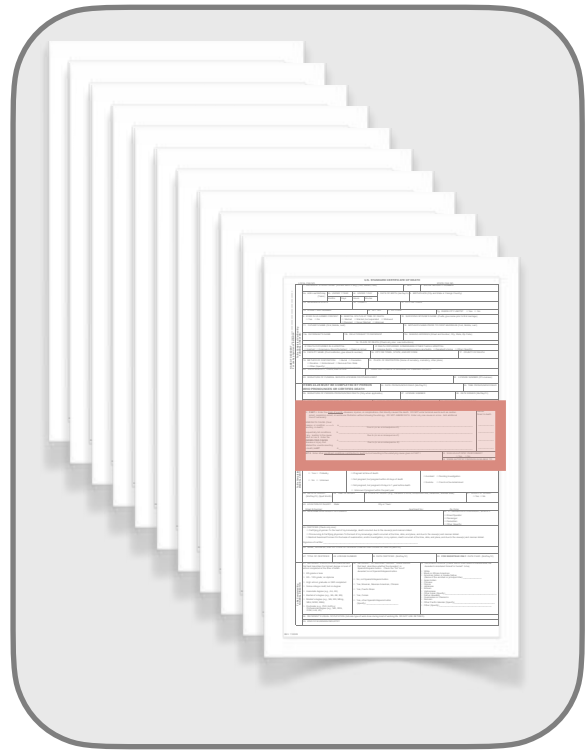
SuperMICAR

- Acute MI → I21.9
- Coronary artery disease → I25.1
- Type 2 diabetes → E11.9
- Hypertension → I10
- CKD stage 3 → N18.3

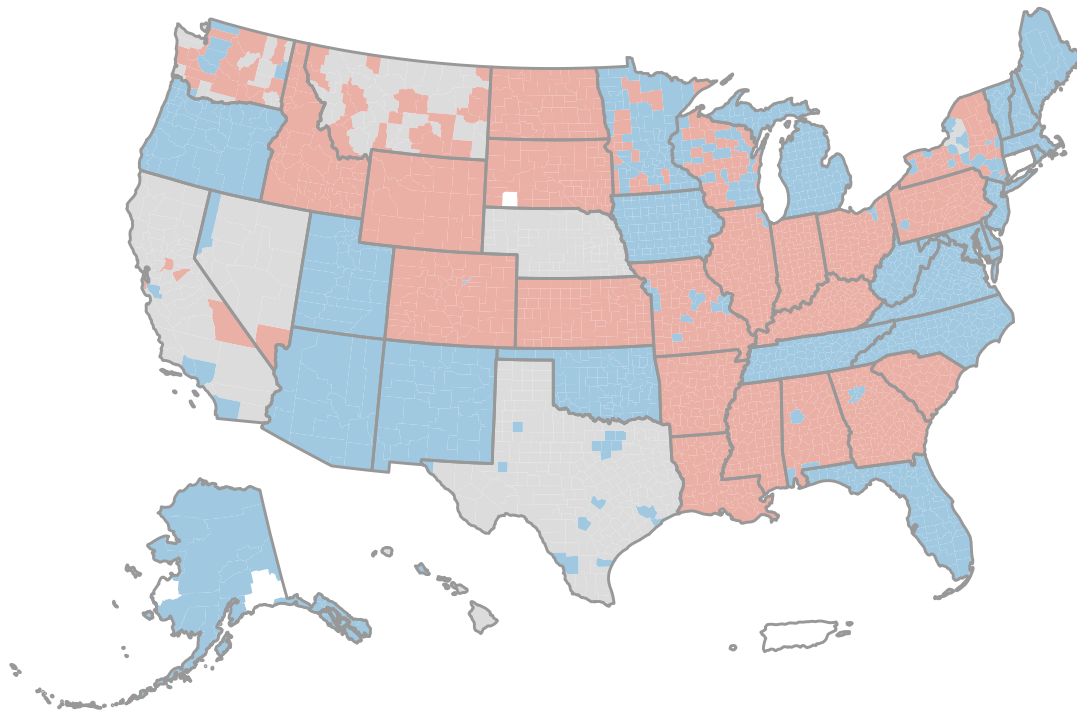
ACME

- **I25.1** → **UCOD**
- I21.9 → MCOD
- E11.9 → MCOD
- I10 → MCOD
- N18.3 → MCOD

CDC then releases the data publicly



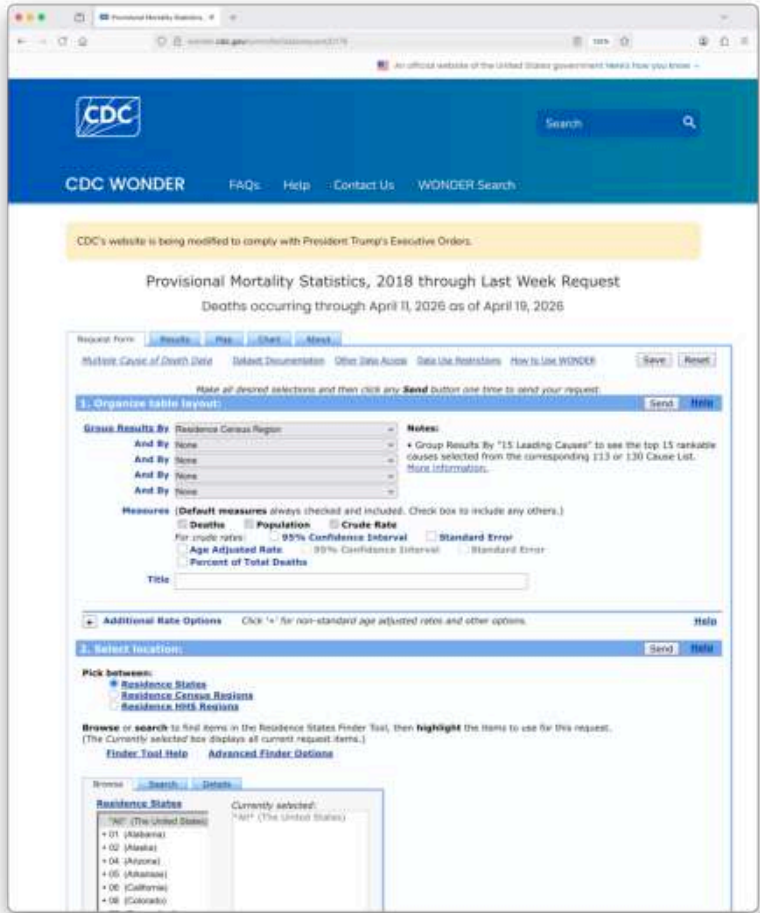
Death certificates are (electronically) filed with the state.



State finalizes death certificates and transmits to CDC NCHS.



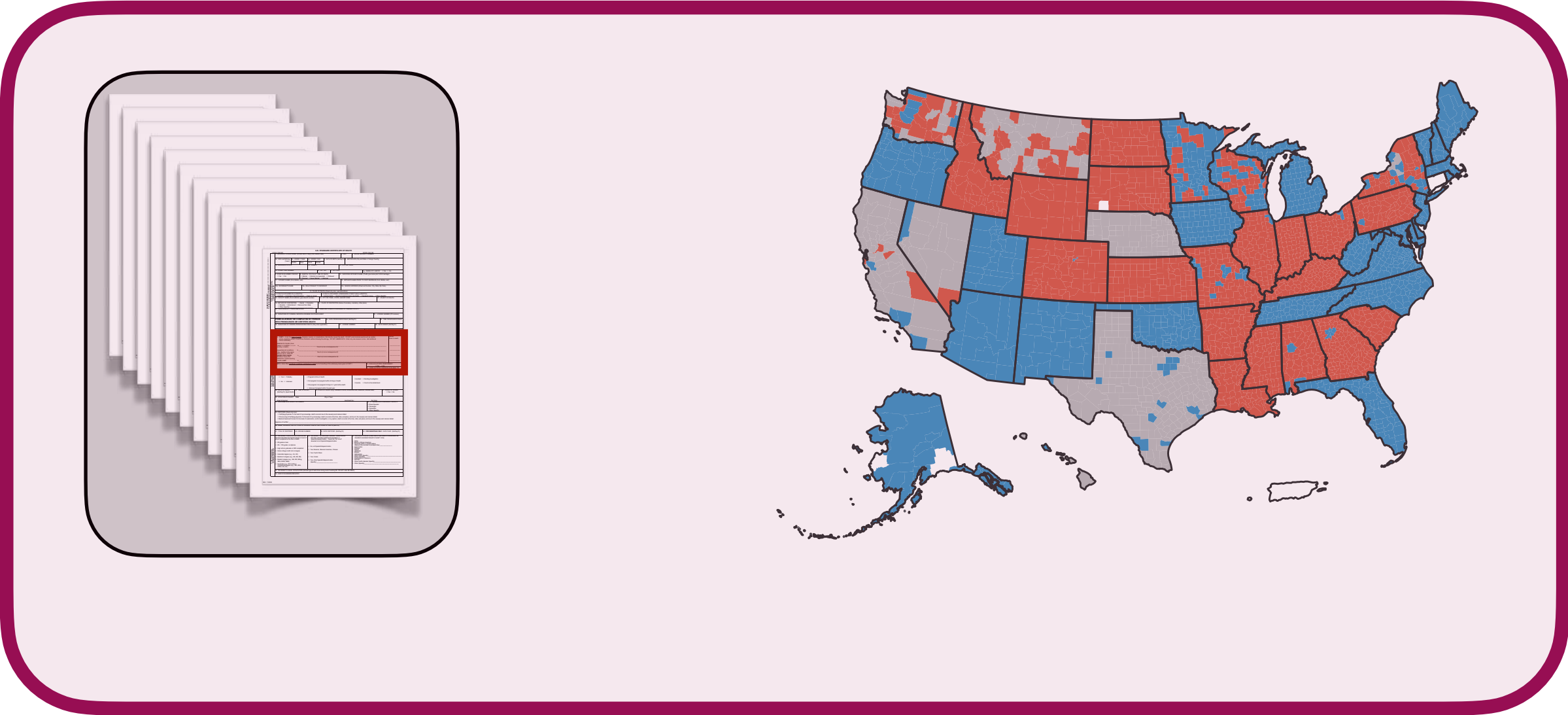
NCHS **codes** deaths using SuperMICAR and ACME



Data are publicly released (e.g., via CDC WONDER)

Phases of death data processing

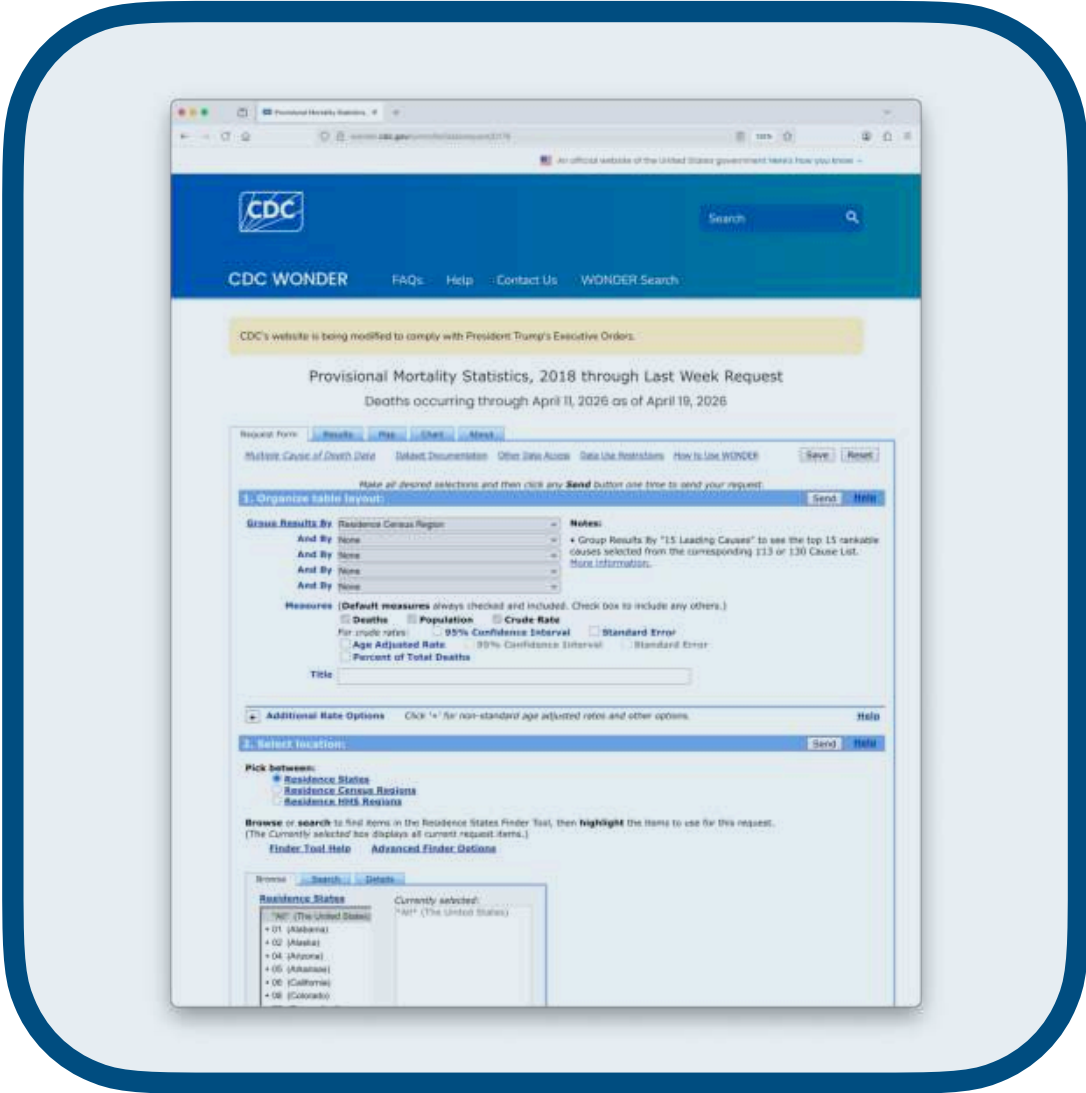
Death reporting



Death coding



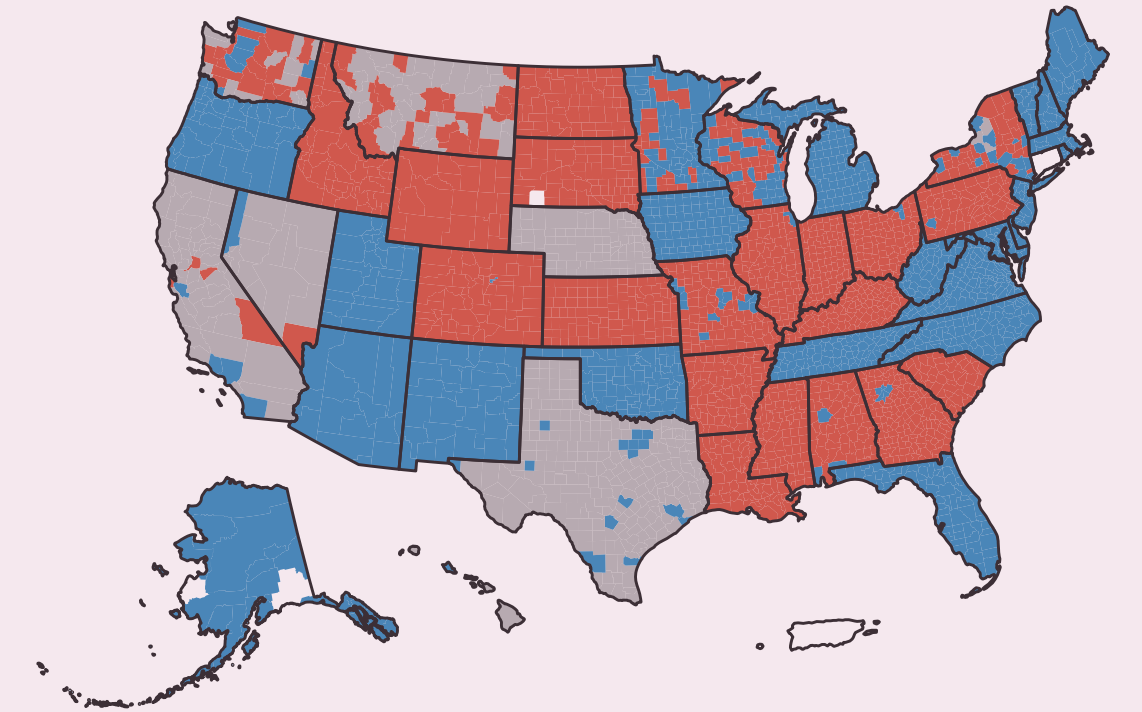
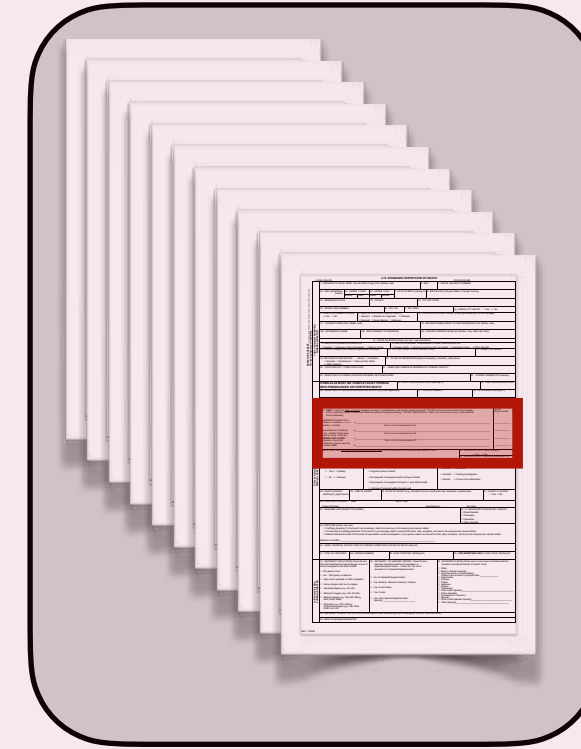
Data release



Death coding is deterministic

- **Death coding** is deterministic: Same free-text input leads to same ICD-10 code. Same set of ICD-10 codes leads to same underlying cause of death.

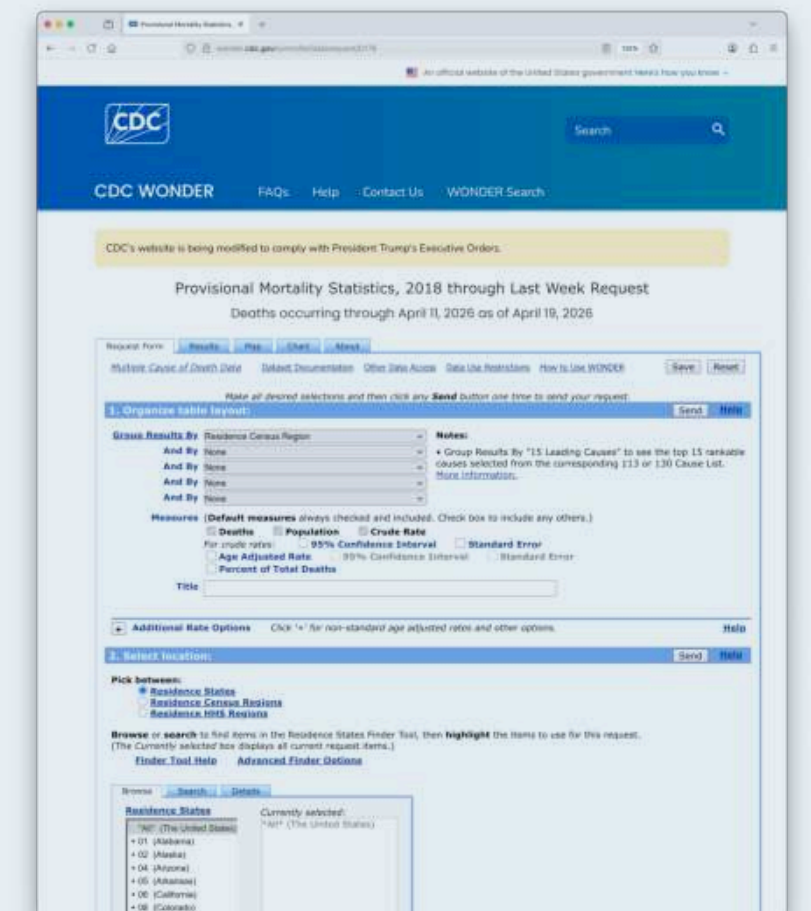
Death reporting



Death coding



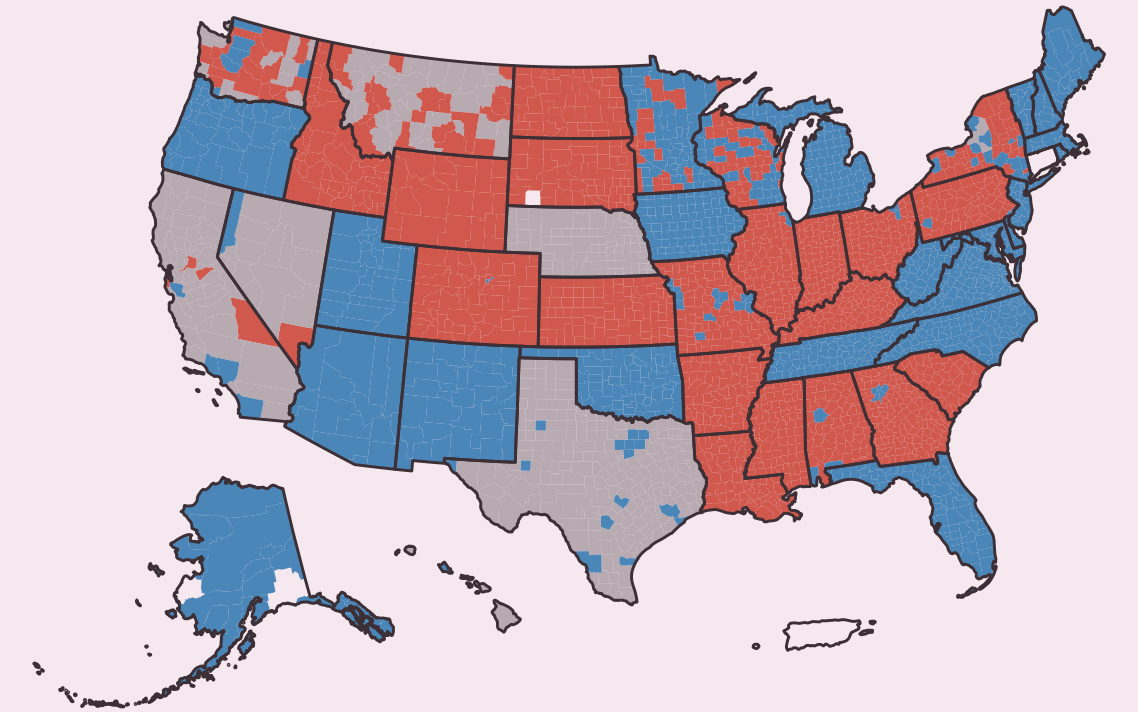
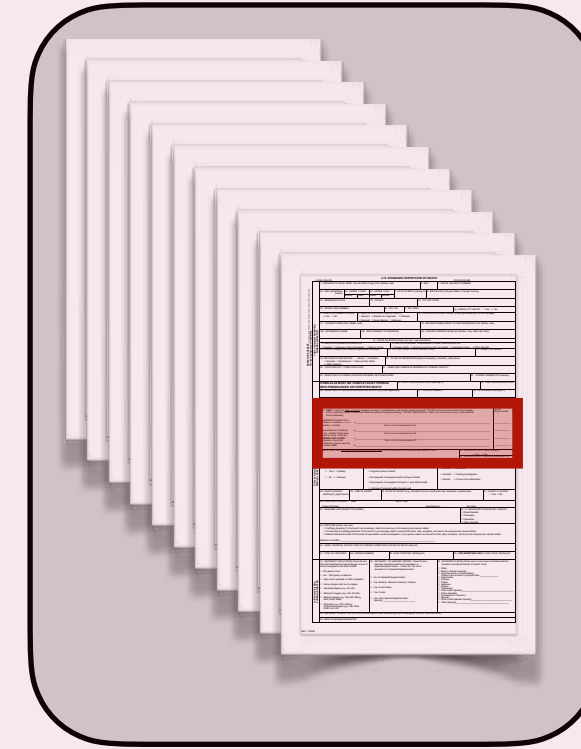
Data release



All variation comes from death *reporting*

- **Death coding** is deterministic: Same free-text input leads to same ICD-10 code. Same set of ICD-10 codes leads to same underlying cause of death.
- Therefore, all variation in death data quality comes from **death reporting**.

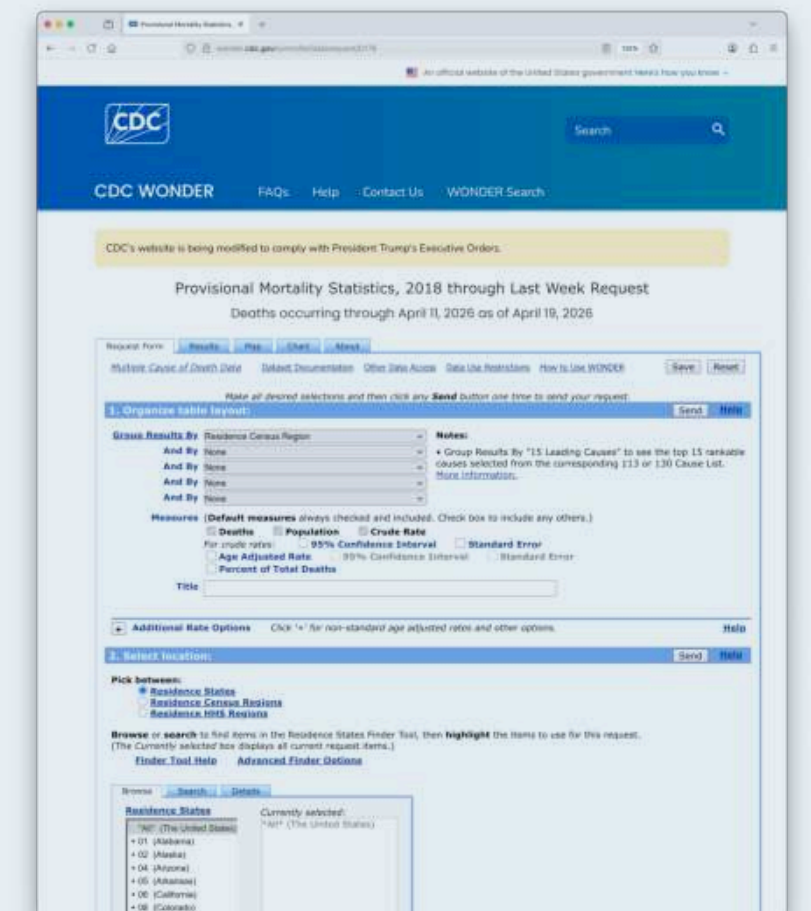
Death reporting



Death coding



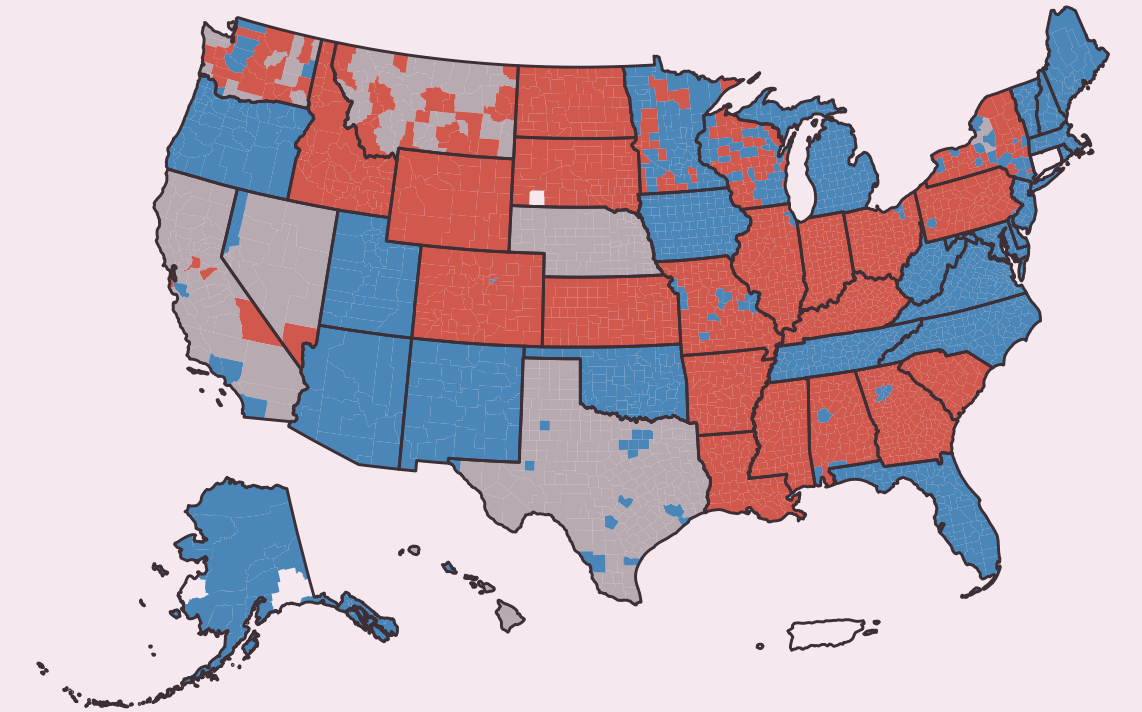
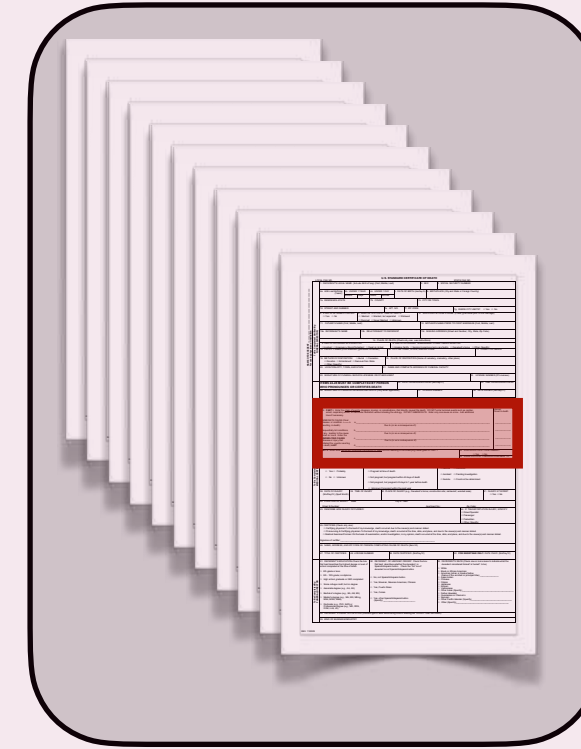
Data release



National death data quality is high

- Like most high-income countries, overall US death data are quite good based on traditional quality metrics.¹⁻³
- **Completeness** (how many deaths are recorded in the system)
- Proportion of "**garbage codes**" (how many registered deaths are specific enough to be useful)
- **Plausibility** (e.g., does the age and sex distribution make sense)

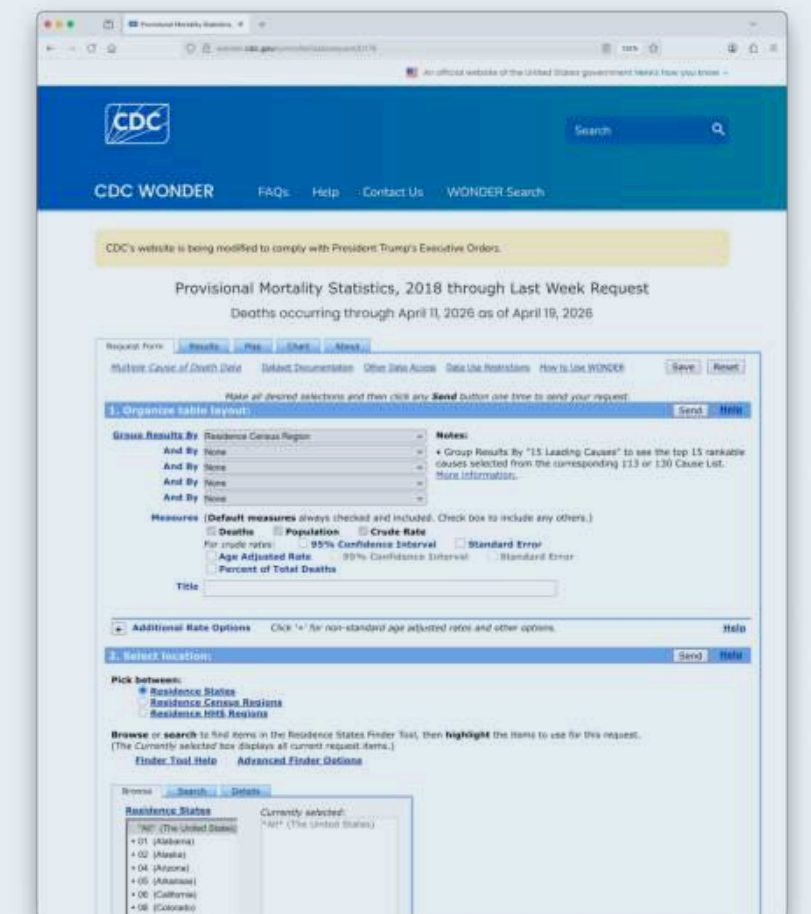
Death reporting



Death coding



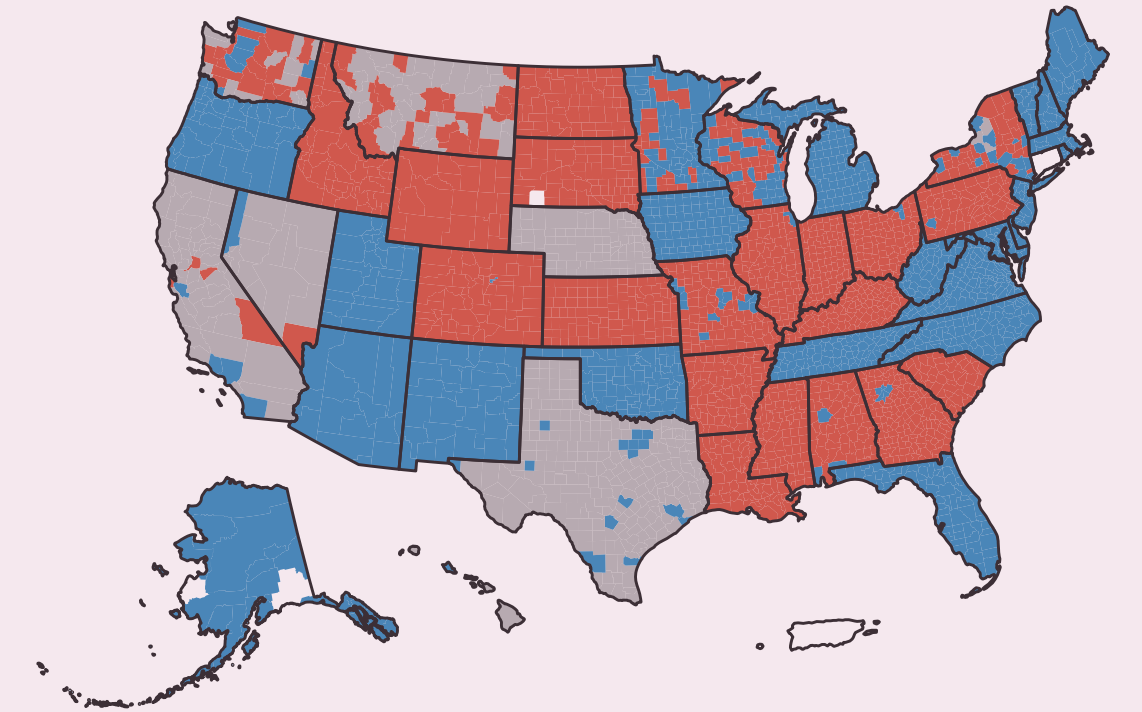
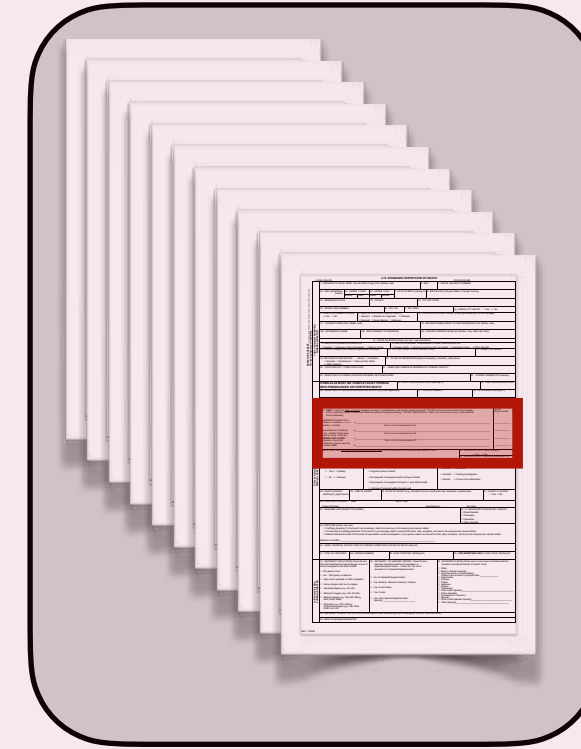
Data release



But far from perfect

- Like most high-income countries, overall US death data are quite good based on traditional quality metrics.¹⁻³
- But there are issues:
 - Lack of specificity for drug overdose deaths⁴⁻⁵
 - Suicide deaths are poorly identified⁶
 - Alzheimer's likely undercounted⁷
 - Maternal mortality like undercounted⁸
 - COVID-19 undercounted⁹
 - Coronary heart disease overcounted¹⁰

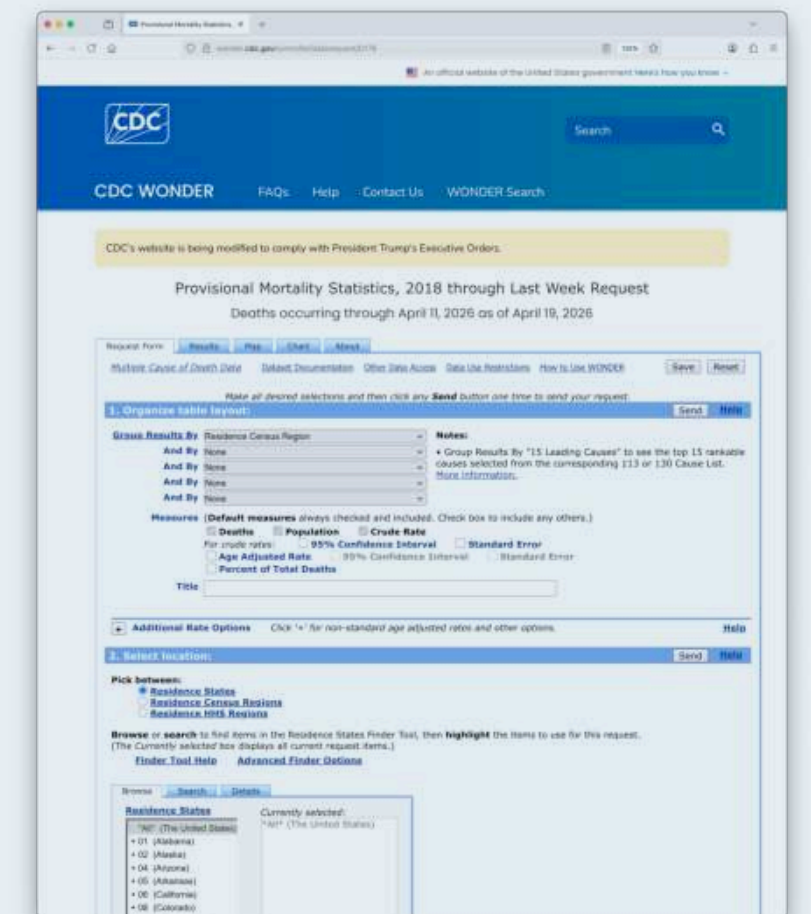
Death reporting



Death coding

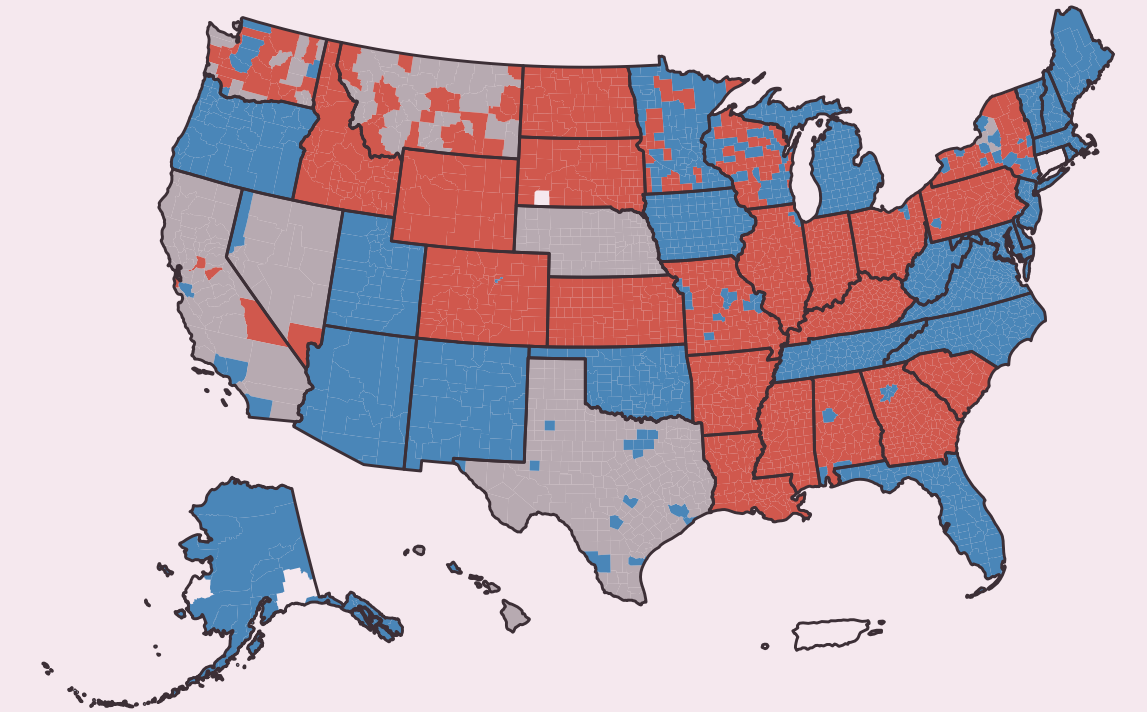
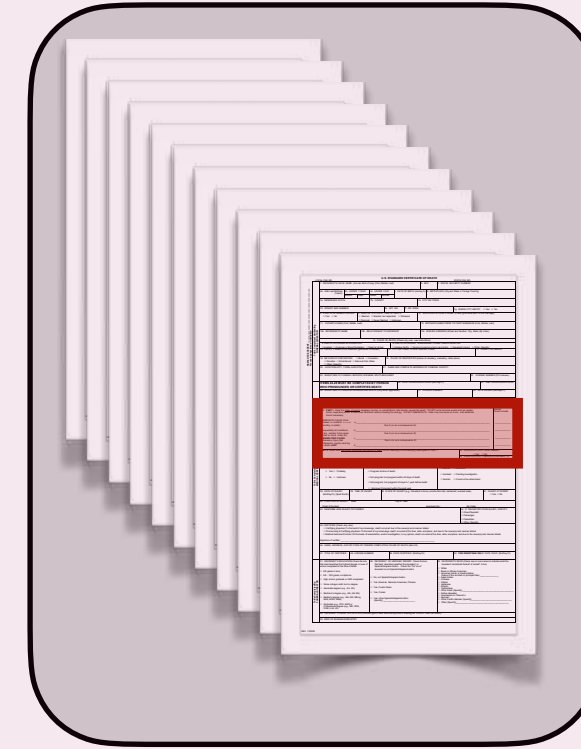


Data release



Research Objective

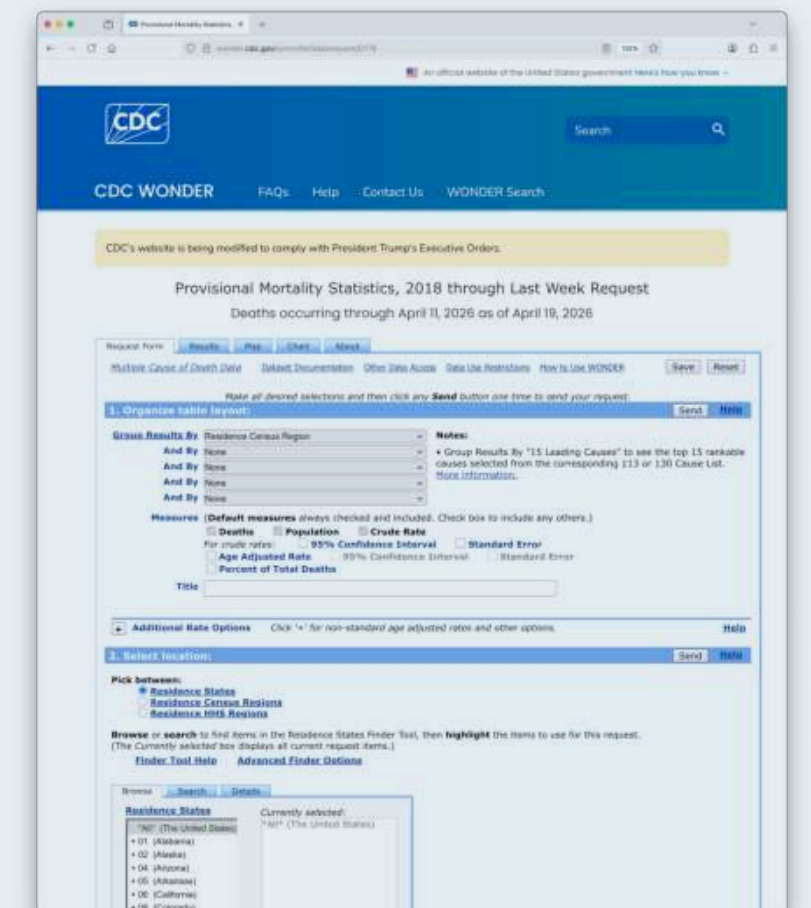
Death reporting



Death coding



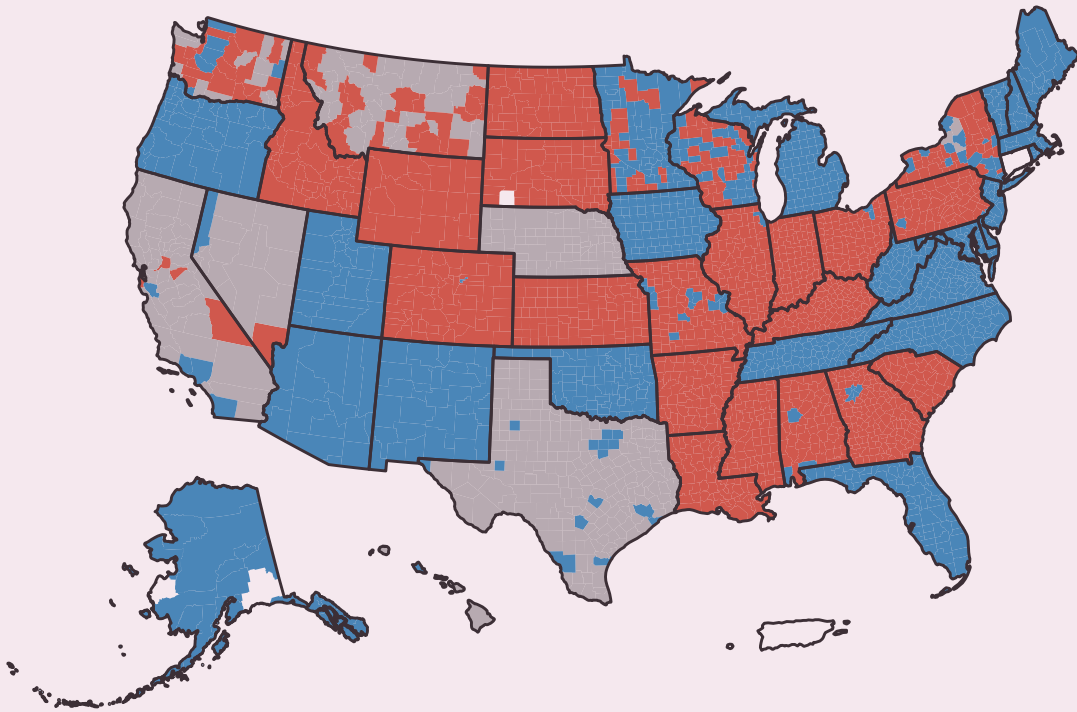
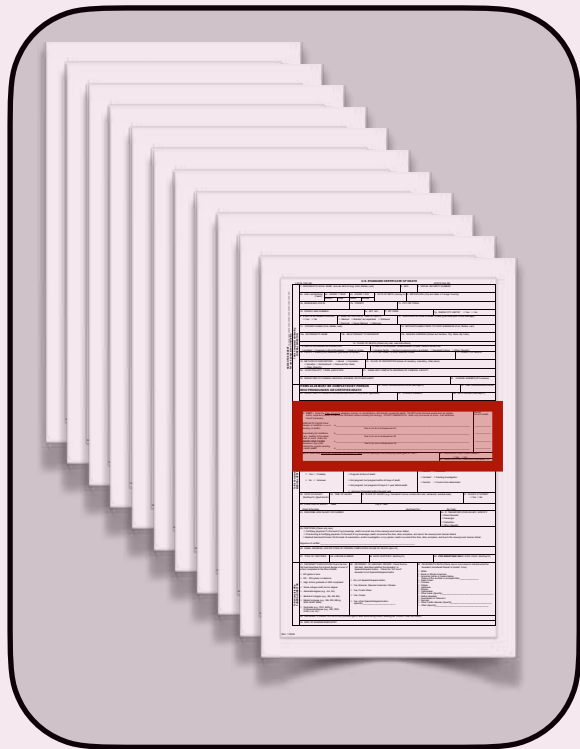
Data release



Research Objective

Develop measures of death reporting quality that are suitable for the context of the United States

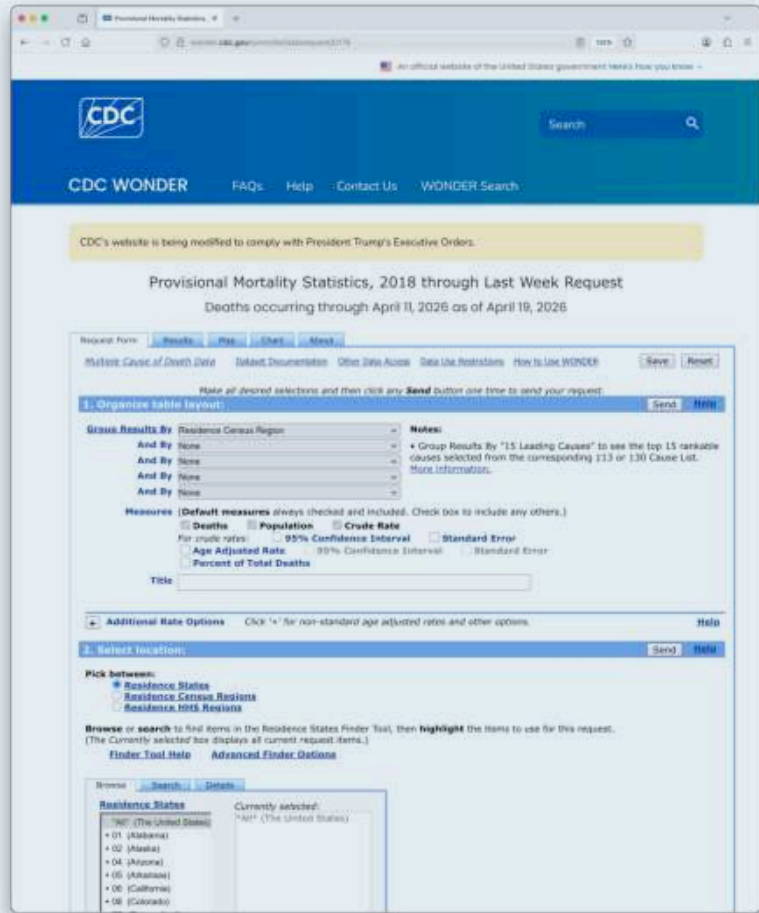
Death reporting



Death coding



Data release



Methods

Methods

- **Data:** NCHS multiple cause of death files, 1999-2022
- **Measures:**
 1. Proportion garbage codes
 2. Level of detail
 3. Re-assignability index
 4. Aggregated quality index
- **Analyses:**
 1. Correlations with spatially-adjusted p-values
 2. Descriptive trends over time and space

Proportion Garbage Codes

- **Intuition:** What proportion of UCOD codes are not useful for public health?

Proportion Garbage Codes

- **Intuition:** What proportion of UCOD codes are not useful for public health?

$$\text{Prop. GC}_c = \frac{\text{Number of deaths with garbage codes}_c}{\text{Total number of deaths}_c}$$

- List of garbage codes developed by IHME, and used as our **baseline measure**
- Examples:
 - R99 "Other ill-defined and unspecified causes of mortality."
 - A41.9 "Sepsis, unspecified organism."

Level of detail

- **Intuition:** How much variation is there in the (non-garbage code) UCOD codes? In other words, do counties "default" to specific reporting habits?

Level of detail

- **Intuition:** How much variation is there in the (non-garbage code) UCOD codes? In other words, do counties "default" to specific reporting habits?
- Examples:
 - Ischemic heart disease (I20-I25)

Level of detail

- **Intuition:** How much variation is there in the (non-garbage code) UCOD codes? In other words, do counties "default" to specific reporting habits?
- Examples:
 - Ischemic heart disease (I20-I25)

ICD-10 Code	Horry, SC (n=538)	East Baton Rouge, LA (n=533)
I25.0 - Atherosclerotic cardiovascular disease, so described	14%	82%
I25.1 - Atherosclerotic heart disease	25%	8%
I25.9 - Chronic ischaemic heart disease, unspecified	53%	7%
Other I20-I25 code	8%	3%

Level of detail

- **Intuition:** How much variation is there in the (non-garbage code) UCOD codes? In other words, do counties "default" to specific reporting habits?
- Examples:
 - Ischemic heart disease (I20-I25)

ICD-10 Code	Horry, SC (n=538)	East Baton Rouge, LA (n=533)
I25.0 - Atherosclerotic cardiovascular disease, so described	14%	82%
I25.1 - Atherosclerotic heart disease	25%	8%
I25.9 - Chronic ischaemic heart disease, unspecified	53%	7%
Other I20-I25 code	8%	3%

Level of detail

- **Intuition:** How much variation is there in the (non-garbage code) UCOD codes? In other words, do counties "default" to specific reporting habits?
- Examples:
 - Ischemic heart disease (I20-I25)

ICD-10 Code	Horry, SC (n=538)	East Baton Rouge, LA (n=533)
I25.0 - Atherosclerotic cardiovascular disease, so described	14%	82%
I25.1 - Atherosclerotic heart disease	25%	8%
I25.9 - Chronic ischaemic heart disease, unspecified	53%	7%
Other I20-I25 code	8%	3%

Step 1 - Get standardized probabilities

- **Within-cause distribution:** For county set k in time t , fraction of deaths that get ICD-10 code d conditional on a death being in cause group c .

$$p_{k,t}^*(d) = \sum_{c \in \mathcal{C}} s_t(c) w_{k,t}(d \mid c)$$

Step 1 - Get standardized probabilities

- **Within-cause distribution:** For county set k in time t , fraction of deaths that get ICD-10 code d conditional on a death being in cause group c .
- **National share weight:** For time t , proportion of deaths are in cause group c .

$$p_{k,t}^*(d) = \sum_{c \in \mathcal{C}} s_t(c) w_{k,t}(d \mid c)$$

Step 1 - Get standardized probabilities

- **Within-cause distribution:** For county set k in time t , fraction of deaths that get ICD-10 code d conditional on a death being in cause group c .
- **National share weight:** For time t , proportion of deaths are in cause group c .
- **Sum across all cause groups** observed in time t .

$$p_{k,t}^*(d) = \sum_{c \in \mathcal{C}} s_t(c) w_{k,t}(d \mid c)$$

Step 1 - Get standardized probabilities

- **Within-cause distribution:** For county set k in time t , fraction of deaths that get ICD-10 code d conditional on a death being in cause group c .
- **National share weight:** For time t , proportion of deaths are in cause group c .
- **Sum across all cause groups** observed in time t .
- **Standardized probability** that code d is reported, after adjusting county set k 's cause-mix to match the **national cause-mix** at time t

$$p_{k,t}^*(d) = \sum_{c \in \mathcal{C}} s_t(c) w_{k,t}(d \mid c)$$

Step 2 – Shannon entropy

- Calculate the **Shannon entropy** of that **standardized probability** distribution.

$$H_{k,t} = - \sum_{d \in \mathcal{D}_t} p_{k,t}^*(d) \log p_{k,t}^*(d)$$

Step 2 – Shannon entropy

- Calculate the **Shannon entropy** of that **standardized probability** distribution.
- Measures variability:
 - High entropy: county set k spreads deaths across many codes
 - Low entropy: county set k "defaults" to a few codes

$$H_{k,t} = - \sum_{d \in \mathcal{D}_t} p_{k,t}^*(d) \log p_{k,t}^*(d)$$

Step 2 – Shannon entropy

- Calculate the **Shannon entropy** of that **standardized probability** distribution.
- Measures variability:
 - High entropy: county set k spreads deaths across many codes
 - Low entropy: county set k "defaults" to a few codes

$$H_{k,t} = - \sum_{d \in \mathcal{D}_t} p_{k,t}^*(d) \log p_{k,t}^*(d)$$

- **(Step 3: Rescale** so range is 0-100.)
- Note: We use county sets because metric is sensitive to small samples

Re-assignability Index

- **Intuition:** Among deaths with garbage code UCODs, how useful is the *contributing* cause information?

Re-assignability Index

- **Intuition:** Among deaths with garbage code UCODs, how useful is the *contributing* cause information?
- Example: Two death certificates with UCOD I64 ("Stroke, not specified..."):

Death Certificate A	Death Certificate B
Part Ia: "Stroke"	Part Ia: "Stroke"
Part II: "Hypertension, AFib, type 2 diabetes"	Part II: "Dementia"

Re-assignability Index

- **Intuition:** Among deaths with garbage code UCODs, how useful is the *contributing* cause information?
- Example: Two death certificates with UCOD I64 ("Stroke, not specified..."):

Death Certificate A	Death Certificate B
Part Ia: "Stroke"	Part Ia: "Stroke"
Part II: "Hypertension, AFib, type 2 diabetes"	Part II: "Dementia"

Re-assignability Index

- **Intuition:** Among deaths with garbage code UCODs, how useful is the *contributing* cause information?
- Example: Two death certificates with UCOD I64 ("Stroke, not specified..."):

Death Certificate A	Death Certificate B
Part Ia: "Stroke"	Part Ia: "Stroke"
Part II: "Hypertension, AFib, type 2 diabetes"	Part II: "Dementia"

Classic risk factors for
ischemic stroke.

Re-assignability Index

- **Intuition:** Among deaths with garbage code UCODs, how useful is the *contributing* cause information?
- Example: Two death certificates with UCOD I64 ("Stroke, not specified..."):

Death Certificate A	Death Certificate B
Part Ia: "Stroke"	Part Ia: "Stroke"
Part II: "Hypertension, AFib, type 2 diabetes"	Part II: "Dementia"

Classic risk factors for
ischemic stroke.

Not informative.
Could be ischemic
or hemorrhagic stroke.

Re-assignability Index

- **Candidate cause set** for garbage code g , as defined in Foreman 2016¹¹.
 - Maps garbage groups g to plausible candidate causes of death k .
 - Example: I46 (cardiac arrest) would not get mapped to lung cancer.

$$h_i = \frac{-\sum_{k \in \mathcal{K}_g} p_i(k) \log p_i(k)}{\log |\mathcal{K}_g|} \in [0,1]$$

Re-assignability Index

- **Candidate cause set** for garbage code g , as defined in Foreman 2016¹¹.
- **Predicted probability** that death i 's true UCOD is candidate cause k
 - We train a regularized multinomial model on non-garbage code deaths where garbage code g appears as a contributing cause and the true UCOD is known. Use this model to predict probability p for each candidate cause k .

$$h_i = \frac{-\sum_{k \in \mathcal{K}_g} p_i(k) \log p_i(k)}{\log |\mathcal{K}_g|} \in [0,1]$$

Re-assignability Index

- **Candidate cause set** for garbage code g , as defined in Foreman 2016¹¹.
- **Predicted probability** that death i 's true UCOD is cause k
- **Per-death normalized entropy** for death i .
 - High entropy = death could be re-assigned to many causes
 - Low entropy = death could be re-assigned to small number of causes

$$h_i = \frac{-\sum_{k \in \mathcal{K}_g} p_i(k) \log p_i(k)}{\log |\mathcal{K}_g|} \in [0,1]$$

Re-assignability Index

- **Candidate cause set** for garbage code g , as defined in Foreman 2016¹¹.
- **Predicted probability** that death i 's true UCOD is cause k
- **Per-death normalized entropy** for death i .
 - High entropy = death could be re-assigned to many causes
 - Low entropy = death could be re-assigned to small number of causes

$$h_i = \frac{-\sum_{k \in \mathcal{K}_g} p_i(k) \log p_i(k)}{\log |\mathcal{K}_g|} \in [0,1]$$

- Take the average across all garbage code deaths in that county

Aggregated Quality Index

- z-score of the **proportion garbage codes** (sign flipped for consistency)
- z-score of **level of detail**
- z-score of **re-assignability index**
- Positive = better than national average; Negative = worse than national average

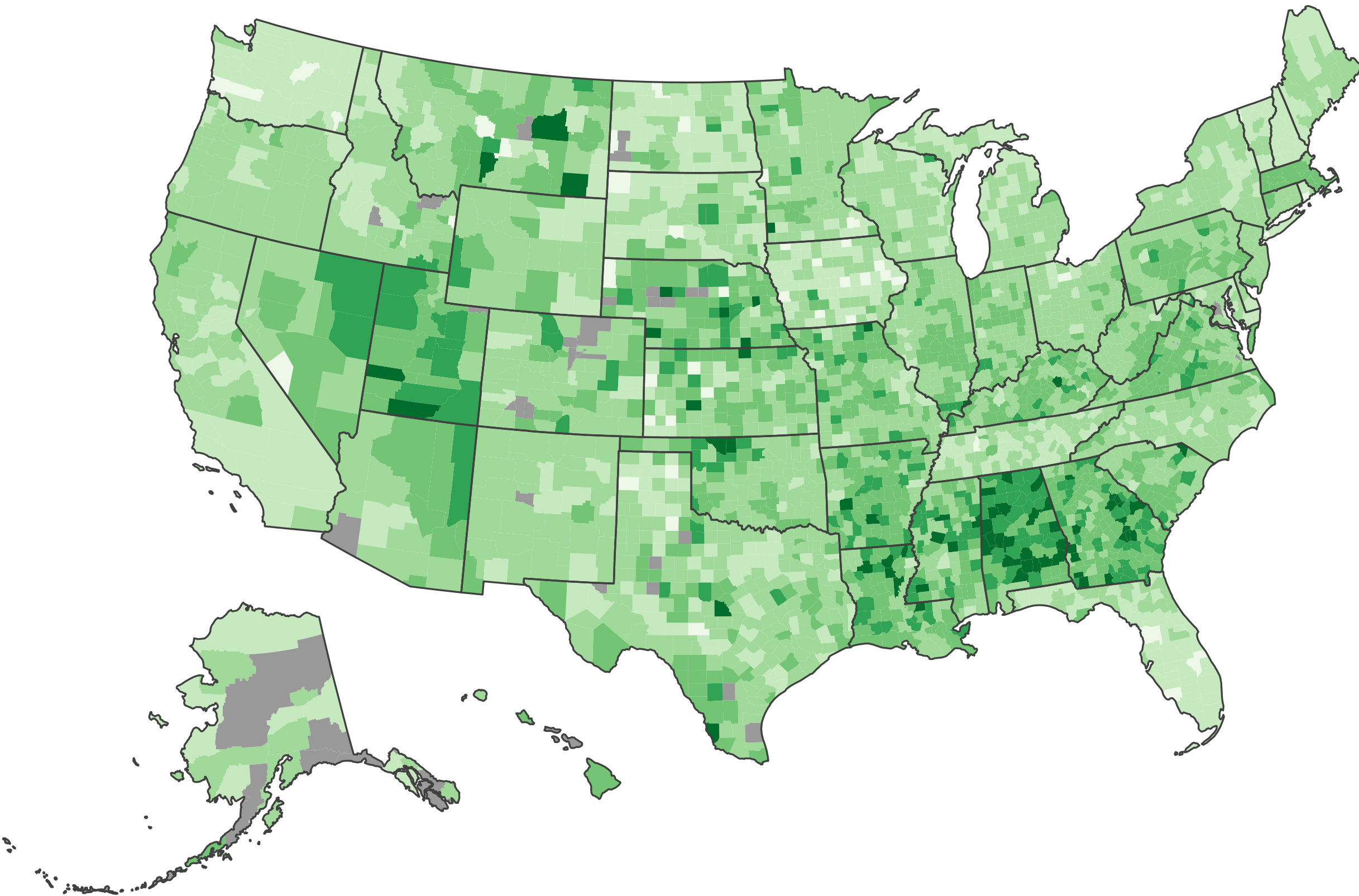
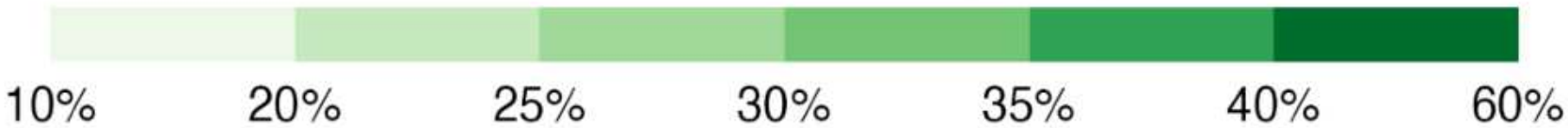
$$A_{c,t} = \frac{1}{3} \left(\frac{\mu_g - g_{c,t}}{\sigma_g} + \frac{d_{c,t} - \mu_d}{\sigma_d} + \frac{r_{c,t} - \mu_r}{\sigma_r} \right)$$

- Not the "right" metric! Just a simple summary.

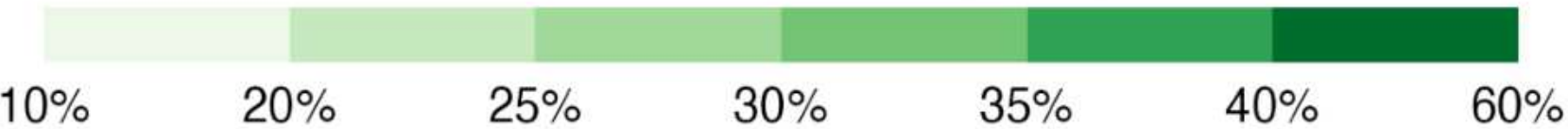
Results

Proportion garbage codes

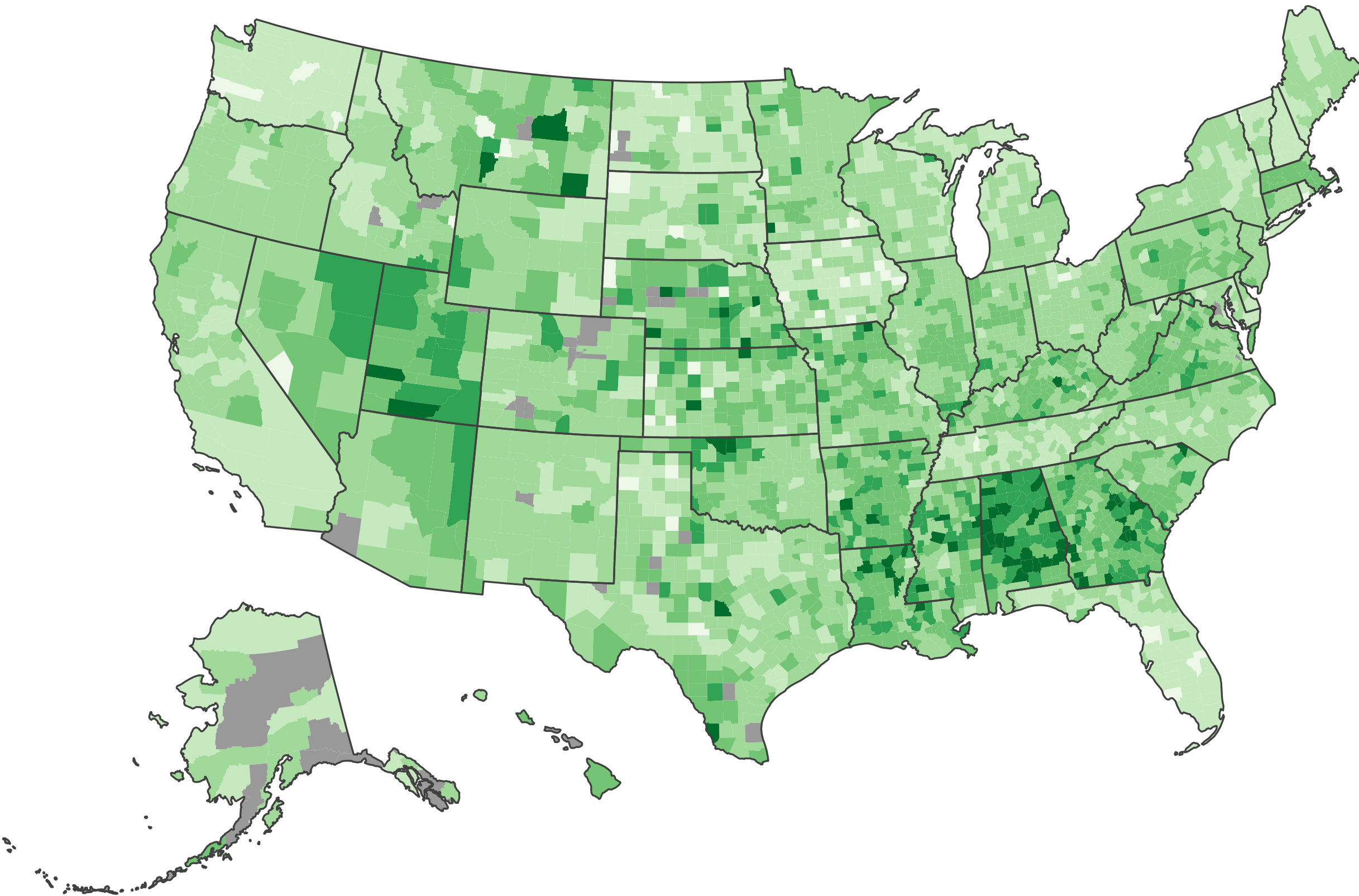
1999–2005



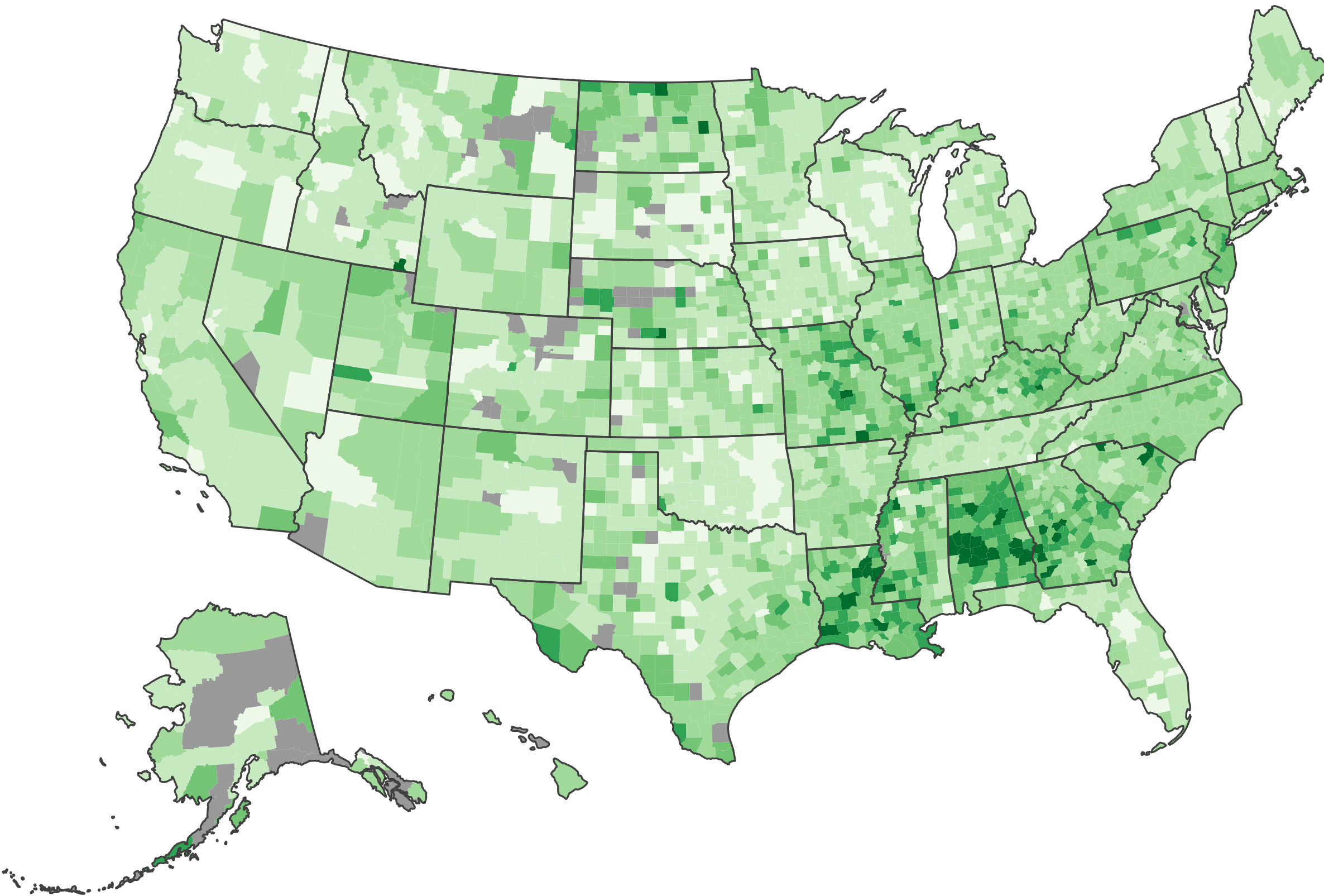
Proportion garbage codes



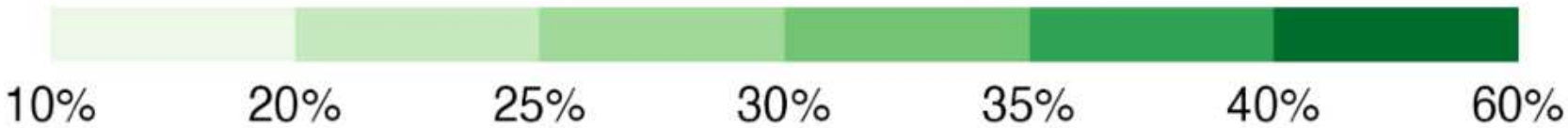
1999–2005



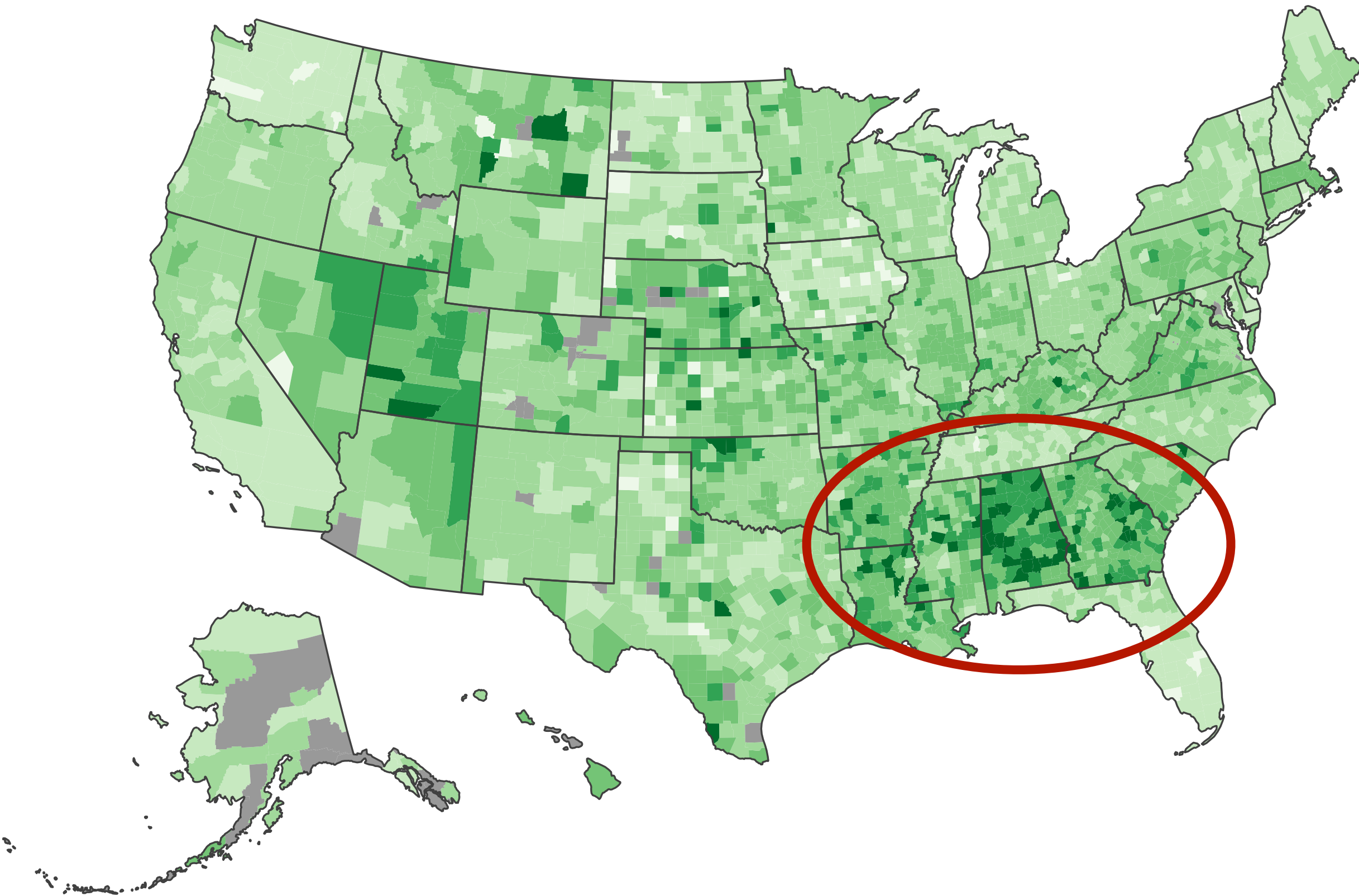
2020–2022



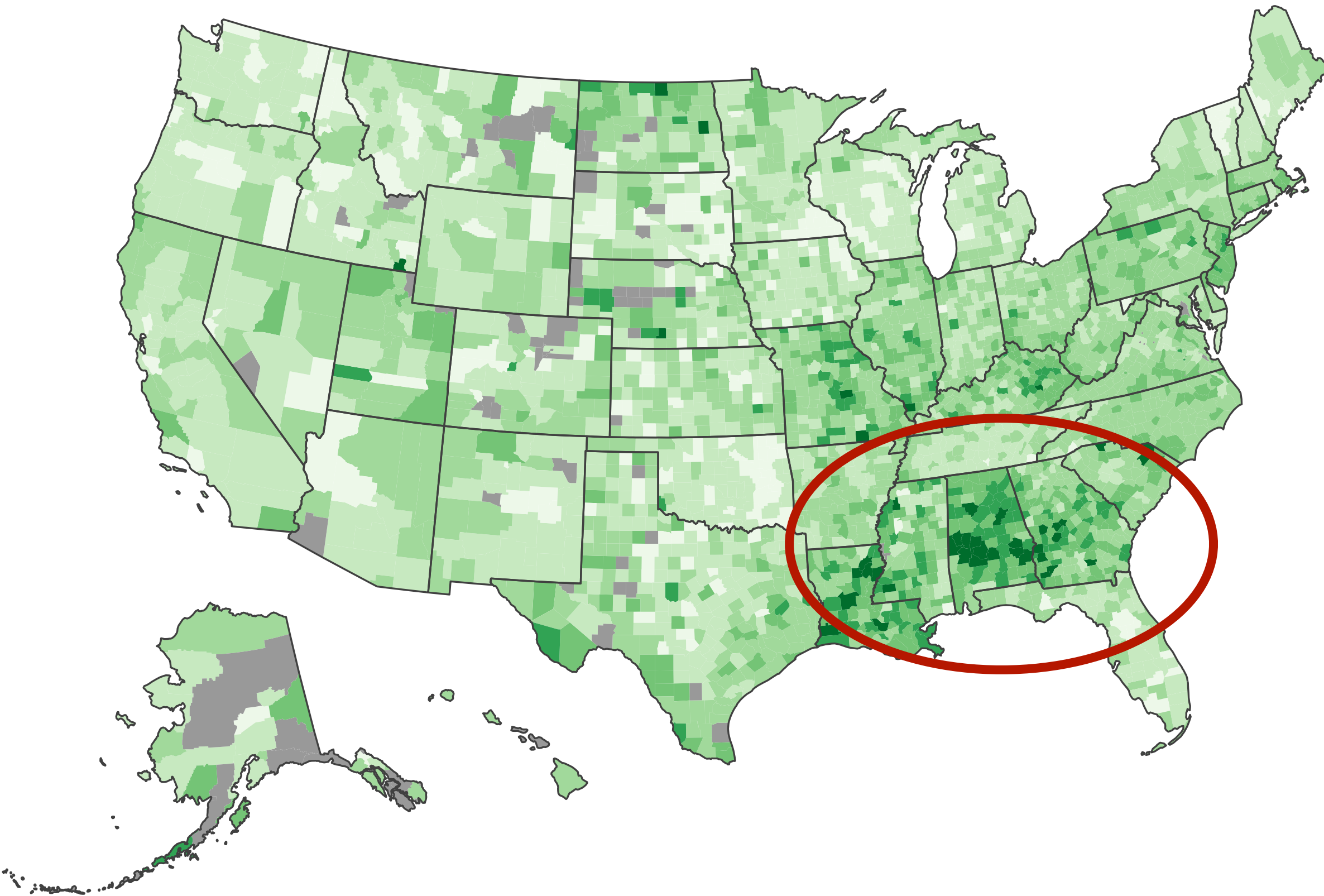
Proportion garbage codes



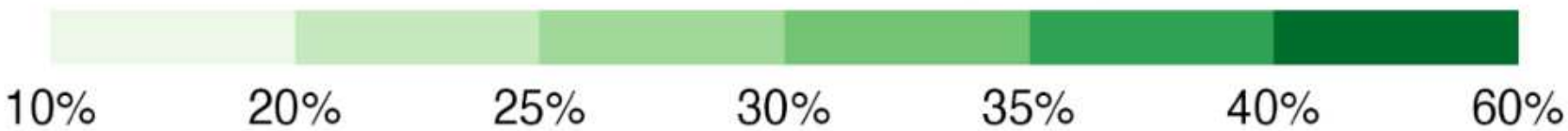
1999–2005



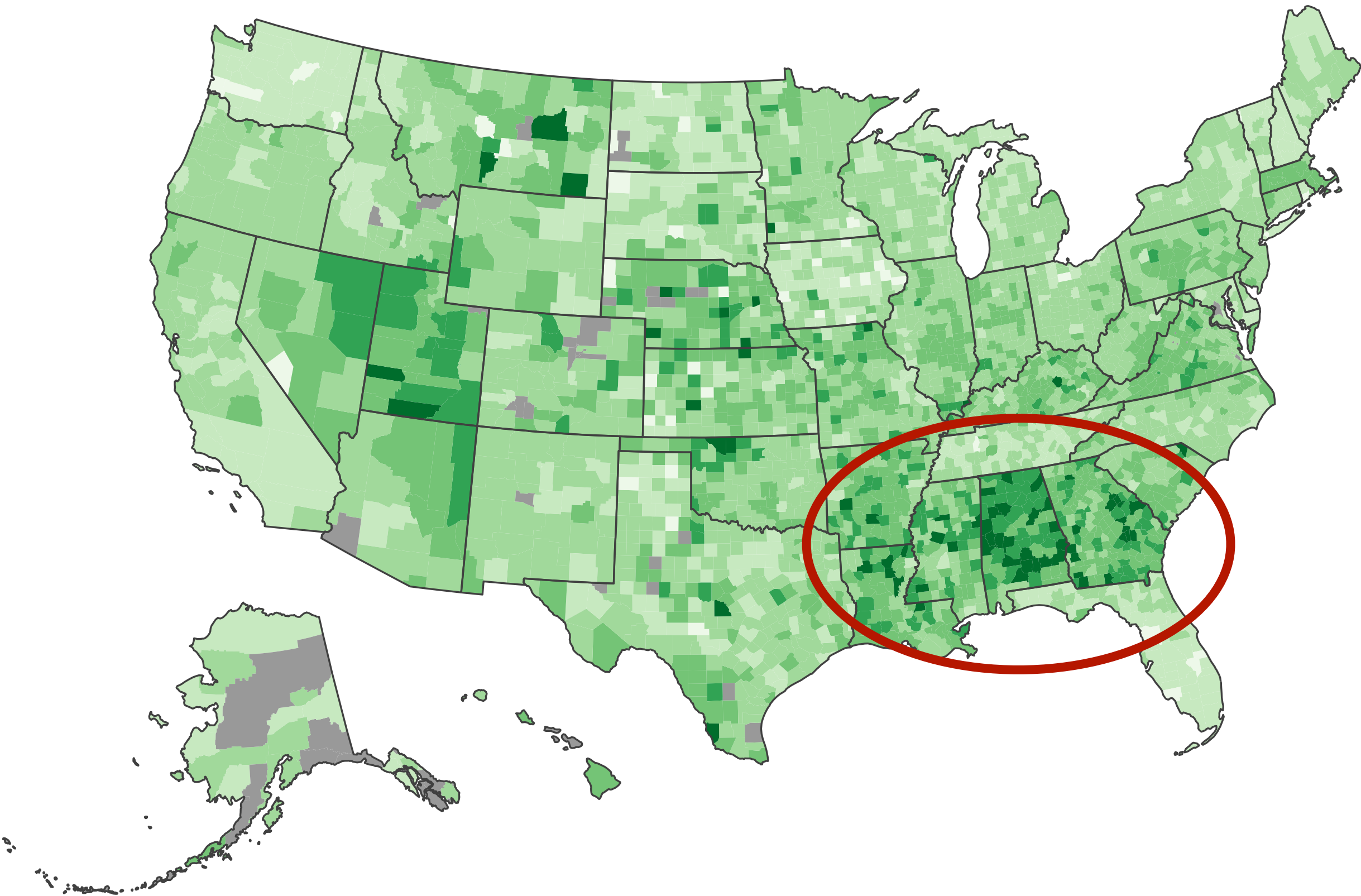
2020–2022



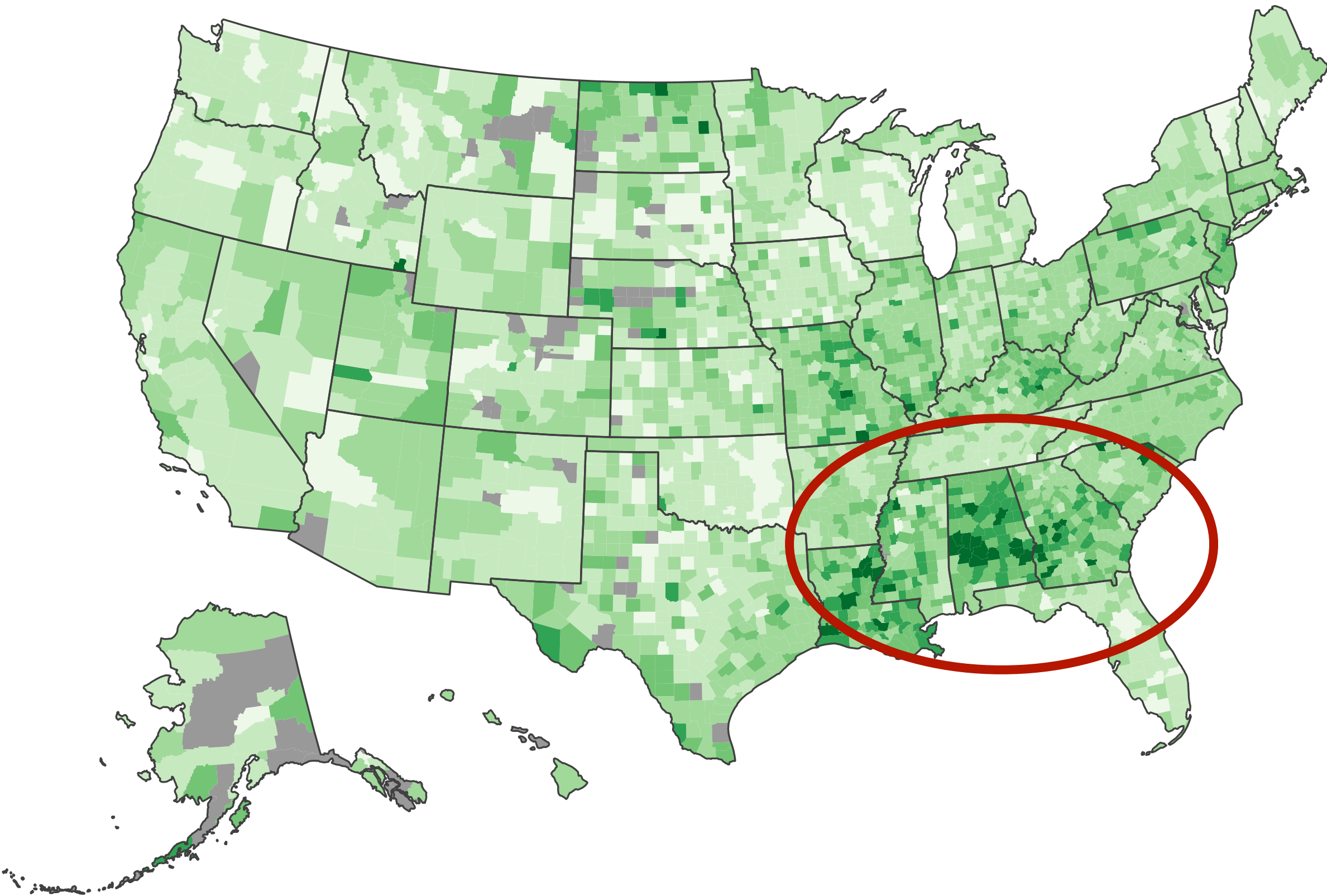
Proportion garbage codes



1999–2005

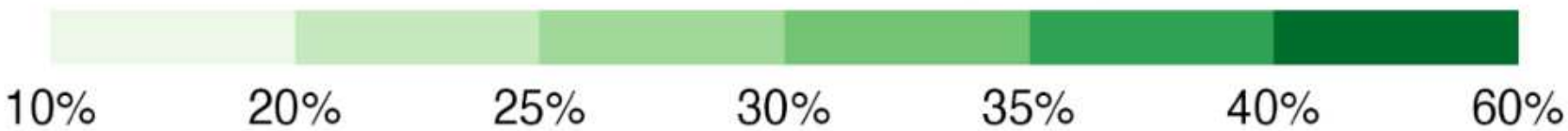


2020–2022

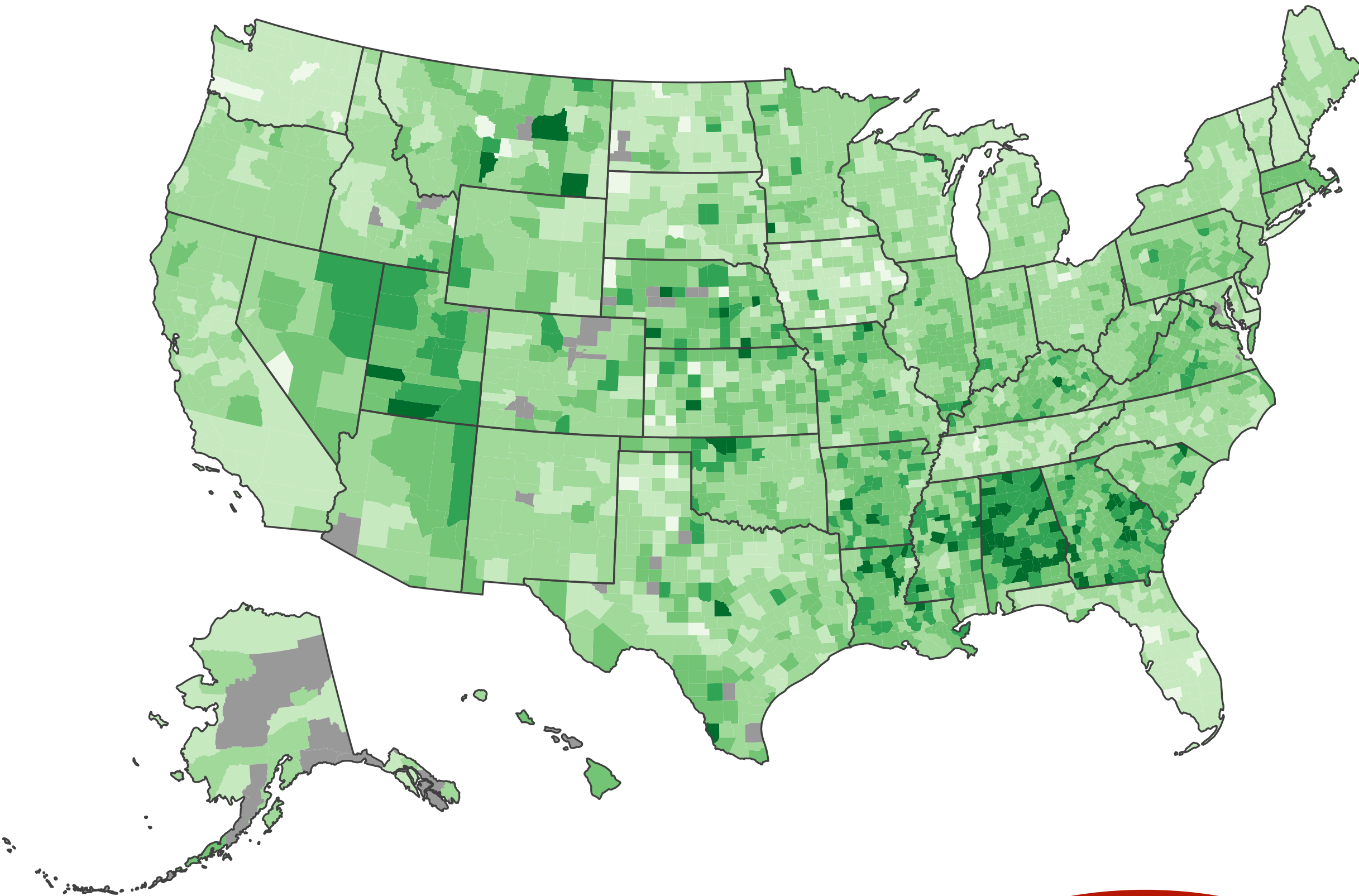


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	−0.24 (p = 0.002)	−0.20 (p = 0.004)
Public health spending per capita	−0.30 (p = 0.002)	−0.18 (p = 0.012)
Death investigation type (ME)	−0.40 (p = 0.002)	−0.23 (p = 0.004)

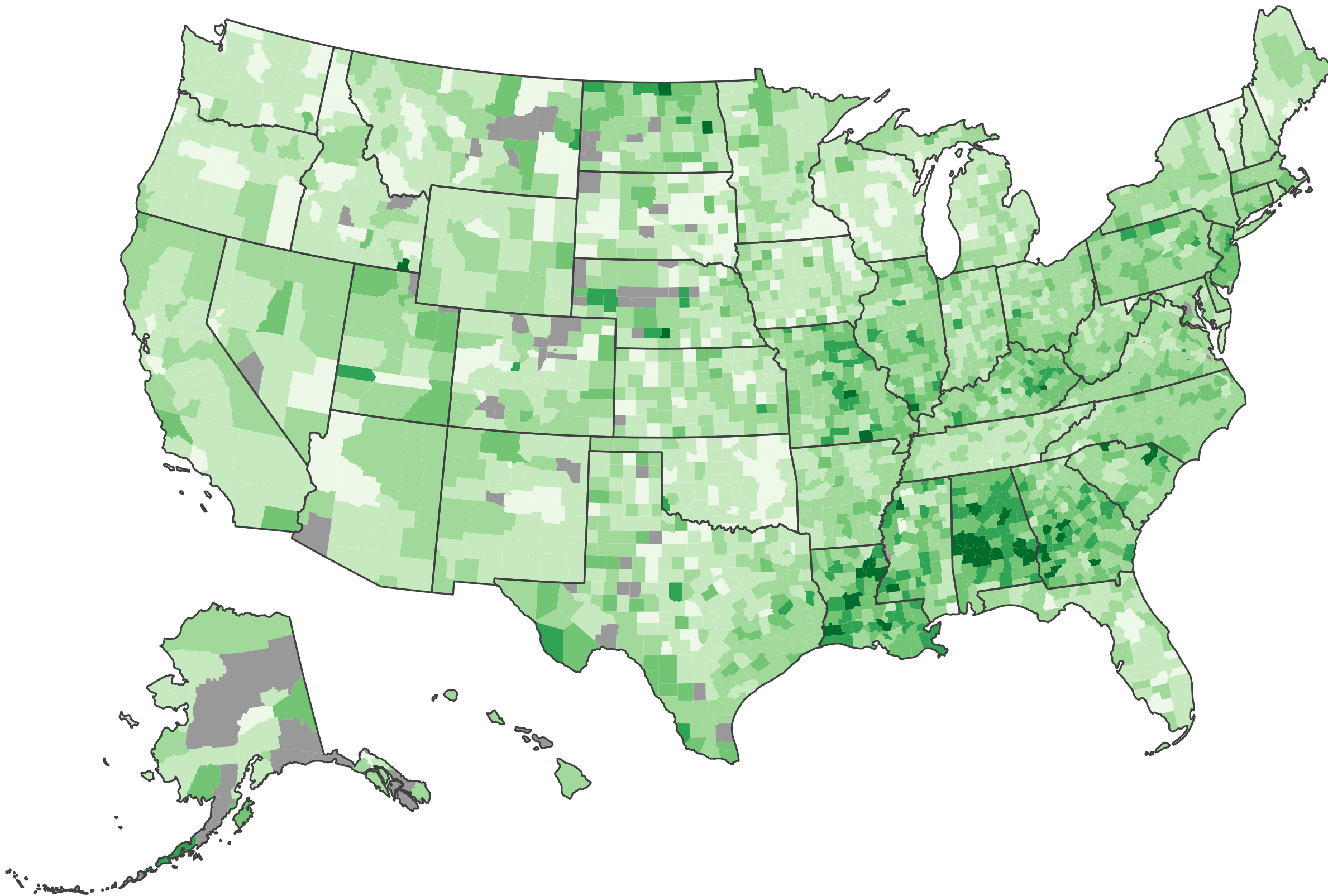
Proportion garbage codes



1999–2005

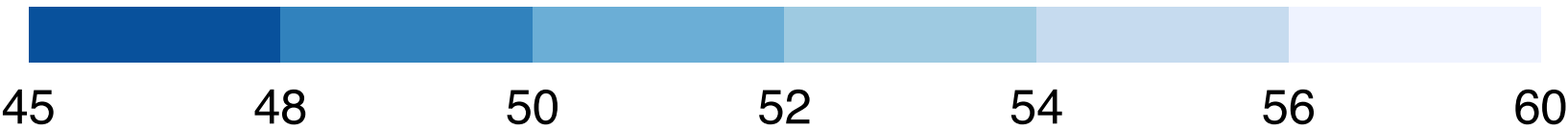


2020–2022

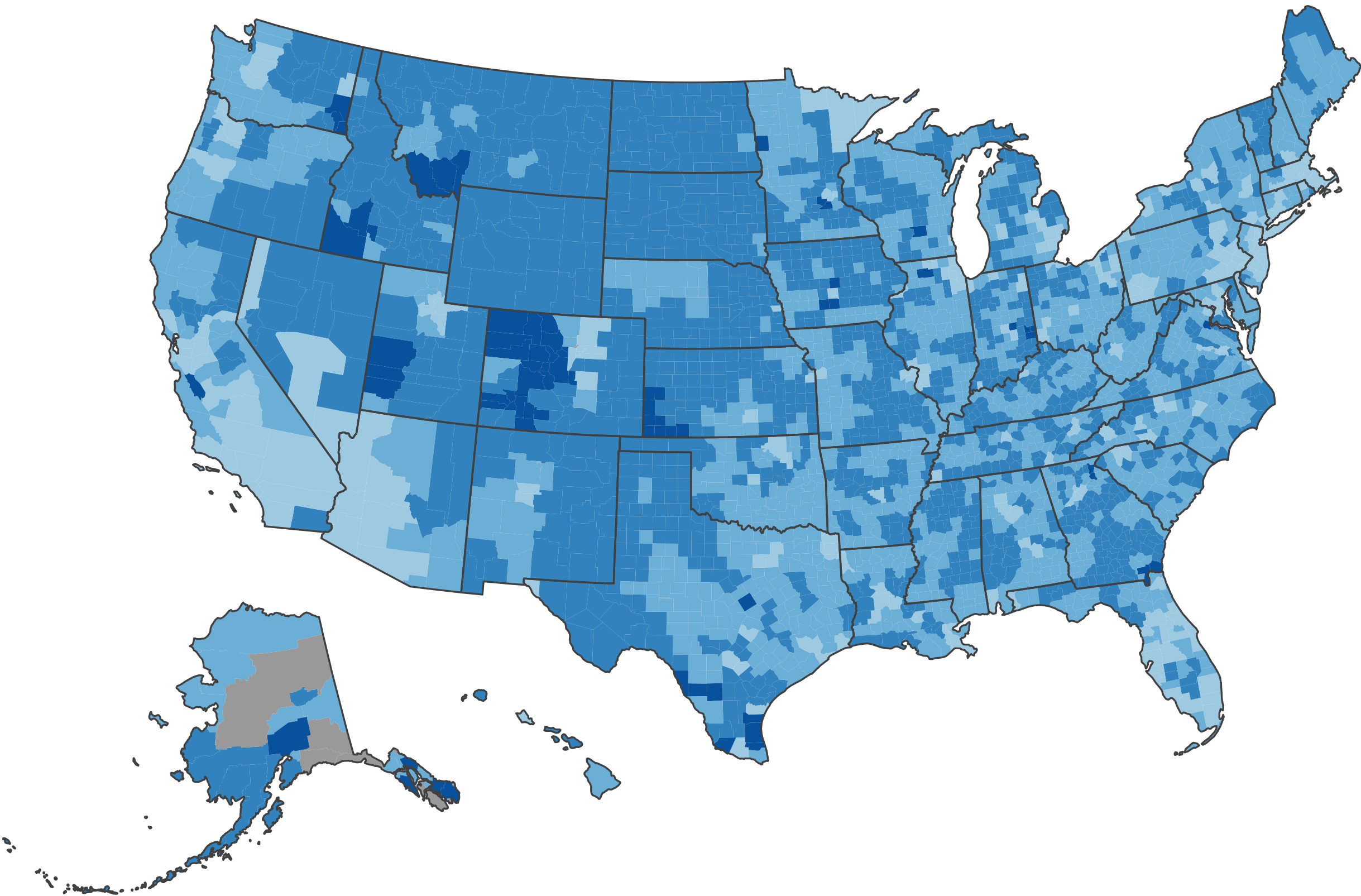


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	−0.24 (p = 0.002)	−0.20 (p = 0.004)
Public health spending per capita	−0.30 (p = 0.002)	−0.18 (p = 0.012)
Death investigation type (ME)	−0.40 (p = 0.002)	−0.23 (p = 0.004)

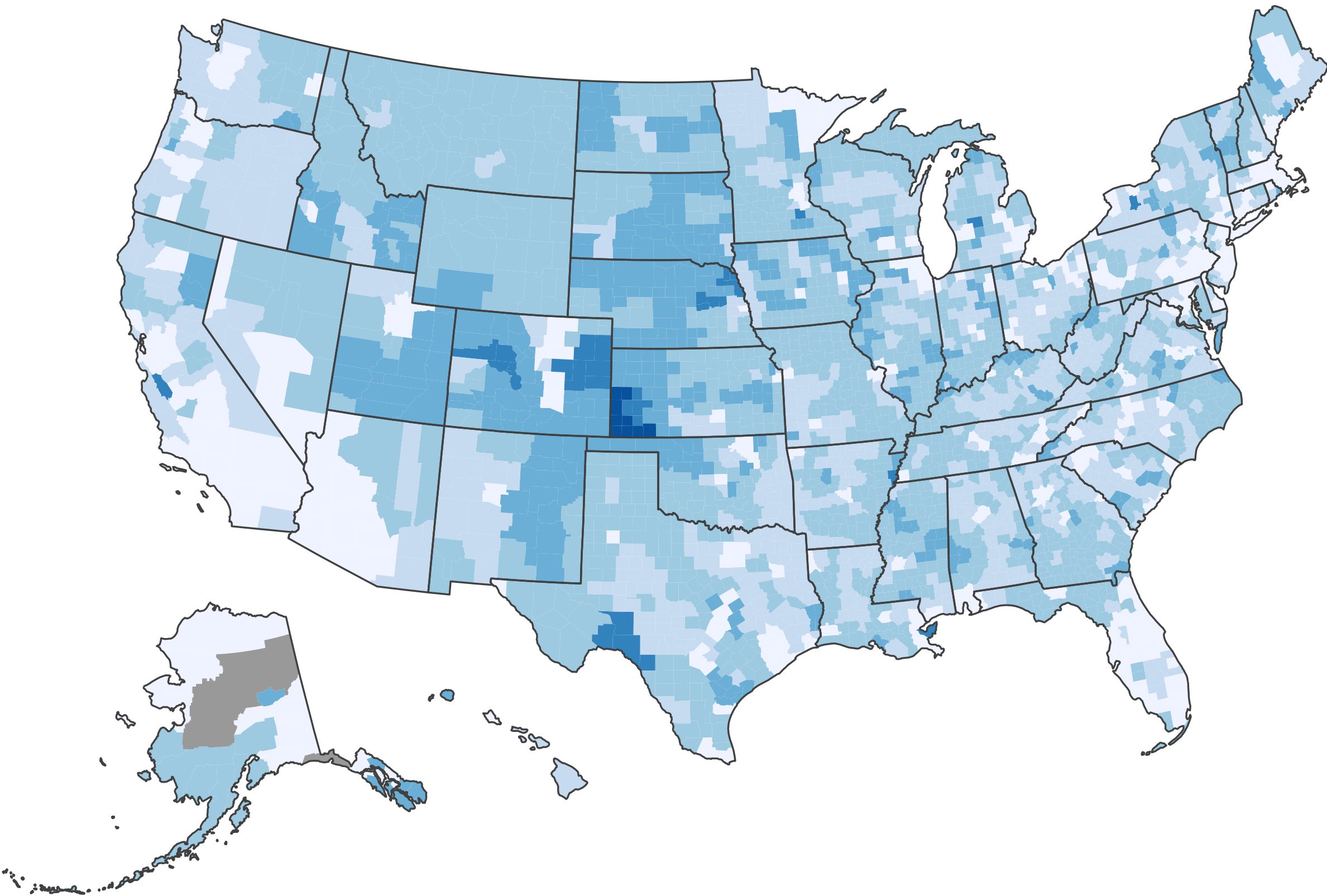
Level of detail



1999–2005

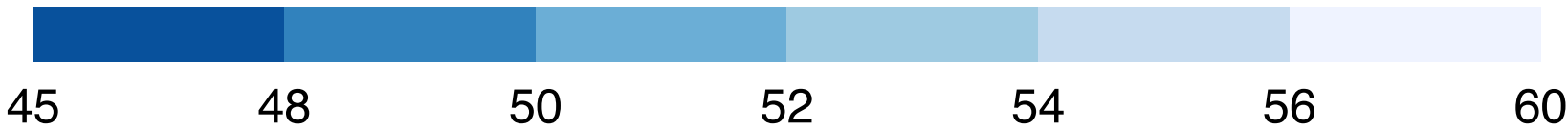


2020–2022

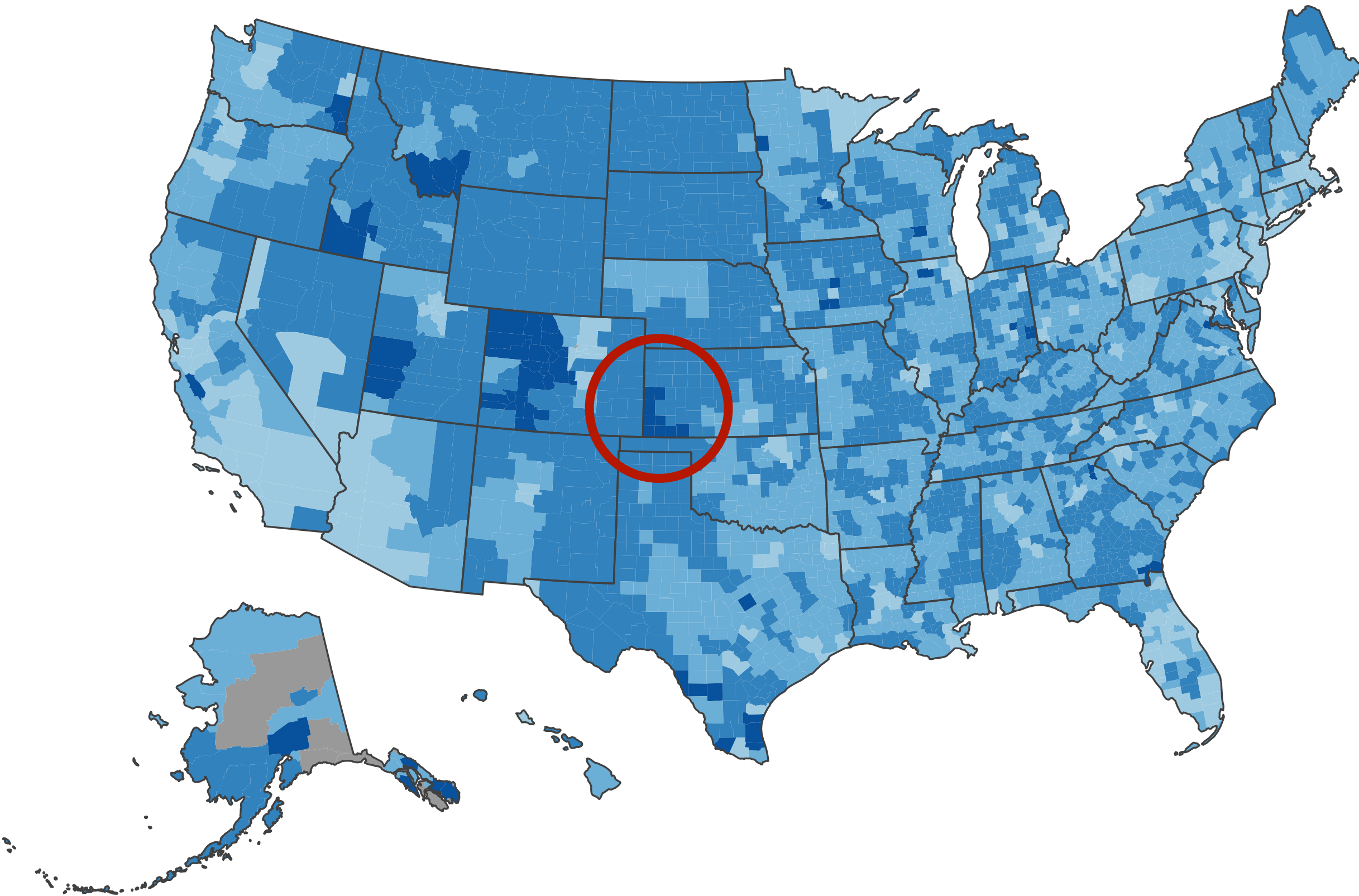


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.34 ($p = 0.004$)	+0.37 ($p = 0.002$)
Public health spending per capita	+0.20 ($p = 0.048$)	+0.17 ($p = 0.10$)
Death investigation type (ME)	+0.32 ($p = 0.002$)	+0.32 ($p = 0.002$)

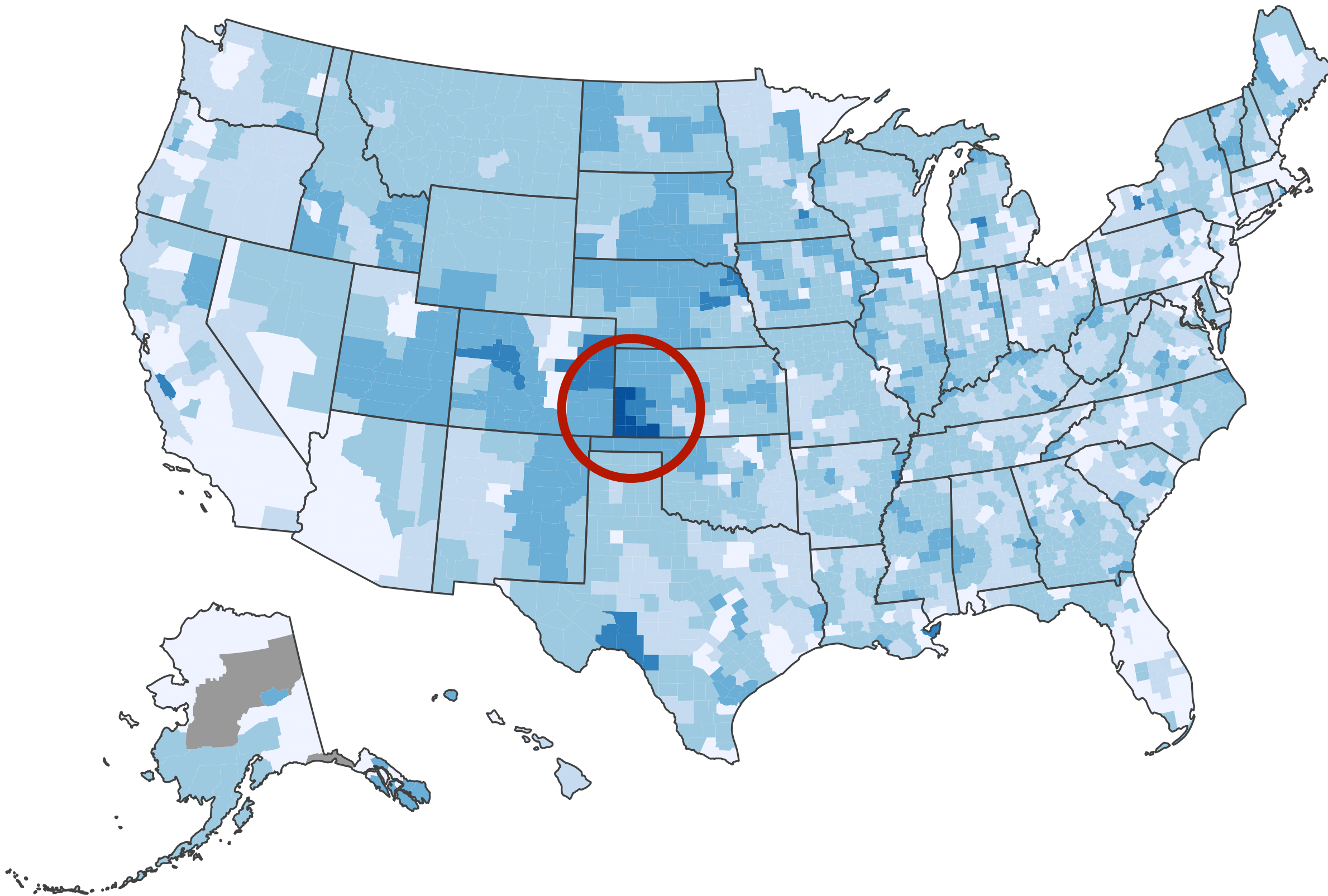
Level of detail



1999–2005

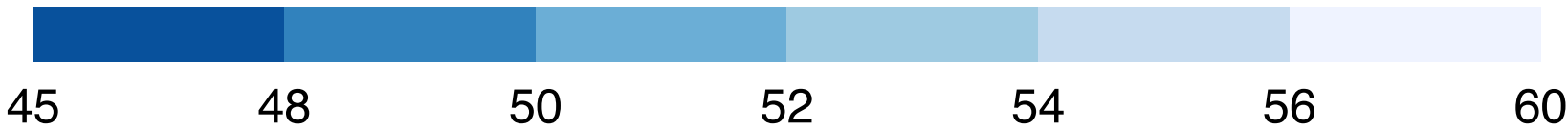


2020–2022

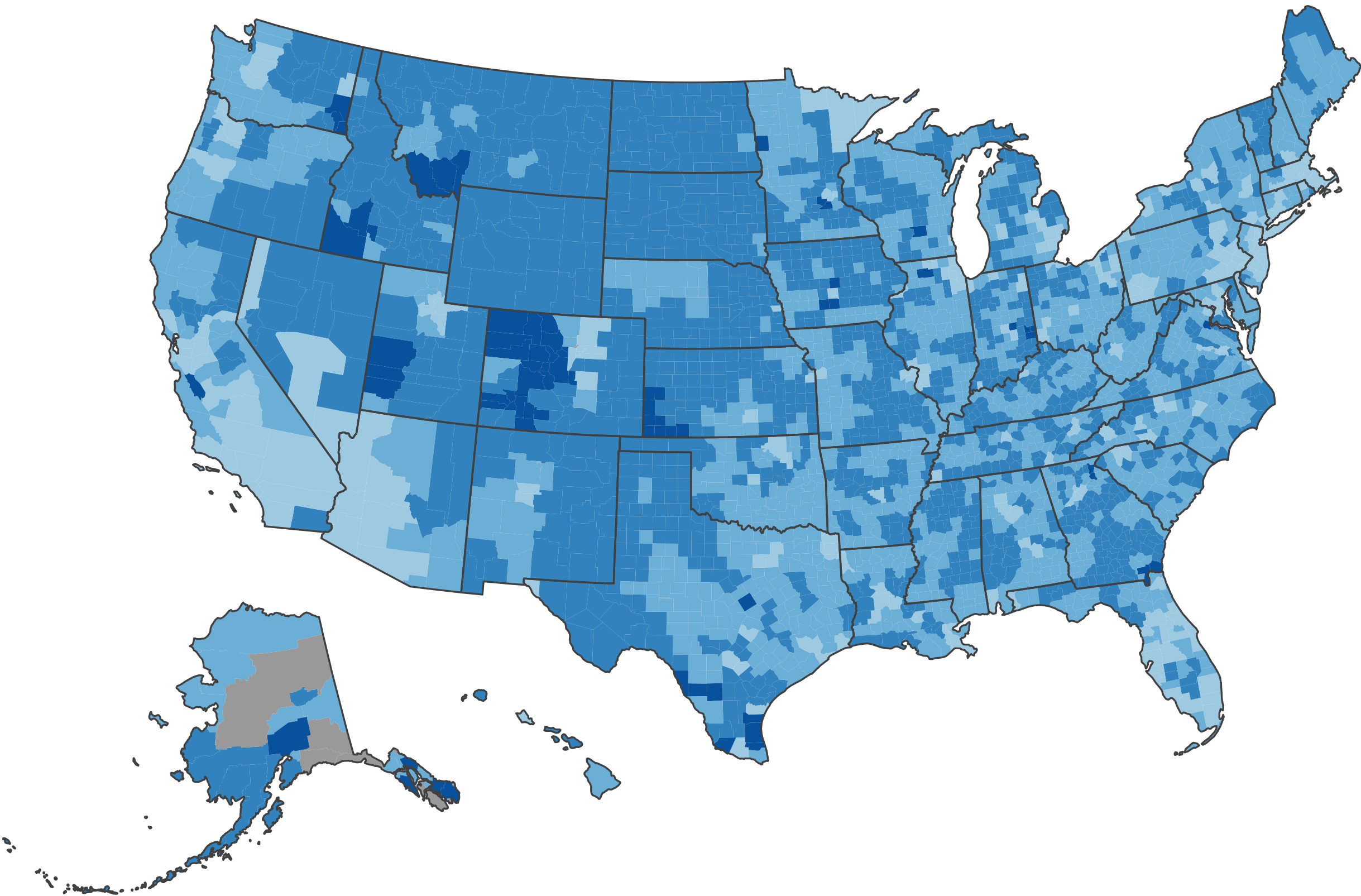


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.34 ($p = 0.004$)	+0.37 ($p = 0.002$)
Public health spending per capita	+0.20 ($p = 0.048$)	+0.17 ($p = 0.10$)
Death investigation type (ME)	+0.32 ($p = 0.002$)	+0.32 ($p = 0.002$)

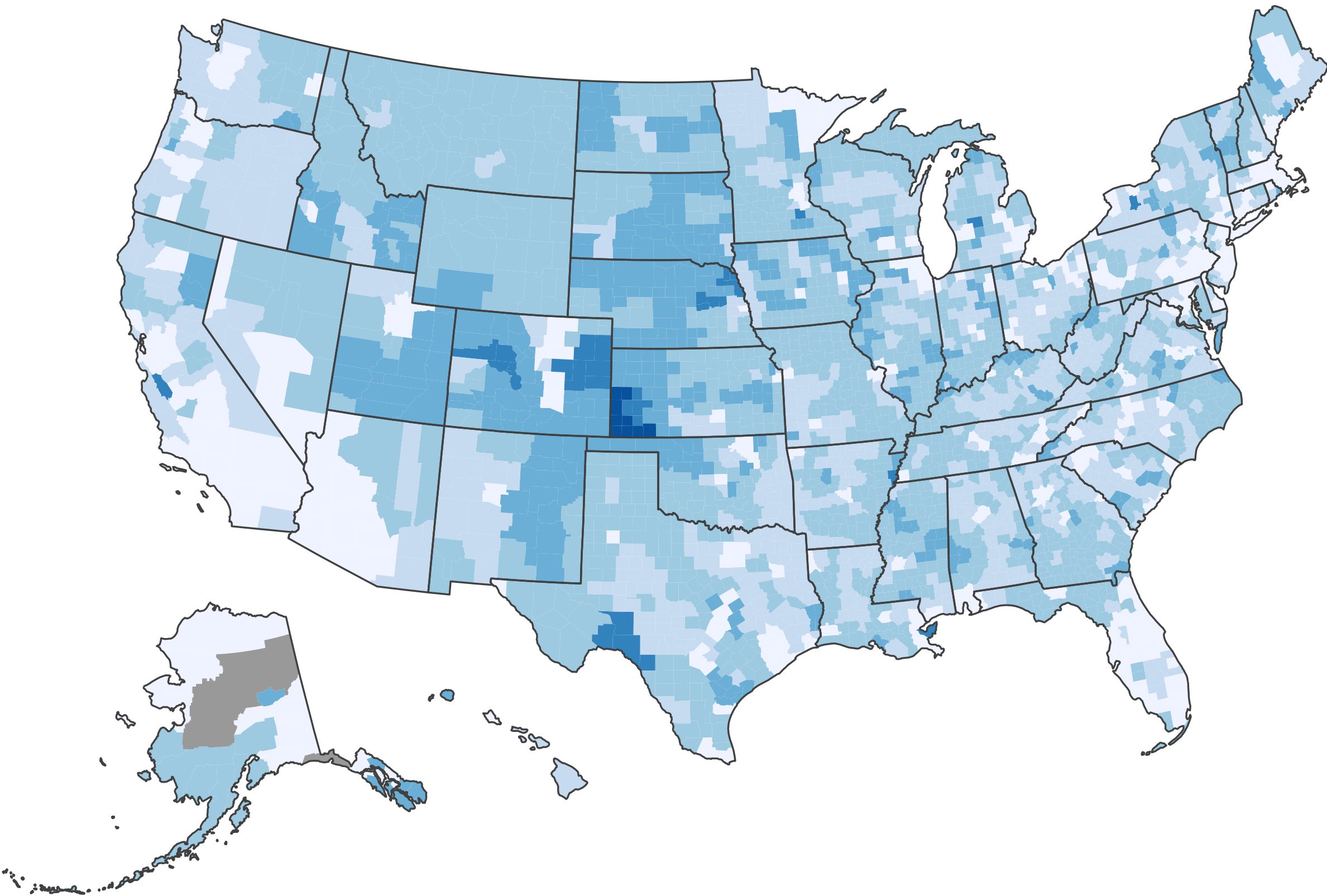
Level of detail



1999–2005

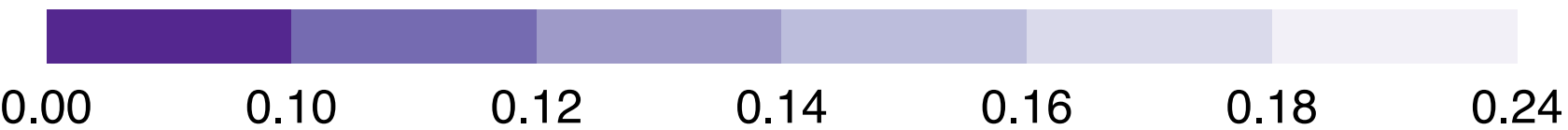


2020–2022

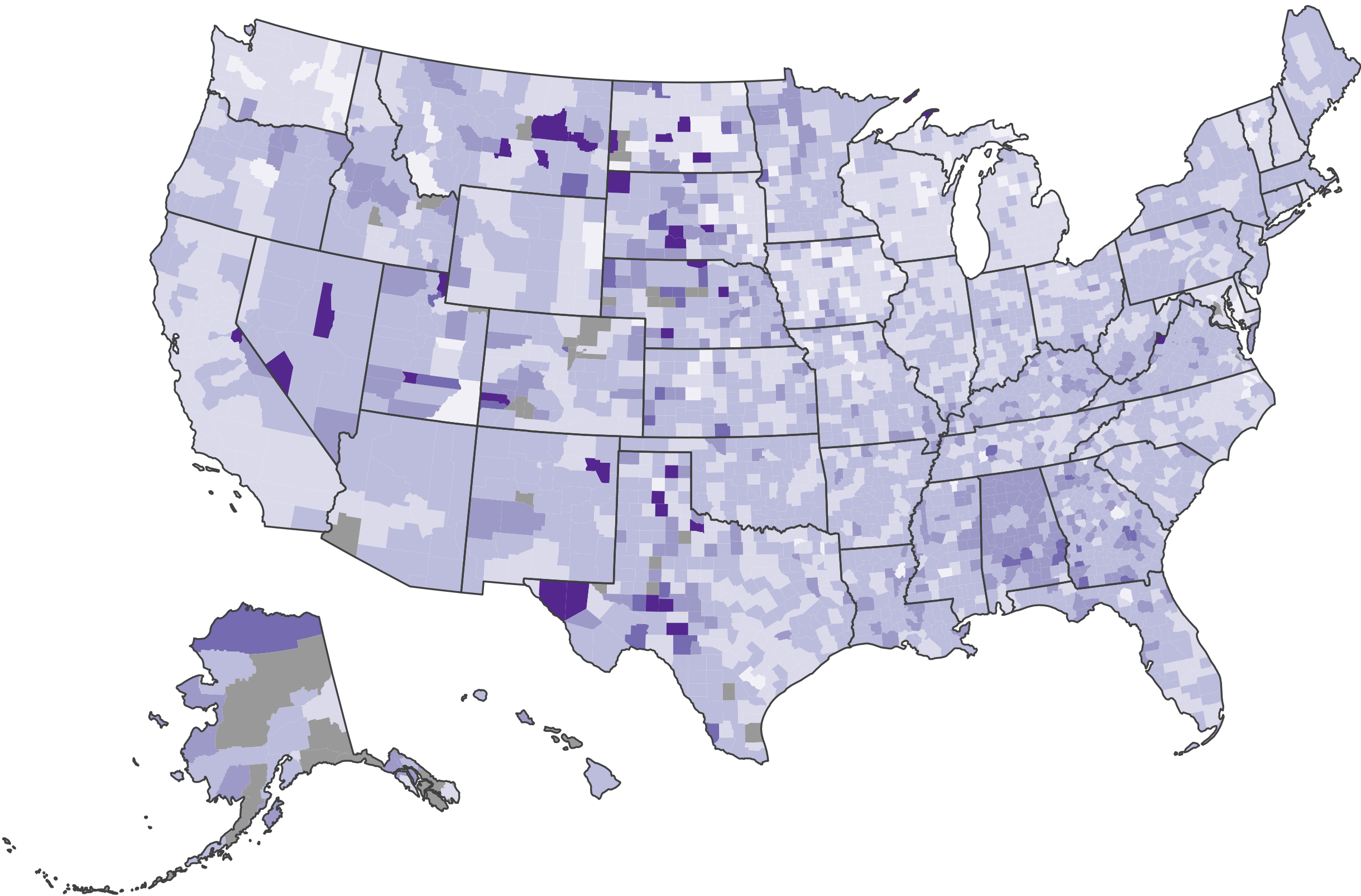


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.34 ($p = 0.004$)	+0.37 ($p = 0.002$)
Public health spending per capita	+0.20 ($p = 0.048$)	+0.17 ($p = 0.10$)
Death investigation type (ME)	+0.32 ($p = 0.002$)	+0.32 ($p = 0.002$)

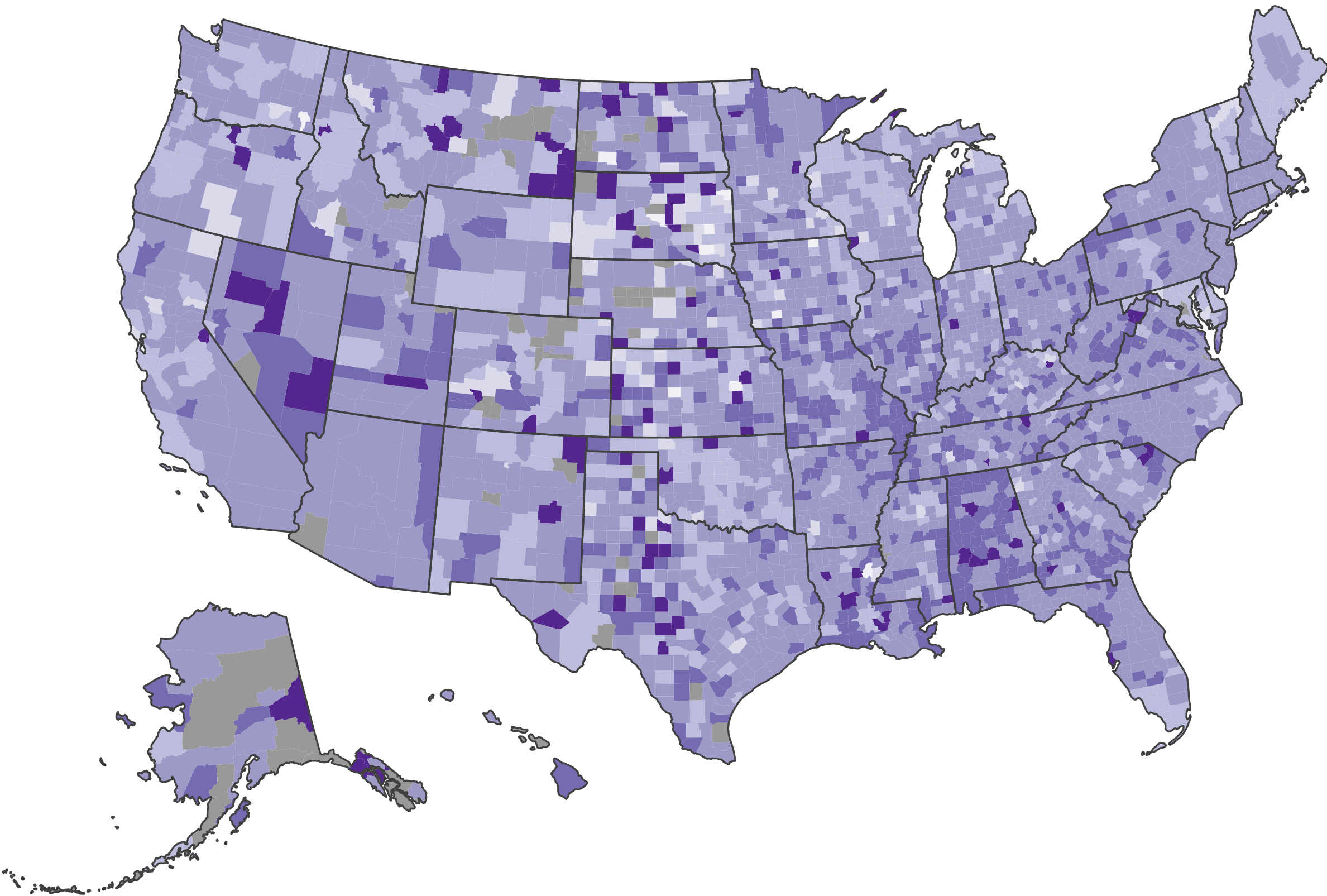
Re-assignability index



1999–2005

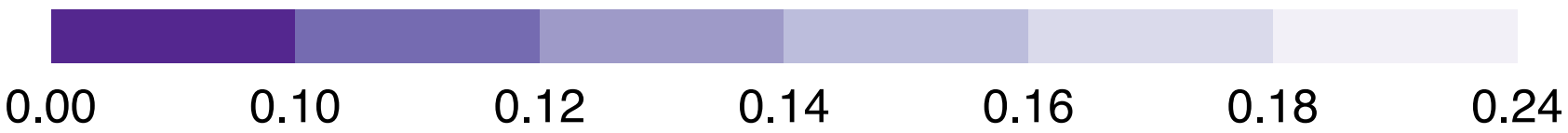


2020–2022

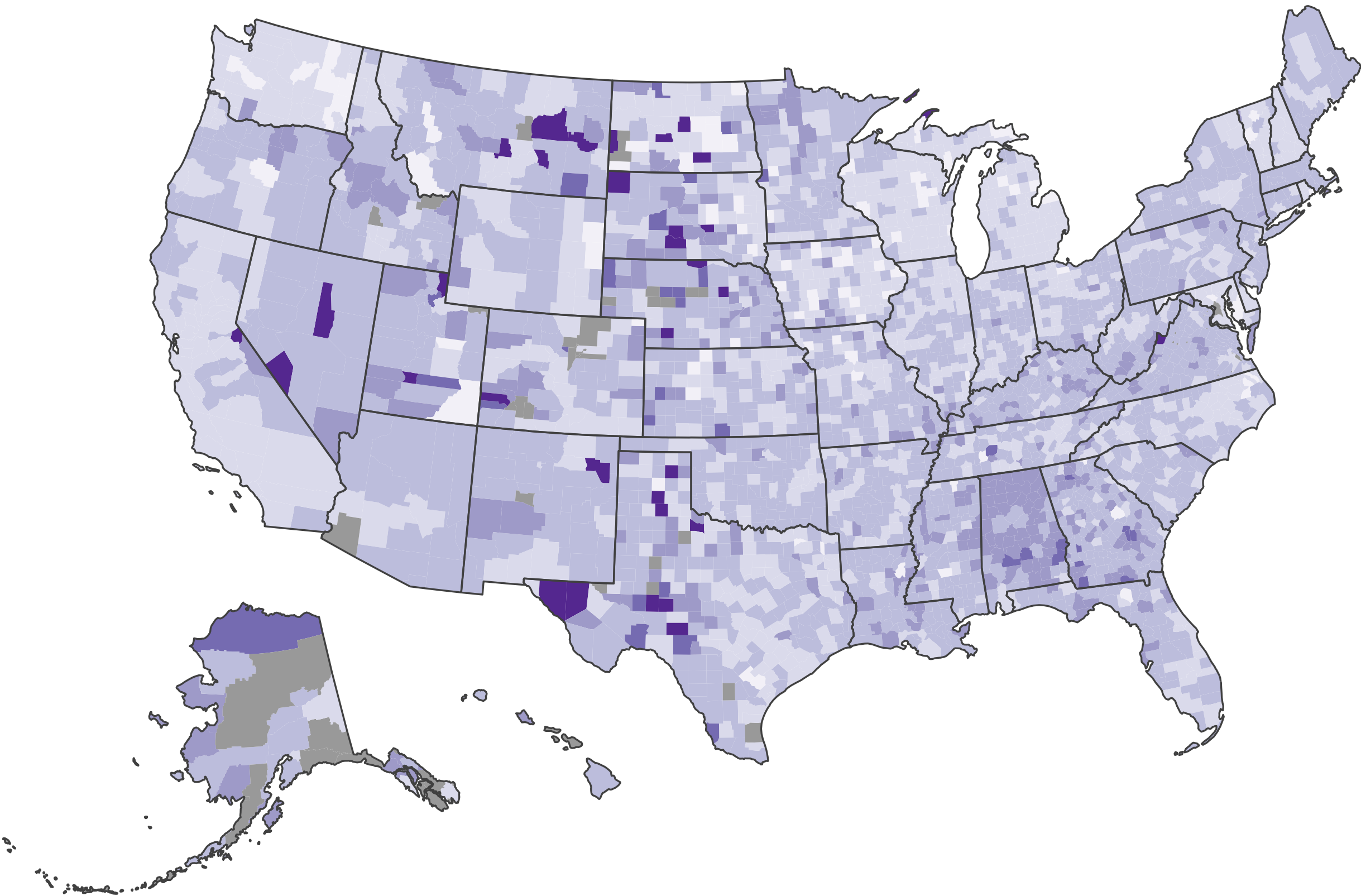


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.23 (p = 0.002)	+0.22 (p = 0.002)
Public health spending per capita	+0.24 (p = 0.004)	+0.12 (p = 0.088)
Death investigation type (ME)	+0.22 (p = 0.002)	+0.17 (p = 0.004)

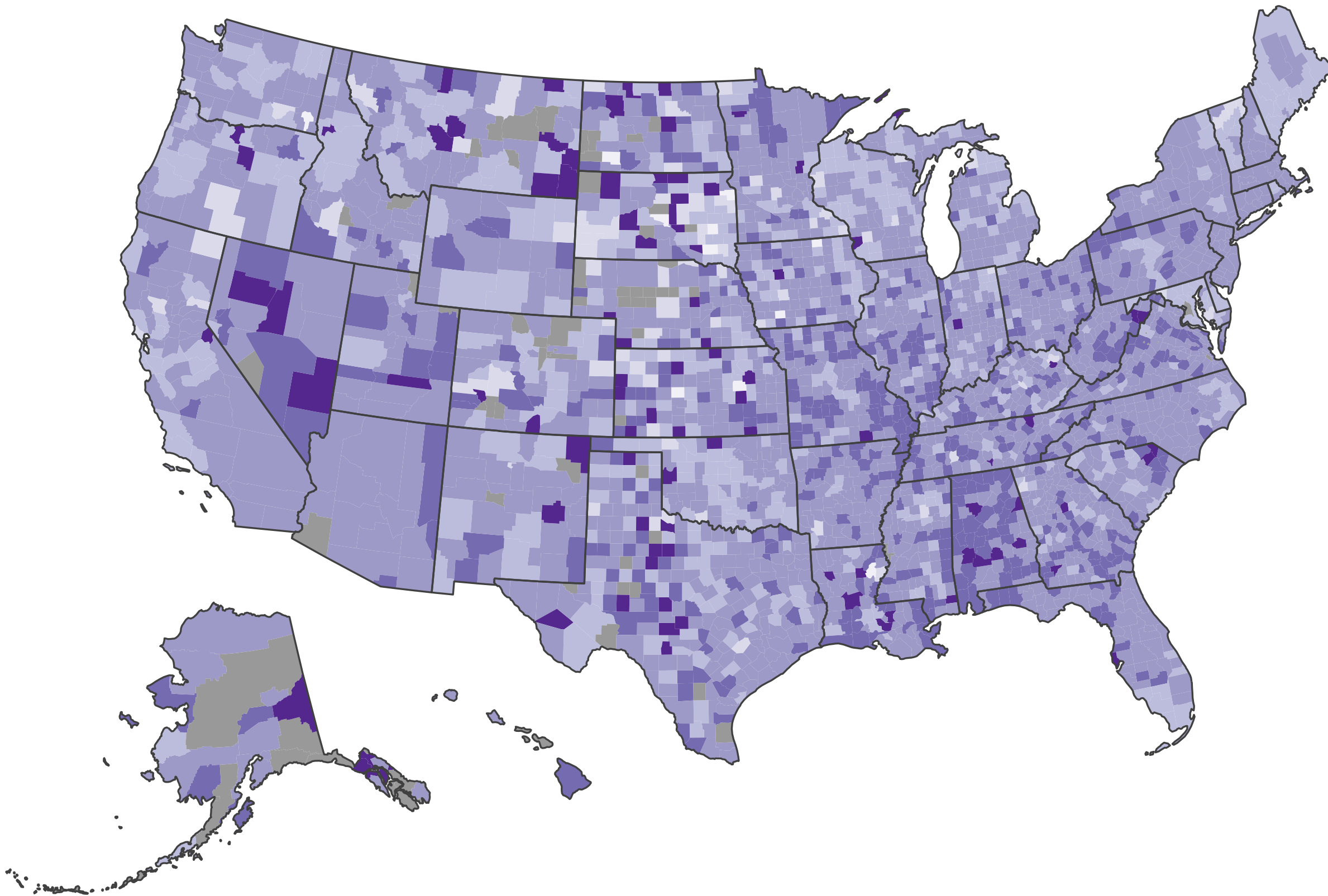
Re-assignability index



1999–2005

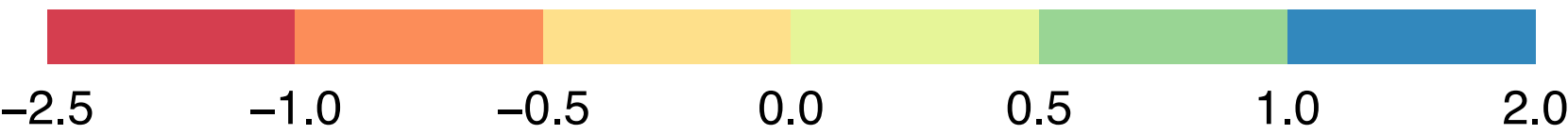


2020–2022

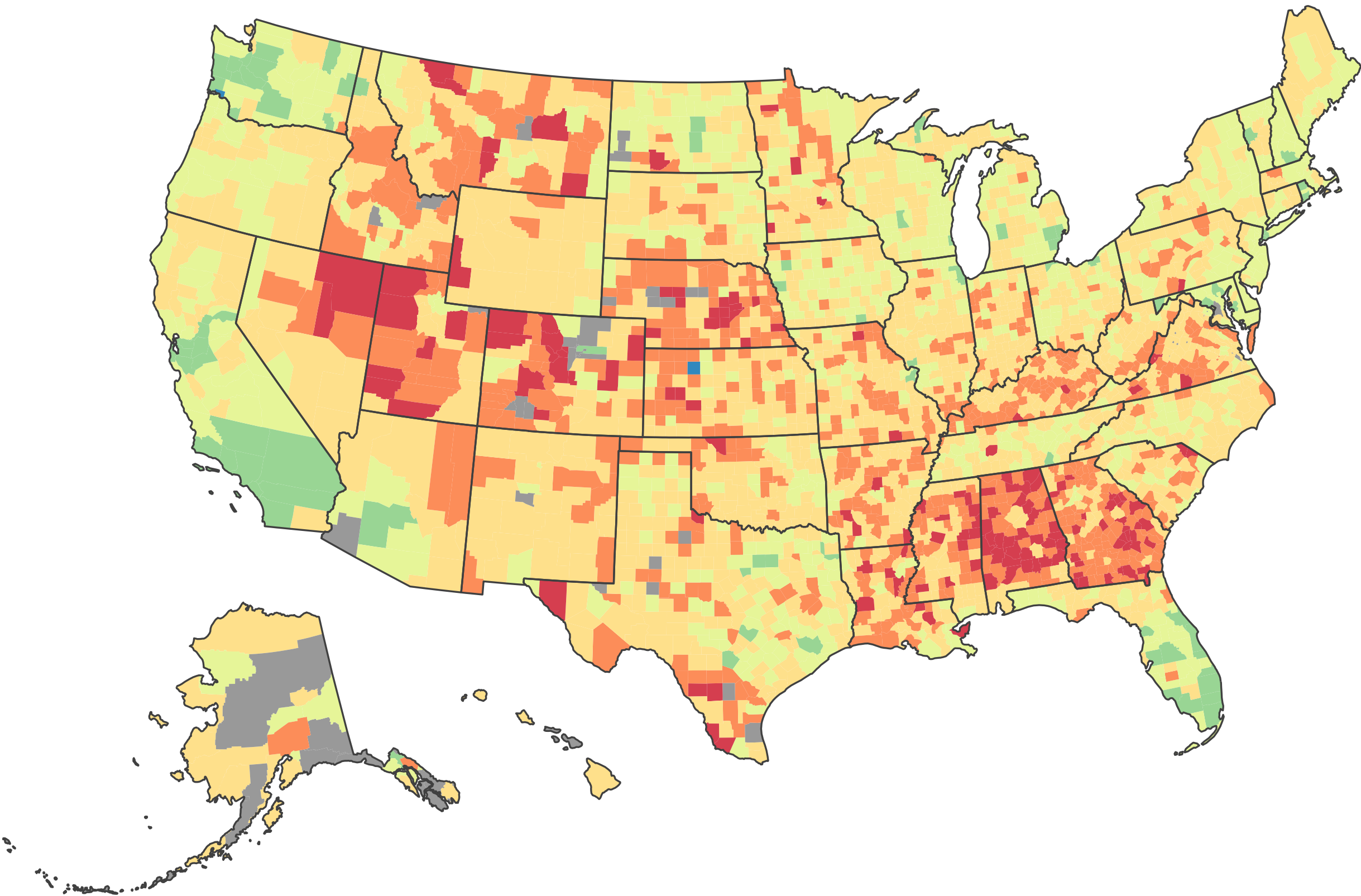


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.23 (p = 0.002)	+0.22 (p = 0.002)
Public health spending per capita	+0.24 (p = 0.004)	+0.12 (p = 0.088)
Death investigation type (ME)	+0.22 (p = 0.002)	+0.17 (p = 0.004)

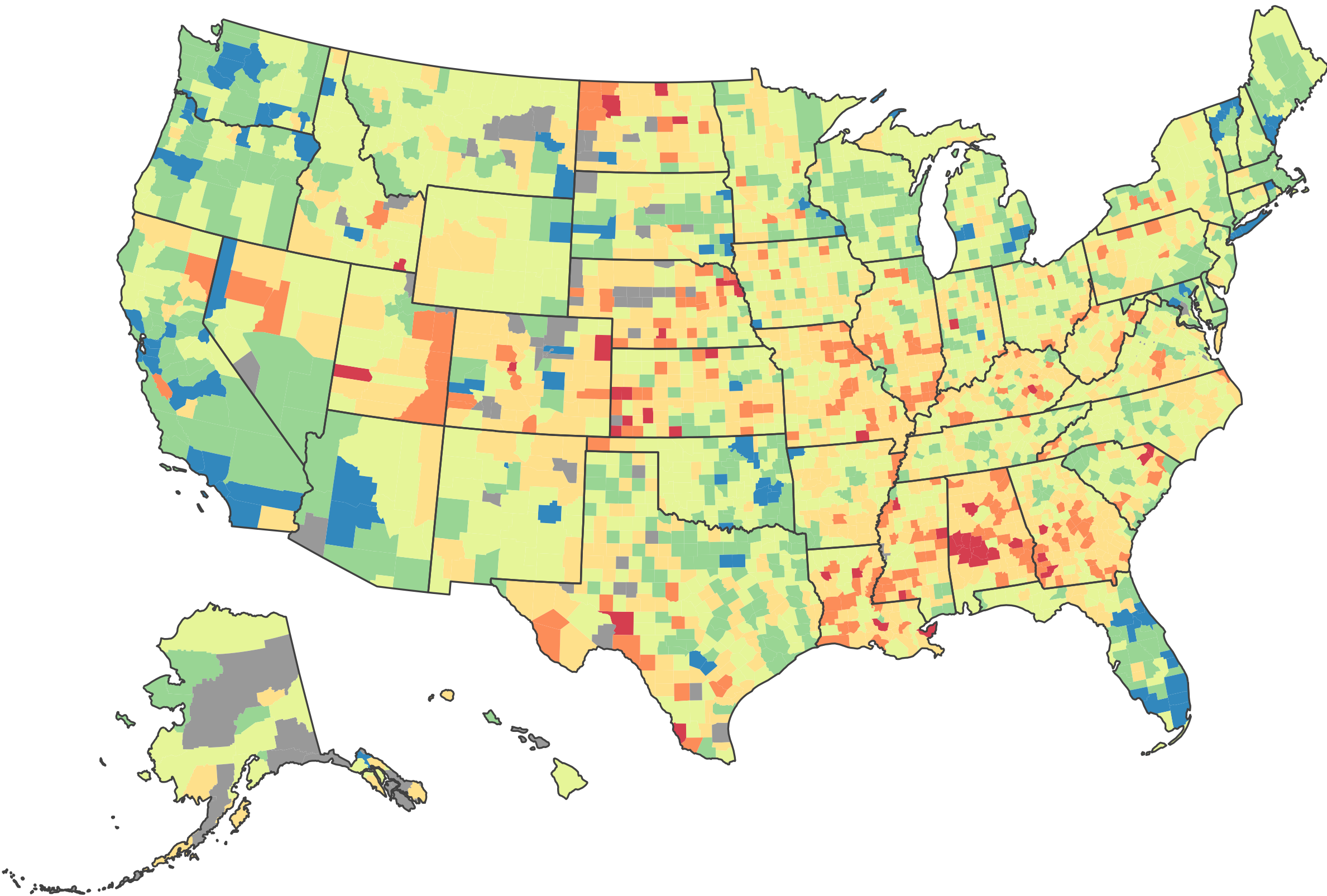
Aggregate index



1999–2005

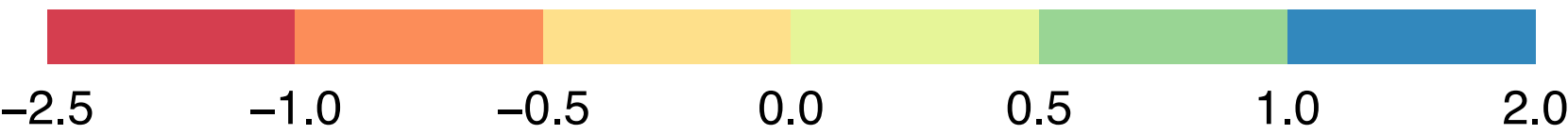


2020–2022

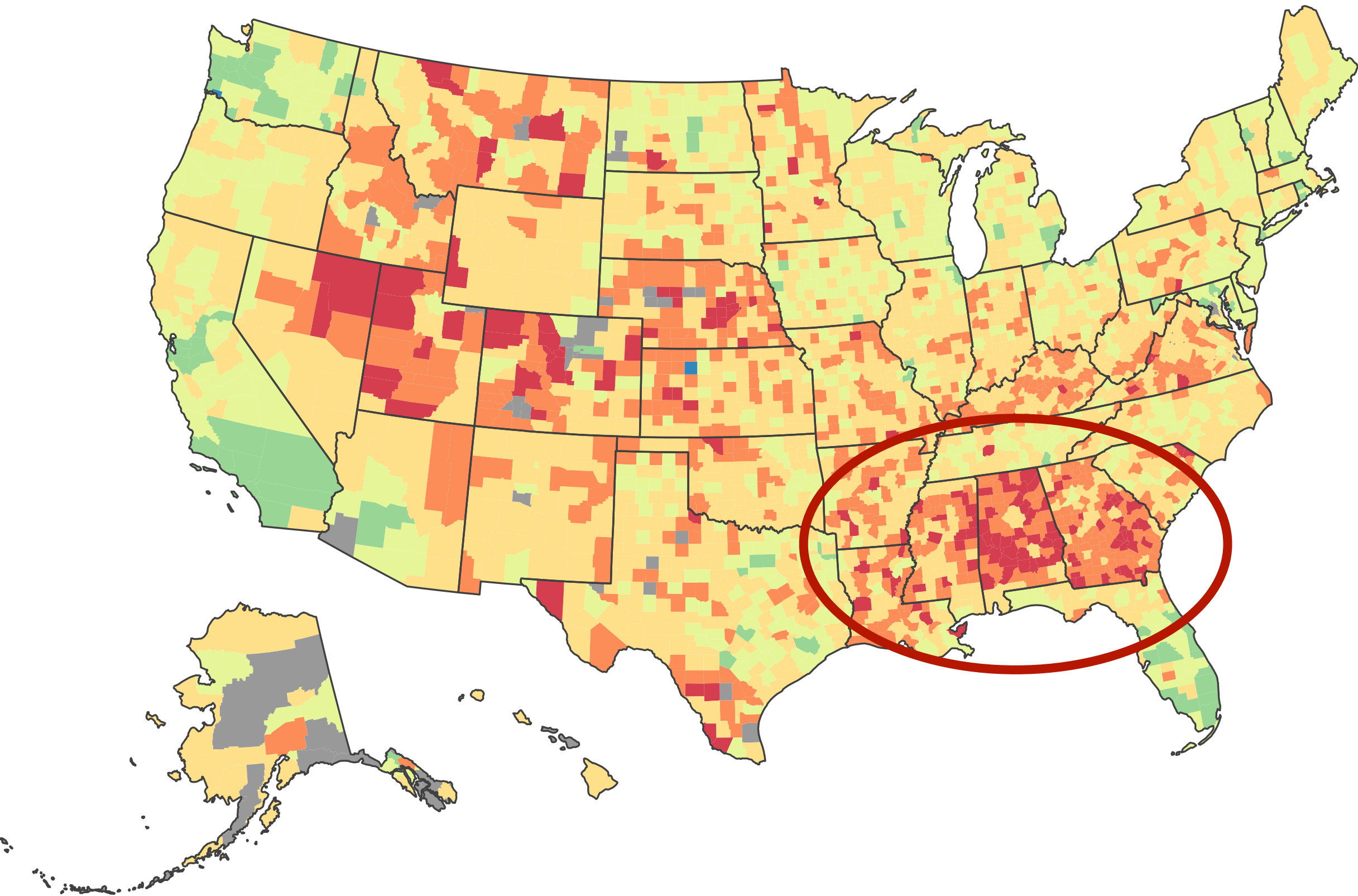


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.41 (p = 0.002)	+0.45 (p = 0.002)
Public health spending per capita	+0.31 (p = 0.002)	+0.25 (p = 0.006)
Death investigation type (ME)	+0.45 (p = 0.002)	+0.42 (p = 0.002)

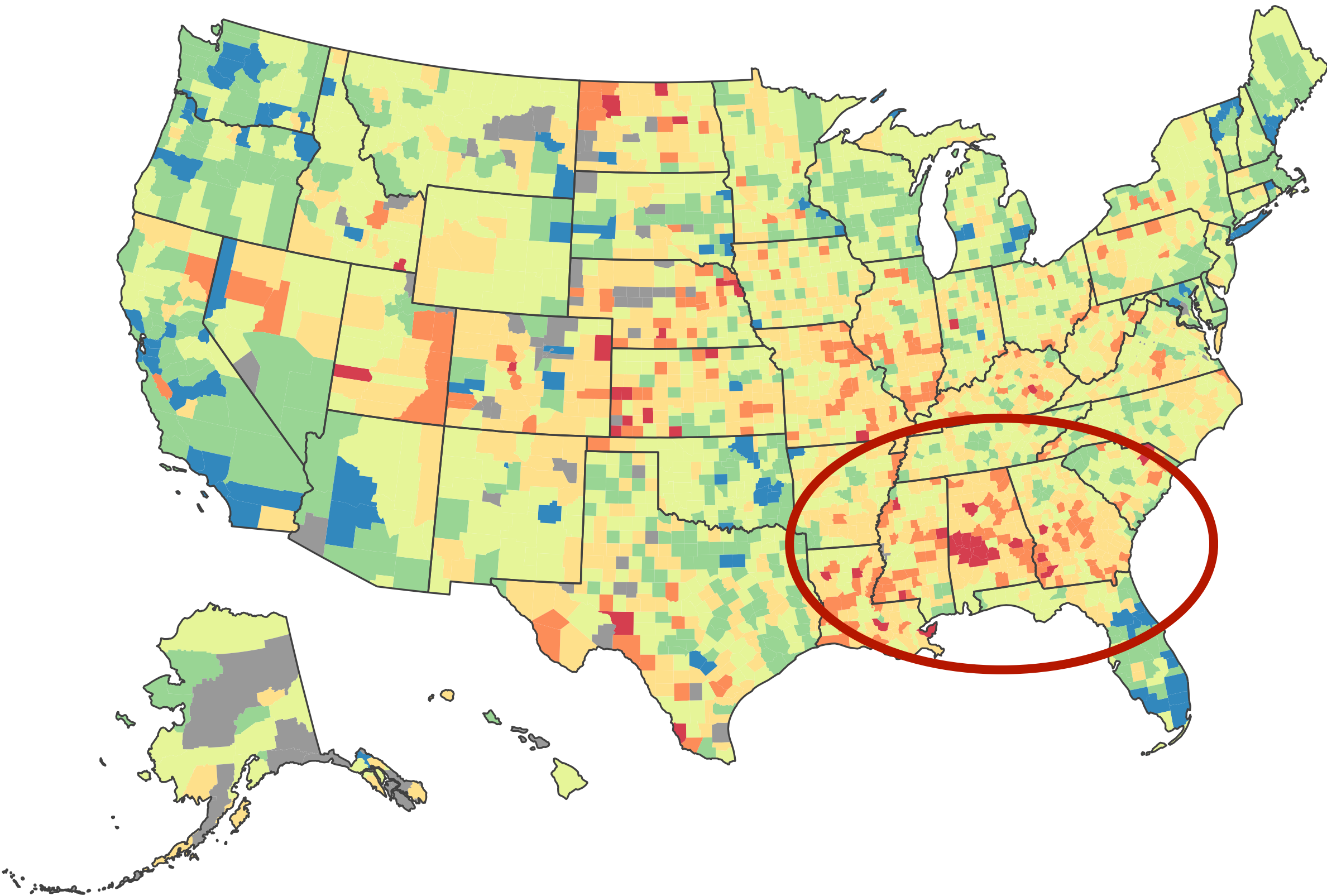
Aggregate index



1999–2005

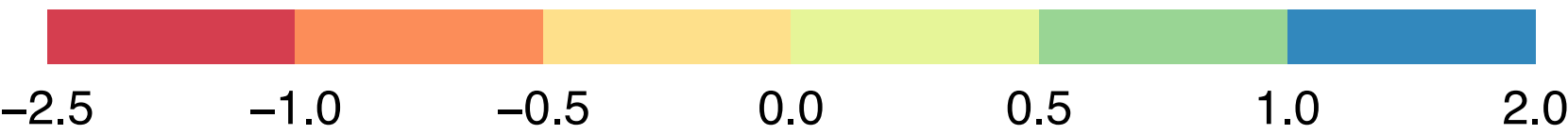


2020–2022

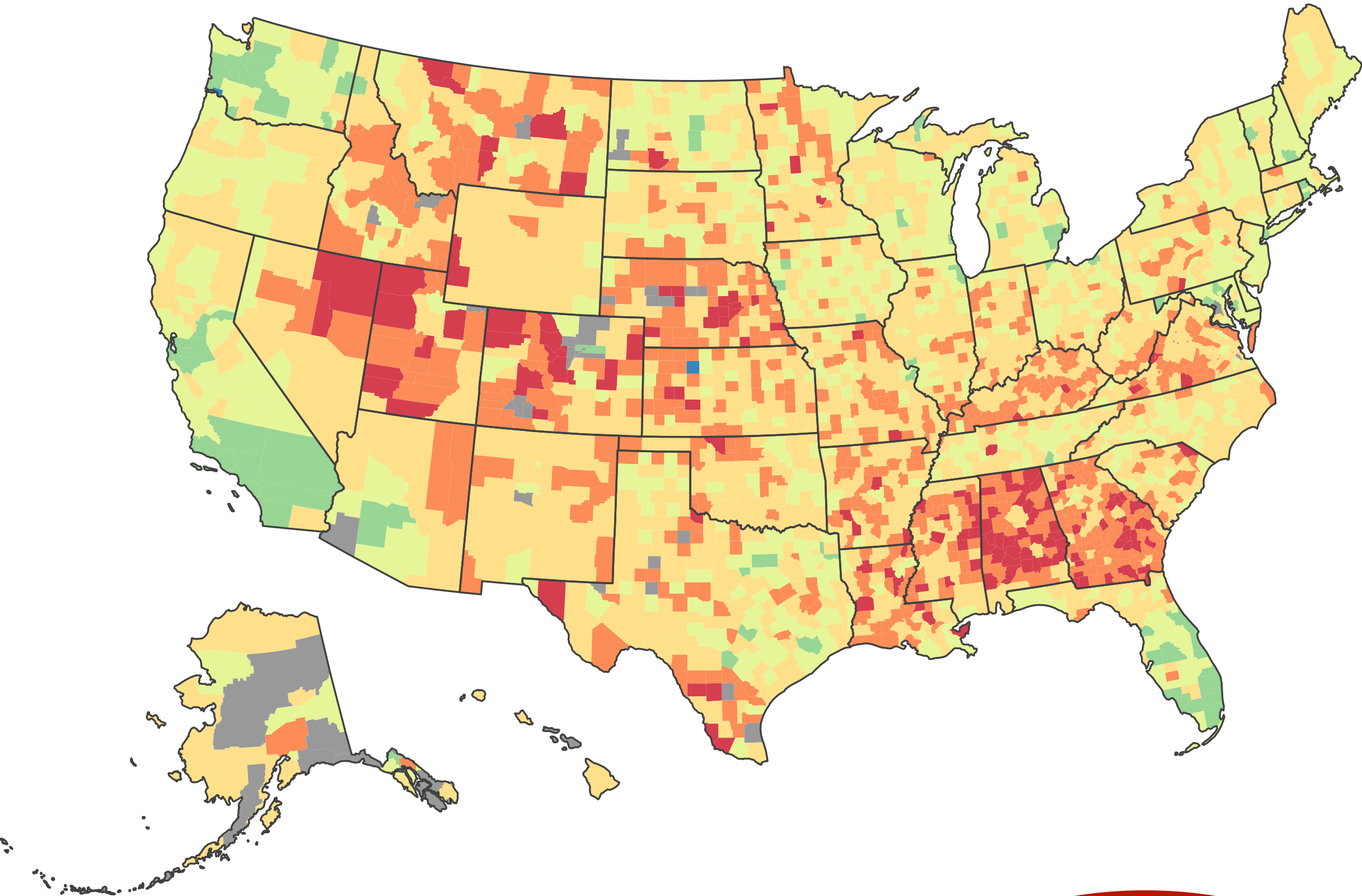


	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.41 ($p = 0.002$)	+0.45 ($p = 0.002$)
Public health spending per capita	+0.31 ($p = 0.002$)	+0.25 ($p = 0.006$)
Death investigation type (ME)	+0.45 ($p = 0.002$)	+0.42 ($p = 0.002$)

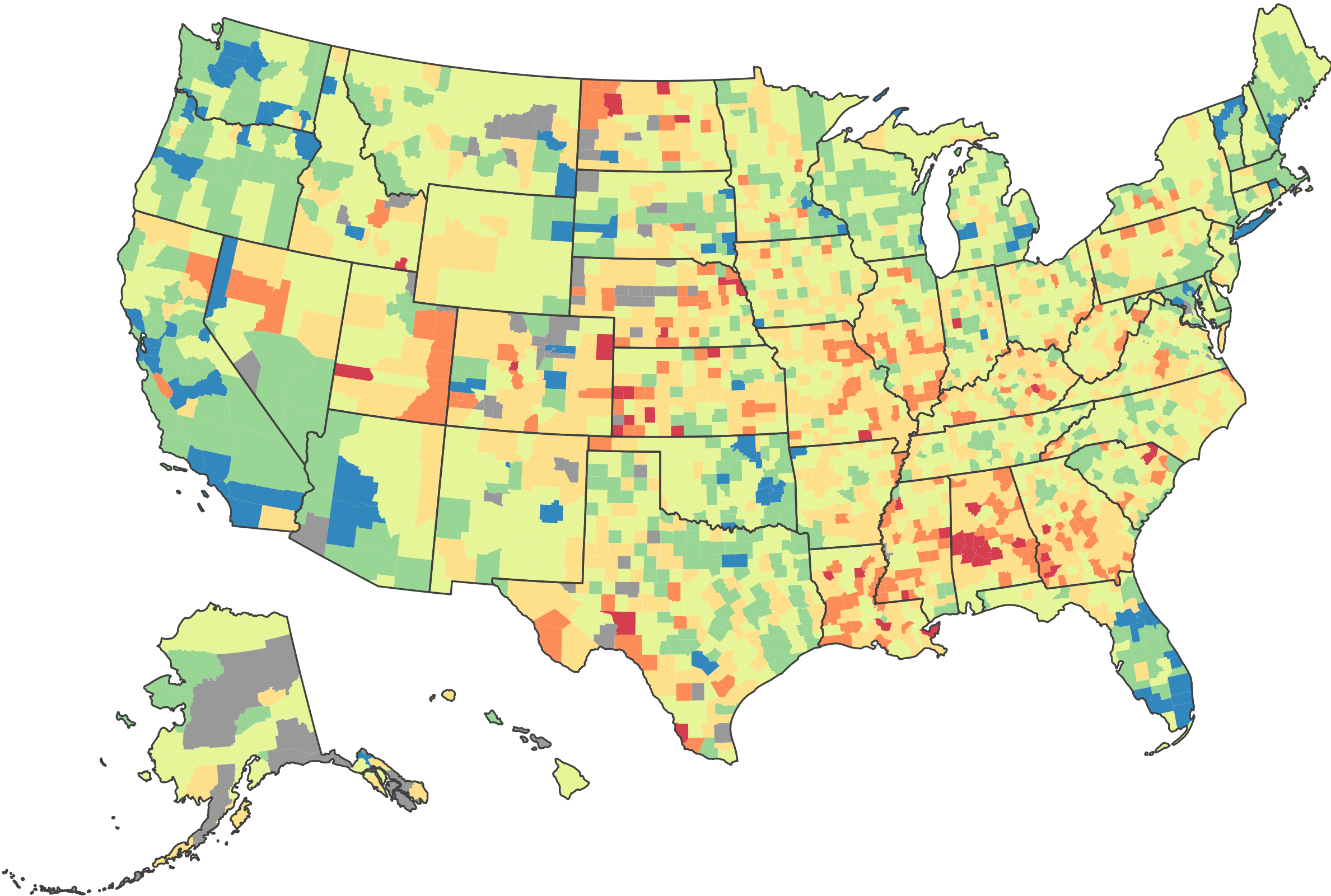
Aggregate index



1999–2005



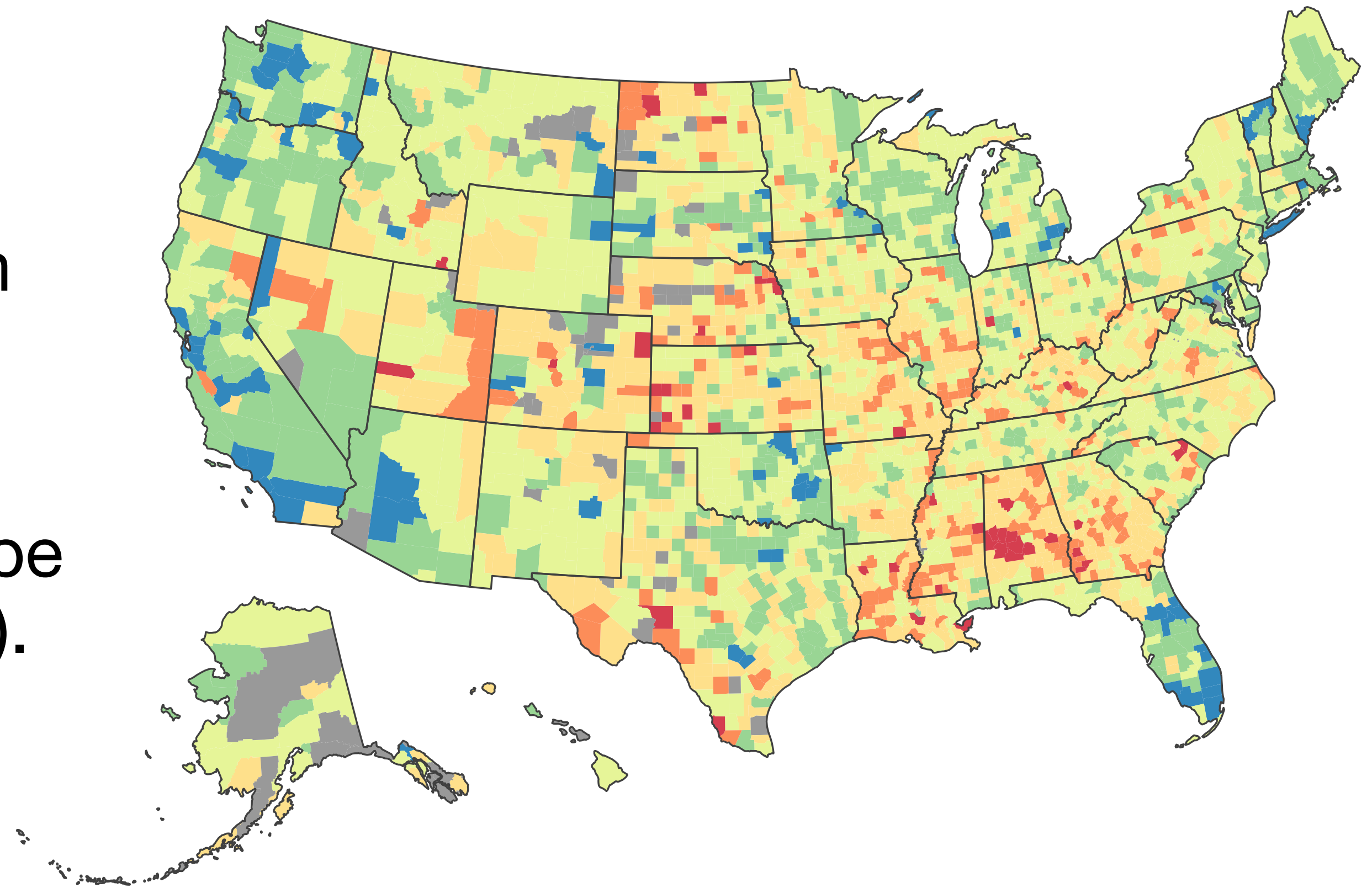
2020–2022



	1999–2005 ρ (MSR p-value)	2020–2022 ρ (MSR p-value)
Median household income	+0.41 (p = 0.002)	+0.45 (p = 0.002)
Public health spending per capita	+0.31 (p = 0.002)	+0.25 (p = 0.006)
Death investigation type (ME)	+0.45 (p = 0.002)	+0.42 (p = 0.002)

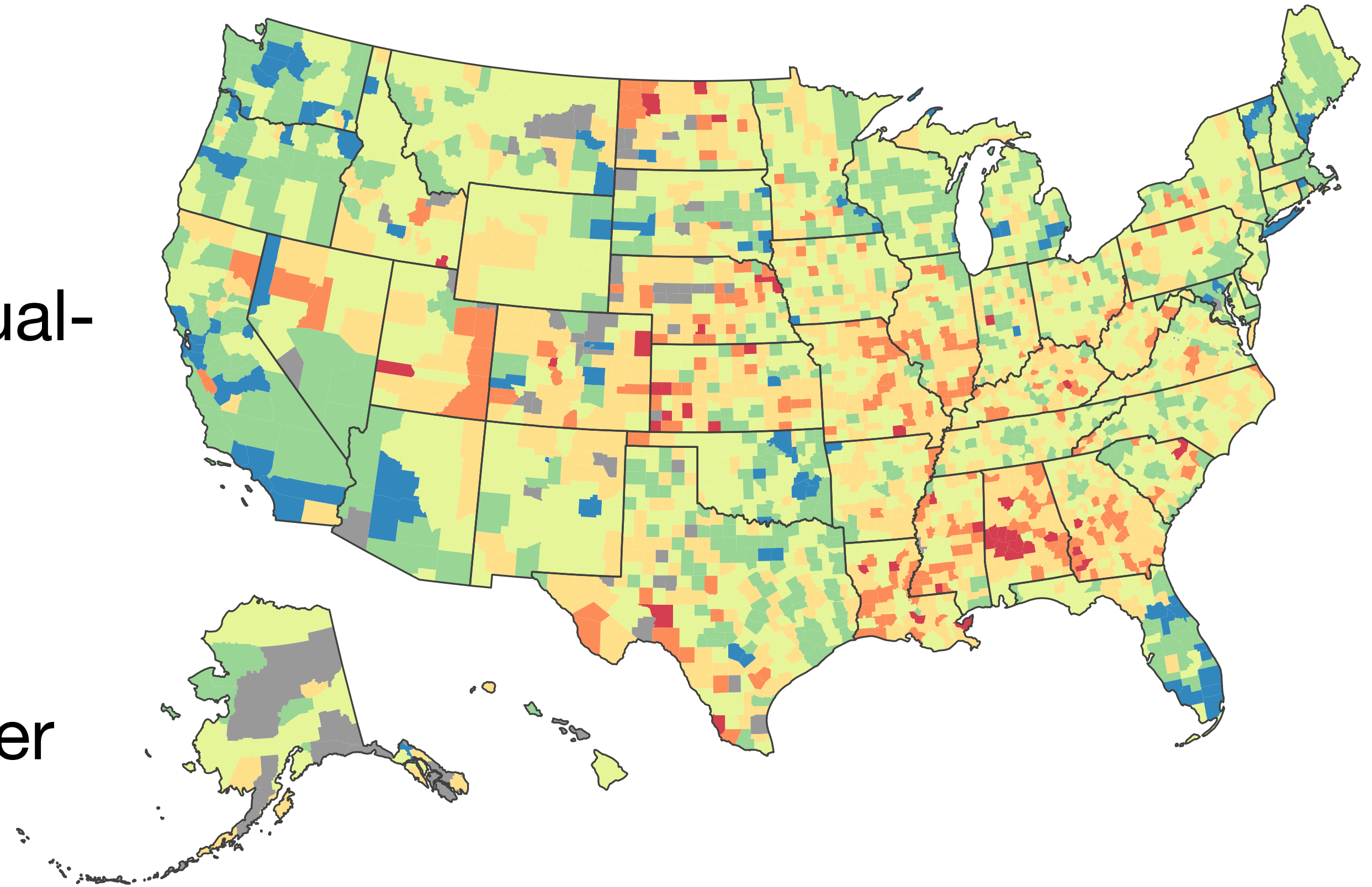
Conclusions

- Our new metrics seem to be capturing **different, complementary aspects** of death reporting quality
- Death reporting quality has **improved over time**
- A lot of **spatial variation** in quality with clear state effects
- **Variation is patterned** by both socioeconomic factors and reporter type (but maybe not public health spending).



Limitations

- We call it "quality" but this is more accurately described as "usability" since we don't know the true cause of death.
- Usability is not the same as accuracy.
- Many correlations are modest
- This is not a diagnostic tool for individual-level death certificates
- **Future work:** Understanding why higher level of detail counties have lower re-assignability index in later periods



Thank you

Monica Alexander, PhD
University of Toronto
monica.alexander@utoronto.ca

Amy Mann, BS
Oxford University
amy.mann@chch.ox.ac.uk

Mathew Kiang, ScD
Stanford University
mkiang@stanford.edu



UNIVERSITY OF
TORONTO



**DATA SCIENCES
INSTITUTE**

References

1. Phillips DE, Lozano R, Naghavi M, Atkinson C, Gonzalez-Medina D, Mikkelsen L, Murray CJL, Lopez AD. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. *Population Health Metrics*. 2014;12:14. doi:10.1186/1478-7954-12-14. PMID: 24982595.
2. Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bulletin of the World Health Organization*. 2005;83(3):171-177. PMID: 15798840.
3. Mikkelsen L, Moesgaard K, Hegnauer M, Lopez AD. ANACONDA: a new tool to improve mortality and cause of death data. *BMC Medicine*. 2020;18(1):61. doi:10.1186/s12916-020-01521-0. PMID: 32146907.
4. Ruhm CJ. Geographic variation in opioid and heroin involved drug poisoning mortality rates. *American Journal of Preventive Medicine*. 2017;53(6):745-753. doi:10.1016/j.amepre.2017.06.009. PMID: 28797652.
5. Boslett AJ, Denham A, Hill EL. Using contributing causes of death improves prediction of opioid involvement in unclassified drug overdoses in U.S. death records. *Addiction*. 2020;115(7):1308-1317. doi:10.1111/add.14943. PMID: 32106355. PMC: PMC9745785.
6. Rockett IRH, Hobbs GR, Wu D, Jia H, Nolte KB, Smith GS, Putnam SL, Caine ED. Variable classification of drug-intoxication suicides across US states: a partial artifact of forensics? *PLoS One*. 2015;10(8):e0135296. doi:10.1371/journal.pone.0135296. PMID: 26295155.
7. James BD, Leurgans SE, Hebert LE, Scherr PA, Yaffe K, Bennett DA. Contribution of Alzheimer disease to mortality in the United States. *Neurology*. 2014;82(12):1045-1050. doi:10.1212/WNL.0000000000000240. PMID: 24598707.
8. MacDorman MF, Declercq E, Cabral H, Morton C. Recent increases in the U.S. maternal mortality rate: disentangling trends from measurement issues. *Obstetrics and Gynecology*. 2016;128(3):447-455. doi:10.1097/AOG.0000000000001556. PMID: 27500333.
9. Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and excess mortality in the United States: a county-level analysis. *PLoS Medicine*. 2021;18(5):e1003571. doi:10.1371/journal.pmed.1003571. PMID: 34014945.
10. Lloyd-Jones DM, Martin DO, Larson MG, Levy D. Accuracy of death certificates for coding coronary heart disease as the cause of death. *Annals of Internal Medicine*. 1998;129(12):1020-1026. doi:10.7326/0003-4819-129-12-199812150-00005. PMID: 9867756.
11. Foreman KJ, Naghavi M, Ezzati M. Improving the usefulness of US mortality data: new methods for reclassification of underlying cause of death. *Popul Health Metr*. 2016;14(1):14. doi:10.1186/s12963-016-0082-6

Extra slides

US Standard Certificate of Death Form

NAME OF DECEDENT
For use by physician or institution

To Be Completed/ Verified By:
FUNERAL DIRECTOR:

LOCAL FILE NO.

U.S. STANDARD CERTIFICATE OF DEATH

STATE FILE NO.

1. DECEDENT'S LEGAL NAME (Include AKA's if any) (First, Middle, Last)				2. SEX		3. SOCIAL SECURITY NUMBER					
4a. AGE-Last Birthday (Years)		4b. UNDER 1 YEAR		4c. UNDER 1 DAY		5. DATE OF BIRTH (Mo/Day/Yr)		6. BIRTHPLACE (City and State or Foreign Country)			
		Months		Days		Hours		Minutes			
7a. RESIDENCE-STATE				7b. COUNTY				7c. CITY OR TOWN			
7d. STREET AND NUMBER				7e. APT. NO.		7f. ZIP CODE			7g. INSIDE CITY LIMITS? ☐ Yes ☐ No		
8. EVER IN US ARMED FORCES? ☐ Yes ☐ No		9. MARITAL STATUS AT TIME OF DEATH ☐ Married ☐ Married, but separated ☐ Widowed ☐ Divorced ☐ Never Married ☐ Unknown				10. SURVIVING SPOUSE'S NAME (If wife, give name prior to first marriage)					
11. FATHER'S NAME (First, Middle, Last)						12. MOTHER'S NAME PRIOR TO FIRST MARRIAGE (First, Middle, Last)					
13a. INFORMANT'S NAME			13b. RELATIONSHIP TO DECEDENT				13c. MAILING ADDRESS (Street and Number, City, State, Zip Code)				
14. PLACE OF DEATH (Check only one: see instructions)											
IF DEATH OCCURRED IN A HOSPITAL: ☐ Inpatient ☐ Emergency Room/Outpatient ☐ Dead on Arrival						IF DEATH OCCURRED SOMEWHERE OTHER THAN A HOSPITAL: ☐ Hospice facility ☐ Nursing home/Long term care facility ☐ Decedent's home ☐ Other (Specify):					
15. FACILITY NAME (If not institution, give street & number)				16. CITY OR TOWN , STATE, AND ZIP CODE				17. COUNTY OF DEATH			
18. METHOD OF DISPOSITION: ☐ Burial ☐ Cremation ☐ Donation ☐ Entombment ☐ Removal from State ☐ Other (Specify):				19. PLACE OF DISPOSITION (Name of cemetery, crematory, other place)							
20. LOCATION-CITY, TOWN, AND STATE				21. NAME AND COMPLETE ADDRESS OF FUNERAL FACILITY							
22. SIGNATURE OF FUNERAL SERVICE LICENSEE OR OTHER AGENT								23. LICENSE NUMBER (Of Licensee)			
ITEMS 24-28 MUST BE COMPLETED BY PERSON WHO PRONOUNCES OR CERTIFIES DEATH						24. DATE PRONOUNCED DEAD (Mo/Day/Yr)			25. TIME PRONOUNCED DEAD		
26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable)						27. LICENSE NUMBER			28. DATE SIGNED (Mo/Day/Yr)		
29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month)				30. ACTUAL OR PRESUMED TIME OF DEATH			31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? ☐ Yes ☐ No				
32. PART I. Enter the <u>chain of events</u> --diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary. IMMEDIATE CAUSE (Final disease or condition -----> resulting in death) a. _____ Due to (or as a consequence of): Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST c. _____ Due to (or as a consequence of): d. _____									Approximate interval: Onset to death _____ _____ _____ _____		
33. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No								34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ☐ Yes ☐ No			
35. DID TOBACCO USE CONTRIBUTE TO DEATH?				36. IF FEMALE: ☐ Not pregnant within past year				37. MANNER OF DEATH			

Cause of Death Section

NAME OF DECEDENT
For use by physician or institution

To Be Completed/ Verified By:
FUNERAL DIRECTOR:

LOCAL FILE NO.

U.S. STANDARD CERTIFICATE OF DEATH

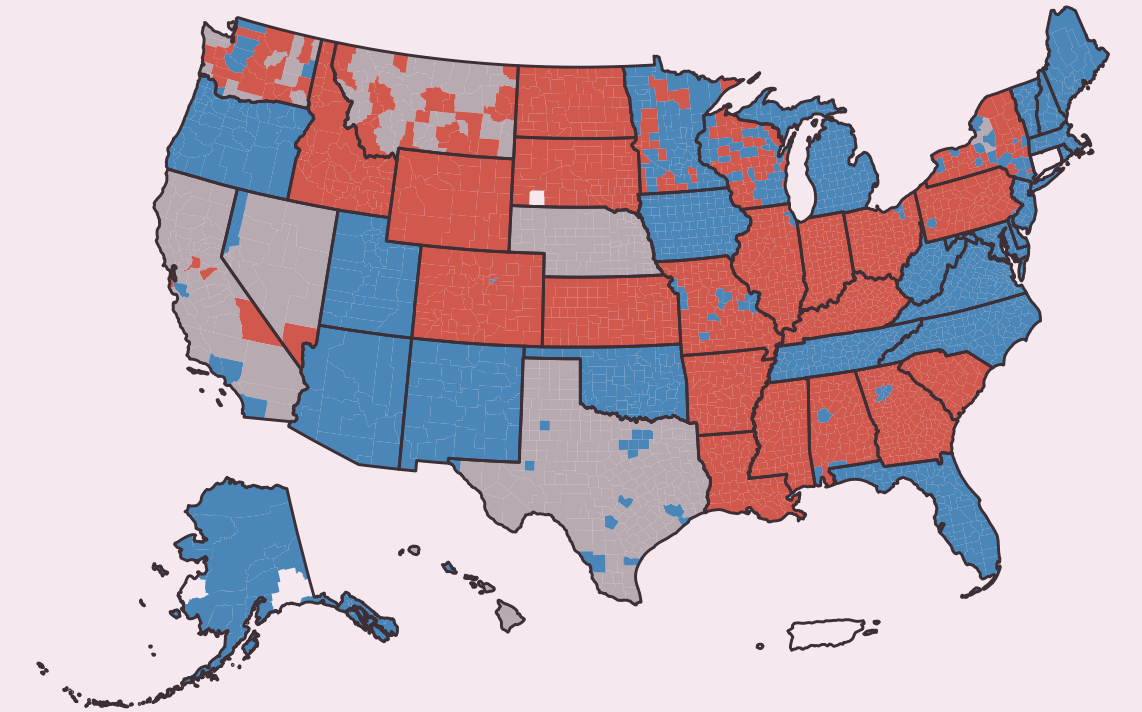
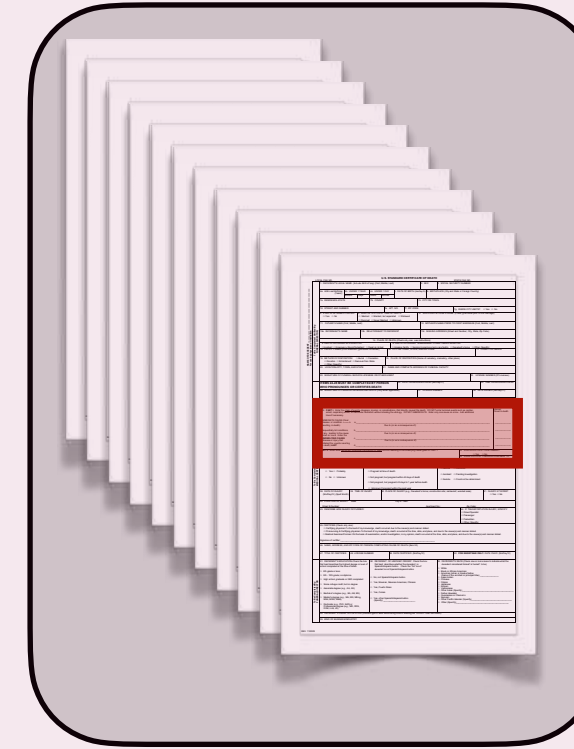
STATE FILE NO.

1. DECEDENT'S LEGAL NAME (Include AKA's if any) (First, Middle, Last)				2. SEX		3. SOCIAL SECURITY NUMBER							
4a. AGE-Last Birthday (Years)		4b. UNDER 1 YEAR		4c. UNDER 1 DAY		5. DATE OF BIRTH (Mo/Day/Yr)		6. BIRTHPLACE (City and State or Foreign Country)					
		Months		Days		Hours		Minutes					
7a. RESIDENCE-STATE				7b. COUNTY				7c. CITY OR TOWN					
7d. STREET AND NUMBER				7e. APT. NO.		7f. ZIP CODE				7g. INSIDE CITY LIMITS? ☐ Yes ☐ No			
8. EVER IN US ARMED FORCES? ☐ Yes ☐ No				9. MARITAL STATUS AT TIME OF DEATH ☐ Married ☐ Married, but separated ☐ Widowed ☐ Divorced ☐ Never Married ☐ Unknown				10. SURVIVING SPOUSE'S NAME (If wife, give name prior to first marriage)					
11. FATHER'S NAME (First, Middle, Last)						12. MOTHER'S NAME PRIOR TO FIRST MARRIAGE (First, Middle, Last)							
13a. INFORMANT'S NAME				13b. RELATIONSHIP TO DECEDENT				13c. MAILING ADDRESS (Street and Number, City, State, Zip Code)					
14. PLACE OF DEATH (Check only one: see instructions)													
IF DEATH OCCURRED IN A HOSPITAL: ☐ Inpatient ☐ Emergency Room/Outpatient ☐ Dead on Arrival						IF DEATH OCCURRED SOMEWHERE OTHER THAN A HOSPITAL: ☐ Hospice facility ☐ Nursing home/Long term care facility ☐ Decedent's home ☐ Other (Specify):							
15. FACILITY NAME (If not institution, give street & number)						16. CITY OR TOWN , STATE, AND ZIP CODE				17. COUNTY OF DEATH			
18. METHOD OF DISPOSITION: ☐ Burial ☐ Cremation ☐ Donation ☐ Entombment ☐ Removal from State ☐ Other (Specify):						19. PLACE OF DISPOSITION (Name of cemetery, crematory, other place)							
20. LOCATION-CITY, TOWN, AND STATE						21. NAME AND COMPLETE ADDRESS OF FUNERAL FACILITY							
22. SIGNATURE OF FUNERAL SERVICE LICENSEE OR OTHER AGENT								23. LICENSE NUMBER (Of Licensee)					
ITEMS 24-28 MUST BE COMPLETED BY PERSON WHO PRONOUNCES OR CERTIFIES DEATH						24. DATE PRONOUNCED DEAD (Mo/Day/Yr)				25. TIME PRONOUNCED DEAD			
26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable)						27. LICENSE NUMBER				28. DATE SIGNED (Mo/Day/Yr)			
29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month)						30. ACTUAL OR PRESUMED TIME OF DEATH				31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? ☐ Yes ☐ No			
32. PART I. Enter the <u>chain of events</u> --diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary. IMMEDIATE CAUSE (Final disease or condition -----> resulting in death) a. _____ Due to (or as a consequence of): _____ Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST b. _____ Due to (or as a consequence of): _____ c. _____ Due to (or as a consequence of): _____ d. _____ PART II. Enter other <u>significant conditions contributing to death</u> but not resulting in the underlying cause given in PART I												Approximate interval: Onset to death _____ _____ _____ _____	
33. WAS AN AUTOPSY PERFORMED? ☐ Yes ☐ No								34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? ☐ Yes ☐ No					
35. DID TOBACCO USE CONTRIBUTE TO DEATH?				36. IF FEMALE: ☐ Not pregnant within past year				37. MANNER OF DEATH					

All variation comes from death *reporting*

- **Death coding** is deterministic: Same free-text input leads to same ICD-10 code. Same set of ICD-10 codes leads to same underlying cause of death.
- Therefore, all variation in death data quality comes from **death reporting**.
- 52 reporting registries (50 states + DC + New York City)
- Medicolegal death investigation system is organized at county-level
 - 50% of counties use coroners, 36% use medical examiners, 14% use other system

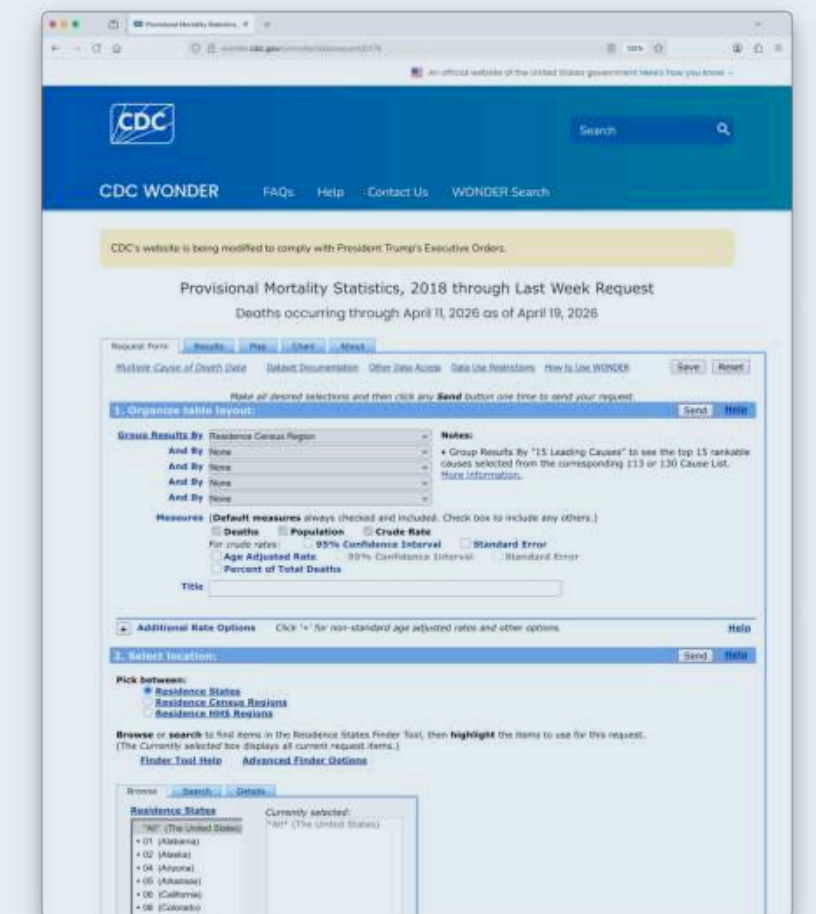
Death reporting



Death coding



Data release

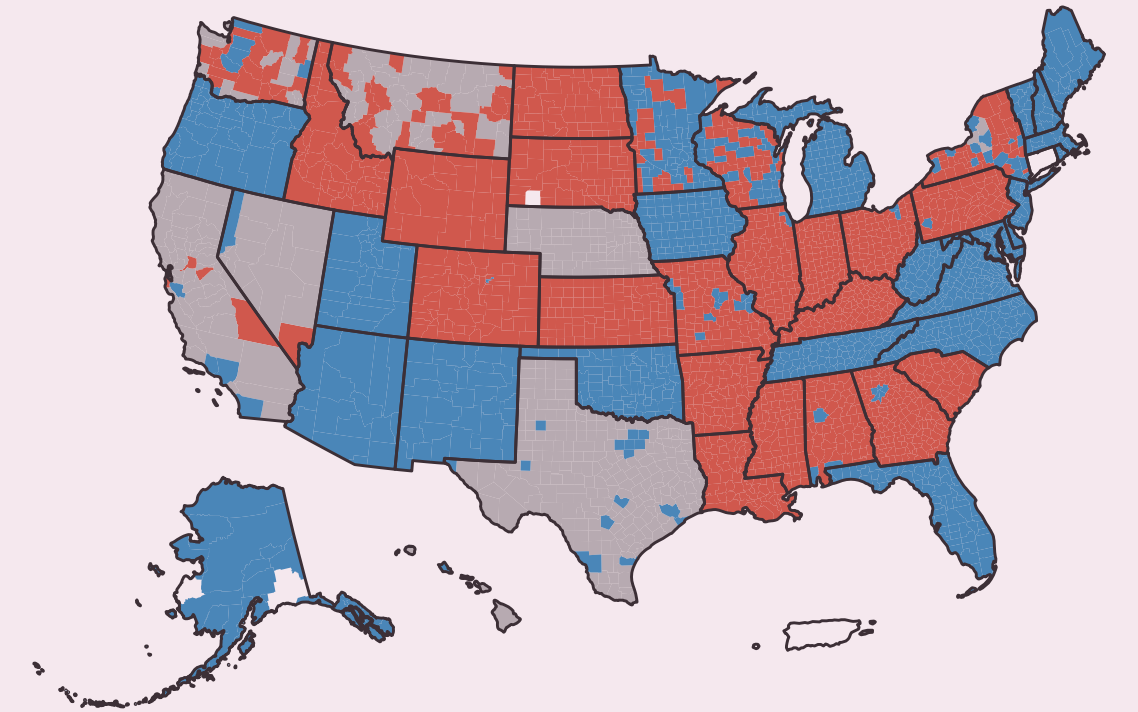
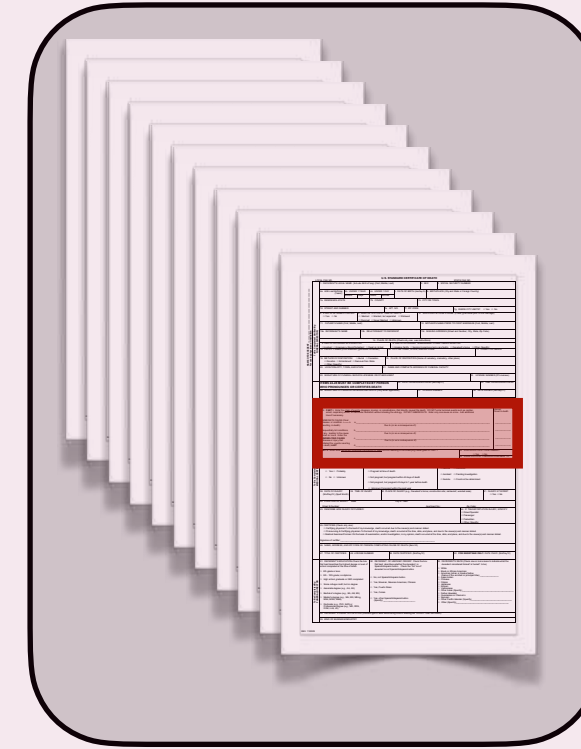


Monitoring quality across time and space

Develop measures of death reporting quality that are suitable for the context of the United States

1. **Scalable monitoring for vital statistics offices.** Track how counties perform over time, resource allocation, etc.

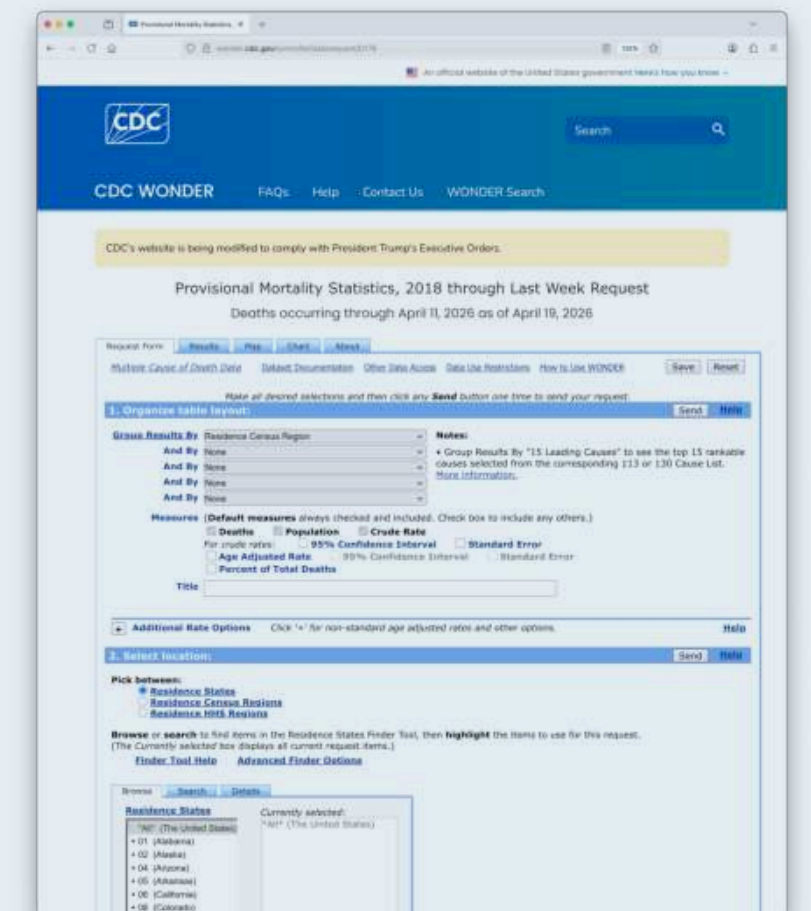
Death reporting



Death coding



Data release

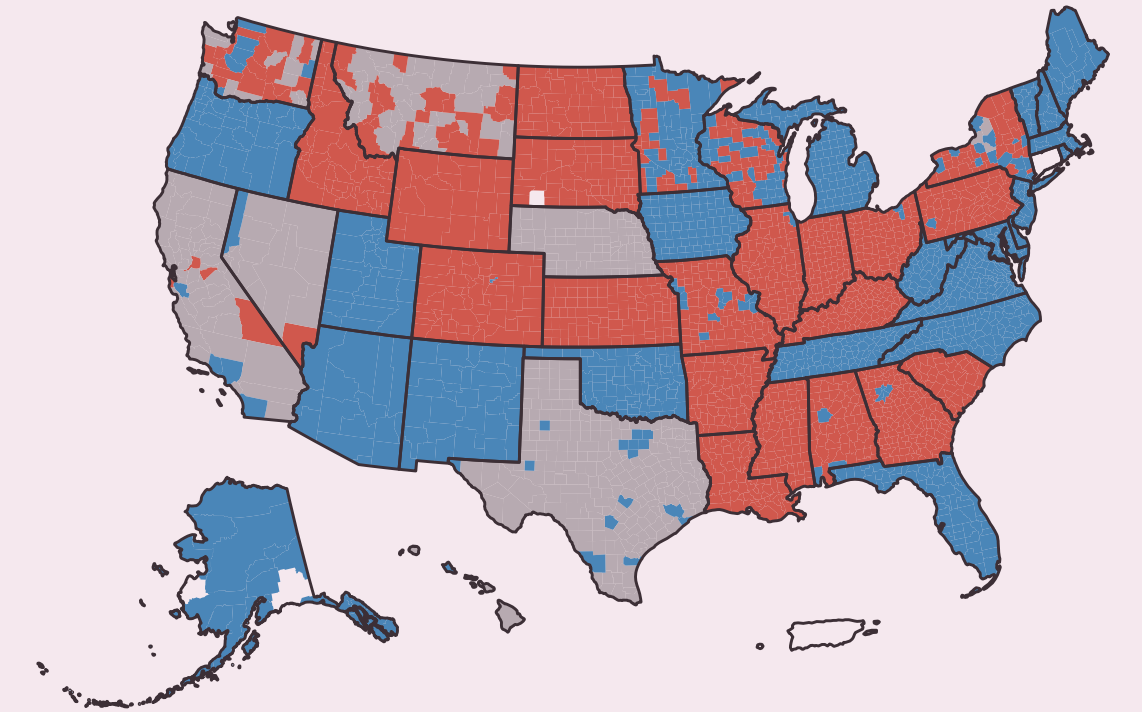
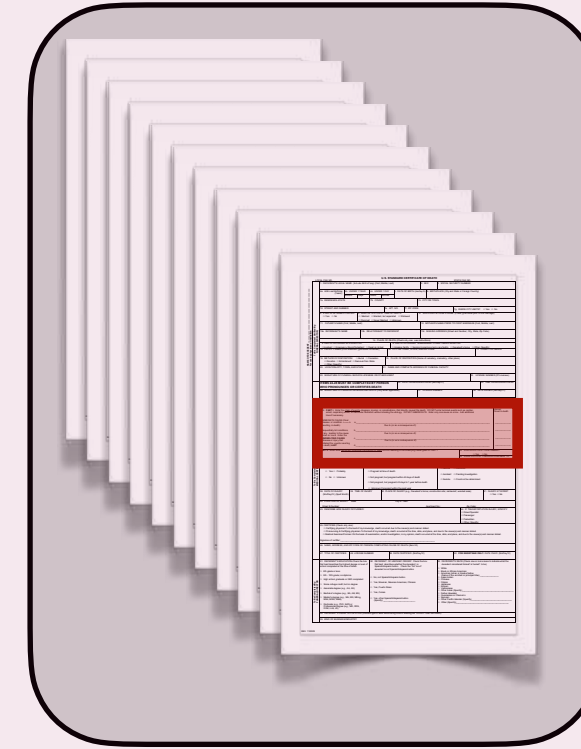


Identifying high-quality research data

Develop measures of death reporting quality that are suitable for the context of the United States

1. **Scalable monitoring for vital statistics offices.** Track how counties perform over time, resource allocation, etc.
2. **Data quality flag for research.** Researchers want to know which death data are most accurate or reliable for their research question.

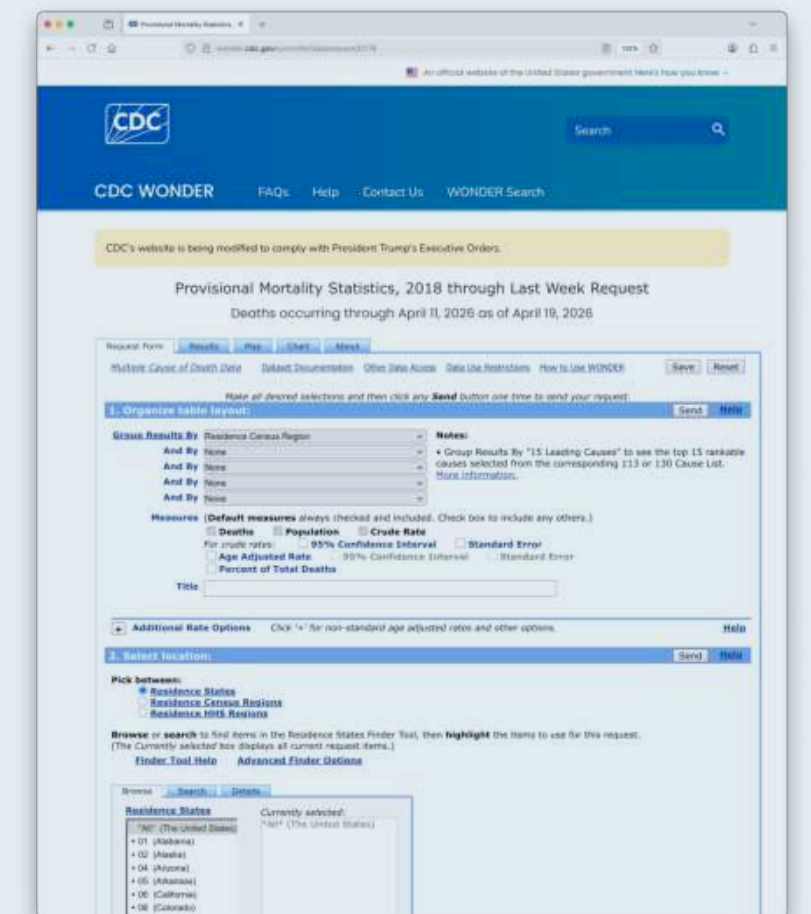
Death reporting



Death coding



Data release



Step 1 – Indirect standardization

- **Problem:** We want to capture variation in codes driven by *reporting differences* rather than by differences in the (true) distribution of causes of death.

Step 1 – Indirect standardization

- **Problem:** We want to capture variation in codes driven by *reporting differences* rather than by differences in the (true) distribution of causes of death.
- **Solution:** Indirect standardization to separate *how certifiers report* from *what people die of*. Forces counties to assume the national cause-mix. Implicitly controls for age-, sex-, and other factors that change the proportion of deaths across large death categories.

$$p_{k,t}^*(d) = \sum_{c \in \mathcal{C}} s_t(c) w_{k,t}(d \mid c)$$

Step 1 – Indirect standardization

- **Problem:** We want to capture variation in codes driven by *reporting differences* rather than by differences in the (true) distribution of causes of death.
- **Solution:** Indirect standardization to separate *how certifiers report* from *what people die of*. Forces counties to assume the national cause-mix. Implicitly controls for age-, sex-, and other factors that change the proportion of deaths across large death categories.

$$p_{k,t}^*(d) = \sum_{c \in \mathcal{C}} s_t(c) w_{k,t}(d \mid c)$$

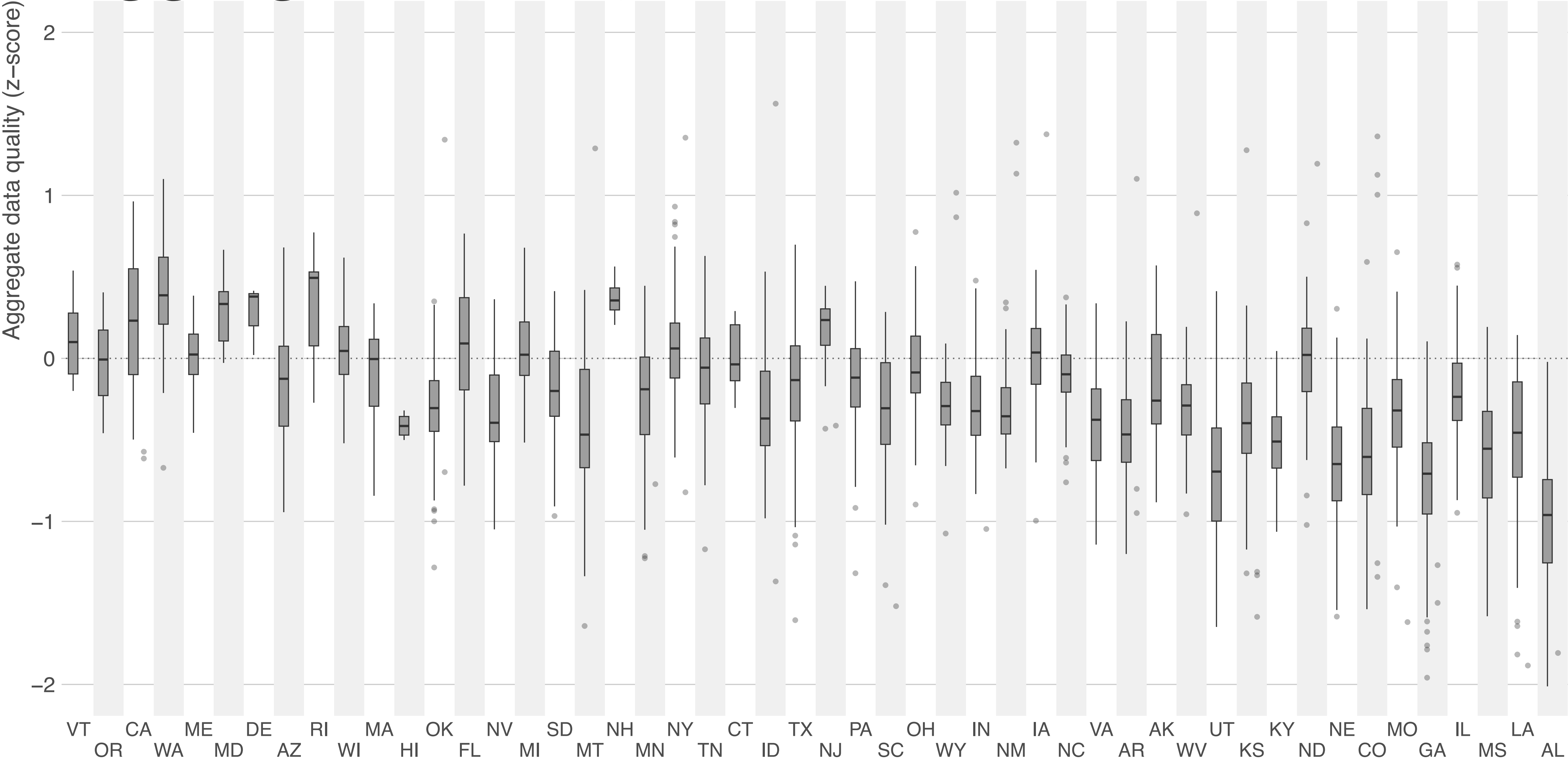
- **Intuition:** *Within* heart disease deaths, how variable are the ICD-10 codes?

Rescaling Level of Detail

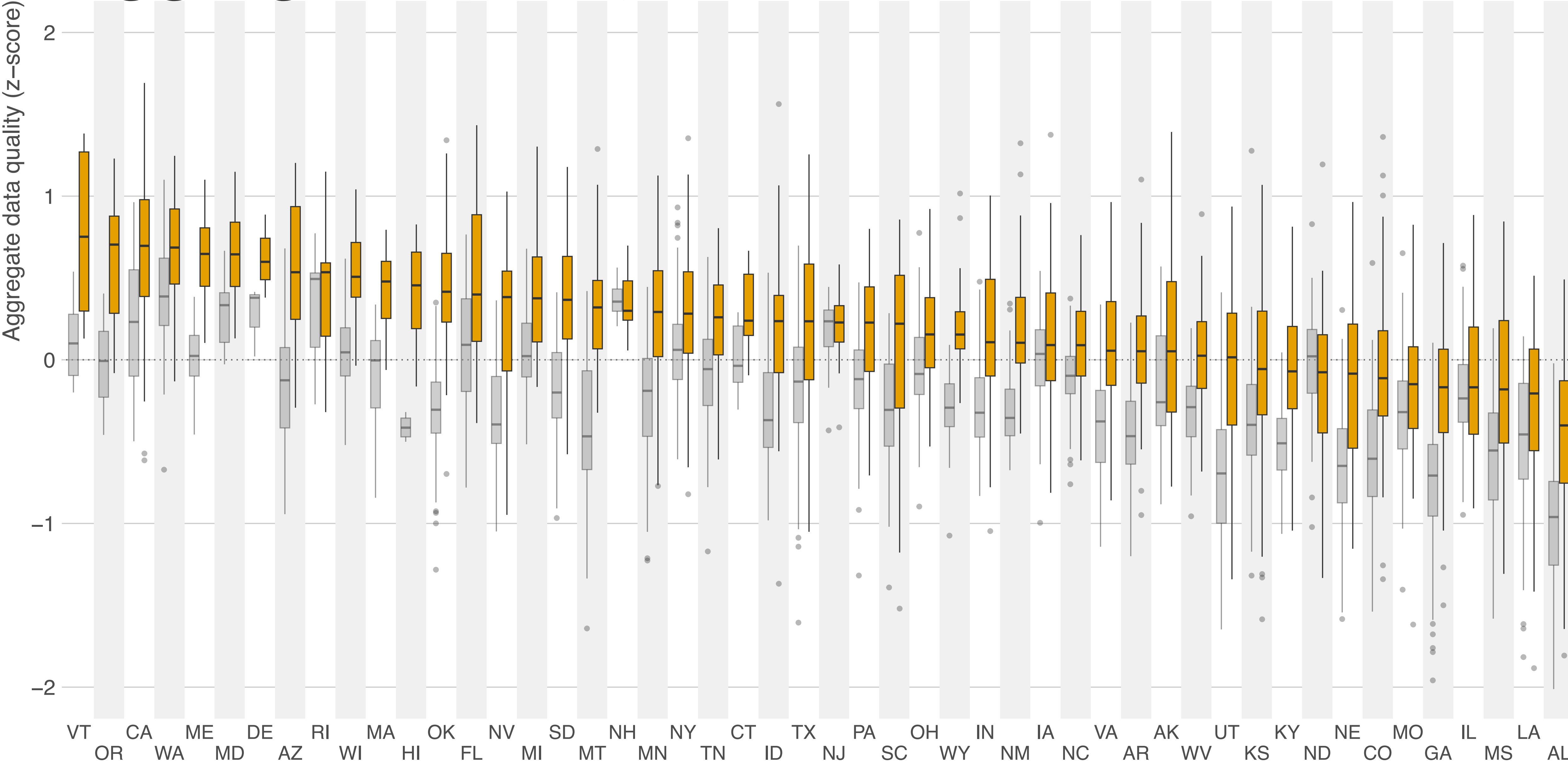
- We rescale level of detail by taking the theoretical maximum entropy across all cause of death codes D observed in time period t and multiply it by 100.

$$\text{Detail}_{k,t} = 100 \times \frac{H_{k,t}}{H_t^{\max}}, \quad H_t^{\max} = \log |\mathcal{D}_t|$$

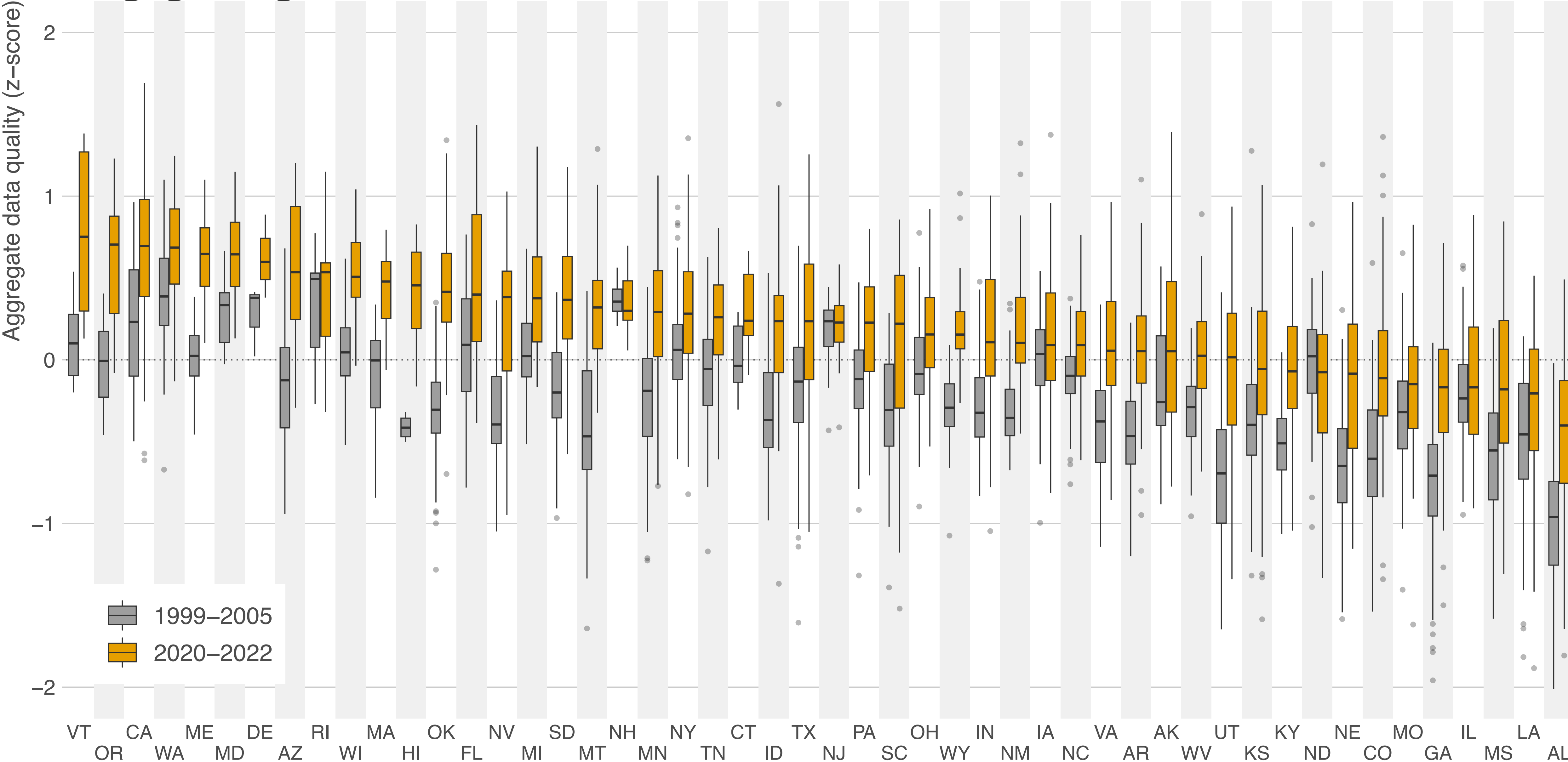
Aggregate quality index, 1999-2005



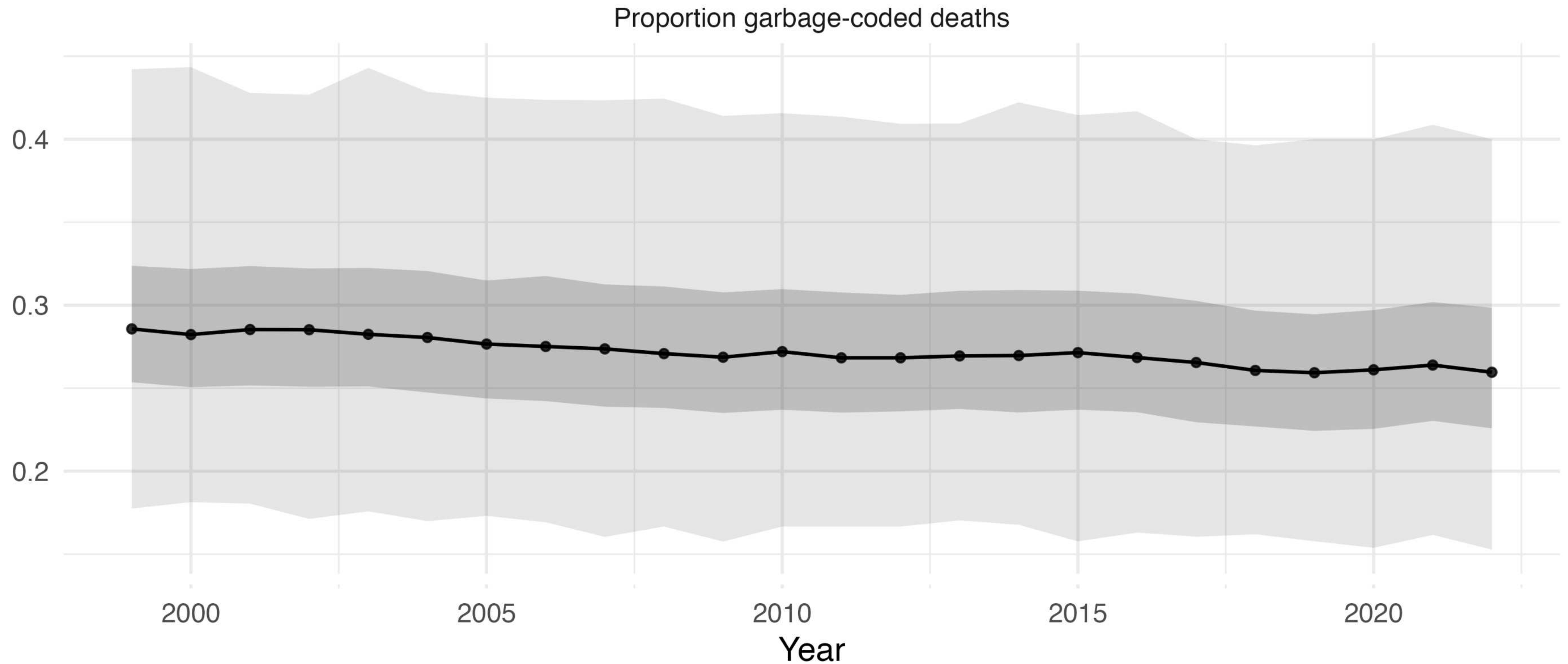
Aggregate quality index, 2020-2022



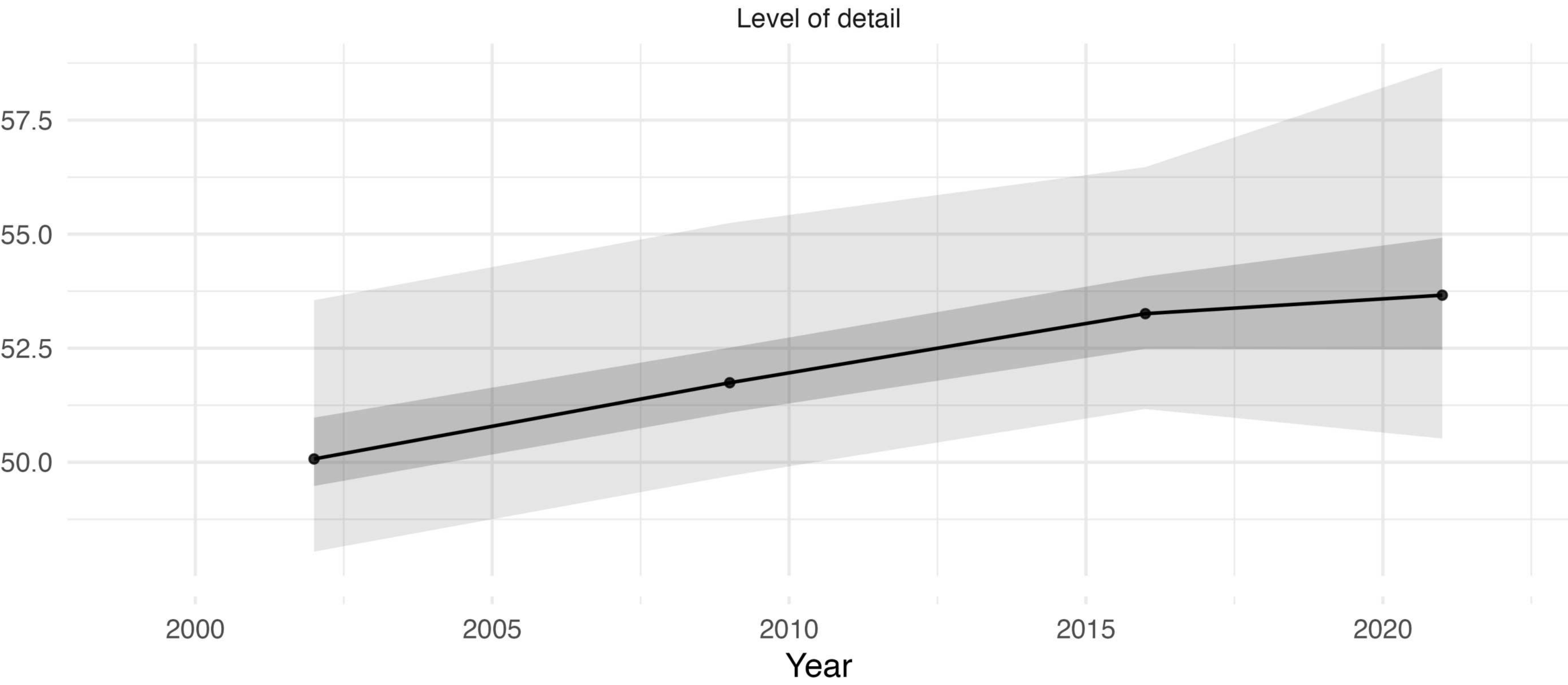
Aggregate qualityindex



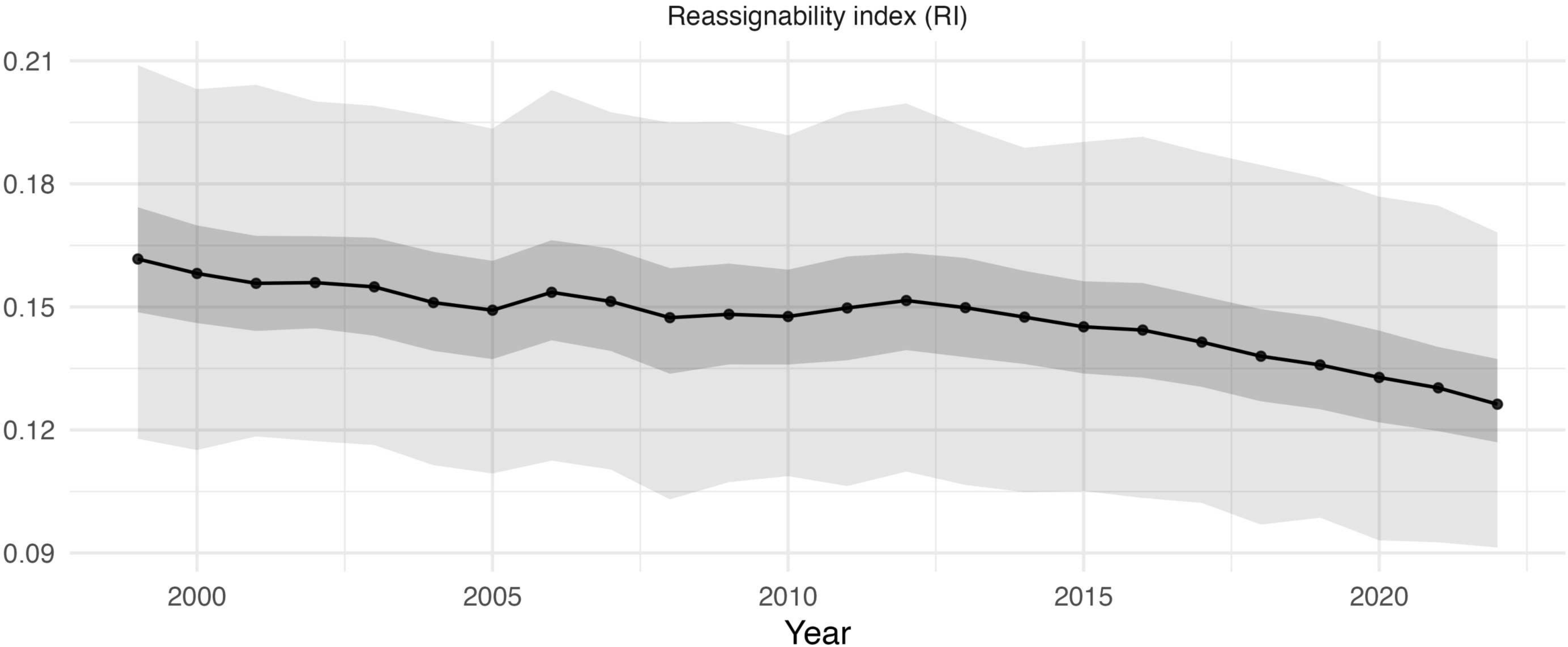
Proportion garbage code over time



Level of detail over time



Re-assignability index over time



Extreme examples - Good

- Counties that are good across all three metrics:
 - Few **garbage codes**
 - Among non-garbage code deaths, lots of **detail**
 - Among garbage code deaths, **informative contributing causes**

County	Prop garb	Level of detail	RI	N deaths
Ventura County, CA	0.205	1.55	0.172	127,300
King County, WA	0.220	1.89	0.162	298,300
Pierce County, WA	0.218	1.91	0.161	146,000

Extreme examples - Bad

- Counties that are good across all three metrics:
 - Lots of **garbage codes**
 - Among non-garbage code deaths, lots of very **little detail**
 - Among garbage code deaths, **uninformative contributing causes**

County	Prop garb	Level of detail	RI	N deaths
Marengo County, AL	0.422	-0.60	0.119	6,300
Talladega County, AL	0.407	-0.36	0.120	23,100
Walker County, AL	0.407	0.10	0.111	23,600