2025 Federal Computer Assisted Survey Information Collection Workshops, April 22, 2025
Session 3A: Data Science and Dissemination

Tracking Cross-State Achievement Gap Trajectory in NAEP Assessments with Dynamic Time Warping Methods

Qiwei (Britt) He

Provost' Distinguished Associate Professor
Data Science and Analytics Program
AI Measurement and Data Science Lab
Georgetown University

17 89
GEORGETOWN UNIVERSITY

Content

- **Overview**
  - National Report Card
  - Cross-State Achievement Gap Trajectories in NAEP
- **Methods**
  - Dynamic Time Warping Similarity Measurement
  - Sequence Clustering
- **Results**
  - Trajectory patterns across states
  - Achievement gap changes across grades and subjects
- **Discussion**

1 Overview

- National Report Card
- Cross-State Achievement Gap Trajectories

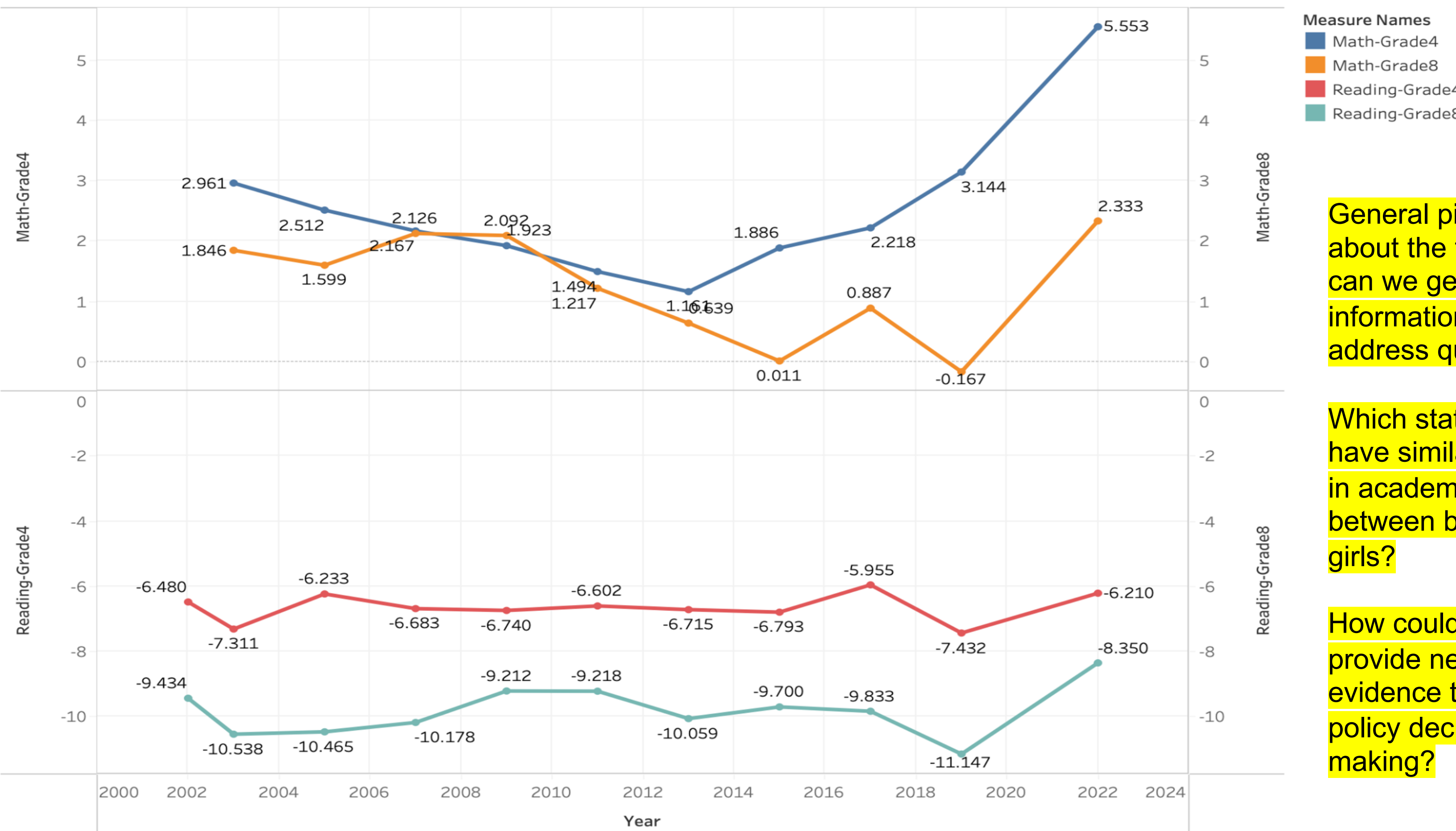
National Report Card

NAEP
NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

- Educational assessment plays **a crucial role** in uncovering learning disparities and shaping data-informed education policy (NCES, 2023).
- The **National Assessment of Educational Progress (NAEP)** is a nationally representative and continuing assessment of what America's students know and can do in various subjects, generally in Grade 4, 8, 12.
- Academic performance by different demographic groups, such as achievement gap between boys and girls, has aroused great attention since the start of NAEP.
- Within the past two decades, **over 42,900 journal papers** were published on the topic with "NAEP and demographic groups".
- However, **very few research** showed a **comprehensive picture through the NAEP development history and identify trajectory changes across states.**

Overview of Academic Gap by Years and Subjects

National Reading VS. Math Gender Gap



The trends of Math-Grade4, Math-Grade8, Reading-Grade4 and Reading-Grade8 for Year. Color shows details about Math-Grade4, Math-Grade8, Reading-Grade4 and Reading-Grade8.

General picture about the trajectory, can we get granular information to address questions:

Which state(s) may have similar patterns in academic gap between boys and girls?

How could we provide new evidence to support policy decision making?

Objectives

- Track cross-state achievement gap trajectories in NAEP Assessments
- Identify trajectory patterns by boys and girls across grades and subjects
- Provide new evidence to policy makers through data-driven approach

2 Methods

- Dynamic Time Warping Similarity Measurement
- Sequence Clustering

Data

- **Subjects:** Math and Reading across 50 states

- **Grades:** Grade 4 and Grade 8

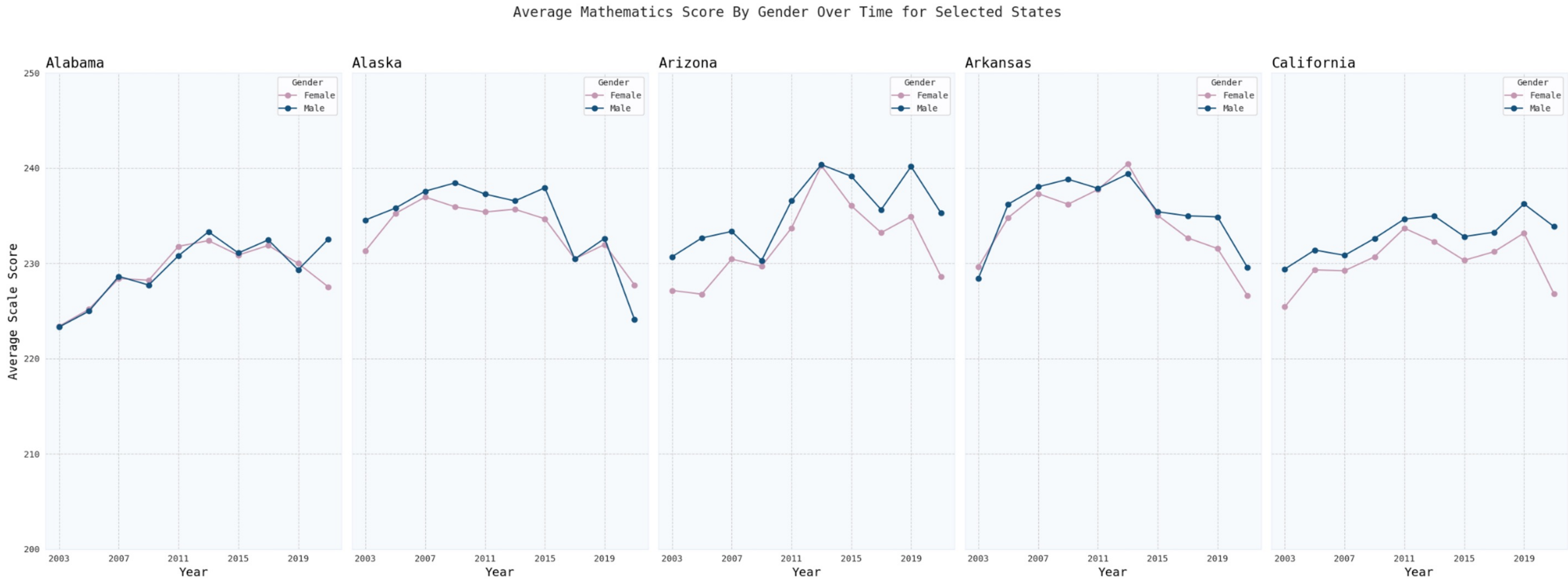
- **Years Covered:** 2002–2022

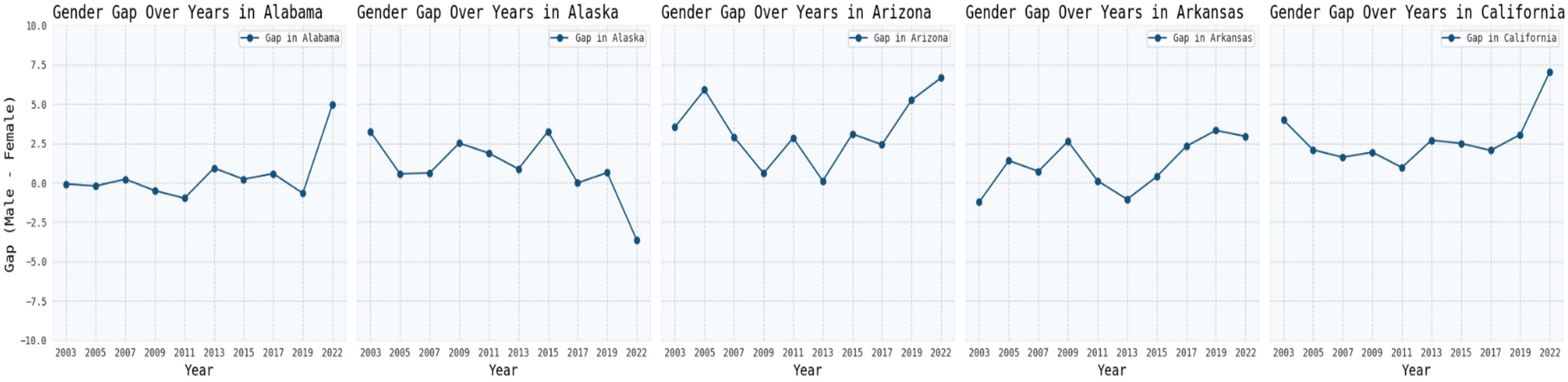
- **Gender gap:** $M_{male} - M_{female}$ per cycle per subject

- Sequence data analysis

NAEP NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Sequence Data by State (Grade 4 Math)

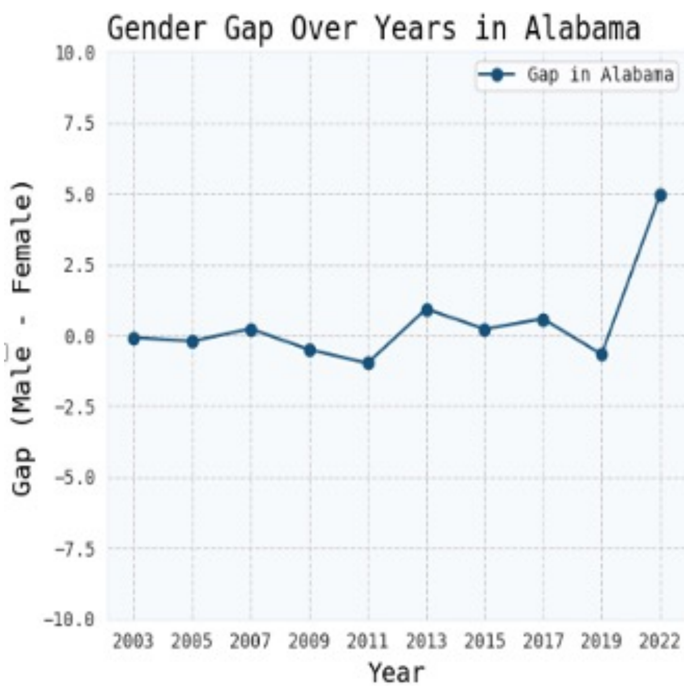




Sequence Data by State

Gender difference across years in Grade 4 Math

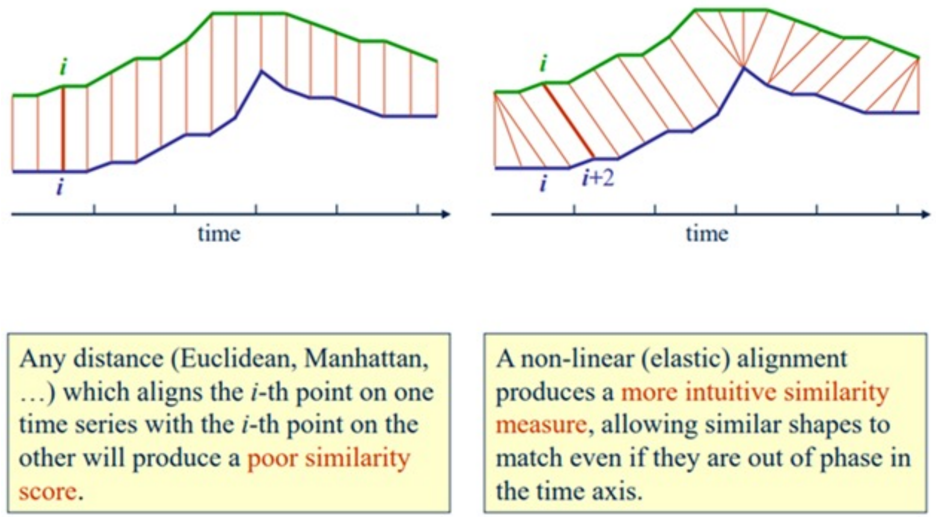
	2003	2005	2007	2009	2011	2013	2015	2017	2019	2022
Alabama	-0.08	-0.21	0.23	-0.51	-0.98	0.93	0.22	0.58	-0.67	4.96
Alaska	3.23	0.56	0.62	2.52	1.87	0.87	3.26	-0.01	0.64	-3.64



Dynamic Time Warping (DTW) (He et al., 2023)

- Dynamic time warping (Sakoe & Chiba, 1978) is a distance measure that searches the optimal warping path between two series.

- Given sequences $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$ with the same or different lengths, a warping path W is an alignment between X and Y, involving one-to-many mappings for each pair of elements.



He, Q., Borgonovi, F., Suárez-Álvarez, J. (2023). Clustering Sequential Navigation Patterns in Multiple-Source Reading Tasks with Dynamic Time Warping Method. *Journal of Computer-Assisted Learning, 39*(3), 719-736.

Dynamic Time Warping Algorithm

- The initial step of DTW algorithm is defined as:

$$DTW(i,j) = \begin{cases} \infty & if\ (i = 0\ or\ j = 0)\ and\ i \neq j \\ 0 & if\ i = j = 0 \end{cases}$$

- The recursive function of DTW is defined as

$$DTW(i,j) = min \begin{cases} DTW(i-1,j) + w_h C(i,j) \\ DTW(i,j-1) + w_v C(i,j) \\ DTW(i-1,j-1) + w_d C(i,j) \end{cases}$$

where (w_h, w_v, w_d) are weights for the horizontal, vertical and diagonal directions, respectively. $DTW(i,j)$ denotes the distance or cost between two sub-sequences $\{x_1, x_2, \ldots, x_i\}$ and $\{y_1, y_2, \ldots, j\}$, and $DTW(N,M)$ indicates the total cost of the optimal warping path.

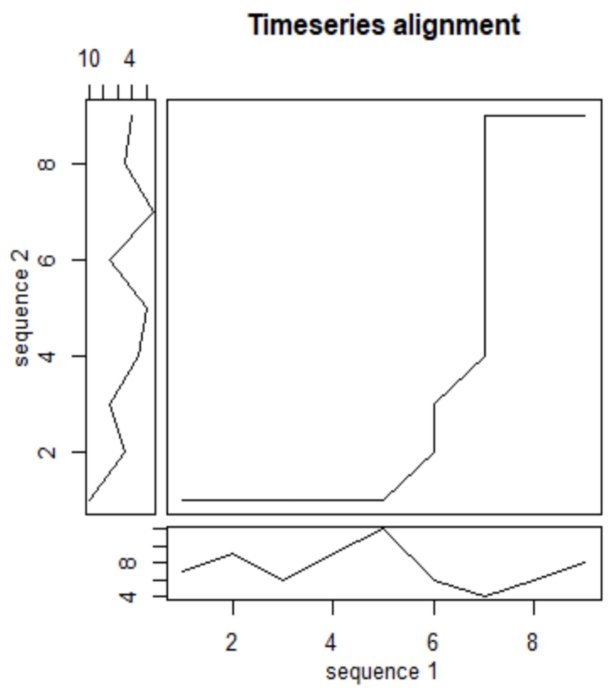
A	**5**	10	6	3	1	2	4	3
	4	6	3	1	0	1	3	3
	3	3	1	0	1	1	2	4
	2	1	0	1	3	4	4	7
	1	0	1	3	6	8	9	13
		1	**2**	**3**	**4**	**3**	**2**	**5**
		B						

$A_i = \{1, 2, 3, 4, 5\}$
$B_j = \{1, 2, 3, 4, 3, 2, 5\}$
$$dtw(i,j) = |A_i - B_j| + \min(D[i-1, j-1], D[i-1, j], D[i, j-1])$$
$$= |5 - 2| + 3$$

Dynamic Time Warping Similarity

Dynamic Time Warping (DTW)

- A time-series alignment algorithm to measure similarity between sequences that may vary in time or speed.
- Useful when states drop from certain cycles, causing unequal sequence lengths.
- DTW constructs a cost matrix and finds the optimal warping path between two sequences.
- Only applicable to numeric sequences, so categorical data needs to be numerically encoded.



Sequence Clustering

Sequence Clustering

- Performed on the DTW distance matrix.

- Uses partition-based clustering (K-medoids), where clusters are formed based on minimizing within-cluster DTW distances.
  - Starts by randomly selecting representative points (medoids).
  - Assigns other sequences to the nearest medoid.
  - Repeats until cluster centers stabilize.

- **Cluster Validity Indices Used**
  - **Silhouette Index:** Measures how similar an object is to its cluster vs. others.
  - **Dunn Index:** High inter-cluster, low intra-cluster distances preferred.
  - **Davies-Bouldin Index:** Measures cluster compactness and separation.

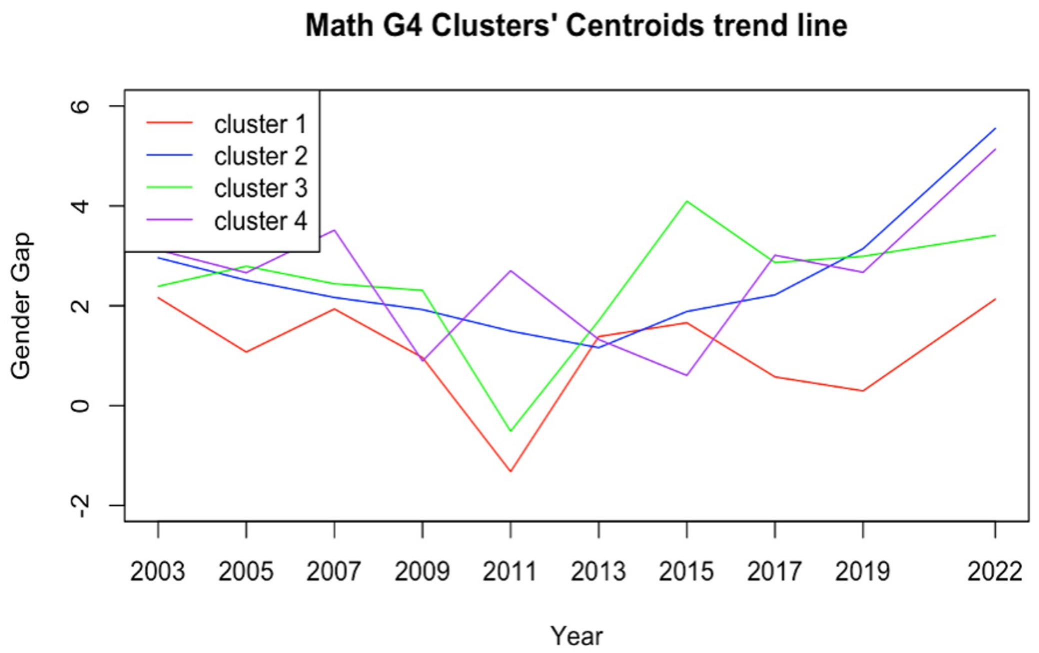
- Iterated values of K = 2 to 10, selected **4 clusters** as optimal

3 Results

- Trajectory patterns across states
- Achievement gap changes across grades and subjects

Trajectory Pattern Across States – Grade 4 Math

- **Cluster 1 (Red):** Significant drop in 2011 and 2015–2019, sharp increase post-2019.
- **Cluster 2 (Blue):** Relatively stable from 2003–2017, but steeply increase after 2017.
- **Cluster 3 (Green):** Similar shape as Cluster 1 but with not significantly impact by COVID.
- **Cluster 4 (Purple):** Up and down in big waves but steeply rises after 2019.



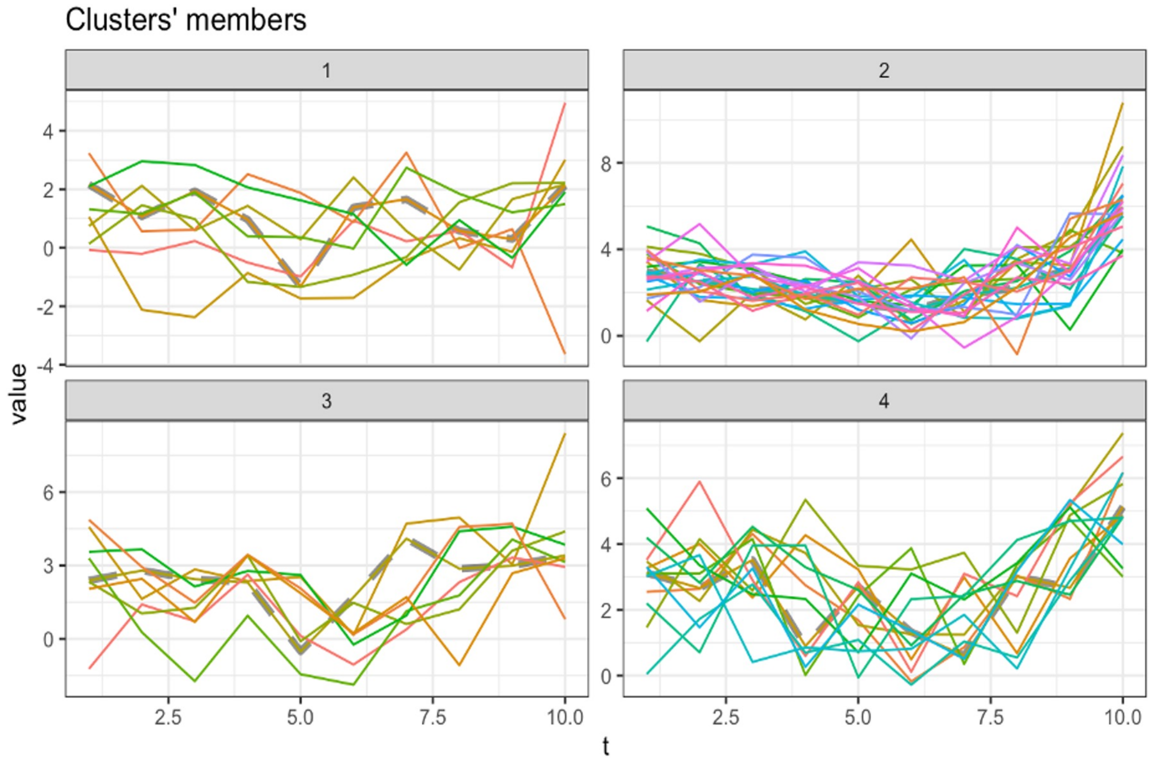
Trajectory Pattern Across States – Grade 4 Math

Cluster 1 : (8 States)
Alabama, Alaska, Georgia, Hawaii, Louisiana, Mississippi, North Carolina, West Virginia

Cluster 2 : (21 States)
California, Colorado, Florida, Idaho, Indiana, Iowa, Kansas, Kentucky, Michigan, Missouri, Nevada, New Jersey, New Mexico, New York, Ohio, Pennsylvania, Utah, Vermont, Virginia, Wisconsin, Wyoming

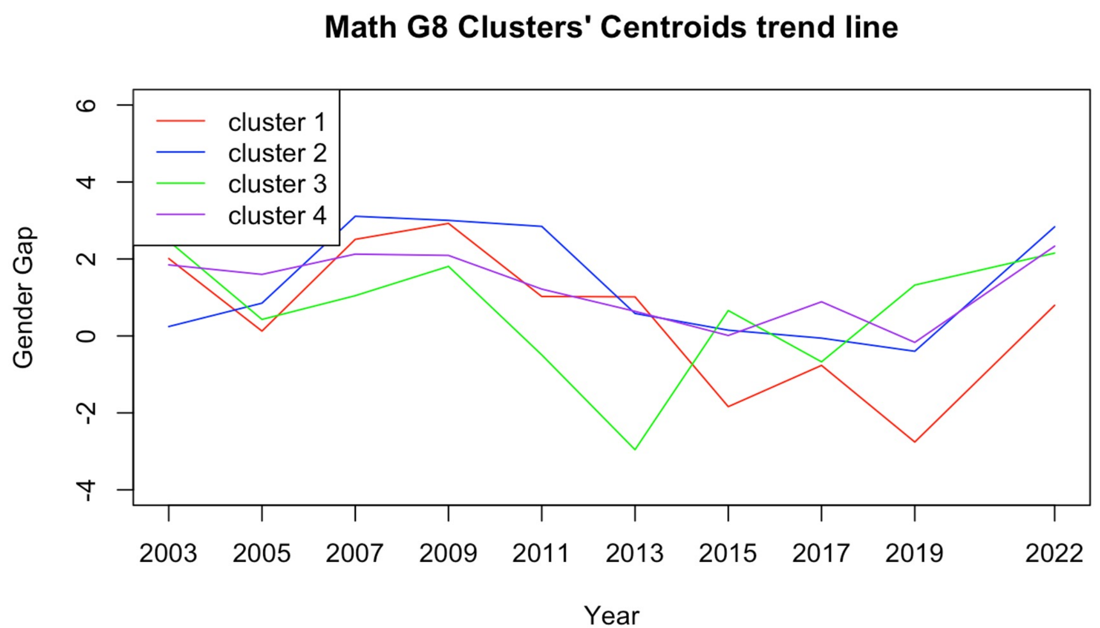
Cluster 3 : (8 States)
Arkansas, Connecticut, Delaware, Massachusetts, Oklahoma, Rhode Island, South Carolina, South Dakota"

Cluster 4 : (13 States)
Arizona, Illinois, Maine, Maryland, Minnesota, Montana, Nebraska, New Hampshire, North Dakota, Oregon, Tennessee, Texas, Washington



Trajectory Pattern Across States – Grade 8 Math

- **Cluster 1 (Red):** significantly dropped between 2015 and 2019, but sharply increased after 2019.
- **Cluster 2 (Blue):** Stable and a bit down between 2003 and 2017, sharply increased after 2019.
- **Cluster 3 (Green):** Gap dropped to the lowest value in 2013, and kept fast increasing afterwards.
- **Cluster 4 (Purple):** Stable and a small dropping trend until 2019, where a noticeable spike appeared after 2019.



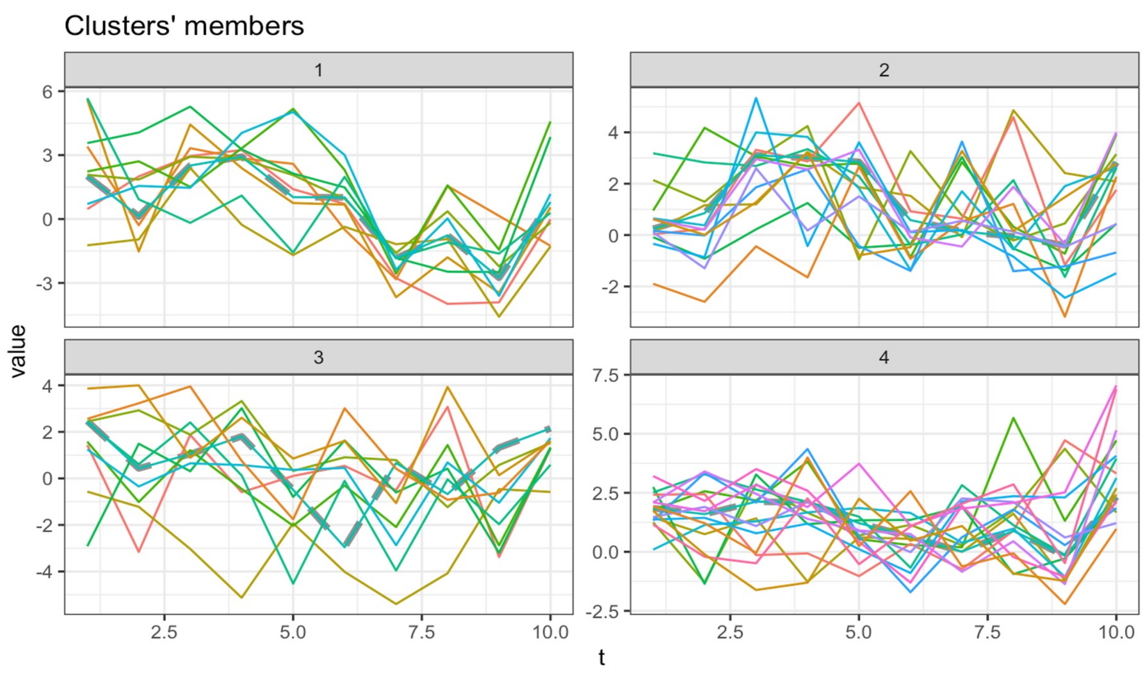
Trajectory Pattern Across States – Grade 8 Math

Cluster 1 : (10 States)
Kentucky, Maryland, Massachusetts, North Carolina, Ohio, Oregon, Pennsylvania, South Carolina, South Dakota, Wyoming

Cluster 2 : (14 States)
Arizona, Arkansas, Colorado, Kansas, Maine, Michigan, Montana, Nebraska, North Dakota, Oklahoma, Tennessee, Vermont, West Virginia, Wisconsin

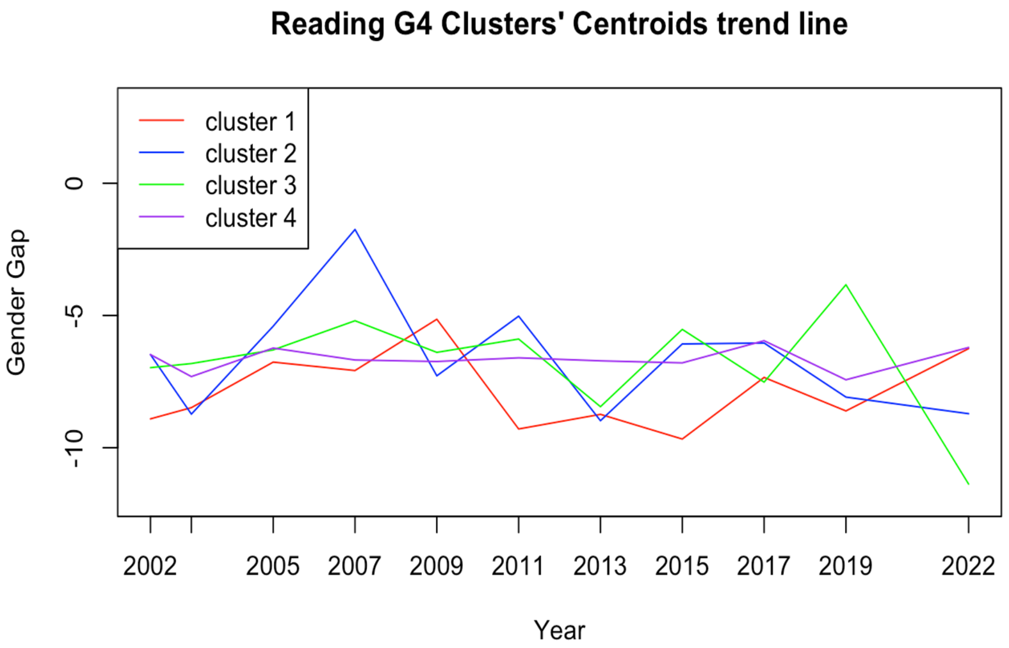
Cluster 3 : (10 States)
Alabama, Delaware, Florida, Hawaii, Indiana, Louisiana, Minnesota, Mississippi, New York, Rhode Island

Cluster 4 : (16 States)
Alaska, California, Connecticut, Georgia, Idaho, Illinois, Iowa, Missouri, Nevada, New Hampshire, New Jersey, New Mexico, Texas, Utah, Virginia, Washington



Trajectory Pattern Across States – Grade 4 Reading

- **Cluster 1 (Red)**: Consistently large gap (girls outperform over boys), gap slightly narrowed after 2019.
- **Cluster 2 (Blue)**: High peak around 2007 indicating the smallest gender gap, curve waves showing in the recent decade.
- **Cluster 3 (Green)**: Fairly stable until 2017, with a sharp spike in 2019 (with the smallest gender gap), followed by a sharp drop in 2022 after COVID.
- **Cluster 4 (Purple)**: Consistently stable with moderate gender gap across all years.



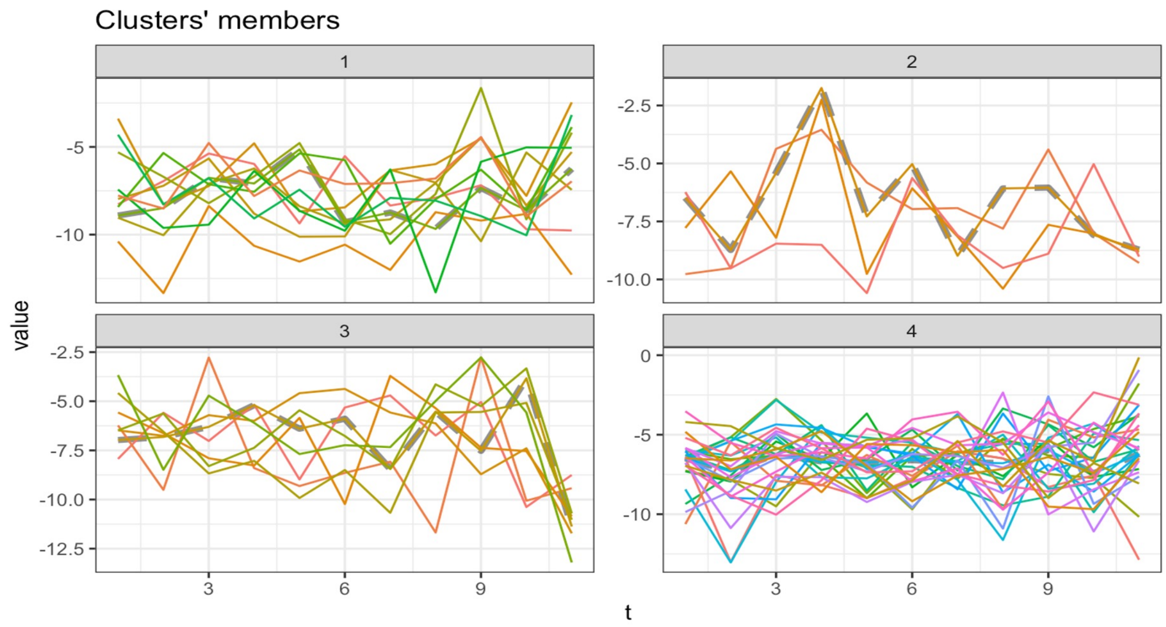
Trajectory Pattern Across States – Grade 4 Reading

Cluster 1 : (11 States)
Alabama, Florida, Hawaii, Indiana, Missouri, Oregon, Pennsylvania, South Carolina, Vermont, Washington, West Virginia

Cluster 2 : (4 States)
Arkansas, Montana, New Mexico, Wisconsin

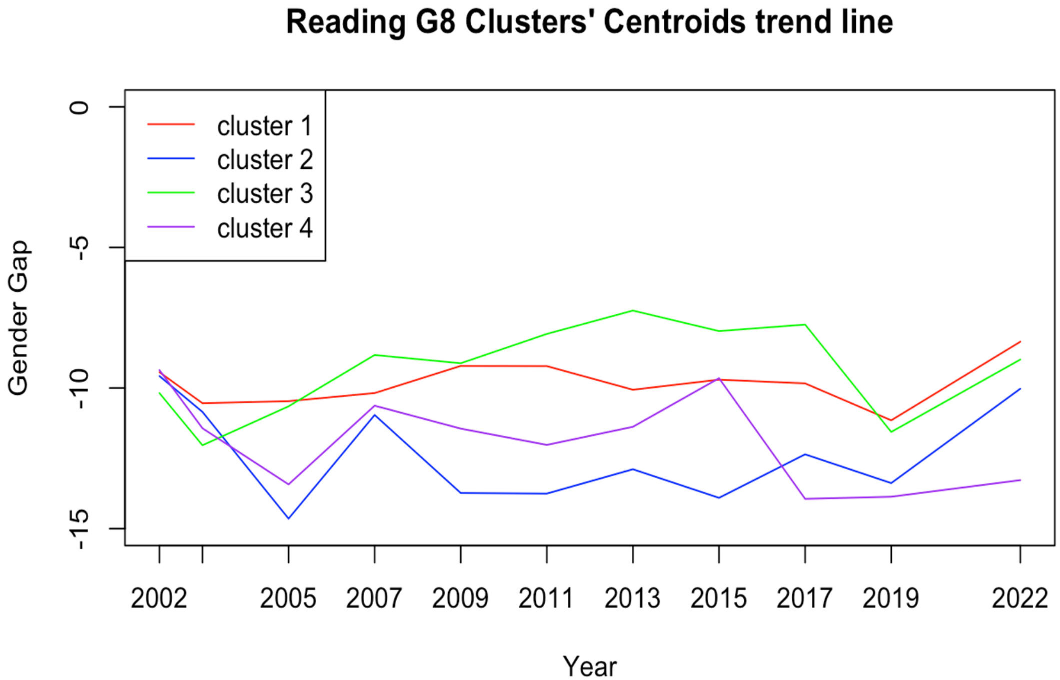
Cluster 3 : (8 States)
Idaho, Louisiana, Mississippi, New Jersey, Oklahoma, Rhode Island, South Dakota, Virginia

Cluster 4 : (27 States)
Alaska, Arizona, California, Colorado, Connecticut, Delaware, Georgia, Illinois, Iowa, Kansas, Kentucky, Maine, Maryland, Massachusetts, Michigan, Minnesota, Nebraska, Nevada, New Hampshire, New York, North Carolina, North Dakota, Ohio, Tennessee, Texas, Utah, Wyoming



Trajectory Pattern Across States – Grade 8 Reading

- **Cluster 1 (Red)**: Consistently wide gender gap across years, with a slight narrowing in 2013–2017 and a small uptick after 2019.
- **Cluster 2 (Blue)**: The largest gender gap, esp. the lowest point at 2005, the gap slightly narrowed down after 2017.
- **Cluster 3 (Green)**: The narrowest gender gap among clusters between 2007 and 2015, a big drop in 2019, with a slight increasing trend (gap narrowing) in 2022.
- **Cluster 4 (Purple)**: Moderate fluctuations with a noticeable drop in 2017, relatively stable afterwards.



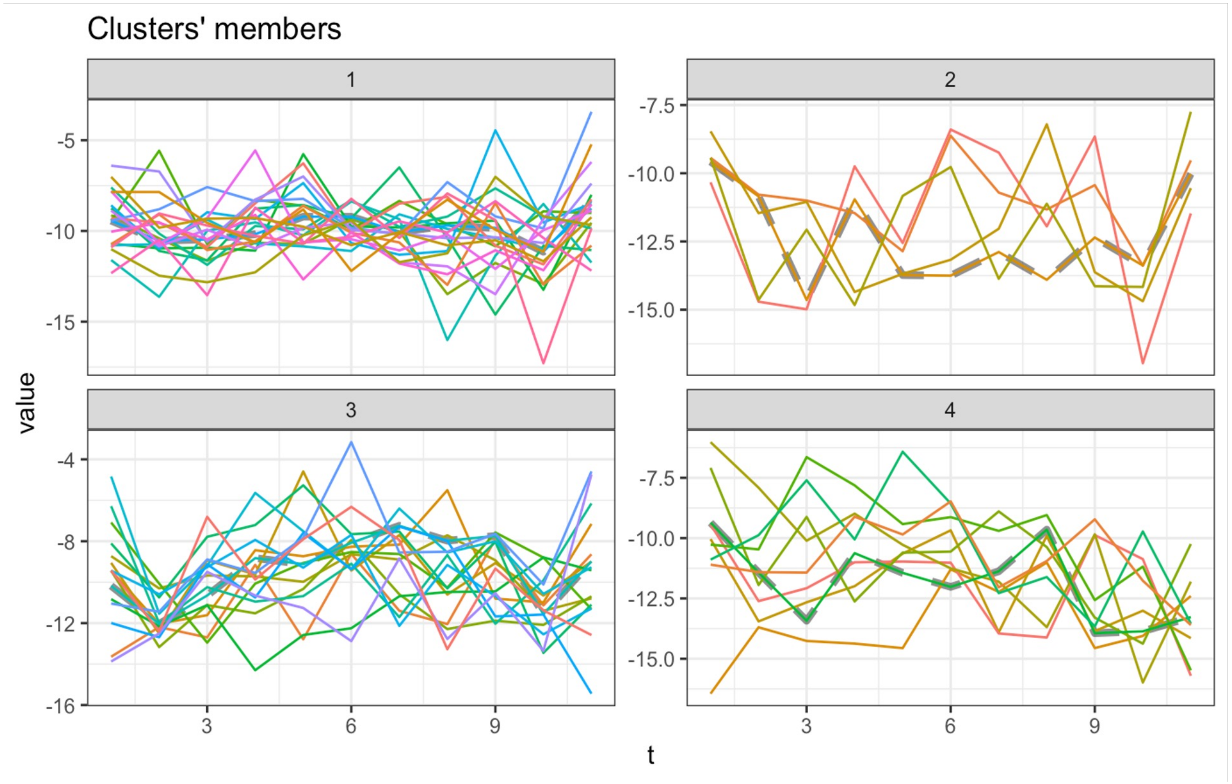
Trajectory Pattern Across States – Grade 8 Reading

Cluster 1 : (20 States)
Arizona, Arkansas, California, Delaware, Florida, Georgia, Illinois, Indiana, Kentucky, Louisiana, Maryland, Michigan, Mississippi, New Jersey, Ohio, Oregon, Rhode Island, South Carolina, Utah, Virginia

Cluster 2 : (5 States)
Alabama, New Hampshire, North Carolina, West Virginia, Wisconsin

Cluster 3 : (16 States)
Colorado, Idaho, Iowa, Kansas, Massachusetts, Minnesota, Nebraska, Nevada, New Mexico, New York, Oklahoma, Pennsylvania, South Dakota, Tennessee, Texas, Washington

Cluster 4 : (9 States)
Alaska, Connecticut, Hawaii, Maine, Missouri, Montana, North Dakota, Vermont, Wyoming



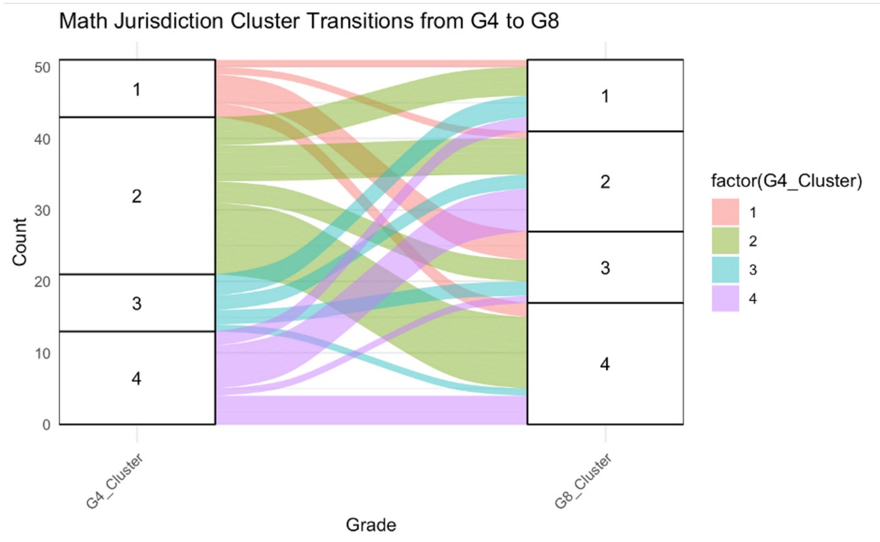
Achievement Gap Changes by Grades and Subjects

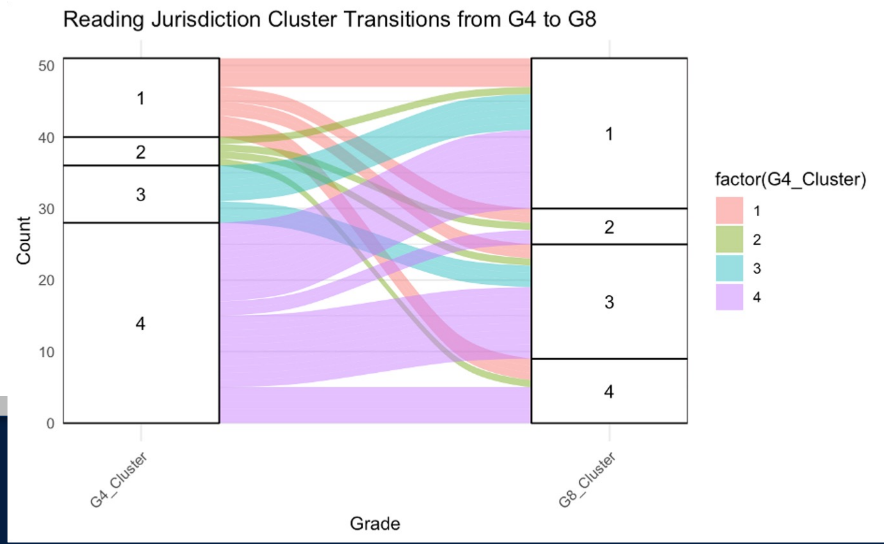
Alluvial Plots used to visualize cluster shifts:

- Transition from Grade 4 to Grade 8
- Math vs Reading.

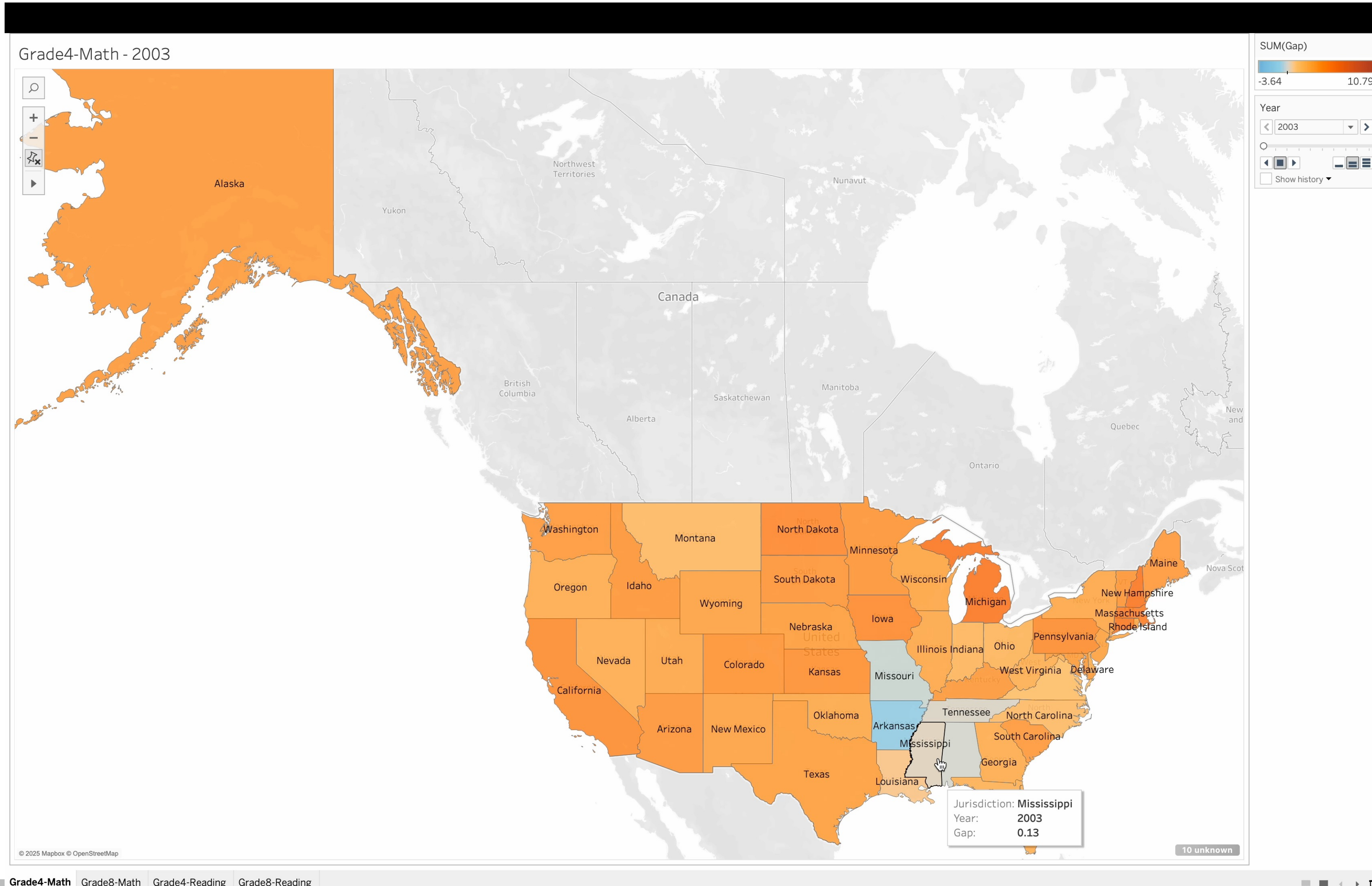
Findings

- **Math** shows many cross-cluster transitions, indicating less stability across grade.
- **Reading** shows more jurisdictions remaining in the same cluster or nearby clusters.
- **Reading is more stable than Math** across grade 4 and grade 8.

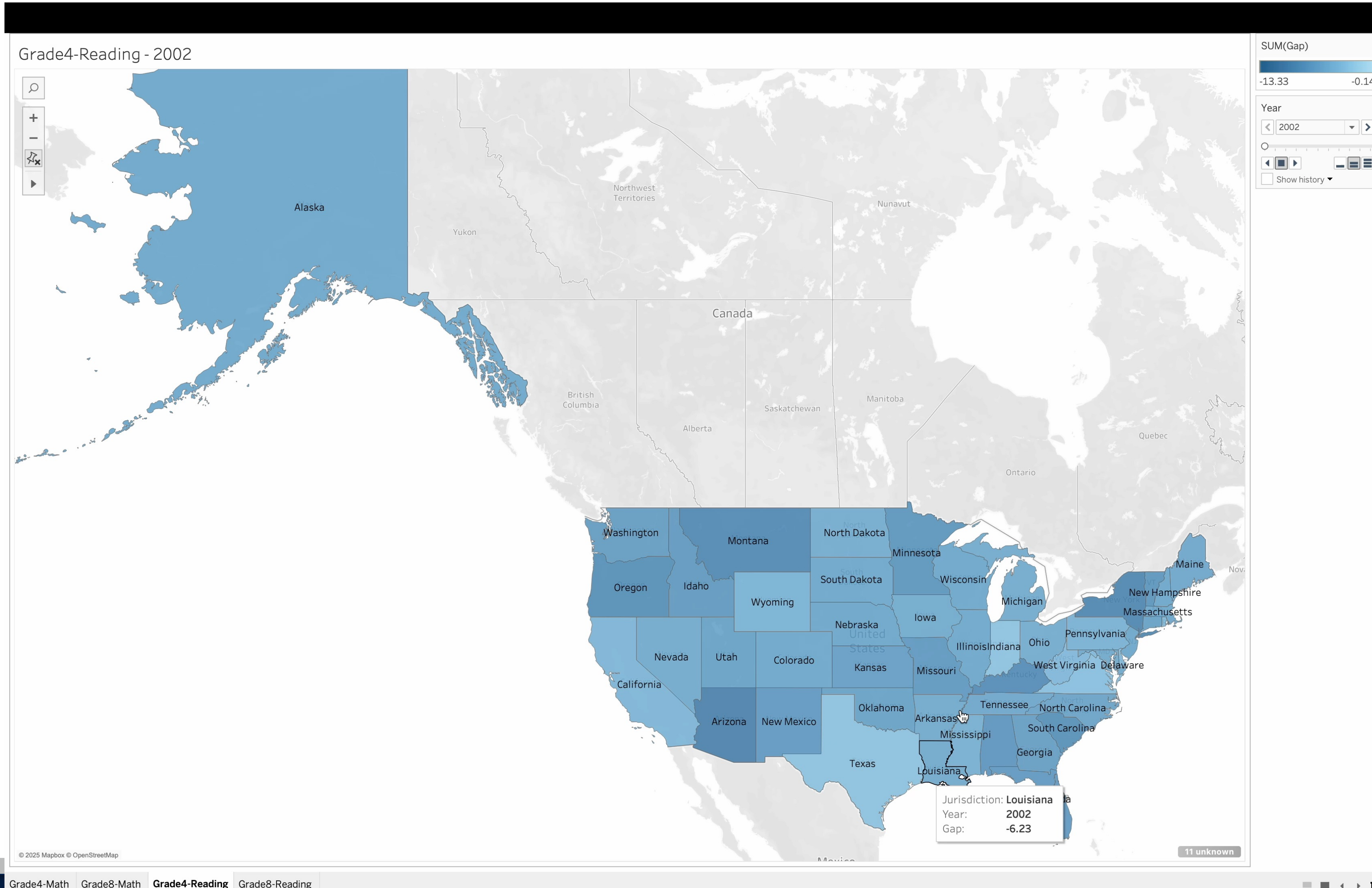




Grade 4 Math Achievement Gap by Gender Across Years



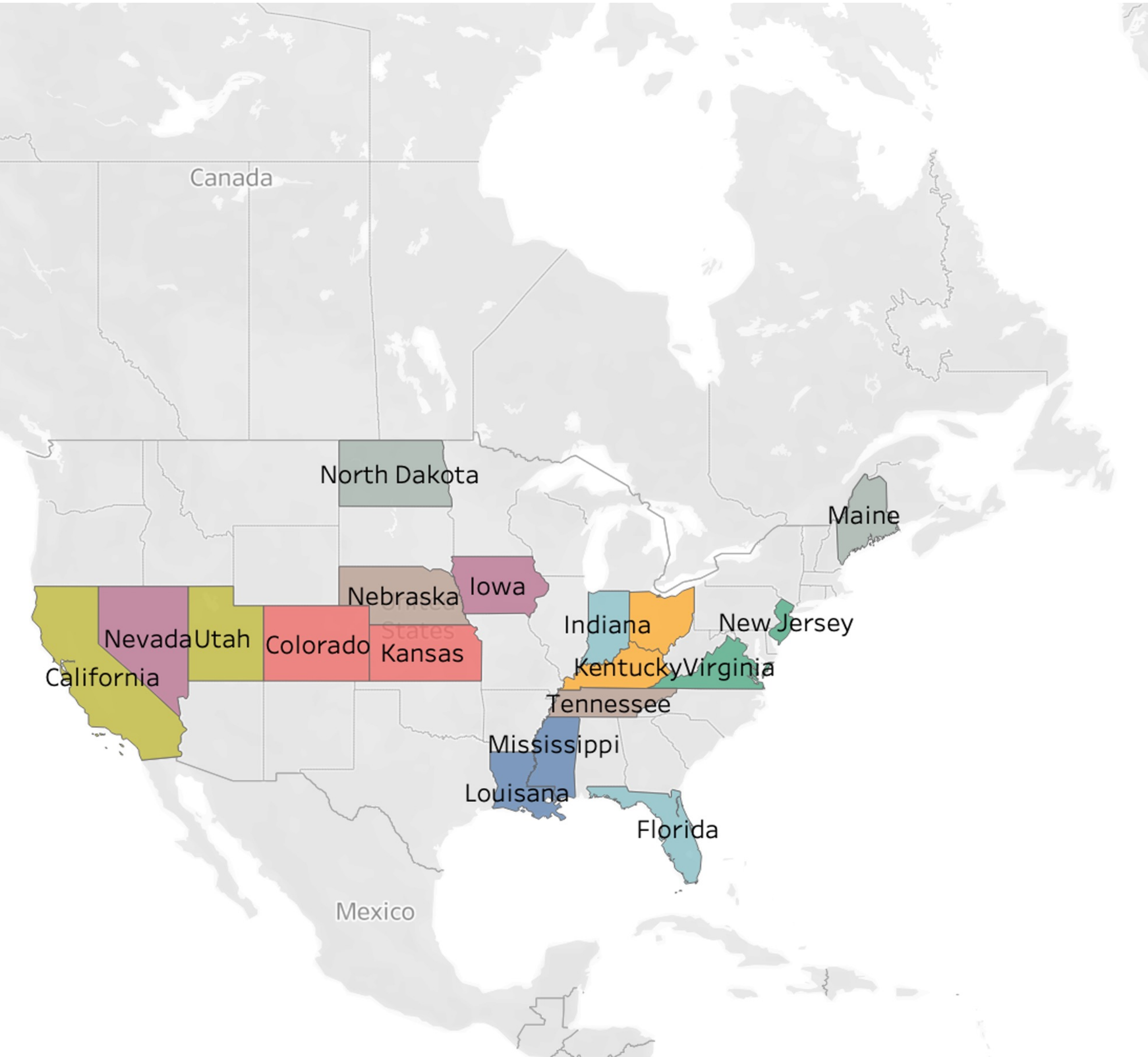
Grade 4 Reading Achievement Gap by Gender Across Years



Geographical Patterns shown in Achievement Gaps

- Nine state pairs consistently grouped together across all settings:
    - Kentucky & Ohio
    - California & Utah
    - Florida & Indiana
    - Louisiana & Mississippi
    - Colorado & Kansas
    - New Jersey & Virginia
    - Iowa & Nevada
    - Nebraska & Tennessee
    - North Dakota & Maine

- Indicates possible shared policies, demographics, or regional factors.



4 Discussion

Summary

- The sequence-based dynamic time warping similarity measurement and sequence clustering support **visualization on trajectory pattern** of achievement gaps across states, and identify **homogenous patterns across state clusters**.

- The achievement gaps between boys and girls significantly **enlarged in Math after COVID**.

- The achievement gaps between boys and girls is **less stable in math than in reading**, suggesting the educational policy or curriculum changes might impact more sensitively in math learning.

- The patterns of achievement gaps trajectory patterns are **more stable in Grade 8 than in Grade 4**, suggesting the younger group might be more impacted with the change of educational policy and curriculum development.

Limitation and Future Work

- This study provides a novel perspective on the impact of educational policies on student performance. Future research could **incorporate additional demographic variables, such as school type, socio-economic status (SES), and region, to analyze trajectory patterns**.

- Methodology developed in this study can be **generally extended to school districts, cities**, to track trajectories of students' growth, academic gaps by different groups.

- Developing a **centralized portal or data platform** could serve as a **valuable tool for tracking performance trends** in NAEP and other assessments over time.

- Findings from this study should be linked **more closely to state-level educational policies** to identify and address the root causes of achievement gaps.

AI Measurement and Data Science Lab

- Process data analytic methods
- AI item generation
- Psychiatry acuity measurement
- ML on missing data prediction
- Eye-tracking for ELL
- VR/AR in assessment

Thank you very much!

Qiwei (Britt) He, PhD
qiwei.he@georgetown.edu

GEORGETOWN UNIVERSITY