INVESTIGATING THE DRIVERS AND CHARACTERISTICS OF HIGHER EDUCATION INSTITUTIONAL CLOSURES POST-2020

BY SEPEHR AKBARI

WHAT PERCENTAGE OF TOTAL U.S. ECONOMIC GROWTH ORIGINATES SOLELY FROM UNIVERSITY-DRIVEN RESEARCH AND INNOVATION?

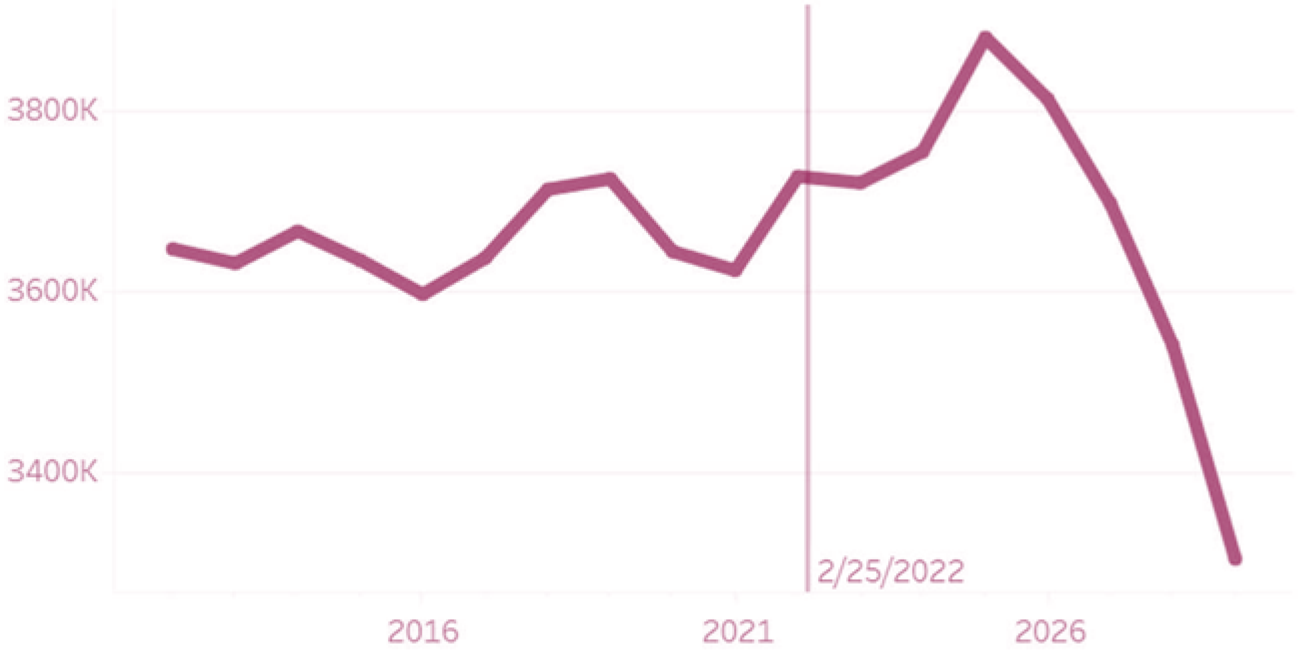
AT LEAST

60%

BROOKINGS (2023)

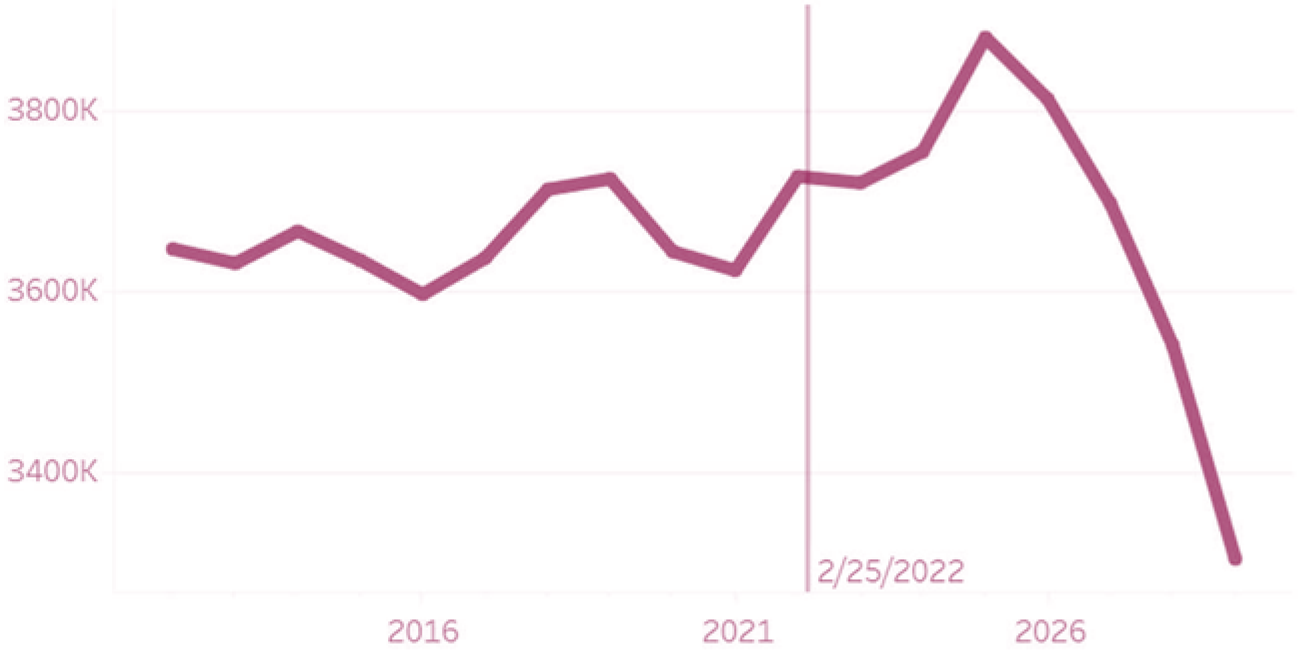
WHAT PERCENTAGE OF TOTAL U.S. ECONOMIC GROWTH ORIGINATES SOLELY FROM UNIVERSITY-DRIVEN RESEARCH AND INNOVATION?

COLLEGES NEED ENROLLMENT



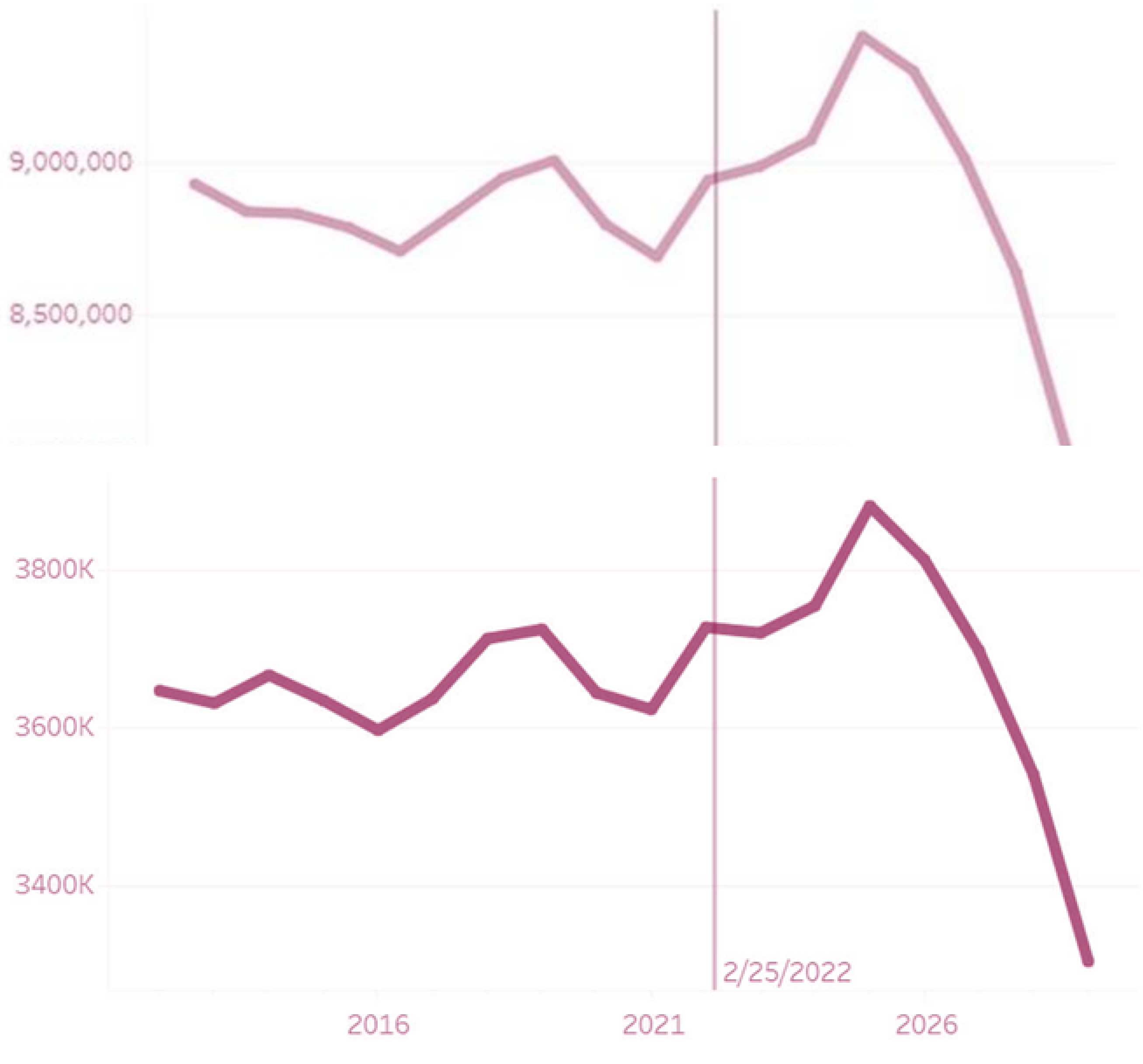
College Enrollment Figures & Projections

Figure(s) Sourced from CUPA-HR



THE CLIFF

- College enrollment projections anticipate a **15% drop after 2025**.
- Between 2025 and 2029 the number of college-bound student will **decline by 400,000 student**.
- Since 2019, the number of undergraduate enrollments has **consistently decreased**.
- Since 2020, **66** higher education institutions have closed; **none opened**.



The Cliff

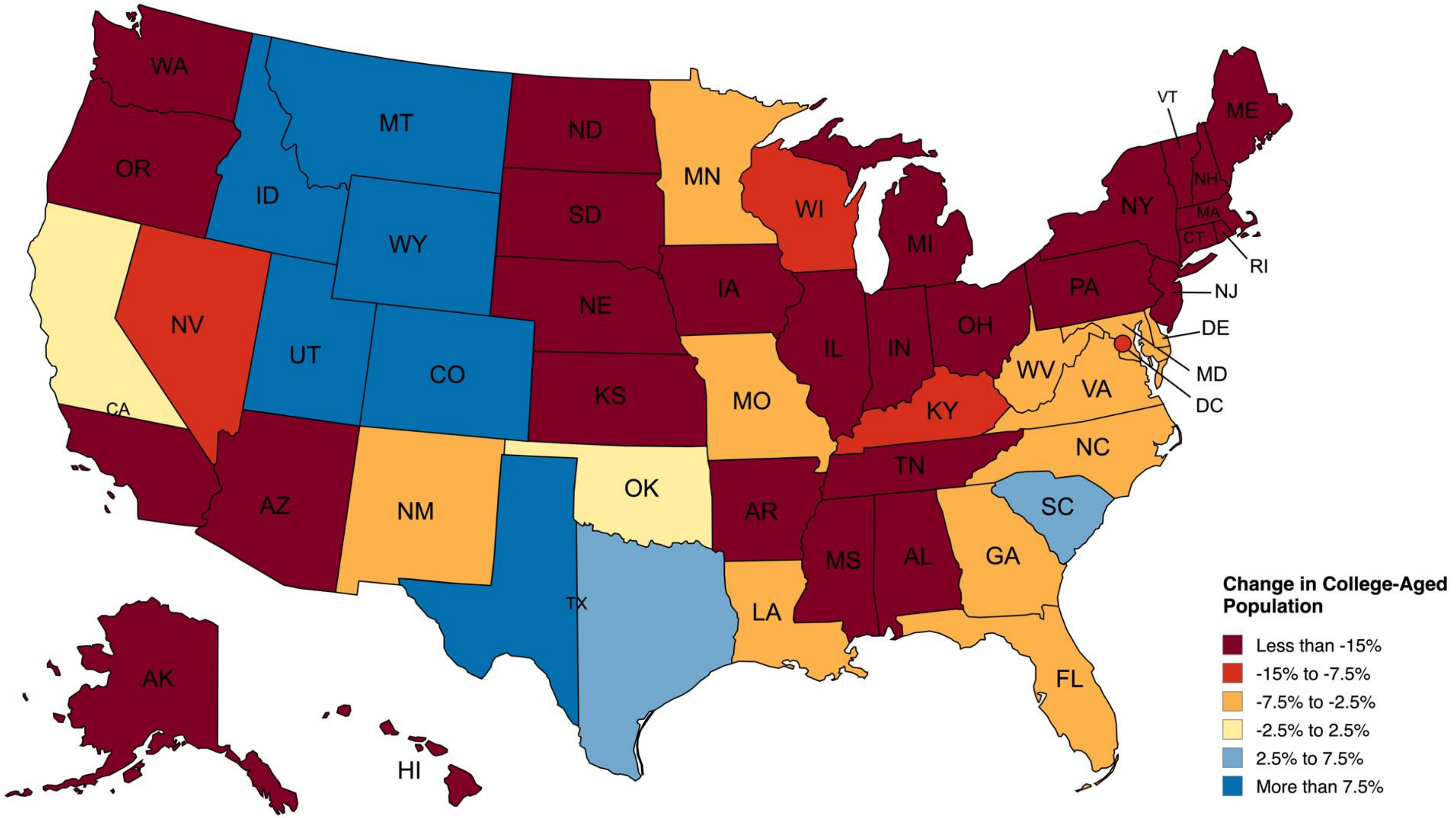
- College enrollment projections anticipate a **15% drop after 2025**.
- Between 2025 and 2029 the number of college-bound student will **decline by 400,000 student**.
- Since 2019, the number of undergraduate enrollments has **consistently decreased**.
- Since 2020, **66** higher education institutions have closed; **none opened**.

Figure(s) Sourced from CUPA-HR

“The number of high school graduates is projected to decline significantly after 2025, leading to a substantial reduction in the pool of traditional college-aged students.”

– NATHAN D. GRAWE

2012-2029 LOSSES AND GAINS IN COLLEGE-GOING STUDENTS



Figure's Credit to Nathan Grawe & Carleton College

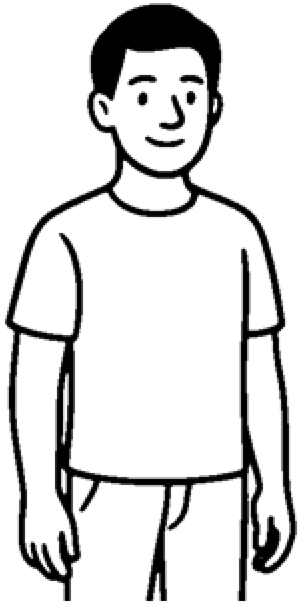
BUT IS THE NUMBER OF PEOPLE THE ONLY FACTOR?

IS URBAN TRAFFIC ONLY CAUSED BY THE NUMBER OF CARS?

Quantifying the Risk: Where are institutions most vulnerable?

- Different states are being affected differently by the demographic changes.
- Grawe argues a regional-level of analysis is beneficial.
- How do we determine how much risk of closure an institution has, by just being in a certain region or state?
- How can we make sure this quantification is general to all institutions in that region/state?

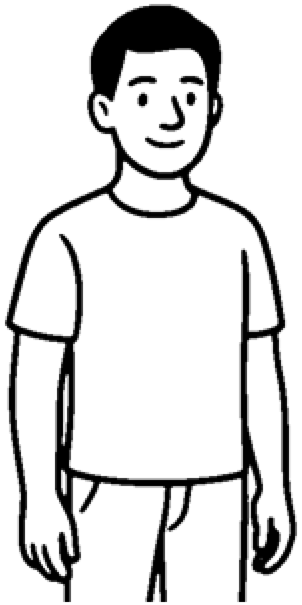
BUT HOW DO WE DO THAT, GIVEN OUR BIASES?



AI-Generated Image/Illustration

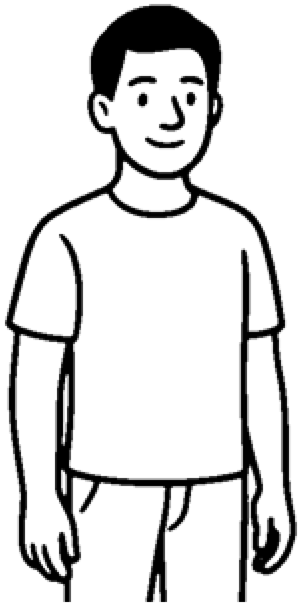
John is very **shy and withdrawn**, invariably helpful but with very little interest in people or in the world of reality, A **meek and tidy soul**, he has a need for order and structure, and a passion for detail.

JOHN



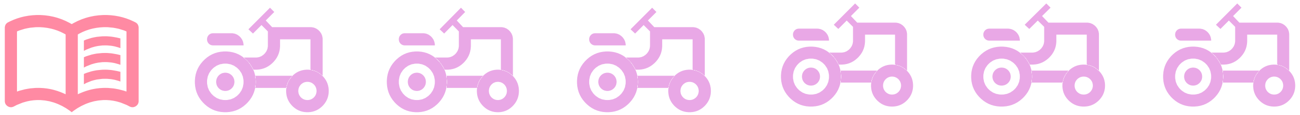
John is very **shy and withdrawn**, invariably helpful but with very little interest in people or in the world of reality, A **meek and tidy soul**, he has a need for order and structure, and a passion for detail.

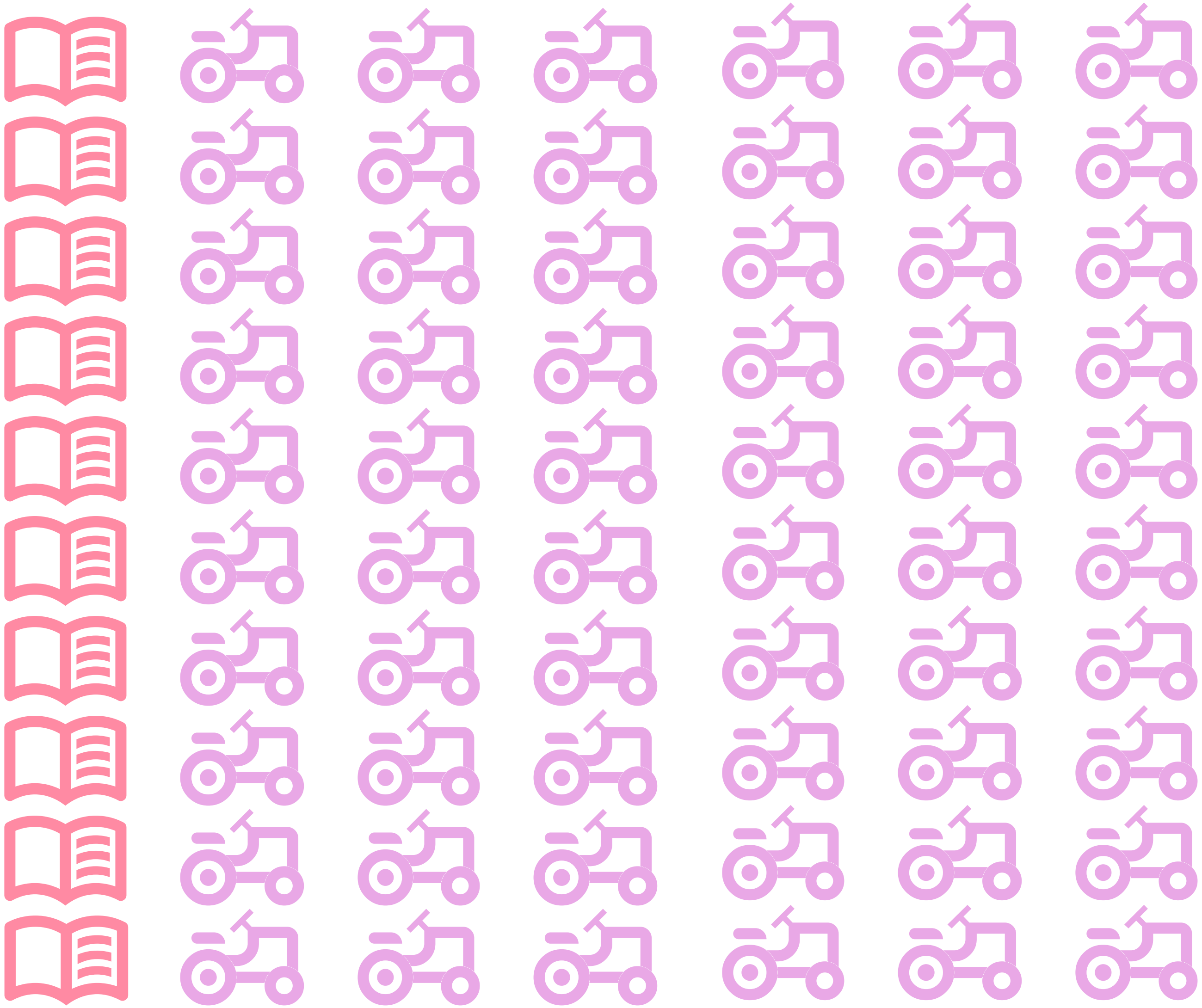
JOHN



AI-Generated Image/Illustration

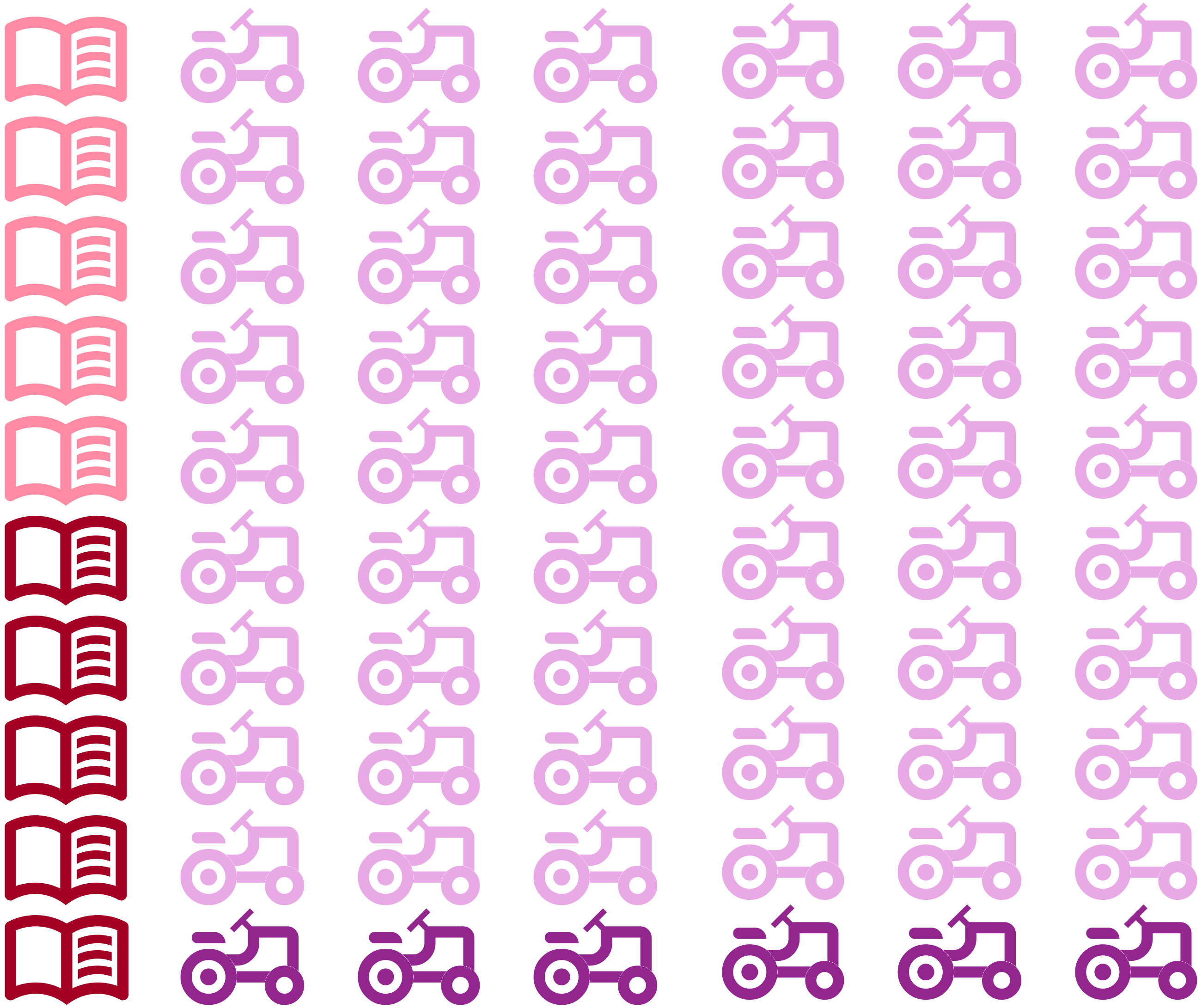
BUT THERE IS **60X** MORE FARMERS THAN LIBRARIANS IN THE US



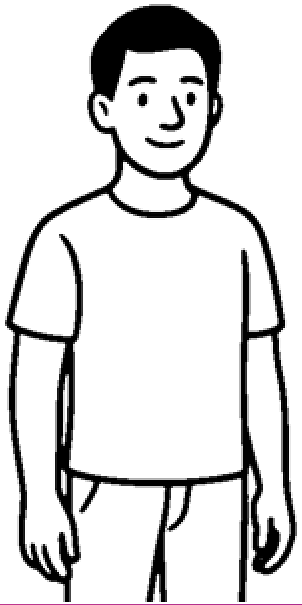


SAMPLE OF
10 LIBRARIANS

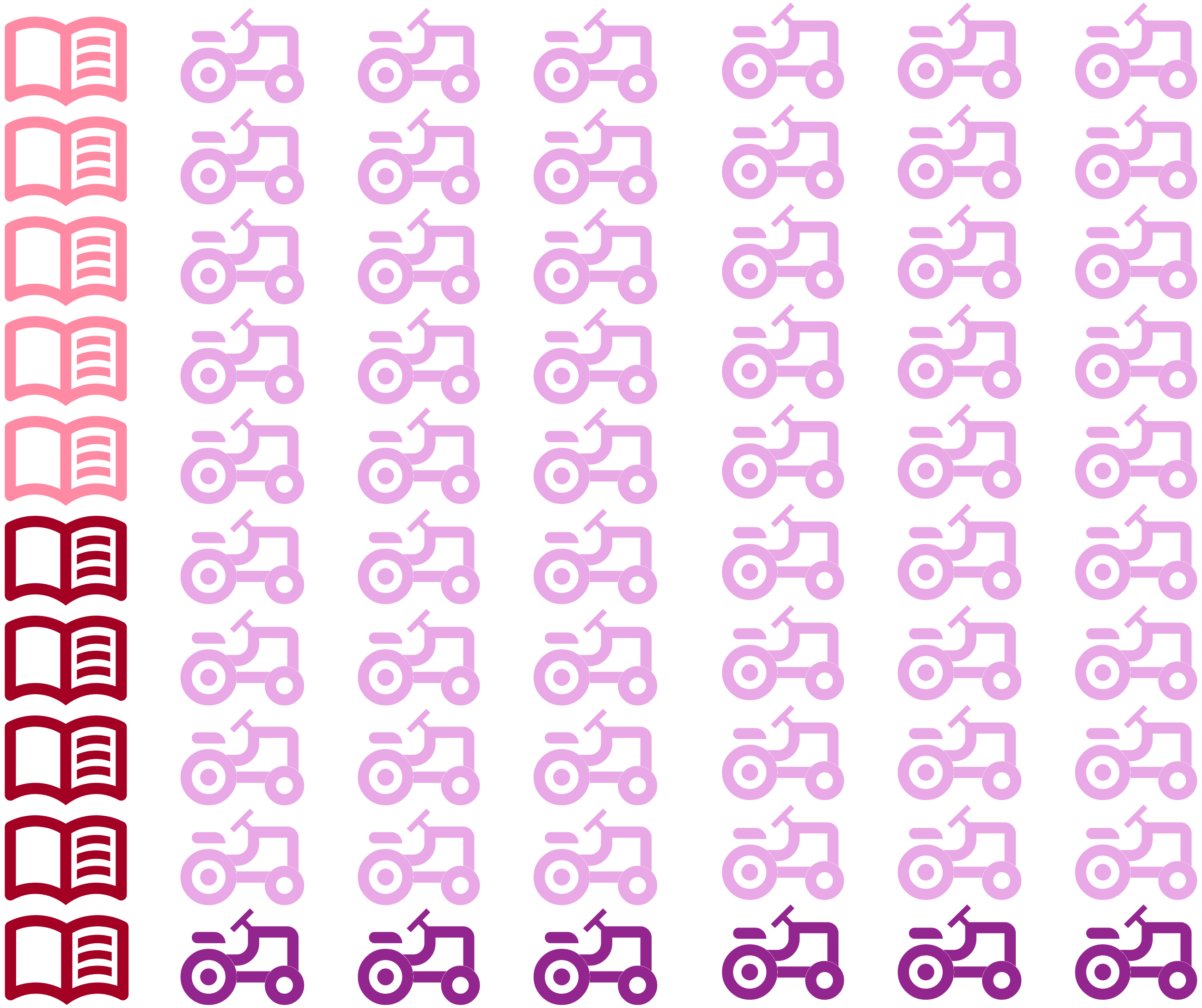
CORRESPONDS TO
600 FARMERS



John is very **shy and withdrawn**, invariably helpful but with very little interest in people or in the world of reality, A **meek and tidy soul**, he has a need for order and structure, and a passion for detail.



AI-Generated Image/Illustration



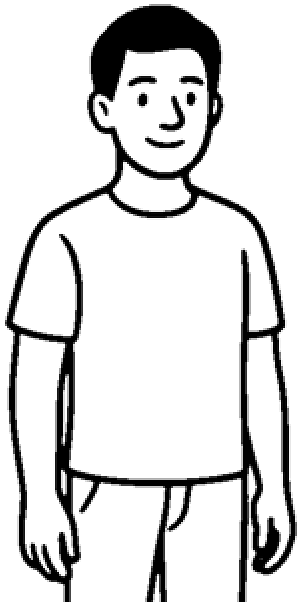
Likelihood of
(Librarian given Description)

$$= \frac{5}{5+60} = 7.6\%$$

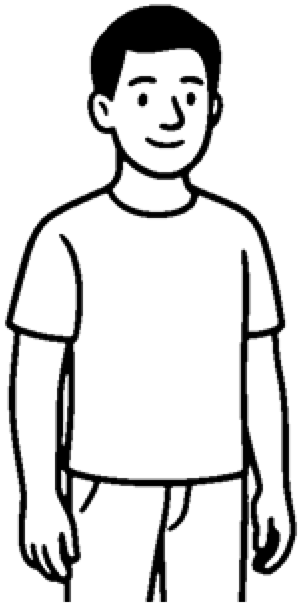
AI-Generated Image/Illustration

AI-Generated Image/Illustration

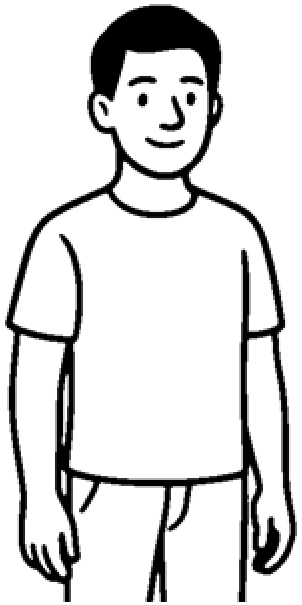
COLORADO



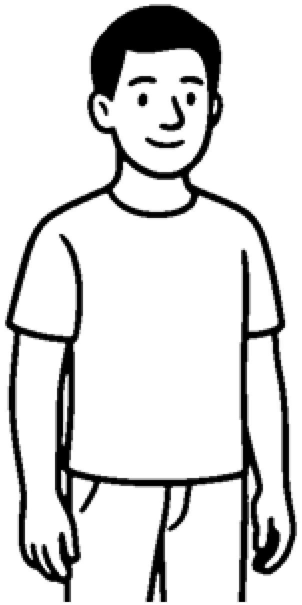
TEXAS



OHIO



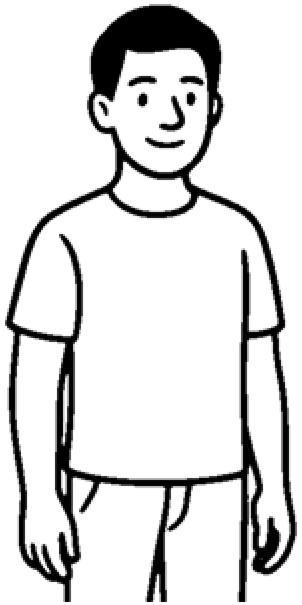
WISCONSIN



AI-Generated Image/Illustration

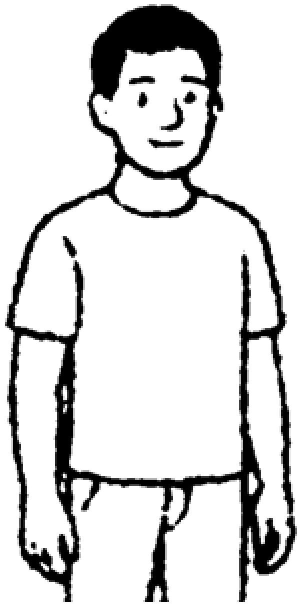
0 CLOSURES

COLORADO



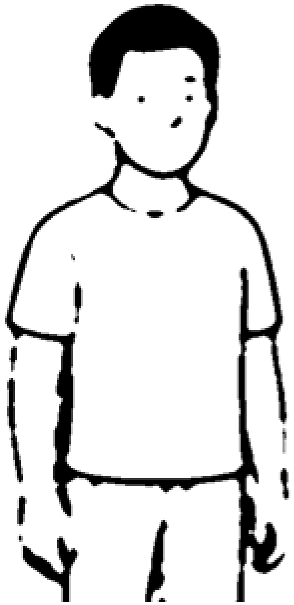
1 CLOSURE

TEXAS



4 CLOSURES

OHIO

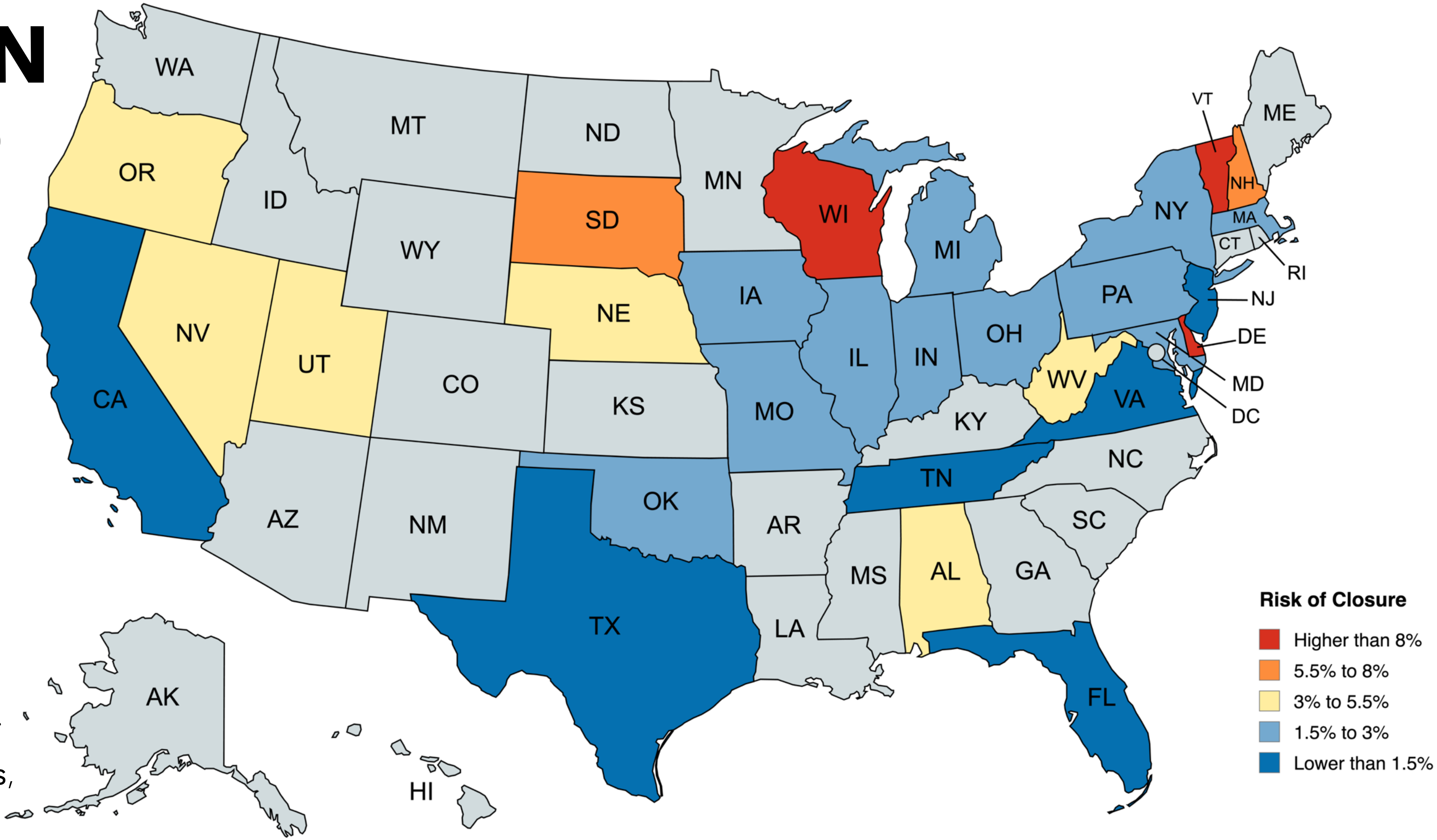


7 CLOSURES

WISCONSIN

AI-Generated Image/Illustration

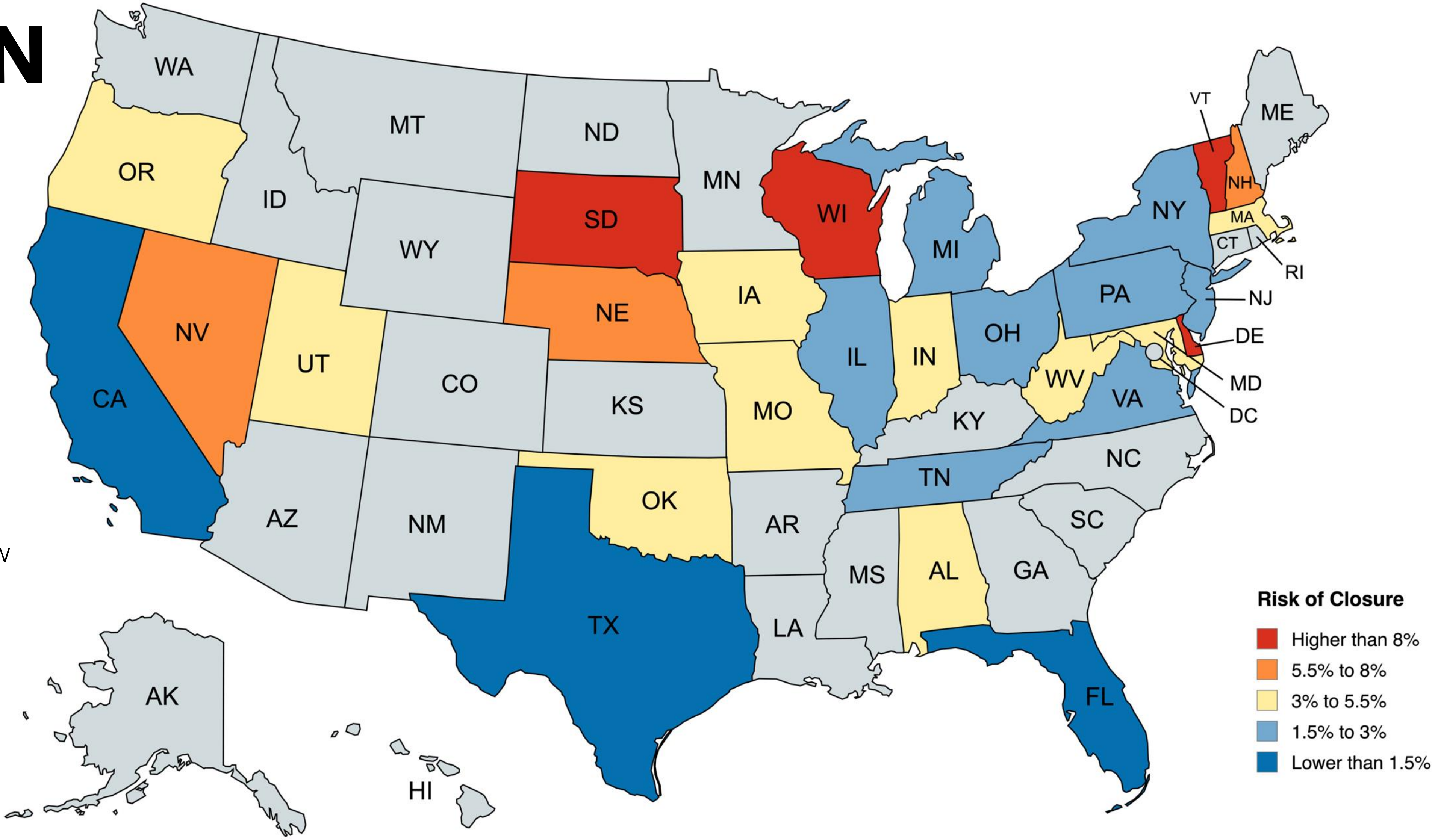
RISK OF CLOSURE IN OBSERVED STATES



- A college in **Wisconsin** has about **9%** risk of closure.
- In **Delaware** its **11%**; just because they were established in a certain state.
- Meanwhile a college in **Texas**, has a **0.5%** risk of closure.

RISK OF CLOSURE IN OBSERVED STATES

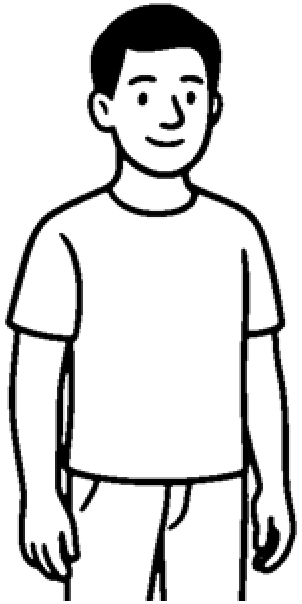
BASED ON CLAIMS



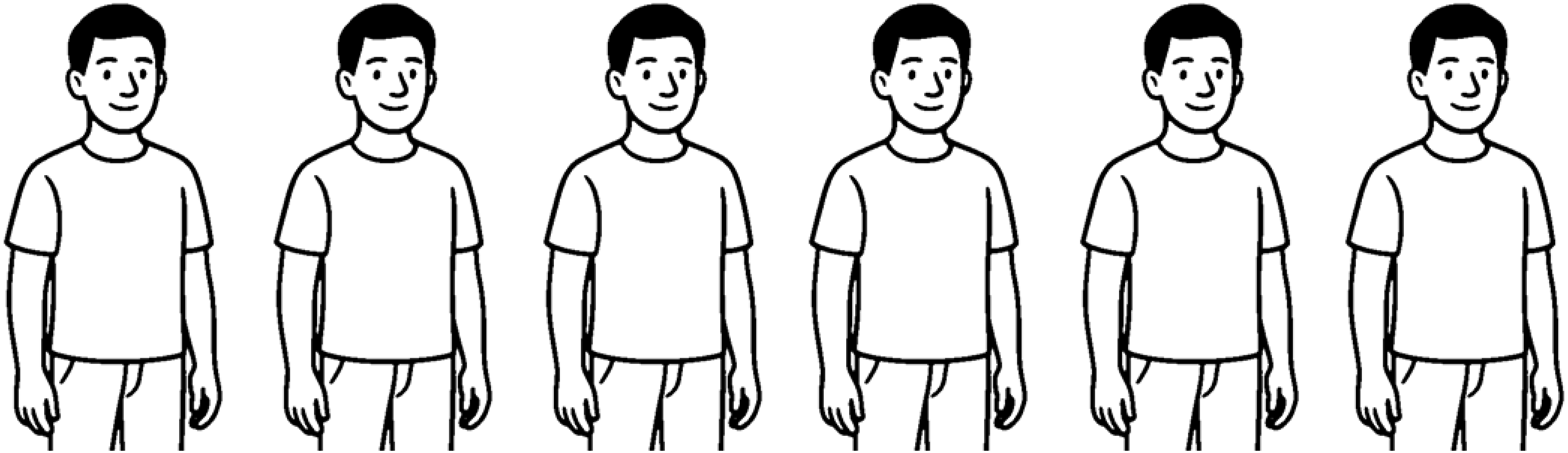
- A college in **Wisconsin** is now about **10%** at risk of closure.
- Now **Delaware** is **15%** at risk.
- While in **Texas**, the risk of closure has only risen by **0.03%**.

How to make a mature model?

JOHN



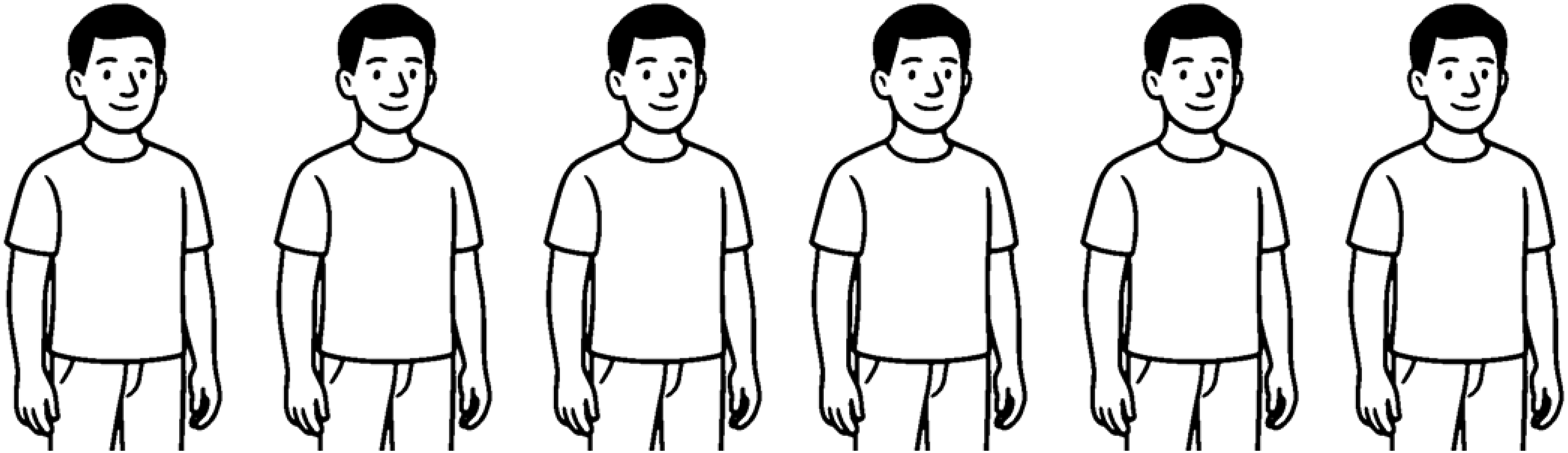
JOHN JOHN JOHN JOHN JOHN JOHN



AI-Generated Image/Illustration

TOWN A TOWN B TOWN C TOWN D

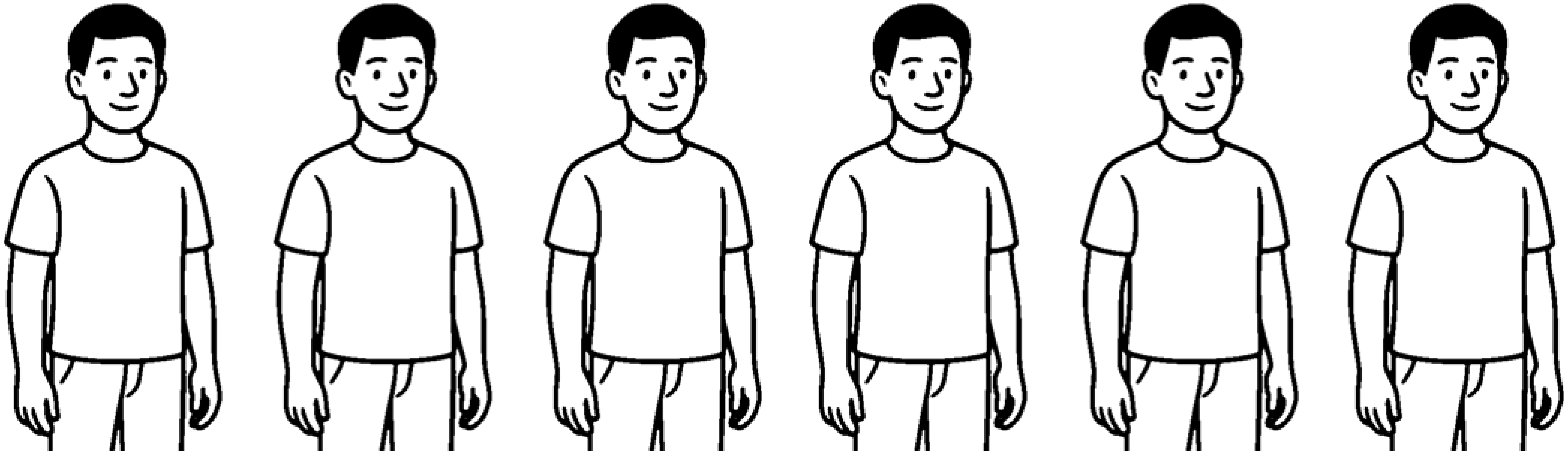
JOHN JOHN JOHN JOHN JOHN JOHN



AI-Generated Image/Illustration

23% Librarian	0.1% Librarian	65% Librarian	7% Librarian
TOWN A	TOWN B	TOWN C	TOWN D
21% Farmer	48% Farmer	0.4% Farmer	4% Farmer

JOHN JOHN JOHN JOHN JOHN JOHN



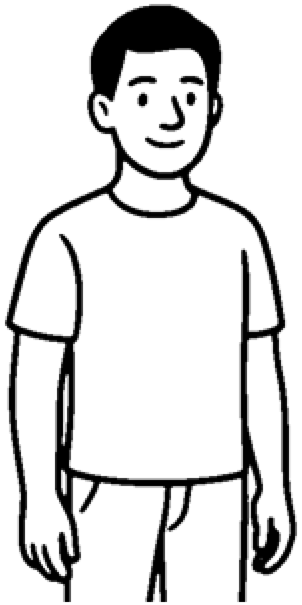
AI-Generated Image/Illustration

65% Librarian

TOWN A TOWN B TOWN C TOWN D

0.4% Farmer

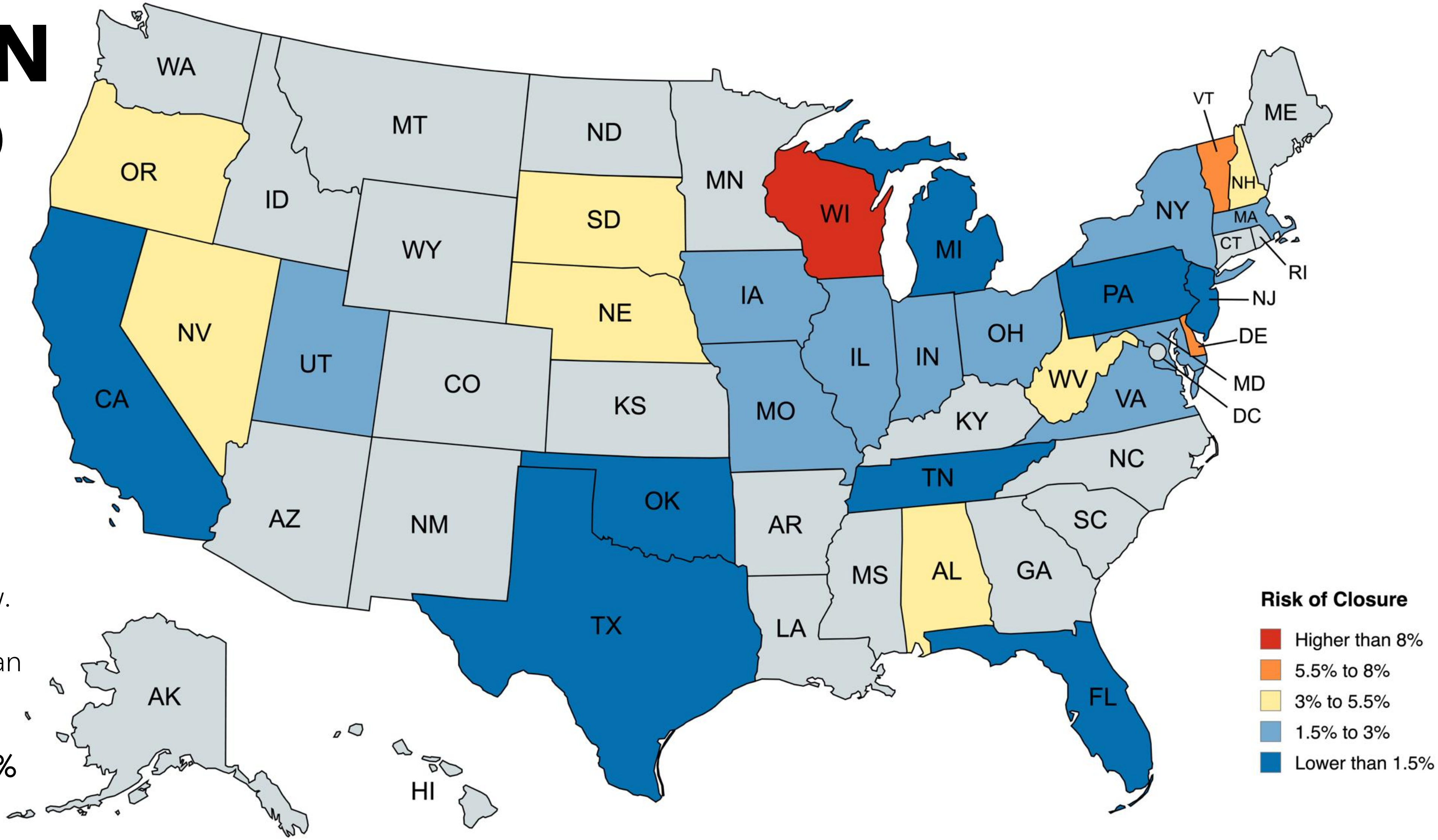
JOHN



AI-Generated Image/Illustration

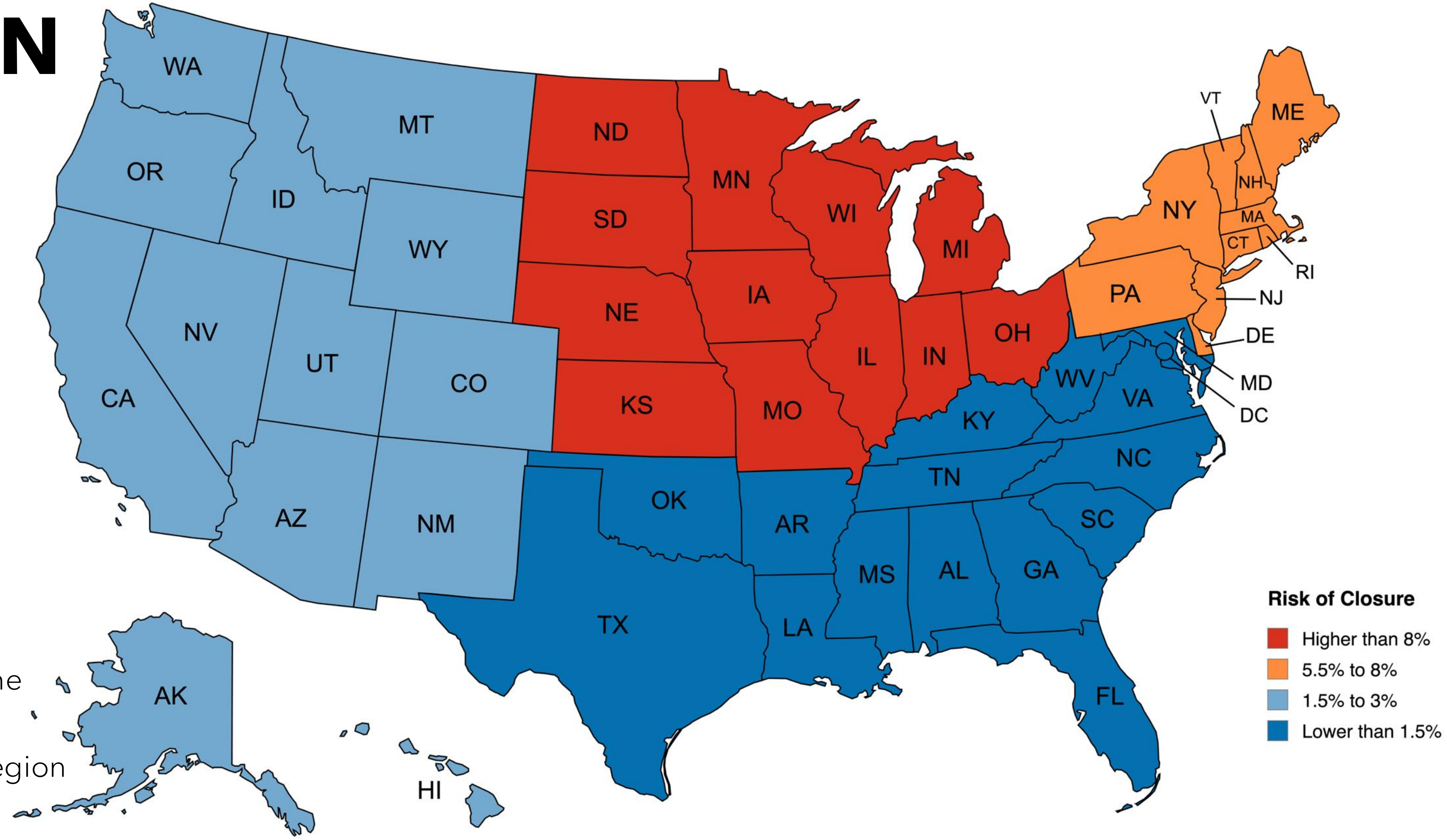
AI-Generated Image/Illustration

RISK OF CLOSURE IN OBSERVED STATES



- A college in **Wisconsin** has about **8%** risk of closure now.
- In **Delaware** its **7%**; lower than Wisconsin!
- A college in **Texas**, has a **0.3%** risk of closure.

RISK OF CLOSURE IN REGIONS



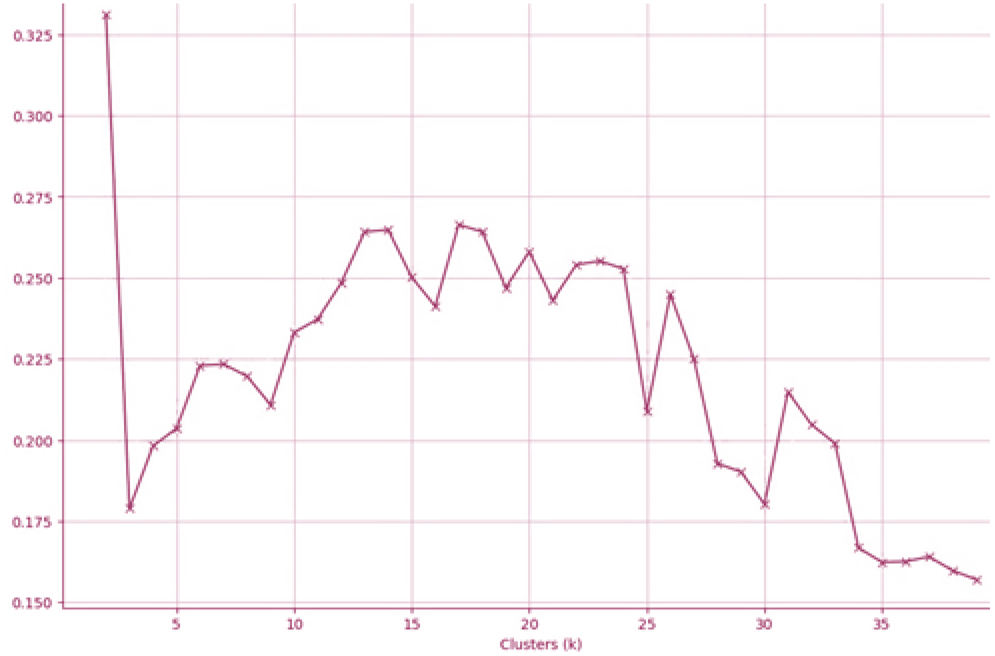
Although even 2%, the risk of closure in the Midwest, is not alarming, when compared to the south or west, there is a clear separation. Based on just the region the college was established in.

HOW DO WE FIND THE CHARACTERISTICS THAT PUT A COLLEGE AT RISK?

Note that : American University of Puerto Rico, closed in 2023, is excluded from this analysis.

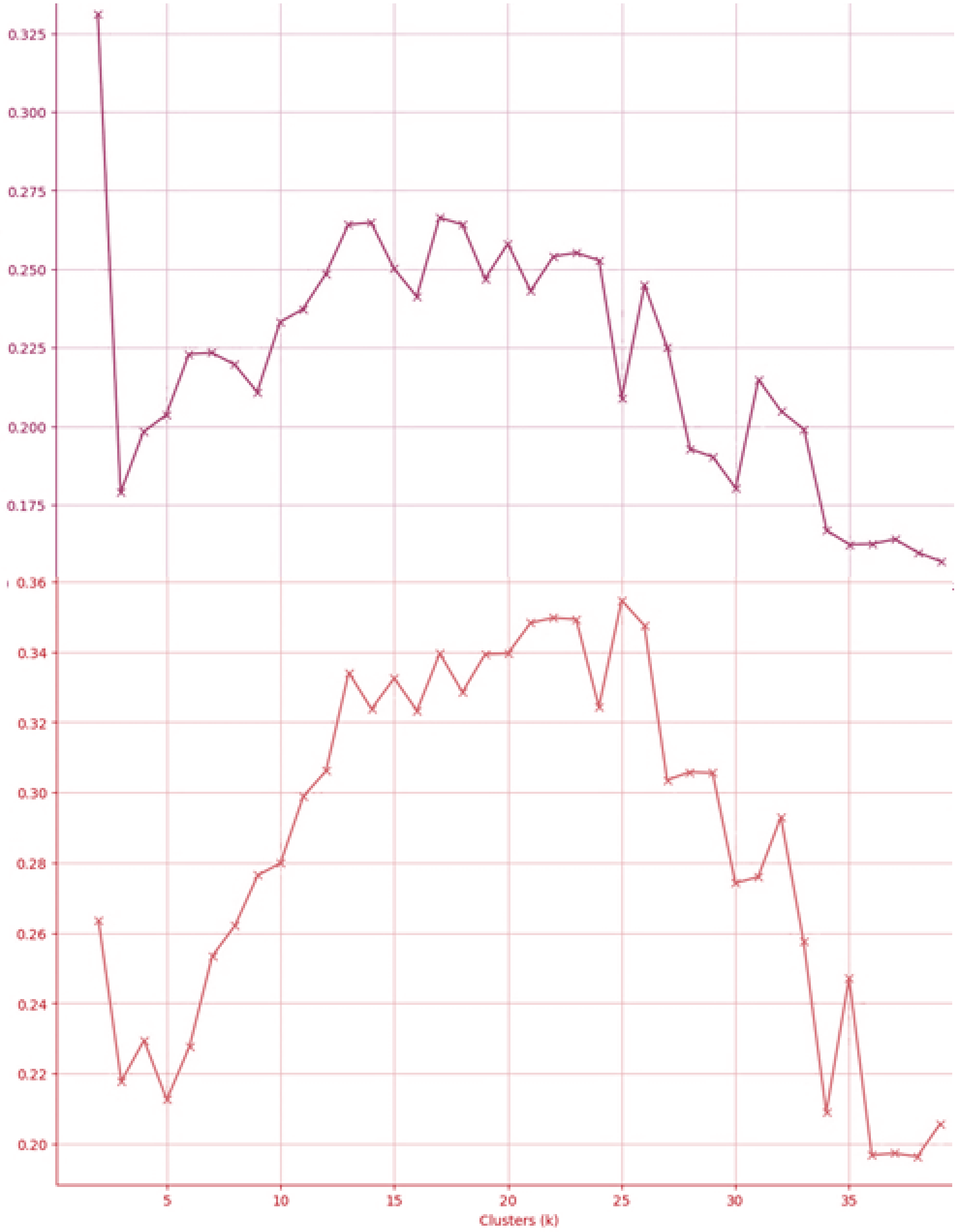
DATASET

2020-2025 CLOSED COLLEGES' CHARACTERISTICS AND INFORMATION + RELEVANT STATE- AND REGIONAL LEVEL DATA



THE CURSE OF DIMENSIONALITY

- The original dataset contained **65 observations** (colleges closed)
- It contained **178 variables** (features such as info and characteristics)
- The dataset's dimension was **reduced** to **66 variables**, excluding dependent and insignificant variables, and adding summary variables.
- The figure shows how the optimal grouping (clustering) results **improved**.



THE CURSE OF DIMENSIONALITY

- The original dataset contained **65 observations** (colleges closed)
- It contained **178 variables** (features such as info and characteristics)
- The dataset's dimension was **reduced** to **66 variables**, excluding dependent and insignificant variables, and adding summary variables.
- The figure shows how the optimal grouping (clustering) results **improved**.

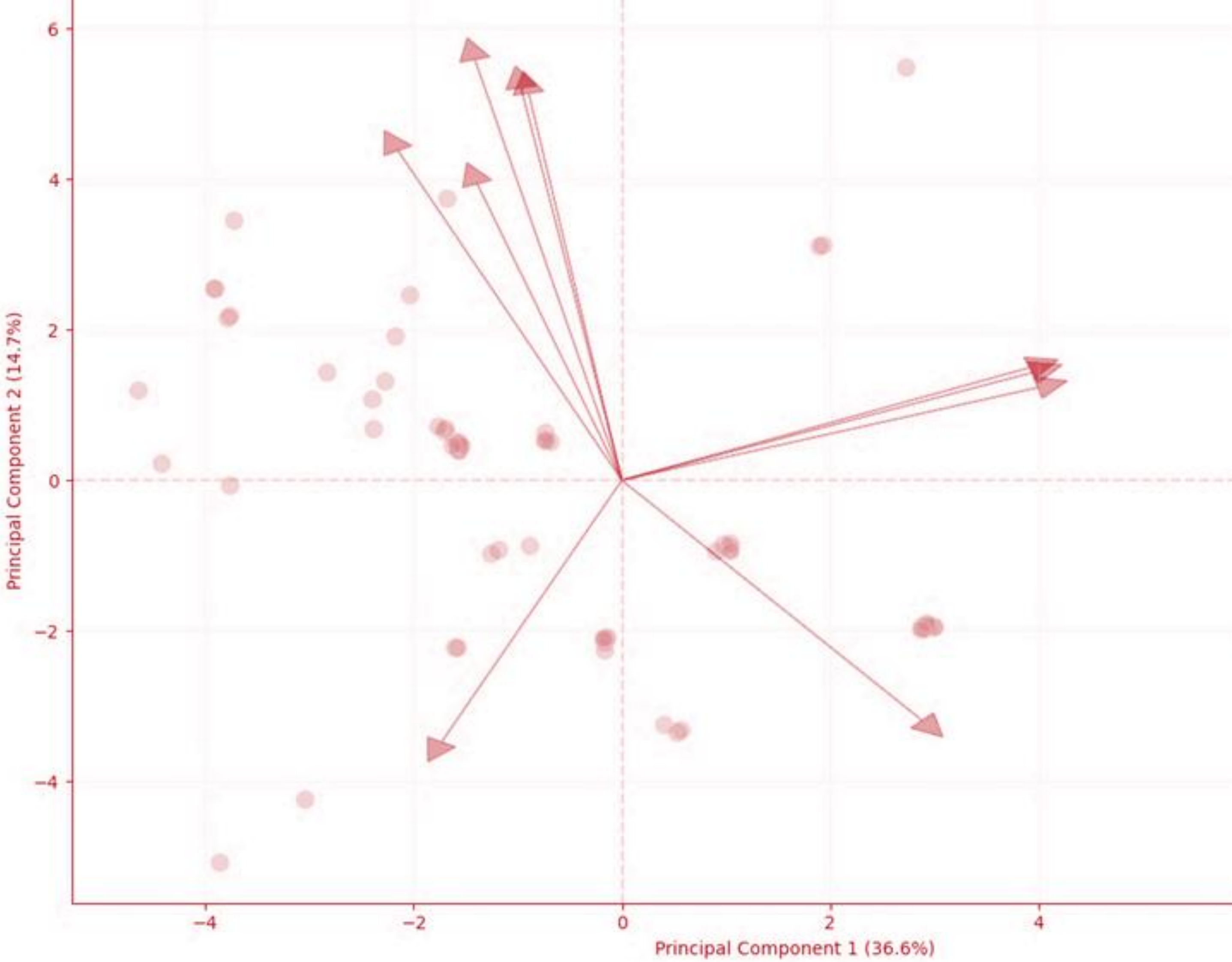
HOW DO WE IDENTIFY THE KEY DRIVERS?

2025: “USE MACHINE LEARNING.”

PRINCIPAL COMPONENT ANALYSIS

Some of the most **influential features** to the first two Principal Components (~51% of variance):

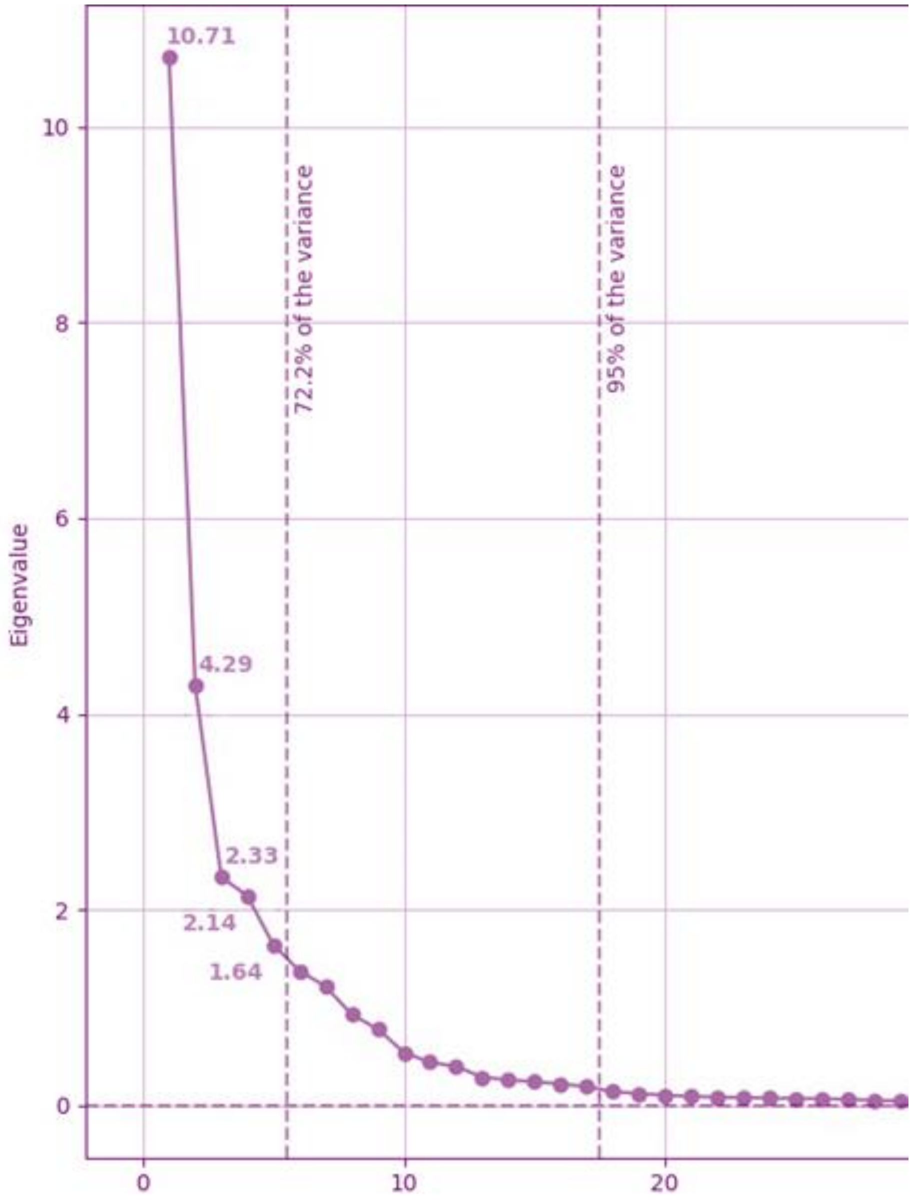
- Region's Birth Rates
- Per Capita Income in State
- State's Population of 18 to 24 yr.
- State's 2-year College's Tuition
- State's Population



PRINCIPAL COMPONENT ANALYSIS

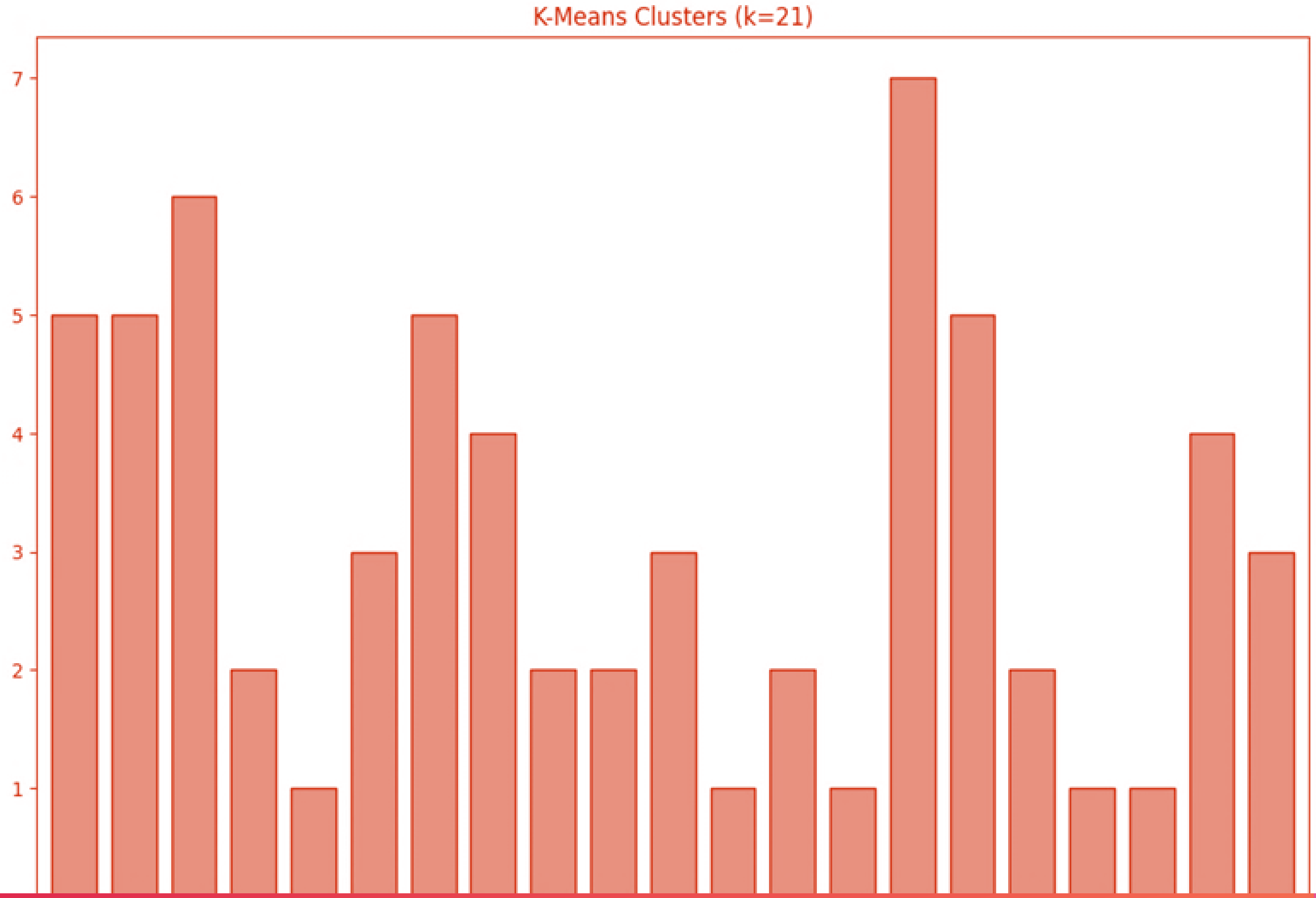
The presence of about 17 Principal with significant high eigenvalues point to the nuance that exists in the dataset; this means:

- The dataset created, and features are multi-faceted, as intended.
- There are many underlying patterns present in the dataset, which supports our hypotheses of college-level characteristics and state and regional data being influential.

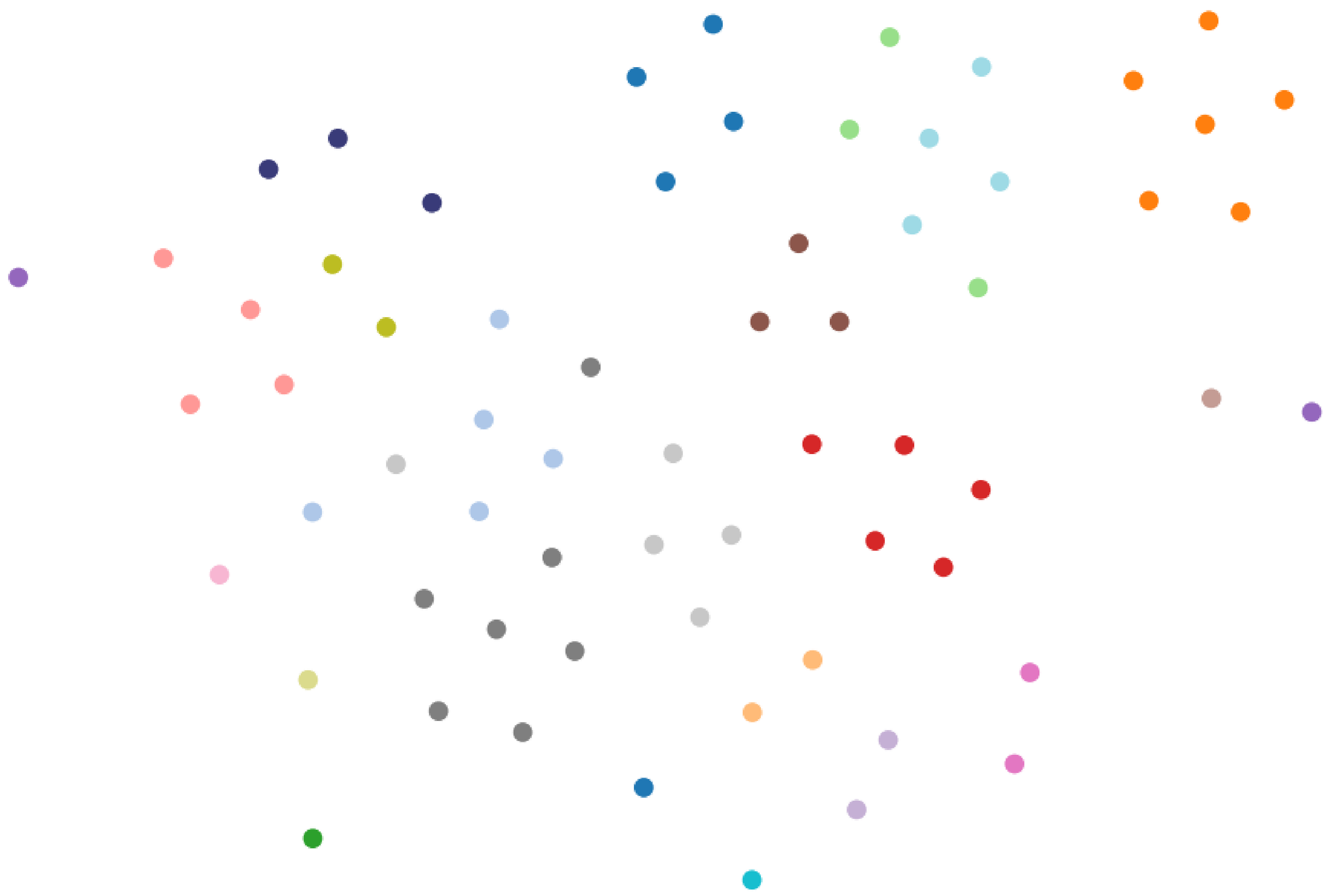


HOW SHOULD WE GROUP THESE INSTITUTIONS?

K-MEANS CLUSTERING

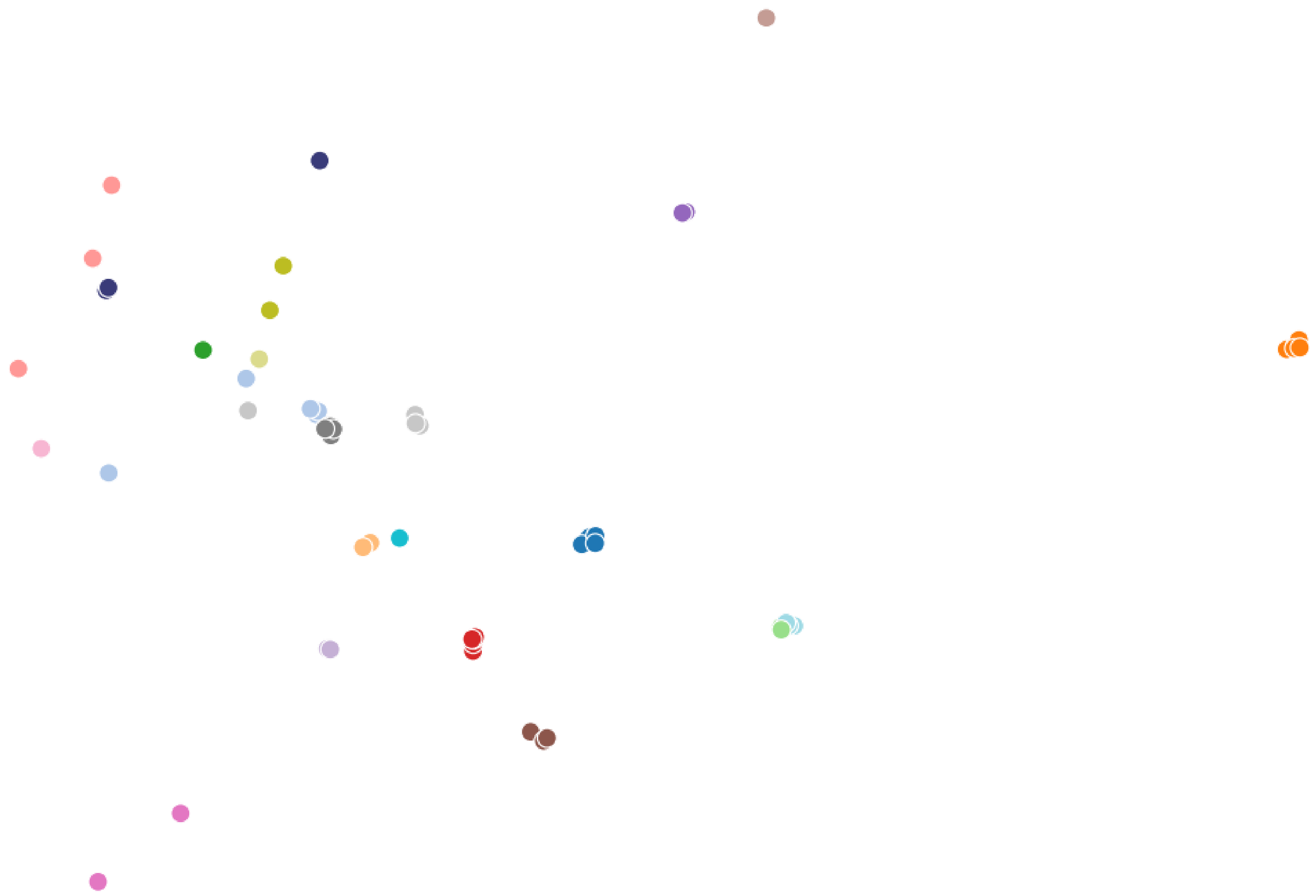


- We use the K-Means, a machine learning algorithm, to group institutions into 21 groups (clusters).
- The purpose of this approach is to see which colleges get grouped together without human intervention and solely based on data.
- We can then dissect each cluster and understand what variables the institutions are getting grouped by, which is a valuable insight for our goal.



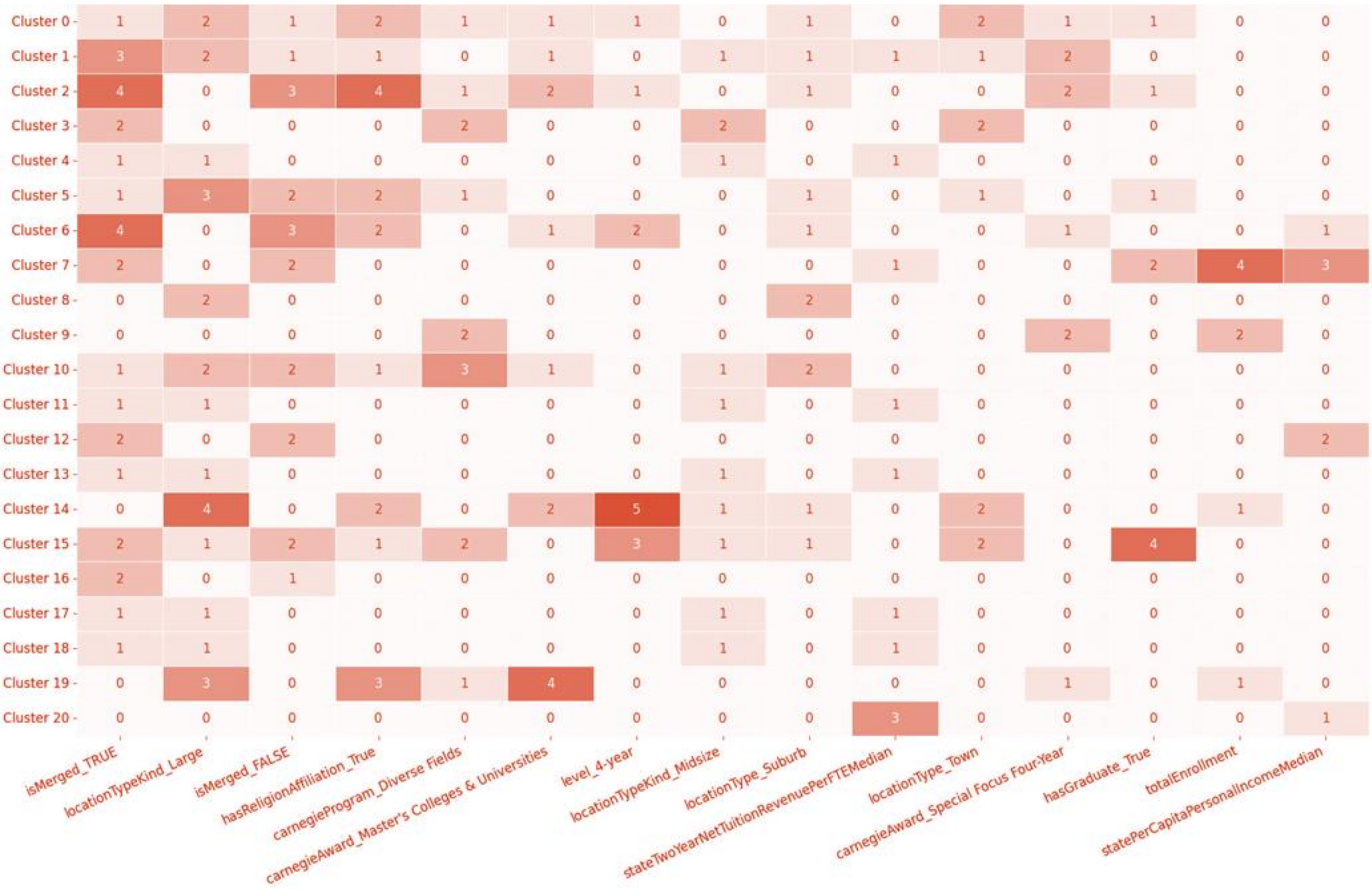
t-SNE VISUALIZATION

- t-distributed Stochastic Neighbor Embedding (t-SNE) is used for representing high dimensional data in lower dimensions.
- The figure proves that the data is not only placed in clusters successfully but is also separatable.
- A clean looking t-SNE like this one, supports the idea of there being strong features that different groups of colleges have in common.



K-MEANS CLUSTERING

- The figure shows 21 k-mean clusters, visualized in a 2D space using PCA.
- Each point is a different closed college, grouped with other colleges based on certain similarities.
- Tighter groups represent the colleges being more similar, looser groups indicate that either the colleges are not very similar, or PC1 and PC2 were not able to capture all the variance.
- Clusters far away from the rest indicate the colleges are distinct from the rest based on underlying features.



FEATURE FREQU-ENCY IN EACH CLUSTER

SHAPELY ADDITIVE EXPLENATIONS

- SHAP values explain, for a specific cluster, which features had the biggest positive or negative impact on colleges being assigned to that cluster.
- This helps us find the defining characteristics of each.
- We can also analyze what the key differentiating features are, and what features are present among all of them.

SHAPELY ADDITIVE EXPLENATIONS

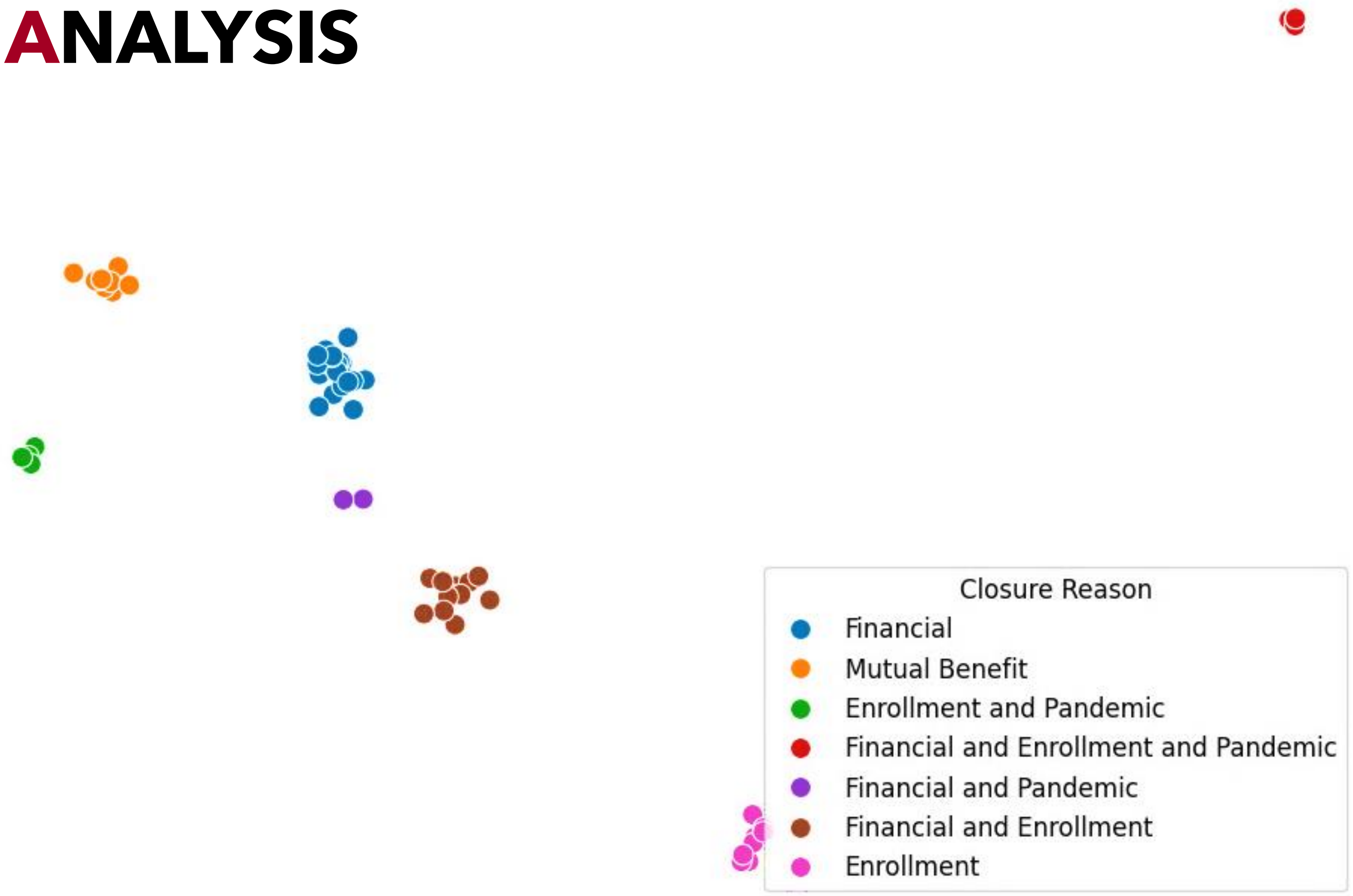
Some of the most influential features underlying the clustering are:

- Appropriation Amounts for 4-year Colleges in States
- Tuition Revenue of States
- Population of the Region
- High School Graduation Rate (ACGR)
- Appropriation Amounts for Community Colleges in States
- Unemployment Rate in the State
- Per Capita Income in State
- Median Endowment of Colleges
- Population of Individuals With Some College Education in State
- Fertility Rate in the Region

NO ONE SHOWED UP TO CLASS… WHY?

LINEAR DISCRIMINANT ANALYSIS

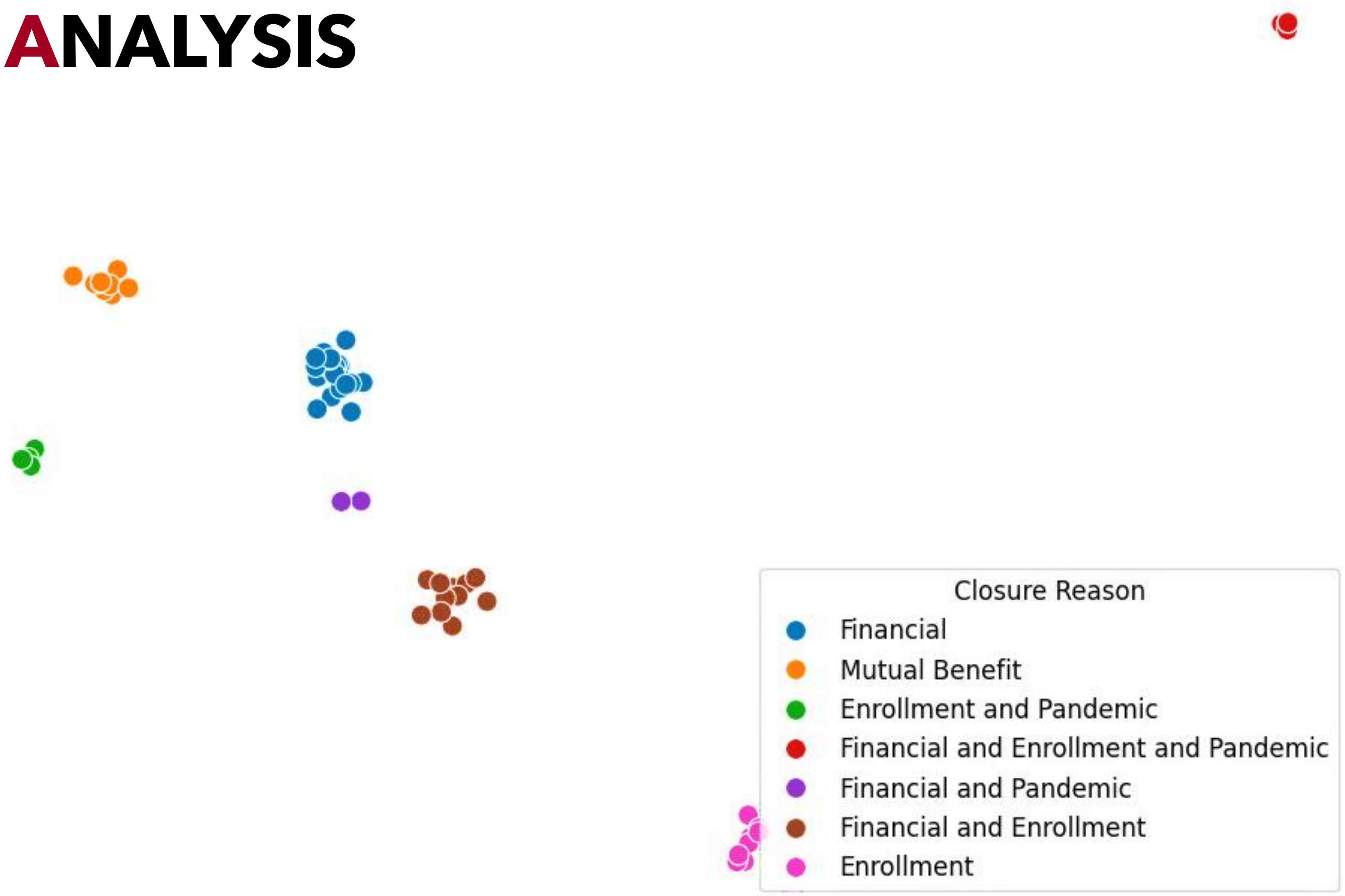
- LDA is a technique that reduces the dimensions of the data by finding the axes (components) that best separate distinct groups, aiming to maximize the distance between the clusters while minimizing the spread within each cluster.
- This figure visualizes in 2D space; since the first two components explain about 90% of the variance between these groups.
- The clear distinct clusters on the plot indicate that LDA was highly successful in separating the closure reasons. Suggesting these reasons for closure have very different underlying characteristics or patterns in the original data.



LINEAR DISCRIMINANT ANALYSIS

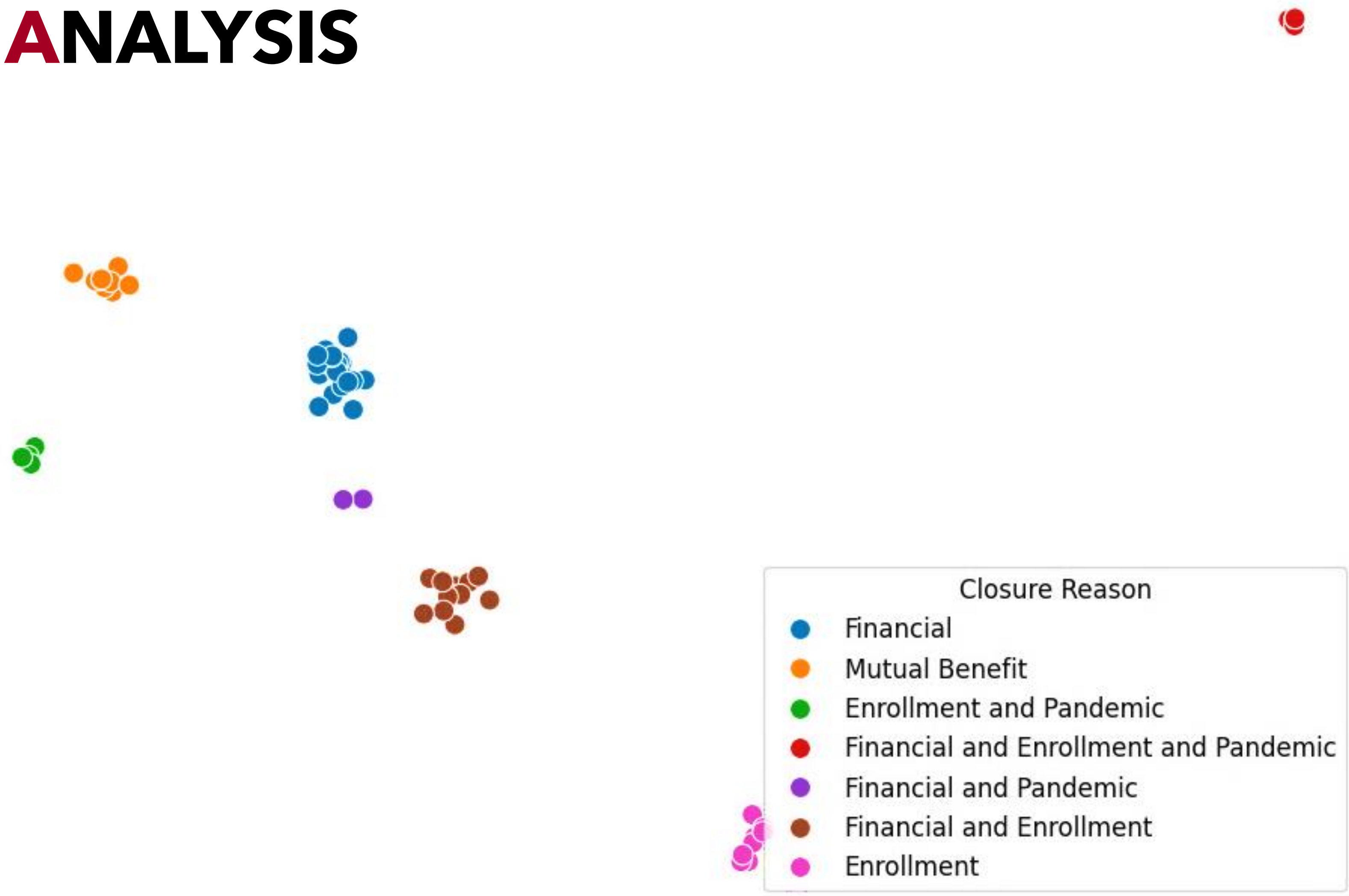
By analyzing the coefficients of our model and their magnitudes we see that:

- The first component shows a strong **negative association with state population**, while **positively correlating with more educated populations**.
- Additionally, programs with **high intensity** but perhaps **less demand** like seminaries, management schools, and doctoral-only colleges are positive drivers.
- Programs defined as **non-traditional** is also a huge driver in this component, bringing everything together well, with an indication that this component differentiates institutions based on **the demand of their academic context**.

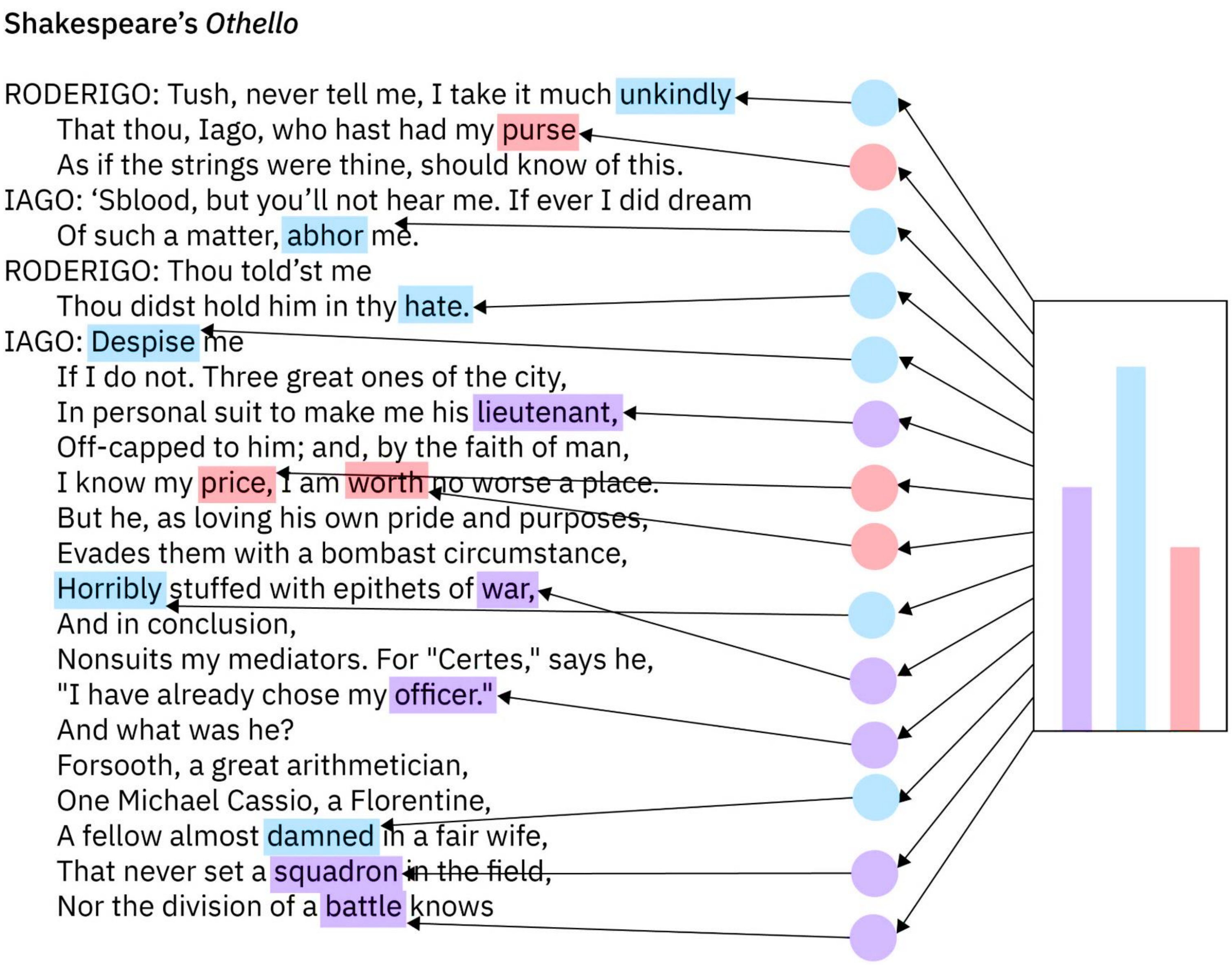


LINEAR DISCRIMINANT ANALYSIS

- The second component has a **strong positive relationship with state population**, but a **strong negative relationship with educated populations,** Even high school graduates.
- The component is also **negatively influenced by graduate programs** like doctoral and research schools, while having positive relationship with associate colleges.
- This indicates that the component gives higher values to institutions in populous states with with a lesser demand for higher education, or rather **a balance of population size and an uneducated population**.



BUT DOES THE "OFFICIAL REASON OF CLOSURE" REALLY CAPTURE ALL THE NUANCE? NOPE...



LATENT DIRICHLET ALLOCATION

- LDA analyzes text to identify hidden "topics" that run through the document.
- It does this by figuring out which words tend to appear together, if these co-occurring words belong to the same underlying topic.
- Simultaneously, for each topic, it estimates a mix of words that are most likely to belong to that topic.
- Each word is technically a "token" and tokenization, which is done before passed in the model, the process of cleaning and preparing words or tokens used to build the model.

Figure by IBM Research

LATENT DIRICHLET ALLOCATION

- Dataset curated included 1 article, mainly from Higher Ed Drive, per closure.
- Tokenization included removing redundant and irrelevant tokens, along with certain tokens that would not add any insights to our analysis.
- In 4 topics captured, ever single one has a word related to “financial” matters, and one related to some “religious” or faith affiliation.
- The figure shows a sample topic (Topic 2) of the LDA model.

THE CLIFF-PROOF INSTITUTION

IN CONCLUSION

UNDERSTANDING THE DEMAND SHIFT

RESTRUCTURING OF ACADEMIC PROGRAMS, FACILITIES, AND STUDENT LIFE

EDUCATION: AN ECONOMIC INVESTMENT

PRIVATE AND PUBLIC FUNDING OF EDUCATION, TO BUFFER INSTITUTIONS AGAINST DEPENDENCIES

A DEEPLY DATA-INFORMED INSTITUTION

A DATA-DRIVEN STRUCTURE FOR ACADEMIC PROGRAMMING, BUDGETING, AND ENROLLMENT CAMPAIGNS

A CHANGING ETHNIC DEMOGRAPHIC

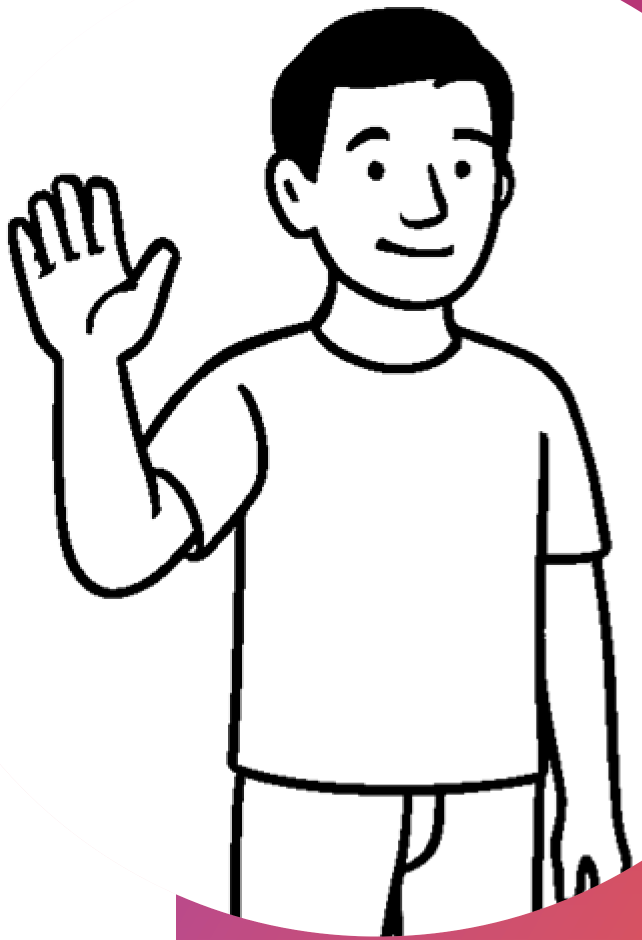
FURTHER RESEARCH DIRECTION

THANK YOU!

Many thanks to the **James Rocco Program** that allowed for this investigation to flourish. Along with **Professor Taylor** for coordinating the program, **Dr. Jamshidi** for supervising the project, and **Bita Aslrousta** for her assistance in curating the project's dataset.

QUESTIONS?

akbaris79@lakeforest.edu



AI-Generated Image/Illustration

SHAPELY ADDITIVE EXPLENATIONS