American College Influence—A Geographical Perspective

Craig E. Wills
Computer Science Department
Worcester Polytechnic Institute
WPI-CS-TR-23-03

May 2023

Abstract

This work employs a geographical perspective to investigate the influence of American colleges. Many aspects of colleges are examined including influence exerted for student enrollment, where their alumni reside, research impact and sports fandom.

We apply a gravity-based model for the determination of spheres of influence for each of these aspects. While this model has typically been applied using one or more metrics of urban markets, in this work we treat colleges and universities as the markets. We model their spheres of influence using an appropriate metric for the weight of attraction of each market. The model accounts for multiple markets in the same vicinity as well as the friction of distance on influence as we consider locations further away from a market. We also account for other considerations such as how to model the influence of public versus private institutions, institutions offering a more specialized set of degrees and the impact of crossing state borders on institution influence.

Not only do we model influence, but we generate U.S. maps showing spheres of influence by displaying a school color of the institution exerting primary influence for each county. In addition, we generate interesting and realistic maps reflecting the intensity of influence as well as a blending of colors for institutions exerting influence in each county.

The map for enrollment shows that regional public institutions generally exert the primary enrollment influence in respective regions of their state with major public institutions in states exerting secondary influence over their entire state. Maps for college football and men's college basketball show rooting areas of college fans. Use of a model allows us to identify schools with more or less fan influence than expected.

Interactive versions of all maps generated and shown in this report can be viewed with a Web browser at `https://web.cs.wpi.edu/%7ecew/collegeinfluencemaps/`. Three types of maps are available. Maps show counties in: 1) color of primary institution exerting influence; 2) color of primary institution reflecting intensity of influence; and 3) a blending of colors for three institutions exerting the most influence.

1 Introduction

Colleges certainly influence American society. This work seeks to understand this influence by applying a gravity-based model for the sphere of influence of college markets. Our work not only develops the model based on actual data, but uses the data to analyze its validity. We initially model geographical influence on enrollment of college students, but apply the model to other aspects of college influence including alumni, research impact and sports fandom. In each case, we generate U.S. maps with each county displayed in school colors corresponding to the college or university exerting the most influence over that county. Based on the developed model, we are able to identify institutions that exert more or less influence than predicted.

Determination of boundaries between urban areas has long been of interest to geographers such as the determination of the boundary between New York City and Boston based on a variety of functional indicators [9]. Identification of urban spheres of interest was done to create planning regions on a national scale based on the spheres of influence model [13], which is a gravity-based model where influence of urban markets decays with distance [14]. Boundaries are created between these markets where influence from two competing urban areas are equivalent. The same model was applied using newspaper circulation data in the U.S. [2] and urban center factors, for example with population as the weight of attraction, in the Republic of Ireland [15]. More recently, this sphere of influence model has been applied to social media data [42].

While this model has typically been applied using one or more metrics of urban markets, in this work we treat colleges as the markets and seek to model their spheres of influence using an appropriate metric for the weight of attraction. In doing so, our work makes contributions from different perspectives:

- **Geography.** We apply the well-known spheres of influence model to a new domain where colleges and universities are the markets exerting influence. We adapt the model to account for when multiple markets are in close proximity to each other and determine model parameters based on ground truth data. We consider additional factors affecting the influence of institutions and are able to use the known data to validate the predictive success of the model.

- **Higher education.** The work measures and visualizes the influence of American colleges across the aspects of enrollment, their alumni, research impact and sports fandom. We combine these separate results to identify regions in the U.S. where a single institution exerts primary influence for all aspects and regions in which primary influence varies among institutions.

- **Sports.** Fans exhibit passionate rooting interests for college sports. Our work unifies disparate data sources for football and men's basketball to model spheres of influence for fan rooting interests of these sports. Use of a model allows us to identify schools with more or less influence than expected.

- **Map visualization.** We generate colorful and enlightening sphere of influence maps for each aspect studied in our work. These visualizations show school colors for the primary institution exerting influence in each county. In addition, we generate two other types of maps showing the intensity of influence as well as a blending of the school colors in proportion to the intensity of their influence. Maps are included as images in this report and available in interactive form at an accompanying website.

In the remainder of this report we pose a set of research questions in Section 2 and describe the model approach we take to answer these questions in Section 3. Section 4 develops a base model for the sphere of influence of college enrollment and Section 5 further tunes the model by considering other factors. We validate the tuned model with actual data in Section 6. Sections 7, 8, 9 and 10 describes how the model is applied to model other aspects of alumni, research, football and men's college basketball influence by colleges and universities. Section 11 compares these different types of influences by identifying regions where a single institution exerts primary influence for all aspects and regions where this influence varies among institutions. We conclude the report in Section 12 with a summary of our results and directions for future work.

2 Research Questions

The motivation of our work is to understand if and how well a model of geographical sphere of influence can be applied to American colleges and universities. We look to examine this influence for a variety of college aspects. This motivation leads to a number of specific research questions that drive our work. These questions are enumerated below.

1. **What parameters are best to use in applying the geographical sphere of influence model for college enrollment?** These parameters include the weight of attraction as well as a distance decay coefficient. Influence results need to be observed for both major (e.g. flagship, land grant) as well as regional public institutions.

2. **What is the impact of other considerations on college market influence?** For example, how does the influence of private compare with public institutions? Does it make a difference if the degrees offered by an institution are specialized or comprehensive? Another consideration is the extent of enrollment influence across state borders.

3. **Once modeled, how well do the predicted college enrollment results match actual data?** With a model in place, the correctness relative to available data can be made. It is also possible to identify specific institutions that are mis-represented in terms of their geographical sphere of influence.

4. **Can geographical influence be modeled for other aspects of colleges? For example, how well can the influence of alumni be modeled and shown?** Understanding the influence of an institution via where its alumni move after leaving is of interest.

5. **Can the research influence of colleges be modeled?** Colleges and universities are engines of research and innovation, which influences their geographic regions. Being able to model the exertion of this influence is of interest.

6. **Major college sports have a significant influence on American society in terms of fan rooting interest. How well can the influence of institutions be modeled for college football?** Can the resulting model be used to identify schools for which the exerted influence is stronger (or weaker) than would be expected?

7. **Similarly, how well can the influence of institutions for men's college basketball be modeled?** Again, can the resulting model be used to identify schools for which the exerted influence is stronger (or weaker) than would be expected?

8. **Can the sphere of influence results be compared for institutions across the many aspects studied in the work?** This research has the potential to identify regions of the country where a single institution exerts primary influence for many aspects as well as regions where the institutions exerting influence vary.

3 Approach

In this section we describe the approach used to answer these questions. We describe the geographical gravity model for the sphere of influence of a market as well as data employed for applying the model.

3.1 Sphere of Influence Model

As presented in works such as [2], the sphere of influence model for markets on surrounding areas can be modeled based upon a combination of the weight of attraction for each market as well as the distance from the market. This parameterized model can be expressed as shown in Equation 1

$$I_{ki} = \frac{A_k}{D_{ki}^{\lambda}} \tag{1}$$

where I_{ki} is the intensity of influence exerted by market k at point i; A_k is the weight of attraction for the market k; D_{ki} is the distance between market k and point i; and λ is the distance decay coefficient indicating the friction of distance. For example, a coefficient of $\lambda = 2$ results in the amount of influence at any point computed based on squaring the distance between the point and market.

This model has typically been applied to urban markets where the weight for the market corresponds to a metric such as its population. However in our work we treat colleges and universities as the markets and seek to model their areas of influence using an appropriate metric for the weight of each market.

In applying the model, the location of a college market is represented using the latitude/longitude coordinates of its campus. Rather than model at the level of points, we model at the U.S. county level (or more specifically at FIPS level where a each county is represented as a single latitude/longitude coordinate within the county[1]. The distance between a market and county is computed based on the difference between their latitude/longitude coordinates. Working at the county granularity for "points" allows us to map the sphere of influence results for college markets by determining the college exerting the most influence over each county in the U.S.

[1]We obtained data for counties from the U.S. Census Bureau [35], which assigns a FIPS (Federal Information Processing Standards) code to each county. The data set of 3143 FIPS codes is primarily of counties (and we describe it as such), but does include a few cities. Conveniently, FIPS codes are also used by the d3.js (`d3js.org`) JavaScript library to visualize a county. As an added benefit, all created visualizations are Web documents both viewable and interactive via a browser.

3.2 College Enrollment Data by State

We initially model the influence of institutions for the enrollment of students. The weight we use for each market is the number of students from the state attending each four-year institution. These data were obtained from the publicly-available Integrated Postsecondary Data System (IPEDS) [22]. In particular, we make use of IPEDS data of each institution with the number of "first-time degree/certificate-seeking undergraduate students" from each state and the District of Columbia. These data are expected to be reported in even-numbered years and we use data from the Fall'20 enrollments gathered as part of previous work [41]. We use these per-state residence counts as the weight of attraction A_k for each institution k enrolling students from a state. We consider roughly 1600 (560 public) 4-year institutions in our work.

4 Enrollment Influence Modeling

The first research question identified in Section 2 seeks to understand how college enrollment can be modeled using geographical sphere of influence, which requires determination of model parameters. Using college enrollment within a state as the weight of attraction for an institution, the primary parameter to determine is the value for the distance decay coefficient λ. However, there is another consideration due to our treatment of institutions as the markets in that multiple markets can and do exist within the same vicinity. This distinction causes us to also determine another parameter for the minimum distance used in the computation of market influence. The parameter ensures that the same distance is associated with points close to more than one institution.

4.1 Enrollment Ground Truth Data

In determining the minimum distance and distance decay coefficient parameters, we were able to obtain per-county enrollment data at state colleges and universities for a number of states. These ground truth data were obtained for Alabama [1], Kansas [16], Louisiana [18], Michigan [19], Minnesota [20], Missouri [21], North Carolina [24], North Dakota [27], Oklahoma [29], South Carolina [32], Texas [34], Utah [37] and Virginia [38]. These data are not always for the same timeframe as the modeled data and are typically only for state public institutions, but the data do afford a means to empirically determine appropriate parameter values. They also allow validation of the resulting spheres of influence for the parameterized model when applied to these states. In determining model parameters we initially focus on the 560 public institutions as more ground truth data are available. The influence of private institutions are subsequently examined in Section 5.

4.2 Determination of Model Parameters

We first determine the minimum distance parameter to use in applying the model. This value is the minimum distance used between coordinates of a county and a college market. If the computed distance is less than this value then the minimum is used.

In determining this minimum distance to use in the model, we use a distance decay coefficient of $\lambda = 2.0$. This value is representative of what we expect for this coefficient. We determine the college market exerting the most influence for each of the 3143 counties using minimum distance values of 10, 15, 20 and 25 miles.

For 92% of the counties we found the same college market exerts the most influence regardless of the minimum distance used. However for the remaining counties in which there is a difference, we use the ground truth data available from the subset of states to determine which minimum distance value results in the best correspondence between actual and predicted market of influence. The results are shown in Table 1.

Table 1: Correspondence of Predicted and Actual Market Influence Based on Minimum Distance Values with $\lambda = 2.0$

Min. Distance (mi)	Pct. Correspondence
10	60%
15	62%
20	66%
25	38%

The results show that a minimum distance of 10, 15 or 20 miles results in similar results with a minimum distance of 25 miles resulting in much worse correspondence between the predicted and actual market exerting the most influence. Based on these results we use a minimum distance of 20 miles as it has the best performance in the range tested. This value is used for all market influence calculations in our work.

We next determine the distance decay coefficient λ. This parameter determines how quickly the influence exerted by a market drops off with distance. Historically, this coefficient has taken on different values depending on the application of the model. For example, Huff [12] noted the coefficient having a value ranging from 1.5 to over 3.0 depending on different types of shopping trips. The same work determined coefficient values of 3.2 and 2.7 for clothing and furniture shopping trips. Work to create planning regions on a national scaled used a coefficient of 2.0 with a two-tiered classification of urban regions [13]. A value of 1.45 was found to be the best fit for newspaper circulation in the U.S. [2]. A study on identifying urban sphere of influence in central China used a distance decay coefficient of 2.0 [4].

Based on this prior work, we consider a range of seven possible values for λ of 1.5, 1.75, 2.0, 2.25, 2.5, 2.75 and 3.0. In each case we use a minimum distance of 20 miles for determination of this coefficient.

We found for 72% of the 3143 counties that the same college market exerts the most influence regardless of the coefficient value used. However for the remaining counties in which there was a difference, we again use the ground truth data described in Section 4.1 to determine which coefficient results in the best correspondence between actual and predicted market of influence. The results are shown in Table 2.

The results show that a distance decay coefficient of 1.5 or 1.75 results in the worst correspondence between predicted and actual market influence. Distance decay coefficients of 2.0 and 2.25 provide a moderate level of correspondence with the remaining coefficients providing the best correspondence. Based on these results we adopt a value of $\lambda = 2.5$ for all market influence calculations in our work as it provides the best correspondence with better performance than values that are higher or lower.

Table 2: Correspondence of Predicted and Actual Market Influence Based on Distance Decay Coefficient (λ) Values with Minimum Distance of 20 Miles

Value of λ	Pct. Correspondence
1.5	33%
1.75	50%
2.0	58%
2.25	63%
2.5	68%
2.75	65%
3.0	66%

We use these base model parameter values (minimum distance of 20 miles and distance decay coefficient of 2.5) to create a map of market influence for public college enrollment. It is shown in Figure 1.

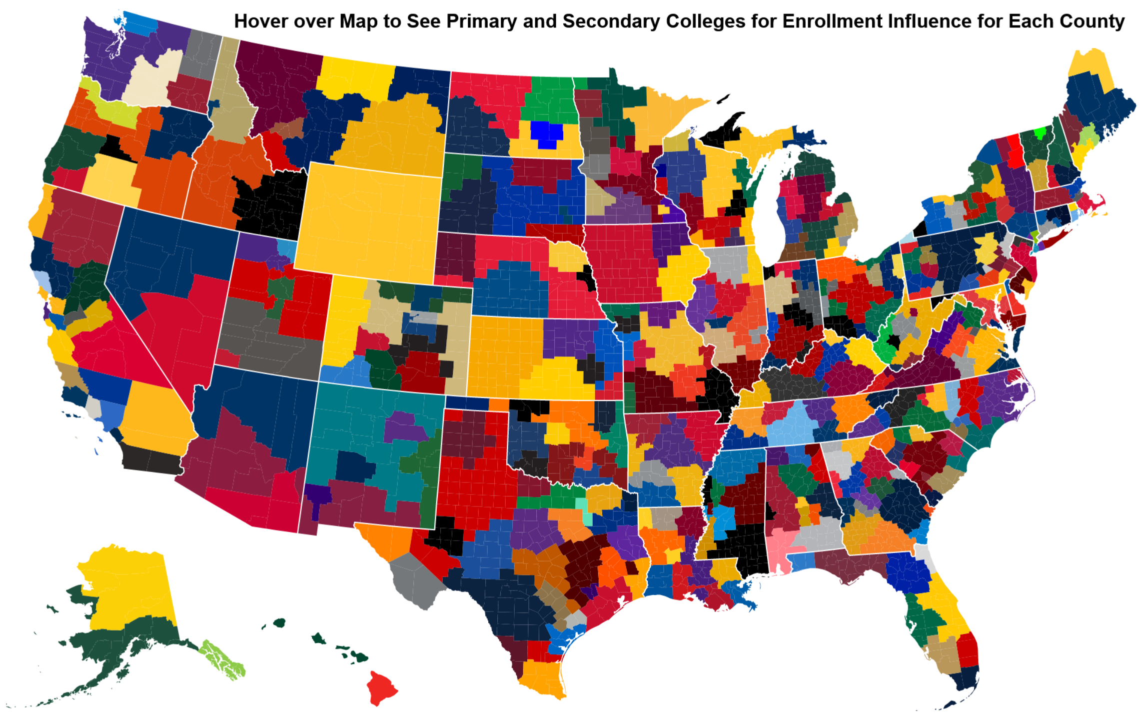

Figure 1: Map of College Enrollment Influence by Public Institutions Using Base Prediction Model

This colorful map shows each U.S. county using school colors corresponding to the public college or university exerting the most influence and thus the institution predicted to enroll the most students from that county. It shows the U. of Wyoming as the only public institution in the state while the state of Nevada is split between two primary institutions (U. of Nevada and UNLV) for college enrollment. In contrast, most states exhibit a mosaic of color as major and regional public institutions exert primary market influences in their region of a state. In these states, major public institutions may enroll the most students of any public institutions, but the distance decay coefficient is relatively high. This distance decay causes the enrollment influence of major public institutions to be overcome by the enrollment influence of regional public institutions in their specific regions of their states.

Table 3 shows the top-10 public institutions for enrollment influence over the largest number of counties using our base prediction model. Note we use shortened institution names with no spaces in tables and the interactive maps to minimize the amount of text while still ensuring unique names for each school. Counties exhibit much variation in size and population across the U.S, but make for a simple means of comparison. These results highlight public institutions with significant predicted enrollment influence. The table shows Iowa State exerting primary enrollment influence over the most number of counties and the state of Nebraska with two institutions in the top 10—the U. of Nebraska and U. of Nebraska-Kearney.

Table 3: Top-10 Public Institutions Exerting Enrollment Influence Over the Largest Number of Counties Using Base Model

State	Institution	Cnt. Counties Exerting Enrollment Influence
IA	IowaSt	61
TX	TexasTech	48
KS	FortHaysSt	41
NE	Nebraska	40
VA	VirginiaTech	33
SD	SouthDakotaSt	31
NE	UNKearney	29
MO	Missouri	29
KY	WesternKentucky	26
GA	GeorgiaSouthern	26

5 Other Considerations for College Market Influence

Our initial map shown in Figure 1 only considers student enrollment to public institutions within the same state. However, there are other factors we investigated that may affect the enrollment influence of colleges. These factors are each considered in the remainder of this section. They include:

- whether an institution is comprehensive in the degrees it offers or is focused in particular areas such as STEM;
- if private institutions have a similar enrollment market influence as public institutions; and
- whether out-of-state institutions at similar distances from a county have similar enrollment market influence as in-state institutions.

5.1 STEM-Focused Institutions

In our initial analysis of the base model for public institution enrollment in states for which we have ground truth data we observe technological institutions that are over-represented in the number of counties in which they are predicted to exert the most influence. These institutions include the Missouri U. of Science & Technology, Michigan Tech and larger institutions such as Virginia Tech and North Carolina State. These institutions are similar in that they are each "STEM focused," with the majority of undergraduate students earning degrees in Science, Technology, Engineering and Math (STEM) disciplines[2]

Table 4 shows the percentage of STEM degrees for public STEM-focused institutions with enrollment influence over at least one county using the base model parameters. We use the ground truth data for the states of Missouri and Michigan to tune the model by reducing the market weight for the most STEM-focused public institutions (>80%) such as Missouri U. Science & Technology and Michigan Tech. Based on experimentation we reduce the enrollment weight by a factor of six for these institutions. The rationale for this reduction is these institutions are not comprehensive in the degrees they offer and thus do not exert the same immediate enrollment influence pattern for students in their region.

The market weight for institutions shown in Table 4 with 50-80% of degrees in STEM is also reduced. The reduction is scaled with almost no reduction for institutions just over 50% and more reduction for institutions closer to 80%. The last column in the table shows the resulting number of counties with enrollment influence after tuning the model. For example, the number of counties for Missouri U. of Science & Technology is reduced from nine counties in the base model to one county in the tuned model. This reduction more closely matches the ground truth for public universities in Missouri. As shown in the table, the influence of other, but not all, STEM-focused institutions is reduced using this tuned modeling. Most importantly, the tuning results is in better alignment between predicted and actual results for institutions in states with ground truth data.

5.2 Private Institutions

Treating private institutions using the same influence enrollment weight as public institutions results in 66 private institutions predicted to exert the most influence over 185 (6%) of the 3143 counties in the U.S. However similar to STEM-focused institutions, we conjecture that the influence pattern for private institutions are less influential in the immediate area of the institution.

As a means to understand the influence of private institutions we again make use of enrollment ground truth data for states described in Section 4.1. Many of these states provide enrollment data only for public institutions, but the data for Minnesota, Oklahoma, Texas and Virginia also contain enrollment data for in-state private institutions. We make use the data from these four states to tune the enrollment weight parameter for private institutions.

Actual enrollment data for Minnesota and Oklahoma show that at least one public institution has more enrollment than any private institution for each county. Similarly, the sphere of influence model does not predict any private institution would exert the most influence. Data for Texas shows

[2]We use IPEDS data to determine the percentage of STEM degrees awarded by each institution. The data provide the percentage of degrees awarded in 38 disciplines. We use the U.S. Immigration and Customs Enforcement list of STEM-Designated Degree Programs [36] to identify the 12 disciplines considered as STEM.

Table 4: Public STEM-Focused Institutions Exerting Enrollment Influence Over Counties Using Base and Tuned Models

		STEM	Cnt. Counties Exerting Enrollment Influence	
State	Institution	Degree Pct.	Base Model	Tuned Model
SD	SouthDakotaMines	98.3	9	1
NM	NewMexicoTech	97.3	1	0
MO	MissouriSci&Tech	93.8	9	1
MI	MichiganTech	86.2	5	3
WI	UWPlatteville	71.8	2	1
SD	DakotaSt	69.1	2	0
WV	WestVirginiaTech	65.1	5	3
PA	Pitt-Johnstown	64.5	1	0
MT	MontanaTech	64.3	3	2
IN	Purdue	63.3	15	14
CA	UCMerced	62.5	3	3
VA	VMI	59.2	2	0
NC	NorthCarolinaSt	58.9	10	8
AL	AlabamaA&M	58.7	1	1
GA	FortValleySt	58.5	4	3
NY	SUNYPoly	57.6	2	2
MA	UMassLowell	56.0	2	0
CA	UCSantaCruz	56.0	3	3
CA	UCDavis	55.2	7	7
MI	LakeSuperiorSt	54.0	2	2
TX	UTDallas	53.8	3	3
IA	IowaSt	53.7	61	56
TX	TAMUKingsville	53.3	3	3
OR	OregonSt	53.0	14	11
NY	StonyBrook	52.9	1	1
IL	Illinois	52.9	21	21
MI	Michigan	52.8	3	3
MS	AlcornSt	52.6	6	6
CA	CalPoly	52.5	1	1
VA	VirginiaTech	51.9	33	31
NY	CUNYCC	51.7	1	1
MD	StMary's(MD)	51.2	1	1
SC	Clemson	51.0	4	4
CA	California	50.6	3	3

that Baylor University enrolls the most students for one county. In contrast, the base model predicts that five institutions (Baylor U., Abilene Christian U. U. of Mary Hardin-Baylor, East Texas Baptist U. and Howard Payne U.) exert the most enrollment influence over at least one county. Data for Virginia shows that the private institutions Liberty U. (10 counties) and Emory & Henry College (2 counties) each enroll the most students for multiple counties, but each have a significant online enrollment. The base model predicts that three institutions—Liberty U., Shenandoah U. and Emory & Henry College—exert the most enrollment influence over at least one county.

Across all private institutions, mismatches between actual and predicted results indicate that the using the same enrollment weight of attraction for private as for public institutions over-represents the influence of these institutions. We thus experimented with changing the enrollment weight for private institutions and as for STEM-focused public institutions reduce the enrollment weight by a factor of six for private institutions. The resulting tuned results predict 12 private institutions exerting the most enrollment influence over 21 ($<1\%$) of all counties.

These private institutions include Morningside College, Northwestern College and Dordt College in Iowa; Augustana College in Illinois; and Southern New Hampshire U. in New Hampshire. We could not find data to confirm or dispute these predictions. BYU-Idaho, affiliated with The Church of Jesus Christ of Latter-day Saints, is predicted to exert the most influence over five eastern Idaho counties with a heavy proportion of Mormons. Baylor U. is predicted to have the most influence over one Texas county, which is confirmed by the data. Abilene Christian U. is predicted for two counties where the actual data shows a significant presence in one county and inconclusive data in another. Liberty U. is predicted for two counties in Virginia where the actual data shows ten counties, but it has a heavy on-line enrollment which may be skewing the actual influence. The U. of Pikeville in Kentucky is predicted to exert the most influence over one county, which could be confirmed [30]. Similarly, school district data for North Dakota [26] confirms the county predicted for the U. of Jamestown. Also in North Dakota, the U. of Mary is predicted to exert influence over four counties. The actual data show that this institution has a presence (or the data are inconclusive) for these counties.

While not complete, we believe these comparisons between tuned predictions and actual data for private institutions demonstrate they provide a relatively small amount of enrollment influence in limited regions of the U.S. Both the predictions and the actual data show the primary enrollment influence is from public institutions. However additional secondary influence from private institutions can be seen in the interactive maps on the website where the three most influential institutions are shown for each county.

5.3 State Borders

Previous work using the same college enrollment dataset shows students also enroll at out-of-state institutions [41]. For example, it shows a relatively high percentage and number of college-bound students from Minnesota and North Dakota that attend a college in the neighboring state.

These results lead to another consideration for our enrollment influence model, which is the effect of crossing a state border on the friction of distance. If the "cost" of a border is zero then the influence prediction would only need to account for distance. However, if there is a reduced influence when crossing a border then the border can be considered to have a "width" that is added to the distance when determining the influence of a market institution. A reduced or even zero border cost could be justified in cases where state institutions have reciprocity agreements with students in neighboring states paying reduced out-of-state tuition, such as between North Dakota and Minnesota [25].

To better understand the impact of crossing a state border on college influence, we re-ran our base model with no distance added for crossing a border. The results reveal a small number of instances where an institution's influence extends into another state. For example, the U. of North Dakota, North Dakota State and South Dakota State are predicted to have the primary enrollment influence over college-bound students of counties in western Minnesota. Utah State exerts influence in counties of both Wyoming and Idaho.

Matching these predicted influences with actual results is difficult due to the lack ground truth data. We were able to use school district data showing where North Dakota high school students attend college (both in-state and out-of-state) [26]. Making use of these ground truth data for North Dakota school districts sending students out-of-state we find that a border cost of zero is too low as the predicted influence of Minnesota institutions is too high compared to North Dakota institutions. Instead a cost of 25-50 miles is a better fit to the actual data. Thus in our tuned model we use a border width of 40 miles for all influence determinations that involve crossing a state border. The resulting tuned model does predict that the U. of North Dakota has primary influence over one northwestern Minnesota county. The interactive maps also show cross-border secondary influences.

6 Validation of Tuned Sphere of Enrollment Influence Model

After consideration for STEM-focused institutions, private institutions and the effect of state borders, Figure 2 shows our tuned prediction influence model for college enrollment. Each county is shown in a school color for the primary institution exerting the most influence for enrollment. For each county, we also determine the secondary institutions exerting the second- and third-most influence for county, which can be viewed in the interactive map that accompanies this report.

Relative to the base model map in Figure 1, the tuned model map in Figure 2 shows a much reduced influence for institutions with a strong STEM focus such as the South Dakota School of Mines. The U. of Jamestown and the U. of Mary are private institutions shown as influential in central North Dakota and the U. of North Dakota is the most influential for Kittson County in the northwest corner of Minnesota.

While the resulting map is interesting (and colorful), we also evaluate the prediction model for each of the 13 states with ground truth data. In the evaluation for each county, we deem a institution influence prediction as correct (or not incorrect) if the actual enrollment for that institution is the highest or within 10% of the highest. Across these thirteen states, the percentage of correct county predictions ranges from 70% (Texas) to 88% (Kansas). Overall the median state prediction success rate is 78% with a mean success rate for all counties across the 13 states of 79%. We observe that many of the mis-predictions occur at the boundaries between areas of influence with some mis-predictions expected due to differences in the timeframes of available data. Overall we believe these results indicate a good correspondence between the predicted and actual results for these states and provide confidence that the predicted results for other states are similarly correct.

The comparison of actual and predicted data also provide the opportunity to identify institutions and types of institutions that do not fit the model well. Observations from this type of comparison for the base model pointed out over representation of STEM-focused institutions. Table 5 shows

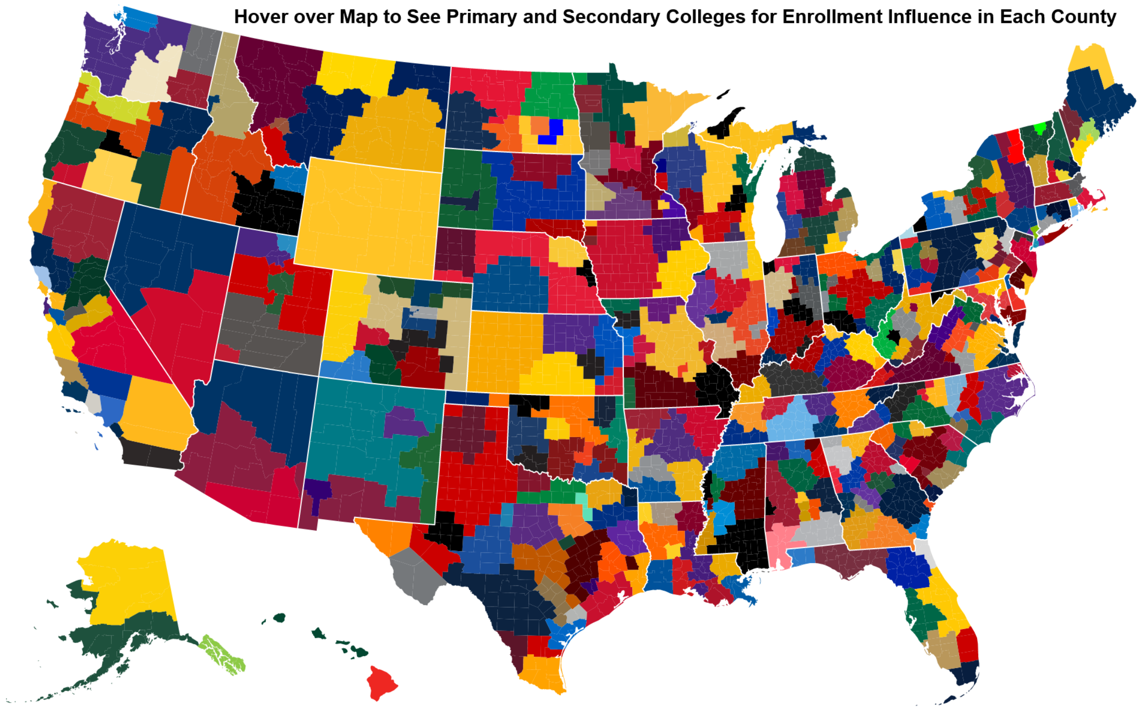

Figure 2: Map of College Enrollment Influence by Public and Private Institutions Using Tuned Prediction Model

institutions with predicted enrollment influence for at least four counties with more than half of those predictions not matching available data.

There is no common theme among the institutions in Table 5. There are a number of major state institutions (e.g. North Carolina State, U. of North Carolina, U. of Texas) where the predicted enrollment influence reach is not as broad as shown in the available data. There are other schools with less influence such as the College of Charleston and William & Mary. There are the urban institutions U. of Missouri in St. Louis and in Kansas City as well as the U. of Houston and U. of New Orleans with an over-representation of predicted counties. There are even a few regional public institutions on the list.

In summary the results across the U.S show that regional public institutions generally exert the primary influence in their respective regions of the state with major public institutions in states exerting secondary influence over their entire state.

As a means to visualize all influences on each county, Figure 3 shows two other interesting and realistic maps of the enrollment influence. The use of three different types of maps to represent college influence is similar to those studied for representing voting patterns [7].

The left map uses color intensity to represent the influence exerted by the primary institution. Areas in which there is particularly weak influence show a correlation with what have been identified as education deserts [10]. These areas include western North Dakota down through eastern New Mexico, western Wyoming down through the Arizona/New Mexico border, and eastern Oregon into northern Nevada.

The right map of the figure blends the colors of institutions exerting influence on each county

Table 5: Institutions Predicted to Exert Enrollment Influence Over At Least Four Counties with More than Half Mis-Predictions

		Cnt. Counties			Miss
State	Institution	Predicted	Correct	Incorrect	%
SC	Charleston	4	0	4	100
MO	UMStLouis	4	0	4	100
VA	William&Mary	5	1	4	80
LA	NewOrleans	5	1	4	80
NC	NorthCarolinaSt	8	2	6	75
NC	NorthCarolina	9	3	6	67
TX	Houston	8	3	5	63
TX	Texas	13	5	8	62
UT	Utah	10	4	6	60
MO	UMKansasCity	5	2	3	60
MN	UMMorris	7	3	4	57
ND	MinotSt	13	6	7	54
OK	Oklahoma	15	7	8	53

in proportion to the influence exerted. Areas in which the influence is primarily exerted by one institution, such as the U. of Wyoming in eastern Wyoming or Texas Tech in western Texas, largely show that school's true colors. However, there is a blurring of colors when multiple institutions exert influence with similar weights.

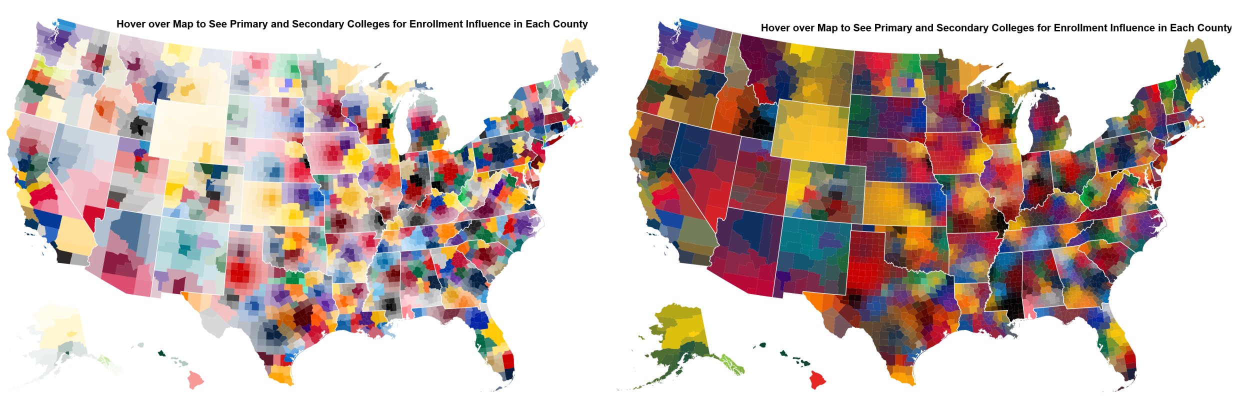

Figure 3: Intensity and Blended Maps of College Enrollment Modeled Influence

7 Alumni Influence

After parameterizing and validating a sphere of influence model for college enrollment, we examine other types of influence exerted by colleges. The first of these influences makes use of an interesting study examining where college graduates move after graduation [6]. This study surveyed college graduates from 445 institutions across the U.S. to understand where they moved and thus where their college exerts alumni influence. The results are reported as the percentage of graduates moving to one of 70 metropolitan areas as well the remaining percentage that move elsewhere.

We make use of the "explore the data yourself" feature of the study's website to obtain the percentages of the most popular of the 70 metropolitan areas for graduates of each institution. We also obtain from the website the percentage of graduates that move outside of these metropolitan areas.

We use these data along with our college enrollment data to model the alumni sphere of influence for each institution. In doing so we assume the number of graduates of an institution is proportional to the number of enrolled students. While we understand that the graduation rate is not the same for all institutions, we believe the number of graduates of an institution correlating to the number of enrolling students is a reasonable approximation for our work.

Making use of this approximation we first compute the number of institution alumni moving to each of the 70 metropolitan areas. We do so by multiplying the percentage of graduates for each institution listed for the metropolitan area by the total number of enrolling students for the institution to estimate the number of graduates of the institution in that area. The institution with the highest number (not necessarily the highest percentage) of graduates is modeled as exerting the most alumni influence for the "central," but not "outlying," counties of the Metropolitan Statistical Area (MSA) [28]. Note for modeling purposes we assume an institutions influence does not necessarily cross state borders in the metropolitan area.

The sphere of influence for counties outside of the 70 metropolitan areas is modeled by multiplying the percentage of graduates outside of the 70 metropolitan areas by the enrollment within the state of the county in which influence is being determined. We use the parameters of the tuned enrollment model for determining this influence.

The resulting map showing a school color for the primary institution exerting the most alumni influence in each county is shown in Figure 4. It shows many of the same institutions influencing areas as for enrollment, but with fewer institutions under consideration than we used for modeling enrollment influence.

As a means for comparison, we re-computed the enrollment sphere of influence map using the tuned model parameters for only the 445 institutions with available alumni data. The enrollment influence areas for this reduced set of institutions is shown in Figure 5.

There are differences between the areas of influence for institution alumni in Figure 4 and enrollment in Figure 5 (intensity and blended versions of each map are available on the website). For example, there are differences between the two maps for the boundary line of influence between the U. of North Dakota and North Dakota State within North Dakota and stretching across the border into western Minnesota. However, making a visual comparison between the two maps is difficult. Rather we determine the number of counties in which influence is exerted in each map and highlight institutions with the largest differences. The institutions with these largest negative differences in counties between alumni and enrollment influence are shown in Table 6.

Each institution shown in the table is located in one of the 70 metropolitan areas of the data. We conjecture that these institutions tend to enroll students from a broader geographic area, but after these students become alumni then they stay in the metropolitan area in relatively larger numbers.

In contrast, Table 7 shows the institutions with the largest positive differences in counties be-

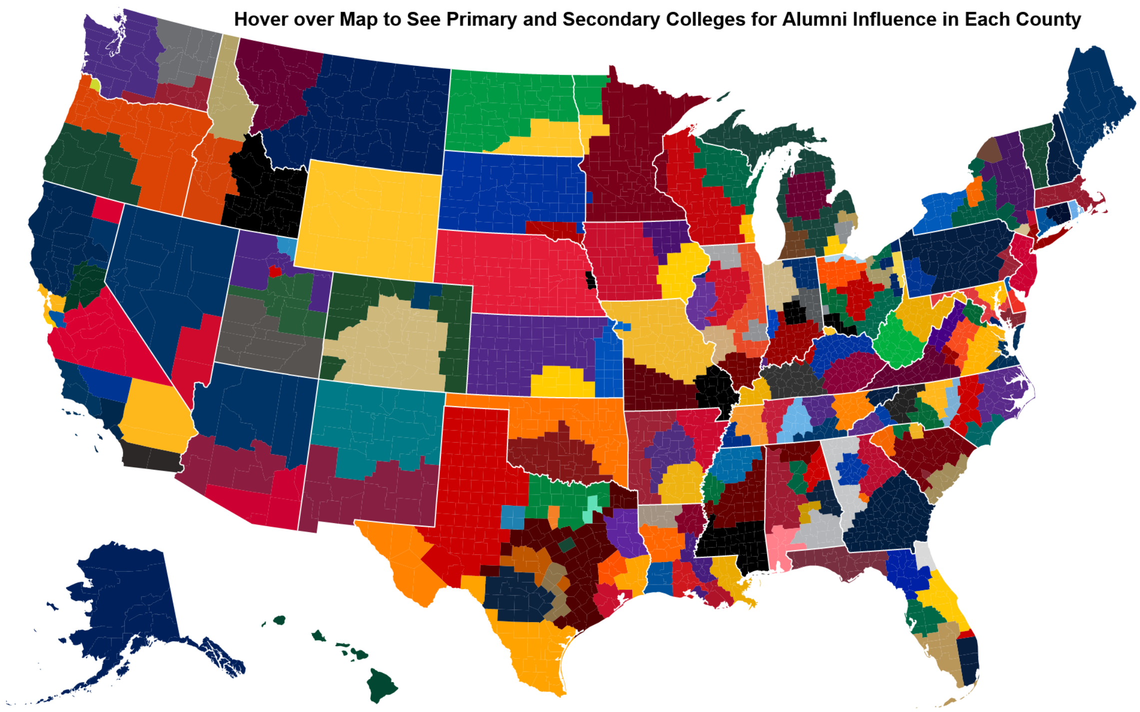

Figure 4: Map of College Alumni Modeled Influence

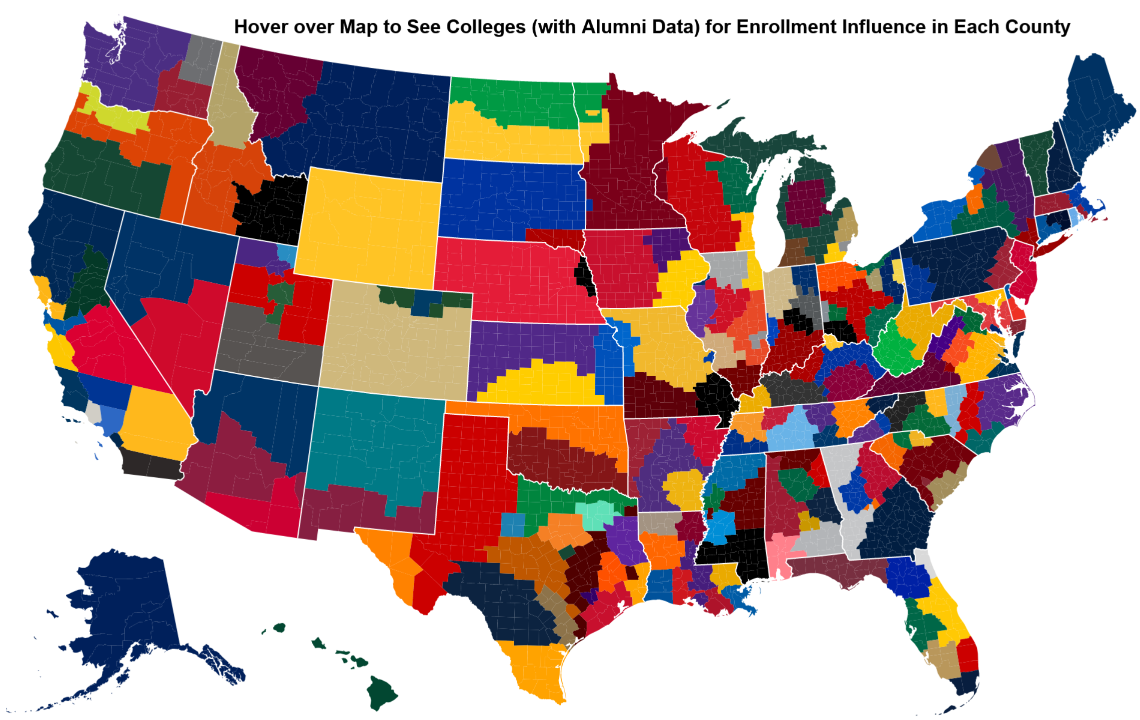

Figure 5: Map of College Enrollment Influence for Institutions with Alumni Data

Table 6: Institutions Exerting Alumni Influence in Fewer Counties Than Enrollment Influence

Institution	Cnt. Counties		
	Enrollment	Alumni	Difference %
WichitaSt	31	12	19
Colorado	57	42	15
NorthDakotaSt	27	12	15
CentralArkansas	31	17	14
GeorgiaSt	19	5	14
UTSA	27	15	12
OhioSt	25	15	10
Texas	22	12	10
Louisville	13	4	9
Utah	10	1	9

tween alumni and enrollment influence. Each of these institutions is outside of one of the 70 metropolitan areas so the resulting alumni are predicted to have less concentration in where they reside after college.

Table 7: Institutions Exerting Alumni Influence in More Counties Than Enrollment Influence

Institution	Cnt. Counties		
	Enrollment	Alumni	Difference %
TexasA&M	20	49	29
KansasSt	50	70	20
NorthDakota	26	41	15
GeorgiaSouthern	57	72	15
ColoradoSt	6	21	15
UTRioGrandeValley	14	23	9
Arkansas	16	25	9
OregonSt	16	24	8
OklahomaSt	38	46	8

8 Research Influence

Another aspect of colleges and universities is the research conducted at these institutions. One metric for quantifying this research is the amount of funding generated by each institution. We apply the research and development expenditures [23] for each institution as their weight of attraction in a sphere of influence model. We investigate the research influence model using the same parameters developed for enrollment influence and similarly reduce the weight for private institutions. We only consider research influence of an institution within its own state as we have no data on how research influence is affected by state borders.

Figure 6 shows the resulting research sphere of influence map with the primary institution shown for each county (intensity and blended versions are available on the website). As expected it includes many of the same institutions that exert enrollment and alumni influences, but with different weights the areas of influence change. For example, a number of STEM-focused institutions exert influence over a broader area than for enrollment.

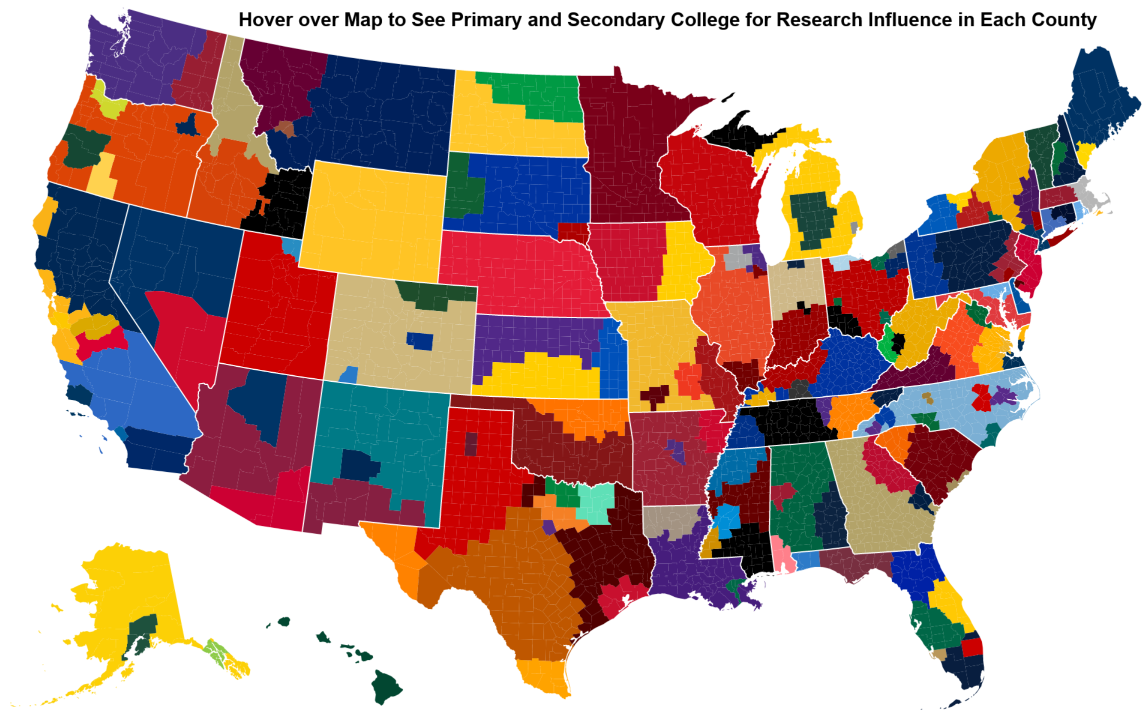

Figure 6: Map of College Research Modeled Influence

In particular, notable institutions exerting relatively more research influence include the U. of Utah, U. of Michigan, U. of Texas, U. of North Carolina, Georgia Tech and the U. of Alabama-Birmingham. Based on the metric of funding, these institutions exert an outsized research influence in their state relative to the number of students they enroll.

9 College Football Influence

We next examine the influence of institutions on rooting interests for college sports beginning with football. We found prior work identifying college football rooting areas based on three types of data: social media followers/likes [8, 31], seats sales for games [5, 11], and informal surveys [17, 3]. While interesting in themselves and providing data for modeling the football fan influence of colleges, these data have shortcomings. These issues include ad hoc survey results with a relatively small number of respondents, differences in the set of schools considered, and even outright contradictions between datasets for the identified rooting areas within different states.

Despite these shortcomings, we do use them to guide the development of a model for the influence of schools on football fan interest. We consider all Division I schools, both in the Football Bowl Subdivision (FBS) and in the Football Championship Subdivision (FCS).

We use the same base parameters as for other models ($\lambda = 2.5$, minimum distance of 20 miles). The primary variable that we control is the weight of attraction assigned to each school. Similar to enrollment influence, we also examine the impact of crossing a state border on fan interest.

We do focus on the FBS schools and first assign the same weight to all schools in the "Power Five" conferences (Atlantic Coast Conference, Big 12 Conference, Big Ten Conference, Pacific-12 Conference, and Southeastern Conference) as well as BYU and Notre Dame as prominent independent schools. By default, the weight for private institutions is reduced as done for other influence models. We next adjust the weights of these schools to best reflect data showing the relative popularity of schools within states. For example, prior data generally shows Alabama having a larger sphere of influence than Auburn. It also shows Virginia Tech having a larger influence area than Virginia. The data show the private institutions Notre Dame, BYU and Southern California have a larger sphere of influence than expected.

We expand on the modeled set of schools by adding the remaining FBS schools with smaller weights that best reflect available data. We also add in schools at appropriate weights from FCS conferences, particularly when FCS schools are in states without any FBS schools. The remaining FCS schools are added with a minimum default weight. Finally we experimented with the "distance" to assign to a state border as we observe the data showing that the influence of a few schools crosses over into an adjacent state. For example, the U. of Oregon exerts influence in northern California while North Dakota State exerts rooting influence in western Minnesota. Based on the data, we use a value of 130 miles in our model. The resulting tuned model is shown in Figure 7 reflecting the school exerting the primary college football fan rooting interest on each county in the U.S.

Not only are the results themselves interesting, but they also identify schools for which the exerted influence is stronger (or weaker) than would be expected. Among the notable outcomes we observe in Figure 7 are:

- the U. of Oregon not only exerts an outsized influence over Oregon State within the state, but it also exerts influence into northern California and Alaska;
- North Dakota State exerts more influence than the U. of North Dakota within the state and its influence crosses the state line into western Minnesota;
- the influence of Penn State is greater than expected over other public institutions in the state and extends into western New York state;
- the influence of the U. of Georgia is particularly strong in the entire state with Georgia Tech not even exhibiting primary influence in the Atlanta area,
- Notre Dame, BYU and USC are private institutions with larger areas of influence than expected; and
- in a number of states the influence of one public institution is stronger than expected in comparison with rival institutions, including U. of Michigan over Michigan State, U. of Alabama over Auburn U., U. of Iowa over Iowa State, U. of Oklahoma over Oklahoma

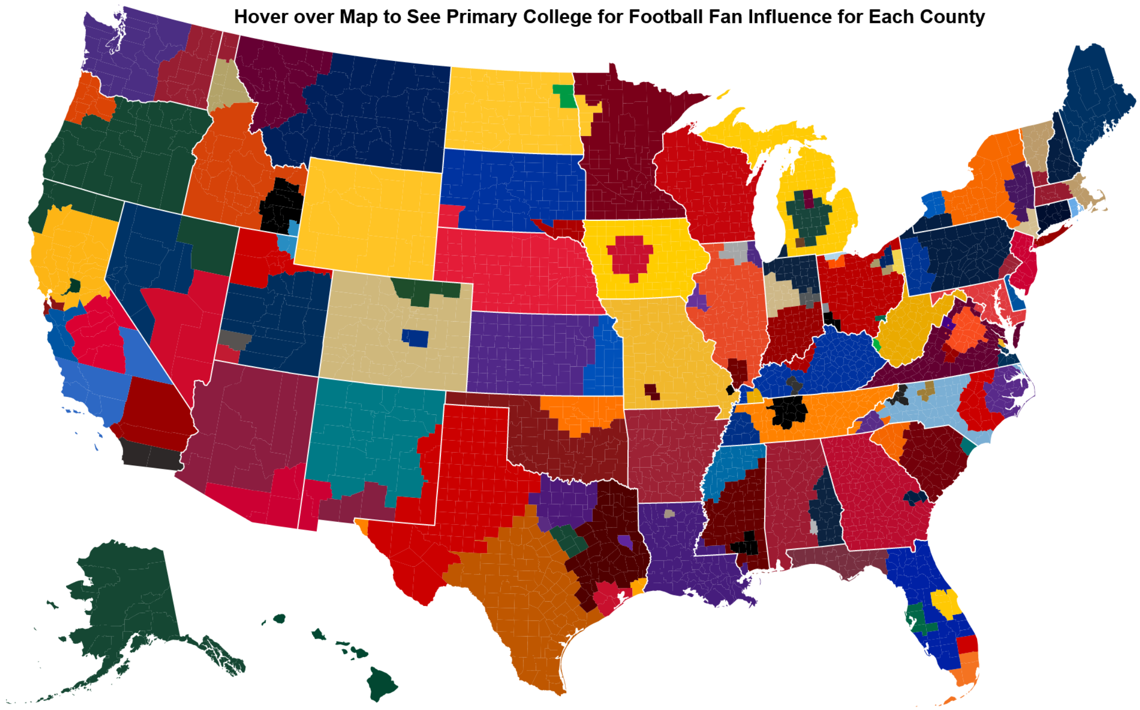

Figure 7: Map of College Football Fan Modeled Influence

State, Virginia Tech over Virginia, U. of North Carolina over North Carolina State, and U. of Texas over other in-state public institutions.

Finally, Figure 8 shows two additional influence maps where the left map shows the intensity of influence by the primary school and the right map blends the colors of schools exerting influence on each county in proportion to the influence exerted. Areas in which the influence is primarily exerted by one school, such as eastern Wyoming or central Georgia, show that school's true color. However, the right map of Figure 8 shows a blurring of colors at the boundaries of influence between multiple schools. This blurring is more realistic as most areas of the country have competing fan interests.

10 Men's College Basketball Influence

We next look to create a similar area of interest map for men's college basketball. We found limited previous work to identify rooting areas for basketball. Maps appear at NCAA tournament time showing predicted fan rooting areas, but only include tournament teams. In their simplest form these maps assume all schools are weighted equally and do not account for state borders, although in previous work we did consider these factors [40]. In terms of consideration of all schools, we found only one study identifying rooting areas across the U.S. [33]. It uses seat sales data. Unfortunately, these data do not include most NCAA Division I schools.

Due to more limited data available for modeling men's college basketball we begin with the model weights developed for college football. We then add in member schools for basketball-only

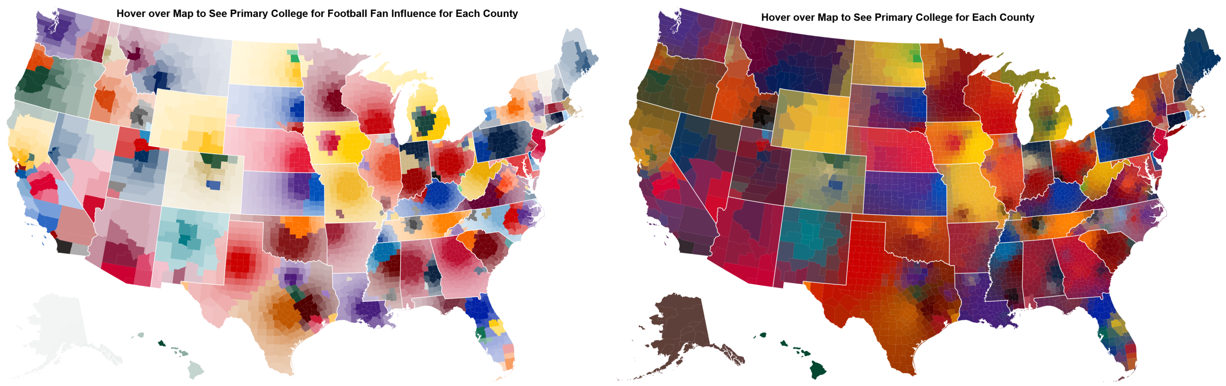

Figure 8: Intensity and Blended Maps of College Football Fan Modeled Influence

conferences such as the Big East and West Coast Conferences. Finally we adjust the weights for schools with available data while still retaining schools in conferences without basketball seat data. The resulting map is shown in Figure 9 reflecting the school exerting the primary men's college basketball fan rooting interest for each county in the U.S.

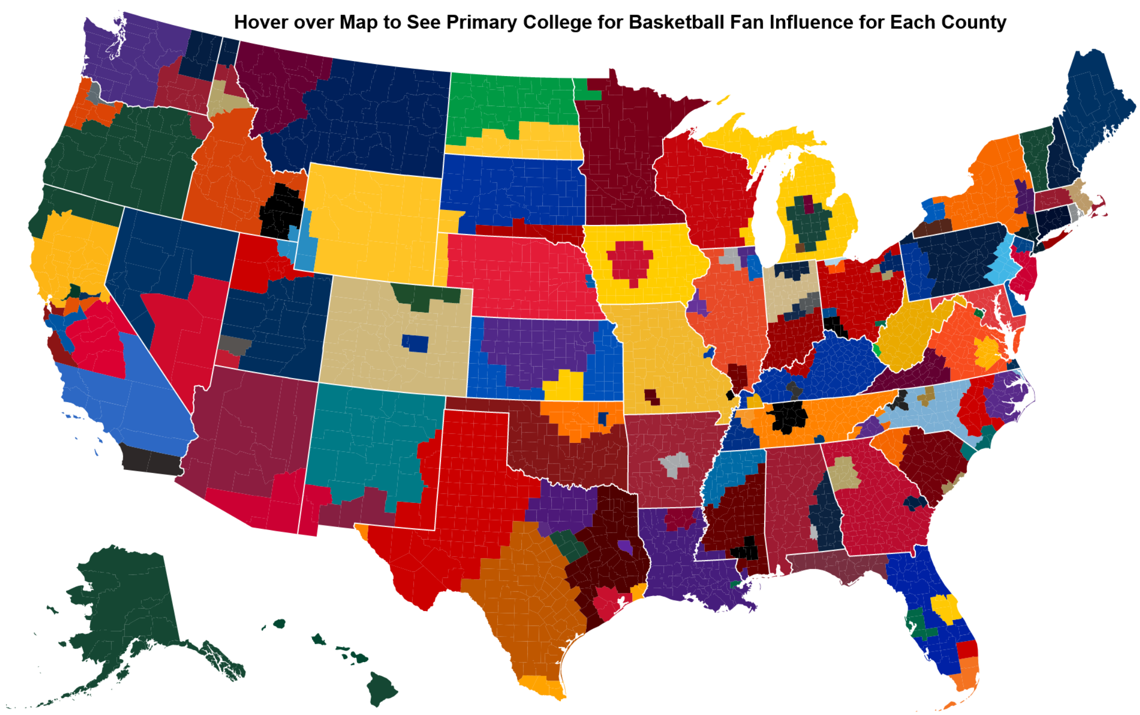

Figure 9: Map of Men's College Basketball Fan Modeled Influence

Similar to football, the map shows a number of interesting results where a school exerts stronger (or weaker) fan influence than expected. Notable observations include:

- the U. of Oregon again exerts outsized influence over Oregon State within the state as well as exerting influence into northern California and Alaska;

- Gonzaga, Villanova, BYU, Syracuse and Seton Hall are private institutions that each exert outsized influence compared with other such institutions;
- the relative influence of the U. of Virginia and VirginiaTech is reversed from football with the U. of Virginia exerting more influence; and
- as in football there are a number of states where the influence of one public institution is stronger than expected in comparison with another institution including U. of Kansas over Kansas State. U. of North Carolina over North Carolina State, U. of Michigan over Michigan State, U. of Iowa over Iowa State, U. of Alabama over Auburn U., and U. of Georgia over Georgia Tech (but not as strongly as football),

As done for football, Figure 10 shows two additional maps. The left map shows the intensity of basketball fan interest. The right map blends colors for schools exerting influence on each county in proportion to the influence exerted. This map shows many areas in which more than one institution has comparable influence.

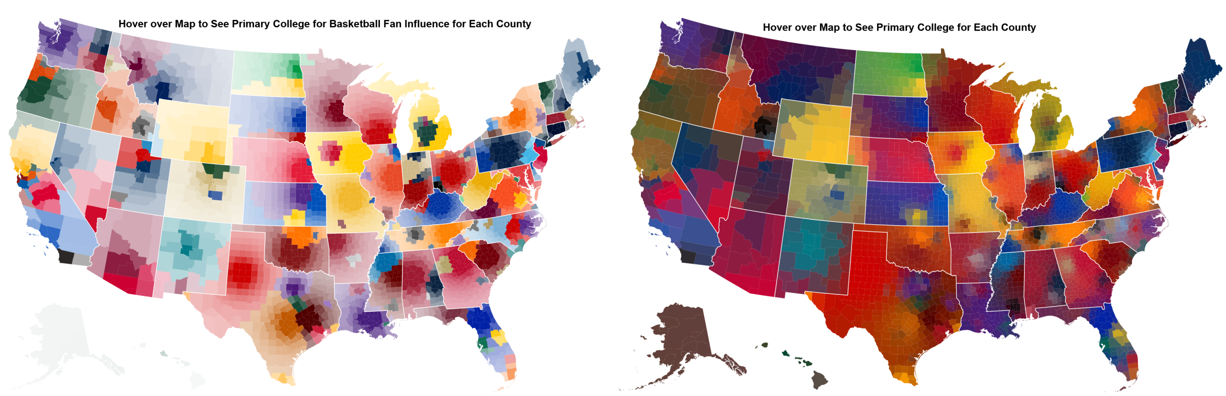

Figure 10: Intensity and Blended Maps of Men's College Basketball Fan Modeled Influence

11 Composite Influence

Having examined influence for the five college aspects of enrollment, alumni, research, football and men's basketball, we have the opportunity to compare influence across these aspects. We can identify areas in which a single institution exerts primary influence for each of these aspects as well as areas in which no institution exerts influence for a majority of these aspects.

The approach we took for this composite influence analysis is to extract the primary institution exerting influence on each county for each of the five aspects. Figure 11 shows a map of the results on this analysis using one of three types of coloring for each county:

1. a school color of an institution is shown if it exerts primary influence for all five studied aspects of influence,

2. dark gray is shown if no institution exerts primary influence for majority of the five studied aspects of influence, and
3. light gray is shown if an institution exerts primary influence for a majority of, but not all, five studied aspects of influence.

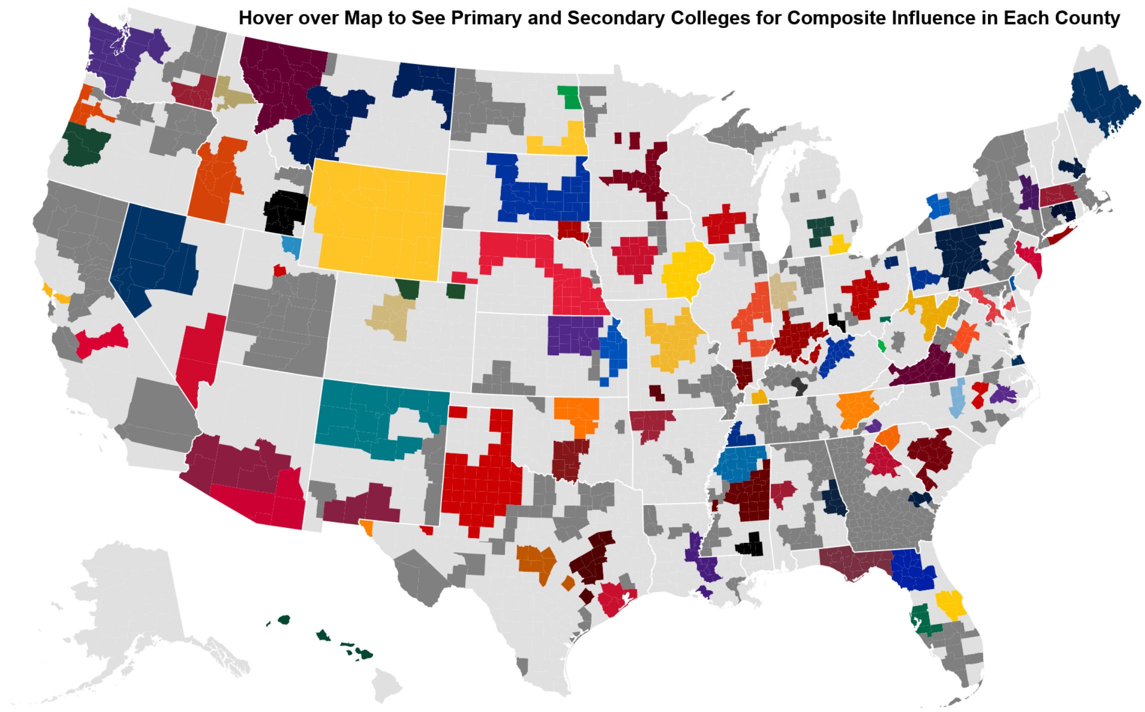

Figure 11: Map of Composite Influences for Five Studied Aspects

The composite map shows that virtually all counties in Wyoming are primarily influenced by the U. of Wyoming while much of the state of Nevada is primarily influenced by the U. of Nevada or UNLV. In contrast, sizable portions of Georgia, New York and Utah do not have a single institution exerting primary influence across a majority of the five aspects.

Figure 12 shows results of a more systematic analysis where the top-10 states for percentage of counties with a singular institution exerting primary influence for all five aspects are rank ordered. It shows Wyoming, Hawaii and Nevada having the highest proportion of counties under a singular institution of influence.

Figure 13 shows the contrasting results where the top-10 states for percentage of counties without any institution exerting primary influence for a majority of the five aspects are rank ordered. It shows the District of Columbia (with only one "county"), Georgia, Massachusetts, New York and Utah having the highest proportion of counties where there is not a single institution clearly exerting influence across multiple aspects. We conjecture that the population in these areas are less likely to identify with a particular institution.

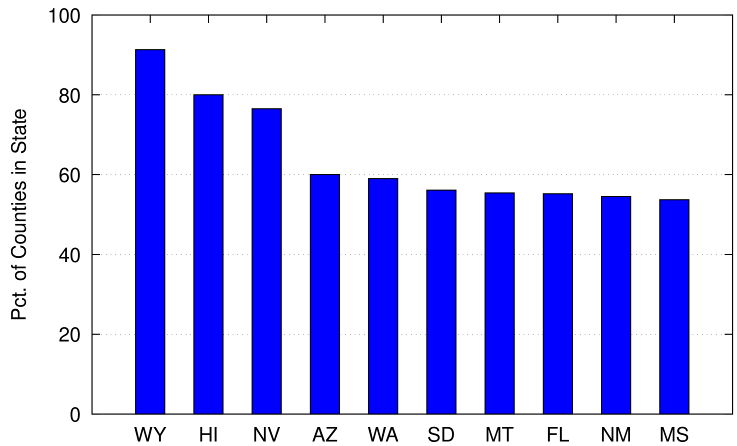

Figure 12: Top-10 States for Pct. of Counties with Singular Institution Exerting Primary Influence for All Five Studied Aspects

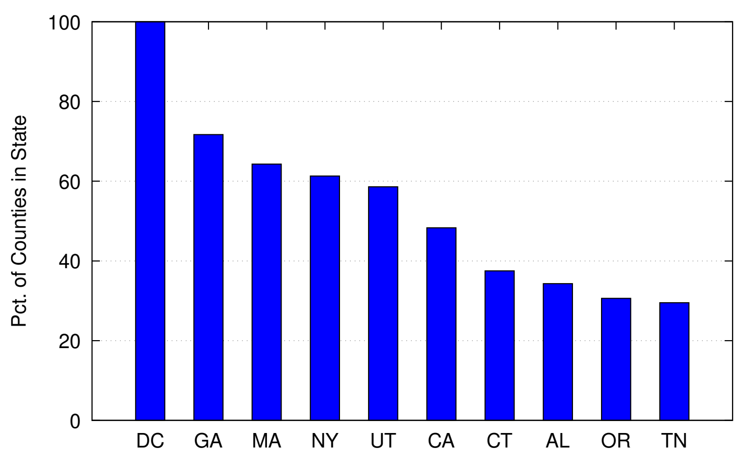

Figure 13: Top-10 States for Pct. of Counties without Any Institution Exerting Primary Influence for Majority of Five Studied Aspects

12 Summary and Future Work

In summary, this work employs a geographical perspective to investigate the influence of American colleges. Many aspects of colleges are examined including influence exerted for student enrollment, where their alumni reside, research impact and sports fandom.

We apply a gravity-based model for the determination of spheres of influence for each of these aspects. While this model has typically been applied using one or more metrics of urban markets, in this work we treat colleges and universities as the markets. We model their spheres of influence using an appropriate metric for the weight of attraction of each market. The model accounts for multiple markets in the same vicinity as well as the friction of distance on influence as we consider locations further away from a market. We also account for other considerations such as how to model the influence of public versus private institutions, institutions offering a more specialized set of degrees and the impact of crossing state borders on institution influence.

The availability of actual enrollment data on a per-county basis for a subset of states both assist us in building the model for this aspect as well as to validate the resulting model. We observe close to an 80% correspondence between the predicted and actual data for counties with available data, which we believe is good given differences in timeframes for these data.

Not only do we model influence, but we generate maps for over 3000 counties in the U.S. showing a school color of the institution exerting primary influence in each county. These maps show the spheres of influence for each institution. In addition, we also show a more interesting and realistic map of the enrollment influence by blending the colors of institutions exerting influence on each county in proportion to the influence exerted. These maps show that regional public institutions generally exert the primary influence in respective regions of their state with major public institutions in states exerting secondary influence over their entire state.

Influence results for the location of alumni show institutions located in metropolitan areas tend to retain more alumni influence in these areas despite enrolling students from a larger area. Research impact measured by funding show that institutions such as U. of Utah, U. of Michigan, U. of Texas, U. of North Carolina, Georgia Tech and the U. of Alabama-Birmingham exert an outsized research influence in their state relative to the number of students they enroll.

Applying the sphere of influence model to the college sports of football and men's basketball both makes use of and unifies disparate data identifying rooting area of college fans. Not only are the results themselves interesting, but they also identify schools for which the exerted influence is stronger (or weaker) than would be expected. Football results show the U. of Oregon, North Dakota State and Penn State exert outsized influence that crosses into adjacent states. Notre Dame, BYU and USC are private institutions with larger areas of influence than expected. There are a number of states in which the influence of one public institution is stronger than expected in comparison with rival institutions. These institutions include the U. of Georgia, U. of Michigan, U. of Alabama and the U. of Texas. Maps of primary and blended influence for college football fan rooting areas are both colorful and interesting.

We similarly model sphere of influence for men's college basketball. These results and maps are similar as for football. Notable results show the U. of Oregon again exerting outsized influence that crosses over into northern California and Alaska. Gonzaga, Villanova, BYU, Syracuse and Seton Hall are private institutions that each exert outsized influence compared with other such institutions. As in football there are states where the influence of one public institution is stronger than expected. These institutions include Kansas U. and the U. of North Carolina.

Finally, we use sphere of influence results from across the five college aspects to identify areas in which a single institution exerts primary influence for each of these aspects as well as areas in which no institution exerts influence for a majority of these aspects. These composite results show the states of Wyoming, Hawaii and Nevada having the highest proportion of counties under a singular institution of influence. In contrast, the states of Georgia, Massachusetts, New York and Utah have the highest proportion of counties where there is not a single institution clearly exerting influence across multiple aspects. We conjecture that the population in these areas are less likely to identify with a particular institution.

This work also points to a number of directions for future work. These directions include: incorporating a more realistic measure of distance such as travel time [39] between locations; considering other factors in enrollment modeling such as more specializations than STEM focus; and modeling of additional aspects and metrics of American college influence.

References

[1] Alabama Commission on Higher Education. ACHE data portal, 2022. https://www.ache.edu/index.php/data-portal/.

[2] Brian JL Berry and RF Lamb. The delineation of urban spheres of influence: Evaluation of an interaction model. *Regional Studies*, 8(2):185–190, 1974.

[3] cinciforthewin. /r/CFB fan map, 2016. https://www.reddit.com/r/CFB/comments/4i51df/rcfb_fan_map/.

[4] Yu Deng, Shenghe Liu, Li Wang, Hanqing Ma, and Jianghao Wang. Field modeling method for identifying urban sphere of influence: A case study on Central China. *Chinese Geographical Science*, 20:353–362, 2010.

[5] Rachael Dottle and Jake Lourim. How widespread is your college football fan base?, October 24, 2019. https://projects.fivethirtyeight.com/college-football-ticket-sales/.

[6] D. Dougherty, B. McGill, D. Chinni, and A. Zitner. Where graduates move after college. *Wall Street Journal*, May 15, 2018. https://www.wsj.com/graphics/where-graduates-move-after-college/.

[7] Rémy A Furrer, Karen Schloss, Gary Lupyan, Paula M Niedenthal, and Adrienne Wood. Red and blue states: dichotomized maps mislead and reduce perceived voting influence. *Cognitive Research: Principles and Implications*, 8(1):1–16, 2023.

[8] Tom Giratikanon, Josh Katz, David Leonhardt, Kevin Quealy, and Marc Tracy. N.C.A.A. fan map: How the country roots for college football. *New York Times*, October 3, 2014. https://www.nytimes.com/interactive/2014/10/03/upshot/ncaa-football-fan-map.html.

[9] Howard L. Green. Hinterland boundaries of New York City and Boston in Southern New England. *Economic Geography*, 31(4):283–300, 1955.

[10] Nick Hillman. Place matters: A closer look at education deserts, May 21, 2019. https://www.thirdway.org/report/place-matters-a-closer-look-at-education-deserts.

[11] Caitlyn Holroyd. Map shows top-selling college football team in each state for 2019. https://www.thescore.com/news/2016991.

[12] David L. Huff. A probabilistic analysis of shopping center trade areas. *Land Economics*, 39(1):81–90, 1963.

[13] David L. Huff. The delineation of a national system of planning regions on the basis of urban spheres of influence. *Regional Studies*, 7(3):323–329, 1973.

[14] David L Huff and George F Jenks. A graphic interpretation of the friction of distance in gravity models. *Annals of the Association of American Geographers*, 58(4):814–824, 1968.

[15] David L. Huff and James M. Lutz. Ireland's urban system. *Economic Geography*, 55(3):196–212, 1979.

[16] Kansas Board of Regents. Higher education statistics, 2022. https://www.kansasregents.org/data/system_data/enrollment_reports.

[17] Jason Kirk. 25 maps that explain college football, August 20, 2014. https://www.sbnation.com/college-football/2014/8/20/6030683/25-maps-that-explain-college-football.

[18] Louisiana Board of Regents. Statewide student profile system, 2023. https://apps.regents.state.la.us/Reports/Report.aspx?reportPath=/SSPS/SPPRSH.

[19] Michigan Senate Fiscal Agency. Study of Michigan public university enrollment patterns by county and institution, 2018. https://www.senate.michigan.gov/SFA/Publications/HiEdEnrollment/PublicUniversityEnrollmentPatterns2018.pdf.

[20] Minnesota Office of Higher Education. Residence of students enrolled in Minnesota institutions, 2022. http://www.ohe.state.mn.us/sPages/student_enroll_data3.cfm.

[21] Missouri Department of Higher Education & Workforce Development. First-time undergraduate enrollment at public baccalaureate and higher degree-granting institutions, by Missouri county, Fall 2014, 2015. https://dhewd.mo.gov/data/statsum/2014-2015%20Statistical%20Summary/table065_1415.pdf.

[22] National Center for Education Statistics. Integrated Postsecondary Data System, 2022. https://nces.ed.gov/ipeds/use-the-data.

[23] National Science Foundation National Center for Science and Engineering Statistics. Rankings by total R&D expenditures, 2021. https://ncsesdata.nsf.gov/profiles/site?method=rankingBySource&ds=herd.

[24] The university of north carolina system county connections fact sheets, 2023. \newblockhttps://www.northcarolina.edu/impact/county-connections/.

[25] North Dakota higher education enrollment projections: Making sense of the data, 2019. https://ndus.edu/wp-content/uploads/sites/6/2021/01/NDUS-Making-Sense-of-Enrollment-Projections.pdf.

[26] Insights into North Dakota education, 2022. https://insights.nd.gov/Education.

[27] North Dakota University System. Fall enrollment report, 2020. https://files.eric.ed.gov/fulltext/ED612582.pdf.

[28] Office of Management and Budget. Revised delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and guidance on uses of the delineations of these areas, March 6 2020. https://www.whitehouse.gov/wp-content/uploads/2020/03/Bulletin-20-01.pdf.

[29] Oklahoma State Regents for Higher Education. Annual headcounts by institution; demographics and geographic origin 2009-10 to 2018-19, 2019. https://www.okhighered.org/studies-reports/enrollment-dashboards/dashboard-GeographicOriginofStudentHeadcountinOklahomaPublicandPrivateHigherEducationAnnualHeadcountbyCountryStateandOklahomaCounty.html.

[30] Council on Postsecondary Education Interactive Data Center. Kentucky County Profile - Graduation Rates, 2019. https://reports.ky.gov/t/CPE/views/KentuckyPostsecondaryEducationInteractiveDataDashboard/Navigation.

[31] Matt Sorenson. College football fan map, Spring 2021. https://public.tableau.com/app/profile/matt.sorenson/viz/CollegeFootballFollowerMap/CollegeFootballMap.

[32] South Carolina Commission on Higher Education. Statistical abstract, 2022. https://che.sc.gov/sites/che/files/Documents/CHE%20Data%20and%20Reports/2022_CHE_Statistical_Abstract_Web.pdf.

[33] Stephen Spiewak. Map: The most popular NCAA college basketball teams by county, March 13, 2023. https://www.vividseats.com/blog/most-popular-college-basketball-teams-map.

[34] Texas Higher Education Coordinating Board. Texas higher education data, 2022. http://www.txhighereddata.org/.

[35] United States Census Bureau. U.S. Gazetteer Files. https://www.census.gov/geo/maps-data/data/gazetteer.html.

[36] U.S. Immigration and Customs Enforcement. STEM designated degree program list, May 2016. https://www.ice.gov/sites/default/files/documents/Document/2016/stem-list.pdf.

[37] Utah System of Higher Education. County of origin, 2023. https://ushe.edu/institutional-data-resources-county-of-origin/.

[38] Virginia State Council of Higher Education. Ld03: Total college and university enrollment, all students, by student origin, 2023. https://research.schev.edu/localities/LD03_Total.asp.

[39] Craig E. Wills. Geographical connectivity in the United States. Technical Report WPI-CS-TR-17-01, Computer Science Department, Worcester Polytechnic Institute, June 2017.

Available at SSRN:
https://ssrn.com/abstract=4351192.

[40] Craig E. Wills. NCAA men's and women's basketball predicted rooting interest areas for 2021 tournament teams, March 2021. Separate maps for the Men's and Women's Tournament.
https://www.cs.wpi.edu/%7ecew/sportsmaps/ncaam21.html
https://www.cs.wpi.edu/%7ecew/sportsmaps/ncaaw21.html.

[41] Craig E. Wills and Chayanne Sandoval-Williams. Migration of American college students. Technical Report WPI-CS-TR-22-06, Computer Science Department, Worcester Polytechnic Institute, 2022. Available at SSRN:
https://ssrn.com/abstract=4344894.

[42] David WS Wong and Qunying Huang. "Voting with their feet": delineating the sphere of influence using social media data. *ISPRS International Journal of Geo-Information*, 6(11):325, 2017.