




A New Lens on High School Dropout: Use of Correspondence Analysis and the Statewide Longitudinal Data System

Kathryn Schaefer Ziemer, Bianica Pires, Vicki Lancaster, Sallie Keller, Mark Orr & Stephanie Shipp

To cite this article: Kathryn Schaefer Ziemer, Bianica Pires, Vicki Lancaster, Sallie Keller, Mark Orr & Stephanie Shipp (2017): A New Lens on High School Dropout: Use of Correspondence Analysis and the Statewide Longitudinal Data System, The American Statistician, DOI: [10.1080/00031305.2017.1322002](https://doi.org/10.1080/00031305.2017.1322002)

To link to this article: <https://doi.org/10.1080/00031305.2017.1322002>

 View supplementary material 

 Accepted author version posted online: 26 Jun 2017.

 Submit your article to this journal 

 Article views: 73

 View Crossmark data 

Running head: A NEW LENS ON HIGH SCHOOL DROPOUT

**A New Lens on High School Dropout: Use of Correspondence Analysis and the Statewide
Longitudinal Data System**

Kathryn Schaefer Ziemer, Bianica Pires, Vicki Lancaster, Sallie Keller, Mark Orr, Stephanie

Shipp

Social and Decision Analytics Laboratory

Bioinformatics Institute of Virginia Tech

Corresponding author:

Kathryn Schaefer Ziemer

Postdoctoral Researcher

Social and Decision Analytics Laboratory

Bioinformatics Institute of Virginia Tech

900 N. Glebe Rd.

Arlington, VA 22203

schaeferkl@gmail.com

507-382-3239

KATHRYN SCHAEFER ZIEMER is a Postdoctoral Researcher at the Social and Decision Analytics Laboratory within the Bioinformatics Institute of Virginia Tech, 900 N. Glebe Rd., Arlington, VA 22203; schaeferkl@gmail.com. Her research interests focus on health decisions, health behaviors, and the impact of social, psychological, and biological factors on well-being.

BIANICA PIRES is a Postdoctoral Researcher at the Social and Decision Analytics Laboratory within the Bioinformatics Institute of Virginia Tech, 900 N. Glebe Rd., Arlington, VA 22203; bpires@vbi.vt.edu. Her research interests involve understanding and exploring complex social systems using computational methods, including agent-based modeling (ABM), social network analysis (SNA), and geographic information systems (GIS) to model human movement and behavior.

VICKI LANCASTER is a Senior Research Scientist at the Social and Decision Analytics Laboratory within the Bioinformatics Institute of Virginia Tech, 900 N. Glebe Rd., Arlington, VA 22203; val20003@vbi.vt.edu. Her research interests focus on applying statistical logic and methodologies across diverse disciplines, including educators, geneticists, chemists, biologists, medical doctors, veterinarians, epidemiologists, demographers, and environmentalists.

SALLIE KELLER is Director and Professor of Statistics for the Social and Decision Analytics Laboratory within the Bioinformatics Institute of Virginia Tech, 900 N. Glebe Rd., Arlington, VA 22203; sallie41@vbi.vt.edu. Her research interests focus on uncertainty quantification,

computational and graphical statistics and related software and modeling techniques, and data access and confidentiality.

MARK ORR is a Research Associate Professor at the Social and Decision Analytics Laboratory within the Bioinformatics Institute of Virginia Tech, 900 N. Glebe Rd., Arlington, VA 22203; morr9@vbi.vt.edu. His research interests focus on understanding dynamic processes and drivers of risky behavior and decision making, primarily in a public health context, at the scale of the individual and populations.

STEPHANIE SHIPP is the Deputy Director and Research Professor for the Social and Decision Analytics Laboratory within the Bioinformatics Institute of Virginia Tech, 900 N. Glebe Rd., Arlington, VA 22203; steph19@vbi.vt.edu. Her research interests focus on innovation and competitiveness, use of big data to advance policy, social network analysis and evaluation of education grant programs, advanced manufacturing, the role of federal laboratories, and funding of high-risk/high-reward research.

ABSTRACT

The combination of loglinear models and correspondence analysis, have long been used to decompose contingency tables and aid in their interpretation (van der Heijden et al. 1989). Until now, this approach has not been applied to the education Statewide Longitudinal Data System (SLDS), which contains administrative school data at the student level. While some research has been conducted using the SLDS, its primary use is for state education administrative reporting. This paper uses the combination of loglinear models and correspondence analysis to gain insight into high school dropouts in two discrete regions in Kentucky, Appalachia and non-Appalachia, defined by the American Community Survey. The individual student records from the SLDS were categorized into one of the two regions and a loglinear model was used to identify the interactions between the demographic characteristics and the dropout categories, push-out and pull-out. Correspondence analysis was then used to visualize the interactions with the expanded push-out categories, boredom, course selection, expulsion, failing grade, teacher conflict, and pull-out categories, employment, family problems, illness, marriage, and pregnancy to provide insights into the regional differences. In this paper we demonstrate that correspondence analysis can extend the insights gained from SDLS data and provide new perspectives on dropouts.

1. INTRODUCTION

To date, our understanding of high school dropout has been limited by the data collected and the methodological approaches used. There are many reasons why students drop out of high school, however, researchers and education stakeholders do not have a clear picture of the pattern of these reasons across student characteristics and geographic locations. This is in part due to the limitations of the data and in part this is a limitation of the analytics used to study dropout reasons. This paper takes a new look at high school dropout by using statewide longitudinal data system (SLDS) education information, combining it with other data, and using correspondence analysis to visualize the results.

The SLDS is a potentially powerful source of data for studying education. It contains administrative education information, such as enrollment information, student characteristics, and dropout status, which is collected by schools and school systems and reported to the state several times a year. While some research has been conducted with this data (e.g., Ahn 2014; Bettinger 2015; Keller et al. 2016), its primary use has been for state educational reporting (e.g., Conaway et al. 2015; Prince and Jenkins 2005). Since the SLDS is comprised of data on the entire population of public school students in the state, it provides a level of geographical breadth and granularity that would be difficult to achieve with traditional education surveys. One purpose of this study is to use SLDS data to explore the reasons why students drop out of high school across different geographical regions.

The reasons for dropping out can be numerous and are usually grouped into larger categories, such as push-out and pull-out reasons (Fine 1986; Rumberger 1987). This binary dropout variable of push-out/pull-out is then used as the outcome variable in logistic regression

analyses (e.g., Bradley and Renzulli 2011; Stearns and Glennie 2006). However, this method risks masking some of the nuances of the specific reasons that students drop out. In this paper, we seek to understand the distribution of specific dropout reasons across key student characteristics. To accomplish this, we use correspondence analysis, an approach not often used in traditional education research, to visualize the complex interaction effects for the specific dropout reasons. We argue that the use of SLDS data and correspondence analysis can provide a greater level of granularity in the area of dropout research.

This article will first describe the literature on push-out and pull-out dropout reasons, explain the characteristics of SLDS data, provide context to the geographical regions of Kentucky, and list the study objectives. Next, the characteristics of the Appalachia region will be reported using Census Bureau data. Then, information about the Kentucky SLDS data source will be provided, including the sample and variables used, and the statistical methods of the loglinear model and the correspondence analysis will be described. The results section breaks down the findings by the three study objectives and the conclusion section describes the overall contribution of the article.

1.1 Push-out and Pull-out Reasons

Dropping out of school carries a large cost to individuals and to society. Students who drop out of school are more likely to experience a number of negative outcomes, including lower earnings, higher unemployment, higher incarceration, and lower life expectancy (Belfield and Levin 2007; Hayes et al. 2002; Martin et al. 2003; Stanard 2003). Understanding why students drop out of school can provide crucial information to inform policy and prevention efforts.

The reasons for dropping out are generally grouped into push-out versus pull-out reasons (Fine 1986; Rumberger 1987). Pull-out reasons are factors external to the school system that pull students out, such as caring for family members or needing to find a job to make money (McNeal Jr 1997; Stearns and Glennie 2006). Push-out reasons are factors associated with the school environment that push students out, such as boredom or failing classes (Fine 1986; Jordan et al. 1996; Rumberger 1987). The key difference is the context of the main factor that causes the student to drop out. Push-out factors lie within the school context and result in a negative relationship between the student and the school environment. Often students who drop out for push-out reasons feel alienated and incapable of succeeding in school, which is only reinforced through suspensions and failing grades (Fine 1986; Jordan et al. 1996; Rumberger 1987).

Alternatively, pull-out factors lie outside of the school within the larger context of an individual's environment. This larger context includes family, community, religious institutions, the job market, and other social forces (Jordan et al. 1996; McNeal Jr 1997; Stearns and Glennie 2006). When these forces come into conflict with the educational objectives of the schools, the external factors may be more compelling than staying in school or may interfere with the student being able to succeed at school.

Previous research comparing pull-out and push-out reasons have found differences based on gender, race, and grade. Males are more likely to drop out for push-out reasons, whereas females are more likely to drop out for pull-out reasons (Bradley and Renzulli 2011; Doll et al. 2013; Jordan et al. 1996; Stearns and Glennie 2006). There also appear to be interactions between gender and race/ethnicity. For example, African-American males are significantly more likely to drop out for disciplinary reasons and less likely to drop out for employment reasons

compared to White males (Stearns and Glennie 2006). In terms of grade, students in later grades are more likely to leave for academic and employment reasons and less likely to leave for disciplinary reasons compared to earlier grades (Stearns and Glennie 2006).

While dropout reasons have been explored according to demographic groups, they have generally not been explored by geography. Instead, studies that compare different geographical areas (e.g., rural versus urban) have focused on dropout rates rather than dropout reasons (Johnson Jr 2013; Leventhal-Weiner and Wallace 2011; Owens 2010; Swanson 2004). Part of this limitation may be the use of survey data, which often have geographical breadth or granularity, but not both.

1.2 Statewide Longitudinal Data System

The SLDS data has several strengths, including geographical breadth and granularity that allows for locating students according to the schools they attend across the entire state. The entire public school student population is contained within the SLDS and students can be linked across years through a unique student identification number, which allows for longitudinal analysis. However, the limitations of SLDS data must also be acknowledged. For example, the data are de-identified making it impossible to link student records with external data sources that provide information on family demographics and student mental and physical health. In addition, variables that are often included in dropout theories, such as student engagement (Archambault et al. 2009), are not included in the data, resulting in a lack of theoretical justification for exploratory modeling.

1.3 Kentucky Geographical Regions

This study focuses on dropouts in Kentucky. Kentucky is largely rural, with 80% of school districts in small towns or rural areas, and has a large number of low socioeconomic status students (Swanson 2004; U.S. Census Bureau 2014). However, Kentucky is not geographically homogenous and contains both Appalachian and non-Appalachian regions. The Appalachia region still has some of the lowest educational attainment in the United States, despite overall declines in the national dropout rate (Pollard and Jacobsen 2015; Stark et al. 2015). The educational disparities between Appalachia and non-Appalachia combined with the negative consequences of dropping out risk creating even greater disparities. This study is the first to our knowledge to explore variations in dropout reasons between the Appalachian and non-Appalachian regions.

1.4 Study Objectives

There are three objectives in this study: first, an examination of student dropout demographics in the two regions; second, an examination of the push-out versus pull-out categories in these two regions broken down by demographic groups (race, grade, and gender); and third, an examination of the association among demographic groups and specific dropout reasons between these two regions using correspondence analysis.

2. CHARACTERIZATION OF APPALACHIA

The Appalachia region of the United States spans a 205,000 square mile area from Mississippi to New York (Appalachian Regional Commission 2016). Central Appalachia, which is primarily comprised of the eastern part of Kentucky, has historically suffered from poverty, health problems, unemployment, low educational attainment, and high dropout rates (Ali and

McWhirter 2006; Bennett 2008; Lemon et al. 1993; Rogers et al. 1997; Sugar 2002). In Kentucky, the Appalachian region has a smaller population per square mile, a greater percentage of White (not Hispanic) residents, higher unemployment, lower household income, higher poverty, a greater percentage of disability, and less mobility as compared to the non-Appalachian region (Pollard and Jacobsen 2015, see Table 1). While educational attainment has improved over the last several decades, this region still lags behind the national average (Pollard and Jacobsen 2015; Stark et al. 2015). One-quarter of people ages 25 and over in Appalachia have less than a high school diploma, as compared to 13.9% for the non-Appalachian region and 14.0% for the U.S (see Table 1; Pollard and Jacobsen 2015).

Appalachia's social and physical isolation has led to a distinct culture characterized by strong family ties, reluctance to change, distrust of outsiders, and a time orientation focused on the present (Bennett 2008; Elam 2002). As such, the Appalachian culture affects educational values, aspirations, and attainment. For instance, Appalachian students tend to have lower educational aspirations, especially when they live in areas with low educational attainment (Bennett 2008; Brown et al. 2009). This appears to be particularly true for males socialized in working class families and communities, which value labor over education (Chenoweth and Galliher 2004). Students in rural central Appalachia who pursued employment after high school rather than college had lower vocational confidence, lower outcome expectations for college, and perceived a greater number of barriers to postsecondary education (Ali and McWhirter 2006). Barriers to educational attainment include taking care of family members and lack of financial resources (Brown et al. 2009).

Given the unique culture of Appalachia and the previous research on education in this region, there is reason to believe that the reasons why students drop out may be different in Appalachia as compared to non-Appalachia. This study sought to explore these potential differences by comparing dropout reasons between the two regions and utilizing correspondence analysis to visualize the results.

3. SLDS DATA

3.1. Data Source

Data for this project are from the Kentucky Statewide Longitudinal Data System (SLDS), which contains data on all public school students in the state of Kentucky reported by school and school district. The data represent the entire population of public school students in the state of Kentucky rather than a sample. Because the data collected through the SLDS are for school system operations, they cannot be changed. As a result, certain types of information were unobtainable, such as psychological factors or family data. In addition, some variables are not readily available to researchers external to the Kentucky Department of Education. This includes free/reduced lunch information that is often used as a proxy for poverty.

3.2. Sample and Variables

This study used longitudinal data on students in grades 9 to 12 for the academic years of 2008-2009 through 2013-2014. On October 1st, each school reports which students dropped out. Therefore, a student documented as a dropout for the 2008-2009 school year either dropped out during the preceding academic year after October 1st or did not enroll by October 1st for the 2009-2010 school year. The reasons for dropping out are obtained from the student or an adult

responsible for the student. If a student was reported as dropping out multiple times, only the first instance of dropout was included in our analyses.

The following variables from the dataset are included in this study: gender, race/ethnicity, grade, dropout status, dropout reason (for those who dropped out), and county in which the student was attending school. There were ten dropout reasons plus the absence of a reason: boredom, course selection, expulsion, failing grade, teacher conflict, employment, family problems, illness, marriage, pregnancy, and “no substantial enrollment.” “No substantial enrollment” was used when the specific dropout reason was unknown. The ten dropout reasons are collapsed into push-out and pull-out categories and included “no substantial enrollment” as a separate category labeled as unknown (see Table 2). A variable indicating whether a student attended school in an Appalachian county or non-Appalachian county is included. The Appalachia region in Kentucky consists of 54 counties in the eastern part of the state, whereas the non-Appalachia region consists of 66 counties in the western and central parts of the state.

Dropouts labeled as unknowns were included in the descriptive analyses but were excluded from the inferential analyses. This was done since including unknowns in the inferential process would not contribute to our understanding of why students leave school. There were differences among unknown dropout percentages for the region by race categories, Appalachia/White = 30%, non-Appalachia/White = 27%, Appalachia/non-White = 58%, and non-Appalachia/non-White=44%. In the final inferential analyses, the focus was on White dropouts due to the small number of non-White students in Appalachia. The current study assumed that the difference between unknown dropout percentages for White students in Appalachia versus non-Appalachia was similar and would not bias any inferences made.

4. STATISTICAL METHODS

4.1. Loglinear Model

Both inferential and descriptive statistics were used to analyze the data. A loglinear model was used to describe the frequency counts for the 64 combinations formed by: region (Appalachia and Non-Appalachia), grade (9th, 10th, 11th, and 12th), gender (female, male), dropout reason category (push-out, pull-out), and race (non-White, White), summed over the five school terms starting in 2009 and ending in 2014. This type of model was selected since it isolates the interactions in multi-way contingency tables. The drawback with this approach is that it requires ≥ 5 observations for each combination. To meet this requirement, the 10 dropout reasons had to be collapsed into two categories, push-out and pull-out, as well as race, which was collapsed into the two categories, non-White and White students.

To understand the interactions among the two regions, the demographic variables, and the ten non-categorized dropout reasons (i.e., boredom, course selection, expulsion, failing classes, teacher conflict, employment, family problems, illness, marriage, and pregnancy) some of the interaction effects from the loglinear model were visualized using correspondence analysis (CA).

Loglinear models are used to describe the association and interaction patterns of cell counts in contingency tables (Agresti 2013). If there is no association between the categorical variables, the log of the cell counts is a function of the additive effects of the categorical variables and there is no statistical association among the categorical variables. If the loglinear model for independence is rejected, then the goal of the inferential analysis is to isolate significant interactions among the categorical variables using the most parsimonious model. Model selection was conducted by starting with the saturated model (region, grade, gender, race,

and dropout reason category and all interactions) and using a backward selection algorithm that employed the Akaike Information Criterion (AIC) (Akaike 1973). The AIC provides a balance between closeness of the predicted values to the true values versus the number of parameters in the model.

4.2. Correspondence Analysis (CA)

CA is a multivariate technique whose geometric interpretation can be used to display categorical data, such as those displayed in two-way or multiway contingency tables (Agresti 2013; Greenacre 1984). Moser (1989) made the astute observation that, “So just as scatter plots are complementary to a linear regression analysis, correspondence analysis is an appropriate technique for graphically exploring contingency tables,”

CA decomposes the chi-square statistic, χ^2 , with the contingency table into orthogonal factors that maximize the separation between the row and column counts. For example, with a 2x2 contingency table each cell is divided by the total count so that the table has a mass of one, the cells are then normalized based on the row and column masses. Thus the cell frequencies are normalized by the contribution of that particular row or column to the total mass of the contingency table. The normalized table is visualized using the principal coordinates that are a function of the right and left singular vectors of the normalized table. For an explanation of the visualization of multiway tables the reader is referred to Greenacre and Blasius (2006).

In this paper, CA plots were used to interpret some of the interaction effects in the final model that were selected based on the AIC. Since small cell sizes (< 5) prohibited the use of the ten dropout reasons in the loglinear model, we substituted the ten reasons for the two dropout categories when constructing the CA plots and focused only on White students. This enabled us

to at least gain a better understanding of the dropout reason differences between the two regions for White students.

All analyses were conducted in R version 3.2.3 (2015-12-10) using RStudio, Inc. version 0.99.489.

5. RESULTS

5.1. Dropout Demographics by Region

To address the first objective of the study, Table 3 presents descriptive statistics on the distribution of dropouts by Appalachia and non-Appalachia regions, for gender, race, and grade. There were a total number of 1,215,183 students in grades 9-12 enrolled in the Kentucky public school system over the years 2009-2014. Of these, 354,686 (29%) were enrolled in the Appalachia region and 860,497 (71%) in non-Appalachia. A slightly larger percentage of students, 13,401 (1.6%), dropped out of the non-Appalachian region than the Appalachian region (4,669/1.3%).

A comparison of the demographics for dropouts in Appalachia and non-Appalachia revealed similarities in terms of grade and gender. For both regions, a higher proportion of males drop out than females, and a higher proportion of 10th and 11th grade students drop out compared to 9th and 12th grade students. The most prominent difference between the two regions was a larger percentage of White students dropping out in Appalachia and a larger percentage of African-American students dropping out in non-Appalachia. This may reflect, in part, the overall demographics of the regions as Appalachia has a higher percentage of White residents and non-Appalachia has a higher percentage of African-American residents.

5.2. Push-out vs. Pull-out Dropouts

To address the second objective of the study, descriptive statistics by push-out and pull-out dropouts are presented in Table 4. The frequencies and percentages are broken down by region, gender, grade, and race (White, African-American, and Other). The percentage of students with an unknown dropout reason was higher for African-American and Other students across all region, gender, and grade combinations. As indicated in Table 4, a higher percentage of non-Appalachia White females and males drop out for push-out reasons as compared to White females and males in Appalachia. For White females in both regions, push-out reasons predominate in the 9th grade, whereas pull-out reasons predominate starting in the 10th grade in non-Appalachia and 11th grade in Appalachia.

In contrast, push-out reasons predominate for White males across all grades in both regions. It is hard to make generalizations for African-American and Other students in Appalachia since the cell numbers in Table 4 are so low. But for female African-American and Other students in non-Appalachia, there is a trend where the percentage of pull-out reasons predominates over push-out in the later grades (10th through 12th); this trend is not seen in non-Appalachia male African-American and Other students.

The inferential analyses were conducted using a loglinear model. The response variable in the model was the log of the cell counts and the predictor variables were the 5 main effects, the 10 two-way interactions, 10 three-way interactions, 5 four-way interactions, and the one 5-way interaction. There were 12,207 observations across the 64 cells. The analyses were run using the saturated model (all 31 effects were included in the model) and then a backward step algorithm was used to find the most parsimonious model based on the AIC.

The final model included the 5 main effects, the 10 two-way interactions, 8 of the 10 three-way interactions (Region:Grade:Race, Region:Grade:Gender, Region:Grade:POPO (Push-out/Pull-Out), Region:Race:Gender, Region:Race:POPO, Grade:Race:Gender, Grade:Race:POPO, Grade:Gender:POPO) and 1 of the 5 four-way interactions (Region:Grade:Race:Gender). The model was selected using the deviance from the saturated model, which is a perfect fit to the data. When compared to the saturated model, the deviance of the final model had a likelihood ratio χ^2 of 10.79, 15 degrees of freedom, and a p-value of 0.77. This was the simplest model that was not statistically different from the saturated model. The model fit was assessed using an F-statistic, p-value < 0.05, an indication the final model fits well. This was verified by evaluating the model assumptions. The standardized residuals were visualized using distribution plots and evaluated for normality using the D'Agostino and Pearson inferential test (D'Agostino and Stephens 1986).

While the final model provides a good fit to the data, the multiple two-, three-, and four-way interactions make the results difficult to interpret. The complex interplay between grade, gender and race indicate that dropout reasons cannot be understood by region alone. This aligns with the literature, which has found interactions between demographic groups and dropout reasons (Bradley and Renzulli 2011; Stearns and Glennie 2006).

5.3. Specific Dropout Reasons

To address the third objective and to better understand the complexity of the model, CA plots were used to explore two of the three-way interactions selected for inclusion in the final model: Region:Race:POPO and Region:Grade:POPO. A CA plot was constructed using the 10 dropout reasons in place of the two push-out/pull-out categories.

Figure 1 is the CA plot across all grades and gender for race (White and non-White, where non-White is African-American and Other combined) by dropout reasons for each region. The distance between race in Figure 1 is a measure of the similarity among their dropout reason frequency profiles. Figure 1 shows that the dropout frequency profiles for White and non-White students in non-Appalachia are more similar than they are for White and non-White students in Appalachia. With regard to the distance between drop out reasons, the pull-out reasons (blue) are more tightly clustered in Appalachia (Figure 1a) meaning their frequencies are similar. For both regions, the distance between the push-out reasons (green) is more varied.

Note that with respect to each other, the positions of the two sets of points for drop out reason and race are not directly comparable and are interpreted here with caution. What can be said is that each dropout reason point lies in the neighborhood of the racial group in which the reason's profile is prominent; therefore, we have enlarged the principles coordinate points (grey circles) for race to aid in interpretation. For example, there are relatively more dropouts due to boredom among non-White students in Appalachia than among White students (Figure 1a); whereas expulsion and course selection dominate for Other students in non-Appalachia. For both regions, push-out reasons dominate for non-White students. Averaged across grade and gender, it appears that pull-out reasons are more closely associated with White students in both regions.

Given the limitations of exploring dropout reasons for non-White students where the numbers in Appalachia are low, we focused on White students in Figure 2. Figure 2 is the CA plot for 9th grade White students for each region by dropout reasons for each gender. The distance between regions is a measure of the similarity between the region frequency profiles across the ten dropout reasons. In Figure 2, the location of the points for Appalachia and non-

Appalachia are on opposite sides of the horizontal axis, an indication their drop out frequency profiles are different, more so for White female students than males.

There are relatively more dropouts due to boredom in the Appalachia region than non-Appalachia for both White females and males. For White males, boredom, employment, family problems, and illness are more strongly associated with Appalachia, whereas course selection, failing classes, marriage, and teacher conflict are more strongly associated with non-Appalachia (Figure 2a); male expulsion rates were similar between the two regions. For females, marriage, expulsion, and boredom reasons for dropouts are more strongly associated with Appalachia; whereas course selection, employment, failing classes, and teacher conflict with non-Appalachia (see Figure 2b). The remaining dropout reasons of family problems, pregnancy, and illness were similar between the two regions for females (Figure 2b).

The CA plots for 10th, 11th, and 12th grade students were constructed and are presented in the supplementary material. This paragraph provides a summary of the plots. For White females, dropout reasons of boredom, marriage, and illness are associated with Appalachia and course selection, failing classes, and family problems with non-Appalachia. For 10th, 11th, and 12th grade White males, marriage and boredom are associated with Appalachia and course selection and failing classes with non-Appalachia. Across all grades the dropout frequency profiles of White females have greater regional differences than those for White males.

6. CONCLUSIONS

The study demonstrated the value of using correspondence analysis (CA) based on contingency tables for understanding high school dropout. Instead of predicting dropout as an outcome, CA allowed us to explore the patterns of dropout reasons across important

demographic characteristics. There were several key aspects of the Statewide Longitudinal Data Systems (SLDS) that supported this analysis. First, the sample size was large enough to explore the numerous dropout categories across multiple student characteristics. Second, the geographic granularity of the data also allowed for exploration of two culturally distinct regions: Appalachia and non-Appalachia. In short, the combination of CA and SLDS provides a new lens with which to view high school dropouts.

The findings from the loglinear model indicate that the reasons for dropping out are embedded in the context of a region. Overall, both push-out and pull-out reasons were common in Appalachia and non-Appalachia. However, there were complex interactions across both regions for the push-out and pull-out reasons and the demographic characteristics. As a result, the findings of the loglinear model were extremely difficult to interpret given the higher order interactions.

When CA plots were used to visualize the data, the nature of the interactions became apparent. For example, push-out reasons dominated for non-White students and pull-out reasons dominated for White students across both regions. Moreover, CA also allowed for the exploration of specific dropout reasons, which provides a more nuanced understanding of why different student groups drop out of high school. The findings demonstrate that while push-out and pull-out were common across both Appalachia and non-Appalachia, the specific dropout reasons differed between the two regions. A good example of this is for White males, where the dropout reasons of boredom, employment, family problems, and illness were more strongly associated with Appalachia, and course selection, failing classes, marriage, and teacher conflict were more strongly associated with non-Appalachia. This highlights the importance of looking at

specific dropout reasons rather than grouping them into larger categories. Moreover, this information can be used to inform dropout prevention efforts and tailor programs to address the specific reasons for dropping out.

The findings also demonstrate the usefulness of SLDS data in exploring dropout patterns across different geographic regions. This study is the first to our knowledge to explore differences in dropout reasons between Appalachia and non-Appalachia, an important geographic distinction since Appalachia is populated by understudied and vulnerable communities with some of the lowest educational attainment in the United States. Overall, the use of both SLDS data and CA provides a greater level of granularity that may be useful for studying disparities and understanding the complex interactions that occur in other areas of education.

References

- Agresti, A. (2013). *Categorical Data Analysis 3rd ed.* Hoboken, NJ John Wiley & Sons, Inc.
- Ahn, T. (2014). "A regression discontinuity analysis of graduation standards and their impact on students' academic trajectories." *Economics of Education Review*. 38, 64-75.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971* (pp. 267-281). Budapest: Akadémiai Kiadó.
- Ali, S. R., & McWhirter, E. H. (2006). "Rural Appalachian Youth's Vocational/Educational Postsecondary Aspirations Applying Social Cognitive Career Theory." *Journal of Career Development*, 33(2), 87-111.
- Appalachian Regional Commission. (2016). The Appalachian Region. Retrieved from http://www.arc.gov/appalachian_region/TheAppalachianRegion.asp
- Archambault, I., Janosz, M., Fallu, J-S., & Pagani, L.S. (2009). "Student engagement and its relationship with early high school dropout." *Journal of Adolescence*, 32(3), 651-670.
- Belfield, C., & Levin, H. (2007). *The Return on Investment for Improving California's High School Dropout Rate*. California Dropout Research Project. UC Santa Barbara.
- Bennett, S. L. R. (2008). "Contextual affordances of rural Appalachian individuals." *Journal of Career Development*, 34(3), 241-262.
- Bettinger, E. (2015). "Need-based aid and college persistence: The effects of the Ohio College Opportunity Grant." *Educational Evaluation and Policy Analysis*. 37, 102S-119S.

- Bradley, C. L., & Renzulli, L. A. (2011). "The complexity of non-completion: Being pushed or pulled to drop out of high school." *Social Forces*, 90(2), 521-545.
- Brown, R., Copeland, W. E., Costello, E. J., Erkanli, A., & Worthman, C. M. (2009). "Family and Community Influences on Educational Outcomes Among Appalachian Youth." *J Community Psychol*, 37(7), 795-808. doi:10.1002/jcop.20331
- Chenoweth, E., & Galliher, R. V. (2004). "Factors influencing college aspirations of rural West Virginia high school students." *Journal of research in rural education*, 19(2), 1-14.
- D'Agostino, R. B., & Stephens, M. (Eds.). (1986). *Goodness-of-fit-techniques*. New York, NY: CRC press.
- Doll, J. J., Eslami, Z., & Walters, L. (2013). "Understanding Why Students Drop Out of High School, According to Their Own Reports: Are They Pushed or Pulled, or Do They Fall Out? A Comparative Analysis of Seven Nationally Representative Studies." *SAGE Open*, 3(4). doi:10.1177/2158244013503834
- Elam, C. (2002). "Culture, poverty and education in Appalachian Kentucky." *Education and Culture*, 18(1), 10-13.
- Fine, M. (1986). "Why urban adolescents drop into and out of public high school." *The Teachers College Record*, 87(3), 393-409.
- Greenacre, M. (1984). *Theory and Application of Correspondence Analysis* Orlando, FL: Academic Press.
- Greenacre, M., & Blasius, J. (2006). *Multiple correspondence analysis and related methods*: CRC Press.

- Hayes, R. L., Nelson, J.-L., Tabin, M., Pearson, G., & Worthy, C. (2002). "Using school-wide data to advocate for student success." *Professional School Counseling*, 6(2), 86-94.
- Johnson Jr, O. (2013). "Is concentrated advantage the cause? The relative contributions of neighborhood advantage and disadvantage to educational inequality." *The Urban Review*, 45(5), 561-585.
- Jordan, W. J., Lara, J., & McPartland, J. M. (1996). "Exploring the causes of early dropout among race-ethnic and gender groups." *Youth & Society*, 28(1), 62-94.
- Lemon, S. D., Newfield, N. A., & Dobbins, J. E. (1993). "Culturally sensitive family therapy in Appalachia." *Journal of Systemic Therapies*, 12, 8-26.
- Leventhal-Weiner, R., & Wallace, M. (2011). "Racial differences in high school dropout rates: An analysis of U.S. Metropolitan areas." *Research in Social Stratification and Mobility*, 29(4), 393-413. doi:10.1016/j.rssm.2010.12.007
- Martin, E. J., Tobin, T. J., & Sugai, G. M. (2003). "Current information on dropout prevention: Ideas from practitioners and the literature." *Preventing School Failure: Alternative Education for Children and Youth*, 47(1), 10-17.
- McNeal Jr, R. B. (1997). "Are students being pulled out of high school? The effect of adolescent employment on dropping out." *Sociology of Education*, 206-220.
- Moser, E.B. (1989). "Exploring contingency tables with corespondence analysis." *CABIOS*, 5(3), 183-189.
- Owens, A. (2010). "Neighborhoods and schools as competing and reinforcing contexts for educational attainment." *Sociology of Education*, 83(4), 287-311.

- Pollard, K., & Jacobsen, L. A. (2015). *The Appalachian Region: A Data Overview from the 2009-2013 American Community Survey* Retrieved from <http://www.prb.org/pdf15/DataOverviewfrom2009-2013ACS.pdf>
- Rogers, C., Mencken, K., & Mencken, F. C. (1997). "Female labor force participation in Central Appalachia: A descriptive analysis." *Journal of Appalachian studies*, 3(2), 189-210.
- Rumberger, R. W. (1987). "High school dropouts: A review of issues and evidence." *Review of Educational Research*, 57(2), 101-121.
- Stanard, R. P. (2003). "High school graduation rates in the United States: Implications for the counseling profession." *Journal of Counseling & Development*, 81(2), 217-221.
- Stark, P., Noel, A. M., & McFarland, J. (2015). *Trends in High School Dropout and Completion Rates in the United States- 1972–2012* (NCES 2015- 015). Retrieved from Washington D.C.:
- Stearns, E., & Glennie, E. J. (2006). "When and why dropouts leave high school." *Youth & Society*, 38(1), 29-57.
- Steinberg, A., & Almeida, C. (2004). "The Dropout Crisis: Promising Approaches in Prevention and Recovery." *Jobs for the Future*.
- Sugar, M. (2002). *Regional identity and behavior*: Springer Science & Business Media.
- Swanson, C. B. (2004). *Who graduates? Who doesn't? A statistical portrait of public high school graduation, class of 2001*. Washington, DC: The Urban Institute.
- U.S. Census Bureau. (2014). *American Community Survey 1-Year Estimates, Table S1701*. Retrieved from

http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_1YR_S1701&prodType=table

van der Heijden P. G. M, de Falguerolles, A., & de Leeuw, J. (1989). "A Combined Approach to Contingency Table Analysis Using Correspondence Analysis and Log-Linear Analysis." *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 38(2), 249-292.

Table 1. Population characteristics of Appalachia/Non-Appalachian Regions of Kentucky

	Appalachia	Non-Appalachia	Kentucky	United States
Population	1,182,515	3,178,818	4,361,333	311,536,594
Population/sq. mile	64.9	149.5	110.5	88.2
Population under age 18	22.6%	23.7%	23.4%	23.7%
White, not Hispanic	95.3%	82.6 %	86.1%	63.3%
Less than high school diploma*	25.2%	13.9%	17.0%	14.0%
Unemployment rate**	9.4%	7.8%	8.2%	8.1%
Median household income (2013)	\$32,854	\$47,182	\$43,036	\$53,046
Population under age 18 in poverty	33.5%	23.4%	26.1%	21.6%
Disability status***	22.3%	13.1%	15.6%	10.1%
Did not move in past year	87.0%	84.0%	84.8%	84.9%

Data Source: Pollard & Jacobsen, 2015; U.S. Census Bureau, 2009-2013 American Community

Survey

*Population ages 25 and over

**Persons ages 25-64 in civilian labor force

***Population ages 18-64

Table 2. Categorization of dropout reasons into push-out and pull-out.*

Push-out Reasons	Pull-out Reasons
Boredom	Employment
Course selection	Family problems
Expulsion	Illness
Failing grade	Marriage
Teacher conflict	Pregnancy

*The category of “No substantial enrollment” was not included as a pull-out or push-out reason.

Table 3. Distribution of Region Dropout Frequencies and Percentages for Gender, Race, and Grade Categories

Variables	<u>Appalachia</u>		<u>Non-Appalachia</u>	
	(N=4,669)		(N=13,401)	
	<i>N</i>	%	<i>N</i>	%
Race				
White	4,164	89	9,200	69
African-American	381	8	3,046	23
Other	124	3	1,155	9
Gender				
Female	1,889	40	5,581	42
Male	2,780	60	7,820	58
Grade				
9th	739	16	2,723	20
10th	1,292	28	3,796	28
11th	1,559	33	3,825	29
12th	1,079	23	3,057	23

White = White, non-Hispanic;

African-American = African-American, non-Hispanic;

Other = Races other than White and African-American (e.g., Hispanic, Asian, Native American)

Table 4. Dropout Category Frequencies and Percentages by Region, Gender, Grade, and Race Combination.

	Appalachia						Non-Appalachia					
	<u>Pull-Out</u>		<u>Push-Out</u>		<u>Unknown</u>		<u>Pull-Out</u>		<u>Push-Out</u>		<u>Unknown</u>	
	(N=707)		(N=563)		(N=619)		(N=1,743)		(N=2,059)		(N=1,779)	
Female	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
9th Grade												
White	58	30	61	32	73	38	156	24	279	43	220	34
African-American	2	10	6	30	12	60	31	13	83	36	117	51
Other	1	14	1	14	5	71	25	23	38	35	46	42
10th Grade												
White	177	41	148	34	105	24	354	34	429	42	247	24
African-American	2	8	7	29	15	62	81	21	131	34	173	45
Other	4	27	4	27	7	47	51	38	38	28	46	34
11th Grade												
White	269	40	201	30	206	30	473	39	458	38	273	23
African-American	1	4	4	17	18	78	61	18	143	42	138	40
Other	1	11	2	22	6	67	60	39	50	33	42	28
12th Grade												
White	187	40	122	26	154	33	367	37	312	31	321	32
African-American	3	19	4	25	9	56	50	20	76	31	119	49
Other	2	14	3	21	9	64	34	37	22	24	37	40
	<u>Pull-Out</u>		<u>Push-Out</u>		<u>Unknown</u>		<u>Pull-Out</u>		<u>Push-Out</u>		<u>Unknown</u>	
	(N=643)		(N=1,243)		(N=894)		(N=1,434)		(N=3,815)		(N=2,571)	
Male	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
9th Grade												
White	75	18	226	53	122	29	128	12	656	60	318	29
African-American	0	0	13	15	74	85	13	3	150	33	286	64
Other	2	20	4	40	4	40	32	18	62	35	83	47
10th Grade												
White	181	26	347	49	177	25	276	18	909	60	318	21
African-American	3	3	25	28	61	69	44	8	215	40	279	52
Other	4	14	10	34	15	52	45	22	90	44	70	34
11th Grade												
White	224	30	344	46	185	25	410	27	786	52	313	21
African-American	6	8	26	35	43	57	37	8	193	43	214	48
Other	6	26	7	30	10	43	50	29	64	37	60	34

12th Grade												
White	139	27	207	40	176	34	332	28	512	43	353	29
African-American	0	0	26	55	21	45	52	13	136	33	224	54
Other	3	18	8	47	6	35	15	14	42	38	53	48

Note: Percentages are for the total number of students within each region, gender, grade, and race combination.

White = White, non-Hispanic;

African-American = African-American, non-Hispanic;

Other = Races other than White and African-American (e.g., Hispanic, Asian, Native American)

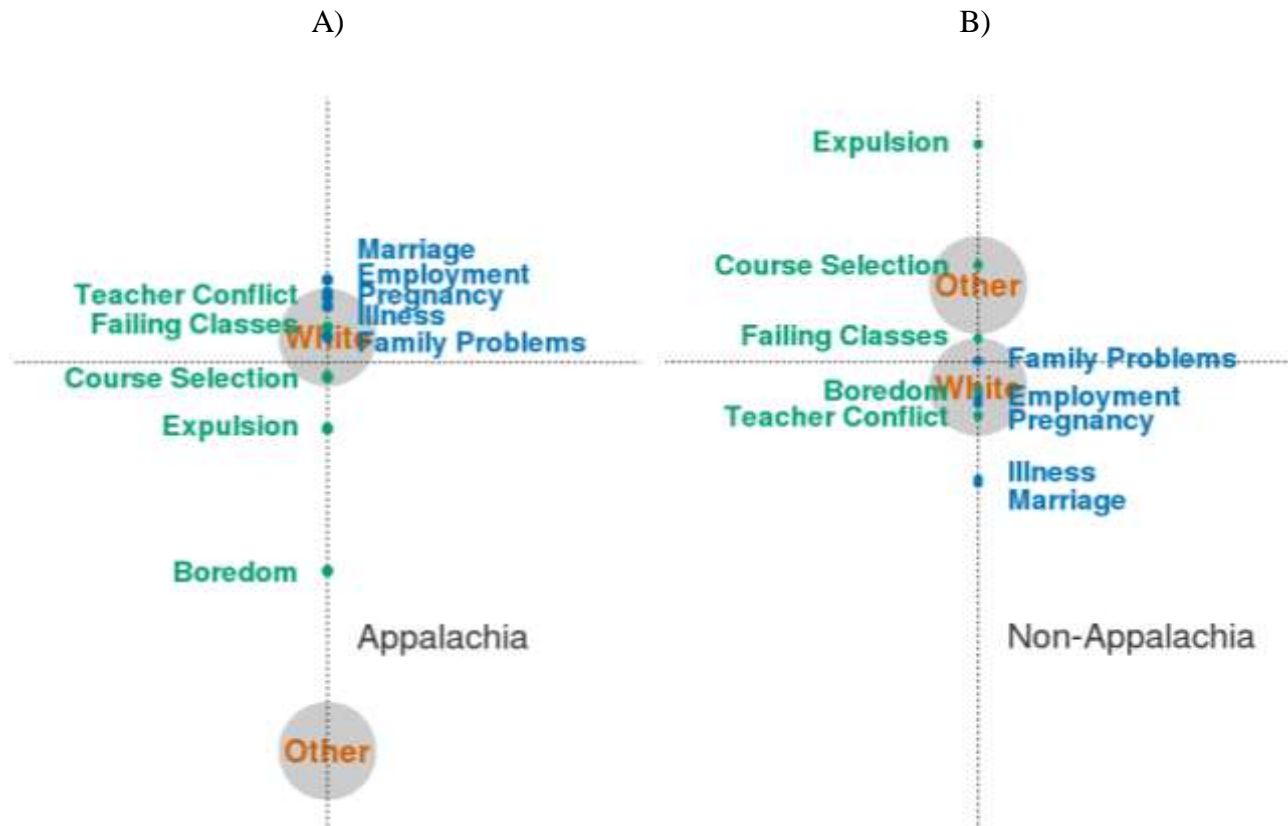


Figure 1. Correspondence Analysis Plot of Dropout Reason by Race for Region (Green = Push-out reasons, left-hand side; and Blue = Pull-out reasons, right hand side)

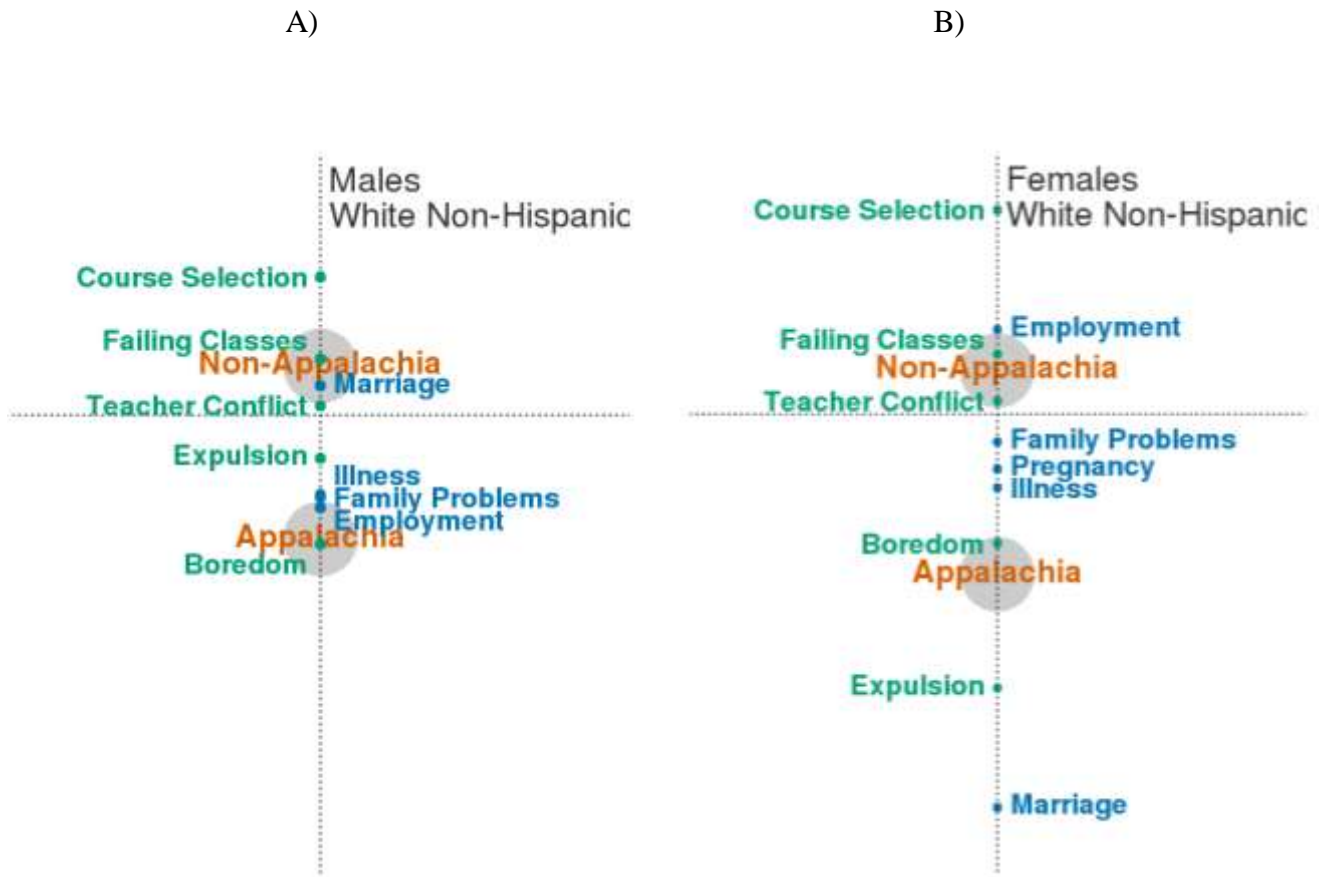


Figure 2. Correspondence Analysis Plot of Dropout Reasons by Region for Male and Female 9th Grade Students (Green = Push-out reasons, left-hand side; and Blue = Pull-out reasons, right hand side)