# Stage 4: Final Results, Evaluation, Reflections

Joya Wheatfall-Melvin

PPOL 5204: Data Science II

Professor Ioannis Ziogas

April 16, 2025

**Research Design & Methodology:** Discuss your analytical plan.

*Alternative Schools SCH_STATUS_ALT*: is a public elementary or secondary school that addresses the needs of students that typically cannot be met in a regular school program and is designed to meet the needs of students with academic difficulties, students with discipline problems, or both students with academic difficulties and discipline problems. Alternative education schools may be sited in locations other than a traditional school building, such as hospitals, mental health centers, jails, or juvenile detention centers.

- While alternative schools serve as a transitionary period in one's educational life (some students stay for days, others for years), they still have a significant effect on one's educational attainment and/or educational perspective. The U.S. Government Accountability Office finds that “certain groups of students attend alternative schools in greater proportions than they do other schools”, specifically black girls and boys.

**“How do school-provided opportunities and climate conditions shape student outcomes for Black girls in alternative schools?”**

**Methodology:**

I. **Characterizing Data - 2017/18 CRDC Data**
    A. Filtering the main dataset for Alternative Schools (Academic, Discipline, and Both)
    B. Storing demographic variables in a list
    C. Identifying merging key ‘ COMBOKEY’
II. **Retrieving Target Variables: Negative Implications of Alternative. Schooling**
    A. Retention (6-12)
        1. Black Female Retention
        2. Total Retention
    B. Black Female Enrollment Population
        1. Overall Abled Black Female Enrollment
        2. Disabled (Under IDEA) Black Female Enrollment
        3. Disabled (Under 504) Black Female Enrollment
        4. Total Enrollment (Male)
        5. Total Enrollment (Female)
    C. Harm Variables
        1. Corporal Punishment *dropped*
        2. Expulsion
        3. Harassment and Bullying
        4. Referrals and Arrests
        5. Restraint and Seclusion

6. Suspensions

**III. Retrieving Feature Variables: School-Provided Opportunity and Structure**

A. Black Female Participation in Programs
   1. Advanced Math *dropped*
   2. Dual Enrollment Program Indicator and Black Female Participation
   3. Gifted and Talented Program Indicator and Black Female Participation
   4. International Baccalaureate Indicator and Black Female Participation
   5. ACT/SAT Test Black Female Participation

B. School-Safety Indicators
   1. Offense Variables (Rape, Batttery, Robberry, Physical Attacks, etc.)

C. Funding Allocation Indicators
   1. Salary expenditure for Total Personnel (funded by the gov.)
   2. Salary expenditure for Teachers (funded by the gov.)
   3. Salary expenditure for Administrators (funded by the gov.)
   4. Salary expenditure for all Support Services staff (funded by the gov.)

D. School-Support Indicators
   1. Count of counselors
   2. Count of social workers
   3. Count of psychologists
   4. Count of nurses
   5. Count of law enforcement officers
   6. Count of security guards
   7. Total count of teachers
   8. Count of teachers not certified

**IV. Pre-Processing Data**

A. Data Cleaning / Aggregation
   1. For Binary indicators, replaced missing values with ‘No'; assuming the incident did not happen
   2. MICE imputation for missing School-Support Indicators; likely that missing values does not indicate no absolute funding
   3. Imputed 0 for all missing Harm, Offense, Personnel count, and Black Female participation values; not trying to “create” a narrative by introducing bias

B. Feature Engineering
   1. Target_df
      a) Normalized Retention by BF Retention / Total Retention per school
      b) Normalized all Harm Variables by BF Harm / Total Harm Type (e.g., BF # of Expulsion / Total # of Expulsions)

c) Combined into all variables into composite harm flag; 10% threshold
   (1) **715 Alt. Schools (=41%) had more than 10% of their Black girls experience harm.**
   (2) **1033 Alt. Schools (=59%) had 10% or fewer of their Black girls experience harm.**

2. Feature_df
   a) Normalize Offense variables by Total Population; rate relative to school size.
   b) After MICE Salary Imputation, normalize Salary by Total Population; salary per student.
      (1) Applied log transformation to reduce skew
   c) Normalize Personnel count by Total Population; rate relative to school size.

## V. Model Selection

A. Logistic Regression
   1. Can interpret the data by showing *how* certain factors influence retention for black female students
   2. However, it could be sensitive to multicollinearity (I believe a lot of my features may be correlated)

B. Random Forest
   1. Information-based classifier
   2. Most necessary in answering *which* factors most influence retention for black female students

C. XGBoost
   1. Good at modelling mixed features in my data: binary indicators, proportions, and log-transformed counts
   2. Great at handling class imbalance and missing values
   3. High performance with small to medium datasets, perfect for my small dataset (n=1748)

## VI. Output Interpretation

***Results:*** *At this point, your results should be polished and your analysis complete. Provide an update to Stage 3 if necessary.*

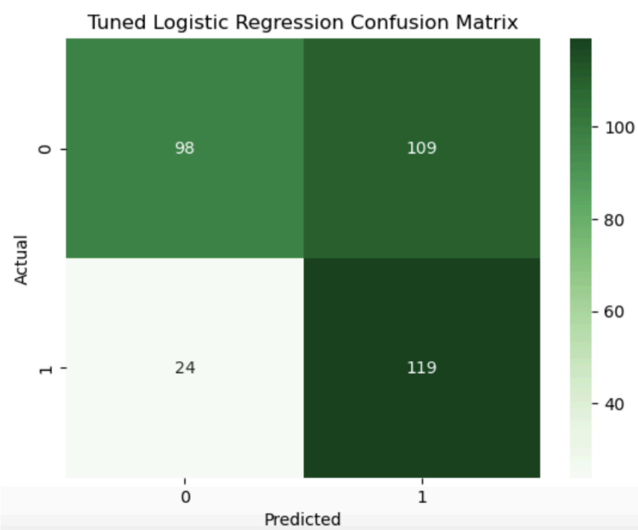
- **Logistic Regression:** Before tuning for hyperparameters, the model correctly predicts 59% of school environments that are not harmful to black girls and 39% of the school environments that are. After tuning for hyperparameters, the model improves significantly by correctly predicting 80% of school environments that are not harmful and 53% of school environments that are. The AUC and recall improve, meaning that the model is more sensitive to detecting harmful environments in schools. There are fewer false negatives, meaning the model is identifying more harmful schools, and higher true positives, meaning the model is better at detecting harmful cases. *Identifying harmful environments is the goal, and this model is moderately lower in detecting harmful events compared to the other models.*

```
Post-Tuning Logistic Regression Performance:
              precision    recall  f1-score   support

           0       0.80      0.47      0.60       207
           1       0.52      0.83      0.64       143

    accuracy                           0.62       350
   macro avg       0.66      0.65      0.62       350
weighted avg       0.69      0.62      0.61       350

Accuracy: 0.62
AUC Score: 0.6319
```

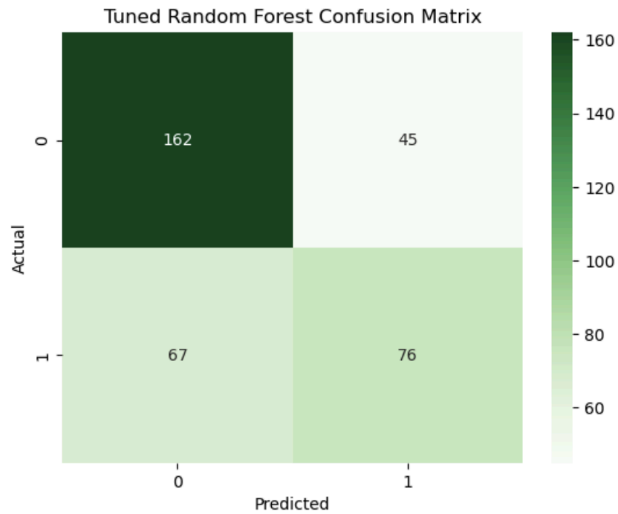



- **Random Forest:** Before tuning for hyperparameters, the model correctly predicts 69% of school environments that are not harmful to black girls and 63% of the school environments that are. After tuning for hyperparameters, the model correctly predicts 71% of school environments that are not harmful and 63% of school environments that are. The AUC and recall improve, meaning that the model is more sensitive to detecting harm in schools and has fewer overfit trees. There are fewer false negatives, meaning the model is identifying more harmful schools, and higher true positives, meaning the model is better at detecting harmful cases. *Since identifying harmful environments is the goal, this model is the best at catching them.*

Post-Tuning Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.71 | 0.78 | 0.74 | 207 |
| 1 | 0.63 | 0.53 | 0.58 | 143 |
| accuracy | | | 0.68 | 350 |
| macro avg | 0.67 | 0.66 | 0.66 | 350 |
| weighted avg | 0.68 | 0.68 | 0.67 | 350 |

Accuracy: 0.68



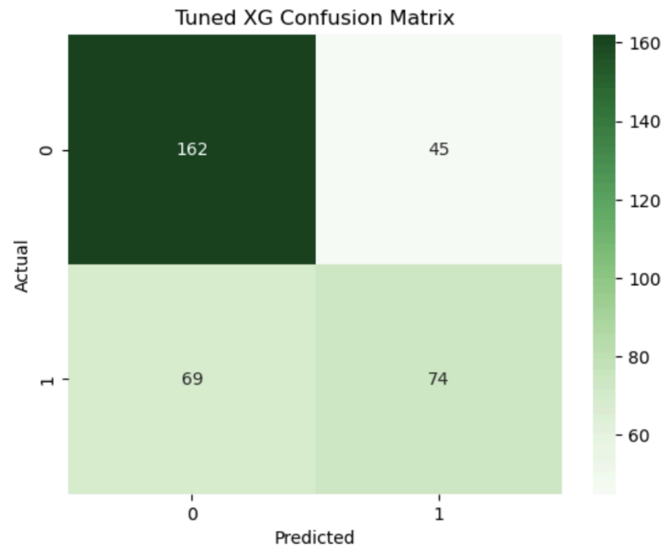

- **XGBoost:** Before tuning for hyperparameters, the model correctly predicts 71% of school environments that are not harmful to black girls and 58% of the school environments that are. After tuning for hyperparameters, the model correctly predicts 70% of school environments that are not harmful and 62% of school environments that are. The AUC increased to 75.8%, indicating the strongest model's ability to discriminate between harmful and non-harmful school environments. The improved recall and precision for the positive class reflect a better balance between identifying risk and avoiding false alarms. In comparison to other models, the XGBoost performs the best even under class imbalance and small sample size.

```
Post-Tuning Classification Report:
              precision    recall  f1-score   support

           0       0.70      0.78      0.74       207
           1       0.62      0.52      0.56       143

    accuracy                           0.67       350
   macro avg       0.66      0.65      0.65       350
weighted avg       0.67      0.67      0.67       350

Accuracy: 0.6742857142857143
```




```
AUC Score: 0.7589
Tuned XGB R²: -0.3479
Tuned XGB RMSE: 0.5707
```

-

***Evaluation:*** *This section should provide the overall evaluation of your selected methodological approach. It should provide comprehensive answers to the following questions:*

*Performance:* *How did your models perform?*

Overall, the **Logistic Regression** models initially favored the majority class and struggled to identify harmful environments. Whereas, the **Random Forest** did a more balanced job at predicting harmful environments, with more realistic feature importances. The **XGBoost** proved to be the strongest overall, achieving the highest AUC score, and it demonstrated greater sensitivity in detecting harmful school environments. After hyperparameter tuning, it became the most robust model for identifying schools with harmful conditions for Black girls in alternative schools.

I honestly expected the results to be worse given the small sample size (n=1748) and the high volume of imputed 0s to handle missing data. However, the results exceeded my expectations, which means that my engineered features captured meaningful patterns despite data sparseness.

*<u>Are you satisfied with your approach? Do you have reservations regarding your methodological choices?</u>*

- One reservation in my preprocessing stage was dealing with the missing values. After realizing how sparse this data was, I did not want to ”create” a narrative by imputing certain harm indicators that were missing. So, I ended up filling in missing values in harm and offense variables as 0s, assuming they may have been underreported or the harm action simply did not happen. This was a virtuous act, but I know it had some negative effects on the predictive accuracy of my models. Inputting so many 0s skewed a lot of the variable distribution and may have had some effects on the accuracy of my models.
- Overall, my methodological approach to feature engineering and normalizing all variables in both target_df and feature_df before training my Logistic Regression model was necessary, but unnecessary in modelling the Random Forest and XGBoost models. I kept those normalized values and proportions and included them in both classification models, due to the time investment in creating those variable proportions, but overall should not have caused any harm in the classification models, and should have helped coefficient interpretation. I do wonder about the differences in precision, recall, and splits if I just included the raw, unnormalized data in those models.

*<u>Answers:</u> After implementing your plan, did you adequately address your research question? Are your answers to your satisfactory? Is your analysis contributing to our overall understanding of your chosen topic? Justify your answers.*

I believe my models accurately provided guidance in addressing: **“How school-provided opportunities and climate conditions shape student outcomes for Black girls in alternative schools?”.** All provided school environment features that were influential in predicting harmful school environments for Black girls in alternative schools included indicators of school safety (e.g., rated of physical offenses, attempted violence, and robbery), staffing resources (e.g. full-time teachers per student, presence of uncertified staff), and access to academic opportunities.

Features such as higher teacher-to-student ratios and academic engagement were associated with reduced harm, while high offense rates and limited supportive staffing were linked to increased harm. These findings are consistent with broader literature on structural inequities in disciplinary policy and education resource allocation; these results reinforce the idea that carceral climate and opportunity are measurable and predictive risk factors for harm exposure among Black Girls leaning in these spaces.

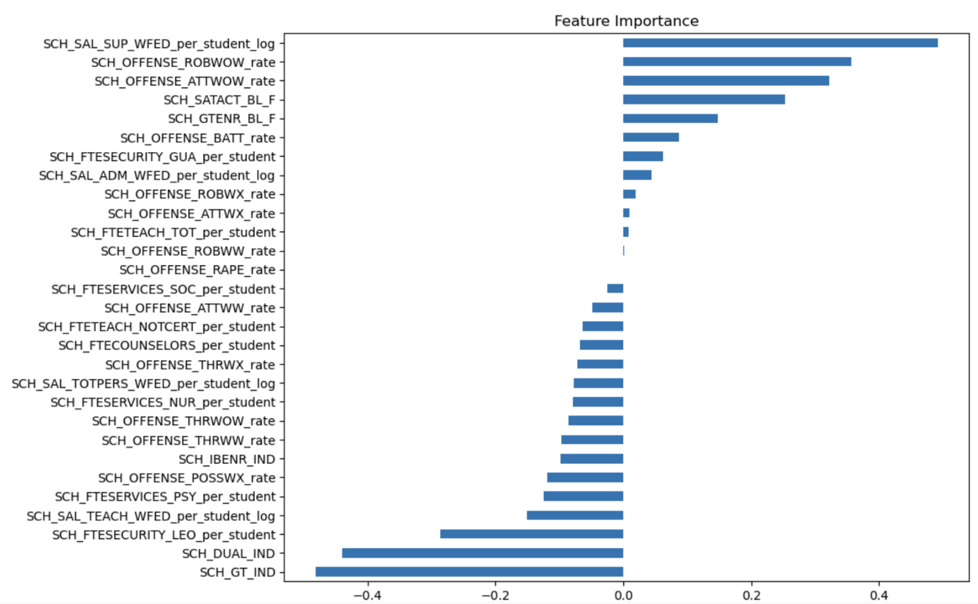
This analysis contributes to our understanding of how machine learning can be used to uncover patterns in current funding allocations and resource distribution in alternative schools and how these patterns affect outcomes for specific demographics, like Black girls. Evaluating the integrity of alternative schooling is inherently challenging due to the high concentration of marginalized students, deeply embedded systemic barriers, and the limited resources often allocated to these institutions. Because alternative schools are typically designated for students facing academic and/or disciplinary challenges, establishing standardized metrics to assess their integrity is both complex and nuanced. This study offers a beginning framework to begin quantifying those conditions and guiding future education policy metrics.

***Reflections:*** *Offer your final thoughts on this exercise, such as:*

*Learning Curve vs Learning Objectives:* My biggest learning curve dealt with feature engineering and making count data understandable and comparative in this context. Continuous data can be messy and has to be transformed correctly to have sound model interpretability. I believe the bulk of my time was spent on pre-processing and analyzing how to deal with count data, creating thresholds, and deciding whether to transform target variables into dichotomous or binary indicators. Also, taking into account the small population, addressing missing data, and what counts as “significant” harm to black female educational attainment were hard factors to consider as well when creating a composite harm indicator.

*Substantive Outcomes:*

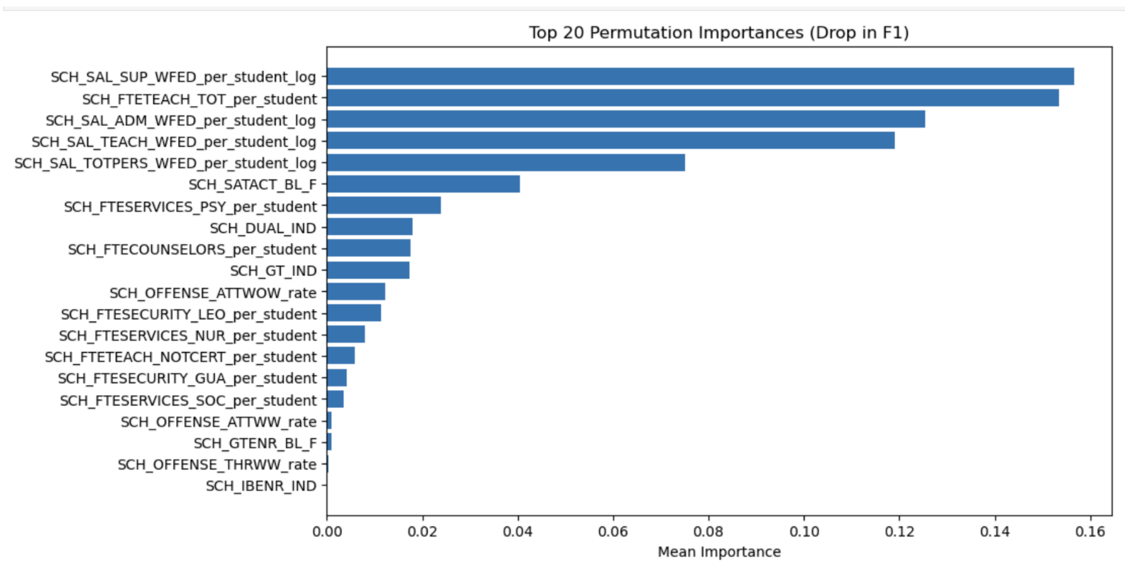
1. **Logistic Regression:**




- Positive Coefficients show features that increase the likelihood of a school being classified as harmful to Black girls. Negative Coefficients show features that decrease the likelihood of harm. Overall, these features surprised me, but I believe they don't hold weight due to how badly the model performed:

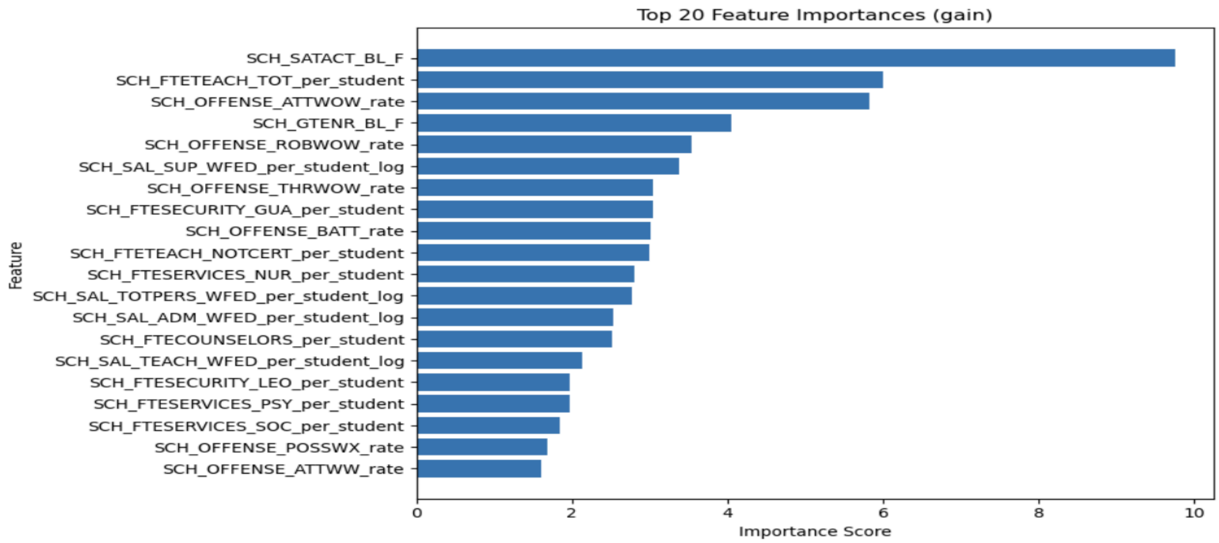
- Top Harm-Predicting Features:
  - Higher support staff salary per student (surprising)
  - Higher rate of robbery with a weapon (without injury) in school
  - Attempted robbery (without injury)
  - More Black girls' participation in SAT/ACT testing (surprising)
  - Higher enrollment of Black girls in the gifted program
- Top Protective Features:
  - Gifted Program programs
  - Dual Enrollment programs
  - Higher staff investment per student (security, psychology, teachers, nurses)

2. **Random Forest:**




- In identifying the 'Top 20 Features from Permutation Importance', higher support services staff salary per student is strongly and negatively associated with harm; I expect this because it suggests that more investment in support staff may influence more equitable protocols and protect the well-being of black girls in alternative schools.
- More teachers per student are also a significant predictor of lesser harm, suggesting that more instructional support may have the potential to provide a caring educational experience for black girls.
- Additionally, Black female students SAT/ACT and Dual Enrollment participation is a significant predictor of less harm towards Black female students in alternative schools.

3. **XGBoost:**




- Reflects how much each feature contributes to the model, based on gain.:
    - Top Predictors of Harmful School Environments:
        - Higher Black girl participation in the SAT/ACT might be negatively associated with harm, or reflect supportive academic environments that reduce harm indicators.
        - More teachers per student are a strong predictor, meaning that better staffing reduces harm.
        - High attempted offenses without weapons are a top harm signal; high offense rates correlate with school safety issues, adding to harm.

*Future directions or research avenues:*

Future research should expand this analysis beyond alternative schools to examine all school environments nationally that contribute to harm experienced by Black girls. While this study counts harmful instances that may have affected the 30,728 Black girls enrolled in alternative education settings, broader analysis could help identify systemic patterns across all school types. Additionally, integrating longitudinal data could offer deeper insight into how school conditions shape student outcomes over time and how interventions may mitigate harm. Further work might also explore the intersectionality of race, gender, and disability status to better understand compounding vulnerabilities in educational settings, in addition to carceral conditions.

*What have you learned? :*

This exercise underscored the critical importance of proper data preprocessing and the significant time investment required to engineer features that yield interpretable, meaningful outcomes. I learned how crucial it is to handle missing data thoughtfully; imputation methods must be chosen carefully to avoid introducing bias, particularly when accounting for harmful incidents towards a specific demographic. Another major insight was recognizing the impact of the COVID-19 pandemic on data quality and collection. I initially relied on 2020–2021 CRDC data to capture the most recent conditions, but it turned out to be extremely sparse, without many essential variables like school funding and staffing counts. This challenge highlighted how external disruptions like a global pandemic (and/or governmental interference in limiting DOE funding) can significantly affect the reliability of datasets and the stability of integral institutions like the Civil Rights Data Collection. I realized that using the more complete 2017–2018 dataset was not only more practical but essential for producing robust models and reliable insights about Black girls' education experiences. This experience has thoroughly deepened my appreciation for understanding the context behind data availability, as well as data visibility.

**References**

Anderson, M. D. (2024, August 13). *How Black Girls Are Harmed in Schools*. Education Writers Association. Retrieved February 2, 2025, from https://ewa.org/data-research-tips/how-black-girls-are-harmed-in-schools

Cabral, B., Annamma, S. A., & Morgan, J. (2023). “When You Carry a Lot”: The Forgotten Spaces of Youth Prison Schooling for Incarcerated Disabled Girls of Color. Teachers College Record, 125(5), 95-113. https://doi.org/10.1177/01614681231181816

Hartocollis, A. (2024, December 16). *Black Student Enrollment at Harvard Law Drops by More Than Half*. The New York Times. Retrieved February 2, 2025, from https://www.nytimes.com/2024/12/16/us/harvard-law-black-students-enrollment-decline.html

National Center for Education Statistics. (n.d.). *Fast Facts: Degrees conferred by race/ethnicity and sex (72)*. National Center for Education Statistics. Retrieved February 2, 2025, from https://nces.ed.gov/fastfacts/display.asp?id=72

U.S. Department of Education. (n.d.). *Civil Rights Data Collection Office for Civil Rights*. Civil Rights Data | U.S. Retrieved February 2, 2025, from https://civilrightsdata.ed.gov/