Racial and Health Disparities on Academic Outcomes

My Le'25
Yaksh Toyesh Ujoodha'25

Context

The School District of Beloit is diverse with students coming from various racial, ethnic, and socio-economic backgrounds. The data shows an **intersection between student race and health level and academic outcomes**.

Research Questions

- Are there significant differences in academic performance between students of different racial/ethnic backgrounds?
- Do health disparities, including serious health issues, contribute to these academic differences?

Why does it matter?

The intersection of race and health is a complex issue with significant implications for academic outcomes. *Research suggests that racial disparities in health can lead to disparities in educational attainment. This case study aims to investigate the specific relationship between race, health, and academic performance.

(*): Montez, J. K., & Cheng, K. J. (2022). *Educational disparities in adult health across U.S. states: Larger disparities reflect economic factors*. Frontiers in public health, 10, 966434. https://doi.org/10.3389/fpubh.2022.966434

Data Cleaning Process

We use Python in Google Colab and Jupyter Notebook for this project:

- Convert A,B,C,D to meaningful values
- Keep scores for Winter 2018-2019 only and removed rest of semesters
- Removed NA values for regression
- One-hot encoding of categorical variables
- Convert regressor features to Float data type

About the dataset

Student Demographics	Academic Information
• Student ID • Age, Sex, DOB (Date of Birth) • **Local Race, His/Lat Ethnicity, Multi-Race, Fed Race** • Special Education and Disability Description • LEP Designation • Lunch Code possibly indicating eligibility for free/reduced lunch. • Rental, Property Value, Rent Price • Income level • **Health level**	• Old School, Entity • Gr (Grade) • Exam • **Test scores (e.g., math and reading scores), including RIT scores and percentages for different seasons (Fall, Winter, Spring)**

About the dataset

Student Demographics	Academic Information
• Majority Group • child_health_level_very_good_compare • parent_relationship • bool_child_access_health_care • bool_child	• WT18_19MathRIT • WT18_19ReadRIT • Old School • Special Education • Eng_Prof_Full

Overview

Dataset statistics

Number of variables	52
Number of observations	221
Missing cells	1025
Missing cells (%)	8.9%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	89.9 KiB
Average record size in memory	416.6 B

Variable types

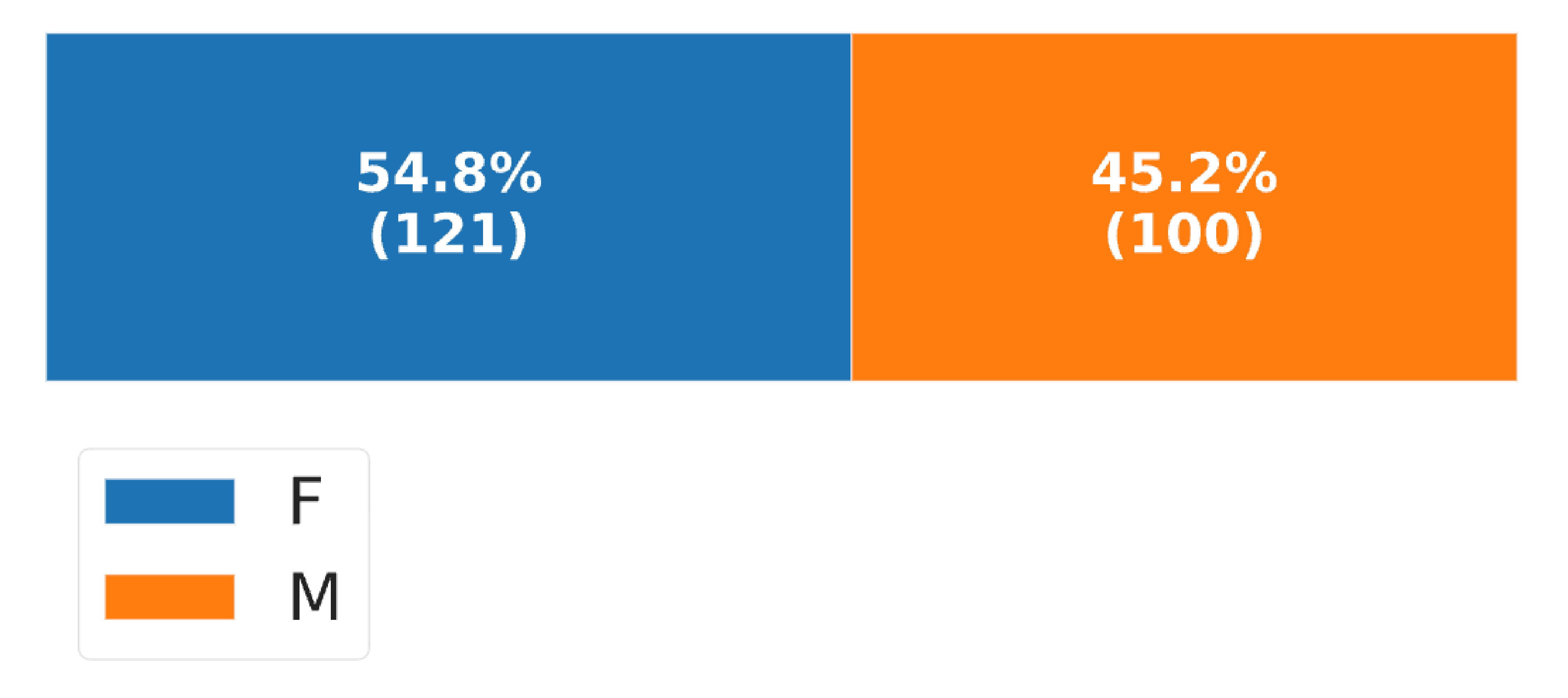
Numeric	5
Categorical	35
Unsupported	7
Boolean	4
Text	1

Students by School

Common Values

Value	Count	Frequency (%)
MCNEEL INTERMEDIATE SCHOOL	67	30.3%
FRUZEN INTERMEDIATE SCHOOL	57	25.8%
CUNNINGHAM INTERMEDIATE SCHOOL	55	24.9%
ALDRICH INTERMEDIATE SCHOOL	42	19.0%

Gender



Income level

Value	Count	Frequency (%)
At or less than $22,459	57	25.8%
$22,460 - $30,451	42	19.0%
At or more than $110,372	25	11.3%
$38,444 - $46,435	18	8.1%
$62,420 - $70,411	14	6.3%
$46,436 - $54,427	11	5.0%
$54,428 - $62,419	10	4.5%
$78,404 - $86,395	10	4.5%
$70,412 - $78,403	8	3.6%

- The distribution is skewed to the left, meaning that a larger proportion of households have lower incomes.
- The highest frequency in income range signals a significant number of households have relatively low incomes.

Mother Education Level

High School diploma or GED	71	32.1%
Some College	49	22.2%
Less than High School	45	20.4%
Associate Degree	26	11.8%
Bachelor Degree	12	5.4%
NULL	5	2.3%
Master Degree, PhD, JD, MD or other advanced degree	3	1.4%

Father Education Level

High School diploma or GED	76	34.4%
Less than High School	61	27.6%
Some College	46	20.8%
Associate Degree	9	4.1%
Bachelor Degree	7	3.2%
Master Degree, PhD, JD, MD or other advanced degree	4	1.8%

Distribution of Race

Value	Count	Frequency (%)
White	102	46.2%
American Indian or Alaskan Native	65	29.4%
Black or African American	53	24.0%
Asian	1	0.5%

Hispanic

Value	Count	Frequency (%)
False	137	62.0%
True	84	38.0%

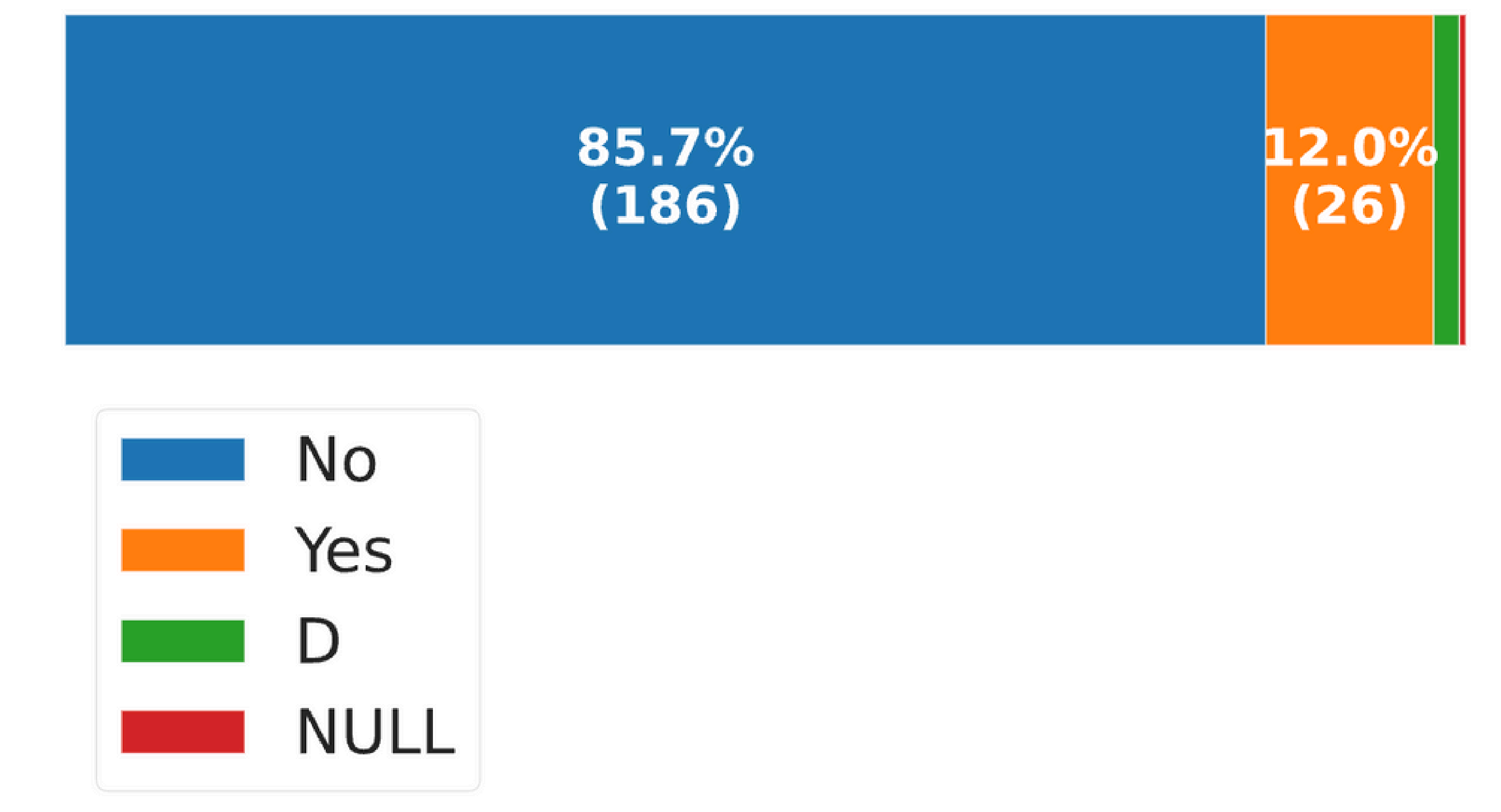
English Proficiency

Value	Count	Frequency (%)
7-Fully English Proficient	156	70.6%
3-Intermediate	28	12.7%
2-Beginning/Production	19	8.6%
4-Advanced Intermediate	8	3.6%
1-Beginning/Preproduction	8	3.6%
5-Advanced	1	0.5%
6-Was LEP/Now Eng Prof	1	0.5%

Health Level

Very Good	135	61.1%
Good	63	28.5%
Average	10	4.5%
NULL	9	4.1%
Very poor	2	0.9%
Poor	2	0.9%

Serious Health Issue

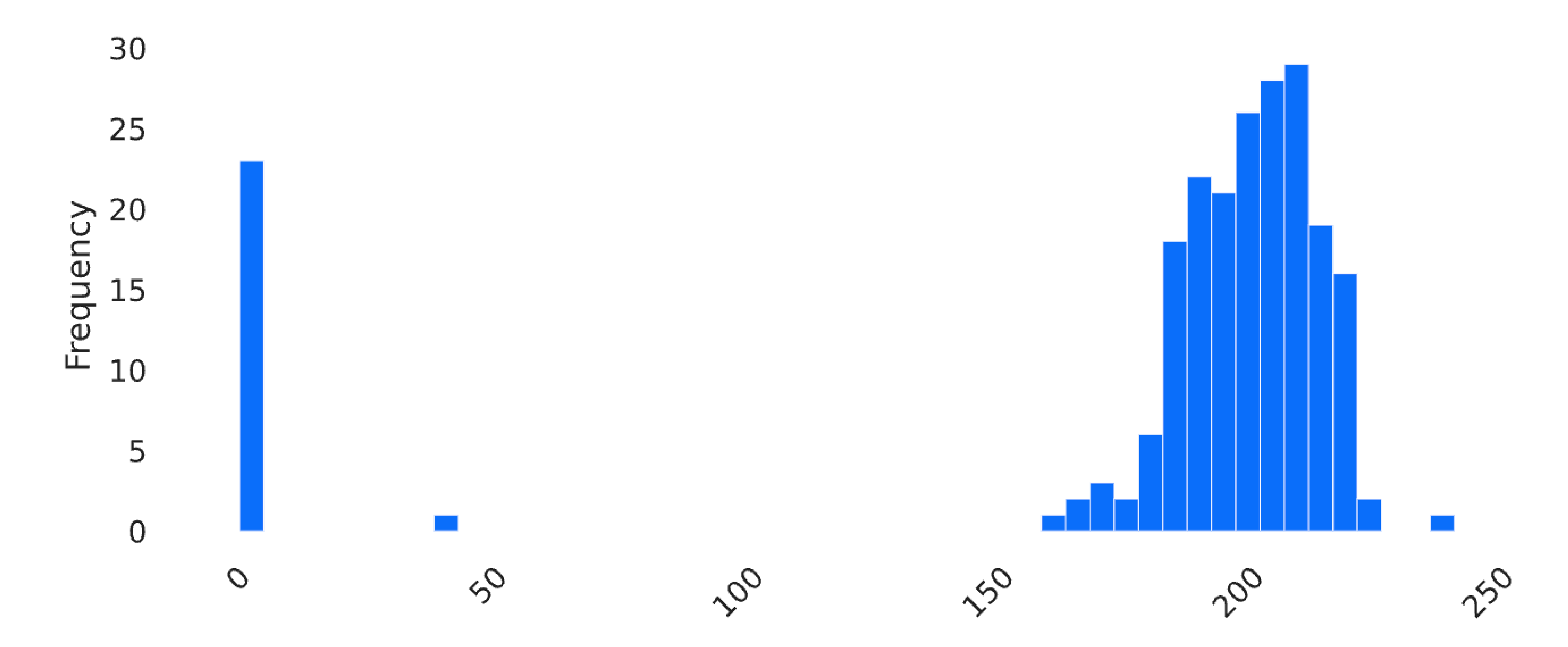


22 students who need Special Education

	Disability
0	183
Speech & Language	19
Other Health Impairment	7
SPECIFIC LEARNING DISABILITY	5
EMOTIONAL BEHAVIORAL DISABILITY	4
Traumatic Brain Injury	1
Significant Development Delay	1

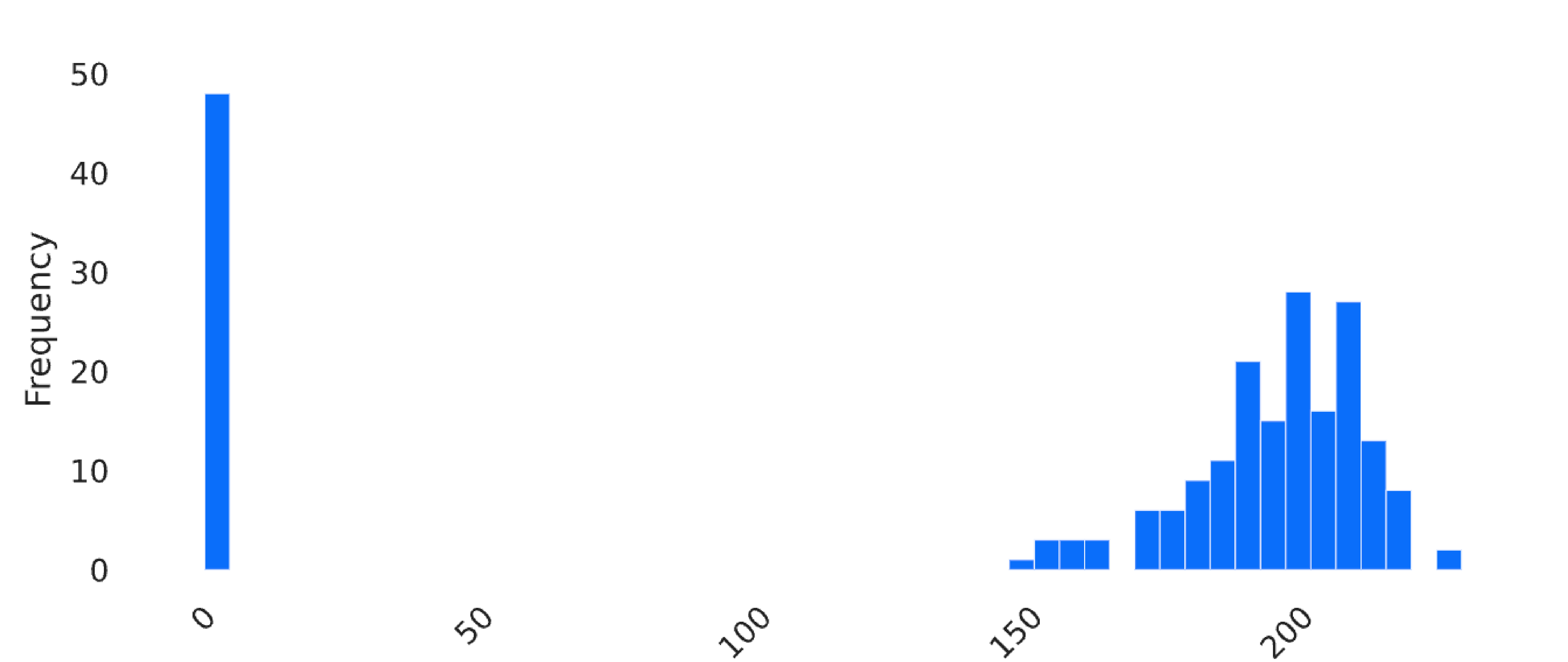
WT18_19 MathRIT

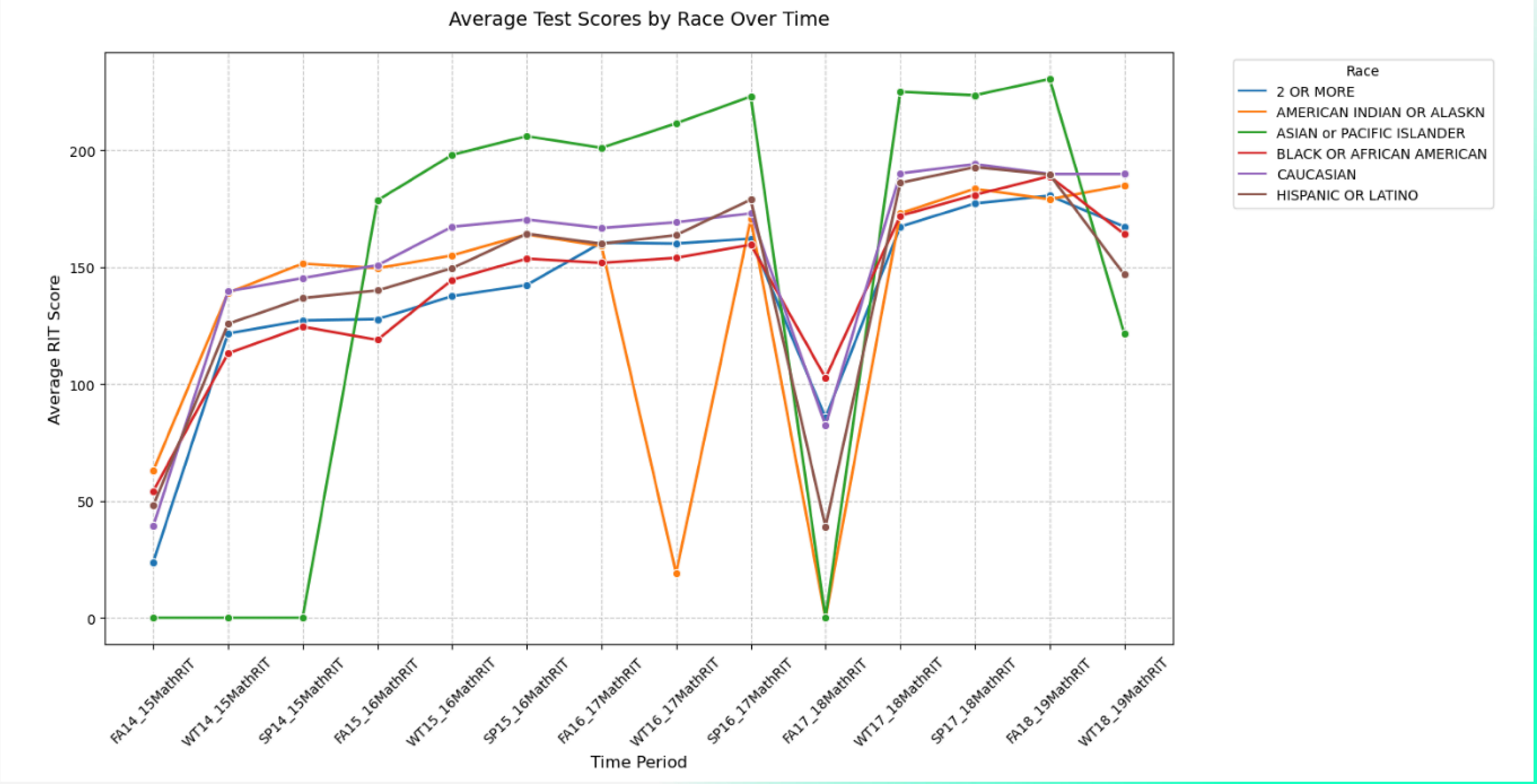
Mean: 180.64545
95-th percentile: 221
Skewness: -2.3648218



WT18_19 ReadRIT

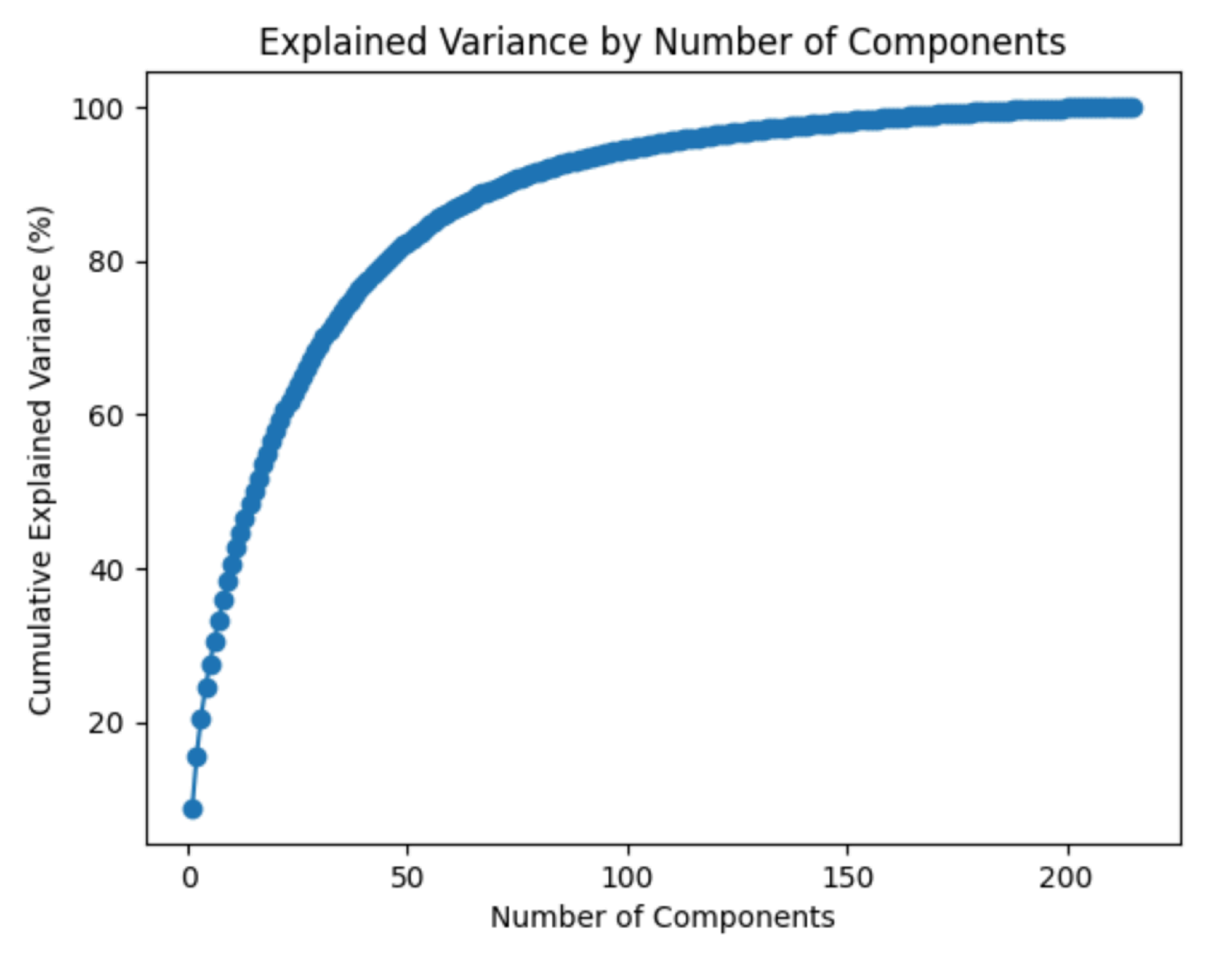
Mean: 155.02727
95-th percentile: 218
Skewness: -1.279984





Average Test Scores by Race Over Time
Average RIT Score
Time Period
0
50
100
150
200
FA14_15MathRIT
WT14_15MathRIT
SP14_15MathRIT
FA15_16MathRIT
WT15_16MathRIT
SP15_16MathRIT
FA16_17MathRIT
WT16_17MathRIT
SP16_17MathRIT
FA17_18MathRIT
WT17_18MathRIT
SP17_18MathRIT
FA18_19MathRIT
WT18_19MathRIT
Race
2 OR MORE
AMERICAN INDIAN OR ALASKN
ASIAN or PACIFIC ISLANDER
BLACK OR AFRICAN AMERICAN
CAUCASIAN
HISPANIC OR LATINO

Principal Component Analysis (PCA)



Based on Lasso regression

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Regression model:

$$WT1819MathRIT = \alpha + \beta_1 ChildHealthLevelGood + \beta_2 ChildHealthLevelNull + \beta 3 ChildHealthLevelPoor + \beta 4 ChildHealthLevelVeryPoor + \beta 5 FedRaceAsian + \beta 6 FedRaceBlack + \beta 6 FedRaceWhite + \epsilon$$

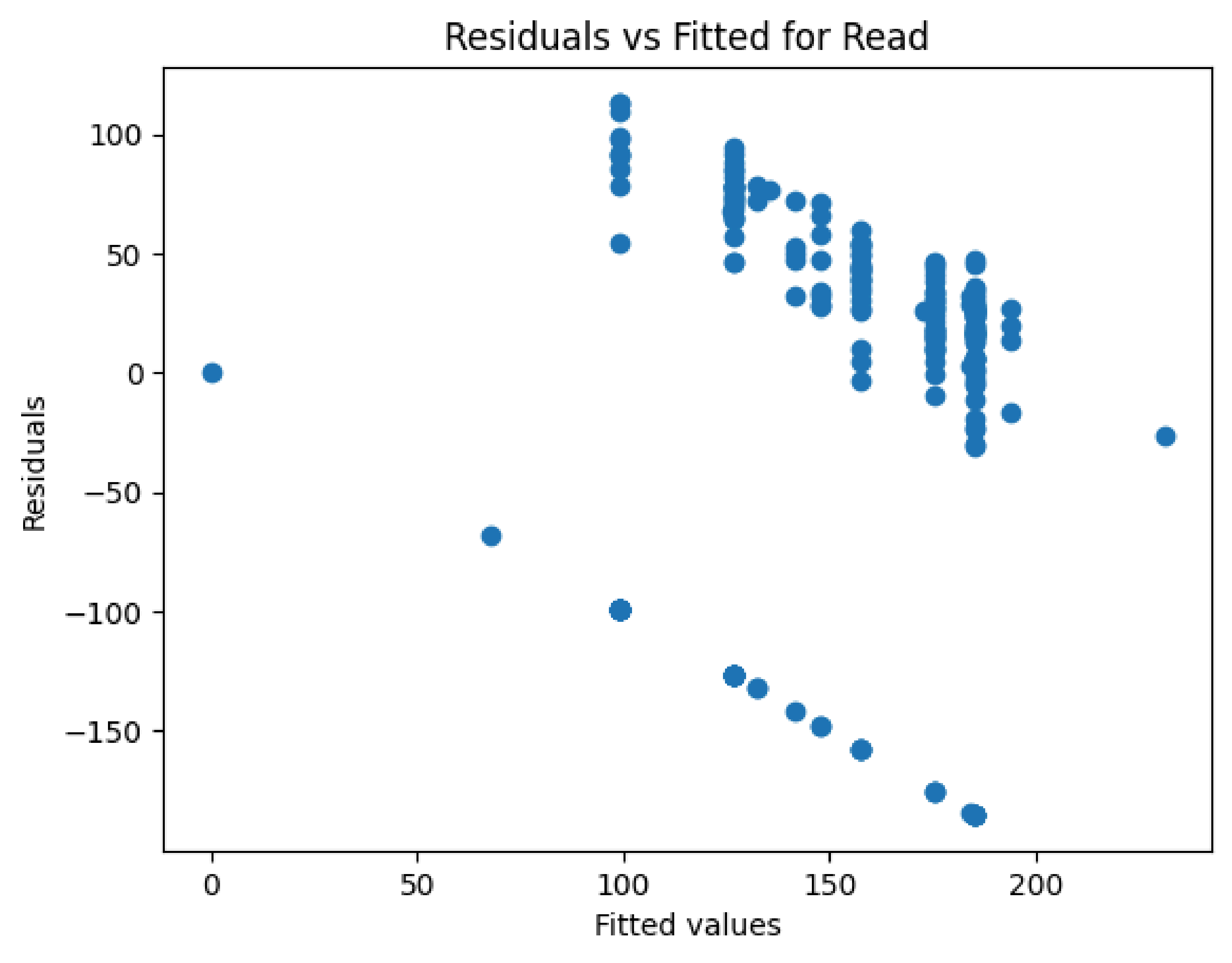
$$WT1819ReadRIT = \alpha + \beta 1 ChildHealthLevelGood + \beta 2 ChildHealthLevelNull + \beta 3 ChildHealthLevelPoor + \beta 4 ChildHealthLevelVeryPoor + \beta 5 FedRaceAsian + \beta 6 FedRaceBlack + \beta 6 FedRaceWhite + \epsilon$$

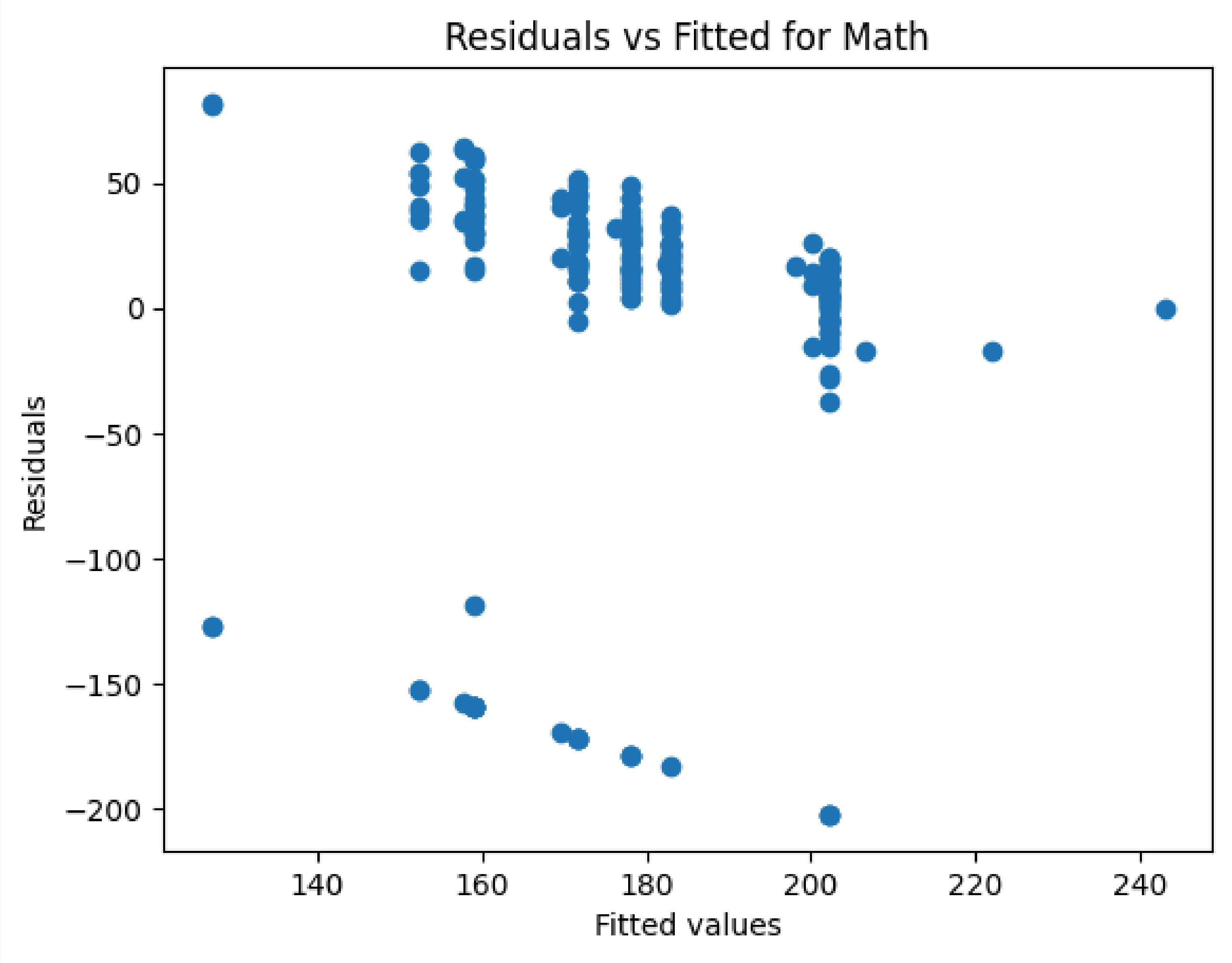
We also use Robust Standard Error to correct the heteroskedasticity of the two models

Base group for health level is: Average
Base group for Race is: American Indian or Alaskan Native

Heteroskedasticity

```
White's Test p-value for Read: 0.16791051206480526
White's Test p-value for Math: 0.3490302204217291
Breusch-Pagan p-value for Read: 0.9346766142435015
Breusch-Pagan p-value for Math: 0.23683623813318913
```

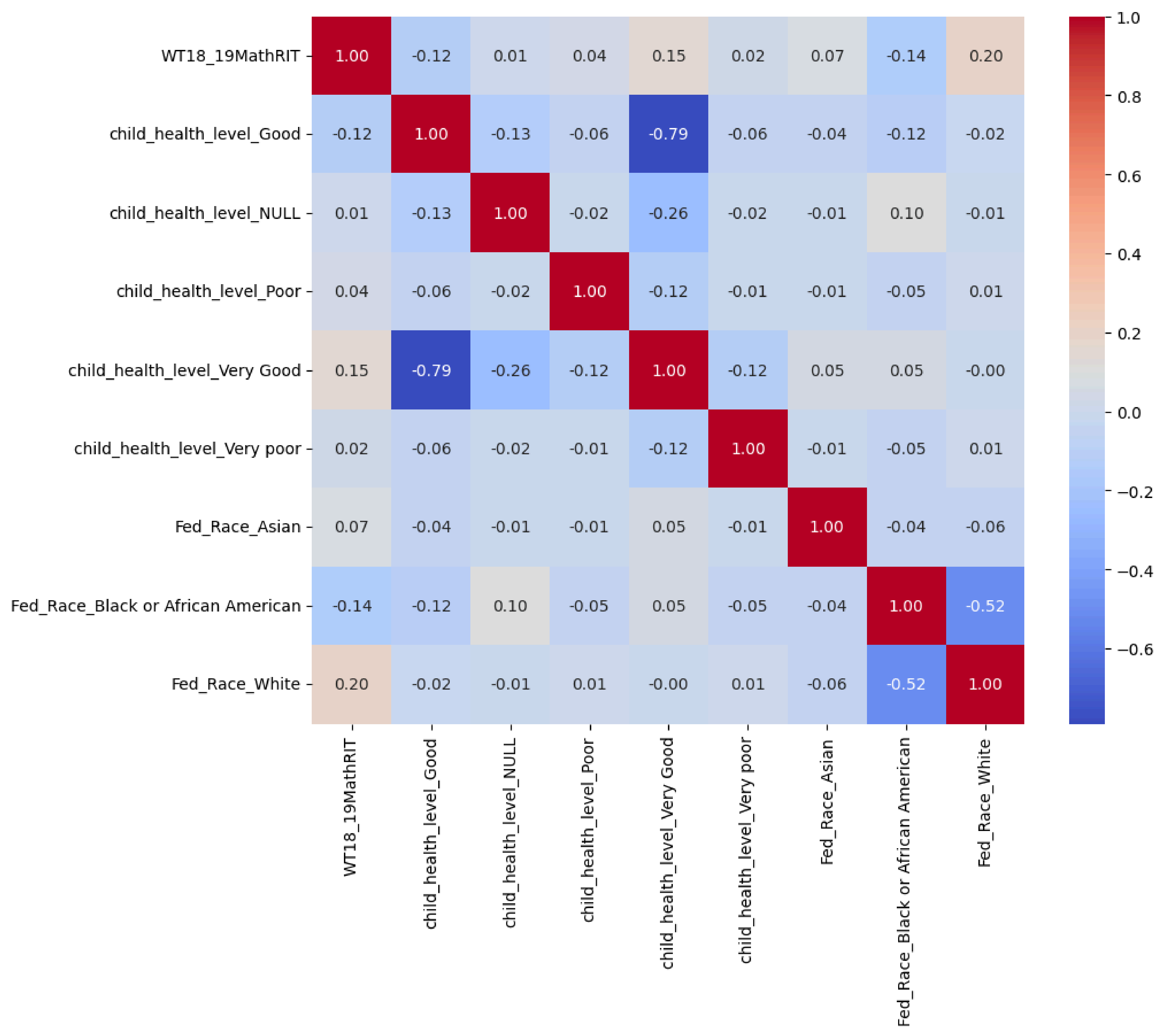




```
                            OLS Regression Results
==============================================================================
Dep. Variable:          WT18_19ReadRIT   R-squared:                       0.152
Model:                             OLS   Adj. R-squared:                  0.120
Method:                  Least Squares   F-statistic:                     4.737
Date:                 Thu, 07 Nov 2024   Prob (F-statistic):           2.32e-05
Time:                         14:52:04   Log-Likelihood:                -1266.2
No. Observations:                  220   AIC:                             2550.
Df Residuals:                      211   BIC:                             2581.
Df Model:                            8
Covariance Type:             nonrobust
=========================================================================================================
                                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------
const                                  83.2070     27.387      3.038      0.003      29.220     137.194
child_health_level_Good                15.7061     27.026      0.581      0.562     -37.569      68.981
child_health_level_NULL                52.2136     35.905      1.454      0.147     -18.566     122.993
child_health_level_Poor                89.6142     60.838      1.473      0.142     -30.313     209.542
child_health_level_Very Good           43.4446     25.816      1.683      0.094      -7.447      94.336
child_health_level_Very poor          -15.3858     60.838     -0.253      0.801    -135.313     104.542
Fed_Race_Asian                       -126.6516     78.815     -1.607      0.110    -282.017      28.714
Fed_Race_Black or African American     48.9462     14.937      3.277      0.001      19.501      78.391
Fed_Race_White                         58.3575     12.555      4.648      0.000      33.609      83.106
==============================================================================
Omnibus:                        32.625   Durbin-Watson:                   1.814
Prob(Omnibus):                   0.000   Jarque-Bera (JB):               43.761
Skew:                           -1.091   Prob(JB):                     3.14e-10
Kurtosis:                        3.097   Cond. No.                         20.0
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          WT18_19MathRIT   R-squared:                       0.084
Model:                             OLS   Adj. R-squared:                  0.049
Method:                  Least Squares   F-statistic:                     2.404
Date:                 Thu, 07 Nov 2024   Prob (F-statistic):             0.0167
Time:                         14:53:07   Log-Likelihood:                -1216.9
No. Observations:                  220   AIC:                             2452.
Df Residuals:                      211   BIC:                             2482.
Df Model:                            8
Covariance Type:             nonrobust
=============================================================================================================
                                           coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
const                                  133.7487     21.890      6.110      0.000      90.597     176.900
child_health_level_Good                 25.1940     21.601      1.166      0.245     -17.388      67.776
child_health_level_NULL                 42.4135     28.699      1.478      0.141     -14.160      98.987
child_health_level_Poor                 64.2580     48.627      1.321      0.188     -31.599     160.115
child_health_level_Very Good            44.4339     20.635      2.153      0.032       3.757      85.111
child_health_level_Very poor            48.7580     48.627      1.003      0.317     -47.099     144.615
Fed_Race_Asian                          64.8174     62.996      1.029      0.305     -59.364     188.999
Fed_Race_Black or African American      -6.6015     11.939     -0.553      0.581     -30.137      16.934
Fed_Race_White                          23.9866     10.035      2.390      0.018       4.205      43.768
==============================================================================
Omnibus:                      101.686   Durbin-Watson:                   1.577
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              282.365
Skew:                          -2.139   Prob(JB):                     4.84e-62
Kurtosis:                       6.535   Cond. No.                         20.0
==============================================================================
```

Multi-colinearity



Policy Recommendations

- **Early Intervention Programs**
    - Regular Health Checkups: Conduct regular health screenings to identify potential health problems and refer children to appropriate specialists.
    - Nutrition Education: Provide education and resources to families on healthy eating habits.
    - Physical Activity: Promote physical activity and outdoor play to improve physical health and cognitive development.
    - Mental Health Support: Offer mental health screenings and counseling services for young children.

- **Culturally Responsive Pedagogy**
    - Train teachers to use culturally responsive teaching methods that address the unique needs of students from diverse backgrounds.
    - Provide language support services for students who are using ESL.
    - Develop targeted interventions to support students who are struggling academically, such as tutoring and mentoring programs.

Conclusion

Our analysis of the provided dataset has revealed **significant disparities** in academic outcomes based on race and health levels. We have identified key factors that contribute to these disparities including **health levels and races**.

To address these issues, we propose **two key policy recommendations**:

1. **Early Childhood Intervention Programs**: By investing in early childhood education and providing comprehensive health services, we can help close the achievement gap and improve long-term outcomes for disadvantaged students.
2. **Culturally Responsive Teaching**: By implementing culturally responsive teaching practices, we can create more inclusive and equitable learning environments for all students.

Limitations of our analysis

- Missing data points, potential response bias in self-reporting data, and limitations on causal inference due to the observational nature of the data.
- The representation of each category is not uniform
- No control variables were included in the regression model
- The limited time prevents us from looking at the time-series data to provide a more in-depth analysis and recommendations for the problems
- There's a need to look at the combined effect of health and race

Future work

- Improve data collection method for uniform representation across different categories.
- Enhance the understanding of the struggle that each marginalized groups might have with the healthcare system
- Conduct qualitative interviews to have better and more comprehensive data

Thank you!