

Sarah Swee BTRY 4100 - Multivariate Analysis Final Paper

Dana Yang

FRPM (Free Reduced-Price Meal) Eligibility in California Schools (2023–2024): Data Cleaning, Exploratory Data Analysis, and Visualizations

Important Files:

- frmp_analysis.R
- Excel dataset: frmp23-24.xlsx
- Run correlation matrix
- Run Shiny App 1, 2, 3

Introduction/Problem Statement:

The data analysis aims to analyze the Free and Reduced-Price Meal (FRPM) eligibility data from California schools to explore the patterns of socioeconomic disadvantage across counties, districts, and school types. By understanding the likelihood of qualifications for the FRPM program, we can understand what regions need more financial assistance, the socioeconomic makeup of students within the area, and where policymakers could further advocate to raise equity. Considering that California became the first state to provide free meals regardless of eligibility to expand on the existing National School Lunch Program (NSLP) in 2022-2023, this data collected is still important for the Department of Education to utilize and understand federal program qualifications, such as Title I Funding, Local Control Funding Formula, and other programs based on socioeconomic makeup.

Motivation for Problem Statement

As a Californian who came from a middle class family where I went to a high school with a wide spectrum of people in the middle class, I was curious to understand which students would buy lunch as opposed to bringing lunch to school. Sometimes, that is an indicator of family frugality, inability to make ends meet financially, or personal choice. However, as I went to school during the COVID years and the NSLP program was established when I returned to school, it was harder to analyze what students qualified.

Research Questions Aimed to Answer with Given Dataset

- What are the distributions of enrolled students that qualify for Free vs. FRPM based on different filters (county, school, region, education type, etc.)?
- How do the distributions differ, what could be said on a region to region basis?

Data Source: How it was collected

Data was collected by the California Department of Education regarding information about the FRPM (Student Poverty) Data. The data reflects the unduplicated counts and percentages of students eligible to receive Free or Reduced-Price Meals (FRPM) under the National School Lunch Program). The file includes school-level FRPM eligible data for K-12 students.

Data Columns:

(Description and Field Provided by cde.ca.gov/ds/ad/filesfpfrpm.asp)

Academic Year	The academic year corresponding to the annual Census Day from which official enrollment counts are determined using data submitted as part of the annual Fall 1 submission to the California Longitudinal Pupil Achievement Data System (CALPADS) by local educational agencies (LEAs), which include school districts, charter schools, and county offices of education. Census Day or Information Day is always the first Wednesday in October.
County Code	A unique two-digit code corresponding to the county.
District Code	A unique five-digit code corresponding to the district.
School Code	A unique seven-digit code corresponding to the school
County Name	County name
District Name	District or Administrative Authority Name
School Name	School Name
District Type	District Ownership Type Description
School Type	School Ownership Type Description
Educational Option Type	Educational Option Type Description
Charter School (Y/N)	A "Y" or "N" value indicating whether a school is a charter school in the current academic year
Charter School Number	A four character value (any combination of numbers and letters) assigned to a charter school. A blank value indicates the school is not a charter school.
Charter School Funding Type	The values in this field indicate the charter school funding type. Values are as follows: Locally funded, Directly funded, Not in Charter School funding model, or Blank
IRC	A "Y" or "N" value indicating whether a school is an independently reporting charter (IRC) school in the current academic year.

Low Grade	Lowest Grade offered
High Grade	Highest Grade offered
Enrollment (TK/K-12)	A total count of TK/K-12 students enrolled (primary or short-term) on Census Day (the first Wednesday in October). These data were submitted to CALPADS as part of the annual Fall 1 submission.
Free Meal Count (TK/K-12)	Of the Enrollment (TK/K-12), a total unduplicated count of students who meet household income or categorical eligibility criteria for free meals based on one or more of the following reasons: (1) applying for the National School Lunch Program (NSLP); (2) submitting alternative household income forms; (3) student homeless or migrant statuses in CALPADS; (4) being "directly certified" through CALPADS as participating in California's food stamp or CalWORKs programs; or (5) being identified through the weekly CALPADS Foster Matching process or matched by the LEA through the CALPADS online match process as being in Foster Placement, Foster Family Maintenance, or Tribal Foster (beginning 2020) on Census day. The Free Meal Count (TK/K-12) is not displayed on any CALPADS report; however, this count represents the official Free Meal Count (TK/K-12) for the academic year.
Percent (%) Eligible Free (TK/K-12)	The percent of students eligible for free meals. [Free Meal Count (TK/K-12) divided by Enrollment (K-12)].
FRPM Count (K-12)	Of the Enrollment (K-12), a total unduplicated count of students who meet household income or categorical eligibility criteria for free or reduced meals (FRPM) based on one or more of the following reasons: (1) applying for the National School Lunch Program (NSLP); (2) submitting alternative household income forms; (3) student homeless, migrant, or tribal foster youth statuses in CALPADS; (4) being "directly certified" through CALPADS as participating in California's food stamp or CalWORKs programs or (5) being identified through the weekly CALPADS Foster Matching process or matched by the LEA through the CALPADS online match process as being in Foster Placement, Foster Family Maintenance, or

	<p>Tribal Foster (beginning 2020) on Census day. The FRPM Count (TK/K-12) generally corresponds to the "Total Unduplicated Eligible Free/Reduced Meal Counts" provided in CALPADS certification report 1.17: FRPM/English Learner/Foster Youth – Count. This count represents the official FRPM Count (TK/K-12) for the academic year.</p>
Percent (%) Eligible FRPM (TK/K-12)	<p>The percent of students eligible for free or reduced-price meals (FRPM). [FRPM Count (TK/K-12) divided by Enrollment (TK/K-12)].</p>
Enrollment (Ages 5-17)	<p>A total count of students ages 5-17 enrolled (primary or short-term) on Census Day (the first Wednesday in October). These data were submitted to CALPADS as part of the annual Fall 1 submission.</p>
Free Meal Count (Ages 5-17)	<p>Of the Enrollment (Ages 5-17), a total unduplicated count of students who meet household income or categorical eligibility criteria for free meals based on one or more of the following reasons: (1) applying for the National School Lunch Program (NSLP); (2) submitting alternative household income forms; (3) student homeless, migrant, or tribal foster youth statuses in CALPADS; (4) being "directly certified" through CALPADS as participating in California's food stamp or CalWORKs programs or (5) being identified through the weekly CALPADS Foster Matching process or matched by the LEA through the CALPADS online match process as being in Foster Placement or Foster Family Maintenance, or Tribal Foster (beginning 2020) on Census day. The Free Meal Count (Ages 5-17) is not displayed on any CALPADS report; however, this count represents the official Free Meal Count (Ages 5-17) for the academic year.</p>
Percent (%) Eligible Free (Ages 5-17)	<p>The percent of students eligible for free meals. [Free Meal Count (Ages 5-17) divided by Enrollment (Ages 5-17)].</p>
FRPM Count (Ages 5-17)	<p>Of the Enrollment (Ages 5-17), a total unduplicated count of students who meet household income or categorical eligibility criteria for free or reduced-price meals (FRPM) based on one or more of the following reasons: (1) applying for the National School Lunch Program (NSLP);</p>

	(2) submitting alternative household income forms; (3) student homeless, migrant, or tribal foster youth statuses in CALPADS; (4) being "directly certified" through CALPADS as participating in California's food stamp or CalWORKs programs or (5) being identified through the weekly CALPADS Foster Matching process or matched by the LEA through the CALPADS online match process as being in Foster Placement, Foster Family Maintenance, Tribal Foster (beginning 2020) on Census day. The FRPM Count (Ages 5-17) generally corresponds to the "Total Unduplicated Eligible Free/Reduced Meal Counts" provided in CALPADS certification report 1.17: FRPM/English Learner/Foster Youth – Count with the 5-17 Year Olds filter applied. This count represents the official FRPM Count (Ages 5-17) for the academic year.
Percent (%) Eligible FRPM (Ages 5-17)	The percent of students eligible for free or reduced-price meals (FRPM). [FRPM Count (Ages 5-17) divided by Enrollment (Ages 5-17)].
CALPADS Fall 1 Certification Status	A Y or N value to indicate whether data were submitted and certified in CALPADS. No data are provided where the certification status is "N".

Summary of Data Cleaning:

- What I noticed in the dataset that needed cleaning
- Visible trends looking at dataset face value
- Generalization and Summary within Data (names of columns, meaning of columns in the data, how many data points,
- Used second row as header from Excel
- Cleaned column names to have underscores as spaces
- Convert columns into numeric values
- Converted Charter School (Y/N) to binary values
- Remove unwanted rows, where School Name == "District Office" and District Type == "State Special Schools" because they were not relevant and consists of 0 values that skewed towards zero
- Dropped Academic Year, County Code, School Code, District Code, CALPADS Fall 1 Certification Status as they were not relevant
- Filtered out rows with 0 or NA in Enrollment (K-12) and Enrollment (Ages 5-17) as they weren't insightful data points

Data Represented After Cleaning:

Districts: 1003, Counties: 58, Schools: 8742, School Types: 16

Linear Model Analysis

Q1: Do charter schools have different rates of FRPM eligibility?

Model: Percent (%) Eligible FRPM (K-12) ~ Charter School (Y/N)

p-value: 0.2051

Interpretation: The result is not statistically significant, suggesting that charter school status alone does not explain differences in FRPM eligibility. Charter status is likely insufficient as a standalone predictor.

Q2: Are school types associated with differences in FRPM eligibility?

Model: Percent (%) Eligible FRPM (K-12) ~ School Type

p-value: 2.2e-16

Interpretation: This is a highly statistically significant result. Different school types are strongly associated with FRPM eligibility, indicating that structural and institutional characteristics may influence economic disadvantage levels across schools.

Q3: Do charter schools tend to be smaller or larger than non-charter schools?

Model: Enrollment (Ages 5–17) ~ Charter School (Y/N)

p-value: 0.003087

Interpretation: This model shows a significant relationship between school size and charter status. Charter schools tend to have lower enrollment on average, suggesting that they may serve smaller student populations.

Q4: Are some counties more disadvantaged than others?

Model: Percent (%) Eligible FRPM (K-12) ~ County Name

p-value: 2.2e-16

Interpretation: Strong evidence suggests that FRPM eligibility varies significantly across counties, underscoring geographic disparities in economic need. County-level context plays an important role in determining school-level eligibility rates.

Q4: What best explains percent FRPM eligibility? (Multivariable model)

Model: Percent (%) Eligible FRPM (K-12) ~ Charter School (Y/N) + School Type + Educational Option Type

p-value: 2.2e-16

Interpretation: When combined, these variables significantly explain variation in FRPM eligibility. This confirms that a multifactor approach—rather than a single predictor—better captures the complexity of socioeconomic disadvantage in schools.

Comparison of Enrollment Predictors for FRPM Eligibility:

Model 6: Percent (%) Eligible FRPM (K-12) ~ Enrollment (K-12) → p-value: 7.90e-13

Model 7: Percent (%) Eligible FRPM (K-12) ~ Enrollment (Ages 5–17) → p-value: 5.67e-12

Interpretation: Both enrollment measures are significantly associated with FRPM eligibility. While each model offers a slightly different scale, they both support the idea that larger enrollments are meaningfully related to FRPM patterns.

Multivariate Analysis:

Q1: Do county and enrollment explain percent eligible for free meals (K–12)?

Model: Percent (%) Eligible Free (K-12) ~ County Name + Enrollment (K-12)

p-value: 2.2e-16

Interpretation: This model finds that both county and enrollment are highly significant predictors of free meal eligibility among K–12 students. This suggests that both regional context and school size impact access to basic nutritional resources.

Q2: Do county and enrollment explain percent eligible for FRPM (K–12)?

Model: Percent (%) Eligible FRPM (K-12) ~ County Name + Enrollment (K-12)

p-value: 2.2e-16

Interpretation: The model demonstrates a strong association between both predictors and FRPM eligibility rates. It reinforces the idea that larger enrollment numbers and county-specific socioeconomic conditions significantly influence school-level economic needs.

Q3: Do county and enrollment explain percent eligible for free meals (Ages 5–17)?

Model: Percent (%) Eligible Free (Ages 5–17) ~ County Name + Enrollment (K-12)

p-value: 2.2e-16

Interpretation: County and enrollment remain statistically significant factors in predicting free meal eligibility for students aged 5–17. This affirms the consistency of these predictors across age bands.

Q4: Do county and enrollment explain percent eligible for FRPM (Ages 5–17)?

Model: Percent (%) Eligible FRPM (Ages 5-17) ~ County Name + Enrollment (K-12)

p-value: 2.2e-16

Interpretation: A highly significant result once again indicates that geographic location and enrollment size are reliable predictors of FRPM eligibility, even when applied to a specific age range.

Ultimately, the p-values tell us that these predictors are statistically significant due to both large sample size and structural differences. County and enrollment are useful indicators to predict free meal and FRPM eligibility for all categories under FRPM vs free meals, and Ages 5-17 vs K-12.

Interpretation of Correlation Matrix/Relevant Linear Models

Relevant Metrics represented: enrollment_k_12, free_meal_count_k_12, percent_eligible_free_k_12, frpm_count_k_12, percent_eligible_frpm_k_12, enrollment_ages_5_17, free_meal_count_ages_5_17, percent_eligible_free_ages_5_17, frpm_count_ages_5_17, percent_eligible_frpm_ages_5_17

These metrics were chosen as these were numerical representative values that could be compared best with linear models. All p-values were found to be <0.001 because our sample size is large (over 10,000 data points).

Correlations that stood out:

Enrollment (K-12) vs Percent Eligible Free (K-12)

Correlation coefficient: -0.109

Interpretation: Weak negative correlation, potentially suggesting that larger schools have a lower % of students eligible for free meals

Free Meal Count (K-12) vs Percent Eligible FRPM (K-12)

Correlation coefficient: 0.348

Interpretation: Although moderate relationship possibly due to the differing of units (one in count and one in percent), it can suggest variability in school size.

Percent Eligible Free (K-12) vs Enrollment (Ages 5-17)

Correlation coefficient: -0.115

Interpretation: Weak negative correlation, however I think this might be a great linear model as percent eligible free does not predict enrollment.

Free Meal Count Ages 5-17 vs Percent Eligible FRPM (Ages 5-17)

Correlation coefficient: 0.335

Interpretation: Slight positive correlation as we are comparing count with percent.

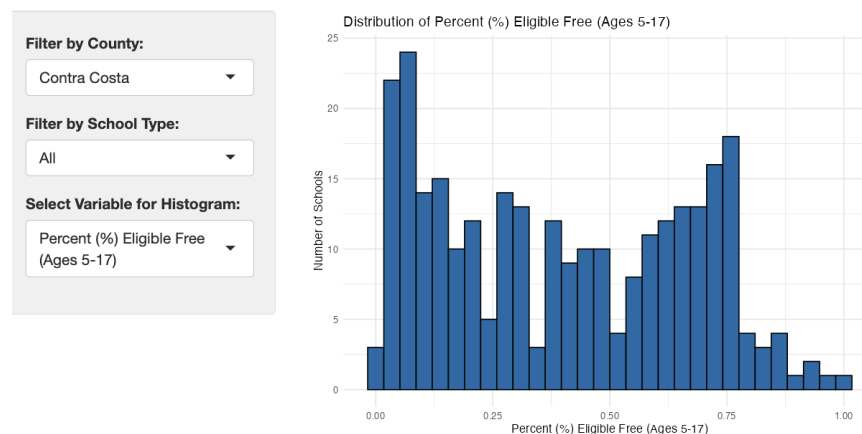
Shiny App #1: Interactive Histogram displaying distribution of Select Variables

Purpose: To explore the distribution of key educational and economic variables (e.g., enrollment, meal counts, eligibility percentages) across schools, filtered by county and school type.

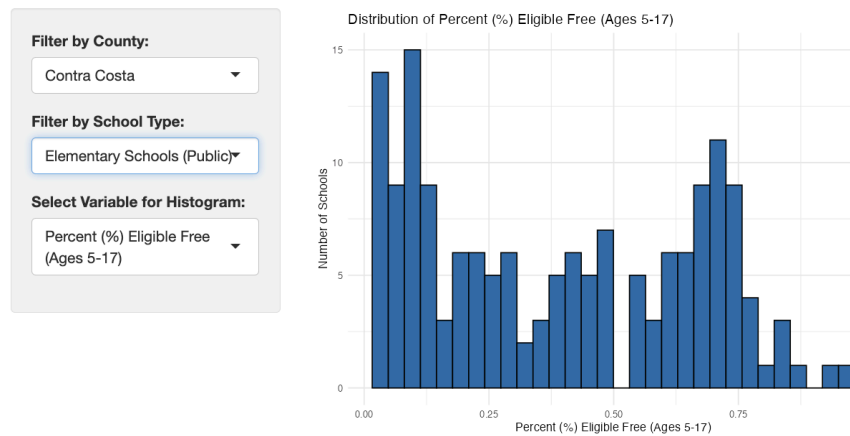
Output: A histogram showing the distribution of the selected metric based on the filtered dataset.

Example Data Visualizations:

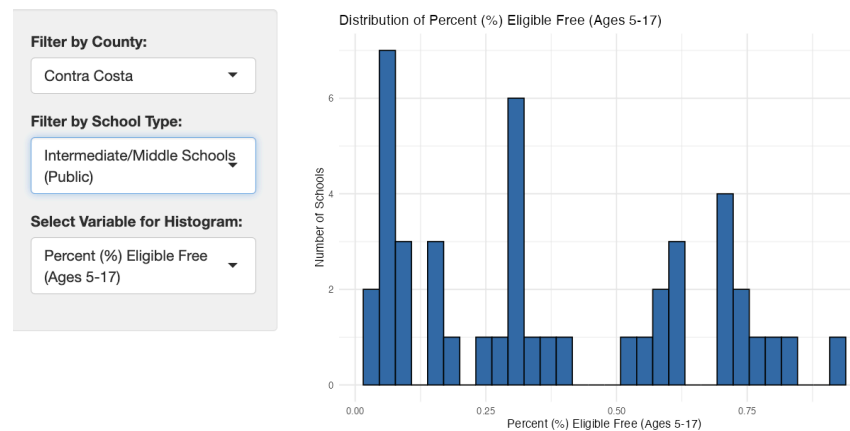
I was interested in exploring my county, Contra Costa, filtering school types by ALL, Elementary Schools (Public), Intermediate/Middle School (Public), and High School (Public), with selected variable as Percent (%) Eligible Free (Ages 5-17). The histogram did not generally shift significantly when other variables relating to Percent Eligibility for Free VS FRPM and Ages 5-17 vs K-12.



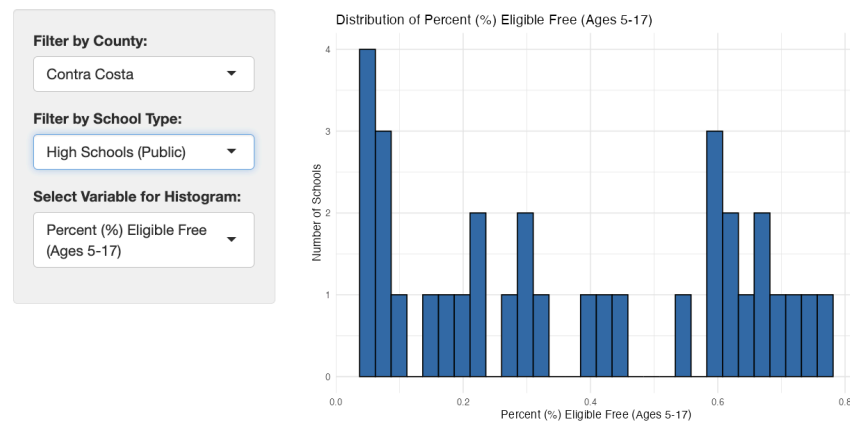
School Type ALL - Distribution is bimodal, where the number of schools with low eligibility rates (under 20%) and the other bump at around 60-80%. This may suggest some possible income disparities and varying needs.



School Type Elementary Schools (Public) - Bimodal, with groups between 0-10% and 60-80% eligibility. This could potentially imply early onset of economic disparities in different school zones.



School Type Intermediate/Middle Schools (Public) - Bimodal and more sparse due to smaller samples. Wide disparity potentially just because of size.



School Type High Schools (Public) - Multiple clusters, low eligibility and fewer in higher range. High schools tend to have lower overall free eligibility, potentially of students being less likely to be in households applying for free meals.

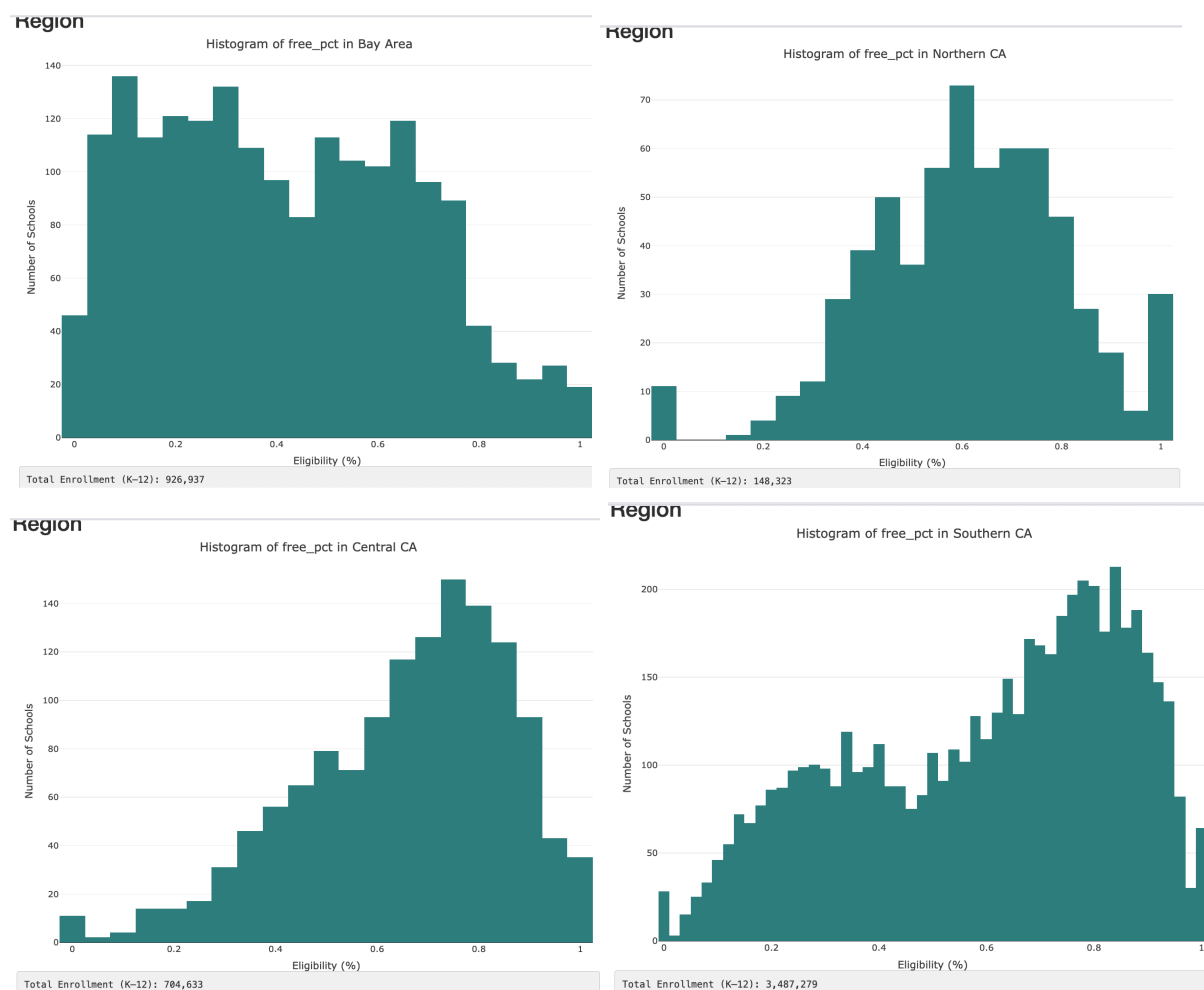
Further possibilities for data analysis:

- Understand how distributions look based on different county, school type, and selected variable for histogram criteria.

Shiny App #2: K-12 Eligibility Distribution by Region

Purpose: To understand the K-12 eligibility rates for Free Meals and FRPM across different regions of California, such as the Bay Area, Northern California, Central California, and Southern California. The goal is to understand the geographical disparities.

Output: A histogram of eligibility percentages across schools, filtered by region and selection metric between Free Eligibility or FRPM Eligibility.



It seems that Central and Southern California have greater left skewed than the Bay Area and Northern California. The distribution of Free eligibility is almost bimodal, which suggests there is polarization between the more affluent and economically disadvantaged communities within the same area. As Southern and Central California demonstrate a higher Free and FRPM eligibility, it suggests that

there is greater need for funding and efforts, as they fall under economic hardship, differences in economic structures, and socioeconomic status. Considering the Bay Area is highly affluent in the tech industry, Central California commonly described as heavily agricultural, and Southern California varying in socioeconomic inequality.



The size of enrollment also plays a role in shaping how these distributions appear. Southern California, with the highest K–12 enrollment, shows a dense and steep histogram, particularly for FRPM eligibility—underscoring both the scale of student need and the sheer number of schools experiencing high economic disadvantage. In contrast, Northern California exhibits a more varied distribution in eligibility levels, with schools spanning from low to high percentages. This could reflect a more heterogeneous mix of school settings and socioeconomic conditions in the region, including both economically struggling counties and those with relatively fewer low-income families. These nuances in regional distribution demonstrate that even within the same state, economic need is not experienced uniformly. The differences between Free and FRPM eligibility rates further illustrate how many students fall into that in-between category—those who may not meet the strict criteria for free meals but are still experiencing financial precarity. These disparities emphasize the importance of using FRPM as a broader socioeconomic indicator and raise questions about how support is allocated across counties and districts.

Further possibilities for data analysis:

- Further exploration could be filtered by different regions and counties.

Shiny App #3: Overlaid Histogram: % Free vs FRPM Eligibility

Purpose: Comparative overlapping histogram distributions between those eligible for Free meals vs those eligible for FRPM, allowing to filter by school types such as elementary, middle, or high school.

Output: Overlaying of 2 histograms to depict Free vs FRPM eligibility.



We see an obvious trend that more people fall in the category of FRPM than just Free in all of these groupings of schools (Elementary School (Public), Intermediate/Middle School (Public), Junior High School (Public), and High School (Public)).

Total enrollment count plays a role in shape and distributions of the skew. The elementary school distribution depicts a smooth and dense right-skewed distribution. Middle School and HS, with ~ 860,000 and ~ 1.6m, respectively, are smaller, but still maintain a clear pattern and greater variance in the middle ranges. Junior high, with ~ 28k seems irregular, considering that fewer schools fall under this category as it carries 7-8 grades.

These graphs all suggest the interpretation that there is a consistent trend of Free and FRPM eligibility across all types of students, calling off potential policy implications of greater funding towards areas with lower economic stability.

Further possibilities for data analysis:

Further exploration could be filtered by different school types such as Continuation High Schools, Alternative Schools of Choice, Special Education Schools (Public), etc.

Conclusion:

Ultimately, we found all of our variables useful indicators to predict FRPM for K-12 and Age 5-17 students, supported by the vastness of the dataset being very large (10k+).

For the restrictions of this final paper, it would have been interesting to further explore the different types of School Types and difference in Free/FRPM eligibility, compare Charter School eligibility statistics, and potentially have created a heatmap to visually depict the proportional saturation of which regions. These findings would be important for educators and policymakers to understand which schools need more allocation of resources beyond affording meals, potentially for school funding, attendance maintenance to qualify to federal funding, and general resource allocations to bridge the inequality gap.

Sources:

Dataset: [Unduplicated Student Poverty – Free or Reduced-Price Meals Data 2023–24](#)

Information on File: [Unduplicated Student Poverty – FRPM Data 2014 to Current Year](#)

File Structure/Variable Definitions: <https://www.cde.ca.gov/ds/ad/fsspfrpm.asp>