

2016 Advanced Placement Scores by Demographic

A Data Visualization Project

Christian H. Nunez

Physics 91SI

5 June 2018

INTRODUCTION

The following project was made possible by the public data set offered on kaggle.com:

<https://www.kaggle.com/collegeboard/ap-scores>

The dataset was not initially workable as it had many blank entries -- the initial challenge was to reformat the dataset so that it was easily manageable for several types of analysis.

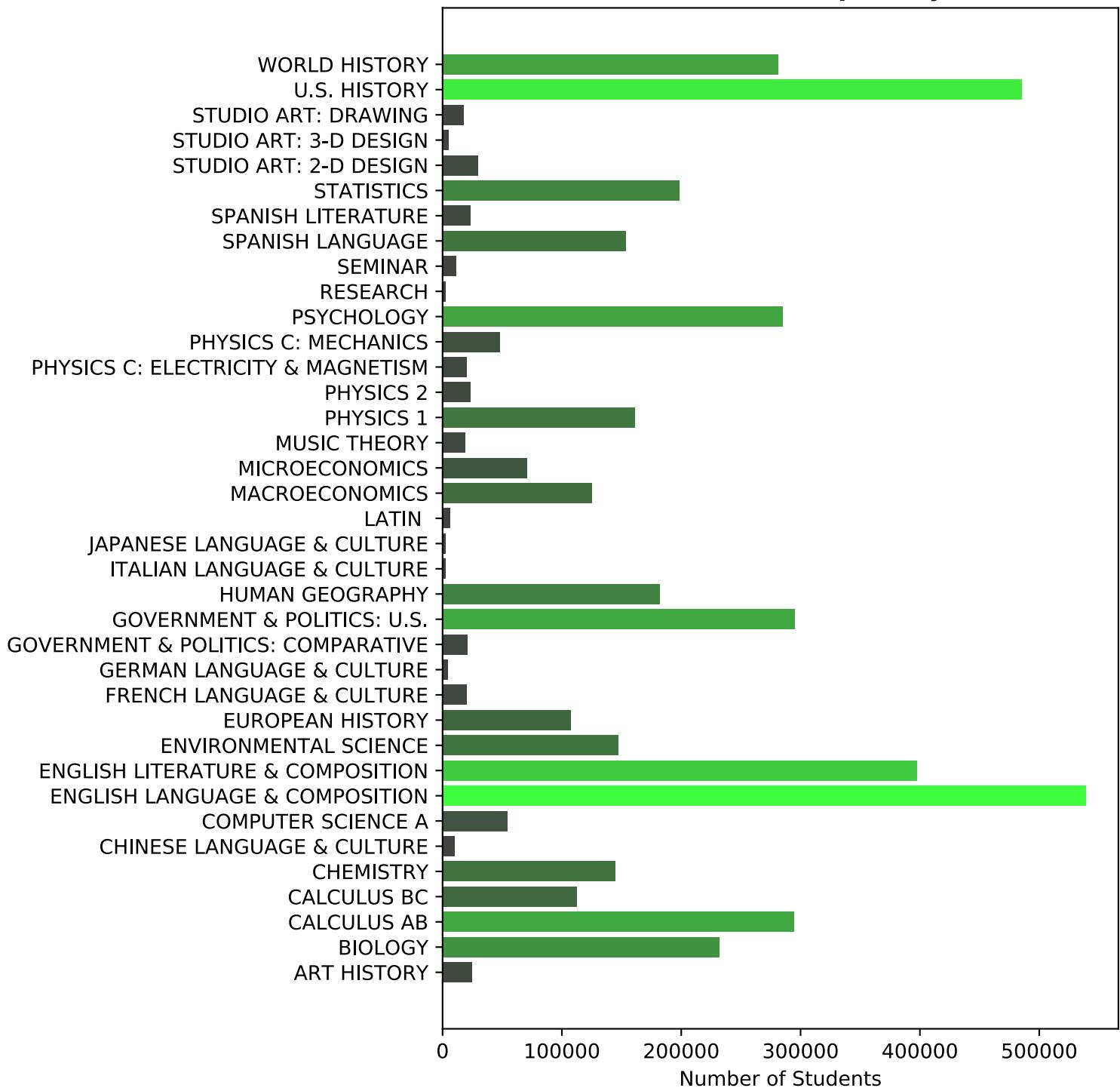
For the 91SI instructors: All of the code for this project can be found in my iPython notebook in my final project repository.

Motivation: For my PWR essays, I studied extensively the roots of standardized testing, tracing it back to the early 20th century eugenics movement. I read Professor Wayne Au of University of Washington's works on what he calls the "zip code effect", which states that a student's success on standardized tests can be accurately predicted by their zip code (which is linked to information about demographics, family financial stability, and so forth). I was interested in exploring a dataset containing information about standardized test scores and student demographics, and thus, this final project was born.

Part I: AP Test Popularity in 2016

The graphic on the following page depicts the popularity (i.e. the number of students taking the exam) of the 37 AP tests offered in 2016 in a horizontal bar graph where the brightness of each green bar is proportional to the bar length.

2016 AP Test Popularity



Part II: Stacked Bar Plots

The graphic on the next page shows the score distribution for each AP test where each bar is split by student demographic.

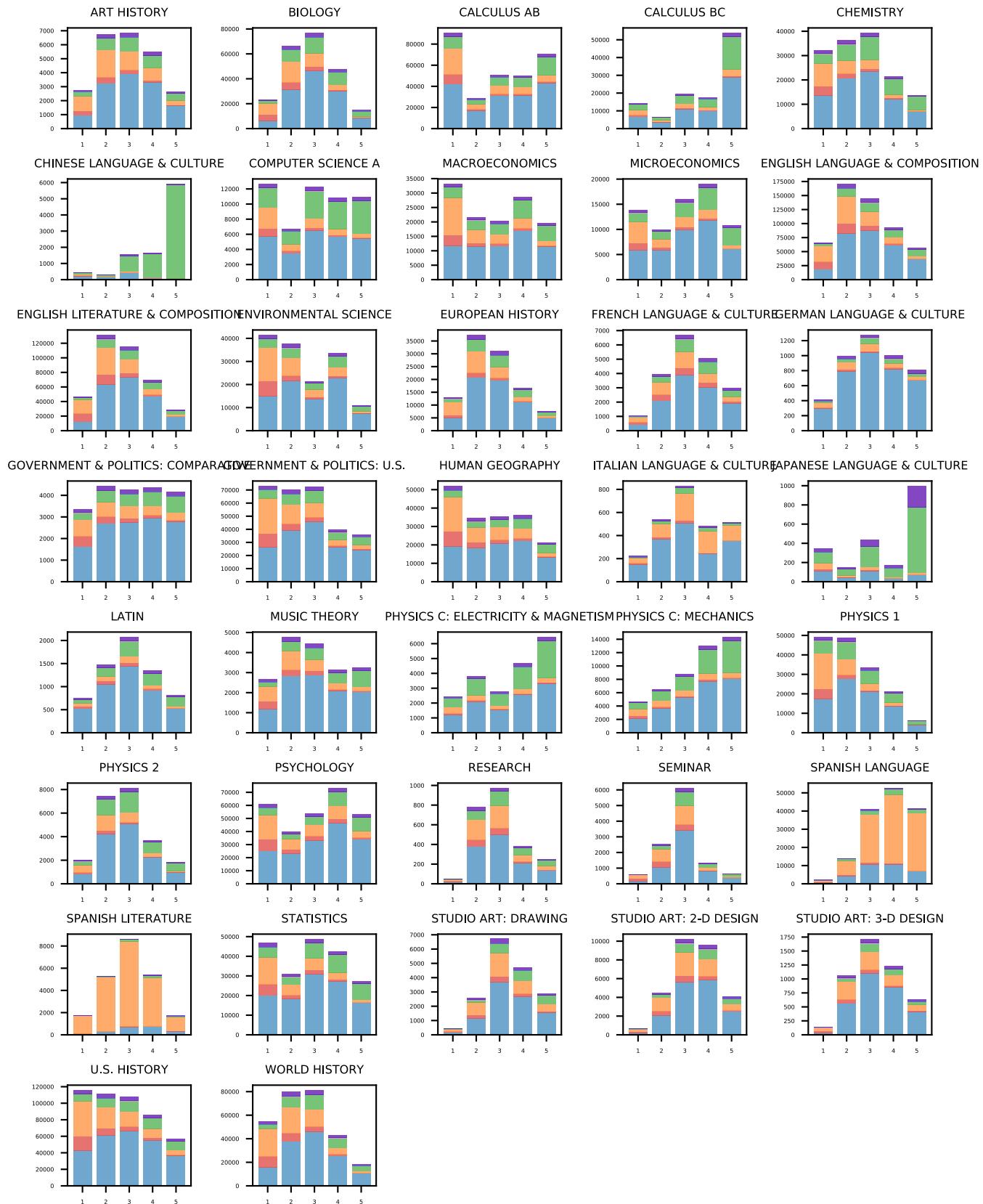
The following snippet of code represents some of the work that went in to forming each subplot of the following graphic. The stacked bar plot was created by tracking the “bottom” of each vertical bar and starting the bars from different demographics at that (continually updated) bottom position.

The legend colors match those in Part III.

```
# Function to create a stacked bar chart
def stacked_bar(xvalues, yvalues, legend_names, ax):
    bottomTracker = 0
    for i in range (0, len(yvalues)):
        ax.bar(xvalues, yvalues[i], label = legend_names[i], bottom = bottomTracker, color=colors_arr[i], alpha=0.8)
        bottomTracker += yvalues[i]
    return ax

def save_stacked(scores, this_class, classnames, i, ax=None):
    legend_names = np.array(["White", "Black", "Latinx", "Asian",
                           "American Indian/Alaska Native", "Native Hawaiian/Pacific Islander",
                           "2+ Races"])
    if(ax==None):
        fig, ax = plt.subplots(1,1)
    ax1 = stacked_bar(scores, this_class, legend_names, ax)
    ax.set_title(classnames[i], fontsize=6)
    ax.tick_params(labelsize=3)
```

2016 AP Test Scores by Student-Identified Race



Part III: Chi-square Test

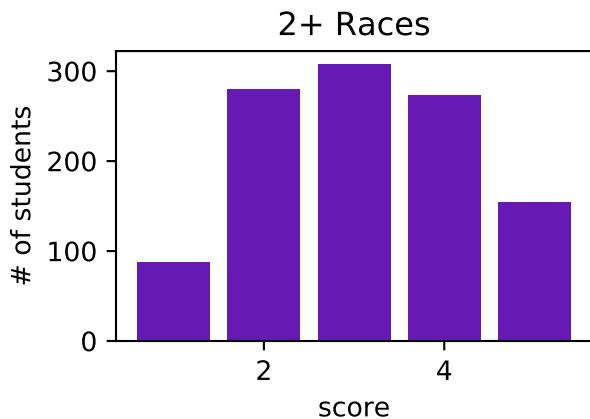
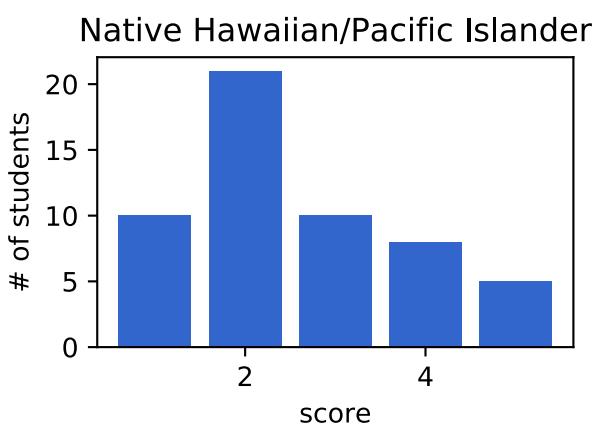
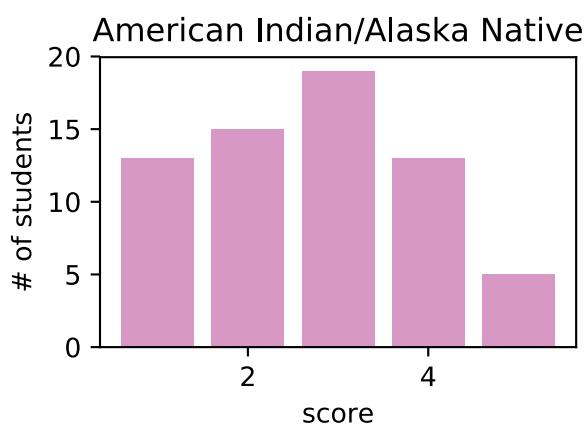
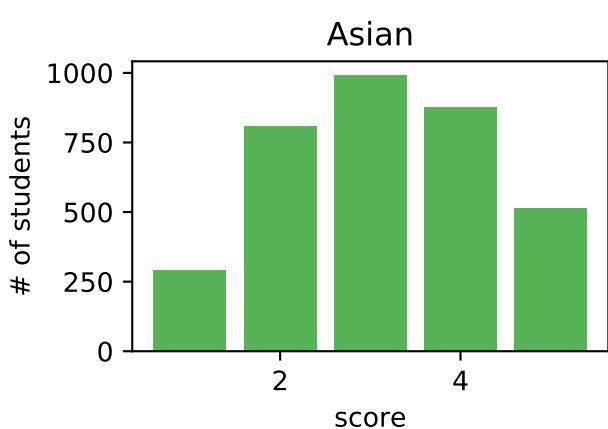
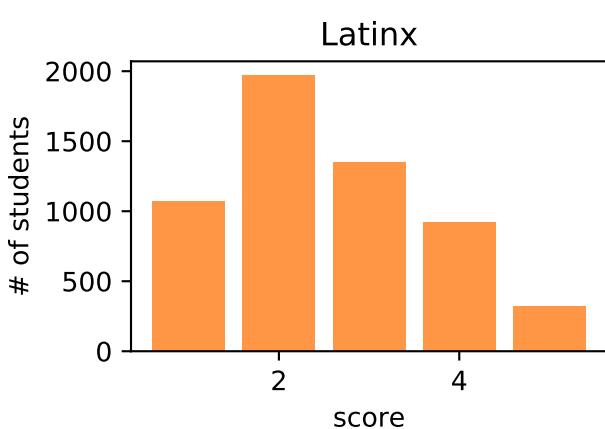
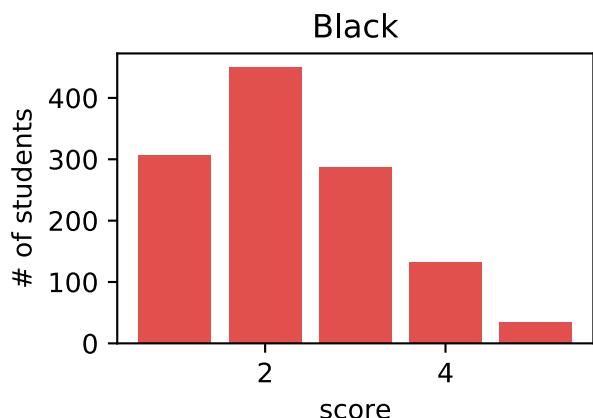
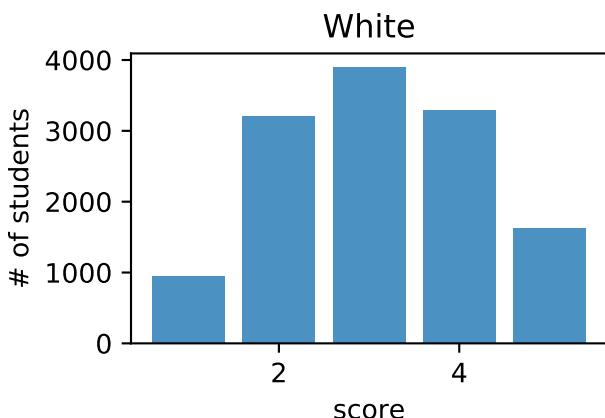
I performed a categorical chi-square test on 37 sub-datasets testing the null hypothesis: "All student demographics score in the same manner on [insert AP test name]". This test makes the assumption that all demographics will log each score with equal frequencies -- in other words, the model assumes that each demographic will perform in the exact same way (proportionally, of course). The chi-square statistic quantifies how much the actual data differs from the model and, when converted to a p-value using a look-up table, provides a p-value representing the probability that the distribution is in the form that it is in purely by chance.

The datasets for each class each have 24 degrees of freedom and the p-value of 0.001 has a chi-square value (assuming 24 degrees of freedom) of 51.179. As shown in the plots in the following tables, the chi-square values differ wildly from this statistic.

For each class, I can reject the null hypothesis, but it is important to be cautious in drawing conclusions. Visually and numerically, there seem to be great differences in the performance of different demographics on the majority of the tests. However, this single statistical test cannot prove anything about the cause of this phenomenon and should not be used to draw concrete conclusions.

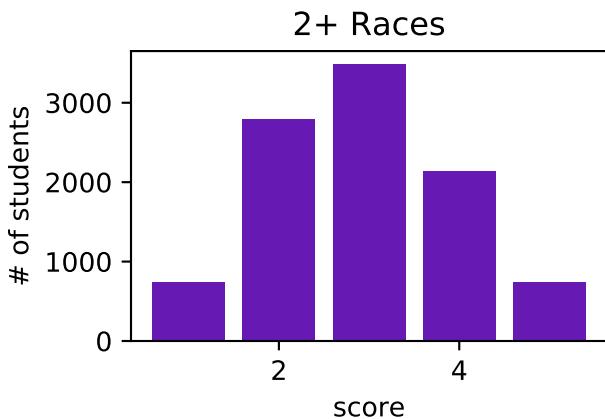
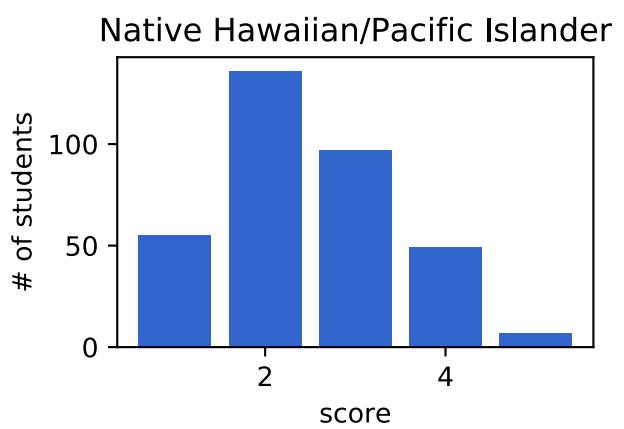
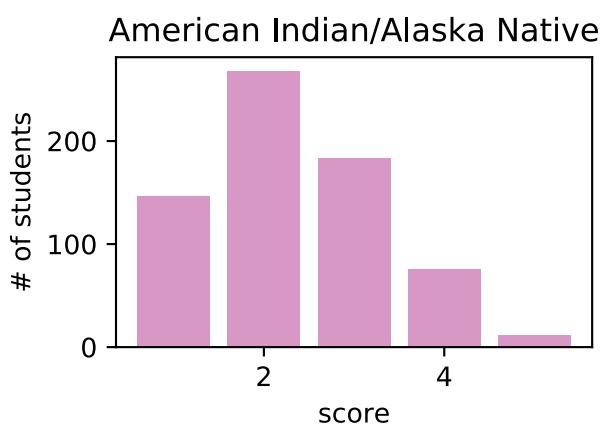
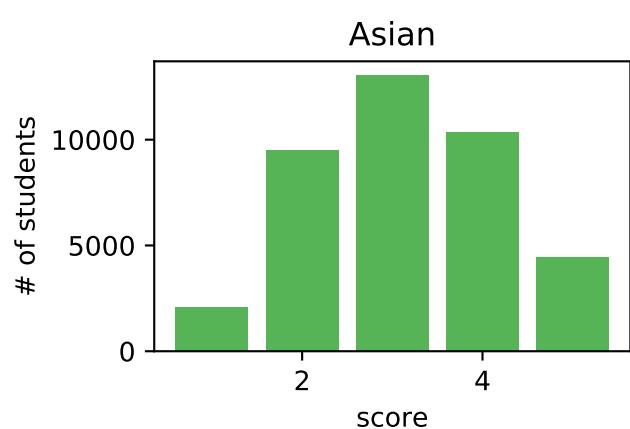
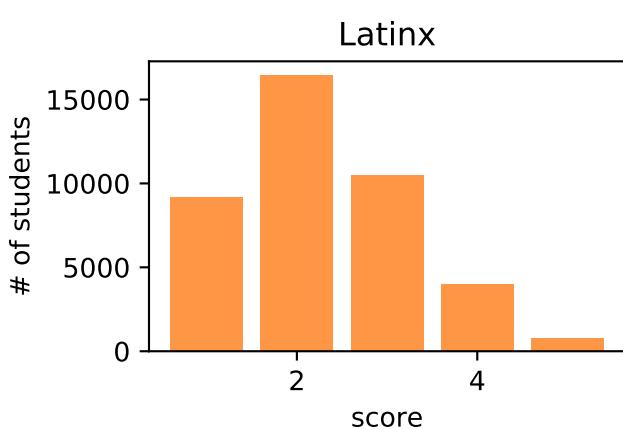
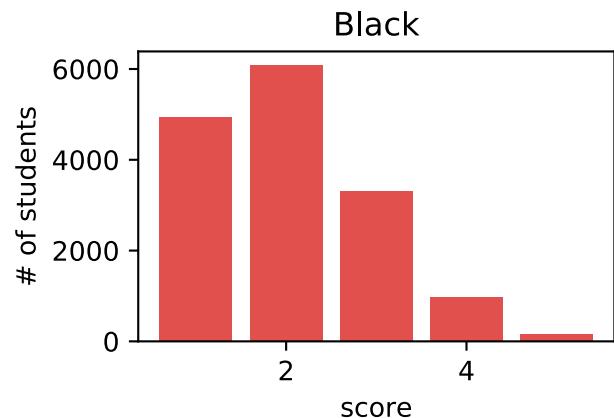
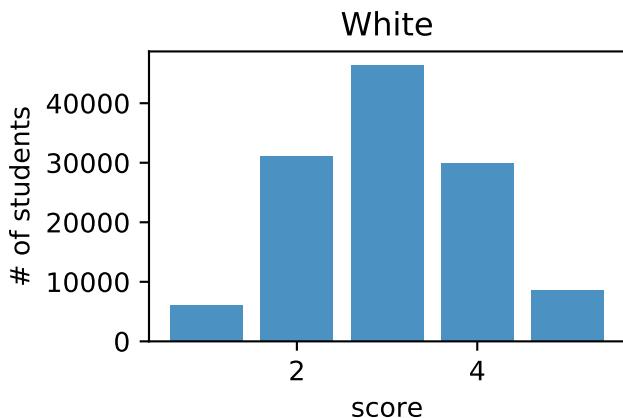
Working on this final project has inspired me to branch out and take more statistics, data science, and CS classes. I would like to learn more about data exploration and performing sound statistical tests.

ART HISTORY



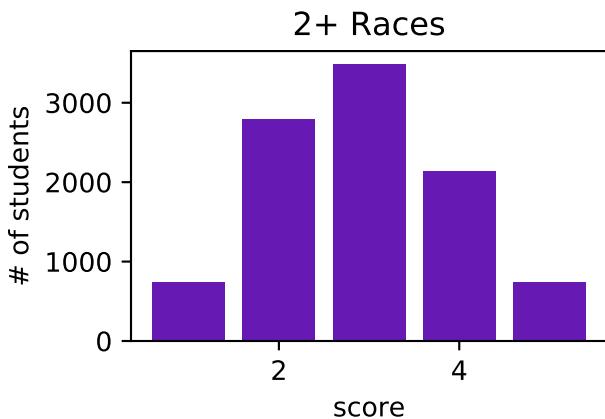
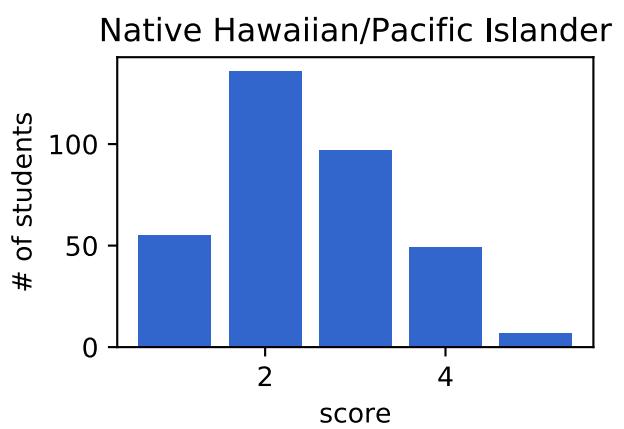
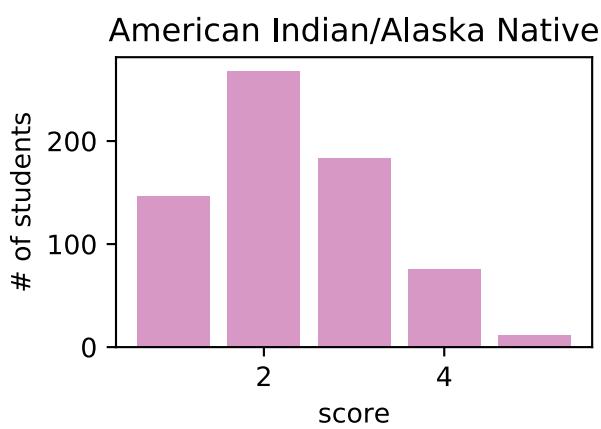
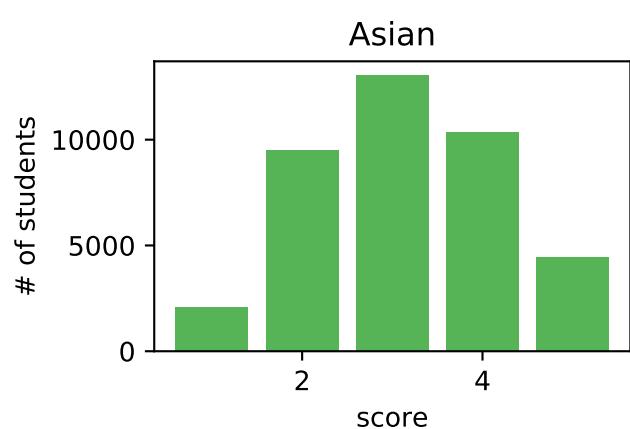
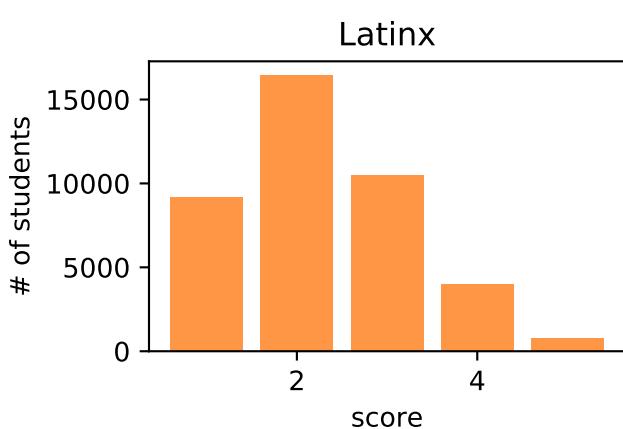
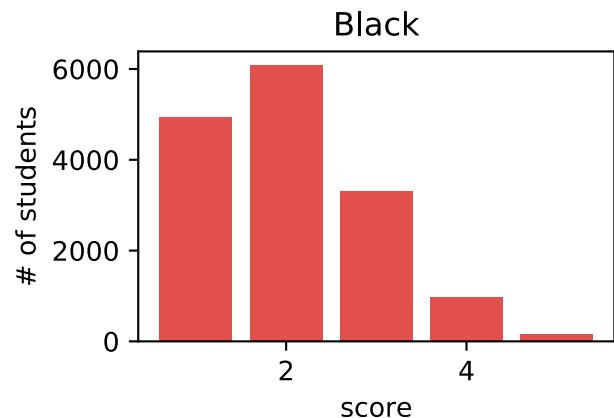
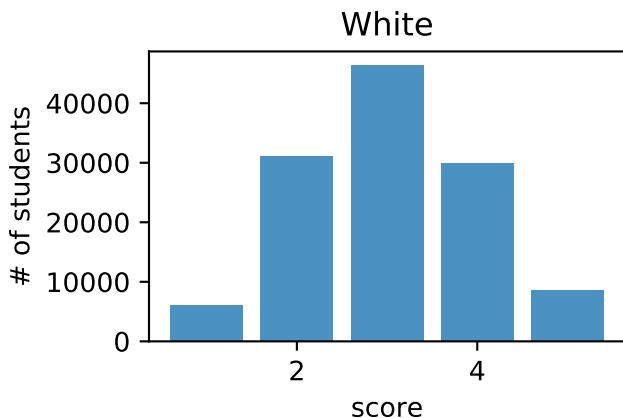
χ^2 statistic = 1572.6298714296893

BIOLOGY



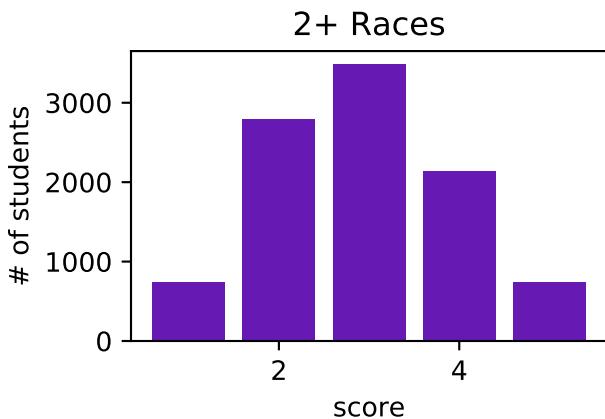
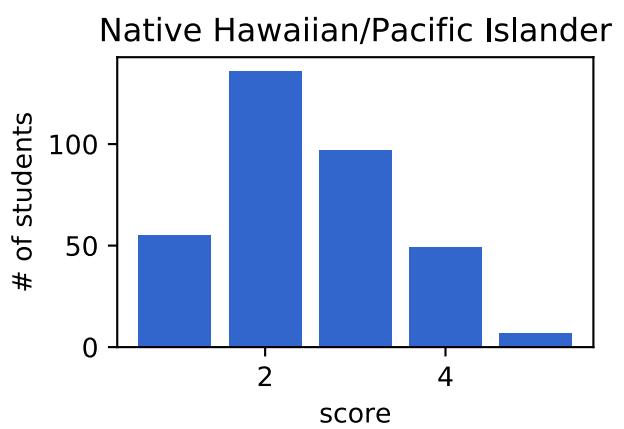
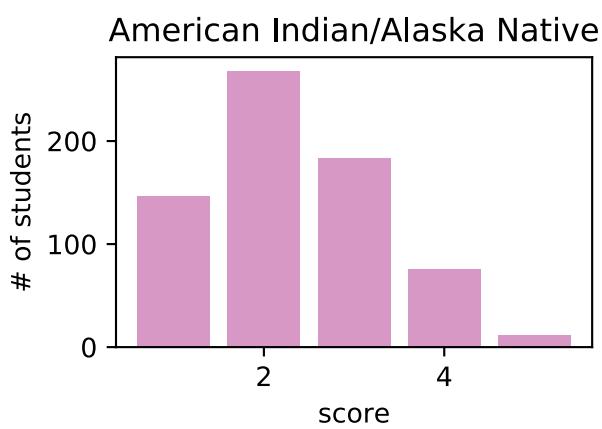
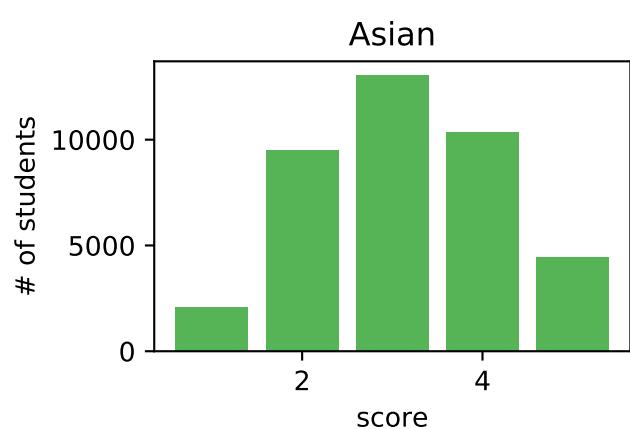
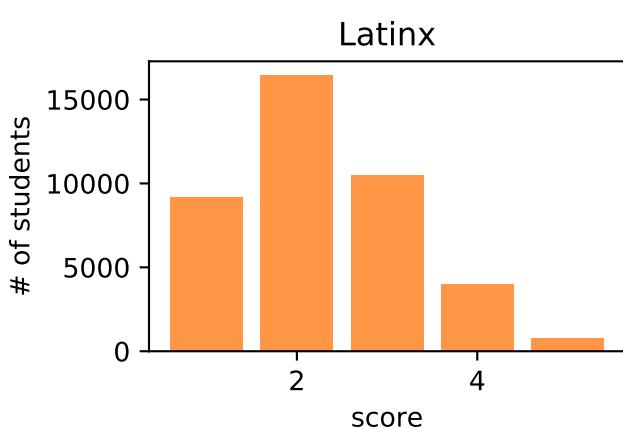
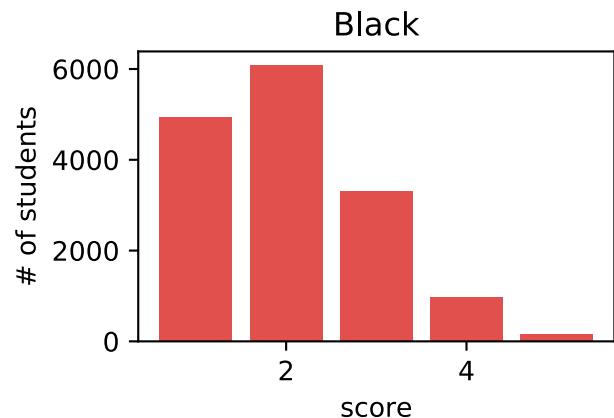
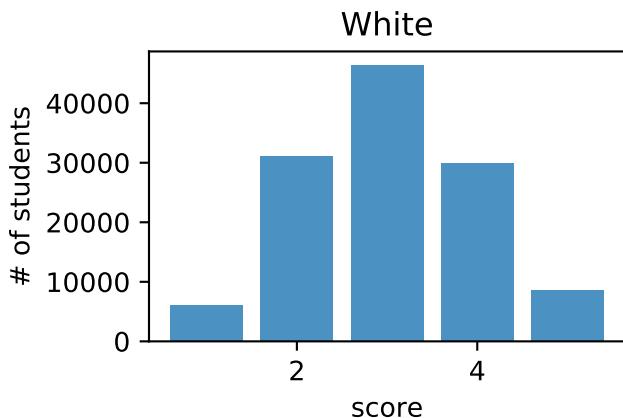
χ^2 statistic = 31852.914287287465

BIOLOGY



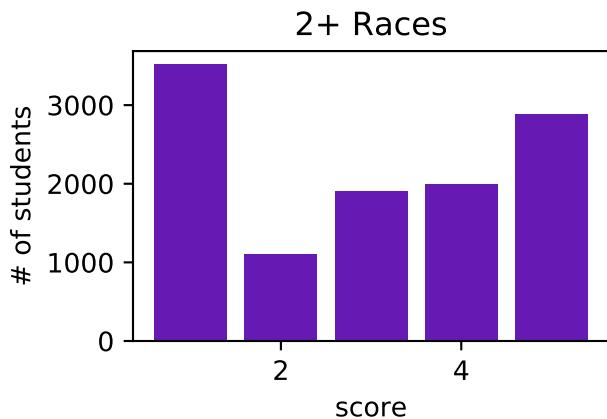
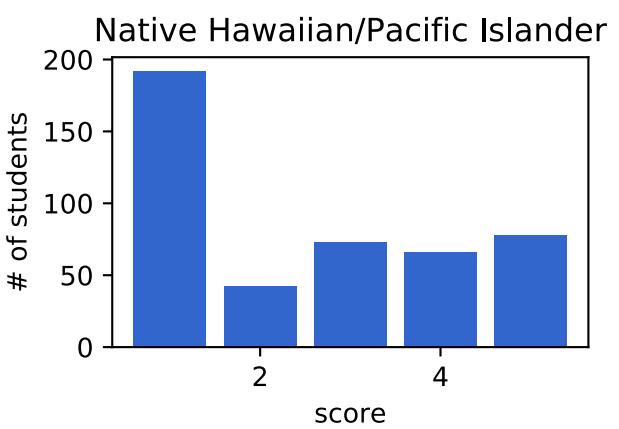
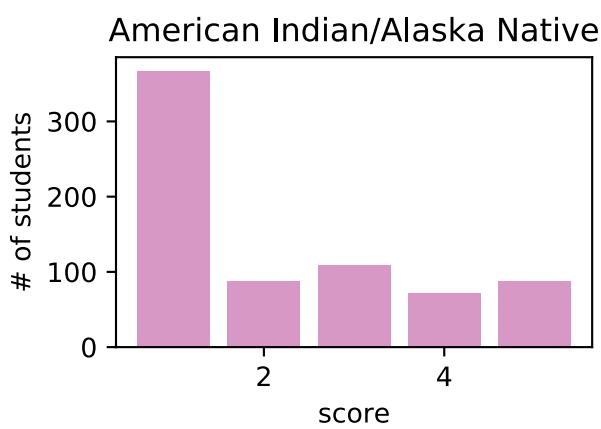
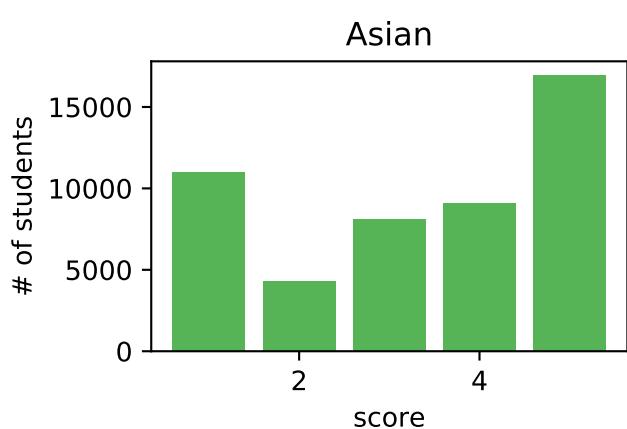
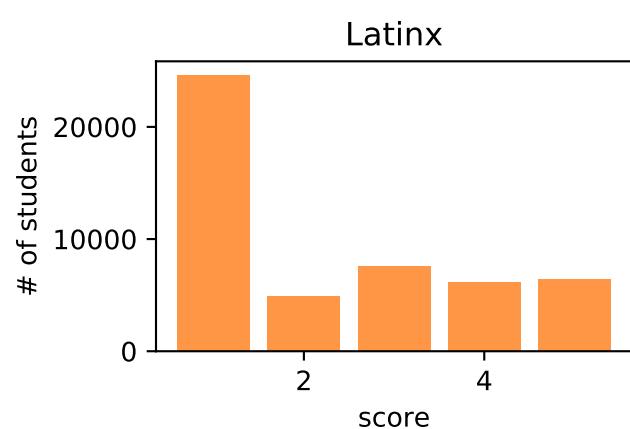
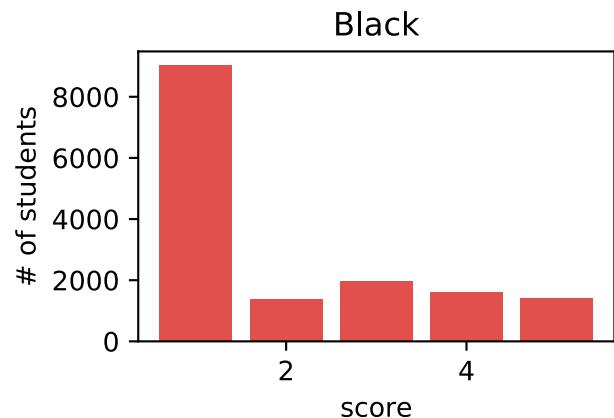
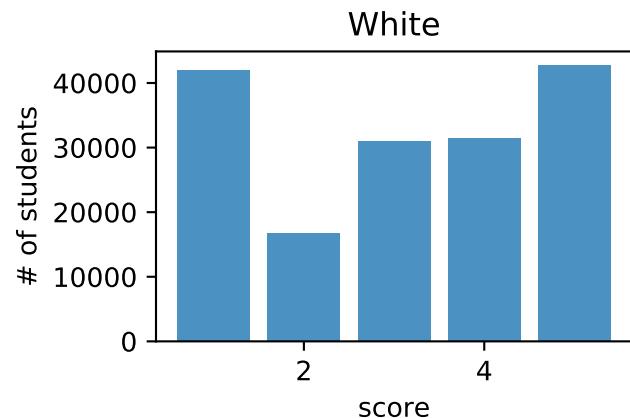
χ^2 statistic = 31852.914287287465

BIOLOGY



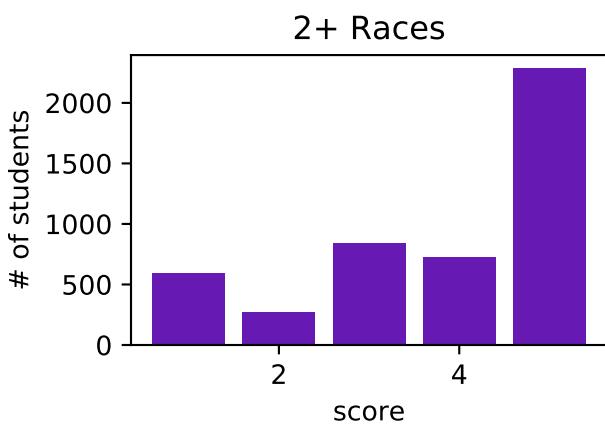
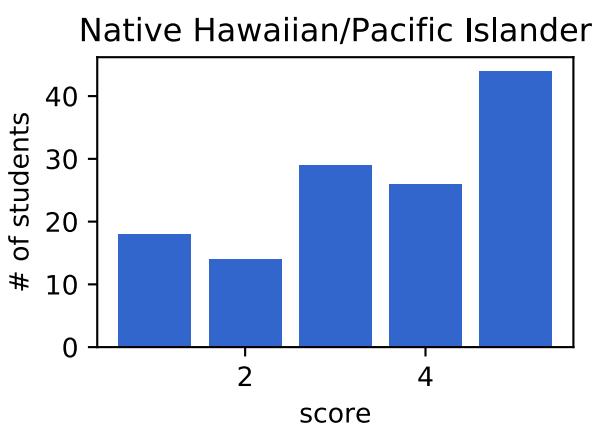
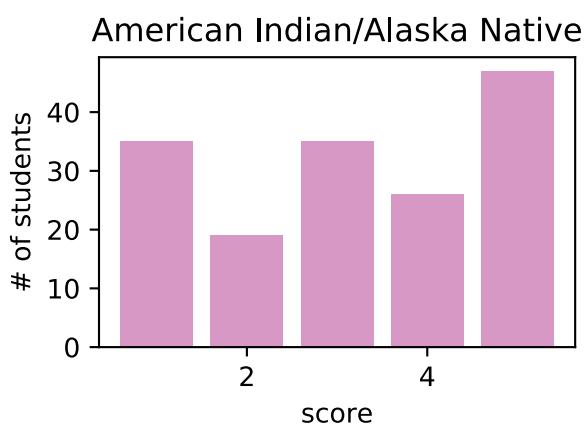
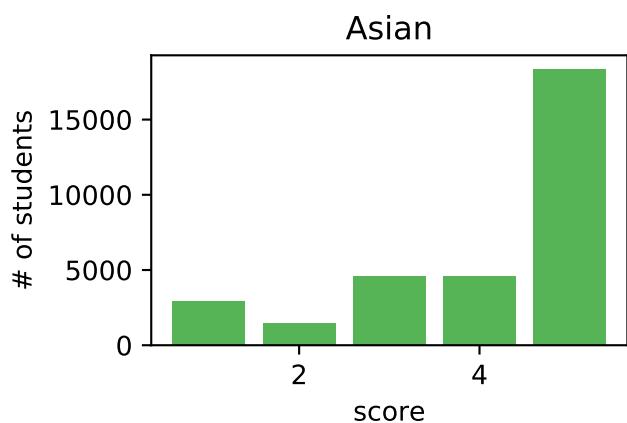
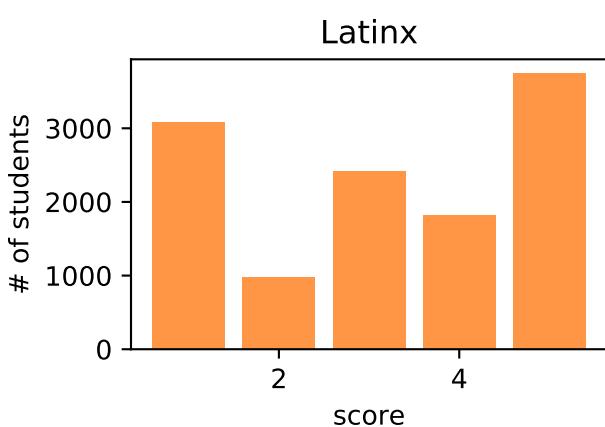
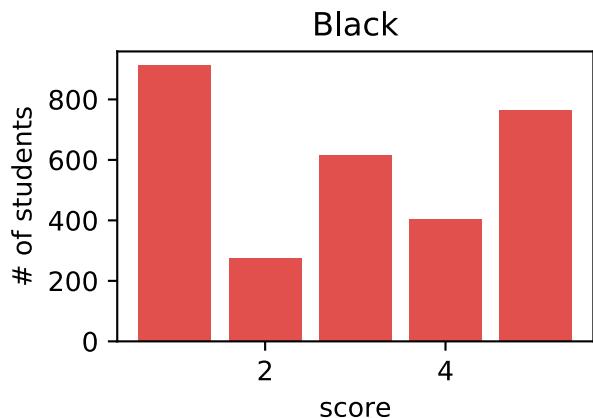
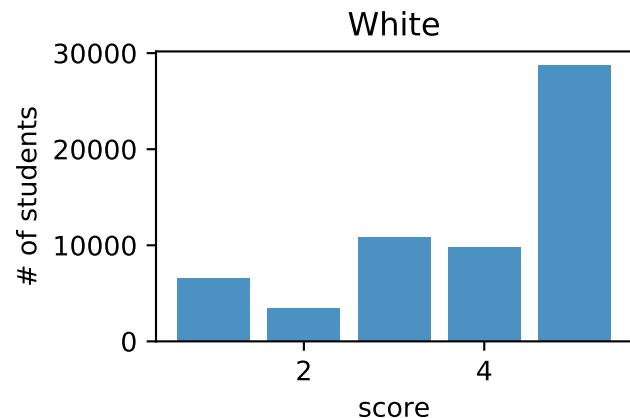
χ^2 statistic = 31852.914287287465

CALCULUS AB



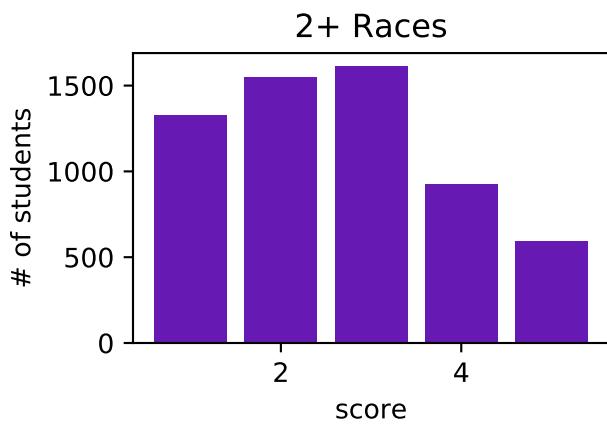
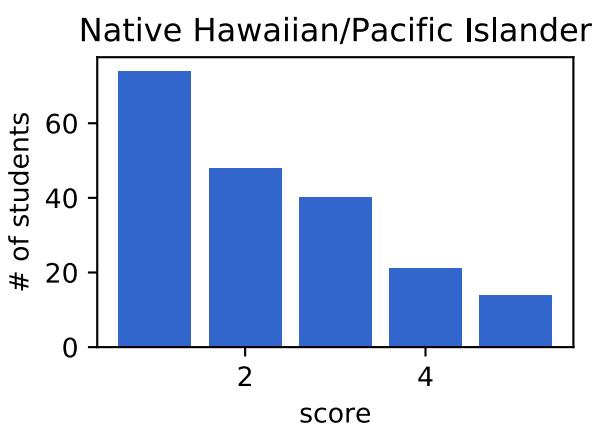
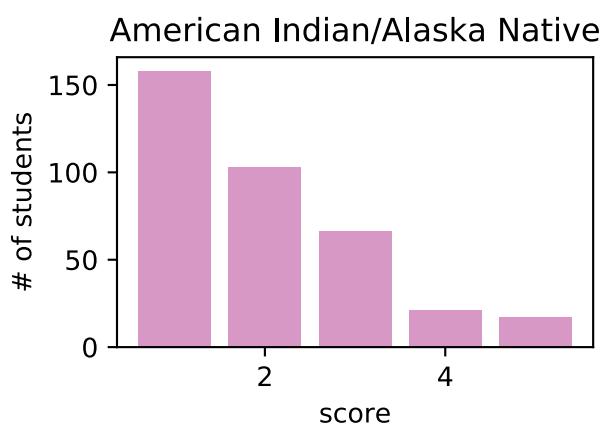
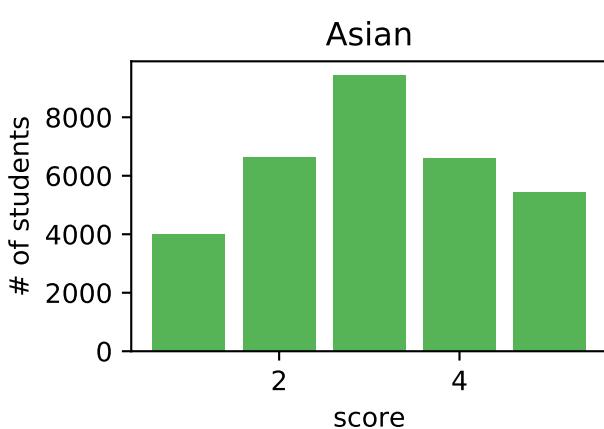
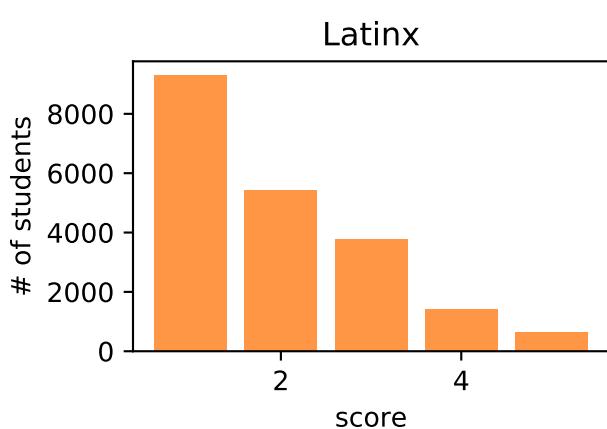
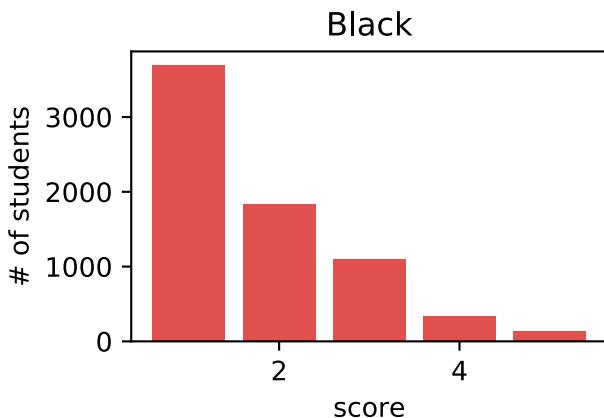
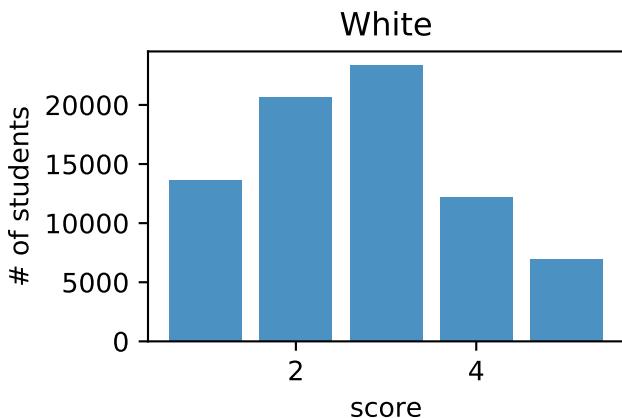
χ^2 statistic = 20658.438021060323

CALCULUS BC



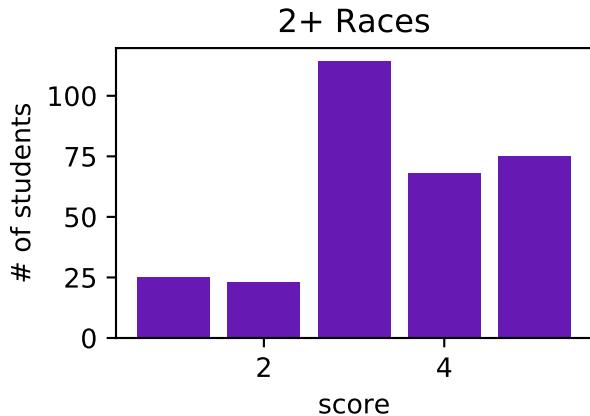
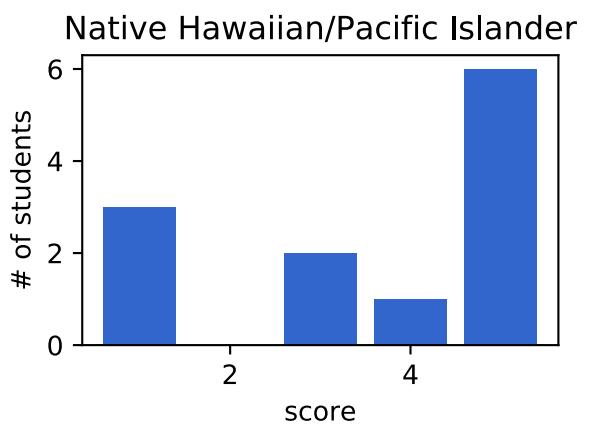
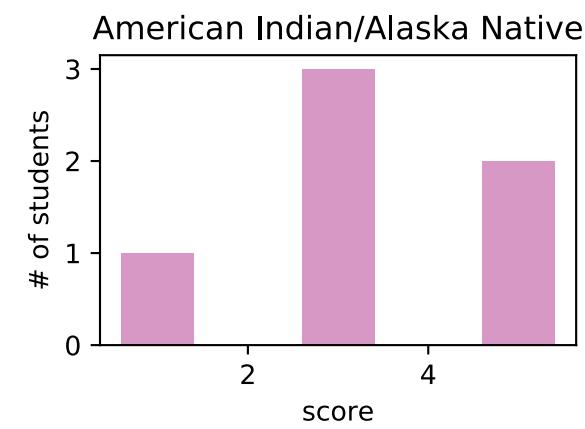
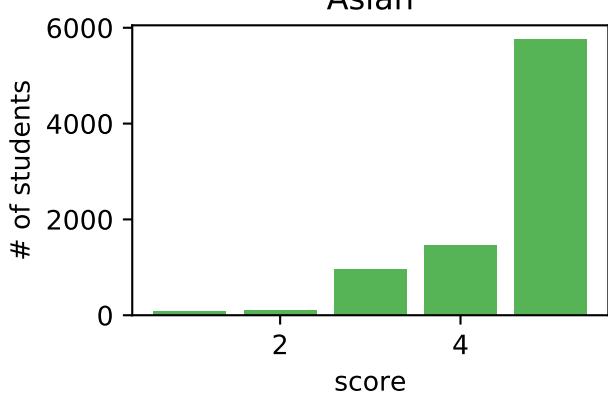
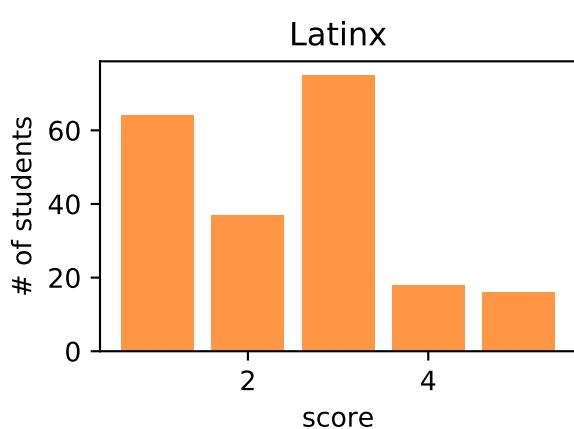
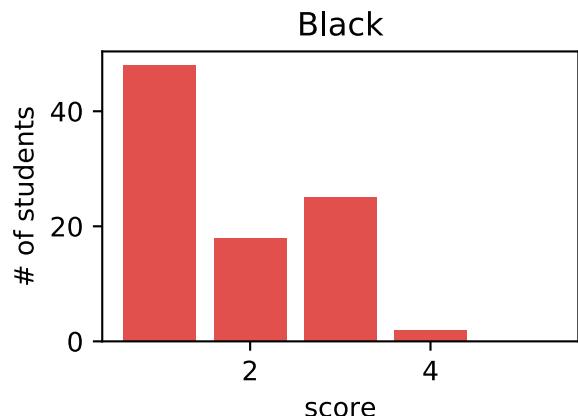
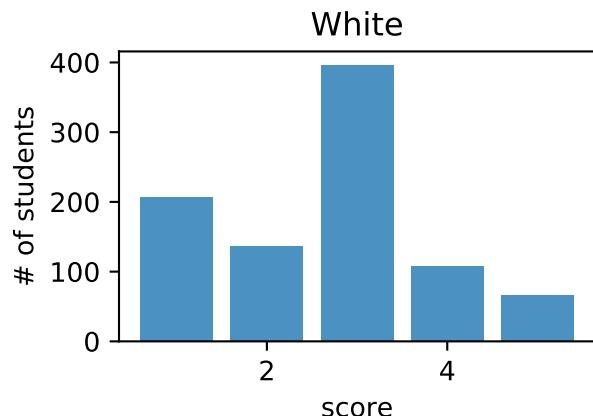
χ^2 statistic = 5020.424515568894

CHEMISTRY



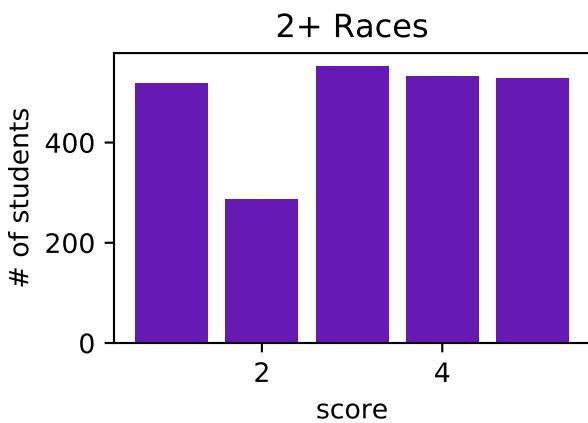
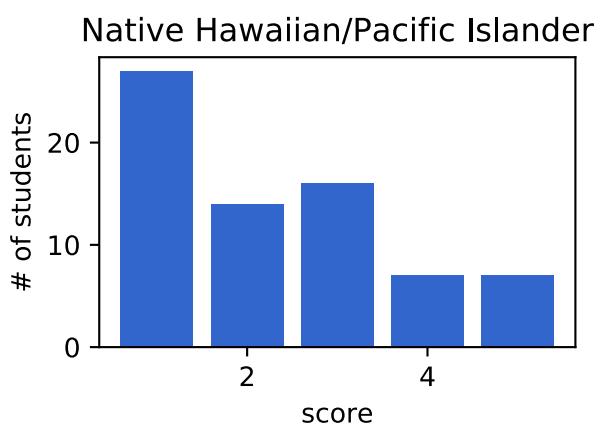
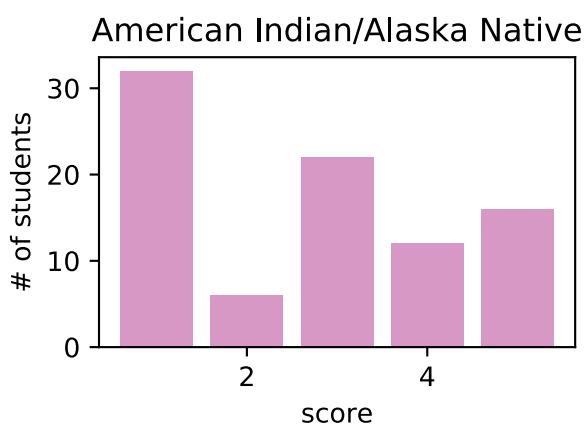
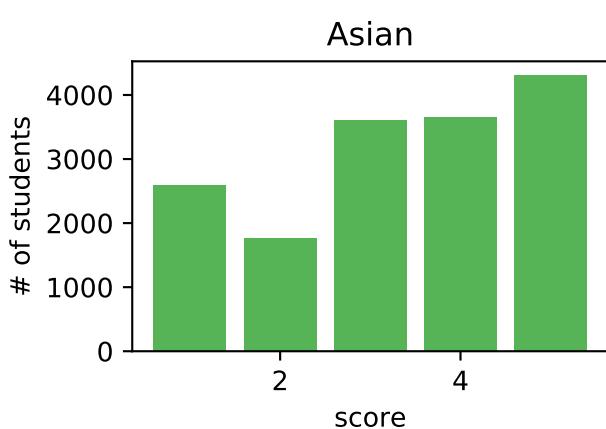
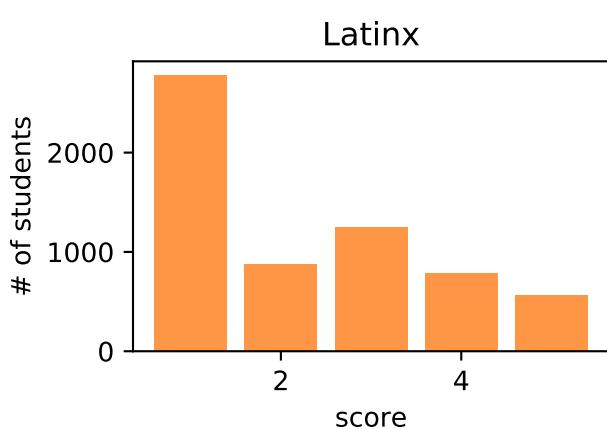
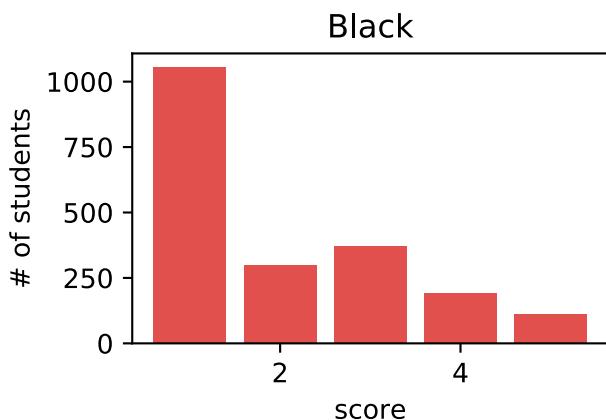
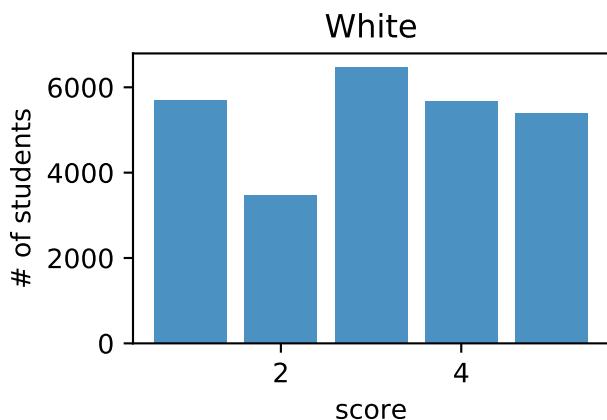
χ^2 statistic = 16705.00798891283

CHINESE LANGUAGE & CULTURE



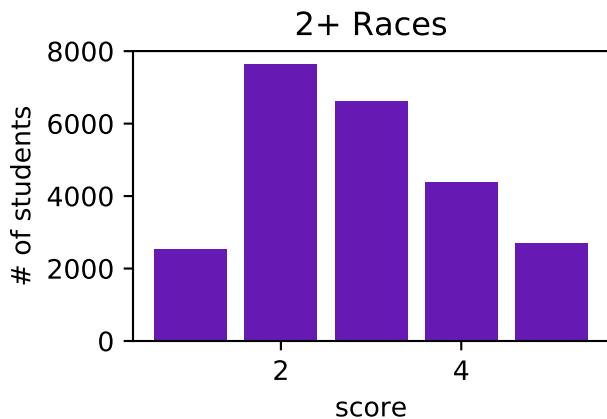
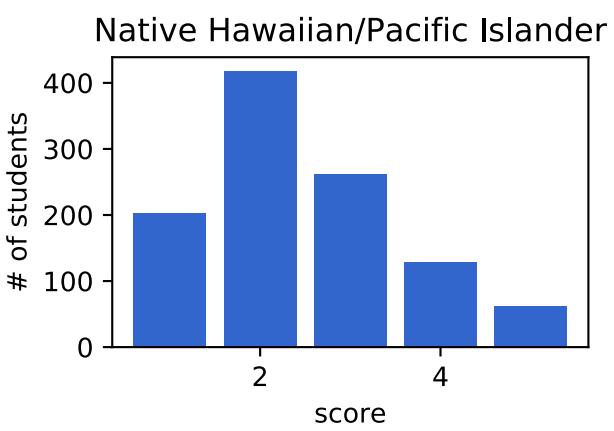
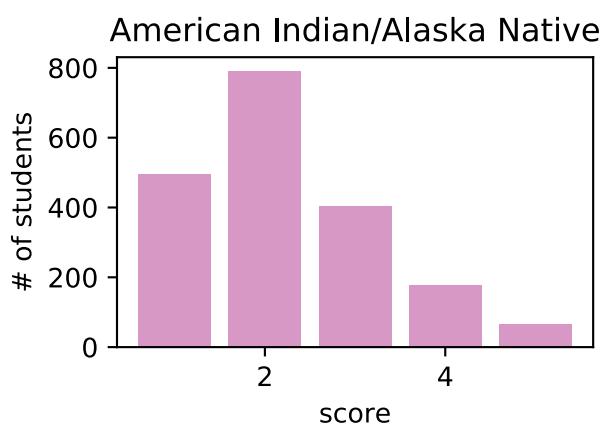
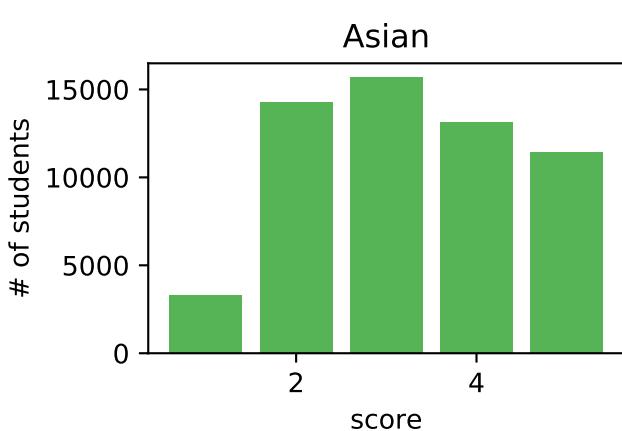
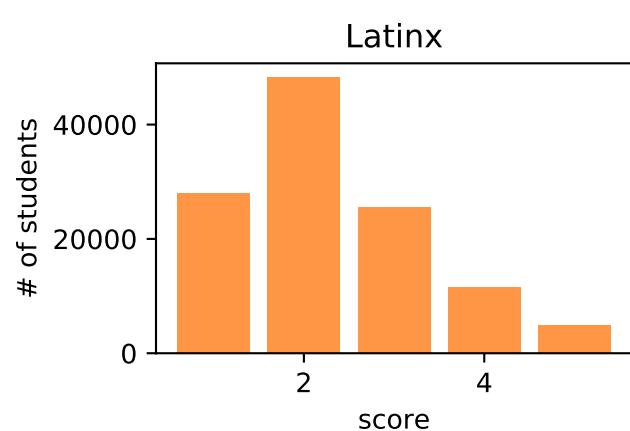
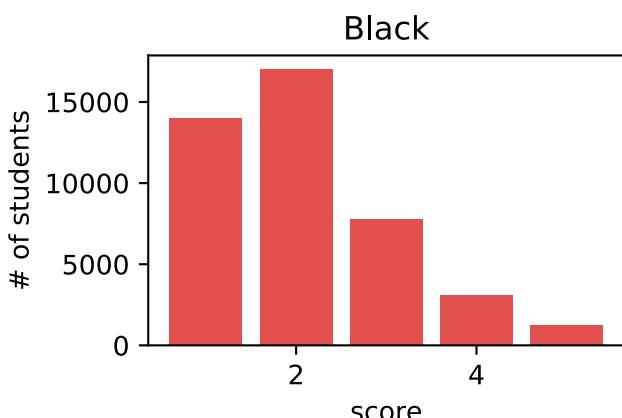
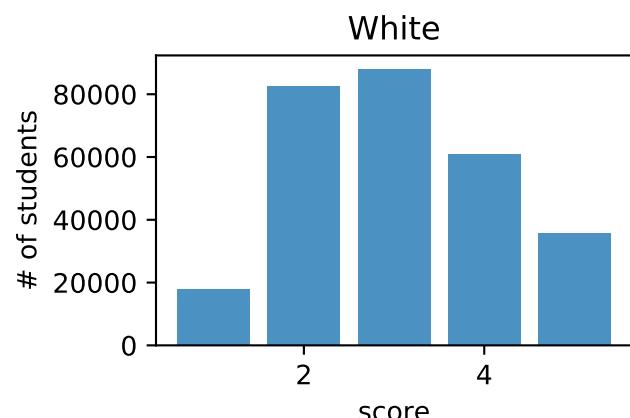
χ^2 statistic = 3947.6243712427604

COMPUTER SCIENCE A



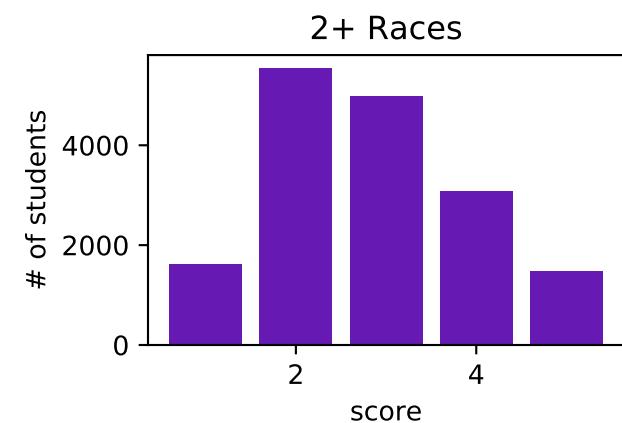
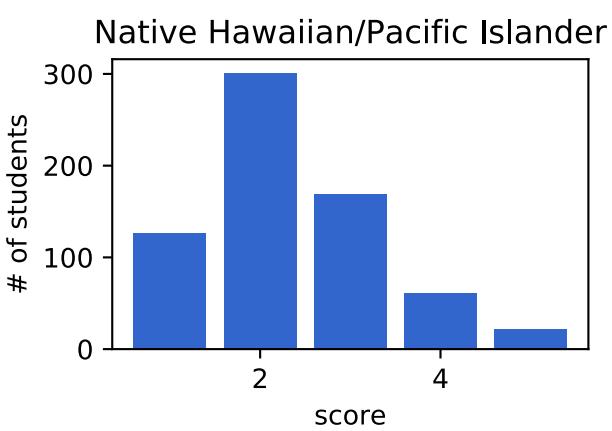
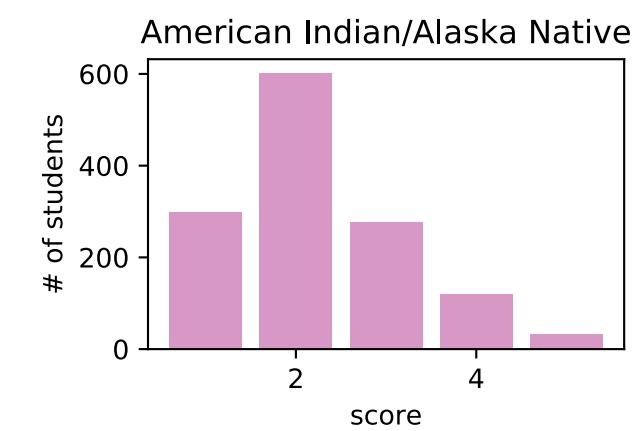
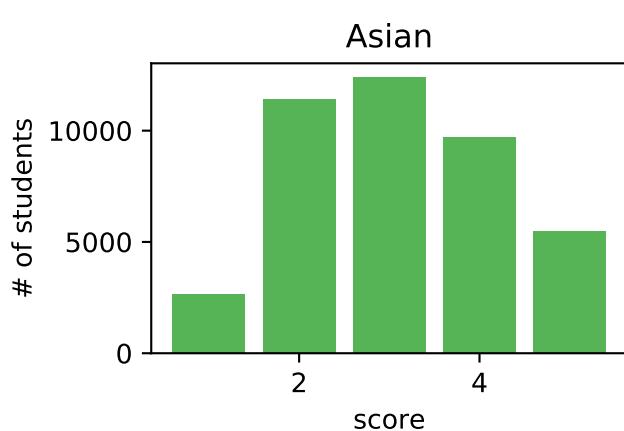
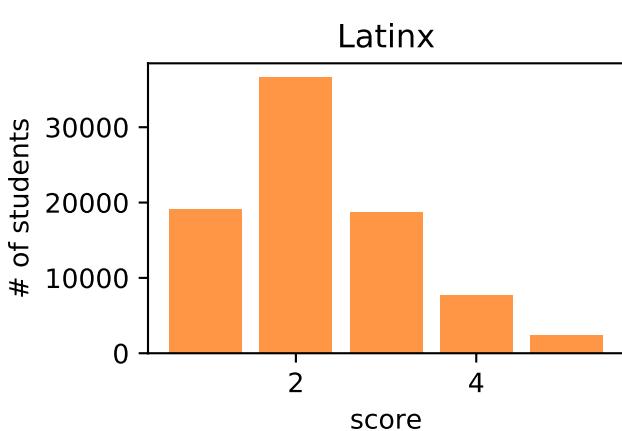
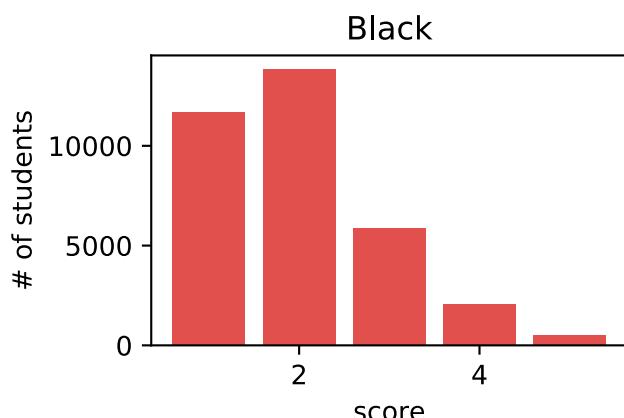
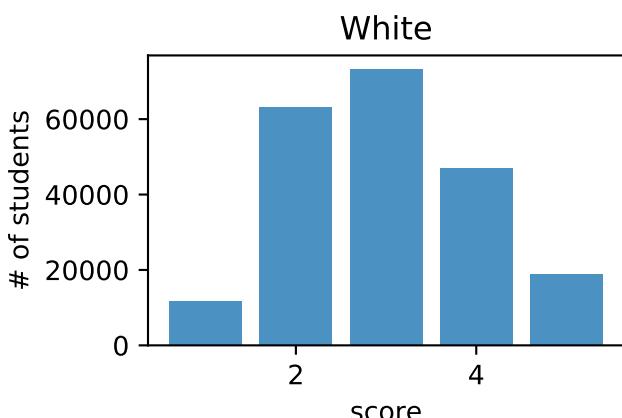
χ^2 statistic = 3744.903379030492

ENGLISH LANGUAGE & COMPOSITION



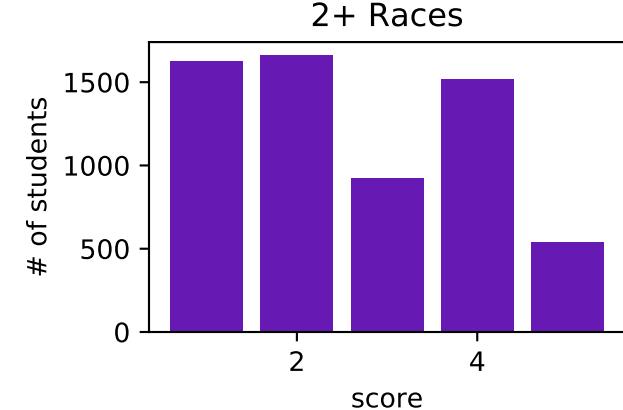
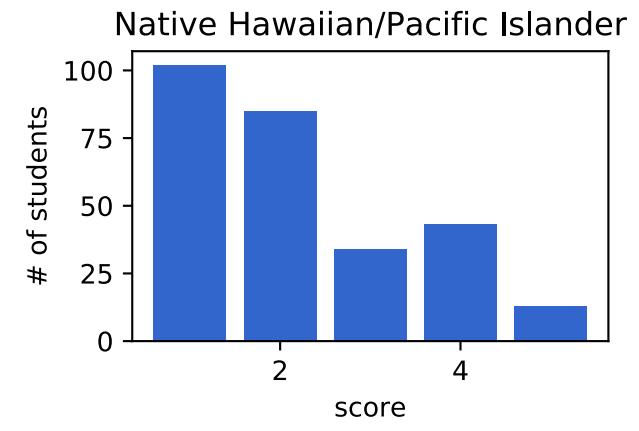
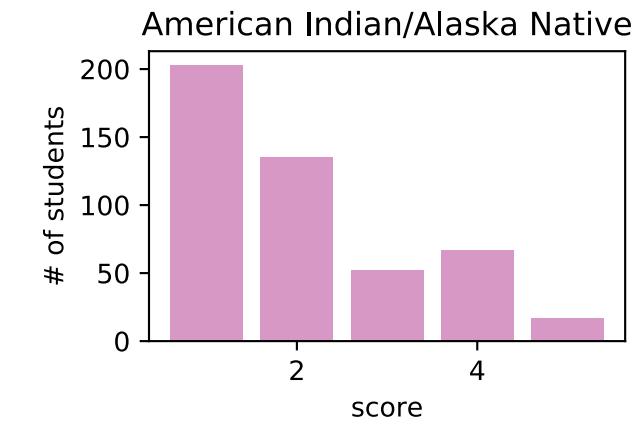
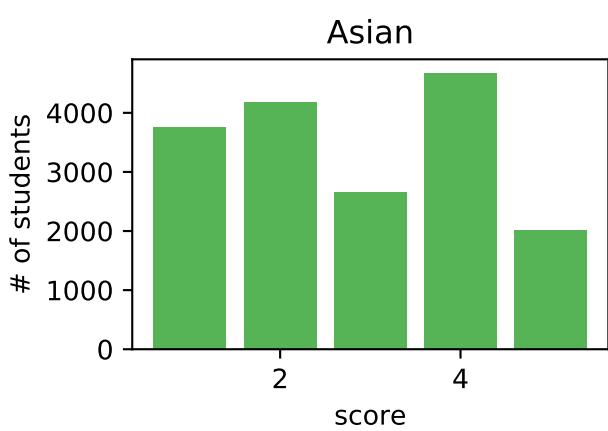
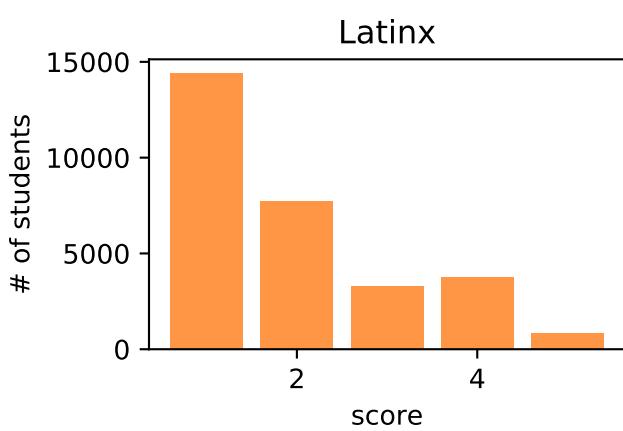
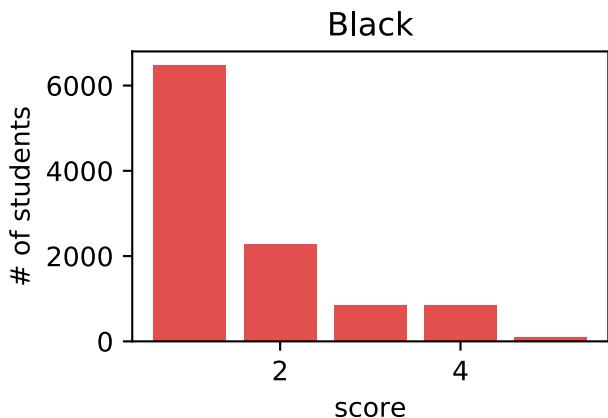
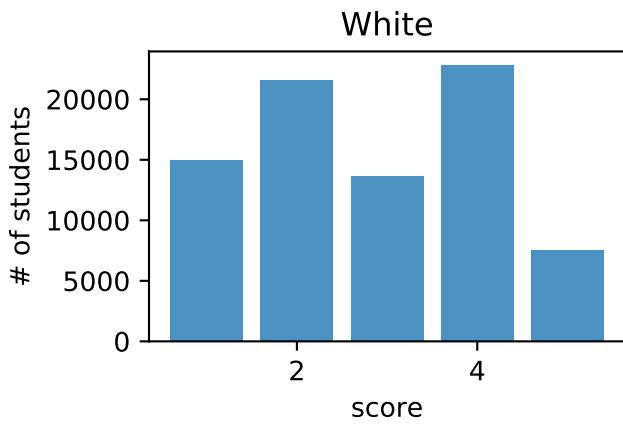
χ^2 statistic = 68849.16222965093

ENGLISH LITERATURE & COMPOSITION



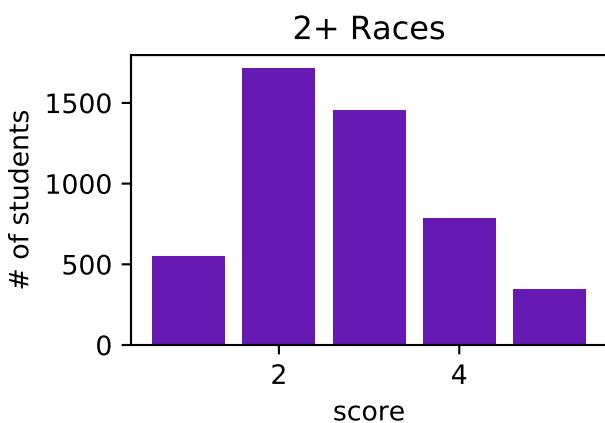
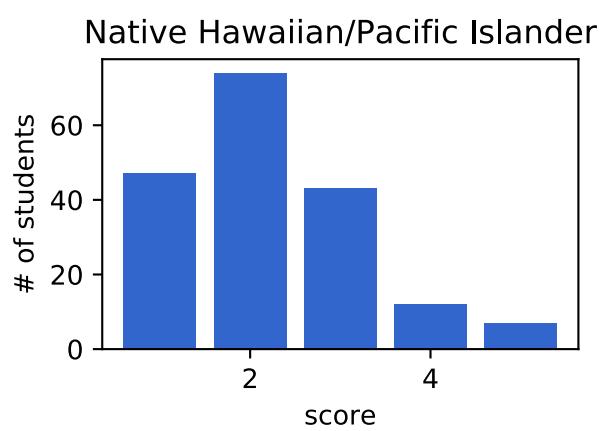
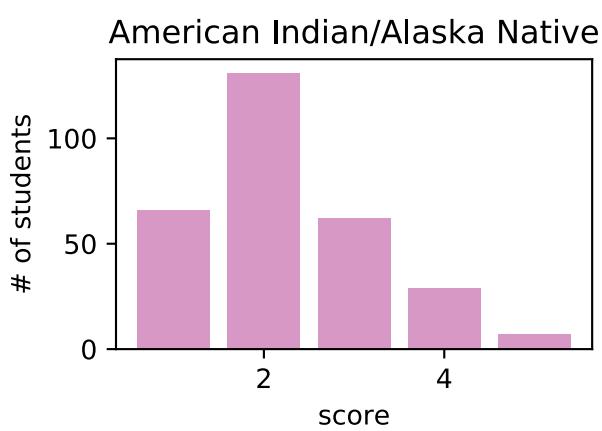
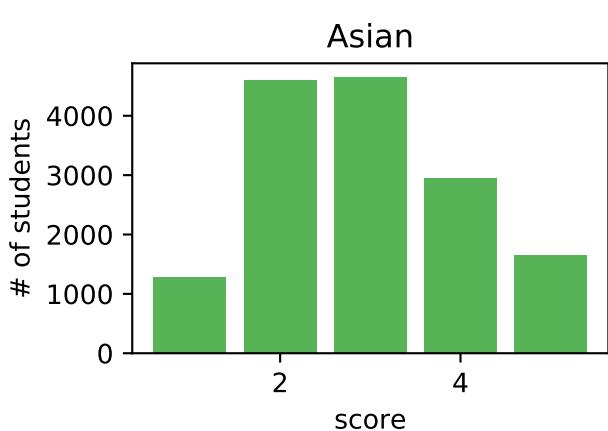
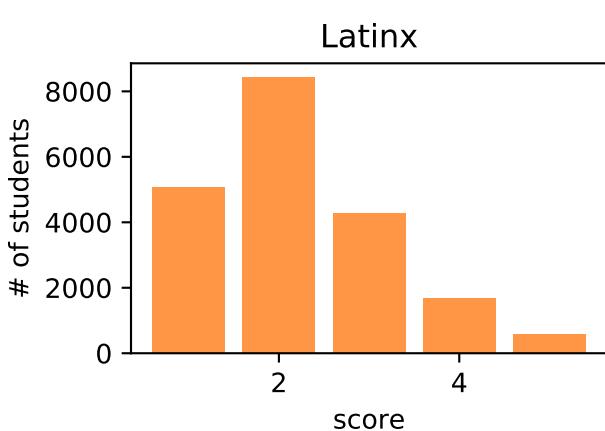
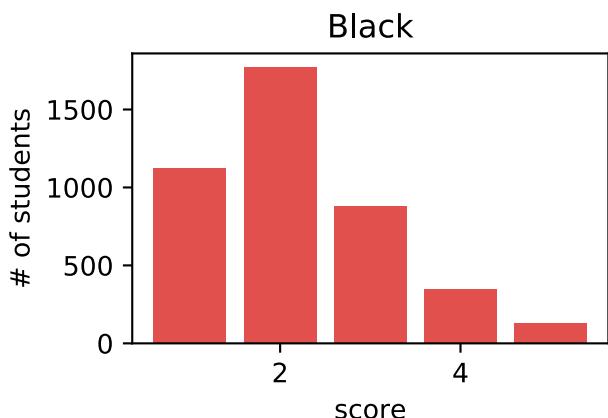
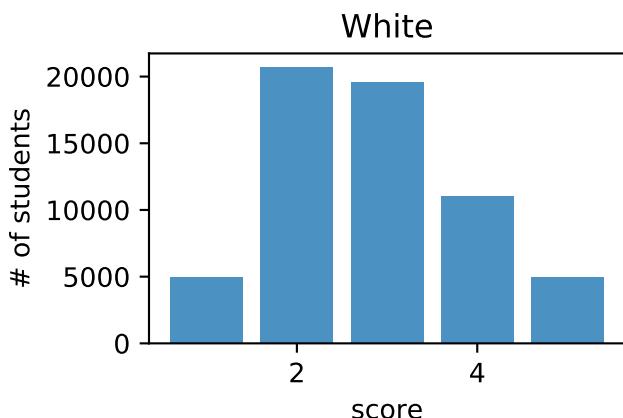
χ^2 statistic = 55961.48027014612

ENVIRONMENTAL SCIENCE



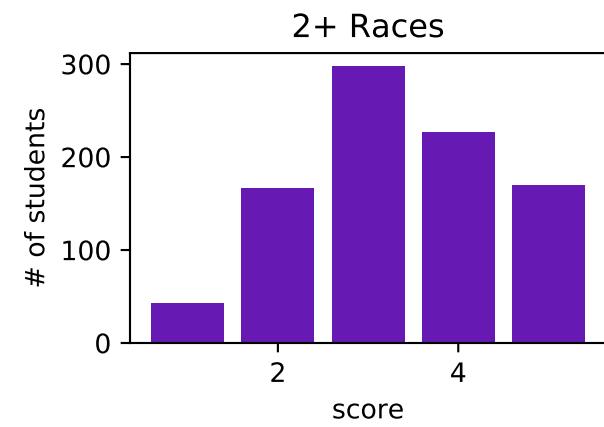
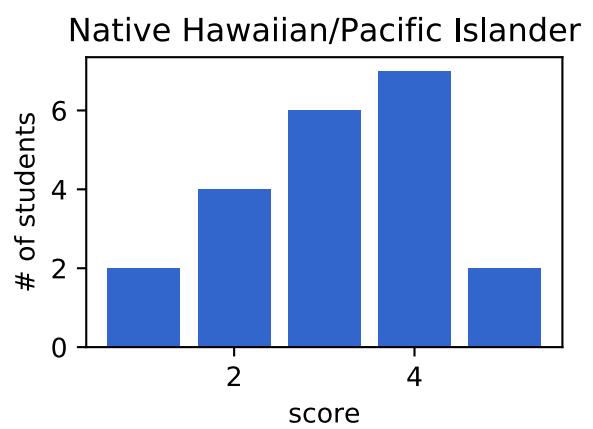
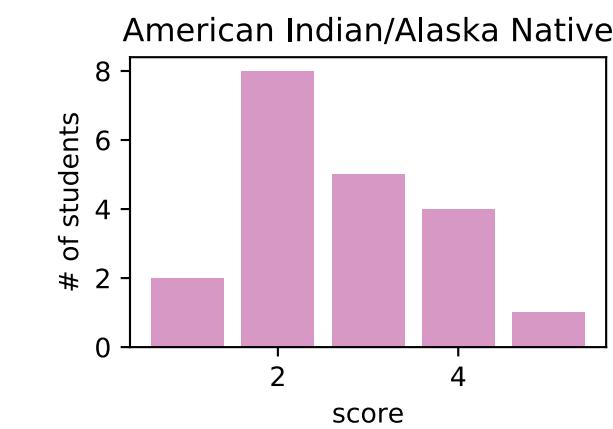
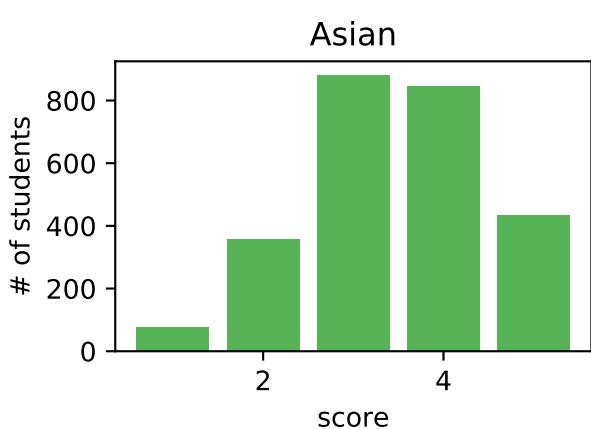
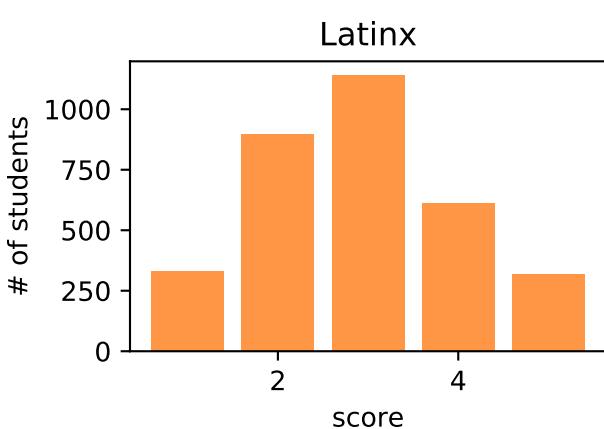
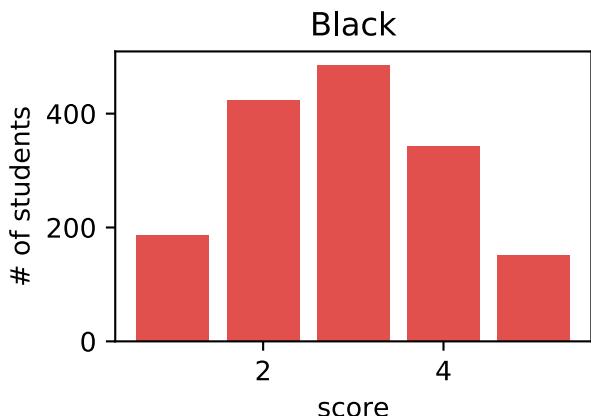
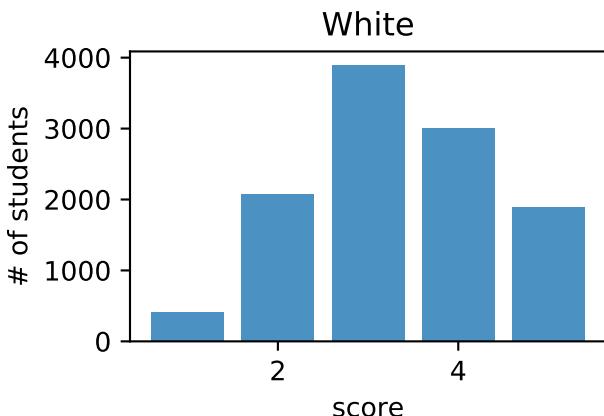
χ^2 statistic = 17868.820030633822

EUROPEAN HISTORY



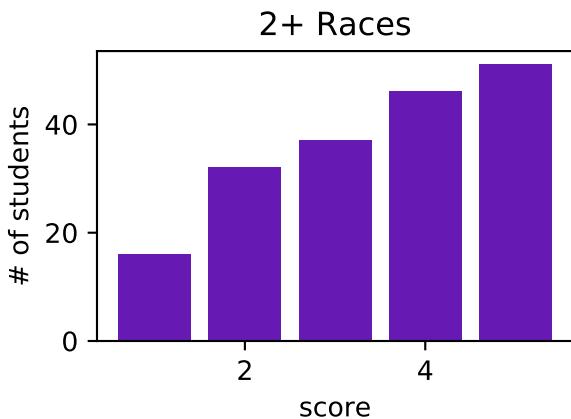
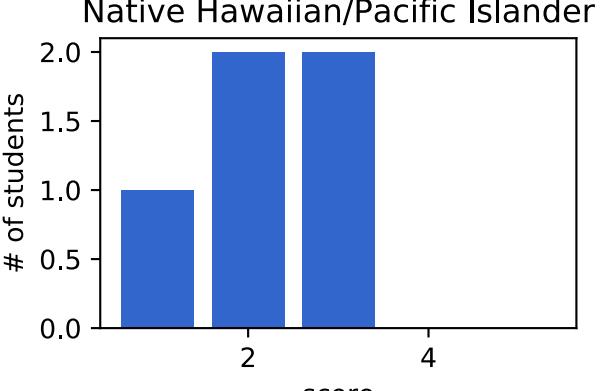
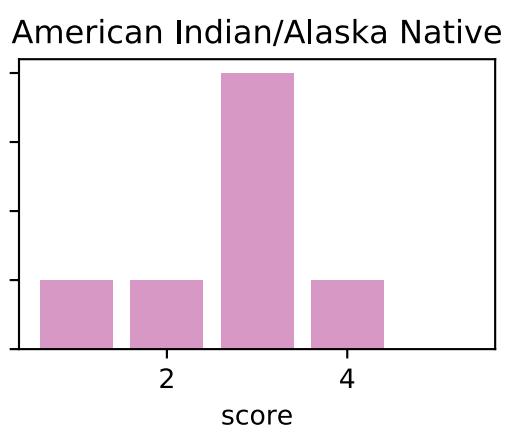
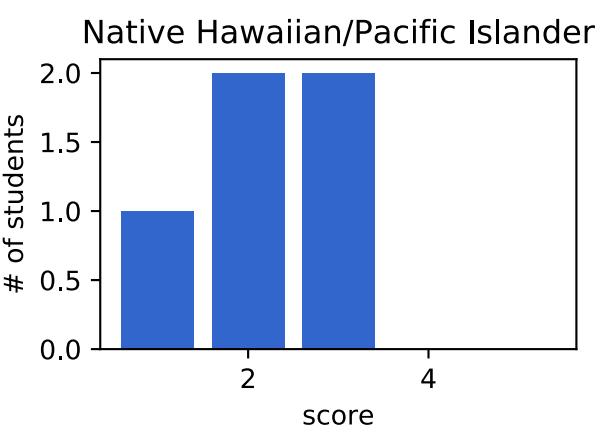
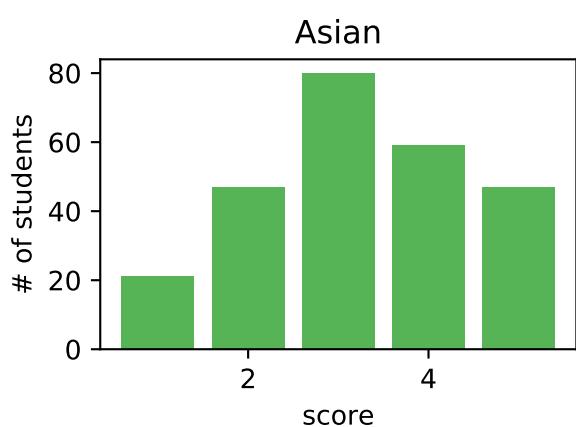
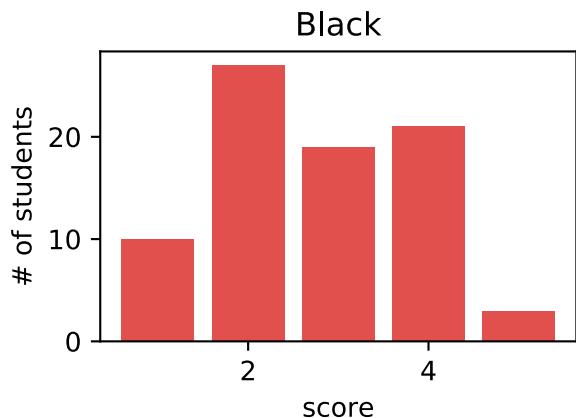
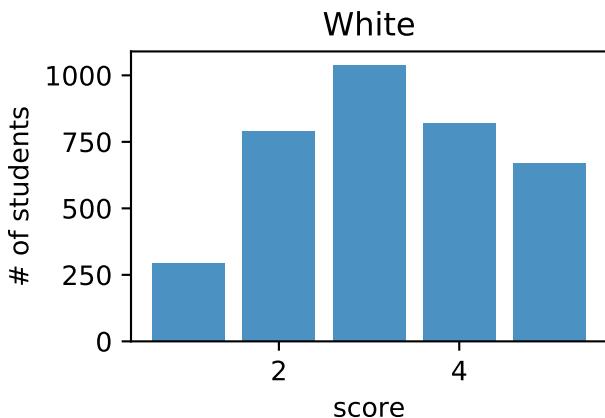
χ^2 statistic = 7851.317015313369

FRENCH LANGUAGE & CULTURE



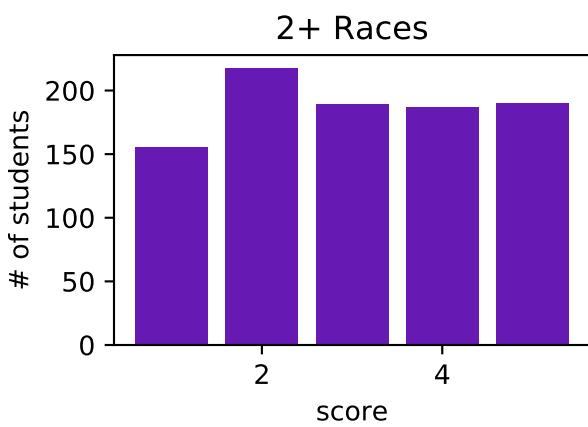
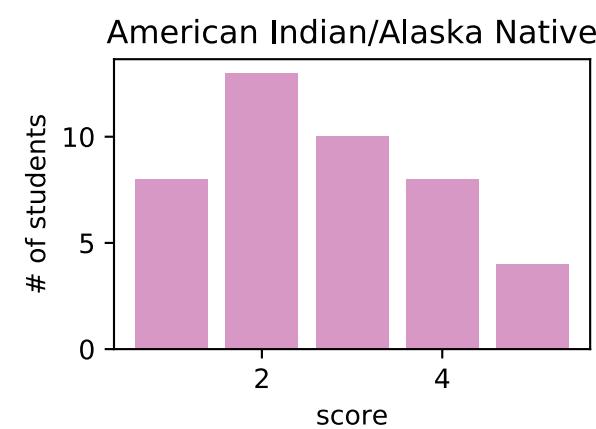
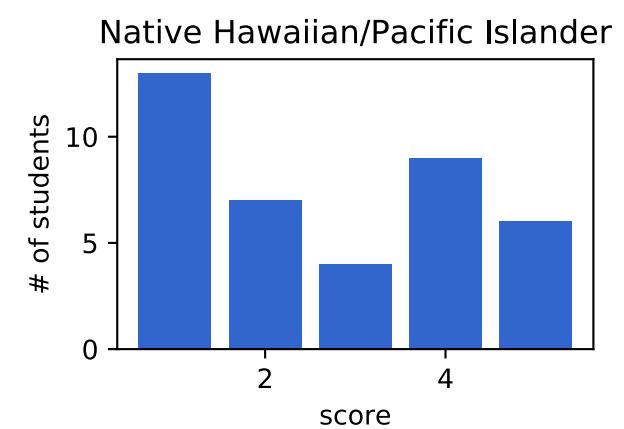
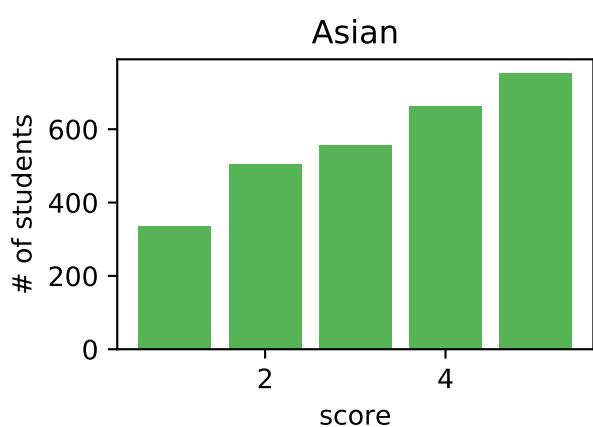
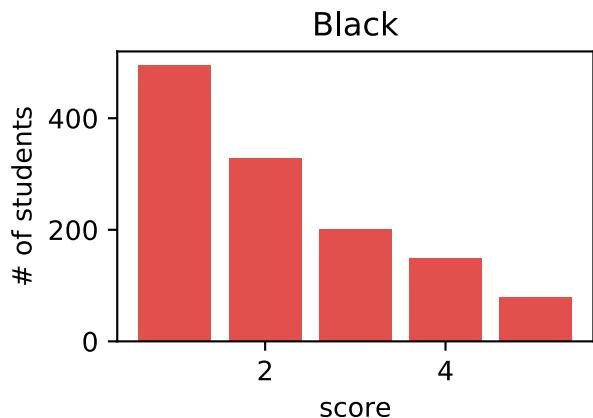
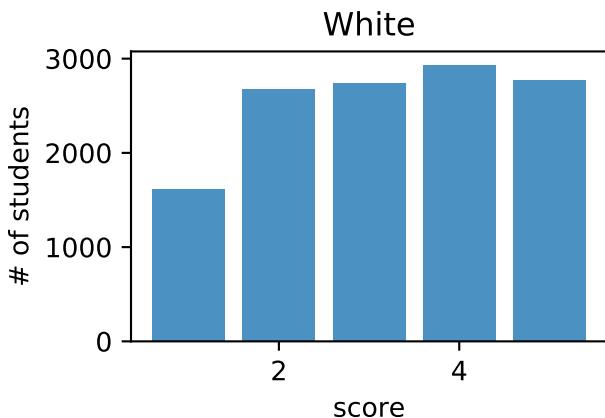
χ^2 statistic = 809.7147731714613

GERMAN LANGUAGE & CULTURE



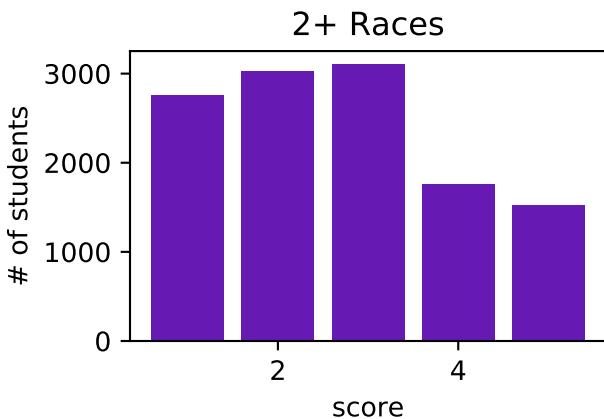
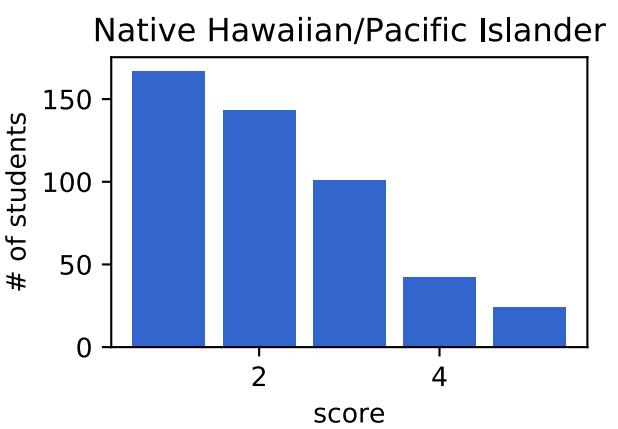
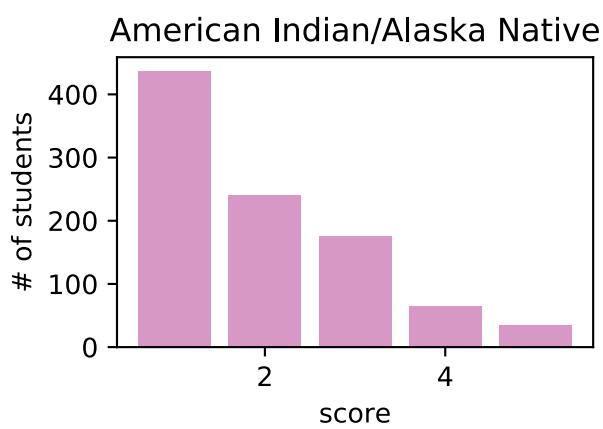
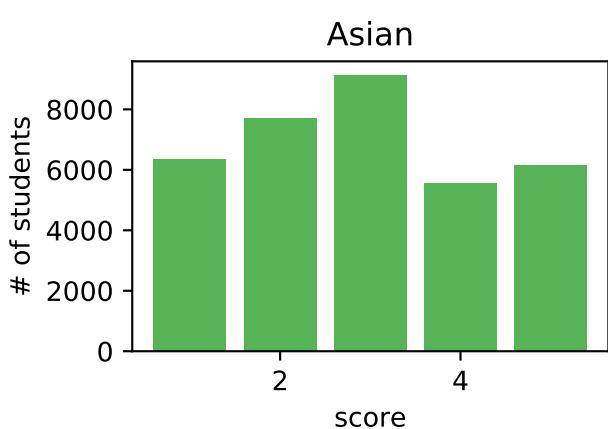
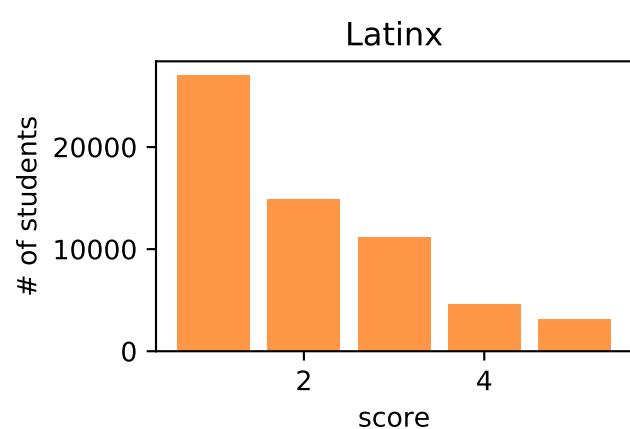
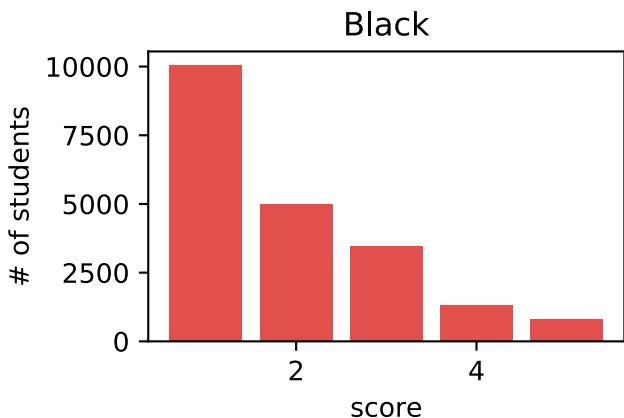
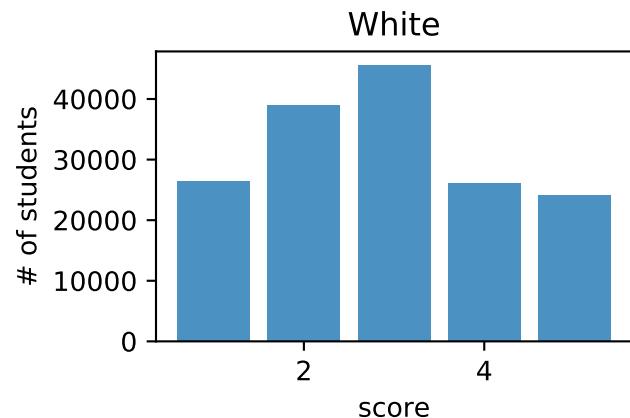
χ^2 statistic = 106.25565206157205

GOVERNMENT & POLITICS: COMPARATIVE



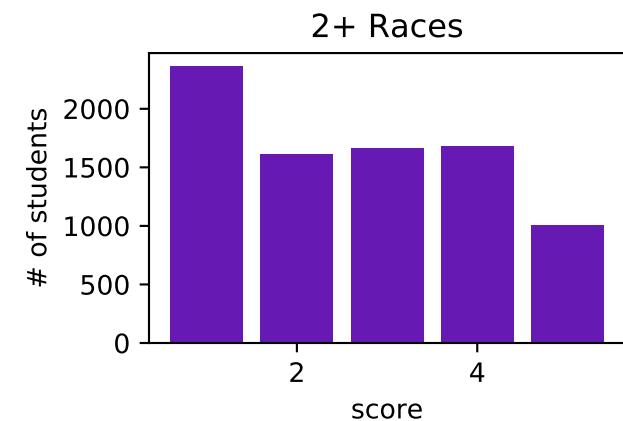
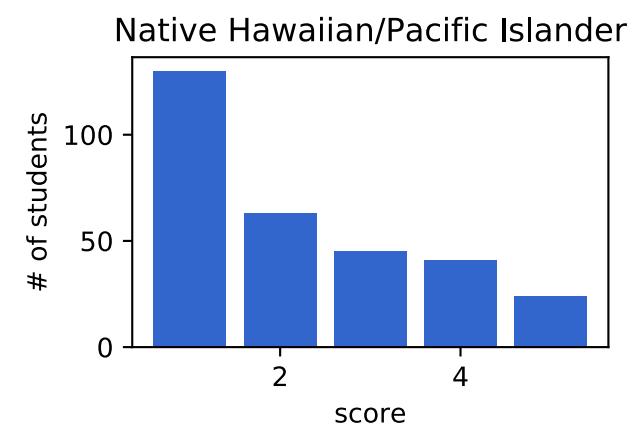
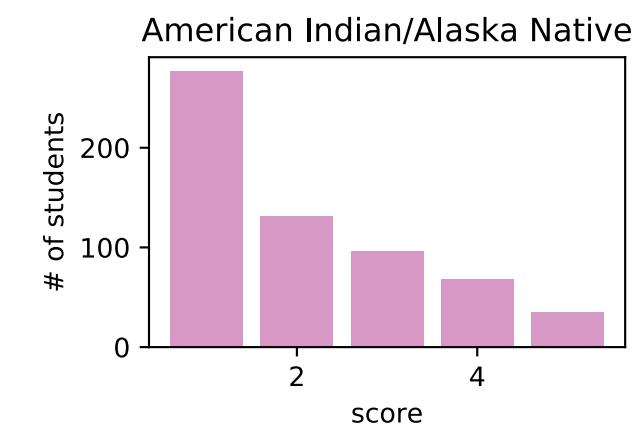
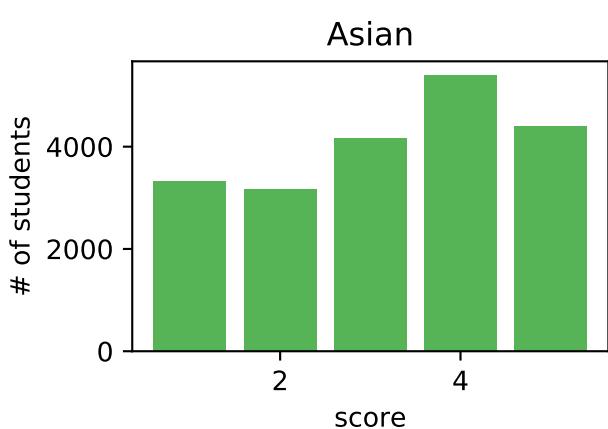
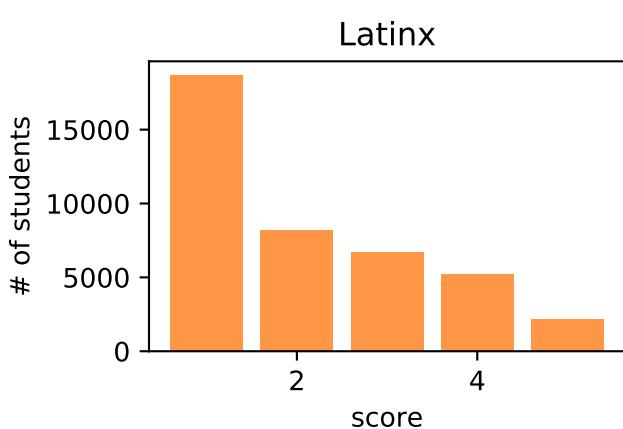
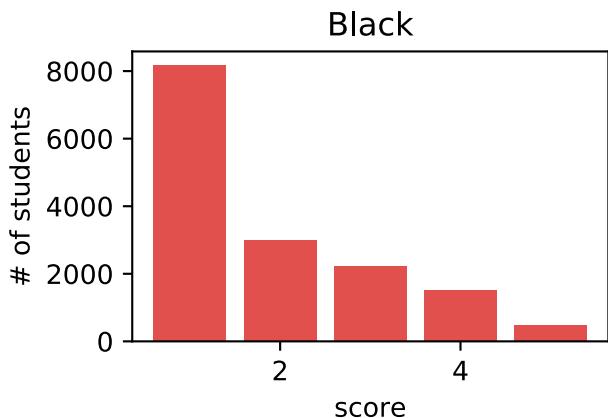
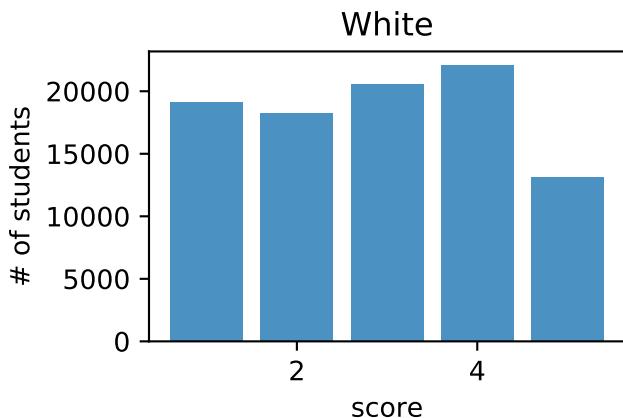
χ^2 statistic = 1230.640144009041

GOVERNMENT & POLITICS: U.S.



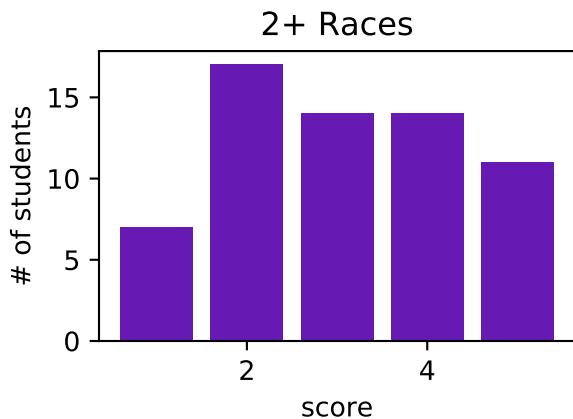
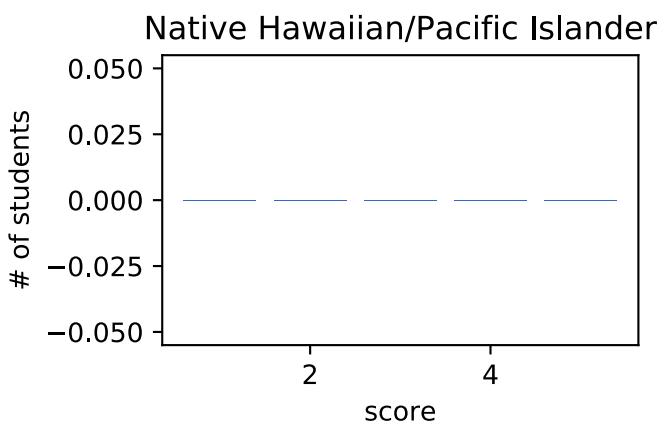
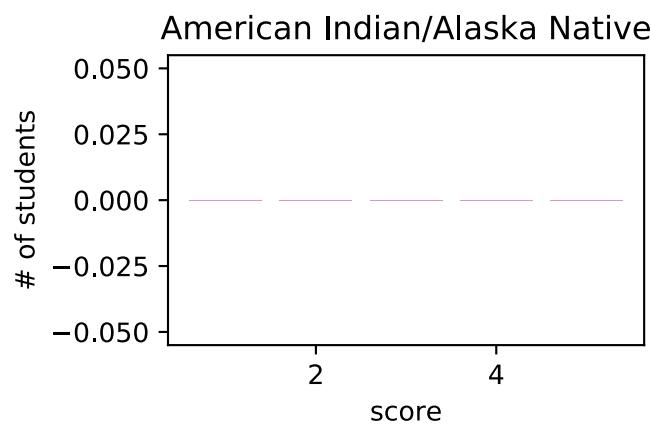
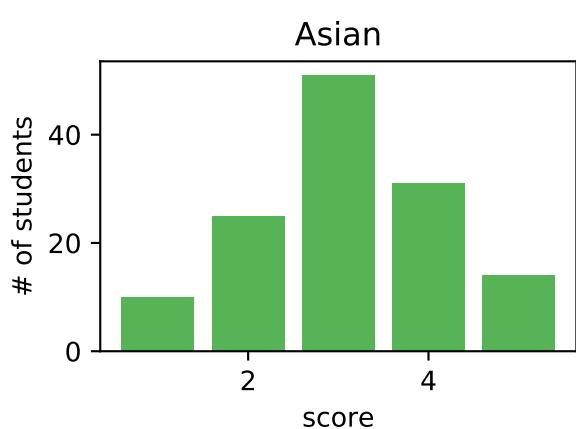
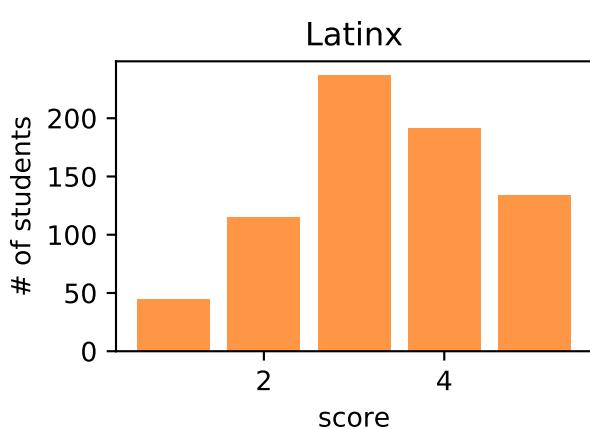
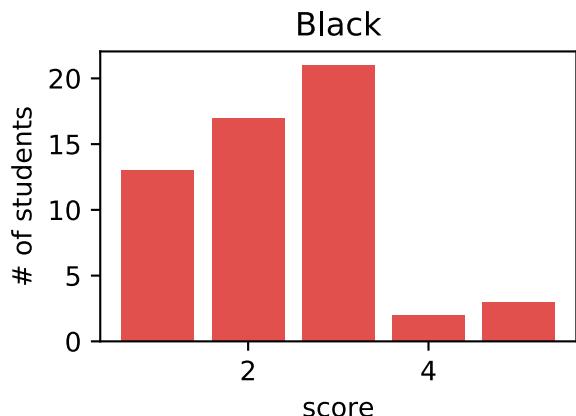
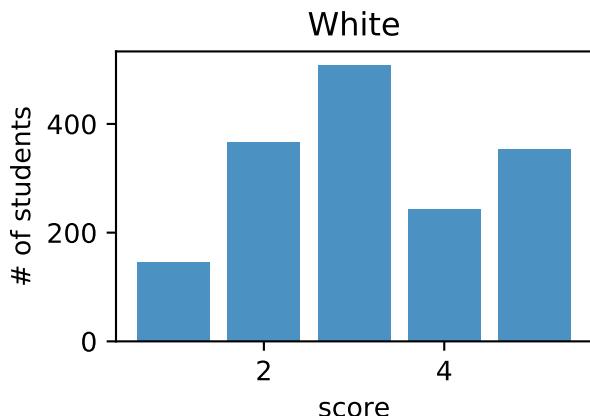
χ^2 statistic = 30704.758451443937

HUMAN GEOGRAPHY



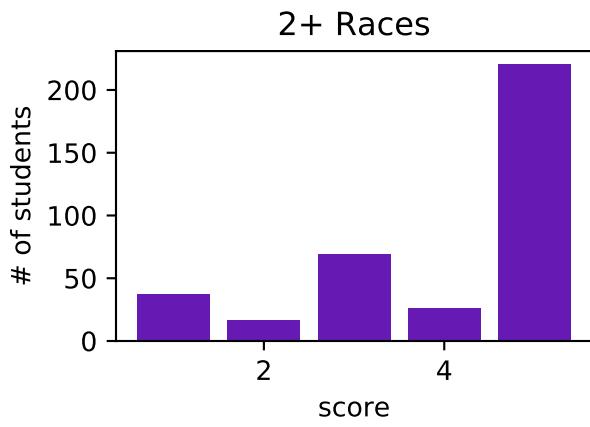
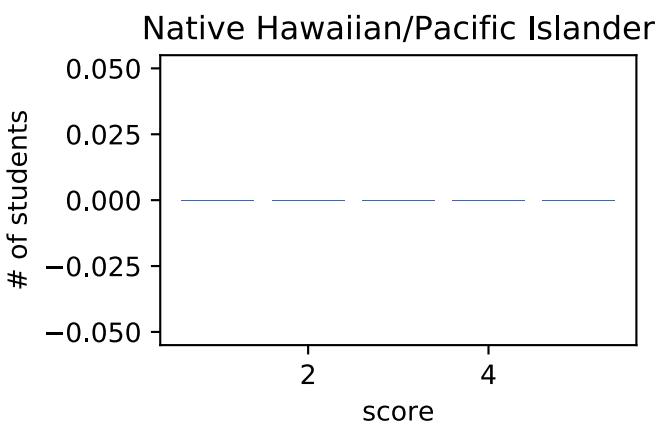
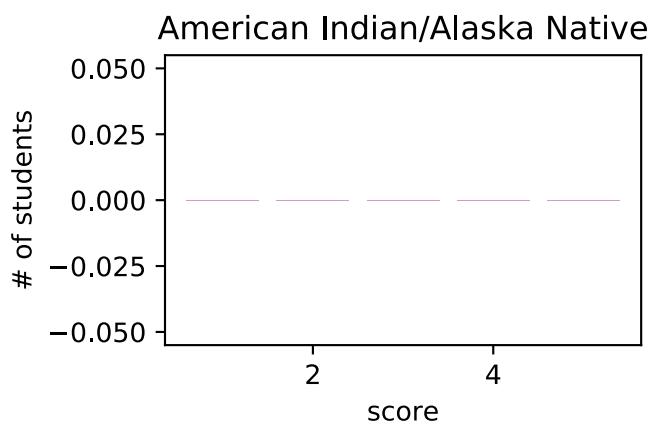
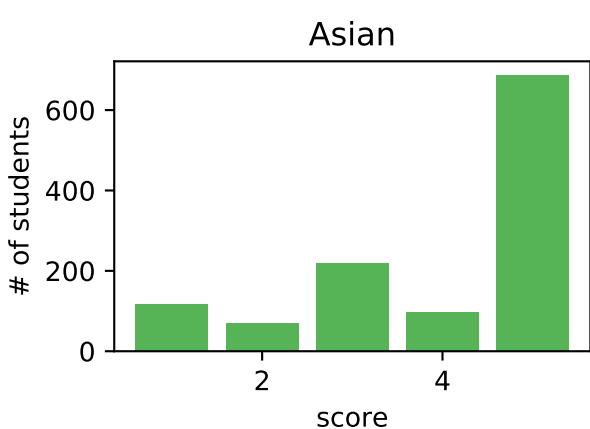
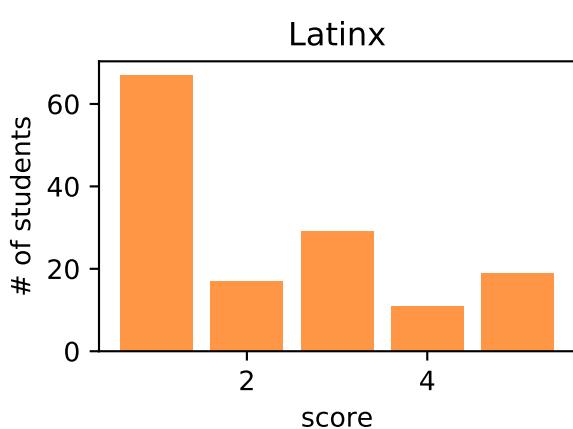
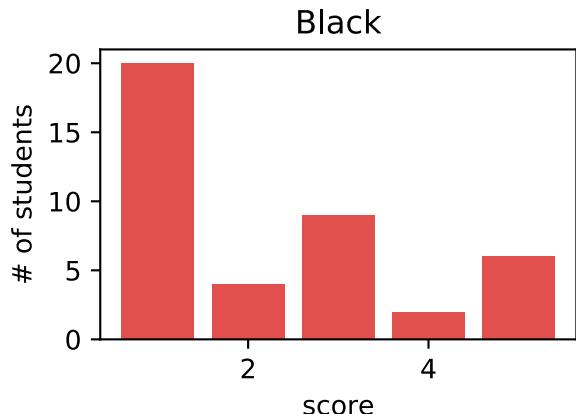
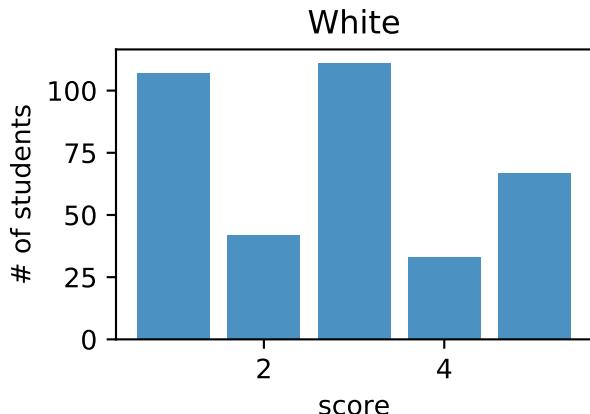
χ^2 statistic = 18820.50291342107

ITALIAN LANGUAGE & CULTURE



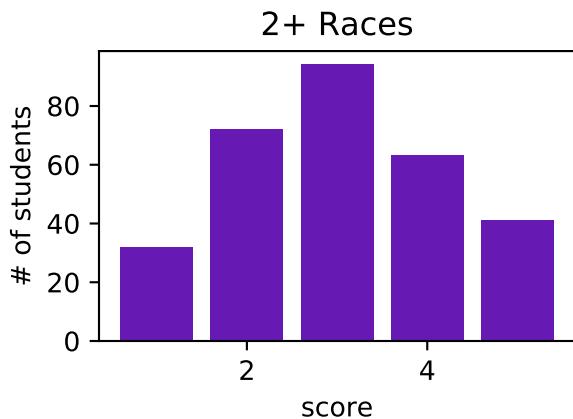
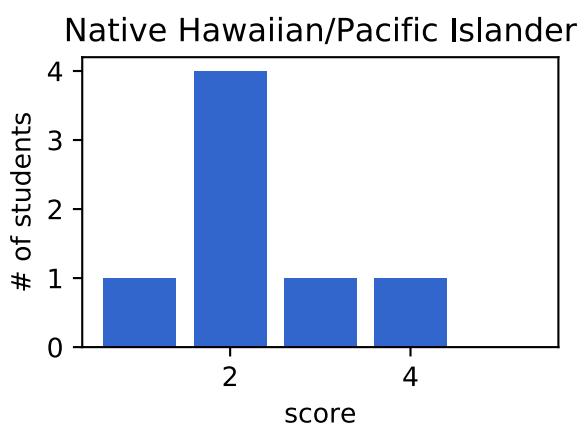
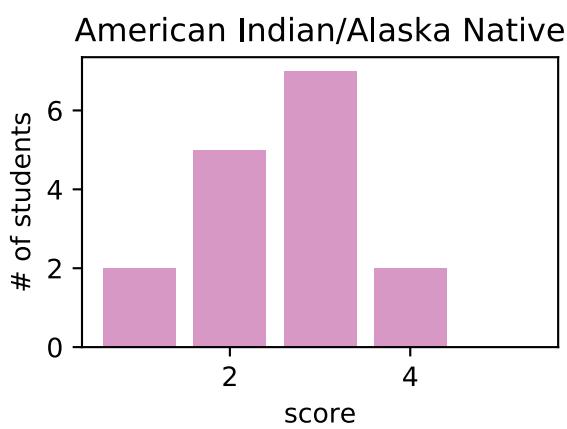
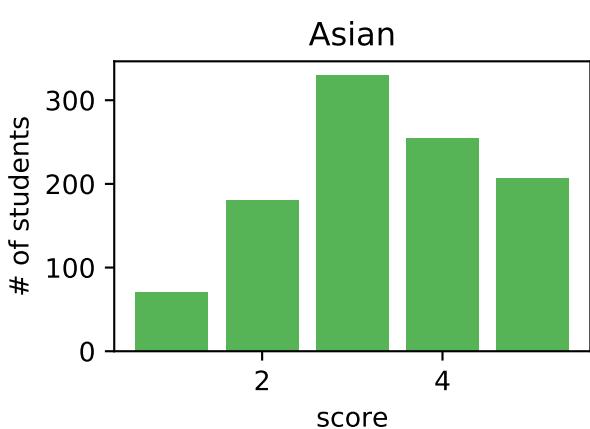
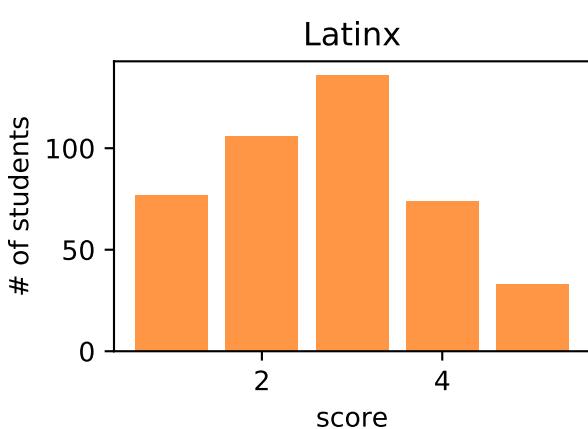
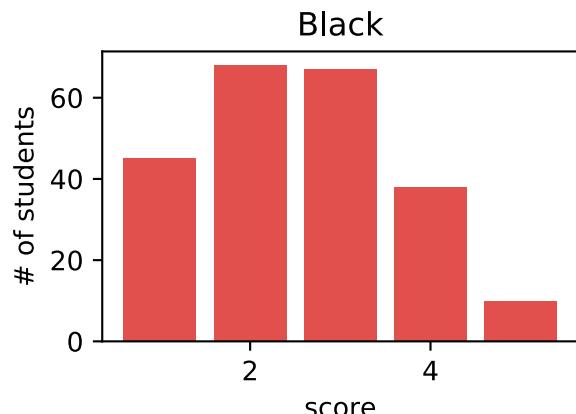
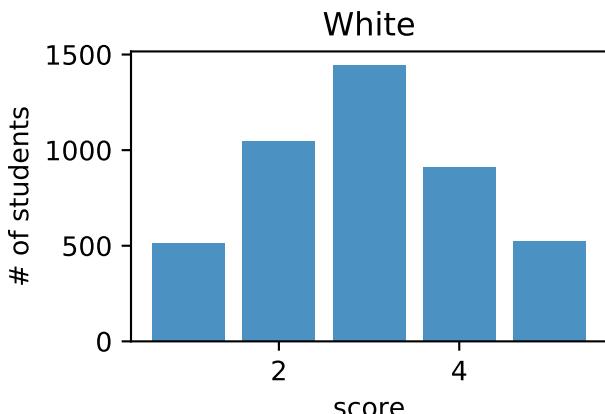
χ^2 statistic = 101.32997190010227

JAPANESE LANGUAGE & CULTURE



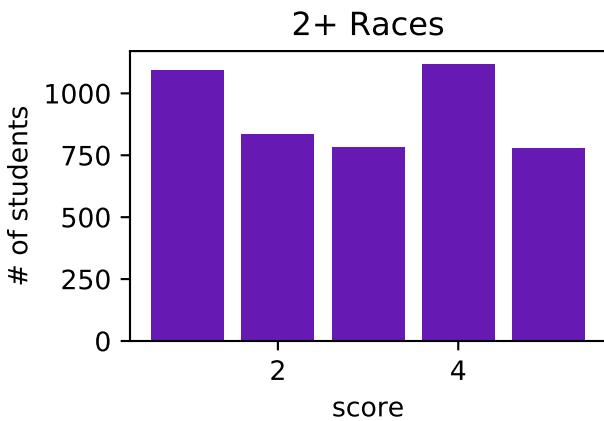
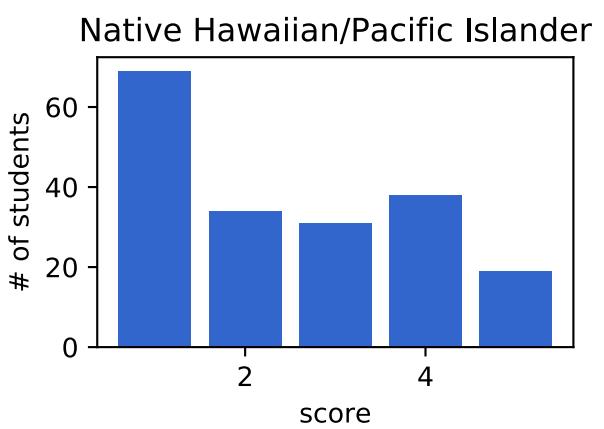
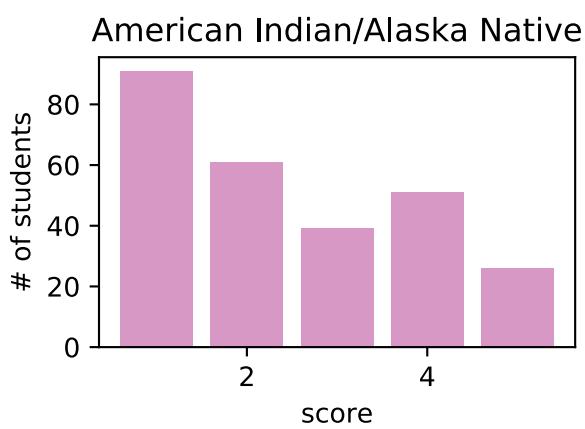
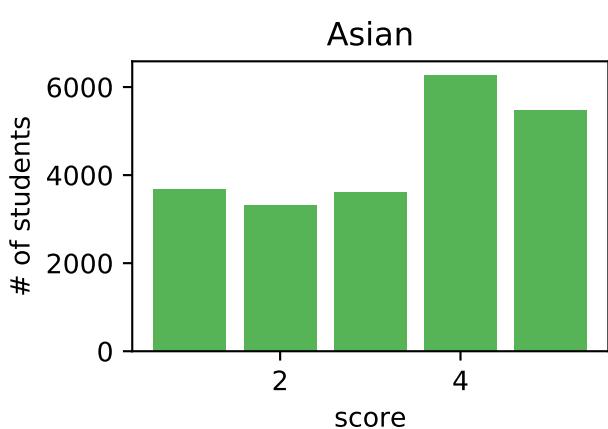
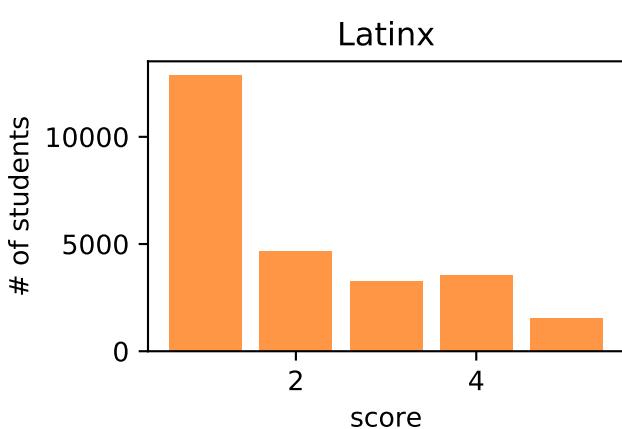
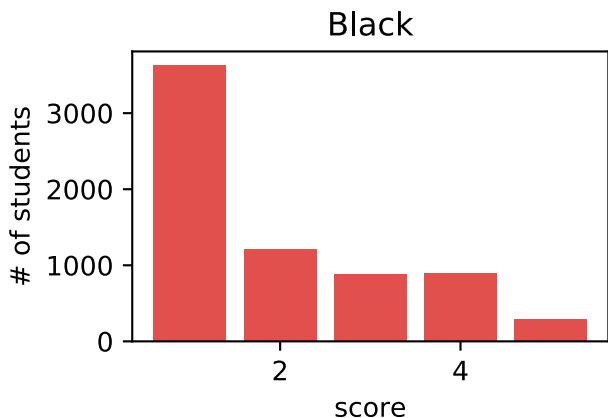
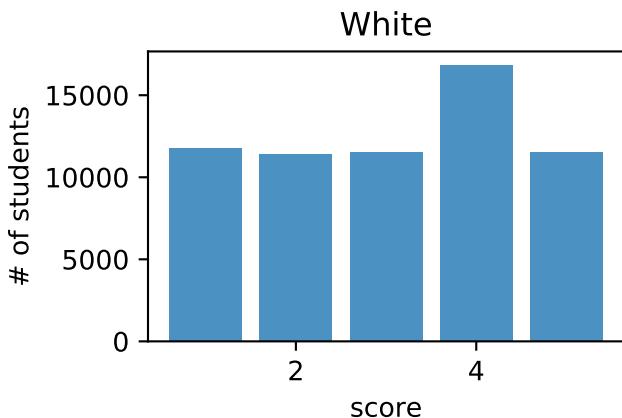
χ^2 statistic = 381.22520704794005

LATIN



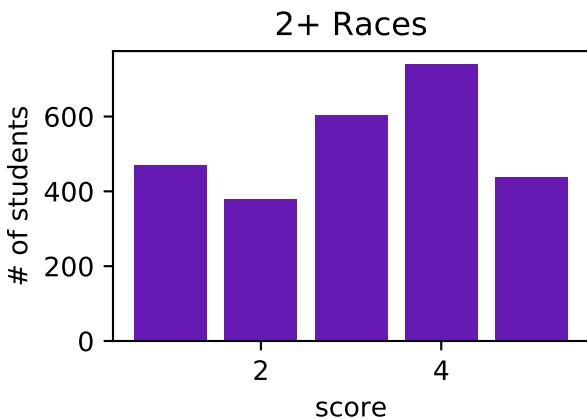
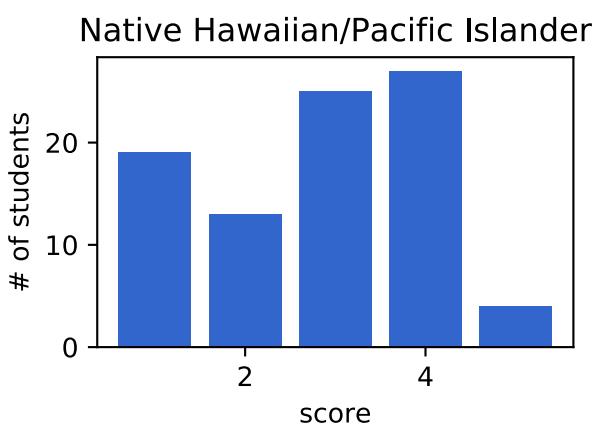
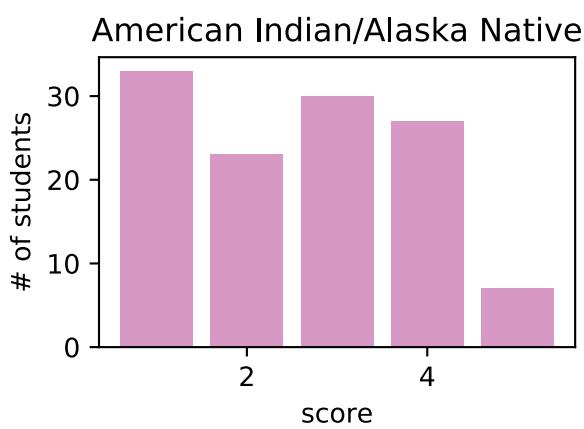
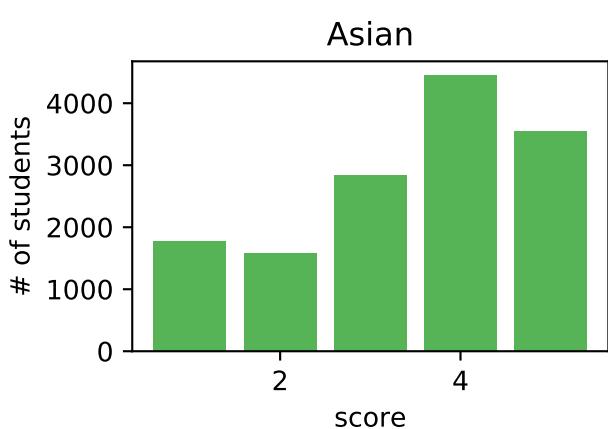
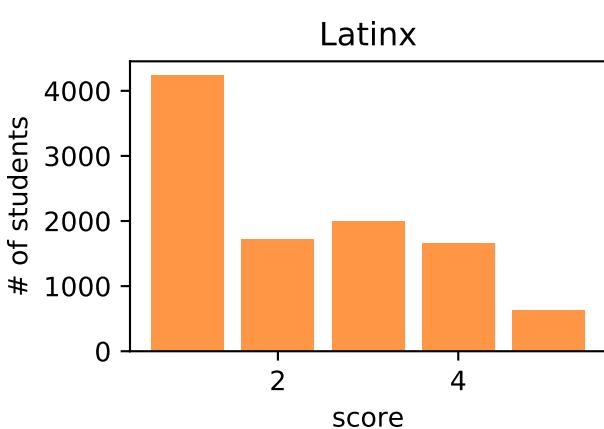
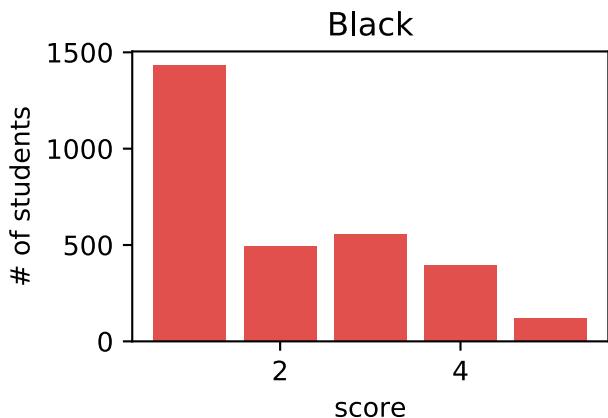
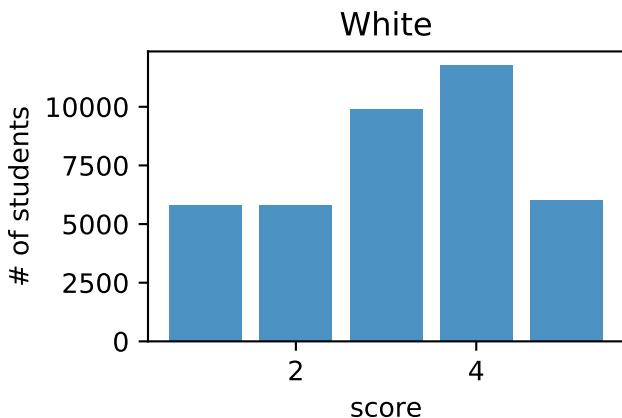
χ^2 statistic = 157.57580816751528

MACROECONOMICS



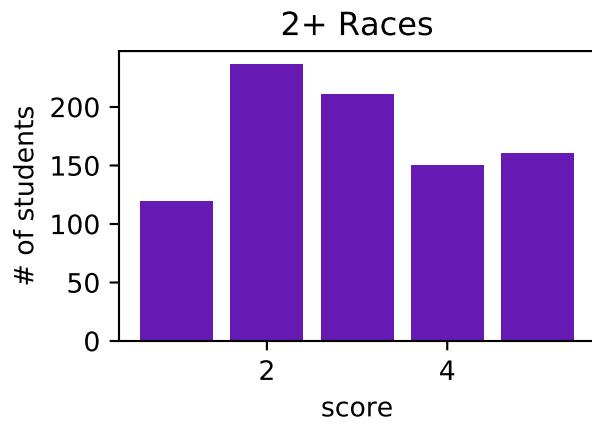
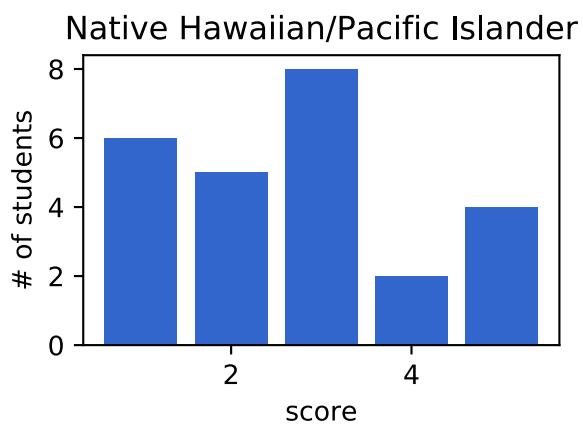
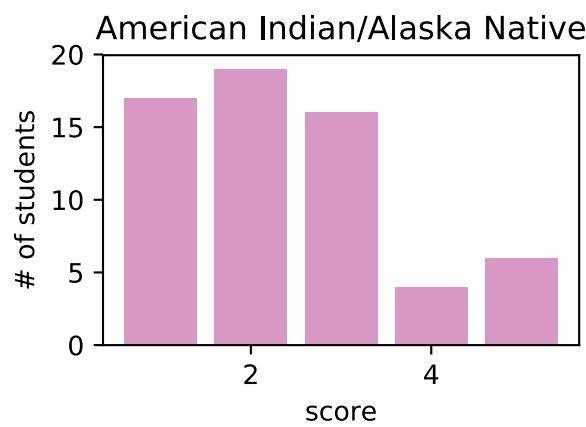
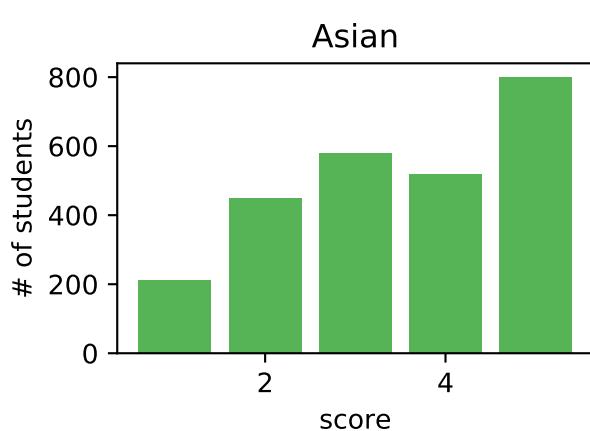
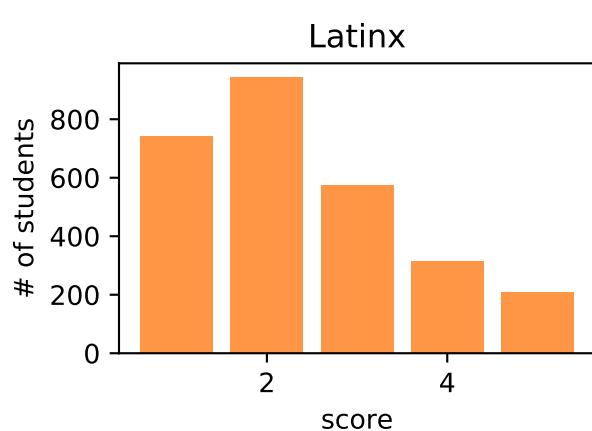
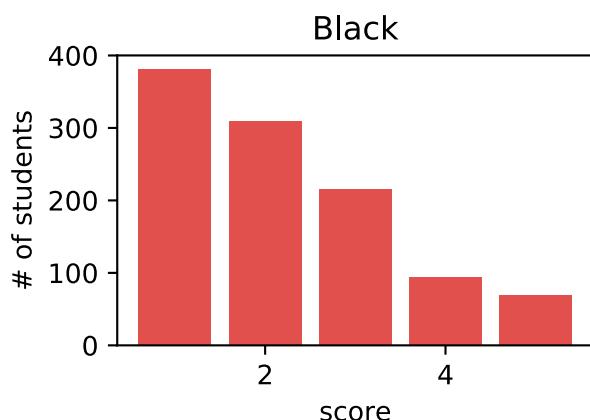
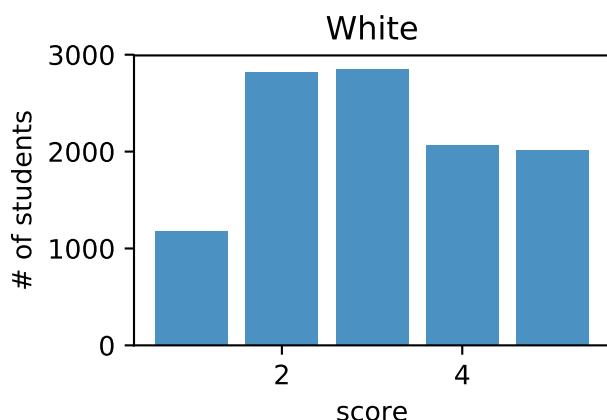
χ^2 statistic = 15098.428644481153

MICROECONOMICS



χ^2 statistic = 7373.346686605634

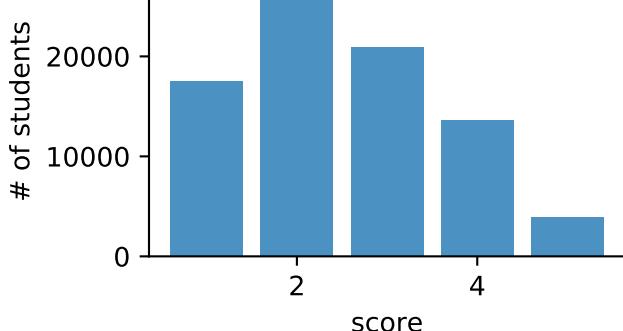
MUSIC THEORY



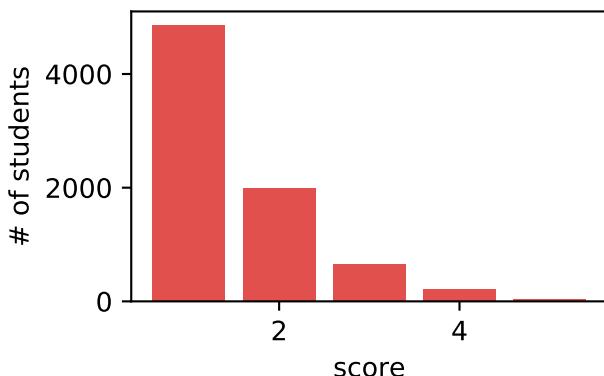
χ^2 statistic = 1624.3396991902569

PHYSICS 1

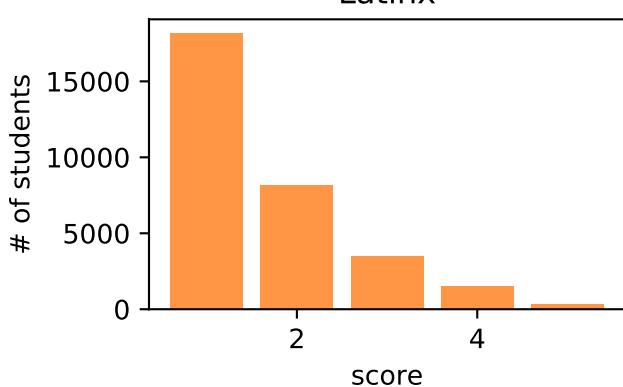
White



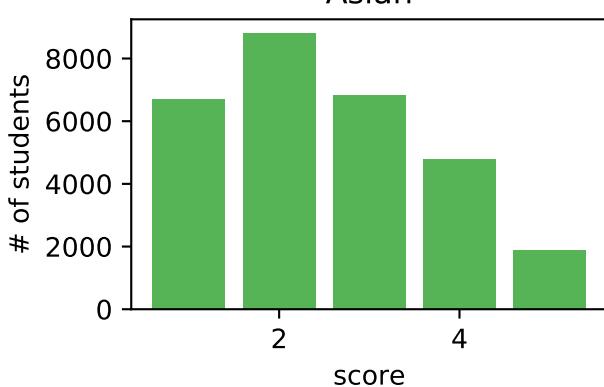
Black



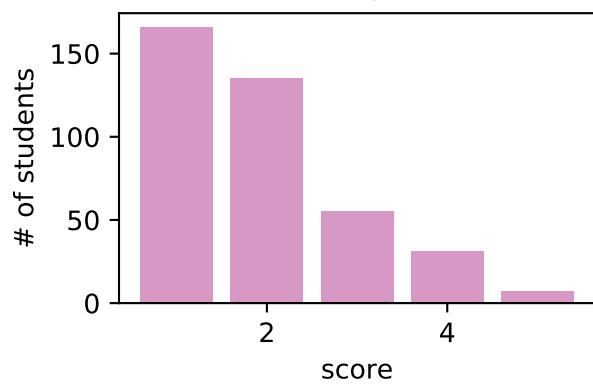
Latinx



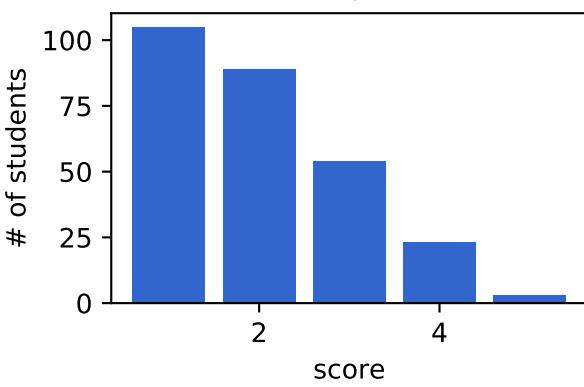
Asian



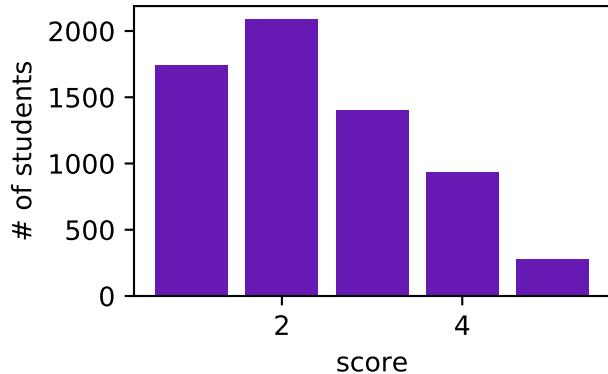
American Indian/Alaska Native



Native Hawaiian/Pacific Islander

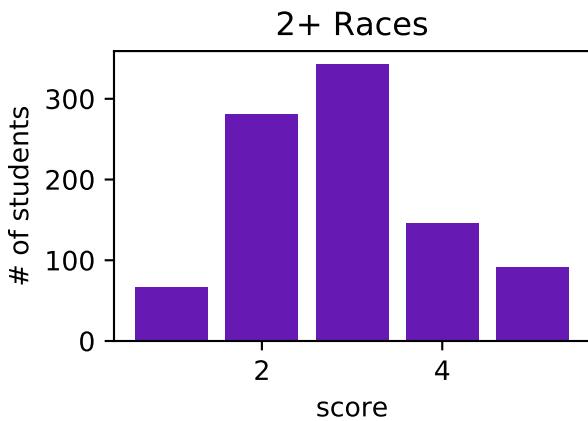
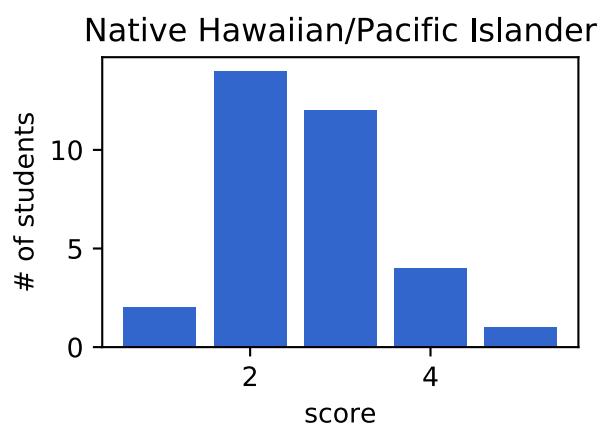
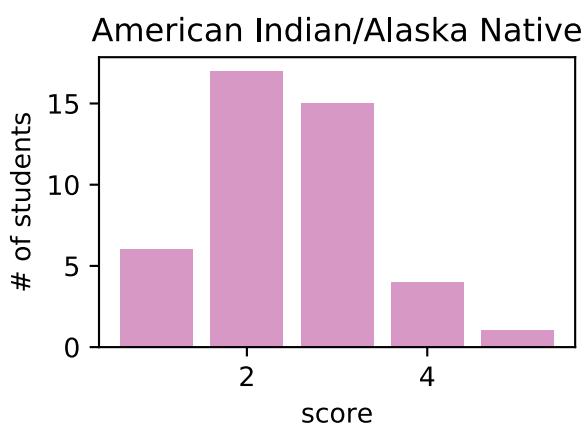
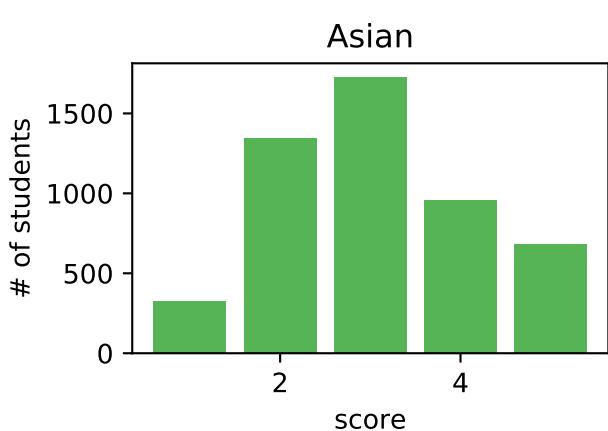
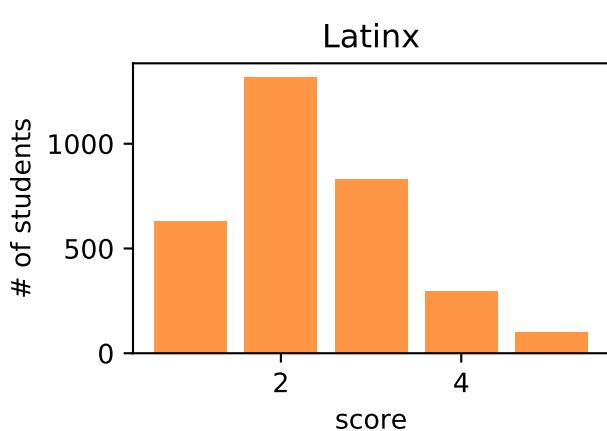
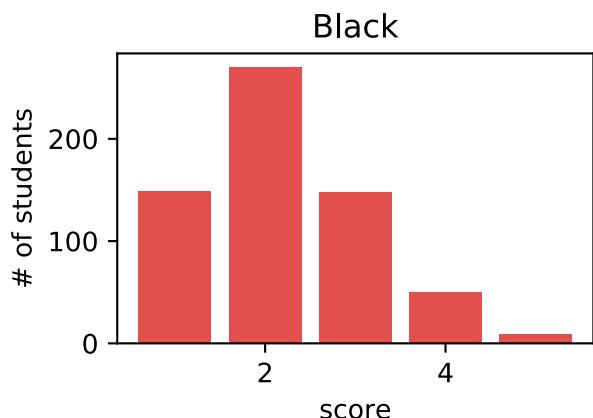
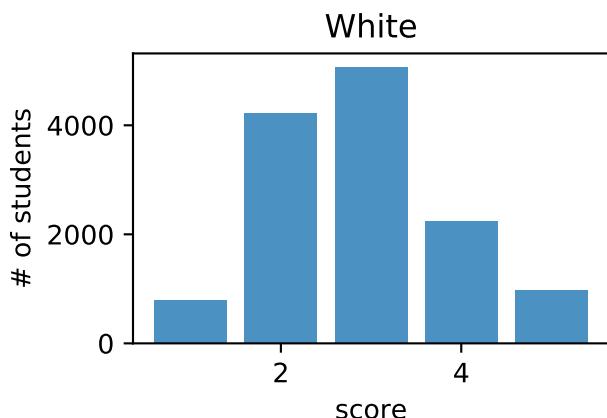


2+ Races



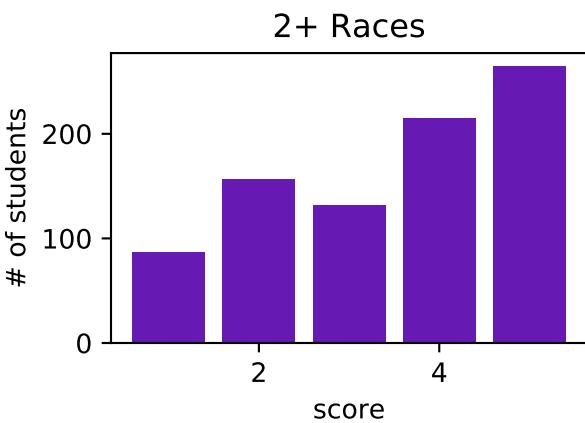
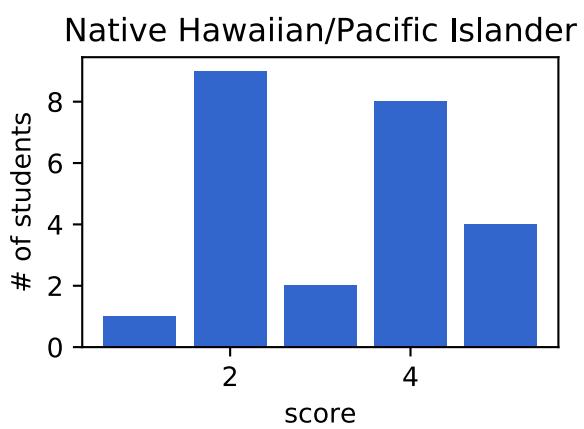
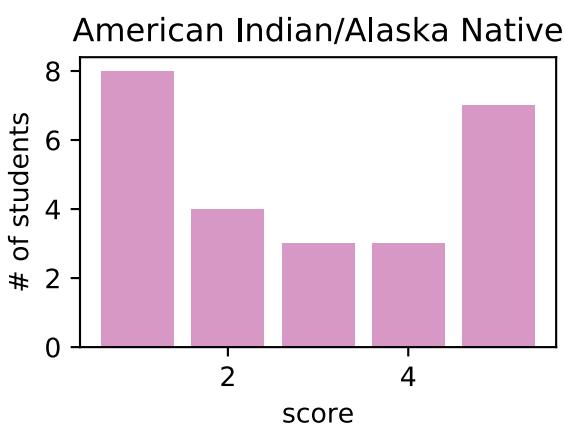
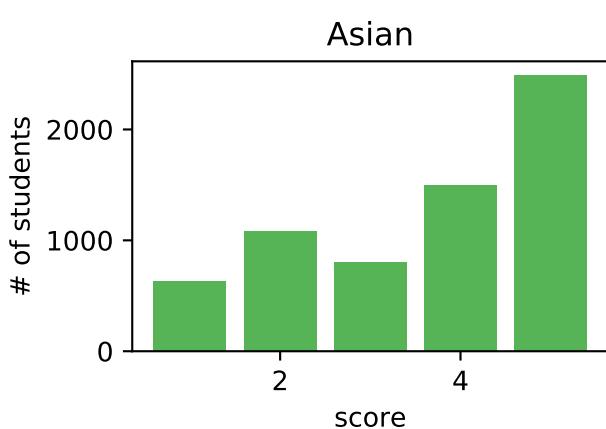
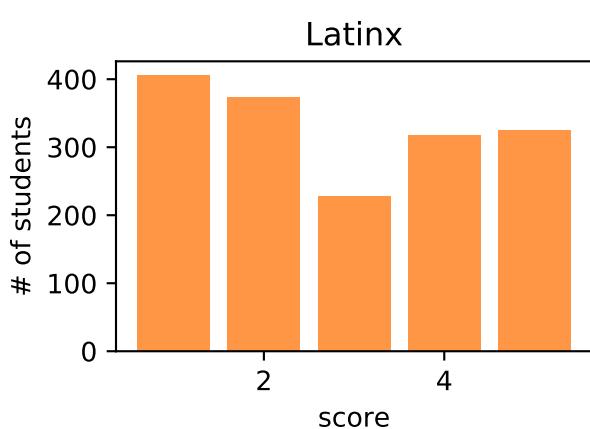
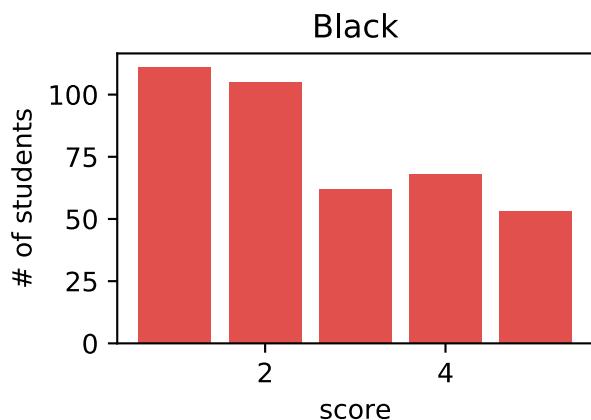
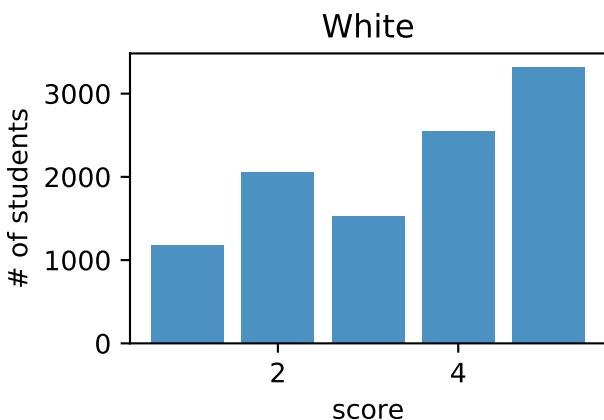
χ^2 statistic = 20997.02170547486

PHYSICS 2



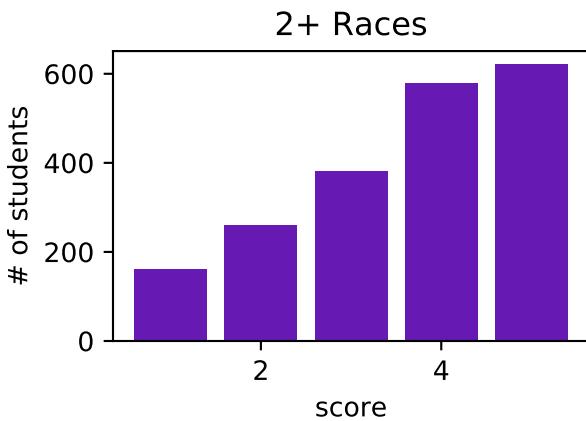
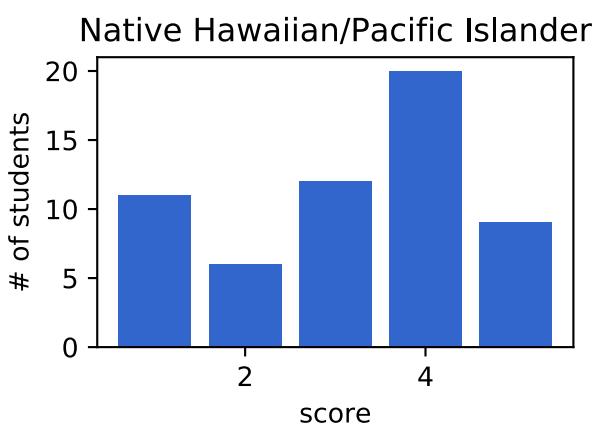
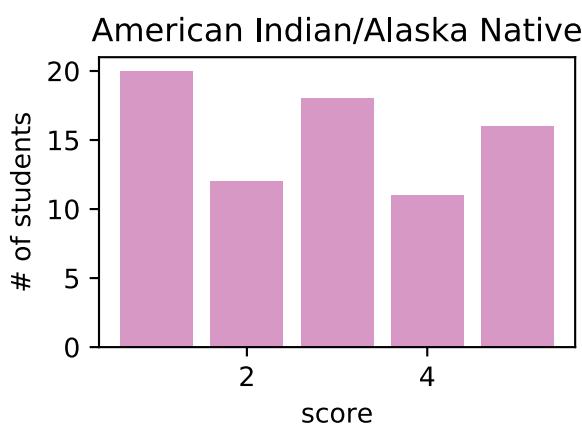
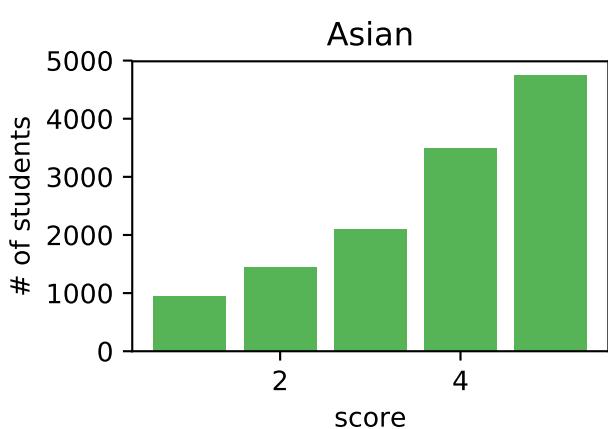
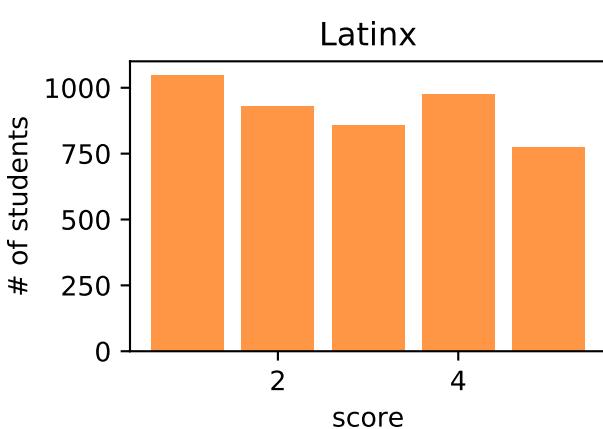
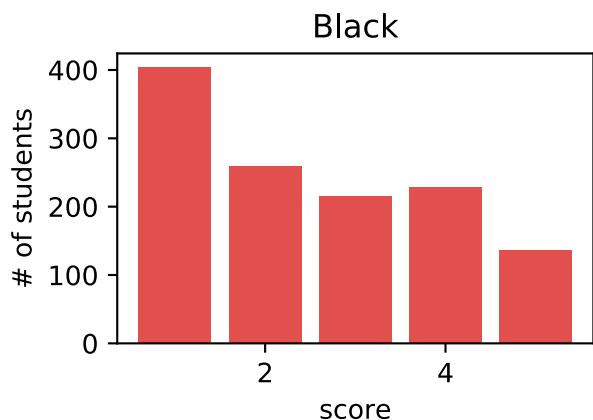
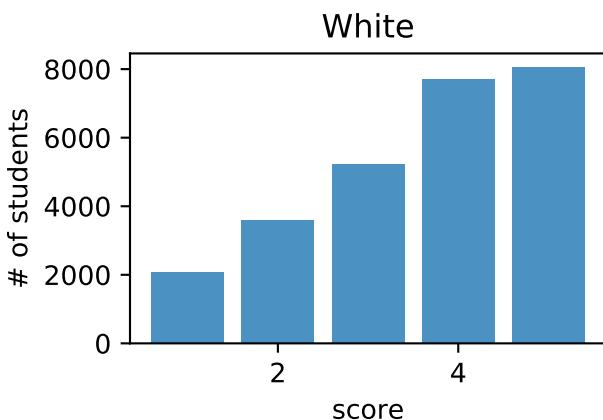
χ^2 statistic = 1539.6358885908082

PHYSICS C: ELECTRICITY & MAGNETISM



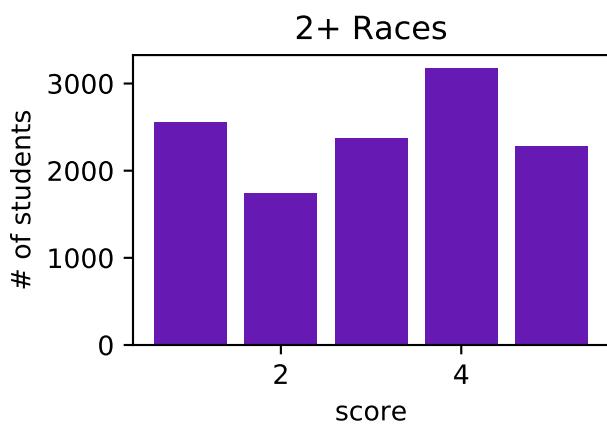
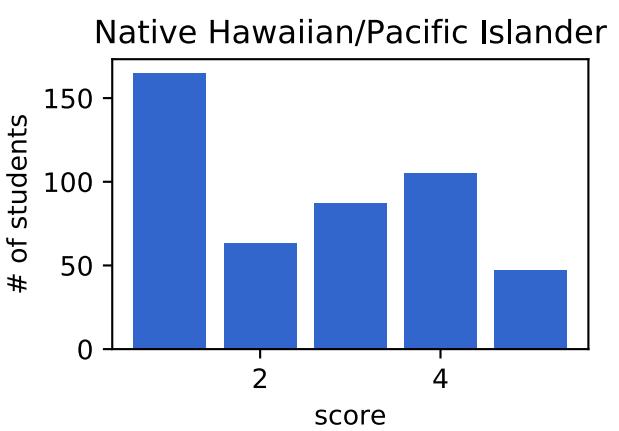
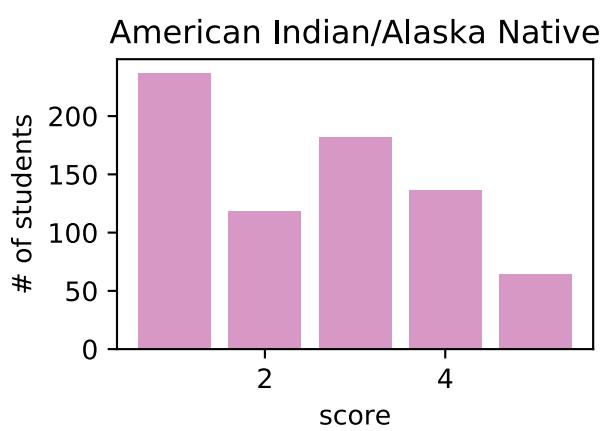
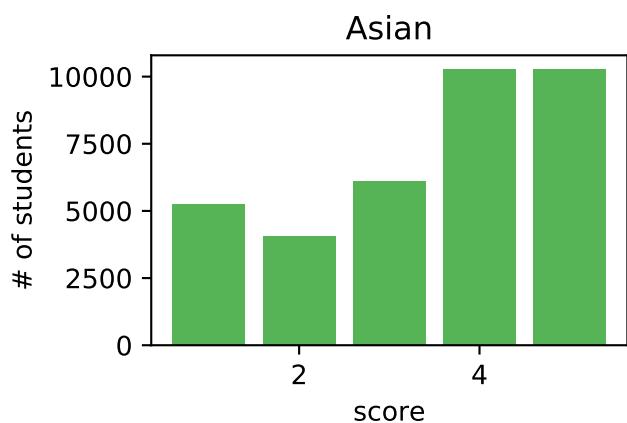
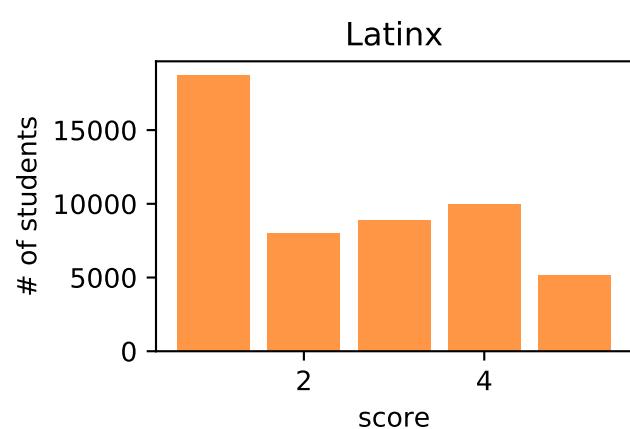
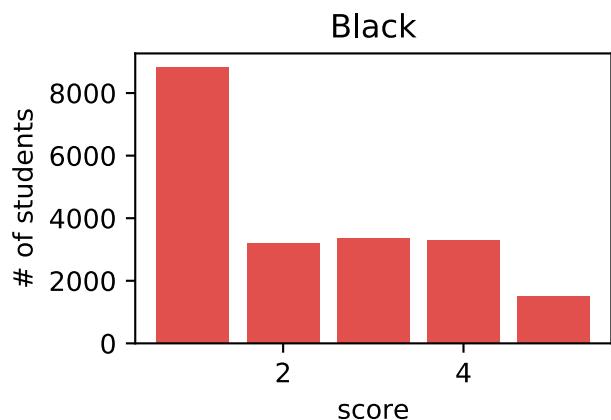
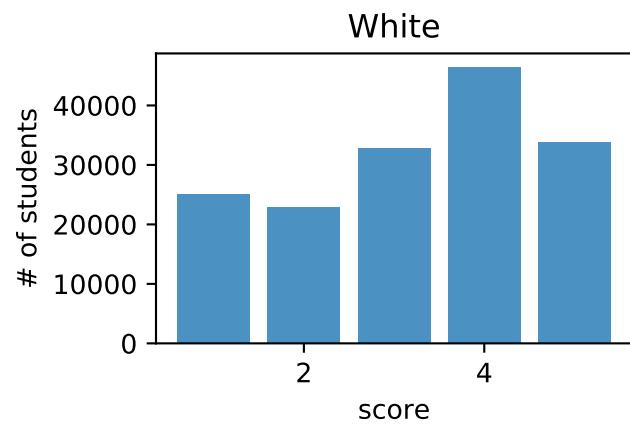
χ^2 statistic = **636.9279539471867**

PHYSICS C: MECHANICS



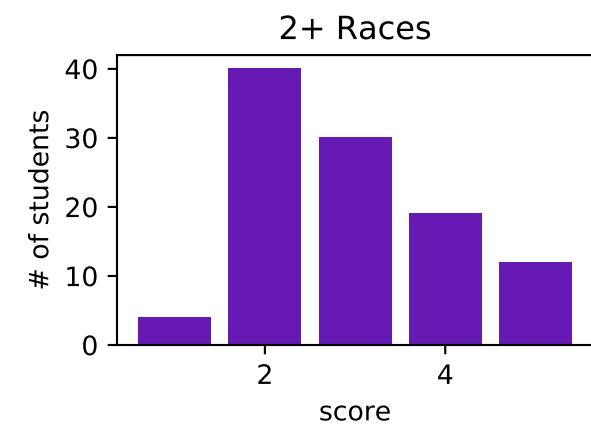
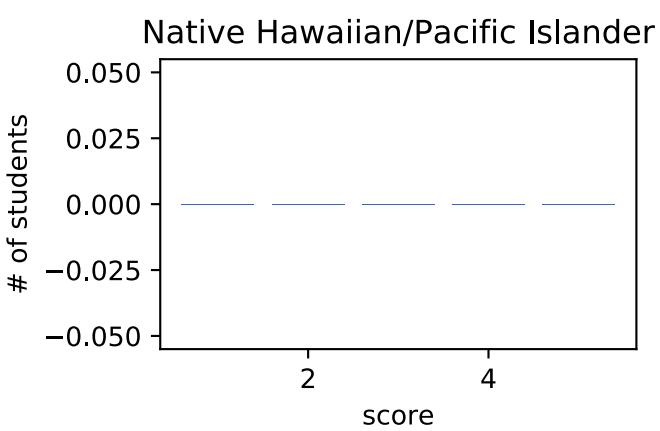
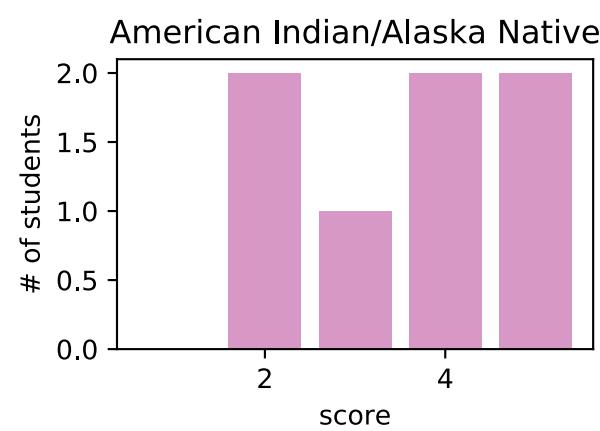
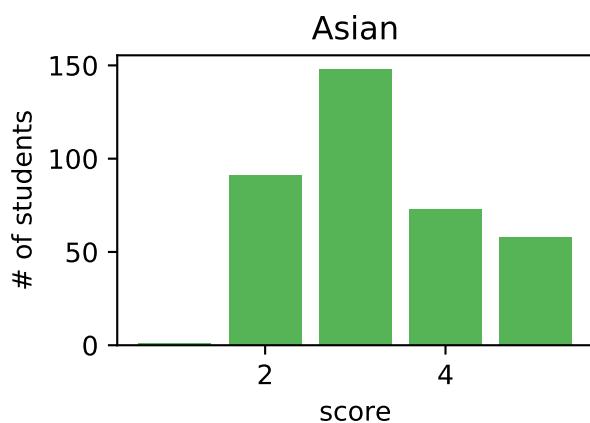
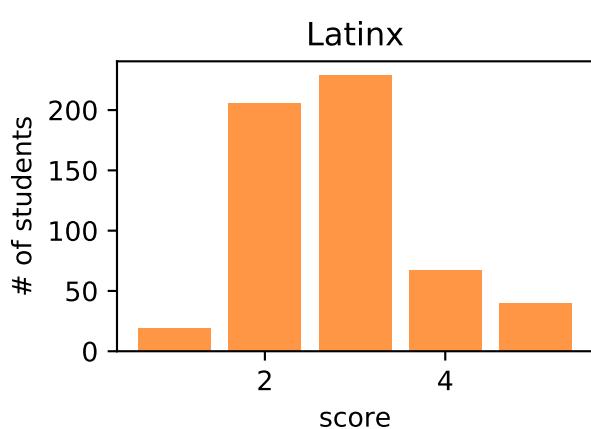
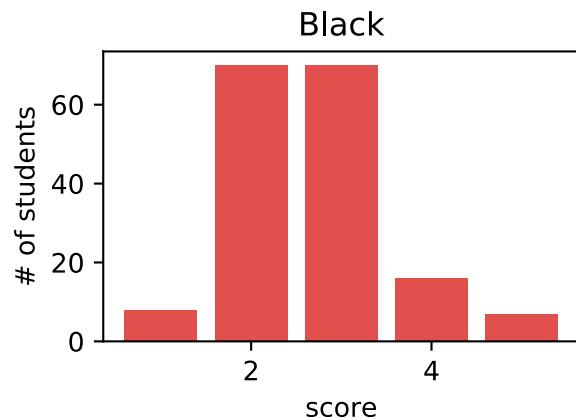
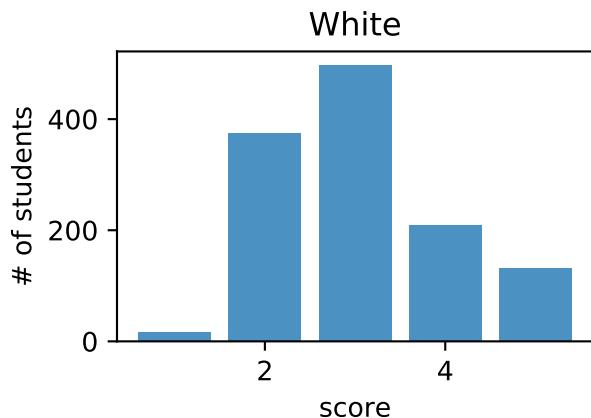
χ^2 statistic = 2706.474310916142

PSYCHOLOGY



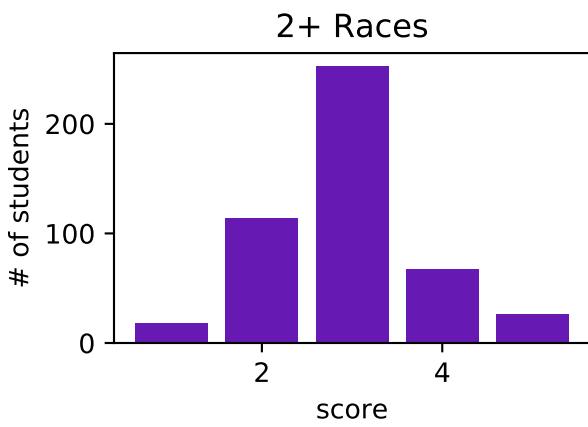
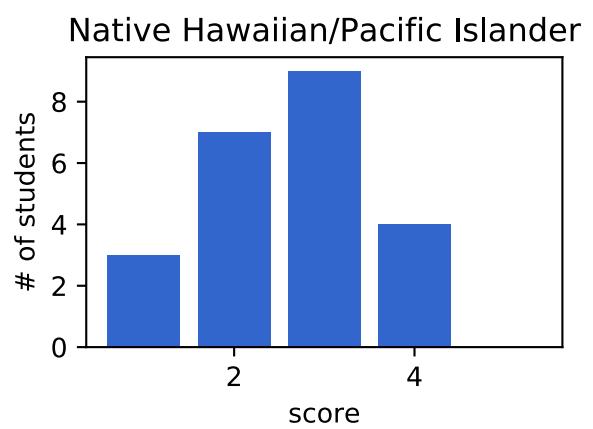
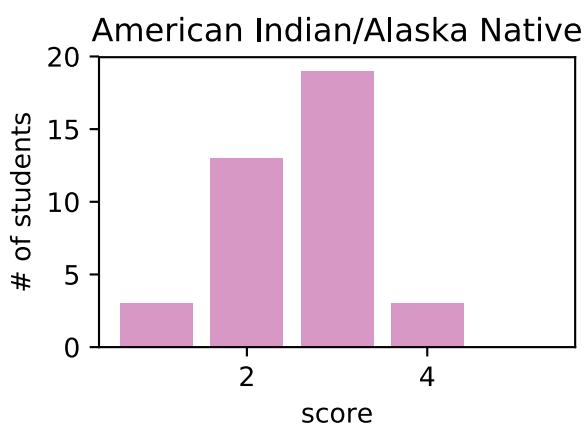
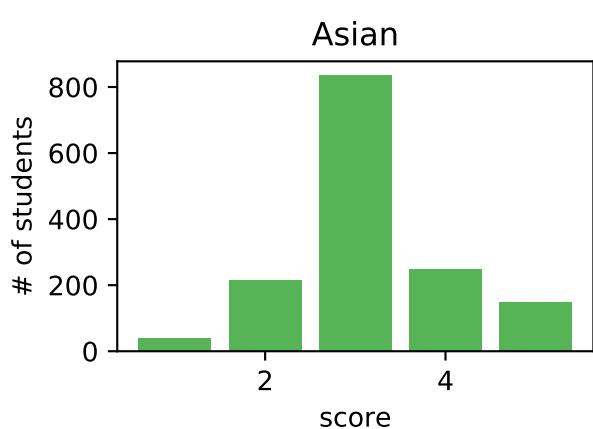
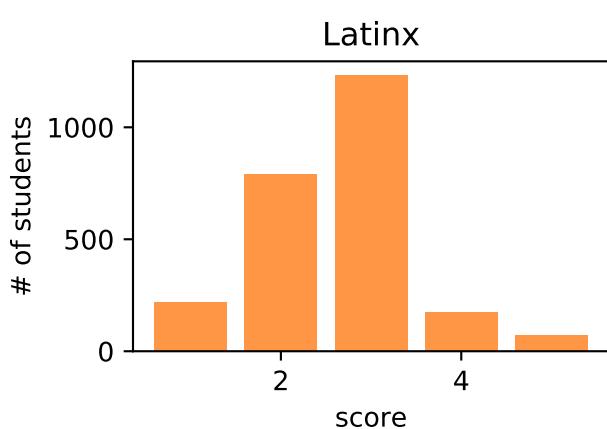
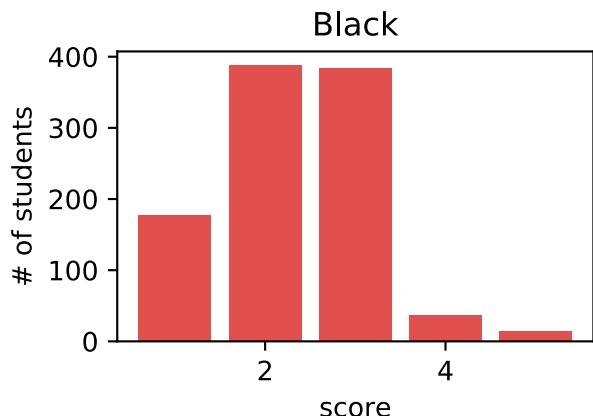
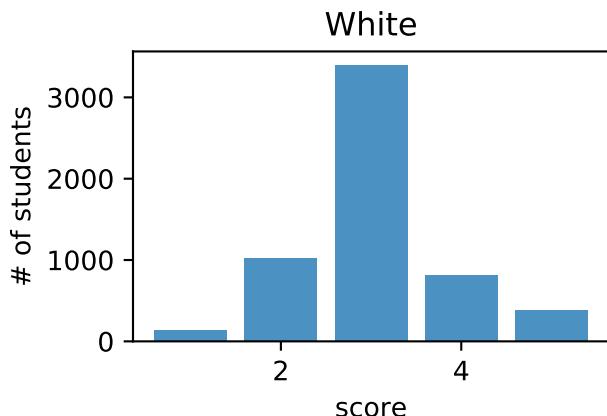
χ^2 statistic = 22093.79996698069

RESEARCH



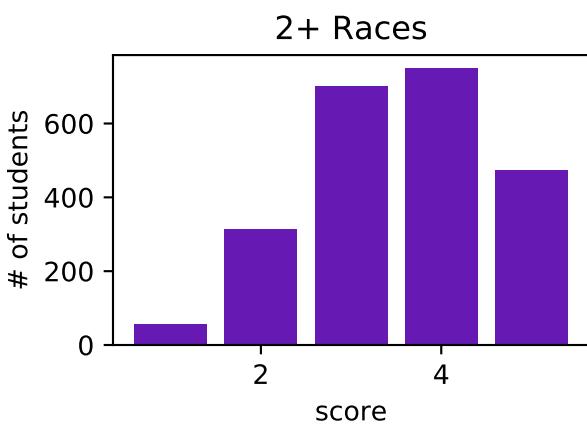
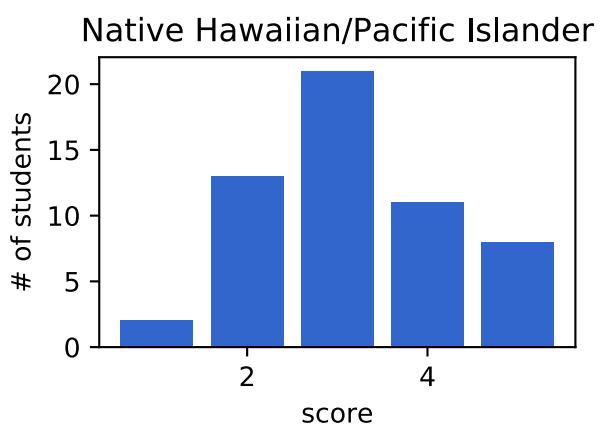
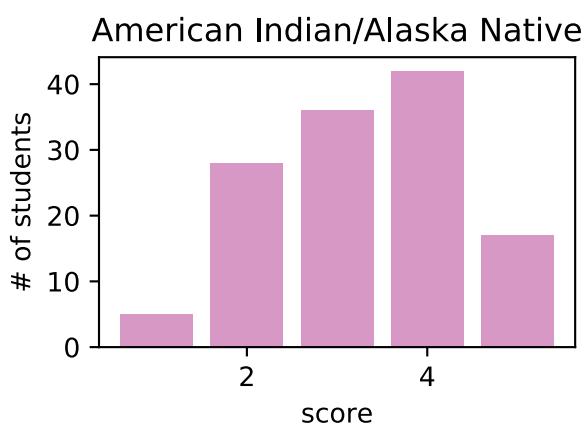
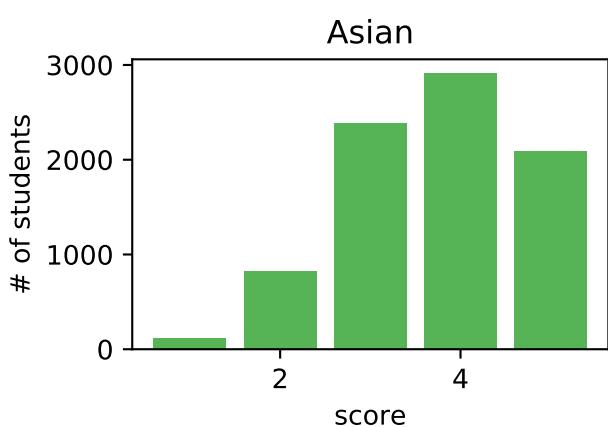
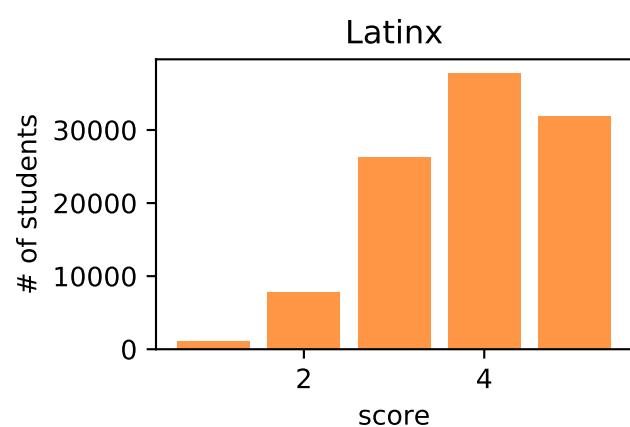
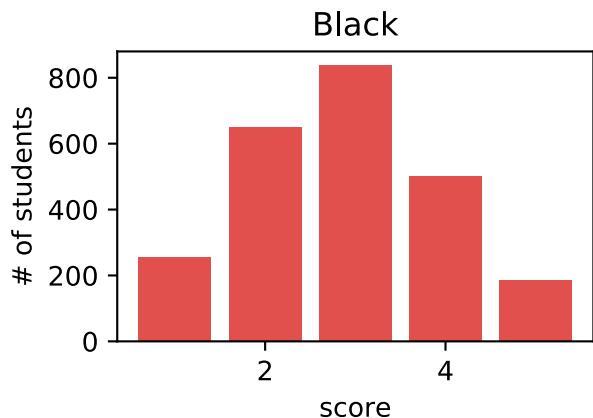
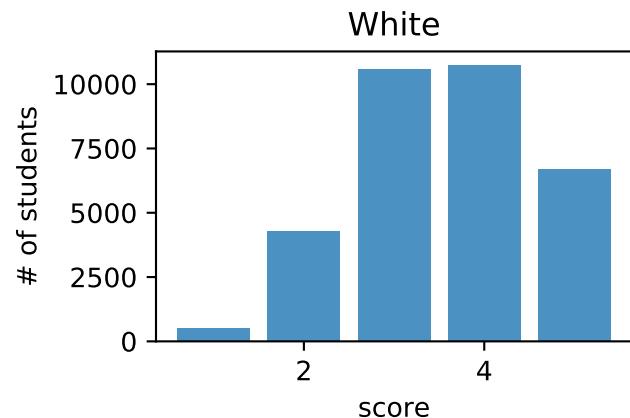
χ^2 statistic = 85.86329953450917

SEMINAR



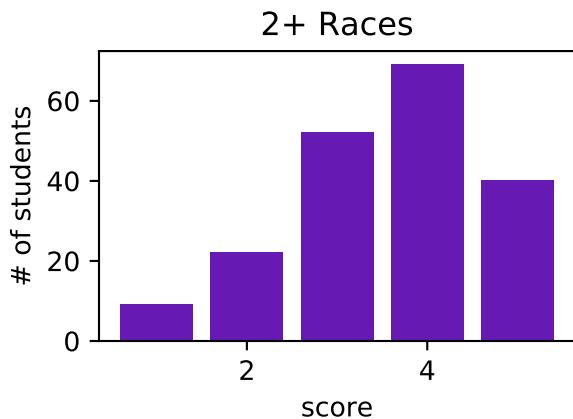
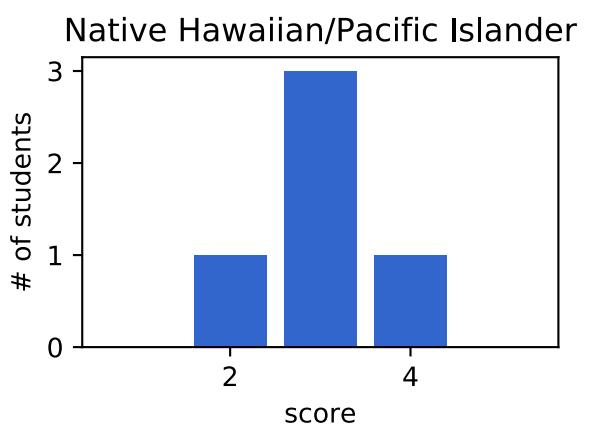
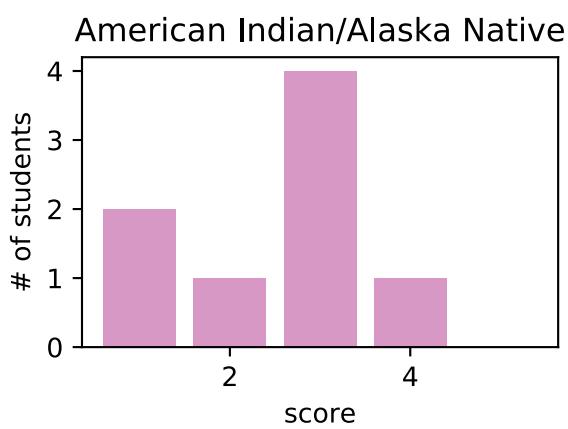
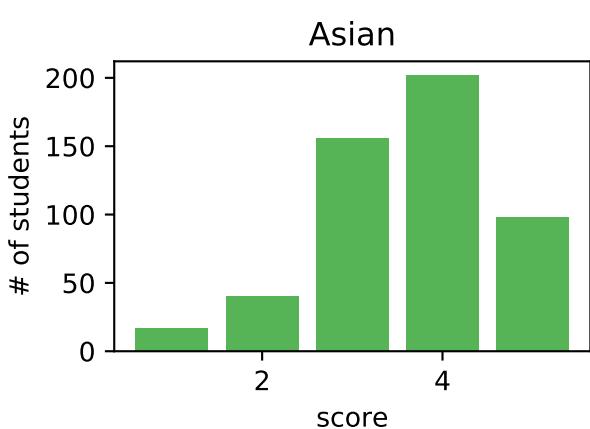
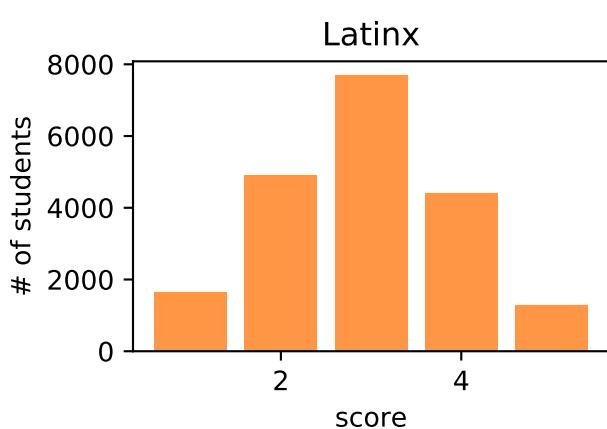
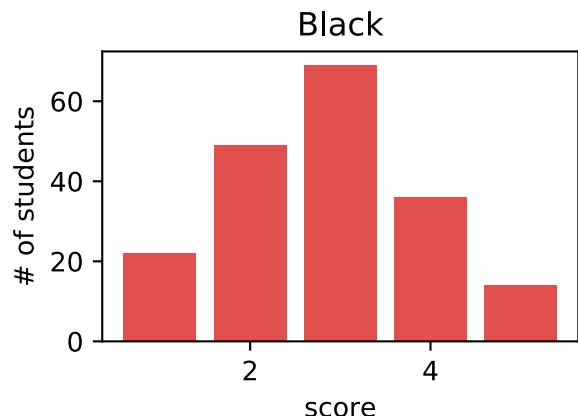
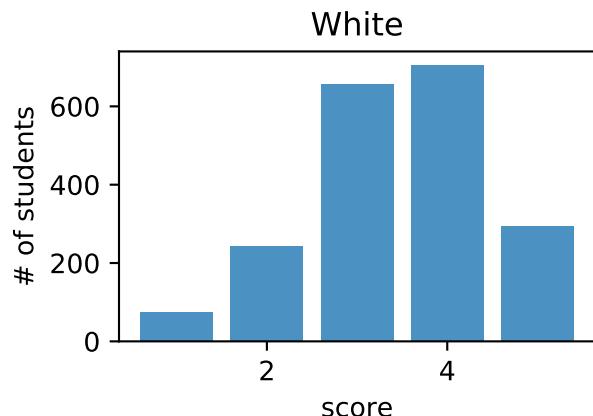
χ^2 statistic = 1125.6680499127926

SPANISH LANGUAGE



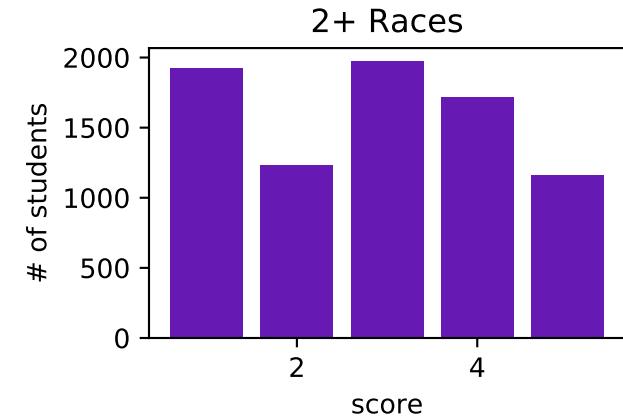
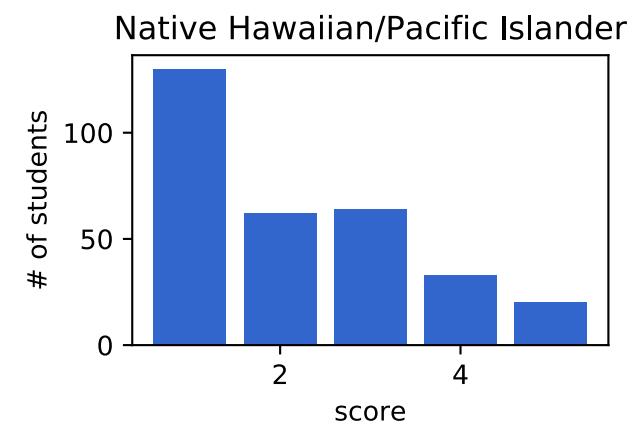
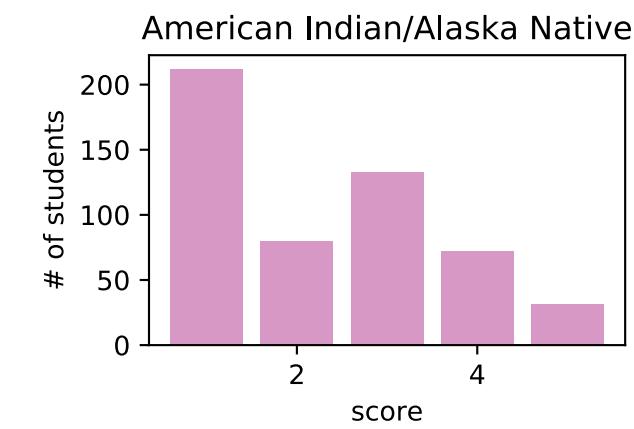
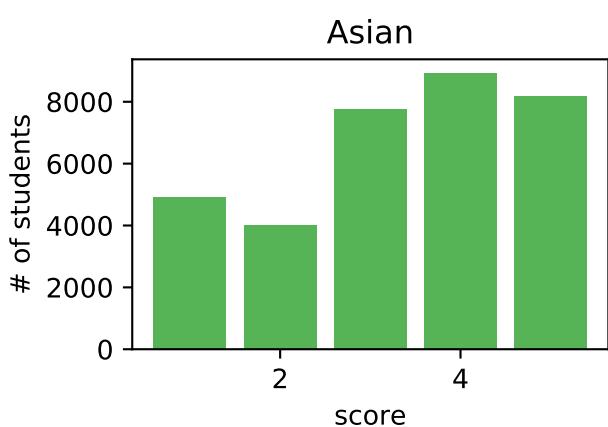
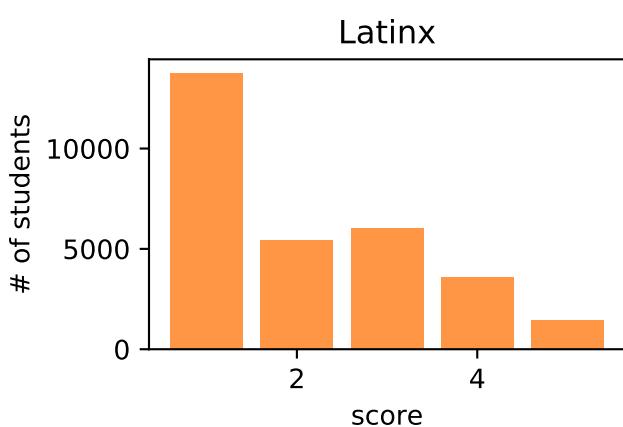
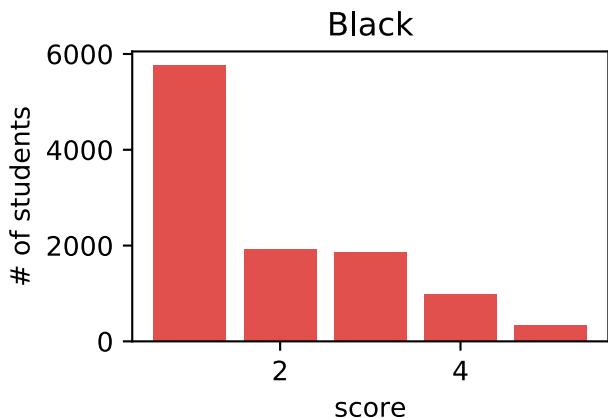
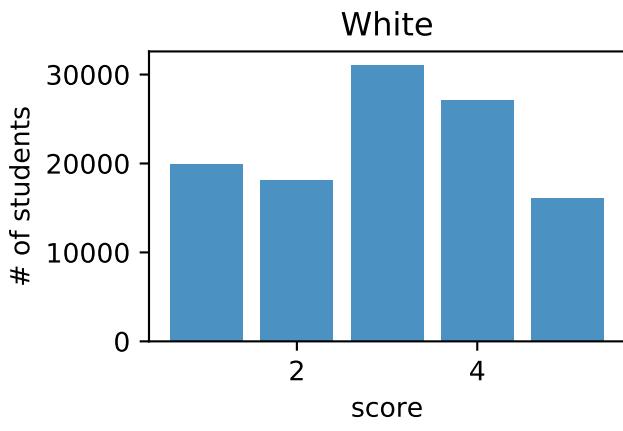
χ^2 statistic = 5480.764108914975

SPANISH LITERATURE



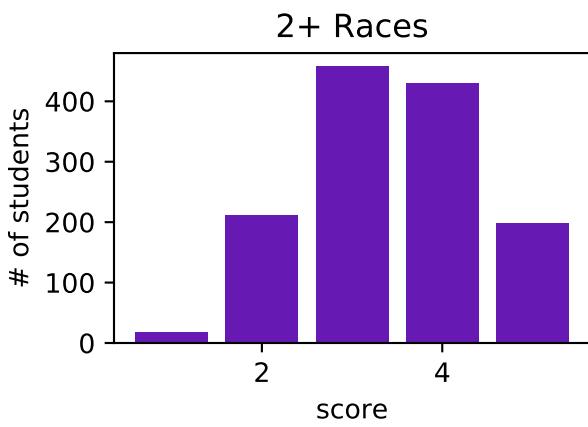
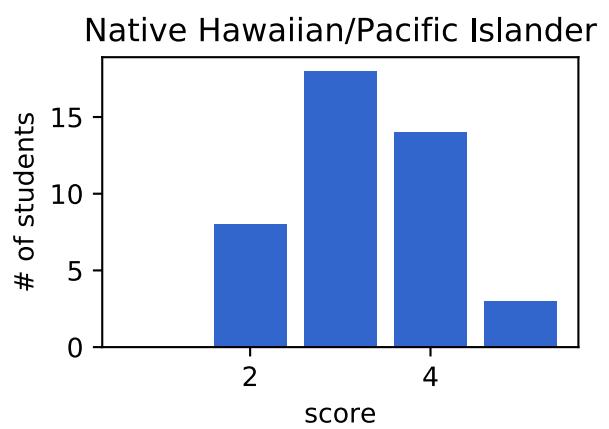
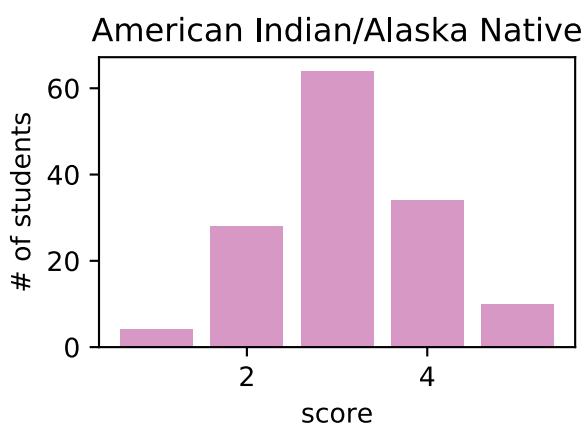
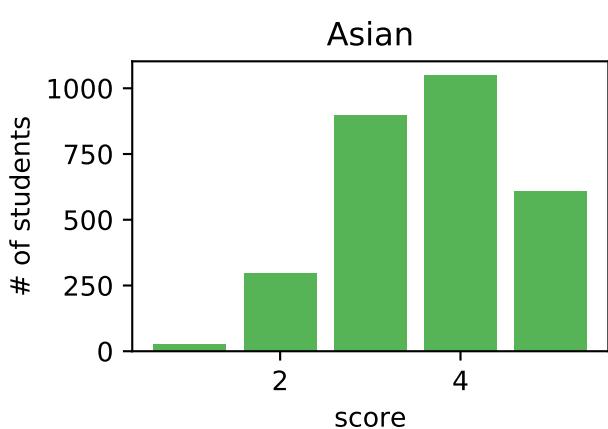
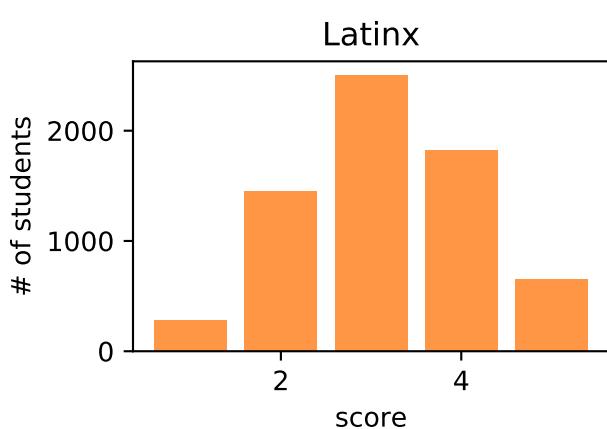
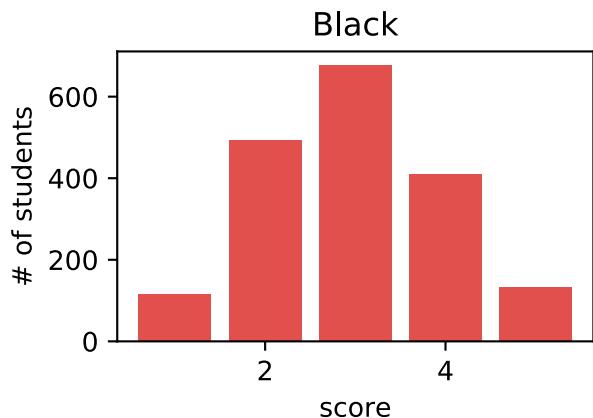
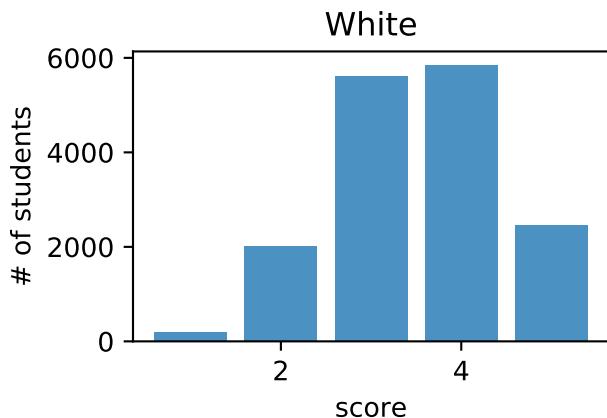
χ^2 statistic = 793.6774474014817

STATISTICS



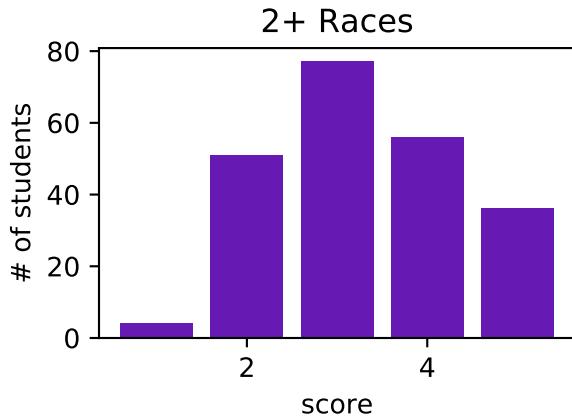
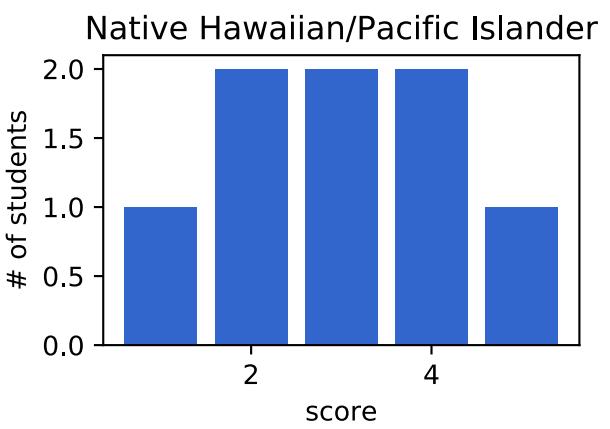
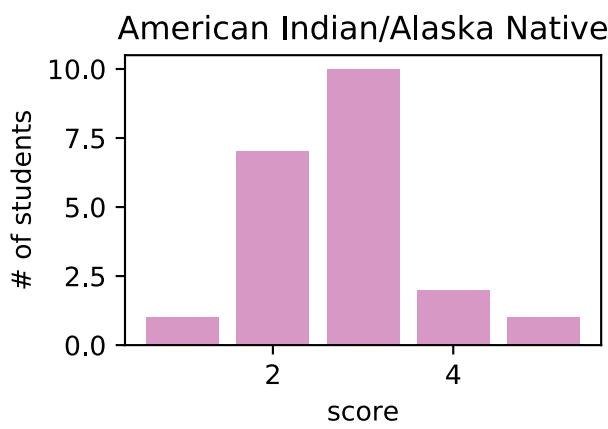
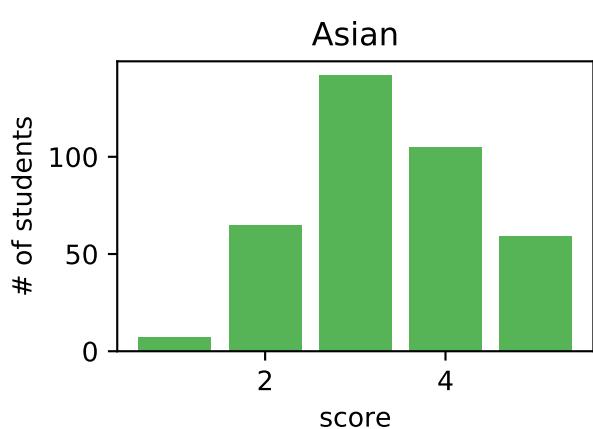
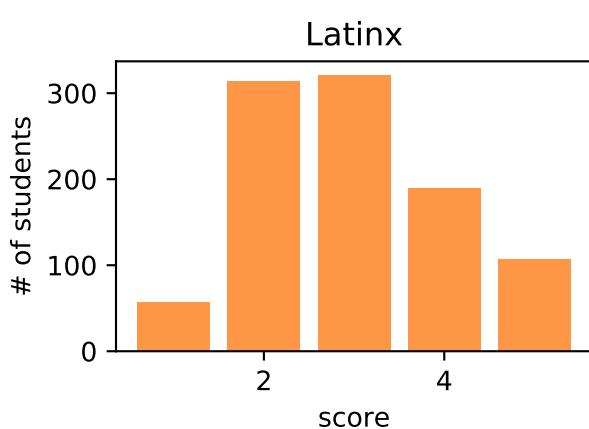
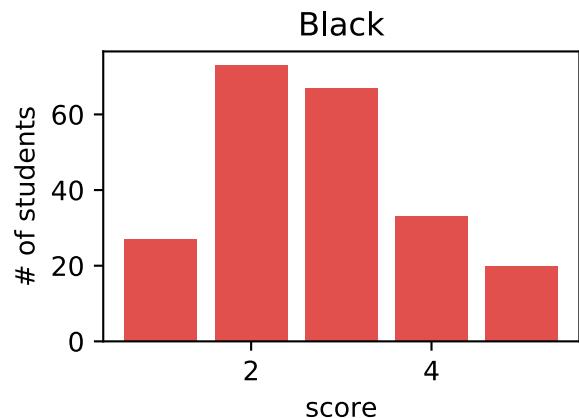
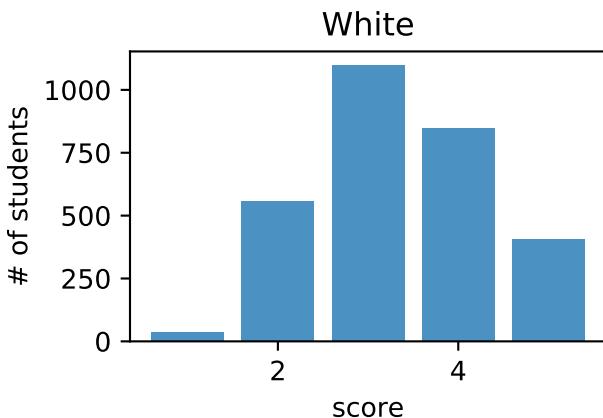
χ^2 statistic = 22532.110792247673

STUDIO ART: 2-D DESIGN



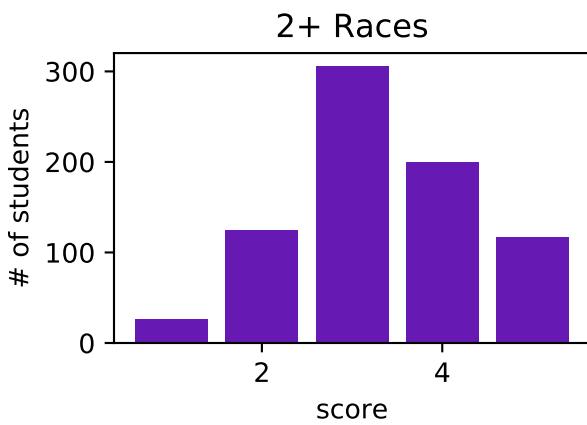
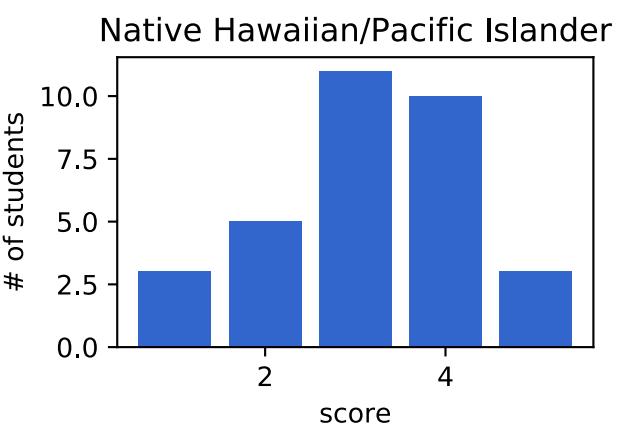
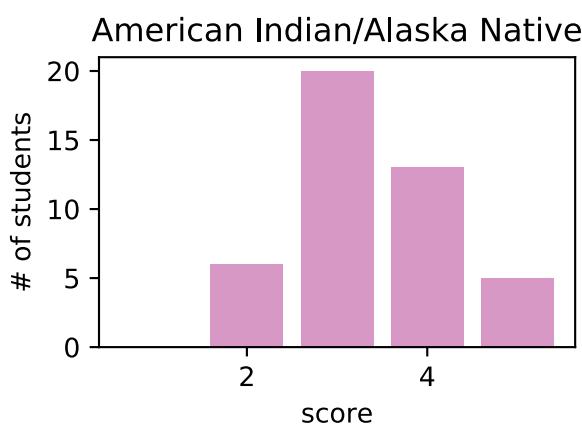
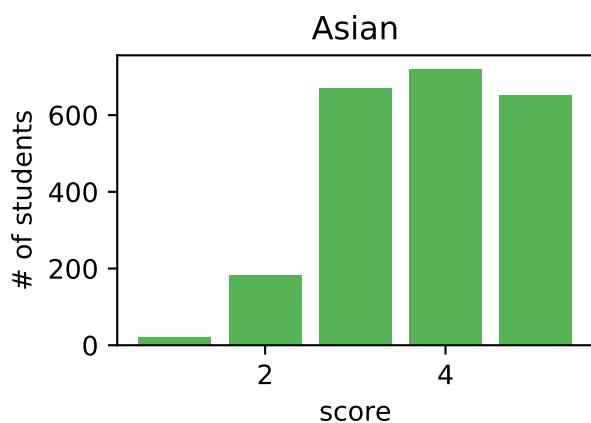
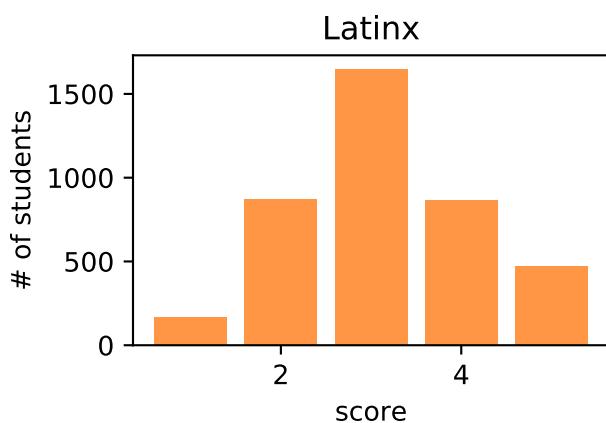
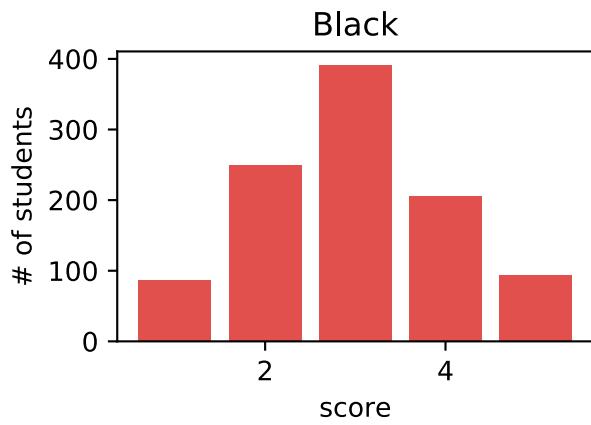
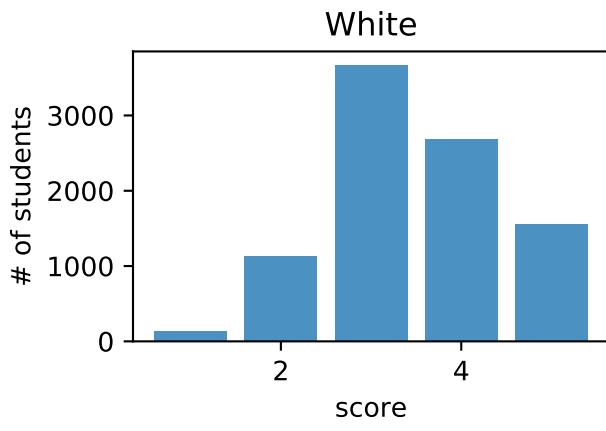
χ^2 statistic = 1318.7670068904279

STUDIO ART: 3-D DESIGN



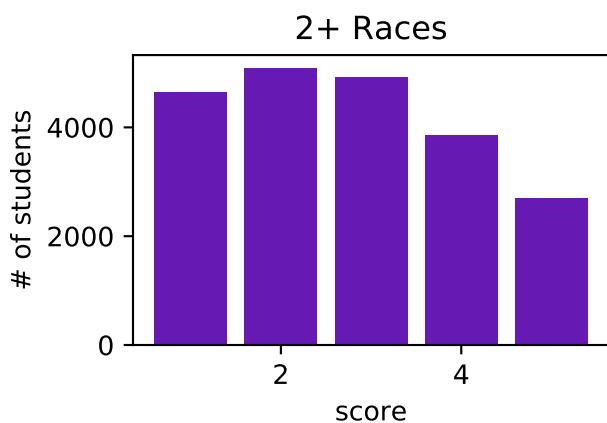
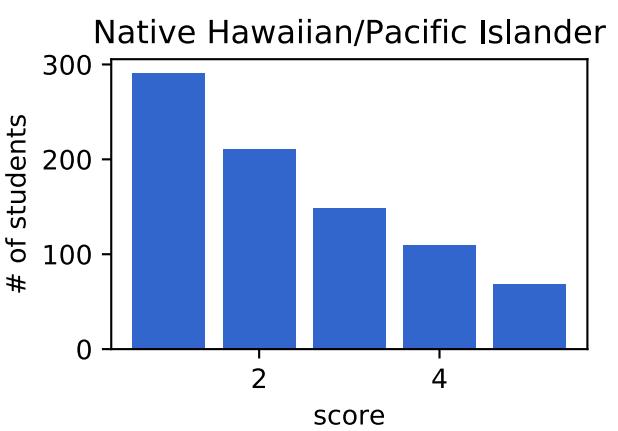
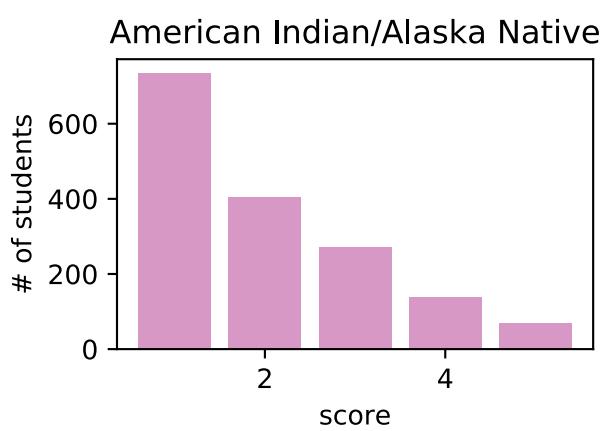
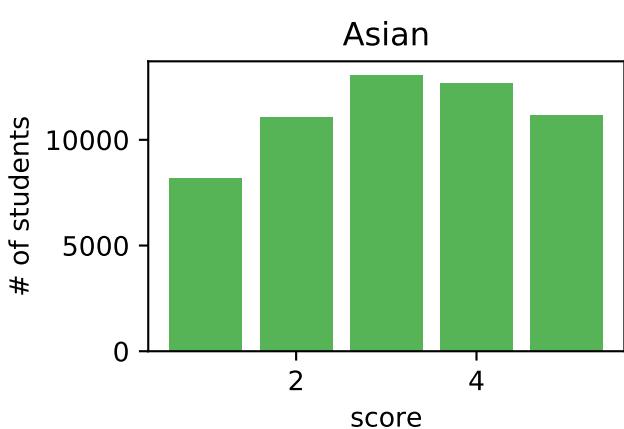
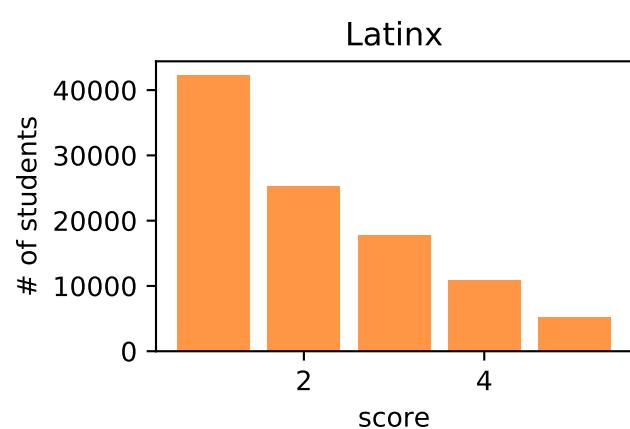
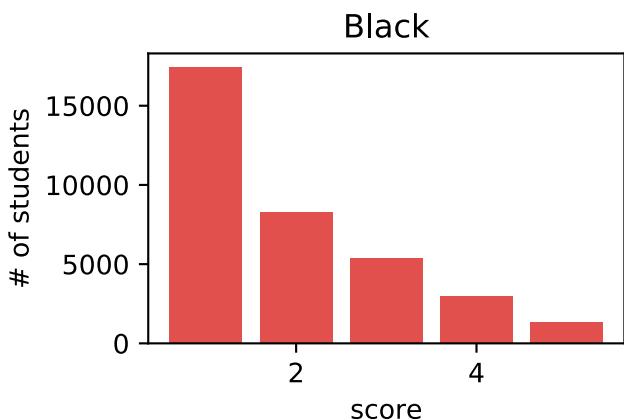
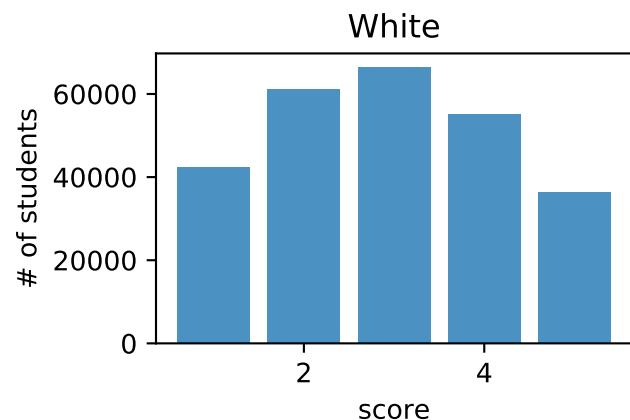
χ^2 statistic = 255.40818185711865

STUDIO ART: DRAWING



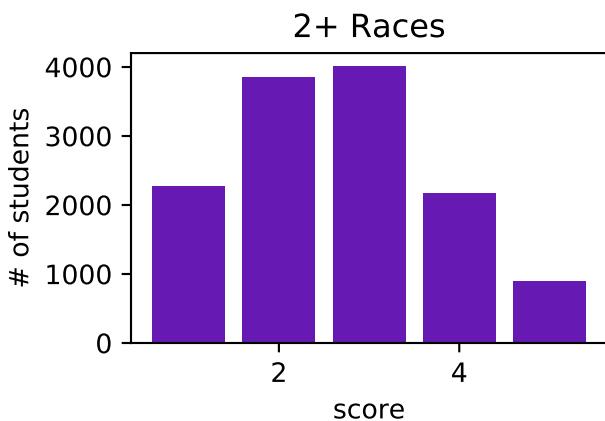
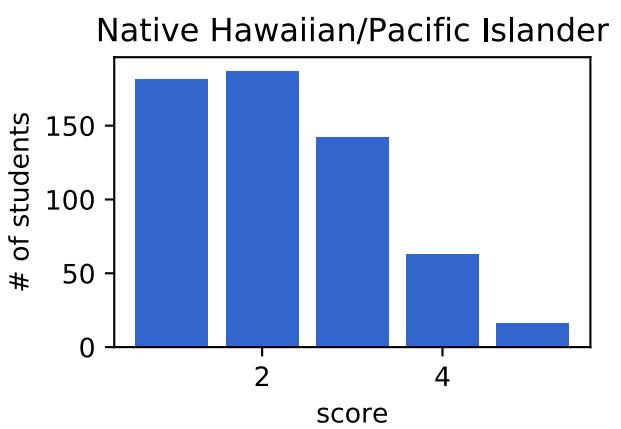
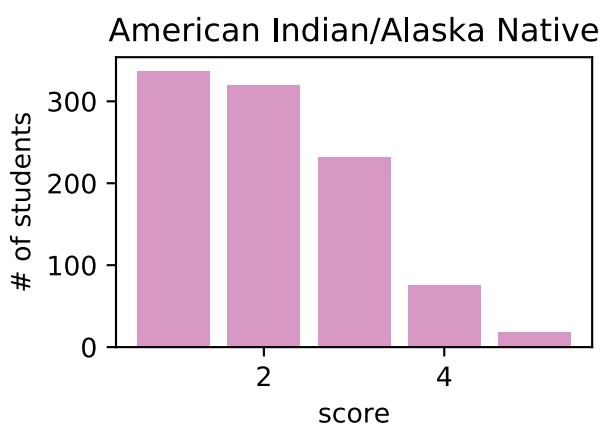
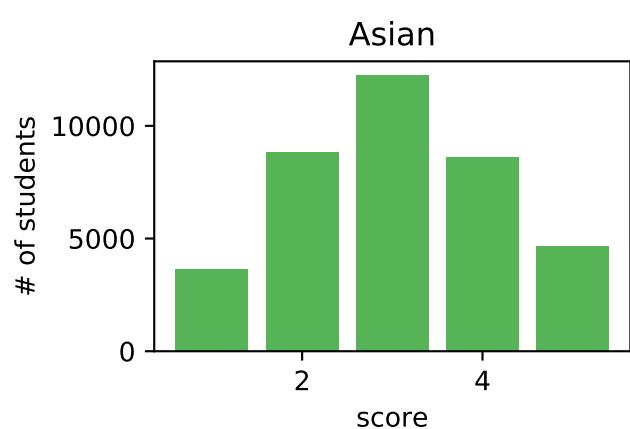
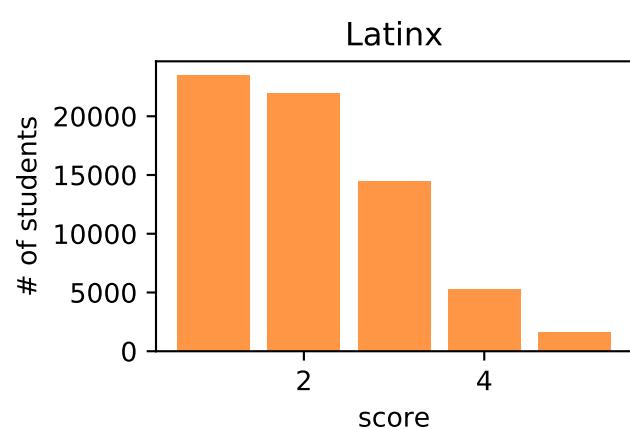
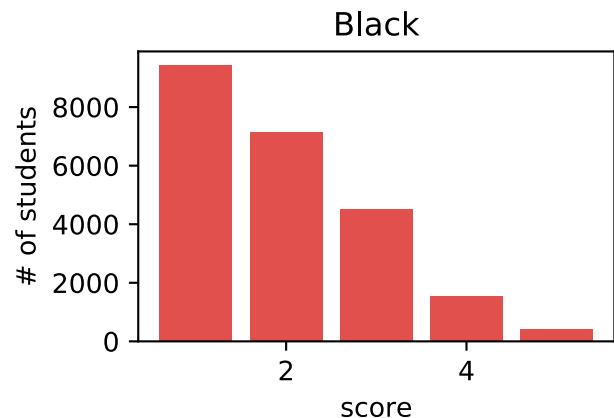
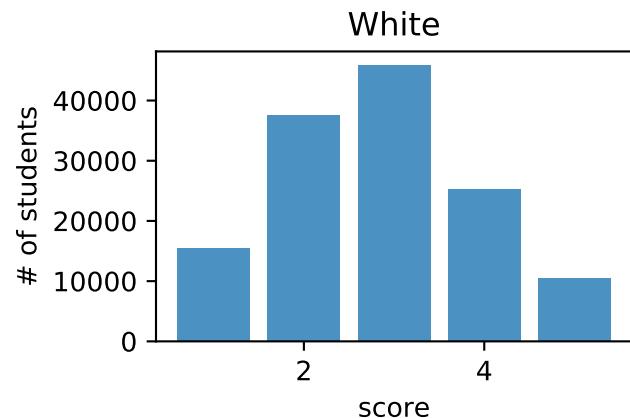
χ^2 statistic = 1013.6301781819227

U.S. HISTORY



χ^2 statistic = 51269.54917217614

WORLD HISTORY



χ^2 statistic = **34841.79695168788**