

Data Analysis of the Factors Impacting ACT scores in Nebraska Public Schools

Zach Benton-Slocum
Haley Briggs
Nathan Horrell
Michael Neman
Emily Overend
Haritha Prasad Rayakota

December 14, 2022

Abstract

The American College Testing (ACT) is a standardized test used for college admissions in the United States and covers four academic skill areas- English, mathematics, reading, and science. The Nebraska State Statute 79-760 was updated by passage of LB 930 in spring of 2016 to require that all public school 11th grade students take the ACT. This report primarily examines how absenteeism, related to COVID, per pupil expenditure, and student demographics affect ACT scores in Nebraska public high schools. The data was collected from the Nebraska education website. It was cleaned and preprocessed using R. For the data product, a dashboard was created using Shiny R with the factors as separate tabs. Based on the analysis, attendance did not significantly impact ACT scores while per-pupil expenditure and demographics like gender and race did. For all the academic years, attendance in all the districts did not change while ACT scores changed. Pearson correlations revealed that the linear relationship between Per-Pupil Expenditure (PPE) and ACT for the 2018-2019 school year was insignificant, however a significant, weak, negative linear relationship between PPE and ACT for the 2020-2021 school year was found. The gender demographics showed a significant drop in ACT scores for the 2020-2021 academic year. For the 2020-2021 academic year, all ethnicity groups also showed a significant decrease in ACT scores except Native Hawaiian race which showed an increase in ACT scores.

1. Introduction

In the spring of 2016, the State of Nebraska passed Legislative Bill 930 which updated State Statute 79-760.03 requiring all 11th graders in public schools to take a national assessment for the purpose of national comparison[2]. The statute also requires that the assessment to be given is also a standard college admission test[2]. The state board of education recommended the assessment be the American College Testing (ACT), and starting in the 2017-2018 school year the ACT was given to every 11th grade public school student, except those identified as needing alternate assessment. The ACT was developed for 11th and 12th grade-students and provides colleges with one common data point that can be used to compare all applicants[1]. There are four multiple choice exam sections in English, Mathematics, Reading, and Science[1]. Each section produces a raw score and is scored independently from 1 to 36[1]. The sections are then averaged to produce the total exam composite score that also ranges from 1 to 36[1].

As of 2020, Nebraska has a total of 1036 schools that are broken up into 244 districts. These schools are populated by 328,724 students and 23,751 teachers. The research questions in this report focused on ACT performance and how factors like attendance ,student expenditures and demographics would affect said performance. Due to this, the focus of this report is on secondary schools in Nebraska as it would still be

many years before elementary students would take the ACT. Once all of the elementary schools and any schools that were deemed as statistically irrelevant were removed the data set shrank to 239 schools.

The average composite scores for Nebraska public school 11th graders are shown in Figure 1. The first and most apparent feature of this plot is that ACT scores have been decreasing since the beginning of ACT administration (the ACT was not administered in the 2019-2020 school year due to the Covid-19 pandemic, by special order of the Governor)[3]. This trend is not unique to Nebraska; ACT scores have been on the decline nationally for the last five years[4]. This report will explore and analyze some of the factors that may impact ACT scores in Nebraska public schools in an effort to determine areas of further study.

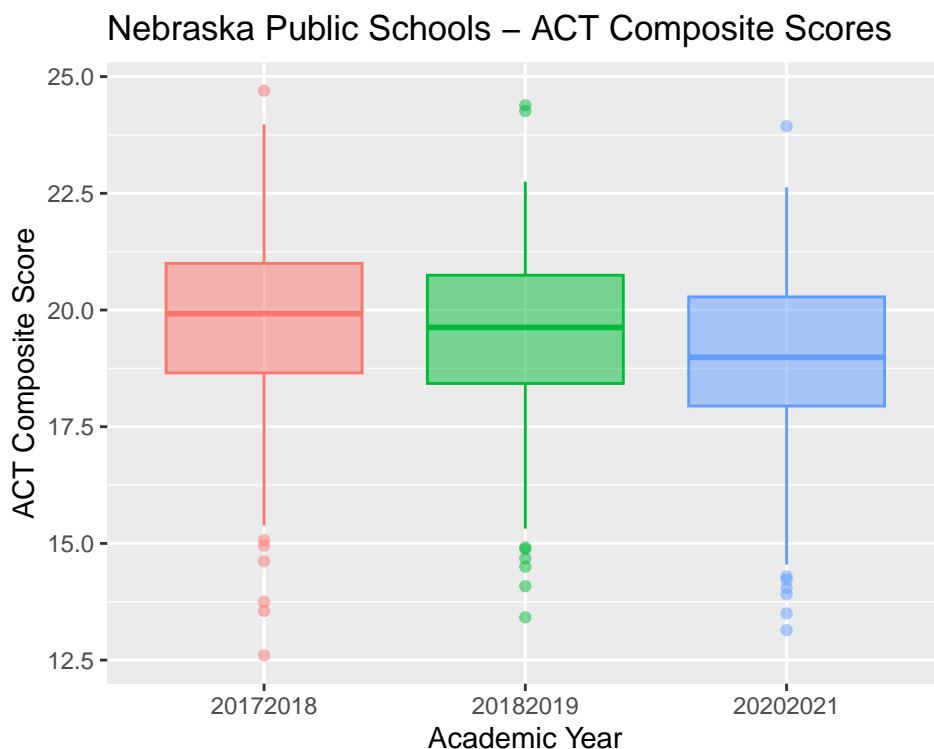


Figure 1: ACT composite scores for the 2017-18, 2018-19, and 2020-21 school years. The average ACT composite score across Nebraska public schools has decreased the last four years.

Currently, the Nebraska Department of Education does not have a data analyst[7]. The NDE has requested this report address a specific research question: does absenteeism, related to COVID, affect ACT performance? During the pandemic, the NDE was well aware that once schools started operating remotely there would be varying approaches to recording attendance, participation, and engagement, and encouraged the continuation of the attendance-taking processes implemented by the district through the rest of the school year[9].

With the observed downward trend of ACT scores, there may be value in determining what, if any, correlation exists between school funding and ACT outcomes. Across the many schools and districts in the state, does an increase in funding correlate with increased ACT scores? This report explores these questions at various levels of abstraction, from the state, district, and school.

This report seeks to utilize NDE datasets to explore what factors may impact ACT outcomes in Nebraska public schools. First, attendance rate is analyzed, with special emphasis on how any variation in attendance during Covid correlates with ACT outcomes per an NDE request. Next, per pupil expenditure is explored to determine what, if any, correlation exists between funding and ACT scores. Finally, demographic factors are examined with respect to ACT performance. Additionally, a data dashboard is introduced so the reader can explore these data independently.

1.1 Document Reproducibility

Every effort has been made to ensure the highest standard for reproducibility of this analysis. The R markdown (Rmd) document pulls raw data directly from the Nebraska Department of Education website so that no interim downloads or manipulations are needed. The entirety of the analysis occurs from a single data source, and all of the code is run from a single file. The code is well-documented and easily read. There are neither random numbers nor machine learning tools utilized in this analysis; therefore, each calculation is functional, with a one-to-one and on-to mapping from independent to dependent variable. As explained, document reproducibility has been a high priority for this analysis. This report was written in RStudio using an *.rmd filetype, and was converted to PDF format with the knitr package. Any RStudio user could run this code and find identical results.

2. About the data

The data used for this report is publicly provided by the Nebraska Department of Education (NDE)[5]. The NDE collects many data and publishes them across several datasets for each school year. The datasets of interest for this report include ACT Composite Scores, ACT Assessments Data All Subjects, Attendance Rate, and Per Pupil Expenditure. ACT Composite Scores and ACT Assessments Data All Subjects are available for the 2017-2018, 2018-2019, and 2020-2021 school years. Statewide assessments for ACT were not administered in the 2019-2020 school year due to covid-19. Attendance Rate dataset includes data from each preceding school year, so the 2020-2021 school year dataset is used. The Per Pupil Expenditure datasets are available for the 2018-2019, 2019-2020, and 2020-2021 school years.

ACT Composite Scores All three of the available ACT Composite Scores datasets are used for this analysis. Each of the data sets have nine common columns, six of which are identification data, one is the school year, one column is the average composite score, and the last column is the number of students with composite scores. Each of the datasets have between 517-518 observations.

ACT Assessments Data All Subjects The dataset consists of demographic information like gender, race, etc. Datasets for the 2016-2017, 2017-2018, 2018-2019 and 2020-2021 school years were considered. Each of these datasets consisted of 15 columns. Columns included: school identification, demographics, and test scores categorized by subject and student testing rate. Demographics information consisted of six different ethnic groups. School identification columns provided information about three different levels: school, district and state level. The student testing rate columns provided information on average ACT test scores, percentage tested and not tested, ACT benchmark percent and developing percent. There were approximately 26,470 observations in each dataset.

Attendance Rate The Attendance Rate dataset has eight columns and 17,217 rows and includes data from the 2006-2007 through 2020-2021 school years. There are observations that contain aggregate data for the State of Nebraska, the Learning Community of Douglas and Sarpy Counties, and Northern Tier, and each school district. There is also school-specific attendance rate data for each school. The first column indicates the aggregation level, the second column specifies the school year of the observation, there are four columns dedicated to identification data, and one column gives the attendance rate as a proportion from 0 - 1. During the 2019-2020 school year, the average attendance rate was reflective of the ratio of the total number of days students are in school compared to the number of days they could have been in school when the schools were physically open[7]. Therefore, the average attendance rates per school does not reflect on the months of April and May.

Per Pupil Expenditure All three of the available Per Pupil Expenditure (PPE) dataset are used for this analysis. The first column indicates the aggregation level which can be at the district level or observations for each school. The second column specifies the school year of the observation. There are four columns

dedicated to identification data. Six columns are per pupil cost categories, and one column is total per pupil cost which is the sum of the other six per pupil cost columns. The PPE data sets included data for grade schools; elementary, middle, junior high, junior-senior high, and high schools; intermediate schools and secondary schools; and county schools, community schools, magnet centers, attendance centers, and public schools. However, conducted data analyses only included PPE data from schools that also recorded ACT Composite scores.

2.1 Preparing data

ACT Composite Scores The NDE collected ACT composite score data for the academic years of 2017-2018 (N= 519, n= 241), 2018-2019 (N= 518, n= 238), and 2020-2021 (N= 519, n=241). N represents the total number of average ACT Composite Scores reported and n represents the number of high schools analyzed in this report. The ACT Composite Score data set included ACT data for public schools, learning communities, community schools, junior-senior high schools, senior high schools, and secondary schools; but only data for junior-senior high schools, high schools, and secondary schools were included in the analyses. Some high schools and secondary schools did not have ACT composite score data, and were removed from the sample.

The cleaned dataset for ACT composite scores has 9 columns and 803 rows. The first and sixth columns list the level and school type, denoting that the school is a highschool. The second column lists the school year. The third column lists the Agency ID for each school. The seventh column lists the school name. The Agency ID and School Name represents the school from which the data is being reported. The eighth column reports the average composite ACT score. The ninth column reports how many students' ACT composite scores were reported for the particular school. Rows 1 to 268 represent the data from the 2020-2021 school year. Rows 269-535 represent the data from the 2018-2019 school year. Rows 536-803 represent the data from the 2017-2018 school year. Table 1 provides a summary about the average ACT scores for different years.

Table 1: Average Composite ACT Score

Year	Mean
20172018	19.75985
20182019	19.58256
20202021	19.03608
20212022	18.73788

ACT Assessments Data All Subjects The NDE collected subject and demographic information for the academic years of 2016-2017, 2017-2018, 2018-2019 and 2020-2021. The all subjects dataset consisted of columns pertaining to student testing rate, ethnicity and gender information, different subject categories and school identification information. Columns pertaining to student testing rates, including: Developing_percent, On_Track_percent, ACT_Benchmark_Percent, Tested_Count, Not_Testing and Not_Testing_Percent, were removed from the data sets for each year. Schools that did not report ACT scores were denoted with '-1'. Rows with these missing ACT scores were also removed. Each yearly dataset consisted of school (SC), district (DI) and state level (ST) information. The state level aggregate scores were used for the demographic analysis. Rows pertaining to district aggregate scores and individual school scores were removed from each yearly dataset. Datasets for the four years collected were combined for analysis. For ethnicity data exploration, a subset of the entire dataset was created using the SUBGROUP DESCRIPTION column. Similarly, a subset for gender category was also created.

The cleaned dataset was broken up by factors being looked at. The dataset for gender has 24 rows and 15 columns. The data set for the race has 84 rows and 15 columns. The columns in both datasets are the same. The first column reports the school level. The second column reports the school year. The third column reports the county of the school. The fourth column is the school district. The fifth column is the

school type. The sixth column is the name where the reported data is from. For both data sets, all data being looked at is from the reported data from the State of Nebraska. The seventh column is the subject of the ACT scale score, English Language Arts, Mathematics, or Science. The eighth column represents the subgroup type, which is “gender” or “race ethnicity” for the different data sets. The ninth column lists the description of the subgroup. The data set for gender has only 24 rows because the data is subgrouped by gender, which breaks up the data by male and female. The data set for race has 84 rows because the data is subgrouped by race, which breaks up the data into 7 race categories: American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native American or Other Pacific Islander, White, or two or more races. The tenth column represents the average ACT scale score. Table 2, 3 provides summary statistics about the average ACT scores based on gender and Ethnicity for different years respectively.

Table 2: Summary statistics of ACT scores based on gender

	vars	n	mean	sd	min	max	range	se
Female	1	12	18.67583	0.4741971	17.78	19.41	1.63	0.1368889
Male	2	12	18.06167	1.1780248	15.95	19.21	3.26	0.3400665

Table 3: Summary statistics of ACT scores based on Race

	vars	n	mean	sd	min	max	range	se
American Indian or Alaska Native	1	12	14.04083	0.8900506	12.43	15.34	2.91	0.2569355
Asian	2	12	18.82667	0.8850612	17.44	19.74	2.30	0.2554952
Black or African American	3	12	14.14750	1.0247938	12.02	15.43	3.41	0.2958325
Hispanic	4	12	15.64083	0.9413381	13.70	16.66	2.96	0.2717409
Native Hawaiian or Other Pacific Islander	5	12	17.02167	0.9703592	15.50	18.66	3.16	0.2801186
White	6	12	19.59500	0.7184263	18.24	20.52	2.28	0.2073918
Two Or More Races	7	12	17.14167	0.7547888	15.70	18.19	2.49	0.2178887

Attendance Rate In the original data set, found and provided by the State of Nebraska Department of Education website, the secondary schools were denoted by the category of “1” in the “school” column. After converting the data set into a data frame and isolating the needed columns (School Year, School, District, Agency Name, Attendance Rate) the schools in the category “1” were selected and made into a subset. Next, the “school” category column was no longer needed and removed. All the columns were converted to lowercase and any spaces were removed, in order to convert the subset from wide to long. Once the data was in long form, the district codes needed to be added back and duplicates removed. Next, any year not relevant was removed. focusing on the school years 2015-2016 through 2020-2021.

While processing and researching the schools in Nebraska, it was quickly realized that several schools had either closed, combined, or opened. During the school year that a school was closed or not yet opened, the attendance rate denoted in the data set as “NA”. This, combined with information provided by the Nebraska Department of Education’s website, allowed the team to accurately, by year, combine schools that were renamed. To combine them, columns needed to be converted from character to numeric, then the rows that needed to be combined were added together, and removed the row that was combined with the new name of the school. The final data set used the most current name of the schools.

A few schools were removed from the data set. First, any school that was new, meaning not combined or renamed, after 2020. Any school that did not report any data in the school year 2019-2020. This data is critical for reviewing the effect covid had on attendance, thus if it was not reported the school was removed from our data set. This was due to the lack of pre-covid data. Lastly, there was a group of schools that all combined in 2020-2021. This combined 3 very small schools into one. Due to the size of the population, very small, and the approach to covid-19 not being uniformly approached it was decided that they would be

removed from the data set. Any schools that didn't report attendance rate for a given year was denoted as "-1". No additional information was given on why this information was not reported. To calculate the mean of the attendance rates, these were changed to NA's. These NA's were ignored to calculate the mean of the attendance rate.

The major school years explored in our data analysis are 2018-2019, 2019-2020, and 2020-2021. In 2018, Lynch Secondary School closed. In 2020, Verdigre High School opened. In 2018, St. Paul High School was renamed St. Paul Jr-Sr High School. There is no data for attendance rate for St. Paul Jr-Sr High School reported by NDE in the years of 2018-2019, 2019-2020, and 2020-2021. Given that there is no data for at least one of the 2018-2019, 2019-2020, and 2020-2021 school years, the data from Lynch Secondary School, Verdigre High School, and St. Paul High School was removed from our data sets.

In 2019, Ewing High School, Clearwater High School and Orchard High School combined to form one public high school named Summerland High School. Due to the differences in school population and years opened, we decided that the data from each of the schools: Ewing High School, Clearwater High School, Orchard High School, and Summerland High School were not significant enough to include in our data set. Therefore, the data from each of the schools was removed from our data set.

The years of schools being opened and closed can be found on the NDE website[8].

The cleaned data has 235 rows and 8 columns. The first column lists the name of the high school with reported attendance rates for the given years. The second through seventh columns represent the school years of 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, 2020-2021. The eighth column listed the school district of each high school. Each row represents the attendance rates for given years and district for each high school. Table 4 provides summary statistics about attendance rate per year.

Table 4: Summary Table for Attendance Rate by Year

	vars	n	mean	sd	median	min	max	range
20152016	1	235	0.9446268	0.0257447	0.9496	0.7236	0.9995	0.2759
20162017	2	235	0.9403349	0.0293064	0.9465	0.7046	0.9753	0.2707
20172018	3	235	0.9378672	0.0285255	0.9442	0.6908	0.9741	0.2833
20182019	4	234	0.9376090	0.0358898	0.9469	0.6419	0.9974	0.3555
20192020	5	235	0.9382574	0.0257071	0.9438	0.7305	0.9926	0.2621
20202021	6	235	0.9300345	0.0341799	0.9361	0.6471	0.9990	0.3519

Per Pupil Expenditure The PPE and ACT datasets were loaded into R directly from the NDE website. Cleaning the datasets consisted of removing the null values (-1) in the ACT dataset as some schools did not report average composite ACT scores. The PPE dataset did not have any null values. Next the PPE and ACT datasets for a given year were merged using "county", "district", and "school" as the keys. Merging the datasets in this way ensured that only observations that had both per-pupil expenditure data and ACT composite score data were included. To ensure that each observation of ACT corresponded with the same observation of PPE, a function was run to compare the 'name' column from each dataset. All names were equal, giving credibility that the datasets were merged appropriately. Next, the ACT/PPE datasets for 2018-2019 and 2020-2021 (the only two common school years that data is available) were merged. Since both datasets had the same column definitions this merge was performed by combining rows. Finally, the 'county' identifier was replaced by its corresponding county name[11].

The cleaned ACT/PPE dataset has 15 columns and 926 observations. There are four columns for identification: 'name', 'county', 'district', and 'school'. The 'type' column is the aggregations level, which indicates whether the observation is for the State, a district, a learning community, or an individual school. There are seven columns for various per-pupil expenditure categories. The 'datayears' column indicates the school year of the observation. Finally, there are two columns for ACT test data, one for average composite ACT score, and the other is the number of test-takers for the observation. Table 5 provides summary statistics about per pupil expenditure and ACT score for different years.

Table 5: Mean Scores and Per Pupil Expenditure by School Year

Year	ACT Score	Per-Pupil Cost
20182019	19.58256	18083.54
20202021	19.03608	19554.48

3. Data Product

Since the data explored in this report is related directly to Nebraska and the various counties or districts in the state, it is reasonable that geospatial visualizations are relevant and valuable. Additionally, with the various factors being explored there would be additional value in building a visualization tool that could allow the reader to explore the various combinations of factors in an effort to discover new and interesting study areas. To that end, the scope of a data visualization dashboard is developed in this section. The web-based data visualization dashboard has several tabs and makes use of the ‘shinydashbaord’ package in R.

For the data product, a dashboard was built to display the data and findings. The audience for this product ranges from administrators to parents and anyone interested in ACT data trends for the State of Nebraska(researchers, etc.). The dashboard has four tabs, corresponding to four areas of focus: ACT Dataset, ACT and per pupil expenditure, ACT and Attendance, and ACT and Demographics.

3.1 ACT Dataset

This tab has two sub-tabs, ‘ACT by County’ and ‘ACT Timeseries’. The ‘ACT by County’ tab allows the user to select the school year. Then, a Nebraska county-level heatmap of ACT composite scores for each county is generated in the main panel. The ‘ACT Timeseries’ tab generates a time series plot of average ACT scores for the selected county. Multiple counties may be selected for comparison of ACT score trends between counties.

3.2 ACT and Per Pupil Expenditure:

The ACT and Per Pupil Expenditure tab item contains visualizations generated from analyses of the ACT and PPE data sets and contains three sub-tabs: ‘Correlation’, ‘Pupil Expenditure Maps’, and ‘ACT vs PPE Animation’.

In the ‘Correlation’ sub-tab, a heat correlation plot of multiple expenditure variables and average ACT composite score is displayed on the first tab of this section. The color of the cells on the correlation plot is scaled and indicates the r-value for each correlation, where dark blue represents -1.0, blue represents 0, green represents 0.5, and yellow represents 1.0. Hovering on a cell displays the variables of that cell and the corresponding correlation. Clicking on a cell of the correlation plot will display a scatter plot corresponding to the variables of the selected cell in the correlation heat map. Furthermore, the school year (2020-2021 or 2018-2019) can be chosen to view different correlations for different years. This heat correlation plot allows the viewer to examine multiple correlations of expenditure and average ACT composite score in one place.

In the ‘Per Pupil Expenditure Maps’ sub-tab, a heat map of Nebraska that displays the average values of different expenditure variables by county is displayed. A box on the left-hand side contains two drop-downs where an expenditure variable can be selected, and a school year can be selected. The color for this heat map is scaled from blue to light pink. The viewer can easily see which counties and areas of Nebraska expenditure is abundant and where it is sparse on this heat map.

The ‘ACT vs PPE Animation’ sub-tab shows an animation of the scatter plots depicting the linear relationships between PPE and ACT for the 2018-2019 and 2020-2021 academic school years. This animated scatter plot was generated to allow the viewer to better visualize the change in trends between the years. Clicking

the “Play” button will run the animation, and the animation is also interactive where the viewer can hover over a data point and view more detailed information about that data point (such as the school name, per pupil cost, average ACT composite score, school year, district, and average teacher salary.)

3.3 ACT and Attendance

The dashboard tab for ACT and Attendance consists of two visualizations. The first tab shows a histogram of attendance rate per school. The histogram can be selected based on the academic years (2015-2021). User input for year choice is created using the select tab. The histogram has attendance on x-axis and count on y-axis. The second tab depicts the visualization of attendance rate for the academic years 2015-2021 for each district. The district selection is created using the select tab. When a particular district is selected, the visualization displays the attendance for that district for each academic year on the same plot.

3.4 ACT and Demographics

The dashboard tab for Demographics consists of 4 sub tabs. The first sub tab depicts the ACT scores based on the gender for years (2020-2021, 2016-2017, 2017-2018, 2018-2019). A select tab is created for user choice for both year and gender. On the right side of the dashboard based on the user input for gender and year a bar plot is generated. The bar plot shows how each gender scored for 3 subjects namely English, Mathematics and Science for a particular year. For example, Males scored on average 16.97 in English, 19.11 in Mathematics and 18.94 in Science in the 2016-2017 academic year. The second sub-tab depicts the average ACT scores based on different ethnicity groups for various years. A select tab is created for user choice for both year and gender. On the right side of the dashboard based on the user input for gender and year a bar plot is generated. The bar plot shows how each group scored for 3 subjects namely English, Mathematics and Science for a particular year. For example, Hispanics scored on average 15.13 in English, 16.55 in Mathematics and 16.34 in Science in the 2016-2017 academic year. Third sub-tab shows summary information based on different demographics for each year. Select tabs are created for user choice for both year and gender/ethnicity selection. On the right side of the dashboard based on the user input for gender/ethnicity and year selection, a box plot is generated. The boxplot shows how each category scored for a particular year. For example, the range of Male scores are between 16.5 and 19.5 for the year 2016-2017. On the same plot Female boxplot is also shown. Fourth sub-tab shows summary information for all the years faceted by gender or ethnicity. A select tab is created for user choice for gender/ethnicity selection. On the right side of the dashboard based on the user input for gender/ethnicity, box plots are generated for 2016-2017, 2017-2018, 2018-2019, 2020-2021 academic years.

4. Results

4.1 Absenteeism Related to Covid

This analysis attempts to answer the following question: Did COVID impact ACT scores during the years of 2018-2019, 2019-2020, and 2020-2021?

Table 6 shows the average attendance rate of all high schools in the school years: 2015-2016, 2016-2017, 2017-2018, 2018-2019, 2019-2020, and 2020-2021. The average for 2019-2020 is comparable to the average of previous school years, given that absences related to COVID for the months of March, April, and May were not recorded in the average for each school. The average for 2020-2021 is lower than the average for previous years, given that the districts began recording absences.

Table 6: Average Attendance for All High Schools Per Year

2015-2016	2016-2017	2017-2018	2018-2019	2019-2020	2020-2021
0.9452562	0.9413828	0.9389169	0.9400763	0.9397159	0.933678

The following histograms in figure 2 display the frequency of average attendance rates by district in the years of 2018-2019, 2019-2020, and 2020-2021. For each histogram, the highest frequencies are found near 0.95 attendance rate. In the 2020-2021 school year, there is more variability of the distributions of attendance rates between 0.85 and 1.0. There are several values for attendance rates between 0.80 and 0.90 in the 2020-2021 school year. During the 2018-2019 and 2019-2020 school years, there are at most two values for attendance rates between 0.80 and 0.90.

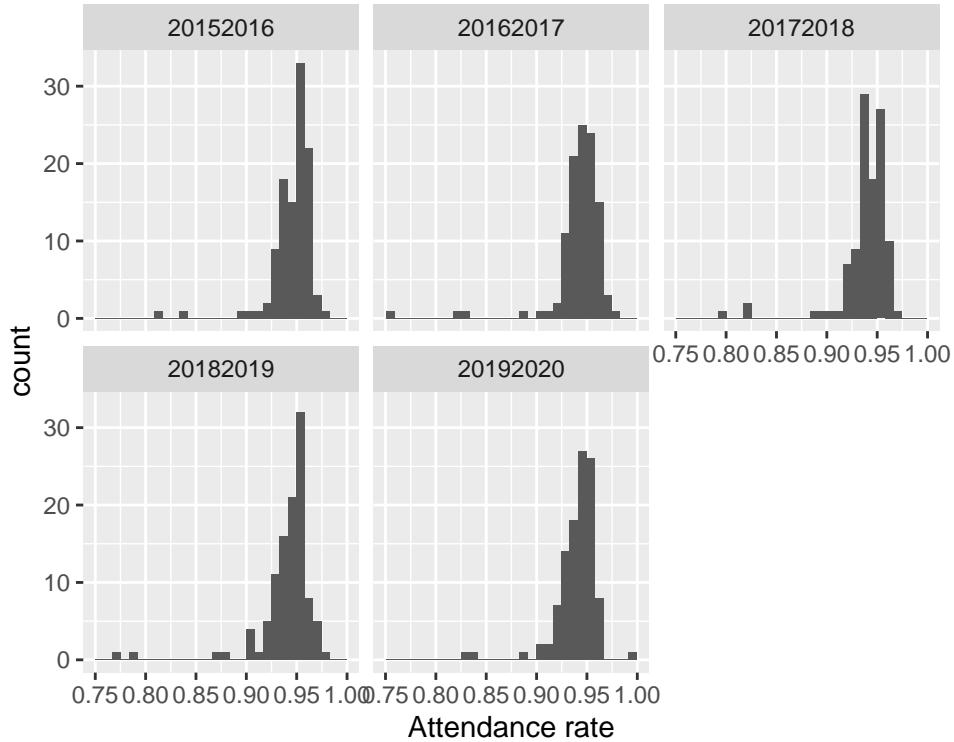


Figure 2: Histogram depicting attendance rate based on each district for 2015-2016,2016-2017, 2017-18, 2018-19,2019-2020 and 2020-2021 school years

The Attendance tab for the dashboard provides insights about how attendance rate varied over different years. For the years analyzed, the plots did not show any significant drop in attendance rate. Based on histogram plots from the dashboard,the distribution of attendance for 2015-2016 academic year was in the range of 89-97%. For the 2016-2017 academic year the attendance rate ranged between 87-96 %. For the 2017-2018 academic year, the range of attendance rate was between 88-94 %. For the 2018-2019 academic year, the range of attendance rate was between 86-97 %. For the 2019-2020 academic year, the range of attendance rate was between 87-98%. And for the 2020-2021 academic year, the attendance rate distribution range was 86-97%.The second visualization depicts attendance rate for different academic years for each district. Upon selection of a particular district, the plot displays the attendance for each year. Most of the districts had attendance rates in the range of 90-98%.

4.2 Per Pupil Expenditure

Does per pupil expenditure impact ACT scores?

To answer this research question, Pearson correlations were conducted to examine the linear relationships between PPE and average ACT composite scores among Nebraska public high schools for the academic school years of 2018-2019 and 2020-2021. It was found that PPE and ACT for the 2018-2019 academic school year do not have a significant, linear relationship ($r = -0.0645$, $p = 0.1668$). Data analysis produced a mean PPE of 18,084 dollars and an average of 19.6 for the average ACT composite scores for that school year. However, the Pearson correlation evaluating PPE and average ACT composite scores for the 2020-2021 school year revealed that there was a significant, linear relationship between these variables ($r = -0.1422$, $p = 0.0021$). The mean PPE for the 2020-2021 school year was 19,554 dollars, and the mean average ACT composite score for 2020-2021 was 19.0. The generated scatter plots show a weak, negative linear relationship for both academic years.

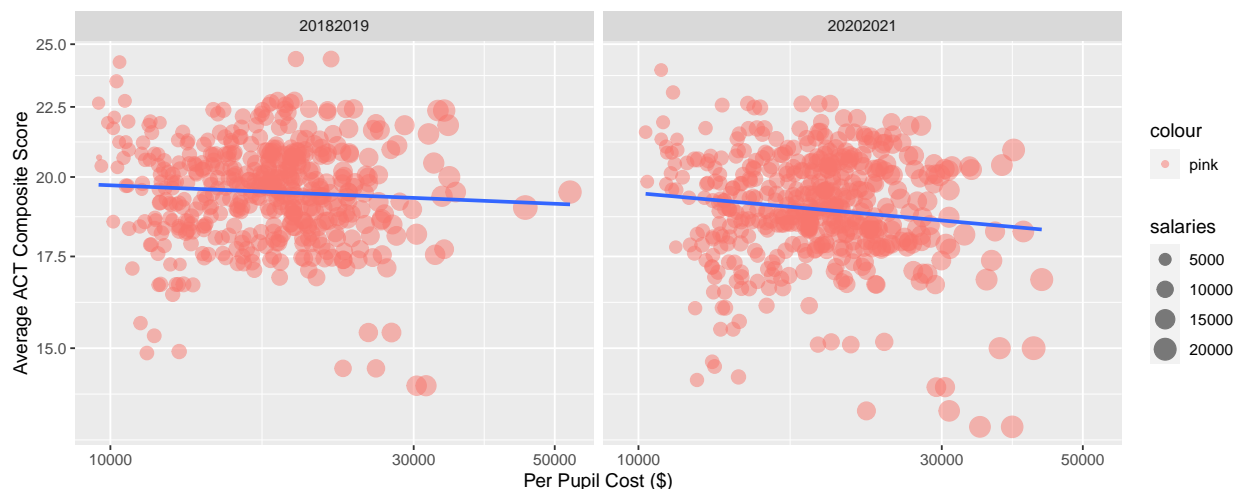


Figure 3: PPE and average ACT composite scores for the 2018-2019 school year produced an insignificant linear relationship. However, a significant, weak, negative linear relationship was found between PPE and average ACT composite score for the 2020-2021 school year.

Comparison of the individual components of per-pupil expenditure and ACT composite score for each school revealed that the overall trend of insignificant or weakly negative correlation between per-pupil expenditure and ACT outcomes remains, as shown in Figure 4. The direction and strength of the observed r -values produced by the correlations of Per Pupil Cost, 3% of Building, State/Local, Federal, Benefits, Salaries, and Admin with average ACT composite score changed very little between the 2018-2019 and 2020-2021 academic school years. Furthermore, Federal continues to be the most strongly correlated with average ACT score for the 2018-2019 and 2020-2021 school years.

4.3 Demographic factors influencing ACT scores

Does gender affect the overall ACT scores?

To answer this research question, the ACT scores over 4 years (2016-2017, 2017-2018, 2018-2019 and 2020-2021), were analyzed using the ACT all subjects dataset. Figure 6 illustrates the average ACT scores grouped for different years faceted by gender.

The Figure 6 provides the following insights. For the academic year 2016-2017, the average ACT scores for Males ranged between 17.0 and 19.1 with median at 18.9. The violin plot was right-skewed. For females the ACT scores ranged between 18.7 to 19.1 with median at 18.8. For the academic year 2017-2018, the average

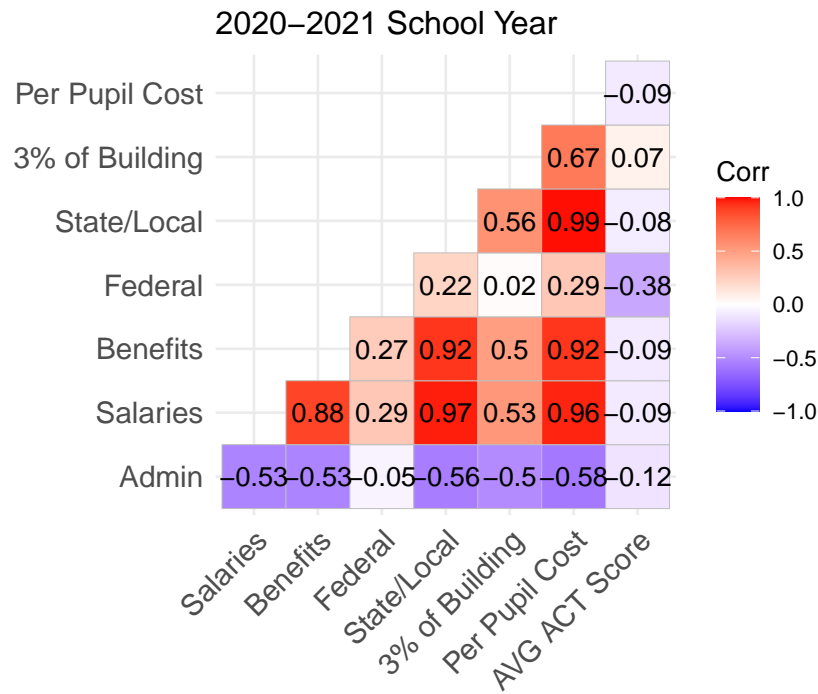


Figure 4: Correlations between ACT Composite scores and various categories of per-pupil expenditure for the 2020-2021 school year.

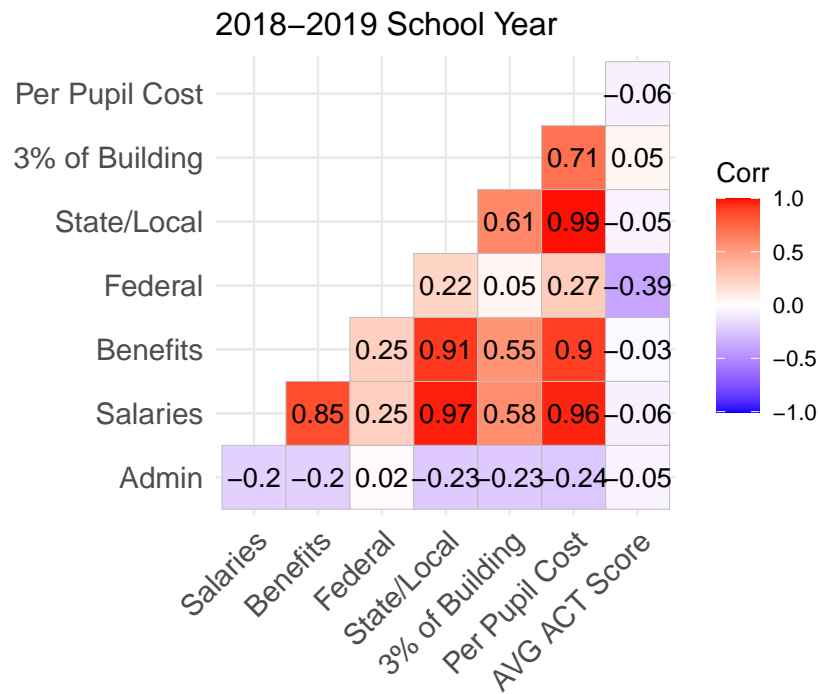


Figure 5: Correlations between ACT Composite scores and various categories of per-pupil expenditure for the 2018-2019 school year.



Figure 6: Violin plots depicting ACT scores for the 2016-2017, 2017-18, 2018-19, and 2020-21 school years faceted by gender

ACT scores for males ranged between 16.6 and 19.2 with median at 18.9, and for females the ACT scores ranged between 18.6 to 19.4 with median at 18.9. For the academic year 2018-2019, the average ACT scores for males ranged between 16.8 and 19.1 with median at 18.9. For females, the ACT scores ranged between 18.6 to 19.2 with median at 18.8. For the academic year 2020-2021, the average ACT scores for males ranged between 15.9 and 18.2 with median at 18.1. For females, the ACT scores ranged between 17.8 to 18.4 with median at 17.9. Based on these plots for the academic years analyzed, females scored higher compared to males. There is a significant drop in ACT scores for the 2020-2021 academic year for both genders. Males scores show a higher median score across all years. The range of ACT scores is considerably larger for male students compared to females. Both male and female scores showed a significant drop for the 2020-2021 school year. Female scores are right-skewed and male scores are left-skewed. The maximum ACT score for female students is higher than male students.

Does Race affect the overall ACT scores?

To answer this research question, the ACT scores over 4 years (2016-2017, 2017-2018, 2018-2019 and 2020-2021) were analyzed using the ACT all subjects dataset. Figure 7 illustrates the average ACT scores grouped for different years faceted by ethnicity. Different ethnic groups noticed in the dataset are American Indian, Asian, Black or African American, Hispanic, Native Hawaiian or Other Pacific Islander, White, Two or more races.

The Figure 7 provides the following insights. For the ethnicity category, in the academic year 2016-2017, the average ACT scores for White ranged between 19.0 to 20.2. For Asian the range was between 17.8 and 19.7. For Native Hawaiian's the ACT scores were in the range of 15.5 and 17. For Hispanic it was between 15.1 and 16.6. For American Indians the ACT score ranged between 13 and 14.6. For the Black group the ACT scores ranged between 13.5 and 15.0. Two or More Races ACT scores ranged between 17.1 and 18.2. In the academic year 2017-2018, the average ACT scores for White ranged between 18.8 to 20.5. For Asian the range was between 17.4 and 19.7. For Native Hawaiian's the ACT scores were in the range of 16.3 and 16.8. For Hispanic it was between 14.9 and 16.7. For American Indians the ACT score ranged between 13.4 and 15.3. For Black group the ACT scores ranged between 13.6 and 15.4. Two or More Races ACT scores

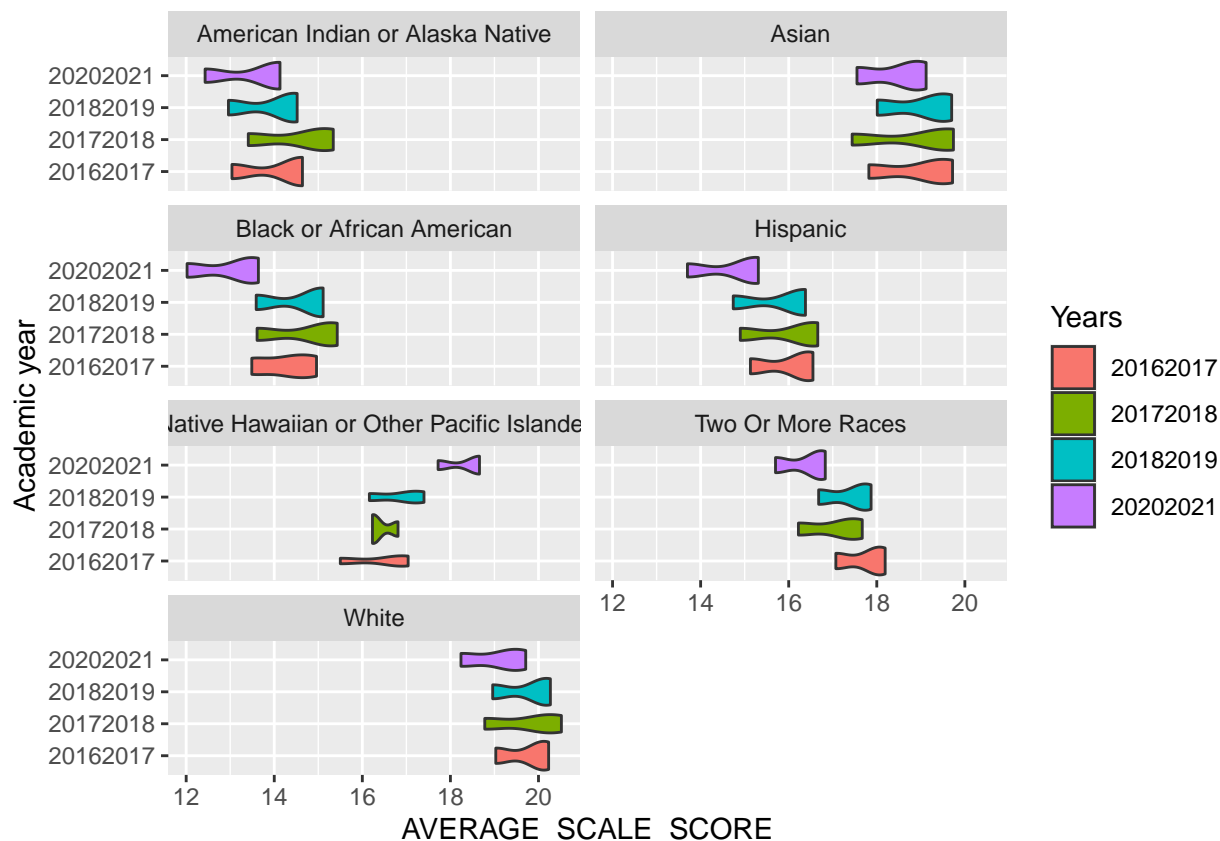


Figure 7: Violin plots depicting ACT scores for the 2016-2017, 2017-2018, 2018-2019, and 2020-2021 school years faceted by ethnicity

ranged between 16.2 and 17.7. In the academic year 2018-2019, the average ACT scores for White ranged between 19 to 20.3. For Asian the range was between 18 and 19.7. For Native Hawaiian's the ACT scores were in the range of 16. 2 and 17.4. For Hispanic it was between 14.7 and 16.4. For American Indians the ACT score ranged between 13 and 14.5. For the Black group the ACT scores ranged between 13.6 and 15.11. Two or More Races ACT scores ranged between 16.7 and 17.9. In the academic year 2020-2021, the average ACT scores for White ranged between 18.2 to 19.7. For Asian the range was between 17.6 and 19.1. For Native Hawaiian's the ACT scores were in the range of 17.7 and 18.7. For Hispanic it was between 13.7 and 15.3. For American Indians the ACT score ranged between 12.4 and 14.11. For the Black group the ACT scores ranged between 12.0 and 13.6. Two or More Races ACT scores ranged between 15.7 and 16.8. Overall, all ethnicities show a relatively consistent distribution of test scores. White and Asian students scored higher than other ethnic groups represented in the data for the four years of data analyzed. All ethnic groups in the dataset showed a drop in ACT scores for the 2020-2021 school year with the exception of Native Hawaiian/Other Pacific Islander which saw an increase in ACT scores.

The dashboard tabs for demographics also included plots for each gender and ethnicity based on year for different subjects. From the dashboard plots for each gender based on year , for the academic year 2016-2017 the highest average ACT scores for English and Sciences were scored by Female (18.73,19.11). While the highest ACT score for Mathematics was by Male (19.11). For the academic year 2017-2018 the highest average ACT scores for English and Sciences were scored by females (18.6,19.41). While the highest score in Mathematics was scored by Male(18.9). For the academic year. For the 2018-2019 academic year, the highest average ACT scores for English and Sciences was (18.56,19.16) scored by females. For Mathematics the highest score was 19.07 by a Male. For the academic year 2020-2021 the highest average ACT scores for English and Sciences were scored by females(17.78,18.39) scored. For Mathematics the highest score was

18.11 by a Male. Based on those plots, for the academic years analyzed, the highest ACT scores for English was scored by Females, Males scored higher Mathematics ACT scores when compared with Females. Females scored higher Sciences ACT scores when compared with males.

The second sub-tab provided plots for each ethnicity based on years. Based on those plots for the academic years analyzed (2016-2019,2020-2021), the highest ACT scores for English, Mathematics and Sciences were scored by White Ethnic groups. The third and fourth sub tab provide summaries on gender and ethnicity using boxplots for different academic years similar to violin plots already discussed above.

4.4 Topics for Prospective Studies

4.4.1 Per Pupil Expenditure Pearson correlations were conducted to evaluate linear relationships between PPE and average ACT composite scores of Nebraska public high schools for the 2018-2019 and 2020-2021 school years. Analyses showed that there was no significant linear relationship between PPE and average ACT composite score for the 2018-2019 school year, however there was a significant, weak, negative linear relationship between PPE and ACT for the 2020-2021 school year. Although the linear relationship for the 2018-2019 year was found to be insignificant, these findings suggest that PPE may have a significant linear relationship with average ACT composite scores for future academic school years. Prospective studies evaluating PPE and average ACT composite scores should continue to evaluate this linear relationship and should also consider PPE as a predictive variable in a linear regression or in a multiple regression, so long as the observed correlation remains significant.

Another area for future study lies in exploring administrative costs and how they are defined in the PPE dataset. It was discovered that the per-pupil expenditure value for a district was, in several cases, an order of magnitude greater than the administrative per-pupil expenditure for any given school in the district. For example, the administrative per-pupil expenditure for Omaha Public Schools at the district level is \$50,288, while the highest administrative per-pupil expenditure for an individual school in that same district is Omaha South High School with \$2,964. From this information, it follows that the district-level per-pupil costs are not an average cost per pupil as measured at the individual school level. One possible explanation could be that school bus costs are evaluated at the district level and that these costs are not included in school level cost information, but clarity was not readily available or identifiable in the Data Definitions webpage on the Nebraska Department of Education website.

4.4.2 Demographics Data analyses regarding student demographics produced compelling results. The results revealed a sharp decline in average ACT scale score between the 2018-2019 and 2020-2021 school years across all races, except for Native Hawaiian or Other Pacific Islander students. Additionally, there was a greater decrease in average ACT scores of Black or African American students compared to students of other races. Prospective research studies and data analyses should absolutely include race as a variable when evaluating relationships between student demographics and ACT scores. Understanding why there was a sharp decline and how demographics impact student outcomes can lead to powerful educational and societal reform recommendations. Results also showed that there is a wider range in ACT score for males, but males tend to be on the higher end of the state average reported results, while females have a lower ACT score range and tend to have a lower median value. Additional research examining the reasons for these student outcomes differences based on gender could provide valuable information to better understand gender differences in secondary education.

4.4.3 Attendance Rate Due to how attendance data was collected for the 2019 - 2020 school year, it is hard to say how attendance affected ACT scores. When we ignore this value, the attendance score changed every little while the ACT score seemed to trend negatively. If attendance does not impact ACT scores, understanding what has changed during school hours could help explain the decline in ACT average scale scores.

5.0 Conclusion

5.1 Findings and Audience The datasets and findings in this study are relevant to high schools, colleges, and parents. ACT scores are often used by high schools to determine the effectiveness of instruction and curriculum. Although there is no direct evidence of higher scores based on particular instruction, in general, individual schools, districts, and administrators can look at the trends in ACT throughout the school years and determine areas for improvement as well as areas of success. Schools can also look at trends based on factors such as race and ethnicity or gender and target students who are at risk of being less successful on the ACT.

ACT scores are used by colleges to help with admission decisions, course placement, and academic advisement[10]. Colleges often have benchmark scores that are to be met in order to take a course. For example, the ACT College Readiness Benchmark for a College Algebra course is a 22 in the Math Subject Area, meaning that students with a 22 or higher on the Math section of the ACT are predicted to have a high success rate in a college-level algebra course. Colleges that receive several students from Nebraska public schools, such as the public universities in Nebraska, could use the information from this report to predict the need of general education courses based on the change in scores in each subject area. Parents are able to use this research to help guide their students based on current trends.

5.2 Limitations and Discussion

As with the majority of studies, the design of the current study is subject to limitations. One limitation is the collection of ACT scores by the State of Nebraska. Although the ACT is a required test for all public school students, students are able to apply to take an alternate summative assessment in place of the ACT, if necessary. This alternate assessment is typically granted to students with severe cognitive disabilities or multi-handicapping conditions and should not exceed 1% of the student population[6]. As the data reported is from the public high schools in Nebraska, there is a lack of data from private high schools, which accounts for a portion of the population. Data reported from the NDE only includes scores from the test taken during eleventh grade at each public high school. The data does not include any data from tests administered throughout the year by ACT[7]. There are students whose scores are not reported because they miss the test completely or they miss a subject of the test, therefore a composite score is not reported[7]. This is the reason for the different numbers on the composite scores and the All Subjects data. For example, a student could only finish the reading section and have a reported score for that section of the test, but not complete the rest of the test and not have an overall composite score[7]. There are no reported ACT Scores for the State of Nebraska during the 2019-2020 academic school year, given that the test is administered during March and April, when several schools were operating remotely. This impacts our data and research because we are missing a major year of ACT scores and how the beginning of a pandemic impacted the ACT score of students across Nebraska. Overall, there are several ACT scores that are not being reported on in the NDE collection or there are several students who are not taking the ACT and not being included within the NDE reports.

5.2.1 Per Pupil Expenditure: The data analyses from the conducted Pearson correlations between PPE and ACT for 2018-2019 and 2020-2021 yielded differing results where 2018-2019 showed an insignificant linear relationship, but 2020-2021 showed a significant, weak negative relationship. As stated previously, these findings suggest that PPE and average ACT composite scores may produce significant linear relationships for future academic years, and this relationship should be considered for future research studies and data analyses.

These findings also suggest that providing schools additional funds for students does not necessarily create more successful students, and due to the small r-values produced, there may be other important factors to consider for improving ACT composite scores and student outcomes.

Overall, the PPE dataset had little limitations, and the collected data was very tidy and very complete. However, some definitions and explanations of the data are missing. For example, some schools had very

low values in Federal funding and it is unclear if these values are due to a data error or if there is an administrative reason for them, and reported district-level per pupil expenditure values for administration was much higher than school-level per pupil expenditure values for administration and it is unknown how this this quantity is defined and why it is different from school-level.

5.2.2 Attendance: A limitation is found in the reporting of attendance rates in the 2019-2020 and 2020-2021 school years. During the 2019-2020 school year, the attendance rate is reported as if the school year ended in March, on whatever date each district closed or went fully remote, but the attendance rates for the first part of the year are valid rates[7]. In 2020-2021, the rates are based on whatever the districts chose to report as absences[7]. Therefore, the attendance rates for the 2019-2020 and 2020-2021 school year do not accurately represent the attendance rate for students for the entirety of each school year. This removed three months' worth of data. Making it difficult to see if lower attendance due to COVID-19 impacted ACT scores. If looked at holistically, either accepting or ignoring 2019-2020 reporting, the attendance rate really did not change in any significant way.

Nebraska's Department of Education does a poor job of highlighting or labeling schools that have closed, been combined, or are simply new. Leading to additional research when a school stops reporting data. In some cases, 2 or more schools were being combined, and it was never noted in the data itself. The schools that combined simply reported nothing or stopped reporting anything. The new school that was formed would then start reporting the year it opened. This could lead to data being misinterpreted or misleading. This would need to be evaluated every year to ensure all schools are being represented correctly.

It was discovered late in our data analysis that an error was made during the data cleaning process and several high schools were left out for the attendance rates. This is a small limitation to our results, but it is a starting point for prospective studies.

5.2.3 Demographics: When taking a closer look at race, it becomes pretty clear that something negatively impacted every student between 2018-2019 and 2020-2021. Interestingly, there is one exception, Native Hawaiian or Other Pacific Islanders' average scale scores went up. The general decrease in scores is not even across all races. Black or African Americans seem to have dropped more than others.

A majority of the data points in the ACT All Subjects dataset are from the White race, and there is a lack of data points from other races. Therefore, the data in our research focuses on the data from the State of Nebraska as a whole and cannot be analyzed per school, county, or district because several schools would be removed due to a lack of reporting scores for each race. Although there are several data points for both genders, we looked at the data points statewide to normalize the data.

When we look at gender, a few things pop out. While female students seem to have a less varied score, they skew left. Meaning that most of the scores for female students are on the low end. While male students have a greater varied score, they are skewed to the right. The reason for this is unknown but worth further study.

References

- [1] Herberger, G., Immekus, J., and Ingle, W. K. (2020). *Student, neighborhood, and school factors and their association with college readiness: Exploring the Implementation of a race- and socioeconomic-based student assignment plan.*, Education and Urban Society, 52(3), 459-488., doi: 10.1177/0013124519858128
- [2] Nebraska Legislature. (n.d.). nebraskalegislature.gov. Retrieved October 31, 2022, from <https://nebraskalegislature.gov/laws/statutes.php?statute=79-760.03>
- [3] Coronavirus FAQ – Nebraska Department of Education. (2020, June 11). Caronavirus FAQ. <https://www.education.ne.gov/publichealth/faq/>
- [4] Average ACT Score For the High School Class of 2022 Declines to Lowest Level in More Than 30 Years. (2022, October 12). ACT. <https://leadershipblog.act.org/2022/10/GradClassRelease2022.html>

- [5] NEP Data Downloads - Nebraska Dept of Education. (n.d.). [Nep.education.ne.gov](https://nep.education.ne.gov/Links). Retrieved October 31, 2022, from <https://nep.education.ne.gov/Links>
- [6] NSCAS Alternate Summative Assessment – Nebraska Department of Education. (2020, August). NSCAS Alternate Summative Assessment. <https://www.education.ne.gov/assessment/nscas-alternate-summative-assessment>
- [7] M. Reiner, personal communication, September 22, 2022.
- [8] NEP - Nebraska Dept of Education. (n.d.). [Nep.education.ne.gov](https://nep.education.ne.gov/). <https://nep.education.ne.gov/>
- [9] School Districts. (n.d.). [Www.nebraskamap.gov](https://www.nebraskamap.gov/datasets/nebraska::school-districts/about). Retrieved October 31, 2022, from <https://www.nebraskamap.gov/datasets/nebraska::school-districts/about>
- [10] ACT Test - Scores - Higher Education Professionals. (n.d.). ACT. Retrieved December 11, 2022, from <https://www.act.org/content/act/en/products-and-services/the-act-postsecondary-professionals/scores.html#:~:text=How>
- [11] Nebraska County Names and Numbers. (n.d.). Retrieved December 14, 2022, from https://revenue.nebraska.gov/sites/revenue.nebraska.gov/files/doc/pad/county_names_numbers.pdf