Analyzing the Relationship Between Poverty and Arts Participation

Jevon Lipsey, Camden Burke, and Rakim Johnson

MA117: Elementary Probability and Statistics, Colorado College

Luis David Garcia Puente

October 15th, 2024

Abstract

This project aims to explore whether socioeconomic status is a reliable predictor of arts participation levels, operating under the assumption that higher overall participation reflects broader youth involvement. The data collected supports the conclusion that arts participation is indeed linked to socioeconomic status. Students from higher-income families had more opportunities to explore various artistic pursuits at least once. Conversely, lower-income participants demonstrated more frequent and sustained engagement in the artistic opportunities available to them. These findings provide justification for further research on the directions of these dependency trends.

Background Information:

As initiatives to increase literacy and STEM education have been largely successful (Amer, 2019), we've seen an overall decline in participation in artistic pursuits (Arts, 2018). This is especially unfortunate when taken into consideration alongside a large body of research that would suggest that participation in the arts is linked with healthier lifestyles. Moreover, we are operating under the assumption that higher overall participation in the arts is linked to the proportion of youth involved in the arts. We are making this assumption because while metrics exist that indicate whether there are arts programs present in K-12, it is much harder to find concrete metrics that accurately explore whether students engage with the arts both outside of instruction time while in school and also once they leave the compulsory education system. While there are a myriad of potential causes for this, our group was more interested in ascertaining whether income played a role in determining if someone's socioeconomic status played a role in their overall participation in artistic pursuits.

Aims:

In essence, the question that we are attempting to answer is: *Is socioeconomic status a relatively accurate predictor of the level and extent of participation in the arts?* This question parses out the relationship between the variables while also leaving room to explore the data in a way that can lead to a more robust understanding of the role of arts in our society. In that pursuit of that goal, in consultation with our own notions of the role of arts, we constructed the following hypothesis statements:

H_0: Arts participation and Socioeconomic status are independent of one another.

H_1: Arts participation and socioeconomic are not independent of one another.

Data:

The data that we used was sourced from two primary sources: a series of surveys conducted by the University of Michigan and the National Center for Education Statistics (NCES). As we thought about bringing these two data sets together, we decided to focus on the year 2012 despite having a variety of years between 1984 and 2012 that contained data from both sources due to the recency paired with the sheer amount of data contained within that year from the University of Michigan data set. Moreover, we chose to use the data from the NCES because of its credibility and the fact that the data was readily available.

The variables that we chose to use were the percent of students that qualified for free or reduced lunch (stratified by state), Participation in the arts (stratified by type of artistic pursuit), and Income of participants (collected in the arts survey conducted by the University of Michigan). We chose to use both the information collected as a part of free and reduced-price

lunch programs in addition to the data collected for the income portion of the arts survey because they are both illustrative of poverty and socioeconomic status broadly across the country. The data collected about the free or reduced-price lunch program gives insight into the population while the data collected as a part of the survey is more reflective of the sample.

Methods:

First we Downloaded lunch and participation data, and filtered the data to focus on the information from 2012. Then we sorted states by high poverty and low poverty. The threshold we used was 50% of students qualifying for free or reduced lunch.

```
# data preparation
# datasets downloaded and cleaned on microsoft excel beforehand
install.packages("readxl")
library(readxl)
participation <- load("Downloads/35596-0001-Data.rda")
lunches <- read_excel("Downloads/lunches2012-13 clean.xlsx")

# Add a new column classifying states by poverty
lunches$`Poverty Category` <- ifelse(lunches$`Percent Eligible` > 50,
                                     "High Poverty", "Low Poverty")
# focus on 2012 participation data to match lunches data and has location
participation2012 <- subset(da35596.0001, YEAR == 2012 & !is.na(FIPS_STATE))
```

We will then figure out which states are high poverty and low poverty by graphing lunch data and creating high poverty and low poverty vectors with respective state codes.
Then, we performed data cleanup and sorted participation into the art types we're focused on. Afterwards, we can sort participation data by high and low poverty area codes. Once sorted, we can calculate means, and then store them in dataframes to conduct non-pooled t-tests for association. Doing non-pooled t-tests in R will provide p-values for each art type, providing evidence as to whether or not there is an association between poverty and arts participation.

Data Analysis Plan:

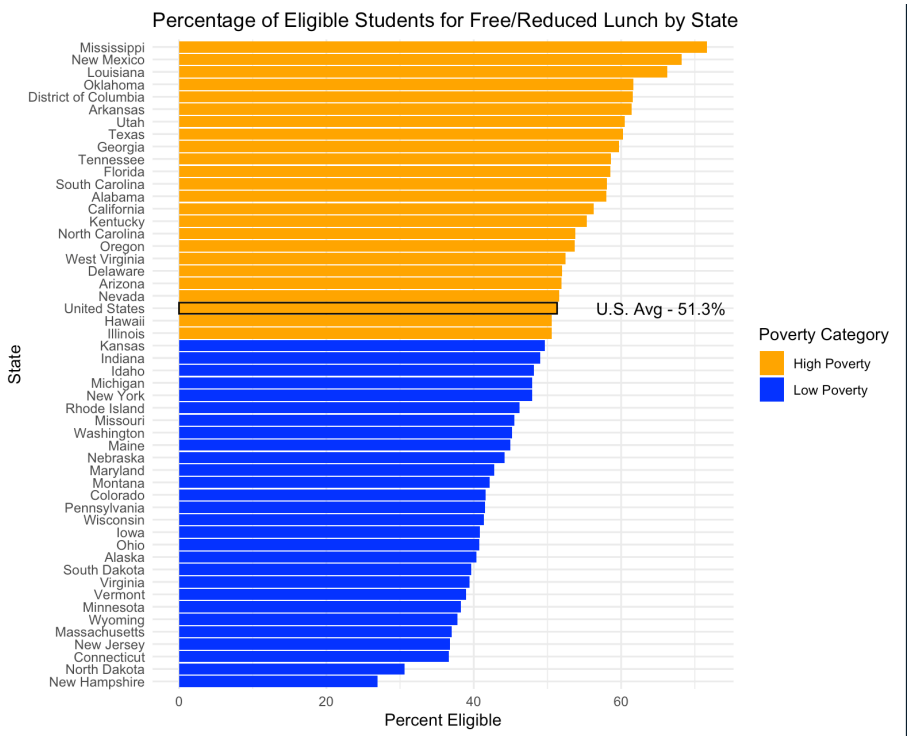
For our first test, we performed a non-pooled t-test. We chose this because high poverty vs. low poverty have different sample sizes and variances. This, along with the assumption that states are not dependent on each other, entails that our samples should be independent from each other. The sample size of high poverty is *n=8467* and the sample size of low poverty is *n=9584*.

Our objective is to test whether or not there is a statistically significant difference in arts participation between low-poverty and high-poverty areas. Low-poverty areas and high-poverty areas are defined by whether more than *50%* of students in the area are eligible for free/reduced lunch. Our hypotheses were as such: Null Hypothesis (H_0): There is no significant difference in arts participation between high- and low-poverty areas. Alternative Hypothesis (H_1): There is a significant difference in arts participation between high- and low-poverty areas.

Results:

First we had to determine which states fell into the high and low poverty categories.

```
# plot lunches percentage of eligible students by state
library(ggplot2)
ggplot(lunches, aes(x = reorder(State, `Percent Eligible`),
                    y = `Percent Eligible`, fill = `Poverty Category`)) +
  # highlight U.S
  geom_bar(stat = "identity", aes(color = ifelse(State == "United States",
                                                 "highlight", "normal"))) +
  scale_color_manual(values = c("highlight" = "black",
                                "normal" = "transparent"), guide = "none") +
  # legend
  scale_fill_manual(values = c("High Poverty" = "orange",
                               "Low Poverty" = "blue"),
                    name = "Poverty Category") +
  coord_flip() +
  # labels
  labs(title = "Percentage of Eligible Students for Free/Reduced Lunch by State",
       x = "State",
       y = "Percent Eligible") +
  theme_minimal() +
  # label U.S. average
  geom_text(aes(label = ifelse(State == "United States", "U.S. Avg - 51.3%", "")),
            hjust = -0.3)
```



Now, we can sort participation data by high poverty and low poverty:

```r
# high poverty states (derived from lunches)
high_poverty <- c("AL", "AZ", "AR", "CA", "DE", "DC", "FL", "GA", "HI", "IL",
                  "KY", "LA", "MS", "NV", "NM", "NC", "OK", "OR", "SC", "TN",
                  "TX", "UT", "WV")
# low poverty states
low_poverty <- c("AK", "CO", "CT", "ID", "IN", "IA", "KS", "ME", "MD", "MA",
                 "MI", "MN", "MO", "MT", "NE", "NH", "NJ", "NY", "ND", "OH",
                 "PA", "RI", "SD", "VT", "VA", "WA", "WI", "WY")
```

We had to clean the data to perform numerical analysis:

```r
# use sub to remove numerical codes from state codes
participation2012$FIPS_STATE <- sub(".*\\)\\s*", "", participation2012$FIPS_STATE)

# sort participation by arts, both yes/no and numerical
arts_vars <- c("FESTIVAL", "CRAFT_FAIR", "ARTMUSEUM", "PARK", "BALLET", "CLASSICAL",
               "DANCE", "JAZZ", "SALSA", "MUSICAL", "PLAY", "OPERA", "BOOKS")
arts_n_vars <- c("ARTMUSEUM_N", "BALLET_N", "CLASSICAL_N", "JAZZ_N", "SALSA_N",
                 "MUSICAL_N", "PLAY_N", "OPERA_N", "DANCE_N", "PARK_N", "CRAFT_FAIR_N")

# turn all variables to numeric - turn (1) Yes and (2) No to 1/0
participation2012[arts_vars] <- lapply(participation2012[arts_vars], function(x) {
  ifelse(grepl("\\(1\\)", x), 1, ifelse(grepl("\\(2\\)", x), 0, NA))
})
participation2012[arts_n_vars] <- lapply(participation2012[arts_n_vars], function(x) {
  as.numeric(x)
})
```

After doing this, we could sort participation data by high poverty and low poverty, allowing us to calculate means and visualize the differences (ignoring all N/A values).

```r
high_participation <- subset(participation2012, FIPS_STATE %in% high_poverty)
low_participation <- subset(participation2012, FIPS_STATE %in% low_poverty)

# calculate means for binary arts participation (yes/no)
high_poverty_means <- colMeans(high_participation[arts_vars], na.rm = TRUE)
low_poverty_means <- colMeans(low_participation[arts_vars], na.rm = TRUE)

# calculate means for frequency-based arts participation
high_poverty_n_means <- colMeans(high_participation[arts_n_vars], na.rm = TRUE)
low_poverty_n_means <- colMeans(low_participation[arts_n_vars], na.rm = TRUE)
```

Then, we could populate data frames with the means to visualize the data in barplots (Figure 1 and 2).

```r
# data frame for binary participation means
arts_means_binary <- data.frame(
  art_type = rep(arts_vars, 2),
  poverty_level = rep(c("High Poverty", "Low Poverty"), each = length(arts_vars)),
  mean = c(high_poverty_means, low_poverty_means)
)

# create bar plot for binary means
ggplot(arts_means_binary, aes(x = art_type, y = mean, fill = poverty_level)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "High Poverty vs Low Poverty Means by Art Type (Binary) Participation in the Last Year",
       x = "Art Type",
       y = "Mean (0-1)",
       fill = "Poverty Level") +
  scale_fill_manual(values = c("High Poverty" = "orange", "Low Poverty" = "blue"))
rm
# data frame for frequency-based participation means
arts_means_frequency <- data.frame(
  art_type = rep(arts_n_vars, 2),
  poverty_level = rep(c("High Poverty", "Low Poverty"), each = length(arts_n_vars)),
  mean = c(high_poverty_n_means, low_poverty_n_means)
)
# Create bar plot for frequency means
ggplot(arts_means_frequency, aes(x = art_type, y = mean, fill = poverty_level)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "High Poverty vs Low Poverty Means by Art Type (Frequency) Participation in the Last Year",
       x = "Art Type",
       y = "Mean Frequency",
       fill = "Poverty Level") +
  scale_fill_manual(values = c("High Poverty" = "orange", "Low Poverty" = "blue"))
```

Another question we asked was “Which types of art are more common in states and are they affected by income?” As seen in Figure 1, the top three most frequent responses within high poverty areas are jazz concerts, art museums, and salsa dancing . The top three most frequent responses for low poverty are jazz concerts, art museums, and classical. However, when using units of yes/no (binary, Figure 2) both high poverty and low poverty have the same top three responses: books, parks, and craft fairs.

Interestingly enough, the bars on Figure 2 show that low poverty areas do every single art type at least once more than higher poverty areas. However, the frequency graph (Figure 1) shows that high poverty areas participate in the arts more frequently than low poverty.

Figure 1:

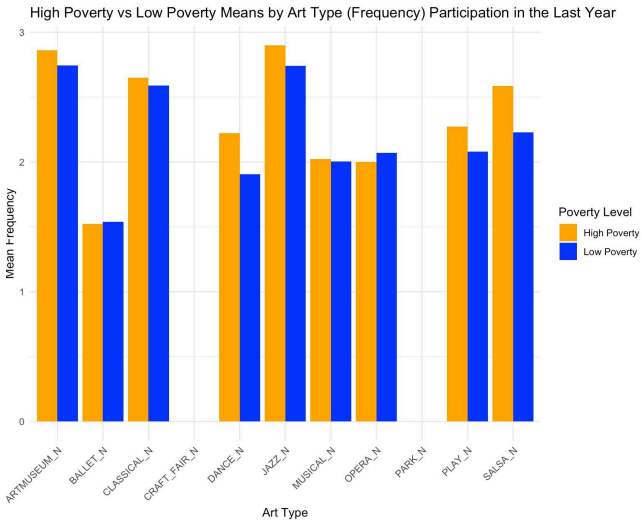
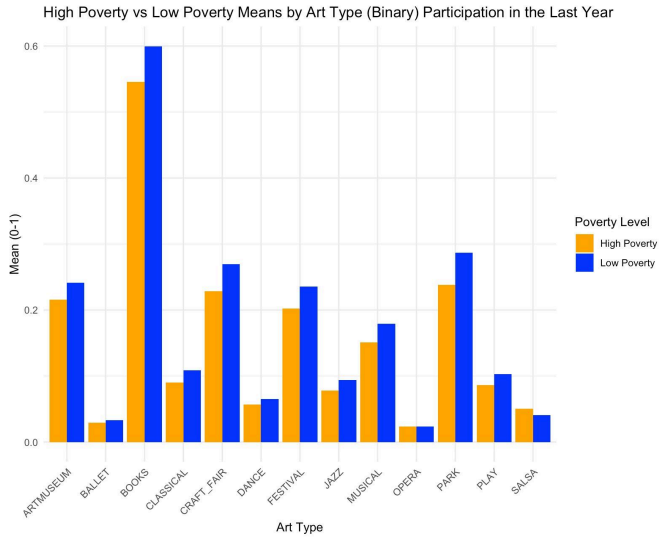


Figure 2:



To test the association between poverty level and arts participation, we conducted two-sided, independent non-pooled t-tests to calculate p-values for each art event. The p-values tell us the likelihood that the observed differences in participation rates between high-poverty and low-poverty areas are due to chance. We ran these tests for both binary and frequency responses.

Frequency responses in participation had null values for park and craft fair attendance, so it was omitted from the data.

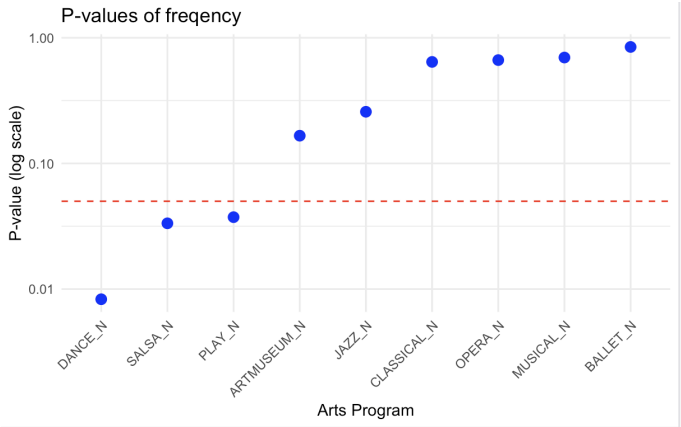
```r
# run t-tests for binary participation variables (art types)
binary_t_test_results <- sapply(arts_vars, function(var) {
  t.test(high_participation[[var]], low_participation[[var]], na.rm = TRUE)
})

# run t-tests for each frequency-based participation variable
# remove two N/A variables from arts_n_vars
clean_arts_n_vars <- arts_n_vars[!(arts_n_vars %in% c("PARK_N", "CRAFT_FAIR_N"))]
# now, run t-tests on the cleaned arts_n_vars
frequency_t_test_results <- sapply(clean_arts_n_vars, function(var) {
  t.test(high_participation[[var]], low_participation[[var]], na.rm = TRUE)
})
```

This gave us the p-values of each art event, which we could then visualize:

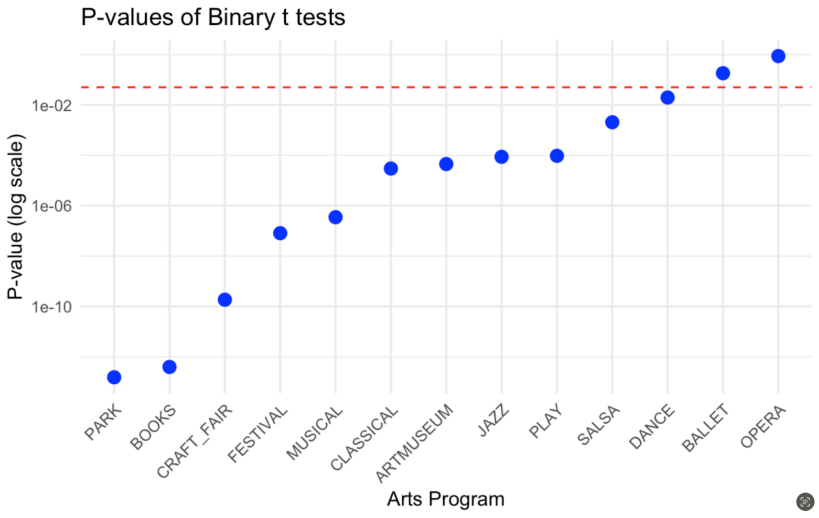
```
> p_values <- c(0.1661165, 0.8446182, 0.6420684,
+               0.2575247, 0.0333655, 0.6957627,
+              0.03734415, 0.6638199, 0.008301096)
> var_names <- c("ARTMUSEUM_N", "BALLET_N", "CLASSICAL_N",
+                "JAZZ_N", "SALSA_N", "MUSICAL_N",
+                "PLAY_N", "OPERA_N", "DANCE_N")
df <- data.frame(var_names, p_values)
> ggplot(df, aes(x = reorder(var_names, p_values), y = p_values)) +
+                    geom_point(size = 3, color = "blue") +
+                    geom_hline(yintercept = 0.05, linetype = "dashed", color = "red") +
+                    scale_y_log10() +  # Log scale to handle small p-values
+                    labs(x = "Arts Program", y = "P-value (log scale)",
+                                        title = "P-values of freqency") +
+                          theme_minimal() +
+                    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Figure 3:




```
> art_names <- c("FESTIVAL", "CRAFT_FAIR", "ARTMUSEUM",
+                "PARK", "BALLET", "CLASSICAL",
+                "DANCE", "JAZZ", "SALSA",
+                "MUSICAL", "PLAY", "OPERA",
+                "BOOKS")
> p_values <- c(8.083188e-08, 1.849201e-10, 4.493992e-05, 1.540362e-13,
+               0.1810027, 2.970914e-05, 0.01959071, 8.695962e-05,
+               0.00205433, 3.487997e-07, 9.500085e-05, 0.8672225,
+               3.958165e-13)
> df <- data.frame(art_names, p_values)
> ggplot(df, aes(x = reorder(art_names, p_values), y = p_values)) +
+     geom_point(size = 3, color = "blue") +
+     geom_hline(yintercept = 0.05, linetype = "dashed", color = "red") +
+     scale_y_log10() +  # Log scale to handle small p-values
+     labs(x = "Arts Program", y = "P-value (log scale)",
+          title = "P-values of Arts Programs") +
+     theme_minimal() +
+     theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Figure 4:



Analysis of p-values from non-pooled t-tests comparing high-poverty and low-poverty participation in various arts programs reveals significant disparities. Most programs in Figure 4 show statistically significant differences ($p < 0.05$), with the degree of significance varying across program types. Participation types like PARK, BOOKS, and CRAFT FAIR exhibit extremely low p-values ($p < 1e\text{-}10$), indicating highly significant differences. Other programs such as FESTIVAL, MUSICAL, and CLASSICAL show strong evidence of disparities ($1e\text{-}6 < p < 1e\text{-}4$). Some performing arts programs (e.g., SALSA, DANCE) demonstrate weaker, yet still significant, differences. A few programs in the frequency analysis (e.g., BALLET_N, MUSICAL_N) show p-values above 0.05 in Figure 3, suggesting potentially non-significant differences. Overall, these results indicate that poverty level is associated across most art participation types, with varying degrees of disparity. Some programs show highly pronounced differences, while others not as much; Further investigation into the nature and causes of these disparities is warranted.

Conclusion:

The data that we collected has allowed us to address our overarching question: *Is socioeconomic status a relatively accurate predictor of the level and extent of participation in the arts?* Based on the information given, there is a clear linkage between arts participation and socioeconomic status. In regards to the ability to try artistic pursuits at least once, our evidence showed that those from affluent backgrounds had the opportunity to attempt artistic pursuits more often. However, for the artistic pursuits available to those from lower-income backgrounds, it was clear that there was more repeated and ongoing participation. While this data doesn't

illuminate why this occurs, we instead can understand this phenomenon in context with our broader understanding of how artistic traditions are often forms of escape and resistance from and against the paradigm of suffering imposed on them by dominant forces. In other words, people from lower economic means may participate in artistic pursuits as a means of catharsis and rebellion against their material and emotional condition. That said, a limitation of this project is a lack of qualitative data to give a fuller picture of this phenomenon but also its origins. A more refined study of this topic may do well to include testimonial questions for participants engaging in the arts question so as to perhaps gain insight into the particularities of each individual's motivations for engaging in the arts.

Contributions:

Jevon: Frequency and Binary graphs and R calculations, Formatting sheets, and cleaning data, Results

Camden: Abstract/intro, Graphs of p-values and R, Methods

Rakim: Intro/ Conclusion, Methods, Powerpoint, Conceptual framework, Background Info

Sources:

Americans for the Arts. Arts Education Navigator: Facts & Figures. 2019, www.americansforthearts.org/sites/default/files/ArtsEd_Navigator_FactsFigures.pdf.

National Endowment for the Arts. The Arts and Achievement in At-Risk Youth: Findings from Four Longitudinal Studies. 2012, www.arts.gov/sites/default/files/Arts-At-Risk-Youth.pdf.

Arts Education Partnership. Music Matters: How Music Education Helps Students Learn, Achieve, and Succeed. 2011, www.aep-arts.org/wp-content/uploads/Music-Matters-Final.pdf.

National Endowment for the Arts. The U.S. Arts Economy. 2020, www.arts.gov/news/2020/us-arts-economy-2020-reports.

World Health Organization. What Is the Evidence on the Role of the Arts in Improving Health and Well-Being? A Scoping Review. 2019, www.euro.who.int/__data/assets/pdf_file/0020/411418/what-is-the-evidence-on-the-role-of-the-arts-in-improving-health.pdf.

National Endowment for the Arts. U.S. Trends in Arts Attendance and Literary Reading: 2002-2017. 2018, www.arts.gov/sites/default/files/2018-sppa-feb2019.pdf.

U.S. Department of Education. Arts Education in Public Elementary and Secondary Schools: 1999-2000 and 2009-10. National Center for Education Statistics, 2012, nces.ed.gov/pubs2012/2012014rev.pdf.

National Endowment for the Arts. U.S. Trends in Arts Attendance. 2018, www.arts.gov/sites/default/files/2018-sppa-feb2019.pdf.

California Alliance for Arts Education. Arts Education at Risk in California. 2011, www.artsed411.org/research/arts_education_at_risk.

National Endowment for the Arts. 2017 Survey of Public Participation in the Arts. 2018, www.arts.gov/sites/default/files/2017-sppa-september2018.pdf.

National Endowment for the Arts. The Arts in Communities. 2018, www.arts.gov/sites/default/files/nea-sppa-feb2019.pdf.