

Statistical Analysis of California Department of Development Service Discrimination Lawsuit

James Palmer - University of North Alabama - 30 November 2024



Abstract

This paper will examine data collected during a discrimination lawsuit against the California Department of Development Service (DDS) and then interpret that data with the purpose of determining the veracity of the claims made against the California DDS. Analysis will consist of the creation of a series of charts and graphs made to show any trends present among the collected data. The analysis performed on the California DDS's funding data shows little evidence of discrimination based on ethnicity or gender. The largest determining factor in funding was found to be the age group each individual belonged to, with older individuals receiving more funding on average.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 2 |
| 2 | Statement of Problem and Statistical Analysis Approach | 2 |
| 3 | Results | 2 |
| 3.1 | Analysis of Expenditures by Ethnicity | 2 |
| 3.2 | Analysis of Expenditures by Gender | 7 |
| 4 | Summary and Conclusions | 9 |
| 5 | Sources | 10 |
| 6 | Appendix: R Code | 10 |

List of Figures

| | | |
|----|---|---|
| 1 | Box Plot of Expenditures by Ethnicity | 3 |
| 2 | Box Plot of Expenditures for Individuals Aged 0 to 5 by Ethnicity | 4 |
| 3 | Box Plot of Expenditures for Individuals Aged 6 to 12 by Ethnicity | 4 |
| 4 | Box Plot of Expenditures for Individuals Aged 13 to 17 by Ethnicity | 5 |
| 5 | Box Plot of Expenditures for Individuals Aged 18 to 21 by Ethnicity | 5 |
| 6 | Box Plot of Expenditures for Individuals Aged 22 to 50 by Ethnicity | 6 |
| 7 | Box Plot of Expenditures for Individuals Aged 51+ by Ethnicity | 6 |
| 8 | Bar Chart of Average Expenditures by Age Group | 7 |
| 9 | Bar Chart of Average Expenditures for Males by Age Group | 8 |
| 10 | Bar Chart of Average Expenditures for Females by Age Group | 8 |
| 11 | Bar Chart of Average Expenditures for Males by Ethnicity | 9 |
| 12 | Bar Chart of Average Expenditures for Females by Ethnicity | 9 |

1 Introduction

A few years ago, a lawsuit was filed against the California Department of Development Service (DDS) that alleged the DDS was using more of its funding to aid White, non-Hispanic individuals than for Hispanic individuals. The purpose of the California DDS is to provide funding for those who are developmentally disabled, so discrimination in funding could have life-changing consequences for those victimized.

The California DDS provided a comma-separated values (CSV) file containing demographic and funding information. That CSV file contains all of the data that was examined and presented in this paper. The CSV file lists the age group, gender, ethnicity, and expenditure amount for 1,000 different cases. The approach taken to analyzing the data provided by the California DDS includes the creation and interpretation of box plots and bar graphs that subset the data to compare funding to different groups.

2 Statement of Problem and Statistical Analysis Approach

The goal of the analysis was to determine if there was a discrepancy in funding between White and Hispanic individuals that would indicate some level of discrimination in the California DDS. The following are the plots and charts created by which the California DDS's data was analyzed:

- Box plot of expenditures by ethnicity
- Box plots for each age group showing expenditures by ethnicity
- Bar chart of average expenditures by age group
- Bar charts for each gender showing expenditures by age group
- Bar charts for each gender showing average expenditures by ethnicity

All of these figures were created using R. The *dbplyr* library in R was also used for splitting the provided CSV data into various subsets. For the bar charts that use average expenditures, the sum of each subset's expenditures was divided by the number of rows in the subset to find the average.

3 Results

3.1 Analysis of Expenditures by Ethnicity

Figure 1 shows a box plot of all 1,000 data points provided by the California DDS, organized by the ethnicity of each individual. Based purely on this plot, it would appear that there is a lack of funding for Hispanic individuals when compared to White individuals. However, it is important to note the number of outliers in the Hispanic data. In total, there are 60

outliers in the Hispanic data points. Considering the fact that 376 of the 1,000 data points are for Hispanic individuals, that many outliers is quite significant. From this plot, it is clear that the median expenditure for Hispanic individuals is noticeably lower than for White individuals.

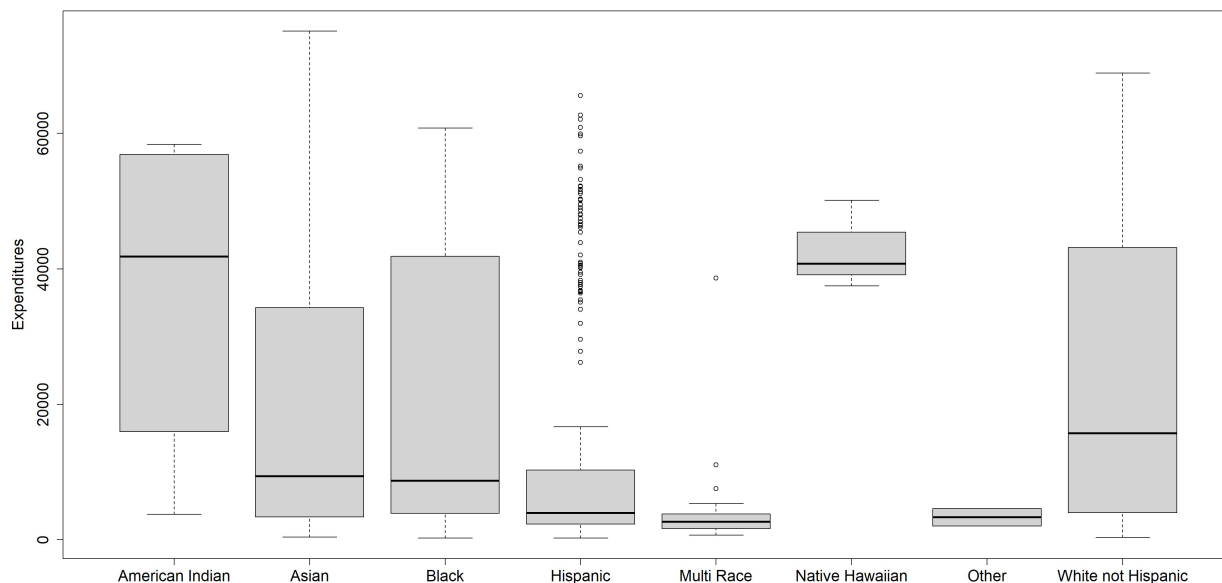


Figure 1: Box Plot of Expenditures by Ethnicity

Figures 2 through 7 show box plots for expenditures by ethnicity for each age group. These plots paint a much different picture than that found in Figure 1. It should be noted that not all of these plots contain the same ethnicities, and this is because some age groups did not have any data points containing those ethnicities. For example, Figure 5 does not contain any data for American Indian or Native Hawaiian individuals. By making subsets of the data by age group, the disparity in funding between White and Hispanic individuals seems to almost completely vanish.

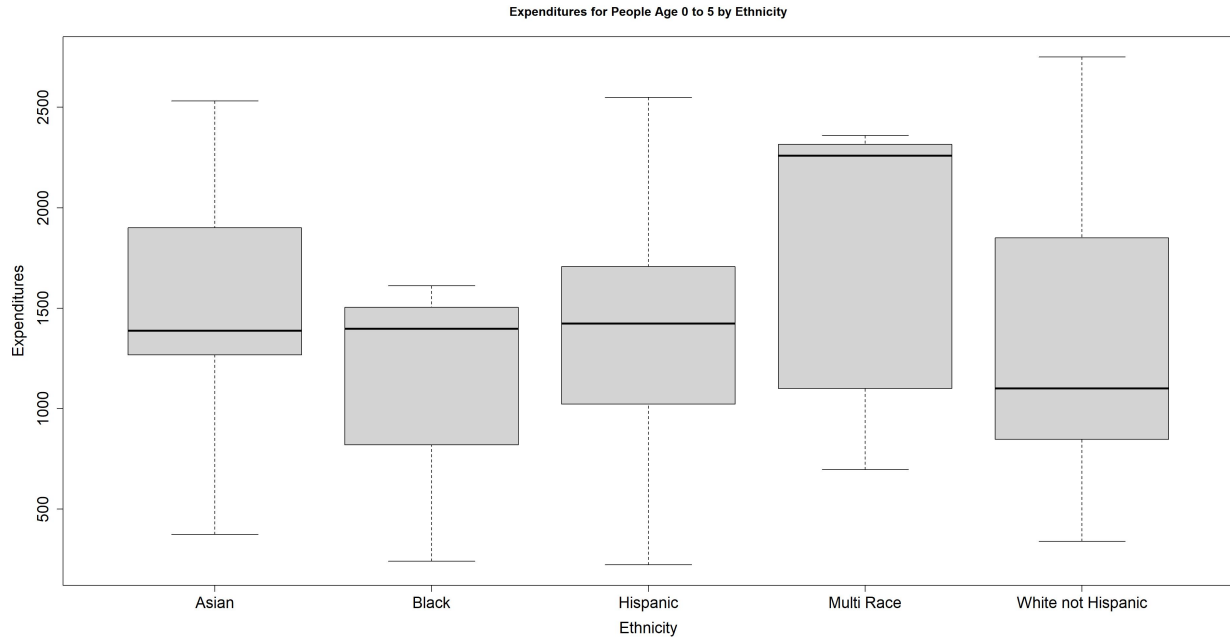


Figure 2: Box Plot of Expenditures for Individuals Aged 0 to 5 by Ethnicity

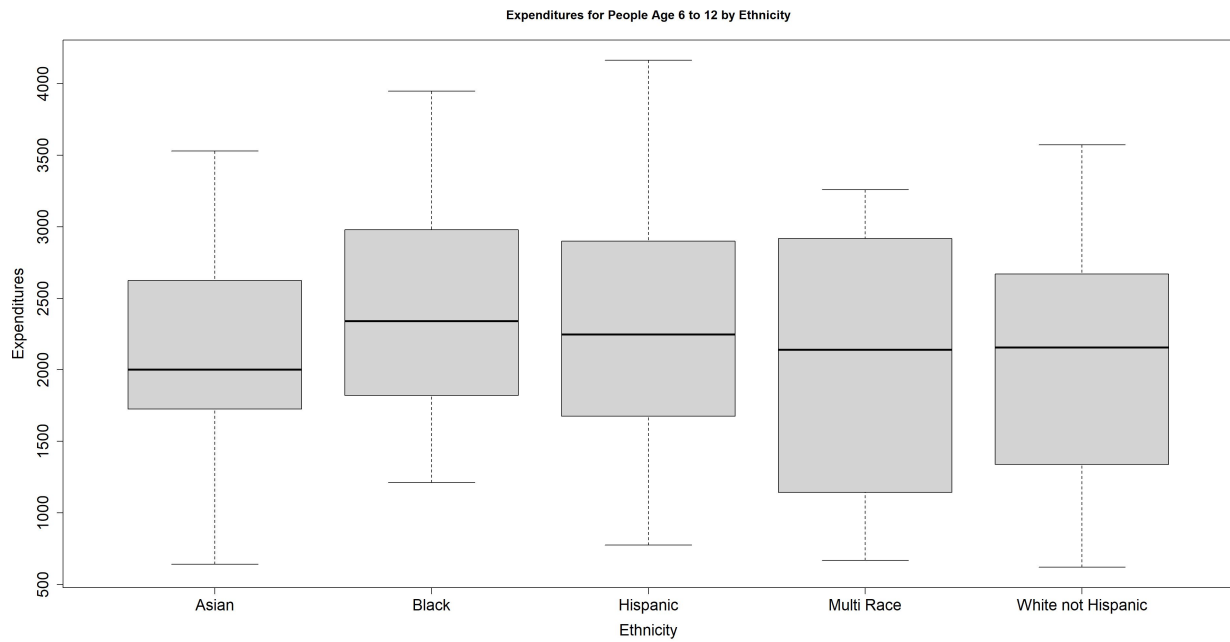


Figure 3: Box Plot of Expenditures for Individuals Aged 6 to 12 by Ethnicity

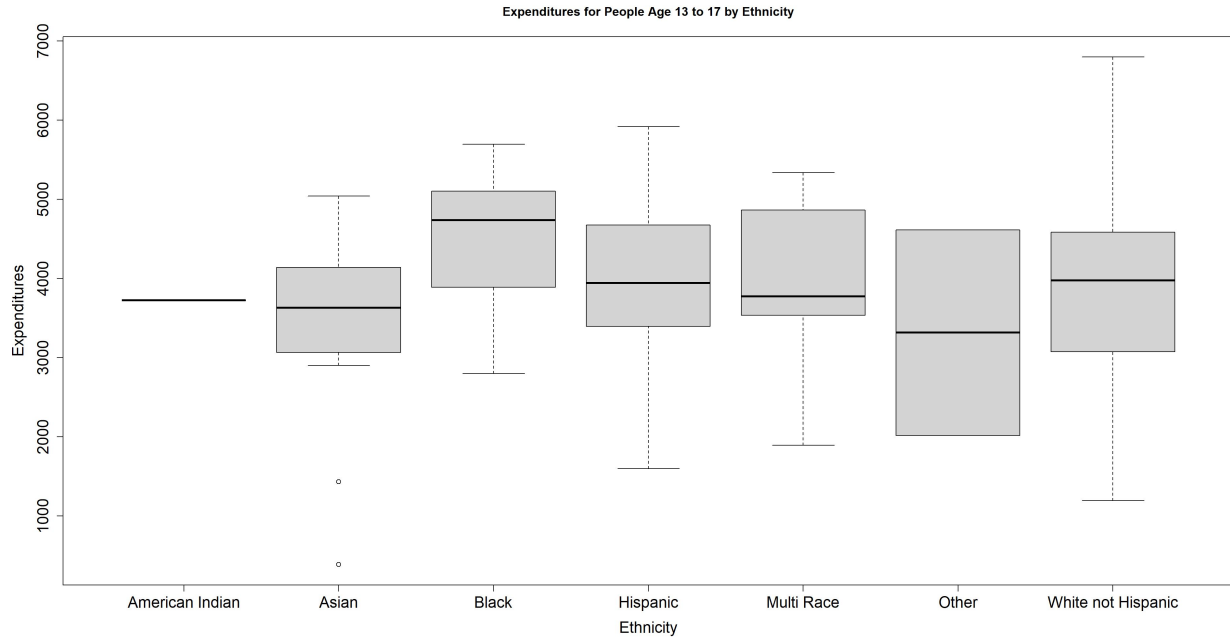


Figure 4: Box Plot of Expenditures for Individuals Aged 13 to 17 by Ethnicity

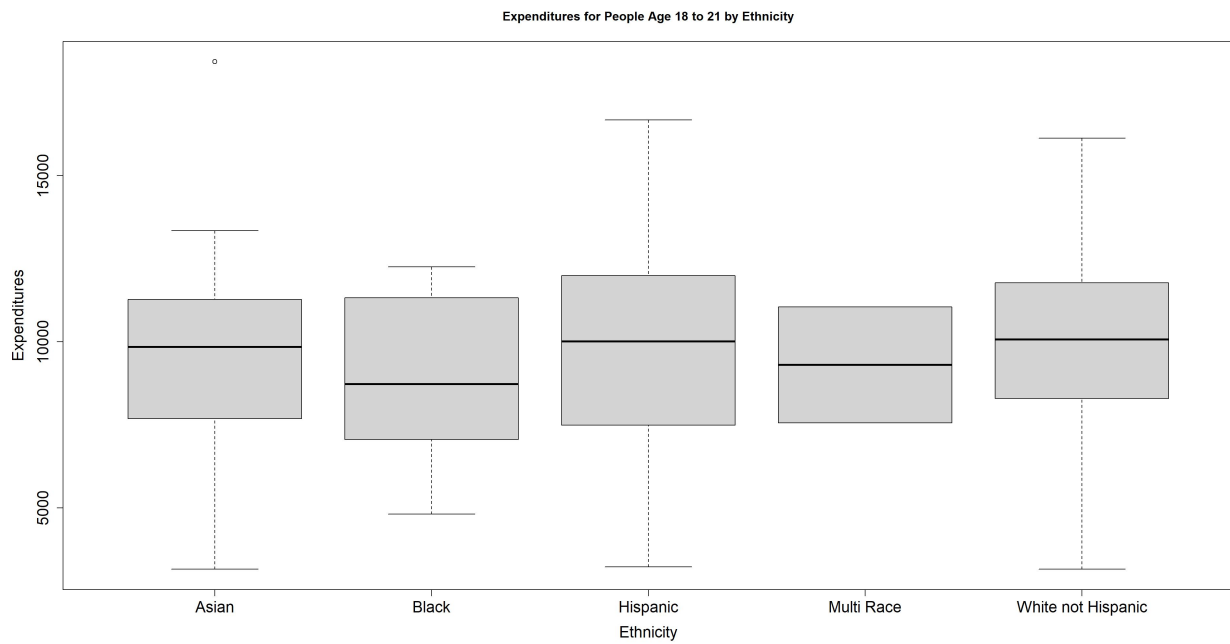


Figure 5: Box Plot of Expenditures for Individuals Aged 18 to 21 by Ethnicity

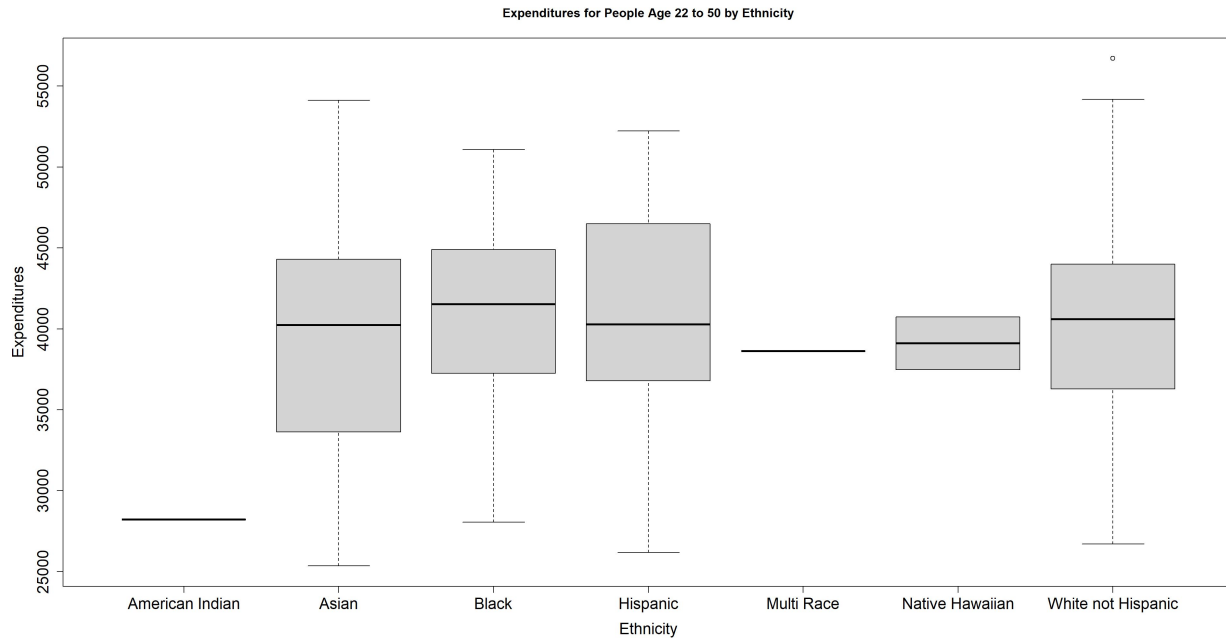


Figure 6: Box Plot of Expenditures for Individuals Aged 22 to 50 by Ethnicity

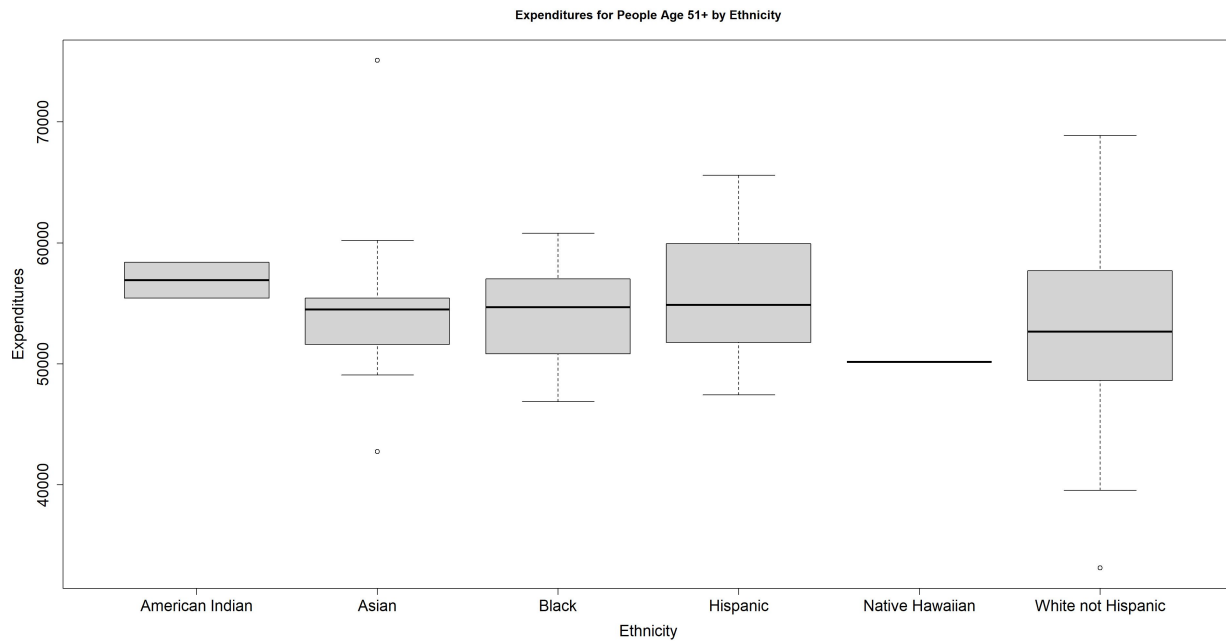


Figure 7: Box Plot of Expenditures for Individuals Aged 51+ by Ethnicity

How can the difference in funding between White and Hispanic individuals seem obvious when looking at Figure 1 but nonexistent when looking at Figures 2 through 7? The difference lies in the amount of funding each age group receives. Figure 8 shows the average expenditures for individuals based on their age group. As one's age increases, the amount of funding they are likely to receive increases dramatically.

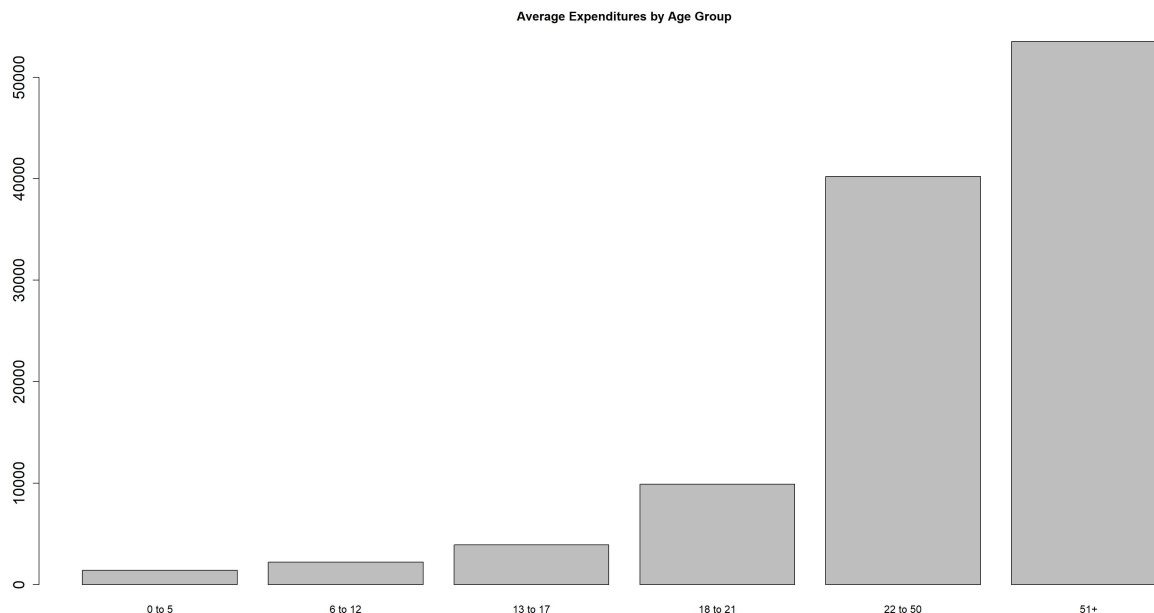


Figure 8: Bar Chart of Average Expenditures by Age Group

This difference in funding based on age group can explain the difference in funding between in Hispanic and White individuals because more of the White individuals that received funding were in the older age groups. Below is a table that lists the number of White and Hispanic individuals that were in each age group:

| Ethnicity | 0-5 | 6-12 | 13-17 | 18-21 | 22-50 | 51+ |
|------------------|------------|-------------|--------------|--------------|--------------|------------|
| Hispanic | 44 | 91 | 103 | 78 | 43 | 17 |
| White | 20 | 46 | 67 | 69 | 133 | 66 |

3.2 Analysis of Expenditures by Gender

Because of anticipation of a future lawsuit alleging funding discrimination based on gender, further research was done into the expenditures in respect to gender. Figure 9 shows a bar chart containing information about the average expenditures for males based on their age group. Similarly, Figure 10 shows a bar chart containing information about the average expenditures for females based on their age group. As discussed earlier, older age groups receive more funding than their younger counterparts. Visually, there is very little difference between the two figures. Females appear to receive slightly more funding than males, with the biggest difference being in the 51+ category, where females on average received about \$2,400 more than males. This difference means that females aged 51+ received on average about 104.6% the funding of their male counterparts.

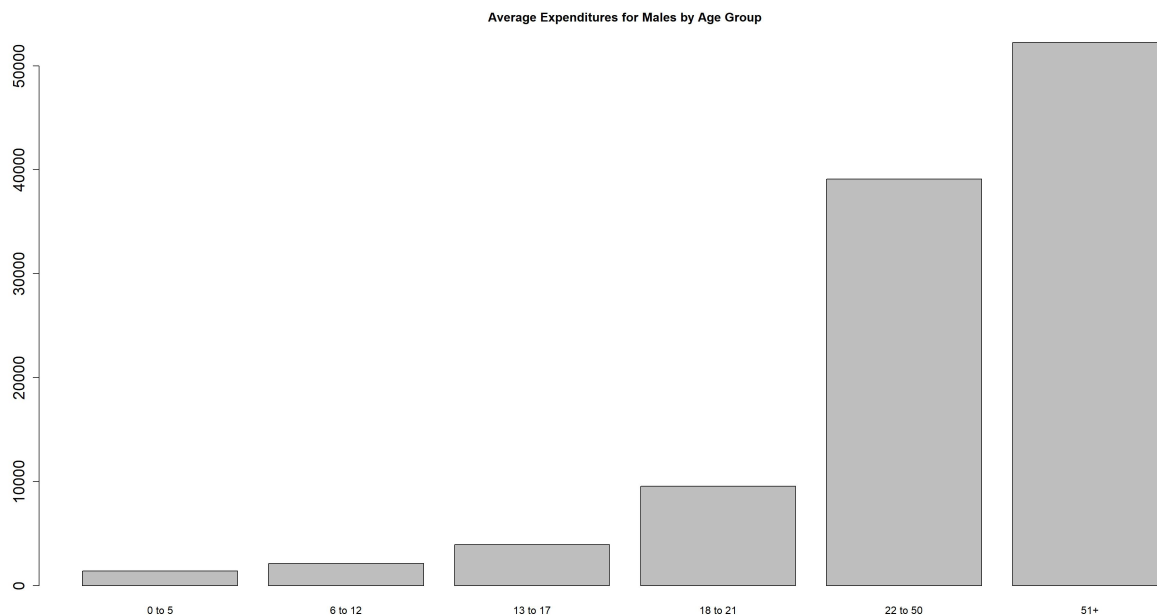


Figure 9: Bar Chart of Average Expenditures for Males by Age Group

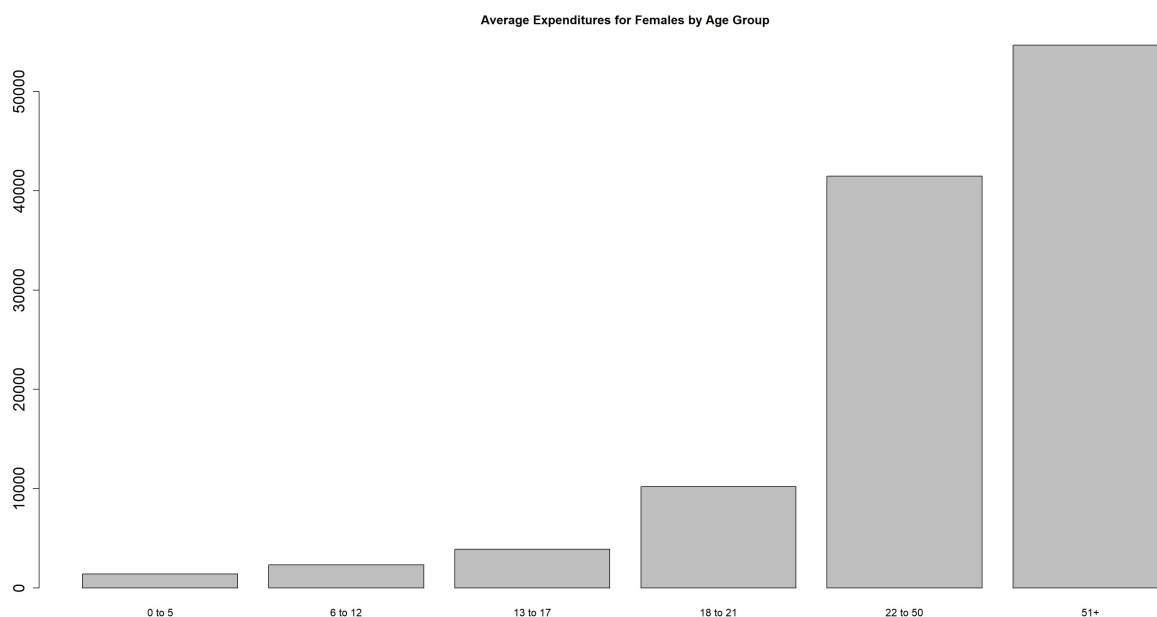


Figure 10: Bar Chart of Average Expenditures for Females by Age Group

Figures 11 and 12 organize the same data by ethnicity rather than age group. The average expenditures for American Indian and Native Hawaiian females were about \$11,000 and \$8,000 higher than males respectively. The average funding for other groups was very similar, with the exception of multi-race and Asian males, whose average expenditures were approximately \$2,000 and \$4,000 higher than their female counterparts respectively.

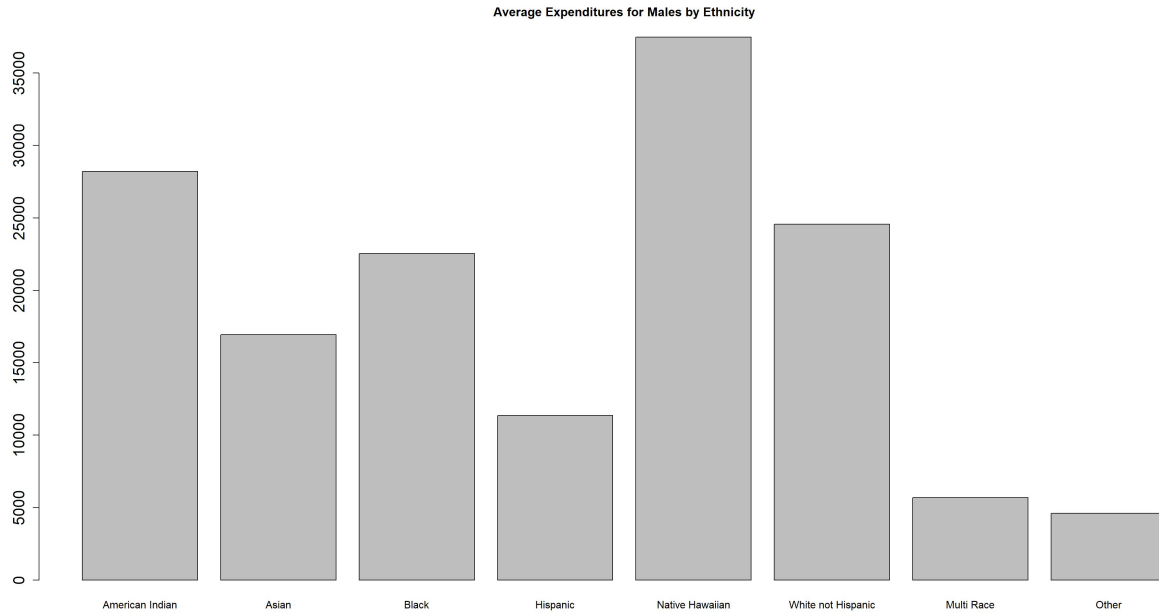


Figure 11: Bar Chart of Average Expenditures for Males by Ethnicity

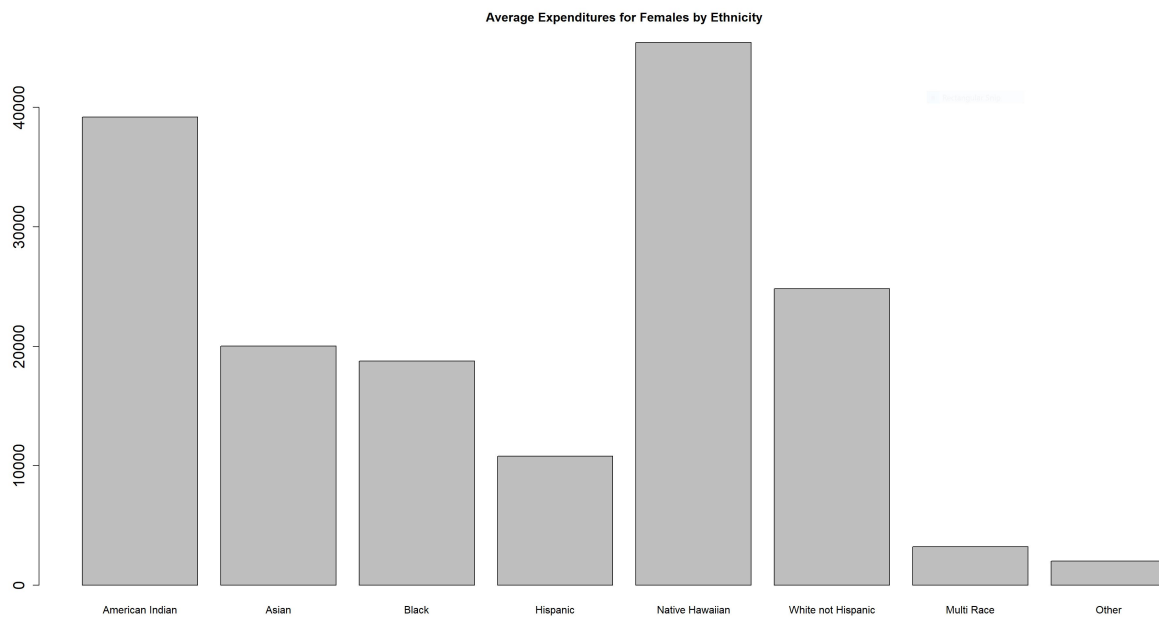


Figure 12: Bar Chart of Average Expenditures for Females by Ethnicity

4 Summary and Conclusions

Based on analysis performed through a series of box plots and bar charts, there is little evidence to suggest that the California DDS underfunded individuals because of their ethnicity or gender. In both cases, the age of the individual is significantly more likely to have a

significant impact on the amount of funding they receive as opposed to ethnicity or gender. The seeming disparity in the median expenditures for Hispanic and White individuals can be explained by the age groups the individuals fit into, with more of the White individuals being in the older age groups that received more funding. While there is some difference in funding between male and female individuals, the difference is too small for conscious, malicious discrimination to be the most likely explanation.

More information on exactly how the California DDS decides how much funding each individual receives would be needed to draw more definitive conclusions about why some differences do exist in the funding of certain individuals. The lawsuit filed against the California DDS for underfunding Hispanic individuals when compared to White individuals should be dismissed, as there is not nearly enough evidence to suggest discrimination took place. Additionally, it is unlikely a further lawsuit would be filed alleging the underfunding of males based on the provided data.

Analyzing the difference in funding between different disabilities the California DDS provides services for could be an excellent avenue for further research. Different disabilities come with unique challenges that would most likely affect the amount of funding needed to adequately aid the individuals affected. This difference in funding based on the individual's condition could also explain some of the variation in funding between individuals seen in the data examined in this paper.

5 Sources

California DDS Expenditure Data CSV:

https://una.instructure.com/courses/91813/files/15725119?module_item_id=2091895

6 Appendix: R Code

```
1  # MA-345 Project Code
2  # James Palmer
3  # November 30, 2024
4
5  library(dbplyr)
6  library(dplyr)
7  input_data = read.csv("datasets/California_DDS_Expenditures.csv")
8  # View(input_data)
9
10 # QUESTION 1
11 boxplot1 = boxplot(input_data$Expenditures~input_data$Ethnicity,
12                   horizontal=F, xlab="Ethnicity",
13                   ylab="Expenditures", cex.lab=1.5, cex.axis=1.5)
14 boxplot1$out
15 boxplot1$group
16 boxplot1$names
17 length(boxplot1$group) - 3
```

```

18 nrow(filter(input_data, Ethnicity=="Hispanic"))
19
20 # QUESTION2
21 unique(input_data$Age.Group)
22 age_group_1 = filter(input_data, Age.Group=="0 to 5")
23 age_group_2 = filter(input_data, Age.Group=="6 to 12")
24 age_group_3 = filter(input_data, Age.Group=="13 to 17")
25 age_group_4 = filter(input_data, Age.Group=="18 to 21")
26 age_group_5 = filter(input_data, Age.Group=="22 to 50")
27 age_group_6 = filter(input_data, Age.Group=="51+")
28 boxplot(age_group_1$Expenditures~age_group_1$Ethnicity,
29         horizontal=F, xlab="Ethnicity", ylab="Expenditures",
30         main="Expenditures for People Age 0 to 5 by Ethnicity",
31         cex.lab=1.5, cex.axis=1.5)
32 boxplot(age_group_2$Expenditures~age_group_2$Ethnicity,
33         horizontal=F, xlab="Ethnicity", ylab="Expenditures",
34         main="Expenditures for People Age 6 to 12 by Ethnicity",
35         cex.lab=1.5, cex.axis=1.5)
36 boxplot(age_group_3$Expenditures~age_group_3$Ethnicity,
37         horizontal=F, xlab="Ethnicity", ylab="Expenditures",
38         main="Expenditures for People Age 13 to 17 by Ethnicity",
39         cex.lab=1.5, cex.axis=1.5)
40 boxplot(age_group_4$Expenditures~age_group_4$Ethnicity,
41         horizontal=F, xlab="Ethnicity", ylab="Expenditures",
42         main="Expenditures for People Age 18 to 21 by Ethnicity",
43         cex.lab=1.5, cex.axis=1.5)
44 boxplot(age_group_5$Expenditures~age_group_5$Ethnicity,
45         horizontal=F, xlab="Ethnicity", ylab="Expenditures",
46         main="Expenditures for People Age 22 to 50 by Ethnicity",
47         cex.lab=1.5, cex.axis=1.5)
48 boxplot(age_group_6$Expenditures~age_group_6$Ethnicity,
49         horizontal=F, xlab="Ethnicity", ylab="Expenditures",
50         main="Expenditures for People Age 51+ by Ethnicity",
51         cex.lab=1.5, cex.axis=1.5)
52
53 # QUESTION 3
54 male_data = filter(input_data, Gender=="Male")
55 female_data = filter(input_data, Gender=="Female")
56 # PART A
57 age_groups = c("0 to 5", "6 to 12", "13 to 17",
58               "18 to 21", "22 to 50", "51+")
59 male_0to5 = filter(male_data, Age.Group=="0 to 5")
60 male_6to12 = filter(male_data, Age.Group=="6 to 12")
61 male_13to17 = filter(male_data, Age.Group=="13 to 17")
62 male_18to21 = filter(male_data, Age.Group=="18 to 21")
63 male_22to50 = filter(male_data, Age.Group=="22 to 50")
64 male_51plus = filter(male_data, Age.Group=="51+")
65
66 male_expenditures =
67   c((sum(male_0to5$Expenditures)/nrow(male_0to5)),
68     (sum(male_6to12$Expenditures)/nrow(male_6to12)),
69     (sum(male_13to17$Expenditures)/nrow(male_13to17)),
70     (sum(male_18to21$Expenditures)/nrow(male_18to21)),

```

```

71     (sum(male_22to50$Expenditures)/nrow(male_22to50)),
72     (sum(male_51plus$Expenditures)/nrow(male_51plus)))
73 male_expenditures
74
75 barplot(male_expenditures, names.arg=age_groups,
76         main="Average Expenditures for Males by Age Group",
77         cex.lab=1.5, cex.axis=1.5)
78
79 female_0to5 = filter(female_data, Age.Group=="0 to 5")
80 female_6to12 = filter(female_data, Age.Group=="6 to 12")
81 female_13to17 = filter(female_data, Age.Group=="13 to 17")
82 female_18to21 = filter(female_data, Age.Group=="18 to 21")
83 female_22to50 = filter(female_data, Age.Group=="22 to 50")
84 female_51plus = filter(female_data, Age.Group=="51+")
85
86 female_expenditures =
87   c((sum(female_0to5$Expenditures)/nrow(female_0to5)),
88     (sum(female_6to12$Expenditures)/nrow(female_6to12)),
89     (sum(female_13to17$Expenditures)/nrow(female_13to17)),
90     (sum(female_18to21$Expenditures)/nrow(female_18to21)),
91     (sum(female_22to50$Expenditures)/nrow(female_22to50)),
92     (sum(female_51plus$Expenditures)/nrow(female_51plus)))
93 female_expenditures
94
95 barplot(female_expenditures, names.arg=age_groups,
96         main="Average Expenditures for Females by Age Group",
97         cex.lab=1.5, cex.axis=1.5)
98
99 #PART B
100 unique(input_data$Ethnicity)
101 ethnicities = c("American Indian", "Asian", "Black", "Hispanic",
102               "Native Hawaiian", "White not Hispanic",
103               "Multi Race", "Other")
104 male_native_am = filter(male_data, Ethnicity=="American Indian")
105 male_asian = filter(male_data, Ethnicity=="Asian")
106 male_black = filter(male_data, Ethnicity=="Black")
107 male_hispanic = filter(male_data, Ethnicity=="Hispanic")
108 male_hawaiian = filter(male_data, Ethnicity=="Native Hawaiian")
109 male_white = filter(male_data, Ethnicity=="White not Hispanic")
110 male_multi = filter(male_data, Ethnicity=="Multi Race")
111 male_other = filter(male_data, Ethnicity=="Other")
112
113 male_race_expenditures =
114   c((sum(male_native_am$Expenditures)/nrow(male_native_am)),
115     (sum(male_asian$Expenditures)/nrow(male_asian)),
116     (sum(male_black$Expenditures)/nrow(male_black)),
117     (sum(male_hispanic$Expenditures)/nrow(male_hispanic)),
118     (sum(male_hawaiian$Expenditures)/nrow(male_hawaiian)),
119     (sum(male_white$Expenditures)/nrow(male_white)),
120     (sum(male_multi$Expenditures)/nrow(male_multi)),
121     (sum(male_other$Expenditures)/nrow(male_other)))
122 sort(male_race_expenditures)
123

```

```

124 barplot(male_race_expenditures, names.arg=ethnicities,
125         main="Average Expenditures for Males by Ethnicity",
126         cex.lab=1.5, cex.axis=1.5)
127
128 female_native_am = filter(female_data, Ethnicity=="American Indian")
129 female_asian = filter(female_data, Ethnicity=="Asian")
130 female_black = filter(female_data, Ethnicity=="Black")
131 female_hispanic = filter(female_data, Ethnicity=="Hispanic")
132 female_hawaiian = filter(female_data, Ethnicity=="Native Hawaiian")
133 female_white = filter(female_data, Ethnicity=="White not Hispanic")
134 female_multi = filter(female_data, Ethnicity=="Multi Race")
135 female_other = filter(female_data, Ethnicity=="Other")
136
137 female_race_expenditures =
138   c((sum(female_native_am$Expenditures)/nrow(female_native_am)),
139     (sum(female_asian$Expenditures)/nrow(female_asian)),
140     (sum(female_black$Expenditures)/nrow(female_black)),
141     (sum(female_hispanic$Expenditures)/nrow(female_hispanic)),
142     (sum(female_hawaiian$Expenditures)/nrow(female_hawaiian)),
143     (sum(female_white$Expenditures)/nrow(female_white)),
144     (sum(female_multi$Expenditures)/nrow(female_multi)),
145     (sum(female_other$Expenditures)/nrow(female_other)))
146 sort(female_race_expenditures)
147
148 barplot(female_race_expenditures, names.arg=ethnicities,
149         main="Average Expenditures for Females by Ethnicity",
150         cex.lab=1.5, cex.axis=1.5)
151
152 # EXTRA CODE FOR FURTHER ANALYSIS
153 age0to5 = filter(input_data, Age.Group=="0 to 5")
154 age6to12 = filter(input_data, Age.Group=="6 to 12")
155 age13to17 = filter(input_data, Age.Group=="13 to 17")
156 age18to21 = filter(input_data, Age.Group=="18 to 21")
157 age22to50 = filter(input_data, Age.Group=="22 to 50")
158 age51plus = filter(input_data, Age.Group=="51+")
159 age_expend = c(mean(age0to5$Expenditures),
160               mean(age6to12$Expenditures),
161               mean(age13to17$Expenditures),
162               mean(age18to21$Expenditures),
163               mean(age22to50$Expenditures),
164               mean(age51plus$Expenditures))
165 age_expend
166 nrow(age0to5)
167 nrow(age6to12)
168 nrow(age13to17)
169 nrow(age18to21)
170 nrow(age22to50)
171 nrow(age51plus)
172
173 age_groups = c("0 to 5", "6 to 12", "13 to 17",
174               "18 to 21", "22 to 50", "51+")
175 barplot(age_expend, names.arg=age_groups,
176         main="Average Expenditures by Age Group",

```

```

177         cex.lab=1.5, cex.axis=1.5)
178
179 hisp_data = filter(input_data, Ethnicity=="Hispanic")
180 hisp_0to5 = filter(hisp_data, Age.Group=="0 to 5")
181 hisp_6to12 = filter(hisp_data, Age.Group=="6 to 12")
182 hisp_13to17 = filter(hisp_data, Age.Group=="13 to 17")
183 hisp_18to21 = filter(hisp_data, Age.Group=="18 to 21")
184 hisp_22to50 = filter(hisp_data, Age.Group=="22 to 50")
185 hisp_51plus = filter(hisp_data, Age.Group=="51+")
186 nrow(hisp_0to5)
187 nrow(hisp_6to12)
188 nrow(hisp_13to17)
189 nrow(hisp_18to21)
190 nrow(hisp_22to50)
191 nrow(hisp_51plus)
192
193 white_data = filter(input_data, Ethnicity=="White not Hispanic")
194 white_0to5 = filter(white_data, Age.Group=="0 to 5")
195 white_6to12 = filter(white_data, Age.Group=="6 to 12")
196 white_13to17 = filter(white_data, Age.Group=="13 to 17")
197 white_18to21 = filter(white_data, Age.Group=="18 to 21")
198 white_22to50 = filter(white_data, Age.Group=="22 to 50")
199 white_51plus = filter(white_data, Age.Group=="51+")
200 nrow(white_0to5)
201 nrow(white_6to12)
202 nrow(white_13to17)
203 nrow(white_18to21)
204 nrow(white_22to50)
205 nrow(white_51plus)

```