

# **California Department of Development Service (DDS) Funding Analysis**

Montana Merkle | University of North Alabama | 11/30/2022



## **Abstract:**

This paper will analyze data from the California Department of Development Service (DDS) on their funding based on race and gender to determine if the lawsuit against them is legitimate or not. The main objectives of this project are to come to a conclusion on if the DDS has provided more funding to White Non-Hispanics than Hispanics, and if they have possibly discriminated based on gender as well. Throughout this project, data sampling and the use of data subsets will be utilized to compare funding based on different categories in the form of box plots and bar charts. Two important results were found; the boxplots of expenditures by ethnicity for each age group were alike when looking at Hispanic and Non-Hispanics and the mean of female and male expenditures only differed by \$129. Due to these results, I have come to the conclusion that the lawsuit against the California DDS was not legitimate and the organization has also not committed gender discrimination.

**Keywords:** California Department of Development Service Discrimination Lawsuit, Racial Discrimination, Gender Bias, Data Sampling, Data Subsets

## Table of Contents

<b>Section 1: Introduction .....</b>	<b>3</b>
<b>Section 2: Statement of Problem and Statistical Analysis Approach .....</b>	<b>3</b>
Description of Statistical Methods .....	4
Figure 1 - Expenditures by Ethnicity .....	4
<b>Section 3: Results .....</b>	<b>4</b>
Analysis Findings .....	4
Figure 2 - Expenditures by Ethnicity Age Group 0-5 .....	5
<b>Section 4: Summary &amp; Conclusion .....</b>	<b>5</b>
Conclusion .....	6
Limitations of Analysis .....	6
Subjects for Further Study .....	6
<b>References .....</b>	<b>7</b>
<b>Appendices .....</b>	<b>7</b>
Tables and Figures .....	7
Computer Listings .....	13

## **Section 1: Introduction**

A lawsuit was filed against the California Department of Development Service (DDS) on the basis that the California DDS was providing more funding towards White Non-Hispanics than Hispanics. A file of the California DDS Expenditures was given as data to come to a conclusion on this lawsuit. This problem is important because the organization provides funding for developmentally disabled individuals within the state. Funding should not be distributed in a discriminatory fashion, since everyone deserves to be given funding in an equal manner, not determined by race or gender. On top of this, if there is significant evidence to support the lawsuit, then the organization could face serious legal and financial ramifications. To come to a conclusion on if the California DDS is providing less funding to Hispanics, I will utilize data sampling to analyze a subset of the data and potentially identify patterns and trends useful to prove/disprove the claim. In anticipation of a future lawsuit about gender discrimination, I will also employ the use of bar charts to help visually discern trends of categorical data (male and female) to come to a conclusion on this possible future problem.

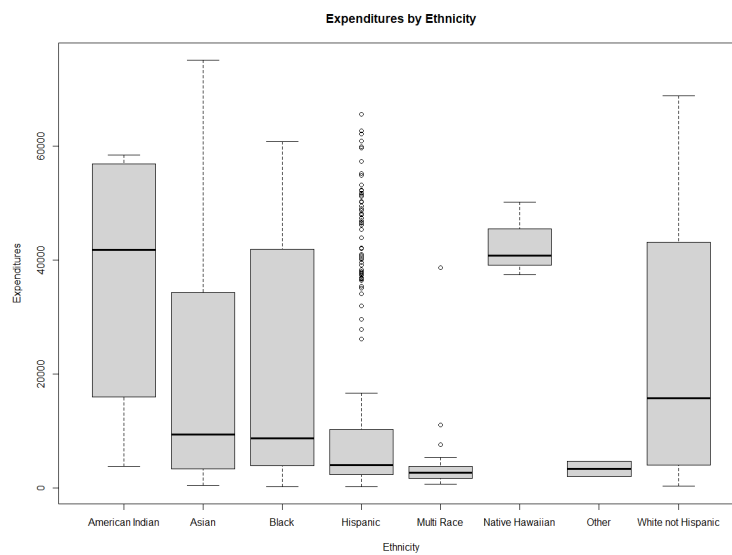
## **Section 2: Statement of Problem and Statistical Analysis**

### **Approach**

The main statistical computing program I used to come to a conclusion on the discrimination lawsuit was RStudio which was useful in developing visual aids. I was able to create multiple figures that displayed the data after watching a number of videos on how to properly utilize the software. The first figure I created was a boxplot of expenditures by ethnicity (Figure 1, displayed below), and then went on to subset the data by age group to take a closer look at the distribution of expenditures by ethnicity (Figures 2 through 7). To further my investigation I also utilized the summary tool in RStudio which allowed me to easily compare between Hispanic and White Not-Hispanic expenditures through their median values. The next step was to look into the possible future gender discrimination lawsuits, so I first subsetted the data by gender and then

created bar charts of expenditures by age group for females and males with the ggplot2 library (Figures 8 and 9). With this subsetting data I then also created bar charts of expenditures by ethnicity of each gender (Figures 10 and 11). The code I used to create all of the figures is found in the appendices section of the report.

**Figure 1 - Expenditures by Ethnicity**

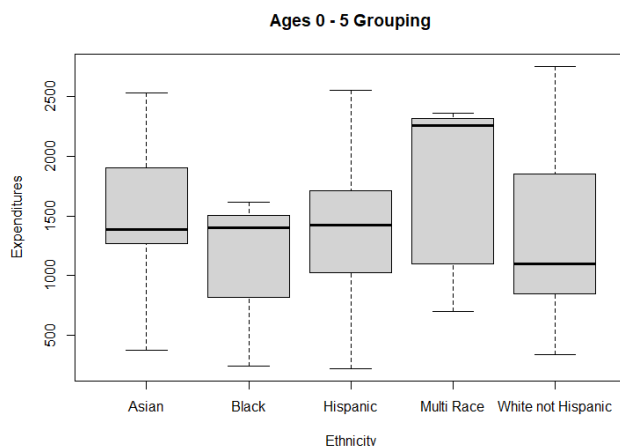


## Section 3: Results

At first glance, it seems like the box plot displaying the expenditures by ethnicity (Figure 1) supports the claim that the California DDS supplies less funding towards Hispanics than White Non-Hispanics by a large difference, but the graphic might be misleading since it has a high number of outliers present (circles) on the Hispanic ethnicity grouping. Due to the presence of outliers, the median is more useful than the mean in comparing the distribution of expenditures because it is less skewed by outliers. The White Non-Hispanic group had a much higher IQR, \$39,157, compared to the Hispanic group, \$7,961 (from the summary of Hispanic data). A higher IQR for the White Non-Hispanic group expenditures shows that the spread of the middle fifty percent of data varies a lot more from the median (\$15,718) than the Hispanic group expenditures (which has a median of \$3,952). Although after subsetting the data by age groupings, the medians and boxplots (Figures 2 displayed below and Figures 3 through

7) were very similar and contained a higher median of expenditures for the Hispanic grouping, surpassing that of the White Non-Hispanic, for ages 6 through 12, 0 through 5, and 50 plus. After examining all the data at my disposal I have come to the conclusion that the California DDS is not discriminately providing more funds to White Non-Hispanics than Hispanics.

**Figure 2 - Expenditures by Ethnicity Age Group 0-5**



Lastly, I analyzed the data to determine if gender discrimination was present in the expenditures. The bar charts of the female and male expenditures by age groups (Figures 8 and 9) and the charts of female and male expenditures by ethnicity (Figures 10 and 11) were very similar. On top of this, the mean of female and male expenditures only differed by \$129 (from the summary of male and female data). These facts suggest that the California DDS is not discriminating based on gender if there is a lawsuit in the future based on this claim.

## Section 4: Summary and Conclusion

A discrimination lawsuit was filed against the California DDS claiming that White Non-Hispanic was receiving more funding than Hispanics. I attempted to come to a conclusion on the validity of the claim and analyze the expenditures to determine if the organization was discriminating based on gender in anticipation of future lawsuits as

well. I utilized RStudio to create figures that helped me visually compare the money received by the White Non-Hispanic and Hispanic groups in the form of bar charts and boxplots. On top of this, I examined the differences between the mean and median of the data groupings (male, female, Hispanic, and White Non-Hispanic). I found that while the boxplots of expenditures by ethnicity (Figure 1) pointed towards discrimination, the boxplots of expenditures by ethnicity subsetted by age group as well as the IQR and median values of the data provided enough evidence that the organization's lawsuit was invalid. I came to this conclusion because multiple boxplots subsetted by age had a higher median for Hispanics than for White Non-Hispanics in addition to their similarity as a whole. I believe the results derived from the figures I created were imperative to drawing a conclusion on the lawsuit and helping to make sure all ethnicities were able to receive equal funding from the organization. Although, there were some limitations of my data analysis method as my choice in utilizing boxplots did not allow me to notice major skews in data or the shape of the distribution.

Another objective of this report was to determine if the California DDS was discriminating based on gender in anticipation of future lawsuits. My approach to determining if there was gender discrimination was to subset the data by gender and create bar charts by both expenditures by ethnicity and by age group for males and females (Figures 8 through 11) on top of analyzing the mean of the groups. The bar charts were extremely similar and also had means that differed by only \$129 which led me to the conclusion that there was no gender discrimination present in the data. These results are important to both the California DDS and those receiving funding and could possibly help the organization protect itself from lawsuits in the future.

In the time ahead a subject that would be beneficial to study could be the correlation between the number of children in families receiving funding and the amount received. When there are older children in the families do they receive more funding? Which ethnicity has the highest number of families that receive support? Is there discrimination towards households with single mothers?

## References

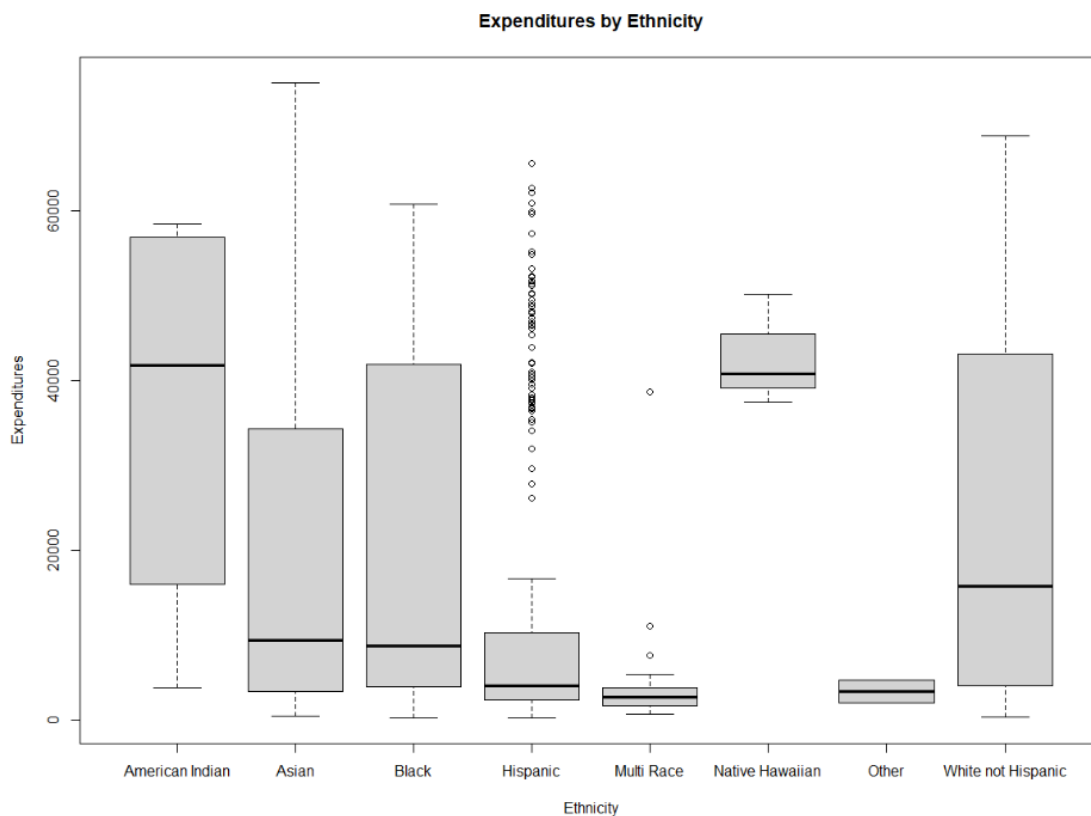
“Box Plot (Box and Whiskers): How to Read One & How to Make One in Excel, TI-83, SPSS.” Statistics How To, 2 Aug. 2022, <https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/box-plot/>.

“Boxplots in R | Creating Box and Whisker Plots in R.” YouTube, YouTube, 18 Feb. 2016, <https://www.youtube.com/watch?v=KQfym04mIDY>.

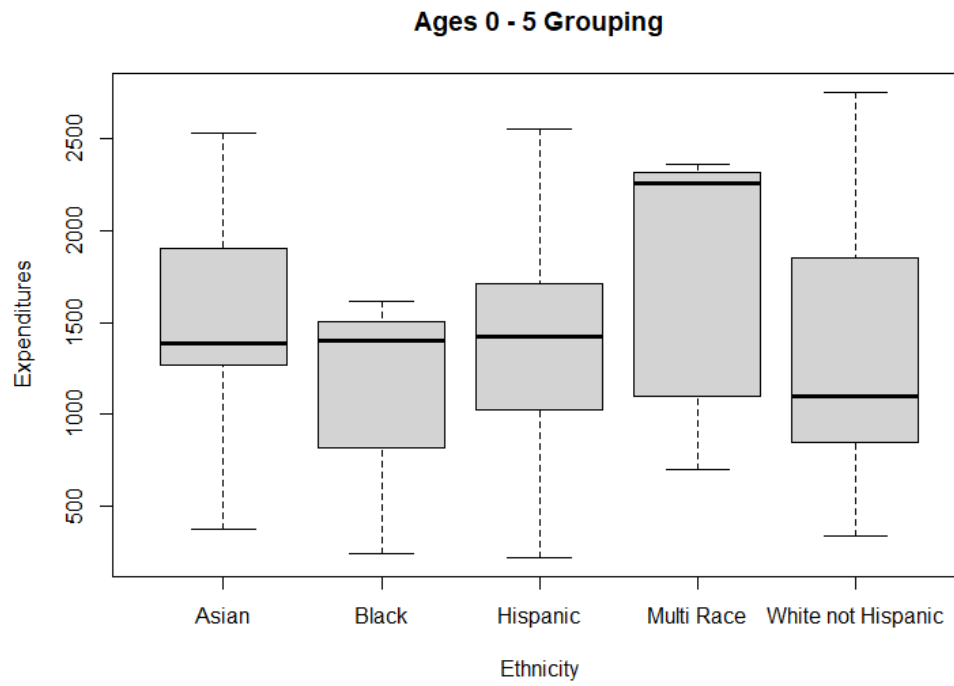
Holtz, Yan. “Basic Barplot with GGPLOT2.” – The R Graph Gallery, <https://r-graph-gallery.com/218-basic-barplots-with-ggplot2.html>.

## Appendices: Table and Figures

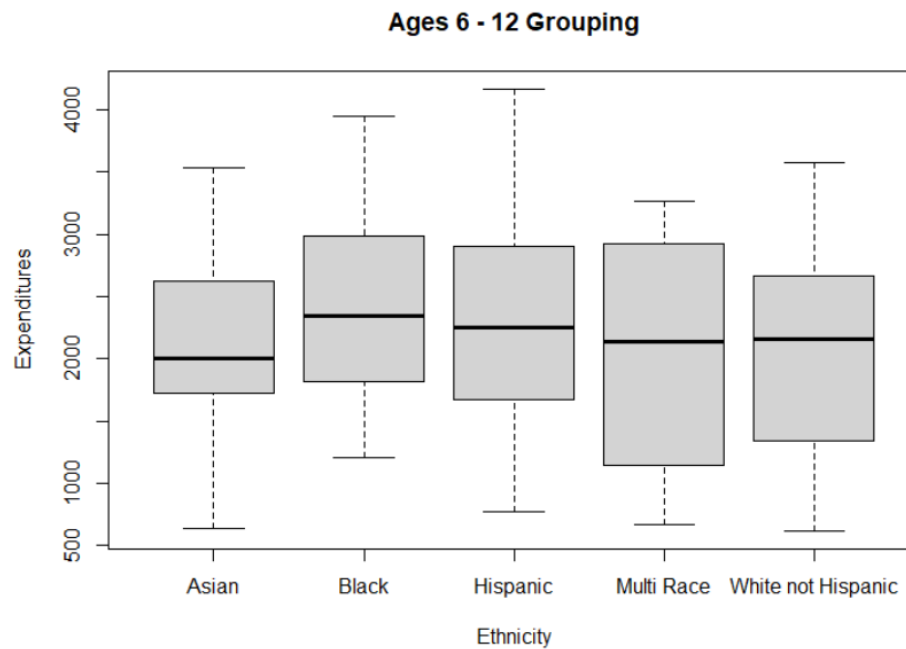
**Figure 1 - Expenditures by Ethnicity**



**Figure 2 - Expenditures by Ethnicity Age Group 0-5**

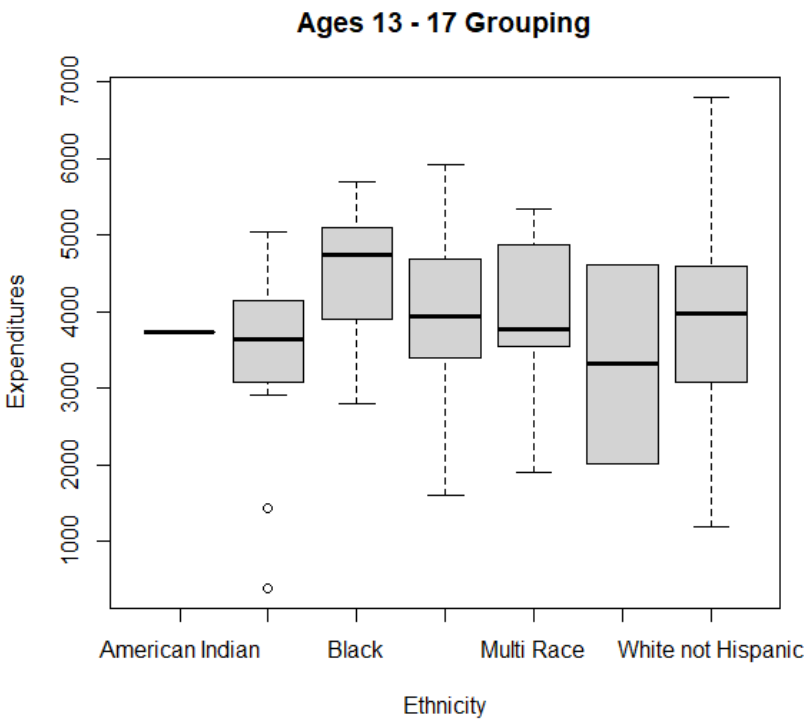


**Figure 3 - Expenditures by Ethnicity Age Group 6-12**

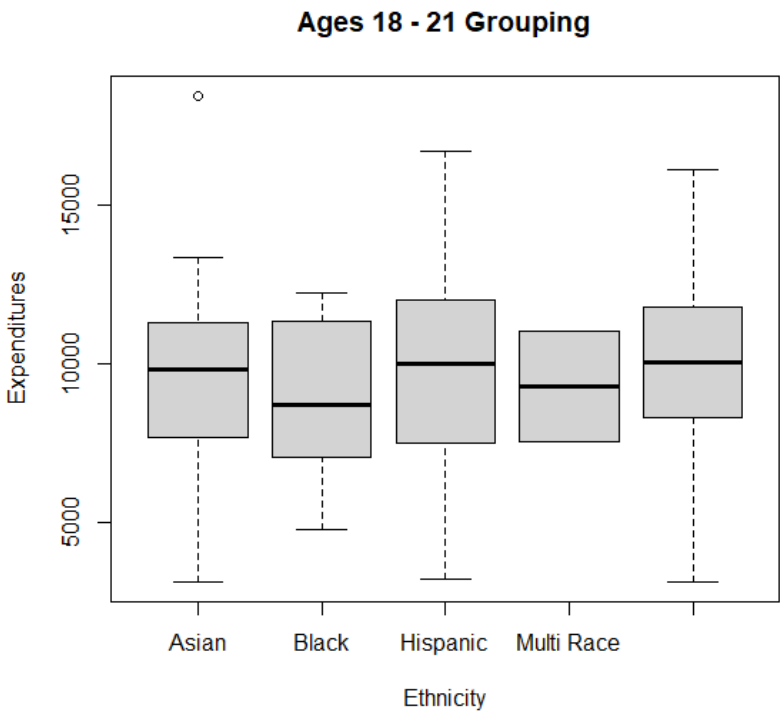




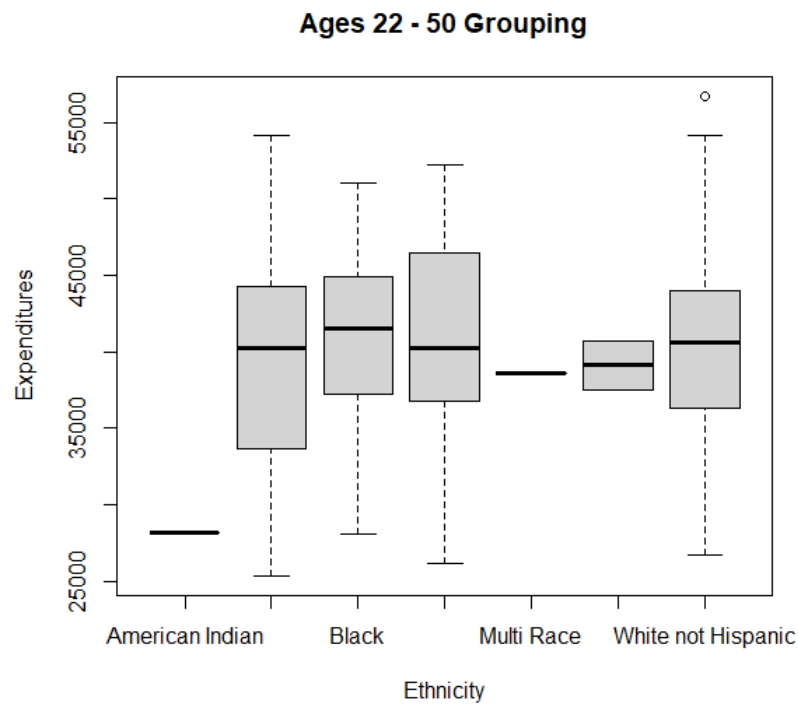
**Figure 4 - Expenditures by Ethnicity Age Group 13-17**



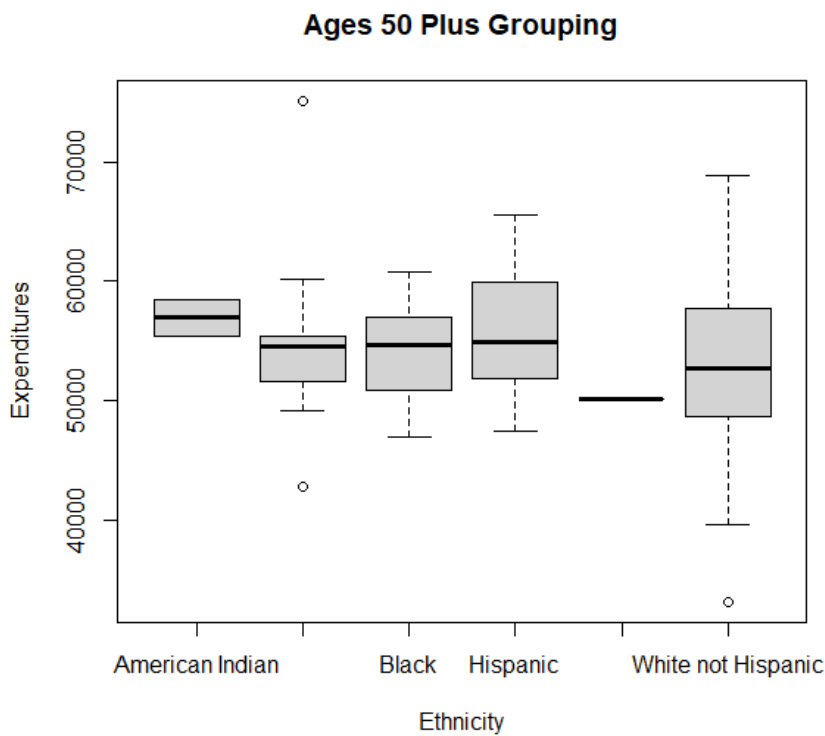
**Figure 5 - Expenditures by Ethnicity Age Group 18-21**



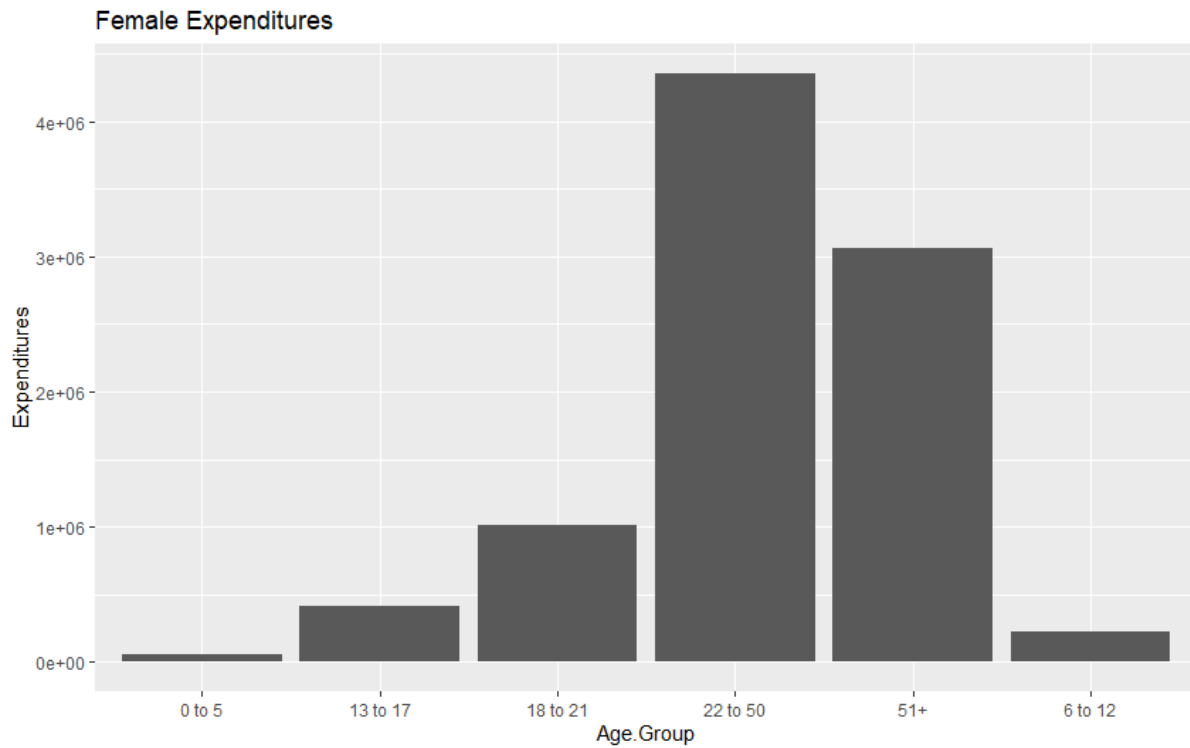
**Figure 6 - Expenditures by Ethnicity Age Group 22-50**



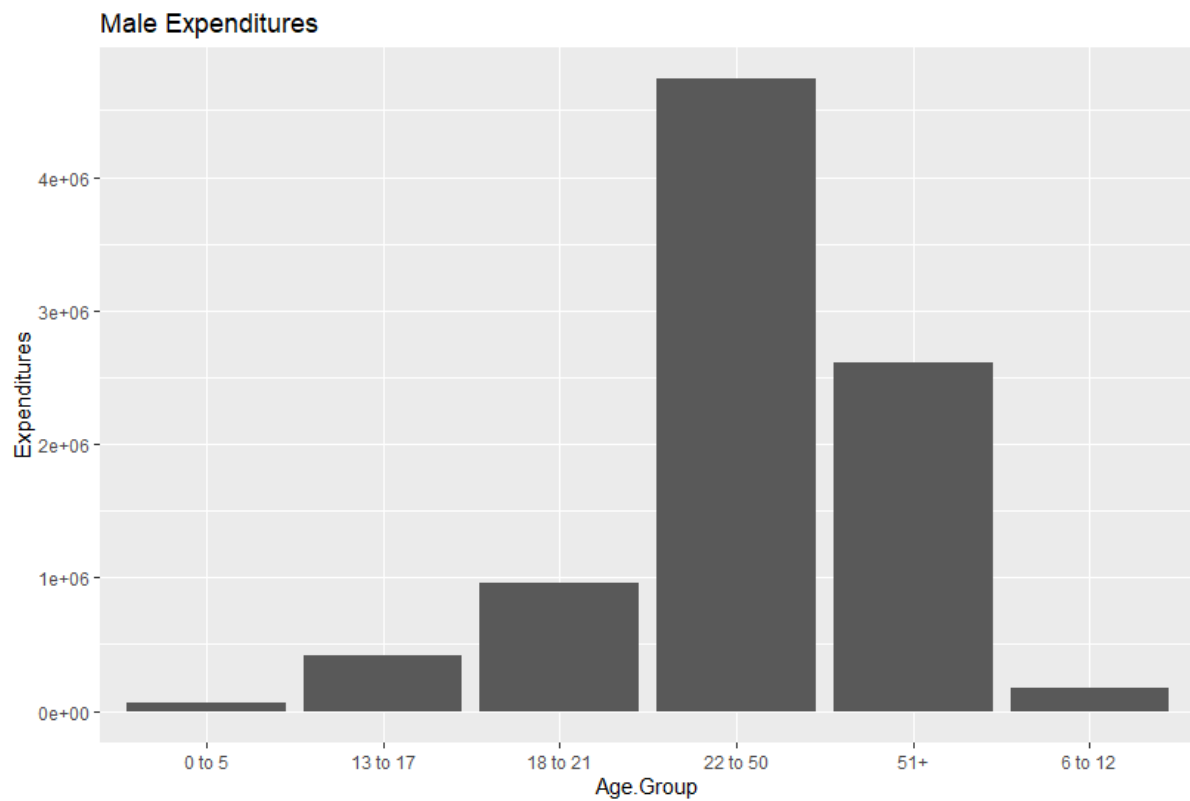
**Figure 7 - Expenditures by Ethnicity Age Group 50+**



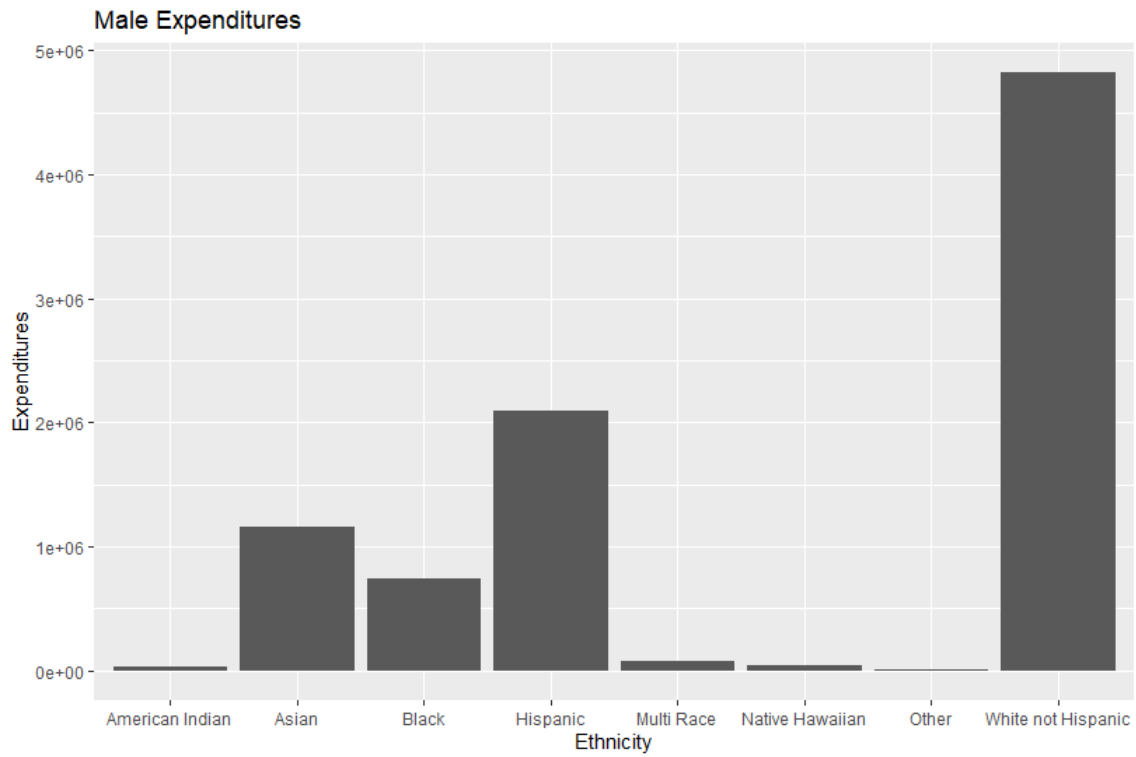
**Figure 8 - Expenditures of Female Recipients by Age Groups**



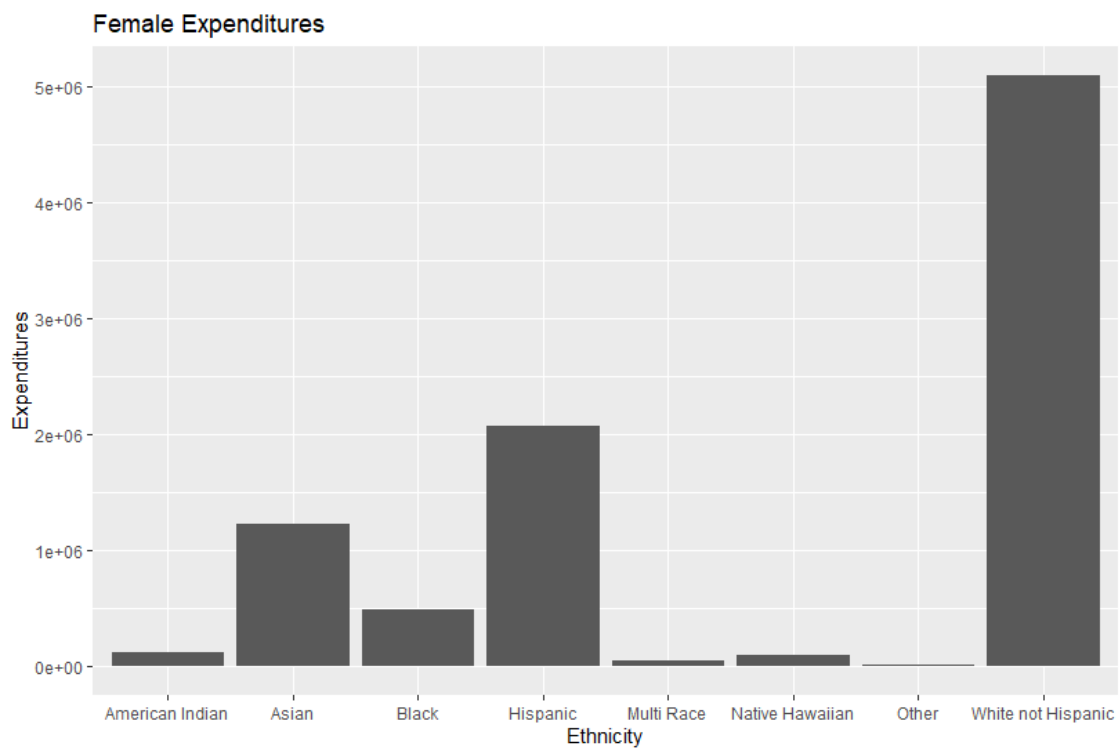
**Figure 9 - Expenditures of Male Recipients by Age Groups**



**Figure 10 - Expenditures of Male Recipients by Ethnicity**



**Figure 11 - Expenditures of Female Recipients by Ethnicity**



## Appendices: Computer Listings

### Barchart of Expenditures of Male Recipients by Ethnicity

```
Boxplot Expenditures by Ethnicity.R* x female x male x data x
1 library(readr)
2 library(ggplot2)
3
4 data <- read_csv("California_DDS_Expenditures.csv")
5
6 ggplot(male, aes(x = Ethnicity, y = Expenditures)) + geom_bar(stat = "identity") + ggtitle("Male Expenditures")
7
8
```

### Barchart of Expenditures of Female Recipients by Ethnicity

```
Boxplot Expenditures by Ethnicity.R* x female x male x data x
1 library(readr)
2 library(ggplot2)
3
4 data <- read_csv("California_DDS_Expenditures.csv")
5
6 ggplot(female, aes(x = Ethnicity, y = Expenditures)) + geom_bar(stat = "identity") + ggtitle("Female Expenditures")
7
8
```

### Barchart of Expenditures of Male Recipients by Age Groups

```
Boxplot Expenditures by Ethnicity.R* x female x male x data x
1 library(ggplot2)
2 library(readr)
3
4 data <- read_csv("California_DDS_Expenditures.csv")
5
6 ggplot(male, aes(x=Age.Group, y = Expenditures)) + geom_bar(stat = "identity") + ggtitle("Male Expenditures")
7
```

### Barchart of Expenditures of Female Recipients by Age Groups

```
Boxplot Expenditures by Ethnicity.R* x female x male x data x
1 library(ggplot2)
2 library(readr)
3
4 data <- read_csv("California_DDS_Expenditures.csv")
5
6 ggplot(female, aes(x=Age.Group, y = Expenditures)) + geom_bar(stat = "identity") + ggtitle("Female Expenditures")
7
```

### Boxplot of Expenditures by Ethnicity Age Group 0-5

```
Boxplot Expenditures by Ethnicity.R* x data x hispanic x wnohispanic x
1 library(readr)
2
3 data <- read_csv("California_DDS_Expenditures.csv")
4
5 agezerotofive <- subset(data, Age.Group == "0 to 5")
6 boxplot(agezerotofive$Expenditures ~ agezerotofive$Ethnicity, xlab = "Ethnicity", ylab = "Expenditures", main = "Ages 0 - 5 Grouping")
7
```

## Boxplot of Expenditures by Ethnicity Age Group 6-12

```
Boxplot Expenditures by Ethnicity.R* | data | hispanic | wnothispanic |
1 library(readr)
2
3 data <- read.csv("California_DDS_Expenditures.csv")
4
5 agesixtotwelve <- subset(data, Age.Group == "6 to 12")
6 boxplot(agesixtotwelve$Expenditures ~ agesixtotwelve$Ethnicity, xlab = "Ethnicity", ylab = "Expenditures", main = "Ages 6 - 12 Grouping")
7
```

## Boxplot of Expenditures by Ethnicity Age Group 13-17

```
Boxplot Expenditures by Ethnicity.R* | data | hispanic | wnothispanic |
1 library(readr)
2
3 data <- read.csv("California_DDS_Expenditures.csv")
4
5 agesthirtoseventeen <- subset(data, Age.Group == "13 to 17")
6 boxplot(ageshirtoseventeen$Expenditures ~ ageshirtoseventeen$Ethnicity, xlab = "Ethnicity", ylab = "Expenditures", main = "Ages 13 - 17 Grouping")
7
```

## Boxplot of Expenditures by Ethnicity Age Group 18-21

```
Boxplot Expenditures by Ethnicity.R* | data | hispanic | wnothispanic |
1 library(readr)
2
3 data <- read.csv("California_DDS_Expenditures.csv")
4
5 ageeightentotwentyone <- subset(data, Age.Group == "18 to 21")
6 boxplot(ageeightentotwentyone$Expenditures ~ ageeightentotwentyone$Ethnicity, xlab = "Ethnicity", ylab = "Expenditures", main = "Ages 18 - 21 Grouping")
7
```

## Boxplot of Expenditures by Ethnicity Age Group 22-50

```
Boxplot Expenditures by Ethnicity.R* | data | hispanic | wnothispanic |
1 library(readr)
2
3 data <- read.csv("California_DDS_Expenditures.csv")
4
5 agestwentytwotofifty <- subset(data, Age.Group == "22 to 50")
6 boxplot(ageswentytwotofifty$Expenditures ~ ageswentytwotofifty$Ethnicity, xlab = "Ethnicity", ylab = "Expenditures", main = "Ages 22 - 50 Grouping")
7
```

## Boxplot of Expenditures by Ethnicity Age Group 50+

```
Boxplot Expenditures by Ethnicity.R* | data | hispanic | wnothispanic |
1 library(readr)
2
3 data <- read.csv("California_DDS_Expenditures.csv")
4
5 agesfiftyoneplus <- subset(data, Age.Group == "51+")
6 boxplot(agesfiftyoneplus$Expenditures ~ agesfiftyoneplus$Ethnicity, xlab = "Ethnicity", ylab = "Expenditures", main = "Ages 50 Plus Grouping")
7
```

## Boxplot of Expenditures by Ethnicity

```

Boxplot Expenditures by Ethnicity.R
1 library(readr)
2 data <- read_csv("California_DDS_Expenditures.csv")
3
4 boxplot(data$Expenditures~data$Ethnicity, xlab = "Ethnicity", ylab = "Expenditures", main = "Expenditures by Ethnicity")
5
6

```

## Summary of Female and Male Data

```

Boxplot Expenditures by Ethnicity.R * female * male * data *
1 library(readr)
2
3 data <- read_csv("California_DDS_Expenditures.csv")
4 female <- subset(data, Gender == "Female")
5 male <- subset(data, Gender == "Male")
6
7 summary(female)
8 summary(male)
9

```

```

> summary(female)
  Id      Age.Group      Age      Gender      Expenditures      Ethnicity
Min. :10210 Length:503 Min. : 0.00 Length:503 Min. : 222 Length:503
1st Qu.:31650 Class :character 1st Qu.:11.50 Class :character 1st Qu.: 2872 Class :character
Median :55442 Mode :character Median :18.00 Mode :character Median : 6400 Mode :character
Mean :54721 Mean :22.74 Mean :18130
3rd Qu.:76533 3rd Qu.:25.00 3rd Qu.:39488
Max. :99718 Max. :95.00 Max. :75098

> summary(male)
  Id      Age.Group      Age      Gender      Expenditures      Ethnicity
Min. :10409 Length:497 Min. : 0.00 Length:497 Min. : 386 Length:497
1st Qu.:31919 Class :character 1st Qu.:13.00 Class :character 1st Qu.: 2954 Class :character
Median :55345 Mode :character Median :18.00 Mode :character Median : 7219 Mode :character
Mean :54604 Mean :22.86 Mean :18001
3rd Qu.:75960 3rd Qu.:26.00 3rd Qu.:37201
Max. :99898 Max. :95.00 Max. :68890

```

## Summary of Hispanic and White Non-Hispanic Data

```

Boxplot Expenditures by Ethnicity.R * data * hispanic * wnothispanic *
1 library(readr)
2
3 data <- read_csv("California_DDS_Expenditures.csv")
4 hispanic <- subset(data, Ethnicity == "Hispanic")
5 wnothispanic <- subset(data, Ethnicity == "white not Hispanic")
6
7 summary(hispanic)
8 summary(wnothispanic)
9

```

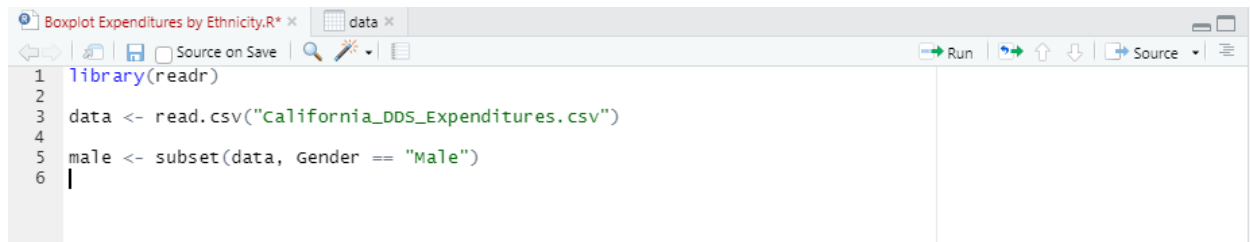
```

> summary(hispanic)
  Id      Age.Group      Age      Gender      Expenditures      Ethnicity
Min. :10486 Length:376 Min. : 0.00 Length:376 Min. : 222 Length:376
1st Qu.:29318 Class :character 1st Qu.: 9.75 Class :character 1st Qu.: 2331 Class :character
Median :55079 Mode :character Median :15.00 Mode :character Median : 3952 Mode :character
Mean :53742 Mean :17.38 Mean :11066
3rd Qu.:74612 3rd Qu.:20.00 3rd Qu.:10292
Max. :99791 Max. :95.00 Max. :65581

> summary(wnothispanic)
  Id      Age.Group      Age      Gender      Expenditures      Ethnicity
Min. :10210 Length:401 Min. : 0.00 Length:401 Min. : 340 Length:401
1st Qu.:31859 Class :character 1st Qu.:15.00 Class :character 1st Qu.: 3977 Class :character
Median :54675 Mode :character Median :21.00 Mode :character Median :15718 Mode :character
Mean :54179 Mean :27.98 Mean :24698
3rd Qu.:76533 3rd Qu.:33.00 3rd Qu.:43134
Max. :99898 Max. :90.00 Max. :68890

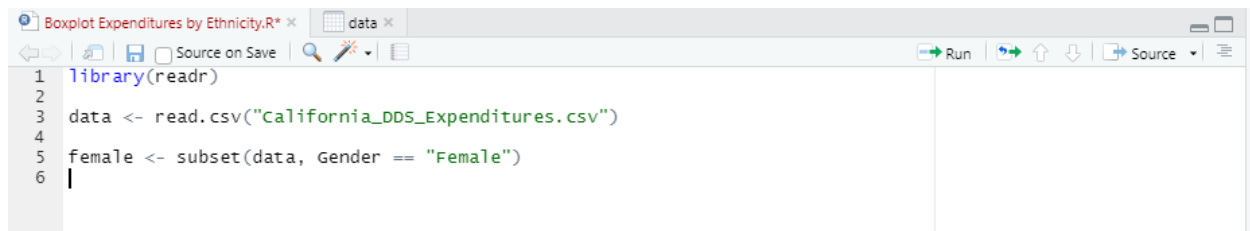
```

## Subset Data by Gender - Male



```
1 library(readr)
2
3 data <- read.csv("California DDS Expenditures.csv")
4
5 male <- subset(data, Gender == "Male")
6 |
```

## Subset Data by Gender - Female



```
1 library(readr)
2
3 data <- read.csv("California DDS Expenditures.csv")
4
5 female <- subset(data, Gender == "Female")
6 |
```