Writeup

Writeup

Project 1

Both the pdfs "Arizona_Department_of_Correction_Inmate_Fact_Sheet_2020.pdf" and "Arizona_Department_of_Corrections_Cost_Report_2020.pdf" which can be found in ./data/Raw_Data are reports released by the Arizona Department of Corrections,Rehabilitation& Reentry.

In order to extract tables from these pdf reports into csv files, I first used Adobe Acrobat to turn the pdf files into xlsx files. The xlsx versions of these files can also be found in ./data/Raw_Data. Then, I used Excel to inspect these files and identify which tables I wanted to extract. Each table was extracted into its own csv file.

Cost Report

The Cost Report could not easily be saved as a csv file because it had images and text that was not in a table. Thus, saving it as a csv would cause information to be lost. Thus, I decided to extract the individual tables that were relevant to me from the xlsx file. The R script "Cost_Report_Data_Cleaning.R" extracts these tables into csv files. The focus of my project is the difference in expenditures between private and public prison units. Thus, although the report has lots of interesting information, I focused on tables directly related to this.

I extracted the Expenditure Summary table on page 8 of the pdf. In this table, I only extracted the information from the Private Prisons. The information on private prisons was organized by unit. I extracted the entire table from pages 10 and 11 of the pdf, which contains information on State Prison Expenditure by Prison Unit. Since both the extracted data for Private and Public prisons was then organized by unit, I could compare the data.

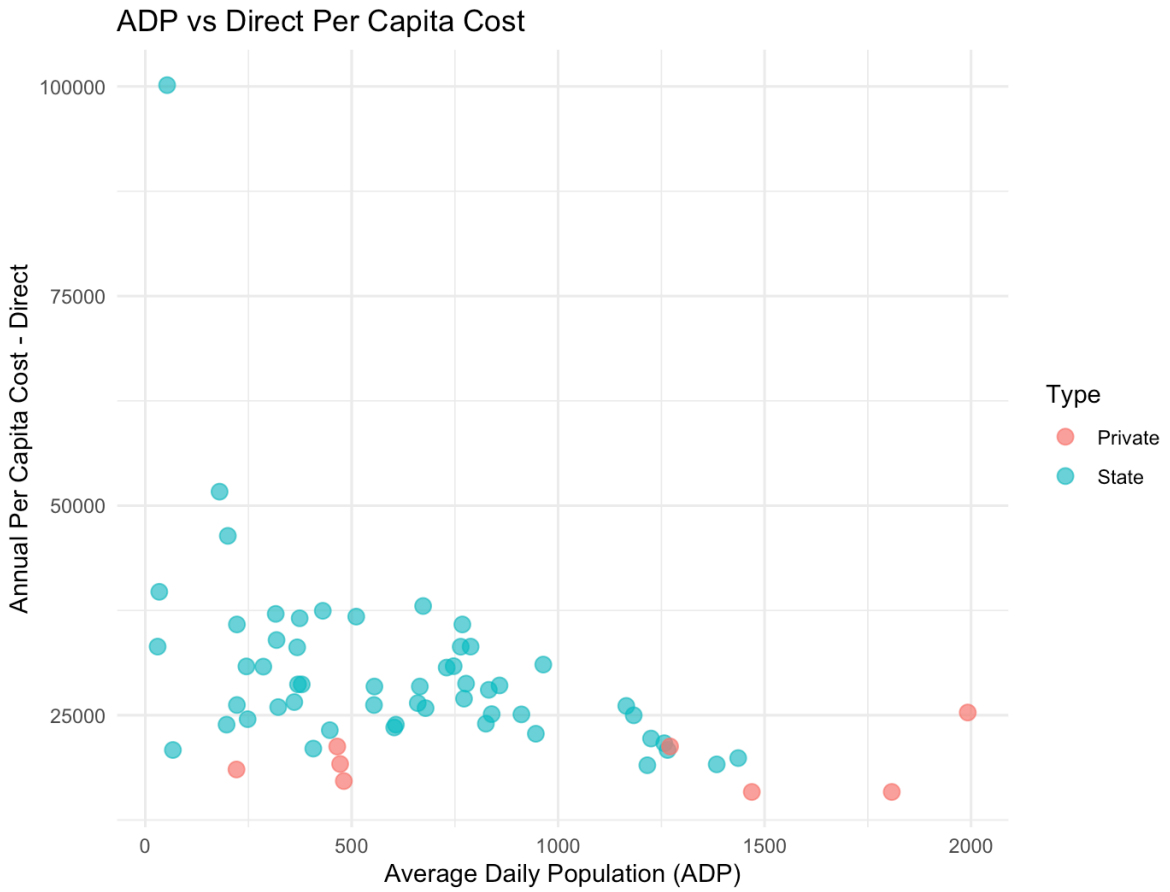
These extracted csv files are saved in ./data/Processed_Data as follows:

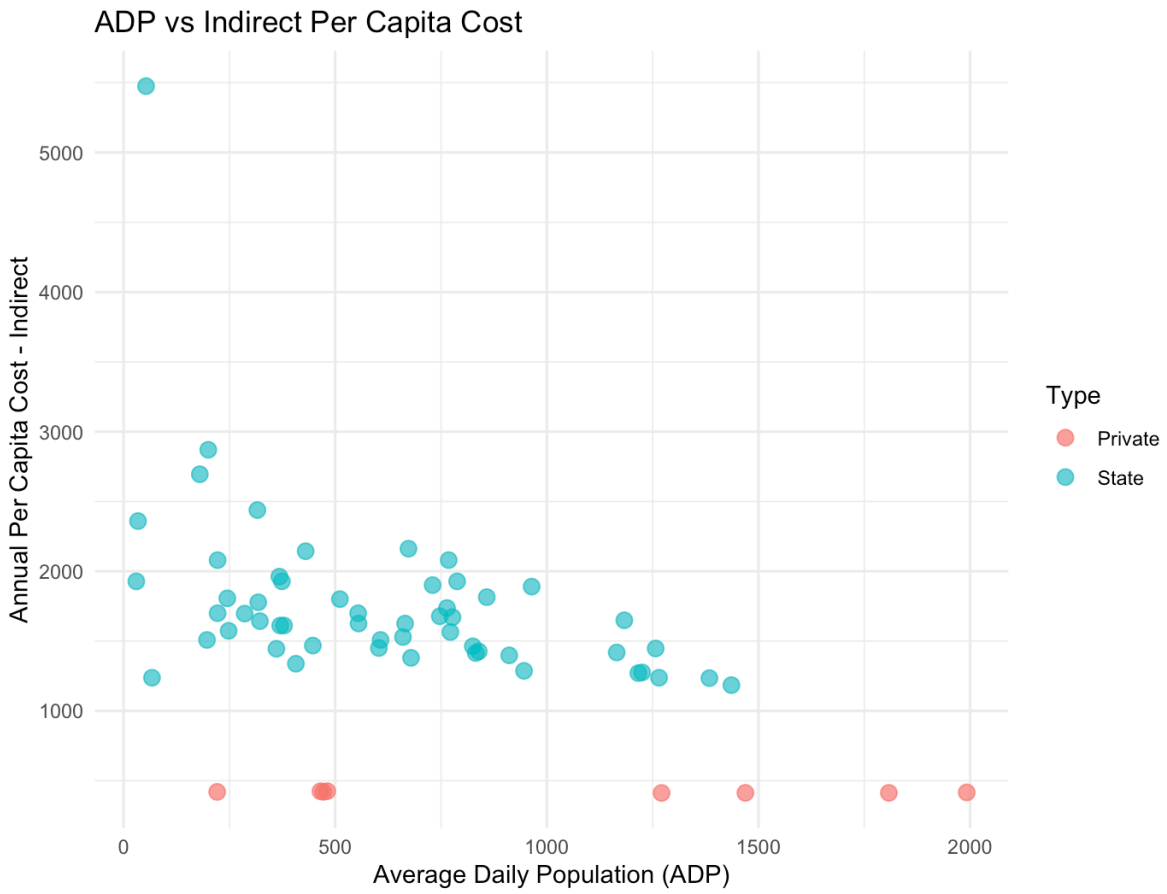
- Private_Prison_Revenue.csv
- State_Prison_Revenue.csv

Project 2

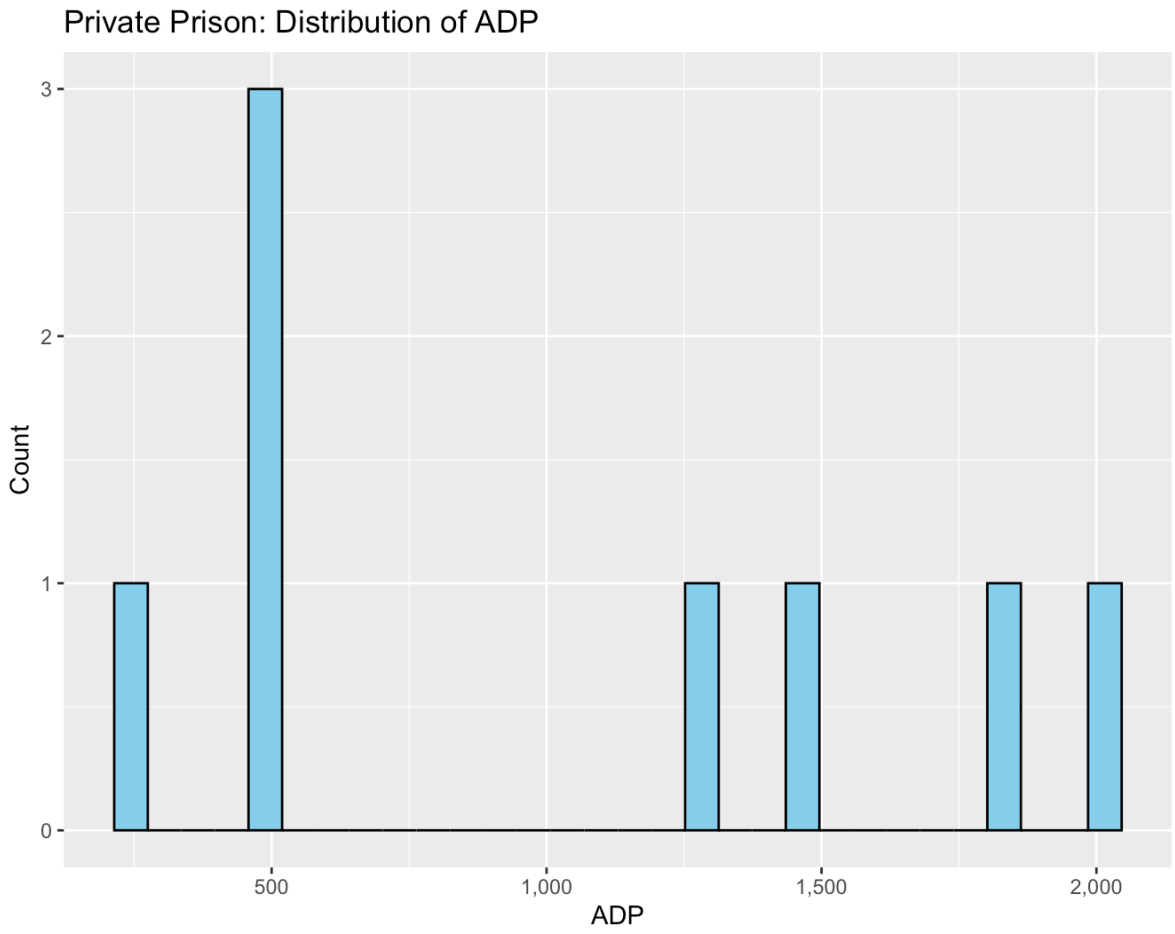
Multivariate Plot:

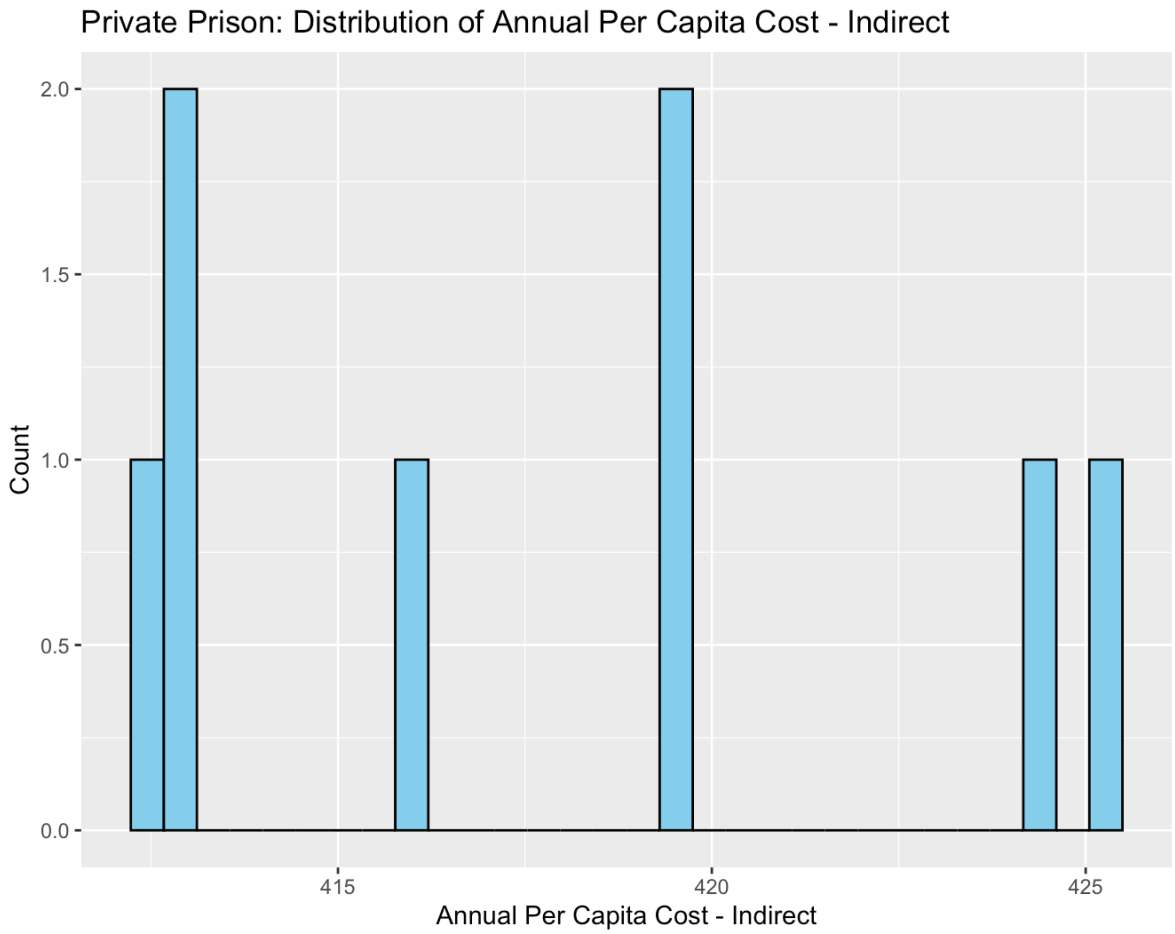
Private prisons have less Direct and Indirect Annual Per Capita Cost than State prisons. Direct Costs are tied to the operation of a specific prison facility or unit. They include salaries for staff; food, clothing, and hygiene products for inmates; utilities (electricity, water, etc.) used in the facility; and facility maintenance. Indirect costs are shared or overhead costs that support the prison system but aren't directly linked to one facility. They include central administration salaries (e.g., the state corrections department headquarters); IT systems that serve multiple facilities; legal services and oversight; state-level contracting or procurement services; and training programs for correctional officers across facilities.

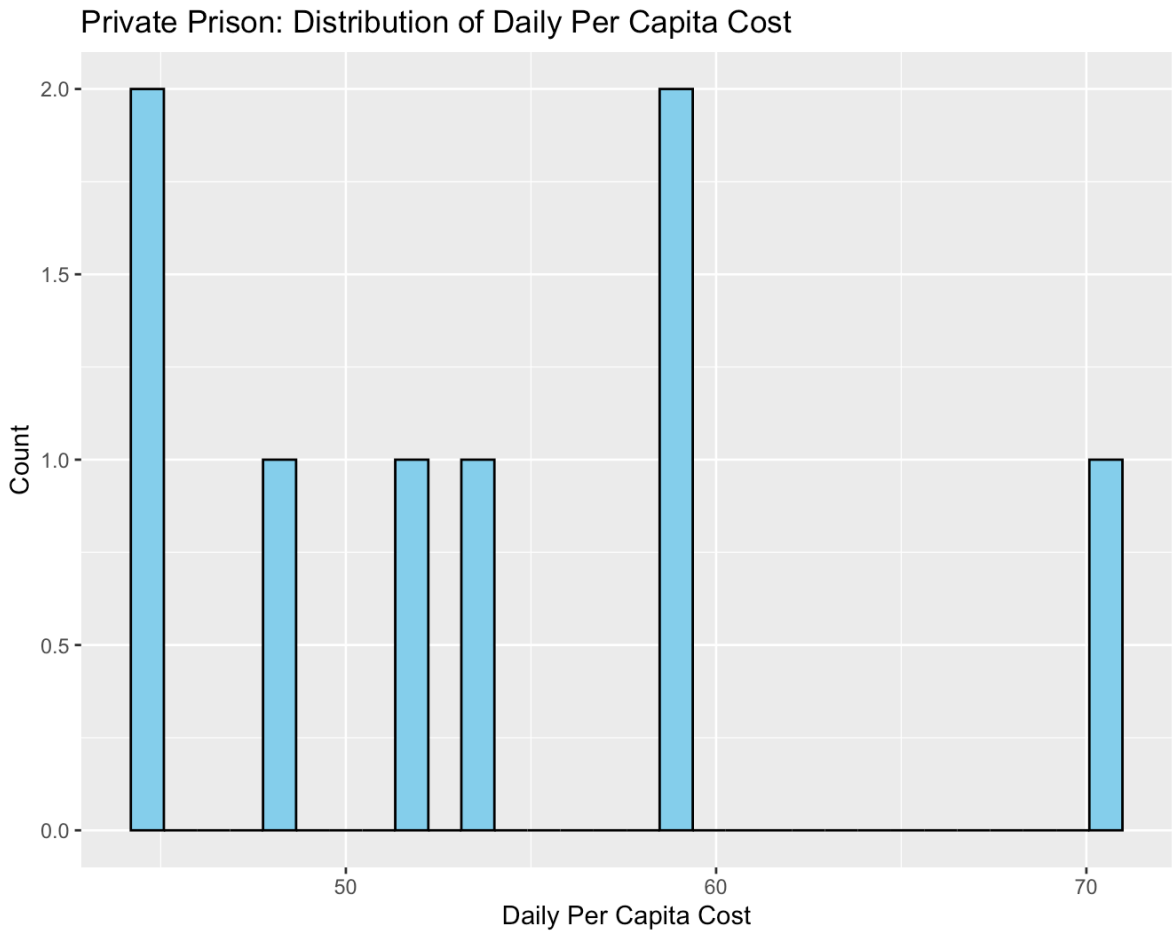


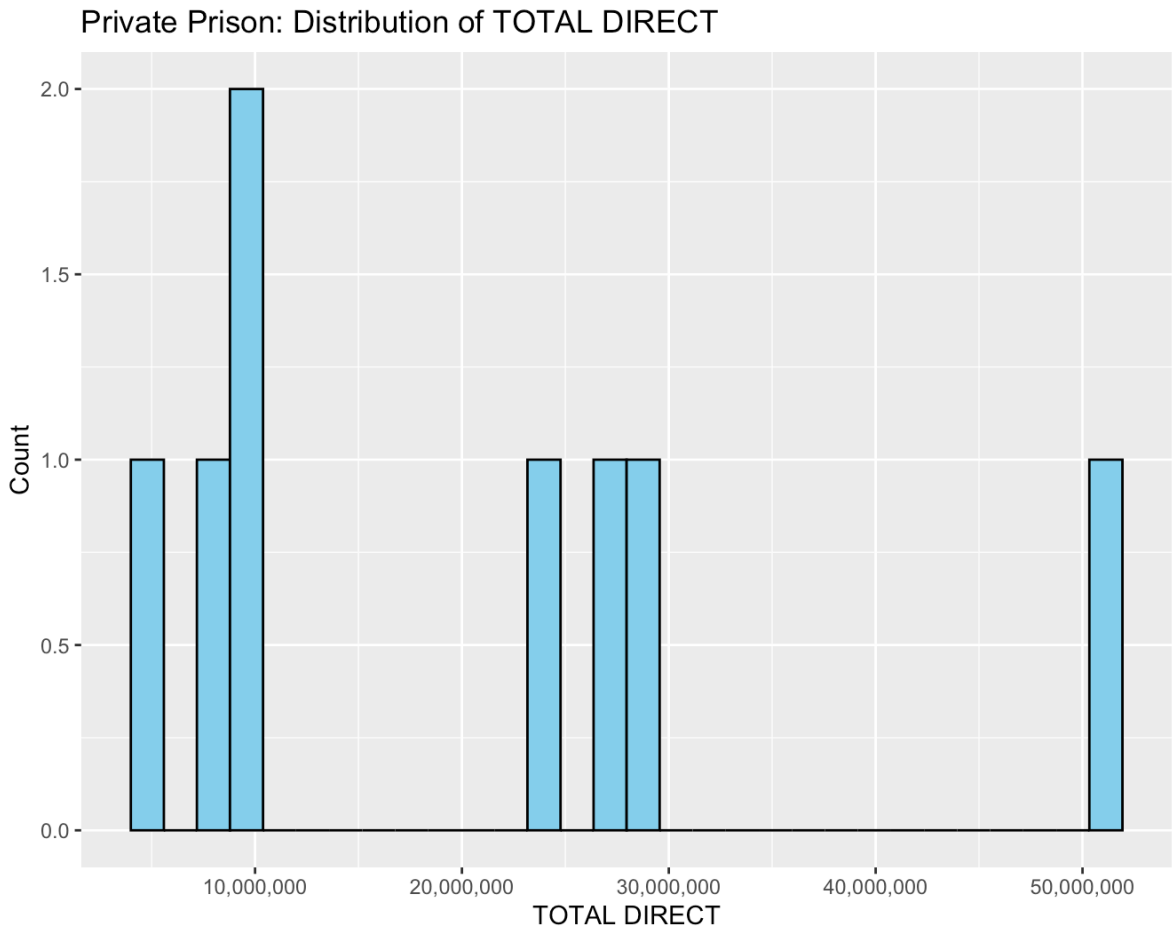


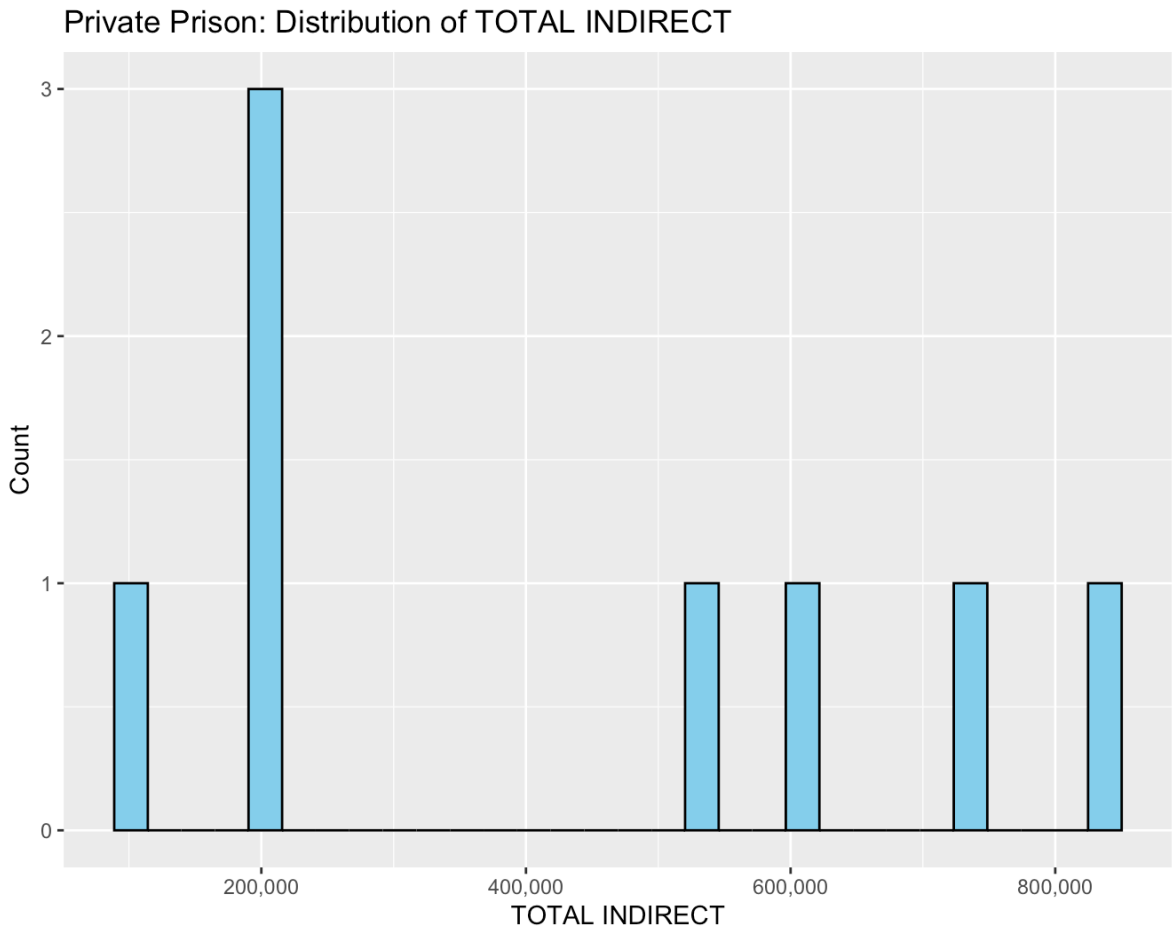
Private Prison Distributions:


Private Prison: Distribution of ADP
Count
3
2
1
0
500
1,000
1,500
2,000
ADP

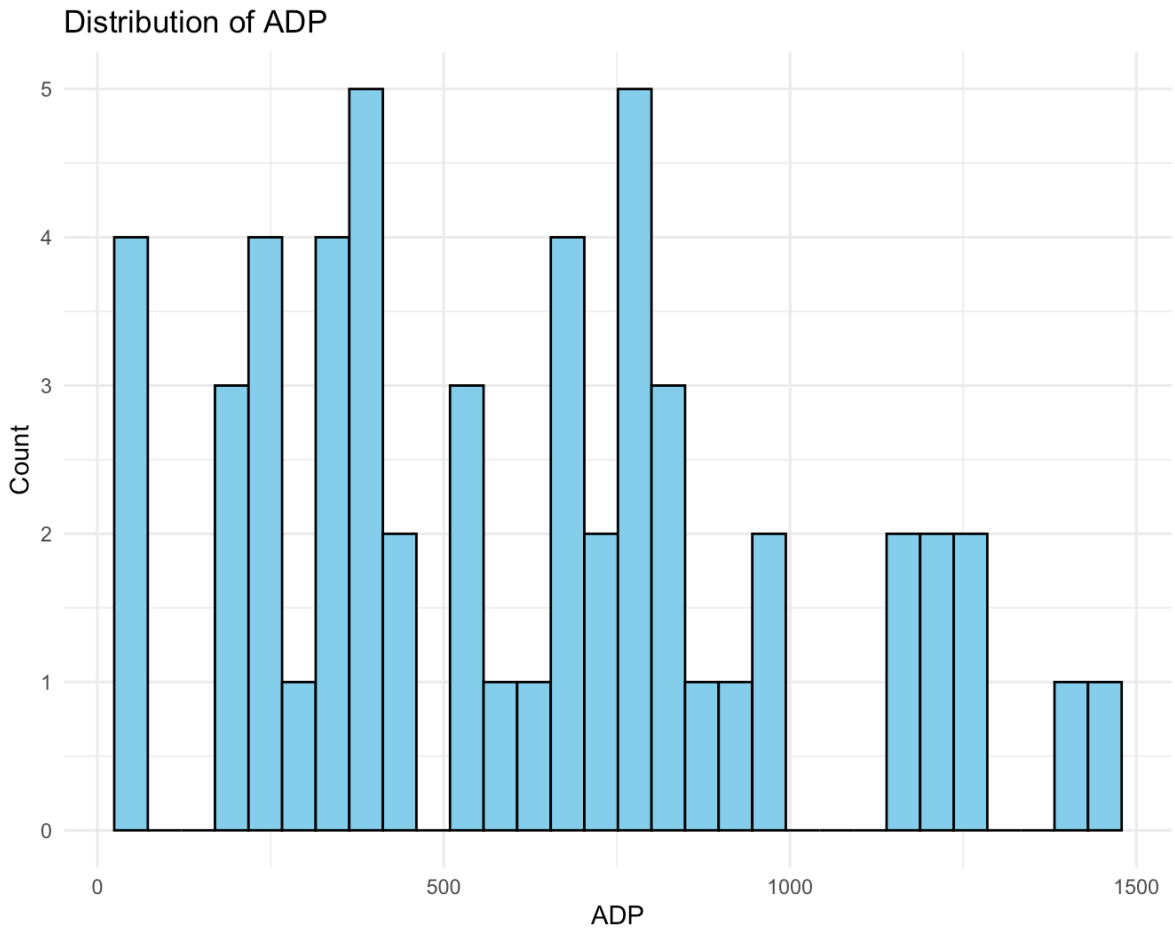

Private Prison: Distribution of Annual Per Capita Cost - Indirect
Count
2.0
1.5
1.0
0.5
0.0
415
420
425
Annual Per Capita Cost - Indirect

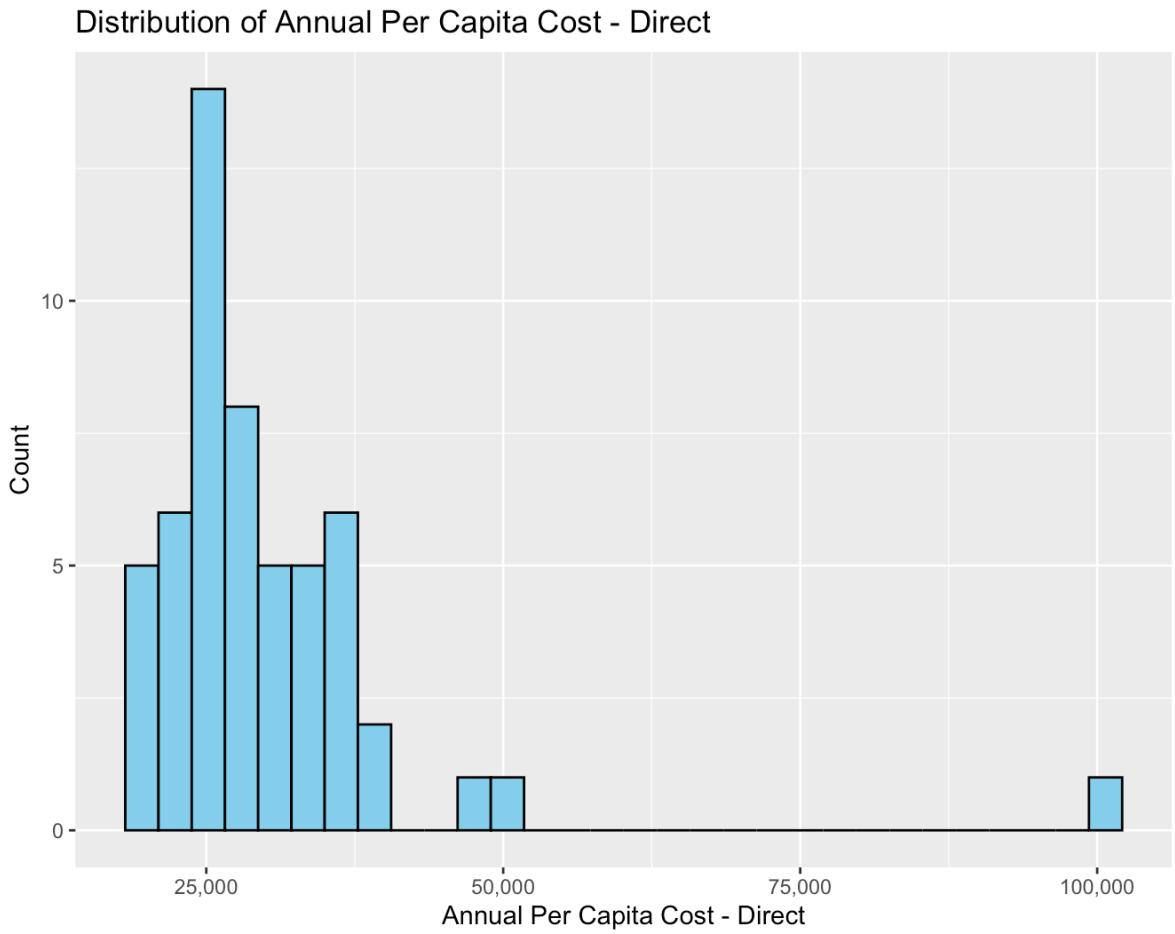

Private Prison: Distribution of Daily Per Capita Cost
Count
2.0
1.5
1.0
0.5
0.0
50
60
70
Daily Per Capita Cost

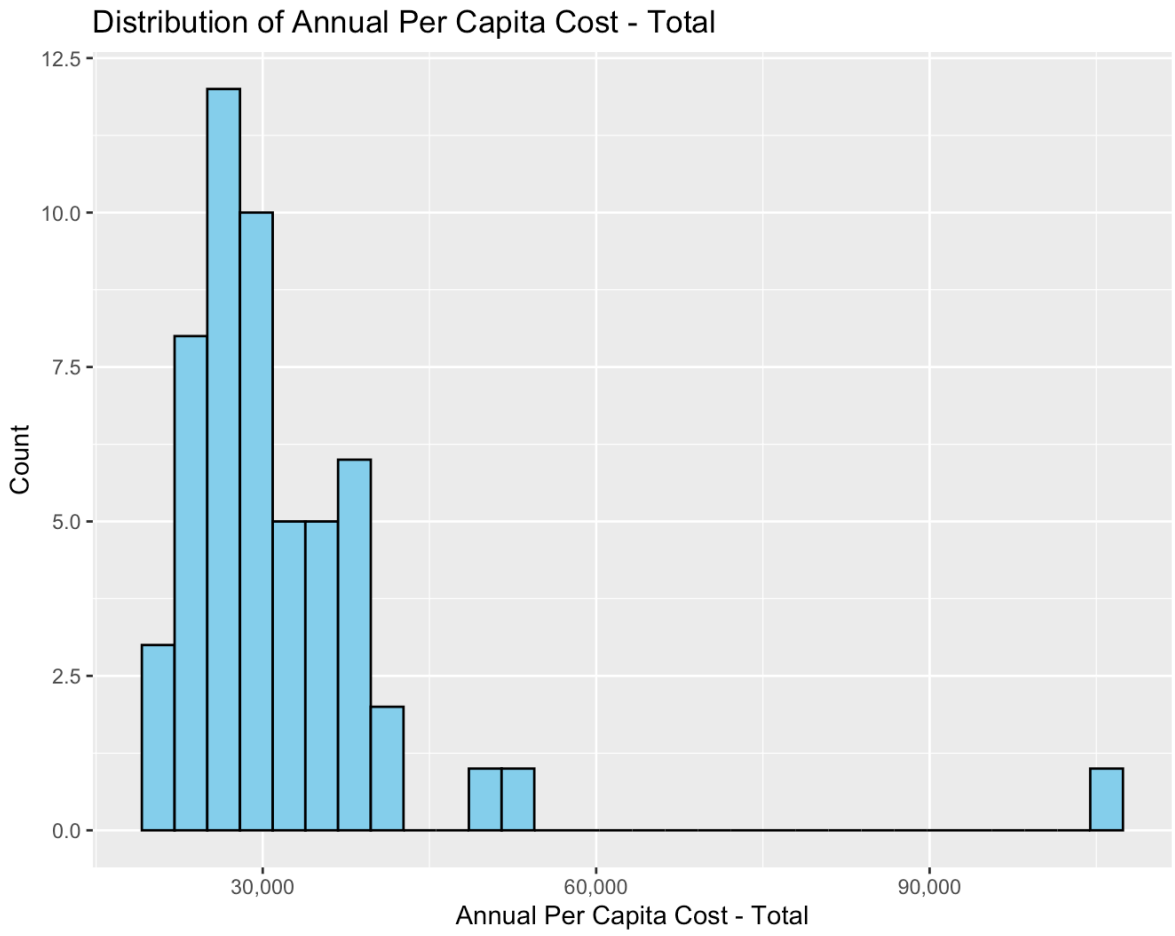

Private Prison: Distribution of TOTAL DIRECT
Count
2.0
1.5
1.0
0.5
0.0
10,000,000
20,000,000
30,000,000
40,000,000
50,000,000
TOTAL DIRECT

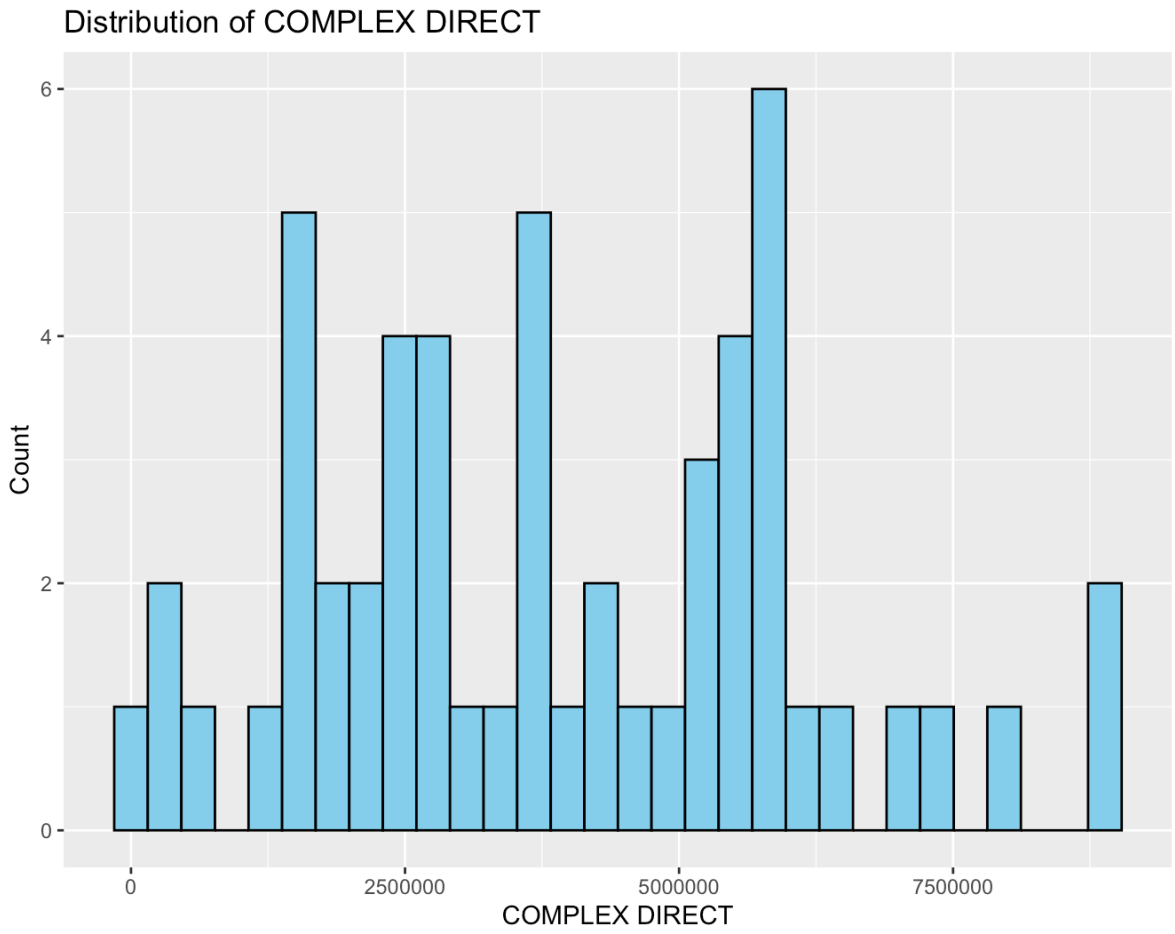


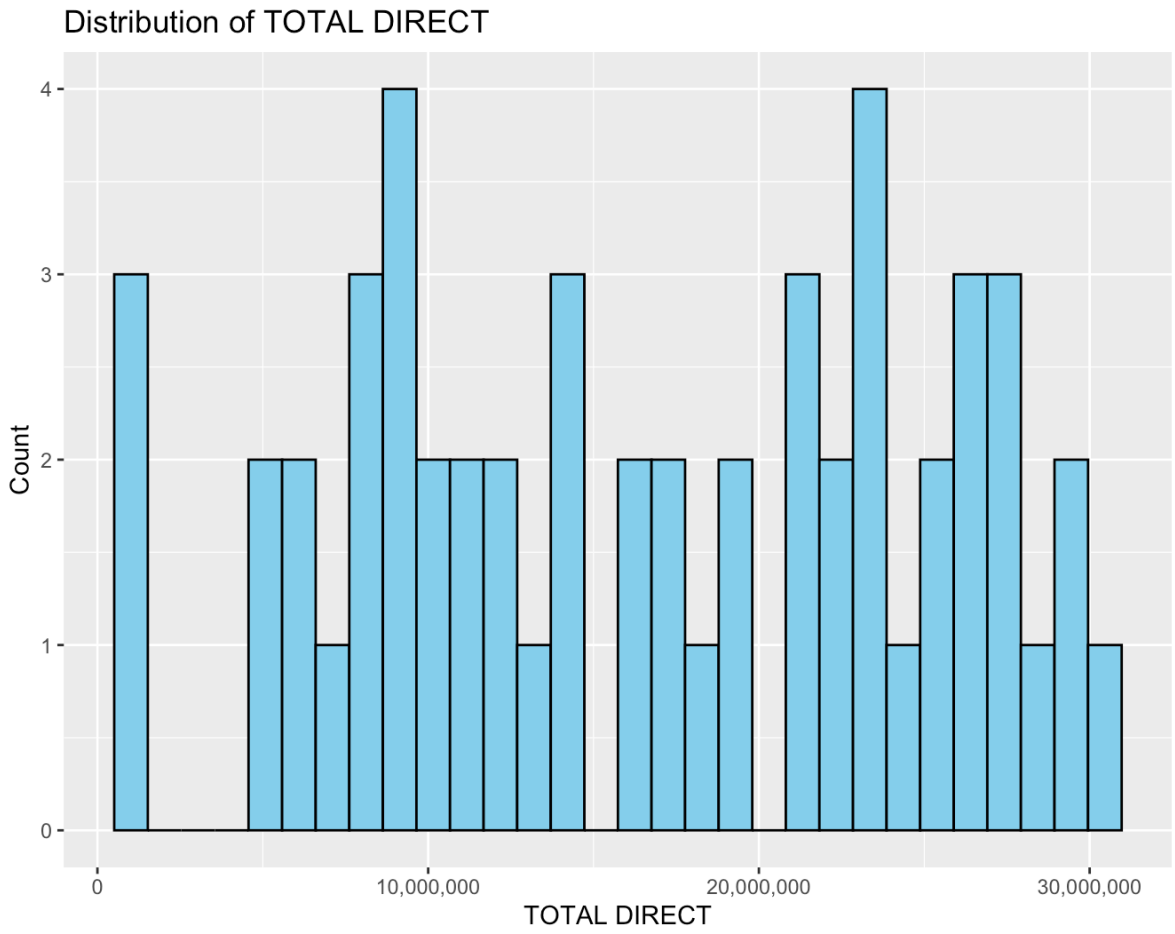
State Prison Distributions:

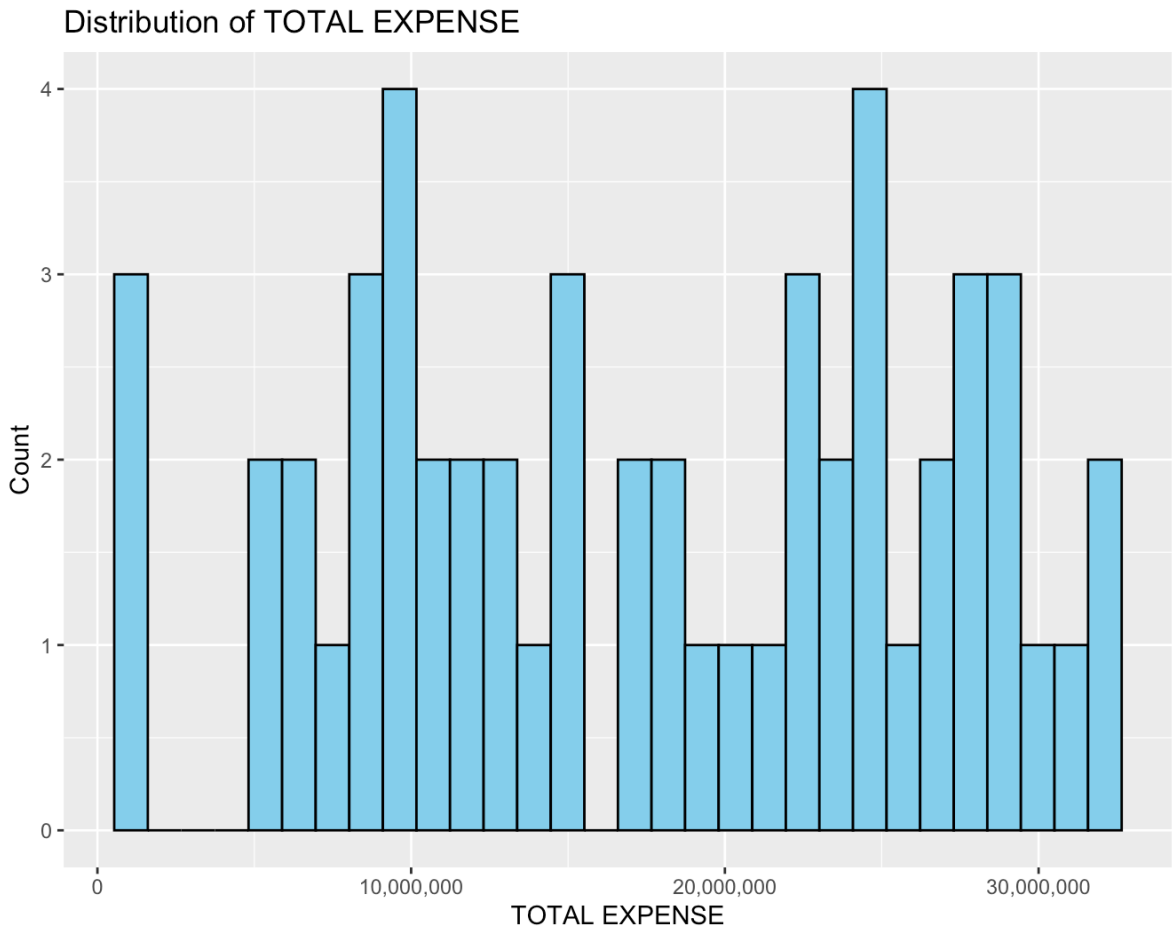

Distribution of ADP
Count
ADP
0
1
2
3
4
5
0
500
1000
1500

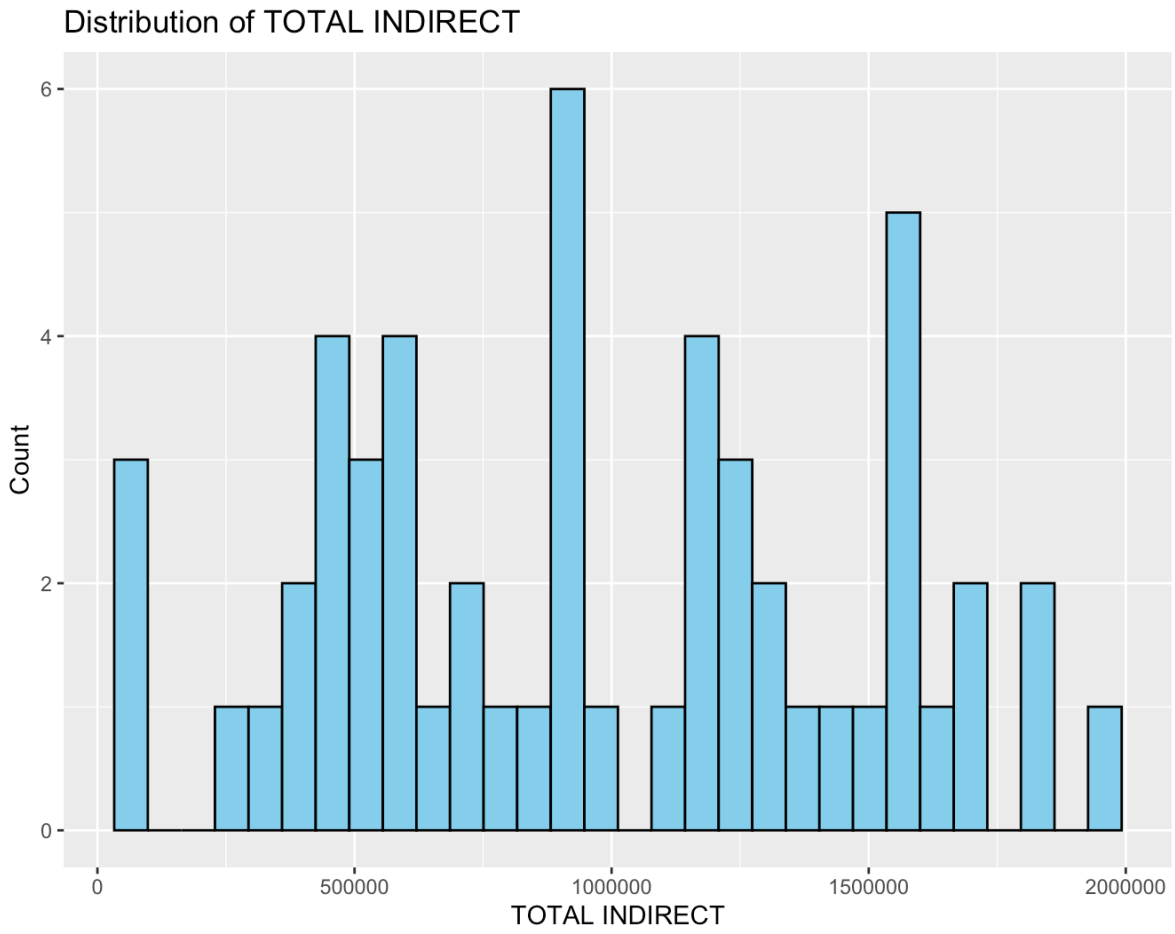

Distribution of Annual Per Capita Cost - Direct
Count
10
5
0
25,000
50,000
75,000
100,000
Annual Per Capita Cost - Direct

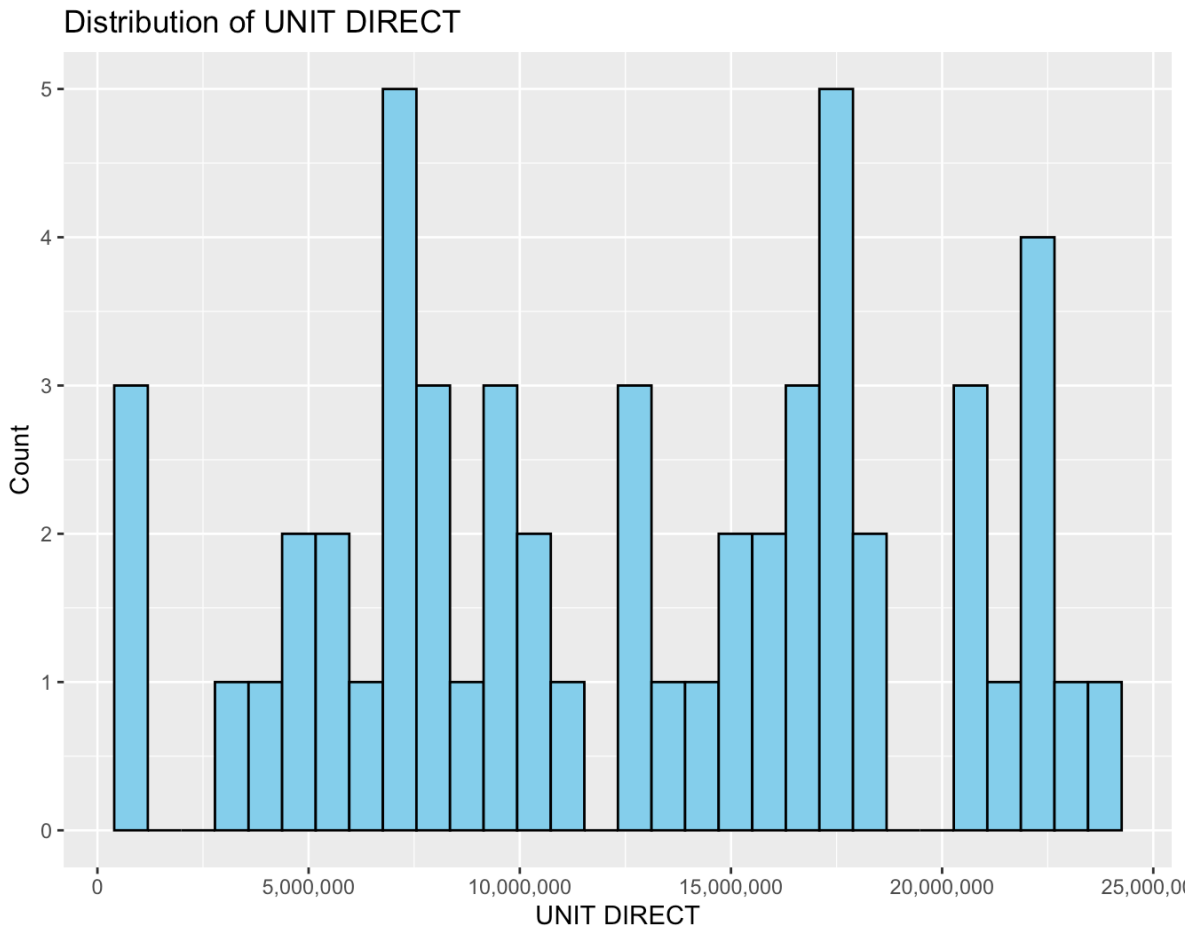

Distribution of Annual Per Capita Cost - Total
12.5
10.0
7.5
5.0
2.5
0.0
Count
30,000
60,000
90,000
Annual Per Capita Cost - Total


Distribution of COMPLEX DIRECT
Count
6
4
2
0
0
2500000
5000000
7500000
COMPLEX DIRECT


Distribution of TOTAL DIRECT
Count
4
3
2
1
0
0
10,000,000
20,000,000
30,000,000
TOTAL DIRECT


Distribution of TOTAL EXPENSE
Count
0
1
2
3
4
0
10,000,000
20,000,000
30,000,000
TOTAL EXPENSE


Distribution of TOTAL INDIRECT
Count
6
4
2
0
0
500000
1000000
1500000
2000000
TOTAL INDIRECT



Project 3

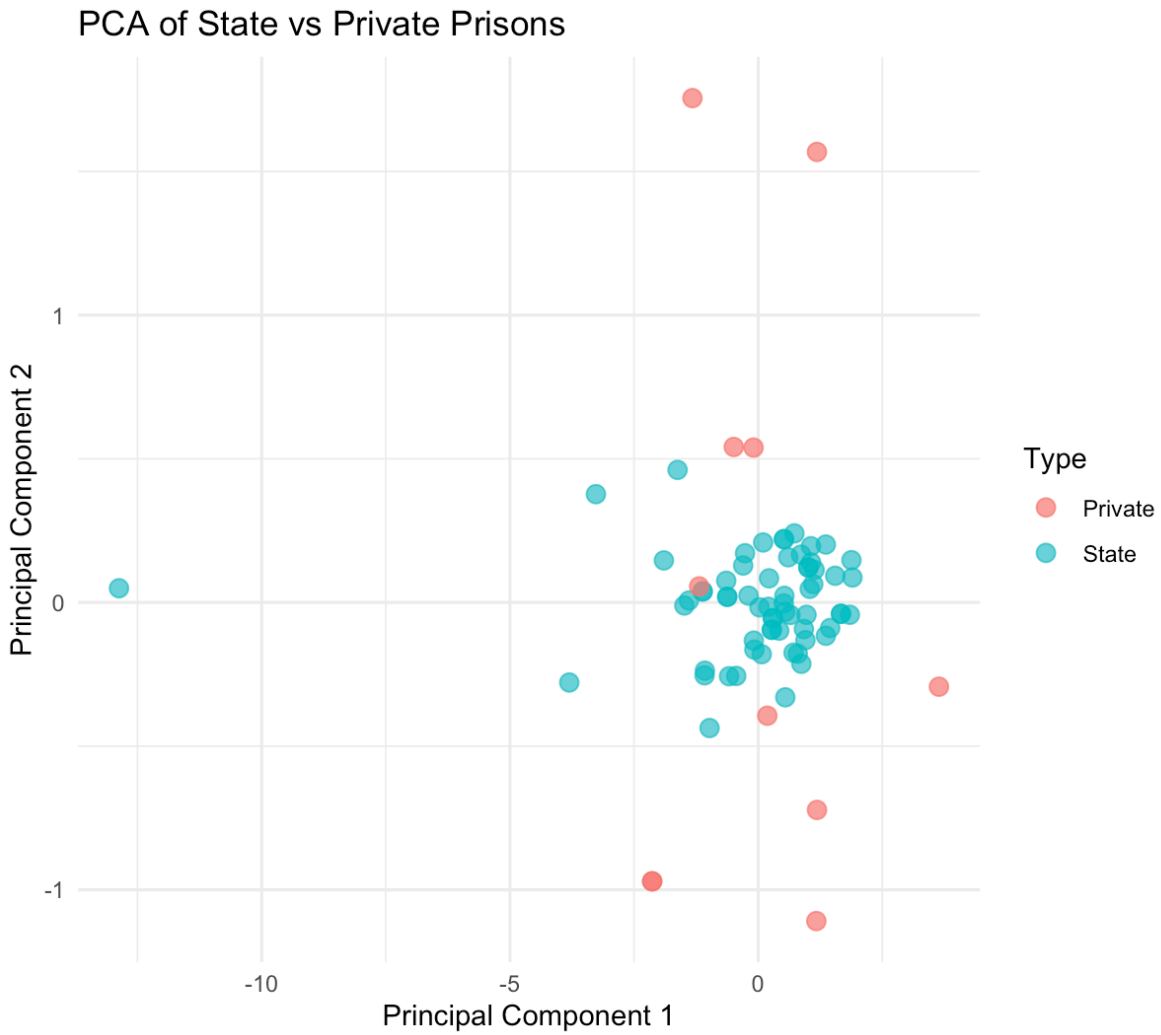
Model

In order to analyze the difference between private and public prisons, I used Principle Component Analysis. After reducing dimensionality with PCA, I created a plot in which private and public prisons are colored differently.

Analysis

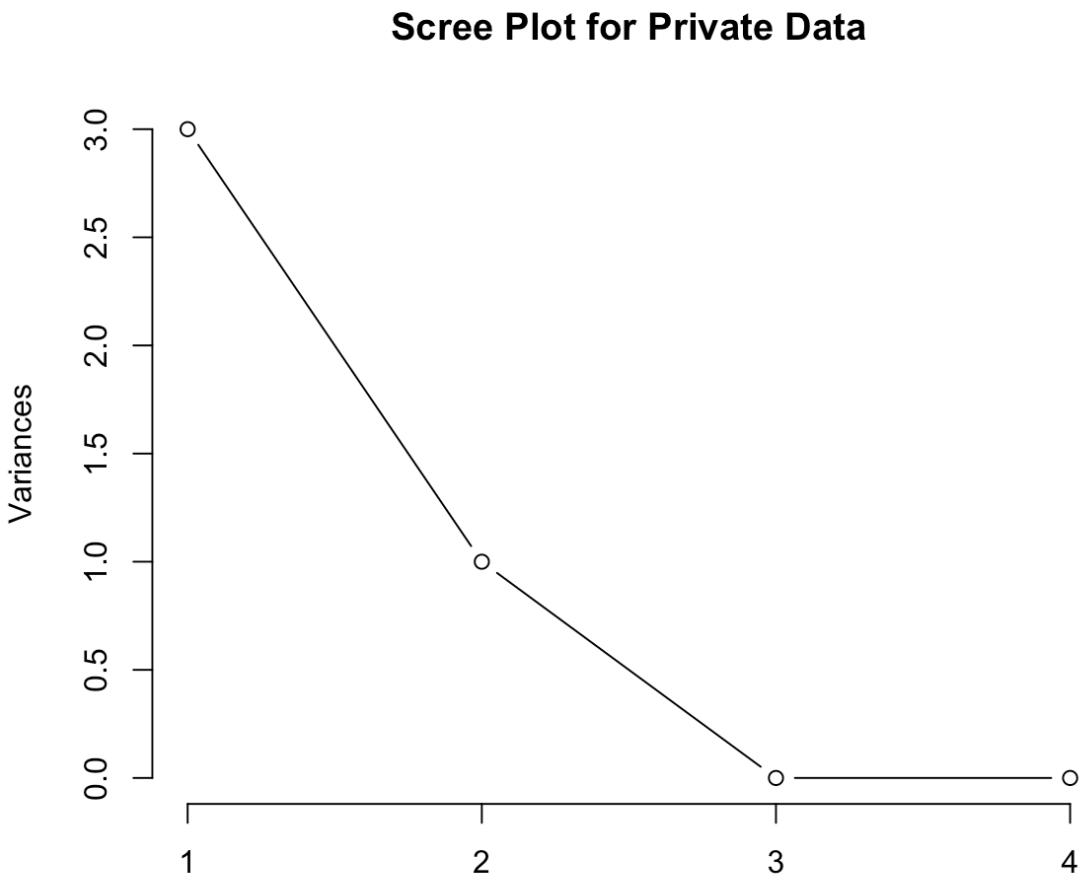
PCA

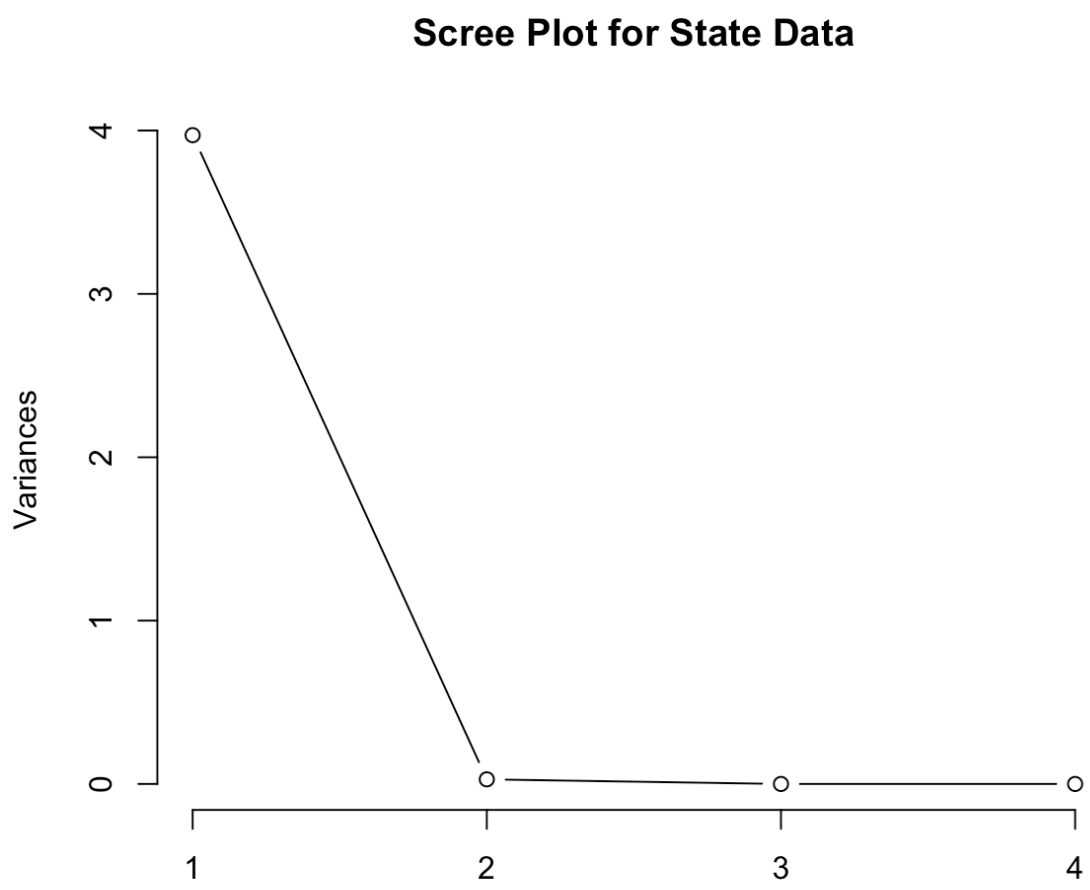
The PCA Graph shows that State prisons are more consistent in how they operate per inmate and Private prisons are more varied and some diverge significantly from the state cluster. The difference is not significant enough to determine that all private prisons are different from all state prisons.



The scree plot for private prisons shows that PCA1 and PCA2 have non-zero variances. This implies PCA found two patterns in the data.

The scree plot for state prisons shows that only PCA1 has a non-zero variance. This implies PCA found only one pattern in the data.


Scree Plot for Private Data
Variances
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1
2
3
4



These are the features in the PCA components:

PC1 (State) and PC1 (Private):

Annual Per Capita Cost - Total

Daily Per Capita Cost

Annual Per Capita Cost - Direct

Annual Per Capita Cost - Indirect

PC2 (Private):

Annual Per Capita Cost - Indirect

Annual Per Capita Cost - Direct

Daily Per Capita Cost

Annual Per Capita Cost - Total

This suggests that Annual Per Capita Cost - Indirect is relevant to Private Prisons but not State Prisons.

Individual Regression

State: Annual Per Capita Cost - Direct ~ ADP

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 39088.833 2796.711 13.977 < 2e-16 ***ADP -14.417 3.897 -3.699 0.000522***

State: Annual Per Capita Cost - Indirect ~ ADP

Project 4