Case Study on the Variability of Slave Transactions

Dataset: The dataset used details the financial records of over 16,000 New Orleans slave sales between 1856 and 1861. This dataset was first used in "Betting on Secession: Quantifying Political Events Surrounding Slavery and the Civil War." (2016) by Calomiris, Charles W. and Jonathan Pritchett and published in American Economic Review.

Link to dataset: https://corgis-edu.github.io/corgis/csv/slavery/

The original dataset contained 21 variables listed below.

```
 [1] "Buyer.County.of.Origin"                          "Buyer.Full.Name"
 [3] "Buyer.Origin"                                    "Buyer.State.of.Origin"
 [5] "Seller.County.of.Origin"                         "Seller.Full.Name"
 [7] "Seller.Origin"                                   "Seller.State.of.Origin"
 [9] "Slave.Age"                                       "Slave.Gender"
[11] "Slave.Name"                                      "Slave.Skin.Color"
[13] "Transaction.Date"                                "Transaction.Number.of.Adult.Slaves"
[15] "Transaction.Number.of.Child.Slaves"              "Transaction.Number.of.Total.Slaves.Purchased"
[17] "Transaction.Sale.Details.Discount.Rate"          "Transaction.Sale.Details.Payment.Method"
[19] "Transaction.Sale.Details.Predicted.Interest.Rate" "Transaction.Sale.Details.Price"
[21] "Transaction.Sale.Details.Prices.Listed"
```

However, since many of the variables were categorical or textual variables and thus incompatible with our project of studying the dataset through Principal Component Analysis, I decided to keep the following 8 variables. I observed that 20 rows in the entry did not have a date, so I removed them.

```
[1] "Slave.Age"                                "Transaction.Number.of.Adult.Slaves"
[3] "Transaction.Number.of.Child.Slaves"       "Transaction.Number.of.Total.Slaves.Purchased"
[5] "Transaction.Sale.Details.Discount.Rate"   "Transaction.Sale.Details.Predicted.Interest.Rate"
[7] "Transaction.Sale.Details.Price"           "Dates"
```

Variable Description:

[1] The age of the slave sold.	[2] The number of adult slaves bought in the transaction.
[3] The number of child slaves purchased.	[4] The total number of slaves purchased.
[5] The discount rate provides an estimate for the present value of future cash flow. It helps to quantify the value of an investment at the moment of a slave purchase based on projections of how much money it will generate in the future.	[5] Slaves could be bought with a cash down payment or credit. The discount rate refers to the interest rate charged by slave owners for having loaned slaves by attributing credit.
[7] The total price of a slave buyer's transaction.	[6] The year of the transaction*.

Note: While the original dataset detailed the day, month, and year of the transaction, I decided to keep only the year to study the impact of year variations on the total variability of the dataset.

As shown below, the data had drastically different mean and standard deviation values.

```
   Slave.Age    Transaction.Number.of.Adult.Slaves Transaction.Number.of.Child.Slaves Transaction.Number.of.Total.Slaves.Purchased
 Min.   :-1.0   Min.   :  0.000                    Min.   : 0.0000                   Min.   :  1.000
 1st Qu.:19.0   1st Qu.:  1.000                    1st Qu.: 0.0000                   1st Qu.:  1.000
 Median :25.0   Median :  1.000                    Median : 0.0000                   Median :  1.000
 Mean   :27.1   Mean   :  1.526                    Mean   : 0.2367                   Mean   :  1.762
 3rd Qu.:35.0   3rd Qu.:  1.000                    3rd Qu.: 0.0000                   3rd Qu.:  2.000
 Max.   :80.0   Max.   :147.000                    Max.   :11.0000                   Max.   :148.000
```

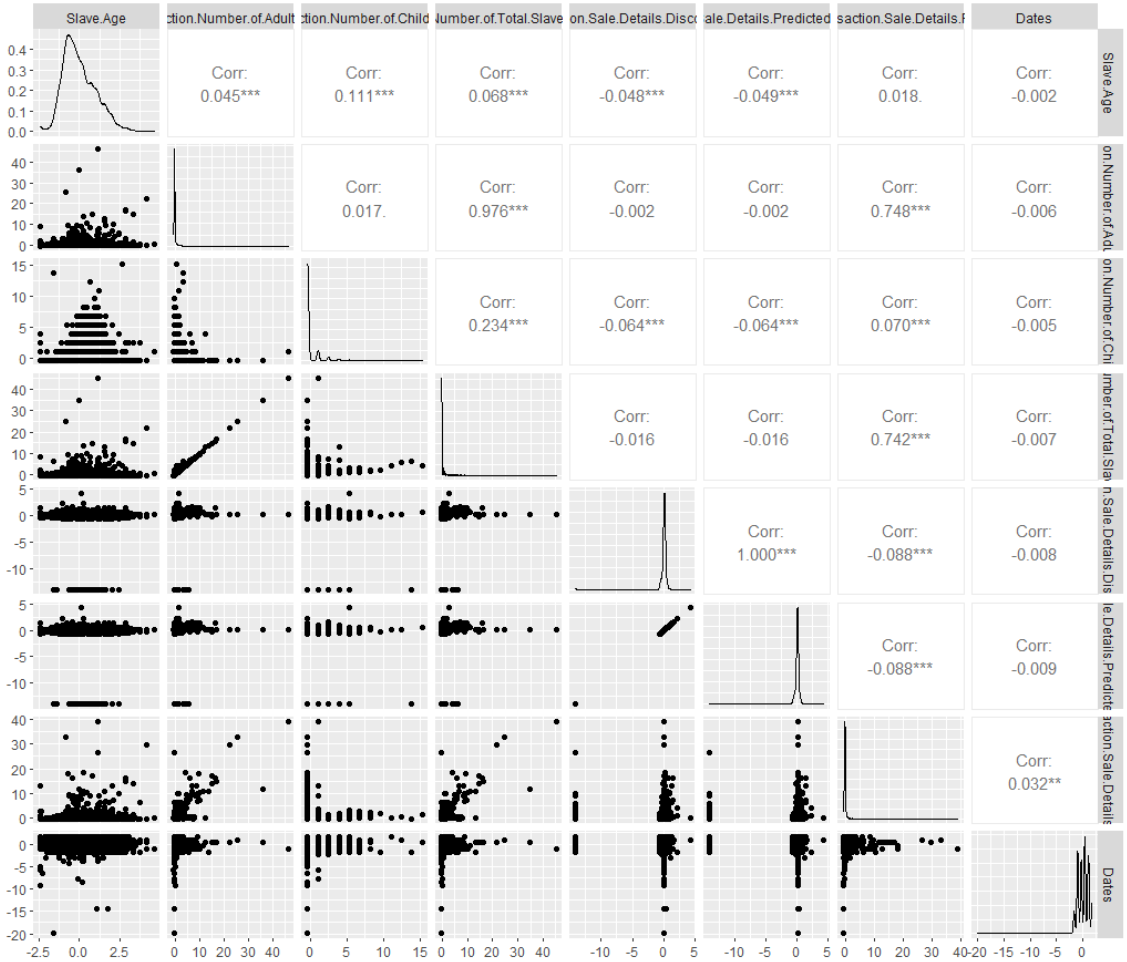
```
Transaction.Sale.Details.Discount.Rate Transaction.Sale.Details.Predicted.Interest.Rate Transaction.Sale.Details.Price      Dates
Min.   :0.000000                       Min.   :0.00000                                  Min.   :    0                    Min.   :1829
1st Qu.:0.006027                       1st Qu.:0.07477                                  1st Qu.:  800                    1st Qu.:1857
Median :0.006027                       Median :0.07477                                  Median : 1100                    Median :1859
Mean   :0.005987                       Mean   :0.07427                                  Mean   : 1310                    Mean   :1859
3rd Qu.:0.006027                       3rd Qu.:0.07477                                  3rd Qu.: 1400                    3rd Qu.:1860
Max.   :0.007767                       Max.   :0.09730                                  Max.   :96500                    Max.   :1861
```

Normalizing the data, I get

```
   Slave.Age       Transaction.Number.of.Adult.Slaves Transaction.Number.of.Child.Slaves Transaction.Number.of.Total.Slaves.Purchased
Min.   :-2.4484   Min.   :-0.4838                    Min.   :-0.3363                   Min.   :-0.2351
1st Qu.:-0.7057   1st Qu.:-0.1667                    1st Qu.:-0.3363                   1st Qu.:-0.2351
Median :-0.1829   Median :-0.1667                    Median :-0.3363                   Median :-0.2351
Mean   : 0.0000   Mean   : 0.0000                    Mean   : 0.0000                   Mean   : 0.0000
3rd Qu.: 0.6885   3rd Qu.:-0.1667                    3rd Qu.:-0.3363                   3rd Qu.: 0.0733
Max.   : 4.6096   Max.   :46.1322                    Max.   :15.2961                   Max.   :45.0971
```

```
Transaction.Sale.Details.Discount.Rate Transaction.Sale.Details.Predicted.Interest.Rate Transaction.Sale.Details.Price      Dates
Min.   :-14.04702                      Min.   :-14.0242                                 Min.   :-0.53777                 Min.   :-19.9290
1st Qu.:  0.09371                      1st Qu.:  0.0943                                 1st Qu.:-0.20938                 1st Qu.: -1.0346
Median :  0.09371                      Median :  0.0943                                 Median :-0.08624                 Median :  0.3150
Mean   :  0.00000                      Mean   :  0.0000                                 Mean   : 0.00000                 Mean   :  0.0000
3rd Qu.:  0.09371                      3rd Qu.:  0.0943                                 3rd Qu.: 0.03690                 3rd Qu.:  0.9898
Max.   :  4.17728                      Max.   :  4.3479                                 Max.   :39.07343                 Max.   :  1.6646
```

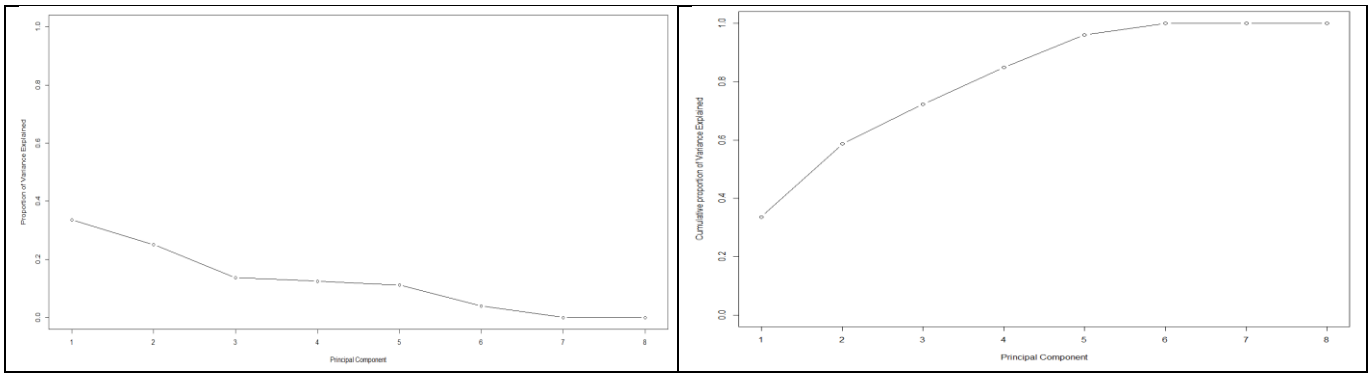
Before performing Principal Component Analysis, I noticed the discount rate and interest rates have a correlation of 1, and the number of adult slaves purchased has correlation 0.976 with the number of slaves purchased. Finally, both the number of slaves purchased, and the number of adult slaves purchased have strong correlations with the transaction price having values 0,748 and 0.742 respectively.



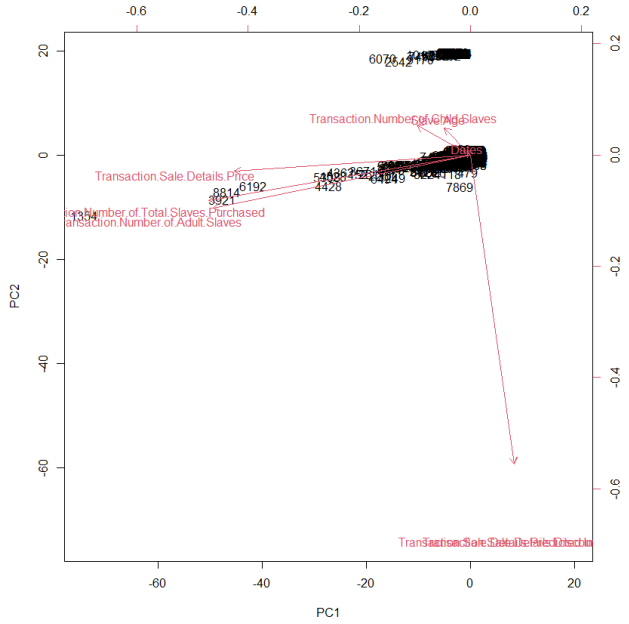
Carrying out Principal Component Analysis on the normalized data I get the following results.

```
> summary(pr.out)
Importance of components:
                          PC1    PC2    PC3    PC4    PC5     PC6      PC7       PC8
Standard deviation     1.6397 1.4154 1.0465 1.0001 0.9467 0.56260 0.005035 1.046e-14
Proportion of Variance 0.3361 0.2504 0.1369 0.1250 0.1120 0.03956 0.000000 0.000e+00
Cumulative Proportion  0.3361 0.5865 0.7234 0.8484 0.9604 1.00000 1.000000 1.000e+00
> pr.out$rotation
                                                 PC1         PC2         PC3         PC4
Slave.Age                                -0.058502933  0.06145919  0.68521777 -0.09492308
Transaction.Number.of.Adult.Slaves       -0.578294739 -0.11977020 -0.11401519  0.02959127
Transaction.Number.of.Child.Slaves       -0.120564861  0.06700821  0.69180351 -0.05560227
Transaction.Number.of.Total.Slaves.Purchased -0.588530017 -0.10193098  0.03924440  0.01671075
Transaction.Sale.Details.Discount.Rate    0.098770532 -0.69475868  0.07315472 -0.01828940
Transaction.Sale.Details.Predicted.Interest.Rate  0.098320283 -0.69486589  0.07283407 -0.01806590
Transaction.Sale.Details.Price           -0.530835537 -0.03661332 -0.12356330 -0.02255493
Dates                                    -0.006345977  0.01136095 -0.10686735 -0.99276055
                                                 PC5         PC6          PC7           PC8
Slave.Age                                 0.716103587  0.03823631  0.0001816843  6.304759e-16
Transaction.Number.of.Adult.Slaves        0.096967589 -0.39165985 -0.0002311752 -6.888795e-01
Transaction.Number.of.Child.Slaves       -0.687715508  0.05204114  0.0001104855 -1.537186e-01
Transaction.Number.of.Total.Slaves.Purchased -0.054935402 -0.36958026 -0.0002008333  7.083895e-01
Transaction.Sale.Details.Discount.Rate   -0.006826440  0.04267089 -0.7071076540 -2.747802e-15
Transaction.Sale.Details.Predicted.Interest.Rate -0.006874309  0.04263212  0.7071057773  3.053113e-15
Transaction.Sale.Details.Price            0.030340294  0.83676515 -0.0001722667 -9.020562e-17
Dates                                    -0.028425898 -0.04503863  0.0001293086 -2.116363e-16
```

Calculating the proportion of variability explained by each principal component and the cumulative proportion of variance explained.



Finally, I get the biplot for PC1 and PC2.



Tentative Analysis:

Note: I have refrained from commenting on the signs associated to difference variables in the principal components because these are completely insignificant. My code run on another computer could give inverse signs because principal components found are unique up to a sign. We can, however, comment on which variables have inverse signs.

PC1: (33.6% variability)

- Strong negative weights for:
  - Number of adult slaves (-0.57831736)
  - Number of slaves purchased (-0.58856073)
  - Price of purchase (-0.53075772)
- Small negative weights:
  - Number of child slaves bought (-0.120564861)
- Small Positive weights for:
  - Discount rate (0.098770532)
  - Interest Rate (0.098320283)

Interpretation:

The first PCA seems to confirm our intuitions that the more slaves are bought, the higher the purchasing cost (it however does not match our intuitions that humans are tradeable goods). The price increases when slaves are bought and when numerous are bought at once. We can see that while both the number of adult slaves and the number of slaves are significantly weighted variables, the number of child slaves seems to hold less significant weight in increasing the final price than the other two. This can be explained by the fact that adult slaves were deemed more valuable since they were stronger and more skilled and thus play a larger role in determining the price.

We can see the correlation between the number of adult slaves bought, number of slaves purchased, and the price of the purchase in the biplot which shows that the arrows of all three variables move in the same direction.

***PC2*:** (25% variability)

- Strong negative weights for
  - Interest rate (-0.69486589)
  - Discount rate (-0.69475868)
- Slight negative weights for
  - Number of child slaves (-0.11977020)
  - Number of slaves purchased (-0.10193098)

Interpretation:

Since interest rate and discount rates measure the same thing, that both are both are identically weighted and that their arrows in the biplot point in the exact same direction makes sense.

The slight negative weights associated to the number of child slaves purchased and the total number of slaves purchased makes sense since the number of slaves bought would increase as the number of child slaves are bought, we cannot make any inference about both variable's relationships with each other and the other variables.

PC3: (13% variability explained)

- Strong positive weights to:
    - Slave age (0.68521777)
    - Number of child slaves (0.69180351)
- Medium Negative:
    - Transaction price (-0.12)
    - Number of adults purchased (-0.11)
    - Date (-0.10)

Interpretation:

PC2 identifies an inverse correlation between price and number of child slaves purchased. While the transaction price is not significantly weighted, it further confirms our interpretation of PC1.

The age variability is explained by the fact that few child slaves were sold. On average, slaves that were sold were 27 years old with a std dev of 11.4. Thus, it seems that the variability in the age of slaves sold is explained in part by the number of child slaves sold.

***PC4*:** (12% variability)

- Extremely strong variability:
    - Dates (-0.99)

Interpretation: The dataset was taken from a paper which argued that there was not a significant correlation between the news between 1856 and 1861. In previous PCAs, the date was given little variability. However, from PCA4 we can infer that the Year was responsible for 12% of the total variability. As argued in the paper, extreme years (ie. 1860 and 1861), not monthly news had significant effects on the price of slavery as southern farmers became increasingly anxious by the possibility of a war between the North and the South. However, Principal Component Analysis seems to contest this conclusion. It shows that other variables like the number of slaves bought, the number of adults bought, and the age of slaves varied in the same way as the price. These appear to be the standard variables that drive transaction prices. The year was given little weight in PC1 in which the price was highly weighted and thus seems to vary independently of all the variables that appear to be related to the price of the transaction.

Tentative Conclusion:

The PCA analysis is far from conclusive but does seem to show that slave purchases prices varied in the most part due to variables such as the number bought, and the age of slaves bought. The year represents a tenth of the variability but varies alone. Thus, it seems to not have had a significant impact on slave prices.