Project 3 Report

JessLMurray

12/1/2019

Urban Ministries of Durham (UMD) Data Analysis

Data was wrangled in Python and exported to csv files to be read into R; statistical analysis and plotting was performed in R. The data sets used in this analysis include the following (provided by UMD):

Data Set	Description
Client	information on client age, gender, race, ethnicity, and veteran status
Disable_Entry	information re. disability type & status at time of entry to UMD services
EE_UDES, Entry_Exit	information re. entry and exit interviews conducted with clients, including housing status and history, domestic violence history, length of stay, and reason(s) for leaving UMD
Health_Ins Entry	health insurance type and coverage status for clients at time of entry to UMD services
Income Entry	information re. cash and non-cash sources of income for UMD clients at time of entry to UMD services

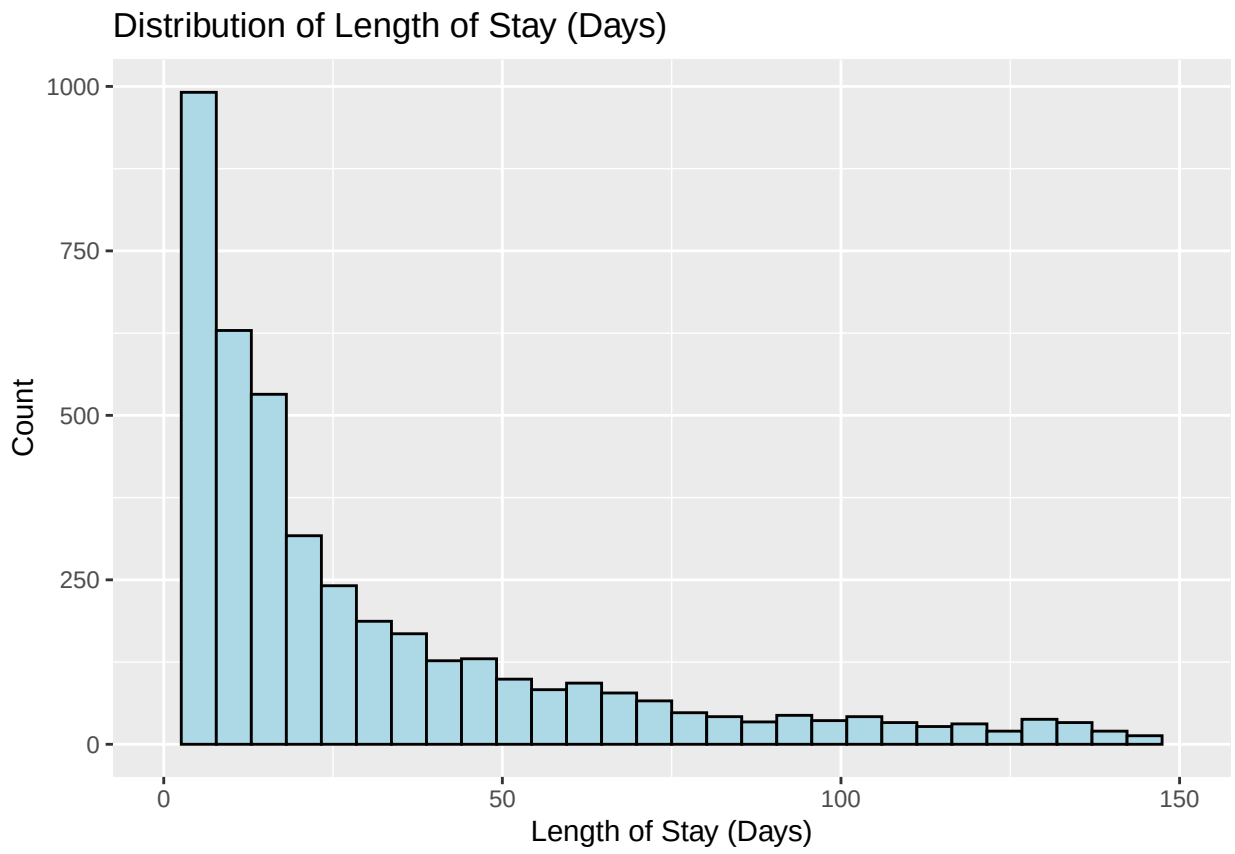
Research Questions

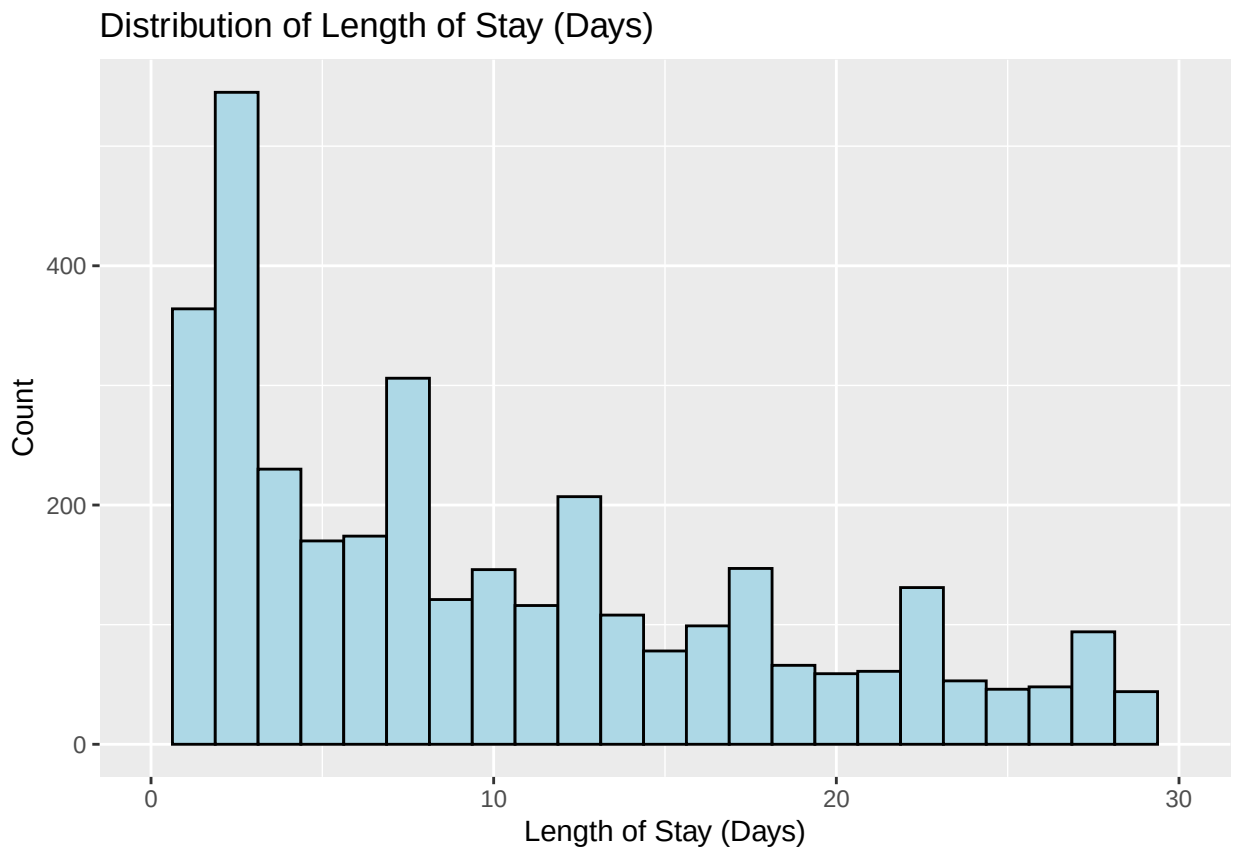
This project will focus on the influence of social determinants of health on the extent of services needed. My research questions are: * What is the general distribution of the clients' age, gender, and race? * Is there a relationship between client demographics and the length of stay at UMD? * Specific demographic variables I am interested in include: age, gender, race, veteran status, disability determination, health insurance coverage, domestic violence victimhood/survivorship, and income source(s).

Exploratory Data Analysis

Distribution of Length of Stay (LOS) at UMD

Here is a simple histogram showing the distribution of length of stay at UMD. I also made a 'zoomed in' version that includes only clients who stayed between 0 and 30 days at UMD:


Distribution of Length of Stay (Days)
Count
1000
750
500
250
0
0
50
100
150
Length of Stay (Days)

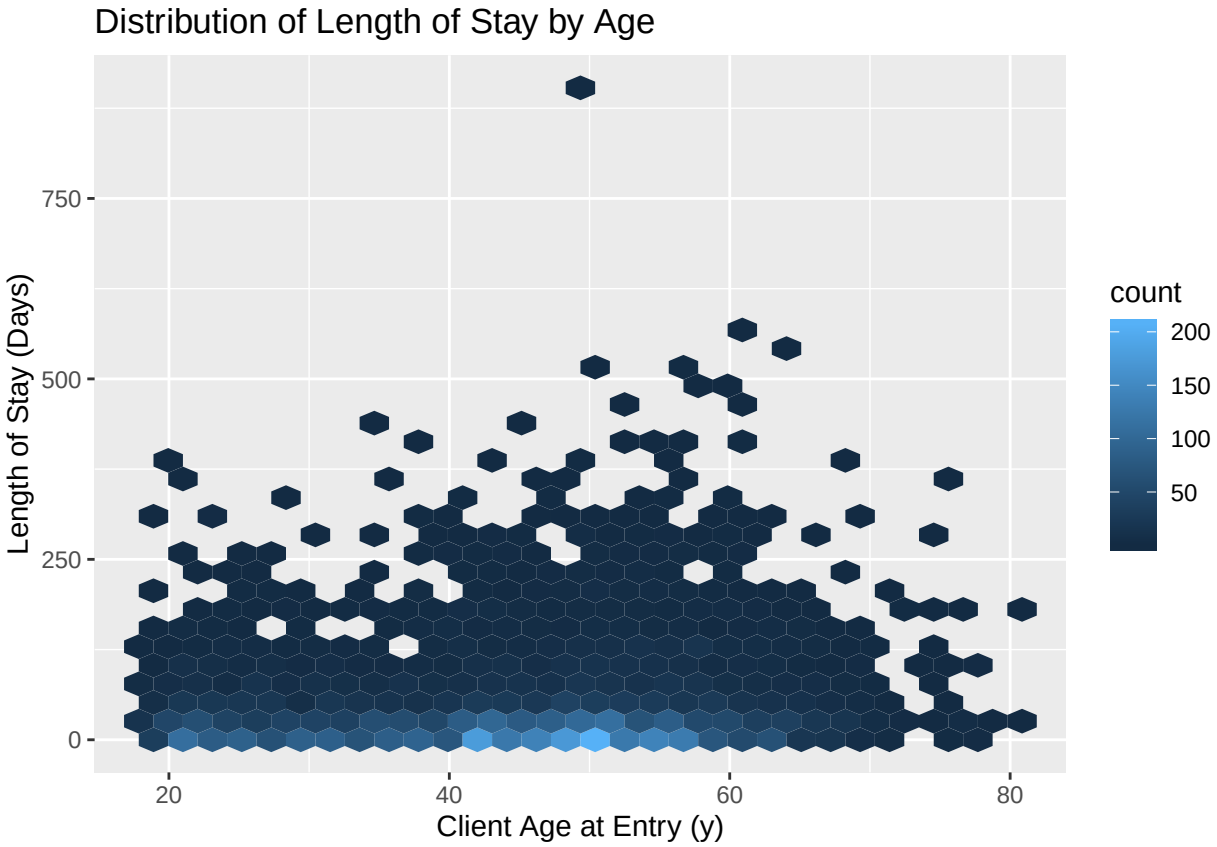



```
summary(entry_exit_LOS$'LOS_days')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00    5.00   16.00   39.59   45.00  894.00     107
```

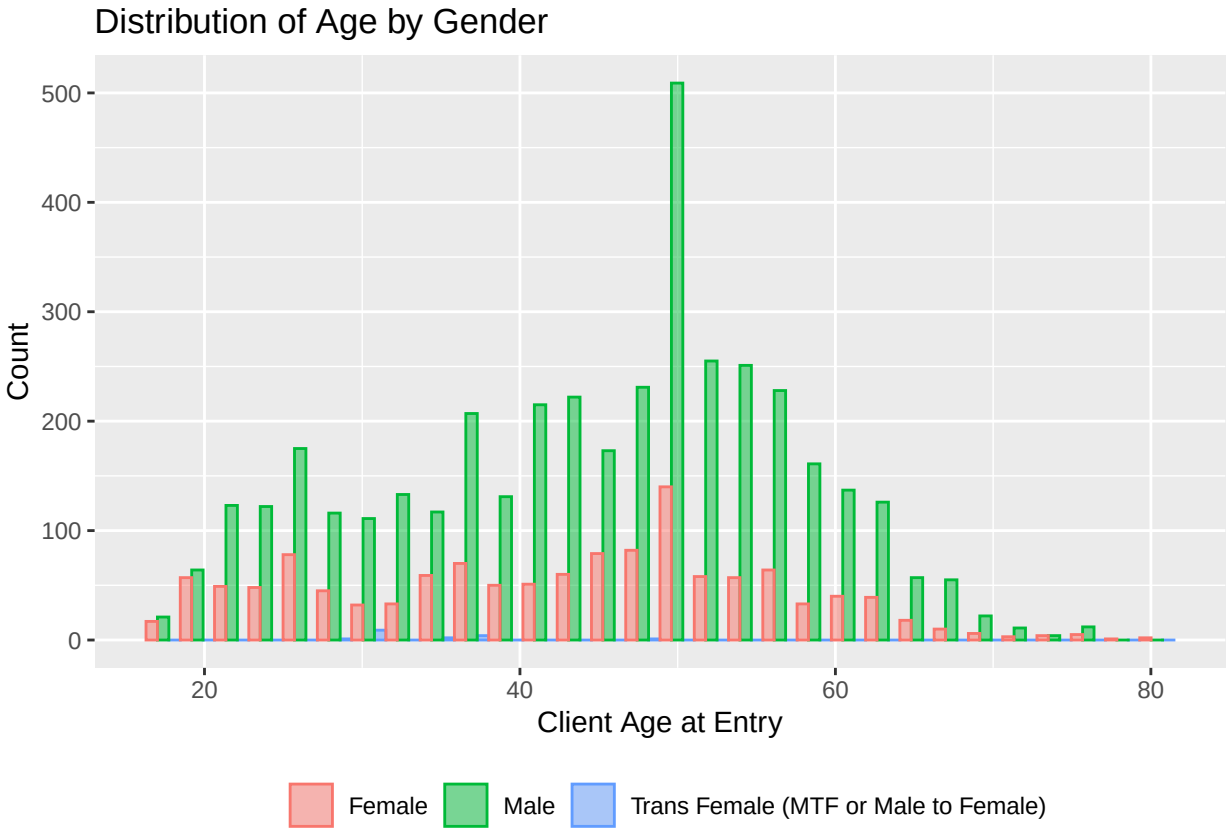
The average length of stay (LOS) at Urban Ministries is **40 days** with range of **0-894 days**. The **median LOS = 16 days** (it is best to use the median since there is right skew to the LOS distribution).

Distribution of Length of Stay by Age



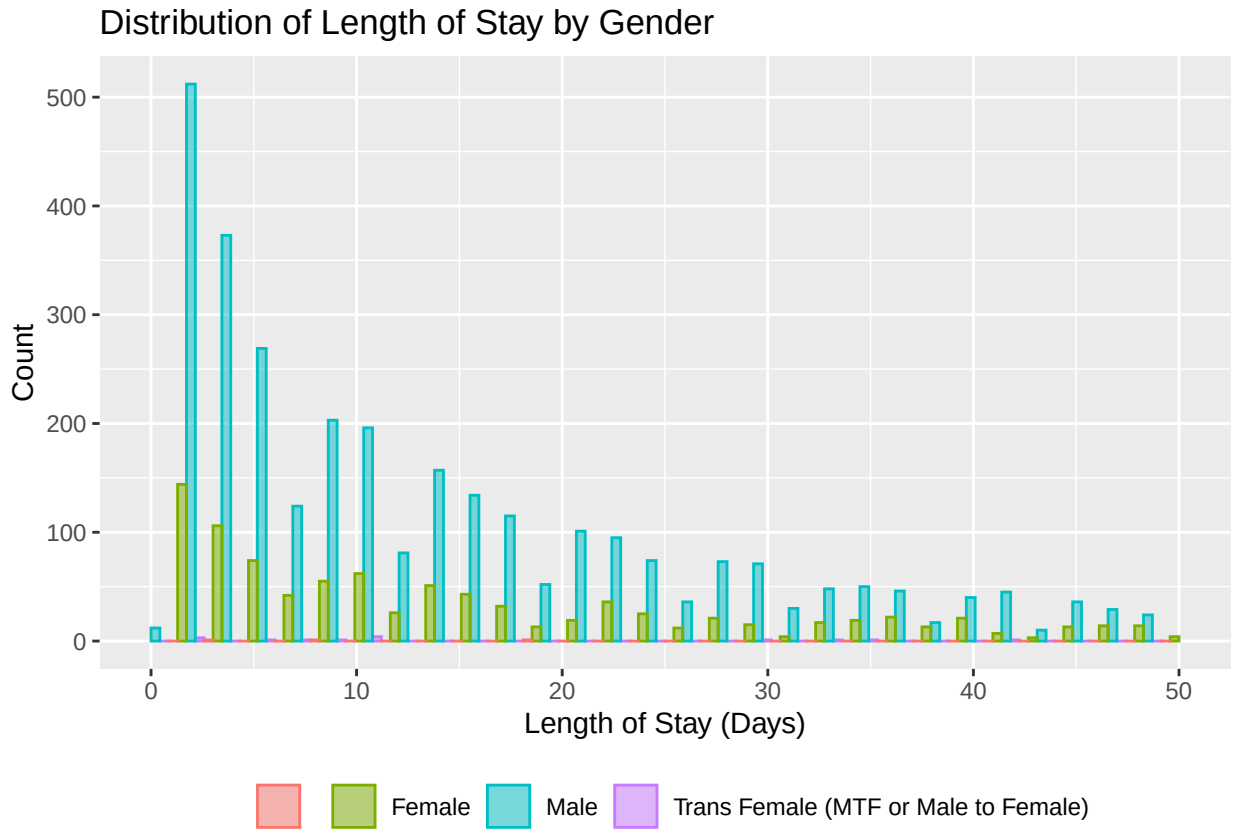
There does not appear to be any relationship between the length of stay and the age of UMD clients.

Are There Differences in Age Distribution by Gender?



More men than women appear to use UMD's services overall; however, there appears to be no difference in the age distribution between the genders (i.e., the average age of women at UMD does not appear to be very different from the average age of men at UMD). There appears to be a bimodal distribution in the ages of those who use UMD's services for both genders, with one peak at around age 25 and another peak at around age 50.

Is Length of Stay Associated with Gender?




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   18.00   35.00   46.00   44.24   54.00   81.00       3
```

```
##
##                                                    Female
##                         0.0005661446                        0.2434421589
##                                 Male Trans Female (MTF or Male to Female)
##                         0.7527835441                        0.0032081525
```

The average age of UMD clients is 44 years old, with a range of 18 - 81. Approximately 75% of clients are male, and approximately 24% of clients are female (17 clients in the data set are transgender, and information is not provided for some clients). Based on the distribution, the average length of stay does not appear to be affected by gender.

Is Length of Stay Associated with Race?

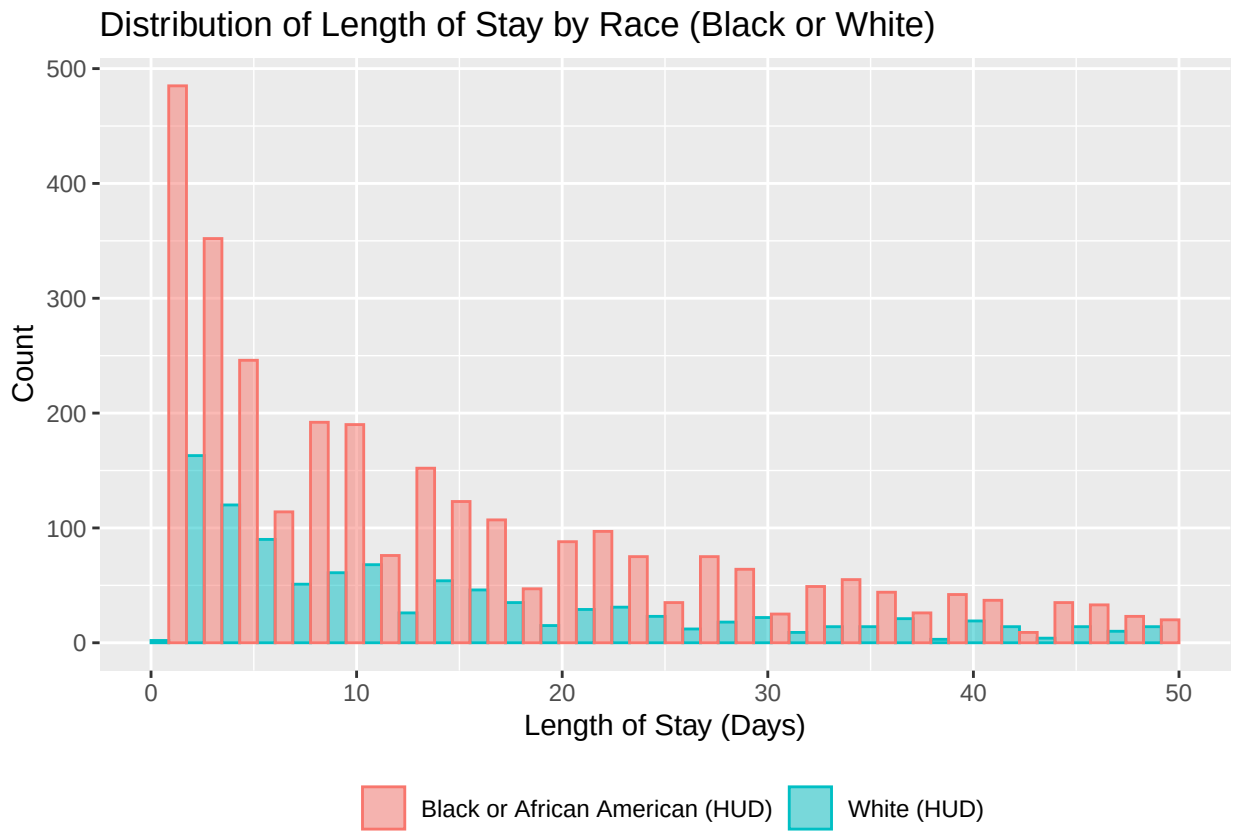
```
racetable = table(na.omit(client$`Client Primary Race`))
prop.table(racetable)
```

```
##
```

```
##
##                                     0.0005661446
##          American Indian or Alaska Native (HUD)
##                                     0.0169843367
##                                      Asian (HUD)
##                                     0.0005661446
##                  Black or African American (HUD)
##                                     0.7291941876
##                        Client doesn't know (HUD)
##                                     0.0007548594
##                             Client refused (HUD)
##                                     0.0009435743
##                           Data not collected (HUD)
##                                     0.0011322891
## Native Hawaiian or Other Pacific Islander (HUD)
##                                     0.0020758634
##                                      White (HUD)
##                                     0.2477826005
```

Approximately 72.9% of clients are black, and 24.8% of clients are white. 1.7% of clients are American Indian or Alaskan Native.

I am interested in knowing if there is a racial difference in the average length of stay. Since the vast majority of clients are black or white, I will remove clients who are not black or white so I can more easily visualize racial differences (code below). *Note: This analysis will therefore be limited by the exclusion of other races.*



I performed a t-test to see if there is a significant difference in the average number of days spent at UMD by race.

```
##
##  Welch Two Sample t-test
##
## data:  LOSbw$LOS_days by LOSbw$`Client Primary Race`
## t = 1.6397, df = 2389.7, p-value = 0.1012
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6227479  6.9791063
## sample estimates:
## mean in group Black or African American (HUD)
##                                     40.36397
##                     mean in group White (HUD)
##                                     37.18579
```

T-test results indicate that there is no significant difference in the average length of stay for black clients vs. white clients:

Length of Stay (black clients)	Length of Stay (white clients)	p-value
40.4 days	37.2 days	0.1012

Is Length of Stay Associated with Veteran Status?

```
vettable = table(na.omit(client$`Client Veteran Status`))
prop.table(vettable)
```

```
##
##                          Data not collected (HUD)                  No (HUD)
##              0.001887149              0.001321004              0.891488960
##                Yes (HUD)
##              0.105302887
```

Approximately 89% of the 5,299 clients in the 'client_LOS' data set are not veterans and approximately 11% of clients are veterans. Is there a significant difference in the length of stay for veterans vs. non-veterans?

```
##
##  Welch Two Sample t-test
##
## data:  vet_LOS$LOS_days by vet_LOS$`Client Veteran Status`
## t = 0.48725, df = 333.82, p-value = 0.6264
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.882415  9.756091
## sample estimates:
##  mean in group No (HUD) mean in group Yes (HUD)
##                40.13460                38.19776
```

There is no significant difference in the average length of stay based on veteran status:

Length of Stay (veterans)	Length of Stay (non-veterans)	p-value
37.8 days	39.8 days	0.4619

Is Length of Stay Associated with Disability Determination?

About 88% of clients are not disabled, and about 12% of clients have some form of disability. Does disability status impact the length of stay at UMD?

```
##
##  Welch Two Sample t-test
##
## data:  disable_LOS$LOS_days by disable_LOS$`Disability Determination (Entry)`
## t = 2.7207, df = 18767, p-value = 0.00652
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.3492626 2.1493718
## sample estimates:
##  mean in group No (HUD) mean in group Yes (HUD)
##                31.73123                30.48192
```

There is a very small yet significant difference in length of stay based on disability status which suggests that abled persons spend longer at UMD on average:

Length of Stay (disabled persons)	Length of Stay (abled persons)	p-value
31.1 days	31.8 days	0.0222

Though this finding is technically statistically significant, it is likely not relevant or useful to those at UMD since the overall difference is less than 1 day. If I were to construct a multivariable model predicting length of stay using demographic variables as predictors, I would **not** include disability status as a predictive variable based on these results.

Is Length of Stay Associated with Health Insurance Coverage?

```
##
##                    Data Not Collected                  No
##       0.0001327962       0.0002434598       0.9428090833
##                Yes
##       0.0568146607
```

94.3% of clients do not have health insurance, and approximately 5.6% of clients have coverage.

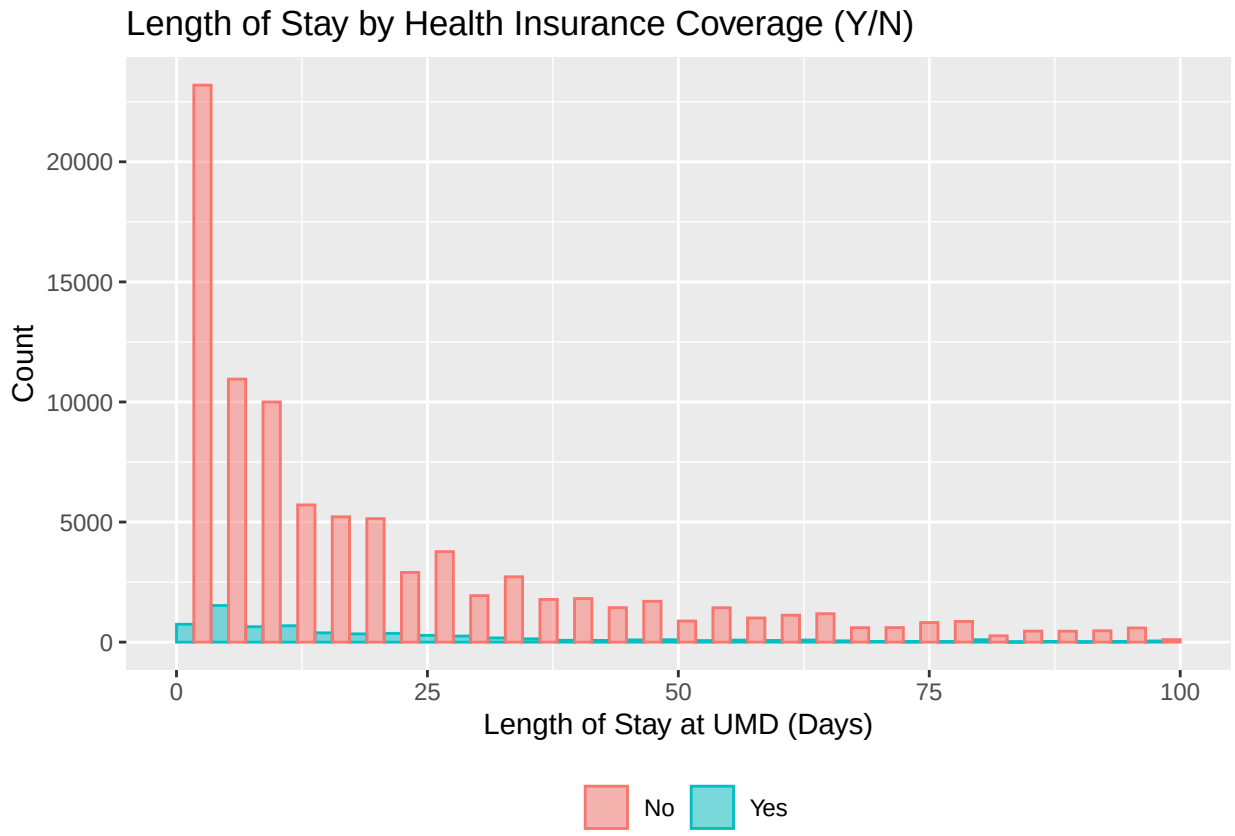
```
##
##  Welch Two Sample t-test
##
## data:  ins_LOS$LOS_days by ins_LOS$`Covered (Entry)`
## t = 5.6162, df = 8344.6, p-value = 2.015e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.073209 4.296460
```

```
## sample estimates:
##  mean in group No mean in group Yes
##          32.25522           29.07039
```

Clients without health insurance stay about three days longer at UMD on average compared to clients with health insurance:

Length of Stay (health insurance)	Length of Stay (no health insurance)	p-value
29.6 days	32.3 days	6.05e-11

Clients without health insurance may have a greater number of medical expenses compared to those with insurance, which may contribute to a longer stay at UMD. Insurance coverage may also serve as a general indicator of overall wealth, which would suggest that clients with more wealth spend less time at UMD.

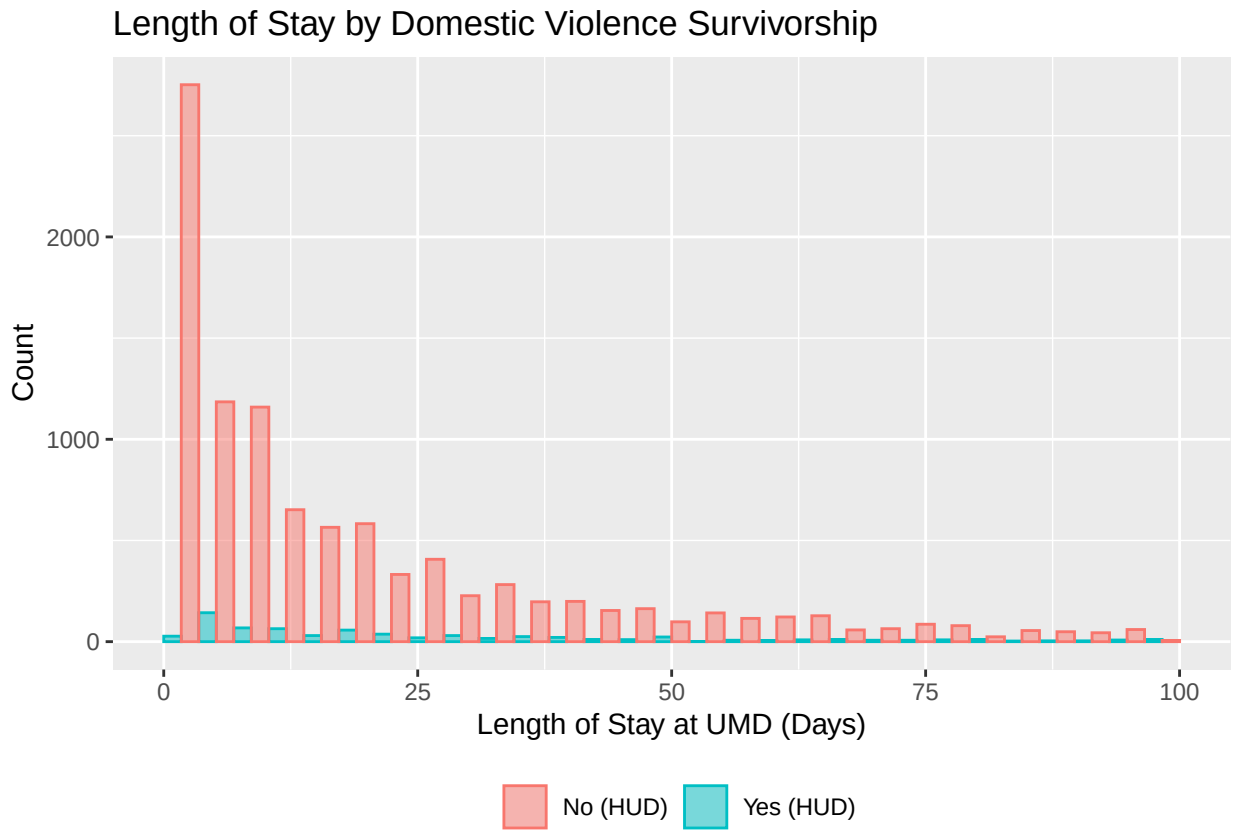


Is Length of Stay Associated with Domestic Violence Victimhood/Survivorship?

Approximately 9.4% of UMD clients are domestic violence victims/survivors.

```
##
##  Welch Two Sample t-test
##
## data:  vio_LOS$LOS_days by vio_LOS$`Domestic violence victim/survivor(341)`
## t = -3.906, df = 853.96, p-value = 0.0001013
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.048437  -3.658326
## sample estimates:
##  mean in group No (HUD) mean in group Yes (HUD)
##                30.43015                 37.78353
```



Clients who have experienced domestic violence stay approximately 8 days longer at UMD compared to those who are not victims of domestic violence:

Length of Stay (domestic violence victims/survivors)	Length of Stay (no history of domestic violence)	p-value
38.2 days	30.5 days	9.59e-9

I believe this large difference in length of stay is indicative of the fact that domestic violence is often correlated with homelessness. This information may be useful to UMD staff: based on this trend, clients who are victims/survivors of domestic violence could be flagged to receive additional follow-up and/or assistance from UMD.

Is Length of Stay Associated with Income Status?

```
##
##                   Data Not Collected                    No
##       0.0003264816       0.0004309557       0.9660067386
```

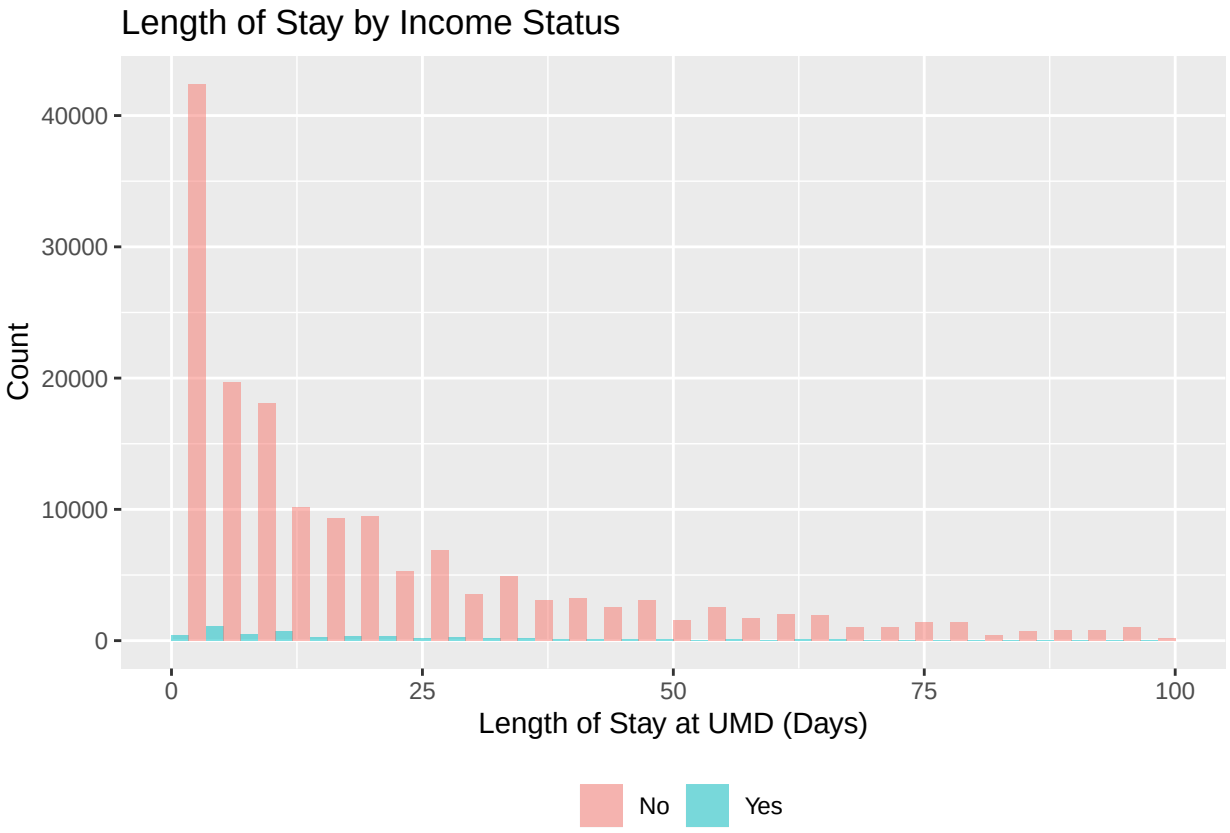
```
##                 Yes
##       0.0332358242
```

About 96.6% of UMD clients are not receiving any form of income at entry.

```
##
##  Welch Two Sample t-test
##
## data:  income_LOS$LOS_days by income_LOS$`Receiving Income (Entry)`
## t = -7.0027, df = 6924.6, p-value = 2.747e-12
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.852019 -3.292211
## sample estimates:
##  mean in group No mean in group Yes
##          31.73913           36.31124
```

Clients with some form of income stay about four more days at UMD on average compared to clients with no income:

Length of Stay (receiving income)	Length of Stay (no income source)	p-value
35.4 days	31.7 days	2.537e-16



This is an interesting and unexpected finding.Further analysis could be performed to explore the source

and amount of income for clients, and whether different forms of income have different relationships to the length of stay. It is possible that the income source is too small to allow clients to live independently, or the fact that the client(s) have an income source disqualifies them from receiving benefits given to those with no income. It is also possible that a lack of income may serve as a barrier to receiving services from UMD (for example, a client who cannot afford to pay a bus fare to return to UMD from across town may be counted as leaving UMD's service). More information from clients and staff at UMD may provide some insight into this finding.

Summary Table & Conclusions

Demographic Variable Tested	Relationship to Length of Stay (LOS)?	Finding
Age	No	Age did not appear to be associated with LOS based on visual inspection of distribution
Gender	No	Gender was not associated with LOS based on visual inspection of distribution
Race	No	No significant difference in LOS by race (black or white)
Veteran Status	No	No significant difference in LOS by veteran status
Disability Determination	Yes*	*Very small yet significant difference in LOS in favor of disabled clients; likely not a 'real' difference
Health Insurance Coverage	Yes	Significantly longer LOS for clients without health insurance (32.3 days vs. 29.6 days, p = 6.05e-11)
Domestic Violence Victimhood/Survivorship	Yes	Significantly longer LOS for clients who are victims/survivors of domestic violence (38.2 days vs. 30.5 days, p = 9.59e-9)
Receiving Income	Yes	Significantly longer LOS for patients with a source of income (35.4 days vs. 31.7 days, p = 2.54e-16)

Demographic variables that appeared to influence the length of stay at UMD include: **health insurance coverage, domestic violence victimhood/survivorship, and presence of an income source**. These demographic data may be used by UMD staff to inform decisions regarding resource allocation for specific clients. This analysis was limited by the lack of knowledge surrounding the data, i.e. how the data was collected, and how the variable responses should be interpreted. With additional time, I would like to investigate the nature of the relationship between length of stay and income source, as well as the possible relationship between SPDAT survey responses and length of stay.