

Project 1 Report: An Analysis of Services Provided by Urban Ministries of Durham

Jessica Murray

Urban Ministries of Durham Data Analysis

How many clients/families use services provided by UMD?

To begin answering this question, I first needed to clean the data. I removed all rows with dates prior to 1983 (when UMD was founded) and after the date when I began this analysis (Sept 15, 2019) - these dates are clearly typos.

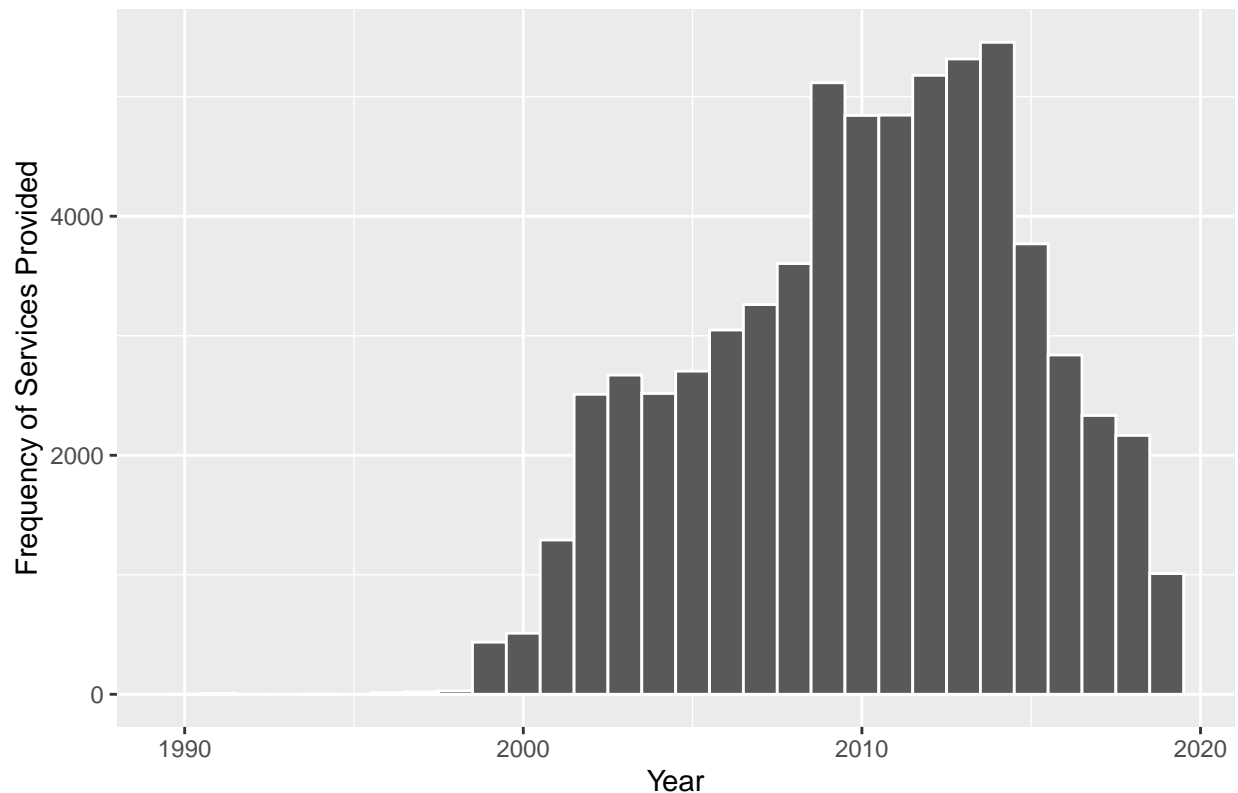
```
#remove all rows with mistyped dates
library(dplyr)
umdclean <- umd %>%
  filter(Date >= "1983-01-01") %>%
  filter(Date <= "2019-09-15")
nrow(umdclean)
```

```
## [1] 65498
```

There were **65,498** instances of services provided in this dataset! Let's plot the number of services provided over time:

```
ggplot(umdclean, aes(x=year)) +
  geom_histogram(color = "white") +
  labs(x = "Year", y = "Frequency of Services Provided",
       title = "Number of Times Service was Provided by UMD")
```

Number of Times Service was Provided by UMD



Based on this histogram, it looks like most services weren't provided or accounted for prior to around the year 2000. I will therefore focus further analysis on data from the past 20 years (1999 to 2019).

```
umd20 <- umdclean %>%
  filter(Date >= "1999-01-01")
```

While the raw number of times service was provided is interesting, it doesn't tell us the number of *unique* clients/families that were assisted by UMD. We must find the number of unique client file numbers to do this:

```
#How many clients/families use services provided by UMD?
#find the number of unique client file numbers

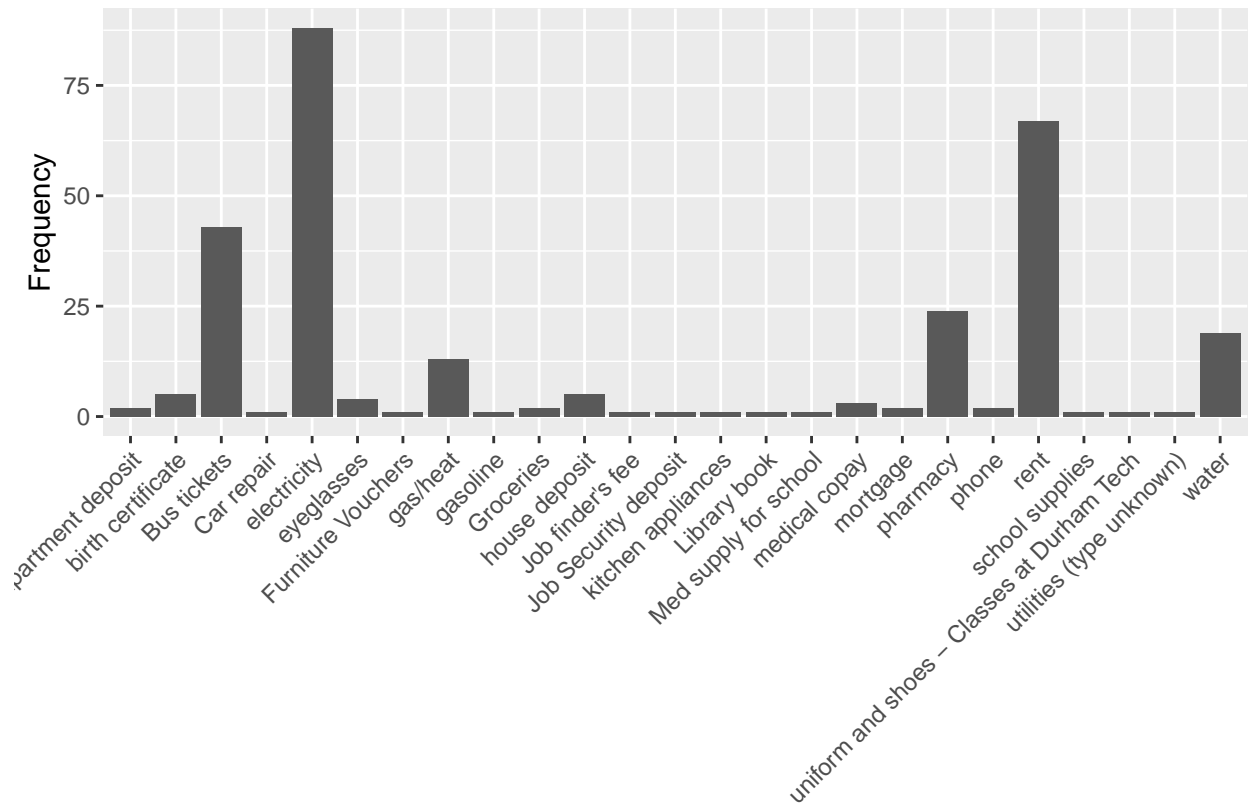
umdunique <- unique(umd20$`Client File Number`)
length(umdunique)
```

```
## [1] 11228
```

A total of **11,228** clients and families have been aided by UMD in the last 20 years!

Which monetary services are provided most frequently?

To determine the frequency of *monetary* services provided most often, I made a bar plot of the type of bills paid by category:



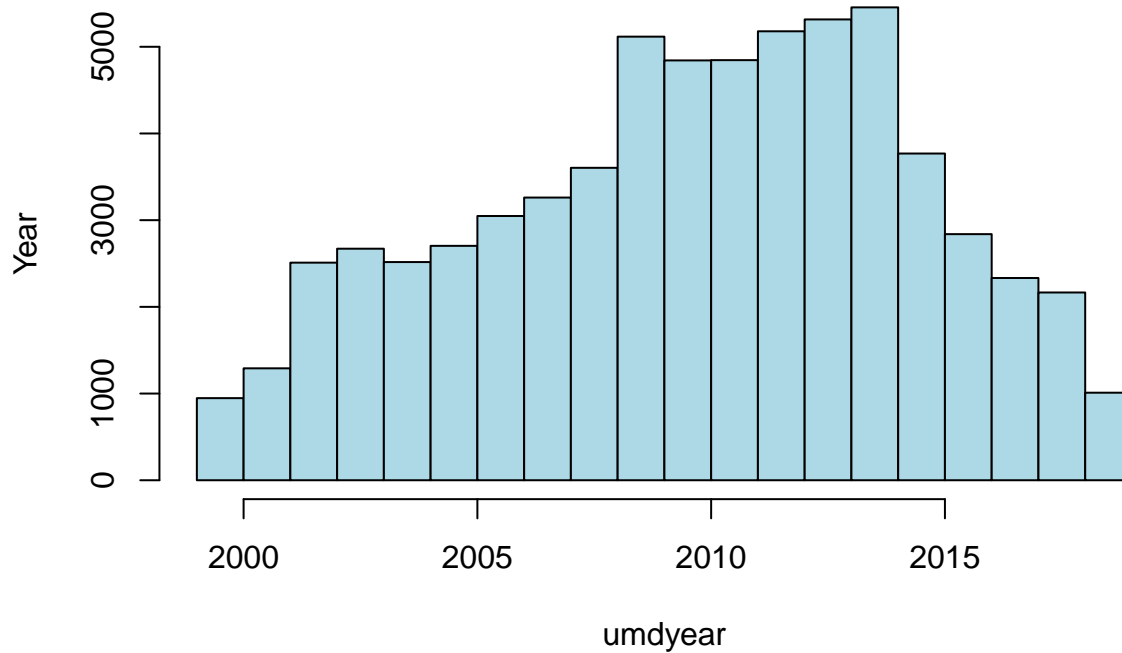
Reason for Monetary Donation from UMD

It appears that for monetary donations provided, electricity bills, rent, and bus tickets were the most common. While we could guess that rent would cost the most followed by electricity bills and bus tickets, we cannot determine which bills/monetary donations are the most expensive without knowing the amount paid in each category. For example, although pharmacy bills are not one of the most frequent categories, it may still be the most expensive depending on the cost of prescriptions purchased without insurance.

Are there any long-term trends in the quantity of services provided?

To answer this question, I created a histogram that displayed the frequency of the number of clients/families served by year:

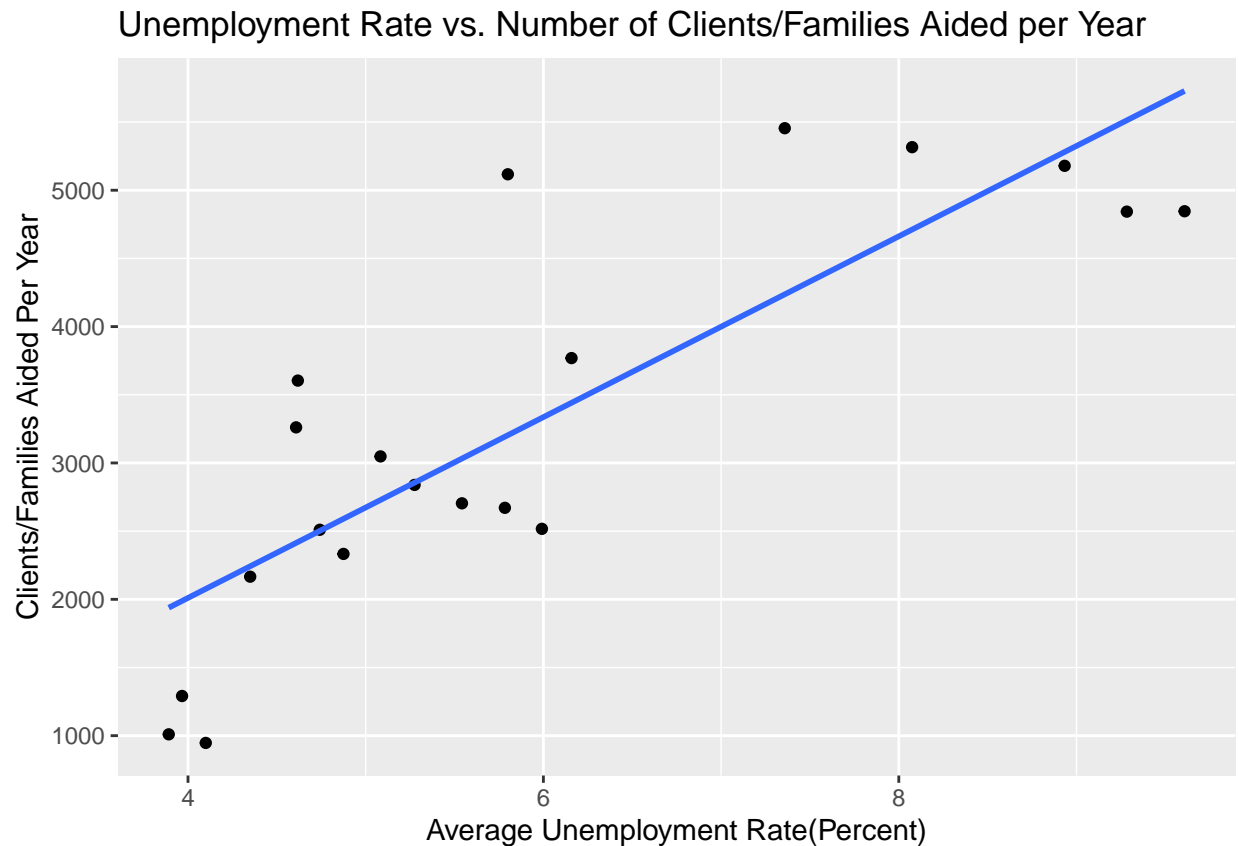
Frequency of Clients/Families Aided by UMD per Year



It appears that the number of services provided tends to increase from 1999 until around 2013, and then decreases from around 2013 to the present. One might guess that the sharp uptick in services provided in 2008 may correlate to the U.S. recession that began in the same year. To test whether unemployment rates are related to the amount of aid given, we may conduct a correlation analysis that compares the unemployment rate to the number of clients/families aided by UMD each year. Unemployment data was obtained from the Bureau of Labor Statistics.

##	Clients.Families.Aided.Per.Year	Unemployment.Rate..Percent.
## 1	947	4.100000
## 2	1291	3.966667
## 3	2510	4.741667
## 4	2671	5.783333
## 5	2517	5.991667
## 6	2704	5.541667
## 7	3048	5.083333
## 8	3261	4.608333
## 9	3604	4.618182
## 10	5117	5.800000
## 11	4843	9.283333
## 12	4846	9.608333
## 13	5179	8.933333
## 14	5316	8.075000
## 15	5455	7.358333
## 16	3769	6.158333
## 17	2839	5.275000
## 18	2333	4.875000

```
## 19          2166          4.350000
## 20          1010          3.891667
```



There appears to be a positive correlation between the national unemployment rate and the number of clients/families helped per year by UMD. Let's find the R-squared value for this data:

```
##
## Call:
## lm(formula = blscorr$Clients.Families.Aided.Per.Year ~ blscorr$Unemployment.Rate..Percent.)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1129.66  -676.57   -89.02   397.00  1913.42
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -641.2     661.7  -0.969   0.345
## blscorr$Unemployment.Rate..Percent.    662.9    107.5   6.169 7.99e-06
##
## (Intercept)
## blscorr$Unemployment.Rate..Percent. ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 843.8 on 18 degrees of freedom
## Multiple R-squared:  0.6789, Adjusted R-squared:  0.6611
```

F-statistic: 38.06 on 1 and 18 DF, p-value: 7.989e-06

Our summary statistics demonstrate that there is a significant positive correlation between the national unemployment rate and the number of clients/families who were helped by UMD ($R^2 = 0.6789$, p-value = 8.0×10^{-6}).

Conclusions

To conclude, we gained the following insights from analysis of the data:

- Most clients and families were served in the last 20 years as opposed to earlier in UMD's history. This may be due to an improvement/change in recordkeeping over time, or due to an increase in UMD's profile due to special events and/or word-of-mouth advertising.
- The the most common reasons for monetary donations from UMD were the following: electricity bills > rent > bus tickets.
- The national unemployment rate provided by the Bureau of Labor Statistics is significantly positively correlated with the number of services provided by UMD from 1999-2019.