Who will donate? A study on Canadian donation statistics

Taojun Wang, Yu Yuan, Biwen Zheng

October 17, 2020

Executive Summary

Donation is a blessing and a gift for charitable purpose. In this study, we investigate the Canadian donation statistics and observe some interesting pattern of donors, that is, older, rich widow or widower with a high education level, are more generous and enthusiastic on donation. The reason of this scenario could be that the young and families are limited on financial ability. Additionally, we discuss about further modification by considering the total number of children as a factor that might be associated with the amount of donate value.

Introduction

Our research focus is the donation of Canadians in 2013. A person's donation choice can be influenced by many factors, including household income, education, gender, religious faith and so on. Some of them will decide whether people donate or not, and some others might affect the amount people finally donate. Our report is aimed to figure out how these factors determine people's donation choice, and construct a prediction model based on the data. We hope the model can explain and predict people's donation behavior, which might help charity organizations to target their donors. We expect that our model will help organizations to target middle-upper class families in their 40s with higher education. They usually have a successful life and generous attitude to give back to the community. We expect the single to donate the least as they are mostly young and with limited financial ability.

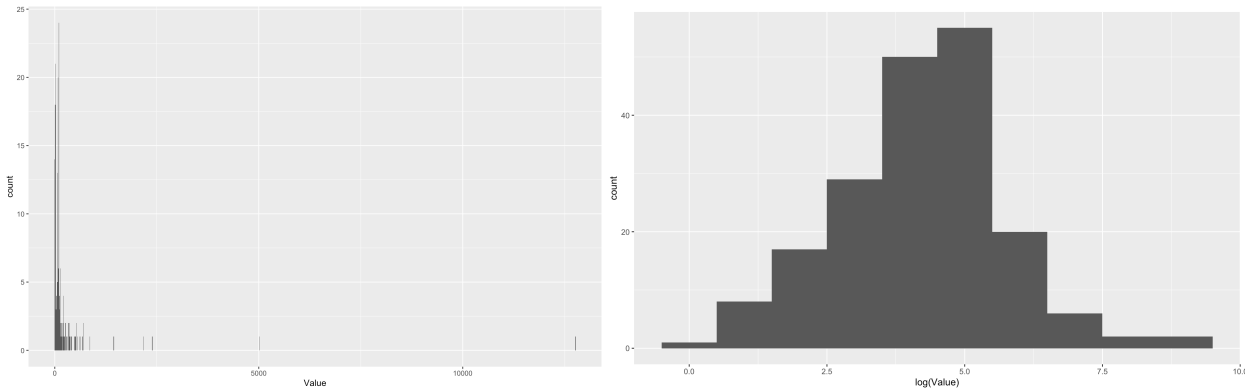
Methodology

The target population is all Canadian citizens and permanent residents aged 15 or older. To sample, the survey uses telephone questionnaires to get responses from only one eligible member from each household. Voluntary participants will get a long interview, while the non-voluntary ones will be divided into two groups, among whom will get long and short interviews respectively. The survey sampling is based on a stratified design, where each province is a stratum. Since different provinces have quite different populations and patterns of donor behavior, stratified sampling ensures that our sample can well represent the target population. The survey question includes their age group, gender, marital status, education level, labor force status, household income and religious attendance. Participants will be divided into different groups according to their answers to each question, and the average annual donations will be calculated. We will use the linear regression model to fit our data, and find the relationship between the average annual donations and those variables.

We consider to constrcut multiple linear regressions with a response variable as donate value and a single variable we are interested. For instance, for discovering the relationship between Sex and donate value, we have a model as

$$log\ (Value_i) = \beta_0 + \beta_1 Sex_i$$

Figure 1. Distribution of Donate Values (original vs, logarithm)

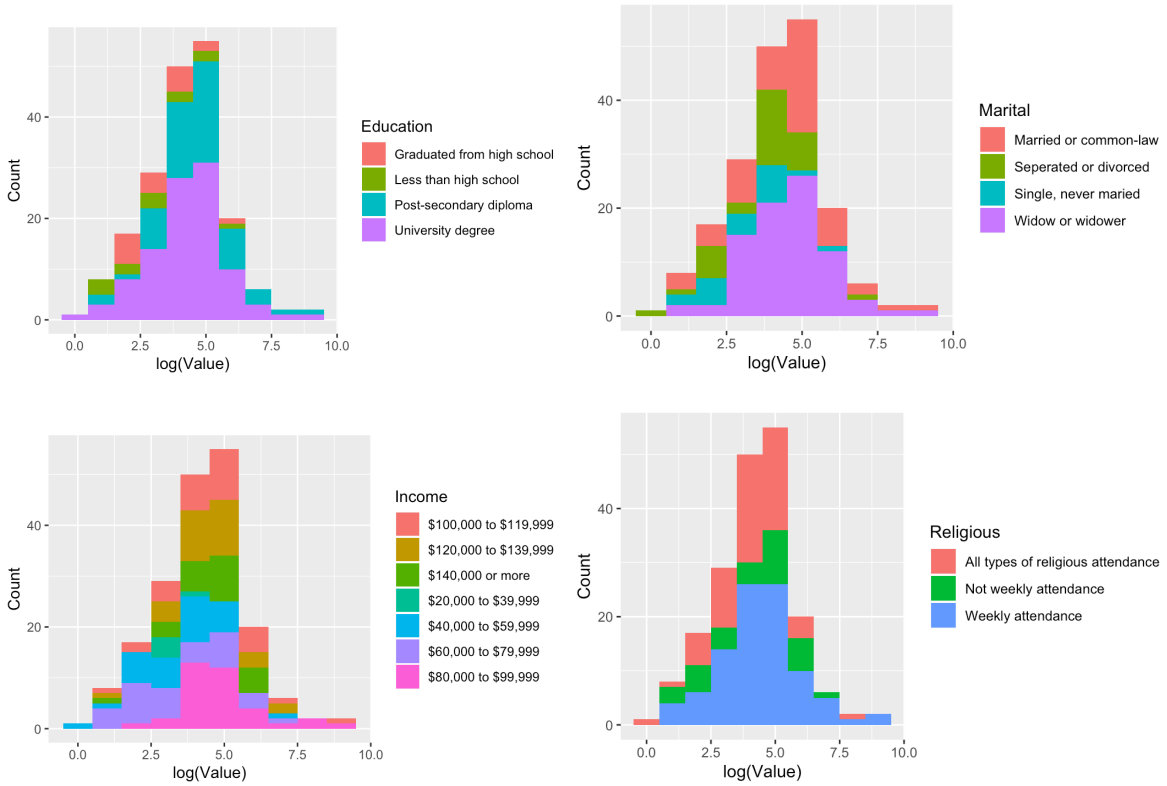


Results

The average value of donation is 253.13 CAD, and the median value is 83.70 CAD. The histogram of original donation values show that it is a highly right skewed distribution with sufficient outliers that may affect our later modeling. Therefore, we transformed our original donation values by taking logarithm. The distribution for new response variable is approximately bell shaped from the second figure. Additionally, we can obtain that most sampled people's donation amount is in the range from 42.52 CAD to 148.41 CAD.

Figure 2. The distribution of independent variables

Figure 2. The distribution of independent variables (Continued)



Furthermore, people aged from 35 to 65 donate the most, and on average the amount that people donate increases linearly as the age increases. The average annual donations of male is 20% more than that of females, but for low-amount donors, genders make very little difference. Also, females, compared to male, seem to be more enthusiastic to donate (higher donor rate). In terms of education, people with higher diplomas tend to donate more. For people who do not have a university degree (i.e. those who graduate from secondary school or less, who graduate from high school or other post-secondary diplomas), the average annual donated value increases slightly and is still approximately in the same level. When a person gets a university degree, he is very likely to donate much more than those who do not. In addition, Household income influence affects people's donation behavior. Median annual donations keeps increasing as the household income increases. The effect is relatively small and shows a linear pattern when the income is relatively low (< $12000). After reaching this threshold, donations begin to increase rapidly and approximately show an exponential pattern. Labor force status is another factor we study. People who are employed have the largest proportion of donors, and highest annual donation value. People who are not in the labor force, including full-time students and retired people are closed to employed ones on average, and have a lower donor rate and larger distribution spread.

Findings on the marital status variable shows people who lose their couples (i.e. widow or widower) tend to donate the most of all, however the cofficient is not statistically

significant at $\alpha = 0.05$ level. Married people rank second in donation amount but they are most likely to donate of all. Divorced people rank third, and single people show the most negative attitude in donations. Finally, we find that there is no obvious difference in the amount of donation between people who have weekly religious attendance and people who do not regardless of the donor rate or the average donations. Generally, when a charity organization wants to target people with higher probability to donate, and with a larger tendency to donate generously. They should target an old, rich widow or widower with a high education level.

```
# single covariate model (example of sex full codes attached in Appendix)
fit.sex <- lm(log(Value) ~ Sex, data = df_tc)
summary(fit.sex)

##
## Call:
## lm(formula = log(Value) ~ Sex, data = df_tc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1712 -1.0264  0.2298  0.7330  5.1008
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.4152     0.7562   3.194  0.00165 **
## SexFemales    1.6673     0.7751   2.151  0.03276 *
## SexMales      1.9383     0.7702   2.517  0.01269 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.512 on 187 degrees of freedom
## Multiple R-squared:  0.03683,    Adjusted R-squared:  0.02652
## F-statistic: 3.575 on 2 and 187 DF,  p-value: 0.02995
```

Weakness

There are several weaknesses we find after constructing our model. First of all, the married ranked second after widow or widower. This is a bit of a surprise for us, as we expected that a happily married family would donate the most among all marital statuses. Thus, we concluded a few possible reasons why widow and widower donate more: they know the pain of losing someone, so they are more generous towards helping others so others' loved ones won't suffer the same tragedy; they might inherit some fortune from their deceased spouses, and spending the fortune on charity is a way of comforting themselves that they did good things with their deceased spouses' fortune. Other than that, the single ranked last when donating, and this is consistent with our expectation.

Secondly, we think there should be another variable: number of kids they have. When people donate, they are helping someone or someone's kids through a hard time. Sympathy is the best incentive for people to be generous. If someone with kids sees other kids in poverty living a struggling life, he/she is more likely to donate to the child charity organization to help out they way they can.

Last but not least, the p-value of most regression models we get is a bit large, which indicates a lack of linearity, so our model might not perform well in prediction. Since a lot of variables are categorical rather than quantitative, a linear regression model might not be a good choice to reveal the characteristics of data.

Reference

- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- General Social Survey: An Overview, 2019, Statistics Canada, https://www150.statcan.gc.ca/
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.
- Weisberg, S. (2005). Applied linear regression (Vol. 528). John Wiley & Sons
- Hossain, B., & Lamb, L. (2012). Does the effectiveness of tax incentives on the decision to give charitable donations vary across donation sectors in Canada?. Applied Economics Letters, 19(15), 1487-1491.

Appendix

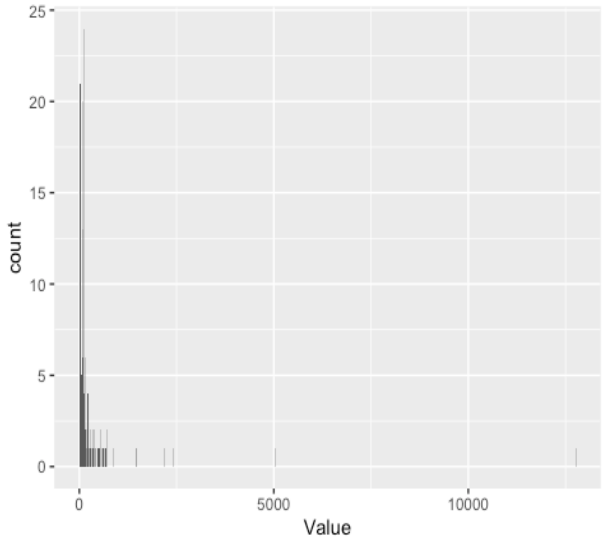
R codes

```
setwd("~/Desktop/Data/STA304")
df_tc <- read.csv("donation data.csv")
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────
──────────── tidyverse 1.3.0 ──

## ✓ggplot2 3.2.1     ✓purrr   0.3.3
## ✓tibble  2.1.3     ✓dplyr   0.8.4
## ✓tidyr   1.0.2     ✓stringr 1.4.0
## ✓readr   1.3.1     ✓forcats 0.4.0

## ── Conflicts ────────────────────────────────────────────────────────
─────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

# Distribution of Donation

df_tc %>%
```

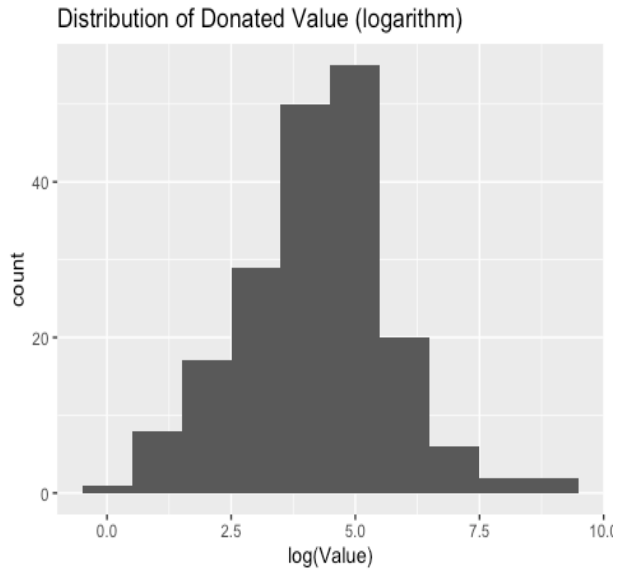
```
  select(Value) %>%
  summary()
```

```
##      Value
##  Min.   :    1.20
##  1st Qu.:   18.65
##  Median :   83.70
##  Mean   :  253.13
##  3rd Qu.:  131.50
##  Max.   :12763.60
```

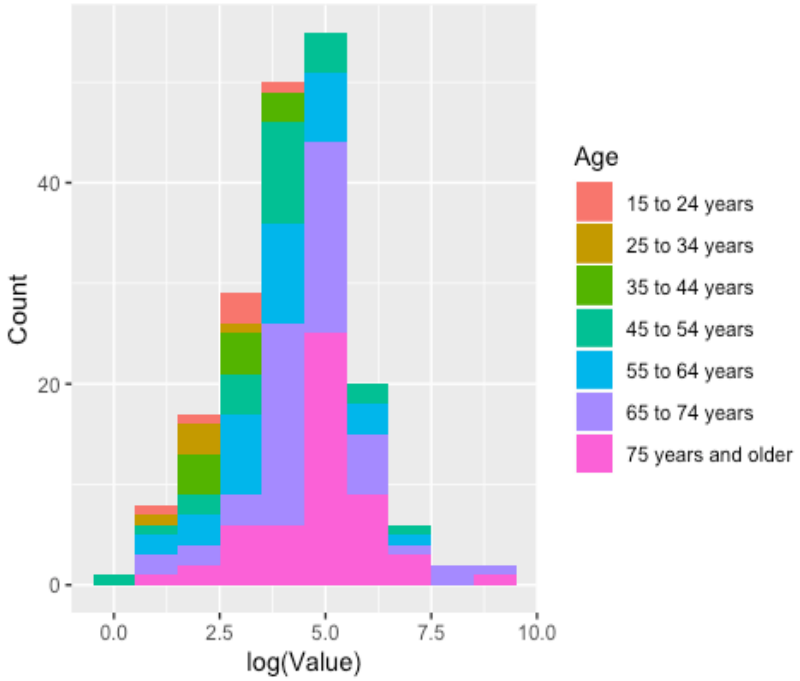
```
df_tc %>%
  ggplot(mapping = aes( x = Value)) +
  geom_histogram(binwidth = 10)
```




```
df_tc %>%
  ggplot(mapping = aes( x = log(Value))) +
  geom_histogram(binwidth = 1) +
  labs(title = "Distribution of Donated Value (logarithm)")
```

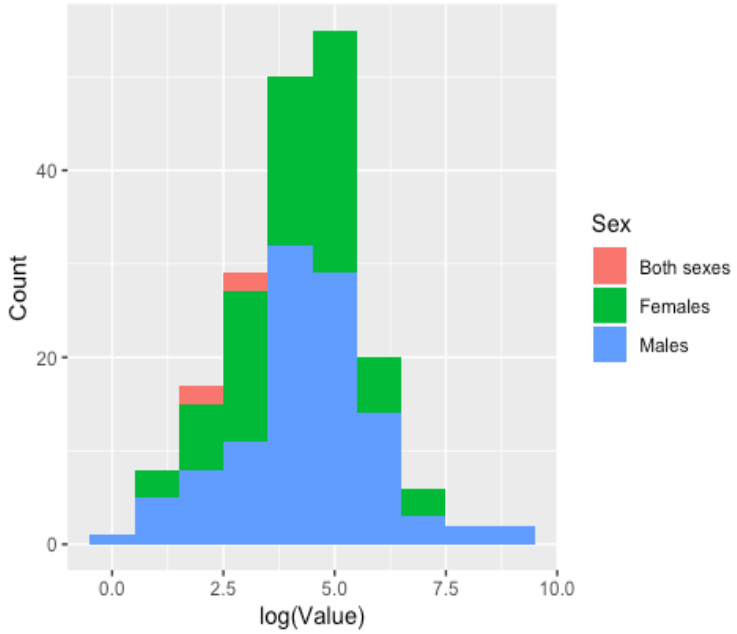



```r
# plot of donation on Age
df_tc %>%
  ggplot(mapping = aes(x = log(Value), fill = Age)) +
  geom_histogram(binwidth = 1) +
  labs(x = "log(Value)",
       y = "Count")
```

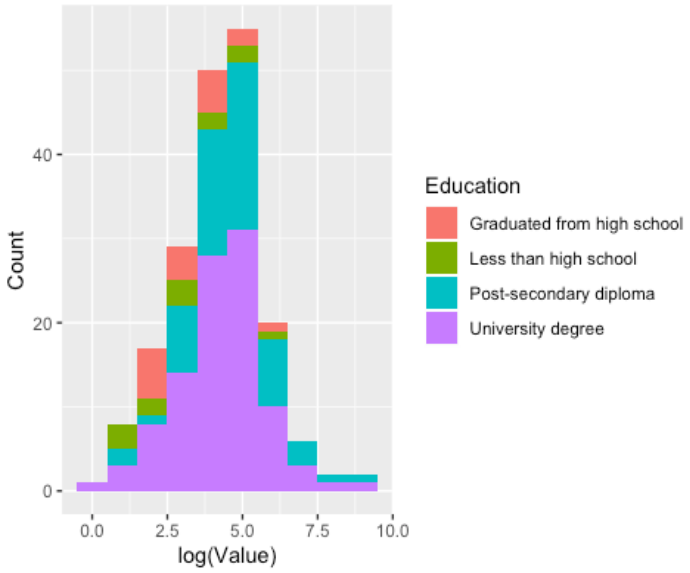



```r
# plot of donation on sex
df_tc %>%
  ggplot(mapping = aes(x = log(Value), fill = Sex)) +
  geom_histogram(binwidth = 1) +
```

```
  labs(x = "log(Value)",
       y = "Count")
```



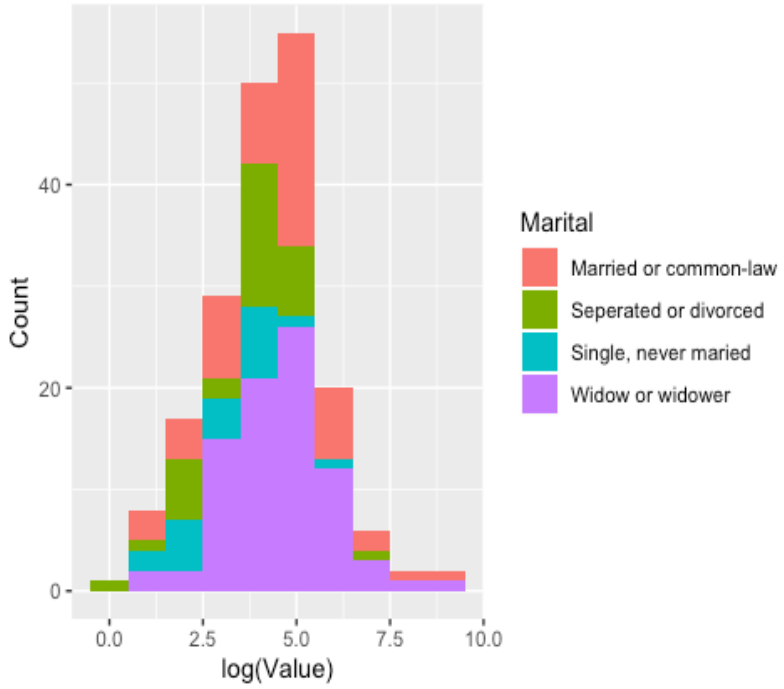

```
# frequency table of donation on sex
df_tc %>%
  group_by(Sex) %>%
  summarize(n = n(), Mean = mean(Value), Median = median(Value)) %>%
  mutate(freq = round(n / sum(n), 4))

## # A tibble: 3 x 5
##   Sex            n  Mean Median   freq
##   <fct>      <int> <dbl>  <dbl>  <dbl>
## 1 Both sexes     4  12.1   12.8 0.0211
## 2 Females       79 131.    83.3 0.416
## 3 Males        107 353.    84.4 0.563

# plot of donation on education
df_tc %>%
  ggplot(mapping = aes(x = log(Value), fill = Education)) +
  geom_histogram(binwidth = 1) +
  labs(x = "log(Value)",
       y = "Count")
```




```r
# table of donation on eduction
df_tc %>%
  group_by(Education) %>%
  summarize(n = n(), Mean = mean(Value), Median = median(Value)) %>%
  mutate(freq = round(n / sum(n), 4))
```

```
## # A tibble: 4 x 5
##   Education                     n  Mean Median   freq
##   <fct>                     <int> <dbl>  <dbl>  <dbl>
## 1 Graduated from high school    18  53.7   24.3 0.0947
## 2 Less than high school         13  62.7   17.2 0.0684
## 3 Post-secondary diploma        59 316.   100   0.310
## 4 University degree            100 277.    84.4 0.526
```

```r
# plot of donation on Marital
df_tc %>%
  ggplot(mapping = aes(x = log(Value), fill = Marital)) +
  geom_histogram(binwidth = 1) +
  labs(x = "log(Value)",
       y = "Count")
```




```
# table of donation on Marital
df_tc %>%
  group_by(Marital) %>%
  summarize(n = n(), Mean = mean(Value), Median = median(Value)) %>%
  mutate(freq = round(n / sum(n), 4))
```

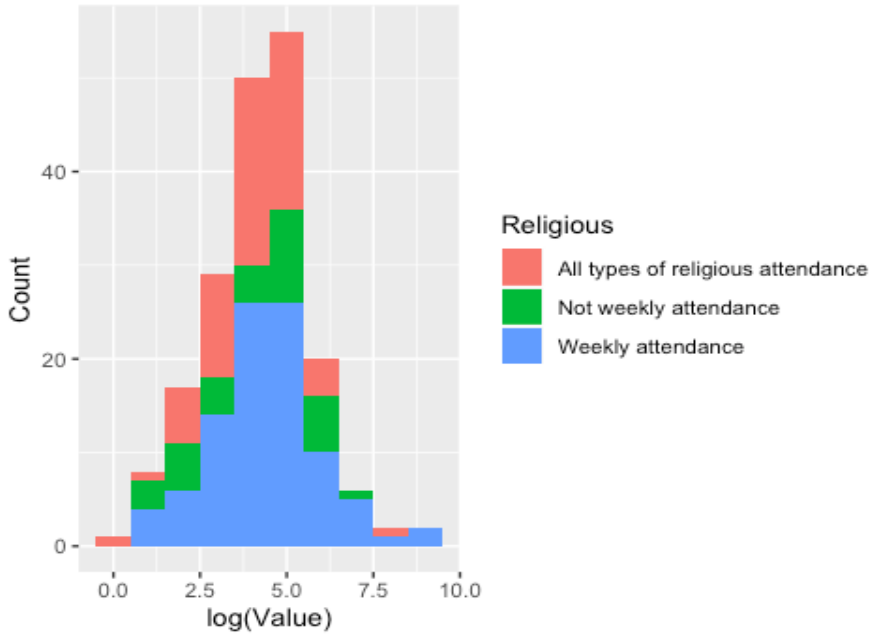
```
## # A tibble: 4 x 5
##   Marital                   n  Mean Median  freq
##   <fct>                 <int> <dbl>  <dbl> <dbl>
## 1 Married or common-law    55 291.    100   0.290
## 2 Seperated or divorced    32  85.5    73.4 0.168
## 3 Single, never maried     20  53.2    17.4 0.105
## 4 Widow or widower         83 341.    100   0.437
```

```
# plot of donation on income
df_tc %>%
  ggplot(mapping = aes(x = log(Value), fill = Income)) +
  geom_histogram(binwidth = 1) +
  labs(x = "log(Value)",
       y = "Count")
```


```
# table of donation on income
df_tc %>%
  group_by(Income) %>%
  summarize(n = n(), Mean = mean(Value), Median = median(Value)) %>%
  mutate(freq = round(n / sum(n), 4))
```

```
## # A tibble: 7 x 5
##   Income                   n  Mean Median   freq
##   <fct>                <int> <dbl>  <dbl>  <dbl>
## 1 $100,000 to $119,999    31 323.   100   0.163
## 2 $120,000 to $139,999    31 178.    95   0.163
## 3 $140,000 or more        24 158.   100   0.126
## 4 $20,000 to $39,999       5  26.7   16.5 0.0263
## 5 $40,000 to $59,999      30 101.    65.6 0.158
## 6 $60,000 to $79,999      33 105.    17.2 0.174
## 7 $80,000 to $99,999      36 615.   100   0.190
```

```
# plot of donation on Religious
df_tc %>%
  ggplot(mapping = aes(x = log(Value), fill = Religious)) +
  geom_histogram(binwidth = 1) +
  labs(x = "log(Value)",
       y = "Count")
```




```
# single covariate model
fit.sex <- lm(log(Value) ~ Sex, data = df_tc)
summary(fit.sex)

##
## Call:
## lm(formula = log(Value) ~ Sex, data = df_tc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1712 -1.0264  0.2298  0.7330  5.1008
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.4152     0.7562   3.194  0.00165 **
## SexFemales    1.6673     0.7751   2.151  0.03276 *
## SexMales      1.9383     0.7702   2.517  0.01269 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.512 on 187 degrees of freedom
## Multiple R-squared:  0.03683,    Adjusted R-squared:  0.02652
## F-statistic: 3.575 on 2 and 187 DF,  p-value: 0.02995

fit.age <- lm(log(Value) ~ Age, data = df_tc)
summary(fit.age)

##
## Call:
## lm(formula = log(Value) ~ Age, data = df_tc)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8826 -0.4932  0.0294  0.6433  4.6895
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.4602     0.5704   4.313 2.63e-05 ***
## Age25 to 34 years      -0.3632     0.8461  -0.429 0.668267
## Age35 to 44 years       0.4957     0.7091   0.699 0.485382
## Age45 to 54 years       1.3960     0.6352   2.198 0.029227 *
## Age55 to 64 years       1.5124     0.6187   2.444 0.015454 *
## Age65 to 74 years       2.1156     0.6002   3.525 0.000535 ***
## Age75 years and older   2.3047     0.6018   3.829 0.000176 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.397 on 183 degrees of freedom
## Multiple R-squared:  0.1955, Adjusted R-squared:  0.1691
## F-statistic: 7.413 on 6 and 183 DF,  p-value: 4.087e-07
```

```
fit.income <- lm(log(Value) ~ Income, data = df_tc)
fit.income
```

```
##
## Call:
## lm(formula = log(Value) ~ Income, data = df_tc)
##
## Coefficients:
##                (Intercept)  Income$120,000 to $139,999
##                    4.52614                     -0.05711
##     Income$140,000 or more    Income$20,000 to $39,999
##                    0.07785                     -1.45599
##   Income$40,000 to $59,999    Income$60,000 to $79,999
##                   -1.00726                     -1.12572
##   Income$80,000 to $99,999
##                    0.34974
```

```
fit.m <- lm(log(Value) ~ Marital, data = df_tc)
summary(fit.m)
```

```
##
## Call:
## lm(formula = log(Value) ~ Marital, data = df_tc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6917 -0.8752  0.1041  0.8186  4.9532
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   4.3849     0.1992  22.011  < 2e-16 ***
```

```
## MaritalSeperated or divorced  -0.6171     0.3285  -1.879  0.06186 .
## MaritalSingle, never maried   -1.2508     0.3858  -3.242  0.00141 **
## MaritalWidow or widower        0.1162     0.2569   0.453  0.65143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.477 on 186 degrees of freedom
## Multiple R-squared:  0.08581,    Adjusted R-squared:  0.07106
## F-statistic:  5.82 on 3 and 186 DF,  p-value: 0.0008025
```

```
fit.e<- lm(log(Value) ~ Education, data = df_tc)
summary(fit.e)
```

```
##
## Call:
## lm(formula = log(Value) ~ Education, data = df_tc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0709 -0.7853  0.0657  0.7624  5.2012
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.2642     0.3447   9.469  < 2e-16 ***
## EducationLess than high school   -0.2402     0.5323  -0.451 0.652366
## EducationPost-secondary diploma   1.3905     0.3938   3.531 0.000522 ***
## EducationUniversity degree        0.9890     0.3745   2.641 0.008968 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.463 on 186 degrees of freedom
## Multiple R-squared:  0.1041, Adjusted R-squared:  0.08963
## F-statistic: 7.203 on 3 and 186 DF,  p-value: 0.0001339
```

```
fit.r <- lm(log(Value) ~ Religious, data = df_tc)
summary(fit.r)
```

```
##
## Call:
## lm(formula = log(Value) ~ Religious, data = df_tc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8456 -1.2088  0.2113  0.8040  5.0605
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     4.02789    0.19261  20.912   <2e-16 ***
## ReligiousNot weekly attendance -0.05117    0.32852  -0.156    0.876
## ReligiousWeekly attendance      0.36594    0.24892   1.470    0.143
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.529 on 187 degrees of freedom
## Multiple R-squared:  0.01586,    Adjusted R-squared:  0.005333
## F-statistic: 1.507 on 2 and 187 DF,  p-value: 0.2243
```