

PERSONAL WEALTH SATISFACTION ANALYSIS

**AN EXAMINATION OF EDUCATION LEVEL,
GENDER, AGE, RISK TOLERANCE, AND
BUDGETING STRATEGIES ON PERSONAL WEALTH
SATISFACTION IN THE UNITED STATES**

June 14, 2023

Abdulla Alameri
The University of Manchester

Contents

1	Introduction	2
2	Literature Review	2
3	Describing the Data	3
3.1	Data Source	3
3.2	Data Variables	3
3.2.1	Personal Wealth Satisfaction	3
3.2.2	Education	4
3.2.3	Gender	5
3.2.4	Age Group	6
3.2.5	Risk Tolerance	7
3.2.6	Household Budgeting	7
3.3	Data Limitaions	8
4	Data Analysis	8
4.1	Data Correlation	8
4.2	Multiple Regression Analysis	9
5	Conclusion	10
	Refernce List	11
	Appendix	11

1 Introduction

This report presents an in-depth analysis of data drawn from a 2015 national study conducted in the United States, aiming to explore the dynamic relationship between multiple key factors and personal wealth satisfaction. The primary factors under scrutiny include education level, gender, age, risk tolerance, and budgeting strategies, each offering unique insights into the subjective perception of wealth.

In a rapidly changing socio-economic environment, personal wealth satisfaction has become an area of intense study and interest. This subjective measure of financial well-being plays an essential role in shaping overall life satisfaction and wellbeing, and understanding its determinants is crucial for policymakers, financial institutions, and individuals alike.

Education level, gender, and age, represent demographic and socioeconomic indicators that have been historically associated with wealth accumulation and satisfaction. Risk tolerance, on the other hand, sheds light on an individual's financial behavior, potentially driving wealth satisfaction independently of actual wealth levels. Lastly, budgeting strategies allow us to comprehend how disciplined financial planning and management impact perceived wealth.

Through a detailed review of the 2015 data, we aim to unravel the interplay between these variables and their collective impact on personal wealth satisfaction across different population segments. The findings of this report will contribute to a more nuanced understanding of financial satisfaction, assist in the development of targeted financial education programs, and aid in formulating policies aimed at enhancing the financial wellbeing of citizens.

In the sections that follow, we present our methodology, discuss our findings in detail, and explore their implications for various stakeholders.

2 Literature Review

The subjective measure of financial well-being, often referred to as wealth satisfaction, has been a topic of substantial research across various disciplines, including psychology, sociology, and economics. This complex and multi-faceted concept hinges not only on an individual's objective financial conditions but also on subjective perceptions of financial success (Diener Biswas-Diener, 2002).

A significant factor influencing wealth satisfaction is the level of education. According to a study by Oreopoulos and Salvanes (2011), higher education directly correlates with improved financial outcomes and wealth accumulation, contributing to higher levels of wealth satisfaction. Further, it has been suggested that individuals with a higher education degree often exhibit better financial literacy, which can further boost their satisfaction with their financial status (Lusardi Mitchell, 2014).

Gender disparities in wealth satisfaction have been a topic of substantial research as well. Women have been found to report lower levels of financial satisfaction compared to men (Bajtelsmit VanDerhei, 1997), possibly due to persistent wage gaps, gender differences in risk tolerance, or unequal access to financial resources and opportunities.

Age is another factor found to significantly impact wealth satisfaction. As individuals age, they often report higher levels of financial satisfaction (Namkoong, Sherraden Sherraden, 2008). This could be due to increased wealth accumulation over time and fewer financial obligations in older age.

Risk tolerance, an individual's willingness to take financial risks, also plays a pivotal role in shaping wealth satisfaction. Individuals with higher risk tolerance are often more open to investment opportunities that, though risky, may result in high financial returns, thereby boosting their wealth satisfaction (Grable, Lytton O'Neill, 2004).

Finally, budgeting behavior significantly influences wealth satisfaction. Regular and consistent personal budgeting has been linked to improved financial outcomes and increased wealth satisfaction as it

enables individuals to manage their finances better and achieve their financial goals (Godwin, 1998).

In conclusion, wealth satisfaction is a complex construct influenced by a variety of demographic, socioeconomic, and behavioral factors. This literature review provides a foundation for our study and informs our investigation into how education, gender, age, risk tolerance, and budgeting behavior impact wealth satisfaction.

3 Describing the Data

variables	n	mean	sd	median	trimmed	mad	min	max
WLTH_SAT	25706	5.82	2.73	6	5.93	2.97	1	10
EDU	25706	4.57	1.69	4	4.61	2.97	1	7
GENDER	25706	1.55	0.50	2	1.56	0.00	1	2
AGE_GR	25706	3.72	1.63	4	3.77	1.48	1	6
RISK_TOL	25706	5.18	2.65	5	5.16	2.97	1	10
BUDGET	25706	1.41	0.49	1	1.39	0.00	1	2

Figure 1: Variable's descriptive statistics table

3.1 Data Source

The data used in this report is from the 2015 National Financial Capability Study conducted by the Financial Industry Regulatory Authority (FINRA). The data is collected from a national census which included all states of the United States of America. The census covered alot of aspects but for the prupose of this report we chose 6 of interest. The observations used in this report excluded individuals who answered questions of the census with either do not know or prefer not to say to any of our variables of interest, which left us with 25706 observations for this report.

3.2 Data Variables

3.2.1 Personal Wealth Satisfaction

The WLTH_SAT variable, representing individuals' satisfaction with their personal financial condition, has been measured on a scale from 1 to 10, with 1 being 'not at all satisfied' and 10 being 'extremely satisfied'. This analysis is based on a sample size of 25,706 respondents ($n = 25706$).

The mean score of WLTH_SAT in the data is 5.86, indicating that, on average, respondents report a moderate level of satisfaction with their wealth status. The median value is 6, suggesting that half of the respondents have a satisfaction level of 6 or higher. This slight difference between the mean and median points to a relatively balanced distribution, though there could be a slight skewness towards lower satisfaction levels. The trimmed mean, which reduces the effect of outliers by cutting off a percentage of the data from either end, is slightly higher at 5.93.

The standard deviation (sd) for WLTH_SAT is 2.73, indicating a relatively wide spread of responses around the mean. In other words, individuals' satisfaction with their personal wealth varies significantly within the sample. This variation is also reflected in the MAD (Median Absolute Deviation) value of 2.97, another measure of dispersion that's more robust to outliers.

The minimum (min) value recorded is 1, indicating that some respondents are not at all satisfied with their personal financial condition, while the maximum (max) value is 10, showing that some are extremely satisfied. This broad range of responses underscores the diversity of wealth satisfaction levels among respondents, highlighting the importance of understanding its determinants.

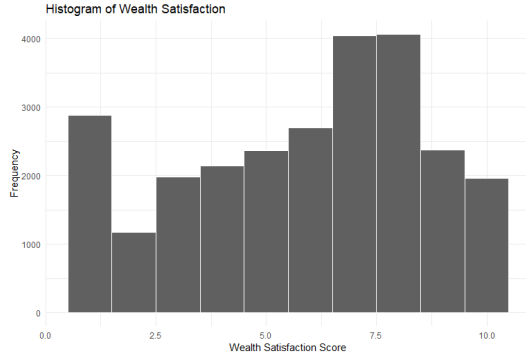


Figure 2: Histogram of Wealth Satisfaction

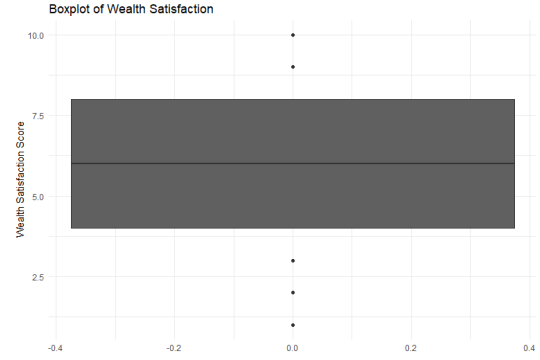


Figure 3: Boxplot of Wealth Satisfaction

3.2.2 Education

The **EDU** variable denotes the education level of individuals on an ordinal scale from 1 (did not complete high school) to 7 (postgraduate degree). The analysis includes a total of 25,706 participants ($n = 25706$).

1. Did not complete high school
2. High school graduate - regular high school diploma
3. High school graduate - GED or alternative credential
4. Some college, no degree
5. Associate's degree
6. Bachelor's degree
7. Postgraduate degree

The mean education level in the sample is 4.57, which falls between “some college, no degree” (4) and “Associate's degree” (5), indicating that, on average, respondents have some college education but may not have a degree. The median is 4, which directly corresponds to “some college, no degree”, suggesting that at least half of the respondents have reached this level of education or higher.

The standard deviation (**sd**) is 1.69, indicating that there's a reasonable spread in the education levels of respondents, with many having education levels above or below the mean. This is supported by the Median Absolute Deviation (**mad**) of 2.97, another measure of dispersion.

The trimmed mean, calculated by trimming a percentage of observations from either end of the sorted data set, is 4.61, which is similar to the mean. This suggests the mean is not being heavily influenced by extreme values (lower or higher education levels).

The minimum value (**min**) is 1, indicating there are respondents who did not complete high school. Conversely, the maximum value (**max**) is 7, showing that some respondents have obtained a postgraduate degree.

The distribution of this variable might suggest an educational trend among the respondents, which can influence wealth satisfaction levels. It also underscores the importance of considering education level as a significant independent variable in predicting wealth satisfaction.

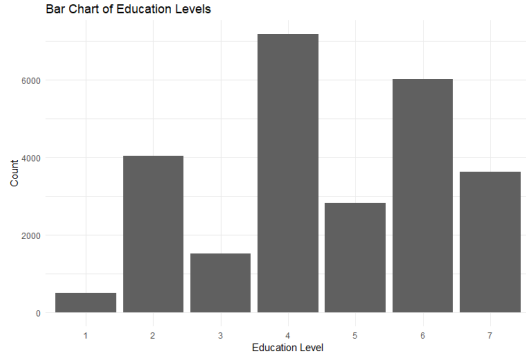


Figure 4: Bar Chart of Education Levels

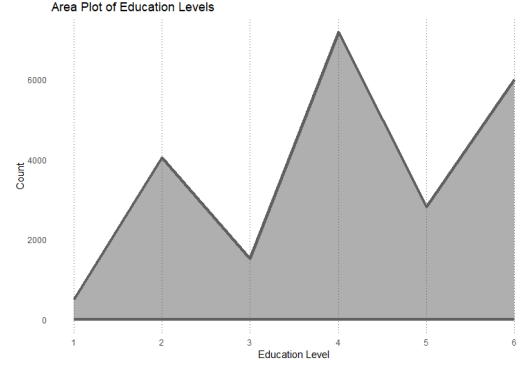


Figure 5: Area Plot of Education Levels

3.2.3 Gender

The **GENDER** variable, as defined in our study, is a dichotomous categorical variable that denotes the gender of the respondents, with '1' representing 'male' and '2' representing 'female'. This analysis encompasses a total of 25,706 participants ($n = 25706$).

The mean value of this variable is 1.55, indicating a slight majority of female respondents in the sample, as a value closer to 2 would suggest more females, while a value closer to 1 would suggest more males.

The median is 2, which corresponds to 'female', suggesting that at least half of the respondents identify as female. The low standard deviation (sd) of 0.50 indicates a tight distribution around the mean, as expected with binary variables.

The trimmed mean, which minimizes the impact of outliers, is 1.56, very close to the actual mean, suggesting there aren't extreme or erroneous gender classifications affecting our mean calculation.

The Median Absolute Deviation (mad) is 0, which is expected for a binary variable and indicates there is no deviation from the median.

The minimum (**min**) and maximum (**max**) values are 1 and 2 respectively, aligning with our defined gender categories.

Given these statistics, it can be inferred that the sample has a slight majority of females. The implications of this distribution on wealth satisfaction will be further explored in the analysis.

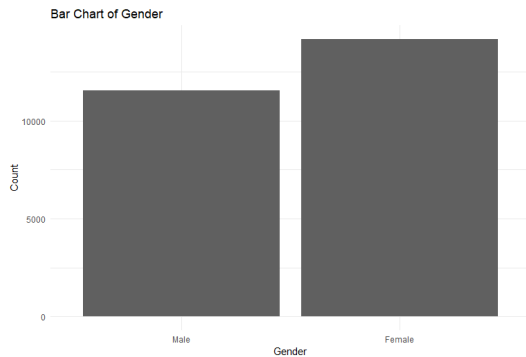


Figure 6: Bar Chart of Gender

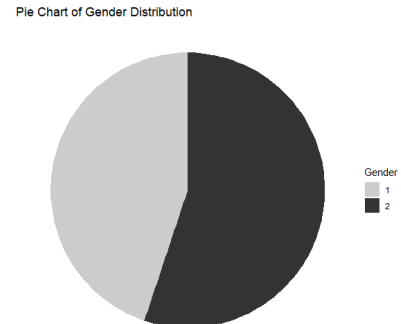


Figure 7: Pie Chart of Gender Distribution

3.2.4 Age Group

The `AGE_GR` variable, representing the age groups of the respondents, is a categorical variable with six classes, ranging from '1' denoting ages 18-24, up to '6' indicating ages 65 and above. In the given sample of 25,706 respondents ($n = 25706$), the statistical characteristics of this variable are as follows:

1. 18-24 Years Old
2. 25-34 Years Old
3. 35-44 Years Old
4. 45-54 Years Old
5. 55-64 Years Old
6. 65+ Years old

The mean value of the variable is 3.72. Since our categories are ordinal, this suggests that the average respondent falls between the '35-44' and '45-54' age groups.

The standard deviation (sd) is 1.63, indicating that the age groups of respondents generally deviate from the mean by approximately 1 to 2 categories. This highlights a fairly broad age distribution amongst respondents.

The median value is 4, representing the '45-54' age group. This suggests that half of the respondents are younger than 45-54 years, and the other half are older.

The trimmed mean, which is calculated by trimming a certain percentage of the extreme values from both ends of the distribution to minimize the impact of outliers, is 3.77, not significantly different from the mean. This suggests that extreme values aren't significantly skewing our mean age group.

The Median Absolute Deviation (mad) is 1.48, indicating that the age groups of respondents deviate from the median by roughly 1.5 categories.

The minimum and maximum values of the variable are 1 and 6, respectively, corresponding to the '18-24' and '65+' age groups. This indicates that our sample includes a wide age range from 18 to 65 and above.

In summary, the `AGE_GR` variable demonstrates a broad distribution, with a majority of the respondents falling within the middle age groups. These observations will help us understand how wealth satisfaction varies across different age groups.

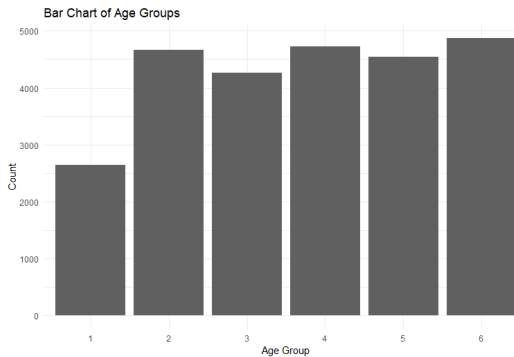


Figure 8: Bar Chart of Age Groups

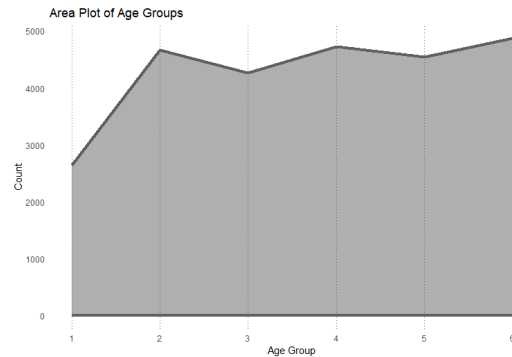


Figure 9: Area Plot of Age Groups

3.2.5 Risk Tolerance

The `RISK_TOL` variable was analyzed using a sample size (n) of 25,706 respondents. The mean risk tolerance score was 5.18, implying that on average, respondents have a moderate level of willingness to take on financial risks. The standard deviation was 2.65, indicating a substantial variation in risk tolerance among the respondents, with some individuals exhibiting a strong willingness to undertake financial risks while others show a much lower tolerance.

The median risk tolerance score, which is the middle score when all are arranged in ascending order, was found to be 5. This is quite close to the mean, suggesting that the distribution is not heavily skewed. The minimum and maximum risk tolerance scores were 1 and 10, respectively, indicating that the dataset encompasses the full range of possible scores.

A trimmed mean (mean of the middle 80%) of 5.61 was calculated, which is a robust measure that minimizes the influence of outliers by focusing on the central part of the distribution. Furthermore, the Median Absolute Deviation (MAD) was 2.97, offering another robust measure of the variability in the risk tolerance scores. These robust statistics further confirm the presence of considerable diversity in the willingness of respondents to accept financial risks.

Taken together, these statistical properties highlight the diverse risk preferences among individuals, paving the way for an interesting exploration of how these differences impact personal wealth satisfaction.

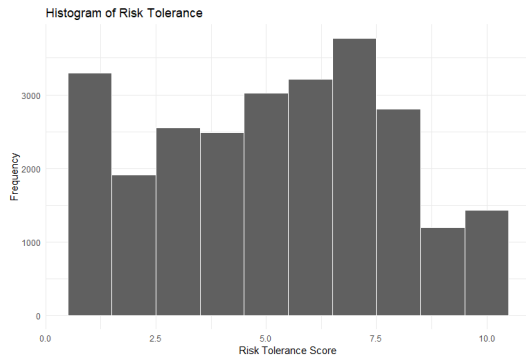


Figure 10: Histogram of Risk Tolerance

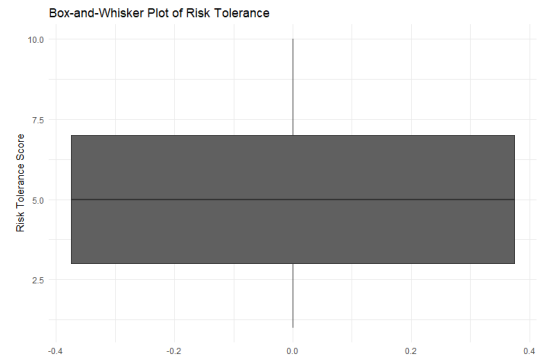


Figure 11: Box-and-Whisker Plot of Risk Tolerance

3.2.6 Household Budgeting

The `BUDGET` variable, which denotes whether or not a participant maintains a household budget, includes responses from 25,706 participants. This binary variable has a mean value of approximately 1.41 and a median value of 1, suggesting that more participants reported having a budget (coded as 1) than not having a budget (coded as 2).

The standard deviation of the `BUDGET` variable is 0.49, indicating a relatively small spread from the mean. This value, combined with a median absolute deviation (MAD) of 0, further points to the preponderance of budget-keepers in our dataset.

The trimmed mean, a robust measure of central tendency, is approximately 1.39 when calculated with a 10% trimming level. The trimmed mean is less sensitive to outliers and extreme values than the standard mean, and this value suggests a similar conclusion about the distribution of our data.

Overall, the `BUDGET` variable ranges from a minimum value of 1 (has a budget) to a maximum value of 2 (does not have a budget), with the majority of participants reporting that they maintain a budget. This characteristic could be an influential factor in our exploration of wealth satisfaction and its determinants.

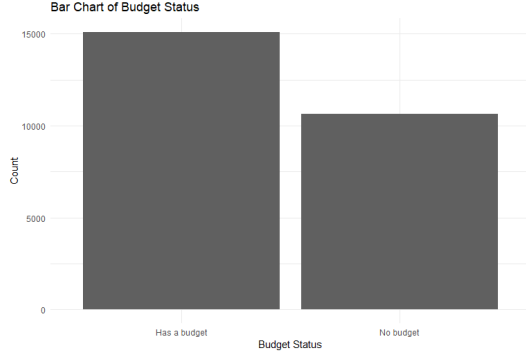


Figure 12: Bar Chart of Budget Status

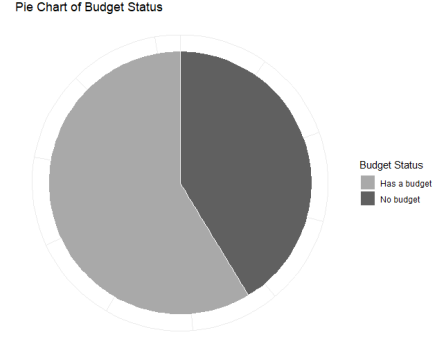


Figure 13: Pie Chart of Budget Status

3.3 Data Limitations

The current report relies on data from the 2015 US National Study, which inherently comes with certain potential limitations. First and foremost, the data, being from 2015, might not accurately reflect the present-day circumstances due to socio-economic and demographic shifts over the years. Another limitation could be the potential for sampling bias. If the national survey was not conducted using a truly randomized and representative sample, the results might be skewed with over or underrepresentation of certain demographic groups.

This study also largely depends on self-reported data, introducing the risk of various response biases. Social desirability bias, for instance, might lead respondents to provide answers they perceive as socially acceptable, rather than reflecting their true opinion or behavior. Similarly, recall bias might pose a problem if respondents find it challenging to accurately remember past behaviors or feelings.

Additionally, the analysis is limited by the variables included in the data. It's possible that other factors affecting personal wealth satisfaction were not captured in the survey. For instance, the measurement of wealth satisfaction (WLTH_SAT) and risk tolerance (RISK_TOL) are on a subjective scale, and the interpretation of these numbers can vary greatly among individuals, potentially leading to discrepancies in the results.

Furthermore, the observational nature of this study makes it difficult to establish definitive cause-effect relationships among the variables. Therefore, even if certain variables are found to be associated with wealth satisfaction, it doesn't necessarily imply that they cause changes in wealth satisfaction. Lastly, the problem of missing data can't be overlooked. Non-responses or incomplete responses to the census could result in missing data that, if not handled appropriately, could potentially bias the results. Given these limitations, the results of the analysis should be interpreted with caution, and findings should be corroborated with other data sources or more recent studies whenever possible.

4 Data Analysis

4.1 Data Correlation

Variables	WLTH_SAT	EDU	GENDER	AGE_GR	RISK_TOL	BUDGET
WLTH_SAT	1.000 00					
EDU	0.170 23	1.000 00				
GENDER	-0.136 43	-0.081 11	1.000 00			
AGE_GR	0.112 42	0.041 29	-0.038 03	1.000 00		
RISK_TOL	0.370 35	0.175 88	-0.215 72	-0.201 42	1.000 00	
BUDGET	-0.070 86	-0.013 64	-0.028 65	0.085 91	-0.054 02	1.000 00

Figure 14: Correlation Matrix of Chosen Variables Table

The variable 'EDU', representing the education level, displays a positive correlation of 0.170 with 'WLTH_SAT', suggesting a weak positive relationship. This indicates that, on average, those with higher education levels tend to report slightly higher wealth satisfaction. However, given the relatively low magnitude of this correlation, the relationship is weak and potentially influenced by other factors. The influence of education on wealth satisfaction could possibly be explained by the opportunities higher education often provides in terms of higher-paying jobs, thus leading to greater financial satisfaction.

'GENDER' displays a correlation of -0.136 with 'WLTH_SAT', indicating a weak negative relationship. This suggests a minor tendency for one gender (gender = 2) to have slightly lower wealth satisfaction than the other (gender = 1). However, this relationship is relatively weak, suggesting that gender may not significantly impact an individual's wealth satisfaction. It's crucial to bear in mind that gender, while influential in many aspects of life, is one of many factors affecting wealth satisfaction and likely interacts with other societal and individual factors not fully represented in this dataset.

Age group, or 'AGE_GR', has a correlation of 0.112 with 'WLTH_SAT', suggesting a weak positive relationship. This implies that as the age group increases (i.e., as individuals get older), they slightly tend to have higher wealth satisfaction. This weak correlation suggests that age, while playing a role, is not a major determinant of wealth satisfaction by itself. Other individual, societal, and economic factors may intersect with age to impact wealth satisfaction.

Risk tolerance, represented by 'RISK_TOL', shows a correlation of 0.370 with 'WLTH_SAT', the strongest among the variables considered. This moderate positive relationship indicates that individuals who have a higher willingness to take financial risks tend to also have a higher level of wealth satisfaction. This relationship might reflect the rewards often associated with financial risk-taking, which may lead to higher wealth satisfaction. Nevertheless, while this relationship is the strongest among the variables considered, it's still not very strong, and the factors contributing to risk tolerance and its effects on wealth satisfaction are likely to be complex.

Finally, 'BUDGET', indicating whether the individual has a household budget, shows a weak negative correlation of -0.071 with 'WLTH_SAT'. This suggests that those who reported not having a budget (Budget = 2) have a slight tendency to report higher wealth satisfaction than those with a budget (Budget = 1). However, this relationship is very weak, suggesting that the presence of a budget does not significantly impact wealth satisfaction. It's important to note that while budgeting is a key financial management strategy, it may be used more frequently by individuals experiencing financial strain, which might partially explain this correlation.

In conclusion, the correlations between 'WLTH_SAT' and the independent variables range from weak to moderate, and none show a strong relationship. It's essential to bear in mind that these correlations do not imply causation, and the relationships among these variables are likely influenced by various other factors not included in this dataset.

4.2 Multiple Regression Analysis

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.819675	0.098824	28.532	< 2e-16
EDU	0.147676	0.009272	15.928	< 2e-16
GENDER	-0.229714	0.031796	-7.225	5.16e-13
AGE_GR	0.316454	0.009685	32.676	< 2e-16
RISK_TOL	0.391661	0.006180	63.380	< 2e-16
BUDGET	-0.369424	0.031363	-11.779	< 2e-16

Figure 15: Regression Results

The intercept, which is 2.82, represents the expected value of WLTH_SAT when all other variables (EDU, GENDER, AGE_GR, RISK_TOL, BUDGET) are held constant at zero. However, since these variables do not naturally start at zero (e.g., GENDER is coded 1 for male and 2 for female, AGE_GR

ranges from 1 to 6 etc.), the intercept value doesn't have a practical interpretation in this case.

Moving on to the other variables, education level (EDU) shows a positive relationship with wealth satisfaction (WLTH_SAT). This suggests that as individuals achieve higher levels of education, they tend to be more satisfied with their financial condition. With every unit increase in education level, wealth satisfaction increases by approximately 0.15 units, holding all other variables constant. The extremely small p-value provides strong evidence to conclude that this relationship is not due to random chance, but instead indicates a statistically significant relationship.

Gender (GENDER) has a negative coefficient, suggesting that being female is associated with lower wealth satisfaction compared to being male, when controlling for all other variables in the model. To be more precise, switching from the category of male to female (as GENDER is coded 1 for male and 2 for female) is associated with a decrease of approximately 0.23 units in wealth satisfaction. The associated p-value, which is nearly zero, indicates that this relationship is statistically significant.

In terms of age groups (AGE_GR), the positive coefficient shows that older individuals are generally more satisfied with their wealth status compared to younger ones, when everything else is held constant. With every increment in age group category, the wealth satisfaction score increases by around 0.32 units. This finding is supported by the extremely small p-value, suggesting that the positive effect of age on wealth satisfaction is statistically significant.

Risk tolerance (RISK_TOL) also shows a positive relationship with wealth satisfaction. This indicates that those who are more willing to take financial risks tend to have higher wealth satisfaction. For each unit increase in risk tolerance, wealth satisfaction increases by approximately 0.39 units, holding all other variables constant. The p-value for this variable is nearly zero, indicating a significant effect of risk tolerance on wealth satisfaction.

Finally, having a household budget (BUDGET) is associated with higher wealth satisfaction. Specifically, switching from having a budget to not having one (as BUDGET is coded 1 for 'has a budget' and 2 for 'no budget') is associated with a decrease of around 0.37 units in wealth satisfaction, when controlling for all other variables. The associated p-value is very small, providing strong evidence of the significant negative effect of not having a budget on wealth satisfaction.

Each variable in this model is significantly associated with wealth satisfaction, and the direction of these relationships (indicated by the sign of the coefficients) is consistent with expectations based on prior knowledge. However, it's crucial to remember that these results represent associations and not causal relationships. These findings shed light on the factors related to wealth satisfaction and can help inform strategies to improve it. Further research could explore these relationships in more depth and could include additional variables that may also influence wealth satisfaction.

5 Conclusion

In this report, we undertook a deep analysis of a dataset containing responses from a wealth satisfaction survey, which encompassed 25,706 respondents. The key variables studied included the respondent's education level (EDU), gender (GENDER), age group (AGE_GR), risk tolerance (RISK_TOL), and the existence of a household budget (BUDGET). The primary aim was to investigate the associations between these variables and wealth satisfaction (WLTH_SAT).

Upon analyzing the data, we found that each of the variables under study had a significant association with wealth satisfaction. Education level showed a positive relationship with wealth satisfaction. This implies that with higher education levels, individuals reported higher wealth satisfaction, likely due to greater earning potential and financial stability that advanced education often brings.

Gender also played a significant role, with female respondents reporting lower wealth satisfaction compared to their male counterparts. While this study does not delve into the reasons behind this disparity, it aligns with broader societal discussions around the gender wage gap and financial confidence.

Older age groups were found to be more satisfied with their wealth status compared to younger ones. This trend could be explained by several factors including longer periods for wealth accumulation and possibly lower financial commitments among older age groups.

Individuals willing to take more financial risks reported higher wealth satisfaction, indicating a potential relationship between financial risk-taking and perceived financial success. Furthermore, the presence of a household budget was associated with higher wealth satisfaction. This highlights the role of personal finance management in creating a sense of satisfaction with one's wealth.

Despite the insights gathered, it's important to recognize the limitations in the data used for this study. These include potential response bias, as the willingness to complete the survey may be influenced by an individual's current financial situation. Additionally, as this is an observational study, we can only infer associations and not causal relationships between the variables.

In conclusion, this report has provided significant insights into the various factors influencing wealth satisfaction. By gaining a clearer understanding of these associations, we can develop more targeted strategies to improve financial wellbeing and wealth satisfaction. These strategies might include promoting financial education, advocating for gender financial equality, and encouraging budgeting practices. Future research could extend this analysis by investigating other potential influences on wealth satisfaction and by exploring the causal relationships identified here.

Reference List

- Bajtelsmit, V. L., & VanDerhei, J. L. (1995). Risk aversion and pension investment choices. In Positioning pensions for the twenty-first century (pp. 45-66).
- Diener, E., & Biswas-Diener, R. (2002). Will money increase subjective well-being?. Social Indicators Research, 57(2), 119-169.
- Godwin, D. D. (1998). Household debt quintiles: Explaining changes 1983-1989. Journal of Consumer Affairs, 32(2), 369-393.
- Grable, J., Lytton, R. H., & O'Neill, B. (2004). Projection bias and financial risk tolerance. Journal of Behavioral Finance, 5(3), 142-147.
- Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature, 52(1), 5-44.
- Namkoong, K., Sherraden, M. S., & Sherraden, M. (2008). Asset accumulation and aging in South Korea. Journal of Family and Economic Issues, 29(1), 129-143.
- Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. Journal of Economic Perspectives, 25(1), 159-84.

Appendix

```
1 # Written by Abdulla Alameri
2 # Last Edit : 12-June-2023
3
4 # Load Packages
5 library(readxl)
6 library(psych)
7 library(ggplot2)
8 library(dplyr)
9 library(corrplot)
10
11 # Set Working Directory
```

```

12 setwd("D:/side project")
13
14 # Import the CSV file from the working directory
15 df <- read_excel("Choosen Data.xlsx")
16
17 # Clean Data from missing values
18 df_clean <- subset(df, !apply(df == 99 | df == 98, 1, any))
19
20 # Generate a Descriptive stats table
21 describe(df_clean)
22
23
24 # Create Histogram for WLLT_SAT
25 ggplot(df_clean, aes(x=WLLT_SAT)) +
26   geom_histogram(binwidth=1, fill="#606060", color="white") +
27   theme_minimal() +
28   labs(title="Histogram of Wealth Satisfaction",
29        x="Wealth Satisfaction Score",
30        y="Frequency")
31
32 # Create Box and Whisker plot WLLT_SAT
33 ggplot(df_clean, aes(y=WLLT_SAT)) +
34   geom_boxplot(fill="#606060", coef=0) +
35   theme_minimal() +
36   labs(title="Boxplot of Wealth Satisfaction",
37        y="Wealth Satisfaction Score")
38
39 # Create a bar chart for EDU
40 ggplot(df_clean, aes(x=factor(EDU))) +
41   geom_bar(fill="#606060") +
42   theme_minimal() +
43   labs(title="Bar Chart of Education Levels",
44        x="Education Level",
45        y="Count")
46
47
48 # Create An Area plot of EDU
49 ggplot(df_clean, aes(x=EDU)) +
50   stat_bin(binwidth=1, geom="area", fill="#606060", color="#606060", size=1.5, alpha=0.5) +
51   geom_vline(xintercept=1:6, linetype="dotted", color="#606060") +
52   scale_x_continuous(breaks=1:6, limits=c(1,6)) +
53   theme_minimal() +
54   theme(panel.grid.major = element_blank(),
55         panel.grid.minor = element_blank()) +
56   labs(title="Area Plot of Education Levels",
57        x="Education Level",
58        y="Count")
59
60 # Create a bar chart for Gender
61 ggplot(df_clean, aes(x=factor(GENDER))) +
62   geom_bar(fill="#606060") +
63   scale_x_discrete(labels=c("1" = "Male", "2" = "Female")) +
64   theme_minimal() +
65   labs(title="Bar Chart of Gender",
66        x="Gender",
67        y="Count")
68
69 # Create a pie chart for Gender
70 df_clean %>%
71   group_by(GENDER) %>%
72   summarise(n=n()) %>%
73   mutate(prop = n/sum(n)) %>%

```

```

74 ggplot(aes(x="", y=prop, fill=factor(GENDER))) +
75 geom_bar(stat="identity", width=1) +
76 coord_polar("y", start=0) +
77 scale_fill_grey(start = 0.8, end = 0.2) +
78 theme_void() +
79 labs(title="Pie Chart of Gender Distribution",
80       fill="Gender")
81
82 # Create a bar chart for AGE_GR
83 ggplot(df_clean, aes(x=factor(AGE_GR))) +
84 geom_bar(fill="#606060") +
85 theme_minimal() +
86 labs(title="Bar Chart of Age Groups",
87       x="Age Group",
88       y="Count")
89
90 # Create a Area plot for AGE_GR
91 ggplot(df_clean, aes(x=AGE_GR)) +
92 stat_bin(binwidth=1, geom="area", fill="#606060", color="#606060", size=1.5, alpha=0.5) +
93 geom_vline(xintercept=1:6, linetype="dotted", color="#606060") +
94 scale_x_continuous(breaks=1:6, limits=c(1,6)) +
95 theme_minimal() +
96 theme(panel.grid.major = element_blank(),
97       panel.grid.minor = element_blank()) +
98 labs(title="Area Plot of Age Groups",
99       x="Age Group",
100      y="Count")
101
102 # Create a histogram for RISK_TOL
103 ggplot(df_clean, aes(x=RISK_TOL)) +
104 geom_histogram(binwidth=1, fill="#606060", color="white") +
105 theme_minimal() +
106 labs(title="Histogram of Risk Tolerance",
107       x="Risk Tolerance Score",
108       y="Frequency")
109
110 # Create a box-and-whisker for RISK_TOL
111 ggplot(df_clean, aes(y=RISK_TOL)) +
112 geom_boxplot(fill="#606060") +
113 theme_minimal() +
114 labs(title="Box-and-Whisker Plot of Risk Tolerance",
115       y="Risk Tolerance Score")
116
117
118 # Create a bar chart for BUDGET
119 ggplot(df_clean, aes(x=factor(BUDGET))) +
120 geom_bar(fill="#606060") +
121 scale_x_discrete(labels=c("1" = "Has a budget", "2" = "No budget")) +
122 theme_minimal() +
123 labs(title="Bar Chart of Budget Status",
124       x="Budget Status",
125       y="Count")
126
127 # Create a pie chart for BUDGET
128 budget_df <- data.frame(Category = factor(df_clean$BUDGET, levels = c(1, 2), labels = ←
129   c("Has a budget", "No budget")))
130
131 ggplot(budget_df, aes(x="", y=..count.., fill=Category)) +
132 geom_bar(stat="count", width=1, color="white") +
133 coord_polar("y", start=0) +
134 scale_fill_manual(values=c("#A9A9A9", "#606060")) +
135 theme_minimal() +

```

```
135 theme(axis.text.x = element_blank(),
136        axis.title.x = element_blank(),
137        axis.title.y = element_blank()) +
138 labs(title="Pie Chart of Budget Status",
139      fill="Budget Status")
140
141 # Create correlation matrix Table
142 cor_matrix <- cor(df_clean[, c("WLTH_SAT", "EDU", "GENDER", "AGE_GR", "RISK_TOL", "BUDGET")])
143 print(cor_matrix)
144
145 # Create Multiple Regression Analysis table
146
147 model <- lm(WLTH_SAT ~ EDU + GENDER + AGE_GR + RISK_TOL + BUDGET, data=df_clean)
148 summary(model)
```
