

# STA304 - Fall 2021

## Assignment 1

Cameron Dietzel - 1003506191

### Part 1

#### Goal

The cryptocurrency industry has remained the highest performing asset class for two straight years, incurring 800% gains during the 2020 calendar year [1]. Even with such impressive recent history, the global cryptocurrency market capitalization of roughly 1.9 trillion(\$USD) is expected to more than triple by 2030 [2]. With a bright outlook by many, the question remains: what type of investor does the young cryptocurrency market attract? Does knowledge level or risk tolerance affect one's decision to invest in this market? Do majority of investments come from a specific gender or income class? In this survey, the goal will be to determine any relation between an individuals exposure to the cryptocurrency market, compared to their economic or personal characteristics and qualifications.

#### Procedure

In order to collect responses, the survey will be posted across main social media platforms, such as Facebook, Twitter, Instagram, and LinkedIn. The survey will specifically be targeting the working class population, as the elderly have likely not submerged into the young cryptocurrency market. Each social media platform targets a different audience, and so each are needed to gain accurate results. The largest age groups (in years) for Facebook, Twitter, Instagram, and LinkedIn are 25-34, 30-49, 25-34 and 46-55 respectively [3]. This overall age range consists of majority of the target population. Therefore, the sample population will be those who have one of the listed social media accounts. Within this population, those completing the survey would likely be individuals whom I have connected with on these platforms; this is the frame population. It is possible for those not connected with me to complete the survey, but unlikely. The sample will consist of those within this frame population who complete the survey. This sampling procedure is the most effective and efficient with a short time horizon for data collection. It targets a high percentage of the desired population, as over 70% of the Canadian citizens are on a social media platform [3]. The potential drawbacks to this approach would be the limited access to mainly individuals whom I have connected with. Majority of this frame population is young (likely mid twenties) and may have different investment preferences to those in a later stage of life. This will mainly consist of those who are university educated which could lead to some bias in the results.

#### The Survey.

Link to survey: <https://www.surveymonkey.com/r/WKHLPRQ> [1].

#### Three Questions From The Survey

**Question 1: Which of the following would you be most likely to invest in?**

1. Cryptocurrency
2. Stocks
3. Bonds
4. Mutual Funds

5. Exchange Traded Funds
6. Real Estate
7. Other
8. Hold in cash

This question gauges the individuals preferences for investments. The listed categories include a variety of asset classes with different risk levels. This question is useful to draw any conclusions on if the recent surge in the cryptocurrency market has caused a greater proportion of individuals to rank cryptocurrency as their most desired asset class, or if the industry remains too young to establish this dominance. This question can also look to discover whether the type of investor the cryptocurrency market attracts truly believes in the future dominance of the industry, or is investing due to the social media hype based around it. A potential downfall to this question would be that the categories mainly put focus on financial investments leaving the 'other' category to be quite broad. Individuals may prefer to invest in memorabilia or rare goods that would all be lumped together. One may also choose to invest in experiences such as vacations, events, or education which are valid responses but of no use to the outcome of the survey.

**Question 2: What is your risk tolerance?**

**Low: 0-20**

**Medium Low: 21-40**

**Medium: 41-60**

**Medium High: 61-80**

**High: 81-100**

The question is meant to gauge the risk appetite of the participant. Cryptocurrencies have a historical 1 year average volatility of roughly 33%, which is significantly higher than any of the other broad asset classes[4]. This question seeks to understand whether only those with high-risk appetite are willing to make the cryptocurrency investment, or if it can attract those with lower risk preferences due to the outstanding market performance over the past several years. It is particularly difficult to gain a metric value for one's risk tolerance, which could be a potential downfall of this question. There is no foundational basis to compare risk tolerance to, so the (0-100) scale is likely viewed differently from person to person. Further risk preference questions could be asked to get a better understanding of one's tolerance levels.

**Question 3: What percent of your liquid investments are currently held in cryptocurrencies?**

The survey seeks to generalize what type of individual is likely to hold cryptocurrency investments. Comparing the individuals response from this question to their income class, gender, knowledge level or academic experience will be particularly useful. This question was formulated to discover the percentage held as oppose to dollar amount, as it provides a better representation of one's portfolio exposure to the cryptocurrency industry. A potential pitfall to the question itself could be the terminology used. Those completing the survey may not have a strong understanding of what a liquid investment is, and so their response to the question may not be accurate. That being said, the question needed to be formulated this way in order to better gain a true sense of exposure such that investments with some difficulty for resale (paintings, shoes, etc) are not included in the calculations. It is also possible that certain people do not know how much of their portfolio is in cryptocurrencies, if their funds are held with a discretionary financial adviser who can trade without the individual's permission.

## Part 2

### Data

In order to collect data, the survey was implemented on Survey Monkey and posted on various social media accounts requesting completion (Facebook, Twitter, Instagram, and LinkedIn). After 36 hours, 52 individuals had completed the survey and the link was taken down so results could be processed. The questions were formulated requiring specific responses (i.e an integer number between 0 and 100, inclusively) or categorical variables with a finite list of potential answers. Therefore all responses were of a desired form and so the data did not need to be cleaned. Some participants chose to skip questions and so the NA's had to be removed before the data could be processed.

Although the social media platforms target the ideal age ranges, those whom I have connected with are on the younger side. Therefore the responses likely do not generalize to the working class population on the whole, as age bias is present. Also, majority of these individuals earn above average income (as they are university educated), which could also skew the results. For greater accuracy, the survey would need to have been implemented on the various social media accounts for a longer period of time and better spread to a more diverse group of individuals.

The survey contains ten questions in total; two are numerical and 8 are categorical. Five of such questions gauge the respondents history and view of the cryptocurrency market with the most important variable for the study being the percentage of assets held in cryptocurrency. The remaining five questions are simple personal questions to discover the gender, income, education, etc. of the individual. The most important variables in the study are listed below. After the NA responses were removed, there are a total of 51 valid respondents for the survey.

1. 'knowledge': categorical variable consisting of four options (high, medium, low, no knowledge) to analyze one's knowledge level of the cryptocurrency market.
2. 'owned': categorical variable where participants have indicated if they have owned, or currently own cryptocurrencies.
3. 'invest\_decision': categorical variable analyzing asset class preferences for investors.
4. 'percent\_assets': the percentage of ones liquid assets held in cryptocurrencies.
5. 'risk\_tolerance': quantifies the investors risk tolerance on a scale from 0-100, inclusive. 0 would be completely risk averse and 100 would be completely risk seeking.
6. 'gender': the gender of the individual.
7. 'income': the income level of the individual.

Table 1 summarizes the important trends in the numerical variables. The average respondent holds 19% of liquid assets in cryptocurrency, and has an average risk tolerance score of 49. A scatterplot of the two numerical variables is plotted below, looking to understand any relation between risk preferences and percentage of assets held in cryptocurrencies.

Table 1: Numerical Variables Statistical Summary

Variable	Sample.Population	Mean	Standard.Error	Min	Max
percent_assets	51	18.96	29.28	0	100
risk_tolerance	51	48.73	19.21	10	90

Figure 1 plots an individuals percentage of assets held in cryptocurrencies compared to their risk tolerance score, with a line of best fit included. As shown, the line of best fit is upward sloping which supports the idea that investors with higher risk preferences have a greater likelihood of holding a higher percentage of cryptocurrency in their portfolio. Cryptocurrency is historically known for being volatile, and even with the historic performance over the past several years, it appears that such performance does not alter the risk preferences of the cryptocurrency investors[4].

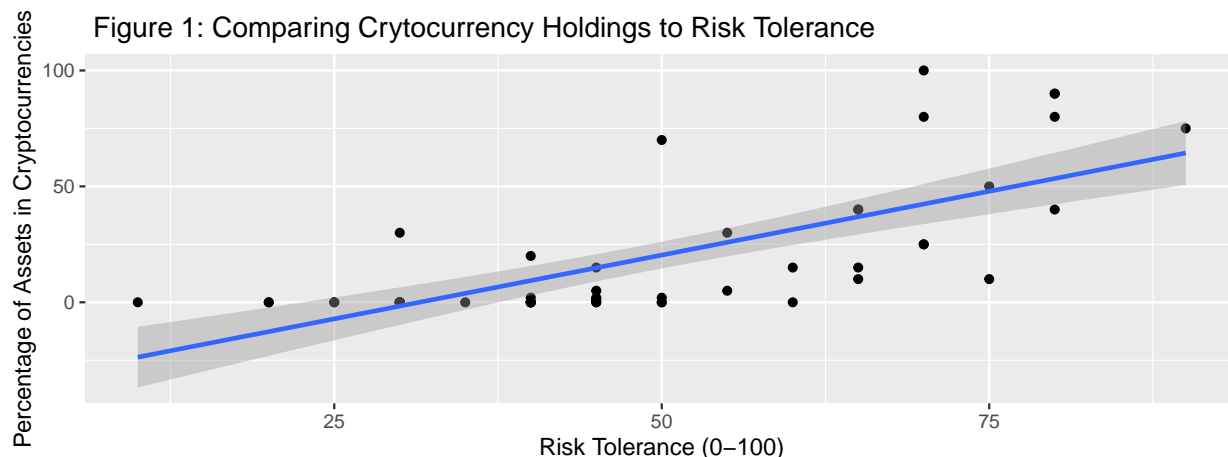


Table 2&3 and Figure 2&3 [7] analyze the important categorical variables in the survey. The tables below indicate the male to female ratio is fairly even, with slightly more males who have completed the survey. It can also be noted that nearly all respondents have at least some knowledge level of the industry, with roughly 63% indicating they have minimally a moderate level of knowledge.

Table 2: Gender Representation

Gender	Male	Female	Other	Undisclosed
Percentage	54.45%	45.55%	0%	0%

Table 3: Knowledge Level

Knowledge	High	Moderate	Low	No.Knowledge
Percentage	19.61%	43.14%	35.29%	1.96%

The results from figure 2[9] indicate that the cryptocurrency market is the most appealing to those in both the lower and upper income brackets. Individuals with an income closer to the Canadian average tend to be far less likely to have owned cryptocurrency in the past. These results make intuitive sense. Lower income individuals may see risky investments as their only possibility for greater financial success, and higher income individuals may have additional cash that they can afford to lose. Thus, both are likely to increase their holdings of riskier assets, such as cryptocurrencies. The more comfortable middle class worker likely does not feel as forced to make riskier investments, and does also not have the additional funds to waste.

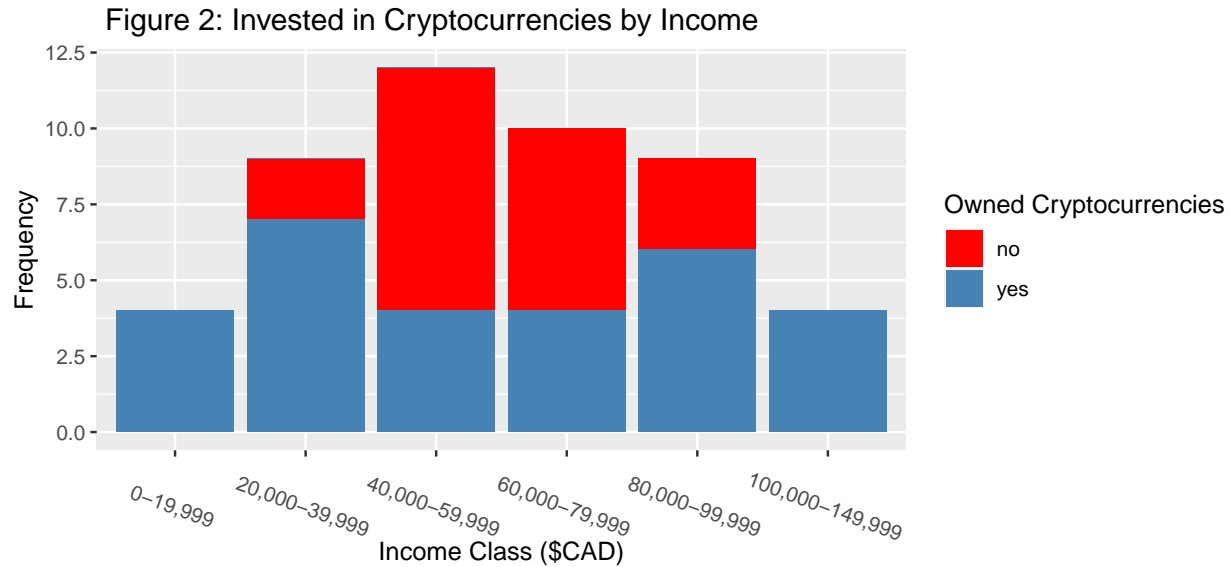
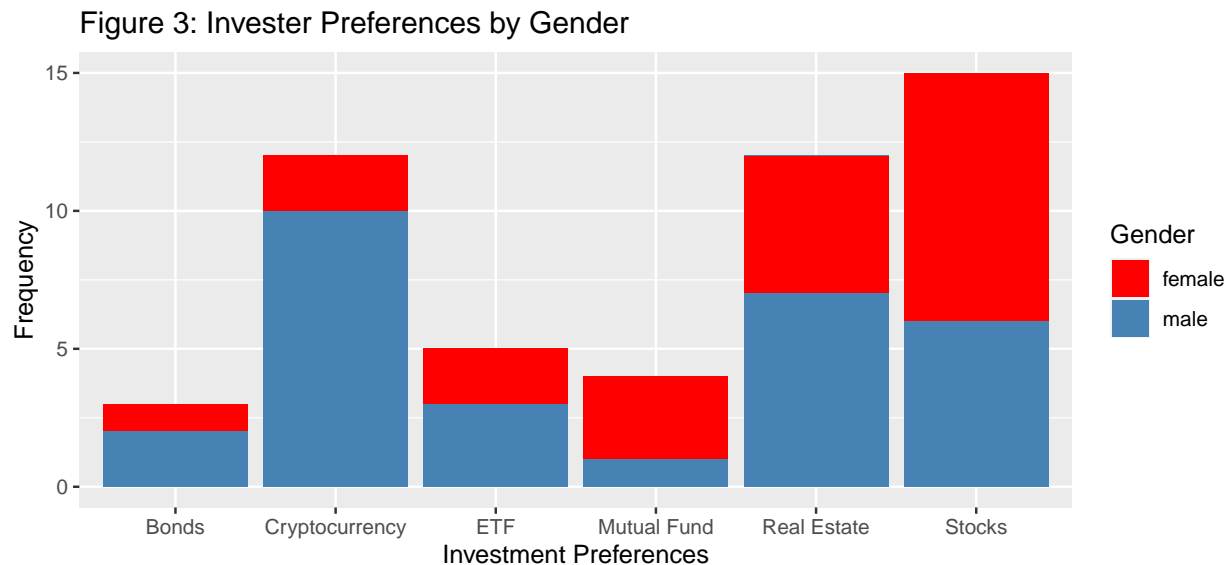


Figure 3 shows the investment preferences of those who completed the survey, compared to the gender of that individual. The results from the figure indicate that majority of asset classes seem to have an even distribution of male and female representation. The cryptocurrency market appears to indicate a gender bias. It can be noted that the male gender consists of 55% of all individuals in the survey, but 85% of those who recognize cryptocurrency as their most desired asset class. This figure appears quite surprising, however cryptocurrency has indeed historically had a gender bias. Roughly 66% of cryptocurrency investors are male [5], which supports the survey's discovery. Given the small sample at hand, the 85% figure would likely decline and converge to the industry average as data is collected from more individuals. It should also be noted that no individuals who completed the survey identified as 'other', or preferred not to disclose.



The key takeaways from figure 1-3 are as follows. For the numerical data, there appears to be a positive linear relationship between risk tolerance and percentage of investments held in cryptocurrencies. That means cryptocurrencies become more attractive as risk tolerance rises. For the categorical data, the lower and upper income brackets have a much higher percentage of individuals who have invested in cryptocurrencies, compared to those in the income brackets near the nation average. There also appears to be some gender bias in the industry, as the male gender consists of 85% of those who have indicated cryptocurrency as their

most desired asset class to invest in.

All analysis for this report was programmed using **R version 4.0.2**.

## Methods

### Hypothesis Test

The first statistical method used will be the One-Sample t-test for the mean risk tolerance. This test will determine whether the sample mean statistically differs from a specified value [10]. Looking at the target population on the whole, the average investor is slightly risk averse [6]. This would correspond to a risk tolerance rating of roughly 35 on the 0-100 scale, and will be the hypothesized risk tolerance score used for the t-test. If the positive linear relation between risk tolerance and percentage of assets held in cryptocurrency holds true (as indicated in Figure 1), and majority of investors are indeed risk averse, then this could be supporting rationale for investors holding a low percentage of assets in cryptocurrency. The mean is calculated as the average risk tolerance score of all 51 individuals, with the standard error being the deviation about this mean. These values will be used to calculate the test statistic, which indicates how closely the observed sample mean matches the hypothesized mean of 35. The test statistic will indicate a corresponding P-value, which measures how compatible the hypothesized mean is with the data. Generally, a P value of less than 0.05 is considered significant and is low enough to reject the null hypothesis [7].

This method is appropriate since the risk tolerance data appears to follow a normal distribution, the scale of 0-100 inclusively is continuous, and each data point is independent of one another, and there does not appear to be any outliers [10]. In other terms, each response does not affect the next, and the results cluster about the sample mean in the pattern of a normal distribution.

### Confidence Interval

To examine a categorical variable, the 'invest\_decision' variable will be converted into a binary indicator variable. This will be used to invoke a 95% confidence interval for the percentage of individuals whose most desired asset class is cryptocurrencies. The variable will indicate a '1', if the participant has chosen cryptocurrency as their top asset class, and '0' otherwise. This test will determine the range of values that we can be 95% certain contain the true mean for the variable. This statistic is important to determine the degree of certainty as to what percentage of individuals prefer the cryptocurrency market to invest in, compared the other asset classes. The sample mean will be the proportion of those with a '1' compared to the entire sample of 51 individuals, while the standard error will capture the deviation about this mean. A t-interval test will be used since the population standard deviation is unknown. This t interval must be calculated at 95% with  $n-1 = 50$  degrees of freedom. The confidence interval is completed under the following assumptions. The data is sampled randomly, the data points are independent and the sample size is sufficiently large [11].

## Results

Table 4 shows the important metrics for the hypothesis test with the null hypothesis being a mean of 35. The test statistic must first be calculated.

$$\text{Test Statistic} = \frac{\bar{x} - \mu}{\frac{SE(x)}{\sqrt{n}}} = \frac{48.32 - 35.00}{\frac{19.20}{\sqrt{51}}} = 5.104$$

The test statistic of 5.1 corresponds to a P-value of  $5.189 * 10^{-6} < 0.05$ . This implies that the P-value is significant, and therefore we reject the null hypothesis that the true mean is 35. With a sample mean of 48.73, this is not a surprising result. This implies that the true risk tolerance mean of the sample is likely not the industry average level. This is possibly due to the majority of the frame population being the risk seeking younger generation, or the number of cryptocurrency investors in the sample.

As the goal of the survey is to discover the type of investor the cryptocurrency market attracts, this helps provide some insight into the risk preferences of such investors. As proven with the one sample t-test, the population mean is not representative of those who completed the sample. This is because the sample average is significantly higher than the estimated population mean. Also, 50% of those who have completed the study indicated cryptocurrency ownership while the same holds true for just 14% of the population altogether[13]. Thus, the high risk tolerance from the sample is likely correlated with the high percentage of respondents indicating they have owned cryptocurrency, implying the cryptocurrency market tends to attract those with increased risk preferences.

Table 4: T-test for True Mean Risk Tolerance of 35

Variable	Sample	Mean	Standard.Error	Null.Hypothesis	Test.Statistic	P.value
risk_tolerance	51	48.324	19.205	35	5.104	5.189e^-6

Table 5 observes the key statistics for the confidence interval about the mean percentage of investors whom prefer cryptocurrencies. As indicated in the table, the percentage of those who have chosen the cryptocurrency market as their most desired asset class is 23.53% with a standard error of 42.84%. This yields a 95% confidence interval of (13.33%,33.73%). This represents the range for which we can be 95% confident contains the true average percentage of those with cryptocurrency as their most desirable asset class to invest in. The Calculation for the 95% Confidence Interval with 50 degrees of freedom is as follows:

$$\bar{x} \pm t(0.05, 50) \frac{SE(x)}{\sqrt{n}} = 0.2353 \pm 1.7 \frac{0.4284}{\sqrt{51}}$$

The large range provides some uncertainty as to how this figure contributes to identify the type of investor the market attracts. The sample mean (identified in figure 2) indicates half the respondents have invested in cryptocurrency. If this holds true, and only 15% (lower portion of the confidence interval) indicate cryptocurrency as a top asset class, then the market is likely enticing those who are speculative and investing with limited knowledge or foundational belief in the industry. However, if half the population invest and 30% (upper portion of the confidence interval) indicate that cryptocurrency is a top asset class, then this investor type likely consists of those with a greater knowledge base and are making informed decisions.

Table 5: Confidence interval for mean percentage of investors who prefer cryptocurrency

Variable	Sample	Observations	Percentage	Standard.Error	Lower.Bound	Upper.Bound
invest_decision	51	23	0.2353	0.4284	0.1333	0.3373

To summarize, the statistical methods implemented above have indicated that the cryptocurrency market attracts individuals with a higher risk tolerance. The large error in the confidence interval implies uncertainty as to what the true percentage of those whom prefer the cryptocurrency market is, while if the true mean is low or high indicates different qualities for the investors.



## Bibliography

1. Novak, D (2021, May 04) *Cryptocurrency Is The Best Performing Asset Class*. <https://www.benzinga.com/21/05/20942550/cryptocurrency-is-the-best-performing-asset-class-for-2-years-straight-but-what-does-this-mean-for-y>. (Last Accessed: Oct. 1 2021)
2. Gottlinch, M. (2021, Aug 15) *Global Crypto Market will more than triple to \$4.94B*. <https://seekingalpha.com/news/3733907-global-crypto-market-will-more-than-triple-to-494b-by-2030-allied-market-research>. (Last Accessed: Oct. 1 2021)
3. Dean, B. (2021, Sept. 01) *How Many People Use Social Media in 2021*. <https://backlinko.com/social-media-users>. (Last Accessed: Oct. 1 2021)
4. Woo, W. (2021, Sept. 30) *Bitcoin Volatility vs Other Assets*. <https://charts.woobull.com/bitcoin-volatility-vs-other-assets/>. (Last Accessed: Oct. 1 2021)
5. Cohen, J. (2021, Aug 30) *Cryptocurrency investing has a big gender problem*. <https://www.cnbc.com/2021/08/30/cryptocurrency-has-a-big-gender-problem.html>. (Last Accessed Oct. 1 2021)
6. Campbell, J. (1997) *Risk Neutrality*. <https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/risk-neutrality>. (Last Accessed Oct. 1 2021)
7. Campbell, J. (1997) *r - Order Bars in ggplot2 graph*. <https://stackoverflow.com/questions/5208679/order-bars-in-ggplot2-bar-graph>. (Last Accessed Oct. 1 2021)
8. Allaire, J.J., et. el. *Center Plot Title in ggplot2*. <https://stackoverflow.com/questions/40675778/center-plot-title-in-ggplot2>. (Last Accessed: Oct. 1, 2021)
9. Grolemond, G. (2014, July 16) *Function Reference ggplot2*. <https://ggplot2.tidyverse.org/reference/>. (Last Accessed: May 5, 2021)
10. Kent State University. (2021, Sept. 29) *One Sample t Test*. <https://libguides.library.kent.edu/SPSS/OneSampletTest>. (Last Accessed: Oct 1, 2021)
11. Bevans, R. (2020, Jul. 17) *Test Statistics*. <https://www.scribbr.com/statistics/test-statistic/>. (Last Accessed: Oct 1, 2021)
12. Zach. (2021, Jun. 22) *The 6 Confidence Interval Assumptions to Check*. <https://www.statology.org/confidence-interval-assumptions/>. (Last Accessed: Oct 1, 2021)
13. Alcott, D. (2021, Apr. 21) *14% of Americans Own Crypto*. <https://ca.finance.yahoo.com/news/study-reveals-crypto-biggest-investors-132102315.html/>. (Last Accessed: Oct 1, 2021)

## Appendix

Here is a glimpse of the data set surveyed:

```
## Rows: 51
## Columns: 11
## $ Participant      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16~
## $ knowledge        <chr> "high", "high", "moderate", "moderate", "low", "low",~
## $ owned            <chr> "yes", "yes", "yes", "yes", "no", "no", "no", "no", "~
## $ outperform       <chr> "yes", "yes", "yes", "yes", "no", "unsure", "unsure",~
## $ invest_likliness <chr> "extremely likely", "extremely likely", "extremely li~
## $ percent_assets   <int> 80, 40, 20, 15, 0, 0, 0, 0, 10, 0, 0, 30, 70, 0, 2, 0~
## $ invest_decision  <chr> "cryptocurrency", "real estate", "cryptocurrency", "r~
## $ risk_tolerance   <int> 70, 80, 40, 65, 30, 25, 50, 45, 75, 60, 50, 30, 50, 3~
## $ education        <chr> "university degree", "masters degree", "university de~
## $ income           <chr> "20,000-39,999", "150,000-199,999", "20,000-39,999", ~
## $ gender           <chr> "male", "male", "female", "male", "female", "male", "~
```