Project 2

STATS 220 Semester One 2024

Jamie Dunwoodie-Rowley

Introduction

This report focuses on the weekly spending and saving habits of people in my sample. It highlights the differences between essential living costs, such as housing and groceries, and personal expenses, such as entertainment and subscriptions and then compares that will the amount people save each week.

When designing the form in this report, a significant guideline that was thought about was to consider my respondents. This was an important decision as I had to consider who was going to be responding to the survey and ensure that the questions were inclusive of all participants. A specific example of my consideration of all respondents was when designing the 'About you' section of the report where I asked for respondents gender. I made sure to include an 'other' option as well as a 'prefer not to say' option for participants who are non-binary or would rather not share their gender identity.

The wording used in the question of the survey, "how much money did you spend on living costs this week?" prompts participants to recall their spending habits that week. Therefore, if the survey were taken each week, the data that is collected would be able to be analysed against the data from previous weeks to see changes in human behaviour. Specifically, we could analyse how much on average, people spend on a week by week basis and analyse how this changes. This could then be analysed in conjunction with world events happening that week and see if there are any changes in human behaviour, and spending based on what is happening around them. For example, we could see if spending goes up around holidays compared to weeks with no holidays.

link to form

Analytics

Three things i learnt about the data:

- The first thing I learnt is that the most common living cost is cellphone, this makes sense as most people now have cell phones and have to pay for plans, even when living at home and having no other expenses. The second highest is groceries as most people pay for groceries, unless living with caregivers or in university accommodation that provide food.
- I learnt that the lowest personal cost category was travel, this is likely because people do not travel on a week to week basis, instead they save up for a few big trips a year. However, a few respondent's did respond that they spent money on travel so it is possible they took a trip that week. It would be interesting to analyse their data in future weeks to see how their behaviour changes and if their travel costs lower.
- Very few people save more than 50% of their income a week, it is the least saved amount.

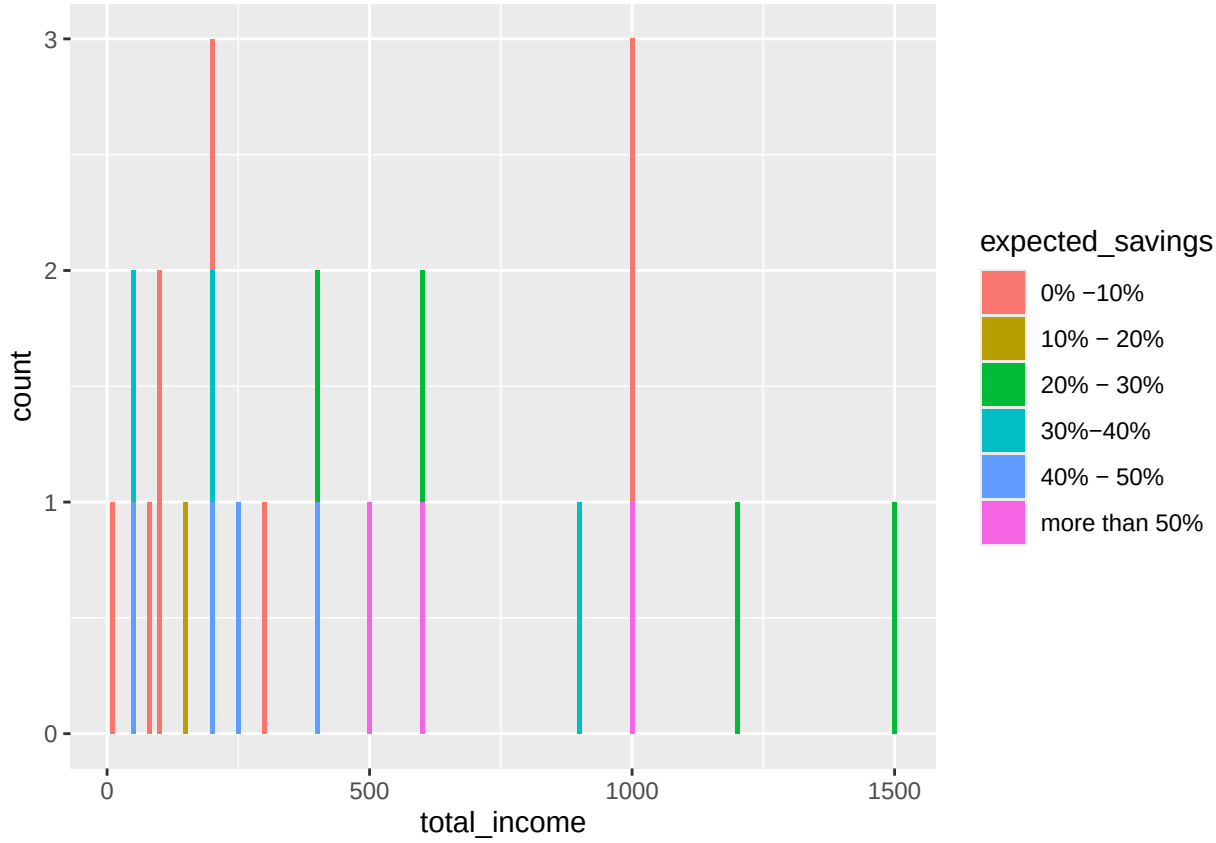
```
#load data...
learning_data <- read_csv("https://docs.google.com/spreadsheets/d/e/2PACX-1vSsjAq84CkLUyX8eM421JZFyTxYD

#Rename cols
learning_data <- learning_data %>%
  rename(total_income=2,living_out=3,living_costs=4,personal_expenses=5,personal_costs=6,expected_savin
```

```r
#calculate total income
total_income <- sum(learning_data$total_income)

# Calculate average personal spend
avr_personal_spend <- round(mean(learning_data$personal_expenses),2)

# Calculate average living spend
avr_living_spend <- round(mean(learning_data$living_out),2)

ti <- paste("$", total_income, sep="")
aps <- paste("$",avr_personal_spend, sep="")
als <- paste("$",avr_living_spend, sep="")
```

Total Income vs Savings

This graph shows the spread of total income vs how much each participant is expecting to save. The bars in the graph are filled with expected savings. I previously predicted there would be a lot more correlation between expected savings and total income, assuming that respondents who earnt more would save more. However, this doesnt seem to be true according to the graph as the respondents with the highest income tend to be saving 20%-30%. This makes sense as people with higher income might also have higher expenses, such as higher housing costs and utilities.

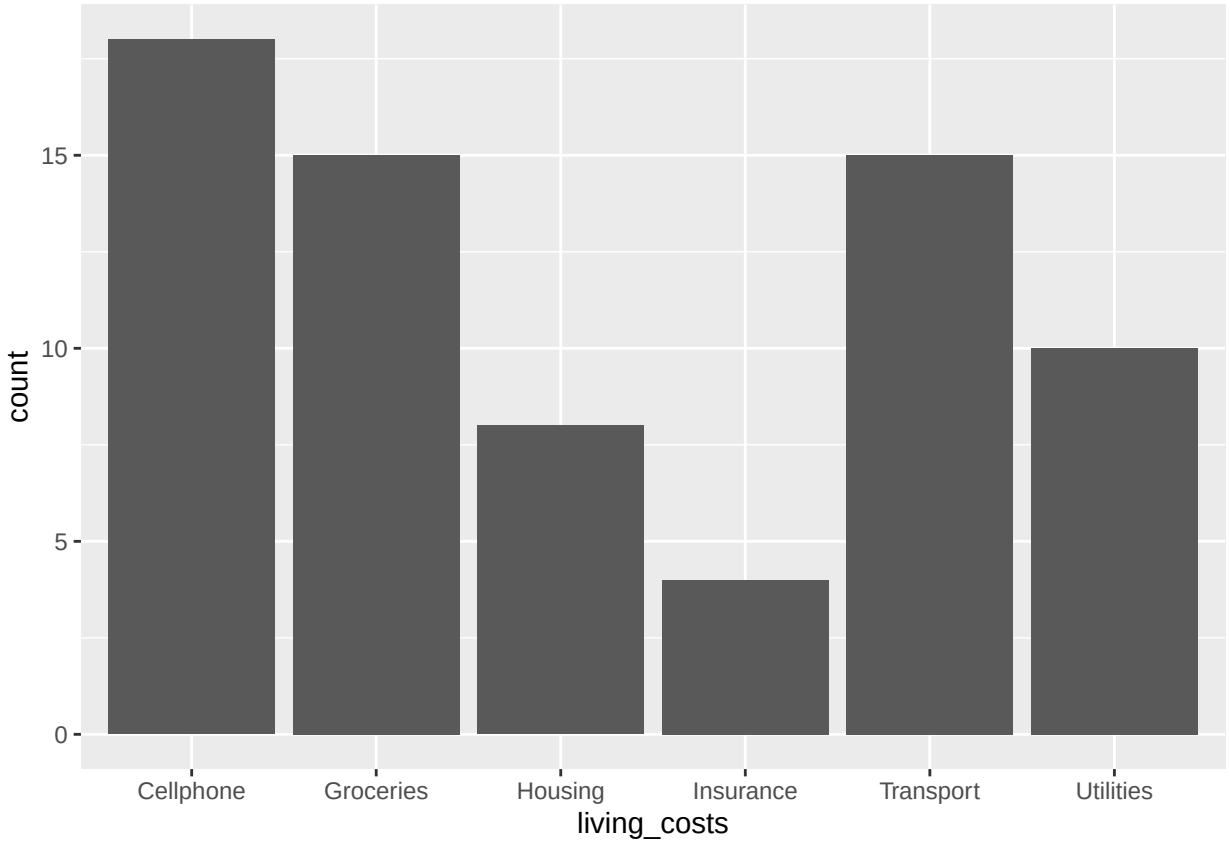
```r
learning_data %>%
  ggplot() +
  geom_bar(aes(x = total_income, fill=expected_savings))
```



Living costs

This graph shows the the count of each living cost broken down, cellphone and transport make the most sense being the highest as most people have to pay these costs, even when living with caregivers etc.

```
utilities <- learning_data %>%
  separate_rows(living_costs, sep = ", ")

utilities %>%
  ggplot() +
  geom_bar(aes(x = living_costs))
```



- The total income of all participants is $10790
- The average living cost of all participants is $201.74.
- The average personal spend of all participants is $85.22.

Creativity

I demonstrated creativity in two main ways;

1. The use of mean and round, both functions i had to find by reading R docs
2. I was also creative in designing the expected savings VS total income graph utilizing the fill to demonstrate the amount that participents save each week in respect to their incomes.

Learning reflection

I learnt a lot throughout the course of this project, one of my favourite things was the use of Google forms/sheets to create a dynamic data pipelines. As a software developer I write a lot of python code and in order to make a data pipeline the same as this I would have to design my own frontend and database

requiring a lot of work, the use of this pipeline greatly speeds up the process. I would be curious to explore the limitations to see why this solution is not used more often. Over all, I thoroughly enjoyed this project and the coding behind it.