Examples of data visualization in R, and quantitative methods from a social survey in Colombia.

1. Luis Fernando Grajales H. Professor Universidad Nacional de Colombia. lfgrajalesh@unal.edu.co
2. Sara Grajales-Morales, Sociologist and MSc Educational management. sagrajalesmo@gmail.com

32° Simposio Internacional de estadística 2023, Ibagué - Colombia

UNIVERSIDAD NACIONAL DE COLOMBIA

Abstract

Data Visualization in R for Statistics in Social Science has increased in recent years due to several applications such as social media, gender studies, and labour market, among others. This work combines some plots of data visualization in R with some initial quantitative methods in order to analyze a survey applied to Colombian families. The database consists on 19,416 homes in 2017. We show some examples of visualization such as ggplot. In the statistical examples we present univariate and bivariate analysis of variables related to financial burden in the families. We compared the result between three main cities in Colombia. Some results indicate that the families use their remaining money from their basic expenses for three objectives: for save, for house purchase and for emergencies.

Objectives

The study aims to characterize the utilization of remaining funds in surveyed households of IEFIC 2017, differentiated by cities. Additionally, it seeks to establish an approach for social sciences through data visualization and identify challenges in data reading and interpretation for social scientists.

Analysis

The IEFIC 2017 dataset was filtered for the head of household. Then, individuals with remaining money after basic expenses were selected.

Individuals	head.ho	w.remai
47347	19416	3684

Taula 1: Dimensions of used data frames

We analyzed the filtered data and formulated research questions based on the sample for each city's financial dynamics, as an example:
1. What is the distribution of expenses among the different categories in each city?

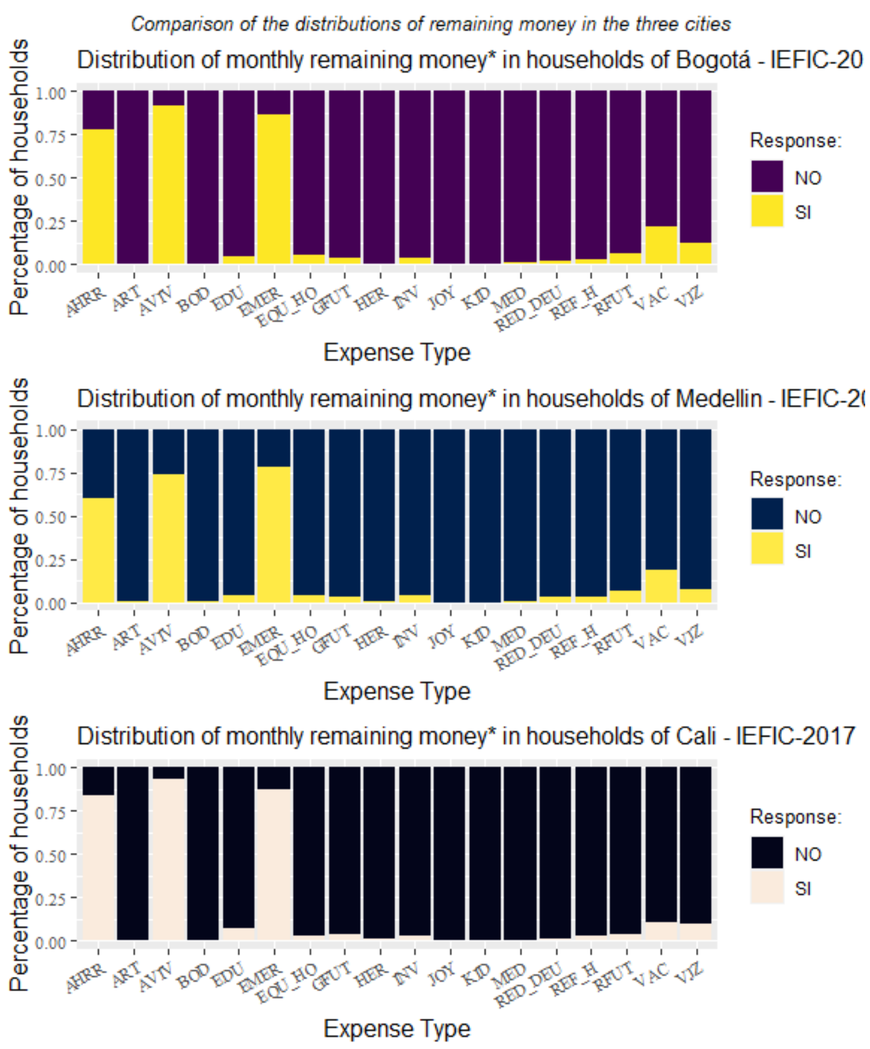

Figura 1: Comparison of expenses in the three cities

The remaining money in the sample from the 3 cities shows a similar pattern, with savings, home purchases, emergencies, and vacations being prominent, while jewelry, artworks, and inheritances are less common.

Results

The initial results analyzed the most common expenses to understand the financial behavior of the sample.

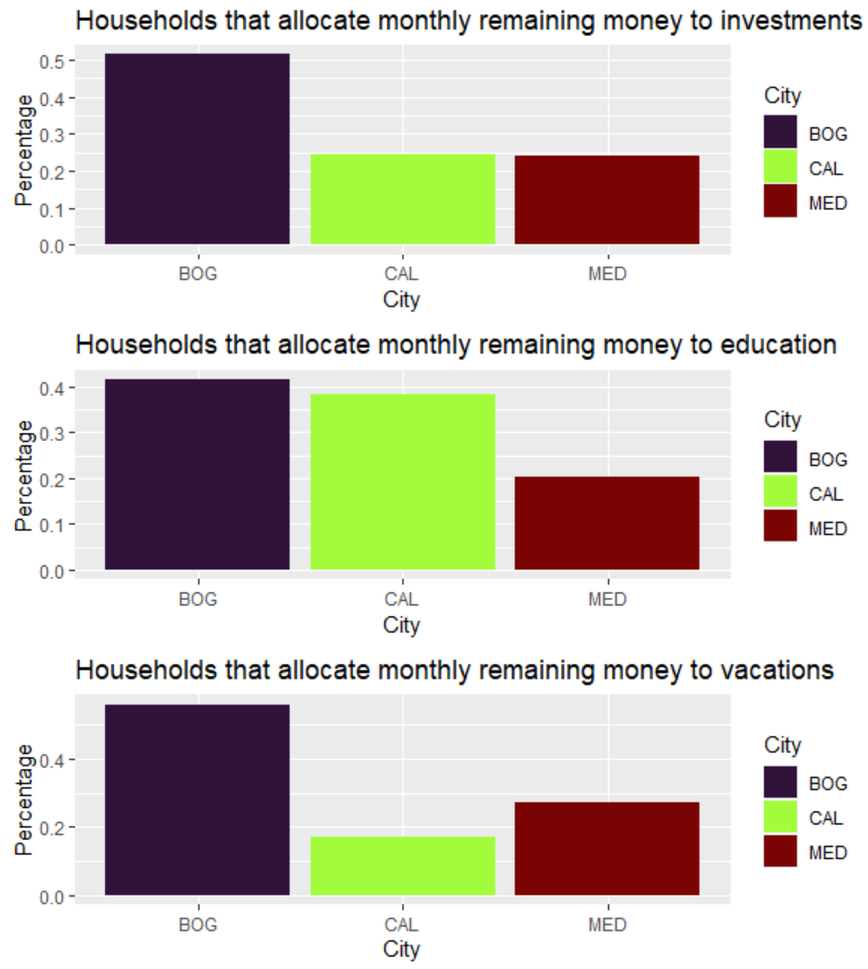

Figura 2: Comparison of expenses in the three cities

```
Inference example - two independent proportions
Proportions of saving for elderly years
HØ: P.bog=P.med
HA: Not equal
Statistics: Chi-square continuity correction
Computing.
Sample proportions:
P.bog=203/1651
P.med=68/951
p. value <0.001
Decision: Rejected HØ
Conclusion: Proportions are not equal, Bogotá and
Medellín.
```

Figura 3: Inference example

Descriptive Conclusions.

1. Only 19% of the households have remaining money to save.
2. Percentage dedicated to saving for elderly years : Bog: 12.2%; Cali: 9.3%; Med: 7.1%.

References

1. R (2023) Version 4.2.3. r-project.org.
2. Zeitlin W and es Auerbach C (2019): Basic statistics for the behavioral and social sciences using R. Oxford University Press.

Acknowledgements

Thanks to "Semillero de Investigación PSEUDOALEATORIOS". Departamento de Estadística. Specially to Venus Celeste Puertas Gualteros.

Apendix.

Some names of expenses: AVIV: Savings for housing purchase. JOY: Purchase of jewelry. VAC: Vacations. VJZ: For old age. EMER: Emergency cases. HER: Leaving an inheritance. AHRR: Saving.