Age and Income: An Analysis of Economic Well-Being Across Generations

1 Introduction

In our research, we explore the relationship between age and income, focusing on how generations differ economically. By identifying trends in income variation across generations, we can inform policies for different age groups. Our research contributes to informed economic planning, ensuring support and opportunities for all ages. Our analysis will consider how education and hours worked relate to age. Education influences earning potential and career opportunities, while hours worked reflect labor participation and financial stability. However, the survey-based data we're using is subject to biases and errors. Selected variables pose issues; for instance, education's dependence on age complicates estimating these variables' true effect. Our findings should be interpreted cautiously, considering the broader context and risks.

2 Brief Description of the Data

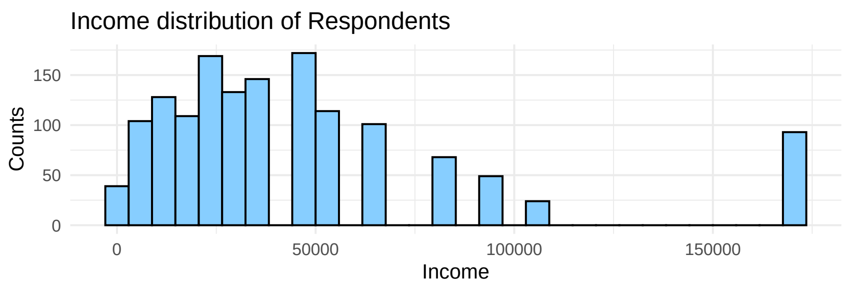
The data used in this analysis is derived from the General Social Survey (GSS), a widely-used survey conducted by the National Opinion Research Center (NORC) at the University of Chicago. The GSS is an observational study that collects data on a range of social, economic, and demographic variables from a nationally representative sample of adults in the United States. Each unit of observation in the GSS dataset is an individual respondent, selected through a multistage sampling method designed to ensure representativeness of the U.S. adult population.

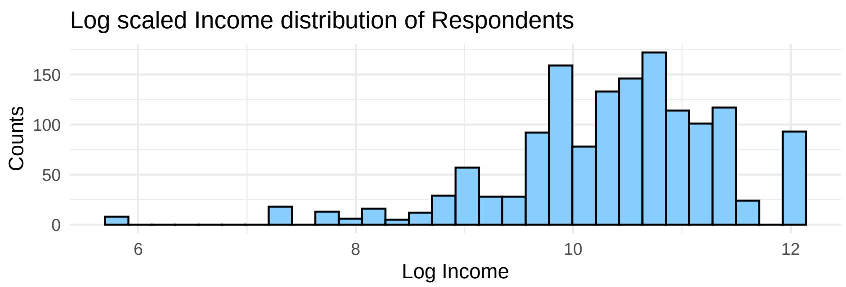
3 Discussion of How Key Concepts are Operationalized

In our analysis, we operationalize the key concepts of age and income using the following variables:
Age: The variable "age" represents the respondent's age at the time of the survey. This variable provides a straightforward measure of individuals' chronological age and serves as our X concept in the analysis.
Income: The variable "conrinc," represents the respondent's income in dollars and aligns with our research question of individual-level income outcomes. "Realrinc" was also considered but not used since it represents inflation-adjusted income in dollars, with a base year of 1986.
Other key concepts are operationalized as follows:
Education: The categorical variable "degree" was used to operationalize the concept of educational attainment. "Degree" provides each respondent's highest completed degree.
Hours Worked: The variable "hrs1," hours worked in the previous week, was used to operationalize the concept of labor force participation. "Hrs2," usual number of hours worked, was also considered but was not used since it contained a significant number of null values and appeared to be used primarily for respondents categorized as employed but not currently working due to sick leave, vacation, or on strike.

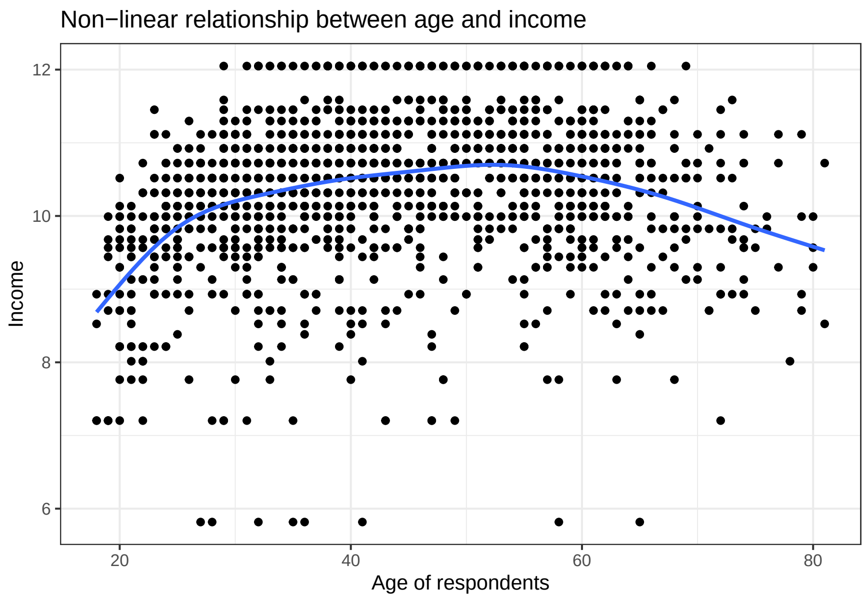
4 An Explanation of Key Modeling Decisions

We made key modeling decisions to ensure consistent statistical analysis. Our initial dataset had many invalid responses in both dependent and independent variables, coded as -100 for "inapplicable," -99 for "no answer," and so on. In "Conrinc", 1,544 of 3,544 observations were invalid; in "Hrs1", 1,606 were invalid. We processed these into dummy variables and then to null values (NA) for data integrity, allowing examination of their effect in our regression model. After omitting nulls, 1,449 observations remained. This refined dataset is for upcoming analysis. Observing "Conrinc"'s long-tail distribution, we applied a natural log transformation to the income data, creating "log_e_income" for our regression model.





5 Visualization



6 Well-Formatted Regression Table

Table 1: The relationship between natural log income and age

	Dependent variable:				
	log_e_income				
	(1)	(2)	(3)	(4)	(5)
AGE	0.013***	0.015***	0.013***	0.006	0.105***
	(0.002)	(0.002)	(0.002)	(0.005)	(0.020)
HRS1		0.025***	0.024***	0.022***	0.020***
		(0.002)	(0.002)	(0.002)	(0.002)
DEGREEAssociate/junior college			0.209**	0.175*	0.151
			(0.080)	(0.081)	(0.081)
DEGREEBachelor's			0.652***	0.636***	0.609***
			(0.054)	(0.053)	(0.053)
DEGREEGraduate			0.868***	0.826***	0.811***
			(0.061)	(0.062)	(0.061)
GenerationGen X				0.184	−0.158
				(0.100)	(0.114)
GenerationGen Z				−0.408	−0.258
				(0.228)	(0.230)
GenerationMillennials				0.007	−0.244
				(0.168)	(0.168)
GenerationSilent Generation				−0.826***	−0.038
				(0.242)	(0.283)
I(AGE^2)					−0.001***
					(0.0002)
Constant	9.790***	8.670***	8.453***	8.842***	7.050***
	(0.093)	(0.133)	(0.129)	(0.358)	(0.508)
Observations	1,449	1,449	1,449	1,449	1,449
R^2	0.032	0.147	0.274	0.300	0.313
Adjusted R^2	0.031	0.146	0.271	0.296	0.309
Residual Std. Error	0.990 (df = 1447)	0.929 (df = 1446)	0.859 (df = 1443)	0.844 (df = 1439)	0.836 (df = 1438)
Note:					*p<0.05; **p<0.01; ***p<0.001

7 Discussion of Results

Age emerged as a **statistically significant** predictor across models, indicating a consistent association with income. Thus, we can reject the null hypothesis that age has no income effect. Regression coefficients show the relationship between age and income, i.e., a positive coefficient suggests that income increases with age. This reflects career progression and experience accumulation, rewarded in the labor market. However, the relationship's magnitude, indicated by R-squared values, varies across models, from 0.0315 for model 1 to 0.3133 for model 5, showing that variance in income explained by models increases with additional variables.

The **practical significance** includes application to policy and workforce development. Model 1, with an R-squared of 0.0315, shows age alone does not account for much income variability, indicating other factors' influence. More complex models, like model 5 with R-squared 0.3133, include additional variables affecting income. This suggests the complexity of income determination and the need for a multi-faceted approach in economic policies. Interventions should consider various factors, emphasizing continuous learning and skill development for income growth.

8 Discussion of Limitations

IID. The design of the GSS suggests a degree of independence in the responses, i.e., it's cross-sectional nature, and stratified random sampling process. However, the use of multiple data collection methods (i.e., web, face-to-face, and telephone interviews) could influence how questions are interpreted and answered.

Linearity. Our models are based on the linearity assumption, supported by scatterplot and residual analysis. Despite this, we acknowledged potential non-linearity by considering transformations and interactions, but up to model 5, we avoided polynomial terms to keep the models interpretable.

Homoscedasticity. Residual analysis hinted at possible heteroscedasticity, indicating that income variance might change with age. We addressed this by transforming income and applying robust standard errors.

Normal Distribution of Residuals. Q-Q plots suggest that residuals mostly conform to normality, but deviations in the tails highlight potential non-normal distribution.

Multicollinearity. High VIF values for AGE and I(AGE^2) suggest caution in coefficient interpretation due to possible multicollinearity. The Generation variable also showed some multicollinearity, but not enough to cause significant concern, while HRS1 and DEGREE appeared unaffected, suggesting their coefficient estimates are reliable.

Omitted Variable Bias. While the models used key covariates, such as education and hours worked, there might be other relevant factors, such as industry or geographic location, that could also impact income. The absence of these variables could potentially lead to biased estimates.

Best Linear Predictor (BLP). Our model is designed to be the best linear representation of how independent variables predict log-transformed income. While transformations and robust error terms help meet BLP criteria, significant multicollinearity, especially with age variables, poses a challenge. Yet, the BLP is largely intact as our model effectively captures key trends.

9 Conclusion

This study reveals a significant, positive relationship between age and income, suggesting that income tends to increase as individuals grow older. This trend has important implications for understanding economic well-being across generations and can inform policies aimed at economic equity. The findings align with the initial research hypothesis that income varies across age groups, reflecting the progression of individuals' career and life stages. These results contribute to a broader discussion on income inequality and economic planning. They emphasize the necessity for nuanced policies that consider the varying economic needs of different age groups, ensuring equitable access to resources and opportunities for all.