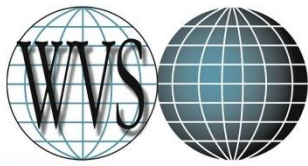


EXPLORATORY DATA ANALYSIS

on



World Value Surveys WAVES 1 to 7 For ARGENTINA

Author: Federico Sarrailh

Georgian College

August 20th, 2024

Presented To: Instructor Mr. Gladstone Johnathan

Table of Contents

| | |
|---|-----------|
| Introduction..... | 3 |
| Background..... | 3 |
| Research Questions..... | 3 |
| Expected Outcomes | 5 |
| Structure of the Analysis..... | 5 |
| Data Source and Considerations | 6 |
| Assumptions & Limitations | 7 |
| Processing the Data..... | 8 |
| PART 1: Extract, filter and preprocessing data for Argentina..... | 8 |
| PART 2: Univariate Analysis | 13 |
| PART 3: Hypothesis Testing & Forecasting | 40 |
| PART 4: Inferential Techniques | 57 |
| Results & Discussion..... | 67 |
| Future Work..... | 68 |
| Tracking this Analysis | 70 |
| Reference List..... | 72 |
| APPENDIX A | 74 |
| APPENDIX B | 77 |

Introduction

Background

The World Values Survey (WVS) is worldwide research that examines people's values, beliefs, and socio-cultural changes over time. The survey is carried out in nearly 100 countries and provides insights into how people's views on a wide range of topics such as social values, economic development, religion, opinions on government and politics, along with demographic measurements. The WVS seek to understand the impact of cultural values on social and political life by collecting data over multiple waves, every five or seven years, depending on the characteristics of the country. These surveys help researchers, governments and organizations to track trends in societal changes and make country comparisons.

WVS in Argentina

Personally, I have always been interested in human behavior, people's motivations, and the reasons that lead them to act or behave in certain ways. This work represents an interesting opportunity to explore these aspects through data analysis.

The behavior of Argentine citizens is often highlighted around the world for thinking "outside the box." Rules of the labor world, politics, social life and human creativity -among many others- in Argentina seem to have an unexpected twist. Sometimes, Argentine logic... seems to have little logic. Argentinians are also known for their resourcefulness, perseverance in the face of adversity, and warmth, even towards someone they have just met.

The curiosities briefly described in the previous paragraphs, along with the no less important fact that I am of Argentine origin, have led me to take an interest in and conduct this data analysis work.

Research Questions

The aim of this paper is to try to understand certain aspects that characterize and the reasons that motivate Argentine citizens. The focus is on selected aspects to analyze, these aspects are Happiness, the importance of Relationships, Religion, Politics, Responsibility, and People's Expectations.

Within them, we could spend hours asking interesting questions to extract information from the available survey data; however, we need to narrow the scope to a series of questions that will facilitate the practical execution of this analysis. Therefore, the following questions are formulated:

HAPPINESS AND SATISFACTION IN LIFE

Research Question:

"What are the characteristics that most influence an Argentine's happiness? "*En mis viejos tiempos éramos más felices!*" (Were we happier in the old days, as my grandfather used to say?)

[Written in my grandfather's original language, Spanish, *Reference 7*]

Supporting Questions:

- a. How do characteristics related to family, friends, and sense of belonging compare to work and income in terms of importance?
- b. Did the importance of these remain constant over the years?
- c. Is there a relationship between religion and the perception of happiness?

POLITICS AND SOCIETY

Research Question:

How does the population perceive the role of politics and its influence on the country and their lives?

Supporting Questions:

- d. Does the Argentine believe that the responsibility of their lives depends mostly on themselves, or do they place their expectations on the government?
- e. To what extent does interest in politics lead Argentine people to take political actions aimed at influencing society?
- f. To what extent does nationalism correlate with interest in politics?

Expected Outcomes

After each study and test conducted on this paper, the reader will find the results and a brief interpretation but, the ultimate goal of this paper is to answer the previous Research Questions as precisely as possible. This is stated in one of the last sections titled 'Results & Discussion'. The answers will emerge from the interpretation of the results from the various tests and their considerations all combined.

Structure of the Analysis

The analysis process on which this current paper is based is called Exploratory Data Analysis (EDA). It guides us from the available data to obtaining knowledge and answers to our questions, and it consists of a series of methods and techniques that must be followed in an orderly manner.

Parts on this EDA:

PART 1: Extract, Filter and Preprocessing data for Argentina

PART 2: Univariate Analysis

PART 3: Hypothesis Testing

PART 4: Inferential Techniques & Forecasting

I use MS Excel for all data manipulation, analysis, and most of the visualization for the entire project, although I must acknowledge that some initial steps of Extraction, Transformation, and Load (ETL) of the dataset were performed through Python in Jupyter Notebook using the Pandas library. Similarly, some 3D graphs during the inferential analysis stage were created using the Matplotlib library in Jupyter Notebook due to the complexity and/or impossibility of creating them in MS Excel. The codes and graphs are available in APPENDIX A and APPENDIX B.

Data Source and Considerations

The complete file constitutes the version V3.0 of the World Values Survey (WVS) time-series dataset for the period 1981-2022. This dataset combines WVS surveys completed in waves 1 (1981-1983); 2 (1990-1992); 3 (1995-1998); 4 (2000-2004); 5 (2005-2008); 6 (2010-2014), and 7 (2017-2022), across 106 countries around the world.

WVS time-series shows how the values of the given country/ society have been changing over time - rather than how the values of a selected group of people (panel) have been changing over their life. Therefore, methodologically, this WVS dataset is a time-series but it is not a panel data-file. National-wide representative samples of the adult population have been surveyed in every country/ territory in every wave, but there is no continuity between the samples across waves. Respondent ID numbers across the waves refer to different respondents and not the same person.

[Most of this text quoted from 'Release notes', *Reference 1*]

The totality of the World Value Surveys conducted in Argentina are considered in this work, starting from 1984 (WAVE 1) to the latest one conducted in 2017 (WAVE 7).

The data was retrieved from the official webpage, you can check it by the following this link:

<https://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp>

However, due to its size, it is not recommended to open the document in MS Excel as it is. Some filtering and extraction processes should be done first to focus on Argentina and the variables of interest, as explained in 'PART 1: Extract, Filter and Preprocess Data for Argentina'.

After that, you should obtain a dataset with 7,402 rows (data points) x 55 columns (variables) that is the starting point and the 'raw data' of this work. It is highly recommended to carefully read the 'Data Dictionary' and the "Tracking this Analysis" sections for a complete understanding. It contains the original codification names for each variable as well as the new given names used for this project and their descriptions.

EXTRA SOURCES of INFORMATION

In order to make a verification of the results found after doing univariate analysis, I gathered some data from Argentine population. I added commentaries on the variables, which are supported by the *Reference 4*, *Reference 5* and *Reference 8*.

Assumptions & Limitations

In this section, I outline some assumptions to consider when processing, analyzing and concluding about the data throughout this work.

- The data instances you will see in this work constitute a sample of the Argentine population over 7 non-consecutive years. Considering that the Argentine population is currently around 40 million inhabitants, 7401 data points may not be sufficient to answer questions about the population with certainty. One must proceed with great caution.
- No information is known about the geographical positioning or distribution of the samples across the country. Therefore, an implicit, interdependent, and unknown variable might be the geographical area of the country where certain parts of the instances were taken.
- There may probably be other unknown variables related to the sampling method used to select the respondents (Sampling Errors). Non-sampling errors may also exist due to interviewer errors, which are unknown in this work.
- The risk of not independence of the answers in the datapoints may exist. That means, since we are dealing with 7 surveys conducted over a period of 33 years, it certainly exists the risk that the same person would be interviewed more than once across the surveys. Then probably, that person answering in a certain survey may have (or not) a relationship with the same person answering in the next conducted survey. That is a risk on the statistic tests that requires independence between the variables. Said that, we are assuming this doesn't happen for the purposes of this work.
- Taking the above into account, and for the purposes of this work, the applied methods and the resulting conclusions will be based on the available data, and it will be assumed:
 - i. That the individuals (datapoints) are equally distributed across the country and in each province.
 - ii. The individual is a representative sample of the area or territory in which he lives.
 - iii. The risk of not independence does not occur for the purposes of this work.

Processing the Data

PART 1: Extract, filter and preprocessing data for Argentina

As already mentioned, it is not practical nor useful for the interest of this work to keep the complete information from the original downloaded dataset. I am focusing only on my country of origin: Argentina, and on just a few selected variables that I consider appropriate and oriented towards investigating into my Research Questions. That filtering process was done in the Jupyter Notebook environment through the Pandas Library, due to its ability to process large datasets quickly and sequentially. Also, it's common for survey questions to be identified with generic names like A001, A0022... X001, Q001, etc.; therefore, in this preprocessing section, meaningful names were also assigned to the variables representing the questions. All of this can be verified in the Python code in APPENDIX A. Or you can go directly to the new coded names and their meanings at the 'Variables' section of this work.

By the end of this first stage, the dataset consists of 7402 rows x 55 columns, and you can see it on the sheet: "WVS_TimeSeries_ARGENTINA" in the MS Excel file.

Following, I did a copy of that 'raw data' and I started working in the re-arrangement of the variables on the main working sheet: "WVS_study_ARGENTINA".

Throughout the development of this work, some additional variables lose relevance for various reasons and are discarded. They are mentioned at the time and in 'Tracking this Analysis' section. Many of the variables are measured in Lickert Scales and, due to the inconsistency in the measurement scales across the waves of the survey, I did some data adaptations and recoding. I acknowledge that this is a dangerous task since it can modify the significance of the data. It was re-checked several times. There was no remedy but to carry out the previous process. to ensure data consistency for the upcoming studies. All data ranges and limitations (e.g. missing and/or invalid values) are displayed under the 'Variables' title.

If you want to follow the exact technical process performed and mentioned in this section, please refer to the 'Tracking this Analysis' section or to Sheet 2: "Tracking Notes" in the MS Excel file. Finally for this section, I have to mention that prior to start working on the different methods in the MS Excel file, I have ordered the independent variables (according to my study) on the far left,

up to the variable 'RESPONDENT_NUM'. From there to the right, the dependent variables begin. The independent variables and the labels for all variables were frozen to facilitate navigation through the file.

About the 'missing values'

A priori, no rows or datapoints with missing or erroneous data in the variables are removed; they will be questioned and removed on a case-by-case basis when used in each study. However, they must be identified and accounted for before conducting the tests in 'PART 2: Univariate Analysis'.

Variables

I show in the following pages a table containing the selected variables for this work, their type, name encoding, range of values, missing or erroneous data, and the question they respond to in the survey.

For a better understanding of the variables, corresponding to the questions, I suggest downloading and reviewing the surveys and/ or check the 'Data Dictionary' sheet in the MS Excel document attached along with this paper.

[You can do that by downloading them from *Reference 1* and reviewing the preprocessing steps in *APPENDIX A*]

| Variable (In alpha order) | Type | Range / Possible Values | Missing / Erroneous Data | Description / Responds to the question of |
|---------------------------|---------------------------------|--|---------------------------------|--|
| AGE | Ratio Discrete | [16 - 93] | NO | Age of the interviewed |
| CONF_ARMED | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | NO | Confidence: Armed Forces |
| CONF_CHURCH | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | NO | Confidence: Churches |
| CONF_GOV | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Confidence: The Government |
| CONF_LABOURUN | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | NO | Confidence: Labour Unions |
| CONF_PARL | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | NO | Confidence: Parliament |
| CONF_POLPART | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Confidence: The Political Parties |
| CONF_PRESS | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | NO | Confidence: The Press |
| DATE_INT | Date/Time (Interval Continuous) | [yyyy-mm-dd] | No Data=-4 | Date interview |
| EMPLOY_STATUS | Nominal Polytomous | [1 - 8], DR = -2 | No Data=-4 | Employment status |
| ETHNIC_GROUP | Nominal Polytomous | [1 - 6], DR = -2 | No Data = -4, Wrong value = '9' | Ethnic group |
| FEEL_OF_HAPP | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | NO | Feeling of happiness |
| FREE_CHO_CON | Ordinal Polytomous | [1 - 10], DK = -1, DR = -2 | NO | How much freedom of choice and control |
| GOV_RESP | Interval Discrete | [1 - 10], DK = -1, DR = -2 | No Data=-4 | Level of Government responsibility on |
| HOW_MANY_CHILD | Ratio Discrete | [0 - 5], DR = -2 | Missing/ NA = -5 | How many children do you have |
| INCOME_SCALE | Ordinal Polytomous | [1 - 10], DK = -1, DR = -2 | No Data=-4, for WAVE 1 & WAVE 5 | Scale of incomes |
| IP_LIFE_FAM | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Important in life: Family |
| IP_LIFE_FRI | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Important in life: Friends |
| IP_LIFE_LEI | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4, mostly WAVE 1 | Important in life: Leisure time |
| IP_LIFE_POL | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Important in life: Politics |
| IP_LIFE_REL | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Important in life: Religion |
| IP_LIFE_WOR | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4, for WAVE 1 | Important in life: Work |
| LEVEL_EDU_COUNTRY | Ordinal Polytomous | [1 - 10] | No Data=-4 | Education (country specific) |
| MARITAL_STA | Nominal Polytomous | [1 - 6], DK = -1, DR = -2 | NO | Marital status |
| MEM_HUM_ORG | Nominal Polytomous | [0 - 2], DR = -2 | No Data=-4 | Active/Inactive membership of charitable/humanitarian organization |
| MEM_LAB_UNI | Nominal Polytomous | [0 - 2], DR = -2 | No Data=-4 | Active/Inactive membership of labour unions |
| MEM_MUS_EDU | Nominal Polytomous | [0 - 2], DR = -2 | No Data=-4 | Active/Inactive membership of art, music, educational |
| MEM_POL_PAR | Nominal Polytomous | [0 - 2], DR = -2 | No Data=-4 | Active/Inactive membership of political party |
| MEM_REL_ORG | Nominal Polytomous | [0 - 2], DR = -2 | No Data=-4 | Active/Inactive membership of church or religious organization |
| MEM_SP_REC | Nominal Polytomous | [0 - 2], DR = -2 | No Data=-4 | Active/Inactive membership of sport or recreation |
| POL_ATT_DEM | Nominal Polytomous | [1 - 3], DK = -1, DR = -2 | NO | Political action: attending lawful/peaceful demonstrations |
| POL_INT | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2 | NO | Interest in politics |
| POL_JOI_BOY | Nominal Polytomous | [1 - 3], DK = -1, DR = -2 | NO | Political action: joining in boycotts |
| POL_SYS_ARMY | Nominal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Political system: Having the army rule |
| POL_SYS_DEM | Nominal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Political system: Having a democratic political system |
| POL_SYS_EXPERT | Nominal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Political system: Having experts make decisions |
| POL_SYS_LEADER | Nominal Polytomous | [1 - 4], DK = -1, DR = -2 | No Data=-4 | Political system: Having a strong leader |
| POL_UNOFF_STR | Nominal Polytomous | [1 - 3], DK = -1, DR = -2 | No Data=-4 | Political action: joining unofficial strikes |
| PROUD_OF_NAT | Ordinal Polytomous | [1 - 4], DK = -1, DR = -2, Not appl.=-3 | NO | How proud of nationality |
| REL_GROUPS | Nominal Polytomous | [0 - 8], DK = -1, DR = -2 | No Data=-4 | Religious denominations - major groups |
| RESPONDENT_NUM | Nominal Polytomous | [120001 - 721003] | | Number assigned to the interviewed |
| SAT_W_LIFE | Ordinal Polytomous | [1 - 10], DK = -1, DR = -2 | NO | Satisfaction with your life |
| SETTL_SIZE | Ordinal Polytomous | [1 - 8] | No Data=-4 | Settlement size |
| SETTL_TYPE | Nominal Polytomous | [1 - 5] | No Data=-4 | Settlement type where interview was conducted |
| SEX | Nominal Dichotomous | 1: MALE, 2: FEMALE | NO | Sex of the interviewed |
| WVS_WAVES | Interval Discrete | [1 - 7] | NO | Chronology of EVS-WVS waves |
| YEAR_OF_BIRTH | Interval Discrete | [1900 - 1999] | NO | Year of birth |
| YEAR_SURVEY | Interval Discrete | [1984, 1991, 1995, 1999, 2006, 2013, 2017] | NO | Year of the survey in ARGENTINA |
| YEAR_SVY_START | Date/Time (Interval Continuous) | [yyyy-mm] | NO | Year/month of start-fieldwork |

Image 1. Characteristics of the variables.

| Original Name | Internal Survey Category | Variable (In alpha order) |
|---------------|---|---------------------------|
| X003 | DEMOGRAPHICS | AGE |
| E069_02 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | CONF_ARMED |
| E069_01 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | CONF_CHURCH |
| E069_11 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | CONF_GOV |
| E069_05 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | CONF_LABOURUN |
| E069_07 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | CONF_PARL |
| E069_12 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | CONF_POLPART |
| E069_04 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | CONF_PRESS |
| S012 | OBSERVATIONS BY THE INTERVIEWER | DATE_INT |
| X028 | DEMOGRAPHICS | EMPLOY_STATUS |
| X051 | DEMOGRAPHICS | ETHNIC_GROUP |
| A008 | HAPPINESS AND WELL-BEING | FEEL_OF_HAPP |
| A173 | HAPPINESS AND WELL-BEING | FREE_CHO_CON |
| E037 | ECONOMIC VALUES | GOV_RESP |
| X011 | DEMOGRAPHICS | HOW_MANY_CHILD |
| X047_WVS | DEMOGRAPHICS | INCOME_SCALE |
| A001 | SOCIAL VALUES, ATTITUDES & STEREOTYPES | IP_LIFE_FAM |
| A002 | SOCIAL VALUES, ATTITUDES & STEREOTYPES | IP_LIFE_FRI |
| A003 | SOCIAL VALUES, ATTITUDES & STEREOTYPES | IP_LIFE_LEI |
| A004 | SOCIAL VALUES, ATTITUDES & STEREOTYPES | IP_LIFE_POL |
| A006 | SOCIAL VALUES, ATTITUDES & STEREOTYPES | IP_LIFE_REL |
| A005 | SOCIAL VALUES, ATTITUDES & STEREOTYPES | IP_LIFE_WOR |
| X025CSWVS | DEMOGRAPHICS | LEVEL_EDU_COUNTRY |
| X007 | DEMOGRAPHICS | MARITAL_STA |
| A105 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | MEM_HUM_ORG |
| A101 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | MEM_LAB_UNI |
| A100 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | MEM_MUS_EDU |
| A102 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | MEM_POL_PAR |
| A098 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | MEM_REL_ORG |
| A099 | SOCIAL CAPITAL, TRUST & ORGANIZATIONAL MEMBERSHIP | MEM_SP_REC |
| E027 | POLITICAL INTEREST & POLITICAL PARTICIPATION | POL_ATT_DEM |
| E023 | POLITICAL INTEREST & POLITICAL PARTICIPATION | POL_INT |
| E026 | POLITICAL INTEREST & POLITICAL PARTICIPATION | POL_JOI_BOY |
| E116 | POLITICAL CULTURE & POLITICAL REGIMES | POL_SYS_ARMY |
| E117 | POLITICAL CULTURE & POLITICAL REGIMES | POL_SYS_DEM |
| E115 | POLITICAL CULTURE & POLITICAL REGIMES | POL_SYS_EXPERT |
| E114 | POLITICAL CULTURE & POLITICAL REGIMES | POL_SYS_LEADER |
| E028 | POLITICAL INTEREST & POLITICAL PARTICIPATION | POL_UNOFF_STR |
| G006 | POLITICAL CULTURE & POLITICAL REGIMES | PROUD_OF_NAT |
| F025 | DEMOGRAPHICS | REL_GROUPS |
| S007 | | RESPONDENT_NUM |
| A170 | HAPPINESS AND WELL-BEING | SAT_W_LIFE |
| X049 | OBSERVATIONS BY THE INTERVIEWER | SETTL_SIZE |
| X050B | OBSERVATIONS BY THE INTERVIEWER | SETTL_TYPE |
| X001 | DEMOGRAPHICS | SEX |
| S002VS | | WVS_WAVES |
| X002 | DEMOGRAPHICS | YEAR_OF_BIRTH |
| S020 | | YEAR_SURVEY |
| S022 | | YEAR_SVY_START |

Image 2. Categories and original name of the variables within the Survey context.

Document Visuals

I present here a visual overview of the initial MS Excel document built as above mentioned.

Raw unorganized data on Sheet 1:

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
|----|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|-------------|-------------|-------------|-------------|------------|--------------|---------------|---------|-------------|-------------|---------------|
| 1 | MAIN_INDEX | IP_LIFE_FAM | IP_LIFE_FRI | IP_LIFE_LEI | IP_LIFE_POL | IP_LIFE_WOR | IP_LIFE_REL | FEEL_OF_HAP | MEM_REL_ORG | MEM_SP_REC | MEM_MUS_EDU | MEM_LAB_UNI | MEM_POL_PAR | MEM_HUM_ORG | SAT_W_LIFE | FREE_CHO_CON | FIGHT_FOR_COU | POL_INT | POL_JOI_BOY | POL_ATT_DEM | POL_UNOFF_STR |
| 2 | 0 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 8 | 10 | -1 | 3 | 3 | 3 | -4 |
| 3 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 10 | 1 | 3 | 3 | 3 | -4 |
| 4 | 2 | 1 | 1 | 2 | 4 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 10 | 0 | 4 | 3 | 3 | -4 |
| 5 | 3 | 1 | 2 | 1 | 3 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | -2 | 1 | 2 | 3 | 1 | -4 |
| 6 | 4 | 1 | 1 | 1 | 3 | 1 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 9 | 0 | 3 | 3 | 3 | -4 |
| 7 | 5 | 1 | 1 | 2 | 1 | 4 | 1 | 2 | 2 | 0 | 1 | 0 | 1 | 0 | 9 | 5 | 0 | 1 | 3 | 2 | -4 |
| 8 | 6 | 1 | 1 | 1 | 2 | 1 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 5 | -1 | 2 | 3 | 1 | -4 |
| 9 | 7 | 1 | 3 | 3 | 2 | 1 | 2 | 2 | 0 | 0 | 0 | 1 | 0 | 1 | 8 | 7 | 1 | 2 | 3 | 3 | -4 |
| 10 | 8 | 1 | 1 | 1 | 2 | 2 | 3 | 2 | 0 | 2 | 1 | 0 | 0 | 0 | 8 | 7 | 1 | 2 | -1 | 2 | -4 |
| 11 | 9 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 1 | 0 | 0 | 1 | 0 | 0 | 8 | 6 | 1 | 3 | 3 | 2 | -4 |
| 12 | 10 | 1 | 2 | 1 | 2 | 1 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | -2 | 0 | 1 | 3 | 1 | -4 |
| 13 | 11 | 1 | 1 | 3 | 2 | 1 | 3 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 8 | 10 | 0 | 3 | 3 | 3 | -4 |
| 14 | 12 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 8 | 10 | 0 | 2 | 3 | 2 | -4 |
| 15 | 13 | 1 | 4 | 2 | 1 | 1 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 8 | 0 | 2 | -1 | 2 | -4 |
| 16 | 14 | 1 | 3 | 1 | 2 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 8 | 0 | 2 | 3 | 3 | -4 |
| 17 | 15 | 1 | 2 | 2 | 2 | 1 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 8 | 0 | 2 | 2 | 1 | -4 |
| 18 | 16 | 1 | 2 | 3 | 2 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 10 | -1 | 2 | 3 | 2 | -4 |
| 19 | 17 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 5 | 1 | 3 | -1 | -1 | -4 |
| 20 | 18 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 10 | 1 | 2 | 3 | 2 | -4 |
| 21 | 19 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 10 | 0 | 3 | 3 | 2 | -4 |
| 22 | 20 | 2 | 1 | 1 | 4 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 7 | 6 | -1 | 4 | 3 | 3 | -4 |
| 23 | 21 | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 5 | 1 | 2 | 2 | 2 | -4 |
| 24 | 22 | 2 | 2 | 2 | 4 | 2 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 4 | 0 | 4 | -1 | 2 | -4 |
| 25 | 23 | 1 | 2 | 2 | 3 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 4 | 1 | 3 | 2 | 1 | -4 |
| 26 | 24 | 1 | 1 | 1 | 3 | 2 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 7 | 0 | 3 | 2 | 2 | -4 |
| 27 | 25 | 1 | 1 | 1 | 3 | 2 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 8 | -1 | 2 | -1 | -1 | -4 |
| 28 | 26 | 1 | 2 | 4 | 3 | 2 | 4 | 4 | 2 | 1 | 0 | 0 | 0 | 0 | 8 | 8 | 0 | 3 | 1 | 1 | -4 |
| 29 | 27 | 2 | 2 | 3 | 4 | 4 | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | 6 | 0 | 4 | 3 | 3 | -4 |
| 30 | 28 | 1 | 1 | 1 | 4 | 1 | 2 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 8 | 8 | 0 | 3 | -1 | -1 | -4 |
| 31 | 29 | 2 | 2 | 2 | 4 | 1 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 8 | -1 | 4 | 2 | 1 | -4 |
| 32 | 30 | 1 | 1 | 1 | 4 | 2 | 4 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 7 | 9 | 0 | 4 | 3 | 1 | -4 |
| 33 | 31 | 1 | 1 | 1 | 4 | 1 | 4 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 9 | 8 | -1 | 4 | 3 | 1 | -4 |
| 34 | 32 | 2 | 3 | 2 | 2 | 3 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 8 | 0 | 2 | 2 | 2 | -4 |
| 35 | 33 | 1 | 1 | 1 | 4 | 2 | 2 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 9 | 8 | 0 | 4 | 2 | 1 | -4 |
| 36 | 34 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 7 | 8 | 0 | 2 | 1 | 1 | -4 |
| 37 | 35 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 6 | 7 | 0 | 2 | 3 | 1 | -4 |
| 38 | 36 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 9 | 10 | -1 | 2 | 3 | 2 | -4 |
| 39 | 37 | 1 | 1 | 1 | 4 | 1 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 7 | -1 | 4 | 3 | 3 | -4 |

<

>

WVS_TimeSeries_ARGENTINA

Tracking Notes

Data Dictionary

7WVS_TimeSeries_ARGENTINA

+

Ready Accessibility: Investigate

Display Settings

<

Image 3. Raw unorganized data on Sheet 1: WVS_TimeSeries_ARGENTINA.

Organized data on Sheet 4:

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V |
|----|----------|-------------|--------------|---------------|-----|-----|----------------|------------|----------------|--------------|---------------|-----------|---------|-------------|----------------|------------|------------|--------------|-------|---|---|
| 1 | WVS_WAVE | YEAR_SURVEY | ETHNIC_GROUP | YEAR_OF_BIRTH | SEX | AGE | RESPONDENT_NUM | DATE_INT | YEAR_SVY_START | INCOME_SCALE | EMPLOY_STATUS | LEVEL_EDU | COUNTRY | MARITAL_STA | HOW_MANY_CHILD | SETTL_SIZE | SETTL_TYPE | FEEL_OF_HAPP | SAT_M | | |
| 2 | 7 | 2017 | 1 | 1966 | 1 | 50 | 720001 | 2017-07-09 | 201707 | 5 | 1 | 5 | 5 | 5 | 5 | -5 | 8 | 1 | 2 | | |
| 3 | 7 | 2017 | 1 | 1963 | 2 | 34 | 720002 | 2017-07-09 | 201707 | 3 | 1 | 7 | 6 | 6 | 6 | 0 | 8 | 1 | 2 | | |
| 4 | 7 | 2017 | 1 | 1982 | 1 | 35 | 720003 | 2017-07-10 | 201707 | 5 | 1 | 7 | 6 | 6 | 6 | 0 | 8 | 1 | -1 | | |
| 5 | 7 | 2017 | 1 | 1946 | 2 | 71 | 720004 | 2017-07-10 | 201707 | 7 | 4 | 5 | 1 | 1 | 8 | 1 | 8 | 1 | 1 | | |
| 6 | 7 | 2017 | 1 | 1979 | 2 | 37 | 720005 | 2017-07-10 | 201707 | 4 | 3 | 6 | 4 | 1 | 8 | 1 | 8 | 1 | 1 | | |
| 7 | 7 | 2017 | 1 | 1959 | 2 | 58 | 720006 | 2017-07-11 | 201707 | 4 | 4 | 4 | 1 | 5 | 8 | 1 | 8 | 1 | 3 | | |
| 8 | 7 | 2017 | 2 | 1985 | 1 | 32 | 720007 | 2017-07-11 | 201707 | 5 | 3 | 5 | 6 | 1 | 8 | 1 | 8 | 1 | 3 | | |
| 9 | 7 | 2017 | 2 | 1983 | 2 | 34 | 720008 | 2017-07-11 | 201707 | 4 | 2 | 5 | 2 | 2 | 8 | 1 | 8 | 1 | 3 | | |
| 10 | 7 | 2017 | 1 | 1952 | 1 | 65 | 720009 | 2017-07-11 | 201707 | 5 | 3 | 5 | 1 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 11 | 7 | 2017 | 1 | 1990 | 1 | 27 | 720010 | 2017-07-11 | 201707 | 5 | 3 | 5 | 6 | 2 | 8 | 1 | 8 | 1 | 1 | | |
| 12 | 7 | 2017 | 2 | 1977 | 2 | 40 | 720011 | 2017-07-13 | 201707 | 4 | 5 | 5 | 1 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 13 | 7 | 2017 | 1 | 1973 | 1 | 44 | 720012 | 2017-07-13 | 201707 | 3 | 1 | 9 | 6 | 1 | 8 | 1 | 8 | 1 | 2 | | |
| 14 | 7 | 2017 | 2 | 1996 | 2 | 21 | 720013 | 2017-07-13 | 201707 | 2 | 6 | 5 | 6 | 1 | 8 | 1 | 8 | 1 | 2 | | |
| 15 | 7 | 2017 | 9 | 1996 | 1 | 21 | 720014 | 2017-07-13 | 201707 | 4 | 1 | 7 | 6 | 4 | 8 | 1 | 8 | 1 | 2 | | |
| 16 | 7 | 2017 | 1 | 1946 | 2 | 71 | 720015 | 2017-07-13 | 201707 | 6 | 4 | 4 | 5 | 2 | 8 | 1 | 8 | 1 | 3 | | |
| 17 | 7 | 2017 | 1 | 1984 | 1 | 33 | 720016 | 2017-07-08 | 201707 | 5 | 1 | 5 | 1 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 18 | 7 | 2017 | 1 | 1955 | 2 | 61 | 720017 | 2017-07-08 | 201707 | 5 | 4 | 8 | 1 | 3 | 8 | 1 | 8 | 1 | 2 | | |
| 19 | 7 | 2017 | 2 | 1976 | 1 | 41 | 720018 | 2017-07-08 | 201707 | 4 | 2 | 4 | 4 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 20 | 7 | 2017 | 1 | 1953 | 1 | 64 | 720019 | 2017-07-08 | 201707 | 5 | 2 | 3 | 2 | 1 | 8 | 1 | 8 | 1 | 2 | | |
| 21 | 7 | 2017 | 1 | 1987 | 2 | 30 | 720020 | 2017-07-08 | 201707 | 6 | 1 | 9 | 4 | 2 | 8 | 1 | 8 | 1 | 1 | | |
| 22 | 7 | 2017 | 2 | 1947 | 2 | 69 | 720021 | 2017-07-11 | 201707 | 5 | 4 | 9 | 5 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 23 | 7 | 2017 | 1 | 1930 | 1 | 87 | 720022 | 2017-07-11 | 201707 | 5 | 4 | 3 | 5 | 1 | 8 | 1 | 8 | 1 | 1 | | |
| 24 | 7 | 2017 | 3 | 1985 | 2 | 31 | 720023 | 2017-07-11 | 201707 | 3 | 1 | 4 | 1 | 2 | 8 | 1 | 8 | 1 | 1 | | |
| 25 | 7 | 2017 | 1 | 1961 | 2 | 36 | 720024 | 2017-07-11 | 201707 | 1 | 1 | 8 | 1 | 3 | 8 | 1 | 8 | 1 | 3 | | |
| 26 | 7 | 2017 | 1 | 1966 | 1 | 51 | 720025 | 2017-07-11 | 201707 | 6 | 3 | 5 | 1 | 1 | 8 | 1 | 8 | 1 | 2 | | |
| 27 | 7 | 2017 | 1 | 1986 | 1 | 31 | 720026 | 2017-07-09 | 201707 | 6 | 3 | 4 | 1 | 2 | 8 | 1 | 8 | 1 | 1 | | |
| 28 | 7 | 2017 | 1 | 1984 | 2 | 33 | 720027 | 2017-07-09 | 201707 | 5 | 1 | 7 | 3 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 29 | 7 | 2017 | 3 | 1962 | 1 | 55 | 720028 | 2017-07-09 | 201707 | 9 | 3 | 9 | 3 | 2 | 8 | 1 | 8 | 1 | 1 | | |
| 30 | 7 | 2017 | 2 | 1975 | 2 | 42 | 720029 | 2017-07-09 | 201707 | 6 | 1 | 4 | 1 | 1 | 8 | 1 | 8 | 1 | 2 | | |
| 31 | 7 | 2017 | 1 | 1964 | 2 | 52 | 720030 | 2017-07-09 | 201707 | 9 | 3 | 5 | 4 | 3 | 8 | 1 | 8 | 1 | 1 | | |
| 32 | 7 | 2017 | 2 | 1990 | 1 | 27 | 720031 | 2017-07-07 | 201707 | 3 | 3 | 6 | 6 | 2 | 8 | 1 | 8 | 1 | 1 | | |
| 33 | 7 | 2017 | 2 | 1995 | 2 | 21 | 720032 | 2017-07-07 | 201707 | 5 | 2 | 7 | 6 | 2 | 8 | 1 | 8 | 1 | -1 | | |
| 34 | 7 | 2017 | 1 | 1995 | 1 | 21 | 720033 | 2017-07-14 | 201707 | 7 | 1 | 7 | 6 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 35 | 7 | 2017 | 1 | 1974 | 2 | 43 | 720034 | 2017-07-07 | 201707 | 4 | 3 | 5 | 1 | 2 | 8 | 1 | 8 | 1 | 3 | | |
| 36 | 7 | 2017 | 2 | 1970 | 2 | 47 | 720035 | 2017-07-07 | 201707 | 5 | 2 | 5 | 6 | 2 | 8 | 1 | 8 | 1 | 2 | | |
| 37 | 7 | 2017 | 1 | 1946 | 2 | 70 | 720036 | 2017-07-04 | 201707 | 4 | 3 | 9 | 5 | 3 | 8 | 1 | 8 | 1 | 2 | | |
| 38 | 7 | 2017 | 1 | 1982 | 2 | 35 | 720037 | 2017-07-11 | 201707 | 6 | 4 | 8 | 1 | 2 | 8 | 1 | 8 | 1 | 1 | | |

Image 4. Organized data on Sheet 4: WVS_study_ARGENTINA.

PART 2: Univariate Analysis

In this Section I will provide a thorough description of each variable of interest, that will help us to understand the sample of the Argentine population we have and, will lead us a step closer to the answers we are looking for in the Research Questions. The examination of these variables through descriptive statistics is the first step to understanding the data provided by the survey.

I will perform cleansing of erroneous data and coding to make data more consistent for calculus purposes here, and in further steps. We will review the numbers in search of outliers and analyze the distributions of the variables to decide if they can be used for further studies, or if they should be discarded. Also, I will look for measures of skewness and kurtosis in frequency distributions, with comparison to 'normal' distributions. So, to summarize, the tasks to be performed for each variable of interest are:

- A. Preliminary Cleaning and Coding
- B. Examination
- C. Visualization

A: Cleaning of missing, erroneous or not applicable data is a fundamental task to be performed diligently so that these values do not affect the real data hidden within the variables. I also apply coding or categorization to simplify, translate, and make the data more consistent to improve statistical analysis.

B: Examination of the results that the statistical methods or the graphs are showing us, deciding whether to conduct further information searches from external sources to compare this data, understanding if we have signs that the sample we have is representative of the population in which the study was conducted. The numerical examination of the methods will also help us decide if we are dealing with normal distributions that will allow us to apply other statistical methods to relate variables in the future, help us identify possible outliers that we need to be careful with, and make the first assumptions about what answers the data we have might lead us to.

C: The visual representation adds fundamental information to the study, it helps us understand how the data is distributed within the context of each variable. It allows us to quickly get a deeper understanding and draw some preliminary conclusions.

Following, a description of each variable of interest separated in categories according to the aspects mentioned in 'Research Questions'.

DEMOGRAPHICS

YEAR_SURVEY & WVS WAVES

Year in which the surveys were conducted and its correspondent wave number.

- A. Cleaning and Coding: No action taken. No missing and no erroneous data.
- B. Examination: The survey was conducted in Argentina in the years shown in the table, at intervals of 4 and 7 years. A similar number of people were interviewed each year, except for a slight increase in WAVE 4.
- C. Visualization:

| WAVES | Count of YEAR_SURVEY | YEAR |
|--------------------|----------------------|------|
| 1 | 1005 | 1984 |
| 2 | 1002 | 1991 |
| 3 | 1079 | 1995 |
| 4 | 1280 | 1999 |
| 5 | 1002 | 2006 |
| 6 | 1030 | 2013 |
| 7 | 1003 | 2017 |
| Grand Total | 7401 | |

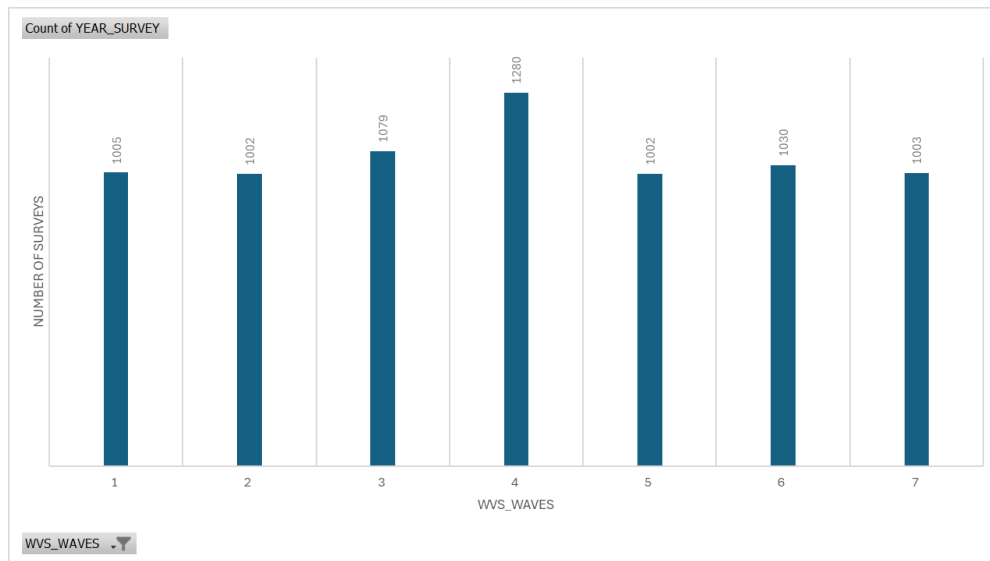


Image 5. Summary table and histogram for WAVES correspondent to the surveys.

AGE

Age of the person interviewed.

- A. Cleaning and Coding: No action taken. No missing and no erroneous data.
- B. Examination: Wide age ranges from 16 to 93 years old, average age is 42.5 years old. $|Skewness| \leq 2$ and $|Kurtosis| \leq 2$ so we can assume normal distribution. No possible outliers were detected since there are no values beyond $1.5 \times IQR$ rule. We can observe that the most frequent age among the interviewees was 24 years. More statistical data can be found in the table below.
- C. Visualization:

| AGE | |
|-------------------------|--------------|
| Mean | 42.49196055 |
| Standard Error | 0.199351473 |
| Median | 40 |
| Mode | 24 |
| Standard Deviation | 17.15002077 |
| Sample Variance | 294.1232124 |
| Kurtosis | -0.849153816 |
| Skewness | 0.420321496 |
| Range | 77 |
| Minimum | 16 |
| Maximum | 93 |
| Sum | 314483 |
| Count | 7401 |
| Confidence Level(95.0%) | 0.390785624 |

Image 6. Descriptive statistics for AGE.

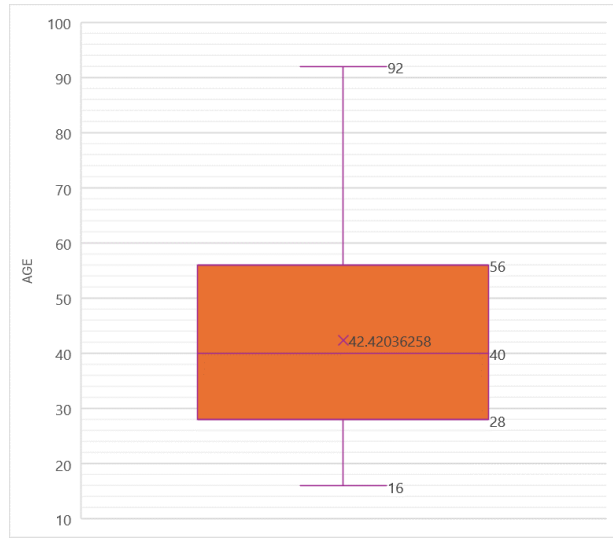


Image 7. Boxplot for AGE.

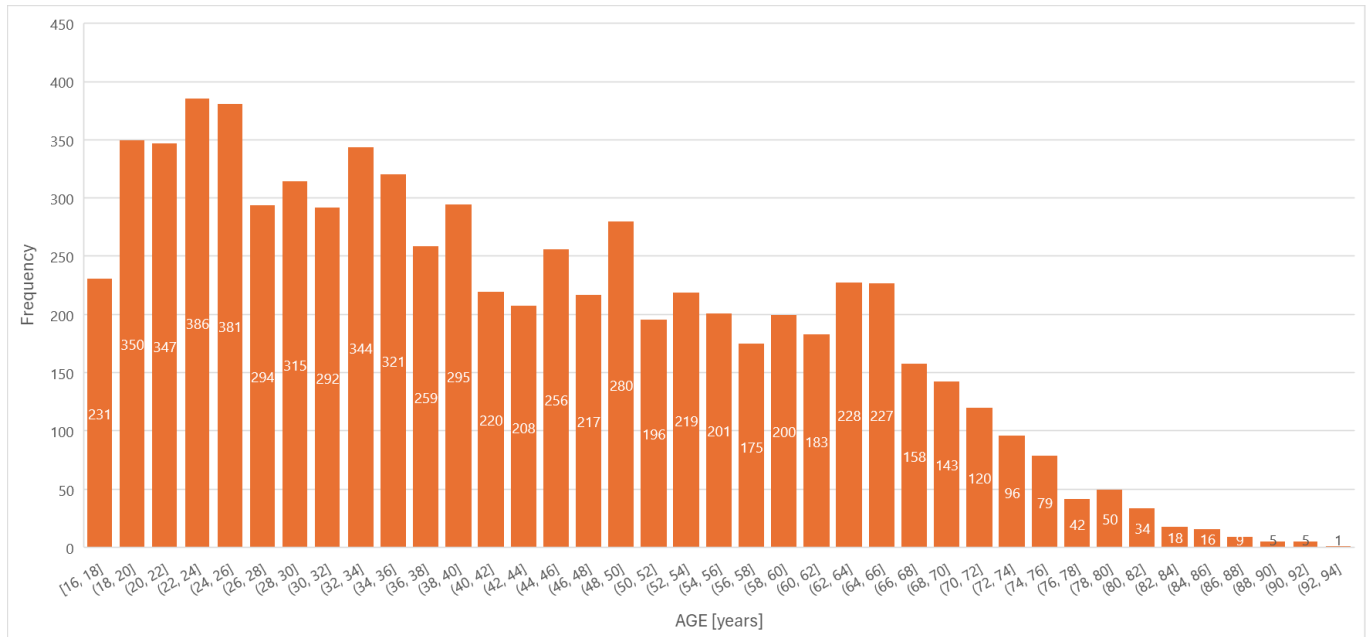


Image 8. Histogram for AGE

Further Analysis for AGE:

It is important to understand if this frequency distribution for AGE is representative of the overall population. Determining this can be an advantage for subsequent analyses and conclusions, as it enables us to make inferences about the Argentine population in general, based on the results of the studies of this sample.

The data for the Argentine population distribution was extracted from *Reference 8*.

There is no data in these surveys from people in the 0 to 14 age groups, therefore they were excluded from the comparison.

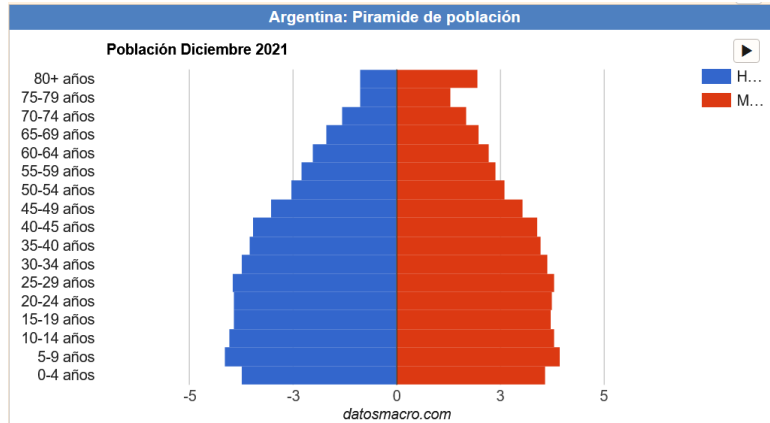


Image 9. Argentina's Population pyramid (December 2021), Reference 8.

| AGE | Sample Freq. | Sample % | Population % | Population % scaled | difference |
|----------------|--------------|----------|--------------|---------------------|------------|
| 15 - 19 | 417 | 5.63 | 7.68 | 10.01956947 | 4.39 |
| 20 - 24 | 897 | 12.12 | 7.70 | 10.0456621 | -2.07 |
| 25 - 29 | 840 | 11.35 | 7.79 | 10.16307893 | -1.19 |
| 30 - 34 | 786 | 10.62 | 7.41 | 9.667318982 | -0.95 |
| 35 - 39 | 713 | 9.63 | 7.07 | 9.223744292 | -0.41 |
| 40 - 44 | 590 | 7.97 | 6.91 | 9.015003262 | 1.04 |
| 45 - 49 | 614 | 8.30 | 6.12 | 7.984344423 | -0.31 |
| 50 - 54 | 554 | 7.49 | 5.18 | 6.757990868 | -0.73 |
| 55 - 59 | 457 | 6.17 | 4.71 | 6.14481409 | -0.03 |
| 60 - 64 | 530 | 7.16 | 4.26 | 5.557729941 | -1.60 |
| 65 - 69 | 446 | 6.03 | 3.72 | 4.853228963 | -1.17 |
| 70 - 74 | 298 | 4.03 | 3.03 | 3.953033268 | -0.07 |
| 75 - 79 | 148 | 2.00 | 2.22 | 2.8962818 | 0.90 |
| >= 80 | 111 | 1.50 | 2.85 | 3.718199609 | 2.22 |
| TOTAL | | 100% | 76,65% | 100% | |
| Scaling Factor | 1.304631442 | | | | |

Image 10. Comparison table between the samples of the surveys vs. Argentine population.

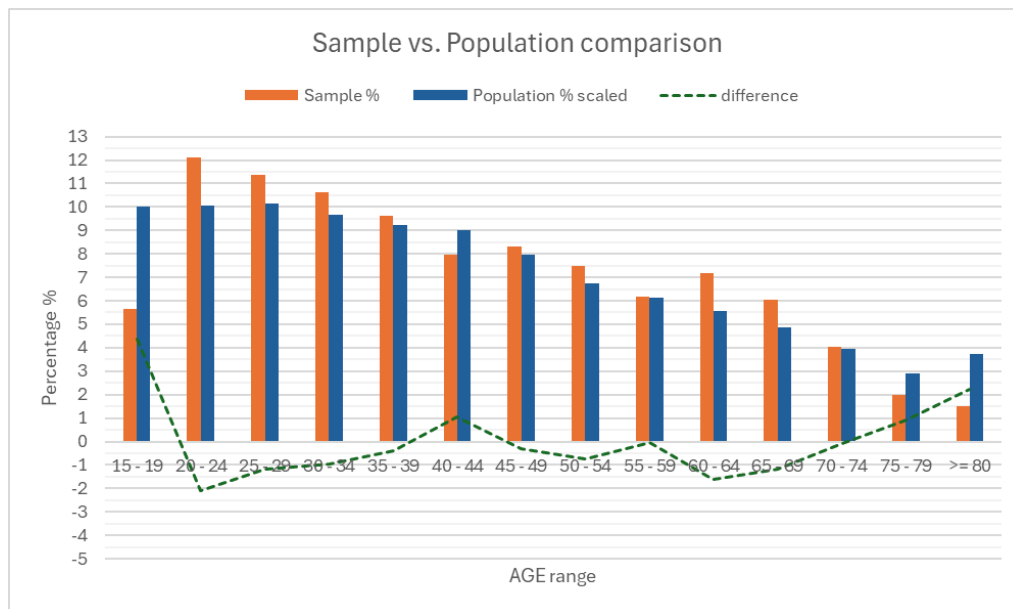


Image 11. Comparison graph between the samples of the surveys vs. Argentine population.

According to the variations shown by the green dashed line, there seem to be significant differences. To confirm this, we can perform a Chi-Square test. This test located here in this univariate analysis section is an exception, given the importance of the issue.

Chi-Square of Sample vs. Population

| | AGE | 15 - 19 | 20 - 24 | 25 - 29 | 30 - 34 | 35 - 39 | 40 - 44 | 45 - 49 | 50 - 54 | 55 - 59 | 60 - 64 | 65 - 69 | 70 - 74 | 75 - 79 | >= 80 | Grand Total |
|-----------------------|-------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--------|-------------|
| SAMPLE = observed | | 417.00 | 897.00 | 840.00 | 786.00 | 713.00 | 590.00 | 614.00 | 554.00 | 457.00 | 530.00 | 446.00 | 298.00 | 148.00 | 111.00 | 7401.00 |
| POPULATION = expected | | 741.55 | 743.48 | 752.17 | 715.48 | 682.65 | 667.20 | 590.92 | 500.16 | 454.78 | 411.33 | 359.19 | 292.56 | 214.35 | 275.18 | 7401.00 |
| | Grand Total | 1158.55 | 1640.48 | 1592.17 | 1501.48 | 1395.65 | 1257.20 | 1204.92 | 1054.16 | 911.78 | 941.33 | 805.19 | 590.56 | 362.35 | 386.18 | 14802.00 |

| | |
|------------------------|----------|
| p-value for Chi-square | 1.58E-73 |
|------------------------|----------|

Image 12. Chi-Square results and p- value of Samples vs. Population.

As it seen, the p-value is much smaller than the pre-selected alpha = 5%, then we reject the Null Hypothesis, and conclude that the Sample CANNOT be considered representative of the Argentine population.

Thus, we cannot use the results and conclusions obtained in this work to generalize about the population. We would need a larger number of samples and to represent the age spectrum analogously.

YEAR_OF_BIRTH

Year in which the interviewee was born.

A. Cleaning and Coding: No action taken. No missing and no erroneous data.

B. Examination: The years in which the interviewees were born cover almost the entire 20th century, this is interesting as it means that the surveys cover a complete perspective of it. $|\text{Skewness}| \leq 2$ and $|\text{Kurtosis}| \leq 2$ so we can assume normal distribution. Just one value was found slightly beyond 1.5*IQR rule (year 1900), but it's not considered an outlier. More statistical data can be found in the table below.

C. Visualization:

| YEAR_OF_BIRTH | |
|-------------------------|----------|
| Mean | 1957.637 |
| Standard Error | 0.233349 |
| Median | 1959 |
| Mode | 1952 |
| Standard Deviation | 20.07477 |
| Sample Variance | 402.9964 |
| Kurtosis | -0.60994 |
| Skewness | -0.18066 |
| Range | 99 |
| Minimum | 1900 |
| Maximum | 1999 |
| Sum | 14488471 |
| Count | 7401 |
| Confidence Level(95.0%) | 0.45743 |

Image 13. Descriptive statistics for YEAR_OF_BIRTH.

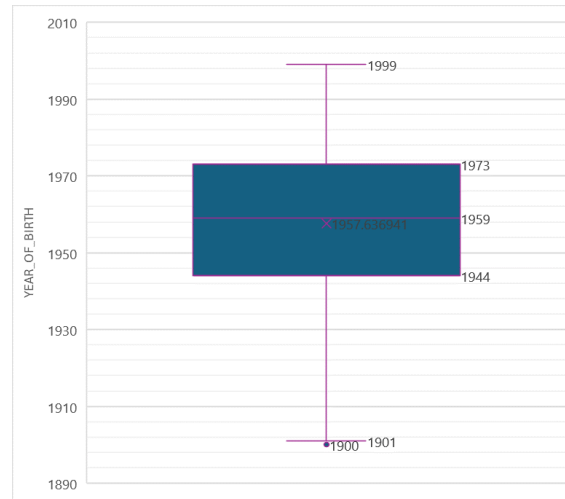


Image 14. Boxplot for YEAR_OF_BIRTH.

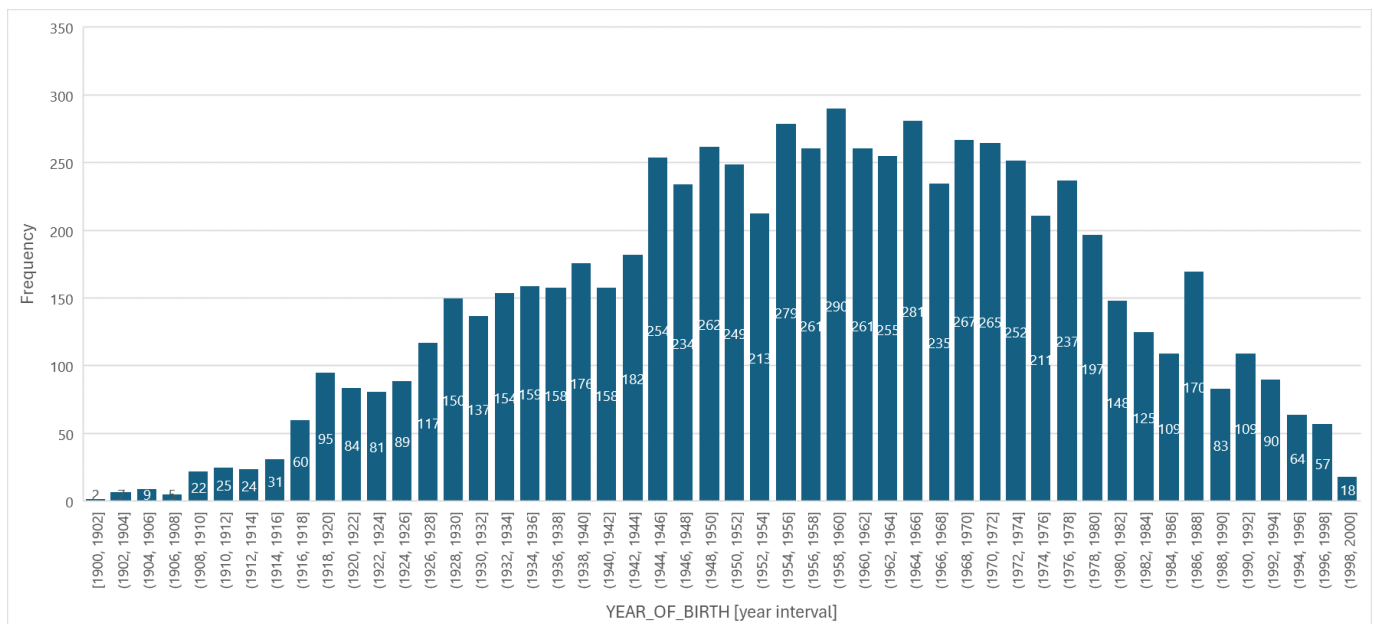


Image 15. Histogram for YEAR_OF_BIRTH.

SEX

Sex of the interviewee.

- A. Cleaning and Coding: 1: MALE, 2: FEMALE. No missing and no erroneous data.
- B. Examination: While there is an equitable distribution of male/female, it would be interesting to see later if this distribution accompanies the entire age range. There are in fact 377 more women across the surveys. It seems to be quite close to the numbers of the Argentine population, as *Reference 4* from the latest census in July 2022 estimates 51.3% women and 48.7% men.
- C. Visualization:

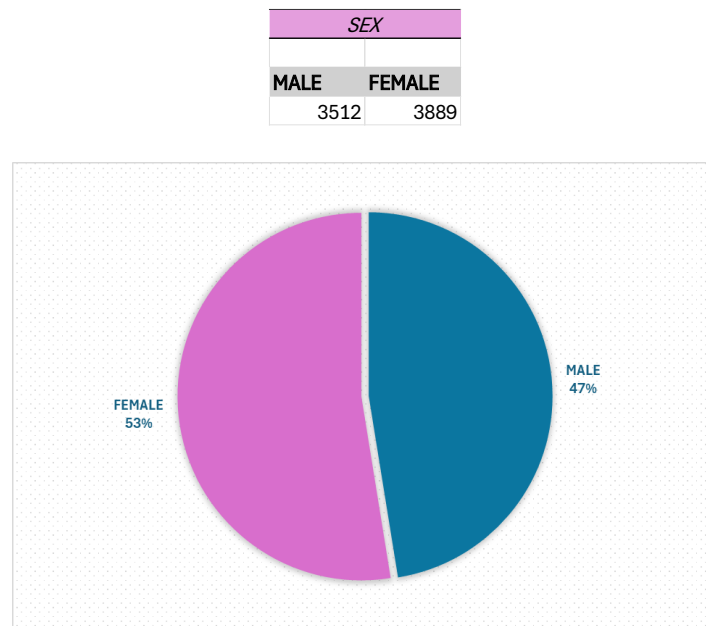


Image 16. Summary table and pie chart for SEX.

ETHNIC_GROUP

Ethnic group of the interviewee.

- A. Cleaning and Coding: Missing values and inconsistencies in the surveys' codification for this field were found. There is no available data for WAVES 1 to 5, only for WAVES 6 and 7. In addition, codification between them is different.
- B. Examination: Only results for 2005 (1-9 responses) out of 7401 (total 7 waves) interviewees. They belong to WAVES 6 and 7, but even between the two surveys, there is only a coding match for the number 1: White. All other values cannot be taken as valid. What I can say with certainty, through a basic percentage calculation is that the white population accounts for 81% of the 2005 interviewees. Nevertheless, this variable is initially NOT relevant in contributing to the answers of the Research Questions.
- C. Visualization:

| | | | | | |
|------------------|------|---|--|---|--|
| | | NOTE: Data available only for WAVES 7 and 6. No data for the WAVES 1 to 4, and 5 surveys. | | FOUND INCONSISTENCIES BETWEEN WAVES | |
| ETHNIC GROUP | | WAVE 7 Spanish | | WAVE 5 | |
| | | Q290. Grupo étnico (Codificar por observación) | | V256. (CODIFICAR GRUPO ÉTNICO POR OBSERVACION) | |
| -4 No Data | 5396 | 1 Blanco | | 1 <input type="checkbox"/> Blanco Caucásico | |
| -2 Don't Respond | 22 | 2 Moreno claro | | 2 <input type="checkbox"/> Raza Negra | |
| 1 White | 1616 | 3 Moreno oscuro | | 3 <input type="checkbox"/> Indio Sudasiático, Paquistaní, etc. | |
| 2 XXXX | 228 | 4 Negro | | 4 <input type="checkbox"/> Chino del Este asiático, Japonés, etc. | |
| 3 XXXX | 66 | 5 Indígena | | 5 <input type="checkbox"/> Árabe, del Asia Central | |
| 4 XXXX | 1 | 6 Otro (anotar): _____ | | 6 <input type="checkbox"/> Otro, especifique | |
| 5 XXXX | 2 | [Cambiar la lista según grupos étnicos en Argentina] | | | |
| 6 Other | 3 | WAVE 7 English | | | |
| 9 XXXX | 67 | Q290. Respondent's ethnic group: | | | |
| SUM | 2005 | Modify the list for your country/ society. | | | |
| % WHITE | 81% | 1 White | | | |
| | | 2 Black | | | |
| | | 3 South Asian Indian, Pakistani, etc. | | | |
| | | 4 East Asian Chinese, Japanese, etc. | | | |
| | | 5 Arabic, Central Asian | | | |
| | | 6 Other (write in): _____ | | | |

Image 17. Summary table for ethnic group and inconsistencies found in the questions across the surveys 7 and 5.

LEVEL_EDU_COUNTRY

Education (country specific)

A. Cleaning and Coding: Missing values and inconsistencies in the surveys' codification for this field were found. There is no available data for the below-mentioned WAVES, only for WAVES 3,4 and 7. In addition, codification among them is different. For further details about the equivalencies analyzed, kindly check the excel file.

B. Examination: As seen below, there is 55% of missing data. This variable is discarded, and the action contributes to a change in the Supporting Question “e”.

C. Visualization:

| | | | | | |
|------------------------------|------|--|--|--|--|
| EDUCATION (COUNTRY SPECIFIC) | | | | | |
| LEVEL_EDU_COUNTRY | | | | | |
| -4 | 4039 | | | | |
| 1 | 31 | | | | |
| 2 | 248 | | | | |
| 3 | 787 | | | | |
| 4 | 664 | | | | |
| 5 | 753 | | | | |
| 6 | 133 | | | | |
| 7 | 338 | | | | |
| 8 | 210 | | | | |
| 9 | 181 | | | | |
| 10 | 17 | | | | |

Waves without data: 1, 2, 5, 6

Waves with data: 3, 4, 7

Image 18. Summary table for LEVEL_EDU_COUNTRY and inconsistencies found across the surveys.

MEASURES OF HAPPINESS AND SATISFACTION

FEEL_OF_HAPP

Feeling of happiness of the interviewee.

- A. Cleaning and Coding: The TOTAL is not considering those who ‘Don’t Respond’ or ‘Don’t Know’ for this variable in the table. I have categorized the rest of the individuals between ‘HAPPY’ and ‘UNHAPPY’. No missing and no erroneous data.
- B. Examination: As seen below, 83% of the interviewees have a feeling of happiness considering all aspects of their lives.
- C. Visualization:

| FEELING OF HAPINESS | | | | |
|---------------------|------|--|-------|---------|
| | | | HAPPY | UNHAPPY |
| -2 Don't Respond | 39 | | | |
| -1 Don't Know | 64 | | | |
| 1 Very Happy | 2168 | | | |
| 2 Rather happy | 3914 | | | |
| 3 Not very happy | 1069 | | | |
| 4 Not at all happy | 147 | | | |
| TOTALS | 7298 | | 6082 | 1216 |

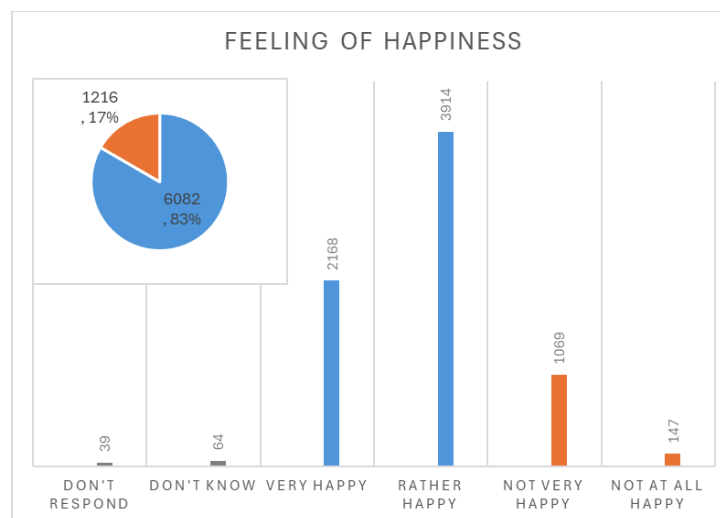


Image 19. Summary table, pie chart and bar chart for FEEL_OF_HAPP.

SAT_W_LIFE

Satisfaction with your life.

- A. Cleaning and Coding: The TOTAL is not considering those who ‘Don’t Respond’ or ‘Don’t Know’ for this variable in the table. I have categorized the rest of the individuals between ‘SATISFIED’ and ‘DISSATISFIED’. The ‘Neutral’ ones are divided in half, with each half going to each part. No missing and no erroneous data.
- B. Examination: As seen below, 86% of the interviewees are at the higher levels of Satisfaction with their lives.
- The scale variable is converted into a Likert scale variable, from [1 - 10] to [-4 - 5], for the purposes of applying descriptive statistics. $|\text{Skewness}| \leq 2$ and $|\text{Kurtosis}| \leq 2$ so we can assume normal distribution. More statistical data can be found in the table below.

C. Visualization:

| SATISFACTION WITH LIFE [re-coded] | | | |
|-----------------------------------|------|---------|------|
| SAT_W_LIFE | | | |
| -2 Don't Respond | 41 | | |
| -1 Don't Know | 32 | DISSAT. | SAT. |
| 1 Completely dissatisfied | 147 | | |
| 2 Level -3 | 74 | | |
| 3 Level -2 | 144 | | |
| 4 Level -1 | 255 | 1003 | |
| 5 Neutral | 765 | | |
| 6 Level 1 | 732 | | 6325 |
| 7 Level 2 | 1396 | | |
| 8 Level 3 | 1779 | | |
| 9 Level 4 | 818 | | |
| 10 Completely satisfied | 1218 | | |
| TOTAL | 7328 | 1003 | 6325 |

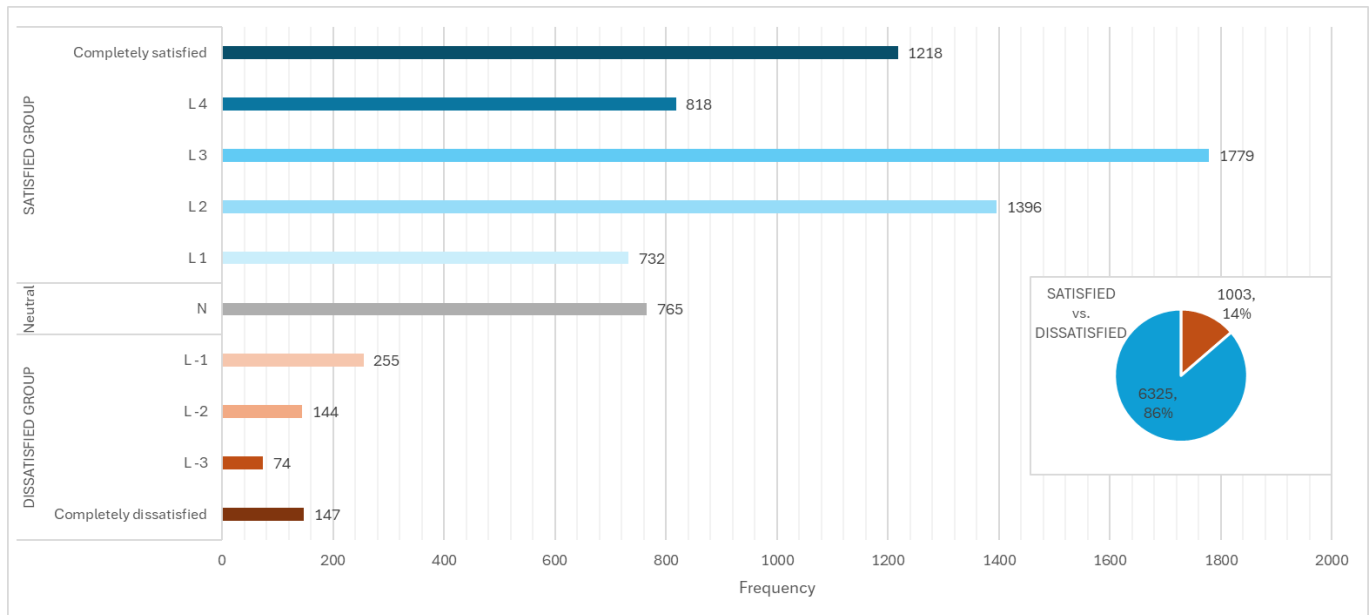


Image 20. Summary table, pie chart and bar chart for SAT_W_LIFE.

| DISSATISFIED GROUP | | | | Neutral | SATISFIED GROUP | | | | |
|--------------------------------------|------|------|------|---------|-----------------|------|------|-----|----------------------|
| Completely dissatisfied | L -3 | L -2 | L -1 | N | L 1 | L 2 | L 3 | L 4 | Completely satisfied |
| New encoding for statistics purposes | | | | | | | | | |
| -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| 147 | 74 | 144 | 255 | 765 | 732 | 1396 | 1779 | 818 | 1218 |

| SAT_W_LIFE | descriptive stats. |
|-------------------------|--------------------|
| Mean | 2.302128821 |
| Standard Error | 0.023980726 |
| Median | 3 |
| Mode | 3 |
| Standard Deviation | 2.052839821 |
| Sample Variance | 4.214151329 |
| Kurtosis | 0.578081447 |
| Skewness | -0.803786436 |
| Range | 9 |
| Minimum | -4 |
| Maximum | 5 |
| Sum | 16870 |
| Count | 7328 |
| Confidence Level(95.0%) | 0.047009125 |

Image 21. Re-coding table and descriptive statistics for SAT_W_LIFE.

FREE_CHO_CON

How much freedom of choice and control. About their actions affecting their lives.

- A. Cleaning and Coding: The TOTAL is not considering those who ‘Don’t Respond’ or ‘Don’t Know’ for this variable in the table. I have categorized the rest of the individuals between ‘WHITHOUT CHOICE’ and ‘CHOICE’. The ‘Neutral’ ones are divided in half, with each half going to each part. No missing and no erroneous data.
- B. Examination: As seen below, 86% of the interviewees consider that they have a high degree of control and choice over what happens in their lives. The scale variable is converted into a Likert scale variable, from [1 - 10] to [-4 - 5], for the purposes of applying descriptive statistics. $|\text{Skewness}| \leq 2$ and $|\text{Kurtosis}| \leq 2$ so we can assume normal distribution. More statistical data can be found in the table below.
- C. Visualization:

| FREEDOM OF CHOICE AND CONTROL [re-coded] | | | |
|--|------|------------|--------|
| FREE_CHO_CON | | | |
| -2 Don't Respond | 143 | | |
| -1 Don't Know | 197 | W/O CHOICE | CHOICE |
| 1 No choice at all | 122 | | |
| 2 Grade -3 | 51 | | |
| 3 Grade -2 | 144 | | |
| 4 Grade -1 | 256 | 1015 | |
| 5 Neutral | 884 | | |
| 6 Grade 1 | 698 | | 6046 |
| 7 Grade 2 | 1180 | | |
| 8 Grade 3 | 1418 | | |
| 9 Grade 4 | 717 | | |
| 10 A great deal of choice | 1591 | | |
| TOTAL | 7061 | 1015 | 6046 |

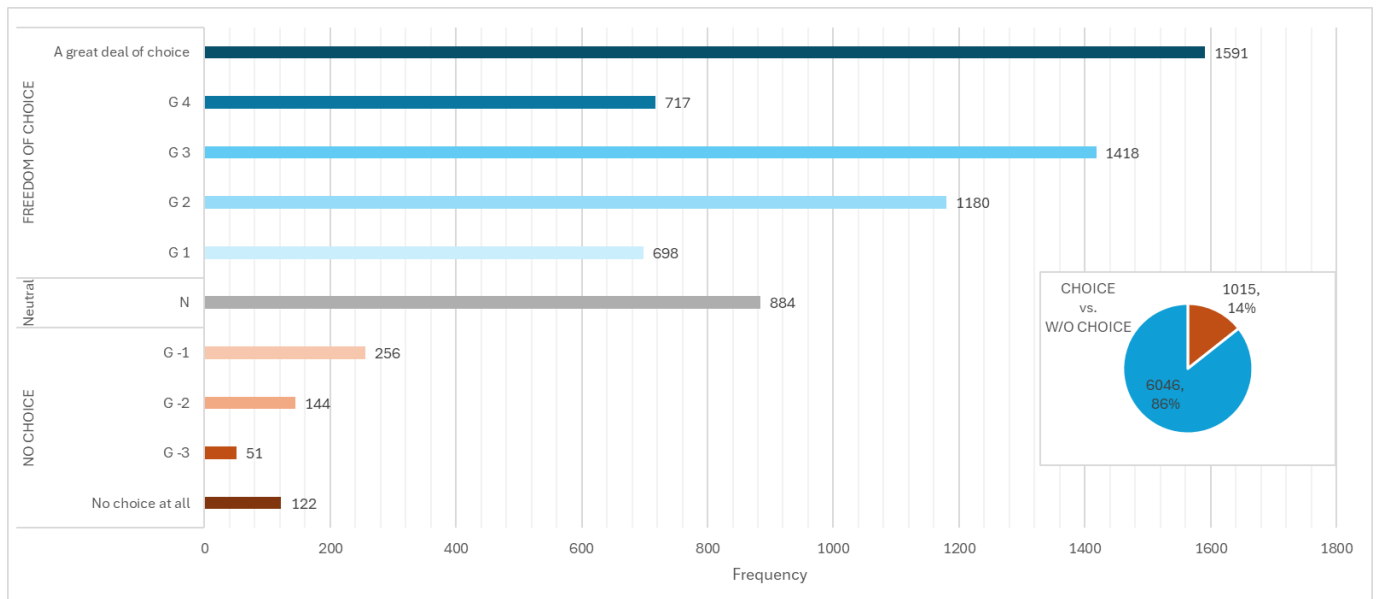


Image 22. Summary table, pie chart and bar chart for FREE_CHO_CON.

| NO CHOICE | | | | Neutral | FREEDOM OF CHOICE | | | | |
|--------------------------------------|------|------|------|---------|-------------------|------|------|-----|-----------|
| Great Deal | G -3 | G -2 | G -1 | N | G 1 | G 2 | G 3 | G 4 | No Choice |
| New encoding for statistics purposes | | | | | | | | | |
| -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| 122 | 51 | 144 | 256 | 884 | 698 | 1180 | 1418 | 717 | 1591 |

| FREE_CHO_CON | descriptive stats. |
|-------------------------|--------------------|
| Mean | 2.400509843 |
| Standard Error | 0.025328706 |
| Median | 3 |
| Mode | 5 |
| Standard Deviation | 2.128364963 |
| Sample Variance | 4.529937417 |
| Kurtosis | 0.079551572 |
| Skewness | -0.667083509 |
| Range | 9 |
| Minimum | -4 |
| Maximum | 5 |
| Sum | 16950 |
| Count | 7061 |
| Confidence Level(95.0%) | 0.049651863 |

Image 23. Re-coding table and descriptive statistics for FREE_CHO_CON.

MARITAL_STA

Marital Status

- A. Cleaning and Coding: The TOTAL does not include those who ‘Don’t Respond’ or ‘Don’t Know’ for this variable in the table. I have categorized the rest of the individuals between ‘ACCOMPANIED’ and ‘ALONE’.
- B. Examination: As seen below, around 56% of the interviewees share their lives with a partner, and 44% are probably in a lonely situation.
- C. Visualization:

| MARITAL STATUS | | | | |
|------------------------------|------|--|-------------|-------|
| MARITAL_STA | | | ACCOMPANIED | ALONE |
| -2 Don't Respond | 0 | | | |
| -1 Don't Know | 3 | | | |
| 1 Married | 3450 | | 4175 | |
| 2 Living together as married | 725 | | | |
| 3 Divorced | 191 | | | 3223 |
| 4 Separated | 405 | | | |
| 5 Widowed | 633 | | | |
| 6 Single | 1994 | | | |
| TOTAL | 7398 | | 4175 | 3223 |

Image 24. Summary table for MARITAL_STA.

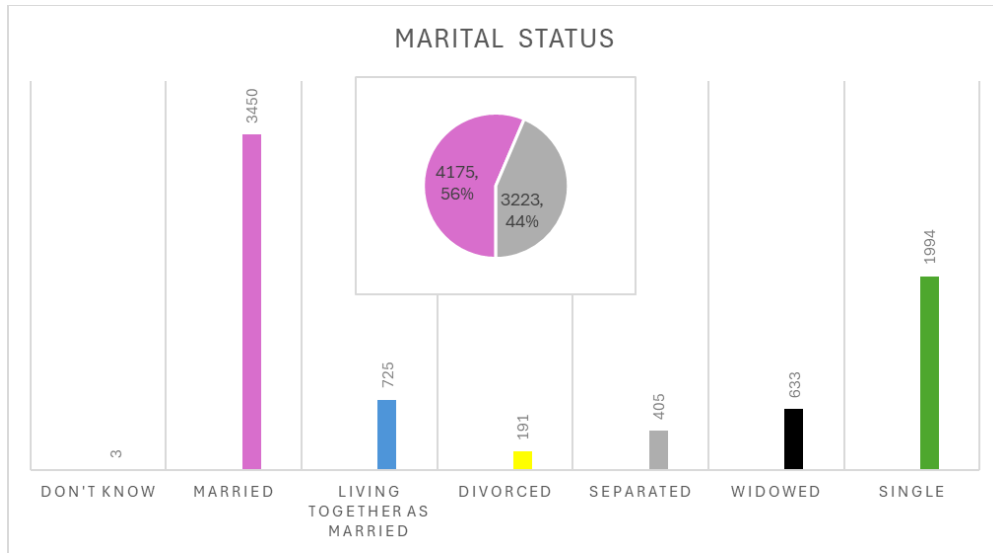


Image 25. Pie chart and bar chart for MARITAL_STA.

HOW_MANY_CHILD

How many children do you have

- A. Cleaning and Coding: 92 entries of missing data (code -5), they are not considered in the calculations or in the percentages.
- B. Examination: As seen below, around 69% of the interviewees have at least 1 child, while 31% claim they have no children. $|\text{Skewness}| \leq 2$ and $|\text{Kurtosis}| \leq 2$ so we can assume normal distribution. No possible outliers. It is observed that the average number of children is 1.705 for the 1984 -> 2017 WVS Surveys. More statistical data can be found in the table below.
- C. Visualization:

| HOW MANY CHILDREN | | | |
|-------------------|------|-------|-----------|
| HOW_MANY_CHILD | | | |
| -5 | 92 | CHILD | W/O CHILD |
| -2 | 26 | | |
| 0 | 2254 | | 2254 |
| 1 | 1196 | 5029 | |
| 2 | 1732 | | |
| 3 | 1126 | | |
| 4 | 494 | | |
| 5 | 481 | | |
| TOTAL | 7283 | 5029 | 2254 |

| | |
|-------------------------|----------|
| Mean | 1.705204 |
| Standard Error | 0.017909 |
| Median | 2 |
| Mode | 0 |
| Standard Deviation | 1.528367 |
| Sample Variance | 2.335907 |
| Kurtosis | -0.6609 |
| Skewness | 0.54124 |
| Range | 5 |
| Minimum | 0 |
| Maximum | 5 |
| Sum | 12419 |
| Count | 7283 |
| Confidence Level(95.0%) | 0.035107 |

Image 26. Summary table and descriptive statistics HOW_MANY_CHILD.

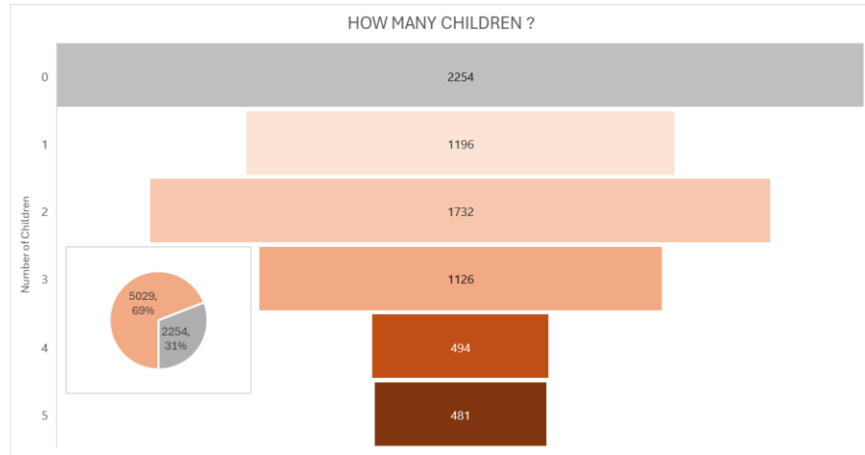


Image 27. Funnel chart for HOW_MANY_CHILD.

IP_LIFE_LEI

Important in life: Leisure time.

- A. Cleaning and Coding: There is no data for WAVE 1 (Year 1984). The TOTAL does not include those who 'Don't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'IMPORTANT' and 'NOT IMPORTANT'.
- B. Examination: As seen below, 79% of the interviewees consider that leisure time is important in their lives.
- C. Visualization:

| IMPORTANT IN LIFE: LEISURE | | | |
|----------------------------|------|-----------|---------------|
| IP_LIFE_LEI | | | |
| -4 No Data | 1005 | IMPORTANT | NOT IMPORTANT |
| -2 Don't Respond | 15 | | |
| -1 Don't Know | 52 | | |
| 1 Very Important | 2280 | 5020 | |
| 2 Rather Important | 2740 | | |
| 3 Not very important | 1083 | | 1309 |
| 4 Not at all important | 226 | | |
| TOTAL | 6329 | 5020 | 1309 |

Image 28. Summary table for IP_LIFE_LEI.

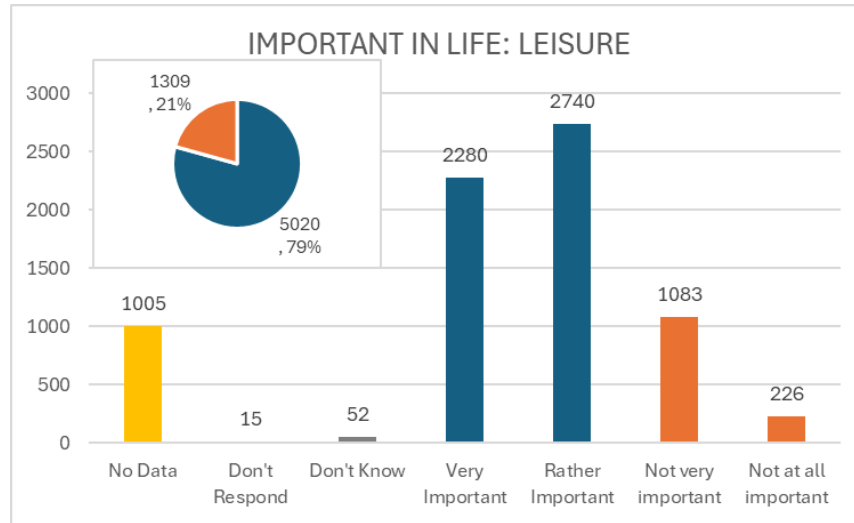


Image 29. Pie chart and bar chart for IP_LIFE_LEI.

IP_LIFE_FAM

Important in life: Family.

D. Cleaning and Coding: There is no data for WAVE 1 (Year 1984). The TOTAL does not include those who 'Don't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'IMPORTANT' and 'NOT IMPORTANT'.

E. Examination: As seen below, 99% of the interviewees consider that Family is important in their lives.

F. Visualization:

| IMPORTANT IN LIFE: FAMILY | | | | |
|---------------------------|------|-----------|---------------|--|
| IP_LIFE_FAM | | | | |
| -4 No Data | 1005 | IMPORTANT | NOT IMPORTANT | |
| -2 Don't Respond | 4 | | | |
| -1 Don't Know | 2 | | | |
| 1 Very Important | 5731 | 6328 | | |
| 2 Rather Important | 597 | | | |
| 3 Not very important | 46 | | 62 | |
| 4 Not at all important | 16 | | | |
| TOTAL | 6390 | 6328 | 62 | |

Image 30. Summary table for IP_LIFE_FAM.

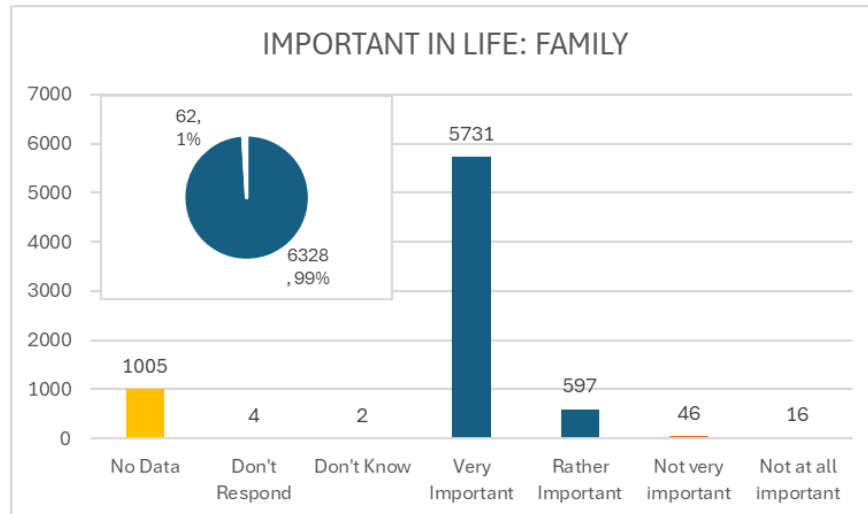


Image 31. Pie chart and bar chart for IP_LIFE_FAM.

IP_LIFE_FRI

Important in life: Friends.

- A. Cleaning and Coding: There is no data for WAVE 1 (Year 1984). The TOTAL does not include those who 'Don't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'IMPORTANT' and 'NOT IMPORTANT'.
- B. Examination: As seen below, 86% of the interviewees consider that their Friends are important in their lives.
- C. Visualization:

| IMPORTANT IN LIFE: FRIENDS | | | |
|----------------------------|------|-----------|---------------|
| IP_LIFE_FRI | | | |
| -4 No Data | 1005 | IMPORTANT | NOT IMPORTANT |
| -2 Don't Respond | 6 | | |
| -1 Don't Know | 25 | | |
| 1 Very Important | 3410 | 5449 | |
| 2 Rather Important | 2039 | | |
| 3 Not very important | 793 | | 916 |
| 4 Not at all important | 123 | | |
| TOTAL | 6365 | 5449 | 916 |

Image 32. Summary table for IP_LIFE_FRI.

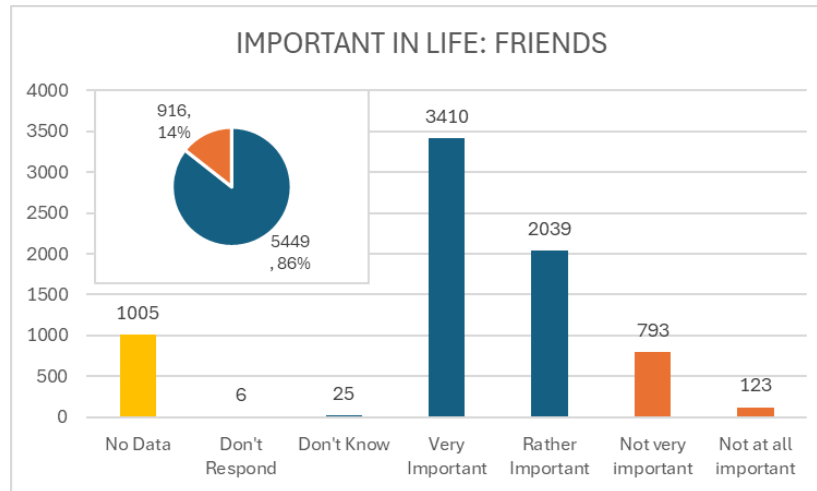


Image 33. Pie chart and bar chart for IP_LIFE_FRI.

REL_GROUPS

Religious denominations - major groups.

- A. Cleaning and Coding: The TOTAL does not include those who 'Didn't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'RELIGIOUS' and 'NOT RELIGIOUS'.
- B. Examination: As seen below, around 83% of the interviewees are religious people, and most of them are Catholics. Also, it is interesting to note that 5% or 354 people refuse to respond. Religious beliefs have been fluctuating quite a bit in Argentina, and according to Reference 5, the Argentine scientific organization CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) publishes that 62.9% identify as Catholics, followed by 18.9% as non-religious, and in third place Evangelicals/ Protestant with 15.3%.
- C. Visualization:

| RELIGIOUS DENOMINATIONS - major groups [re-coded] | | | | | |
|---|----|------|--|---------------|-----------|
| REL_GROUPS | | | | NOT RELIGIOUS | RELIGIOUS |
| Don't Respond | -2 | 354 | | | |
| Don't know | -1 | 5 | | | |
| Don't belong to any | 0 | 1181 | | 1181 | |
| Catholic | 1 | 5511 | | | |
| Protestant | 2 | 154 | | | |
| Othodox | 3 | 28 | | | |
| Jew | 4 | 74 | | | |
| Muslim | 5 | 6 | | | |
| Hindu | 6 | 12 | | | |
| Buddhist | 7 | 76 | | | |
| Other | 8 | 0 | | | |
| TOTAL | | 7042 | | 1181 | 5861 |
| | | | | 16.77% | 83.23% |

Image 34. Summary table for REL_GROUPS.

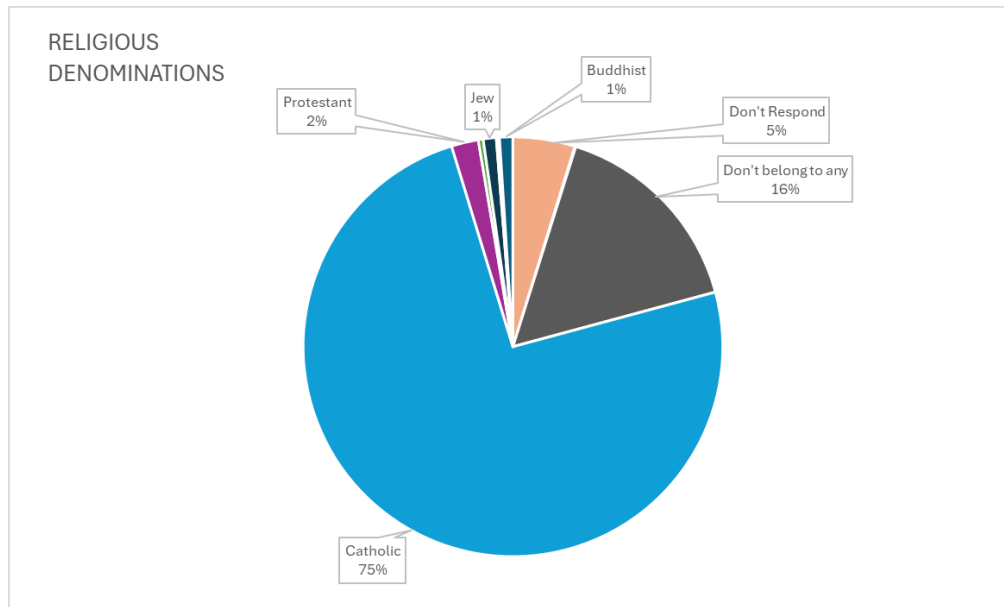


Image 35. Pie chart for REL_GROUPS.

IP_LIFE_REL

Important in life: Religion.

- A. Cleaning and Coding: There is no data for WAVE 1 (Year 1984). The TOTAL does not include those who 'Don't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'IMPORTANT' and 'NOT IMPORTANT'.
- B. Examination: As seen below, 63% of the interviewees consider that religion is important in their lives; while 37% think that is not very important or not important at all.
- C. Visualization:

| IMPORTANT IN LIFE: RELIGION | | | | |
|-----------------------------|------|-----------|---------------|--|
| IP_LIFE_REL | | | | |
| -4 No Data | 1005 | IMPORTANT | NOT IMPORTANT | |
| -2 Don't Respond | 21 | | | |
| -1 Don't Know | 41 | | | |
| 1 Very Important | 2127 | 4013 | | |
| 2 Rather Important | 1886 | | | |
| 3 Not very important | 1471 | | | |
| 4 Not at all important | 850 | | 2321 | |
| TOTAL | | 4013 | 2321 | |

Image 36. Summary table for IP_LIFE_REL.

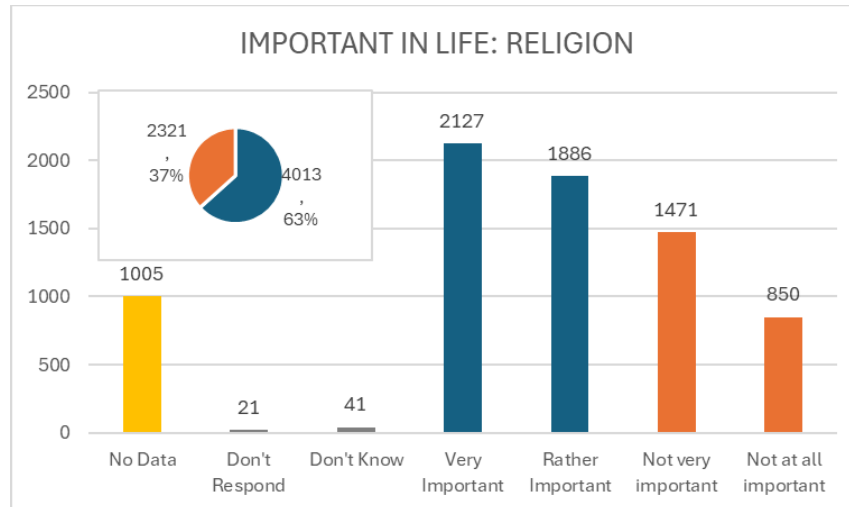


Image 37. Pie chart and bar chart for IP_LIFE_REL.

MEASURES OF WORK & MONEY

INCOME_SCALE

Scale of incomes

- A. Cleaning and Coding: There is no data for WAVE 1 (Year 1984) and for WAVE 5 (Year 2006).
- B. Examination: $|\text{Skewness}| \leq 2$ and $|\text{Kurtosis}| \leq 2$ so we can assume normal distribution. No possible outliers. It is observed that almost half of the total sample falls within the middle-income groups: 4, 5, and 6. The richest groups, 9 and 10, represent 8%, while the poorest groups, 1 and 2, represent 18% of the sample. More statistical data can be found in the table below.
- C. Visualization:

| SCALE OF INCOME | | |
|------------------|--|------|
| INCOME_SCALE | | |
| -4 No Data | | 2007 |
| -2 Don't Respond | | 329 |
| -1 Don't Know | | 213 |
| 1 Lowest group | | 376 |
| 2 Group 2 | | 451 |
| 3 Group 3 | | 560 |
| 4 Group 4 | | 714 |
| 5 Group 5 | | 896 |
| 6 Group 6 | | 681 |
| 7 Group 7 | | 481 |
| 8 Group 8 | | 306 |
| 9 Group 9 | | 189 |
| 10 Highest group | | 198 |

| | |
|-------------------------|----------|
| Mean | 4.920857 |
| Standard Error | 0.033447 |
| Median | 5 |
| Mode | 5 |
| Standard Deviation | 2.329824 |
| Sample Variance | 5.428079 |
| Kurtosis | -0.54066 |
| Skewness | 0.235741 |
| Range | 9 |
| Minimum | 1 |
| Maximum | 10 |
| Sum | 23876 |
| Count | 4852 |
| Confidence Level(95.0%) | 0.065572 |

Image 38. Summary table and descriptive statistics table for INCOME_SCALE.

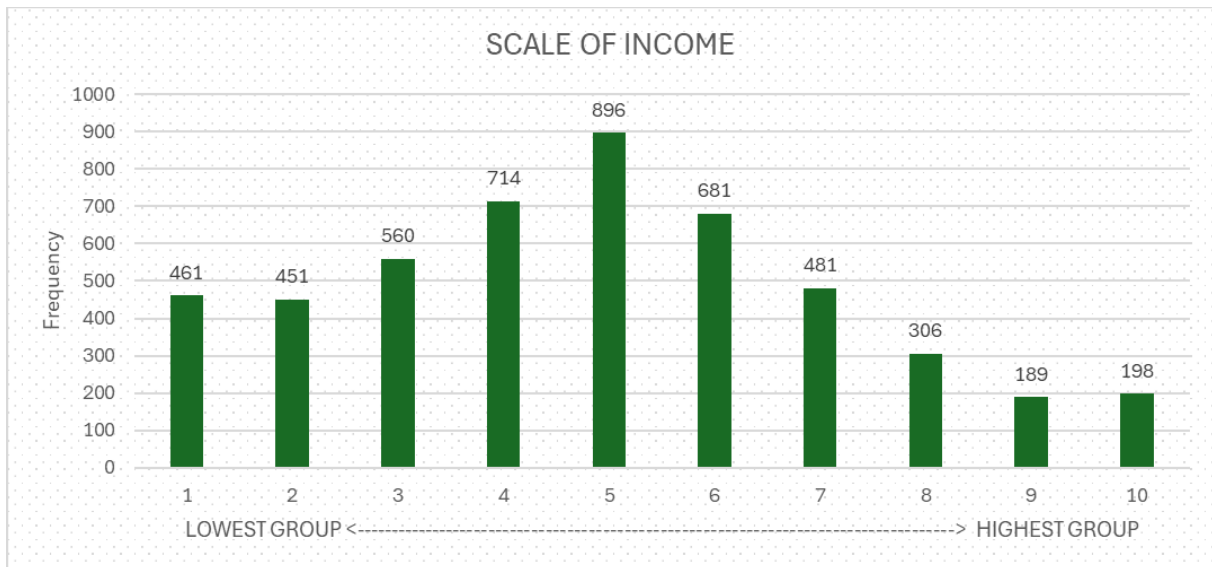


Image 39. Bar chart for INCOME_SCALE.

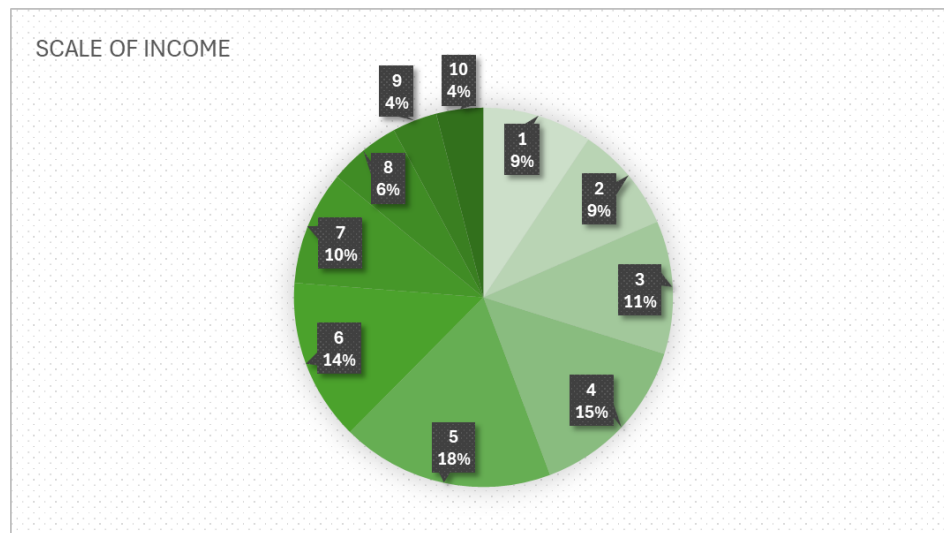


Image 40. Pie chart for INCOME_SCALE.

IP_LIFE_WORK

Important in life: Work.

- A. Cleaning and Coding: There is no data for WAVE 1 (Year 1984). The TOTAL does not include those who 'Don't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'IMPORTANT' and 'NOT IMPORTANT'.
- B. Examination: As seen below, 95% of the interviewees consider that their Work is important in their lives; while just the 5% think that is not very important or not important at all.
- C. Visualization:

| IMPORTANT IN LIFE: WORK | | | | |
|-------------------------|--|------|-----------|---------------|
| IP_LIFE_WORK | | | | |
| -4 No Data | | 1005 | IMPORTANT | NOT IMPORTANT |
| -2 Don't Respond | | 17 | | |
| -1 Don't Know | | 25 | | |
| 1 Very Important | | 4347 | 6014 | |
| 2 Rather Important | | 1667 | | |
| 3 Not very important | | 228 | | |
| 4 Not at all important | | 112 | | 340 |
| TOTAL | | 6354 | 6014 | 340 |

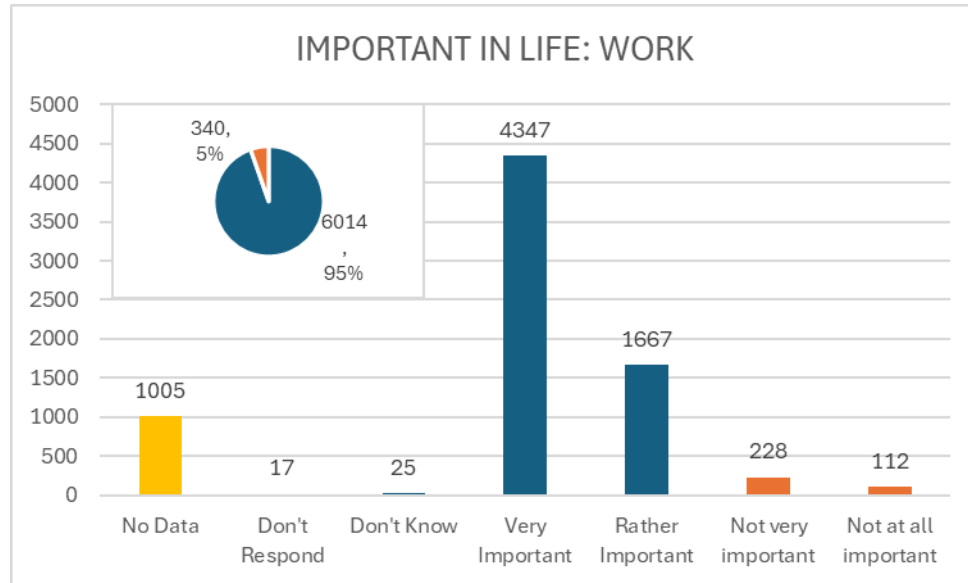


Image 41. Summary table, pie chart and bar chart for IP_LIFE_WOR.

POLITICAL CULTURE & TRUST

IP_LIFE_POL

Important in life: Politics.

- Cleaning and Coding: There is no data for WAVE 1 (Year 1984). The TOTAL does not include those who 'Don't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'IMPORTANT' and 'NOT IMPORTANT'.
- Examination: As seen below, 31% of the interviewees consider that their Politics is important in their lives; while 69% think that is not very important or not important at all.
- Visualization:

| IMPORTANT IN LIFE: POLITICS | | | | | |
|-----------------------------|------|-----------|---------------|--|--|
| IP_LIFE_POL | | | | | |
| -4 No Data | 1005 | IMPORTANT | NOT IMPORTANT | | |
| -2 Don't Respond | 13 | | | | |
| -1 Don't Know | 35 | | | | |
| 1 Very Important | 658 | 1945 | | | |
| 2 Rather Important | 1287 | | | | |
| 3 Not very important | 2098 | | 4403 | | |
| 4 Not at all important | 2305 | | | | |
| TOTAL | 6348 | 1945 | 4403 | | |

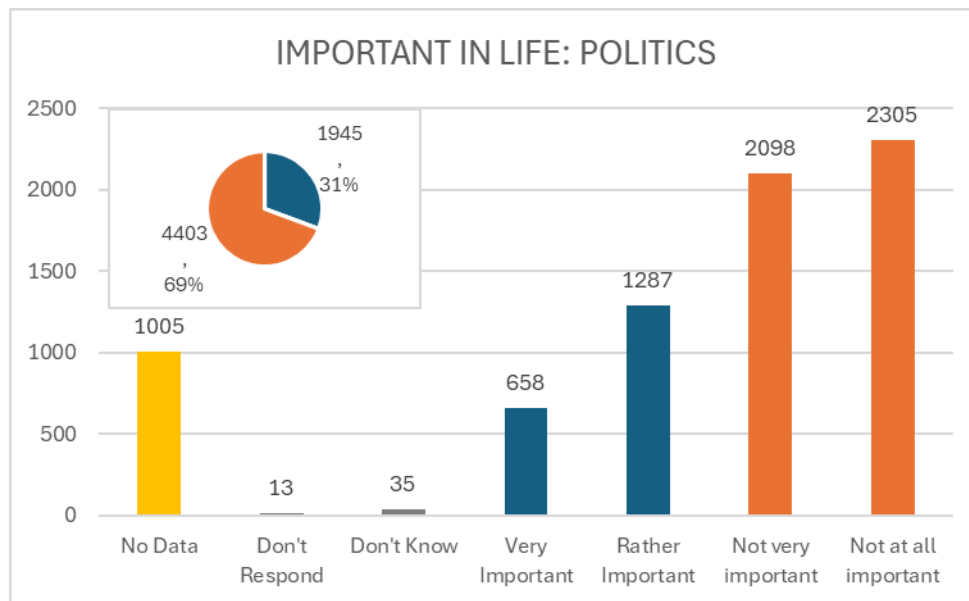


Image 42. Summary table, pie chart and bar chart for IP_LIFE_POL.

POL_INT

Interest in Politics.

- Cleaning and Coding: The TOTAL does not include those who 'Don't Respond' or 'Don't Know' for this variable in the table. I have categorized the rest of the individuals between 'INTERESTED' and 'NOT INTERESTED'.
- Examination: As seen below, 30% of the interviewees are interested in Politics; while 70% are not very interested or not interested at all.
- Visualization:

| INTEREST IN POLITICS | | | | |
|-------------------------|------|--|------------|----------------|
| POL_INT | | | INTERESTED | NOT INTERESTED |
| -2 Don't Respond | 32 | | | |
| -1 Don't Know | 30 | | | |
| 1 Very interested | 570 | | 2206 | |
| 2 Somewhat interested | 1636 | | | |
| 3 Not very interested | 2362 | | | 5133 |
| 4 Not at all interested | 2771 | | | |
| TOTAL | 7339 | | 2206 | 5133 |

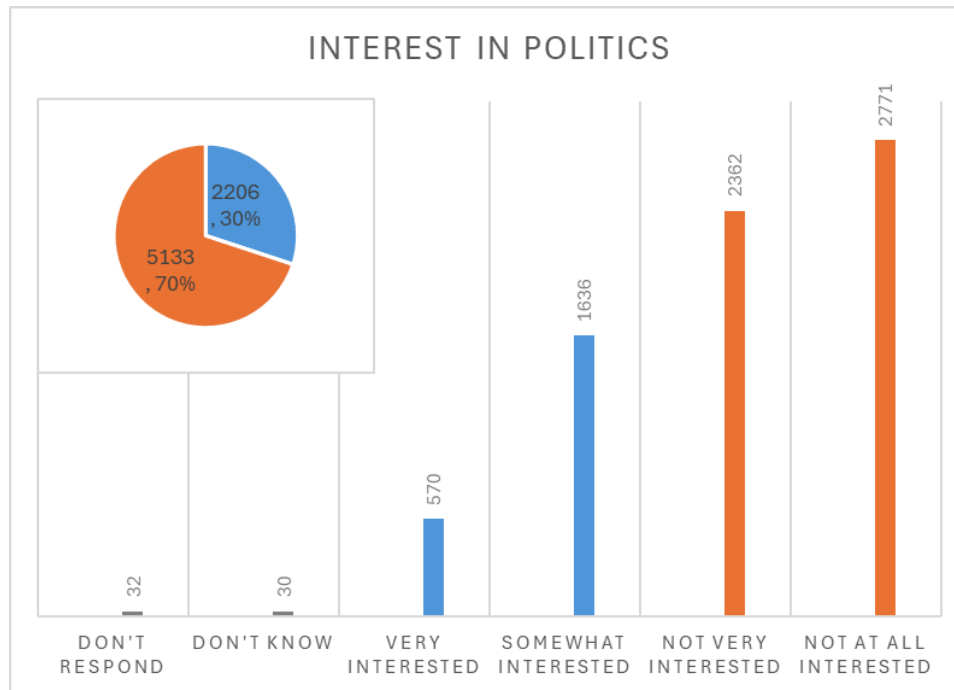


Image 43. Summary table, pie chart and bar chart for POL_INT.

POL_JOI_BOY

Political action: joining in boycotts

POL_UNOFF_STR

Political action: joining unofficial strikes

POL_ATT_DEM

Political action: attending lawful/peaceful demonstrations

- A. Cleaning and Coding: There is no data for WAVE 5 (Year 2006).
- B. Examination: Most people state that they would never join any type of political demonstrations. However, if we look specifically at the category of "Peaceful Demonstrations," the situation is different from the others: 50% say they "Would Never Do", while the remaining 50% might participate or have already participated in such demonstrations.
- C. Visualization:

| JOINING IN BOYCOTTS | | JOINING UNOFFICIAL STRIKES | | ATT. LAWFUL/ PEACEFUL DEMONSTRATIONS | |
|---------------------|------|----------------------------|--|--------------------------------------|--|
| POL_JOI_BOY | | POL_UNOFF_STR | | POL_ATT_DEM | |
| -4 No Data | 0 | 1002 | | 0 | |
| -2 Don't Respond | 221 | 231 | | 195 | |
| -1 Don't Know | 435 | 243 | | 253 | |
| 1 Have done | 184 | 558 | | 1178 | |
| 2 Might do | 970 | 1214 | | 2244 | |
| 3 Would never do | 5591 | 4153 | | 3531 | |

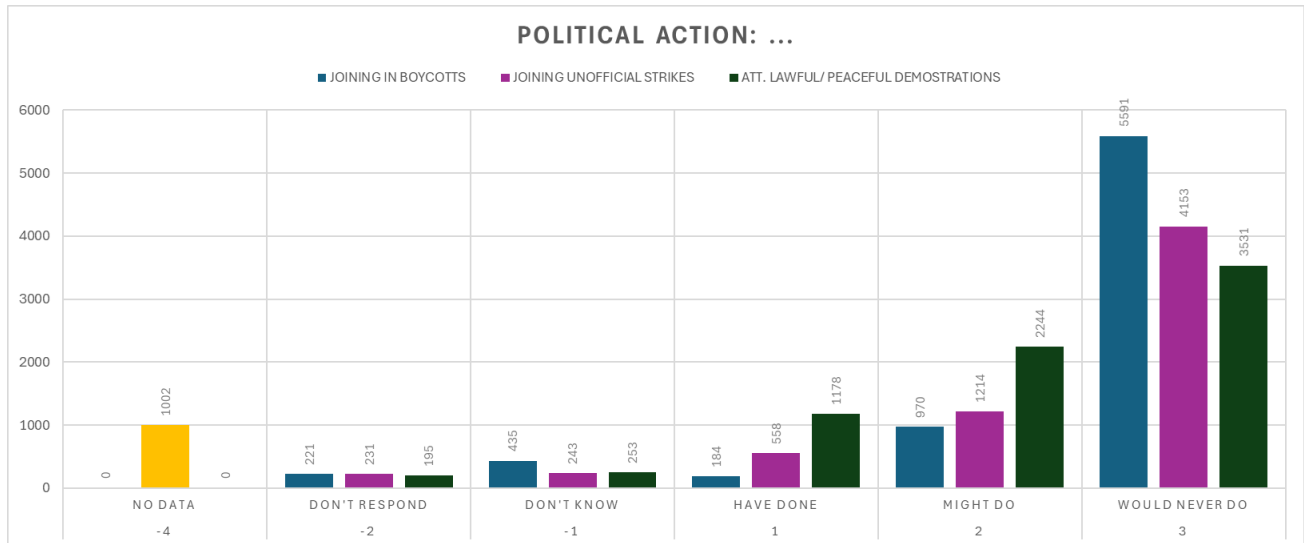


Image 44. Summary table and bar chart for joining political demonstrations.

PROUD_OF_NAT

Proud of Nationality.

- A. Cleaning and Coding: The TOTAL does not include those who 'Don't Respond', those who 'Don't Know', or the "Not Applicable" ones for this variable in the table. I have categorized the rest of the individuals between 'PROUD' and 'NOT PROUD'.
- B. Examination: As seen below, 89% of the interviewees are proud of their nationality, while 11% are not very proud or not proud at all.
- C. Visualization:

| PROUD OF NATIONALITY | | | |
|----------------------|------|-------|-----------|
| PROUD_OF_NAT | | | |
| -3 Not Applicable | 161 | | |
| -2 Don't Respond | 78 | PROUD | NOT PROUD |
| -1 Don't Know | 97 | | |
| 1 Very proud | 4056 | 6323 | |
| 2 Quite proud | 2267 | | |
| 3 Not very proud | 545 | | 742 |
| 4 Not at all proud | 197 | | |
| TOTAL | 7065 | 6323 | 742 |

Image 45. Summary table for PROUD_OF_NAT.

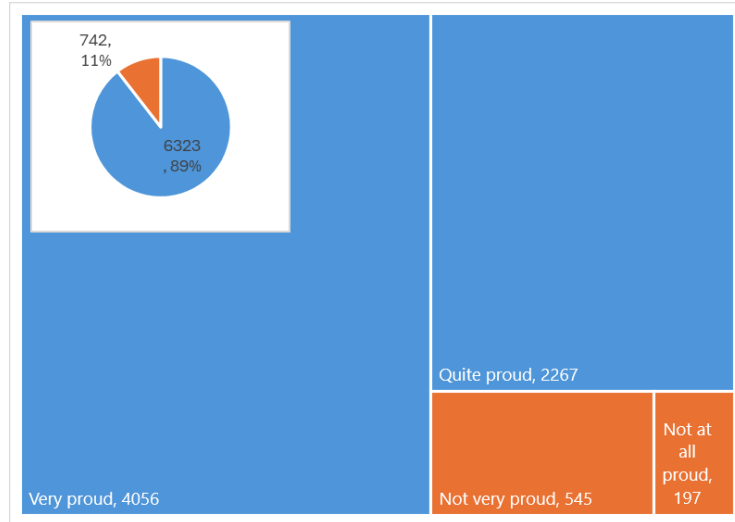


Image 46. Tree map for PROUD_OF_NAT.

GOV_RESP

Responsibility to provide, by the government or by themselves.

A. Cleaning and Coding: There is no data for WAVE 1 (Year 1984).

B. Examination: The scale variable is converted into a Likert scale variable, from [1 - 10] to [-4 - 5], for the purposes of applying descriptive statistics. $|\text{Skewness}| \leq 2$ and $|\text{Kurtosis}| \leq 2$ so we can assume normal distribution. An average of 0.95 is observed, which means there is a balance in terms of the level of responsibility in people's opinions, with a slight tendency towards individual responsibility. More statistical data can be found in the table below.

C. Visualization:

| LEVEL OF GOVERNMENT RESPONSIBILITY ON | | |
|---------------------------------------|--|------|
| GOV_RESP | | |
| -4 No Data | | 1005 |
| -2 Don't Respond | | 70 |
| -1 Don't Know | | 185 |
| 1 GOV. Responsibility | | 754 |
| 2 Level -3 | | 292 |
| 3 Level -2 | | 525 |
| 4 Level -1 | | 447 |
| 5 Neutral | | 579 |
| 6 Level 1 | | 889 |
| 7 Level 2 | | 391 |
| 8 Level 3 | | 630 |
| 9 Level 4 | | 480 |
| 10 INDIVIDUAL Responsibility | | 1154 |

Image 47. Summary table for GOV_RESP.

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| 754 | 292 | 525 | 447 | 579 | 889 | 391 | 630 | 480 | 1154 |

| Descriptive statistics | |
|-------------------------|--------------|
| Mean | 0.95456766 |
| Standard Error | 0.038639926 |
| Median | 1 |
| Mode | 5 |
| Standard Deviation | 3.027999748 |
| Sample Variance | 9.168782472 |
| Kurtosis | -1.188972293 |
| Skewness | -0.183953118 |
| Range | 9 |
| Minimum | -4 |
| Maximum | 5 |
| Sum | 5862 |
| Count | 6141 |
| Confidence Level(95.0%) | 0.075747795 |

Image 48. Re-coding table and descriptive statistics for GOV_RESP.

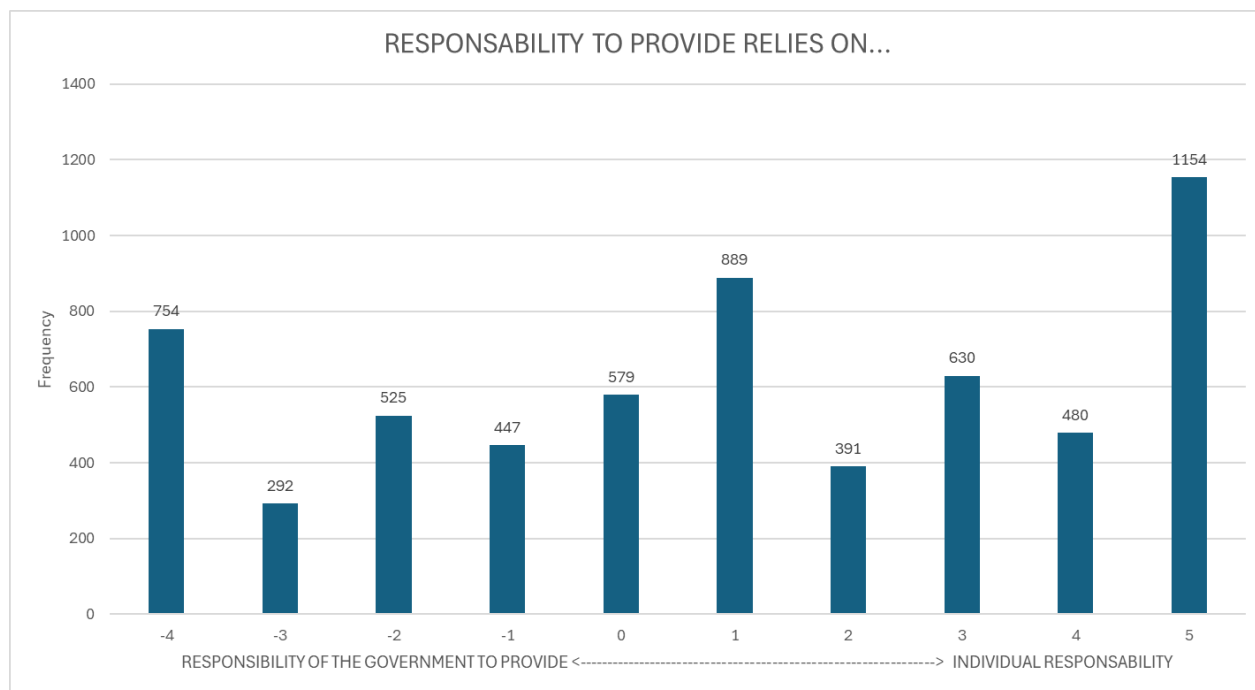


Image 49. Bar chart for GOV_RESP.

MEASURES OF TRUST & ORGANIZATIONAL MEMBERSHIP

MEM_HUM_ORG

Active/Inactive membership of charitable/humanitarian organization

MEM_SP_REC

Active/Inactive membership of sport or recreation

MEM_LAB_UNI

Active/Inactive membership of labour unions

MEM_MUS_EDU

Active/Inactive membership of art, music, educational

MEM_POL_PAR

Active/Inactive membership of political party

MEM_REL_ORG

Active/Inactive membership of church or religious organization

- A. Cleaning and Coding: There is no data for multiple WAVES across the different memberships. Kindly check the Excel document.
- B. Examination: We can highlight that most of the surveyed people are not members of any organization, with only a small minority being members (active or inactive).
- C. Visualization:

| CHARITABLE/ HUMANITARIAN ORG. | | SPORT OR RECREATION | | LABOUR UNIONS | | ART/ MUSICAL/ EDUCATIONAL ORG. | | POLITICAL PARTY | | CHURCH OR RELIGIOUS ORG. | |
|-------------------------------|-------------|---------------------|------------|---------------|-------------|--------------------------------|-------------|-----------------|-------------|--------------------------|-----------------|
| | MEM_HUM_ORG | | MEM_SP_REC | | MEM_LAB_UNI | | MEM_MUS_EDU | | MEM_POL_PAR | | MEM_REL_ORG |
| -4 No Data | 2282 | | 3287 | | 2282 | | 2282 | | | 2282 | No Data |
| -2 Don't Respond | 2 | | 0 | | 0 | | 0 | | | 1 | Don't Respond |
| 0 Don't belong | 4542 | | 3343 | | 4673 | | 4426 | | | 4623 | Don't belong |
| 1 Inactive member | 375 | | 431 | | 271 | | 401 | | | 325 | Inactive member |
| 2 Active member | 200 | | 340 | | 175 | | 292 | | | 170 | Active member |

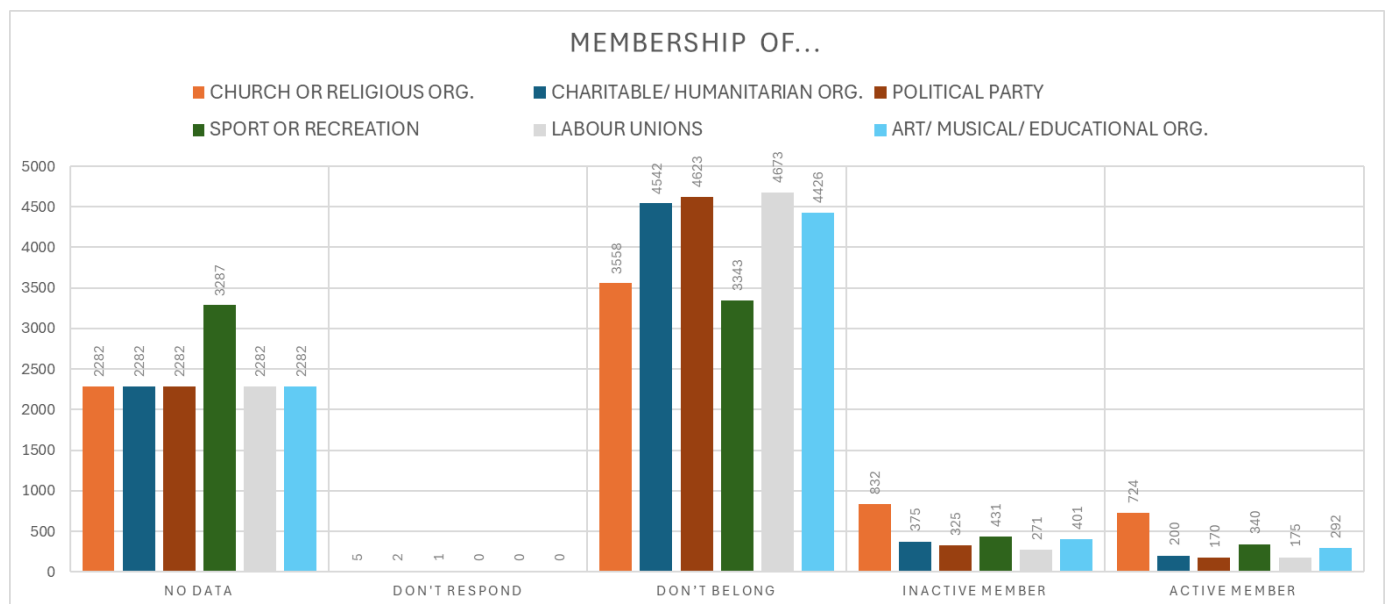


Image 50. Summary table and bar chart for Memberships of different organizations.

PART 3: Hypothesis Testing & Forecasting

The goal in this processing step is to evaluate the data by applying varieties of statistical tests and interpreting the results in order to find out (or not) relationships. That is, to examine if there are statistically significant differences or correlations among groups of variables. These tests may help us determine if there is enough evidence in the sample data to support a certain hypothesis.

By choosing the appropriate variables for each test, it is possible that they will guide us closer to the answers of the questions that motivated this study.

I am going to use statistical methods such as Odds Ratio (OR), Chi-Square, T-test and ANOVA, depending on the case to analyze, the type of variables involved, and the question it seeks to clarify. For each test conducted, I will explain: the Question the test is aimed at and the Results & Interpretation. We have questions related to HAPPINESS AND SATISFACTION IN LIFE aspects and questions related to POLITICS AND SOCIETY aspects, so I will sub-divide this section into these categories.

BUT it is important here to clarify that the final and definitive answers to the Research and the Supporting Questions will not be made explicit until the 'Results & Discussion' section of this paper.

TESTS to analyze HAPPINESS AND SATISFACTION IN LIFE

Question a: How do characteristics related to family, friends, and sense of belonging compare to work and income in terms of importance?

Tests:

- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: FAMILY (IP_LIFE_FAM).
- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: FRIENDS (IP_LIFE_FRI).
- **Chi- Square** of MEMBERSHIPS (MEM_HUM_ORG, MEM_SP_REC, MEM_LAB_UNI MEM_MUS_EDU, MEM_POL_PAR, MEM_REL_ORG) vs. FEELING OF HAPINESS (FEEL_OF_HAPP)

- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: WORK (IP_LIFE_WOR).
- **T-test** of HAPPY? vs. SCALE OF INCOME (INCOME_SCALE)

Results & Interpretation:

- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: FAMILY (IP_LIFE_FAM).

| | | FEEL_OF_HAPP | | |
|---------------------------|---------------|---------------------|------------|-------------|
| | | FEELING OF HAPINESS | | |
| IP_LIFE_FAM | | HAPPY | UNHAPPY | Grand Total |
| IMPORTANT IN LIFE: FAMILY | IMPORTANT | 5273 | 978 | 6251 |
| | NOT IMPORTANT | 35 | 27 | 62 |
| | Grand Total | 5308 | 1005 | 6313 |
| | | O.R. | 4.15924628 | |

| CI of OR | | | | |
|---|-----------|---|--------|--|
| exp(ln(OR)±(Z _{α/2})*SE(ln(OR))) WHERE SE(ln(OR)) is sqrt(1/a+1/b+1/c+1/d)) | | | | |
| ln(OR) = | 1.4253339 | SE(ln(OR)) = | 0.2585 | |
| Z _{α/2} = | 1.96 | for CL 95%, Z _{α/2} is 1.96 SD | | |
| Lower bound: 2.5059806 | | Upper bound: 6.903217732 | | |

Image 51. Table and results for the Odds Ratio between IP_LIFE_FAM and FEEL_OF_HAPP.

Since the C.I. doesn't overlap 1 we may reject the Null Hypothesis, and we can say that there is a possible association, with a C.L.of 95%. The odds are 4,16 times greater that the people who give more importance to family are happier compared to those who consider that family is not important in life.

- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: FRIENDS (IP_LIFE_FRI).

| | | FEEL_OF_HAPP | | |
|----------------------------|---------------|---------------------|------------|-------------|
| | | FEELING OF HAPINESS | | |
| IP_LIFE_FRI | | HAPPY | UNHAPPY | Grand Total |
| IMPORTANT IN LIFE: FRIENDS | IMPORTANT | 4608 | 771 | 5379 |
| | NOT IMPORTANT | 682 | 228 | 910 |
| | Grand Total | 5290 | 999 | 6289 |
| | | O.R. | 1.99806018 | |

P-value of 23% is bigger than Alpha = 5%, so we must accept the Null Hypothesis. There is no statistically significant relationship between the memberships and the happiness. But also, since in two cells we have values < 10, it is not recommended to apply this test, the results may not be reliable.

- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: WORK (IP_LIFE_WOR).

| | | FEEL_OF_HAPP | | |
|-------------------------|---------------|---------------------|---------|-------------|
| | | FEELING OF HAPINESS | | |
| IP_LIFE_WOR | | HAPPY | UNHAPPY | Grand Total |
| IMPORTANT IN LIFE: WORK | IMPORTANT | 4998 | 945 | 5943 |
| | NOT IMPORTANT | 277 | 57 | 334 |
| | Grand Total | 5275 | 1002 | 6277 |
| | | O.R. | | 1.08832732 |

| CI of OR | | | |
|---|------------|---------------------------------------|-------------|
| $t(Z_{\alpha/2}) * SE(\ln(OR))$ WHERE $SE(\ln(OR))$ is $\sqrt{1/a+1/b+1/c+1/d}$ | | | |
| $\ln(OR) =$ | 0.08464195 | $SE(\ln(OR)) =$ | 0.149707211 |
| $Z_{\alpha/2} =$ | 1.96 | for CL 95%, $Z_{\alpha/2}$ is 1.96 SD | |
| Lower bound: 0.81157036 | | Upper bound: 1.459462299 | |

Image 55. Table and results for the Odds Ratio between IP_LIFE_WOR and FEEL_OF_HAPP.

Since the C.I. overlaps 1 we must accept the Null Hypothesis, we can say that there is NO association between happiness and work in peoples' lives, with a C.L. of 95%.

- **T-test** of HAPPY? vs. SCALE OF INCOME (INCOME_SCALE)

| RESPONDENT_NUM | INCOME_SCALE | HAPPY? |
|----------------|--------------|--------|
| 720001 | | 5 Yes |
| 720002 | | 3 Yes |
| 720004 | | 7 Yes |
| 720005 | | 4 Yes |
| 720006 | | 4 No |
| 720007 | | 5 No |
| 720008 | | 4 No |
| 720009 | | 5 Yes |
| 720010 | | 5 Yes |
| 720011 | | 4 Yes |

Image 56. Portion of the table created with the new dichotomous variable HAPPY? utilized to build the Pivot Table for the T-test.

| t-Test: Two-Sample Assuming Equal Variances | | |
|---|-------------|------------|
| | Yes | No |
| Mean | 5.095996021 | 4.03488372 |
| Variance | 5.327598392 | 5.15919552 |
| Observations | 4021 | 774 |
| Pooled Variance | 5.300438906 | |
| Hypothesized Mean Difference | 0 | |
| df | 4793 | |
| t Stat | 11.74217756 | |
| P(T<=t) one-tail | 1.03937E-31 | |
| t Critical one-tail | 1.645171604 | |
| P(T<=t) two-tail | 2.07874E-31 | |
| t Critical two-tail | 1.960459052 | |

Image 57. T-test table for INCOME_SCALE vs. HAPPY?

Variances and Observations:

| | | |
|-------------|---|--|
| 1.032641304 | Factor of 2 different? NO, similar variances. | |
| 5.195090439 | Factor of 2 different? YES | |

P- value is much much less than Alpha = 5%, we may reject the Null Hypothesis, and we could say that there is a possible association between the income and the happiness in peoples' lives, with a C.L. of 95%.

BUT, analyzing both distributions below, there seems to be a total overlap between both normal curves, which also leads me to reason that this would indicate there is NO significant difference between the groups.

So, I'm in doubt here. Maybe the detail lies in the number of observations; the sample of the UNHAPPY is 5 times smaller than that of the HAPPY.

The test, in my opinion, is inconclusive for now and should be rerun when a larger number of responses are collected. Another alternative could be to reduce the number of observations in the HAPPY group, but this is very delicate, and the best sampling method for selecting the datapoints would need to be carefully determined.

| INCOME of HAPPY | | INCOME of NOT HAPPY | |
|-------------------------|--------------|-------------------------|--------------|
| Mean | 5.095996021 | Mean | 4.034883721 |
| Standard Error | 0.036399775 | Standard Error | 0.081643293 |
| Median | 5 | Median | 4 |
| Mode | 5 | Mode | 5 |
| Standard Deviation | 2.308159092 | Standard Deviation | 2.271386256 |
| Sample Variance | 5.327598392 | Sample Variance | 5.159195523 |
| Kurtosis | -0.513977818 | Kurtosis | -0.378449931 |
| Skewness | 0.190382189 | Skewness | 0.551810147 |
| Range | 9 | Range | 9 |
| Minimum | 1 | Minimum | 1 |
| Maximum | 10 | Maximum | 10 |
| Sum | 20491 | Sum | 3123 |
| Count | 4021 | Count | 774 |
| Confidence Level(95.0%) | 0.071363735 | Confidence Level(95.0%) | 0.160268856 |

Image 58. Comparative of the descriptive statistics for both groups of people vs. their income.

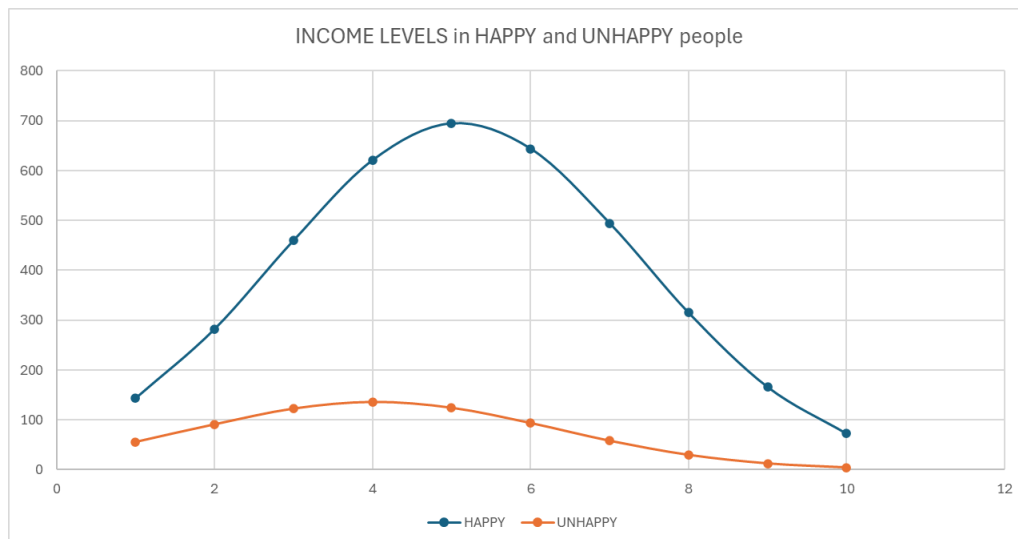


Image 59. Comparative of the bell curves for both groups.

Question b: Did the importance of these remain constant over the years?

Tests:

- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: FAMILY (IP_LIFE_FAM).

Results & Interpretation:

| | | FEEL_OF_HAPP | | |
|---------------------------|---------------|---------------------|---------|-------------|
| | | FEELING OF HAPINESS | | |
| IP_LIFE_FAM | | HAPPY | UNHAPPY | Grand Total |
| IMPORTANT IN LIFE: FAMILY | IMPORTANT | 875 | 180 | 1055 |
| | NOT IMPORTANT | 7 | 6 | 13 |
| | Grand Total | 882 | 186 | 1068 |
| | | O.R. | | 4.1666667 |

Image 60. Odds Ratio of the variables for year 1995, WAVE 3.

| | | FEEL_OF_HAPP | | |
|---------------------------|---------------|---------------------|---------|-------------|
| | | FEELING OF HAPINESS | | |
| IP_LIFE_FAM | | HAPPY | UNHAPPY | Grand Total |
| IMPORTANT IN LIFE: FAMILY | IMPORTANT | 851 | 126 | 977 |
| | NOT IMPORTANT | 6 | 6 | 12 |
| | Grand Total | 857 | 132 | 989 |
| | | O.R. | | 6.7539683 |

Image 61. Odds Ratio of the variables for year 2006, WAVE 5.

| | | FEEL_OF_HAPP | | |
|---------------------------|---------------|---------------------|---------|-------------|
| | | FEELING OF HAPINESS | | |
| IP_LIFE_FAM | | HAPPY | UNHAPPY | Grand Total |
| IMPORTANT IN LIFE: FAMILY | IMPORTANT | 891 | 120 | 1011 |
| | NOT IMPORTANT | 1 | 6 | 7 |
| | Grand Total | 892 | 126 | 1018 |
| | | O.R. | | 44.55 |

Image 62. Odds Ratio of the variables for year 2013, WAVE 6.

| | OR |
|-------------|-------------|
| YEAR = 1995 | 4.166666667 |
| YEAR = 2006 | 6.753968254 |
| YEAR = 2013 | 44.55 |

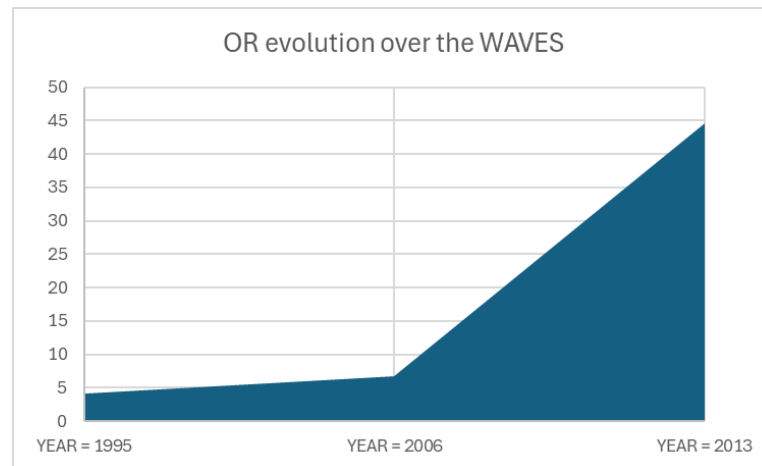


Image 63. Evolution of the Odds Ratio over the WAVES.

Odds Ratio evolution over 3 different years (WAVES 3, 5 and 6). Due to time constraints, only the importance of family was analyzed. The reason I chose WAVE 6 and no 7 (no respecting year separation) is because in WAVE 7 (year 2017) there were no available datapoints for IP_LIFE_FAM option 3 AND FEEL_OF_HAPP levels 3 or 4, so the Pivot Table and results fail.

The trend described in the graph regarding the relationship between happiness and the importance of family seems to be increasing, slightly from years 1995 to 2006, and then increases to reach the WAVE 6 value: that result in my opinion, cannot be considered valid. To be stricter, I would say that good results cannot be obtained in individual years due to the lack of a sufficient number of observations to populate all fields adequately.

Then, I cannot conclude that the results obtained in this question tests are good enough.

Question c: Is there a relationship between religion and the perception of happiness?

Tests:

- **Odds Ratio** of FEELING OF HAPINESS (FEEL_OF_HAPP) vs. IMPORTANT IN LIFE: RELIGION (IP_LIFE_REL).

Results & Interpretation:

| | | FEEL_OF_HAPP | | |
|-----------------------------|---------------|---------------------|---------|-------------|
| | | FEELING OF HAPINESS | | |
| IP_LIFE_REL | | HAPPY | UNHAPPY | Grand Total |
| IMPORTANT IN LIFE: RELIGION | IMPORTANT | 3336 | 632 | 3968 |
| | NOT IMPORTANT | 1923 | 368 | 2291 |
| | Grand Total | 5259 | 1000 | 6259 |
| | | O.R. | | 1.01013053 |

| CI of OR | | | |
|--|------------|---------------------------------------|--------------------------|
| $\exp(\ln(OR) \pm (Z_{\alpha/2}) * SE(\ln(OR)))$ WHERE $SE(\ln(OR))$ is $\sqrt{1/a+1/b+1/c+1/d}$ | | | |
| $\ln(OR) =$ | 0.01007956 | $SE(\ln(OR)) =$ | 0.071550337 |
| $Z_{\alpha/2} =$ | 1.96 | for CL 95%, $Z_{\alpha/2}$ is 1.96 SD | |
| Lower bound: | | 0.87795574 | Upper bound: 1.162204024 |

Image 64. Table and results for the Odds Ratio between IP_LIFE_REL and FEEL_OF_HAPP.

Since the C.I. overlaps 1 we must accept the Null Hypothesis, we can say that there is NO association between happiness and religion in peoples' lives, with a C.L. of 95%.

Question (none): -test made to test Satisfaction with Life in 3 Age separated groups-

Tests:

- **ANOVA** between 3 groups of people separated by AGES (AGES_cat) vs. Satisfaction with your life. (SAT_W_LIFE).

Results & Interpretation:

GROUP 1: People AGE < 32 yo

GROUP 2: People 50 >= AGE >= 32 yo

GROUP 3: People 50 > AGE

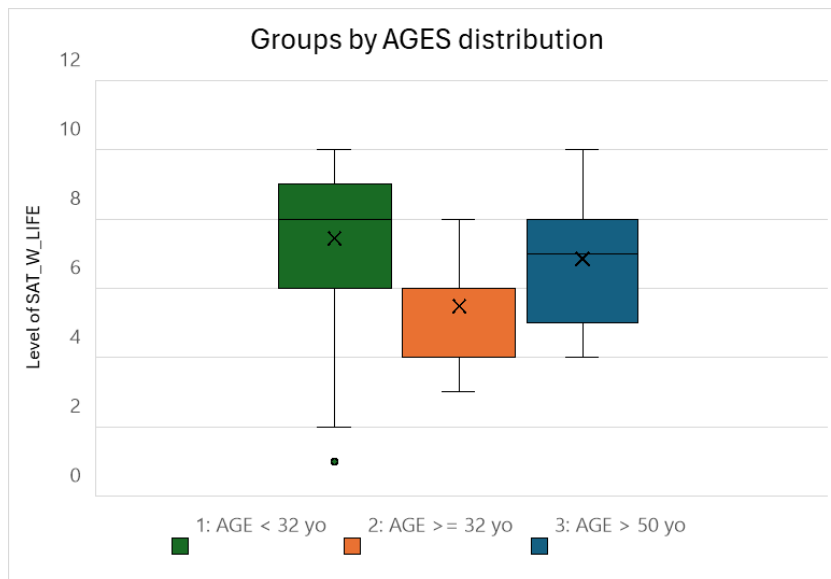


Image 65. Box plot for the 3 groups of people categorized by AGE.

| 3: AGE > 50 yo | | 2: AGE >= 32 yo | | 1: AGE < 32 yo | |
|----------------|----------|-----------------|----------|----------------|----------|
| Mean | 7.207929 | Mean | 7.257278 | Mean | 7.439983 |
| Standard Error | 0.045853 | Standard Error | 0.040068 | Standard Error | 0.03848 |
| Median | 8 | Median | 8 | Median | 8 |
| Mode | 8 | Mode | 8 | Mode | 8 |
| Standard Devi | 2.232715 | Standard Devi | 2.020141 | Standard Devi | 1.891388 |
| Sample Variar | 4.985017 | Sample Variar | 4.080968 | Sample Variar | 3.577349 |
| Kurtosis | 0.343626 | Kurtosis | 0.601894 | Kurtosis | 0.597661 |
| Skewness | -0.80809 | Skewness | -0.77083 | Skewness | -0.76097 |
| Range | 9 | Range | 9 | Range | 9 |
| Minimum | 1 | Minimum | 1 | Minimum | 1 |
| Maximum | 10 | Maximum | 10 | Maximum | 10 |
| Sum | 17090 | Sum | 18448 | Sum | 17975 |
| Count | 2371 | Count | 2542 | Count | 2416 |

Image 66. Comparative of the descriptive statistics for the groups of people vs. their satisfaction with life.

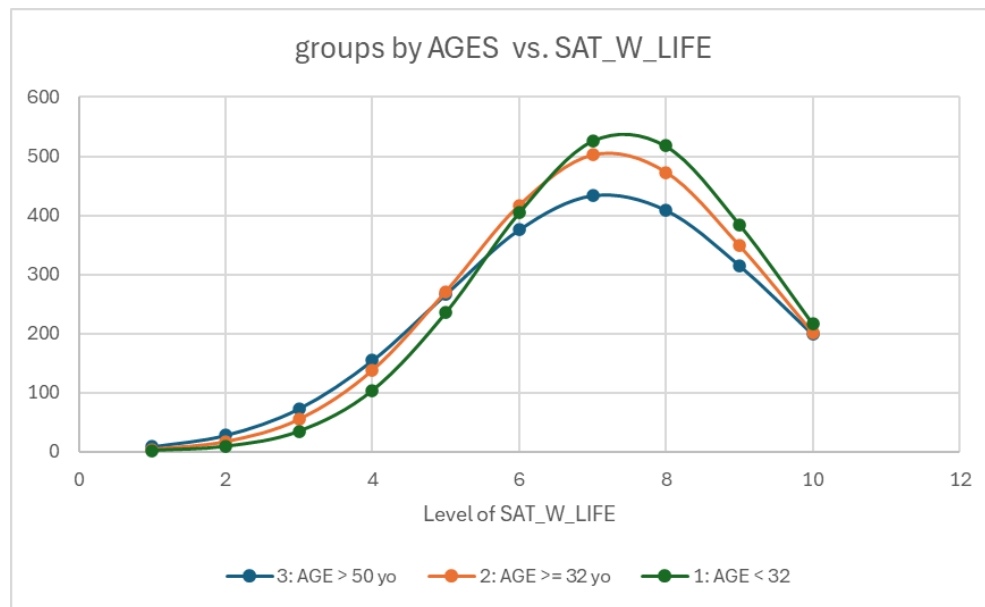


Image 67. Comparative of the bell curves for the 3 groups.

| | | | | | | |
|---|--------------|------------|----------------|-----------------|----------------|---------------|
| Anova: Single Factor | | | | | | |
| SUMMARY | | | | | | |
| <i>Groups</i> | <i>Count</i> | <i>Sum</i> | <i>Average</i> | <i>Variance</i> | <i>SD</i> | |
| 3: AGE > 50 yo | 2370 | 17087 | 7.209705 | 4.979644 | 2.231512 | |
| 2: AGE >= 32 yo | 2542 | 18448 | 7.257278 | 4.080968 | 2.020141 | |
| 1: AGE < 32 yo | 2416 | 17975 | 7.439983 | 3.577349 | 1.891388 | |
| Apparently the younger group is the most satisfied with their life. | | | | | | |
| ANOVA | | | | | | |
| <i>Source of Variation</i> | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> | <i>P-value</i> | <i>F crit</i> |
| Between Groups | 71.27204 | 2 | 35.63602 | 8.473525 | 0.000211 | 2.996958 |
| Within Groups | 30805.81 | 7325 | 4.205572 | | | |
| Total | 30877.09 | 7327 | | | | |

Image 68. Groups descriptive statistics and ANOVA table.

P-value of is much much less than $\alpha = 5\%$, so we reject the Null Hypothesis. There is a statistically significant relationship between the groups in terms of their feeling of satisfaction in their lives.

Despite the possible outlier seen in the green boxplot (AGE < 32 years), it is possible to see a clear difference between the means of the distributions. Looking up the bell curves and the mean values, they tell us interesting information: overall the 3 groups seem to be satisfied since their means are around a good high value of 7, but the most satisfied group is the 1: AGE < 32, followed by the 2: 50 >= AGE >= 32 and in the last place the group 3: AGE > 50.

TESTS to analyze POLITICS AND SOCIETY

Question d: Does the Argentine believe that the responsibility of their lives depends mostly on themselves, or do they place their expectations on the government?

Tests:

- **ANOVA** between 3 groups of people separated by AGES (AGES_cat) vs. Responsibility of the Government to Provide (GOV_RESP_coded).

Results & Interpretation:

GROUP 1: People AGE < 32 yo

GROUP 2: People 50 >= AGE >= 32 yo

GROUP 3: People 50 > AGE

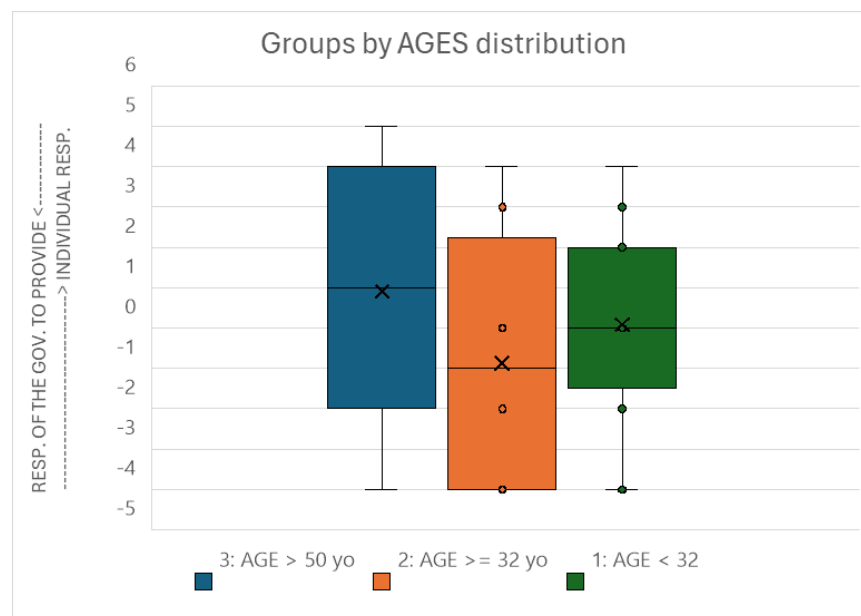


Image 69. Box plot for the 3 groups of people categorized by AGE.

| 3: AGE > 50 yo | | 2: AGE >= 32 yo | | 1: AGE < 32 | |
|----------------|----------|-----------------|----------|----------------|----------|
| Mean | 0.90567 | Mean | 0.973227 | Mean | 0.981647 |
| Standard Error | 0.068433 | Standard Error | 0.064997 | Standard Error | 0.067569 |
| Median | 1 | Median | 1 | Median | 1 |
| Mode | 5 | Mode | 5 | Mode | 5 |
| Standard Devi | 3.055072 | Standard Devi | 2.999019 | Standard Devi | 3.033831 |
| Sample Variar | 9.333467 | Sample Variar | 8.994114 | Sample Variar | 9.204129 |
| Kurtosis | -1.21488 | Kurtosis | -1.15087 | Kurtosis | -1.20184 |
| Skewness | -0.15143 | Skewness | -0.18983 | Skewness | -0.209 |
| Range | 9 | Range | 9 | Range | 9 |
| Minimum | -4 | Minimum | -4 | Minimum | -4 |
| Maximum | 5 | Maximum | 5 | Maximum | 5 |
| Sum | 1805 | Sum | 2072 | Sum | 1979 |
| Count | 1993 | Count | 2129 | Count | 2016 |

Image 70. Comparative of the descriptive statistics for the groups of people vs. their thinking about the responsibility to provide.



Image 71. Comparative of the bell curves for the 3 groups.

| Anova: Single Factor | | | | | | |
|----------------------|----------|------|----------|----------|----------|----------|
| SUMMARY | | | | | | |
| Groups | Count | Sum | Average | Variance | SD | |
| 1 | 2017 | 1983 | 0.983143 | 9.204081 | 3.033823 | |
| 2 | 2130 | 2072 | 0.97277 | 8.990334 | 2.998389 | |
| 3 | 1994 | 1807 | 0.906219 | 9.329384 | 3.054404 | |
| ANOVA | | | | | | |
| Source of Variation | SS | df | MS | F | P-value | F crit |
| Between Groups | 7.01396 | 2 | 3.50698 | 0.382414 | 0.682229 | 2.997195 |
| Within Groups | 56289.31 | 6138 | 9.170627 | | | |

Image 72. Groups descriptive statistics and ANOVA table.

P-value is much much bigger than Alpha = 5%, so we must accept the Null Hypothesis. There is no statistically significant relationship between the groups in terms of their thinking about if the Government should take responsibility to provide, or if People have to take the responsibility to provide for themselves.

But also, looking up the bell curves of the groups, they tell us interesting information: the fact that there is no difference among them, and the three have their mean around the value 1, tell us that the people slightly tend to believe that the responsibility to provide falls on themselves.

Question e: To what extent does interest in politics lead Argentine people to take political actions aimed at influencing society?

Tests:

- **Chi-Square** of How many POLITICAL ACTIONS people might do or have done (POL_JOI_BOY, POL_UNOFF_STR, and POL_ATT_DEM) vs. their INTEREST IN POLITICS (POL_INT)

Results & Interpretation:

| VALUE | CONCLUSION |
|--------|---|
| 1016.3 | People interested in politics are more likely to NOT to participate in any political demonstration. |
| 1255.1 | People participating in one type or more of political demonstrations are more likely to not to |
| 903.76 | be interested in politics in fact. |
| 440.4 | |

Image 75. Table of some expected values analysis

These results are both interesting and puzzling to me. Could there be a confounding variable, or could this be a fact about the Argentine people?

Question f: To what extent does nationalism correlate with interest in politics?

Tests:

- **Odds Ratio** of INTEREST IN POLITICS (POL_INT) vs. PROUD OF NATIONALITY (PROUD_OF_NAT).

Results & Interpretation:

| | | POL_INT | | |
|----------------------|-------------|----------------------|----------------|-------------|
| | | INTEREST IN POLITICS | | |
| PROUD_OF_NAT | | INTEREST | NOT INTERESTED | Grand Total |
| PROUD OF NATIONALITY | PROUD | 1858 | 264 | 2122 |
| | NOT PROUD | 4415 | 472 | 4887 |
| | Grand Total | 6273 | 736 | 7009 |
| | | O.R. | | 0.75240743 |
| | | OR inverse | | 1.32906716 |

| CI of OR | | | | | |
|--|-------------|---|-----------------------|--|--|
| exp(ln(OR) ± (Z _{α/2} * SE(ln(OR)))) WHERE SE(ln(OR)) is sqrt(1/a+1/b+1/c+1/d)) | | | | | |
| ln(OR) = | 0.284477311 | SE(ln(OR)) = | 0.081678 | | |
| Z _{α/2} = | 1.96 | for CL 95%, Z _{α/2} is 1.96 SD | | | |
| Lower bound: | | 1.13245647 | Upper bound: 1.559812 | | |

Image 76. Table and results for the Odds Ratio between PROUD_OF_NAT and POL_INT.

Since the OR is below 1, that suggests negative association. I take the inverse and observing the value and the C.I., they both are- narrowly- above 1, so it may be a slight association. We could say that the people most interested in politics are those who are not very proud of their nationality. BUT, since the values are very close to 1, in fact, I have more doubt than certainty in saying that.

PART 4: Inferential Techniques

Through the application of inferential statistical techniques, I will examine possible relationships between numerical variables of interest, according to my Research Questions. These techniques may allow us to make predictions on the basis of observed data and to determine if it is possible to use sample data to interpolate/ extrapolate behaviors in the population.

Additionally, I will discuss the significance of the R^2 coefficient for each technique performed, the necessary assumptions and the possible risks. Also, through certain calculations, we can determine how appropriate the assumptions enabling the test are; or if the test itself is invalid.

Another interesting procedure we can perform on the data is called Forecasting. It allows us to make inferences about future behaviors or to perform output prediction through the application of different methods and according to aspects such as the type and amount of available data.

For the specific case of this study and its data, we can perform forecasting based on linear regressions. I will use Multiple Linear Regression with sets of three variables. Then, given two of them and using the equation of the best fit plane, it is possible to calculate the third variable (within certain limits).

Important Note:

Some of the following tests may not seem directly related to the Research Questions, but they undoubtedly contribute to clarifying their context and will help us in the final conclusions.

Similar to the structure presented on the previous sections, I will mention the aspect or area of interest and the inferential statistics method applicable to the variables related to that area.

TESTS for HAPPINESS AND SATISFACTION IN LIFE

Test:

- **Linear Regression** of SATISFACTION WITH YOUR LIFE (SAT_W_LIFE_coded) vs. FREEDOM OF CHOICE AND CONTROL (FREE_CHO_CON_coded).

Results & Interpretation:

| NO CHOICE | | | | | Neutral | FREEDOM OF CHOICE | | | | |
|--------------------------------------|------|------|------|-----|---------|-------------------|------|-----|------|-----------|
| Great Deal | G -3 | G -2 | G -1 | | N | G 1 | G 2 | G 3 | G 4 | No Choice |
| New encoding for statistics purposes | | | | | | | | | | |
| -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | |
| 122 | 51 | 144 | 256 | 884 | 698 | 1180 | 1418 | 717 | 1591 | |

| DISSATISFIED GROUP | | | | | Neutral | SATISFIED GROUP | | | | |
|--------------------------------------|------|------|------|-----|---------|-----------------|------|-----|------|----------------------|
| Completely dissatisfied | L -3 | L -2 | L -1 | | N | L 1 | L 2 | L 3 | L 4 | Completely satisfied |
| New encoding for statistics purposes | | | | | | | | | | |
| -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | |
| 147 | 74 | 144 | 255 | 765 | 732 | 1396 | 1779 | 818 | 1218 | |

Image 77. Recoding tables for FREE_CHO_CON_coded and SAT_W_LIFE_coded.

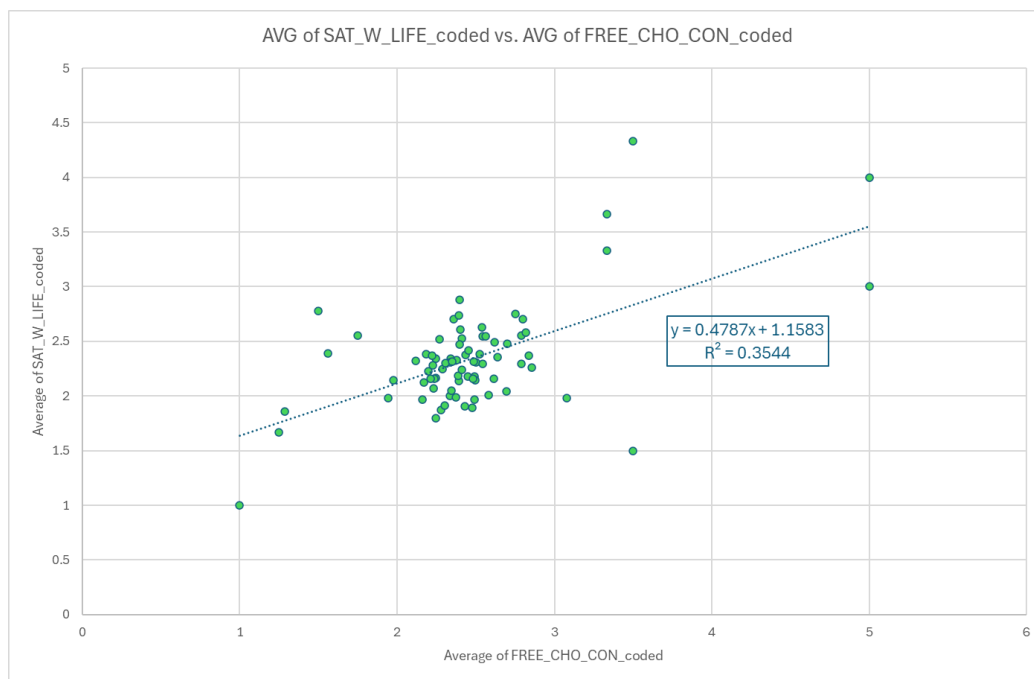


Image 78. Linear Regression of FREE_CHO_CON_coded vs. SAT_W_LIFE_coded.

| Regression Statistics | | | | | | | | |
|-----------------------|--------------|----------------|----------|----------|----------------|------------|-------------|-------------|
| Multiple R | 0.595315012 | | | | | | | |
| R Square | 0.354399964 | | | | | | | |
| Adjusted R Square | 0.345905226 | | | | | | | |
| Standard Error | 0.4778915 | | | | | | | |
| Observations | 78 | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 1 | 9.528012698 | 9.528013 | 41.71994 | 9.00222E-09 | | | |
| Residual | 76 | 17.35690174 | 0.22838 | | | | | |
| Total | 77 | 26.88491444 | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 0.736418253 | 0.273616138 | 2.691428 | 0.008746 | 0.191464486 | 1.28137202 | 0.191464486 | 1.28137202 |
| Average of SAT_W_L | 0.74028544 | 0.114611277 | 6.459098 | 9E-09 | 0.5120173 | 0.96855358 | 0.5120173 | 0.968553581 |

Image 79. Summary output table.

There could be an apparent relationship between the variables, as explained by the best fit line. It is possible to observe a positive relationship, meaning that as respondents report having more freedom of choice in life, they seem to be more satisfied with their lives. However, we also notice that most of the points are concentrated around the point [2.4, 2.3], which means that few of the more distant points could be considered 'outliers' and might be affecting the relationship. Analyzing the R^2 factor, we can say that only 35% of the variation in the dependent variable around is explained by the regression line. This is not an acceptable value to confirm the relationship, but it gives us an idea to focus on in future surveys after obtaining more data points.

Overall, given the concentration around the centroid [2.4, 2.3], we can affirm that the majority of respondents are at good levels of Satisfaction and Freedom of Choice.

Test:

Linear Regression of SATISFACTION WITH YOUR LIFE (SAT_W_LIFE_coded) vs. AGE.

Results & Interpretation:

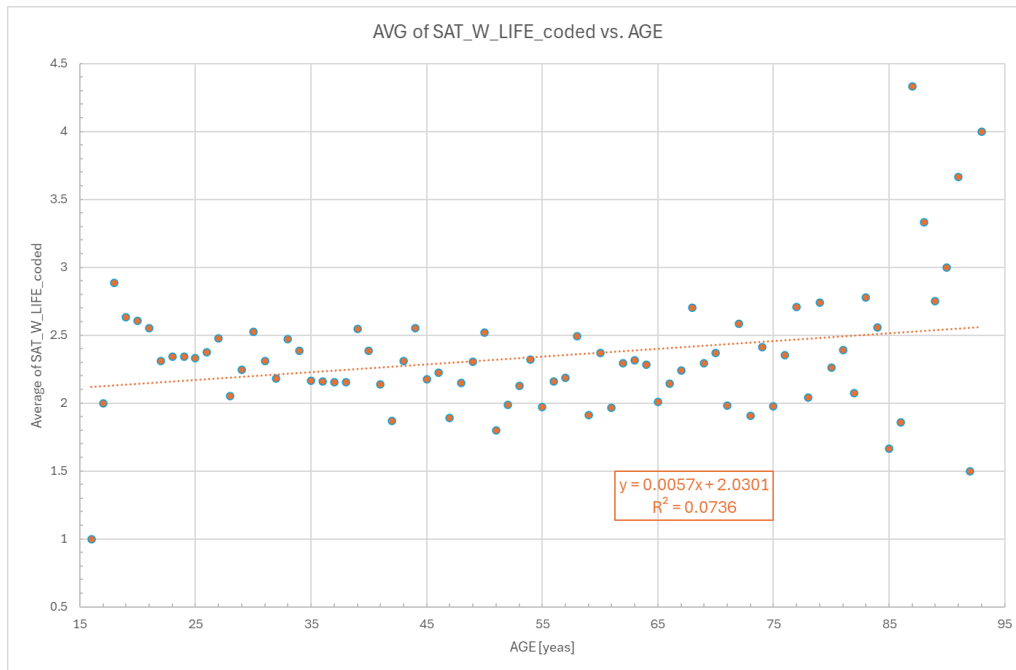


Image 80. Linear Regression of SAT_W_LIFE_coded vs. AGE.

| Regression Statistics | | | | | | | | |
|-----------------------|--------------|----------------|-------------|-------------|----------------|-------------|-------------|-------------|
| Multiple R | 0.271292897 | | | | | | | |
| R Square | 0.073599836 | | | | | | | |
| Adjusted R Square | 0.06141036 | | | | | | | |
| Standard Error | 0.460356703 | | | | | | | |
| Observations | 78 | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 1 | 1.279619233 | 1.279619233 | 6.037982036 | 0.016280999 | | | |
| Residual | 76 | 16.10655036 | 0.211928294 | | | | | |
| Total | 77 | 17.38616959 | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 2.03014865 | 0.136518541 | 14.87086394 | 3.31595E-24 | 1.758248411 | 2.302048889 | 1.758248411 | 2.302048889 |
| AGE | 0.005688854 | 0.002315149 | 2.457230562 | 0.016280999 | 0.001077836 | 0.010299872 | 0.001077836 | 0.010299872 |

Image 81. Summary output table.

As seen on the graph, there is no good correlation between the variables. The R^2 tell us that the percentage of data explained by the regression is as low as 7,36%.

We can't do anything here but accept that maybe we don't have enough data to achieve greater accuracy; or we can simply conclude that it is not possible to model the satisfaction of Argentinians in this sample based on their age. Overall, we can certainly say that the level of satisfaction is relatively high since most Argentinians in the sample fall between level 2 and level 3.

Test:

Linear Regression of FREEDOM OF CHOICE AND CONTROL (FREE_CHO_CON_coded) vs. AGE.

Results & Interpretation:

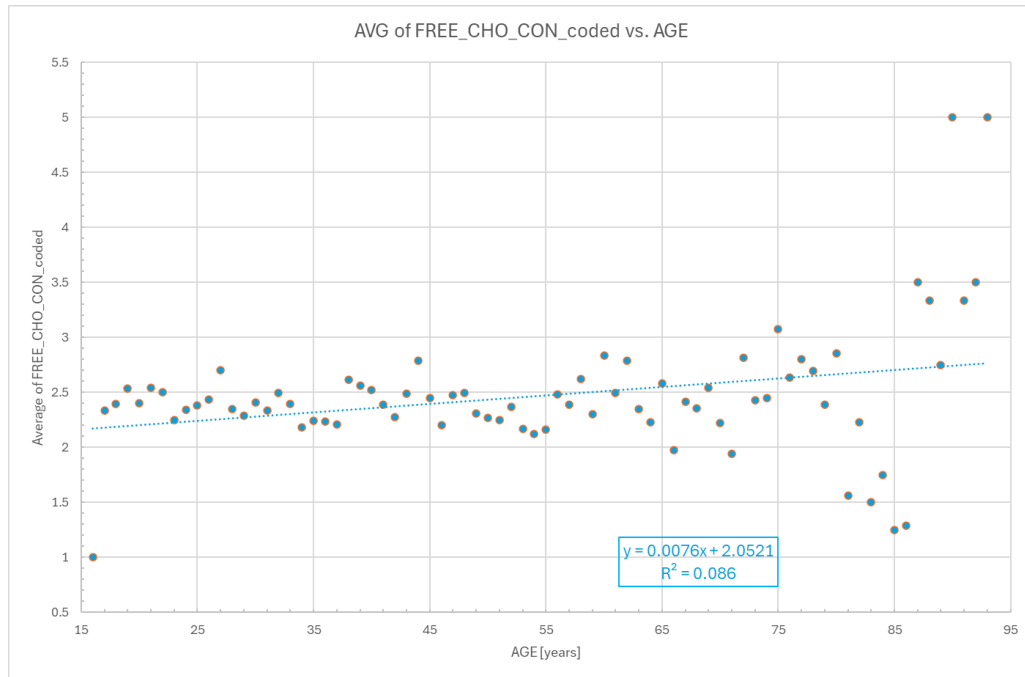


Image 82. Linear Regression of FREE_CHO_CON_coded vs. AGE.

| Regression Statistics | | | | | | | | |
|-----------------------|--------------|----------------|-------------|-------------|----------------|-------------|-------------|-------------|
| Multiple R | 0.293247687 | | | | | | | |
| R Square | 0.085994206 | | | | | | | |
| Adjusted R Square | 0.073967814 | | | | | | | |
| Standard Error | 0.568619838 | | | | | | | |
| Observations | 78 | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 1 | 2.31194687 | 2.31194687 | 7.150457576 | 0.009170471 | | | |
| Residual | 76 | 24.57296757 | 0.323328521 | | | | | |
| Total | 77 | 26.88491444 | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 2.052083258 | 0.168623918 | 12.16958595 | 1.59019E-19 | 1.71623962 | 2.387926896 | 1.71623962 | 2.387926896 |
| AGE | 0.007646687 | 0.002859607 | 2.674033952 | 0.009170471 | 0.001951286 | 0.013342089 | 0.001951286 | 0.013342089 |

Image 83. Summary output table.

Again, as seen on the graph, there is no good correlation between the variables. The R^2 tell us that the percentage of data explained by the regression is as low as 8,6%.

We can't do anything here but accept that maybe we don't have enough data to achieve greater accuracy; or we can simply conclude that it is not possible to model the freedom of choice and

control of Argentinians in this sample based on their age. Overall, we can certainly say that the level of freedom of choice and control is relatively high since most Argentinians in the sample fall between level 2 and level 3.

Test:

Linear Regression of SATISFACTION WITH YOUR LIFE (SAT_W_LIFE_coded) vs. SCALE OF INCOME (INCOME_SCALE).

Results & Interpretation:

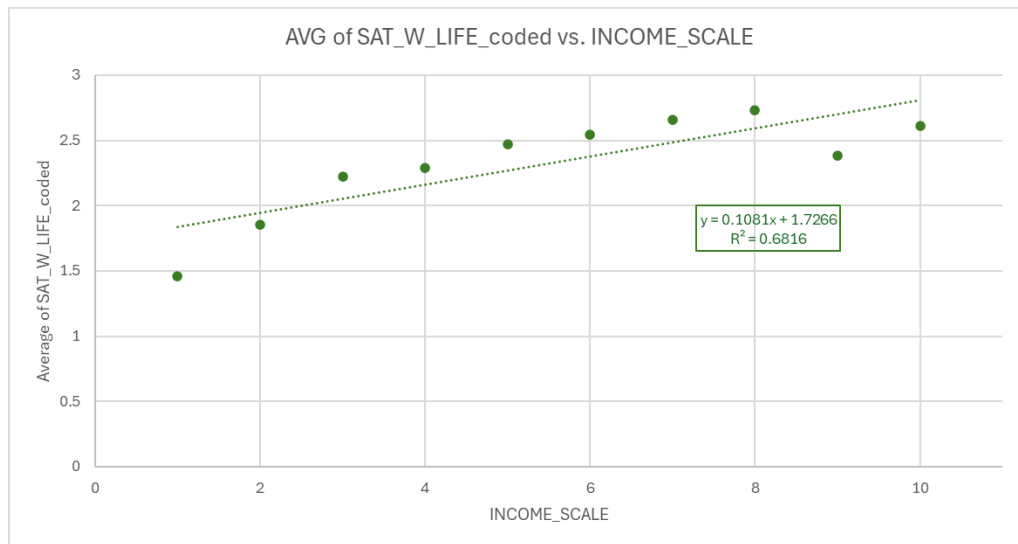


Image 84. Linear Regression of SAT_W_LIFE_coded vs. INCOME_SCALE.

| | | | | | | | |
|------------------------------|-------------|----------------|----------|----------|----------------|--------------|-------------------------|
| Regression Statistics | | | | | | | |
| Multiple R | 0.825619 | | | | | | |
| R Square | 0.681647 | | | | | | |
| Adjusted R | 0.641853 | | | | | | |
| Standard E | 1.811909 | | | | | | |
| Observations | 10 | | | | | | |
| ANOVA | | | | | | | |
| | df | SS | MS | F | Significance F | | |
| Regression | 1 | 56.23589 | 56.23589 | 17.12935 | 0.003258928 | | |
| Residual | 8 | 26.26411 | 3.283014 | | | | |
| Total | 9 | 82.5 | | | | | |
| Coefficients | | | | | | | |
| | Coefficient | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% Upper 95.0% |
| Intercept | -9.14028 | 3.583463 | -2.55068 | 0.034138 | -17.40376358 | -0.876803677 | -17.4038 -0.87680368 |
| Average SAT_W_LIFE_coded | 6.307984 | 1.524123 | 4.138762 | 0.003259 | 2.793348816 | 9.822618396 | 2.793349 9.822618396 |

Image 85. Summary output table.

Observing the regression graph in this case, we see that according to the R^2 , the percentage of data explained by the regression is 68%, which represents, in my opinion, a fairly acceptable value. We can say that as income increases, the satisfaction with life reported by the Argentinians in this sample also increases. Another interesting effect to note is that by observing the lowest income level (1), the satisfaction level still remains in a moderate value (1.5). Remember that, although not shown on the vertical axis of the graph, the SAT_W_LIFE_coded variable has up to 4 negative levels (-1 to -4).

It is also observed that the slope is quite moderate, meaning that for an increase in income by one point, satisfaction only increases by 11%. Income has a positive but smooth influence as it increases, while despite this, the level of life satisfaction remains moderate to high.

Another interesting experiment I tried in this test was removing the extreme income levels, such as levels 1, 9, and 10, and it was observed that the model's accuracy, as indicated by the R^2 , increases to 93%.

Test:

Multiple Linear Regression of SCALE OF INCOME (INCOME_SCALE) vs. SATISFACTION WITH YOUR LIFE (SAT_W_LIFE) vs. AGE.

Results & Interpretation:

| Y | X1 | X2 |
|-----|-------------------------|-----------------------------|
| AGE | Average of INCOME_SCALE | Average of SAT_W_LIFE_coded |
| 18 | 5.577181208 | 2.939597315 |
| 19 | 5.642857143 | 2.598214286 |
| 20 | 5.724489796 | 2.571428571 |
| 21 | 5.153846154 | 2.623931624 |
| 22 | 5.403846154 | 2.336538462 |
| 23 | 5.495049505 | 2.306930693 |
| 24 | 5.118881119 | 2.461538462 |
| 25 | 5.176100629 | 2.421383648 |
| 26 | 5.302083333 | 2.4375 |
| 27 | 5.038095238 | 2.485714286 |
| 28 | 5.068965517 | 2.195402299 |
| 29 | 5.052083333 | 2.104166667 |
| 30 | 5.009615385 | 2.461538462 |
| 31 | 5.293478261 | 2.413043478 |
| 32 | 4.896907216 | 2.360824742 |

Image 86. Part of the Pivot Table considering the three variables of interest.

Case 1: Considering ALL AGES in the sample

| ALL SAMPLES | | | | | | | | |
|---------------------------------------|---|----------------|----------|----------|----------------|-----------|-------------|-------------|
| SUMMARY OUTPUT | | | | | | | | |
| Regression Statistics | | | | | | | | |
| Multiple R | 0.591 | | | | | | | |
| R Square | 0.350 | | | | | | | |
| Adjusted R Square | 0.332 | | | | | | | |
| Standard Error | 18.282 | | | | | | | |
| Observations | 77.000 | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 2 | 13305.00029 | 6652.5 | 19.90398 | 1.21112E-07 | | | |
| Residual | 74 | 24732.99971 | 334.2297 | | | | | |
| Total | 76 | 38038 | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 95.566 | 13.200 | 7.240 | 0.000 | 69.264 | 121.868 | 69.264 | 121.868 |
| Average of INCOME_SCALE | -13.799 | 2.242 | -6.154 | 0.000 | -18.267 | -9.331 | -18.267 | -9.331 |
| Average of SAT_W_LIFE_coded | 9.384 | 4.143 | 2.265 | 0.026 | 1.128 | 17.639 | 1.128 | 17.639 |
| y = AGE | FIT PLANE EQUATION | | | | | | | |
| x ₁ = AVG INCOME_SCALE | y = -13.79 x ₁ + 9.38 x ₂ + 95.57 | | | | | | | |
| x ₂ = AVG SAT_W_LIFE_coded | | | | | | | | |
| example: | | | | | | | | |
| income | 5 | | | | | | | |
| satisfaction | 2.6 | | | | | | | |
| AGE | 50.96696417 | | | | | | | |

Image 87. Summary output table for the MLR considering samples from ALL AGES.

Through this Multiple Linear Regression, we study the combined effects of the variables; this represents a 3-dimensional space, where the best relationship among them can be fitted on a plane. In this method and for our COMPLETE data samples, we found that the R² is 35%, meaning that only that percentage of the relationships can be explained by the best fit plane of the presented equation.

Both predictors have p-values less than 0.05, which indicates that they are statistically significant and that there is sufficient evidence to assert that both predictors have a real effect on the dependent variable. Looking at the coefficients: For INCOME_SCALE, the coefficient value of -13.799 indicates an inverse relationship, meaning that people with higher incomes tend to be younger, while those with lower incomes tend to be older. For SAT_W_LIFE_coded, the coefficient value of +9.384 indicates that people who are more satisfied with their life tend to be older.

Additionally, making use of the plane's equation, I present a simple example of how it could be used for prediction. For example: Given a high-income value (5) and a medium life satisfaction (2.6), we could say that the person is approximately 51 years old.

Case 2: deleting AGES=16, 17 and >= 86 years old from the sample

| DELETING 16, 17 and 86 >= YO samples | | | | | | | | |
|---------------------------------------|---|----------------|----------|----------|----------------|-----------|-------------|-------------|
| SUMMARY OUTPUT | | | | | | | | |
| | | | | | | | | |
| Regression Statistics | | | | | | | | |
| Multiple R | 0.846 | | | | | | | |
| R Square | 0.715 | | | | | | | |
| Adjusted R Square | 0.706 | | | | | | | |
| Standard Error | 10.714 | | | | | | | |
| Observations | 68.000 | | | | | | | |
| | | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 2 | 18735.43814 | 9367.719 | 81.60513 | 1.87841E-18 | | | |
| Residual | 65 | 7461.561862 | 114.7933 | | | | | |
| Total | 67 | 26197 | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 162.736 | 11.144 | 14.603 | 0.000 | 140.479 | 184.992 | 140.479 | 184.992 |
| Average of INCOME_SCALE | -22.451 | 1.842 | -12.185 | 0.000 | -26.130 | -18.771 | -26.130 | -18.771 |
| Average of SAT_W_LIFE_coded | -3.121 | 4.125 | -0.757 | 0.452 | -11.358 | 5.117 | -11.358 | 5.117 |
| | | | | | | | | |
| y = AGE | FIT PLANE EQUATION | | | | | | | |
| x ₁ = AVG INCOME_SCALE | y = -22.45x ₁ + 3.12 x ₂ + 162.74 | | | | | | | |
| x ₂ = AVG SAT_W_LIFE_coded | | | | | | | | |
| | | | | | | | | |
| | example: | | | | | | | |
| income | 5 | | | | | | | |
| satisfaction | 2.6 | | | | | | | |
| AGE | 42.36846278 | | | | | | | |

Image 88. Summary output table

In this case, if we consider the ages mentioned as 'outliers' due to the small number of samples in them (<10), the results are the following. The $R^2 = 71.5\%$ means that the model is quite good at predicting AGE. Analyzing the p-value of 0.452 for the SAT_W_LIFE_coded variable, indicates that there is not enough evidence to say that it has a significant impact on AGE. The p-value for INCOME_SCALE again confirms that this variable has a significant impact on age. This model suggests that the SAT_W_LIFE_coded predictor could be excluded or investigated further to understand why it is not having a significant impact.

Test:

Multiple Linear Regression of FREEDOM OF CHOICE AND CONTROL (FREE_CHO_CON_coded) vs. SATISFACTION WITH YOUR LIFE (SAT_W_LIFE_coded) vs. AGE.

Results & Interpretation:

| Row Labels | Average of SAT_W_LIFE_coded | Average of FREE_CHO_CON_coded |
|------------|-----------------------------|-------------------------------|
| 16 | 1 | 1 |
| 17 | 2 | 2.333333333 |
| 18 | 2.882882883 | 2.398148148 |
| 19 | 2.630434783 | 2.536723164 |
| 20 | 2.607361963 | 2.403726708 |
| 21 | 2.55 | 2.542857143 |
| 22 | 2.309090909 | 2.50310559 |
| 23 | 2.341772152 | 2.246753247 |
| 24 | 2.342222222 | 2.339285714 |
| 25 | 2.331858407 | 2.380090498 |
| 26 | 2.374193548 | 2.436241611 |
| 27 | 2.477707006 | 2.699346405 |
| 28 | 2.051851852 | 2.345864662 |
| 29 | 2.245398773 | 2.288343558 |
| 30 | 2.523809524 | 2.409722222 |
| 31 | 2.310077519 | 2.333333333 |
| 32 | 2.18125 | 2.493670886 |

Image 89. Part of the Pivot Table considering the three variables of interest.

| ALL SAMPLES | | | | | | | | |
|-----------------------|--------------|----------------|----------|-------------|----------------|-----------|-------------|-------------|
| SUMMARY OUTPUT | | | | | | | | |
| Regression Statistics | | | | | | | | |
| Multiple R | 0.316991597 | | | | | | | |
| R Square | 0.100483672 | | | | | | | |
| Adjusted R Square | 0.07649657 | | | | | | | |
| Standard Error | 21.77657024 | | | | | | | |
| Observations | 78 | | | | | | | |
| ANOVA | | | | | | | | |
| | df | SS | MS | F | Significance F | | | |
| Regression | 2 | 3973.074159 | 1986.537 | 4.189070941 | 0.018851218 | | | |
| Residual | 75 | 35566.42584 | 474.219 | | | | | |
| Total | 77 | 39539.5 | | | | | | |
| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
| Intercept | 18.4606448 | 13.0488128 | 1.414738 | 0.161283928 | -7.533923275 | 44.45521 | -7.53392327 | 44.45521288 |
| SAT_W_LIFE_coded | 7.144273168 | 6.499886176 | 1.099138 | 0.275223583 | -5.804164083 | 20.09271 | -5.80416408 | 20.09271042 |
| FREE_CHO_CON_coded | 7.825734391 | 5.227010565 | 1.497172 | 0.138545333 | -2.587004614 | 18.23847 | -2.58700461 | 18.2384734 |

Image 90. Summary output table

In the MLR conducted between the previous variables, it can be observed that the $R^2 = 0.1$ or 10%. It is a very low value indicating that most of the variability in AGE is not explained by the variables included in the model. Observing the p-values we can say that none of the predictors are statistically significant, suggesting that these variables do not have a clear impact on the prediction of age in your sample.

Final Note for this Section: A few additional tests were performed, which can be verified in the MS Excel document. They were not included in this paper as they did not provide any further clarity beyond that provided by the included tests and images.

Results & Discussion

It is reiterated that the conclusions in this section only apply to the survey sample, as it was determined that the available data does not allow for demonstrating that the sample is representative of the Argentinian population.

There are some conclusions that can be established with certainty in response to our questions, while others cannot be answered precisely.

Re-writing our initial questions:

HAPPINESS AND SATISFACTION IN LIFE

Research Question:

"What are the characteristics that most influence an Argentine's happiness? *"En mis viejos tiempos éramos más felices!"* (Were we happier in the old days, as my grandfather used to say?)

It can be said without a doubt that the value placed on family and friends is indisputably high (99% and 83% consider them important) and has a direct relationship with higher levels of happiness compared to those who do not consider these factors important.

Regarding membership in different organizations, it cannot be concluded that this affects happiness. While work is considered an important factor (95%), it does not appear to affect happiness.

Regarding income, it is observed that it does contribute positively to higher levels of happiness, satisfaction, and freedom and control in life, but only in a very slight way; it does not seem to be a determining factor. The Argentinians in these surveys, whether they belong to high-income or low-income groups, remain satisfied with their lives. Interestingly, the tests indicate that as individuals age, their level of satisfaction decreases, but this cannot be concluded with certainty since there are also fewer samples in older age groups.

Regarding religion, most respondents indicate that they belong to a religion (83%), but a smaller number consider religion important in their lives (63%); and in the end, after the tests, it is not possible to determine that religion influences happiness.

For all the reasons mentioned above, without a doubt, for the Argentinians in this survey, Family and Friends are source of happiness and the most important factors in their lives.

POLITICS AND SOCIETY

Research Question:

How does the population perceive the role of politics and its influence on the country and their lives?

The results are not conclusive in demonstrating a significant difference in opinions, as all age groups tend to believe more in individual responsibility to provide rather than placing expectations on the government. The interest in and importance of politics in the lives of the surveyed Argentinians are relatively low (30% each), and those who participate in socio-political demonstrations are also a minority. On the other hand, it is noted that nearly 90% of respondents consider themselves proud of their nationality, but this seems to have no relation to the importance they place on politics.

Future Work

Another interesting technique to apply, but it is beyond the scope of this paper, could be to implement PCA Factor Analysis (Principal Component Analysis). Some possible implementation is to separate the data points into two clusters -those who feel happy with their lives and those who do not (FEEL_OF_HAPP)- and analyze them considering the numerical variables SAT_W_LIFE, FREE_CHO_CON, MARITAL_STA, HOW_MANY_CHILD, and INCOME_SCALE. Through this process, we could attempt to determine whether any combination of variables (and which ones to a greater extent) can explain the variations in the happiness or satisfaction of the respondents. In fact, numerous additional analyses can be conducted if we combine implementations of the technique with the creation of various data clusters based on the available polytomous variables. A few more examples, Cluster 1 could include those who consider family important and work less important (IP_LIFE_FAM + IP_LIFE_WOR) vs. the previously mentioned numerical variables. Cluster 2 could include those who consider work important and give less importance to family. Cluster 3 could consist of those who are interested in politics and participate in demonstrations (POL_INT + POL_ATT_DEM / POL_UNOFF_STR / POL_JOI_BOY). We could continue generating a wide variety of clusters with different characteristics, determine the factors that most likely affect each one, and even compare clusters to draw interesting conclusions. This task

remains to be implemented by the author of this paper on some boring and cold winter afternoon, or by any reader interested in delving further.

Tracking this Analysis

Tracking notes on the actions performed on the raw data to allow any analyst to replicate the analysis or any reader without a background to understand it.

| Number | NOTES |
|--------|--|
| a | Where to find the data ? I got the dataset from: |
| b | Download it as .csv format |
| c | Imported to jupyter notebook, processing with Pandas library |
| d | Filtered and Trimmed by Country = Argentina and, by variables -questions- prioritizing those present in the majority of the seven surveys. |
| e | Filtered and Trimmed by variables of personal interest |
| f | Renaming the variables labels -check Data Dictionary- |
| g | Saved as a new .csv as you see it in Sheet 1 |
| h | The size of the dataset is 7402 instances by 55 columns |
| 1 | Creation of the WORKING SHEET named 'WVS_study_ARGENTINA' |
| 2 | Re-Ordering and freezing of the columns by dependant and independent variables |
| 3 | Removing the numbers '32' or '320' at the beginning of "RESPONDENT_NUM" column because the coding represents the country. New column "RESPONDENT_NUM" created. Hidden the previous column 'J' |
| 4 | Removing the '32000' at the beginning of "ETHNIC_GROUP" column because the coding represents the country. New column "ETHNIC_GROUP" created. Hidden the previous column 'E'. |
| 5 | I discard interesting variables as: 'WEIGHT', 'PROF_JOB' due to inconsistencies . |
| 6 | ETHNIC_GROUP' available data for Waves 5 and 7, changed ethnic numer 4 to 6 in Wave 5 to ensure data consistency. |
| 7 | Change numbers 1 -> 2 and 2 -> 1 in all 'MEM_XXX_XXX' columns in WAVE 3 and WAVE 1 because the meaning is inverted compared to WAVES 5, 6 AND 7. Hidden old columns: 'AF', 'AH', 'AJ', 'AL', 'AN', 'AP'. |
| 8 | I discard the variables as: 'LEVEL_EDU', 'LEVEL_EDU_RESP' because this data is better codified in the column 'LEVEL_EDU_COUNTRY'. |
| 9 | Removing the numbers '3200' OR '320' at the beginning of "LEVEL_EDU_COUNTRY" column because the coding represents the country. New column "LEVEL_EDU_COUNTRY" created. Hidden the previous column 'N'. |
| 10 | Corrected the format on column 'DATE_INT', hidden the previous column 'C'. |
| 11 | Removing of the column 'FIGHT_FOR_COU' due to inconsistencies . |
| 12 | SHEETS RE-ORGANIZED |

| | | | |
|----|--|--|--|
| 13 | Ccolumns of the variables "DATE_INT" "YEAR_SVY_START" and "ETHNIC_GROUP" in the main sheet, unfreezed and moved to the right to give viewers a broader range of visible variables. | | |
| 14 | "Introduction & Scope" and "Data Source and Considerations " sections updated. | | |
| 15 | "Processing the Data " and "Variables" section updated. | | |
| 16 | "Dataset Exploration Part 1: Excel Visuals " section updated. | | |
| 17 | "Dataset Exploration Part 2: Variables Univariate Analysis" new section to compile Part 2 study of the variables. | | |
| 18 | Add column 'Missing/ Erroneous Data' to 'Data Dictionary' | | |
| 19 | Recoding/ Categorization of data in the following variables: SEX, ETHNIC_GROUP, FEEL_OF_HAPP, SAT_W_LIFE,FREE_CHO_CON, IP_LIFE_LEI, REL_GROUPS, INCOME_SCALE | | |
| 20 | ETHNIC_GROUP: inconsistencies and missing values were found among the surveys. | | |
| 21 | Performing the multivariable analysis. | | |
| 22 | 9 or 99 interpreted as Don't Answer for consistency between surveys. | | |
| 23 | RELIGION: Recode 9 from surveys 6, 5 & 2 as -2 for consistency | | |
| 24 | "Tracking this Analysis " section updated. | | |
| 25 | Reordering of the useful variables in the main sheet,. | | |
| 26 | Created Extra column for recoding | | |
| 27 | Added new variables and univariate analysis to the report | | |
| 28 | MEM_HUM_ORG' Column BH created, string coding of MEMBERSHIP for Chi-Square test purpose | | |
| 29 | MEM_SP_REC' Column BI created, string coding of MEMBERSHIP for Chi-Square test purpose | | |
| 30 | MEM_LAB_UNI' Column BJ created, string coding of of MEMBERSHIP for Chi-Square test purpose | | |
| 31 | MEM_MUS_EDU' Column BK created, string coding of of MEMBERSHIP for Chi-Square test purpose | | |
| 32 | MEM_POL_PAR' Column BL created, string coding of of MEMBERSHIP for Chi-Square test purpose | | |
| 33 | MEM_REL_ORG' Column BM created, string coding for Chi Square table purposes | | |
| 34 | NUM_MEMBERSHIPS' variable created in column BN, it count the number of MEMBERSHIPS, for Chi-Square test purpose | | |
| 35 | HAPPY ? variable created in column BO, dichotomous for T-test purposes | | |
| 36 | AGES_cat variable created in column BP, for ANOVA test purpose | | |
| 37 | GOV_RESP_coded variable created in column BQ, recoding for ANOVA test purpose | | |
| 38 | POL_JOI_BOY_willing variable created in column BR, dichotomous for Chi-Square test purpose | | |
| 39 | POL_UNOFF_STR_willing variable created in column BS, dichotomous for Chi-Square test purpose | | |
| 40 | POL_ATT_DEM_willing variable created in column BT, dichotomous for Chi-Square test purpose | | |
| 41 | NUM_POLITICAL_ACTION variable created in column BU, it count the number of joining the POLITICAL ACTIONS, for Chi-Square test purpose | | |
| 42 | SAMPLEvs.POPULATION sheet created to check if this survey samples are representative to Argentina's population | | |
| 43 | Created new sheets on this document for the analysis through inferential methods | | |
| 44 | Re-order of the tests sheets to match the order of appearance on the paper. | | |
| 45 | Column with the numerical variables used in inferential tests re-coloured to light grey and grouped | | |

Reference List

1. World Values Survey (WVS),

Inglehart, R., C. Haerper, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen (eds.). 2022. World Values Survey: All Rounds - Country-Pooled Datafile. Madrid, Spain & Vienna, Austria: JD Systems Institute & WWSA Secretariat. Dataset Version 3.0.0. [doi:10.14281/18241.17](https://doi.org/10.14281/18241.17)

<https://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp>

2. BDAT1005-24S-30681 Math for Data Analytics, classes material.

Author: Mr. Jonathan Gladstone, BES, BASc, P.Eng.

3. Cover image taken from World Values Survey Association LinkedIn, <https://www.linkedin.com/company/world-values-survey-association/posts/?feedView=all>

4. Argentine population, official Argentina's website, <https://www.argentina.gob.ar/pais/poblacion>

5. "Beliefs, values and attitudes in Argentine society" official study from CONICET, website: <https://www.conicet.gov.ar/creencias-valores-y-actitudes-en-la-sociedad-argentina/>

6. "World Values Survey Wave 7 (2017-2022) Variables Report V6.0"

<https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>

It is recommended to check this link and download the .pdf if you want

7. "En mis viejos tiempos éramos más felices!" -*sentence in spanish language-*

"In my old times we were happier!". Assertion made by my grandfather Juan Carlos Sarrailh, on occasions when there was talk of dissatisfaction with life situations or material desires.

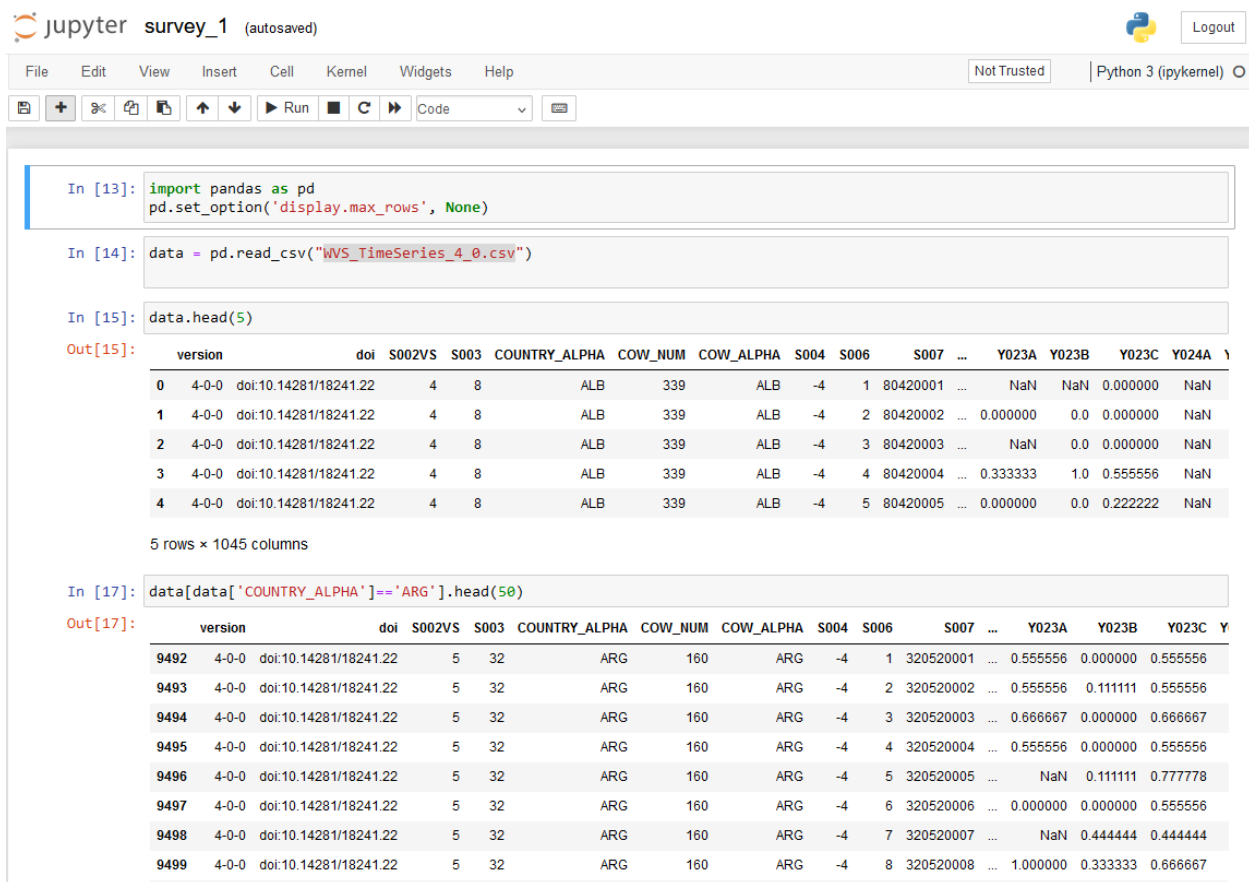
8. Argentina's Population pyramid. Datosmacro.com

<https://datosmacro.expansion.com/demografia/estructura-poblacion/argentina>

APPENDIX A

Some screenshots of the Python code to extract data applicable to Argentina and to rename the variables. If you want the exact and full code, feel free to ask for it to the following email address: fedesarraih@hotmail.com

Filtering process for the original extracted dataset: “WVS_TimeSeries_4_0.csv”



The screenshot shows a Jupyter Notebook interface with the following code and output:

```
In [13]: import pandas as pd
pd.set_option('display.max_rows', None)
```

```
In [14]: data = pd.read_csv("WVS_TimeSeries_4_0.csv")
```

```
In [15]: data.head(5)
```

Out[15]:

| | version | doi | S002VS | S003 | COUNTRY_ALPHA | COW_NUM | COW_ALPHA | S004 | S006 | S007 | ... | Y023A | Y023B | Y023C | Y024A |
|---|---------|-----------------------|--------|------|---------------|---------|-----------|------|------|----------|-----|----------|-------|----------|-------|
| 0 | 4-0-0 | doi:10.14281/18241.22 | 4 | 8 | ALB | 339 | ALB | -4 | 1 | 80420001 | ... | NaN | NaN | 0.000000 | NaN |
| 1 | 4-0-0 | doi:10.14281/18241.22 | 4 | 8 | ALB | 339 | ALB | -4 | 2 | 80420002 | ... | 0.000000 | 0.0 | 0.000000 | NaN |
| 2 | 4-0-0 | doi:10.14281/18241.22 | 4 | 8 | ALB | 339 | ALB | -4 | 3 | 80420003 | ... | NaN | 0.0 | 0.000000 | NaN |
| 3 | 4-0-0 | doi:10.14281/18241.22 | 4 | 8 | ALB | 339 | ALB | -4 | 4 | 80420004 | ... | 0.333333 | 1.0 | 0.555556 | NaN |
| 4 | 4-0-0 | doi:10.14281/18241.22 | 4 | 8 | ALB | 339 | ALB | -4 | 5 | 80420005 | ... | 0.000000 | 0.0 | 0.222222 | NaN |

5 rows × 1045 columns

```
In [17]: data[data['COUNTRY_ALPHA'] == 'ARG'].head(50)
```

Out[17]:

| | version | doi | S002VS | S003 | COUNTRY_ALPHA | COW_NUM | COW_ALPHA | S004 | S006 | S007 | ... | Y023A | Y023B | Y023C | Y |
|------|---------|-----------------------|--------|------|---------------|---------|-----------|------|------|-----------|-----|----------|----------|----------|---|
| 9492 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 1 | 320520001 | ... | 0.555556 | 0.000000 | 0.555556 | |
| 9493 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 2 | 320520002 | ... | 0.555556 | 0.111111 | 0.555556 | |
| 9494 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 3 | 320520003 | ... | 0.666667 | 0.000000 | 0.666667 | |
| 9495 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 4 | 320520004 | ... | 0.555556 | 0.000000 | 0.555556 | |
| 9496 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 5 | 320520005 | ... | NaN | 0.111111 | 0.777778 | |
| 9497 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 6 | 320520006 | ... | 0.000000 | 0.000000 | 0.555556 | |
| 9498 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 7 | 320520007 | ... | NaN | 0.444444 | 0.444444 | |
| 9499 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 8 | 320520008 | ... | 1.000000 | 0.333333 | 0.666667 | |

Filtering for Argentina:

```
In [18]: data2=data[data['COUNTRY_ALPHA']=='ARG']

In [19]: data2.info()

<class 'pandas.core.frame.DataFrame'>
Index: 7401 entries, 9492 to 16892
Columns: 1045 entries, version to SurvSagg
dtypes: float64(53), int64(988), object(4)
memory usage: 59.1+ MB

In [21]: #grabo el dataframe a un nuevo csv solo de ARG
data2.to_csv('WVS_TimeSeries_ARGENTINA.csv')
```

Filtering of the variables of interest

```
In [2]: data = pd.read_csv("WVS_TimeSeries_ARGENTINA.csv")

In [3]: data.head(5)

Out[3]:
```

| | Unnamed: 0 | version | doi | S002VS | S003 | COUNTRY_ALPHA | COW_NUM | COW_ALPHA | S004 | S006 | ... | Y023A | Y023B | Y023C | Y024A |
|---|------------|---------|-----------------------|--------|------|---------------|---------|-----------|------|------|-----|----------|----------|----------|-------|
| 0 | 9492 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 1 | ... | 0.555556 | 0.000000 | 0.555556 | 0.00 |
| 1 | 9493 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 2 | ... | 0.555556 | 0.111111 | 0.555556 | 0.33 |
| 2 | 9494 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 3 | ... | 0.666667 | 0.000000 | 0.666667 | 0.00 |
| 3 | 9495 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 4 | ... | 0.555556 | 0.000000 | 0.555556 | 1.00 |
| 4 | 9496 | 4-0-0 | doi:10.14281/18241.22 | 5 | 32 | ARG | 160 | ARG | -4 | 5 | ... | NaN | 0.111111 | 0.777778 | 0.66 |

5 rows x 1046 columns

```
In [4]: columns_to_select1= ['A001', 'A002', 'A003', 'A004', 'A005', 'A006', 'A008', 'A009', 'A100', 'A101', 'A102', 'A105',
In [5]: columns_to_select2= ['E012', 'E023', 'E026', 'E027', 'E028', 'E037']
In [6]: columns_to_select3= ['E069_01', 'E069_02', 'E069_04', 'E069_05', 'E069_07', 'E069_09', 'E069_11', 'E069_12', 'E114', 'E115',
In [7]: columns_to_select4= ['F025', 'G006', 'S002VS', 'S007', 'S012', 'S020', 'X025', 'X025A_01', 'X025CSWVS']
In [8]: columns_to_select5= ['X001', 'X002', 'X003', 'X007', 'X011', 'X028', 'X036', 'X047_WVS', 'X049', 'X050B', 'X051']
In [9]: columns= columns_to_select1 + columns_to_select2 + columns_to_select3 + columns_to_select4 + columns_to_select5
In [10]: len(columns)
Out[10]: 54
```

```
In [13]: data2.head(7)

Out[13]:
```

| | A001 | A002 | A003 | A004 | A005 | A006 | A008 | A009 | A009 | A100 | ... | X002 | X003 | X007 | X011 | X028 | X036 | X047_WVS | X049 | X050B | X051 |
|---|------|------|------|------|------|------|------|------|------|------|-----|------|------|------|------|------|------|----------|------|-------|---------|
| 0 | 1 | 2 | 2 | 3 | 1 | 1 | 2 | 2 | 0 | 0 | ... | 1984 | 21 | 6 | 0 | 2 | 25 | -4 | 8 | -4.0 | 32001.0 |
| 1 | 1 | 3 | 3 | 3 | 3 | 3 | 1 | 0 | 0 | 0 | ... | 1971 | 35 | 2 | 0 | 1 | 51 | -4 | 8 | -4.0 | 32001.0 |
| 2 | 1 | 1 | 2 | 4 | 1 | 2 | 1 | 0 | 0 | 0 | ... | 1980 | 26 | 1 | 2 | 5 | 61 | -4 | 8 | -4.0 | 32001.0 |
| 3 | 1 | 2 | 1 | 3 | 1 | 2 | 2 | 0 | 0 | 0 | ... | 1958 | 48 | 4 | 2 | 1 | 33 | -4 | 8 | -4.0 | 32001.0 |
| 4 | 1 | 1 | 1 | 3 | 1 | 3 | 1 | 0 | 0 | 0 | ... | 1974 | 32 | 2 | 2 | 1 | 25 | -4 | 8 | -4.0 | 32001.0 |
| 5 | 1 | 1 | 2 | 1 | 4 | 1 | 2 | 2 | 0 | 1 | ... | 1936 | 70 | 6 | 0 | 4 | 21 | -4 | 8 | -4.0 | 32001.0 |
| 6 | 1 | 1 | 1 | 2 | 1 | 4 | 3 | 0 | 0 | 0 | ... | 1947 | 59 | 1 | 1 | 3 | 21 | -4 | 8 | -4.0 | 32001.0 |

7 rows x 54 columns

Renaming of the variables

RENAMING

```
In [14]: data2.rename(columns={'A001': 'IP_LIFE_FAM', 'A002': 'IP_LIFE_FRI', 'A003': 'IP_LIFE_LEI', 'A004': 'IP_LIFE_POL', 'A005': 'IP_LIFE_MOR', 'A006': 'IP_LIFE_REL', 'A008': 'FEEL_OF_HAP', 'A009': 'MEM_REL_ORG', 'A009': 'MEM_SP_REC', 'A100': 'MEM_MUS_EDU', 'A101': 'MEM_LAB_UNI', 'A102': 'MEM_POLI_PAR', 'A105': 'MEM_HUM_ORG', 'A170': 'SAT_W_LIFE', 'A173': 'FREE_CHO_CO_N'}, inplace=True)
```

```
C:\Users\Federico\AppData\Local\Temp\ipykernel_4896\3732586336.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
```

```
In [15]: data2.rename(columns={'E012': 'FIGHT_FOR_COU', 'E023': 'POL_INT', 'E026': 'POL_JOI_BOY', 'E027': 'POL_ATT_DEM', 'E028': 'POL_UNOOF_STR', 'E037': 'GOV_RESP', 'E069_01': 'CONF_CHU', 'E069_02': 'CONF_ARMED', 'E069_04': 'CONF_PRESS', 'E069_05': 'CONF_LABOURUN', 'E069_07': 'CONF_PARL', 'E069_09': 'CONF_SECSYS', 'E069_11': 'CONF_GOV', 'E069_12': 'CONF_POLPART'}, inplace=True)
```

```
C:\Users\Federico\AppData\Local\Temp\ipykernel_4896\2886563540.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
```

```
In [16]: data2.rename(columns={'E114': 'POL_SYS_LEADER', 'E115': 'POL_SYS_EXPERT', 'E116': 'POL_SYS_ARMY', 'E117': 'POL_SYS_DEM', 'E118': 'POL_SYS_REL', 'E119': 'POL_SYS_REL', 'E120': 'POL_SYS_REL', 'E121': 'POL_SYS_REL', 'E122': 'POL_SYS_REL', 'E123': 'POL_SYS_REL', 'E124': 'POL_SYS_REL', 'E125': 'POL_SYS_REL', 'E126': 'POL_SYS_REL', 'E127': 'POL_SYS_REL', 'E128': 'POL_SYS_REL', 'E129': 'POL_SYS_REL', 'E130': 'POL_SYS_REL', 'E131': 'POL_SYS_REL', 'E132': 'POL_SYS_REL', 'E133': 'POL_SYS_REL', 'E134': 'POL_SYS_REL', 'E135': 'POL_SYS_REL', 'E136': 'POL_SYS_REL', 'E137': 'POL_SYS_REL', 'E138': 'POL_SYS_REL', 'E139': 'POL_SYS_REL', 'E140': 'POL_SYS_REL', 'E141': 'POL_SYS_REL', 'E142': 'POL_SYS_REL', 'E143': 'POL_SYS_REL', 'E144': 'POL_SYS_REL', 'E145': 'POL_SYS_REL', 'E146': 'POL_SYS_REL', 'E147': 'POL_SYS_REL', 'E148': 'POL_SYS_REL', 'E149': 'POL_SYS_REL', 'E150': 'POL_SYS_REL', 'E151': 'POL_SYS_REL', 'E152': 'POL_SYS_REL', 'E153': 'POL_SYS_REL', 'E154': 'POL_SYS_REL', 'E155': 'POL_SYS_REL', 'E156': 'POL_SYS_REL', 'E157': 'POL_SYS_REL', 'E158': 'POL_SYS_REL', 'E159': 'POL_SYS_REL', 'E160': 'POL_SYS_REL', 'E161': 'POL_SYS_REL', 'E162': 'POL_SYS_REL', 'E163': 'POL_SYS_REL', 'E164': 'POL_SYS_REL', 'E165': 'POL_SYS_REL', 'E166': 'POL_SYS_REL', 'E167': 'POL_SYS_REL', 'E168': 'POL_SYS_REL', 'E169': 'POL_SYS_REL', 'E170': 'POL_SYS_REL', 'E171': 'POL_SYS_REL', 'E172': 'POL_SYS_REL', 'E173': 'POL_SYS_REL', 'E174': 'POL_SYS_REL', 'E175': 'POL_SYS_REL', 'E176': 'POL_SYS_REL', 'E177': 'POL_SYS_REL', 'E178': 'POL_SYS_REL', 'E179': 'POL_SYS_REL', 'E180': 'POL_SYS_REL', 'E181': 'POL_SYS_REL', 'E182': 'POL_SYS_REL', 'E183': 'POL_SYS_REL', 'E184': 'POL_SYS_REL', 'E185': 'POL_SYS_REL', 'E186': 'POL_SYS_REL', 'E187': 'POL_SYS_REL', 'E188': 'POL_SYS_REL', 'E189': 'POL_SYS_REL', 'E190': 'POL_SYS_REL', 'E191': 'POL_SYS_REL', 'E192': 'POL_SYS_REL', 'E193': 'POL_SYS_REL', 'E194': 'POL_SYS_REL', 'E195': 'POL_SYS_REL', 'E196': 'POL_SYS_REL', 'E197': 'POL_SYS_REL', 'E198': 'POL_SYS_REL', 'E199': 'POL_SYS_REL', 'E200': 'POL_SYS_REL', 'E201': 'POL_SYS_REL', 'E202': 'POL_SYS_REL', 'E203': 'POL_SYS_REL', 'E204': 'POL_SYS_REL', 'E205': 'POL_SYS_REL', 'E206': 'POL_SYS_REL', 'E207': 'POL_SYS_REL', 'E208': 'POL_SYS_REL', 'E209': 'POL_SYS_REL', 'E210': 'POL_SYS_REL', 'E211': 'POL_SYS_REL', 'E212': 'POL_SYS_REL', 'E213': 'POL_SYS_REL', 'E214': 'POL_SYS_REL', 'E215': 'POL_SYS_REL', 'E216': 'POL_SYS_REL', 'E217': 'POL_SYS_REL', 'E218': 'POL_SYS_REL', 'E219': 'POL_SYS_REL', 'E220': 'POL_SYS_REL', 'E221': 'POL_SYS_REL', 'E222': 'POL_SYS_REL', 'E223': 'POL_SYS_REL', 'E224': 'POL_SYS_REL', 'E225': 'POL_SYS_REL', 'E226': 'POL_SYS_REL', 'E227': 'POL_SYS_REL', 'E228': 'POL_SYS_REL', 'E229': 'POL_SYS_REL', 'E230': 'POL_SYS_REL', 'E231': 'POL_SYS_REL', 'E232': 'POL_SYS_REL', 'E233': 'POL_SYS_REL', 'E234': 'POL_SYS_REL', 'E235': 'POL_SYS_REL', 'E236': 'POL_SYS_REL', 'E237': 'POL_SYS_REL', 'E238': 'POL_SYS_REL', 'E239': 'POL_SYS_REL', 'E240': 'POL_SYS_REL', 'E241': 'POL_SYS_REL', 'E242': 'POL_SYS_REL', 'E243': 'POL_SYS_REL', 'E244': 'POL_SYS_REL', 'E245': 'POL_SYS_REL', 'E246': 'POL_SYS_REL', 'E247': 'POL_SYS_REL', 'E248': 'POL_SYS_REL', 'E249': 'POL_SYS_REL', 'E250': 'POL_SYS_REL', 'E251': 'POL_SYS_REL', 'E252': 'POL_SYS_REL', 'E253': 'POL_SYS_REL', 'E254': 'POL_SYS_REL', 'E255': 'POL_SYS_REL', 'E256': 'POL_SYS_REL', 'E257': 'POL_SYS_REL', 'E258': 'POL_SYS_REL', 'E259': 'POL_SYS_REL', 'E260': 'POL_SYS_REL', 'E261': 'POL_SYS_REL', 'E262': 'POL_SYS_REL', 'E263': 'POL_SYS_REL', 'E264': 'POL_SYS_REL', 'E265': 'POL_SYS_REL', 'E266': 'POL_SYS_REL', 'E267': 'POL_SYS_REL', 'E268': 'POL_SYS_REL', 'E269': 'POL_SYS_REL', 'E270': 'POL_SYS_REL', 'E271': 'POL_SYS_REL', 'E272': 'POL_SYS_REL', 'E273': 'POL_SYS_REL', 'E274': 'POL_SYS_REL', 'E275': 'POL_SYS_REL', 'E276': 'POL_SYS_REL', 'E277': 'POL_SYS_REL', 'E278': 'POL_SYS_REL', 'E279': 'POL_SYS_REL', 'E280': 'POL_SYS_REL', 'E281': 'POL_SYS_REL', 'E282': 'POL_SYS_REL', 'E283': 'POL_SYS_REL', 'E284': 'POL_SYS_REL', 'E285': 'POL_SYS_REL', 'E286': 'POL_SYS_REL', 'E287': 'POL_SYS_REL', 'E288': 'POL_SYS_REL', 'E289': 'POL_SYS_REL', 'E290': 'POL_SYS_REL', 'E291': 'POL_SYS_REL', 'E292': 'POL_SYS_REL', 'E293': 'POL_SYS_REL', 'E294': 'POL_SYS_REL', 'E295': 'POL_SYS_REL', 'E296': 'POL_SYS_REL', 'E297': 'POL_SYS_REL', 'E298': 'POL_SYS_REL', 'E299': 'POL_SYS_REL', 'E300': 'POL_SYS_REL', 'E301': 'POL_SYS_REL', 'E302': 'POL_SYS_REL', 'E303': 'POL_SYS_REL', 'E304': 'POL_SYS_REL', 'E305': 'POL_SYS_REL', 'E306': 'POL_SYS_REL', 'E307': 'POL_SYS_REL', 'E308': 'POL_SYS_REL', 'E309': 'POL_SYS_REL', 'E310': 'POL_SYS_REL', 'E311': 'POL_SYS_REL', 'E312': 'POL_SYS_REL', 'E313': 'POL_SYS_REL', 'E314': 'POL_SYS_REL', 'E315': 'POL_SYS_REL', 'E316': 'POL_SYS_REL', 'E317': 'POL_SYS_REL', 'E318': 'POL_SYS_REL', 'E319': 'POL_SYS_REL', 'E320': 'POL_SYS_REL', 'E321': 'POL_SYS_REL', 'E322': 'POL_SYS_REL', 'E323': 'POL_SYS_REL', 'E324': 'POL_SYS_REL', 'E325': 'POL_SYS_REL', 'E326': 'POL_SYS_REL', 'E327': 'POL_SYS_REL', 'E328': 'POL_SYS_REL', 'E329': 'POL_SYS_REL', 'E330': 'POL_SYS_REL', 'E331': 'POL_SYS_REL', 'E332': 'POL_SYS_REL', 'E333': 'POL_SYS_REL', 'E334': 'POL_SYS_REL', 'E335': 'POL_SYS_REL', 'E336': 'POL_SYS_REL', 'E337': 'POL_SYS_REL', 'E338': 'POL_SYS_REL', 'E339': 'POL_SYS_REL', 'E340': 'POL_SYS_REL', 'E341': 'POL_SYS_REL', 'E342': 'POL_SYS_REL', 'E343': 'POL_SYS_REL', 'E344': 'POL_SYS_REL', 'E345': 'POL_SYS_REL', 'E346': 'POL_SYS_REL', 'E347': 'POL_SYS_REL', 'E348': 'POL_SYS_REL', 'E349': 'POL_SYS_REL', 'E350': 'POL_SYS_REL', 'E351': 'POL_SYS_REL', 'E352': 'POL_SYS_REL', 'E353': 'POL_SYS_REL', 'E354': 'POL_SYS_REL', 'E355': 'POL_SYS_REL', 'E356': 'POL_SYS_REL', 'E357': 'POL_SYS_REL', 'E358': 'POL_SYS_REL', 'E359': 'POL_SYS_REL', 'E360': 'POL_SYS_REL', 'E361': 'POL_SYS_REL', 'E362': 'POL_SYS_REL', 'E363': 'POL_SYS_REL', 'E364': 'POL_SYS_REL', 'E365': 'POL_SYS_REL', 'E366': 'POL_SYS_REL', 'E367': 'POL_SYS_REL', 'E368': 'POL_SYS_REL', 'E369': 'POL_SYS_REL', 'E370': 'POL_SYS_REL', 'E371': 'POL_SYS_REL', 'E372': 'POL_SYS_REL', 'E373': 'POL_SYS_REL', 'E374': 'POL_SYS_REL', 'E375': 'POL_SYS_REL', 'E376': 'POL_SYS_REL', 'E377': 'POL_SYS_REL', 'E378': 'POL_SYS_REL', 'E379': 'POL_SYS_REL', 'E380': 'POL_SYS_REL', 'E381': 'POL_SYS_REL', 'E382': 'POL_SYS_REL', 'E383': 'POL_SYS_REL', 'E384': 'POL_SYS_REL', 'E385': 'POL_SYS_REL', 'E386': 'POL_SYS_REL', 'E387': 'POL_SYS_REL', 'E388': 'POL_SYS_REL', 'E389': 'POL_SYS_REL', 'E390': 'POL_SYS_REL', 'E391': 'POL_SYS_REL', 'E392': 'POL_SYS_REL', 'E393': 'POL_SYS_REL', 'E394': 'POL_SYS_REL', 'E395': 'POL_SYS_REL', 'E396': 'POL_SYS_REL', 'E397': 'POL_SYS_REL', 'E398': 'POL_SYS_REL', 'E399': 'POL_SYS_REL', 'E400': 'POL_SYS_REL', 'E401': 'POL_SYS_REL', 'E402': 'POL_SYS_REL', 'E403': 'POL_SYS_REL', 'E404': 'POL_SYS_REL', 'E405': 'POL_SYS_REL', 'E406': 'POL_SYS_REL', 'E407': 'POL_SYS_REL', 'E408': 'POL_SYS_REL', 'E409': 'POL_SYS_REL', 'E410': 'POL_SYS_REL', 'E411': 'POL_SYS_REL', 'E412': 'POL_SYS_REL', 'E413': 'POL_SYS_REL', 'E414': 'POL_SYS_REL', 'E415': 'POL_SYS_REL', 'E416': 'POL_SYS_REL', 'E417': 'POL_SYS_REL', 'E418': 'POL_SYS_REL', 'E419': 'POL_SYS_REL', 'E420': 'POL_SYS_REL', 'E421': 'POL_SYS_REL', 'E422': 'POL_SYS_REL', 'E423': 'POL_SYS_REL
```

APPENDIX B

3-D plots made in Python and Matplotlib codes for the **Multiple Linear Regression** of SCALE OF INCOME (INCOME_SCALE) vs. SATISFACTION WITH YOUR LIFE (SAT_W_LIFE) vs. AGE.

ALL AGES considered: best fit plane in blue

Deleted AGES=16, 17 and ≥ 86 years old: best fit plane in orange

