

# Causes of Divorce

Pablo X Zumba – Data Analyst

Ruth M Ojeda – Psychologists

01/18/2022

## ASK

### Introduction / Topic to explore

This project will explore the possible causes of divorce using the knowledge of a psychologist as a stakeholder and a reliable database offered by the government. The stakeholder recommends the following type of information would be very valuable for the analysis; however, the stakeholder is aware that there might be limitations to get specific data.

Data requested: Race, age, sex, religion, highest level of education, marital status over the years, psychological and pathological background before marriage such as: emotional dependence, esteem, depression, eating disorders. Drug use, family background such as: income, catastrophic and psychiatric diseases. Infidelity if married, domestic violence (verbal, psychological or sexual), number of kids, family income, level of communication between couple, dating time before marriage, married years, sexual life.

### What problems we are trying to solve or investigate?

We attempt to provide a psychologist with information about the possible causes of divorce (focusing on psychological factors such as depression) so that she can use this information to improve or modify the marital therapy session that she provides to her patients. Some specific questions the stakeholder wants to answer are:

- Why don't marriages last too long?
- What are the factors that lead from marriage to divorce?
- Marital failure or divorce is a consequence of some psychological factors such as depression or anxiety?
- Are infidelity, physical or psychological abuse or quality of life predominant factors when deciding a divorce?
- What are the patterns that stand out in a group of divorced people? Age, religion, years of marriage, time they were in love before getting married?

### What Metrics are use to measure the data to achieve the objective. Who are the stakeholders?

We use and filter a dataset that meets most of the stakeholder's criteria and then use filtering techniques to separate the dataset into different groups, for example: people who have been divorced at least once and people who are still married or were still married at the end of the time

period provided by the survey in the database. At that point, the main differences between those groups will be the metrics. For example: if we find that drug use is more common in people who have been divorced, we will take this information as a metric.

### **Who is the audience for this analysis and how does this affect the analysis process and presentation?**

The audience is the psychologists who give marital therapy. These results should be understandable to any psychologist with a basic understanding of statistics.

### **How does this data help the stakeholder make decisions?**

The results of the study will help the stakeholder to better understand the possible causes of divorce in a specific group of people so that they can make improvements in marital therapy sessions or focus on key aspects of individuals who have been divorced.

### **Deliverables:**

- A clear summary of the business task
- Cleaning and manipulation script
- A summary of the analysis process
- Visualizations and key findings
- Insights based on the analysis

### **Activity Summary**

The project consists of 6 different phases:

- The Ask phase that is present at the beginning of this document.
- The Prepare phase that consists mostly of getting the data that fits stakeholders' criteria and cleaning the process to make the data useful and reliable.
- The Process phase focuses on ensuring that the data is clean and uses filtering techniques to get the data ready for the analysis.
- The analysis phase looks for patterns and trends that are useful to answering stakeholders' questions and may also find other unexpected patterns.
- The Share phase uses different charts to make the data findings more understandable.
- The Act phase gives insights and conclusions to the stakeholder using the analysis and charts given in the previous phases.

### **PREPARE**

The chosen dataset contains information of people that were born between 1980 and 1984 and are currently 38-42 years old. The time-frame survey is from 1997 to 2020. The dataset was downloaded from the United States Department of Labor | National Longitudinal Surveys. There were some limitations to find some data requested by the stakeholder but after a meeting, we decided that the most relevant and available information in the database are the following: dysfunctional family, gender, age, religion, female depression, male depression, ethnicity, marital

status at the end of the survey, parents divorced, personality scale, medication for depression, medication for sleep disorders, reason first marriage ended and highest level of education.

Link: <https://www.nlsinfo.org/investigator/pages/search?s=NLSY97>

We also made some assumptions based on the interpretation table provided by the Department of Labor. The assumptions are described in each step of the analysis.

### **Where is your data located?**

The data is called NLSY97 and is located at the 'NLSY97 National Longitudinal Surveys' into the program of the US Bureau of Labor Statistics. Link: <https://www.bls.gov/nls/nlsy97.htm#topical-guide>

### **How is the data organized?**

- Each person surveyed has a unique code that has 18 different variables.
- The survey last 22 years and made observations each year from 1997 to 2011 and after that, each two years: 2013, 2015, 2017, 2019.
- The number of observations and the total number of participants during the survey are: 8,984
- All the variables are categorical and not continuous.

### **Are there issues with bias or credibility in this data? Does the data ROCCC?**

ROCCC stands for Reliable, Original, Comprehensive, Current, and Cited.

- Reliable | HIGH because the source is the United States Department of Labor.

- Original | HIGH due the source reputation.

- Comprehensive | MEDIUM due the limitations on some psychological parameters requested by the stakeholder.

- CURRENT | HIGH since the last round of survey was at 2020.

- CITED | HIGH due the source reputation.

### **Licensing, privacy, security, and accessibility?**

The NLSY97 is made public-use and addresses licensing, privacy and accessibility using a tool called 'Investigator' to find the desired data before to download.

### **Verifying the data's integrity?**

We assumed the integrity of the data based on the source reputation and after the cleaning process, the data has been reviewed to check inconsistencies.

### **Does the dataset helps to answer the main question?**

The dataset has some limitations, however, the stakeholder states that the the current information is very useful and can be taking into account.

## Problems with the data?

There are some parameters requested that are not in the dataset or the stakeholder decided to do not analyze it:

- Psychological and pathological factors: emotional dependence, esteem, eating disorders, drug use
- Psychiatric and catastrophic family illnesses.
- Income and number of kids varies too much over the time and we decided to not analyze it any more.
- Infidelity if married cannot be determined.
- Domestic violence verbal, psychological or sexual cannot be determined.
- It is not possible to determine the level of communication between the couple during dating and sexual life. It's a sensitive factor that cannot be trusted in this situation and must be taken into account directly with the patient.

## PROCESS

### Tools chosen to work

We chose RStudio to clean, analyze data, and create an HTML/PDF document using R Markdown files. For charts and graphics we used Tableau.

### Steps to ensure data is clean

After to get the data downloaded as .csv file, the tidyverse library was installed for the cleaning process. The data was imported to Rstudio and explored using `str()` and `summary()` commands. After exploration the data was transformed to a tibble and a function called 'vallabels' was created to modify numeric values for categorical data. Consecutively, renaming columns was necessary to better understand the context of data. Missing values were set to NA and we proceed to check for the existence of Null data, duplicates and mismatched data type using the summary function.

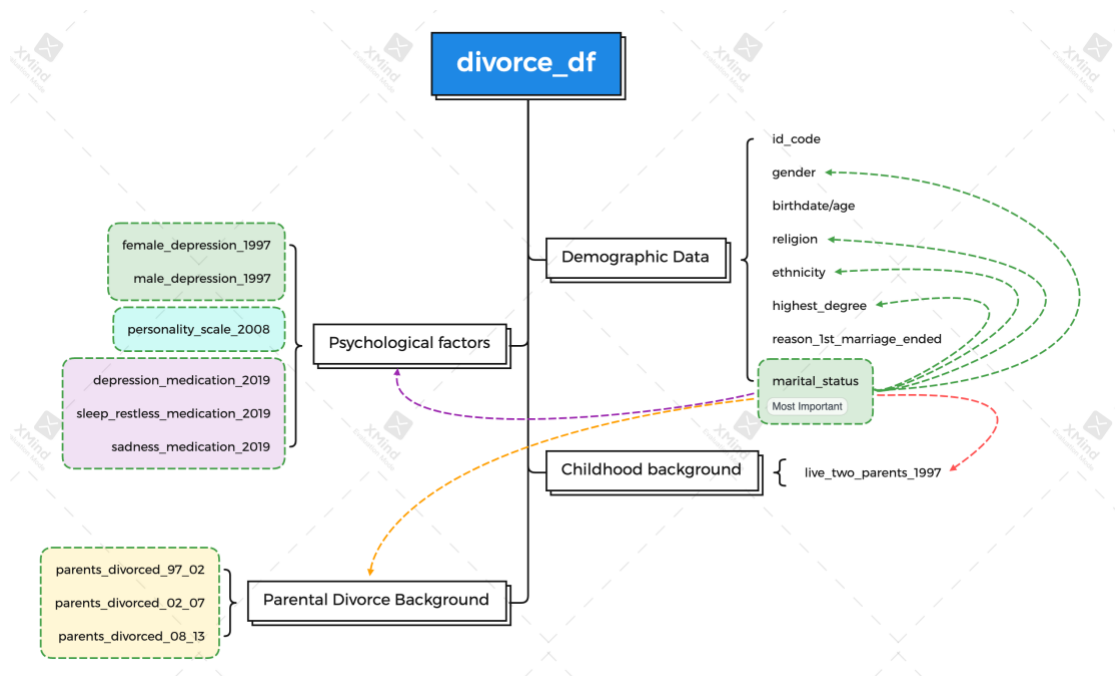
### How to verify the data is clean and ready to analyze

We can state the data is reasonable clean if each variable is one column, each observation is one row, datatypes of each column has been checked and duplicates and blank fields were handled. The cleaning processes and verification is detailed and documented on the "Cleaning\_Process" script for future reference. Cleaning process available on github at:

## ANALYZE

### How is the data organized to perform analysis on it?

The variables has been organized in four different groups: Childhood background, Parental Divorce Background, Demographic Data and Psychological factors. It is clear that this project focuses on the possible causes of divorce so the "marital\_status" variable is key to analyze versus the other variables. Here is a how we organize the data to perform analysis and the relationship between marital\_status with other variables:



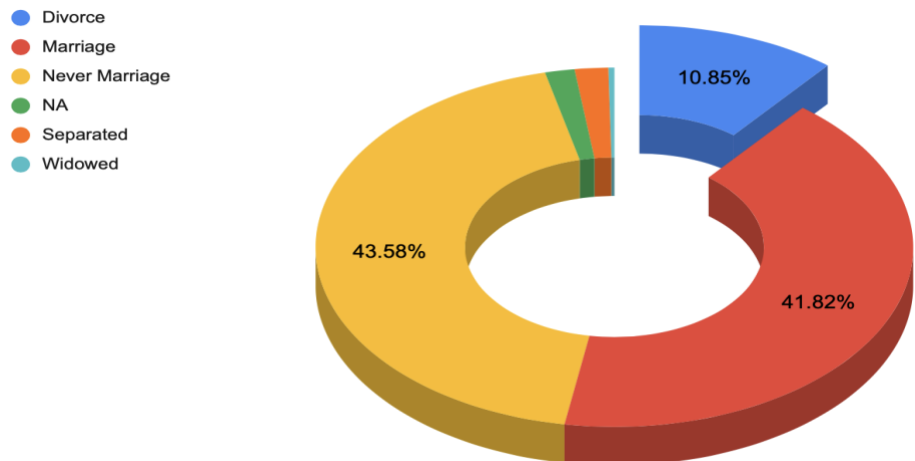
*Variables organized in four different groups*

### How is the data formatted?

All the variables has been transformed in categorical data and only the `id_code` variable is continuous so we can say the our data is discrete and has only a limited amount of possible states. For example, the 'marital\_status' variable has only 6 possible states: Never-married, Married, Separated, Divorced, Widowed and NA.

**What surprises did you discover in the data?**

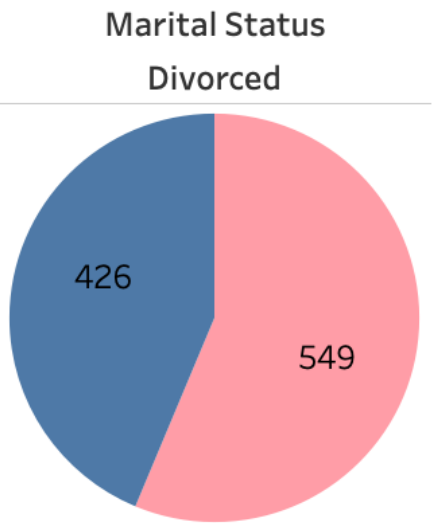
About 11% of those surveyed during 1997-2019 are divorced people.



*Figure 1 Percentage of people divorced compared with a different marital status*

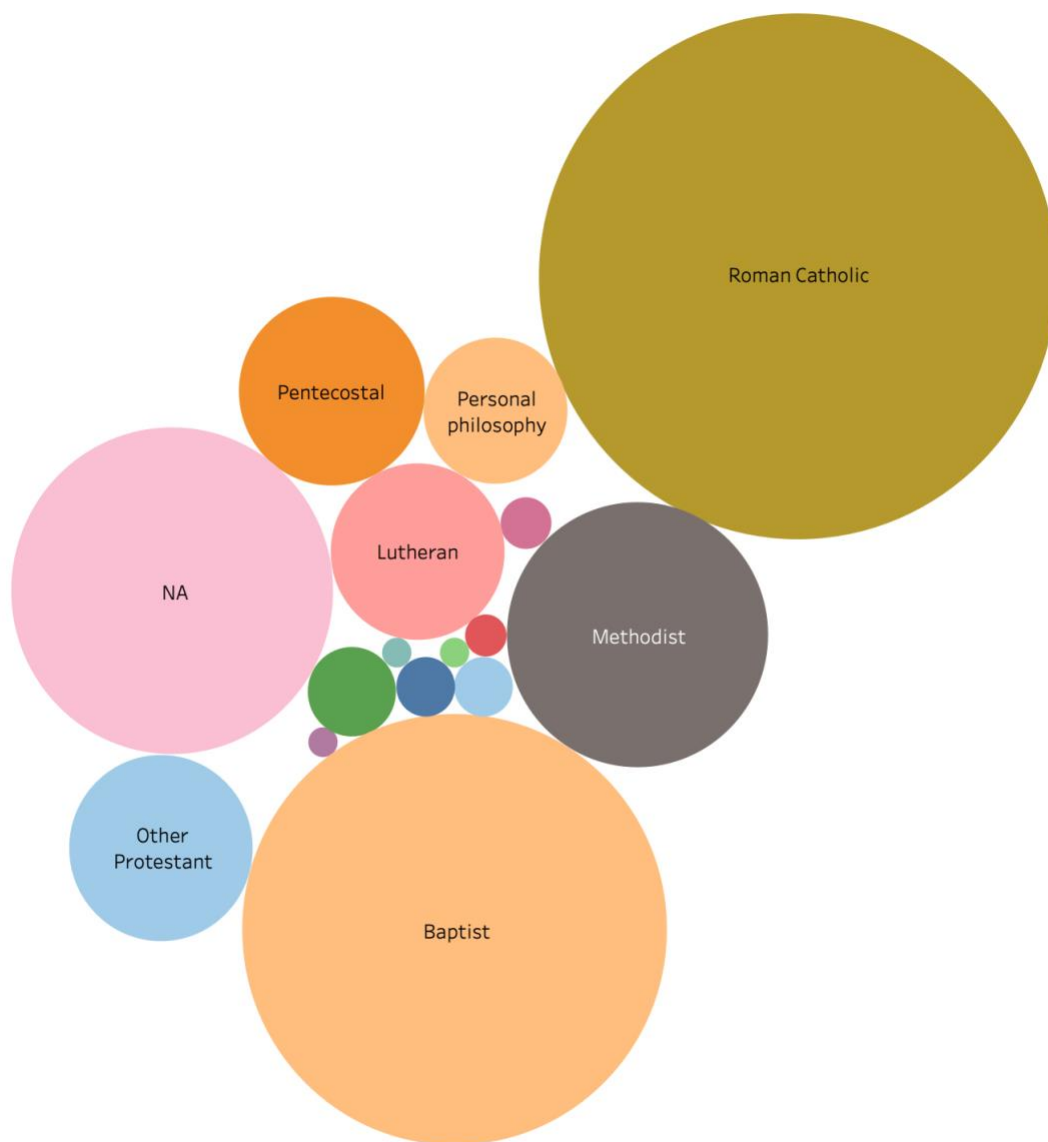
At first glance, it was surprising that although the number of men surveyed (4,599) is higher than the number of women surveyed (4,385), the results show that more women reported being divorced than men.

Total number of divorces by Gender



*Figure 2 Number of divorced people by gender.*

There are more Roman Catholic people divorced than any other religion.



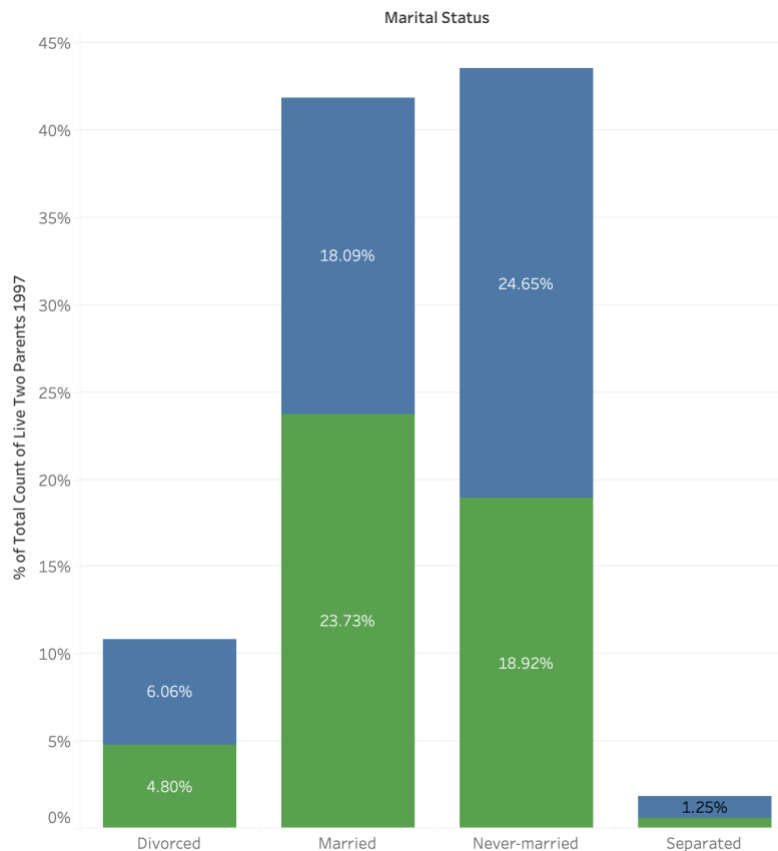
*Figure 3 Dispersion of the type of religion among the divorced people.*

The data show that having a higher academic degree decreases the ratio between marriage and divorce, that is, having a higher academic degree leads to less chance of getting divorced.

Male Marital Status	None	GED	Associate	High School Diploma (12 Years)	Bachelor	Master	PhD
Married	72	113	196	567	396	149	18
Divorced	31	53	54	146	48	15	1
Ratio	0.43	0.47	0.28	0.26	0.12	0.10	0.06
Female Marital Status	None	GED	Associate	High School Diploma (12 Years)	Bachelor	Master	PhD
Married	73	77	226	491	469	253	25
Divorced	43	36	84	183	91	35	3
Ratio	0.59	0.47	0.37	0.37	0.19	0.14	0.12

*Figure 4 Highest academic degree by gender.*

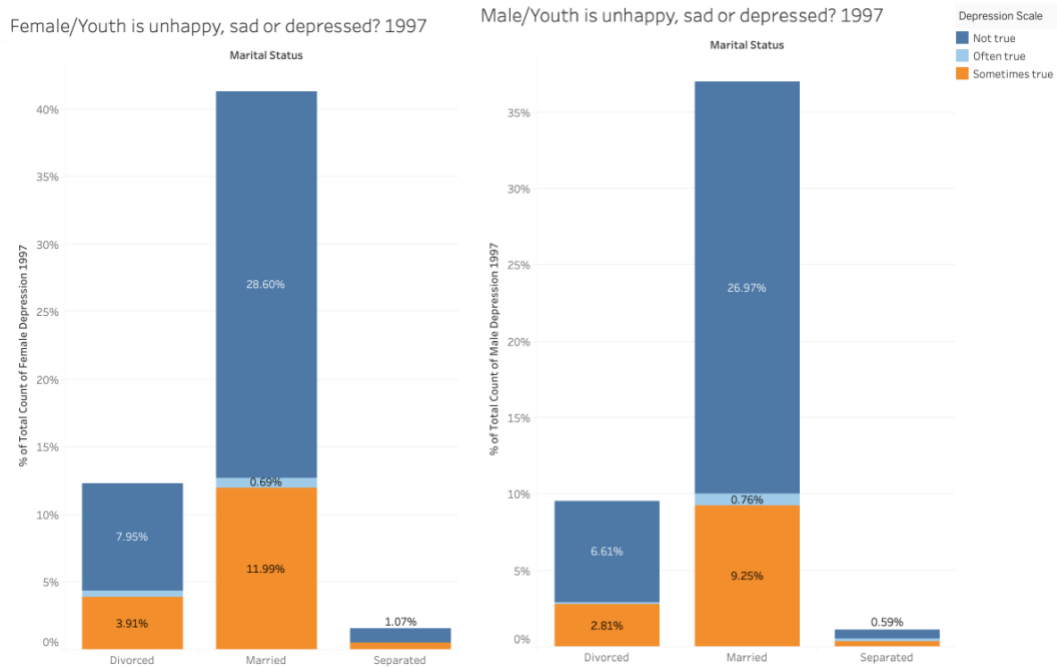
There is a higher percentage of divorced, never-married and separated people who said they had not lived with both parents in 1997.



*Figure 5 Percentage of individuals who were living with both parent in 1997. Green=Yes / Blue=No.*

The depression factor when divorced respondents were teenagers (1997) does not appear to be directly related to divorce.

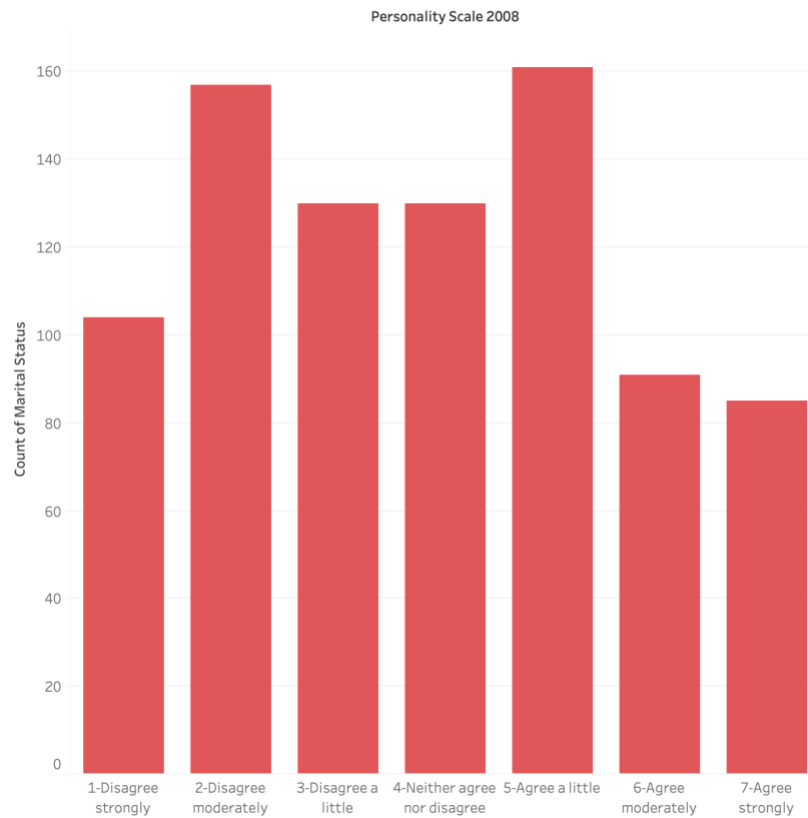




*Figure 6 Depression in 1997 by Gender*

When respondents were given a personality scale question in 2008, there was a tendency for most divorcees to say they agreed a little that they felt anxious or easily upset, that is, there is a greater number of divorcees who admit to having problems with anxiety and temper control in 2008.

Does youth feel Anxious or easily upset? 2008

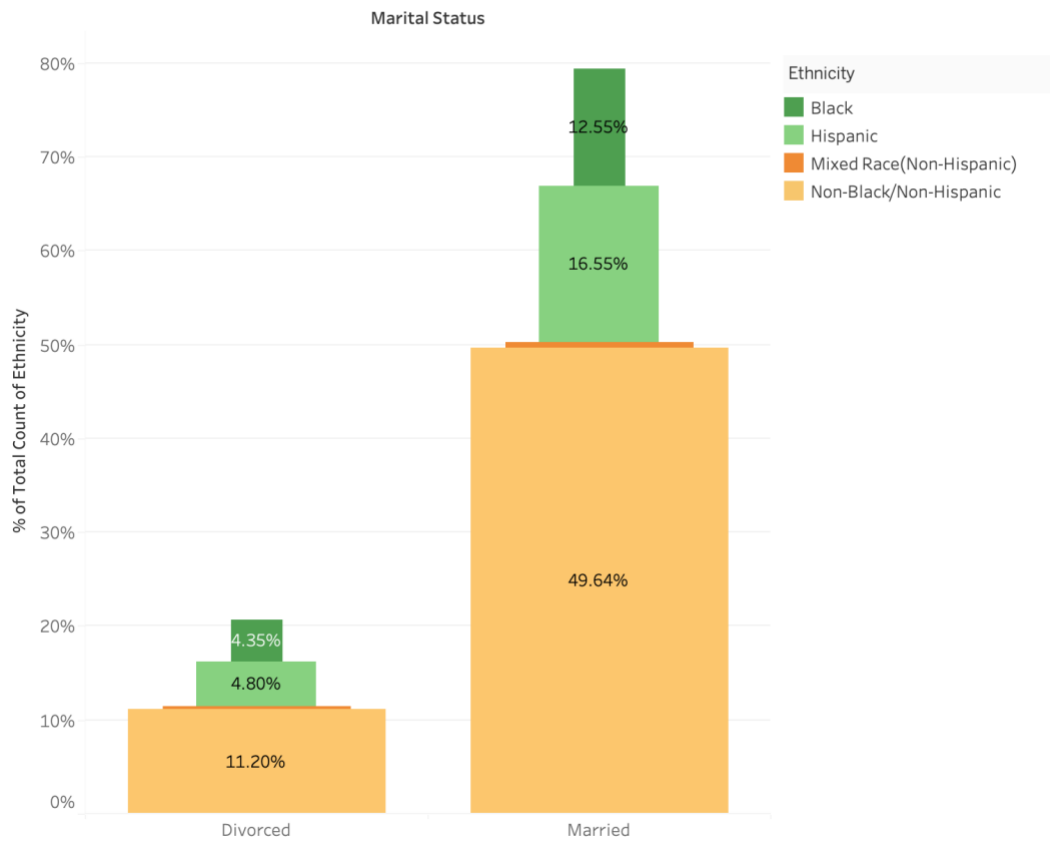


*Figure 7 Anxiety/Easy upset scale 2008*

**What trends or relationships have you found in the data?**

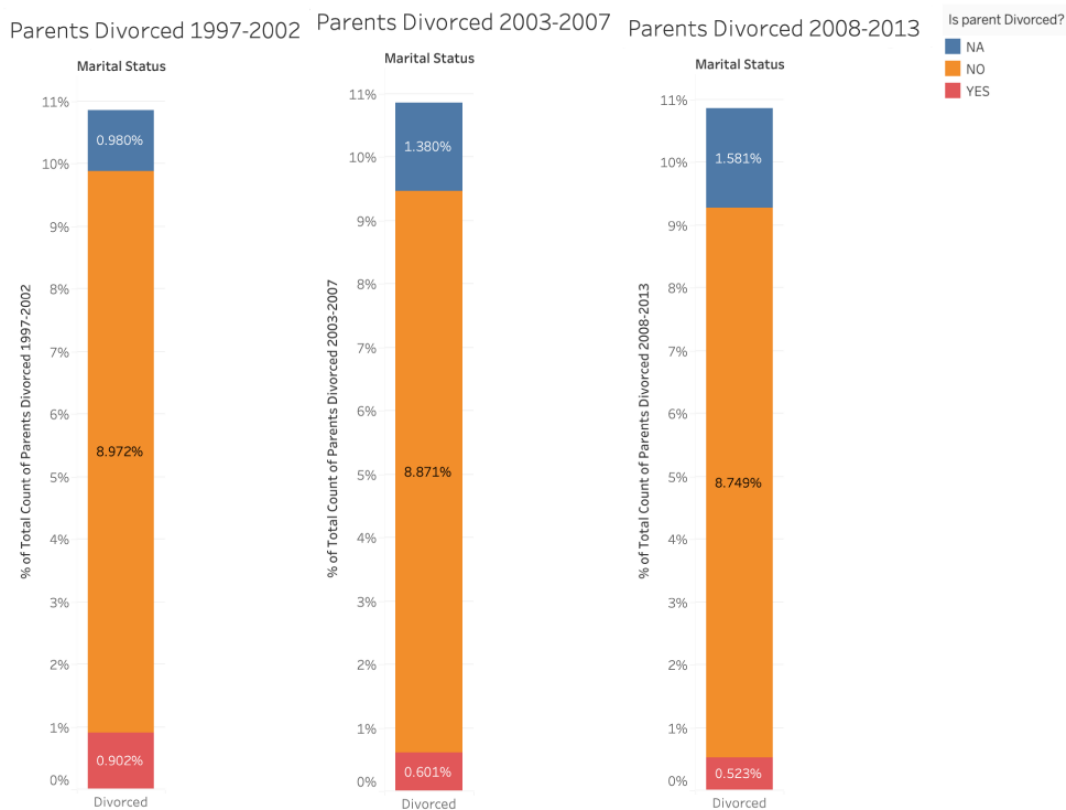
- There is a higher percentage of Non-Black/Non-Hispanic that had been divorce.

## Ethnicity viewed from the perspective of Marital Status



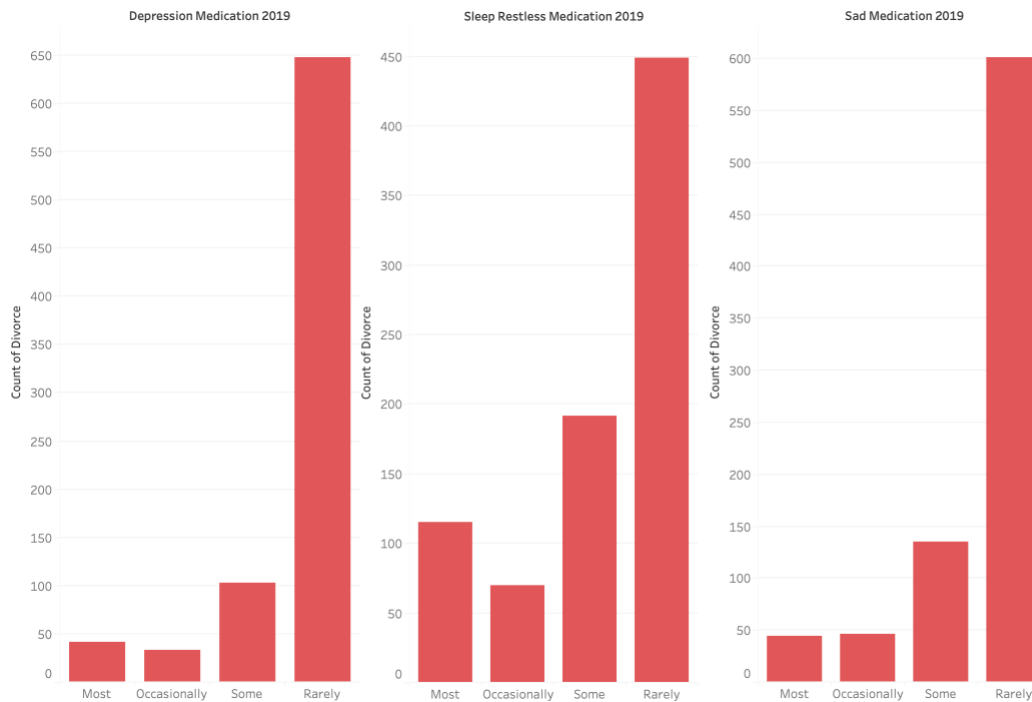
*Figure 8 Ethnicity vs Race*

The data shows that the percentage of divorced people who reported that their parents divorced over the years is decreasing while the percentage of NA is increasing, so we can conclude that divorced people could be hiding information about their parents' divorce status from 1997 to 2013.



*Figure 9 Parent-Divorce Background 1997-2013*

Among divorced people in 2019 survey, the factors that predominate in medication intake are: Firstly, medication intake due sleep restless, secondly because they feel sad and thirdly because they suffer from depression. Due to this we can say that in general, mostly divorce people can directly be affected by sleep restless problems.



*Figure 10 Cause of medications intake 2019*

### How do these insights answer your question or solve the problem?

- Why don't marriages last too long?

To answer that question, a more detailed analysis must be conducted that takes into account more factors, such as marital status over time, compared to other psychological and demographic factors. It was not possible to speculate on a direct factor that led to the divorce due to the complexity of data-correlation and the short deadline set to complete this project.

- What are the factors that lead from marriage to divorce?

Based on the data, it appears that not having an academic degree could lead to divorce, and the higher the degree, the lower the marriage/divorce ratio. We can also say that not having lived with both parents during adolescence could contribute to a future divorce.

- Marital failure or divorce is a consequence of some psychological factors such as depression or anxiety?

It is likely that sleep restless is a consequence of being divorced. Depression and anxiety medication intake are not to relevant between surveyeeyes as sleep restless medication is.

- Are infidelity, physical or psychological abuse or quality of life predominant factors when deciding a divorce?

Infidelity, physical and psychological abuse could not be analyzed due to the availability of data, but quality of life seems to be related to the highest academic degree achieved as long as we understand

quality of life as a factor of personal satisfaction. A person having overcome an academic degree, can feel a high personal satisfaction in addition to obtaining other benefits that contribute to marital stability, making a divorce less likely.

- What are the patterns that stand out in a group of divorced people? Age, religion, years of marriage, time they were in love before getting married?

We cannot establish any relationship between the age of divorced people since the data provided by the survey is from people who were born between 1980-1984 so the time-frame is too short to analyze. Regarding religion, we can say that the majority of divorced people are Roman Catholic, Baptist and Methodist. Years of marriage and dating before marriage were not analyzed in this project due to what was mentioned in the first question. However, there is data available that can be downloaded and re-analyzed with a specific focus on marital status over time and dating.

## SHARE

This document contains the findings and visualizations to answer the questions established by the stakeholder and it is accessible to the audience since it has an consecutive order to be read.

## ACT

### **What is the final conclusion based on this analysis?**

Family education is related with academic education and all psychological factor and a right management of all this factors could decrease the probability of divorce in the future.

### **How the stakeholder may apply this insights?**

The stakeholder could focus her marital therapy sessions by analyzing the patient academic education, parental-divorce background and could repeat rigorously personality and depression tests in addition of paying attention at the medication intake for sleep problems.

### **Are there any next steps you or your stakeholders can take based on your findings?**

There could be a future study focused on determining why more women are divorced than men.

### **Is there additional data you could use to expand on your findings?**

Changes in marital status over time could be analyzed to determine at what point in their lives people are most likely to divorce.