**Successful Marriage**

Tarik Mitchell

Shelton Hartley

Benjamin Dragenice

12/20/2020

CTEC 128

Dr. Miller

**Background**

Marriage, according to Britannica.com, is a legally and socially sanctioned union that is regulated by laws, rules, customs, beliefs, and attitudes that prescribe the rights and duties of the partners and accords status to their possible offspring. For this project our group used data provided by the American Community Survey. Within the data the ACS had an expectation of five groups married, separated, divorced, widowed, and never married. As a group we decided to use data specifically from 2019 although the ACS had data predating back to 1996. With this data we focused on what factors lead to a successful marriage, and if these factors actually make a difference. As a team our brainstorming process was the most important part of making this project. Personality traits were going to be hard to determine why people get married with data, so we focused on external things that could be tracked. We started with four major groups: economics , social, housing, and demographic. Economics covered subtopics such as one's employment status, income , or occupation. For social reasons we thought it would be best to talk about education , status, or previous relationships. As for housing touched points such as if someone had a house, and transportation. Demographics had to be the most important topic out of them all because we knew that the statistics could change for each race. At the end we decide to narrow down our factors down to three out of many for our presentation. Our focus was directed to education level , employment status, and race.

After the brainstorming portion of our project came execution. As a team we had to implement the tools that Professor Miller provided the class with. First we started with the critical thinking in data science aspect. She provided one vital step that would help our team achieve this goal. This was to think in a manner that employs curiosity, creativity, skepticism, analysis, and logic. With this critical tip we had to learn to separate what we knew as a team about marriage to what we could prove. The data had its own limitations on how much it could provide the team with but with limitations on what information it was able to provide it had the right amount of data needed in order to sufficiently complete our goal . The data answered our hypothesis on certain factors The data answers the most important points such as if it answers what we are looking for, and if it tells the “story” that we wanted to present to our audience.

The next lesson we used to help our project come along was about data ingestion. In this lecture Professor Miller used a quote that stated, “If we have data, let’s look at data. If we all have opinions , let go with mine.” from Jim Barksdale. At first we did not understand exactly what Jim or Professor Miller was trying to convey at first, but when we started doing more work with the data we individually began to understand the quote. Essential when doing a big data project with multiple people it is hard for everyone's opinions to be the same in order for us to choose the right direction. Data outside from our own research is unbiased and can not be disputed when it comes to the direction we wanted to go. As a team we had to wrangle out data collectively in order to make the spreadsheet more understandable for us to explain to others. With each file in the ACS’ came a spreadsheet in excel with another document that gives the reader a complete synopsis of what to expect upon examination. Leading to the next lesson Professor Miller taught us.

After thoroughly scrubbing through our excel for hours, our data looked much simpler than before. This was what she called data wrangling. She kept enforcing that you should never believe that the data that we were working with was “perfect” . In order to wrangle our data we used six steps that were provided. Step one was *discovery*. In the discovery step you had to understand what your problem was and know what you needed to find. Step two was *structuring*. In the structuring phase we reconstruct, remove, and combine data. We had to look for inconsistencies like duplicates, missing or incomplete slots, or typos. In step three we had to “*clean*” the data. While cleaning data you had to remove any data that could possibly mess up the results that were intended for the project. Step four was *enriching.* This step was meant to decide if other data is useful enough to achieve our objective. Step five consisted of *validating.* With this process we had to ensure that our data’s quality was sufficient and consistent. The final step, step six, consisted of *publishing*. Publishing was the planning for and preparing the data for future projects.

After our data wrangling crash course with Professor Miller came exploratory data analysis tools. We used data from the American Community Survey with a dataset holding 50,000 records. The dataset is from 2019 and it covers marriage status, education level, racial status, and employment status.

**Data Ingestion**

The first step of the process was the data ingestion for the project. We retrieved our data from the American Community Survey (ACS). The ACS provides information about the nation and its people on a yearly basis. For this project our group decided to use the data from 2019. As time has changed the expectation of people and their choice of spouses have changed. Using the most recent data will reflect on how people choose their spouse today and what is considered important. The subtopics we have decided to focus on are people’s social and economic success. Social success deals with how good a person has of reputation. If a person is able to function in a public setting are they suitable enough for a private setting with a spouse? The economic factor plays apart in the marriage topic as well, because each person must contribute to help keep the couple surviving. If one person is contributing more than the other is it ok and will the marriage continue to be successful? With this data we plan on answering small questions such as these in order to help predict a successful marriage.

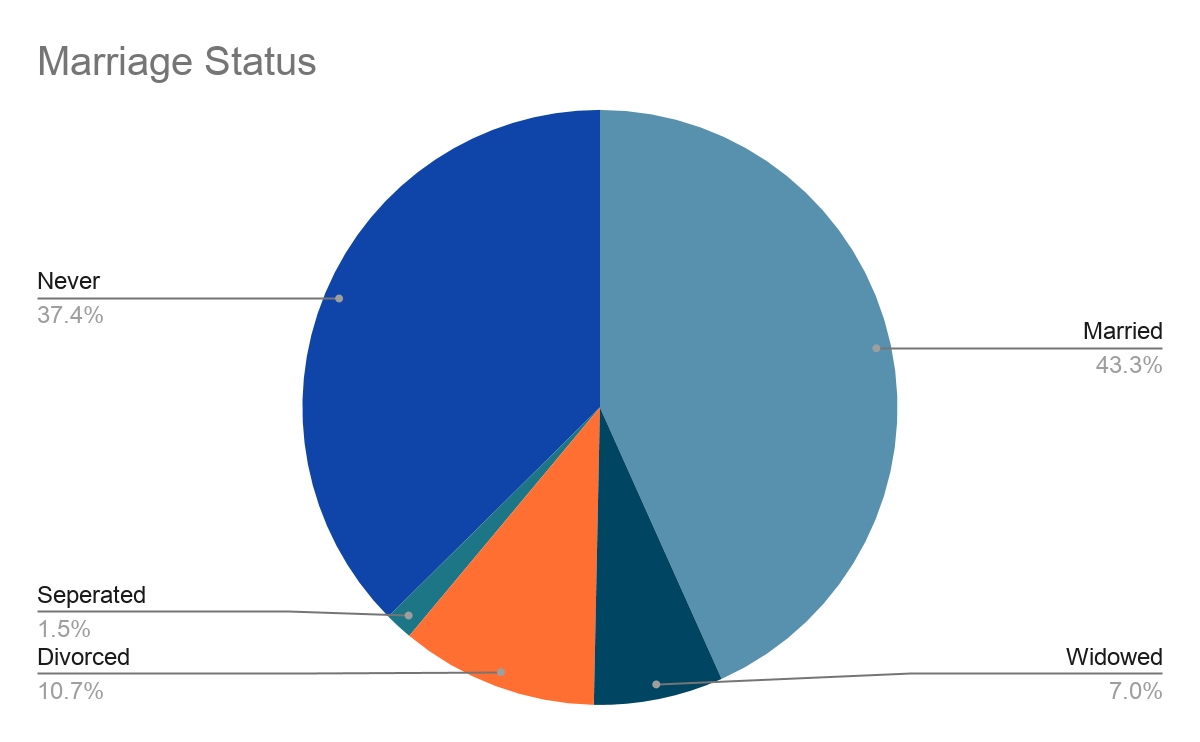
**Data Wrangling**

Our team has chosen to use the data for the population of the entire nation. We chose to do this because in order to get an accurate estimate of what it takes to have a successful marriage the entire nation needs to be surveyed. Specific areas may have higher divorce rates because of the environment and we don’t want that to be a factor when determining the variables that create a successful marriage, we want the variables to depict the nation as a whole. Therefore, we are using the broad data that covers the entire nation. We have also decided to narrow the data down to people the age of 18 and above. This will help to rule out the factor that the couple was obviously too young to get married. For the final analytic data set, the number of instances is 3733 and the number of features is 209.

We had issues with accessing the data in the PUMS data in order for us to download the data and conduct our data wrangling. We were getting error messages when trying to download the data. So we tried using a different browser and the errors stopped and we were able to download the data. It was little difficult in wrangling the data due to the lack of experience with excel but with google and other resources we were able to filter and search for the exact data we needed.

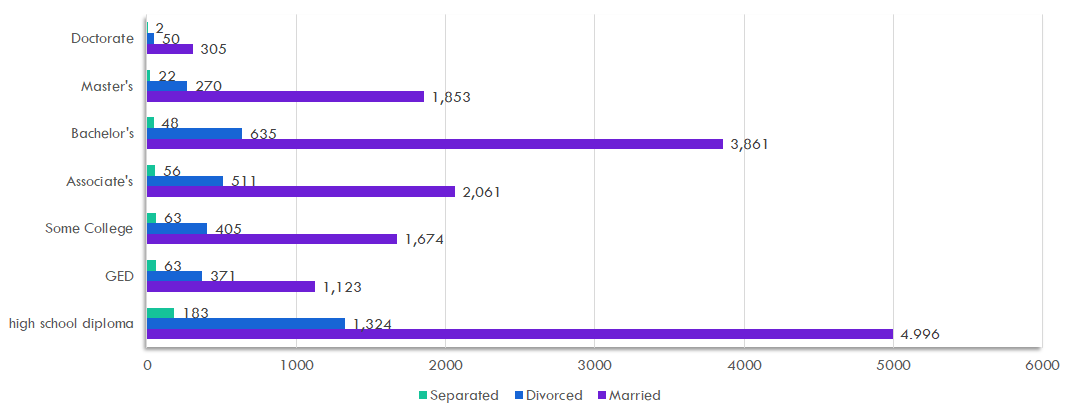
We don’t have to do much feature engineering because the data that we are using is very informative. The ACS data gives us a wide variety of features to choose from and our topic fits well with these features. We will be changing engineering the data to only display people that have a successful marriage. This will help us to be more specific in our approach of analyzing and wrangling the data. After we engineer that we will be able to wrangle the data and find all of the variables we need to complete our research.

**Exploratory Data Analysis (EDA)**



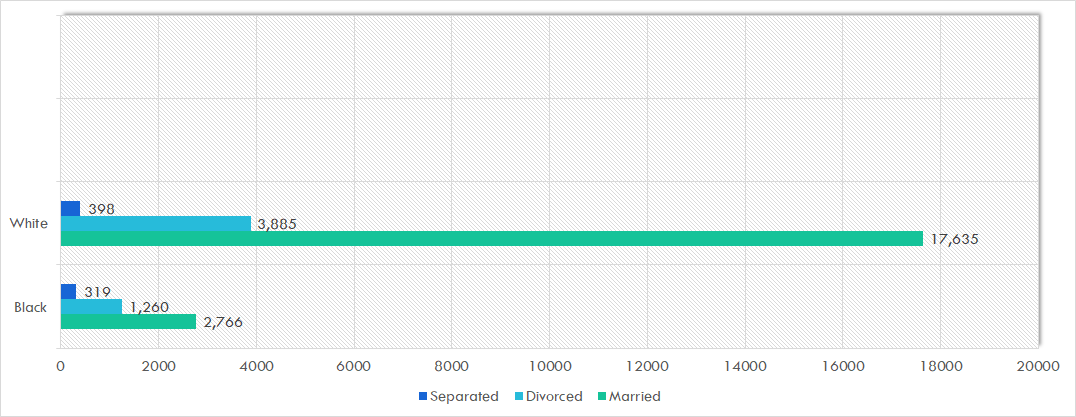
* This pie chart shows the marriage status of the 50,00 people in the dataset.
* 43.3% of people in the dataset were married
* 37.4% of people in the dataset were never married
* 10.7% of people in the dataset were divorced
* 7.0% of people in the dataset were widowed
* 1.5% of people in the dataset were separated

Education Level x Marriage Status



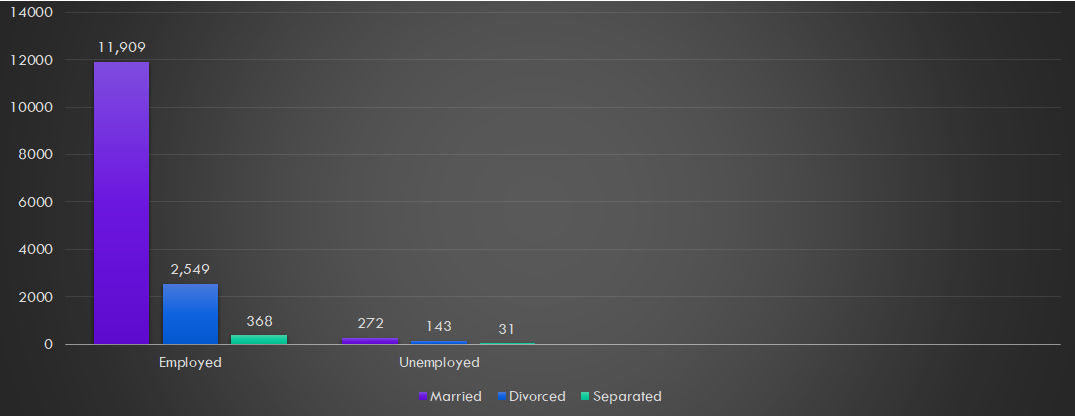
* This horizontal bar graph shows how people education level correlates with their marriage status
* People who are married tend to either have a bachelor’s degree or a high school diploma
* People who are separated tend to have just a high school diploma
* People who are divorced tend to have just a high school diploma

Racial Status x Marriage Status



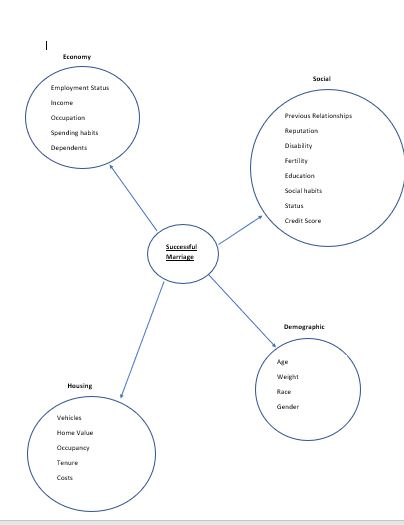
* This horizontal bar graph shows how racial status correlates with marriage status
* White people tend to be married more than black people by a huge differential
* White people tend to be separated more than black people by a small amount
* White people tend to be divorced more than black people by a very tiny amount

Employment Status x Marriage Status



* This vertical bar graph shows how employment status correlates with marriage status
* When someone is employed they’re most likely to be married compared to being separated or divorced
* When someone is unemployed they’re most likely to be married compared to being separated or divorced.
* Employed people are more likely to be married, separated, or divorced than unemployed people

**Data Insights**

There are many factors that can create a successful marriage as you can see in our mind map:

After in depth research we narrowed the factors down to three. These three are employment status, educational attainment and race. The results of our project led us to multiple conclusions. When looking at the effect of employment status on a marriage we have concluded that people who are employed have the highest chances of attaining a successful marriage. When deciding what level of education has the biggest benefit on a marriage we have concluded that people with a master’s degree have the highest chance of attaining a successful marriage. Lastly, when looking at what effect a person’s race has on a marriage we have concluded that white people have the highest chance of attaining a successful marriage

**Recommendations**

As a group we would run into a couple problems. The researching aspect of this project was fine. When we started looking at the data the transcript provided did not answer the majority of the questions we had. We ended up taking another couple of hours to research what each group menat in order to fully understand the big data. Also when it the transcripts provided did not fully go into each race within each topic. With those few minor tweaks the understanding and overall timing of the project could have been a little more simpler and time efficient.

**References**

Herston, K. (2016, September 30). Biggest Factor in Divorce Is Husband's Employment Status, Says Harvard Study. Retrieved December 20, 2020, from <https://herstontennesseefamilylaw.com/2016/10/05/biggest-factor-in-divorce-is-husbands-employment-status-says-harvard-study/>

Crowley, J. C. (2020, November 3). What Are Common Reasons for Divorce? Survive Divorce. <https://www.survivedivorce.com/common-reasons-for-divorce#8>

US Census Bureau. (2020a, October 16). About the Survey. The United States Census Bureau. <https://www.census.gov/programs-surveys/acs/about.html>

Marriage. Retrieved December 20, 2020, from https://www.britannica.com/topic/marriage