Predicting Divorce in the United States

Karan Rao
PES University
Bangalore, India 560085
Email: karanrao3@gmail.com

Nitish J. Makam
PES University
Bangalore, India 560085
Email: nitishmakam@gmail.com

***Abstract*—High divorce rates are an issue in many countries, perhaps most famously in the United States of America. According to the American Psychological Association, 40 to 50 percent of marriages end in divorce. As marriage is arguably the basis of civilization, this is a serious problem. Another issue is the psychological effect of divorce on parents and children, which persists for years with far-reaching consequences. The purpose of this analysis is to detect a relationship between dissolving marriages and possible demographic and economic causes, and also the types of people most and least likely to get divorced. We also attempt to determine the statistics for different age groups, and various minor questions such as whether there is a positive correlation between income and the choice to be separated rather than divorced. We will be using predictive methods such as decision trees, naive Bayes and logistic regression on US census data. The objective is to determine what leads to divorce at the level of the individual and to increase the number of successful marriages by making people aware of these causes.**

***Index Terms*—divorce, data analysis, prediction, census**

I. Introduction

a) : Marriage (or monogamy, to be precise) is arguably the basis of civilization. Monogamy is important because if we regressed, in an extreme example, to a polygamous society, we would often be under conflict. As men are allowed a number of wives, they spend time fighting among themselves and invading neighboring tribes for more. This creates an environment that is not conducive to collaboration because men begin to trust each other less. Monogamy, on the other hand, is egalitarian and consequently peaceful because everyone is given a chance to mate. [1]

b) : This point is relevant because the rates of divorce remain high in the US even though they have been dropping since the 1980s, and we are slowly moving away from monogamy, or at least two-parent households that are a stable environment for children. Though polygamy has not been accepted in the West, there has been ample promotion of another non-monogamous lifestyle: the single parent household. William Tucker argues that it is a sort of "state polygamy", a societal phenomenon involving the increase of single women raising kids without fathers while relying on the government to meet the children's needs. [1]

c) : The aforementioned two-parent household is the best for children in terms of education, development and psychology. Though correlation does not equal causation, there are undeniable associations.

Divorce has lasting repercussions for the children that have to live through it. In the United States, children with divorced parents are three times likelier to be expelled or become teen parents, six times as apt to be impoverished, and twelve times more likely to go to jail when compared to their peers from intact homes, according to the Heritage Foundation. Another Heritage study reports that children from fatherless homes are 15 times more likely to be jailed, 33 times more likely to be physically abused, and 73 times more likely to be killed.

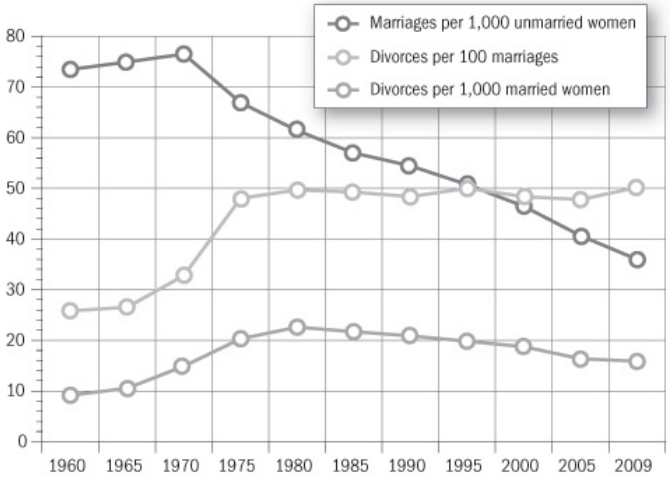


Fig. 1. Marriage and divorce rates

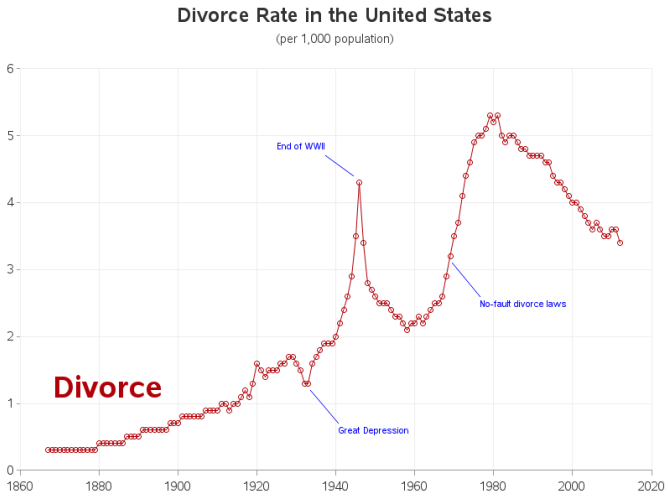


Fig. 2. Divorce rates explained

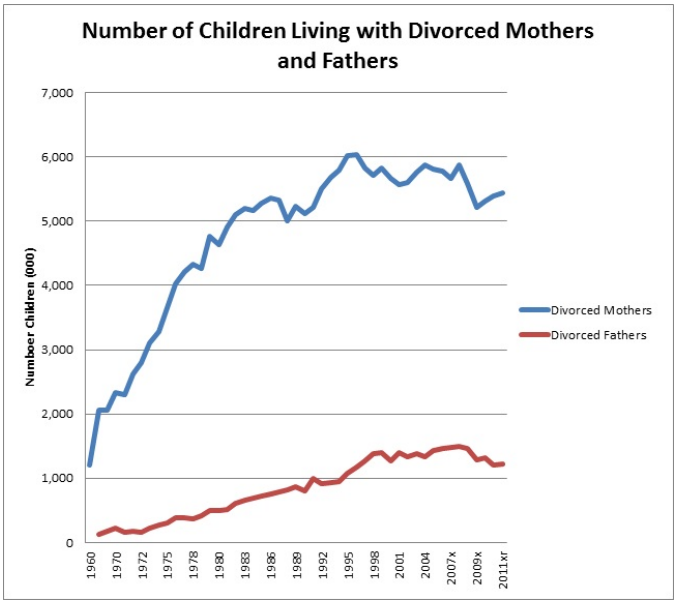
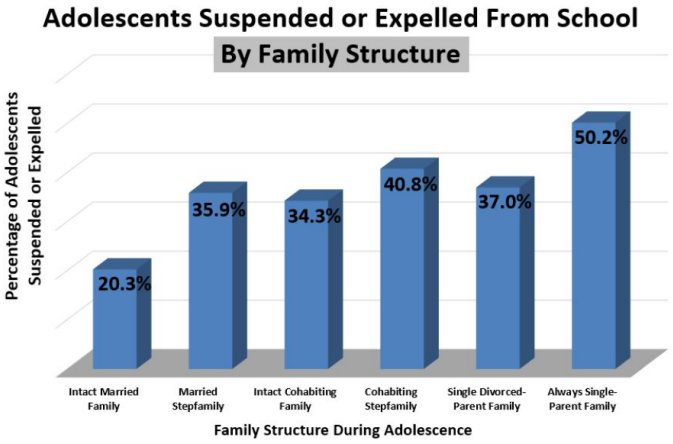


Fig. 3. Custody



Source: National Longitudinal Study of Adolescent Health. Adolescents Grades 7-12.

Fig. 4. Expulsion and suspension

Children in two-parent households are 44 percent less likely to be physically abused, 47 percent less likely to suffer physical neglect, 43 percent less likely to suffer emotional neglect, and 55 percent less likely to suffer some form of child abuse. Furthermore, a 2010 Family Research Council study finds that children of intact, married parents live five years longer than those whose parents are divorced. [11]

d) : The aforementioned fatherless homes are common because in the event of divorce, custody is usually awarded to the mother. In 2014, about five of every six custodial parents were mothers (82.5 percent) and one of every six were fathers (17.5 percent), proportions that were not statistically different from those in 1994 (from which our census data is drawn). [2]

e) : Naturally, divorce is devastating for the adults themselves. Divorcees report emotional and psychological issues, loneliness, the loss of confidence, financial trouble and in some cases the loss of relationships with friends and relatives. Fathers who do not get custody feel that they are no longer centrally important to their children. [3] [4] Also, prospects for remarriage are bleak. The divorce rates for second and third marriages are 67% and 74% respectively. [5]

f) : The objective is to determine what leads to divorce and to increase the number of successful marriages by making people aware of these causes.

g) : Predictors include age at marriage, religious compatibility, income, education, race, culture, urbanization, the number of women in the workforce and also the length of time for which no-fault divorce has been allowed. The introduction of no-fault divorce (first in California in 1969), where there is no need to show that your spouse breached the marriage contract, increased the US divorce rate by 20-25% according to a 1989 study by Justec Research in Virginia. Whether no-fault divorce should be allowed is beyond the scope of this project, but it has both pros and cons. One of the cons is that it may be used as a solution to trivial problems that can be settled with a little work.

h) : Another speculative cause is that laws about the distribution of assets after a divorce may motivate it.

i) : After this increase, the divorce rate is in the neighborhood of 40-50%. However, divorce demography can be misleading if not defined clearly. Some methods are crude and compare the number of marriages in a year to the number of divorces in a year (which could be from older marriages), while some are longitudinal studies that track the same marriages over a long period of time.

II. Literature Survey

A. Assumptions

a) : Others have made assumptions while working with the same dataset. One is that Capital Gain and Capital Loss may be discarded, as only a few rows have values other than 0; or they can be scaled logarithmically. They have also ignored fnlwgt, which is the final sampling weight of each record. They have assumed that conditions are similar in certain categories of countries or jobs so that we can group them together instead of dealing with so many levels. Levels can be grouped if they are related and low in frequency. [6] [7]

b) : Following the above, we may need to combine Divorced and Separated while cleaning the data, since both imply some kind of rift. The others can be grouped into Not-divorced.

c) : Some researchers started out with assumptions about the coefficients of each predictor. In a 1995 study, researchers concluded that the income level played a significant role in divorce rates, exhibiting a positive correlation. [9] But as many marital fights stem from the lack of money, the coefficient is ambiguous. [10] One would think that education would lead one to weigh decisions carefully, resulting in a positive coefficient. However, income usually rises with education. If

this were true, based on the theoretical backing for the income variable [9], one should expect divorce rates to rise. Thus, the education coefficient should have an ambiguous sign. [10] Urban areas tend to have less strict social codes and norms and religious backings. These effects can lead to an increase in divorce. Also, the heterogeneity of metropolitan areas is much higher, so there are more "fish in the sea" for people of urban areas. The contrary is true in rural areas. [10]

B. Approaches

a) : In The Determinants of Divorce Rate: An Econometric Study [10], the number of years since no-fault divorce was enacted, median income, percentage of women 16+ who work, percentage in poverty, proportion of the population in metropolitan areas, education and religion were considered to be the most common predictors of divorce.
Methods like decision trees, naive Bayes, logistic regression and SVMs have been used for predicting variables in the same dataset. [12] [13]

b) : The distribution of the percentage of people with an income greater than or equal to $50,000 across different categories of a parameter (like for each different marital status) was displayed for a better understanding of the data. [13]

c) : Graphs of the variable to be predicted can be plotted against every other parameter in the dataset. [12]

d) : The data can also be cleaned and simplified by making groups of similar levels. [6] [7]

C. Results

a) : The most significant variable in both equations that were formulated was religion according to the measure of the t-test. Changing the education rate would produce the largest effect on the divorce rate because the education coefficient is larger than the other independent variables. [10]
The following observations are from projects that used the same dataset to predict income, which was its intended use.

b) : It was found that the NBTree classifier was the best performer (lowest error) out of 17 algorithms, a hybrid of naive Bayes and decision trees. [12]

c) : One interesting finding was that using a non-machine learning model outperforms all machine learning models if a certain process based on the properties on the dataset is used.
If an individual is a white male wth a college degree, then we predict that the individuals income is greater than $50K. We learn from exploratory analysis that the data is biased toward white males with college degrees having an income greater than $50,000 per year. [13]

d) : By estimating the relationships among marriage, divorce, work effort, and wage rates, researchers found that being married and having high earnings reinforce each other over time. Others looked at the how income affects the marriage and divorce decisions of young Americans; they found that high earnings capacity increases the probability of marriage and decreases the probability of divorce for young men, but decreases the probability of marriage for young women and has no effect on the likelihood of divorce. [14]

e) : For an example of results based on race: it was found that compared to both white and Hispanic women, black women marry later in life, are less likely to marry at all, and have higher rates of marital instability. [8]

D. Lacunae

Previous researchers have not used parameters such as hours worked per week, native country, and success of investments. They have done time-series analysis of divorce rates but not prediction of divorce at the individual level.

III. Proposed Problem Statement

Our primary aim is to predict the likelihood of an individual being divorced using data that concerns that particular individual. It does not concern couples and compatibility between different types of people; we are only making predictions at the level of the individual. Additionally, we are looking at the likelihood of separation compared to divorce for different income groups and races, the divorce statistics for Indians in the United States, and the relationships between marital status and every other variable in the dataset.

IV. Proposed System

See figure 5 for the block diagram of the project.

V. The Components of the System

A. Cleaning

a) : The dataset included a large number of countries for the native country parameter. As we were not focusing on individual countries except for India, we grouped the countries into zones based on economics, geography, and political organization. For example, Euro 1 included the more affluent European countries and Euro 2 included the less affluent ones.

b) : As for work classes, they were cleaned to private, federal government, self-employed and not working (among others) and occupations were reduced to blue collar and white collar (among others). Some instances of our dataset had marital status reduced simply to divorced and not divorced (with 'divorced' including separation because they often involve the same issues) and other times, the entire range of possibilities was used. The training and test datasets (separate and provided by UCI) were both cleaned in this manner.

B. Exploratory Analysis

With a correlation plot, we established that the numerical attributes of age, education level and hours worked per week were not correlated and were not influencing each other. We plotted box plots, scatterplots, stacked bar charts, pie charts, and histograms to assess the data and discover relationships. We also decided which variables to drop, such as capital gain and capital loss (the success of investments) as they were mostly blank for the dataset. The plots revealed that there was no noticeable correlation between occupation or work class and marital status, so they could be ignored. We expected hours worked per week to be correlated with divorce, but the stacked bar plot proved otherwise. Interestingly, we found that there were more than twice as many adults who had never married than those who had been divorced.

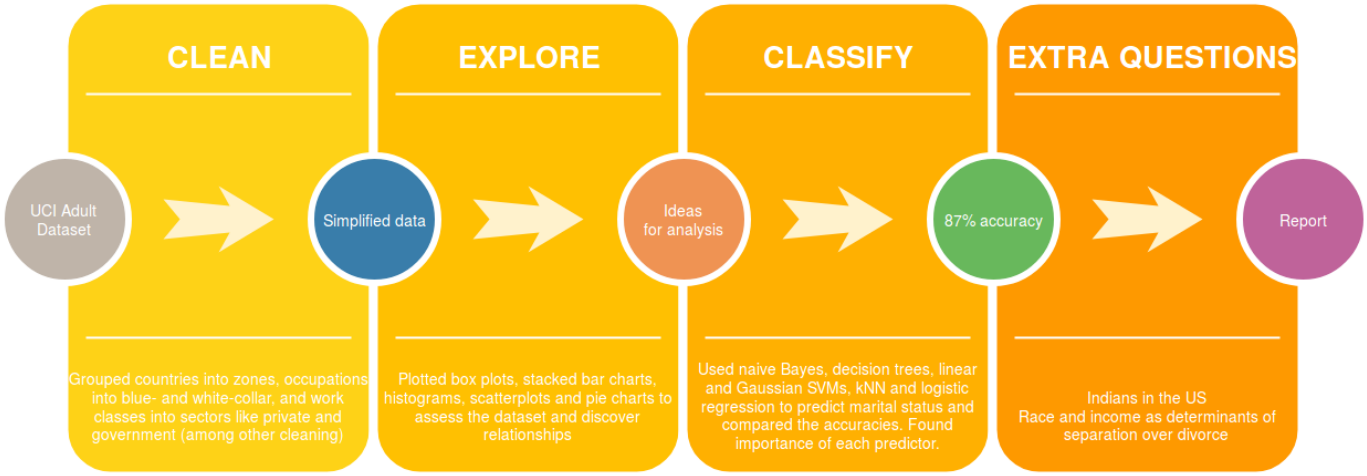


Fig. 5. Block diagram

C. Classification

We used naive Bayes, k-nearest neighbours, linear and Gaussian support vector machines (SVMs), logistic regression, and J48 decision trees and compared the results. Our metrics for comparison were ROC curves, accuracy and f-score. Naive Bayes with some columns removed had the best ROC curve and had the second best accuracy (86%). The Gaussian SVM was the most accurate at 87% and the decision tree had a notable accuracy of 84.7%.

D. Other Questions

We intended to check if people in the military were more likely to be divorced, but there were only nine of them in the dataset and none were divorced. But, interestingly, six out of those nine had never been married.

There were only 100 Indians in the dataset, out of which two were divorced and two were separated. This is a low rate compared to the other native countries. Again, this was not much to go on, but it followed some of the patterns we had found earlier. The ones with lower incomes were separated instead of divorced. The youngest one to be separated/divorced had an 11th grade education.

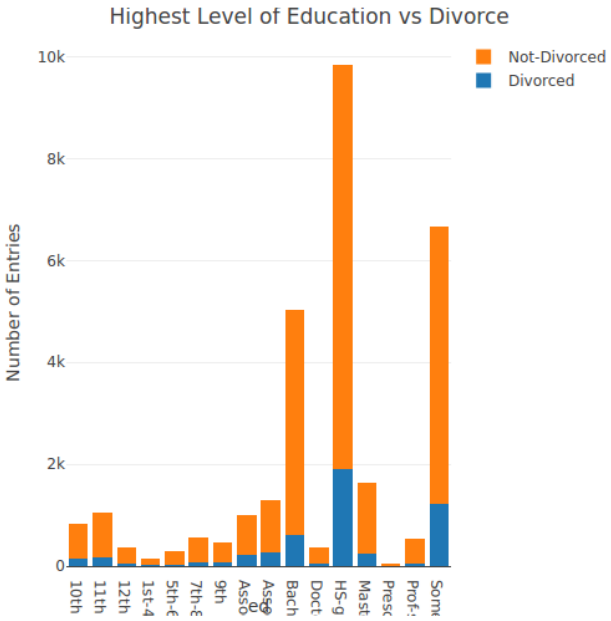


Fig. 6. Education and Divorce

VI. Experiments and Results

Our plot of education against divorce showed that the percentage of divorcees dropped sharply at the Bachelor's degree bar and remained low at higher education levels. (Figure 6)

We observed that there were more female divorcees than male at virtually every age, with very large differences from ages 35-50. (Figure 7)

We saw that the largest ratio of divorced:not divorced occurred in the 41-50 age bin. There seemed to be no divorcees in the 81-90 bin and few beyond 61, confirming our earlier conjecture that a longer a marriage lasts after 40, the stronger it is. (Figure 8)

24.5% of the Black population is divorced, compared to 16.4% of the White population. (Figure 9)

We saw a much lower divorce rate for incomes above $50,000 a year. (Figure 10)

We found that many more Black people were separated as compared to White people, in terms of percentage. One possible reason for this is financial. Divorces are expensive and a ten-year marriage (counting separation) qualifies divorcees for Social Security benefits, which is why some people only divorce after they cross ten years and stay separated before that. There are also health insurance and tax benefits which will help those with low incomes. (Figure 11)

Confirming this, we found that the ratio of separated:divorced

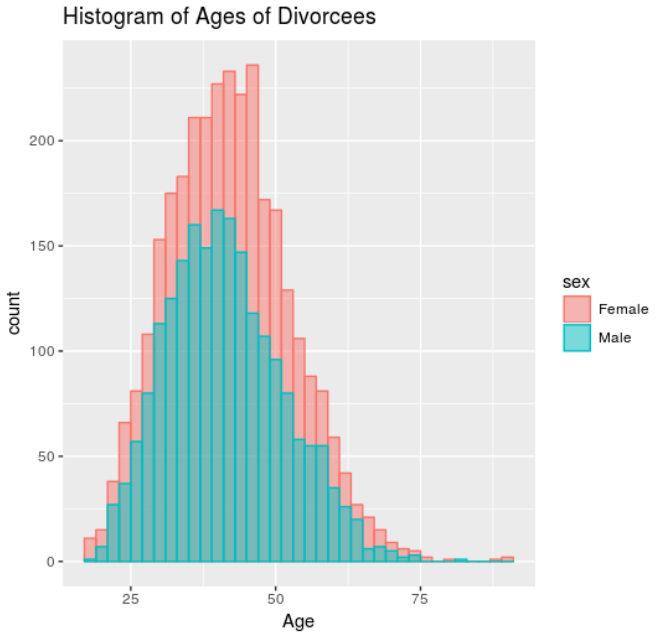


Fig. 7. Gender, age and divorce

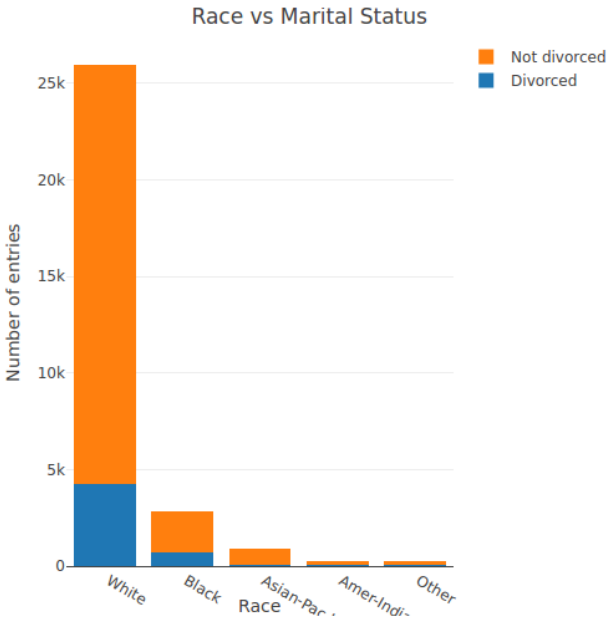


Fig. 9. Race and divorce

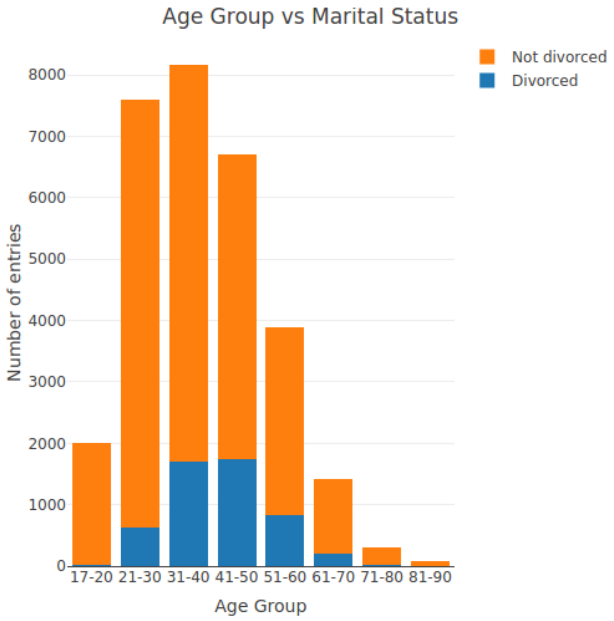


Fig. 8. Age and Divorce

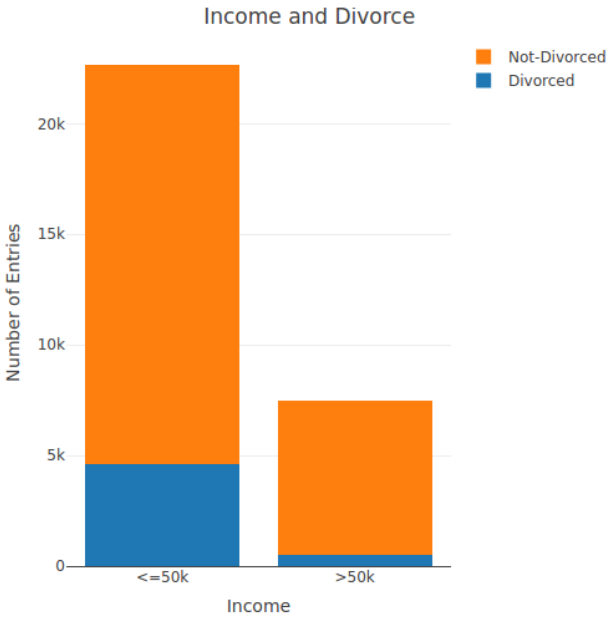


Fig. 10. Income and divorce

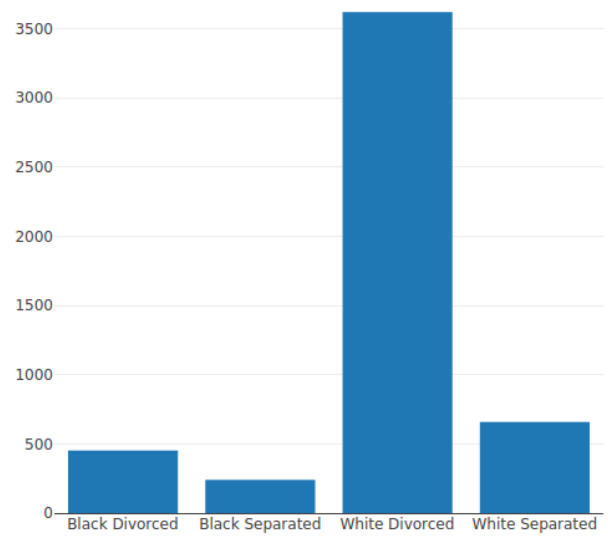


Fig. 11. Race, separation and divorce

was 232:1000 for incomes below $50,000 a year while it was 146:1000 for incomes higher than $50,000.
Using the relaimpo package, we found the relative importance of predictors and learned that age and income were the biggest determinants.

VII. Conclusion

Some of the work done here confirms theories that have been proposed earlier, such as the fact that education means divorce is less likely. Some of it disagrees with theories: we found that blue-collar workers had the lowest divorce rate but they are supposed to have one of the highest. In the end, most of it was in accord with the real world. Divorce is a problem that we know the predictors of and it is something we can bring down if we make careful decisions. Marriage is arguably the foundation of civilization and we must try to make it as successful as possible.

VIII. Roles

The work was not cleanly divided; we collaborated on most of the project.
Karan: did more of the literature survey and planning than plots and models.
Nitish: did more plots and models than paper-reading.
That's about it.

References

[1] William Tucker *Marriage and Civilisation: How Monogamy Made Us Human* 2014
[2] Timothy Grall *Custodial Mothers and Fathers and Their Child Support, https://www.census.gov/content/dam/Census/library/publications/2016/demo/P60-255.pdf* 2013
[3] Elizabeth Marquardt *https://thrivingcouples.com/2011/07/20/the-effects-of-divorce-on-children-and-parents/*
[4] ALFI *http://alfi.org.ph/2015/04/effects-of-divorce-on-parents/*
[5] *www.wevorce.com/blog/why-do-second-marriages-fail/*
[6] *http://blog.pangyanhan.com/posts/2017-02-15-analysis-of-the-adult-data-set-from-uci-machine-learning-repository.ipynb.html*
[7] *http://scg.sdsu.edu/dataset-adult_r/*
[8] R. Kelly Raley et al. *The Growing Racial and Ethnic Divide in US Marriage Patterns*
[9] Paul A. Nakonezny et al. *The Effect of No-Fault Divorce Law on the Divorce Rate Across the 50 States and Its Relation to Income, Education, and Religiosity*
[10] Rachel N. Ruth *The Determinants of Divorce Rate: An Econometric Study*
[11] Patrick Fagan *http://www.heritage.org/testimony/the-impact-marriage-and-divorce-children* 2004
[12] Chet Lemon et al. *Predicting if income exceeds $50,000 per year based on 1994 US Census Data with Simple Classification Techniques*
[13] Tracy Nham *Classifying Income from 1994 Census Data*
[14] *https://www.bls.gov/opub/mlr/2013/article/marriage-and-divorce-patterns-by-gender-race-and-educational-attainment.htm*