

# **Project Proposal: Predicting Divorce for Young Adults**

**Celeste Chen and Sarah Unbehaun, March 29, 2017**

## **Introduction**

In the past few decades, American views on love, sex, marriage, and divorce have undergone drastic changes. The sexual freedom of the 1960s gave way to greater conservatism with the end of the Vietnam war and the growing politicization of sex. A 2014 study found that baby boomers who had been part of the free-love generation not only had (1) “unprecedented divorce incidence in 1980 and 1990 when they were in their 20s and 30s” but that (2) this pattern continued well into the present. Researchers noted that the “baby Boom generation was responsible for the extraordinary rise in marital instability after 1970. They are now middle-aged, but their pattern of high marital instability continues.”

Within the past decade, perceptions of sex, marriage, and love have transformed again. With fewer Americans getting married each year, divorce rates are dropping but continue to be high. Are millennials headed towards a similar path when they do decide to tie the knot? With the new legality of same-sex marriage and changing attitudes about marriage in general, has the concept of marriage been so fundamentally changed that its participants may be more or less primed for eventual divorce? As future generations build, participate in, and redefine relationships, what sorts of beliefs and life experiences can we use to predict whether or not a couple will divorce?

## **Hypothesis and Research Questions**

We expect that certain social and political views affect the health and length of a marriage. Those characteristics will allow us to predict who is most likely to get a divorce or remain married. Research question: What personal characteristics or views are most highly associated with divorce?

## **Data Experiment Design**

For married individuals ages 25 to 45, we will predict the probability of getting a divorce within the first seven years of marriage based on the respondent’s social views and history, include attitudes about pornography, fundamentalism of religious views, and instances of premarital sex or extra-marital sex. Our analysis will also use standard factors such as education, race, religion, political views and having children.

## **Data, Processing and Analytical Methods**

We will use data from the General Social Survey from years 1990 to 2014, approximately a generational cohort. The data is available as a cumulative file of cross-sectional samples from all years 1972 to 2014, but we are limiting our selection based on the availability of specific variables. The data should be fairly complete for the years selected and is already formatted. We may need to use weights in our analysis in accordance with the accompanying codebook. Using the train-validate-test method, we will separate our available data in to a 70-15-15 split in order to train our algorithm and assess its accuracy. We will test three methods of analyzing the data for predictions: (1) logistic regression, (2) decision trees, and (3) random forest. Based on the accuracy results of each method, we will determine the best model for prediction.

## **Ethical Considerations**

This could affect real-life relationships, depending on how the results are used, especially if marriage counselors included it as a tool in their work. Unfortunately, someone might be influenced to consider a divorce who wouldn’t have or a couple might not get married who had thought about it, if the prediction indicates that their marriage is unlikely to last. On the other hand, this could provide support for dating services that cater to those looking for long-term relationships by identifying personal factors that contribute to marriage longevity.