

# Project Proposal

Danny Klinenberg 7626054

Michael Topper 95788708

March 09, 2020

## Research Question

Using a large dataset from the IPUMS CPS, we propose to answer the following questions:

- 1) *Can we predict who gets a divorce based on characteristics from a sample?* Since (Becker 1974), economists have been interested in why people get married, and what factors influence divorce. Therefore, our question will be if we can successfully create a prediction model that predicts divorce for U.S. citizens.
- 2) *Do younger people divorce more than older people?* Economic theory tells us that extremely young adults have a lower cost of divorce than older adults (Brassiolo 2016). The mechanism behind this result is that younger people (without children) have a greater likelihood of finding a partner if they are divorced. For example, if a 25 year old divorces, they have a much larger pool of potential people to remarry than if a 55 year old divorces. Hence, it is much less costly for the young adult to divorce than the older adult. Empirical evidence on this topic is limited, yet we can explore this mechanism using our dataset.
- 3) *Does having young children impact divorce decisions?* Similar argument as question (2). The costs of getting divorces (theoretically) rise when a young child is present and therefore we can expect less divorces for couples with young children. Is this statement true? Does our data follow the model?
- 4) *Does religion influence divorce decisions?* Religion contributes to individual's values. For instance, a person who is a devout Christian may be less inclined to get a divorce than an Atheist because their religion is opposed to divorce. Once again, this can be attributed to economic theory: the cost of divorce for a Christian may be higher than the cost of divorce for an Atheist. However, religions are likely to have heterogeneous effects across divorce. In particular, some religions have more lenient views on divorce than others. We will be able to disentangle these effects with our data.
- 5) *Do the predictors of divorce change over time?* In other words, are the factors that influence divorce dynamic? According to the literature, the aggregate risk of marriage ending in divorce has declined from the peak that occurred in the 1980s (Cherlin 2010). We seek to analyze whether there were different factors that were a heavy influence of divorce in past years that no longer carry weight in today's society. Social norms change rapidly over time, and we argue that what predicts divorce today is different than what predicted divorce a decade ago. Many scholars have tried to explain the dynamic divorce rates with a variety of explanations. A study by (Cherlin 2004) found that social attitudes towards divorce have changed drastically over time. On the other hand, a study by (Ruggles 1997) suggests that rising female labor force participation may have caused fluctuations in divorce. We can explicitly test these theories by forming predictive models using a base year and then testing the model on a variety of different years. We can then plot the success rate over time to get an idea of the dynamics.

## Data

Our main dataset will be the Integrated Public Use Microdata Series (IPUMS) Current Population Survey (CPS) for 2009 and 2019. We use the subset of the data called Annual Social and Economic Supplement (ASEC) for CPS. The survey asks members of households a variety of social and economic questions about the previous year. The Census Bureau randomly selects about 60,000 households a year. There are 388022 observations and 22 features. The features are:

Table 1: Features

Name
Relationship to household head
Age
Sex
Race
Adult civilian, armed forces, or child
Asian subgroup
Veteran status
Number of own children in household
Number of own children under age 5 in hh
Birthplace
Foreign birthplace or parentage
Hispanic origin
Employment status
Labor force status
Occupation
Educational attainment recode
Occupation last year
Industry last year
Worked last year
Total family income
Total personal income
Wage and salary income

The variable of interest, *Marital Status* is classified into 9 categories:

Table 2: Classification of Marital Status

Value	Label
1	Married, spouse present
2	Married, spouse absent
3	Separated
4	Divorced
5	Widowed
6	Never married/single
7	Widowed or Divorced
9	NIU

We will convert this variable into a binary outcome defined as divorced if an individual is **Divorced** or **Widowed or Divorced**. Defining divorce as so, 7.26% of the population is divorced. While this appears problematic at first, this equates to 28163 individuals being classified as divorced.

## Analysis Plan

We will preface this section by warning the reader that this analysis plan is extremely tentative and subject to change. However, we plan on using supervised learning techniques including k-nearest neighbors, decision

trees, logistic regression, and neural networks to answer our research questions. We plan to use cross validation to find optimal number of neighbors in k-nearest neighbors and the optimal size of our tree when using decision trees. Due to the inaccuracy of decision trees, we will supplement this with random forests. We plan to compare the accuracy of random forests with the accuracy of our decision trees. Exploratory analysis will be performed with a variety of different graphs to find “good” predictors for our outcomes of interest. In addition, we will provide summary statistics to give a broad picture of the data that we are working with (i.e. average health score, most frequent health violations etc.). For model validation, we will randomly split our dataset into two sections: a training set and a test set. We will train the model on the training data and then test it on our testing data.

Additionally, to answer question (5), we plan to create a model based on a “base” year. This model will be tested and trained on the same year to get predictions on divorce for that particular year. After doing this, we will test the model on multiple years to find how the model predicts *over time*. Furthermore, if time permits, we can do this with a variety of base years and see how each of prediction models vary over time. Since decision trees give a nice representation of which covariates are weighted the most, we can analyze the trees within each of these base years to see what are the most important factors of divorce.

## References

- Becker, Gary S. 1974. “A Theory of Marriage: Part Ii.” *Journal of Political Economy* 82 (2, Part 2): S11–S26.
- Brassiolo, Pablo. 2016. “Domestic Violence and Divorce Law: When Divorce Threats Become Credible.” *Journal of Labor Economics* 34 (2): 443–77.
- Cherlin, Andrew J. 2004. “The Deinstitutionalization of American Marriage.” *Journal of Marriage and Family* 66 (4): 848–61.
- . 2010. “Demographic Trends in the United States: A Review of Research in the 2000s.” *Journal of Marriage and Family* 72 (3): 403–19.
- Ruggles, Steven. 1997. “The Rise of Divorce and Separation in the United States, 1880–1990.” *Demography* 34 (4): 455–66.