FinalProjectStep3

Santana, Monica

2023-10-24

Introduction

The research topic I plan on solving using data science is on divorce. Why would someone be interested in this? Well, marriage is a technology that has been around since 2000 B.C. It has been around since then and although there are many people for it, there are also many people against being married. Are they entirely wrong, when statistically marriages end up in divorce 54% of the time. Why are those divorce rates so high? What are we doing now more than ever, that are causing divorce rates to just continue to go up. Maybe marriage is an outdated technology or maybe there are more outlets for marriages to fall apart easier, such as social media. This data would interest those who potentially want to get married, and those who do not believe in it because statisically marriages are more likely to fail then survive. Divorce is not something that is easily dealt with or easy to go through especially with children involved. With data science, I can look at datasets involving divorces and the percentages of different issues or reasons why people fail to overcome the divorce statistic.

Research Questions

1. Does economic similarities or disimilarities play a role in a couple divorcing?

2. Does cultural play a role in a failed marriage, if they are culturally the same or not?

3. What factors affect marriages the most to the point of divorce?

4. Does age play a factor in a divorce, or how old someone gets married?

5. Does having common interests help or are opposites more likely to stay married?

6. If there are children involved from a previous relationship, does that have a role in divorce?

7. When a couple divorces, how willing were they to be married in the first place?

8. Does religion play an effect in whether couples tend to not divorce as often?

9. Does having a good relationship with the spouse's family make a difference in a marriage?

10. Those who have mental health disorders, will that have a big enough affect to divorce?

Approach

With these questions, it will directly address the problem of a high divorce rate by seeing which one of these factors contribute to failed marriages the most. Also can help struggling couples in relationships see if the reason they think their marriage is falling apart is not as important as it may seem. Although every relationship is different, and every person is unique, this data will show reuccurring factors that play a role in divorce since it is so common, there are bound to be patterns. This data can reinforce why marriage is not something to just do and throw around, rather a careful decision to avoid future problems that lead to divorce. Using these research questions, I plan on using a dataset that have asked couples the same questions I chose for research, or similiar and compare the different variables to see whether or not they are big factors in divorce. With the datasets I can compare visually using histograms and scatterplots to plot the different variables. The data will statistically confirm the divorce rate of 54% but also show insight on how NOT to fall into that statistic by either not getting married at all or making sure the person you are married with is the real deal.

Discuss how your proposed approach will address (fully or partially) this problem

My proposed approach will address this problem due to answering the research questions with insight on why divorce rates are so high, and if marriage is truly the right decision. My approach will definitely partially address this problem as I am sure there are several other factors that contribute to divorces and unsuccessful marriages. Seeing the different factors through comparison graphs will help understand the reasons behind divorce and how it is much greater than just "falling out of love".

Data

https://www.kaggle.com/datasets/hosseinmousavi/marriage-and-divorce-dataset

I plan on using this dataset from Kaggle, in csv form that was collected in 2017. The original purpose of the dataset is to target divorce probability using several inputs. The inputs(variables) are 30, and include age gap, education, economic similarity, social similarities, cultural similarities, social gap, common interests, religion compatibility, number of children from a previous marriage, desire to marry, independency, relationship with the spouse's family, trading in, engagement time, love, commitment, mental health, the sense of having children, previous trading , previous marriage, proportion of common genes, addiction, loyalty, height ratio, good income, self confidence, relation with non-spouse before marriage, spouse confirmed by family, divorce in the family of grade 1, and socializing with the opposite sex age. The data was imputed by retrieving this data from a marriage counselor using consultations before marriage from 1975-2016.

Required Packages

library(readr) to open and use the csv file

library(dplyr) to convert to a data frame and use the dplyr functions to manipulate the data

library(ggplot2) to plot and create graphs with the variables

Plots and Table Needs

The plots and tables that will help me compare the variables are histograms, scatterplots, probability plots, and linear regression models. I also want to use covariance and correlation matrices.

Questions for Future Steps

I need to figure out how to put several variables to compare in histograms and how to properly cite the data for the final result.

How to import and clean my data

Import data

```
library(readr)
divorce_csv <- read.csv('Downloads/Marriage_Divorce_DB.csv')
head(divorce_csv)
```

```
##     Age.Gap Education Economic.Similarity Social.Similarities
## 1 0.1116326  1.915111            10.99868           76.456065
## 2 3.3553842  2.957842            82.13812           48.656031
## 3 6.5273650  2.772463            26.33783           59.356238
## 4 5.2030749  1.729242            66.95603            5.472612
## 5 6.8649618  4.370290            76.24503           26.797234
## 6 6.0724157  1.805129            48.92965           79.325712
##   Cultural.Similarities Social.Gap Common.Interests Religion.Compatibility
## 1             47.847460   50.31766         88.09990                83.73807
## 2             30.188517   54.11461         57.02097                98.40813
## 3             10.340252   76.59538         80.59099                41.74346
## 4              1.003407   55.07143         99.71808                70.49301
## 5             93.291581   73.73624         52.89620                11.72973
## 6             49.980833   86.31503         84.53861                37.11711
##   No.of.Children.from.Previous.Marriage Desire.to.Marry Independency
## 1                              4.402822       22.868019     1.269738
## 2                              4.367024       40.336843     1.658179
## 3                              1.197120       45.941845     1.766594
## 4                              3.392041        2.924863     1.065769
## 5                              2.373553       89.851492     1.103652
## 6                              4.240947       45.083740     1.455919
##   Relationship.with.the.Spouse.Family Trading.in Engagement.Time     Love
## 1                            73.20695   79.26236        5.424734 70.47223
## 2                            91.66609   63.76329        4.371315 60.81839
## 3                            17.92650   65.03750        6.514788 52.60183
## 4                            17.03675   32.49325        1.701053 80.60009
## 5                            81.65681   82.31090        9.012912 72.73072
## 6                            38.12058   37.14541        4.687885 37.34356
##   Commitment Mental.Health The.Sense.of.Having.Children Previous.Trading
## 1   76.10683      70.24180                      86.13846        39.437387
```

```
## 2   70.94377      80.02900                     60.66823        26.251231
## 3   96.09261      80.38394                     28.22565         7.228832
## 4   64.42502      63.75505                     45.34834        30.312101
## 5   73.94405      61.69622                     98.68880        58.777743
## 6   71.24099      64.59292                     88.86391        46.790771
##   Previous.Marriage The.Proportion.of.Common.Genes Addiction  Loyalty
## 1          2.766927                     21.032882  3.134119 49.64848
## 2          1.308314                     41.257627  2.067377 75.22070
## 3          4.203533                     23.917319  3.599095 22.55187
## 4          2.117970                     46.056440  1.549274 99.17214
## 5          1.151556                      2.743993  4.031738 21.62947
## 6          3.033611                     37.622215  2.547224 84.50858
##   Height.Ratio Good.Income Self.Confidence
## 1     30.82295    94.49916        45.96482
## 2     68.26822    41.10260        65.38771
## 3     59.13487    23.05358        84.27190
## 4     40.98412    43.40004        96.08123
## 5     89.12238    51.61551        53.33082
## 6     71.17455    72.52237        57.64234
##   Relation.with.Non.spouse.Before.Marriage Spouse.Confirmed.by.Family
## 1                                 2.032610                    1.719332
## 2                                 1.053402                    1.456192
## 3                                 8.268308                    7.095241
## 4                                 5.852371                    6.570749
## 5                                 9.717223                    7.609152
## 6                                 1.183741                    2.388009
##   Divorce.in.the.Family.of.Grade.1
Start.Socializing.with.the.Opposite.Sex.Age
## 1                         2.262242
24.35677
## 2                         9.795998
19.66715
## 3                         9.986173
15.52252
## 4                         5.099396
34.66593
## 5                         1.294295
22.54576
## 6                         4.664718
16.73969
##   Divorce.Probability
## 1            2.760190
## 2            1.962979
## 3            2.858803
## 4            1.404621
## 5            1.318819
## 6            2.329512
```

Cleaning data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

divorce_data <- select(divorce_csv, Age.Gap, Economic.Similarity,
Cultural.Similarities, Common.Interests, Religion.Compatibility,
No.of.Children.from.Previous.Marriage, Desire.to.Marry,
Relationship.with.the.Spouse.Family, Mental.Health, Commitment,
Divorce.Probability)
head(divorce_data)

##     Age.Gap Economic.Similarity Cultural.Similarities Common.Interests
## 1 0.1116326            10.99868             47.847460         88.09990
## 2 3.3553842            82.13812             30.188517         57.02097
## 3 6.5273650            26.33783             10.340252         80.59099
## 4 5.2030749            66.95603              1.003407         99.71808
## 5 6.8649618            76.24503             93.291581         52.89620
## 6 6.0724157            48.92965             49.980833         84.53861
##   Religion.Compatibility No.of.Children.from.Previous.Marriage
Desire.to.Marry
## 1               83.73807                                4.402822
22.868019
## 2               98.40813                                4.367024
40.336843
## 3               41.74346                                1.197120
45.941845
## 4               70.49301                                3.392041
2.924863
## 5               11.72973                                2.373553
89.851492
## 6               37.11711                                4.240947
45.083740
##   Relationship.with.the.Spouse.Family Mental.Health Commitment
## 1                            73.20695      70.24180   76.10683
## 2                            91.66609      80.02900   70.94377
## 3                            17.92650      80.38394   96.09261
## 4                            17.03675      63.75505   64.42502
## 5                            81.65681      61.69622   73.94405
## 6                            38.12058      64.59292   71.24099
##   Divorce.Probability
## 1            2.760190
## 2            1.962979
## 3            2.858803
## 4            1.404621
```

```
## 5             1.318819
## 6             2.329512
```

What does the final data set look like?

Final data is cleaning my data by selecting only the columns I plan on using in my research for a total of 11 variables. I have 10 research questions, while this dataset looks at 30 different variables of divorce so cutting it down to 10 variables and 1 being the total divorce probablity for 11 total is crucial to be condensed. The variables are already named by the divorce factors, so I did not have to change anything there.

What information is not self-evident?

I plan on uncovering new data by comparing certain variables to one another and pin point the top factors of divorce. The dataset shows the probability by couple, but the percentage on each variable differs, and it is hard to tell how each variable/factor affected each individual couple. I also want to compare what each couple that was least likely to divorce has in common, and what the couples that were most likely to divorce have in common as well. I assume there are variables such as commitment that play a role but couples will stay together and not divorce if children are involved. So I don't think this information is self-evident without comparing the variables/factors of divorce in the data.

What are different ways you could look at this data?

The different ways I could look at the data are using different graphs and plots to compare the variables/factors of divorce. I can use histograms to select a few of the variables to see which of those affect couples the most then slowly narrow it down to one or two of the top factors of divorce. I also would like to see the variables that played the least in a divorce and see if there are any unpredictable ones, that most would not assume to be the least to play a role. The data shows the factors that affected each couple the most, but it does not compare the factors between all the 90 couples.

How do plan to slice and dice the data?

I plan to slice the data by variable/column and compare each one to one another. I plan on using at least 3-4 variables per comparison for a total of 3-4 different histograms and scatterplots. The data was already sliced from the original 30 variables in the dataset to 11 variables for my analysis. Although the other variables are impactful in the divorce, it is not what I covered in my research questions.

How could you summarize your data to answer the key questions?

I would summarize the data by variables/factors that answer my specfic research questions. Then to answer my one question of what factors affects marriage the most to point of divorce, I would summarize the data with the highest divorce probability based on variables comparing all 90 of the couples that were questioned by the marriage counselor. I want to include my assumptions on which variables affect marriages those most, such as commitment and desire to marry. But this data may prove that to be otherwise and show an unpredictable result so that will be really interesting to see.

What types of plots and tables will help you to illustrate the findings to your questions?

The plots and tables that will help me illustrate the findings are histograms to compare 10 variables for each couple, and see what variable affected each couple the most. Then scatterplots to compare each variable to one another and see the highest points and if there are any patterns. I might also use probability plots, and linear regression models to see any patterns or correlations there. I also want to use covariance and correlation matrices to compare each variable as a second validation to show validity to my data analysis.

Do you plan on incorporating any machine learning techniques to answer your research questions? Explain

I do not plan on it only because I am not comfortable enough with these techniques yet to include it in my analysis. This dataset I am using however, is included in a paper where machine learning techniques were used. So I was able to see how it would look using evolutionary algorithms, specifically the adaptive neuro-fuzzy expert system.

Questions for future steps

To make the data easier to read, I want to rearrange the data where the divorce probability column is in order for most likely to divorce, to least likely, or vice versa. Right not it is in order of 1-90 of the different couples that were asked these questions but it would be much easier to read the data when it is in order of the probability. That way if it is easier to see a trend just by looking at the cleaned dataset. I am not sure what function to use for that yet, but I assume there is something in the dplyr package that can rearrange from highest number to lowest in one column or vice versa.

A story / narrative that emerged from your data. Follow this structure.

Introduction

For this final part of my research, I will use several plots and graphs to show the different factors in divorce and how they played a role compared to the divorce probability rate, or just within the factor itself. Divorce is becoming part of society just as much as marriage, and instead of it becoming normal, it should be something rare. Maybe marriage is not the best for a couple's situation, and that is perfectly fine. Couples do not have to become married to build a foundation for themselves and/or a family. It is the fact that the percent of divorces is greater than marriages. This proves that there is a problem within marriages, and it is matter of finding the greatest factors that play a role in divorce to help with this continous problem of the increasing divorce rate all over the world.

The problem statement you addressed

The problem I am addressing is what factors contribute to divorce the most. The divorce rate is currently at 54% and has only continued to go up. I chose a dataset that a marriage counselor used on couples before they got married, but later ended up in divorce. The questions asked are regarding the variables in the dataset, such as "desire to marry", this is the percentage the couple said they had before marriage. A variable I also included was the "divorce probability rate" that the marriage counselor had imposed based off their responses.

How you addressed this problem statement

I addressed the problem statement by conducting analysis on each research question individually using histograms, simple linear regression, and finding the means of each factor. Some variables depending on what their individual outputs were, were either used alone or with the divorce probability rate to compare and see if the variable increases or decreases with the divorce probability rate. A model I would recommend to create would have all the variables together, to see which is the highest factor for divorce. My recommendation would be a multiple regression model to see how each variable affects the divorce probability rate and see there regression lines all in one graph.

Analysis

I will go through all research questions #1-10 and summarize the analysis from each.

Research question #1: Does economic similarities or disimilarities play a role in a couple divorcing?

I used a histogram on this variable alone to show the percentage of economic similarities, and where the percentage was most frequent. This histogram shows the percentage with the highest frequency being around 60% and 79% of economic similarities. I find this insight interesting because I think most would assume if the economic similarities % was very low, that would be more likely to cause issues in a marriage, over having very similar economic situations. This does play a role in divorce.

Research Question #2: Does cultural play a role in a failed marriage, if they are culturally the same or not?

I also used a histogram on this variable and the most frequent percentage for cultural similarities is in 15%, 16%, and 20%. So most couples did not have similar cultures. This provides insight on a possibility of being a high factor of divorce. I can see this being a factor, as culture can play a huge role in morals and values where a lot of disagreement can occur, so it does in fact play a role in marriage.

Research Question #3: What factors affect marriages the most to the point of divorce?

I analysed the mean of each variable to see which one had the highest mean, and it was a very close call between the variables "common interests" and "mental health". Both with a mean of about 75%. I think both of those variables are not suprising, to have the highest means. What is insightful is even having 75% of common interests, will not change the fact that the couple still got a divorce at the end. It will be difficult to see which factor is the greatest for divorce without the multiple regression model.

Research Question #4: Does age play a factor in a divorce, or how old someone gets married?

I analysed this variable using a simple linear regression model. This output showed the larger the age gap in years, the higher the divorce probablity rate was. Vice versa, the smaller the age gap, the lower the divorce rate. I think most would assume this to be true before analysis, but it is nice to see the evidence since some would argue that age does not play a role. The mean for age gap is 5 years.

Research Question #5: Does having common interests help or are opposites more likely to stay married?

I did another simple linear regression with common interests and divorce probability rate variables. However, this regression line was almost completely horizontal so it is harder to analyze and conclude whether common interests plays a big role in divorce. There were also no couples that said they had 0% in common interests, so it wouldn't even account for those marriages. Therefore, it is hard to say whether more common interests helps than having less.

Research Question #6: If there are children involved from a previous relationship, does that have a role in divorce?

This analysis was the most difficult since this dataset showed no frequencies in the 0 children so assuming that this particular dataset every marriage had children from a previous marriage. The histogram also showed frequencies in between each whole number, for example 1.5 children had data which can be faulty, since there cannot be half of a child. So this variable is not the most reliable with the data, but I did get a whole number for the mean which was 3 children from a previous marriage. It is hard to say whether it has a role in divorce when the dataset doesn't accoount for marriages with no children from a previous marriage.

Research Question #7: When a couple divorces, how willing were they to be married in the first place?

I did a simple linear regression model for this analysis and the variable compared to the divorce probability rate was negative. The regression line shows the higher the desire to marry percentage is, the lower the divorce probability rate. Vice versa, the lower the desire to marry %, the higher the divorce rate. I would have assumed this to be true even before looking at the data but an interesting insight, is the mean for the desire to marry is about 48%. I would not expect that number to be so low, but it makes sense when they all end up in divorce regardless. I would hope that number to be a lot higher for couples that do not ever divorce.

Research Question #8: Does religion play an effect in whether couples tend to not divorce as often?

I used a histogram to analyze this variable and it shows a good amount of points distributed. There was two high frequencies in the lower 30% and the higher 75%. The mean for this variable was 52%, so a good half had religion compatibility. This sounds reasonable for a divorce dataset, and would assume it to be a high factor for divorce but not the greatest factor.

Research Question #9: Does having a good relationship with the spouse's family make a difference in a marriage?

I used a simple linear regression model for this variable and the divorce probability rate and the positive regression line confirms that the lower percentage for relationship with the spouse's family, the higher the divorce probability rate. Vice versa, the higher the % of a good relationship with the spouse's family, the lower the divorce rate. This is insightful for those families that tend to stay away from their loved ones and aren't close to their spouse's family. This data confirms it does make a difference in a marriage.

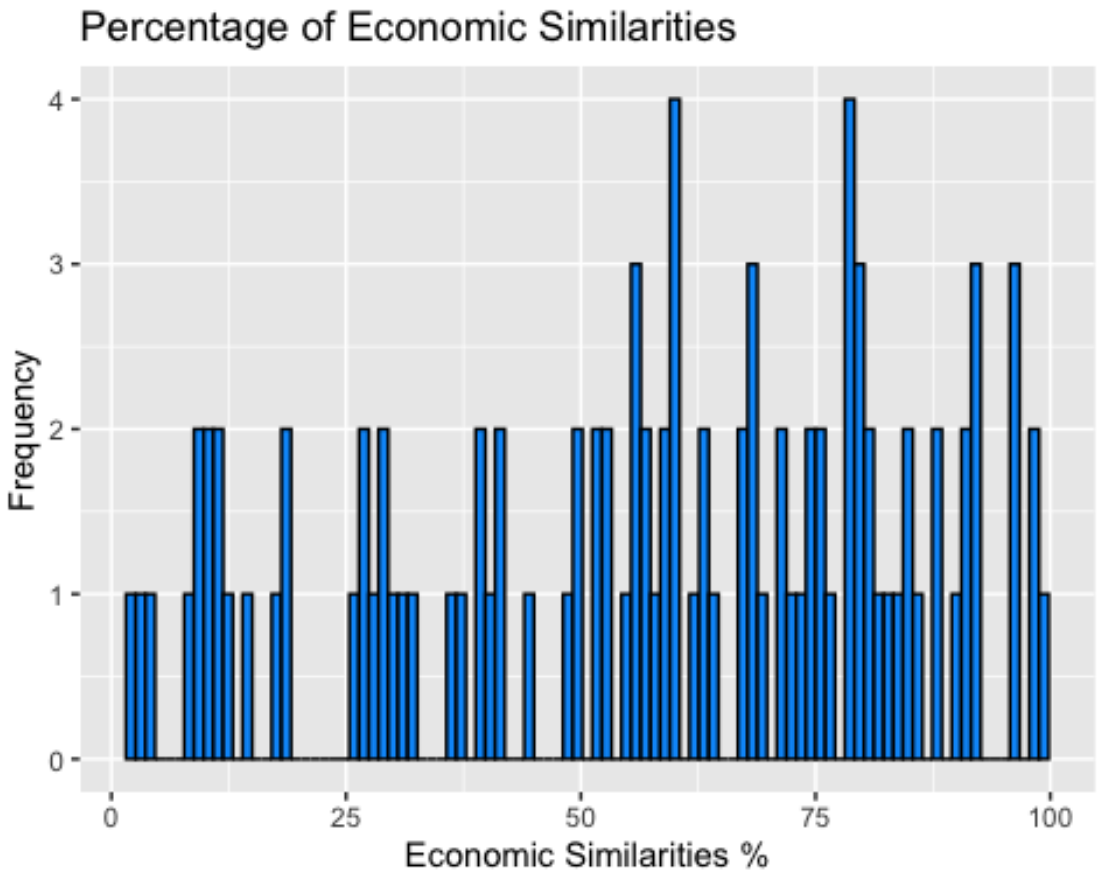
Research Question #10: Those who have mental health disorders, will that have a big enough affect to divorce?

This linear regression model shows a regression line with the higher the divorce probability rate, the higher the mental health percentage was. Vice versa, the lower the mental health %, the lower the divorce rate. This data is a bit tricky though, since the lowest mental health % in the dataset does not go below 50%. So this does not account for marriages with no mental health problems. It is safe to assume mental health does have a role in divorce, when it is present. The mean also was one of the highest at 75%. But this dataset does not cover no mental health issues, so it is hard to say it is the greatest factor in divorce.

Coding for Analysis

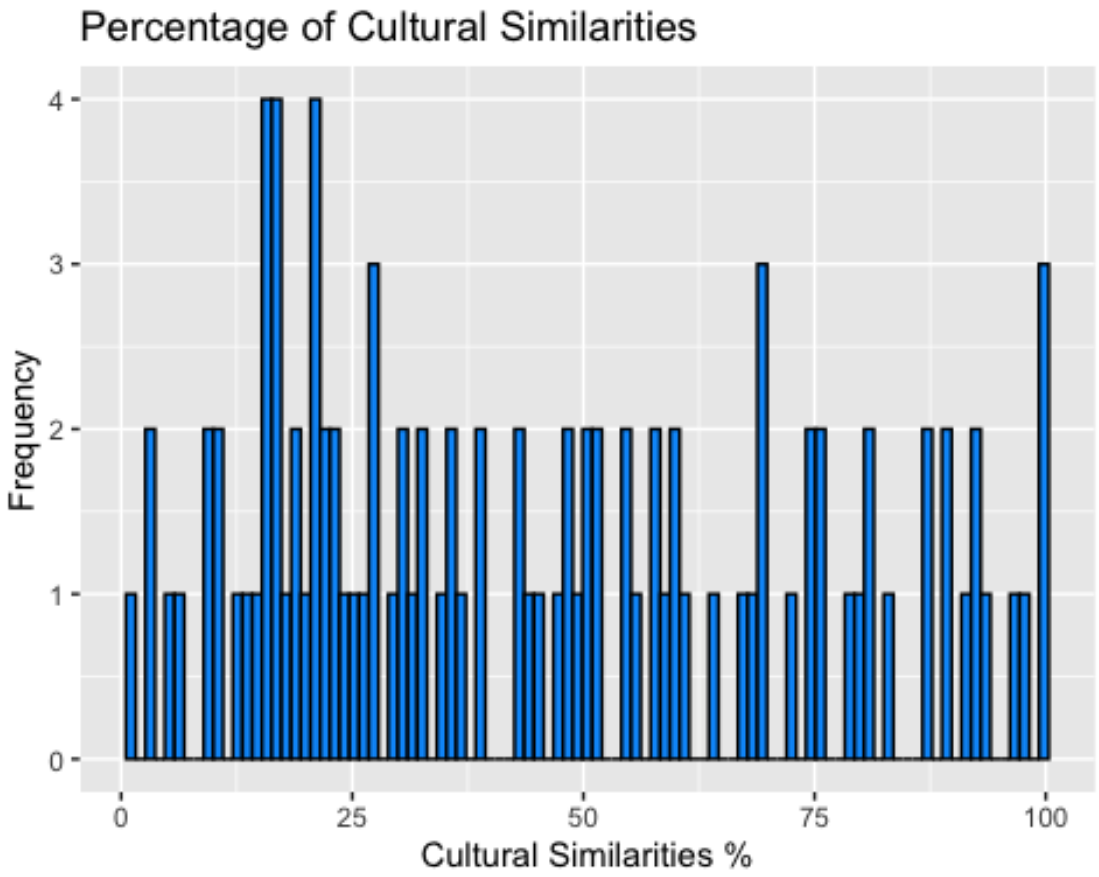
1. Does economic similarities or disimilarities play a role in a couple divorcing?

```
library(ggplot2)
ggplot(data=divorce_data, aes(x=Economic.Similarity)) + geom_histogram(bins =
95, color = "#000000", fill = "#0099F8") + labs(title = 'Percentage of
Economic Similarities', x = 'Economic Similarities %', y = 'Frequency')
```



2. Does cultural play a role in a failed marriage, if they are culturally the same or not?

```
ggplot(data=divorce_data, aes(x=Cultural.Similarities)) + geom_histogram(bins = 95, color = "#000000", fill = "#0099F8") + labs(title = 'Percentage of Cultural Similarities', x = 'Cultural Similarities %', y = 'Frequency')
```



3. What factors affect marriages the most to the point of divorce?

```
factor1mean <- summarize(divorce_data, mean(Economic.Similarity))
factor1mean

##   mean(Economic.Similarity)
## 1                  57.03479

factor2mean <- summarize(divorce_data, mean(Cultural.Similarities))
factor2mean

##   mean(Cultural.Similarities)
## 1                   45.35473

factor3mean <- summarize(divorce_data, mean(Age.Gap))
factor3mean

##   mean(Age.Gap)
## 1      5.037214

factor4mean <- summarize(divorce_data, mean(Common.Interests))
factor4mean

##   mean(Common.Interests)
## 1               74.99878
```

```r
factor5mean <- summarize(divorce_data,
mean(No.of.Children.from.Previous.Marriage))
factor5mean
```

```
##   mean(No.of.Children.from.Previous.Marriage)
## 1                                    3.004066
```

```r
factor6mean <- summarize(divorce_data, mean(Desire.to.Marry))
factor6mean
```

```
##   mean(Desire.to.Marry)
## 1              47.69228
```

```r
factor7mean <- summarize(divorce_data, mean(Religion.Compatibility))
factor7mean
```

```
##   mean(Religion.Compatibility)
## 1                     52.16547
```

```r
factor8mean <- summarize(divorce_data,
mean(Relationship.with.the.Spouse.Family))
factor8mean
```

```
##   mean(Relationship.with.the.Spouse.Family)
## 1                                  53.47605
```

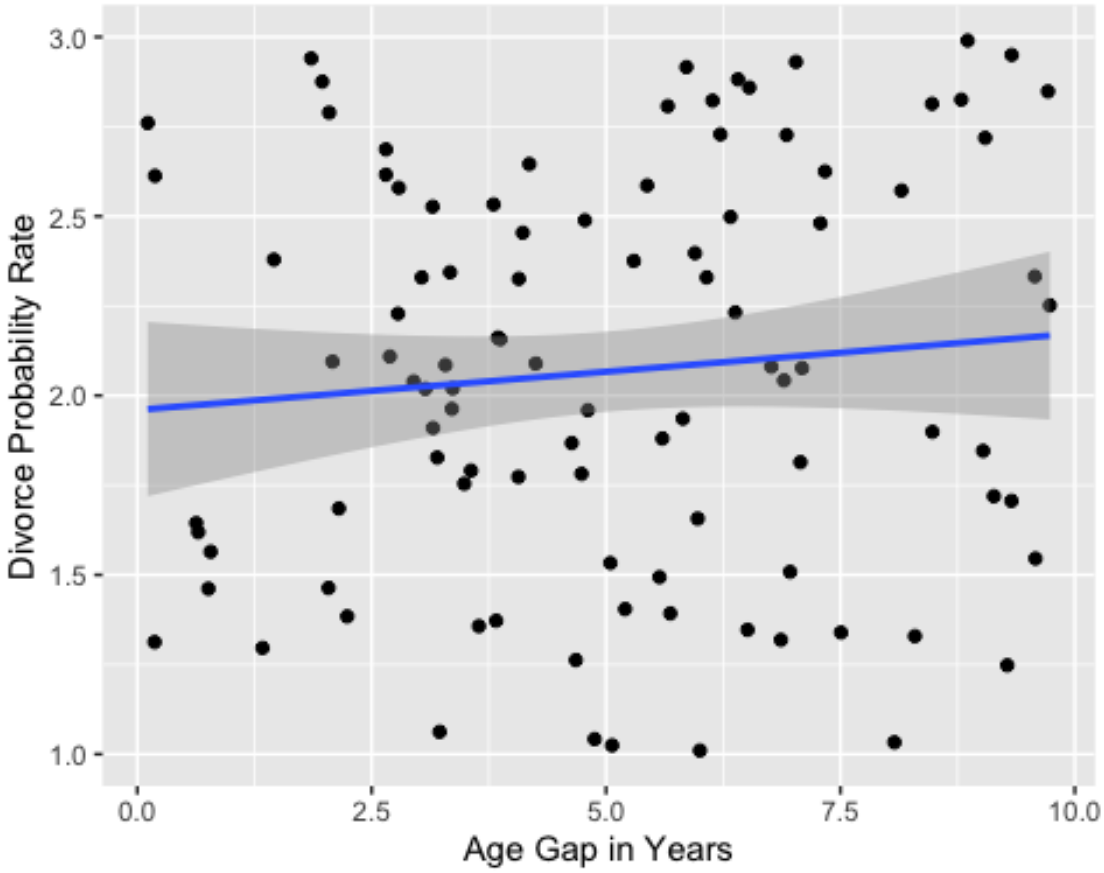
```r
factor9mean <- summarize(divorce_data, mean(Mental.Health))
factor9mean
```

```
##   mean(Mental.Health)
## 1            74.35657
```

4. Does age play a factor in a divorce, or how old someone gets married?

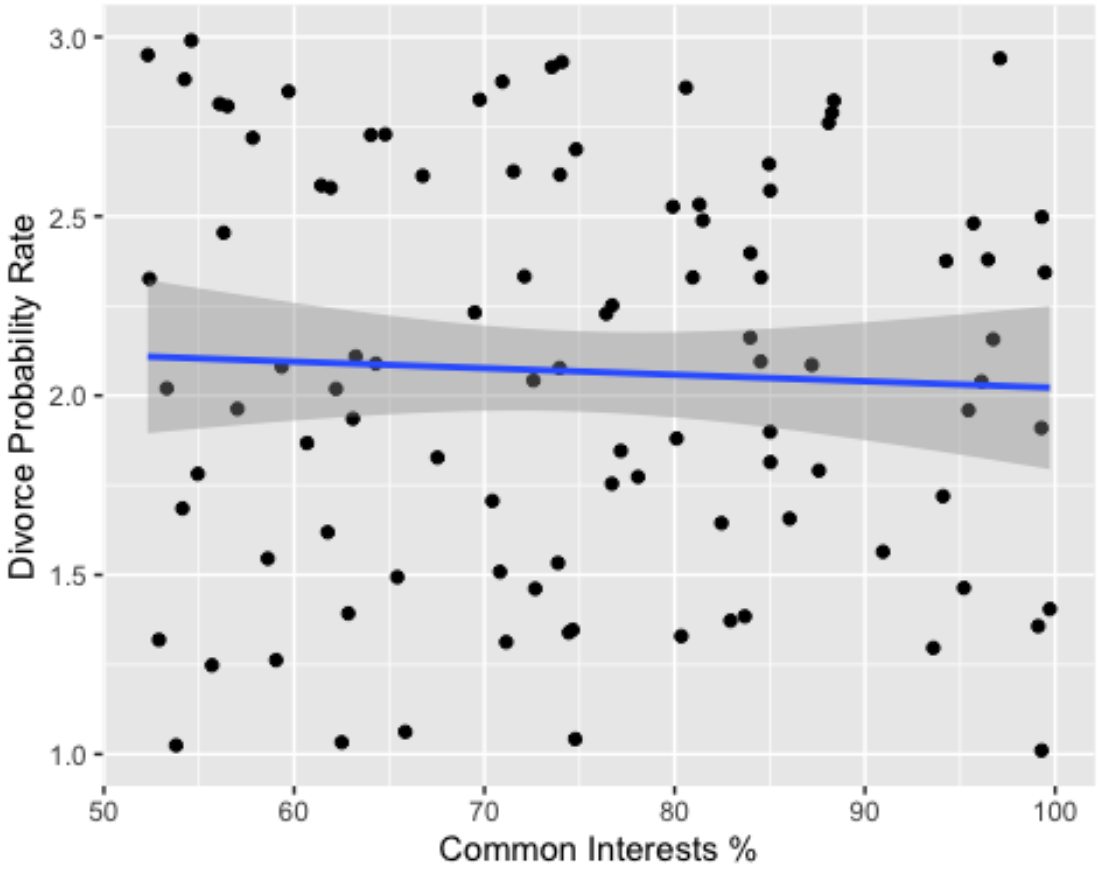
```r
ggplot(data=divorce_data) + aes(x=Age.Gap, y=Divorce.Probability) +
geom_point() + labs(x='Age Gap in Years', y='Divorce Probability Rate') +
geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



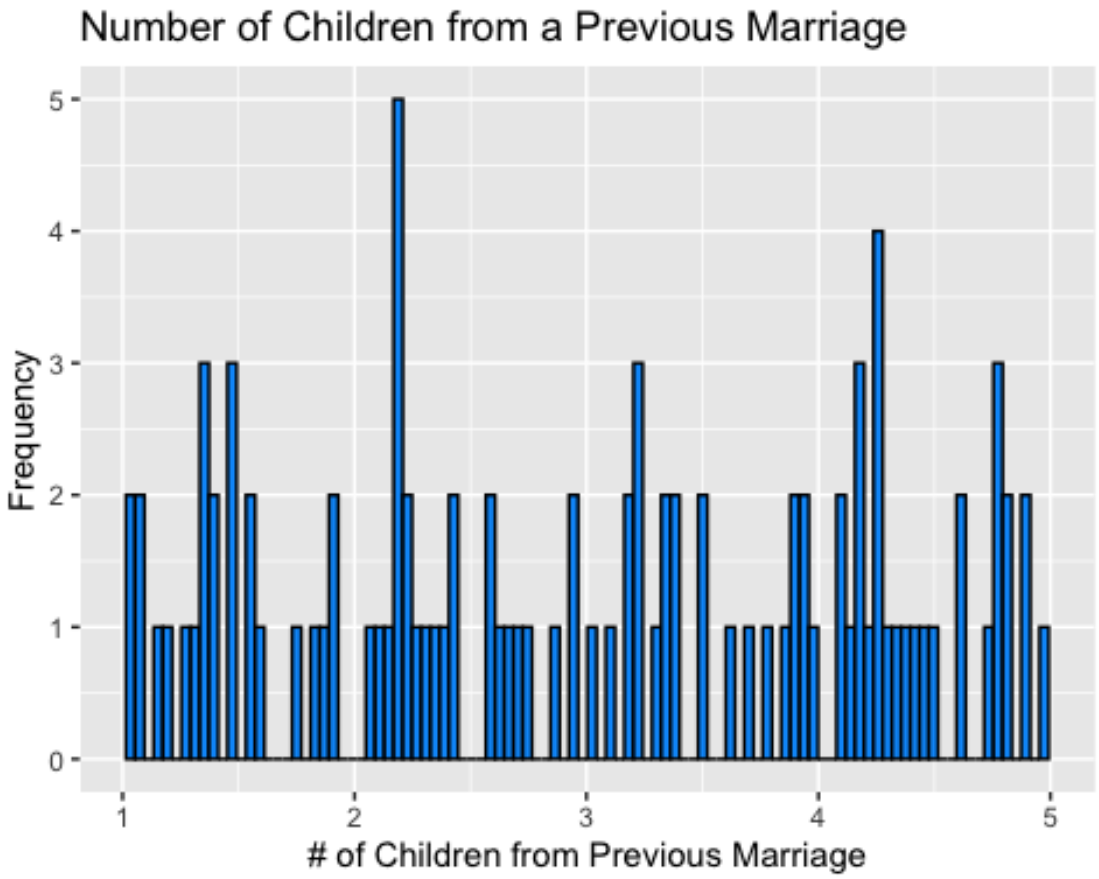
5. Does having common interests help or are opposites more likely to stay married?

```
ggplot(data=divorce_data) + aes(x=Common.Interests, y=Divorce.Probability) +
geom_point() + labs(x='Common Interests %', y='Divorce Probability Rate') +
geom_smooth(method = "lm")

## `geom_smooth()` using formula = 'y ~ x'
```



6. If there are children involved from a previous relationship, does that have a role in divorce?

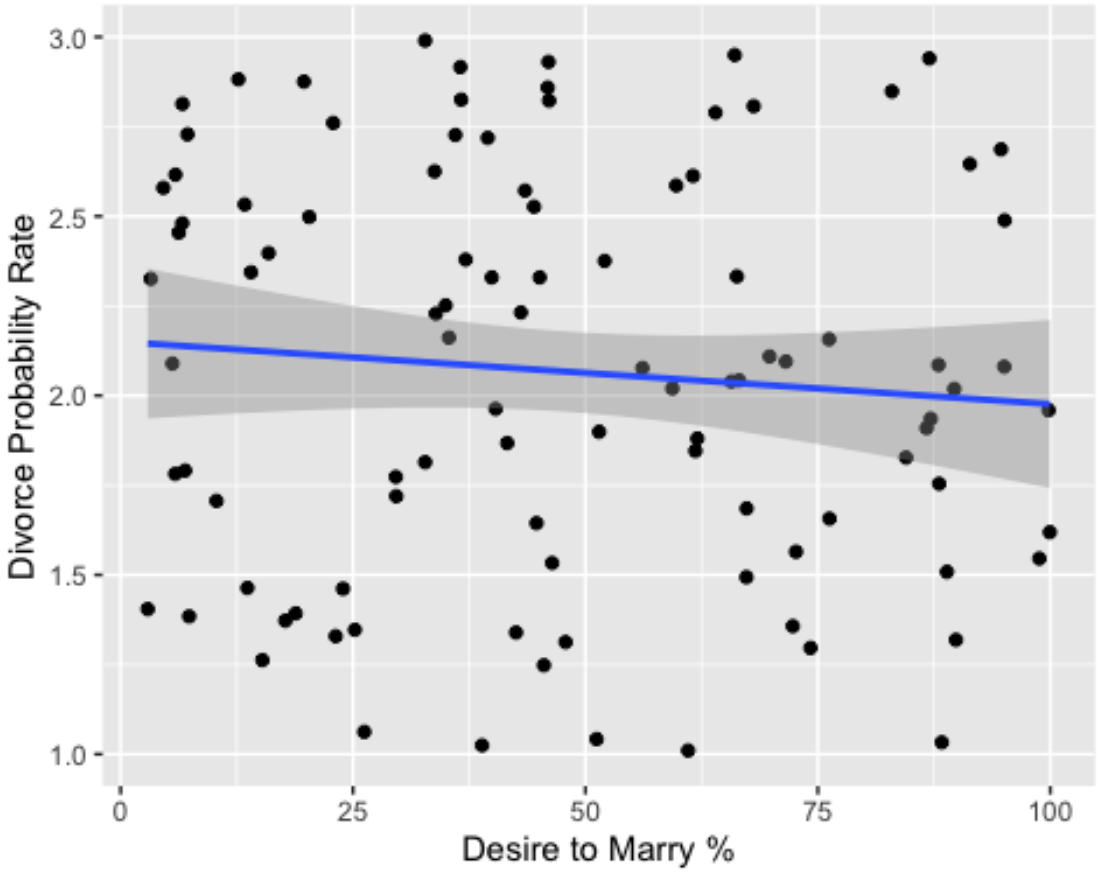
```
ggplot(data=divorce_data, aes(x=No.of.Children.from.Previous.Marriage)) +
geom_histogram(bins = 100, color = "#000000", fill = "#0099F8") + labs(title
= 'Number of Children from a Previous Marriage', x = '# of Children from
Previous Marriage', y = 'Frequency')
```



7. When a couple divorces, how willing were they to be married in the first place?

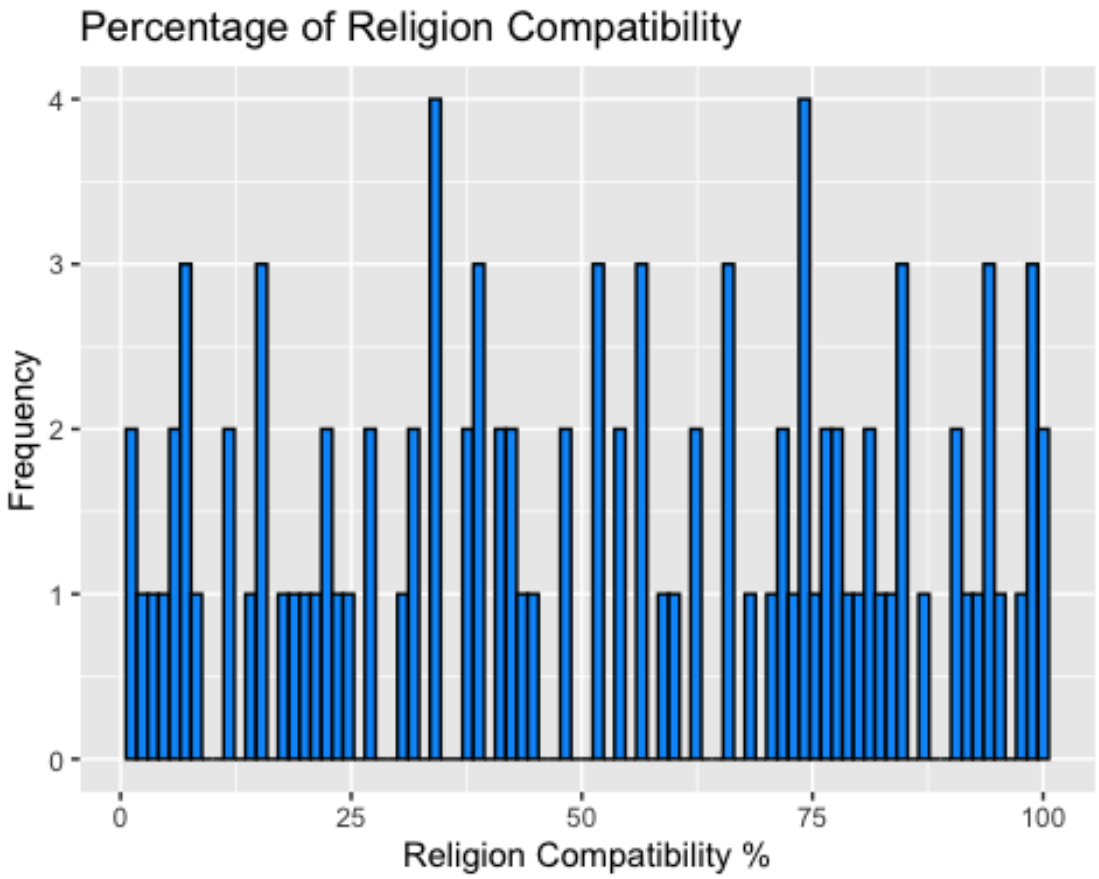
```
ggplot(data=divorce_data) + aes(x=Desire.to.Marry, y=Divorce.Probability) +
geom_point() + labs(x='Desire to Marry %', y='Divorce Probability Rate') +
geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



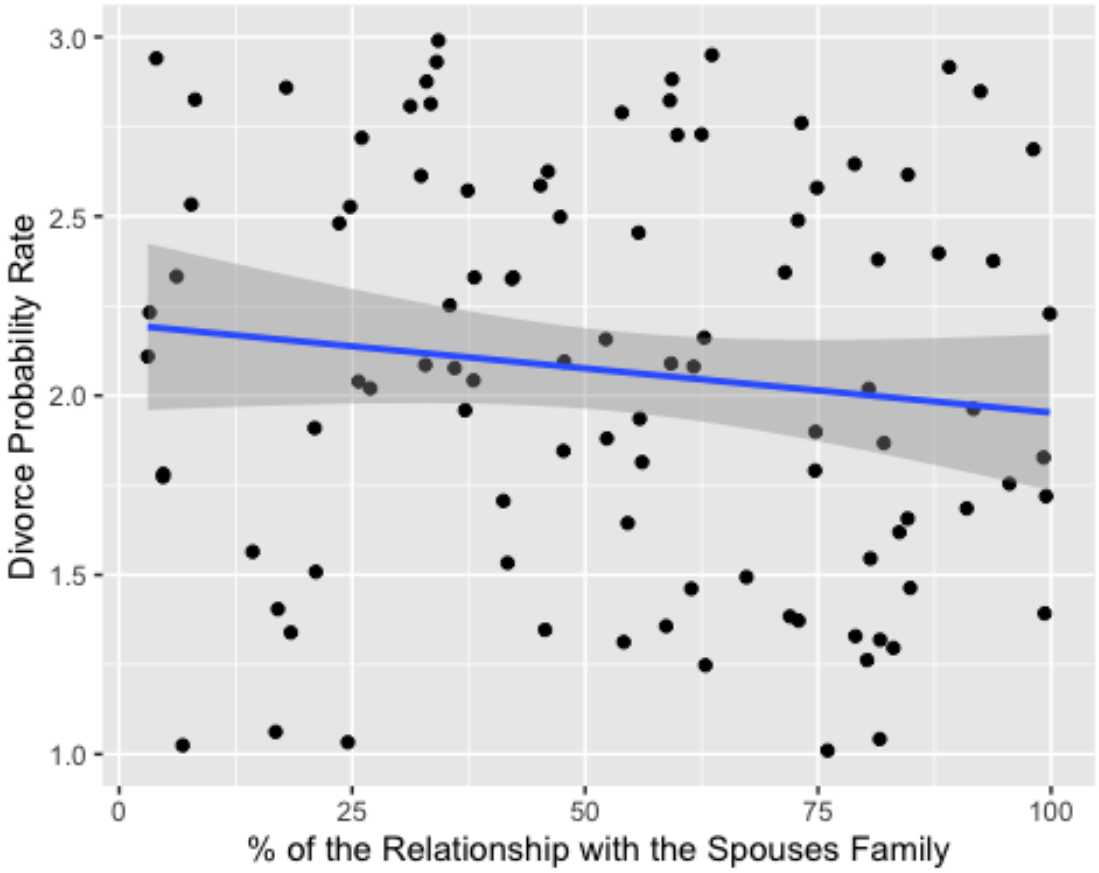
8. Does religion play an effect in whether couples tend to not divorce as often?

```
ggplot(data=divorce_data, aes(x=Religion.Compatibility)) +
geom_histogram(bins = 85, color = "#000000", fill = "#0099F8") + labs(title =
'Percentage of Religion Compatibility', x = 'Religion Compatibility %', y =
'Frequency')
```



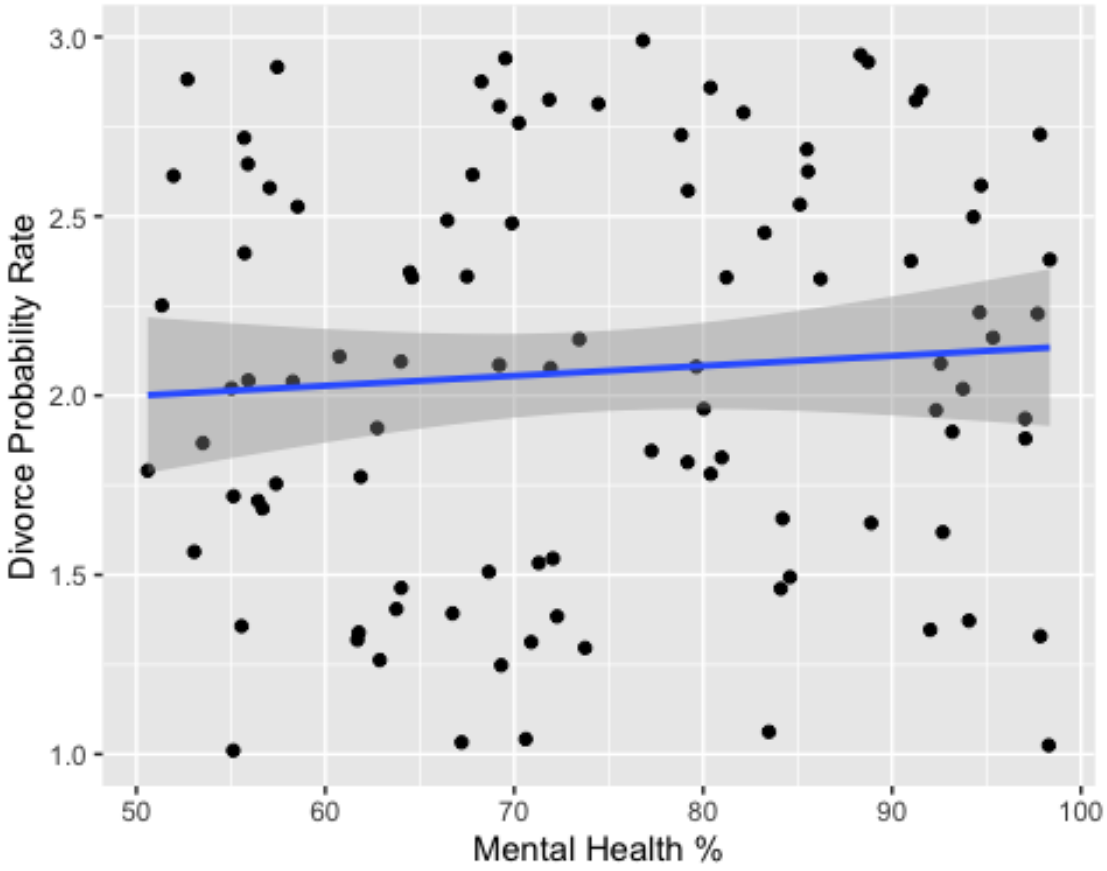
9. Does having a good relationship with the spouse's family make a difference in a marriage?

```
ggplot(data=divorce_data) + aes(x=Relationship.with.the.Spouse.Family, y=Divorce.Probability) + geom_point() + labs(x='% of the Relationship with the Spouses Family', y='Divorce Probability Rate') + geom_smooth(method = "lm")

## `geom_smooth()` using formula = 'y ~ x'
```



10. Those who have mental health disorders, will that have a big enough affect to divorce?

```
ggplot(data=divorce_data) + aes(x=Mental.Health, y=Divorce.Probability) +
geom_point() + labs(x='Mental Health %', y='Divorce Probability Rate') +
geom_smooth(method = "lm")

## `geom_smooth()` using formula = 'y ~ x'
```


Divorce Probability Rate
3.0
2.5
2.0
1.5
1.0
50
60
70
80
90
100
Mental Health %

Implications

My implications to the target audience would be, overall my analysis shows a good idea of what factors to look for in marriage. It provides insight on those questioning if have nothing in common can make or break a relationship, when this analysis shows majority of couples had 75% common interests yet still ended in divorce so this factor although having the highest percentage out of all the variables, shows that opposites can attract and weren't as common in this divorce dataset. Many of my research questions provide insight like that to a target audience but it is difficult to provide an overall answer to my initial problem statement of the largest factor in divorce without creating a more detailed and intricate model.

Limitations

Limitations would be the sample size. This dataset covers about 100 marriages which is very small in the entire population. I think running my research questions and analysis on other datasets from other countries can provide even greater insight and build on this small dataset. Arranged marriages are popular in some cultures, so even datasets on that could be built and expanded. There are many countries with extremely high divorce rates so looking at those datasets would be the most ideal to improve my analysis.

Concluding Remarks

Overall, this analysis for this dataset provided reinforcing insights on some assumptions that were safe to make such as having a good relationship with your spouse's family leads to a lower probability of divorce. But also new insights such as having as high as 75% common interests does not guarantee a successful marriage, but was more common for divorce in this specific dataset. With a good model, this anaylsis can be built upon, and improved to show more detailed insight to this growing divorce rate.