Divorces in Texas 2011-2013

STAT405 Group14: Hongyu Mao, Rylie Gao, Richard He, Qingyang Liu

11/30/2017

Group Member

- Yucheng He (yh43), Hongyu Mao (hm27), Qingyang Liu (ql12), Rylie Gao (yg38)

Introduction

- Main Dataset:
  Our main dataset is the Report of Divorce or Annulment Indexes from Texas Department of State Health Services in 2011, 2012, and 2013. The data set records information of each divorce in the state happened in 2011, 2012, and 2013, including each couple's husband name & age, wife name & age, number of child, date of marriage & divorce, and county name.

- Question for the Main Dataset:
  What are the common characteristics among the divorced population? In the following analysis, we will study divorced couples' husband & wife age, average age, age difference, marriage length, number of children, and the weather(precipitation). And we will also study the interrelationship among some of the factors.

- Supporting Dataset & Question for the Supporting Dataset:
  Our supporting dataset is the Weather data by Zip Code, which describes the daily amount of precipitation recorded by every station. We aggregated this dataset in SQLite to get monthly precipitation.

Content

- Analysis No.1: Husband's Age v.s. Wife's Age
  This section explores the characteristic of husband's age and wife's age in the divorced population.

- Analysis No.2: Divorce Counts by Four Varaibles
  1. Age Difference (in Years) v.s. Number of Divorces
  Explores the characteristic of age difference for divorced couples.
  2. Average Age v.s. Number of Divorces
  Explores the characteristic of average age for divorced couples.
  3. Marriage length (in Years) v.s. Number of divorces
  Explores the characteristic of marriage length (in years) for divorced couples.
  4. Number of Children v.s. Number of divorces
  Explores the characteristic of number of children for divorced couples.

- Analysis No.3: Number of Children v.s. Marriage Length (in Years)
  This section explores how the number of children affects marriage length (in years).

- Analysis No.4: Marriage length (grouped by average age) v.s. Number of Divorces + analysis of Harris County

This section explores how much proportion of the divorces at each marriage length (in year) is coming from a certain range of couple average age. Therefore, we can explore how the couple average age affects the number of divorces at different marriage lengths.

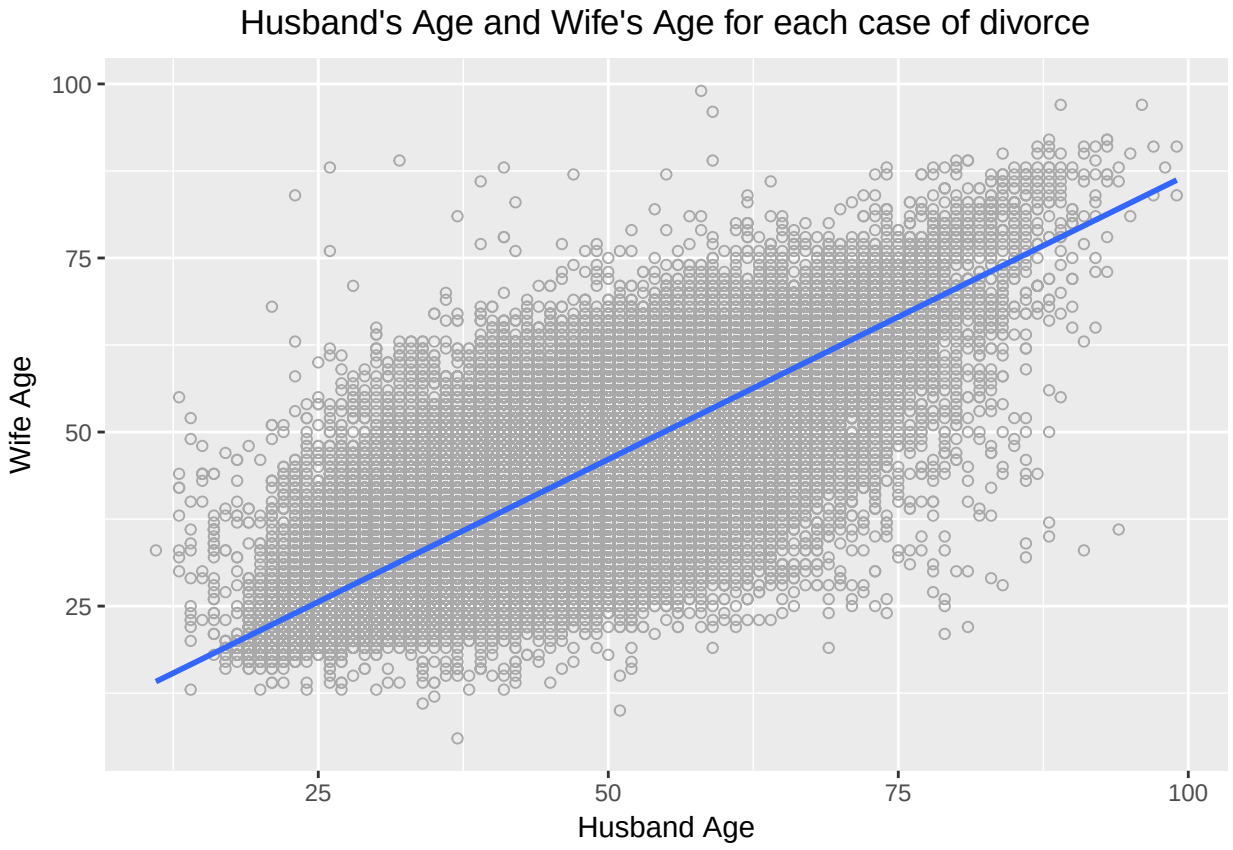
- Analysis No.5: Precipitation of each month v.s. Number of Divorces in that month
  This section explores how different months in a year affect the number of divorces, how different months affect precipitation of that month, and how the aggregated monthly precipitation affect the aggregated monthly number of divorces.

- Analysis No.6: Characteristics of Different Age Groups (Conceptual Plot)

Example of Raw Data

- *div1113_SQL*: The table containing all the marriage and divorce information from 2011 to 2013 in Texas. The columns are: SFN, H_NAME(husband's name), H_AGE(husband's age), W_NAME(wife's name), W_AGE(wife's age), NUM_CHILD(the number of children each family has), MARR_DATE(the marriage date), DIV_DATE(the divorce date), COUNTY_ID(ID for different counties), COUNTY_NAME(name of county in Texas).
- *Weather by Zip Code*: It records the weather condition on each day in each zip code area. We would like to check if there is any relationship between the number of divorces in a certain time period in each county and the weather condition in this time period in this county.

Analysis No.1: Husband's Age v.s. Wife's Age



In this plot, each gray point represents the husband's age on horizontal axis and the wife's age on the vertical axis. We have removed all the zero ages and the NAs so that the data we focus on is valid and makes sense.

The blue line is the linear regression line computed for the observations. We can see that there is a positive relationship between these two variables, with a slope of 0.818591. Which means that the estimated wife's age will be the husband's age times 0.8. For example, if a divorced man from Texas is 40 years old, we would predict that his ex-wife is around 32 years old.

In this plot, actually, we also include a 95% confidence interval, but it is invisible because the standard error is too small, about 0.001. Thus, the confidence interval is too narrow to see. We think that there are two main reasons that cause this. The first one is that, we have a large amount of data, more than 200 thousand points in this plot, and therefore, we have gained more certainty about where we believe the true value is. The second one is that, our linear model has a very good precision.

The conclusion we get here is that, generally, among divorced couples in Texas, if the man's age increases, his wife's age will also increase. In other words, the older the man is, the older his wife will be. And the wife's age is usually four fifths of the husband's age.

Analysis No. 2: Count Divorces by Four Variables

In this part, we are going to analyze the characteristic of divorced couples from four factors: age difference, average age, marriage length and number of children.

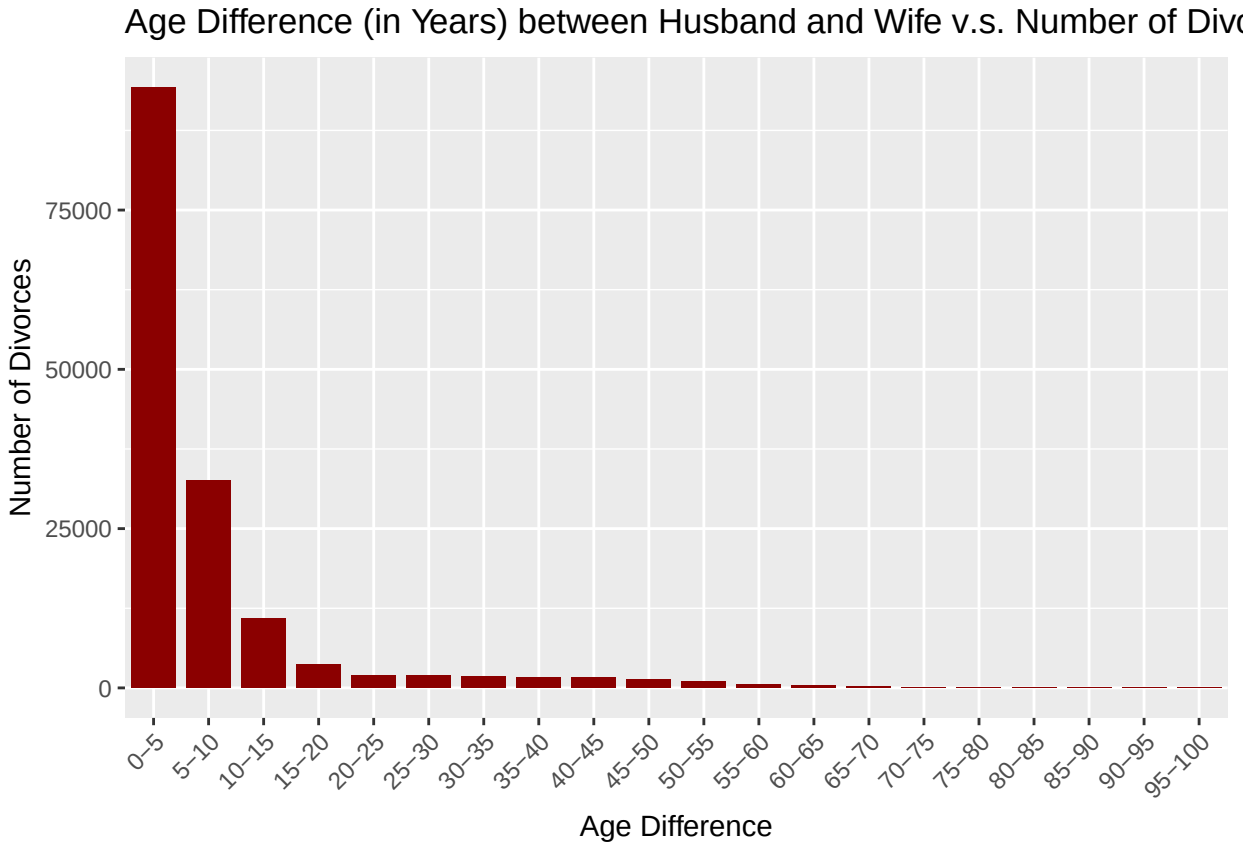
In the first plot, we will talk about the age difference between divored couples. We can see that most divorces happen when the age difference is under 15 and the number of divorces decreases suddenly when the age difference increases above 15. This is because the number of married couples whose age difference is greater than 15 is much smaller; therefore, the number of divorced couples with age difference greater than 15 is also small. And we can also find that the number of divorces decreases sharply when the age difference is greater than 5.

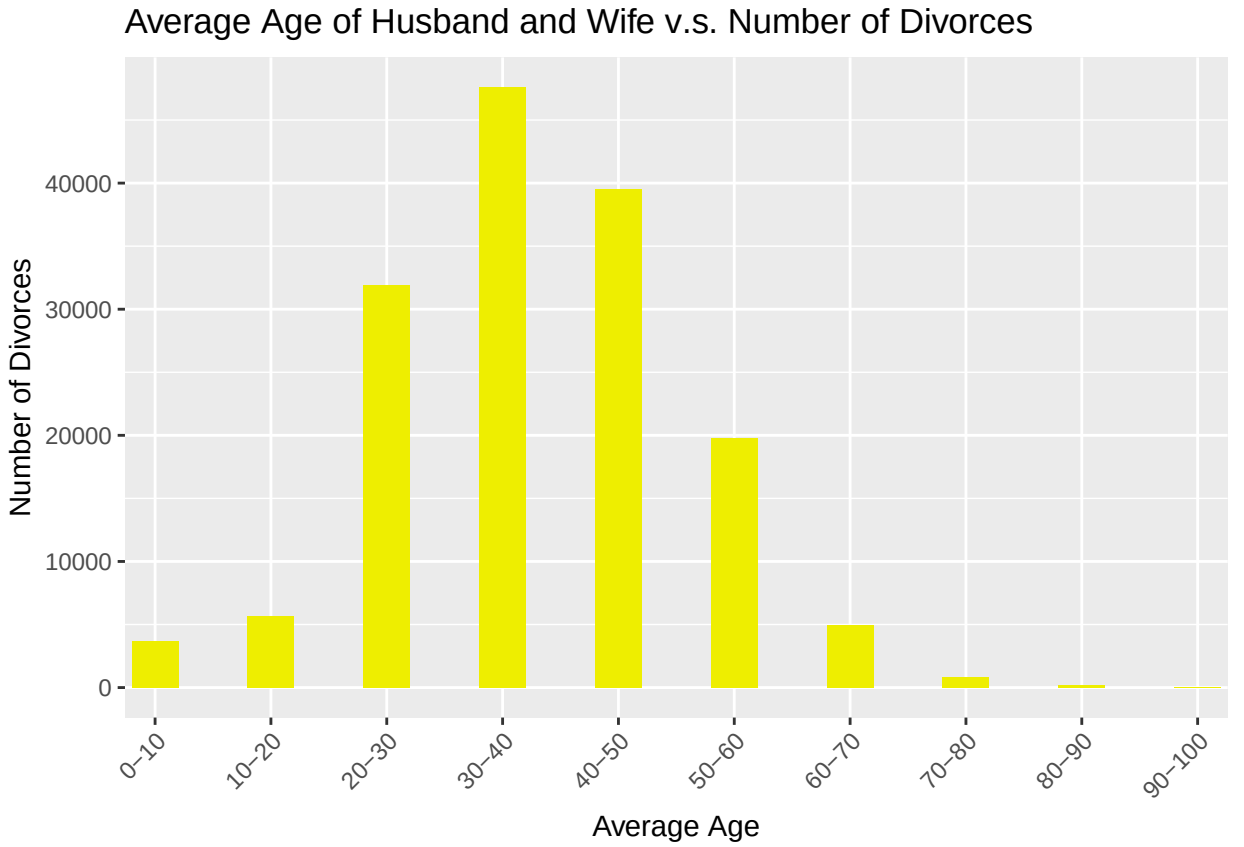
The second plot is about the average age of the divorced couples. From the plot, we can see that the number of divorces is normally distributed across average ages. Most divorces happen when the couples' age is around 30-50 years old, and fewer happens when the couples get younger or older. This is easy to understand, because few people get married under 20 and most couples cannot live up to 90. And in most cases, the middle-aged couples tend to divorce.

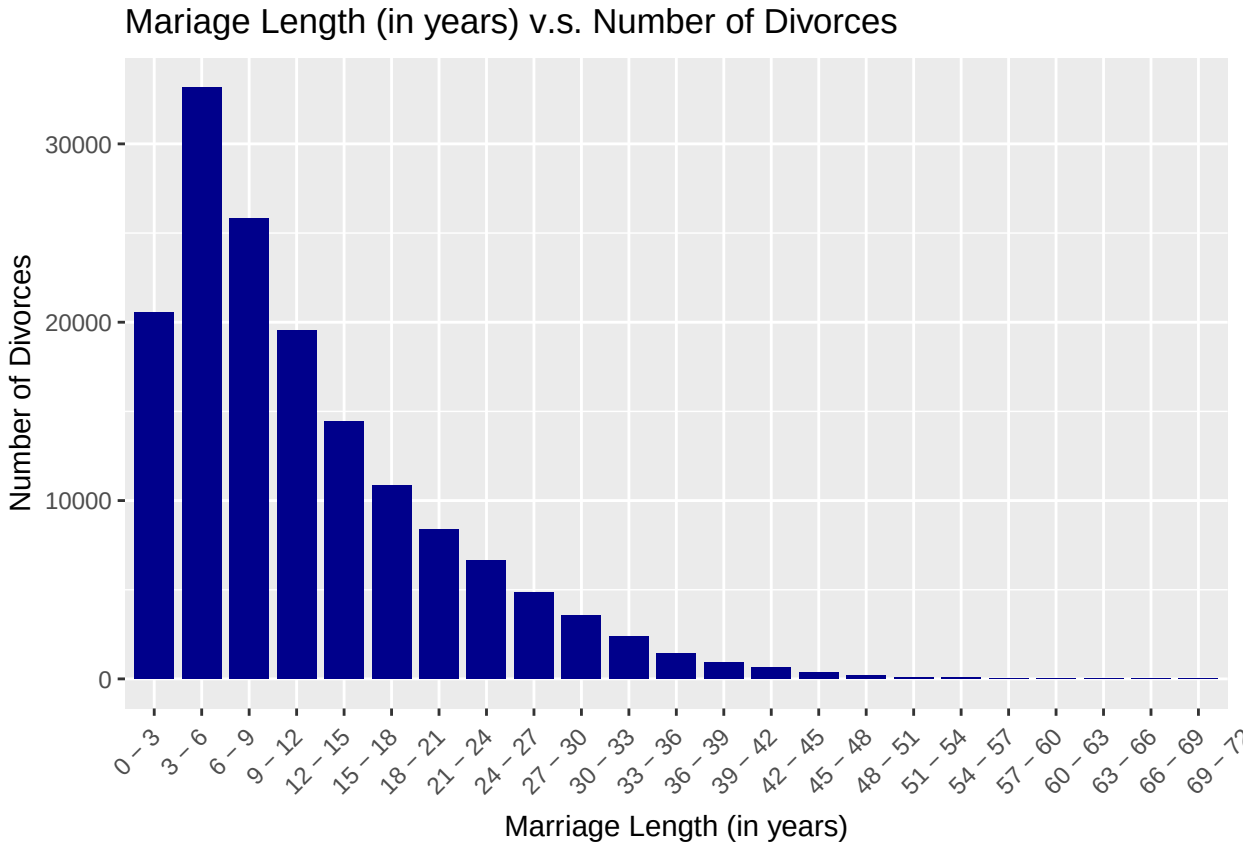
In the thrid plot, we will focus on the marriage length(in years) for the divorced couples. The number of divorces here follows a chi-square distribution. It increases when the marriage length increases from 0 to 6 years, and gradually decreases when the marriage length is greater than 6 years.

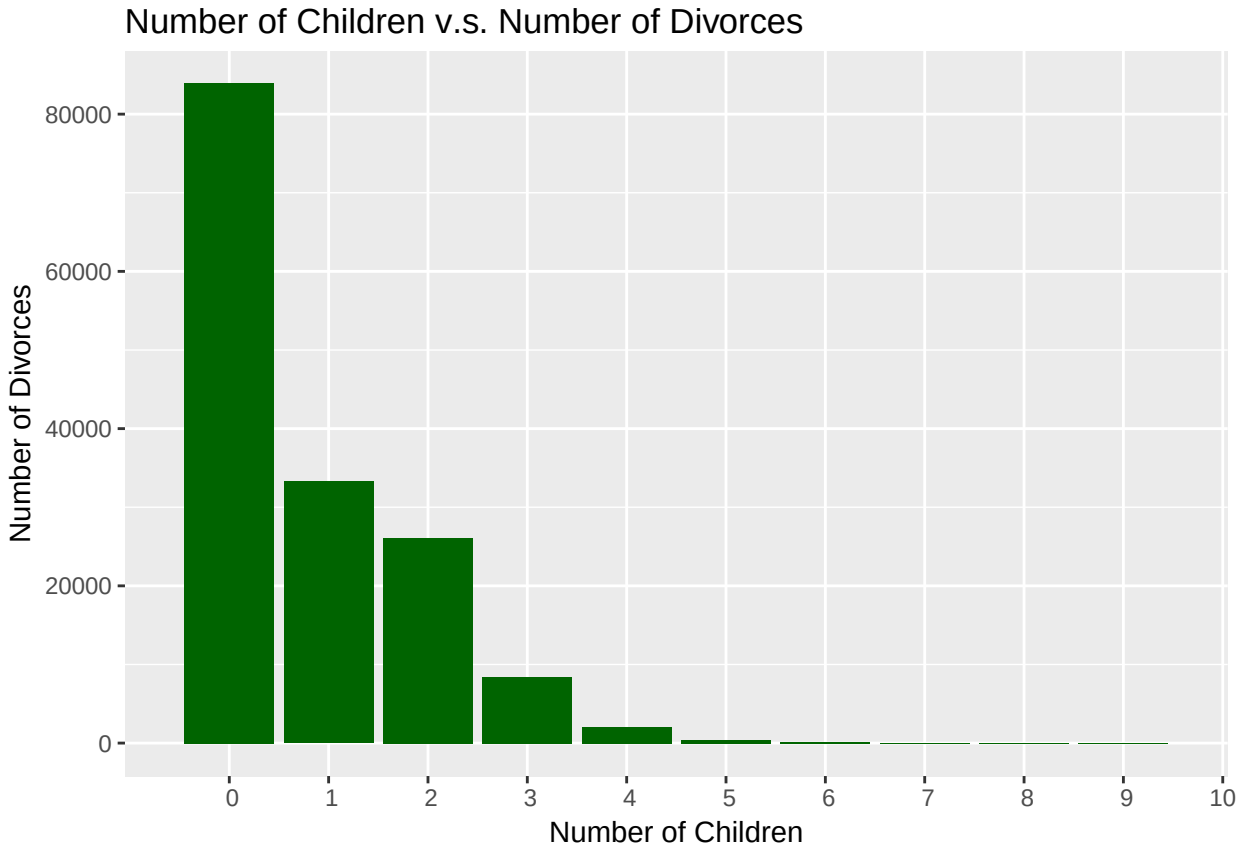
In the last plot, we will study the number of children for the divorced couples. We find that most divorced couples have no kids, and fewer than one half of the divorced couples have one or two kids. Nearly no couples divorce when they have more than 3 kids. The reason is easy to understand. First, few couples have more than 3 kids. Second, for sake of the children, most parents tend to maintain the marriage and they will not divorce. Therefore, most divorces happen when the couple has no kids.

After analyzing the four barplots, we can conclude that most divorced couples are between 30 and 50 years old; the husband and wife's age difference is smaller than 5 years old; their marriage length is between 3 and 6 years, and it is more likely that they have no kids.

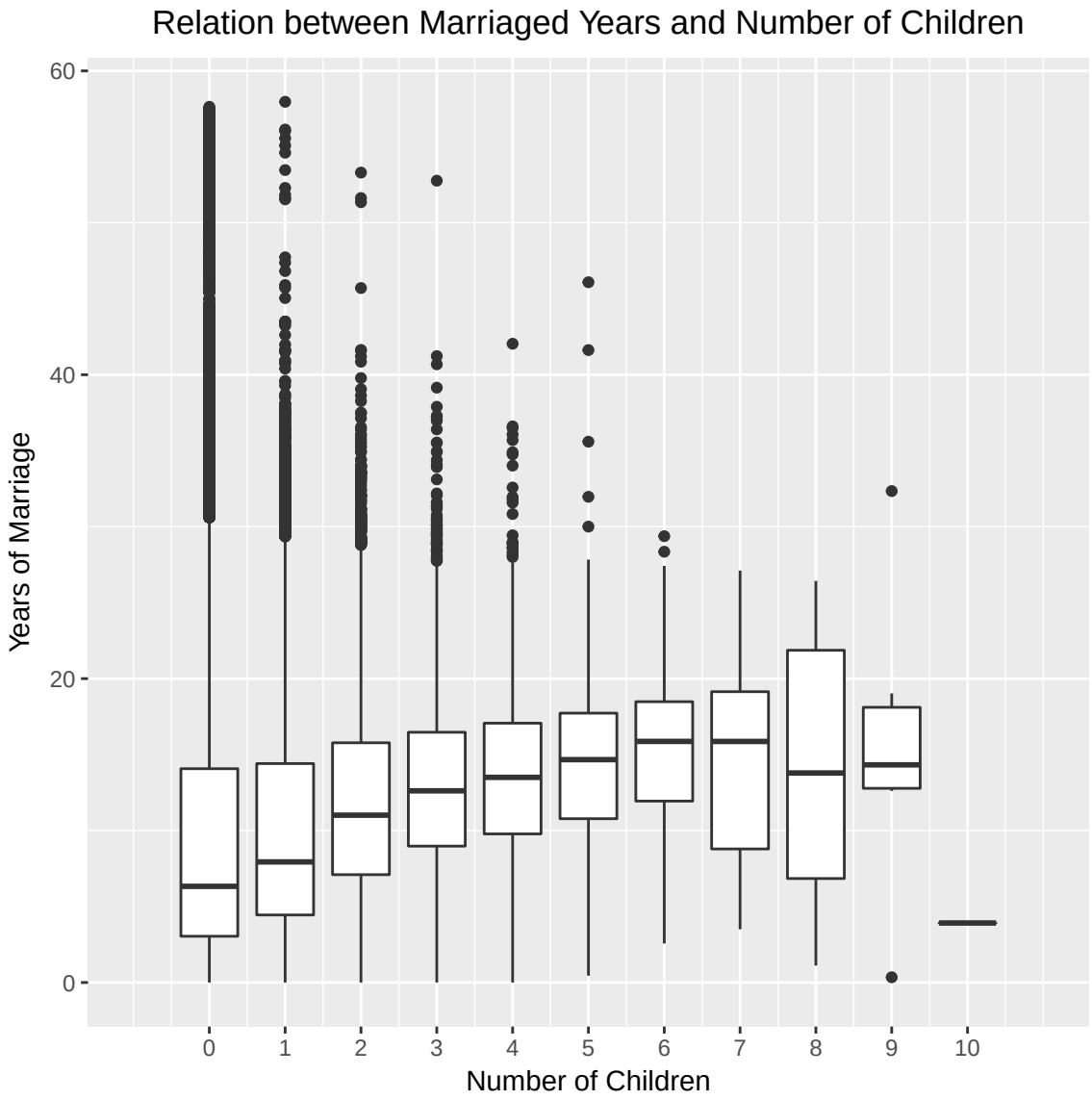

Age Difference (in Years) between Husband and Wife v.s. Number of Divo
Number of Divorces
75000
50000
25000
0
0–5
5–10
10–15
15–20
20–25
25–30
30–35
35–40
40–45
45–50
50–55
55–60
60–65
65–70
70–75
75–80
80–85
85–90
90–95
95–100
Age Difference


Average Age of Husband and Wife v.s. Number of Divorces
Number of Divorces
40000
30000
20000
10000
0
0–10
10–20
20–30
30–40
40–50
50–60
60–70
70–80
80–90
90–100
Average Age


Mariage Length (in years) v.s. Number of Divorces
Number of Divorces
30000
20000
10000
0
0 – 3
3 – 6
6 – 9
9 – 12
12 – 15
15 – 18
18 – 21
21 – 24
24 – 27
27 – 30
30 – 33
33 – 36
36 – 39
39 – 42
42 – 45
45 – 48
48 – 51
51 – 54
54 – 57
57 – 60
60 – 63
63 – 66
66 – 69
Marriage Length (in years)


Number of Children v.s. Number of Divorces
Number of Divorces
80000
60000
40000
20000
0
0
1
2
3
4
5
6
7
8
9
10
Number of Children

Analysis No.3: Number of Children v.s. Marriage Length (in years)

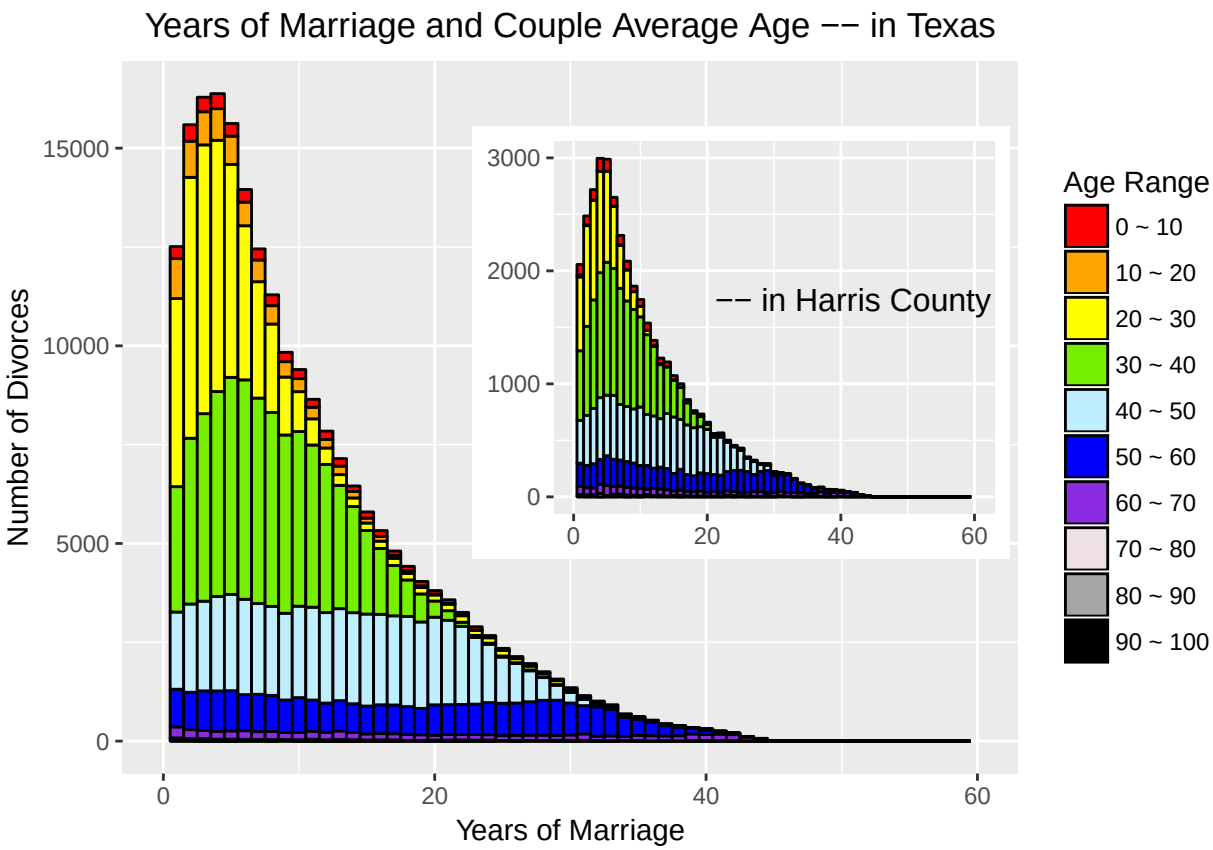


In this box plot, each point represents a record of a divorced couple. We divide them into different groups according to their number of children, and for each group, we analyze their years of marriage.

From this box plot, we concluded three observations:

1. The maximum married years for a group decreases as the number of children increases.
2. From the sample medians, we can see that the marriage length of a couple generally increases as the number of children this couple has increases. From a sentiment perspective, we can also see that since the more children a couple has, the more likely this couple will stay married for more years.
3. The spread of marriage length shrinks as the number of children increases. From a statistical perspective, this is more likely due to the number of observations is fewer for divorce records with fewer children.
4. The number of outliers decreases as the number of children increases, meaning fewer cases of abnormal data points.

Analysis No.4: Couple Average Age v.s. Years of Marriage



Talking of couple's age and their marriage length, most of us would believe that older divorced couples tend to have a longer marriage. However, this is not what we found from the divorce data in Texas from 2011 - 2013. As we can see from the two blue parts in the plot, the bar height stays almost the same horizontally, which means that older couples (couples whose average age equals to or greater than 40 years old) have the same number of divorces across different marriage lengths. This contradicts to our previous assumption.

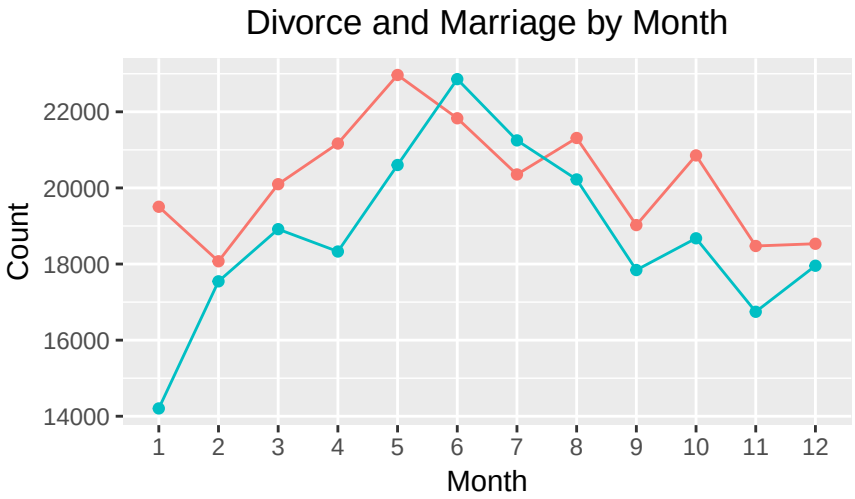
Next, we look at the yellow and the green parts. They represent the couples with an average age between 20 to 40 years old. According to the plot, the bar height increases first, then decreases. This resembles a chi-square distribution similar to the distribution of the whole data set. We believe the reason is that, the couples between 20 to 40 years old constitute the main part of the whole dataset, and thus their distribution has a strong influence on the distribution of the whole dataset.

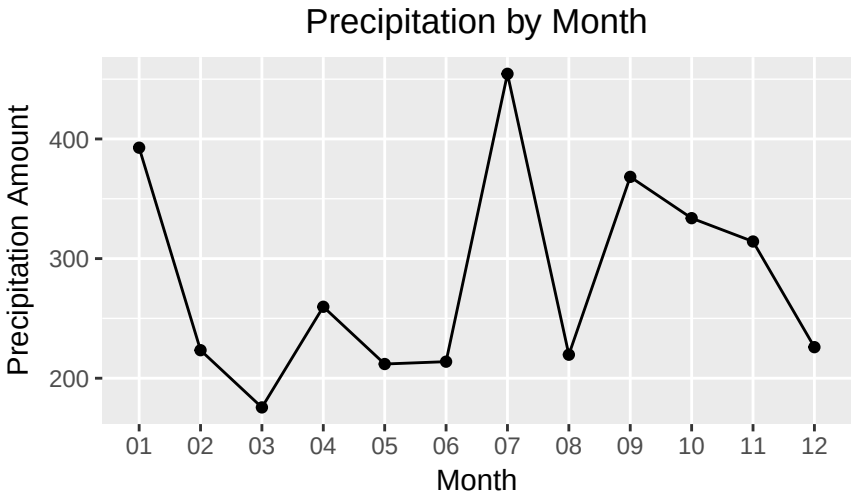
For the red and orange parts (couples under 20 years old), the number of divorces stays almost the same over the first few marriage lengths, then it decreases. In addition, we can also see that couples' marriage lengths are limited by their ages, which to some extent proves the validity of our data.

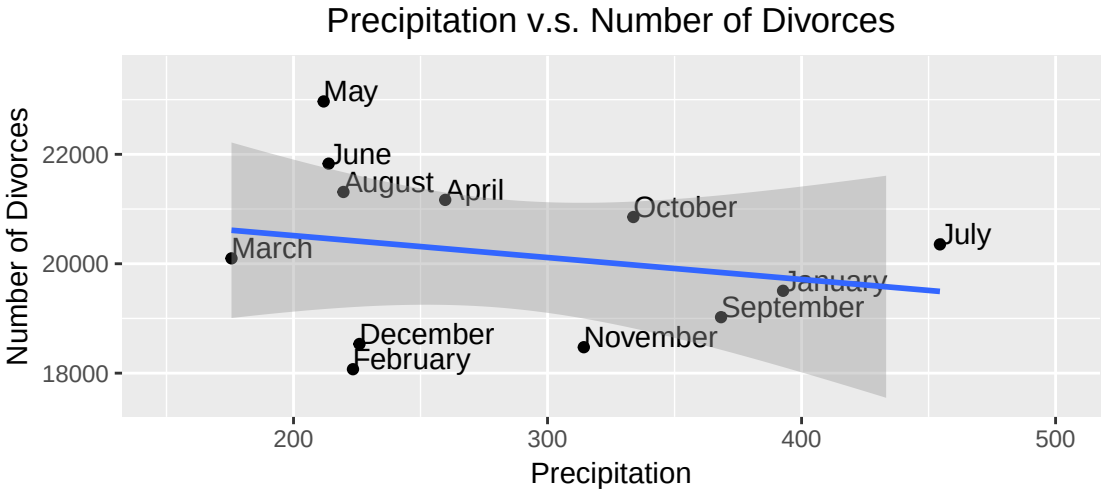
Also, the data records from just Harris County follows a same trend and pattern as the data from the whole Texas.

Notes: There is one thing that worth our attention, which is, there appears to be a decent number of couples under 20 or even 10 years old. This probably is because of wrong entering of data, but after doing some research online, we found out that, according to the former marriage law in Texas, any child of any age could get married with parental and judicial approval. But, this year, thanks to a former Texas child bride, who changed the marriage laws in Texas, state lawmakers have officially made it illegal to marry under the age of 18.

Analysis No.5: Precipitation v.s. Number of Divorces







We were curious about how months will affect the number of divorces and we made the first plot. In this plot, the green dots represent number of divorces in each month and the red dots represent the number of marriages for the divorced couples in Harris county. We notice that most divorces happen in the middle of the year, and fewer divorces happen at the beginning or end of the year.

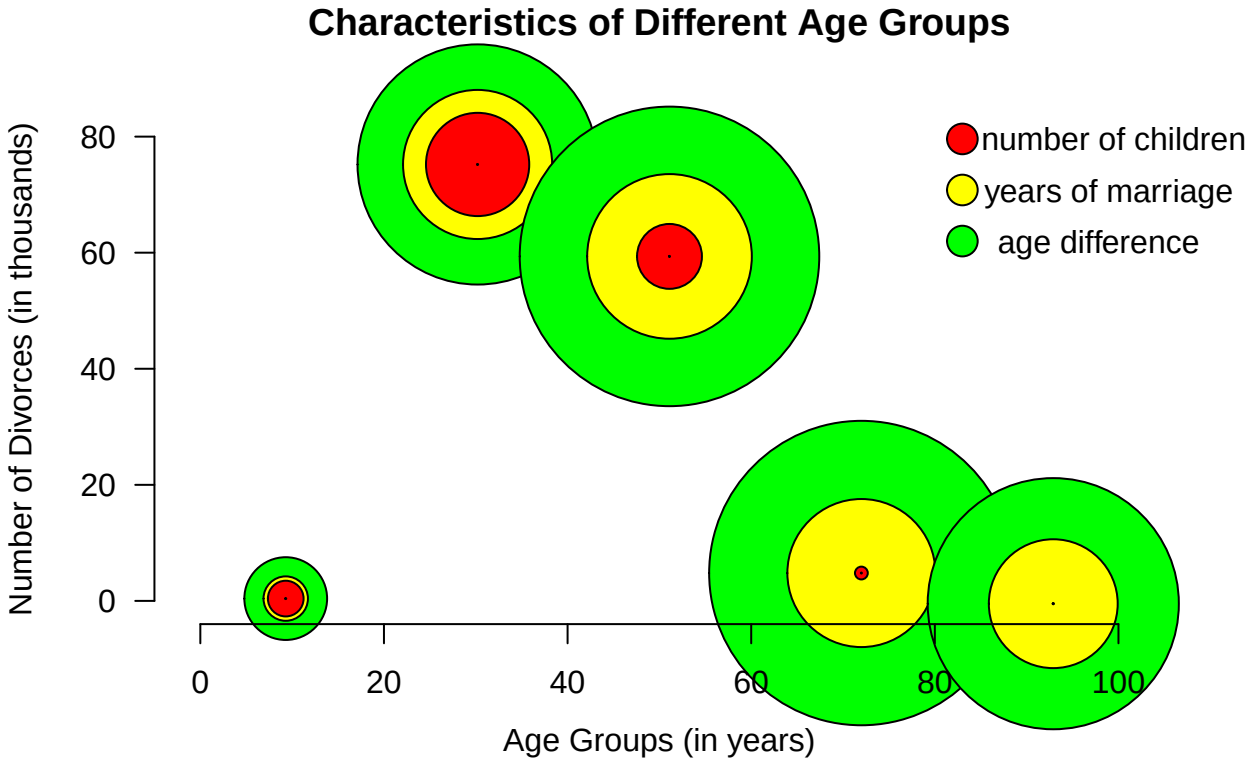
In the second plot, we introduce our auxiliary dataset – the precipitation in Harris county in each month. And this plot shows the different amount of precipitation in that month.

We are also interested in the relationship between amount of precipitation and number of divorces. For example, it is possible that, in the rainy days, people get sad easily and have more quarrels with their spouse. Consequently, they get mad with each other and go to Texas Department of State Health Services to get divorce. In our analysis, we used the aggregate amount of precipitation between 2011 and 2013 for each month as the x-axis, and the aggregate number of divorces between 2011 and 2013 for each month as the y-axis. Then we will have 12 points in the third plot, each of which represents the aggregate number of divorces for a given precipitation amount.

In the third plot, the blue line is the linear regression line and the shaded area is the 95% confidence interval. From this plot, we can find there is a weak negative linear relationship between the amount of monthly precipitation and the aggregate number of divorces.

But, since the slope for the regression line is too small, this is not a good linear model and we cannot use it to predict the number of divorces for a given amount of precipitation. Although the linear model is not good, it helps us reject our previous assumption. At first, we assumed that more couples get divorced during rainy days. But from the plot, it seems like there are more divorces when the precipitation amount is low.

Analysis 6: Characteristics of Different Age Groups



In our conceptual plot above, each age group (from 0-20 years old, from 20-40 years old, from 40-60 years old, from 60-80 years old, from 80-100 years old) has a circle consisted of 3 rings of 3 different colors. The width of each ring represents the relative value of 3 variables for a specific age group: number of children, years of marriage, and age difference. We can clearly see the following points from our conceptual graph:
1. As the widths of the green rings barely change among different age groups except the first group (the group from 0-20 years old), we can see that the age difference is almost the same across couples of different ages.
2. The width of the yellow ring increases as the group's age becomes larger and larger. Thus, marriage lengths increase as the couples become older, which makes sense intuitively, too. For the red wings, their widths shrink as the groups become older and older. For the oldest group, the red wing is even invisible,

which means that most of the divorced couples who are 80-100 years old don't have children. We can conclude that the older the couples, the more likely that they don't have any child.
3. The vertical axis represents number of divorces for each group and the horizontal axis represents the age for each group, which determine the position of the circle's center for each age group. When we connect the five center points by straight line, the line's trend resembles the pattern we found in Analysis 2 (Average Age v.s. Number of Divorces).

Conclusions to the Question

Re-cap of the project question: How are different factors (like husband's age and wife's age, marriage length, number of children, monthly precipitaion ext.) related to the number of divorces in Texas throughout 2011-2013? And are there any relationship between some of those factors?

According to our five analysis above, we've found that:
1. There is a positive relationship between the husband's age and wife's age. And for any divorced couples in Texas, the estimated wife's age = husband's * 0.8 + 5.
2. The number of divorces is influenced by variables (age difference between husband and wife, length of marriage, couple's average age and number of children) in different ways.
3. Relationship between number of children and marriage length: the more children the couple has, the more likely they will stay in marriage.
4. We cannot find a strong relationship between precipitation and number of divorces.
5. Divorced couples from different age groups have different divorced characteristics.

References:

1. Texas Department of State Health Services. (2015). Report of Divorce or Annulment Indexes (2011, 2012, 2013 versions). [The data set records information of each divorce in the state happened in 2011, 2012, and 2013, including each couple's husband name & age, wife name & age, number of child, date of marriage & divorce, and county name]. Retrieved from https://www.dshs.texas.gov/vs/marriagedivorce/dindex.shtm

2. PromptCloud (user on the website called "kaggle"). (2017). Airbnb Property Data from Texas [The dataset records each Airbnb host's location, price, and description in Texas]. Retrieved from https://www.kaggle.com/PromptCloudHQ/airbnb-property-data-from-texas

3. MARIA (user on the website called "gaslamp media"). 2013. Zip Code Latitude Longitude City State County CSV. [The data set records all the Texas zip codes and their associated latitude, longitude, city, state, and county]. Retrieved from https://www.gaslampmedia.com/download-zip-code-latitude-longitude-city-state-county-csv/