

How Couples Meet and Stay Together (HCMST)

Dawn Chen, James Diao, Jørn Emborg, Amy Zhao

May 4, 2017

Contents

1	Introduction	1
2	Data	2
2.1	Variable Descriptions	2
2.2	Exploratory Plots	3
3	Analysis	5
3.1	Methods of Meeting and the Rise of Online Dating	5
3.2	Exploring correlations between meeting online and political party, education level, and parental approval	6
3.3	Examining how religion and race affect partner selection	7

1 Introduction

Romantic relationships are a crucial part of life. Love, dating, marriage, and divorce constitute some of the most significant events in many people’s lives. With emerging data, it may be possible to better understand these important interpersonal interactions, and how they have changed over time.

Michael Rosenfeld’s group at Stanford has compiled a dataset entitled “How Couples Meet and Stay Together” (HCMST), containing 534 survey responses from 4002 respondents on demographic and relationship information (<https://data.stanford.edu/hcmst>).

HCMST contains 5 sequential rounds of surveys from 2009, 2010, 2011, 2013, and 2015. Due to response bias in follow-up surveys, we have refrained from speculating on which couples have stayed together and why. Instead, we have focused on the earlier datasets, aiming to draw conclusions from the couples as they reported themselves in 2009.

Using this data, we hope to examine the factors associated with relationship satisfaction and longevity, including:

1. How has the way couples meet changed over time?
2. What factors predict whether a respondent met their partner online?
3. How do people select partners based on religion and race?

2 Data

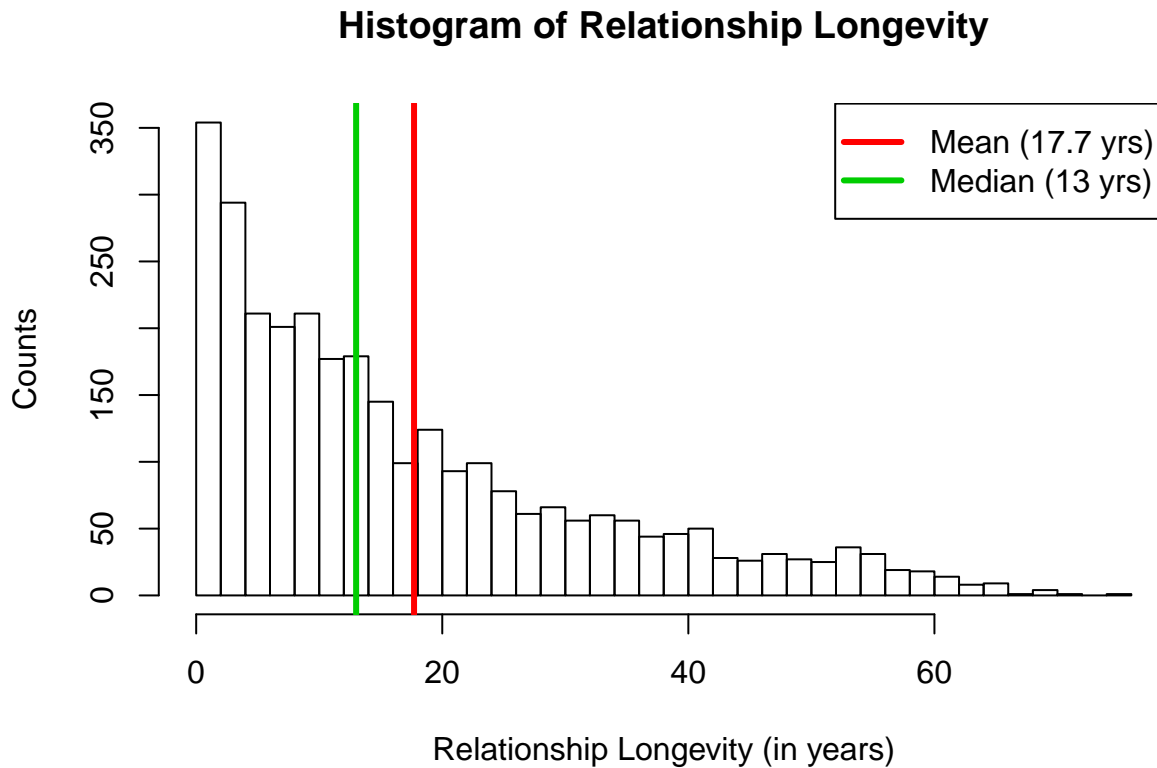
2.1 Variable Descriptions

1. PPAGE: respondent age at time of HCMST wave I survey (from 19-95); integer: range from 19-95
2. PPGENDER: respondent gender; binary: range from 1-2 (1 = male, 2 = female)
3. CHILDREN_IN_HH: number of children in household; integer: range from 0-7
4. PPMARIT: marital status; categorical factor: married, widowed, divorced, separated, never married, living with a partner
5. HHINC: dollar value household income; integer: range from 2500-200000
6. PPWORK: current employment status; categorical factor: working – as a paid employee, working – self-employed, not working – on temporary layoff from a job, not working – looking for work, not working – retired, not working – disabled, not working – other
7. PPEDUCAT: educational status of respondent; categorical factor: less than high school, high school, some college, bachelor’s degree or higher
8. PAPRELIGION: identified religion of respondent; categorical factor: Baptist, Protestant, Catholic, Mormon, Jewish, Muslim, Hindu, Buddhist, Pentecostal, Eastern Orthodox, other Christian, other non-Christian, None
9. PPPARTYID3: political party affiliation; categorical factor: republican, democrat, independent, another party, no preference
10. RESPONDENT_RACE: race that the respondent identifies with; categorical factor: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic American Indian, Non-Hispanic Asian/Pacific Islander, Non-Hispanic Other, Hispanic
11. PARTNER_RACE: race that the respondent’s partner identifies with; categorical factor: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic American Indian, Non-Hispanic Asian/Pacific Islander, Non-Hispanic Other, Hispanic
12. RESPONDENT_YRSED: educational attainment of respondent; numeric: range from 0-20
13. PARTNER_YRSED: educational attainment of respondent’s partner; numeric: range from 0-20
14. W4_ATTRACTIVE: respondent’s rated attractiveness of themselves; integer: range from 1-4 (4 = very attractive, 3 = moderately attractive, 2 = slightly attractive, 1 = not at all attractive)
15. W4_ATTRACTIVE_PARTNER: respondent’s rated attractiveness of their partner; integer: range from 1-4 (4 = very attractive, 3 = moderately attractive, 2 = slightly attractive, 1 = not at all attractive)
16. HOW_LONG_RELATIONSHIP: current relationship duration; integer: range from 0-76
17. GENDER_ATTRACTION: sexual preference; categorical factor: opposite gender only, mostly opposite, both genders equally, same gender mostly, only same gender
18. SAME_SEX_COUPLE: sexual preference; binary: 0 = opposite-sex couple, 1 = same-sex couple
19. PARENTAL_APPROVAL: parental approval or disapproval of respondent’s relationship with partner; binary: 0 = disapproval or unknown, 1 = approval
20. RELATIONSHIP_QUALITY: respondent’s assessment of relationship quality with partner; integer: range from 1-5 (1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = excellent)
21. HOW_LONG_RELATIONSHIP: length of respondent’s most recent relationship; integer: range from 0-76
22. AGE_DIFFERENCE: absolute difference in age between the respondent and their partner; integer: range from 0-69
23. RESPONDENT_RELIG_16_CAT: respondent’s identified religion at 16; categorical factor: Protestant or oth Christian, Catholic, Jewish, Neither, No religion
24. PARTNER_RELIG_16_CAT: partner’s identified religion at 16; categorical factor: Protestant or oth Christian, Catholic, Jewish, Neither, No religion
25. HOW_LONG_AGO_FIRST_MET: time since respondent first met partner; numeric: 0-76
26. HOW_LONG_AGO_FIRST_ROMANTIC: time since respondent was first romantically involved with partner; numeric: 0-76
27. MET_?: whether the respondent met their partner in the specified fashion; binary: 1 = Yes, 0 = No
28. YEAR_MET: indicates the year that the respondent first met their partner; numeric: 1933-2009

2.2 Exploratory Plots

2.2.1 Histogram of Relationship Longevity

How long do people typically stay together? This question may be answered by looking at the distribution of relationship lengths for polled respondents.

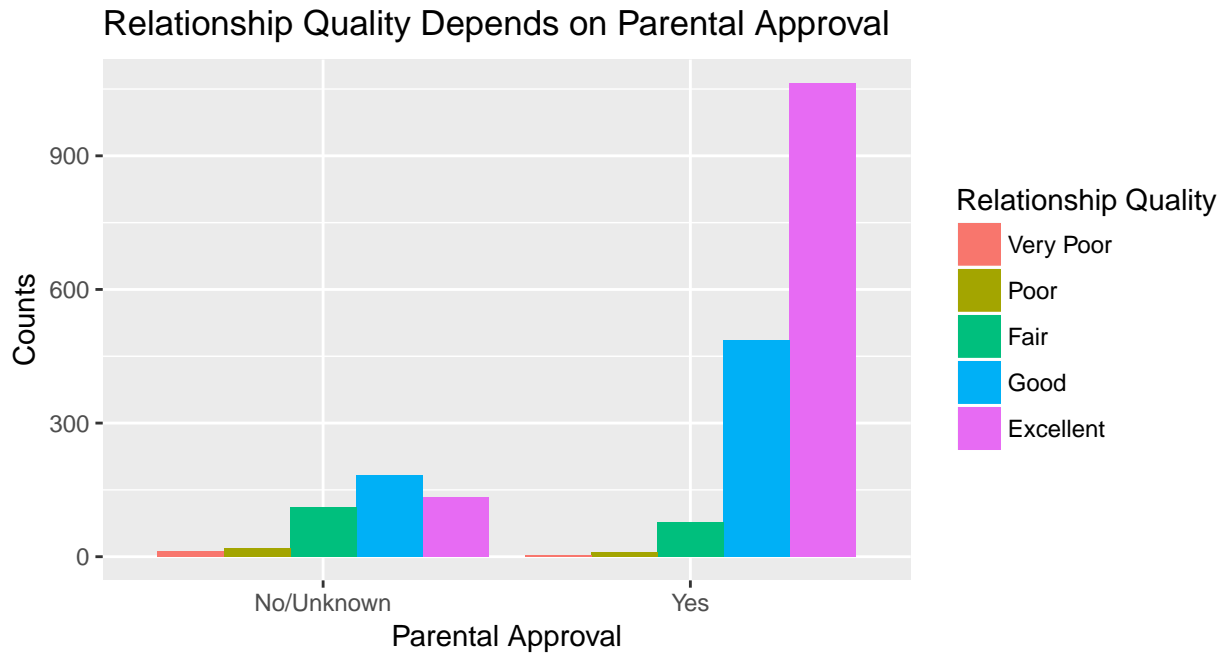


By plotting the relationship longevity of 2983 respondents, we find that most of the relationships have been short-term, and fewer relationships last longer. In fact, the median and mean of the respondents' relationship length are 13 and 17.7 years, respectively, the difference attesting to the right-skew. The relationship length in years for increments of 10% of respondents are given by:

##	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
##	0.0	1.5	4.0	7.0	10.0	13.0	17.0	22.4	30.0	42.0	76.0

2.2.2 Parental Approval and Relationship Quality

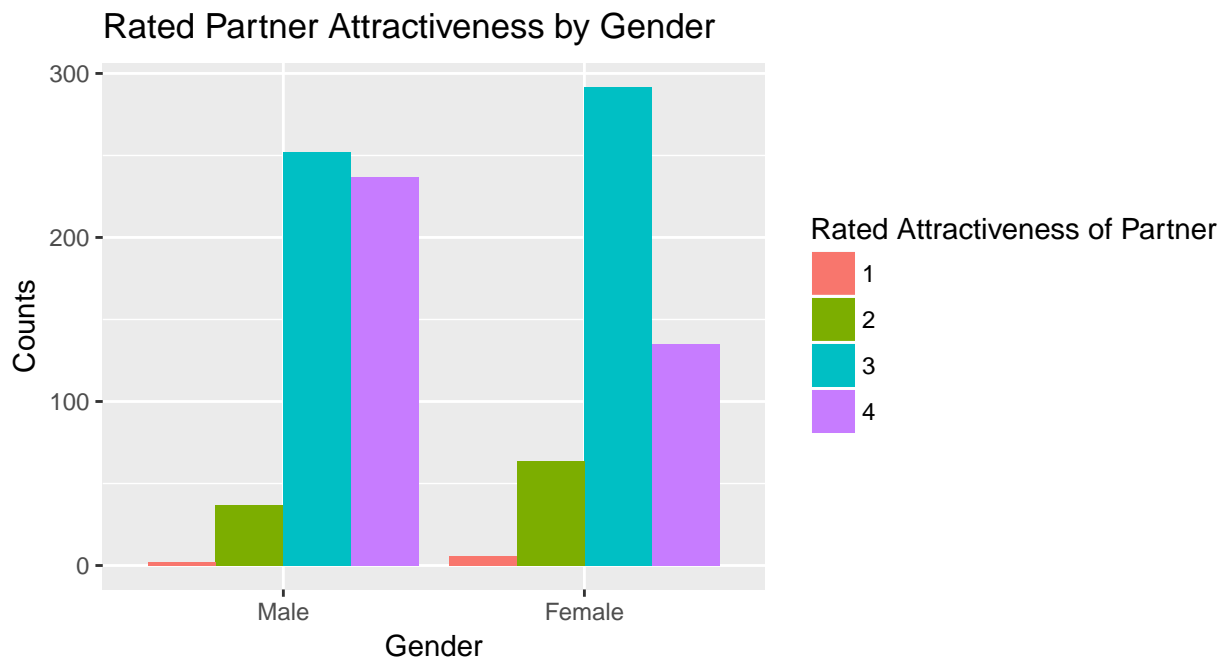
A preliminary linear model showed that parental approval was the strongest linear predictor of relationship quality, among the variables described. This can be visualized by plotting the distribution of relationship quality for the two groups.



Couples with parental approval are much more likely to rate their relationship quality as “excellent” (65%), compared to those without or with unknown parental approval (29%). In fact, those without or with unknown parental approval report a higher percentage of “good” relationships than “excellent” relationships. This result is interesting but requires a better understanding of confounding variables to verify.

2.2.3 Importance of Attractiveness by Gender

We were interested to find that the attractiveness rating of the partner was a survey question. We aimed to stratify this by various demographics, and found that the most robust difference was by gender.



It is clear that men are much more likely to rate their partners as more attractive, and women are much more likely to rate their partners as less attractive. In fact, men are 66% more likely to rate their partners as

“very attractive”, while women are 3 times more likely to rate their partners as “not attractive at all.” This suggests that men are much more likely to value attractiveness in a partner as see this as a precondition for a relationship.

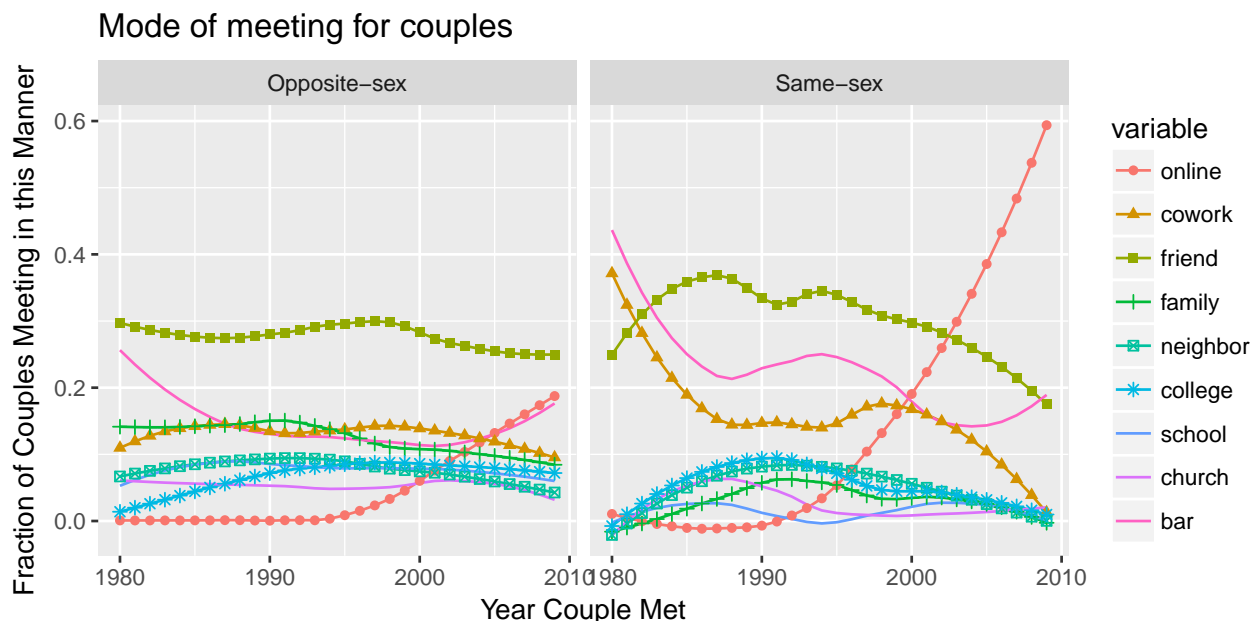
3 Analysis

3.1 Methods of Meeting and the Rise of Online Dating

There are many ways that couples meet, such as traditional ways, including through family, friends, the church or the neighbourhood, or non-traditional ways, such as online. With the rise of the Internet, communications within existing social networks have become more prevalent, and the efficiency of searching for and finding new people outside of one’s pre-existing social network has also increased drastically. Indeed, the number of users of online dating platforms - such as Tinder - has been on the rise.

As the dataset contains data from couples who met at diverse time points, we decided to analyze how the ways in which couples meet have changed over time. This is based on their response to the question “How did you meet partner_name?”, and the year they met was calculated from the variable “How long ago romance with partner began”. The data was then smoothed with lowess regression, using a bandwidth of 0.8. Friends, family and co-workers can belong to either respondent or partner. The meeting trends for both same-sex and opposite-sex couples from 1980 to 2009 are shown below.

Note: N = 2462 for heterosexual couples (coded as 1 in the SAME_SEX_COUPLE variable), N = 462 for same-sex couples (coded as 2 in the SAME_SEX_COUPLE variable).



For couples who met in 1990 or before, there is essentially 0 percentage who met online, since the World-Wide Web has not been invented. Between 1995 and 2005, there is an exponential growth in the proportion of respondents who met their partners online. For opposite-sex couples, 20% of them meet online in 2009. This rise is even greater in same-sex couples, as more than 60% of the couples now meet online by 2009. We can see that online meeting has become the dominant and preferred method of meeting for same-sex couples, maybe because individuals are less inhibited online and are more likely to reveal that they are looking for a same-sex relationship.

For opposite-sex couples, other than the rise of online platform as a way to meet, the other ways of how couples meet have remained steady from 1980 to today. Meeting romantic partners through friends still

makes up a dominant portion of our respondents, hovering at about 30%. Meeting through school, family, neighbor, church or coworkers have remained constant. There is a slight drop in the percentage of people who meet at bars, maybe because the internet has replaced going to bars as a new way of meeting strangers.

For same-sex couples, there is much more variability in the trend, which can be attributed to the smaller sample size. We can see that there is a sharp drop in the fraction of couples who met at bar or met through coworkers, maybe because individuals are now more sensitive towards the same-sex community, and are less willing to recommend potential partners.

3.2 Exploring correlations between meeting online and political party, education level, and parental approval

In the previous section, we noted a dramatic increase in the number of couples that met online, especially for same-sex couples. This finding motivated us to further explore relationships between meeting online and other variables, such as religion, political party, and education level. Specifically, we were interested in the distribution of percentages of couples of different religions, political parties, and education levels who either (1) met online or (0) did not meet online. We could visualize this “distribution” of different percentages through the use of tables.

Let us first look at the distribution of couples who met online or offline among different political parties.

```
##          PPPARTYID3
## MET_ONLINE (1) republican (2) other (3) democrat
##          0          0.40578078 0.02402402    0.57019520
##          1          0.31111111 0.01111111    0.67777778
```

From this table, it looks like, of the people who meet their partners online, more are democrats as compared to republicans (67.8% vs. 31.1%). Furthermore, there is an increase in representation of democrats who meet online as opposed to offline. Using Pearson’s chi-squared test - which is a statistical test that looks at how likely observed difference between categorical variables occurs by chance - with two degrees of freedom, we find that this difference among democrats, republicans, and other political parties in meeting online and offline is statistically significant (with a p-value of 0.002226, which is statistically significant at a significance level of 0.05). This finding may be due to the fact that democrats are more liberal, and are thus more open to online dating.

Next, we investigated the relationship between meeting online and the respondent’s education level:

```
##          PPEDUCAT
## MET_ONLINE      1          2          3          4
##          0 0.11036036 0.25750751 0.28078078 0.35135135
##          1 0.03703704 0.14444444 0.38518519 0.43333333
```

Interestingly, there is more representation of highly educated people in the couples that meet online as compared to those who meet offline (p-value for Pearson’s Chi-Squared test of 1.67×10^{-8} with three degrees of freedom). In fact, of the people who meet their partners online, 43% are highly educated (4) as compared to not very educated (1), which represent about 0.03% of this total cohort of people. On the other hand, of the couples who do not meet online, the education levels are more evenly represented (35% vs 11% for the same two subgroups). This finding may be a result of access to technology. For example, the more educated someone is, the more likely they may be able to afford computers and smartphones, which consequently allow for easy access to online dating sites.

Finally, let us look at the correspondence between parental approval and meeting online:

```
##          PARENTAL_APPROVAL
## MET_ONLINE      1          2
##          0 0.2019651 0.7980349
##          1 0.3183857 0.6816143
```

The difference among groups here is statistically significant; Pearson's Chi-Squared test yields a p-value of 1.577e-8 with three degrees of freedom for this data. Of the people who meet online, there is a decrease in those with parental approval as compared to people who meet offline. This may be a consequence of less inhibition that is correlated in people who try online dating; due to this lack of inhibition, they may be less likely to take their parent's opinions into account. Therefore, parental approval may matter more for the people who meet offline as compared to those who meet than online.

3.3 Examining how religion and race affect partner selection

In this section, we will investigate whether couples tend to adhere to the same religious and racial groups when choosing partners.

3.3.1 Selecting only respondents involved in relationships at the time of the survey

As there was no variable that explicitly stated whether the respondent was in a relationship or not (the PPMARIT variable states marital status, but couples can be in a relationship and not necessarily married), we estimated people who were not in relationships as those respondents with missing values (i.e. NAs) in the partner columns. For example, only 0.26% of values were missing for the respondents while about 25% of values were consistently missing from the variables corresponding to partners. Therefore, we claim that these missing values may indicate that the respondent is single.

Using this conservative method, we selected for people who had or currently have partners with information on race and religion, which we will investigate further in the following sections.

3.3.2 Categorizing Religion

To explore whether respondents tend to end up with partners that are of the same religion, we created a binary variable, RELIGION_SAME that has value 0 if the couple does not believe in the same religion or 1 if they do indeed believe in the same religion. Note that if both respondent and partner are non-religious, this variable still counts them as being of the same religion.

The gap could be due to singles (simply have no partner). Hence it might be reasonable to simple exclude the columns with missing values for partner religion.

Creating the binary variable for same religion:

```
## [1] 0.5656971

##
##      0      1 <NA>
## 1299 1692      0

##
##                                     (1) Protestant or oth Christian
## (1) Protestant or oth Christian                                     1103
## (2) Catholic                                                         310
## (3) Jewish                                                            17
## (4) Neither Christian nor Jewish                                     24
## (5) No religion                                                       180
##
##                                     (2) Catholic (3) Jewish
## (1) Protestant or oth Christian                                     285    19
## (2) Catholic                                                         412    19
## (3) Jewish                                                            21    32
## (4) Neither Christian nor Jewish                                     9     1
```

```

##      (5) No religion                92          5
##
##                                     (4) Neither Christian nor Jewish
##      (1) Protestant or oth Christian                31
##      (2) Catholic                                17
##      (3) Jewish                                   4
##      (4) Neither Christian nor Jewish              13
##      (5) No religion                              23
##
##                                     (5) No religion
##      (1) Protestant or oth Christian                131
##      (2) Catholic                                80
##      (3) Jewish                                   9
##      (4) Neither Christian nor Jewish              14
##      (5) No religion                              132
## [1] 0.5656971

```

First we investigate how many people form couples within the boundaries of their religion. Please note that the category ‘other’ is somewhat problematic since by this measure hindu’s marrying muslims is of same religion. However the category of ‘others’ forming couples with ‘others’ is small so we choose to ignore this complication.

```
## [1] 0.5656971
```

I.e. 57% of couples form couples within their religion. However, this number doesn’t tell us much on its own. To make sense of it we can compare it to what we would expect in a sample where people do not care about the religion of their partners at all.

To do this we run a bootstrap/ permutation test. Where we randomly sample the religion of the partner based on the original sample.

Results from the test

```

## [1] 0.5656971
## [1] 0.3833071
## [1] 0.4758325

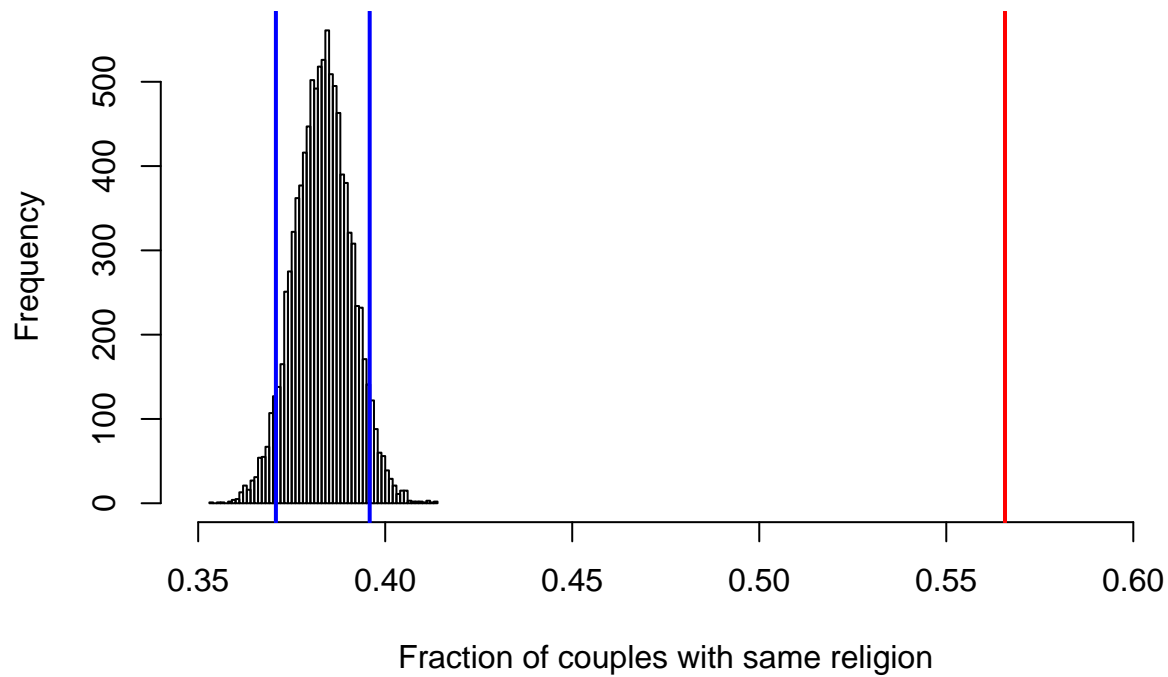
```

Hence 48 % more couples are within the same religion than we would expect from a random assignment.

How confident are we in this result being significantly different from the null-hypothesis? As the following shows - we are basically as confident as we can get. In the 10,000 simulations not a single one randomly generated a distribution that skewed. As the diagram shows

```
## [1] 1
```


Histogram of randomised couples



3.3.3 Race

3.3.4 Categorising Race

Same kind of variable is created for race:

```
## [1] 0.8288198
```

```
##
##               (1) NH white (2) NH black (3) NH Amer Indian
## (1) NH white           2151          27          19
## (2) NH black            32         178           2
## (3) NH Amer Indian      13           1           5
## (4) NH Asian Pac Islander 24           3           1
## (5) NH Other            17           4           2
## (6) Hispanic            89          19           1
##
##               (4) NH Asian Pac Islander (5) NH Other
## (1) NH white                31          16
## (2) NH black                 1           3
## (3) NH Amer Indian           0           0
## (4) NH Asian Pac Islander    24           1
## (5) NH Other                 0           4
## (6) Hispanic                 5           2
##
##               (6) Hispanic
## (1) NH white                156
## (2) NH black                 24
## (3) NH Amer Indian           3
```

```
## (4) NH Asian Pac Islander 7
## (5) NH Other 6
## (6) Hispanic 117
## [1] 0.8288198
```

Then we carry out same analysis for race.

```
## [1] 0.8288198
```

I.e. 82% of couples form within the same race, which may sound like a lot, however this number on its own without a comparison does not tell us much. Hence we compare it to the case where the partners' race is randomly sampled.

Doing Permutation Test On Race

The results from our simulation were as follows:

```
## [1] 0.8288198
```

```
## [1] 0.6393305
```

```
## [1] 0.2963872
```

I.e. the number of couples that stick within their race is 30% larger than we would expect from a completely randomly coupled sample.

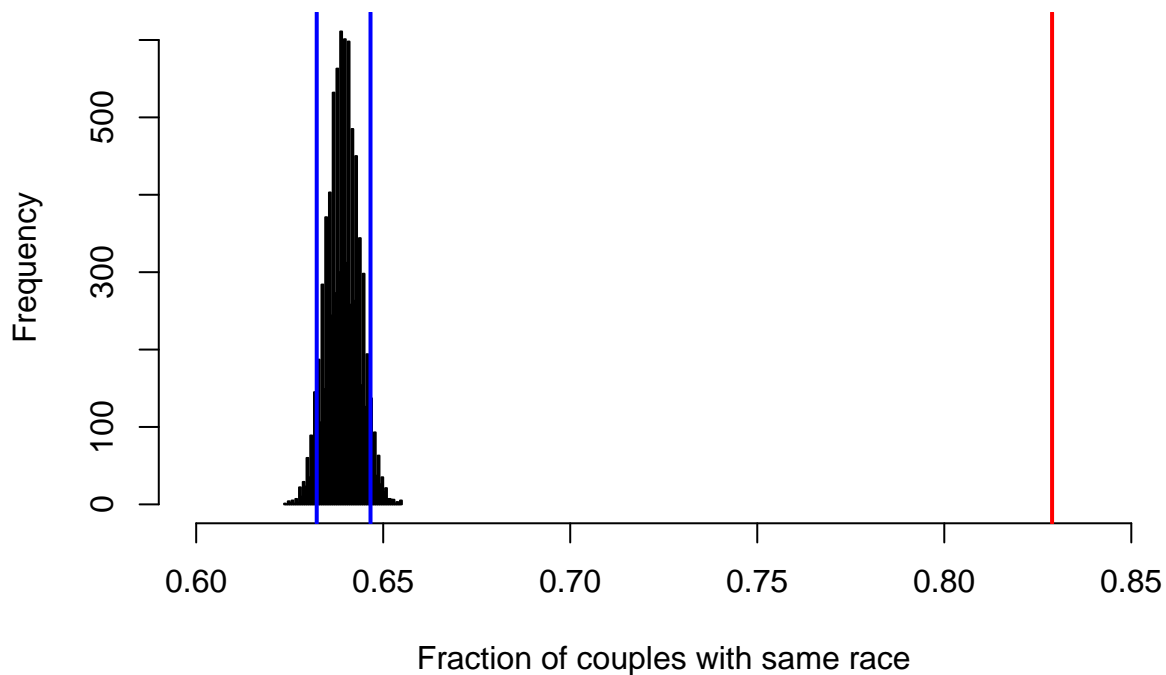
Confidence:

```
## [1] 1
```

```
##      5%      95%
```

```
## 0.6322300 0.6466065
```

Histogram of randomised couples



Again as suggested by the histogram and the empirical continuous distribution function not a simple sample of our 10,000 simulations were as skewed as the observed distribution. I.e. we are very confident that people

are more likely to form couples within their race than the randomised sample would suggest.

(here could include break down of different religions if space?)

```
## [1] (6) Hispanic          (1) NH white
## [3] (2) NH black            (3) NH Amer Indian
## [5] (5) NH Other            (4) NH Asian Pac Islander
## [7] <NA>
## 6 Levels: (1) NH white (2) NH black ... (6) Hispanic
## [1] 0.9247635
## [1] 0.07523646
## [1] 0.7776663
## [1] 0.7672414
## [1] 0.2327586
## [1] 0.07756603
## [1] 0.1666667
## [1] 0.8333333
## [1] 0.01003009
## [1] 0.3934426
## [1] 0.6065574
## [1] 0.02039452
## [1] 0.3738019
## [1] 0.6261981
## [1] 0.1046473
```