Econ/Demog C175 Lab 11: Marriage – Do Opposites Attract, or Do Likes Like Likes?

In this lab we will look at patterns of intermarriage among traditional opposite-sex couples and more recent same-sex couples in California. Our data comes from the American Community Survey from 2015. The Census recently changed its coding so that same-sex married couples would not have their responses edited to "non-married." (See http://www.pewresearch.org/fact-tank/2014/05/29/census-says-it-will-count-same-sex-marriages-but-with-caveats/)

```
# Do not edit this chunk, but *do* press the green button to the answer key for the quiz info (the unre
tot = 0
answer.key = "eJytUrtuwzAM3PMVhJe0gCHUj7VD0MeYAGn6AYpD2ORtKRHlpP77Sorj1vWaRQJJ8Hh3pDiprt2jSUQCz+DehSi0M
library(quizify)
source.coded.txt(answer.key)
```

Reading in data

```
df <- read.csv("/data175/ca_marriage_out.csv", as.is = T)
nrow(df)
```

```
## [1] 34214
```

```
## we see there are about 34,000 couples in our sample
```

Let's look at the variables

```
head(df)
```

```
##      sex sex_sp  x x_sp   college college_sp racere racere_sp incwage
## 1   male female 41   32   college  nocollege      w         w   50000
## 2 female   male 34   34   college    college      w         w       0
## 3   male female 58   49 nocollege  nocollege      w         w   29000
## 4 female   male 45   46 nocollege  nocollege      w         w   67000
## 5 female   male 43   47 nocollege  nocollege      w         w   95000
## 6 female   male 58   56 nocollege  nocollege      w         w   30000
##   incwage_sp
## 1          0
## 2     230000
## 3          0
## 4      67000
## 5          0
## 6      50000
```

Each line is a couple. Variables ending in "_sp" are for the spouse of the person answering the questionnaire. Only one person per household filled out the questionnaire.

- 'x' is age in years
- 'college' takes the value "college" for those with a 4-year degree and "nocollege" for those without.
- 'racere' is a recode of a more detailed race variable into Asian ('a'), Black ('b'), and White ('w').

- 'wageinc' is wage income in the previous 12 months in dollars

To make our analysis easier, we've restricted the sample to - California only - respondents and spouses aged 19-59 - races: Asian, White and Black

Let's just check to see if there are same sex marriages.

```
head(df[df$sex == df$sex_sp,])
```

```
##        sex sex_sp  x x_sp   college college_sp racere racere_sp incwage
## 38    male   male 46   46   college    college      w         w   73000
## 105   male   male 45   47   college  nocollege      w         w   50000
## 248 female female 54   55 nocollege  nocollege      w         a   75000
## 296 female female 39   19 nocollege  nocollege      w         w   16000
## 387   male   male 24   22   college    college      w         w       0
## 395   male   male 54   53   college    college      b         b       0
##     incwage_sp
## 38      150000
## 105     160000
## 248      60000
## 296          0
## 387          0
## 395          0
```

Let's save the variables in a form that's easier to type

```
sex = df$sex
sex_sp = df$sex_sp # sex of spouse
age = df$x
age_sp = df$x_sp
college = df$college
college_sp = df$college_sp
incwage = df$incwage
incwage_sp = df$incwage_sp
race = df$racere
race_sp = df$racere_sp
```

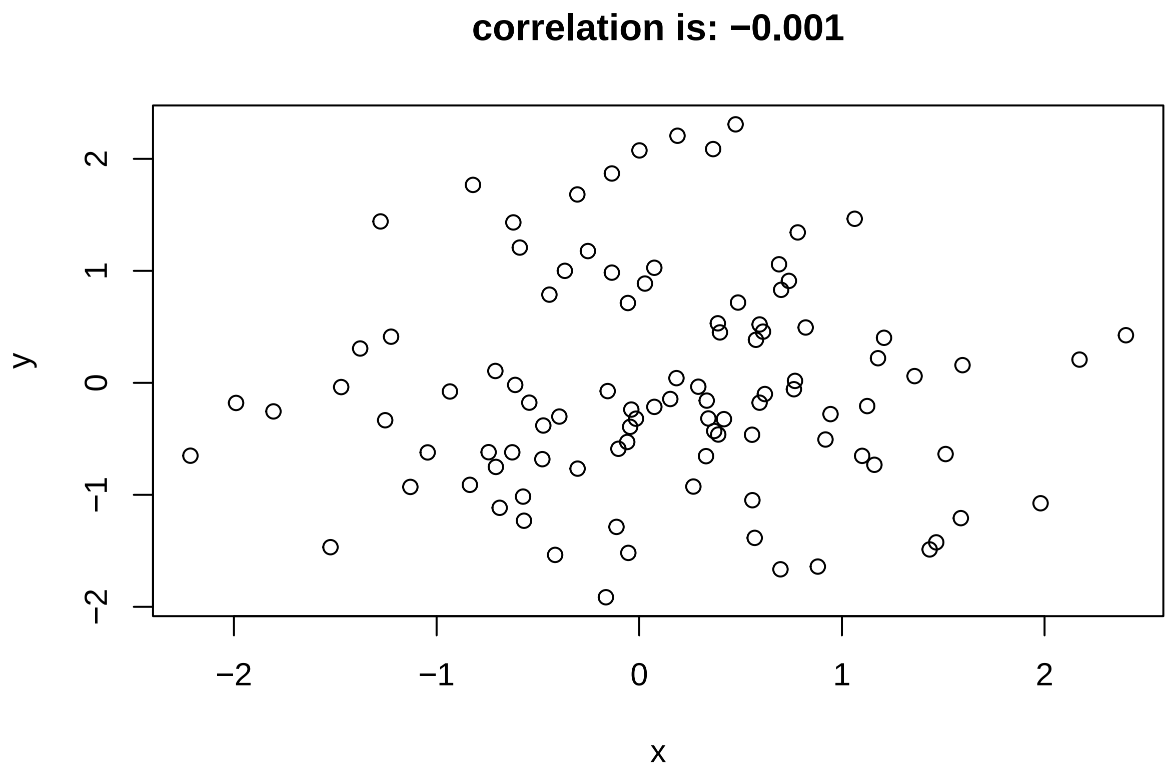
Measuring positive and negative assortative mating

Correlation: for a continuous variable taking numeric vales

The correlation between two vectors of numbers x and y tell you how the x & y are "co-related". A value of 1 means that x and y are perfectly aligned, so that like is with like. A value of -1 means that they are negatively related, so that like and unlike pair together. A value of 0 means that they are unrelated to each other.

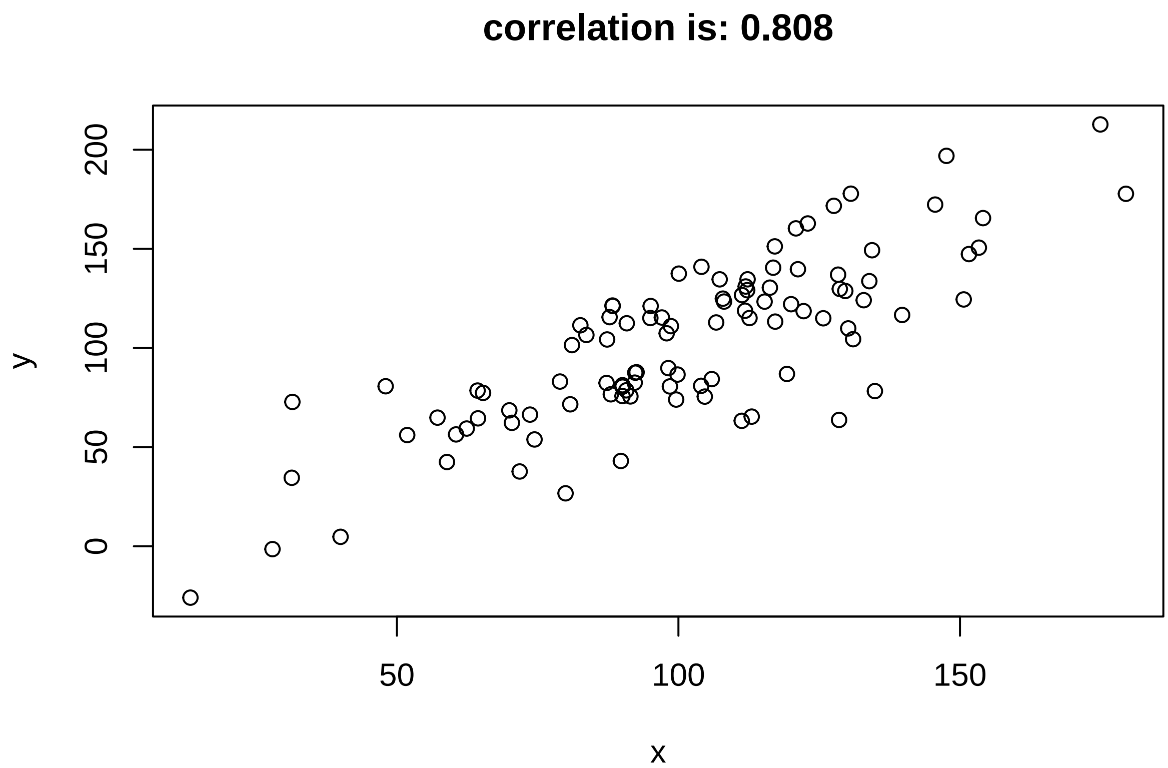
An example of zero correlation would be two independent sets of randomly generated numbers

```
set.seed(1)
x = rnorm(100)
y = rnorm(100)
plot(x,y)
title(paste("correlation is:", round(cor(x,y),3)))
```



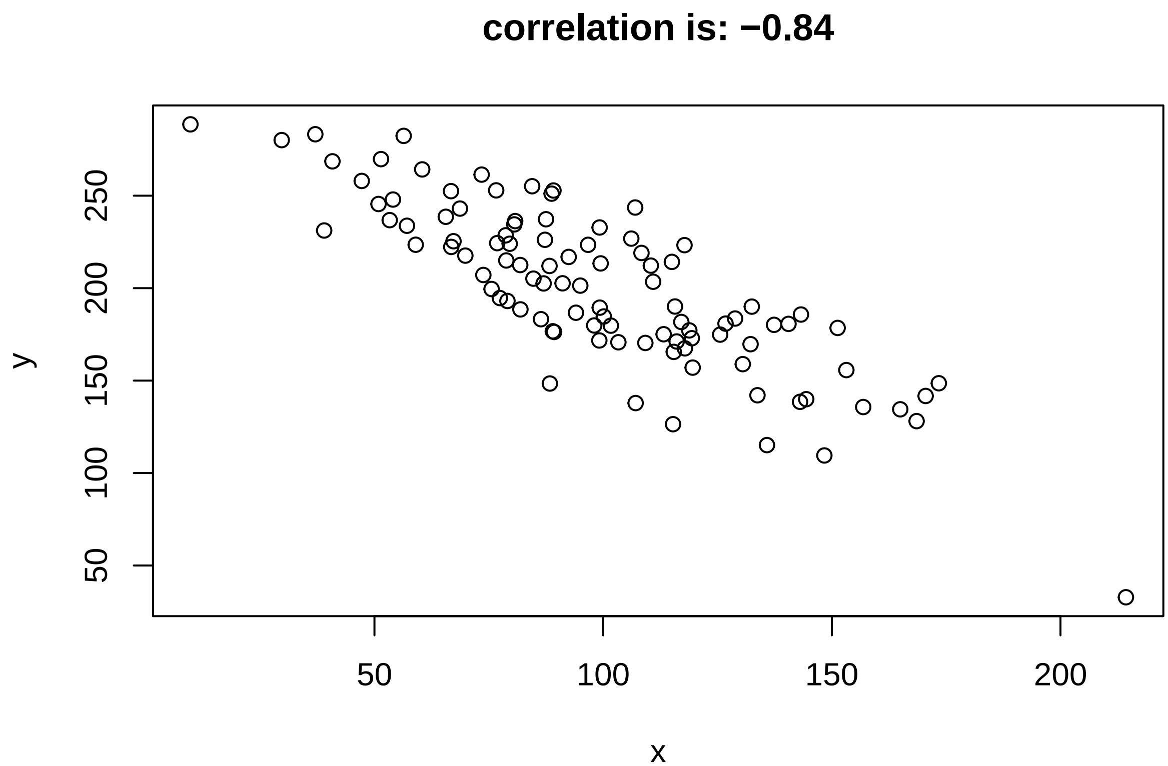
An example of positive correlation

```
x = rnorm(100, mean = 100, sd = 30)
epsilon = rnorm(length(x), mean = 0, sd = 25)
y = x + epsilon
plot(x,y)
title(paste("correlation is:", round(cor(x,y),3)))
```



An example of negative correlation

```
x = rnorm(100, mean = 100, sd = 30)
epsilon = rnorm(length(x), mean = 0, sd = 25)
y = 300 -x + epsilon
plot(x,y)
title(paste("correlation is:", round(cor(x,y),3)))
```



Q1.1 What is NOT true about the correlation between married partners?

A. A negative value means opposites attract.

B. A positive value means like prefers like.

C. A value of 0.5 means that there is no tendency for married partners to be similar.

D. A value of 0 means there is no tendency for married partners to be similar.

```
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.1 = "C"
quiz.check(answer1.1)
```

```
## Your  answer1.1 : C
## Correct.
## Explanation:  The only NOT TRUE answer is C
```

Measuring association in a 2x2 table

Correlation is useful for continuous numeric variables. Sometimes we have variables that take on only two values, which we call binary. The relationship between two binary variables can be summarized with the odds ratio.

Say we have a 2x2 table of the sexes of spouses in California

```
table(sex, sex_sp)
```

```
##         sex_sp
## sex      female  male
##   female    215 13402
##   male    20366   231
```

To measure the association we can calculate the odds ratio, which is defined as

$$\theta = \frac{(F[1,1]/F[1,2])}{(F[2,1]/F[2,2])}$$

In our case this would be

```
theta = (215/13402) / (20366/231)
print(theta)
```

```
## [1] 0.0001819596
```

The odds-ratio can take values from 0 to positive infinity. A value of 1 indicates a neutral relationship, with no positive or negative sorting. A value close to 0 indicates very negative sorting; high values (10, 100, or even 1000) indicate very positive sorting.

The reason that this is called the odds-ratio is because the numerator is the odds of being in cell [1,1] compared to [1,2], and the denominator is odds of being in cell [2,1] compared to [2,2]. If there were no tendency to marry one sex rather than the other, the odds of marrying a "male" would be the same for females and males and the ratio would be 1.

Here's an example of a table with an odds ratio of 1.

3 5 6 10

A value of $\theta > 1$ tells us that there is positive association, as with a correlation greater than 0. An example would be

10 3 2 20

Q1.2: The odds ratio in this table is: A. 200/ 6 B. 5 / 200 C. 20/ 60 D. 30 / 40

```
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.2 = "A"
quiz.check(answer1.2)
```

```
## Your  answer1.2 : A
## Correct.
## Explanation:  The odds ratio is 33.3, indicating positive association.
```

A value of $\theta < 1$ tells us that there is negative association. An example is our above table of spouses by sex in California. We expect a number that is much less than 1, because we know that most marriages in California are between people of opposite sexes.

Here's a function to calculate the odds ratio

```
get.odds.ratio <- function(my.table)
{
  if (!all(dim(my.table) == c(2,2)))
    stop("table needs to be 2x2")
  theta = (my.table[1,1]/ my.table[1,2]) / (my.table[2,1] / my.table[2,2])
  return(theta)
}
```

Let's check our calculation of the odds ratio of our table by sex

```r
sex.tab <- table(sex, sex_sp)
print(sex.tab)
```

```
##         sex_sp
## sex      female  male
##   female    215 13402
##   male    20366   231
```

```r
sex.theta <- get.odds.ratio(sex.tab)
print(paste("theta = ", round(sex.theta, 5)))
```

```
## [1] "theta =  0.00018"
```

Q1.3: Which of the following is true for the odds ratio found for the table of marriages by combination of sex. A. There's no tendency for marriages to be same- or opposite- sex. We know this because the odds ratio is close to 0. B. There's some tendency for marriages to be same-sex. We know this because the odds ratio is positive. C. There's a strong tendency for marriages to be opposite-sex. We know this because the odds ratio is much less than 1.

```r
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.3 = "C"
quiz.check(answer1.3)
```

```
## Your  answer1.3 : C
## Correct.
## Explanation:  Odds ratios go from 0 to positive infinity, with 1 signifying
## .no relationship.
```

Sorting for opposite-sex married couples

Here we will look at assortative pairing for opposite-sex couples. At the end of the lab, you will repeat this exercise for same-sex couples to see if same-sex spouses appear to be more similar to each other (or more different) than opposite-sex couples.

1. Education

Our variable "college" tells us if the person has a 4-year college degree. The predictions from Becker's theory are ambiguous. On the one hand, we would expect education to be associated with market wages, which Becker thinks should be negatively correlated across spouses. On the other hand, there are many reasons to think that education should positively sort, including complementarity in investments in children and complementary consumption (of books, films, and conversation). And then there is the issue of search costs: since future spouses often meet in an educational or work setting, and since these are highly sorted by education, we might expect likes to marry likes simply out of convenience, even if there were no preference for assortative marriage in education.

In order to compare same-sex to opposite sex marriages, we create an index variable.

```r
ss <- sex == sex_sp ## index for "same sex" ("ss")  couples
os <- !ss ## index for "opposite sex" ("os") couples
```

We can now look at the association of education for opposite sex couples

Now we can construct two tables, one for the college status of same-sex couples and one for the college status of opposite sex couples

```r
college.tab.os <- table(college[os], college_sp[os])

print(college.tab.os)
```

```
##
##              college nocollege
##   college      11342      4890
##   nocollege     2973     14563
```

```r
print(get.odds.ratio(college.tab.os))
```

```
## [1] 11.36153
```

Q1.4: Would you agree or disagree with the following statement

> We see an odds ratio of 11.4, which is much bigger than one, showing strong positive sorting between opposite-sex spouses.

A. Agree B. Disagree

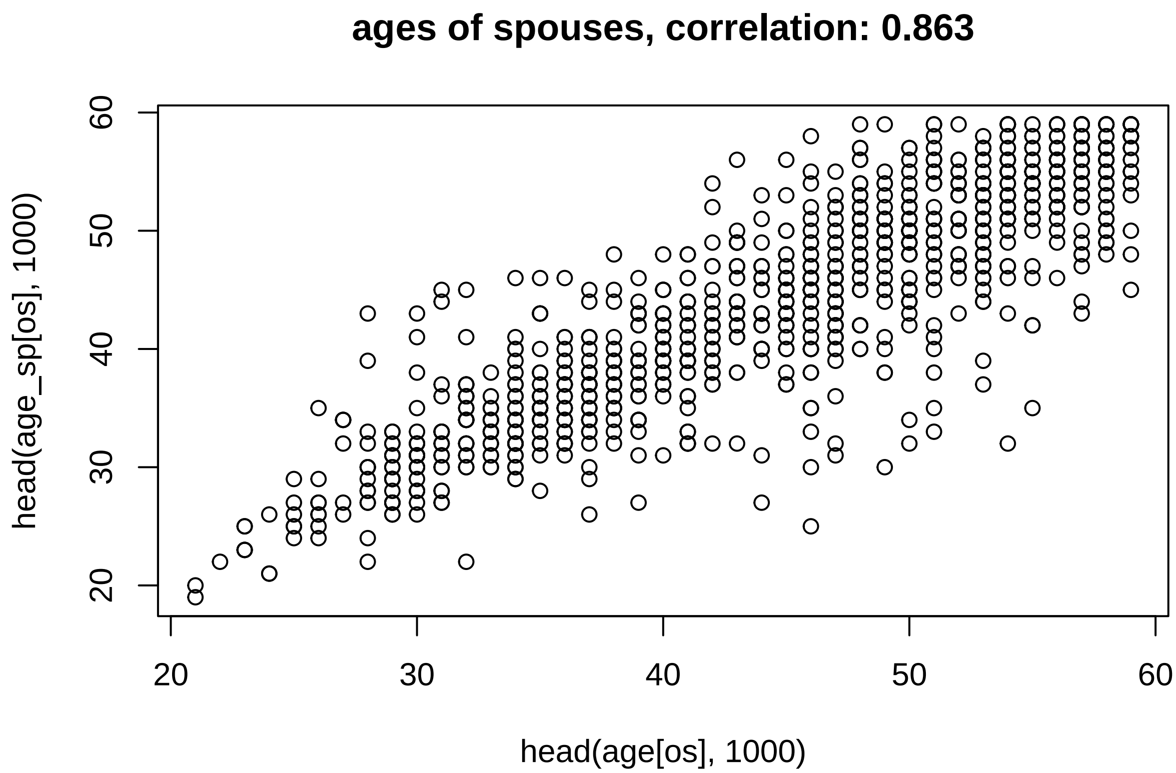
```r
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer1.4 = "A"
quiz.check(answer1.4)
```

```
## Your  answer1.4 : A
## Correct.
## Explanation:  11.4 is much greater than 1.
```

2. Age

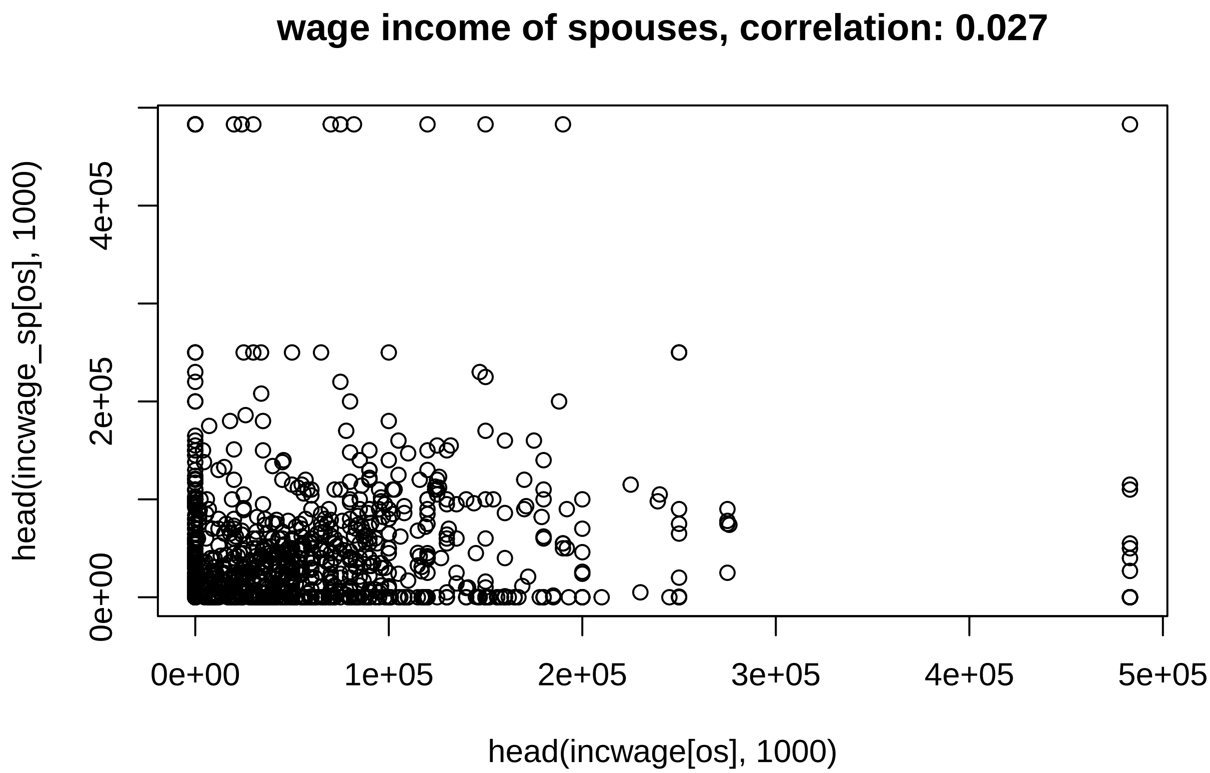
We expect ages to be similar between spouses both because of the relative ease of meeting people of one's own age group and also because age will be correlated with a number of other traits including cultural taste, political outlook, energy level, and other characteristics that we would expect to positively sort.

```r
plot(head(age[os], 1000), head(age_sp[os], 1000)) # we only plot first 1000 points
title(paste("ages of spouses, correlation:", round(cor(age[os], age_sp[os]),3)))
```



3. Wage income

```
plot(head(incwage[os], 1000), head(incwage_sp[os], 1000))
title(paste("wage income of spouses, correlation:",
            round(cor(incwage[os], incwage_sp[os]),3)))
```

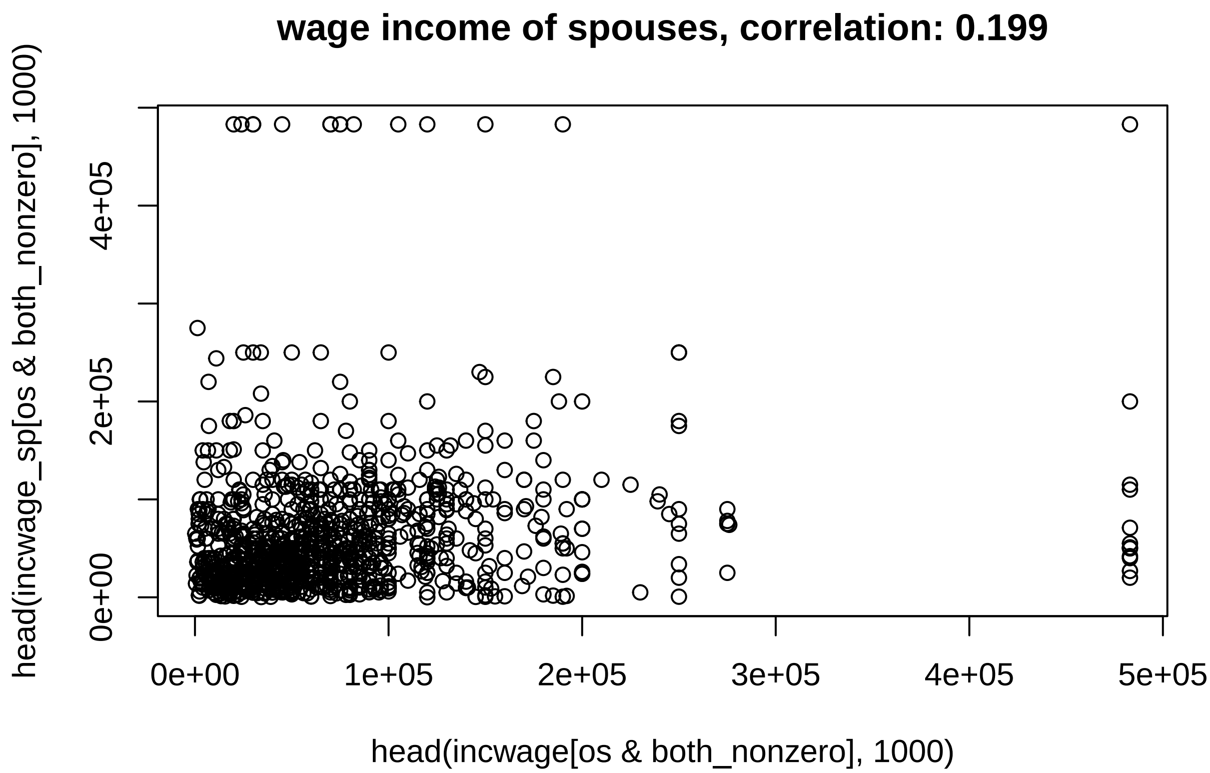


Note: The data is top-coded, so that all values above a certain level are assigned the same value, in this case \$483,000, which is the average of the top-coded values.

Here we see a more complicated story. We see a large number of points in which one of the spouses has zero income and the other a positive income, so this is consistent with the specialization argument of Becker. But overall there is no strong positive or negative relationship.

We could plot only couples where both partners had income greater than zero.

```
both_nonzero = incwage > 0 & incwage_sp > 0
plot(head(incwage[os & both_nonzero], 1000),
     head(incwage_sp[os & both_nonzero], 1000))
title(paste("wage income of spouses, correlation:",
            round(cor(incwage[os & both_nonzero],
                      incwage_sp[os & both_nonzero]),3)))
```



But Becker would probably say this is uninformative, since where there are big differences in potential earnings, one spouse will choose to specialize in home production and have zero earnings. So if we leave out the zero earners, we're missing the important part of the story.

Perhaps a better test would be to see if zero earnings tend to be only in one spouse.

```
zerowage = incwage == 0
zerowage_sp = incwage_sp == 0
zerowage.tab.os <- table(zerowage, zerowage_sp)
print(zerowage.tab.os)
```

```
##          zerowage_sp
## zerowage FALSE  TRUE
##    FALSE 19704  7991
##    TRUE   4625  1894
```

```
print(get.odds.ratio(zerowage.tab.os))
```

```
## [1] 1.009768
```

Surprisingly, we get an odds ratio close to 1. So even here, we don't see negative assortative marriage. We could further refine the analysis, say by considering people of a particular age range. But the bigger point is probably that potential wages before marriage, not observed wages after marriage, are probably what people are considering when they choose a spouse. Education is a great measure of potential wages, because it is observed both for those in and out of the labor force.

3. Race

Our data set has been restricted to people who identify as Asian (“a”), Black (“b”), or White (“w”). In order to calculate the association, we will consider three possible 2x2 combinations (“a” & “w”), (“b” & “w”), and (“a” & “b”)

First we make our tables of the race of spouses of each 2x2 combination:

```
os.wb <- os & race %in% c("w", "b") & race_sp %in% c("w", "b")
os.wa <- os & race %in% c("w", "a") & race_sp %in% c("w", "a")
os.ab <- os & race %in% c("b", "a") & race_sp %in% c("b", "a")
race.wb.tab.os <- table(race[os.wb], race_sp[os.wb])
print(race.wb.tab.os)
```

```
##
##         b     w
##   b   839   155
##   w   245 24023
```

```
race.wa.tab.os <- table(race[os.wa], race_sp[os.wa])
print(race.wa.tab.os)
```

```
##
##         a     w
##   a  6636   655
##   w  1117 24023
```

```
race.ab.tab.os <- table(race[os.ab], race_sp[os.ab])
print(race.ab.tab.os)
```

```
##
##        a    b
##   a 6636   43
##   b   55  839
```

We see that most marriages in all three tables are “same race”. But it is hard to tell by looking at the numbers which pairs of races have the highest tendency to in-marry and inter-marry. For this we use the odds-ratio.

```
print(get.odds.ratio(race.wb.tab.os))
```

```
## [1] 530.7517
```

```
print(get.odds.ratio(race.wa.tab.os))
```

```
## [1] 217.8909
```

```
print(get.odds.ratio(race.ab.tab.os))
```

```
## [1] 2354.167
```

So it appears that crossing the White-Asian racial divide is easier than crossing the White-Black racial divide, and both are easier than crossing the Asian-Black racial divide.

Conclusions for opposite-sex spouses

Our findings are consistent with Becker’s statements about positive sorting by race, age, and education, but not particularly supportive of the negative assortative pattern he predicted (in the 1970s) for market wages.

Same-sex couples

Becker's main interest is in explaining why opposite-sex couples form. But his theory has implications for same-sex couples. He writes

> Households with only men or only women are less efficient because they are unable to profit from the sexual difference in comparative advantage.
>
> Source: Becker (1991), *Treatise_on_the_Family*, p. 37-38, cited by Jepsen and Jepsen(2001).

This inability to "profit from the sexual difference in comparative advantage" could mean that same-sex couples would seek to profit more from non-sexual differences, perhaps in demographic characteristics like age or education. Or it might mean that same-sex couples may derive their utility less from comparative advantages in production and more from complementarities in consumption.

Another factor to think about is limited choice. If one is looking for a partner in a large pool, then it may be easier to find exactly what one wants. On the other hand, if there is a decisive constraint – like the person must be over 6 feet tall or the person must be interested in a same-sex marriage – then a less than ideal match on another characteristic might well be worth it.

Replicating our analysis for same-sex couples

We can use the "ss" index we created to look at assortative marriage patterns for same sex couples. For example, for education, we can calculate

```
college.tab.ss <- table(college[ss], college_sp[ss])
print(college.tab.ss)
```

```
##
##             college nocollege
##   college       202        74
##   nocollege      49       121
```

```
print(get.odds.ratio(college.tab.ss))
```

```
## [1] 6.740761
```

We see that our odds ratio of 6.7 for same sex couples is much less than the odds ratio of 11.4 observed for opposite-sex couples. This means that there appears to be a greater tendency to cross educational divides for same-sex couples.

Graded Questions

0. Do you expect same-sex marriage spouses to be more like each other or less like each other than opposite-sex spouses? Is your expectation the same for age, race, income and education? If not, specify further. Explain your reasoning. Make sure to write down somewhere your expectations BEFORE you do the analysis. (You won't be graded on this. The goal is for you to get the most out of your data analysis and to prevent you from coming up with a purely post hoc explanation after you've seen the result.)

I expect them to be more like each other. There is tradeoffs in sex and gender roles, so I feel like you'll see more similarities between the same sex.

1. Replicate our analysis of opposite-sex marriages for same-sex marriages. Complete the (5) cells marked "X" in the following table. Transfer your answers to bCourses.

Characteristic	Opposite-sex	Same-sex	Opposite-sex	Same-sex
Measure	odds ratio	odds ratio	correlation	correlation
Education	11.4	6.7		
Age			0.86	0.73
Wage Income			0.03	0.12
Race ("w","b")	531	17		
Race ("w","a")	218	38		
Race ("a","b")	2354	inf		

2. How does positive sorting differ for same-sex spouses, compared to opposite-sex spouses? (For example, "The smaller odds ratios for education mean that same-sex marriages sort less positively on education than do opposite-sex marriages."). [Describe the remaining 5 relationships in a sentence each.]

- The infinity odds ration for Asian and Black for same sex marriage sort much more twards race than oppposite sex marriage, meaning that it is near impossible for same sex marriage to happen in same sex compared to opposite sex.
- The smaller odds ratio for White and Asian mean that same sex marriage sort less positively than opposite sex, meaning you'll see more White and Asian same sex marriage than white-asian opposite sex marriage.
- White and Black mean that same sex marraiges sort less positively on same sex marraige, thus, same effect as above.
- For wage income, there is higher correlation for same sex marriage, but since it's still near 0, there is no positive or negative relationship
- For age, there is higher correlation in general, but same sex has lower correlation. This means it is less likely for same sex marraige to be around the same age compared to opposite sex.

3. Describe a potential explanation for the pattern of your results. Try to use the words "complementarity" and "comparative advantage" in your answer. If some characteristics sort more and some less, try to make sense of the pattern. (There is no wrong explanation here. Just make sure your explanation is consistent with your evidence.) [< 75 words]

It seems like the data and analysis shows that the relationship of opposite sexes seems to be more complimentary for each other compared to same sex. The preference for same gender, age and education is apparent in opposite sex, but wage seems to indication that there was no comparative advantage and was not a strong factor in determining which role served better.

4. We didn't distinguish same-sex *male* couples from same-sex *female* couples. If we did, which would you expect to have a less positive correlation in wage income: same-sex female couples or same-sex male couples? Explain briefly why. [< 25 words] (Hint: you don't need to calculate anything, just explain your logic. There is no single right answer.)

I would expect same sex male couples to have less positive correlation in wage income. This is because of today's wage gap in men and women.

5. Let's extend the example we did in class (on about slide 20 of the lecture on assortative pairing). Let the payoffs have the form

$$Z(F_i, M_j) = (F_i^r + M_j^r)^{1/r}$$

and let $F_1 = M_1 = 1$, $F_2 = M_2 = 4$, and $F_3 = M_3 = 5$, with $r = 3$. Here the subscripts $1, 2, 3$ refer to some characteristic, which could be income, personality type, ethnic background, religion, education, height, or anything else. When a marriage is formed of people with the same-subscript, they have the same value of that characteristic, as we can see from the assignments made above.

a) Fill in missing cells of the 3×3 matrix of payoffs.

```
M1    M2    M3
```

F1 1.26 4.02. 5.01 F2 4.02 5.04 5.74 F3 5.01 5.74 6.30

b) What would be the optimal sorting? (Hint: choose the sorting that maximizes the sum of the payoff of all three marriages.) Is this an example of "like marrying like," or is this an example of "opposites attracting"?

M1F3 5.01 M2F1 4.02 M3F2 5.74 Sum 14.77

c) Tell a story of why the marriage pattern might look this way if the characteristic were "college major". (For example, if the example showed likes marrying likes, then you would want to give a reason why a marriage of people from the same college major would have a higher "payoff" than from different majors). Please limit your answer to a few sentences.

Marriage pattern might look this we if the characteristic was a college major as people with the same or similar major will often talk about academia or industry, thus they'll often collaborate together and get a higher payoff. Different majors will have a lower payoff as you'll often have passions or ideas that could be completely orthogonal to each other and don't compliment each other.

Congratulations! You are finished with Lab 10.