

# Relationship Survey Data

DATA 557 Final Project

March 11, 2020

Andreia Sodre Nichols

Karl Stavem

Maggie Weatherly

Natasha Halfin

Sanjana Gupta

# Dataset Overview

- **Survey: “How Couples Meet and Stay Together (HCMST)”**
- Surveyed 4,002 U.S. respondents over five waves from 2009 to 2015
- Each respondent reported on their demographic data as well as their partner's, in addition to rating relationship quality
  - i.e. one row = one individual representing the couple
- Respondents recruited by random digit dial phone survey (randomized selection)
- Heterosexual and LGBTQ respondents sampled
- Coupled, uncoupled, married, unmarried, etc. respondents sampled

# Assumptions/Caveats

- Focusing only on wave one data
- Reduced project dataset to filter out respondents who were not in a relationship (25% of population)
- May or may not be representative of broader U.S. population
- LGBTQ populations oversampled
- Self-reported responses; we assume accuracy, but may not always be the case

# Research Questions

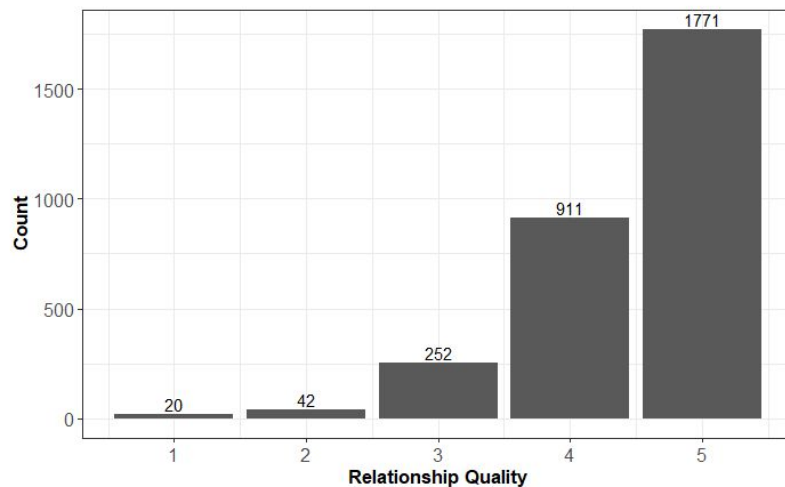
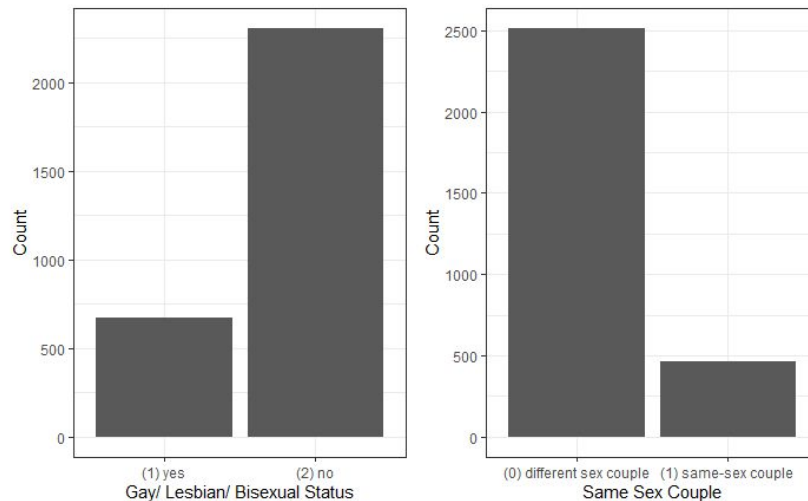
We want to understand how the following variables affect relationships:

**Race, education, and income**

**Metric for evaluation:** relationship quality (scale of 1-5, ranging from poor to excellent)

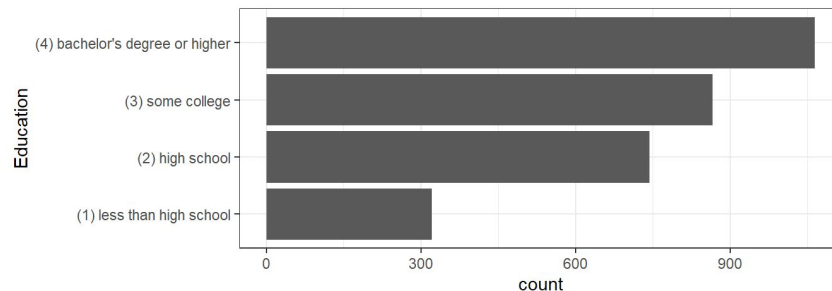
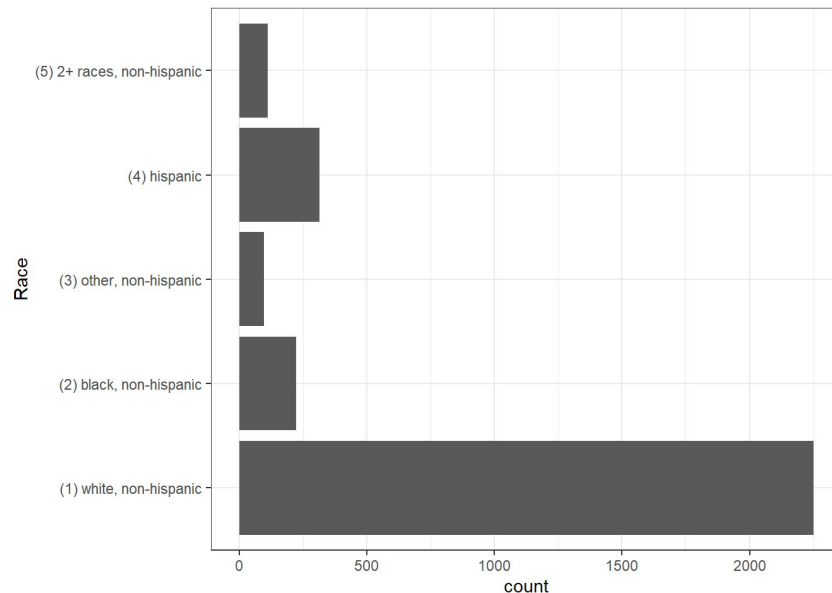
# Descriptive Statistics

- 1/5 respondents identify as LGBTQ
- 15% of our research population are part of a same sex couple
- Even split between males and females
- Majority of respondents rated relationship quality at 5 (excellent); self-selection bias?



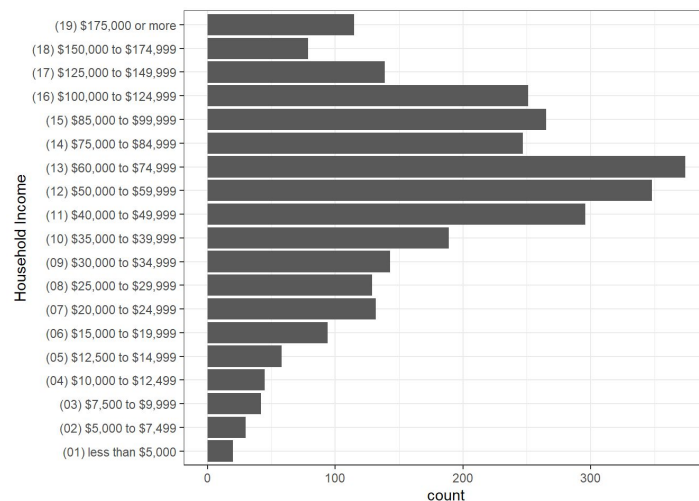
# Descriptive Statistics

- Significantly higher representation of white respondents; skews population
- Respondents with bachelor's degree or higher have the highest representation
- Still reasonable representation across different education levels



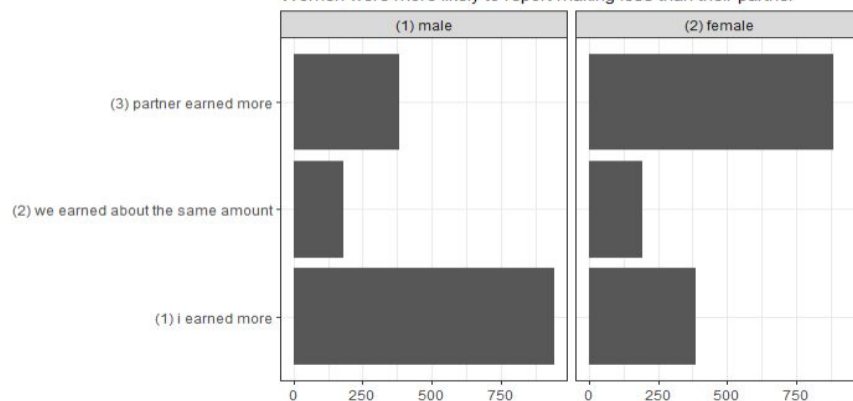
# Descriptive Statistics

- Combined household income is mostly normally distributed; ideal
- Centered around \$75K bin
- When looking at couples with income disparity, women were more likely to report earning less than their partner



## Earnings among Partners in the US

Women were more likely to report making less than their partner



# Race

We looked at three questions:

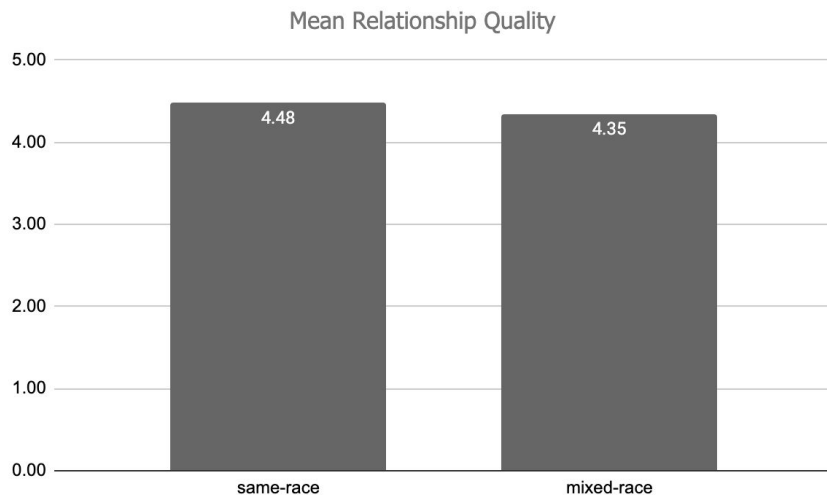
- Do **same-race couples** report higher quality of relationship than **mixed-race**?
- Do **same-race couples report equal level** quality of relationship?
- **Which same-race couple reports higher** quality of relationship?



# Race

Do same-race couples report higher quality relationship than mixed-race couples?

- **Null Hypothesis:** same-race quality  $\leq$  mixed-race quality
- **Alternative Hypothesis:** same-race quality  $>$  mixed-race quality



# Race

**Method:** Welch T-test

**Result:** p-value of 0.0002

- Therefore, we would reject the null hypothesis → **sufficient evidence to suggest that same-race couples appear to report higher relationship quality than mixed-race couples**

# Race

**Do same-race couples report equal levels of relationship quality?**

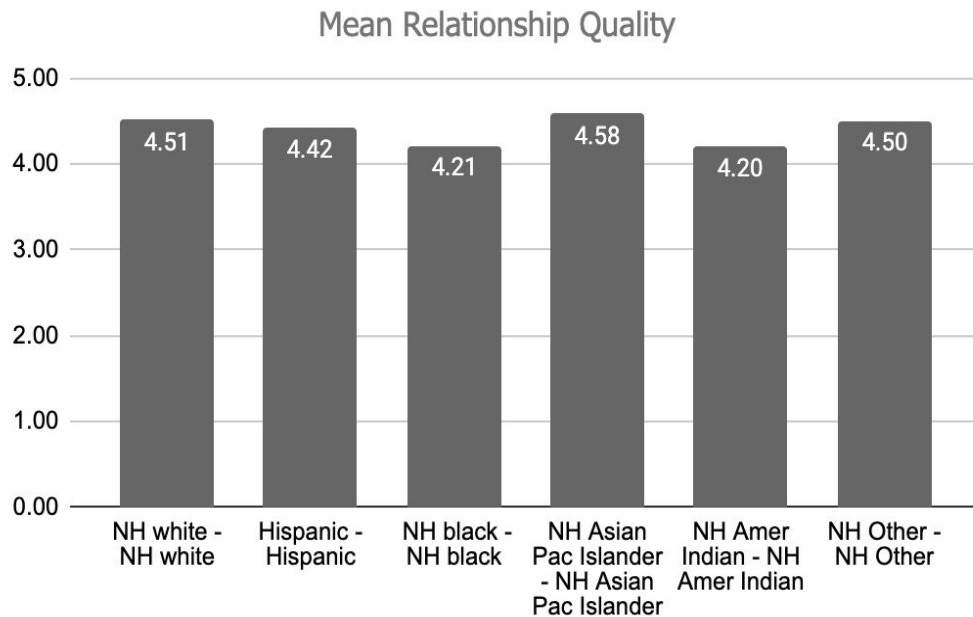
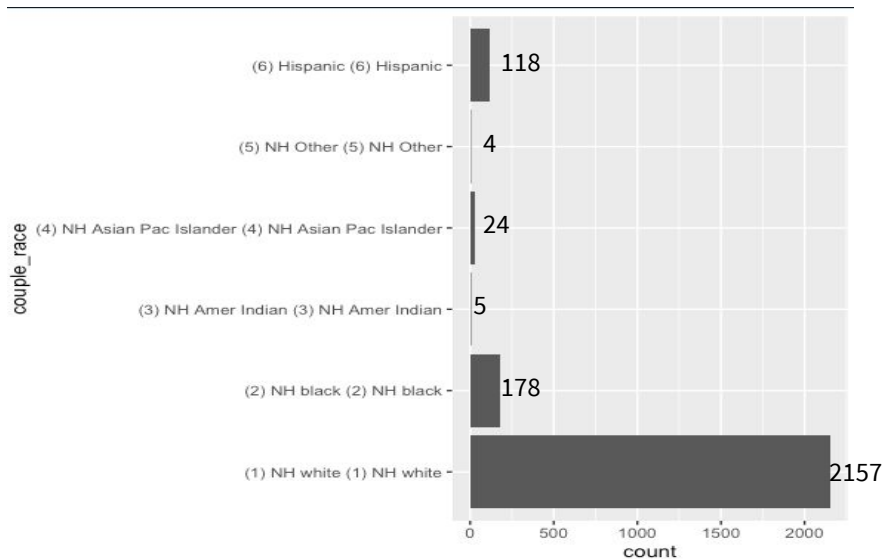
- **Null Hypothesis:** all types of same-race couples report the same mean relationship quality
- **Alternative Hypothesis:** at least one type of same-race couple does not report the same mean relationship quality

**Method:** ANOVA

**Result:** p-value of  $3.66e-05$  \*\*\*

- Therefore, we would reject the null hypothesis → **sufficient evidence to suggest that within same-race couples, race itself may have an effect on relationship quality.**

# Race



# Race

Which same-race couple reports higher quality of relationship?

- **Null Hypothesis:** [specific group] same-race quality  $\leq$  all other same-race quality
- **Alternative Hypothesis:** [specific group] same-race quality  $>$  all other same-race quality

**Method:** Welch T-test

Race	NH white	Hispanic	NH black	Asian Pac Islander	NH Amer Indian	NH Other
p-value	5.25E-05	0.817317	0.9999844	0.2046487	0.7548558	0.4878145

- Therefore, we would reject the null hypothesis (in green)→ **sufficient evidence to suggest that same-race white couples report higher quality of relationship**

# Education

We looked at three questions:

- **Does education disparity** between partners affect relationship quality?
- **How does the amount of education disparity** between partners affect relationship quality?
- **Does it matter who in the relationship has more education?** Does this affect relationship quality?

# Education

- **Null Hypothesis:** Mean relationship quality is no different for couples that have the same education vs. the couples that have different education
- **Alternative Hypothesis:** Mean relationship quality is different for couples that have the same education vs. the couples that have different education
- **Method:** Welch t-test to compare mean relationship quality between both groups
  - Mean relationship quality for couples that have same education : 4.477
  - Mean relationship quality for couples that have different education : 4.451
- **Result:** p-value of 0.3576
  - Therefore, we would fail to reject the null hypothesis→ **Insufficient evidence to suggest that education disparity has an effect on relationship quality**

# Education

- **Null Hypothesis:** The number of years of education disparity in a relationship does not have an effect on mean relationship quality
- **Alternative Hypothesis:** The number of years of education disparity in a relationship has an effect on mean relationship quality
- **Method:** Linear Regression with number of years of education disparity as predictor variable and relationship quality as target variable
- **Result:** coefficient: -0.0178, p-value: 0.000906
  - Therefore, we would reject the null hypothesis → **Sufficient evidence to suggest that number of years of education disparity has an effect on relationship quality**

Education disparity = Respondent Education - Partner Education



# Education

- **Null Hypothesis:** The absolute value of the number of years of education disparity in a relationship does not have an effect on mean relationship quality
- **Alternative Hypothesis:** The absolute value of the number of years of education disparity in a relationship has an effect on mean relationship quality
- **Method:** Linear Regression with absolute value of number of years of education disparity as predictor variable and relationship quality as target variable
- **Result:** coefficient: -0.0064, p-value: 0.3734
  - Therefore, we would not reject the null hypothesis→ **Not sufficient evidence to suggest that absolute value of number of years of education disparity has an effect on relationship quality**

Absolute value of education disparity =  $\text{abs}(\text{Respondent Education} - \text{Partner Education})$

# Income

We looked at three questions:

- Does combined household income have an effect on reported relationship quality?
- Do couples that make above \$75K (*Happiness Threshold*) report higher relationship quality than those who do not?
- What is the association between income disparity within couples and relationship quality?

# Income – Total Household Income

- **Null Hypothesis:** Total combined household income has no effect on mean reported relationship quality.
- **Alternative Hypothesis:** Combined household income has an effect on mean relationship quality.
- **Method:** Household income is reported in ranges (e.g. \$60,000 - \$74,999). Use ANOVA to model income range vs. relationship quality.
- **Result:**  $p\text{-value} = 0.000271$ 
  - Therefore, we would reject the null hypothesis → **There is sufficient evidence to suggest that overall household income has an effect on relationship quality**

# Income – Happiness Threshold

- **Null Hypothesis:** Mean relationship quality is no different for couples that earn greater than \$75k vs. those who earn less.
- **Alternative Hypothesis:** Mean relationship quality is greater for those in the higher income group.
- **Method:** One-sided, Welch T-Test comparing samples' mean relationship quality between both groups.
- **Result:**
  - Mean score for >\$75k group = 4.52. Mean relationship quality for <\$75k group = 4.42
  - p-value = .0003428
  - Therefore, we would reject the null hypothesis→ **There is sufficient evidence to suggest that overall household income above \$75k has an effect on mean relationship quality**

# Income – Income Disparity

- **Question:** Does relationship quality tell us anything about the odds of having an income disparity between partners?
- **Method:** Logistic regression using binary calculated field based on couples' responses:
  - GROUP ONE: *"We earned about the same amount."*
  - GROUP TWO: *"Partner earned more."* & *"I earned more."*
- **Result:**
  - Interpretation of the coefficient: With each unit increase in relationship quality, the odds of having an income disparity as a couple decreases by a factor of 0.881.
  - P-value = 0.09859. Therefore, **not a statistically significant result.**

# Conclusions

- Same-race couples appear to report higher relationship quality
- Race itself appears to have an effect on relationship quality
- Education disparity by itself does not necessarily have an effect on relationship quality, but which partner has more education in the relationship might
- Greater combined income does appear to correlate with relationship quality
- Earning more than \$75K appears to have a significant effect on relationship quality as well

**Essentially: start relationships with highly-educated individuals of the same race that make a lot of money\***

**\*Irony warning**

# Additional Analysis Opportunities

- Survival Analysis; to estimate how long relationships may last
- Look at change over time (i.e. compare responses across waves)
- Analyze further the effect of which partner has more education
- Is there any interaction between the variables we examined?

# Questions

?

---

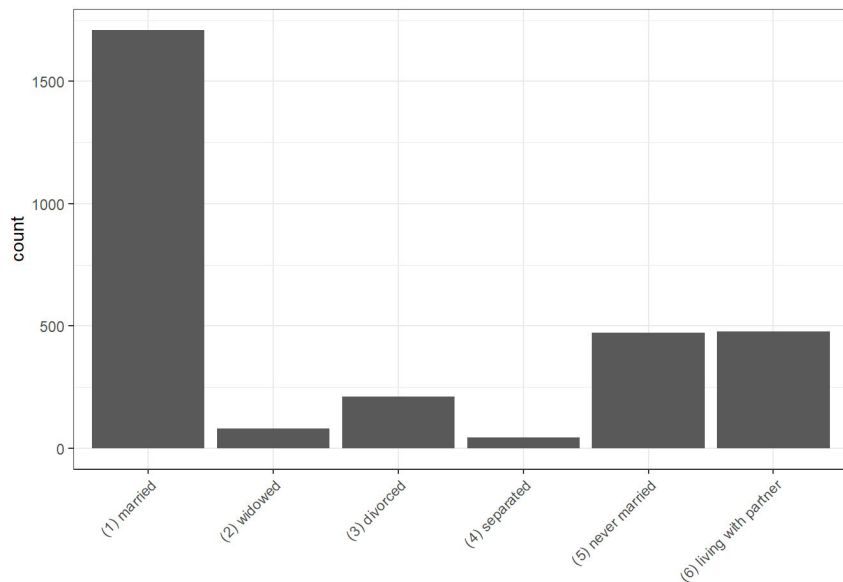


# Appendix



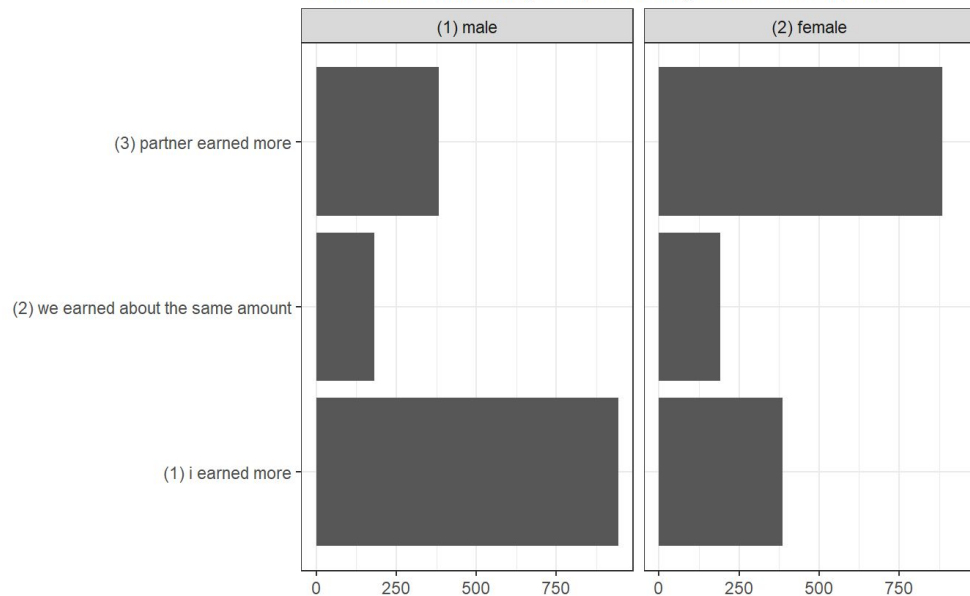
# Descriptive Statistics

Marital Status



Earnings among Partners in the US

Women were more likely to report making less than their partner



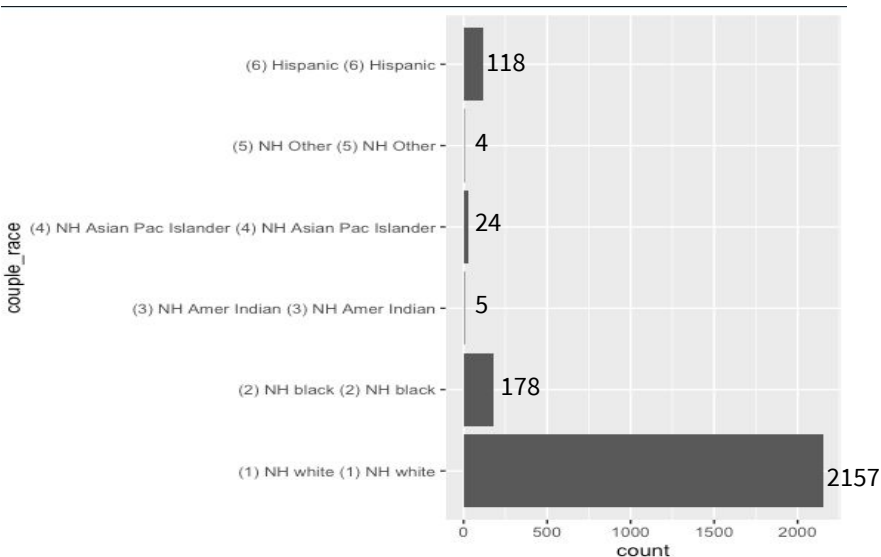
Source: HCMST 2017

# Race

Which same-race couple reports higher quality of relationship?

- **Null Hypothesis:** same-race A quality  $\leq$  same-race B quality
- **Alternative Hypothesis:** : same-race A quality  $>$  same-race B quality

**Method:** Pairwise T-test



# Race

Main / Comparison	NH white	Hispanic	NH black	Asian Pac Islander	NH Amer Indian	NH Other
NH white		0.1153555	1.22E-05	0.7294108	0.2278712	0.4935064
Hispanic	0.8846445		0.01831096	0.877174	0.2973202	0.5589214
NH black	0.9999878	0.981689		0.9952632	0.4898751	0.6976784
Asian Pac Islander	0.2705892	0.122826	0.004736823		0.1875794	0.4402512
NH Amer Indian	0.7721288	0.7026798	0.5101249	0.8124206		0.6758922
NH Other	0.5064936	0.4410786	0.3023216	0.5597488	0.3241078	

# Income – Income Disparity

Modeled using logistic regression for several reasons:

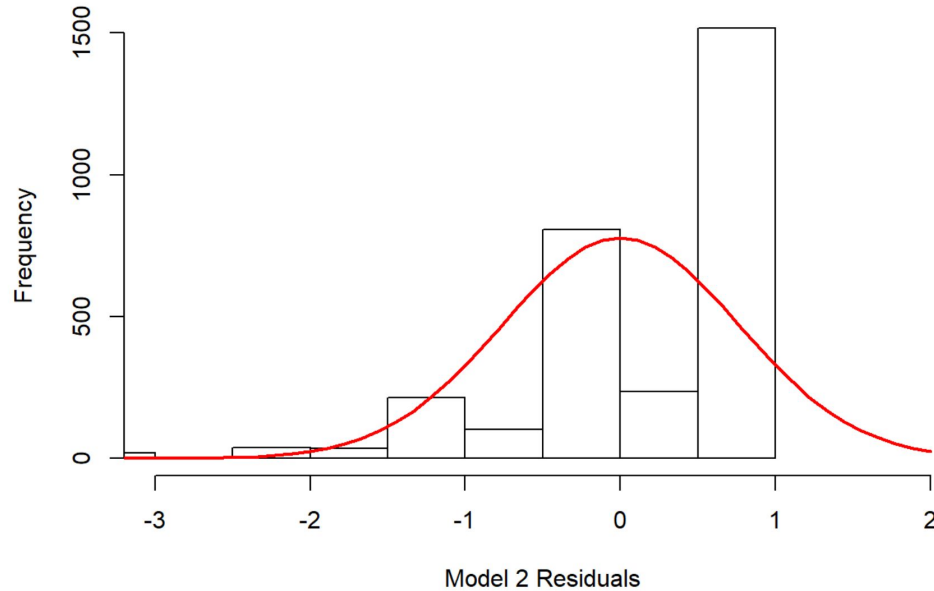
1. Logistic regression does not require a linear relationship.
2. The residuals do not need to be normally distributed.
3. Homoscedasticity is not required.

Requires that we flip the question around in order to have binary outcome variable:

- *“Does relationship quality tell us anything about the odds of having an income disparity between partners?”*

# Income – Income Disparity

- Data is *not* normal:



# Income – Happiness Threshold

- Testing the *Happiness Threshold*: Household income of at least \$75k

**BUSINESS** 07/17/2014 10:29 am ET | Updated Dec 06, 2017

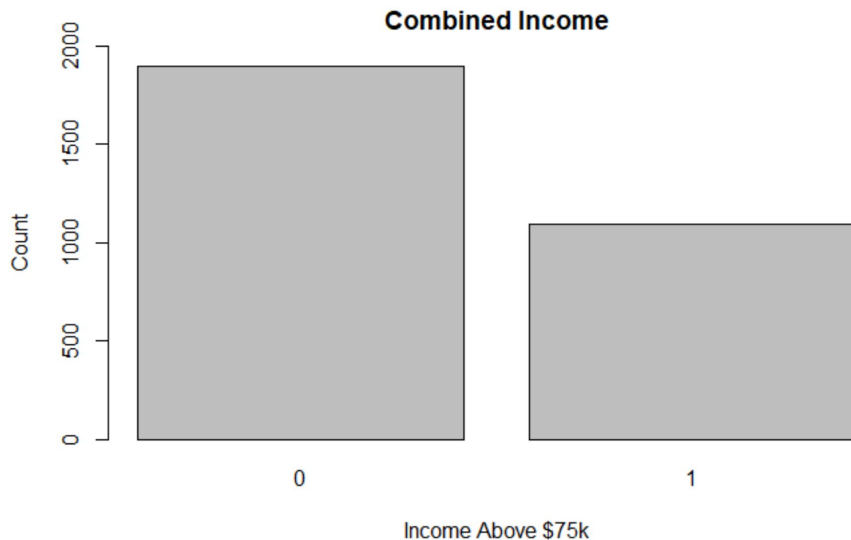
## Here Is The Income Level At Which Money Won't Make You Any Happier In Each State

By Kevin Short

Source: [https://www.huffpost.com/entry/map-happiness-benchmark\\_n\\_5592194](https://www.huffpost.com/entry/map-happiness-benchmark_n_5592194)

# Income – Happiness Threshold

- ≈1,900 couples earn more than \$75,000.
- ≈1,000 couples earn less.





# Education

- Used a Welch T-test on a calculated field:

```
ed_difference <- df$relationship_val[df$diff_in_education != 0]
no_ed_difference <- df$relationship_val[df$diff_in_education == 0]

modell <- t.test(ed_difference, no_ed_difference, var.equal = FALSE)
modell
```

```
##
##  Welch Two Sample t-test
##
## data:  ed_difference and no_ed_difference
## t = -0.92011, df = 2165.5, p-value = 0.3576
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08315131  0.03004213
## sample estimates:
## mean of x mean of y
##  4.450941  4.477495
```

# Education

```
df$diff_in_education = df$RESPONDENT_YRSED - df$PARTNER_YRSED  
summary(lm(relationship_val ~ diff_in_education, data=df))
```

```
##  
## Call:  
## lm(formula = relationship_val ~ diff_in_education, data = df)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -3.5485 -0.4592  0.5051  0.5408  0.7550   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    4.459199   0.013899  320.839 < 2e-16 ***  
## diff_in_education -0.017850   0.005374   -3.322 0.000906 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.7597 on 2987 degrees of freedom  
## (20 observations deleted due to missingness)  
## Multiple R-squared:  0.00368,    Adjusted R-squared:  0.003346   
## F-statistic: 11.03 on 1 and 2987 DF,  p-value: 0.000906
```

# Education

```
df$abs_diff_in_education = abs(df$diff_in_education)
summary(lm(relationship_val ~ abs_diff_in_education, data=df))
```

```
##
## Call:
## lm(formula = relationship_val ~ abs_diff_in_education, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4709 -0.4645  0.5291  0.5355  0.6215
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.470882   0.018511  241.53  <2e-16 ***
## abs_diff_in_education -0.006372   0.007158   -0.89   0.373
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.761 on 2987 degrees of freedom
## (20 observations deleted due to missingness)
## Multiple R-squared:  0.0002652, Adjusted R-squared: -6.947e-05
## F-statistic: 0.7924 on 1 and 2987 DF, p-value: 0.3734
```

# Education

- **Null Hypothesis:** The calculated categorical variable of which partner has more education (partner, respondent, same) does not have an effect on mean relationship quality
- **Alternative Hypothesis:** The variable of which partner has more education does have an effect on mean relationship quality
- **Method:** ANOVA with calculated categorical variable as predictor variable and relationship quality as target variable
- **Result:** p-value of 0.0028
  - Therefore, we would reject the null hypothesis→ **Sufficient evidence to suggest that which partner has higher education has an effect on relationship quality**

# Education

```
df$respondent_vs_partner = df$diff_in_education

df$respondent_vs_partner[df$diff_in_education < 0] = 'partner'
df$respondent_vs_partner[df$diff_in_education > 0] = 'respondent'
df$respondent_vs_partner[df$diff_in_education == 0] = 'same_ed'

summary(aov(relationship_val ~ factor(respondent_vs_partner), data=df))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## factor(respondent_vs_partner)    2    6.8   3.400    5.89 0.0028 **
## Residuals                2986 1723.7   0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 20 observations deleted due to missingness
```