

BST270_Final_Proj

2025-01-22

Reproducing “Marriage Isn’t Dead — Yet” Figures

The aim is to reproduce figures presented in the article published on FiveThirtyEight: *Marriage Isn’t Dead — Yet*. The article is available here: <https://fivethirtyeight.com/features/marriage-isnt-dead-yet/>.

Data are available via on the FiveThirtyEight GitHub (<https://github.com/fivethirtyeight/data/tree/master/marriage>), but are also included in the `data` folder of this repo. First we load the data:

```
# read in data
both_sexes_df <- read.csv("./data/both_sexes.csv")
divorce_df <- read.csv("./data/divorce.csv")
women_df <- read.csv("./data/women.csv")
```

Figure 1: Marriage Rates by Education and Race - Ages 25-34

We first use the `both_sexes_df` dataframe.

Next we simplify dataframes by removing unneeded columns (ie. columns unrelated to education level and race/ethnicity). We also drop na entries and convert divorce rates to marriage rates.

The columns we use are `HS_2534` - High school graduate or less (`EDUCD < 65`), age between 25 and 34, `SC_2534` - Some college (`EDUCD >= 65 & <= 100`), age between 25 and 34, and `BAp_2534` - Bachelor's degree or more (`EDUCD > 100`), age between 25 and 34 for the education plot and `White_2534` - Non-Hispanic white, age between 25 and 34, `Black_2534` - Black or African-American, age between 25 and 34, and `Hisp_2534` - Hispanic of any race for the race/ethnicity plot.

```
# Filter both_sexes.csv

# Education
# remove unneeded columns
both_sexes_edu_filt_df <- both_sexes_df[, c("year", "HS_2534",
                                             "SC_2534", "BAp_2534")]

# remove na's from HS_2534, SC_2534, BAp_2534 columns
both_sexes_edu_filt_df <- both_sexes_edu_filt_df[!is.na(both_sexes_edu_filt_df$HS_2534) |
                                                  !is.na(both_sexes_edu_filt_df$SC_2534) |
                                                  !is.na(both_sexes_edu_filt_df$BAp_2534),]

# convert to marriage rates
both_sexes_edu_filt_df$HS_2534 <- 1 - both_sexes_edu_filt_df$HS_2534
both_sexes_edu_filt_df$SC_2534 <- 1 - both_sexes_edu_filt_df$SC_2534
both_sexes_edu_filt_df$BAp_2534 <- 1 - both_sexes_edu_filt_df$BAp_2534

# Race & Ethnicity
# remove unneeded columns
```

```

both_sexes_race_eth_filt_df <- both_sexes_df[, c("year", "White_2534",
                                                "Black_2534", "Hispanic_2534")]
# remove na's from White_2534, Black_2534, Hispanic_2534 columns
both_sexes_race_eth_filt_df <- both_sexes_race_eth_filt_df[
  !is.na(both_sexes_race_eth_filt_df$White_2534) |
  !is.na(both_sexes_race_eth_filt_df$Black_2534) |
  !is.na(both_sexes_race_eth_filt_df$Hispanic_2534),]

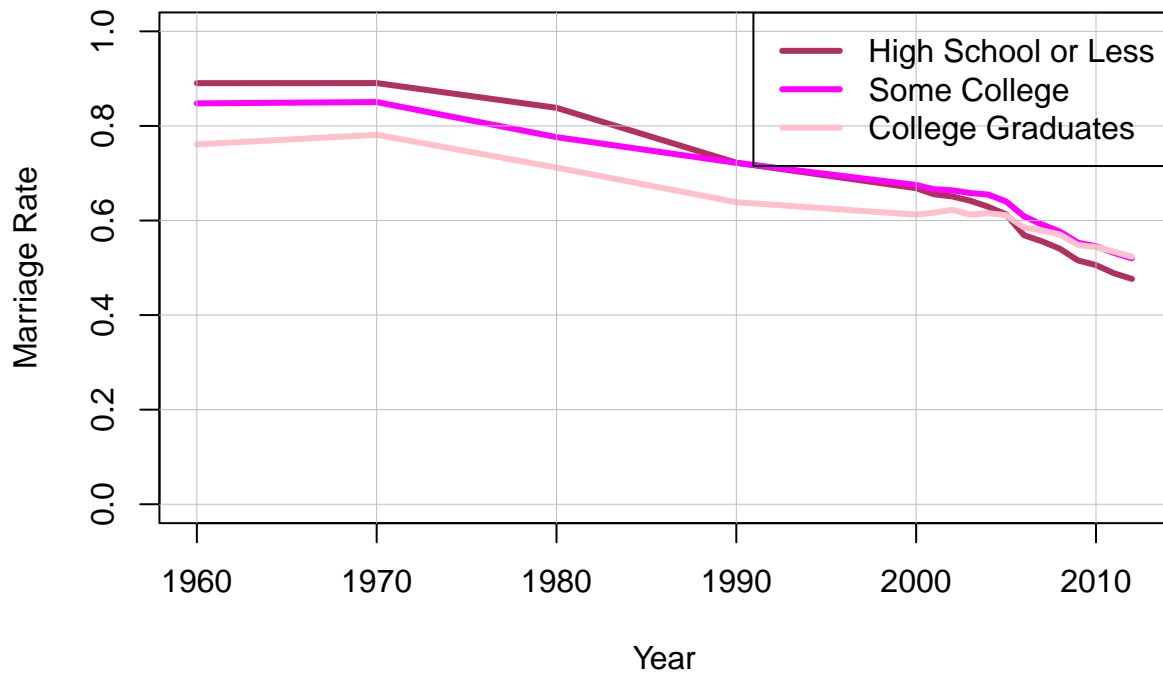
# convert to marriage rates
both_sexes_race_eth_filt_df$White_2534 <- 1- both_sexes_race_eth_filt_df$White_2534
both_sexes_race_eth_filt_df$Black_2534 <- 1- both_sexes_race_eth_filt_df$Black_2534
both_sexes_race_eth_filt_df$Hispanic_2534 <- 1- both_sexes_race_eth_filt_df$Hispanic_2534

# Plot Marriage Rates by Education Level
matplot(both_sexes_edu_filt_df$year, cbind(both_sexes_edu_filt_df$HS_2534,
                                           both_sexes_edu_filt_df$SC_2534,
                                           both_sexes_edu_filt_df$BAp_2534),

        type = "l", lty = 1,lwd=3,
        col = c("maroon", "magenta", "pink"), xlab = "Year",
        ylab = "Marriage Rate", main = "Marriage Rates by Education",
        ylim = c(0,1))
grid(nx = NULL, ny = NULL,
     lty = 1,
     col = "gray",
     lwd = 0.5)
legend("topright", legend = c("High School or Less", "Some College",
                              "College Graduates"),
      col = c("maroon", "magenta", "pink"),
      lty = 1, lwd=3)

```

Marriage Rates by Education



```
# Plot Marriage Rates by Race/Ethnicity
matplot(both_sexes_race_eth_filt_df$year, cbind(both_sexes_race_eth_filt_df$White_2534,
                                                both_sexes_race_eth_filt_df$Black_2534,
                                                both_sexes_race_eth_filt_df$Hispanic_2534),

        type = "l", lty = 1, lwd=3,
        col = c("forestgreen", "darkgreen", "lightgreen"), xlab = "Year",
        ylab = "Marriage Rate", main = "Marriage Rates by Race",
        ylim = c(0,1))
grid(nx = NULL, ny = NULL,
     lty = 1,
     col = "gray",
     lwd = 0.5)
legend("topright", legend = c("Non-hispanic White", "Black", "Hispanic"),
     col = c("forestgreen", "darkgreen", "lightgreen"),
     lty = 1, lwd=3)
```

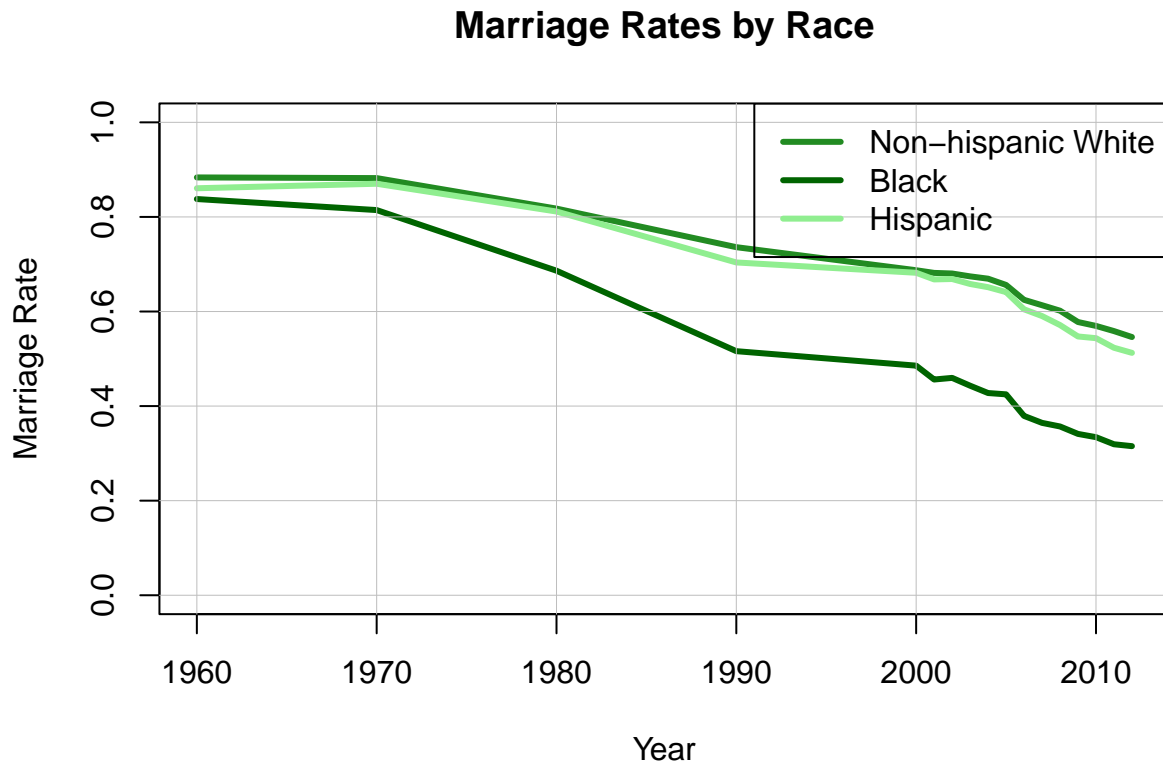


Figure 3: Marriage Rates for Women by Education - Ages 25-34

We use the `women_df` dataframe here.

The columns we use for women with children's education levels are `kids_HS_2534` - High school graduate or less (`EDUCD < 65`), age between 25 and 34, `kids_SC_2534` - Some college (`EDUCD >= 65 & <= 100`), age between 25 and 34, and `kids_BAp_2534` - Bachelor's degree or more (`EDUCD > 100`), age between 25 and 34 and the columns we use for women without children's education levels are `nokids_HS_2534` - High school graduate or less (`EDUCD < 65`), age between 25 and 34, `nokids_SC_2534` - Some college (`EDUCD >= 65 & <= 100`), age between 25 and 34, and `nokids_BAp_2534` - Bachelor's degree or more (`EDUCD > 100`), age between 25 and 34.

```
# Filter women.csv

# Women with Children
# remove unneeded columns
women_df_with_children <- women_df[, c("year", "kids_HS_2534", "kids_BAp_2534")]

# remove na's from kids_HS_2534 and kids_BAp_2534 columns
women_df_with_children <- women_df_with_children[
  !is.na(women_df_with_children$kids_HS_2534) |
  !is.na(women_df_with_children$kids_BAp_2534),]

# convert to marriage rates
women_df_with_children$kids_HS_2534 <- 1 - women_df_with_children$kids_HS_2534
women_df_with_children$kids_BAp_2534 <- 1 - women_df_with_children$kids_BAp_2534
```

```

# Women without Children
# remove unneeded columns
women_df_without_children <- women_df[, c("year", "nokids_HS_2534",
                                           "nokids_BAp_2534")]

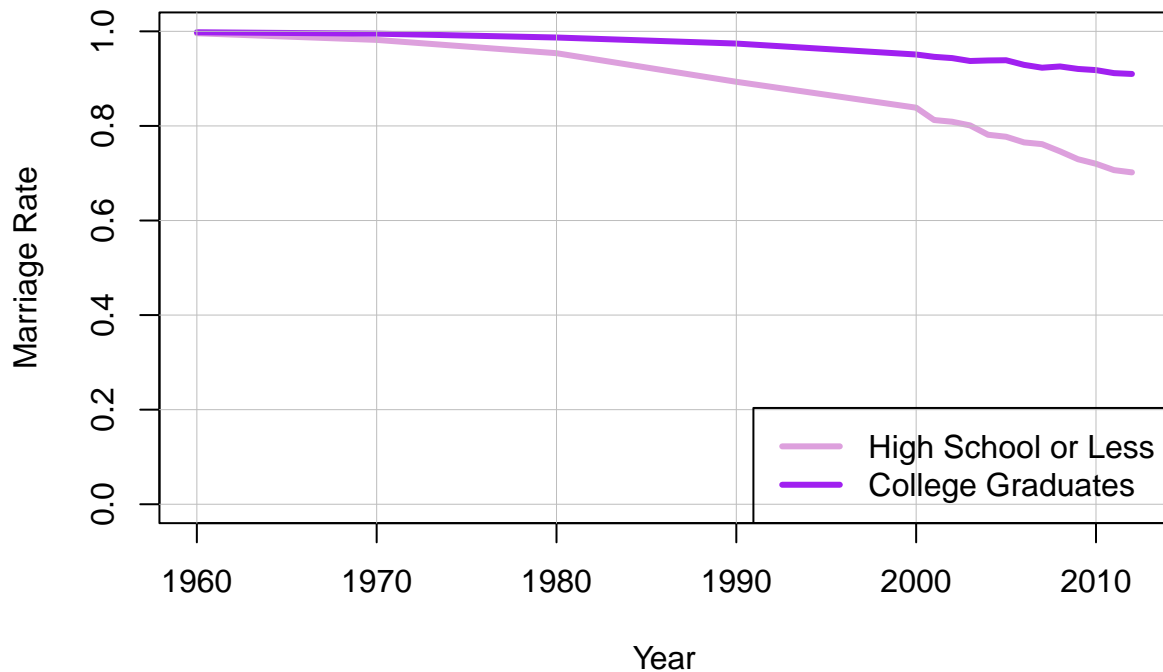
# remove na's from nokids_HS_2534 and nokids_BAp_2534 columns
women_df_without_children <- women_df_without_children[
  !is.na(women_df_without_children$nokids_HS_2534) |
  !is.na(women_df_without_children$nokids_BAp_2534),]

# convert to marriage rates
women_df_without_children$nokids_HS_2534 <- 1- women_df_without_children$nokids_HS_2534
women_df_without_children$nokids_BAp_2534 <- 1 - women_df_without_children$nokids_BAp_2534

# Plot marriage rates for women with children
matplot(women_df_with_children$year, cbind(women_df_with_children$kids_HS_2534,
                                           women_df_with_children$kids_BAp_2534),
        type = "l", lty = 1, lwd=3,
        col = c("plum", "purple"), xlab = "Year",
        ylab = "Marriage Rate", main = "Marriage Rates for Women - With Children",
        ylim = c(0,1))
grid(nx = NULL, ny = NULL,
     lty = 1,
     col = "gray",
     lwd = 0.5)
legend("bottomright", legend = c("High School or Less",
                                 "College Graduates"),
     col = c("plum", "purple"),
     lty = 1, lwd=3)

```

Marriage Rates for Women – With Children



```
# Plot marriage rates for women without children
matplot(women_df_without_children$year, cbind(women_df_without_children$nokids_HS_2534,
                                              women_df_without_children$nokids_BAp_2534),
        type = "l", lty = 1, lwd=3,
        col = c("plum", "purple"), xlab = "Year",
        ylab = "Marriage Rate", main = "Marriage Rates for Women - Without Children",
        ylim = c(0,1))
grid(nx = NULL, ny = NULL,
     lty = 1,
     col = "gray",
     lwd = 0.5)
legend("bottomright", legend = c("High School or Less",
                                "College Graduates"),
      col = c("plum", "purple"),
      lty = 1, lwd=3)
```

Marriage Rates for Women – Without Children

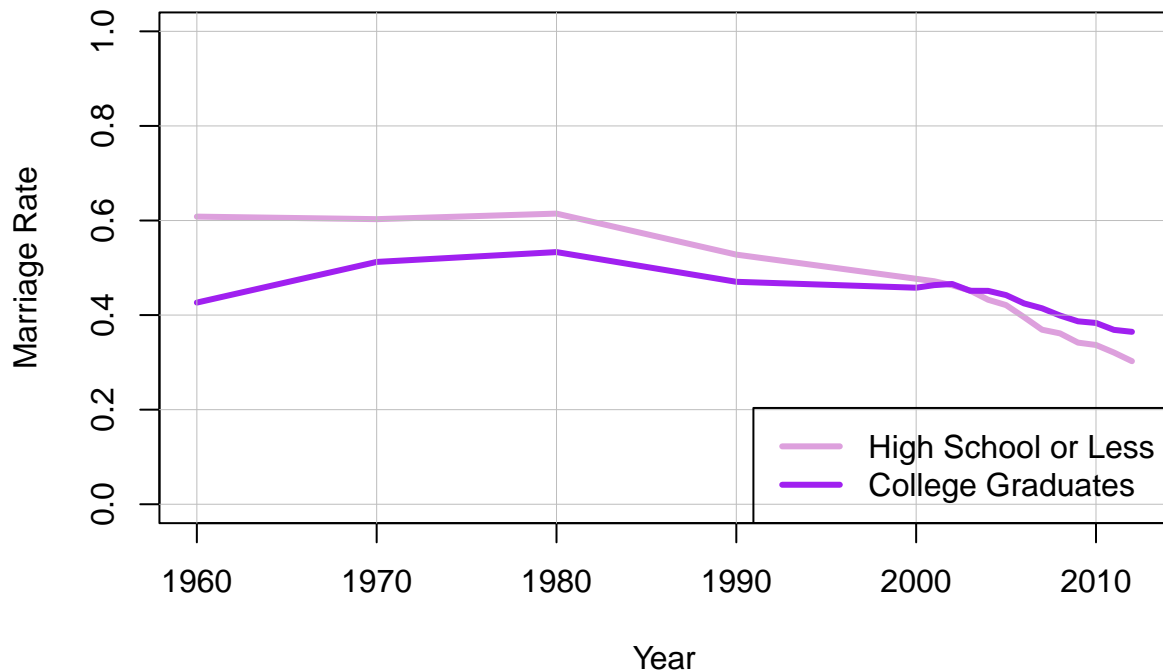


Figure 4:

Divorce Rates by Education - Ages 35-44

We use the `divorce_df` dataframe here.

The columns we use for women with children's education levels are `HS_3544` - High school graduate or less (`EDUCD < 65`), age between 35 and 44, `SC_3544` - Some college (`EDUCD >= 65 & <= 100`), age between 35 and 44, and `BAp_3544` - Bachelor's degree or more (`EDUCD > 100`), age between 35 and 44.

```
# Divorce rates
# remove unneeded columns
divorce_edu_df <- divorce_df[,c("year", "HS_3544", "SC_3544", "BAp_3544")]

# remove na's from HS_3544, SC_3544, BAp_3544 columns
divorce_edu_df <- divorce_edu_df[
  !is.na(divorce_edu_df$HS_3544) |
  !is.na(divorce_edu_df$SC_3544) |
  !is.na(divorce_edu_df$BAp_3544) ,]

# Divorce Plot
matplot(divorce_edu_df$year, cbind(divorce_edu_df$HS_3544,
  divorce_edu_df$SC_3544,
  divorce_edu_df$BAp_3544),
```

```

type = "l", lty = 1, lwd=3,
col = c("turquoise", "blue", "lightblue"), xlab = "Year",
ylab = "Divorce Rate", main = "Divorce Rates by Education - Ages 35-44",
ylim = c(0,0.25))
grid(nx = NULL, ny = NULL,
     lty = 1,
     col = "gray",
     lwd = 0.5)
legend("topleft", legend = c("High School or Less",
                             "Some College",
                             "College Graduates"),
      col = c("turquoise", "blue", "lightblue"),
      lty = 1, lwd=3)

```

