Keyla Pereira

- Lab 6

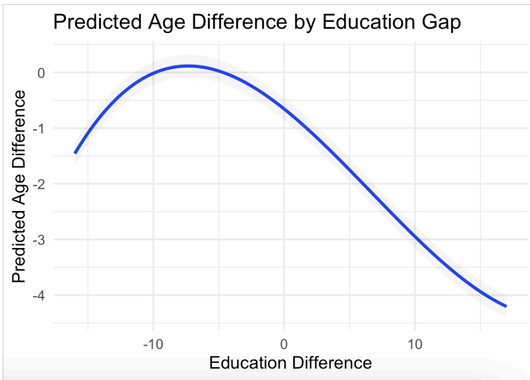
Nancy Reyes Soto, Irania Matos, Keyla Pereira, Kerra Sinanan, Heidy Collado.

- Subgroup: Adults aged 25–55 who are in the labor force and worked around 35 hours per week during the year.
- Most couples in the sample are similar in both age and education.
- On average, differences are small, but some couples show larger age or education gaps, which motivates the regression analysis.
- A linear model was first fit between age difference and education difference.
- Used model comparison of Wald test
- The relationship between education and age differences varies depending on the size of the education gap.
- Model predictions remain realistic and within a few years, showing that larger education gaps are less common but still meaningful in explaining age differences.

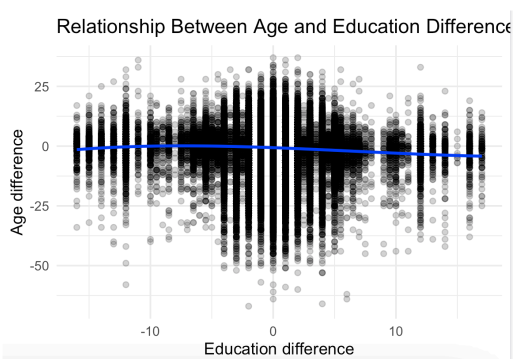
Using the ACS couples dataset, I compared each person's age and education to their partners using age_diff and educ_diff. Focusing on full-time working adults ages 25–55, I found most couples are similar in age and schooling but with some variation. Regressing age_diff on educ_diff gave a coefficient of about −0.15 ($p < 0.001$), showing that the more educated partner tends to be slightly younger. Adding squared and cubic terms significantly improved the fit (Wald $F \approx 91.5$, $p < 0.001$), meaning the relationship is nonlinear. The predicted age gaps are small and realistic, suggesting that within working couples, higher education is linked to being the younger partner.

	Model 1	**Model 2**
(Intercept)	-0.719***	-0.656***
	(0.012)	(0.012)
educ_diff	-0.148***	-0.188***
	(0.005)	(0.006)
I(educ_diff^2)		-0.008***
		(0.001)
I(educ_diff^3)		0.000***
		(0.000)
Num.Obs.	244874	244874

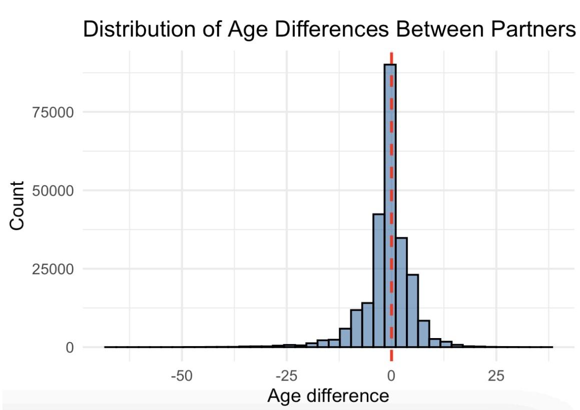
This curved line shows model predictions from the cubic regression. As education differences grow, the predicted age gap changes slightly, confirming a small but statistically significant relationship found in the regression test.



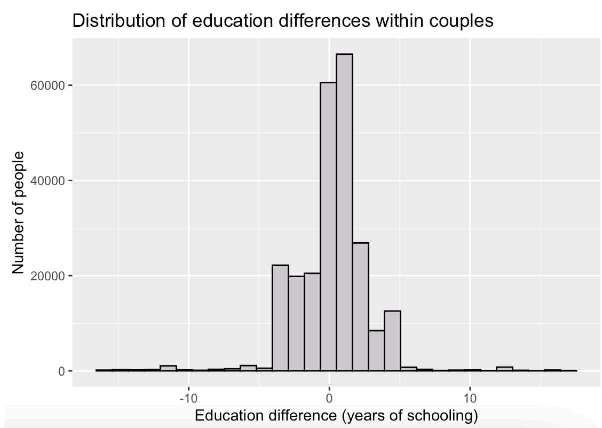
The scatterplot shows that age and education gaps are weakly related. The mostly flat blue line means that differences in schooling don't strongly predict age gaps, though a small curved shape hints at a mild nonlinear trend.



Most couples have very small age gaps centered around zero, meaning partners are usually close in age. The small left tail suggests that in some cases, the respondent is slightly younger than their partner.



Education levels between partners are also very similar, with most couples sharing nearly the same years of schooling. This pattern shows that people tend to partner with others who have similar educational backgrounds.



```
load("/Users/keylapereira/Downloads/ACS_2021_couples.RData")

head(acs2021_couples)
summary(acs2021_couples)
str(acs2021_couples)

library(ggplot2)
library(tidyverse)
library(haven)

setwd("..//ACS_2021_PUMS//") # your directory structure will be different
load("ACS_2021_couples.RData")


acs2021_couples$age_diff <- acs2021_couples$AGE - acs2021_couples$h_age


summary(acs2021_couples$age_diff[(acs2021_couples$SEX == "Female")&(acs2021_couples$h_sex == "Male")])
summary(acs2021_couples$age_diff[(acs2021_couples$SEX == "Male")&(acs2021_couples$h_sex == "Female")])
summary(acs2021_couples$age_diff[(acs2021_couples$SEX == "Male")&(acs2021_couples$h_sex == "Male")])
summary(acs2021_couples$age_diff[(acs2021_couples$SEX == "Female")&(acs2021_couples$h_sex == "Female")])

summary(acs2021_couples$AGE[(acs2021_couples$SEX == "Female")&(acs2021_couples$h_sex == "Male")])
summary(acs2021_couples$h_age[(acs2021_couples$SEX == "Female")&(acs2021_couples$h_sex == "Male")])

acs2021_couples$educ_numeric <- fct_recode(acs2021_couples$EDUC,
                                   "0" = "N/A or no schooling",
                                   "2" = "Nursery school to grade 4",
                                   "6.5" = "Grade 5, 6, 7, or 8",
                                   "9" = "Grade 9",
                                   "10" = "Grade 10",
                                   "11" = "Grade 11",
                                   "12" = "Grade 12",
                                   "13" = "1 year of college",
                                   "14" = "2 years of college",
                                   "15" = "3 years of college",
```

```
                                        "16" = "4 years of college",
                                        "17" = "5+ years of college")

acs2021_couples$educ_numeric <- as.numeric(levels(acs2021_couples$educ_numeric))[acs2021_couples$educ_numeric]

acs2021_couples$h_educ_numeric <- fct_recode(acs2021_couples$h_educ,
                                          "0" = "N/A or no schooling",
                                          "2" = "Nursery school to grade 4",
                                          "6.5" = "Grade 5, 6, 7, or 8",
                                          "9" = "Grade 9",
                                          "10" = "Grade 10",
                                          "11" = "Grade 11",
                                          "12" = "Grade 12",
                                          "13" = "1 year of college",
                                          "14" = "2 years of college",
                                          "15" = "3 years of college",
                                          "16" = "4 years of college",
                                          "17" = "5+ years of college")

acs2021_couples$h_educ_numeric <- as.numeric(levels(acs2021_couples$h_educ_numeric))[acs2021_couples$h_educ_numeric]

acs2021_couples$educ_diff <- acs2021_couples$educ_numeric - acs2021_couples$h_educ_numeric


acs_subgroup <- acs2021_couples %>% filter((AGE >= 25) & (AGE <= 55) &
                                          (LABFORCE == 2)
                                        & (WKSWORK2 > 4)
                                        & (UHRSWORK >= 35) )


library(AER)

m1 <- lm(age_diff ~ educ_diff,
         data = acs2021_couples)

m2 <- lm(age_diff ~ educ_diff + I(educ_diff^2) + I(educ_diff^3),
         data = acs2021_couples)

coeftest(m1, vcov = vcovHC)



waldtest(m1, m2, vcov = vcovHC)


library(modelsummary)

models <- list(
  "M 1" = lm(age_diff ~ educ_diff, data = acs2021_couples),
  "M 2" = lm(age_diff ~ educ_diff + I(educ_diff^2) + I(educ_diff^3), data = acs2021_couples) )

modelsummary(models, stars = TRUE)



----------------------------

acs2021_couples$age_diff <- acs2021_couples$AGE - acs2021_couples$h_age

acs2021_couples$educ_numeric <- fct_recode(acs2021_couples$EDUC,
                                        "0"="N/A or no schooling",
                                        "2"="Nursery school to grade 4",
                                        "6.5"="Grade 5, 6, 7, or 8",
                                        "9"="Grade 9",
                                        "10"="Grade 10",
                                        "11"="Grade 11",
                                        "12"="Grade 12",
                                        "13"="1 year of college",
                                        "14"="2 years of college",
```

```
                                        "15"="3 years of college",
                                        "16"="4 years of college",
                                        "17"="5+ years of college")

acs2021_couples$educ_numeric <- as.numeric(as.character(acs2021_couples$educ_numeric))

acs2021_couples$h_educ_numeric <- fct_recode(acs2021_couples$h_educ,
                                         "0"="N/A or no schooling",
                                         "2"="Nursery school to grade 4",
                                         "6.5"="Grade 5, 6, 7, or 8",
                                         "9"="Grade 9",
                                         "10"="Grade 10",
                                         "11"="Grade 11",
                                         "12"="Grade 12",
                                         "13"="1 year of college",
                                         "14"="2 years of college",
                                         "15"="3 years of college",
                                         "16"="4 years of college",
                                         "17"="5+ years of college")

acs2021_couples$h_educ_numeric <- as.numeric(as.character(acs2021_couples$h_educ_numeric))

acs2021_couples$educ_diff <- acs2021_couples$educ_numeric - acs2021_couples$h_educ_numeric

acs_subgroup <- acs2021_couples %>%
  filter(AGE >= 25, AGE <= 55, LABFORCE == 2, WKSWORK2 > 4, UHRSWORK >= 35)

m1 <- lm(age_diff ~ educ_diff, data = acs_subgroup)
m2 <- lm(age_diff ~ educ_diff + I(educ_diff^2) + I(educ_diff^3), data = acs_subgroup)

coeftest(m1, vcov = vcovHC)
waldtest(m1, m2, vcov = vcovHC)

models <- list("Model 1" = m1, "Model 2" = m2)
modelsummary(models, stars = TRUE)


---------------------


  educ_seq <- seq(min(acs_subgroup$educ_diff, na.rm = TRUE),
                max(acs_subgroup$educ_diff, na.rm = TRUE),
                length.out = 200)

pred_data <- data.frame(
  educ_diff = educ_seq,
  age_diff_hat = predict(m2, newdata = data.frame(educ_diff = educ_seq))
)

ggplot(pred_data, aes(x = educ_diff, y = age_diff_hat)) +
  geom_line(color = "blue", size = 1.2) +
  geom_ribbon(aes(
    ymin = age_diff_hat - 0.2,
    ymax = age_diff_hat + 0.2
  ), alpha = 0.15, fill = "gray") +
  labs(
    title = "Predicted Age Difference by Education Gap",
    x = "Education Difference",
    y = "Predicted Age Difference"
  ) +
  theme_minimal(base_size = 14)


------------

  ggplot(acs_subgroup, aes(x = educ_diff, y = age_diff)) +
  geom_point(alpha = 0.2, color = "black") +
  geom_smooth(
    method = "lm",
    formula = y ~ poly(x, 3, raw = TRUE),
```

```
  color = "blue",
  se = FALSE,
  linewidth = 1.2
 ) +
 labs(
  title = "Relationship Between Age and Education Differences (Cubic Fit)",
  x = "Education difference",
  y = "Age difference"
 ) +
 theme_minimal(base_size = 14)
```

```
 ggplot(acs_subgroup, aes(x = age_diff)) +
 geom_histogram(
  bins = 40,
  fill = "steelblue",
  color = "black",
  alpha = 0.7
 ) +
 geom_vline(
  xintercept = 0,
  color = "red",
  linetype = "dashed",
  linewidth = 1
 ) +
 labs(
  title = "Distribution of Age Differences Between Partners",
  x = "Age difference",
  y = "Count"
 ) +
 theme_minimal(base_size = 14)
```

```
 ggplot(acs_subgroup, aes(x = educ_diff)) +
 geom_histogram(bins = 30, color = "black", fill = "gray80") +
 labs(
  x = "Education difference (years of schooling)",
  y = "Number of people",
  title = "Distribution of education differences within couples"
 )
```

Topic: Does the price level of fashion brands significantly influence customer satisfaction ratings?

The first article I chose is “Consumer Perceptions of Price, Quality, and Value” by Valarie Zeithaml (1988). This study explores how consumers connect price, quality, and value when deciding what to buy. Zeithaml’s main question is whether price serves as a signal of product quality and how people form value judgments based on what they receive compared to what they give up. The paper reviews earlier research and develops a model showing that consumers often rely on price as an indicator of quality when other information, such as brand or product details, is limited.

The second article I chose is “Low Prices Are Just the Beginning: Price Image in Retail Management” by Ryan Hamilton and Alexander Chernev (2013). This study looks at how customers form a general “price image” of a retail store, meaning how expensive or inexpensive they believe it is overall. The authors use survey and experimental data to analyze how specific pricing cues like discounts, everyday low prices, and high priced items shape consumers’ overall price perceptions and satisfaction. They use econometric techniques such as regression and analysis of variance to measure how these factors affect the store’s image and customer attitudes. The main question addressed is how retailers can influence customer satisfaction and value perception not only through actual prices but also through the image of pricing they create.