Problem Set 1 - Part II

Alin Ierima

Immigration attitudes: the role of economic and cultural threat

[***Hint: Take your time to read the description of the assignment and the questions carefully.**]

Why do the majority of voters in the U.S. and other developed countries oppose increased immigration? According to the conventional wisdom and many economic theories, people simply do not want to face additional competition on the labor market (*economic threat* hypothesis). Nonetheless, most comprehensive empirical tests have failed to provide evidence in support of this hypothesis and it appears that people often support policies that are against their personal economic interest. At the same time, there has been growing evidence that immigration attitudes are rather influenced by various deep-rooted ethnic and cultural stereotypes (*cultural threat* hypothesis). Given the prominence of workers' economic concerns in the political discourse, how can these findings be reconciled?

This exercise is based in part on Malhotra, N., Margalit, Y. and Mo, C.H., 2013. "Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact." *American Journal of Political Science*, Vol. 38, No. 3, pp. 393-433.

The authors argue that, while job competition is not a prevalent threat and therefore may not be detected by aggregating survey responses, its *conditional* impact in selected industries (e.g the tech industry) may be quite sizable. To test their hypothesis, they conduct a unique survey of Americans' attitudes toward H-1B visas. The plurality of H-1B visas are occupied by Indian immigrants, who are skilled but ethnically distinct, which enables the authors to measure a specific skill set (high technology) that is threatened by a particular type of immigrant (H-1B visa holders). The data set `immig.csv` has the following variables:

Name	Description
`age`	Age (in years)
`female`	`1` indicates female; `0` indicates male
`employed`	`1` indicates employed; `0` indicates unemployed
`nontech.whitcol`	`1` indicates non-tech white-collar work (e.g., law)
`tech.whitcol`	`1` indicates high-technology work
`expl.prejud`	Explicit negative stereotypes about Indians (continuous scale, 0-1)
`impl.prejud`	Implicit bias against Indian Americans (continuous scale, 0-1)
`h1bvis.supp`	Support for increasing H-1B visas (5-point scale, 0-1)
`indimm.supp`	Support for increasing Indian immigration (5-point scale, 0-1)

- The main outcome of interest (`h1bvis.supp`) was measured as a following survey item:

"Some people have proposed that the U.S. government should increase the number of H-1B visas, which are allowances for U.S. companies to hire workers from foreign countries to work in highly skilled occupations (such as engineering, computer programming, and high-technology). Do you think the U.S. should increase, decrease, or keep about the same number of H-1B visas?"

- Another outcome (`indimm.supp`) similarly asked about the *"the number of immigrants from India." Both variables have the following response options: `0` = "decrease a great deal", `0.25` = "decrease a little", `0.5` = "keep about the same", `0.75` = "increase a little", `1` = "increase a great deal".*

- To measure explicit stereotypes (`expl.prejud`), respondents were asked to evaluate Indians on a series of traits: capable, polite, hardworking, hygienic, and trustworthy. All responses were then used to create a scale lying between `0` (only positive traits of Indians) to `1` (no positive traits of Indians).

- Implicit bias (`impl.prejud`) is measured via the *Implicit Association Test* (IAT) which is an experimental method designed to gauge the strength of associations linking social categories (e.g., European vs Indian American) to evaluative anchors (e.g., good vs bad). Individual who are prejudiced against Indians should be quicker at making classifications of faces and words when *European American* (*Indian American*) is paired with *good* (*bad*) than when *European American* (*Indian American*) is paired with *bad* (*good*). If you want, you can test yourself here. [*Comment: It is actually quite fun to do the test, try it out if you have a few minutes.]

Question 1 (25 points)

Start by examining the distribution of immigration attitudes. Shortly discuss your findings in the text. *(5 points)*

- What is the proportion of people who are willing to increase the quota for high-skilled foreign professionals (`h1bvis.supp`)?
- What is the proportion of people who are willing to support immigration from India (`indimm.supp`)?

Now compare the distribution of two distinct measures of cultural threat: explicit stereotyping about Indians (`expl.prejud`) and implicit bias against Indian Americans (`impl.prejud`).

- In particular, create a scatterplot, where `impl.prejud` is on the y axis and `expl.prejud` on the x axis. Use solid circles, and color these in *steelblue2* with transparency level 0.5. Since the variable `expl.prejud` is not exactly continuous, to make the points more dispersed for presentation reasons you can add a small amount of noise to this variable with the function `jitter()`. Try out `jitter(immig$expl.prejud, 3)`, to see how the variable `expl.prejud` changes. In the below scatterplot we passed `jitter(immig$expl.prejud, 3)` to the `plot()` function. *(5 points)*
- Run a linear regression between the two prejudice variables, where `impl.prejud` is the outcome variable and `expl.prejud` the predictor. Save the linear regression model in an object called `fit1`. Add a linear regression line from this model to the scatterplot. Color it in *orangered*, set line width equal to 3. *(5 points)*
- Calculate the correlation coefficient between the two prejudice variables and then add this correlation coefficient as text to the top right corner of the scatterplot. *(4 points)*
- Taking into account the scatterplot, correlation coefficients and the linear regression model, what can you say about the relationship between the implicit and explicit prejudice variables? Do you find any evidence that these two variables represent the same concept or they rather capture different constructs of prejudice? *(6 points)*

Answer 1

```
immig <- read.csv("data/immig.csv")
```

Proportion of people who are willing to increase the quota for high-skilled foreign professionals (`h1bvis.supp`)

```
table(immig$h1bvis.supp) / length(immig$h1bvis.supp)
```

```
##
##         0      0.25       0.5      0.75         1
## 0.3045013 0.2250662 0.2956752 0.1015004 0.0635481
```

Proportion of people who are willing to support immigration from India (`indimm.supp`)

```
table(immig$indimm.supp) / length(immig$h1bvis.supp)
```

```
##
##          0       0.25        0.5       0.75          1
## 0.28508385 0.17917034 0.39452780 0.09885260 0.03265666
```

Insert your discussion here: There is an overall anti-migration perception for both high-skilled professionals and Indians, however a good part of the respondents also hold a neutral position (0.5).

Regression results from model `fit1` (for your information)

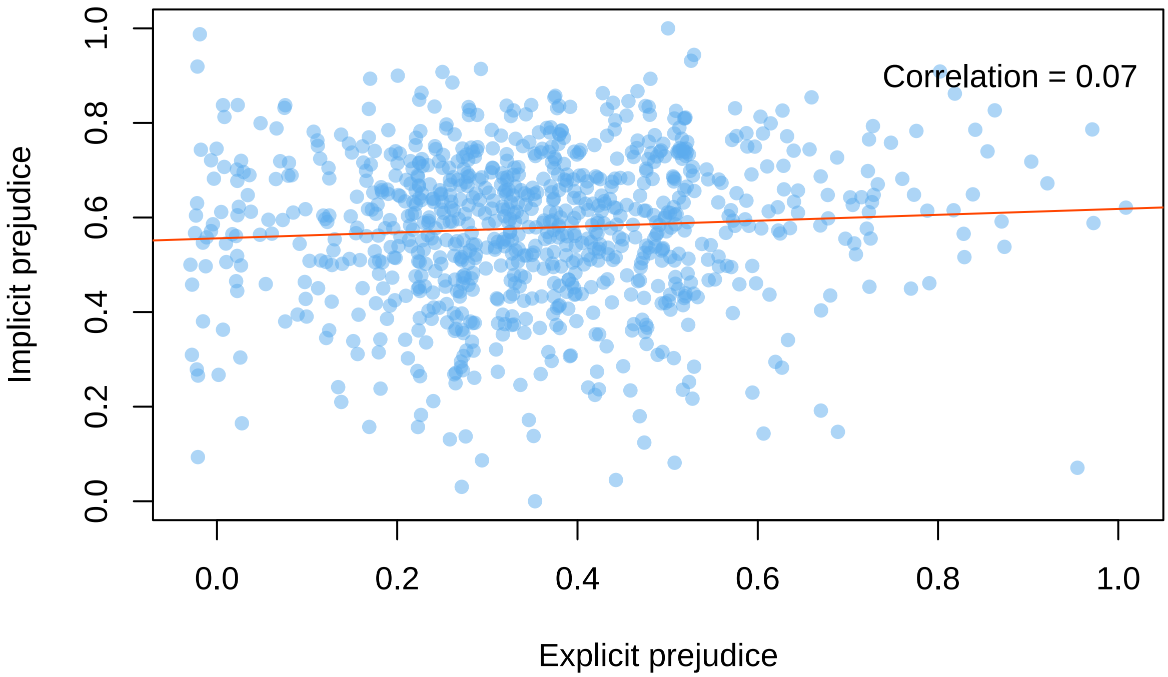
```
(fit1 = lm(immig$impl.prejud ~ immig$expl.prejud))
```

```
##
## Call:
## lm(formula = immig$impl.prejud ~ immig$expl.prejud)
##
## Coefficients:
##       (Intercept)  immig$expl.prejud
##           0.55588            0.06227
```

Plot

```
#Omitting NAs for cor()
immig_no_na = na.omit(immig)
correlation = cor(immig_no_na$expl.prejud, immig_no_na$impl.prejud)


plot(jitter(immig$expl.prejud, 3), immig$impl.prejud, pch = 16,
     col = adjustcolor("steelblue2", alpha.f = 0.5),
     xlab = "Explicit prejudice",
     ylab = "Implicit prejudice")
abline(fit1, col = "orangered")
text(x = 0.88, y = 0.9,
     labels = paste("Correlation =", round(correlation, 2)))
```



Insert your discussion here: The correlation sugests the two types of prejudice go very slightly together, indicating that they represent different concepts. Explicit prejudice does not necessarily correspond to high levels of implicit prejudice to a great deal, however, and this in turn proves that the two are largely disconnected. This is further supported by the scatterplot which does not indicate a significant correlation between the two.

Question 2 (20 points)

[This part of the question is for 5 points]

- Compute the correlations between all four immigration attitude and cultural threat measures. Rely on the same observations for all 3 pairs. Provide a short discussion of your findings. In particular, given the correlation results, do you agree that cultural threat (captured by the two prejudice variables) is an important predictor of immigration attitudes (captured by the H-1B visa support and Indian immigration support variables) as claimed in the literature?

[This part of the question is for 10 points]

- If the labor market hypothesis is correct, opposition to H-1B visas should also be more pronounced among those who are economically threatened by this policy such as individuals in the high-technology sector. At the same time, tech workers should not be more opposed (than non tech workers and than unemployed) to general Indian immigration because of any *economic* considerations.
  - To test these expectations, run two linear regression models, where you regress H-1B and Indian immigration attitudes on the indicator variable for tech workers (`tech.whitcol`). Save these models in objects called `fit2` and `fit3`.

- Shortly discuss the findings from model `fit2` and model `fit3`. Do the results support the labor market hypothesis?
- Is the relationship different from the one involving cultural threat (consider the correlation matrix for the relationship between cultural threat and immigration attitudes). If so, how?

[This part of the question is for 5 points]

- Present the results from the regression models (`fit2` and `fit3`) in a nicely formatted regression table. To achieve this pass these models to the function `stargazer::stargazer()`. NOTE: For the **R chunk** that calls `stargazer`, please use the option results = 'asis' in the R chunk specifications or else the formatting will be off. When calling `stargazer::stargazer()`, you can use the following format:

`stargazer()` will output a nicely formatted regression table when you knit the file to a pdf (not html). If you would like to have a nicely formatted regression table in html, set the argument *type = "html"*. For this assignment ignore the p-values in your discussion of the results.

Answer 2

Correlation matrix

```
na_omitting_data = na.omit(immig[ ,c("expl.prejud", "impl.prejud", "h1bvis.supp", "indimm.supp")])
cor(na_omitting_data)
```

```
##             expl.prejud impl.prejud h1bvis.supp indimm.supp
## expl.prejud  1.00000000  0.06864096  -0.1570602  -0.3214425
## impl.prejud  0.06864096  1.00000000  -0.1133121  -0.1296632
## h1bvis.supp -0.15706018 -0.11331209   1.0000000   0.6107116
## indimm.supp -0.32144247 -0.12966322   0.6107116   1.0000000
```

Insert your discussion here: As already previously seen, explicit and implicit prejudice go together to a very small extent (0.07). Additionally, explicit prejudice has a negative correlation of -0.16 for visa support and -0.32 for Indian immigration support, whereas implicit prejudice has a negative correlation of -0.11 for visa and -0.13 for Indian immigration. This indicates that these correlations are still too weak to claim that cultural threat predicts immigration attitudes, with one exception (the correlation between explicit prejudice and Indian immigration support is -0.32, a moderate correlation and the strongest of the four).

Regression Results

```
(fit2 = lm(immig$h1bvis.supp  ~ immig$tech.whitcol))
```

Support for H-1B visas (model `fit2`)

```
##
## Call:
## lm(formula = immig$h1bvis.supp ~ immig$tech.whitcol)
##
## Coefficients:
##        (Intercept)  immig$tech.whitcol
##            0.34995             -0.05334
```

```r
(fit3 = lm(immig$indimm.supp ~ immig$tech.whitcol))
```

Support for Indian immigration (model `fit3`)

```
##
## Call:
## lm(formula = immig$indimm.supp ~ immig$tech.whitcol)
##
## Coefficients:
##        (Intercept)  immig$tech.whitcol
##           0.352540           -0.005082
```

Insert your discussion here: It is true that the opposition to the visa is more pronounced for the high-tech workers than the opposition to the Indian immigrants, however the effect is still not much. If the base support is 0.35 for non-white collar tech people, then the difference between them and white collar tech people is only around 0.05. Still, this slope is three times bigger than the one existing for the Indian immigrants. Therefore, there is some support for the labour market hypothesis. On the other hand, the correlations of cultural threat seem more convincing in terms of magnitude than this hypothesis.

Nice regression table

```r
stargazer::stargazer(fit2, fit3,
                     title = "Regression Models for Immigration Attitudes",
                     intercept.top = T, intercept.bottom = F,
                     covariate.labels = c("High-Skilled Worker", "High-Skilled Worker"),
                     dep.var.labels = c("Support for H-1B Visas", " Support for Indian Immigration"),
                     header = FALSE,
                     type = "latex")
```

Table 2: Regression Models for Immigration Attitudes

	Dependent variable:	
	Support for H-1B Visas	Support for Indian Immigration
	(1)	(2)
High-Skilled Worker	0.350***	0.353***
	(0.009)	(0.008)
High-Skilled Worker	−0.053	−0.005
	(0.040)	(0.037)
Observations	1,122	1,122
R^2	0.002	0.00002
Adjusted R^2	0.001	−0.001
Residual Std. Error (df = 1120)	0.300	0.276
F Statistic (df = 1; 1120)	1.764	0.019
Note:		*p<0.1; **p<0.05; ***p<0.01

Question 3 (30 points)

[This part of the question is for 15 points]

When examining hypotheses, it is always important to have an appropriate comparison group. One may argue that comparing tech workers to everybody else as we did in Question 2 may be problematic due to a variety of confounding variables (such as skill level and employment status). To overcome this issue, run a regression where support for H-1B visas is a function of the categorical variable `group` which captures different job categories and work status. To do so, follow the below steps:

- Create a factor variable `group` .The `group` variable should take a value of *Tech* if someone is employed in tech, *Whitecollar* if someone is employed in other "white-collar" jobs (non tech high-skilled jobs, e.g. in law or finance), *Other* if someone is employed in any other sector, and *Unemployed* if someone is unemployed. *(5 points)*
  - *Hint: If you run a regression model using the `group` variable as predictor, R will automatically choose the reference (baseline) category. You can check all categories in a given factor variable using the function `levels()`. For example `levels(immig$group)` will give you all categories in this variable. R takes the first category in this list as the reference category. You can set the reference category of factor variable with the function `relevel()`. For example, if you would like to have *Other* as a baseline group (reference group), you can do the following:

```
immig$group <- relevel(immig$group, ref = "Other")
```

- Then, compare the support for H-1B across these `group` categories using a linear regression. Save the results from the linear regression in an object called `fit4`. *(4 points)*
- Shortly discuss the results from the regression model. Is this comparison of immigration attitudes between the categories in the `group` variable more or less supportive of the labor market hypothesis than your findings from Question 2? *(3 points)*
- Indicate in the text what is the average support lever for each category in the `group` variable. Do not copy-paste the coefficients in the text. Use instead 'r'. Remember that you can extract the coefficients from the model using `coef(model_name)["Coefficient Name"]`. *(3 points)*
  - For example, if you would like to extract and round the intercept coefficient, you can do the following:
  - If you would like to use this information within the text, write 'r round(coef(fit3)["(Intercept)"],2)' within the text.

```
coef(fit4)["(Intercept)"]
## rounded
round(coef(fit4)["(Intercept)"],2)
```

[This part of the question is for 8 points]

Now, one may also argue that those who work in the tech sector are disproportionately young and male which may confound our results.

- To account for this possibility, fit another linear regression (model `fit5`), where in addition to the `group` variable you also include `age` and `female` as pre-treatment covariates. (4 points)
- Shortly discuss your findings. Does it change the results and, if so, how? (4 points)

[This part of the question is for 7 points]

Finally, fit a linear regression model with all threat indicators (`group`, `expl.prejud`, `impl.prejud`) and calculate its R^2. Save the model output in an object called `fit_full`. *(4 points)*

- Shortly discuss your findings. *(3 points)*
  - How much of the variation is explained?
  - Based on the model fit, what can you conclude about the role of threat factors (cultural and economic threat) in predicting support for H-1B visas?

Answer 3

```
process_group = function(x) {
  ifelse(x$employed == 0, "Unemployed",
         ifelse(x$tech.whitcol == 1, "Tech",
                ifelse(x$nontech.whitcol == 1 & x$tech.whitcol == 0, "Whitecollar", "Other")))}

immig$group =  immig  |>
  process_group() |>
  as.factor() |>
  relevel(ref = "Tech")
```

Regression results - impact of group on support for H-1B visas (model `fit4`)

```
(fit4 = lm(immig$h1bvis.supp ~ immig$group))
```

```
##
## Call:
## lm(formula = immig$h1bvis.supp ~ immig$group)
##
## Coefficients:
##           (Intercept)       immig$groupOther   immig$groupUnemployed
##               0.29661                0.04946                 0.05217
## immig$groupWhitecollar
##               0.09563
```

Our discussion here: The results show us that all the other groups are more open to H-1B workers than the tech group, strengthening the hypothesis of the labour market hypothesis. The value of tech workers is 0.3 whereas the predicted support is 0.4 for white collars, 0.35 for unemployed, and 0.35 for others. The results are more supportive than the ones from Q2 due to the fact that, through directly examining the job categories with their visa support, we can notice that the tech people are on average predicted to have the lowest score.

Regression results - impact of group on support for H-1B visas controlling for female and age) (model `fit5`)

```
(fit5 = lm(immig$h1bvis.supp ~ immig$group + immig$female + immig$age))
```

```
##
## Call:
## lm(formula = immig$h1bvis.supp ~ immig$group + immig$female +
##     immig$age)
##
## Coefficients:
##           (Intercept)        immig$groupOther   immig$groupUnemployed
##               0.43338                 0.07598                 0.08984
## immig$groupWhitecollar            immig$female               immig$age
##               0.13127                -0.07536                -0.00248
```

Insert your discussion here: The model shows that, in fact, being a female is predicted to influence the openness to the visa negatively (-0.08), and also age has a negative slight effect as well (-0.002). While including gender and age as covariates enhance the explanatory power of the model, tech workers still remain the least open to the visa. Since the intercept for fit5 (including only males and controlling for age) is higher than the intercept for fit4 (both genders and no age control), it implies that including age and gender as covariates may explain some of the variation, but the effect of job category remains significant even when considering only male workers.

Regression results (`fit_full` model)

```
(fit_full = lm(immig$h1bvis.supp ~ immig$group + immig$expl.prejud + immig$impl.prejud))
```

```
##
## Call:
## lm(formula = immig$h1bvis.supp ~ immig$group + immig$expl.prejud +
##     immig$impl.prejud)
##
## Coefficients:
##           (Intercept)        immig$groupOther   immig$groupUnemployed
##               0.50239                 0.04719                 0.04099
## immig$groupWhitecollar       immig$expl.prejud       immig$impl.prejud
##               0.10270                -0.25268                -0.19169
```

```
summary(fit_full)$r.squared
```

```
## [1] 0.03833731
```

Insert your discussion here: Both coefficients for explicit and implicit prejudice are negative, indicating that higher level of prejudice are associated with lower support for the visas. However, the R^2 is very low, suggesting that only around 4% of the variation is explained by these predictors.

Question 4 (40 points)

Besides economic and cultural threat, many scholars also argue that gender is an important predictor of immigration attitudes. While there is some evidence that women are slightly less opposed to immigration than men, it may also be true that gender conditions the very effect of other factors such as cultural threat.

- To see if it is indeed the case, fit a linear regression of H-1B support (`h1bvis.supp`) on gender (`female`), implicit prejudice (`impl.prejud`), and the interaction between gender and implicit prejudice. Call the model `fit_h1b_gender`. *(5 points)*
- Then, create a plot with the **predicted** level of H-1B support (y-axis) across the full range (from its minimum to maximum value) of implicit prejudice (x-axis) by gender (for male and female). *(10 points)* [Hint: You can use the `predict()` function we covered in class.]
- Considering the results, would you agree that gender alters the relationship between cultural threat and immigration attitudes? Discuss why. *(5 points)*

Age is another important covariate.

- Fit two regression models. In the first model H-1B support (`h1bvis.supp`) is a linear function of age. In the second model H-1B support (`h1bvis.supp`) is quadratic function of age. Save the output of these models in the objects `fit_age` and `fit_age_sq` respectively. *(5 points)*
- Present the results from **both** models (`fit_age`, `fit_age_sq`) by plotting the predicted levels of support (y-axis) across the whole age range (x-axis). ***(10 bonus points)***
- Shortly discuss the results when immigration support is a linear function of age and when immigration is a quadratic function of age. Would you say that people become more opposed to immigration with age? Which model do you trust more - which model fits the data better? ***(5 bonus points)***

Answer 4

Regression results (`fit_h1b_gender` model)

```
(fit_h1b_gender = lm(h1bvis.supp ~ impl.prejud + female + impl.prejud * female, data = immig))
```

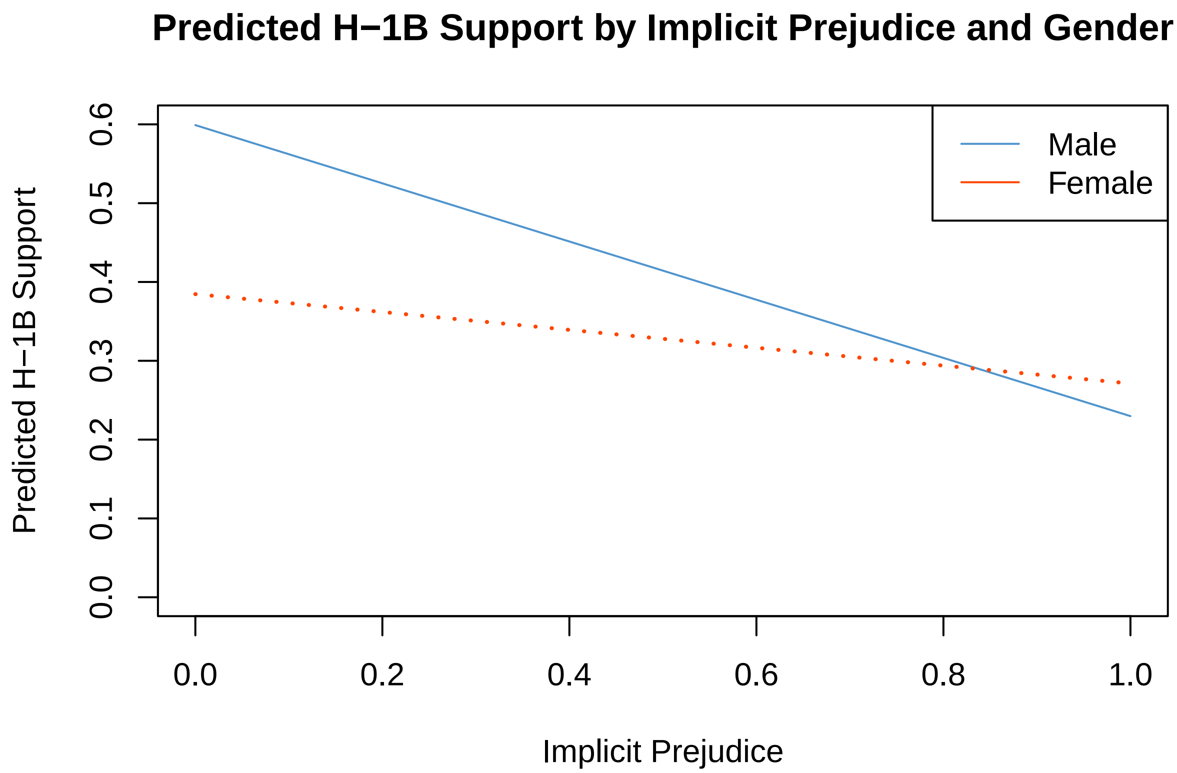
```
##
## Call:
## lm(formula = h1bvis.supp ~ impl.prejud + female + impl.prejud *
##     female, data = immig)
##
## Coefficients:
##        (Intercept)         impl.prejud              female  impl.prejud:female
##             0.5990             -0.3692             -0.2144              0.2558
```

Plot with predicted H-1B support for different values of `impl.prejud` and gender (`female`)

```
implicit_prejudice_range = seq(min(immig$impl.prejud, na.rm = TRUE), max(immig$impl.prejud, na.rm = TRU

predictions_gender =
  data.frame(
    impl.prejud = implicit_prejudice_range,
     male = predict(fit_h1b_gender, newdata = data.frame(female = 0, impl.prejud = implicit_prejudice_ra
    female = predict(fit_h1b_gender, newdata = data.frame(female = 1, impl.prejud = implicit_prejudice_
)


plot(predictions_gender$impl.prejud, predictions_gender$male, type = "l", xlab = "Implicit Prejudice", y
lines(predictions_gender$impl.prejud, predictions_gender$female, col = "orangered", lty = 3, lwd = 2)
legend("topright", legend = c("Male", "Female"), col = c("steelblue3", "orangered"), lty = 1)
```



Insert your discussion here: This graph shows that gender does alter the relationship between implicit prejudice and immigration attitudes, as the effect of cultural threat on immigration is of a greater steepness for males, the more implicitly prejudiced they are. The trend is negative for females as well; however, increasing the implicit prejudice does not lead to a steep slope; rather, the support for visas remains around the same. Therefore, I would agree to the hypothesis that gender influences the relationship between cultural threat and immigration atitudes.

Regression results (model `fit_age`)

```
(fit_age = lm(h1bvis.supp ~ age, data = immig))
```

```
##
## Call:
## lm(formula = h1bvis.supp ~ age, data = immig)
##
## Coefficients:
## (Intercept)          age
##    0.451677    -0.002169
```

Regression results (model `fit_age_sq`)

```r
(fit_age_sq = lm(h1bvis.supp ~ age + I(age^2), data =immig))
```

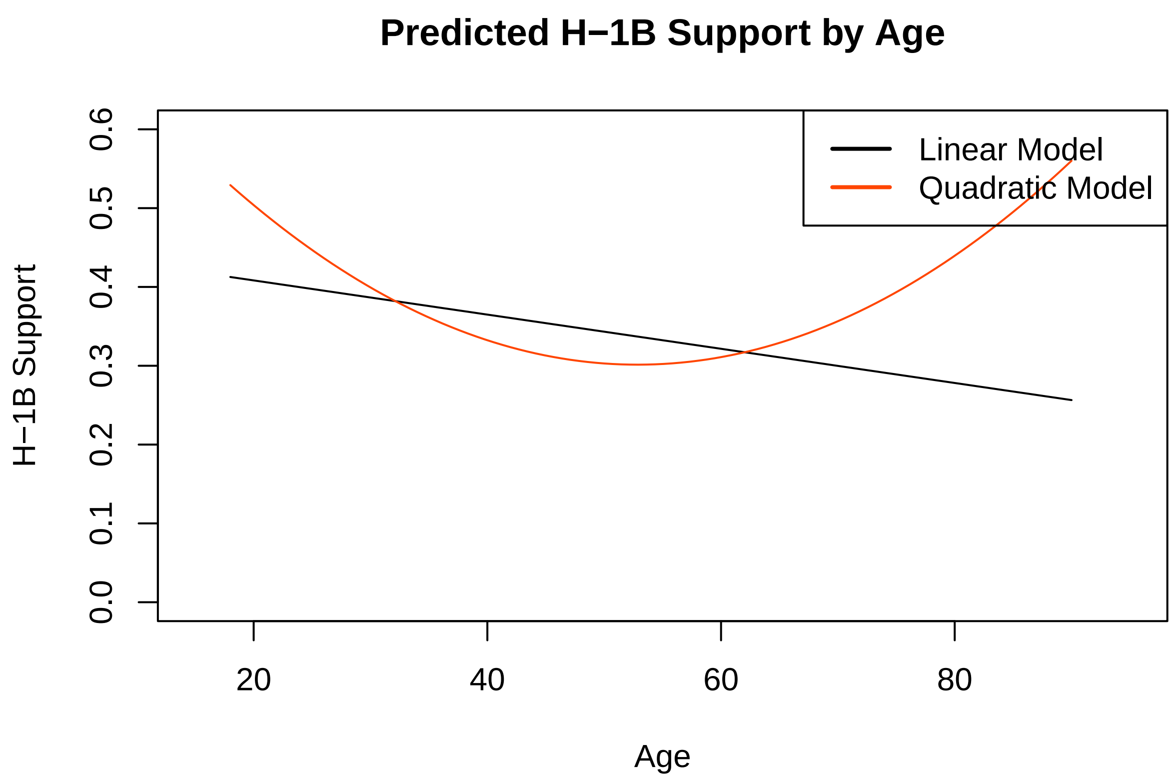
```
##
## Call:
## lm(formula = h1bvis.supp ~ age + I(age^2), data = immig)
##
## Coefficients:
## (Intercept)          age     I(age^2)
##   0.8253377   -0.0198254    0.0001875
```

Plot with predicted H-1B support for different values of age (linear and quadratic function)

```r
age_range = seq(min(immig$age, na.rm = TRUE), max(immig$age, na.rm = TRUE), length.out = 100)

predicted_fit_age = predict(fit_age, newdata = data.frame(age = age_range))
predicted_fit_age_sq = predict(fit_age_sq, newdata = data.frame(age = age_range, `age^2` = age_range^2))

# Plot predicted levels of support for both linear and quadratic models
plot(age_range, predicted_fit_age, type = "l", xlab = "Age", ylab = "H-1B Support", col = "black", main
lines(age_range, predicted_fit_age_sq, col = "orangered")   # Quadratic model
legend("topright", legend = c("Linear Model", "Quadratic Model"), col = c("black", "orangered"), lty =
```



Insert your discussion here: The normal linear model works under the assumption of linearity. If we judge based on this model, we would believe that the predicted levels of support for immigration decrease steadily with increasing age. However, the quadratic model is able to observe non-linearity as well, and it shows us that the levels of support initially decrease with age but then start to increase after a certain point. In this case, the latter might be more appropriate.

Evaluation

- 4 questions for a total of 100 points plus 15 bonus points (see question 4).