Instructions for the Project

* Analysis related to the paper
  + Children and their Parents Labour Supply (Angrist and Evans, 1998)
* ML application in the paper
  + GENERALIZED RANDOM FORESTS (Susan Athey, Julie Tibshirani and Stefan Wager, 2018)
* Data
  + SAS dataset available at <https://economics.mit.edu/faculty/angrist/data1/data>
* Variables for analysis using causal forests (Y , X , W)

List of Yi

1. Yi = “If worked for pay” for all women sample
2. Yi = “If worked for pay” for married women sample
3. Yi = “If worked for pay” for husbands of married women sample
4. Yi = “weeks worked” for all women sample
5. Yi = “weeks worked” for married women sample
6. Yi = “weeks worked” for husbands of married women sample
7. Yi = “hours worked” for all women sample
8. Yi = “hours worked” for married women sample
9. Yi = “hours worked” for husbands of married women sample
10. Yi = “annual labour income” for all women sample
11. Yi = “annual labour income” for married women sample
12. Yi = “annual labour income” for husbands of married women sample
13. Yi = “log of family income” for all women sample
14. Yi = “log of family income” for married women sample
15. Yi = “log of non-wife income” for married women sample

List of Xik  (Covariates: common to all Yi specifications noted above.)

1. Xi1 = “age” of unit “i” i.e. age of woman (if considering female labour supply) or age of man (if considering husband’s labour supply)
2. Xi2 = “age at first birth” of unit “i” i.e. age of woman at birth of first child (if considering female labour supply) or age of man at birth of first child (if considering husband’s labour supply)
3. Xi3 = “years of education” of unit “i” i.e. age of woman at birth of first child (if considering female labour supply) or age of man at birth of first child (if considering husband’s labour supply)

Race and ethnic background indicators: Please recode this information as needed for the decision tree

1. Xi4 = “White” is race and ethnic background of unit “i” i.e. race and ethnic background of woman (if considering female labour supply) or race and ethnic background of man (if considering husband’s labour supply). This is an indicator variable that takes the value 1 if race =white and 0 otherwise.
2. Xi5 = “Black” is race and ethnic background of unit “i” i.e. race and ethnic background of woman (if considering female labour supply) or race and ethnic background of man (if considering husband’s labour supply). This is an indicator variable that takes the value 1 if race = black and 0 otherwise.
3. Xi6 = “Hispanic” is race and ethnic background of unit “i” i.e. race and ethnic background of woman (if considering female labour supply) or race and ethnic background of man (if considering husband’s labour supply). This is an indicator variable that takes the value 1 if race =hispanic and 0 otherwise.
4. Xi7 = “Other Race” is race and ethnic background of unit “i” i.e. race and ethnic background of woman (if considering female labour supply) or race and ethnic background of man (if considering husband’s labour supply). This is an indicator variable that takes the value 1 if race =other race and 0 otherwise.

List of Wi  (treatment: common to all Yi specifications noted above i.e. looking at impact of fertility decision on different measures of labour supply decision (by different set of economics agents))

1. Wi  = “more than two children”

The Angrist and Evans paper has another Wi  = “children ever born” but we will not use it for this project.

* Analysis: **Restrict analysis to 1980 data**
  + Calculate means and standard deviations for Yi , Xi , Wi
    - Check if you have the right sample by comparing your descriptive statistics against those in Table 2 and 6 of Angrist and Evans
  + OLS estimation of Yi on X’s and W
    - Results should line up with columns 1, 4 and 7 of Table 7.
    - OLS does not take into account endogenity of Wi (non-random treatment assignment of treatment i.e. fertility decision). So results cannot be interpreted as causal effects.
  + Run causal forest for each Yi
    - Causal tree helps us analyze heterogeneity in a parameter (the effect of a variable)
    - Causal trees do not take into account endogeneity of Wi (non-random treatment assignment of treatment i.e. fertility decision). So results cannot be interpreted as showing causal effects but the exercise will help you learn how to implement causal trees.
  + Optional: IV estimation of Yi on X’s , W, Z
    - Zi = “same sex” is an indicator variable that takes on value 1 if the first two children have the same gender
    - Results should line up with columns 2, 5 and 8 of Table 7.
    - <https://pypi.org/project/linearmodels/>
  + IV forest
    - To be discussed in final class.