1. Introduction
   1. Abstract
   2. Background
2. Hypothesis and referencing Deming’s paper on headstart
   1. What are the **short and long term effects** of **head start program on youth** based on a **child's ethnicity**.
3. Methodology
   1. Data processing
      1. Defining what headstart is represented in the dataset
         1. Most likely determining whether or not an individual participated in any head start program, rather than a set of years. Need to generate a new variable.
      2. Defining what is “short and long term effects” based on dataset
         1. Education Enrollment
            1. High school
            2. College
         2. Family income
            1. We can do a before/after. Taking an individual, is their income higher than their peers holding all other factors fixed?
      3. Defining the “ethnicities” that we are classifying
         1. Hispanic
         2. black
         3. white
      4. Head Start
         1. If ever participated in head start, would mean yes
      5. Other preschool
      6. No preschool
      7. Tracking individual people through the years
         1. Perhaps have a subset of people who graduate high school early
         2. Subset of people who don’t graduate high school
      8. Variable List
         1. head\_start: 0 or 1
         2. sibdiff : 0 or 1
         3. mom\_id: unique from the dataset
         4. Hispanic: 0 or 1
         5. Black: 0 or 1
         6. Male: 0 or 1
         7. Firstborn: 0 to 1
         8. lninc\_0to3: log income, 0 to 3,
         9. momed: years of mom edu
         10. dadhome\_0to3: research
         11. ppvt\_3: IDK
         12. lnbw: IDK
         13. comp\_score\_5to6: some test score
         14. comp\_score\_7to10: some test score
         15. comp\_score\_11to14: some test score
         16. repeat: if you ad to repeat a grade
         17. learndis: 0 or 1
         18. hsgrad: 0 or 1
         19. somecoll: 0 or 1
         20. idle: IDK
         21. fphealth: IDK
         22. Age of mother when child was born (Teen pregnancy): numeric
         23. Age of child when attended HS (pre school age): numeric
         24. If child has mental disability (to control for): 0 or 1
         25. If child has limiting conditions (in case it might result in repetition): 0 or 1
         26. Birthweight: numeric
             1. Done, straight from the dataset.
   2. Data
4. Results
   1. Figure 1: recreate linear model from the paper, present table.
   2. Figure 2: present the best model to fit our research questions.
   3. Figure 3: we found that figure one (pros and cons)
   4. Figure 4: why our model would best answer our hypothesis.
5. Conclusion
6. Appendix