**Linear Modeling (Re-Analysis)**

* Re-implement model(s) in Table 3
  + Including full model
  + **This will be done with non-imputed, non-missing data rows as in the paper**
* Re-implement appendix tables (using imputed data here)
* Pinpoint the coefficients across these models

**Causal Analysis (New Stuff)**

* **We do a stratification-analysis similar to pgs 9-10 - Lecture 06** (outcome regression)
  + The “treatment” variable Z here is the racial category of the person
  + For these models, we use the **imputed** dataset
    - Note that the tercile-level analysis won’t work properly if data isn’t imputed (would have to use non-imputed instead)
  + Worth considering: Binarizing the race variable to be 0/1, with 1 (White) set to 0 and 2 and 3 (Black and Other) set to 1, I think binarizing would be better for interpretation purposes.
* With subsets, analyze the race coefficient at:
  + Cohort-level
  + Tercile-level
  + No time for cohort+tercile levels simultaneously, can mention this as other direction for the project
* Using the fit of first model, and full-model across all **(pared is not used in either model here, as in the paper)**, compare and consider the differences of the variable here.

Note for both: **No gender difference analysis as done in the original paper,** no time and would imply needing a second “treatment variable” set to gender. Would have to include gender-race interaction term coefficient if this were the case .