Team 2:**BLUE**

Duke MIDS
Master in Interdisciplinary Data Science

EFFECTS OF JOB TRAINING ON WAGES
PART 2

September, 25th 2020

Team Members:
Jiaman “Betty” Wu
Dapo Adegbile
Sutianyi “Bill” Wen
Christopher Oblak

INTRODUCTIONS

WE WANT TO FIND OUT….

If we can find a relationship between positive wages and if an individual received treatment?

Key Variables Collected

- treat — 1 IF RECEIVED JOB TRAINING, 0 IF DID NOT
- training — 0 IF PARTICIPANT DID NOT RECEIVE JOB TRAINING
- age — AGE IN YEARS
- educ — YEARS OF EDUCATION
- black — 1 IF RACE IS BLACK, 0 OTHERWISE
- hisp — 1 IF HISPANIC ETHNICITY, 0 OTHERWISE
- married — 1 IF MARRIED, 0 OTHERWISE
- nodegree — 1 IF PARTICIPANT DROPPED OUT OF HIGH SCHOOL, 0 OTHERWISE
- re74 — REAL ANNUAL EARNINGS IN 1974
- re75 — REAL ANNUAL EARNINGS IN 1975
- re78 — REAL ANNUAL EARNINGS IN 1978

Article Of Original Study :

- Lalonde, R. J. (1986), Evaluating the econometric evaluations of training programs with experimental data

Do You Need Better Job Training?

DATA ASSUMPTIONS AND INITIAL MANIPULATION

- **Created New Column –> Wage Factor**
  Based on the value of rev78:
  1 = Positive Wage
  0 = No Wage

- **Factored and Relabeled "nodegree"**
  - **"Not Drop out" = 0**
  - **"Drop Out" = 1**

- **Created A Race Factor Variable**
  0 = Neither, 1 = Hispanic, 2 = Black
  Base Factor for Race = 0

TABLE OF WAGE FACTOR		
Factor Variable	**Frequency**	**Percentage**
"Zero Wage" - 0	143	0.232
"Pos Wage" - 1	471	0.767

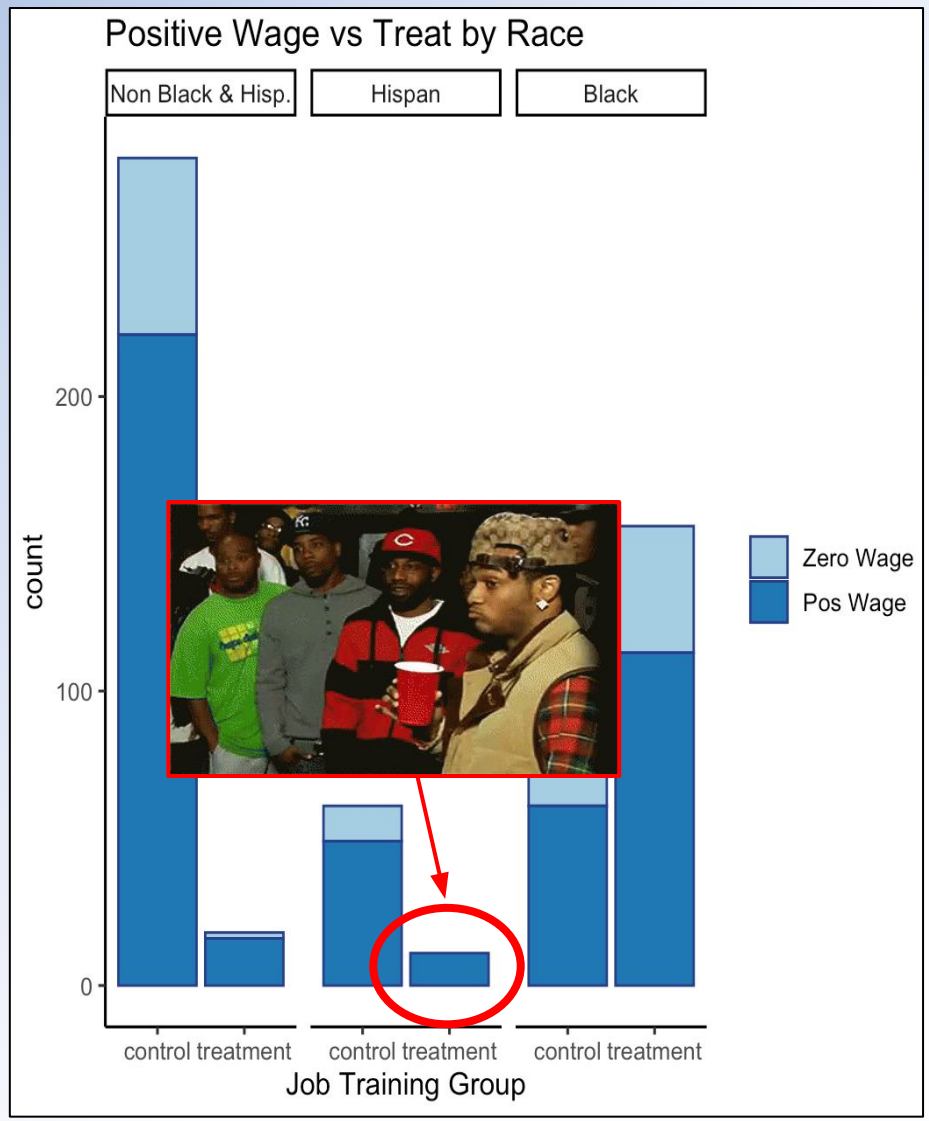
TABLE OF RACE DISTRIBUTION			
	Non Black and Hisp	**Hispan**	**Black**
Percentage	0.487	0.117	0.396
Var Count	299	72	243

Team Members: Jiaman "Betty" Wu, Dapo Adegbile, Sutianyi "Bill" Wen, Christopher Oblak

Team 2:BLUE

EDA FINDINGS (1 of 2)

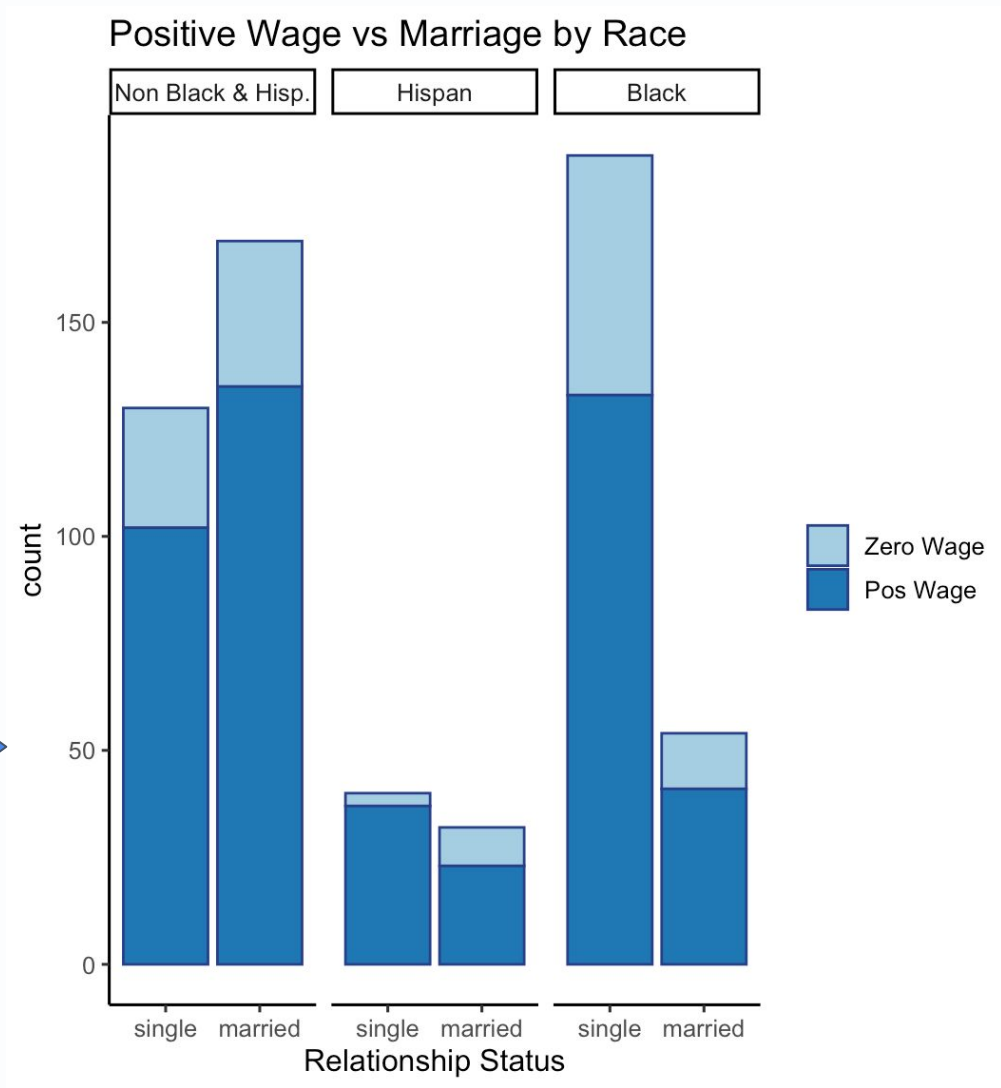
Duke MIDS
Master in Interdisciplinary Data Science



Worth noting that there are less observations of just hispanics.

Possible interaction because proportion of zero wages to pos wage shifts across race.

Another potential interaction because the proportion both zero and pos wages for single and married people varies by race

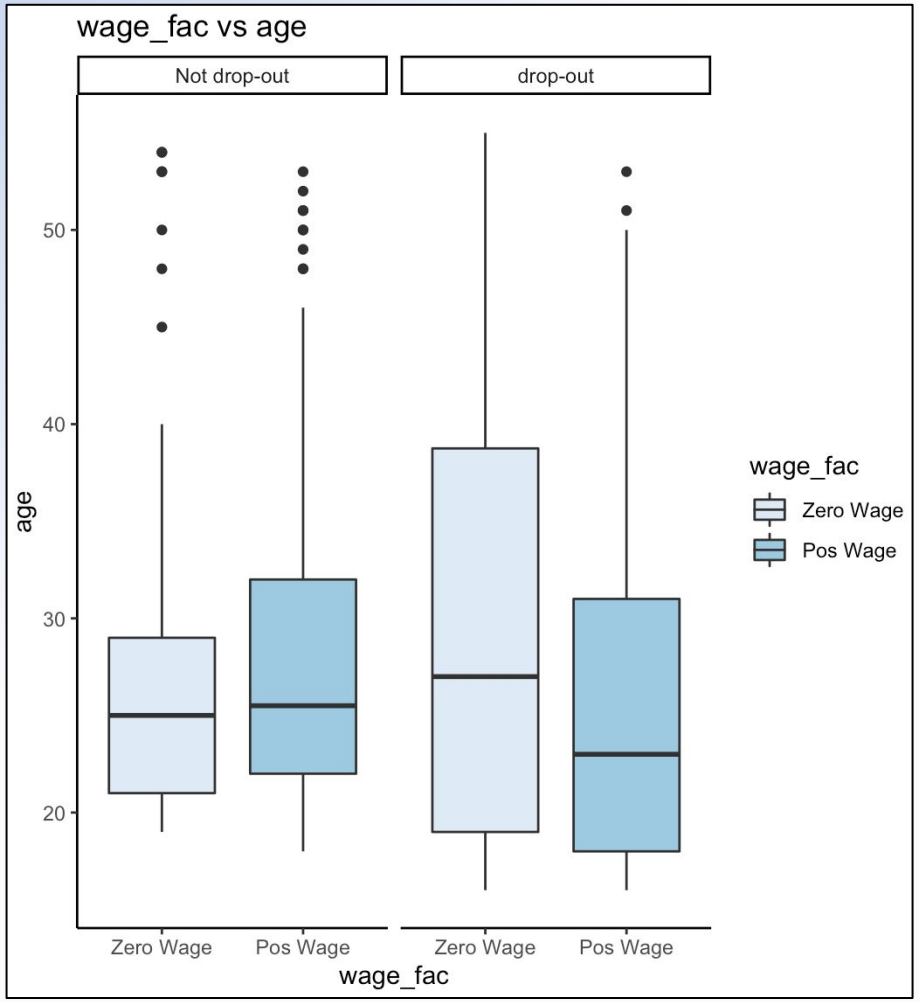


Team Members: Jiaman “Betty” Wu, Dapo Adegbile, Sutianyi “Bill” Wen, Christopher Oblak

Team 2:**BLUE**

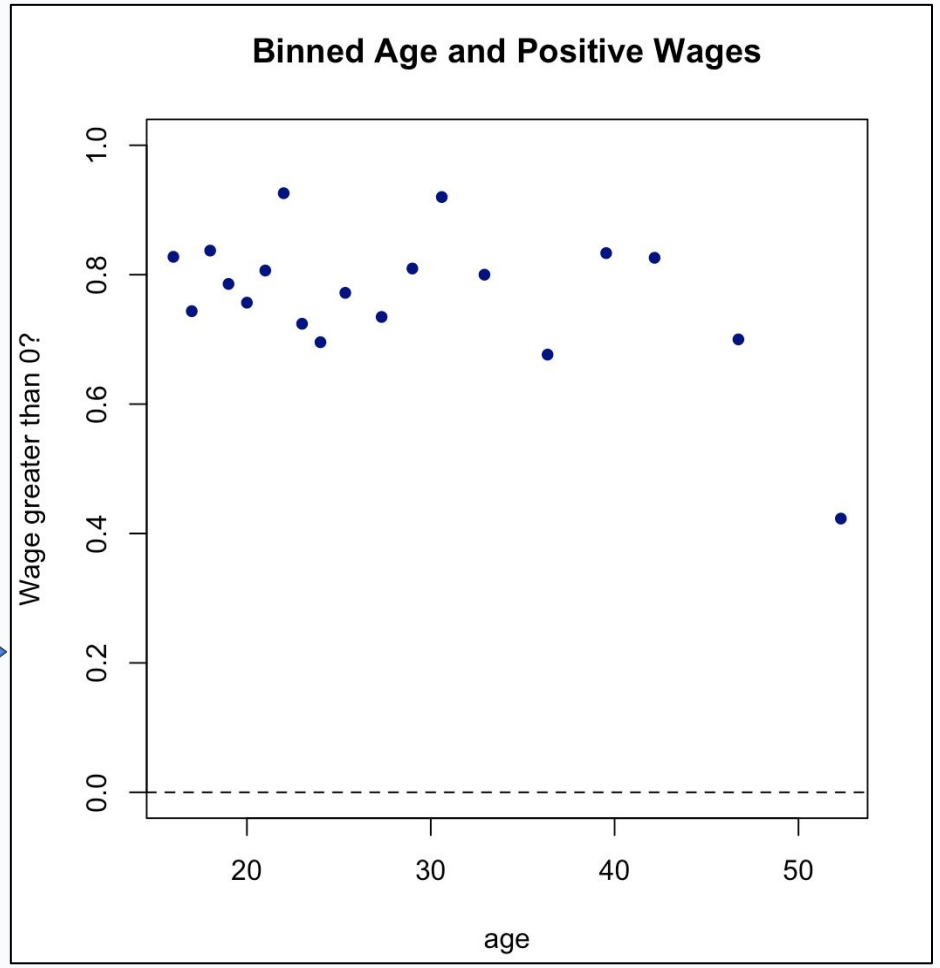
Duke MIDS
Master in Interdisciplinary Data Science

EDA FINDINGS (2 of 2)



The trend of ages vs wages is not consistent across degree status, leading us to believe this might be a relevant interaction.

This plot shows us that there is a large fluctuation of whether or not the wages are greater than 0 as a function of age, leading us to believe a polynomial transformation might be necessary.



MODEL DEVELOPMENT

INITIAL FULL MODEL BASED ON EDA FINDINGS

	wage		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	3.40	0.75 – 15.30	0.112
agec	1.03	0.98 – 1.07	0.264
agec_sq	1.00	1.00 – 1.00	**0.022**
educ	1.03	0.92 – 1.15	0.609
nodegree [1]	1.10	0.62 – 1.96	0.739
treat [1]	1.14	0.66 – 1.95	0.641
married [1]	1.15	0.71 – 1.86	0.578
race [Hispan]	1.22	0.60 – 2.49	0.587
race [Black]	0.57	0.34 – 0.95	**0.031**
agec * nodegree [1]	0.96	0.92 – 1.00	**0.045**
Observations	614		
R^2 Tjur	0.053		

AIC Forward Selection

NULL MODEL =
Wage ~ Treatment
FULL MODEL =
Wage ~ ***See Table

AIC FORWARD SELECTION MODEL

	wage		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	5.39	3.82 – 7.60	**<0.001**
treat [1]	1.15	0.69 – 1.93	0.591
agec_sq	1.00	1.00 – 1.00	**<0.001**
race [Hispan]	1.13	0.57 – 2.26	0.730
race [Black]	0.53	0.32 – 0.87	**0.012**
Observations	614		
R^2 Tjur	0.044		

Team Members: Jiaman “Betty” Wu, Dapo Adegbile, Sutianyi “Bill” Wen, Christopher Oblak

FINAL MODEL (1 of 2)

$$y_i \mid x_i \sim \ Bernoulli(\pi_i)$$

$$\log(\frac{\pi_i}{1-\pi_i}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5},$$

FINAL MODEL SELECTION/ASSESSMENT

- Using AIC Forward + Age Centered
- F-Test to verify our Final Model would not be significantly better with the full dependent variables/interactions we started with. **Pr(>Chi): 0.255**

	wage		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	5.29	3.67 – 7.63	**<0.001**
treat [1]	1.16	0.69 – 1.96	0.571
agec_sq	1.00	1.00 – 1.00	**0.012**
race [Hispan]	1.12	0.56 – 2.25	0.739
race [Black]	0.52	0.32 – 0.87	**0.012**
agec	1.00	0.97 – 1.02	0.783
Observations	614		
R^2 Tjur	0.044		

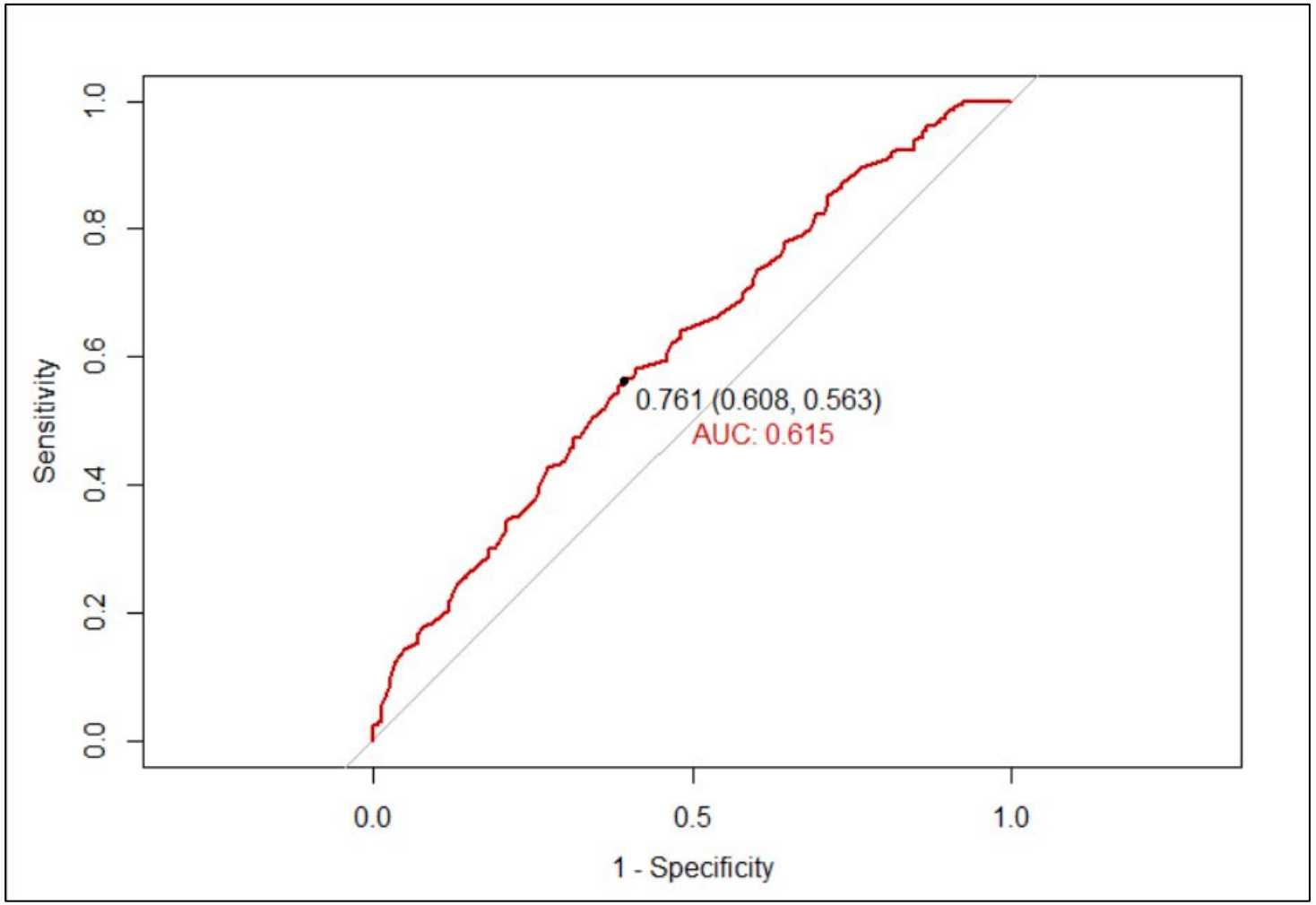
FINAL MODEL (2 of 2)

FINAL MODEL ASSESSMENT

- **Confusion Matrix**

	REFERENCE	
PREDICTION	**0**	**1**
0	11	2
1	132	469

- **Accuracy** = 0.782
- **Sensitivity** = 0.996
- **Specificity** = 0.077



CONCLUSIONS

FINDINGS

- Is there evidence that workers who receive job training tend to be more likely to have positive (non-zero) wages than workers who do not receive job training?
  a. There is **no evidence** showing that workers who receive job training tend to be more likely to have positive wages compared to workers without training because the variable “treat” is not significant according our final model.

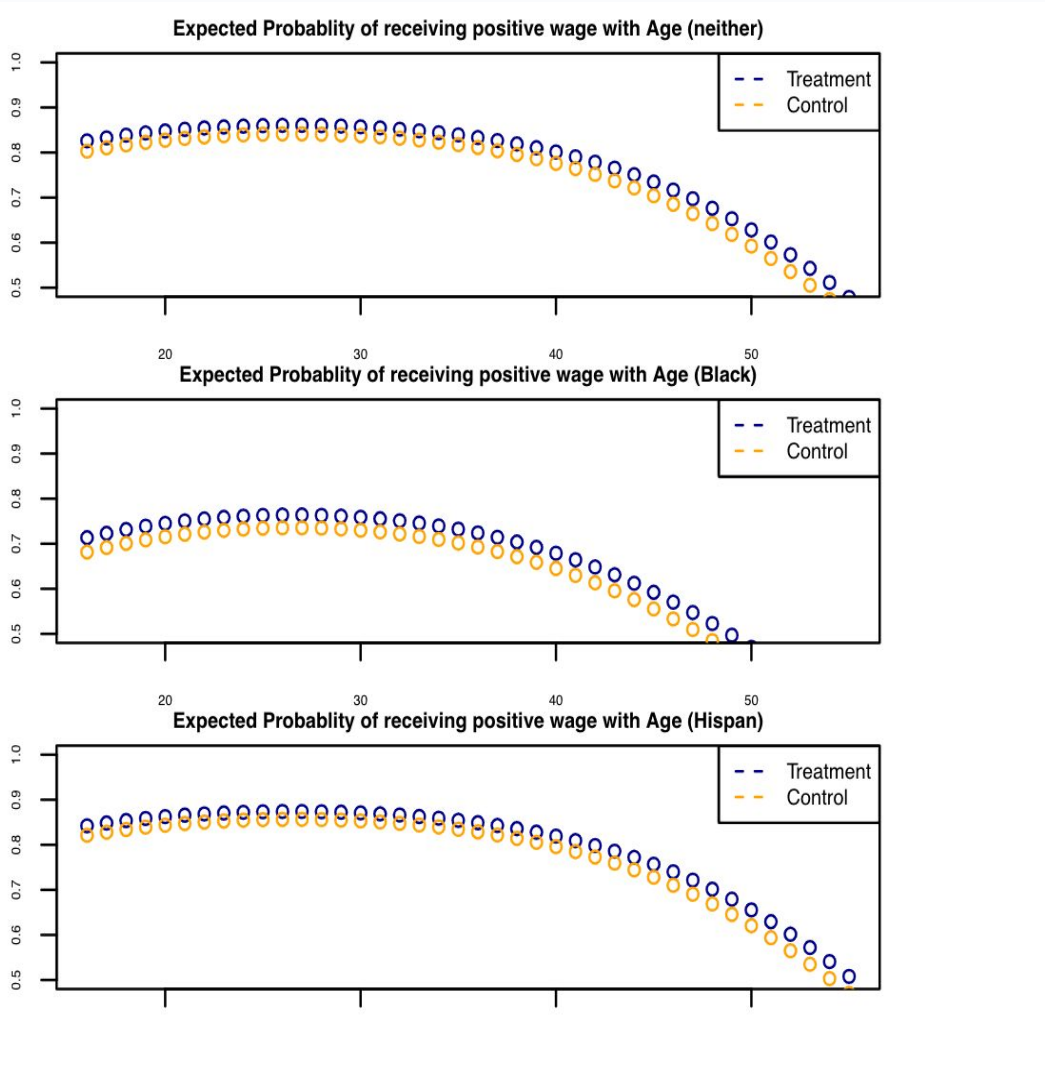
RESULTS

- Quantify the effect of the treatment, that is, receiving job training, on the odds of having non-zero wages.
  a. With everything staying the same, the odds of having a positive wage will be multiplied by **1.15** if the workers receive training.

- What is a likely range for the effect of training?
  a. With everything staying the same, the odds of having a positive wage after receiving the job training will be multiplied by between **0.69 and 1.96** compared to the control group, in a 95% confidence interval.

CONCLUSIONS

- Is there any evidence that the effects differ by demographic groups?
  a. From our EDA, it appears that the training effect is different by demographic groups. However, **we can't conclude** whether the effect by different demographic group is significant because there is 0 observation in treatment group that is hispanic having zero-wage.
- Are there other interesting associations with positive wages that are worth mentioning?
  a. There is evidence that the effects **differ** by demographic groups. Our model shows us that, compared to non blacks & non hispanics and hispanics, if someone is black, the odds of them having non-zero wages will be multiplied by 0.52. This is deemed statistically significant because our p-value of 0.012 is below the 0.05 threshold.



THANK YOU FOR WATCHING