Self Perception & Participation in the Arts

Wenxuan Zhu & Abbie Natkin

01
Data

02
Regression Analysis & IPW

03
Sensitivity Analysis

1. Does higher self-perception of creativity cause higher arts participation?

01

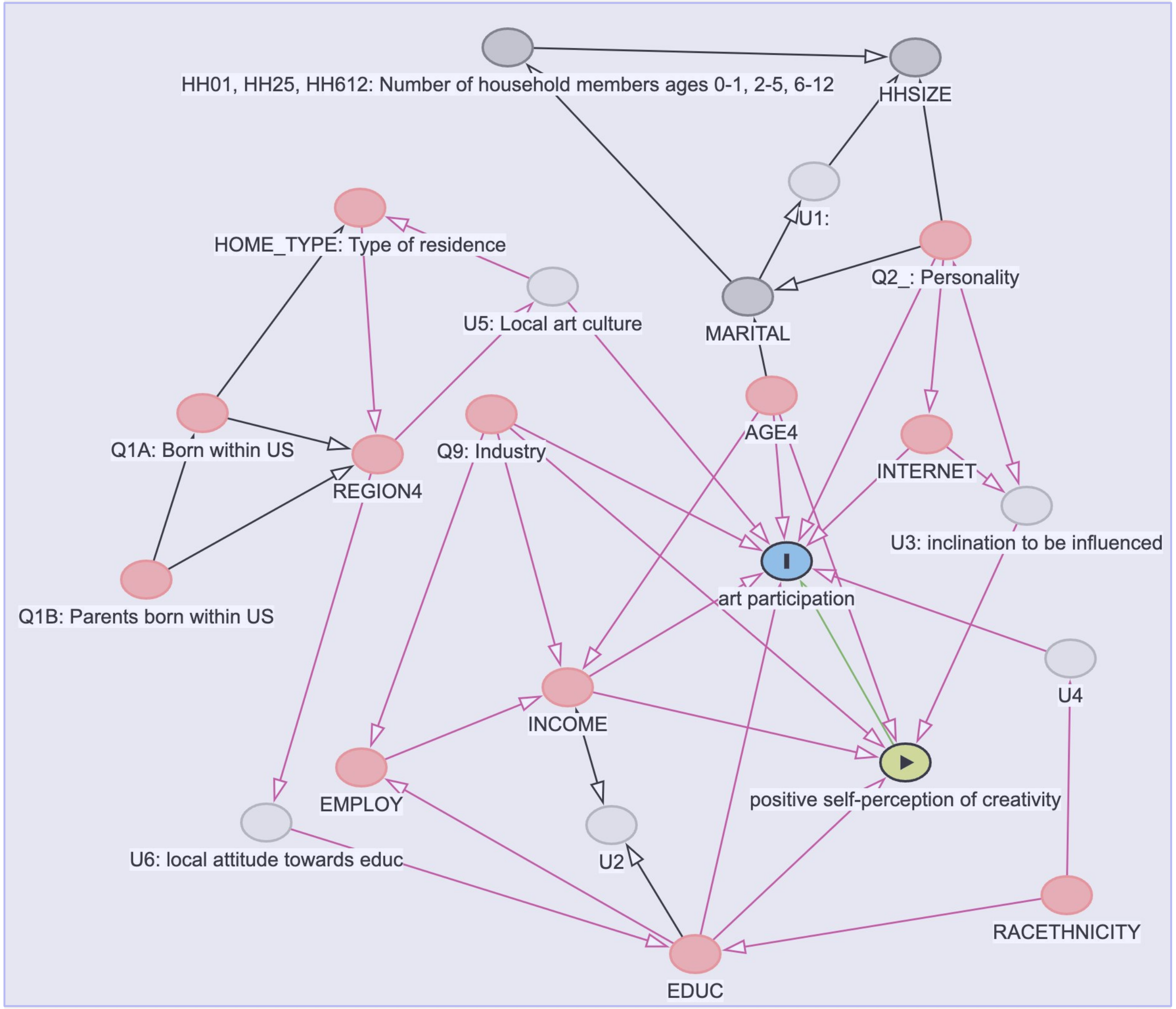
Data!

Data Overview

“Self perceptions of Creativity & Arts Participation,” 2018

- ★ Survey given to 3,447 adults in the US of differing demographics and socioeconomic status
- ★ “The primary objective of the national survey is to measure the ways that American adults experience and exercise creativity in their daily lives.”
- ★ “...self-perceptions of creativity across six creative "domains": artistic creativity, creativity in math/science, creativity in business/entrepreneurship, creativity in social settings, creativity in civic settings, and creativity in "everyday" activities”

DAG


HH01, HH25, HH612: Number of household members ages 0-1, 2-5, 6-12
HHSIZE
U1:
HOME_TYPE: Type of residence
Q2_: Personality
U5: Local art culture
MARITAL
AGE4
Q1A: Born within US
Q9: Industry
REGION4
INTERNET
U3: inclination to be influenced
Q1B: Parents born within US
art participation
U4
INCOME
EMPLOY
positive self-perception of creativity
U6: local attitude towards educ
U2
RACETHNICITY
EDUC

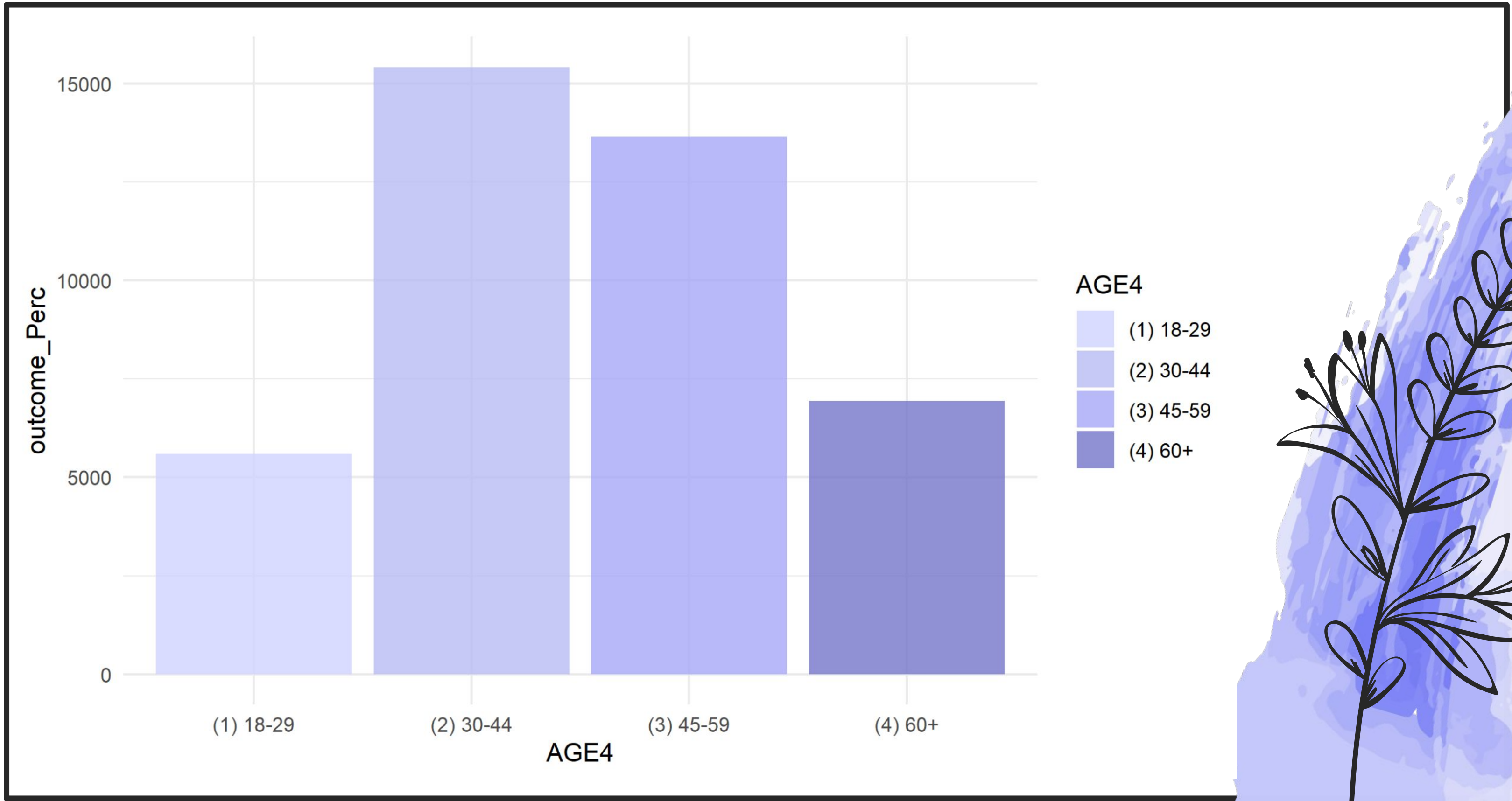
Important Variables

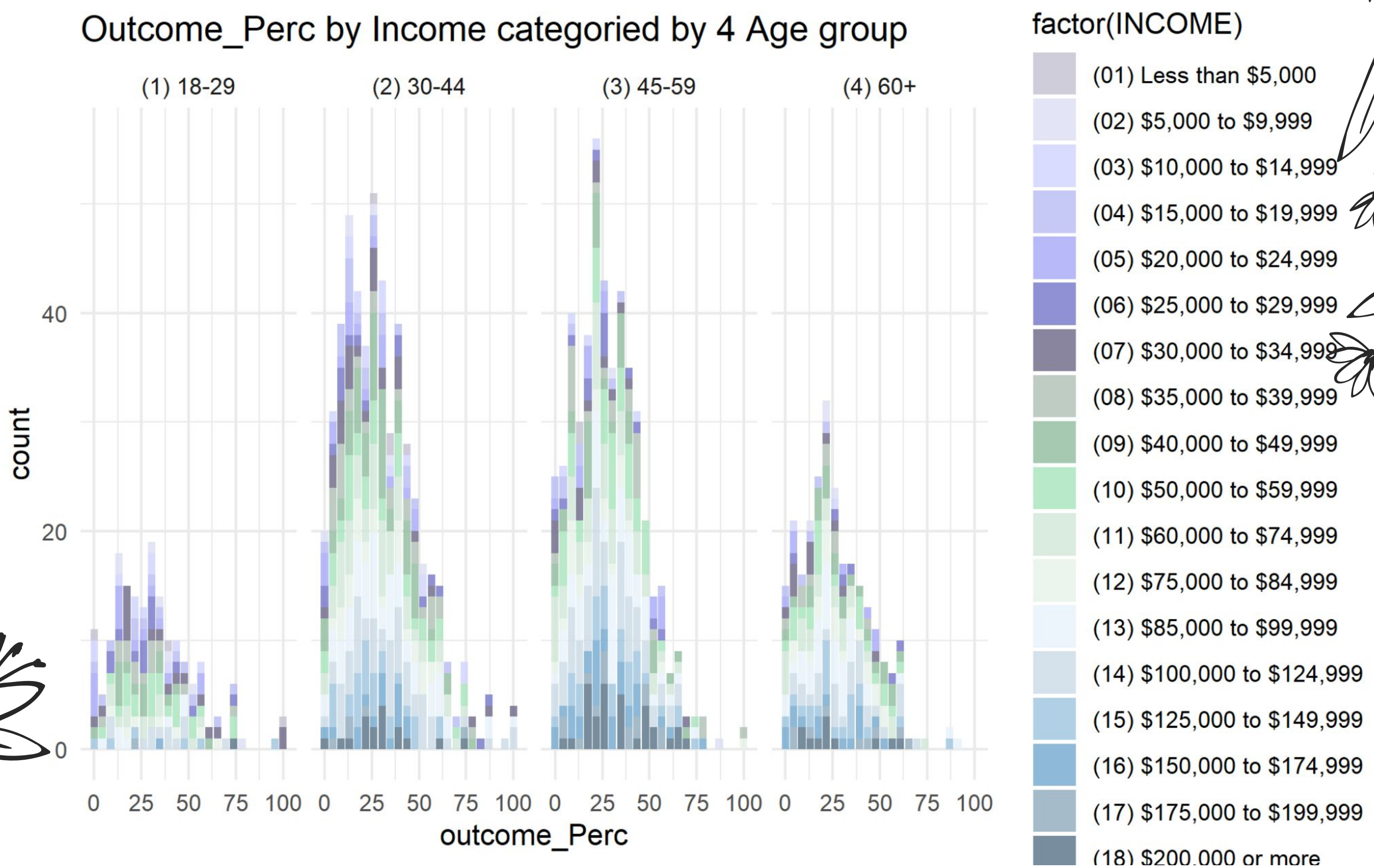
- **Outcome variable:**
  - ie. During the last 12 months did you go to a musical, play an instrument, go to an art exhibit, etc.
- **Treatment**
  - ie. Compared to people of approximately your age and life experience, how relatively creative are you in making up lyrics to a song, making up dance moves, solving puzzles, etc.

Minimal Sufficient Condition

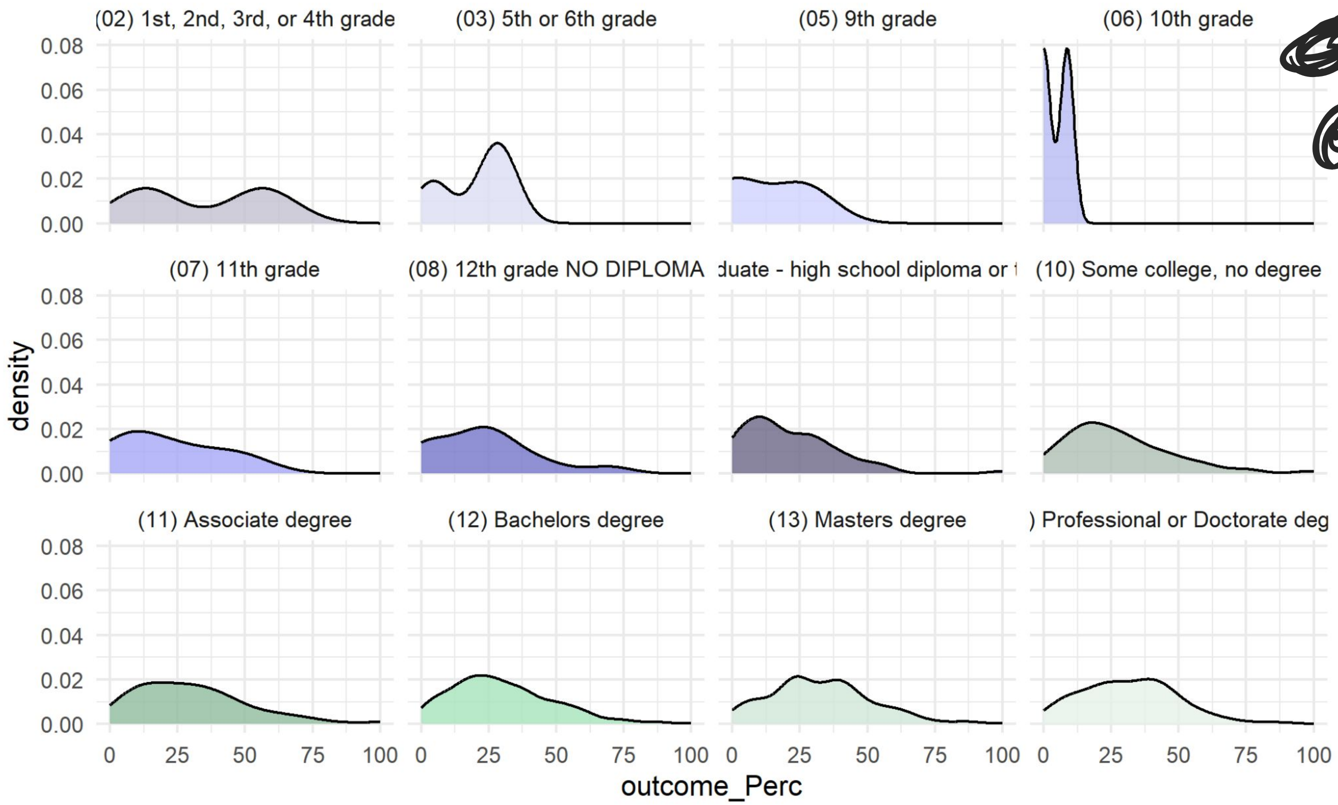
Minimal sufficient adjustment sets:

- AGE4
- EDUC
- INCOME
- INTERNET
- Personality_Perc

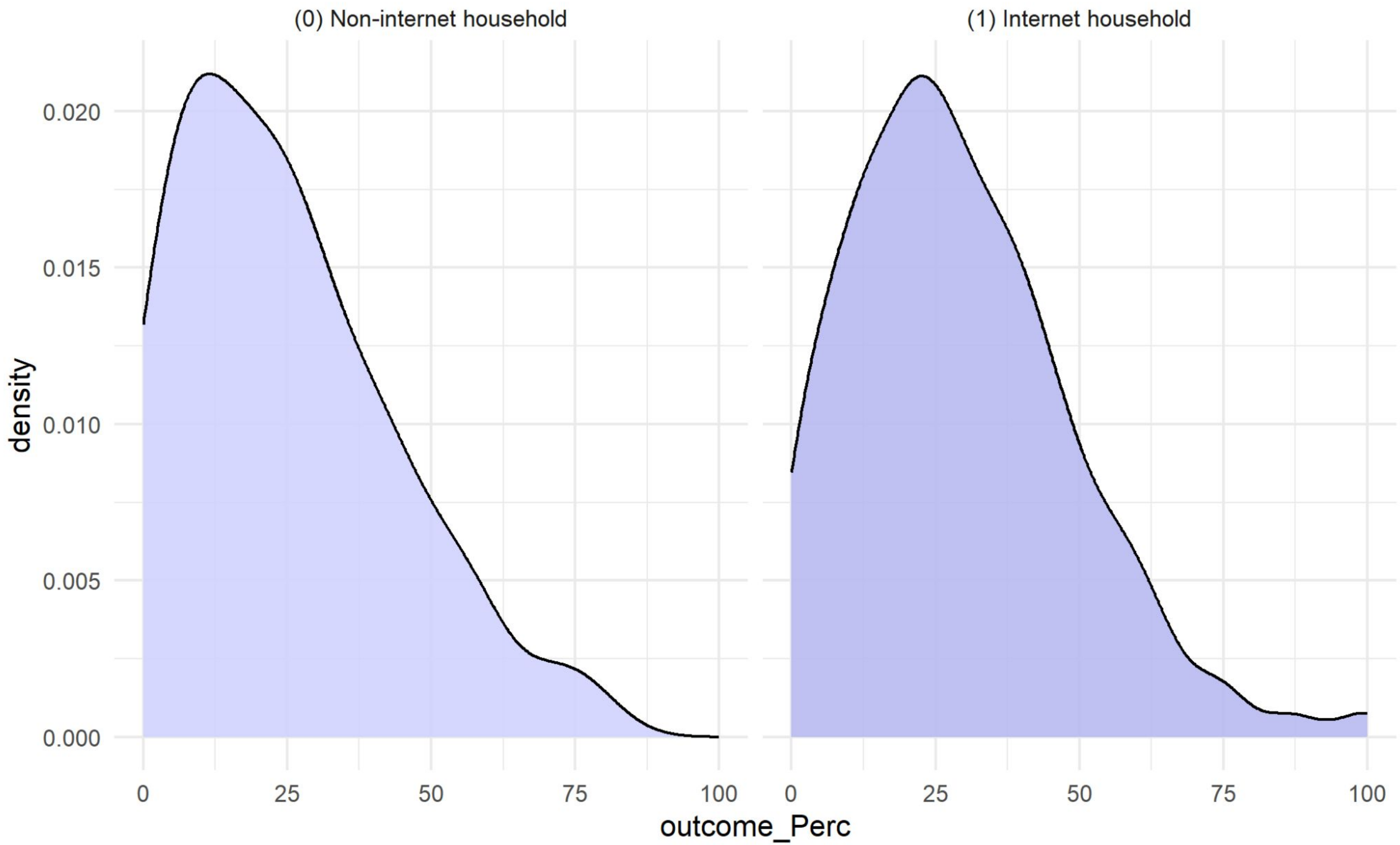

15000
10000
5000
0
outcome_Perc
(1) 18-29
(2) 30-44
(3) 45-59
(4) 60+
AGE4
AGE4
(1) 18-29
(2) 30-44
(3) 45-59
(4) 60+



Outcome_Perc by Income categoried by 4 Age group
(1) 18-29
(2) 30-44
(3) 45-59
(4) 60+
count
0
20
40
0 25 50 75 100
outcome_Perc
factor(INCOME)
(01) Less than $5,000
(02) $5,000 to $9,999
(03) $10,000 to $14,999
(04) $15,000 to $19,999
(05) $20,000 to $24,999
(06) $25,000 to $29,999
(07) $30,000 to $34,999
(08) $35,000 to $39,999
(09) $40,000 to $49,999
(10) $50,000 to $59,999
(11) $60,000 to $74,999
(12) $75,000 to $84,999
(13) $85,000 to $99,999
(14) $100,000 to $124,999
(15) $125,000 to $149,999
(16) $150,000 to $174,999
(17) $175,000 to $199,999
(18) $200,000 or more



(02) 1st, 2nd, 3rd, or 4th grade
(03) 5th or 6th grade
(05) 9th grade
(06) 10th grade
(07) 11th grade
(08) 12th grade NO DIPLOMA
duate - high school diploma or
(10) Some college, no degree
(11) Associate degree
(12) Bachelors degree
(13) Masters degree
) Professional or Doctorate deg
density
0.00
0.02
0.04
0.06
0.08
0
25
50
75
100
outcome_Perc

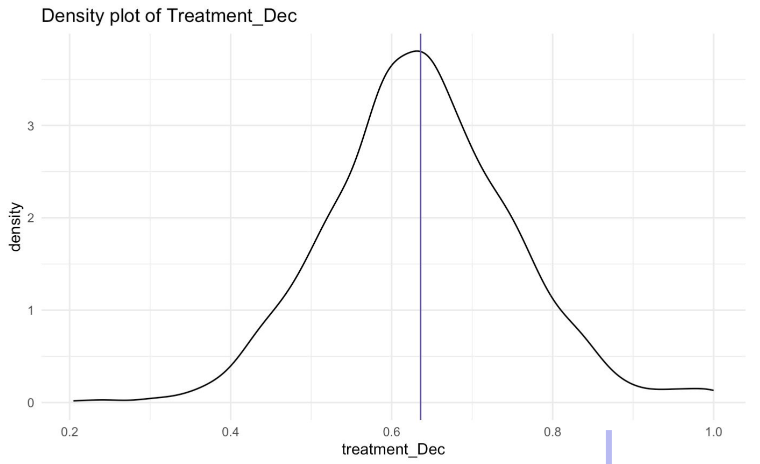

(0) Non-internet household
(1) Internet household
density
0.000
0.005
0.010
0.015
0.020
0
25
50
75
100
outcome_Perc

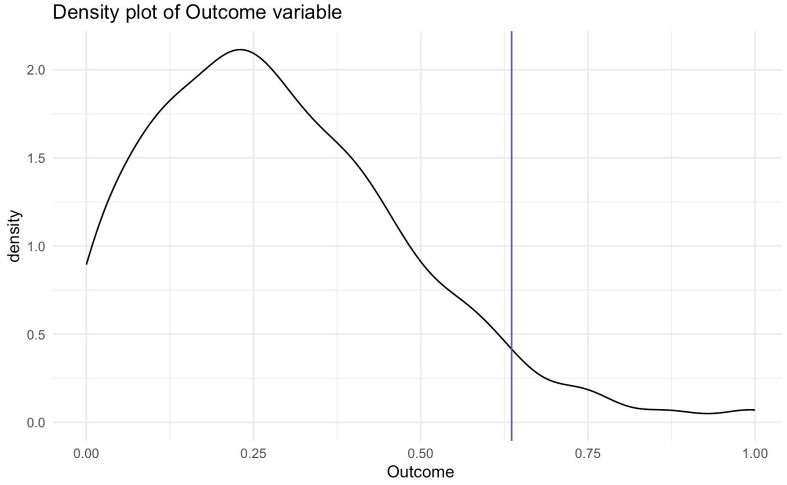
02

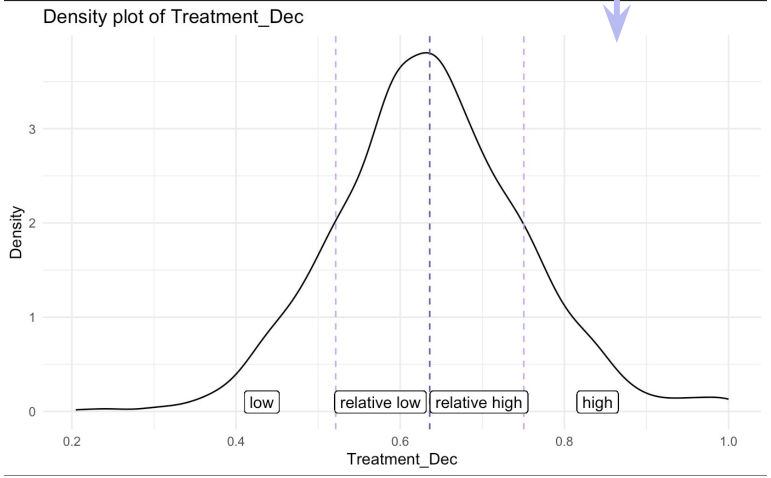
Regression Analysis!

Data Manipulation

- Treatment variable → 4 categories
- Outcome variable → binary
- Income → 3 categories







Outcome Regression

Goal:

- Condition on confounding variables to disable non-causal paths

Model Code:

- mod_overall <- glm(Outcome ~ treatment_Dec + AGE4 + EDUC + income_3 + ns(personality_Perc, 2) + INTERNET + Q9, family = "binomial", data = df)
- summary(mod_overall)

Output →

```
## Call:
## glm(formula = Outcome ~ treatment_Dec + AGE4 + EDUC + income_3 +
##     ns(personality_Perc, 2) + INTERNET + Q9, family = "binomial",
##     data = df)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.52401  -0.27665  -0.01838   0.20438   1.84755
##
## Coefficients:
##                                                          Estimate
## (Intercept)                                             -3.028368
## treatment_Dec                                            3.634765
## AGE4(2) 30-44                                           -0.120705
## AGE4(3) 45-59                                           -0.200314
## AGE4(4) 60+                                             -0.299955
## EDUC(03) 5th or 6th grade                               -0.590374
```

Average Causal Effect Estimates

```
## # A tibble: 1 × 8
##   term        estimate std.error statistic  p.value odds_ratio ci_lower ci_upper
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>      <dbl>    <dbl>    <dbl>
## 1 treatment_…     3.63     0.559      6.50 7.77e-11       37.9     12.7     113.
```

→ A strong positive causal relationship!

IPW Analysis

Random Forest:

- Recipe: data_rec <- recipe(treatment_Cat4_Cha ~ AGE4 + EDUC + income_3 + personality_Perc + INTERNET + Q9, data = df_rf)

Optimal workflow (`mtry` = 6) →

```
## ══ Workflow [trained] ══════════════════════════════════════════════════
## Preprocessor: Recipe
## Model: rand_forest()
##
## ── Preprocessor ────────────────────────────────────────────────────────
## 0 Recipe Steps
##
## ── Model ───────────────────────────────────────────────────────────────
## Ranger result
##
## Call:
##  ranger::ranger(x = maybe_data_frame(x), y = y, mtry = min_cols(~6,      x), num.trees = ~1000, min.node.size
= min_rows(~2, x), probability = ~TRUE,      importance = ~"impurity", num.threads = 1, verbose = FALSE,      see
d = sample.int(10^5, 1))
##
## Type:                             Probability estimation
## Number of trees:                  1000
## Sample size:                      1444
## Number of independent variables:  6
## Mtry:                             6
## Target node size:                 2
## Variable importance mode:         impurity
## Splitrule:                        gini
## OOB prediction error (Brier s.):  0.5453149
```

IPW Analysis Weighting

Code:

- rf6_predict <- as.data.frame(predict(data_fit_mtry6, new_data = df, type = "prob"))
- df_rf <- cbind(df, rf6_predict)
- df <- df_rf %>%

```
        mutate(
          ps = case_when(treatment_Cat4_num == 0 ~ .pred_low,
                 treatment_Cat4_num == 0.25 ~ `.pred_relative low`,
                 treatment_Cat4_num == 0.75 ~ `.pred_relative high`,
                 treatment_Cat4_num == 1 ~ .pred_high),
```

- ipw = 1/ps)

IPW Analysis Result

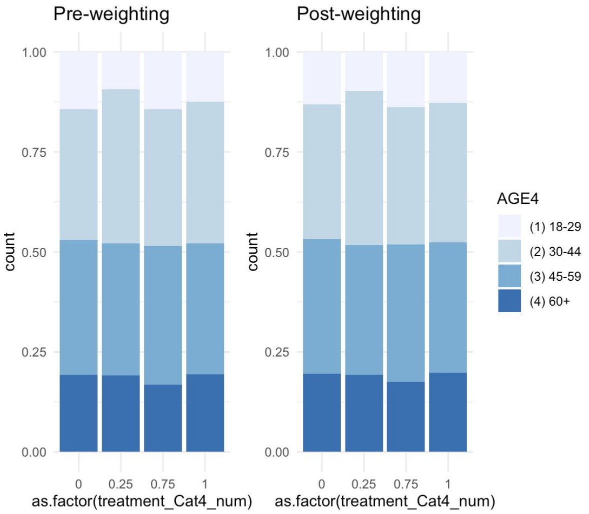
```
## Call:
## svyglm(formula = outcome_Cat2 ~ as.factor(treatment_Cat4_num),
##     design = design, family = "quasibinomial", data = df_subs)
##
## Survey design:
## svydesign(ids = ~0, weights = df_subs$ipw, data = df_subs)
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         -1.6902     0.2034  -8.310  < 2e-16 ***
## as.factor(treatment_Cat4_num)0.25    1.1288     0.2248   5.022 5.75e-07 ***
## as.factor(treatment_Cat4_num)0.75    2.0443     0.2241   9.124  < 2e-16 ***
## as.factor(treatment_Cat4_num)1       2.6553     0.2596  10.227  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.000693)
##
## Number of Fisher Scoring iterations: 4
```

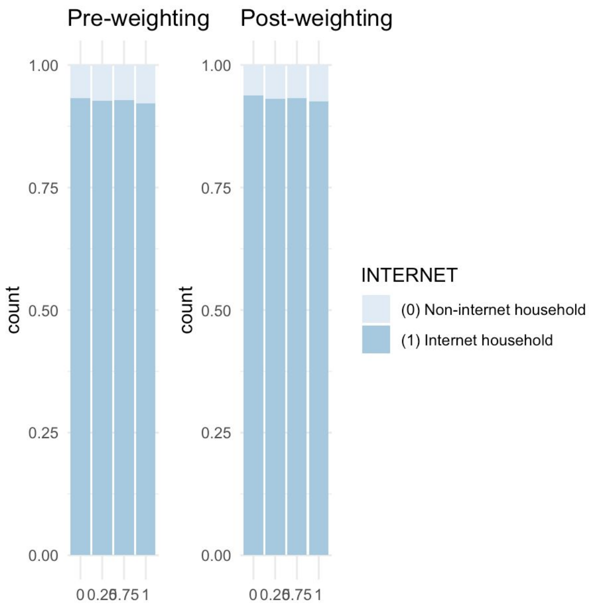
```
##                                       2.5 %    97.5 %
## (Intercept)                       -2.0892093 -1.291263
## as.factor(treatment_Cat4_num)0.25  0.6878818  1.569714
## as.factor(treatment_Cat4_num)0.75  1.6047517  2.483801
## as.factor(treatment_Cat4_num)1     2.1459538  3.164554
```

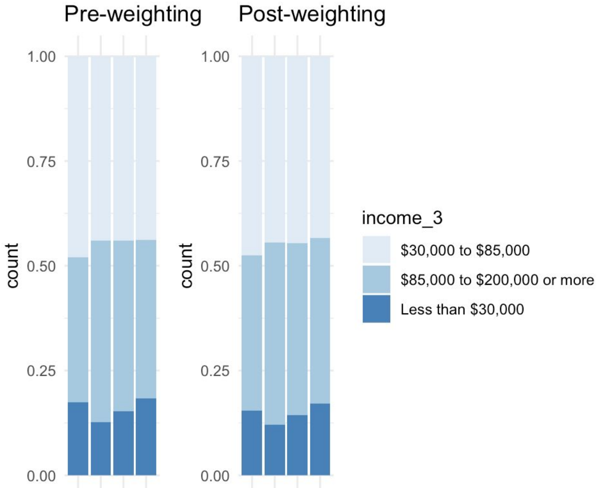
→ Consistent with outcome regression model

Balance Checking

- Nuanced differences between pre & post-weighting





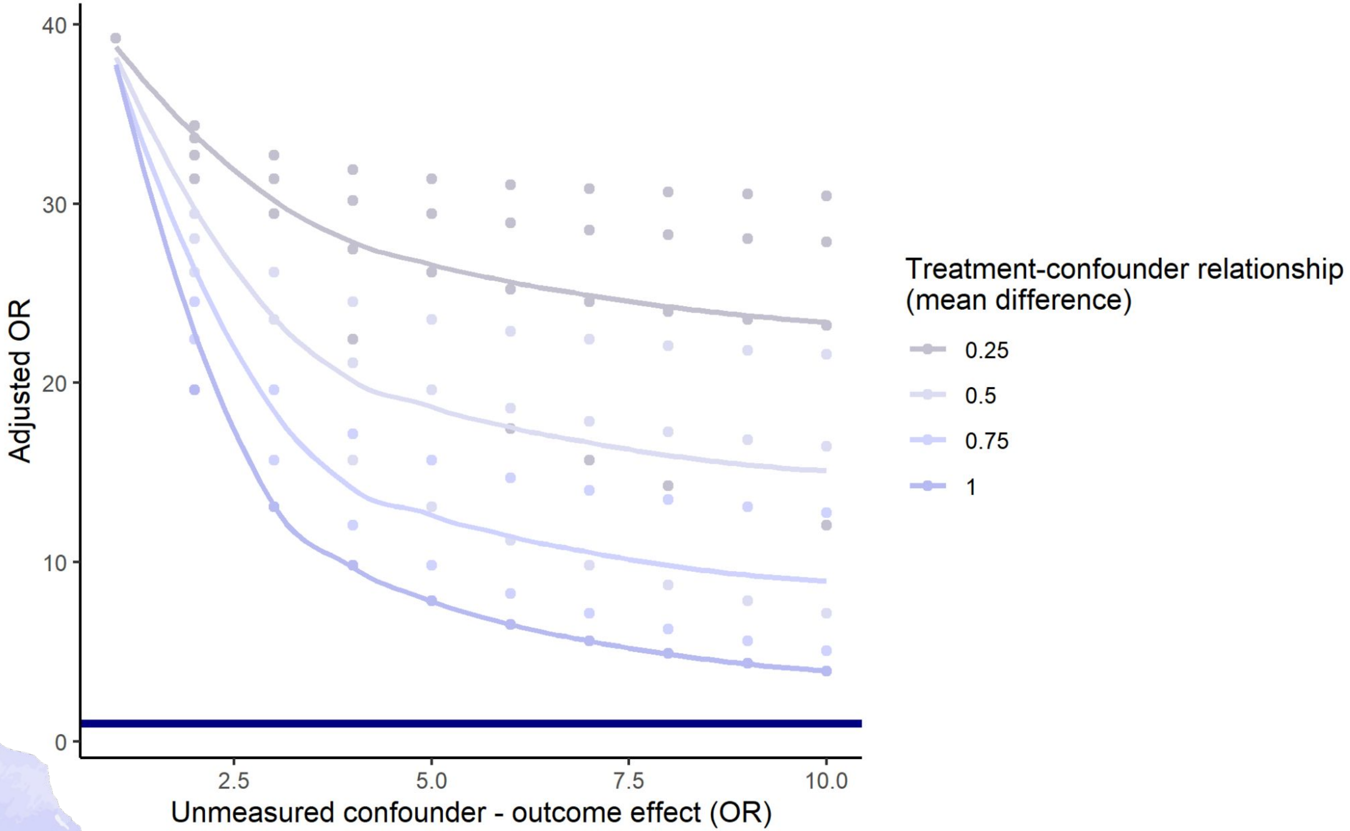


Statistical weights for the study eligible respondents were calculated using panel base sampling weights to start.

03

Sensitivity Analysis!

Sensitivity Analysis



2. Does a positive self-perception cause higher arts participation?

New Treatment Variable (Perception of Positive Personality Traits)

- ★ Thorough
- ★ Helpful
- ★ Energy
- ★ Clever
- ★ Enthusiastic
- ★ Trusting
- ★ Confident
- ★ Kind
- ★ Outgoing

```r
mod_PP <- glm(Outcome ~ secondPositivePersonality_Perc + AGE4 + EDUC + income_3 +
ns(personality_Perc, 2) + INTERNET + Q9, family = "binomial", data = df)
```

```
                 (Intercept) secondPositivePersonality_perc
                 -1.48401940                     0.01758004
[1] 0.02422978
```

Odds Ratio = 1.0177

Thanks!