

# Pilot Data

Tiffany Cheng, Matt Kinkley, Prachi Varma

7/11/2022

We created a survey in Qualtrics, which can be found [here](#). We then had workers on Amazon's MTurk complete the survey. We are worried about data quality, but this plot uses all data collected until we can refine what data to filter out. We got 146 total responses, although some may be duplicates.

```
d <- fread("qualtrics_mturk_data.csv")[-1,][-1,] %>% janitor::clean_names()

d[, treatment := 1 * (gender_t != "")]
d[, gender := ifelse(gender_t != "", gender_t, gender_c)]

d[, sc0 := as.numeric(sc0)]
d[, score := sc0/10]

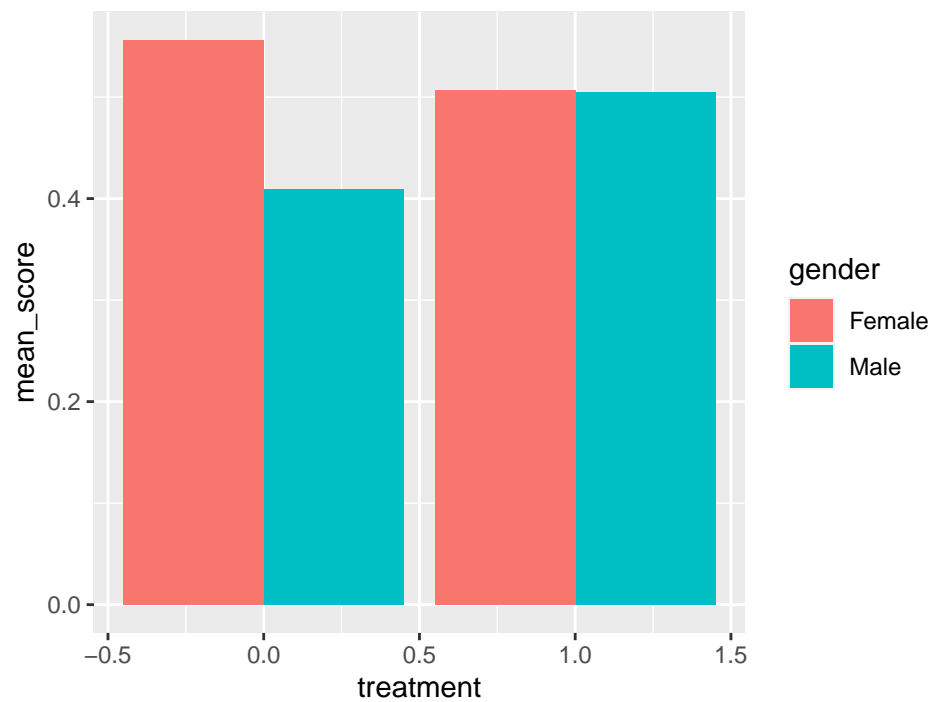
d[, age := as.numeric(age_1)]
d[, race := as.factor(race)]
d[, education := as.factor(education)]
d[, income := as.factor(income)]

d[, source := "mturk"]
```

## Plot of Data

```
d[, .(mean_score = mean(score)), by = .(treatment, gender)] %>%
  ggplot(aes(x = treatment, y = mean_score, fill = gender)) +
  geom_col(position = position_dodge()) +
  labs(title = "Avg. Score by Treatment Group and by Gender")
```

Avg. Score by Treatment Group and by Gender



## Covariate Balance

Conducting a covariate balance check by regressing the treatment variable on each of our four covariates results in a p-value of 0.0702 from the F-test. Since we fail to reject the null hypothesis at an  $\alpha = 0.05$  significance level, we can conclude that the covariate differences between treatment and control are all equal to 0.

```
cov_bal <- d[, lm(treatment ~ gender + age + race + education + income)]
```