

# R Coding Demonstration

## Week 4: Reducing Transphobia (Revisited)

Matthew Blackwell

Gov 51 (Harvard)

# Introduction

- Today we're going to cover some tools for exploring bivariate relationships.
- We'll use the data from the Broockman & Kalla (2016) transphobia study.
- Basic summary of experiment:
  - Randomly assigned door-to-door canvassers to two conditions
  - Conditions: perspective-taking script (treatment) or recycling script (placebo)
  - Follow up surveys at 3 days, 3 weeks, 6 weeks, and 3 months.

```
phobia <- read.csv("data/transphobia_all.csv")
```

---

Variable Name	Description
age	Age of the respondent in years
female	1=respondent marked "Female" on voter registration, 0 otherwise
voted_gen_14	1 if respondent voted in the 2014 general election
vote_gen_12	1 if respondent voted in the 2012 general election
treat_ind	1 if respondent was assigned to treatment, 0 for control
racename	character name of racial identity indicated on voter file
democrat	1 if respondent is a registered Democrat
therm_trans_t0	0-100 feeling therm. about transgender people at baseline
therm_trans_tX	0-100 feeling therm. about transgender people in Wave X after treatment
therm_obama_t0	0-100 feeling therm. about Barack Obama at baseline
therm_obama_tX	0-100 feeling therm. about Barack Obama in Wave X after treatment

---

# Question 1

Plot the average thermometer scores for transgender people in the treated and control group as a function of time.

- $t_1 = 3$  days,  $t_2 = 21$  days,  $t_3 = 42$  days,  $t_4 = 90$  days.

# Answer 1

```
treated <- subset(phobia, treat_ind == 1)
placebo <- subset(phobia, treat_ind == 0)

treated_avgs <- c(mean(treated$therm_trans_t0, na.rm = TRUE),
                  mean(treated$therm_trans_t1, na.rm = TRUE),
                  mean(treated$therm_trans_t2, na.rm = TRUE),
                  mean(treated$therm_trans_t3, na.rm = TRUE),
                  mean(treated$therm_trans_t4, na.rm = TRUE))

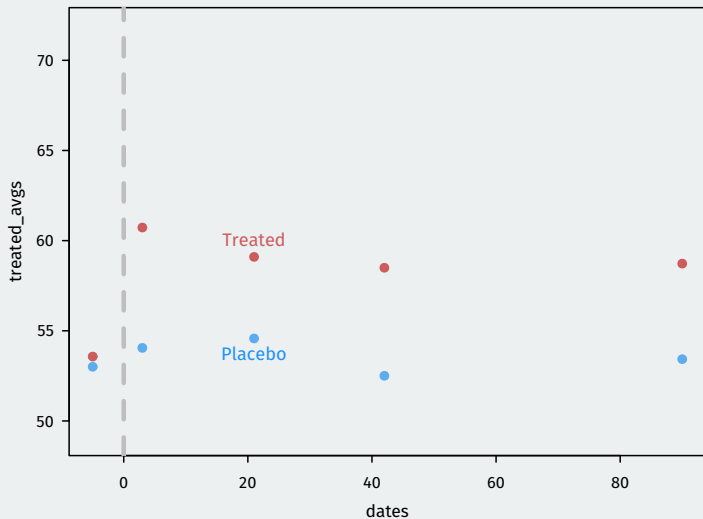
placebo_avgs <- c(mean(placebo$therm_trans_t0, na.rm = TRUE),
                  mean(placebo$therm_trans_t1, na.rm = TRUE),
                  mean(placebo$therm_trans_t2, na.rm = TRUE),
                  mean(placebo$therm_trans_t3, na.rm = TRUE),
                  mean(placebo$therm_trans_t4, na.rm = TRUE))
```

# Answer 1 (cont'd)

```
dates <- c(-5, 3, 21, 42, 90)
plot(x = dates, y = treated_avgs, pch = 19,
     col = "indianred", ylim = c(49, 72))
points(x = dates, y = placebo_avgs, pch = 19, col = "steelblue2")
abline(v = 0, lty = 2, col = "grey", lwd = 3)

## Write labels next to the first points
## "pos" here places them below (1) and above (3) the
## points indicated by x/y
text(x = dates[3], y = treated_avgs[3], labels = "Treated",
     col = "indianred", pos = 3)
text(x = dates[3], y = placebo_avgs[3], labels = "Placebo",
     col = "dodgerblue", pos = 1)
```

# Answer 1 (cont'd)



## Question 2

Create a similar plot of average thermometer scores about Obama (`therm_obama_tX`) by treatment group against the date.

Note that the Obama thermometer score was not measure in wave 3.



## Answer 2

```
## more compact approach using colMeans() function
## first we make a vector of the obama therm variable names
obama_names <- c("therm_obama_t0", "therm_obama_t1", "therm_obama_t2",
                "therm_obama_t4")

## colMeans calculate the mean of each column of a data frame/matrix
bo_treated_avgs <- colMeans(treated[, obama_names], na.rm = TRUE)

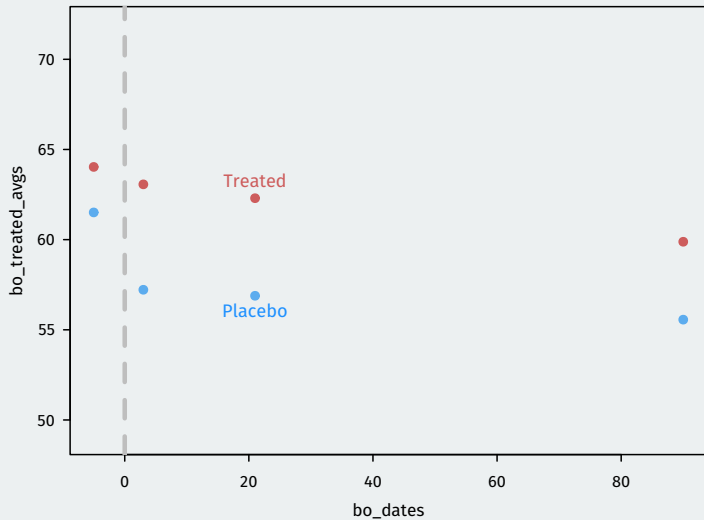
bo_placebo_avgs <- colMeans(placebo[, obama_names], na.rm = TRUE)
```

## Answer 2 (cont'd)

```
bo_dates <- c(-5, 3, 21, 90)
plot(x = bo_dates, y = bo_treated_avgs, pch = 19,
     col = "indianred", ylim = c(49, 72))
points(x = bo_dates, y = bo_placebo_avgs, pch = 19,
       col = "steelblue2")
abline(v = 0, lty = 2, col = "grey", lwd = 3)

text(x = bo_dates[3], y = bo_treated_avgs[3], labels = "Treated",
     col = "indianred", pos = 3)
text(x = bo_dates[3], y = bo_placebo_avgs[3], labels = "Placebo",
     col = "dodgerblue", pos = 1)
```

## Answer 2 (cont'd)



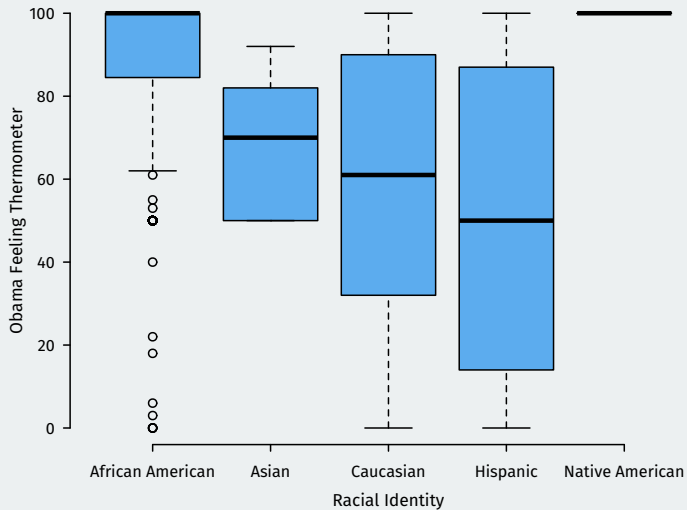
## Question 3

Let's see how feelings about Obama vary by race. Make separate boxplots of Obama feeling thermometers at baseline for each racial group in the sample.

# Answer 3

```
boxplot(therm_obama_t0 ~ racename, data = phobia,  
        xlab = "Racial Identity", ylab = "Obama Feeling Thermometer",  
        las = 1, col = "steelblue2", frame = FALSE)
```

# Answer 3 (cont'd)



## Question 4

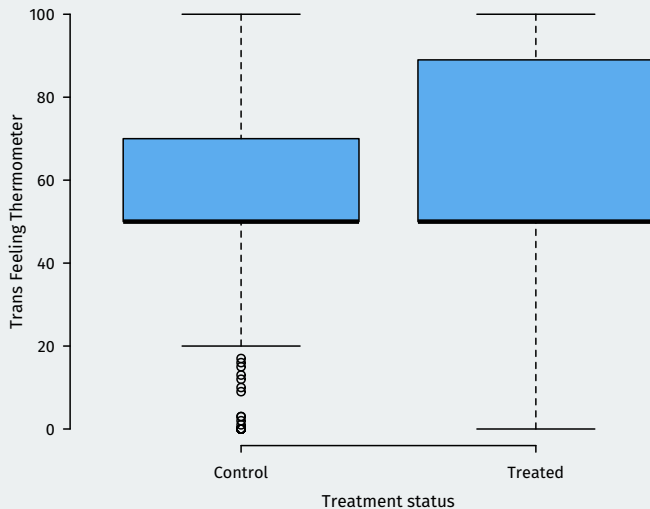
Let's compare more than just the average effects of treatment. Create two side-by-side boxplots of thermometer scores of transgender people in Wave 1 for both the treated and control groups. Use names to give informative labels to the boxplots.

## Answer 4

```
boxplot(therm_trans_t1 ~ treat_ind, data = phobia,  
        names = c("Control", "Treated"),  
        xlab = "Treatment status", ylab = "Trans Feeling Thermometer",  
        las = 1, col = "steelblue2", frame = FALSE)
```



# Answer 4 (con't)



## Question 5

Let's play with correlations. What is the correlation between all of the Obama thermometer scores?

# Answer 5

```
obama_names <- c("therm_obama_t0", "therm_obama_t1",  
                "therm_obama_t2", "therm_obama_t4")
```

```
## hm, doesn't work NAs  
cor(phobia[, obama_names])
```

```
##           therm_obama_t0 therm_obama_t1 therm_obama_t2 therm_obama_t4  
## therm_obama_t0           1             NA             NA             NA  
## therm_obama_t1           NA            1             NA             NA  
## therm_obama_t2           NA            NA             1             NA  
## therm_obama_t4           NA            NA             NA             1
```

## Answer 5 (cont'd)

```
## calculate any correlation with the observed values for that pair
cor(phobia[, obama_names], use = "pairwise")
```

```
##           therm_obama_t0 therm_obama_t1 therm_obama_t2 therm_obama_t4
## therm_obama_t0           1.000           0.803           0.765           0.632
## therm_obama_t1           0.803           1.000           0.700           0.541
## therm_obama_t2           0.765           0.700           1.000           0.606
## therm_obama_t4           0.632           0.541           0.606           1.000
```

```
## drop any rows with NAs
cor(phobia[, obama_names], use = "complete.obs")
```

```
##           therm_obama_t0 therm_obama_t1 therm_obama_t2 therm_obama_t4
## therm_obama_t0           1.000           0.785           0.761           0.646
## therm_obama_t1           0.785           1.000           0.663           0.553
## therm_obama_t2           0.761           0.663           1.000           0.601
## therm_obama_t4           0.646           0.553           0.601           1.000
```

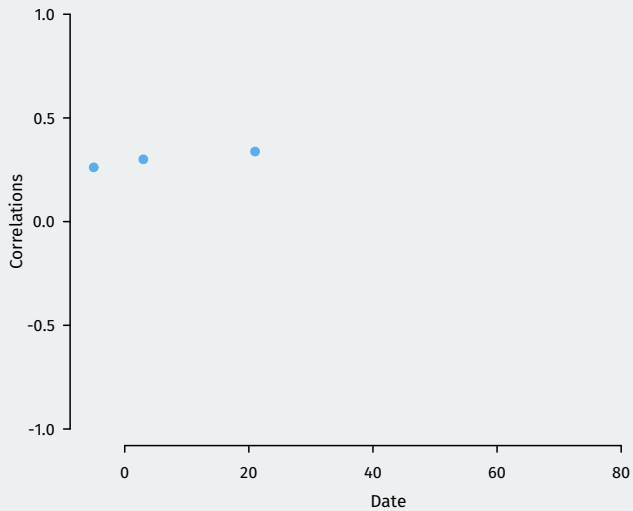
## Question 6

Calculate the correlation between transgender thermometer scores and Obama thermometer scores in each survey (except wave 3). Use pairwise observed values to calculate these correlations. Save them as a vector and plot them in a scatter plot against time.

## Answer 6

```
cor_t0 <- cor(phobia$therm_trans_t0, phobia$therm_obama_t0,  
             use = "pairwise")  
cor_t1 <- cor(phobia$therm_trans_t1, phobia$therm_obama_t1,  
             use = "pairwise")  
cor_t2 <- cor(phobia$therm_trans_t2, phobia$therm_obama_t2,  
             use = "pairwise")  
cor_t4 <- cor(phobia$therm_trans_t4, phobia$therm_obama_t4,  
             use = "pairwise")  
  
obama_trans_cors <- c(cor_t0, cor_t1, cor_t2, cor_t4)  
plot(x = bo_dates, y = obama_trans_cors, ylim = c(-1, 1),  
     xlab = "Date", ylab = "Correlations", pch = 19,  
     col = "steelblue2", las = 1, frame = FALSE)
```

# Answer 6



## Question 7 (Challenge question)

Calculate one correlation from this data set without using the `cor` function.



# Answer 7

```
cor_data <- na.omit(phobia[, c("therm_obama_t0", "therm_obama_t1")])
n <- nrow(cor_data)

mean_obama_t0 <- mean(cor_data$therm_obama_t0)
sd_obama_t0 <- sd(cor_data$therm_obama_t0)
z_obama_t0 <- (cor_data$therm_obama_t0 - mean_obama_t0) / sd_obama_t0

mean_obama_t1 <- mean(cor_data$therm_obama_t1)
sd_obama_t1 <- sd(cor_data$therm_obama_t1)
z_obama_t1 <- (cor_data$therm_obama_t1 - mean_obama_t1) / sd_obama_t1

sum(z_obama_t0 * z_obama_t1) / (n - 1)
```

```
## [1] 0.803
```

```
cor(phobia$therm_obama_t0, phobia$therm_obama_t1, use = "pairwise")
```

```
## [1] 0.803
```