Pilot data analysis

Social media team

```
library(ggplot2)
library(data.table)
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(pwr)
library(lsr)
```

Read in data and reformat

```
d1 <- read.csv("~/Documents/berkeley/W241/final-project/Berkeley Social Commitment: Follow-up_December
                   stringsAsFactors = F, na.strings=c("","NA"))
ppl <- read.csv("~/Documents/berkeley/W241/final-project/241 Participant List - Participants.csv",
                   stringsAsFactors = F, na.strings=c("","NA"))
ppl <- ppl[c(1,3,9:10)]

## karen deane twice
## joann podracky not mapping

d1 <- d1[c("Q3", "Q4", "Q17", "Q5","Q6","Q7", "Q9","Q10","Q11", "Q13","Q14","Q15")]
d1 <- d1[3:nrow(d1),]
d1 <- d1[rowSums(is.na(d1)) != ncol(d1),]
d1 <- d1[(d1$Q3 != "s" & d1$Q3 != "test"),]

d2 <- merge(d1[c("Q3", "Q4", "Q17", "Q6","Q7", "Q10","Q11", "Q14","Q15")], ppl[c(2,3)], by.x = "Q3", by
d2 <- d2[!is.na(d2$Q6) & !is.na(d2$Q7),]
d2 <- d2[!duplicated(d2),]

d2[d2$Q3 == "JoAnn Podracky ",]$Treatment.Seq <- 2
d2[d2$Q3 == "Melissa Tyburczy ",]$Treatment.Seq <- 6

d2$Q7 <- as.numeric(gsub("\\,", "", d2$Q7))
d2$Q11 <- as.numeric(gsub("\\,", "", d2$Q11))
d2$Q15 <- as.numeric(gsub("\\,", "", d2$Q15))
```

```
## Warning: NAs introduced by coercion
```

```r
# Not applicable = 0
# Through digital means = 1
# In person = 2

d2[d2$Q6 %like% "Not applicable", ]$Q6 <- 0
d2[d2$Q10 %like% "Not applicable", ]$Q10 <- 0
d2[d2$Q14 %like% "Not applicable", ]$Q14 <- 0

d2[d2$Q6 %like% "In person", ]$Q6 <- 2
d2[d2$Q10 %like% "In person", ]$Q10 <- 2
d2[d2$Q14 %like% "In person", ]$Q14 <- 2

d2[d2$Q6 %like% "Through digital means", ]$Q6 <- 1
d2[d2$Q10 %like% "Through digital means", ]$Q10 <- 1
d2[d2$Q14 %like% "Through digital means", ]$Q14 <- 1

names(d2) <- c("name", "email", "know_us", "day1_treatment", "step_day1", "day2_treatment", "step_day2"

# for treatment received on what day below
# 1 = day1
# 2 = day2
# 3 = day3

d2$treatment1_day_received <- NA
d2$treatment2_day_received <- NA
d2$treatment3_day_received <- NA

d2[d2$treatment_seq == 1,]$treatment1_day_received <- 1
d2[d2$treatment_seq == 1,]$treatment2_day_received <- 2
d2[d2$treatment_seq == 1,]$treatment3_day_received <- 3

d2[d2$treatment_seq == 2,]$treatment1_day_received <- 1
d2[d2$treatment_seq == 2,]$treatment2_day_received <- 3
d2[d2$treatment_seq == 2,]$treatment3_day_received <- 2

d2[d2$treatment_seq == 3,]$treatment1_day_received <- 2
d2[d2$treatment_seq == 3,]$treatment2_day_received <- 1
d2[d2$treatment_seq == 3,]$treatment3_day_received <- 3

d2[d2$treatment_seq == 4,]$treatment1_day_received <- 2
d2[d2$treatment_seq == 4,]$treatment2_day_received <- 3
d2[d2$treatment_seq == 4,]$treatment3_day_received <- 1

d2[d2$treatment_seq == 5,]$treatment1_day_received <- 3
d2[d2$treatment_seq == 5,]$treatment2_day_received <- 1
d2[d2$treatment_seq == 5,]$treatment3_day_received <- 2

d2[d2$treatment_seq == 6,]$treatment1_day_received <- 3
d2[d2$treatment_seq == 6,]$treatment2_day_received <- 2
d2[d2$treatment_seq == 6,]$treatment3_day_received <- 1
```

Checking for ordering/priming effect For this part, we're interested in seeing if receiving certain treatment on what day has any effect on the outcome, so the outcome will be whether or not the subject had more than 5000 steps on the third day. Analysis will be limited to those who actually followed their treatment assignment.

```
d2_a <- d2
d2_a$outcome <- ifelse(d2_a$step_day3 > 5000, 1, 0)

# limit to those who followed directions
d2_a <- rbind(d2_a[d2_a$treatment_seq == 1 & d2_a$day1_treatment == 0 & d2_a$day2_treatment == 1 & d2_a$
              d2_a[d2_a$treatment_seq == 2 & d2_a$day1_treatment == 0 & d2_a$day2_treatment == 2 & d2_a$
              d2_a[d2_a$treatment_seq == 3 & d2_a$day1_treatment == 1 & d2_a$day2_treatment == 0 & d2_a$
              d2_a[d2_a$treatment_seq == 4 & d2_a$day1_treatment == 1 & d2_a$day2_treatment == 2 & d2_a$
              d2_a[d2_a$treatment_seq == 5 & d2_a$day1_treatment == 2 & d2_a$day2_treatment == 0 & d2_a$
              d2_a[d2_a$treatment_seq == 6 & d2_a$day1_treatment == 2 & d2_a$day2_treatment == 1 & d2_a$

# interested in interaction between treatment and day
m_a <- lm(outcome ~ day1_treatment:treatment1_day_received + day2_treatment:treatment2_day_received + da
             step_day1 + step_day2, data = d2_a)
m_a
```

```
##
## Call:
## lm(formula = outcome ~ day1_treatment:treatment1_day_received +
##     day2_treatment:treatment2_day_received + day3_treatment:treatment3_day_received +
##     step_day1 + step_day2, data = d2_a)
##
## Coefficients:
##                             (Intercept)
##                               5.555e-01
##                               step_day1
##                               2.691e-05
##                               step_day2
##                               1.003e-05
## day1_treatment0:treatment1_day_received
##                              -8.404e-02
## day1_treatment1:treatment1_day_received
##                              -3.326e-02
## day1_treatment2:treatment1_day_received
##                                      NA
## day2_treatment0:treatment2_day_received
##                              -4.473e-02
## day2_treatment1:treatment2_day_received
##                              -2.387e-02
## day2_treatment2:treatment2_day_received
##                                      NA
## day3_treatment0:treatment3_day_received
##                                      NA
## day3_treatment1:treatment3_day_received
##                                      NA
## day3_treatment2:treatment3_day_received
##                                      NA
```

```
coeftest(m_a)
```

```
##
```

```
## t test of coefficients:
##
##                                          Estimate  Std. Error t value
## (Intercept)                            5.5554e-01  3.6426e-01  1.5251
## step_day1                              2.6910e-05  2.9564e-05  0.9102
## step_day2                              1.0032e-05  3.9445e-05  0.2543
## day1_treatment0:treatment1_day_received -8.4036e-02  2.3928e-01 -0.3512
## day1_treatment1:treatment1_day_received -3.3258e-02  1.5266e-01 -0.2179
## day1_treatment2:treatment1_day_received         NA          NA      NA
## day2_treatment0:treatment2_day_received -4.4726e-02  2.5048e-01 -0.1786
## day2_treatment1:treatment2_day_received -2.3869e-02  1.4937e-01 -0.1598
## day2_treatment2:treatment2_day_received         NA          NA      NA
## day3_treatment0:treatment3_day_received         NA          NA      NA
## day3_treatment1:treatment3_day_received         NA          NA      NA
## day3_treatment2:treatment3_day_received         NA          NA      NA
##                                         Pr(>|t|)
## (Intercept)                               0.1393
## step_day1                                 0.3711
## step_day2                                 0.8012
## day1_treatment0:treatment1_day_received   0.7283
## day1_treatment1:treatment1_day_received   0.8292
## day1_treatment2:treatment1_day_received       NA
## day2_treatment0:treatment2_day_received   0.8597
## day2_treatment1:treatment2_day_received   0.8743
## day2_treatment2:treatment2_day_received       NA
## day3_treatment0:treatment3_day_received       NA
## day3_treatment1:treatment3_day_received       NA
## day3_treatment2:treatment3_day_received       NA
```

Sequence of treatment isn't likely to impact outcome.

T-test and power calculations

```r
# get data frame in workable format
d <- rbindlist(list(d2_a[1:nrow(d2_a),c("day1_treatment","step_day1")],
                    d2_a[1:nrow(d2_a),c("day2_treatment","step_day2")],
                    d2_a[1:nrow(d2_a),c("day3_treatment","step_day3")]))
names(d) <- c("treatment","steps")
d$outcome <- ifelse(d$steps > 5000, 1, 0)

###  Control vs digital
t.test(d[treatment == 0]$outcome, d[treatment == 1]$outcome, paired = T)
```

```
##
##  Paired t-test
##
## data:  d[treatment == 0]$outcome and d[treatment == 1]$outcome
## t = -0.29732, df = 32, p-value = 0.7681
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2379101  0.1773041
## sample estimates:
## mean of the differences
##             -0.03030303
```

```r
cohensD(d[treatment == 0]$outcome, d[treatment == 1]$outcome, method = "paired")
```

```
## [1] 0.05175636
```

```r
### Control vs in person
t.test(d[treatment == 0]$outcome, d[treatment == 2]$outcome, paired = T)
```

```
##
##  Paired t-test
##
## data:  d[treatment == 0]$outcome and d[treatment == 2]$outcome
## t = 0.57143, df = 32, p-value = 0.5717
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1554323  0.2766444
## sample estimates:
## mean of the differences
##              0.06060606
```

```r
cohensD(d[treatment == 0]$outcome, d[treatment == 2]$outcome, method = "paired")
```

```
## [1] 0.09947295
```

```r
### In person vs digital
t.test(d[treatment == 2]$outcome, d[treatment == 1]$outcome, paired = T)
```

```
##
##  Paired t-test
##
## data:  d[treatment == 2]$outcome and d[treatment == 1]$outcome
## t = -0.82808, df = 32, p-value = 0.4138
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3145301  0.1327119
## sample estimates:
## mean of the differences
##             -0.09090909
```

no significant difference in both comparisons