Final Study Data Analysis

April Kim, Jennifer Podracky, Saurav Datta

```
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)
library(cobalt)
library(stringr)
library(AER)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(pander)
```

Read in data and reformat

```
assigned_treatment_seq <- data.frame(seq_id = c(1,2,3,4,5,6),
                                     day1 = c(0,0,1,1,2,2),
                                     day2 = c(1,2,0,2,0,1),
                                     day3 = c(2,1,2,0,1,0))
d2 <- fread("241 Participant List - Final Study Results - 20181215.csv", na.strings=c("","NA"))
d2[UserId == 65,]$Q10 <- "In person"
d2[UserId == 13,]$Q6 <- "Through digital means"
# stringsAsFactors = F)
names(d2) <- str_replace_all(names(d2), c(" " = "." , "," = "" ))

# Not applicable = 0
# Through digital means = 1
# In person = 2
# Both in person and through digital means = 3

d2 <- d2[, .(userId = UserId,
             treatment_seq = as.integer(Treatment.Seq),
             day1_treatment = as.integer(as.character(factor(Q6, levels = c('Not applicable', 'In perso
                                                                            'Through digital means'),
                                                             labels = c(0, 2, 1)))),
             day2_treatment = as.integer(as.character(factor(Q10, levels = c('Not applicable', 'In pers
                                                                            'Through digital means',
                                                                            'Both in person and throug
                                                             labels = c(0, 2, 1, 3)))),
             day3_treatment = as.integer(as.character(factor(Q14, levels = c('Not applicable', 'In pers
                                                                            'Through digital means',
                                                                            'Both in person and throug
                                                             labels = c(0, 2, 1, 3)))),
             day1_steps = as.numeric(gsub("\\,", "", Q7)),
             day2_steps = as.numeric(gsub("\\,", "", Q11)),
             day3_steps = as.numeric(gsub("\\,", "", Q15)),
             age_range = as.integer(as.character(factor(Age, levels = c('18 - 24',
                                                                        "25 - 34",
                                                                        "35 - 44",
                                                                        "45 - 54",
                                                                        "55 - 64",
                                                                        "65+"),
                                                         labels = c(0, 1, 2, 3, 4, 5)))),
             # gender = factor(Gender),
             gender = as.integer(as.character(factor(Gender, levels = c('Male', 'Female', 'Gender non-c
                                                     labels = c(0, 1, 2)))),
             lives_with_others = as.integer(as.character(factor(Living.Situation, levels = c('Alone', '
                                                                labels = c(0, 1)))),
             # know_us = factor(Q17),
             know_us = as.integer(as.character(factor(Q17, levels = c('No', 'Yes'),
                                                      labels = c(0, 1)))),
             location_lat = as.double(LocationLatitude),
             location_long = as.double(LocationLongitude)
)]
```

```
## Warning in eval(jsub, SDenv, parent.frame()): NAs introduced by coercion

## Warning in eval(jsub, SDenv, parent.frame()): NAs introduced by coercion
```

```
## Warning in eval(jsub, SDenv, parent.frame()): NAs introduced by coercion
```

```
d2$gender[is.na(d2$gender)] <- 2
d2$age_range[is.na(d2$age_range)] <- 6
d2$lives_with_others[is.na(d2$lives_with_others)] <- 2
d2$know_us[is.na(d2$know_us)] <- 2

head(d2, 5)
```

```
##    userId treatment_seq day1_treatment day2_treatment day3_treatment
## 1:     82             6              0              1              0
## 2:     57             3              1              0              2
## 3:     89             4             NA             NA             NA
## 4:     69             3              1              0              2
## 5:     85             3              1              0              2
##    day1_steps day2_steps day3_steps age_range gender lives_with_others
## 1:         NA       5040       3788         1      0                 1
## 2:      21290      13959      13717         0      0                 1
## 3:         NA         NA         NA         1      0                 1
## 4:       6343       3247      10198         1      0                 1
## 5:      13624       5406       7851         1      1                 1
##    know_us location_lat location_long
## 1:       1     41.89250      -87.7895
## 2:       1     37.75101      -97.8220
## 3:       1     37.97240     -122.3369
## 4:       1     40.37070      -74.0084
## 5:       1     42.41730      -71.1087
```

```
#Covariate Balance Check
bal.tab(treatment_seq ~ gender + age_range + lives_with_others + know_us + location_lat + location_long
        data = d2)
```

```
## Balance Measures
##                      Type Corr.Un
## gender            Contin. -0.0625
## age_range         Contin. -0.0099
## lives_with_others Contin. -0.0105
## know_us           Contin.  0.0214
## location_lat      Contin.  0.0157
## location_long     Contin. -0.0480
##
## Sample sizes
##     Total
## All    75
```

```
cov_check <- lm(treatment_seq ~ gender + age_range + lives_with_others + know_us + location_lat + locati
                data = d2)
summary(cov_check)
```

```
##
## Call:
```

```
## lm(formula = treatment_seq ~ gender + age_range + lives_with_others +
##     know_us + location_lat + location_long, data = d2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7912 -1.4863  0.1883  1.4598  2.6363
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.1350717  4.1690323   0.272    0.786
## gender            -0.1639576  0.3999806  -0.410    0.683
## age_range         -0.0003737  0.1750913  -0.002    0.998
## lives_with_others -0.1171429  0.7099934  -0.165    0.869
## know_us            0.0212714  0.3840969   0.055    0.956
## location_lat       0.0465410  0.0841290   0.553    0.582
## location_long     -0.0082195  0.0124256  -0.661    0.511
##
## Residual standard error: 1.784 on 68 degrees of freedom
## Multiple R-squared:  0.01086,    Adjusted R-squared:  -0.07641
## F-statistic: 0.1245 on 6 and 68 DF,  p-value: 0.993
```

Checking for ordering/priming effect

Is previous day's treatment highly predictive of how many steps are taken today?

```
# n = 75
df <- d2

# remove subjects/rows who were non-compliant
# n = 24
df <- df[rowSums(is.na(df[,c(3:8)])) != ncol(df[,c(3:8)]), ]

head(df, 5)
```

```
##    userId treatment_seq day1_treatment day2_treatment day3_treatment
## 1:     82             6              0              1              0
## 2:     57             3              1              0              2
## 3:     69             3              1              0              2
## 4:     85             3              1              0              2
## 5:     66             4              1              2              0
##    day1_steps day2_steps day3_steps age_range gender lives_with_others
## 1:         NA       5040       3788         1      0                 1
## 2:      21290      13959      13717         0      0                 1
## 3:       6343       3247      10198         1      0                 1
## 4:      13624       5406       7851         1      1                 1
## 5:       7016       1211       5717         0      0                 1
##    know_us location_lat location_long
## 1:       1     41.89250      -87.7895
## 2:       1     37.75101      -97.8220
## 3:       1     40.37070      -74.0084
## 4:       1     42.41730      -71.1087
## 5:       1     42.35760      -71.0514
```

```r
# day 3 steps using day 1 and 2 treatment
m1 <- lm(day3_steps ~ day1_treatment + day2_treatment, df)
summary(m1)
```

```
##
## Call:
## lm(formula = day3_steps ~ day1_treatment + day2_treatment, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7119.9 -2316.1  -462.4  1377.2  8580.4
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7505.7      921.5   8.145 2.12e-10 ***
## day1_treatment   -385.7      672.7  -0.573    0.569
## day2_treatment   -365.5      600.8  -0.608    0.546
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3440 on 45 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.0142, Adjusted R-squared:  -0.02961
## F-statistic: 0.3241 on 2 and 45 DF,  p-value: 0.7248
```

```r
# ATE (standard error)
print(paste0("Estimated effect of day1 treatment: ", signif(m1$coefficients[2], 3),
" (", signif(coef(summary(m1))[2,2], 3), ")"))
```

```
## [1] "Estimated effect of day1 treatment: -386 (673)"
```

```r
print(paste0("Estimated effect of day2 treatment: ", signif(m1$coefficients[3], 3),
" (", signif(coef(summary(m1))[3,2], 3), ")"))
```

```
## [1] "Estimated effect of day2 treatment: -366 (601)"
```

```r
# include days1,2 steps as covariates to understand
# subjects' step counts have as a function of
# treatment against waht they would typically do
m2 <- lm(day3_steps ~ day1_treatment + day2_treatment + day1_steps + day2_steps, df)
summary(m2)
```

```
##
## Call:
## lm(formula = day3_steps ~ day1_treatment + day2_treatment + day1_steps +
##     day2_steps, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8183.6 -1375.5    61.2  1536.6  5057.0
##
```

```
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2602.4071  1128.7513   2.306  0.02614 *
## day1_treatment -484.8396   517.7869  -0.936  0.35444
## day2_treatment -296.7856   460.9784  -0.644  0.52319
## day1_steps        0.2442     0.1230   1.985  0.05373 .
## day2_steps        0.4542     0.1341   3.386  0.00155 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2623 on 42 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.455,  Adjusted R-squared:  0.4031
## F-statistic: 8.766 on 4 and 42 DF,  p-value: 3.076e-05
```

```
print(paste0("Estimated effect of day1 treatment: ", signif(m2$coefficients[2], 3),
             " (", signif(coef(summary(m2))[2,2], 3), ")"))
```

```
## [1] "Estimated effect of day1 treatment: -485 (518)"
```

```
print(paste0("Estimated effect of day2 treatment: ", signif(m2$coefficients[3], 3),
             " (", signif(coef(summary(m2))[3,2], 3), ")"))
```

```
## [1] "Estimated effect of day2 treatment: -297 (461)"
```

We do not see that the previous days' treatment assignments to predict the last day's step count is highgly predicitive and significant, which is super for us!

```
stargazer(m1, m2,
          dep.var.labels=c("Steps - Day 3"),
          covariate.labels=c("Treatment - Day 1", "Treatment - Day 2", "Steps - Day 1", "Steps - Day 2"
          omit.stat=c("all"))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 20, 2018 - 16:26:40

Condense treatment sequence to 1 treatment

```
df1 <- df[,-c(4,5,7,8)]
df2 <- df[,-c(3,5,6,8)]
df3 <- df[,-c(3,4,6,7)]
names(df1)[names(df1) == "day1_treatment"] = "treatment"
names(df1)[names(df1) == "day1_steps"] = "steps"
names(df2)[names(df2) == "day2_treatment"] = "treatment"
names(df2)[names(df2) == "day2_steps"] = "steps"
names(df3)[names(df3) == "day3_treatment"] = "treatment"
names(df3)[names(df3) == "day3_steps"] = "steps"
d <- rbind(df1, df2, df3)
# combine digital and in person treatment as one
d$treatment2 <- ifelse(d$treatment == 0, 0, 1)
```

Table 1:

	Dependent variable:	
	Steps - Day 3	
	(1)	(2)
Treatment - Day 1	−385.744	−484.840
	(672.663)	(517.787)
Treatment - Day 2	−365.548	−296.786
	(600.840)	(460.978)
Steps - Day 1		0.244*
		(0.123)
Steps - Day 2		0.454***
		(0.134)
Constant	7,505.694***	2,602.407**
	(921.528)	(1,128.751)
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
d$outcome <- ifelse(d$steps > 5000, 1, 0)

head(d, 5)
```

```
##    userId treatment_seq treatment steps age_range gender lives_with_others
## 1:     82             6         0    NA         1      0                 1
## 2:     57             3         1 21290         0      0                 1
## 3:     69             3         1  6343         1      0                 1
## 4:     85             3         1 13624         1      1                 1
## 5:     66             4         1  7016         0      0                 1
##    know_us location_lat location_long treatment2 outcome
## 1:       1     41.89250      -87.7895          0      NA
## 2:       1     37.75101      -97.8220          1       1
## 3:       1     40.37070      -74.0084          1       1
## 4:       1     42.41730      -71.1087          1       1
## 5:       1     42.35760      -71.0514          1       1
```

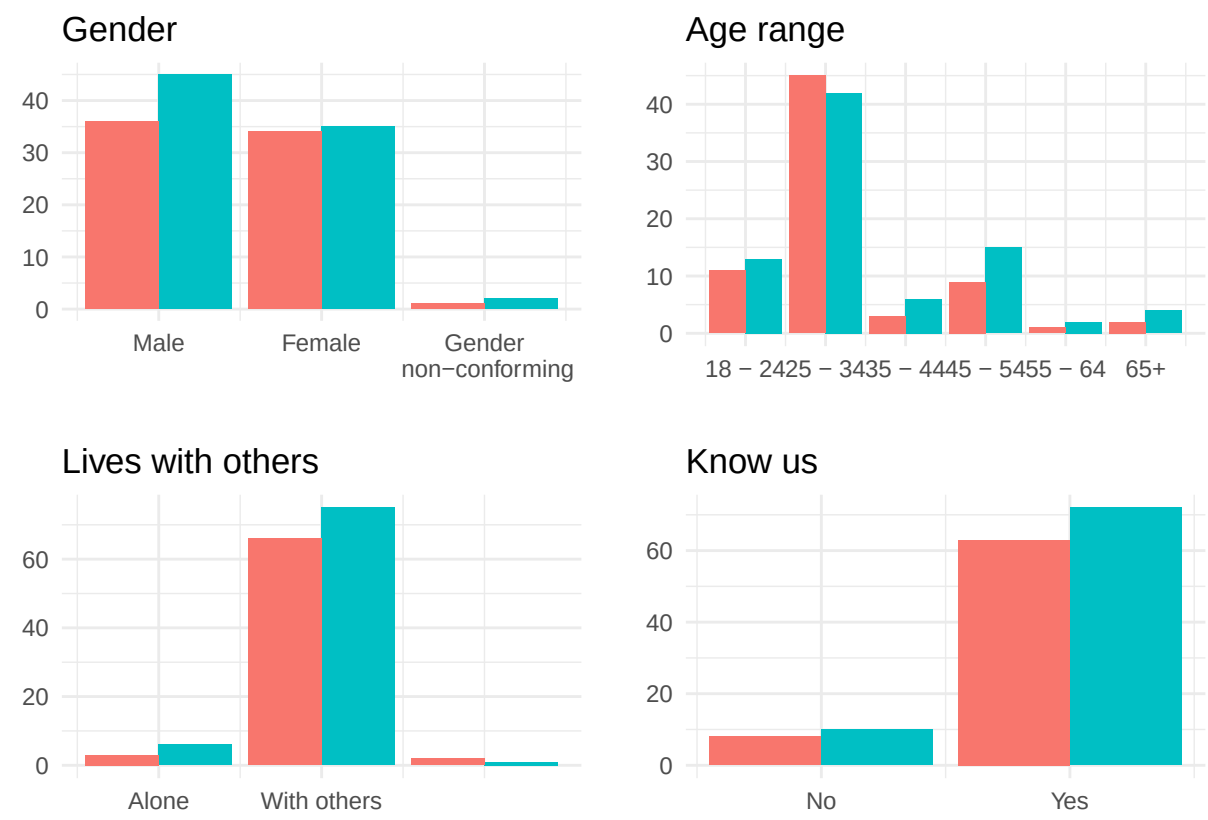
Make some pretty plots to show distribution, populatin etc.

```
# population that actually responded to data collection survey
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```r
d.gender <- d[, c("gender", "treatment2")]
p_gender <- ggplot(d.gender, aes(x=gender, fill = factor(treatment2))) +
  geom_bar(stat="count", position=position_dodge()) +
  theme_minimal() + theme(legend.position="right") +
  xlab("") + ylab("") + ggtitle("Gender") +
  guides(fill = guide_legend(title = "Assignment")) +
  scale_fill_discrete(labels = c("Control", "Treatment")) +
  scale_x_continuous(breaks = c(0, 1, 2),
                     labels = c('Male', 'Female', 'Gender\n non-conforming'))

p_gender_no_legend <- ggplot(d.gender, aes(x=gender, fill = factor(treatment2))) +
  geom_bar(stat="count", position=position_dodge()) +
  theme_minimal() + theme(legend.position="none") +
  xlab("") + ylab("") + ggtitle("Gender") +
  # guides(fill = guide_legend(title = "Assignment")) +
  # scale_fill_discrete(labels = c("Control", "Treatment")) +
  scale_x_continuous(breaks = c(0, 1, 2),
                     labels = c('Male', 'Female', 'Gender\n non-conforming'))


d.age <- d[, c("age_range", "treatment2")]
p_age <- ggplot(d.age, aes(x=age_range, fill = factor(treatment2))) +
  geom_bar(stat="count", position=position_dodge()) +
  theme_minimal() + theme(legend.position="none") +
  xlab("") + ylab("") + ggtitle("Age range") +
  # guides(fill = guide_legend(title = "Assignment")) +
  # scale_fill_discrete(labels = c("Control", "Treatment")) +
  scale_x_continuous(breaks = c(0, 1, 2, 3, 4, 5),
                   labels = c('18 - 24',
                              "25 - 34",
                              "35 - 44",
                              "45 - 54",
                              "55 - 64",
                              "65+"))

d.others <- d[, c("lives_with_others", "treatment2")]
p_others <- ggplot(d.others, aes(x=lives_with_others, fill = factor(treatment2))) +
  geom_bar(stat="count", position=position_dodge()) +
  theme_minimal() + theme(legend.position="none") +
  xlab("") + ylab("") + ggtitle("Lives with others") +
  # guides(fill = guide_legend(title = "Assignment")) +
  # scale_fill_discrete(labels = c("Control", "Treatment")) +
  scale_x_continuous(breaks = c(0, 1),
                     labels = c('Alone', 'With others'))

d.know_us <- d[, c("know_us", "treatment2")]
p_know_us <- ggplot(d.know_us, aes(x=know_us, fill = factor(treatment2))) +
  geom_bar(stat="count", position=position_dodge()) +
  theme_minimal() + theme(legend.position="none") +
  xlab("") + ylab("") + ggtitle("Know us") +
  # guides(fill = guide_legend(title = "Assignment")) +
  # scale_fill_discrete(labels = c("Control", "Treatment")) +
  scale_x_continuous(breaks = c(0, 1),
```

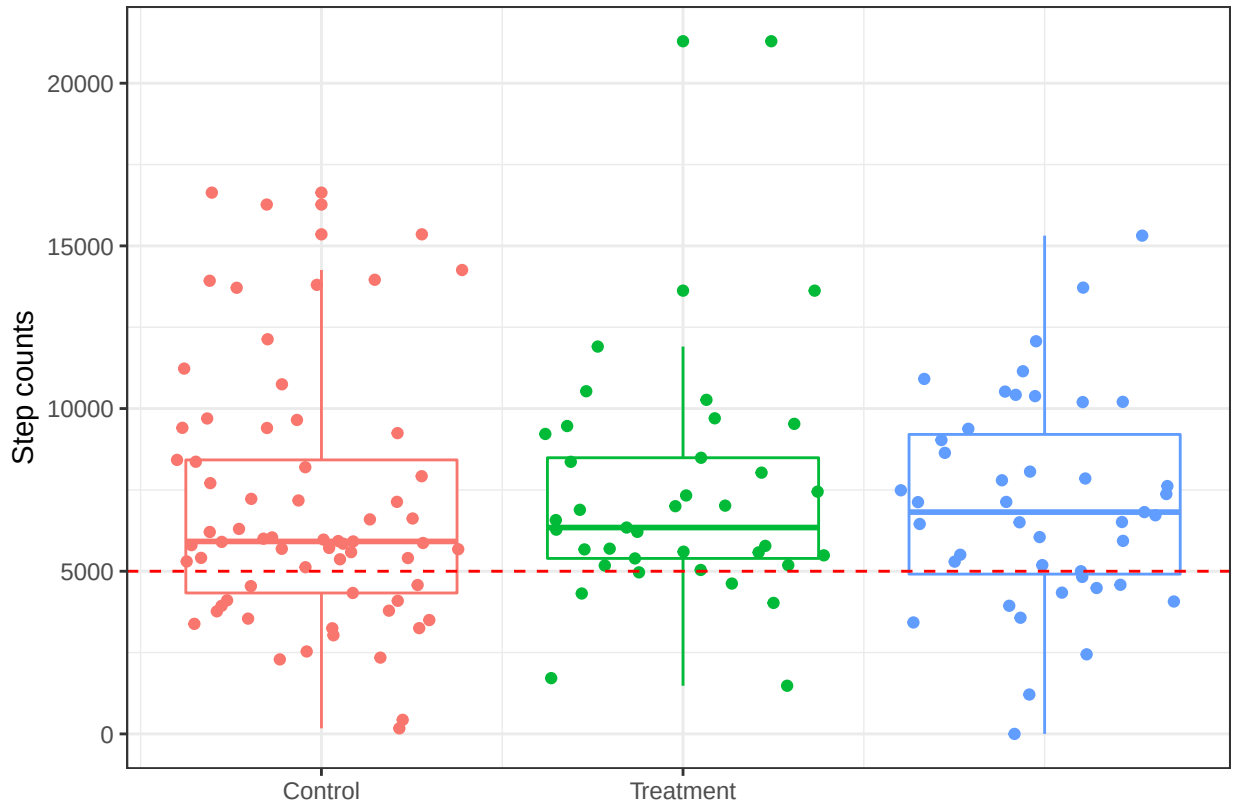
```
                    labels = c('No', 'Yes'))

# p_gender
grid.arrange(p_gender_no_legend, p_age, p_others, p_know_us,
             ncol = 2)
```




```
# control and digital and in person distribution
ggplot(d, aes(x=treatment, y=steps, colour = factor(treatment))) +
geom_boxplot() + geom_jitter() +
geom_hline(yintercept=5000, linetype="dashed", color = "red") +
xlab("") + ylab("Step counts") + theme_bw() +
    scale_x_continuous(breaks = c(0, 1),
                       labels = c('Control', 'Treatment')) +
  theme(legend.position="none")
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```

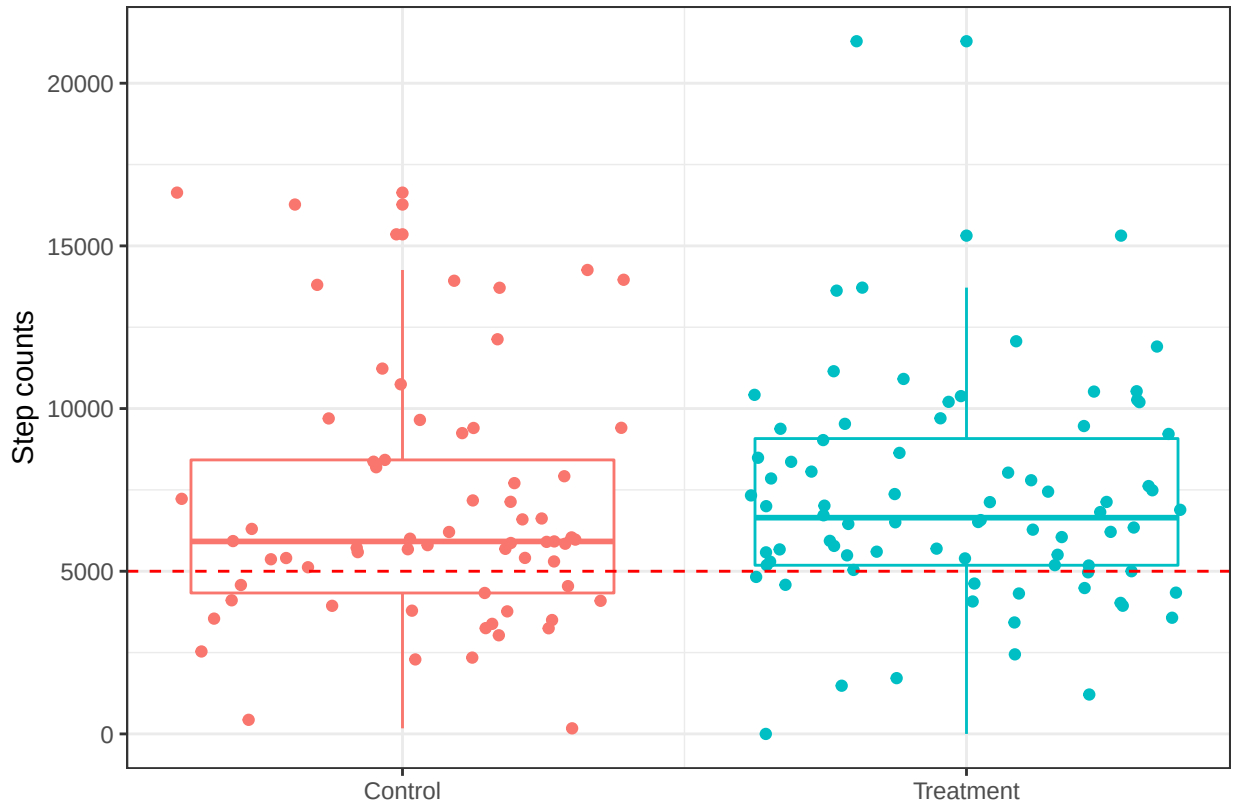
```
## Warning: Removed 8 rows containing missing values (geom_point).
```




```
# control and treatment (digital+in person) when time component removed
ggplot(d, aes(x=treatment2, y=steps, colour = factor(treatment2))) +
geom_boxplot() + geom_jitter() +
geom_hline(yintercept=5000, linetype="dashed", color = "red") +
xlab("") + ylab("Step counts") + theme_bw() +
    scale_x_continuous(breaks = c(0, 1),
                       labels = c('Control', 'Treatment')) +
  theme(legend.position="none")
```

```
## Warning: Removed 8 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 8 rows containing missing values (geom_point).
```



For control vs digital and control vs in person

```
d$treatment <- factor(d$treatment)
fit_3 <- lm(outcome ~ treatment + userId , d)
# robust se
se_3 <- sqrt(diag(vcovHC(fit_3, type = 'HC')))

# remove intercept
fit_3.1 <- lm(outcome ~ treatment + userId -1, d)
# robust se
se_3.1 <- sqrt(diag(vcovHC(fit_3, type = 'HC')))

fit_3_covariates <- lm(outcome ~ treatment + age_range + gender + lives_with_others + know_us + locatio
# robust se
se_3_covariates <- sqrt(diag(vcovHC(fit_3_covariates, type = 'HC')))

# ATE (standard error)
print(paste0("Estimated effect of treatment (control, in person, digital): ", signif(fit_3$coefficients
" (", signif(se_3[2], 3), ")"))
```

```
## [1] "Estimated effect of treatment (control, in person, digital): 0.125 (0.0819)"
```

```
print(paste0("Estimated effect of treatment (control, in person, digital) + covariates: ", signif(fit_3,
" (", signif(se_3_covariates[2], 3), ")"))
```

```
## [1] "Estimated effect of treatment (control, in person, digital) + covariates: 0.0327 (0.0845)"
```

```
stargazer(fit_3.1,
          se=list(se_3),
          omit = c("treatment0"),
          dep.var.labels=c("Steps > 5000"),
          covariate.labels=c('Commit in person', 'Commit digitally', "User ID"),
          omit.stat=c("all"))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 20, 2018 - 16:26:42

Table 2:

	Dependent variable:
	Steps > 5000
Commit in person	0.744***
	(0.082)
Commit digitally	0.629***
	(0.088)
User ID	0.002
	(0.001)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
stargazer(fit_3, fit_3_covariates,
          se=list(se_3, se_3_covariates),
          dep.var.labels=c("Steps > 5000"),
          covariate.labels=c("Treatment", "User ID", "Age range", "Gender", "Has housemate", "Knows us"
          omit.stat=c("all"))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 20, 2018 - 16:26:42

test hypothesis that telling others make it more likely to take >5000 steps (control vs treatment)

```
#suppress intercept term
fit_2 <- lm(outcome ~ treatment2 + userId -1, d)
# robust se
se_2 <- sqrt(diag(vcovHC(fit_2, type = 'HC')))
```

Table 3:

	Dependent variable:	
	Steps > 5000	
	(1)	(2)
Treatment	0.125	0.129
	(0.082)	(0.085)
User ID	0.010	0.033
	(0.088)	(0.087)
Age range	0.002	
	(0.001)	
Gender		−0.017
		(0.036)
Has housemate		−0.048
		(0.077)
Knows us		0.200
		(0.154)
Latitute		0.122
		(0.134)
Longitude		−0.002
		(0.014)
location_long		0.001
		(0.003)
Constant	0.619***	0.593
	(0.078)	(0.800)
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
fit_2_covariates <- lm(outcome ~ treatment2 + age_range + gender + lives_with_others + know_us + locati
# robust se
se_2_covariates <- sqrt(diag(vcovHC(fit_2_covariates, type = 'HC')))

# ATE (standard error)
print(paste0("Estimated effect of treatment (control, treatment): ", signif(fit_2$coefficients[2], 3),
" (", signif(se_2[2], 3), ")"))
```

```
## [1] "Estimated effect of treatment (control, treatment): 0.00932 (0.000987)"
```

```
print(paste0("Estimated effect of treatment (control, treatment) + covariates: ", signif(fit_2_covariat
" (", signif(se_2_covariates[2], 3), ")"))
```

```
## [1] "Estimated effect of treatment (control, treatment) + covariates: 0.0776 (0.0735)"
```

```
stargazer(fit_2,
          se=list(se_2),
          dep.var.labels=c("Steps > 5000"),
          covariate.labels=c("Social commitment", "User ID"),
          omit.stat=c("all"))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 20, 2018 - 16:26:42

Table 4:

	Dependent variable:
	Steps > 5000
Social commitment	0.353***
	(0.067)
User ID	0.009***
	(0.001)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
stargazer(fit_2, fit_2_covariates,
          se=list(se_2, se_2_covariates),
          dep.var.labels=c("Steps > 5000"),
          covariate.labels=c("Treatment", "User ID", "Age range", "Gender", "Has housemate", "Knows us"
          omit.stat=c("all"))
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 20, 2018 - 16:26:42

power calculations

Table 5:

	Dependent variable:	
	Steps > 5000	
	(1)	(2)
Treatment	0.353***	0.078
	(0.067)	(0.073)
User ID	0.009***	
	(0.001)	
Age range		−0.019
		(0.037)
Gender		−0.046
		(0.077)
Has housemate		0.200
		(0.149)
Knows us		0.131
		(0.135)
Latitute		−0.002
		(0.014)
Longitude		0.001
		(0.003)
Constant		0.598
		(0.805)
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
###  Control vs treatment (digital+in person)
# since we fail to reject the null hypothesis,
# let's calculate number of subjects needed for 80% power
effect_size_outcome <- cohensD(d[treatment2 == 0]$outcome, d[treatment2 == 1]$outcome)
#power we got from our experiment
pwr.t2n.test(n1 = 71, n2 = 82, d = effect_size_outcome, sig.level = 0.05)
```

```
##
##      t test power calculation
##
##              n1 = 71
##              n2 = 82
##               d = 0.1537758
##       sig.level = 0.05
##           power = 0.1563328
##     alternative = two.sided
```

```
# 80% powered test
pwr.t.test(power = 0.8, d = effect_size_outcome, sig.level = 0.05, type = "two.sample")
```

```
##
##      Two-sample t test power calculation
##
##               n = 664.7977
##               d = 0.1537758
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
### extra plots
# day1
pd1 <- ggplot(df, aes(x=day1_treatment, y=day1_steps, colour = factor(day1_treatment))) +
  geom_boxplot() + geom_jitter() +
  geom_hline(yintercept=5000, linetype="dashed", color = "red") +
  xlab("") + ylab("Step counts") + theme_bw() +
  scale_x_continuous(breaks = c(0, 1, 2),
                     labels = c(0, 1, 2)) +
  # labels = c('Control', 'In person', 'Through digital means')) +
  theme(legend.position="none") + ggtitle("Step count - day 1")
# day2
pd2 <- ggplot(df, aes(x=day2_treatment, y=day2_steps, colour = factor(day2_treatment))) +
geom_boxplot() + geom_jitter() +
geom_hline(yintercept=5000, linetype="dashed", color = "red") +
xlab("") + ylab("Step counts") + theme_bw() +
    scale_x_continuous(breaks = c(0, 1, 2),
                       labels = c(0, 1, 2)) +
    #                   labels = c('Control', 'In person', 'Through digital means')) +
  theme(legend.position="none") + ggtitle("Step count - day 2")
# day3
pd3 <- ggplot(df, aes(x=day3_treatment, y=day3_steps, colour = factor(day3_treatment))) +
```

```
geom_boxplot() + geom_jitter() +
geom_hline(yintercept=5000, linetype="dashed", color = "red") +
xlab("") + ylab("Step counts") + theme_bw() +
    scale_x_continuous(breaks = c(0, 1, 2),
                       labels = c(0, 1, 2)) +
    #                   labels = c('Control', 'In person', 'Through digital means')) +
  theme(legend.position="none") + ggtitle("Step count - day 3")
```