

# TV Watching Habits Study

*Cameron Bell, Alyssa Eisenberg, Sarah Cha*

*4/6/2018*

## Abstract

Television is consumed through a growing number of channels today (cable, streaming networks like Hulu, Amazon, Netflix, and Youtube) while increasingly the mould for what looks like television is becoming more fluid (e.g. video clips on Youtube). We suspect some individuals don't actually realize how much television they're watching on a daily basis and we set out to understand how social pressures around television watching would alter or impact individual behavior. In our study, we randomly assigned treatment emails to half of the participants that highlights how much their peers are watching TV on average. XXX

Nielsen Total Audience report for the fourth quarter of 2016 discussed the steady trend of decline in live TV watching alongside increased use of smartphones to view video and the increasing penetration of subscription video on demand. While live viewing decreased to 4 hours and 23 minutes a day among adults 18 years and up (from 4 hours and 27 minutes in the prior year), smartphone usage rose to 2 hours and 32 minutes a day from 1 hour and 15 minutes a year ago. Needless to say, individuals are consuming video content in more formats today and given the rate at which online and on-demand content is growing, consumers can access endless amount content right at their fingertips. This points to a trend of increased video/screen watching and could mean that consumers are watching more TV than they actually realize. As a result, we're curious to understand if we could change TV watching habits if we focused the attention on how many hours people are watching vs. their peers.

We set up an experimental study where we randomly assigned treatment emails to half of the participants that summarizes how much their peers are watching television on average. We wanted to understand if social/peer pressure has the ability to impact how much television a person watches (e.g. "Wow, my peers are watching only 2 hours of TV a week and I'm watching at least 10. Shame on me."). We begin by discussing the experimental design we developed to test whether social pressures/influence can actually impact individual behavior as it relates to TV watching. Next, we discuss audience for our experiment, the nature and frequency of the treatment, and our outcome measures. Statistical analysis from our study suggests that XXXXXXX taking into account non-compliance and attrition aspects of the experiment. We conclude by discussing the implications for generalizability to a broader population at large and future potential avenues for exploration.

## Experimental Design

Our aim was to develop a means for studying the influence of social pressures on TV watching, recognizing that the means by which Americans consume TV is very diverse today (cable, online video streaming, YouTube, etc) and what people classify as television can vary quite a bit (e.g. television shows, news, sports, movies, YouTube clips, etc). We hypothesized that when we carefully defined what we classify as "television" watching and participants could compare apples to apples their consumption vs. peers, they'd be impacted if their peers were watching less than they were. We also hypothesized that we'd see heterogeneous treatment effect from our participants based on amount of TV watched prior to experiment (ie. those in 1st quartile not likely to reduce amount of TV and may actually increase instead).

We decided on surveys as the means by which we would collect all the necessary information including outcome measures from our participants. Our experiment consisted of two surveys - one pre and one post treatment - where we would measure pre-survey TV behavior for all participants along with what we thought could be key covariates of interest (age, gender, marital/employment status, etc) and then we'd randomly assign individuals to treatment vs. control where the individuals in the treatment group would be aware of how much their peers were watching. Then for the actual calculation of treatment effect using the post-survey results we did a difference in difference comparison where we compare the hours of TV watched between

treatment and control group for causal estimate conditioning on pre-experiment TV watching measure as a covariate.

In designing our study, we were very mindful of how to maximize participation to get a large enough sample size, get accurate reportings, and how we could extract the most information in the least intrusive way. On the latter, conversations with our peers made us realize that there was clearly a balance between incentives for participants and the time commitment involved. We offered 5 Amazon gift cards for \$20 each as a reward for participating, with the stipulation that participants had to complete a 3 minute pre-survey and a 1 minute post-survey in order to be entered into the drawing.

## Pilot Study

From that we had a good framework to start with and we implemented a pilot study with 9 individuals to start that would help identify what potential problems with the design could be and also serve as a good reference point for a power calculation that would help us understand how many subjects we needed in order to draw an actual treatment effect if there was one. The pilot took place over the course of 1 week (Feb 26 - Mar 4) where friends and family took answered a pre-survey (see Appendix for full questionnaire), were randomly assigned to control or treatment, treatment group was sent an email about peer usage, and then everyone filled out a post-survey on their TV consumption.

### Some highlights of what we learned from the pilot:

1. Attrition is likely to be high. Of the 9 close family and friends that we targeted for the survey, we needed up with full observation sets for only 7 of 9. Of course the short observation time had an impact in the case of the pilot.
2. At least for our pilot study, the quality of the data coming out of the Chrome extension reading was low - 3 of 7 individuals had 0 or NA output even though they indicated they watched TV during that time.
3. We emailed the pre- and post-surveys out via Qualtrics with the sender name as “UCBerkeley TV Habits Study” but the treatment email came from Alyssa’s gmail, raising the question if this created confusion and/or potential for individuals to ignore Alyssa’s emails (non-compliance).
4. We ran a power calculation from the results of the pilot survey. From the simulation we ran, we would need a treatment effect of around -4 hours/week to find a statistically significant result 80% of the time with 80 observations. If we assume pre-experiment TV watching is related to post-experiment TV watching, using this covariate would reduce our standard error and could thus help our power calculation - 80% of the time we would find a statistically significant result for about half the treatment effect (-2 hours/week).

*Subjects* In an ideal scenario, we would survey a large sample of the US adult population in order to get a representative sample for our experiment but with the key constraint being time and resources (we have only about 10 weeks and \$500), we implemented a much more limited study where we, ourselves, were tasked with finding our respondents by word of mouth and social media but we leverage the power calculation give us a strong confidence that if there were a statistically significant result, we’d be able to detect it with the size of the data ( $n$ = number of participants) and the potential effect size.

We canvassed coffee shops and college campuses and put up flyers on social media to get participants. From our power calculation we knew we wanted around 80-100 individuals when it was all said and done but also knew we’d get a high level of attrition between the time we recruited the subjects to the time that the post-surveys were due. Even though we used different methods of outreach including word of mouth, social media, and email, we tried to put together a standardized pitch with key talking points to ensure that all three of us were giving out the same details about the experiment and that more importantly none of us in the process to giving too much detail on what we were aiming to do/analyze to avoid participants skewing their answers. Here is the standardized pitch:

```
##fix!!
#png(file="pitch.png",width=400,height=350)
```

Initially we took down email addresses of all interested participants and then those participants were emailed a pre-survey that they had to fill out to be a part of the actual experiment. Just as with the pilot, the pre-survey asked all participants about their initial TV watching habits and covariates at the beginning of the experiment. Useful potential covariates included age, marital status, and pre-existing TV watching habits.

#### *Random assignment of treatment and controls*

In terms of randomization once we had our set list of participants from the pre-survey, we used a simple random assignment to get half of our subjects into control and the other half in treatment in an effort to prevent against any selection bias. For a check, we did a covariate balance check and found that the randomization had worked - there were no statistically significant differences between the control and treatment group nor did the F-test of joint significance return anything of significance.

*Outcome Measures* We carefully define in our communication with participants what we consider to be TV consumption (cable, streaming online, Youtube videos, sports, movies) and ask survey respondents to estimate and record how much they watch. Another potential option was to create a Chrome Netflix extension that would track all Netflix usage an individual had if they installed it and then the individual could input the numbers directly in the survey at the end of the survey period. We anticipated that attrition could be a potential problem with the set-up the survey given with pre-survey and treatment emails respondents could probably take a pretty good guess at what outcome measure we were most interested in while individuals at the core of what we want to study (ie. individuals who watch a lot of TV and whether or not they are influenced by treatment emails) could become most discouraged from continuing to participant through to the end. In an effort to entice individuals to be willing participants in our survey, we offered five \$20 Amazon gift cards with the stipulation that both the pre and post survey needed to be filled out and completed on-time in order to be entered into the drawing.

## Calculating Power

### *Base Case*

First, create the base case where we are running a simple t-test on difference in means between treatment and control, and the treatment effect is constant for everyone. Use the mean and variance from the pilot pre-survey for self-reported TV since the Netflix data was minimal. Since the pilot was so short, we did not get a good estimate of the treatment effect. Instead, run across a variety of treatment effects and check the power value.

```
individual_simulation <- function(
  units, mean_control, sd_control, tau) {
  ## units is our sample size
  ## mean_control is the mean TV watched with no treatment
  ## sd_control is the standard deviation in TV watched with no treatment
  ## tau is the treatment effect (change in amount of TV for treatment)

  ## Assume simple randomization among all units with 50% probability of treatment
  urn <- c('treat', 'control')
  d <- data.table(id = 1:units)
  d[, condition := sample(urn, size = .N, replace = TRUE)]

  ## Assumes a normal distribution of potential outcome to control
  ## Mean and SD are supplied to function
  ## Assumes a constant treatment effect across all participants
  d[, Y0 := pmax(rep(0, units), rnorm(units, mean=mean_control, sd = sd_control))]
  d[, Y1 := pmax(rep(0, units), Y0 + tau)]
  d[condition == 'control' , outcome := Y0]
```

```

d[condition == 'treat' , outcome := Y1]

res <- list(
  'pvalue' = d[ , t.test(outcome ~ condition)$p.value],
  'tau' = tau,
  'baseline_mean' = d[condition == 'control', mean(outcome)],
  'baseline_n' = d[condition == 'control', .N],
  'alt_mean' = d[condition == 'treat', mean(outcome)],
  'alt_n' = d[condition == 'treat', .N],
  'df_n' = d[ , .N],
  'ate' = d[condition == 'treat', mean(outcome)] - d[condition == 'control', mean(outcome)]
)

return(res)
}

```

If you want to, simulate with particular input values.

```

# Simulate experiment many times for particular input values
# calculate % of cases where results are statistically significant at 5% level
results <- replicate(1000, individual_simulation(80, 8.6, 6.25, -1), simplify = FALSE)
results <- rbindlist(results)
power <- results[ , mean(pvalue < 0.05)]
power

```

```
## [1] 0.103
```

From the simulation below, we see that we would need a treatment effect of around -4.0 hrs/week in order to find a statistically significant result in 80% of cases.

```

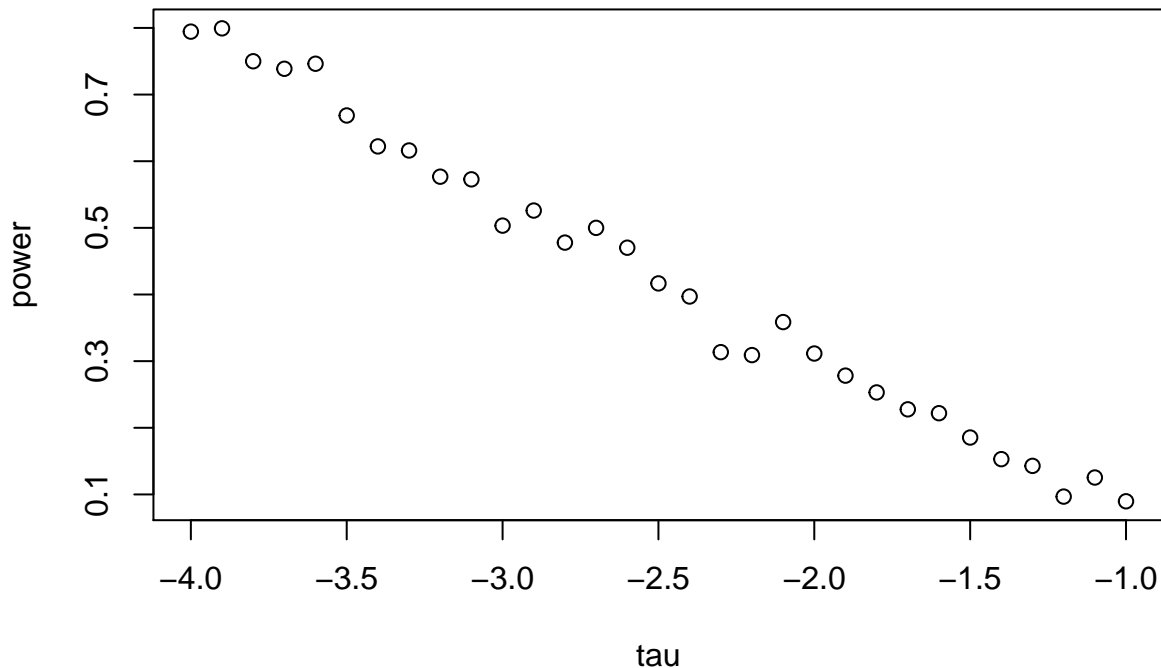
# Allow treatment effect to vary and run many simulated experiments
# Note this takes several minutes to run
moving_tau <- seq(from=-1, to= -4, by=-.1)

results <- replicate(10000,
  individual_simulation(80, 8.6, 6.25, sample(moving_tau, size = 1)),
  simplify = FALSE)
results <- rbindlist(results)
results$sig <- as.numeric(results$pvalue<0.05)

# Calculate power at each treatment effect level and plot them
results_agg <- aggregate(sig ~ tau, FUN=mean, data=results)
plot(sig ~ tau, data=results_agg,
  ylab="power",
  main = "Power by Treatment Effect Size")

```

## Power by Treatment Effect Size



### *Covariate Case*

Now, we are assuming that pre-experiment average time watching TV will be highly correlated with post-experiment weekly time spent watching TV, our outcome. Using this as a covariate should reduce our SE and provide more power, so we will run a t-test on a regression including this covariate. We will make the same assumptions for our pre-experiment values as we made in the base case for our potential outcomes to control. Then, we will assume that each individuals potential outcomes to control are normally distributed with mean of that individuals TV prior to experiment, and a SD value provided as an input. We will test against several values of this SD since we have no data for it.

```
individual_simulation_covariate <- function(
  units, mean_control, sd_control, tau, sd_individual) {
  ## units is our sample size
  ## mean_control is the mean TV watched prior to experiment
  ## sd_control is the standard deviation in TV watched prior to experiment
  ## tau is the treatment effect (change in amount of TV for treatment)
  ## sd_individual is the SD for an individuals weekly TV watched with no treatment

  ## Assume simple randomization among all units with 50% probability of treatment
  urn <- c('treat', 'control')
  d <- data.table(id = 1:units)
  d[, condition := sample(urn, size = .N, replace = TRUE)]
  d[, treated := as.numeric(condition == "treat")]

  ## Assumes a normal distribution of pre-experiment avg weekly TV
  ## Mean and SD are supplied to function
  ## Assumes each individual has a normal distribution centered on their pre-experiment value
  ## Assumes a constant treatment effect across all participants
  d[, preTV := pmax(rep(0, units), rnorm(units, mean=mean_control, sd = sd_control))]
  d[, Y0 := pmax(rep(0, units), rnorm(units, mean=preTV, sd = sd_individual))]
  d[, Y1 := pmax(rep(0, units), Y0 + tau)]
}
```

```

d[condition == 'control' , outcome := Y0]
d[condition == 'treat' , outcome := Y1]

## Run a regression including preTV as a covariate
model <- lm(outcome ~ treated + preTV, data=d)

res <- list(
  'pvalue' = coeftest(model)[2,4],
  'tau' = tau,
  'sd_individual' = sd_individual,
  'baseline_n' = d[condition == 'control', .N],
  'alt_n' = d[condition == 'treat', .N],
  'df_n' = d[, .N],
  'ate' = unname(model$coefficients[2])
)

return(res)
}

```

If you want to, simulate with particular input values.

```

# Simulate experiment many times for particular input values
# calculate % of cases where results are statistically significant at 5% level
results <- replicate(1000, individual_simulation_covariate(80, 8.6, 6.25, -1, 1), simplify = FALSE)
results <- rbindlist(results)
power <- results[, mean(pvalue < 0.05)]
power

```

```
## [1] 0.982
```

From the simulation below, we can see the combinations of individual standard deviation in amount of TV watched per week and treatment effect in order to find a statistically significant result in 80% of cases. For example, if the individual SD was about half that of the overall population (3), we would need a treatment effect of -2 hrs/week to have 80% power.

```

# Allow treatment effect and sd_individual to vary and run many simulated experiments
# Note this takes over ten minutes to run
moving_tau <- seq(from=-1, to= -4, by=-.1)
moving_sd_individual <- seq(from=1, 5, by=0.2)

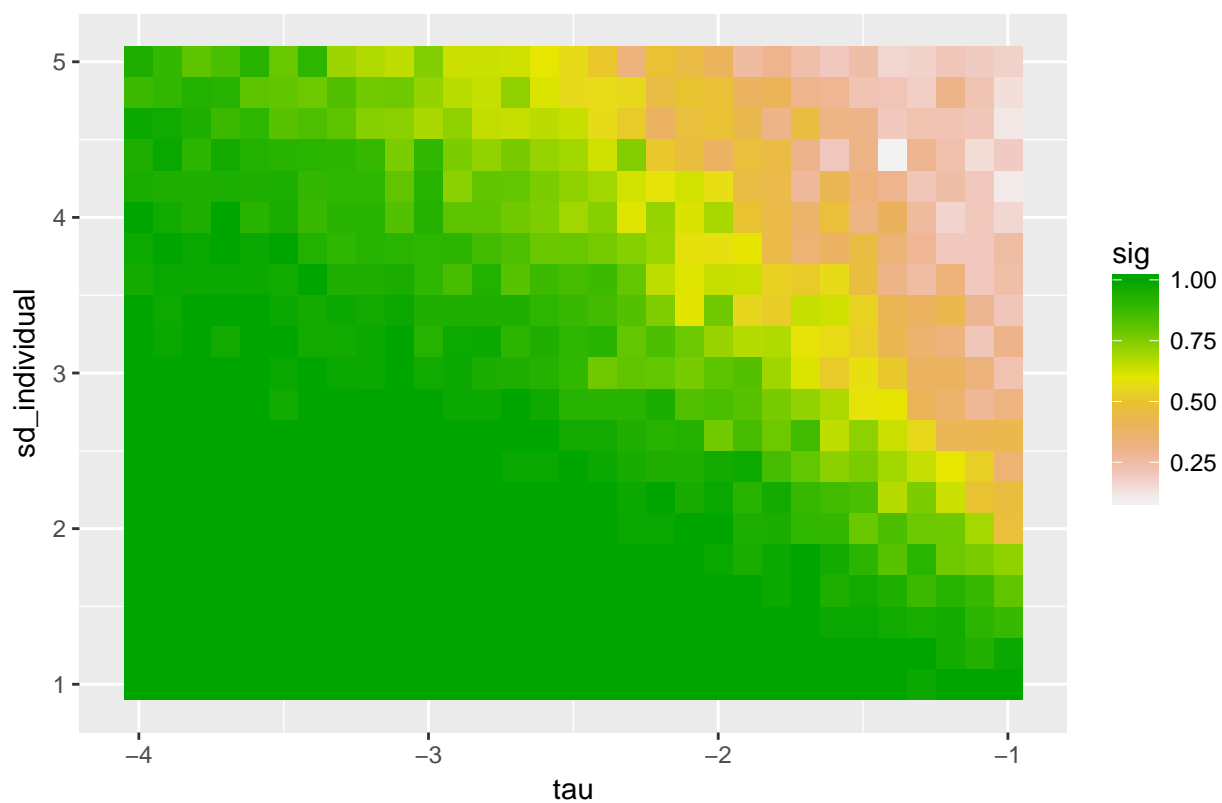
results <- replicate(50000,
  individual_simulation_covariate(80, 8.6, 6.25, sample(moving_tau, size = 1), sample(moving_sd_individual, size = 1),
  simplify = FALSE)
results <- rbindlist(results)
results$sig <- as.numeric(results$pvalue<0.05)

# Calculate power at each treatment effect level and plot them
results_agg <- aggregate(sig ~ tau + sd_individual, FUN=mean, data=results)
results_agg$over80 <- factor(as.numeric(results_agg$sig>=0.8))

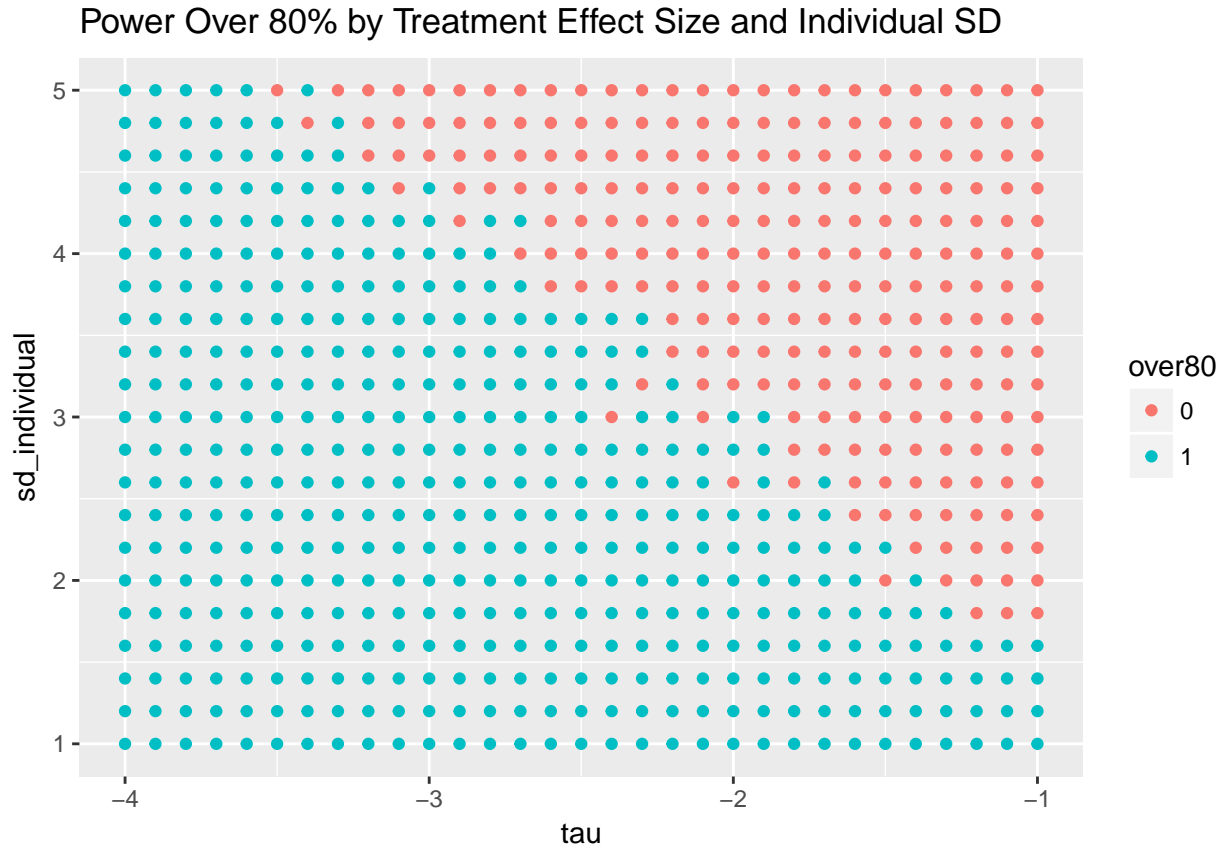
ggplot(results_agg, aes(x=tau, y=sd_individual)) +
  geom_raster(aes(fill = sig)) + scale_fill_gradientn(colours = rev(terrain.colors(10))) +
  ggtitle("Power Level by Treatment Effect Size and Individual SD")

```

Power Level by Treatment Effect Size and Individual SD



```
ggplot(results_agg, aes(x=tau, y=sd_individual, color=over80)) +  
  geom_point() +  
  ggtitle("Power Over 80% by Treatment Effect Size and Individual SD")
```



## The Real Experiment

### *Adjusting the design for the real experiment*

Using the takeaways from the pilot study, we adjusted the design for our experiment to incorporate more optional language around the Chrome Netflix extension. Learning that the quality of the data received through the extension output during the pilot wasn't quite what we had hoped, made us realize we didn't want to make installation of that extension and its output such a strong requirement. We also created a gmail account with the same title as the Qualtric email sender which is "UCBerkeley TV Habits Study." Finally, we incorporated more time in the timeline for both gathering participants (to increase the power of the experiment) but also to give individuals ample time to fill out both the pre- and post-survey in order to maximize the inference ability of the experiment. This is the timeline we worked with:

- Mar 8-11 - gather participants
- Mar 12 - send out pre-survey early in the morning
- Mar 15 - close pre-survey, randomize intervention, and covariance imbalance check
- Mar 18 - treatment/control email [completed 11pm EST]
- Apr 2 - start final data collection
- Apr 6 - end final data collection
- Apr 7 - start data analysis

We gathered 136 email addresses of potential participants of which 42 came from BYU, 38 from social media websites like Facebook and Instagram, 37 from UCBerkeley Slack, and 19 friends of Alyssa, Cameron, and Sarah. There was about 40% attrition thereafter as 83 of the 136 individuals filled out the pre-survey for the experiment.



## Processing the Pre-Survey Data

```
# Load the data
data <- read.csv("UCBerkeley_TV_Habits_Study_Pre-Survey_March_2018_13.01_CLEAN.csv.csv")

# Remove extraneous columns and rename remaining columns
data <- data[,c(12,18:36)]
colnames(data) <- c("linkedEmail", "enteredEmail", "gender", "age", "region", "employment", "maritalStatus", "children", "hoursTV", "binge", "primaryChannel", "moreTimeThanWanted", "watchAlone", "shareProfile", "netflixAccountAndChrome")

# Generate factors for all multiple choice columns
for (i in c(3:8,10:11,13:15,19)){
  data[,i] <- factor(data[,i])
}

# Label answers
levels(data$gender)=c("male", "female")
levels(data$age)=c("21-", "22-34", "35-44", "45-54", "55-64", "65+") [as.numeric(levels(data$age))]
levels(data$region)=c("midwest", "northeast", "southeast", "southwest", "west", "outsideUS") [as.numeric(levels(data$region))]
levels(data$employment)=c("full", "part", "looking", "unemployed", "student", "retired", "homemaker", "self", "other") [as.numeric(levels(data$employment))]
levels(data$maritalStatus)=c("single", "married", "widowed", "divorced", "Separated") [as.numeric(levels(data$maritalStatus))]
levels(data$children)=c("yes", "no") [as.numeric(levels(data$children))]
levels(data$binge)=c("once a week", "once a month", "once every couple months", "once a year", "no") [as.numeric(levels(data$binge))]
levels(data$primaryChannel)=c("netflix", "HBO", "hulu", "amazon", "youtube", "cable", "other") [as.numeric(levels(data$primaryChannel))]
levels(data$moreTimeThanWanted)=c("once a year", "couple times a year", "once a month", "couple times a month", "no") [as.numeric(levels(data$moreTimeThanWanted))]
levels(data$watchAlone)=c("alone", "withOthers") [as.numeric(levels(data$watchAlone))]
levels(data$shareProfile)=c("yes", "no") [as.numeric(levels(data$shareProfile))]
levels(data$netflixAccountAndChrome)=c("both", "noNetflix", "neither", "noChrome")

summary(data)
```

```
##               linkedEmail               enteredEmail
## ace9312@gmail.com      : 1   ace9312@gmail.com      : 1
## alianardo@gmail.com   : 1   alianardo@gmail.com   : 1
## amodeo@berkeley.edu    : 1   amodeo@berkeley.edu    : 1
## amycall7@gmail.com     : 1   amycall7@gmail.com     : 1
## Ariana.viera@gmail.com : 1   ariana.fonnesbeck@gmail.com: 1
## bedoukiane@email.chop.edu: 1   bryanmoore@berkeley.edu : 1
## (Other)                :75   (Other)                :75
##   gender      age      region      employment maritalStatus
## male :36  21- : 8  midwest : 4  full :50  single :41
## female:45  22-34:61 northeast:26  part : 5  married :39
##           35-44: 8  southeast: 4  looking : 1  divorced: 1
##           45-54: 4  southwest: 2  student :19
##           west :42  homemaker: 3
##           outsideUS: 3  self : 3
##
## children  hoursTV      binge  primaryChannel
## yes:17  Min. : 0.50  once a week : 7  netflix:54
## no :64  1st Qu.: 4.00  once a month :13  HBO : 1
##           Median : 7.00  once every couple months:28  hulu : 7
##           Mean : 8.58  once a year :12  amazon : 1
##           3rd Qu.:12.00  no :21  youtube: 4
##           Max. :30.00
##           cable : 4
##           other :10
## allMethods  moreTimeThanWanted  watchAlone shareProfile
```

```
## 1,4 : 9 once a year :14 alone :39 yes:70
## 1,3,4 : 7 couple times a year :27 withOthers:42 no :11
## 1,3,4,5: 7 once a month :12
## 1,4,5 : 6 couple times a month:17
## 1,2,3,4: 4 once a week :11
## 1,5 : 4
## (Other):44
## netflixDays netflixHours netflixMin netflixAccountAndChrome
## 1 :11 Na :10 Na :10 both :57
## Na : 9 : 6 : 5 noNetflix:11
## : 6 7 : 4 36 : 3 neither : 1
## 0 : 6 0 : 3 21 : 2 noChrome :12
## 2 : 3 21 : 3 6 : 2
## (Other):10 (Other):20 (Other):25
## NA's :36 NA's :35 NA's :34
## source
## BYU :14
## Friend :13
## Online (social media):26
## UC Berkeley Slack :28
##
##
##
```

## Randomize treatment

```
# Simple random assignment (treat=1 mean it is in the treatment group)
# set seed so that results of random process are reproducible
set.seed(569320)
data$treat <- sample(c(1,0), size = nrow(data), replace = TRUE)
summary(data$treat)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0000 0.0000 1.0000 0.5185 1.0000 1.0000
```

## Covariate balance check

```
# Statistical F-test
model <- lm(treat ~ gender + age + region + employment + maritalStatus + children + hoursTV + binge + p
summary(model)
```

```
##
## Call:
## lm(formula = treat ~ gender + age + region + employment + maritalStatus +
## children + hoursTV + binge + primaryChannel + moreTimeThanWanted +
## watchAlone + shareProfile + source, data = data)
##
## Residuals:
## Min 1Q Median 3Q Max
```

```

## -0.83086 -0.28271 -0.01469 0.31640 0.90032
##
## Coefficients:
## Estimate Std. Error t value
## (Intercept) 0.592958 0.634099 0.935
## genderfemale -0.015549 0.169939 -0.091
## age22-34 -0.539260 0.308423 -1.748
## age35-44 -0.600747 0.424798 -1.414
## age45-54 -1.306264 0.528002 -2.474
## regionnortheast 0.267982 0.341704 0.784
## regionsoutheast -0.519488 0.471135 -1.103
## regionsouthwest -0.295409 0.874745 -0.338
## regionwest 0.034080 0.349848 0.097
## regionoutsideUS -0.151552 0.527816 -0.287
## employmentpart 0.303626 0.343494 0.884
## employmentlooking -0.066978 0.682308 -0.098
## employmentstudent 0.072856 0.294871 0.247
## employmenthomemaker 0.318297 0.624702 0.510
## employmentself 0.271009 0.446795 0.607
## maritalStatusmarried 0.158713 0.181386 0.875
## maritalStatusdivorced 0.150948 0.758664 0.199
## childrenno 0.079323 0.236747 0.335
## hoursTV 0.003951 0.015881 0.249
## bingeonce a month -0.341357 0.323328 -1.056
## bingeonce every couple months -0.102676 0.336912 -0.305
## bingeonce a year -0.402906 0.391381 -1.029
## bingenno 0.074337 0.359537 0.207
## primaryChannelHBO -0.595417 0.607137 -0.981
## primaryChannelhulu -0.090349 0.276323 -0.327
## primaryChannelamazon 0.936428 0.621561 1.507
## primaryChannelyoutube 0.121141 0.385527 0.314
## primaryChannelcable -0.066489 0.346772 -0.192
## primaryChannelother 0.176490 0.232485 0.759
## moreTimeThanWantedcouple times a year -0.016908 0.246426 -0.069
## moreTimeThanWantedonce a month 0.071240 0.271740 0.262
## moreTimeThanWantedcouple times a month 0.139692 0.254396 0.549
## moreTimeThanWantedonce a week -0.124344 0.288242 -0.431
## watchAlonewithOthers -0.074714 0.173103 -0.432
## shareProfileno 0.074293 0.302336 0.246
## sourceFriend 0.241597 0.406697 0.594
## sourceOnline (social media) 0.312077 0.294752 1.059
## sourceUC Berkeley Slack 0.432445 0.345034 1.253
## Pr(>|t|)
## (Intercept) 0.3549
## genderfemale 0.9275
## age22-34 0.0875 .
## age35-44 0.1645
## age45-54 0.0174 *
## regionnortheast 0.4372
## regionsoutheast 0.2763
## regionsouthwest 0.7372
## regionwest 0.9228
## regionoutsideUS 0.7754
## employmentpart 0.3816

```

```

## employmentlooking                0.9223
## employmentstudent                0.8060
## employmenthomemaker              0.6130
## employmentself                   0.5473
## maritalStatusmarried              0.3864
## maritalStatusdivorced            0.8432
## childrenno                       0.7392
## hoursTV                          0.8047
## bingeonce a month                0.2970
## bingeonce every couple months    0.7620
## bingeonce a year                 0.3090
## bingen                             0.8372
## primaryChannelHBO                 0.3322
## primaryChannelhulu                0.7453
## primaryChannelamazon              0.1392
## primaryChannelyoutube             0.7549
## primaryChannelcable              0.8489
## primaryChannelother               0.4519
## moreTimeThanWantedcouple times a year 0.9456
## moreTimeThanWantedonce a month    0.7944
## moreTimeThanWantedcouple times a month 0.5858
## moreTimeThanWantedonce a week    0.6683
## watchAlonewithOthers              0.6682
## shareProfileno                   0.8071
## sourceFriend                     0.5556
## sourceOnline (social media)       0.2956
## sourceUC Berkeley Slack           0.2169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5192 on 43 degrees of freedom
## Multiple R-squared:  0.4269, Adjusted R-squared:  -0.06623
## F-statistic: 0.8657 on 37 and 43 DF,  p-value: 0.6711

# Checked F-stat and p-value. Null hypothesis is that the coefficients are jointly equal to 0
# We cannot reject the null that the variables are jointly insignificant

# Additionally, examine levels of key covariates by treatment/control

se_diff_means <- function(treatment, control) {
  round(sqrt(sd(control)^2/length(control) + sd(treatment)^2/length(treatment)),2)
}

# hours of TV
t1 <- round(mean(data$hoursTV[data$treat==1]),2)
c1 <- round(mean(data$hoursTV[data$treat==0]),2)
diff1 <- t1-c1
se1 <- se_diff_means(data$hoursTV[data$treat==1],data$hoursTV[data$treat==0])

# male
t2 <- round(mean(data$gender[data$treat==1]=="male"),2)
c2 <- round(mean(data$gender[data$treat==0]=="male"),2)
diff2 <- t2-c2
se2 <- se_diff_means(as.numeric(data$gender[data$treat==1]=="male"),as.numeric(data$gender[data$treat==0]=="male"))

```

```

# marital status = married
t3 <- round(mean(data$maritalStatus[data$treat==1]=="married"),2)
c3 <- round(mean(data$maritalStatus[data$treat==0]=="married"),2)
diff3 <- t3-c3
se3 <- se_diff_means(as.numeric(data$maritalStatus[data$treat==1]=="married"),as.numeric(data$maritalStatus[data$treat==0]=="married"),2)

# no children
t4 <- round(mean(data$children[data$treat==1]=="no"),2)
c4 <- round(mean(data$children[data$treat==0]=="no"),2)
diff4 <- t4-c4
se4 <- se_diff_means(as.numeric(data$children[data$treat==1]=="no"),as.numeric(data$children[data$treat==0]=="no"),2)

# moreTimeThanWanted: couple times a year
t5 <- round(mean(data$moreTimeThanWanted[data$treat==1]=="couple times a year"),2)
c5 <- round(mean(data$moreTimeThanWanted[data$treat==0]=="couple times a year"),2)
diff5 <- t5-c5
se5 <- se_diff_means(as.numeric(data$moreTimeThanWanted[data$treat==1]=="couple times a year"),as.numeric(data$moreTimeThanWanted[data$treat==0]=="couple times a year"),2)

# moreTimeThanWanted: once a month
t6 <- round(mean(data$moreTimeThanWanted[data$treat==1]=="once a month"),2)
c6 <- round(mean(data$moreTimeThanWanted[data$treat==0]=="once a month"),2)
diff6 <- t6-c6
se6 <- se_diff_means(as.numeric(data$moreTimeThanWanted[data$treat==1]=="once a month"),as.numeric(data$moreTimeThanWanted[data$treat==0]=="once a month"),2)

# moreTimeThanWanted: couple times a month
t7 <- round(mean(data$moreTimeThanWanted[data$treat==1]=="couple times a month"),2)
c7 <- round(mean(data$moreTimeThanWanted[data$treat==0]=="couple times a month"),2)
diff7 <- t7-c7
se7 <- se_diff_means(as.numeric(data$moreTimeThanWanted[data$treat==1]=="couple times a month"),as.numeric(data$moreTimeThanWanted[data$treat==0]=="couple times a month"),2)

# moreTimeThanWanted: once a week
t8 <- round(mean(data$moreTimeThanWanted[data$treat==1]=="once a week"),2)
c8 <- round(mean(data$moreTimeThanWanted[data$treat==0]=="once a week"),2)
diff8 <- t8-c8
se8 <- se_diff_means(as.numeric(data$moreTimeThanWanted[data$treat==1]=="once a week"),as.numeric(data$moreTimeThanWanted[data$treat==0]=="once a week"),2)

# Put into a table for display
d <- data.frame(variable = c("hours TV", "male", "married", "no children", "watched more than wanted: couple times a year", "watched more than wanted: once a month", "watched more than wanted: couple times a month"),
  control = c(t1, t2, t3, t4, t5, t6, t7, t8),
  treatment = c(c1, c2, c3, c4, c5, c6, c7, c8),
  diff = c(diff1, diff2, diff3, diff4, diff5, diff6, diff7, diff8),
  se = c(se1, se2, se3, se4, se5, se6, se7, se8) )
knitr::kable(d)

```

variable	control	treatment	diff	se
hours TV	8.74	8.41	0.33	1.40
male	0.48	0.41	0.07	0.11
married	0.43	0.54	-0.11	0.11
no children	0.81	0.77	0.04	0.09
watched more than wanted: couple times a year	0.29	0.38	-0.09	0.11
watched more than wanted: once a month	0.14	0.15	-0.01	0.08
watched more than wanted: couple times a month	0.24	0.18	0.06	0.09

variable	control	treatment	diff	se
watched more than wanted: once a week	0.14	0.13	0.01	0.08

Finally, output the treatment assignments along with emails.

```
output_data = data[,c("linkedEmail", "treat")]
head(output_data)
```

```
##           linkedEmail treat
## 1 sarahkelley1759@gmail.com 1
## 2 jennyq.wu@gmail.com      0
## 3 j.fallentine@gmail.com   0
## 4 phat.t.doan@gmail.com    0
## 5 krissy1734@outlook.com   0
## 6 amodeo@berkeley.edu      1
```

```
write.csv(output_data, file = "ExperimentTreatmentAssignment.csv")
```

```
#load the survey data
```

```
survey_data <- read.csv("UCBerkeleyTVHabitsStudyPost-Survey_April2018CLEAN.csv")
survey_data$Q3_1 <- as.numeric(survey_data$Q3_1)
survey_data$Q3_2 <- as.numeric(survey_data$Q3_2)
survey_data$Q3_3 <- as.numeric(survey_data$Q3_3)
```

```
#first just pull the relevant columns from survey data
```

```
cols <- c(12, 18:24)
```

```
sub.data <- survey_data[,cols]
```

```
sub.data$Finished<- 1 #add a variable for if they finished the post-survey
```

```
colnames(sub.data)=c("linkedEmail", "post_survey_ID", "reportedhrs_P", "netflixDays_P", "netflixHours_P")
```

```
#join the pre-survey & survey data set using emails
```

```
all_data <- merge(data, sub.data, by="linkedEmail", all=TRUE)
```

```
all_data$Finished <- ifelse(is.na(all_data$Finished), 0, 1)
```

```
all_data$Assigned <- all_data$treat
```

```
all_data[, -21]
```

```
##           linkedEmail           enteredEmail gender  age
## 1 ace9312@gmail.com ace9312@gmail.com male 22-34
## 2 alianardo@gmail.com alianardo@gmail.com female 21-
## 3 amodeo@berkeley.edu amodeo@berkeley.edu male 35-44
## 4 amycall7@gmail.com amycall7@gmail.com female 45-54
## 5 Ariana.viera@gmail.com ariana.fonnesbeck@gmail.com female 22-34
## 6 bedoukiane@email.chop.edu Ebedoukian@gmail.com female 22-34
## 7 bryanmoore@berkeley.edu bryanmoore@berkeley.edu male 22-34
## 8 carmen.east@berkeley.edu carmen.easterwood@gmail.com female 22-34
## 9 carolinenbell@gmail.com carolinenbell@gmail.com female 22-34
## 10 Chrisells21@gmail.com ELLSWORTH_21@HOTMAIL.COM male 22-34
## 11 collin.whitmore@gmail.com collin.whitmore@gmail.com male 22-34
## 12 Conor.b.fitzgerald@gmail.com conor.b.fitzgerald@gmail.com male 22-34
## 13 danrasband@berkeley.edu danrasband@berkeley.edu male 22-34
## 14 Desiree.chavez28@gmail.com desiree.chavez28@gmail.com female 22-34
## 15 dianaxsim@gmail.com dianaxsim@gmail.com female 22-34
## 16 dkent@berkeley.edu rrdkent@gmail.com male 22-34
## 17 dpaulin@berkeley.edu Dpaulin@berkeley.edu male 35-44
```

## 18	dzlarson332@gmail.com	dzlarson332@gmail.com	male	21-
## 19	e.ellsworth4@gmail.com	e.ellsworth4@gmail.com	female	22-34
## 20	Ericpfreeman@gmail.com	ericpfreeman@gmail.com	male	35-44
## 21	g.s.chahal@berkeley.edu	g.s.chahal@berkeley.edu	male	22-34
## 22	gerard.john.kelly@gmail.com	gkelly@berkeley.edu	male	22-34
## 23	gubler12@hotmail.com	gubler12@hotmail.com	male	22-34
## 24	harrison.ostler@byu.edu	harrisonko9@gmail.com	male	22-34
## 25	highamjosh@gmail.com	highamjosh@gmail.com	male	22-34
## 26	Hilary.nicholson@gmail.com	Hilary.nicholson@gmail.com	female	22-34
## 27	hule1195@gmail.com	hule1195@gmail.com	female	22-34
## 28	igelfenboym@berkeley.edu	igelfenboym@berkeley.edu	female	22-34
## 29	j.fallentine@gmail.com	j.fallentine@gmail.com	male	22-34
## 30	jasonmogle@gmail.com	jasonmogle@gmail.com	male	22-34
## 31	jeannette.sue@gmail.com	jeannette.sue@gmail.com	female	22-34
## 32	jennyq.wu@gmail.com	jennyq.wu@gmail.com	female	35-44
## 33	jill.rosok@gmail.com	jill.rosok@gmail.com	female	22-34
## 34	jonahhsmith@berkeley.edu	jonahhsmith@berkeley.edu	male	35-44
## 35	joseph.simps@gmail.com	joseph.simps@gmail.com	male	22-34
## 36	Jpark1127@gmail.com	jpark1127@gmail.com	male	22-34
## 37	julia.r.kapoor@gmail.com	julia.r.kapoor@gmail.com	female	22-34
## 38	juliab@berkeley.edu	juliab@berkeley.edu	female	22-34
## 39	kamber.mccormick@gmail.com	kamber.mccormick@gmail.com	female	22-34
## 40	kanitha@berkeley.edu	kanitha@berkeley.edu	female	22-34
## 41	katebaird2763@gmail.com	Katebaird2763@gmail.com	female	21-
## 42	kathrynhamilton@berkeley.edu	kathrynhamilton@berkeley.edu	female	22-34
## 43	kaylee.dudley@gmail.com	kaylee.dudley@gmail.com	female	22-34
## 44	krissy1734@outlook.com	krissy1734@outlook.com	female	22-34
## 45	kyle.m.zeller@gmail.com	kyle.m.zeller@gmail.com	male	22-34
## 46	Lisa@merrillimages.com	lisa@merrillimages.com	female	45-54
## 47	ltteichert@gmail.com	ltteichert@gmail.com	male	21-
## 48	machupicchu2014@gmail.com	machupicchu2014@gmail.com	male	21-
## 49	Mackenziehh@hotmail.com	mackenziehh@hotmail.com	female	22-34
## 50	mariaafabiano@gmail.com	mariaafabiano@gmail.com	female	22-34
## 51	markdthomp@gmail.com	markdthomp@gmail.com	male	22-34
## 52	Mathaholic@berkeley.edu	mathaholic@berkeley.edu	female	35-44
## 53	Meiazn@yahoo.com	Meiazn@yahoo.com	female	22-34
## 54	mells7@hotmail.com	mells7@hotmail.com	female	35-44
## 55	michelle.m.kim@berkeley.edu	michelle.m.kim@berkeley.edu	female	22-34
## 56	miketp333@gmail.com	Miketp333@gmail.com	male	22-34
## 57	mramsay2017@gmail.com	mramsay2017@gmail.com	female	21-
## 58	mursil@gmail.com	mursil@gmail.com	male	22-34
## 59	nsorensen64@gmail.com	nsorensen64@gmail.com	female	22-34
## 60	paigecarver@gmail.com	paigecarver@gmail.com	female	21-
## 61	pdurkin@berkeley.edu	pdurkin@berkeley.edu	male	45-54
## 62	phat.t.doan@gmail.com	phat.t.doan@gmail.com	female	22-34
## 63	Rachaelwolfe35@gmail.com	Rachael	female	22-34
## 64	rachelaedo@gmail.com	rachelaedo@gmail.com	female	22-34
## 65	Rosche.kenny@gmail.com	rosche.kenny@gmail.com	male	22-34
## 66	rutikasb@gmail.com	rutikasb@gmail.com	female	22-34
## 67	samibangerter1@gmail.com	samibangerter1@gmail.com	female	21-
## 68	Samman1108@gmail.com	samman1108@gmail.com	male	22-34
## 69	sarahkelley1759@gmail.com	sarahkelley1759@gmail.com	female	22-34
## 70	sarahsheppard@gmail.com	sarahsheppard@gmail.com	female	22-34
## 71	scott.george@berkeley.edu	scott.george@berkeley.edu	male	35-44

## 72	seddon@berkeley.edu	seddon@berkeley.edu	male	45-54
## 73	Shahbakht.hamdani@gmail.com	shahbakht.hamdani@gmail.com	male	22-34
## 74	sharadv@berkeley.edu	sharadv@berkeley.edu	male	22-34
## 75	sophia.a.chang@gmail.com	sophia.a.chang@gmail.com	female	22-34
## 76	Stefperez@sbcglobal.net	Stefperez@sbcglobal.net	female	22-34
## 77	Summergirl1992@gmail.com	summergirl1992@gmail.com	female	22-34
## 78	valkuo@gmail.com	Valkuo@gmail.com	female	22-34
## 79	wakystuf@gmail.com	wakystuf@gmail.com	male	22-34
## 80	Whitney.sam@gmail.com	whitney.sam@gmail.com	male	22-34
## 81	xiaokiiwii@gmail.com	xiaokiiwii@gmail.com	female	22-34
##	region	employment	maritalStatus	children hoursTV
## 1	west	student	married	no 10.0
## 2	west	part	single	no 5.0
## 3	northeast	student	divorced	yes 15.0
## 4	west	part	married	yes 3.0
## 5	northeast	homemaker	married	no 5.0
## 6	northeast	full	married	no 10.0
## 7	west	full	married	no 4.5
## 8	west	full	single	no 12.0
## 9	west	student	single	no 8.0
## 10	midwest	full	single	no 14.0
## 11	west	student	single	no 10.0
## 12	northeast	full	married	no 12.0
## 13	west	full	married	yes 10.0
## 14	west	student	married	no 3.0
## 15	northeast	full	married	no 25.0
## 16	west	full	single	no 6.0
## 17	west	full	married	yes 12.0
## 18	west	student	single	no 3.0
## 19	west	full	single	no 11.0
## 20	northeast	full	married	yes 25.0
## 21	west	full	single	no 6.0
## 22	outsideUS	part	married	yes 2.0
## 23	west	full	single	no 5.0
## 24	west	student	married	no 1.0
## 25	northeast	full	married	yes 3.0
## 26	northeast	full	married	no 4.0
## 27	west	full	single	no 5.0
## 28	west	full	married	no 8.0
## 29	west	student	single	no 4.0
## 30	west	student	single	no 6.0
## 31	northeast	full	single	no 20.0
## 32	northeast	full	married	yes 5.0
## 33	northeast	full	single	no 14.0
## 34	northeast	looking	married	yes 10.0
## 35	west	student	married	no 15.0
## 36	northeast	full	single	no 6.0
## 37	northeast	full	married	no 10.0
## 38	northeast	full	single	no 7.0
## 39	southwest	homemaker	married	yes 1.0
## 40	southeast	full	married	no 15.0
## 41	west	student	single	no 2.0
## 42	midwest	full	single	no 3.0
## 43	west	student	single	no 30.0



## 44	west	full	married	no	3.0
## 45	northeast	self	single	no	20.0
## 46	west	self	married	no	4.0
## 47	west	part	single	no	1.0
## 48	west	student	single	no	0.5
## 49	west	full	married	yes	14.0
## 50	west	part	married	no	9.0
## 51	northeast	full	single	no	5.0
## 52	west	full	married	no	7.0
## 53	southeast	full	married	no	3.0
## 54	southwest	homemaker	married	yes	10.0
## 55	southeast	full	married	no	5.0
## 56	northeast	full	single	no	10.0
## 57	west	student	single	no	5.0
## 58	west	full	married	no	12.0
## 59	northeast	full	single	no	10.0
## 60	west	student	single	no	3.0
## 61	outsideUS	student	married	yes	10.0
## 62	west	full	married	yes	5.0
## 63	northeast	full	single	no	5.0
## 64	west	student	single	no	5.0
## 65	midwest	full	single	no	4.0
## 66	west	full	single	no	15.0
## 67	west	student	single	no	10.0
## 68	northeast	full	single	no	18.0
## 69	west	full	single	no	3.0
## 70	northeast	full	married	no	1.0
## 71	west	full	single	no	17.5
## 72	southeast	self	married	yes	10.0
## 73	outsideUS	full	single	no	25.0
## 74	northeast	full	single	no	12.0
## 75	northeast	full	married	no	2.0
## 76	midwest	student	single	no	3.5
## 77	west	full	married	no	7.0
## 78	northeast	full	married	yes	2.0
## 79	northeast	full	single	no	15.0
## 80	west	full	married	yes	8.0
## 81	west	full	single	no	10.0
##		binge	primaryChannel	allMethods	
## 1	once every couple months		netflix	1,4,5,7	
## 2		no	netflix	1,4,5	
## 3		no	netflix	1,2,4,6,7	
## 4		no	netflix	1,3,4	
## 5	once every couple months		netflix	1,4	
## 6		once a month	netflix	1,3,4	
## 7		once a year	netflix	1,2,4,5,7	
## 8		once a month	netflix	1,2,4,5	
## 9	once every couple months		netflix	1,4,5	
## 10		no	netflix	1,3,4,5	
## 11		once a month	netflix	1,3,4,5,6,7	
## 12	once every couple months		cable	1,2,3,5,6	
## 13	once every couple months		netflix	1,4	
## 14		once a year	netflix	1,3,4,5	
## 15		once a week	hulu	1,3,4,5	

## 16	once a year	netflix	1,3,4,5
## 17	once a month	netflix	1,2,4,6
## 18	once every couple months	youtube	4,5,7
## 19	once a month	netflix	1,3,4
## 20	once a week	netflix	1,4,6
## 21	once every couple months	netflix	1,3,4,5
## 22	once every couple months	other	1,5,7
## 23	once a week	netflix	1,2,4,5,7
## 24	no	netflix	1
## 25	no	netflix	1,2,3,4
## 26	once every couple months	netflix	1,3
## 27	once a year	amazon	1,4,6
## 28	once a year	netflix	1,4,7
## 29	once every couple months	netflix	1,4
## 30	once every couple months	other	1,3,4,5,7
## 31	once a month	netflix	1,5
## 32	once a month	netflix	1,2,3,4
## 33	once every couple months	youtube	1,2,4,5,6
## 34	no	other	1,2,4,7
## 35	once a month	netflix	1,4,5
## 36	no	netflix	1,2,3,5
## 37	once every couple months	netflix	1,2,3,4,5,6
## 38	once a year	netflix	1,6
## 39	no	netflix	1,4,5
## 40	once every couple months	other	1,5,7
## 41	no	netflix	1,5,7
## 42	once every couple months	netflix	1,4
## 43	once a week	netflix	1,5
## 44	once every couple months	netflix	1,4
## 45	once a week	netflix	1,4,7
## 46	no	netflix	1,2,4,5,6
## 47	no	youtube	5,6,7
## 48	no	other	7
## 49	once every couple months	hulu	3,4
## 50	once every couple months	hulu	1,3,4
## 51	once a year	other	1,2,4,6,7
## 52	once every couple months	netflix	1,3,4
## 53	once a year	netflix	1,2
## 54	no	cable	1,6
## 55	once every couple months	other	1,6,7
## 56	once a year	cable	1,2,6
## 57	once a month	netflix	1,3,4
## 58	once a week	netflix	1,3,4
## 59	once every couple months	hulu	1,4
## 60	once every couple months	netflix	1,5
## 61	no	netflix	1,7
## 62	once a year	hulu	1,3,4,5
## 63	once every couple months	cable	1,4
## 64	no	netflix	1,3,6
## 65	once every couple months	netflix	1,2,4,7
## 66	once every couple months	netflix	1,5
## 67	once every couple months	netflix	1,4
## 68	once a month	hulu	1,2,3,5,7
## 69	no	netflix	1,4

## 70	once every couple months	HBO	1,2,3,4		
## 71	once a week	youtube	1,2,4,5		
## 72	no	netflix	1,2,4,6		
## 73	once a month	netflix	1,4,5		
## 74	no	netflix	1,2,4,5		
## 75	no	other	1,7		
## 76	once every couple months	netflix	1,2,3,4		
## 77	once a month	hulu	1,3,4,5		
## 78	once a year	other	1,2,5,7		
## 79	once a month	netflix	1,2,4,7		
## 80	no	netflix	1,4,5		
## 81	once a year	other	1,2,3,4,5,7		
##	moreTimeThanWanted watchAlone	shareProfile	netflixDays	netflixHours	
## 1	couple times a month withOthers	yes	0	8	
## 2	once a week withOthers	yes	<NA>	<NA>	
## 3	once a week alone	no	3	2	
## 4	couple times a year withOthers	no			
## 5	couple times a year withOthers	yes	1	21	
## 6	once a month alone	yes	Na	Na	
## 7	couple times a month alone	yes	<NA>	<NA>	
## 8	couple times a month withOthers	yes	1	10	
## 9	once a week alone	no	3	11	
## 10	couple times a month alone	no	0	20	
## 11	couple times a month alone	no	0	23	
## 12	once a month withOthers	yes	<NA>	<NA>	
## 13	couple times a month withOthers	yes	1	7	
## 14	once a month withOthers	yes	1	21	
## 15	once a week withOthers	yes	Na	Na	
## 16	once a year withOthers	yes		8	
## 17	couple times a year withOthers	yes	Na	Na	
## 18	couple times a month alone	no	<NA>	<NA>	
## 19	once a week alone	yes	<NA>	<NA>	
## 20	once a year alone	yes	7	2	
## 21	couple times a year withOthers	yes	<NA>	<NA>	
## 22	couple times a year alone	yes	0	0	
## 23	once a year withOthers	yes	<NA>	<NA>	
## 24	once a month withOthers	yes	Na	Na	
## 25	couple times a year withOthers	yes			
## 26	once a year withOthers	yes	<NA>	<NA>	
## 27	couple times a year alone	yes	<NA>	<NA>	
## 28	once a year withOthers	yes		7	
## 29	couple times a year alone	yes	<NA>	Na	
## 30	couple times a month alone	yes	<NA>	<NA>	
## 31	once a week alone	yes	2	21	
## 32	couple times a year withOthers	yes			
## 33	couple times a month alone	yes	1	4	
## 34	once a year withOthers	yes	<NA>	<NA>	
## 35	couple times a year alone	yes	1		
## 36	once a month withOthers	yes	<NA>	<NA>	
## 37	once a year alone	yes	Na	Na	
## 38	couple times a month alone	yes	1	17	
## 39	once a month withOthers	yes	NA	<NA>	
## 40	once a month withOthers	yes	<NA>	<NA>	
## 41	once a week alone	yes	<NA>	<NA>	

## 42	once a month withOthers	yes	0	0
## 43	once a month alone	yes	5	9
## 44	couple times a year withOthers	yes	<NA>	<NA>
## 45	couple times a month alone	yes	3	9
## 46	once a year withOthers	no	<NA>	<NA>
## 47	couple times a year alone	no	<NA>	<NA>
## 48	couple times a year withOthers	no	<NA>	<NA>
## 49	couple times a year withOthers	yes	N/a	
## 50	couple times a month withOthers	yes	2	4
## 51	couple times a year withOthers	yes	<NA>	<NA>
## 52	couple times a year withOthers	yes	4	5
## 53	couple times a month alone	yes	Na	Na
## 54	once a week alone	yes	na	na
## 55	once a year alone	yes	<NA>	<NA>
## 56	couple times a year withOthers	yes	<NA>	<NA>
## 57	couple times a month withOthers	yes	Na	Na
## 58	once a year alone	yes	<NA>	<NA>
## 59	once a month withOthers	yes	<NA>	<NA>
## 60	couple times a year withOthers	yes	0	7
## 61	once a month withOthers	no	<NA>	<NA>
## 62	couple times a year withOthers	yes	na	na
## 63	once a year alone	yes	Na	Na
## 64	once a week alone	yes	<NA>	<NA>
## 65	couple times a year alone	yes	1	7
## 66	couple times a month alone	yes	1	13
## 67	couple times a year alone	yes	<NA>	<NA>
## 68	couple times a year alone	yes	<NA>	<NA>
## 69	once a year withOthers	yes	<NA>	<NA>
## 70	couple times a year withOthers	yes	<NA>	na
## 71	once a week alone	no	2	17
## 72	couple times a year withOthers	yes	<NA>	<NA>
## 73	couple times a year alone	yes	1	0
## 74	once a week alone	yes	<NA>	<NA>
## 75	once a year withOthers	yes	<NA>	<NA>
## 76	once a month alone	yes		
## 77	couple times a month withOthers	yes	<NA>	<NA>
## 78	couple times a year alone	yes	Na	Na
## 79	couple times a month alone	yes	1	15
## 80	once a year withOthers	yes	<NA>	<NA>
## 81	couple times a year alone	yes	<NA>	<NA>
##	netflixMin netflixAccountAndChrome		source	
## 1	36 both		BYU	
## 2	<NA> noNetflix		BYU	
## 3	50 both		UC Berkeley Slack	
## 4		neither	Online (social media)	
## 5	17 both		Online (social media)	
## 6	Na both		Friend	
## 7	<NA> both		UC Berkeley Slack	
## 8	55 both		UC Berkeley Slack	
## 9	24 83 both		Online (social media)	
## 10	26 both		Online (social media)	
## 11	47 both		Online (social media)	
## 12	<NA> both		Online (social media)	
## 13	33 both		UC Berkeley Slack	

## 14	24	both	Online (social media)
## 15	Na	both	Friend
## 16	30	both	UC Berkeley Slack
## 17	Na	both	UC Berkeley Slack
## 18	<NA>	noNetflix	BYU
## 19	<NA>	both	Online (social media)
## 20	9	both	UC Berkeley Slack
## 21	<NA>	both	UC Berkeley Slack
## 22	31	both	UC Berkeley Slack
## 23	<NA>	both	BYU
## 24	Na	both	BYU
## 25		both	Online (social media)
## 26	<NA>	noChrome	Online (social media)
## 27	<NA>	both	Online (social media)
## 28	10	both	UC Berkeley Slack
## 29	Na	both	BYU
## 30	<NA>	noNetflix	BYU
## 31	36	both	Friend
## 32		both	Friend
## 33	48	both	UC Berkeley Slack
## 34	<NA>	noNetflix	UC Berkeley Slack
## 35	35	both	BYU
## 36	<NA>	noNetflix	Online (social media)
## 37	Na	both	Friend
## 38	6	both	UC Berkeley Slack
## 39	NA	noChrome	Online (social media)
## 40	<NA>	noNetflix	UC Berkeley Slack
## 41	<NA>	both	BYU
## 42	21	both	UC Berkeley Slack
## 43	43	both	Online (social media)
## 44	<NA>	noChrome	UC Berkeley Slack
## 45	36	both	Friend
## 46	<NA>	noChrome	Online (social media)
## 47	<NA>	noNetflix	BYU
## 48	<NA>	noNetflix	BYU
## 49		noNetflix	Online (social media)
## 50	6	both	Online (social media)
## 51	<NA>	both	Friend
## 52	9	both	UC Berkeley Slack
## 53	Na	both	Friend
## 54	na	both	Online (social media)
## 55	<NA>	noChrome	UC Berkeley Slack
## 56	<NA>	noChrome	UC Berkeley Slack
## 57	Na	both	BYU
## 58	<NA>	both	UC Berkeley Slack
## 59	<NA>	both	Online (social media)
## 60	23	both	Online (social media)
## 61	<NA>	noChrome	UC Berkeley Slack
## 62	na	both	UC Berkeley Slack
## 63	Na	noChrome	Online (social media)
## 64	<NA>	both	BYU
## 65	22	both	Friend
## 66	11	both	UC Berkeley Slack
## 67	<NA>	both	BYU

## 68	<NA>	both	Online (social media)		
## 69	<NA>	noChrome	UC Berkeley Slack		
## 70	nA	noChrome	Friend		
## 71	21	both	UC Berkeley Slack		
## 72	<NA>	noChrome	UC Berkeley Slack		
## 73	2	both	UC Berkeley Slack		
## 74	<NA>	both	UC Berkeley Slack		
## 75	<NA>	noNetflix	Online (social media)		
## 76		noChrome	Online (social media)		
## 77	<NA>	both	Online (social media)		
## 78	Na	both	Friend		
## 79	19	both	Friend		
## 80	<NA>	noNetflix	Online (social media)		
## 81	<NA>	both	Friend		
##	post_survey_ID	reportedhrs_P	netflixDays_P	netflixHours_P	
## 1	ace9312@gmail.com	20.0	1	1	
## 2	alianardo@gmail.com	10.0	NA	NA	
## 3	amodeo@berkeley.edu	20.0	1	6	
## 4	amycall7@gmail.com	14.0	3	15	
## 5	ariana.viera@gmail.com	9.0	1	11	
## 6	bedoukiane@email.chop.edu	10.0	NA	NA	
## 7	<NA>	NA	NA	NA	
## 8	carmen.east@berkeley.edu	8.0	1	10	
## 9	carolinenbell@gmail.com	13.0	1	2	
## 10	chrisells21@gmail.com	30.0	NA	NA	
## 11	collin.whitmore@gmail.com	35.0	1	14	
## 12	Conor.b.fitzgerald@gmail.com	16.0	1	13	
## 13	danrasband@berkeley.edu	20.0	1	9	
## 14	desiree.chavez28@gmail.com	20.0	1	4	
## 15	dianaxsim@gmail.com	40.0	5	16	
## 16	dkent@berkeley.edu	30.0	1	12	
## 17	<NA>	NA	NA	NA	
## 18	<NA>	NA	NA	NA	
## 19	<NA>	NA	NA	NA	
## 20	ericpfreeman@gmail.com	80.0	2	3	
## 21	g.s.chahal@berkeley.edu	17.0	NA	NA	
## 22	gerard.john.kelly@gmail.com	3.0	1	1	
## 23	gubler12@hotmail.com	22.0	NA	NA	
## 24	harrisonko9@gmail.com	15.0	NA	NA	
## 25	<NA>	NA	NA	NA	
## 26	Hilary.nicholson@gmail.com	15.0	NA	NA	
## 27	hule1195@gmail.com	6.0	NA	NA	
## 28	igelfenboym@berkeley.edu	15.0	5	14	
## 29	J.fallentine@gmail.com	6.0	6	17	
## 30	jasonmogle@gmail.com	5.0	NA	NA	
## 31	jeannette.sue@gmail.com	40.0	1	7	
## 32	Jenny.wu@gmail.com	5.0	NA	16	
## 33	jill.rosok@gmail.com	10.0	1	10	
## 34	<NA>	NA	NA	NA	
## 35	joseph.simps@gmail.com	20.0	1	1	
## 36	<NA>	NA	NA	NA	
## 37	<NA>	NA	NA	NA	
## 38	juliab@berkeley.edu	20.0	1	12	
## 39	Kamber.mccormick@gmail.com	10.0	NA	NA	

## 40	kanitha@berkeley.edu	9.0	NA	NA
## 41	katebaird2763@gmail.com	14.0	NA	NA
## 42	kathryn.hamilton13@gmail.com	12.0	6	17
## 43	kaylee.dudley@gmail.com	59.0	1	5
## 44	krissy1734@outlook.com	9.0	NA	NA
## 45	kyle.m.zeller@gmail.com	3.0	2	10
## 46	lisa@merrillimages.com	4.0	NA	NA
## 47	ltteichert@gmail.com	12.5	NA	NA
## 48	machupicchu2014@gmail.com	6.0	NA	NA
## 49	mackenziehh@hotmail.com	17.0	4	1
## 50	mariaafabiano@gmail.com	30.0	1	1
## 51	<NA>	NA	NA	NA
## 52	nikkilovesmath@gmail.com	24.0	2	1
## 53	<NA>	NA	NA	NA
## 54	<NA>	NA	NA	NA
## 55	Michelle.m.kim@berkeley.edu	15.0	NA	NA
## 56	miketp333@gmail.com	24.0	NA	NA
## 57	mramsay2017@gmail.com	17.0	NA	NA
## 58	mursil@gmail.com	25.0	NA	NA
## 59	nsorensen64@gmail.com	14.0	NA	NA
## 60	paigecarver@gmail.com	4.0	1	10
## 61	pdurkin@berkeley.edu	10.0	NA	NA
## 62	phat.t.doan@gmail.com	10.0	5	16
## 63	<NA>	NA	NA	NA
## 64	racheaedo@gmail.com	5.0	1	12
## 65	<NA>	NA	NA	NA
## 66	rutikasb@gmail.com	15.0	2	7
## 67	samibangerter1@gmail.com	15.0	6	17
## 68	Samman1108@gmail.com	15.0	NA	NA
## 69	Skelley@berkeley.edu	2.0	6	17
## 70	sarahsheppard@gmail.com	4.0	NA	NA
## 71	scott.george@berkeley.edu	30.0	1	1
## 72	<NA>	NA	NA	NA
## 73	<NA>	NA	NA	NA
## 74	sharadv@berkeley.edu	13.0	NA	NA
## 75	Sophia.a.chang@gmail.com	3.0	7	18
## 76	stefperez@sbcglobal.net	3.5	NA	NA
## 77	summergirl1992@gmail.com	48.0	NA	NA
## 78	<NA>	NA	NA	NA
## 79	wakystuf@gmail.com	13.0	1	8
## 80	whitney.sam@gmail.com	60.0	6	17
## 81	xiaokiiwii@gmail.com	16.0	NA	NA
##	netflixMin_P Treated Finished NA Assigned			
## 1	22	1	1	0
## 2	NA No	1	1	1
## 3	5 Yes	1	1	1
## 4	23	1	1	0
## 5	5 No	1	1	1
## 6	NA No	1	1	1
## 7	NA <NA>	0	NA	0
## 8	18 Yes	1	1	1
## 9	4 Yes	1	1	1
## 10	NA No	1	1	1
## 11	3	1	1	0

## 12	7	No	1 1	1
## 13	20	Yes	1 1	1
## 14	13	No	1 1	1
## 15	24	No	1 1	1
## 16	12		1 1	0
## 17	NA	<NA>	0 NA	0
## 18	NA	<NA>	0 NA	1
## 19	NA	<NA>	0 NA	0
## 20	9	No	1 1	1
## 21	NA		1 1	0
## 22	1	Yes	1 1	1
## 23	NA		1 1	0
## 24	NA		1 1	0
## 25	NA	<NA>	0 NA	1
## 26	NA		1 1	0
## 27	NA	Yes	1 1	1
## 28	12		1 1	0
## 29	25		1 1	0
## 30	NA	No	1 1	1
## 31	11		1 1	0
## 32	24		1 1	0
## 33	19	No	1 1	1
## 34	NA	<NA>	0 NA	1
## 35	10	No	1 1	1
## 36	NA	<NA>	0 NA	1
## 37	NA	<NA>	0 NA	1
## 38	22	No	1 1	1
## 39	NA	Yes	1 1	1
## 40	NA		1 1	0
## 41	NA	Yes	1 1	1
## 42	25	No	1 1	1
## 43	8		1 1	0
## 44	NA		1 1	0
## 45	15	Yes	1 1	1
## 46	NA		1 1	0
## 47	NA	No	1 1	1
## 48	NA	Yes	1 1	1
## 49	1		1 1	0
## 50	1	Yes	1 1	1
## 51	NA	<NA>	0 NA	1
## 52	2	No	1 1	1
## 53	NA	<NA>	0 NA	0
## 54	NA	<NA>	0 NA	0
## 55	NA		1 1	0
## 56	NA		1 1	0
## 57	NA		1 1	0
## 58	NA	No	1 1	1
## 59	NA		1 1	0
## 60	14	Yes	1 1	1
## 61	NA		1 1	0
## 62	24		1 1	0
## 63	NA	<NA>	0 NA	1
## 64	6		1 1	0
## 65	NA	<NA>	0 NA	0



```
## 66      21      No      1  1      1
## 67      25      Yes      1  1      1
## 68      NA      No      1  1      1
## 69      25      No      1  1      1
## 70      NA      NA      1  1      0
## 71      17      Yes      1  1      1
## 72      NA      <NA>      0 NA      0
## 73      NA      <NA>      0 NA      0
## 74      NA      NA      1  1      0
## 75      26      Yes      1  1      1
## 76      NA      NA      1  1      0
## 77      25      NA      1  1      0
## 78      NA      <NA>      0 NA      0
## 79      16      NA      1  1      0
## 80      25      No      1  1      1
## 81      NA      NA      1  1      0
```

```
#Post-survey questions
str(all_data)
```

```
## 'data.frame':   81 obs. of  30 variables:
## $ linkedEmail      : Factor w/ 81 levels "ace9312@gmail.com",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ enteredEmail     : Factor w/ 81 levels "ace9312@gmail.com",...: 1 2 3 4 5 17 6 7 8 18 ...
## $ gender           : Factor w/ 2 levels "male","female": 1 2 1 2 2 2 1 2 2 1 ...
## $ age              : Factor w/ 4 levels "21-","22-34",...: 2 1 3 4 2 2 2 2 2 2 ...
## $ region           : Factor w/ 6 levels "midwest","northeast",...: 5 5 2 5 2 2 5 5 5 1 ...
## $ employment       : Factor w/ 6 levels "full","part",...: 4 2 4 2 5 1 1 1 4 1 ...
## $ maritalStatus    : Factor w/ 3 levels "single","married",...: 2 1 3 2 2 2 2 1 1 1 ...
## $ children         : Factor w/ 2 levels "yes","no": 2 2 1 1 2 2 2 2 2 2 ...
## $ hoursTV          : num  10 5 15 3 5 10 4.5 12 8 14 ...
## $ binge            : Factor w/ 5 levels "once a week",...: 3 5 5 5 3 2 4 2 3 5 ...
## $ primaryChannel    : Factor w/ 7 levels "netflix","HBO",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ allMethods       : Factor w/ 36 levels "1","1,2","1,2,3,4",...: 25 24 13 18 23 18 11 9 24 19 ...
## $ moreTimeThanWanted : Factor w/ 5 levels "once a year",...: 4 5 5 2 2 3 4 4 5 4 ...
## $ watchAlone       : Factor w/ 2 levels "alone","withOthers": 2 2 1 2 2 1 1 2 1 1 ...
## $ shareProfile     : Factor w/ 2 levels "yes","no": 1 1 2 2 1 1 1 1 2 2 ...
## $ netflixDays      : Factor w/ 12 levels "", "0", "1", "2",...: 2 NA 5 1 3 11 NA 3 5 2 ...
## $ netflixHours     : Factor w/ 19 levels "", "0", "10", "11",...: 15 NA 8 1 10 19 NA 3 4 9 ...
## $ netflixMin       : Factor w/ 28 levels "", "10", "11", "17",...: 17 NA 21 1 4 27 NA 22 11 12 ..
## $ netflixAccountAndChrome: Factor w/ 4 levels "both","noNetflix",...: 1 2 1 3 1 1 1 1 1 1 ...
## $ source           : Factor w/ 4 levels "BYU","Friend",...: 1 1 4 3 3 2 4 4 3 3 ...
## $ treat            : num  0 1 1 0 1 1 0 1 1 1 ...
## $ post_survey_ID   : Factor w/ 65 levels "ace9312@gmail.com",...: 1 2 3 4 5 6 NA 7 8 9 ...
## $ reportedhrs_P    : num  20 10 20 14 9 10 NA 8 13 30 ...
## $ netflixDays_P    : num  1 NA 1 3 1 NA NA 1 1 NA ...
## $ netflixHours_P   : num  1 NA 6 15 11 NA NA 10 2 NA ...
## $ netflixMin_P     : num  22 NA 5 23 5 NA NA 18 4 NA ...
## $ Treated          : Factor w/ 3 levels "", "No", "Yes": 1 2 3 1 2 2 NA 3 3 2 ...
## $ Finished         : num  1 1 1 1 1 1 0 1 1 1 ...
## $ NA               : num  1 1 1 1 1 1 NA 1 1 1 ...
## $ Assigned         : num  0 1 1 0 1 1 0 1 1 1 ...
```

```
#Self-reported TV watching
```

```
t9 <- round(mean(all_data$reportedhrs[all_data$Assigned==1 & all_data$Finished==1], na.rm=T),2)
c9 <- round(mean(all_data$reportedhrs[all_data$Assigned==0 & all_data$Finished==1], na.rm=T),2)
```

```

diff9 <- t9-c9
se9 <- se_diff_means(as.numeric(complete.cases(all_data$reportedhrs[all_data$Assigned==1])),as.numeric(

#Netflix extension readings
all_data$NFLXreading <- (all_data$netflixDays_P*24) + all_data$netflixHours_P + (all_data$netflixMin_P/
t10 <- round(mean(all_data$NFLXreading[all_data$Assigned==1], na.rm=T),2)
c10 <- round(mean(all_data$NFLXreading[all_data$Assigned==0], na.rm=T),2)
diff10 <- t10-c10
se10 <- se_diff_means(as.numeric(complete.cases(all_data$NFLXreading[all_data$Assigned==1])),as.numeric(

#Received treatment email
#recode Yes = 1, No = 0, Blank/NA = NA
levels(all_data$Treated) <- c("Yes", "No", "")
all_data$Treated <- as.numeric(all_data$Treated)
all_data$Treated[is.na(all_data$Treated)] <- 0
all_data$Treated <- ifelse(all_data$Treated == 3, 1, 0)
all_data$Treated

## [1] 0 0 1 0 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0
## [36] 0 0 0 1 0 1 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0
## [71] 1 0 0 0 1 0 0 0 0 0 0

## % of assigned that were actually treated
compl <- round(mean(all_data$Treated[all_data$Assigned == 1]), 2) #compliance rate is low --> 36% of th
compl2 <- round(mean(all_data$Treated[(all_data$Assigned == 1) & (all_data$Finished == 1)]), 2) #actual
compl

## [1] 0.36

compl2

## [1] 0.43

# Put into a table for display
outcomes <- data.frame(variable = c("Self-ReportedTV", "NetflixExt"),
  control = c(c9, c10),
  treatment = c(t9, t10),
  diff = c(diff9, diff10),
  se = c(se9, se10))
knitr::kable(outcomes)

```

variable	control	treatment	diff	se
Self-ReportedTV	17.48	17.73	0.25	0.09
NetflixExt	70.39	67.16	-3.23	0.11

## Summary of initial findings

Why are the netflix readings so much bigger than the self-reported #s?

```

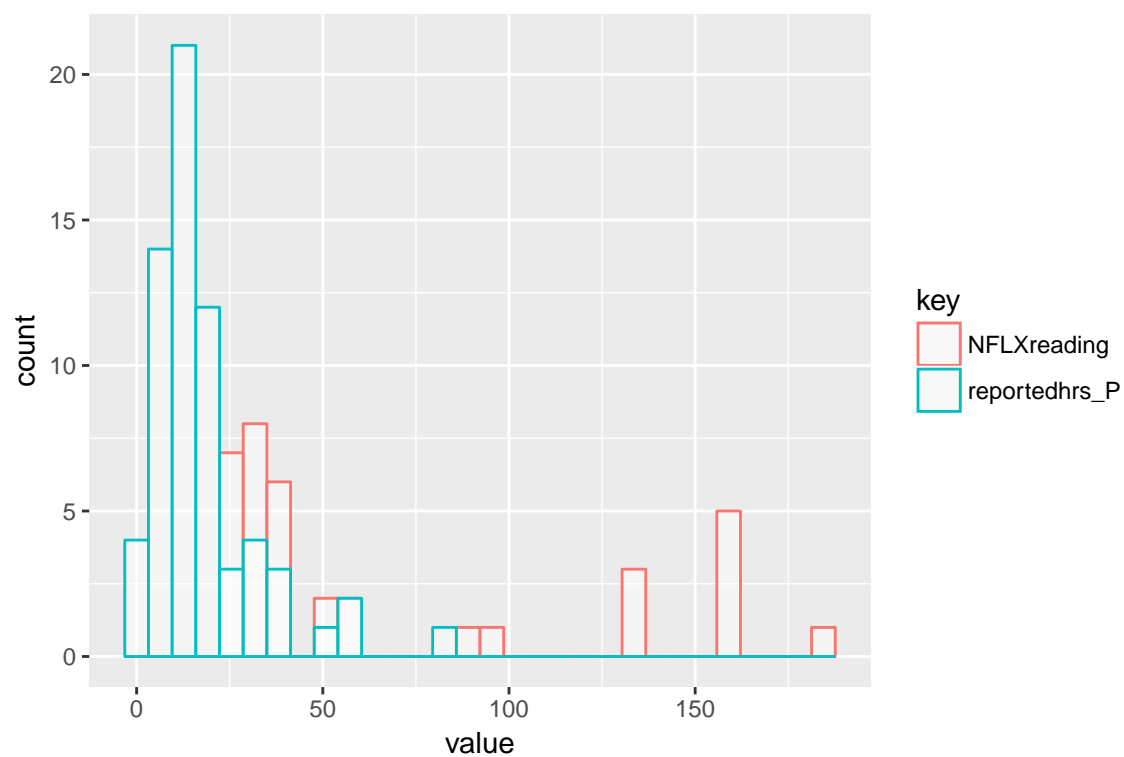
hours_watched <- all_data[ , c("reportedhrs_P", "NFLXreading")]

t <- hours_watched %>%
  gather(key, value)

ggplot(t, aes(x = value, color = key)) +

```

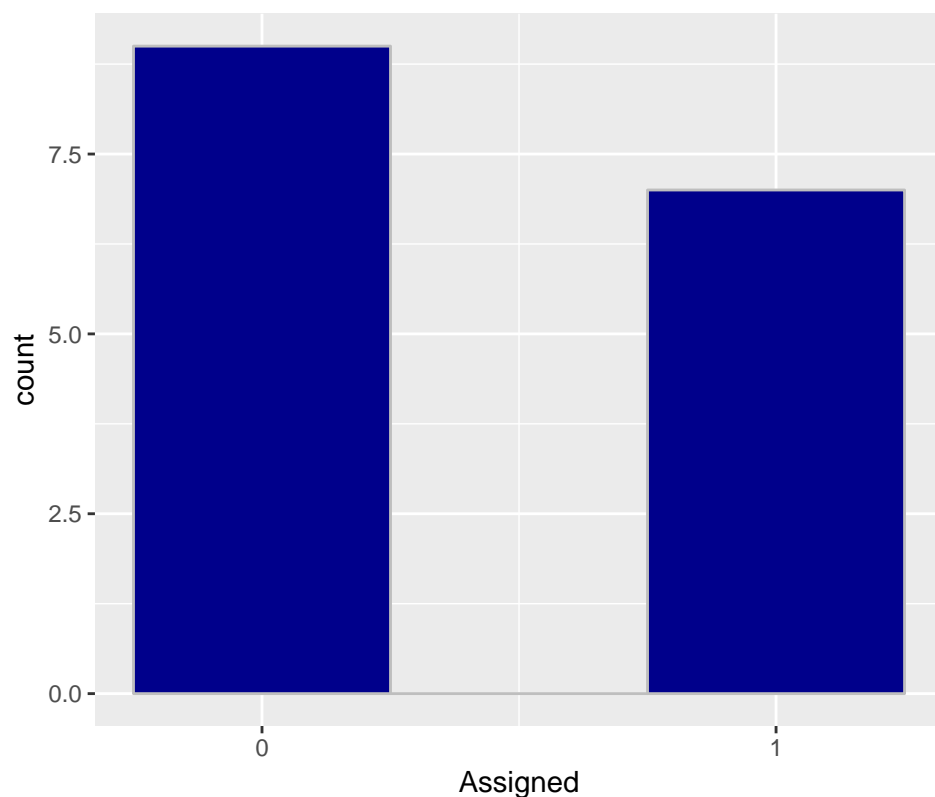
```
geom_histogram(fill="white", alpha=0.5, position="identity")
```



### Attrition

```
attrit_cases <- all_data[all_data$Finished == 0 ,]
ggplot(attrit_cases, aes (x=Assigned)) +
  geom_histogram(binwidth = 0.5, fill="darkblue", col="grey") +
  scale_x_continuous(breaks = seq(0,1,1)) +
  ggtitle("Comparing attrition for control(0) vs. treatment (1)")
```

## Comparing attrition for control(0) vs. treatment (1)



```
#percent attrition by assignment
xtabs(~ Finished + Assigned, data=all_data)

##          Assigned
## Finished  0    1
##          0  9  7
##          1 30 35

round(prop.table(xtabs(~ Finished + Assigned, data=all_data)),2)

##          Assigned
## Finished    0    1
##          0 0.11 0.09
##          1 0.37 0.43

attrit_by_assigned <- lm(Finished ~ Assigned, data=all_data)
summary(attrit_by_assigned)

##
## Call:
## lm(formula = Finished ~ Assigned, data = all_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8333  0.1667  0.1667  0.2308  0.2308
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.76923    0.06435  11.955  <2e-16 ***
```

```
## Assigned      0.06410    0.08936    0.717    0.475
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4018 on 79 degrees of freedom
## Multiple R-squared:  0.006472,    Adjusted R-squared:  -0.006104
## F-statistic: 0.5146 on 1 and 79 DF,  p-value: 0.4753

attrit_covariate_check <- lm(Assigned ~ gender + age + region + employment + maritalStatus + children +
summary(attrit_covariate_check)

##
## Call:
## lm(formula = Assigned ~ gender + age + region + employment +
##      maritalStatus + children + hoursTV + binge + primaryChannel +
##      moreTimeThanWanted + watchAlone + shareProfile + source,
##      data = all_data[all_data$Finished == 1, ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7925 -0.2648  0.0000  0.1926  0.7656
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   0.780280   0.742524   1.051
## genderfemale                   0.011888   0.222420   0.053
## age22-34                      -0.730372   0.347550  -2.101
## age35-44                      -0.617413   0.501151  -1.232
## age45-54                      -1.488938   0.612322  -2.432
## regionnortheast               -0.088633   0.443604  -0.200
## regionsoutheast              -1.473855   0.716649  -2.057
## regionsouthwest              -0.249889   0.976121  -0.256
## regionwest                   -0.099272   0.418580  -0.237
## regionoutsideUS              -0.368271   0.788755  -0.467
## employmentpart                0.230664   0.406877   0.567
## employmentstudent            -0.067821   0.366886  -0.185
## employmenthomemaker          0.616241   0.701882   0.878
## employmentself               0.306065   0.533251   0.574
## maritalStatusmarried         0.357057   0.214726   1.663
## maritalStatusdivorced        0.122356   0.851637   0.144
## childrenno                   0.108830   0.296727   0.367
## hoursTV                      0.010582   0.019846   0.533
## bingeonce a month            -0.104093   0.400094  -0.260
## bingeonce every couple months 0.014922   0.398400   0.037
## bingeonce a year             -0.261701   0.442231  -0.592
## bingenno                     0.100424   0.411776   0.244
## primaryChannelHBO            -0.181192   0.689853  -0.263
## primaryChannelhulu           -0.017222   0.332850  -0.052
## primaryChannelamazon         0.994517   0.740814   1.342
## primaryChannelyoutube        -0.004435   0.445836  -0.010
## primaryChannelcable          0.142984   0.539320   0.265
## primaryChannelother          0.412018   0.328701   1.253
## moreTimeThanWantedcouple times a year -0.079962   0.337992  -0.237
## moreTimeThanWantedonce a month 0.146948   0.351999   0.417
```

```

## moreTimeThanWantedcouple times a month 0.156936 0.342818 0.458
## moreTimeThanWantedonce a week 0.021999 0.401548 0.055
## watchAlonewithOthers -0.266044 0.218361 -1.218
## shareProfileno 0.117118 0.374235 0.313
## sourceFriend 0.062644 0.504726 0.124
## sourceOnline (social media) 0.222885 0.332739 0.670
## sourceUC Berkeley Slack 0.472965 0.396673 1.192
## Pr(>|t|)
## (Intercept) 0.3023
## genderfemale 0.9578
## age22-34 0.0447 *
## age35-44 0.2282
## age45-54 0.0217 *
## regionnortheast 0.8431
## regionsoutheast 0.0492 *
## regionsouthwest 0.7998
## regionwest 0.8143
## regionoutsideUS 0.6442
## employmentpart 0.5753
## employmentstudent 0.8547
## employmenthomemaker 0.3874
## employmentself 0.5706
## maritalStatusmarried 0.1075
## maritalStatusdivorced 0.8868
## childrenno 0.7165
## hoursTV 0.5981
## bingeonce a month 0.7966
## bingeonce every couple months 0.9704
## bingeonce a year 0.5587
## bingenno 0.8091
## primaryChannelHBO 0.7947
## primaryChannelhulu 0.9591
## primaryChannelamazon 0.1902
## primaryChannelyoutube 0.9921
## primaryChannelcable 0.7929
## primaryChannelother 0.2204
## moreTimeThanWantedcouple times a year 0.8147
## moreTimeThanWantedonce a month 0.6795
## moreTimeThanWantedcouple times a month 0.6506
## moreTimeThanWantedonce a week 0.9567
## watchAlonewithOthers 0.2333
## shareProfileno 0.7566
## sourceFriend 0.9021
## sourceOnline (social media) 0.5084
## sourceUC Berkeley Slack 0.2431
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5406 on 28 degrees of freedom
## Multiple R-squared: 0.4935, Adjusted R-squared: -0.1577
## F-statistic: 0.7578 on 36 and 28 DF, p-value: 0.7853

```

## Model Building

### \*Findings

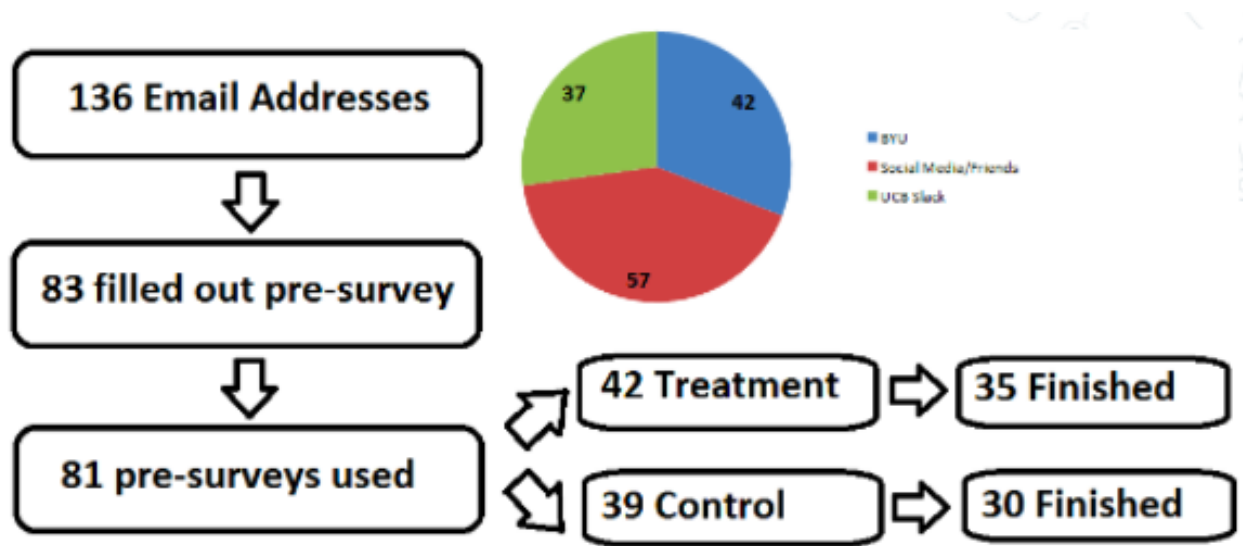


Figure 1: Sample Treatment email

## Appendix

*The pre-survey in the pilot consisted of the following questions:*

1. Provide us with an email address that will be used to contact you throughout the course of this survey.
2. Are you male or female?
3. How old are you?
4. What region do you currently live in?
5. What is your current employment status?
6. What is your marital status
7. Do you have children under the age of 18?
8. How much television do you watch on a weekly basis? This includes cable, online streaming, Youtube, and can consist of movies, tv shows, and video clips.
9. Do you tend to binge watch TV shows (more than 4 hours of consecutive TV watching)?
10. What is your primary method of watching television?
11. Check off all the ways in which you consume television - Netflix, HBO, Hulu, Amazon, Youtube, Cable, Other.
12. How often do you find that after watching television you spent more time on it than you would have liked to?
13. Do you tend to watch television alone or with others?
14. Do you share your Netflix profile with someone else?
15. A link to installing the Netflix Chrome instruction as well as a text box for inputting what the extension outputs as usage over the last two week.

### Summary of pilot pre-survey takeaways:

1. 8 of 9 individuals filled out the pre-survey in the designated time. Given the quick turnaround time (2 days), we expected some level of non-response as work, travel, etc can get in the way.
2. of the 8 individuals who did respond, 5 were Female and 3 male; 7 of the 8 were in the 22-34 age group, 7 of the 8 were in the Northeast region, 7 of 8 were full-time employed; 6 of 8 were married and 2 were single; and none of the 8 individuals had children under the age of 18.
3. At the mid-point of the ranges, individuals in the pilot reported watching 7.75 hours of TV a week.

**CONTROL EMAIL BODY:**

Thank you for completing the pre-survey to enter into our study. We really appreciate you taking the time out of your day to help us out.

As a reminder, we will be sending a final brief survey on April 2nd asking you to report your amount of television watched during the study period. If you are using the extension, it tracks Netflix usage regardless of the medium - whether you watch it on a smart TV, your phone, or a computer, it will show up. Fill out that final survey by April 6th and we will enter you into the drawing for five \$20 Amazon gift cards!

Thanks again,  
Alyssa, Cameron, and Sarah

Figure 2: Sample Control email

Netflix was the primary method of watching TV for half of the individuals while cable and other were the primary method for the other half while 7 of the 8 individuals said they consumed television through Netflix.

4. 3 of 8 individuals were unable to install the extension either because they don't have a Netflix account or they don't use Chrome as their web browser. Of the individuals that did input a reading from the extension, the average Netflix consumption over the prior two weeks was 787.2 minutes or just over 13 hours of TV.

Over the next couple of days individuals randomly assigned to the treatment group were sent an email with a bar chart showing peer consumption trends while those in the control group did not receive any email. Then all 8 individuals were asked to fill out a *post-survey* on how much TV they watched during the pilot period.

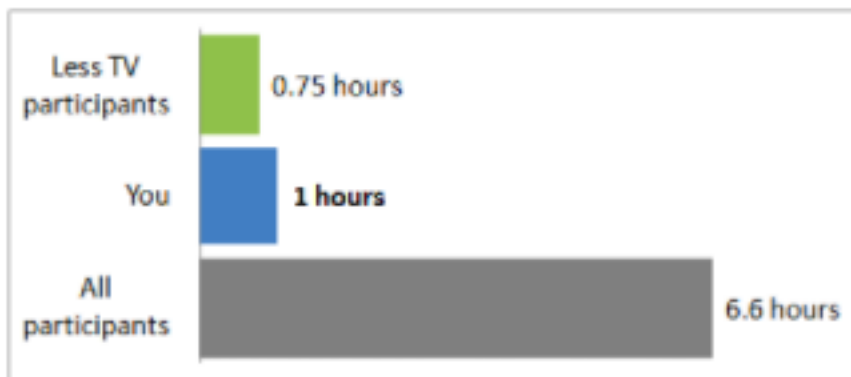
*The post-survey in the pilot consisted of the following questions:*

1. Provide us with your email address. This is the same identifier you used on the first survey and how we have been contacting you.
2. How much television did you watch during the study period from Feb 28 through March 1, inclusive? This includes cable, online streaming, and YouTube, and can consist of movies, tv shows, and video clips.
3. Input the time reading from your Chrome Netflix extension.



**TREATMENT EMAIL BODY (V1):**

Thank you for completing the pre-survey to enter into our study. We really appreciate you taking the time out of your day to help us out, and wanted to share some interesting tidbits about TV watching habits from our data collection so far.



As a reminder, we will be sending a final brief survey on April 2nd asking you to report your amount of television watched during the study period. If you are using the extension, it tracks Netflix usage regardless of the medium - whether you watch it on a smart TV, your phone, or a computer, it will show up. Fill out that final survey by April 6th and we will enter you into the drawing for five \$20 Amazon gift cards!

Thanks again,  
Alyssa, Cameron, and Sarah

Figure 3: Sample Treatment email