

# News Articles and Happiness

Mihir Sathe, Rahul Vaswani, Saurabh Jaju, Sophia Huang

## Contents

0.1	Abstract . . . . .	1
0.2	Introduction . . . . .	1
0.3	Experiment Design . . . . .	2
0.4	Data Anlaysis . . . . .	6
0.5	Discussion of Results and Inference . . . . .	15
0.6	Limitations . . . . .	16
0.7	Conclusion . . . . .	16
0.8	Appendix . . . . .	17

## 0.1 Abstract

*We tested our hypothesis that news articles can influence a person’s happiness by conducting a field experiment. Each day we randomly assigned subjects into control and treatment groups. An article would be sent out at 6 AM PST either by email or text, and the subject would answer baseline questions and a Likert scale to establish mood. After the initial survey, the subject would be required to read a news article with either a positive theme (treatment) or a neutral theme (control). A follow-up survey was sent out at 11 AM PST and the subject would be required to answer other survey questions as well as fill in a Likert scale to record their current mood. We analyzed the results by subtracting the prior mood by the baseline mood in order to establish an outcome variable of **Mood Difference**. Randomization inference and a two sample Mann-Whitney U Test were conducted to calculate an ATE and record any significance. Ultimately, we failed to reject the sharp null hypothesis that news articles do not effect a subjects’ mood.*

## 0.2 Introduction

In 2014, Adam Kramer, Jamie Guillory, and Jeffrey Hancock published a paper called Experimental evidence of massive-scale emotional contagion through social networks. In this study, Kramer and others experimented on Facebook users by using sentiment based manipulation on a users Newsfeed to attempt to manipulate a user’s mood. The outcome of this experiment caused some of Facebook’s users to become happier, while inversely causing other users to fall into a depressive like state. While we disagree with almost everything about why or how this experiment was conducted and the effects on its participants, it did allow us to question if we can influence a person;s happiness.

The way an event influences a person’s happiness has been a topic of research in social science for a long time. Defining and measuring happiness is a subjective and difficult problem; however, a person’s happiness heavily affects not just the person but the surroundings of a person as well. One theory is that a happy person could be highly efficient employee with good team work and another theory is that a happy person might vote in a particular way and buy a particular set of items. We believe that a happier population would benefit society as a whole.

This belief caused us researchers to formulate a question – **Can we make someone happier through a mass scale medium? To what extent does reading positive news affect one’s mood?**

We chose to conduct an experiment in order to help understand and answer this question. First, we needed to define our “mass scale medium.” Ultimately, we settled on a medium that the majority of Americans consume each morning, every day: news articles. Thus, if an event as minute as reading a certain news article can influence a person’s happiness then it can be useful to economists, psychologists, political parties, employers

and marketers to measure and induce certain types of actions and decisions taken by people. Our hypothesis is that an exposure to a positive news article would alter a person’s mood.

$$H_0 : \mu_{\text{Mood Difference}} = 0$$

$$H_1 : \mu_{\text{Mood Difference}} \neq 0$$

In this paper we are presenting the results of an experiment conducted through the surveys administered to the subjects both before and after asking them to read a news article early in the morning. After researching the various approaches to measuring true happiness such as Roger Tourangeau et. al have described in their paper, we have taken a more direct approach of directly asking the subjects how did they feel immediately before reading a news article and then surveying their mood again a few hours after the initial survey. In addition to collecting a subjects’ mood, we also collect more data regarding their sleep, caffeine consumption, and whether they performed an act of kindness such as helping another person in the day. These extra data points were collected to factor in possible covariates and to verify the data regarding happiness the subjects had provided. We further expand on the experiment in the Experiment Design section.

### 0.3 Experiment Design

#### 0.3.1 Initial Design

As discussed earlier, happiness (or any other emotion) is difficult to measure. We decided to split the subjects into two groups, treatment and control, and administer a “happy” news article to the treatment group and a “neutral” news article to the control group. Our source of happy news articles was Reddit’s Uplifting News subreddit. Filtering on the most upvoted news articles of the month allowed us to get a consensus on what a group of people (redditors) deem to be the “most uplifting news”. The control group’s news article would be sourced from sites such as National Geographic and Scientific American. We chose these sources in an attempt to avoid sending negative articles such as political news. We proofread the articles and developed a data bank of questions.

Our initial experimental design consisted of sending participants two surveys each day which would allow us to gauge mood in the morning prior to reading the article, and a few hours later in the day. A single question would be emailed to the subject at 6 AM: How do you feel right now? The question would be answered with a numeric scale with 1 being “Sad” and 10 being “Happy”. This first survey would give us a baseline mood. We would then send out a link to a news article and have the subject read this article. At around 5 PM, we would send a follow-up survey with the same question and answer choices in addition to questions such as, “If you feel sad, please write a brief sentence on why you feel this way?”. The original belief was that we could parse out responses from the text answers and look for keywords that could be confounding variables. Understandably, a Likert scale can yield quite subjective results; however, given the constraints of the experiment, we deemed it our best option. The experiment would have lasted 6 days and we would be able to hopefully see some sort of treatment effect.

#### 0.3.2 First Experiment Re-Design

There were many problems with our initial design. The 11 hours between the surveys was too long to measure the effect of reading a positive article. The longer this gap was, the more potential influencers there could be on a person’s mood throughout the day, and if we waited that long we would capture a lot of noise in our data. Additionally, we realized that having subjects physically type out answers may cause attrition. SurveyMonkey wrote in a blog post that a person is willing to spend 1 minute and 15 seconds on a one question survey and a two question survey should take around 2 minutes. If people were forced to think about why they were sad and then type out their response, it may take longer and then we would face drop-offs in our participation rate. We needed to find another way to measure covariate effects. A solution to this problem was to address covariate collection via multiple choice questions. Thus we decided to ask subjects:

- How much water did you drink today?
- How much caffeine did you drink today?
- Did you help someone today?
- Did you purchase a non-food item today?

The later two questions were taken from Tourangeau’s paper, while the water and caffeine question helped disguise our true treatment measure so that we could reduce answer bias if a subject caught on to what we were trying to measure. Additionally, we decided to include a summary of the article along with the link. This would allow participants who did not want to click the article to have a quick way to get the premise of the article we were sending out. Lastly, we theorized that we would have high attrition if we sent out three different emails to subjects, since people might interpret our experiment as spam, so we chose to reduce our means of communications to two emails a day and include the article as a second page in the first survey. Note that we had not figured out a way to deal with noncompliance at this point in our initial survey design; e.g, did a subject actually read the news article (or even the summary of the article)? We discuss how we dealt with this after our initial pilot experiment.

### 0.3.3 Recruiting Subjects

In order to have an effective experiment, we decided to reach out to friends, coworkers, and family members. We decided to forego tools such as Amazon’s Mechanical Turk and Qualtrics survey responders in favor of having some sort of connection with our subjects. Incentives of \$100 gift card raffles and Google Home raffles were used in order to get people to sign up for the experiment. Below is an example of a Slack message sent out to one of UC Berkeley MIDS channels:

Hi All, apologies in advance for the impending spam coming from all the W241 groups in Slack. My team is looking for participants to sign up for our W241 Experiment. The experiment will last for 6 days. Each morning you’ll receive a survey, a brief article, and a follow-up survey later in the afternoon. At a minimum, you can read the article title and fill the 2 surveys each day (which should take <1 minute each). The article and surveys will be sent via email or text. We will be raffling off prizes of gift cards and a Google Home for participation. Thanks! Signup link: [https://berkeley.qualtrics.com/jfe/form/SV\\_2lgkG3WevFgGDMF](https://berkeley.qualtrics.com/jfe/form/SV_2lgkG3WevFgGDMF)

Our signup form asked for the following information:

- Name
- Email
- Phone number
- Would you prefer to receive a text or email?
  - Text/Email
- Gender
  - Male/Female/Prefer not to say
- Age Range
  - 25 and under/26-30/.../60 and over/Prefer not to say
- Employment Status
  - Employed/Self-Employed/Student/Retired/Unemployed/Other
  - A participant could choose multiple options for this question.

Questions for gender, age range, and employment status were included in case we needed to use any blocking or clustering on our subjects. This is discussed in more detail in the Descriptive Analysis section.

### 0.3.4 Randomizing Subjects

We needed to maximize our sign ups in order to have high statistical power (since statistical power increases as the square root of sample size increases). Signups were difficult to obtain because the duration of the experiment was quite long, but we managed to get 50 people to sign up, which was lower than our initial

expectation. In order to increase statistical power and generate more observations, we adjusted our design to treat each day like a new experiment. **Each day we randomly assigned subjects to treatment and control and we collected the subject’s baseline mood during our first survey in order to control for any spillover effect. This process of randomly assigning subjects to control and treatment each day and then gathering the baseline mood for the day allowed us to treat each subject as a ‘new’ subject (establishing independence) for each day of the survey** (thus allowing us to have 50\*4 observations).

Additionally, upon looking at the sign up distribution, we had a fairly equal amount of women and men sign up for the survey; however, the majority of these subjects were between the ages of 26-35 and were also employed. **Because of the ambiguous boundries between the data and the sparseness of subjects, we chose not to conduct and blocking or clustering** in our experiment.

### 0.3.5 Technology

Each day, a python script would randomize our subjects into test and control groups. Since our experiment required emails or text messages sent at exact times, we implemented a scheduler (Boomerang) to send both the initial and post survey emails. For text messages, we purchased a Berkeley area code phone number from Twilio and utilized the Twilio python API to automatically send text messages at our designated times. We used Qualtrics to create the surveys and record the results. Now that our experiment was set up, we conducted a small pilot study to work out potential implementation errors the week before our actual study.

### 0.3.6 Articles Used

#### 0.3.6.1 Treatment

Below are articles used for the treatment group (uplifting articles):

- Dave Grohl Delivers Food to First Responders Fighting California Wildfires
- Powerball winner, a single mom, shares her wealth with wounded veterans
- Garbageman insists on finishing his route during Camp Fire so he can check on elderly, evacuates a 93 year old woman
- Batkid Is Now Cancer Free After Stealing Everyone’s Hearts In San Francisco

#### 0.3.6.2 Control

Below are articles used for the control group. We manually selected these articles looking for neutral articles that were not politically charged:

- Why Don’t We Forget How to Ride a Bike?
- How cat tongues work—and can inspire human tech
- Amazon is not too big to fail, and one day Amazon will fail, says Amazon CEO Jeff Bezos
- Not all carbs are bad: Study shows high-carb diets can promote healthy brain aging

### 0.3.7 Pilot Experiment

We conducted a pilot experiment on six people for a duration of two days. This would allow us to see how exhausted a subject would become after being exposed to 2 surveys a day for a consecutive two-day period. We began each day with the Initial Survey at 6 AM. This survey had the following questions:

- Name
- As of right now, have you had more than 2 glasses of water today?
- As of right now, have you had any caffeine today?
- As of right now, how many news articles have you read today?

- As of right now, what is your current mood?
- How many hours did you sleep last night?

The next page of the survey consisted of the article for treatment and control used for Day 1. Please refer to the appendix to see a list of articles used.

At 10 AM we sent a follow-up survey with the following questions:

- Name
- As of right now, have you had more than 2 glasses of water today?
- As of right now, how many cups of caffeine have you had?
- As of right now, how many news articles have you read today?
- As of right now, what is your current mood?
- Have you bought anything other than food today?
- Did you help someone today?

### 0.3.8 Pilot Feedback

We gathered feedback from our pilot study subjects. They answered all the questions and submitted responses for both days; however, they did express that if they were to receive this survey for 6 days consecutively, it would be exhausting. We decided that we would have sufficient statistical power with 200 (50x4 days) observations and thus this may mitigate attrition during our study, so we reduced the experiment to 4 days. The pilot subjects mentioned that it was too easy to skip reading the article (or even the summary) and just take the post survey, which would result in noncompliance. **We added an easy question about the news article as a 3rd page of the 6am survey to test whether each subject read the assigned article. This allowed us to control for noncompliance in case participants were just clicking through to complete the survey without reading questions.** At this point, we felt that we had too many questions on the early morning survey so we removed the water question. We noticed during the pilot study that one subject did not answer the initial survey the first day until 10am, when the second survey was pushed out to them. We pushed out the second survey to 11am to give sufficient time to answer the first survey each day.

### 0.3.9 Final Experiment

After taking the feedback of the pilot experiment, we informed our participants that we were reducing the time of the experiment to four days. We sent an email briefly describing the experiment half a week before the start day, and then we followed up with a reminder email the night before. The following day, we had our experiments sent out via email and text.

The initial email and text at 6 AM consisted of the following questions:

- Name
- As of right now, have you had any caffeine today?
- As of right now, how many news articles have you read today?
- As of right now, what is your current mood?
- How many hours did you sleep last night?
- **NEW PAGE**
- News Article
- **NEW PAGE**
- Question about news article.

The post-survey at 11 AM consisted of the following questions:

- Name
- As of right now, how many cups of caffeine have you had?
- As of right now, how many news articles have you read today?

- As of right now, what is your current mood?
- Have you bought anything other than food today?
- Did you help someone today?

Visuals of the experiment are included in the index.

## 0.4 Data Analysis

```
# Load in Data
## Raw data available in the data folder;
## however, we are importing in a cleaned and joined dataframe.
library(WilcoxCV)
d = read.csv('data/finalDF.csv')
```

### 0.4.1 Exploratory Data Analysis

**Sample:** We had 59 signups for our experiment in total, of which 52 responded to at least one survey. 49 participants completed the full 4-day experiment. Though we had attrition for these 9 cases, we do not believe that this biased our results since our randomization controlled for this attrition by having participants randomized into control or treatment each day. We dropped observations where outcome was missing.

If a subject failed to complete a survey, we removed only that day from our results. Due to randomization and independence of each day, the subject was still counted on days when they filled the pre and post surveys.

```
d = d[is.na(d$MoodDiff) == FALSE, ]
```

The remaining analysis focuses on the 52 respondents that submit at least one survey. 47.1% (n= 24) were female and 52.9% (n=27) were male. 52.9% (n= 27) were under 30-year-old and 47.1% (n= 25) were in the age group of 30 years old or over. Below we examine the distributions of gender and age. Note that we counted each day's response as an independent observation, so our final dataset consists of 202 observations, from 52 subjects over the 4 days.

**Variables:** The outcome variable of the study was the difference in the mood of participants between the first survey, which was sent in the morning and taken prior to reading the article, and the second survey which was given 5 hours after the first survey and article were sent. The independent variables includes dummy variables of gender, age range, and the treatment.

Our table below shows the distribution of the difference in mood for the control and treatment group:

Group	Felt less happy	No Change	Felt Happier
Control	19 (20.7%)	55 (59.8%)	18 (19.5%)
Treatment	12 (16.5%)	51 (56.4%)	33 (34.4%)

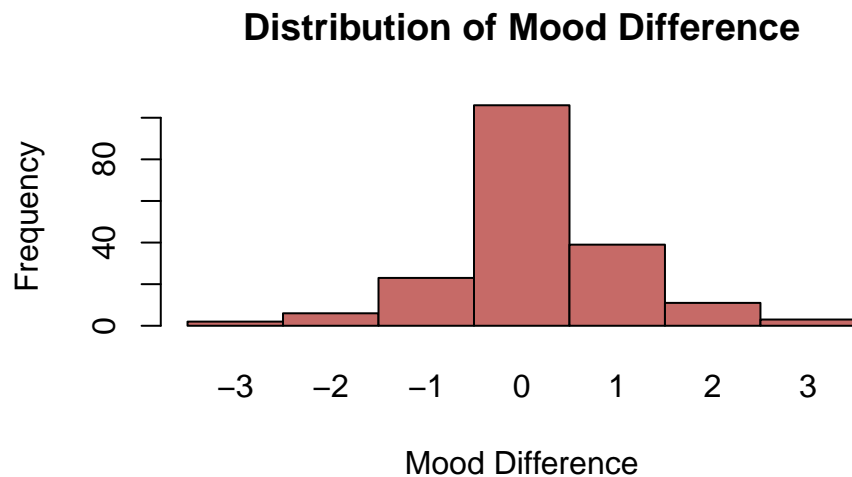
We explore the distribution of the outcome variable and a select number of covariates:

**Outcome:**

```
int.hist = function(x,ylab="Frequency",...) {
  barplot(table(factor(x)),space=0,ylab=ylab,...);
}

int.hist(d$MoodDiff,
  main="Distribution of Mood Difference",
  xlab="Mood Difference",
```

```
col = "#c66a67")
```



The distribution table above displays that our outcome is fairly normally distributed with a mean of 0.1526316 and a standard deviation of 0.9500033. The negative numbers represent a negative mood change, while the positive numbers represent a change toward happiness. Zero represents no change. We also note that the numbers are ranked and discrete. The difference in the distributions between treatment and control is examined in the Average Treatment Effect Analysis section.

#### Covariates:

*Note: The graphs below denote subjects in the treatment group as yellow and subjects in the control as blue. If there is any overlap in the distribution, it will be green.*

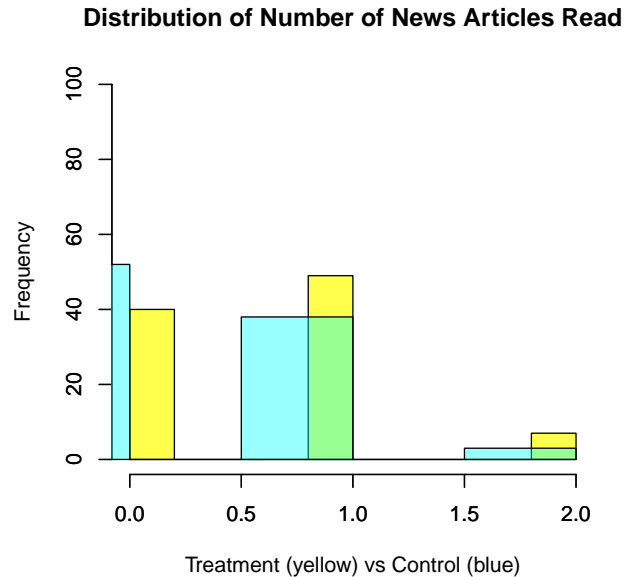
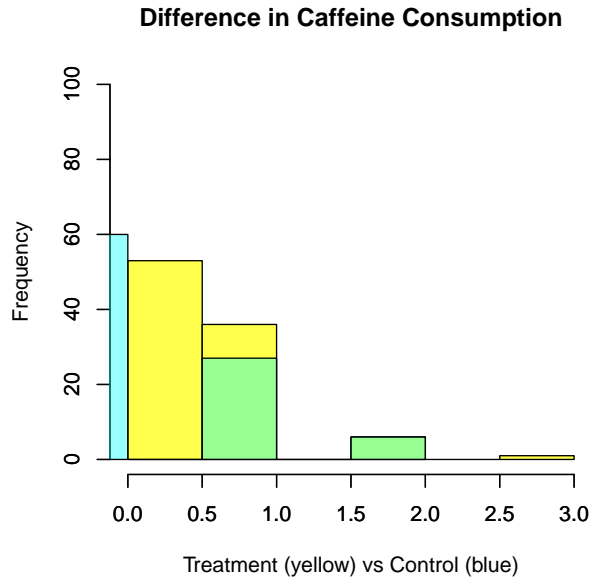
```
par(mfrow=c(1,2))
hist(d[d$Treatment == 1,]$coffee_dif,
     xlim=c(0,3),
     ylim=c(0,100),
     breaks=10,
     col=rgb(1,1,0,0.7),
     main="Difference in Caffeine Consumption",
     xlab="Treatment (yellow) vs Control (blue)")

par(new=TRUE)
hist(d[d$Treatment == 0,]$coffee_dif,
     xlim=c(0,3),
     ylim=c(0,100),
     breaks=10,
     col=rgb(0,1,1,0.4),
     main="", xlab="", ylab="")

hist(d[d$Treatment == 1,]$news_articles_dif,
     xlim=c(0,2),
     ylim=c(0,100),
     breaks=10,
     col=rgb(1,1,0,0.7),
     main="Distribution of Number of News Articles Read",
     xlab="Treatment (yellow) vs Control (blue)")

par(new=TRUE)
hist(d[d$Treatment == 0,]$news_articles_dif,
```

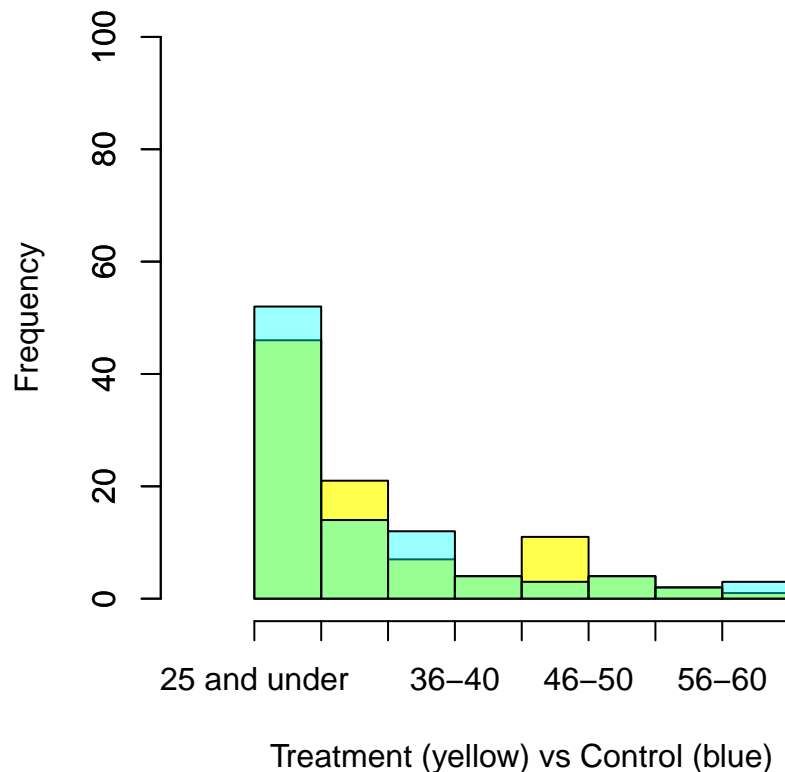
```
xlim=c(0,2),
ylim=c(0,100),
breaks=10,
col=rgb(0,1,1,0.4),
main="",xlab="",ylab="")
```



```
hist(as.numeric(d[d$Treatment == 1,]$Age.Range),
     xlim=c(0,10),
     ylim=c(0,100),
     breaks=10,
     col=rgb(1,1,0,0.7),
     main="Distribution of Age Ranges",
     xlab="Treatment (yellow) vs Control (blue)",
     xaxt = "n")
par(new=TRUE)
hist(as.numeric(d[d$Treatment == 0,]$Age.Range),
     xlim=c(0,10),
     ylim=c(0,100),
     breaks=10,
     col=rgb(0,1,1,0.4),
     main="",
     xlab="",ylab="", xaxt = "n")
axis(1, at=1:length(unique(d$Ag)), labels=levels(unique(d$Age.Range)))
```



## Distribution of Age Ranges



We have examined and displayed the distribution of the difference between the post survey and pre survey for coffee consumption and news articles read in both treatment and control to show that the covariates have been evenly distributed between treatment and control as a part of our randomization.

There is an anomaly in the `news_articles_difference` distribution – a negative number. This means that someone might have accidentally marked that they read more news articles in the morning than they had cumulatively read for the entire day. We choose to keep this data point in our analysis, but make note that it may affect certain tests.

We perform the Mann-Whitney U Test to check for how similar the distributions are between treatment and control in order to check for covariate balance.

```
wilcox.test(coffee_dif~Treatment, mu=0, paired=FALSE, conf.int=TRUE, data = d)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  coffee_dif by Treatment
## W = 4059, p-value = 0.1688
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -8.310116e-05  4.729328e-05
## sample estimates:
## difference in location
## -4.847271e-05
```

```
wilcox.test(news_articles_dif~Treatment, mu=0, paired=FALSE, conf.int=TRUE, data = d)
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
##
## data: news_articles_dif by Treatment
## W = 3768.5, p-value = 0.02744
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -2.630071e-05 -3.889115e-05
## sample estimates:
## difference in location
## -4.715592e-05
```

The above tests allow us to fail to reject the null hypothesis that the distributions for treatment and control are similar for caffeine consumption which means we have an similar representation of subjects for treatment and control. Interestingly, we reject the null hypothesis for news article difference which implies that there is some difference between treatment and control in terms of how many news articles they have read. This could be caused by subjects continuing to read news articles after we sent them the initial article or due to some other confounds. We will need a better way to measure this for future experiments.

Additionally, we examine the distribution in age range. In the histogram, we note that there is a disproportionate amount of subjects that are 35 years old or under. We ultimately classify these as “millenials” and will do further analysis on this group after our primary result evaluations. Conducting a Mann-Whitney U Test results in:

```
wilcox.test(as.numeric(d$Age.Range)~Treatment, mu=0, paired=FALSE, conf.int=TRUE, data = d)

##
## Wilcoxon rank sum test with continuity correction
##
## data: as.numeric(d$Age.Range) by Treatment
## W = 4155, p-value = 0.3259
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -9.999650e-01 4.030071e-05
## sample estimates:
## difference in location
## -4.015943e-05
```

which allow us to fail to reject the null hypothesis.

## 0.4.2 Average Treatment Effect Analysis

### 0.4.2.1 Randomization Inference

Since our outcome is a categorical variable, a linear regression would not allow us to gain an accurate assesment of the Average Treatment Effect. Thus, in order for us to ascertain levels of significance and perform analysis, we choose to conduct randomization inference. We will conduct 10,000 simulated random assignments under the sharp null hypothesis to assess whether our ATE from the sample is significant.

First, we calculate the experiment’s ATE:

```
library("data.table")
d = data.table(d)
sample_ate = d[Treatment == 1, mean(MoodDiff),] - d[Treatment == 0, mean(MoodDiff),]
sample_ate

## [1] 0.1968085
```

We have an Average Treatment Effect of 0.1968 from our experiment. In order to see the significance of this value, we conduct randomization inference:

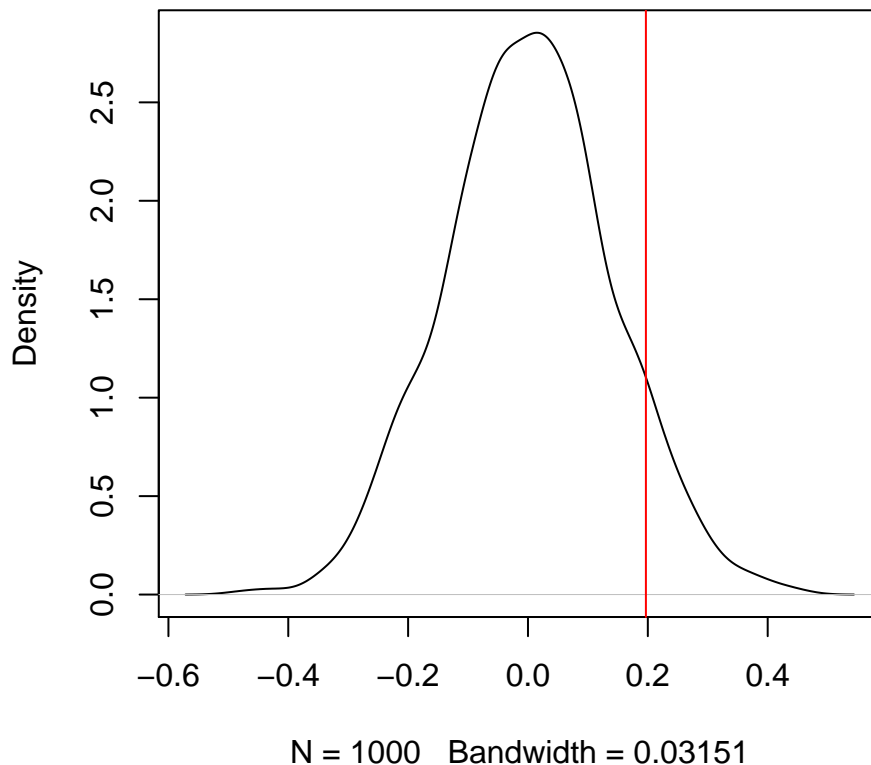
```
randomize = function() sample(c(rep(0,nrow(d[Treatment == 0])),rep(1,nrow(d[Treatment == 1]))))

est.ate = function() {
  d$treat_rand = randomize()
  mean(d[treat_rand == 1, MoodDiff]) - mean(d[treat_rand == 0, MoodDiff])
}

distribution.under.sharp.null = replicate(1000, est.ate())

plot(density(distribution.under.sharp.null),
     main = "Randomization Inference Results")
abline(v=sample_ate, col = "red")
```

## Randomization Inference Results



```
p_val = mean(sample_ate < distribution.under.sharp.null)
p_val
```

```
## [1] 0.069
```

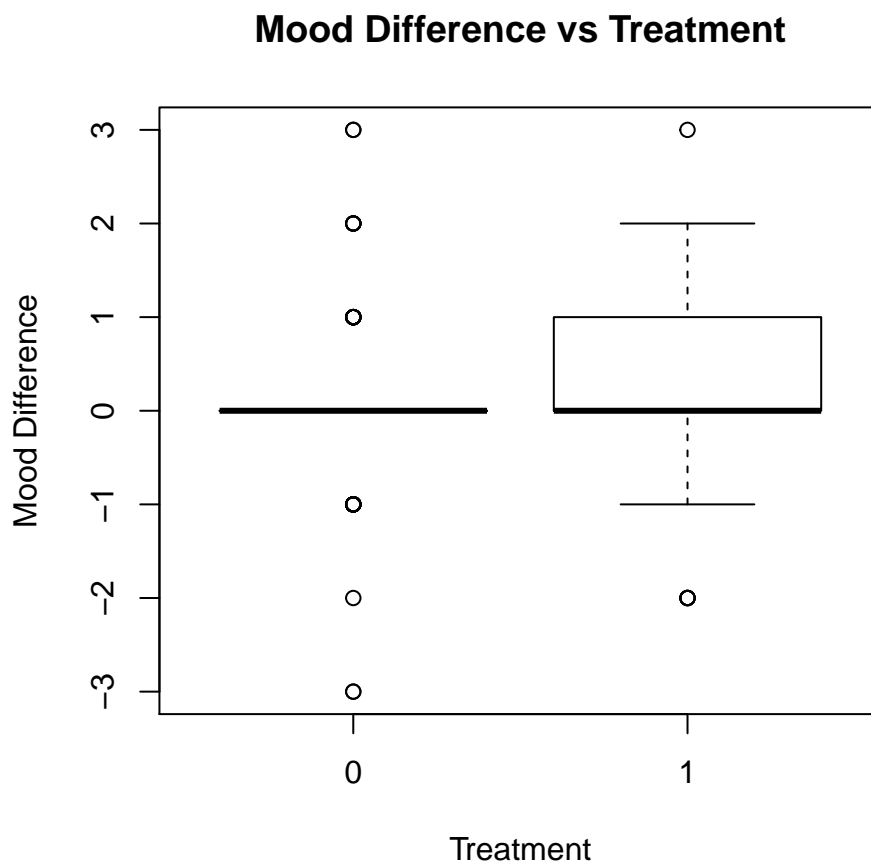
Randomization inference tells us that we do not meet the standard significance level of  $p \leq 0.5$ . We can say that our results are not statistically significant as shown through randomization inference, so we fail to reject the sharp null hypothesis. With that being said, we do see signs of marginal significance with a p-value of 0.069. This gives us some evidence that we should conduct another experiment to further examine this treatment effect.

#### 0.4.2.2 Mann-Whitney U Test

We will conduct a Mann-Whitney U Test next to examine confidence intervals. Our experiment tested the difference between two conditions, either a change in mood or no change in mood, with participants randomly assigned to treatment or control for each of the 4 days. Our outcome variable was a categorical variable of mood difference. Having a mood difference of -2 from a 4 to a 2 on our scale (moving from okay to sad) may have a more significant impact than having a mood difference of +2 from a 4 to a 6 on our scale (okay to happy) since the impact of emotional states aren't linear. Therefore, we used the non-parametric Mann-Whitney U test.

**Test on Overall Participants:**

```
boxplot(MoodDiff~Treatment, data = d,  
        main="Mood Difference vs Treatment",  
        xlab="Treatment",  
        ylab="Mood Difference")
```



```
wilcox.test(MoodDiff~Treatment, mu=0, paired=FALSE, conf.int=TRUE, data = d)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: MoodDiff by Treatment  
## W = 3840, p-value = 0.04979  
## alternative hypothesis: true location shift is not equal to 0  
## 95 percent confidence interval:  
## -4.679254e-05 1.054435e-05  
## sample estimates:  
## difference in location
```

```
## -4.470203e-06
```

The boxplot above shows the distribution of the mood difference for the control (0) and treatment (1) groups. Most control group participants had no change in their mood. The mean between the control and treatment groups did not differ, however, the boxplot shows that the distribution of the treatment group is larger than that of the control group. The Mann-Whitney U Test is needed so that we can for us to reject or fail to reject the null hypothesis that the distributions are similar in the control and treatment groups.

We have a p-value of .04979 with a 95% confidence interval of -0.000047 to 0.000011. Note that a mood difference of 0 is included in this confidence interval, therefore we see that we do not satisfy rejecting the sharp null hypothesis, despite a statistically significant p-value of .04979. Despite no statistical significance, we can argue that there is still marginal significance in the results, thus we should conduct another experiment to gain more data and insights.

### 0.4.3 Specific Average Treatment Effect Analysis

We see above that our main hypothesis that positive news articles affect mood yielded insignificant results when testing the entire sample. As noted in the EDA, we had a disproportionate amount of subjects under the age of 35. Below we explore the potential significance of this at the risk of p-hacking.

#### 0.4.3.1 Average Treatment Effect For Millenials

We will denote the age group that is 26-35 as millenials, isolate this group, and conduct tests to verify the statistical significance of the average treatment effect. We begin first with the general case of randomization inference:

```
d_m = d[d$AGE26To35 == 1, ]

sample_ate = d_m[Treatment == 1, mean(MoodDiff),] - d_m[Treatment == 0, mean(MoodDiff),]
print(paste0("The millenial ATE is : ", sample_ate))

## [1] "The millenial ATE is : 0.304545454545455"

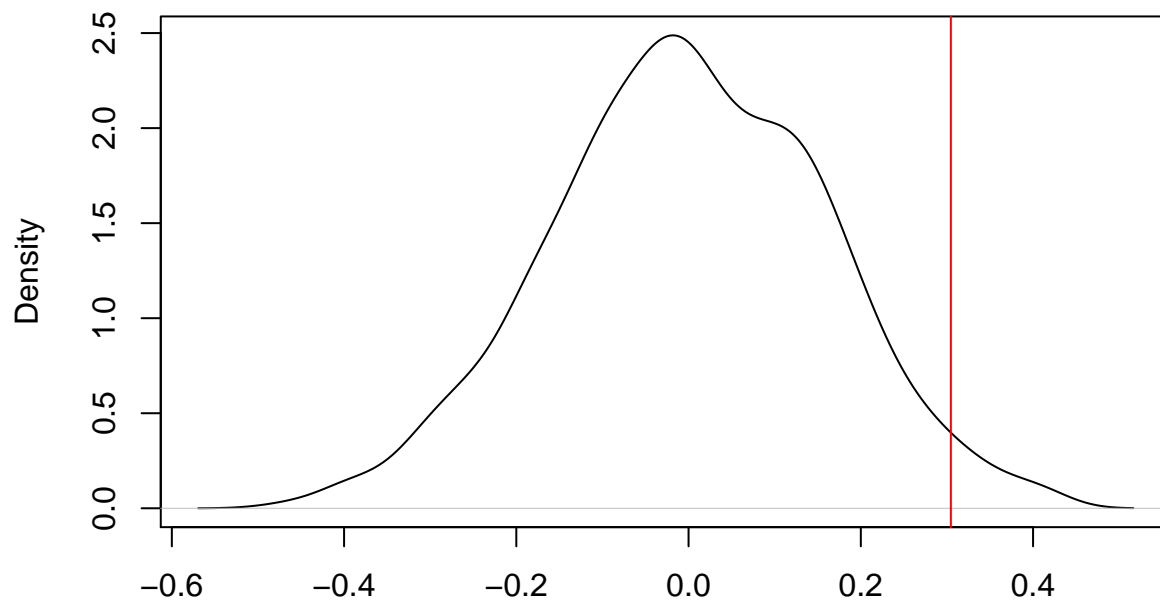
randomize = function() sample(c(rep(0,nrow(d_m[Treatment == 0])),rep(1,nrow(d_m[Treatment == 1]))))

est.ate = function() {
  d_m$treat_rand = randomize()
  mean(d_m[treat_rand == 1, MoodDiff]) - mean(d_m[treat_rand == 0, MoodDiff])
}

distribution.under.sharp.null = replicate(1000, est.ate())

plot(density(distribution.under.sharp.null),
     main = "Randomization Inference Results")
abline(v=sample_ate, col = "red")
```

## Randomization Inference Results



N = 1000 Bandwidth = 0.03576

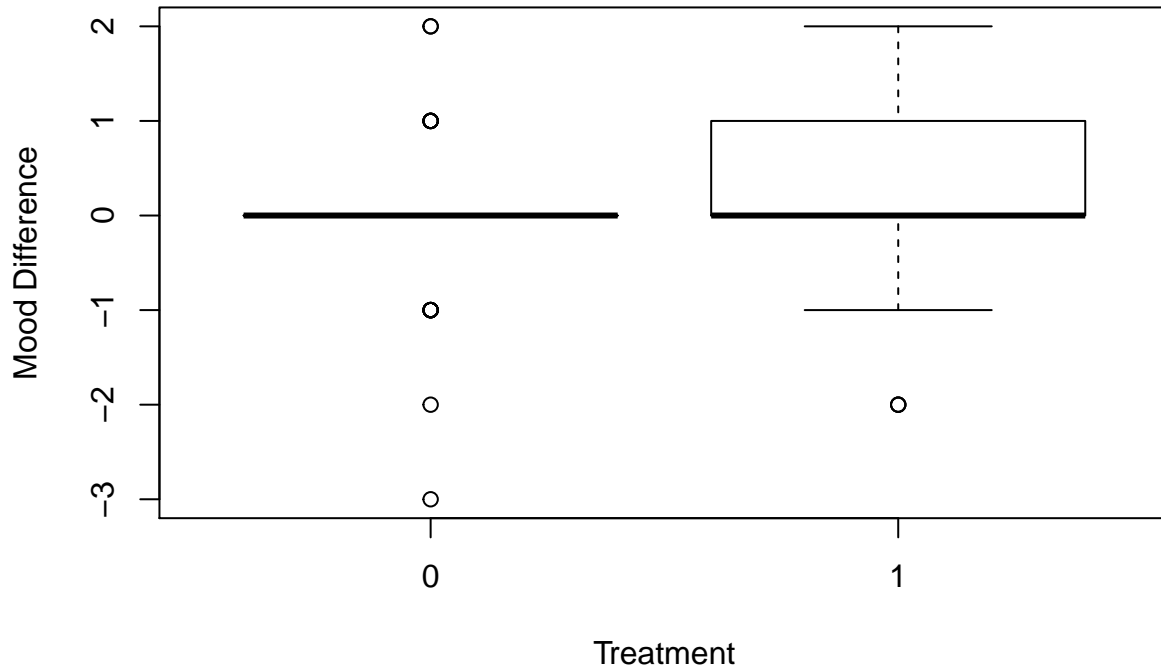
```
p_val = mean(sample_ate < distribution.under.sharp.null)
print(paste0("Millenial p-value : ", p_val))
```

```
## [1] "Millenial p-value : 0.02"
```

The randomization inference test yields a p-value that is statistically significant ( $< .05$ ). Therefore, under randomization inference we can reject the sharp null hypothesis; however, let us also check the Mann-Whitney U Test:

```
boxplot(MoodDiff~Treatment, data = d_m,
        main="Mood Difference vs Treatment",
        xlab="Treatment",
        ylab="Mood Difference")
```

## Mood Difference vs Treatment



```
wilcox.test(MoodDiff~Treatment, mu=0, paired=FALSE, conf.int=TRUE, data = d_m)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: MoodDiff by Treatment
## W = 1284.5, p-value = 0.02341
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -9.999476e-01 -3.399253e-05
## sample estimates:
## difference in location
## -4.575571e-05
```

With a p-value of 0.02341 and a 95% confidence interval of -.9999476 to -.0000340, we can reject the sharp null hypothesis. The Wilcoxon rank-sum (Mann-Whitney U Test) suggests that the uplifting news does affect the subject's mood ( $W = 1284.5$ ,  $p = 0.02341$ ). Note that the result is a slightly negative effect on mood in the treatment group.

## 0.5 Discussion of Results and Inference

In our main hypothesis test about the entire sample size, randomization inference suggested that there was no significant result ( $p = 0.6$ ). The Mann-Whitney U test suggested a significant result ( $p = 0.04979$ ). However, the confidence interval was close to zero (0) and range from -0.000047 to 0.000011. Therefore, we conclude that there was no statistically significant treatment effect of the uplifting news on participants happiness at the end of the day. The result suggested the happiness level of treatment group did not differ significantly from the control group. There were participants in the treatment group that indicated their mood (happiness level) changed negatively (e.g. from happy to sad). The results from these participants might contribute to the negative confidence interval. It was also an indication that other activities might occur during the day that our survey did not capture. Additionally, due to the categorical nature of our outcome variable, it is very

difficult to test the effect of any of the covariates we captured without using a regression. Randomization inference does not allow for such tests, thus we were only able to test the immediate treatment effect.

After collecting results and performing EDA, we realized that the age ranges were clumped around 26-35 years old. This allowed us to form a “side-hypothesis” that there might be a significant ATE if we isolated this age group. Upon isolation, the result suggested that the uplifting news had significant effect on participants’ happiness level. Interestingly we observed a confidence interval that was bounded below zero (negative to negative). This leads us to believe that while we did observe a significant ATE, there is still some indication that happiness would decrease.

Both tests on overall participants and age group of 26 to 35 suggested that we might fail to capture significant activities that happened during the day that might affect participants happiness level. Our survey asked how much sleep and coffee participants had prior night and during the day. However, these factors alone did not show significant correlation with participants happiness. Ultimately, when experimenting with a hard to measure feature such as happiness, we expect there to be a large amount of covariates that would be very difficult to control for.

## 0.6 Limitations

Despite not yielding any statistically significant results in our calculated treatment effect for the sample population group, our findings suggest that exploring age groups or employment status may result in statistically significant treatment effects. Unfortunately, given the time and budgetary constraints of the experiment, we were unable to generate a large enough sample and test this hypothesis. *If* we had the time and budgetary capabilities, we would collect a larger sample size and retest. We would also examine measuring a subject’s engagement to determine happiness rather than asking them to fill out a survey. We would need to remove as much subjectivity from the results as possible. A medium like Reddit or Buzzfeed would be an interesting place to conduct an experiment like this. We could have two webpages, one with only uplifting articles and another one with the standard article layout. We can place an advertisement similar to Wikipedia’s “Please Donate”, and see if people engage with the ad. While not perfect (and could potentially cause public backlash), this would allow us to measure some form of happiness in a non-subjective manner.

## 0.7 Conclusion

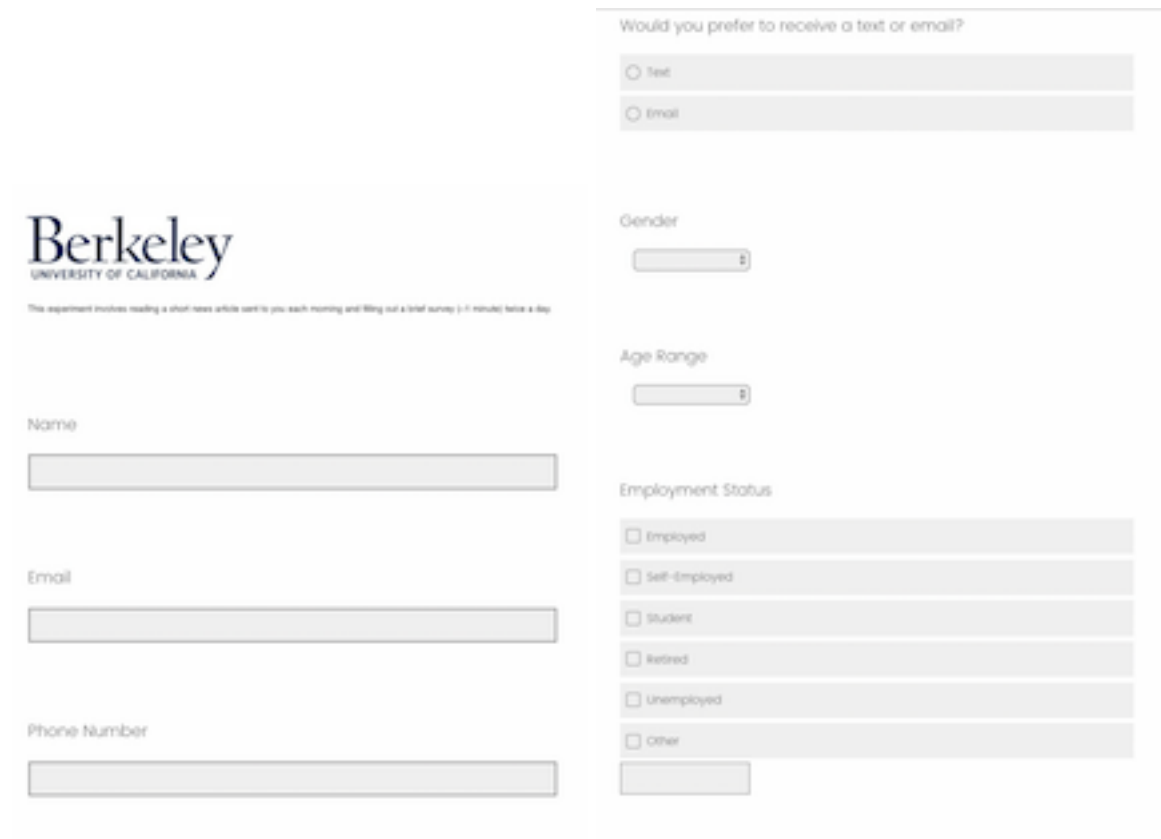
Ultimately, our original experiment did not allow us to reject the sharp null hypothesis; however, we did see promise in perhaps reconducting the experiment with a larger sample size and examining a smaller subgroup of age or employment. There was temptation to dive further into exploring these subsets within our current study, but we tried to be as careful as possible to avoid p-hacking or any fishing expeditions. An interesting outcome of the actual experiment is that subjects would message us and ask us to continue sending them news articles. They enjoyed being delivered news each morning that they could read. It was gratifying to hear that those that partook in the experiment enjoyed participating in it.

In conclusion, we were able to test our question of whether we could influence a person’s happiness by treating them with a positive news article. An experiment provided us to intervene on one’s group news consumption and establish a treatment and control group in order to conduct a field test of our hypothesis. Our hypothesis that we would reject the sharp null hypothesis  $H_0 : \mu_{\text{pre-article mood}} = \mu_{\text{post-article mood}}$  and that we would increase a subjects mood  $H_1 : \mu_{\text{pre-article mood}} < \mu_{\text{post-article mood}}$ . We split groups into a treatment and control group. The treatment group received an uplifting news article while the control group received and neutral news article. Our measurements consisted of friends, classmates, and family. Each day, we randomized our subjects into treatment and control which ultimately proved to provide an accurate representation of randomization. Our ultimate outcome was **Mood Difference** which was calculated by subtracting the post survey mood by the pre survey mood. This distribution looked fairly normal with a mean centered around zero, indicating no change in mood. We conducted randomization inference and Mann-Whitney U test in order to check for significance and ultimately deemed that our results were not statistically significant.



## 0.8 Appendix

### 0.8.1 Images of sign up form



The image shows a sign-up form for Berkeley University of California. The form is divided into two main sections. The left section contains the Berkeley logo, a description of the experiment, and input fields for Name, Email, and Phone Number. The right section contains a question about preferred communication method (Text or Email), and dropdown menus for Gender and Age Range. Below these are checkboxes for Employment Status (Employed, Self-Employed, Student, Retired, Unemployed, Other) and a final empty input field.

**Berkeley**  
UNIVERSITY OF CALIFORNIA

This experiment involves reading a short news article sent to you each morning and filling out a brief survey (5-7 minutes) twice a day.

Name

Email

Phone Number

Would you prefer to receive a text or email?

☐ Text  
☐ Email

Gender

Age Range

Employment Status

☐ Employed  
☐ Self-Employed  
☐ Student  
☐ Retired  
☐ Unemployed  
☐ Other

# Berkeley

UNIVERSITY OF CALIFORNIA

Instructions:

- Page 1: Answer all questions
- Page 2: Read the text
- Page 3: Complete last question

Submit the survey to ensure that you're entered in the raffle

Name

As of right now, have you had any caffeine today?

☐ Yes ☐ No

As of right now, how many news articles have you read?

☐ none

☐ 3 articles or less

☐ More than 3 articles




<https://www.juulfeednews.com/article/nb-wellbeing/cancer-free-remission>

☐ Ironman ☐ Batman

Click the arrow below to submit your survey and be entered into the raffle!



### 0.8.3 Post-Survey Images



UNIVERSITY OF CALIFORNIA

Instructions:

- Submit the survey to ensure that you're entered in the raffle!

Name

Roughly how many servings of caffeine have you had today?

☐ None

☐ 2 cups or less

☐ 3 to 4 cups

☐ 5 cups or more

Have you bought anything other than food today?

☐ Yes

☐ No

Did you help someone today?

☐ Yes

☐ No

Submit

How many news articles have you read today?

- ☐ None
- ☐ 3 articles or less
- ☐ More than 3 articles

As of right now, what is your current mood?


- ☐ Very upset
- ☐ Upset
- ☐ Slightly upset
- ☐ Average
- ☐ Slightly happy
- ☐ Happy
- ☐ Very happy

## 0.8.4 Example of Articles

Why Don't We Forget How to Ride a Bike?

The way memories are anchored in the brain plays a role, neuropsychologist Boris Suchan explains

By Boris Suchan on November 16, 2016



READ THIS NEXT

- Paradox of Memory
- When a Moment of Memory

BuzzFeed News

Batkid Is Now Cancer-Free After Stealing Everyone's Heart in San Francisco

The 10-year-old, who stole everyone's heart when his one wish was to be a superhero for one day, has now been in remission for five years.

Michelle Brader Van Dyke  
BuzzFeed News Reporter

Posted on November 14, 2016, at 10:11 p.m. ET

Twitter Facebook Email

