A Study of the Effects of Personalization on Response Rates for Employee Feedback Requests

Winston Lin, Carlos Rodriguez Castillo and Sean Underwood

MIDS W241 - Field Experiments, Spring 2016

1. Introduction

AppNexus is a medium-sized, global ad technology company. Its primary offering is an exchange platform where buyers and sellers of online advertising opportunities can interact in real-time through auction-based transactions. A rapidly growing company with over 1,100 employees, AppNexus is continuously looking to meaningfully engage with its workforce. As such, the strategy and operations groups at AppNexus conduct quarterly "All-Hands" meetings where all the employees at the three major business units of the company are invited to go through a retrospective of the quarter, a look ahead into the upcoming quarter and get a sense of where the company stands from the perspective of our company leaders.

These "All-Hands" meetings have a high production cost: several hundred hours of preparations in aggregate go into making each of them. They also incur a cost of hundreds of aggregate working hours for those individuals that attend the meetings. In order to evaluate the effectiveness of these meetings, the AppNexus operations team sends an email containing a link to a feedback survey to all attendees the day after each meeting. Attendees are asked various questions about the extent to which they recall information presented at the meeting as well as their reactions to it; general feedback is also requested. These feedback surveys play a significant role in helping management to evaluate how informative the meetings actually are and shape content for future meetings. To ensure that the feedback is informative and valid as an input to the meeting planning process, an adequate number of responses must be collected from employees around the company. Thus, interventions that increase response rates to the All-Hands meeting feedback surveys will help AppNexus leadership improve efforts to inform its employees about the company's status, major initiatives, as well as enhance its insight into important qualitative aspects of the workforce such as morale.

This paper looks to answer the research question: *Do more personalized requests for feedback increase the quality and quantity of feedback responses from subjects?* To do so, we conducted two analogous experiments on the "All-Hands" meetings of separate AppNexus business unit populations: the Advertiser Technology Group (ATG) and the Publisher Technology Group (PTG). This paper is divided into five main parts: an introduction, a detailed explanation of the experimental design, the results from the experiments conducted, a conclusion and an acknowledgement of those that helped make the experiment possible.

1.1. Previous Research

A number of studies have found that personalizing email survey invitations by including the recipient's name in the salutation yields a higher response rate than using a generic invitation (Fan & Yan 2009, Sánchez-Fernández et al. 2012 and Sauermann & Roach 2013). Social Exchange theory, which explains people's actions as being motivated by the expectation that the benefits of conferring a favor will ultimately be outweighed by the costs, is a popular conceptual framework for explaining this phenomenon. Social exchange is distinct from economic exchange in that it takes a more expansive view of the possible benefits that people might consider. Even a sense of fulfillment at helping others may be a sufficient reward. (Dillman et al, 2014)

In keeping with this theory, we hypothesize that the effect of personalization might be particularly strong in a smaller company like AppNexus, where an informal atmosphere and strong company culture will serve to amplify the perception of social reward that comes with personalization. We also hypothesize that the effect is strongest when the personalization uses the individual's first name rather than their full name as well as when informal language is used. Both of these hypotheses inform our study design. However, we do not attempt to experimentally verify either of them, and have instead sought the more modest goal of confirming that personalization improves survey response rates in a corporate setting.

2. Experimental Design

2.1 Definition of Treatment and Treatment Effect

The strategy and operations teams at AppNexus are primarily interested in studying the rate at which employees who attend "All-Hands" meetings respond to requests for feedback. The requests are emails sent to meeting attendees, and the responses submitted via an online survey. Thus, the outcome of interest is whether or not a given employee submits a response to the online feedback survey. This outcome is expressed as a binary variable for each attendee of the "All-Hands" meetings in our study. These events can then be aggregated for the groups studied and divided by the total number of subjects in each group to calculate a response rate for the subject groups of interest (specifically treatment and control) which as an outcome maps well to the primary interest of the AppNexus team mentioned above.

The treatment in this study is personalization of a feedback request. This personalization is defined as a modification of a feedback request to include (1) the first name of the recipient of the feedback request as well as (2) the more personal language "thanks". This is in contrast to "no personalization" in control

where the feedback requests are directed to "team", as opposed to the subject, and the more formal "thank you" is used instead of "thanks" (see Appendix A for examples of the treatment and control feedback requests).

We identified the treatment effect through linear regression estimation of treatment (personalization) on outcome (feedback form response submission). Our experiment had a simple null and alternative hypothesis design. Our null hypothesis was that personalization of the feedback request would result in no change in response rates; our alternative hypothesis was that such personalization would result in an increase in response rates. Our model and hypotheses can formally be expressed as:

Linear regression model and null and alternative hypotheses

$$Y_i = \beta_0 + \beta_1 \times d_i + \varepsilon_i$$

where: $Y_i = 1$ if subject submits response to feedback request; 0 if no response
β_1 = average treatment effect of personalization
(change in response rate due to personalization, in percentage points)
$d_i = 1$ if subject i receives treatment; 0 if in control

$H_0: \beta_1 = 0$ (no change in response rate)
$H_A: \beta_1 > 0$ (positive change in response rate)

2.2. Sampling

For the two AppNexus studies, we were able to obtain a census of all meeting invitees. The operations team supplied a list of all email addresses in the mailing lists they typically use for distributing surveys, and Human Resources supplied additional information about employees' organizational unit, region and gender. There were 352 total employees in the ATG list, and 379 in the PTG list. Each of the AppNexus email lists included eleven people who could not be cross-referenced the data supplied by HR. Missing data included the person's given name, so it was not possible to apply the experimental treatment to these individuals. We believe these were people who left the company, but whose email addresses had not yet been removed from the distribution lists. In keeping with AppNexus practice, we sent them a generic survey invitation, but we chose to exclude them from analysis.

2.3. Assignment of Treatment

Block random assignment was used to place subjects into treatment and control. Subjects were blocked by organizational unit, geographic region and gender. AppNexus defines two major organizational units, Engineering and Services. There are three geographic regions: Americas (AMS), Asia/Pacific (APAC) and Europe/Middle East/Africa (EMEA). Out of the 12 possible combinations of these variables, we assigned nine blocks for the ATG survey and eight for the PTG survey. Two blocks were unassigned in the ATG data due to there being no male or female APAC engineers, and the two blocks for EMEA engineers were merged into a single one because only one was female. In PTG, there again no male or female EMEA engineers. There were also no female EMEA engineers or APAC services staff.

After blocking was performed, complete random assignment was applied to subjects within each block. To verify that our block randomization procedure was successful, we checked for covariate balance between treatment and control groups in each block. Block definitions and summary statistics for samples from the ATG and PTG studies are shown below in Table 1.

Table 1. Blocks, covariates, values, and block size

ATG Survey Blocks

Org	Region	Gender	Treat	N
Engineering	AMS	F	0	10
Engineering	AMS	F	1	10
Engineering	AMS	M	0	47
Engineering	AMS	M	1	46
Engineering	EMEA	NA	0	3
Engineering	EMEA	NA	1	4
Services	AMS	F	0	33
Services	AMS	F	1	32
Services	AMS	M	0	52
Services	AMS	M	1	51
Services	APAC	F	0	1
Services	APAC	F	1	2
Services	APAC	M	0	7
Services	APAC	M	1	8
Services	EMEA	F	0	9
Services	EMEA	F	1	9
Services	EMEA	M	0	13
Services	EMEA	M	1	14

PTG Survey Blocks

Org	Region	Gender	Treat	N
Engineering	AMS	F	0	14
Engineering	AMS	F	1	14
Engineering	AMS	M	0	64
Engineering	AMS	M	1	65
Engineering	EMEA	M	0	1
Engineering	EMEA	M	1	2
Services	AMS	F	0	31
Services	AMS	F	1	31
Services	AMS	M	0	40
Services	AMS	M	1	39
Services	APAC	M	0	4
Services	APAC	M	1	4
Services	EMEA	F	0	11
Services	EMEA	F	1	10
Services	EMEA	M	0	24
Services	EMEA	M	1	25

2.4. Experimental Protocol

In order to test our procedures and identify possible issues ahead of the actual experiments, we performed a pilot study in collaboration with the datascience@berkeley staff, who graciously agreed help us to test our procedures. The study involved sending emails inviting MIDS students to take an online survey designed to measure their awareness, engagement, and satisfaction with respect to the *Applications in Data Science* webinar series offered by the MIDS program. The decision to conduct the pilot study with MIDS students was meant to preserve the sample for the actual experiments and ensure sufficient power to detect a treatment effect if one existed. The details of the pilot study can be found in Appendix A.

For the actual experiments, we collaborated with the AppNexus team to implement and execute the treatment procedure. Typically, the operations team created feedback surveys in SurveyMonkey, each with a single response collector, and a member of the operations team would then send an email with a request for feedback and a link to the survey to a company distribution list. For our experiment, the operations team created multiple response collectors for each feedback survey in order to bucket responses for individual treatment or control groups in each block. The operations team then used the mail merge feature in Microsoft Outlook to send the email invitation to each subject individually.

Since AppNexus collects survey responses anonymously, in principle it is not possible to tie responses to individuals after the fact. They also usually do not include demographic questions in their surveys, and we did not wish to introduce them for fear of being incongruous with typical AppNexus feedback questionnaires. By blocking on our covariates of interest and using separate SurveyMonkey collectors for each block, we were able to relate the response rates to our covariates of interest in a way that worked with AppNexus' requirements.

We should note, however, that we failed to completely preserve respondent anonymity in our treatment assignment. In both studies, we had cases where a single person was assigned to a single collector. If any of the people assigned to these collectors had responded then it would have been possible for AppNexus staff to identify the person by cross-referencing their response with the mail merge list that we supplied.

Past email invitations to take feedback surveys were sent from individual email accounts via Microsoft Outlook, so we designed our protocol to mimic this and limit the actual treatment to personalization of the feedback request. Emails for both treatment and control were implemented with

Outlook's Mail Merge feature, which enables users to quickly create and send personalized emails in batch from their individual Outlook accounts to a list of email contacts. The feature required an Outlook account, Word template containing the email body and custom fields for the salutation, and comma-separated value file with the names and email addresses to be merged into each email. The AppNexus team provided a draft of the email body with the survey link, which was reproduced in a Word template. After we tested the process of personalizing and sending the emails with Mail Merge, we provided the AppNexus team with the final email templates and step-by-step instructions to execute the treatment protocol. Examples of the final treatment and control emails can be found in Appendix B.

Since AppNexus is a small company, there was some risk of spillover effects due to employees talking about the survey email with their coworkers. We initially considered cluster randomization as a way to counter this, but ultimately decided against it for several reasons. It was uncertain that clustering on any of the available covariates would work. We expected that any discussion of the survey would happen within social groups rather than professional ones, and we were uncertain that any of the information we had would correspond well with the social networks. Geographic region was a likely candidate, but, with only three major regions provided, it was reported to us at too coarse a level to be useful for cluster random assignment. We also expected the treatment to go largely unnoticed, or at least not to be seen as interesting enough to merit much conversation among survey recipients. Finally, our initial statistical power estimates indicated that our experiment was already underpowered, and we did not want to reduce it any further by introducing clustering when we could not be certain that it would confer any benefit.

Since we used mail merge, individual emails were addressed to each person in both the treatment and control groups. Past practice was to send the email to a distribution list. Since the "to" address on an email is visible to recipients, all participants in the experiment would see that the email is individually addressed to themselves in that line. This could be considered a form of personalization, but we did not consider it to be a form of experimental treatment because it was applied uniformly across both the treatment and control groups. Our initial expectation was that this difference would largely go unnoticed by the email recipients, and would therefore have a negligible influence on our results.

As it happens, though, several people both in the ATG and PTG groups noticed the difference, and were interested enough to reply directly to the survey invitation with inquiries or to express surprise at the change. Some asked if the email was accidentally sent to only them. We instructed the operations team to respond that there was no error and that they were trying out a different method to send the survey out this

time. This made it somewhat more difficult to interpret our results, which we will discuss more in the Results section.

The ATG All-Hands meeting was held on April 7, 2016, and the feedback survey was sent the following morning to all attendees. The PTG All-Hands meeting took place on April 20, 2016, and the survey was sent the following afternoon to all attendees. Survey responses for both meetings were tallied for each treatment group after seven days.

3. Results & Discussion

3.1. Experiment 1: ATG All-Hands Meeting

Results from the ATG All-Hands feedback survey are summarized in Table 2. The survey had an overall response rate of 8.0%. The response rate among recipients of the generic survey invitation was 4.0%. Recipients of the personalized email were nearly three times as likely to respond: the treatment group had an overall response rate of 11.9% which translates to an estimated treatment effect of a 7.9% lift in response rate. This result is significant at the 1% level ($p = .006$, 95% CI = [2.5%, 13.7%].

We saw zero responses among the group that was excluded due to NA covariates. This lends some additional credence to our belief that those individuals are no longer at the company. However, since this group was sent the control email, and we saw zero responses from the control group in many of the blocks, it is not a conclusive confirmation of our supposition.

3.2. Experiment 2: PTG All-Hands Meeting

Results for the PTG All-Hands meeting are can be found in Table 2. We found an overall response rate of 8.9%; the response rate in the control group was 4.9% and the response rate in treatment was 13%. We estimated a treatment effect of a 8.1% lift in response rate, which is significant at the 1% level ($p = .006$, 95% CI = [2.3%, 13.9%]).

Similarly to what was observed in the ATG experiment, no responses were recorded for the group that was excluded due to NA covariates. Like in the analysis of the ATG results, we consider this an indication that we may indeed be dealing with employees that no longer work at the company.

Table 2. Regression analysis for ATG and PTG feedback survey experiments

	Dependent variable:	
	Responded to Survey	
	ATG	PTG
	(1)	(2)
Personalization	0.079***	0.081***
	(0.029)	(0.029)
Constant	0.040**	0.049**
	(0.020)	(0.021)
Observations	351	370
R^2	0.021	0.020
Adjusted R^2	0.019	0.018
Residual Std. Error	0.269 (df = 349)	0.283 (df = 368)
F Statistic	7.641*** (df = 1; 349)	7.599*** (df = 1; 368)
Note:		*p<0.1; **p<0.05; ***p<0.01

3.4. Applications and Future Research

One avenue for future research associated to this study could focus on determining to what extent the findings in the AppNexus experiments generalize to other populations of interest. This is particularly interesting given the very different results in the pilot vis-a-vis the ATG and PTG experiments. Future experiments could be for populations analogous to those in the pilot or the AppNexus experiments.

That said, within the AppNexus population there is also the interesting possibility of continuing to study the populations that attend the "All-Hands" meetings. A future research opportunity that may be relevant to the strategy and operations teams at AppNexus is how to increment the response rate for feedback forms across specific subsections of the population. Specifically, they may be interested in finding mechanisms to better engage and obtain higher response rates from smaller geographical, gender and organizational populations (e.g. EMEA-Engineering, Female-Services-EMEA, Male-Services-APAC). This could potentially be achieved through incorporating standard error reduction techniques into the experimental framework such as deploying higher dosages of treatment and/or implementing a within subjects design that would permit the experimenters to expand the sample size in the blocks with less people.

These additional research opportunities at AppNexus would be greatly facilitated by the experimental framework that we developed as it provides features that significantly reduce the operational

burden of executing them (e.g. mail merge procedure) and provide techniques to extract significant and relevant results (e.g. blocked random assignment).

AppNexus could also evolve the experimental design to incorporate better privacy safeguards for situations where blocks are merged into larger blocks given that a single observation is in the block. One option to tackle this issue would be to randomly and blindly allocate the single observation to a contiguous block that would maintain the anonymity of the individual (changing the block coding of the observation at the time of merging).

Yet another source of further research would be in the realm of experimental treatments. We would like to determine if there are other forms of treatment that could be effective in stimulating responsiveness: different kinds of personalization (salutations, signatures, etc.), subject line/body content, incentives for respondents, timing of invites/reminders (see previous research, Sauermann and Roach (2012)).

Finally, rather than explore different kinds of treatments, future work could alternatively look into different outcomes of interest. Other studies could look into estimating other responsiveness metrics such as survey completion rates or quality of feedback responses (e.g. average number of words).

4. Conclusion

While our initial results are encouraging, we feel that it is best to approach them with a measure of skepticism. One of our core assumptions was that our treatment would only be subtly noticeable to the email recipients. Since people noticed artifacts of our experiment and thought them to be noteworthy enough to mention to the person who sent out the survey invitations, we have to consider the possibility that at least some of our observed treatment effect is due to the Hawthorne effect. The reaction we observed was to the change in how emails were addressed (sending individual emails rather than using a distribution list), which was applied equally to both the treatment and control groups. We are not aware of anybody commenting specifically on the personalized salutation, but we also have no compelling reason to believe it would have been any less noticeable.

If participants were talking to the operations team about the changes to the survey format then they were likely also talking to each other. This potential source of spillover would have reduced our estimate of the ATE, but it is doubtful that it would have been great enough to counteract any overestimation due to the novelty of the treatment. We still think it is most likely that our results overestimate the impact that AppNexus could expect over the long run by personalizing their survey invitation emails.

In a non-experimental setting the treatment would have been inexpensive to apply. The bulk of the effort in running our experiment was applying the mechanisms for observing the experiment's outcome, not applying the treatment. Even accounting for the likelihood that the majority of the increase in response rate would dissipate over time, our results still indicate that relatively minor personalization of the invitations would be a cost-effective way to improve AppNexus' employee survey response rates.

5. Acknowledgements

We would like to thank AppNexus for their collaboration in this study. Without their collaboration and support the study would not have been successful or even possible. We would also like to thank Alex Hughes and David Reiley who gave us invaluable insight into on how to proceed with some of the trickier parts of this experiment (and who taught us everything we know about field experimentation!). Finally, we would like to thank Drew Paulin and Paloma Tamminga who patiently helped us implement our pilot study.

Appendices

Appendix A. Pilot Study

For the pilot study, a list of all 270 students currently enrolled in the MIDS program was supplied by 2U. In cases where members of our team were members of the survey audience, we assigned ourselves to a known treatment group so that we could subtract them from the data prior to analysis. This allowed us to assist our collaborators by participating in their surveys without biasing our own results.

One significant departure from the primary experiment's design was that we did not use block random assignment for the pilot study. Gender was not included because we did not have easy access to that information, and department was not included because that variable does not apply in this case. This did unfortunately limit our ability to use the pilot as a full dress rehearsal, since it involved significantly less logistical complexity.

The pilot treatment was roughly the same as in the primary study, with control participants receiving an email with a generic salutation and treatment participants receiving a personalized salutation that included their first name. In order to remain consistent with existing precedent, the exact wording of the email salutation was slightly different: "Hello Everyone" for control, and "Hello [student's first name]" for treatment. After the salutation and a short message, all subject received a link to the MIDS webinar survey.

The pilot survey and email for were implemented by the 2U team using Clicktools, an online survey platform. We provided a draft of the email and performed random assignment of subjects based on the email list provided by 2U. The emails were sent to all subjects on April 1, 2016, and a reminder email to take the survey was sent one week later. Anonymized responses for our pilot analysis were provided by 2U the same day that the reminder email was sent. In addition, to help draw additional responses, we also posted a reminder to take the survey in the general MIDS Slack team the day after we received results.

Results from the pilot study are listed in Table 3. The response rate was 3.0% in the control group and 4.5% in the treatment group. Our point estimate of the ATE was not significant, with a 95% confidence interval of [-3%, 6%] estimated using randomization inference under the sharp null hypothesis.

Table 3. Pilot Survey Results

Group	N	Responses	Rate
Treatment:	134	6	4.5%
Control:	133	4	3.0%
Total:	267	10	3.7%

Appendix B. Examples of Emails for Treatment and Control Groups

CONTROL EMAIL

Hi team!

Thank you again for joining the ATG All Hands yesterday.

For access to the Vidyo recording and Powerpoint presentation, please visit the All ATG Chatter Group.

As we are always trying to improve, please also take a few moments to provide us with your feedback via this survey.

Thanks,
Kelsey

TREATMENT EMAIL

Hi John!

Thanks again for joining the ATG All Hands yesterday.

For access to the Vidyo recording and Powerpoint presentation, please visit the All ATG Chatter Group.

As we are always trying to improve, please also take a few moments to provide us with your feedback via this survey.

Thanks,
Kelsey

Appendix C. R Code for Experiment

```
library(data.table)
library(ggplot2)
library(RCurl)
library(stargazer)
library(ri)


############################
###### ATG EXPERIMENT ######
############################

###### ATG EXPERIMENT: EXPLORATORY ANALYSIS ######

# read data from remote git repo
csv =
getURL('https://raw.githubusercontent.com/winlingit/w241-project-csw/master/rcode/a
tg_results.csv')
dt.atg = data.table(read.csv(textConnection(csv)))

# estimate overall ATE
dt.atg[ , .(y = sum(Responses)/sum(N)), by = treat][ , y[1]-y[2]]

# calculate response rates in each block
dt.atg[, Rate := Responses / N]
dt.atg[ treat == 1, Treat := 'treatment']
dt.atg[ treat == 0, Treat := 'control']

# Plot the chart of winning
ggplot(dt.atg, aes(x=Block, y=Responses, fill=Block)) +
  geom_bar(stat='identity') +
  geom_text(aes(x=Block, y=Responses, label=Responses), vjust=-.1, size=3) +
  facet_wrap( ~ Treat) +
  ggtitle("ATG Responses by Block") +
  xlab("") +
  ylab("Responses") +
  theme( axis.line=element_blank(),
         axis.text.x=element_blank(),
         axis.ticks.x=element_blank(),
         axis.title.x=element_blank(),
         panel.border=element_blank(),
         panel.grid.major.x=element_blank(),
         panel.grid.minor.x=element_blank())

# Plot the chart of proportional winning
ggplot(dt.atg, aes(x=Block, y=Rate, fill=Block)) +
  geom_bar(stat='identity') +
  geom_text(aes(x=Block, y=Rate, label=round(Rate, digits=2)), vjust=-.1, size=3) +
  facet_wrap( ~ Treat) +
  ggtitle("ATG Response Rates by Block") +
  xlab("") +
```

```
  ylab("Response Rate") +
  theme( axis.line=element_blank(),
         axis.text.x=element_blank(),
         axis.ticks.x=element_blank(),
         axis.title.x=element_blank(),
         panel.border=element_blank(),
         panel.grid.major.x=element_blank(),
         panel.grid.minor.x=element_blank())


###### ATG EXPERIMENT: REGRESSION ANALYSIS ######

# regression models
m1.atg = lm(responded ~ treat, data = dt.atgx) # treatment only
m2.atg = lm(responded ~ treat + Female, data = dt.atgx) # treatment + female
m3.atg = lm(responded ~ treat + Female + Org, data = dt.atgx) # treatment + female
+ org
m4.atg = lm(responded ~ treat + Female + Org + Region, data = dt.atgx) # treatment
+ female + org + region

# show all regression models
stargazer(m1.atg, m2.atg, m3.atg, m4.atg, type = 'text', title = 'Regression
Analysis for ATG Feedback Survey Experiment')

t.test(dt.atgx$responded ~ dt.atgx$treat)$p.value



############################
###### PTG EXPERIMENT ######
############################

###### PTG EXPERIMENT: EXPLORATORY ANALYSIS ######

# read data from remote git repo
csv =
getURL('https://raw.githubusercontent.com/winlingit/w241-project-csw/master/rcode/p
tg_results.csv')
dt.ptg = data.table(read.csv(textConnection(csv)))

# recover observations
dt.ptgx = dt.ptg[rep(seq(.N), N)]  # expand table to 352 rows
dt.ptgx[ , responded := c(rep(1, max(Responses)), rep(0, .N - max(Responses))), by
= CollectorName]  # add var for responded
dt.ptg$Responses == dt.ptgx[ , sum(responded), by = CollectorName]$V1  # checksums
for total responses for each collector

# estimate overall ATE
dt.ptg[ , .(y = sum(Responses)/sum(N)), by = treat][ , y[1]-y[2]]

# calculate response rates in each block
dt.ptg[, Rate := Responses / N]
dt.ptg[ treat == 1, Treat := 'treatment']
dt.ptg[ treat == 0, Treat := 'control']
```

```
# Plot the chart of winning
ggplot(dt.ptg, aes(x=Block, y=Responses, fill=Block)) +
        geom_bar(stat='identity') +
        geom_text(aes(x=Block, y=Responses, label=Responses), vjust=-.1, size=3) +
        facet_wrap( ~ Treat) +
        ggtitle("PTG Responses by Block") +
        xlab("") +
        ylab("Responses") +
        theme( axis.line=element_blank(),
               axis.text.x=element_blank(),
               axis.ticks.x=element_blank(),
               axis.title.x=element_blank(),
               panel.border=element_blank(),
               panel.grid.major.x=element_blank(),
               panel.grid.minor.x=element_blank())

# Plot the chart of proportional winning
ggplot(dt.ptg, aes(x=Block, y=Rate, fill=Block)) +
        geom_bar(stat='identity') +
        facet_wrap( ~ Treat) +
        geom_text(aes(x=Block, y=Rate, label=round(Rate, digits=2)), vjust=-.1,
size=3) +
        ggtitle("PTG Response Rates by Block") +
        xlab("") +
        ylab("Response Rate") +
        theme( axis.line=element_blank(),
               axis.text.x=element_blank(),
               axis.ticks.x=element_blank(),
               axis.title.x=element_blank(),
               panel.border=element_blank(),
               panel.grid.major.x=element_blank(),
               panel.grid.minor.x=element_blank())


###### PTG EXPERIMENT: REGRESSION ANALYSIS ######

# regression models
m1.ptg = lm(responded ~ treat, data = dt.ptgx) # treatment only
m2.ptg = lm(responded ~ treat + Female, data = dt.ptgx) # treatment + female
m3.ptg = lm(responded ~ treat + Female + Org, data = dt.ptgx) # treatment + female
+ org
m4.ptg = lm(responded ~ treat + Female + Org + Region, data = dt.ptgx) # treatment
+ female + org + region

# show all models
stargazer(m1.ptg, m2.ptg, m3.ptg, m4.ptg, type = 'text', title = 'Regression
Analysis for PTG Feedback Survey Experiment')

t.test(dt.ptgx$responded ~ dt.ptgx$treat)$p.value



####################
```

```
###### EXTRAS ######
####################

###### ATG EXPERIMENT: RANDOMIZATION INFERENCE #######

# Note: moved from main analysis

# recover observations, from:
http://stackoverflow.com/questions/2894775/replicate-each-row-of-data-frame-and-spe
cify-the-number-of-replications-for-each
dt.atgx = dt.atg[rep(seq(.N), N)]  # expand table to 352 rows
dt.atgx[ , responded := c(rep(1, max(Responses)), rep(0, .N - max(Responses))), by
= CollectorName]  # add var for responded
dt.atg$Responses == dt.atgx[ , sum(responded), by = CollectorName]$V1  # checksums
for total responses for each collector

### 1. Analysis with "ri" package

## NOTE: This code is largely copy/pasted and hacked from the example code for the
dt package
y = dt.atgx$responded
Z = dt.atgx$treat
block = dt.atgx$blocknum

# Estimate ATE without blocking
perms.noblock <- genperms(Z, maxiter=100000) # Generate 100k additional assignments
probs.noblock <- genprobexact(Z) # probability of treatment
ate.noblock <- estate(y, Z, prob=probs.noblock) # estimate the ATE
ate.noblock

Ys <- genouts(y, Z ,ate=0) # generate potential outcomes under sharp null of no
effect
distout.noblock <- gendist(Ys, perms.noblock, prob=probs.noblock) # generate
sampling dist. under sharp null
dispdist(distout.noblock, ate.noblock)  # display characteristics of sampling dist.
for inference

Ys <- genouts(y,Z,ate=ate.noblock) ## generate potential outcomes under tau = ATE
distout.noblock <- gendist(Ys,perms.noblock, prob=probs.noblock) # generate
sampling dist. under tau = ATE
dispdist(distout.noblock, ate.noblock)  ## display characteristics of sampling
dist. for inference

# Estimate ATE with blocking
perms <- genperms(Z, blockvar=block, maxiter=100000) # Generate 100k additional
assignments
probs <- genprobexact(Z, blockvar=block) # probability of treatment
ate <- estate(y, Z, prob=probs) # estimate the ATE
ate

# Conduct Sharp Null Hypothesis Test of Zero Effect for Each Unit

Ys <- genouts(y, Z, ate=0) # generate potential outcomes under sharp null of no
effect
```

```
distout <- gendist(Ys, perms, prob=probs) # generate sampling dist. under sharp
null
dispdist(distout, ate)  # display characteristics of sampling dist. for inference

# Generate Sampling Distribution Around Estimated ATE

### 2. Analysis with manual RI

# from randomization.R
assign.treatment <- function(n) {
        n.treat = round(n / 2)  # Uses banker's rounding
        n.control = n - n.treat
        data = sample(c(rep(1, n.treat), rep(0, n.control)))
        data
}

# generate sampling distribution for ATE
sim.ate = function(dt) {
        dt[ , treat.ri := assign.treatment(.N), by = block] # randomize assignments
in each block
        ate = dt[ , .(y = sum(responded)/.N), by = treat.ri][ , y[1]-y[2]] #
estimate ATE for each assignment
        ate
}

ate.dist = replicate(10000, sim.ate(dt.atgx)) # generate sampling distribution for
ATE
plot(density(ate.dist), main = "Distribution of ATE", col = "black", xlab = NA)
abline(v = ate, col = 'red')

p.value = mean(ate.dist >= ate) # return 1-tailed p-value for estimated ATE
p.value # p-value < 0.01
```

Works Cited

Dillman, D. A., Smyth, J. D., & Christian, L. M., 2014. Internet, Phone, Mail, an Mixed-Mode Surveys: The Tailored Design Method, 4th Edition. John Wiley & Sons, Inc.

Fan, W. & Yan, Z., 2010. Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior,* 26, 132-139.

Sanchez-Fernández, J., Muñoz-Levia, F., & Montoro-Ríos, F. J., 2012. Improving retention rate and response quality in Web-based surveys. *Computers in Human Behavior* 28, 507-514.

Sauermann, H. & Roach, M., 2013. Increasing web survey response rates in innovation research: An experimental study of static and dynamic contact design features. *Research Policy,* 42, 273-286.