# Do animated graphs increase the effectiveness of email-based event recruitment at Yale?

A Final Project for S&DS315

Deon Ababio, Alex Gordon, and Nathan Kim

December 11th, 2020

## Introduction

The coronavirus pandemic has affected nearly every aspect of the Yale student experience. Most notably, classes have been moved onto Zoom, but students are still learning. It is what takes place beyond the classroom that is of particular interest: extracurricular participation. The traditional activities bazaar was held virtually, with clubs making online booths; however, due to the lack of in-person recruitment tactics, such as food events, welcome meetings, and getting to know club members through socials, clubs needed to get creative in how they marketed to the incoming first-year class.

The Yale Entrepreneurial Society, the Yale Scholars of Finance, and the Yale Undergraduate Consulting Group, for example, sent out mass emails to the student body in hopes of increasing attendance at their recruitment events. In an online environment, email appeared to be the best approach to connecting with Yalies.

But what kind of email will get the most attendance? To find out, we conducted a randomized field experiment to compare the efficacy of static and animated email graphics. We developed a new club on campus, the Yale Students for Civic Engagement, and invited the entire first-year class to attend our virtual election watch party.

We randomized 1265 first-years into one of four treatments: 1) two static graphs (control group), 2) one static graph and one animated graph, 3) one animated graph and one static graph, and 4) two animated graphs. In the end, we found no statistically significant difference in the RSVP rate between the various assignments.

This paper will answer our research question: Do animated graphs increase the effectiveness of email-based event recruitment at Yale? We will cover our hypotheses, treatments and outcomes, analysis and interpretation of data, limitations of our experiment, and a concluding discussion about our results.

## Hypotheses

It stands to reason that the design of the graph that gets the most response should be the one that is the most attractive. Literature on the use of data visualization testifies that animated graphics can be exciting ways to show trends in data with multiple dimensions. Robertson, et al confirm "[trend animation] is the fastest technique for presentation and participants find it enjoyable and exciting", but they also found that the motion can occasionally lead to participant confusion (Robertson et al. 2008). [1]

Insofar as the graphs we are animating are straightforward bar and line charts, we think that the tradeoff in confusion is minimal compared to the increased attention-getting ability of the animation. The application of our research is not to increase understanding of information, but to increase event attendance by getting subjects to read an email. If a viewer as drawn in by the animated graph, we expect them to continue reading through the email, and to be more likely to RSVP.

Therefore, we hypothesize that the treatment group shown two animated graphs in the body of the recruitment email will yield the greatest share of event RSVPs. The treatment groups with one animated and one static graph will yield the second greatest share of event RSVPs. And the control group shown two static graphs will yield the least share of event RSVPs. More generally, we can write out this hypothesis as the following:

H1: Emails with animated graphs will receive a different proportion of RSVPs than emails with static graphs.

We also wanted to test whether there was an impact of animating just the bar graph or just the line graph, under the theory that a line graph that moves along an x-axis of time will be easier to follow than a bar graph that moves along a y-axis of percent turnout. We test the following hypothesis:

H2: Emails with animated line/static bar graphs will receive a different proportion of RSVPs than emails with static line/animated bar graphs.

[1] Robertson, George et al. "Effectiveness of animation in trend visualization." IEEE transactions on visualization and computer graphics vol. 14,6 (2008): 1325-32. doi:10.1109/TVCG.2008.125

# Methods

The purpose of this project was to discover if animated graphics lead to an increase in signups. We had four different treatment groups:

1. A static line graph of voter turnout rates for the midterm and presidential elections from 1880 to the present.
2. An animated line graph of voter turnout rates for the midterm and presidential elections from 1880 to the present
3. Static bar graph of registered voters and turnout among racial lines in 2016
4. Animated bar graph of registered voters and turnout among racial lines in 2016

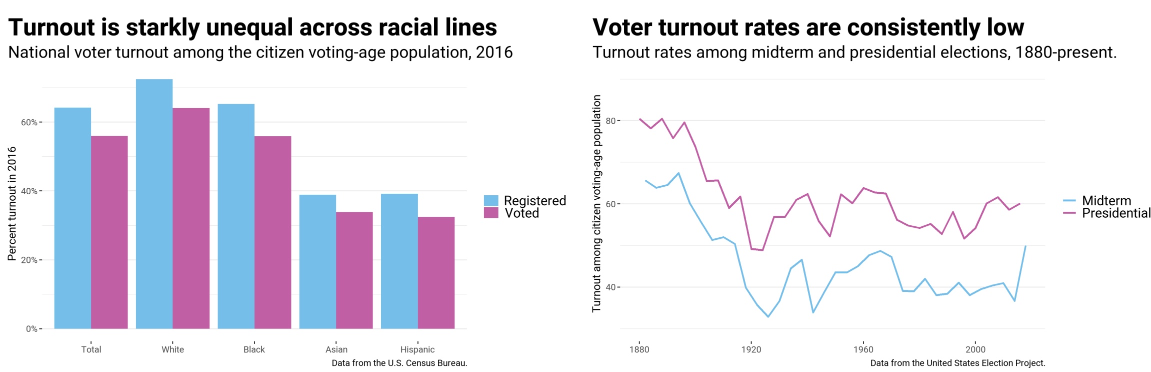


Figure 1: The static versions of the bar and line plots.

These graphics were generated with the `ggplot2` and `showtext` packages, with the `gganimate` and `magick` packages in R to convert the static images into animated plots.[1] The voter turnout trends since 1880 in the line graph was taken from the United States Elections Project, and the trends by race seen in the bar graphs came from the U.S. Census Bureau. Our animations can be viewed here.

[1] An annoying issue was had in formatting, where custom fonts implemented via the `showtext` package somehow made random frames distorted in the `gganimate`d plots. These frames were manually removed and then the `magick` package was used to stitch the remaining frames back together, with the end result being a GIF that misses a few frames and therefore could have appeared rocky. After watching both sets of GIFs several times I decided this was a negligible issue.

When deciding how to place people in certain treatment groups, we used a simple random sample. We used a random number generator to equally place the same number of people in each group. There are many statistical concepts that we could have used to increase our precision, but decided that it was not necessary. For example, we could have used blocked random assignment, which is arranging of experimental units in groups that are similar to one another, but we did not see any covariates that could play a factor in affecting our results. Initially, we thought about blocking by major because we believed political science majors would be more interested in civic engagement in comparision to STEM majors; however, we realized that many first year students were undecided, making blocking inappropriate.

We ran a balance check across residential colleges and whether students are on leave, followed by a check for treatment noncompliance. Treatment noncompliance for our study was defined as the rate at which participants did not open the email at all, given they would not receive the treatment.

We used Mailchimp and the Mailchimp API (through the `chimpr` package in R) to contact students and track several important variables like the email open rate, the time of opening, whether students opened the email twice, and whether the students clicked the link for the Google Form where they could sign up for our "event." Finally, we ran our main analyses to study whether any treatments had a significant effect on form sign-ups. We specifically ran a set of t-tests with Bonferroni corrections search for differences across the four treatment groups, and then ran a linear regression with only the static-only and animated-only groups.

### Balance check

To confirm that our randomization process was sound, we ran a balance check across residential college. If the treatment was distributed completely randomly, within each residential college, an even percentage of students should have received the treatment.

As we see from Table 1 and Figure 2 below, though there are some discrepancies where some colleges had a disproportionate amount of students in one treatment group over another, we are overall reassured that the treatment was distributed randomly across college.

Table 1: Percentage in treatment group across each residential college

| Residential College | Both Animated | Static Bar, Animated Line | Animated Bar, Static Line | Control (No animation) |
|---|---|---|---|---|
| Benjamin Franklin | 23.71% | 28.87% | 23.71% | 23.71% |
| Berkeley | 24% | 29.33% | 25.33% | 21.33% |
| Branford | 23.6% | 17.98% | 24.72% | 33.71% |
| Davenport | 23.81% | 21.9% | 24.76% | 29.52% |
| Ezra Stiles | 22.89% | 27.71% | 25.3% | 24.1% |
| Grace Hopper | 30.68% | 18.18% | 21.59% | 29.55% |
| Jonathan Edwards | 21.18% | 31.76% | 24.71% | 22.35% |
| Morse | 31.46% | 21.35% | 19.1% | 28.09% |
| Pauli Murray | 23.64% | 27.27% | 28.18% | 20.91% |
| Pierson | 23.86% | 27.27% | 22.73% | 26.14% |
| Saybrook | 23.17% | 26.83% | 26.83% | 23.17% |
| Silliman | 24.1% | 19.28% | 34.94% | 21.69% |
| Timothy Dwight | 25.53% | 25.53% | 26.6% | 22.34% |
| Trumbull | 26.32% | 27.37% | 22.11% | 24.21% |

### Treatment noncompliance

We then checked for treatment nonconcompliance; as we see in Figure 3, treatment noncompliance was unexpectedly a problem for our research design. Because the subject line of the emails were the same across all four treatment groups and each email was sent at the same time, we did not expect any difference to

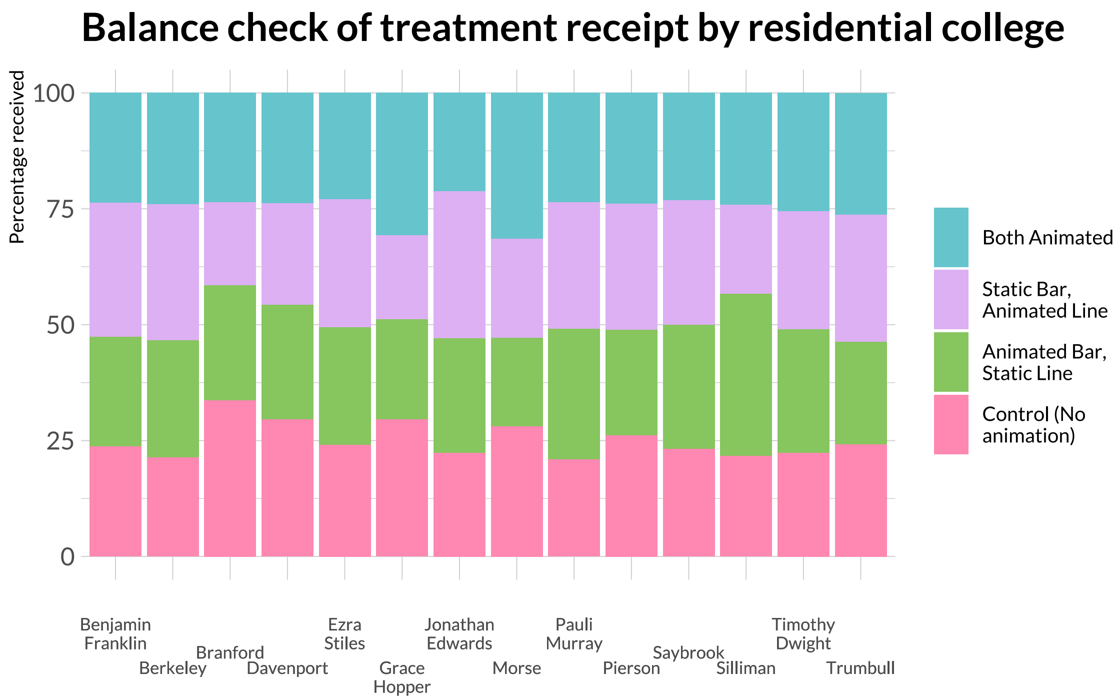


Figure 2: Our balance check in graphical form.

appear across the four treatment groups in terms of treatment receipt. However, as the chart below shows, the treatment group receiving an email with an animated bar chart and a static line chart opened the initial email at a much lower rate than the other three treatment groups. Just as confusingly, this trend disappears in the follow-up email.

We aren't sure what happened here; as some classmates noted in their final project presentation, Yale SIS or Gmail could have labeled our email as spam in this case and many students may not have received the treatment. This still would not explain why only one treatment group was marked as spam, or why this did not happen with our follow-up emails. We proceed onward with a grain of salt at the ready, and limit our study from here to specifically those candidates that opened at least one email.

# Results

### Summary statistics of response variables

Some initial summary statistics in Table 2 show a far bigger "weakness" of our study: engagement with the email to fill out the form was far too low to be able to discern a statistically significant test of whether animated graphics assist with event registration. We present three metrics here of engagement: whether students returned to the email a second time, whether students clicked on the form, whether they actually filled out the form. Only the metric of whether the students actually filled out the form would be relevant for a club aiming to boost event registration.

As we see from these statistics, although a number of people returned to either email a second time, far fewer people actually clicked on the form and an even smaller number of people filled out the form. Trends from the three metrics also do not match each other, given the animated bar and static line treatment group opened the follow-up email a second time at a noticeably lower rate than other groups but had the highest percentage of form clicks and form fills.

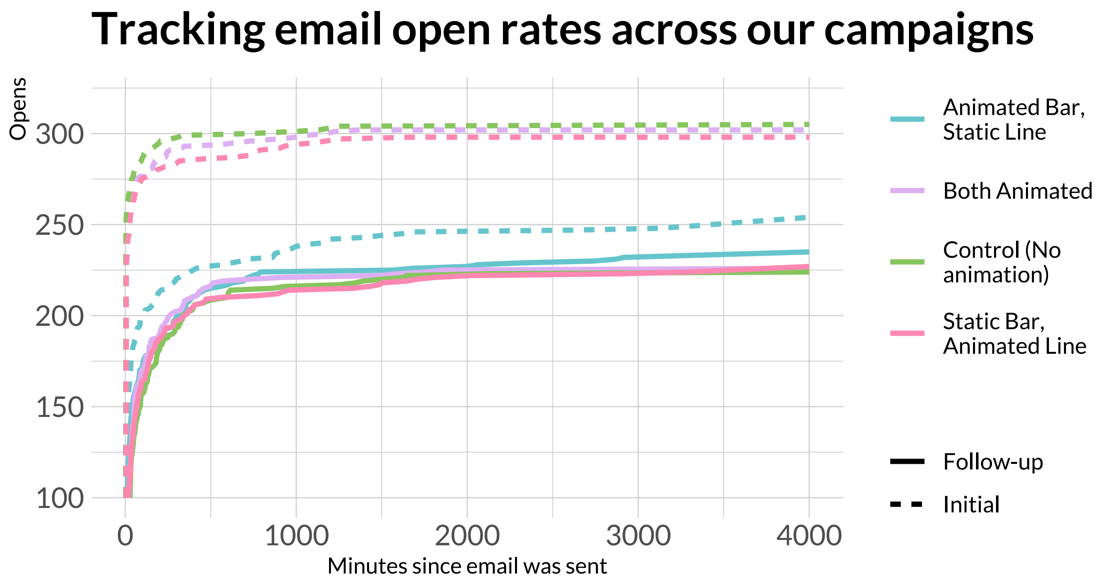


Figure 3: Cumulative opens since the emails were sent.

These statistics also suggest that lack of engagement is not the only issue in our study. Even for the proportion opening the emails a second time, where there is a substantial amount of data, a treatment effect is still not really present. The slight difference from the control group to the both-animated group that appears is also confusingly in the opposite direction as we hypothesized.

Table 2: Summary of Response Statistics Across Treatment Groups

| | Treatment Group | Opened either email a second time | Form click rate | Form fill rate |
|---|---|---|---|---|
| 1 | Control (No animation) | 49.72% | 0.95% | 0.76% |
| 2 | Animated Bar, Static Line | 33.13% | 2.53% | 0.82% |
| 3 | Static Bar, Animated Line | 44.19% | 0.63% | 0.19% |
| 4 | Both Animated | 46.21% | 0.32% | 0.19% |

**Pairwise comparisons and linear regression**

Going further with a set of pairwise t-tests with Bonferroni corrections (Table 3), we confirm our suspicions that there is no significant difference in form response rates across all four treatment groups. A linear regression (Table 4) using just the control group that received only static images and the both-animated group, which performs a similar test without the Bonferroni correction, returns a similar result that animations did not have a significant impact on event registration, be it positive or negative, compared to emails without graphics.

**Power analysis**

Evidently, our research design did not find a significant difference between the animation group and the control group. But instead of being a problem of how our treatment was implemented, it could have been simply an issue of not having enough participants. We purposefully cast quite a large net (the entire first

Table 3: P-values from pairwise comparisons using t-tests, pooled SD, and Bonferroni corrections

| | Comparison Group | Control (No animation) | Animated Bar, Static Line | Static Bar, Animated Line |
|---|---|---|---|---|
| 1 | Animated Bar, Static Line | 1 | NA | NA |
| 2 | Static Bar, Animated Line | 1 | 0.8995 | NA |
| 3 | Both Animated | 1 | 0.892 | 1 |

Table 4: Model-based comparison among just animated and static (control) treatment groups.

| | Fill rate |
|---|---|
| Treatment Group = Both Animated | $-0.006$ |
| | (0.004) |
| Constant | $0.008^{**}$ |
| | (0.003) |
| Observations | 1,057 |
| $R^2$ | 0.002 |
| Adjusted $R^2$ | 0.001 |
| Residual Std. Error | 0.069 (df = 1055) |
| F Statistic | 1.803 (df = 1; 1055) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

year class, a total of 1,265 students) to avoid this issue, but maybe a future experiment could go even larger and thus find significance. We ran a power analysis to find how increasing sample size could help our search for power, specifically testing fifteen sample sizes from 2,000 to 20,000 participants and a hypothesized treatment effect of .006, with 200 simulated experiments each.

As we see in Figure 4, although brute-forcing the experiment and increasing sample size would eventually lead us to a power above .8, such an experiment would likely require around 10,000 participants. In addition to being practically difficult, it also likely would still be theoretically strange (as the current observed effect favors static images over animated images in terms of student engagement and event registration) and the small magnitude in this experiment would mean this effect is probably inconsequential even if statistically significant. Instead of increasing the sample size, future experiments on this question should change research design as noted below.

# Conclusion

With the information given, we were unable to provide enough evidence that animated graphics increase the effectiveness of email-based event recruitment at Yale. Regardless of which treatment was assigned to the subjects, it does not look like animated graphics had a significant effect on recruitment/RSVPS.

The experiment had many limitations, some of which were out of our control. Yale Students for Civic Engagement was a random club that nobody has heard of. We didn't have any name recognition prior to the emails that we sent to first years in late October; thus, many students were probably not interested in other club activities. Also, with clubs such as Yale democrats, the federalist party, and the centrist party being popular on Yale's campus, our club's topic of civic engagement may have been redundant or not engaging to first years. If we were to repeat this experiment again, we would have established the club around the time students were looking to join clubs earlier in the fall to make our experiment "more fieldy".

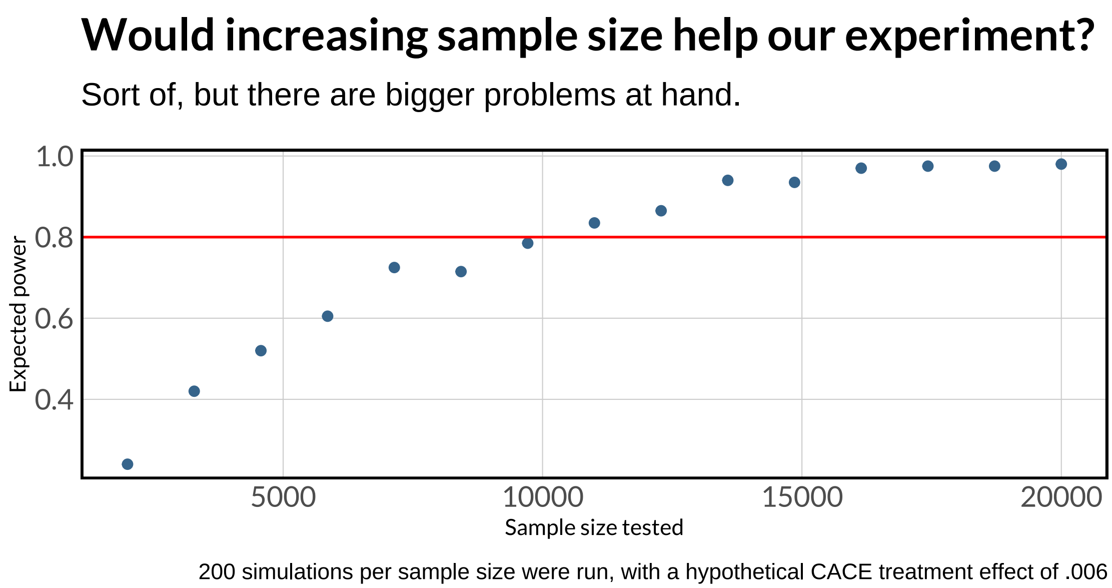


Figure 4: Power analysis.