Analysis of 54 Activate Good Weekly Newsletters, 2019-2020

STATCOM: Alvin Sheng, Naomi Giertych, Jimmy Hickey

Abstract

There are four main sections in this final report: 1) Exploration of Aggregated Metrics, 2) Trends in the Open Probability, 3) Trends in the Click Probability, and 4) Link Characteristics that Promote Clicks. Each main section is headed by an executive summary that describes i) the questions of interest that the main section aims to answer, ii) the statistical methods used to answer the questions, and iii) what we believe are the most pertinent takeaways of the main section.

Contents

1 Exploration of Aggregated Metrics: Executive Summary

1.1 Questions of Interest

- What are some general trends regarding the weekly newsletters' aggregated metrics (open, click, bounce and unsubscribe rates)?

1.2 Statistical Analysis

- Correlations between the metrics
- Scatterplots of the metrics vs. various factors

1.3 Takeaways

- Clicks are positively correlated with opens, but negatively correlated with unsubscribes.
- The start of the COVID pandemic seemed to have caused a temporary spike in engagement.
- The next two reports, "Trends in the Open Probability" and "Trends in the Click Probability," will delve further into how various factors affect the open and click rate.

2 Exploration of Aggregated Metrics: Report

2.1 Overview of the Data Used

We focus on 54 newsletters, sent roughly weekly in the years 2019 and 2020. For this initial report, we look at four aggregated metrics for the newsletters: the open, click, bounce, and unsubscribe percentages.

The definition of the metrics are as follows:

$$\text{Open } \% = \frac{\text{number of contacts who opened the newsletter}}{\text{number of contacts sent to}} \times 100\%$$

$$\text{Click } \% = \frac{\text{number of contacts who clicked a link in the newsletter}}{\text{number of contacts sent to}} \times 100\%$$

$$\text{Bounce } \% = \frac{\text{number of contacts who could not be reached}}{\text{number of contacts sent to}} \times 100\%$$

$$\text{Unsubscribe } \% = \frac{\text{number of contacts who used newsletter to unsubscribe}}{\text{number of contacts sent to}} \times 100\%$$

We also examine the following characteristics of the newsletters:

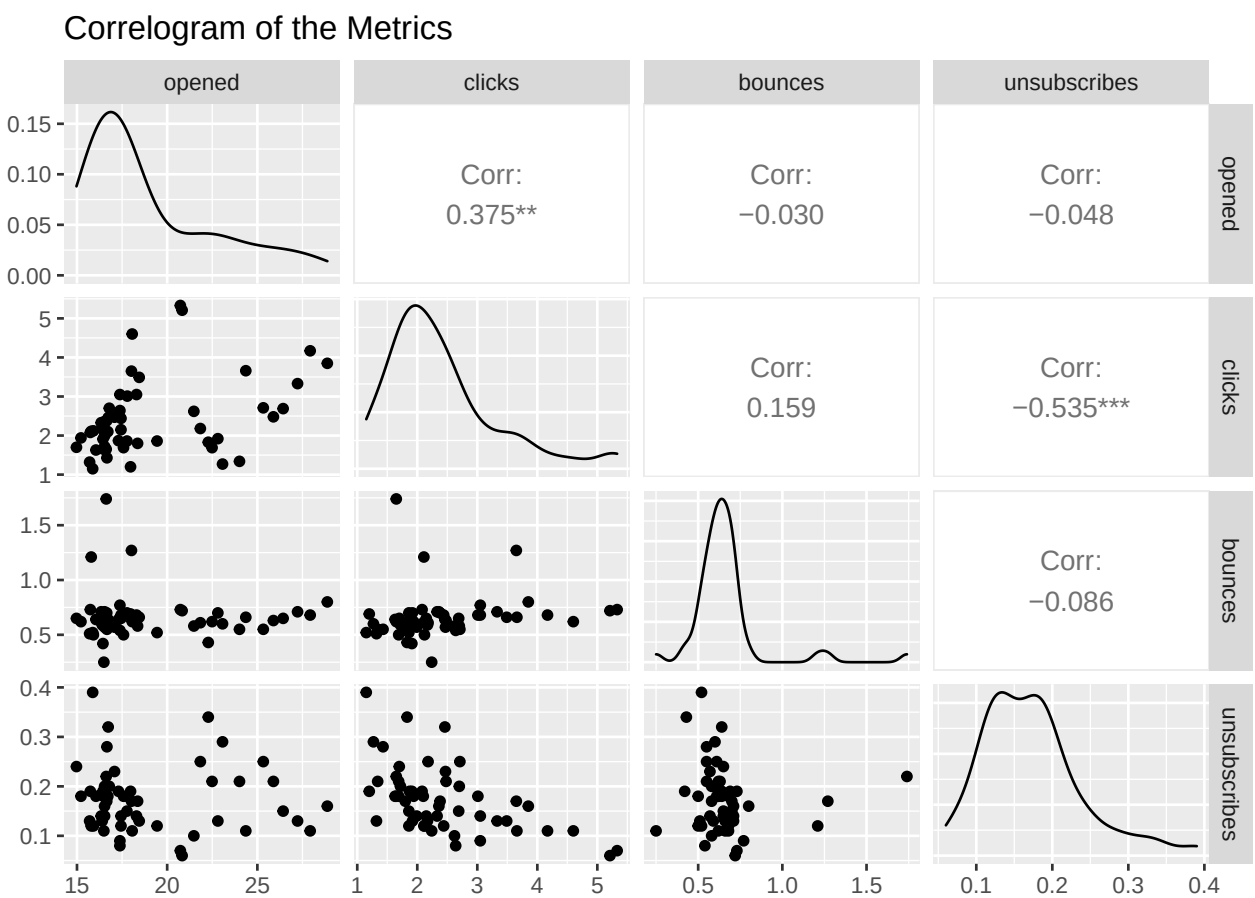
- Trend over the date newsletter was sent out (2019-01-01 to 2020-12-31).
- Whether newsletter was sent before or after start of the COVID pandemic on 2020-03-12.
- Time of day newsletter was sent out (6:30 am to 8:40 pm).
- Length of subject by number of characters.

Below are 5 example newsletters from the dataset, i.e. the 5th, 8th, 25th, 35th, and 47th newsletters sent out.

	datetime	opened	clicks	bounces	unsubscribes
5	2019-05-01 15:21:00	17.97	1.20	0.69	0.19
8	2019-08-05 07:50:00	16.66	1.43	0.55	0.28
25	2020-06-05 07:50:00	17.38	2.64	0.54	0.08
35	2020-08-19 07:05:00	16.53	1.99	0.57	0.14
47	2020-11-11 06:51:00	18.30	3.05	0.68	0.14

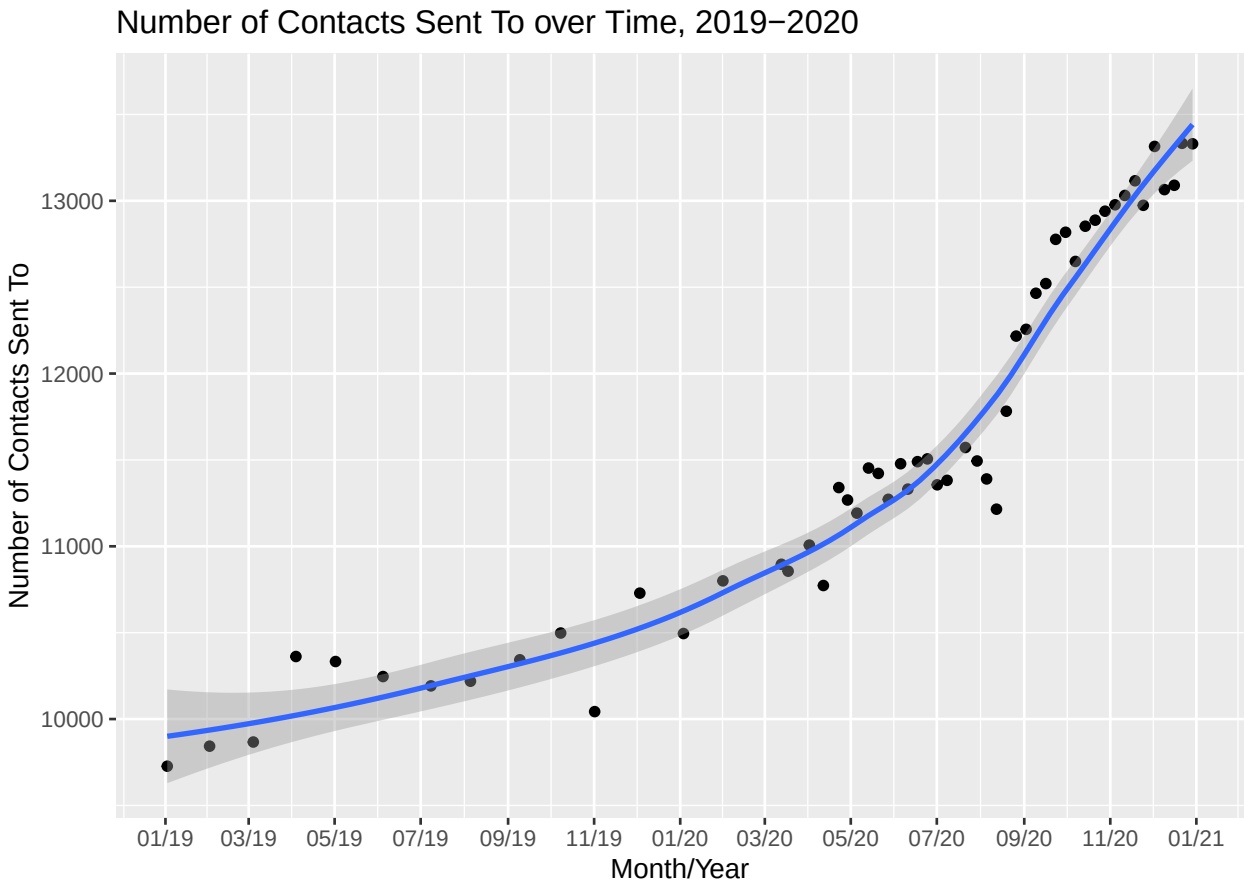
	contacts_sent_to	covid	mins_since_midnight	subject_length
5	10333	Before	921	75
8	10219	Before	470	79
25	11478	After	470	53
35	11782	After	425	76
47	13031	After	411	68

2.2 Correlation of the Metrics

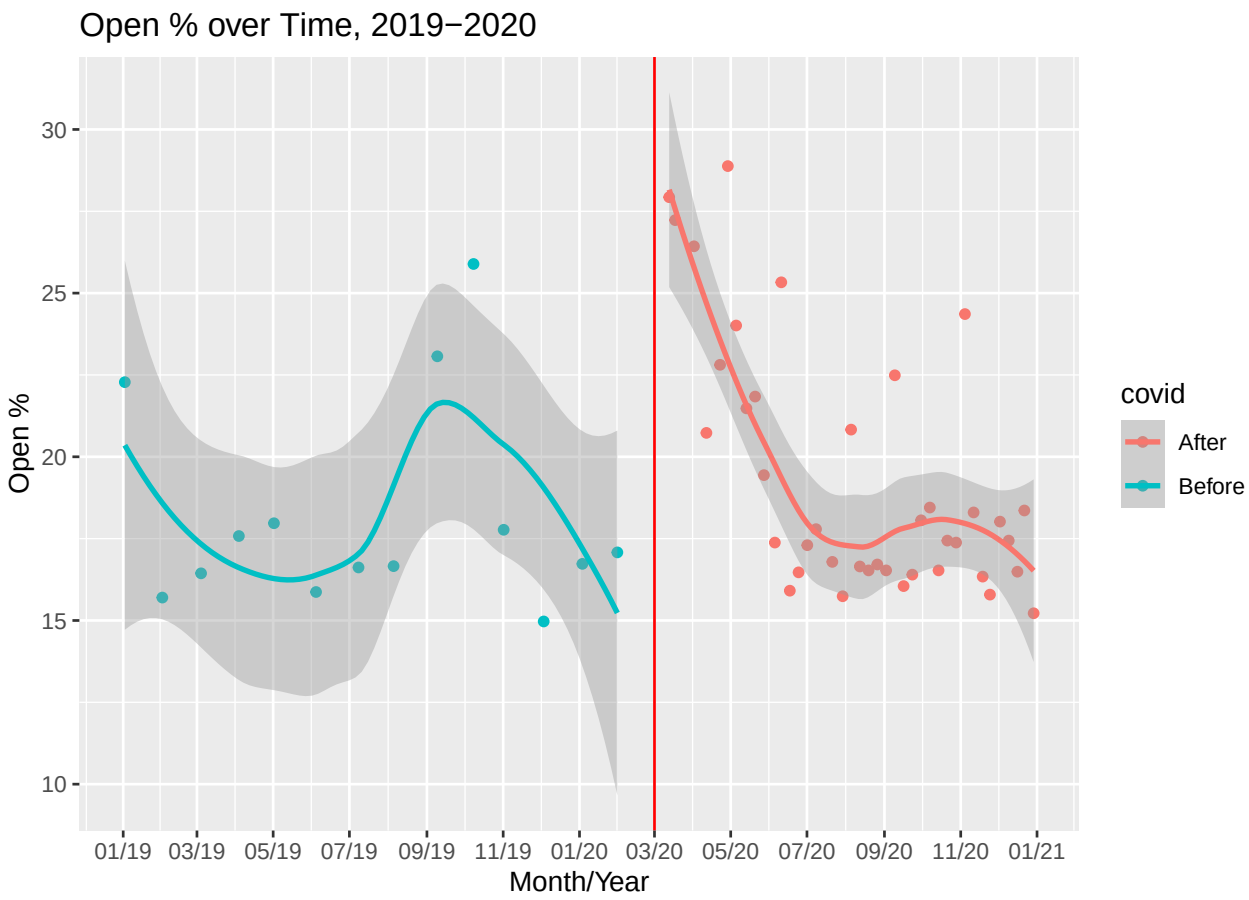


There are two significant correlations: a positive one between opens and clicks, and a negative one between unsubscribes and clicks.

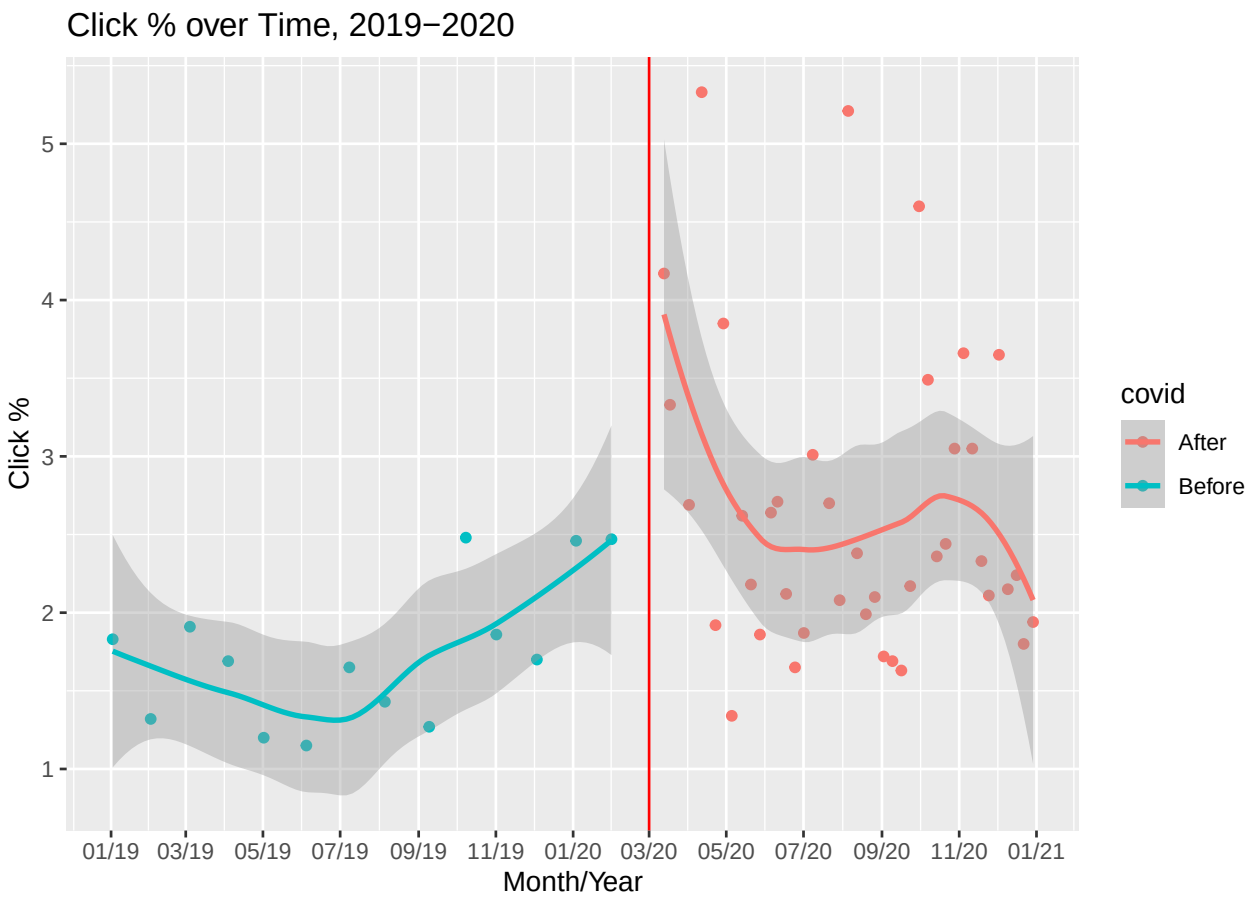
2.3 Summary Statistics over Time, 2019-2020

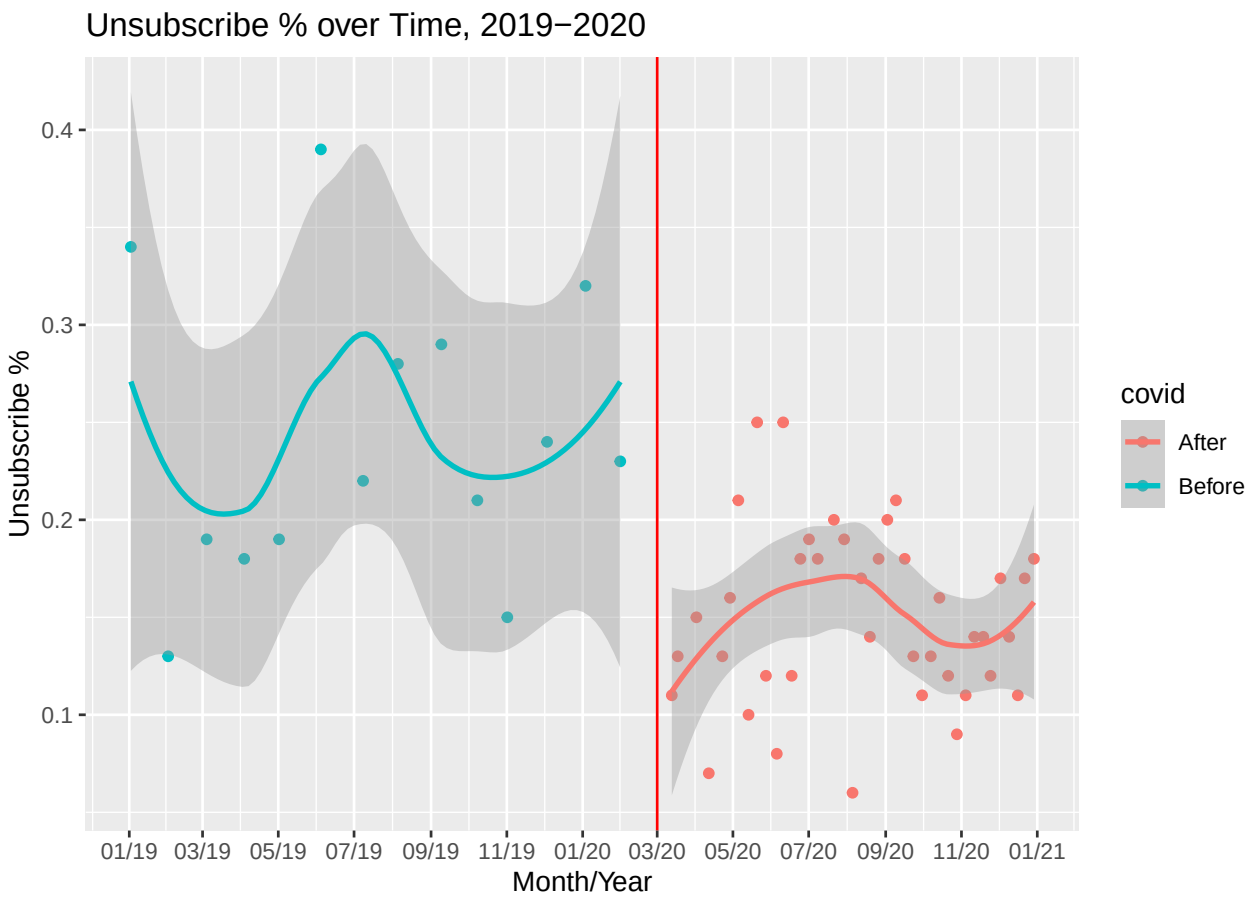


When did the pandemic start changing things? The March 12 weekly newsletter was the first one to mention the COVID-19 pandemic and remote volunteering opportunities.

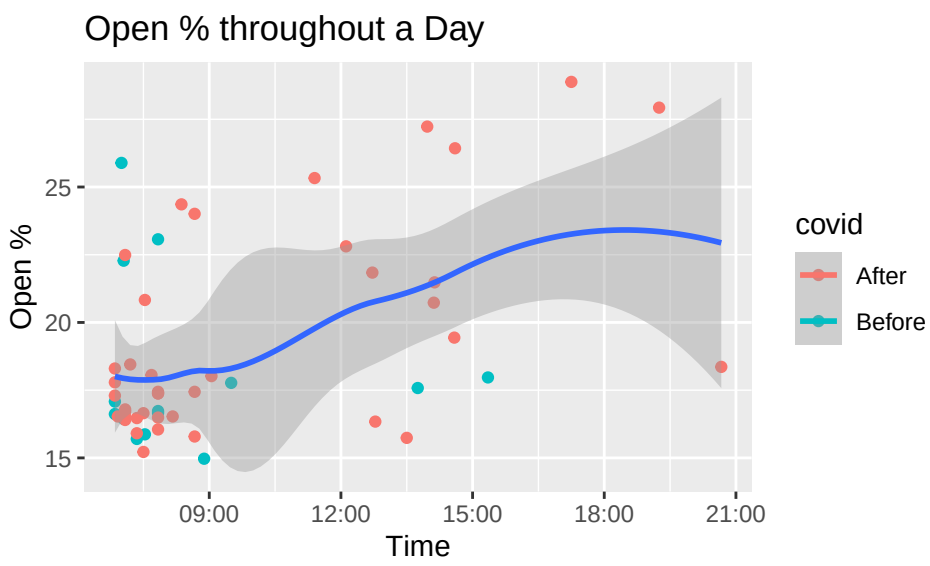


There is a spike in the open % after March.


Click % over Time, 2019−2020
Click %
5
4
3
2
1
01/19
03/19
05/19
07/19
09/19
11/19
01/20
03/20
05/20
07/20
09/20
11/20
01/21
Month/Year
covid
After
Before

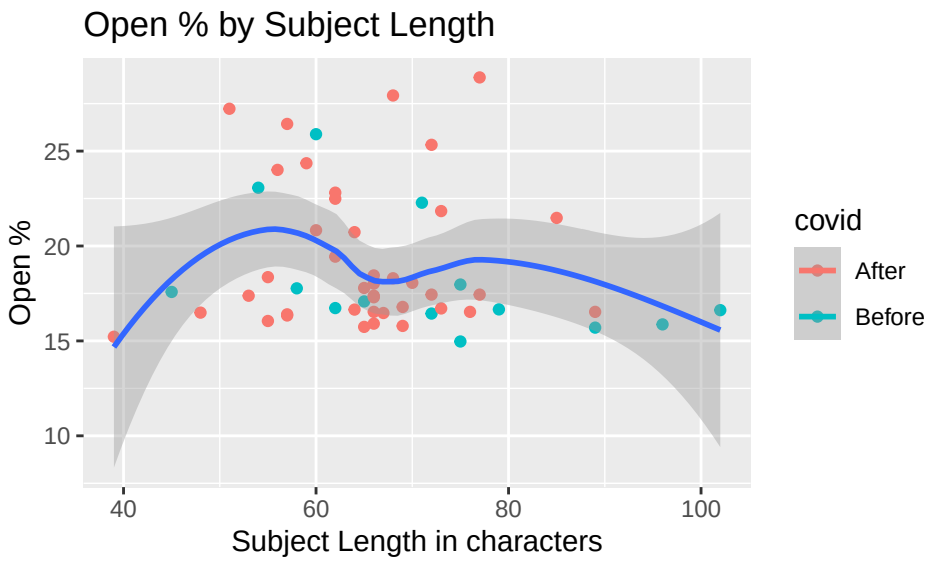


2.4 Summary Statistics throughout a Day



There may be a slight upward trend.

2.5 Effect of Subject Headings



3 Trends in the Open Probability: Executive Summary

3.1 Questions of Interest

- What are the factors that increase the probability of a subscriber opening a newsletter within a week of its send date?
  - What is the best time of day to send a newsletter?
  - Does the email subject affect the open rate?
- How has the COVID pandemic affected the open rate?

3.2 Statistical Analysis

- Barplots to visually display trends
- Statistical model to examine how factors interact
- Results from the model to confirm the trends shown in the barplots

3.3 Takeaways

- Before the pandemic, there was a dip in the open probability when the newsletter was sent at around 10 am. During the pandemic, 10:30 am and 5:10 pm seem to be optimal times for sending the newsletter.
- Shorter subject headings (in terms of number of characters) are better.

4 Trends in the Open Probability: Report

4.1 Overview of the Data Used

There are 16,291 unique subscribers in the data set, with 622,614 observations.

We focused on the probability that a given subscriber will open a weekly newsletter within a week of its sent date. If the subscriber opens the newsletter after a week or uses the newsletter to unsubscribe, we consider it a non-open. We examined several factors affecting the open probability:

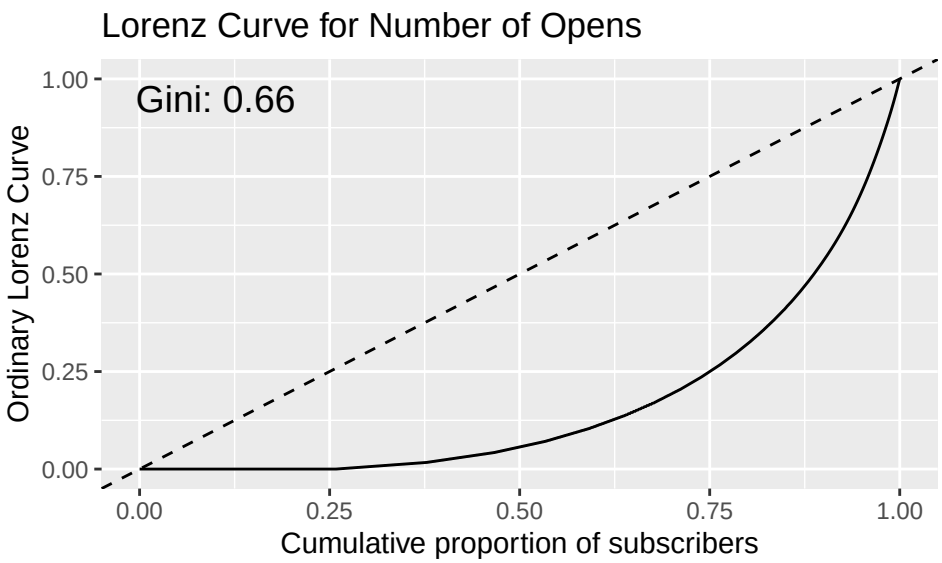
- Trend over the date newsletter was sent out (2019-01-01 to 2020-12-31).
- Whether newsletter was sent before or after start of the COVID pandemic on 2020-03-12.
- Time of day newsletter was sent out (6:30 am to 8:40 pm).
- Length of subject by number of characters.

Below are 10 sample observations from the dataset. Each row corresponds to one of the 622,614 newsletter-subscriber pairs. The last variable, week_open, is the response variable of interest. 1 indicates that the subscriber opened the newsletter within the week; 0 indicates that the subscriber received the newsletter but didn't open it.

date_sent	subscriberid	covid	mins_since_midnight	subject_length	week_open
2020-07-08 06:51:04	70392049	After	411	65	0
2020-07-21 07:05:50	67928039	After	425	69	0
2020-11-18 12:47:35	71244202	After	767	57	0
2020-12-29 07:30:43	70293552	After	450	39	1
2020-06-17 07:21:09	70811658	After	441	66	0
2020-09-23 07:05:42	71567218	After	425	57	1
2020-07-29 13:30:36	67927884	After	810	65	0
2019-12-03 08:53:03	61625847	Before	533	75	0
2019-12-03 08:53:03	57442167	Before	533	75	0
2020-12-16 07:50:15	71567206	After	470	48	0

4.2 Lorenz Curve

The Gini Index ranges from 0 to 1, with 1 being perfect inequality. In this case, the distribution of opens among the subscribers seems unequal; according to the curve, the top 25% of people account for 75% of opens.



4.3 Overview of Analysis

For every factor of interest, we plotted a barplot to display any trends.

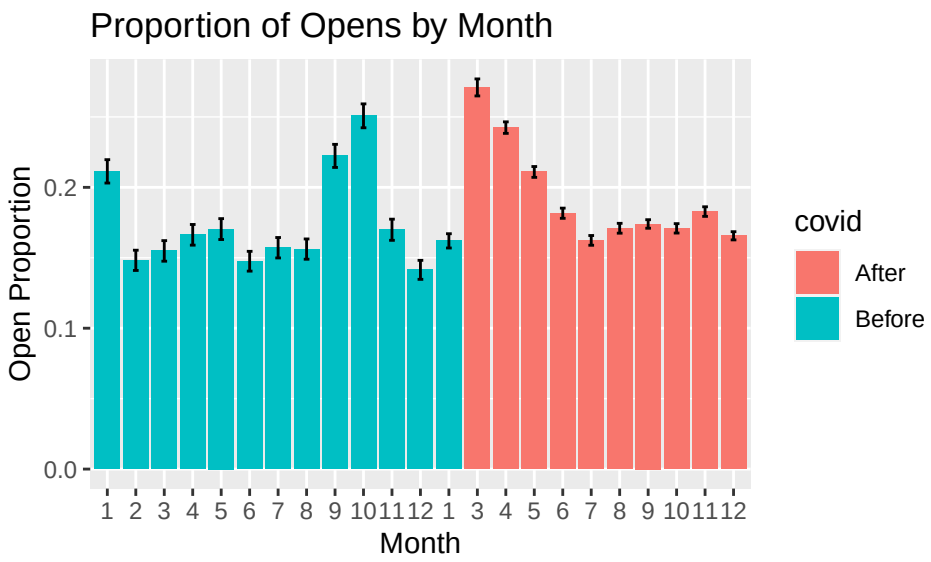
We also fitted a generalized additive mixed model to the data to confirm the trends shown in the barplots. Results from the model are more reliable than just using the barplots, because the model accounts for confounding and dependence between opens for the same subscriber.

The model uses a random sub-sample of 1,629 subscribers (63,897 observations) so it can finish in a reasonable amount of time.

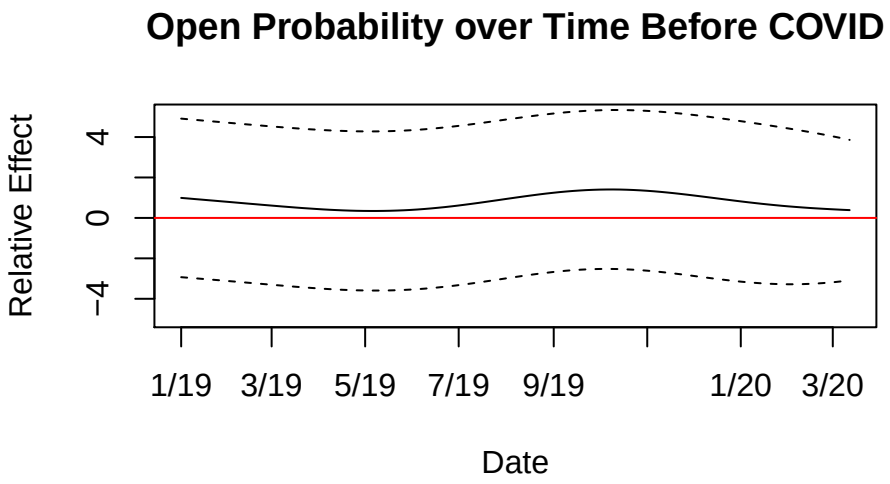
All plots show 95% confidence intervals of the estimates; two standard errors above and below the estimates are indicated on top of the bars in the barplots and by the dashed lines or shading in the line graphs. The true value can be expected to lie within two standard errors from the estimate.

4.4 Trend over Date

The below barplot shows the proportion of subscribers that opened the newsletter, given the month the newsletter was sent to them.

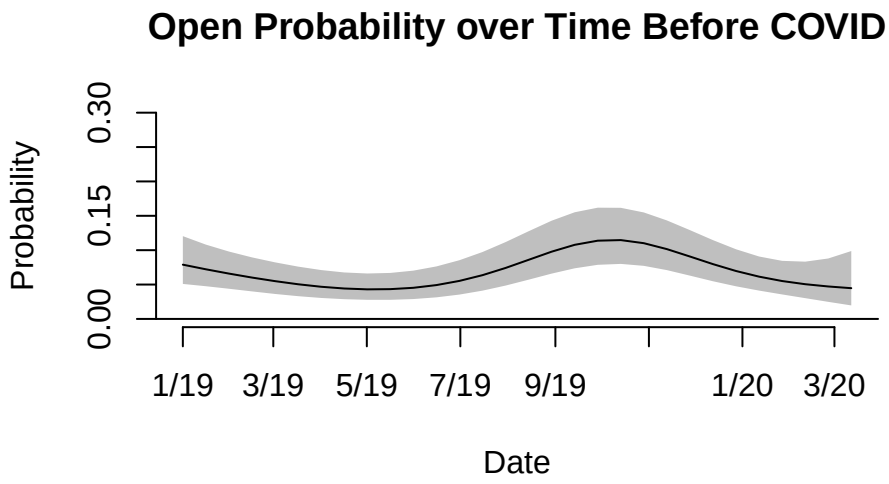


The following plot shows the relative effect of the date the newsletter is sent out on the open probability (a negative relative effect corresponds to a decrease in probability, and a positive relative effect corresponds to an increase in probability) before the pandemic. The red line at zero indicates no effect. There appears to be a seasonal trend.

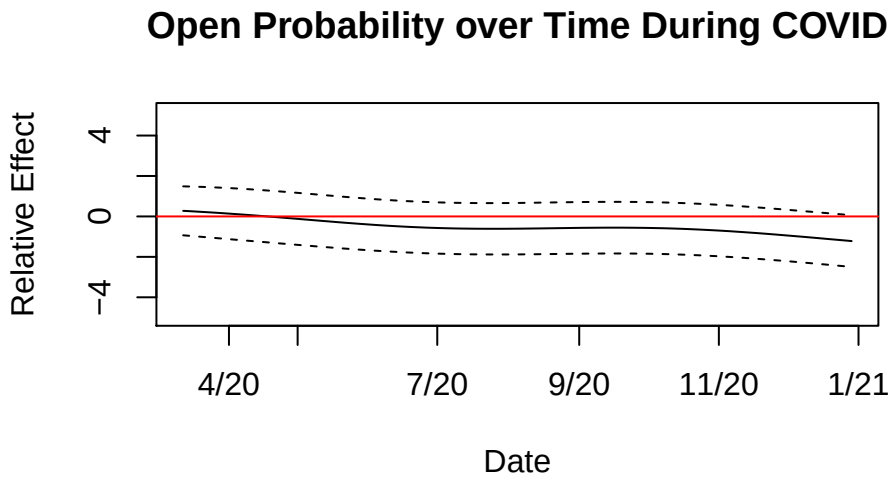


The above plot shows the partial effect of the date alone, without considering other covariates. The following plot shows the actual estimated probabilities over time under the following specific scenario:

- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.

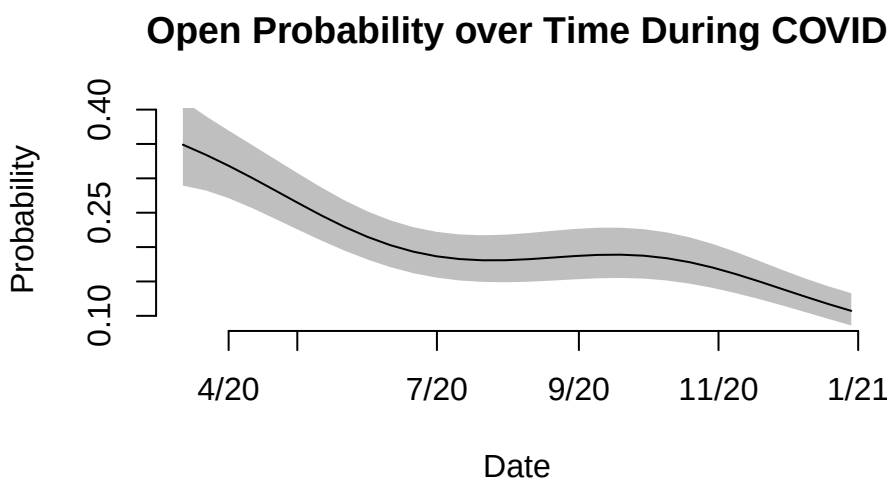


After the pandemic, there is a downward trend in open probability.



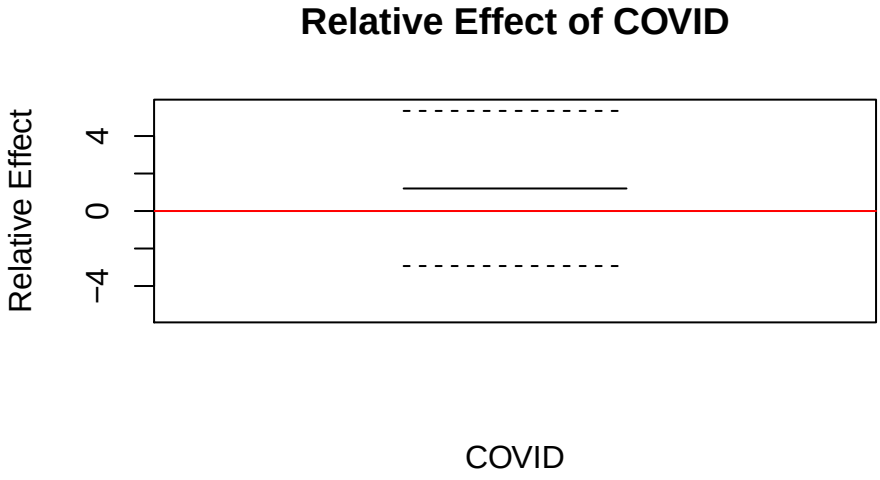
The above plot shows the partial effect of the date alone, without considering other covariates. The following plot shows the actual estimated probabilities over time under the following specific scenario:

- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.



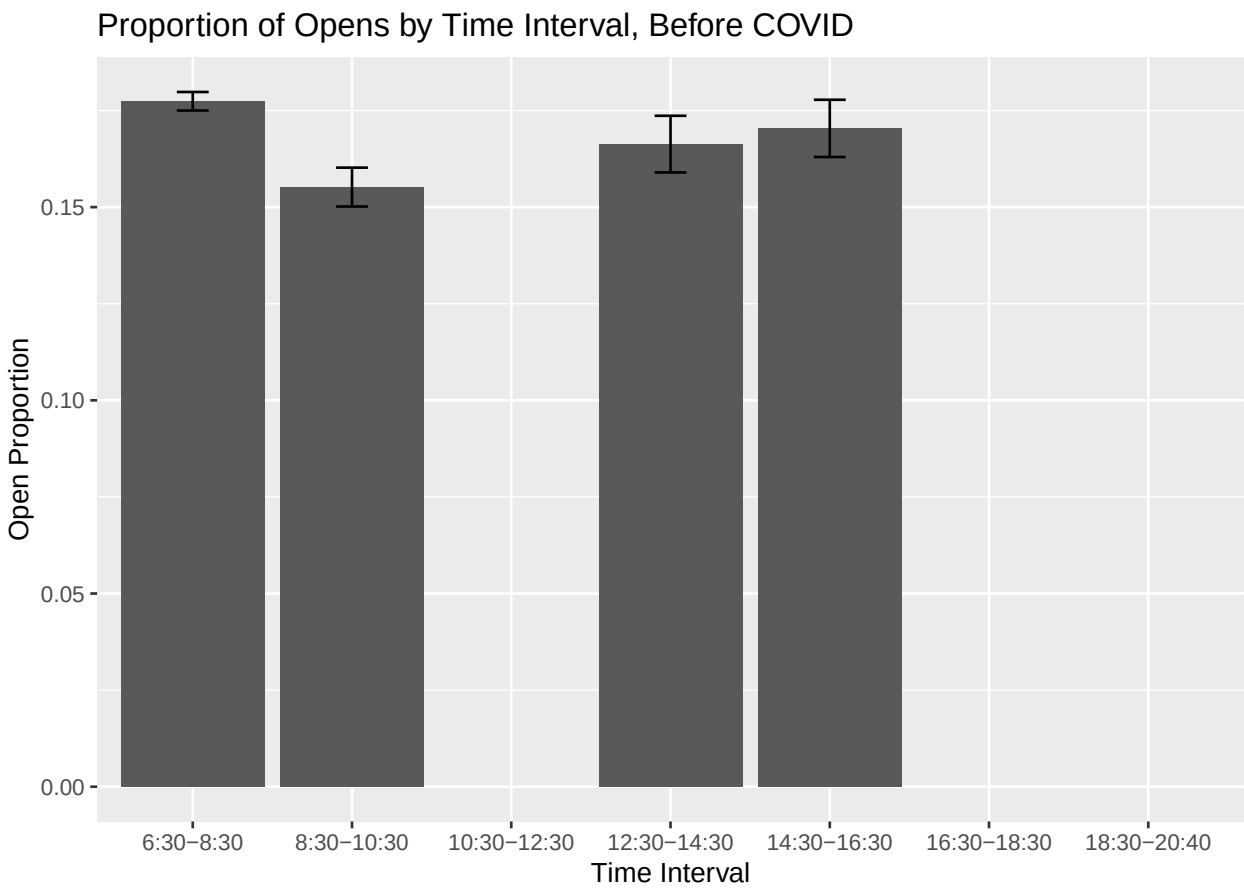
Below shows the relative effect of COVID (solid line), i.e. whether the newsletter was sent after the pandemic started. It appears that the open probability rises after the pandemic starts, but the effect is not significant

(see the dashed standard error bars). However, COVID significantly affects how the open probability varies by date or hour of day the newsletter was sent.

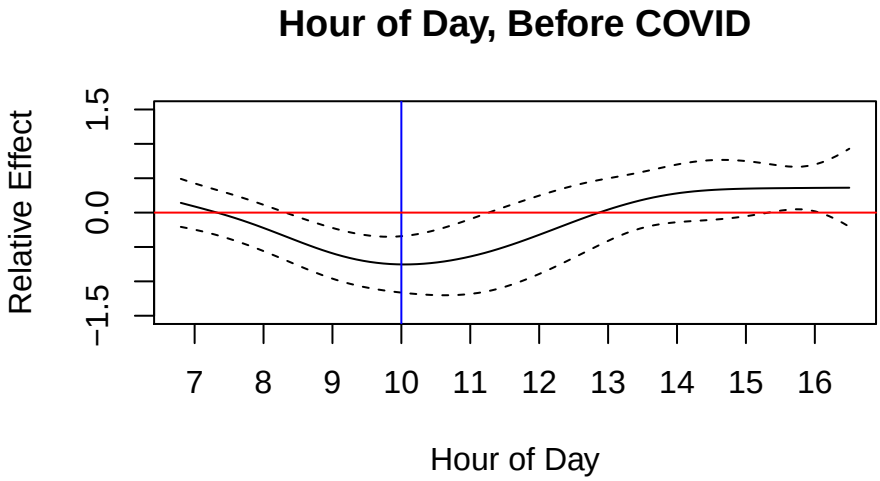


4.5 Time of Day Trend, Before COVID

The below barplot shows the proportion of subscribers that opened the newsletter sent before the pandemic, given that the newsletter was sent to them within a specific time interval. There were no newsletters sent in three of the time intervals, so the bars are absent.

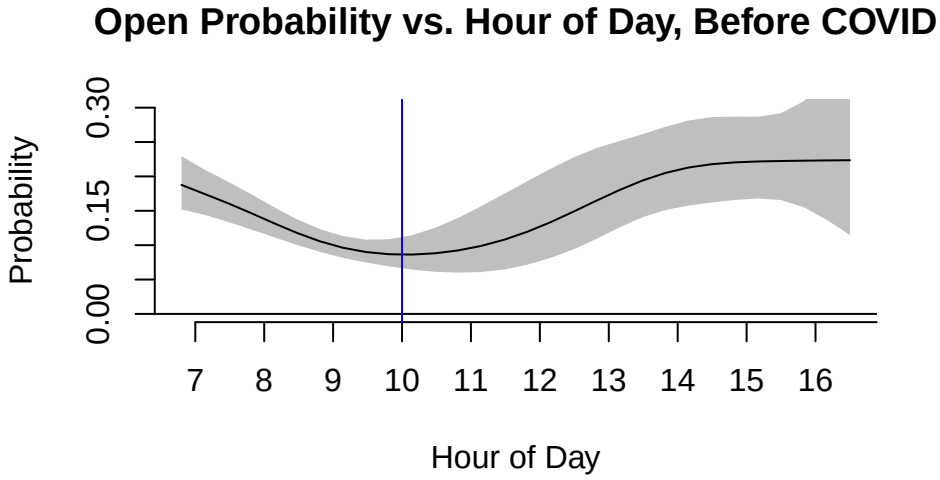


The following plot shows the relative effect of the time of day the newsletter is sent out on the open probability (a negative relative effect corresponds to a decrease in probability, and a positive relative effect corresponds to an increase in probability) before the pandemic. It appears that there is a dip in the open probability at about 10 am.



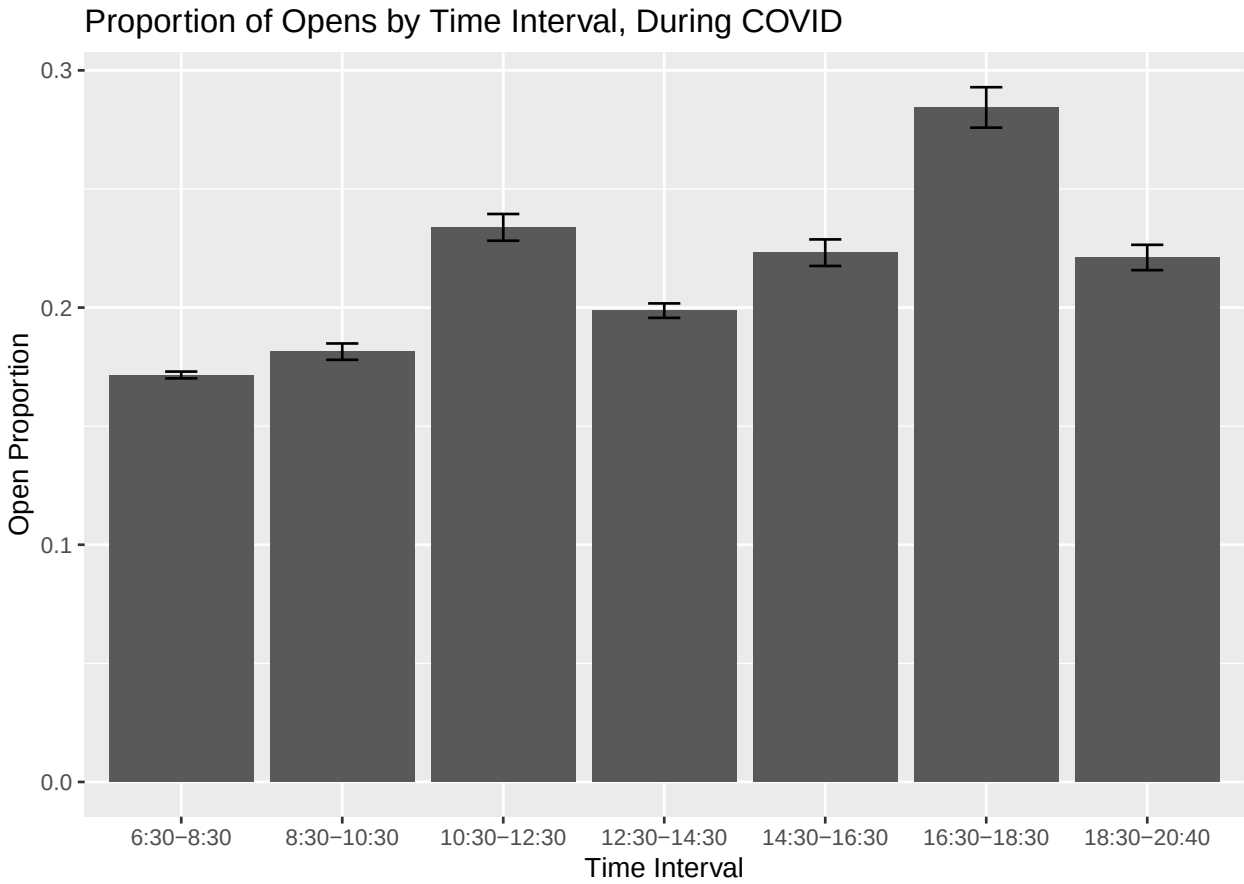
The above plot shows the partial effect of the time of day alone, without considering other covariates. The following plot shows the actual estimated probabilities by time of day under the following specific scenario:

- The newsletter was sent out on December 1, 2019.
- The newsletter has the median subject length of 66 characters.

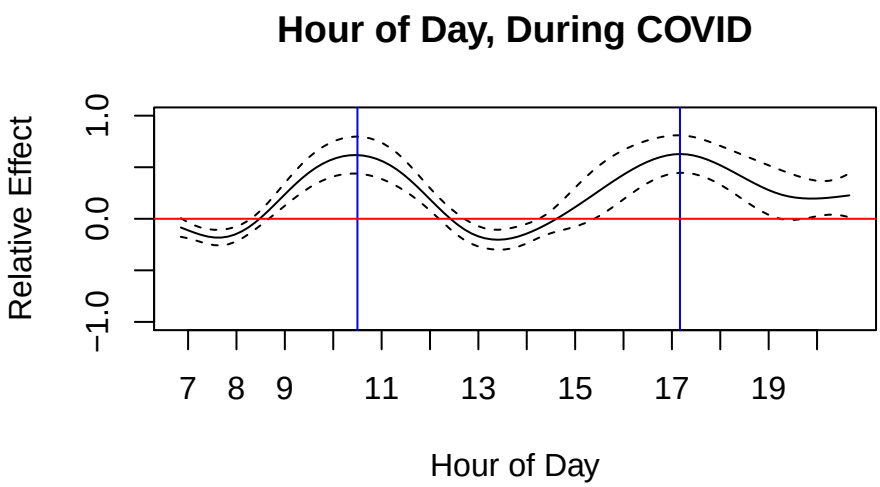


4.6 Time of Day Trend, During COVID

The below barplot shows the proportion of subscribers that opened the newsletter during the pandemic, given that the newsletter was sent to them within a specific time interval. The barplot suggests that there are two time intervals with higher open proportions.

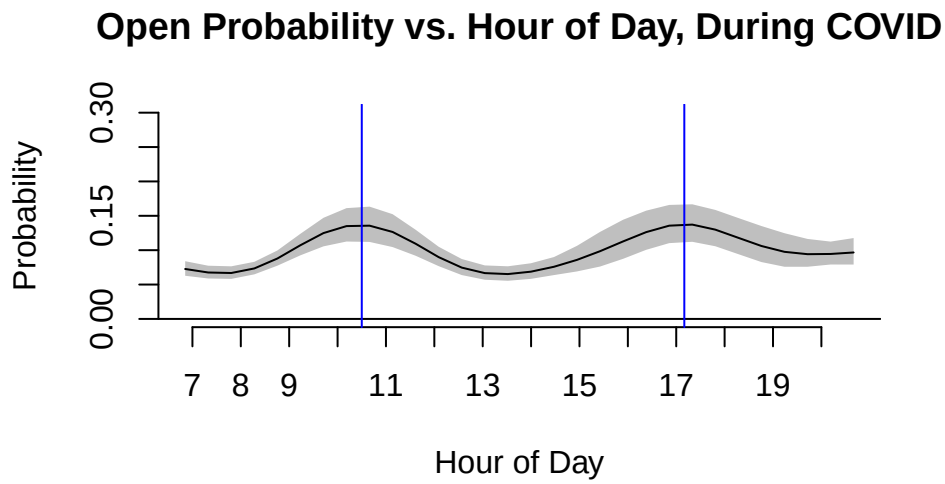


The following plot shows the relative effect of the time of day the newsletter is sent out on the open probability during the pandemic. It appears that the optimal times are about 10:30 in the morning and 17:10 in the evening.



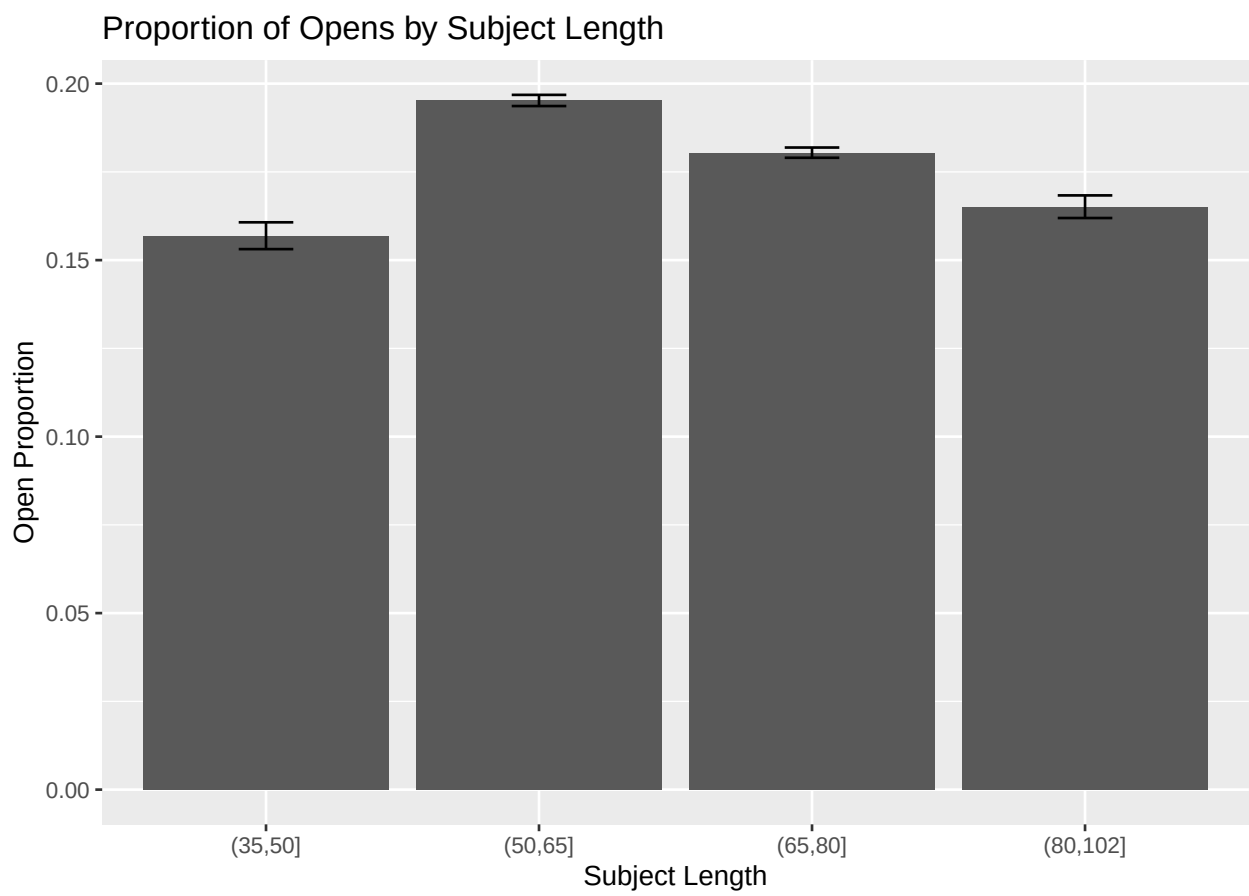
The above plot shows the partial effect of the time of day alone, without considering other covariates. The following plot shows the actual estimated probabilities by time of day under the following specific scenario:

- The newsletter was sent out on December 1, 2020.
- The newsletter has the median subject length of 66 characters.

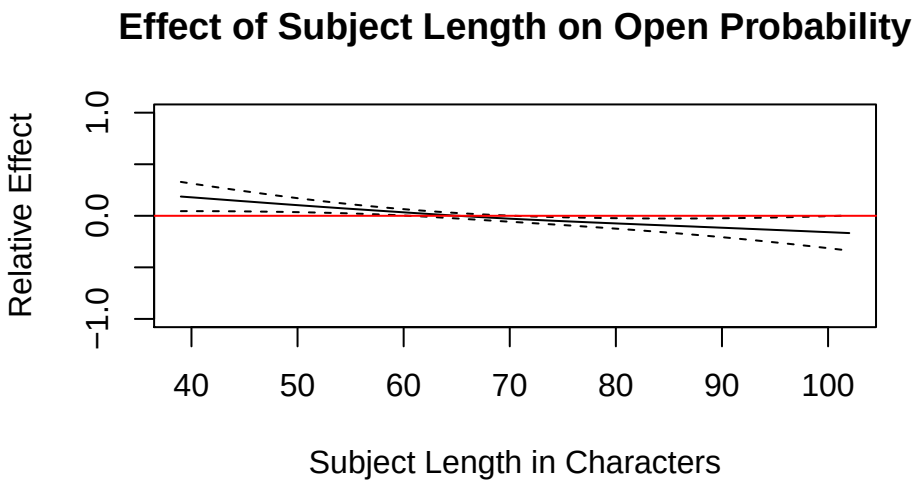


4.7 Subject Length Trend

The below barplot shows the proportion of subscribers that opened the newsletter, given that the subject length was within a specific interval.

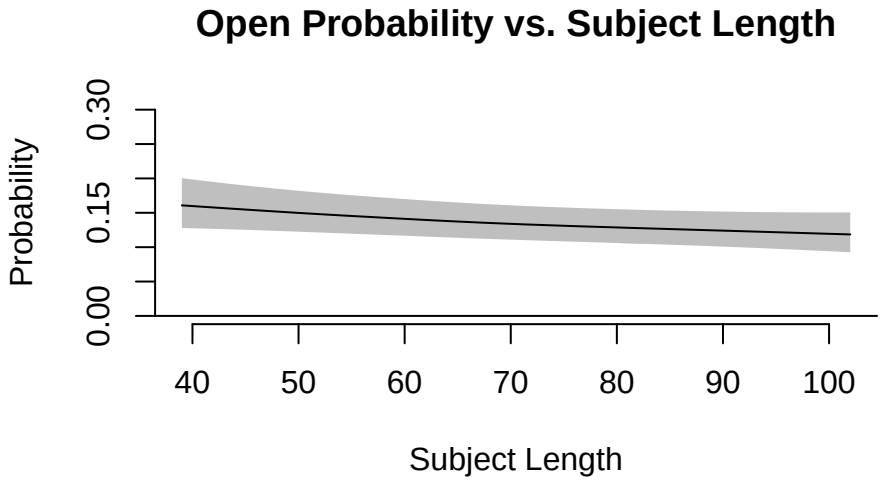


The following plot shows the relative effect of the subject length on the open probability. There appears to be a downward trend in open probability as the subject length increases.



The above plot shows the partial effect of the subject length alone, without considering other covariates. The following plot shows the actual estimated probabilities by subject length under the following specific scenario:

- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.



5 Trends in the Click Probability: Executive Summary

5.1 Questions of Interest

- What are the factors that increase the probability of a subscriber clicking on any link in a newsletter?
  - Does the time of day the newsletter is sent affect the click rate?
  - Does the email subject affect the click rate?
  - Does the content of the newsletter (word count, number of links, number of pictures) affect the click rate?
- How has the COVID pandemic affected the click rate?

5.2 Statistical Analysis

- Barplots to visually display trends
- Statistical model to examine how factors interact
- Results from the model to confirm the trends shown in the barplots

5.3 Takeaways

- Because the click probability is already quite low, changing the time of day or subject length would not affect the click probability as much as it would affect the open probability. Thus, we recommend trying to increase the open rate before trying to change the click rate with those factors.
- Before the pandemic, there were less clicks if the newsletter was sent around 9 am and more clicks if the newsletter was sent around 1 pm. However, during the pandemic, the time of day the newsletter was sent didn't affect the click rate.
- A subject length of 70 characters leads to the most clicks.
- A higher word count is associated with less clicks.
- More text or image links lead to more clicks.
- The optimal number of clickable pictures appears to be 8. More clickable pictures is not necessarily better.

6 Trends in the Click Probability: Report

6.1 Overview of the Data Used

There are 16,291 unique subscribers in the data set, with 622,614 observations.

We focused on the probability that a given subscriber will click on at least one link in a given newsletter. In addition to the variables used in the open probability model,

- Trend over the date newsletter was sent out (2019-01-01 to 2020-12-31).
- Whether newsletter was sent before or after start of the COVID pandemic on 2020-03-12.
- Time of day newsletter was sent out (6:30 am to 8:40 pm).
- Length of subject by number of characters.

we also look at the following factors related to the content within the newsletter:

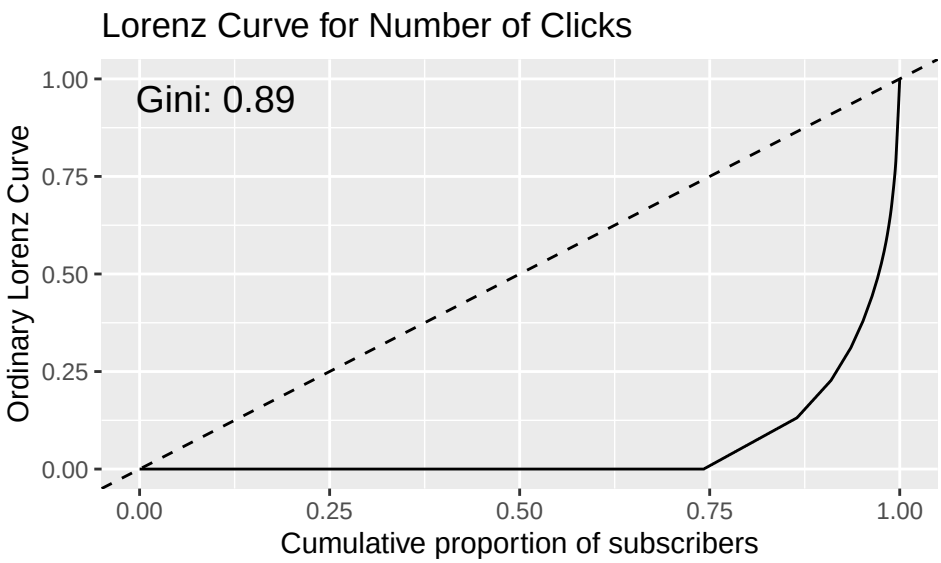
- Number of words in newsletter.
- Number of links (text or image) in newsletter.
- Number of clickable pictures in newsletter.
- Number of unclickable pictures in newsletter.

Below are 10 sample observations from the dataset, containing the four "content" variables mentioned above. Each row corresponds to one of the 622,614 newsletter-subscriber pairs. The last variable, clicks, is the response variable of interest. 1 indicates that the subscriber clicked on any link in a given newsletter; 0 indicates that the subscriber didn't click on any link in a given newsletter.

num_words	num_links	num_clickable_pics	num_unclickable_pics	clicks
347	23	6	0	0
410	23	6	1	0
402	28	8	0	0
512	30	9	0	0
333	20	6	0	0
357	30	13	0	0
382	26	6	0	0
484	31	7	3	0
484	31	7	3	0
513	27	8	0	0

6.2 Lorenz Curve

The Gini Index ranges from 0 to 1, with 1 being perfect inequality. In this case, the distribution of clicks among the subscribers seems unequal; according to the curve, only 25% of subscribers click at all, and the top 10% of people account for 75% of clicks.



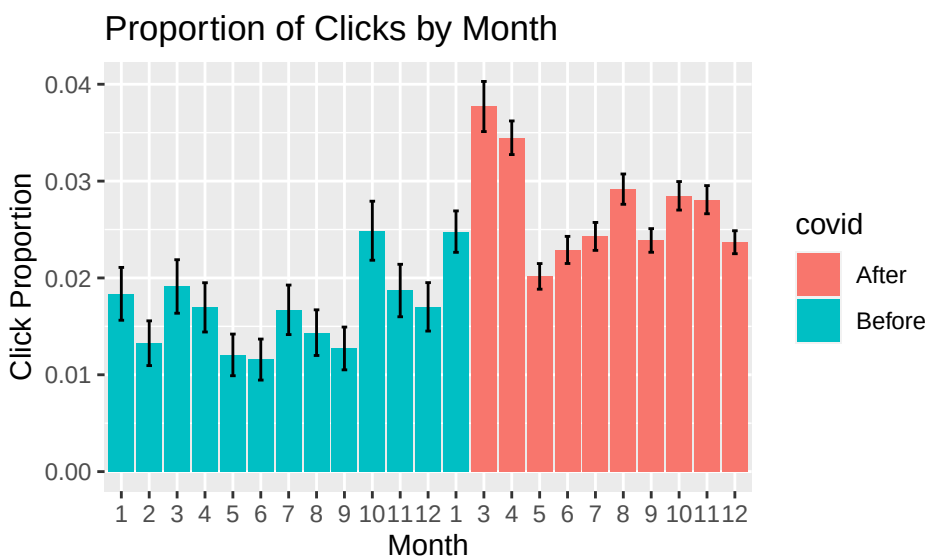
6.3 Overview of Analysis

The generalized additive mixed model uses a random sub-sample of 1,629 subscribers (63,897 observations) so it can finish in a reasonable amount of time.

All plots show 95% confidence intervals of the estimates; two standard errors above and below the estimates are indicated on top of the bars in the barplots and by the dashed lines or shading in the line graphs. The true value can be expected to lie within two standard errors from the estimate.

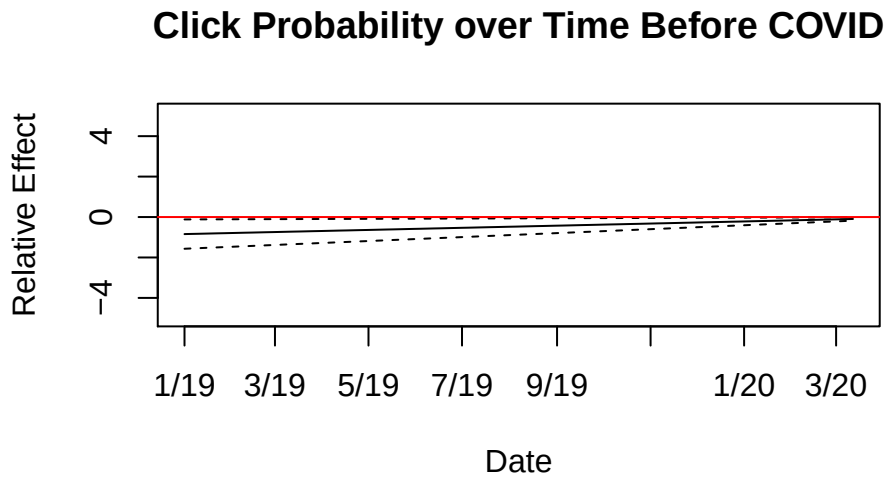
6.4 Trend over Date

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given the month the newsletter was sent to them.



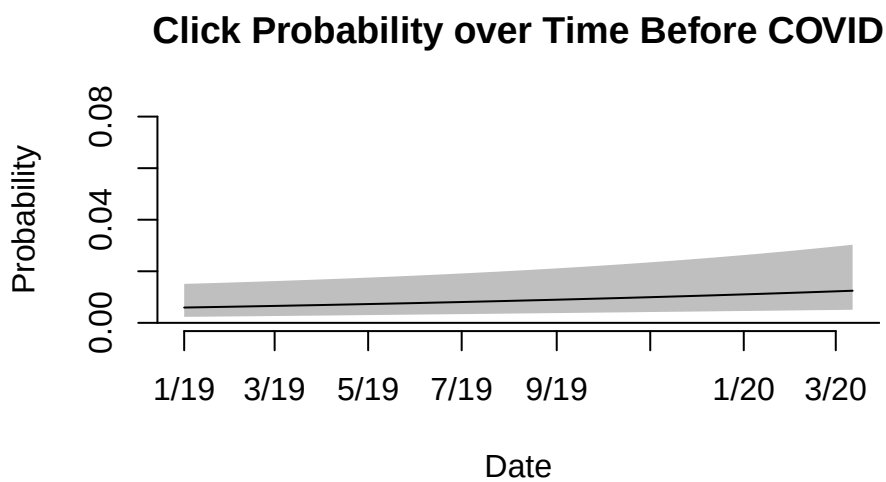
The following plot shows the relative effect of the date the newsletter is sent out on the click probability (a negative relative effect corresponds to a decrease in probability, and a positive relative effect corresponds

to an increase in probability) before the pandemic. There is a slight increase over time leading up to the pandemic.

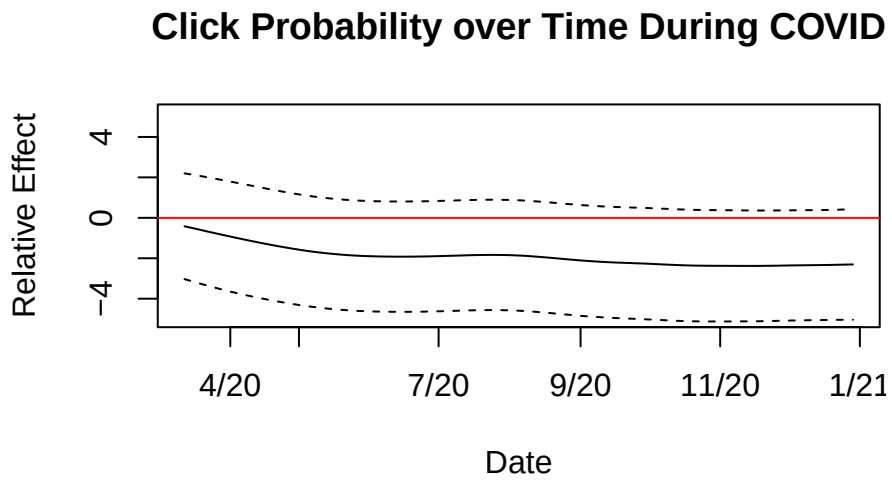


The above plot shows the partial effect of the date alone, without considering other covariates. The following plot shows the actual estimated probabilities over time under the following specific scenario:

- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

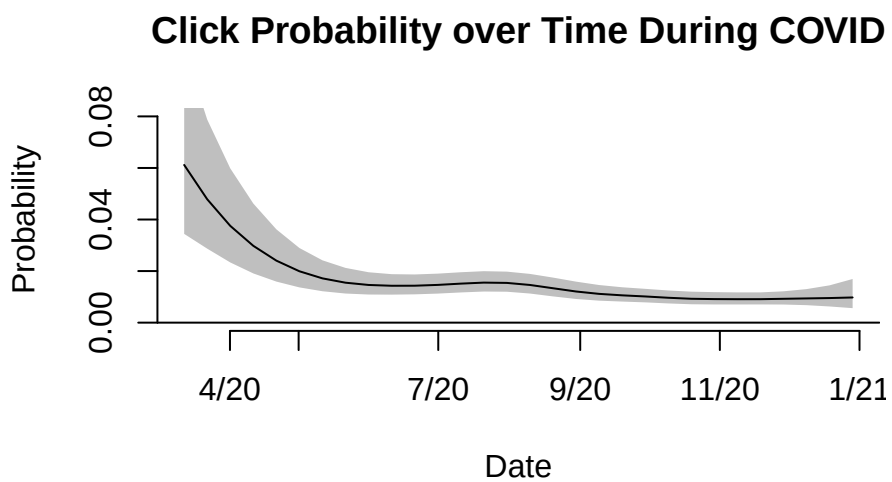


After the pandemic, there is a downward trend in click probability.

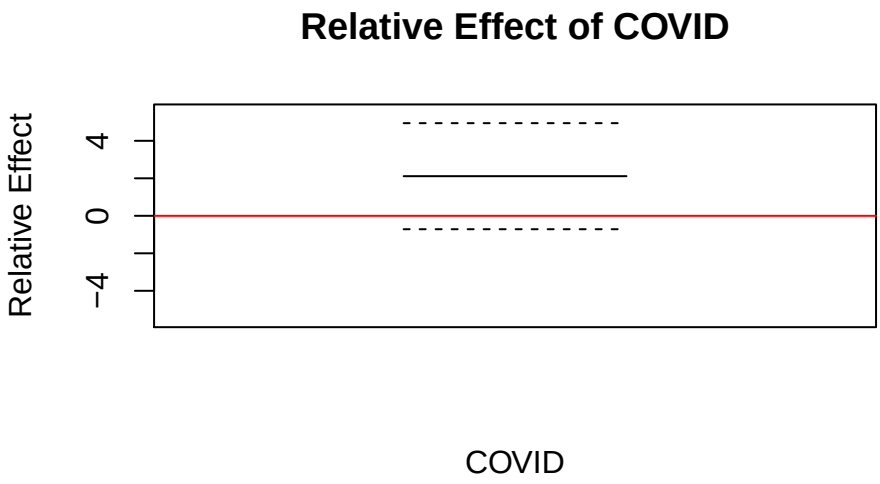


The above plot shows the partial effect of the date alone, without considering other covariates. The following plot shows the actual estimated probabilities over time under the following specific scenario:

- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

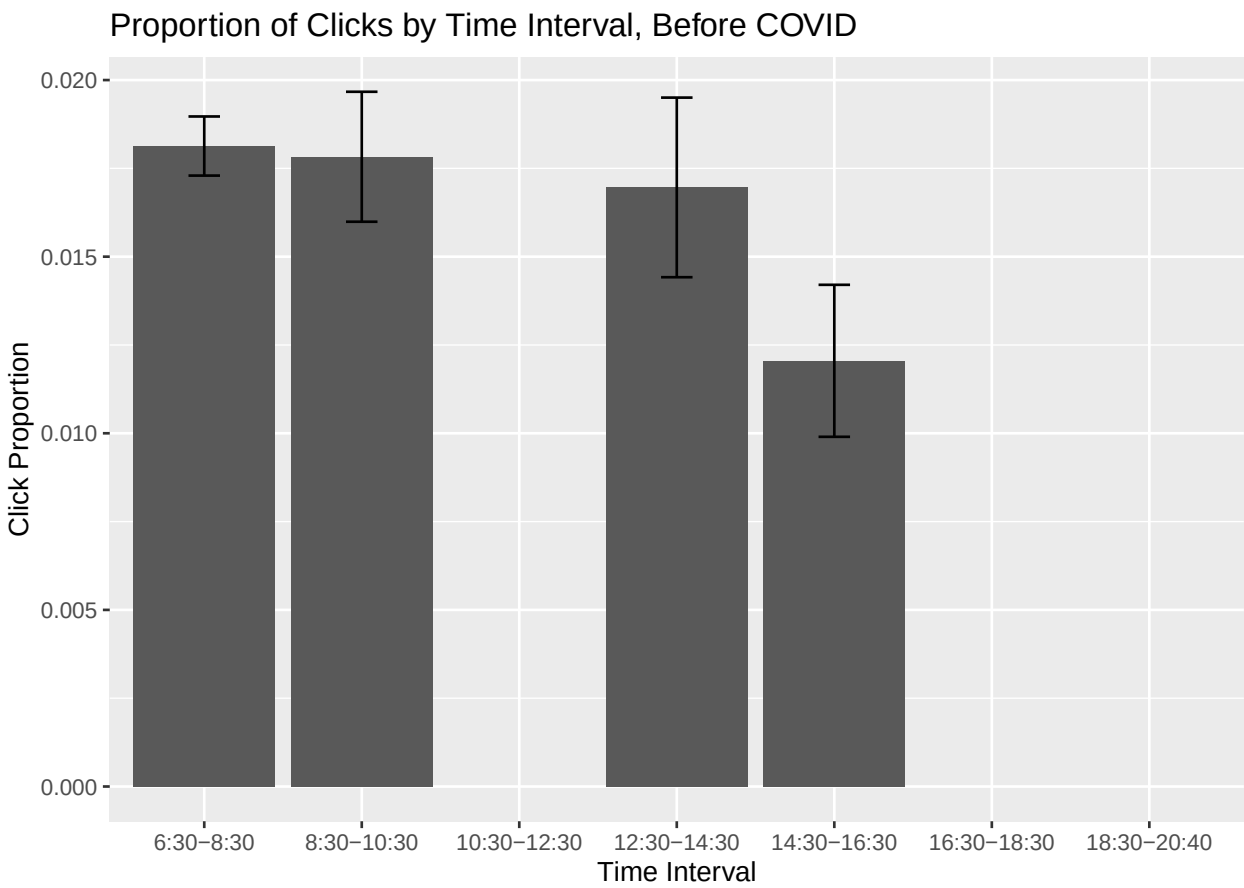


Below shows the relative effect of COVID (solid line), i.e. whether the newsletter was sent after the pandemic started. It appears that the click probability rises after the pandemic starts, but the effect is not significant (see the dashed standard error bars). However, COVID significantly affects how the click probability varies by date or hour of day the newsletter was sent.



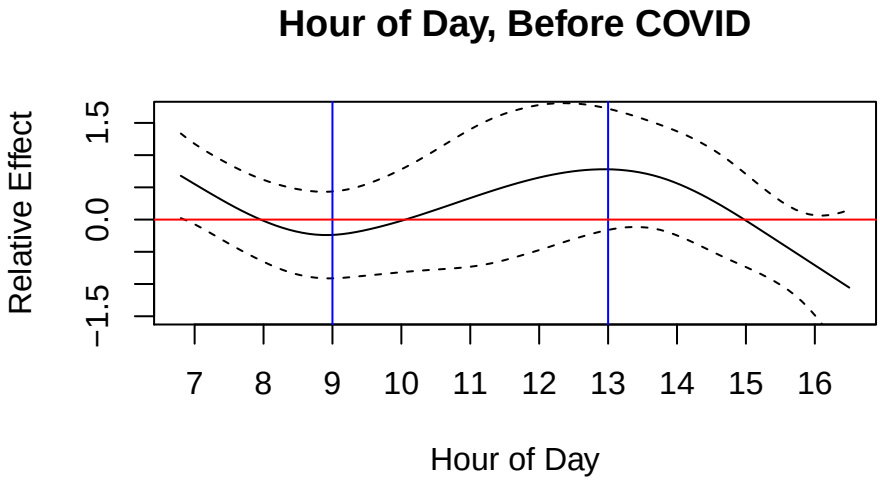
6.5 Time of Day Trend, Before COVID

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter sent before the pandemic, given that the newsletter was sent to them within a specific time interval. There were no newsletters sent in three of the time intervals, so the bars are absent.



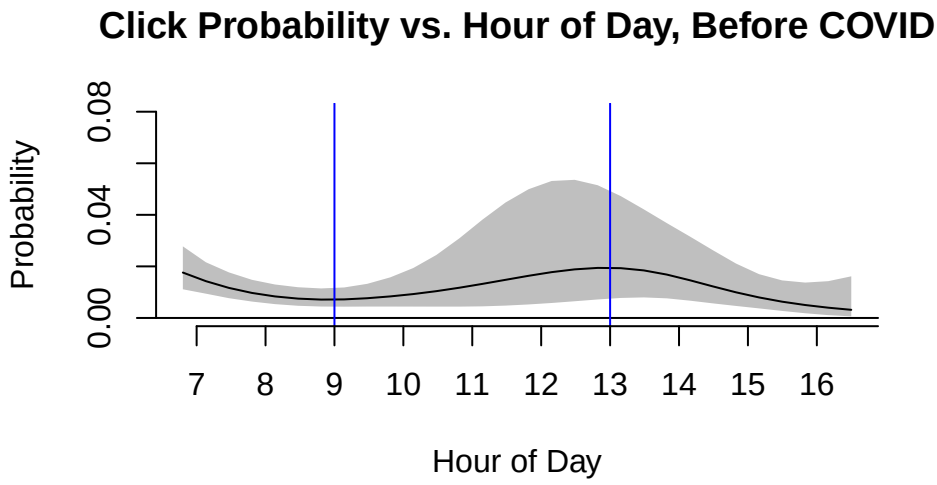
The following plot shows the relative effect of the time of day the newsletter is sent out on the click probability (a negative relative effect corresponds to a decrease in probability, and a positive relative effect corresponds to an increase in probability) before the pandemic.

It appears that there is a dip in the click probability at about 9 am, whereas there is a rise at about 1 pm.



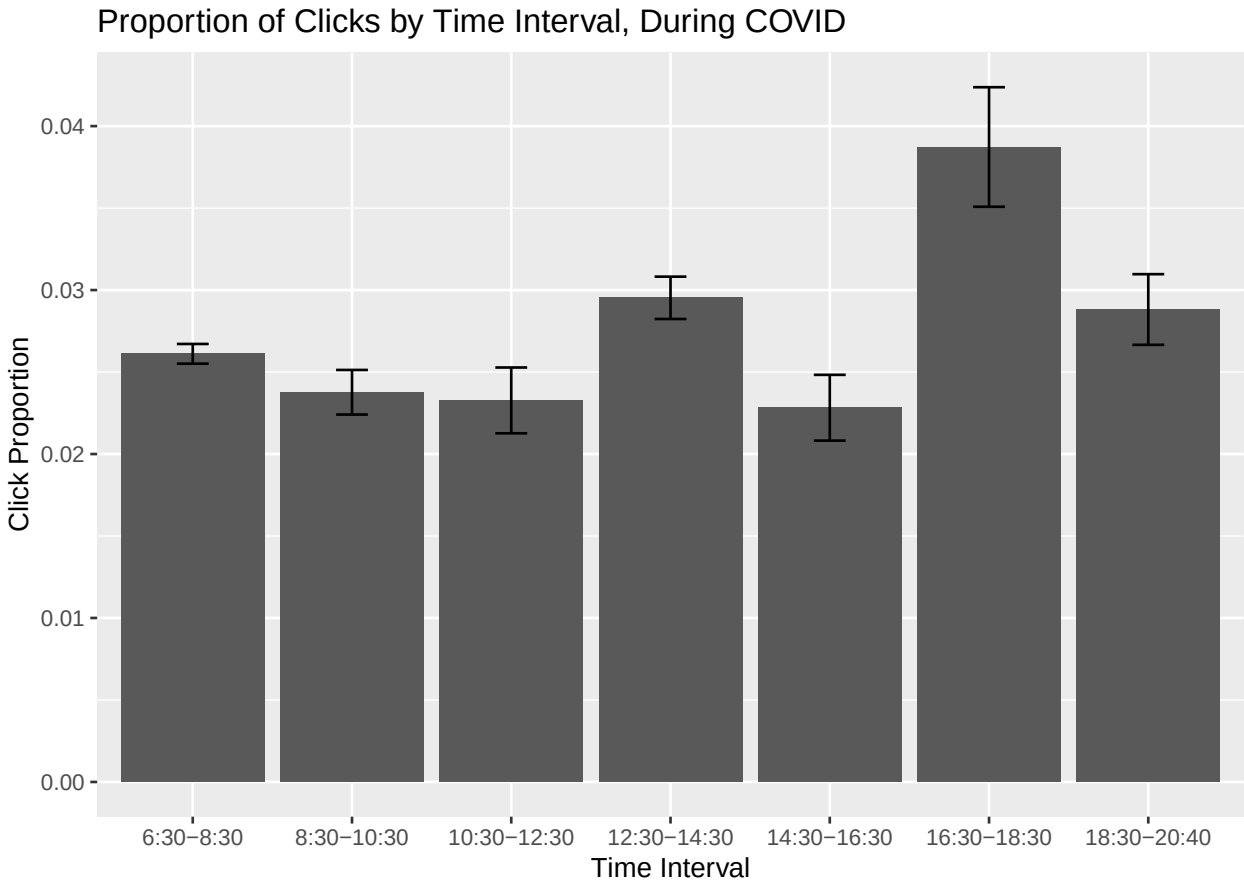
The above plot shows the partial effect of the time of day alone, without considering other covariates. The following plot shows the actual estimated probabilities by time of day under the following specific scenario:

- The newsletter was sent out on December 1, 2019.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

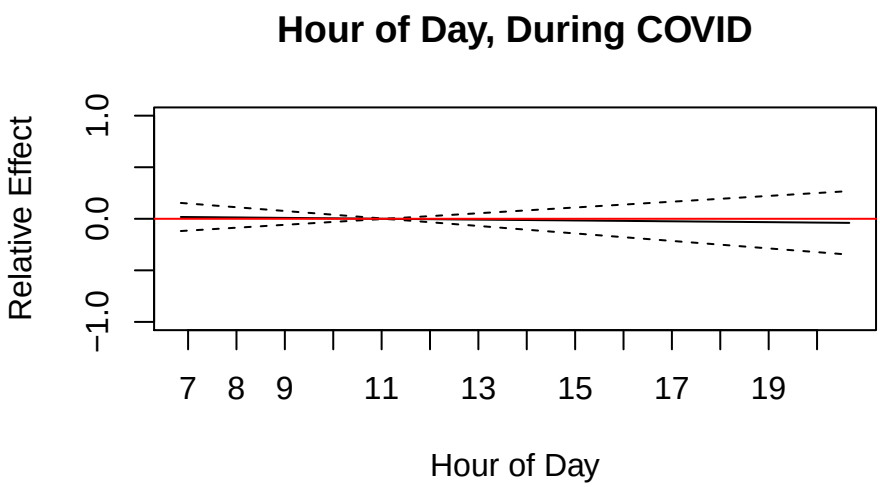


6.6 Time of Day Trend, During COVID

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter sent during the pandemic, given that the newsletter was sent to them within a specific time interval.



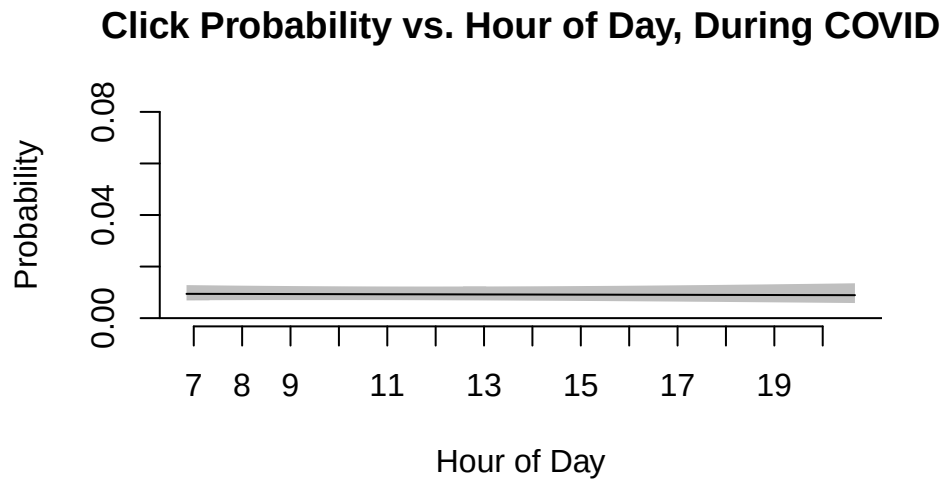
The following plot shows the relative effect of the time of day the newsletter is sent out on the click probability during the pandemic. During the pandemic, the time of day the newsletter is sent has no discernible effect on the click probability.



The above plot shows the partial effect of the time of day alone, without considering other covariates. The following plot shows the actual estimated probabilities by time of day under the following specific scenario:

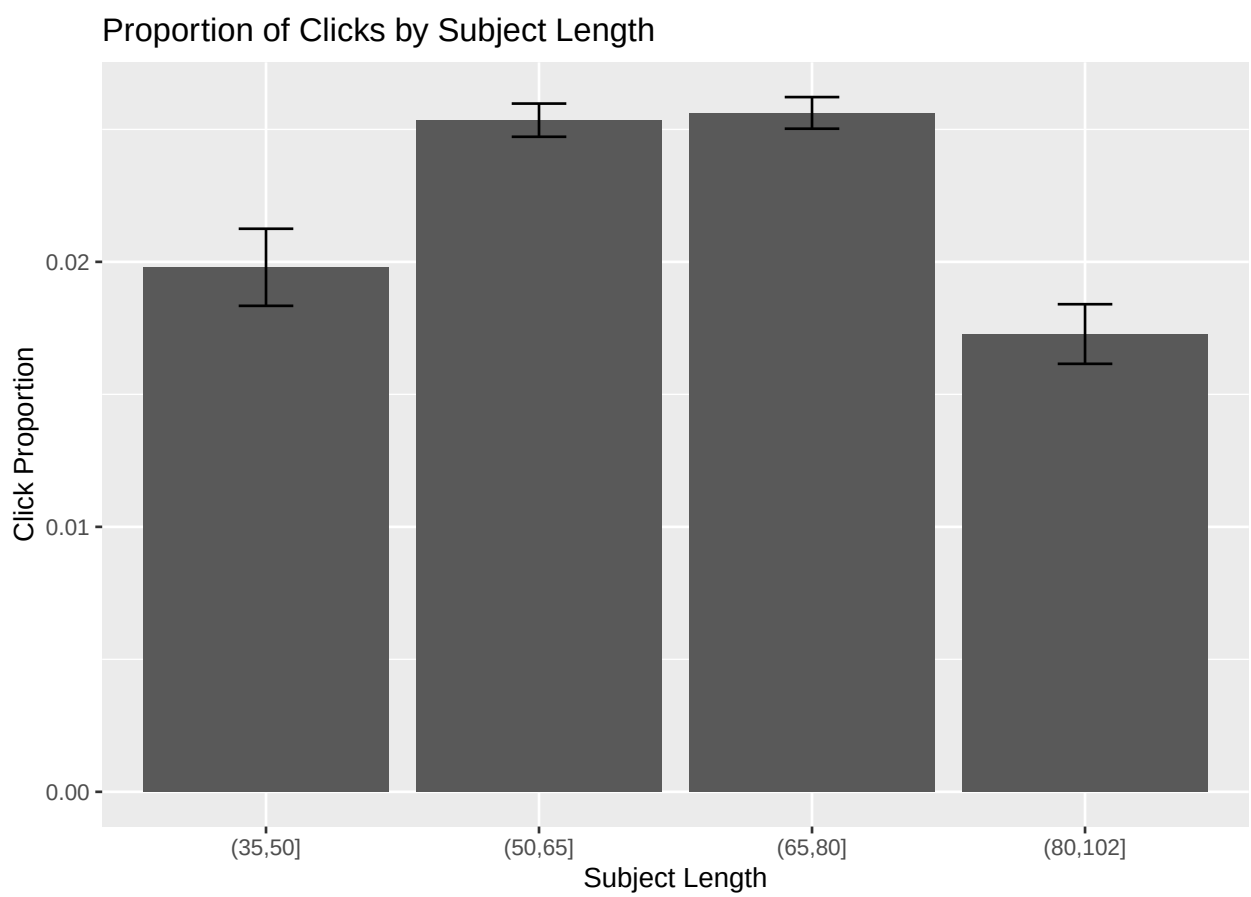
- The newsletter was sent out on December 1, 2020.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.

- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.



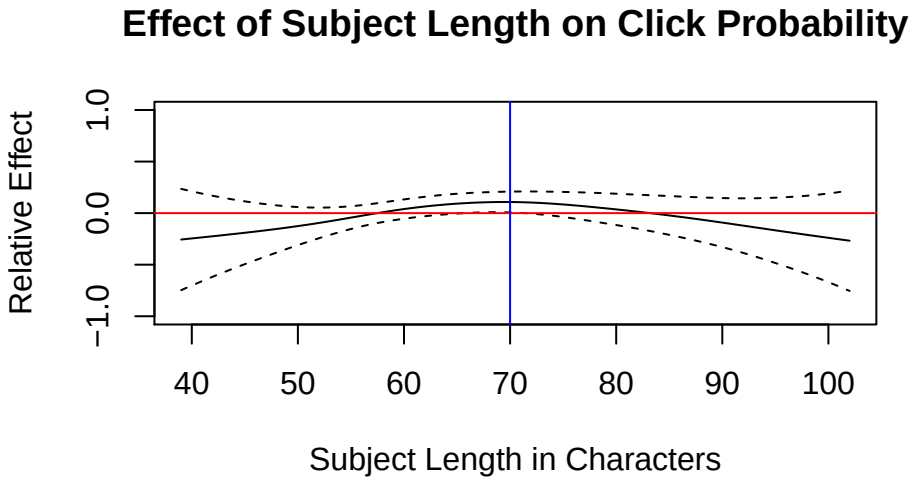
6.7 Subject Length Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given that the subject length was within a specific interval.



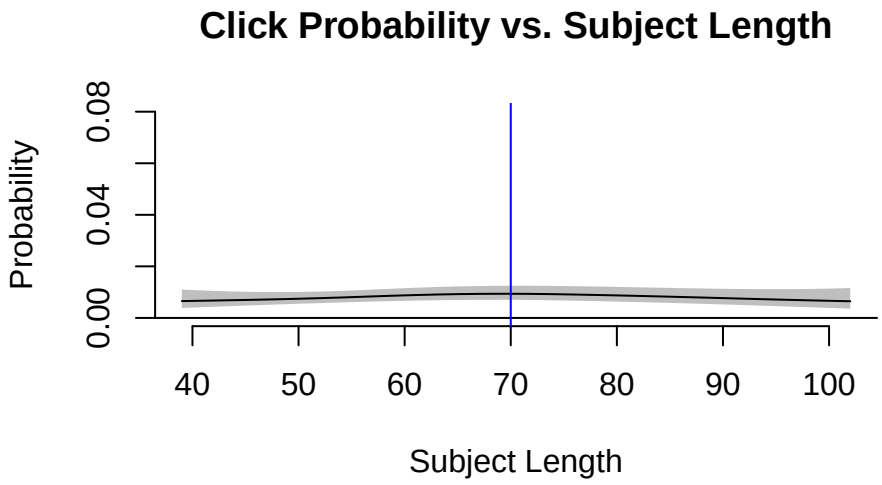
The following plot shows the relative effect of the subject length on the click probability. The maximum click

probability corresponds to a subject length of approximately 70 characters.



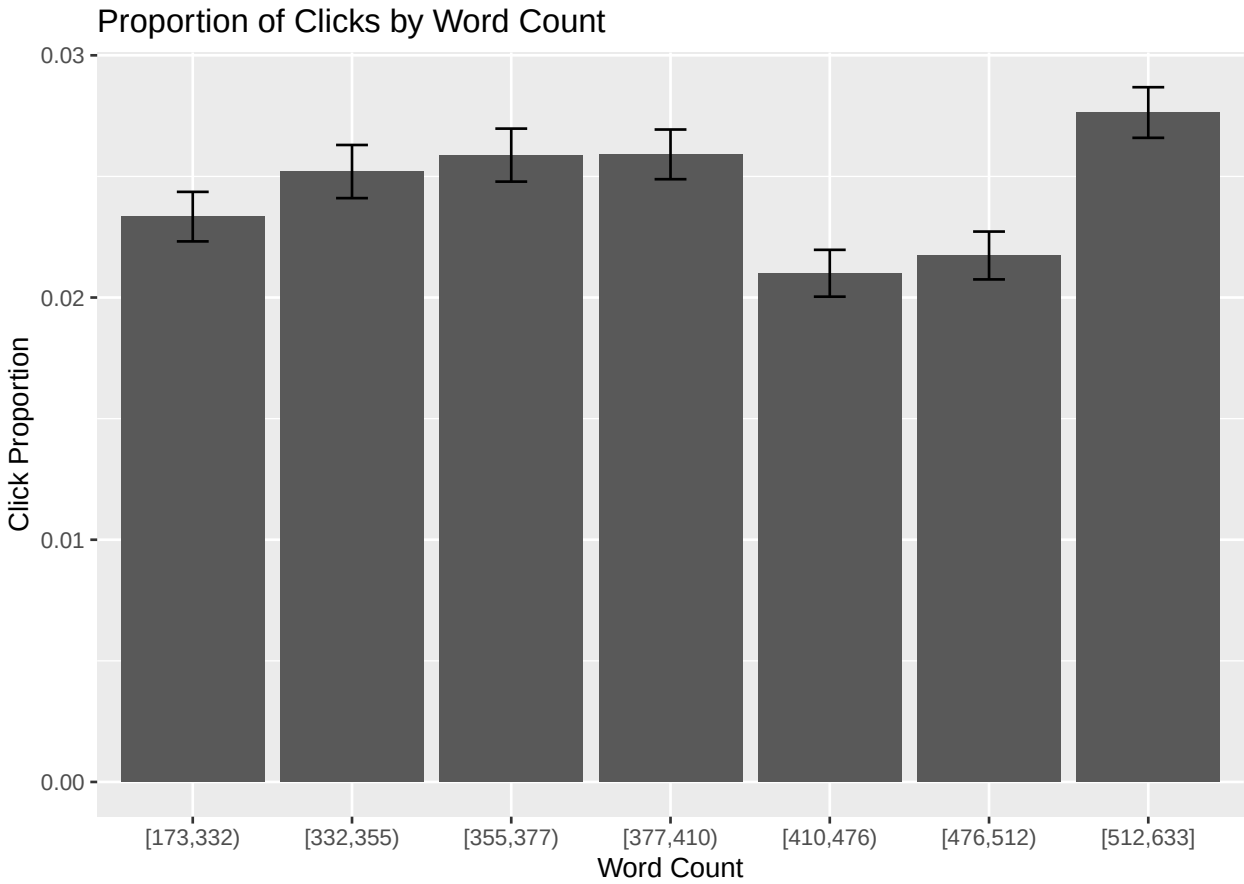
The above plot shows the partial effect of the subject length alone, without considering other covariates. The following plot shows the actual estimated probabilities by subject length under the following specific scenario:

- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

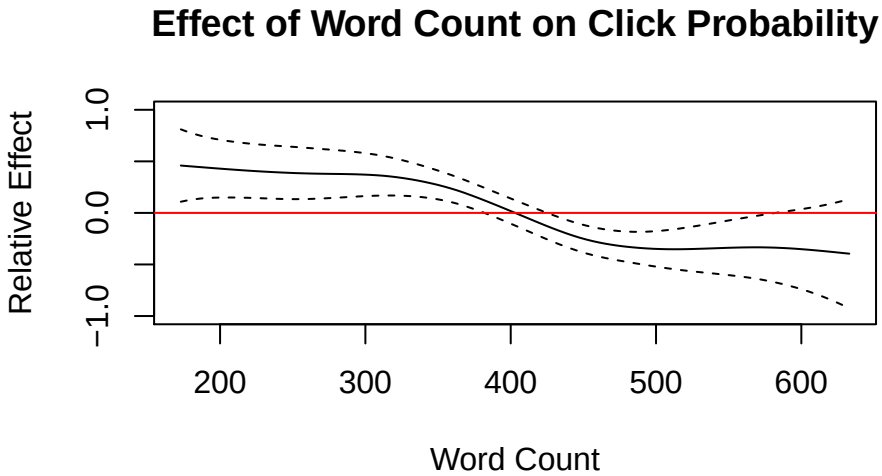


6.8 Word Count Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given that the newsletter word count was within a specific interval.



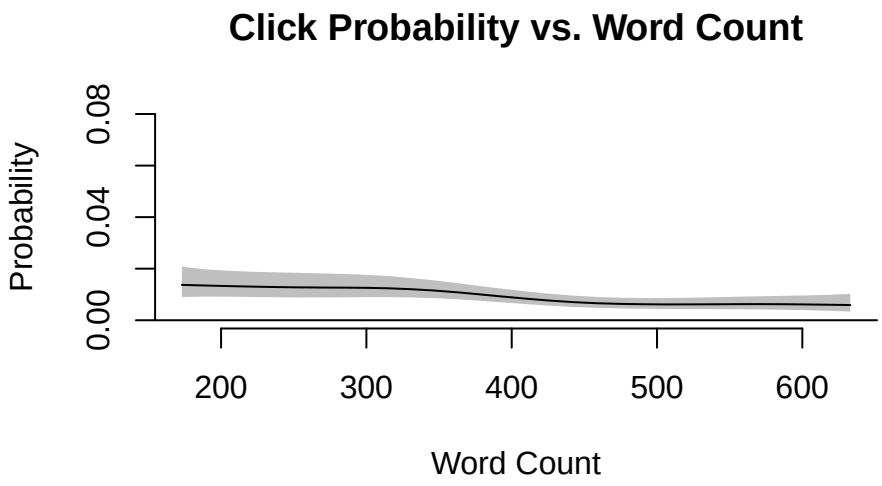
The following plot shows the relative effect of the word count on the click probability. There appears to be a downward trend in click probability as the word count increases.



The above plot shows the partial effect of the word count alone, without considering other covariates. The following plot shows the actual estimated probabilities by word count under the following specific scenario:

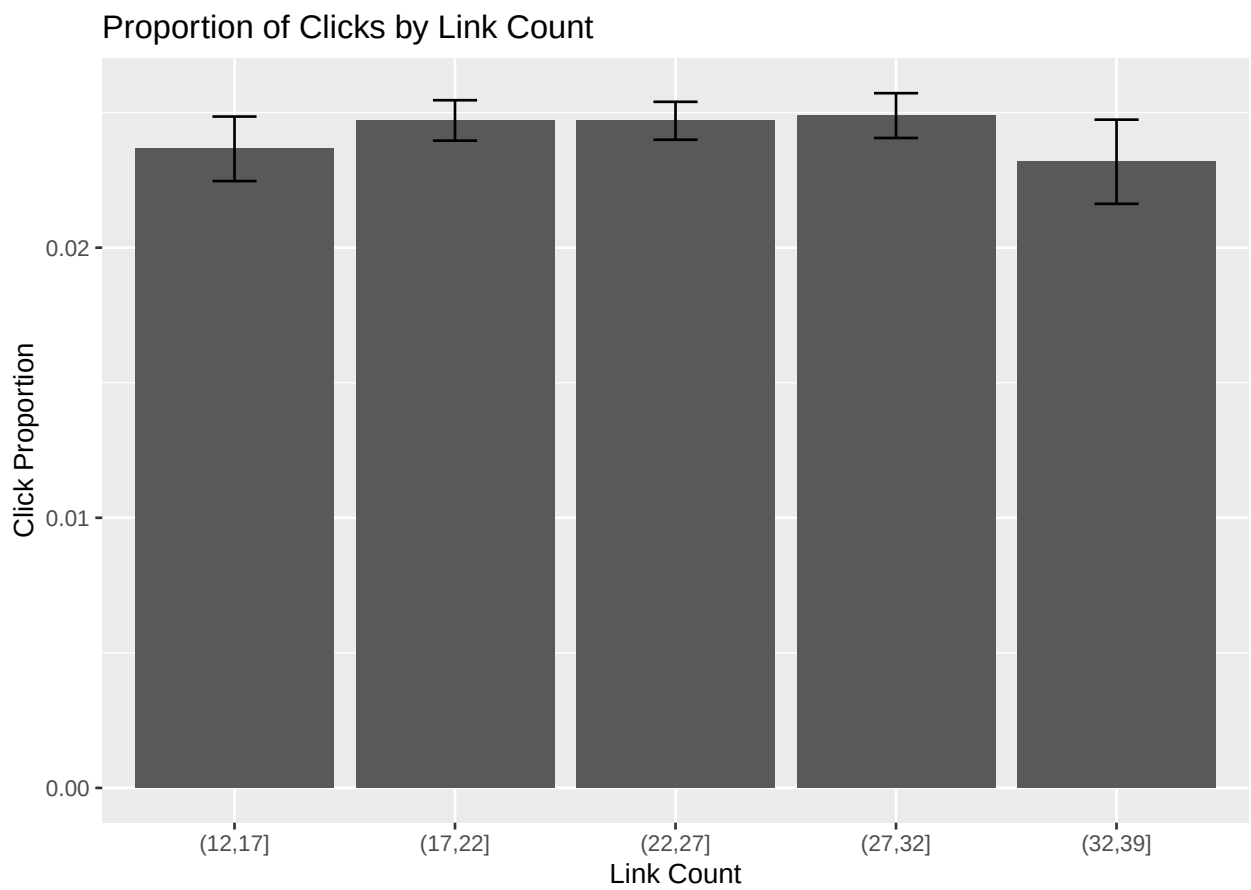
- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.

- The newsletter has the median number of unclickable pictures at 0.

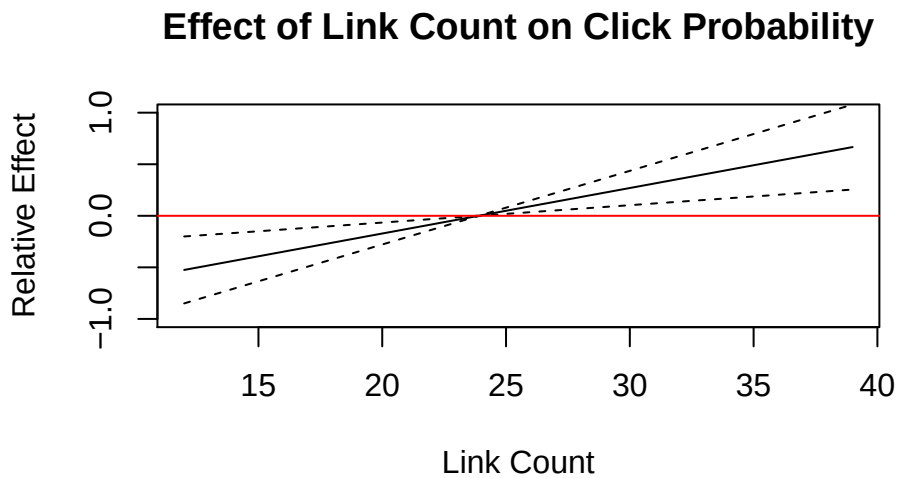


6.9 Number of Links Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given that the newsletter link count was within a specific interval.

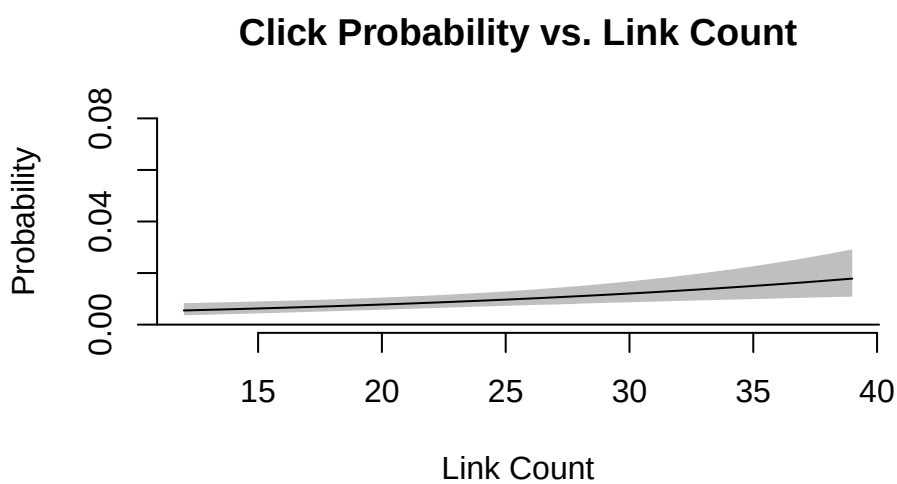


The following plot shows the relative effect of the link count on the click probability. More links is associated with a higher click probability.



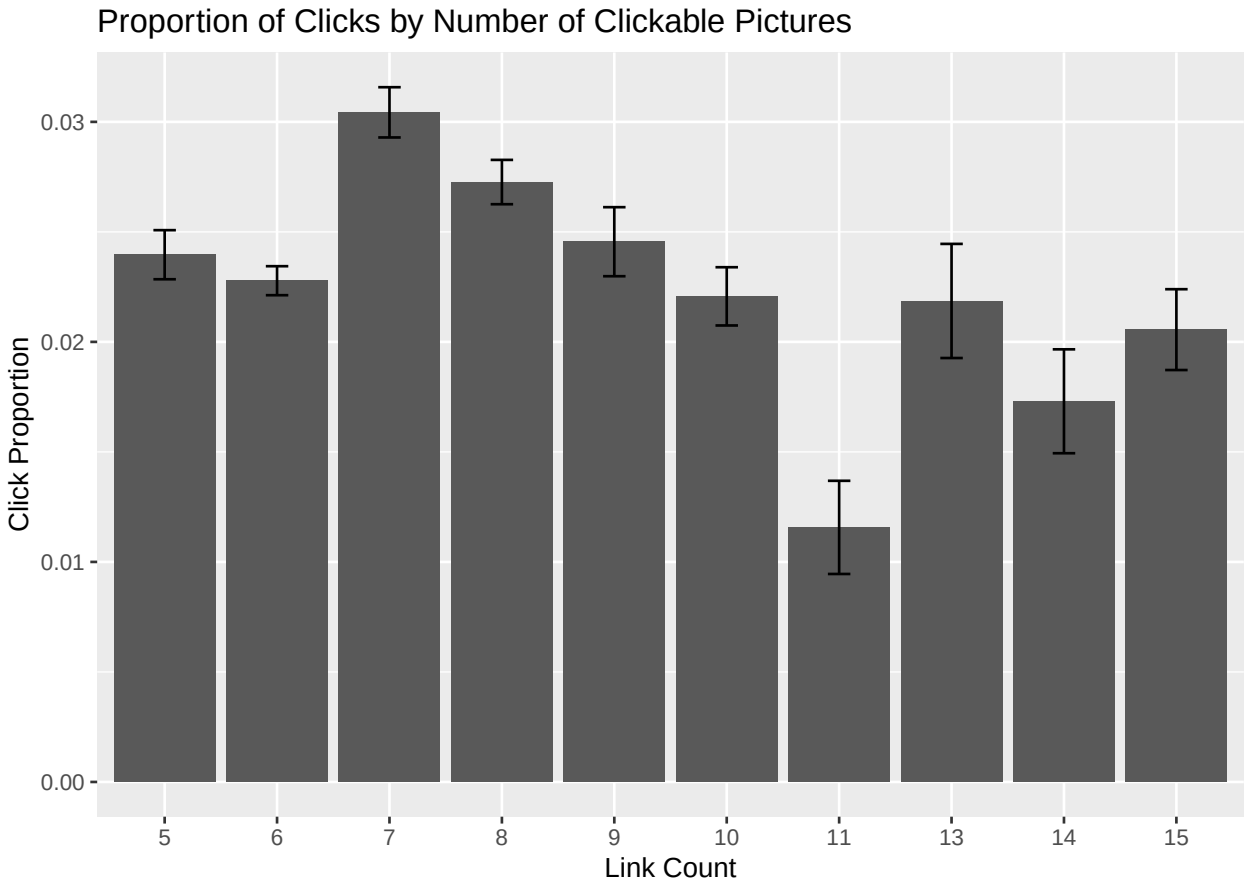
The above plot shows the partial effect of the link count alone, without considering other covariates. The following plot shows the actual estimated probabilities by link count under the following specific scenario:

- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of clickable pictures at 7.
- The newsletter has the median number of unclickable pictures at 0.

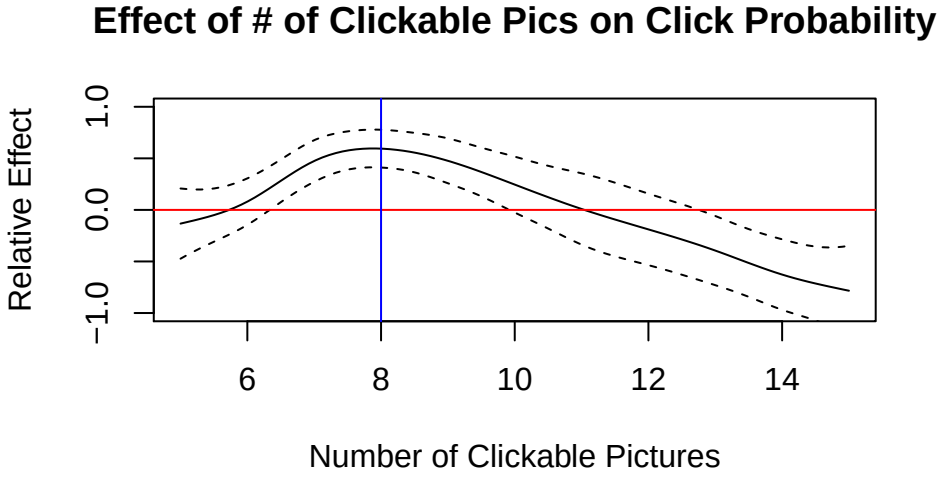


6.10 Number of Clickable Pictures Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given the number of clickable pictures in the newsletter.



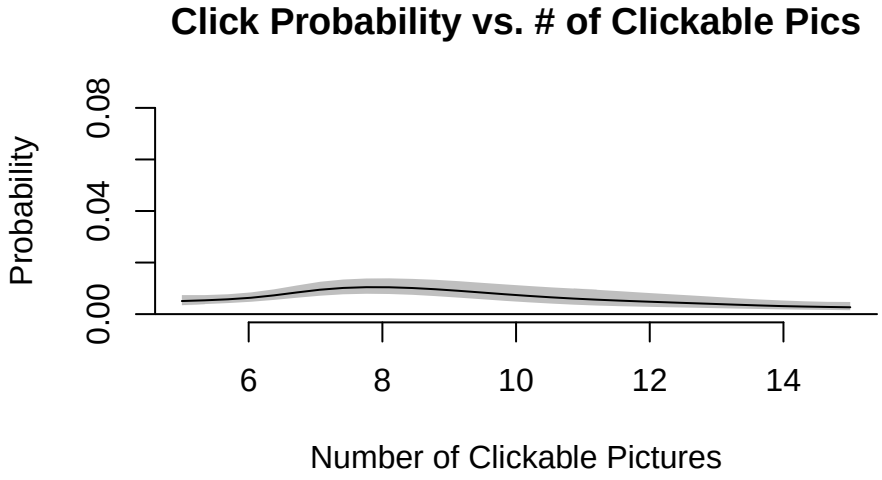
The following plot shows the relative effect of the number of clickable pictures on the click probability. The optimal number of clickable pictures appears to be 8.



The above plot shows the partial effect of the number of clickable pictures alone, without considering other covariates. The following plot shows the actual estimated probabilities by number of clickable pictures under the following specific scenario:

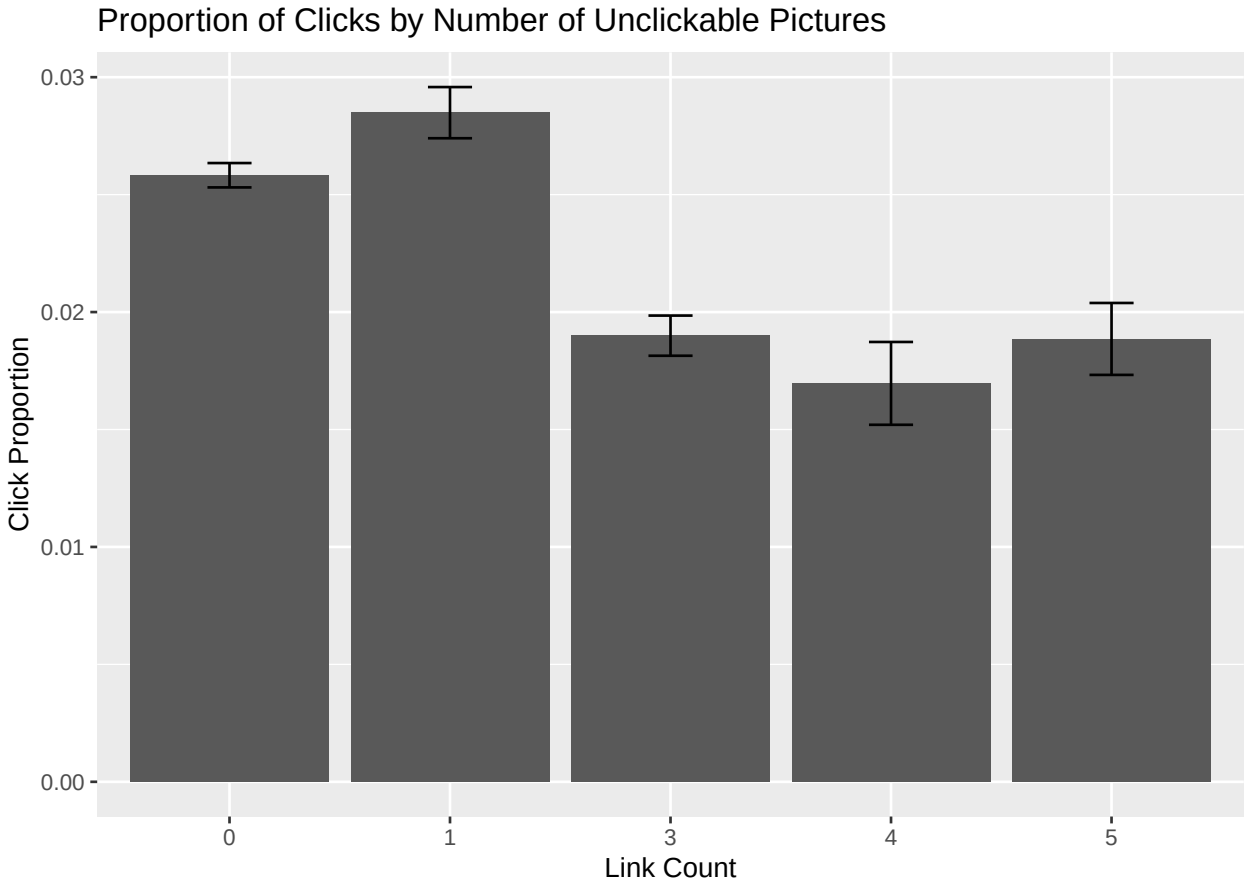
- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.

- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of unclickable pictures at 0.



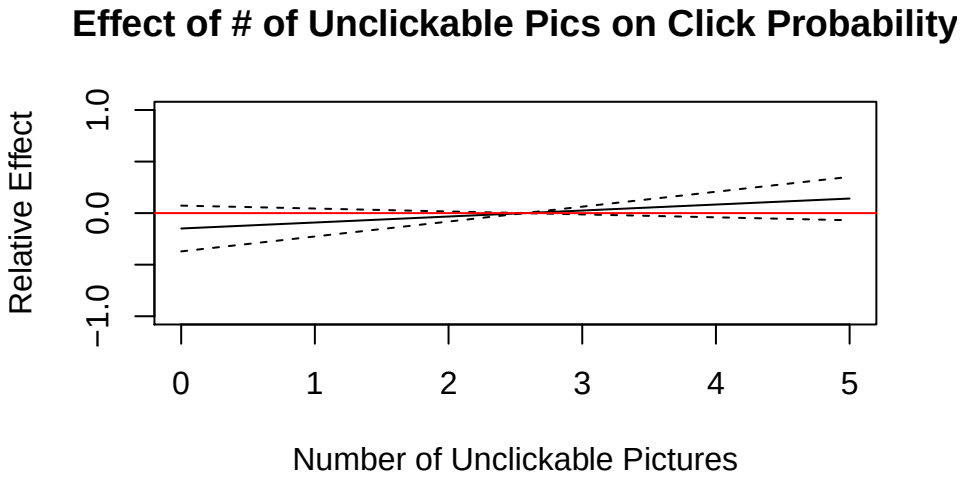
6.11 Number of Unclickable Pictures Trend

The below barplot shows the proportion of subscribers that clicked on a link in the newsletter, given the number of unclickable pictures in the newsletter.



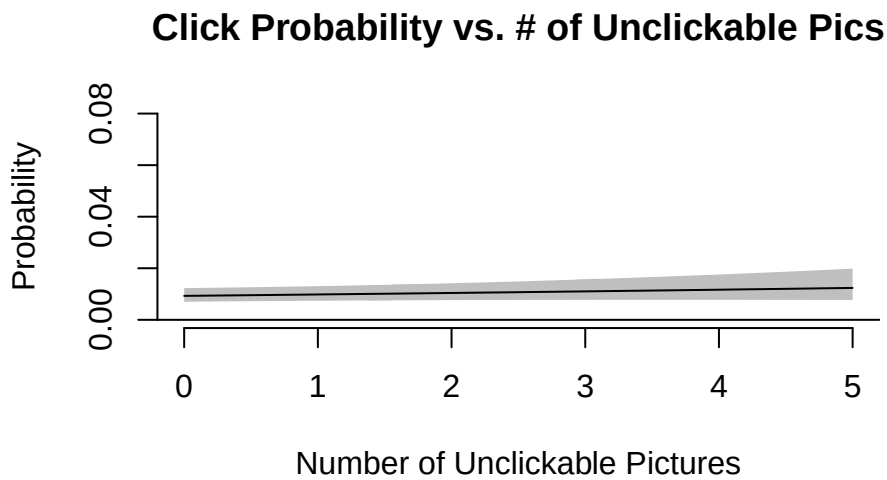
The following plot shows the relative effect of the number of unclickable pictures on the click probability.

The number of unclickable pictures does not have a significant effect on click probability



The above plot shows the partial effect of the number of unclickable pictures alone, without considering other covariates. The following plot shows the actual estimated probabilities by number of unclickable pictures under the following specific scenario:

- The newsletter was sent out on December 1, 2020.
- The newsletter was sent out at 10:30 am.
- The newsletter has the median subject length of 66 characters.
- The newsletter has the median number of words at 392 words.
- The newsletter has the median number of links at 24 links.
- The newsletter has the median number of clickable pictures at 7.



7 Link Characteristics that Promote Clicks: Executive Summary

7.1 Questions of Interest

- Which link characteristics (bolded, font size, font color, location within newsletter) lead to more clicks?
- Which topics mentioned in the links (e.g. animals, art, family, hunger) tend to attract clicks?

7.2 Statistical Analysis

- Bar plots and scatter plots to visually display click response based on link characteristics
- Statistical model to examine how factors interact and support the trends shown in the plots
- Qualitative analysis on the words associated with a link

7.3 Takeaways

- The location of the link in the newsletter affects how often it is clicked. Specifically, the further down the link is in the newsletter, the less likely it is to be clicked on. Therefore, we suggest placing the most important or time-sensitive links at the top of the newsletter.
- Links that are bolded get slightly more clicks than links that do not have bolded text.
- Links that are orange [Mandarin Orange (146, 46, 33), Tangerine (242, 136, 0) or (238, 135, 2), Orange Peel (255, 151, 9)], gray [Charcoal (67, 67, 67)], or blue [Denim (17, 85, 204), Danube (85, 142, 190)] colored get more clicks than other colors.
- Having the link address appear in the newsletter multiple times, whether as text or image links, increases the number of clicks. However, this effect is not significant, according to the model.
- People tend to click on text links rather than image links.
- Opportunities with baby chicks and sharing one's skills are significantly popular.

8 Link Characteristics that Promote Clicks: Report

In this report, we investigate the characteristics of links that make them more likely to be clicked on.

The rest of the report is organized as follows. First, we give a brief description of how the data was obtained and a synopsis of the assumptions we made to analyze the click data. Then, we introduce the features used in the model and analyze how click rates were affected by these features separately. Finally, we fit a statistical model to the data and interpret the results.

8.1 Data Description

The data comes from a few sources: the CSV files generated from iContact and the plain-text and HTML source code of each of the newsletters. From the CSV files, we determine the unique number of times a link was clicked on. We define a unique click to be a unique combination of subscriber ID, newsletter date, and link; in other words, if a subscriber clicked on the same link from the same newsletter, we do not count that click.

Additionally, we only know the click count for each link address, not the actual link, within a newsletter. Thus, if there are multiple links with the same address in a newsletter, we don't know how many clicks each of those separate links received. To alleviate this issue of duplicate addresses, we assume that the first text link with a given address received all the clicks associated with that address. We account for whether the address is duplicated in our model.

We identify the time of day the newsletter was sent and whether it was before or after the COVID-19 pandemic was declared (03/20/20) from the CSV files. The plain-text files are used to get the text associated with a link and an approximation of how far down the newsletter the link is, e.g. a link that is about half-way down the newsletter would be assigned 50%. Finally, we obtain style characteristics and whether the link was an image or had an image associated with it from the HTML source code.

Below is a summary of the features we created for the text links:

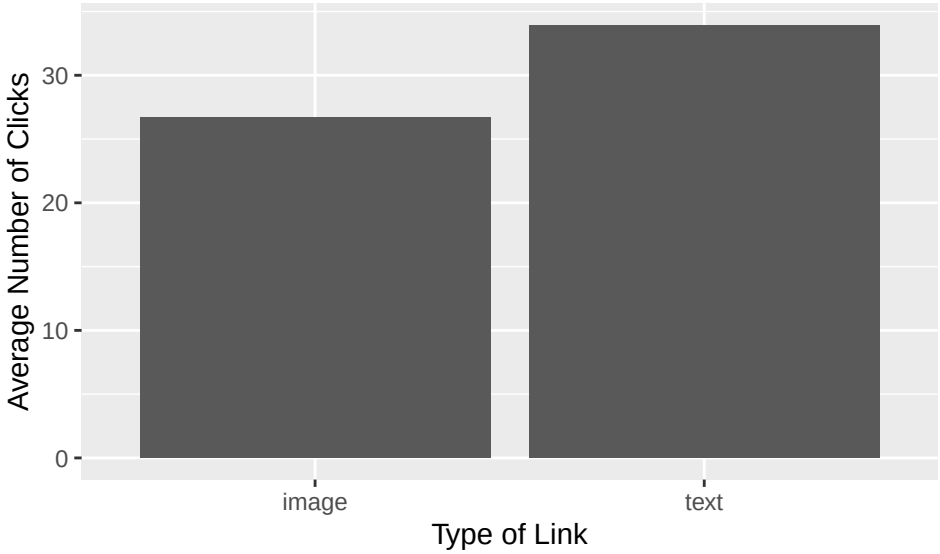
1. Bolded: whether the link text is bolded.
2. Font Size: ranges from 10-48 point.
3. Font Color: 26 possible colors. See Appendix for RGB values.
4. Image Associated: indicator for whether there is an image within the newsletter with the same link address.
5. Hour: hour of when the newsletter was sent
6. COVID: indicator for whether the COVID-19 pandemic was underway
7. Location within document: cumulative percentage of the document prior to a link
8. Duplicate: indicator for whether the link address appears more than once in a newsletter

8.2 Data Exploration

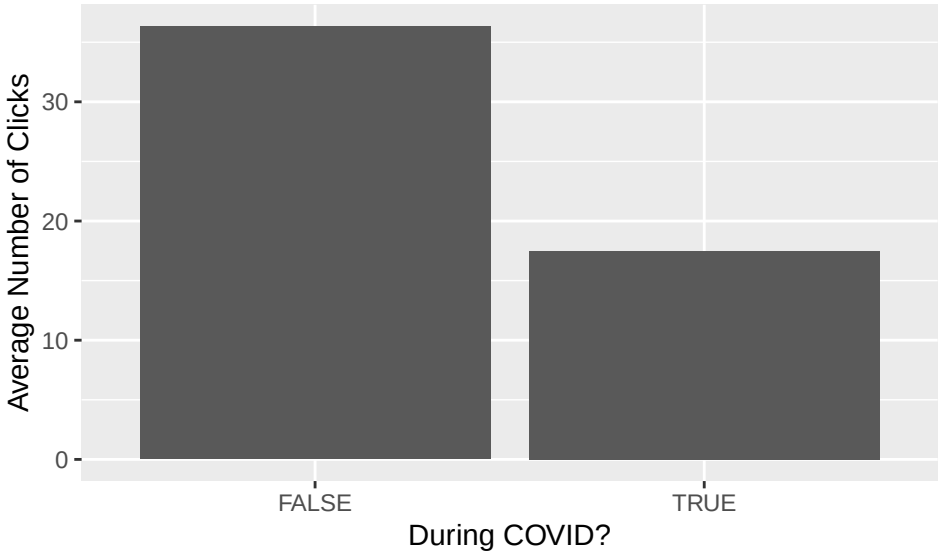
Before fitting any models to the data, we explore how the number of clicks a link address receives depends on the variables mentioned above. It is important to note that in doing this, we do not control for how many times a link address was used within the same newsletter. For each of the categorical variables, we graph the category and the average number of times a link of that category was clicked below.

8.2.1 All Links

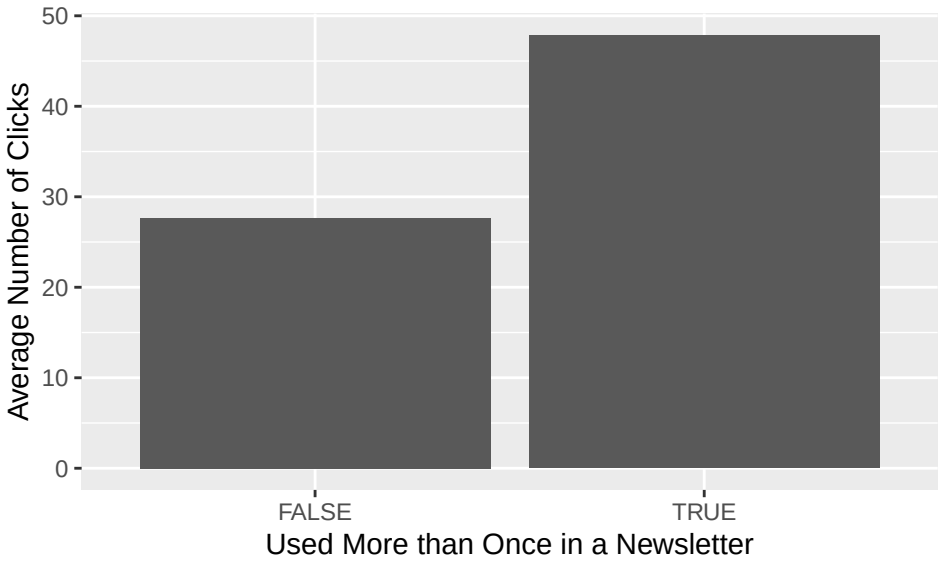
The following barplots pertain to all links, whether they are text or images.



Text links are clicked on more often than pictures. However, because our model only examines text links, we cannot verify this trend with the model.



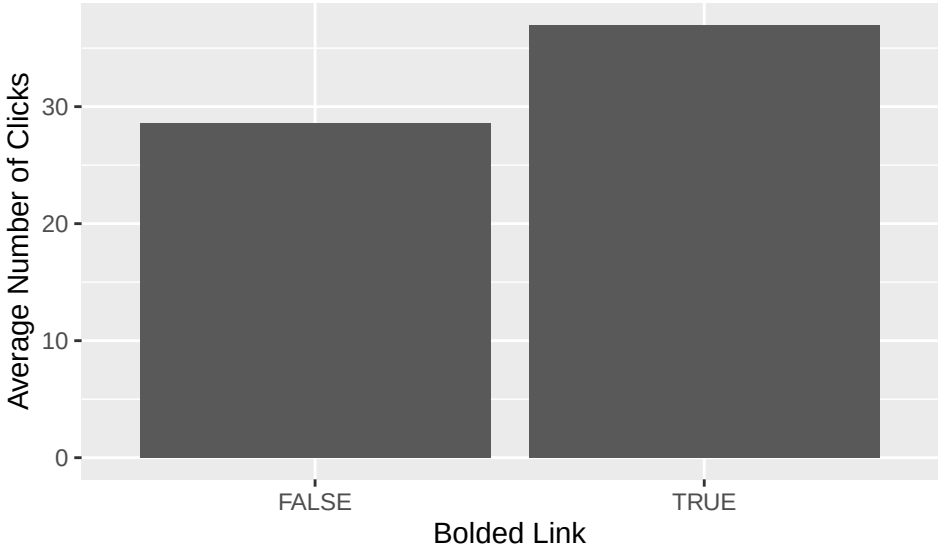
Based on the bar plot above, it seems COVID impacted how much subscribers choose to interact with the newsletters. This is not all that surprising given the challenges everyone was facing during the pandemic.



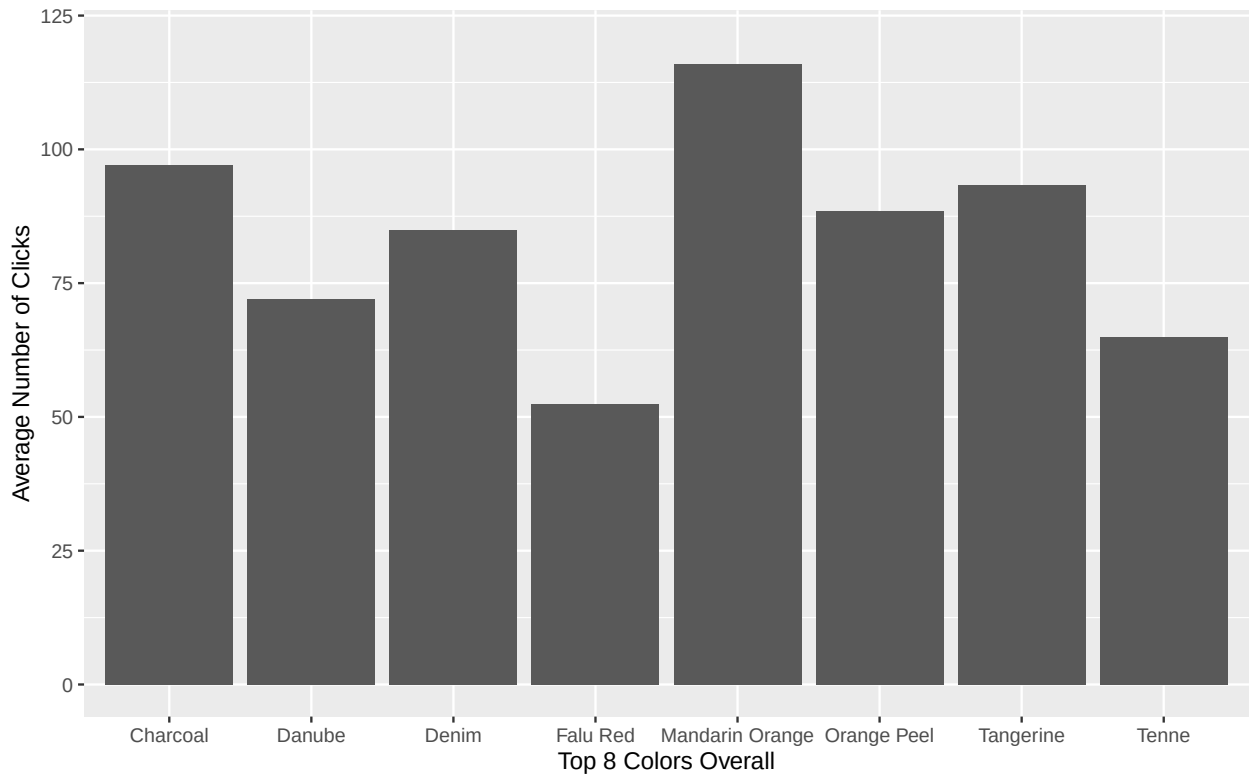
The above bar plot indicates that addresses that appear multiple times in a newsletter tend to be clicked on more often, as expected. Even so, this effect is not significant, according to the below model.

8.2.2 Text Links

The following barplots pertain to only text links.

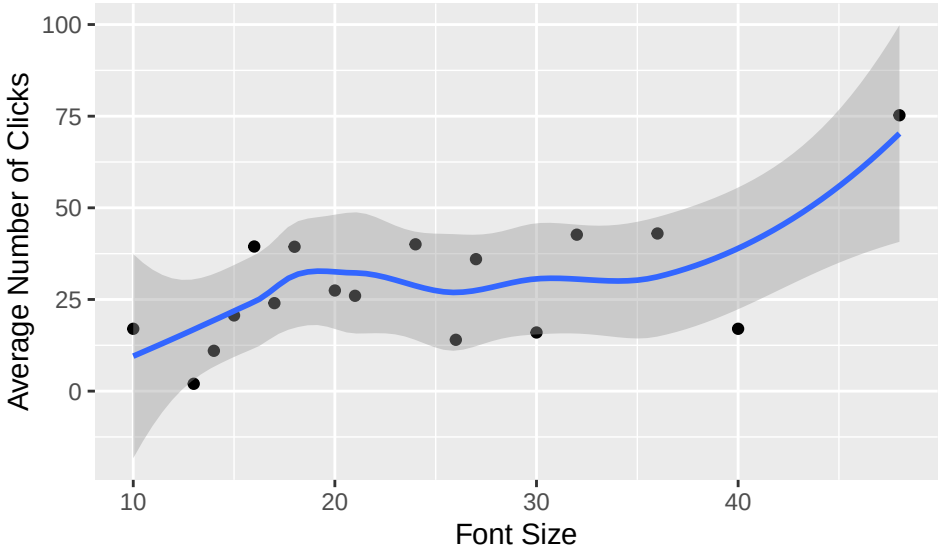


Based on the bar plot above, it appears that bolding the text associated with the link also increases the chance that someone clicks on it.

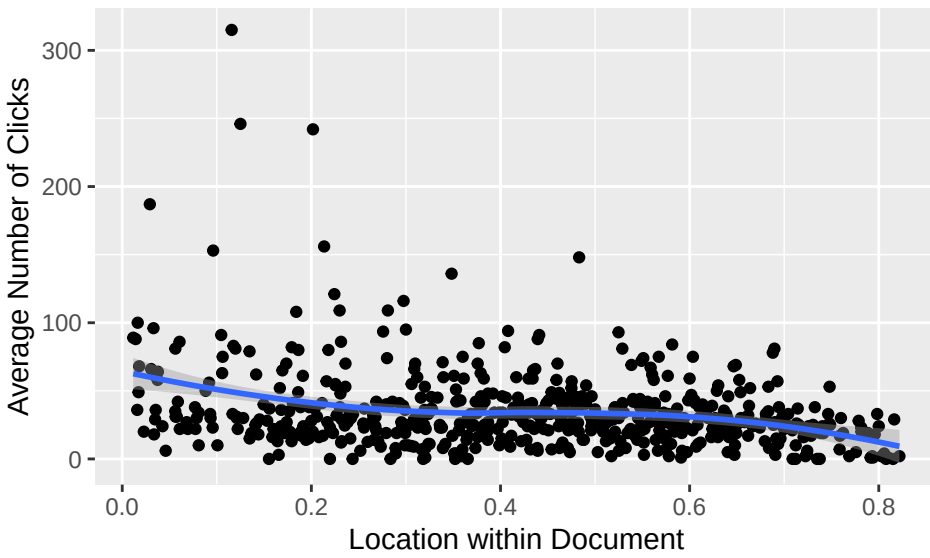


This bar plot focuses on the top eight most clicked-on text color choices, on average. See Appendix for RGB values.

Any color of orange seems to grab people's attention! Mandarin Orange only appeared in the newsletter promoting the Remote Volunteer Project: DIY Family Essentials Kits opportunity so it is tempting to think the large number of clicks this color received may have more to do with the highly-relatable project. However, this project was advertised in four different newsletters using links colored as cinnabar and falu red (both are different tints of red) and these links were not clicked on as often. While there are more factors at play than just the link color, the fact that the links advertising the same opportunity in red got fewer clicks suggests that a text color of orange is more impactful.



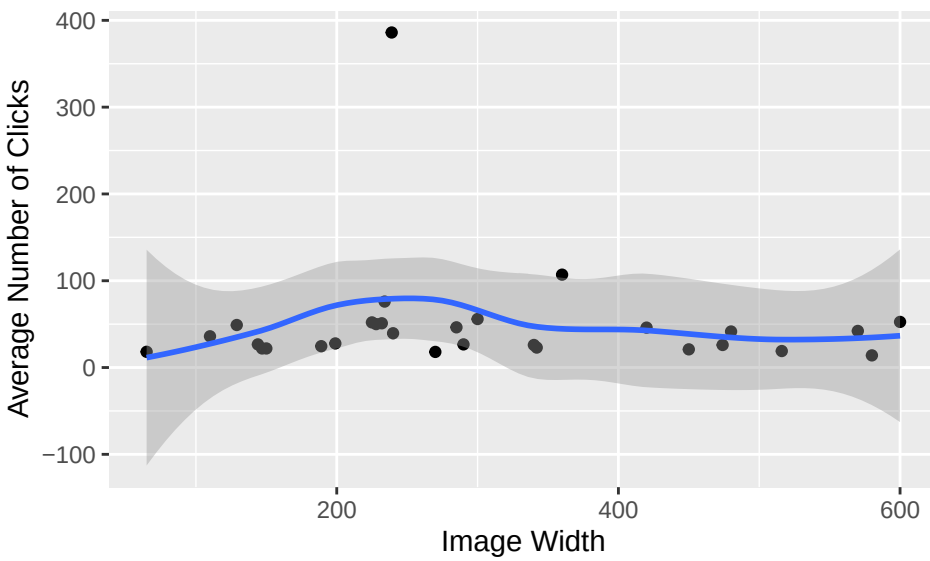
Larger links seem to be clicked on more. However, according to the below model, this effect is not significant.



The further down the newsletter link is, the less likely it is to be clicked on.

8.2.3 Image Links

The following barplots pertain to only image links.



The dimensions of the image is determined by the image width–the image height is adjusted automatically. According to the scatter plot, the image width doesn't appear to affect the click rate.

8.3 Model Fitting

We fit a zero-inflated beta regression model to uncover trends regarding the text links. This model does not include image links.

The outcome variable of interest, the number of clicks, was standardized by dividing the click count by the number of subscribers the newsletter containing the link was sent to. The beta regression allows us to model proportion data (data that's bounded between zero and one, non-inclusive). The "zero-inflated" in the name

refers to extending the beta regression to include observations with a value of zero (many links received no clicks). The zero-inflated beta regression fits three parameters: mu, sigma, and nu. The mu variable corresponds to the mean of the click proportion (relative to the number of subscribers) and is modeled in a similar manner to simple linear regression.

The variables in our model are the following: doc_prop, bolded, color_name, font_size, hour, covid_ind, image_assoc, and dup. "doc_prop" is the proportion down the document a link is; in other words, a link that is about halfway down a newsletter will be about 50%. "bolded" indicates whether a link was bolded. "color_name" is the color of the link as named by https://www.color-blindness.com/color-name-hue/.

Below we give a histogram of click proportions and the fitted model parameters for mu. From Table 1 below, we see that the link's location in the document, whether the link is bolded, and the color of the link make a statistically significant difference on whether the link is clicked or not. Additionally, we see that the top eight colors shown above are also statistically significant, except for Tenne. Interestingly, the indicator for whether the link address is duplicated or not, nor the indicator for whether the link address has an image associated with it, are not significant in the model.

Finally, the mu coefficients given in the table below are, unfortunately, uninterpretable in their raw form. Luckily, a transformation of these coefficients gives the odds ratio of each variable. For the variables that are statistically significant at or below the 0.01 level, we give the odds ratios in Table 2. As an example of how odds ratios are interpreted, the doc_prop odds ratio of 0.46 means that when the location of the link in the document increases by one percentage point, the odds of it being clicked is 0.46 times the odds of it being clicked in the original position. Overall, an odds ratio greater than one indicates a positive association and an odds ratio less than one indicates a negative association; note that these agree with the signs of the coefficients in Table 1.

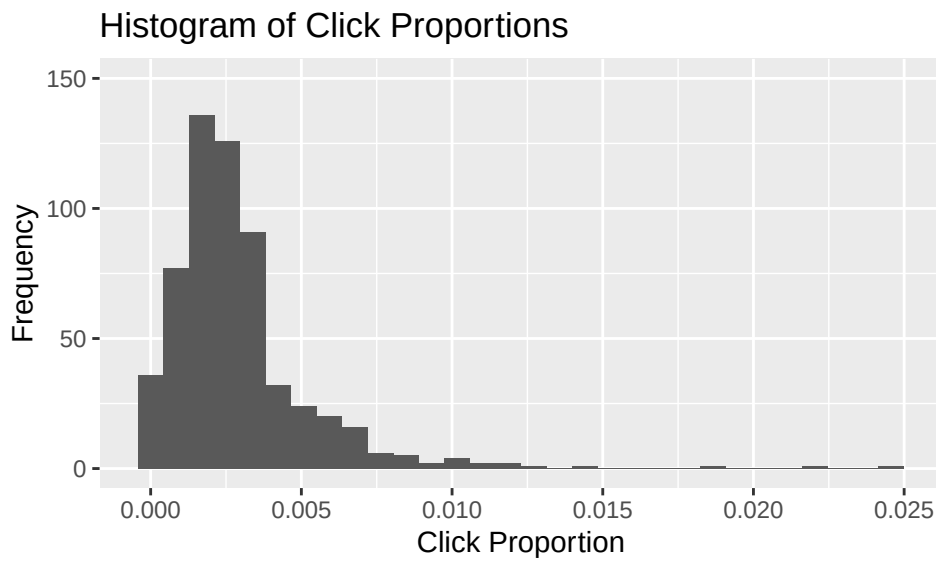


Table 1: Mu Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-5.96	0.26	-23.18	< 2e-16	***
doc_prop	-0.78	0.14	-5.56	0.00	***
boldedTRUE	0.22	0.06	3.50	0.00	***
color_nameBlack	0.61	0.36	1.71	0.09	.
color_nameBlack Pearl	-0.09	0.44	-0.21	0.84	
color_nameCharcoal	1.06	0.31	3.39	0.00	***
color_nameChocolate	0.47	0.24	1.93	0.05	.
color_nameCinnabar	0.07	0.33	0.23	0.82	
color_nameCitron	0.13	0.30	0.45	0.66	
color_nameDanube	1.62	0.34	4.72	0.00	***
color_nameDenim	1.49	0.31	4.84	0.00	***
color_nameDim Gray	-0.23	0.73	-0.32	0.75	
color_nameEastern Blue	0.28	0.25	1.09	0.28	
color_nameEclipse	0.37	0.24	1.58	0.12	
color_nameFalu Red	0.64	0.24	2.70	0.01	**
color_nameGamboge	0.15	0.53	0.29	0.77	
color_nameGrey	0.18	0.24	0.77	0.44	
color_nameMandarin Orange	1.35	0.38	3.52	0.00	***
color_nameMariner	0.26	0.59	0.43	0.67	
color_nameNero	0.18	0.68	0.27	0.79	
color_nameOrange Peel	1.04	0.34	3.08	0.00	**
color_nameSlate Blue	0.45	0.61	0.75	0.46	
color_nameTangerine	0.94	0.28	3.42	0.00	***
color_nameTeal	0.44	0.24	1.82	0.07	.
color_nameTenne	0.85	0.44	1.92	0.06	.
color_nameTyrian Purple	0.28	0.55	0.51	0.61	
color_nameWhite	0.42	0.24	1.72	0.09	.
font_size	-0.01	0.01	-1.03	0.31	
hour	0.01	0.01	1.29	0.20	
covid_indTRUE	-0.70	0.08	-9.17	< 2e-16	***
image_assocTRUE	0.03	0.10	0.32	0.75	
dupTRUE	0.12	0.09	1.30	0.19	

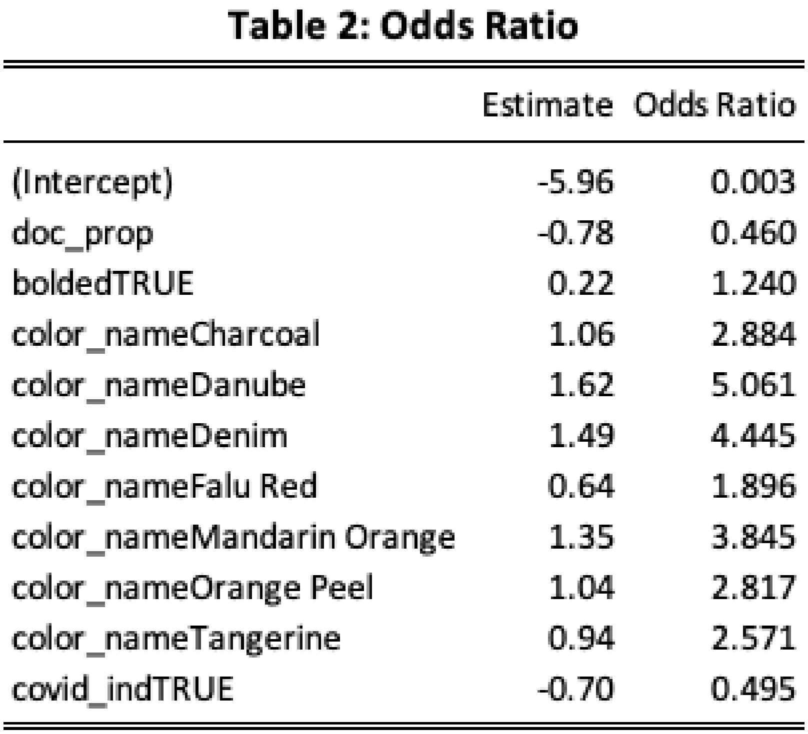


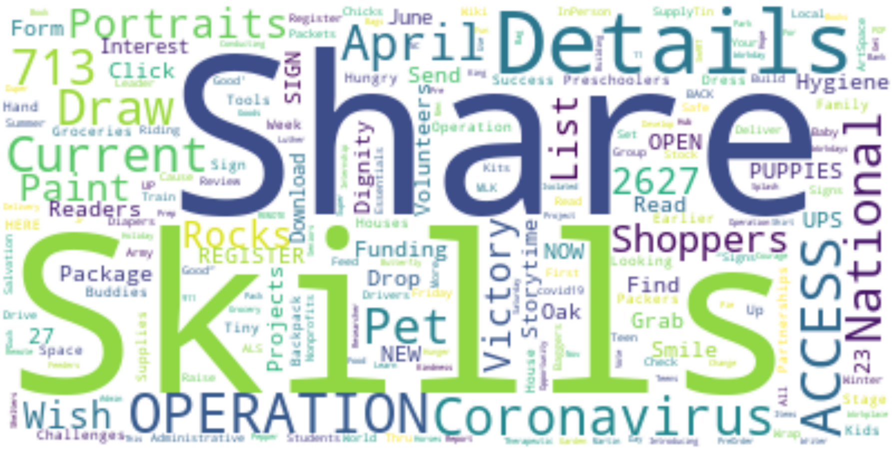
Table 2: Odds Ratio

	Estimate	Odds Ratio
(Intercept)	-5.96	0.003
doc_prop	-0.78	0.460
boldedTRUE	0.22	1.240
color_nameCharcoal	1.06	2.884
color_nameDanube	1.62	5.061
color_nameDenim	1.49	4.445
color_nameFalu Red	0.64	1.896
color_nameMandarin Orange	1.35	3.845
color_nameOrange Peel	1.04	2.817
color_nameTangerine	0.94	2.571
covid_indTRUE	-0.70	0.495

8.4 Qualitative Text Analysis

Finally, we explored what words encouraged subscribers to click on a link by creating word clouds. The word clouds are composed of the capitalized words that were contained in each link. By focusing on capitalized words, we attempted to ignore most of the unimportant filler words. The size of the words corresponds to the proportion of unique clicks relative to the number of total clicks a newsletter obtained. Note that this is slightly different than the proportion of clicks defined above for the zero-inflated beta model.

The first word cloud below is for any links that were not social media for Activate Good.



The second word cloud below is for links that were for opportunities as defined by the link containing "opportunity" in the address; these links correspond to volunteer opportunities for subscribers.

As we can see, opportunities with baby chicks and sharing one's skills are significantly popular compared to other opportunities!

9 Appendix (Color RGB Values and Opportunity Link Click Frequency Table)

Here we provide the RGB values for all of the colors that were present in the news letter links. The names were assigned according to https://www.color-blindness.com/color-name-hue/.

Color Names
Bahia (179,183,27)
Black (0,0,0)
Black (10,10,10)
Black Pearl (29,33,41)
Charcoal (67,67,67)
Chocolate (228,104,16)
Chocolate (233,93,20)
Cinnabar (231,93,38)
Citron (152,154,38)
Danube (85,142,190)
Denim (17,85,204)
Dim Gray (97,97,97)
Eastern Blue (0,136,168)
Eclipse (55,55,55)
Eclipse (57,57,57)
Falu Red (148,45,27)
Gamboge (228,134,9)
Grey (127,127,127)
Mandarin Orange (146,46,33)
Mariner (56,88,152)
Mortar (85,85,85)
Mortar (89,89,89)
Nero (34,34,34)
Orange Peel (255,150,0)
Orange Peel (255,151,9)
Slate Blue (102,94,208)
Tangerine (238,135,2)
Tangerine (242,124,0)
Tangerine (242,136,0)
Tangerine (248,118,0)
Teal (0,109,131)
Tenne (206,86,0)
Tyrian Purple (13,0,0)
White (255,255,255)

This table shows the number of unique people that clicked each "opportunity link" (https://activategood.org/opportunity/##) in the newsletters. Click the title of the link to go to the opportunity page.

Title	Dates Sent	Click Count
IN PERSON - "Operation Mask Up" - Assemble Mask Kits	09/30/20, 12/21/20	383
Help us Fight Hunger with the Food Bank (Raleigh)	06/05/20, 06/10/20, 07/01/20, 07/29/20, 08/05/20, 08/26/20	229
Help Feed Hungry Kids - food distribution to kids & families	11/24/20, 12/02/20	198
Christmas Toy and Clothing Center	11/24/20, 12/09/20, 12/16/20	172
Train to be a Project Leader with Activate Good!	07/29/20, 08/05/20	158
REMOTE - Create and Mail Cards to Men in the Shelter This Holiday Season	12/02/20, 12/16/20	157
Delivery Drivers	12/02/20	136
At Home Project: Hatching/ Caring for Baby Chicks	04/11/20	122
Building Nature Trails at Roberts Chapel Conservation Area	10/21/20, 11/18/20	95
Join Us "in the Dog House"!	10/08/19	91
Home Project: Garden Tool & Wishlist Collection	05/27/20, 06/17/20	88
Teens + Parents: Help Package Hygiene Supplies for Dignity Week	12/29/20	86
WAKE Up and Read Grab, Go, and Read! Book Bags	07/21/20	85
FIGHT HUNGER WITH THE FOOD BANK (RALEIGH)	10/28/20, 11/11/20	82
WAKE Up and Read Grab, Go, and Read! Book Bags	08/05/20	80
**** Weekly Packets for Children Needed ****	04/28/20	78
HOLIDAY HEROES are Needed In the Triangle - YMCA Virtual Santas SIGN UP!!	11/24/20, 12/09/20	78
Pick up and Deliver Family Essentials Kits to Nonprofits Serving Families	10/14/20	75
Marbles Kids Museum PreK Pack assembly	12/16/20	75
Prep for the Super Duper Duck Splash	07/01/20, 07/08/20	74
Courage Cards in Spanish	07/08/20, 07/29/20, 08/12/20	73
Career & Workplace Videos for Students Served by Junior Achievement & WELL	09/02/20, 09/09/20, 09/16/20	73
UWH - Fight Hunger with the Food Bank (Raleigh)	09/02/20, 09/09/20	72
Backpack Buddies Packers	02/01/19, 03/04/19	71
Supply Drive	03/17/20, 05/05/20	70
Puppy Time	11/01/19	69
REMOTE Create BOOST Videos About Academic Concepts	08/26/20, 09/09/20, 09/23/20	67
Thanksgiving Bag-athon!- Filled!	10/28/20, 11/11/20	67
Dress For Success Sale event	06/24/20	66
Courage Cards for TEENS!	03/17/20, 05/05/20	65

Title	Dates Sent	Click Count
Wednesday Build Day	07/21/20	65
Prepare items for our upcoming Family Volunteer Day	11/11/20, 11/18/20	65
Grant Finders and Donation Requests	03/17/20, 05/05/20	64
Volunteer Driver for Meal Delivery to Kids and Families	11/18/20	63
Diaper Wrapping!	10/14/20	62
2020 Catfish Classic	10/21/20	60
Warmth for Wake Volunteer	01/02/19, 11/01/19, 12/03/19	59
Help us Fight Hunger with the Food Bank (Durham)	05/27/20	59
Rock Painting / Garden Volunteers - Kid Friendly	06/24/20	59
Volunteer Workdays	11/04/20	59
Board of Advisors	04/22/20	58
Update Client Information In Our Database	07/01/20, 07/21/20	58
Christmas Client Registration	10/08/19	57
MLK Day of Service with TABLE	12/29/20	55
Monthly Adoption Event	06/04/19, 07/08/19, 08/05/19	53
Aftercare Committee Volunteer	10/07/20	53
Remote Volunteer Project: DIY Family Essentials Kits (Supplies Needed)	11/11/20	49
Administrative Assistant	10/21/20	48
Monthly Adoption Event	01/03/20	46
Sort Goods for Good with the Raleigh Rescue Mission Donation Center	10/28/20	46
Warehouse Pantry Volunteer	10/07/20	45
Trail Building with TLC	01/03/20	43
Pantry Makeover with Alliance of AIDS Services	07/29/20, 08/12/20	43
Special Event Volunteer	05/13/20	42
Families on the Farm: Pickup Rocks & Sticks	09/16/20	42
Help Set Up the Pride for Parents Christmas Store - Downtown Raleigh	11/18/20	42
Fight Hunger with the Food Bank in 2021 (Raleigh)	12/29/20	42
2020 Int'l Festival - Merch & Souvenirs Volunteer	01/03/20, 12/03/19	41
Internship	11/11/20	41
Volunteer Artists Need: ART IN THE GARDEN	04/22/20	40
Food Pantry & Distribution volunteers	09/23/20	40
11th Annual Book Bag Drive "Thru"	08/12/20	40
Triangle Golf Tournament Event Volunteer	10/07/20	40
IN PERSON - Wrap Diapers for Babies at the Diaper Bank (Durham)	12/21/20	40
Special Olympics North Carolina Spring Internships	01/03/20, 01/31/20	39
BUMP Website Updates	03/17/20	38
Build a Stage for Graduation Ceremonies!	06/24/20	38
Stock food for families in crisis with Wake Relief	08/26/20	38
2020 First Friday Fun at Artspace!	01/02/19	37
Web Developer/E-Commerce Developer	05/13/20	37
FIGHT HUNGER WITH THE FOOD BANK (DURHAM)	10/28/20, 11/11/20	37
Power Hour Volunteers	01/03/20	34
Medical Interpreters (Spanish)	03/04/19, 06/10/20	33
NC Therapeutic Riding Center	01/02/19	33
Event Photographer	10/08/19	33
Meet the Heroes Black History Festival	01/31/20	33

Title	Dates Sent	Click Count
Virtual Volunteer Orientation	03/17/20	33
Deliver Clothes from Bins Off-site to Bin On-site	03/17/20	33
Dress For Success Sale event	01/31/20	32
ReStore Volunteer	11/04/20	31
Sort Donations at The Women's Center	07/08/19, 08/05/19	30
Housing Advisor	01/03/20	30
Sorting donations	12/03/19	30
February Workday	01/31/20	30
13th Annual Pepper Festival	09/16/20	30
JA of Eastern NC - Virtual Volunteer Career Speaker Series Video	04/28/20	30
UWH - Fight Hunger with the Food Bank (Durham)	09/02/20, 09/09/20	30
Food Pantry Distribution Setup!	08/19/20	30
Community Service @ Farm	03/04/19, 06/04/19, 07/08/19	29
Seeking New Board Members	11/01/19, 12/03/19	29
Greeter/Donations Monitor - MORNING	11/11/20	29
Volunteer in the Gardens	01/03/20	27
Office Assistant	01/31/20	27
REMOTE - Create BOOST Videos About Academic Concepts	11/04/20	27
Sorting Donations	11/01/19	26
Volunteer Bookkeeper	01/03/20	26
Garden Group Non Profit Interest Meeting	07/08/20	26
Durham Grocery Bags for Seniors	01/02/19	25
Grocery Bags for Seniors Meal Delivery	05/01/19	25
Budgeting Advisor	10/08/19	25
Social Media Volunteer	08/19/20	25
Volunteer Administrative Assistant	12/03/19	24
REMOTE - National Voter Registration Day Calls	09/02/20	24
Garden Work Day at Alliance!	12/03/19	23
Grant writer needed ASAP	01/31/20	23
Educational Farm- Social Media and Email Help	01/31/20	23
Refrigerator Expert Needed!	08/05/20	23
Fundraising/Development/Grant-Writing Volunteer	01/02/19	22
Free Food Fridays	01/02/19	21
Recycle Drop-Off Volunteer	06/04/19, 07/08/19	21
Nonprofit Event Planning Committee	01/03/20	21
POSTPONED Until 2020 Youth Tournament @ NCSU	12/03/19	21
Program Data and Evaluation	01/31/20	21
School Tour Education Volunteer	01/31/20	21
Meals On Wheels lunch meal delivery	11/01/19	21
Volunteers Transportation Drivers	10/08/19	20
Sports & Recreation	10/08/19	20
Pick up items in Raleigh/Cary deliver to Durham	12/03/19	20
Community Calendar Postings	05/20/20	20
Summer Art Camp - Volunteer Teaching Assistants!	04/03/19, 07/08/19, 08/05/19	20
Fundraiser & Event Coordinator	04/03/19, 05/01/19	19
Remote Opportunity for Social Media Ambassadors!	09/02/20	19
Tree Giveaway at NC Museum of Natural Sciences	03/04/19	18
North Carolina Museum Park Workday	02/01/19	17
iCan Bike Camp	07/08/19	17

Title	Dates Sent	Click Count
Social Media/Mixed Media Assistance	12/03/19	17
Campus Beautification Day	10/08/19	17
Women's Theatre Festival Volunteers	07/08/19	16
AIDS Walk and 5K Run 2019	10/08/19	16
Wildcat Station Volunteer	10/08/19	14
Movers needed	08/19/20	14
Raleigh St. Patrick's Day Parade and Festival 2019	03/04/19	13
Kitten Schmoozer Training Sesh	04/03/19	13
YMCA 175th Anniversary Day of Service	06/04/19	13
Arts & Crafts	11/01/19	13
Basketball and Tennis Instructors Needed	01/02/19	12
Provide Childcare at Families Moving Forward	02/01/19	12
Youth Small Biz Coach	03/04/19	11
Volunteer Opportunities-Miracle League	01/02/19	11
Hugh O'Brian You Leadership Facilitator	06/04/19	11
Helpers needed for Crosby Community Day 7/19/19	07/08/19	11
Horse Leader for Therapeutic Riding Lessons	05/01/19	10
March Workday at the Hub Farm	02/01/19	10
Art Program Volunteer	02/01/19	10
Stocking and Warehouse Assistance	06/04/19	10
Photographer needed for special event	05/01/19	10
Volunteer at the Triangle Walk to Defeat ALS®!	04/03/19	9
Host Families for International Youth Exchange	08/05/19	9
Help a Refugee Settle in our Community	05/01/19	9
Business Development Volunteer	11/01/19	9
Front Desk Attendant	04/03/19	8
Silent Auction Assistants!	03/04/19	8
Volunteer With High School Students for CIS	02/01/19	8
Live & Local Spring Fest Hillsborough Street	06/04/19	8
Assist teachers with the check out process	02/01/19	7
Artsplosure 2019 - May 18 and 19	05/01/19	7
Hoopla Hyperspace Party in the Park at the NCMA	05/01/19	7
Childcare Volunteers	08/05/19	7
Plant Seedlings / Flowers at the Community Farm	08/05/19	6
Cary ReStore - Skills Based Volunteer Opportunity	02/01/19	6
Ledge Creek Forest Conservation Area II Trail Day	03/04/19	6
Beautify a School (Durham, 4/12, Morning)	04/03/19	6
Youth Robotics Clinic	06/04/19	6
Certified Dietitian or Nutritionist needed	05/01/19	5
Families Together volunteer orientation	08/05/19	5
The Arc of the Triangle Family Fun Night Volunteer	08/05/19	5
Theater Seat Restoration at The ArtsCenter	11/01/19	5
Attorney for A Nonprofit	01/02/19	4
Boys High School Ultimate Frisbee Coach	02/01/19	4
Sound the Alarm with the American Red Cross	04/03/19	3
2019 Packapalooza Intercept Survey	08/05/19	3
Director of Fundraising	11/01/19	3
Volunteer Board Member	04/03/19	2
Saturday Garden Work Day at Alliance!	07/08/19	2