Email Response Time

FINAL PROJECT DESCRIPTION

Junkai Qiu (jq480 jq480@nyu.edu)
Jiahao Huang(jh4849 jh4849@nyu.edu)
Gulin Fu (gf926 gf926@nyu.edu)
Jun Li(jl7338 jl7338@nyu.edu)

Project page (on Github): https://github.com/NYU-CS6313-Fall16/Email-Response-Time-8
Video: https://vimeo.com/196799748
Working demo: https://junkaiqiu.github.io/info_visual/

What is the problem you want to solve and who has this problem?

The problem we want to solve is how long it takes to people to respond to emails in a given email archive. And for each user, who is the most active replier and who is the least.

What are the driving analytical questions you want to be able to answer with your visualization?

1. Do people who reply mails fast also receive quick responses?
2. What's the percentage of quick/slow repliers for a certain user?
3. How the email response time distributed among all the repliers?
4. Who has the most active relationship with a certain user? And the least active relationship?

What does your data look like? Where does it come from? What real-world phenomena does it capture?

Data:


```
"Andrea_Ferguson@inbox.com":
{"repliedTime":51342600,
"recNum":60,
"replyTime":46096592.59259259,
"responseObj":{
        "Louis_Oliver@inbox.com":64426261.9047619,
        "Tina_Sanchez@gmail.com":27497100,
        "Paula_Fernandez@gmail.com":5438250,
        "Margaret_Hughes@outlook.com":31737000,
        "Martin_Murphy@hotmail.com":23096000}}
```


attribute name	attribute type	description	value range	derived or not

user address	string	identify a certain user	categories	Yes, random generate from hashed address
repliedTime	numeric	average response time other users reply to this user	[0, 1200000000] Unit: ms	Yes, calculated from given response time
recNum	numeric	the number of email this user received	>0	Yes, count given response emails
replyTime	numeric	average response time this user reply to other users	[0, 1200000000] Unit: ms	Yes, calculated from given response time
responseObj	array of object	users who email to this address and their response time	categories	yes

What have others done to solve this or related problems?
Paper 1: *A Comparison of Response Rate, Response Time, and Costs of Mail and Electronic Surveys* by David M. Shannon & Carol C. Bradshaw

Link: http://www.tandfonline.com/doi/abs/10.1080/00220970209599505

The survey focused on percentage of response time over time for email which inspire us percentage to analyze the distribution of quick and slow repliers.

Paper 2: *SeeMail: Visualizing Email Response* by Laura Dabbish & Zachary Wise

Link: http://www.personalinformatics.org/docs/chi2011/dabbish.pdf
Explanation: SeeMail processes user email headers to generate visual summaries of response behavior.
Image: Average response times for a specified user.

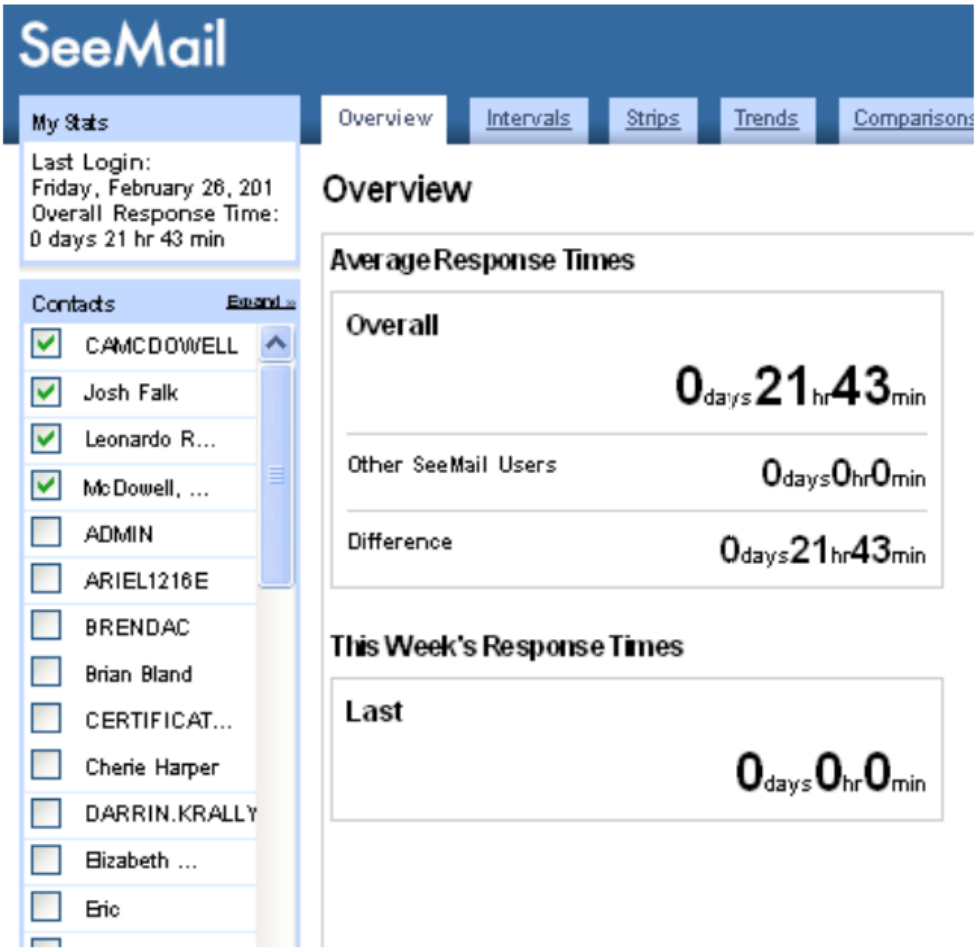
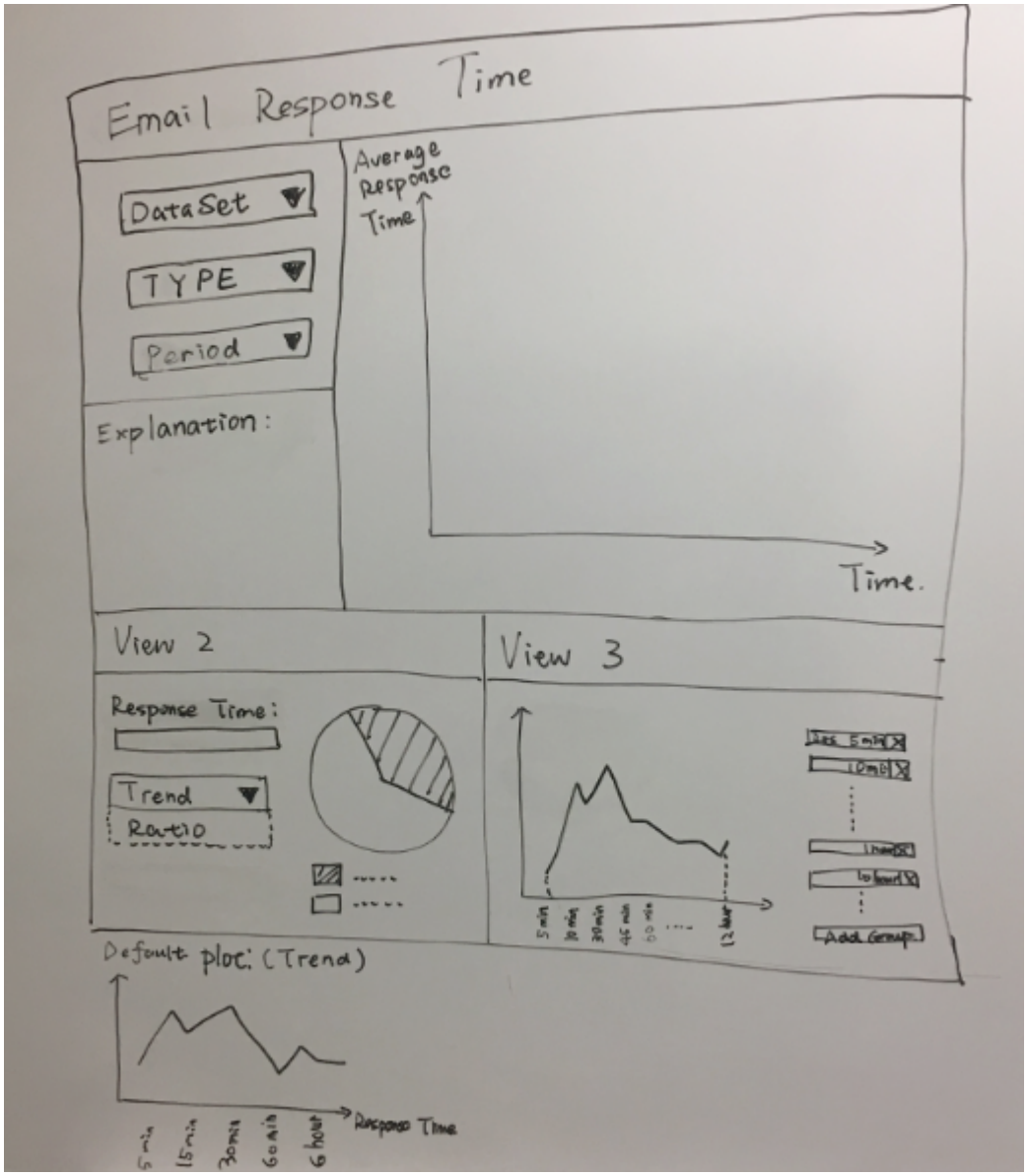


Figure 1: SeeMail Overview page presenting aggregated response time across all messages.

This paper inspire us to focus on analyzing the relationship between user average response time and people response email habits.

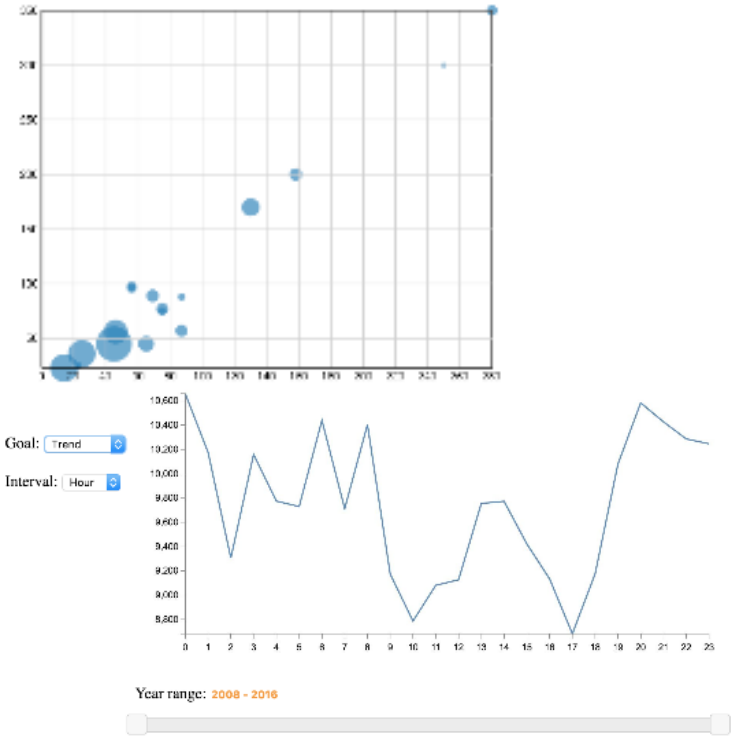
Design Iterations

First iteration:



1.We focused on the relationship between response time from other and response time to others.
We used three dropdowns to allow user to change the period, dataset or the chart type (line chart or bar chart). Then the mentor opinion is the drop downs can't directly reflect the questions. The user may need to think what this button mean and what he/she should do with these.
2. For the second chart we also implement 2 kind of charts: pie and line chart. User can input their own definition of quick replier.
3. For the third view we implement a line chart that indicate how the response time change in different time period. Like the response time trend in a day/week or even year. User can also edit the x axis and the line chart will change according to the input.

Second Iteration:



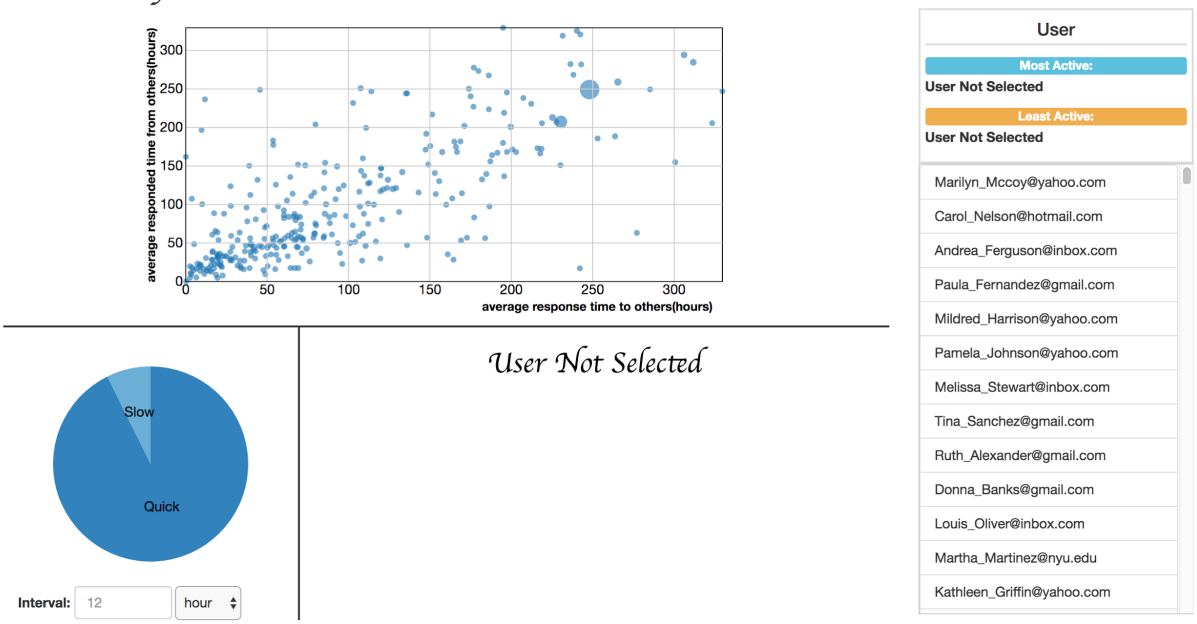
Based on the feedback in first mock up, we change the drop down into showing the time interval and what goal this chart tends to complete, which can directly reflect the visualization target.

But during second mock up, we received feedback that people already knew the pattern between email response time and day time period. For example, people know that you have to wait a long time for response if you send an email during midnight.

The professor/TA are more interested in the response time relationship with people. So we change our direction into find out the information about different users, which require user to input, (in this case select), an username and we can dig into the relationship between this user and other users.

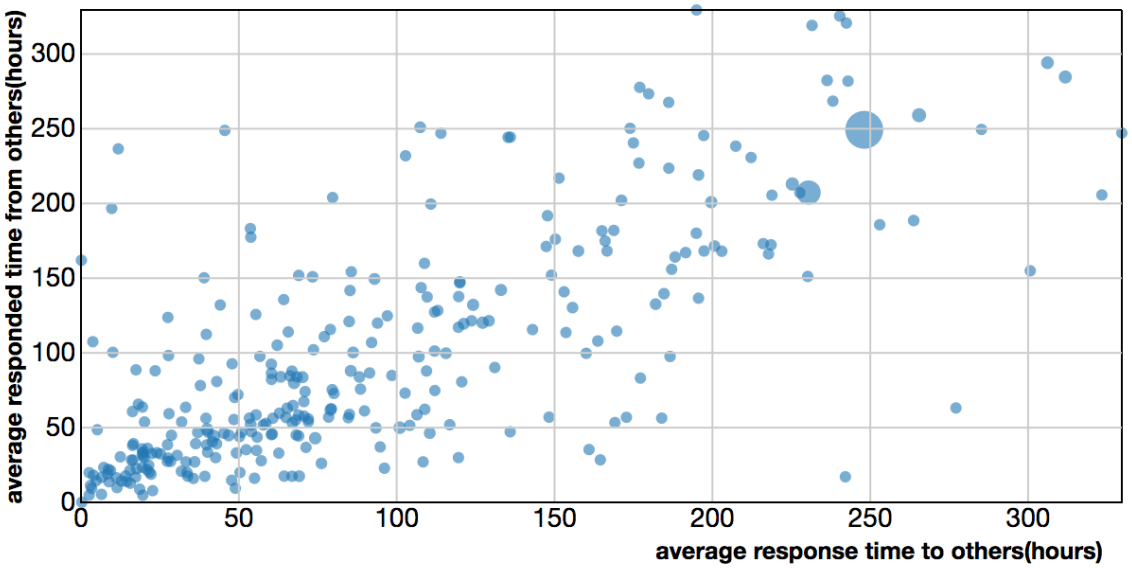
The main question remains the same: show if there is certain relationship between response time to others and response time from others.

Final Visualization



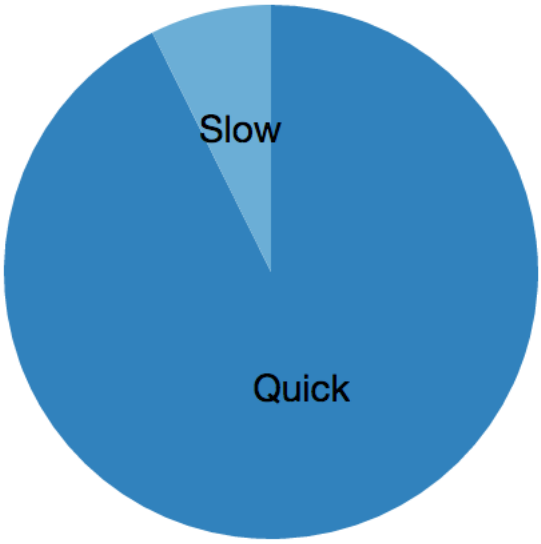
After two iterations, we made such final visualization.

Main view:



The x axis is the average response time user reply to others, the y axis is the average response time others reply to user. If x / y is 45 degrees, we can have conclusion if user reply fast, he or she can also receive email quick.

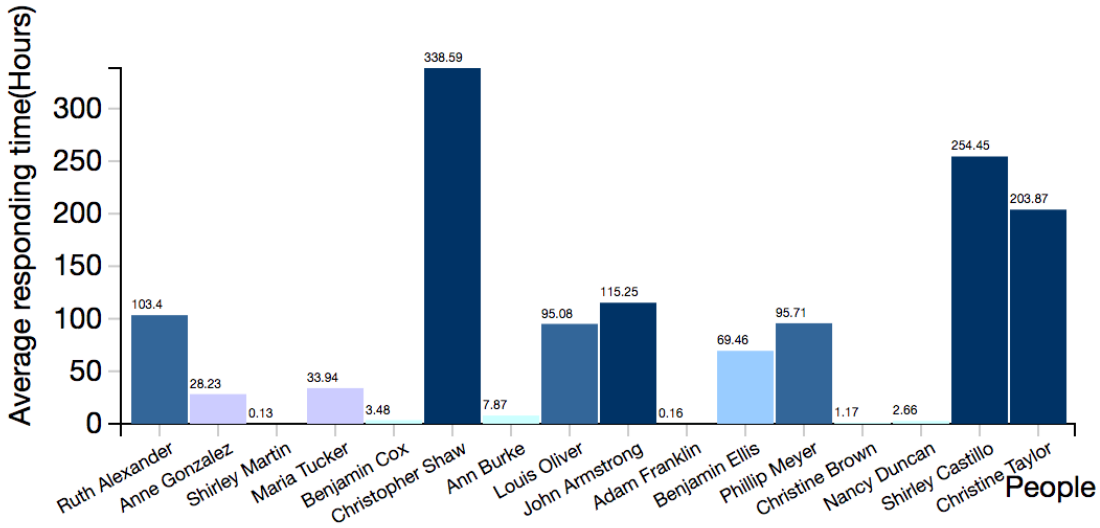
Second View:



Interval: 12 hour

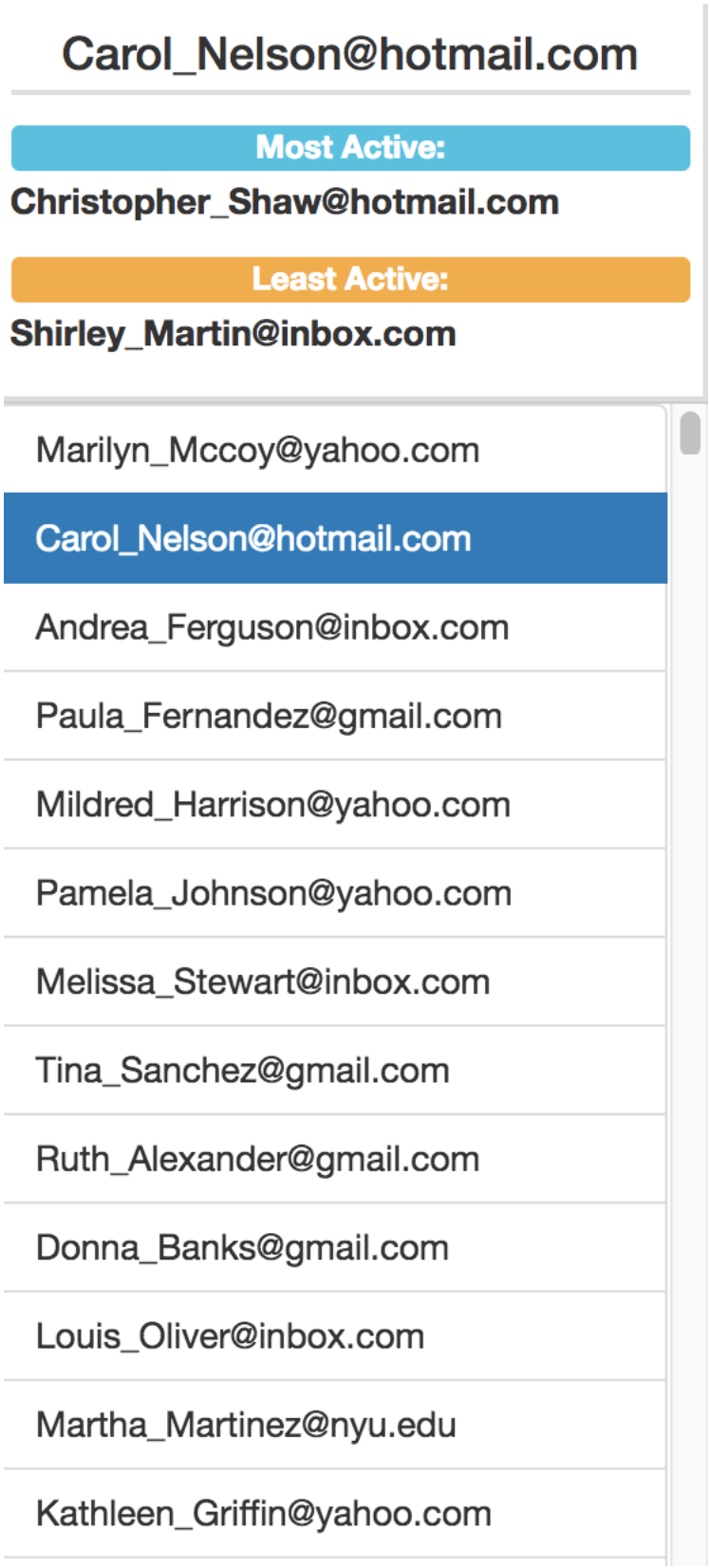
If users who response email in 12 hours, they are defined as quick repliers. So this chart shows us the distribution of slow and quick repliers.

Third view:



The x axis lists all the users' response to the selected user, the y axis shows average responding time in hour. So the higher a bar is, the slow he/she response.

User list:



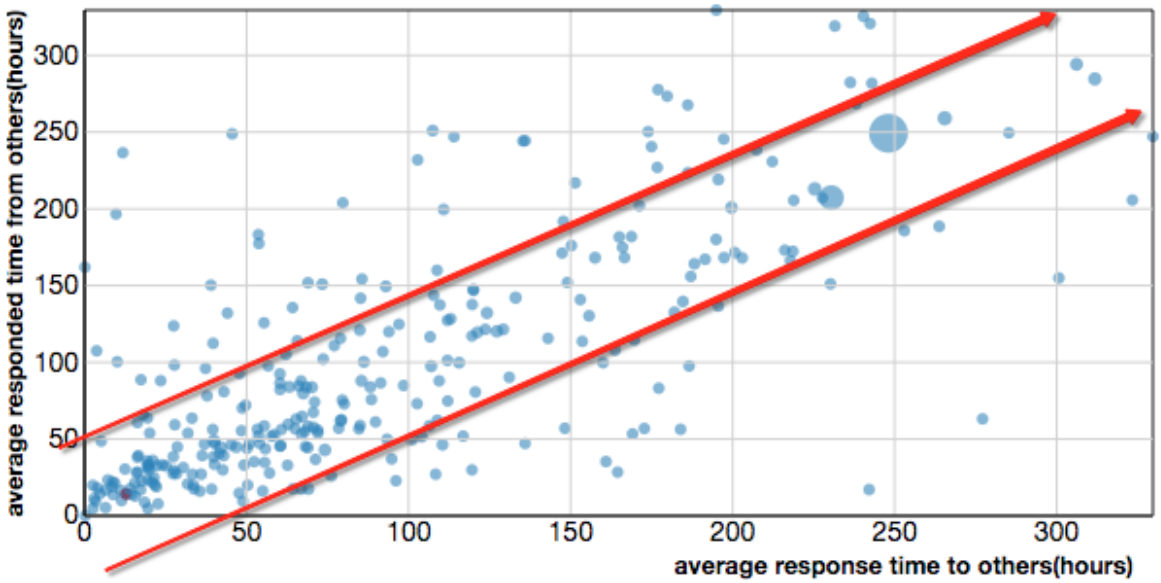
The first line is address of selected user.
Then displays the most active and least active user related to the selected user.

And the rest is a list of all users, once a user is selected, it will roll to the position of this user and highlight both the name and the circle he/she represent in the first chart. Also the second and third charts will use data of selected user to regenerate.

Insert images of your final visualization, describe how to read it

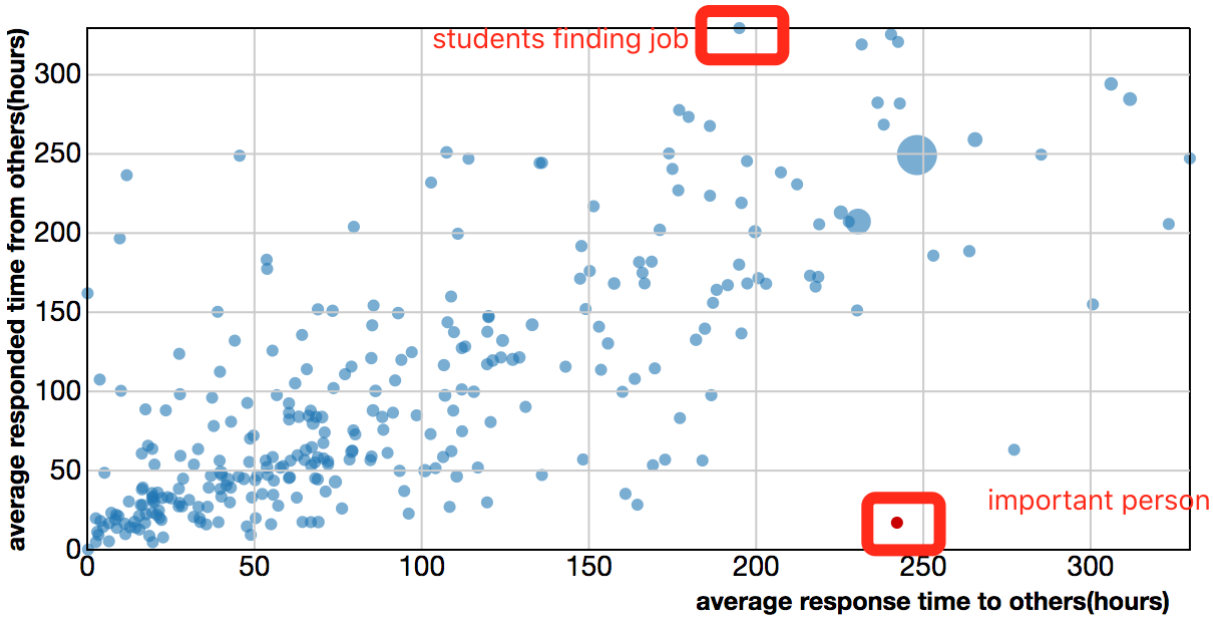
Findings

1. Our major conclusion is there exists a pattern indicates that users who reply fast also receive quick responses.



As you can see from the graph, majority of bubbles gather in and form an approximate 45 degree space, which means the response time from others and response time to other has linear relationship.

2. Email response time vary based on who you send or receive email with. For example, someone important may reply slowly even when you reply him/her quickly.



Using our develop tool, you can simply select the bubble that depart from the main pattern. For example, the lower side bubble represents an important person, and for the upper side bubble, it just like us student sending email to HR in huge companies.

Limitations and Future Works

Limitations:

1. The raw dataset only contains hashed address, so we randomly generate user names, which means that the user name in visualization represent nothing but placeholder.
2. The dataset is a static json file, we cannot dynamically analyze user response time without real time data input.
3. After filtering the dataset (eliminating emails that never received any response or emails contains invalid characters), we have only 300 Users and 5000 emails, which is too small to support our conclusion.

Future Works:

1. We can set up a back-end server, writing a auto web crawler to fetch data from the internet. But since email content may involve personal information, we need to be very careful in the fetching process.
2. The original data sometime contain invalid character and cause unfixable date conversion error, but we may do some research on the data structure or build a invalid data standard to solve such problem.
3. We may attach this visualization with current data analyzer like google analytics solution, which has huge email dataset in its database, and bring out better performance (may be dynamic data process and visualization).