Electronic communication has taken a massive role in modern human communication. It has infiltrated the workplace, our social lives and many other facets of our life. Text messaging, email, and social networks have changed the way we communicate and interact. Supporters of these technologies laud the way they in which they improve our efficiency, productivity, organization, and ability to connect with a larger network. Critics suggest that the new forms of communication have reduced the quality of social interaction and are distracting. Whether good or bad, these technologies look to be here to stay, and it is both interesting and necessary to study their effects on society.

One of the great things about electronic communication is that it is easily and automatically stored, and therefore lends itself well to analysis. My project, therefore, is to analyze and visualize this data. The analysis will initially be based on text message data, which, for Apple’s Messages app, is stored automatically both on the iPhone as well as Mac computers. This data is stored in “~/Library/Messages/chat.db” and includes fields such as the date and time of the message, whether the message was sent or received, the text of the message, the length of the message (in characters), and the contact who sent the message, among other things. There are also iPhone backups stored in “~/Library/Application Support/MobileSync/Backup/”. In theory these should two data sources should be identical up to a certain point in history, but in reality they often are not. I will
develop and test my project using my own personal data set, but the goal of the project will be to create a program by which any user can run these visualizations on his or her own data. From text messages, I hope to expand to other forms of communication, both social and professional, such as email data (available through the Gmail API) and Facebook messenger data (available through Facebook's chat API). Both of these data sources will have fields similar to those taken from the chat.db: date, time, sent/received, text, length, and contact. The project is meant to serve as both a source of enjoyment for the user and as an insight into his or her relationships as well as the nature of electronic communication in general. The bulk of the project will be centered on the analysis of and, especially, the visualization of the text message data. The other sources of communications data will mainly serve as a contrast to the text message data.

Definite deliverables for the project are:

• Aggregation, analysis, and compression of the various forms of data
  o Code will involve selecting the data from ~/Library/Messages/chat.db and from iPhone backups, and organizing into pandas dataframes
• Visualizations, including:
  o A depiction of the contacts with which the user communicates most
    ▪ This will be in the form of a scatter plot, with the x-axis representing the number of messages sent to the contact and the y-axis representing the number of messages received
  o Changes of communication levels by contacts by date and by time of day
  o The subject of communications (a word cloud and a word tree)
  o Response time plot
• Cross filter options for these visualizations
• The ability to search for conversations by contact and by date
• A user-friendly experience that enables easy collection of the data and browsing of the visualizations

Areas that I would like to explore after writing the functionality above:
• Ability to compare your visualizations with other users
  o For example, a visualization showing the difference between the temporal patterns in two users texting habits
• Intelligent labeling of conversations into categories such as: logistics, collaboration, conversation, etc.
• Relationship analysis
• Integrations with other communications apps, such as those mentioned above
• Sentiment analysis

Questions I would like to answer and patterns I would like to see:
• Are your closest contacts those you communicate with at constant density and frequency, or varying density/frequency?
• How does the time of day and date affect your communication?
• What is the focus of your conversations, and how does this change over time?
• Do people communicate in similar ways (i.e., are there significant differences in the patterns discovered by the visualizations)?

The main technical challenge will be learning d3.js and creating a wide variety of visualizations. A related challenge will be preparing the data to be visualized using these tools. Also, the storage of the data will be a large part of the project, particularly if the user-to-user comparison feature comes into fruition.