LTS-2 System:
Exploratory Analysis of Service Usage Over Time and Space

The System:

The LTS-2 is an SMS information system designed to help Mongolian Nomadic herders to improve lives outcomes. The system works by providing weather forecast and pasture information to herders by text message. Anyone with a Mongolian phone number can text into the system, though the information provided is tailored for nomadic herders. In particular, the system aims to reduce livestock deaths due to severe weather, and ensure long-term sustainability of herding by reducing overgrazing.

Users can request weather or pasture information via text. One to three, and four day forecasts are available, in addition to pasture information pulled from the Livestock Early Warning system run by Texas A&M University. To receive a text message response, users must request information for a specific bagh (subdistrict). The data we analyzed is comprised of all the incoming and outgoing texts associated with the LTS-2 system. Incoming texts consist of a bagh-level area code and an integer from 1-3 indicating request type. 1 and 2 represent one- to three-day and four- to six-day forecast respectively, while 3 represents a request for pasture information. The formatting of an incoming text must conform to fairly strict rules for a user to receive a response. The text must begin with a 5 digit bagh area code, followed by at least one space and the request type. Any change in formatting, or addition of other information will mean a user does not receive a response. In general, outgoing weather forecasts contain the starting date of the forecast interval, one of four predicted meteorological conditions (sun, clouds, rain, snow). When available, higher resolution forecasts that include high and low temperature are provided.

The Data

As described above the system is comprised of incoming and outgoing messages in the system. Due to the system's cataloging method, the dataset does not contain a response for every request, or a request for every response. The first is due to user error. Improperly formatted messages are catalogued but do not receive a response. The second, is a system error. In the dataset, these errors appear to be messages sent to phone numbers that never sent requests for information; however, these messages are due to transcription errors in the system's data collection method, and no actual messages were sent out. An update to the system in the fall of 2016 mostly solved the problem. These phantom messages appear only infrequently after September 2016.

The raw data we analyzed contained 125,000 observations spanning from June of 2016 to December of 2018. Observations were of individual messages, either ingoing and outgoing, and contained five variables: telephone number, date, time, message, and message type (incoming/outgoing). Since telephone numbers are traceable to individuals, we replace

telephone numbers with random unique identifiers. There are a roughly equivalent number of incoming and outgoing messages, with 61,000 incoming messages and 63,000 outgoing.

31	14278	2017-04-01	1899-12-31 08:11:21	in	62267 2
32	14278	2017-04-01	1899-12-31 08:12:17	out	4sar4: uulerheg
33	24056	2017-04-01	1899-12-31 08:18:57	in	23177 2
34	24056	2017-04-01	1899-12-31 08:19:18	out	4sar4: uulerheg
35	50171	2017-04-01	1899-12-31 08:36:27	in	62131 1
36	50171	2017-04-01	1899-12-31 08:37:25	out	4sar1: uulerheg

Fig 1: (above) Several lines of the de-identified, raw dataset.

Analysis:

Our analysis of the LTS-2 dataset focused identifying and describing patterns in system usage over time and space. In addition to generally describing patterns in system usage, the following analysis looks at the characteristics of individual users, with some attention to user and system errors. Our analysis included mapping of spatial usage of the system over time using an interactive shiny dashboard (URL).

In analyzing the ways the system is used, we focused on three major areas. First, characterizing the average user. Second, describing in general trends in system usage over time, and finally looking at the spatial distribution of usage.

Individual Users:

We began by characterizing the average user. In total, 11,520 different users requested information from the system between June 2016 and December. The majority of these users requested information relatively few times.

Mean	Median	Min	Max
4.63	2	1	492

Table 1: (above) Descriptive statistics of the number of the distribution of times individual users requested information.

The above table shows the descriptive statistics of the distribution of the number of times individual users requested information. With a mean of 4.63 and median of 2, the distribution of the number of requests by individual users, suggesting that many of the system's users stop using the system after a short time.

Type of Request	**Number of Requests**
1-3 day weather	36,188
4-6 day weather	14,628
Pasture	1,349

Table 2: Total counts of requests for each of the three types of information offered by the LTS-2

Users seem mostly interested in weather information. Of the approximately 52,000 correctly formatted messages (messages that contain only a five-digit area code, and a request code between 1 and 3), about 36,000 were for 1-3 day weather forecasts and 14,000 were for 4-6 day weather forecasts. In comparison, only about 1300 requests for pasture information, suggesting that weather information is the main type of information users are interested in.

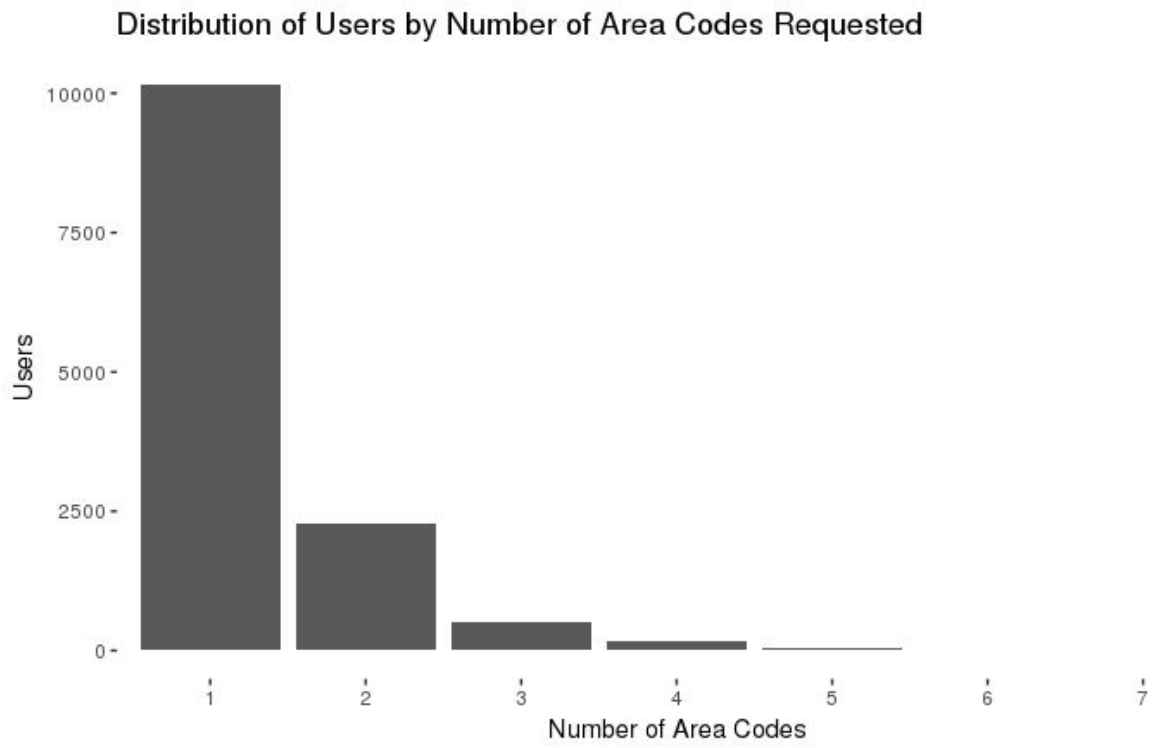


Fig. 2: The distribution of number of area codes requested by individual users. Most users only requested information for a single area code.

The above bar plot shows the distribution of number of area codes for which individual users requested information. Most users only requested information for one area code (around 10,000). Number of area codes drops off significantly after that, with about 2,500 users requesting 2. It's difficult to know whether this pattern is at all related to the geographical location of users, because many users only queried the system a single time; however, using the shiny dashboard mapping geographical usage, it's possible to gain some insight into the spatial distribution of users.

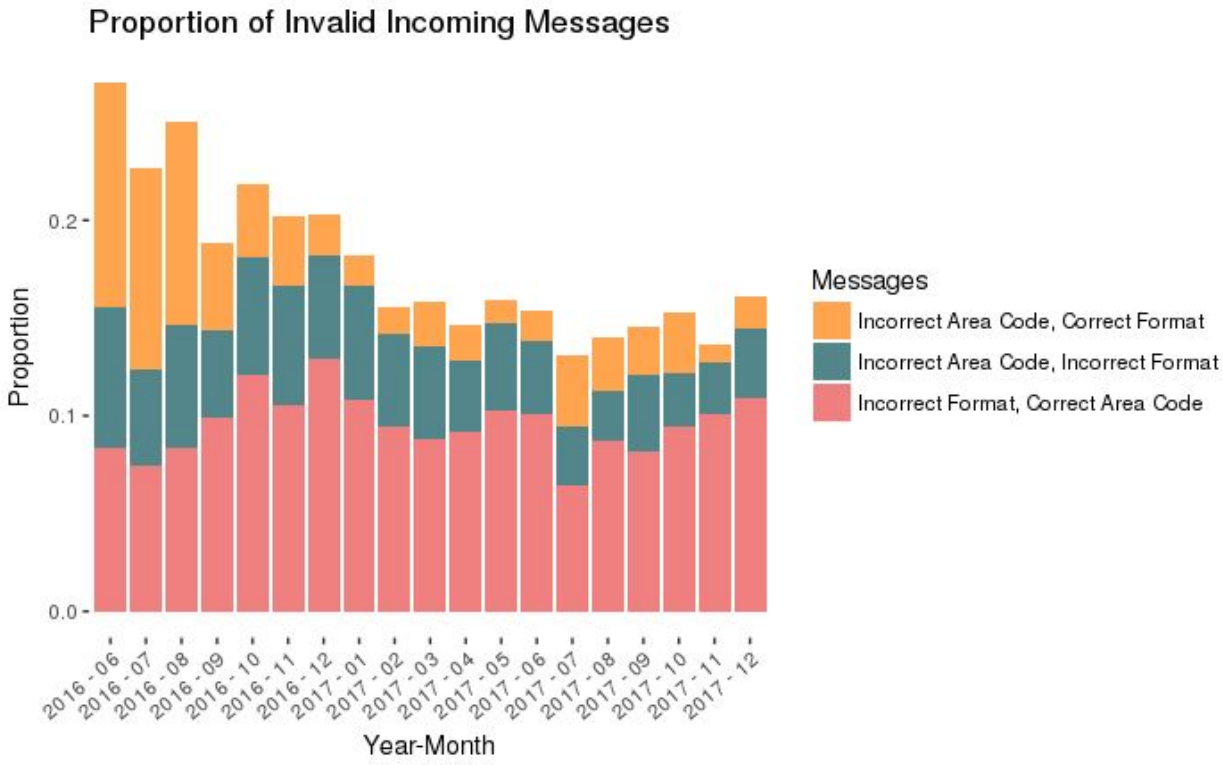
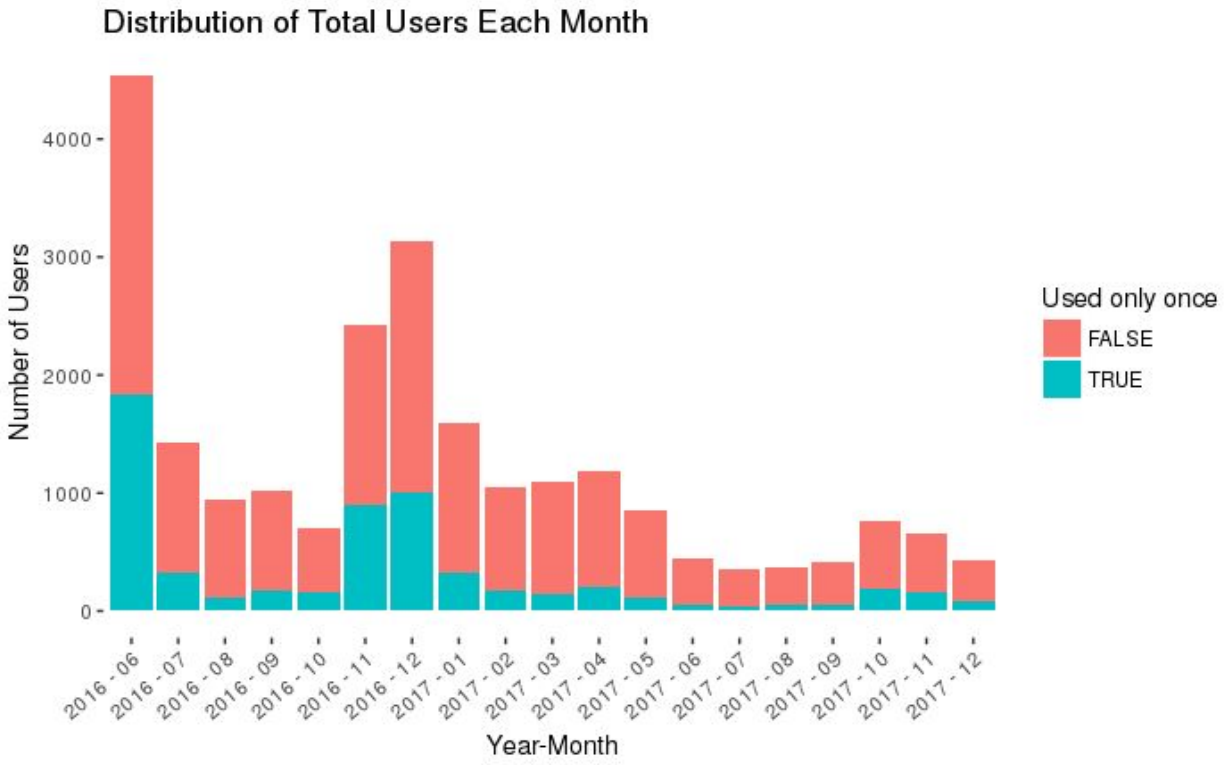


Fig. 3: Proportion

The graph above shows the proportions of invalid messages each month, which are caused by incorrectly formatting and/or area codes that do not correspond to actual places. We see an overall decreasing pattern in the total proportion of invalid requests. The proportion of messages with incorrect formatting does not change much over time, while the use of non-existing area codes in messages tends to decrease. Incorrect formatting appears to be the most common and consistent type of user errors.

Users were creative in their formatting of messages. Incorrect messages involved anything from requests for multiple types of information at a time or multiple area codes at a time, to entire

sentences describing the information a user wanted. Given the relatively high proportion of incorrectly formatted messages over time, it may be helpful for the service to respond with formatting instructions to any unreadable messages. Currently, incorrectly formatted messages do not receive a response, so providing a response to incorrectly formatted messages may improve users' interactions with the system.



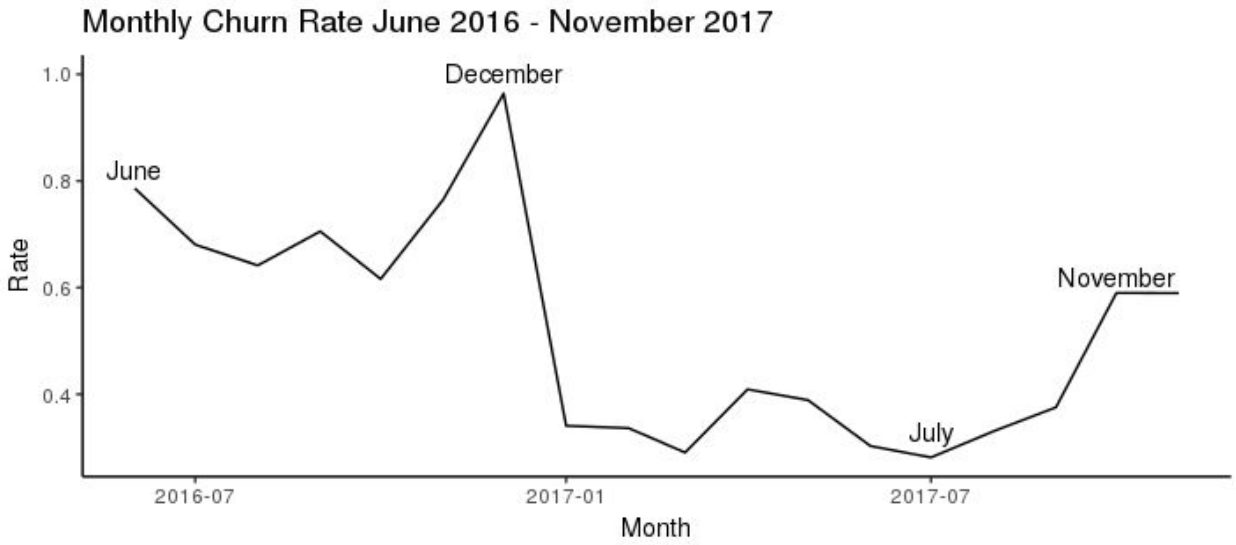
The graph above shows the total number of users that texted into the system each month, along with the proportion of users that only ever text into the system once. Clearly, a significant portion of users text into the LTS-2 system and then never use it again. Any conclusion about seasonal usage is difficult given the newness of the LTS system, though the apparent trend seems to lower usage in the summer months (June, July, August, and September). Three months, in particular, stand out both as having the most number of users and a large proportion of one-time users: June, November, and December 2016.

Looking at the ways individual users queried the system, two trends stand out. The first is that most users used the system a small number of times. The second is that the number of users is inconsistent over time.

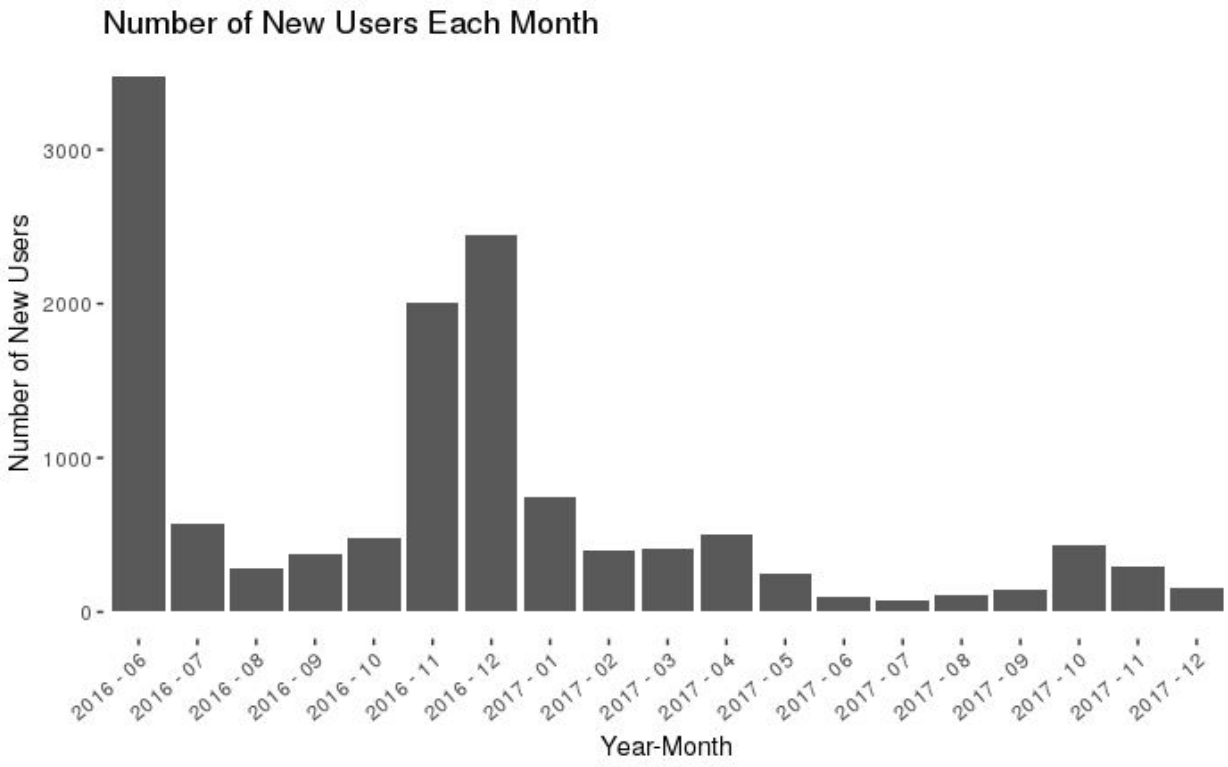
General Patterns in Usage

To get a better understanding of the ways users enter and exit the system, and to understand aggregate usage patterns, we looked at spatio-temporal usage.

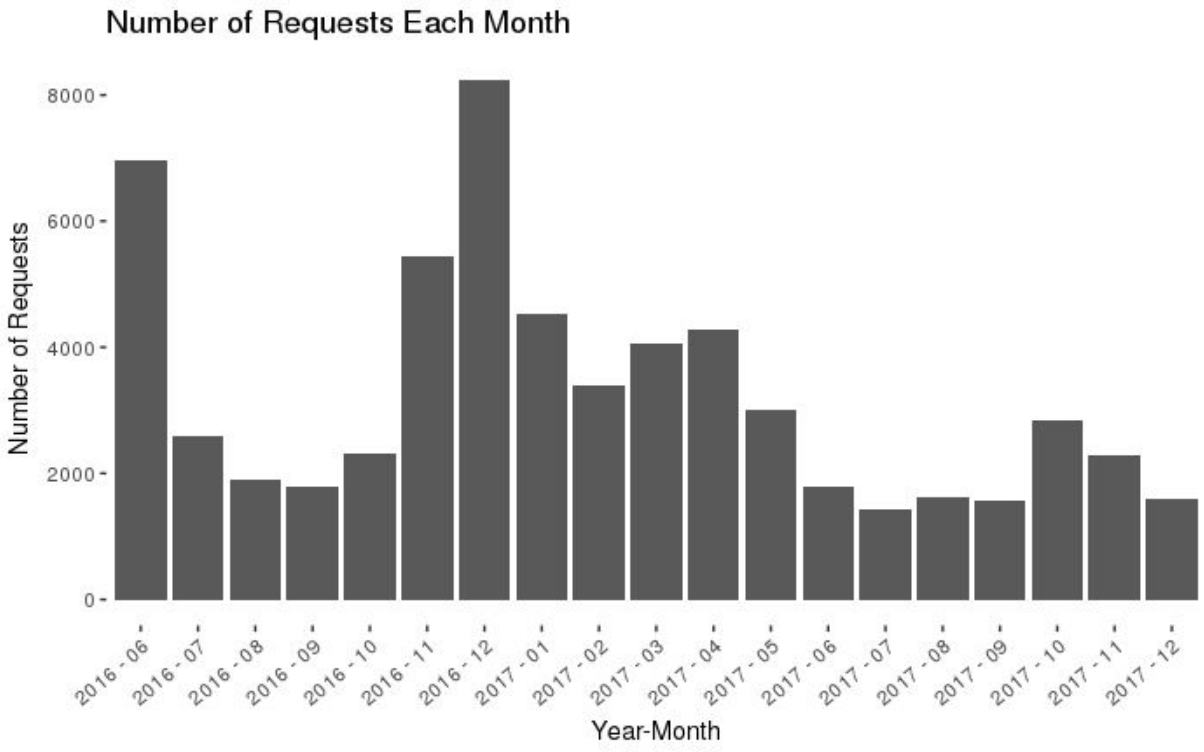
Since the dataset covers a relatively short period of time, we calculated monthly churn rates by dividing the number of users who used the system for the last time in a given month by the total number of users that month. The timeplot below shows this churn rate over time. Since we have no data for January 2018, December 2017 was excluded in the analysis.



The system's churn rate was relatively high, indicating that people are constantly entering and exiting the system. 2016 had consistently high churn rates, with a range between 60% and 80% between June and November of 2016, peaking in December at 96%, which was the highest churn rate throughout the year. Relative to 2016, 2017 saw low churn rates, with a range from



The above graph shows the number of new users over time. Generally, the number of new users entering the system has decreased over time, particularly from December 2016 onward. Taken in conjunction with the timeplot of churn rate (fig #), number of users seems to have declined over time, with a slight uptick in usage around winter time.



Mirroring the above graph of new users per month above (fig #), the number of requests for information directed at the service decreased over time, again with a slight uptick around late fall and winter. The number of messages per month shows a more substantial drop in usage around summer time than number of new users. Still, the plots seem to map fairly well onto each other, as well as onto the above plot of the total number of users per month over time (fig #). From this we conclude that there may be a trend in usage over time, but the time frame is too short to determine what processes might be behind that trend.

We expected to see a seasonal cycle in the usage of the LTS-2 system due to the seasonal nature of nomadic herding. The LTS-2 system experienced a usage spike during the winter of 2016. However, while there was also an uptick in requests during the winter of 2017, it was of a significantly smaller magnitude. The graph above also shows that usage during the summer months seems fairly consistent from 2016 to 2017.

Sources:
“Leveraging Tradition and Science in Disaster Risk Reduction in Mongolia - 2 (LTS2 Extension).” http://mercycorps.org.mn/beta/index.php/en/projects/lts-ii.