Data Exercise

Jennifer Benson

1. Data cleaning

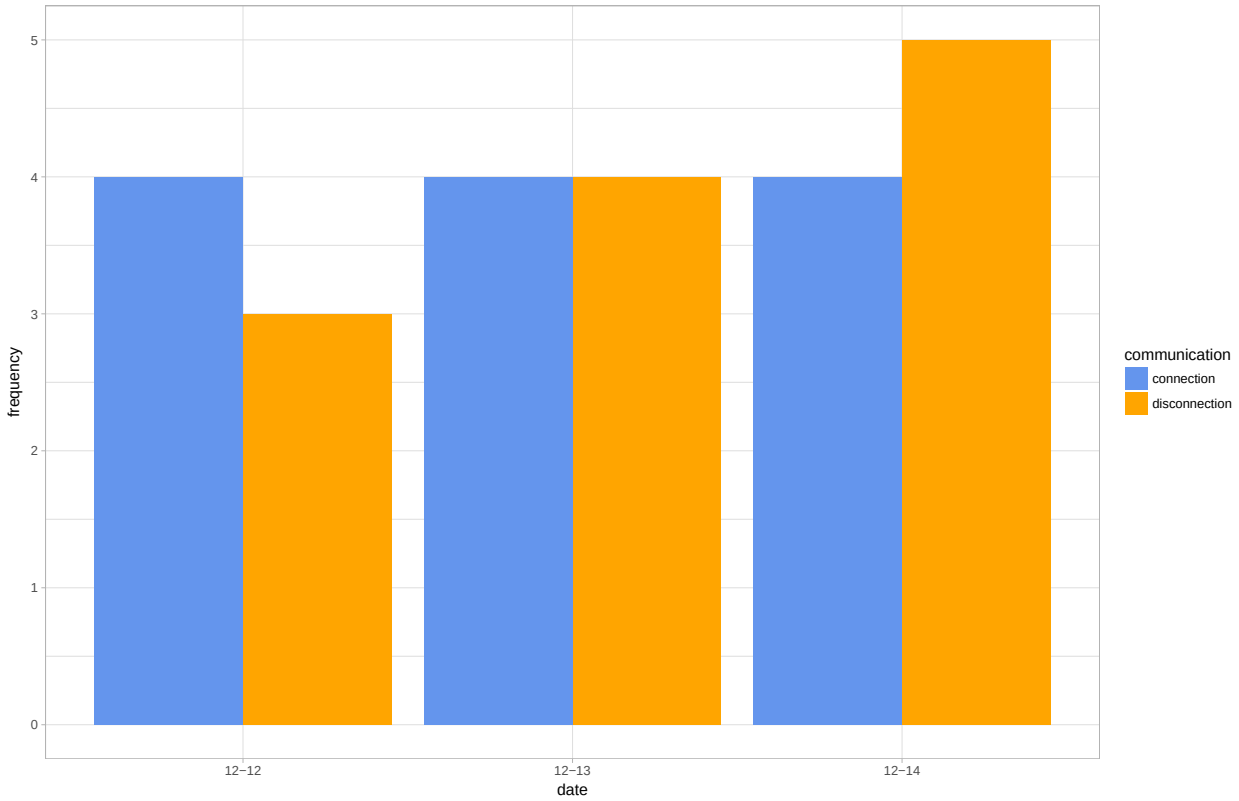
Before use in analysis, the data provided for the coding excerise had to be proccessed. Once the data was imported into R, the following files were merged: results-20171214-145849.csv and results-20171214-145922.csv. This was done by standardizing the datasets, which included creating a message type column denoting whether the message was a jsonPayload_event or jsonPayload_log to be able to differentiate between the data sets. Furthermore, the date column was converted to a date object to allow for data manipulation. There are no missing values in the data set and column types include string and dates; also, the data is representative of 12/12 through 12/14.

2. Connections

2.1 How often does the printer connect and disconnect?

Figure 1 shows connections and disconnections of the printer per day.

Figure 1: frequency of printer connections and disconnections



Furthermore, the printer connects an average of every 152.881 seconds and disconnects an average of every 185.9528 seconds.

Table 1: time elapsed between printer connections and disconnections

communication	mean	median	range
connection	152.8810	29.54102	1.90755 - 1001.000717
disconnection	185.9526	108.70587	1.596317 - 1001.0012

The time between connections was calculated using results-20171214-145849.csv, which contains logs; the message "WebSocketCleint: connection established" was used to determine how often the printer connects. How often the printer connects was calculated using two different metrics: by grouping the messages by day and calculating the time between connections as well as frequency counts of connections per day. The aformentioned methods were applied to results-20171214-145922.csv using the message "ws:close".

2.2 What is the average connection time?

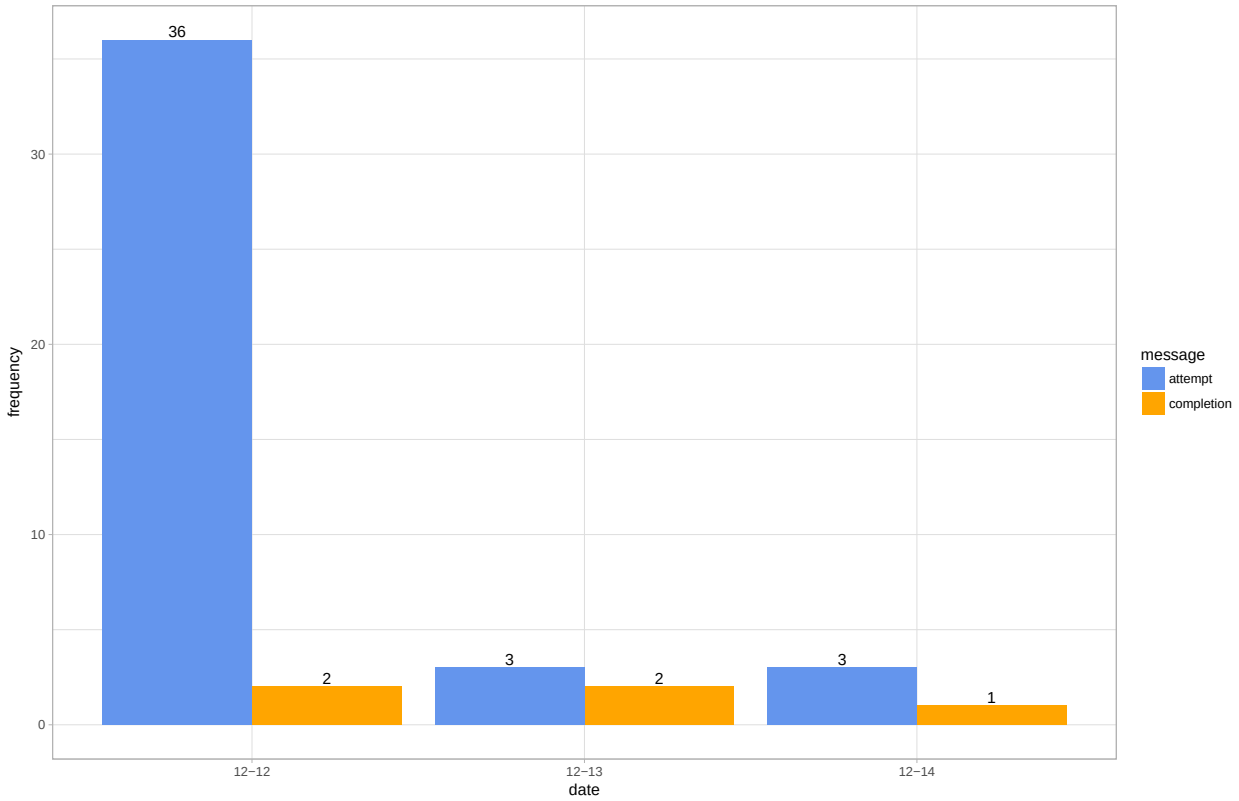
The average connection time is 11.3442 seconds. Average connection time was calculated by taking the average of the time elapsed between sequential connections and disconnections. Since the data is a snapshot of continuous data collection, there were some assumptions that had to be made with "extra" connections and closes that could not be matched. For example, a connection was intitiated on 2017-12-14 at 22:41:07.401000 with no associated closure in this snapshot of the data, which was subsequently removed. Furthermore, there was a disconnection that could not be matched with a connection; in this case, the first disconnection was removed from the data used for this calculation with the reasoning that the extra disconnection came from a connection implemented in a previous snapshot.

3. Head-hunts

3.1 How often (over what period) does the printer perform head-hunt?

The printer performs head-hunt an average of 8.4 times per request, which was calculated by dividing the frequency of head-hunt actions performed by the frequency of successful completions of head-hunts. Futhermore, Figure 2 shows the frequency of head-hunt requests and successful completions per day. As shown in the table, the head-hunts performed per request has a minimum of 1.5 on 12/13, which indicates more efficiency, to a maximum of 18 on 12/12.

Figure 2: frequency of head−hunt requests and completions



3.2 Is the head-hunt result always the same?

The head-hunt result is either a success or failure, but if it fails the head-hunt is repeated until success is achieved. Thus, in the end the same result, a successful head-hunt, is achieved. As shown in Table 2, the printer re-attempts to perform the head-hunt until a successful outcome occurs.

Table 2: Subset of sequential hunt failure and success calls

time	message	messageType
2017-12-14 00:00:12.335000	hunt:received	jsonPayload_event
2017-12-14 00:00:12.430000	hunt:failed	jsonPayload_event
2017-12-14 00:00:16.213000	hunt:received	jsonPayload_event
2017-12-14 00:00:16.326997	hunt:failed	jsonPayload_event
2017-12-14 00:00:31.897000	hunt:received	jsonPayload_event
2017-12-14 00:00:40.973000	hunt:finished:succeeded	jsonPayload_event

3.3 Does it always require the same amount of time to complete a hunt?

There are two parts to this question:
* time elapsed between a successful pair of “hunt:received” and “hunt:finished:succeeded” calls
* time elapsed between an initital, failed “hunt:received” call and “hunt:finished:succeeded”

Both of these questions are relevant and inform different aspects of the head-hunt procces.

As seen in Table 3, it does not always require the same amount of time to complete a hunt; the mean time to

complete a hunt is 15.00318 seconds, the median is 8.96598, and the range is 8.731 to 28.638.

Table 3: time in seconds elapsed to complete a hunt

time
19.788000
8.965998
8.892917
8.731000
28.638000

This variation in time to complete a hunt is driven by multiple failed attempts at performing the head-hunt action. This can be interpreted from Table 4, which shows that the average time between a successful "hunt:received" to a "hunt:finished:succeeded" message is consistent with a range of 8.965998-9.449998, mean of 9.051799, and median of 9.036.

Table 4: time in seconds elapsed to complete a successful hunt

time
9.449998
8.965998
9.036000
8.731000
9.076000

4. Change in machine settings

4.1 What is the rate of change of the "machine settings" over time?

The average rate of change of the "machine settings" was determine by the average time between completions as defined by a "settings:completed" message, which resulted in an average rate of change of 145.24 seconds.

4.2 How often does the server request "machine settings", and how does that compare with how often the printer updates them?

The server requests "machine settings" with approximately the same rate of change that they get updated; both have an average rate of change of 145.24 seconds.