Assignment 2 Social Media

Gijs Lakeman (s3383180)
Erik Storm (s3715671)

June 25, 2020

1 Questions

1.1 Q1 Efficiency

To measure the efficiency of the communication method. We used the duration of a game to compare the interfaces. This measures failed games as well as successful games. It takes this into account, because a game only fails because of overtime. So when a game has a long duration it negatively impacts the measure making it less efficient, like it should be. We aggregated the individual durations of the messages sent by their game number. We then categorised those numbers by their interface type and in what kind of experiment they were done. This data was then exported as a .csv file and imported into RStudio to make graphs and aggregate the data on different classes.

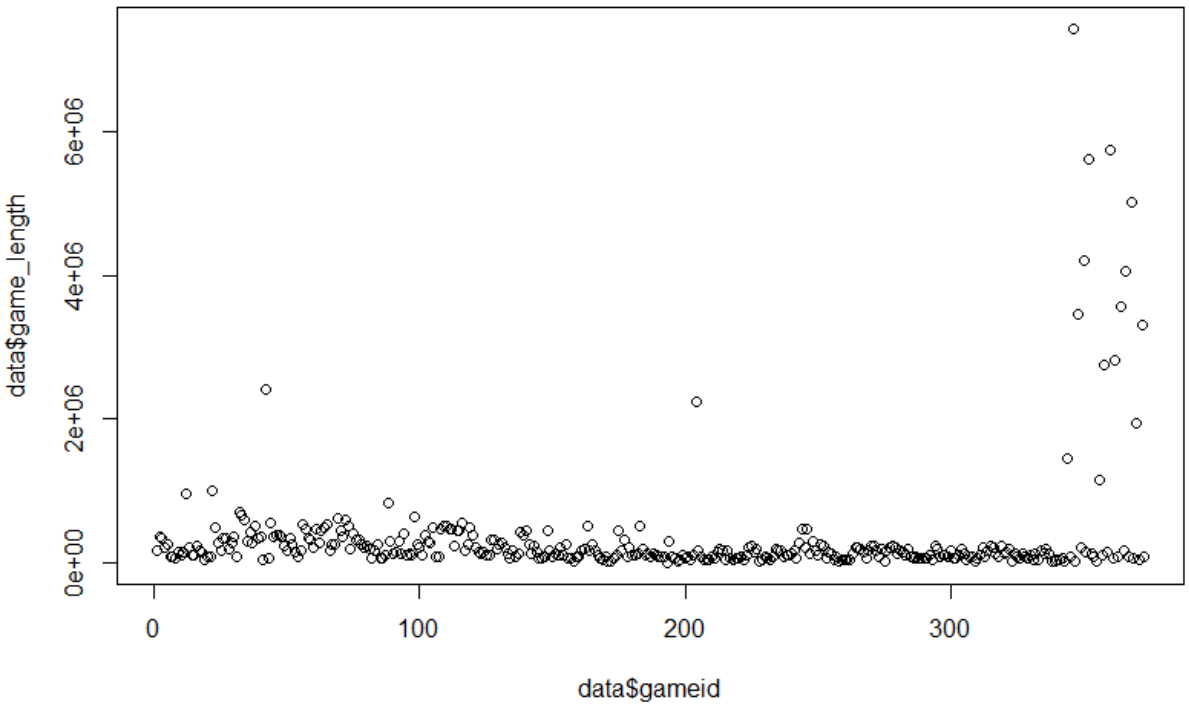

Figure 1: A scatterplot of the length of each game. It is clear that there are a lot of outliers.

The data shows that there are many outliers. Before we created this plot, we found that there was a large difference of gametime between interfacetypes and this scatterplot assured us that it was due to the outliers and not a valid reason.

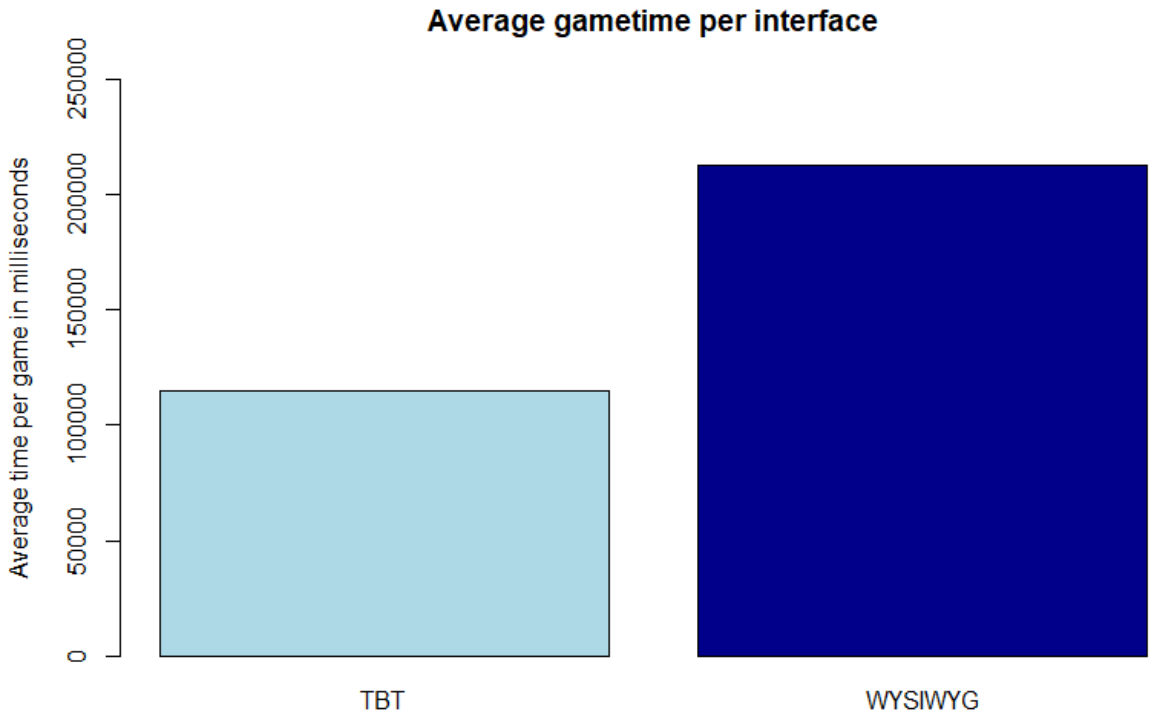

Figure 2: Barplots of the interface used and what experiment type the game was in. Time measured in milliseconds.

Once the outliers were removed at a breaking point of 500.000 milliseconds. We could then see that there was a big difference between the two interface types. The TBT method has almost half the time played compared to the WYSIWYG interface. Meaning that the TBT method is far more efficient than the WYSIWYG interface.

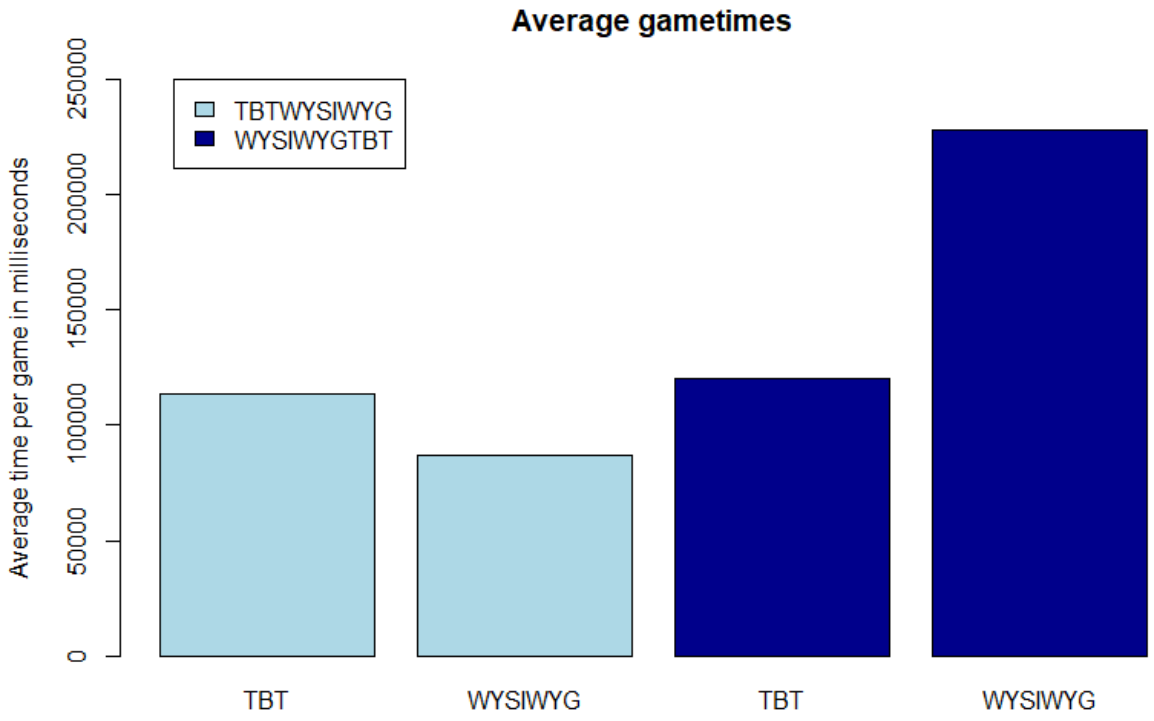

Figure 3: Average time it took to send a message per interface. Measured in milliseconds.

But this does not tell the whole story. There also were 2 types of experiments. One were the TBT method was tested first and one were it was tested last on a group. Which leads us to figure 3. Wherein you can see that on each experiment the second interface which is used by a group, has a lower game-time as compared to the first interface used. Which can be caused by the group better understanding how the game worked. What can also be seen is that TBT has roughly the same game-time in both experiments. WYSIWYG in the WYSIWYGTBT experiment has extremely longer game-time as compared to the other bars. This is most likely because participants do not know how the game works and the interface is unfamiliar. But when people know how the game works and people get to use the WYSIWYG interface, as in the TBTWYSIWYG experiment, participant outperform all other bars and making it the most efficient messenger.

1.2 Q2 Turn-Taking

For the measurement of the turn-taking, we used the average amount of messages sent before the other participant intervened. We chose this measure, because we think that a good conversation has participants taking turns after every message. This makes clear that the participants are actively interacting with each other and it is not like sending a long mail where after you return another long mail.

To calculate this, we counted the amount of messages a person sent before the other

person intervened. The count of the messages would then be added to a list of the specific interface type. Once the list was full, we then averaged the list and got the average amount of consecutive messages sent per interface type. The following barplot was generated in RStudio.

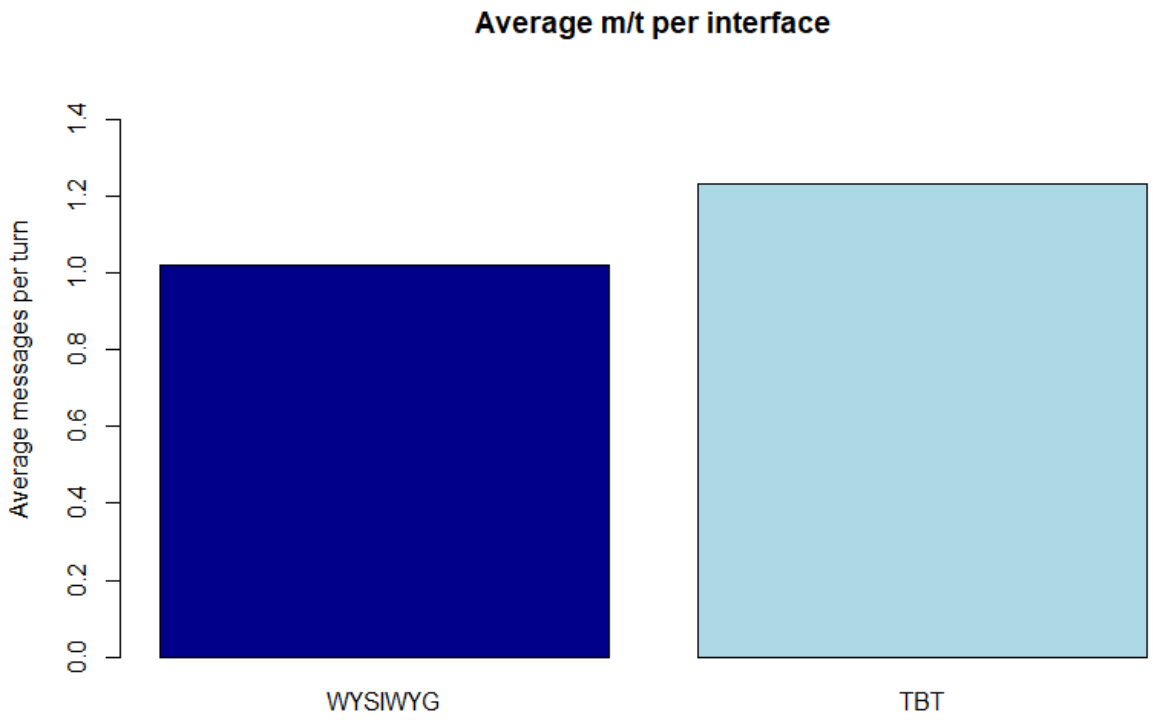

Figure 4: The average messages sent by a participant before the other participant interrupts with a message.

The figure above shows that the TBT interface on average uses more messages before the other user interrupts.

1.3 Q3 Language Copying

For this measure we calculated the percentage of words occurring in the previous turn compared to the next turn. To calculate this we took all the words per turn and compared them to the turn before the current turn. Then the number of overlapping words was counted and added to a list of overlapping word turn counts. Ranging from 0 to more than 4. All these counts were then divided by the total number of turns of the interface. This gives us the average number of overlapping words per turn.

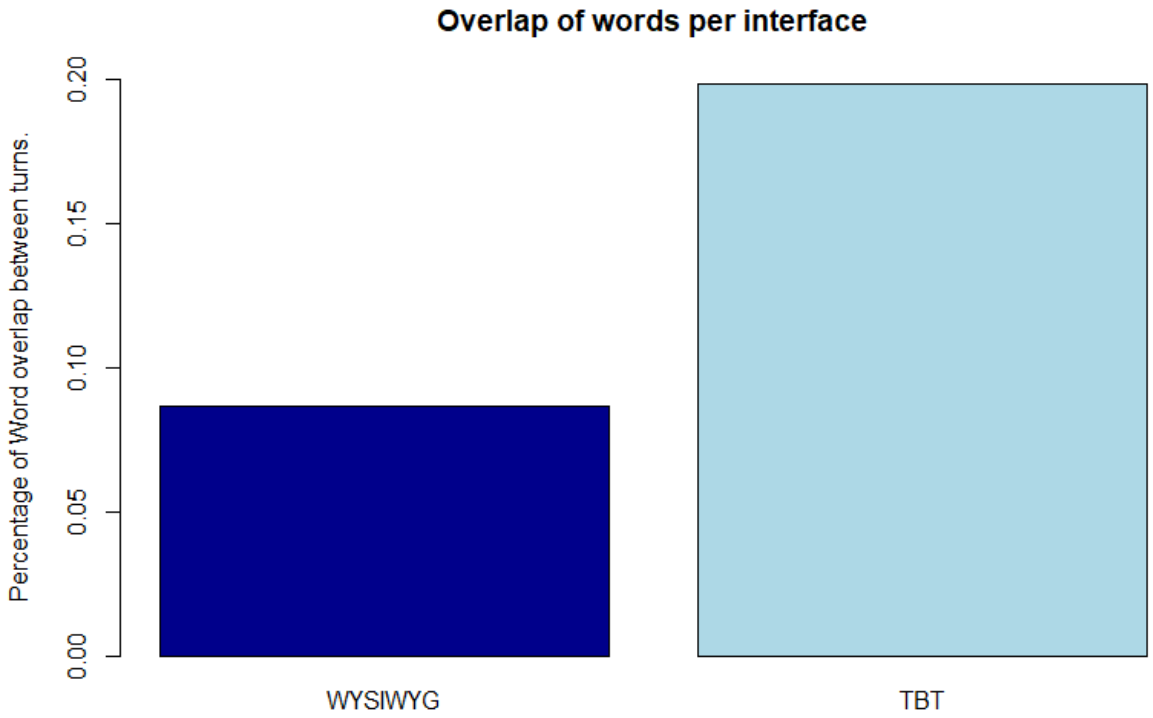

Figure 5: The percentage of overlapping words measured between turns.

As can be seen in the barplot, the TBT uses more overlapping words than the WYSIWYG interface. Having, on average, 0.2 overlapping words compared to less than 0.1 overlapping words per turn for the WYSIWYG interface.

1.4 Q4 Happiness

Our measure for happiness is the average positive score per turn. This score can indicate if an interface is frustrating to work with or not. The positive score is generated by NLTK's Vader which generates then based on the meaning of words but also the . The first step is translating the sentences to English using Google translate, which is not perfect but better than not translating. We translated the messages because Vader only works properly on English sentences. We used a separate script, translations.py, as it took a long time to run the translation script we decided to add it to our data set so we do not have to translate every time we run the other calculation script. In term of pre-processing we removed the '((NEWLINE))' tags and replaced them with spaces as they both do not have any meaning. We then took the sum of the positive scores per interface type and then calculated the average positive scores per turn by dividing the sum by the total. We excluded the data lines in this calculation. A higher average score indicates a more positive or happier interface. In this case TBT has the better score, but both interfaces do not differ too much (just 0.08).

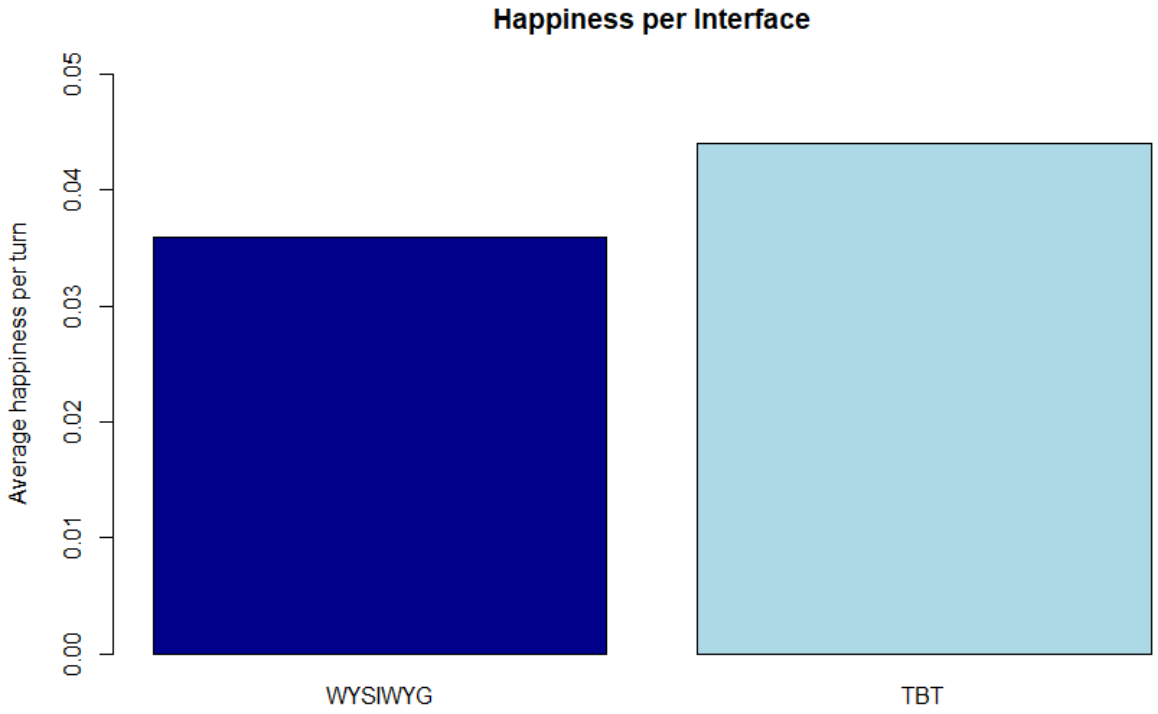

Figure 6: The average positive score per turn (higher is better).

1.5 Q5 Emotional Synchrony

Our measure for emotional synchrony is the average amount of emotional changes per turn. The same data set (with the translation added) as in Q4. We do this by comparing the negative and positive sentiment per turn and then selecting the highest one as the current main sentiment, if it differs from the previous turn then it is counted a change in sentiment. We added this to a sum and then calculated the average by dividing in by the total amount of turns. A lower score means less times the sentiment was changed. In this case WYSIWYG scored better than TBT. The difference is once again not very large (0.016). If a higher or lower score is better is of course debatable because on the one hand a lower score indicates a more consistent experience but on the other hand the lower score might also indicate that people cannot express themselves in a proper way.

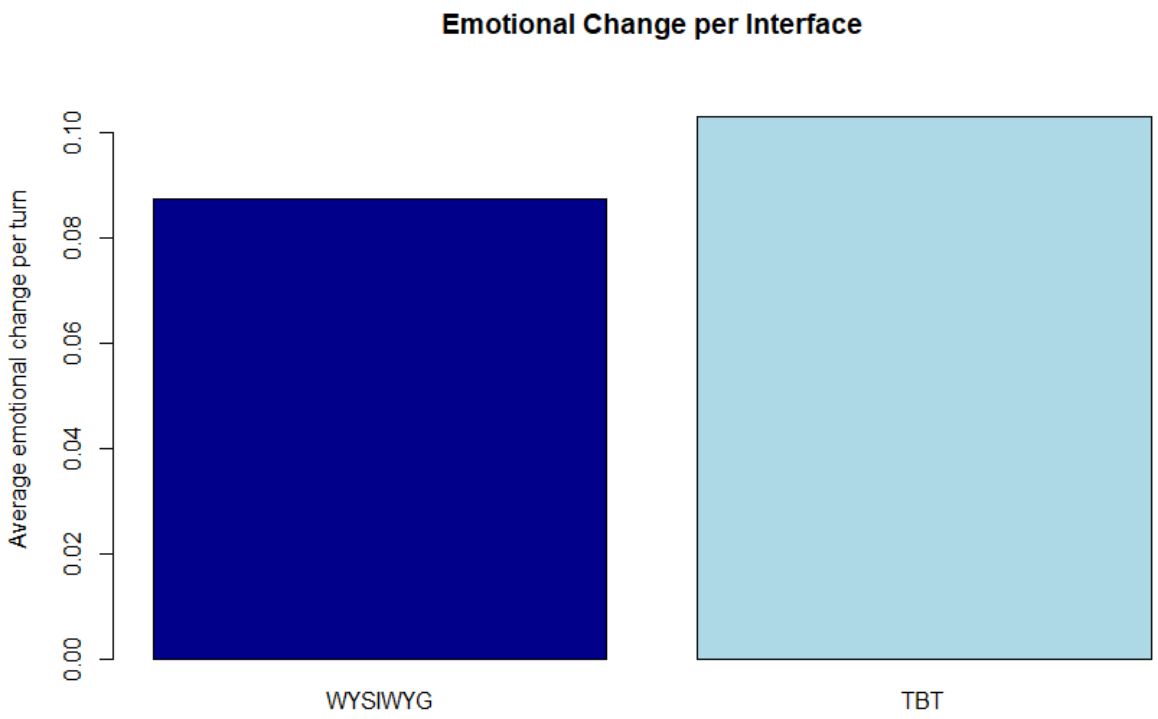

Figure 7: The average amount of times the primary emotional sentiment changes per interface (lower is better).

1.6 Q6 Data Usage

A good interface does in some cases, not require a minimal usage of data. Data usage can also be a way to express how many information a person needs in order to understand a certain instruction or gain knowledge. In this case the goal is to complete the level and in order to achieve that, you need information of the other player. Depending on how much data you send, you can determine how well the interface represents information. It also naturally shows how efficient it is in data usage. Some games end up being overtime, which would seem like a problem for this measure, because it assumes the game has been successfully ended. However, because failed games tend to already have more characters, it will negatively impact the measure and compensate for that. We used the measure average characters per turn. Because every character is encoded in the same way, meaning that they are equal units of data. Therefore we can use it as a measurement.

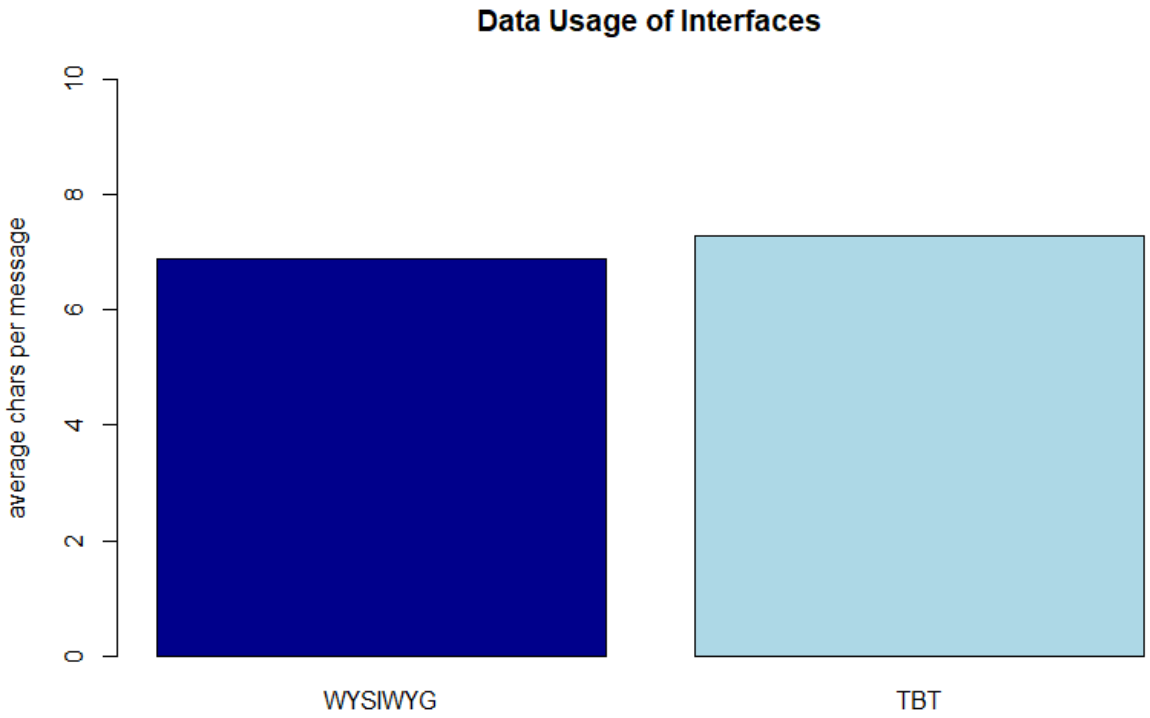

Figure 8: The average number of characters used per interface.

The barplot shows that WYSIWYG is slighty better than TBT, leading by 0.4 fewer characters. Meaning that TBT needs marginally less data to get the point accross.