Designing a Novel Method for Personalizing Recommendations to Decrease Plastic Pollution

Seung Ah Choi, Monta Vista High School, Cupertino, California

I. INTRODUCTION
Individuals of first world countries that already have streamlined waste management systems, who do not wish to contribute to the plastic pollutants that harm oceans and wildlife health, but also do not have enough belief that individual action can make a significant change, thus continuing to use single-use plastics out of convenience, need a mobile application that encourages these individuals to work towards zero-waste through small actions (ex. using a mug at Starbucks or using a reusable fork at school) by visibly quantifying all the small actions (through possibly a point system) that happen in a day regarding how much plastic was saved. The first analysis is to compare whether the call for governmental/individual action, in general, is higher in first or third world countries. After analyzing multiple Reddit threads that correspond to the topic, turns out that the count of words that call for governmental/individual action is greater in response to plastic pollution in third-world countries rather than in first-world countries and the count of words that are apathetic towards governmental/individual action is greater in response to plastic pollution in first-world countries than in third-world countries. The second analysis is to compare whether zero waste action is more prominent in first world countries or third world countries. After analyzing multiple Reddit threads that correspond to the topic, turns out that the count of supporting/positive words will be greater in response to actions for zero waste than to actions against zero waste and the count of opposing/negative words will be greater in response to actions against zero waste than to actions for zero waste.

II. ABSTRACT
According to Our World in Data, Third World countries (i.e. India) tend to have a higher share of plastic waste that is inadequately managed while First World countries (i.e. The US) have higher plastic waste generation per person. This difference in the characteristics of plastic pollution depending on the country's standing results in varying optimal recommendations for users depending on which country they live in. Through Big Text and OSOME meme analysis, I constructed a list with optimal recommendations for First World and Third World countries. Based on the list, I designed a User Interface (UI) with Google Apps Scripts that provide
personalized recommendations based on the country’s standing and user’s preferred difficulty and reassessed the code based on the six qualities of code. The User Interfaces’ purpose is to aid people who wish to help solve plastic pollution by offering a set of personalized tasks for each user and keeping their progress accountable through a point tracking system. With a significant number of users, the application could eventually contribute to solving plastic pollution.

**Big Text Data Analysis**

The first analysis is to compare whether the call for governmental/individual action, in general, is higher in first or third world countries. After analyzing multiple Reddit threads that correspond to the topic, turns out that the count of words that call for governmental/individual action is greater in response to plastic pollution in third-world countries rather than in first-world countries and the count of words that are apathetic towards governmental/individual action is greater in response to plastic pollution in first-world countries than in third-world countries.

The second analysis is to compare whether zero waste action is more prominent in first world countries or third world countries. After analyzing multiple Reddit threads that correspond to the topic, turns out that the count of supporting/positive words will be greater in response to actions for zero waste than to actions against zero waste and the count of opposing/negative words will be greater in response to actions against zero waste than to actions for zero waste.

**First Big Text Data Analysis**

**Hypothesis:** The count of words that call for governmental/individual action will be greater in response to plastic pollution in third-world countries rather than in first-world countries and the count of words that are apathetic towards governmental/individual action will be greater in response to plastic pollution in first-world countries than in third-world countries.

**Null Hypothesis:** The count of words that call for governmental/individual action and the count of words that are apathetic towards governmental/individual action will be approximately the same in response to plastic pollution in first-world countries and in third-world countries.

**Figure 1. Word Cloud for Plastic Pollution in Third-World Countries and in First-World countries**

Both word clouds were generated by a text that was a compilation of comments from a Reddit thread which was about plastic pollution in each first-world and third-world country. The body of the text was both around 3.2k words, and I chose the comments from most upvoted in order so that there was no bias involved in my selection of the comments. I chose the particular Reddit threads because it was the most upvoted/discussed topic amongst plastic pollution in...
third-world and first-world countries. At first glance, the most frequently used words in both word clouds (or the words that are the largest size) are general topics like ‘people,’ ‘India,’ ‘plastic,’ ‘water’ etc. Therefore I will focus more on the words that are related to the alternative and null hypothesis that I have generated. I believe that the word clouds support my alternative hypothesis more than the null hypothesis. The first Word Cloud that depicts people’s responses to plastic pollution in third-world countries (in this case India, as it is one of the most used words in the word cloud) mainly points to how people think that the main problem regarding plastic pollution in third-world countries is that people are uneducated as shown by the repetitive use of the words ‘ignorant’ and ‘dumb’. Moreover, as depicted by the repetitive use of the words ‘government’ and ‘education,’ people strongly advocate for governmental action in India to educate the ‘people’ (another word that shows up frequently) the importance of zero waste and reduction of plastic pollution in rivers. This also implies that people generally think that individuals themselves must also take the initiative to learn and get information about how bad plastic pollution is in their country. On the contrary, the second Word Cloud that depicts people’s responses to plastic pollution in first-world countries includes words like ‘pointless’ which is apathetic towards an individual or even governmental change. In this context, the words ‘much’ and ‘can’ also seem to support the idea that people in first-world countries believe that there is not much they can do individually or even as a country. In addition, as depicted by the frequent use of ‘global’ and ‘research,’ the people of the first-world country called for global action and more scientific studies on how to decrease the amount of plastic waste.

|                      | Plastic pollution in third-world countries | Plastic pollution in first-world countries |
|----------------------|-------------------------------------------|-------------------------------------------|
| Call for individual/governmental action (government, education, ignorant, dumb) | 39                                        | 3                                         |
| Apathetic towards governmental/individual action (pointless, irrelevant, impossible) | 2                                         | 11                                        |

This means that the alternative hypothesis, that the count of words that call for governmental/individual action will be greater in response to plastic pollution in third-world countries rather than in first-world countries and the count of words that are apathetic towards governmental/individual action will be greater in response to plastic pollution in first-world countries than in third-world countries, is supported.

**Conclusion:** Now that I have two kinds of analysis (word clouds and chi-squared), I feel that the alternative hypothesis is supported. From the chi-squared test, the p-value is <.00001, and is significant at p<.05. From the word cloud, I saw that the majority of the words that call for governmental and individual action in the word cloud for third-world countries, and I saw the majority of the words that were apathetic towards governmental and individual action in the word cloud for first-world countries. I believe that both the word cloud and the chi-squared test confirms the alternative conclusion.
Second Big Text Data Analysis

**Hypothesis:** The count of supporting/positive words will be greater in response to actions for zero waste than to actions against zero waste and the count of opposing/negative words will be greater in response to actions against zero waste than to actions for zero waste.

**Null Hypothesis:** The count of supporting/positive words and the count of opposing/negative words will be approximately the same in response to actions for zero waste and to actions against zero waste.

**Figure 2. World Cloud Comparison between the Actions for Zero Waste (Positive) and the Actions against Zero Waste (Negative)**

From the chi-squared test, the p-value is <.00001 and is significant at p<.05, which means that the alternative hypothesis from above is supported.

Osome Meme Analysis

I discussed what other topics are discussed when #plasticpollution is discussed on social media. I used an Osome tool to draw the following conclusions.

**Hashtags/Topic:** #zerowaste, #cancer, #plasticpollution, #plastic (impacts society)

**Null Hypothesis:** There will not be much of a correlation between #zerowaste or #cancer and #plasticpollution (plastic)

**Alternative Hypothesis:** There will be a strong correlation between #zerowaste or #cancer and #plasticpollution as these two topics have been discussed together frequently.

**Figure 3. Co-Occurrence of tags (networks) #zerowaste, #plasticpollution**
Figure 4. Time plot

Figure 5. The same graph as Fig 4, but the natural log version is underneath

Conclusion: From the network graphs for #zerowaste and #plasticpollution, the hashtag that showed the largest incidents of co-occurrence was #cancer. Since the larger the circle in the network graph indicates that the particular hashtag is discussed more by people alongside the given hashtags, this means that #cancer is frequently discussed when people talk about #zerowaste and #plasticpollution. This makes sense to me because health is one of the most significant discussion topics for people. Other hashtags related to #plasticpollution was #ocean since the oceans and rivers are largely impacted by plastic pollution. Environmentally friendly advocacy hashtags were also found like #ecofriendly, #waronplastic, and #gogreen. Hashtags related to zero waste were mostly advocating for various methods to go zero waste like #recycling, #handmade, #vegan, and #circularconomy.

One hashtag that I thought was particularly significant was #plasticfreejuly. When I did research to find the meaning of this hashtag apparently the national zero waste day takes place every year on July 3rd. Based on this information, I created time plot graphs from June 23rd, 2019 to July 7th, 2019, the time period that I thought these hashtags would be discussed most frequently amongst netizens. Turns out the usage of zero waste did significantly increase (around 50%) on July 3rd in particular. The number of users for the hashtag #cancer also increased right before July 3rd significantly, which I believe was a method for internet users to bring about awareness for #plasticpollution and the importance of zero waste before the national zero waste day. The usage of #plastic suddenly increased on June 26th, and while I do not know exactly why, I believe there must have been some law passed about plastic, or some trend regarding recycling or remaking plastic was trending on social media. It may have also been influenced by some viral picture related to plastic, however, it is not possible to know if plastic was used in a negative or positive connotation. Regardless, my guess is that it was used positively because the use of #plasticpollution simultaneously decreased during that time, showing the opposite of co-occurrence. In general, the #cancer was used more than #plastic because people usually care a lot about their health, and these two hashtags were used more on average than #plasticpollution and #zerowaste because it is less specific to a certain topic and can be used on a general basis.
In-Sheet Calculation Analysis

I calculated the average values for the share of plastic waste that is inadequately managed around the world which turned out to be 34.43% on average and the average value for plastic waste generation per person which turned out to be 0.18 ton per year.

**Objective:** Use Google Apps Script in Google Sheets to calculate the average, minimum, maximum, median, and standard deviation value of the share of plastic waste that is inadequately managed in each country and plastic waste generation per person by country. I have calculated the average, minimum, median, maximum, and standard deviation of the two different data sets that I have found each with 186 data points. I have included the calculation formulas that I have used in Table 3.

### Conclusion

The formulas I used in Google Apps Script made it significantly easier for me to calculate the average, minimum, median, maximum, and standard deviation of 186 data points because it allows selecting a group or range of the data sets that I want to explore. If I had to input each data point in a calculator and used the known standard mathematical formulas, the process would have taken much longer.

### Algorithm Overview

The program’s purpose is to help people decrease plastic pollution by recommending specific tasks like avoiding using plastic (if their country has high plastic waste generation) or pushing for better waste management systems (if their country has high plastic waste mismanagement). My video explains the significance of optimizing recommendations based on their country and describes the following process.

Once the user inputs their country, the code determines the country’s standing. Then, the user is asked if they want recommendations. If they answer NO, the app terminates with a thank-you message. If they answer YES, the program asks the user for the preferred difficulty of recommendations. Based on country standing and preferred difficulty,
personalized recommendations are called. Once tasks are completed, a specific number of points based on task difficulty is awarded. Finally, the user receives an encouraging message along with the total points they gained.

My peers who test-ran my rough application thought it lacked user interactivity because my code mainly incorporated constants. Upon receiving this feedback, I independently fixed the problem. First, I included method getCurrentCell() to assign variable userAnswer to whichever cell the user selects. This way, the stored value will alter real-time based on the user’s choice. However, I soon realized when I run the following methods the currently active cell would change, storing a different value for userAnswer. Thus, I decided to copy the user’s initially selected value to a constant range, so the variable stays constant throughout running the code. Second, whenever I rerun the copyOverforPoints() method, the new set of recommendations would paste over the old set because variable targetRow (indicating the row to paste in the recommendations in the ‘Point Tracker’ Spreadsheet) is constant. I wanted the variable to change based on the user’s decisions, but couldn’t find a way to save values from a previous run since data resets every run. Therefore, I added pseudocode, which has guidelines on how to change targetRow necessarily. Although directly changing the code may be inefficient and time-consuming, I concluded it’s the best way to preserve user interactivity while accumulating data.

[Refer to the code below for the circle] My partner and I worked independently and later combined our methods in proper sequence so the code works together. The following methods can run individually. determineCountryStanding() categorizes the searched country into First-World, Third-World, or Neither. This method has if-else if blocks that contain if-statements that use inequalities based on percentValue (country’s percent mismanaged plastic waste) which is searched by my partner’s getCountryInfoByName() method. Then, askUserAbtRecommendations() asks if the user wants recommendations to help solve plastic pollution. If-else blocks of this method ensure different outputs based on user response making the code polymorphic. The response NO terminates the program with a thank you message and YES asks for the user's preferred difficulty (HARD, MEDIUM, or EASY). If answered YES, activateGetRec() can correctly run since it can store the user's preferred difficulty (determined by askUserAbtRecommendation()) to variable userAnswer. activateGetRec() calls getRecommendations(userAnswer) which outputs personalized recommendations. This method contains if-statements and boolean values which ensures only the recommendations that match both countryStanding (output of determineCountryStanding()) and userAnswer is copied over from ‘Recommendations Data Set’ Spreadsheet. For-loops of this method make sure to check every row of the data set by iteratively increasing foundRow and prevent overlaps of pasted recommendations by incrementally increasing targetRow.

[Refer to the code below for the rectangle] My method pointSystem() implements my abstraction. My
abstraction is a 2D array. pointSystem() uses logic through if-else if blocks which are divided based on the difficulty levels of the recommendations. This is indicated by the first column of the array (column C in the pointTrackerSpreadsheet). The second column (column D in the pointTrackerSpreadsheet) determines if the task is completed by checking if the user inputted an O. If both criteria are met, point values are outputted accordingly: ten points for HARD levels, five points for MEDIUM, and one point for EASY. pointSystem() uses math when foundRow value incrementally increases by one in the for-loop which causes the program to move onto the next row of the matrix. Math also shows when it calculates the total points in sourceRange using the ‘SUM’ formula Google Apps Script provides. This abstraction makes code more efficient by helping manage the complexity of using two different variables at the same time. Using a 2D array allows me to store both difficulty level and user input. Thus, I can print out corresponding point values dynamically using a for loop, by simply changing the index numbers, making the final total calculation easier without creating many different variables.
The code for the program is high quality as it encompasses the six qualities of code and uses abstractions like the 2D array that maximizes the efficiency of the program. Also, the code incorporates for-loops and boolean values that help minimize the steps needed to get an output.

Decidable: In all cases of an input, there will be a corresponding output. In the method `getCountryInfoByName()`, the user can input any value (even if it is not a valid country name) and still get some type of output when the code is run. This way, the user has the option to choose to explore information about different countries. The following method `determineCountryStanding()` uses the `getRecommendations()` method. In the method `askUserAbtRecommendations()`, the user can also input any value other than YES or NO when they are asked if they want recommendations and still get an output. In the method `getRecommendations()`, the user can also choose any other cell in the spreadsheets other than HARD, MEDIUM, or EASY and the code will still get some output. The code is provable on paper. The code is complete because we can make a flowchart on paper as shown in Figure 6.

**Figure 6. Flow Chart**

---

**Six Qualities of Code**

The code for the program is high quality as it encompasses the six qualities of code and uses abstractions like the 2D array that maximizes the efficiency of the program. Also, the code incorporates for-loops and boolean values that help minimize the steps needed to get an output.

Decidable: In all cases of an input, there will be a corresponding output. In the method `getCountryInfoByName()`, the user can input any value (even if it is not a valid country name) and still get some type of output when the code is run. This way, the user has the option to choose to explore information about different countries. The following method `determineCountryStanding()` uses the percentValue found in the previous method. In the method `askUserAbtRecommendations()`, the user can also input any value other than YES or NO when they are asked if they want recommendations and still get an output. In the method `getRecommendations()`, the user can also choose any other cell in the spreadsheets other than HARD, MEDIUM, or EASY and the code will still get some output. The code is provable on paper. The code is complete because we can make a flowchart on paper as shown in Figure 6.

**Figure 6. Flow Chart**

---

**Six Qualities of Code**

The code for the program is high quality as it encompasses the six qualities of code and uses abstractions like the 2D array that maximizes the efficiency of the program. Also, the code incorporates for-loops and boolean values that help minimize the steps needed to get an output.

Decidable: In all cases of an input, there will be a corresponding output. In the method `getCountryInfoByName()`, the user can input any value (even if it is not a valid country name) and still get some type of output when the code is run. This way, the user has the option to choose to explore information about different countries. The following method `determineCountryStanding()` uses the percentValue found in the previous method. In the method `askUserAbtRecommendations()`, the user can also input any value other than YES or NO when they are asked if they want recommendations and still get an output. In the method `getRecommendations()`, the user can also choose any other cell in the spreadsheets other than HARD, MEDIUM, or EASY and the code will still get some output. The code is provable on paper. The code is complete because we can make a flowchart on paper as shown in Figure 6.

**Figure 6. Flow Chart**

---

**Six Qualities of Code**

The code for the program is high quality as it encompasses the six qualities of code and uses abstractions like the 2D array that maximizes the efficiency of the program. Also, the code incorporates for-loops and boolean values that help minimize the steps needed to get an output.

Decidable: In all cases of an input, there will be a corresponding output. In the method `getCountryInfoByName()`, the user can input any value (even if it is not a valid country name) and still get some type of output when the code is run. This way, the user has the option to choose to explore information about different countries. The following method `determineCountryStanding()` uses the percentValue found in the previous method. In the method `askUserAbtRecommendations()`, the user can also input any value other than YES or NO when they are asked if they want recommendations and still get an output. In the method `getRecommendations()`, the user can also choose any other cell in the spreadsheets other than HARD, MEDIUM, or EASY and the code will still get some output. The code is provable on paper. The code is complete because we can make a flowchart on paper as shown in Figure 6.

**Figure 6. Flow Chart**

---

**Six Qualities of Code**

The code for the program is high quality as it encompasses the six qualities of code and uses abstractions like the 2D array that maximizes the efficiency of the program. Also, the code incorporates for-loops and boolean values that help minimize the steps needed to get an output.

Decidable: In all cases of an input, there will be a corresponding output. In the method `getCountryInfoByName()`, the user can input any value (even if it is not a valid country name) and still get some type of output when the code is run. This way, the user has the option to choose to explore information about different countries. The following method `determineCountryStanding()` uses the percentValue found in the previous method. In the method `askUserAbtRecommendations()`, the user can also input any value other than YES or NO when they are asked if they want recommendations and still get an output. In the method `getRecommendations()`, the user can also choose any other cell in the spreadsheets other than HARD, MEDIUM, or EASY and the code will still get some output. The code is provable on paper. The code is complete because we can make a flowchart on paper as shown in Figure 6.
Sound
The code/program is logically entailed and provable. As shown in the truth tables, when we input true values, true values come out. For getCountryInfoByN
ame() and determineCountryStanding, when we input a country then the program will determine the country’s standing with proper reasoning. For askUserAbtRecommendations(), the inputs YES or NO will not give an error message and either terminate the code or move on to the next section of the code. For getRecommendations() & copyOverforPoints(), when we input the cells with HARD, MEDIUM, and EASY, the corresponding set of recommendations will show. Finally for pointSystem(), a certain point value shows if the user says that they have completed the task.

Correct
There seem to be no errors in our code when there is a correct input, the code will give information about the country, a set of recommendations, or the correct amount of total points (awarded for tasks that are accomplished) as we designed the algorithm.

Efficient
In our code, we optimized for efficiency. Our code is easy to follow and not cluttered with a lot of words. In each of the methods (except for copyOverforPoints()) uses an if-else block which ensures that code is organized into categories. For-loops and boolean values are used when choosing which recommendations meet the criteria which allow us to
check for every single data point in the ‘recommendation data set’ array efficiently. I also used a 2D array in my last method to make code more efficient by helping manage the complexity of using two different variables at the same time. Using a 2D array allows me to store both difficulty level and user input. Thus, I can print out corresponding point values dynamically using a for loop, by simply changing the index numbers, making the final total calculation easier without creating many different variables.

Polymorphic
In our code, a variety of inputs can cause a method to behave differently which makes the code polymorphic, as shown in the truth tables. For example, depending on the user’s country choice (affecting the country standing) and preferred difficulty level, the set of recommendations will be different. Also, depending on the preferred difficulty level of the recommendations and whether or not the user has accomplished those tasks, the number of total points awarded would be different.

III. CONCLUSION & SUGGESTION FOR FUTURE RESEARCH
For the current application, I have arbitrarily set the inequality values that determine if percent mismanaged plastic waste is high or low when determining if the country is first or third-world. For future modifications, I believe some background research would be needed to correctly determine what value and above percent mismanaged plastic should be considered third-world instead of estimating the value. I believe it would be worth adding more suggested recommendations especially for the countries that are in the average range because right now the code will randomly select recommendations made both for the first world countries and the third world countries. Also to statistically analyze exactly how efficient and motivating the application is for people to contribute to decreasing plastic pollution, I would like to do run an experiment in which users from both first-world and third-world countries use the application for a certain amount of time and rate their motivation levels to see if the application actually contributed to a significant change. The testing can be done simply through a google form and then we can use chi-squared analysis to see how significant the results were. The testing can also be done by calculating the change in total points earned to analyze in a more detailed fashion.

In terms of modifying the code, I suggest adding reset mechanisms on the code so that the spreadsheet automatically resets after running the entire application once. The tricky part though is that I will have to figure out a way to save the points values from the previous run into the next run so that the point values accumulate over time.

REFERENCES
[1] “Plastic Waste Generation per Person.” Our World in Data, ourworldindata.org/grapher/plastic-waste-per-capita. Accessed 18 April 2020.

[2] “r/UrbanHell - People Offering Prayers at River Yamuna, India, which is Frothing from Industrial Waste.” Reddit, https://www.reddit.com/r/UrbanHell/comments/9zmaj0/people_offering_prayers_at_river_yamuna_india/. Accessed 29 March 2020.

[3] “r/news - It's Raining Plastic: Microscopic Fibers Fall from the Sky in Rocky Mountains | US News.” Reddit, https://www.reddit.com/r/news/comments/cpr
[4] “Share of Plastic Waste That Is Inadequately Managed.” *Our World in Data*, ourworldindata.org/grapher/inadequately-managed-plastic. Accessed 18 April 2020.