COMP3008 Project 2

Quantitative Usability Evaluation

CGJB

Clara Au

Gilbert Lam

Justin Ward

Brandon Ward

Table of Contents

Part 1 : Descriptive Statistics

I. Exploration of Text21 and Image21

Text21

While exploring the Text21 password scheme demonstration through creating a password with the automated generator and testing the successful attempts of password entries it was apparent that the scheme itself provided a secure strong set of strings and numbers that would not be easy to break through a brute force attack. However, the memorability of these passwords is not the best for users, hindering the usability of this scheme for realistic use. The combination of letters and numbers used in the password scheme provides a large password space despite only taking up five characters. Although this short combination and large password space might be an advantage towards a strong password scheme, it may be difficult to remember in the long term. The passwords may be too complex and the cognitive chunking technique humans naturally use for memorability would not be as useful in this case since the password generated is a random order of letters and strings and doesn't follow a specific formatted order for all the passwords generated with this scheme. As a result, it would be difficult for an individual to remember their passwords for three different accounts unless the individual had very high memorability skills.

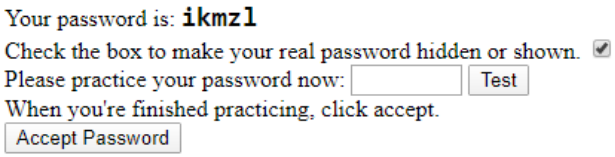


Fig 1.1 Text21 password scheme features a random generated order of lower case letters and numbers of a total length of 5 characters.

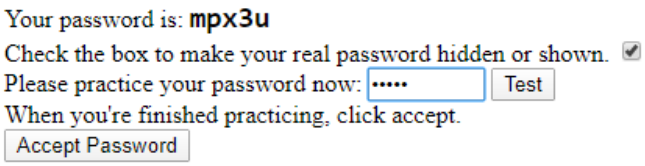


Fig 1.2 The interface prompts you to practice inputting the password before it is accepted to increase memorability.

Image21

While exploring the Image21 password scheme demonstration and being given a set of certain cells from a grid overlaying an image, it was found to be a lot more difficult to remember than the Text21 password scheme. The password scheme is presented on an 8 cell by 6 cell grid with a total of 48 cells. Out of the 48 cells, the password scheme is 5 random selected cells. Although the password space is very large it would be extremely difficult to remember for an individual to memorize given the size of the grid and how many cells highlighted. In addition, the highlighted cells of the image do not select anything identifiable in the making it more difficult to memorize as there is no significance or anything to reference. Another disadvantage is accessibility for the visually impaired as the visually impaired would have a near to impossible method of memorizing the password scheme and inputting it.

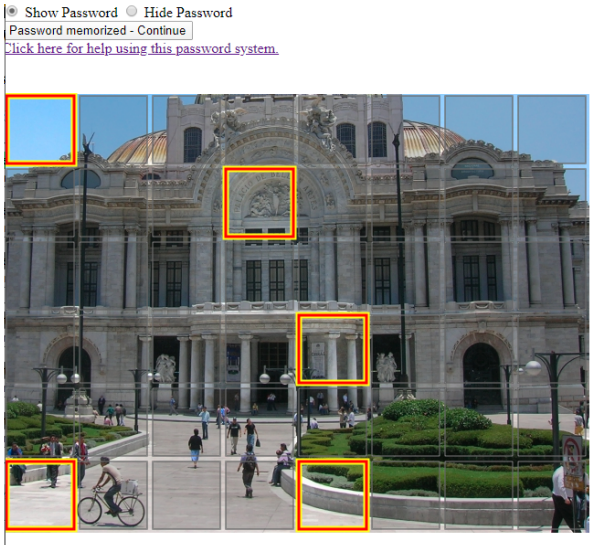


Fig 2.The Image21 random generated password scheme features a random image and a grid overlay with highlighted cells being the password.

II. Documentation of Log Data Processing Software

The log data parser was created to analyze the logs from Text21 and Image21. The main task of the parser was to clean the data of any inconsistencies or anomalies, but it is also used to produce some basic statistical calculation.Calculations include, number of attempts, number of successful attempts, number of failed attempts, average time for successful attempt and average time for failed attempts. The aforementioned calculations are computed for each user and then exported to the output.csv

The parser works by first sorting all entries by date to ensure the log file is in chronological order. Then the parser iterates through each csv file creating mappings between userIds and a list of their entries. Finally, the program iterates through the entries for each user, determining the amount of attempts, as well as computing the time for each attempt.The parser program was created using python, and can be found in the project archive. Along with the parser you will find a README with run instructions as well as the source code and the pseudocode for the program. Lastly, a copy of the output.csv with previously parsed data can also be found in the archive.

III. Comparing Usability of Text21 and Image21

The statistics below were calculated using R, the R script is entitled “stats.R” and is located in the project archive for your reference. The output.csv from the parser was imported to R as a dataframe to produce all the statistics and graphs below.

A. Text21 Statistics

Statistic	Number of Logins
Total Login Mean	16.611
Successful Login Mean	14.055
Failed Login Mean	2.555

Total Login Median	16.000
Successful Login Median	15.000
Failed Login Median	1.000
Total Login Standard Deviation	4.900
Successful Login Standard Deviation	3.438
Failed Login Standard Deviation	3.329

Fig. 1. Mean,Median,Standard Deviation of Number of Logins per user (total,successful and unsuccessful)

Statistic	Time (in seconds)
Successful Login Time Mean	9.954
Failed Login Time Mean	6.017
Successful Login Time Median	9.066
Failed Login Time Median	5.500
Successful Login Time Standard Deviation	4.244
Failed Login Time Standard Deviation	6.896

Fig. 2. Mean,Median,Standard Deviation of Login Time per user (successful and unsuccessful)

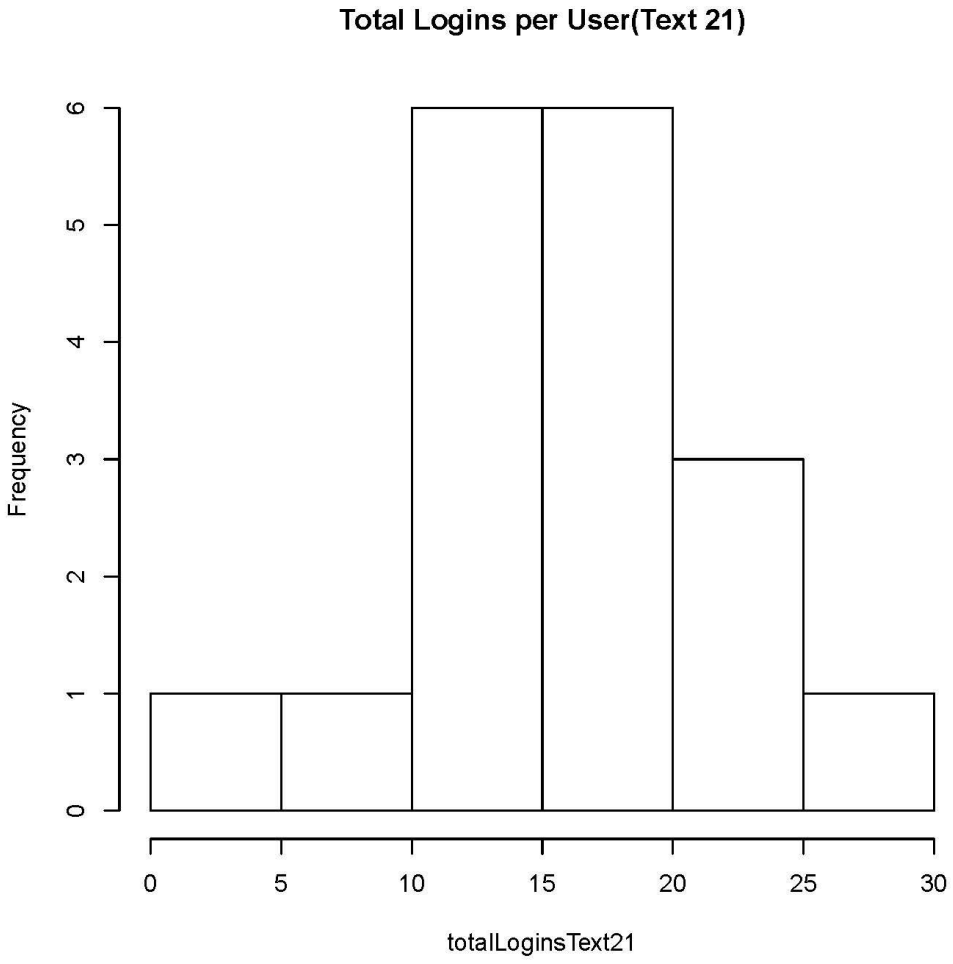


Fig. 3. Number of Total Logins per User (Text 21)

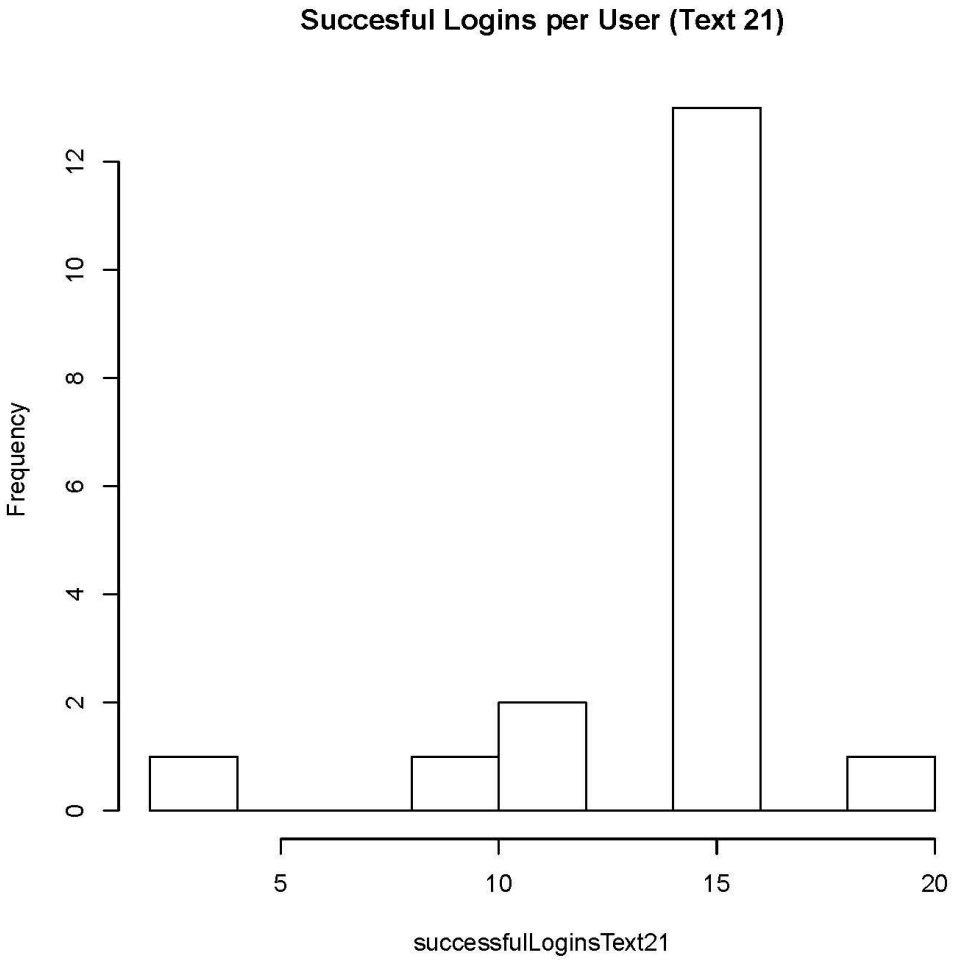


Fig. 4. Number of Successful Logins per User (Text 21)

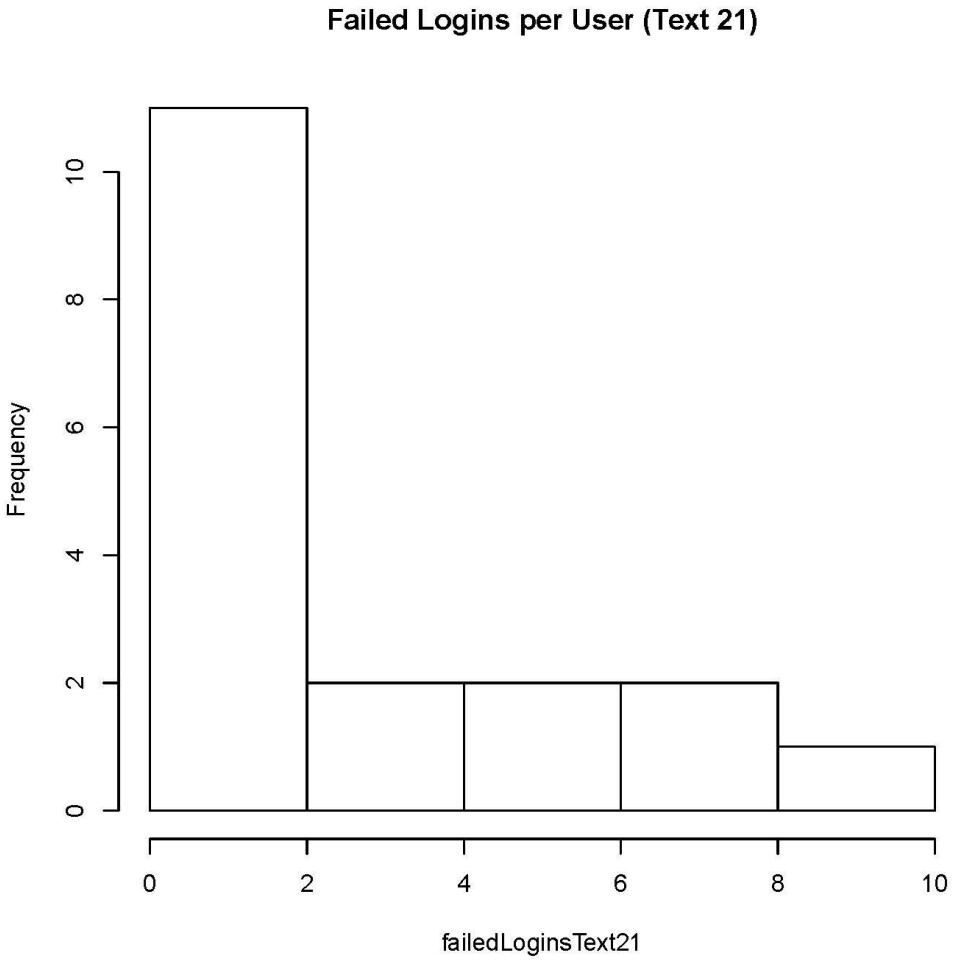


Fig.5. Number of Failed (Unsuccessful) Logins per User (Text 21)

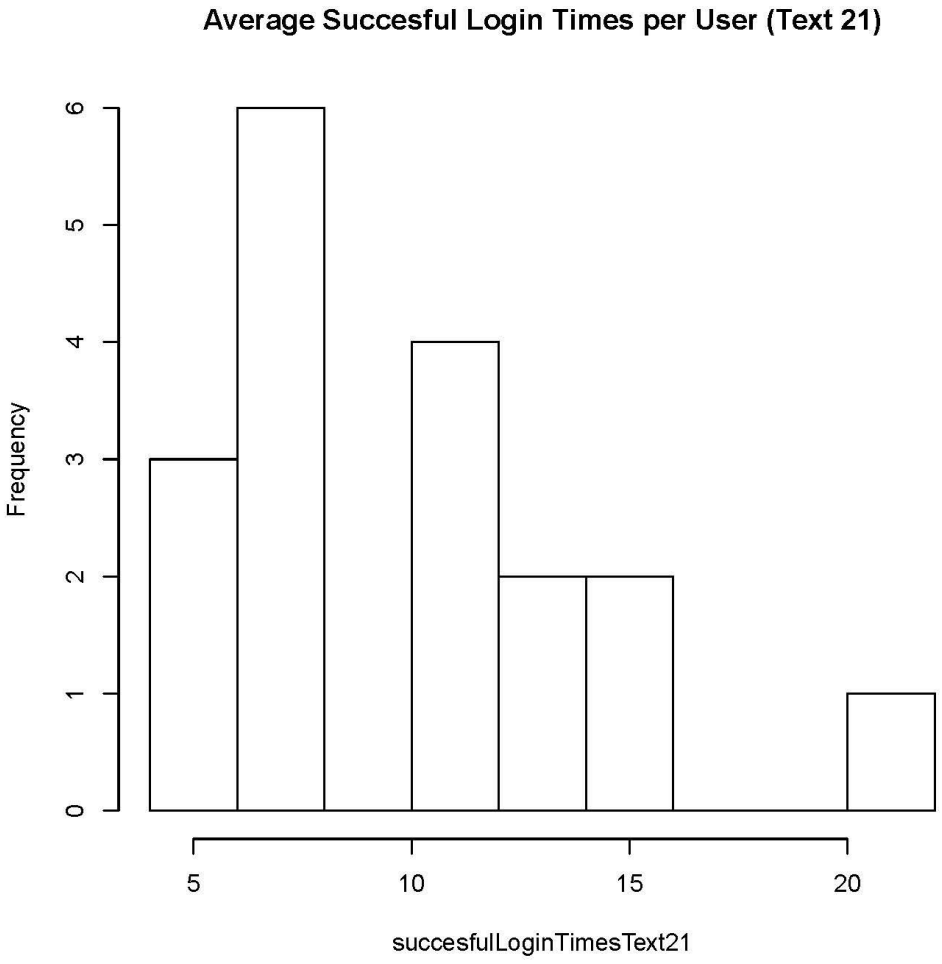


Fig.6. Average Successful Login Times per User (Text 21)

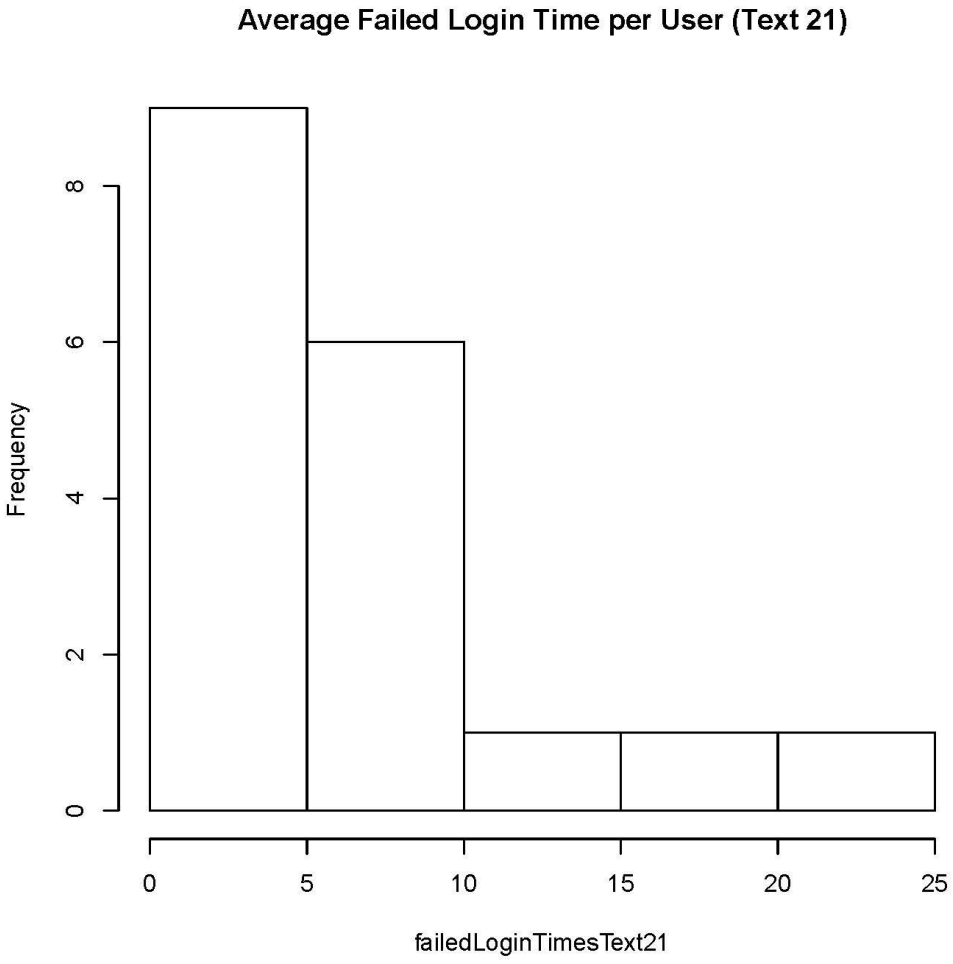


Fig.7. Average Failed (Unsuccessful) Login Times per User (Text 21)

B. Image21 Statistics

Statistic	Number of Logins
Total Login Mean	19.333
Successful Login Mean	14.867
Failed Login Mean	4.466
Total Login Median	18.000
Successful Login Median	15.000
Failed Login Median	3.000
Total Login Standard Deviation	5.287
Successful Login Standard Deviation	1.355
Failed Login Standard Deviation	4.437

Fig. 8. Mean,Median,Standard Deviation of Number of Logins per user (total,successful and unsuccessful)

Statistic	Time (in seconds)
Successful Login Time Mean	18.482
Failed Login Time Mean	22.895
Successful Login Time Median	19.562
Failed Login Time Median	17.333
Successful Login Time Standard Deviation	7.536
Failed Login Time Standard Deviation	14.608

Fig. 9. Mean,Median,Standard Deviation of Login Time per user (successful and unsuccessful)

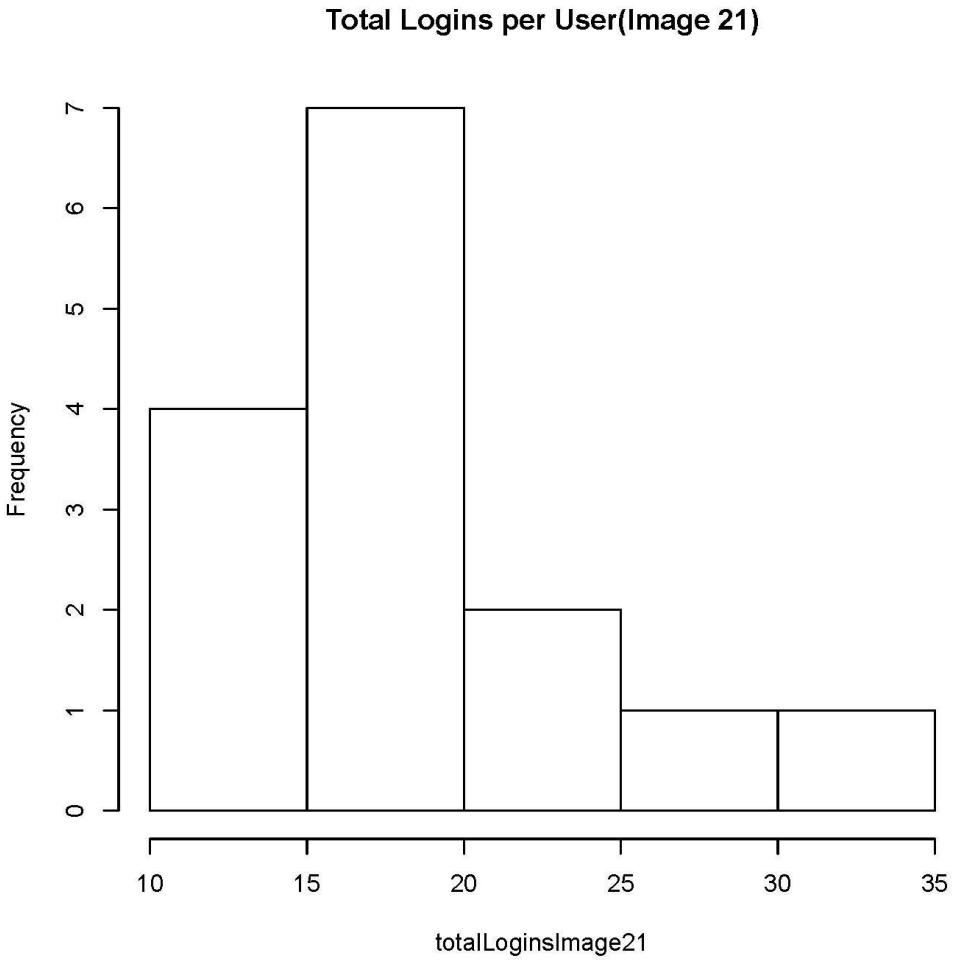


Fig.10. Number of Total Logins per User (Image 21)

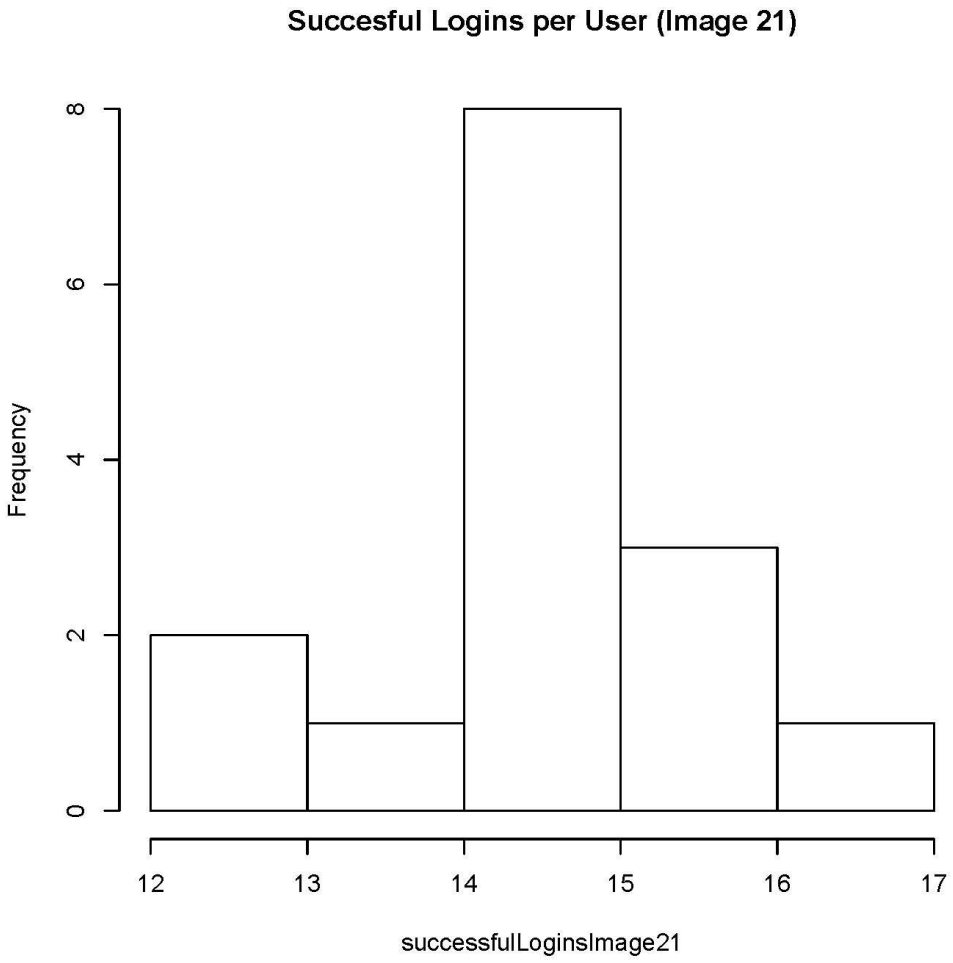


Fig.11. Number of Successful Logins per User (Image 21)

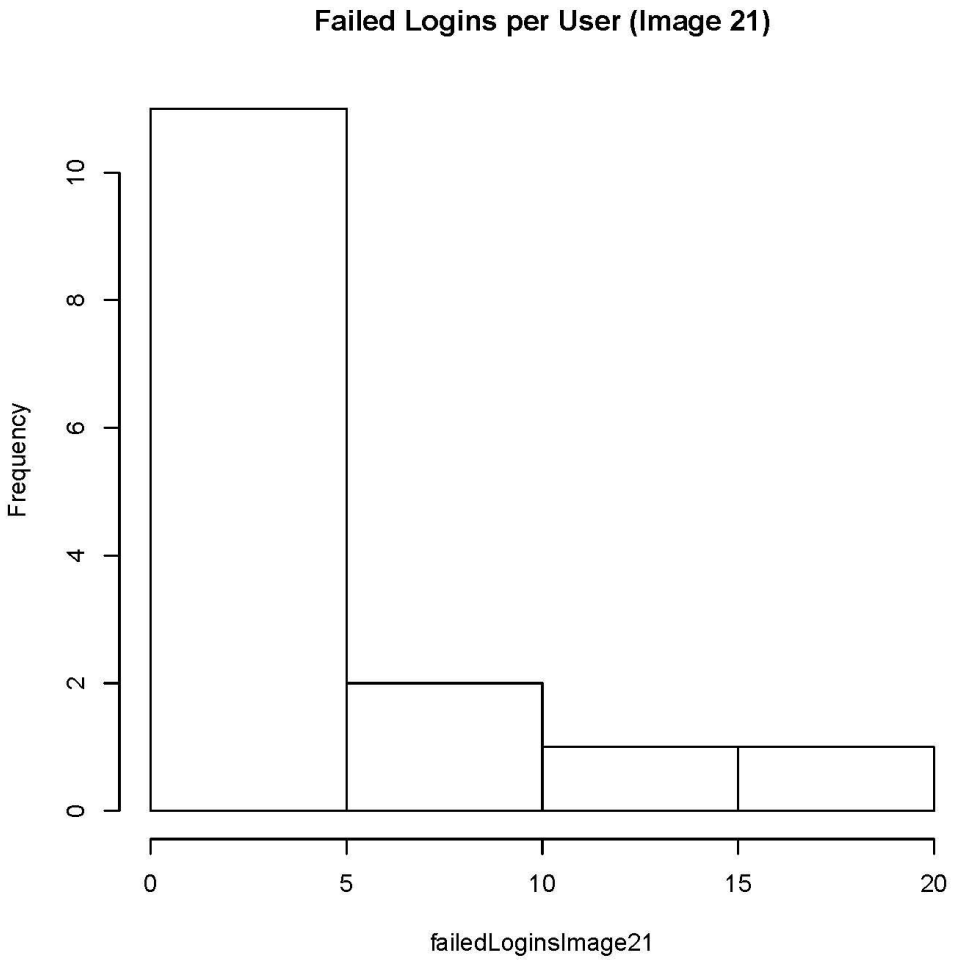


Fig.12. Number of Failed (Unsuccessful) Logins per User (Image 21)

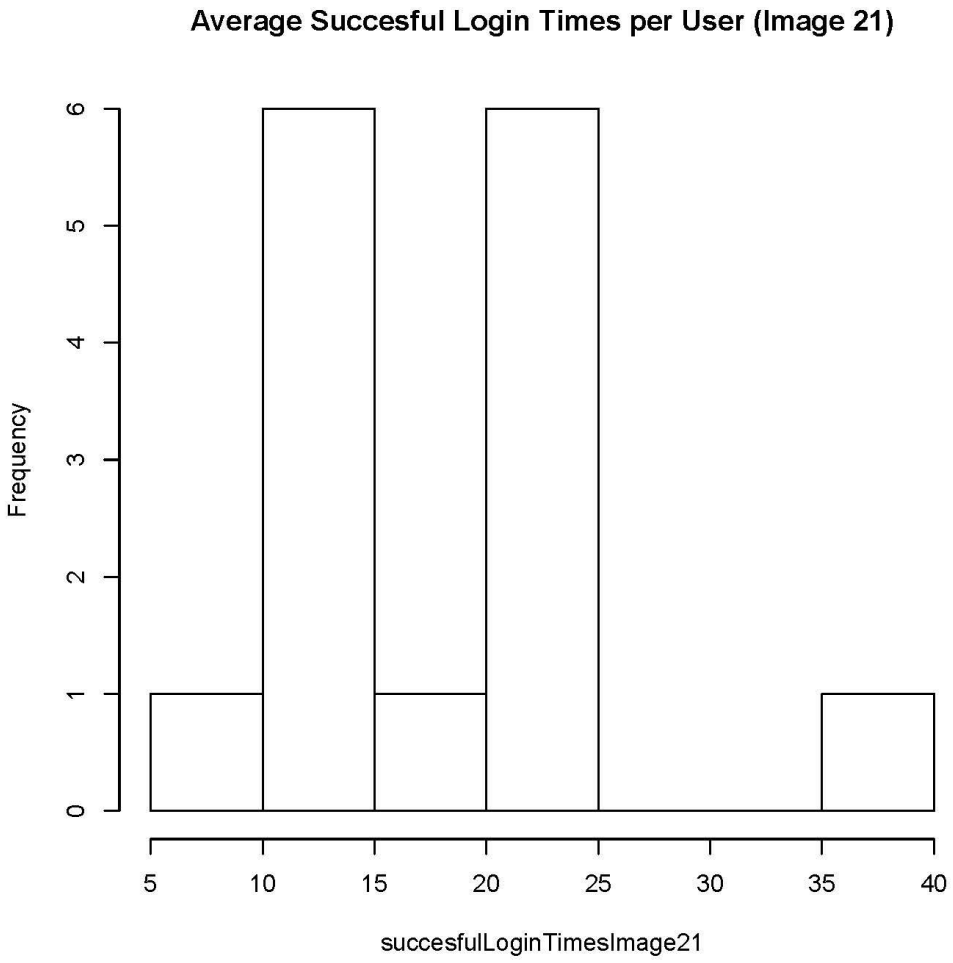


Fig.13. Average Successful Login Times per User (Image 21)

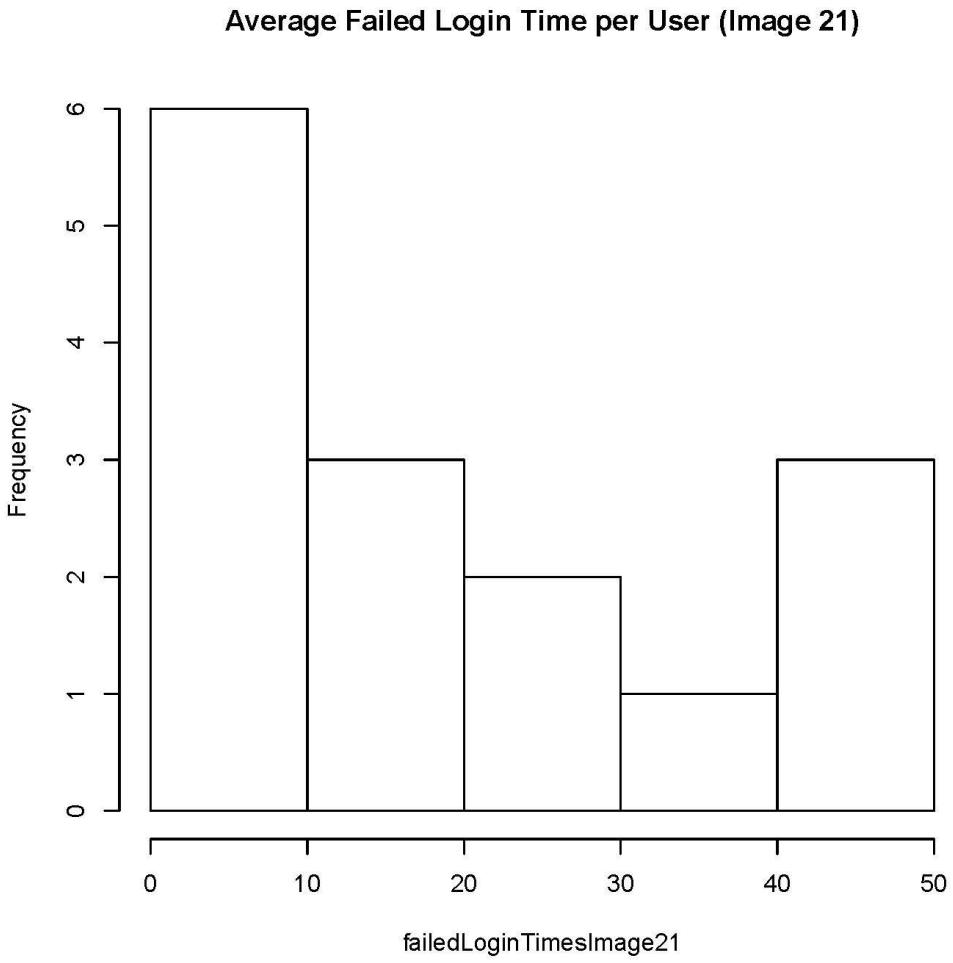


Fig.14. Average Failed (Unsuccessful) Login Times per User (Image 21)

C. Combined Statistics

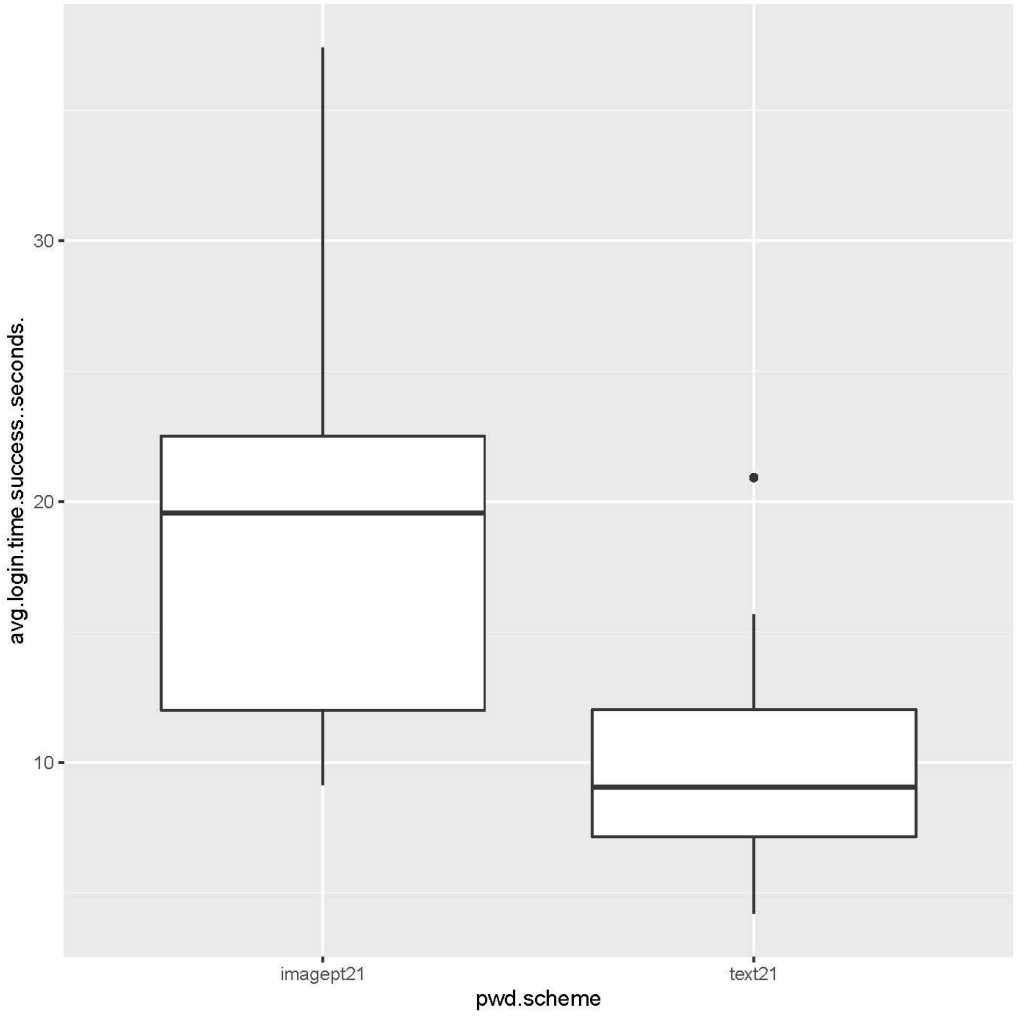


Fig.15. Boxplot Comparing relation between password scheme and average successful login time

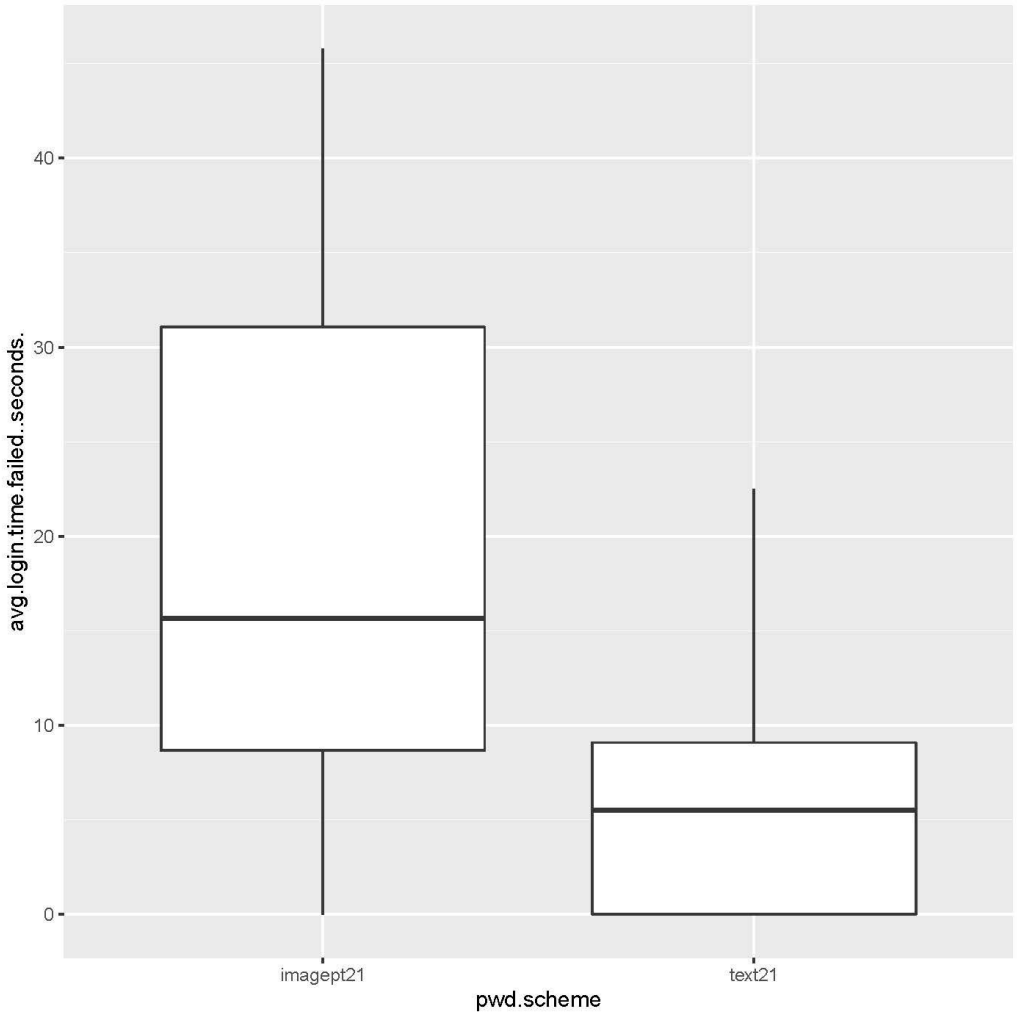


Fig.16. Boxplot Comparing relation between password scheme and average failed (unsuccessful) login time

In comparison between the statistics of Text21 and Image21 and upon further analysis, Text21 proves to be a better password scheme overall. The statistics of failed login attempts between Image21 (M = 4.466 , MD = 3.000, SD = 4.437) and Text21 (M = 2.555 , MD = 1.000, SD = 3.329) concludes that Image21 had a greater amount of failed logins. The difference between the two medians, Text21 (MD = 1.000) and Image21 (MD = 3.000), indicates that users of Image21 were more prone to failed login attempts compared to the Text21 password scheme, thus implying that users of Image21 have a greater difficulty in successfully entering the image based password. This greater difficulty of successful login attempts in Image21 indicates that the

image based password scheme ranks low on learnability. In conjunction, the mean of failed and successful login times of Text21(FLM = 6.017, SLM = 9.954,) were significantly lower than Image21(FLM = 22.895 , SLM = 18.482). The shorter login times of Text21 imply that the text based password scheme ranks higher on memorability as users spent less time attempting to recall the correct password in comparison to the Image21 password scheme. We can interpret this data as Text21 having more advantages in usability compared to Image21 and assume that images and components of an image with no significance are harder to memorize as opposed to characters and numbers. Although Text21 generated passwords have no significance either it still proves to be easier to remember as opposed to remembering random located cells over a large grid image. In order to increase the memorability of something a natural cognitive process is rehearsal of the subject. In regards to Text21 and Image21 password schemes, Text21 can be rehearsed easily because of the familiarity of characters and numbers but Image21 cannot be rehearsed easily because of the unfamiliarity of the image presented. Therefore the statistical interpretations of both password scheme tests prove Text21 to be the more usable password scheme.

Part 2 : Design, Implementation, Statistical Inference

I. Design Rationale

Uno is a simple card game with card values ranging from 1 to 9, and 4 colors of red, green, yellow and blue. A typical uno deck has 108 unique cards including certain "special cards" and duplicates. For the purpose of our "Uno" style password scheme these cards will not be included. Leaving us with a total of 36 unique cards, the password scheme will have users select 5 cards as their password. Each card "slot" will have a choice from the full deck of 36 cards, ie. 36 to the power of 5.

$$36^5 = 60,466,176$$

From the above calculation one can see there would be 60,466,176 total password possibilities. This can be converted to bits by taking the log of the total number of passwords.

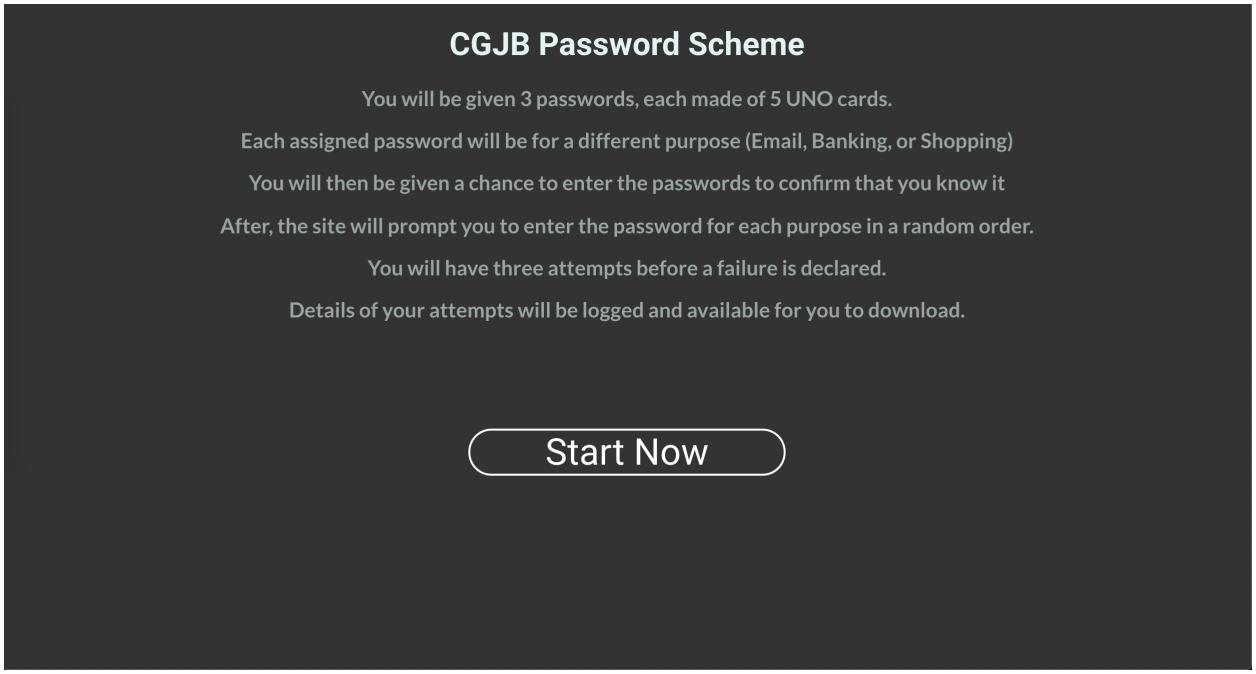
$$\log_2(36^5) = 25.84$$

Therefore, one can conclude that the password scheme that has been derived is a 25.84 bit password scheme which exceeds the standard of a 21 bit password scheme.

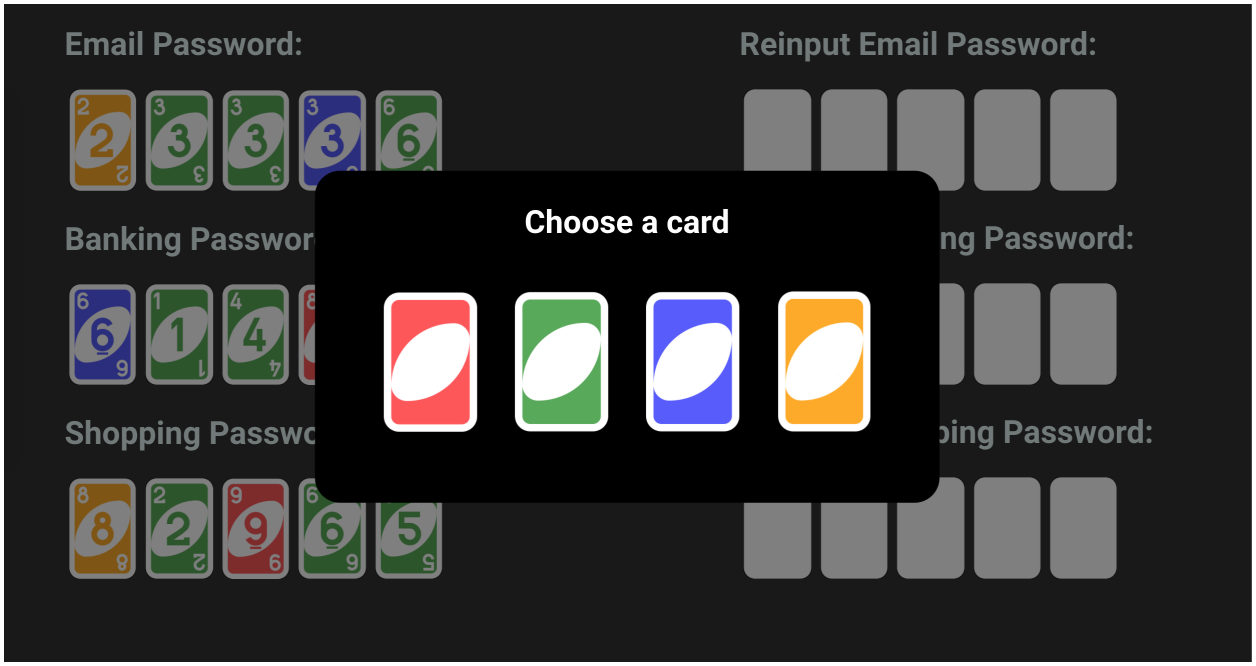
The “Uno” style scheme was chosen with hopes of improving memorability of passwords over a standard text based password entry scheme. Due to the simplicity there is an increase in the learnability of the scheme, making this type of password scheme ideal for children and the elderly. The memorization of these passwords is dependent on being able to remember the two key aspects of the cards, their colour and their value. When entering the password, the user is asked to enter each card. This is done through promoting the user with 5 blank cards and having them click on them one at a time. Upon clicking on a card a user will first be shown the 4 options of uno card colours, red, blue, green, and yellow. A user will then click on which colour they’re looking for and the application will display the array of values of that colour of cards. Selecting the right value will snap the chosen card into the blank space and the user can continue entering the remaining cards.

The bright colours aid in both separating additional values while simultaneously improving the memorability in the segregation of cards. A user simply needs to remember the value of each colour in proper order. Without capitalizations, special characters, and variable lengths, the passwords should be simpler and thus, more memorable.

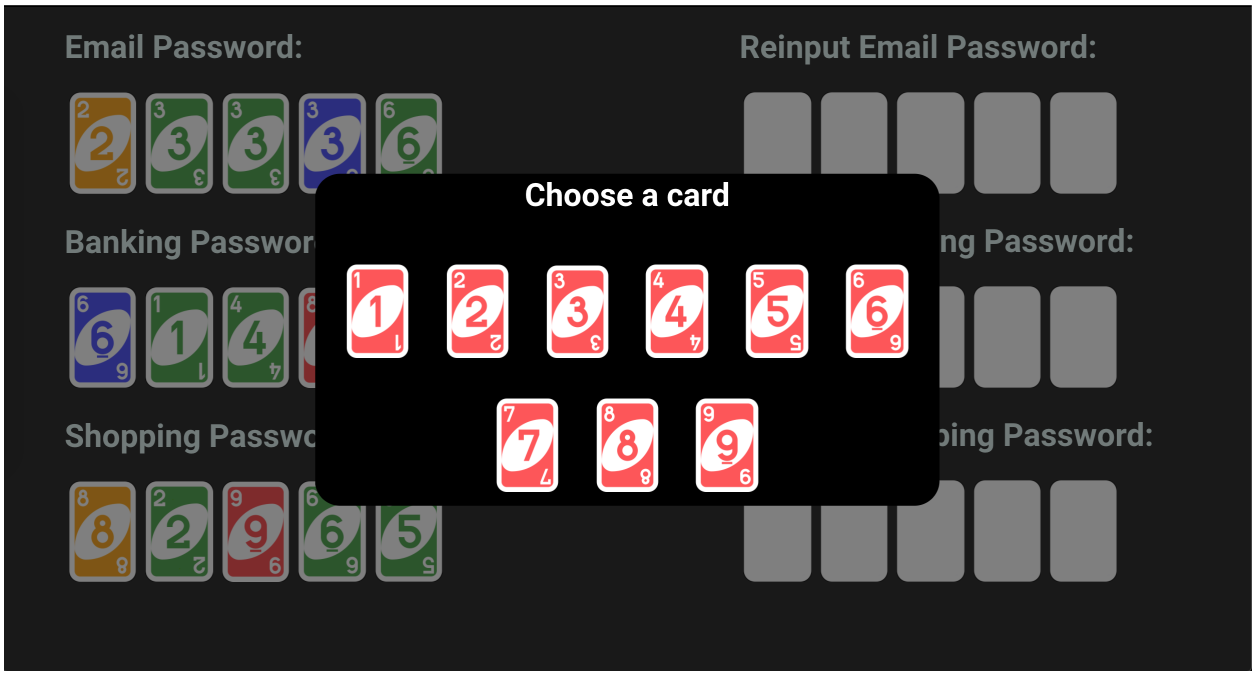
II. Screenshots of Program



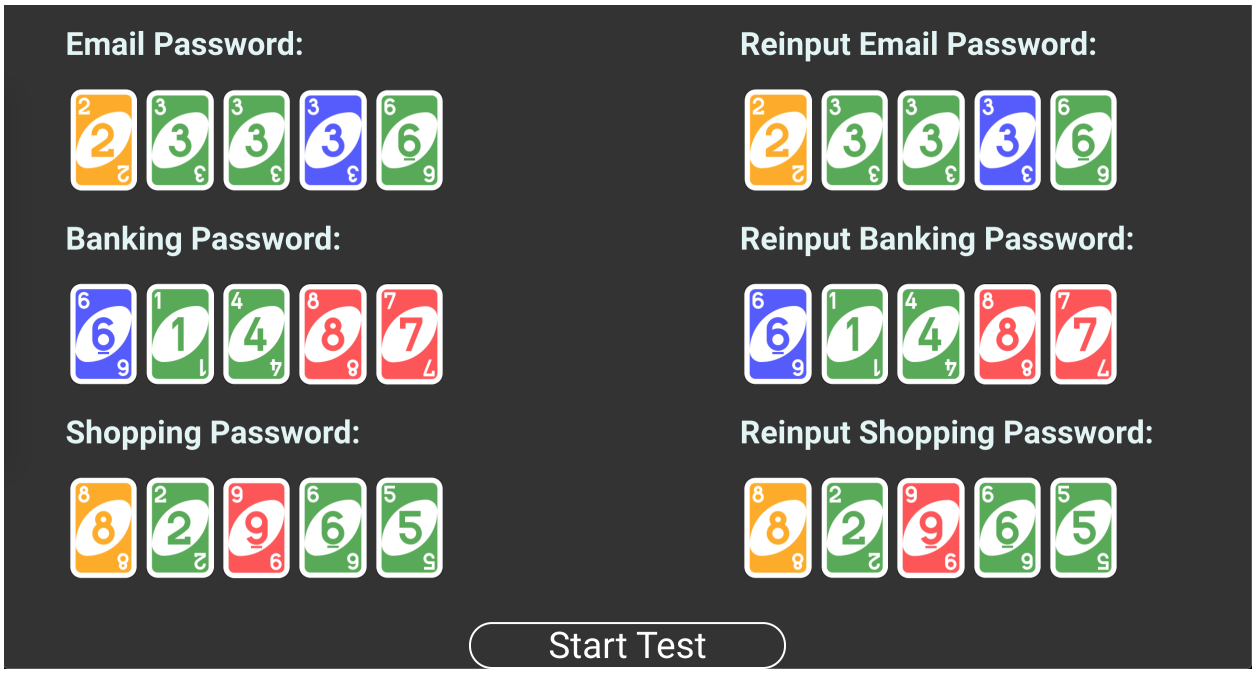
The application landing page is a brief description of the purpose of the program and what events will be logged. The 'Start Now' button initializes the password scheme program.



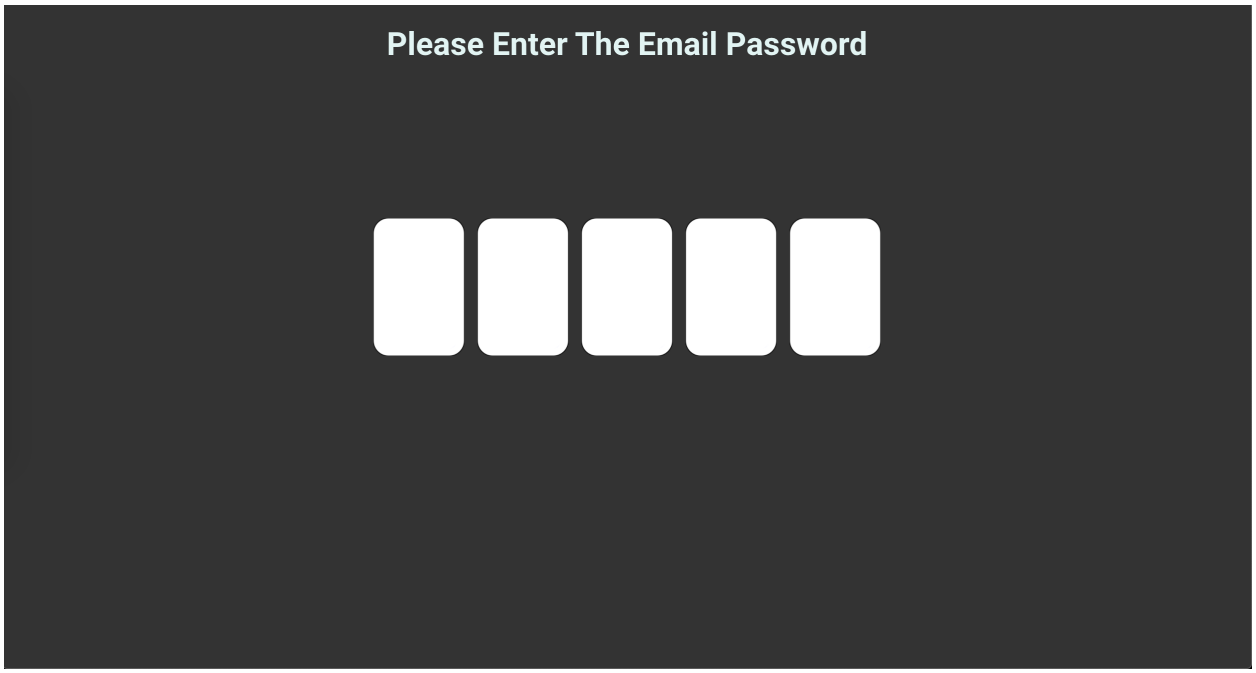
The user can then begin entering the random passwords they've been assigned by selecting the corresponding blank space on the right half of the screen. Once clicked, a dialog appears prompting the user to select which colour card they're looking for.



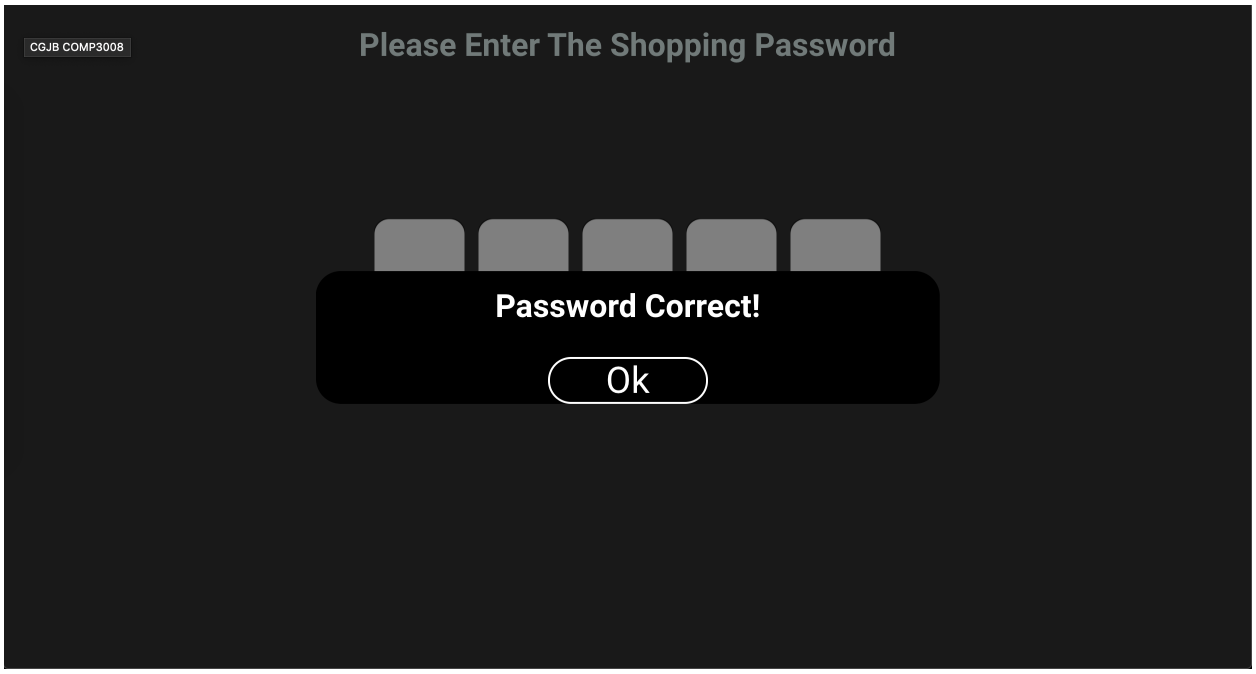
Once they've selected the colour, an array of the possible values is displayed for the user to choose from.



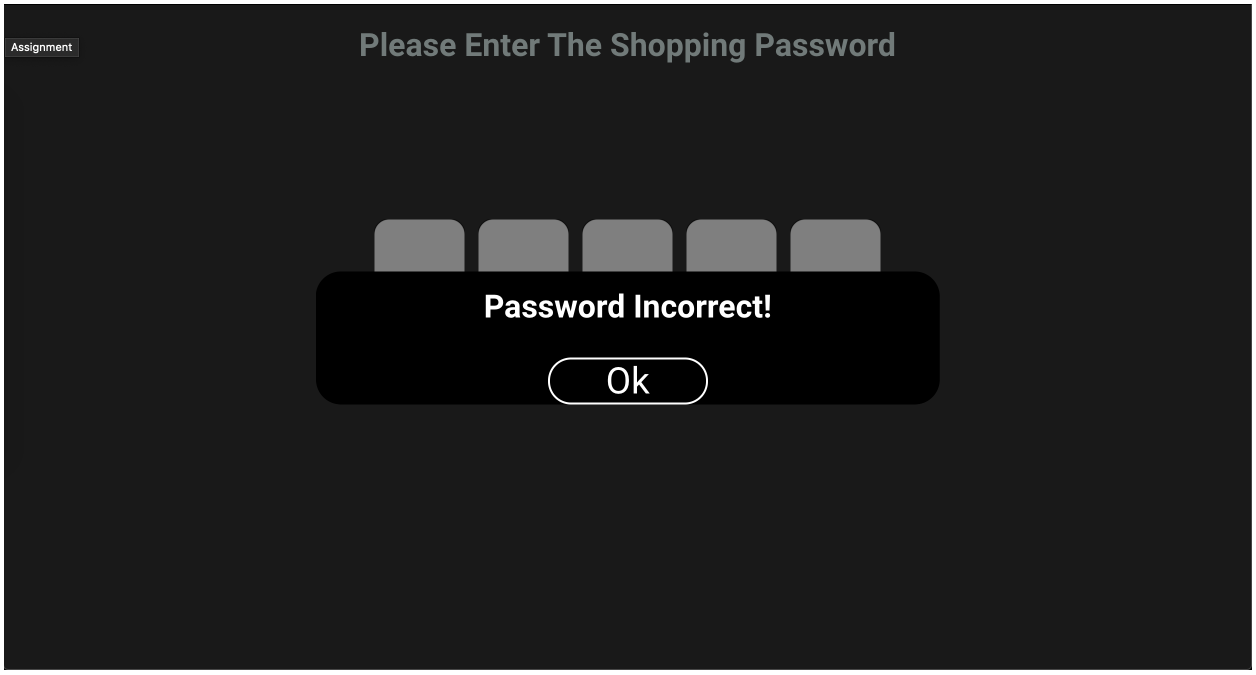
The user completes each of the 3 assigned passwords before beginning the actual tests. Once they've entered every password correctly, a 'Start Test' button appears to allow the user to begin testing the usability and memorability of their assigned password.



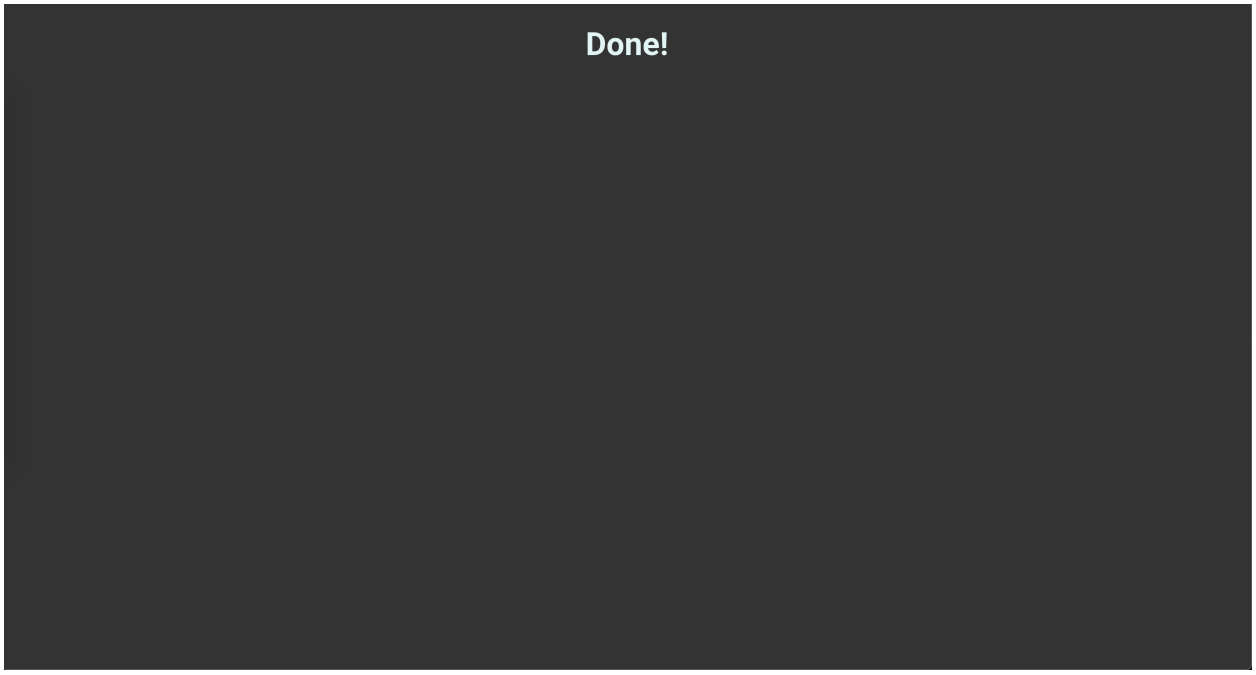
First, the user must enter their 'Email' password. The card selection process is the same as described before.



On a successful entry, the application informs the user of their success and prompts them with the next password.



On a failed attempt, the application lets the user know through a similar dialog.



Once completed, the application lets the user know that they're finished through a simple page with the word 'Done!'

III. Questionnaire

The questionnaire for user testing can be found at the following link :
https://hotsoft.carleton.ca/comp3008limesurvey/index.php/681262?lang=en
Attached in the zip is a PDF copy of the likert scale user testing questions.
https://drive.google.com/file/d/1KjD8tXhUV6jAysEN_6RfdYqqo15cd978/view?usp=sharing

IV. Results Interpretation

UnoStyle Statistics

The statistics below were calculated using R, the R script is entitled “statsUno.R” and is located in the project archive for your reference. The outputUno.csv from the parser was imported to R as a dataframe to produce all the statistics and graphs below.

Statistic	Number of Logins
Total Login Mean	6.272727
Successful Login Mean	1.545455
Failed Login Mean	4.727273
Total Login Median	7
Successful Login Median	1
Failed Login Median	6
Total Login Standard Deviation	2.412091
Successful Login Standard Deviation	1.29334
Failed Login Standard Deviation	3.608072

Fig. 17. Mean,Median,Standard Deviation of Number of Logins per user (total,successful and unsuccessful)

Statistic	Time (in seconds)

Successful Login Time Mean	19
Failed Login Time Mean	18.44444
Successful Login Time Median	21
Failed Login Time Median	19.91667
Successful Login Time Standard Deviation	5.410161
Failed Login Time Standard Deviation	7.659935

Fig. 18. Mean,Median,Standard Deviation of Login Time per user (successful and unsuccessful)

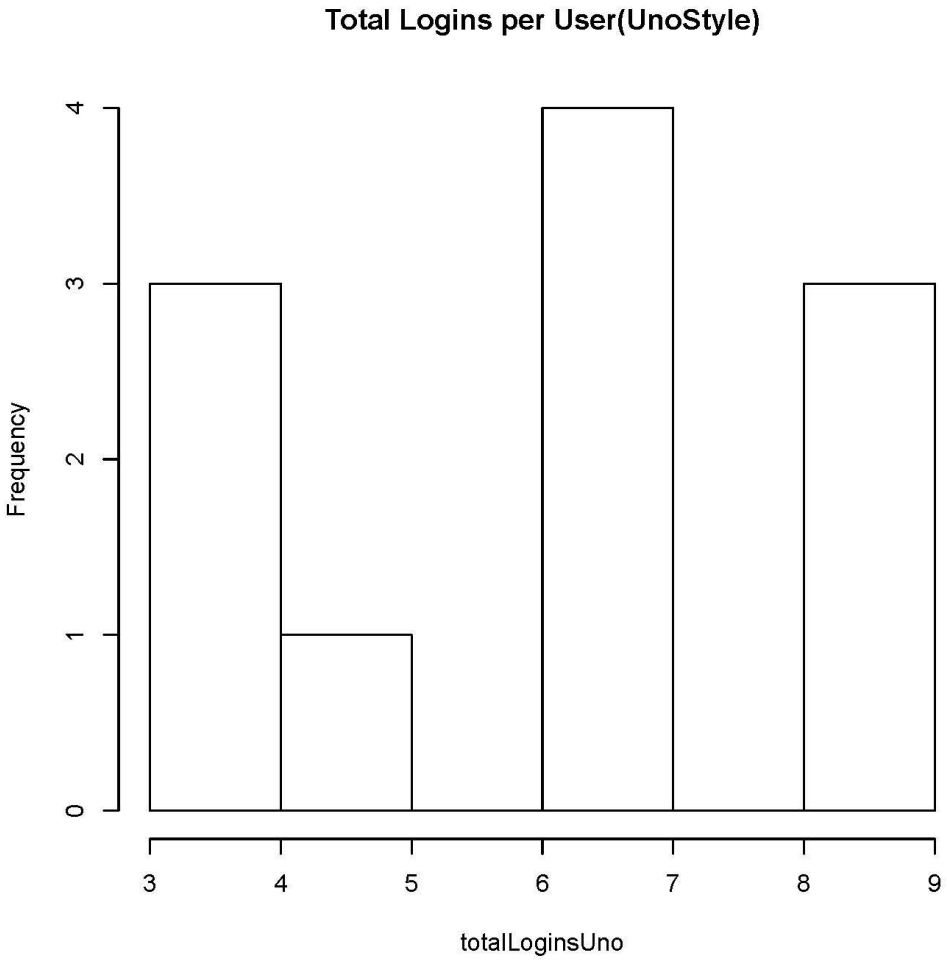


Fig. 19. Number of Total Logins per User (UnoStyle)

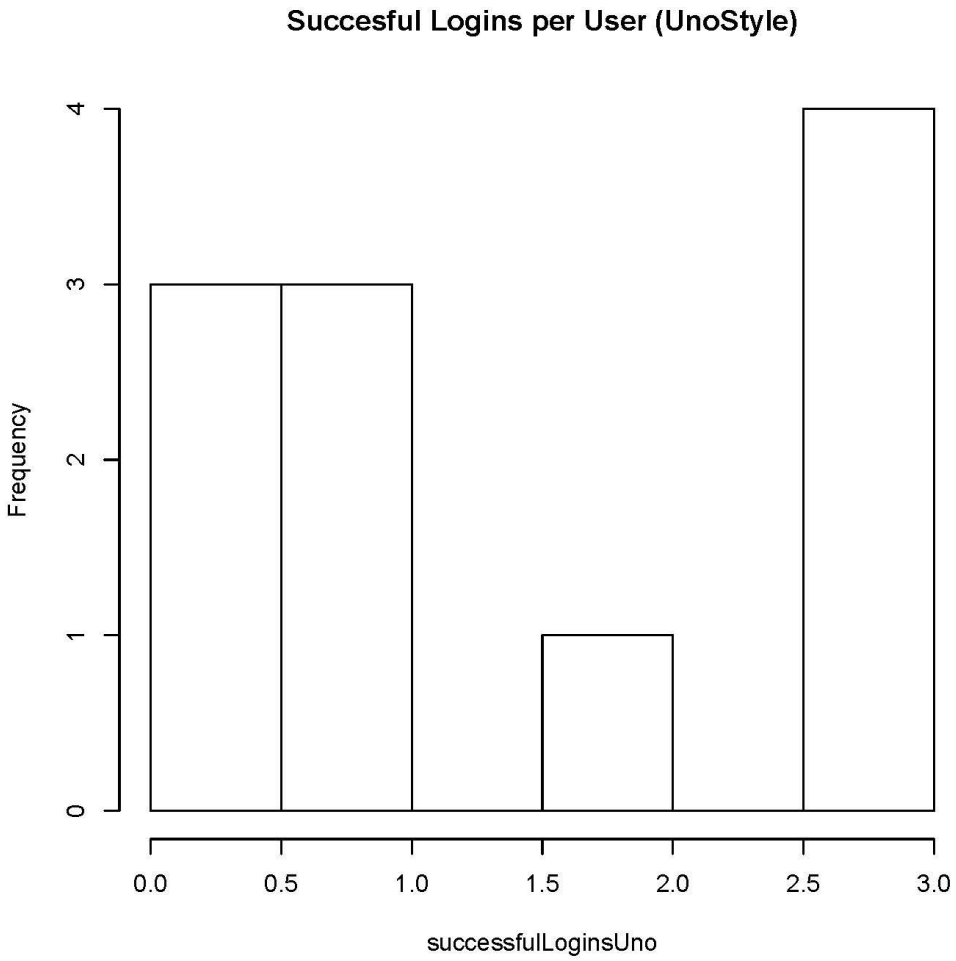


Fig. 20. Number of Successful Logins per User (UnoStyle)

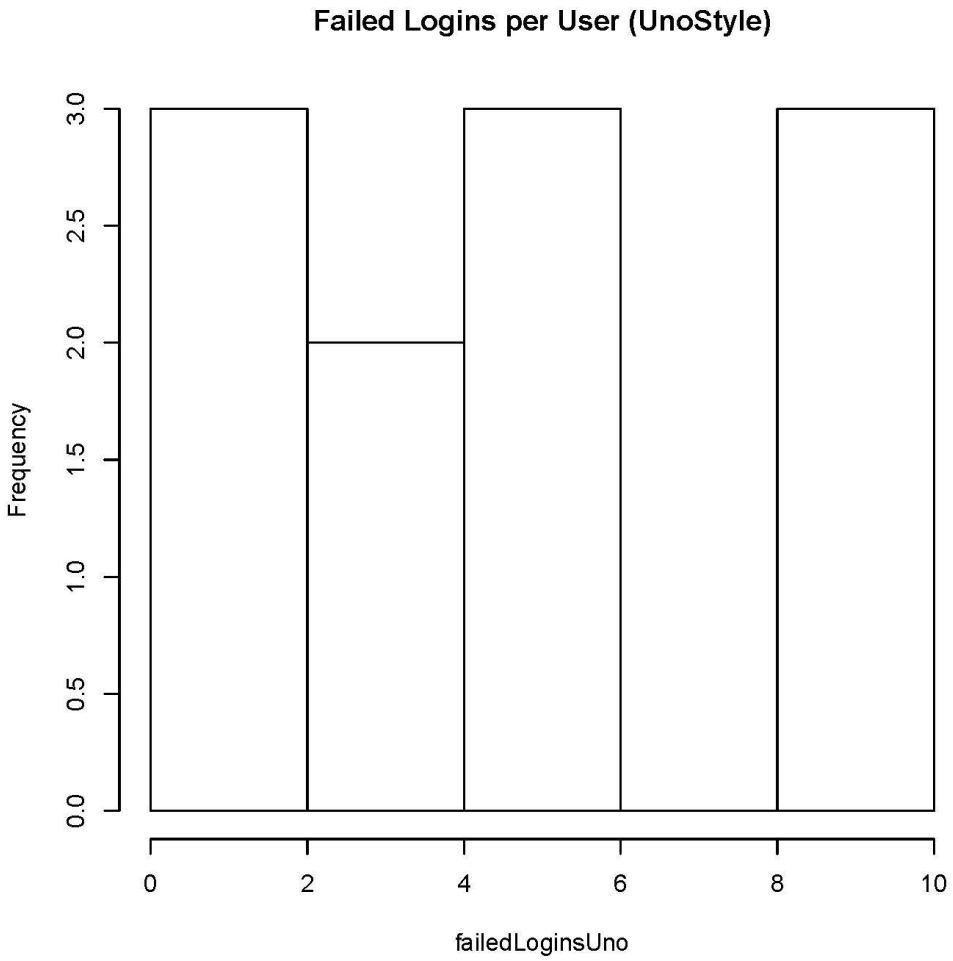


Fig 21. Number of Failed (Unsuccessful) Logins per User (UnoStyle)

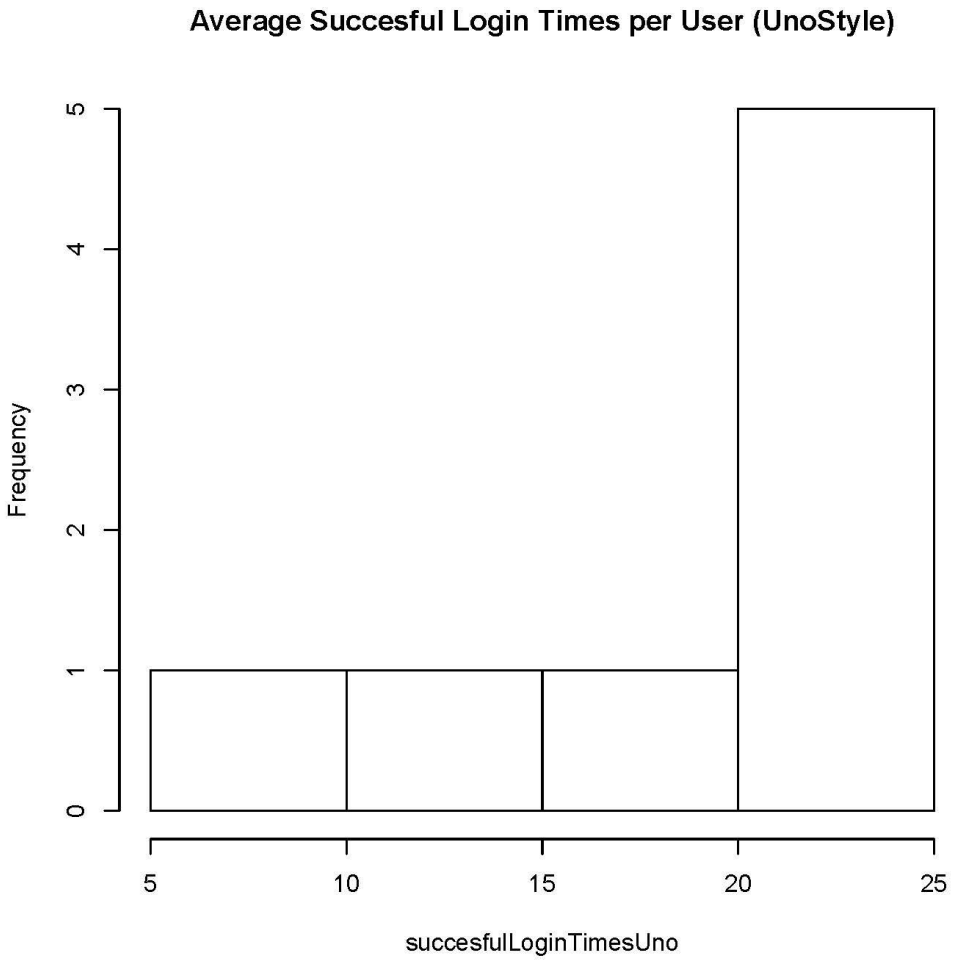


Fig. 22. Average Successful Login Times per User (UnoStyle)

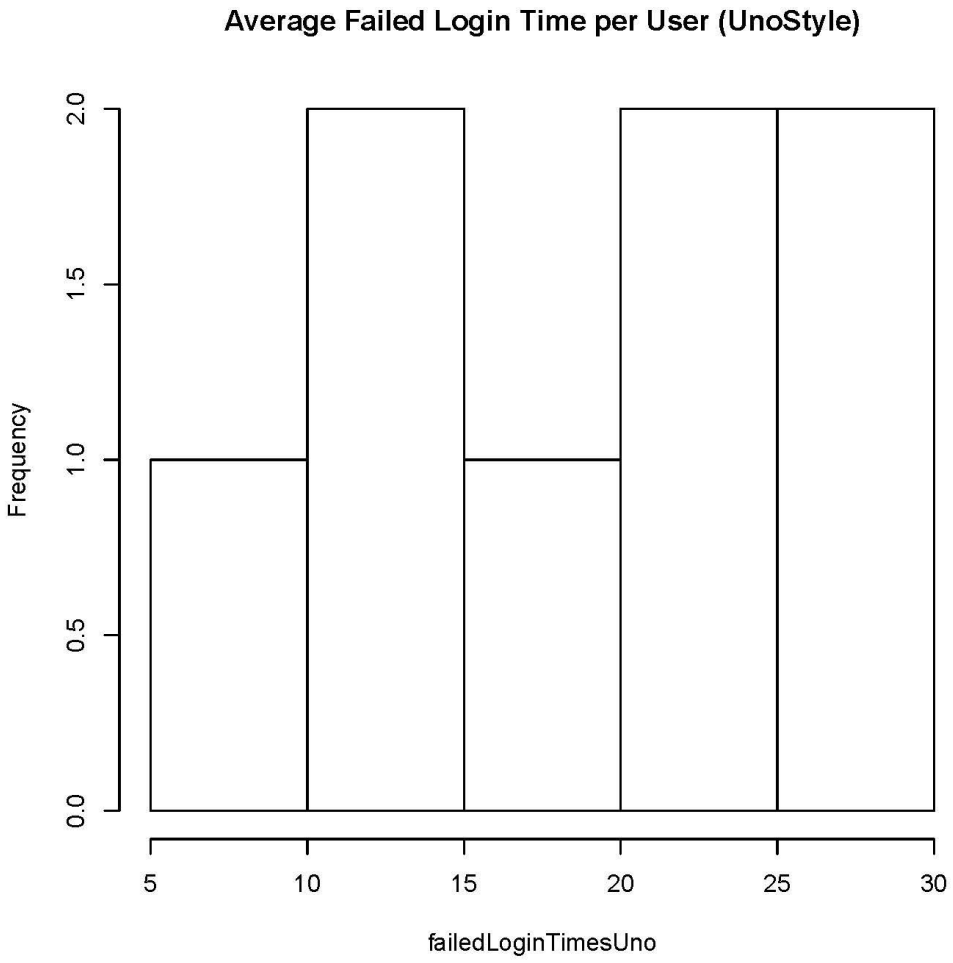


Fig. 23. Average Failed (Unsuccessful) Login Times per User (UnoStyle)

D. Combined Statistics

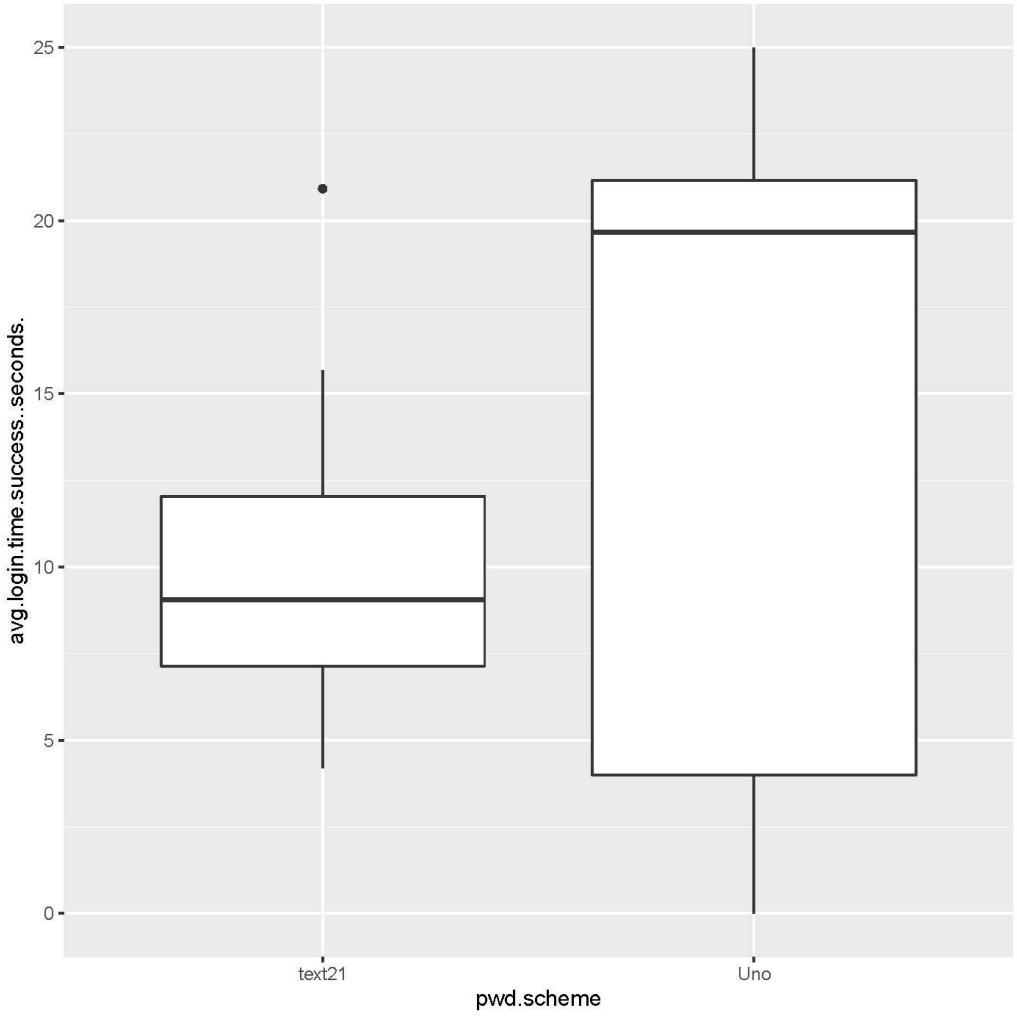


Fig. 24. Boxplot Comparing relation between password scheme and average successful login time

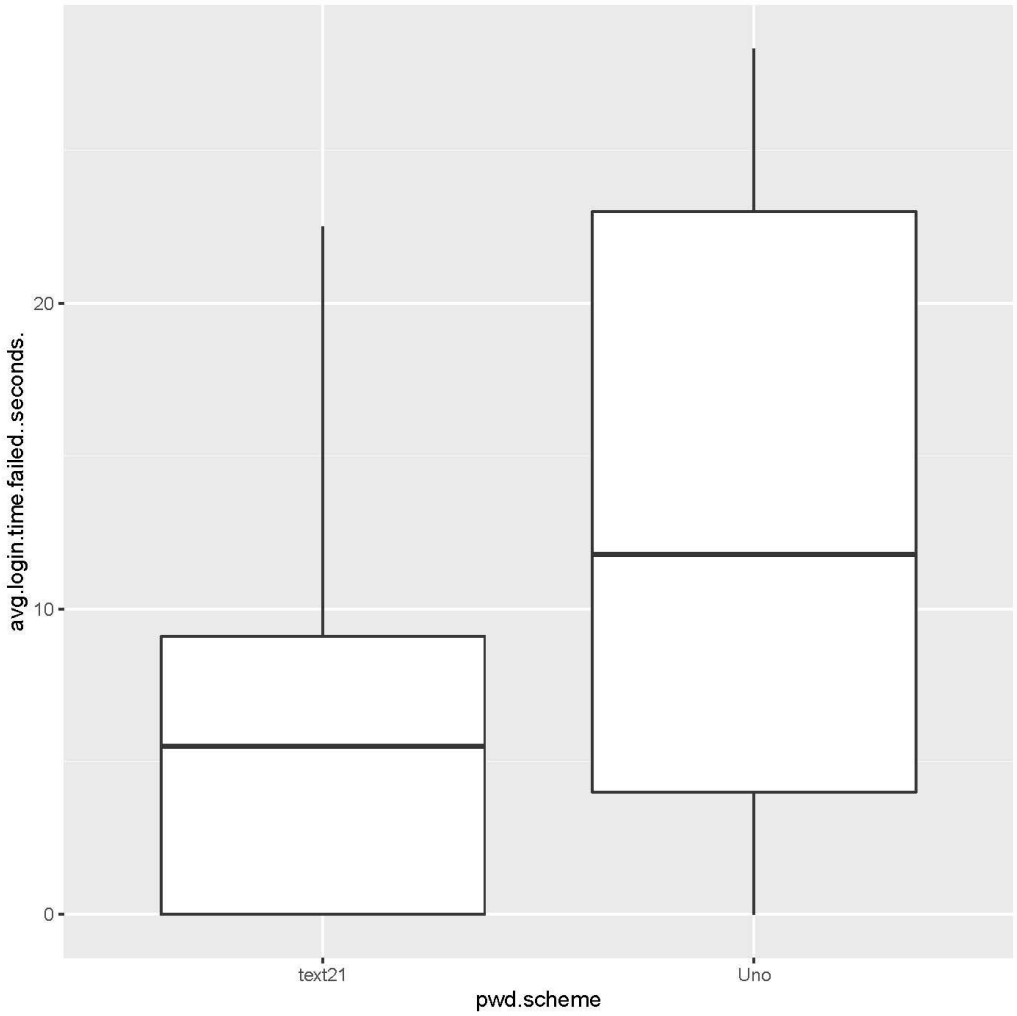


Fig. 25. Boxplot Comparing relation between password scheme and average failed (unsuccessful) login time

Null Hypothesis: There will be no significant difference in login times and no significant difference in successes versus failures.

Alternative Hypothesis: There will be a difference in login times and the number of successes versus failures.

Data Set	Text-21 p-value	Uno p-value	Normalized (p-value > 0.05)
Total Logins	0.115	0.03179	No
Successful Logins	9.84e-05	0.01854	No
Failed Logins	0.0007387	0.07207	No
Successful Login Time	0.1565	0.07293	Yes
Failed Login Time	0.0947	0.4482	Yes

Fig. 26. Data from Shapiro-Wilk tests in R

Data Set	Test Type	p-value
Total Logins	Mann-Whitney	5.562e-05
Successful Logins	Mann-Whitney	8.702e-06
Failed Logins	Mann-Whitney	0.1569
Successful Login Time	Unpaired t-test	0.0007527
Failed Login Time	Unpaired t-test	0.01803

Fig. 27. Data from Tests

From this data, we can reject the null hypothesis as we see a clear significant difference in the time taken to login as well as a significant difference in the number of failed attempts.

System Log Data

In order to either support our hypothesis, our team had 10 individuals partake in using the test system developed. Each participant was asked to sign a consent form informing them of the purpose of these tests, input each of the 3 randomly assigned passwords while they were displayed on screen, then immediately attempt to remember those passwords. Once completed, the individuals completed a short survey composed of 12 likert-scale questions meant to judge the effectiveness and efficiency of the new password scheme.

From the data logged by our systems and analyzed through R, it has become clear that our hypothesis has been disproven. As is clear in figure 25, the Uno style password scheme takes much more time to enter than the Text21 scheme. The mean of the Uno style was 19 while Text21 averaged out at 14.055. This indicates that there is a 5.055 second difference in the average time it takes to input a password successfully. Moreover, when comparing figures 22 and 23, there is consistency to the Uno scheme's successful password entry time which is not present when looking at the failure time. This is likely due to the system requiring a password entry to move on. When a user knows they've forgotten the password, the constraint likely frustrates them leading them to give up and enter the wrong password as fast as they can.

The number of successful password entries is quite low in the Uno style statistics. With a successful login mean of 1.545455 and failed login mean of 4.727273, it's clear that users frequently failed at remembering their passwords. In comparison with Text21's mean successes of 14.055 and failures of 2.555, the Uno style password scheme is much harder for users to remember. This error in memory is likely due to the difficulty in remembering a number paired with a colour. The vibrant colour is visually appealing, however in order to remember each one a user must remember the colour name and the value of that card.

When remembering characters, a user simply needs to remember the character itself. As human memory tends to recognize things better than it recalls them, the issue is likely due to the values being more recognizable than the colours themselves. Additionally, having 5 cards of different values and colours leads to a total of 10 elements to remember. This doubles the length of the

password in memory. Furthermore, we've become trained to remember characters for all of the traditional passwords we've ever used. This has led our brains to be conditioned into thinking of passwords as characters, not colours. While it's easy to recognize an 8 from a 9, it can be harder to recognize a blue 8 from a red 8. When you start mixing different values into these colours, the colours and values tend to blend together in memory.

Survey Data

The statistics below were calculated using R, the R script is entitled "statsUnoSurvey.R" and is located in the project archive for your reference. The survey-result.csv was exported from LimeSurvey with the following settings :

1. Export as CSV
2. Export questions as: Question Codes
3. Export responses as: Answer Codes
4. Completion state: Completed response only

The survey-results.csv was imported to R as a dataframe to parse and convert the likert-scale data to produce the statistic below.

Question (Strongly Disagree: 1, Disagree: 2, Neutral: 3, Agree: 4, Strongly Agree: 5)	Mean	Median	Standard Deviation
Q1: Do you often find yourself incorrectly entering a password scheme that contains only letters and numbers because you have forgotten it or mistaken the entered password for one that belongs to another account?	2.9	3	1.197219
Q2: Do you often forget new passwords for logins that you don't frequently use?	3.9	4	1.197219
Q3: Do you often reuse the same text and number combination passwords for different accounts for the sake of not having to remember another password?	4.6	5	0.5163978
Q4: Do you find it's easier to memorize and remember a set of characters and numbers more easily than images and colours?	3.7	4	0.9486833
Q5: Do you prefer choosing your own password of characters	4.8	5	0.421637

and numbers as opposed to a computer generated one?			
Q6: Do you believe the passwords you create are secure and strong that won't be easily broken into?	2.9	2.5	0.9944289
Q7: Did you find this password scheme that has visual features in combination with numbers more efficient to use when logging into an account as opposed to character and number based password schemes?	1.9	2	0.9944289
Q8: In regards to safety and security, do you feel this password scheme with visual features combined with numbers is more secure than a just character and numbers based password scheme?	3.9	4	0.7378648
Q9: How likely would you recommend this password scheme for everyone of all ages to use?	1.7	1.5	0.9486833
Q10: If you were to use this password scheme for one account as opposed to many, would you find it easier to remember and efficient to use?	3.4	4	1.577621
Q11: Did you find it harder to memorize the order of colours in the password scheme as opposed to the order of numbers?	3.6	4	1.173788
Q12: If you were to choose between using a normal text and number based password scheme and this UNOstyle password scheme how likely would you choose to use the UNOstyle password scheme?	1.7	2	0.4830459

Fig. 28. Survey Data Mean, Median, Standard Deviations

From the survey data, it becomes even more clear that our Uno password scheme is disliked by the users who have tested the system. Only one of the users said that they disagreed that characters are easier to remember than colours and images. The Q4 mean of 3.7 and median of 4 with a low standard deviation indicate that their answers support our conclusions on colour and

value memorization from the system log data. Furthermore, every user frequently uses the same password for different accounts and fully believe it's much easier to remember passwords they've chosen, as shown by the median of 5 and mean of 4.6 for Q3. These two correlate as their ability to choose their passwords will often have them recycling passwords from other platforms, though a mean of 2.9 and median of 2.5 for Q6 shows that roughly half of the users understand that this meant a sacrifice in the security of their online presence.

When asked if our password scheme was more efficient, we found a mean response of 1.9 and median of 2. A lower standard deviation of .994 indicates that most of these answers are similarly low. This and the log data regarding password entry times makes it clear that selecting the card colours and values take much longer than entering a character from the keyboard. This problem may have been resolved had we used hotkeys for card values, however it would reduce the learnability of the scheme. After many users specified their distaste for our password scheme, we were surprised to find that almost everyone believed it to be more secure than the standard password schemes which exist today. A Q8 mean of 3.9 and Median of 4 with an even lower standard deviation of 0.738 suggests this to be true. It is possible, however, that this data was skewed due to the randomly assigned passwords being more secure as well. After the analysis, we weren't surprised to find that no one would choose the Uno style password scheme over the traditional text-based ones out there now, nor would they recommend it to their friends. Both of these questions have some of the lowest values, Q9 with a mean of 1.7 and median of 1.5 and Q12 with 1.7 and 2, respectively. It's clear from this data that Text21 is a more efficient and usable password scheme, however it's possible that the Uno style is more secure.