Waterstons
we're with you

+44 345 094 0945
www.waterstons.com

Article

Jan 2022

Why your Wordle word won't work

Let's look at using data to try to play the perfect game of Wordle!

Wordle is a simple-looking word game – every morning, the game chooses a word and you are given 6 guesses to find it. With each guess, you are given clues as to how correct you are, allowing your follow up guess to (hopefully) be closer to the answer. The game, which is very similar to Mastermind, has become incredibly popular with the internet. While it started as a pet project, a game for the developer's partner to play, it now has over 2 million players trying to guess each day's word. If you've not played it, you may at least have seen the cryptic symbols that are an interpretation of your days attempt posted being on Twitter.

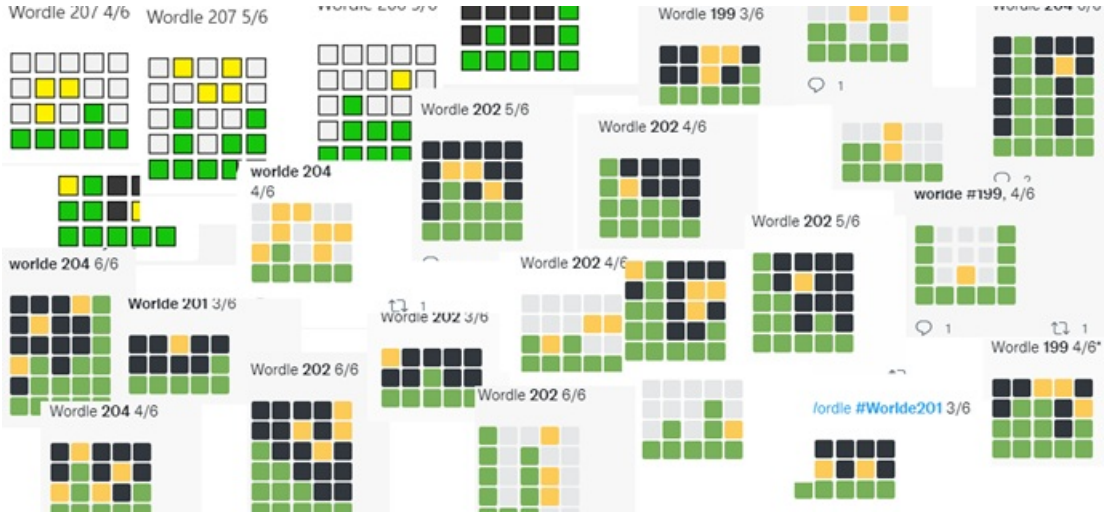


Figure 1: One day scientists may finally discover what these strange patterns and symbols mean

Wordle is deceptively simple, it can be quick to play, and you usually always reach your answer in your 6 goes. However, trying to find the best strategy means you have to answer some interesting questions: what is the absolute best word at the beginning, and then what is the best word to play at every other step? This is what the Waterstons' data team have been thinking about in meetings, pubs and chats over the last few weeks. The process of "solving" Wordle, thoroughly with a computer, may lead to a strategy that is not the most intuitive for a human to try to play the game with. However, by thinking algorithmically, we can start to try to build a computer better at Wordle-ing than us.

So, where should I start my game?

Typically, the most popular method of deciding the "best" first word comes from counting the vowels and consonants [1]. The argument goes, if we know that more words have "S" as the first letter than "Z", our first guess should not start with "Z". If almost no words have "uu" in them, we should not choose a word with this in it first. This is fairly intuitive, and by searching through the dictionary (specifically, the custom dictionary wordle uses to decide if your guess is a valid word) you can come up with a list of words that would appear to be the finest first guesses you could make. Doing this results in words like "Soare", "Saree" and "Seare" being suggested. Aside from this, I've also seen it noted that playing "Wader", "Bumph", "Stock" then "Vinyl" tests almost every letter, and maybe they should be played. Becky from Waterstons' marketing department (the genius who managed it Wordle #208 in her 2nd guess!) comes out to bat pretty hard for "About" as a first choice.

For a human, these ideas make a lot of sense – they play on how we understand words are constructed and what we instinctively know about often letters and combinations show up. If you were not a human though and looked at the problem with purely Spock-like logic, is there a better way of doing it? Instead of considering popular letters in words, we could focus on reducing the number of remaining viable words. This arguably leads you to win the game quicker. There are around 2,500 words that are acceptable answers in Wordle, at the beginning, you have no insight into what word will be correct. As you make a guess, the list of possible answers is made smaller and smaller and smaller, until you have 1 remaining option: the correct word. Thinking of the problem like this – a race to reduce the number of viable answers – gives us a separate way to try to pick the best word. Maybe, the best first word is the word that allows us to discard as many wrong guesses as possible.

Throwing away words

If you play "kites" as a first word you will get back a pattern of colours depending on how correct you are. For example, the clue might be [green, orange, green, black, black], or [black, orange, orange, black, green]. In fact, there are 243 different patterns they can be, each informing your next choice differently. Getting one of these patterns will therefore rule out large amounts of the dictionary, as you know there are words you can definitely no longer play. Our goal is to now find the word that guarantees you remove as many words from the dictionary as possible.

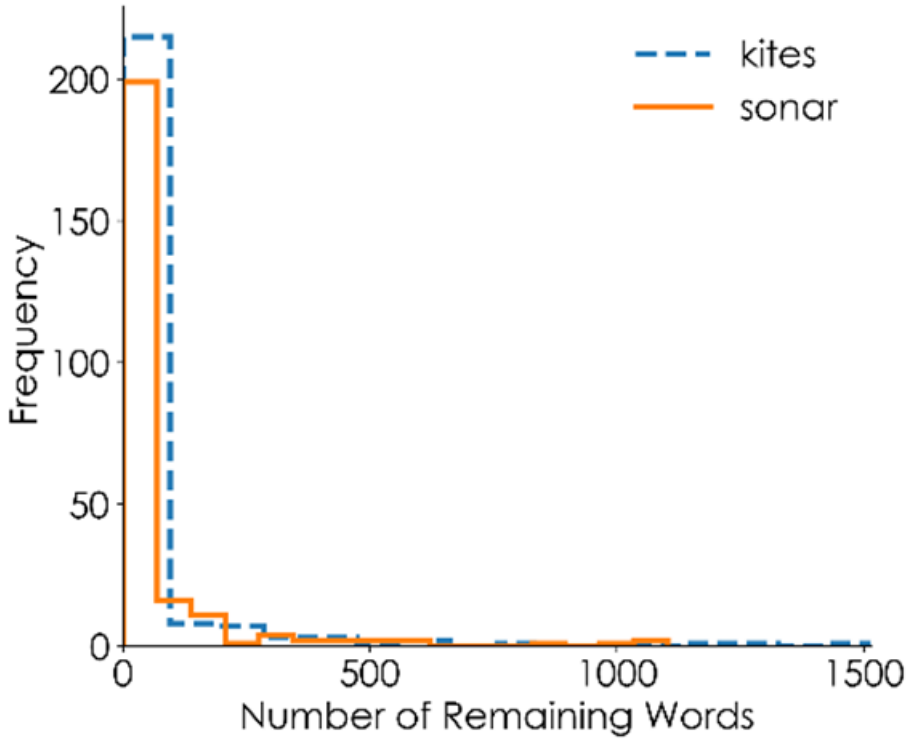


Figure 2: We can look at how many words can be left in the dictionary, depending on what word we play. Here, we compare all the options for "kites" against all the options for "sonar". We are gonna argue "Sonar" is the better word.

We are going to use this plot to argue that "Sonar" is a better first word than "Kites", but why? Well, If you play Sonar, even in the worst-case scenario, you are left with around 1000 words. If you play Kites, you are left with just over 1500. Now, we can apply this logic to the entire list of words you can play in Wordle. This gives us a new suggestion for the best first word, the word that is guaranteed to remove as much from the dictionary as it can. Doing this gives us the word "Serai" (see the dictionary at the end), followed by "Reais", "Soare", "Paseo" and then "Aeros".

Now, armed with a new best word ("Serai"!) we can try to Wordle with it! Let's try today's game: Wordle #209. Playing Serai gives us back:



Now, knowing this specific clue, we also know how many words are left in the dictionary (697 words long) and what they are. From here, we will perform a similar task as before. From this reduced dictionary, which choice is guaranteed to reduce the dictionary the most? Well, now that's "Banya". Playing this gives:



Knowing this clue, we can check how many words are left to choose from – it's 29. Following this procedure, of playing a word, getting the clue, and finding the word that reduces the list of words as much as possible, we can play until we get to a situation where only 1 word remains in the dictionary. This will be the correct word. Our next plays look like this:



Look at that! We've won! This method of algorithmically playing Wordle works by reducing the space as much as we can, as fast as we can. This is actually very similar to how solvers for the game Mastermind work [2]! Over time, the space collapsed like:

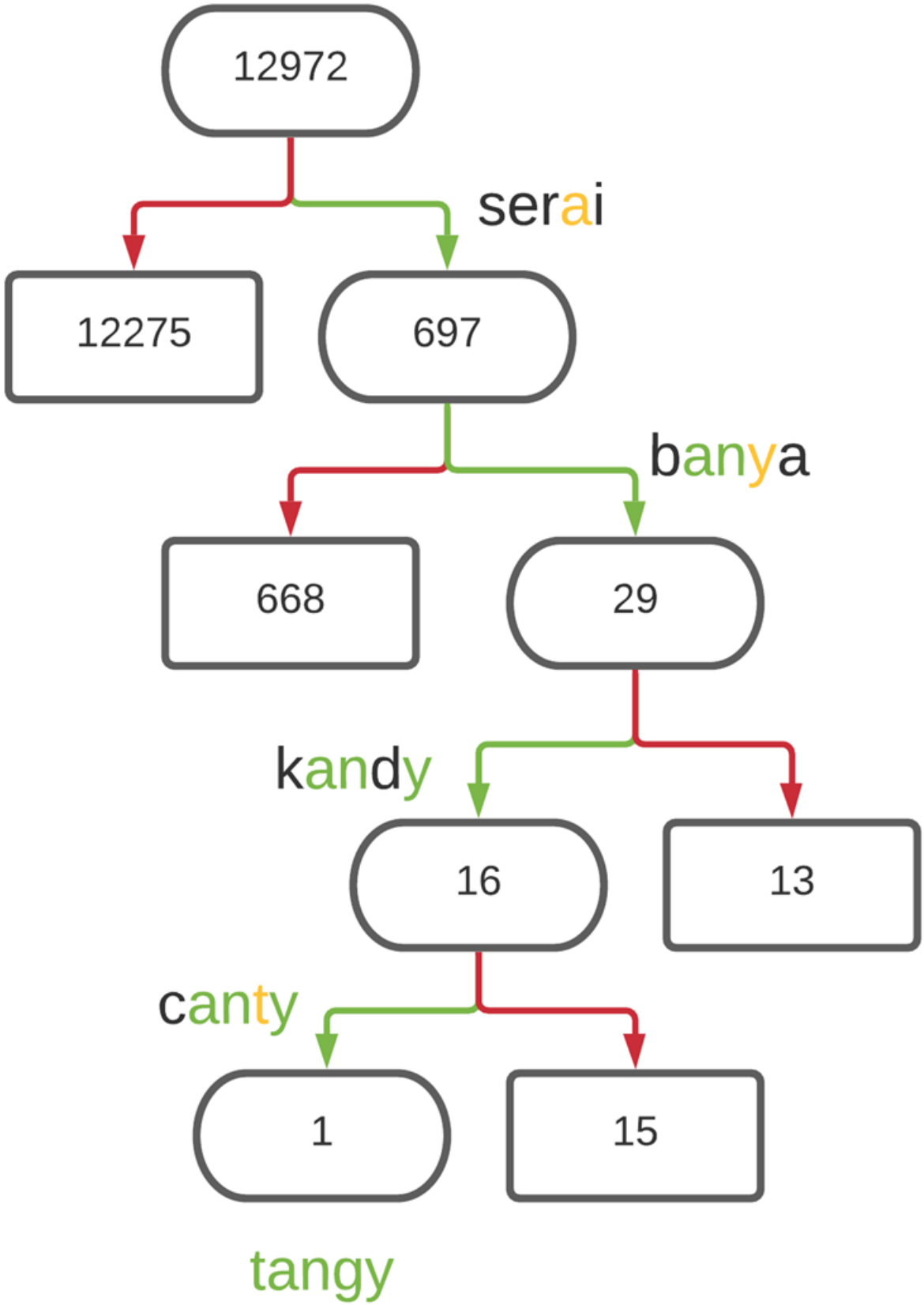


Figure 3: How the number of words left in the dictionary changes as guesses are played and clues are received.

Our “best” first word looks at how much the answer space is cut. However, we can do better! What if, for each first word, we play every possible game of Wordle? [AB1] Instead of going down the tree above, we go down every single possible tree. This is a computationally expensive task, requiring a lot of comparisons. However, by checking every guess you could make at every step, for every possible answer, we can find on average how quickly each starting word will complete the game. In our next post, this is what we will attempt! There are a host of optimisations that can be done to improve performance, and a lot of care needs to be taken to ensure you are correctly emulating Wordles rules. However, by using the concepts we’ve just talked about in this post, can we find an even better best first word?

Earlier this week, Alex (the head of innovation and leadership here) sent a company-wide emailing giving Waterstons’ employees a new task: play Wordle. From this important project, we will be able to find the number of attempts the average Waterstons’ employee takes to play Wordle. Hopefully, in the next article, this will give our method a worthy competitor!

Wordle is a simple-looking game that has inside it a fun and interesting problem. It shows that what is simple and intuitive to us, may not be the optimal way to solve a problem. There are some problems a computer can solve in ways that are not immediately obvious to us, with the end result being better than our intuitive first guesses. Next time we will try to push what we’ve done further – to find, arguably, the best first and second and third and fourth and fifth and sixth word you could ever play.

Ps. Here are some words I learned making this:

Serai: *noun*: another term for a caravanserai (an inn with a central courtyard) Seare: *adjective*: dry and withered Bumph: *noun*: documents containing uninteresting information Reais: *noun*: plural of real Soare: *noun*: a young hawk Paseo: *noun*: a walk or stroll *Banya*: *noun*: a Russian steam bath Kandy: A city in central Sri Lanka Canty: *adjective*: lively or brisk

[1] - https://bert.org/2021/11/24/the-best-starting-word-in-wordle/

[2] - https://en.wikipedia.org/wiki/Mastermind_(board_game)#Worst_case:_Five-guess_algorithm