



# BIODS253 W24: Statistical Wordle Assistant

Joseph Boen<sup>1</sup>

<sup>1</sup>Department of Biomedical Data Science, Stanford University



Stanford  
University

## How To Play

Guess the Wordle in 6 tries.

- Each guess must be a valid 5-letter word.
- The color of the tiles will change to show how close your guess was to the word.

### Examples

**W** **E** **A** **R** **Y**

**W** is in the word and in the correct spot.

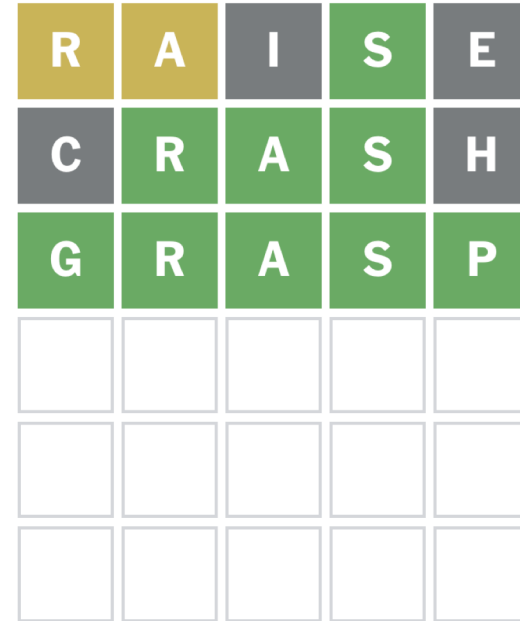
**P** **I** **L** **L** **S**

**I** is in the word but in the wrong spot.

**V** **A** **G** **U** **E**

**U** is not in the word in any spot.

## Wordle



# (Abridged) Design Considerations

## **Aims:**

A lightweight wordle assistant to help users make statistically informed guesses.

## **Non-Aims:**

This isn't a graphical wordle simulator, it's a statistical assistant. We will be doing some evaluations, but we aren't trying to design an optimal algorithm.

## **Requirements:**

Fast enough to run interactively, (with caveats for evaluations). Guesses and behavior should replicate behavior of NYT app, not necessarily other wordle copies.

## **Dependencies:**

This is designed to be super lightweight, so written mostly in pure python with some common libraries (NumPy, pandas). Nonetheless, packaged everything in a reproducible conda env.

# Intuition

The **quality** of a guess is how well it *filters* possible answers (2315 valid answers)

F	U	Z	Z	Y
---	---	---	---	---

# What Makes a Good Guess?

The **quality** of a guess is how well it *filters* possible answers (2315 valid answers)

F	U	Z	Z	Y
---	---	---	---	---

F	U	Z	Z	Y
---	---	---	---	---

Not much information gained, many words fit this.

F	U	Z	Z	Y
---	---	---	---	---

A lot of information gained, very few words fit this

# What Makes a Good Guess?

The **quality** of a guess is how well it *filters* possible answers (2315 valid answers)

F	U	Z	Z	Y
---	---	---	---	---

Quality = likelihood x information gained

F	U	Z	Z	Y
---	---	---	---	---

Not much information gained, many words fit this.

F	U	Z	Z	Y
---	---	---	---	---

A lot of information gained, very few words fit this

# Measuring Guess Quality with Information Theory

The **quality** of a guess is how well it *filters* possible answers (2315 valid answers)

F	U	Z	Z	Y
---	---	---	---	---

F	U	Z	Z	Y
---	---	---	---	---

1352 remaining,  
~58%

F	U	Z	Z	Y
---	---	---	---	---

2 remaining\*  
~0.08%

$$\begin{aligned} E[I] &= \sum_x p(x) I(x) \\ &= \sum_x p(x) \log_2 \frac{1}{p(x)} \\ &= - \sum_x p(x) \log_2 p(x) \end{aligned}$$

\*flyer or foyer

# Strategies

## Best Guess:

select *word* with highest possible expected information

R	A	I	S	E
---	---	---	---	---

C	O	U	N	T
---	---	---	---	---

4.91 bits,  
not possible answer

## Best Answer:

select possible *answer* with highest possible expected information

F	O	R	T	Y
---	---	---	---	---

4.33 bits,  
possible answer

Tradeoff between maximizing expected information and getting “lucky” on the next turn

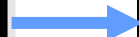


# Demo

C	R	A	N	E
---	---	---	---	---



```
python wordle_solver.py  
--board "crane" "GYXXX"
```



flexible usage, can help with first guess, or after (multiple) previous turns. ~0.5 sec latency enables interactive assistance.

There are 9 possible answers remaining.

All Guesses Ranked by Information (\* = possible answer)

ivory	3.170	tulip	2.948	sorry	2.948	slurp	2.948	court	2.948	*
viola	3.170	guile	2.948	quirk	2.948	folio	2.948	hovel	2.948	
lorry	3.170	could	2.948	intro	2.948	rumor	2.948	furor	2.948	
glory	3.170	quail	2.948	worry	2.948	ovoid	2.948	fluid	2.948	

... top 20/2315 shown

Possible Answers Ranked by Information

court	2.948	choir	2.642	curly	2.419	chirp	2.419	curvy	2.059
chord	2.725	curry	2.503	curio	2.419	color	2.419		

... top 9/9 shown

Using the strategy: best\_guess, we suggest picking one of:  
ivory viola lorry glory

Using the strategy: best\_answer, we suggest picking one of:  
court

# Evaluation

Play wordle over all 2315 possible answers with 10 repetitions\*, counting the number of guesses until correct. Compare against baselines

## Random Answer:

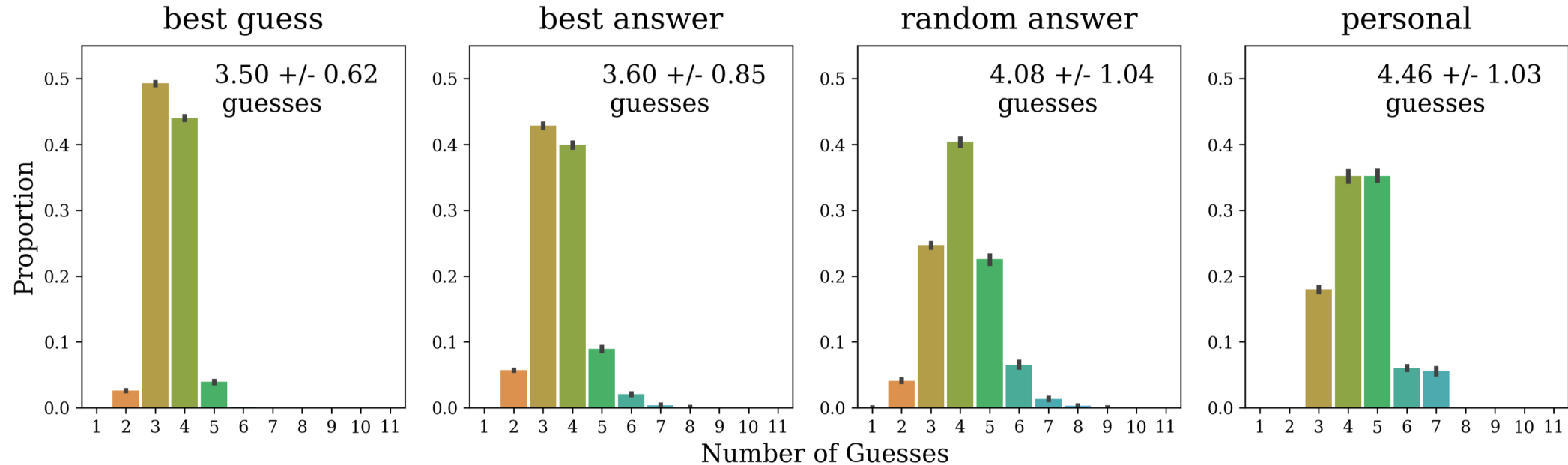
Select possible answers at each turn randomly.

## Human (aka me):

Bootstrap resample my historical performance on the NYT app.

\*if there is a tie in expected information, both strategies will randomly pick a guess

# Results



Using a paired\* t-test, we have with statistical significance that the strategies in order of least to most number of guesses are:

**best guess, best answer, random answer, and me (☹)**

\*paired per secret word



# (Some) Engineering Details

## Testing:

Wordle is surprisingly ambiguous with its rules, particularly concerning duplicate letters. I ran tests on these edge cases to ensure I replicated the NYT app's behavior.

C	H	I	L	D
---	---	---	---	---

wordle

C	O	U	C	H
---	---	---	---	---

guess

## Speed:

The solver is fast enough for interactive applications, with  $\sim 0.5$  sec/guess.

For simulations, too slow (2 secs/game, 1 hour per strategy). After profiling, introduced intermediate caching to yield 0.35 secs/game,  $\sim 13$  mins per strategy)



# Thanks for Listening!

Still polishing repo for final submission, but check it out here:  
[https://github.com/tboen1/biods253\\_final\\_project](https://github.com/tboen1/biods253_final_project)



Stanford  
University