Note: This writeup is intended for educational purposes only and may not accurately reflect the safety of mushrooms. Do not use any information in this writeup or infographic to make any decisions, especially decisions for consuming mushrooms.

i. A brief recap of your data, goals, and tasks, focusing on those that most directly influence your design.

In my project, I utilized the Mushroom Overload 6.7M dataset, a comprehensive collection of data points about different mushroom species. This dataset includes a wide array of categorical variables that describe the physical characteristics and habitats of mushrooms. The primary focus of the project was to leverage this dataset to teach the reader how to distinguish between edible and poisonous mushrooms using data science techniques. The vast amount of data available in the dataset allowed for a detailed exploration of patterns and correlations that can inform safer foraging practices.

The main goal was to use the dataset to identify key characteristics that differentiate edible mushrooms from poisonous ones. To achieve this, I focused on analyzing the distributions of mushrooms categorized as edible or poisonous across various features. Specifically, I examined the overall percentage of mushrooms that are edible versus those that are poisonous, which provided a foundational understanding of the dataset's composition. This overarching analysis helped frame the subsequent, more detailed investigations into specific mushroom features and their impact on edibility.

The tasks I undertook were centered on understanding the distribution of edibility in relation to specific features like habitat and cap color. By analyzing these features

independently, I aimed to identify trends and patterns that could serve as reliable indicators of mushroom toxicity. For instance, I explored how the likelihood of a mushroom being edible or poisonous varied depending on its habitat, which could suggest environmental factors influencing edibility. Similarly, I investigated the relationship between cap color and edibility, providing insights into whether certain visual traits are associated with higher risks. These analyses were crucial in building a data-driven approach to mushroom identification, helping to distill complex datasets into actionable knowledge.

ii. Screenshots of and/or a link to your visualization implementation (see below for additional guidance)

NOTE: I have split the visualization onto two separate pages to improve readability in the writeup. It is one continuous document in the real world, and the continuous version is appended in the notes at the end of the writeup.

NOTE 2: The PDF of the infographic is also included in the following GitHub repository, which has even better readability. Thank you!

https://github.com/emontj/DataVisualizationFinal/

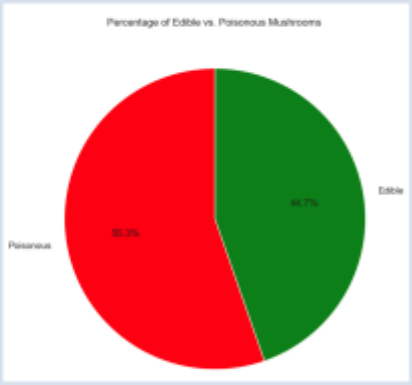
(see next page for start of visualization)

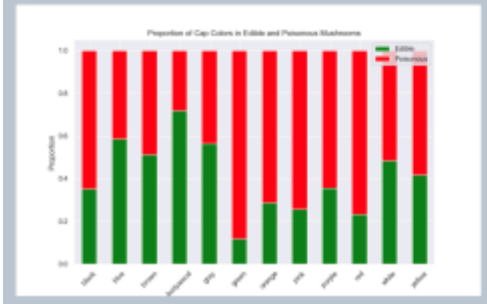
WHAT MUSHROOMS CAN I EAT?

Identifying Wild Mushrooms as Edible or Poisonous with Data Science

Worse than a coinflip

Roughly 55% of wild mushrooms are poisonous. If knowing nothing more about mushrooms, it is more likely that a mushroom is poisonous.



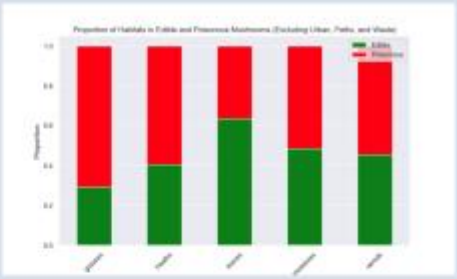


When considering cap color, some are safer bets

Cap color alone does not indicate safe-to-eat mushrooms, but burlywood, gray and blue are amongst the safest. Green and red are almost always poisonous.

Avoid mushrooms growing in the grass

The best place to pick mushrooms is from leafy areas. Meadows and woods are tossup habitats for mushrooms with nearly 50/50 splits, and grass and hearth-habitat mushrooms are more often dangerous.



When taking a gamble:

Bet on mushrooms that are brown or burlywood and grow in leafy areas. Avoid mushrooms that grow in grass and are red or green.

And, do your own research!

Disclaimer: This chart is intended for educational purposes only and may not accurately reflect the safety of mushrooms. Always consult a reliable source or expert before consuming any wild mushrooms.

Data Source: BwandoWando. (2024). 🍄Mushroom Overload | 6.7M Rows [Dataset]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/5490971

by Ethan Montgomery

iii. A summary of the key elements of your design and accompanying justification

For my project, I designed an infographic-style presentation to communicate the key findings from the Mushroom Overload 6.7M dataset. This format was chosen because it allows for a clear separation of the multiple tasks involved while ensuring that the information is presented in a visually appealing and easy-to-digest manner. It also builds iteratively with a simple stat followed by multiple more complex stats. The infographic format also enables a structured progression of ideas, guiding the reader step by step through the analysis. By breaking down the data into manageable sections, I aimed to make the information more accessible, especially for those who may not be familiar with mushrooms (I was not familiar with mushrooms going into this project, either).

I began the infographic with a broad statistic that highlights the overall distribution of poisonous versus edible mushrooms. This initial overview serves as a simple foundation, providing the reader with a big-picture understanding of the dataset. By presenting this general statistic first, I aimed to establish the context for the analysis and give the reader a clear sense of the scope of the issue at hand—how prevalent poisonous mushrooms are compared to edible ones. This starting point is essential for framing the more detailed analysis that follows, and teaching the reader how this visualization intends to think about the mushroom edibility question.

Following the broad overview visualization, I introduced more specific analyses that build on this foundational information. The next section of the infographic focuses on the likelihood of a mushroom being edible or poisonous based on its cap color. By

isolating cap color as a variable, I provided insights into how certain visual characteristics can be indicative of a mushroom's edibility. This section was designed to help the reader understand the visual cues that might suggest whether a mushroom is safe to eat, making the information both practical and relevant for real-world application.

The analysis then progresses to examine the relationship between mushroom habitat and edibility. This section of the infographic explores how the environment in which a mushroom is found can influence its likelihood of being poisonous or edible. By adding this layer of analysis, the reader can see how external factors, beyond just physical appearance, play a role in determining mushroom safety. This iterative buildup of information, from general distribution to specific characteristics, ensures that the reader is gradually deepened into the analysis, enhancing their overall understanding.

Finally, the infographic combines the findings from the cap color and habitat analyses to present scenarios where mushrooms are most likely to be safe for consumption. This synthesis of information offers actionable insights that can help the reader make informed decisions in real-life foraging situations. By bringing together the different elements of the analysis, the final section provides a comprehensive summary that ties all the findings together, offering clear guidelines for identifying edible mushrooms. This structured, step-by-step design ensures that the information is not only conveyed effectively but also leaves the reader with a practical understanding of the factors influencing mushroom edibility.

One last thing to add is that I included a disclaimer stating that this chart should not be used, as this is just an educational project and the information in here is not

necessarily “valid” without further consultation from experts or reliable sources. I included this to ensure that the visualization is ethical and does not mislead anyone. I also ensured viewers of the infographic were aware of this disclaimer in my testing mentioned in subsequent sections of this writeup.

iv. A discussion of your final evaluation approach, including the procedure, people recruited, and results. Note that, due to the difficulty of recruiting experts, you can use colleagues, friends, classmates, or family to evaluate your designs if experts or others from your target population are unavailable.

In my final evaluation approach, I conducted a two-part testing procedure to assess both the informativeness and memorability of the mushroom infographic I created. I recruited participants amongst friends, family and colleagues and first showed them the infographic, followed by a test set of 10 mushroom images, each known to be either edible or poisonous. During this first test, I allowed the participants to refer to the infographic as they decided whether each mushroom was safe to eat. This initial test was designed to see how well the infographic communicated the necessary information and whether it effectively guided the participants in making accurate decisions.

After the first test, I had the participants complete a second test using a different set of 10 mushroom images. This time, however, I did not allow them to look back at the infographic. My aim here was to evaluate how well they remembered the information they had learned from the infographic. By comparing their performance on this test with the first one, I could assess whether the infographic's information was retained and could

be applied without direct reference. This step was valuable in determining the long-term impact of the infographic as a learning tool.

v. A synthesis of your findings, including what elements of your approach worked well and what elements you would refine in future iterations.

In synthesizing my findings from the evaluation, I discovered several key takeaways about what worked well and areas that could be refined in future iterations. The infographic proved to be both effective and memorable, as evidenced by participants' ability to recall and apply the information, particularly the statistic that 55% of mushrooms are poisonous. This statistic was frequently recited first, indicating that it made a strong impression. However, while the infographic was successful in conveying core information, it lacked the depth needed to fully equip users with all the knowledge necessary for accurately determining mushroom edibility. There were additional factors within the dataset that could have been included to provide more comprehensive guidance, which would have added meaningful value to the tool.

One element that posed a challenge was the way habitats were presented. The dataset included various habitats where mushrooms could grow, but some of these, like "hearth," were confusing for participants. Additionally, there was a general lack of awareness that mushrooms could grow in places other than grass, which highlighted a gap in the audience's knowledge that the infographic could have addressed more clearly. This suggests that in future iterations, I should consider either simplifying or more thoroughly explaining these habitat categories to ensure they are easily understood and correctly interpreted by users.

Interestingly, I found that participants were familiar with the color "burlywood," a label from the dataset that I initially thought might be too obscure. My own unfamiliarity with this color had led me to worry that it might confuse others, but this concern was unfounded. This insight reassures me that certain seemingly obscure details can be included without detracting from the overall clarity of the infographic. Moving forward, I would refine the design by incorporating a broader range of factors that influence mushroom edibility, and by providing clearer explanations of complex or less commonly known elements, such as habitat types. This would enhance the infographic's depth and utility, making it a more robust tool for learning.

Additionally, I found that a uniform sample of 10,000 mushrooms held very similar relationships to the full 6.7M dataset. I was able to quickly iterate on my charts by narrowing down to a dataset of 10,000 mushrooms and test out different charts. This was valuable in terms of workflow and ultimately worked well, because producing charts for the full 6.7M dataset was extremely slow! This was a key learning point and something I will keep in mind for future large-scale projects, even if just to foster the development process and not to draw any insights from.

vi. Sources

Dataset: BwandoWando. (2024). Mushroom Overload | 6.7M Rows [Dataset]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/5490971

vii. Notes

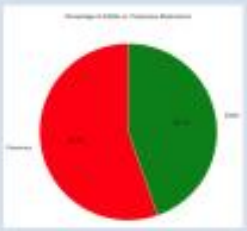
This writeup is intended for educational purposes only and may not accurately reflect the safety of mushrooms. Do not use anything in this writeup or report to make decisions regarding consuming mushrooms.

WHAT MUSHROOMS CAN I EAT?

Identifying Wild Mushrooms as Edible or Poisonous with Data Science

Worse than a coinflip

Roughly 55% of wild mushrooms are poisonous. If knowing nothing more about mushrooms, it is more likely that a mushroom is poisonous.

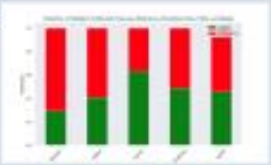


When considering cap color, some are safer bets

Cap color alone does not indicate safe-to-eat mushrooms, but burlywood, gray and blue are amongst the safest. Green and red are almost always poisonous.

Avoid mushrooms growing in the grass

The best place to pick mushrooms is from leafy areas. Meadows and woods are tossup habitats for mushrooms with nearly 50/50 splits, and grass and hearth-habitat mushrooms are more often dangerous.



When taking a gamble:

Bet on mushrooms that are brown or burlywood and grow in leafy areas. Avoid mushrooms that grow in grass and are red or green.

And, do your own research!

Disclaimer: This chart is intended for educational purposes only and may not accurately reflect the safety of mushrooms. Always consult a reliable source or expert before consuming any wild mushrooms.

Data Source: BeaverMania. (2024). 🍄Mushroom Overload | 6.7M Rows [Dataset]. Kaggle. https://doi.org/10.34740/KAGGLE/DS/5349311

by Ethan Montgomery