September 9 and 11, 2024

- Questions/clarifications from the Zybooks
- Why numpy?
    - Loops are slow!
    - Everything is linear algebra!
    - Manipulating arrays and doing math on them is a lot of data science!
    - Turn pokedata into arrays and manipulate.

Details

Get into groups. I want you to go to colab.research.google.com and upload the pokemon.csv file to your workspace, as we did on Friday. I can help with this.

1. Import all the relevant packages (pandas and numpy)
2. Make a dataframe that has the columns corresponding to just the stats of the Pokemon (hp, attack, defense, special_attack, special_defense, speed).
3. Convert this dataframe to a numpy array.
4. Sort this data along the rows. What do you notice? What questions can you ask from here? Ask one and answer it.
5. Sort this data along the columns. What do you notice? What questions can you ask from here? Ask one and answer it.
6. Invent your own Pokemon and add its stats to the last row of the array.
7. Give a one-dimensional array that gives the *stat total* of each Pokemon. The stat total is the sum of the stats for a Pokemon.
8. Give a one-dimensional array that gives the *stat product* of each Pokemon. The stat product is the product of the stats for a Pokemon.
9. Can you figure out, using `pandas`, which Pokemon has the largest stat product? Which has the largest stat total? Is it the same Pokemon?
10. Give an array that shows the largest stat of each Pokemon.
11. Give an array that shows the largest value of each stat in the Pokemon dataset (there should only be six numbers in this array).
12. Can you figure out, using `pandas`, which Pokemon has the largest attack? The largest defense? Etc?
13. Similary, find the Pokemon with the smallest attack? The smallest defense?
14. Which stat is largest, on average, across the Pokemon in this dataset? Which is smallest?

15. Which stat has the largest median across the Pokemon in this dataset? Which is has the smallest median? Are the answers the same as the previous question?

16. Write the code to turn the array you have into a one-dimensional array, sorted so that the data is in the order you would get by starting at the top left number and working your way down the first row, then jumping to the second row at the end of the first row, and so on.

17. Reshape the array to be any compatible shape you want. Does this mean anything in the context of the Pokemon dataset?

    - What happens if you use an incompatible shape?

18. Pick two stats (your choice) and two `type1`s (your choice) and use `matplotlib` make a scatterplot of the two stats against each other. Give your plot an appropriate title, label your axes, and color the points according to the `type1`. Create a legend that indicates which colors are used for each `type1`. *You can use the old faithful example in the text as a guide here.*

19. Pick two more stats and make a scatterplot of these with the same `type1`s. Make a figure that includes your two plots (this one and the one from the previous question) as subplots. Make sure each subplot has a caption.