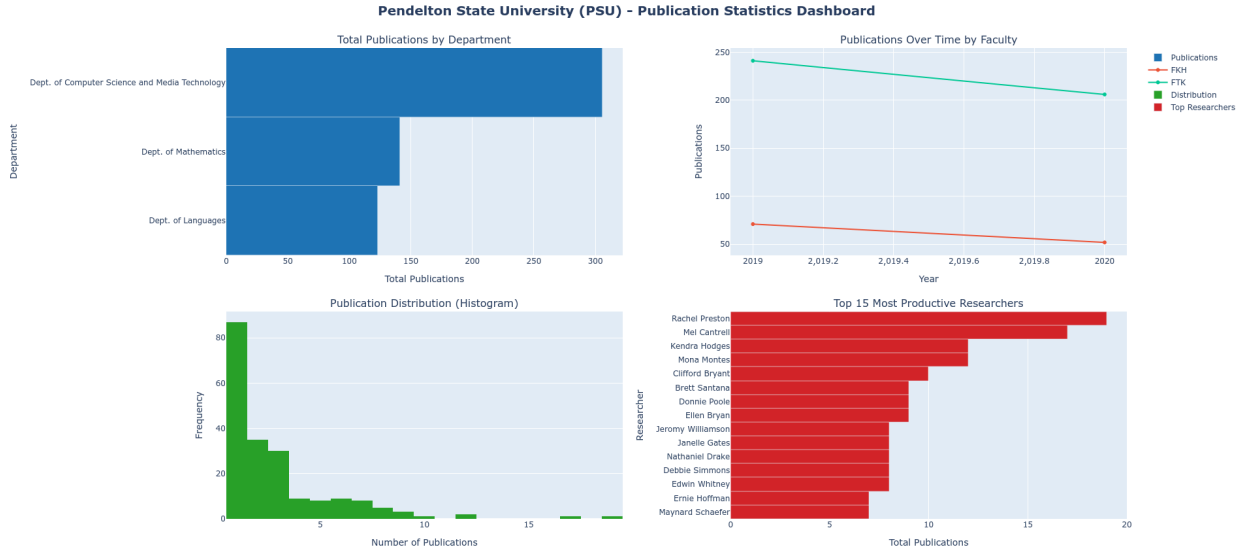
Task 1

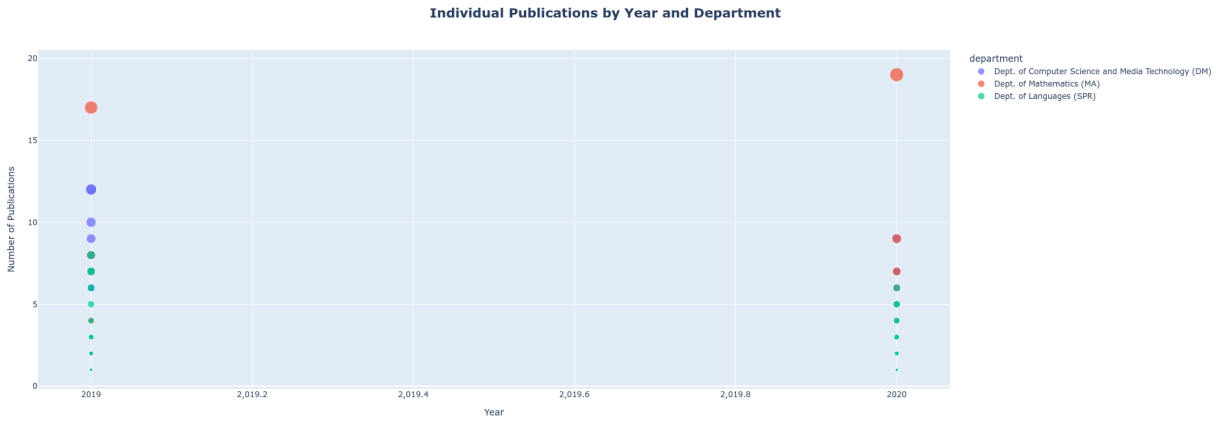
When I first looked at the data, it instantly made sense to me and its structure wasn't hard to internally visualize. The data consisted of several entries in .json format where each entry had 5 strings and 2 integers. It is supposed to show how many publiciations different researchers has made and from which university, faculty and department they are from.

As a solution for this assignment i created a simple python script and used pyplot in order to create different plots for html pages. In order to run it you can follow the README. The first page generated from the script “publications_dashboard.json” just shows some simple plots that i have juxtaposed that shows for example bar charts of how many publications each department has made and which researchers has made the most contributions. I also showed a lineplot over how many publications has been made over time but since we only have 2 years worth of data it is hard to dra insights from this.

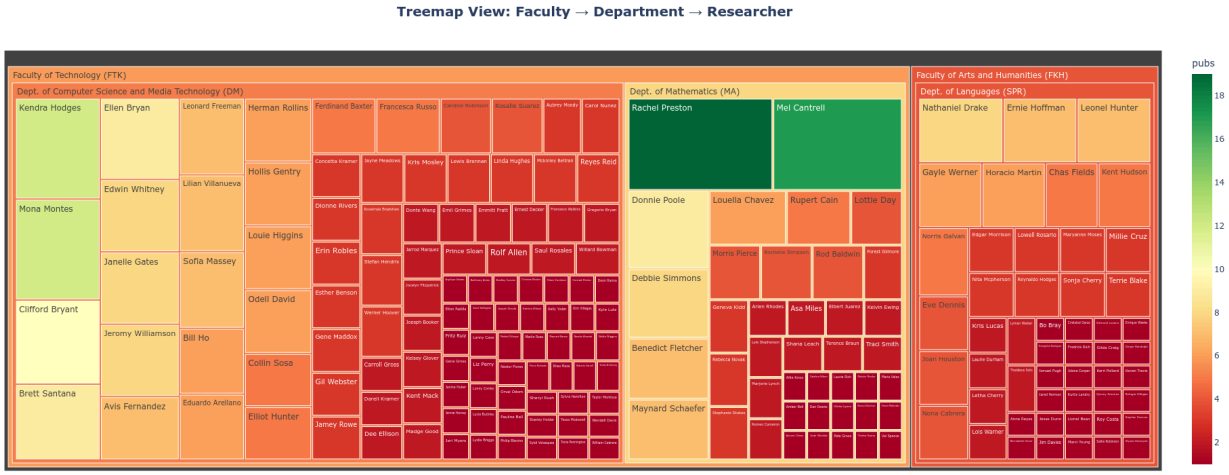


I also included some interaction tools like zooming and selection in order to highlight and look deeper in to the distributions that i created.

As a purely experimental part of this assignment i also wanted to include a scatterplot but since the data is distributed in such a way it is hard to make a valuable scatterplot so i didnt draw any conclusions from this but it is a good way to try out the lasso selection, box selection and zooming features.



Then i created my personal favorite plot which is a hierarchichall heatmap of the entire dataset which shows the top contributions from each researcher, department and faculty respectively.



Task 2

Pros:
What i really enjoy about the tool is that i can easily pick and chose the tensorset and metadata and i also like that i can easily chose the parameters of each algorithm with a slider instead of choosing the numerical value which gives me perspective on how much im decreasing or increasing the parameter. Also i like that the tool follows basic infovis principles like overview, zoom and detail on demand.

Cons:
Sometimes there are some performance issues especially with algorithms like T-sne but that might just be because i have a low-performance computer and computations might be made on client-side.

Design choices to reconsider
When clicking a specific datapoint i would like it if the specific information appeared right along side it and that i could with a click select multiple ones and see fine-grain differences between them instead of relying on the selection tool but it is a bit nit-picky since i dont know if this is a realistic problem.

Use cases and stakeholders:
I feel like this tool would be used for ML and NLP engineers or researchers who want to make a sanity check on if their dataset is good and if they have good separation of their dataset and how different words are bunched together as well as their respective relationships.

However i do feel like there is a good use case for beginners in this field to get more familiar and to interact with the algorithms in an intuitive way and learn new things.