GSoC - 2023 : Wrapped

Amartya Nambiar

2023-08-24

|  |
| --- |
|  |

My contribution at [**EMBL - EBI**](https://www.ebi.ac.uk/) for [**Google Summer of Code**](https://summerofcode.withgoogle.com/)(GSoC) involved creating visualizations that would help in comparative metagenomics. Additionally, I used Vega which introduced an approach that would allow a common grammar for creating visualisations. This was done to enhance code re-usability in both React and Jupyter Notebooks, regardless of programming language.

After the results were released, my mentors [**Martin Beracochea**](/The%20Project/mentors.html#martin-beracochea) and [**Sandy Rogers**](/The%20Project/mentors.html#sandy-rogers) remained in close contact with me during the community bonding period. Subsequently, the coding period unfolded, which I’ve structured into three distinct phases to align with the three projects I undertook as part of my contributions.

### Phase 1 30th May - 30th Jun

* Delved into relevant research papers, understand the intricacies of assembly runs, samples, studies and analyses, which form the foundational concepts of MGnify.
* Documented all visualizations present on the MGnify website which was used in Phase 2 - [Viz\_Recon](http://localhost:4843/posts/Viz_Recon/)
* Significant chunk of Phase 1 was spent contributing and collaborating with [Alejandra Escobar](https://www.ebi.ac.uk/people/person/alejandra-escobar-zepeda/) (MGnify Bioinformatician) on the enhancement of her Jupyter notebook - **Drawing presence/absence KOs for one metagenomic sample**, specifically focusing on Pathways Visualization.
* The objective was to streamline the notebook, making it more concise, interactive, self-explanatory, and immediately usable.
* **Technologies Utilized** **:**
  + [Docker](https://www.docker.com/) + [JupyterLab](https://jupyterlab.readthedocs.io/en/latest/) + [R](https://www.r-project.org/)
  + [Quarto](https://quarto.org/)
  + Packages like [KEGGREST](http://www.bioconductor.org/packages/release/bioc/vignettes/KEGGREST/inst/doc/KEGGREST-vignette.html), [Pathview](https://pathview.uncc.edu/overview#kegg_view)
* **Pull Request** **:**
  + Worked on the existing PR by Alejandra, **Link :** <https://github.com/EBI-Metagenomics/notebooks/pull/26>
  + **Status :** **Under Review** 🟡

: : : : : : : : : : : : : : : : : :

### Phase 2 1st Jul - 14th Aug

* The second phase deals with exploring and building visualizations for the MGnify website.

#### **Phase 2A** **1st** **Jul - 16th** **Jul**

* This phase dealt with exploration of visualization technologies that could be used both in javascript and python.
* This involved testing packages like plotly and highcharts preferaablly highcharts bcos mgnify website uses Highcharts.js for visualization in both JS and Python.
* Then cam up the idea of exploring a common visualization grammar that would solve thise problem efficiently.
* Tested out pandas.js, danfo.js, vegalite
* This is important because visualizations then could be easily ported to jupyter notebooks where the user would be able to use data on the same visualization with custom or wrangled data that serve their purpose.
* By the end of this phase, we decided to go ahead with vega ecosystem bacuase it serves our interest welll.

. . .

#### **Phase 2B** **17th** **Jul - 14th** **Aug**

* This phase dealt with exploration of visualization technologies that could be used both in javascript and python.
* This involved testing packages like plotly and highcharts preferaablly highcharts bcos mgnify website uses Highcharts.js for visualization in both JS and Python.
* Then cam up the idea of exploring a common visualization grammar that would solve thise problem efficiently.
* Tested out pandas.js, danfo.js, vegalite
* This is important because visualizations then could be easily ported to jupyter notebooks where the user would be able to use data on the same visualization with custom or wrangled data that serve their purpose.
* By the end of this phase, we decided to go ahead with vega ecosystem bacuase it serves our interest welll.

: : : : : : : : : : : : : : : : : :

### Phase 3 14th Aug

* Working on Varsha Kale’s idea of Study Summary, more specifically detailed analysis GO abundamces - GO-slim and GO data.
* Involves working ontologie web format of gene ontologies
* The visualizations for the same include a bar chart for go slim, a tree for parent and subclass GOs and a heatmap for GO annotaions over asembly runs for a parent and subclass GOs. tree & heatmap use GO full abundance
* note that altair is used which is the py package for generating vegalite viz
* **Technologies Utilized** **:**
  + [JupyterLab](https://jupyterlab.readthedocs.io/en/latest/) + Python
  + Packages like Altair, graphviz, owlready..

### Unexplored Horizons