**Formicidae**

**Visualized**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

[The Process Book]

The Paleobiology Journey of three young coders,

hungering for adventure.

Visualized by: Asas Husain Book by: Brendan Liu

Aaron Fei

CSCI – 14a, hosted by Zona Kostic, Jared Jessup, Brian Chu, & Frederec Im.

[Synopsis]

The Paleobiology has long been known as one of the messiest fields of all science, due to the absurd amounts of fossil records which make management excessively difficult. As a response to common data mismanagement made within, we came up with a visual solution to solve issues such as species being confused with one another, tables being far too large to sort manually, and as a result, publications resulting in mass discrepancy from such tiring work.

Our visualization sought to compile all this data automatically, pinpointing and plotting their exact location of discovery, age, length, etc. It then makes it super easy for viewers to locate such data plots via an interactive user interface. How did it do it? Keep reading and you’ll know how we came up with such a great tool :)

A close up of a piece of paper

Description automatically generated

[This is about 1/20th of the data we’re reading. Shocked? Don’t worry! We’re only looking at 1/100th of the whole database for just ants, and yes, that includes columns.]

[Goal-posting]

Hearing news about a final project while taking CSCI-14a, the three of us, Asas, Brendan, and Aaron, connected on Discord for a get-together brainstorm. This is when Asas brought in his love for Paleontology, which would tremendously differentiate the soon to come project/visualization. That’s right, we were diving into the world of ants! (Ants were Brendan’s idea blame him)  
  
From there, the starting point and direction of progress was established. It was decided that we would put together a graph that both demonstrated ant genus’ growth over time (millions of years to be exact), as well as our takeaways from everything taught over 7 informative weeks. Perfect, now all that was needed to start was a little data collection.

[Finding Purpose]

Journeying into the Paleobiology Database, we soon came face to face the true purpose of our project. Biology databases… were freaking messy. Thousands of records, all with a massive array of attributes, equaled a chaos of a database to sort through. Since this was the case, we were forced to limit and filter our data twice, else our graph would end up overflowed, overwhelmed, and possibly even unreadable (worst nightmare!).

It was also in this moment, that we truly knew what we had to do. We were building this graph, for the greater good of every viewer out there. We’re humans, we like everything short and sweet after all.

[Proto-Procedure]

The first prototype produced from the filtered database (now a local .csv file), was a simple numpy/Pandas-based Python file, capable of displaying at matplotlib scatterplot of every ant’s length over time and… not much else. The file could only be run successfully via Jupyter Notebook, and had yet to see its days as a true app. It was built mainly by Asas Husain, who would soon take his graph to further levels of development.

[Intermission]

Following their Milestone 2 submission, the boys had their future in foresight, but they realized that their little project could definitely use a greater touch of charm. This called for… a 2nd Brainstorm! There must be some way to better sort a mass of 200+ records…

And then… it happened. Brendan Liu sprang up the idea that would give the project life. The boys would be plotting the records on an interactive Earth. Enticed, Aaron Fei took it up as his duty to construct as user-friendly a World for locating the many curious records of ant fossils around the world.

Alongside the planned world graph came Asas’ grand rebranding, via a spaghetti chart! A chart that would take his previous scatterplot and take it to greater detail, seeking to more thoroughly meet the original goal!

Impressed by his teammates, Brendan decided to take a more creative approach to his role, which would be building a PowerPoint as a bridge connecting the graphs to the audience. Though not as heavily involved as the others, he attempted to pitch in assistance whenever he could, researching certain topics and gathering more possible ideas.

How do all these look like? Turn the page to find out.

[[CORE]]

[ZA WARUDO]

A picture containing grass, covered, many, bunch

Description automatically generatedThis, is the World Map. It’s an entire map of the world, but plotted within are all the data plots of our favorite ants! It comes complete with a working zoom and pan function!

All the data is stored on a local database, which holds all the plots and their attributes. The app.py & index.html retrieve the code from such in order to display this lovely.

Each plot has a latitude and longitude associated with their site of recovery, deciding their location on the map. Along with that, the plots are differently sized depending on how many specimens of the same species were uncovered in that very area. Each plot also has a flavor text box, with a variable based description that further entails the details associated with each plot.

A later change introduced the ability to work with a Prominent Edge, allowing polygons to be drawn in the map and function as a sort of measurement tool. This allows measurement of area and perimeter of the area selected. Furthermore, it adds even more interactivity to the User interface!

Fun Faccs: Following the app’s completion and deployment to Heroku, Aaron Fei, the madlad, had considered employing the same trick except with COVID-19 Cases around the world. Unfortunately, time crunches cut it short, and it was then that he decided to add the extra features to this very map.

A screen shot of a computer

Description automatically generated[ SKITTER SCATTER ]

Following up, what would’ve been a spaghetti chart. What’s a spaghetti chart, you ask? OverSimplified, a multi-line chart. Explained more thoroughly, gives every unique genus their own distinct line, showing each of their associated lengths changing over time.

Following an unfortunate first incident with the spaghetti chart, later on it was remodeled back into a scatterplot, which highlighted ant size changes over time in a cleaner “condensed” manner. The data is optimized before being displayed, adding all body part lengths for the most accurate representation. It now serves as the direct answer to the original goal, showing ant size changes over time!

[[Previous Endeavors]]

[ Extra: Reinforcement-based Age of Empires II A.I. ]

Before the AntMap came forth, the team had their eyes set on building an AI player with Reinforcement-based Machine Learning. The target game? The real-time strategy classic, Age of Empires II. This project was led mainly by Asas, whom with Brendan proposed the process of replicating algorithms based off of human professional players. The AI would act off a sort of decision tree, replicating net-gain behaviors while backtracking/abandoning net-loss behaviors. Via many trials, it would eventually compile a dominantly beneficial algorithm. However, High usage of outside modules made the project near impossible in their amount of time given, so the project was dropped in the name of paleontology! (and in favor of simple visualization.

[Roles]

* **Brendan Liu:**
  + Built the first iteration of the PowerPoint, and designed the entire Process Book you’re reading right now! :D
  + Edited together the screencast for compatibility and clarity!
  + Minor but Major: put together the citations.
  + Creative Outlet, brainstorming a whole bunch of the ideas that soon came to life.
* **Asas Husain:**
  + Kick-started the initial process single-handedly, building our first initial proposal for machine-learning Age of Empires II.
  + Even after time crunches revealed themselves, he kept strong for the next phase tinkering with a stubborn spaghetti/scatterplot graph without giving out.
  + Voiced over the screencast, the trooper he is!
  + And of course, he finalized our beautifully proud Presentation on Slides ^-^
* **Aaron Fei:**
  + Core-drive of the 2nd & 3rd phase, building and adding all the amazing features to the amazing interactive world map!
  + Kept a strong spirit all the way through, even in troubling times. Synopsis: he’s our hero.

[[Farewell]]

Everything good must come to an end… and sadly, so must our precious little process book. But, to leave off on a good note, we’ll go re-summarize everything you’ve been introduced to, and show how much we’ve grown as students, and together, as a team.

The AntMap was set out to provide as friendly a UI for all curious Paleontologists seeking to plot their unorganized blocks of data. Did it succeed in this field? Of course, and we’re super happy of how far we’ve managed to come with this tiny project of ours. Sure enough, it didn’t truly answer the original concept: showing ants and their growth over time. But… scatterplot to the rescue; we accomplished everything we set out to do.

[[Farewell, Cont.]]

As for us? Phew… what a journey. I have an almost certain feeling that none of us have ever experienced teamwork this intense. Calling to build things together, sharing our progress, lifting each other up when we were down, having a baller, I, Brendan, am overjoyed to have shared this moment with my new friends. It’s been a long 7 weeks, but no matter the stress, it’s been learning curve I was so happy to hit my stride on. From me, from all of us, thank you all. I wish you the best for all your future experiences together as instructors.

* Brendan Liu

Well I think it’s an interesting thing that all three of us who do not have great programming and data science experience are in a group. Fortunately, this helps us to learn a lot more than others.

* Aaron (Dele) Fei

[References]

PBDB Team, “Fossil Records of Formicidae in North America”, *The Paleobiology Database,* <https://paleobiodb.org/classic/displayDownloadGenerator>

Master of Cartography and Geographic Information System, “leaflet实现简单量算工具功能(附源码下载)”, Southwest University, *Zhihu.com*, <https://www.zhihu.com/column/p/156919183>

Kostic, Zona, 2020, CSCI S-14a (34493), assisted by Jared Jessup, Brian Chu, and Frederec Im, Harvard University, <https://www.summer.harvard.edu/course-catalog/ssp-courses/building-interactive-web-applications-for-data-analysis/34493>

Holtz, Yan, 2018, “#124 Spaghetti Plot”, *The Python Graph Gallery*, <https://python-graph-gallery.com/124-spaghetti-plot/#:~:text=A%20Spaghetti%20plot%20is%20a,propositions%20to%20make%20it%20better>

plotly Team, “Plot CSV Data in Python”, *plotly,* <https://plotly.com/python/plot-data-from-csv/> <https://plotly.com/python/plot-data-from-csv/>

Riccardo, 2016, “Filter Leaflet Maps with a Slider” , *Digital Geography*, <https://digital-geography.com/filter-leaflet-maps-slider/>

codeofsumit, 2015, “leaflet-geoman”, *GitHub,* v2.6.0,Accessed August 2020, <https://github.com/geoman-io/leaflet-geoman>