

Welcome to DataFirst midterm event

Fall 2023



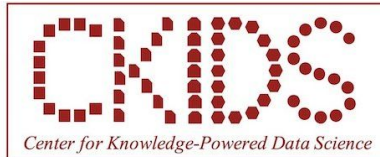
Manu Sharma, Medium

We hope that you are all having fun and learning more about **data science**, **teamwork** and the **domain you get to work on** this semester!

90 students

18 projects

USC



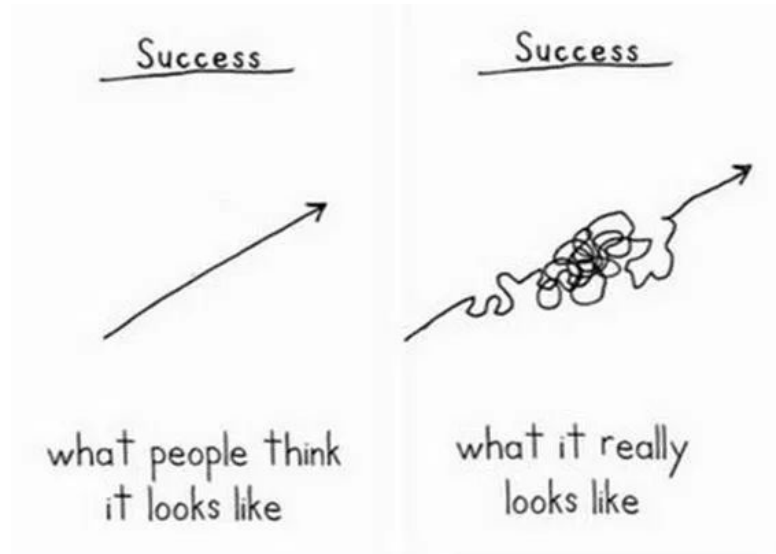
GRIDS @ USC

Midterm: How does it work?

- After our presentation, students will present their results on a poster
- Navigate through the rooms to talk to your fellow DataFirst participants
- Get to know other projects and don't be shy!

Slides adapted from Fred Morstatter, Deborah Khider

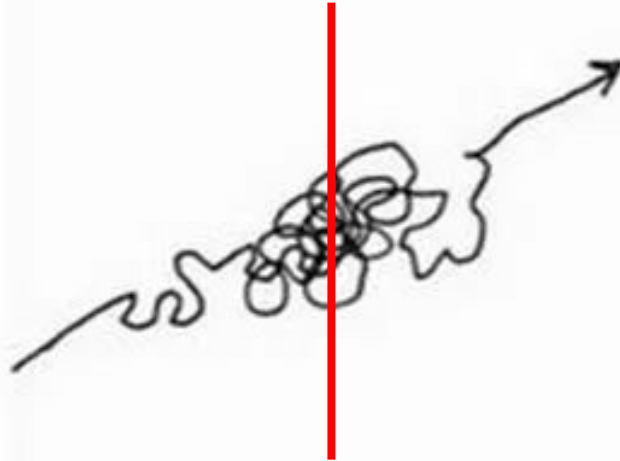
Progress is built on failure



Success

Midterm (you are here)

Final



Also...
Papers you've read
Talks you've heard
etc.

what it really
looks like

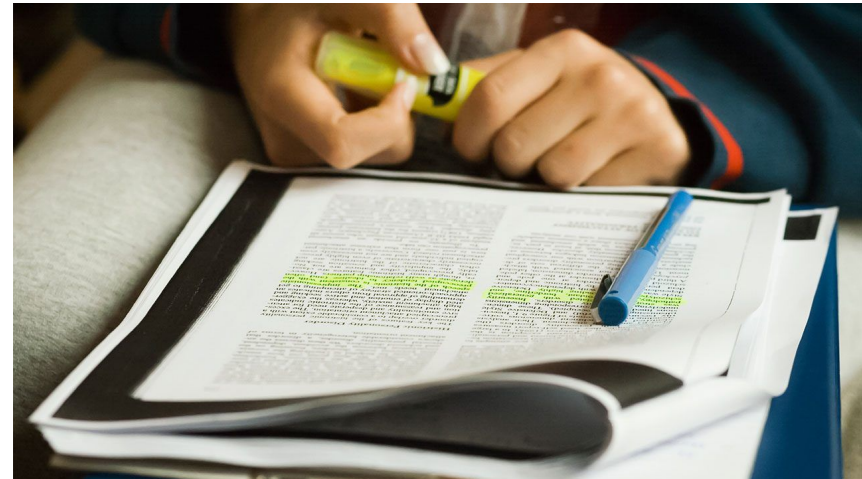
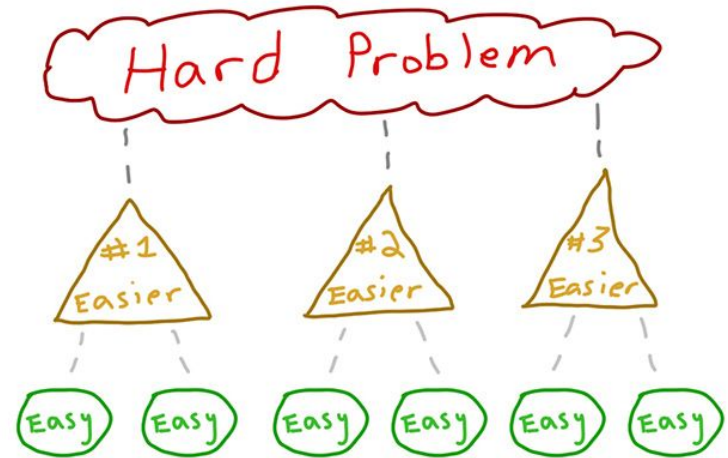
State of your DataFirst project

- Hopefully, everything is going fine. Likely, that isn't the case.
 - Missing or flawed data
 - Problems need to be more refined
 - Initial approach didn't work
 - Initial hypotheses invalidated
-
- Don't worry! We've all been there.
 - Some tips for moving forward.



Debugging the problem

- List all the components involved. Think about where the weak link might be.
- Design a logical and simple troubleshooting process to find the problem.
- Ask for help. Ask the internet, ask colleagues, ask other labs. There's no shame in being stuck, even on a basic technique.
- Read. Think, and read again.



Tip: It could be the data

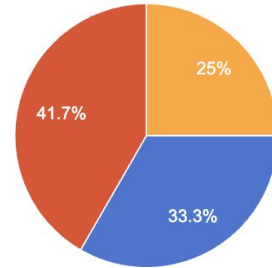
Know your data!

- Some projects have dataset provided.
 - Ensure that this data is actually representative of the problem.
 - Labels, too!
- Check data collection process for:
 - bias
 - sufficient size
 - variety
- All data needs cleaning before being usable!



Work with your teammates

- Figure out how best to work with your diverse team -- who can you collaborate with and learn from
- Tip: Find someone who can help/collaborate. Working with other students in small sub-groups (2-3 students) can help to foster teamwork
- Communicate with fellow team mates, to figure out how you can help them and viceversa



- Very collaborative
- Somewhat collaborative, I am working with at least one other peer
- I am working on my own



Rethinking the problem.

- Make sure your role is clearly scoped -- what do you want to accomplish?
- Tip: blend ideas. Try multiple, different, new approaches.
- Put yourself into an environment where you can be exposed to new ideas.
- Read, experiment



Get feedback from outsiders!

- You don't have all of the answers.
- Your mentors probably don't either.
- Use this midterm as an opportunity to get feedback!
- Use today to ask questions to *other* groups. They may be working on a different problem but their solution could work for yours.



Looking Forward

Save the Date: Final Presentations

Friday, December 1, 2023, 5-7 pm

- ~5 minute presentation / per group
- Award Ceremony

What to include in presentation:

- Motivation and research question
- Outline of the research
- Major findings
- Overall conclusion

Create a project website before Final Presentation!

Why create a website?

- Create a digital and visual archive for your project
- Share your project's findings with a broad community
- Add project website to your CV
- Enter to win "Best Website" Award!

Information to include:

- Project motivation
- Specific problem you are addressing
- Data and methods
- Results and some fancy visualizations
- What you've learned
- About (mentor and student information)

Bootstrap

markdown



GitHub Pages



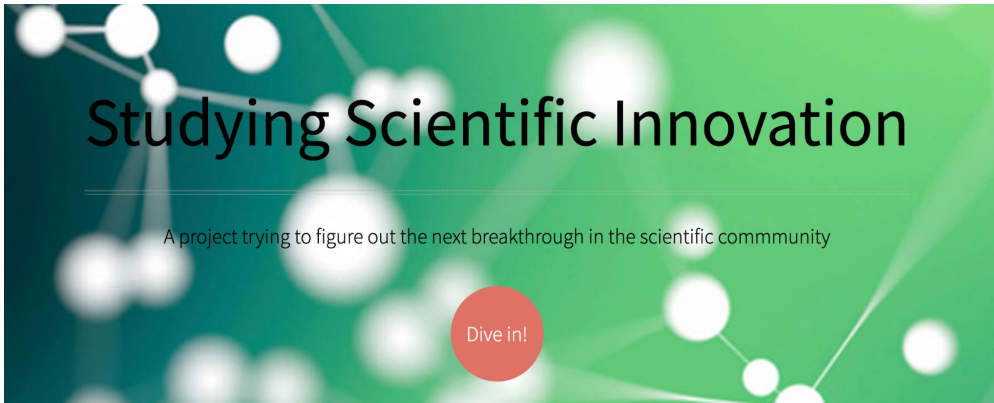
Google Sites

Example project websites



Homeless Encampments in Los Angeles

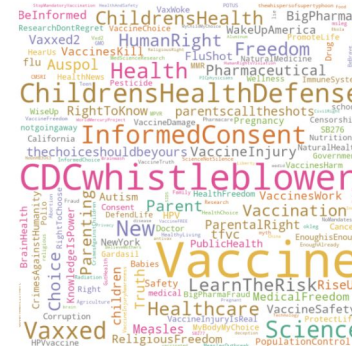
<https://sites.google.com/usc.edu/homeless-encampments-in-la>



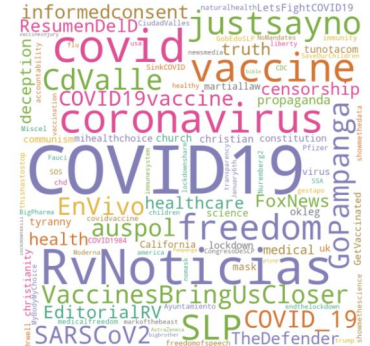
Award for Best Project Website Spring 2022:

<https://vineet-agarwa11.github.io/>

[studying-scientific-innovation.github.io/index.html](https://vineet-agarwa11.github.io/studying-scientific-innovation.github.io/index.html)



Anti-vaxxer pre-covid
hashtag wordcloud



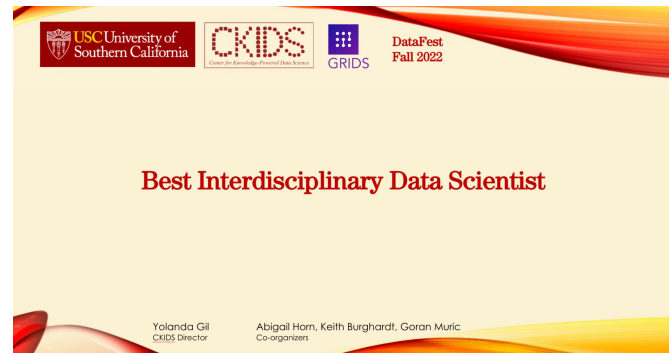
Anti-vaxxer post-covid
hashtag wordcloud

Covid Vaccine Misinformation

<https://sites.google.com/usc.edu/covidvaccinemisinfo>

Incentives for Participation: Awards!

- At the end of the semester, faculty mentors nominate their students for awards
- There are 9 types of awards, for individuals and for groups (including for project website)
- Awards are provided when there are ties as well
- Each award receives an **\$100 Amazon gift card**
- Something else to add to your CV!



Next Steps

Share your mid-term feedback!

Sharing your experience will help us to help you!

This is an anonymous form.

USC GRIDS CKIDS DataFest Midterm
Feedback

<https://forms.gle/TiDbnne7i4YA7nCLA>

Upload your slides onto GitHub

Product Solutions Open Source Pricing

Search or jump to... Sign in

ckids-datafirst / 2023-fall-paleoclimate Public

generated from ckids-datafirst/template-project-website

Code Issues 3 Pull requests Actions Projects Security Insights

Files

main

Go to file

- .github
- assets
- config
- content
 - approach
 - authors
 - data
 - people
 - problem-statement
 - results

Pyleoclim.pdf

index.md

2023-fall-paleoclimate / content / results / Pyleoclim.pdf

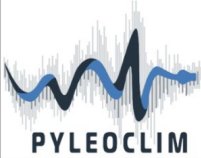
uhermjacob Add midterm poster ✓ bdf0988 · 2 days ago History

778 KB

Pyleoclim: A Python Package for the Analysis of Paleoclimate Data

Introduction

Pyleoclim is an object-oriented Python package for analyzing and visualizing time-series paleoclimate data, which offer unique challenges to the analyst, as they usually come in the form of timeseries with missing values and age uncertainties. Our goal for this semester is to increase its functionalities such as anomaly detection, optimizing time-series analysis, and visualization styles.

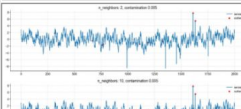


Methods

- Understand the functionalities on generated data/examples from blog post
- Write functions in Pyleoclim that apply methods (ML for anomaly detection/visualization) to paleoclimate data
- Write CI tests
- Write Documentations
- Write a tutorial notebooks on how to use the method/functionality

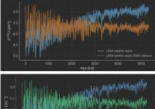
Issue 1: KNN on anomaly detection (Sunny Lee)

This issue aims to explore how KNN works for timeseries data especially on paleoclimate data. Moreover, we also implement automate auto-tuning of the parameters using the silhouette method to heuristically determine all parameters



Issue 2: SciencePlots Functionality (Ginny Barnes)

SciencePlots - API for developing standardized plot styles using Matplotlib (plots for IEEE, science, nature articles)



Approach

- Generate notebook with Pyleoclim/SciencePlot plot comparisons (IP)
- Add SciencePlots into Pyelo dependencies and environment (IP)
- Append new SciencePlot styles to Pyelo Plot.py file
- Create Unit Test to ensure smooth integration of SciencePlots

Send us links to these uploaded slides

Midterm DataFirst 2023

- Date: Friday, October 20, 5-7pm
- Location: SAL 101

Presentations

Paleoclimatology

We will upload these presentations to the website

