

# 2024 Spring Data Science Boot Camp

Introduction

# Welcome!

- Welcome to the Spring 2024 Data Science Boot Camp!
- In this boot camp we will:
  - Learn some python
  - Learn some data science
  - Complete a data science project

# Top two resources

- Boot Camp Website,  
<https://www.erdosinstitute.org/programs/spring-2024/data-science-boot-camp/>
- Erdős Institute Slack
  - spring-2024-cohort is a private channel.
    - You should already be a member!
  - spring-2024-data-science is a public channel you should join.

# Lecturer

- Steven Gubkin, PhD
  - Head of Training and Assessment at Erdős since 01/01/24
  - First time lecturing this bootcamp!
- Graduated from OSU Math in 2016
- Taught math at Cleveland State from 2016 - 2023



# Group Project Coordinator

- Alec Clott, PhD
  - Head of Data Science Projects
  - Sr. Principal, Quantitative Analytics and Data Science at Gartner
- Graduated from OSU Political Science in 2021



# Alumni Advisor

- Matthew Osborne, PhD
  - Boot Camp Lecturer from 2020 to 2023
  - Senior Operational Analyst at NetJets
- Graduated from OSU Math in 2020



# Your contact for access

- Amalya Lehmann, PhD
  - PhD in Music History, Literature, and Theory from UC Berkeley
- Your top contact for:
  - Slack channel access
  - GitHub repository access



# The Erdős Institute Projects

Spring 2024



# Goals

- An opportunity to work with real-world data and produce findings in a short time-span
- Focus on substantive areas (environment, health, health, finance, etc.) using techniques from the bootcamp.
  - Note that the focus should be on using what we learn.
  - Okay to use more advanced methods. Just make sure to compare their performance to the best model you could make using methods covered in the bootcamp.
- Building your portfolio is crucial in the data science market, provides a framework for job interviews

# Projects

- Portfolio-worthy data science project/product
- Includes:
  - 5-minute overview video and presentation
  - Annotated GitHub
  - Executive Summary
- Reviewed by project judges
- Top 5 projects will present to all participants in our closing ceremony for the Spring 2024 Bootcamp

# Team Formation

# Background of Bootcamp attendees

- Hundreds of students from all over the world
- Some of you may know other attendees, others of you won't
- Many different backgrounds (subject areas, experience with coding)\*
- Various types of data science career goals
- Various goals for the bootcamp
- Various goals for the projects

\*And that is totally fine and expected!

# Read these documents

<https://www.erdosinstitute.org/programs/spring-2024/data-science-boot-camp/>

(Project Information at Bottom)

# Project Expectations

# Overall Structure

- **Team size: 3-4 people**
  - Can choose group or...
  - Can be matched with others based on your job goals or data science interests
- **Goals: “portfolio” project**
  - Can be used in job interviews (when the time comes)
  - Results have business value
  - Communicate to lay-people and team of data scientists
- **Structure**
  - Group meetings -- each group decides how much time they want to spend
  - Weekly office hour w/ Alec (open to all)
    - Weekly 12:00 - 12:30 ET and by appointment
  - Check-in with project mentor on a regular basis (15-30 min)

# Project Requirements

- **Instructions at the bottom of the Spring 2024 Data Science Bootcamp Page**
- **In order to get an Erdős certificate, you must complete a data science project start to finish**
  - Project must be coded in Python
  - Have an annotated GitHub repository
  - Executive summary of your project results and implications
  - ***For presentation day:***
    - **5-min** pre-recorded PowerPoint presentation detailing project process from start to finish
    - Judges will vote on winners!
    - More info will be given closer to project day



# Your To-Do List

# First Important Dates:

## Schedule

Click on any date for more details

<b>Lecture 1: Introduction</b> February 5, 2024 at 3:00:00 PM	EVENT	<b>Math Hour (at 10) / Office Hour (at 11)</b> February 6, 2024 at 10:00:00 AM	EVENT	<b>Problem Solving Session 1</b> February 8, 2024 at 3:00:00 PM	EVENT
<b>Lecture 2: Data Collection</b> February 12, 2024 at 3:00:00 PM	EVENT	<b>Math Hour (at 10) / Office Hour (at 11)</b> February 13, 2024 at 10:00:00 AM	EVENT	<b>Problem Solving Session 2</b> February 15, 2024 at 3:00:00 PM	EVENT

**Feb 19, 2024** 11:59 PM **Watch 3 Previous Top Projects**  
Consult the project database, and watch at least 3 previous top projects from Erdos Alumni.

**Feb 29, 2024** 11:59 PM **Watch video about Project Formation**  
This should help answer any Q's you may have going into project formation

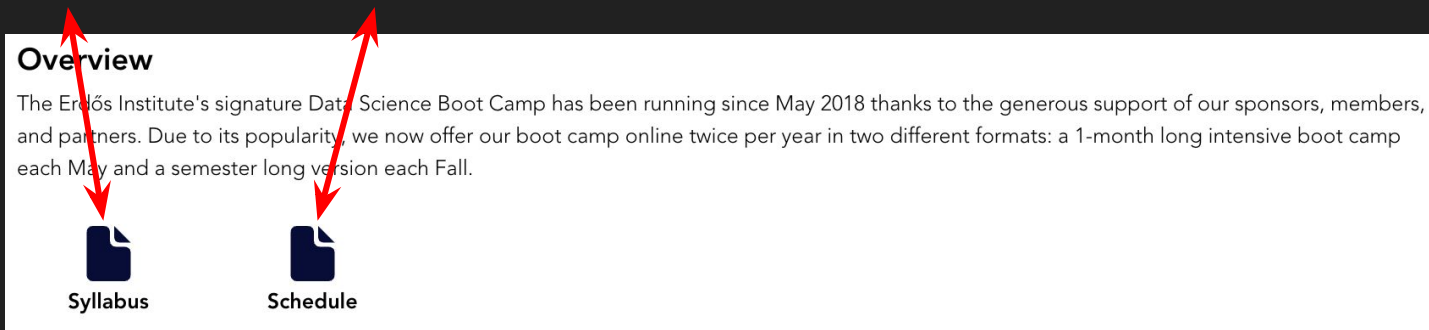
**Mar 1, 2024** 4:30 PM **Project Pitch Hour**  
Opportunity to meet with other Erdos Fellows and form teams and propose topics.

**Mar 8, 2024** 11:59 PM **Submit Team Proposal or Idea to Project Formation Page**  
If you want to propose a project, or have an idea for a project, submit it by this date.

Note: You can find these dates at the bottom of the course website



# Boot Camp Format: Non-Project Portion

- 12 Live Lectures
- 12 Math Hours
- 12 Problem Solving Sessions
- All Zoom links can be found in your Erdős profile or on the course website
- Syllabus and Schedule can be found on the course website



**Overview**

The Erdős Institute's signature Data Science Boot Camp has been running since May 2018 thanks to the generous support of our sponsors, members, and partners. Due to its popularity, we now offer our boot camp online twice per year in two different formats: a 1-month long intensive boot camp each May and a semester long version each Fall.

 **Syllabus**       **Schedule**

Two red double-headed arrows are overlaid on the image. One arrow connects the 'Overview' heading to the 'Syllabus' link. The other arrow connects the 'Overview' heading to the 'Schedule' link.

# Lectures

- Live lectures 3:00 - 4:30 PM ET every Monday until 04/22/2024
  - Will be recorded and uploaded to the website
- Every lecture jupyter notebook already has a pre-recorded lectures on the website.
  - I will be updating these with newer versions the week before each lecture.

# Problem Sessions

- One hour to work on problem sets in small groups
- Every Thursday 3:00 PM - 4:00 PM ET
  - Will not be recorded
- TAs will rotate between groups to assist and observe
- Each problem session also has a “prep notebook” with prerequisite practice.
- First one 02/08/2024 will make sure you are set up:
  - Github access, pulling the repo
  - Ability to run notebooks
  - Review of basic python

# Math Hour and Office Hour

- Math hour is every Tuesday 10:00 to 11:00
  - This is *optional* extra content.
  - Check out the self-assessment in “week 0”. Try crash course if you need to brush up.
  - Will usually cover the mathematical underpinnings of techniques covered the following week.
- Office Hour is every Tuesday 11:00 to 12:00 and by appointment.
  - Ask anything about course content, projects, debugging, etc.

# Getting Set Up

- Clone the repository
- Be able to open a jupyter notebook

# The GitHub Repository

- Link can be found on the course website
- Contains all of the educational content for the boot camp

The screenshot shows the GitHub interface for the repository 'data-science-spring-2024'. At the top, there are buttons for 'Edit Pins', 'Unwatch' (2), 'Fork' (0), and 'Star' (0). Below this, the repository name is followed by 'Private'. The main navigation bar includes 'main' (selected), '1 Branch', and '0 Tags'. A search bar 'Go to file' and buttons 'Add file' and 'Code' are also present. The repository description states: 'This repository contains the materials for the Spring 2024 incarnation of the Erdős Institute Data Science boot camp.' The 'About' section lists links for 'Readme', 'View license', 'Activity', and 'Custom properties', along with statistics: '0 stars', '2 watching', and '0 forks'. The 'Releases' section indicates 'No releases published' and provides a link to 'Create a new release'. The file list shows a recent update to 'README.md' by StevenGubkin, and a directory structure including 'bonus-content', 'data', and several 'week-' prefixed folders.

**data-science-spring-2024** Private

Edit Pins Unwatch 2 Fork 0 Star 0

main 1 Branch 0 Tags

Go to file Add file Code

**StevenGubkin** Update README.md 609403b · 1 hour ago 34 Commits

bonus-content	Assembling bonus content.	2 hours ago
data	Reorganized into weeks rather than lectures	2 weeks ago
week-01-introduction	Fixing a few bugs	yesterday
week-02-data-collection	Fixing a few bugs	yesterday
week-03-regression-I	Week 3 reviewed.	2 weeks ago
week-04-regression-II	Starting work on week 5	2 weeks ago
week-05-regression-III	Assembling bonus content.	2 hours ago
week-06-time-series-I	Week 6 ready to go.	2 weeks ago

**About**

This repository contains the materials for the Spring 2024 incarnation of the Erdős Institute Data Science boot camp.

Readme View license Activity Custom properties

0 stars 2 watching 0 forks

**Releases**

No releases published  
[Create a new release](#)



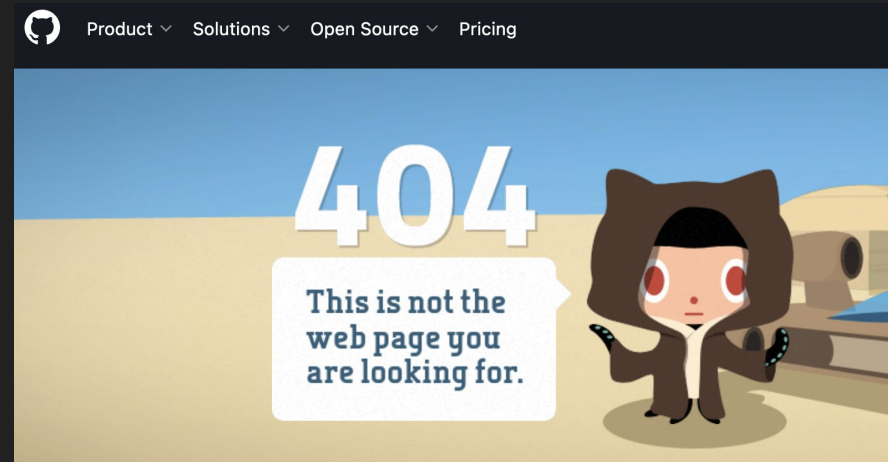
# The GitHub Repository - Steps

- Sign into your GitHub account
- Clone the repository onto your computer
  - Can find instructions in the “First Steps” section of the website
- Everyday of the boot camp you will need to “pull” the updates to the repository
  - Look for “Getting Started with GitHub” in the “First Steps” section of the website
- Can be useful to make a copy of the folder in your computer for your edits

# The GitHub Repository - 404 Issue

If you receive the 404 Error when clicking repo link:

- Check you are signed in
- Check that you have added your GitHub link to your Erdős profile
- Message Amalya about being added to the repository



# Jupyter Notebooks

- All educational content contained in jupyter notebooks
- Allows combination of Word document style text and python code
- Let's look at an example

# Jupyter Notebooks - Getting Set Up

- Follow Step 3 Under “First Steps” on website
- Lots of options:
  - Visual Studio Code: I use this one
  - Jupyter Notebook
  - Anaconda Navigator
  - Many other options

Questions & Concerns?