HPC in the City: Pandemics

SC23
Denver, CO | i am hpc.

Team Progress and Technology

November 5, 2023

HackHPC
Hackathons and Codeathons

HTTPS://HACKHPC.GITHUB.IO/HPCINTHECITY23

The University of Texas at Austin
Center for Pandemic Decision Science

STAR PARTNERSHIP PROGRAM

TACC TEXAS ADVANCED COMPUTING CENTER

Omnibond Engineering • Trust • Identity

VOLTRON DATA

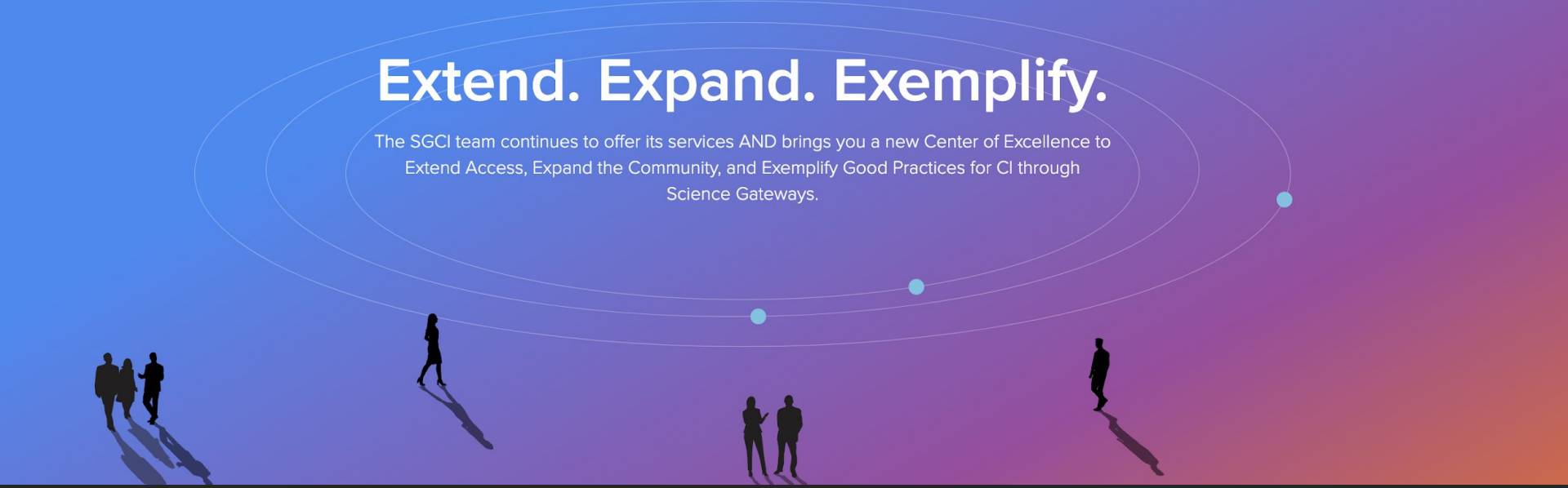
SGX3 Extend. Expand. Exemplify.

SGX3

Extend. Expand. Exemplify.

A Center of Excellence to Extend Access, Expand the Community, and Exemplify Good Practices for CI through Science Gateways.

Award #2231406



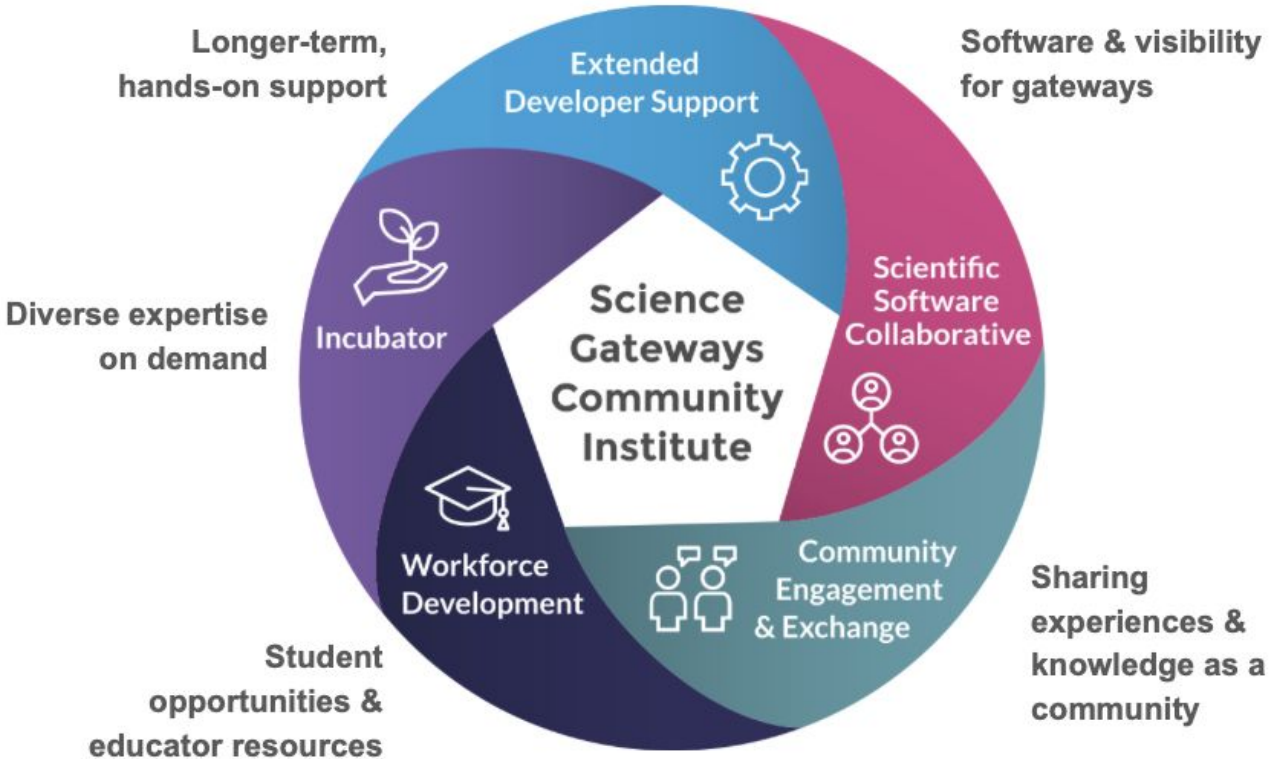
2023 Hack HPC in the City: Pandemic SGCI/SGX3 Update

Linda Bailey Hayden, Co-PI
LBHAYDEN@ECSU.EDU

NSF awards
1547611
2231406

SGX3 SGCI

Science Gateways Community Institute



Leadership Team

Michael Zentner
Director

Claire Stirm
Project Manager
Incubator Lead

Maytal Dahan
Scientific Software Collaborative Lead

Sandra Gesing
Community Engagement Lead

Linda Hayden
Workforce Development Lead

Nancy Maron
Sustainability Blueprint Factory Lead

Paul Parsons
User Experience Consulting Lead

939 webinar attendees

2056 participants in SGCI's events

504 student & **41** faculty participants

154 letters of collaboration

240 Focus Week Attendees

13 Affiliates

32 success stories

63 Consultations

14 partners

48 Extended Developer Support projects

$1,384,325 additional funding to SGCI by external projects

600 Gateway Catalog entries

SGX3 | SGCI

SGX3

Extend. Expand. Exemplify.

A Center of Excellence to Extend Access, Expand the Community, and Exemplify Good Practices for CI through Science Gateways.

Award #2231406

SGX3
SGCI
A Center of Excellence and a Software Institute to serve the Science Gateways Community
Helping…
• Community Focused
• Workforce Focused
• Future Focused
• Working Toward Preparing for the Future
Building…
• Development Service Focused
• Operations Service Focused
• Heavy Touch Consulting
• Working Toward Self-

SGX3

A new Cyberinfrastructure Center of Excellence for Science Gateways

SGX3 – NSF Funded

- ✓ Community building activities
- ✓ Workforce development activities
- ✓ Light-touch consulting / advisory services
- ✓ Envisioning the future through Blueprint Factories
  - ✓ ACCESS
  - ✓ PATh
  - ✓ Materials Genome Initiative
  - ✓ Sustainability practices
  - ✓ more to come…

SGCI – Client Funded

- ✓ Heavy-touch consulting / other services
- ✓ Software team augmentation / outsourcing
- ✓ Professional science gateway operations

SGX3 is $7.5 million over 5 years beginning September 2022

Workforce Development Overview

https://sciencegateways.org/faculty-focused

SGX3's workforce development activities contribute to broader impact by enriching existing and forming new relationships with minority serving institutions and organizations to bring gateway development into curricula, bring domain-specific gateways to relevant classrooms and research settings, and train faculty to scale these efforts to grow and live beyond SGX3.

The SGX3 Faculty Program builds a supportive HPC/Gateways community for the faculty while providing them the training and support needed to succeed. SGX3 staff assist faculty in establishing HPC accounts for their classes and consult with them through the implementation phase of their curriculum changes.

- HPC/SG Curriculum Enhancement Efforts
- Faculty workshops at ADMIUSA.ORG Symposiums
- Faculty Hackathons
- Faculty Poster Session at Gateways conference
- Gateway Community mentors assigned to faculty

SGX3 | SGCI

Partnership with ADMIUSA.org has been a key ingredients to being successful.

2023 ADMI Symposium

SGCI /SGX3 involvement included:

Faculty Session:
Hackathon HPC Education
Charlie Dey, *Director, Training and Professional Development*
Je'aime Powell, *Sr. Systems Admin.* TACC

Student Workshop
HPC and Science Gateway Opportunities
Charlie Dey, *Director, Training and Professional Development*
Je'aime Powell, *Sr. Systems Admin.* TACC
Faculty Session:
Initiative for HBCUs/MSIs
Dr. Elijah Maccarthy, HPC Engineer
Systems Acceptance and User Environment
Oak Ridge National Laboratory

Student and Faculty Session: *The ACCESS Program: Research Computing Resources for All* **Ms. Virginia Do**, *Outreach Manager & SIParCS Internship Director,* NCAR

SGX3 SGCI

2023 Coding Institute & Hackathon

- Sixteen students participated in the virtual 2023 Coding Institute. All were computer science majors. Weeks one and two of the Coding Institute focused on building non-technical and basic technical skills. Week three was devoted to specific gateway technology led by TACC. Finally, week four involved team projects via the hackathon.

- The Hackathon was co-sponsored by SGCI/SGX3, Omnibond Systems, Texas Advanced Computing Center and Amazon Web Services June 26th - 29th. All team projects focused on using UX Design techniques to revamp ADMIUSA.org and the HACKHPC.org sites. SGCI staffer Ali Baigelenov abaigele@purdue.edu served as a consultant and judge for the event.

Ali Baigelenov

Professional development seminar speakers: Dan Dietz, Suzanne Prentice and Jacqueline Jackson.

SGX3 SGCI

Student Programs. Hackathons, professional development seminars, and coding institutes that have a focus on participants from traditionally underrepresented populations will be continued from SGCI.

- Coding Institute June 5-29, 2023
- ADMI Symposium April 13-15 , 2023
- ADMI Hackathon June 26 -30, 2023
- Gateways Conference Mentors
- Rising Stars Award
- HPC in the City Hackathon@ SC
- Internships at TACC

SGX3 | SGCI

SGX3 Internships
https://sciencegateways.org/internships

Each year, our Workforce Development team offers summer internships for students interested in developing their gateway development skills. Interns are placed at the Texas Advanced Computing Center (TACC).

Eligible applicants include graduate students majoring in computer science or computer engineering (or related fields). The student will be funded by SGX3 to join the TACC science gateway team for the summer, working on live, impactful gateways.

Stipend

Participants will receive a $5,400 stipend (scholarship), housing (if not local to Austin), and meal card at The University of Texas at Austin, and travel arranged by TACC.

Additionally, travel grants to present research at an annual conference will be available to selected participants.

SGX3 | SGCI

Intern Presenters: (L to R) Jackson, Dhanny, Swathi, Prithul, Steven

- **Dhanny Indrakusuma** - working on Tapis with the Cloud and Interactive Computing team on creating a machine learning hub application that aims to enhance the experience of non-technical individuals involved in machine learning research. **Dhanny will continue his work at TACC funded by TAPIS.**
- **Jackson Thetford and Steven Oh** - working on the SCOPED (Seismic COmputational Platform for Empowering Discovery) project, an organization that advances research for seismic analysis to create custom Tapis applications using the Tapis UI infrastructure.
- **Prithul Sarker** - working with the web mobile applications team and the project primarily focuses on the backend operations of applications in high-performance computing.
- **Swathi Vallabhajosyula** - working with the Tapis team on extending the platform to include microservices to profile applications for resource consumption and recommend walltime.
- Jackson and Steven (undergraduates) will be funded by an NSF Scoped project and will be working with Ian Wang at UT Austin and TACC staff to continue the project they have been working on.

SGX3 SGCI

Theme Song: Welcome Back

Coreen Mullen

Xinyi Miao

Evans Etrue Howard

REDWARN

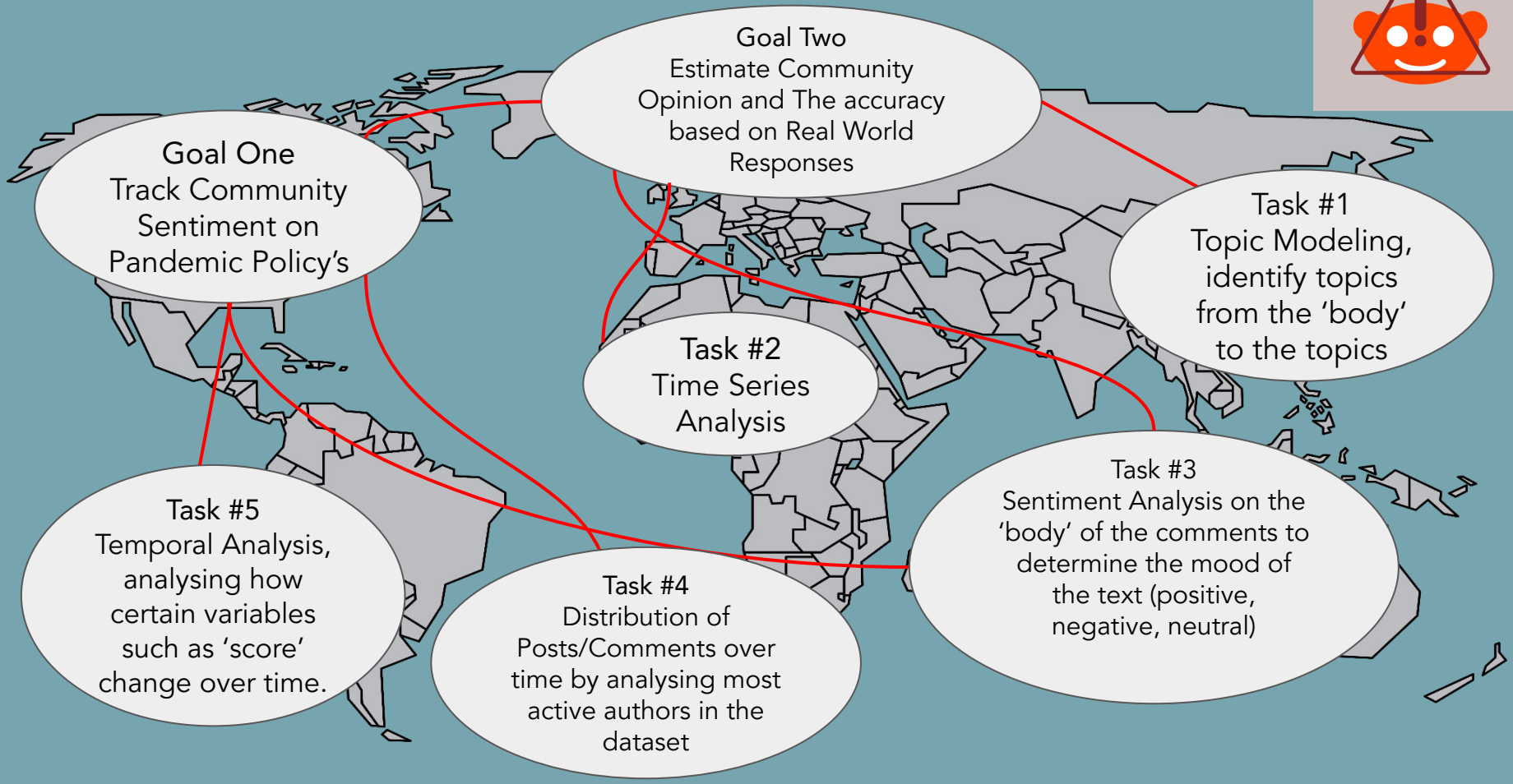
Mentors

- Emily Javan
- Oluwasegun Ibrahim
- Lydia Fletcher

Qimora Mason

Ahmad Samyono

REDWARN (Reddit Data for Early Warning and Response to Pandemics)



Project Plan

Evans

- Github Lead
- Code Collaboration
- Data Analysis
- Task #3 :

- Analyse frequency of posts
- Distribution of posts over time by authors
(5hrs)

Qimora

- Code collaboration
- Analysing data
- Task #1

- Clean “text” column
-Handle Missing Values
(5hrs)

Ahmad

- Analysing Data
- Code Collaboration
- Task #4

-How semantics change over time
-Study how the semantics changed after policy changed
(5hrs)

Xinyi

- Code Collaboration
- Visualisation
- #Task #5

-Prepare text for topic modeling
- Topic correlation
(5hrs)

Coreen

- Poster and Slide creation
- Code Collaboration
- Analysing the Data
- #Task 2

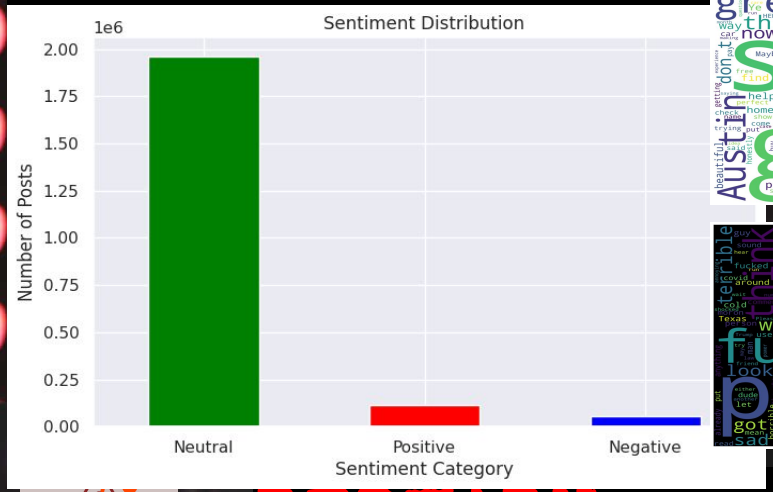
-Aggregate Sentiment scores
(5hrs)

Hackers

Mentors

Technology Used:
Google Colab.
Python and R.

Bottlenecks/Issues:
We had to use Eureka due to the size of the data, we pre-processed it then uploaded it to our Google Colab.



Status Updates

Task #1: Topic Modeling - 5 hours.
Task #2: Time Series Analysis - Finished.
Task #3: Sentiment Analysis - Finished.
Task #4: Distribution of Comments/Posts Over Time - 5 hours.
Task #5: Temporal Analysis - 5 hours.

—Analyzing the results is our next major task.
—Combining Time Series Analysis + Topic Modeling.

REDWARN (Reddit Data for Early Warning and Response to Pandemics)

TEAM RENDER MAGES

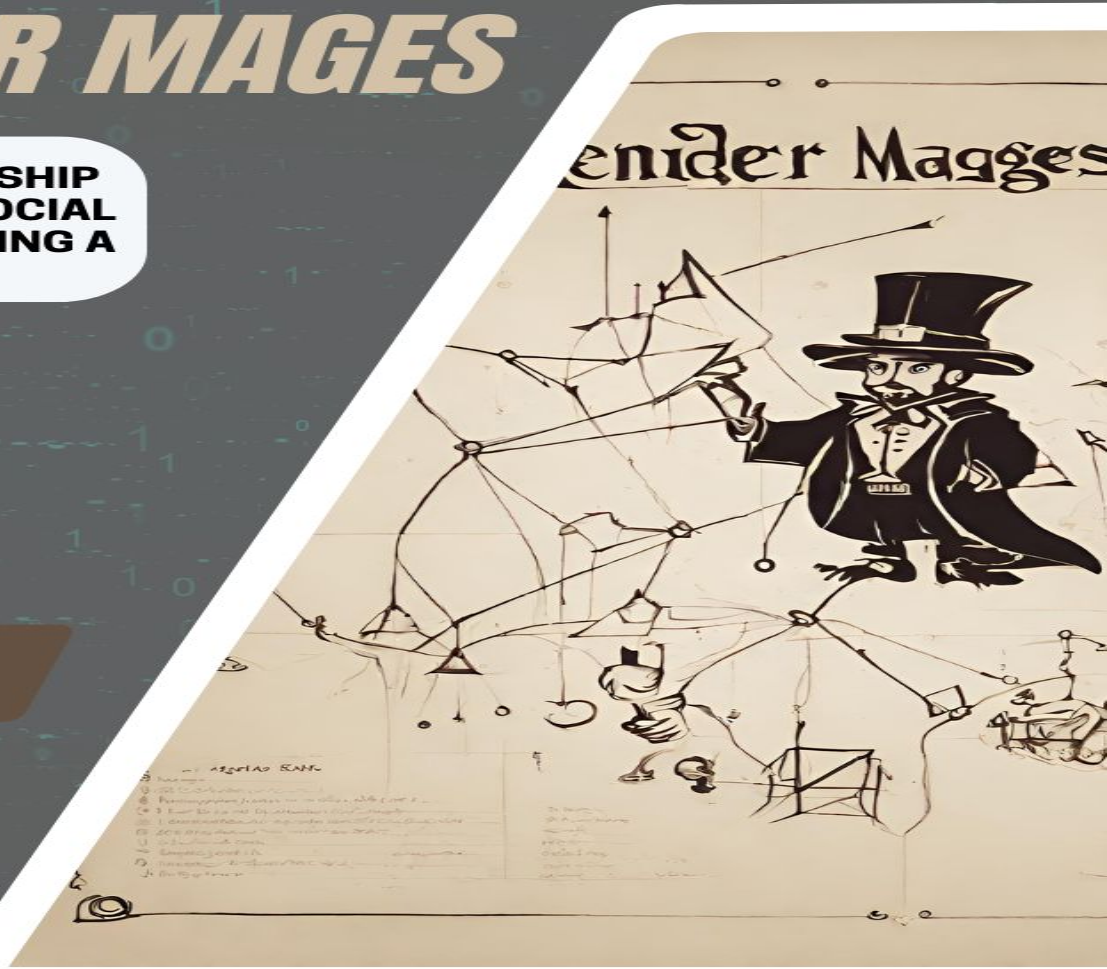
EXAMINING THE RELATIONSHIP BETWEEN MOBILITY AND SOCIAL VULNERABILITY INDEX DURING A PANDEMIC

GIDEON OSEI BONSU
JOSHUA HARRELL
CLARENCE CONNER
DANEISHA HARRIS
SUSAN GARZA

JOSE HERRERA

EMMA BUKOWSKI

Theme Music: Believer



Render Mages

Goals	▪ Explore the relationship between changes in mobility and social vulnerability score (SVI) in Austin, Texas ▪ Determine if this relationship is dependent on mobility restrictions
Tasks	▪ Select the appropriate time period ▪ Compare how mobility is different from selected dates to other times ▪ Evaluate the relationship between SVI and Mobility ▪ Verify the evaluations to show a general pattern ▪ Review calculations to improve results ▪ Visualize results

MEET THE TEAM

JOSHUA HARRELL
GitHub Operator

Susan garza
Poster Lead

GIDEON OSEI BONSU
Coding Lead

Clarence Conner
Documentation Lead

DaNeisha Harris
PowerPoint Designer

JOSE HERRERA
Mentor

Emma Bukoswki
Mentor

Render Mages status update

Completed
(Sub goal)

- Select time period (Susan & Joshua)
- Define areas of Austin by north, east, south, west, or central (Susan)
- Locate zip codes with inpatient admission facilities (Susan)
- Bar graph: mean SVI score per area of city

In Progress
(Sub goal)

- Compare how mobility is different in years 2019 compared to 2020 (Clarence, Gideon, Joshua)
- Slide presentation (Da'Neisha & Susan)

Needs to be completed
(Sub goal)

- Implement linear regression (C, D, &J)
- Evaluate for patterns (All)
- Optimize calculations (All)
- User Interface and testing (J, D, S)
- Poster (S &D)

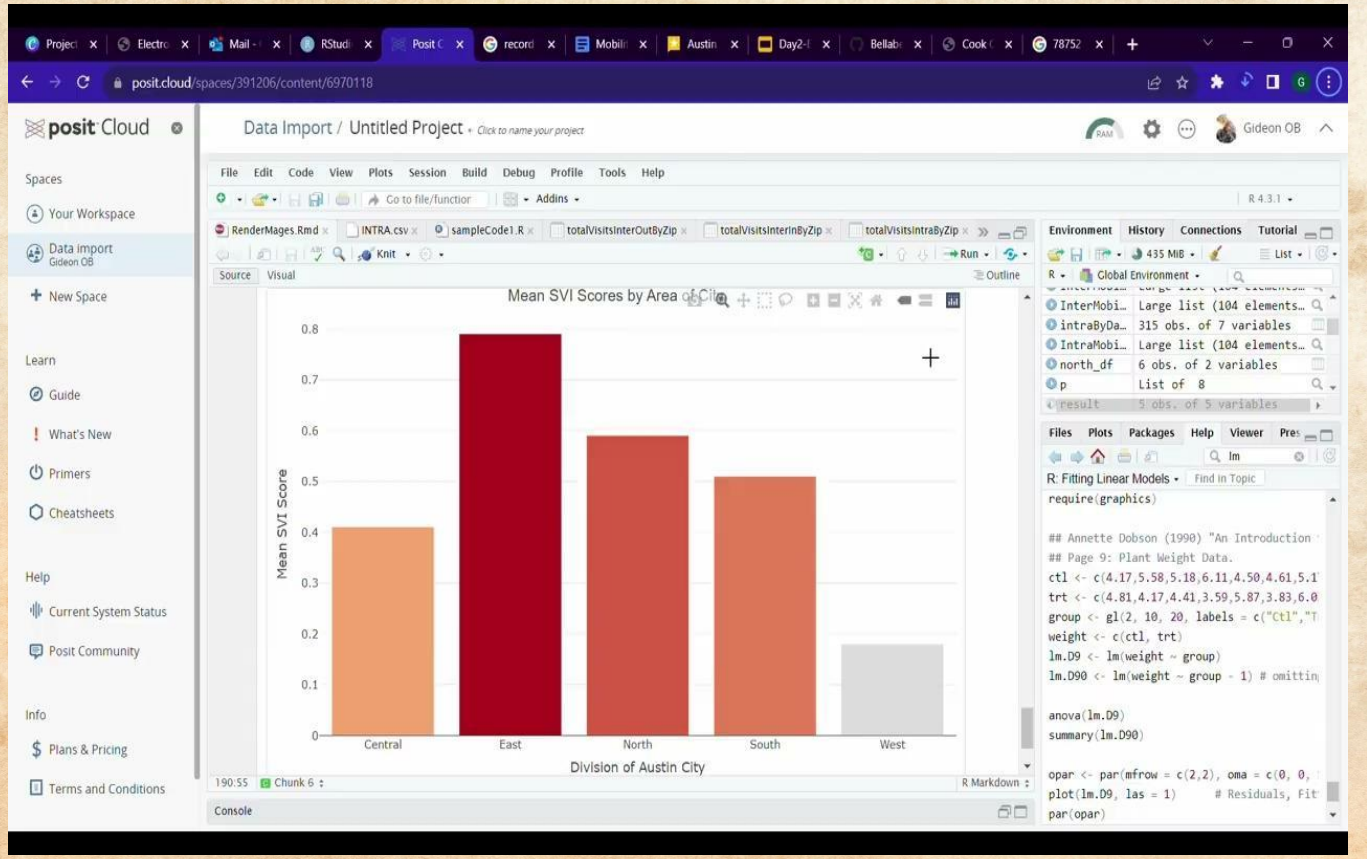
Technology Used:

- R & R studio
- Google Maps
- Github
- Plotly
- Dplyr

Bottlenecks/ issues

- R and Rstudio is not familiar to us yet, to have the ability create maps Interactive Choropleth & Scatter Maps with Plotly

SVI Score Based on Area of Austin City



MASSY SITUATION:
Mobility Data Analysis

Team name: Party Animals

Mentor: Kelly Gaither
Co-Mentor: Gladys Chen
Hackers: Leah Monet Morgan, Yamonta Gaines, Michael Olubode, Lisa Phan, Alex Gutierrez

Theme_song: WE ARE ONE

Party Animals Goals and Plans

- **Overall Party Goal: Enhance public safety and situational awareness by analyzing mobility data from Safegraph to identify and visualize mass gatherings that occurred from 2018 through February 2022.**
  - **1st Party Task:**

    Get a list of actual historical mass gathering events – date, location, and size
  - **2nd Party Task:**

    Find 1st Party Task events in the mobility data
  - **3rd Party Task:**

    Compare actual mass gathering events to representations in the mobility data.
  - **4th Party Task:**

    Analyze the data to identify recurring patterns and trends in mass gatherings, such as the frequency, size, and locations of events, in order to gain better understanding of the dynamics involved.
  - **5th Party Task:**

    Investigate the relationship between mass gatherings & superspreader events

REGGIE BALLESTEROS photography

Roles of the Party Animals

Primary personnel:

Visualization: Michael Olubode

Coding: Lisa Phan

Statistics: Yamonta Gaines & Alex Gutierrez

Github: Alex Gutierrez

Ground Truth Research: LeahMonet Morgan & Yamonta Gaines

Census Data Expert: Whole Group

Safegraph Data Expert: Whole Group

Shared Spaces:

- Documents/Presentations/Data
  - Google Drive
- Comms
  - Discord Channel #massysituation
- Source Code Repository
  - Github
    - Team Repo URL

Party task 5:

Understanding the relationship between mass gatherings and super-spreader events during Covid-19

+ Subtask 1: Looking at pattern of attendees leaving mass-gatherings (~ 0.5 day)

- Using census data, device usage, map the reported visit count to number of attendees (done)
- Install packages for this to run on TACC - not done

+ Subtask 2 : Map this pattern to covid spread from CDC (1 day)
- Step1: writing function to get date-time events

+ Subtask 3: Choropleth map visualization (~ 0.5 day)

Party Animals Party Goal:

Enhance public safety and situational awareness by analyzing mobility data from Safegraph to identify and visualize mass gatherings that occurred from 2018 through February 2022.

Task	Description	Name + Assignment	Progress/ Bottlenecks
1.	Get a list of actual historical mass gathering events – date, location, and size [Google Sheets]	LeahMonet + Yamonta + Alex: - Collect Data on Spreadsheet (Date, Location, approximate # of people)	- Researching real world events and inputting into spreadsheet
2.	Find 1st Party Task events in the mobility data [Python/R]	LeahMonet Yamonta + Alex + Lisa: Finding mass events based on # of visitors	-Successfully accessed Safegraph & Census Data
3.	Compare actual mass gathering events to representations in the mobility data.	Michael: Creating setup for Dashboard on JavaScript to load data when ready	
4.	Analyze the data to identify recurring patterns and trends in mass gatherings, such as the frequency, size, and locations of events, in order to gain better understanding of the dynamics involved.	Lisa: Step1: Go from sampled device information to an estimate of actual people present at event Step 2: Functions to query time specific events to match superspreaders to real-time data Step 3: Data visualization	-R Package Installation Error -Running on Cloud but not on Terminal -Tricky census terminology and mapping
5.	Understand the relationship between super-spreader events and mass gatherings during Covid		

REGGIE BALLESTEROS photography

QUESTIONS ??

Next Session:

- **DAY 4 FINAL EVENING CHECK-IN:**
  Team Progress
  Mentor Trailers
  [Monday, *11/6/23 @ 5:00pm CST*]

Schedule:

https://hackhpc.github.io/hpcinthecity23/schedule.html

The University of Texas at Austin
Center for Pandemic Decision Science

STAR PARTNERSHIP PROGRAM

Omnibond Engineering ▪ Trust ▪ Identity

TACC TEXAS ADVANCED COMPUTING CENTER

SGX3 Extend. Expand. Exemplify.

VOLTRON DATA