ECS 351L and 851L : Geospatial Data Science

Bonus lecture: Previous class projects

Email: jonathan.ryan@duke.edu
Office: 3101 Grainger Hall
Office hours: Mondays and Wednesdays 15:00-16:00

Week	Date	Lecture	Homework
Unit 1: Geospatial data analysis			
1	Aug 25	Introduction	Assignment 1
	Aug 27	Vector data	
2	Sep 1	No class (Labor Day)	Assignment 2
	Sep 3	Rivers of the world activity	
3	Sep 8	Network data	Assignment 3
	Sep 10	Residential flooding activity	
4	Sep 15	Gridded data	Assignment 4
	Sep 17	Census activity	
Unit 2: Geospatial machine learning			
5	Sep 22	Machine learning fundamentals	Assignment 5
	Sep 24	Wine activity	
6	Sep 29	No class (baby)	
	Oct 1	No class (baby)	
7	Oct 6	Neural networks	Assignment 6
	Oct 8	Penguin activity	
8	Oct 13	No class (Fall break)	Assignment 7
	Oct 15	Previous class projects	
9	Oct 20	Specialized neural networks	Assignment 8
	Oct 22	EuroSAT activity	

Assignment 7 (due Oct 22)

Now we have explored some different types of geospatial data (e.g. table, network, and gridded) in different contexts (e.g. urbanization, habitat loss, mobility), we should be ready to come up with an idea for a final project.

Assignment 7 (due Oct 22)

Now we have explored some different types of geospatial data (e.g. table, network, and gridded) in different contexts (e.g. urbanization, habitat loss, mobility), we should be ready to come up with an idea for a final project.

There is intentionally a lot of flexibility here. This can be an individual or group project, though we encourage you to consider a **group project** for two reasons. (1) it is more fun to work in teams, and (2) it will provide you with first-hand experience of collaborative software development in Github, a key skill for many careers.

Assignment 7 (due Oct 22)

Now we have explored some different types of geospatial data (e.g. table, network, and gridded) in different contexts (e.g. urbanization, habitat loss, mobility), we should be ready to come up with an idea for a final project.

There is intentionally a lot of flexibility here. This can be an individual or group project, though we encourage you to consider a **group project** for two reasons. (1) it is more fun to work in teams, and (2) it will provide you with first-hand experience of collaborative software development in Github, a key skill for many careers.

There are several options for choosing a final project:

- If you're already engaged in **independent research**, propose a project that will move that work forward
- Explore a **side project** that is related to your independent research
- Propose a **new research question** that you have been interested in for a while but never had the chance to investigate. Hopefully some of the concepts we've covered in class will help you carry out the project.
- **Join forces** with others in the class who already have a project idea.
- If independent research is new to you or you are still struggling, **talk to the instructor, TA, or grad students** in the class to define a suitable final project.

Final project report (due Dec 5)

The final report should be about 2000 words and contain at least the following sections:

- Introduction
- Methods
- Results
- Discussion/Conclusion

Final project report (due Dec 5)

The final report should be about 2000 words and contain at least the following sections:

- Introduction
- Methods
- Results
- Discussion/Conclusion

The report should include multiple figures and references, where appropriate. Please include a **link** to your **GitHub repository** in the final report. We will be thinking about the following questions when grading the final project reports:

Final project report (due Dec 5)

The final report should be about 2000 words and contain at least the following sections:

- Introduction
- Methods
- Results
- Discussion/Conclusion

The report should include multiple figures and references, where appropriate. Please include a **link** to your **GitHub repository** in the final report. We will be thinking about the following questions when grading the final project reports:

Goals: Did the project investigate an interesting topic? Was the motivation of the project clearly outlined?

Implementation: Did the project demonstrate effective usage of geospatial data? Were appropriate Python packages used? Did the project carry out geospatial analysis? To what extent did the project go beyond what we covered in the labs?

Reflection: Did the report include a thorough discussion of the strengths and weakness? Were future directions/improvements discussed?

Collaboration: Is there evidence of using GitHub to manage code development? If so, how many commits and pull requests? Did all members of the team contribute to the GitHub repository?

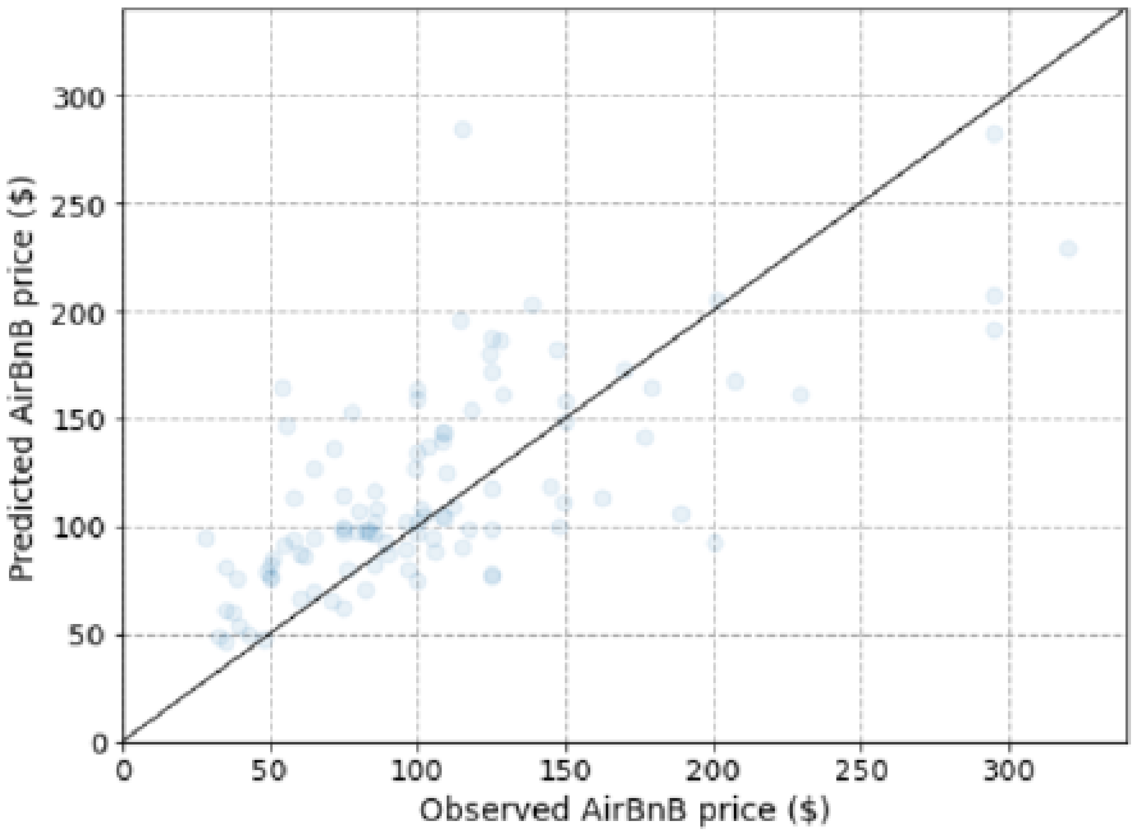
Presentation: Was the report logically structured and organized? Did it contain figures? Were the figures tidy and labelled properly?

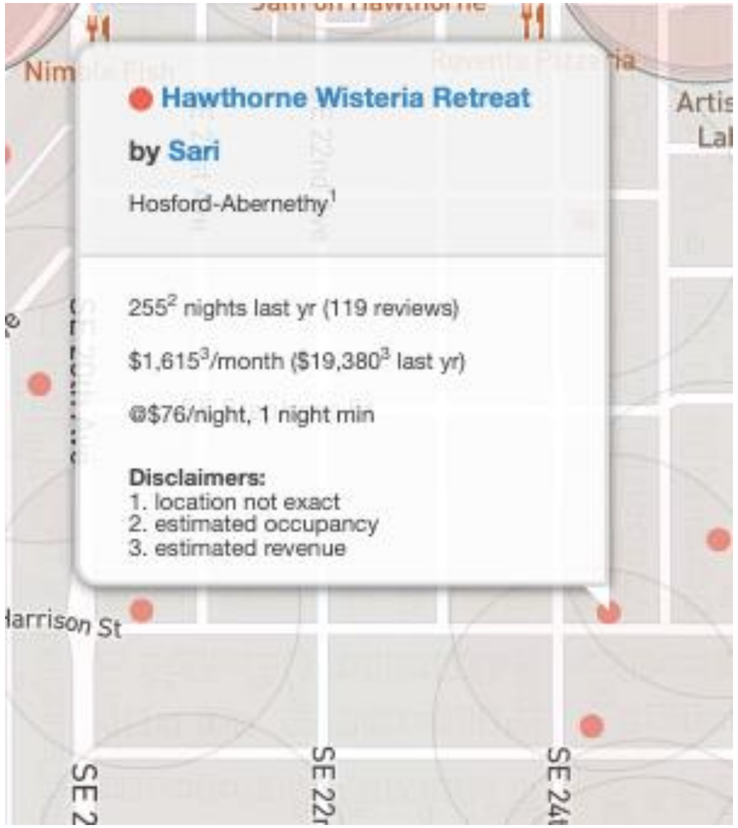
Final project presentations (week starting Dec 1)

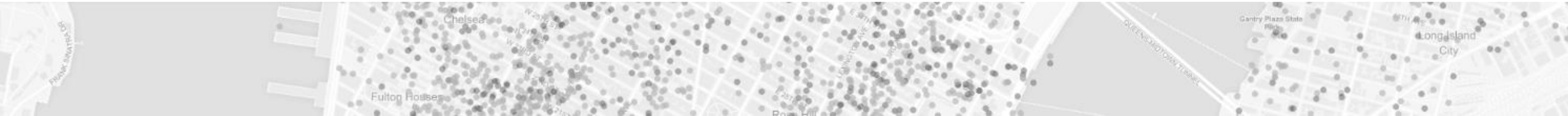
- 10-minute presentation with 2 minutes for questions
- 10 slides, one minute per slide
- Similar structure to final project reports

Machine learning AirBnB rates in Portland

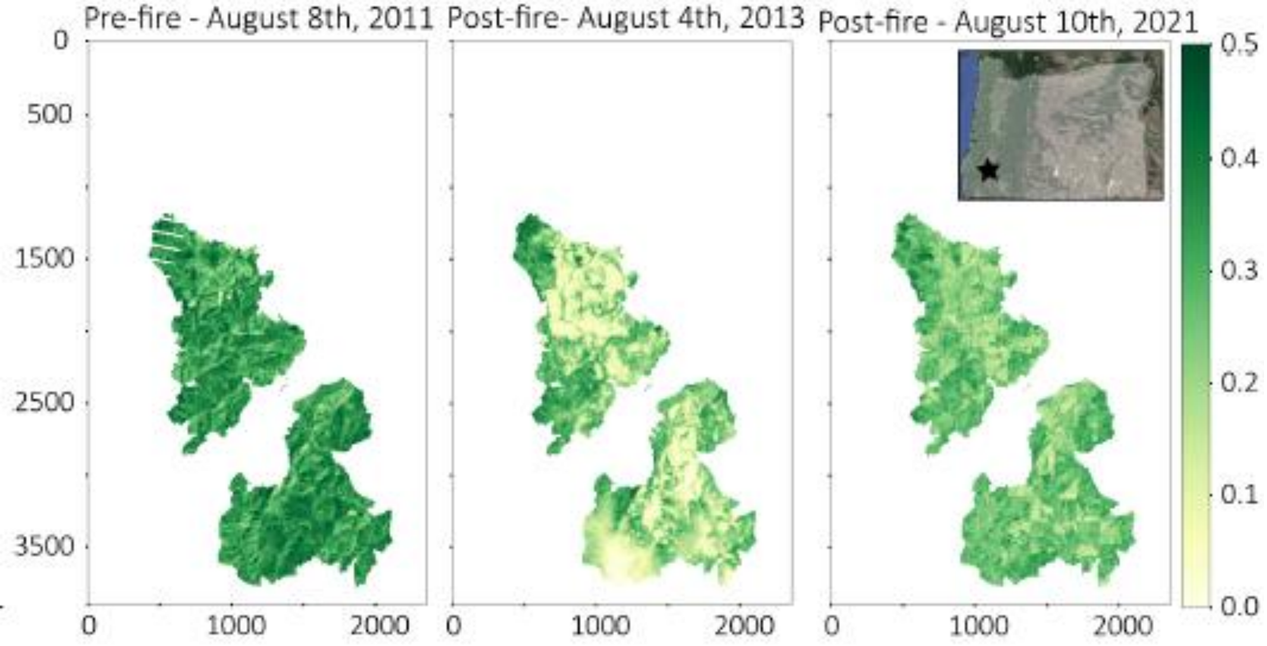
https://insideairbnb.com/

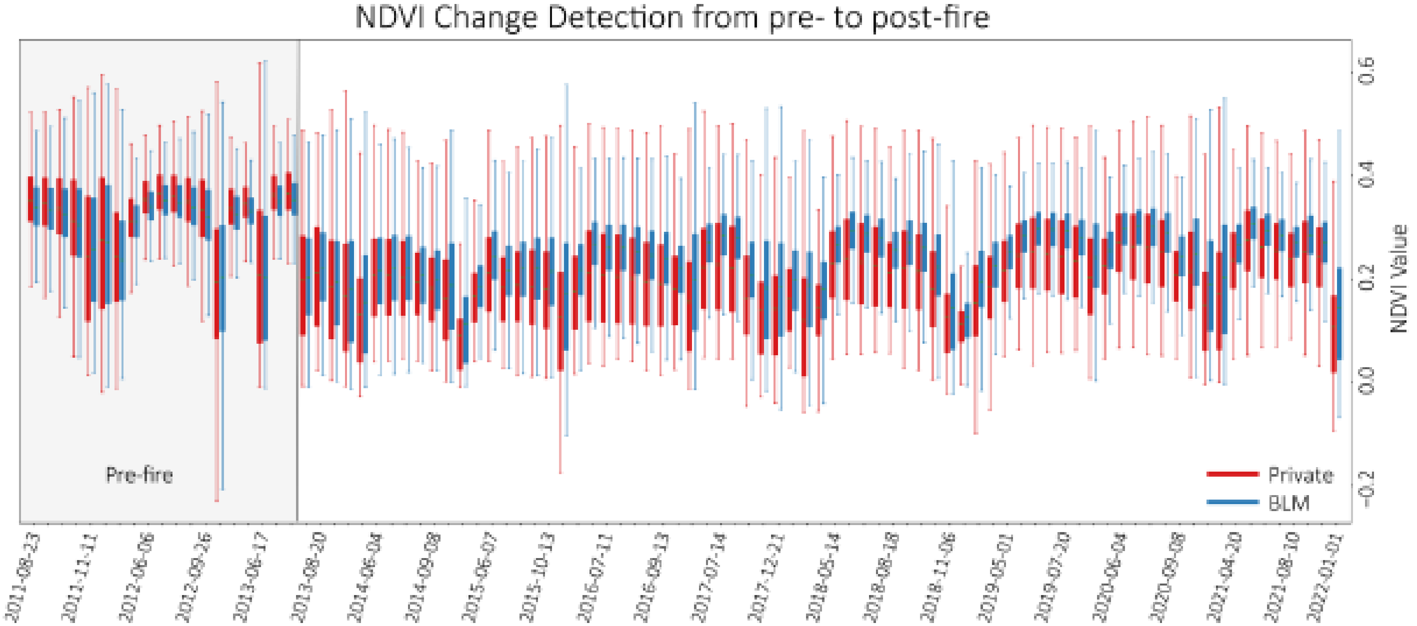






Post-fire Vegetation Recovery for the Douglas Fire Complex, Oregon

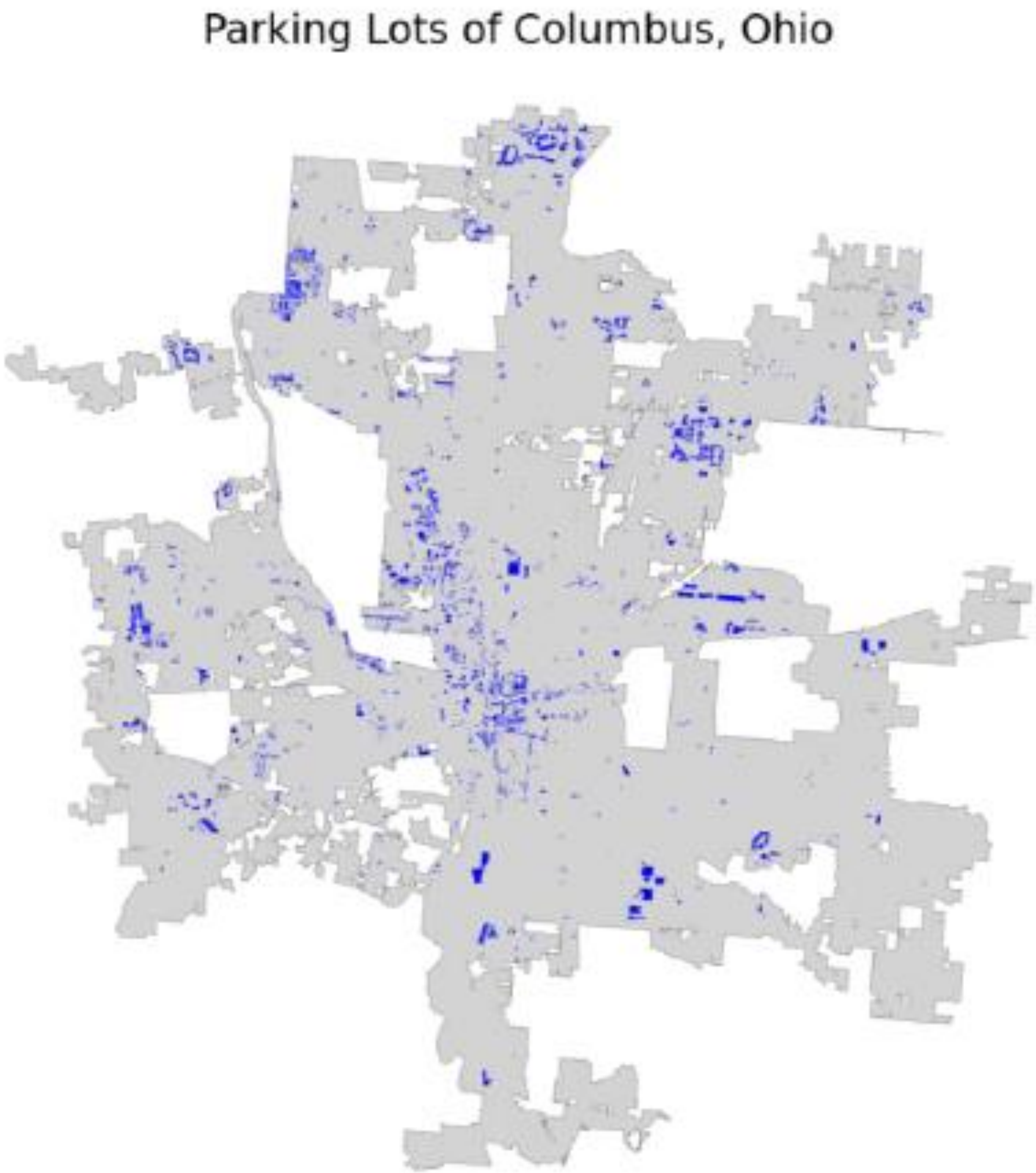




Parking lot surface area coverage in U.S. Cities

OSMnx to download geospatial data from OpenStreetMap that contained the tag “parking”

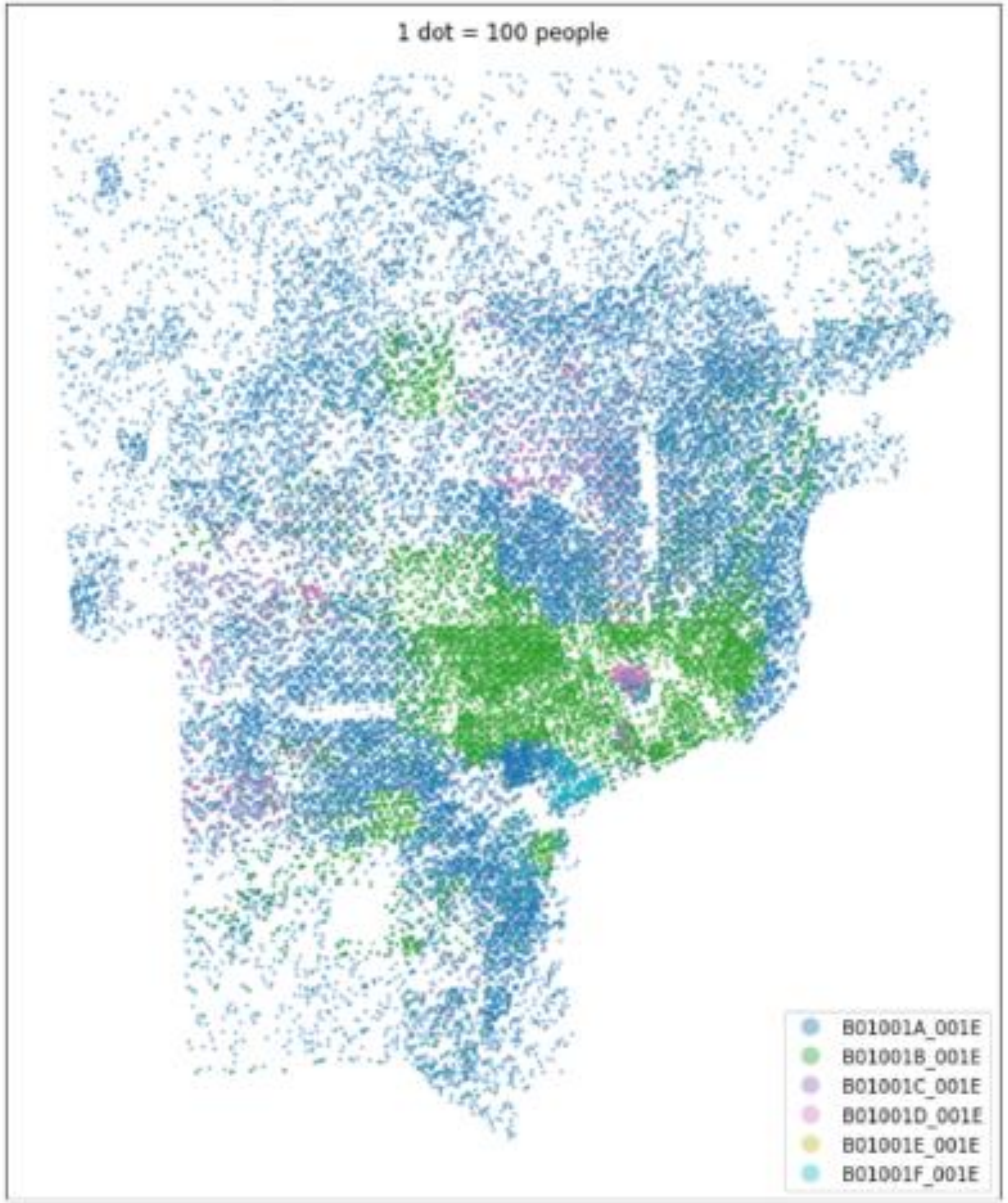
	City Name	Parking Lot Cover (%)
0	Indianapolis	3.477085
1	Jacksonville	0.422643
2	Seattle	1.103871
3	Houston	1.162967
4	Columbus	2.300840
5	Philadelphia	2.212235
6	Dallas	0.877751
7	El Paso	0.430330
8	Chicago	2.625216
9	Los Angeles	1.202591
10	Austin	1.732617

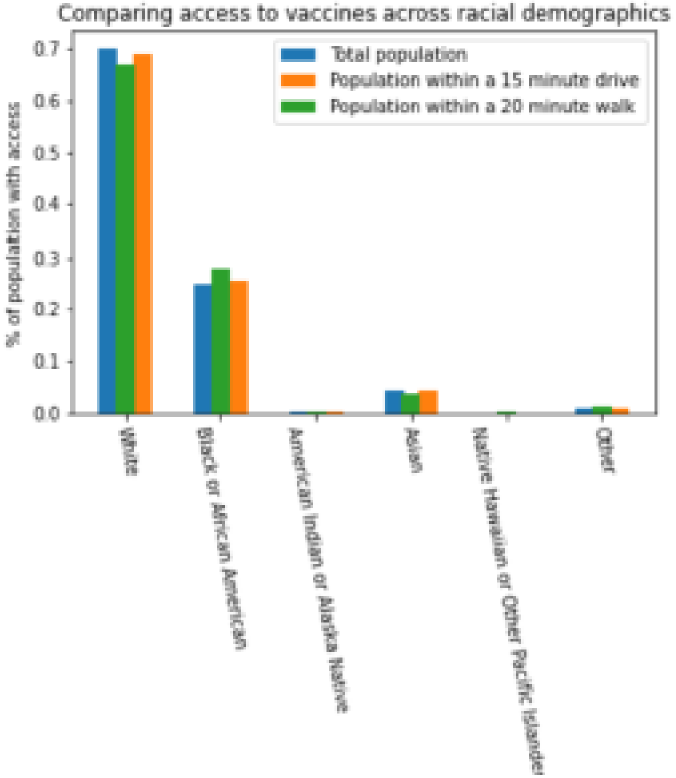


Spatial distributions of COVID-19 vaccine availability according to race

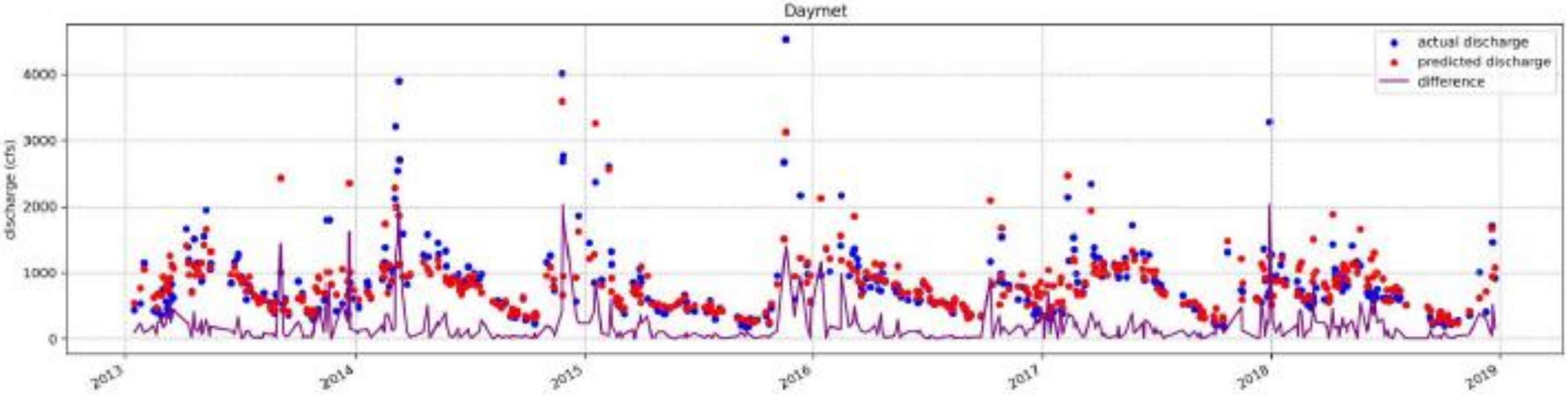
OSMnx to download street network data from OpenStreetMap to compute walking and driving distances to nearest vaccination clinic.

census package to download population and race/ethnicity data for Detroit, Michigan



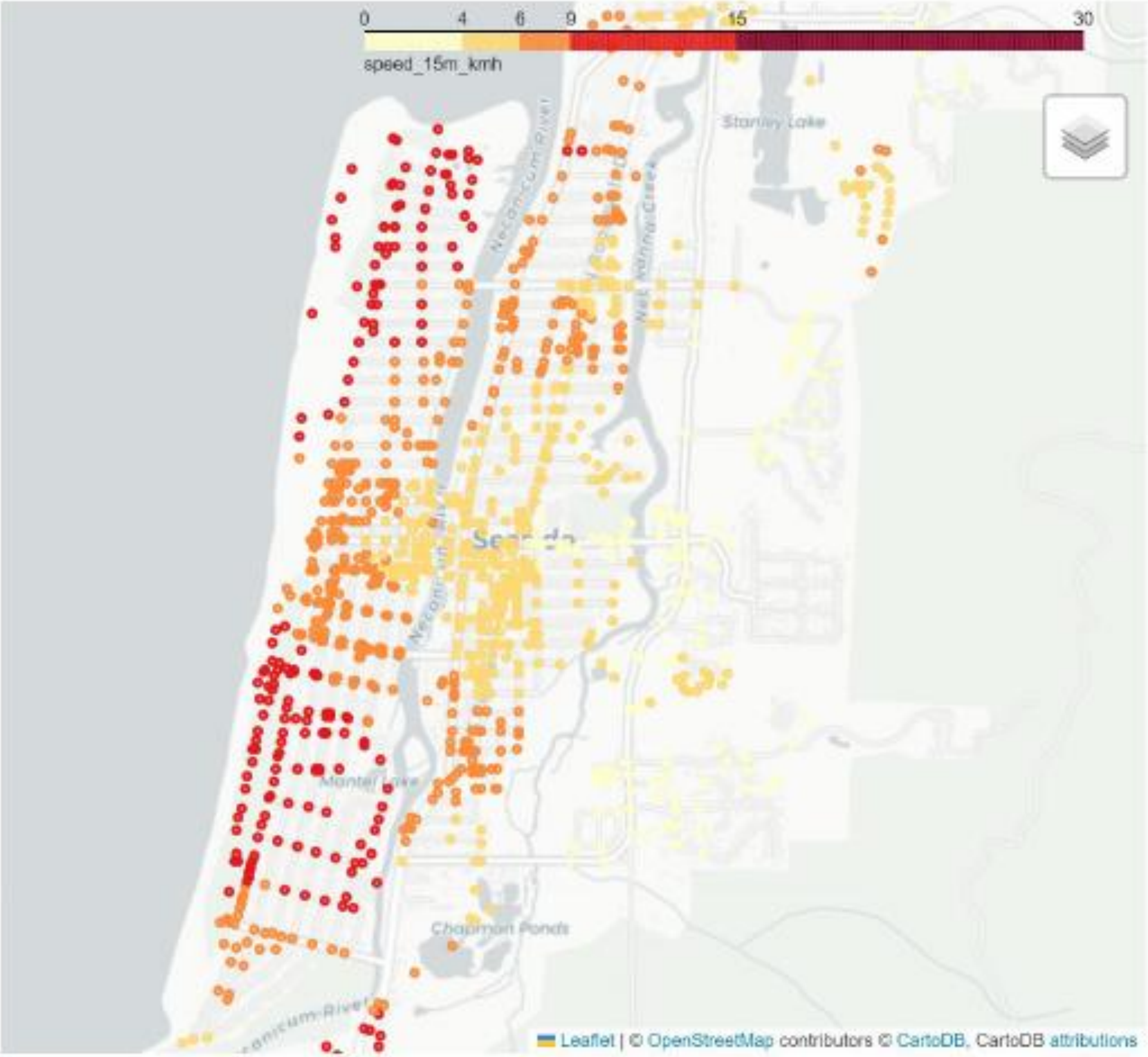


Machine learning the discharge of the Nisqually River

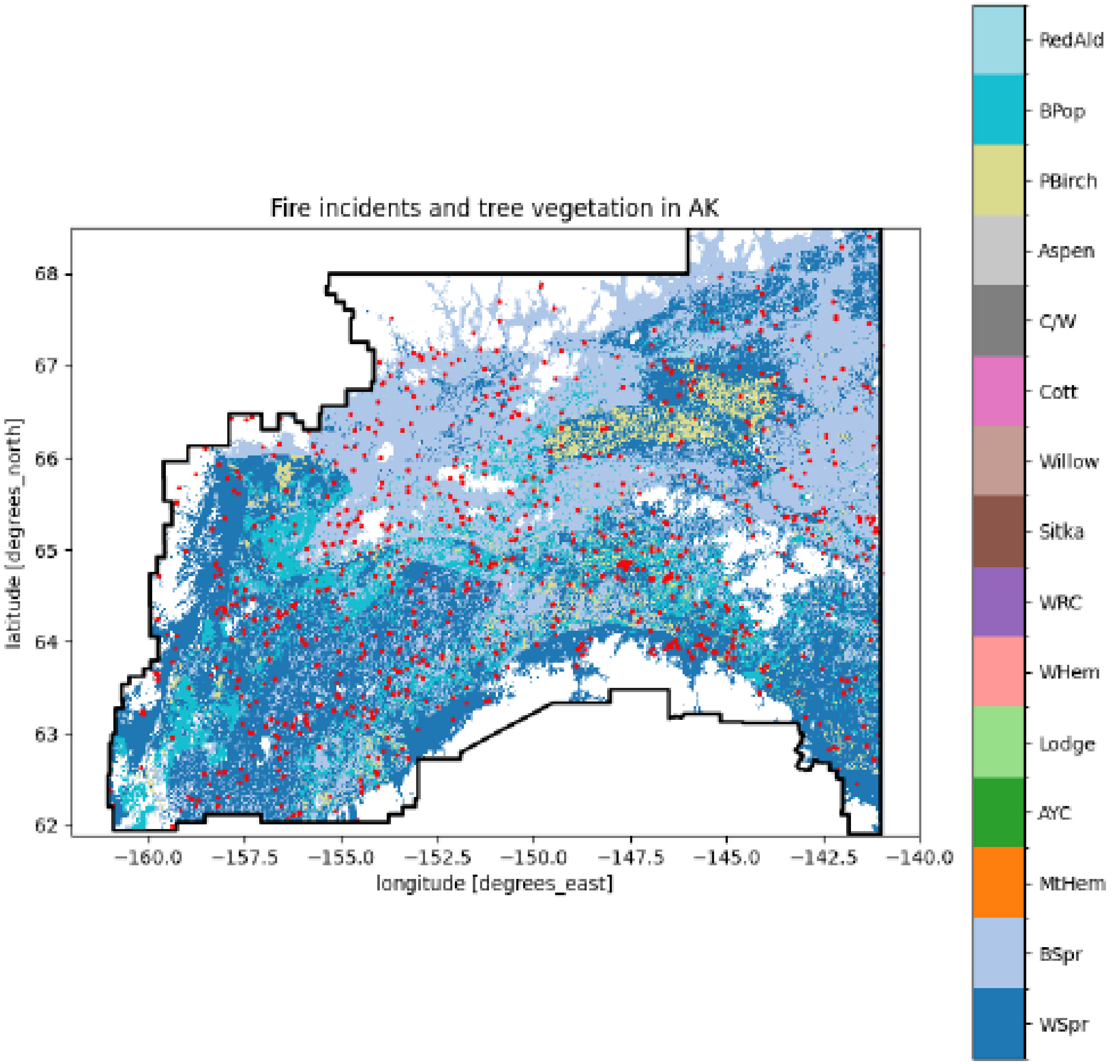


Evacuation times for the XXL Cascadia Tsunami

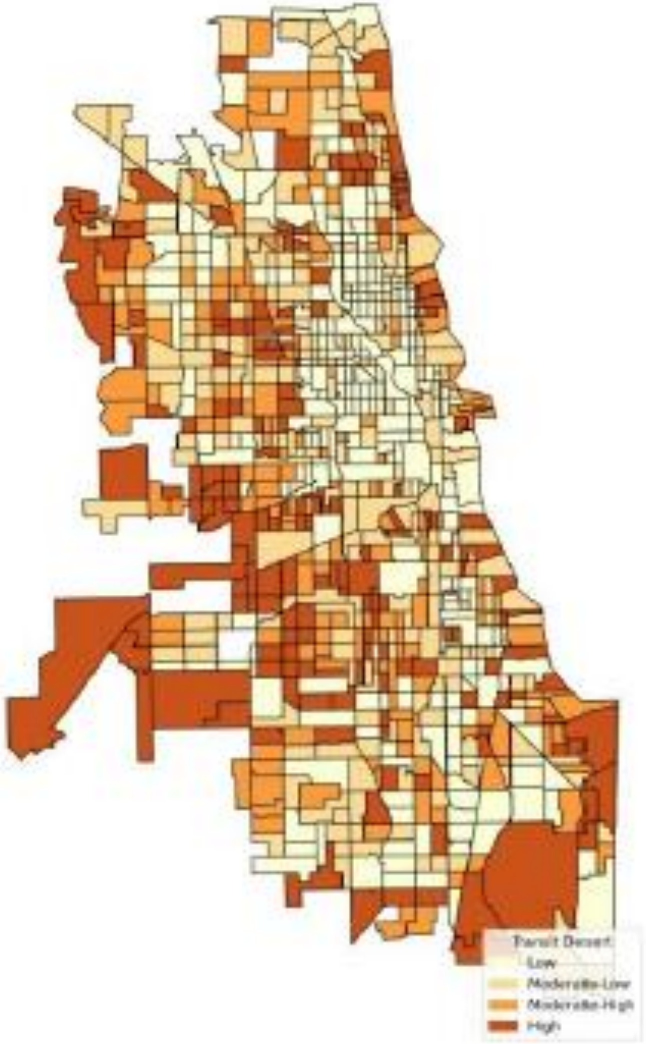
Seaside, OR

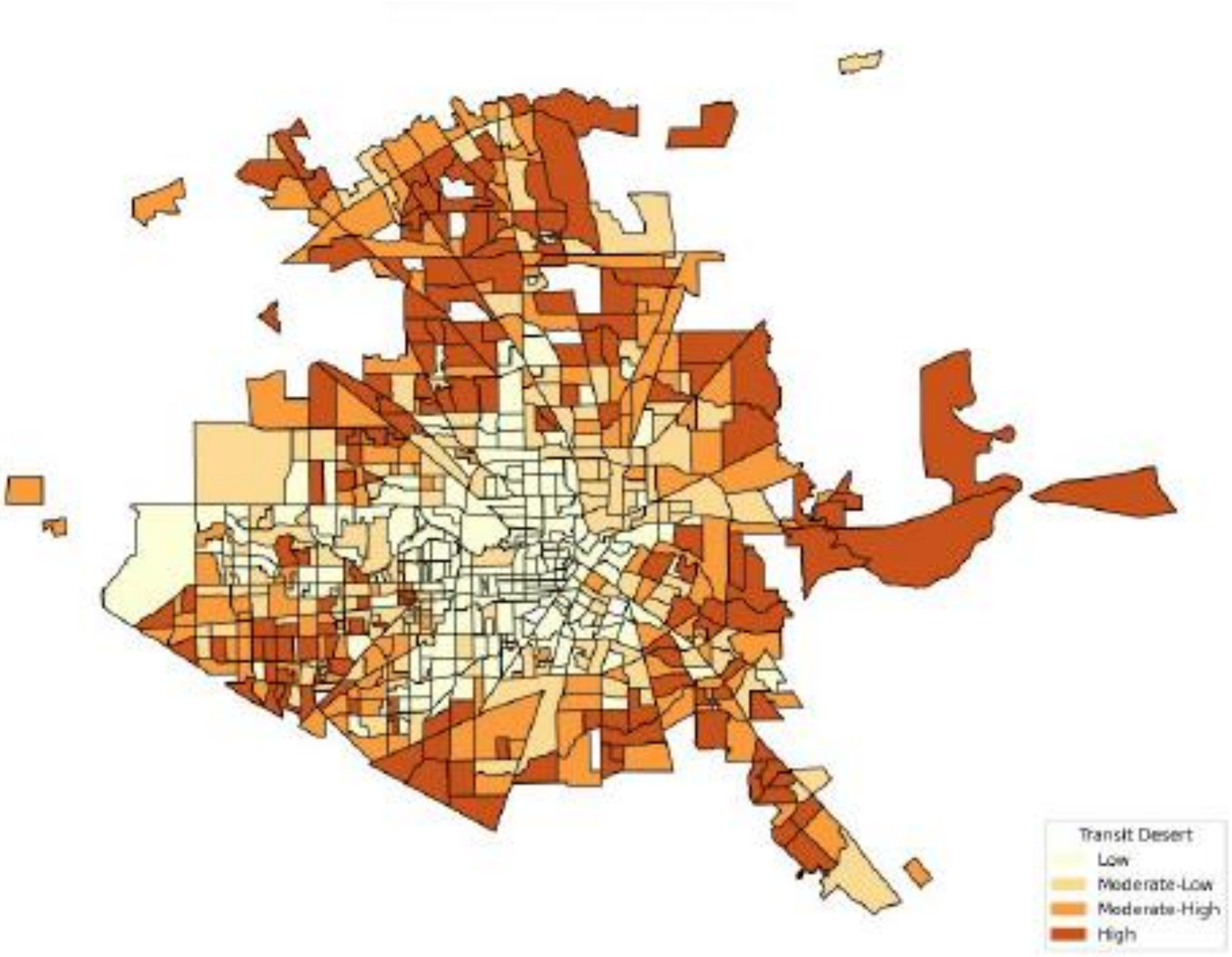


Predicting wildfire size in Alaska using machine learning



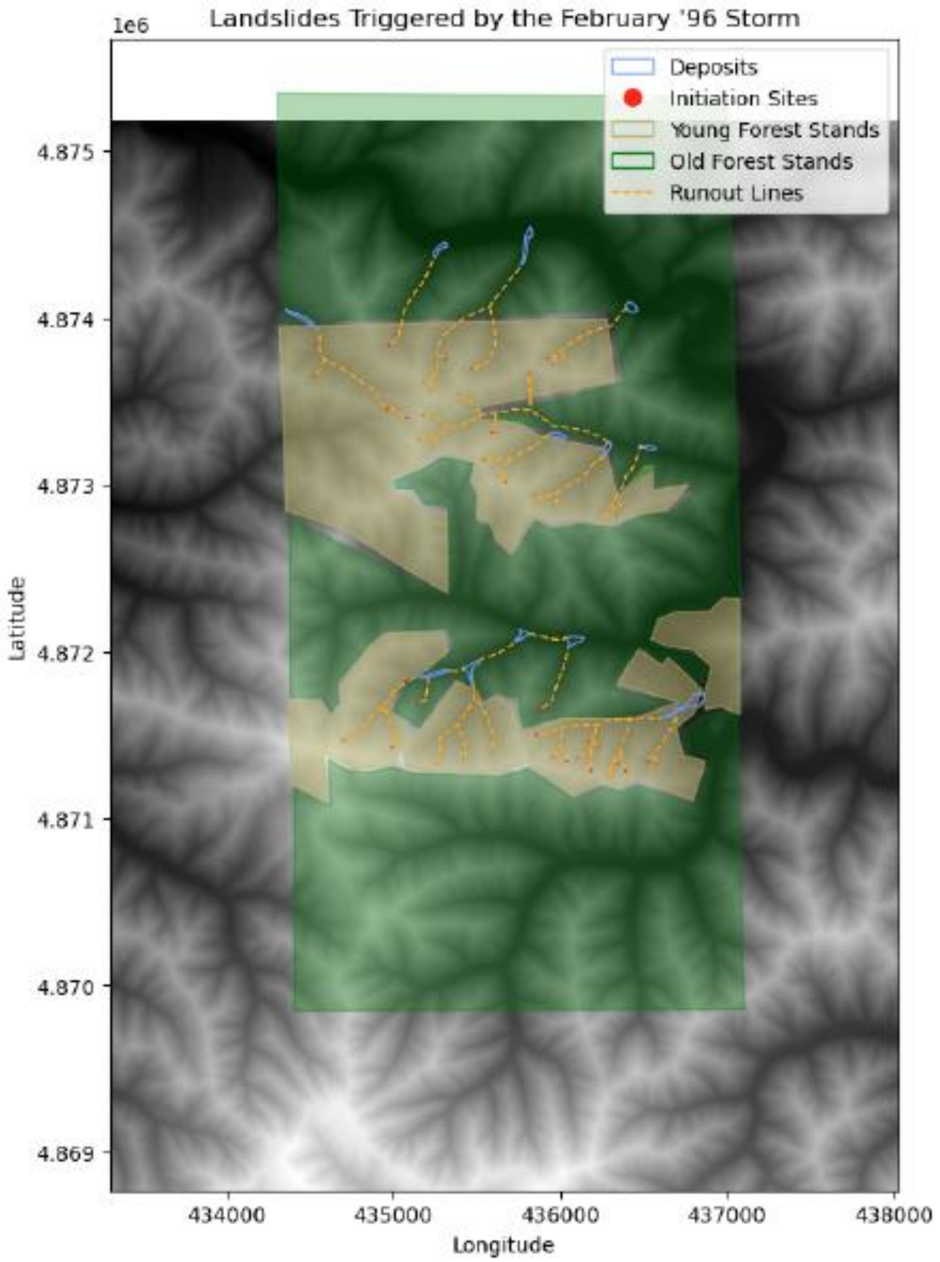
Identifying Transit Deserts: A Case-Study of Chicago and Houston



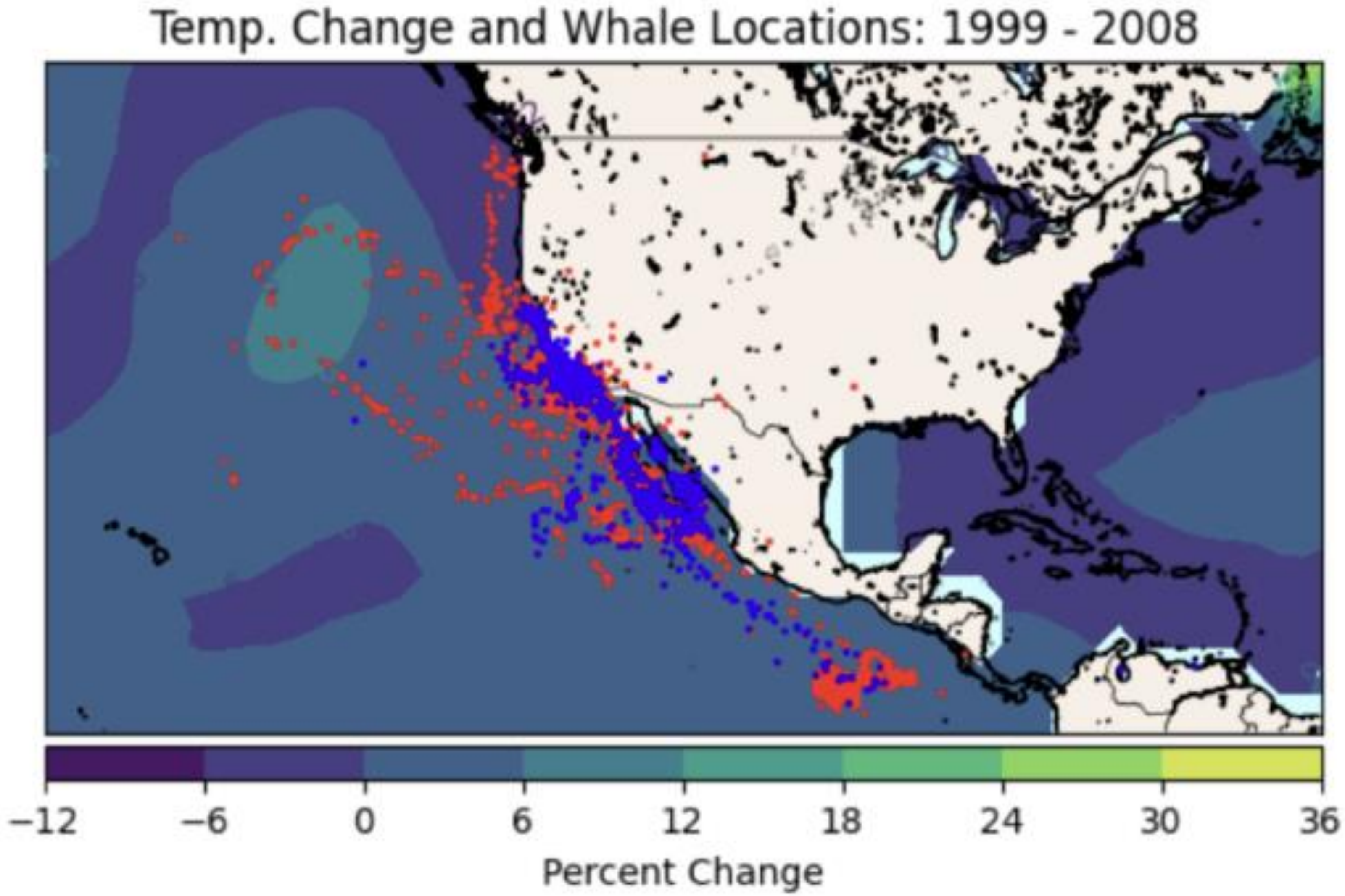


Transit desert = transit demand – transit supply

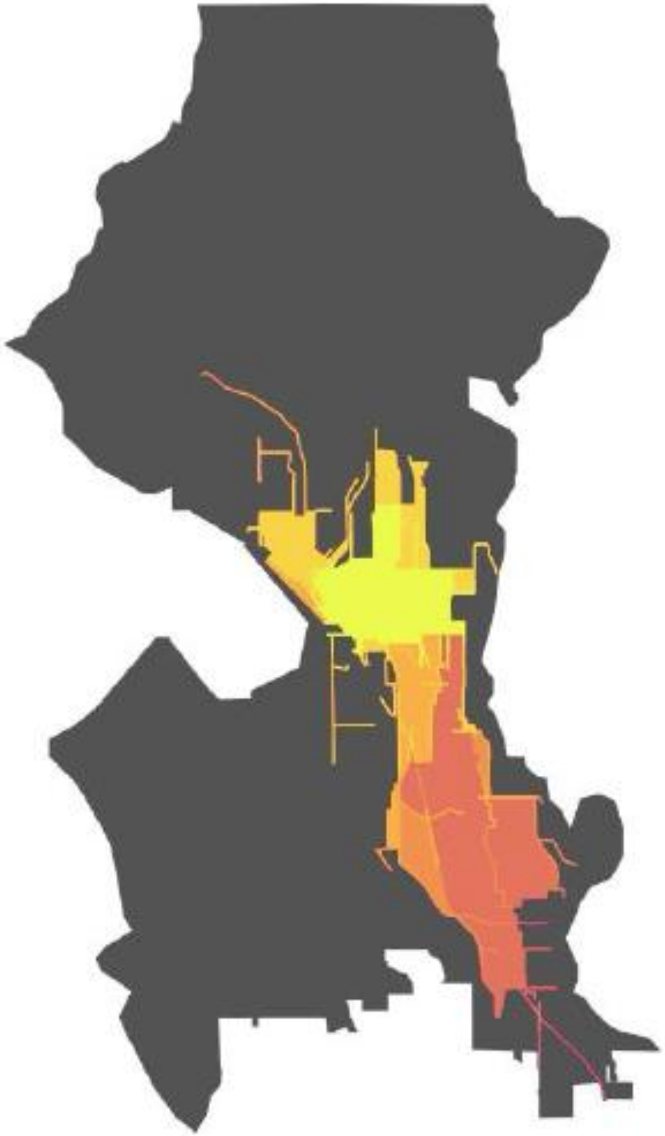
Vegetation influence on debris flow initiation in Oregon

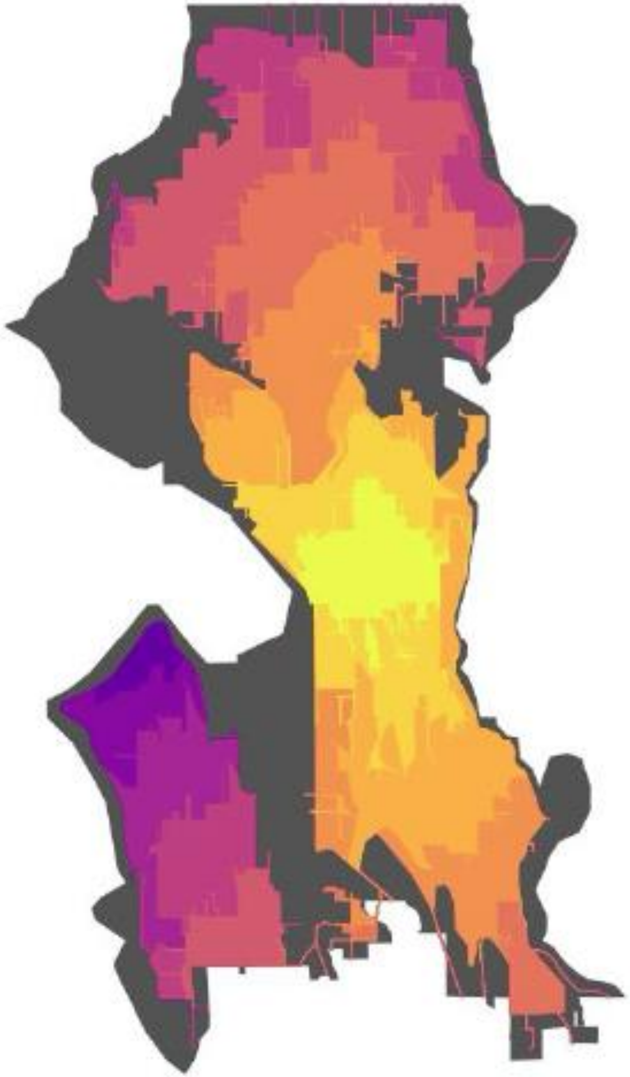


Blue Whale migration in the North Pacific Ocean between 1999 and 2008



Bike connection and accessibility in Seattle





Some takeaways

- Generate a clear and specific research question or goal

- Define the objectives you need to address the research question or goal

- Read broadly to find articles that have done something similar and either:
    - Adopt their methods to a new location or study site
    - Approach their research question or goal with new methods (e.g. deep learning)

- Examples of titles that are too descriptive and limited:
    - "*Temporal analysis of NDVI in Northern California*"
    - "*Land cover change in Bentonville, Arkansas*"

- Examples of titles that are too complex:
    - "*Modeling with Stale Data: Does Disaster Insurance Need an Overhaul in the Face of Climate Change?*"
    - "*Deep seated landslide kinematics*"

Next time: Specialized neural networks

Email: jonathan.ryan@duke.edu
Office: 3101 Grainger Hall
Office hours: Mondays and Wednesdays 15:00-16:00