

# Project Pitches!

DATA515 // 26wi // 2/3/2026

# Sarah: Identifying Hidden Flood Risk in Seattle

## Project: reusable analysis tool

Task: Compare the observed number of flooding reports for drainage basins in storm events to expected range learned from historical data, given rainfall, elevation, and drainage infrastructure.

- Use drainage basins, storm event pairs as the unit of data.

### Goal:

1. Detect where reported flooding is inconsistent with historical and physical expectations.
2. Build an interpretable model to examine the influence of key factors on flooding.

## Data

### Observed flooding:



#### Seattle Customer Service Requests

(Filter for flooding / drainage-related request types)

#### Seattle Infrastructure Data Layers, Sewer and Drainage (GIS)

- Rain Gages ([Rainfall measurements granular](#))
- SPU DWW Drainage Basins ([Where water flows / collects](#))
- SPU DWW Catch Basins / Inlets / Drainage Points ([Where surface water enters](#))

#### [Rainfall measurements city wide:](#)

#### NOAA Hourly Precipitation Data (Seattle)

#### [Elevations in Seattle:](#)

#### 2021 USGS Lidar DEM: King County, WA

# Player evaluation tool for basketball

Type: Research/tool

Goal: Identify potential prospects given market value, player characteristics, team needs

Data: basketball-reference, nba api

Ian Chang

# Examining IMDb Trends – Daniel Yan

Data: <https://developer.imdb.com/non-commercial-datasets/>

IMDb provides a list of datasets with a lot of information available on the site with metrics & information about media hosted. Let's examine how different genres' popularity changes & examine whether score inflation is an actual phenomenon on the website.

*Project Type*

Interactive Tool

## Data Sources

1. Medicare Part D Prescribers by Geography & Drugs (CMS)
2. Medicare Part D Enrollment by Demographics (CMS)



# Medicare Rx Equity Dashboard

## Short pitch

I am developing an interactive dashboard analysing Medicare Part D prescribing patterns across states (2018-2023).

It normalizes prescription rates by beneficiary demographics like age groups, low-income subsidy status, ESRD patients, and rural vs. urban divides for fair state comparisons.

Researchers filter by drug class (statins, antidiabetics) and population subgroups to explore interactive choropleth maps, time trend charts, and comparison tables. States colour-code by prescribing intensity: Texas red (high volume), California green (low), instantly revealing geographic disparities, preventive care gaps, and plan performance differences.

Name: Fatima Fazil

# Parking Pal - Presenting Data

- Pitch: Wouldn't it be great to figure out where you're least likely to get ticketed when parking in Seattle?

Output: dashboard-like tool allowing users to explore when and where tickets were given: most common streets/areas tickets tended to be awarded and times when most tickets tended to be given.

- Data Sources:
  - Seattle Department of Transportation
  - Seattle Municipal Court
- Ideator: Jason Cao

# Drug Overdose and Opioid Prescribing in WA

**Project Type:** Research

**Proposal:** Washington has detailed data on opioid prescribing<sup>1,2</sup>, and overdose EMS responses<sup>3</sup>, ER visits<sup>4</sup>, hospitalizations<sup>5</sup>, and deaths<sup>6</sup> from 2012 through 2025. This research project would combine data about overdose outcomes and opioid prescribing rates across specific quarterly time periods and counties across the state. The intent of the research is to better understand how prescribing patterns are correlated with overdose deaths and track the changing pattern of overdose deaths as opioid prescribing diminished. This research is valuable to understand how overdose patterns in Washington have shifted over time and how they were influenced by opioid prescribing.

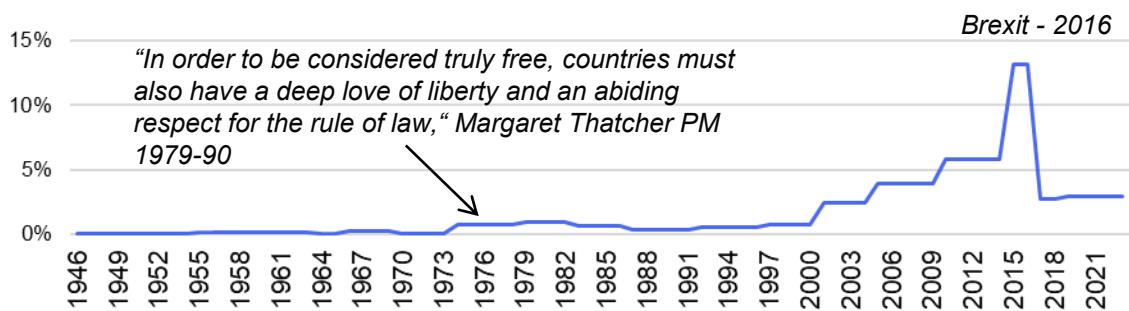
**Data Sets:** 1: Prescription Monitoring Public Use Data; 2: Prescription Demographics; 3: Overdose EMS responses; 4: Overdose ED visits; 5: Overdose Hospitalization visits; 6: Overdose deaths

**Data Sources:** 1: [https://data.wa.gov/health/Prescription-Monitoring-Program-PMP-Public-Use-Dat/8y5c-ekcc/about\\_data](https://data.wa.gov/health/Prescription-Monitoring-Program-PMP-Public-Use-Dat/8y5c-ekcc/about_data); 2-5: <https://doh.wa.gov/data-and-statistical-reports/washington-tracking-network-wtn/opioids/overdose-dashboard>

**Name:** Scott Fry

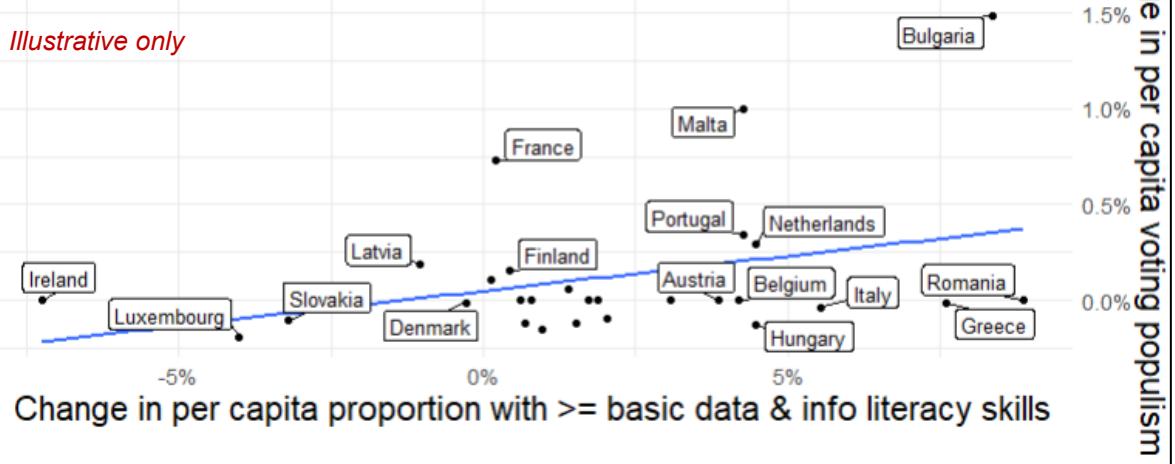
Dive into authoritarian populism and the accessibility of information and technology in Europe! Presently, there's still relative transparency (i.e. access to data in Europe) and 20-30 countries [rows] to build a wide file underpinning a tool, or maybe a research question?

### Proportion of UK voting populist, 1946-2023



### Proportional changes in basic digital skills and authoritarian populism. 2023 vs. 2021

Illustrative only



Note: Parties classified as authoritarian if they openly advocate non-democratic ideologies – e.g., Marxism-Leninism, fascism, Nazism – or if they have demonstrated a willingness to abandon democratic principles while in power.  
Source: Timbro Populism Index, Goodreads, RSF, Eurostat

### Timbro Authoritarian Populism

- Vote for Authoritarian
- Support for Radical Right
- Support for Radical Left
- Ideologic support (various parties)

### Eurostat Digital Skill Index

- Information & data literacy
- Online communication and collaboration
- Digital content creation
- Online safety awareness
- Online problem solving

**LR, KNN, PCA, etc.? Time Series?**

### More Features???

- Timbro Authoritarian Populism Index: <https://populismindex.com/>
- Eurostat Digital & Information Skills: [https://publications.jrc.ec.europa.eu/repository/bitstream/JRC140617/JRC140617\\_01.pdf](https://publications.jrc.ec.europa.eu/repository/bitstream/JRC140617/JRC140617_01.pdf)
- Reporters w/out Borders "Press Freedom Index": <https://rsf.org/en/index?year=2016>
- United Nations: <https://datahub.itu.int/data/?e=AUT>
- EIU: <https://viewpoint-eiu-com.offcampus.lib.washington.edu/data/>
- Anthropic Claude Usage: <https://www.anthropic.com/economic-index#country-usage>
- Something else??

**Jeremy Wise**

# SILENT DRIFT: DETECTING HIDDEN CHANGES IN PUBLIC DATA

W

PROJECT TYPE: TOOL / REUSABLE DATA INFRASTRUCTURE

## Project Pitch

Public datasets frequently change across releases. Columns shift, values are revised, and definitions evolve, often without clear notice. What if we built a tool that compares multiple versions of the same dataset and its documentation to automatically detect schema, distributional, and semantic drift? This helps analysts avoid silent errors and improves reproducibility when working with real-world open data.

## Data Sources (tentative)

**Primary Dataset:** U.S. Census American Community Survey (ACS) yearly releases (structured tabular data, multiple years)

**Secondary Dataset:** ACS technical documentation / codebooks (unstructured metadata describing variables, definitions, and methodology)

Goal: Pair the structured dataset with the unstructured metadata/codebook and flag changes in meaning, not just numbers.

Maanya Cola Bharath

# Why can't I find a job?? (In Seattle)

Research Project: Are mass layoffs followed by hiring sprees? Are layoffs in Seattle followed by hiring recovering, or do they signal permanent job losses? Any industries that drive these layoffs (eg. Tech)?

What I will do: plot whether there is a hiring rebound after layoffs and generally explore the health and patterns of Seattle and Washington's labor market.

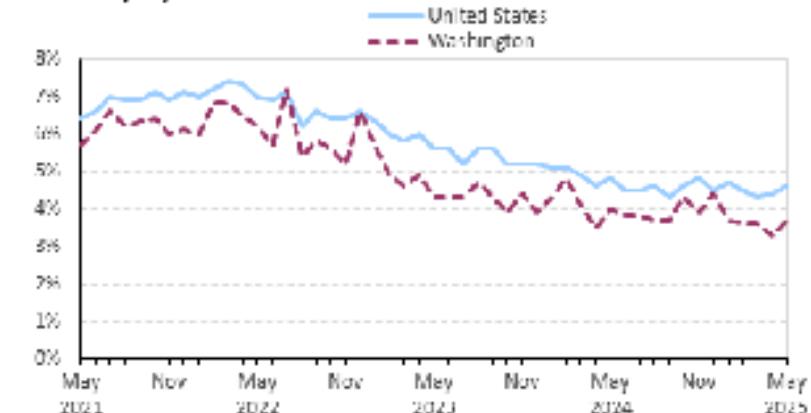
Data sources: [Worker adjustment and retraining notification \(WARN\) layoff and closure database](#)

BLS: [Seattle-Tacoma-Bellevue, WA](#)

U.S. Bureau of Labor Statistics

[Job Openings and Labor Turnover Survey \(JOLTS\)](#) program of the Bureau of Labor Statistics (BLS) produces monthly and annual estimates of job openings, hires, and separations for the nation. The JOLTS program also produces monthly [state estimates](#) for all 50 states and the District of Columbia at the total nonfarm industry level.

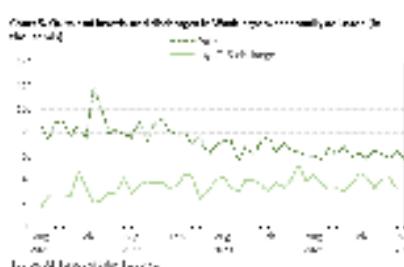
Chart 1. Job openings rates for the United States and Washington, seasonally adjusted



Source: U.S. Bureau of Labor Statistics.

This WARN table includes:

- Name of the employer.
- Business location.
- Number of affected workers.
- Type (layoff or closure).
- Date layoff or closure takes effect.
- Downloadable WARN notice.



# Vulgar Plates and the Poor DMV Workers

Vanity license plates are pretty cool! But what if you are a *terrible* person? So you request a vanity plate with a swear word in it! Or a hateful dog whistle!

You wake up a few days later sad because Janice at the DMV wasn't born yesterday and noticed that you used an *obvious* hate symbol. But what if she hadn't noticed?

California manually reviews every license plate submitted to them. Can we help them reject plates more easily? Can *Twitter*, of all places, help? **Let's create a tool to make Janice's life easier.**

**Data:** <https://github.com/veltman/ca-license-plates/tree/master> (California License Plates)

<https://www.kaggle.com/datasets/muhammadatef/english-profanity-words-dataset> (Hate Speech and Profanity on Twitter)



# MediQuery: Multimodal Health Assistant

## The Pitch

### Research Project

A multimodal AI system that allows patients to describe symptoms using text, images, or voice, and receives personalized health insights. By combining vision models (for analyzing skin conditions or injuries), NLP (for symptom descriptions), and audio processing (for cough analysis), the system provides preliminary assessments and relevant medical information.

## Why It's Cool

Demonstrates true multimodal fusion by processing visual, textual, and audio data simultaneously. Addresses healthcare accessibility by reducing language barriers and enabling remote preliminary screening. Advances research in cross-modal medical AI and user-friendly health tech interfaces.

## Data Sources

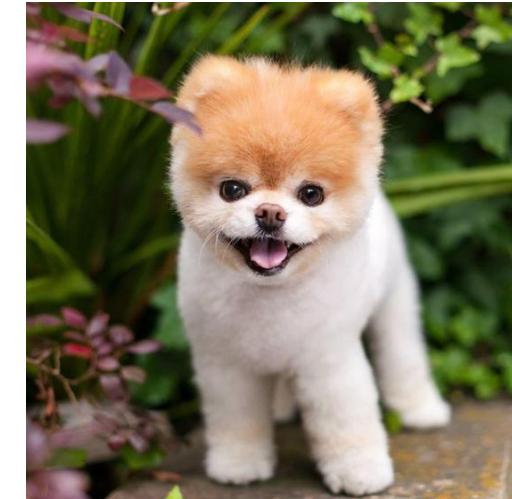
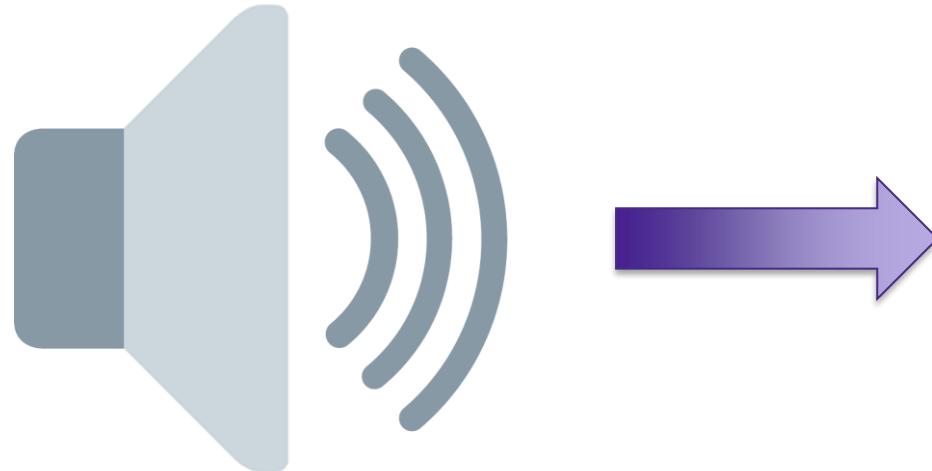
### 1. NIH MedPix Medical Image Database

10,000+ medical images with clinical diagnoses and annotations

### 2. CDC Symptom & Disease Dataset

Text-based symptom descriptions paired with disease outcomes

# Sound2Image Generator



Tool

Sound data: <https://www.kaggle.com/datasets/rushibalajiputthewad/sound-classification-of-animal-voice>

Caption data:

<https://huggingface.co/datasets/d0rj/audiocaps>

DH Lee

W

# Predictive Visibility Risk Assessment Tool for Seattle Urban Traffic Safety

- Project Type: Research or Create a Tool
- What to do: Develop a decision support tool using sunset data, weather data, and historical collision records to identify and predict high-risk time zone across Seattle neighborhoods. We will calculate risk scores by combining low-visibility periods, precipitation, and high-traffic intervals. (Approaches: Weight-based, Regression, or ML)
- Why it is cool: Many study focus on nighttime risks or comparing nighttime vs daytime, but our analysis focus on the 30-minute window after sunset. Through this study, we would explore when human eyes are adjusting to shifting light and collisions peak. Having risk scores can help city planners can take precise safety measures.

[Sunset data](#)  
[Weather data](#)  
[SDOT Collisions All Years](#)  
[SDOT Collisions Vehicles](#)  
[SDOT GIS Datasets](#)

<https://www.fcc-inc.com/dangers-of-driving-during-the-sunset-twilight-hours-2/>

[https://wtsc.wa.gov/wp-content/uploads/dlm\\_uploads/2019/03/Daytime-and-Night-Darkness-and-Light\\_Mar2019.pdf](https://wtsc.wa.gov/wp-content/uploads/dlm_uploads/2019/03/Daytime-and-Night-Darkness-and-Light_Mar2019.pdf)

# Movie & TV Rating Predictor

## Project output: a simple web app

- Enter a new movie/TV show's details (director, cast, genre, plot, etc.)
- Predict a final IMDb rating:

Using ML on structured metadata (cast, director, genre, year), and an agentic AI score based on the plot/background description.

## Datasets:

- Netflix IMDb Scores dataset  
Titles + IMDb ratings
- Netflix Movies & TV Shows dataset  
Netflix metadata: director, cast, categories, description, year, duration, etc.

# What's Cooking?

- William Hu



## Idea/Pitch

**1 What** - A tool that facilitates/encourages people to cook more efficiently. It will have a database of recipes that users can pick and choose from. After picking a recipe, it will automatically go to your designated (Kroger) grocery store and find all the ingredients needed for the recipe along with the cost. The goal is to help encourage people to make that quick pit stop at the grocery store after clocking off work. The ingredients will be compiled for you with a price estimate.

**2 Why/Motivation** - It helps remove more barriers/excuses to not cook (which is a nice life skill and saves money). I would also like a tool like this.

## Data

**4 Kroger Development API** - Allows access to find ingredients, stock/availability, and pricing of food ingredients for Kroger stores. There's a generous request limit so it should work for the goals of this project

**5 Recipes Dataset** - Any large recipe database works (some filtering/tuning would be necessary regardless)

# Polymarket Watchdog -Aaron Kann

Preview    Code    Blame



On January 2nd, 2026, Nicolas Maduro was kidnapped from his home in Caracas by the American military and brought into the American homeland to be put on trial for his crimes. Political affiliations aside, most commentors agreed that the American mission was executed swiftly and effectively, and, most agreed, came as a surprise to nearly everyone.

Well, everyone except for one person.

A couple hours before the Maduro kidnapping, a user on Polymarket placed ~\$33,000 on four different actions that would cash if the American government invaded Venezuela the way it did. It had all the typical features of a trade made by someone on the inside: first ever trade on an account made just before the action, an unusually large wager, occurring hours before the United States military action took place.

And while insider trading is, technically, expressly forbidden on these sites, the nature of the blockchain means that the true identity of this user is unknown to all. And due to the novel nature of these prediction markets, such a trade isn't breaking any federal laws in the same way insider trading on the NYSE. While the previously mentioned trade caused Congressman Richie Torres to call for legislation more explicitly banning insider trading on prediction markets, such legislation looks doomed to stall.

I've always stayed away from putting money on current events in these prediction markets. While I could potentially see the value in betting against a stock as a more profitable hedge than a short position, the legality of insider trading has always pushed me away. However, if we could use Machine Learning on the publicly available data given by Kalshi and Polymarket to catch these actions, this could help us catch the people cheating the system before it happens.



# TradeRewind: Time Machine for your Trading Strategies

**Project Type:** Tool – Algorithmic Trading Strategy Backtester.

**What We're Building:** Python toolkit that tests trading strategies on historical market data.

## Why It's Cool:

- Turns expensive trial-and-error into safe, instant validation.
- Simulate years of trading in seconds, without risking real money.
- Saves traders from costly mistakes before entering real markets.

**Data Sources:** Multi-source data integration ([Yahoo Finance](#) and [Federal Reserve APIs](#))



Priyal Jain

# Breathe Better

## PROBLEM STATEMENT

“How does air pollution relate to **chronic health outcomes** across U.S. communities?”



### Data Sources



#### CDC PLACES

Local chronic disease prevalence



#### EPA Air Quality (AQS)

PM2.5 & Ozone Concentration Levels



#### Census ACS (5-year)

Demographics & Socioeconomic data



### Methodology

Aggregate EPA Air Quality Data  
to County Level

Merge via FIPS Codes

Explore associations using  
interactive visualizations



### Project Output



#### Interactive Explorer

Visual dashboard for insights



#### Reproducible Pipeline

Documented data workflow

# Malicious URL Detection

---

Andrew Jumanca

## ICML-09 Malicious URL Dataset

~2.4M URLs over 120 days

Labels: benign vs. malicious URLs

Will join on:

## Tranco Top Sites Ranking

An open, research-grade list of the **most popular domains on the internet**

## What is the goal?

<https://tranco-list.eu/>

How can we join large-scale URL feature data against independent domain popularity signals to better understand and detect malicious URLs?

<https://www.sysnet.ucsd.edu/projects/url/#datasets>

# Seattle 911 Call Predictor

Aaron Lee

## Project

Uses 10M+ police calls and crime reports spanning two decades to create a **tool** used by emergency dispatchers and law enforcement to help them predict elements of emergency (emergency type, response time, # of officers needed, etc).

## Purpose

- Help law enforcement anticipate and prepare for emergency calls, hopefully leading to more efficient, safe, and effective responses.
- Streamline and efficiently allocate resources for future responses.

## Resources

- [Seattle Real Time Fire 911 Calls](#) ([data.seattle.gov](http://data.seattle.gov))
- [SPD Crime Data: 2008-Present](#) ([data.seattle.gov](http://data.seattle.gov))
- [SPD Officer Involved Shooting \(OIS\) Data](#) ([data.gov](http://data.gov))

# Public Libraries and Schools

Project type: Research

- Compare library usage with local school literacy rates or standardized test scores
- Find any patterns among public libraries and public schools

Data (tentative):

- Seattle Public Library Data
- School District data - literacy rates or scores on standardized tests

Jolene Pern

# Project Proposal - Eva Reutercrona



- Title: What is Seattle Reading?
- Project Type: Answer a Research Question
- Pitch: Are Seattle Public library users checking out all of the popular books on Goodreads? Were the people of Seattle checking out Twilight and other vampire books in 2005 to fight the seasonal depression? I want to take a deeper look at the types of books the Seattle readers are checking out from the library by combining the Seattle Public Library data set with data sets from Goodreads that contain plots, reviews, and userbook interactions from a larger audience.
- Data sets:
  - Seattle Public Library: <https://www.kaggle.com/datasets/seattle-public-library/seattle-library-checkout-records>
  - Goodreads: <https://cseweb.ucsd.edu/~jmcauley/datasets/goodreads.html>

# Urban Heat Island Explorer

## RESEARCH

*"Climate justice meets data science"*



### THE IDEA

Analyze how urban development affects local temperatures and identify neighborhoods most vulnerable to extreme heat. Combine pre-processed satellite temperature data with demographic information to reveal patterns of environmental inequity.

### RESEARCH QUESTIONS

1. Which Seattle neighborhoods experience the highest heat island intensity?
2. Is there a correlation between income levels and heat exposure?
3. How has heat island intensity changed over the past 35 years?

### DATA SOURCES (Verified & Available Now)



#### USGS Surface Urban Heat Island

Pre-processed land surface temperature for 50 US cities including Seattle (1985-2020). Annual heat intensity metrics ready to analyze.



#### US Census ACS Demographics

Income, race, age by census tract via free API. Python libraries (censusdis). 5-year estimates provide reliable statistics.

### ACCESS LINKS:

[data.usgs.gov/datacatalog](http://data.usgs.gov/datacatalog) • [api.census.gov/data.html](http://api.census.gov/data.html) • [censusdis.readthedocs.io](https://censusdis.readthedocs.io)

**Why it's cool: Real environmental justice insights with data immediately available!**

# Data-Driven Survival Strategy for U.S. Small Businesses

Project Type: Create a Tool + Present Data



## Data sources

- **Census Tract Boundaries**  
*U.S. Census TIGER/Line Shapefiles*
- **Demographics & Income**  
*American Community Survey (ACS)*
- **Business Supply & Industry Density**  
*County Business Patterns (CBP)*
- **Competition Intensity**  
*Yelp Open Dataset (reviews, ratings, price levels)*
- **Demand Proxy (Foot Traffic)**  
*SafeGraph Places & Patterns (open/free subsets)*

## What & Why

- Over 50% of U.S. small businesses fail due to poor location and industry choice
- Census-Tract-level tool integrating demographics, supply, competition, and foot traffic
- Models local demand-supply mismatch using LQ and competition intensity
- Identifies survival-friendly industries and locations for entrepreneurs and policymakers

## Expected Output

Interactive map & analytics showing:

- Underserved industries by neighborhood
- High-demand / low-competition trade areas
- Data-backed recommendations for small business survival



Intermission



# Sarah: Identifying Hidden Flood Risk in Seattle

## Project: reusable analysis tool

**Task:** Compare the observed number of flooding reports for drainage basins in storm events to expected range learned from historical data, given rainfall, elevation, and drainage infrastructure.

- *Use drainage basins, storm event pairs as the unit of data.*

### Goal:

1. Detect where reported flooding is inconsistent with historical and physical expectations.
2. Build an interpretable model to examine the influence of key factors on flooding.

## Data

### Observed flooding:

#### Seattle Customer Service Requests

(*Filter for flooding / drainage-related request types*)



#### Seattle Infrastructure Data Layers, Sewer and Drainage (GIS)

- Rain Gages ([Rainfall measurements granular](#))
- SPU DWW Drainage Basins ([Where water flows / collects](#))
- SPU DWW Catch Basins / Inlets / Drainage Points ([Where surface water enters](#))

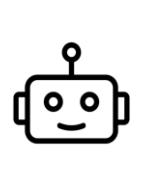
### Rainfall measurements city wide:

#### NOAA Hourly Precipitation Data (Seattle)

### Elevations in Seattle:

#### 2021 USGS Lidar DEM: King County, WA





# WA EV Charging Smart Assistant

- My name :Aiden Tan
- Project Output: a web chatbot
- Build a natural-language chatbot that helps EV owners and planners in Washington State answer questions like:
  - "Which are the best charging station near UW?"
  - "Is EV adoption growing rapidly in King County? Are the charging stations keeping up?"
  - "I drive a Tesla; where are the easiest places to charge while traveling in Washington State?"
- Data
  - WA Electric Vehicle Population Data
    - This dataset shows information of all Electric Vehicles that are currently registered through Washington State Department of Licensing (DOL).
    - [Electric Vehicle Population Data | Data.WA | State of Washington](#)
  - Alternative Fueling Stations
    - EV charging stations(street\_address, zip code, pricing, etc...)
    - [EV-MAP Site Data page | EV Mapping and Planning Tool Site](#)



# Crop Resiliency

Project Type: Research Question

What are the most and least resilient crops harvested in the US?

This does not only help farmers know which crops make the safest investments. It also helps the USDA, or other, government agency predict and prepare to assist farmers when unpredictable weather conditions damage their profits.

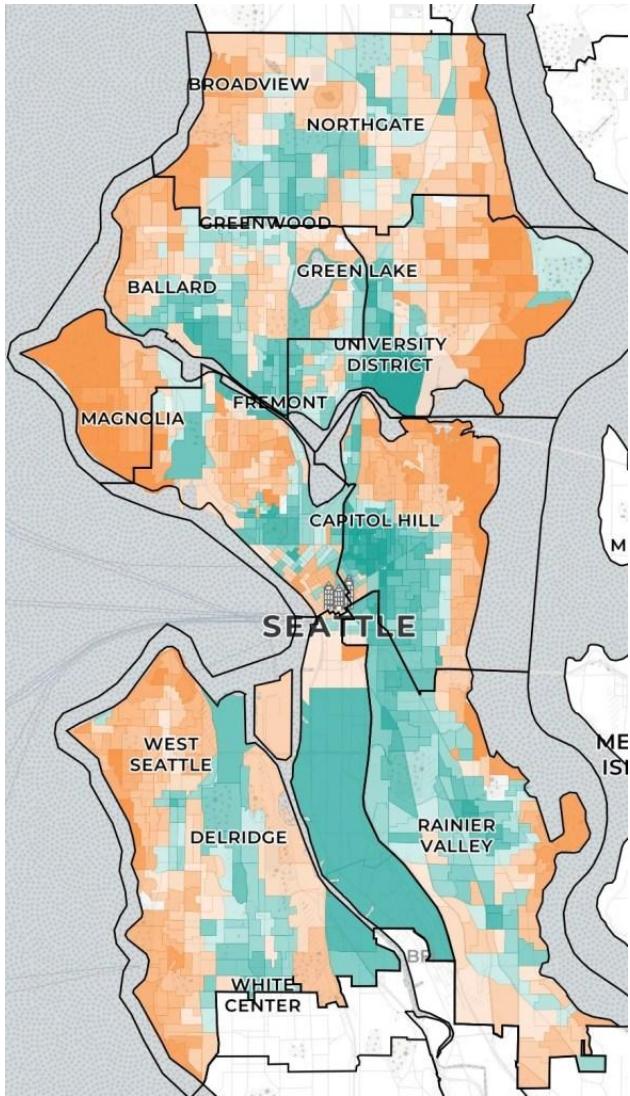
USDA Quick Stats: <https://quickstats.nass.usda.gov/>

NOAA Local Climatological Data: <https://www.ncei.noaa.gov/products/land-based-station/local-climatological-data>

James Holtz

# Seattle Elections

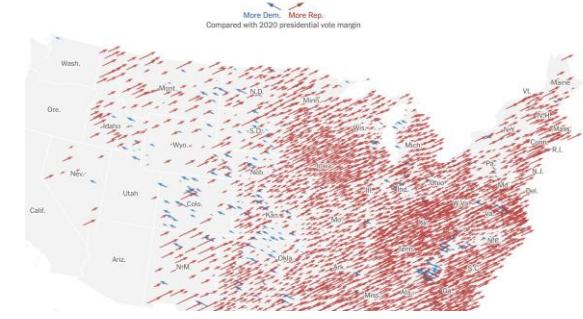
Research Project



2025 Mayoral  
Election Results

Data Source 1: Seattle  
Election Results

Data Source 2: Seattle  
Voter File



**Questions to be  
answered:**

Which areas are the most  
"valuable"?

Which areas are the most  
"swingy"?

Which areas are trending  
towards one side or  
another?

# The Data Scientist Toolkit Tracker

Project Type: Tool

## WHAT DOES THE TOOL DO?

This tool will track commonly used models, libraries, and tools in a data scientist's workflow, using signals like community adoption, development activity, and stability to show how active and reliable each tool is.

## WHAT IS THE VALUE?

Data science and AI tools change fast, and it's hard to know which ones are actually worth using. This tracker helps data scientists keep up with the field, decide which tools make sense to adopt, and update their workflows in a way that can improve their productivity.

## DATA SOURCE

- [GitHub AI Repository Rankings](#)
- [GitHub Release Pages for Key Libraries](#) (e.g., LangChain, LlamaIndex)
- [Awesome Open-Source AI Tools Lists](#)
  - tool discovery and categorization
- [Hugging Face GitHub AI Project Docs Dataset](#) – documentation quality and tool descriptions

By: Arlette Ngabonzima

# Plane Your Trip!

A Web Tool for Predicting Plane Ticket Prices

- In 2023, almost 90% of Americans have taken a commercial flight. Many of us do so regularly
- What are the factors that affect plane ticket prices? Can we predict pricing?
- Data sources
  - “Airfare ML : Predicting Flight Fares” (Kaggle) – Flights prices
  - U.S. Energy Information Administration (EIA) – Fuel prices
  - Other datasets for sources of ticket pricing
- Interested? Come find Henry after this

## US Emissions Playground

- Project type: Tool
- Description: Create an interactive tool which allow users to learn more about sources of US emissions. Users can adjust sector level emissions and see how these hypothetical changes impact emission trajectories, and check if we are on track to meet our climate targets. I think this project is cool because it makes learning about a complex and important topic accessible and interesting.
- Data sources:
  - CO2 and Greenhouse Gas Emissions (Our World in Data)
  - EPA US Greenhouse Gas Inventory
- Catherine Wu

# The Data Scientist Toolkit Tracker

Project Type: Tool

## WHAT DOES THE TOOL DO?

This tool will track commonly used models, libraries, and tools in a data scientist's workflow, using signals like community adoption, development activity, and stability to show how active and reliable each tool is.

## WHAT IS THE VALUE?

Data science and AI tools change fast, and it's hard to know which ones are actually worth using. This tracker helps data scientists keep up with the field, decide which tools make sense to adopt, and update their workflows in a way that can improve their productivity.

## DATA SOURCE

- [GitHub AI Repository Rankings](#)
- [GitHub Release Pages for Key Libraries](#) (e.g., LangChain, LlmalIndex)
- [Awesome Open-Source AI Tools Lists](#)
  - tool discovery and categorization
- [Hugging Face GitHub AI Project Docs Dataset](#) – documentation quality and tool descriptions

By: Arlette Ngabonzima

# Calamity-Watch

## A Real-Time Global Multi-Hazard Monitoring Dashboard

### Project Type: Geospatial Tool / Interactive Map

I am building an interactive web application that aggregates real-time natural disaster data into a single global dashboard. The system combines NASA FIRMS wildfire hotspot data with GDACS multi-hazard alerts (earthquakes, floods, cyclones, and volcanoes) to provide a unified, map-based early-warning view. This is cool because it turns live satellite and disaster-alert feeds into an accessible visual monitoring tool useful for researchers, planners, and travelers.

#### Data Sources:

- NASA FIRMS API - Near real time wildfire hotspot data from MODIS and VIIRS satellites
- GDACS API - Global alerts for earthquakes, floods, tropical cyclones, and volcanic eruptions

Name: Sreeraj Parakkat

## **StreamSurge**

### **A Streaming Data Monitoring Tool for Music Popularity**

Project Type: Tool project

StreamSurge is a Python-based tool that simulates real-time music streaming data, compares it against historical baselines, and detects unusual spikes or drops in popularity. The project focuses on software and system design—data ingestion, monitoring, anomaly detection, and visualization—rather than prediction or research analysis. It's cool because it turns static music data into an interactive monitoring system similar to what analysts use in real-world platforms.

Data Sources (Tentative):

- Spotify Charts — daily streaming counts by track and country (historical batch data)
- Simulated Live Stream — incremental replay of historical streaming data to emulate real-time ingestion

# Predicting Flight Delays

## Johnny Liang

- Project type: Tool
  - Create a tool that uses weather pattern data (temperature, air pressure, etc.) to predict the length of delays upcoming flights.
  - Help individuals plan their trips, help airlines prepare in advance.
- Sources:
1. Flight delay data: <https://www.transtats.bts.gov/ontime/>
  2. Weather data: <https://www.ncdc.noaa.gov/products/land-based-station/integrated-surface-database>
  3. Airport data: <https://datahub.io/core/airport-codes>

# EV Adaptation

Katelyn Wu

How can we optimally establish electric vehicle chargers in seattle?

## Output

Present data that allows developers to input a region and get:

- Predicted EV adaptation likelihood
- Recommended Charger placement and insights on where infrastructure may be needed.

Extra resources: [How planning relies on road and traffic data](#)

## Data

State of Washington Open Data:  
[Electric Vehicle Registration in Washington](#)

Seattle GeoData Transportation:  
[Traffic Flow data](#)



**OPEN** or **CLOSED** ?

## A Business Survival Simulator for NYC

Walk along the Ave: you'll see packed shops, empty storefronts, closed businesses. Some areas are busier than others. Why do they survive while others close?

**Can neighborhood conditions be used to estimate business survival?**

**Data Sources** ([data.cityofnewyork.us](http://data.cityofnewyork.us)):

- **NYC Issued Business Licenses:** regularly updated with business category, location, license issue and expiration date
- **NYC 311 Service Requests (2020 to present):** resident-reported neighborhood complaint type, time, location



**Project Type: Tool Project - estimated probability of surviving past 3 years, 5, etc.**



Hannah Sun

# Early Warning Signals from Aviation Near-Miss Narratives

Anna Tender

*Project Output: Research & Analytics Tool*

**NASA ASRS:** Voluntary aviation safety; report narratives; near-misses, coded fields (phase of flight, anomaly type, reporter role)

**NTSB accident & incident data:** Occurrence categories; Event dates and outcomes

- Analyze aviation safety narratives to detect emerging aviation hazards
- Group narratives using NLP clustering
- Compare hazard trends to official accident and incident data
- Test whether increases in near-miss reports precede formal accident and incident outcomes



# FOOD DESERT PROJECT

JIANNA WONG

- Project Output: map visualization + data tool
  - Show the intersection of income and access to supermarkets vs. fast food
  - Find the riskiest food desert areas to live in
  - Show over the years how this has changed
- Data:
  - USDA Food Access data
  - Census Bureau ACS 5-Year Estimates
  - Local Data for Better Health, Census Tract Data



# Calamity-Watch

## A Real-Time Global Multi-Hazard Monitoring Dashboard

### Project Type: Geospatial Tool / Interactive Map

I am building an interactive web application that aggregates real-time natural disaster data into a single global dashboard. The system combines NASA FIRMS wildfire hotspot data with GDACS multi-hazard alerts (earthquakes, floods, cyclones, and volcanoes) to provide a unified, map-based early-warning view. This is cool because it turns live satellite and disaster-alert feeds into an accessible visual monitoring tool useful for researchers, planners, and travelers.

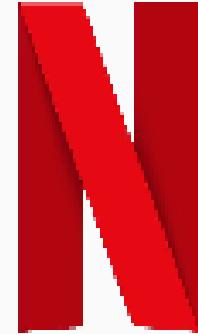
#### Data Sources:

- NASA FIRMS API - Near real time wildfire hotspot data from MODIS and VIIRS satellites
- GDACS API - Global alerts for earthquakes, floods, tropical cyclones, and volcanic eruptions

Name: Sreeraj Parakkat

# User Behavior Prediction

Determining Netflix viewer  
retention based on...



Combine Datasets (kaggle):

Watch history, Search logs, Recommendations,  
Reviews, User info

Goal:

- Create a visual tool that easily answers potential business questions such as:
  - Which customers are higher risk of cancelling?
  - Which customers should receive a retention offer?
- Work with large, messy data and understand its tradeoffs

# Swimsafe WA

Project output: reusable data (a simple web app)

- Advise whether it is safe to swim at a Washington beach, lake, or swim spot
- Interactive map visualization with zoom and click
- Search feature to find specific swim locations

Data:

- Washington State Department of Ecology – Beach Program water quality data
  - Bacteria measurements and swim advisories
- USGS environmental data
  - Water temperature and other context



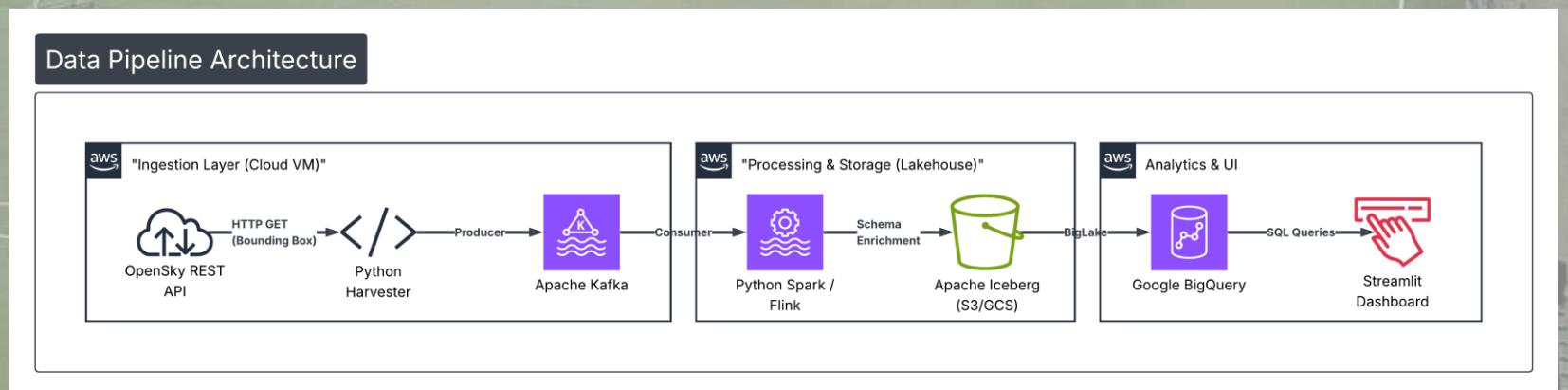


# Book Club / Hobby Group Manager

- Category: Tool/ Website /App? I am open to making any of these.
- Pitch: I love book clubs but it's almost impossible to maintain a book club and coordinate what book to read, when to meet up, deadlines to set, etc. My solution: a dedicated web platform (or app) that centralizes everything book clubs need: propose and vote on books/other decisions democratically, set synchronized reading schedules with progress tracking, host forum discussions, and coordinate meetings. I also imagine and would like to make the tool able to support other hobby groups too, whether that be sports watching, DND + board games, music/bands, etc.
- Data Sources + APIs: Google Books, Open Library, Goodreads, Google calendar API, Meet-up API
- Possible Features: Group creation, voting, progress setting, discussion forums, member profiles, meeting scheduling
- Ha Tien Nguyen

# SUPER BOWL SKY

*To build a real-time Geospatial Lakehouse that monitors all aviation activity within a 50-mile radius of Levi's Stadium during Super Bowl LX. The system will automate the ingestion of high-frequency state vectors, calculate estimated carbon emissions per flight, and utilize Iceberg Time-Travel to audit the 'Celebrity vs. Commercial' congestion ratio.\**



# Segment Hunting



**City of Seattle**



- Q: How do traffic features affect people's ability to run/ride fast on a segment? (research project)
  - Choose a handful of similar segments (length, elevation gain) to analyze (fastest time, spread over top 10, total number of efforts) in slightly different places around the city
  - Implications: safety, weakest segments
- Two(+) datasets: strava segment details and geodata from the city of Seattle (join on location)
  - Access segment details via strava api or web scraping
  - Potentially use SDOT sidewalks, intersections, crosswalks data

## "Under the Weather: Does the Environment Dictate our Fitness and Sleep?"

- **Project type :** Research
- **Question :** How fitness and sleep quality varies with weather
- **Goal:** Integrate biometric data with environmental data to evaluate how weather shifts and daylight changes affect physical activity levels and sleep quality.
- **Datasets:** [PMDataset](#) - Fitbit data from 16 participants for five months (includes calories burned, steps, exercise, sleep scores, sleep breakdown into light, deep, rem) and Weather data from [OpenWeather](#) based on the location where the study was done (Norway).
- **Name:** Urvashi Jha

# Outlier Detection in Real-World Health Data Using All of Us Research Program

## Project: Research + reusable analysis workflow (tool)

### Data:

#### All of Us Research Program

Pros: high external validity (real clinical variation), longitudinal EHR + program-collected measures + surveys, large scale for rare patterns

Cons / constraints: cannot remove or copy participant-level data (no screenshots/downloads); cannot attempt re-identification; cannot link participant-level external datasets without explicit permission; some fields are suppressed/generalized to reduce re-identification risk

### Methods

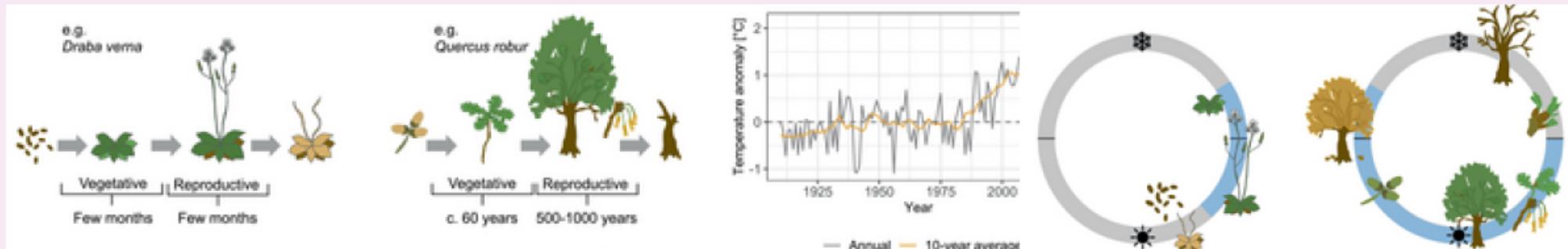
- Build an outlier detection pipeline for patient-level physiology and lab profiles using Mahalanobis distance as a baseline, then compare to a few alternatives
- Analyses run in the cloud-based Researcher Workbench: you create a project workspace and do SQL + notebook analyses against curated datasets
- Example datasets: EHR labs/vitals, collected measurements like BMI, blood pressure

# Modeling How Weather Patterns Predict Plant Bloom Timing (Phenology) \*

**Project Type:** Answer Research Questions

**What we'll do:** Combine phenology observations with NOAA weather data to predict bloom timing across years and species.

**Why it's cool:** Shows climate change in real life by tracking shifts in bloom timing.



**Datasets:**

**USA National Phenology Network (USA-NPN)**

- Plant bloom / first leaf / flowering observations across the U.S.

**NOAA National Centers for Environmental Information (NOAA NCEI)**

- Daily temperature, precipitation, and weather history by location

Emily Tran

*Seattle!*  
*the city of coffee*



# Washington Coffee Matcher

Review data I scraped  
from my coffee  
subscription service

ft  
&

Additional roaster information  
we decide on (mission, store  
availability, etc)

Julia  
Glasser



# WA Hiking Trail Patterns

By Elsie Wang



## Tool + Data Presentation Project

### Description

Interactive analytics and visualization tool exploring when and where people hike in WA and seasonal trail conditions; provides hiking windows

- When does hiking activity decrease and increase in WA?
- How does seasonality differ by region?
- When's the best time to hike a certain trail?

### Data Sources

- Washington Trail Association Trip Reports
  - Timestamps, trail names, regions, condition tags
- Washington Trail Geometry & Metadata
  - OpenStreetMap, WA State Parks GIS
- National Weather Service API
  - Real-time weather conditions

Title: Global Olympic Impact

Project Type: Answer Research Questions

Pitch:

- Do countries that host the olympics medal more often?
- How is a country affected by hosting the olympics?
- Can look at:
  - Health
  - Nutrition
  - Mortality
  - economy
- Potentially try to predict how the LA olympics will affect the US.

Data Sources:

<https://www.kaggle.com/datasets/harshvgh/olympics>

<https://www.kaggle.com/datasets/carolinacappato/olympic-games-info>

<https://www.kaggle.com/datasets/miguelroca/global-health-nutrition-mortality-economic-data>

Name: Sophia Koehn

# Golf Tournament Prediction - Josh Tseng

Predicting golf tournaments winner week in week out.

Data Source: [Data Golf, PGA TOUR](#)