SmokeSignal: Wildfire Smoke and Respiratory Burden in Oregon Counties

DATA 510 M1 Practice Proposal (Critique Activity)

Lucas P. Cordova Teo Webster Ben Mendoza

2026-06-01

*This document is a **practice proposal** for the week 4 in-class critique activity. It is not a real M1 submission. Use it only to practice merit/shortcoming analysis before you submit your own proposal.*

1 Title and Team Metadata

Project title: SmokeSignal: Wildfire Smoke and Respiratory Burden in Oregon Counties

Name	Willamette email
Lucas P. Cordova	lpcordova@willamette.edu
Teo Webster	twebster@willamette.edu
Ben Mendoza	bmendoza@willamette.edu

Peer Stakeholder POs: Alex Kim, Morgan Lee, Riley Santos (Shared Theme: *Public health and environmental exposure*)

Studio Session: 2

Team size: 3

2 Problem Framing and Research Questions

2.1 Background: Oregon, smoke, and health at county scale

Western Oregon used to treat wildfire smoke as an eastern-Cascades problem. That is no longer true. Since about 2015, **Willamette Valley and southwest Oregon** communities have repeatedly logged **Air Quality Index** values in the "Unhealthy" or worse bands during late-summer wildfire events, sometimes for **weeks at a time**, not two-day spikes [2]. Fine particulate matter (PM2.5) from wildfire smoke is small enough to penetrate deep into the lungs; epidemiologic work links

wildfire-specific PM2.5 to **respiratory hospital admissions and ED visits**, with effects that vary between urban and rural counties [1].

At the same time, **county public health agencies** operate on thin staffing and fragmented tools: EPA monitoring maps, school district judgment calls, and anecdotal clinic surges. Oregon Health Authority (OHA) publishes guidance, but **county-level situational awareness** (which counties are worsening relative to their own baselines, and on what timeline) is still hard to assemble quickly from open data alone. We are not trying to replace OHA; we want to show whether a **reproducible county-month panel** built from public (and optionally restricted) sources can support better seasonal decision-making.

2.2 Why this project matters to our team

We are a three-person owner team with different entry points into the same problem:

- **Lucas** spent time in **Jackson County** during the 2020 smoke season and remembers Medford's AQI staying above 150 while outdoor school sports were still debated day-by-day. That pushed us toward a capstone with **visible Oregon geography**, not a generic Kaggle tabular set.
- **Teo** worked on **air-quality sensor QA** in an undergraduate environmental science lab and wants this capstone to include a credible **ingestion and versioning** story (not a one-off CSV download).
- **Ben** is interested in **forecasting** for operations (when to staff clinics, when to pre-position communications) and has family in **Deschutes County** who receive county alert texts that sometimes arrive after the worst smoke has already passed.

What unifies us: we care about **rural and mid-size Oregon counties** that lack the monitoring density of Portland and that still have to make **school, clinic, and outreach** decisions under uncertainty. If we can quantify even an **associational** link between smoke exposure proxies and respiratory burden proxies, and explore whether **short-horizon forecasts** are feasible, we will have a portfolio piece that speaks to **public-sector data science**, not only model accuracy on a leaderboard.

2.3 Problem statement (consequential framing)

Problem: Oregon counties face more frequent and longer wildfire smoke episodes, but decision-makers lack timely, county-specific evidence connecting **smoke exposure** to **respiratory health burden** and to **actionable lead times** for communication and capacity planning.

Consequence if unsolved: Counties continue to rely on regional averages, late alerts, and anecdotal clinic reports; vulnerable populations (children with asthma, older adults, outdoor workers) bear uneven risk [2].

What success would change (this term): A documented data pipeline, an analysis-ready county-month panel for Oregon, and clear answers (with stated limits) to the research questions below, communicated in forms county staff could skim (maps, plain-language summary), not only in a modeling notebook.

2.4 Research questions

Primary research questions (as stated in this draft):

1. **RQ1:** Do Oregon counties with more wildfire smoke days (PM2.5 above EPA 24-hour standard) show higher rates of asthma-related emergency department visits than counties with fewer smoke days, after adjusting for season and population?
2. **RQ2:** Can we forecast county-level "high respiratory burden weeks" one to two weeks ahead using smoke, weather, and historical claims patterns so public health teams can act early?

We also plan to "apply machine learning for insights" across merged county panels.

Charter alignment note: Our committed `CHARTER.md` (week 3) is unchanged from chartering; see Section 10 for repo link.

3 Impact and Stakeholders

Who benefits:

- **Oregon Health Authority (OHA)** epidemiologists comparing counties for resource planning.
- **County public health departments** (e.g., Deschutes, Jackson, Multnomah) deciding when to issue smoke guidance.
- **School district operations staff** weighing outdoor practices during smoke events.

Impact (plausible, not hype): If we produce a reproducible county-month panel and clear association or forecast metrics, OHA-style reviewers could use the workflow as a template for seasonal situational awareness. We are not claiming causal policy evaluation without stronger design.

Shared Theme and peer PO influence: Our peer POs are matched on *public health and environmental exposure*. We expect them to push us on **ecological inference** (county aggregates), **ethics of using claims-like outcomes**, and whether our communication plan reaches non-technical county staff.

4 Data Sources and Access Plan

Source	Description	Access	Volume (est.)	Refresh	Instructor approval
EPA Air Quality System (AQS)	County-linked PM2.5 daily summaries	Public web + API (links 1 and 2 below)	~500k rows national; OR subset smaller	Daily	**Approved** (assumed)

Source	Description	Access	Volume (est.)	Refresh	Instructor approval
CDC PLACES	County chronic disease and ED visit proxies	Public download (link 3)	County-year tables	Annual	**Approved** (assumed)
NOAA NCEI	Temperature, wind, humidity for confounding	Public archives (link 4)	Station-level; aggregate to county	Daily	Pending
NIFC / NWCG open data	Fire perimeters for context flags	Open GIS portal (link 5)	Event polygons	As published	Pending
Oregon All Payer Claims Database (APCD)	County-month pediatric asthma ED rates	OHA researcher application (link 6)	~2M claim lines if granted	Quarterly	**Approved**

Documented access URLs:

1. EPA AQS: https://www.epa.gov/aqs
2. EPA AQS Data API: https://aqs.epa.gov/aqsweb/documents/data_api.html
3. CDC PLACES: https://www.cdc.gov/places
4. NOAA NCEI: https://www.ncei.noaa.gov/
5. NIFC Open Data: https://data-nifc.opendata.arcgis.com/
6. Oregon APCD (OHA): restricted; application via OHA researcher process

We will merge on county FIPS and month. APCD gives finer outcomes than PLACES alone; we listed PLACES as backup.

Note: We have not yet posted an approval request in `#blockers` for APCD or NOAA pulls.

5 Data Engineering Plan

Repo bootstrapped from LucasCordova/ds3template:

- `data/raw/aqs/` : API pulls by county FIPS and parameter code.
- `data/raw/places/` : downloaded CSV snapshots versioned by download date.
- `data/interim/` : harmonized county-month spine.
- `data/processed/` : analysis-ready panel for modeling.
- `src/ingest/aqs_pull.py` : reproducible AQS download script.
- `notebooks/01_panel_qc.ipynb` : row counts, missingness, duplicate checks.

We will use **DVC** or dated folders for versioning (team still deciding). README documents refresh commands. Raw data never overwritten; processed sets tagged in Iteration Reviews.

6 Planned Analysis

Descriptive: Map smoke days and PLACES asthma prevalence by county; time-series plots for 2016 to 2024.

RQ1 (association): Multilevel or county fixed-effect regression: smoke-day count vs asthma ED rate proxy, with seasonal indicators. Consider negative binomial if counts.

RQ2 (forecast): LSTM ensemble (PyTorch) on county-month sequences with smoke, weather, and lagged outcomes; compare to naive seasonal baseline (maybe).

Evaluation: RMSE for forecast; R-squared or coefficient stability for association. Train through 2022, validate 2023, holdout 2024.

Success: We have "answered" RQ1 if smoke coefficient is positive and robust to basic weather controls. RQ2 if LSTM beats baseline on holdout weeks.

Backlog mapping (partial): We created three GitHub issues titled "Get data," "Train model," and "Make dashboard" without full Create / Observe / Analyze triples yet.

7 Program Integration (Five Pillars)

Pillar	Plan
Data engineering	AQS API scripts, county spine, raw/interim/processed layout (Section 5).
Analytics / ML	Regression for RQ1; LSTM ensemble for RQ2.
Visualization and communication	County choropleth maps and a Shiny-style dashboard for "smoke risk week" (tool TBD).
Ethics and responsible use	We will follow fair AI principles and protect privacy.
Statistics / research design	We will use appropriate statistical tests.

8 Ethics and Risk Notes

- **Privacy:** APCD is restricted; we will only report aggregates.
- **Fairness:** We will ensure equitable outcomes for all communities.
- **Limitations:** Data may be incomplete.

(Charter ethics bullets mentioned ecological fallacy at county level and APCD sensitivity; this section does not expand on them.)

9 Timeline and Milestones

Week	Plan
4 (now)	Submit proposal; finalize APCD application
5	Production AQS API pipeline; start LSTM architecture search
6	Full panel merged; LSTM training complete
7	**M2-data-summary** deliverable
8 to 9	Dashboard polish; extra models
10	**M3-poster-draft**
11	User testing with "stakeholders"
12	**M4-writeup-draft**
13	Rehearsal
14	**M5-final** poster and write-up

Risks: APCD delay (mitigation: use PLACES only). Compute for LSTM (mitigation: Colab Pro).

10 DS3 Process and Tooling

- **GitHub repo:** `https://github.com/example-wu/smokesignal-data510` (placeholder; board not linked yet)
- **Projects board:** `https://github.com/users/example-wu/projects/1` (columns created; **3 issues** in Backlog, tags inconsistent)
- **Discord:** Category `#smokesignal-*` confirmed for team; peer project categories "joined"
- `CHARTER.md`: Committed. **Vision:** *Oregon communities have timely evacuation routing during wildfire events.* **Mission:** *Build an interactive evacuation route dashboard for rural Oregon counties by M5.*
- **Backlog:** Seeded in `BACKLOG.md` with 3 PBIs; milestone tags mostly `M1-proposal`
- **Studio loop:** We will post Iteration Reviews in `README.md` before class and respond to peer Briefs within "a reasonable time."

10.1 Appendix: References

1. Liu, J. C., Mickley, L. J., Sulprizio, M. P., et al. (2016). Wildfire-specific fine particulate matter and risk of hospital admissions in urban and rural counties. *Epidemiology*, 27(3), 350-357. https://doi.org/10.1097/EDE.0000000000000456
2. Reid, C. E., Brauer, M., Johnston, F. H., et al. (2016). Critical review of health impacts of wildfire smoke exposure. *Environmental Health Perspectives*, 124(8), 1334-1343.
3. U.S. Environmental Protection Agency. (2024). Air Quality System (AQS) documentation. https://www.epa.gov/aqs
4. Centers for Disease Control and Prevention. (2024). PLACES: Local Data for Better Health. https://www.cdc.gov/places
5. National Interagency Fire Center. (2024). NIFC Open Data. https://data-nifc.opendata.arcgis.com/