Air Sensor Data Analysis

Webinar 2:

Gaining Insights from your Data: Visualization and Interpretation

Webinar Logistics

Automated closed captioning

"Switch to Phone Audio" if your computer's microphone/speakers are not working

Submit questions for Presenters, Zoom Technical Assistance

- **Closed captioning** is available by clicking the "CC" button in your control panel
- This webinar is being recorded.
- **To ask a question:** Type your question in the Q&A Box.
- **Technical difficulties:** If you are having technical difficulties, please send a message through the Q&A Box

Overview of Webinar Series

Webinar #1:

Data Wrangling: Accessing Data, Data Formats, and Quality Control

Webinar #2:

Gaining Insights from your Data: Visualization and Interpretation

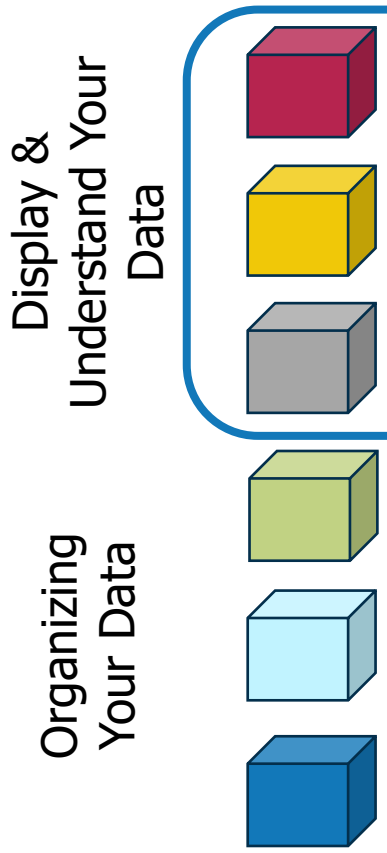
Webinar 2 Agenda

1. Objectives & Analysis
2. Visualizations
3. Matching Objectives & Visuals
4. Interpreting Data & Drawing Conclusions
5. Analysis Tools
6. Recap
7. What's Next
8. Q&A

Objectives & Analysis

Critical Elements of Data Analysis

Building blocks of data analysis



Interpreting (What does it show? Not show?)

Visualizing (How to show data & information?)

Creating Metrics (How to average data? How to aggregate data?)

Today's focus

Supporting Data (What else is needed? Who else needs to be involved?)

Quality Control (How robust? How to operate & maintain? How are data validated?)

Data Collection (How long? How many sites?)

Align your analysis with your project objective!

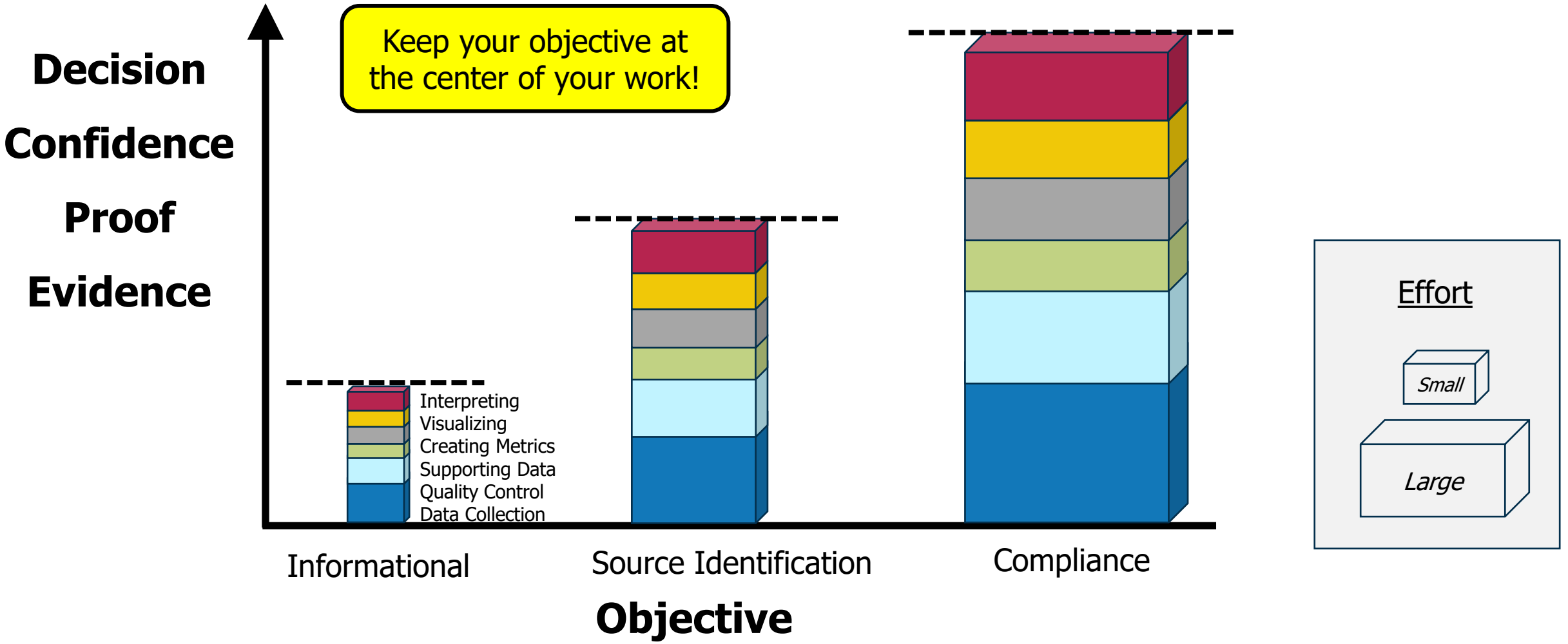
There are many ways to analyze and visualize data, but it ultimately should be based on your objective

As we explore different types of visuals, revisit your objective and ask:

- What are we hoping to learn?
- What information do we want to share?
- Who is our audience?
- What are the limitations of our data? (e.g., don’t have certain data)

Critical Elements of Data Analysis

Building blocks of data analysis



Visualizations

Visualizations

Data visualizations are ways we explore, investigate, and share our data.

It's how we investigate data across:

- **Time** – time series, diurnal/weekly trends, calendar plots
- **Space** – maps, wind and pollution roses, correlation plots

And summarize our data - averages, pie charts, bar charts, etc.

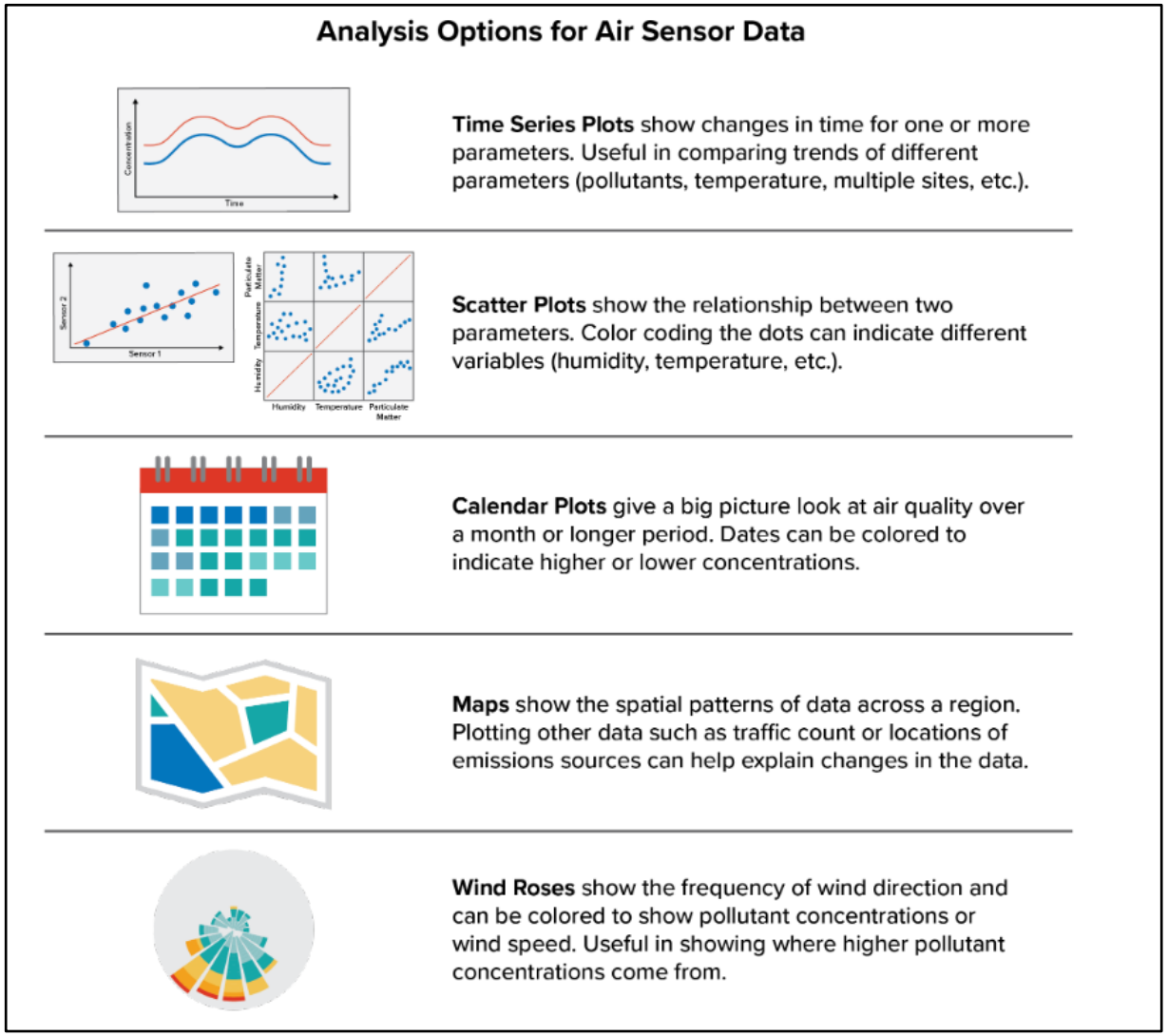


Image Source: EPA

Visuals often can combine information across time and space as well!

Data across Time - Time Series

Explore pollutant changes over time

- Good visualization to start with
- Useful for showing both specific details (spikes, incidents) and general trends
  - Across different parameters (pollutants, temperature, multiple sites, etc.)
  - Against health-based thresholds (e.g., AQI, others)

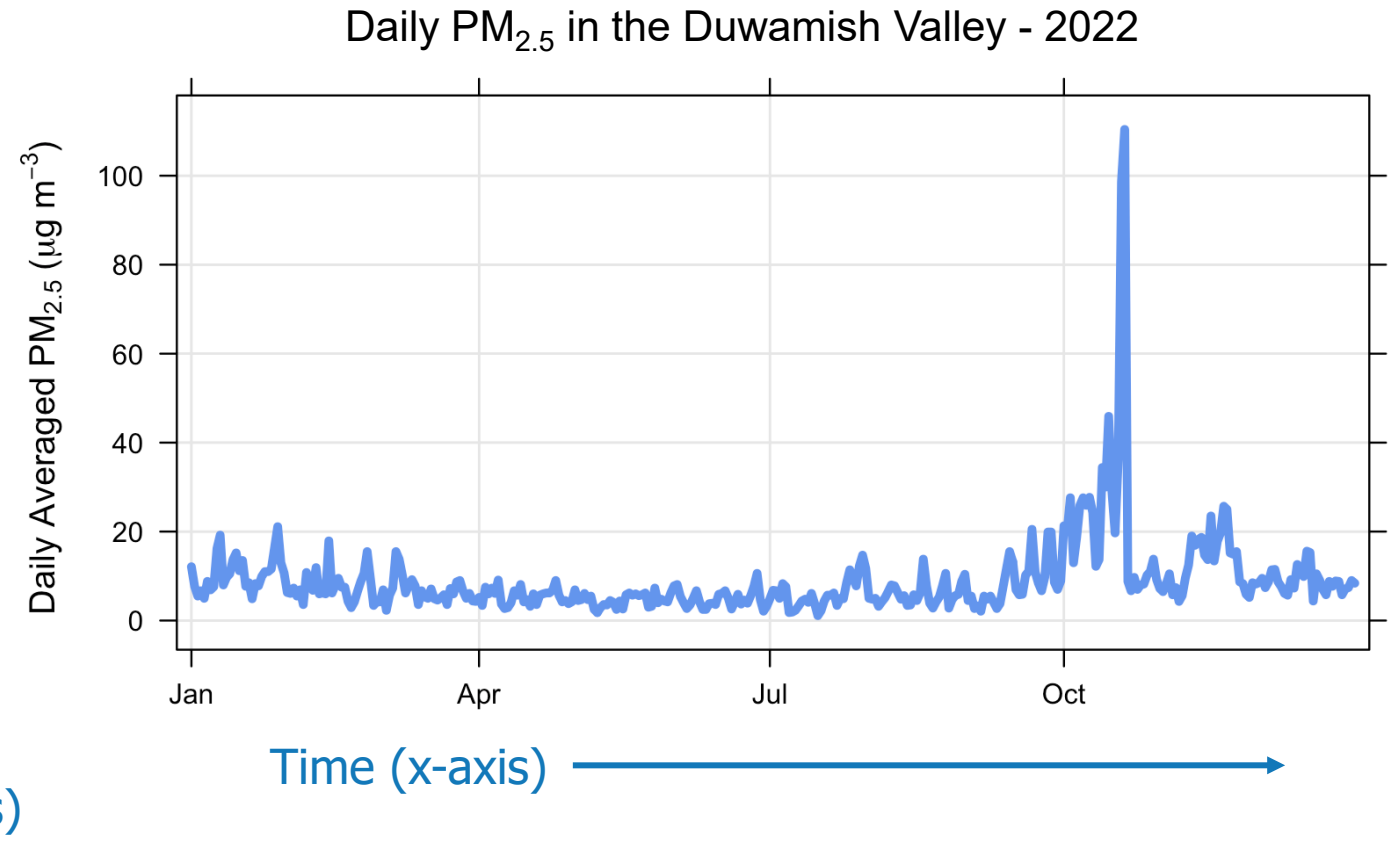
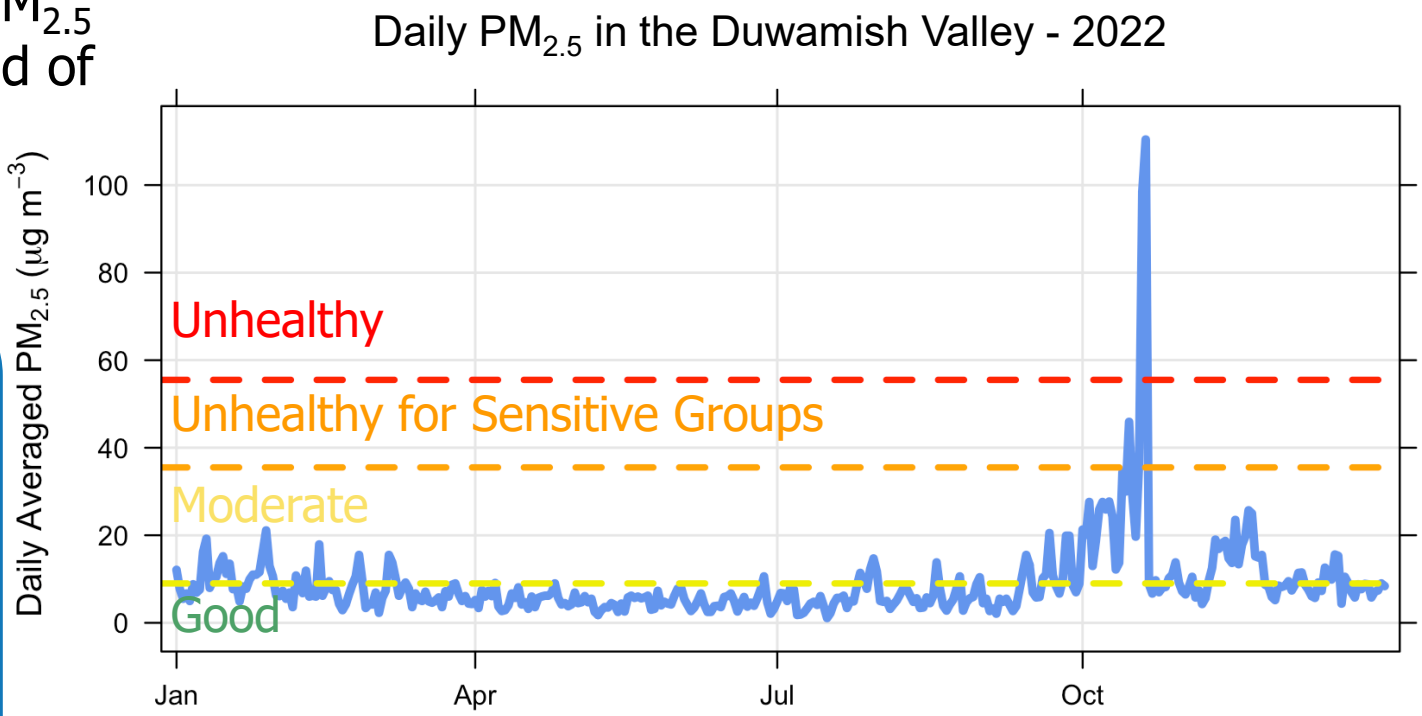


Image Source: TD Enviro

Data across Time - Time Series

Same time series, but now with AQI thresholds to add health context! Show majority of $PM_{2.5}$ in Good and Moderate range, with a period of Unhealthy levels in the fall.



Questions to Ask:

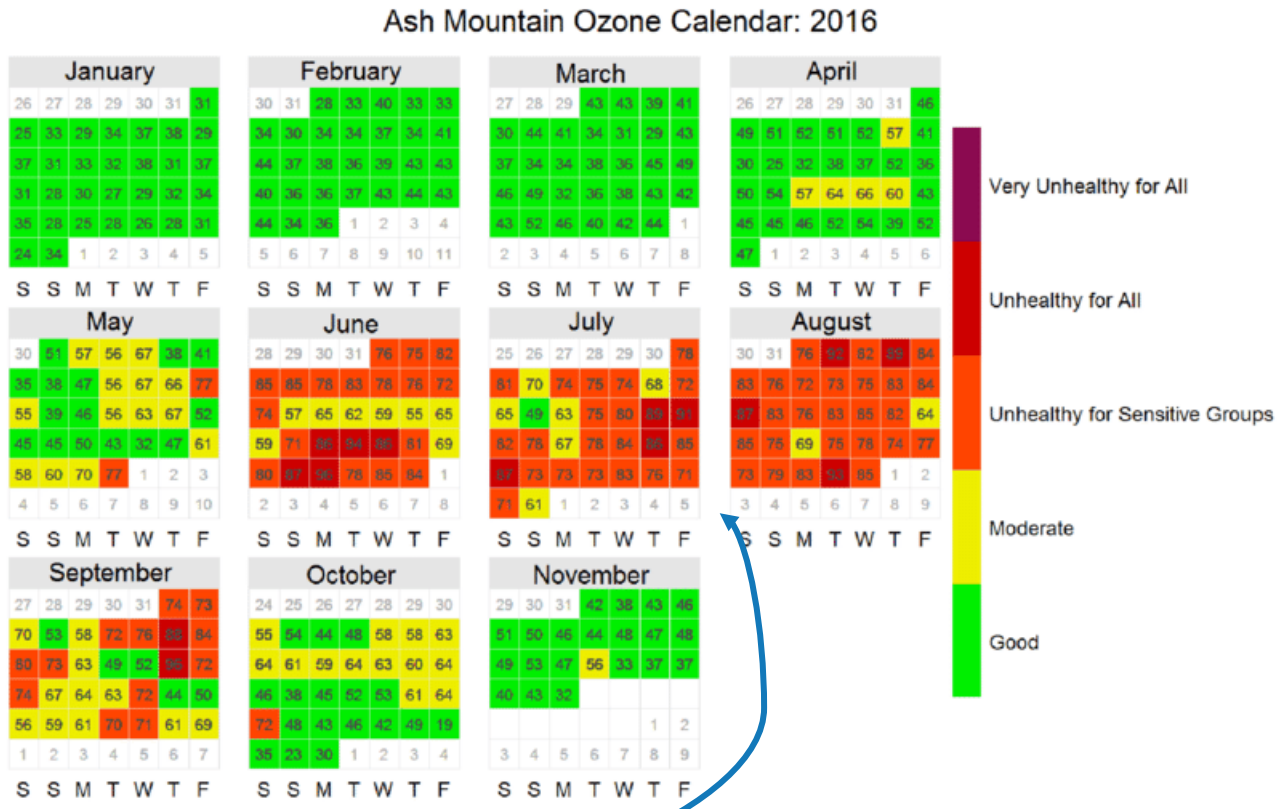
- Are there patterns when data is higher or lower over time?
- What trends appear seasonally? (e.g., during winter inversions)
- What trends could be caused by short-term events?

Image Source: TD Enviro

Data across Time – Calendar Plot

A 'big picture' look at air quality over a month or longer period

- More accessible "time series"
- Days can be described in terms of a range of concentrations or AQI
- Helps illustrate trends across seasons, but less granular information (e.g., exact concentrations, hourly spikes)
- One calendar plot per site/device



Calendar plot of ozone concentrations at Ash Mountain in Sequoia and Kings Canyon National Parks shows summer months experiencing the highest levels of ozone.

Image Source: Meyer, E. and Esperanza, A. National Park Service, 2017

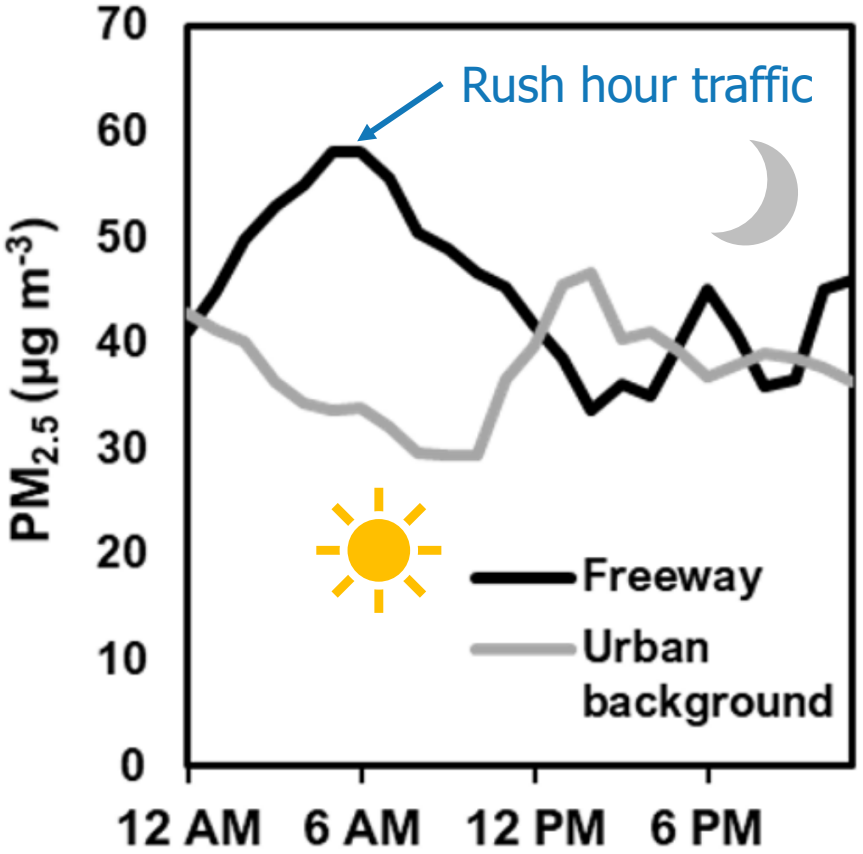
Data across Time – Diurnal/Weekly Trends

Group data into relevant periods, summarizing patterns across time

- Allows you to explore trends across:
  - Time of day (diurnal)
  - Days of the week
  - Months
  - Seasons

Questions to Ask:

- What scheduled activities/behaviors could be affecting pollutant levels (e.g., traffic, school, work, seasonal changes)?
- Do all sites show similar patterns over time?



Diurnal variation of $PM_{2.5}$ concentrations in Alhambra and Monterey Park, CA in 2019.

Image Source: Mousavi, Wu 2021

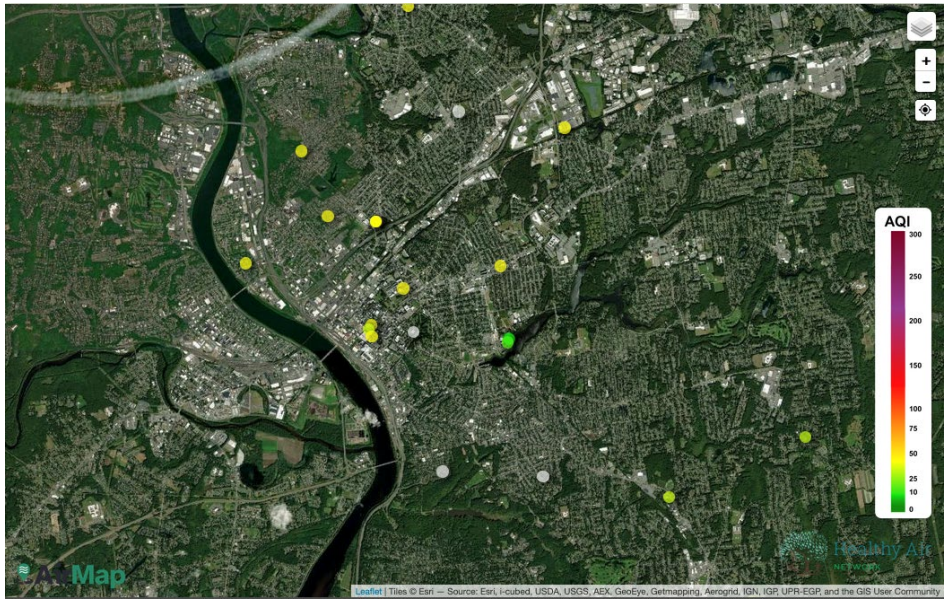
Data across Space - Maps

Show the spatial patterns of data across the project area

- Puts data into spatial context
- Can be enriched with other spatial data (e.g., emissions source locations)
- Critical for viewing mobile data!

Questions to Ask:

- Where do high (or low) concentrations occur? Are there patterns or clusters?
- Does supporting data (e.g., local sources, nearby roads) provide insight into the patterns that you see?



$PM_{2.5}$ AQI levels across Massachusetts on 01/05/2025

Image Source: Healthy Air Network

Mobile $PM_{2.5}$ inside waste trucks along routes in New York, 2016

Image Source: Transform Don't Trash, 2016

Data across Space – Pollution Roses

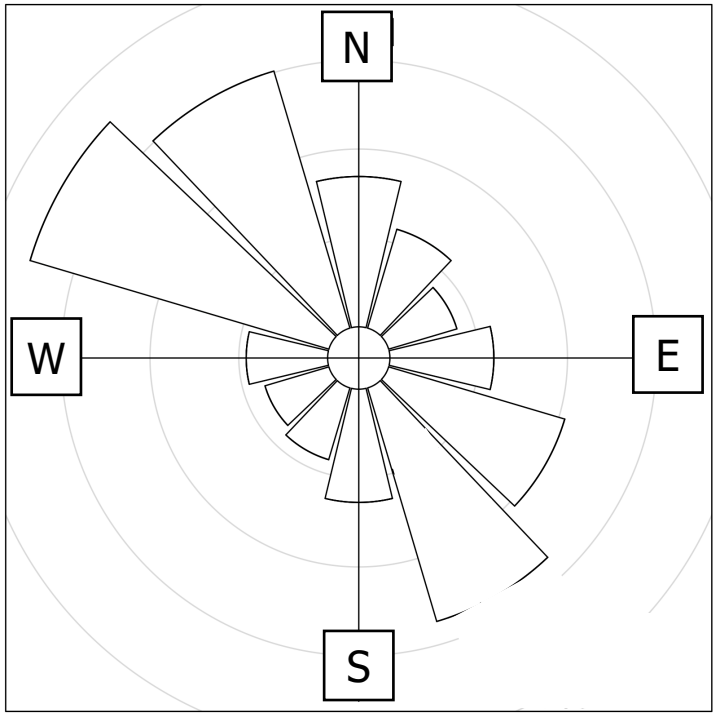
Explore where pollutants may be coming from

- Shows the relationship between pollutant concentrations and wind direction
- Can provide insights into pollution sources when paired with maps/spatial context
- Require reliable local weather data
- Not always easy to interpret
- Requires months/seasons of data to summarize trends (matching amount of pollutant data)

Let’s build a pollution rose step by step

Data across Space – Pollution Roses

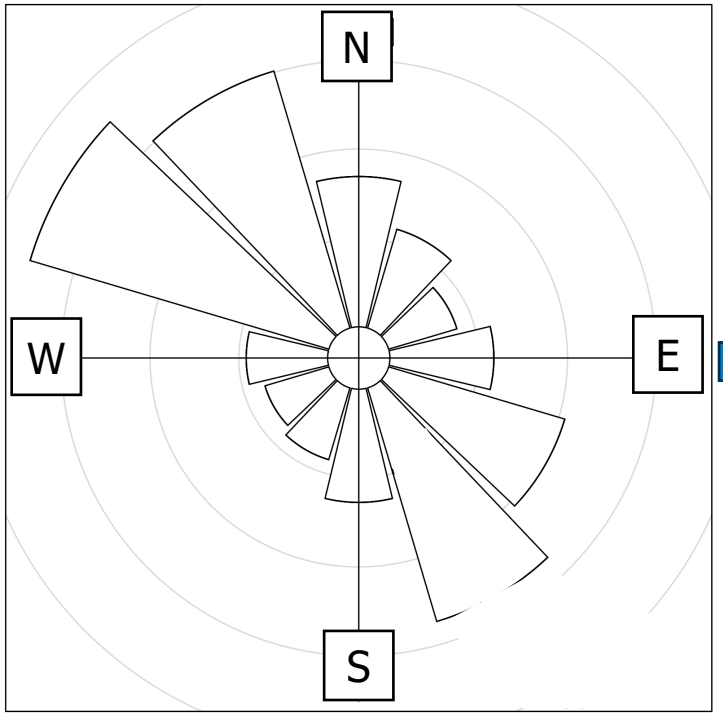
Basic wind rose



Basic wind rose shows that winds predominantly come from Northwest and Southeast

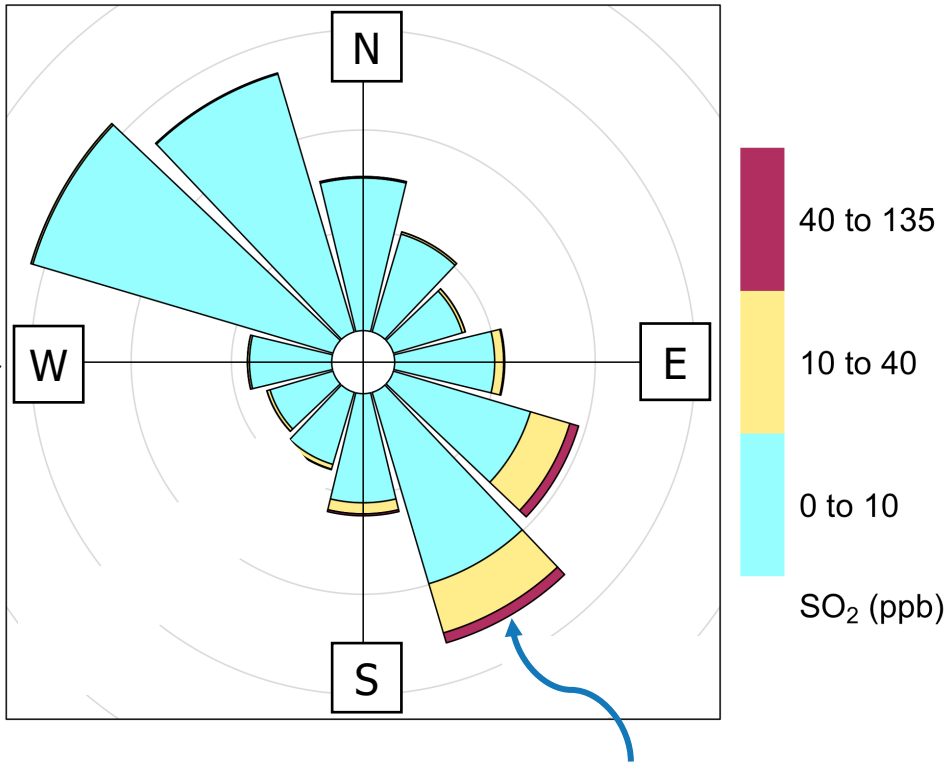
Data across Space – Pollution Roses

Basic wind rose



Basic wind rose shows that winds predominantly come from Northwest and Southeast

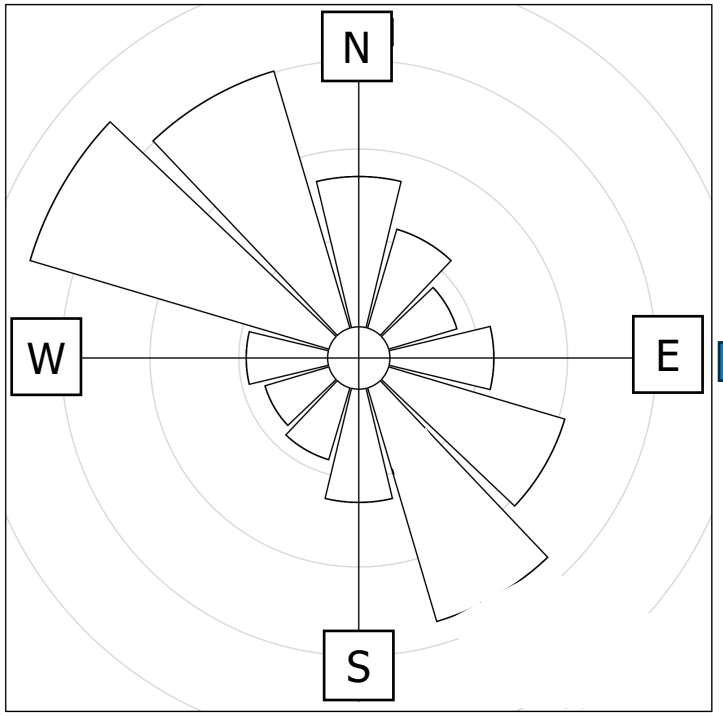
Pollution rose - SO_2



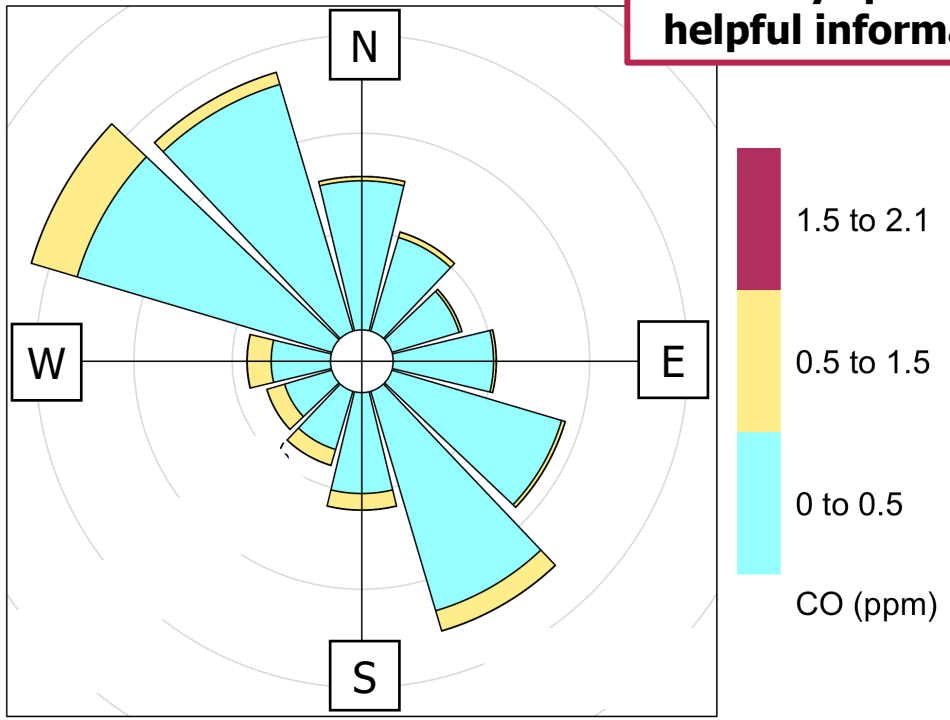
Pollution rose shows that higher SO_2 levels often are measured when winds are coming from the Southeast – could indicate the source of the pollutant!

Data across Space – Pollution Roses

Basic wind rose



Pollution rose - CO



Pollution roses won't always provide helpful information!

Basic wind rose shows that winds predominantly come from Northwest and Southeast

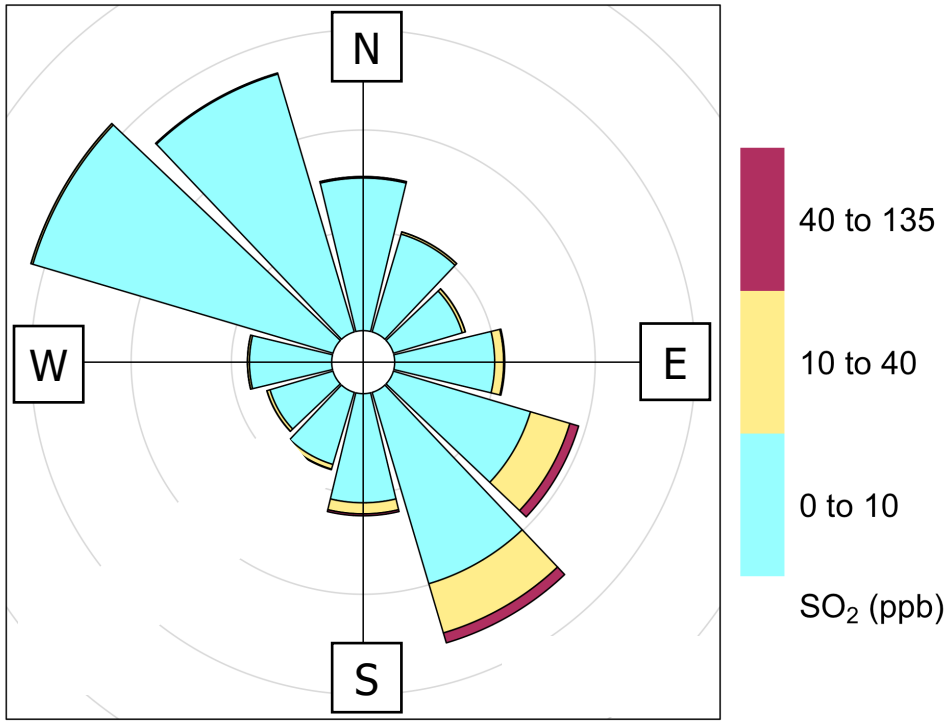
Pollution rose shows that higher CO levels come from a variety of wind directions, which doesn't help isolate pollutant source locations.

Data across Space – Pollution Roses

Questions to ask:

- What direction is the wind predominantly coming from? How does this vary across seasons?
- What is the common wind direction associated with higher pollutant levels?
- Are all pollutants coming from the same wind direction?

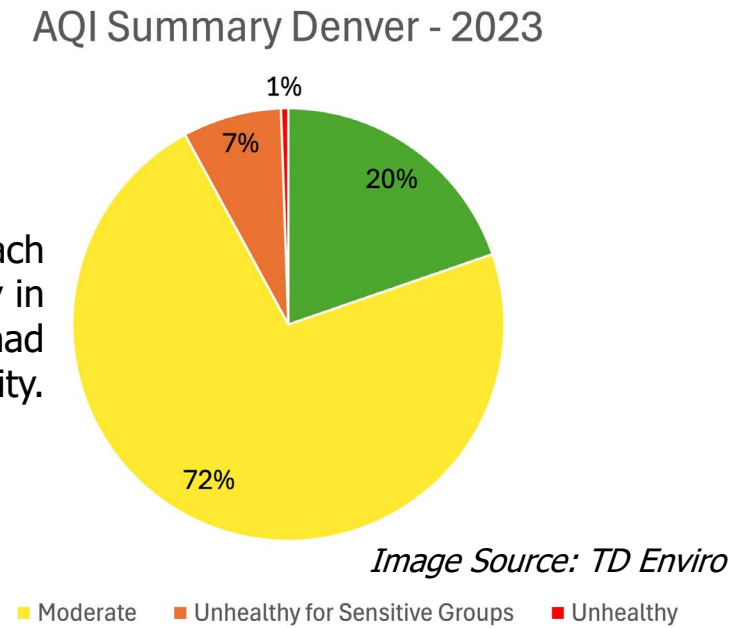
Pollution rose - SO_2



Summaries and Other Visuals

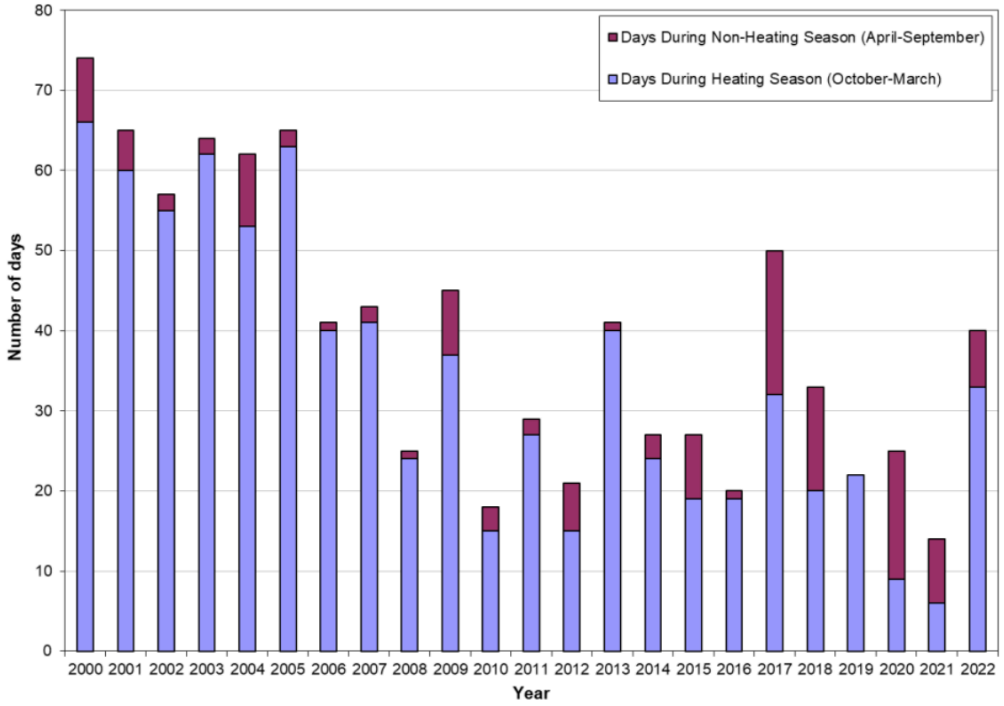
Useful visuals for summarizing air quality information:

- Bar charts – summarize data across groups and/or time, similar to time series, but can be more visually accessible.
- Pie charts – show percentages of data across relevant categories (e.g., AQI)



Pie chart showing the % of days in each AQI category for Denver air quality in 2023. The majority of days had moderate air quality.

Image Source: TD Enviro



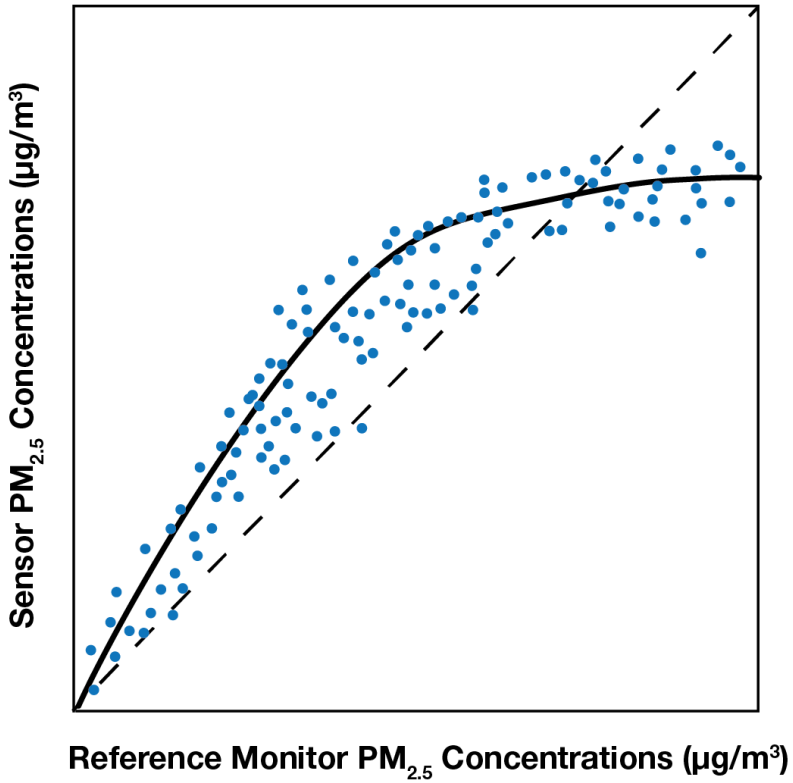
Bar chart showing the # of days $PM_{2.5}$ exceeded Seattle's health goal across the years, separating heating/non-heating seasons.

Image Source: Puget Sound Clean Air Agency

Summaries and Other Visuals

Scatter Plots

- Show the relationship between two parameters
- Helpful for evaluating collocation data and investigating relationships
- Technical to interpret



Scatter plot showing that an air sensor has a linear response at lower concentrations and a non-linear response at higher concentrations

Image Source: EPA

Visualization Best Practices

- Are your units and time stamps correct?
- Do you have clear and descriptive labels?
- Do the colors make sense?
  - Colors can lead people to certain conclusions – e.g. red = bad and green = good
- Is this visually accessible for my audience?
  - Font size, color contrast, alt text
- Before sharing widely, show your visuals to someone who knows little about your project. Are they able to understand the point of the visual?

Chart Your Path

How do we choose which visualizations to use?

With many possible visuals, it can be challenging to know where to start.

Revisit your objective and go through these thought exercises with your team:

1. What are some desired outcomes from sharing data?
   a. What information is most important to share?
   b. What main comparisons are you making? (across time, locations, to thresholds?)
2. Who is your audience?
3. What resources (e.g., data visualization tools) are available?
4. What is your team's capacity and experience?
5. Based on this discussion, what possible visualizations would support your objective?

Identifying your analysis from your objective…
Objective 1: Informational/Awareness

Project example objective:

A technical data team from a local air agency has partnered with a neighborhood association in a large metro area and is monitoring $PM_{2.5}$ and NO_2 levels in their neighborhood. The air agency is concerned that levels measured at the nearest regulatory monitors 10 miles away may not be representative of the local air quality. The local air agency hopes that data collection will (1) help increase awareness around local air quality issues, (2) promote engagement with local youth, elderly, parents, and other local officials through air quality data and (3) help others inform their personal decisions.

Questions:

1. What are some desired outcomes from sharing data?
   a. What information is most important to share?
      What main comparisons are you making?
2. Who is your audience?
3. What resources are available? What is your team's capacity?

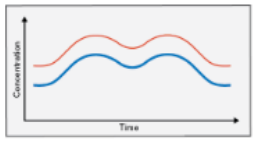
Identifying your analysis from your objective…
Objective 1: Informational/Awareness

Project example objective:

A technical data team from a local air agency has partnered with a neighborhood association in a large metro area and is monitoring $PM_{2.5}$ and NO_2 levels in their neighborhood. The air agency is concerned that levels measured at the nearest regulatory monitors 10 miles away may not be representative of the local air quality. The local air agency hopes that data collection will (1) help increase awareness around local air quality issues, (2) promote engagement with local youth, elderly, parents, and other local officials through air quality data and (3) help others inform their personal decisions.

Map summarizing pollutant data across space

Time series of recent (month) levels, including comparison to nearby regulatory data



Accessible visuals, using AQI, and summaries of data where possible

Real-time pollutant data to inform action

Example: AirNow Fire & Smoke Map

Public map that shows $PM_{2.5}$ data from stationary long-term and temporary monitoring stations as well as sensors.

Data can be explored through:

- Time series
- Maps
- AQI and concentrations

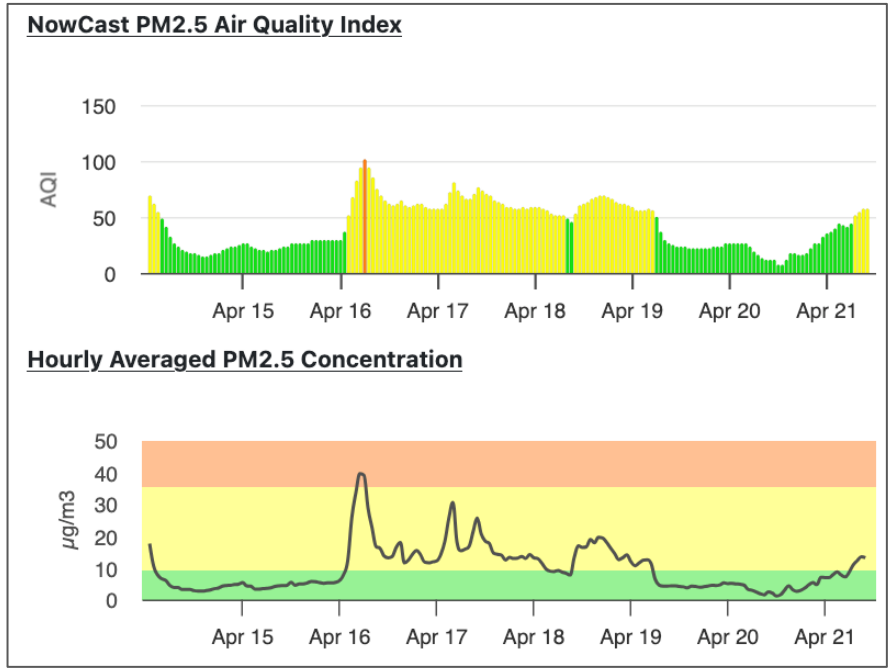


Image Source: AirNow Fire and Smoke Map

Identifying your analysis from your objective
Objective 2: Source Information

Project example objective:

A local air agency installed black carbon sensors in a neighborhood where regulatory data is lacking to better understand levels across space and understand how local sources of concern may impact black carbon concentrations. There is also a local weather station that provides wind speed and direction data. The air agency has a technical data analysis team and intends to share project results at city council and neighborhood meetings to raise awareness around local pollution sources and support further monitoring.

Questions:

1. What are some desired outcomes from sharing data?
   a. What information is most important to share?
      What main comparisons are you making?
2. Who is your audience?
3. What resources are available? What is your team's capacity?

Identifying your analysis from your objective
Objective 2: Source Information

Project example objective:

A local air agency installed black carbon sensors in a neighborhood where regulatory data is lacking to better understand levels across space and understand how local sources of concern may impact black carbon concentrations. There is also a local weather station that provides wind speed and direction data. The air agency has a technical data analysis team and intends to share project results at city council and neighborhood meetings to raise awareness around local pollution sources and support further monitoring.

Map of and average black carbon levels and related pollution sources in the neighborhood (e.g., freeways, truck stops)

How black carbon levels vary with wind direction, informing possible pollutant transport and source information

Accessible data visuals for meetings, and description of analysis and QA/QC for officials

Example: Colorado Fenceline Monitoring

The Colorado Department of Public Health and Environment deployed monitors to measure total volatile organic compounds (tVOCs) around well pads to identify pollution from these sources.

Data was explored by researchers on the project team, visualizing data through:

- Pollution roses
- Maps
- Time series

Pollution roses show higher tVOC concentrations when winds are coming from the direction of both well pads

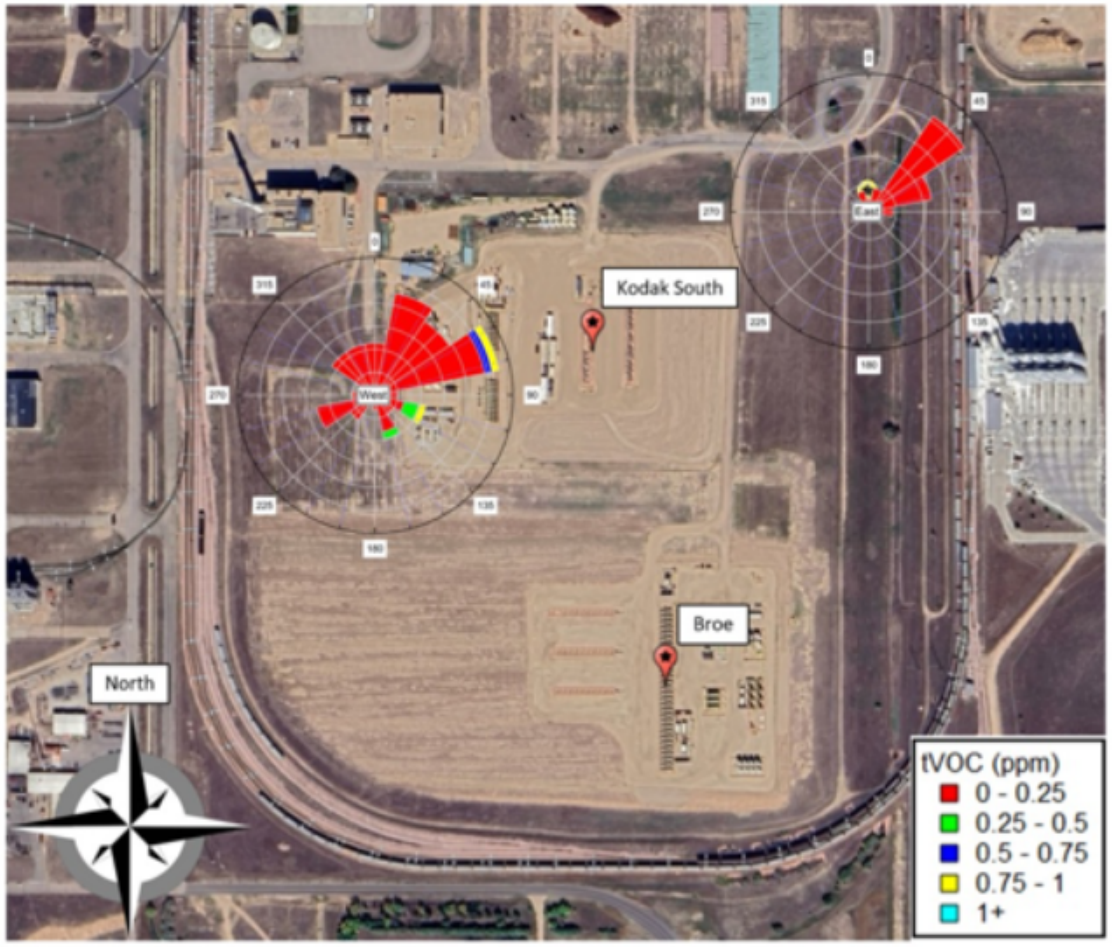


Image Source: CDPHE

Tips

- Start discussions about visualizing data early on - before you collect it.
- More complex ≠ more effective
- Explore your data – analysis isn’t straightforward!
- Don’t rush this process
- Include your whole team (context and feedback is crucial!)
- Revisit your objective before sharing your visuals - ask “Does the visual communicate the key project insights effectively?”

Break

Interpreting, Drawing Conclusions, & Communication

Three Ways to Interpret Data

1. Big-to-small (or small-to-big)
2. Forensic
3. Comparison

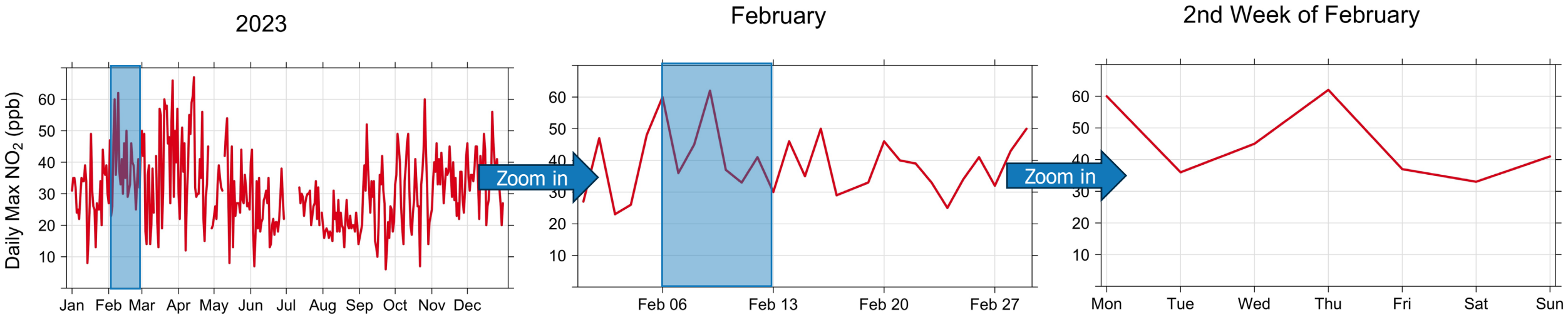
Interpret #1: Big-to-Small & Small-to-Big

Explore the data, looking at the "big picture" and progressively focusing on more detail and the finer picture of your data. Vice versa, start small and work to more larger views of your data.

Approach (big-to-small)

1. Look at long-term time series plots (yearly or seasonal)
2. Focus on monthly, then weekly, then daily, and hourly
3. Look for:
   - Trends over time – are they explainable? How does air quality change? And why?
   - Where are the consistencies/inconsistencies?
   - Are there any anomalies? What might have caused them?
   - Are there seasonal, weekday vs. weekend, or day-night (diurnal) differences?
   - Does air quality change rapidly? Are there any unexpected changes?
   - How does air quality change with the weather?

Interpret #1: Big-to-Small & Small-to-Big



Lots of variation with higher levels in the winter/spring and dips in NO_2 in the summer.

Variation across the month with some higher levels of NO_2 in the first half.

Monday and Thursday see the highest NO_2. The weekend is lower than the weekdays.

Interpret #2: Forensic

Forensic data analysis involves examining your data for clues that support (or refute) your question or concern. Like any forensic analysis, you will need to reconstruct what happened with your air quality data and other supporting data (weather, traffic, source locations, etc.).

Approach

1. Describe how and when the clues or "fingerprints" might appear in your data.
2. Gather other supporting information to support your case.
3. Generate plots of your data.
4. Look for:
   - Times when clues appear. When do they not appear?
   - Does your supporting information provide more insights and help explain the air quality?
   - Are there other times and locations that support your case?
   - Is it convincing? Are there alternative explanations?
   - What evidence is not convincing? How could you address those factors?
   - Can you reconstruct the event and what happened?

Interpret #2: Forensic - Example

AQEarth West Atlanta study

Observed high spikes in vehicle-related pollution (nitric oxide, NO) in neighborhood

Forensic Steps

1. Looked at surrounding pollution sources → found busy truck stops and roadways
2. Compared daily traffic levels across nearby roads → lower traffic density near monitor

Outcome: Neighborhood NO levels were sometimes comparable to highways during certain conditions.

Not conclusive → more forensics required (next step, black carbon data collection)

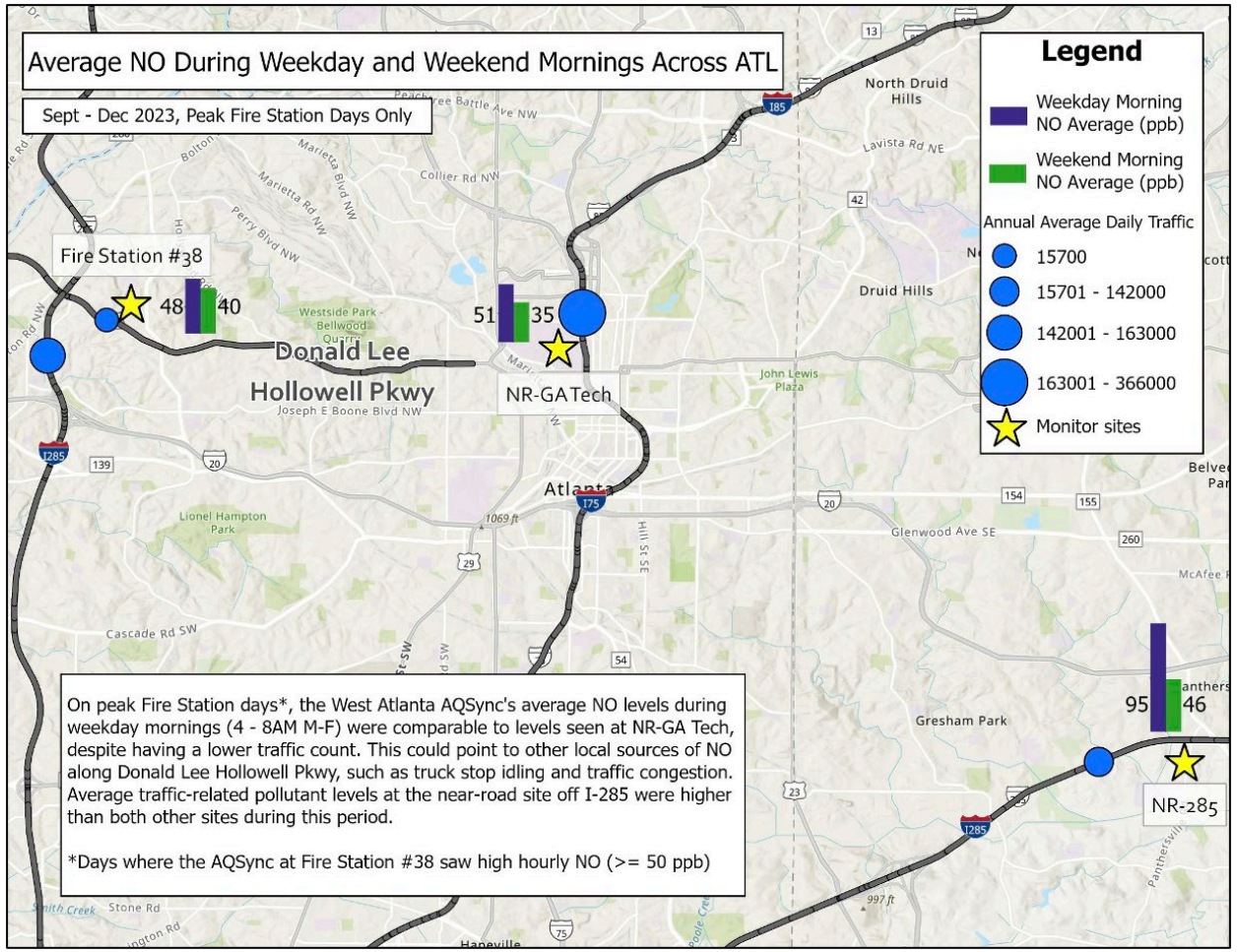


Image Source: TD Enviro

Interpret #3: Comparison

Comparisons allow you to detect differences in air quality over time, across space, and relative to standards. This information can help you determine which times and areas have higher or lower air quality or how your air quality compares to standards.

Approach

1. Over time: Compare one period to another.
2. Across space: Compare one location to another (monthly, seasonal, and annual averages helps draw out differences).
3. To standards: Compare your data to benchmark concentrations. Ensure you process data according to the ”form” of the standard. The “form” describes the statistical method for averaging and processing data to compare it to the standard.

Ask (across space example):

- How do locations compare?
- What potential sources affect one location vs another location?
- Are there other possible explanations?

Interpret #3: Comparison - Example

Neighborhood Air Sensor Analysis

- Belle Haven neighborhood of Menlo Park
- Wanted to know what air sensors showed for this impacted neighborhood
- 15 air sensors within 1 mile
- Processing:
  - Acquired, QC'd, and averaged data
  - Annual average chosen
  - Averaged data for three areas
- Presented to neighborhood and elected officials

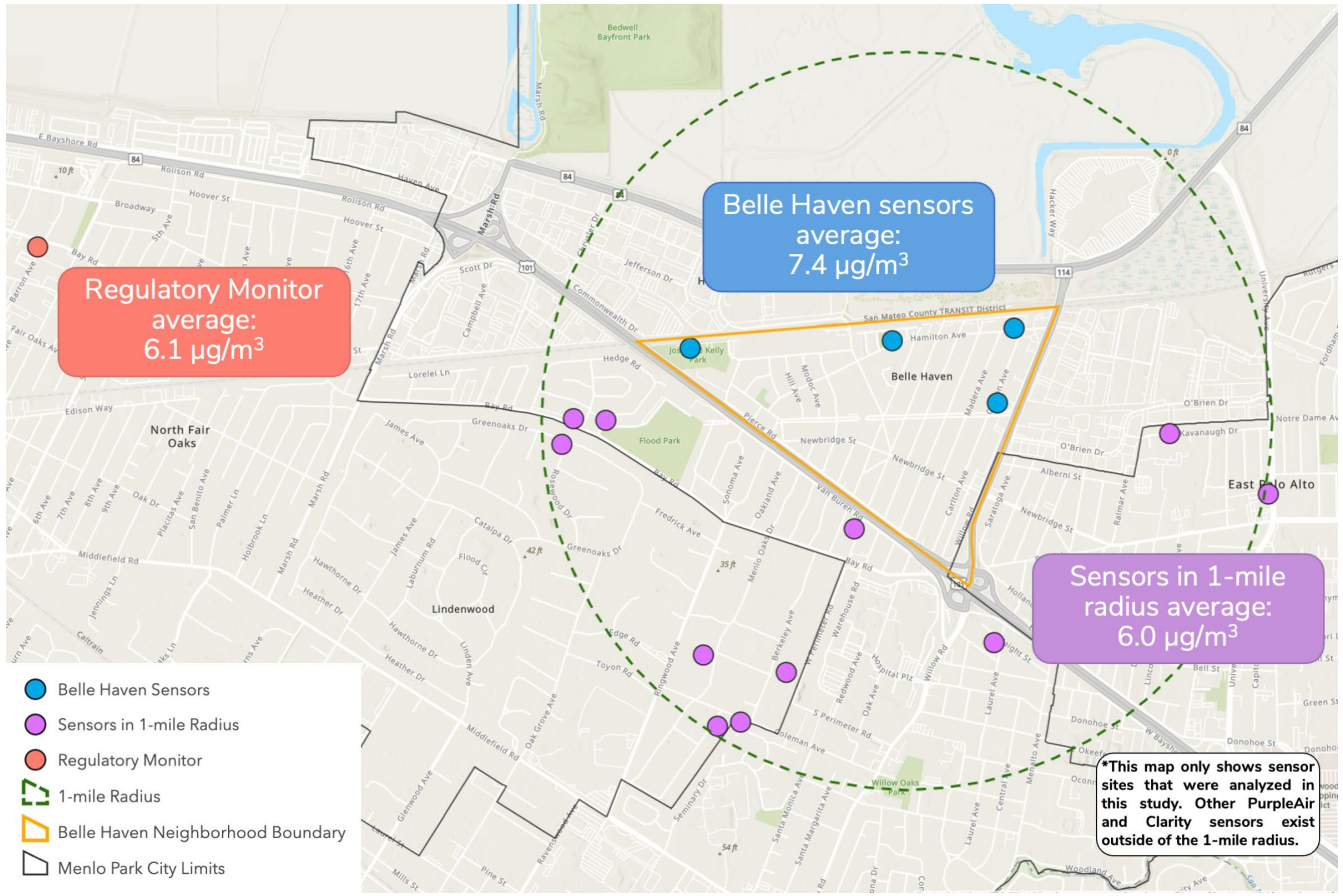


Image Source: Bay Air Center

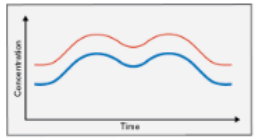
Tips for Interpreting Data

1. Don’t rush to conclusions.
2. Let the data speak for itself and be careful not to over-interpret (or put too much weight on the results).
3. Find someone to show and describe what you see in the data – ask them to be critical of your insights.
4. Label the plots with the conclusion(s).

Communicating Results:

All these steps allow others to gain confidence in your data and results!

1. **State your Purpose/Objective** - Communicate why you conducted the study and why you made some of your decisions. Describe how your device data support meeting the project objective(s).
2. **Describe the Monitoring Setup and Data Collection** - Provide a description of where sensors were located and data collected. Make sure you include supporting information (e.g., device locations, site photographs, QC checks, time stamps, units).
3. **Describe the Data Processing and Analysis** - Share the methods used for data cleaning, corrections/adjustments, QC, data analysis and interpretation, and limitations of the data and your devices.
4. **Visualize Data and Share Results** - Be sure to include main take-aways and call to action or next steps for the project.



Tip: Know your audience before presenting your results!

- Who is in your audience? For example, a scientist may want technical details while others may want a big-picture summary.
- How much information do they want?
- What type of visuals may work best for sharing results?

Project Narrative

+

Visuals

Sharing Project Story, Findings and Next Steps

Analysis Tools

Tons of Tools: The Highlight

- Instrument vendor or data management dashboards
- EPA tools
- Excel & Google sheets
- Programming tools

- **What can they do?**
- **What are they useful for?**
- **What level of effort or skills are required to use them?**

Dashboard Analysis

Instrument dashboards and data management platforms

What can they do?

Depends on the provider:

- Map air monitoring sites
- Make basic graphs (time series, bar charts)
- Create reports
- Download data
- Notifications if suspect data/sensor offline

What are they useful for?

- Regular data review (are there any issues?)
- Reviewing a spike or other incident
- Comparing across time and between sites

What level of effort or skills are required to use them?

Minimal

Fire and Smoke Map - fire.airnow.gov

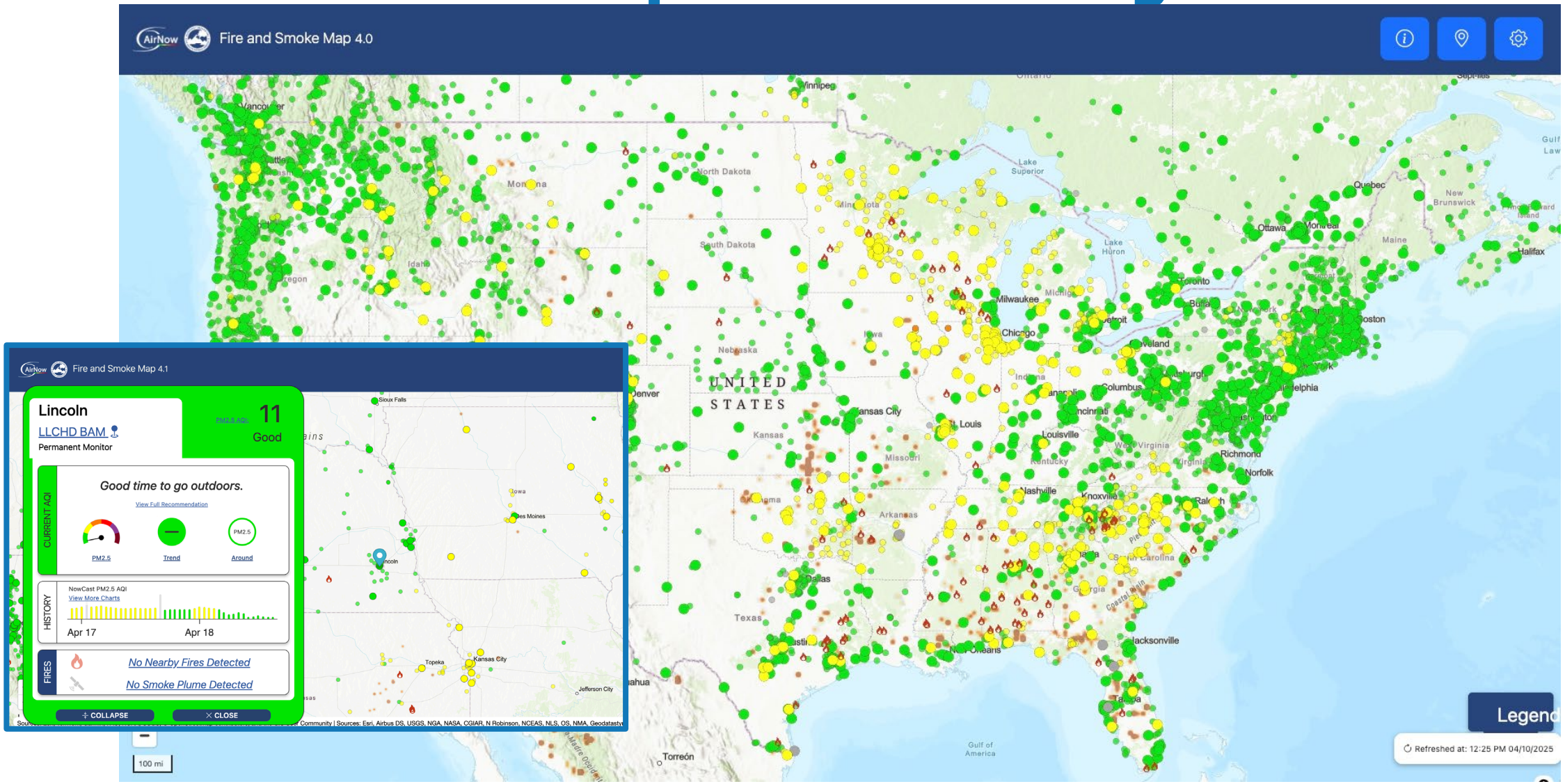


Image source: EPA

EPA Tools

EPA has developed several free helpful analysis tools.

What can they do?

AirData:

- Access and download *regulatory* data
- Data visualization tools
- Create summary reports

Real Time Geospatial Data Viewer (RETIGO):

- Web-based tool with clickable interface
- Explore stationary or mobile air quality data you've collected (need to QC)
- Create maps, timeseries, scatterplots, wind-pollution plots

Air Sensor Network Analysis Tool (ASNAT):

- Data review and QC
- Data visualization

Air Sensor Data Unifier (ASDU):

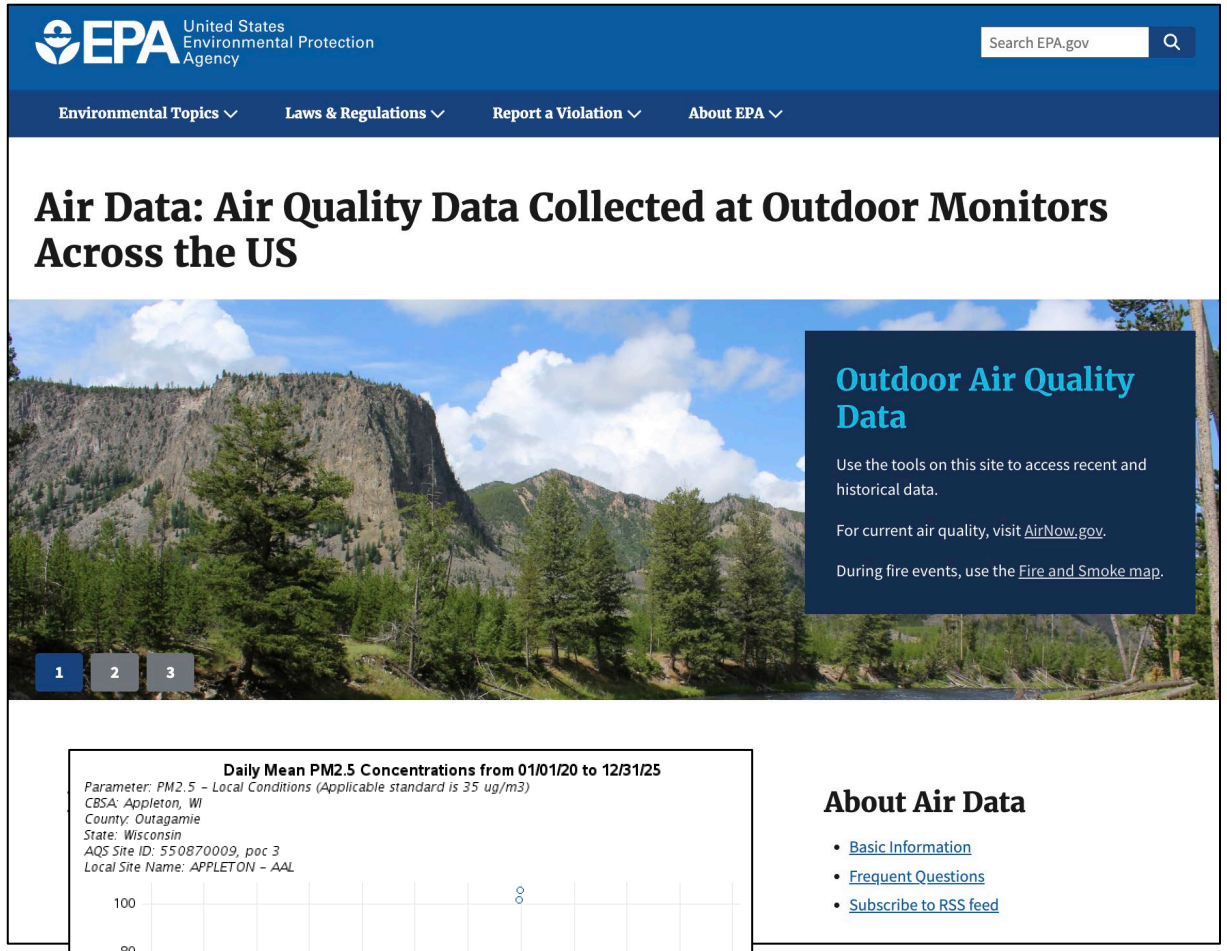
- Data formatting into ASNAT, RETIGO, and other formats

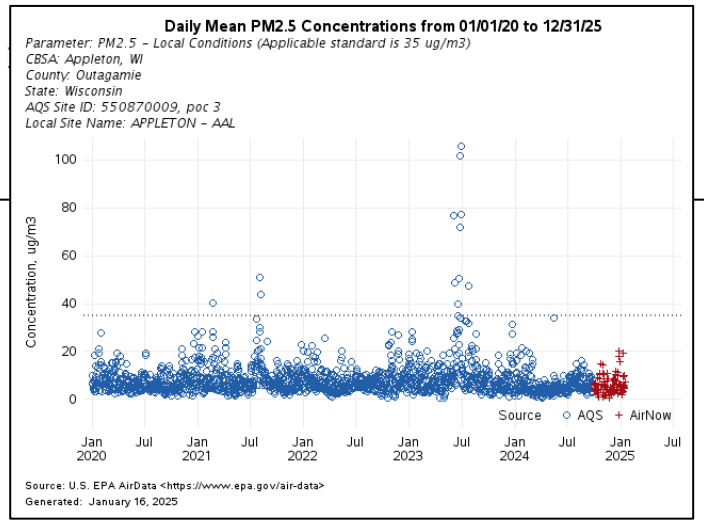
What are they useful for?

- Understanding local air quality (AirData, RETIGO, ASNAT)
- Comparing sites and variables (AirData, RETIGO, ASNAT)
- Visualizing & investigating mobile data (RETIGO)
- Understanding sensor accuracy (ASNAT)

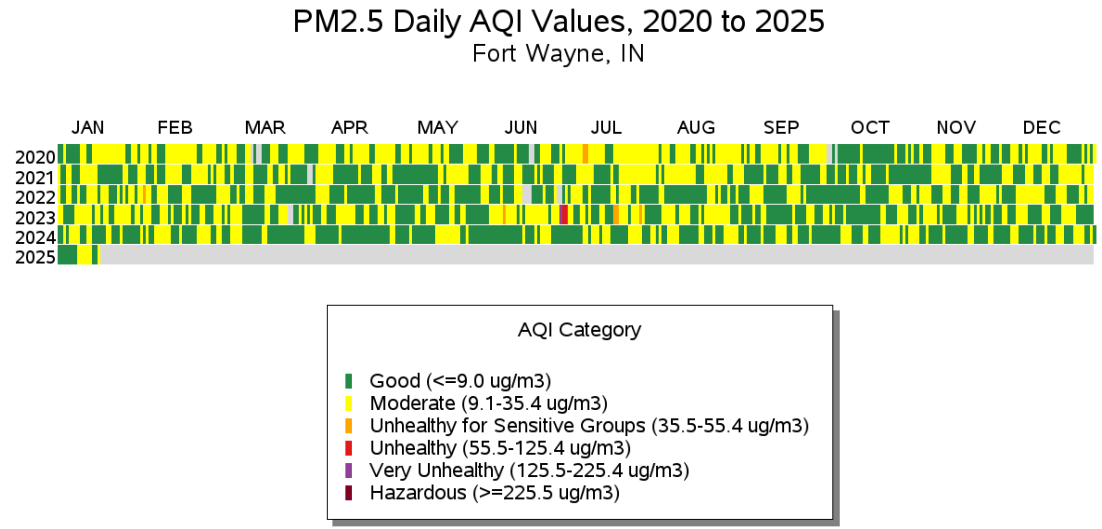
What level of effort or skills are required to use them?

- Some background knowledge about air quality is needed to make correct assumptions and use results (ASNAT)
- R and Rtools installation required (ASNAT, ASDU)
- May need to adjust tricky data formats (RETIGO, ASNAT)





- Access and analyze all US regulatory data
- View where regulatory monitors are located and what pollutants are measured
- Use analysis tools – calendar/tile plot, concentration time series, daily air quality tracker



Source: U.S. EPA AirData <https://www.epa.gov/air-data>
Generated: January 16, 2025

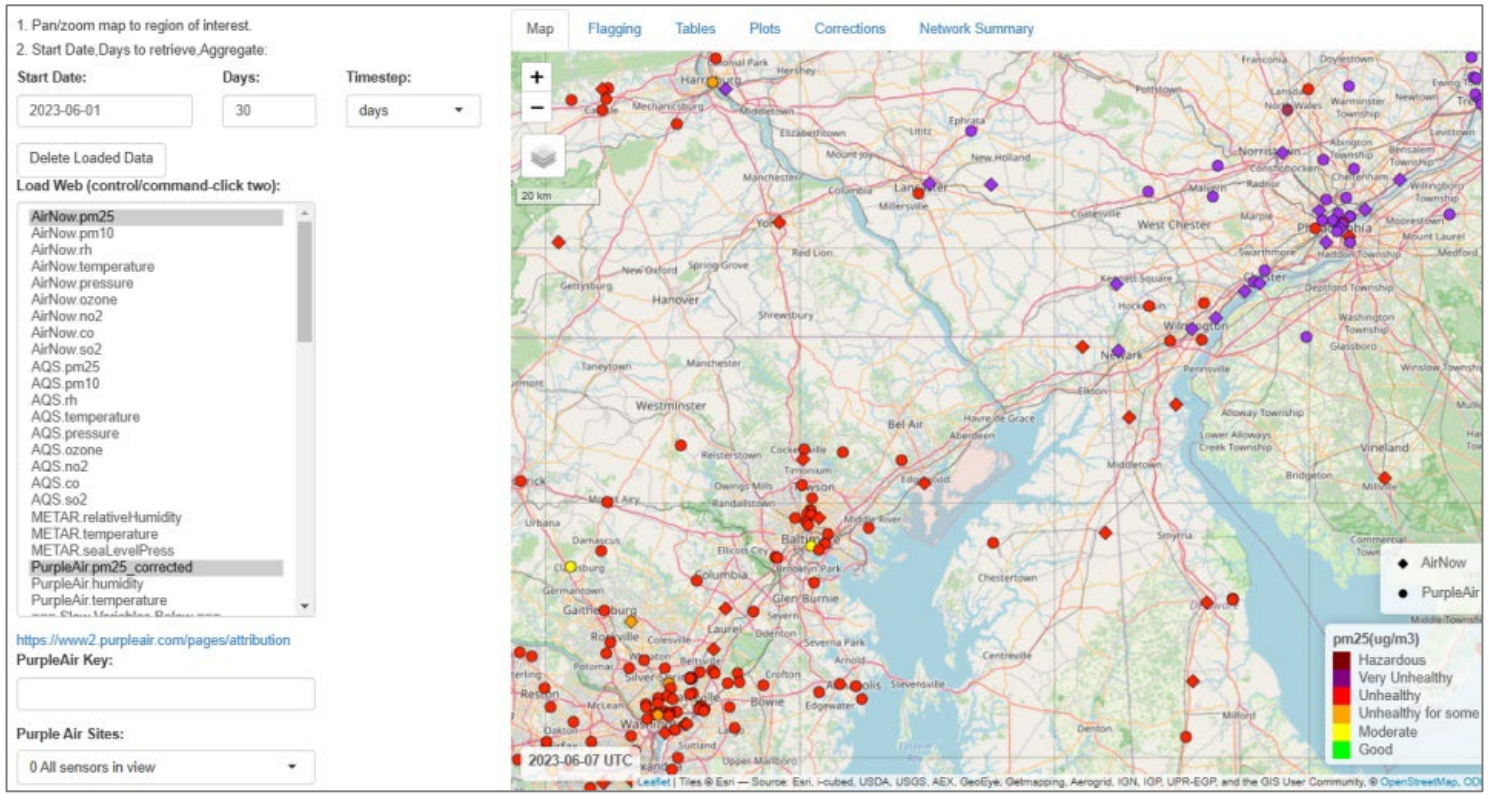
Image source: US EPA

RETIGO

Compare measurements with other publicly available data (including regulatory data, some sensor data, weather stations, satellite data products and imagery)

Image source: US EPA

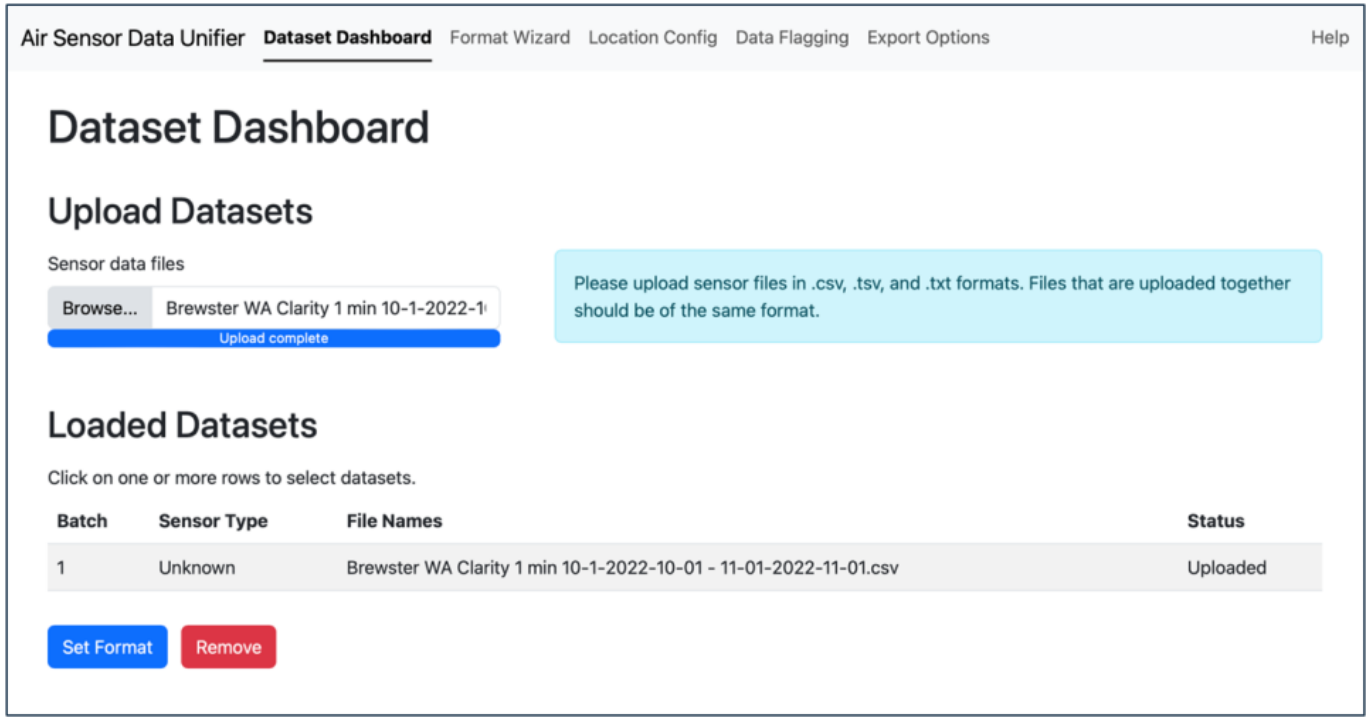
Air Sensor Network Analysis Tool



- Time series, maps, summaries to explore to air quality data from air sensors, government monitors, and weather stations
- Primarily used to evaluate sensor performance with nearest air monitors
- Basic quality control functions
- Target audience: air quality professionals (some background knowledge needed)
- Requires R (free, open-source)

Image source: US EPA

Air Sensor Data Unifier



- Accepts text format data (i.e., .csv, .txt, .tsv).
- Data flagging
- Data averaging
- Export options
  - ASNAT
  - RETIGO
  - Keyhole markup language (KML)
- Requires R (free, open-source)

Image source: US EPA

Excel & Google Sheets

Spreadsheet tools can do lots of different analyses and be very helpful in both determining averages and plotting data. Excel and Google Sheets are two common tools that have very similar functions.

What can they do?

- Analysis from raw data
- Many tools: pivot tables, bar charts, time series, pie charts
- Lots of resources including EPA's Macro Analysis Tool for comparing to regulatory data

What are they useful for?

- Averaging data (determining hourly/daily/annual averages to compare to NAAQS or other sites)
- Creating plots to help investigate one site or compare multiple sites

What level of effort or skills are required to use them?

Some familiarity is helpful, but there are lots of online resources to help you learn.

Analysis can range from limited skill level to advanced.

Programs

Free, open-source programming languages that support statistical computing and graphics. Common languages used to analyze air quality include R and Python.

What can they do?

- Analysis from raw data
- Nearly endless tools – you create what works for you!
- Openair, AirSensor – R packages
- EPA Sensortoolkit - Air Sensor Data Analysis Library (Github Python code library for evaluating air sensor data)

What are they useful for?

- Custom plots/charts
- Air quality-specific tools
- Intricate analysis and visualizations

What level of effort or skills are required to use them?

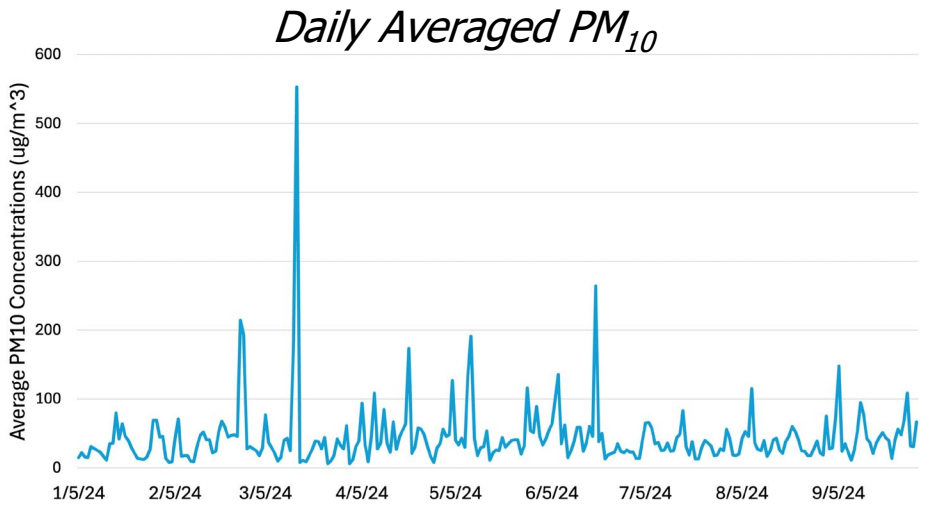
Programming languages take effort to learn. If you choose these tools, expect to spend significant time learning.

There are also many online courses to help you learn: EdX and Coursera, others.

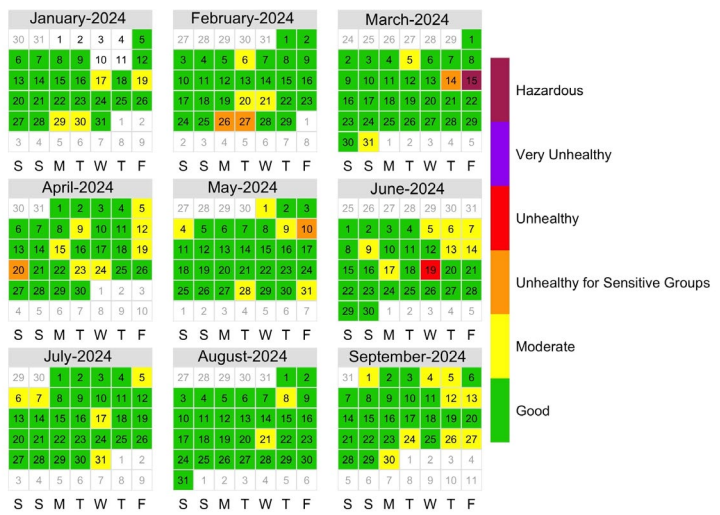
Data Visualization Quiz

An air agency has been monitoring PM_{10} at a neighborhood site for almost a year and wants to inform residents about general levels in their area at a neighborhood meeting. Residents are interested in trends over time and have less technical air quality background. Which data visualization would best support this objective?

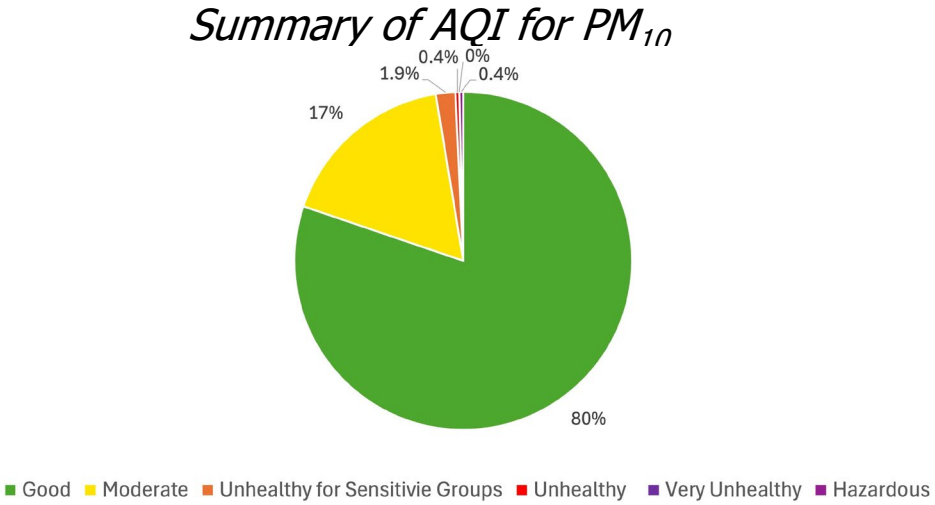
a)



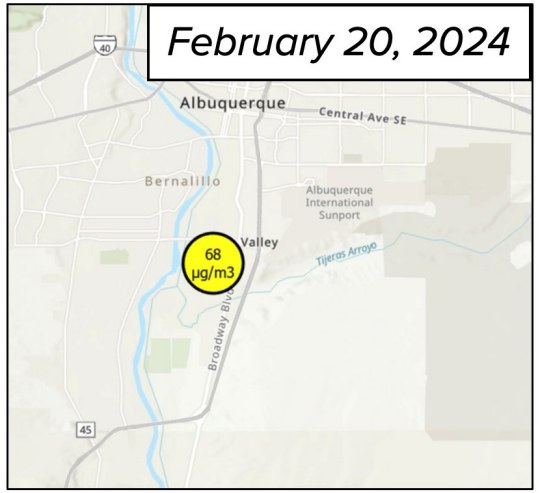
b)



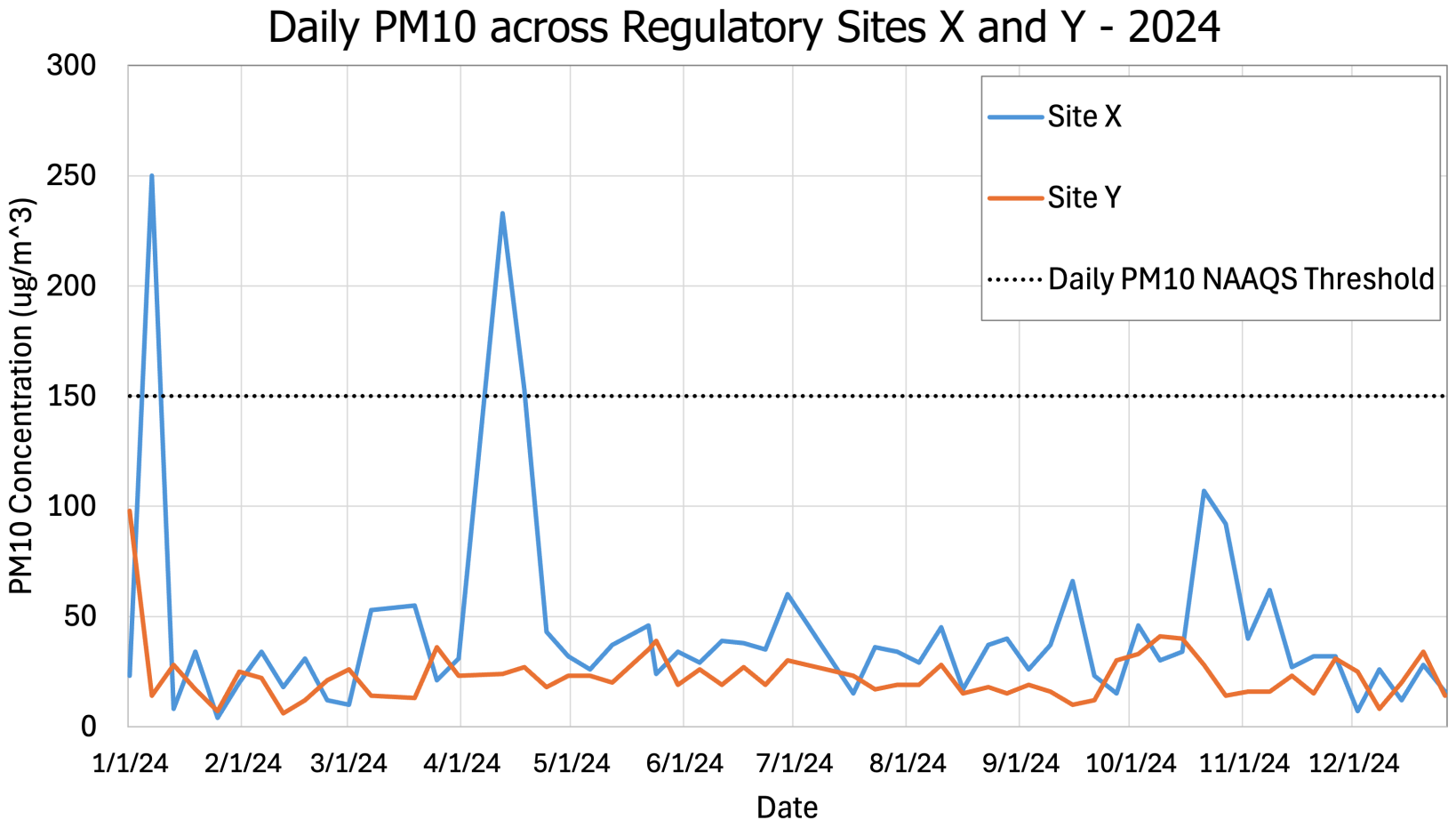
c)



d)

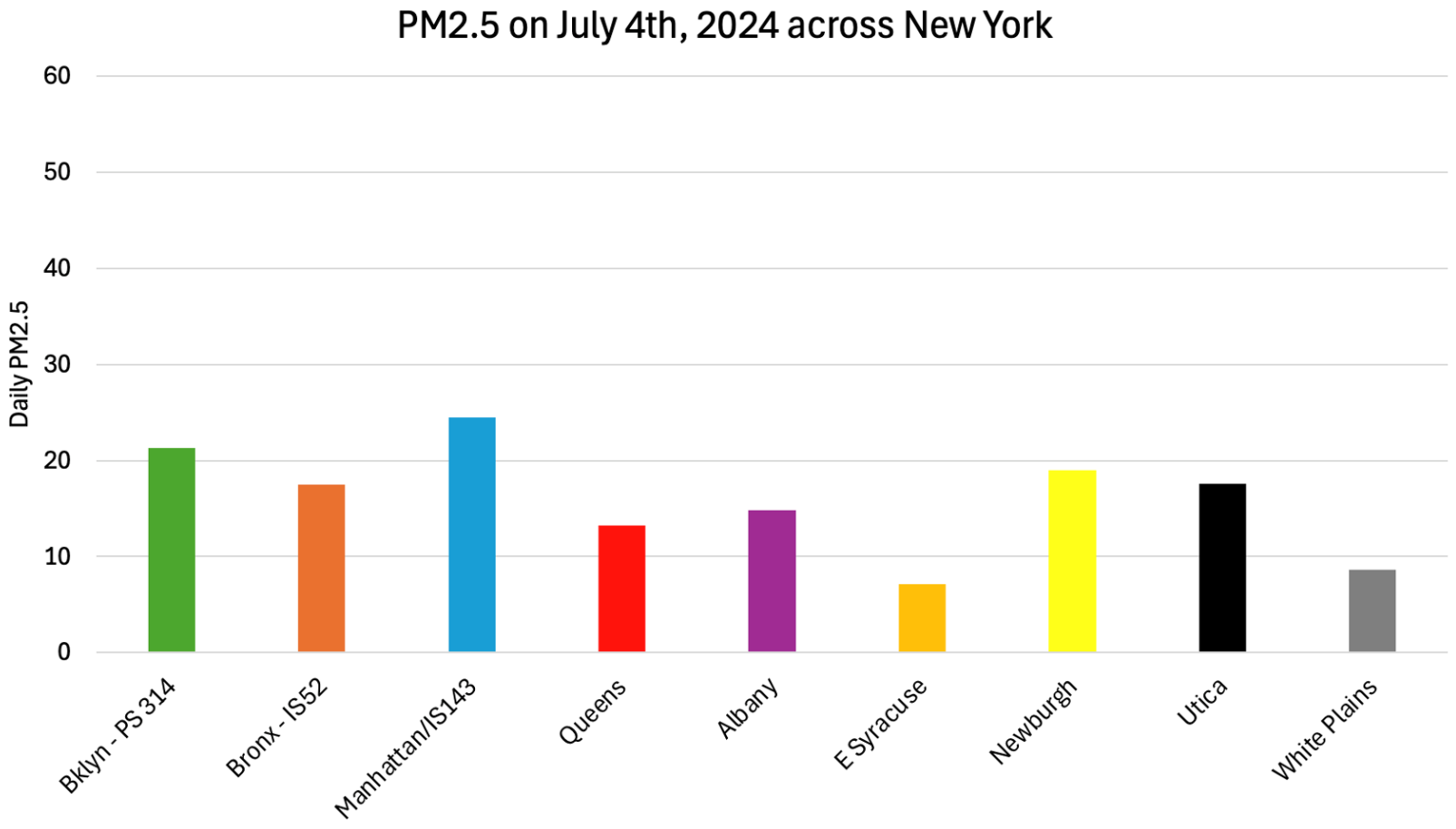


The time series below shows regulatory-grade PM_{10} monitoring data across two sites in 2024. What conclusion(s) can you draw from this data analysis?



a) The spikes in PM_{10} at Site X were caused by emissions from the nearby industry.
b) Something significantly impacted PM_{10} in January, April, and November, leading to higher levels than Site Y.
c) The levels at Site X are frequently higher than Site Y and frequently higher than the daily PM_{10} NAAQS health threshold.
d) All of the above.

What would help improve this bar plot?



a) Add pollutant units to y label
b) Change color of bars to relevant AQI category based on concentration
c) Lower scale (zoom in more)
d) All of the above

Recap

- Objectives & analysis
- Different types of visualizations
- Matching visualizations with your objective
- Interpretation and conclusions
- Data analysis tools

Interactive End-of-Session Feedback

What's the most useful thing you learned today?

Recap of Webinar Series

Webinar #1:

Data Wrangling: Accessing Data, Data Formats, and Quality Control

Webinar #2:

Gaining Insights from your Data: Visualization and Interpretation

Thank You!

Q & A