

# AirSensor Package for PurpleAir Data

Shared tools for a community of practice.



# Institutional Support

The initial development of this package was funded by the [Air Quality Sensor Performance Evaluation Center](#) at the South Coast Air Quality Management District with funds from an EPA STAR grant.

Additional funding was provided by the US Forest Service [AirFire](#) group in support of the Interagency Wildland Fire Air Quality Response Program.

[Mazama Science](#) develops and maintains the package as part of its ongoing relationships with federal, state and local air quality agencies.

# Shared Tools for Communities of Practice

R packages for air quality analysis.

# Why support open, shared tools?

Science at its best is:

- Open
- Transparent
- Reproducible

Shared tools improve communication:

- Among scientists
- Between scientists and citizens
- Among citizen scientists

***Vetted algorithms and standard plots build familiarity and trust.***

# Why support open, shared tools?

The “Data Deluge” is real.

Looking at PM2.5 measurements only:

AirNow 2018:

$$1112 \text{ sta.} \times \frac{1 \text{ float}}{\text{sta.} \cdot \text{hour}} \times \frac{4 \text{ bytes}}{\text{float}} \times \frac{24 \text{ hours}}{\text{day}} \times \frac{365 \text{ days}}{\text{year}} \approx \mathbf{39 \text{ Megabytes}}$$

PurpleAir 2018:

$$7123 \text{ sen.} \times \frac{90 \text{ float}}{\text{sen.} \cdot \text{hour}} \times \frac{4 \text{ bytes}}{\text{float}} \times \frac{24 \text{ hours}}{\text{day}} \times \frac{365 \text{ days}}{\text{year}} \approx \mathbf{22 \text{ Gigabytes}}$$

***New tools are needed to work efficiently with large datasets.***



# R & RStudio

## R statistical programming language

- Free
  - Open source
  - Cross platform
  - Powerful
  - Cutting edge
  - Popular
  - **Analyst centered**
- 
- Thousands of packages
  - High quality graphics

*Steep learning curve!*

## RStudio GUI for R

- Free
  - Open source
  - Cross platform
  - Powerful
  - Cutting edge
  - Popular
  - **Analyst centered**
- 
- Interactive graphics
  - Markdown documents
  - Debugging tools
  - Packaging tools
  - Documentation tools

*Makes everything easier!*

# AirSensor R package goals

## Data Model

- Synoptic data model
- Time series data model

## Data Ingest

- Synoptic data access from Purple Air
- Spatial metadata enhancement
- Time series data access from PA

## Data Analysis

- Outlier detection
- Smoothing, NowCast, etc.
- Linear fits
- State-of-Health metrics
- Comparison with federal monitors

## Data manipulation

- Subset sensors by time/location
- Subset sensors by data/metadata
- Convert sensors to hourly axis
- Work with “openair” package functions

## Data Visualization

- Maps
- Interactive maps
- Timeseries plots
- Interactive timeseries plots
- Community videos

## Ease of Use

- Pipeline style coding

# Working with PurpleAir Synoptic (PAS) data

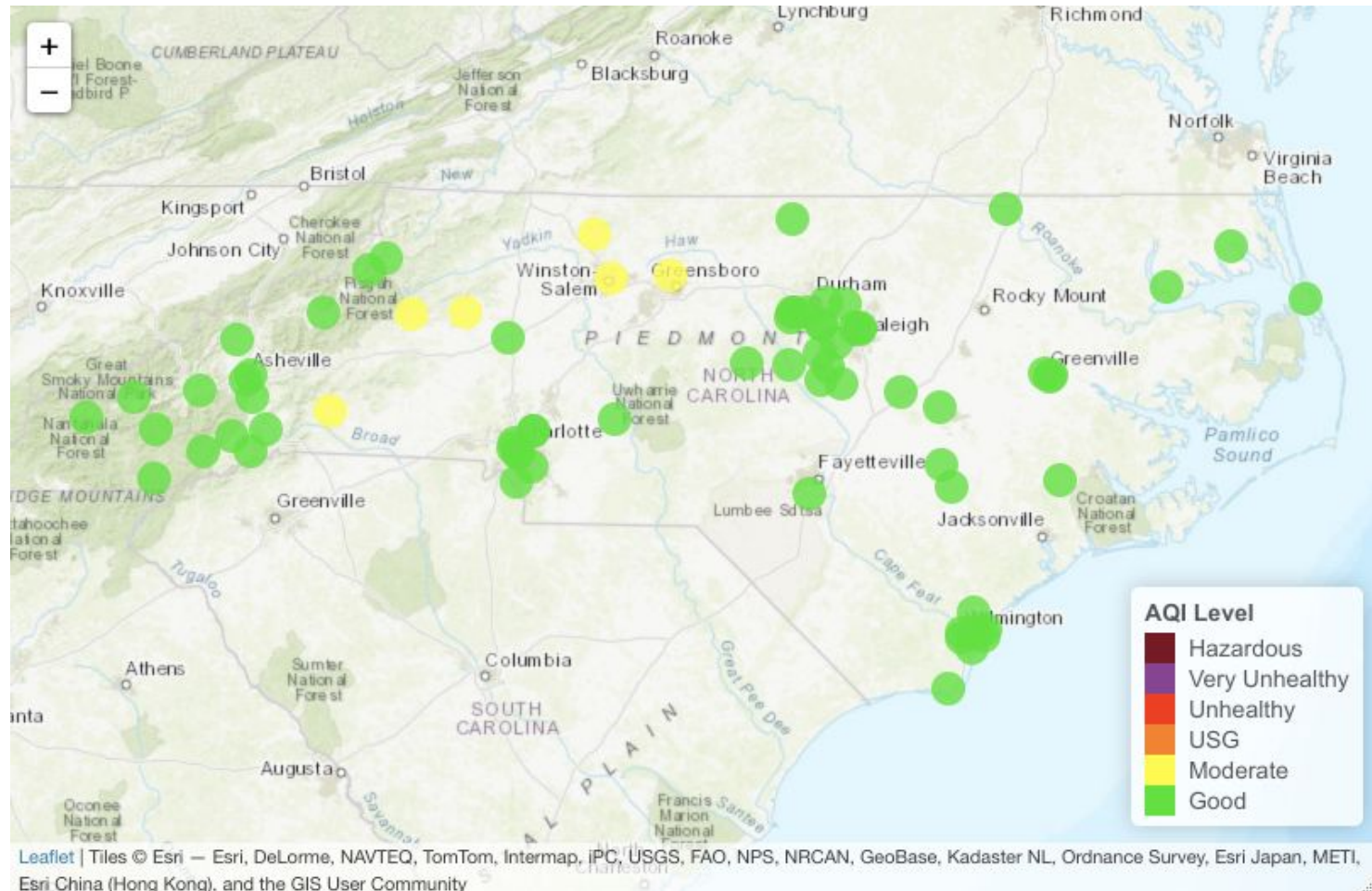
- Same data as shown on PurpleAir map
- Enhanced with additional spatial metadata
- Interactive map



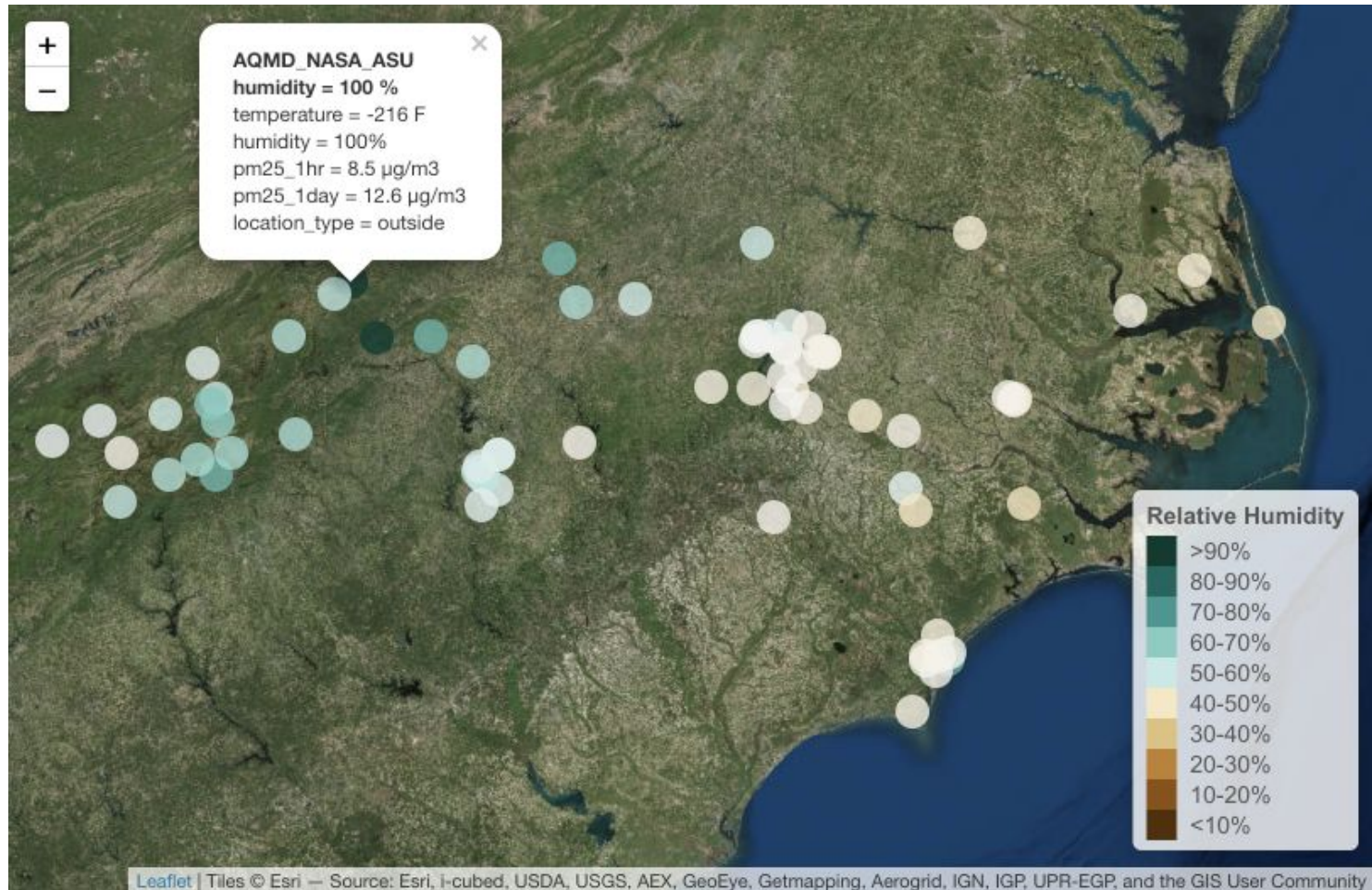
```

pas_load() %>%
  pas_filter(stateCode == 'NC') %>%
  pas_leaflet()

```

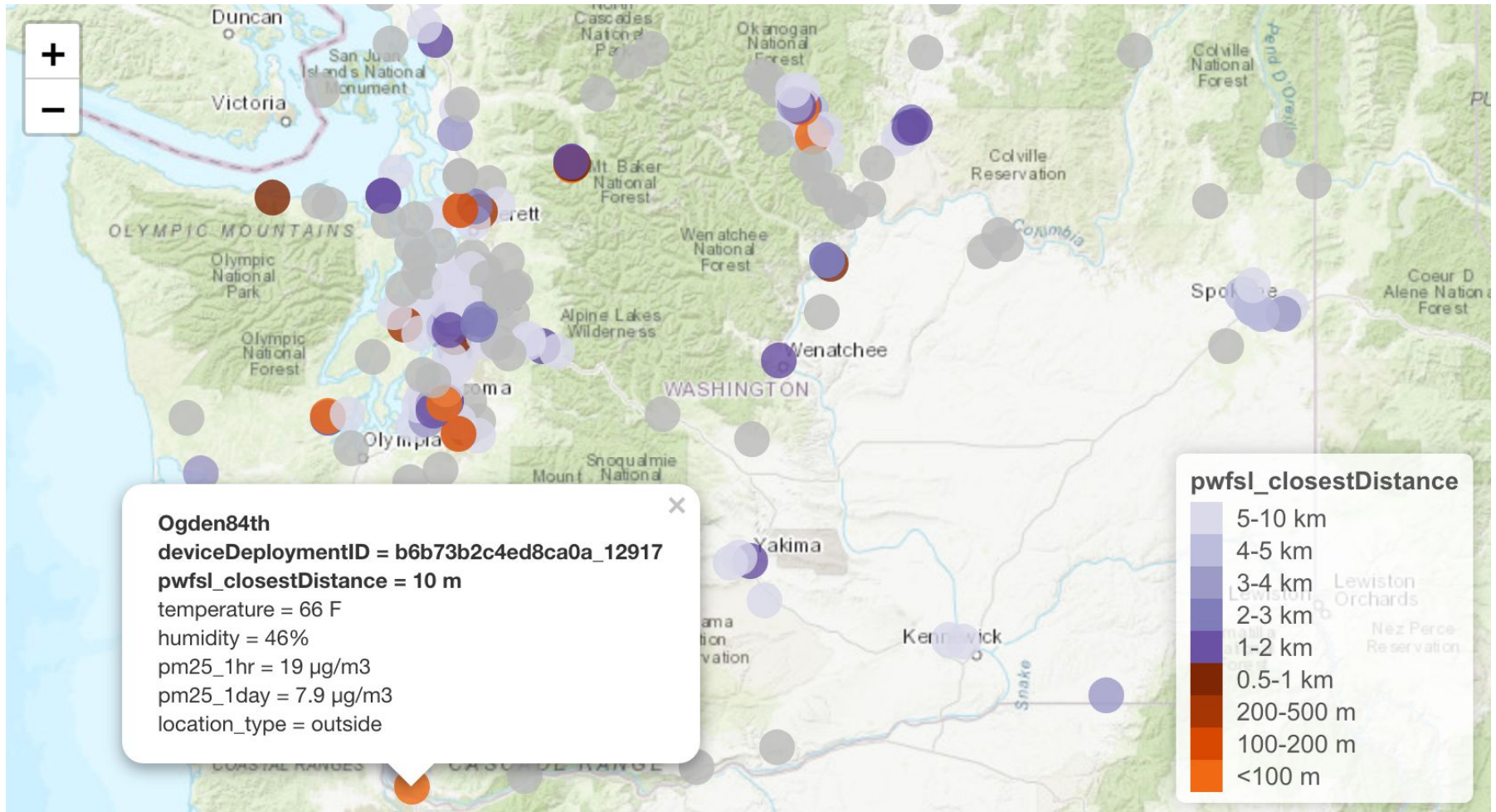


```
pas_load() %>%  
  pas_filter(stateCode == 'NC') %>%  
  pas_leaflet(param = 'humidity', maptype = 'satellite')
```





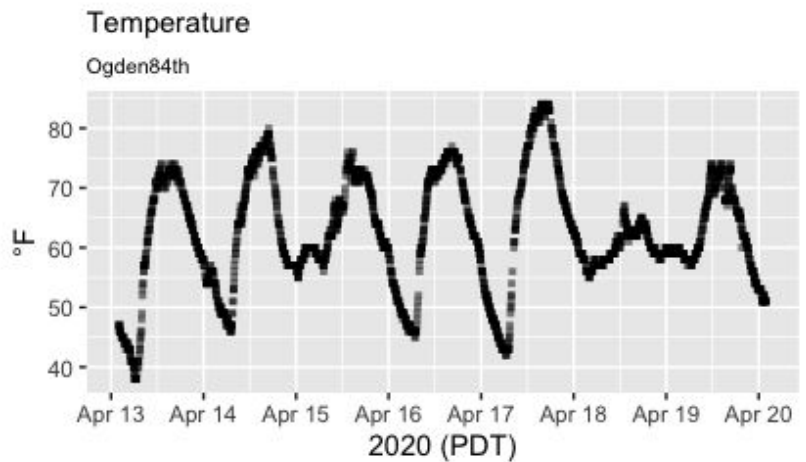
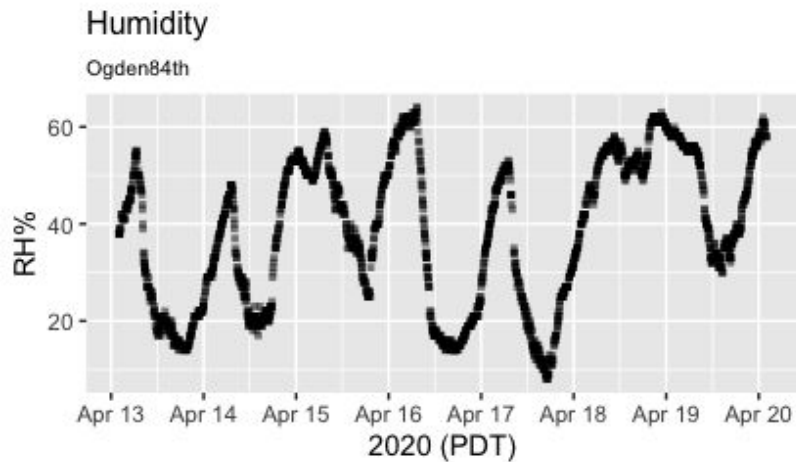
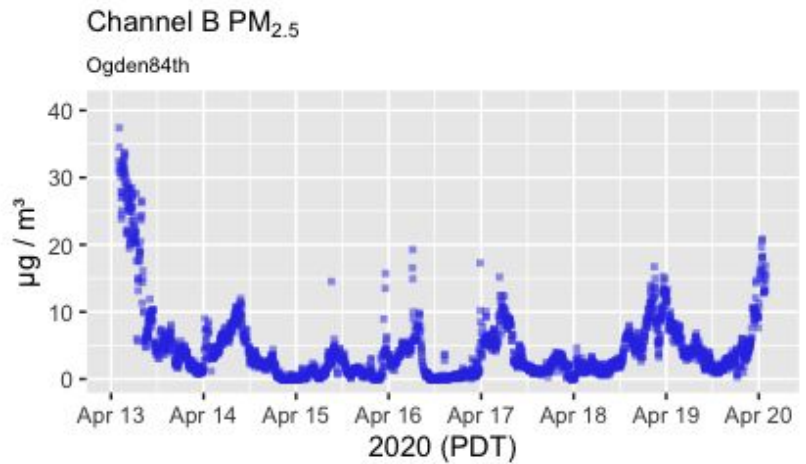
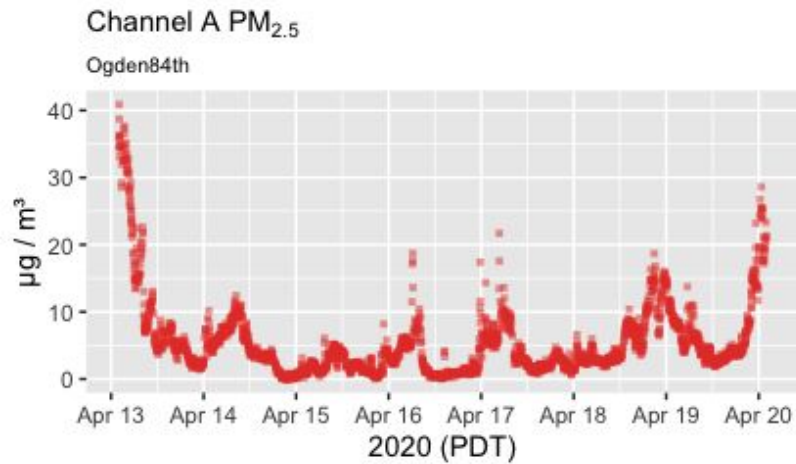
```
pas_load() %>%  
  pas_filter(stateCode == 'WA') %>%  
  pas_leaflet(param = 'pwfsl_closestDistance')
```



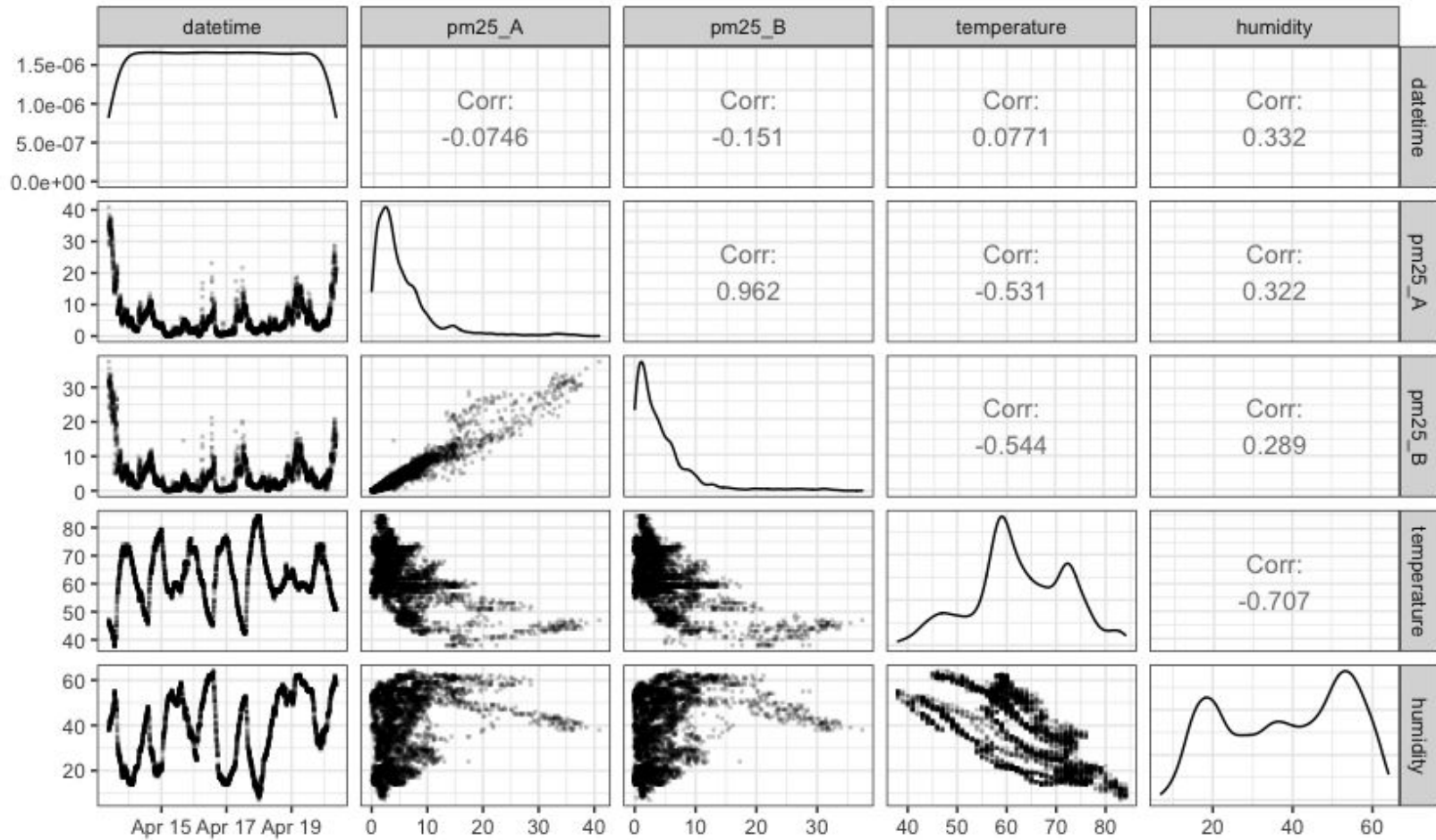
# Working with PurpleAir Timeseries (PAT) data

- Data from ThingSpeak API
- Lots of data analysis
- Lots of data visualization
- Makes common tasks easy
- Makes difficult tasks straightforward

```
pat <- pat_load("b6b73b2c4ed8ca0a_12917")  
pat %>% pat_multiplot()
```

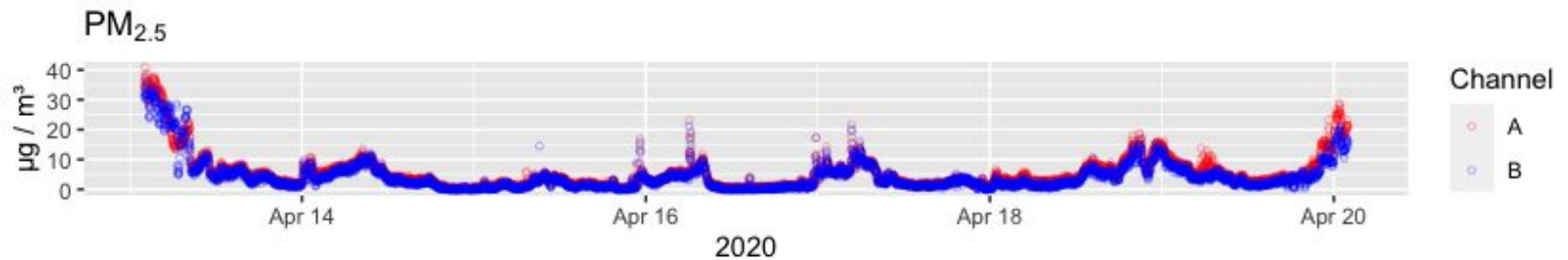
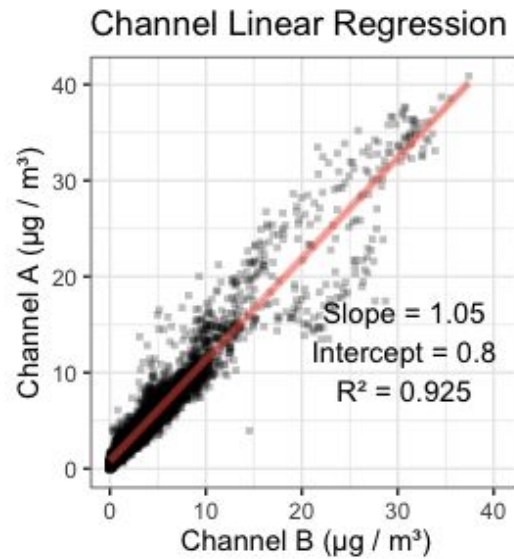


```
pat %>% pat_scatterMatrixPlot()
```



```
pat %>% pat_internalFit()
```

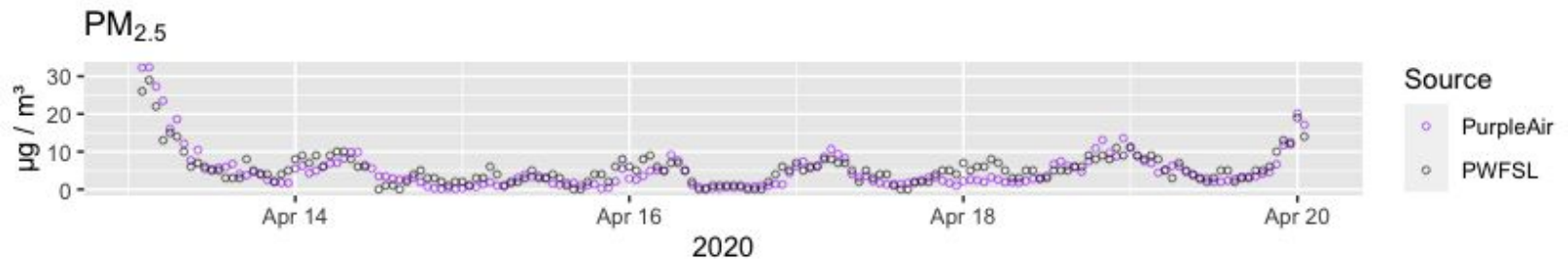
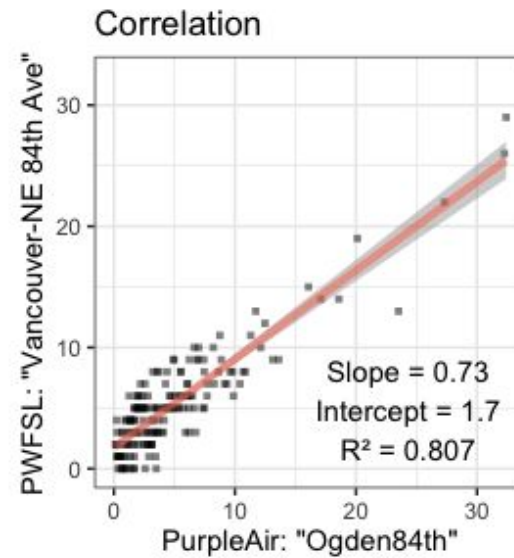
## A / B Channel Comparison -- Ogden84th





```
pat %>% pat_externalFit()
```

## Sensor / Monitor Comparison -- Distance: 0km

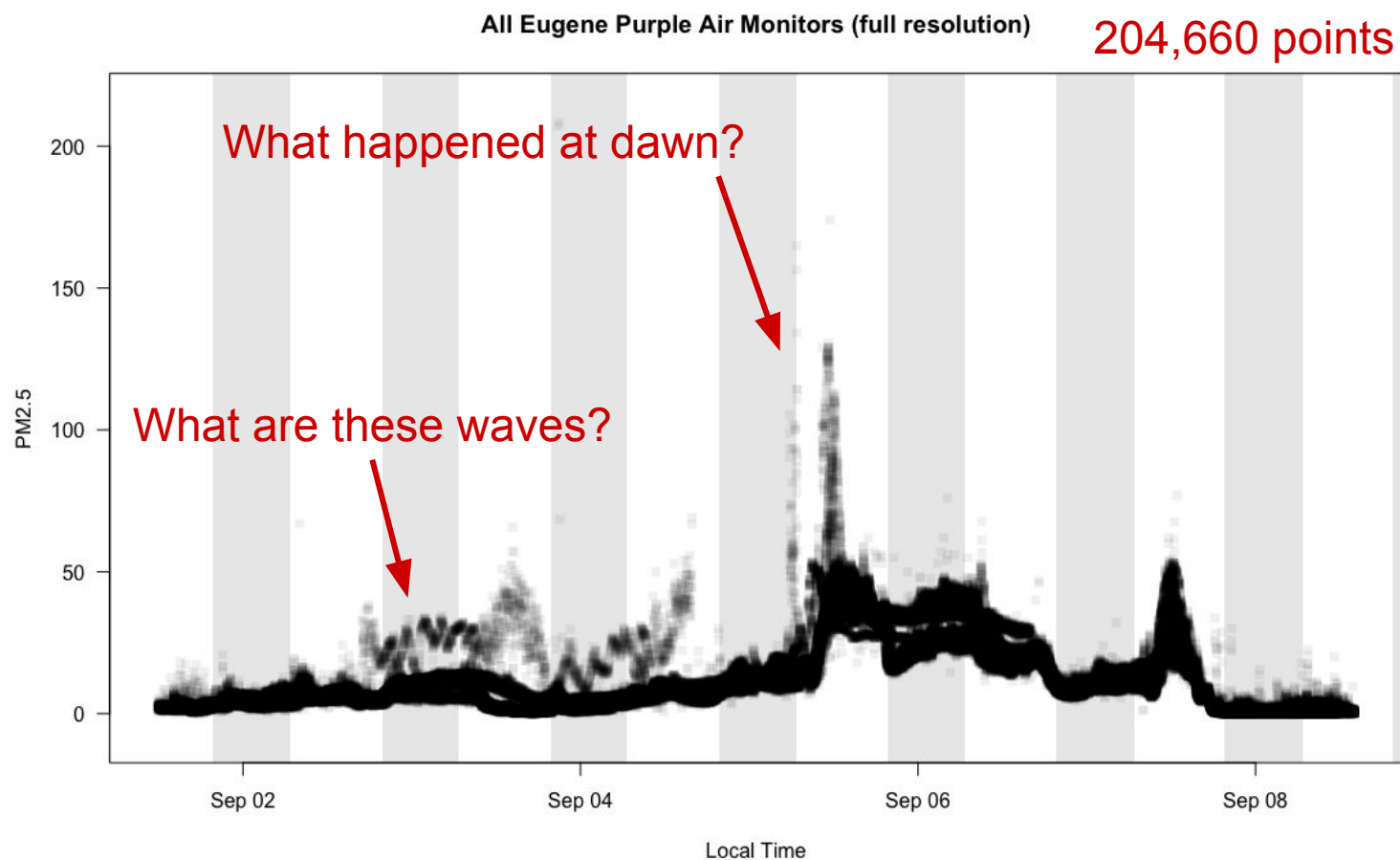




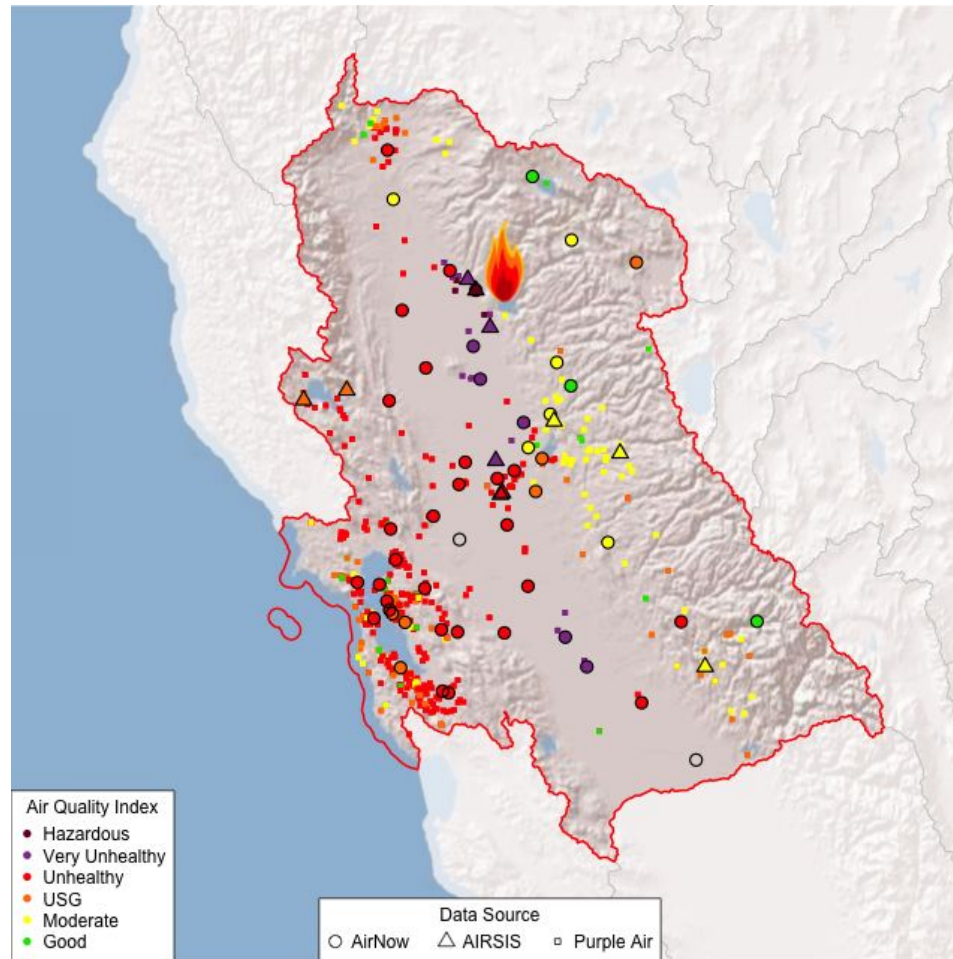
# Advanced Examples

- Compatibility with PWFSLSmoke R package
- Mapping
- Meta-analysis

# One week in Eugene.

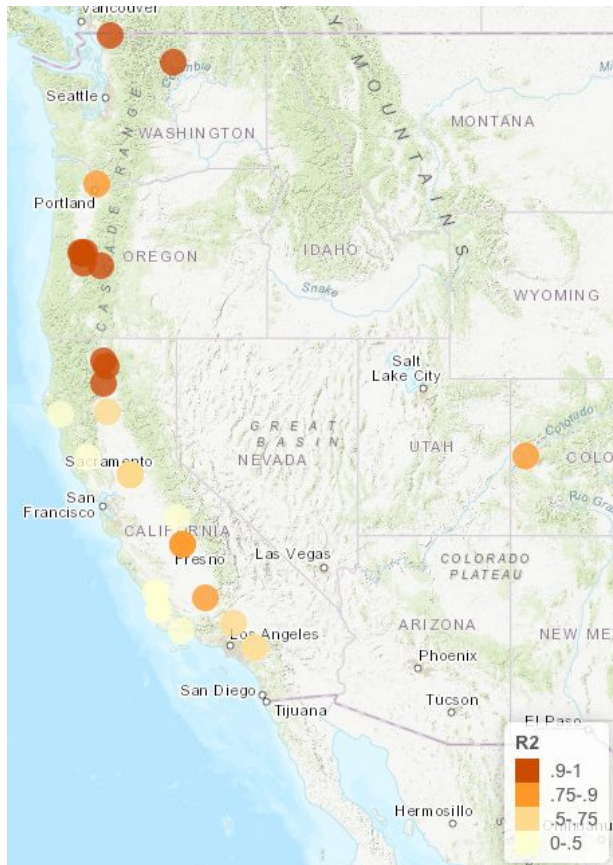


# Camp Fire monitors and sensors

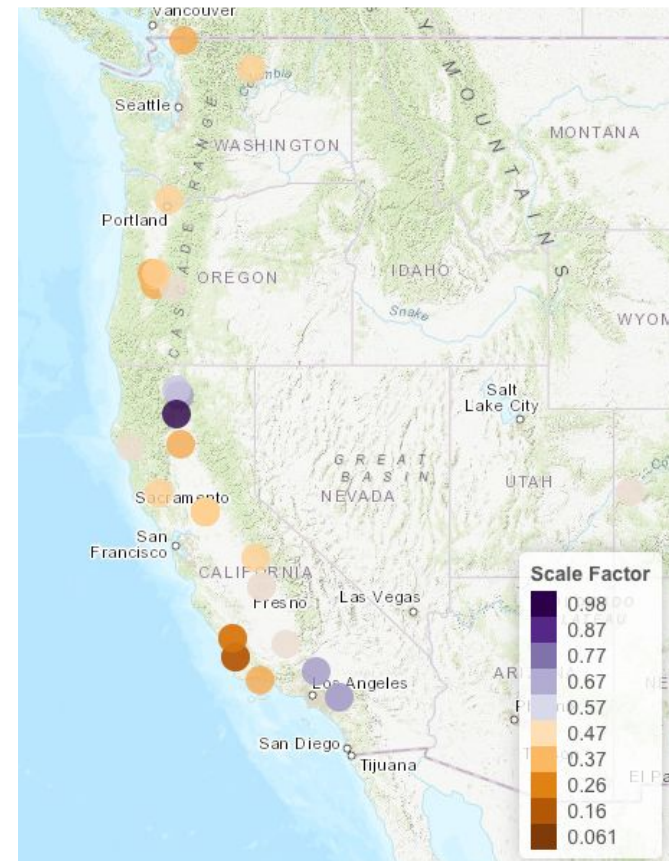


# Meta-analysis: sensor-monitor comparison

How good is the fit?



What is the scale factor?



# State-of-Health Index

Multi-metric index of individual SoH metrics, calculated daily

Reviewed lots of data, tried lots of combinations

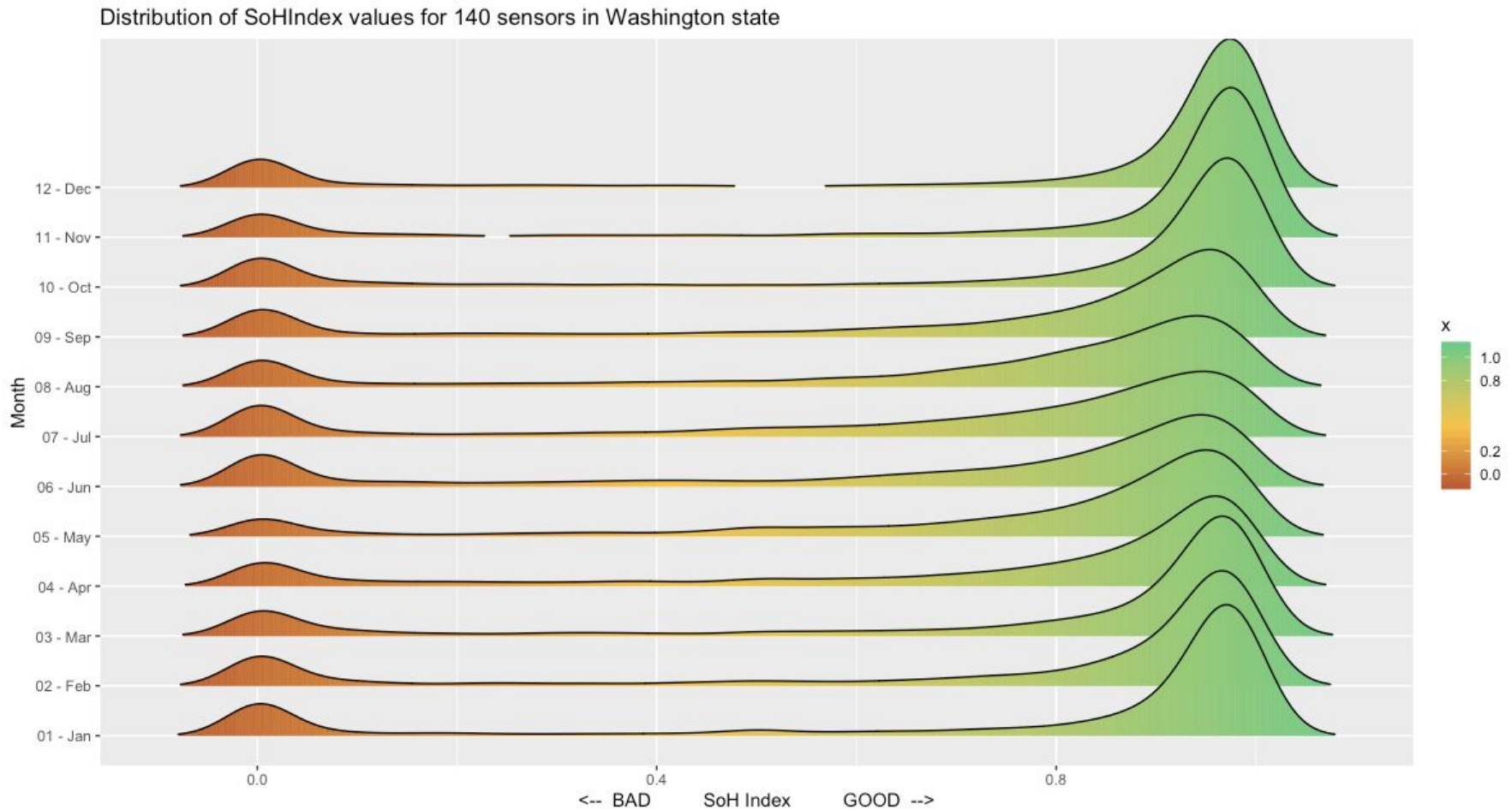
Current best version is:

**PurpleAirSoH\_dailyToIndex\_00()**

1. If the A or B channel percent reporting is  $< 50\%$ , index = 0
2. Otherwise, index =  $\text{pm25\_A\_pm25\_B\_rsquared}$
3. Poor/Fair/Good breaks =  $\text{c}(0, 0.2, 0.8, 1)$

Ends up tossing out 10-20% of sensor-days

# Distribution of SoHIndex by month for Washington



Thanks for listening!

<http://mazamascience.com/presentations>

<http://smoke.mazamascience.com>

For information about the EPA STAR Grant at South Coast AQMD contact Dr. Vasileios Papapostolou:

[vpapapostolou@aqmd.gov](mailto:vpapapostolou@aqmd.gov) or (909) 396-2254



# GitHub Branches!

 [MazamaScience](#) / [AirSensor](#)

 Unwatch ▾ 4

 Star 3

 Fork 1

 Code

 Issues 12

 Pull requests 0

 Actions

 Projects 7

 Wiki

 Security 1

 Insights

 Settings

Overview

Yours


Active

Stale

All branches

Search branches...

## Default branch

 **master** Updated 2 days ago by jonathancallahan



Default

Change default branch

## Your branches

**version-0.5** Updated 18 days ago by jonathancallahan

143 | 5

 New pull request



**hans** Updated 2 days ago by hansmrtn

6 | 0

 New pull request



## Active branches

**hans** Updated 2 days ago by hansmrtn

6 | 0

 New pull request



**version-0.5** Updated 18 days ago by jonathancallahan

143 | 5

 New pull request

