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.

MAZAMA SCIENCE
SEE YOUR DATA

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.

MAZAMA SCIENCE
SEE YOUR DATA

Why support open, shared tools?

The "Data Deluge" is real.

Looking at PM2.5 measurements only:

AirNow 2018:

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

PurpleAir 2018:

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

New tools are needed to work efficiently with large datasets.

MAZAMA SCIENCE
SEE YOUR DATA

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

MAZAMA SCIENCE
SEE YOUR DATA

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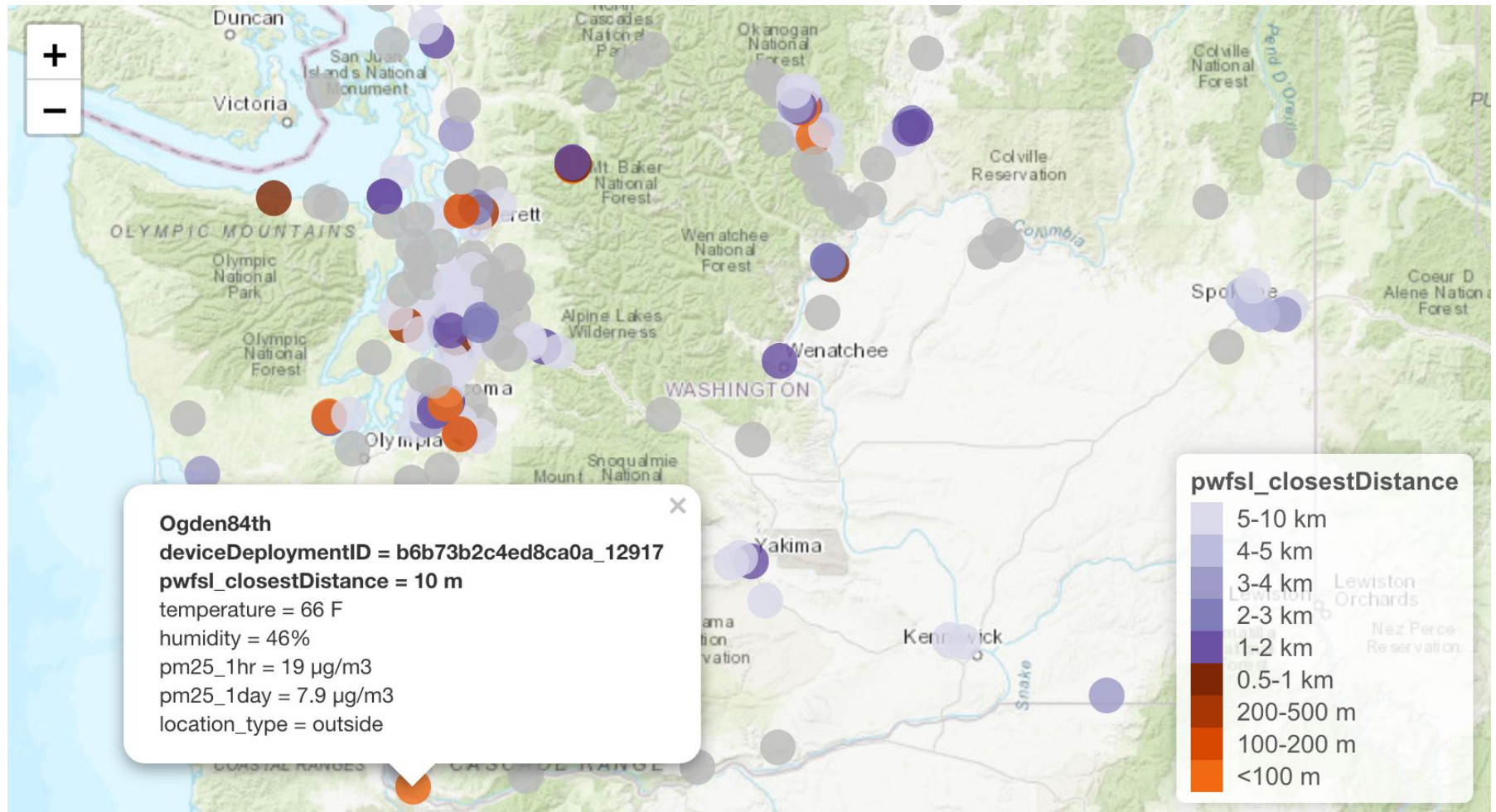

MAZAMA SCIENCE
SEE YOUR DATA
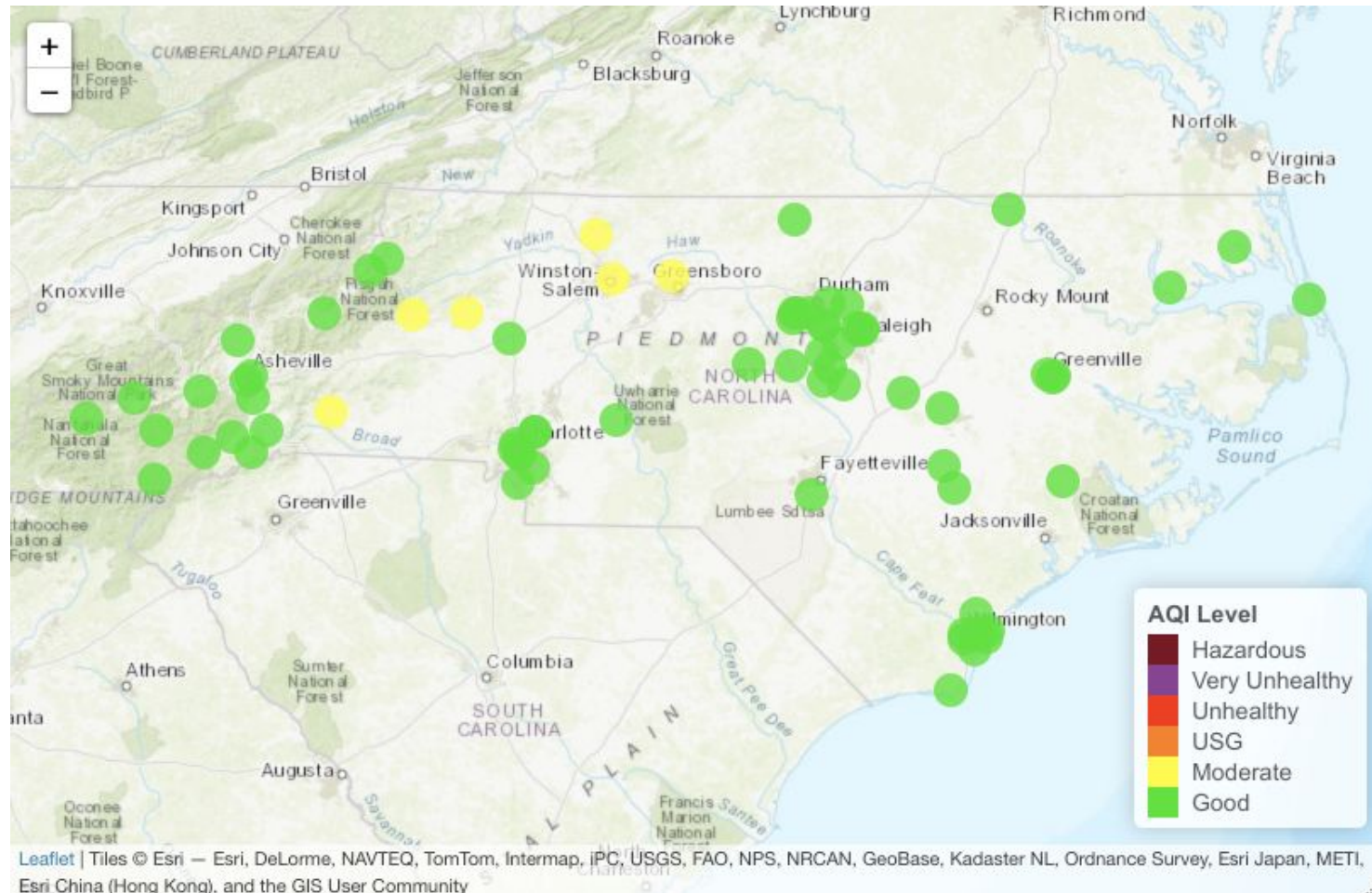
```
pas_load() %>%
  pas_filter(stateCode == 'NC') %>%
  pas_leaflet(param = 'humidity', maptype = 'satellite')
```
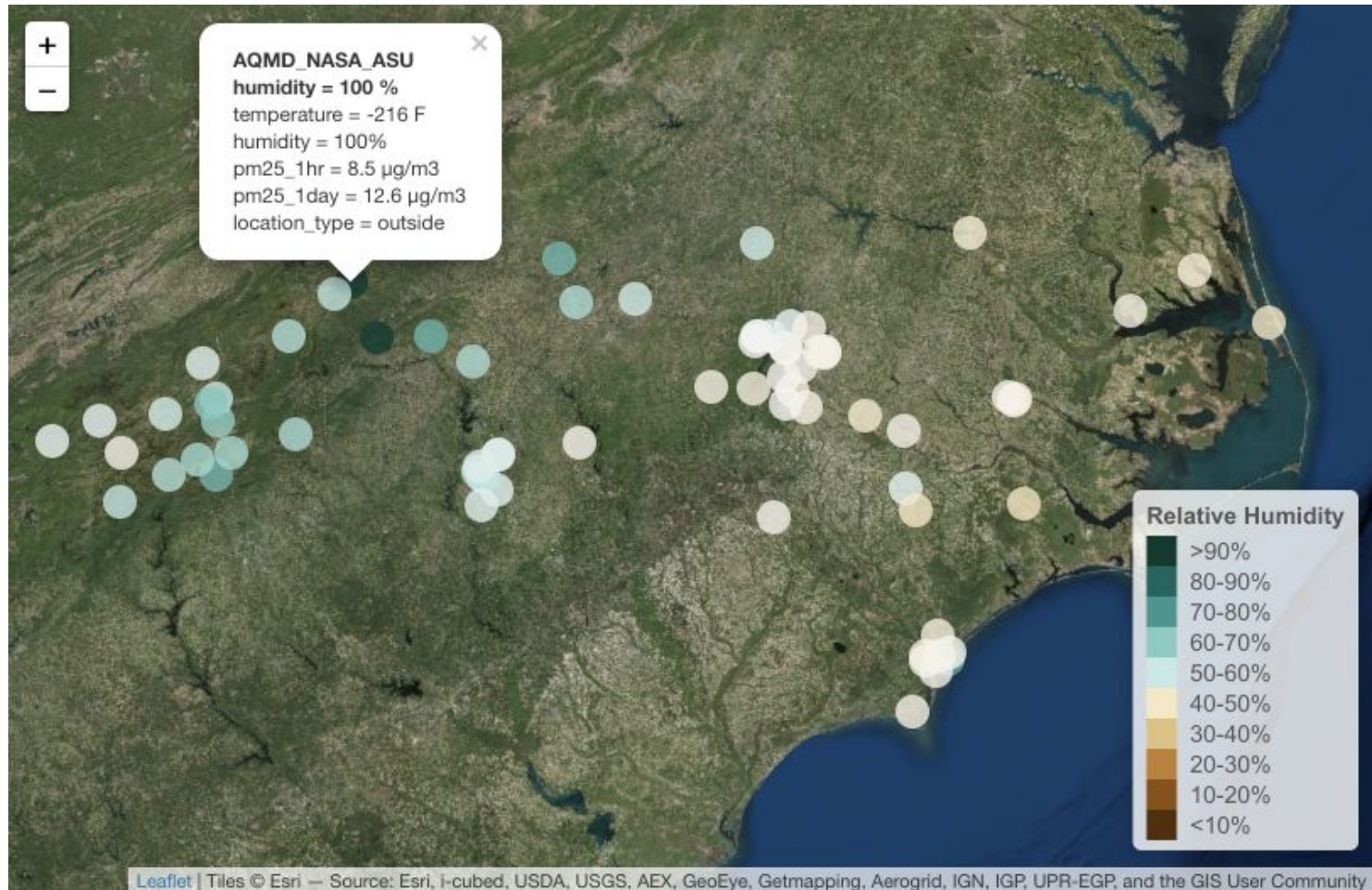
```
pas_load() %>%
  pas_filter(stateCode == 'WA') %>%
  pas_leaflet(param = 'pwfsl_closestDistance')
```


Ogden84th
deviceDeploymentID = b6b73b2c4ed8ca0a_12917
pwfsl_closestDistance = 10 m
temperature = 66 F
humidity = 46%
pm25_1hr = 19 µg/m3
pm25_1day = 7.9 µg/m3
location_type = outside

pwfsl_closestDistance
5-10 km
4-5 km
3-4 km
2-3 km
1-2 km
0.5-1 km
200-500 m
100-200 m
<100 m

MAZAMA SCIENCE
SEE YOUR DATA

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

MAZAMA SCIENCE
SEE YOUR DATA

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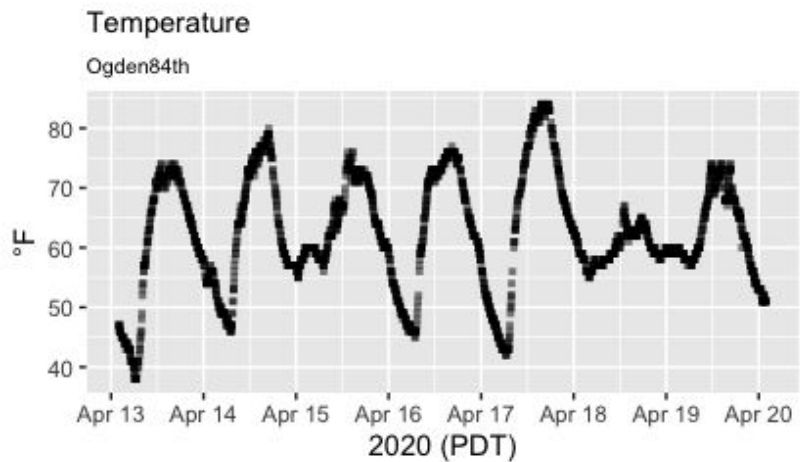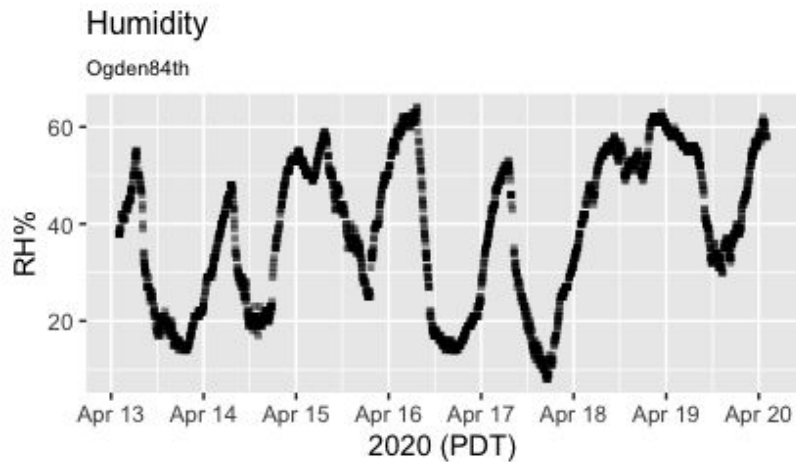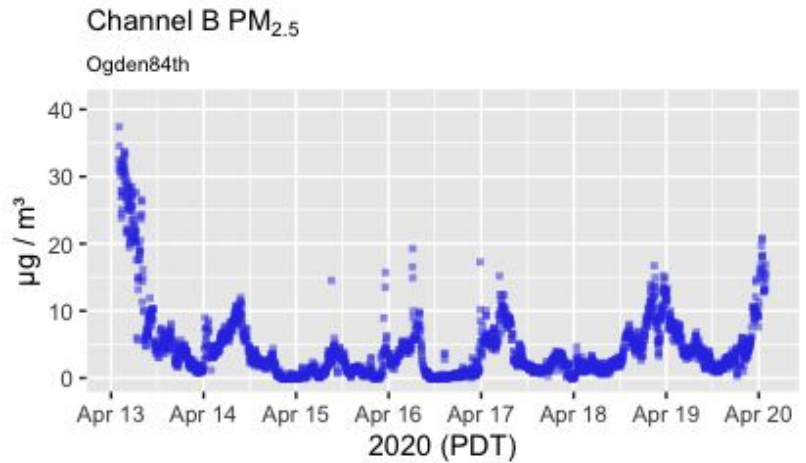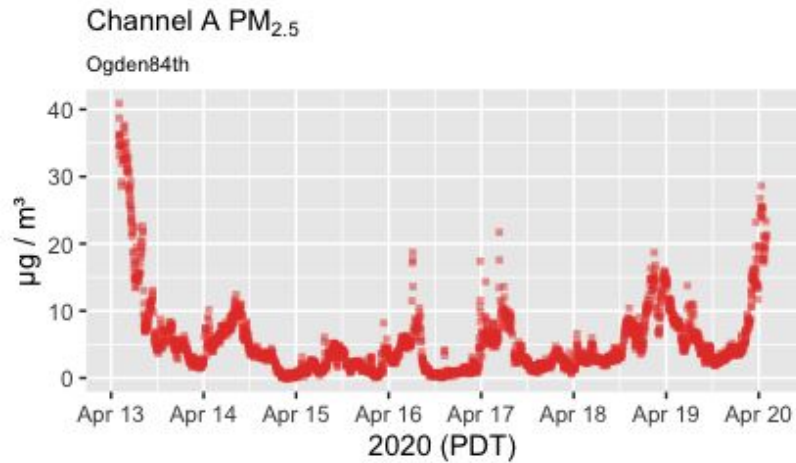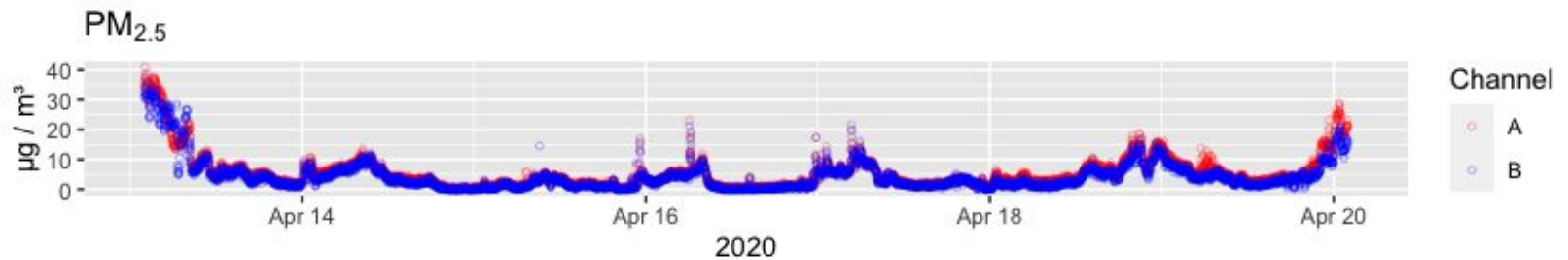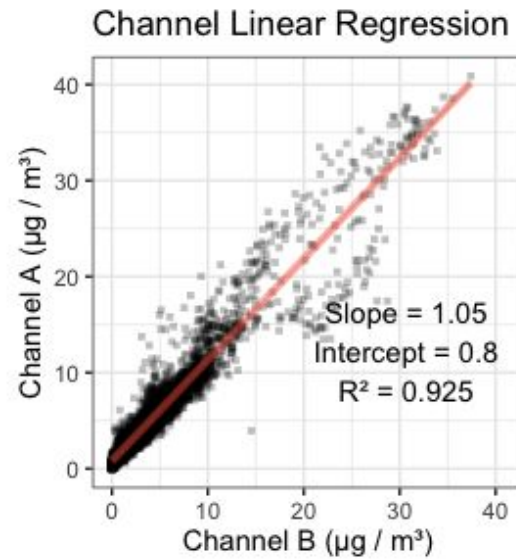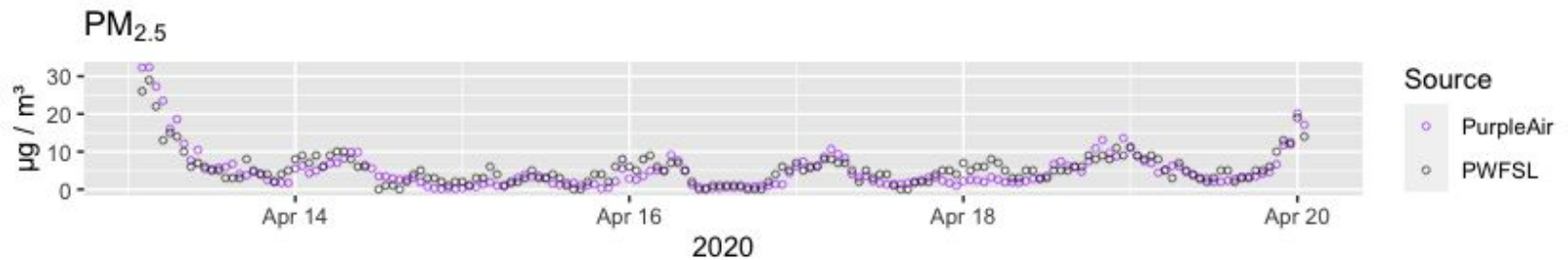
```
pat %>% pat_scatterMatrixPlot()
```
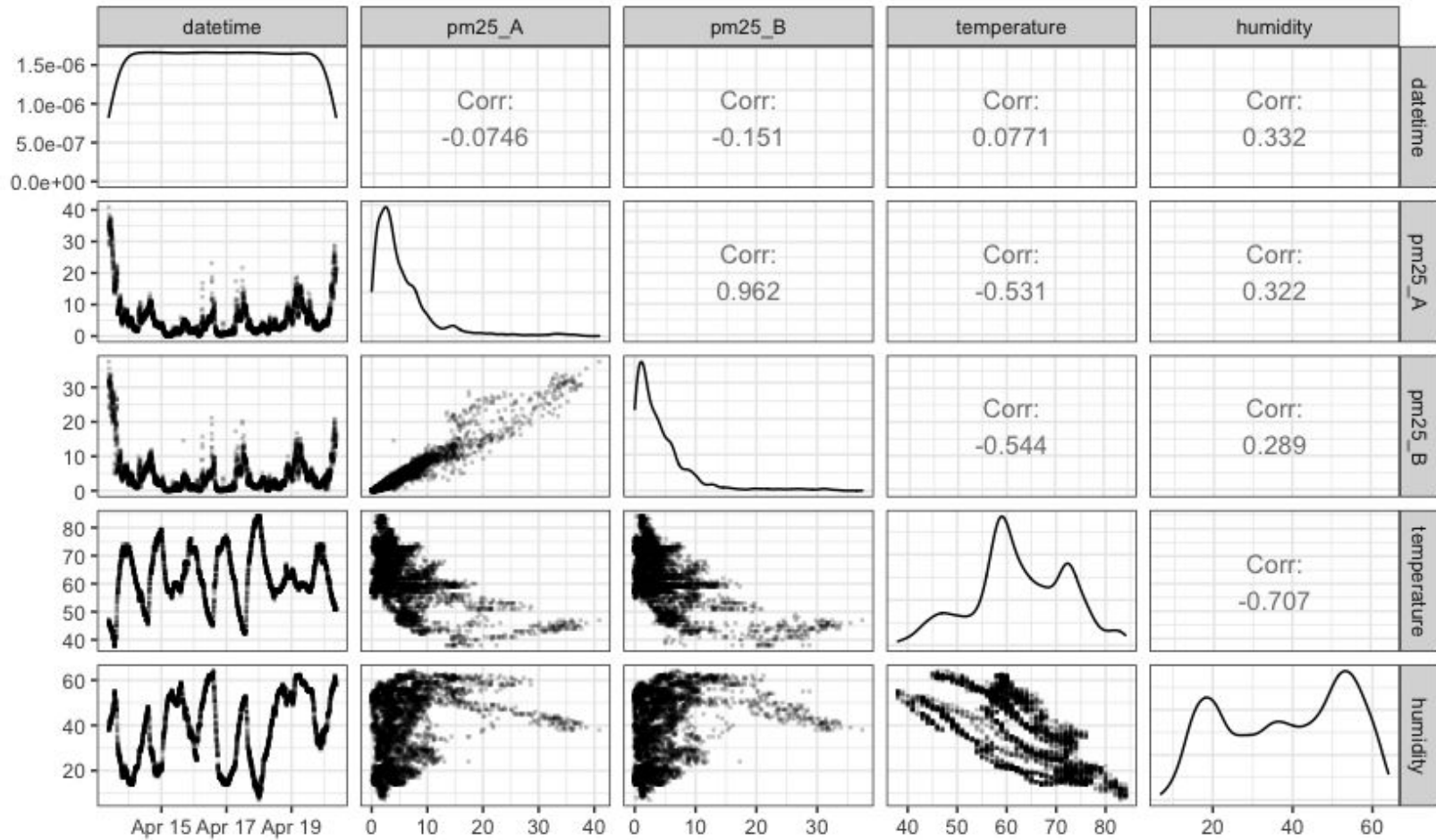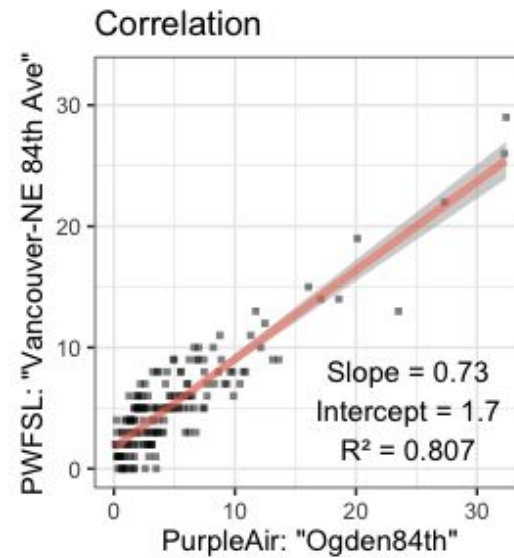
```
pat %>% pat_internalFit()
```


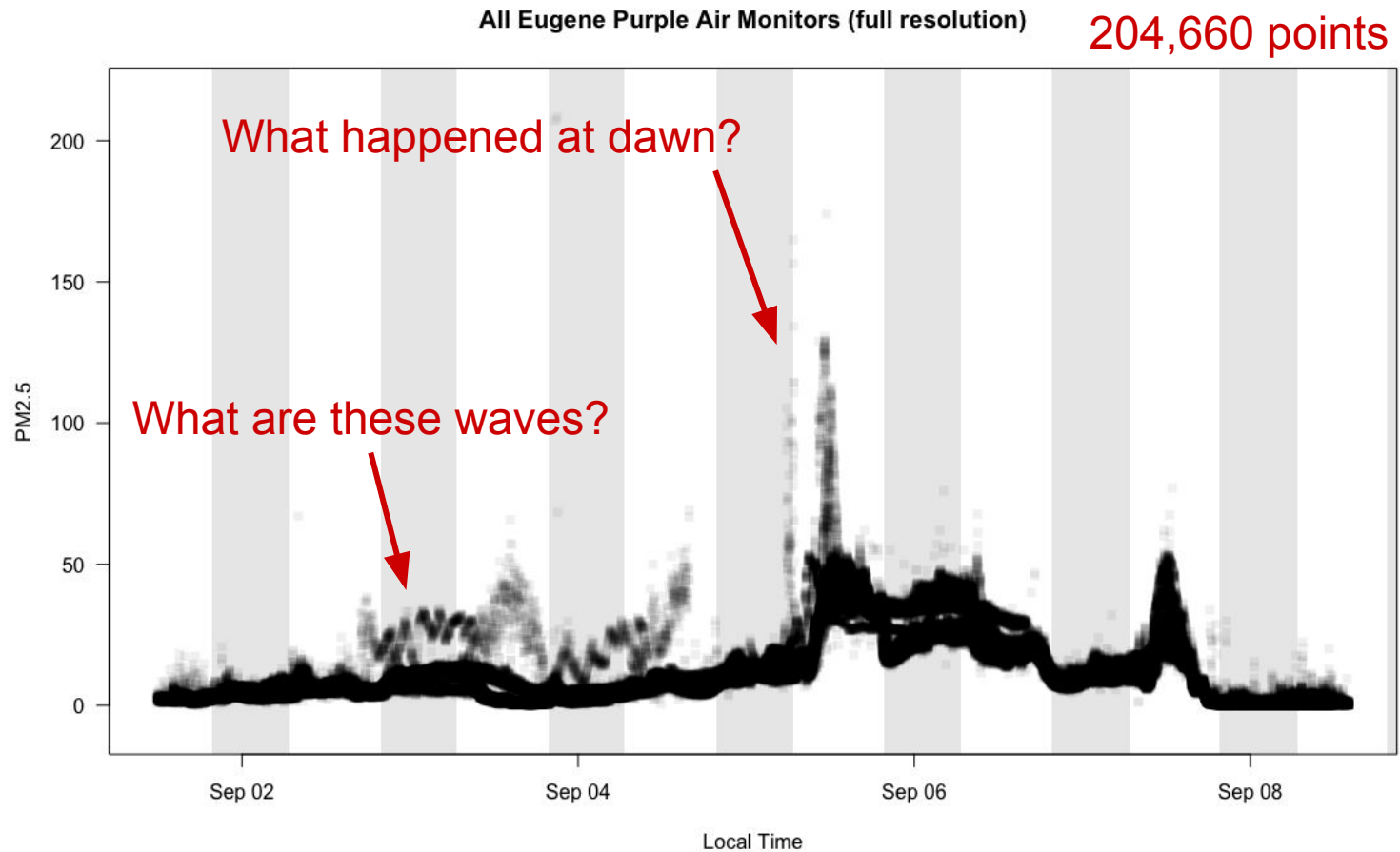A / B Channel Comparison -- Ogden84th

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

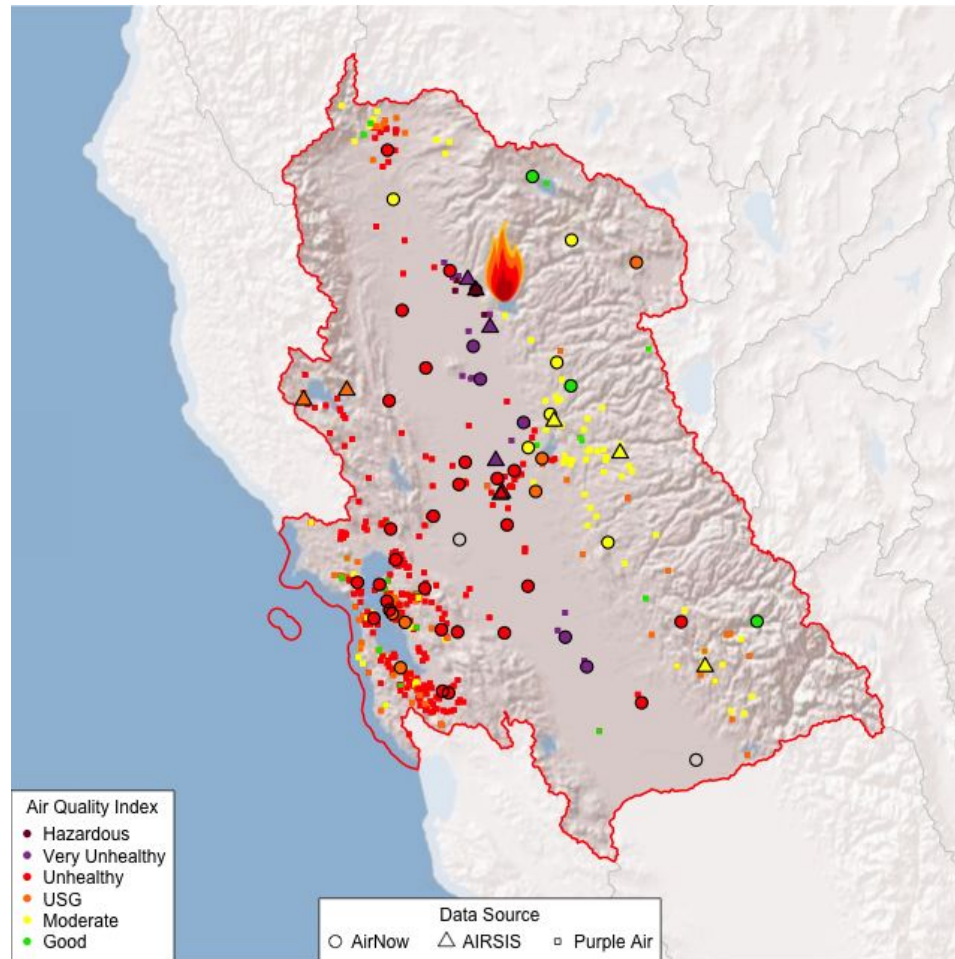

Sensor / Monitor Comparison -- Distance: 0km

Advanced Examples

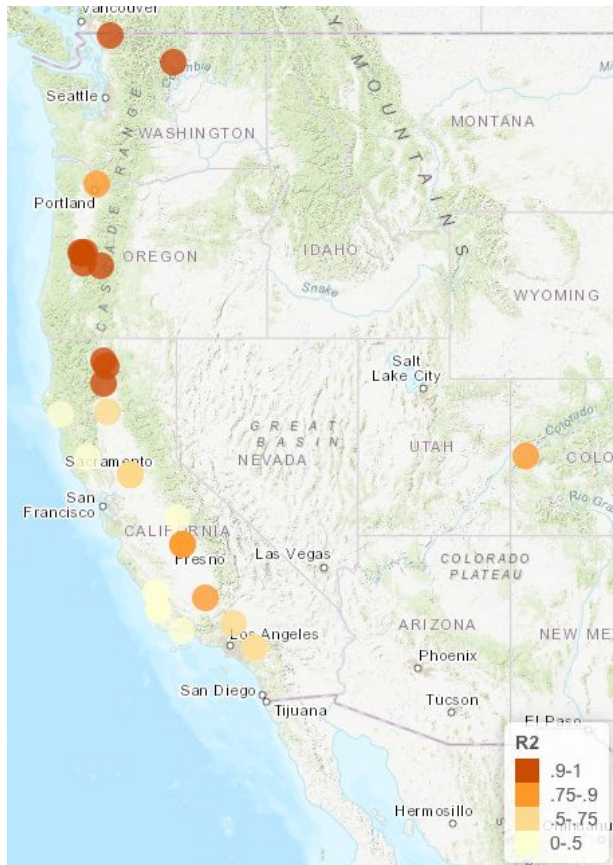- Compatibility with PWFSLSmoke R package

- Mapping

- Meta-analysis

MAZAMA SCIENCE
SEE YOUR DATA

One week in Eugene.

All Eugene Purple Air Monitors (full resolution)

204,660 points

What happened at dawn?

What are these waves?

MAZAMA SCIENCE
SEE YOUR DATA

Camp Fire monitors and sensors

Air Quality Index
- Hazardous
- Very Unhealthy
- Unhealthy
- USG
- Moderate
- Good

Data Source
○ AirNow △ AIRSIS □ Purple Air

MAZAMA SCIENCE
SEE YOUR DATA

Meta-analysis: sensor-monitor comparison

How good is the fit?

What is the scale factor?

MAZAMA SCIENCE
SEE YOUR DATA
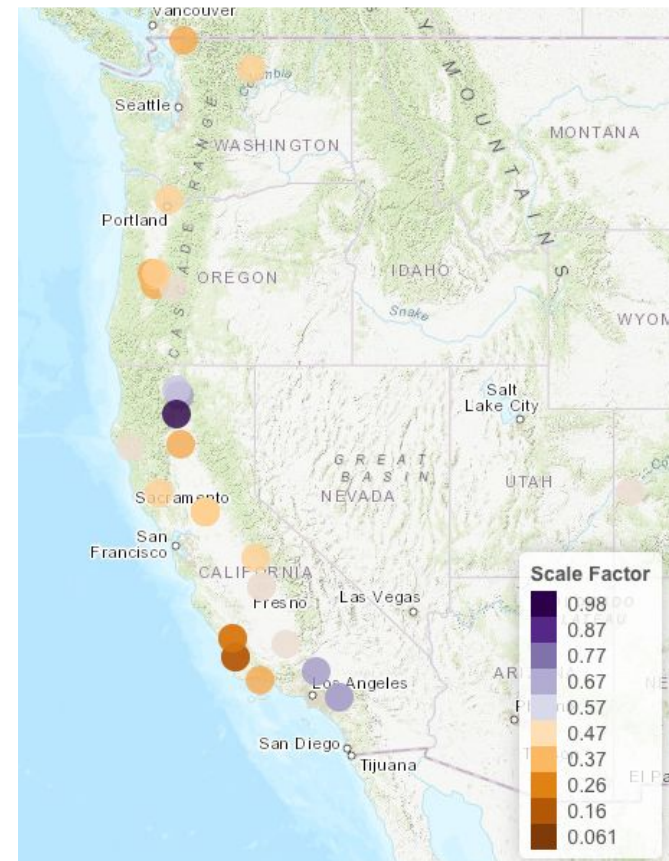
State-of-Health Index

Multi-metric index of individual SoH metrics, calculated daily
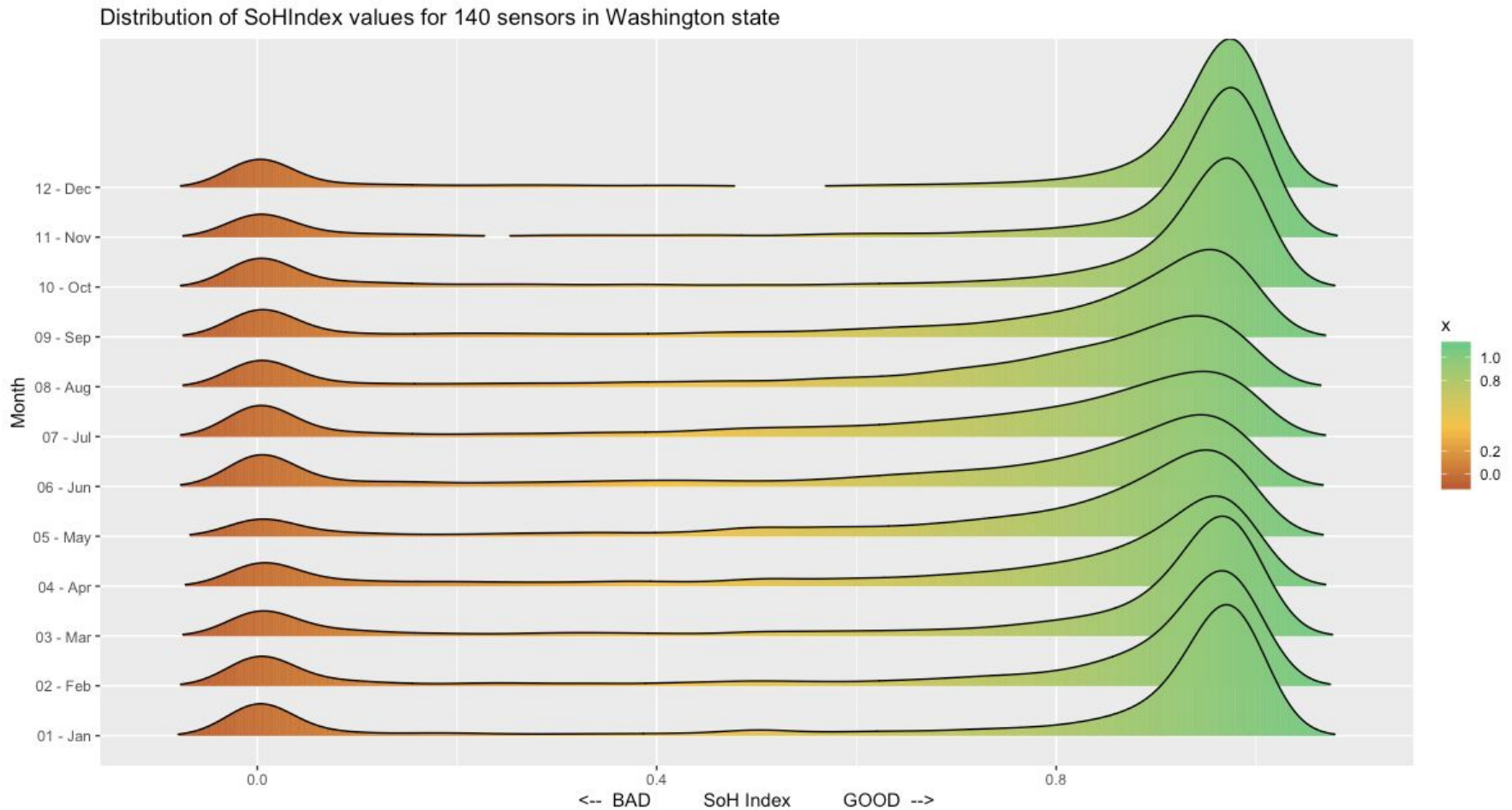
Reviewed lots of data, tried lots of combinations

Current best version is:

`PurpleAirSoH_dailyToIndex_00()`

1. If the A or B channel percent reporting is < 50%, index = 0

2. Otherwise, index = pm25_A_pm25_B_rsquared

3. Poor/Fair/Good breaks = c(0, 0.2, 0.8, 1)

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

Distribution of SoHIndex by month for Washington

Distribution of SoHIndex values for 140 sensors in Washington state

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 or (909) 396-2254

MAZAMA SCIENCE
SEE YOUR DATA

GitHub Branches!