# Southeast Air Quality Analysis

Southeastern Data Users Group

- Purpose of this presentation is to inform you about the formation of a data users group for the Region 4 states and to seek individuals who might be interested in joining.
    - Discuss the "openair" package in R and show some of the analysis tools available.
    - South Carolina and EPA have developed code using the R statistical programming package to help analyze ambient air data.
- Vision for the group:
    - to promote and grow the data analysis capabilities of Southeastern US air quality data through sharing and collaboration between air quality agency staff (and other stakeholders when appropriate).
    - Who would benefit –anyone who does/has done air quality data analysis and is interested in learning more; planning staff or staff responsible for developing nonattainment boundaries, or writing State Implementation Plans; meteorological staff that might be interested in combining weather and ambient data, etc.

Data Analysis Process and Available Tools

SOUTHEAST AIR QUALITY ANALYSIS

# General data analysis process
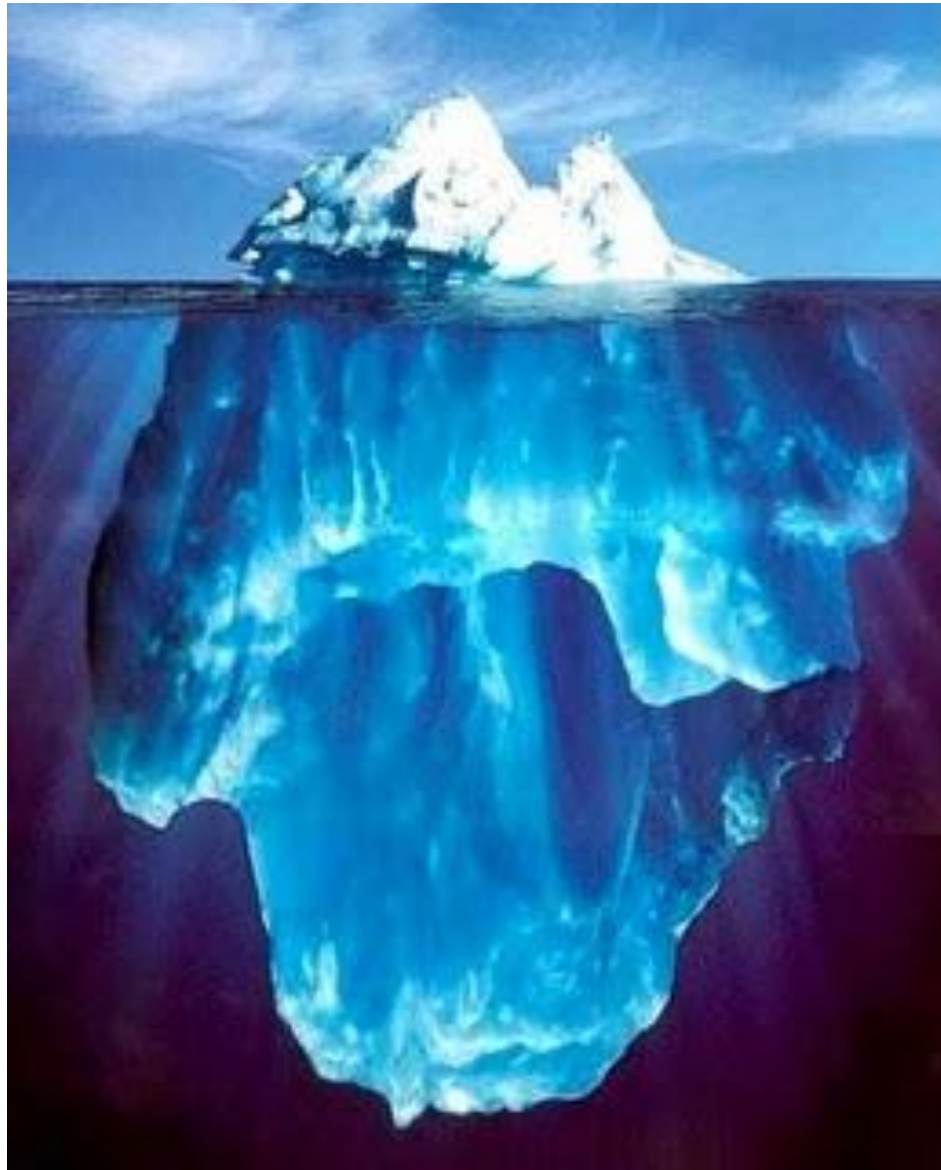
final write up and/or graphics

- - - - - - -

cleaning data, exploratory analysis, specific analysis techniques, error checking, tweaking graphics to look better

Working with data in spreadsheets: issues to be aware of

- Data integrity - Easy to accidentally modify your data; difficult to identify these changes

- Lack of flexibility - To reshape or reanalyze your data can be time consuming or very time consuming

- Lack of transparency – can be hard to determine assumptions and data dependencies

Working with data using programming language
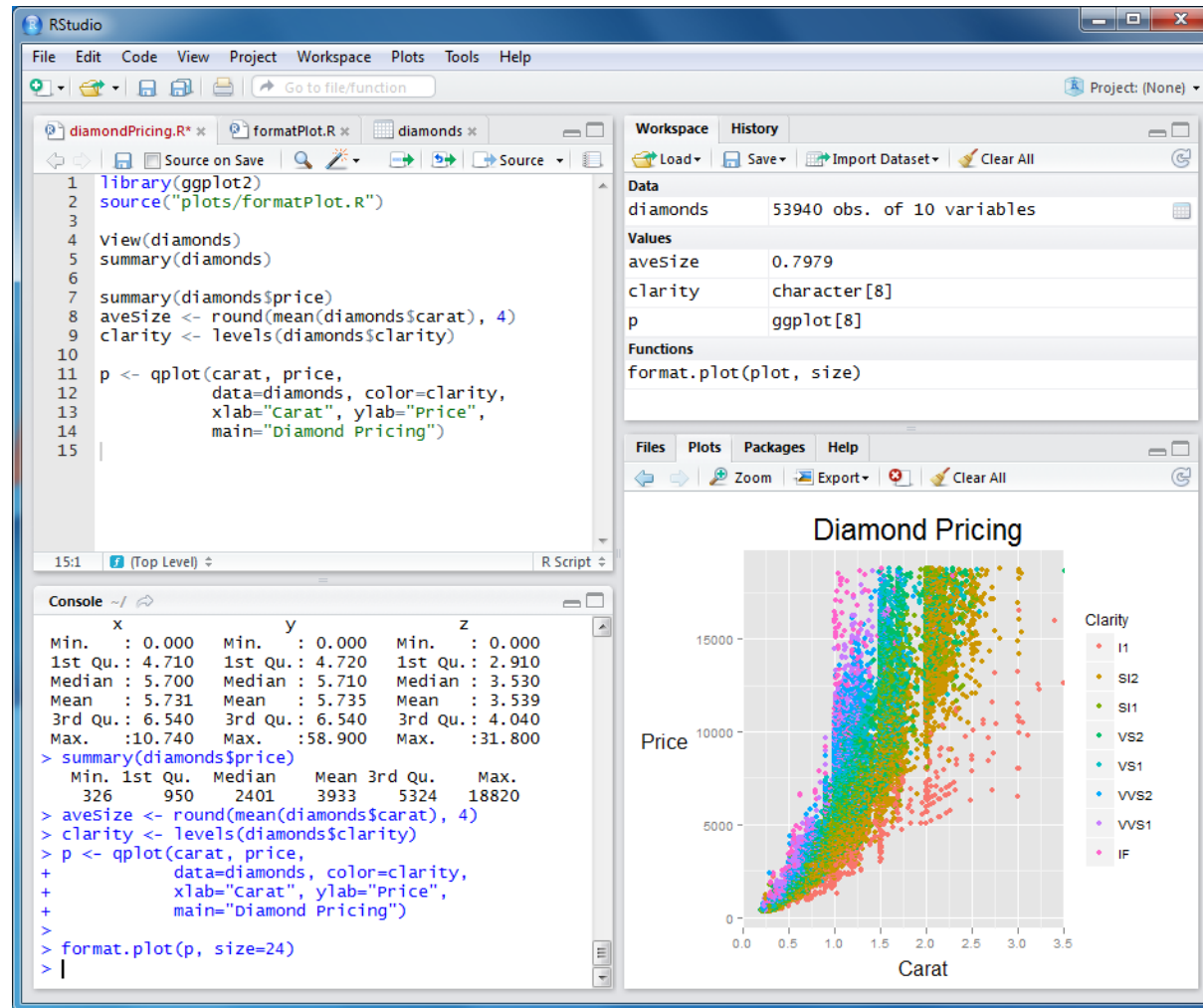
Working with data using programming language

- Typically best for working with 10,000 – millions+ of data records and doing more complex analysis on large data
- Flexible
  - the most options for different analyses
  - can be reused on other analyses
  - can be built in pieces that are useful for other
- Transparent
  - Can comment and show assumptions and methods in the code
  - Can easily share an analysis
- Repeatability
  - Can do the same thing on different data

R – open source statistical package

- Large user base -> well developed libraries for almost any statistical technique.

- Generate presentation quality graphics.

- Better than Python for quick interactive data analysis.

- Python is better for developing reusable code that will be improved and reused for automating processes.

openair package

- Package was developed by Dr. David Carslaw with King's College London. (He is very responsive to questions and bug reports).

- Constant updates to the code.

- User's manual is extensive (currently 287 pages long) and has a good tutorial on how to use R.

- The purpose of openair is to make available a consistent set of tools for analyzing and understanding air pollution data in a free, open-source environment.

Example plots created inside of the R statistical software package

SOUTHEAST AIR QUALITY ANALYSIS

Calendar Plots

450790007 O$_3$ in 2002
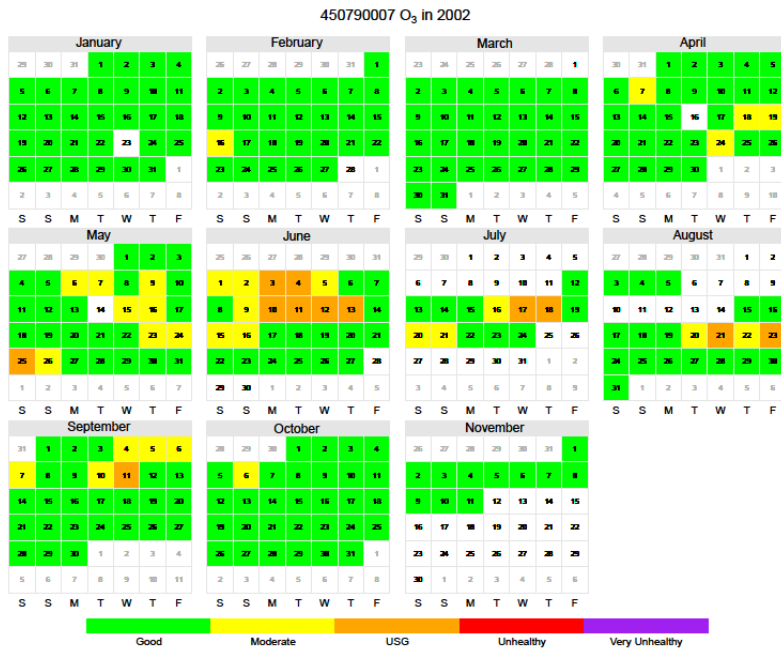
- Can give a quick overview of the AQI values for the criteria pollutants.

- Episodes can quickly be identified.

- Graph can be modified to provide individual monitor or statewide maximum concentration values.

Trend graphs

- Design values for each criteria pollutant.
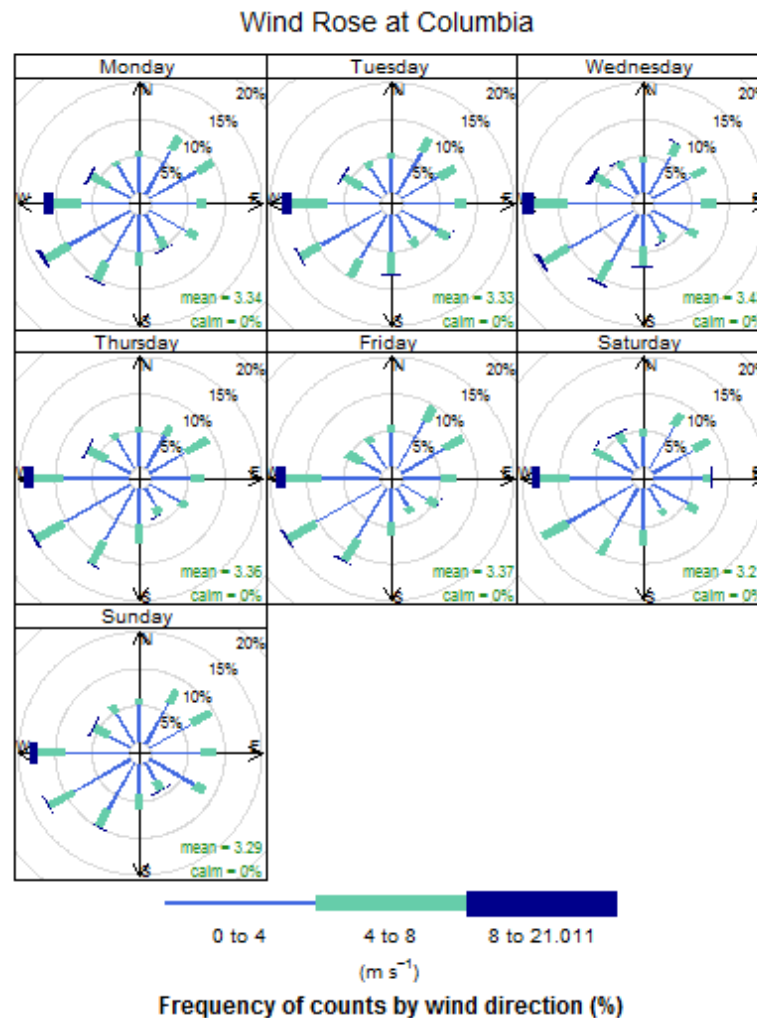- Graph can be customized to display one or more monitoring sites.

Wind Rose

Wind Rose at Columbia
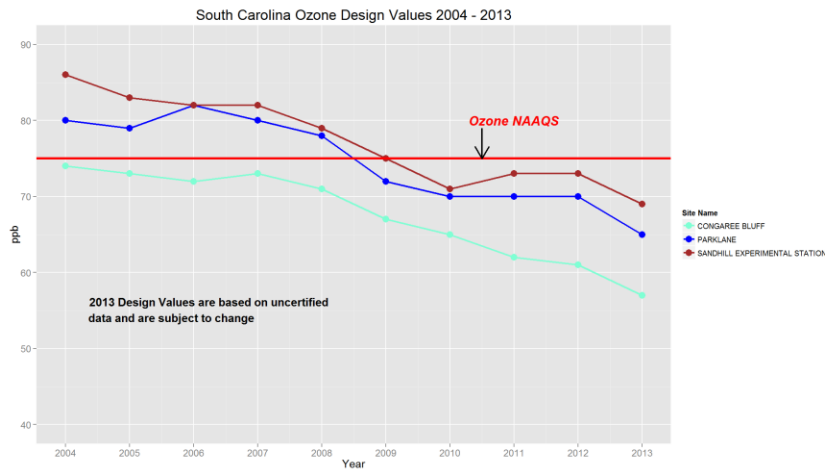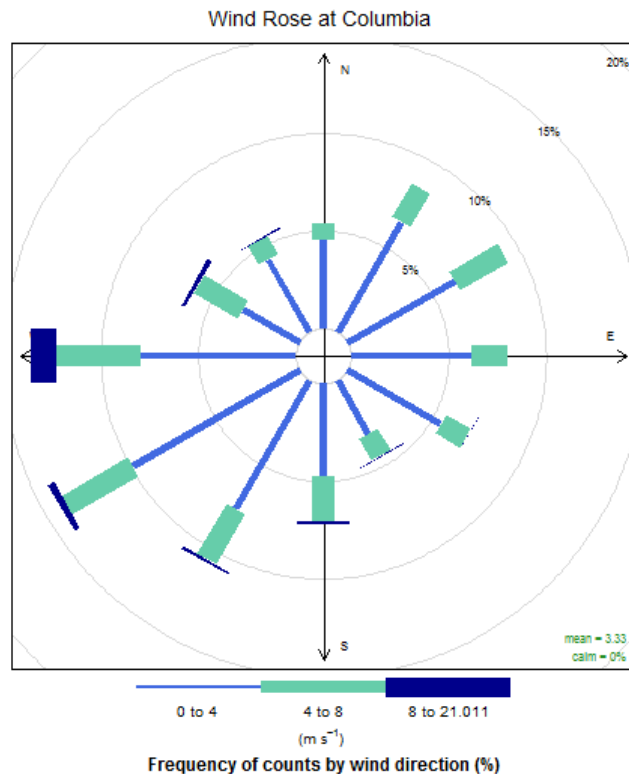
- Traditional windrose. Can substitute wind speed for pollutant concentration to create a pollution rose.

- Wind/Pollution roses can be conditioned to show winds by year, month, day of week, daytime/nighttime.

# Example of wind rose by day of week

Trends in pollutant concentration by wind direction

Ozone Trends (in units/year) By Wind Direction at Parklane

- This graphic shows the trends in pollutant concentration by wind direction.

- Can also calculate a percent or unit change in concentrations for that wind direction. Also has an indicator if the trend is statistically significant.

Concentration "polar" plots

Polar Plot of SO₂ at Irmo
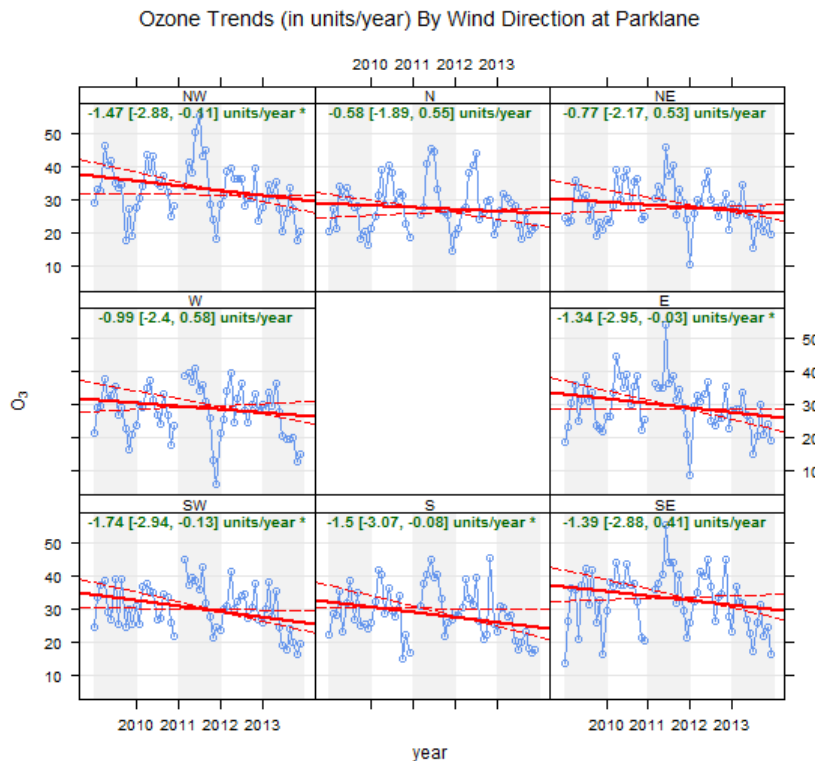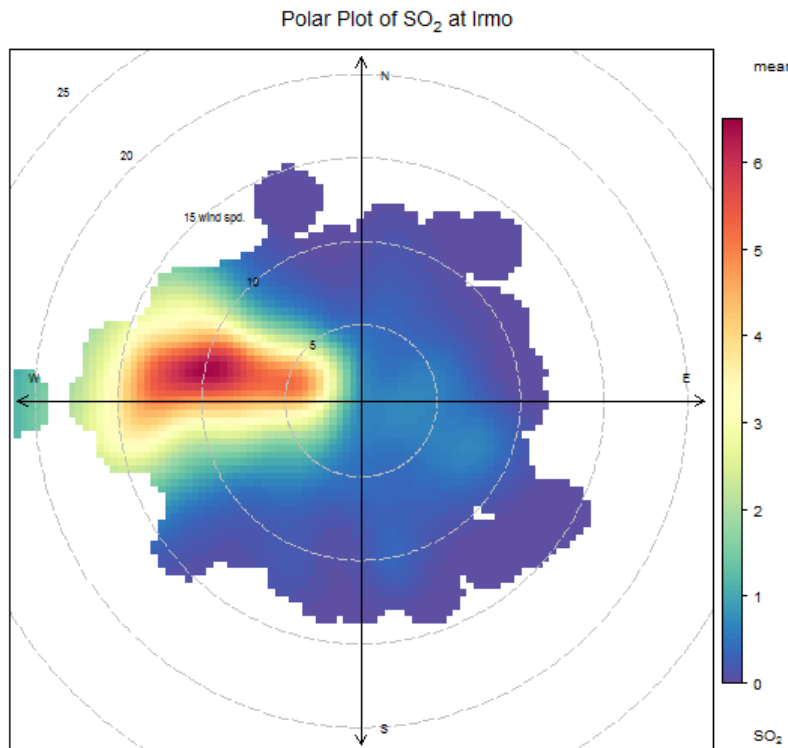
- This plot is a type of pollution rose that combines wind direction, wind speed and pollutant concentration.

- Polar plots can be conditioned to show concentrations by year, month, day of week, daytime/nighttime.

Percent of Standard

Percent of NAAQS: Ozone 2009 – 2013

- This graphic normalizes the design value (by year) based on the NAAQS.
  - (DV/NAAQS)*100
- The dashed vertical lines represent ±15% of the standard.
- This graphic is used in the network assessment to help us determine the value of a site (a site within the vertical dashed lines is more valuable than one outside the lines).

Trajectory Analysis

Backtrajectory frequency for Upstate area monitors
2011 - 2014
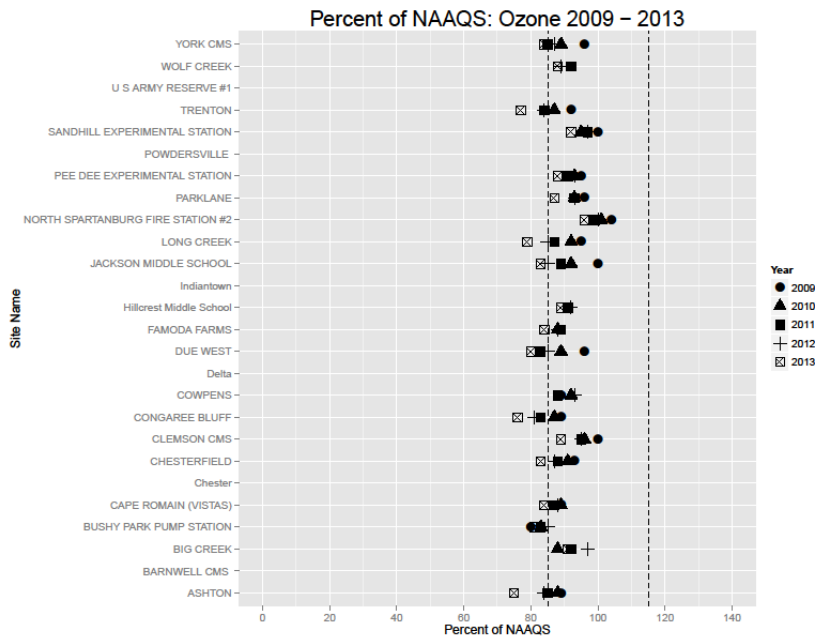
- Openair has a trajectory analysis function which can grid the backtrajectories.

- User can condition the data to look at trajectories on days in which ozone concentrations exceeded a certain threshold.
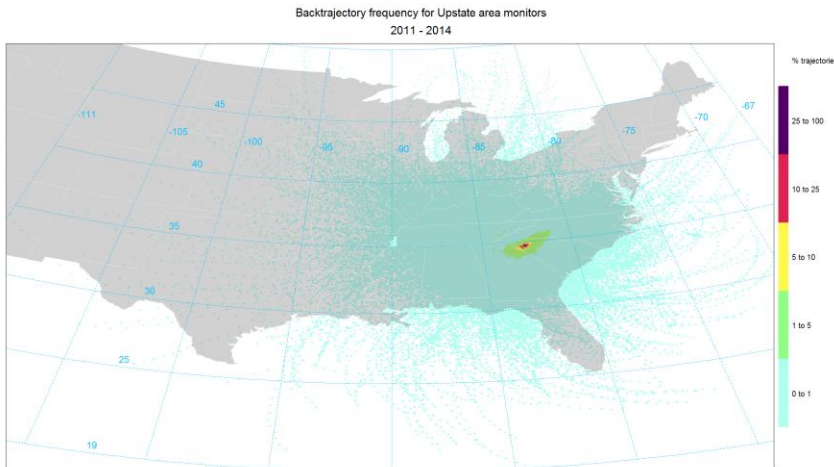
Questions and contact information?

SOUTHEAST AIR QUALITY ANALYSIS

How to join

- If there are you or other staff in your state are interested in joining or learning more please have them contact me, Ryan Brown, Daniel Garver or Darren Palmer.

- Will try to schedule the first call in May.

- A github account (https://github.com/southeast-air-data) has been established to help share and update code.

- This group is not just for R users. Anyone who has developed tools for analyzing ambient air data in any software package (including Excel) are welcome to join. We want this group to be a resource for all to regardless of coding skill.

Contacts

- Tommy Flynn
    - flynntj@dhec.sc.gov (803) 898-3251
- Ryan Brown
    - Brown.Ryan@epa.gov (404) 562-9147
- Daniel Garver
    - Garver.Daniel@epa.gov (404) 562-9839
- Darren Palmer
    - Palmer.Darren@epa.gov (404) 562-9052