EPA Source Apportionment

Spring 2026 Capstone

Deeya Datta, Prateek Mishra, & George Whittington
Client: Mike Cyterski, EPA

Table of contents

01

Recap

Our Project Data and New Client

Source Apportionment

Source apportionment is a "collection of techniques to provide information regarding how much a source (usually a generalized category) contributes to the overall pollutant concentration at receptor (usually a monitoring site) (Rizzo, 2010).

The overall goal is to extract latent factors from the data. We can then interpret these factors to identify specific source profiles.

The Datasets

Baltimore

focuses on fine particulate matter (PM2.5) chemical composition (n=630)

Baton Rouge

focuses on volatile organic compounds (VOCs) rather than particulate matter (n=307)

St. Louis

focuses on fine particulate matter (PM2.5) chemical composition (n=418)

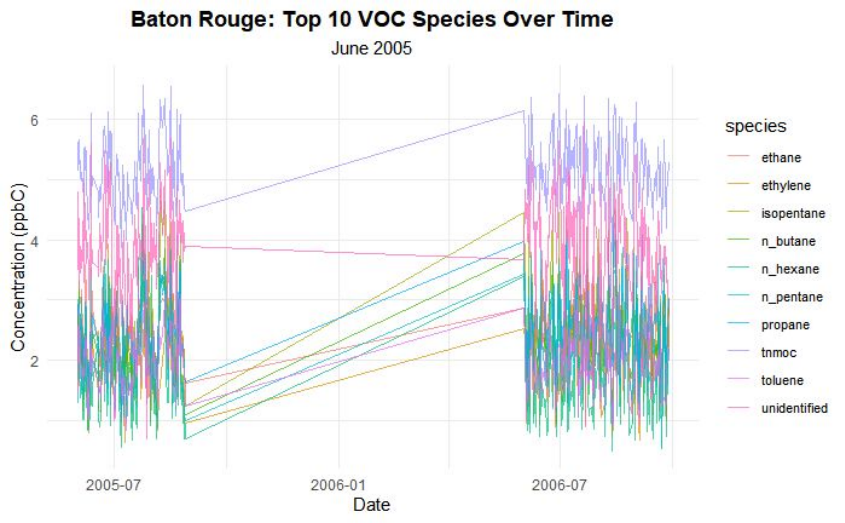
Variables Across Cities

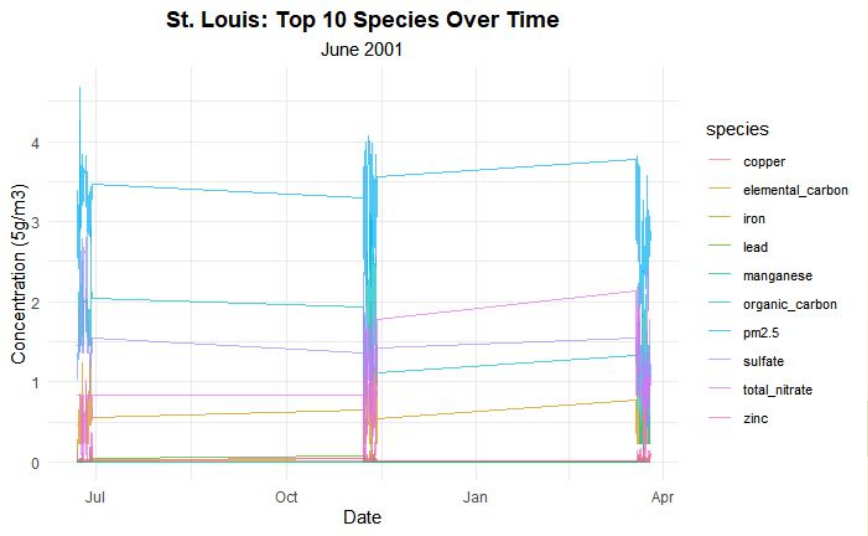
City	Category	Count	Specific Species Included
Baltimore	Trace Elements	17	Al, As, Ba, Br, Ca, Cl, Cr, Cu, Fe, Mn, Ni, Pb, Se, Si, Ti, V, Zn
	Ions	5	NH_4^+, K^+, Na^+, SO_4^{2-}, NO_3^-
	Carbon & Mass	4	EC, OC, OM, PM2.5
Baton Rouge	VOCs	41	BTEX, Alkanes, Alkenes & Isoprene
St. Louis	Trace Elements	8	Cd, Cu, Fe, Mn, Ni, Pb, Se, Zn
	Ions	2	SO_4^{2-}, NO_3^-
	Carbon & Mass	3	EC, OC, PM2.5

02

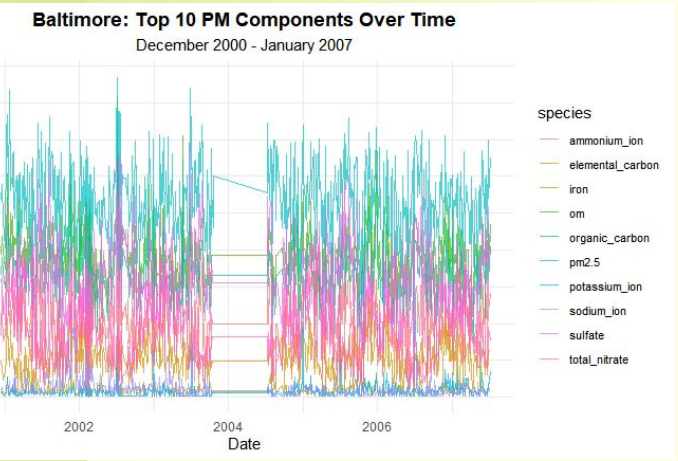
Time Series

Temporal Patterns in the Data

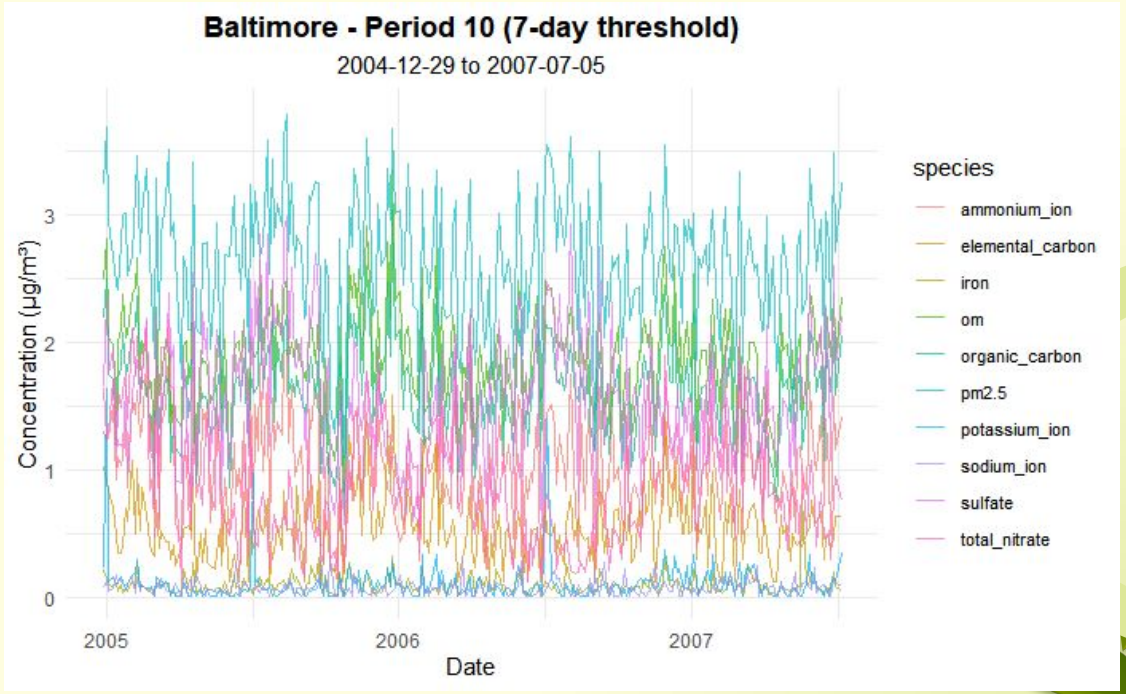




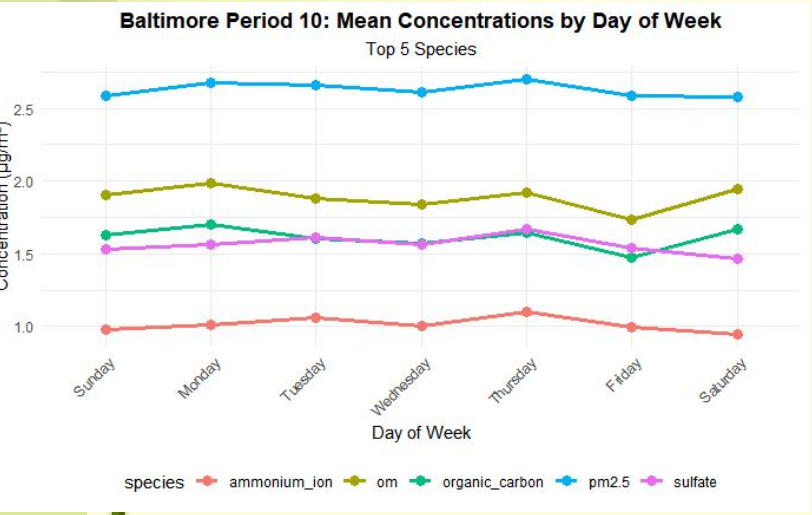
- All three of our datasets had gaps in the data
- Broke all of them up into more continuous chunks
- Split Baton Rouge 3am and 6am



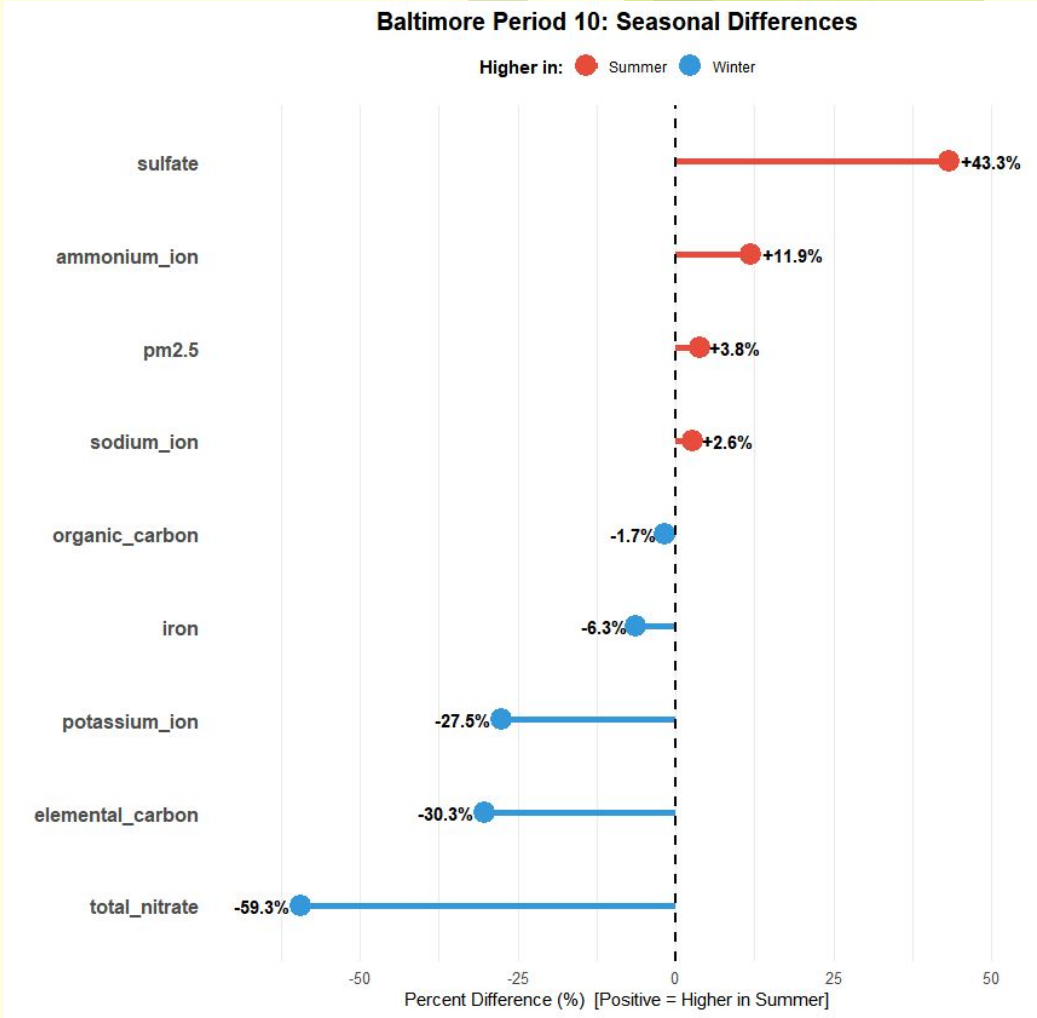
Example: Baltimore



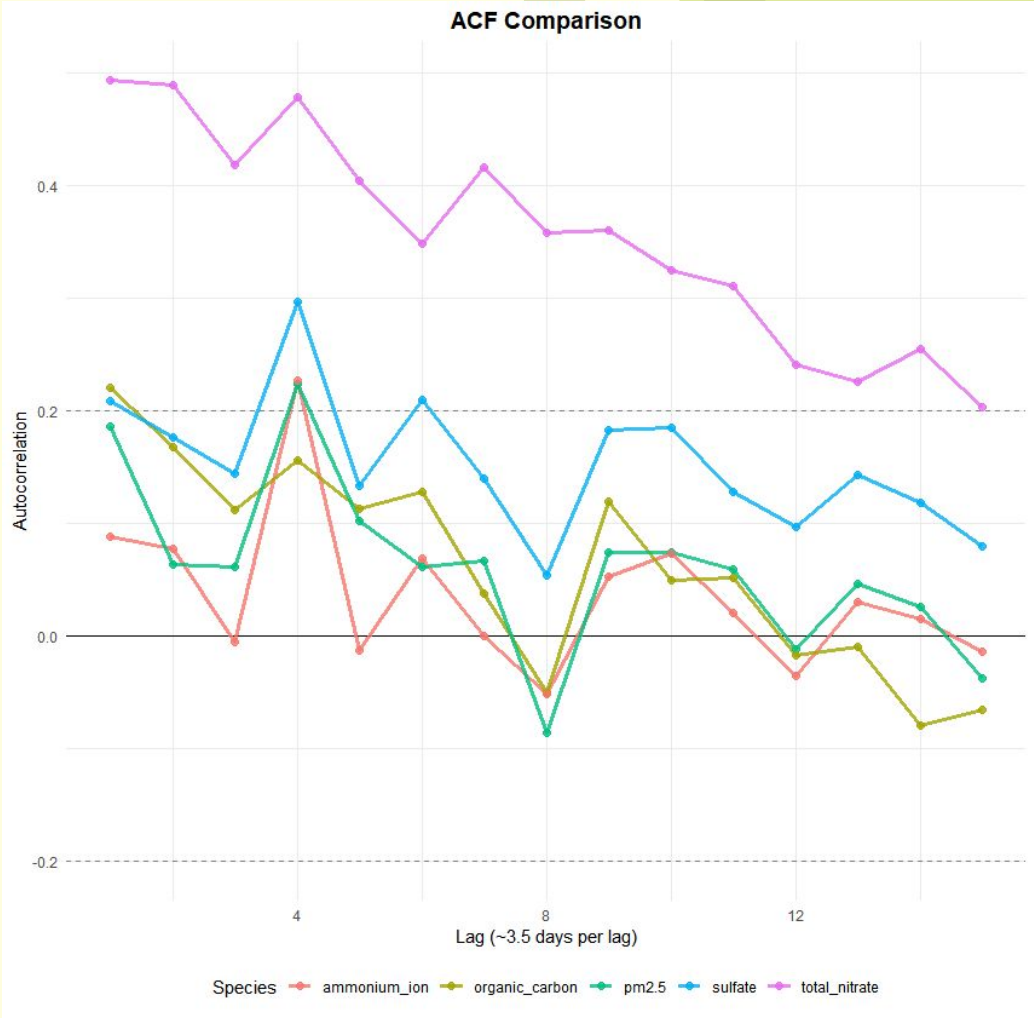
- Broke Baltimore into a total of 10 continuous periods
- Period 10 was the most robust so it is our starting point for time series analysis



- Looked at periodic trends within the period
- No significant periodicity looking at days of the week
- However seasonal differences were found



- Currently working on further time series analysis
  - ACF and PACF
- Started with top 5 species in just Period 10 of Baltimore
- Will expand to some more species, then replicate with the other periods in Baltimore
- Hopefully repeat a similar process with other two regions

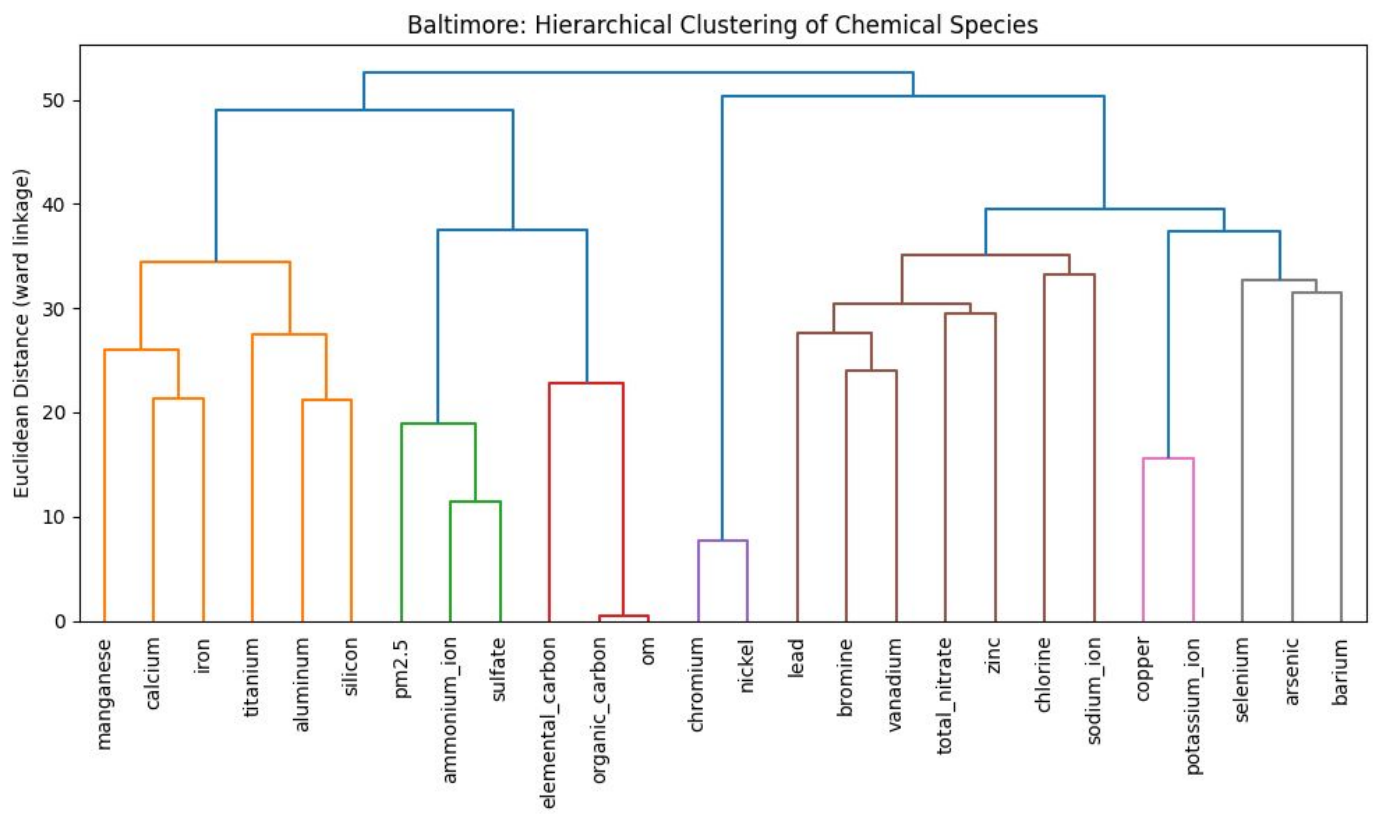


03

Clustering

Grouping by similarity rather than time

Baltimore

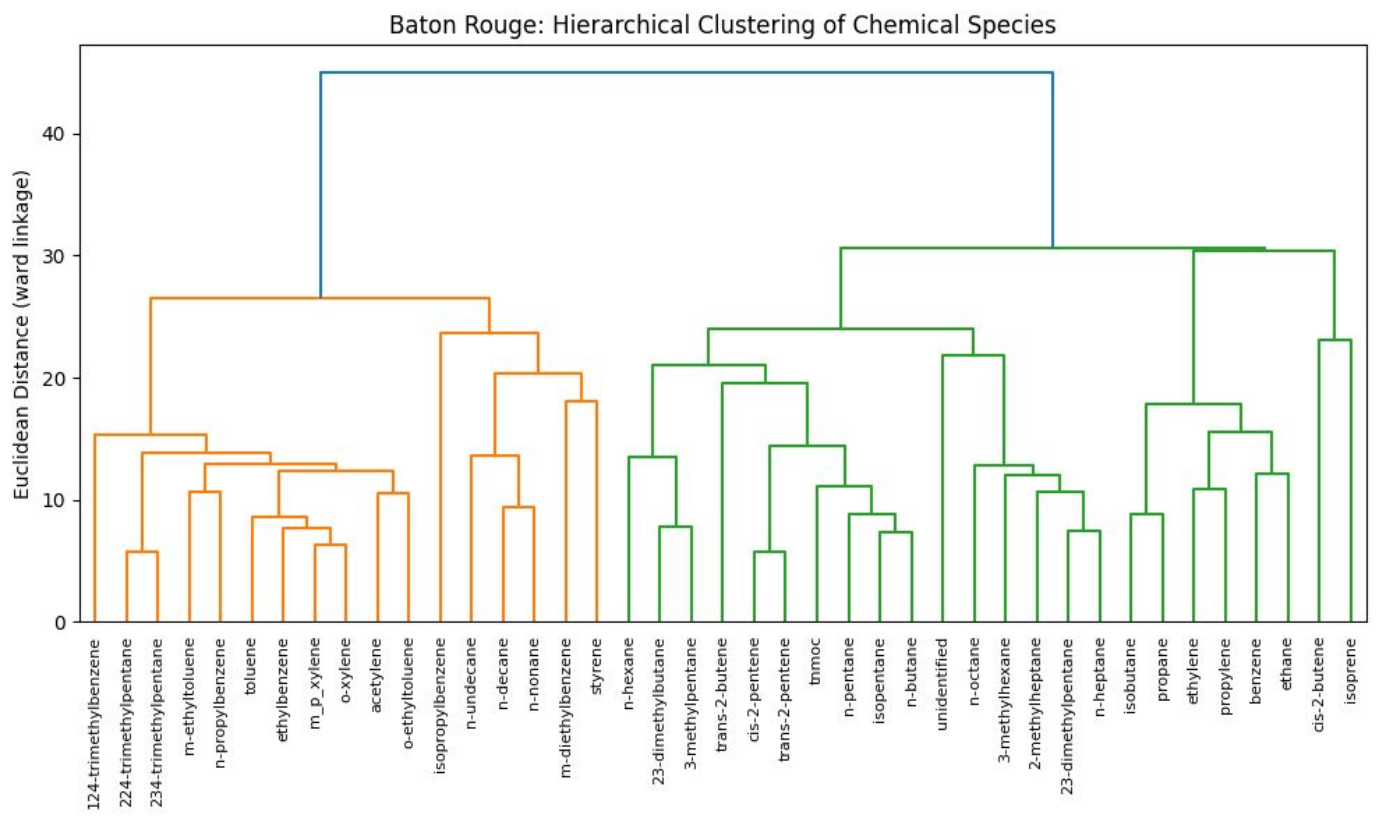


- Organic matter and Organic Carbon cluster at around 0 dist

- Sulfate, Ammonium cluster with PM
- Signs of secondary aerosol formation

- Chromium + Nickel form a distinct, isolated cluster.

Baton Rouge

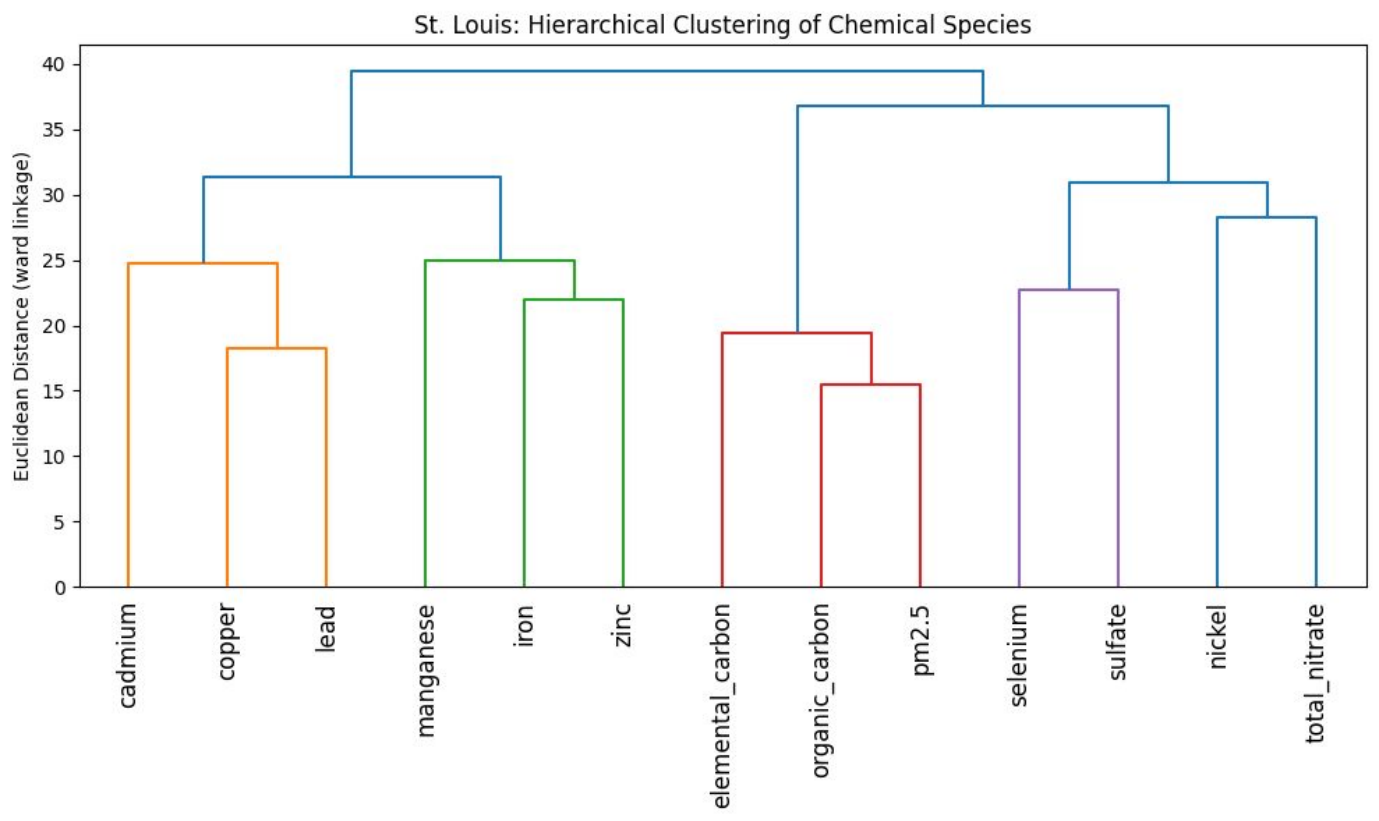


- Aromatics (Benzene, Toluene, Xylenes) cluster together, standard vehicle exhaust signature

- Light Alkanes (Ethane, Propane) form a completely separate group

- Simple model successfully distinguishes refinery/petro-chemical leaks from car exhaust.

St. Louis



- Elemental Carbon + Organic Carbon cluster together
- Burning stuff is the primary driver of particle pollution event

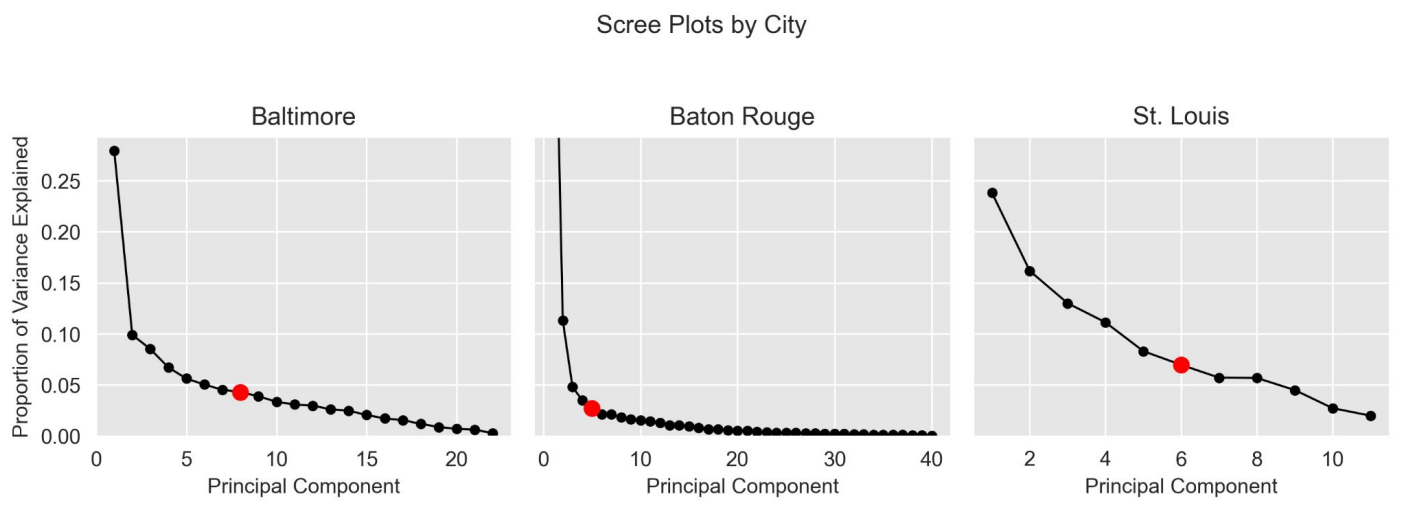
- Sulfate groups specifically with Selenium
- Selenium is a unique tracer for coal

04

PCA

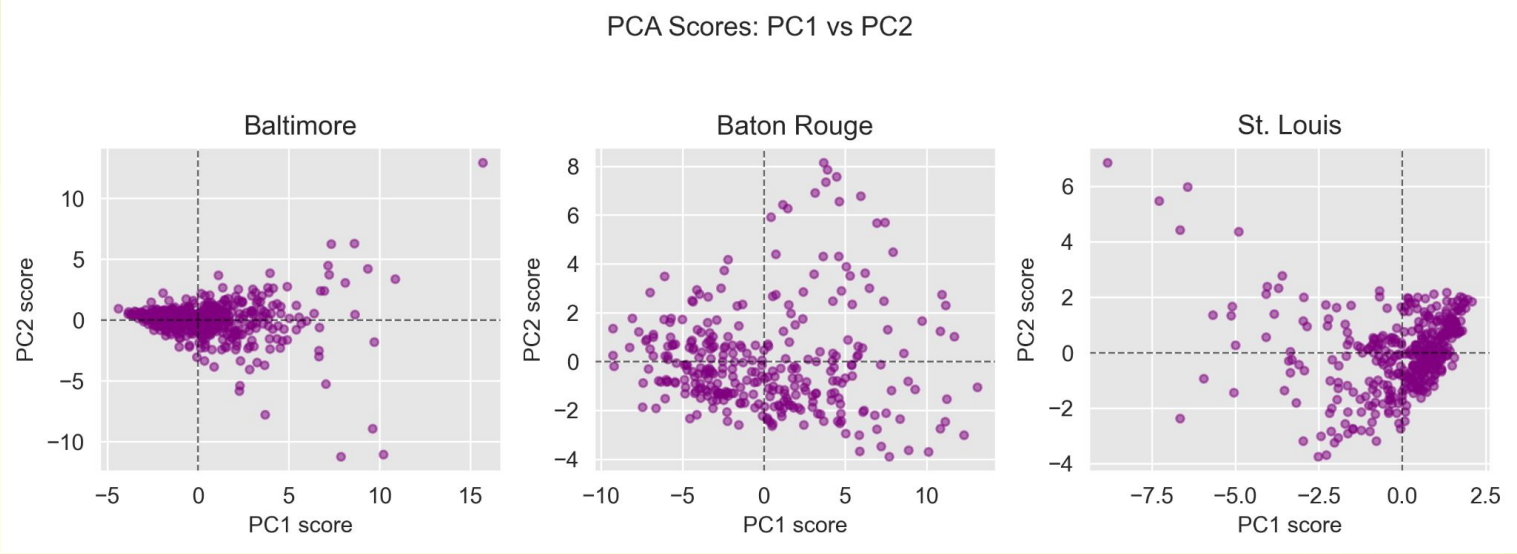
Finding components of variance

Choosing Principal Components



City	Baltimore	Baton Rouge	St Louis
Number of PCs	8	5	6
Total Variance	72.5%	77.1%	79.4%

PC1 vs PC2



Score distributions along the first two principal components, with Baton Rouge exhibiting greater dispersion in the PC1–PC2 space.

PC1 Factor

City	Dominant Contributors	Interpretation
Baltimore	Fe, OC, Br (+); Na^+, Ba, Ni (–)	Contrast between combustion metals *(Fe, OC, Br)* and marine influenced species *(Na^+, Ba, Ni)*
Baton Rouge	TMPs, n-Heptane, DMPs (+); Isoprene, light alkenes (–)	Industrial fuel VOCs *(TMPs, n-Heptane)* contrasted with biogenic hydrocarbons *(Isoprene, alkenes)*
St Louis	SO_4^{2-}, Se, Ni (+); Pb, Zn, Fe (–)	Contrast between sulfate and trace metals *(SO_4^{2-}, Se, Ni)* from primary metals *(Pb, Zn, Fe)*

This shows us covariance that guides factor selection and structure

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**