CIVIC FIELD SCHOOL PROJECT: THE CITIZEN DATA SCIENTIST

Investigating Air Quality of Philadelphia

by Tinashe M. Tapera

MAJOR THEMES

1. The volume of open data available for Philadelphia

2. The poor navigability of data platforms

3. The *citizen data scientist*

AIMS FOR THIS CLASS

Before being familiarised with the scope of the class…

- Get rid of a history elective
- Take an accelerated course

After being familiarised with the scope of the class…

- Contribute to The Asthma Files big data platform
- Develop an understanding of environmental health in Philadelphia
- Conduct novel research in the region of environmental health governance
- Exercise data mining and analysis skills as a *citizen scientist*.
- …Asthma!

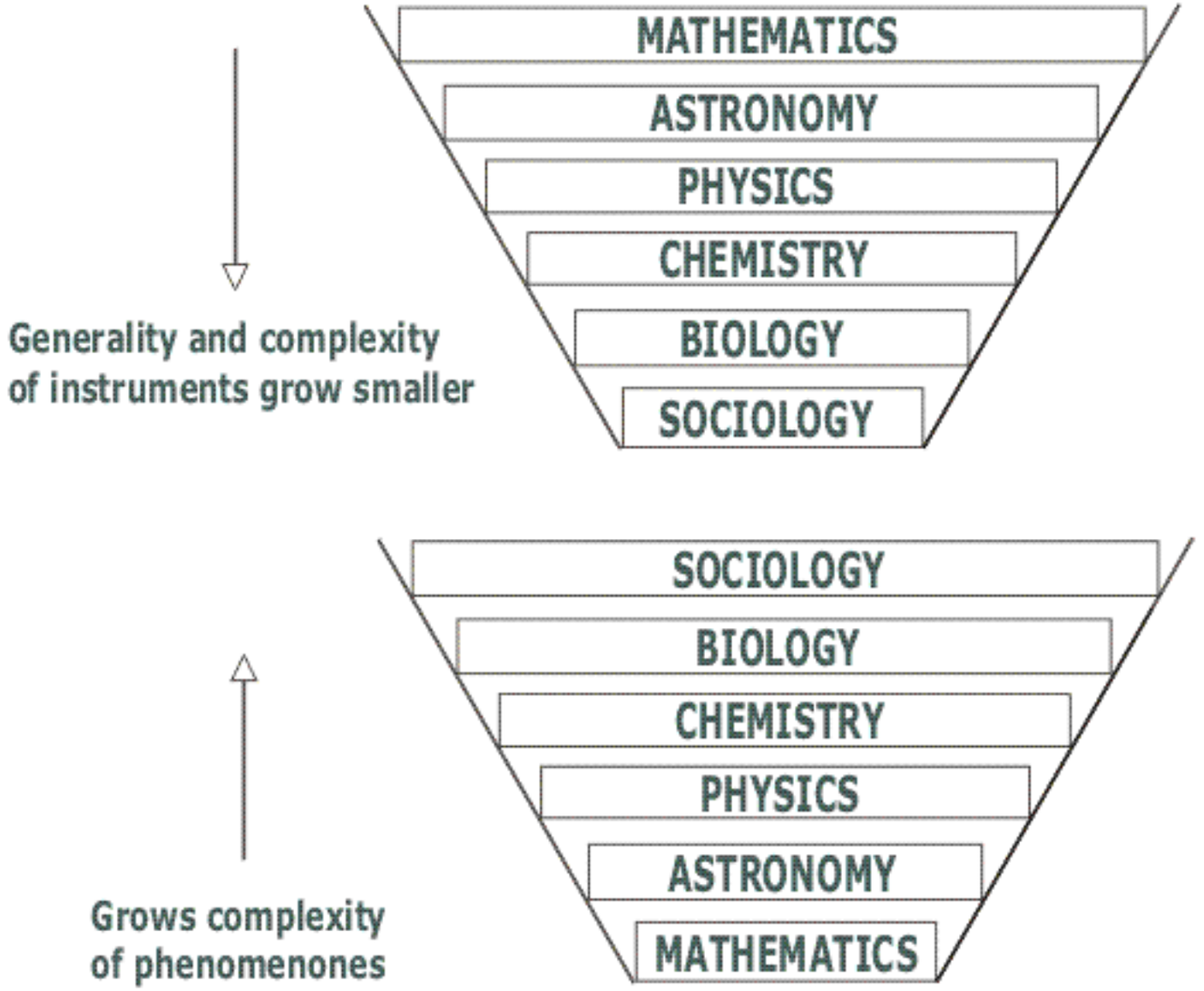
"Citizen science is the involvement of the public in scientific research — whether community-driven research or global investigations."

–CitizenScience.org

"Street Science engages several issues that remain largely unaddressed in the environmental-policy and public-health literatures."

–Jason Corburn

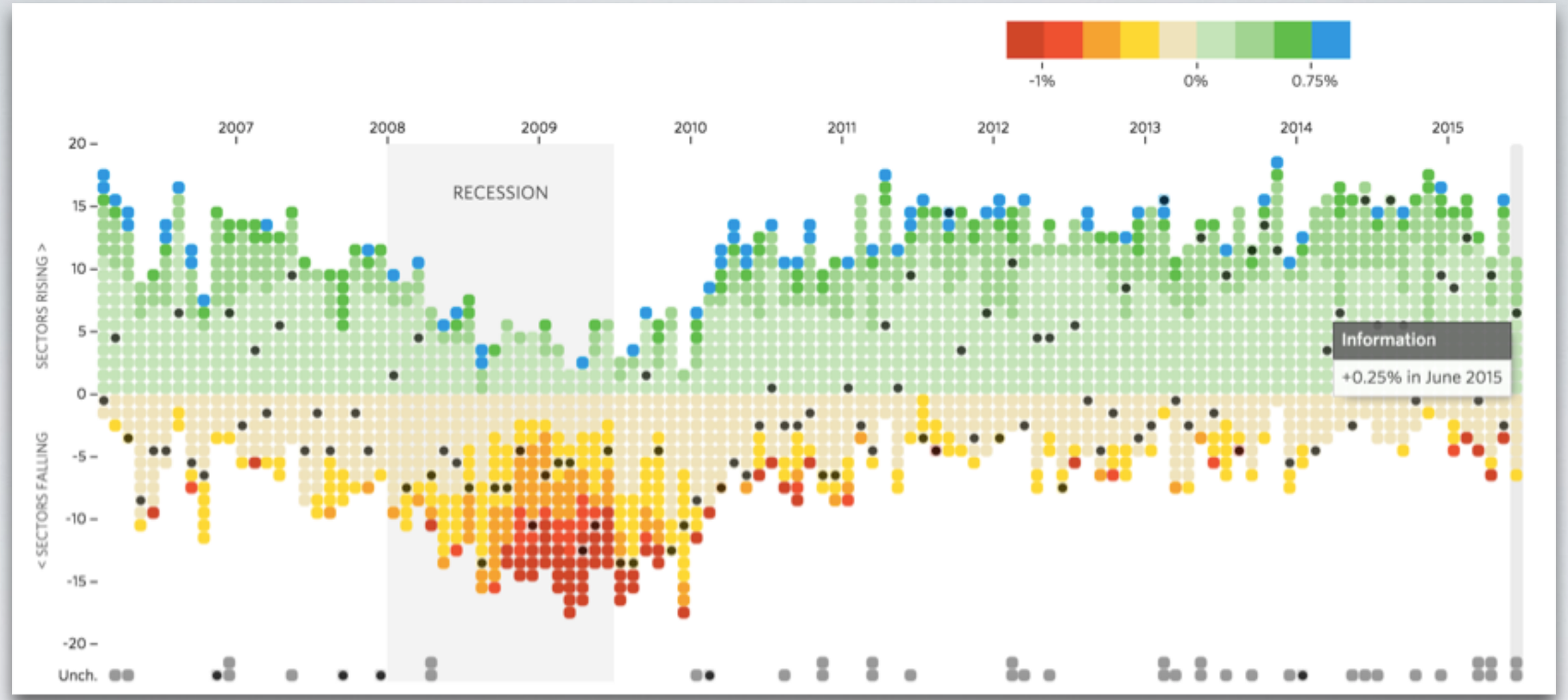
Issues of environmental health affect people individually and uniquely…



COMTE'S THEORY OF SCIENCE

(Environmental health governance would generally fit underneath sociology).

How do we equip more citizens to be street scientists?



BIG DATA & DATA SCIENCE

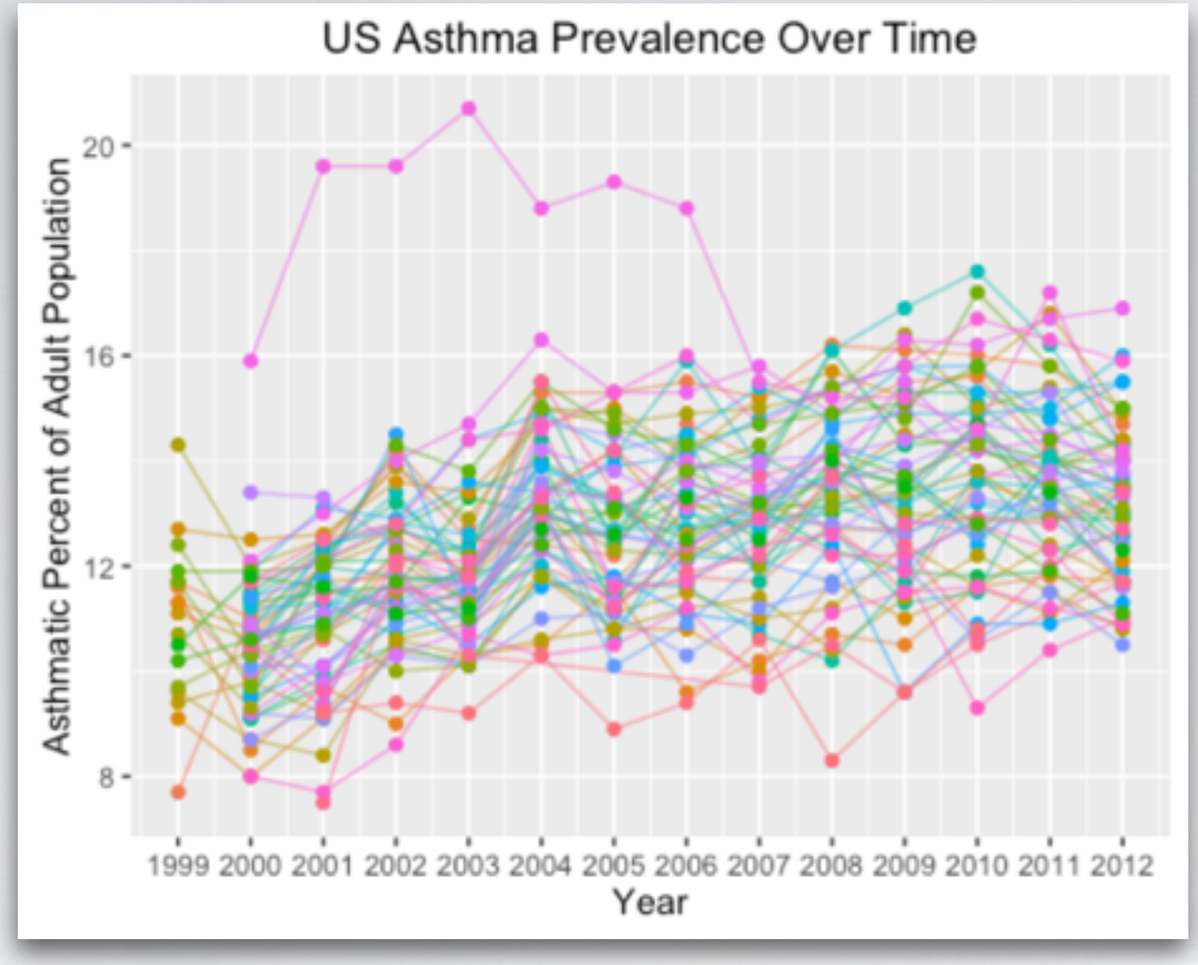
DATA SCIENCE IN BRIEF…

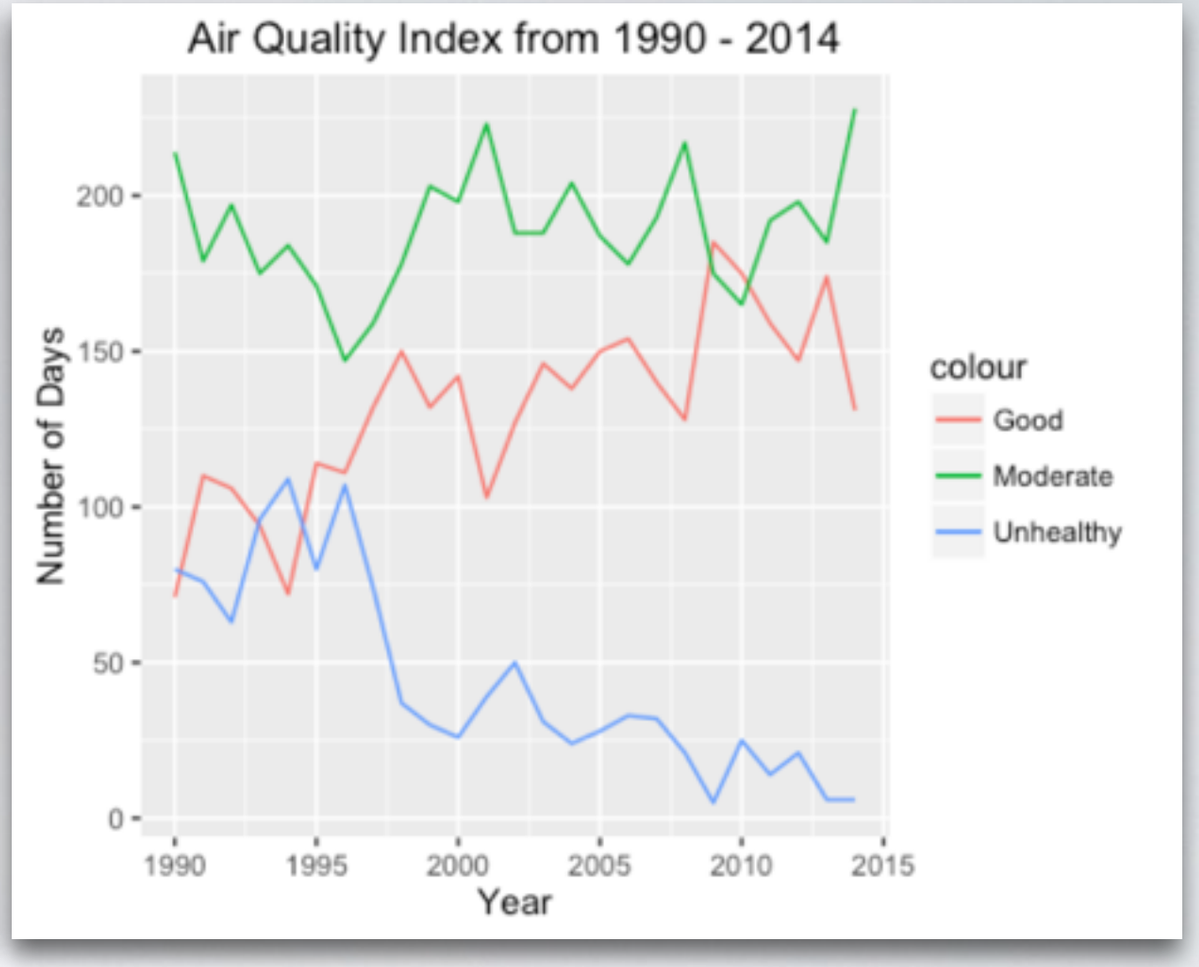
- The volume, velocity, and variety of data being generated by human beings is ever increasing.
    - "2.5 exabytes - that's 2.5 billion gigabytes (GB) - of data was generated **every day in 2012.**" — CNN.com
- This data requires specialists to ***collect, warehouse, analyse, and derive insights from it***.
- These specialists are known as ***data scientists*** and are typically experts in hacking/programming, statistical analysis/machine learning, and domain knowledge.

If resources for learning data science are open-source,

And data are open-source…

It means citizens now have the opportunity to be data scientists and participate in street science.





USE CASE: AIR QUALITY DATA

Investigating open source air quality data in Philadelphia

AIR QUALITY INDEX

Air quality index (AQI): a number used by government agencies (such as the EPA) to communicate regularly to the public how polluted the air currently is or how polluted it is forecast to become.

EPA United States Environmental Protection Agency

AQI AIR QUALITY INDEX

Air Quality Guide for Particle Pollution

Good	0-50	None
Moderate	51-100	Unusually sensitive people should consider reducing prolonged or heavy exertion
Unhealthy for Sensitive Groups	101-150	People with heart or lung disease, older adults, and children should reduce prolonged or heavy exertion.
Unhealthy	151 to 200	People with heart or lung disease, older adults, and children should avoid prolonged or heavy exertion. Everyone else should reduce prolonged or heavy exertion
Very Unhealthy Alert	201 to 300	People with heart or lung disease, older adults, and children should avoid all physical activity outdoors. Everyone else should avoid prolonged or heavy exertion.

$$IAQI_p = \frac{IAQI_{Hi} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}}(C_p - BP_{Lo}) + IAQI_{Lo}$$

Q1: Why does the asthmatic percent of the population increase despite AQI showing improved air quality?

Q2: Why does the final reported AQI only take into account the highest AQI from all the pollutants?

ANALYTICAL APPROACH

- Explore and visualise public environmental health data:

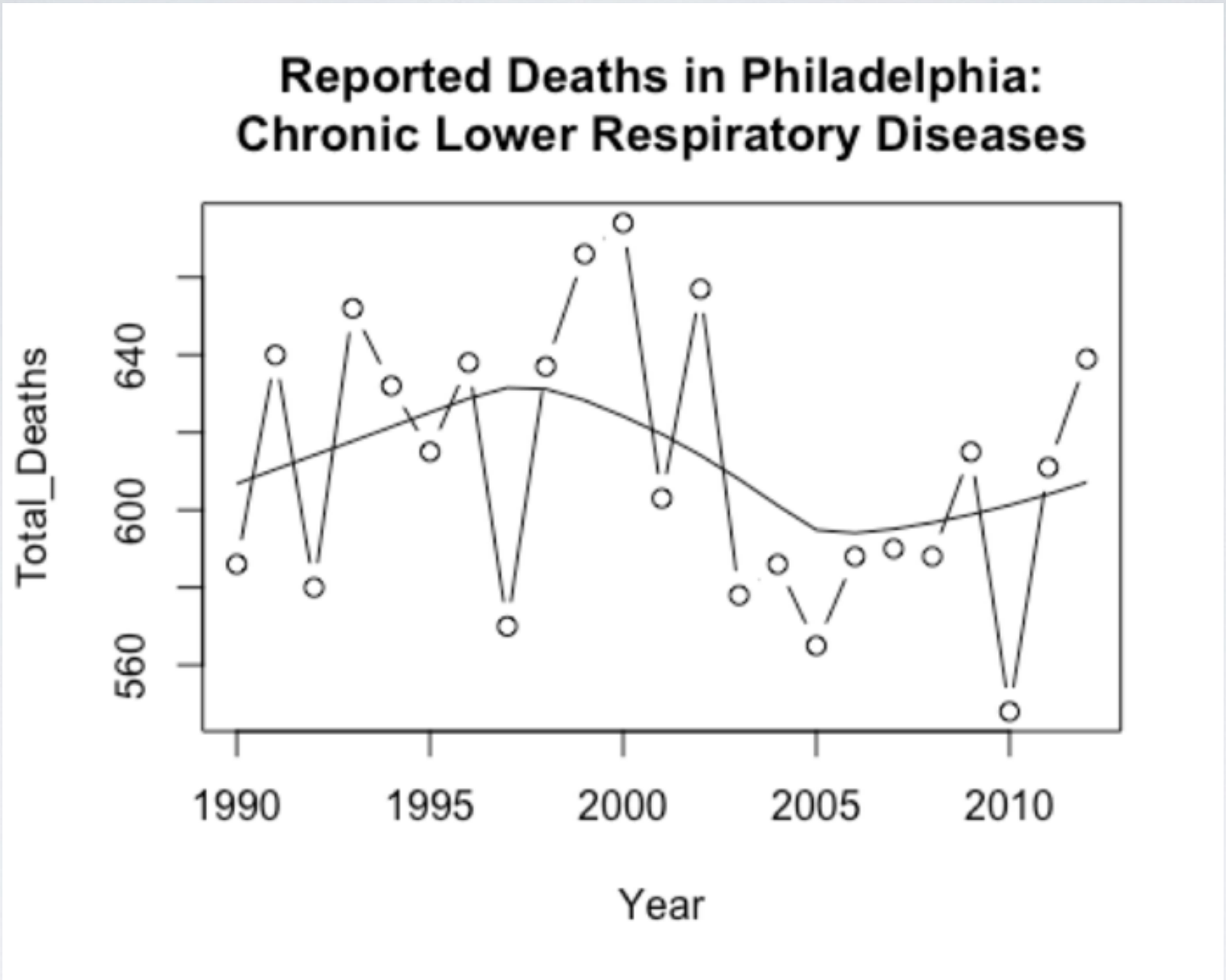
Sources: Centre for Disease Control, Environmental Protection Agency, phila.gov, PA.gov, OpenDataPhilly.org

- Identify interesting statistical trends:

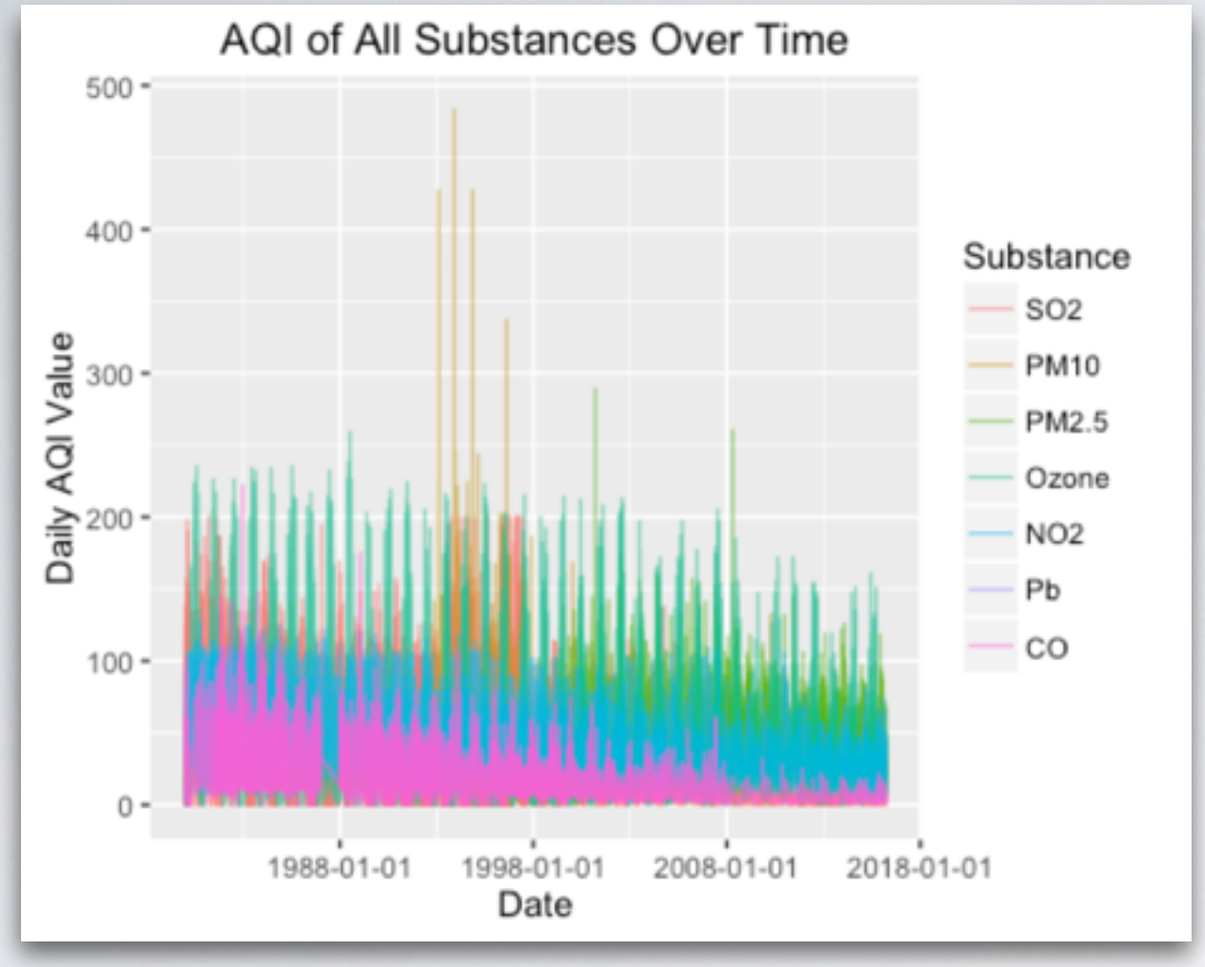
“Child hospitalisations due to chronic lower respiratory disease have more than doubled in the last decade.” — Phila.gov Vital Statistics Report 2014

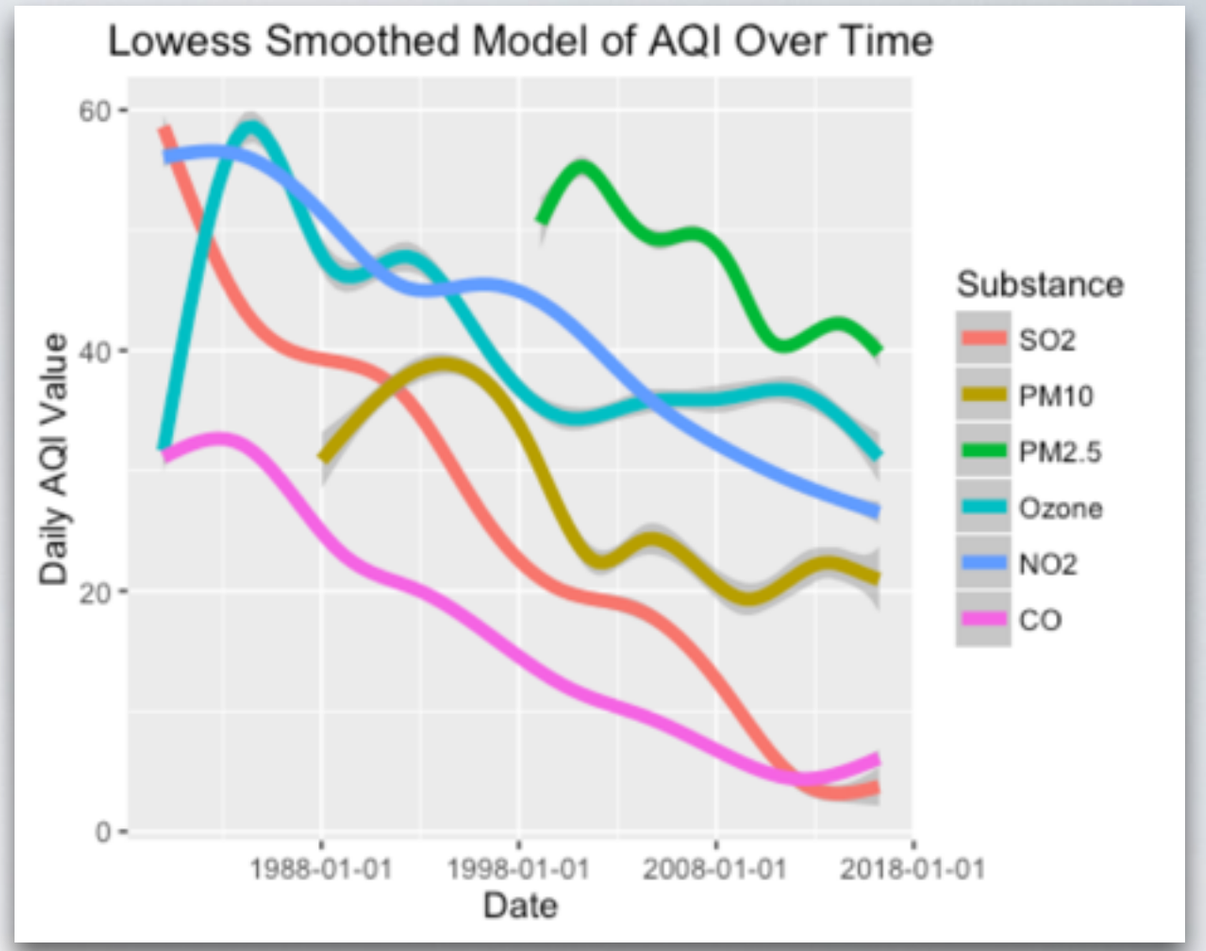
- Statistically investigate an aspect of EH governance that can be addressed through the data:

AQI that is reported may be unsound. Conduct literature reviews and data analysis



An unfortunate lack of cogent data…





SOME DATA WAS WORKABLE

Notice:
Differing slopes of decline
Slight rise in CO and SO2

MAJOR THEMES

1. The volume of open data available for Philadelphia: **A LOT**

2. The poor navigability of data platforms: **DEFINITELY**

3. The *citizen data scientist*: **There are discrepancies in the data and something is not statistically sound. Warrants further investigation. We may believe that AQI is falling from reports, but is our air quality really improving?**

CONCLUSION

- Recall Comart & Pasher (Corburn Ch.4)
- Citizen and street science take time and dedication
- Data scientists could be the future citizen scientists conducting street science (meetup.com provides meeting opportunities for collaboration on different interests)
- A level of being well-informed and learned, or at least able to access knowledge and instruction
- Openness between data providers, analysts, and citizens

Works Cited

Auguste Comte. (2016). Wikipedia. Retrieved 24 August 2016, from https://en.wikipedia.org/wiki/Auguste_Comte

Corburn, J. (2005). Street science. Cambridge, MA: MIT Press.

DUCSTeach. (2016). Retrieved 24 August 2016, from https://www.cs.drexel.edu/~dmz38/ducsteach/

Great Tour of Philadelphia — visitphilly.com. (2016). Philadelphia - Official Visitor Site - visitphilly.com. Retrieved 24 August 2016, from http://www.visitphilly.com/events/philadelphia/great-tour-of-philadelphia/

Phila.Gov | Public Health. (2016). Phila.gov. Retrieved 24 August 2016, from http://www.phila.gov/health/commissioner/VitalStatistics.html

Team:Bielefeld-CeBiTec/StreetScience - 2015.igem.org. (2016). 2015.igem.org. Retrieved 24 August 2016, from http://2015.igem.org/Team:Bielefeld-CeBiTec/StreetScience

Wall, M. (2016). Big Data: Are you ready for blast-off? - BBC News. BBC News. Retrieved 24 August 2016, from http://www.bbc.com/news/business-26383058

All analyses were run in R Studio Version 0.99