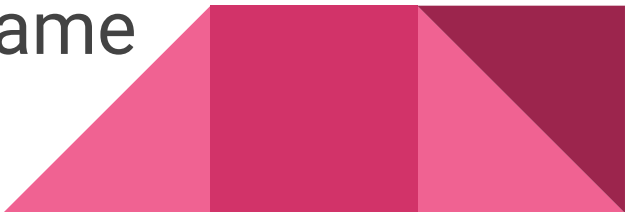




Matthew Manuel, Thien Hua, Leon Lin, Ramon Santos

## Intro: Deliverable

- Dataset was taken from airsofia, a website that tracks air quality
  - Was made in response to Bulgaria for failing to implement significant anti-pollution measures
  - Gives location coordinates, air quality (pressure, temperature, humidity) and timeframe
- 

# Background

- 2013 Sofia was singled out for a study by the European Environment Agency
- This was due to Sofia's presence of several air pollutants in the atmosphere
- Caused by gas and carbon monoxide from burning wood and coal.

# Background (cont)

- Natural factors to air pollution
  - Geographic location
  - Stagnation of air
- Human activity is the main source
  - Transportation with the use of diesel
  - Old heating system

## Background Cont.

- Two groundbreaking decisions made
  - Jnecek Case
  - ClientEarth Case
- These decisions allowed
  - The Right of Standing
  - The Right of Legal Remedy
  - The Right to a Substantive Review



## Related Work

- European Environment Agency in 2013
- World Health Organization in 2016
- Both studies focus on the rampant spread of air pollution
- We used big data obtained from a data source from an agency that tracks down air quality.
- Focus on temperature of a specific region and noting changes in values over time

# Specifications

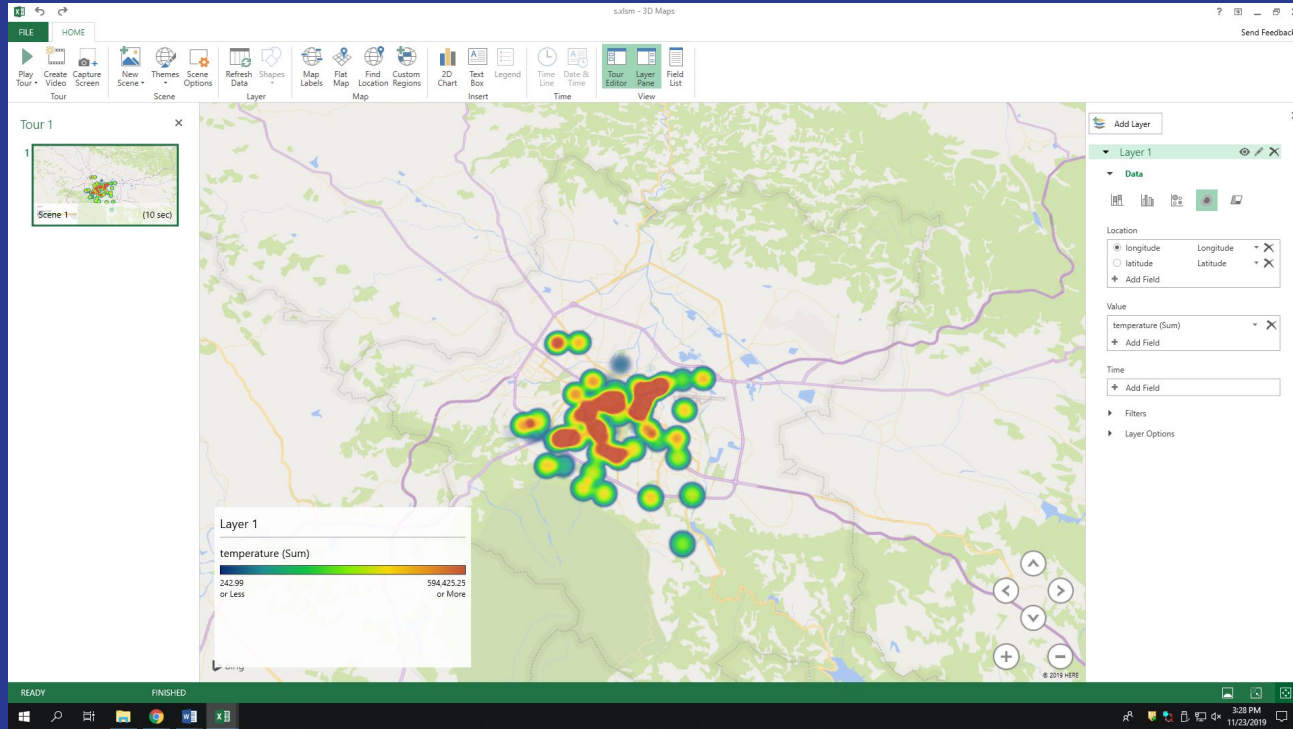
EMR cluster with 3 nodes: m3.xlarge

**Release label:**emr-5.27.0

**Hadoop distribution:**Amazon 2.8.5

**Applications:**Hive 2.3.5, Pig 0.17.0, Hue 4.4.0,  
Ganglia 3.7.2, Tez 0.9.2, Sqoop 1.4.7, HCatalog 2.3.5

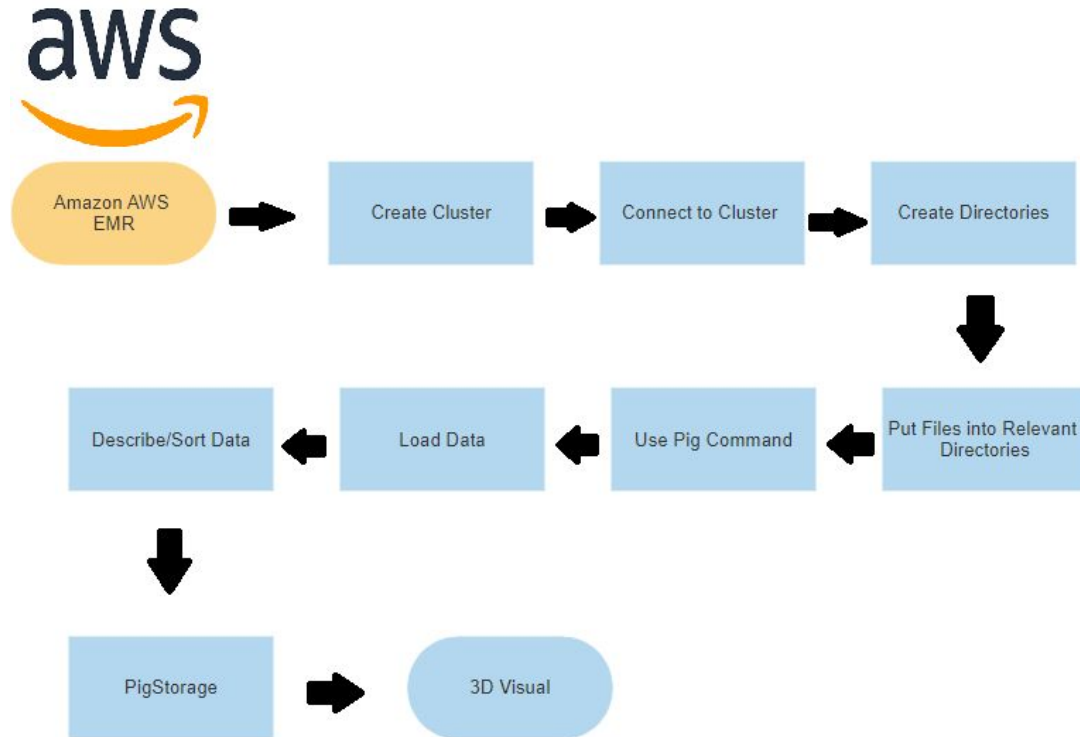
# Our Work



Temperatures in a specific location due to air quality



# Data Flow Chart



Data Url

Data Size: 14 GB

<https://www.kaggle.com/hmavrodiev/sofia-air-quality-dataset>