EnVis - Project Presentation

A visualization tool for environmental quality

Laura Papi, Francesca Restante

Visual Analytics A.Y. 2022-2023

Environmental pollution is an incurable disease. It can only be prevented.

- Barry Commoner

Introduction

- Recent growing awareness of environmental problems
- Several studies assess the necessity for a way to visualize environment data

State of the art

- Designed for domain experts
- Mainly focused on air quality

System Design

- A tool for non-expert users to grasp information and gain insights on pollution in its geographical area

- Analyzing the relations of primary pollution factors (air and water) with many other secondary factors that can still have an impact on the environmental quality

What is "Environmental Quality"?

- Primary factors
  - Air pollution
  - Water pollution
- Secondary factors
  - Presence of green areas
  - Types of circulating vehicles
  - Accessibility to public transport
  - Industrial and commercial noise
  - Waste management

Data Collection

FACTOR	DATA SOURCE
Air pollution	European Environment Agency, Annual AQ statistics https://discomap.eea.europa.eu
Water pollution	ISPRA, Portale Pesticidi https://sinacloud.isprambiente.it
Presence of green areas	ISTAT, Verde urbano (Tav 9.1) https://www.istat.it/it/archivio/281184
Types of circulating vehicles	ISTAT, Mobilità Urbana (Tav 4.1) https://www.istat.it/it/archivio/281184
Accessibility to public transport	ISTAT, Mobilità Urbana e trasporto pubblico (Tav 14) https://www.istat.it/it/archivio/93269
Industrial and commercial noise	ISTAT, Rumore (Tav 5.1 | 6.1) https://www.istat.it/it/archivio/281184
Waste management	ISTAT Rifiuti Urbani (Tav 7.1)https://www.istat.it/it/archivio/281184

Data Pre-processing

- Air pollution
  - data filtering with EEA's online tool
  - compute the mean value for each city and for each pollutant
- Water pollution
  - merge the dataset from each region into a national one
  - compute the max value of each pesticide for each city
  - dimensionality reduction

Dimensionality Reduction - PCA

- Dataset Transformation

  The dataset obtained by merging all region datasets is in the form:
  [City, Pesticide, Value]

  To be able to represent it we need to transform it into:
  [City, Pest1, Pest2, ..., PestN]

  Where each column corresponds to a pesticide and contains its measured values.
  (Operation performed using pivot_table function from pandas library).

Dimensionality Reduction - PCA

- Principal Component Analysis
  - Dropped the column containing the City name
  - Scaled the values using a StandardScaler ($(x\text{-}\mu)/\sigma$)

```
scaled = StandardScaler().fit_transform(transformed_data)
```

  - Compute PCA on 2 components

```
pca = PCA(n_components=2)
components = pca.fit_transform(scaled)
```

  - Reattach the City name column
  - Save to a csv file in order to render the data in a graph. Now in the form: **[City, PC1, PC2]**

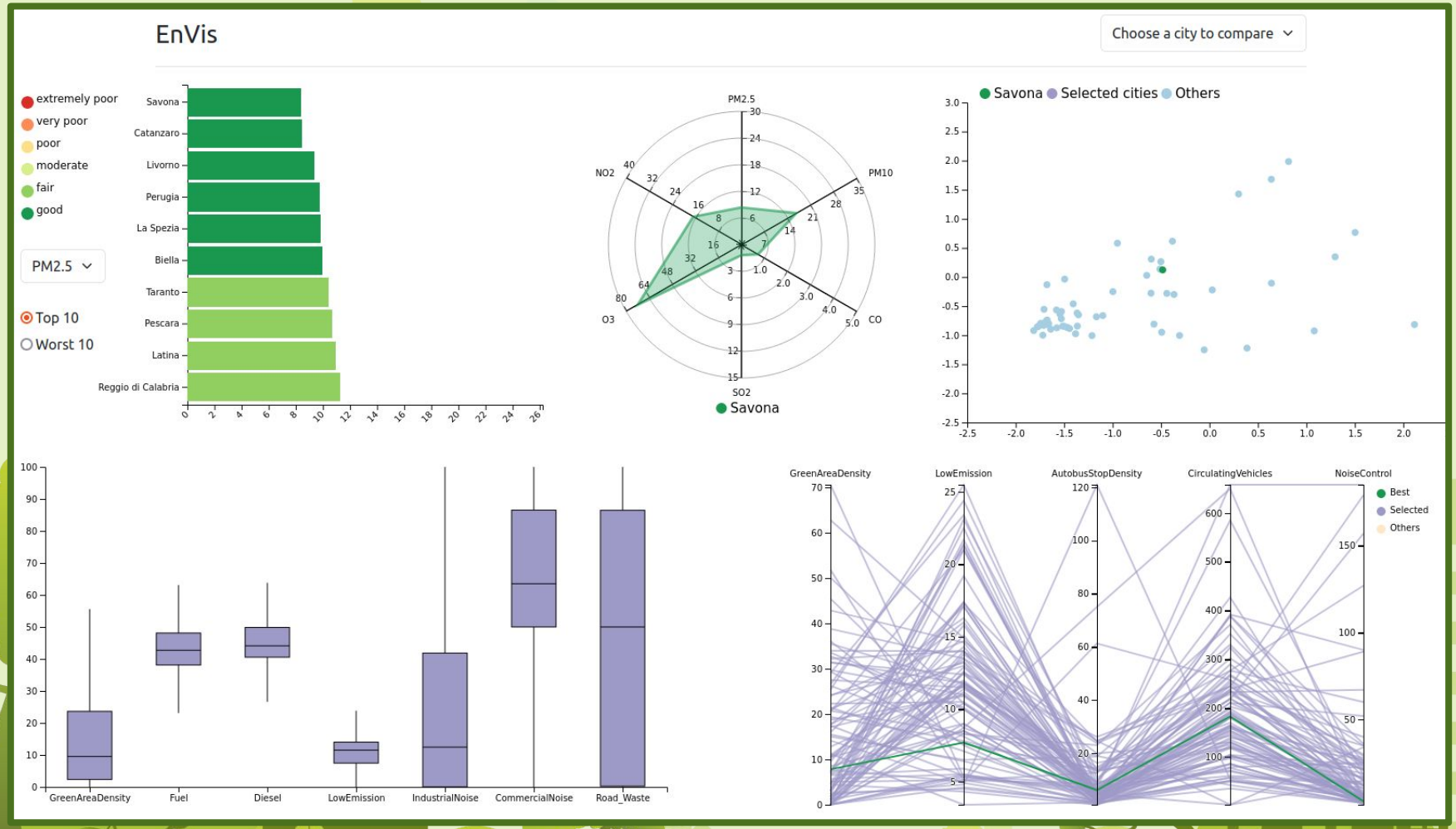
Data Pre-processing

- Presence of green areas
- Types of circulating vehicles
- Accessibility to public transport
- Industrial and commercial noise
- Waste management

All these factors are combined in files concerning the last two graphs. In particular, in the file for the boxplot categories are assigned as a field in the respective column “Category” to distinguish the type of data and then use it for computation.

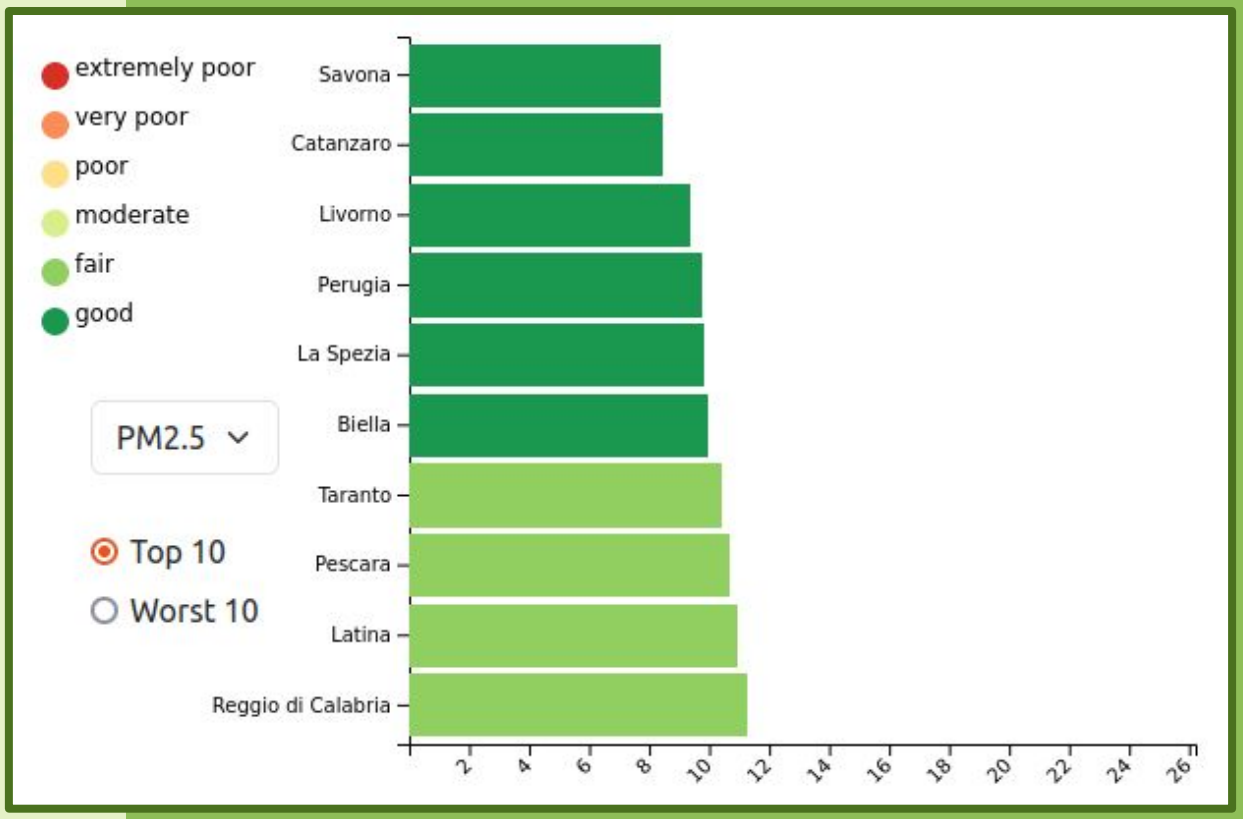
Visualization System

Preview of the complete visualization



Air Quality Ranking

Bar chart that ranks cities based on the mean pollution level of a specific substance in the air.



The user can choose one between 6 of the most impacting pollutants (the ones defining the U.S. Air Quality Index).

He can also choose to display the best 10 cities or the 10 worst.

Air Quality Benchmark

Radar chart that displays at the same time the best ranked city and the city chosen by the user for comparison.

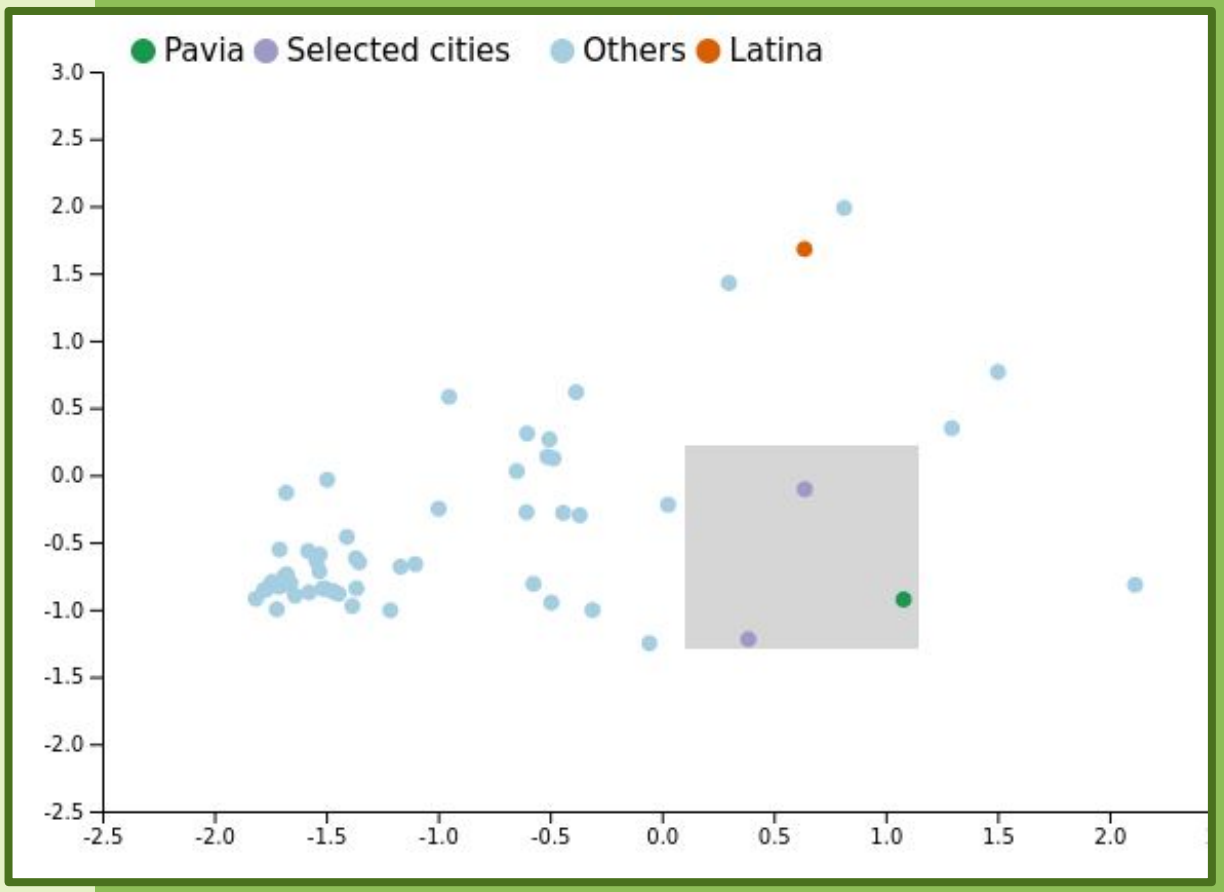
Why the radar chart?

- It is a great tool to summarize data of a single city, by displaying the area extent, and also the values for multiple parameters

- Provides a very immediate and intuitive way to compare two cities

Water Quality Analysis

Scatter plot to display the data generated through the PCA algorithm on pesticide presence in surface water.

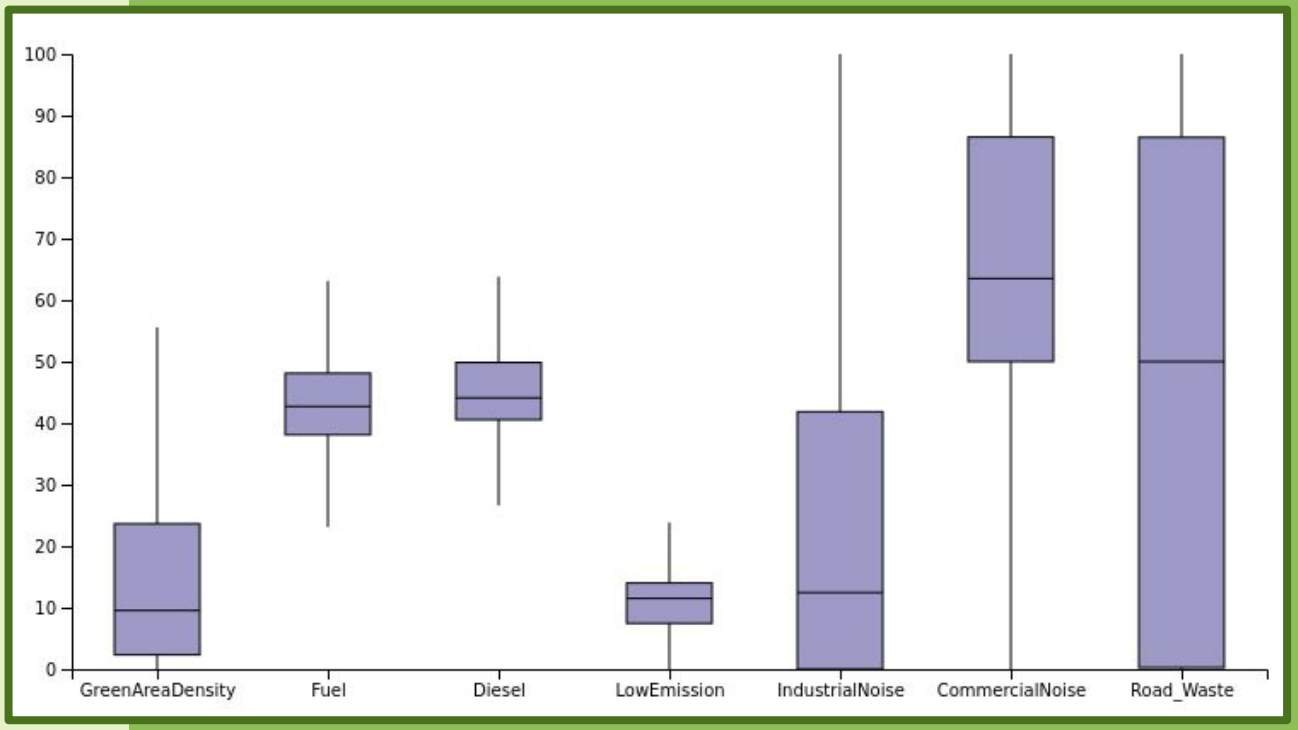


Allows to see which cities are more similar, and identify clusters.

Provides user interaction, it can be brushed to select a subset of the cities on which to recompute all the other visualizations.

Secondary Factors Statistics

Box plot to visualize and examine the distribution and statistical summary of secondary factors data.



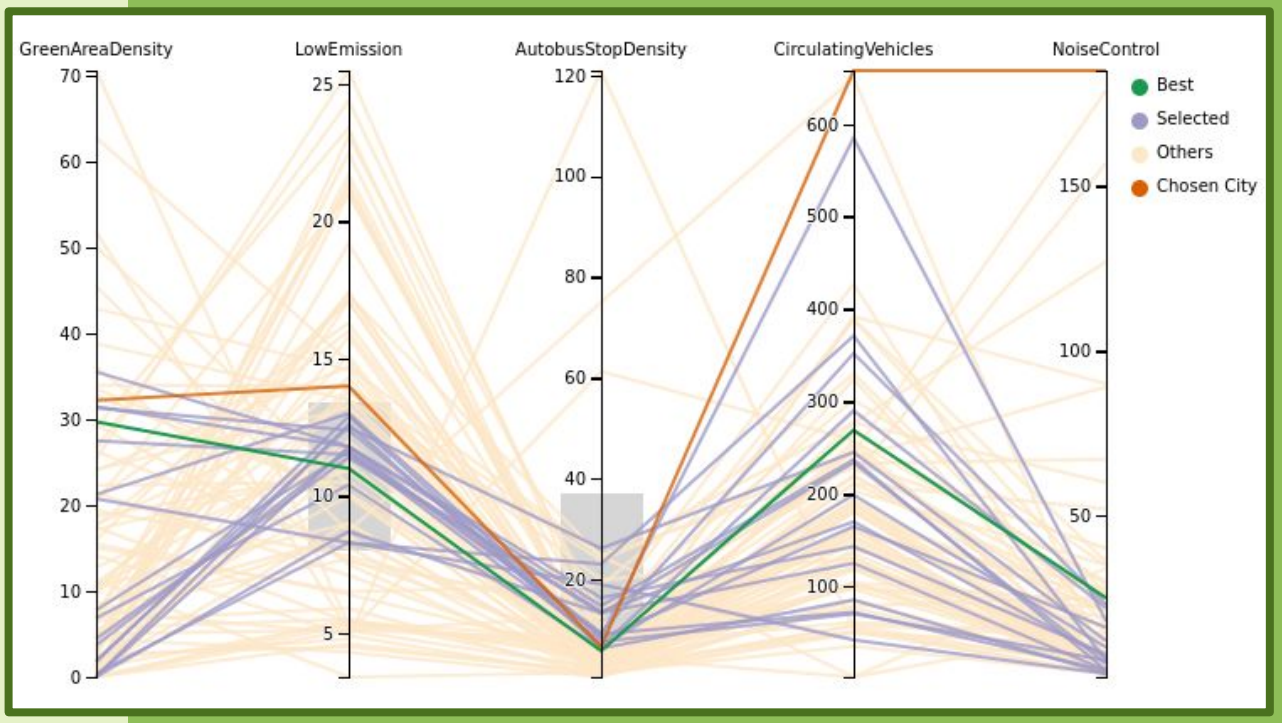
It is useful for a quick assessment of the data's central tendency and spread.

All the values are expressed in percentage.

It will be updated based on the set of cities selected in the others graphs.

Secondary Factors Analysis

Parallel coordinates plot to explore and understand multivariate data based on secondary factors.



Each factor has its own scale and a vertical axis, and a series of connected lines or curves are drawn across these axes to represent individual data points

It is possible to select a range of cities along the axis and update consequently all other graphs.

Analytics

The main components are represented by the boxplot, which do a dynamical computation of statistical factors such as:

- Median: the middle value of the dataset when it is sorted in ascending order.
- Quartiles: divide the dataset into four equal parts. The first quartile (Q1) represents the lower 25% of the data, while the third quartile (Q3) represents the upper 25% of the data. The interquartile range (IQR) is the difference between Q3 and Q1
- Whiskers: The whiskers of a boxplot represent the range of the data beyond the quartiles. They are computed as:

Upper Whisker: Q3 + (1.5 * IQR)

Lower Whisker: Q1 - (1.5 * IQR)

Live Demo

Conclusions

The proposed system provides a way for a non-expert user to gain informations about the environment he lives in.
It is designed to help a user decide where to live, or understand how to improve its own city.

Future Work

Our work presents an overview of environmental quality using the most recent available data at the moment.
It might be interesting to introduce the visualization of data over time, to allow a comparisons and forecasts.

References

Related Works

Z. Deng et al., "AirVis: Visual Analytics of Air Pollution Propagation," in IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 1, pp. 800-810, Jan. 2020, doi: 10.1109/TVCG.2019.2934670.

Y. -R. Zeng, Y. S. Chang and Y. H. Fang, "Data Visualization for Air Quality Analysis on Bigdata Platform," 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 2019, pp. 313-317, doi: 10.1109/ICSSE.2019.8823437.

R. Gupta, N. Singh and B. Singh, "Neural network based GIS application for air quality visualization," 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Dehradun, India, 2016, pp. 169-172, doi: 10.1109/NGCT.2016.7877409.

Niharika, Venkatadri M, and Padma S.Rao, A survey on Air Quality forecasting Techniques, vol. 5(1). International Journal of Computer Science and Information Technologies, 2014, pp.103-107

Lu F, Xu D, Cheng Y, Dong S, Guo C, Jiang X, Zheng X. Systematic review and meta-analysis of the adverse health effects of ambient PM2.5 and PM10 pollution in the Chinese population. Environ Res. 2015 Jan;136:196-204. doi: 10.1016/j.envres.2014.06.029. Epub 2014 Nov 25. PMID: 25460637.

Li H, Fan H, Mao F. A Visualization Approach to Air Pollution Data Exploration—A Case Study of Air Quality Index (PM2.5) in Beijing, China. Atmosphere. 2016; 7(3):35. https://doi.org/10.3390/atmos7030035

References

Libraries

Michael Bostock, D3.js (Version 4.13.0) [Source Code]. https://d3js.org/

Wes McKinney, Pandas (Version 0.25.3) [Source Code]. https://pandas.pydata.org/

Cynthia Brewer, ColorBrewer (Version 2.0) [Source Code]. https://colorbrewer2.org

Datasets

European Environment Agency, Annual AQ statistics https://discomap.eea.europa.eu

ISPRA, Portale Pesticidi https://sinacloud.isprambiente.it

ISTAT, Ambiente urbano https://www.istat.it/it/archivio/281184

Published Work

All the work described in this presentation is publicly available at the following github repository https://github.com/parwal-lp/EnVis

The live version of the presented system is available at https://parwal-lp.github.io/EnVis/

A report documenting this work can be found at https://github.com/parwal-lp/EnVis/blob/main/docs/report.pdf

Thank you for your attention!