Air Quality Analysis in DC

Naomi and Marjorie

2023-12-13

Use Case

The "Air Quality in DC" Shiny app is designed to address the complex challenge of understanding and analyzing air quality data for the District of Columbia. The primary users of this app are environmental researchers, policy makers, public health officials, and urban planners who are interested in the patterns and sources of air pollution in the city.

In urban areas like DC, where traffic congestion, industrial activities, and other pollution sources are prevalent, it is crucial to have a tool that can simplify the analysis of air quality data. This app provides a user-friendly interface for exploring data related to various air pollutants, such as sulfur dioxide, nitrogen dioxide, carbon monoxide, particulate matter, and others that are commonly found in urban environments.

Users can utilize the app to:

- Perform Exploratory Data Analysis (EDA) to uncover patterns, trends, and outliers in the air quality data.
- Generate visualizations such as histograms, density plots, box plots, and scatter plots to visualize the distribution and relationships between different pollutants.
- Conduct hypothesis testing to statistically compare pollution levels across different times of the year or between different monitoring sites.
- Run simple and multiple linear regression analyses to investigate the impact of various factors on air quality indicators.

The app is particularly useful for assessing the effectiveness of environmental regulations, understanding seasonal variations in pollutant levels, and planning interventions to improve air quality. By providing insights into the data through interactive visualizations and statistical analyses, the "Air Quality in DC" app serves as a valuable tool for making informed decisions aimed at enhancing the environmental health of the District of Columbia.

Required Packages

- library(shiny) Used for interactive web applications in R primarily utilizing widgets

  - version 1.7.5

- library(tidyverse) A collection of packages that allows users to easily manipulate and tidy data

  - version 2.0.0

- library(magrittr) A package that builds on the natural pipe

  - version 2.0.3

- library(bslib) Custom themes for shiny and rmarkdown

– version 0.5.1

- library(DT) imports the DataTables library from JavaScript

  – version 0.30

- library(scales) Scale function for visualizations

  – version 1.2.1

- library(thematic) ggplot2 themes

  – version 0.1.3

- library(broom) Converts statistical analysis objects to tidy outputs that can be easily interpreted

  – version 1.0.5

Data Source

Air Quality History ** https://opendata.dc.gov/datasets/515f2c8ff3534302a34bf47f6902ac0d__77/explore

- data published December of 2021 and most recently updated August of 2022
- 20 Variables
- 555,859 observations
- seasonal and diurnal pattern of pollutants in DC
- from Open Data DC where data is DOEE certified and comes from 5 official monitoring sites in DC

Workflow

There are four tabs for users to choose based on the analysis they would like to see.

On the simple linear regression lab, users select one X variable and one Y variable. Options include every pollutant observed, as well as temporal variables for season, year, date, or site name. Then, the user clicks the run regression button to run the model with the given variables and see the plot visualization between the two variables. Then, the user can choose to show the summary statistics of both variables and the summary of the linear model.

On the multiple regression analysis, users can run a multiple regression and make predictions based on these models. For the model, users chose year(s), month(s), response variable, and predictor variable(s) to run the model on. The app runs this regression and shows the user the output, with the option to apply a logistic transformation on the x variable to better interpret the data. Then users can make predictions based on the predictor variables previously selected and a given confidence interval.

On the hypothesis testing tab, users first select a variable for testing and input their chosen null hypothesis. Then when clicking run test, the output of a one sample t-test is shown. For the two sample t-test, users can select two variables to run a t test on and then the two sample t test output is shown.

On the Plotting tab, users select which kind of plot they would like show: histogram, density plot, box plot, or scatter plot. Then they select the x and y variable they would like to visualize. Clicking the generate plot button runs the selected model and shows the visualization to the user.

Scenario Overview for Multiple Regression and Statistical analysis

Imagine you are an environmental scientist studying the impact of various factors on air quality in Washington D.C. You are particularly interested in understanding how different seasons and locations within the city influence the levels of pollutants like nitrogen dioxide and ozone.

Step 1: Launching the app After launching the Shiny app starts the user on the simple linear regression tab which is intuitively designed to guide you through selecting variables, and displaying results.

Step 2: Setting Up Your Analysis Let's say you want to conduct a multiple linear regression analysis to explore the relationship between nitrogen dioxide levels and temperature across different locations in the city.

Step 3: Performing Multiple Regression Analysis For a more comprehensive analysis, you might want to consider multiple factors simultaneously, such as wind speed, humidity, and time of year.

Select multiple predictors in the "Multiple Regression Analysis" tab. Run the analysis and interpret the results, considering how each factor contributes to the model and their individual significance.

Step 4: Hypothesis Testing Suppose you hypothesize that the average level of ozone is higher in summer than in winter. You can test this hypothesis using the app:

Navigate to the "Hypothesis Testing" tab. Select ozone as your variable and specify summer and winter months for comparison. The app will perform a two-sample t-test and provide you with the test statistics and p-value.

Step 7: Making Data-Driven Decisions With the analysis complete, you can now use these insights to make informed decisions or recommendations. For instance, if you find a significant relationship between temperature and nitrogen dioxide levels, you could advocate for policies aimed at reducing emissions during warmer months.

Summary This vignette has walked you through a typical workflow within the app, from selecting variables and running analyses to interpreting results. With these tools at your disposal, you can effectively analyze air quality data and contribute valuable insights to environmental studies and policy making in Washington D.C.

Literature Review

Ambient Air Quality Trends Report 1996-2019. Department of Energy and Environment for the District of Columbia, 2020.

- Emissions and pollution have improved over the last 20 years
- Main sources of pollution include transportation emissions and pollution from other southern states
- DC is in compliance with National Ambient Air Quality Standard for particulate matter, carbon monoxide, sulfur dioxide, nitrogen dioxide, and lead
- Ozone levels are above the standard

Calma, Emilia. "The Geography of Environmental Toxins in the District of Columbia." D.C. Policy Center, 15 October 2020, https://www.dcpolicycenter.org/publications/environmental-toxins/. Accessed 30 October 2023.

- The most common air pollutants are ozone and particulate matter
- Ward 1 has the highest ozone levels and Ward 7 has the highest particulate matter levels
- DC's child and health asthma rates are higher than the national average with Black children having asthma rates around four times higher than white children.
- Air quality is worsened by urban heat islands

Castillo, Maria, et al. "Estimating Intra-Urban Inequities in PM2.5-Attributable Health Impacts: A Case Study for Washington, DC." GeoHealth, vol. 5, no. 11, 2021. AGU, https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GH000431. Accessed 30 October 2023.

- Although air quality has improved over the last few decades, health impacts from PM2.5 (particulate matter) are very unequal in DC
- Southeast region is most at risk for illnesses caused by air pollution including asthma, cardiovascular issues, stroke
- Look at five neighborhood level factors to analyze the inequtiy: education, unemployment, income, race and ethnicity, life expectancy at birth