IJC 445 Data Visualisation 2025-2026

Visualising Air Quality: A Data Visualisation Analysis of Pollution Patterns in Sheffield and Hull

Student Registration Number: 250124882

Word count: 1564

January 9, 2025

University of Sheffield

Composite Visualisation

Figure 1: Seasonal Average Bar Plot

Figure 2: Daily Average Scatter Plot

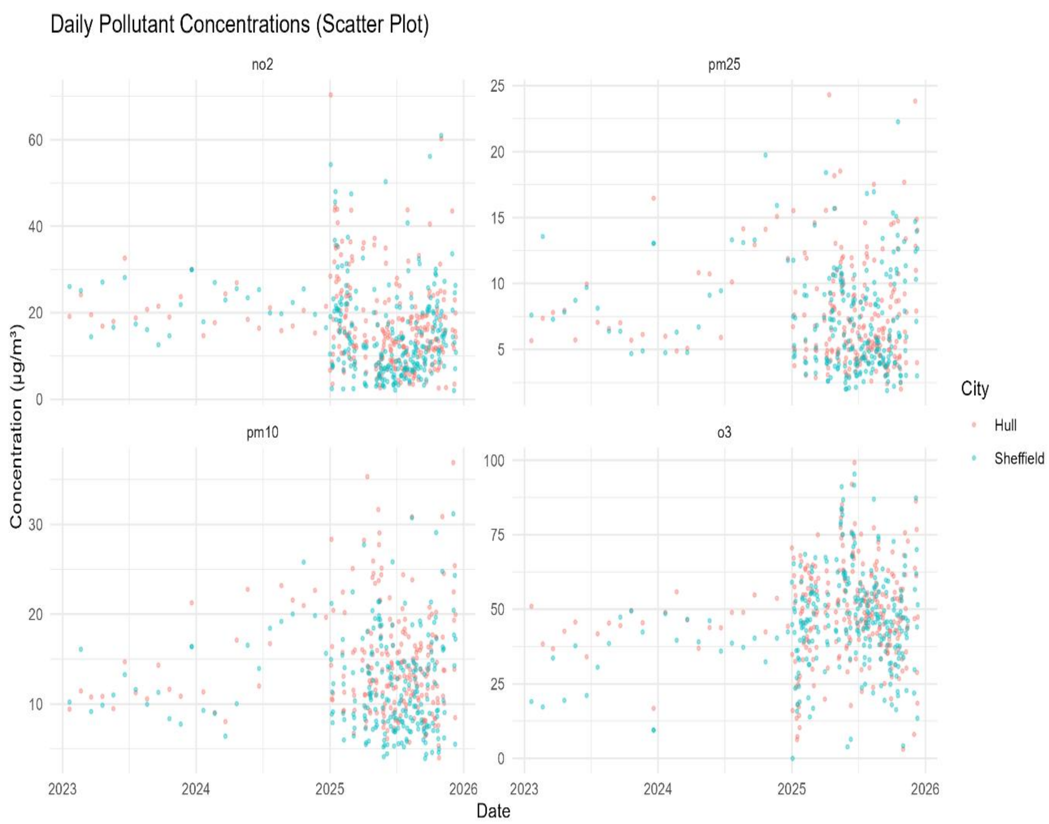


Figure 3: LOESS Trend

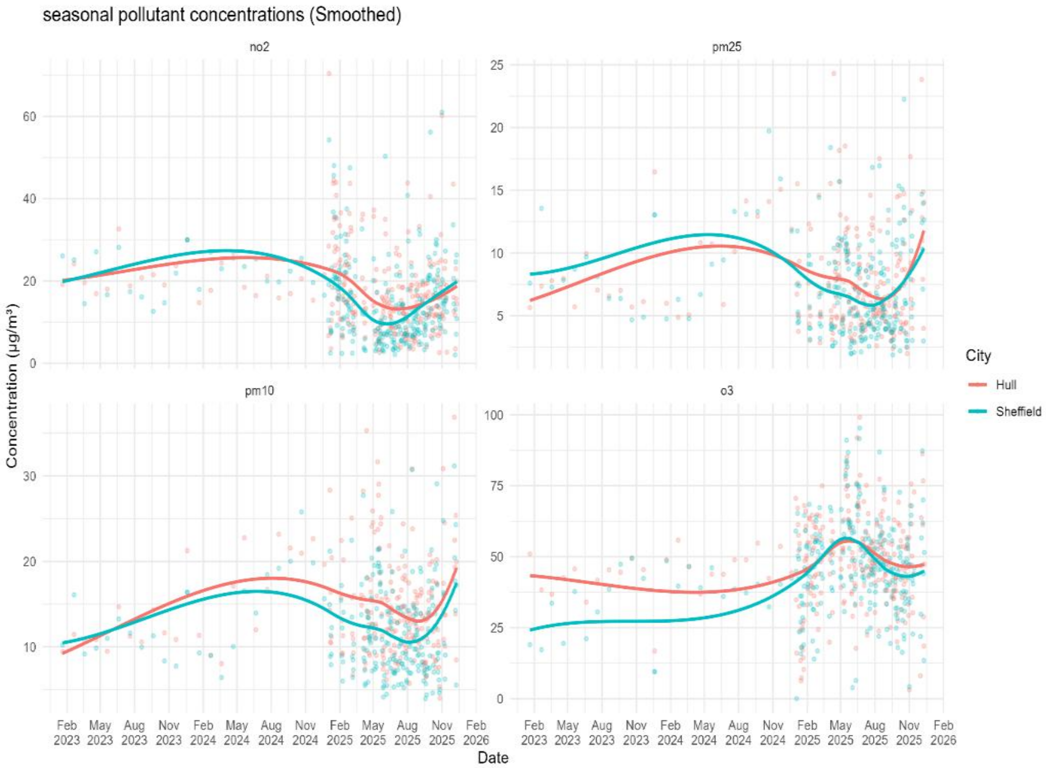
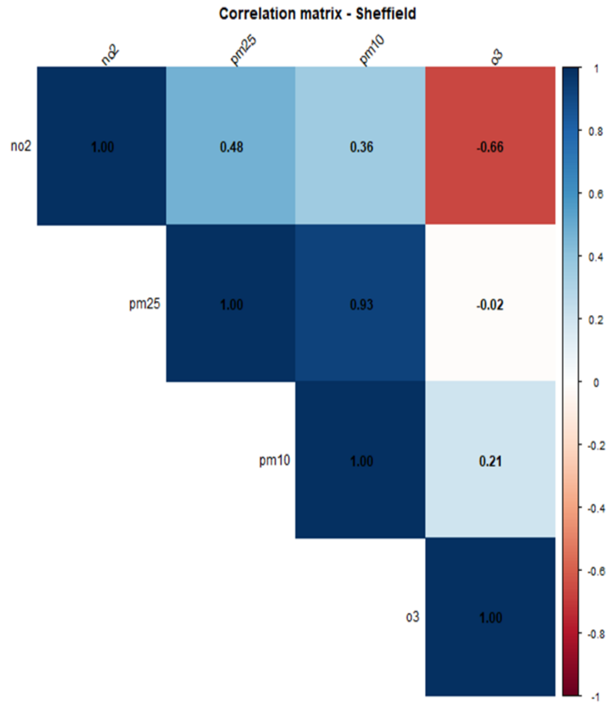


Figure 4: Correlation Matrix Sheffield



Chapter 1

Knowledge Building

This chapter illustrates how a composite visualisation provides insight into air quality dynamics by integrating multiple temporal and relational perspectives.

Instead of displaying charts separately, the visualisation incorporates seasonal averages, daily variability, smoothed trends, and pollutant relationships to facilitate a more comprehensive understanding of pollution patterns in Sheffield and Hull.

1.1 Topic and Question

Air pollution is still a major environmental and public health problem in the United Kingdom. It leads to early deaths and long-term respiratory and heart diseases (WHO, 2021; DEFRA, 2023). In cities, pollutants like nitrogen dioxide (NO_2), ozone (O_3), and particulate matter ($PM_{2.5}$ and PM_{10}) are especially concerning because they are emitted by traffic, industry, and atmospheric changes. While overall emissions have decreased in recent decades, significant differences remain among cities due to local sources, geography, and weather (Monks et al., 2015).

Sheffield and Hull provide a practical comparative case study. Sheffield is an inland, densely populated city with road traffic and combustion-related emissions (Sheffield City Council, 2010). Hull, by contrast, is a coastal port city influenced by maritime activity, industrial sources, and greater atmospheric dispersion (Hull City Council, 2023). These contrasting settings make the two cities well-suited to examining how pollutant behaviour varies over time.

The composite visualisation addresses the following question:

1. How do seasonal fluctuations, daily variability, long-term trends, and pollutant interactions differ between Sheffield and Hull from 2023 to 2025, and what do these combined patterns reveal about underlying emission sources and atmospheric processes?

This question is important because air quality dynamics operate across different time periods. Seasonal averages may mask short-term exposure risks, while daily observations can obscure longer-term trends. Integrating multiple temporal perspectives aligns with calls in the literature for multi-scale environmental analysis.

1.2 Composite Visualisation

The composite visualisation comprises four panels, each offering a distinct analytical perspective.

Figure 1 presents the seasonal average concentrations of NO_2, $PM_{2.5}$, PM_{10}, and O_3. Both cities show clear seasonal patterns. NO_2 and particulate matter levels are higher in autumn and winter, likely due to reduced atmospheric dispersion and increased combustion emissions. O_3 levels are highest in summer because of photochemical formation (Monks et al., 2015). Sheffield has higher NO_2 and $PM_{2.5}$, which matches its traffic-heavy emissions, while Hull has higher PM_{10} and O_3, likely due to its industry and coastal location.

Figure 2 displays daily pollutant concentrations using scatter plots. The panel reveals considerable short-term variability and frequent pollution spikes, especially in 2025, when data coverage is extensive. While both cities exhibit broadly similar patterns of variability, the large number of observations and associated noise hinder the identification of clear trends. These findings underscore the limitations of interpreting raw daily data.

Figure 3 addresses this issue by applying Locally Estimated Scatterplot Smoothing (LOESS) to the daily time series. The smoothed lines make it easier to see seasonal cycles and trends that are hard to spot in the scatter plots. Levels of NO_2 and particulate matter mostly decline in 2023 and 2024, then rise in winter. In contrast, O_3 is higher in the warmer months. These patterns match what is known about atmospheric chemistry, such as ozone titration and the buildup of traffic-related pollutants in winter (Carslaw & Beevers, 2005). The cities still show differences, with Sheffield regularly having higher levels of traffic-related pollutants.

Figure 4 presents a correlation matrix summarising the relationship between pollutants in Sheffield. Strong positive correlations between $PM_{2.5}$ and PM_{10} suggest shared emission sources or influences. The pronounced negative correlation between NO_2 and O_3 reflects ozone titration.

Taken together, the four figures show that the differences between Sheffield and Hull are not caused by just one pollutant or timescale. Instead, these differences result from the combined effects of seasonal cycles, short-term changes, long-term trends, and atmospheric chemistry. The composite visualisation helps turn complex environmental data into clear, understandable information, demonstrating its usefulness for both analysis and communication.

Chapter 2

Theoretical Frameworks

This chapter evaluates the composite visualisation using the ASSERT framework and the Grammar of Graphics, illustrating how theoretical principles informed its design and supported knowledge building.

2.1 Applying the ASSERT Framework

1. Ask a Question

The visual design was guided by a clear analytical question:

- How do seasonal patterns, daily variability, long-term trends, and pollutant interactions differ between Sheffield and Hull from 2023 to 2025?

Addressing this question required a thoughtful approach, since no single chart could compare pollutants, cities, and time periods.

2. Search for Information

Hourly air-quality measurements for NO_2, $PM_{2.5}$, PM_{10}, and O_3 were obtained from the OpenAQ database and aggregated to daily, monthly and seasonal values. Domain literature on urban air pollution and atmospheric chemistry informed variable selection and interpretation (Monks et al., 2015; Carslaw & Beevers, 2005).

3. Structure the Data

Data were organised across multiple perspectives to support different analyses. Seasonal averages provided a high-level comparison, daily values captured short-term variability, and smoothed series facilitated trend interpretation. Pollutants were faceted to enable comparison while avoiding overplotting.

4. Encode the Data

Pollutant concentration is represented by position along a common scale, while colour consistently encodes the city across all figures. Temporal progression is mapped to the x-axis in the time-series panels, enabling intuitive interpretation and efficient cross-panel comparison.

5. Refine the Data

Transparency was applied to daily scatter points to reduce overplotting, while LOESS was added as a secondary layer to reveal underlying structure without obscuring raw data. Axis scales, labels, and colour mappings were kept consistent throughout.

6. Tell the Story

The figures are arranged to guide the viewer from overview to detail and finally to synthesis. Seasonal summaries establish context, daily scatter plots reveal variability, smoothed trends clarify structure, and correlation matrices explain pollutant interactions. Together, they form a coherent visual narrative.

2.2 Grammar of Graphics Analysis

The composite visualisation follows the Grammar of Graphics framework by mapping data to visual elements. In the temporal figures, time is mapped to the x-axis and pollutant concentration to the y-axis, with the city encoded by colour.

Different geometric objects support different analytical tasks: bars facilitate categorical comparison, points reveal variability, lines emphasise trends, and tiles encode relational strength through colour intensity.

Faceting by pollutant produces helps reduce visual clutter while preserving shared scales, supporting accurate comparison. Statistical transformation is applied through LOESS smoothing, showing how computation and visuals combine to reveal patterns not visible in the raw, noisy data. Through this integration, the visualisation becomes a structured analytical system rather than a collection of charts.

Chapter 3

Accessibility

Accessibility ensures that visualisations can be interpreted without unnecessary perceptual or cognitive barriers (Ware, 2013; Cairo, 2016). This chapter evaluates Figure 2, the scatterplot.

Figure 2 is easy to interpret because it shows raw daily observations as points and uses distinct colours for both cities. Still, there are so many observations that the points overlap, which makes it hard for non-experts to spot trends. The time axis spans several years, making it even harder to interpret without clear breaks or smoothing.

Relying only on colour can make it harder for individuals with colour vision deficiencies to interpret the data. While using consistent colours helps, adding features such as different shapes can make the information more accessible (W3C, 2018). Figure 2 focuses on showing accurate data, but it also requires viewers to put in extra effort to understand it. This highlights why it is important to include Figures 3 and 4 as well.

Chapter 4

Visualisation Choice

4.1 Justification of the Selected Visualisation

Figure 3 shows a faceted LOESS line chart that highlights both medium and long-term trends while still showing the daily data points. Line charts help us see changes over time clearly. Faceting avoids overplotting and makes it easier to compare different pollutants. The LOESS method smooths out short-term noise, making gradual changes easier to spot without obscuring the underlying data. Using distinct colours for both cities also helps make the chart easier to read.

4.2 Alternatives

Calendar heatmaps highlight seasonality and reduce overplotting, but are less effective for comparing multiple pollutants or cities and make long-term trends harder to track. Ridgeline plots show distributional shifts but reduce temporal precision and may be challenging for non-experts to interpret. In contrast, the faceted LOESS line chart provides a more precise and intuitive representation that better supports the objectives.

Chapter 5

Implications and Improvements

5.1 Ethical Implications

Figure 1: Using seasonal averages helps with comparison, but it can hide short-term trends that might have negative effects. As a result, aggregation may cause viewers to underestimate the severity of exposure. Treating cities as uniform also hides differences within cities related to traffic and socioeconomic factors (Sicard et al., 2020).

Additionally, seasonal changes may be misinterpreted as evidence of policy effectiveness without accounting for external influences. Ethical visualisation requires transparency about these limitations (Cairo, 2016).

5.2 Future Improvements

Future improvements could strengthen the analysis by showing variability alongside seasonal averages, for example, through interquartile ranges or counts of valid observations. This would help readers understand how representative the seasonal summaries are.

Including meteorological variables such as temperature or wind speed would help distinguish between changes driven by emissions and those shaped by weather. Interactive features such as tooltips or basic filters could help users explore specific periods or values without cluttering the main visuals (Few, 2012).

Finally, accessibility could be improved by using colour palettes suitable for colour-vision deficiencies, adding concise annotations, and increasing text size to keep the visualisation interpretable for a wider audience (W3C, 2018).

Reference

Carslaw, D. C., & Beevers, S. D. (2005). Estimations of road vehicle primary NO_2 exhaust emission fractions using monitoring data in London. *Atmospheric Environment, 39*(1), 167–177. https://doi.org/10.1016/j.atmosenv.2004.08.053

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. New Riders.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*(387), 531–554. https://doi.org/10.1126/science.229.4716.828

Department for Environment, Food & Rural Affairs (DEFRA). (2023). *Air quality statistics in the UK*. UK Government. https://www.gov.uk/government/statistics/air-quality-statistics

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Analytics Press.

Hull City Council. (2023). Air quality annual status report (ASR).

Monks, P. S., Archibald, A. T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., Mills, G. E., Stevenson, D. S., Tarasova, O., Thouret, V., von Schneidemesser, E., Sommariva, R., Wild, O., & Williams, M. L. (2015). Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics, 15*(15), 8889–8973. https://doi.org/10.5194/acp-15-8889-2015

OpenAQ. (2024). OpenAQ platform: Open air quality data. https://openaq.org

Sheffield City Council. (2010). Air quality in Sheffield. https://www.sheffield.gov.uk/pollution-nuisance/air-quality

Sicard, P., De Marco, A., Agathokleous, E., Feng, Z., Xu, X., Paoletti, E., & Calatayud, V. (2020). Amplified ozone pollution in cities during the COVID-19 lockdown. *Science of the Total Environment, 735*, 139542. https://doi.org/10.1016/j.scitotenv.2020.139542

Ware, C. (2013). *Information visualisation: Perception for design* (3rd ed.). Morgan Kaufmann.

World Health Organisation (WHO). (2021). *WHO global air quality guidelines: Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulphur dioxide and carbon monoxide*. https://www.who.int/publications/i/item/9789240034228

World Wide Web Consortium (W3C). (2025). *Web Content Accessibility Guidelines (WCAG) 2.1*. https://www.w3.org/TR/WCAG21/

Appendix

FIG 7.1:Seasonal Average Bar Plot

```r
#plot
p_seasonal_month_2025 <- monthly_seasonal_2025 %>%
  ggplot(aes(x = season, y = season_mean, fill = city)) +
  geom_col(position = "dodge") +
  facet_wrap(~ pollutant, scales = "free_y") +
  labs(
    title = "Seasonal average pollutant concentrations 2025 (from monthly data)",
    x = "Season",
    y = "Mean concentration (µg/m³)",
    fill = "City"
  ) +
  theme_minimal()

p_seasonal_month_2025

ggsave("C:/Users/Tanur/Documents/Sheffield MSC Data Science/INTRODUCTION TO DATA SCIENCE IJC 437/assessment/r studio/Assessment Intro to Data Science/plots/seasonal_means_by_city_2025.png",
       p_seasonal, width = 10, height = 6, dpi = 300)
```

FIG 7.2:Scatter Plot

```r
# Scatter plot only
p_seasonal_scatter <- ggplot(seasonal_long,
                              aes(x = date, y = value, colour = city)) +
  geom_point(alpha = 0.4, size = 0.7) +      # scatter plot
  facet_wrap(~ pollutant, scales = "free_y") +
  labs(
    title = "Daily Pollutant Concentrations (Scatter Plot)",
    x = "Date",
    y = "Concentration (µg/m³)",
    colour = "City"
  ) +
  theme_minimal()

p_seasonal_scatter

# Save the scatter plot
ggsave("C:/Users/Tanur/Documents/Sheffield MSC Data Science/INTRODUCTION TO DATA SCIENCE IJC 437/assessment/r studio/Assessment Intro to Data Science/plots/scatter_only.png",
       p_seasonal_scatter, width = 10, height = 6, dpi = 300)
```

FIG 7.3:LOESS Plot

```r
#----------------------------------------------
#Section 7
# Time-Series with LOESS (for EDA)

# ---- 1. DAILY TIME SERIES + LOESS ----

p_daily_loess <- ggplot(daily_long,
                        aes(x = date, y = value, colour = city)) +
  geom_point(alpha = 0.2, size = 0.4) +
  geom_smooth(method = "loess", se = FALSE) +
  facet_wrap(~ pollutant, scales = "free_y") +
  labs(
    title = "Daily pollutant concentrations (LOESS-smoothed)",
    x = "Date",
    y = "Concentration (µg/m³)",
    colour = "City"
  ) +
  theme_minimal()

p_daily_loess

ggsave("C:/Users/Tanur/Documents/Sheffield MSC Data Science/INTRODUCTION TO DATA SCIENCE IJC 437/assessment/r studio/Assessment Intro to Data Science/plots/daily_LOESS_sheffield_vs_hull.png",
       p_daily_loess, width = 10, height = 6, dpi = 300)
```

FIG 7.4: Correlation Matrix Plot

```r
# ---- 7. CORRELATION MATRIX ----

plot_city_cor_matrix <- function(df, city_name) {
  sub <- df %>%
    filter(city == city_name) %>%
    select(no2, pm25, pm10, o3)   # order as you like

  cor_mat <- cor(sub, use = "pairwise.complete.obs")
  print(cor_mat)

  corrplot(
    cor_mat,
    method = "color",
    type = "upper",
    addCoef.col = "black",
    tl.col = "black",
    tl.srt = 45,
    title = paste("Correlation matrix -", city_name),
    mar = c(0, 0, 2, 0)
  )

  invisible(cor_mat)
}
```

FIG 7.5: Correlation Matrix Plot (ctd)

```r
# Sheffield
png("C:/Users/Tanur/Documents/Sheffield MSC Data Science/INTRODUCTION TO DATA SCIENCE IJC 437/assessment/r studio/
    Assessment Intro to Data Science/plots/correlation/corr_sheffield.png", width = 800, height = 600)
cor_sheffield <- plot_city_cor_matrix(daily, "Sheffield")
dev.off()

# Hull
png("C:/Users/Tanur/Documents/Sheffield MSC Data Science/
    INTRODUCTION TO DATA SCIENCE IJC 437/assessment/r studio/Assessment Intro to Data Science/plots/correlation/corr_hull.png", width = 800, height = 600)
cor_hull <- plot_city_cor_matrix(daily, "Hull")
dev.off()

# Save numeric matrices
write_csv(as.data.frame(cor_sheffield), "C:/Users/Tanur/Documents/Sheffield MSC Data Science/INTRODUCTION TO DATA SCIENCE IJC 437/assessment/r studio/Assessment
write_csv(as.data.frame(cor_hull), "C:/Users/Tanur/Documents/Sheffield MSC Data Science/INTRODUCTION TO DATA SCIENCE IJC 437/assessment/r studio/Assessment Intro
```

GITHUB

As part of the project workflow, all R scripts, datasets, and visualisation figures were uploaded to a GitHub repository. The repository was organised for easy navigation, with separate sections for the IJC437 and IJC445 project pages, and a code directory containing well-commented R scripts. Clear instructions for running the analysis were also included.

GitHub repository: https://github.com/tanure1999/Air-Quality-Analysis-Sheffield-Hull-2023-2025

Figure 7.6: GitHub Repo Page