assignment_1

Stepan Kropachev

2025-01-02

The task

Consider the data set pollutants_22.csv.

Using suitable graphical instruments and computing confidence intervals, check if the three types of pollutants seem to have different distributions with respect:

(a) to the location of the station (city center/suburbs); and
(b) with respect to the year (2000/2020).

Write a short report with the results and with your comments.

Location Analysis

We start from calculating the confidence intervals for **PM10**, **Ozone** and **NO2** for the city center and suburbs using t-test.

```
## PM10 = [ -14.29852 , 13.56372 ]
## Ozone = [ -11.82851 , 5.3112 ]
## NO2 = [ -6.921138 , -0.3747374 ]
```

The confidence intervals show that **PM10** levels do not significantly differ between city center and suburbs [-14.30, 13.56] g/m3, as the interval contains zero.

Similarly, **Ozone** shows no significant difference between locations [-11.83, 5.31] g/m3.

However, **NO2** exhibits significantly higher concentrations in central locations, with a difference of [-6.92, -0.37] g/m3.

Let's lookup to the boxplots of the pollutions by locations.

| PM10 by Location | Ozone by Location | NO2 by Location |
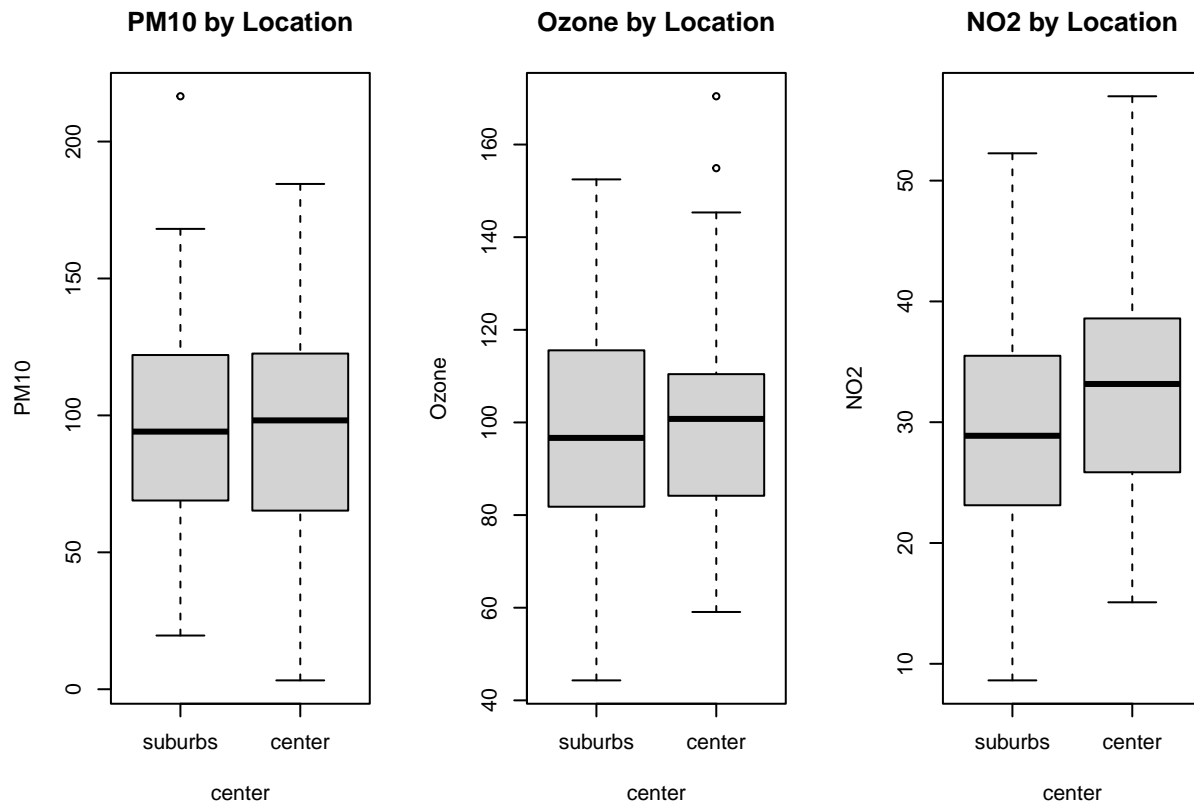
Boxplots for the **PM 10** and the **Ozone** levels show no significant differences in interquartile spreads and median variance between suburbs and the center.

We have an outlier in the **PM10** level in suburbs, and 2 outliers in the **Ozone** levels in the city center. The level of **NO2** significantly differs in suburbs and center, as we also revealed this with the t-test.

Temporal Analysis

Let's proceed with the temporal analysis with the t-test.

```
## PM10 = [ -18.14113 , 9.892734 ]
## Ozone = [ -10.56837 , 6.203189 ]
## NO2 = [ -5.264109 , 1.240917 ]
```

Comparing the years 2000 and 2020, we observe no significant changes in any pollutant:

- PM10: [-18.14, 9.89] g/m3
- Ozone: [-10.57, 6.20] g/m3
- NO2: [-5.26, 1.24] g/m3

All confidence intervals contain zero, indicating no statistically significant changes over this period.

Building the plots for visual analysis.
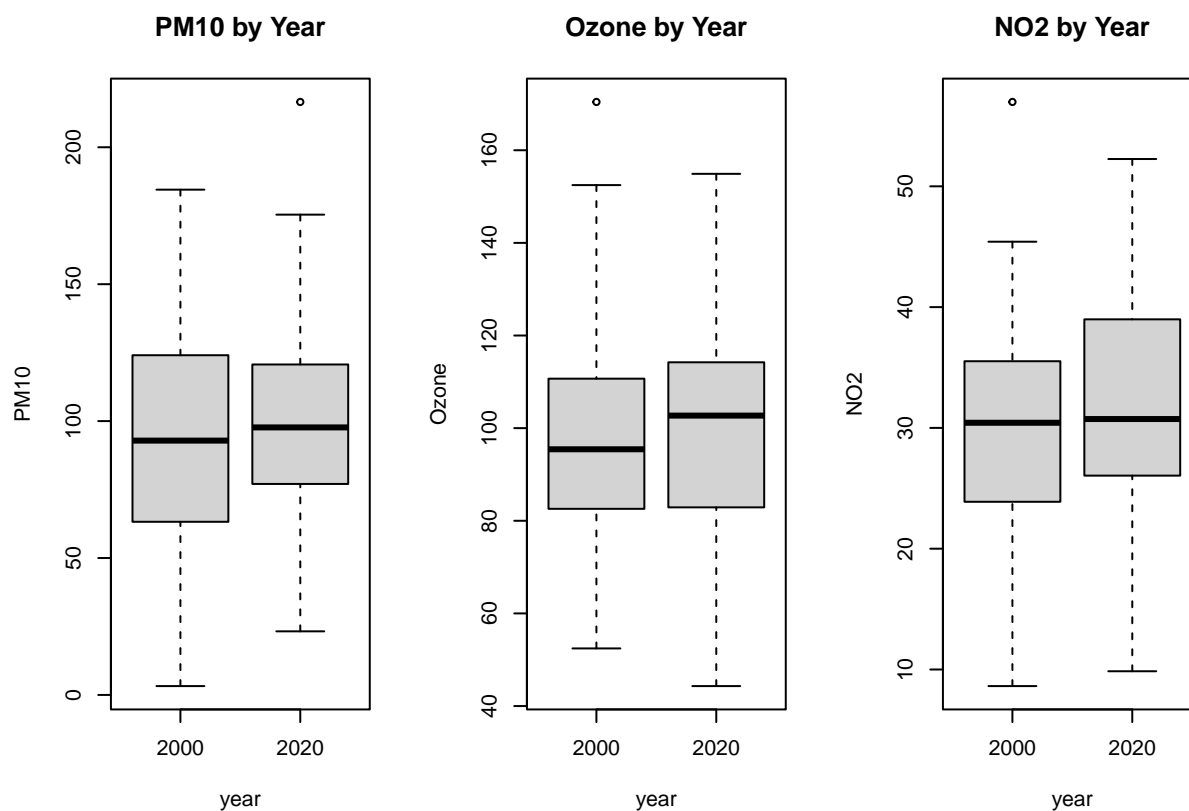
The boxplots reveal relatively similar median levels across years for all pollutants, with comparable spread. Also **PM10** shows some high disperse in 2000. **NO2** displays slightly higher variability in 2020, while **Ozone** distributions remain fairly consistent between the two time periods.

These findings suggest that while location has some impact on pollutant concentrations (particularly for **NO2** in the city center), the temporal evolution between 2000 and 2020 shows no clear statistical evidence of change in pollution levels.