Combining Air Pollution Data with Government Parties

Available Data

I am combining historic data on air pollution in German states (concentration of particulates PM10 in µg/m³) with state governments election data.

Air Pollution

Using the yearly reports on air pollution by the Umweltbundesamt of Germany, the PM10 concentration data was fetched as CSV files from the Umweltbundesamt of Germany website. The files had to be downloaded individually per year by selecting from a drop down menu. Due to this procedure rather difficult to automate and given the small number of downloads I downloaded the files manually.
Each CSV file contained the

- yearly average of PM10 in µg/m³,
- number of days with concentrations above 50 µg/m³,
- number of days with concentrations above 50 µg/m³ (cleaned for usage of road salt etc.)

for every measuring station in Germany for every year from 2002 to 2020. A measuring station is described with the state, a station code, a station name, the kind of the surrounding of the station (e.g. urban area, sub-urban area, ...) and kind of station (e.g. industry, traffic, ...).

Election and Government Data

The election data was a two-fold operation.

The historic election results for each state in Germany (down to every vote for each party) could be downloaded from the regional statistics database of the nation and state governments, Statistische Ämter des Bundes und der Länder. This required a few clicks per state file and could have been automated, but given the small number of files to download, it was quicker to download them manually.
The data was available as CSV files among others, but not designed to be loaded as such. It had no proper header and contained non-CSV-conform metadata at the top and bottom. Between, was CSV-parseable data. Each file contained the election results of all state elections within one German state as far back as the 1990s, down to the individual district result, with votes for the major parties CDU/CSU, SPD, Grüne, FDP, Linke and others.

These files alone do not provide information on which parties eventually make up the government of the respective state. Unfortunately, I could not find such a database for free. Hence, I had to retrieve the data from Wikipedia.
Wikipedia offers in-depth articles for every German state election. Unfortunately, their structure is very different. Some provide tables with the possible coalitions, highlighting the realized one, some state the coalition in continuous text in the first paragraph of the article, some in a dedicted section somewhere else, some only provide graphics. It is impossible to quickly and easily parse this information from the articles. This is why I had to manually read through many articles and write out the parties making up the respective governments into a JSON file, structured by state and year.

Evaluation

The mentioned data files were processed into a final JSON file, that had the following structure:

```json
{
    "state": {
        "year": {
            "pollution": {
                "year_average": 1234.0,  // cumulative average values of all
stations (this year and state)
                "year_average_counter": 60,  // number of aggregated stations for
`year_average`
                "days_above_limit": 1029.0,
                "days_above_limit_counter": 60,
                "days_above_limit_cleaned": 0.0,
                "days_above_limit_cleaned_counter": 60
            },
            "election": {
                "government": [
                    "cdu_csu",
                    "fdp"
                ],
                "date": "DD.MM.YYYY",
                "eligible_voters": 12345678,
                "valid_votes": 12345000,
                "cdu_csu": 1200000,
                "spd": 1200000,
                "grüne": 1200000,
                "fdp": 1200000,
                "linke": 1200000,
                "afd": 0,
                "other": 1200000
            }
        },
        "2003": {
            "pollution": {
                "year_average": 1459.0,
                ...
            },
            "election": {}  // empty as there was no election that year
        },
        ...
    },
    ...
}
```

Visualizations

I implemented three different chart types. The following presents them shortly. If opened in a browser, the chart files provide further information on mouse hover. You find the charts in `output/`.

Absolute Concentration

As found in output/absolute/.*

The first chart shows the absolute yearly average concentration of PM10 for one state next to the nation-wide average. The color depicts the government leader party.
This example shows the chart of Schleswig-Holstein.
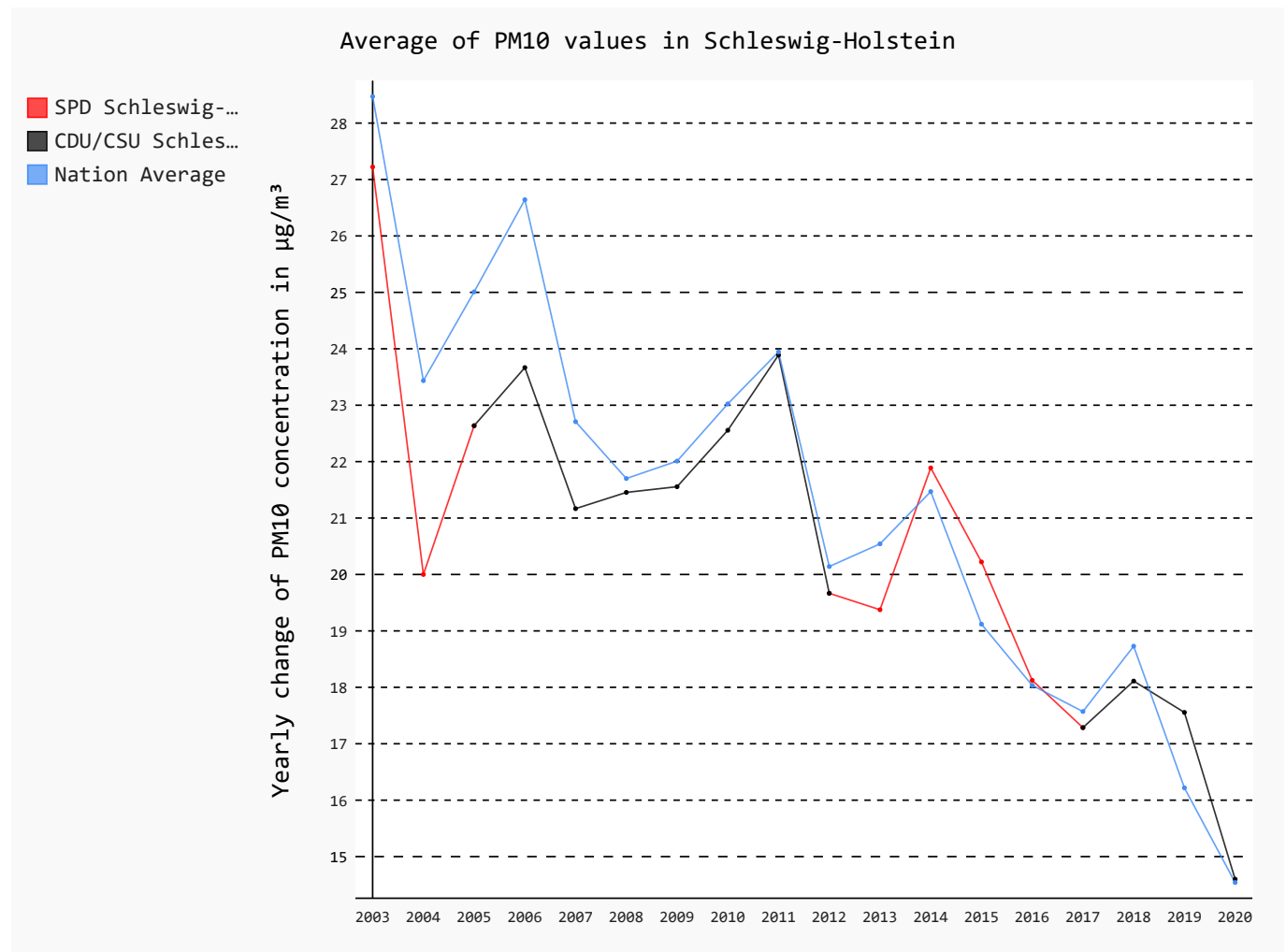
This chart type allows us to get a quick overview of the general situation and validate our processed data.

Relative Concentration Changes

As found in output/changes/.*

The second chart type displys the yearly change of the PM10 concentration relative to the value of the previous year. The color again depicts the government leader party.
Here the chart of Baden-Württemberg.

Average change of PM10 values in Baden-Württemberg
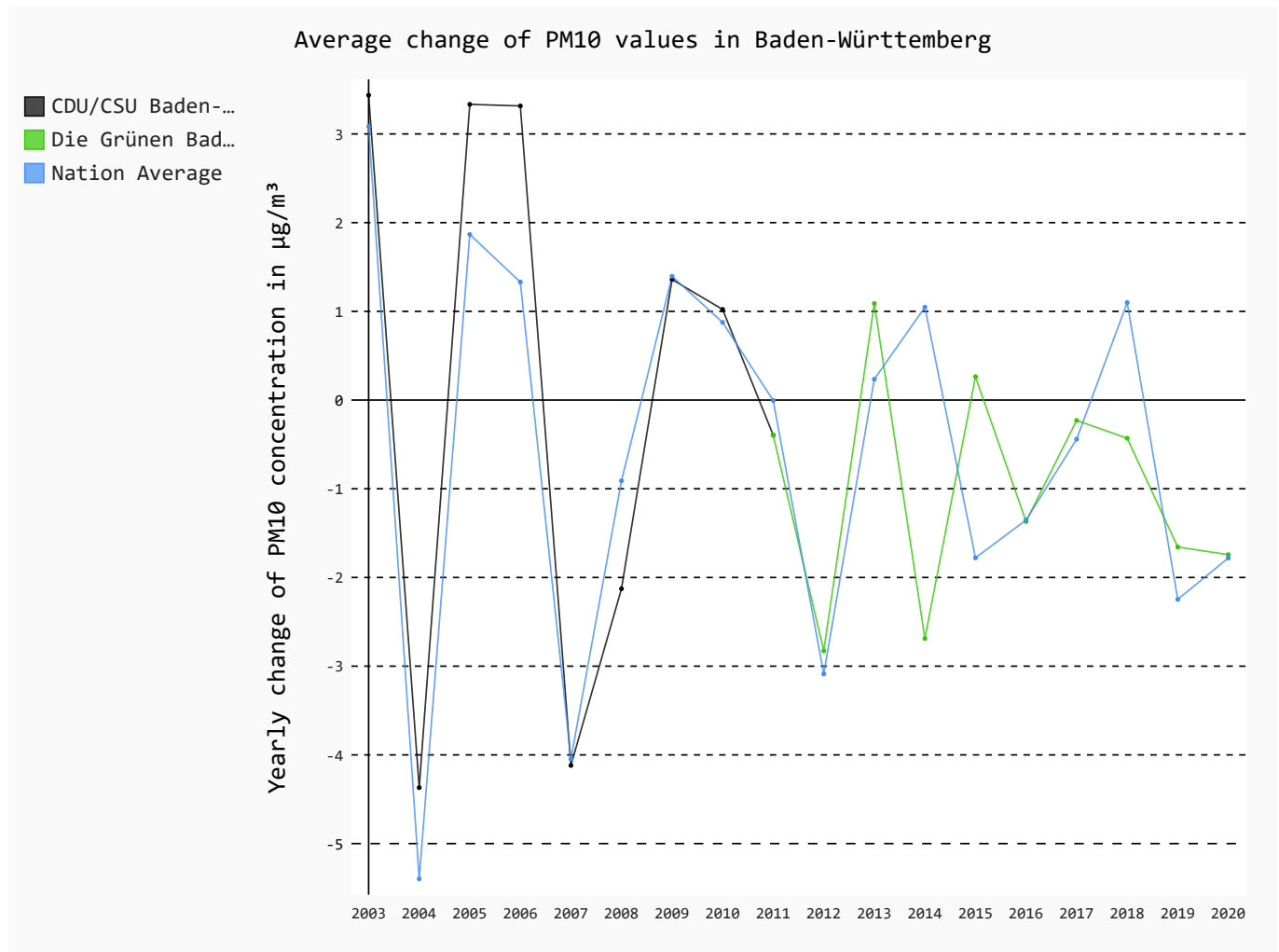
While interesting to see, this chart type is prone to convey little actual information, if not confusing and/or misleading information (the viewer may have the impression that the CDU and Grüne may have a similar, yet differently steady, trend. The absolute chart shows that only after the governing of Grüne, the trend is goes clearly and steady downwards).

Our aim is to efficiently evaluate the performance of parties concerning air pollution. Therefore, the last chart type may be helpful.
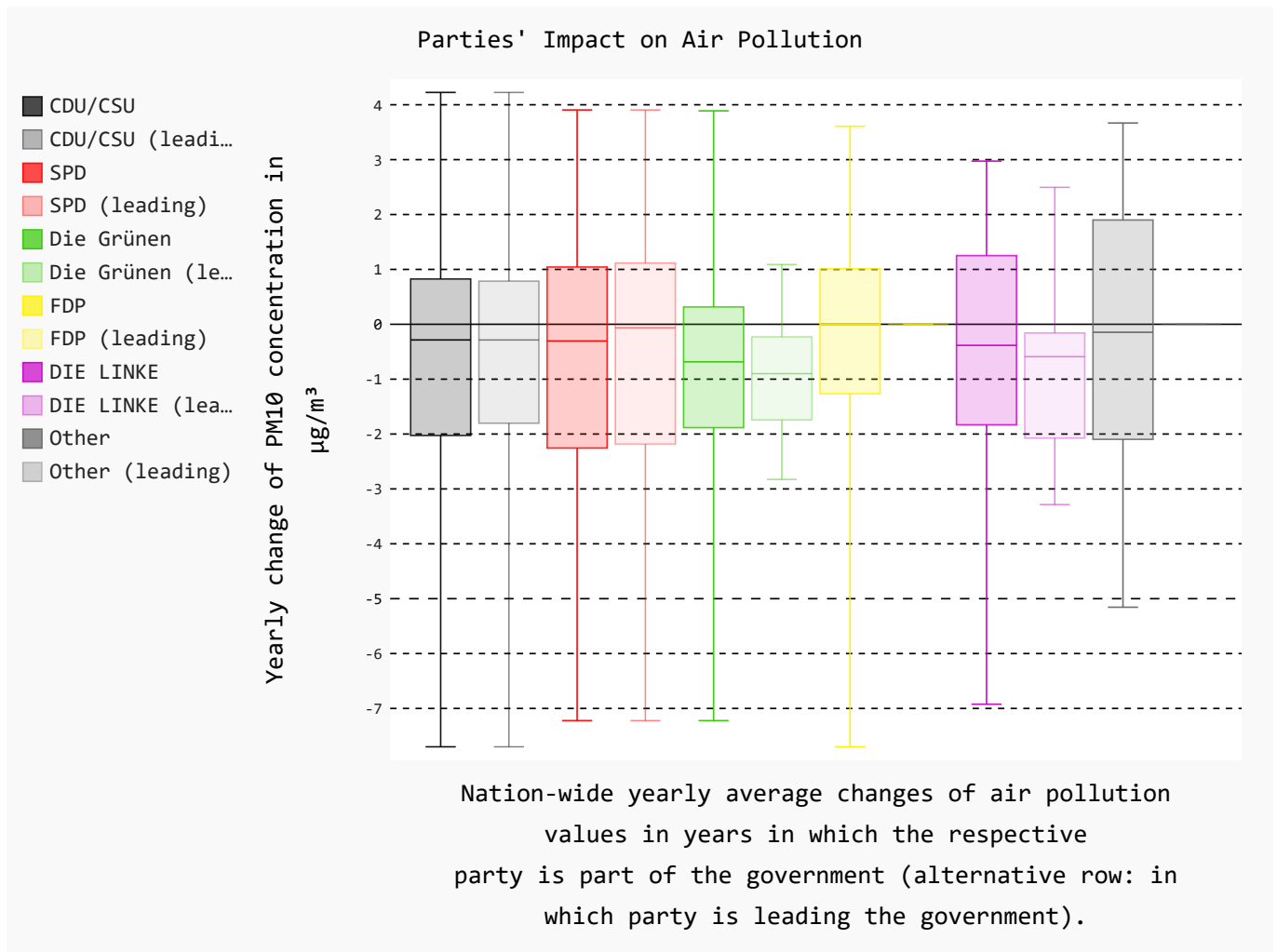
Impact on Air Pollution

As found in output/misc/pollution_impact_all_parties.svg.

You may want to open this chart type in a browser, as it can show more details on mouse hover.

This last chart type shows the impact of a party on air pollution, measured in all states and over all available years.

Each box depicts the yearly changes of PM10 concentration in the air (relative to the year before) among all states during the years in which the respective party was part of the state government (strong color) or biggest government party/government leader (light color).

Parties' Impact on Air Pollution

Nation-wide yearly average changes of air pollution values in years in which the respective party is part of the government (alternative row: in which party is leading the government).

This gives us a good impression of the performance of each party. While rather close, the differences are still clearly visible. It is worth noting that the number of data points varies significantly. While the CDU has governed as leader for 194 years in total, Linke only governed 7 years as leader, moreover solely in a single state.