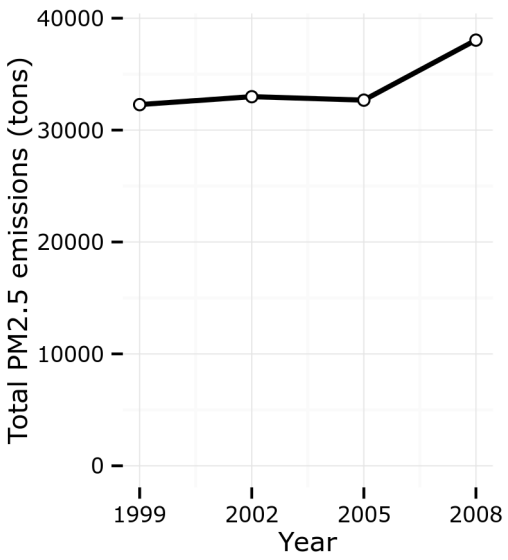
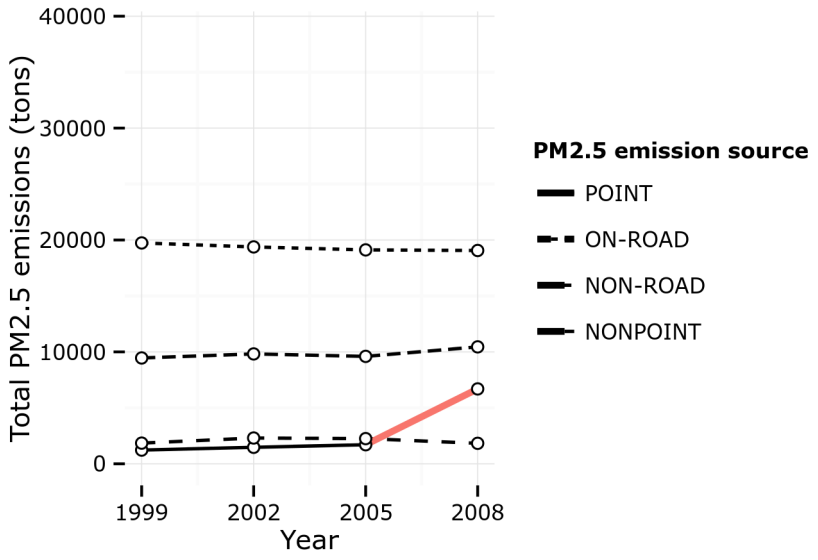
$PM_{2.5}$ emissions through ten years in the USA

Fine particulate matter ($PM_{2.5}$) is an ambient air pollutant with diameter of 2.5 micrometers or less, for which there is a strong evidence that is harmful to human health. This work explores the change in $PM_{2.5}$ emissions through ten years in the USA. As a result, those counties are highlighted which experienced suprisingly high increase of emissions between 1999–2008.

Besides gaining experience with the visualisation package ggplot2 the goal of this work is to explore how the $PM_{2.5}$ emissions have changed in the USA through 1999–2008.



We can observe an increase in the total emissions between 2005 and 2008.



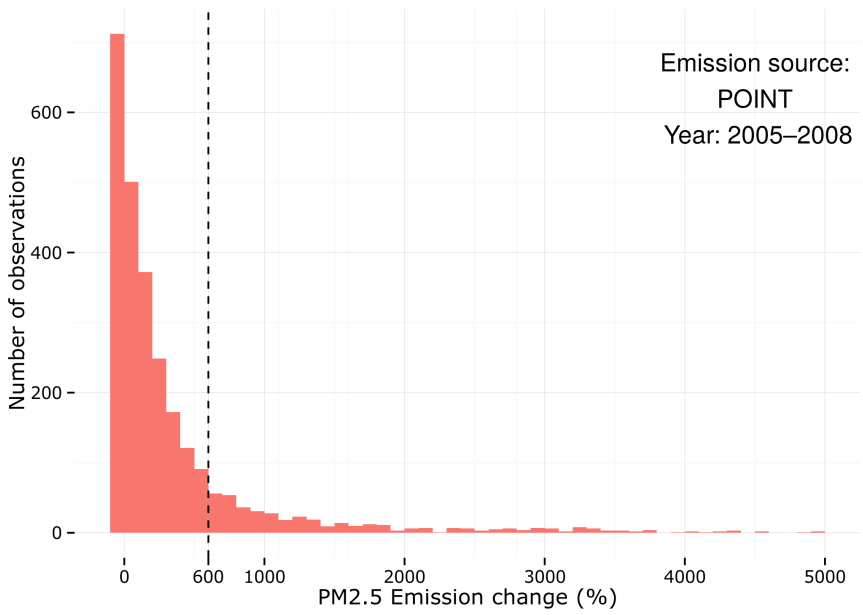
It is clear that the increased emission between 2005 and 2008 originated mainly from POINT type of sources.

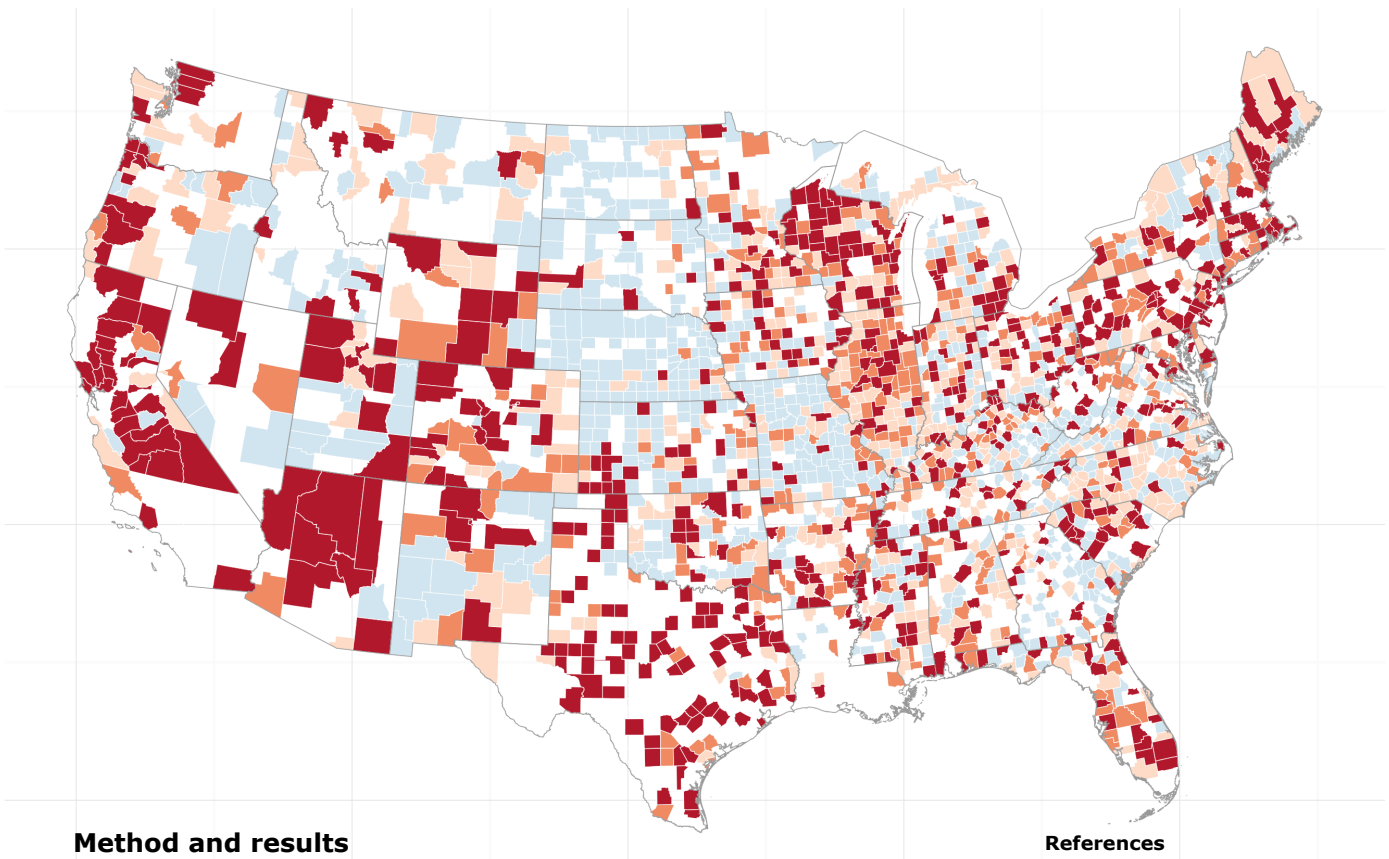
Data

The data originates from the National Emissions Inventory (NEI) which is a comprehensive and detailed estimate of air emissions of air pollutants from all air emissions sources. The NEI is prepared every three years and based primarily upon emission estimates and emission model inputs provided by State, Local, and Tribal air agencies.

POINT: Emissions estimates for sources that are individually inventoried and usually located at a fixed, stationary location. For example industrial facilities, electric power plants, gas stations, etc.

The histogram on the right shows the percentage change in emissions from POINT sources between 2005 and 2008. Because the histogram drops at 600% and common sense dictates that a change in the thousand percent range is unlikely, the values above 600% were excluded from the further analysis.





Map of the USA identifying those counties that experienced a surprisingly high change in emissions from POINT sources between 2005 and 2008. These cases would require closer analysis to discover the cause of the high amount of change. Furthermore, the amount of change was categorized using common sense, but with more information on this topic, the categories could look very different, also with regard of what is considered an outlier.

Method and results

As mentioned above, the visualisations were produced using R with the package ggplot2. For the data preparation and map production besides custom functions two additional packages were required, "raster" and "rgdal".

Personally, the biggest challenge was the data preparation process due to the high number of outliers in the data set. On the other hand, the most rewarding part was the process of exploratory analysis, while a high number of graphics were generated which are not presented here. Through this process I could experience and understand the meaning and place of exploratory data analysis.

References

Chang, W., 2013. R graphics cookbook: practical recipes for visualizing data, 1st ed. ed. O'Reilly, Beijing.

Dorman, M., 2014. Learning R for geospatial analysis: leverage the power of R to elegantly manage crucial geospatial analysis tasks. Packt Publishing, Birmingham.

Particulates, 2015. Wikipedia Free Encycl. Retrieved 03.05.2015.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer New York.

Data: http://www.epa.gov/ttn/chief/net/2008inventory.html, retrieved March 2015

Code: https://github.com/balazsdukai/PM2.5-emissions_visualisation.git

Author: Balázs Dukai
Module supervisor: Prof. Dr. Susanne Bleisch
Module: MSE Information Visualisation
Date: 06.05.2015