The DoMiNO (Data Mining & Neonatal Outcomes) Project
(CIHR/NSERC Collaborative Health Research Program 2013-2018)
https://sites.google.com/a/ualberta.ca/domino/

This exploratory project investigates, through spatial data mining, the colocation of adverse birth outcomes and environmental variables (industrial emissions, SES) in Canada, with a focus on Alberta.

The study is interdisciplinary and involves a team of computing scientists, environmental health specialists, epidemiologist, geographers, perinatologists, obstetricians, representatives of government and non-government organizations and various students at all levels.

The study objectives are to:

- Develop a novel spatial data mining approach to:
  - Identify non-trivial colocations of environmental and maternal variables with adverse birth outcomes.
  - Identify potential patterns of associations among variables for new hypotheses.
- Assess the newly developed Data Mining methods for broader application.
- Evaluate the utility of existing databases used in the study.
- Train a new generation of interdisciplinary high-quality personnel in environmental health.
- Evaluate the interdisciplinary strategies employed in this research, to contribute to the field of Knowledge Translation science.

After five years of working together, the following outcomes emerged.

- Four Data Mining approaches and a visualization tool that allowed the identification of statistical significant **non-trivial rules** associating groups of chemicals with diverse adverse birth outcomes.
- Those groups of chemicals include previously identified pollutants (e.g., particulate matter, CO, NO_2), but also some others that were suspected, but not yet proven, to be associated with adverse birth outcomes (e.g., individual volatile organic compounds).
- Complementary results include the observation of a higher prevalence of adverse birth outcomes in urban settings, where maternal and SES variables do not explain adequately the risk associated with adverse birth outcomes.
- Validation of the association rules obtained by data mining identified rules that relate with increased odds ratios for adverse birth outcomes after taking into consideration maternal and SES variables.
- We trained three master students; four Ph.D. students are close to their defense, one already obtained his degree, and three postdoctoral fellows had share experiences in the team.

We would like to share our findings to determine if they resonate with others and develop a study where we could explore the associations with more precision. To do that, we could use our visualization tool called ViZAR.

Principal Investigators:

Osmar Zaiane (zaiane@ualberta.ca) and Alvaro Osornio Vargas (osornio@ualberta.ca)

August 2017