# Prediction of ADRs using NMF and Weighted NMFs

Arshad Ahamed Shajahan, Sai Shasank Dandu[1]

Ollscoil na Gaillimhe
University of Galway

## Background and Objectives

Adverse drug reactions (ADRs) are a common problem in clinical and pharmacovigilance research and can lead to serious patient harm and biased conclusions if modelled poorly. Sparse, noisy, and highly imbalanced drug ADR data often cause standard machine learning methods to perform no better than naïve frequency-based approaches, unless appropriate low-rank and kernel-based methods are used with clear assumptions. [^1]: https://doi.org/10.1093/bioadv/vbae009

### Objectives

1. Explain the ADR profile prediction problem and its statistical challenges in imbalanced, noisy health data.
2. Explore low-rank and kernel-based methods (e.g. NMF+kernel regression and alternatives) for ADR prediction.
3. Apply these methods to updated drug–ADR datasets with additional molecular and interaction features.

## Data Sources and Datasets

The primary source of data are collected from DGIdb 4.0, SIDER 4.1 and PubChem database.

1. Drug-gene interaction pair: Intersection of drugs from DGIdb 4.0 and SIDER 4.1, generated to the binary matrix form.
2. Chemical fingerprints: Data from PubChem database, generated to the binary matrix form.
3. Drug & side effects: Drug along with it's side effects are extracted from SIDER 4.1 database.

The syntax in this poster template and the `posterdown` package uses the same workflow approach as the R Markdown you know and love.

You can even use the bibliography the same way: Our data were taken from a cluster-randomised trial [1], available from the Irish Social Science Data Archive.

## Early Results / Descriptive Statistics of Datasets

Usually you want to have a nice table displaying some important results that you have calculated. In `posterdown` this is as easy as using the `kable` table formatting you are probably use to as per typical R Markdown formatting.

You can reference tables like so: Table 1. Some basic summaries of the dataset are below:

Table 1: Table caption.

| Sepal.Length | Sepal.Width | Petal.Length | Petal.Width |
|---|---|---|---|
| 5.1 | 3.5 | 1.4 | 0.2 |
| 4.9 | 3.0 | 1.4 | 0.2 |
| 4.7 | 3.2 | 1.3 | 0.2 |
| 4.6 | 3.1 | 1.5 | 0.2 |
| 5.0 | 3.6 | 1.4 | 0.2 |
| 5.4 | 3.9 | 1.7 | 0.4 |
| 4.6 | 3.4 | 1.4 | 0.3 |
| 5.0 | 3.4 | 1.5 | 0.2 |
| 4.4 | 2.9 | 1.4 | 0.2 |
| 4.9 | 3.1 | 1.5 | 0.1 |

Figure 1, and Figure 2 below show the patterns in our dataset. Make sure that all the details in your plots will be legible when printed (legend text, axis text, and any labels)

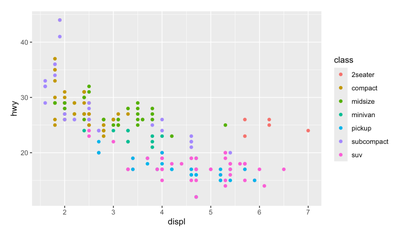


Figure 1: Great figure!

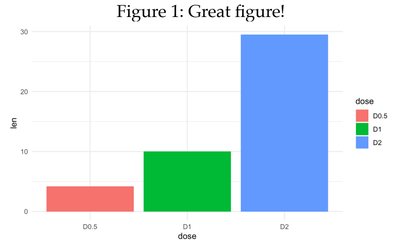


Figure 2: Amazing, right?!

You can even make your plots interactive for the HTML version of the poster. You can use the HTML poster for the presentation session, and the PDF poster will be printed - so be sure the static version looks okay.

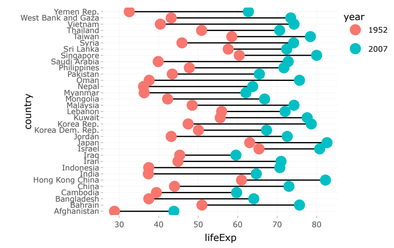


Figure 3: Amazing, right?!

## Next Project Steps

We plan to conduct further analysis using:

- Variable discombobulation [2]
- Expand our minds with explosive machine learning [3].

We will use the `plasticanalysis` package for this.

## GitHub

The code and datasets for this project can be viewed at our GitHub repository here: https://github.com/

## References

1. Murphy et al. 2005 doi: 10.1186/1468-6708-6-11↩
2. Massey et al. 2005 doi: 15.36.413↩
3. Smith et al. 1991 doi: 12.36.486↩