

**H**ome **W**ork

**Jorge 01/04/20**

Cristina Leal  
Gianluca Mazzoni

# Drug databases characterization - part 2

What?	How?	Why?	Goal
Information consistency	Datasets DDI information intersection	Will provide insights of what is the more resourceful database for DDI, which are the most common and shared DDI, and will leave room for finding new interactions reported in very few databases.	Investigate the current status of consistency across databases as well as the degree of standardization of terminology (how many different identifiers are used, how easy are able to be mapped to each other, etc.)
Score	<b>TF (Term-frequency) statistics</b> <b>Frequency score</b> $_{ddi} = \frac{\text{Number of datasets}_{ddi}}{N \text{ total datasets}}$	Commonly used in text mining and information retrieval to see how frequently a term occurs in a document. Similarly, you can apply the same concept to see how 'frequently' a drug interaction is occurs divided by the the total number of databases as a way of normalization.	Score the relevance / frequency of a DDI in the drug database compendia analysed.
Drug ranking	Summary statistics (quartiles, median, mean, sdev)	Interpretability of data	Ranking drugs based on how many interactions they have.
Enrichment	Chi-squared, Fisher test	Interpretability of data	Find over-represented classes/therapeutic group of drugs with DDI information / studied.

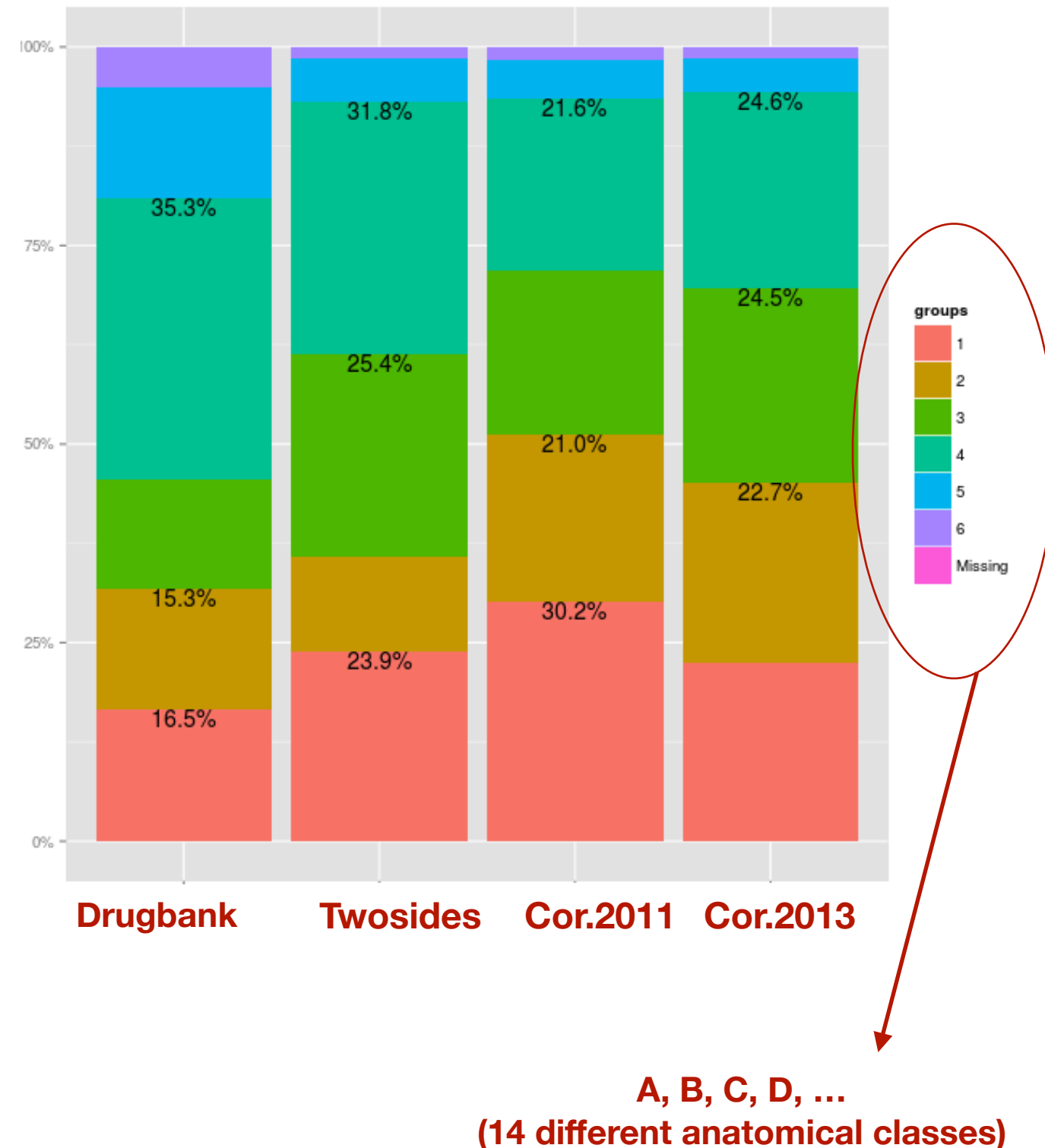
# Visualisation resources and tools:

- **R graph gallery** <https://www.r-graph-gallery.com/>
- **From data to viz** <https://www.data-to-viz.com/index.html#treeanchor>
- **Data viz catalogue** <https://datavizcatalogue.com/index.html>
- **Data exploration and viz** <https://towardsdatascience.com/data-exploration-and-visualization-with-r-ggplot-7f33c10ec1c>
- **Big data visualization techniques** <https://towardsdatascience.com/big-data-information-visualization-techniques-f29150dea190>
- **R for Data Science: Data visualization** <https://r4ds.had.co.nz/data-visualisation.html>
- **Top 50 data viz** <http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html#4.%20Distribution>
- **How and when to use viz** <https://medium.com/@EvanSinar/7-data-visualization-types-you-should-be-using-more-and-how-to-start-4015b5d4adf2>
- **Set intersections**
  - <https://intervene.readthedocs.io/en/latest/introduction.html>
  - <https://github.com/hms-dbmi/UpSetR>
  - <https://github.com/const-ae/ggupset>
  - <http://caleydo.org/tools/upset/>
- **Sankey diagrams**
  - <https://plotly.com/~alishobeiri/1591/plotly-sankey-diagrams/#/>
  - <https://github.com/timelyportfolio/parsetR>
  - <http://sankeymatic.com/>
- **Circos plots** <http://mkweb.bcgsc.ca/tableviewer/settings/>
- **Cytoscape:**
  - [http://manual.cytoscape.org/en/stable/Export\\_Your\\_Data.html](http://manual.cytoscape.org/en/stable/Export_Your_Data.html)
  - Lars Juhl Jensen online workshop: <https://m.youtube.com/watch?v=IH75WJgLeoo&feature=youtu.be>

# Working example 1

## Visualisation of the different Drug interaction classes across datasets

- xaxis: datasets (i.e. Drugbank, TwoSides, etc.)
- yaxis: counts of drugs with reported DDI stacked by anatomical group



## Working example 2

### Sankey diagram

Another cool visualisation that would be very nice to see how information is shared across databases

