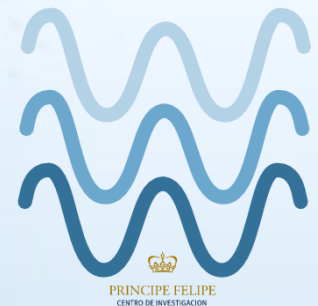


# Visualization of *omics* data

José Francisco Català Senent  
Bioinformatics and Biostatistics Unit, CIPF

16 Oct 2019



# Index

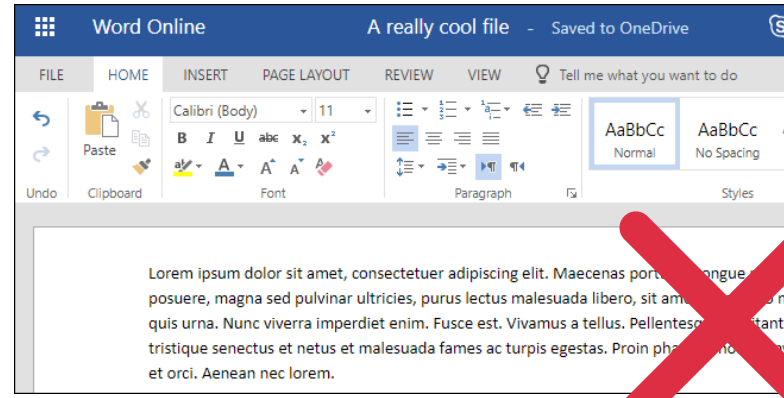
---

- Input files
- Data distribution (BoxPlotR)
- Data matrix visualization (Heatmapper)
- Exercises

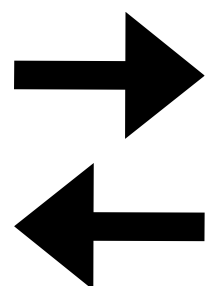
# Input files

- A text file is a file formed by alphanumeric characters, line breaks and tabs.
- Text files can be opened without any special software.

```
>SEQUENCE_1
MTEITAAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKAAKKADRLAAG
LVSVKVSDDFITIAAMRPSYLSYEDLDMTFVENEYKALVAELEKENEERRRLKDPNKPEHK
IPQFASRKQLSDAILKEAEKIKEELKAQKPEKIWDNIIPGKMNSFIADNSQLDSKLTLL
MGQFYVMDDKKTVEQVIAEKEKEFGGKIKIVEFICFEVGEGLKKTEDFAAEVAAQL
>SEQUENCE_2
SATVSEINSETDFVAKNDQFIALTKDTHAHIQSNLSQSVEELHSSTINGVKFEFYLKS
ATIGENLVVRRFATLKAGANGVVNGYIHTNGRVGVVIAAACDSA EVASKSRD...
```



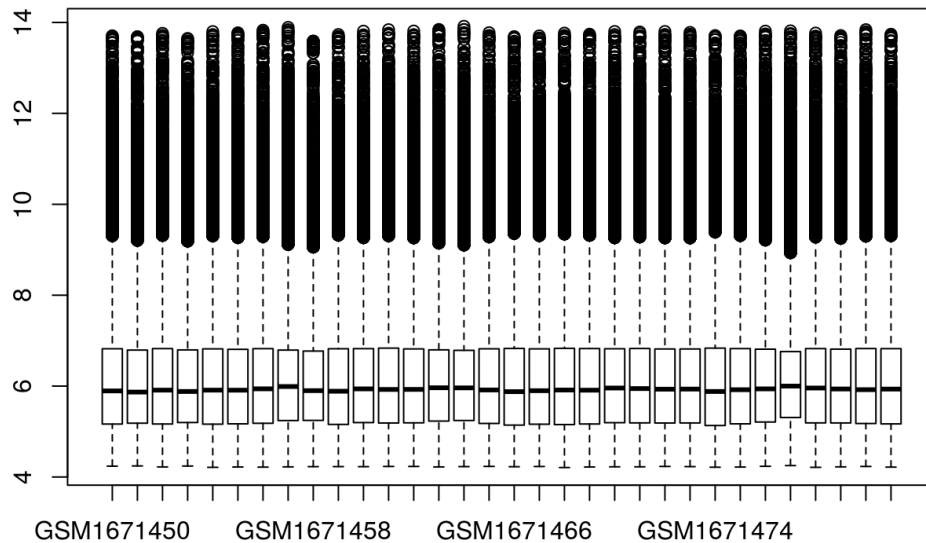
```
SAMPLE.1, SAMPLE.2, SAMPLE.3, SAMPLE.4, SAMPLE.5, SAMPLE.6
129.7, 134.4, 120.8, 115, 105.3, 99.3
664.7, 805.1, 687.6, 723.4, 794.1, 660.5
107.1, 181.7, 122.3, 137.1, 131.9, 142.5
455.6, 1015.7, 554.5, 683.2, 489.2, 580
88.6, 104.6, 92.7, 94.1, 94.8, 89.7
99.3, 133.5, 109.5, 113.6, 112.8, 96.9
292, 660, 369.4, 468.8, 332.6, 347
91.4, 106.7, 90.5, 97, 83.8, 84.8
95, 92.9, 89.5, 96.5, 82.7, 89
1322.8, 2540.1, 1607.9, 1879.9, 3035.8, 3507.4
107.1, 119.6, 94.1, 104.5, 104, 103.6
106.6, 108.7, 98.2, 109.7, 85.3, 81.6
219.1, 395.5, 276.2, 292.5, 199.4, 236.7
629, 1004.9, 674.3, 795.8, 897.8, 937.2
106.1, 119.5, 96.6, 102.5, 90.4, 106
```



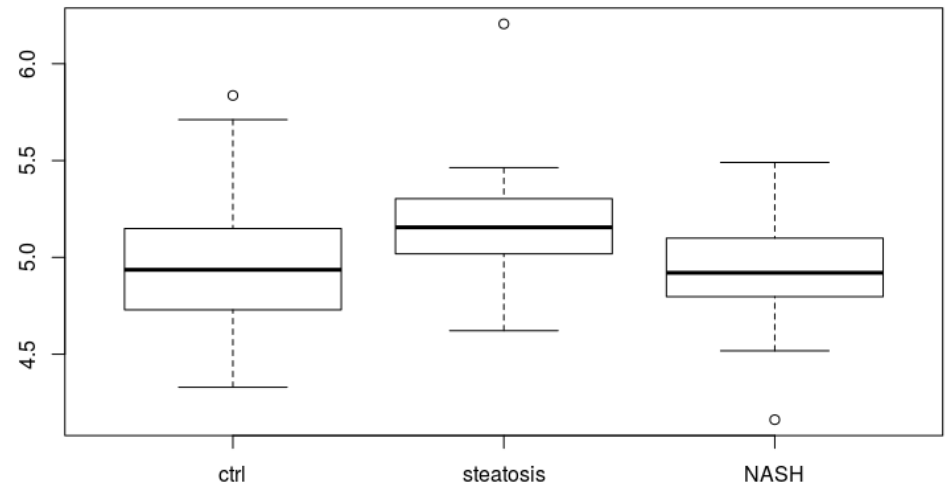
|    | A        | B        | C        | D        | E        | F        |
|----|----------|----------|----------|----------|----------|----------|
| 1  | SAMPLE.1 | SAMPLE.2 | SAMPLE.3 | SAMPLE.4 | SAMPLE.5 | SAMPLE.6 |
| 2  | 129.7    | 134.4    | 120.8    | 115      | 105.3    | 99.3     |
| 3  | 664.7    | 805.1    | 687.6    | 723.4    | 794.1    | 660.5    |
| 4  | 107.1    | 181.7    | 122.3    | 137.1    | 131.9    | 142.5    |
| 5  | 455.6    | 1015.7   | 554.5    | 683.2    | 489.2    | 580      |
| 6  | 88.6     | 104.6    | 92.7     | 94.1     | 94.8     | 89.7     |
| 7  | 99.3     | 133.5    | 109.5    | 113.6    | 112.8    | 96.9     |
| 8  | 292      | 660      | 369.4    | 468.8    | 332.6    | 347      |
| 9  | 91.4     | 106.7    | 90.5     | 97       | 83.8     | 84.8     |
| 10 | 95       | 92.9     | 89.5     | 96.5     | 82.7     | 89       |
| 11 | 1322.8   | 2540.1   | 1607.9   | 1879.9   | 3035.8   | 3507.4   |
| 12 | 107.1    | 119.6    | 94.1     | 104.5    | 104      | 103.6    |
| 13 | 106.6    | 108.7    | 98.2     | 109.7    | 85.3     | 81.6     |
| 14 | 219.1    | 395.5    | 276.2    | 292.5    | 199.4    | 236.7    |
| 15 | 629      | 1004.9   | 674.3    | 795.8    | 897.8    | 937.2    |
| 16 | 106.1    | 119.5    | 96.6     | 102.5    | 90.4     | 106      |
| 17 |          |          |          |          |          |          |

# BoxPlotR

Box plot (or whisker plot) are commonly used to compare datasets.



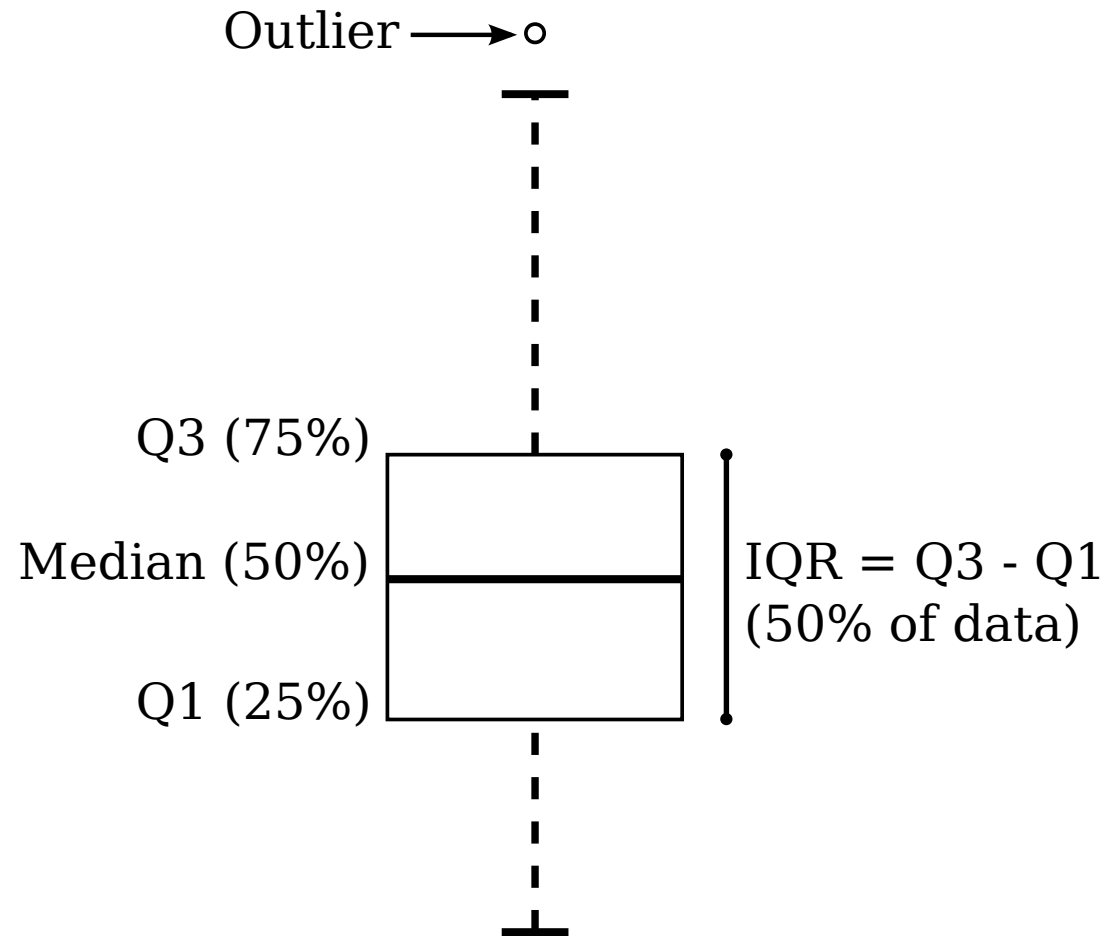
Normalized data from microarray experiment GSE68421



Expression of the gene ISG15 among several group of patients (GSE48452)

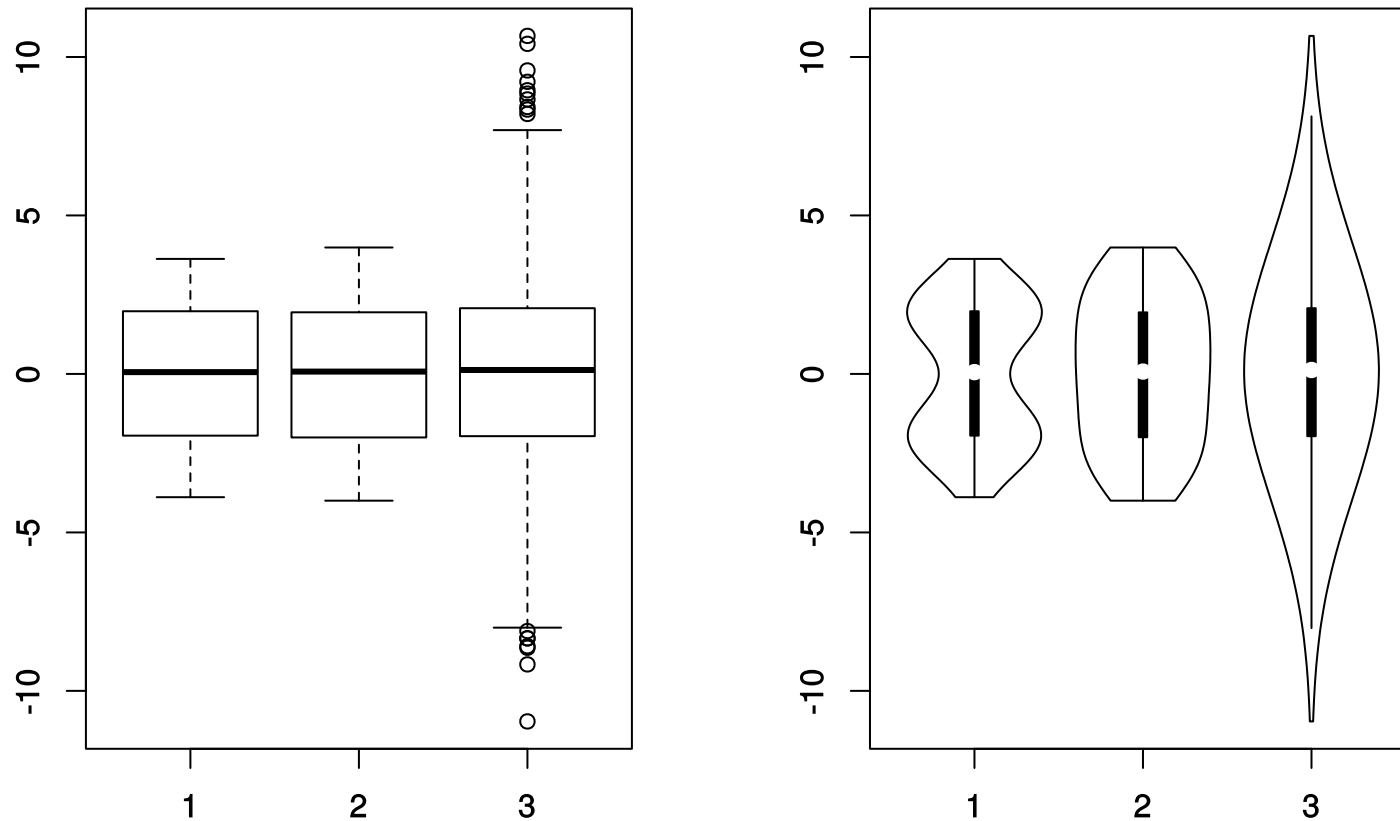
# BoxPlotR

## Understanding Boxplots



# BoxPlotR

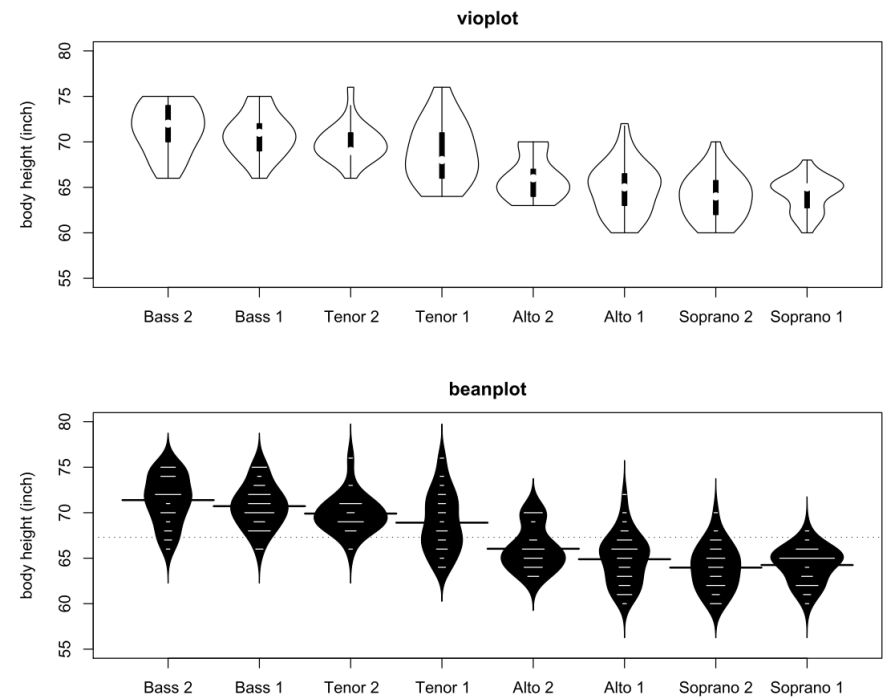
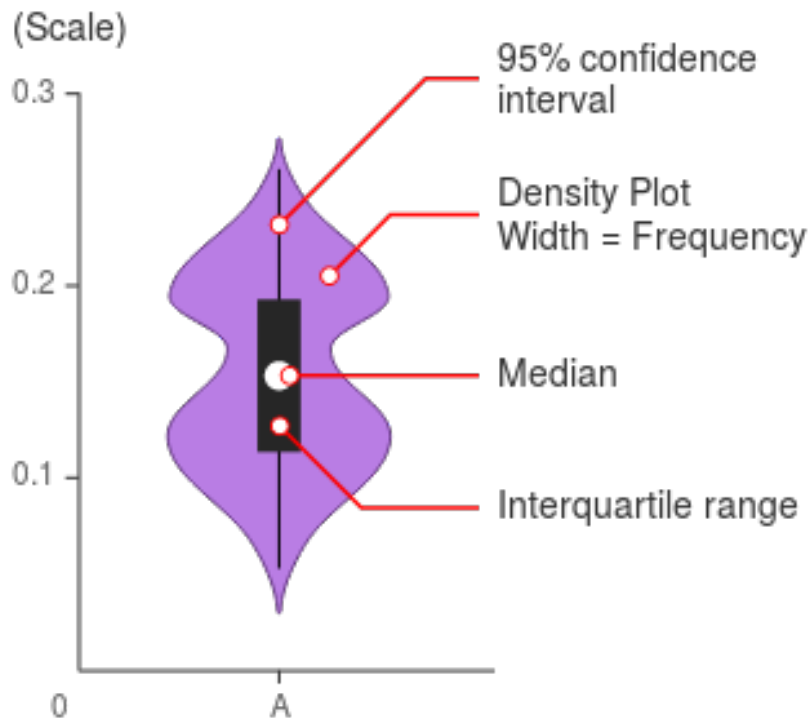
Variants of box plot: violin plot and bean plot



Box plot (left) and violin plot (right) from bimodal (1), uniform (2) and normal (3) distributions

# BoxPlotR

## Variants of box plot: violin plot and bean plot



# BoxPlotR

---

- Accepted file formats:
  - Tab, comma or semicolon separated files
- Data Structure
  - The first row is used to assign column names.
  - The other rows only with numeric values
- It is possible to paste data directly.

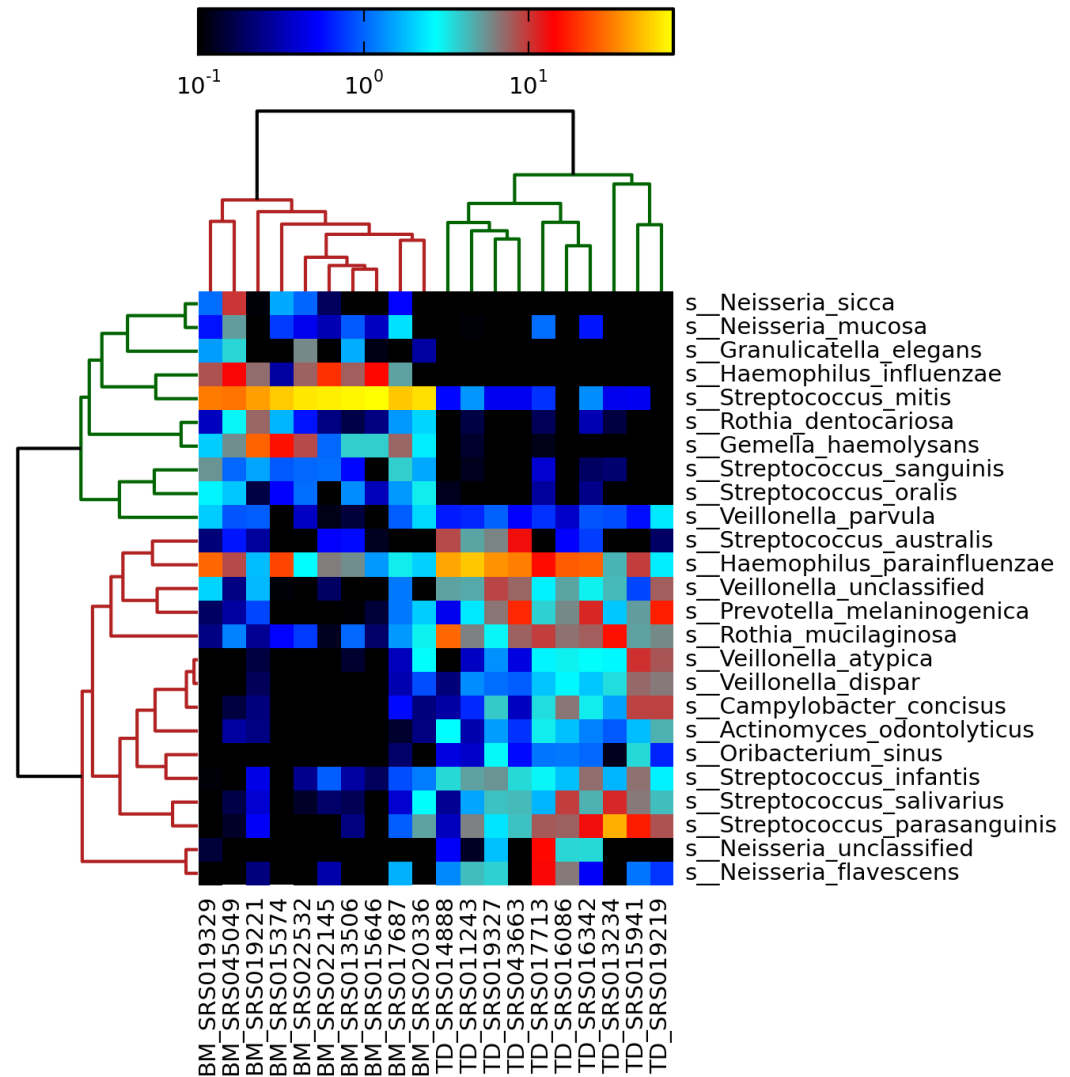
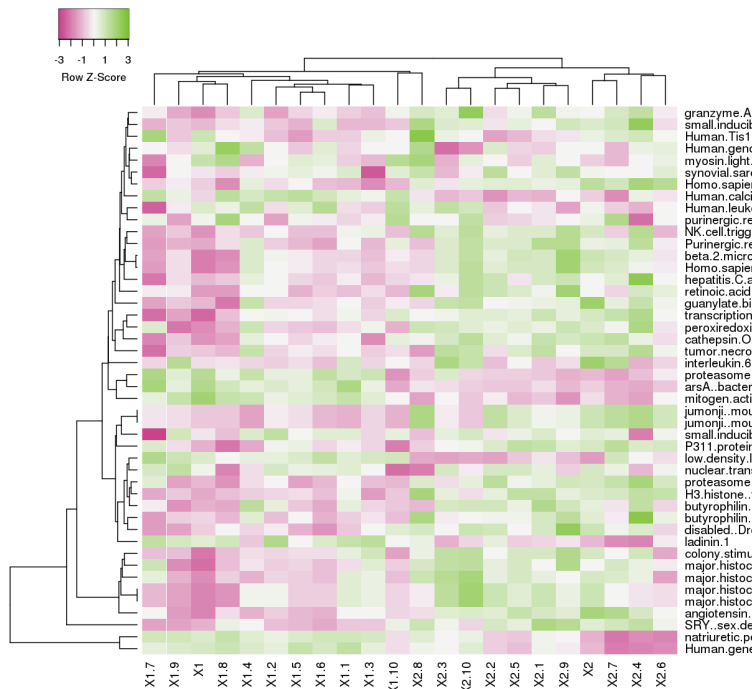
<http://shiny.chemgrid.org/boxplotr/>

<http://bioinfo.cipf.es/BoxPlotR.shiny/>



# Heatmapper

A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colors



# Heatmapper

---

- Accepted file formats:
  - Tab separated files with extensions .txt, .dat, .tsv or .tab.
  - First sheet of a Excel file with extension .xlsx
- Data Structure
  - The first (header) row is used to assign column names.
  - A column labeled UNIQID is required. (or a NAME column)
  - If a column labeled NAME exists (optional), it will be used to assign row names.
  - Any additional columns containing non-numeric data will be ignored.
  - Data values can be positive or negative numbers.
- Input data can have up to 2500 rows and 300 columns

<http://heatmapper.ca/>

# Exercises

---

