UpSetR: A novel R package for visualizing intersecting sets and their attributes
Harvard Medical School, Department of Biomedical Informatics
Jake Conway & Nils Gehlenborg, PhD

Abstract
Understanding the relationships and interactions between data is imperative in bioinformatics. With copious amounts of biological data being generated today, there is high demand for tools that can visualize large numbers of sets and their intersections. Here we present UpSetR, a novel R package that visualizes large numbers of sets and intersections via a matrix-based technique. Along with visualizing data on a set-based level, UpSetR allows users to explore and visualize their data on an element-based level through the implementation of queries and attribute plots. Through its seamless integration with ggplot2 and the ability to apply virtually any query to the data, UpSetR is an extremely powerful tool for data exploration and producing publication quality visualizations.

Approach
• UpSetR was implemented using ggplot2, and arranged using a grid layout (100 x 100)
• Built in intersection and element queries to target points of interest.
  • User can create their own query to operate on rows of data
  • User can create their own plot to display the data

Related Works and Future Features
• There is a Shiny application online that supports matrix portion of UpSetR
  • https://upsetr.shinyapps.io/UpSetR-shiny
• In a future version of the application, queries and attribute plots will be introduced.
• Currently working on a Shiny app linked to Ensembl gene database.
• Implementation of summary statistics for sets, intersections, and attributes in a future release
• UpSetR source code can be found at:
  • https://github.com/hms-dbmi/UpSetR
• An interactive version of the original UpSet and its paper can be found at: http://vcg.github.io/upset/about/
• Implementation of UpSetR inspired by:

Other Set Visualization R Packages
venneuler :

```r
myfunc <- function(row, release, rating){
data <- (row"ReleaseDate" %in% release) &
(row"AvgRating" > rating)
}
```

• Built in histogram and scatter plot functions to display element based data.
  • User can create their own plot to display the data
  • User can create their own query to operate on rows of data

```r
myplot <- function(mydata,x,y){
  Myfunc <- function(row, release, rating){
    data <- (row"ReleaseDate" %in% release) &
    (row"AvgRating" > rating)
  }
  
  # Built in histogram and scatter plot functions to display element based data.
  # User can create their own plot to display the data
  # User can create their own query to operate on rows of data

  geom_point() + scale_color_identity()
}
```

• Need to know what the intersection sizes are
• Becomes difficult to interpret when large number of intersections present. In UpSetR each intersection is represented as a column in the matrix.
• Need to know what the intersection sizes are prior to plotting.
• Need to enter intersections manually. UpSetR only needs a correctly formatted data set.
• Circlize can only display up to 2 way interactions.
• Igraph/Rgraphviz only show which sets interact. Can’t display intersections.
• Don’t allow for queries or additional plots of element data.

Shortcomings Compared to UpSetR

- Becomes difficult to interpret when large number of intersections present. In UpSetR each intersection is represented as a column in the matrix.
- Need to know what the intersection sizes are prior to plotting.
- Need to enter intersections manually. UpSetR only needs a correctly formatted data set.
- Circlize can only display up to 2 way interactions.
- Igraph/Rgraphviz only show which sets interact. Can’t display intersections.
- Don’t allow for queries or additional plots of element data.

Related Works and Future Features

- There is a Shiny application online that supports matrix portion of UpSetR
  - https://upsetr.shinyapps.io/UpSetR-shiny
- In a future version of the application, queries and attribute plots will be introduced.
- Currently working on a Shiny app linked to Ensembl gene database.
- Implementation of summary statistics for sets, intersections, and attributes in a future release
- UpSetR source code can be found at:
- An interactive version of the original UpSet and its paper can be found at: http://vcg.github.io/upset/about/
- Implementation of UpSetR inspired by:

This work was supported by the NIH Big Data to Knowledge (BD2K) Initiative (Award #U54HG007963)