Mastering upset plots: a comprehensive tutorial for effective data visualization in pharmacometrics and drug development

Mahmoud A. Afifi (1), Samer Mouksassi (1)
1. Certara; Radnor, NJ, USA .

Background & Objective

In data analysis, examining variable intersections is crucial for gaining insights essential for decision-making in various fields, such as pharmacometrics, clinical research, and healthcare management. Traditional visualization methods like Venn diagrams often fail with complex datasets, especially when the number of sets exceeds a manageable threshold, causing the "combinatorial explosion" problem. Upset plots have emerged as a superior alternative, effectively visualizing and interpreting complex data intersections. This tutorial guides you through mastering upset plots, demonstrating their practical applications with case studies.

Methods

Each case study illustrates the utility of upset plots in different domains:

1.Oncology: Understand how upset plots aid in exploring the pattern of pathology features associated with certain clinical interventions.

2.Diagnosis: Understand how upset plots help in exploring the pattern between the overlapping symptoms or positive microbiologic results within patient cohorts and supporting the decision making in clinical practice and clinical research.

3.Adverse Event Analysis: Discover how upset plots facilitate adverse event monitoring and management by visualizing the co-occurrence of adverse events within patient populations.

4.Missing Data Analysis: Gain insights into addressing missing data challenges using upset plots, identifying patterns and potential biases to ensure study validity.

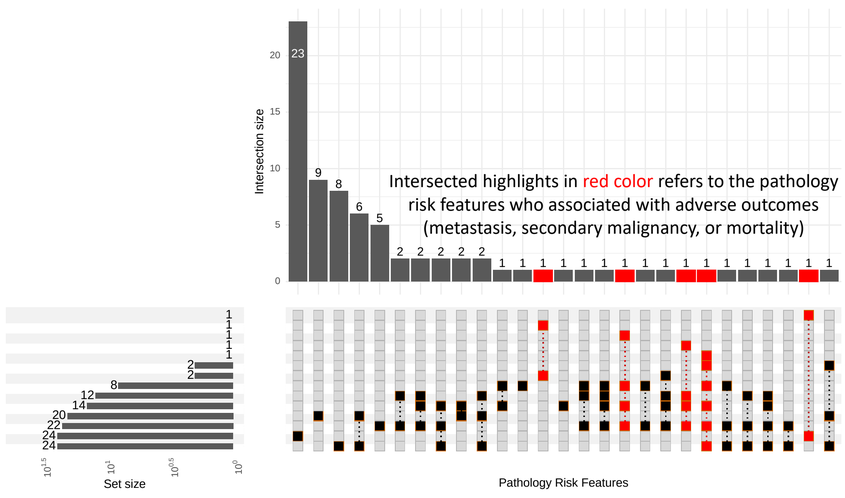
5.Medication Usage Profiling: Understand how upset plots elucidate medication usage patterns, concomitant drugs, and drug-drug interaction.

6. Covariate Selection: Show covariate ranking and selection patterns across a number of simulated datasets to facilitate the comparison of covariate selection methods.

Results

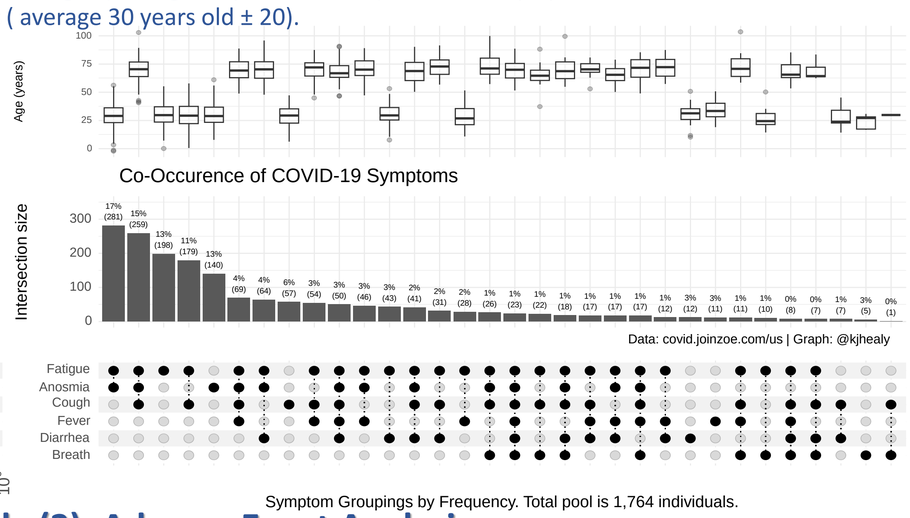
Case Study (1): Oncology:

In retinoblastoma, data analysis showed that additional chemotherapy may not be needed for high-risk patients (those with high-risk pathology features) who had their eye removed within a year after finishing other conservative eye therapy.



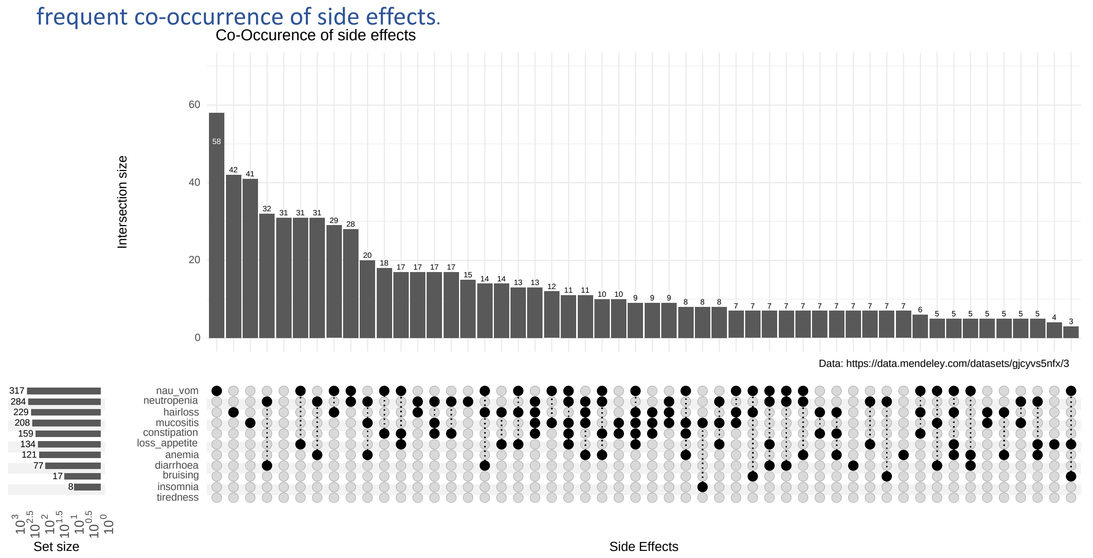
Case Study (2): Diagnosis:

The most common symptoms in COVID-19 was fatigue and anosmia (281 patients); it represents 17% of all other combinations of symptoms. The occurrence of three symptoms or more was correlated with old ages (average 70 years old ± 20), while two symptoms or less was correlated young patients (average 30 years old ± 20).



Case Study (3): Adverse Event Analysis:

Nausea and vomiting was the most frequent side effect. Neutropenia and diarrhea was the most frequent co-occurrence of side effects.



References:

1.Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph. 2014 Dec;20(12):1983-92. doi: 10.1109/TVCG.2014.2346248. PMID: 26356912; PMCID: PMC4720993.
2.Michal Krassowski. (2020). krassowski/complex-upset. Zenodo. http://doi.org/10.5281/zenodo.370059
3.El-zomor, H., Alieldin, A., Afifi, M. A., El-hussieny, A., Taha, H., Refaat, A., Elhussein, A., Ahmed, S., & Zaghloul, M. (2018). Outcome of unilateral retinoblastoma: A 10-year experience at Children Cancer Hospital in Egypt (CCHE). *SIOP 2019*, V447 SIOP19-1220
4. Asiimwe, I.G. Ndzamba, B.S., Mouksassi, S., Pillai, G., Lombard, A., Lang, J. Machine-Learning Assisted Screening of Correlated Covariates: Application to Clinical Data of Desipramine. AAPS J. 2024 May 30;26(4):63. doi: 10.1208/s12248-024-00934-6.

CERTARA

Mastering upset plots is crucial for unlocking *hidden insights*

Steps to build advanced upset plot

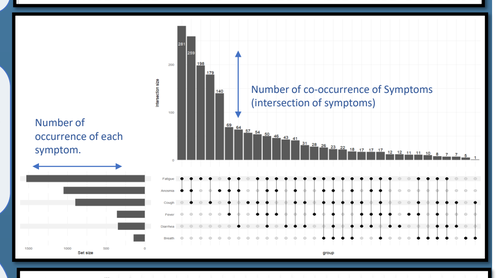
Step 1

- Make sure that the intersection variables are in the binary format (TRUE/FALSE OR 1/0)
- print(BinaryData)

```
> print(BinaryData)
   Anosmia Cough Fatigue Diarrhea Breath Fever       age
1     TRUE FALSE   FALSE    FALSE  FALSE FALSE  9.099591
2     TRUE FALSE   FALSE    FALSE  FALSE FALSE 25.032675
3     TRUE FALSE   FALSE    FALSE  FALSE FALSE 48.826537
4     TRUE FALSE   FALSE    FALSE  FALSE FALSE 24.933151
```

Step 2

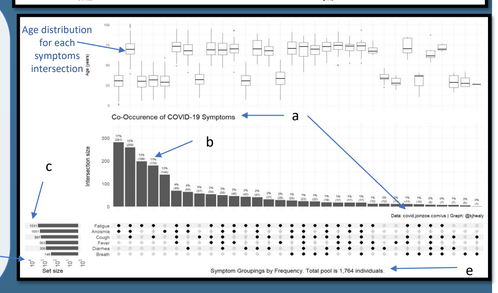
- Create simple/basic upset plot using ComplexUpset::upset() function with two aurgments
  1- the previous dataset (binary data)
  2- the name of intersection variables.
- upset(data = BinaryData, intersect = symptoms)



Step 3

- Using annotations feature, you can add another ggplot object to explore the pattern of another variable for each intersection.
- In this example, we will add the 'Age at Diagnosis' distribution (boxplot) for each intersection symptom.

```
isize =
upset_annotate('age',geom_boxplot(na.rm=TRUE))
isize$geom = c(
 geom_boxplot(na.rm=TRUE, alpha=0.3),
 scale_y_continuous(name="Age (years)"))
upset(data = BinaryData, intersect = symptoms,
      annotations = list('age'=isize))
```



Step 4

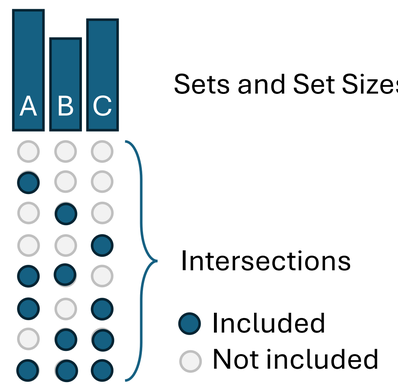
- **More additional Features:**

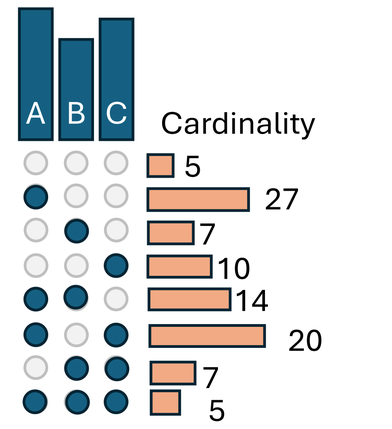
a) adding title and caption.
b) adding percentages over the bars
c) adding the numbers for set size.
d) changing the Set Size to the power format.
e) adding titles for x and y axes.
f) playing with the theme:
- the size of text.
- the size of the plot.
- The size and shape of intersections points.
- Highlighting the intersections and bars.

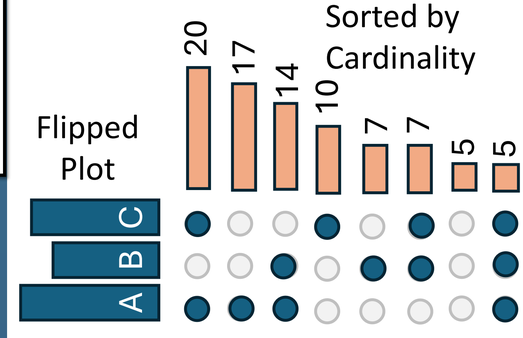
Scan me

Want to learn more? Scan Here all code can be shared

https://upset.app/







Additional data info can be added for each intersection or margin see examples

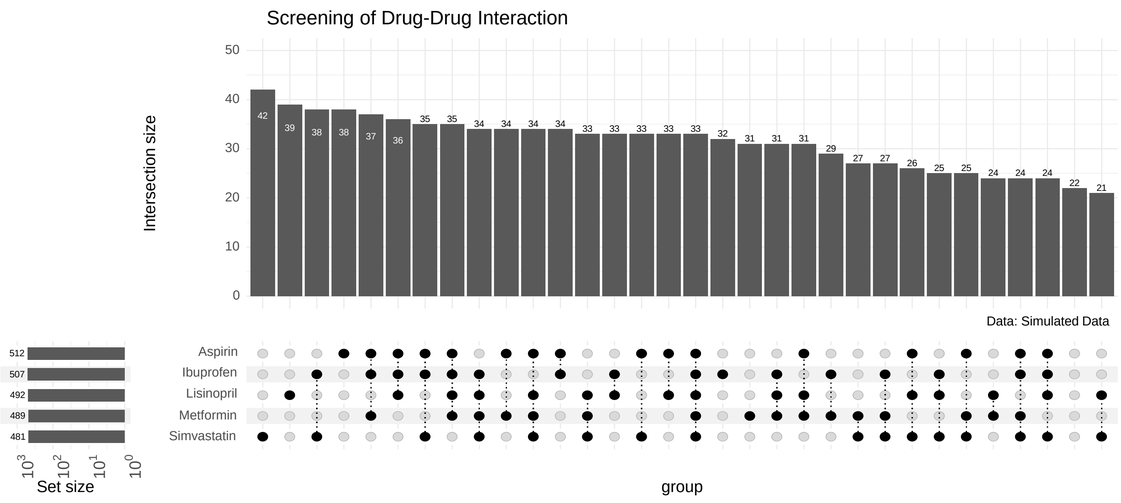
Applications of Upset Plot

Case Study (4): Missing Data Analysis:

Missing in reported side effects was Missing Not At Random (MNAR). The missing was correlated with old physicians.

Case Study (5): Medication Usage Profiling:

Ibuprofen + Simvastatin is the most frequent concomitant medications (38 patients).



Case Study (6): ML Covariate Selection:

Use NONMEM and Caret R Package (implements >= 200 methods)
- Linear regression, GAM, stepwise linear methods (AIC/RMSE), penalized regression, Random Forest, Neural Networks, XGBoost.