Module 5.1: Learn - Biology

Overlap analysis for higher number of comparisons

In the previous module, we used tools to compare results from two data sets such as a Venn diagram. In this study, we have 9 combinations of trimming parameters we are using to cover a range of 2 trimming parameters. As you can see in the figure below, as you increase the number of groups you are comparing, representing overlap using Venn diagrams gets very complicated:

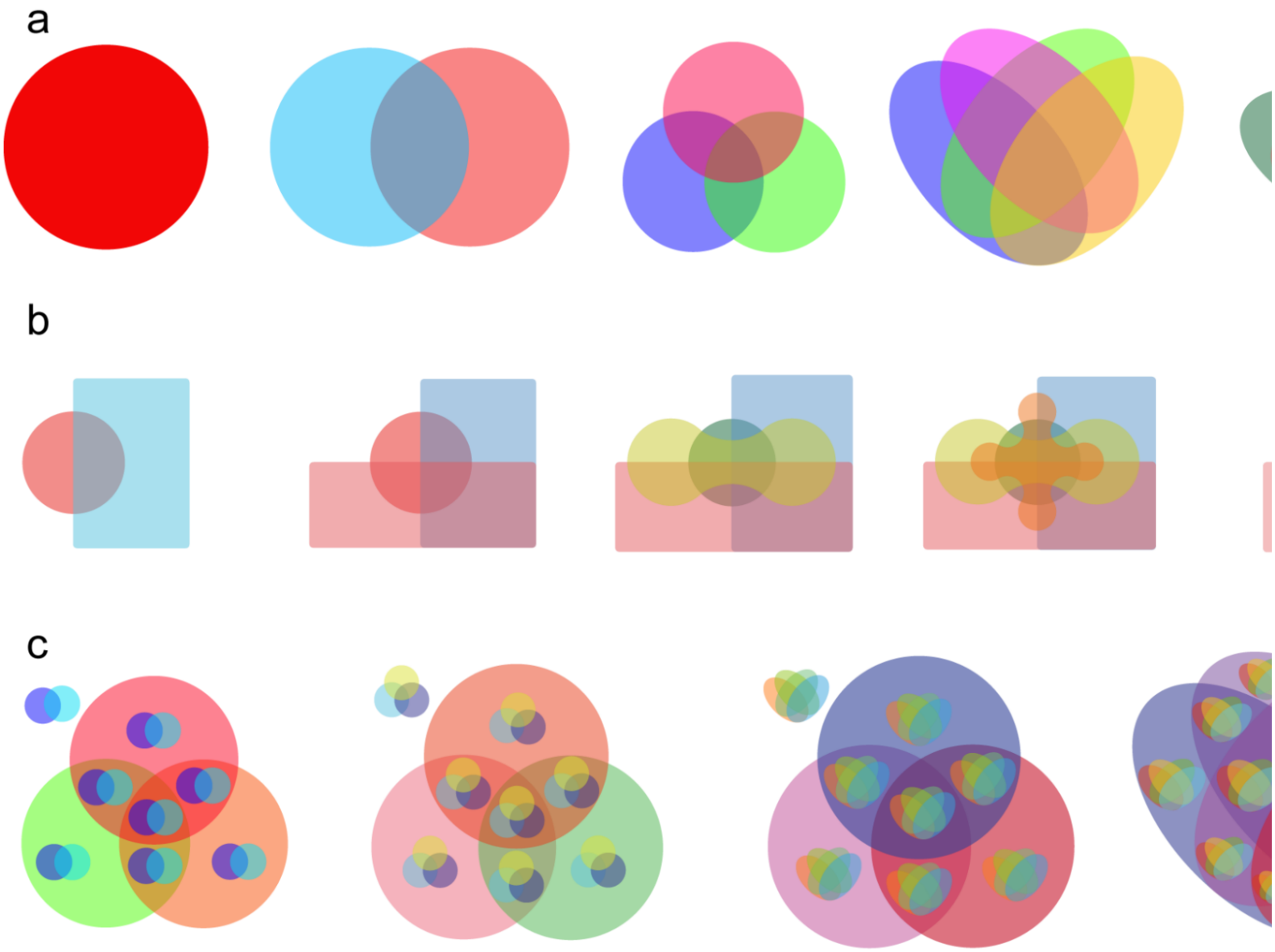


Figure. Increasing complexity of Venn diagrams.

Source. (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154315)

For this reason we are going to use another type of plot called an upset plot to represent overlap between differe
lists from our 9 data sets. An upset plot uses bar graphs to show the values of intersections you would put on a V

lines to represent membership between multiple groups. In the example shown below, we have an upset plot tha
membership to 5 different genres of a collection of movies: Crime, Thriller, Fantasy, Comedy, and Drama. The m
lines to define intersections in groups; a single unconnected dot represents movies uniquely belonging to one ge
multiple rows represent movies belonging to multiple genres. In this way, we can make observations like the higl
that belong to two genres are Comedy and Drama (163 movies).

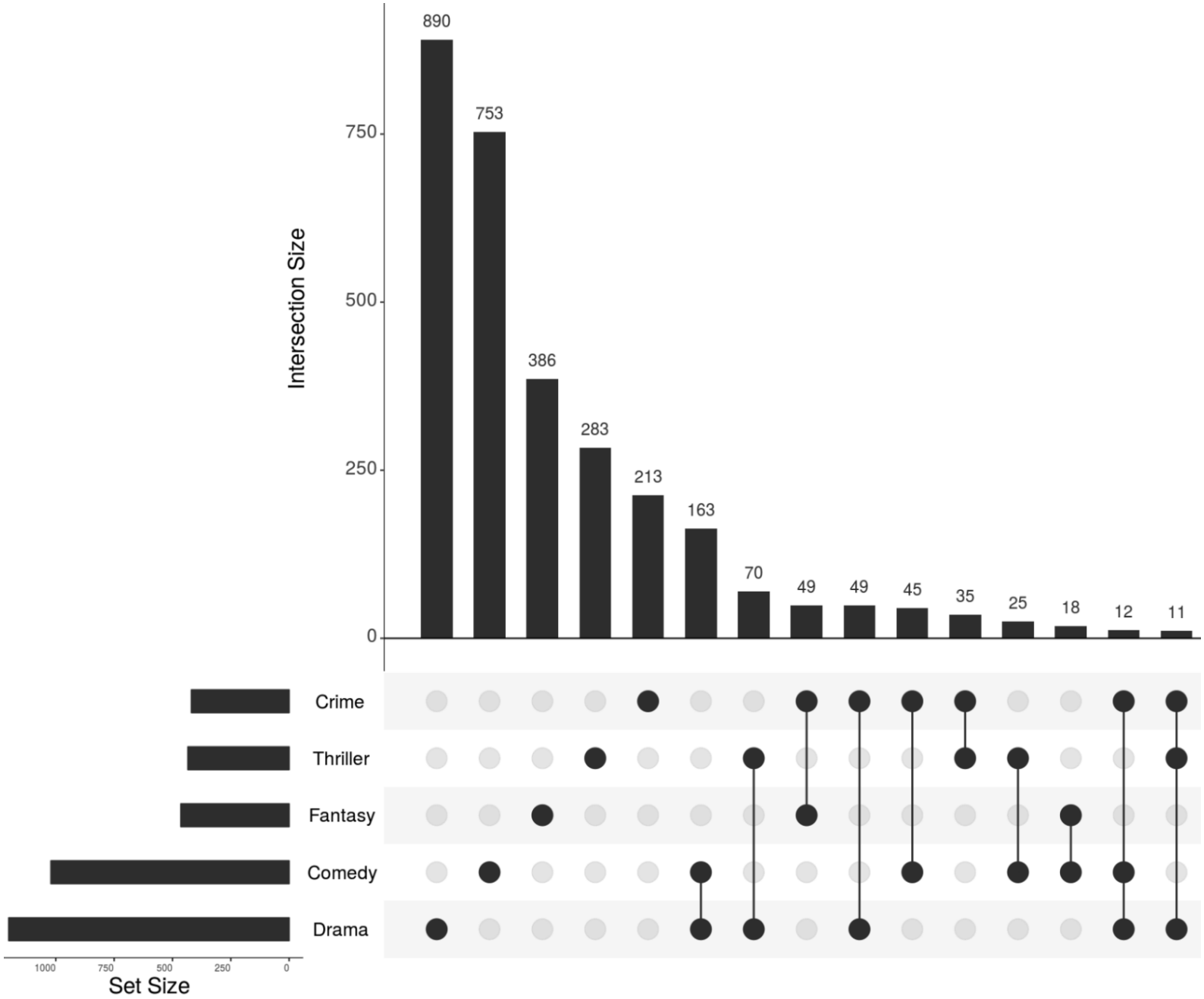
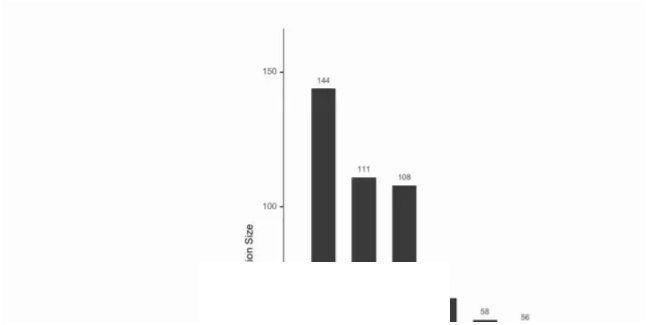


Figure. Upset plot.

An upset plot showing the intersection of movie genres in a collection of movies.

This video describes another simple example of how to read an upset plot and shows the R function we will be us



Video. Upset Plots in R demonstration.

This video describes what an upset plot is and how to create one in R.

View transcription. (https://canvas.asu.edu/courses/122165/files/55608284?wrap=1) **(https://canvas.asu.edu/courses/122165/files/55608284/download?download_frd=1)**

Starting a list of questions we can ask about this data

The aim of this research is to determine the effect of trimming parameters trimq and minlen on differential gene e
question can be broken down into many smaller questions, the answers to which will be what you will be determi
the final project of this course.

Here are some questions to get you started:

1. What metrics of differential expression are affected after varying `trimq`? How?
2. What metrics of differential expression are affected from varying `minlen`? How?
3. Are there any common themes with genes that are robustly differentially expressed no matter what trimming p

    - Higher expression levels?
    - Less noise?
    - Higher fold change between groups?
4. Which genes changed the most between untrimmed and trimmed? Between various trimming parameters?
5. Do we see changes in the MDS plots showing differences between sample?

To answer these questions, you will have to do some programming. R can definitely be confusing at times, so pl
reach out for help.

Module 5.1 Additional Resources

- **Explanation of upset plots from the originators (https://upset.app/)**