

A Study Instruments

Here we show simplified versions of the study instruments used in the survey and interview studies

A.1 Survey Study

An anonymized version of our survey is available at <https://submitterpaper.github.io/study/Upset-Alttext-User-Survey/0>. This includes a full copy of the instructions participants were shown and supports browsing each of the study conditions.

A.2 Interview Study

A.2.1 Demographic Information.

- (1) What is your date of birth?
- (2) What is your gender?
- (3) What is the highest level of education you have completed or are in the process of completing?
- (4) Which screen reader do you use with your computer or other devices (e.g., NVDA, JAWS, VoiceOver, etc.)
- (5) How long have you been using a screen reader?
- (6) Do you use other accessibility devices or software in combination with a screen reader, such as screen magnification or a braille display? If yes, please describe.
- (7) What is your preferred rate of speech when using a screen reader?
- (8) How many hours are you on a computer each day?
- (9) Would you consider your career to be data-intensive or numbers-driven (e.g., regularly work with large datasets, perform statistical analyses, or make decisions based on quantitative information)?
- (10) How often do you interact with data visualizations, such as those for work, from news articles, in video games, etc.? And in what context?
- (11) How would you describe your vision-loss level?
- (12) How would you describe your vision level (e.g., no remaining vision, light perception, central vision, etc.)?
- (13) What is your corrected visual acuity in either Snellen (e.g., 20/200) or LogMAR (e.g., 1.3)?

A.2.2 UpSet Alt-Text Questions.

- (1) Can you describe what you learned about the dataset in your own words?
- (2) What is the dataset about?
- (3) How many sets are shown?
- (4) What is the largest intersection?
- (5) How would you describe this dataset to a friend?
- (6) How did this dataset increase your understanding of COVID symptoms?
- (7) Do you feel like you have a good sense of the dataset? Why, why not?
- (8) What was difficult for you to understand about the dataset?
- (9) What would have been helpful to provide additionally? Was anything missing?
- (10) Do you have comments on the style of the text description? For example, was it too long, too short, too verbose?
- (11) Do you have any other feedback or comments that we did not touch on today?
- (12) In your opinion, what differentiates a great text description from other descriptions?
- (13) Do you have any experiences with accessible data visualizations? If so, please elaborate.

A.2.3 UpSet Plot Questions.

- (1) Does your mental model from the text description match the chart?
- (2) Do you have any other feedback or comments now that you have seen the corresponding visualization?

A.3 Interview Stimuli

Page 1

Introduction

For this study, you will see a text description generated for an UpSet plot. An UpSet plot is a set visualization technique similar to Venn diagrams, but unlike Venn diagrams, UpSet works for more than three sets.

Our research aims to make data visualizations more accessible to people with visual impairments. We want to understand whether text can convey similar amounts of information as a chart. We first will introduce what an UpSet plot is, and how to interpret data from the plot. UpSet Explained

UpSet plots the intersections of sets in a table. Each column corresponds to a set. Bars at the top of the columns show the size of the sets. The row corresponds to an intersection: marks in the cells show which set is included in the intersection. The number of sets that participate in the intersection is referred to as degree. If there is no mark in any of the cells, then it is the intersection of no set, which is also referred to as the empty intersection, with a degree of 0. If there is a mark in every cell of a row, then it is the intersection of all sets.

UpSet plots the size of the intersections as bar charts to the right of the table. The table is also useful because it can be sorted in various ways. A common way is to sort by size, but it's also possible to sort by degree or sets.

Imagine an UpSet plot that shows movie data. Movies have genres like Drama, Comedy, Thriller, Mystery, or Crime. A movie can have a single genre, or it can have multiple genres. In this example, the genres are the sets. Some sets are bigger: there are more Drama movies than Mystery movies, for example. And some intersections will be more common: Thriller and Mystery might be a popular combination, while the combination of Drama, Comedy, and Thriller might be rare. Glossary of Terms

Here are some terms we use in the text description:

Movies that don't fall into any genre are an intersection of no set/the empty intersection. A row with only one mark for drama (movies that are just dramas and have no other genre) is an independent set intersection. A row with 2-3 marks like "Drama-Comedy" corresponds to a low-degree set intersection. A row with 3-5 marks (e.g., Mystery-Crime-Thriller) corresponds to a medium-degree set intersection. A row with even more marks is a high-order set intersection. The last case is that set containing all sets (i.e. all movie genres), which we call an all-set intersection.

That's it for the introduction! Please stop now and ask the interviewers if you might have any clarifying questions.

Next, we'll explore text descriptions generated for a variety of UpSet Plots. Click on the link below to see: UpSet Plot Description.

Visualizing co-occurrence of CoVID 19 Symptoms with UpSet.

```

1613 3 This is an UpSet plot that shows covid cases and their symptoms. A covid case
1614     can have multiple symptoms. The sets are covid symptoms. The items are
1615     covid cases. The intersections show how many covid cases have exactly the
1616     same symptoms. The plot shows intersections of 6 sets. All major
1617     intersections involve the set Fatigue, and Cough. The largest intersection
1618     is Anosmia, and Fatigue, with 281 elements. Other large intersections also
1619     involve Cough, Anosmia, and Fatigue. The intersection of all sets is
1620     present with 23 elements.
1621
1622
1623 4
1624 5 # Dataset Properties
1625
1626 6
1627 7 The dataset contains 6 sets and 4340 elements, of which 6 sets are shown in the
1628     plot.
1629
1630 8
1631 9 # Set Properties
1632
1633 10
1634 11 The set sizes are diverging a lot, ranging from 148 to 1531. The largest set is
1635     Fatigue with 1531 elements, followed by Anosmia with 1051, Cough with 897,
1636     Fever with 363, Diarrhea with 350, and Shortness of Breath with 148.
1637
1638 12
1639 13 # Intersection Properties
1640
1641 14
1642 15 The plot is sorted by size in descending order. There are 32 non-empty
1643     intersections, all of which are shown in the plot. The largest 5
1644     intersections are Anosmia, and Fatigue (281), Cough, Anosmia, and Fatigue
1645     (259), Just Fatigue (198), Cough, and Fatigue (179), and Just Anosmia (140)
1646     .
1647 16 Statistical Information
1648
1649 17
1650 18 The average intersection size is 55, and the median is 24. The 90th percentile
1651     is 179, and the 10th percentile is 7. The largest set, Fatigue, is present
1652     in 78.1% of all non-empty intersections. The smallest set, Shortness of
1653     Breath, is present in 34.4% of all non-empty intersections.
1654
1655 19
1656 20 # Trend Analysis
1657
1658
1659
1660
1661
1662
1663
1664 Manuscript submitted to ACM

```

The intersection sizes peak at a value of 281 and then drastically flatten down to 1. An all set intersection is present with a size of 23. The individual set intersections are large in size. The low degree set intersections lie in the largest sized intersections. The medium degree set intersections can be seen among small and medium and large sized intersections. Among the medium sized intersections, the high order set intersections are significantly present.

B Text Description Examples

Here we provide examples of the text descriptions used in our studies for the Movies dataset as an example of general trend.

B.1 Movies

Short Description no Configuration

This is an UpSet plot which shows the intersections of 6 sets. All major intersections involve the set Action, and Adventure. The largest intersection is Thriller, and Action, with 104 elements. Other large intersections also involve Action and Thriller.

Long Description with no Configuration

UpSet Introduction

This is an UpSet plot that visualizes set intersection. To learn about UpSet plots, visit REDACTED.

Dataset Properties

The dataset contains 17 sets and 6303 elements, of which 6 sets are shown in the plot.

Set Properties

The set sizes are diverging a lot, ranging from 68 to 503. The largest set is Action with 503 elements, followed by Thriller with 492, Adventure with 283, Children with 251, War with 143, and Western with 68.

Intersection Properties

The plot is sorted by size in descending order. There are 28 non-empty intersections, all of which are shown in the plot. The largest 5 intersections are Just the empty inter (2569), Just Thriller (349), Just Action (218), Just Children (160), and Thriller, and Action (104).

```

171713 # Statistical Information
171814 The average intersection size is 138, and the median is 7. The 90th percentile
1719 is 218, and the 10th percentile is 1. The largest set, Action, is present
1720 in 50.0\% of all non-empty intersections. The smallest set, Western, is
1721 present in 25.0\% of all non-empty intersections.
1722
172315
172416 # Trend Analysis
1725 The intersection sizes peak at a value of 2569 and then drastically flatten
172617 down to 1. Just the empty inter is the largest by a factor of 7. The empty
1727 intersection is present with a size of 2569. An all set intersection is not
1728 present. The individual set intersections are large in size. The low
1729 degree set intersections lie in small and medium sized intersections. The
1730 medium degree set intersections can be seen among medium sized
1731 intersections. No high order intersections are present.
1732
1733
1734
1735
1736 Long Description with Configuration
1737
17381 # UpSet Introduction
1739 This is an UpSet plot that visualizes set intersection. To learn about UpSet
17402 plots, visit REDACTED.
1741
17423
17434 # Dataset Properties
1744 The dataset shows attributes of movie genres and ratings. The dataset contains
17455 17 sets and 6303 elements, of which 6 sets are shown in the plot.
1746
17476
17487 # Set Properties
1749 The set sizes are diverging a lot, ranging from 68 to 503. The largest set is
17508 Action with 503 movies, followed by Thriller with 492, Adventure with 283,
1751 Children with 251, War with 143, and Western with 68.
1752
17539
175410 # Intersection Properties
1755 The plot is sorted by size in descending order. There are 28 non-empty
175611 intersections, all of which are shown in the plot. The largest 5
1757 intersections are Just the empty inter (2569), Just Thriller (349), Just
1758 Action (218), Just Children (160), and Thriller, and Action (104).
1759
1760
176112
176213 # Statistical Information
176314 The average intersection size is 138, and the median is 7. The 90th percentile
1764 is 218, and the 10th percentile is 1. The largest set, Action, is present
1765 in 50.0% of all non-empty intersections. The smallest set, Western, is
1766 present in 25.0% of all non-empty intersections.
1767
1768 Manuscript submitted to ACM

```

Trend Analysis

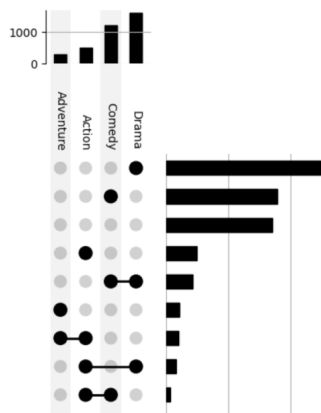
The intersection sizes peak at a value of 2569 and then drastically flatten down to 1. Just the empty inter is the largest by a factor of 7. The empty intersection is present with a size of 2569. An all set intersection is not present. The individual set intersections are large in size. The low degree set intersections lie in small and medium sized intersections. The medium degree set intersections can be seen among medium sized intersections. No high order intersections are present.

C Additional Figures

Finally, we include several figures which are relevant to this but were out of place in the main text. Figure 13 shows an example of our descriptive text integrated with an UpSet plot in a notebook. Figure 14 is a snapshot of our coding of in-the-wild UpSet examples. Figure 15 displays post-survey study responses about the participant preferences for UpSets and our text descriptions of them.

```
In [4]: plot1.plot()
plt.suptitle("UpSet plot with grammar and text description")
Out[4]: Text(0.5, 0.98, 'UpSet plot with grammar and text description')
```

UpSet plot with grammar and text description



Fetch the generated text description

```
In [5]: text_description = plot1.get_alt_text()
```

The text description comes as a JSON format with 3 fields.

- Technique Description: A brief technique of UpSet Plots
- Short Description: A summary of the UpSet Plot. This is similar to a caption
- Long Description: Markdown formatted. In-depth description of many aspects of the plot

Technique Description:

This is an UpSet plot that visualizes set intersection. To learn about UpSet plots, visit <https://upset.app>.

Short Description:

This is an UpSet plot which shows the intersections of 4 sets. All major intersections involve the set Adventure, and Comedy. The largest intersection is Comedy, and Drama, with 215 elements. Other large intersections also involve Action and Comedy.

Fig. 13. A screen shot of a notebook using our automated text generation system.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

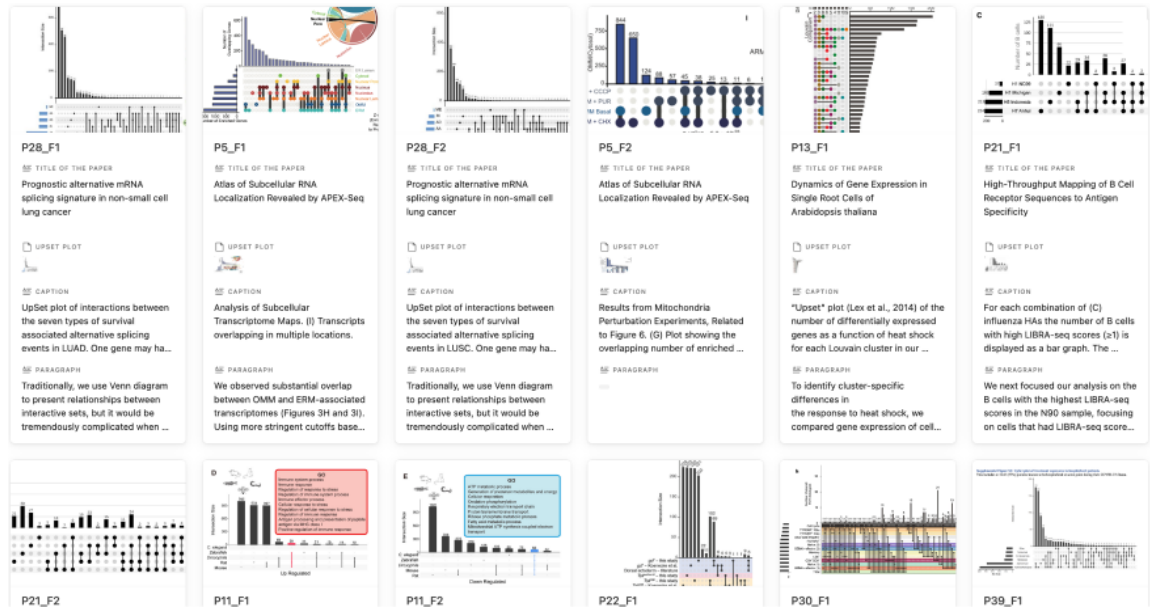


Fig. 14. Selected thumbnails of UpSet plots collected for our classification of patterns found in UpSet plots.

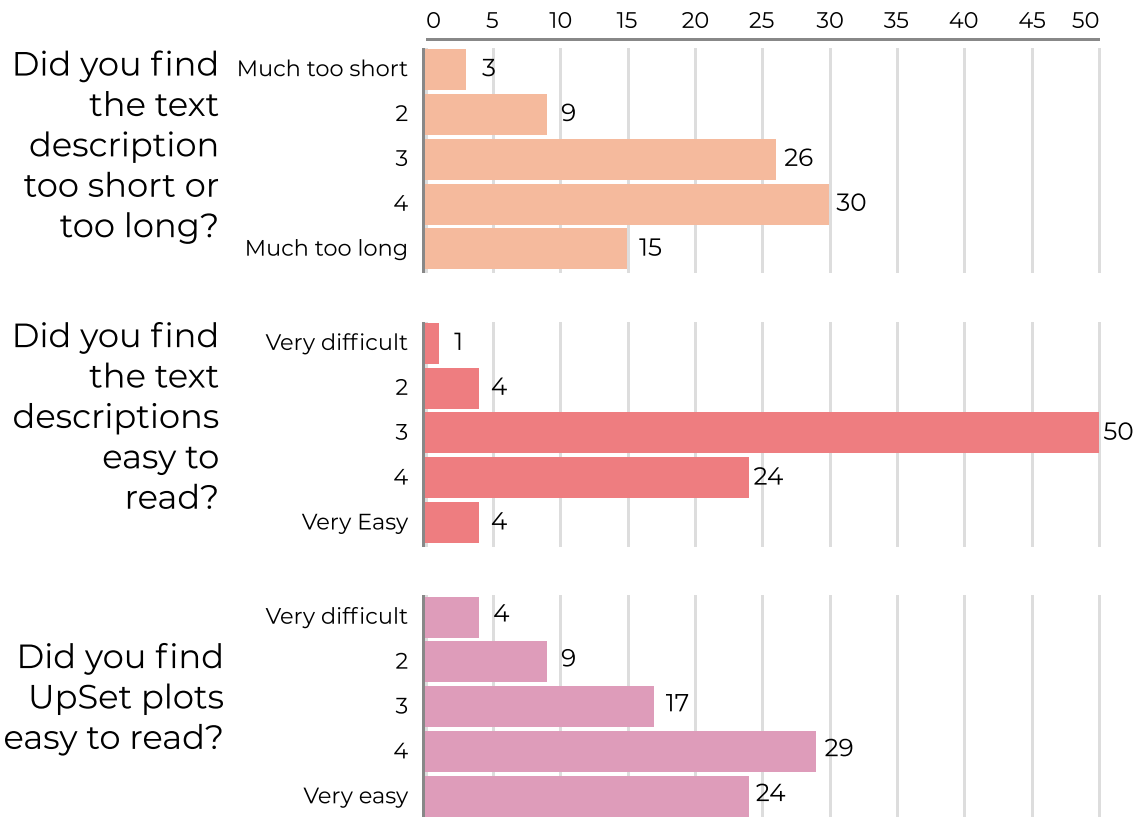


Fig. 15. Espoused preferences from survey respondents in the post-experiment survey.