

# Prompts

## Strong prompt:

You are Dr. Alexander Lex, PhD in data visualization and creator of the Upset plot. You have been tasked with writing an alt text to describe this Upset plot to users with visual impairments. *You have been provided with both an image of the plot and a JSON file containing each row of the plot in textual form.* Using your decades of experience with interactive data analysis, generate a text description that has three key emphases: First, it should describe the conclusions that a sighted users would draw from the plot at a high semantic level. Describing the appearance of the plot and what the components mean is irrelevant and unuseful; assume that your reader already understands what an upset plot is and how it works. Secondly, focus on specifics. Do not resort to vague references or generalized conclusions. Instead, cite specific pieces of data visualized on the plot to support the high-level semantic deductions and conclusions you draw from your visual analysis. Finally, focus on accuracy. It is not acceptable for even a single number to be wrong in your text description. Make sure to double-check the numbers you include in your result to ensure that your overarching conclusions are accurate. *Use the data file to check whether your visual conclusions are correct.* While focusing on these three emphases, use your acumen and the power of interactive visual data analysis to make the conclusions supported by this plot accessible to low-vision users via a high-level, detailed, specific, and accurate text description.

## Pattern analysis prompt:

To assist your analysis of this upset plot, you've compiled a list of patterns that you've seen across many upset plots, while keeping in mind that this plot may represent a new pattern that you haven't seen before. The patterns are as follows:

- Intersection Patterns: Are there many independent sets (intersections containing only 1 set). Are there many low-set intersections (intersections containing 2-3 sets), medium-degree set intersections (intersections containing 3- $n/2$  sets), or high order set intersections (intersections containing  $n/2$  to  $n$  sets).
- Intersection Size Patterns: what are the intersection sizes associated with the intersection patterns? Are the higher-order intersections (involving most sets) large or small? Are the low-set intersections large or small? We classify these into small, medium, large, and largest.
- Specific Attributes: Is the all-set intersection present? Is the empty intersection present? Are the all-set and empty intersections the largest, among the largest, or small?
- Set Sizes: Are set sizes roughly equal, diverging a lot, or moderately diverging?

While the plot may contain a pattern that you haven't seen before, you know that thinking about these types of patterns is a great starting point for reaching complex conclusions about the data represented by the plot.

# Methods

Conditions are entered into the AI in the order presented in the spreadsheet. For API generations, the first sentence of the strong prompt is used for the system prompt. Then, the user prompt is filled with the image, “Here is the JSON representation of the plot:”, accessible processed data (APD), “Here is an example of another alttext generated for a different plot:”, example alttext in quotes, strong prompt excluding first sentence, then pattern prompt. It should be noted that the APD includes attribute information but the image does not; that might need to be changed.

For the Ablation study, coding was first done separately by 2 authors who then converged to agree on final coding. Their preliminary results are [here](#) and [here](#).

The pilot study used a slightly different codebook and is not entirely consistent with the following studies, ie it was done before the 2 authors converged and agreed on a codebook.

## Pre-Study Result Coding

Level 1

Level 2

Level 3

Level 4

Misleading

Heading numbers correspond to rows in the LLM study spreadsheet. Duplicate facts are not removed, so some fact counts represent duplicated information; ie both LLMs prefer to finish with a concluding paragraph that duplicates what they’ve said previously. The Accessible Processed Data included attribute information which made it into 1 generation; since the images and our alttext generator don’t account for attributes, parts of the result that mention attributes (other than degree and deviation) are struck out and not considered.

### 2: Claude, movies, all conditions

L1: 1

L2: 14

L3: 4

L4: 10

---

This UpSet plot visualizes the genre combinations of movies in a dataset, focusing on six main genres: **Thriller, War, Western, Action, Adventure, and Children**. The analysis reveals several key insights about the distribution and overlap of these genres:

Genre Dominance: **Thriller is the most prevalent genre, with 349 movies classified solely as Thrillers. This is followed by Action (218 movies) and Children's movies (160 movies).** This suggests a strong market presence for these individual genres.

Genre Combinations: **The most common genre combination is Thriller and Action, with 104 movies.** This indicates a significant overlap between these two genres, likely representing action-packed thrillers.

Multi-Genre Trends: As the number of combined genres increases, the number of movies in each category generally decreases. For example, there are only 23 movies that combine Thriller, Action, and Adventure. This suggests that movies tend to focus on one or two primary genres rather than spanning multiple categories.

Niche Combinations: Some genre combinations are quite rare. For instance, there is only one movie that combines Thriller, Western, and Action. Similarly, there's only one movie each that combines War with Children, and Western with Adventure and Action. This highlights the uniqueness of certain genre blends.

Genre Independence: Western and War genres show a tendency to stand alone more often than combine with others. There are 52 pure Westerns and 88 pure War movies, compared to their relatively low numbers in combinations.

Children's Genre Crossovers: Children's movies show interesting patterns. While there are 160 pure Children's movies, there are also 75 that combine Adventure and Children, suggesting a strong affinity between these genres for younger audiences.

High-Order Combinations: The highest-order combination in the dataset involves four genres: Thriller, War, Action, and Children, but this is extremely rare with only one such movie.

Set Sizes: The set sizes are diverging significantly. The largest set is the "Unincluded" category with 2,569 movies, suggesting many movies in the dataset don't fall into these six main genres. Among the analyzed genres, Thriller is the largest with 349 solo entries, while Western is the smallest with 52.

Intersection Patterns: The plot shows a predominance of low-set intersections (1-2 genres), with decreasing frequency as more genres are combined. This suggests that most movies tend to focus on one or two primary genres rather than spanning multiple categories.

In conclusion, this UpSet plot reveals a complex landscape of movie genre classifications, with clear preferences for certain genres and combinations, while also highlighting the rarity of movies that span multiple genre categories. The data suggests that while there are dominant individual genres, the film industry also produces a variety of cross-genre movies, catering to diverse audience preferences.

---

## 5: GPT, anon movies, all conditions

L1: 4

L2: 23

L3: 6

L4: 0

E: 3

---

This UpSet plot visualizes the intersections of 17 sets, with a focus on the specific attributes and deviations of each subset. Here is a detailed description of the key insights and patterns observed from the plot:

### Intersection Patterns and Sizes:

1. **\*\*Independent Sets\*\***: The plot shows several independent sets (intersections containing only one set). The largest independent set is "Unincluded" with 2569 elements, followed by "Just Set\_15" with 349 elements, "Just Set\_1" with 218 elements, and "Just Set\_3" with 160 elements. Other independent sets have smaller sizes, such as "Just Set\_16" with 88 elements and "Just Set\_17" with 52 elements.

2. **\*\*Low-Set Intersections\*\***: There are numerous low-set intersections (intersections containing 2-3 sets). Notable low-set intersections include:

- "Set\_1, and Set\_15" with 104 elements.
- "Set\_1, and Set\_2" with 90 elements.
- "Set\_2, and Set\_3" with 75 elements.
- "Set\_1, and Set\_16" with 34 elements.
- "Set\_1, Set\_2, and Set\_15" with 23 elements.

3. **\*\*Medium-Degree Intersections\*\***: Medium-degree intersections (involving 3-4 sets) are present but less frequent. Examples include:

- "Set\_1, Set\_2, and Set\_16" with 7 elements.
- "Set\_1, Set\_3, Set\_15, and Set\_16" with 1 element.

4. **\*\*High-Order Intersections\*\***: High-order intersections (involving more than 4 sets) are rare and generally small in size. The plot does not show any high-order intersections with significant sizes.

#### ### Specific Attributes:

1. **\*\*All-Set Intersection\*\***: The all-set intersection (involving all 17 sets) is not present in the plot.
2. **\*\*Empty Intersection\*\***: The empty intersection is represented by the "Unincluded" subset, which is the largest with 2569 elements.

#### ### Set Sizes:

- The set sizes diverge significantly. The largest set is "Unincluded" with 2569 elements, while other sets have much smaller sizes, such as "Just Set\_1" with 218 elements and "Just Set\_3" with 160 elements. This indicates a wide range of set sizes.

#### ### Deviation Analysis:

- The deviation values for each subset vary, with some subsets showing positive deviations and others showing negative deviations. Notable deviations include:

- "Unincluded" with a deviation of 3.78%.
- "Just Set\_1" with a deviation of -3.67%.
- "Set\_1, and Set\_15" with a deviation of 1.33%.
- "Set\_2, and Set\_3" with a deviation of 1.59%.

#### ### Conclusion:

The UpSet plot reveals a diverse range of intersection patterns, with a significant number of independent sets and low-set intersections. The set sizes diverge considerably, with the "Unincluded" subset being the

largest by a substantial margin. The deviation analysis highlights varying trends across different subsets, with some showing notable positive or negative deviations. Overall, the plot provides a comprehensive view of the intersections and attributes of the 17 sets, emphasizing the diversity and complexity of the data.

---

#### 4: Claude, anon movies, all conditions

L1: 0

L2: 20

L3: 7

L4: 6

---

This UpSet plot reveals several interesting patterns in the dataset, which appears to represent different subsets or categories within a larger population. Here are the key insights:

Set Sizes and Distribution:

The set sizes are highly divergent, ranging from 349 for Set\_15 to 2569 for the unincluded elements. This suggests a significant imbalance in the distribution of elements across the different categories. Set\_1, Set\_2, and Set\_3 have moderate sizes (218, 66, and 160 respectively), while Set\_16 and Set\_17 are smaller (88 and 52).

Intersection Patterns:

The plot shows a mix of low-degree and medium-degree intersections, with very few high-order intersections. The largest intersections tend to involve only 2-3 sets, indicating that elements are more likely to belong to a small number of categories simultaneously.

Dominant Sets:

Set\_1 and Set\_15 appear in many of the larger intersections, suggesting they are central categories that often overlap with others. For example, the intersection of Set\_1 and Set\_15 contains 104 elements, which is one of the largest intersections.

Unique Elements:

There's a significant number of elements that belong to only one set. Set\_15 has the most unique elements (349), followed by Set\_1 (218) and Set\_3 (160). This indicates that these sets have distinct characteristics not shared by other categories.

Complex Relationships:

While most large intersections involve only 2-3 sets, there are some notable medium-degree intersections. For instance, the intersection of Set\_1, Set\_2, and Set\_15 contains 23 elements, showing that some elements share characteristics across multiple categories.

Absence of High-Order Intersections:

The largest intersection involving 4 sets (Set\_1, Set\_3, Set\_15, and Set\_16) contains only 1 element. This suggests that it's extremely rare for elements to belong to many categories simultaneously.

Empty Set:

The plot shows a large number of elements (2569) that don't belong to any of the displayed sets. This "unincluded" group is actually the largest, indicating that a significant portion of the population doesn't fit into the defined categories.

Attribute Patterns:

Looking at the "Deviation" column, we can see that some intersections show significant positive or negative deviations. For example, the intersection of Set\_1 and Set\_2 has a deviation of 1.59, while the intersection of just Set\_1 has a deviation of -3.67. This suggests that elements belonging to certain combinations of sets have distinct characteristics compared to the overall population.

In conclusion, this UpSet plot reveals a complex system where elements tend to belong to a small number of categories simultaneously, with Set\_1 and Set\_15 playing central roles. The large number of unincluded elements and the presence of significant unique elements in several sets suggest a diverse population with many distinct subgroups. The varying deviations across intersections indicate that membership in specific combinations of sets is associated with notable differences in characteristics or behavior compared to the overall population.

---

### 3: GPT, Movies, all conditions

L1: 2

L2: 14

L3: 2

L4: 4

E: 9

---

This UpSet plot visualizes the intersection of six movie genres: Action, Adventure, Children, Thriller, War, and Western. The dataset contains 2,569 movies, with the largest individual genre being Thriller (349 movies) and the smallest being Western (52 movies). The plot reveals several key insights about the distribution and intersection of these genres.

### Intersection Patterns and Sizes:

1. **\*\*Independent Sets\*\***: The largest independent set is Thriller with 349 movies, followed by Action (218), and Children (160). The smallest independent set is Western with 52 movies. These independent sets show a significant number of movies belonging to only one genre.

2. **\*\*Low-Set Intersections\*\***: The most notable low-set intersections (involving 2-3 genres) include:

- Thriller and Action (104 movies)
- Action and Adventure (90 movies)
- Adventure and Children (75 movies)
- War and Action (34 movies)
- Thriller, Action, and Adventure (23 movies)

These intersections indicate a moderate number of movies that span across two or three genres.

3. **Medium-Degree Intersections**: There are very few medium-degree intersections (involving 3-4 genres), with the largest being Thriller, War, Action, and Children (1 movie).
4. **High-Order Intersections**: The plot does not show any high-order intersections involving more than four genres.

#### ### Specific Attributes:

1. **All-Set Intersection**: There is no intersection that includes all six genres.
2. **Empty Intersection**: The empty intersection (movies that do not belong to any of the six genres) is the largest subset with 2,569 movies, indicating that a significant portion of the dataset does not fall into these specific genres.

#### ### Set Sizes:

The set sizes diverge significantly, with Thriller being the largest set (349 movies) and Western being the smallest (52 movies). This divergence suggests that some genres are much more prevalent in the dataset than others.

#### ### Statistical Information:

1. **Average Ratings**: The average ratings across different intersections vary, with the highest mean rating found in the War and Adventure intersection (4.095) and the lowest in the Action and Children intersection (1.94).
2. **Watches**: The number of watches also varies significantly, with the highest mean watches in the War, Action, and Adventure intersection (1,239.43) and the lowest in the Action, Adventure, and Children intersection (181.83).
3. **Release Dates**: The median release dates for movies in different intersections range from 1969 (Western) to 1996 (Thriller, Action, and Adventure).

#### ### Deviation:

The deviation values indicate how much the average rating of each subset deviates from the overall mean rating. The largest positive deviation is found in the Adventure and Children intersection (1.59), while the largest negative deviation is in the Action subset (-3.67).

#### ### Conclusion:

The UpSet plot reveals that the majority of movies in the dataset do not belong to the six specified genres. Among those that do, Thriller is the most common genre, while Western is the least common. The intersections show that movies often belong to only one genre, with fewer movies spanning multiple genres. The average ratings and number of watches vary significantly across different intersections, indicating diverse audience engagement and reception for movies in different genre combinations.

---

# Study

## Conditions

Included in the user prompt field in the order: Image, APD, example alttext, strong (user portion), pattern. When the strong prompt is excluded, the system prompt is left blank, and the strong prompt in the user section is replaced with "You have been tasked with writing an alt text to describe this Upset plot to users with visual impairments". Temperature is set to 0, max tokens 1000. For GPT, no penalties are used and top P is left at 1

## Example alttext prompt:

Here is an example of another alttext generated for a different plot:

"This is an UpSet plot that shows covid cases and their symptoms. A covid case can have multiple symptoms. The sets are covid symptoms. The items are covid cases. The intersections show how many covid cases have exactly the same symptoms. The plot shows intersections of 6 sets. All major intersections involve the set Fatigue, and Cough. The largest intersection is Anosmia, and Fatigue, with 281 elements. Other large intersections also involve Cough, Anosmia, and Fatigue. The intersection of all sets is present with 23 elements.

### Dataset Properties

The dataset contains 6 sets and 4340 elements, of which 6 sets are shown in the plot.

### Set Properties

The set sizes are diverging a lot, ranging from 148 to 1531. The largest set is Fatigue with 1531 elements, followed by Anosmia with 1051, Cough with 897, Fever with 363, Diarrhea with 350, and Shortness of Breath with 148.

### Intersection Properties

The plot is sorted by size in descending order. There are 32 non-empty intersections, all of which are shown in the plot. The largest 5 intersections are Anosmia, and Fatigue (281), Cough, Anosmia, and Fatigue (259), Just Fatigue (198), Cough, and Fatigue (179), and Just Anosmia (140).

### Statistical Information

The average intersection size is 55, and the median is 24. The 90th percentile is 179, and the 10th percentile is 7. The largest set, Fatigue, is present in 78.1% of all non-empty intersections. The smallest set, Shortness of Breath, is present in 34.4% of all non-empty intersections.

### Trend Analysis

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."



## Strong prompt:

System: You are Dr. Alexander Lex, PhD in data visualization and creator of the Upset plot.

User: You have been tasked with writing an alt text to describe this Upset plot to users with visual impairments. You have been provided with both an image of the plot and a JSON file containing each row of the plot in textual form. Using your decades of experience with interactive data analysis, generate a text description that has three key emphases: First, it should describe the conclusions that a sighted users would draw from the plot at a high semantic level. Describing the appearance of the plot and what the components mean is irrelevant and unuseful; assume that your reader already understands what an upset plot is and how it works. Secondly, focus on specifics. Do not resort to vague references or generalized conclusions. Instead, cite specific pieces of data visualized on the plot to support the high-level semantic deductions and conclusions you draw from your visual analysis. Finally, focus on accuracy. It is not acceptable for even a single number to be wrong in your text description. Make sure to double-check the numbers you include in your result to ensure that your overarching conclusions are accurate. Use the data file to check whether your visual conclusions are correct. While focusing on these three emphases, use your acumen and the power of interactive visual data analysis to make the conclusions supported by this plot accessible to low-vision users via a high-level, detailed, specific, and accurate text description.

## Accessible Processed Data (APD):

Here is the JSON representation of the plot:

*[JSON of the accessibleProcessedData field, exported from Upset2 JS. The field name "accessibleProcessedData" is renamed "rows" and wrapped with {} to make valid JSON for the ablation study; for best effort, the field accessibleProcessedData.values is renamed "rows" and only it & its contents are included and wrapped with {}. Attribute values are manually cleaned from the JSON and the "attributes" property of each row is replaced with a "deviation" property which contains only the deviation for that row ]*

## Pattern Prompt

To assist your analysis of this upset plot, you've compiled a list of patterns that you've seen across many upset plots, while keeping in mind that this plot may represent a new pattern that you haven't seen before. The patterns are as follows:

- Intersection Patterns: Are there many independent sets (intersections containing only 1 set). Are there many low-set intersections (intersections containing 2-3 sets), medium-degree set intersections (intersections containing 3- $n/2$  sets), or high order set intersections (intersections containing  $n/2$  to  $n$  sets).
- Intersection Size Patterns: what are the intersection sizes associated with the intersection patterns? Are the higher-order intersections (involving most sets) large or small? Are the low-set intersections large or small? We classify these into small, medium, large, and largest.
- Specific Attributes: Is the all-set intersection present? Is the empty intersection present? Are the all-set and empty intersections the largest, among the largest, or small?

- Set Sizes: Are set sizes roughly equal, diverging a lot, or moderately diverging?

While the plot may contain a pattern that you haven't seen before, you know that thinking about these types of patterns is a great starting point for reaching complex conclusions about the data represented by the plot.

## Image

*A PNG screenshot of the Upset Plot rendered in Upset2 JS online ([upset.multinet.app](https://upset.multinet.app)). Plot is cropped so that the leftmost extent is just past the matrix showing set intersections, topmost extent is just above the list of unincluded sets, rightmost extent is just past the deviation column (showing size, degree (#), and deviation columns, but not attribute columns), and bottommost extent includes all rows but excludes the footer.*

## Notes on codes

- Level 4 REQUIRES domain-specific or contextual knowledge
- Lists of numbers: each number is counted as a separate fact
- L3 needs to be easier to see w/ visual analysis & not just a raw fact from the plot ie trends that upset plots are designed to show
- Statements of extrema without some sort of pattern synthesis are L2. IE "there are low-size intersections" is L2 but "most high-degree intersections are low-size" is L3
- A sentence can be divided into independent facts when they're in clauses that don't rely on each other to make sense, ie the sentence could be divided into two sentences stating the same things & still make sense
  - Would you put it as multiple facts if it was bulleted out? If so, they're separate in the sentence
- Noting that low-order intersections exist or that some intersections are absent is an L2 fact
- If removing a word/clause doesn't change the meaning of the sentence/paragraph, don't code it
- When the text uses a fact twice in a sentence mostly for wording purposes (ie one of the statements could be removed without affecting the meaning/conclusion of the sentence), it's not a duplicate fact
- Any stat that is categorically wrong or off by +/- 10% is misleading
- A higher-order fact supported by lower-level facts (most commonly L3 conclusion supported by L2 data points from the plot) is split into different facts (that example is 1 L3 and 2 L2)
- Things inferred from the left-side matrix of the plot (ie "intersection A frequently appears with other intersection B", "B does not commonly intersect", etc) are L3
- The LLMs frequently confuse overall set size (which they do not have access to) with the size of degree-1 intersections for those sets. In situations where this occurs as a list of degree-1 intersection sizes, we agreed it's more useful to put this down as a single falsehood as long as the degree-1 intersection sizes are correct

L1

L2

L3

L4

False/Misleading

Duplicate

## Ablation

### All conditions

This UpSet plot visualizes the distribution of movie genres across a dataset, focusing on six main categories: Action, Adventure, Children, Thriller, War, and Western. The plot reveals several key insights about genre combinations and their prevalence:

1. Genre Exclusivity: The most striking feature is the large number of movies (2,569) that don't fall into any of these six categories, suggesting a significant portion of the dataset consists of other genres or unclassified films.
2. Single-Genre Dominance: Among the specified genres, single-genre films are the most common. Thrillers lead with 349 movies, followed by Action (218), Children's movies (160), War films (88), Adventure (66), and Westerns (52).
3. Genre Combinations: The most frequent multi-genre combinations are:
  - Thriller and Action (104 movies)
  - Action and Adventure (90 movies)
  - Adventure and Children (75 movies)
4. Complex Genre Mixing: There are several instances of movies combining three or more genres, but these are relatively rare:
  - Thriller, Action, and Adventure (23 movies)
  - War, Action, and Adventure (7 movies)
  - Action, Adventure, and Children (6 movies)
5. Rare Combinations: Some genre combinations are extremely uncommon, with only 1 or 2 movies each. For example:
  - Thriller, War, Action, and Children (1 movie)
  - Western and Children (2 movies)
  - War and Western (1 movie)
6. Genre Flexibility: Thriller and Action appear to be the most flexible genres, frequently combining with others. In contrast, Western and Children's movies tend to stand alone or have fewer combinations.
7. Set Sizes: The set sizes are moderately diverging, with Thriller being the largest individual set (349 movies) and Western the smallest (52 movies) among the specified genres.

8. Intersection Patterns: The plot shows a predominance of low-set intersections (1-2 genres), with fewer medium-degree intersections and very rare high-order intersections.
9. Deviation Analysis: Some combinations show interesting deviations from expected frequencies. For instance, Action-only movies have a large negative deviation (-3.67%), while the Action and Adventure combination has a positive deviation (1.59%), suggesting these genres pair well together more often than expected.

This analysis reveals a complex landscape of genre classifications, where single-genre films dominate but specific multi-genre combinations also play a significant role in the dataset's composition.

## No Image

Based on the provided JSON data for this UpSet plot, I can offer the following high-level semantic description focusing on specific insights and accurate data:

This UpSet plot reveals interesting patterns in movie genre combinations, focusing on Thriller, War, Western, Action, Adventure, and Children's movies. The data suggests that these genres often exist in isolation, with significant overlaps primarily occurring between action-oriented genres.

The most striking observation is the dominance of single-genre movies. The largest subset by far is movies that don't belong to any of these six genres, with 2,569 films. Among the specified genres, "Just Thriller" is the most common with 349 films, followed by "Just Action" with 218, and "Just Children's" with 160.

Interestingly, the most common genre combination is Thriller and Action, with 104 films. This pairing significantly outperforms other two-genre combinations, such as Action and Adventure (90 films) or Adventure and Children's (75 films). This suggests a strong affinity between Thriller and Action genres.

The data also reveals some unexpected patterns. For instance, there are more War films (88) than Adventure films (66) that don't overlap with other genres. This suggests that War films often stand alone in their category, while Adventure tends to blend with other genres.

High-order intersections (movies belonging to many of these genres simultaneously) are relatively rare. The largest such intersection is Thriller, Action, and Adventure, with only 23 films. This indicates that while some genres commonly pair (like Thriller and Action), it's unusual for a film to span across many of these specific genres.

Some genre combinations are notably absent or rare. For example, there's only one film that combines Thriller, Western, and Action, and only one that spans Thriller, War, Action, and Adventure. This suggests that certain genre combinations, particularly those involving Western or War with multiple other genres, are exceptionally uncommon.

The Children's genre shows interesting patterns. While there are 160 pure Children's films, it most commonly combines with Adventure (75 films) rather than Action (only 6 films). This suggests that when Children's films incorporate elements from other genres, they tend to lean more towards Adventure than Action.

In summary, this data reveals that while these genres often exist in isolation, there are specific, preferred combinations (notably Thriller and Action) when they do overlap. The rarity of high-order intersections suggests that movies tend to focus on one or two of these genres rather than attempting to span many simultaneously.

## No APD

Thank you for providing this UpSet plot for analysis. As the creator of UpSet plots, I'm excited to delve into the insights this visualization offers about movie genres. Let me provide a detailed description focusing on the high-level semantic conclusions, supported by specific data points.

This UpSet plot reveals interesting patterns in movie genre combinations:

1. Genre Dominance: Action is by far the most prevalent genre, appearing in the majority of significant intersections. This suggests that action elements are commonly incorporated across various movie types.
2. Genre Purity vs. Hybrids: The largest category by far (2569 movies) consists of films that don't fit into any of the specified genres, indicating a substantial "other" category. In contrast, pure genre films are relatively rare, with Adventure being the most common single-genre category at 349 movies.
3. Common Genre Combinations:
  - Action-Thriller is the most frequent two-genre combination (160 movies)
  - Action-Adventure-Thriller is the most common three-genre mix (90 movies)
  - Action-Adventure-Children-Fantasy is the largest four-genre combination (7 movies)
4. Genre Flexibility:
  - Action shows the most versatility, combining frequently with other genres
  - Western appears the least flexible, rarely combining with other genres
5. Complex Genre Mixing: While present, movies spanning 4 or more genres are relatively uncommon. The most complex combination (Action-Adventure-Children-Fantasy-Western) only accounts for 1 movie.
6. Unexpected Combinations: Some seemingly contradictory genre mixes exist, such as Children-Thriller (3 movies) and Western-Children (1 movie), highlighting niche or experimental filmmaking.

7. Genre Clustering: Adventure, Action, and Thriller frequently appear together, suggesting a common thematic grouping in modern filmmaking.

8. Deviations from Expected Patterns:

- The Action-Thriller combination shows a notable negative deviation (-1.85%), suggesting it's slightly less common than one might expect given the individual genre frequencies.

- The pure Adventure category shows a significant positive deviation (2.15%), indicating it's more common as a standalone genre than predicted.

This analysis reveals a complex landscape of genre interactions in filmmaking, with Action serving as a versatile backbone for many multi-genre productions, while also highlighting the significance of films that defy easy genre categorization.

## No Example Alttext

Based on my analysis of this UpSet plot, here's a high-level semantic description focusing on the key patterns and insights:

This UpSet plot reveals a dataset of movie genres with a strong emphasis on single-genre classifications and limited genre combinations. The data shows several notable patterns:

1. Dominance of single-genre movies: The largest category by far is movies not classified in any of the displayed genres (2,569 movies). Among the specific genres, single-genre classifications are most common, with Thriller (349), Action (218), and Children's (160) movies being the top three.
2. Limited genre combinations: While there are some genre combinations, they are generally much smaller in size compared to single-genre categories. The most common combinations are:
  - Thriller and Action (104 movies)
  - Action and Adventure (90 movies)
  - Adventure and Children's (75 movies)
3. Rarity of complex genre combinations: As the number of combined genres increases, the frequency dramatically decreases. For example:
  - Three-genre combinations like Thriller, Action, and Adventure only account for 23 movies.
  - Four-genre combinations are extremely rare, with only 1 movie each in categories like Thriller, War, Action, and Children.
4. Unexpected deviations: Some genre combinations show interesting deviations from expected frequencies:
  - Action and Adventure movies are more common than expected (+1.59% deviation)
  - Adventure and Children's movies are also more frequent than expected (+1.59% deviation)

- Conversely, single-genre Action (-3.67% deviation) and Adventure (-3.20% deviation) movies are less common than expected, suggesting these genres often appear in combination.
5. War and Western genres: These appear to be relatively niche categories, with War having 88 single-genre movies and Western having 52. They rarely combine with other genres, indicating they are often distinct categories.

In conclusion, this dataset suggests a movie classification system that tends to favor single-genre categorizations, with a particular emphasis on Thriller and Action movies. While some genre combinations exist, they are generally less common, and highly complex genre classifications are extremely rare. This could indicate either a preference for clear-cut genre definitions in the dataset's source or a tendency for movies to strongly align with a single genre in practice.

## No Strong Prompt

Here's an alt text description for this UpSet plot:

This UpSet plot visualizes the intersections of six movie genres: Thriller, War, Western, Action, Adventure, and Children. The plot shows how movies are categorized across these genres, with each intersection representing films that belong to specific combinations of genres.

Dataset Properties: The dataset contains 6 genre sets and 3877 movies in total.

Set Properties: The set sizes diverge significantly. Action is the largest individual genre with 349 movies, followed by Thriller (349), Children (160), War (88), Adventure (66), and Western (52).

Intersection Properties: The plot displays 28 non-empty intersections, sorted by size in descending order. The largest intersections are:

1. Movies not belonging to any of the six genres (2569)
2. Just Thriller (349)
3. Just Action (218)
4. Just Children (160)
5. Thriller and Action (104)

Statistical Information: The average intersection size (excluding the unclassified movies) is about 47, with a median of 6. The largest genre, Action, appears in 39% of all non-empty intersections, while the smallest genre, Western, appears in 21%.

Trend Analysis:

- There's a very large number of movies (2569) that don't belong to any of these six genres.
- Single-genre classifications (low-set intersections) are generally the largest, indicating that many movies fit into just one of these categories.

- Medium-degree intersections (2-3 genres) are present but smaller in size, showing some genre mixing.
- High-order intersections (4+ genres) are rare and very small, suggesting that movies rarely span across many of these genres simultaneously.
- The all-set intersection (movies belonging to all six genres) is not present.
- Action and Thriller have the most overlap, with 104 movies in both genres.
- Children's movies have limited overlap with other genres, mainly intersecting with Adventure.

This plot reveals that while these genres are distinct, there is some crossover, particularly between Action and Thriller. However, most movies in the dataset tend to fall into a single genre category among these six, if they belong to any of them at all.

## No Pattern Prompt

This UpSet plot visualizes the distribution of movie genres across a dataset, focusing on Action, Adventure, Children, Thriller, War, and Western categories. The plot reveals several key insights about genre combinations and prevalence:

1. Genre Exclusivity: The majority of movies (2,569) do not fall into any of these specific genres, suggesting a large portion of the dataset consists of other genres like Drama, Comedy, or Romance.
2. Single Genre Dominance: Among the featured genres, Thrillers are the most common single-genre category with 349 movies, followed by Action (218) and Children's movies (160).
3. Popular Combinations:
  - Thriller and Action is the most frequent two-genre combination (104 movies)
  - Action and Adventure pairing is also common (90 movies)
  - Adventure and Children's movies often overlap (75 movies)
4. Multi-genre Complexity: As genre combinations become more complex, their frequency generally decreases. For example:
  - Thriller, Action, and Adventure combination appears in 23 movies
  - Four-genre combinations are rare, with only 1 movie each for categories like "Thriller, War, Action, and Children" or "Thriller, War, Action, and Adventure"
5. Genre Relationships:
  - Action frequently combines with other genres, appearing in many of the top multi-genre categories
  - War movies, while not numerous on their own (88), often combine with Action (34 movies)
  - Western is the least frequent genre among those shown, with only 52 pure Western movies and few combinations
6. Unexpected Combinations: Some unusual genre pairings exist, such as War and Children (1 movie) or Western and Children (2 movies), highlighting the diversity of the dataset.



7. Data Distribution: The plot shows a long-tail distribution, with many rare genre combinations represented by only a few or even single movies, illustrating the variety and specificity in film categorization.

This analysis provides insights into genre trends, popular combinations, and the complexity of movie categorization within the dataset, emphasizing both common patterns and unique outliers in genre classification.

## Best Effort

### Claude, Movies

This UpSet plot visualizes the genre combinations of movies, focusing on six main genres: Action, Adventure, Children, Thriller, War, and Western. The plot reveals several interesting patterns in how these genres intersect:

1. Genre Independence: The most striking observation is the high number of movies that belong to only one genre. The largest subset by far is movies that don't fall into any of these six genres (2,569 movies). Among the single-genre subsets, Thriller (349 movies) and Action (218 movies) are the most common.
2. Limited Genre Mixing: There are relatively few movies that combine multiple genres from this set. The largest multi-genre combination is Thriller and Action with 104 movies, followed by Action and Adventure with 90 movies.
3. Genre Affinities: Some genres show stronger tendencies to combine:
  - Action frequently pairs with Thriller (104 movies) and Adventure (90 movies)
  - Adventure often combines with Children's movies (75 movies)
  - War movies sometimes mix with Action (34 movies)
4. Rare Combinations: Some genre combinations are quite rare:
  - Only 1 movie combines Thriller, War, Action, and Children
  - Only 1 movie combines Thriller, War, Action, and Adventure
  - Western rarely combines with other genres in this set
5. Genre Sizes: The individual genre sizes vary considerably:
  - Thriller is the largest with 349 movies
  - Action follows with 218 movies
  - Western is the smallest with only 52 movies
6. Complex Combinations: Higher-order combinations (3 or more genres) are generally small:
  - The largest 3-genre combination is Thriller, Action, and Adventure with 23 movies
  - 4-genre combinations are extremely rare, with only 1 movie each

7. Deviations: Some combinations show interesting deviations from expected frequencies:

- Action as a single genre is underrepresented (-3.67% deviation)
- Adventure as a single genre is also underrepresented (-3.20% deviation)
- Action and Adventure together are overrepresented (+1.59% deviation)

This analysis suggests that while these genres are often distinct, there are specific patterns of genre blending, particularly between Action, Thriller, and Adventure. The data also indicates that War and Western genres tend to remain more isolated from the others in this set.

## GPT, Movies

This UpSet plot visualizes the intersections of six movie genres: Action, Adventure, Children, Thriller, War, and Western. The dataset contains 2569 elements, with the largest subset being the "Unincluded" category, which contains movies that do not belong to any of the six genres, comprising 2569 elements.

### Set Sizes

The set sizes vary significantly:

- Thriller: 349
- Action: 218
- Children: 160
- Adventure: 66
- War: 88
- Western: 52

### Intersection Patterns

The plot shows a variety of intersection patterns, with a notable presence of both independent sets and low-degree intersections (involving 2-3 sets). There are fewer medium-degree and high-order intersections.

### Intersection Sizes

- The largest intersection is the "Unincluded" category with 2569 elements.
- The largest single-genre intersection is "Just Thriller" with 349 elements.
- The largest two-genre intersection is "Thriller and Action" with 104 elements.
- The largest three-genre intersection is "Thriller, Action, and Adventure" with 23 elements.
- The smallest intersections, involving combinations of three or more genres, generally have sizes ranging from 1 to 8 elements.

### Specific Attributes

- The all-set intersection (involving all six genres) is not present.
- The empty intersection (movies that do not belong to any of the six genres) is the largest subset.
- The largest intersections are primarily single-genre or two-genre combinations, with the largest being "Just Thriller" and "Thriller and Action."

### Statistical Information

- The average intersection size is relatively small, with many intersections containing fewer than 10 elements.
- The largest set, Thriller, is present in several significant intersections, indicating its common pairing with other genres.
- The smallest set, Western, has fewer intersections and is less frequently combined with other genres.

### ### Trend Analysis

- The intersection sizes peak with the "Unincluded" category and then drop significantly for genre-specific intersections.
- Single-genre intersections are generally larger than multi-genre intersections.
- Low-degree intersections (2-3 genres) are more common and larger than medium or high-degree intersections.

In summary, the plot reveals that most movies do not belong to any of the six genres, and among those that do, single-genre and low-degree intersections are the most common and largest. The Thriller genre is the most prevalent, often intersecting with other genres, while the Western genre is the least common.

## Claude, Anon Movies

This UpSet plot reveals interesting patterns in the relationships between six sets, labeled Set\_1 through Set\_3 and Set\_15 through Set\_17. Here are the key insights:

1. Set Dominance: Set\_1 and Set\_15 are the most prominent sets, appearing in many of the largest intersections. The largest intersection is the "Unincluded" category with 2569 elements, suggesting a significant portion of the data doesn't belong to any of the six sets.
2. Set Sizes: The set sizes vary considerably. Set\_1 (218 elements), Set\_15 (349 elements), and Set\_3 (160 elements) are the largest individual sets, while Set\_16 (88 elements) and Set\_17 (52 elements) are notably smaller.
3. Intersection Patterns:
  - Low-degree intersections dominate the larger sizes. For example, Set\_1 & Set\_15 (104 elements), Set\_1 & Set\_2 (90 elements), and Set\_2 & Set\_3 (75 elements) are among the largest intersections.
  - Medium-degree intersections (3 sets) are present but smaller, such as Set\_1 & Set\_2 & Set\_15 (23 elements).
  - High-degree intersections (4 or more sets) are rare and very small, with only two instances of 1 element each.
4. Unique Elements: Each set has a significant number of unique elements not shared with other sets, ranging from 66 (Set\_2) to 349 (Set\_15).
5. Set Relationships:

- Set\_1 and Set\_15 have the strongest relationship, with their intersection (104 elements) being larger than expected given their individual sizes.

- Set\_2 and Set\_3 also show a strong relationship (75 elements in intersection).

- Set\_17 has relatively weak connections with other sets, its largest intersection being with Set\_1 (8 elements).

6. Deviations: Some intersections show significant deviations from expected values. For instance, Set\_1 & Set\_15 and Set\_1 & Set\_2 show positive deviations, indicating stronger than expected relationships.

7. Complexity: While there are many intersections, most elements belong to either no set or just one set, with decreasing frequency as the number of sets in the intersection increases.

This plot suggests a system where elements tend to belong to one or two sets at most, with rare instances of elements belonging to many sets simultaneously. The relationships between Set\_1, Set\_2, Set\_3, and Set\_15 appear to be the most significant in this dataset.

## GPT, Anon Movies

This UpSet plot visualizes the intersections of 6 sets: Set\_1, Set\_2, Set\_3, Set\_15, Set\_16, and Set\_17. The dataset contains 2569 elements, with the largest subset being the "Unincluded" set, which contains 2569 elements that do not belong to any of the 6 sets.

### Set Sizes:

The sizes of the sets vary significantly:

- Set\_1: 218 elements
- Set\_2: 66 elements
- Set\_3: 160 elements
- Set\_15: 349 elements
- Set\_16: 88 elements
- Set\_17: 52 elements

### Intersection Patterns:

- \*\*Independent Sets:\*\* There are several independent sets (intersections containing only 1 set), with the largest being Set\_15 (349 elements), followed by Set\_1 (218 elements), Set\_3 (160 elements), Set\_16 (88 elements), Set\_2 (66 elements), and Set\_17 (52 elements).

- \*\*Low-Set Intersections:\*\* There are multiple low-set intersections (intersections containing 2-3 sets). The largest low-set intersection is Set\_1 and Set\_15 with 104 elements, followed by Set\_1 and Set\_2 with 90 elements, and Set\_2 and Set\_3 with 75 elements.

- \*\*Medium-Degree Intersections:\*\* There are fewer medium-degree intersections (intersections containing 3 sets). The largest medium-degree intersection is Set\_1, Set\_2, and Set\_15 with 23 elements.

- \*\*High-Order Intersections:\*\* There are no high-order intersections (intersections containing 4 or more sets) with significant sizes.

#### Intersection Size Patterns:

- \*\*Largest Intersection:\*\* The largest intersection is the "Unincluded" set with 2569 elements.
- \*\*Large Intersections:\*\* The largest intersections involving the sets are Set\_15 (349 elements) and Set\_1 (218 elements).
- \*\*Medium Intersections:\*\* Medium-sized intersections include Set\_1 and Set\_15 (104 elements), Set\_1 and Set\_2 (90 elements), and Set\_2 and Set\_3 (75 elements).
- \*\*Small Intersections:\*\* Smaller intersections include Set\_1, Set\_2, and Set\_15 (23 elements), Set\_1 and Set\_16 (34 elements), and Set\_1 and Set\_17 (8 elements).

#### Specific Attributes:

- \*\*All-Set Intersection:\*\* There is no intersection that includes all 6 sets.
- \*\*Empty Intersection:\*\* The "Unincluded" set is the largest, indicating a significant number of elements do not belong to any of the 6 sets.

#### Statistical Information:

- \*\*Deviation:\*\* The deviations for the intersections vary, with the largest positive deviation being for Set\_1 and Set\_2 (1.5880190573031079) and the largest negative deviation being for Set\_1 (-3.669159671059875).

#### Trend Analysis:

- The plot shows a significant number of elements that do not belong to any of the sets, indicating a large "Unincluded" set.
- The independent sets have varying sizes, with Set\_15 being the largest and Set\_17 being the smallest.
- The low-set intersections are more common and larger in size compared to medium-degree and high-order intersections.
- The deviations indicate that some intersections are more or less common than expected, with Set\_1 and Set\_2 being more common and Set\_1 being less common than expected.

In summary, this UpSet plot reveals a dataset with a large number of elements not belonging to any of the 6 sets, significant variation in set sizes, and a predominance of low-set intersections. The deviations highlight specific intersections that are more or less common than expected, providing insights into the relationships between the sets.

Claude, World Organizations

This UpSet plot visualizes the membership of countries in various international organizations, focusing on UNESCO, Interpol, WHO, UN, WTO, and NATO. The plot reveals several key insights about the relationships between these organizations:

1. High degree of overlap among major international bodies:

The largest intersection, comprising 143 countries, belongs to UNESCO, Interpol, WHO, UN, and WTO simultaneously. This indicates a strong core of nations participating in multiple major international organizations, excluding NATO.

2. NATO's distinct membership:

The second-largest intersection (32 countries) includes all six organizations, including NATO. This suggests that while NATO has significant overlap with other major bodies, it remains a more exclusive organization with a distinct membership base.

3. Limited exclusive membership:

Very few countries belong to only one organization. For instance, only 7 countries are members of Interpol exclusively, 3 for UNESCO, and just 1 each for NATO and WTO. This underscores the tendency for nations to participate in multiple international bodies rather than isolating themselves to a single organization.

4. UN, WHO, and UNESCO core group:

There's a notable intersection of 5 countries that are members of UNESCO, WHO, and UN, but not the others. This could represent a group of nations focused on cultural, health, and general international cooperation without engaging in specific law enforcement (Interpol) or trade (WTO) organizations.

5. Unaffiliated countries:

Interestingly, there are 22 countries that do not belong to any of these six major international organizations. This represents a significant number of nations operating outside these global cooperative frameworks.

6. Interpol's broad reach:

Interpol appears in many of the larger intersections, suggesting it has a wide-reaching membership that often overlaps with other major international bodies.

7. WTO's integration:

The WTO is present in the largest intersections, indicating that countries involved in major international organizations tend to also participate in global trade frameworks.

8. Complexity of international relations:

The presence of numerous small intersections (e.g., 2 countries in UNESCO & WHO only, 2 in UNESCO, Interpol & WTO only) highlights the complex and varied nature of countries' participation in international organizations, likely reflecting diverse national priorities and geopolitical considerations.

This analysis reveals a complex web of international cooperation, with a strong core of nations participating across multiple organizations, while also showing the nuanced ways countries engage with global bodies based on their specific needs and priorities.

## GPT, World Organizations

This UpSet plot visualizes the membership of countries in various international organizations, specifically NATO, WTO, UN, WHO, Interpol, and UNESCO. The plot reveals several key insights about the intersection of these memberships:

- \*\*Dominant Intersection\*\*:** The largest intersection, with 143 countries, includes membership in UNESCO, Interpol, WHO, UN, and WTO, but not NATO. This indicates that a significant number of countries are part of these five organizations, suggesting a commonality in international cooperation among these entities.
- \*\*NATO Membership\*\*:** The second-largest intersection, with 32 countries, includes all six organizations: UNESCO, Interpol, WHO, UN, WTO, and NATO. This shows that a smaller subset of countries is part of NATO in addition to the other five organizations, highlighting NATO's more exclusive membership.
- \*\*Non-Members\*\*:** There are 22 countries that are not members of any of the six organizations listed. This subset is the third largest, indicating that a notable number of countries do not participate in these major international organizations.
- \*\*Other Significant Intersections\*\*:**
  - Nine countries are members of UNESCO, Interpol, WHO, and UN, but not WTO or NATO.
  - Seven countries are members of only Interpol.
  - Five countries are members of UNESCO, WHO, and UN, but not Interpol, WTO, or NATO.
  - Four countries are members of both UNESCO and Interpol, but not WHO, UN, WTO, or NATO.
- \*\*Smaller Intersections\*\*:** There are several smaller intersections, each with one to three countries, showing various combinations of memberships. For example:
  - Two countries are members of UNESCO and WHO.
  - Two countries are members of UNESCO, WHO, UN, and WTO.
  - One country is a member of just NATO.
  - One country is a member of just WTO.
  - One country is a member of Interpol, UN, and WTO.
  - One country is a member of Interpol, WHO, and WTO.
  - One country is a member of Interpol, WHO, UN, and WTO.

6. **Set Sizes**: The set sizes for UNESCO, Interpol, WHO, UN, and WTO are relatively large and similar, indicating that these organizations have broad international participation. NATO has a smaller membership size compared to the other five organizations.

7. **Intersection Patterns**: The plot shows a mix of low-set intersections (involving 2-3 sets) and medium-degree set intersections (involving 4-5 sets). The highest-order intersection (involving all six sets) is relatively large, while the low-set intersections vary in size, with some being quite small.

In summary, the plot highlights that UNESCO, Interpol, WHO, UN, and WTO have widespread international membership, with a significant number of countries participating in all five. NATO's membership is more exclusive, with fewer countries being part of all six organizations. Additionally, there is a notable number of countries that do not participate in any of these major international organizations.

## Control, Anon Movies

### # UpSet Introduction

This is an UpSet plot which shows the intersections of 6 sets. To learn about UpSet plots, visit <https://upset.app>. All major intersections involve the sets Set-4, Set-1, and Set-2. The largest non-empty intersection is between Set-1 and Set-2, with 101 elements. Other large intersections also involve Set-4 and Set-1.

### # 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 127 to 1200. The largest set is Set-4 with 1200 elements, followed by Set-1 with 503, Set-2 with 283, Set-3 with 251, Set-5 with 211, and Set-6 with 127.

### # Intersection Properties

The plot is sorted by size in descending order. There are 24 non-empty intersections, all of which are shown in the plot. The largest 5 intersections are the empty intersection (1777); just Set-4 (1003); just Set-1 (283); just Set-5 (131); and just Set-6 (123).

### # Statistical Information

The average intersection size is 161, and the median is 31. The 90th percentile is 283, and the 10th percentile is 3. The largest set, Set-4, is present in 45.8% of all non-empty intersections. The smallest set, Set-6, is present in 8.3% of all non-empty intersections.

### # Trend Analysis



The intersection sizes peak at a value of 1777 and then drastically flatten down to 1. The empty intersection is present with a size of 1777. An all set intersection is not present. The individual set intersections are large in size. The low degree set intersections lie in medium sized intersections. The medium degree set intersections can be seen among small sized intersections. No high order intersections are present.

## Control, World Orgs

### # UpSet Introduction

This is an UpSet plot which shows the intersections of 6 sets. To learn about UpSet plots, visit <https://upset.app>. All major intersections involve the set UN. The largest non-empty intersection is between UNESCO, Interpol, WHO, UN, and WTO, with 143 countries. Other large intersections also involve Interpol, WHO, UN, NATO, UNESCO, and WTO. The intersection of all sets is the second largest with 32 elements.

### # Dataset Properties

The dataset shows attributes of world organization membership by country. The dataset contains 18 sets and 1745 elements, of which 6 sets are shown in the plot.

### # Set Properties

The set sizes are diverging a lot, ranging from 33 to 202. The largest set is UNESCO with 202 countries, followed by Interpol with 200, WHO with 195, UN with 193, WTO with 183, and NATO with 33.

### # Intersection Properties

The plot is sorted by size in descending order. There are 16 non-empty intersections, all of which are shown in the plot. The largest 5 intersections are UNESCO, Interpol, WHO, UN, and WTO (143); UNESCO, Interpol, WHO, UN, WTO, and NATO (32); the empty intersection (22); UNESCO, Interpol, WHO, and UN (9); and just Interpol (7).

### # Statistical Information

The average intersection size is 14, and the median is 2. The 90th percentile is 32, and the 10th percentile is 1. The largest set, UNESCO, is present in 56.2% of all non-empty intersections. The smallest set, NATO, is present in 12.5% of all non-empty intersections.

### # Trend Analysis

The intersection sizes peak at a value of 143 and then drastically flatten down to 1. Unesco, Interpol, WHO, UN, and WTO is the largest by a factor of 4. The empty intersection is present with a size of 22. An all set intersection is not present. The individual set intersections are small in size. The low degree set intersections lie in medium sized intersections. The medium degree set intersections can be seen among small and medium sized intersections. In large and largest sized intersections, the high order set intersections are significantly present.

## Control, Movies

### # UpSet Introduction

This is an UpSet plot which shows the intersections of 6 sets. To learn about UpSet plots, visit <https://upset.app>. All major intersections involve the set Action. The largest non-empty intersection is between Drama and Comedy, with 208 movies. Other large intersections also involve Action and Comedy.

### # Dataset Properties

The dataset shows attributes of movie genre overlap. 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 143 to 1603. The largest set is Drama with 1603 movies, followed by Comedy with 1200, Action with 503, Adventure with 283, Children with 251, and War with 143.

### # Intersection Properties

The plot is sorted by deviation in descending order. There are 38 non-empty intersections, all of which are shown in the plot. The largest 5 intersections are just Drama (1191); just Comedy (809); the empty intersection (762); just Action (220); and Drama and Comedy (208).

### # Statistical Information

The average intersection size is 102, and the median is 13. The 90th percentile is 220, and the 10th percentile is 1. The largest set, Drama, is present in 36.8% of all non-empty intersections. The smallest set, War, is present in 34.2% of all non-empty intersections.

### # Trend Analysis

The empty intersection is present with a size of 762. An all set intersection is not present. The individual set intersections are in large and largest intersections. The low degree set intersections lie in medium sized intersections. The medium degree set intersections can be seen among small and medium sized intersections. No high order intersections are present.

