Analyzing Online Sentiment Towards Marvel TV Shows

Rohan Mehta

Disney+ has now released seven Marvel TV shows. The episodic format of these most recent additions to the MCU has enabled Marvel to pursue more experimental styles, evidenced by series such as WandaVision, Moon Knight, and Ms. Marvel. As such, it is not uncommon for these shows to be met with a broad range of opinions. However with the release of Ms. Marvel, it seems that these viewpoints have become more polarized than ever, eliciting widespread reports of review-bombing [1, 2, 3, 4, 5, 6, 7, 8, 9]. Are these claims true? Does the distribution of user ratings for Ms. Marvel differ significantly from those of other shows? Does it differ significantly from the distribution of ratings according to critics? And what information can we glean about why it received the ratings it did – from both critics and users – by studying the language that features most prominently in those reviews which rated it favorably compared to those which did not?

Downloading and cleaning the data

I surveyed many sites for reviewing movies and TV shows including IMDB, Rotten Tomatoes, and Metacritic. Of all of these Metacritic appeared to be the easiest to scrape, since it had already split up reviews from critics and users into separate pages, and had a consistent structure for dividing reviews. After manually recording the URLs for the user and critic reviews of the seven Marvel TV shows released so far, the data was then imported, cleaned, and formatted. Note that although "users" is a somewhat uncomfortable term to use in this setting, we will just take it to mean all people who are not critics.

In[•]:=
```
marvelShows = {…};
```

In[•]:=
```
reviewLinks = {…};
```

In[•]:=
```
criticReviews = StringSplit[#, "Critic score Publication By date"] & /@
    Import /@ (# <> "/critic-reviews" & /@ reviewLinks);
```

In[•]:=
```
userReviews = StringSplit[#, "User score By date Most helpful"] & /@
    Import /@ ( # <> "/user-reviews" & /@ reviewLinks);
```

In[•]:=
```
allReviews = AssociationThread[
    marvelShows → Map[#[[2]] &, MapThread[List, {criticReviews, userReviews}], {2}]];
```

In[•]:=
```
allReviews = {StringSplit[#[[1]], "Read full review"],
      StringSplit[#[[2]], "All this user's reviews"]} & /@ allReviews;
```

In[ ]:=
```
allReviews = Map[Drop[#, -1] &, allReviews, {2}];
```

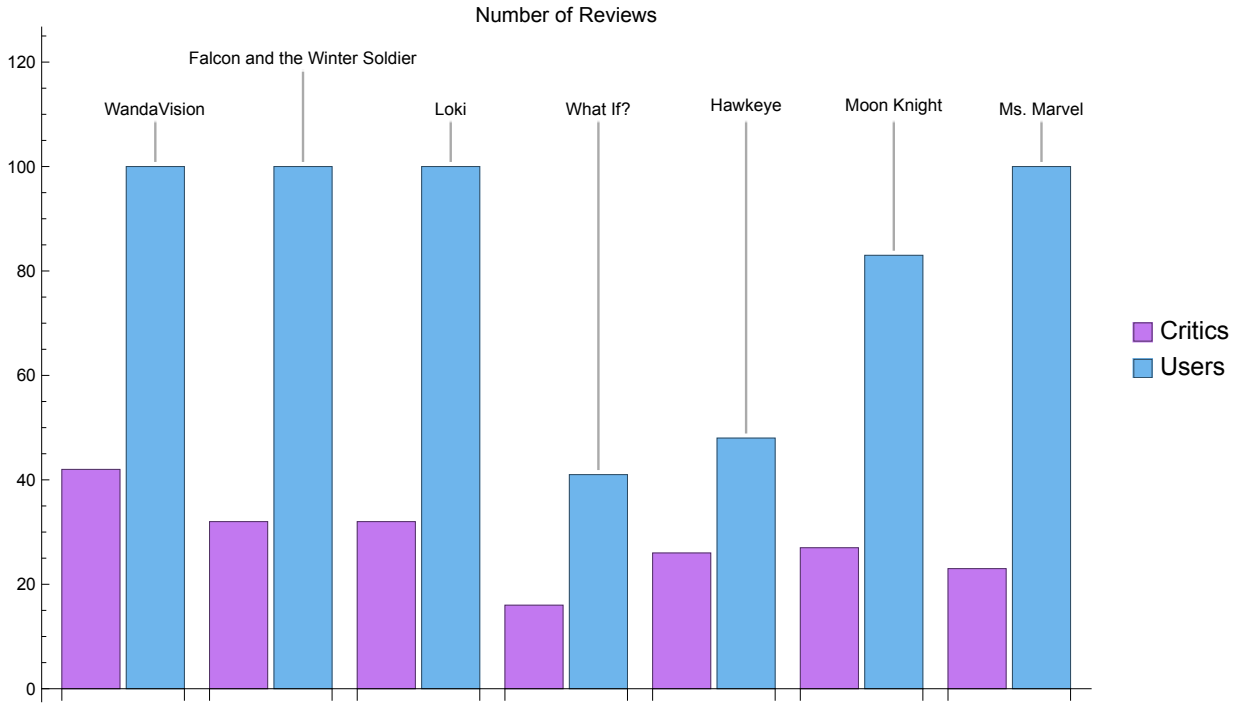
How many reviews did each show get?

One of the easiest metrics to compute is how many reviews exist for each show. As may be expected, the oldest shows (all released early last year) have the greatest number of user reviews (~100), since they have had the most time to accrue users. They also have a slightly greater number of critic reviews, though this difference is not nearly as great as the former (which also makes sense, as critics review all shows almost immediately, since it is their job).

The only discrepancy in this pattern are the shows What If? and Hawkeye, which despite being released well before both Moon Knight and Ms. Marvel, received far less reviews from both users and critics. This is explained by the fact that both shows did not generate much interest (plus, with What If? being an animated show, it likely turned off many critics as well) and were generally poorly received (we will soon see that critics ranked them lowest of all seven shows).

In[ ]:=
```
reviewCount = Map[Length, allReviews, {2}];

BarChart[reviewCount, ChartLegends → {"Critics", "Users"}, ChartStyle → "Pastel",
 PlotLabel → "Number of Reviews", ImageSize → Large, LabelingFunction →
  (If[EvenQ[Last[#2]], Callout[Keys[allReviews][[First[#2]]], None] &)]
```

Out[ ]=



We can also determine what percentage of each show's total reviews came from users versus critics. This metric does not turn out to be particularly interesting, as the breakdown is quite similar for all shows. However, we may consider that it gives us a rough heuristic for how strong an experience the

show evoked in users. All users who posted their review for a given show had some motivation to do so, likely due to a strong reaction they had while watching the show. So the show whose percentage of reviews due to users is the greatest could be interpreted as the most provocative.

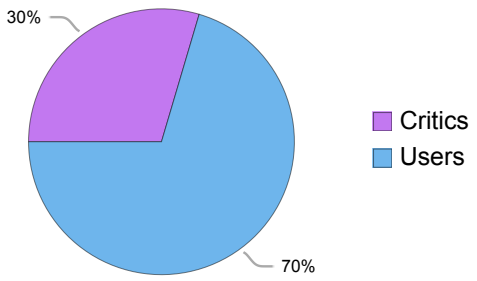
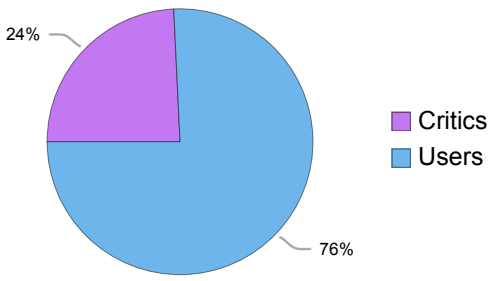
With 81% of its total reviews coming for users, Ms. Marvel technically earns this title, although both Loki and Falcon and the Winter Soldier are right on its heels at 76%, and nearly ever other show is only a few percentage points removed (except Hawkeye, which dips below 70% for this metric). This difference isn't large enough for us to make any useful distinctions save between Hawkeye and the other TV shows (and that's hardly something we need statistics to do, anyway!)

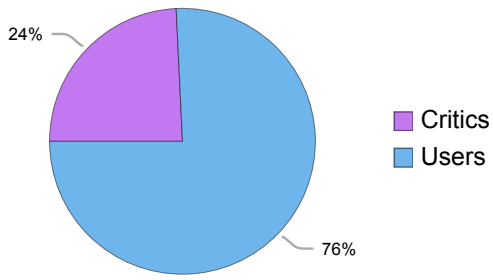
In[ ]:=
```
reviewBreakdown[reviewCount_] :=
 Row@{Spacer[50], PieChart[reviewCount, ...]}
```

In[ ]:=
```
{titles, charts} = {Keys@reviewCount, reviewBreakdown /@ Values@reviewCount};
```

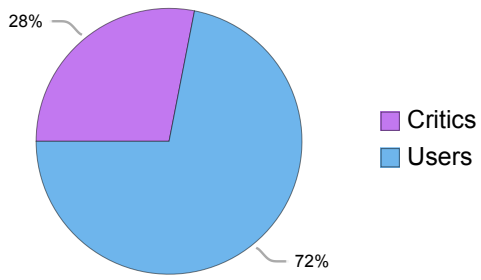
In[ ]:=
```
Grid[Flatten[Transpose[Partition[#, UpTo[2]] & /@ {titles, charts}], 1], Frame → All]
```

Out[ ]=

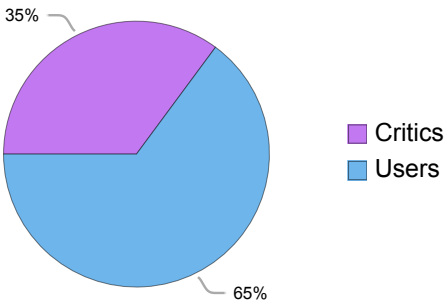
<table>
<tr><td>WandaVision</td><td>Falcon and the Winter Soldier</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>Loki</td><td>What If?</td></tr>
<tr><td></td><td></td></tr>
</table>



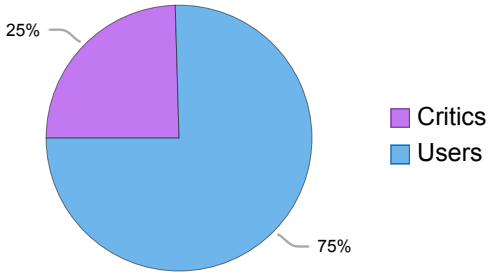
Hawkeye

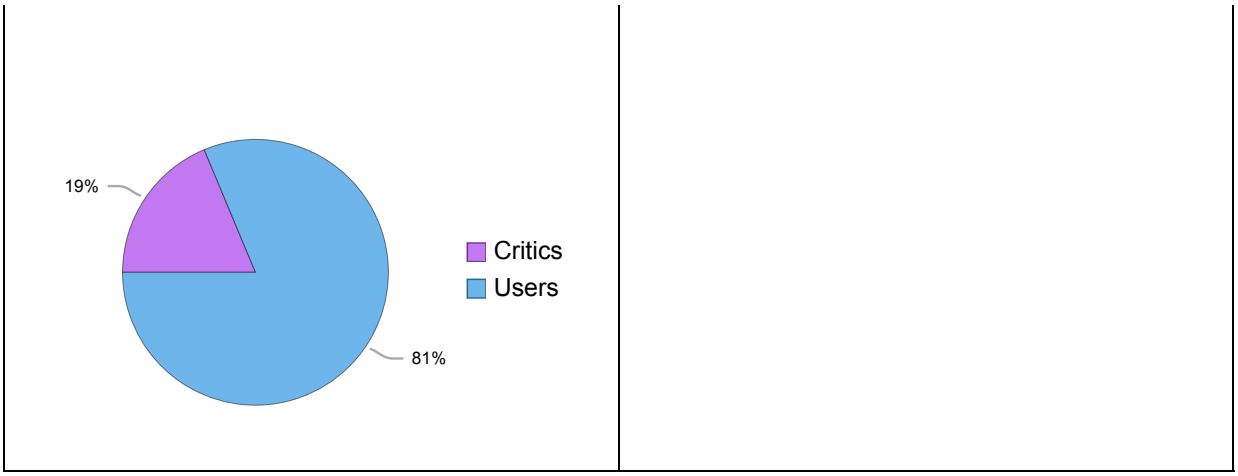


Moon Knight



Ms. Marvel





Finally, by looking up the release dates for all seven shows, we can also plot the quantity of user and critic reviews for each show in chronological order. This corroborates the trend we saw earlier in the bar graph, although it does reveal much more clearly that the number of critic reviews peaked for WandaVision, Marvel's first TV show release. It also demonstrates that the user interest generated by Ms. Marvel – as measured by the number of user reviews – seems to have returned to where it was for Marvel's first three TV releases, after suffering a steep decline with What If?

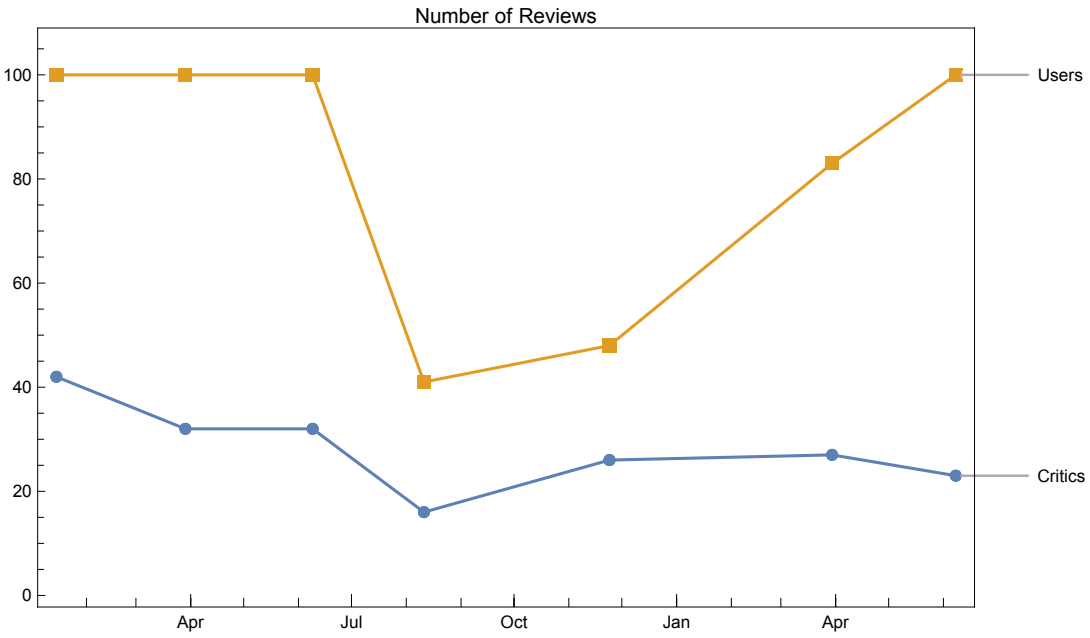
In[ ]:=
```
releaseDates = <| ··· |>;
```

In[ ]:=
```
userCount = {releaseDates[#], reviewCount[#][[1]]} & /@ Keys@reviewCount;
```

In[ ]:=
```
criticCount = {releaseDates[#], reviewCount[#][[2]]} & /@ Keys@reviewCount;
```

In[ ]:=
```
DateListPlot[{userCount, criticCount}, ···]
```

Out[ ]=

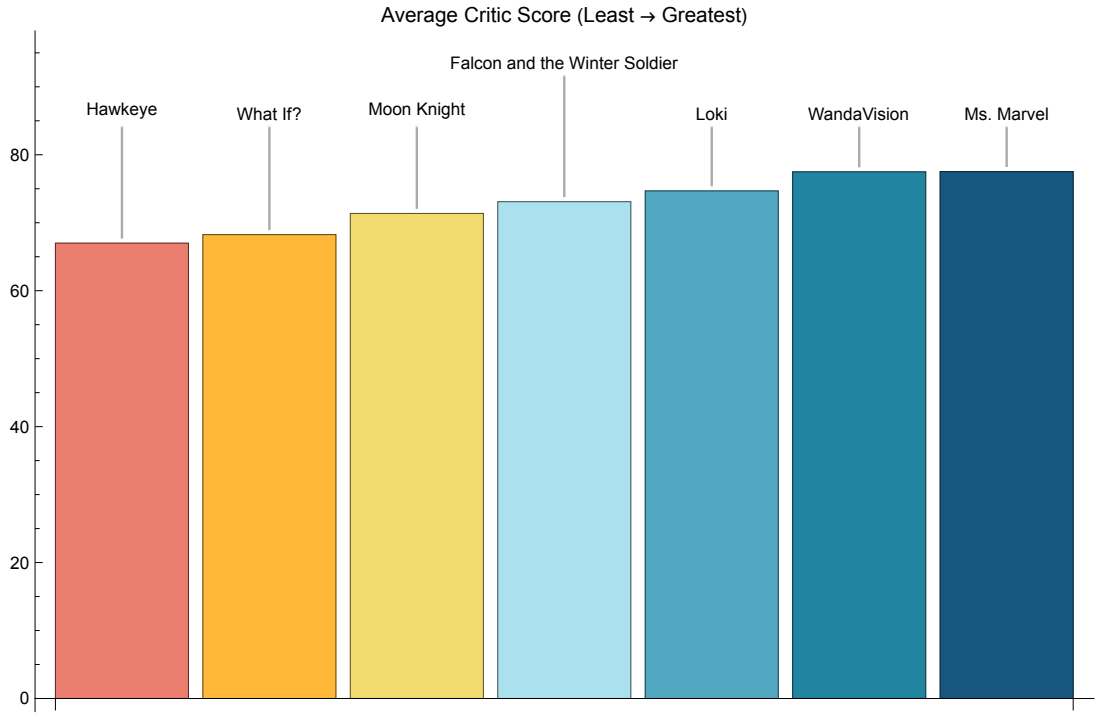


What was the average score for both critics and users?

Now we can actually look at the scores for each review! To do this, we need to extract all the numbers that appear in a given review, and then take the first one (since each review always begins with its score). It should be noted that critics' scores range from 0-100 while users' scores range from 0-10. When we visualize the scores from critics, we see that Ms. Marvel is rated the highest. This is in accordance with reports that it is the highest rated Marvel motion picture (i.e., movie or TV show) ever released [2, 3, 10]. WandaVision takes second place at just a tenth of a percentage point below Ms. Marvel, and Loki takes third place clocking in at a few percentage points below that. As stated before, What If? and Hawkeye rank last (in that order), in the high 60s. Overall though, everything is pretty close here, with the top five shows all within ten percentage points of each other.

```
In[•]:= extractScore[text_] := StringCases[text, "\n" ~~ x : DigitCharacter .. :> x];

In[•]:= scores = AssociationThread[{"Critics", "Users"} → #] & /@
    Map[ToExpression@*Flatten, Map[extractScore, allReviews, {3}], {2}];

In[•]:= scores = <|"Critics" → Select[#[[1]], LessEqualThan[100]],
      "Users" → Select[#[[2]], LessEqualThan[10]]|> & /@ scores;

In[•]:= avgScores = {N@Mean@#["Critics"], N@Mean@#["Users"]} & /@ scores;

In[•]:= BarChart[#[[1]] & /@ SortBy[avgScores, First], ···]
```

Out[•]=

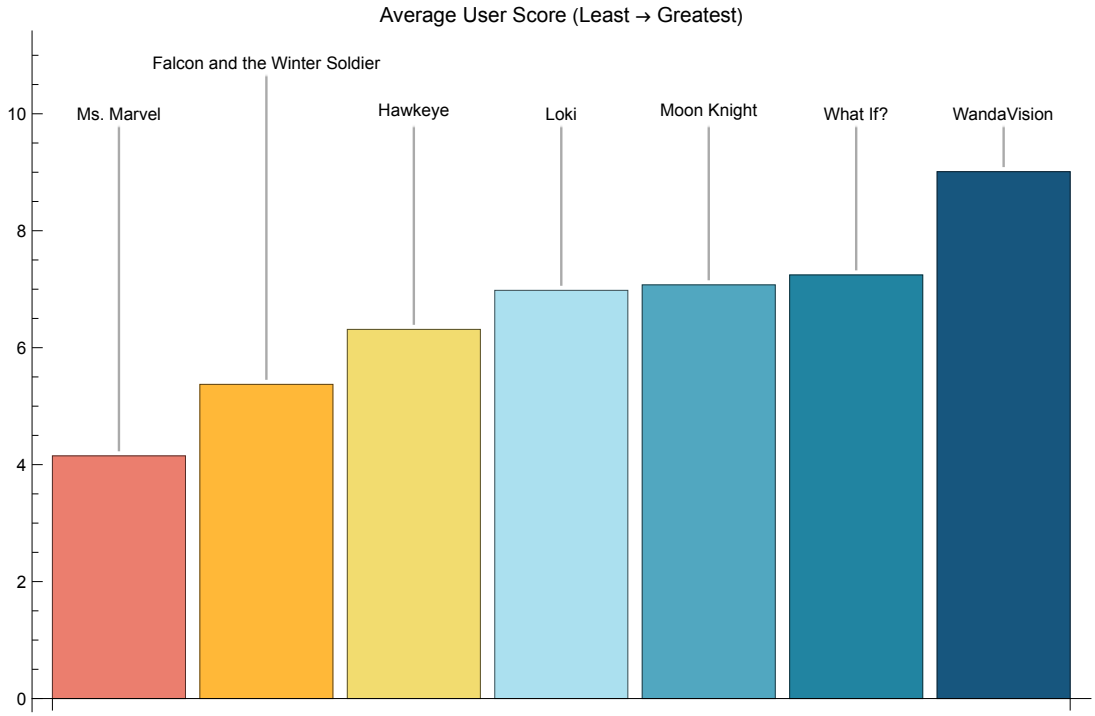


However, when we look at user ratings Ms. Marvel is neither first, second, or third! In fact, it is dead last (and funnily enough, has dropped even further during my most recent web scrape). WandaVision, with

whom it is neck in neck in terms of critic score, towers over it by a margin of fifty percentage points when it comes to user score. Before we try and investigate *why* this might be the case, you have to admit, it's pretty surprising. Usually we expect critics to be more particular and less forgiving than users, as fans place less of a premium on the artistry of a motion picture as long as it keeps them engaged. In fact, we can see how this works in the favor of shows like What If?, which moved all the way up to second place (from second to last) when comparing user score versus critic score. Ms. Marvel appears to be suffering from the reverse effect, where fans are judging it more harshly than critics – punishingly harsh, I would say. So we do have some initial evidence that review-bombing could be happening.

```
In[ ]:= BarChart[#[[2]] & /@ SortBy[avgScores, Last], ···]
```

Out[ ]=



If we actually plot the distributions of each show's user and critic scores as a histogram (and try to reconstruct a PDF from which this distribution might be generated, which we'll call a "smooth histogram"), this discrepancy becomes even more clear. First, let's note that the distribution of user scores for Ms. Marvel is extremely bimodal – more so than that of any other show – with 24 users giving it a perfect score, while 40 rank it somewhere in the 0-2 range. So it seems like our hypothesis that Ms. Marvel is the most polarizing Marvel TV show released to date is in fact correct. Some users really love the show, but for some reason, more hate it. This is much more convincing evidence of review-bombing. A good portion of users agree with critics' favorable estimation of the show, but another group is completely bucking this trend and reacting in the exact opposite way. Plus, the fact that this group is around twice the size of the group generating positive reviews is also notable. Review-bombers will leave as many reviews as they possibly can since their goal is making the show look bad.

Taking a look at the smooth histograms clears some things up as well. For nearly every other show, the major peaks of these curves more-or-less match up between the user and critic score distributions (Falcon and the Winter Soldier is somewhat of an exception), indicating that on average critics and users basically agree on how good the show was overall, even if their specific scores differ on a case-by-case basis. On the other hand, the smooth histogram for Ms. Marvel as generated by users' scores doesn't match up with the one generated by critics' scores at all. The critic distribution is heavily centered around 80%, with a peak whose height is far above that of anything else we see in the critic distributions of other shows. The user distribution, meanwhile, is almost centered at 0%! Moreover, whereas the critic distribution of most other shows observe a slow, sloping descent towards zero after peaking (indicating that a good portion of reviewers gave it a medium-ish score), Ms. Marvel's critic distribution exhibits an almost vertical, 90-degree drop (meaning that reviews with scores far below its peak at 80% are essentially nonexistent). This provides some more quantitative backing to our claim that it really is the best reviewed Marvel motion picture ever.

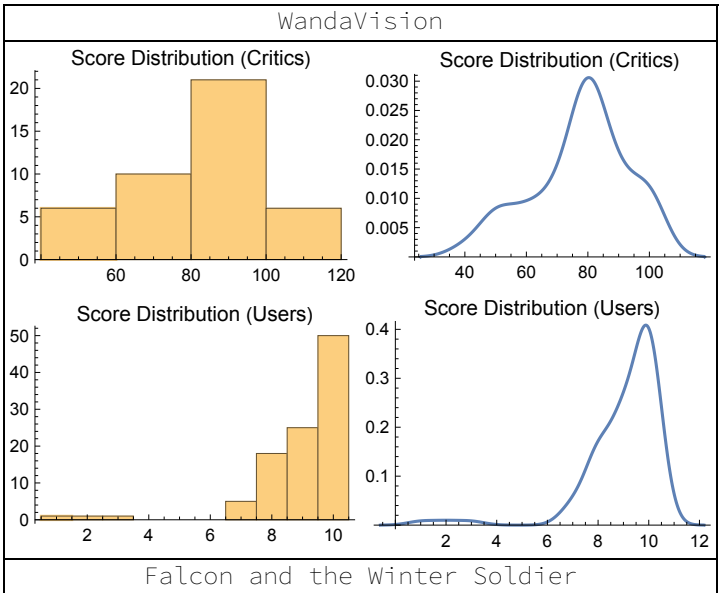
In[ ]:=
```
getPlots[plotFn_] := Module[{newScores, newAssoc},
  newScores = AssociationThread[{"Critics", "Users"} → #] & /@
    (KeyValueMap[(Append[#2, "Score Distribution (" <> ToString@#1 <> ")"] &),
       #] & /@ scores);
  newAssoc = Values /@ Map[plotFn[Drop[#, -1], ···] &, newScores, {2}];
  newAssoc]
```

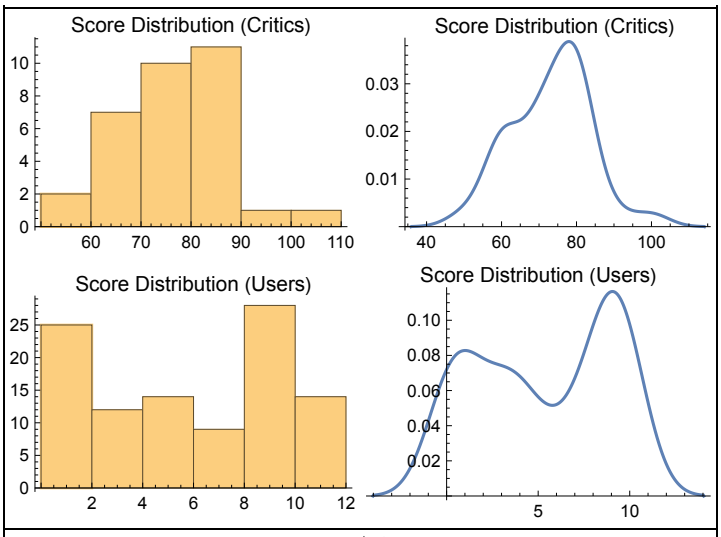
In[ ]:=
```
{hard, smooth} = getPlots /@ {Histogram, SmoothHistogram};
```

In[ ]:=
```
combined = Grid[{{#[[1]][[1]], #[[2]][[1]]}, {#[[1]][[2]], #[[2]][[2]]}}] & /@
   Merge[{hard, smooth}, Identity];
```

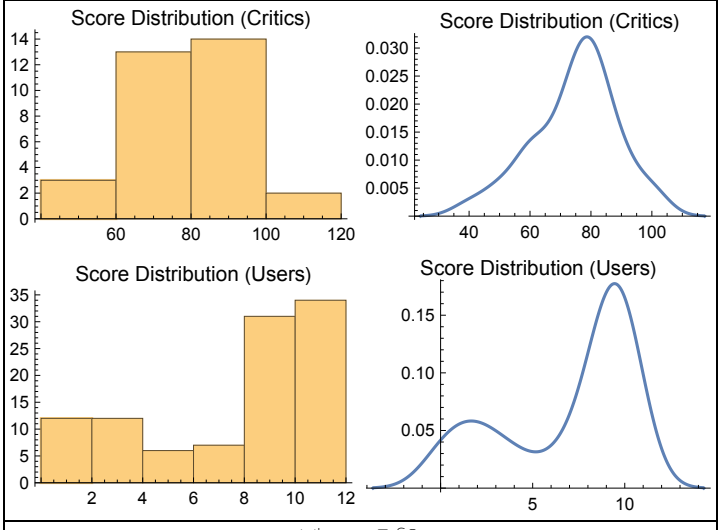
In[ ]:=
```
Grid[Flatten[MapThread[{{#2}, {#1}} &, {Values@combined, Keys@combined}], 1],
 Frame → All]
```

Out[ ]=

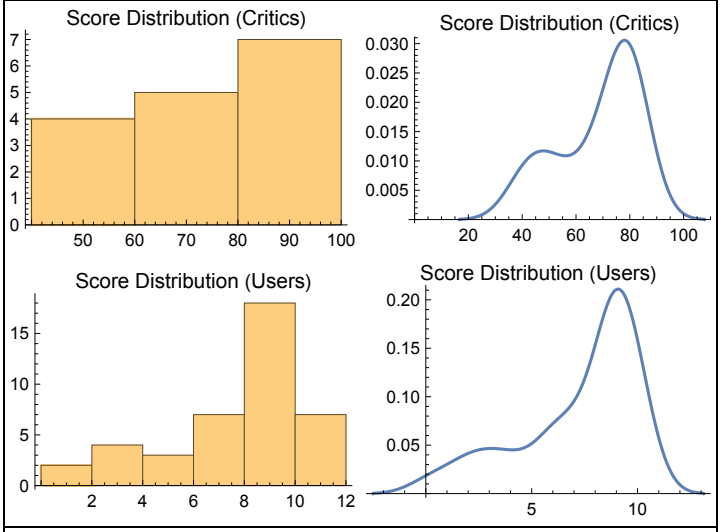




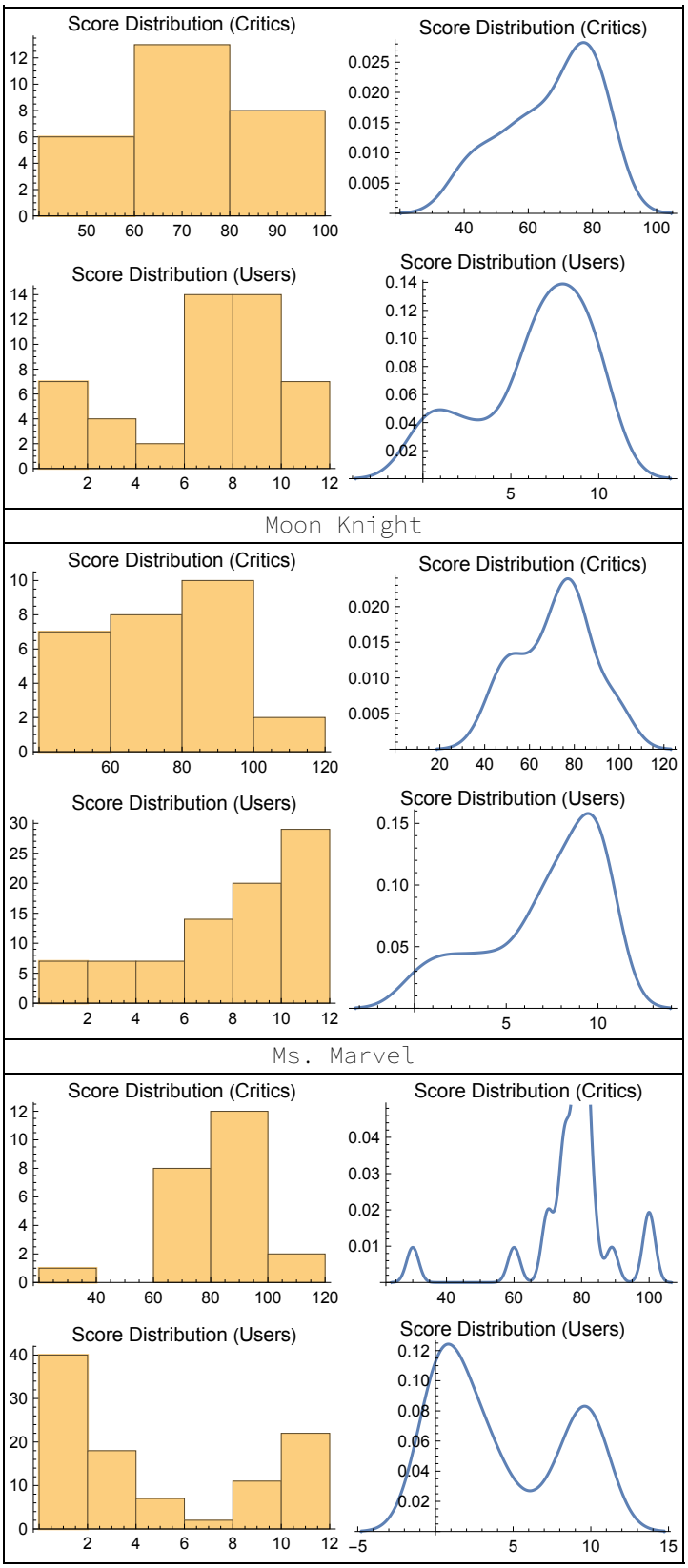
Loki



What If?



Hawkeye

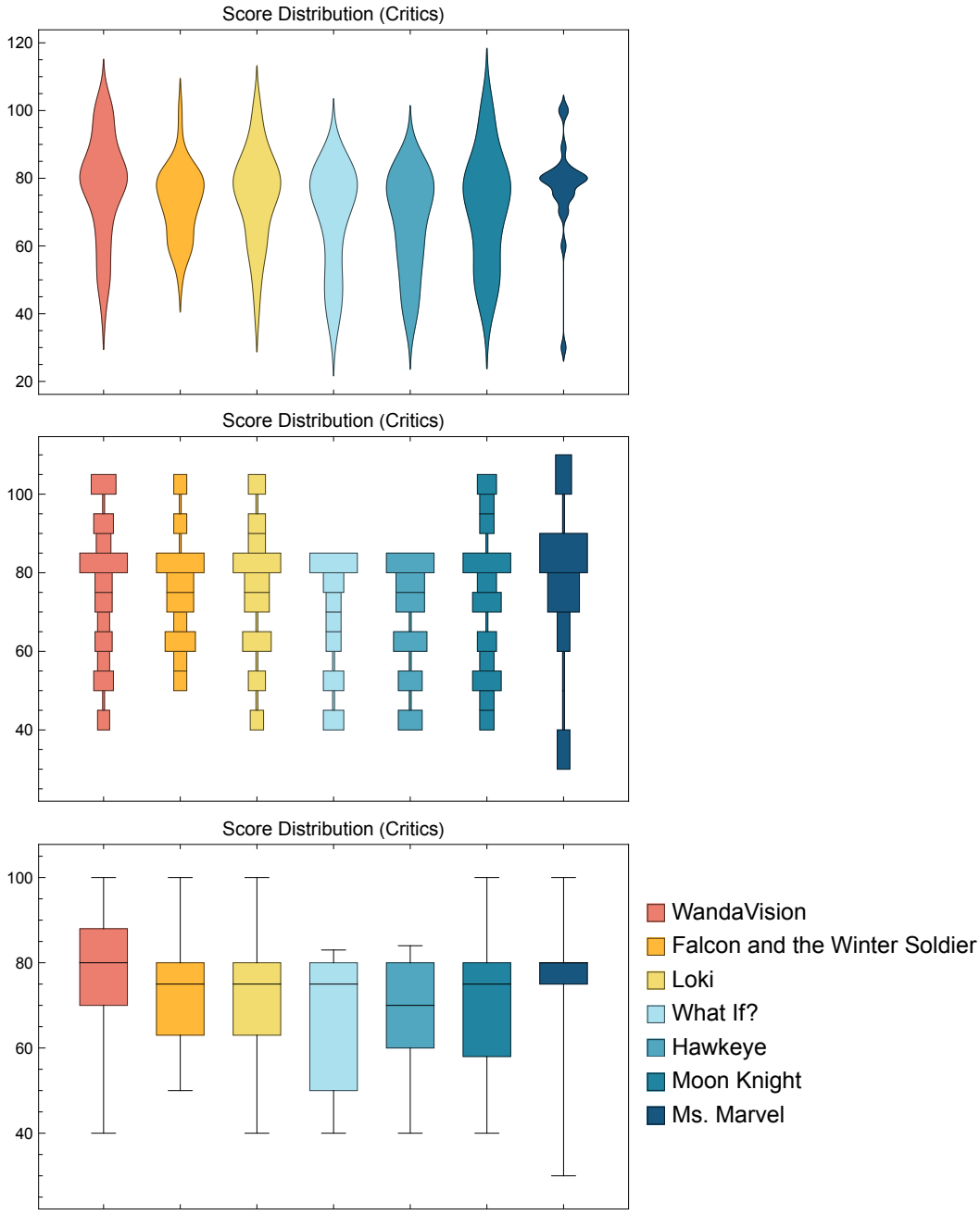


We can also look at density plots of each show's score distributions. Starting off with the critic distribution, notice just how unusual Ms. Marvel looks. Like we said before, it has essentially no reviews below

its peak, corresponding to an extremely, instantly thin tail (as opposed to the fatter, gradually thinning tails of the other shows' distributions). In terms of actual median score, it ties WandaVision for first place at 80%. If we re-render the plot by boxing together scores in a similar range, we see that Ms. Marvel is being weighed down by a single, extremely low critic score, whereas WandaVision is suffering the net effect of several, more mildly negative reviews. Rendering as a box-and-whisker chart reveals that the score weighing Ms. Marvel down – one that is more than 30 percentage points lower than its next lowest score – is a 30%, the lowest score given to any show by any critic. In fact, we can calculate that it (and every other "low" critic score Ms. Marvel receives) is a statistical outlier by a longshot (since they are below Q1 - 1.5 x IQR).

In[ ]:=
```
Column[Table[DistributionChart[Values@(#[[1]] & /@ scores), ···],
  {style, {Automatic, "HistogramDensity", "BoxWhisker"}}]]
```

Out[ ]=

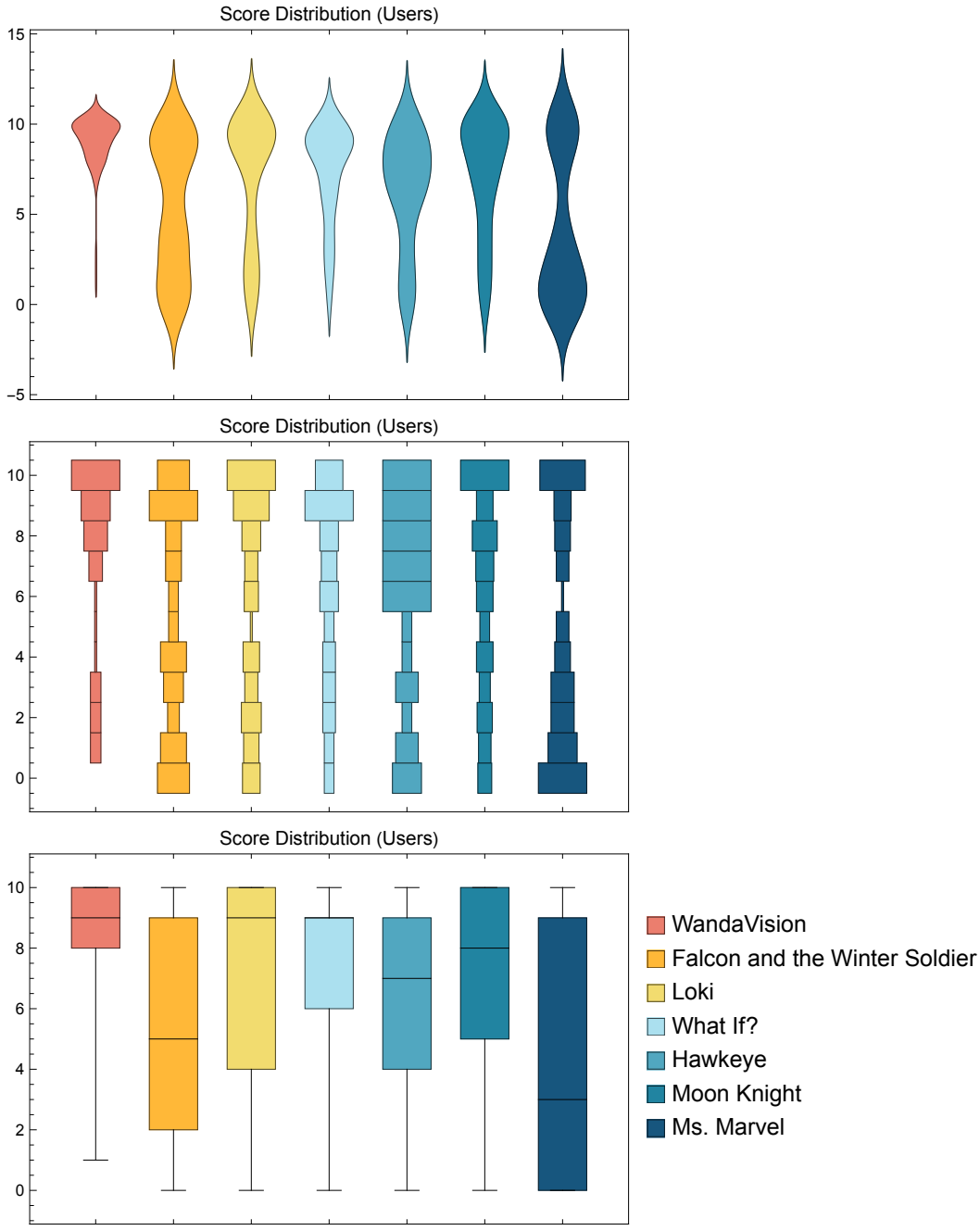


The density plots for the user distributions mostly approximate their corresponding critic distributions (just as we saw with the smooth histograms), save for WandaVision, Falcon and the Winter Soldier, and Ms. Marvel. Here, WandaVision takes on the instantly thinning tail Ms. Marvel exhibited in its critic distribution – a testament to just how well it was received by fans! It also boasts an impressive median score of 90%, compared to Ms. Marvel's dismal 30%. Unlike in its critic distribution. Ms. Marvel now has a fat, expanding tail. This is notable: unlike most other shows, where the probability of drawing a score decreases as it gets lower (hence the thinning tail), the probability of drawing a score from the Ms. Marvel user distribution actually increases the lower it is (up to a certain point, of course). When we

switch to a boxed perspective, we also see the emergence of a sort of “double pyramid”, where boxes are heavily stacked at the bottom (low score) and top (high score), but absent in the middle (medium score). This is just a discretized version of the same phenomenon, and further proves the show’s polarizing nature. By now, we can confidently say that the distribution of user scores for Ms. Marvel is being subjected to a tug-of-war between two extreme opinions, with the negative side mostly winning out. We should also note that the user distribution for Falcon and the Winter Soldier mirrors many of the same trends, although to a far lesser extent.

```
Column[Table[DistributionChart[Values@(#[[2]] & /@ scores), ···],
  {style, {Automatic, "HistogramDensity", "BoxWhisker"}}]]
```

Out[ ]=



Ok, I know I'm belaboring the point by now, but I want to explore one final question. How many perfect scores (100/100 or 10/10) and failing scores (0/100 or 0/10) did each show receive? WandaVision crushes all competition with a total of 50 perfect scores assigned by users, and 6 assigned by critics, the highest of any show. Loki, Moon Knight, and Ms. Marvel all follow up with 2 perfect scores assigned by critics, and 34, 39, and 24 perfect scores assigned by users, respectively.

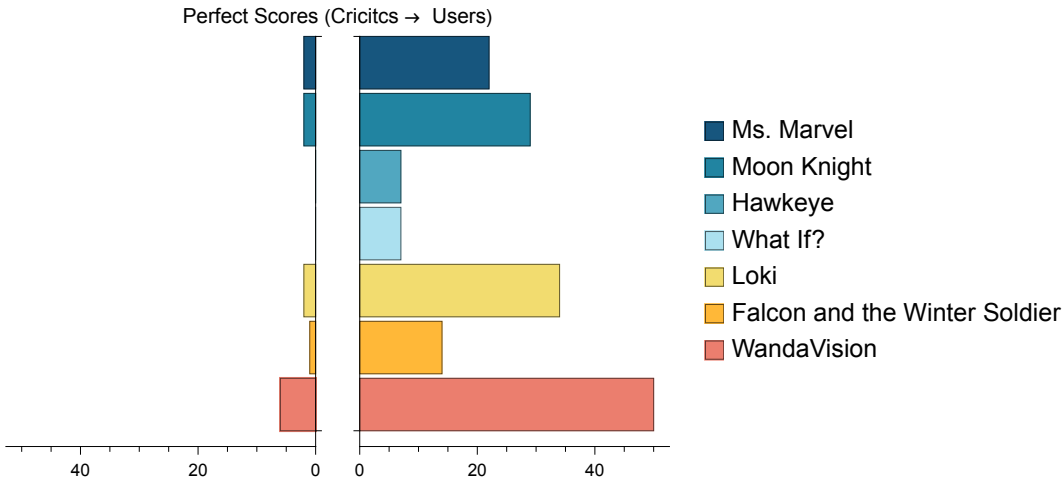
In[ ]:=
```
tabulatedScores = Map[{Counts@#[["Critics"]], Counts@#[["Users"]]} &, scores, {1}];
```

In[ ]:=
```
perfect = # /. Missing[__] → 0 & /@ {#[[1]][100], #[[2]][10]} & /@ tabulatedScores;
```

In[ ]:=
```
fails = # /. Missing[__] → 0 & /@ Map[Lookup[0], tabulatedScores, {2}];
```

In[ ]:=
```
PairedBarChart[First /@ perfect, Last /@ perfect, ···]
```

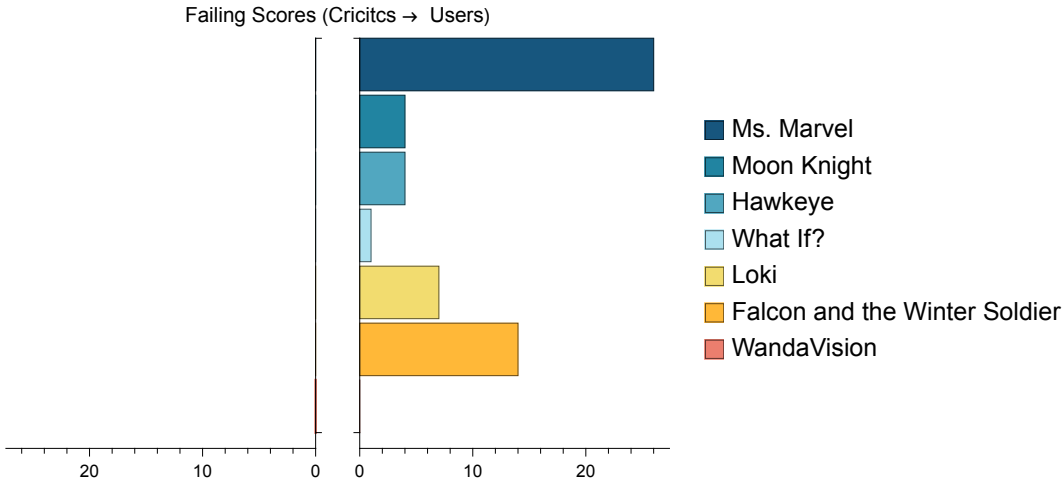
Out[ ]=



Now let's take a look at failing scores. First, we should note that critics never gave a single show a 0. This is probably due to some combination of the greater control they can exercise when assigning a score (100 options versus 10) as well as the fact that this is their profession, and so they are less likely to be impulsive and biased in their rating. Users, on the other hand, had no such qualms. While WandaVision received no failing scores, every other show received at least one. Ms. Marvel leads the pack with a whopping 26, almost twice as much as the next show, Falcon and the Winter Soldier, at 14. And even this is a bit extreme, compared to the 1, 4, or 7 failing scores assigned to all the other shows.

In[ ]:=
```
PairedBarChart[First /@ fails, Last /@ fails, ···]
```

Out[ ]=



So what can we say so far? The distribution of critic scores assigned to Ms. Marvel is very strange. It indicates that, almost unanimously, every critic enjoyed the show. So it seems to have hit a real jackpot in terms of what critics are looking for. The distribution of user scores assigned to Ms. Marvel is also very strange, if not for the same reason. It is being pulled in opposite directions by a group of users

who, in accordance with critics, are rating the show very highly, and a larger group of users who are rating it terribly. Moreover, there tends to be little to no in-between area here. Taken together, this data does paint a very suggestive picture of review-bombing. The show is high-quality and so liked by a good number who come across it. However for an even larger portion of people, something about the show drives them to purposefully give it an unreasonably low score. It is the easiest and most realistic explanation of the data. Of course, we can't necessarily "prove" this. After all, people can always just say they "didn't like it". Here though, we'll take it to be true. But this only opens up an even more interesting question. *What reasoning do users give for rating the show so horribly?*

What types of language are prominent in reviews?

We can assume that users will at least try and explain the rationale behind their score within their review (though we can't always assume their full honesty). So let's look at the actual text that features in these reviews. If we count the occurrences of all words in the critic and user reviews for each show (disregarding "stopwords" as well as those that don't appear at least 10 times), we can generate word-clouds for each show according to both critics and users (note that in the visualization below, the critic wordclouds are on the top and the user wordclouds are on the bottom). The critic wordclouds reveal very little, since critics' reviews use very distinct and variegated language, and so the only words that appear frequently across all reviews are things such as the show title, the release date of the show, and the names(s) of main actor(s). User wordclouds, meanwhile, tend to be quite a bit more descriptive. Notably, since there are less possible scores for users to assign, the wordclouds also show the most common score(s). Again, notice how Ms. Marvel's wordcloud is the only one that captures both a "0" and "10", another testament to its polarizing nature.

In[ ]:=
```
topWords = Map[Select[(# // StringJoin // DeleteStopwords // WordCounts),
      GreaterEqualThan[10]] &, allReviews, {2}];
```

In[ ]:=
```
Grid[Flatten[Transpose[Partition[#, UpTo[1]] & /@
     ({Keys@#, Values@#} &@(Column[#, Center, Spacings → 2] & /@
         Map[WordCloud[#, ImageSize → Medium] &, topWords, {2}]))],
  1], Frame → All, Spacings → {Automatic, 0}]
```

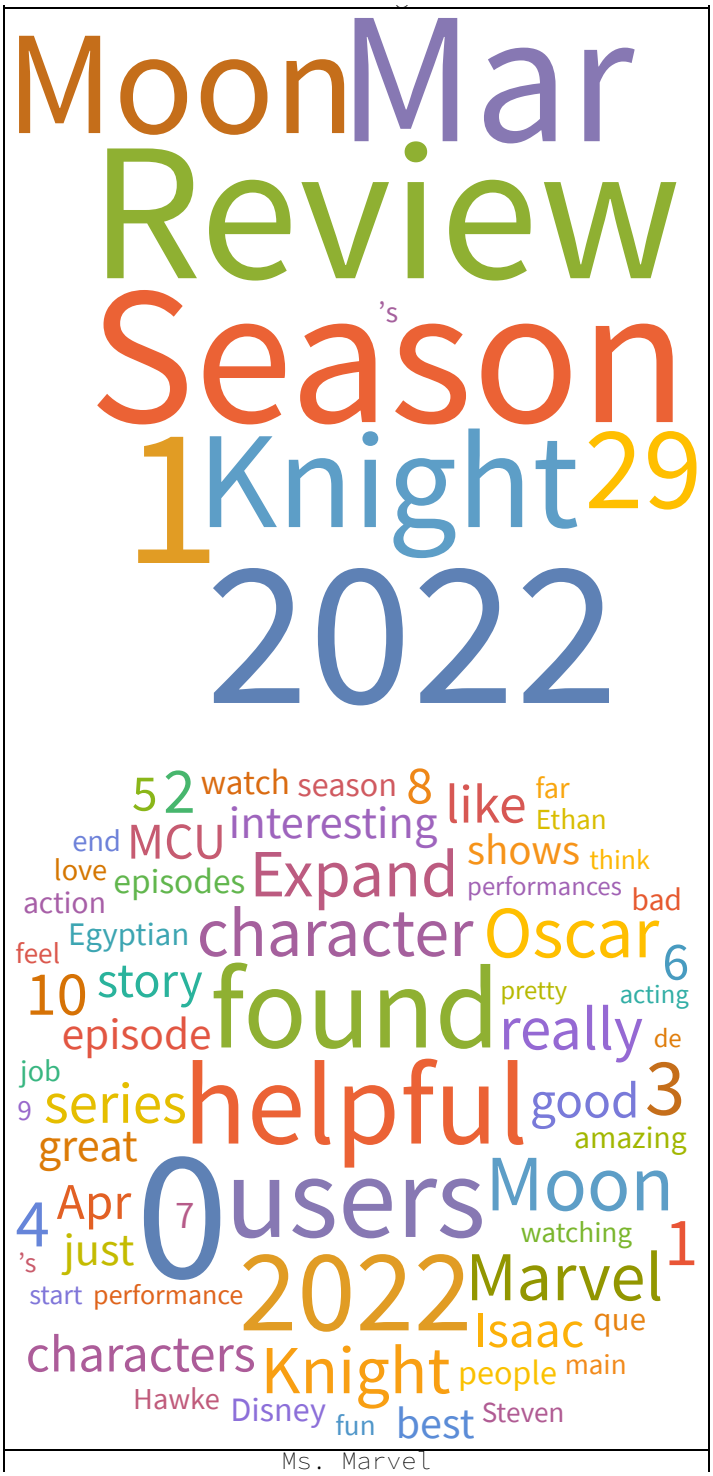
Out[ ]=


Falcon and the Winter Soldier
's
18.
Mar
Review
Season
2021
1
Soldier
series
Winter
Falcon
serie
23
black
better
make
26
thing
woke
17
MCU
episode
interesting
Winter
contains
click
episodes
characters
political
end
28
Captain
bad
25
Soldier
Expand
America
22
20
serum
18
new
5
Bucky
found
Marvel
Sep
Apr
great
6
liked
say
Zemo
Disney
7
watch
super
action
el
helpful
4
1
story
y
like
Sam
19
time
racism
view
want
really
9
good
users
2
Falcon
8
just
10
la
un
29
con
people
series
2021
0
3
16
spoilers
de
13
WandaVision
que
best
review
character
entertaining
21
movies
makes
terrorist
Walker
Loki


2021
Review
Season
1
8
Loki
's
Marvel
Jun
series
24
15
thing
female
22
music
Martino
watch
16
review
better
bad
12
11
makes
multiverse
Marvel
4
best
series
2021
8
Wilson
acting
13
characters
20
interesting
18
amazing
de
Hiddleston
found
just
far
Sep
good
story
2
9
make
view
TV
23
que
Owen
time
helpful
great
3
people
Di
Jul
episode
Sophia
7
Aug
10
like
MCU
0
Loki
spoilers
5
Sylvie
1
really
Jun
development
season
Disney
users
Tom
shows
click
contains
episodes
14
6
TVA
17
Expand
character
love
lot
What If?


Review
Season
Aug1
2021
Marvel
2021
1
episodes
3
like
6
de
4
story
helpful
Sep
Good
Expand
10
la
feel
episode
really
found
0
9
interesting
7
8
series
users
en
just
stories
MCU
5
2
Hawkeye


Hawkeye
Review
Season
23
2021
's
1
Nov
2022
shows
best
like
Marvel
Christmas
Hailee
character
Jeremy
Clint
episodes
really
10
8
Kate
found
0
1
fun
just
6
's
great
Disney
la
2
que
Hawkeye
de
le
4
9
good
helpful
Dec
3
new
7
MCU
2021
users
Bishop
series
episode
Expand
characters
action
moments
better
Steinfeld
Moon Knight


MoonMar
Review
Season
's
1Knight29
2022
52 watch season 8 like far Ethan
end MCU interesting shows think
love episodes Expand performances bad
action
feel Egyptian character Oscar 6
10 story found pretty acting
episode really de
job
9 series helpful good 3
great amazing
4 Apr 7 0users Moon
's just watching 1
start performance 2022Marvel
Isaac que
characters Knight people main
Hawke Disney fun best Steven

Ms. Marvel

Let's take a closer look at Ms. Marvel in particular. We are interested in how the wordclouds differ between positive and negative reviews. We can generate four types of wordclouds: a wordcloud of the words that feature prominently in positive reviews (defined as a review whose score is greater than or equal to 7), a wordcloud of the words that feature prominently in negative reviews (defined as a review

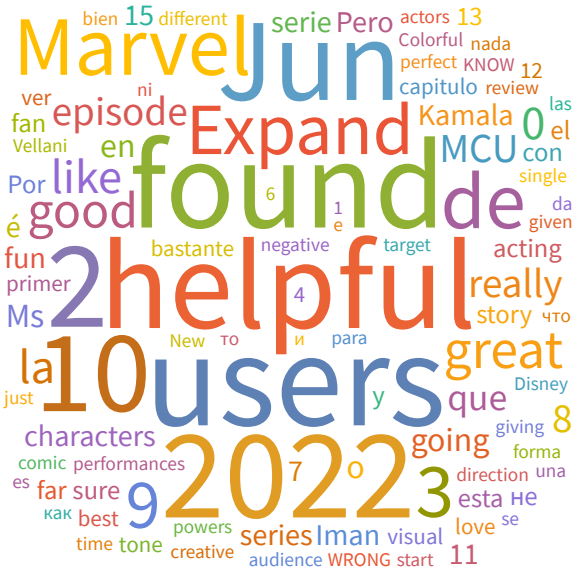
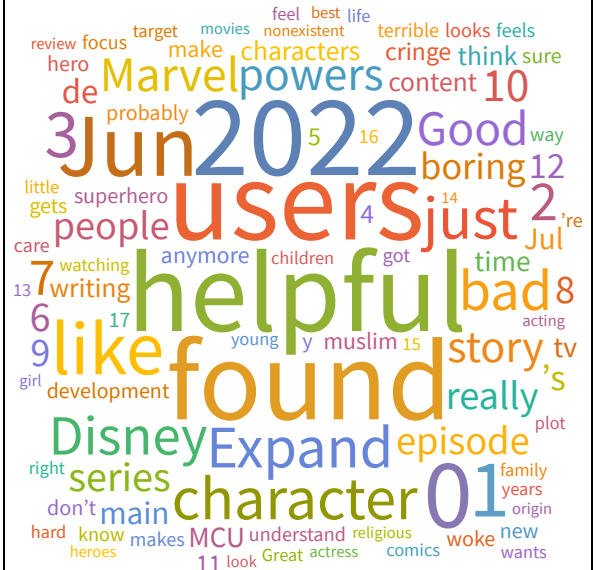
whose score is less than or equal to 3), and wordclouds of words that are exclusive to positive and negative reviews up (so we can get an even better sense of how they differ). Since we are now restricting ourself to a smaller set of reviews, we only require a word to be mentioned twice to be included in one of our wordclouds (this also means that our wordclouds for exclusivity are only sensitive up to a count of two).

```
In[ ]:= scoresToReview = Map[Merge[#, Identity] &,
          Map[Association[First@*extractScore@# → #] &, allReviews, {3}], {2}];

In[ ]:= scoresToTopWords =
        Map[Select[WordCounts[DeleteStopwords@StringJoin[#]], GreaterEqualThan[2]] &,
         scoresToReview, {3}];

In[ ]:= scoresToTopWords = Map[KeyMap[ToExpression], scoresToTopWords, {2}];

In[ ]:= {userGood, userBad} =
        {Last /@ Map[Merge[#, Total] &@*Values, Map[KeySelect[#, GreaterEqualThan[7]] &,
            scoresToTopWords, {2}], {2}], Last /@ Map[Merge[#, Total] &@*Values,
           Map[KeySelect[#, LessEqualThan[3]] &, scoresToTopWords, {2}], {2}]};

In[ ]:= goodVBad[show_] := Module[{good, bad},
        {good, bad} = {userGood[show], userBad[show]};
        Grid[{{"Good User Keywords (All)", "Bad User Keywords (All)"},
          {WordCloud[good, ImageSize → 300], WordCloud[bad, ImageSize → 300]},
          {"Good Users Keywords (Unique)", "Bad User Keywords (Unique)"},
          {WordCloud[Association@(# → good[#] & /@ Complement[Keys@good, Keys@bad]),
            ImageSize → 300],
           WordCloud[Association@(# → bad[#] & /@ Complement[Keys@bad, Keys@good]),
            ImageSize → 300]}}, Frame → All]]
```

Doing this for Ms . Marvel is very revealing. First, let's examine both the wordcloud of all good reviews, as well as the wordcloud containing words exclusive to good reviews. There is a lot of overlap here. Interestingly, we find lots of Spanish and Russian language in both of these categories. Upon further review of the data, there are quite a few reviews left by Spanish speakers (~5), but only one review left by a Russian speaker (with lots of repetitive language). Overall, we see the generic positive language we would expect, such as "good", "great", "creative", "super", "fun", "love", etc . There is also lots of mention the characters and cast (specifically, Iman Vellani, the show's lead actor) and descriptors such as a "charismatic", "colorful", and "relatable". If we look we at the wordcloud of all negative reviews, we see – in addition to generic negative language (e.g., "bad", "trash", "boring", "terrible", "cringe") – words like "Muslim","powers", "comics", "religious", "woke", "political", and "VFX", all of which are further amplified in the wordcloud containing language exclusive to negative reviews.

In[ ]:= `goodVBad["Ms. Marvel"]`

Out[ ]=

Good User Keywords (All)	Bad User Keywords (All)
	
Good Users Keywords (Unique)	**Bad User Keywords (Unique)**
	

So let's break this down. Firstly, a large number of negative reviews mention themes of race and religion. Actually reading some negative reviews makes this even more obvious. Many users invoke "superhero fatigue" as their rationale for rating the show badly, saying that the only thing special about Ms. Marvel is that she is a "Pakistani" or "Muslim" superhero and criticizing the show "for spending too much time around her roots" ("she's Indian" – someone wasn't paying attention! – "we get it"). Of course, this logic is quick to break down when we apply it to other heroes in the MCU...we have been adding white superhero after white superhero for quite a while now without any mention of superhero fatigue. And that's because we have never taken ethnicity to be synonymous with uniqueness. Instead

it is the perspective of a superhero as informed by their worldview, skills, and the people they surround themselves with that allows them to leave a distinct mark among the panoply of Marvel heroes. Ethnic background undoubtedly colors a part of this perspective, but it does not constitute it. In the case of Ms. Marvel, this perspective is that of a teenager growing up in a world of superheroes whom she endlessly admires and fantasizes about, until getting powers of her own forces her to rethink what being a superhero actually means. I don't mean to imply it's a revolutionary concept (really, it's a pretty time-tested comic book trope), but it's hardly devoid of purpose either. And casting it in a new cultural and religious context *does* promise that we won't be watching Spiderman 2.0, something that many review-bombers say they don't want, but perhaps secretly desire.

Secondly, some users point to the change in Ms. Marvel's power set from the comics ("power", "comics") as a reason for disliking the show. To be fair, I can't really relate to this as I have little familiarity with the source material. From some light research, it seems that in the comics Ms. Marvel's powers are more akin to shapeshifting, which better resonate with the themes of struggling to figure out her identity that penetrate both the comics and the show. I do agree this is definitely the most extensive reworking of powers (shapeshifting to energy manipulation) we've seen in the MCU so far, but it's interesting that we've never seen people get so upset over power changes before. I mean, the whole point of WandaVision was to retcon the origin and scale of Wanda's abilities, and correct the watered-down Scarlet Witch we had from Age of Ultron to Endgame. Similarly, in Black Widow, Taskmaster's identity and capabilities underwent serious modifications for the sake of the story (for example, making her abilities due to an external suit, and removing the downside of memory loss that comes when copying new abilities). The list goes on and on. Of course, this is definitely is a valid criticism of the show, and I respect anyone who feels like they could not enjoy the show because of it. But it also seems like the sort of thing that would-be review-bombers could easily use as false justification for their scoring. The same goes for complaints about poor cinematography ("VFX").

Finally, words like "woke" and "political" criticize the "agenda pushing" and "virtue signaling" that many assume to be implicit in any show with a diverse cast. Obviously, Hollywood is no stranger subliminal messaging (just think about Jurassic World: Dominion's very conspicuous decision to cast the film's "villain" as a Tim Cook lookalike), but just because a show has a diverse cast does not mean it is making any particular criticism of certain groups of people and their beliefs. Personally, I do not feel Ms. Marvel did.

Overall, it seems that the language profile of negative reviews focused mainly on the racial and religious identities of the show's main characters and its perceived political ideology, as well as, among a smaller portion of users, deviations from the comics. Positive reviews on the other hand focused more on the show's content, namely its story and actors. This is as close to proving review-bombing we can get: we have identified a large group of users who is giving the show terrible scores despite critics (and many other users) rating it extremely highly, as well as identified a bias that they share. We should also note that the only other Marvel TV show with a largely non-white cast, Falcon and the Winter Soldier, exhibits many of the same statistical trends we have found to be associated with review-bombing in Ms. Marvel. However, they are not as extreme as they are in Ms. Marvel though, and Falcon and the Winter Soldier does not have the same robust baseline in terms of critic score.

How did the length of reviews vary?

There's one more thing I'm curious about though. We define review-bombing as the practice of large numbers of users making a show look bad through the sheer quantity of their negative reviews. So are the "bombed" reviews for Ms. Marvel much shorter than those of other shows, in an effort to promote quantity over quality? On the other hand, it also seems possible that they could be long, incoherent rants.

We can plot smooth histograms for the character length of good and bad reviews for each show according to both critics and users. First, let's note that critic reviews never go above 800 characters, while we see some user reviews going into the low thousands. This is because critic reviews are usually just the first paragraph of a more in-depth review which is linked to at the end of the paragraph. Furthermore, the length distributions for critic reviews, whether good or bad, tend to look very symmetrical. The length distributions for user reviews on the other hand exhibit much more skew, though there does not seem to be any noticeable patterns. Ms. Marvel does have a large right skew in the length distribution for its negative user reviews (indicating a bias towards lots of shorter reviews), but so do many other shows. Perhaps digging a bit deeper into the data would reveal something of note, but at first glance it doesn't appear to be particularly interesting. And since we have already gone on long enough, I will leave this as an exercise to any particularly engaged readers!

In[ ]:=
```
scoresToLength =
  Map[KeyMap[ToExpression], Map[x ↦ StringLength /@ x, scoresToReview, {3}], {2}];
```

In[ ]:=
```
criticLengthGood =
  First /@ Map[KeySelect[GreaterEqualThan[80]], scoresToLength, {2}];
```

In[ ]:=
```
criticLengthBad = First /@ Map[KeySelect[LessEqualThan[50]], scoresToLength, {2}];
```

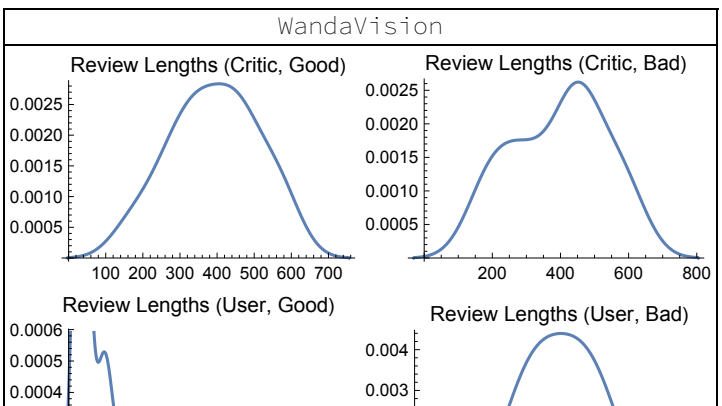
In[ ]:=
```
userLengthGood = Last /@ Map[KeySelect[GreaterEqualThan[7]], scoresToLength, {2}];
```

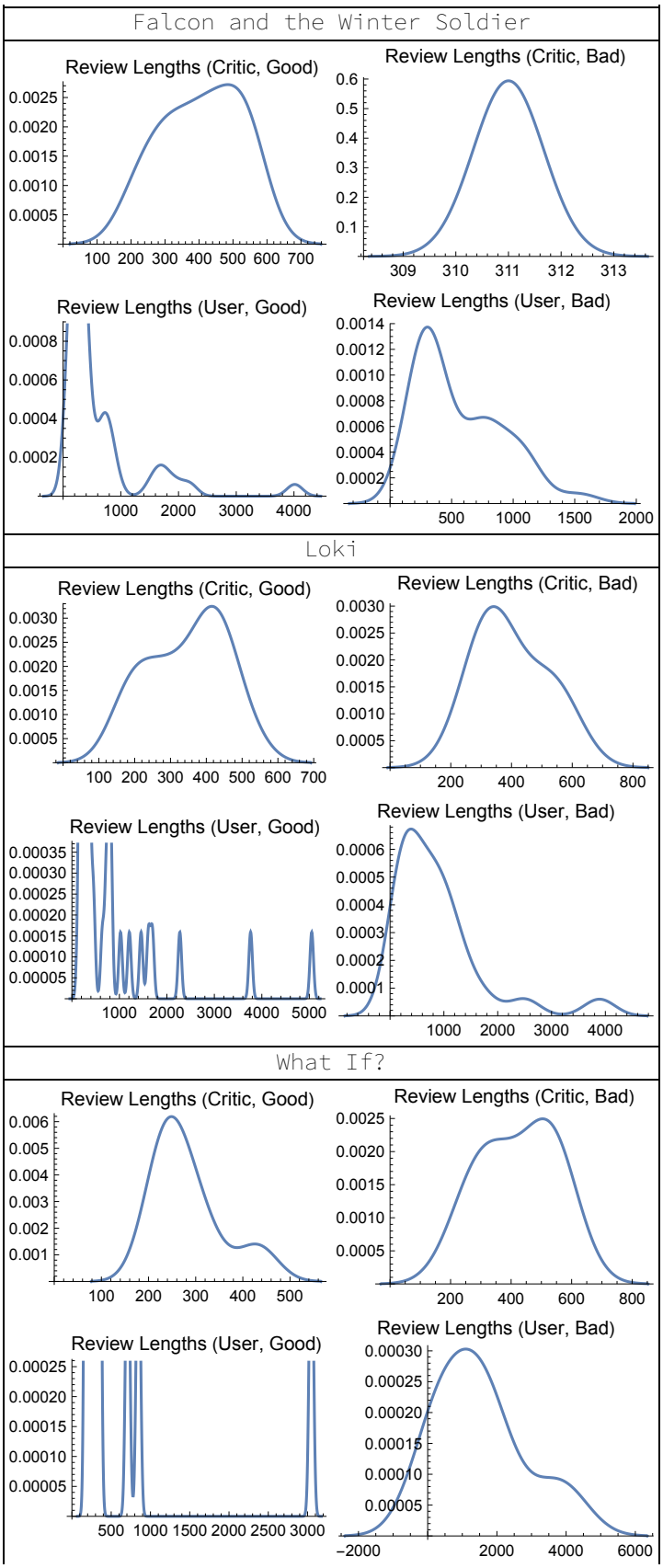
In[ ]:=
```
userLengthBad = Last /@ Map[KeySelect[LessEqualThan[3]], scoresToLength, {2}];
```

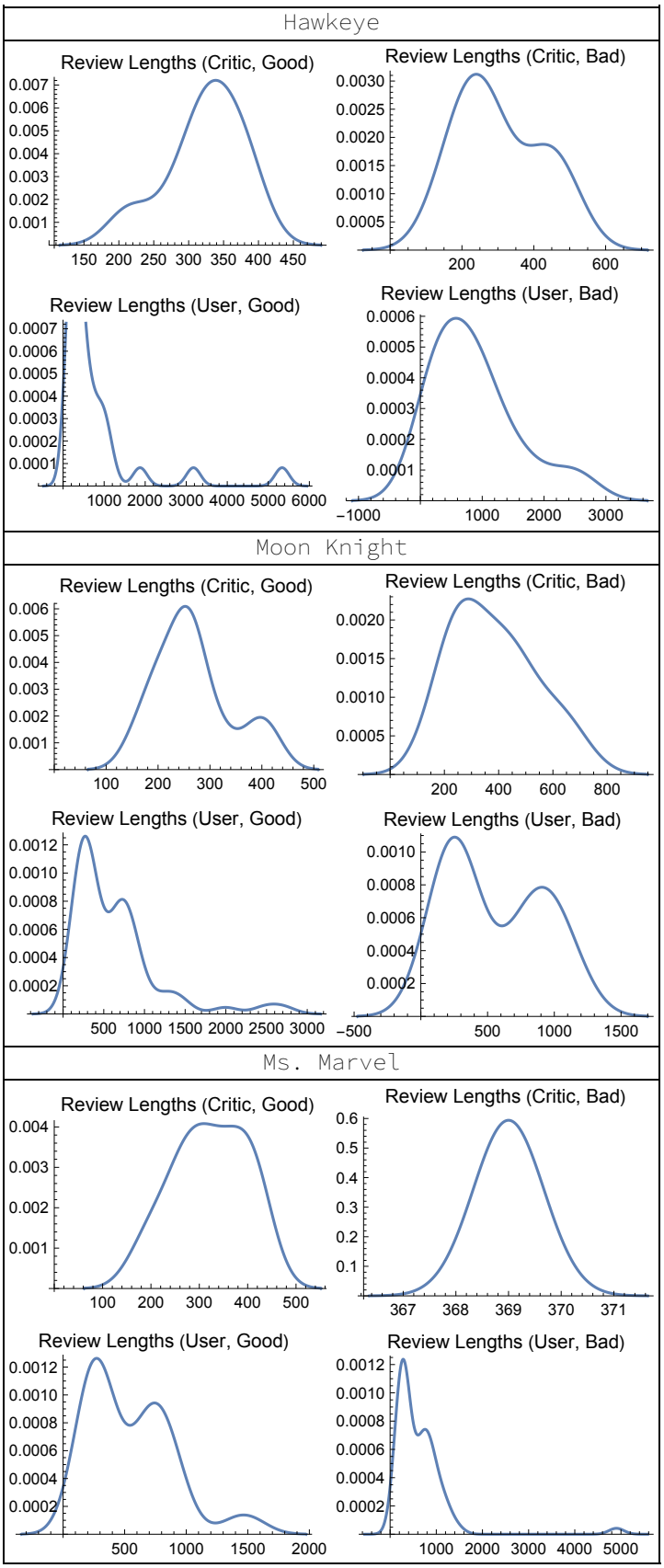
In[ ]:=
```
collected = (Partition[#, {2}] & /@
    Merge[Map[x ↦ SmoothHistogram[#, PlotLabel → x[[2]]] & /@ Flatten /@ Values /@ x[[1]],
      {{criticLengthGood, "Review Lengths (Critic, Good)"},
       {criticLengthBad, "Review Lengths (Critic, Bad)"} ,
       {userLengthGood, "Review Lengths (User, Good)"},
       {userLengthBad, "Review Lengths (User, Bad)"}}], Identity]);
```

In[ ]:=
```
Grid[Flatten[MapThread[{{#1}, {#2}} &,
   {Keys@collected, Grid /@ Values@collected}], 1], Frame → All]
```

Out[ ]=




Falcon and the Winter Soldier
Review Lengths (Critic, Good)
Review Lengths (Critic, Bad)
Review Lengths (User, Good)
Review Lengths (User, Bad)
Loki
Review Lengths (Critic, Good)
Review Lengths (Critic, Bad)
Review Lengths (User, Good)
Review Lengths (User, Bad)
What If?
Review Lengths (Critic, Good)
Review Lengths (Critic, Bad)
Review Lengths (User, Good)
Review Lengths (User, Bad)


Hawkeye
Review Lengths (Critic, Good)
Review Lengths (Critic, Bad)
Review Lengths (User, Good)
Review Lengths (User, Bad)
Moon Knight
Review Lengths (Critic, Good)
Review Lengths (Critic, Bad)
Review Lengths (User, Good)
Review Lengths (User, Bad)
Ms. Marvel
Review Lengths (Critic, Good)
Review Lengths (Critic, Bad)
Review Lengths (User, Good)
Review Lengths (User, Bad)

What statistical conclusions can we ultimately draw?

- Critics liked the show a lot
- Some users liked the show a lot, but more hated it a lot
- Ms. Marvel is the most polarizing TV show Marvel has released to date
- Themes of race, religion, and the show's politics / ideology are highly cited reasons for low scores
- Taken together, these factors paint a highly suggestive picture of review-bombing
- This review-bombing appears to be an attack on the various ways in which the show is more diverse and different than other Marvel motion pictures

Why I did this project

To be honest, the motivation I provided at the outset of this journey was a little disingenuous. After all, we went through a bunch of statistical overkill to "prove" something that could be divined with a little bit of common sense. I mean, Ms. Marvel had a one star rating on Rotten Tomatoes within three minutes of its release [9], not nearly enough time for anyone to have watched a substantial part of the 50 minute episode. And this reaction isn't new either – it mirrors the exact same one that happened several years ago when the comics were first released [7], before Ms. Marvel was able to win hearts and minds. Moreover, the point really wasn't to change anyone's opinion on the topic either (although if it did, that's great!). As you can tell, the people review-bombing the show are deeply entrenched in some pretty pernicious biases, and so I doubt any of the data-driven insights we've discovered here today will mean anything to them, or will inspire them to elevate their own discourse to a commensurate level of accuracy. So what was the point then? Well, it was really just a passion project of mine. It combined my interest in data science with my own strong reaction to the show, and the way I related to many of its themes as an American-born Indian. Plus, the statistician in me wanted to see if I could get any pretty plots out of this! So if anybody needs some evidence to back up their claims of review-bombing, you now have a good 20-or-so pages worth!

Additionally, the framework used here is very general and could be a useful tool for analyzing reactions to Marvel movies and future Marvel shows as well. I encourage you to dig into the code, modify it to your own needs, and see what else you can discover! Make sure to let me know, or send any other thoughts, opinions, and advice my way!

References

- [1]: https://www.yahoo.com/video/ms-marvel-praised-fans-flooded-140258354.html
- [2]: https://in.ign.com/ms-marvel/173635/news/ms-marvel-is-the-highest-rated-mcu-project-ever-on-rotten-tomatoes-despite-the-review-bombing
- [3]: https://www.looper.com/928487/ms-marvel-just-claimed-a-shocking-mcu-title/
- [4]: https://www.nme.com/news/tv/ms-marvel-directors-respond-to-review-bombing-3248481

- [5]: https://www.newsweek.com/ms-marvel-mcu-disneyplus-review-bombing-racist-white-nonsense-1714538
- [6]: https://screenrant.com/ms-marvel-show-review-bomb-backlash-iman-vellani/
- [7]: https://www.digitalspy.com/tv/ustv/a40563393/ms-marvels-iman-vellani-responds-disney-marvel-show/
- [8]: https://tribune.com.pk/story/2361695/ms-marvel-is-being-review-bombed-on-imdb-for-its-south-asian-muslim-superhero
- [9]: https://www.indiatimes.com/explainers/entertainment/explained-what-is-review-bombing-why-is-ms-marvel-at-the-receiving-end-of-this-practice-571956.html
- [10]: https://www.koimoi.com/hollywood-news/ms-marvel-is-now-mcus-highest-critically-rated-series-on-rotten-tomatoes-takes-over-agents-of-s-h-i-e-l-d-wandavision-others/