

Report 7 2.0: How does media coverage change through time?

A casual report

Ricky Heinrich for Mélanie Méthot

2023-09-03

Contents

what's new	1
Overview	1
Count Bar Plot	2
Categories	4
Coverage Through Time	5
Coverage by State	8
With Incomplete	8
Without Incomplete	9

what's new

- (in bold in text)
- 24.4% of the cases in the corpus don't have anything recorded for “# of articles”.
- found in some cases, says ‘0’ in “# of articles” but then there is something in link1 ...
- changed categories after sept 2 meeting: remove 0 as large proportion due to ‘not done yet’, 1, 5-9, 10-29 (public interest), 30+ (sensation); check ‘categories’ subtitle
- check coverage through time -> decades
- coverage by states w/o incomplete -> check analysis

Overview

It looks like there are 747 cases that don't have the number of articles recorded. **We are assuming that these cases are not yet completed yet, ie areas where research still needs to be done. These are 24.4% of the cases in the corpus. These will not be included in the analysis.**

The cases with empty cells for number of articles occur across five of the states, shown in the table below. A quick review in the original spreadsheet of Queensland, Western Australia, and NSW shows that at least

some of these cases do have links. The lack of value in this cell thus comes from a deficiency in the transcriber. The two cases in South Australia don't have any links, and the first few in Victoria also don't, but I didn't look at all of the cases.

Var1	Freq
NSW	535
Queensland	164
South_Australia	2
Victoria	23
Western_Australia	23

I don't have a quick an easy way right now to check the number of links and input a value, although I am sure it is doable. I may try to think about it and find a way later.

~~There are 647 cases where no entry was recorded for “# of articles” AND there was nothing recorded in “link1”. I am adding the entry ‘0’ in those cases.~~

There are 4 cases where ‘0’ was recorded for “# of articles” AND there was something recorded in “link1”. Annoying. I’m putting it under insufficient research for now.

There are 65 cases where there's no entry in the number of articles column, but there is one link or what seems like an article title. 39 cases had something along the lines of “yorick's info” but didn't include any links, so for now have remained as having no entry for # of articles. Since there there isn't an easy straightforward way to count how many links there are per cases, and a quick overview showed that some included more than 5 links, I will not change these values to ‘1’ article as it would skew the results. I will instead just not include them in the analysis, and leave them as ‘no entry’, which is valid since it is a small proportion of the data. Maybe later I'll go back and take the time to manually add the number of links since I guess it would be manageable.

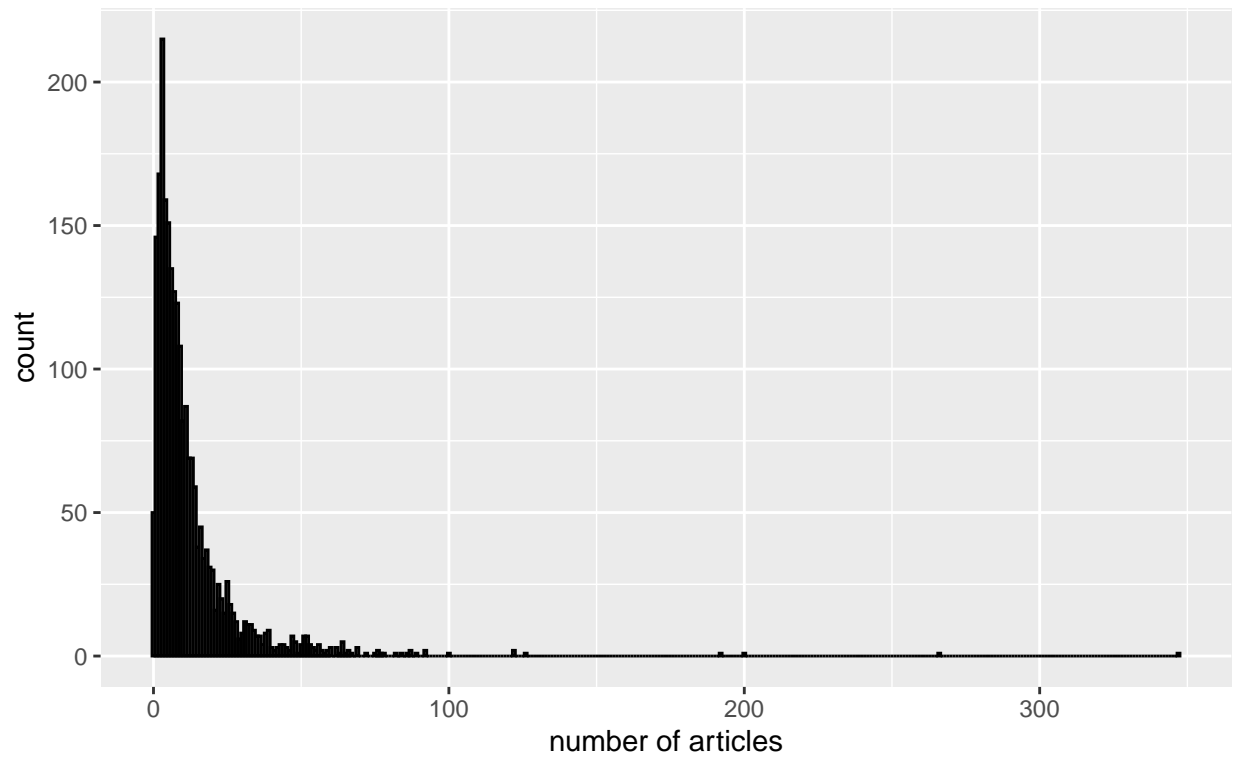
There are a few cases where the entry in “# of articles” is not just a numerical digit, shown in the table below. For those that have a variation of ‘more than x’, I'll just put in the x. For ‘cannot find articles’ and ‘none’ I will put ‘0’.

Var1	Freq
50+	1
6 (but 20 for his youth offences!) will also ask for a divorce later on	1
9 (2 in 1862 and 7 in 1867)	1
at least 12	1
cannot find articles on Trove currently	1
MORE THAN 122	1
more than 92	1
none	9

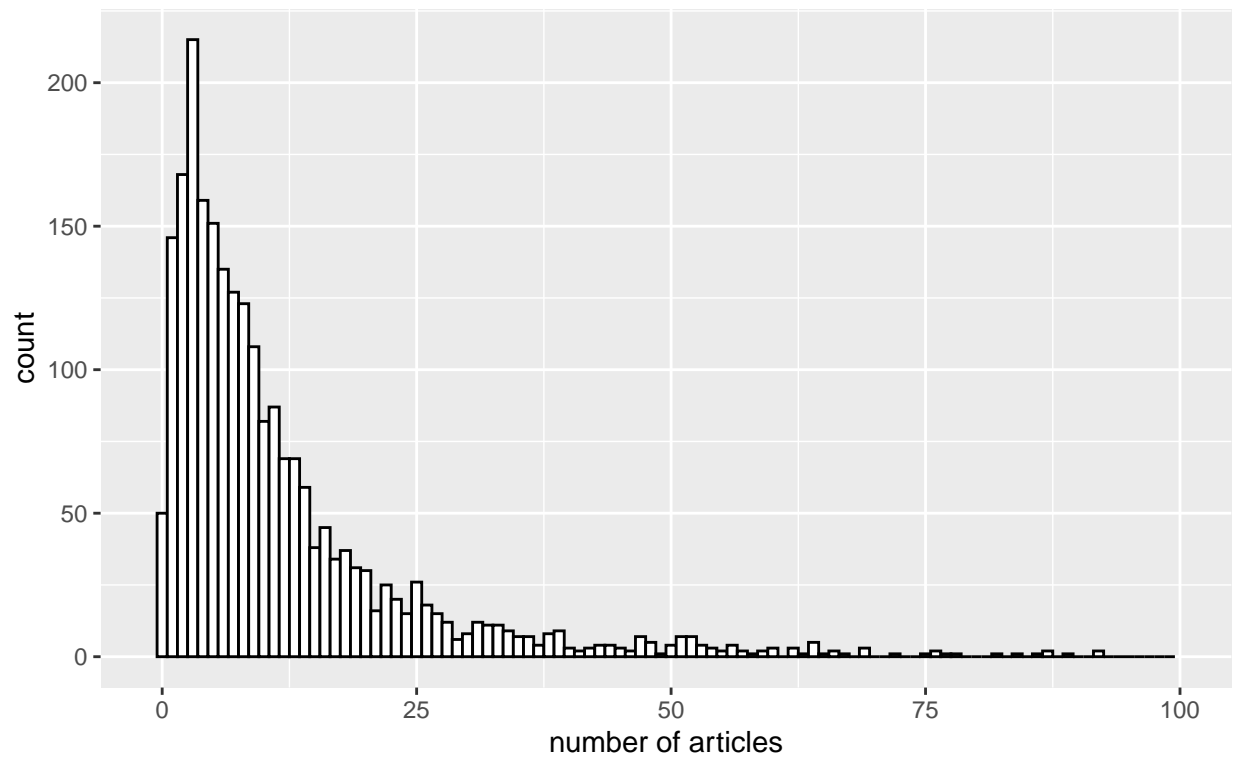
Count Bar Plot

The maximum number of articles is 347. The plot below shows the count of article numbers, and the two after show it zoomed in. **We see that the most occurring is ‘3’, which occurs 215 times.** There's somewhat of a peak at 3 articles, then a somewhat gradual decline, with small peaks at 11, 16, 17, 22, 25, ... these peaks are relatively small differences: the number of cases with 10 articles is 82, and that with 11 is 87.

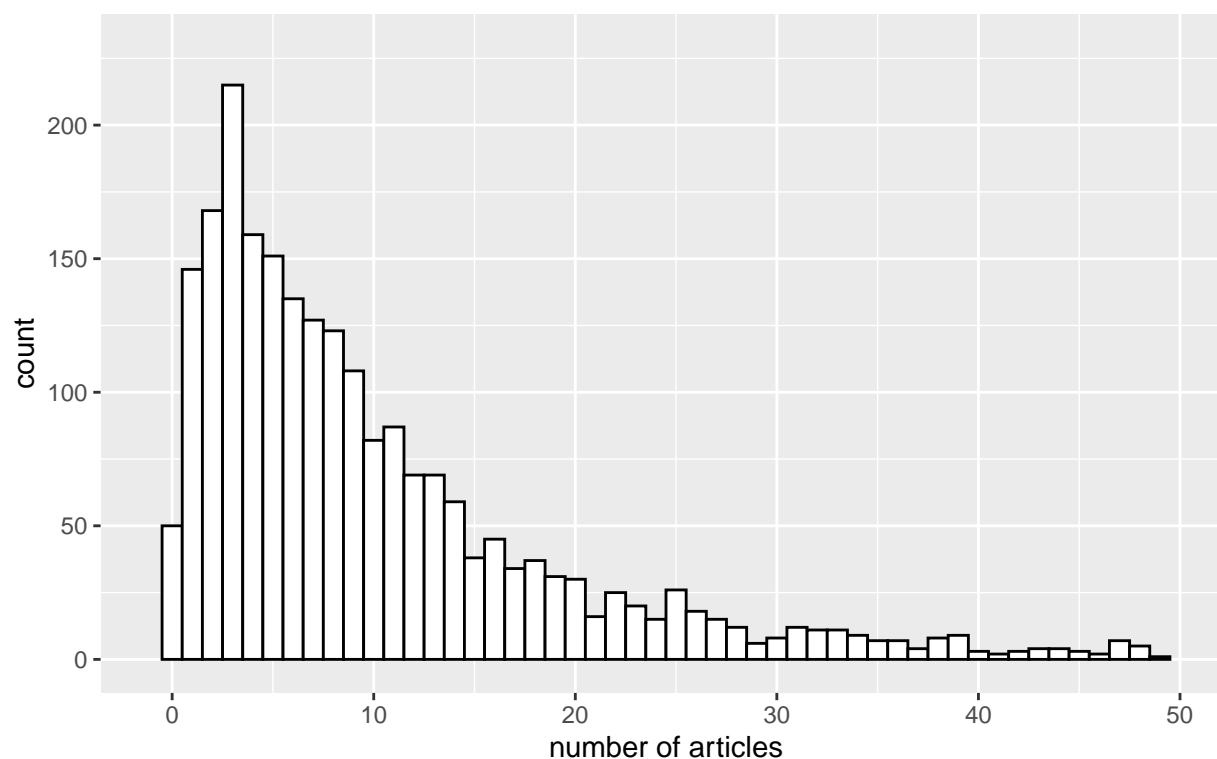
Count of Media Coverage per Case



Count of Media Coverage per Case



Count of Media Coverage per Case

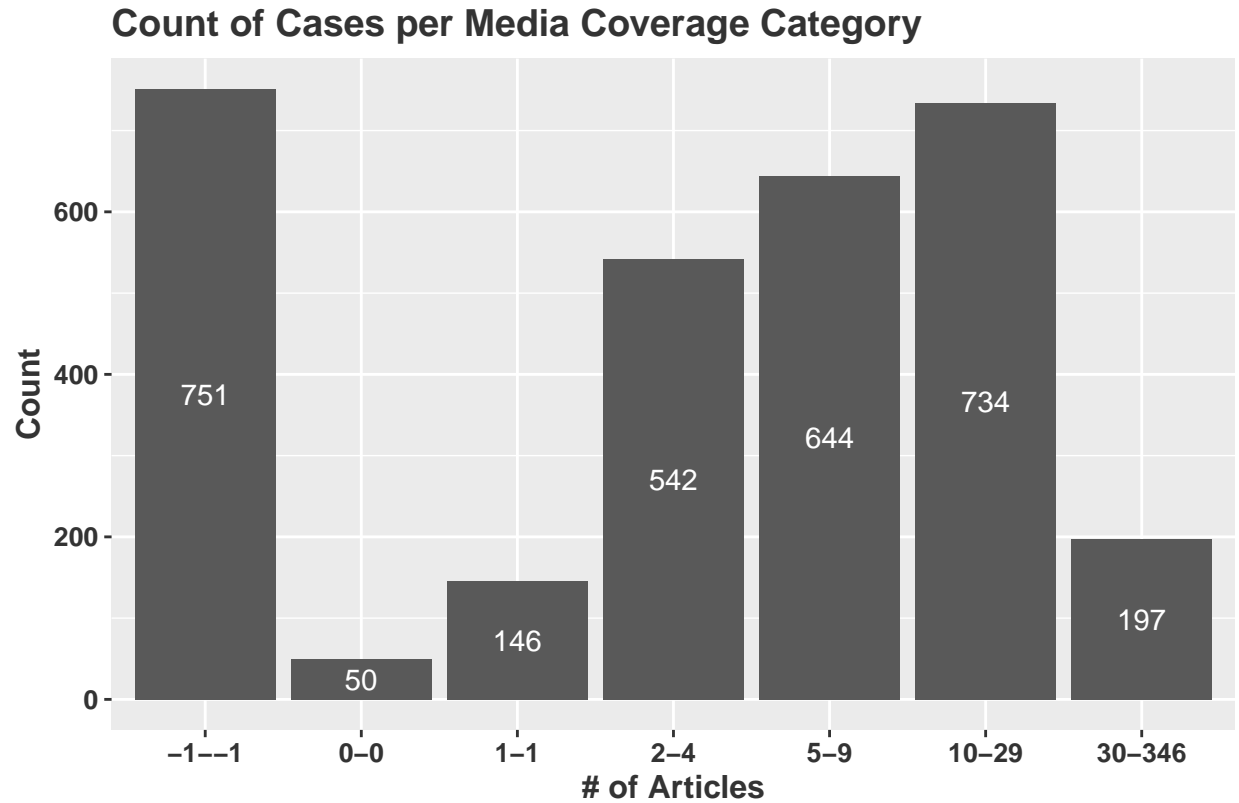


Here is the exact numbers in case you wanted to refer to them directly; every second row is the count.

```
##
##  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
## 50 146 168 215 159 151 135 127 123 108 82 87 69 69 59 38 45 34 37 31
## 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
## 30 16 25 20 15 26 18 15 12 6 8 12 11 11 9 7 7 4 8 9
## 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
## 3 2 3 4 4 3 2 7 5 1 4 7 7 4 3 2 4 2 1 2
## 60 62 63 64 65 66 67 69 72 75 76 77 78 82 84 86 87 89 92 100
## 3 3 1 5 1 2 1 3 1 1 2 1 1 1 1 1 2 1 2 1
## 122 126 192 200 266 347
## 2 1 1 1 1 1
```

Categories

After our discussion on Sept 2, we are categorizing as: 0, 1, 2-4, 5-9, 10-29, 30+. The following plot shows how many cases fall in each category. The most occurring is 10-29 (public interest).

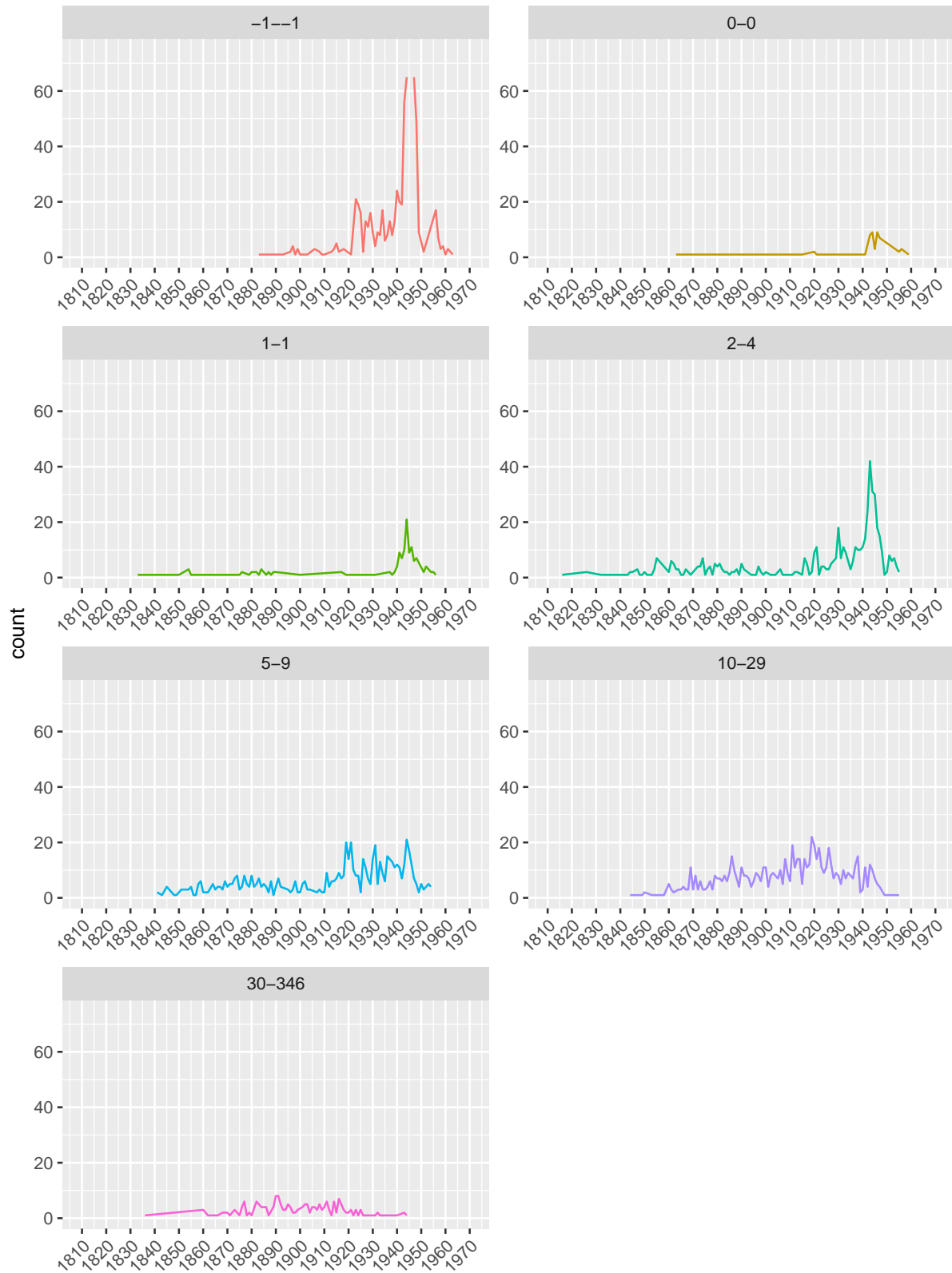


Coverage Through Time

Next we have these groups in time. We see that the groups with the largest difference in media coverage through time is the 1-4: there is an increase of cases with only 1-4 articles near the WWII years. The following plot is by decade: we see that there's a pretty drastic increase in cases with 2-4 articles in the 1940s, with 200 cases. There's also increases across the decades in cases with 5-9 articles and 10-29, although both have peaks that occurred somewhat earlier: case with 5-9 articles peaked at around 110 in the 1930s, whereas 10-29 peaked jointly in 1910s and 1920s at around 125. There were only about 50 cases in the 1940s with 10-29 articles. Sensations (30+ articles) occurred very infrequently in those years: sensations appeared to have peaked in the 1890s, with nearly 50 cases.

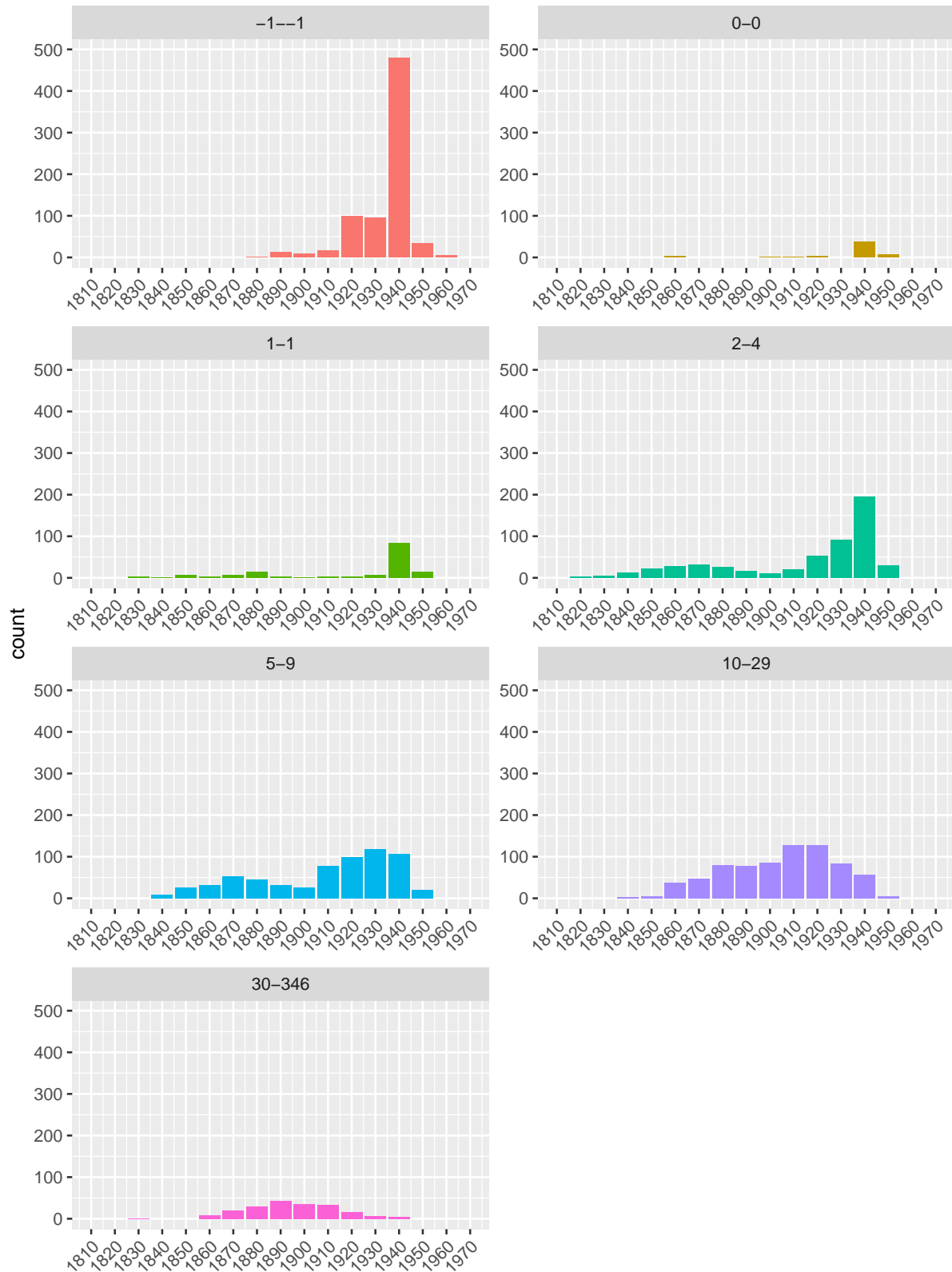
Count by 'number of articles' groups through time

-1--1 represents unfinished cases



Count by 'number of articles' groups by Decade

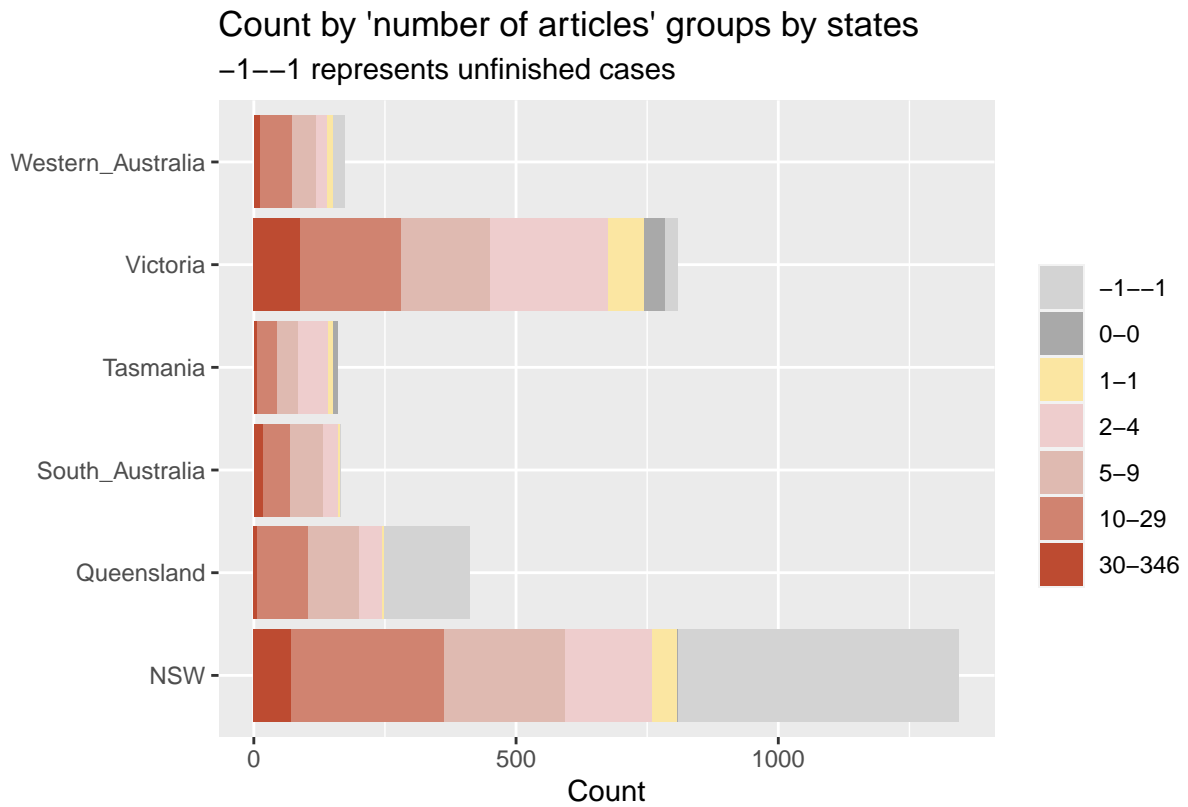
-1--1 represents unfinished cases

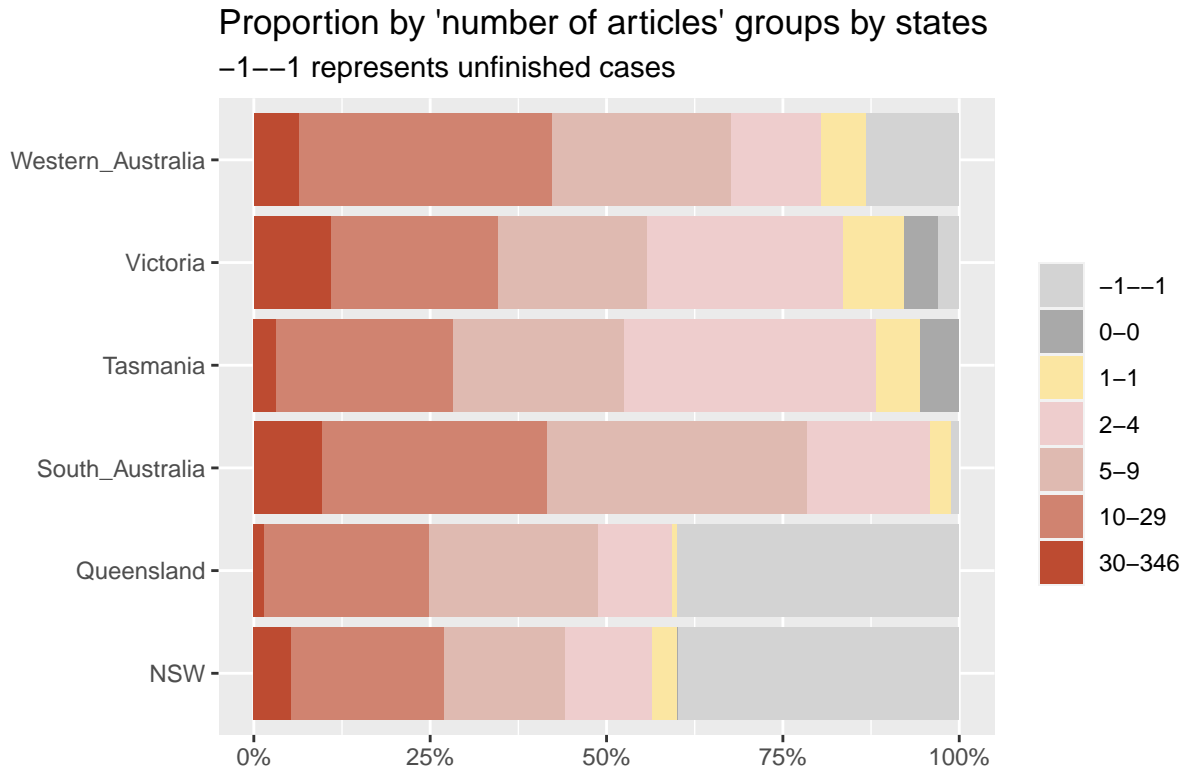


Coverage by State

With Incomplete

The first set of plots here include incomplete cases. I looks like a third of NSW and Queensland cases are incomplete (-1-1).

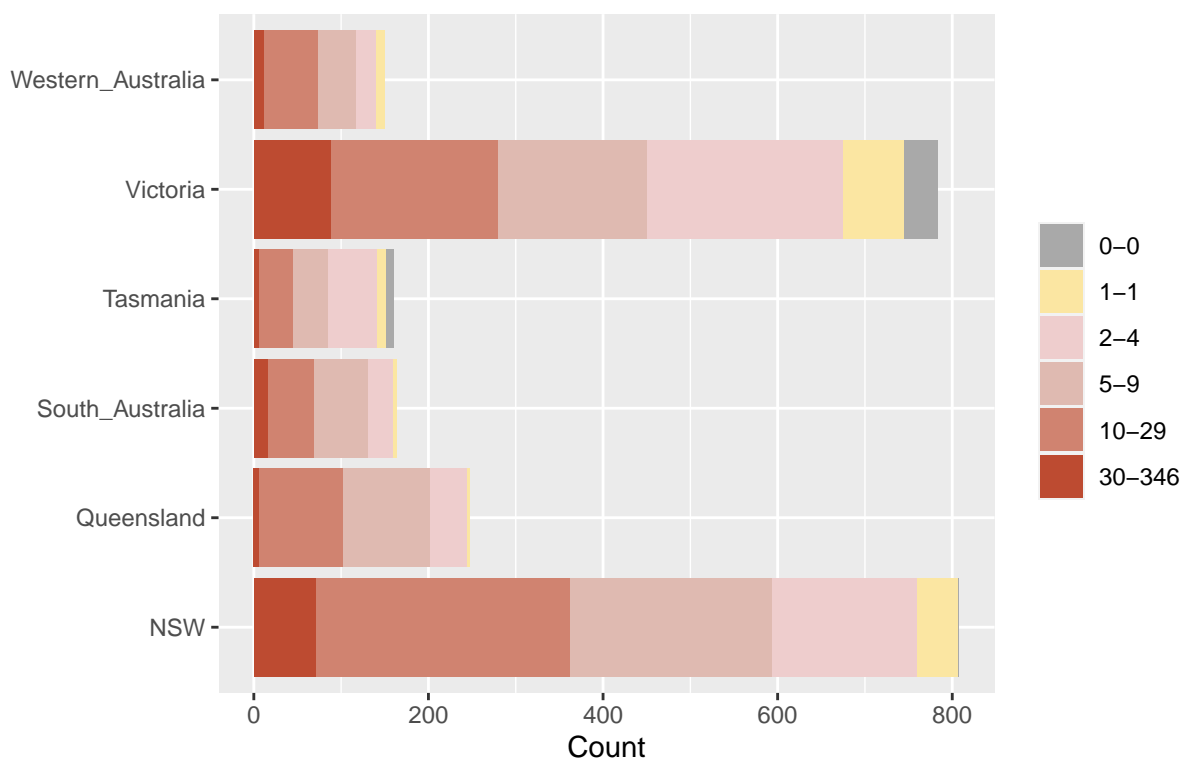




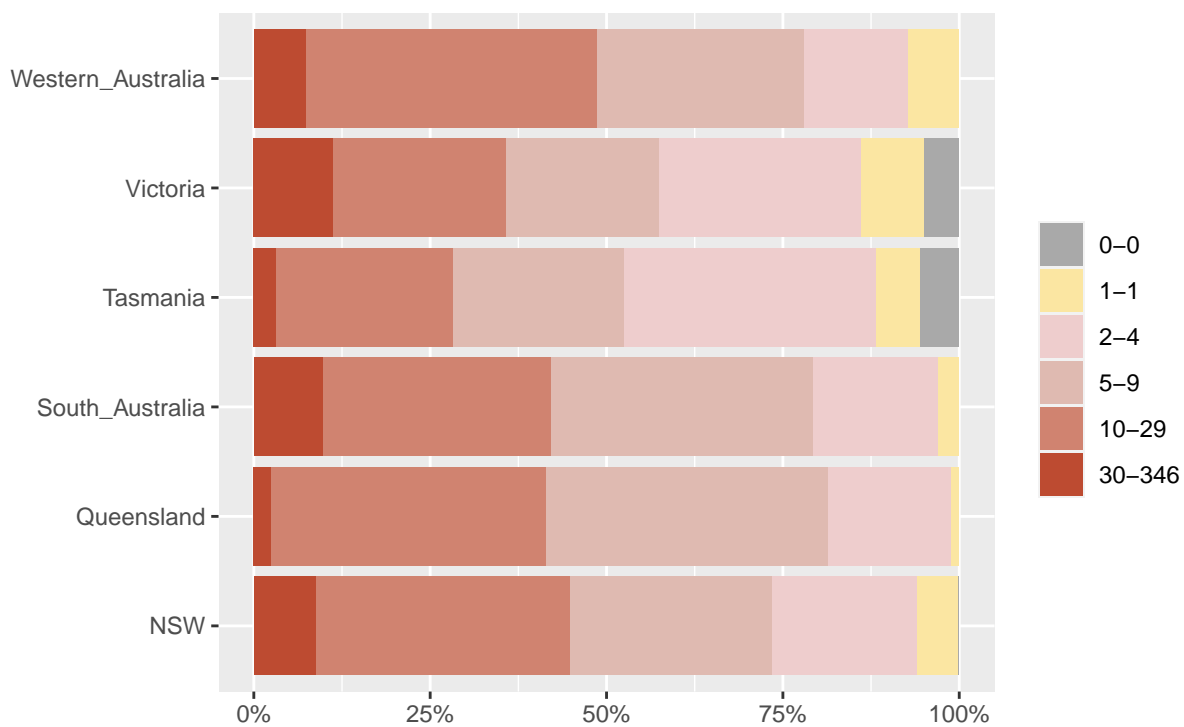
Without Incomplete

There are more cases in Victoria with 2-4 articles than in NSW, and it looks like Victoria has somewhat more cases than NSW that were sensations (30+ articles). Queensland seems to have an equal amount of cases with 5-9 articles than with 10-29 articles. Queensland also has a very small number of cases with only 1 article (3), even less than South Australia (5) (Exact number found in table below). You can see that the sensations (30+ articles) occurs in every state.

Count by 'number of articles' groups by states



Proportion by 'number of articles' groups by states



The exact numbers are shown in the table below.

	-1-1	0-0	1-1	2-4	5-9	10-29	30-346
NSW	537	1	48	166	231	291	71
Queensland	165	0	3	43	99	96	6
South_Australia	2	0	5	29	61	53	16
Tasmania	0	9	10	57	39	40	5
Victoria	24	40	69	225	170	192	88
Western_Australia	23	0	11	22	44	62	11

The following shows the proportions for easier quantifying. For example, the value intersecting 2-4 and NSW 12.3 means that 12.3% of NSW's cases have article counts between 2 and 4.

	-1-1	0-0	1-1	2-4	5-9	10-29	30-346
NSW	39.9	0.1	3.6	12.3	17.2	21.6	5.3
Queensland	40.0	0.0	0.7	10.4	24.0	23.3	1.5
South_Australia	1.2	0.0	3.0	17.5	36.7	31.9	9.6
Tasmania	0.0	5.6	6.2	35.6	24.4	25.0	3.1
Victoria	3.0	5.0	8.5	27.8	21.0	23.8	10.9
Western_Australia	13.3	0.0	6.4	12.7	25.4	35.8	6.4