

Determining historical spread of eBird India data

Karthik Thrikkadeeri

2022-08-07

Quick analysis to determine whether or not historical eBird data in the country is spatially biased, at two levels:

- pre-2000
- pre-1990 & 1990-2000

All these analyses are using the base dataset, without any of the SoIB filters, but with removal of duplicate lists.

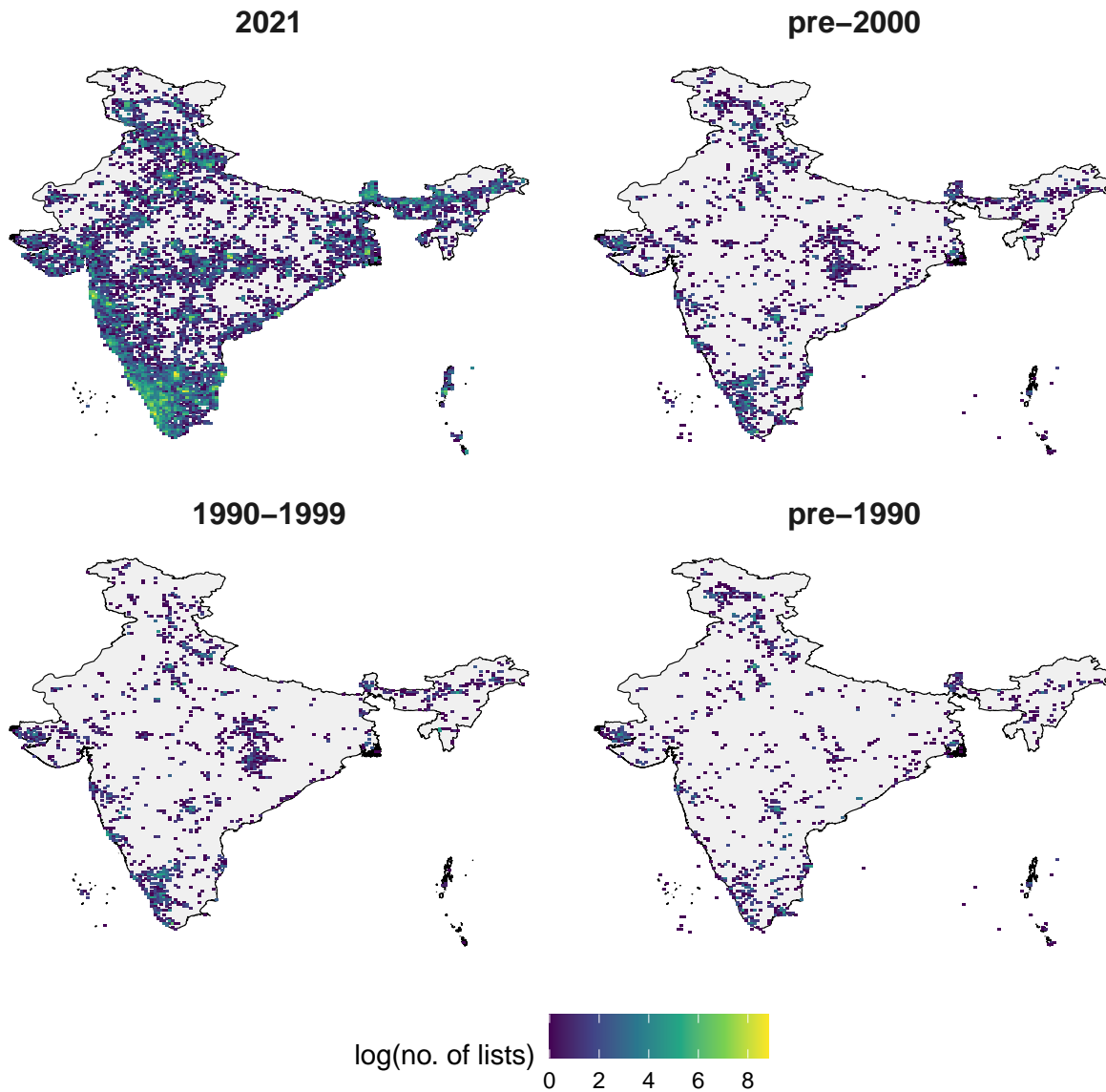
Contents

Maps of spatial spread	2
Histograms	3
Data across geographical regions	6
Progression of data with time periods	7
Conclusion	8

Maps of spatial spread

Firstly, are there any immediate biases that pop up when historical lists are plotted on a map of the country? How does the spread of the data look when the historical period is split into two, compared to without the split? (Zoom in for cell-level detail.)

Birding intensity across grid cells

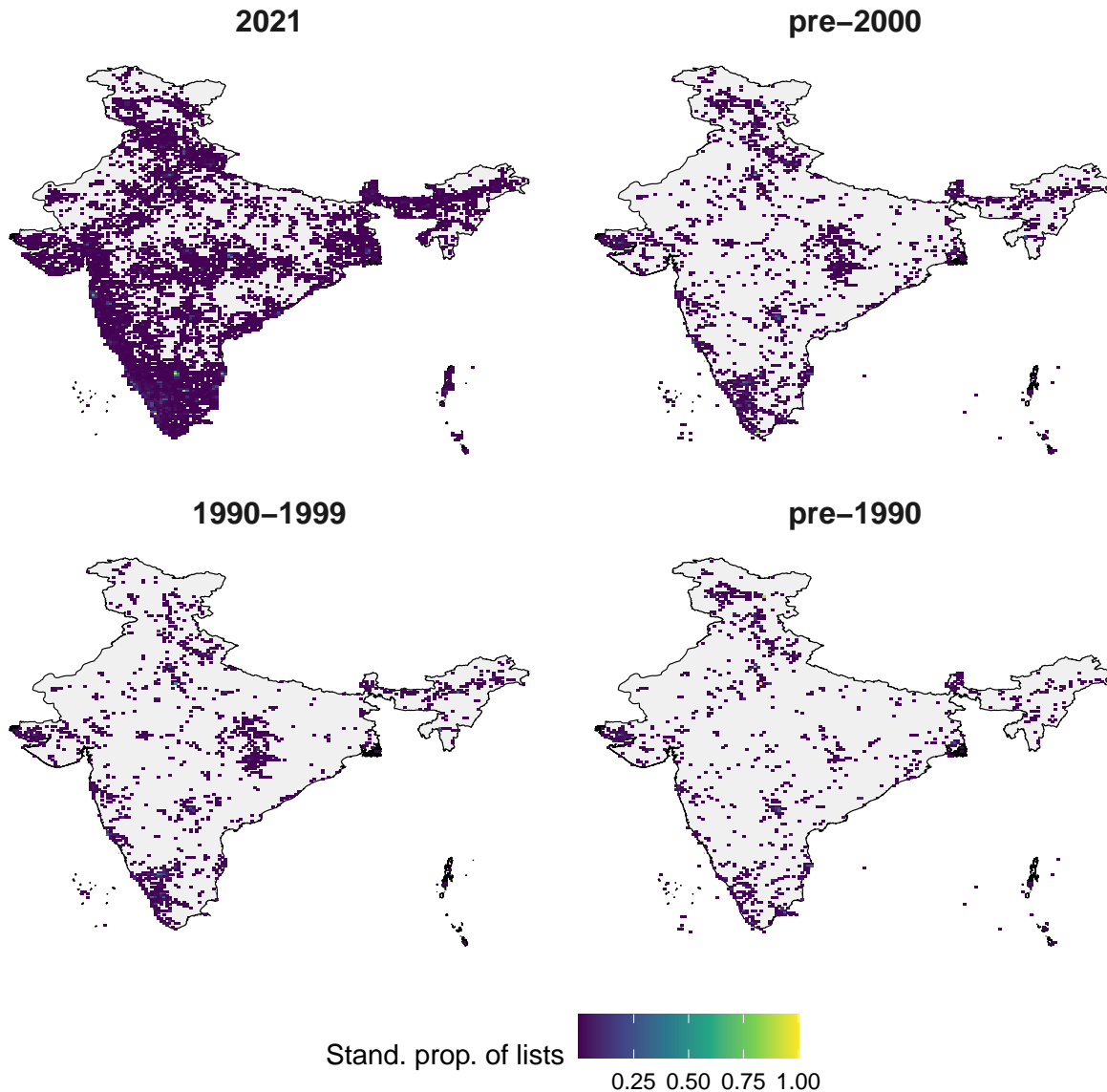


While sheer numbers of lists naturally differ considerably, there doesn't seem to be much of a spatial bias in pre-2000 birding, i.e., concentrations of birding effort across the country before 2000 were not very different from the current scenario, except for central India and West Bengal which lose a large spread of data.

Since sheer numbers are meaningless here, we calculate the proportion of birding in each cell and standardise it to the most active cell (cell with highest number of lists). This standardised

proportion allows meaningful comparison of *concentrations* of birding across *both time periods*: here the actual values are not so important, but comparing the locations of light-coloured cells will tell us if spatial concentrations changed.

How was birding spatially concentrated in different periods?

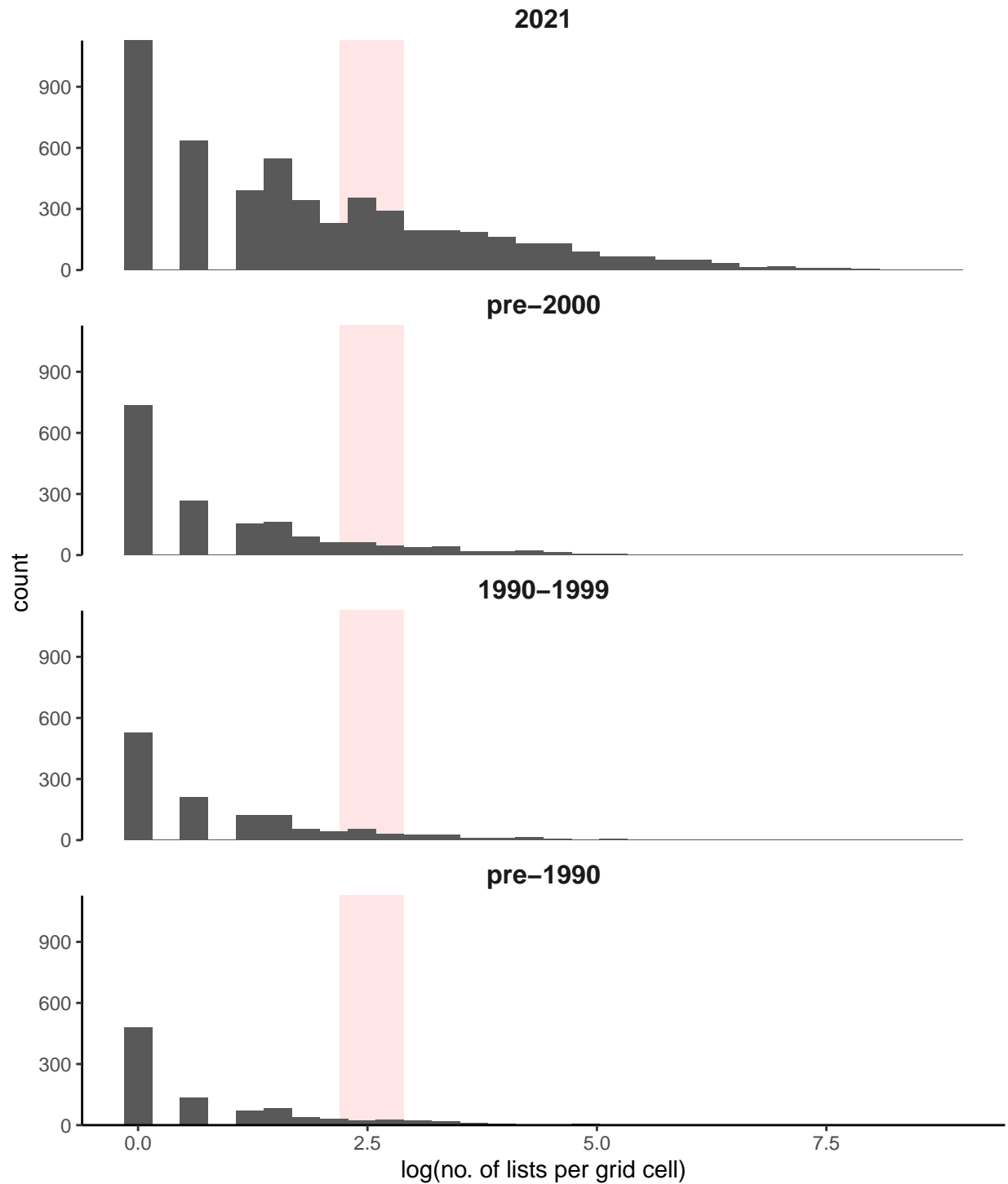


After splitting, some of the data spread remains consistent, especially in Western India. However, a majority of the data in Central India and southern Western Ghats (Kerala, Karnataka) are restricted to 1990-99, while Jammu Kashmir data is pre-1990.

Histograms

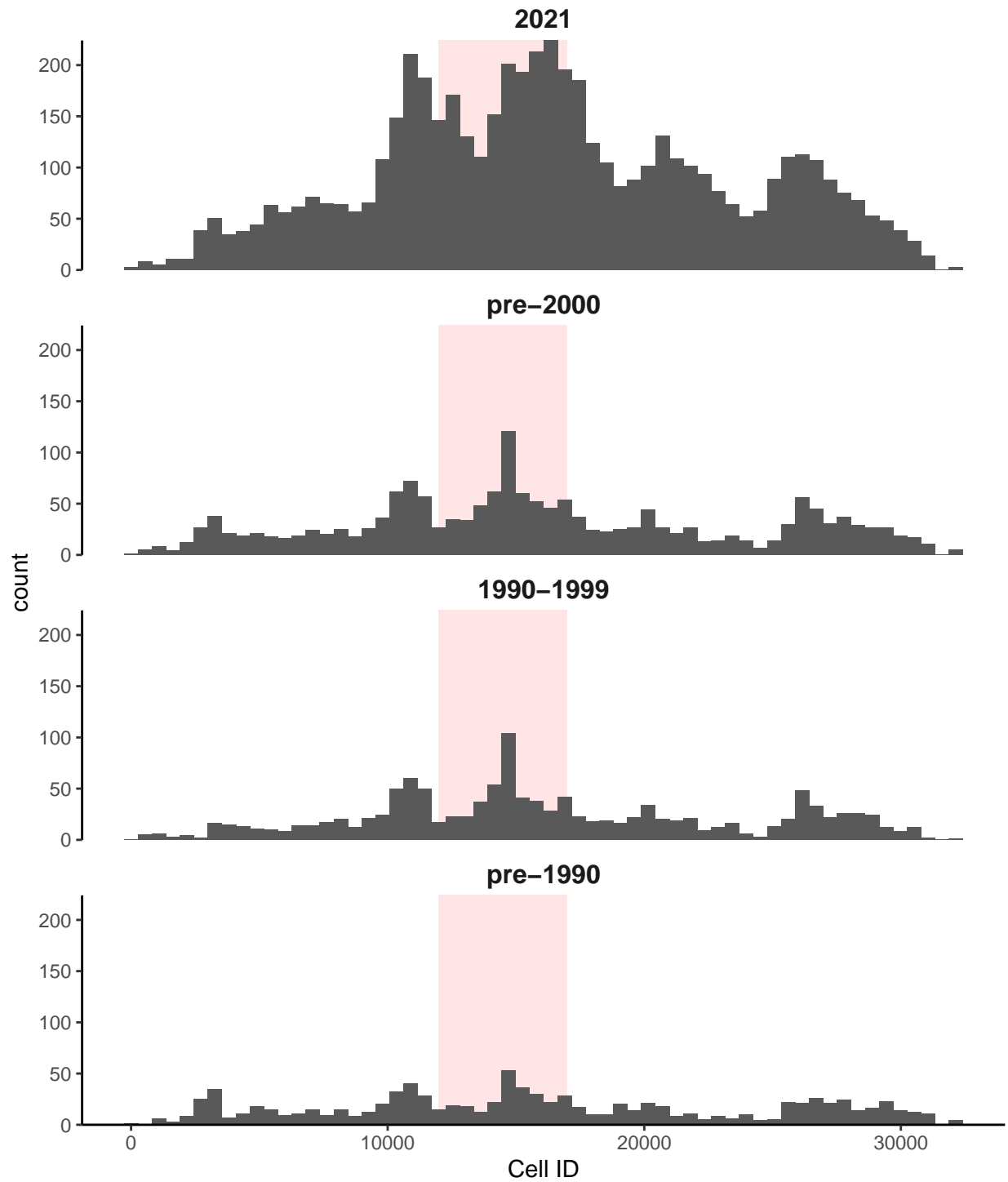
How do the histograms of birding intensity look across the two time periods?

Distribution of birding intensities



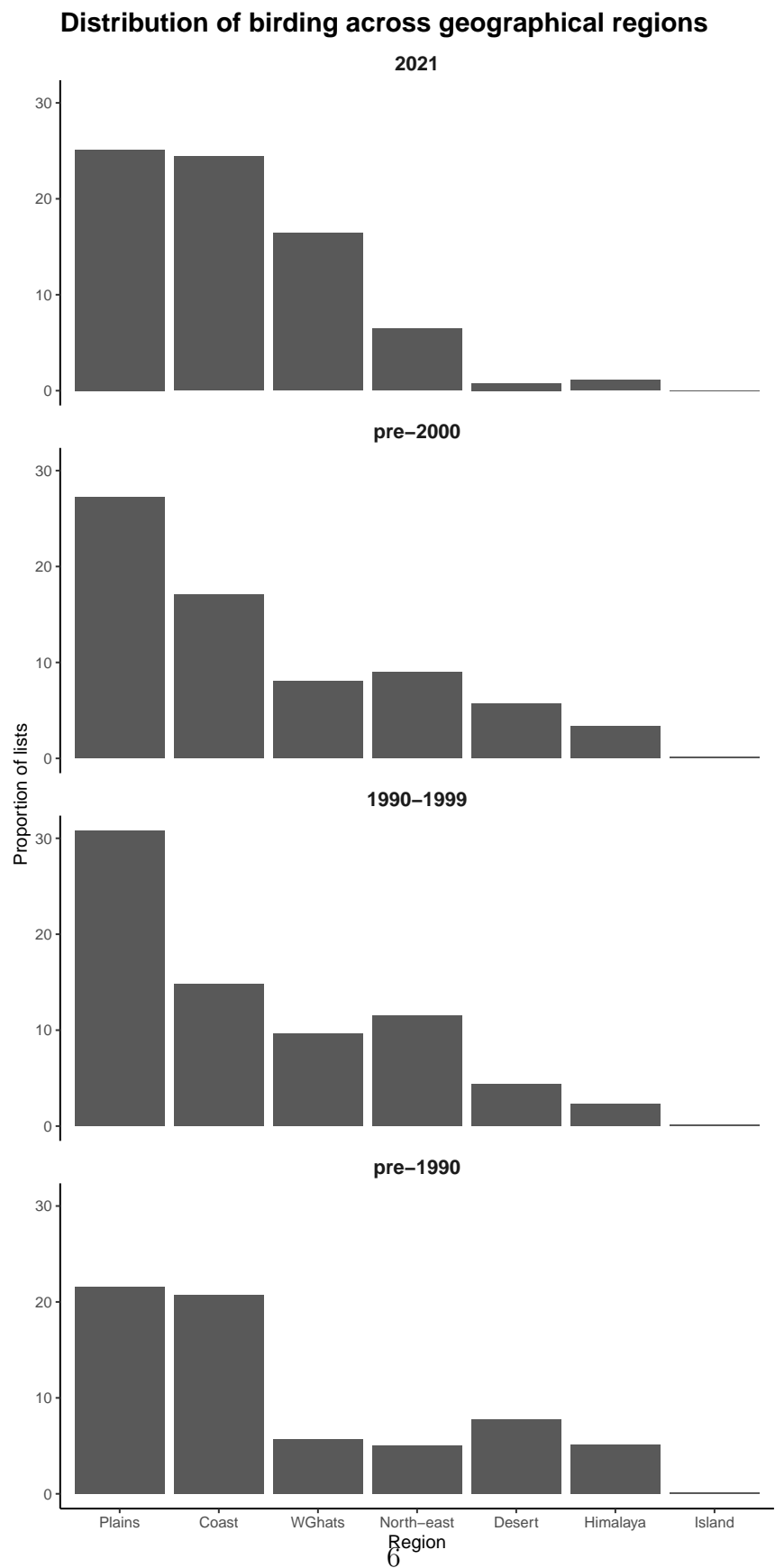
Bar heights are obviously different, but overall shapes are largely similar, except for a minor discrepancy (highlighted in red).

Distribution of spatial spread



Once again, there is a slight discrepancy, but there are 5 distinct humps in all periods.

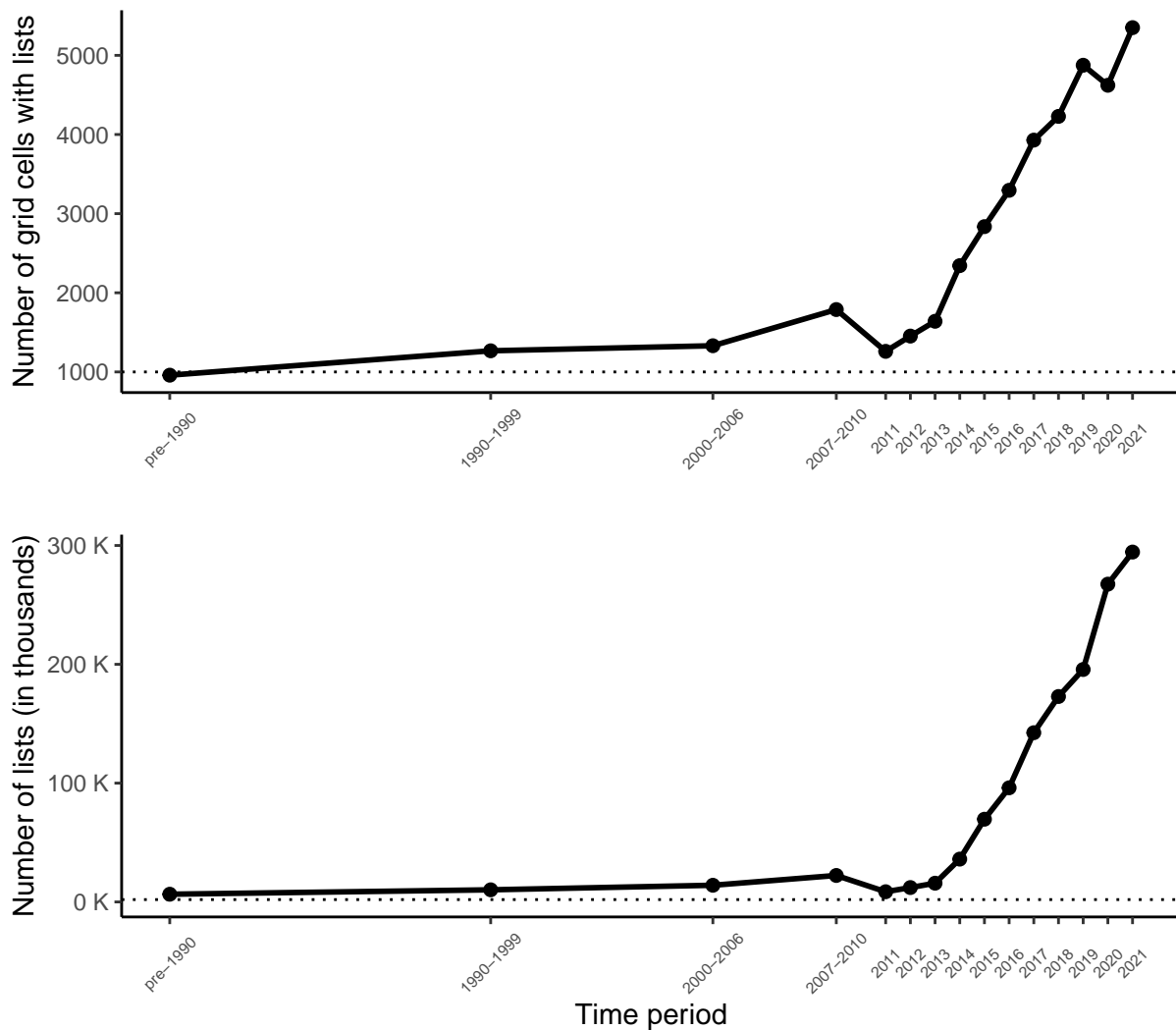
Data across geographical regions



Due to updated districts, our region information is not complete: there are several districts that do not map to any region. But ignoring the NAs, concentrations of birding across the various regions do not seem notably different. A slight difference is that the concentration in WG and coastal regions was lower in pre-2000s, with more focus instead in desert, Himalayan, island and northeast regions.

Progression of data with time periods

How much data in each time period?



All time periods have more than the minimum (2000, the dotted line) number of lists, but since this is before all the filters, it is likely not a good idea to split historical lists. Similarly, 2011-2013 should also be kept unsplit. It is a similar story with spatial spread: pre-1990 does not even have 1000 cells (dotted line) with data, before filtering.

Conclusion

It seems that we are currently **not** at the point where we can split historical data (pre-2000) into finer time intervals, due to insufficient data. However, **keeping all pre-2000 observations together**, there seems to be **no major bias** in spatial spread of the data.