22403800 c7083 Project

Katie Taris

2023-03-08

Bee Colonies in the USA

GitHub Link: https://github.com/ketaris/c7083_Project

00 Table of Contents

01 Introduction

The source of the data is: https://github.com/rfordatascience/tidytuesday/tree/master/data/2022/2022-01-11. The data is presented in two data sets named 'colony' and 'stressor'. The first dataset contains information of colony numbers per state per quarter from the period 2015 to 2021. The stressor dataset contains information regarding the stressors put on colonies and percentage impact they have.

`colony.csv` 10 variables as listed below.

Variable	Class	Description
year	character	year
months	character	month
state	character	State Name (note there is United States and Other States)
colony_n	integer	Number of colonies
colony_max	integer	Maximum colonies
colony_lost	integer	Colonies lost

Variable	Class	Description
colony_lost_pct	integer	Percent of total colonies lost
colony_added	integer	Colonies added
colony_reno	integer	Colonies renovated
colony_reno_pct	integer	Percent of colonies renovated

`stressor.csv` 5 variables as listed below.

Variable	Class	Description
year	character	Year
months	character	Month range
state	character	State Name (note there is United States and Other States)
stressor	character	Stress type
stress_pct	double	Percent of colonies affected by stressors anytime during the quarter, colony can be affected by multiple stressors during same quarter.

02 Data visualisation #1

```
## Warning: Removed 18 rows containing missing values ('geom_label_repel()').
```

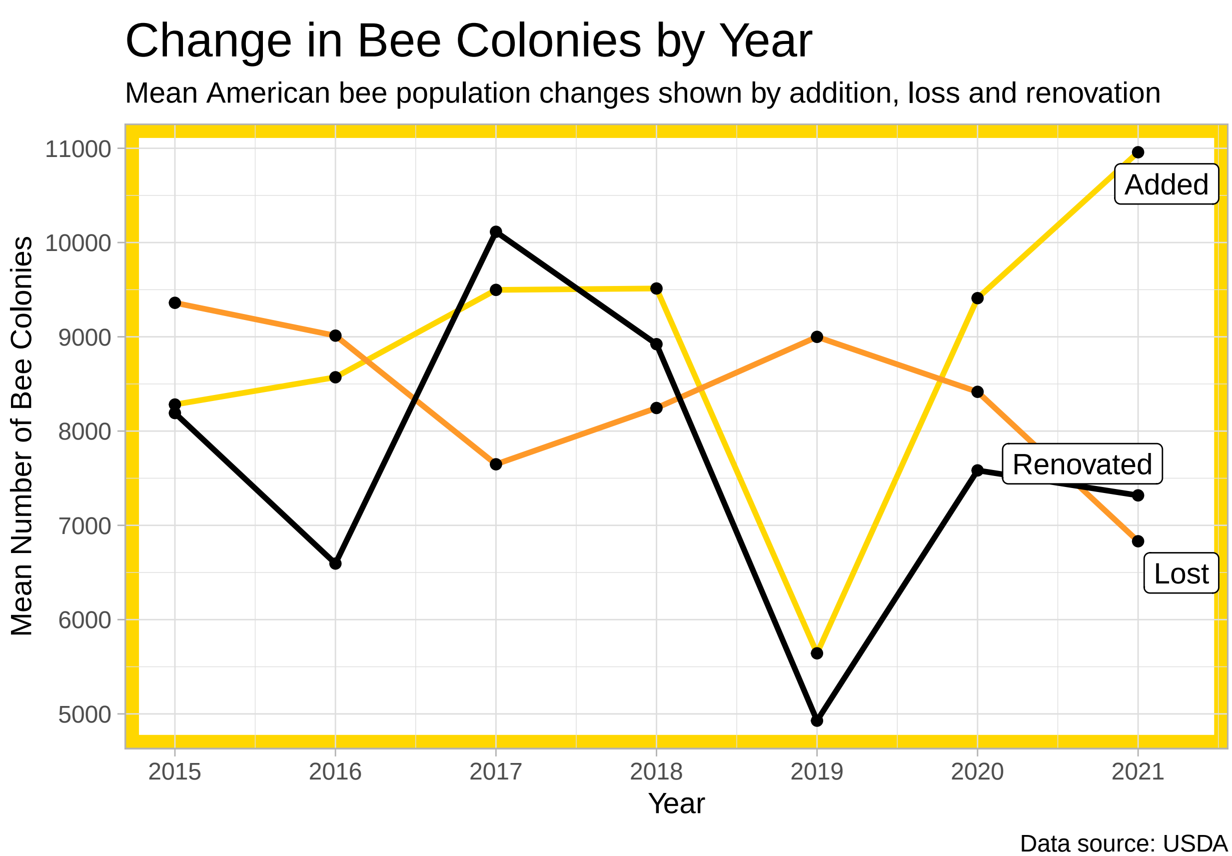

Figure 1: This line graph shows the rise and fall of the mean number of colonies added, renovated and lost over the 7-year period from 2015 to 2021. This graph is supposed to show overall trends and is not specific by state, but nonetheless some large trends can be gleaned, including a large reduction of addition and renovation to colonies and a large loss occurring in 2019. Some of the stressors that cause loss can be found in the next graph. The package ggrepel was used to place labels on the lines, instead of using a legend.

03 Data visualisation #2

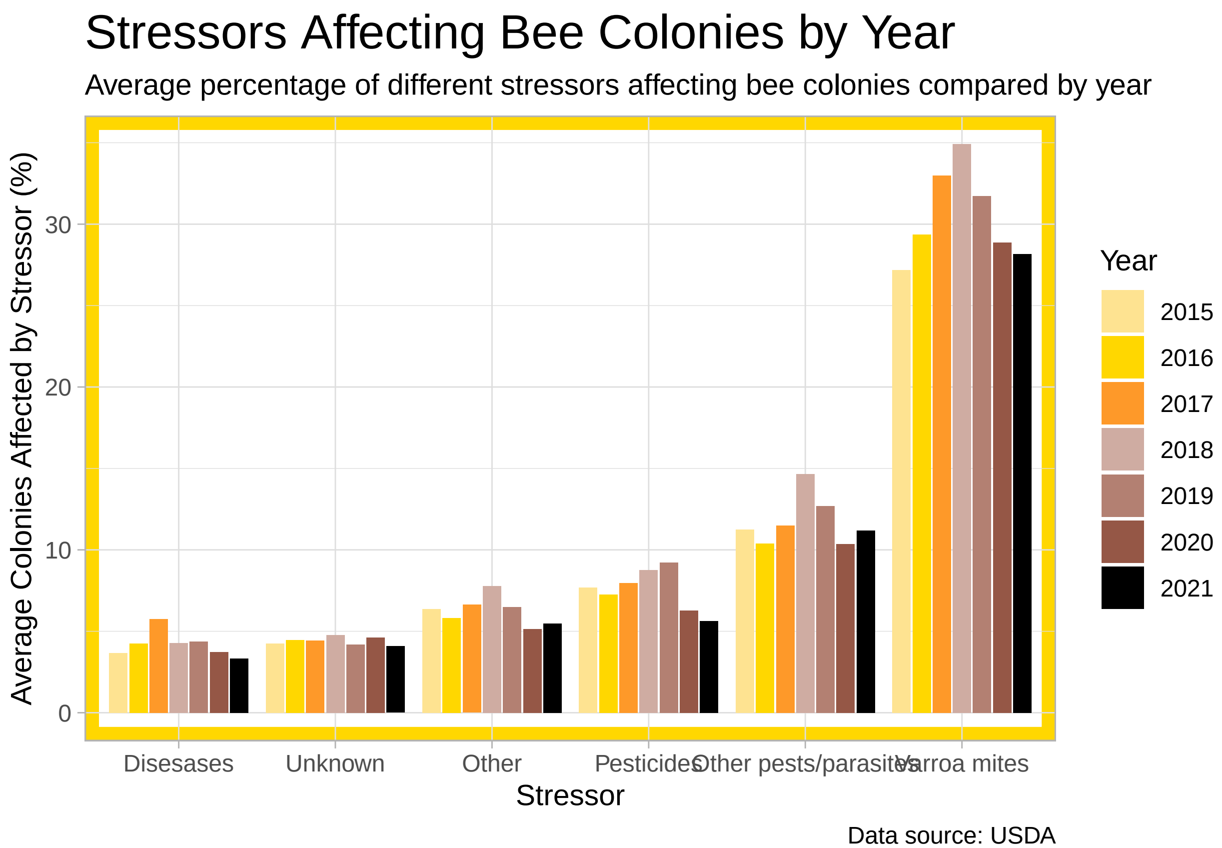

Figure 2: This bar graph demonstrates the average number of colonies affected by a given stressor over 7 years. The graph retains the categories of "unknown" and "other" to demonstrate that there is research that needs to be done in these areas, but also shows that diseases are even less prominent than these causes. The takeaway message here is that Varroa mites are the largest stressor on bee colonies in the US.

04 Data visualisation #3

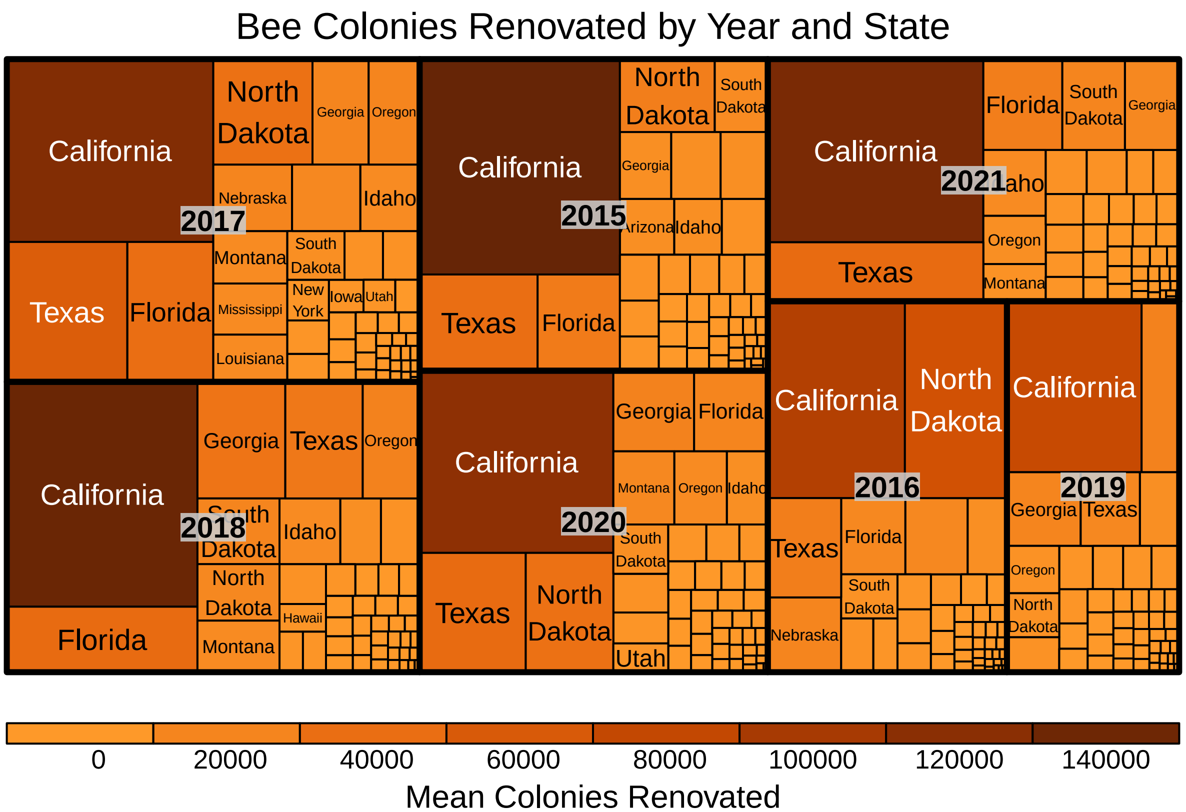

Figure 3: This tree map gives an insight into which states are renovating bee colonies and when this has been occurring. It is apparent that California has renovated the most colonies each year, with a few other states floating around the top 5 mark. This data visualisation is most effective to pick out the largest squares and make an immediate impact in picking out the top 5, but it is not intended that the reader is then able to rank each state in order by judging if one square is marginally larger than its neighbour.

05 Data visualisation #4

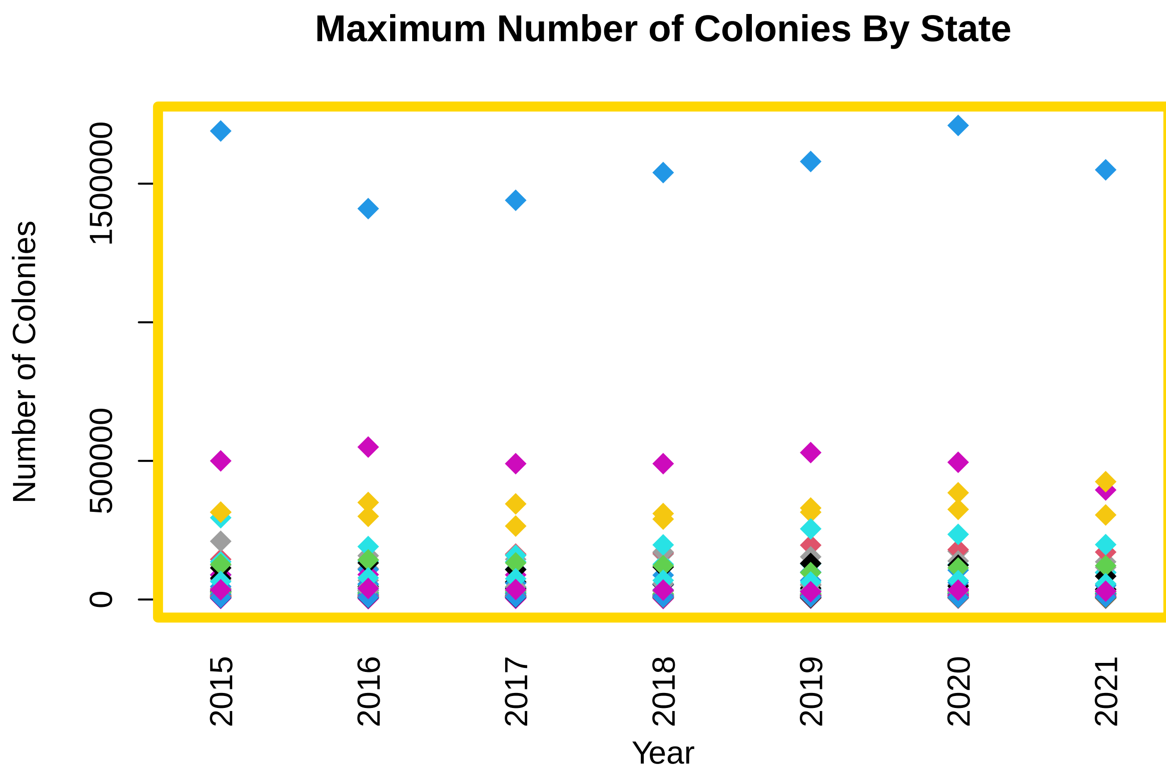

Figure 4: Having previously seen how many colonies were being added and subtracted, we can now use this interaction plot to get an overall picture of how many colonies there are. This graph does not have a legend, as the goal is not to match each colour to its state, but the takeaway message here is that there is one state that has an overwhelmingly higher number of colonies compared to the other states, and this higher number is consistent from the year 2015 up to 2021. The graph was created using base R.

06 Data visualisation #5

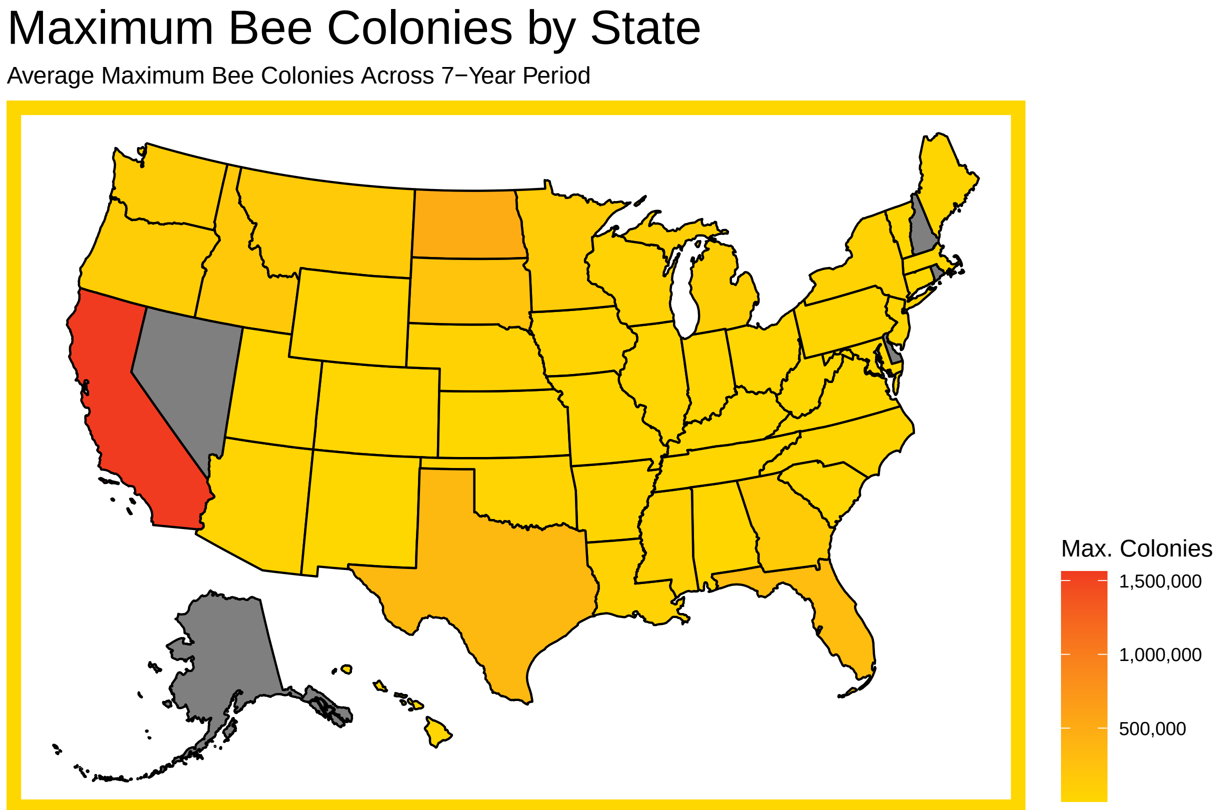

Figure 5: A chloropleth map of the US shows that averaged across 7 years, California has the highest number of bee colonies compared to all the other states. Since no other states are showing any red, this is consistent with the large difference of California and other states in the previous figure. Texas, North Dakota and Florida would be the next highest numbers, which is also consistent with the renovation numbers shown in the previous tree map.

07 Data visualisation #6

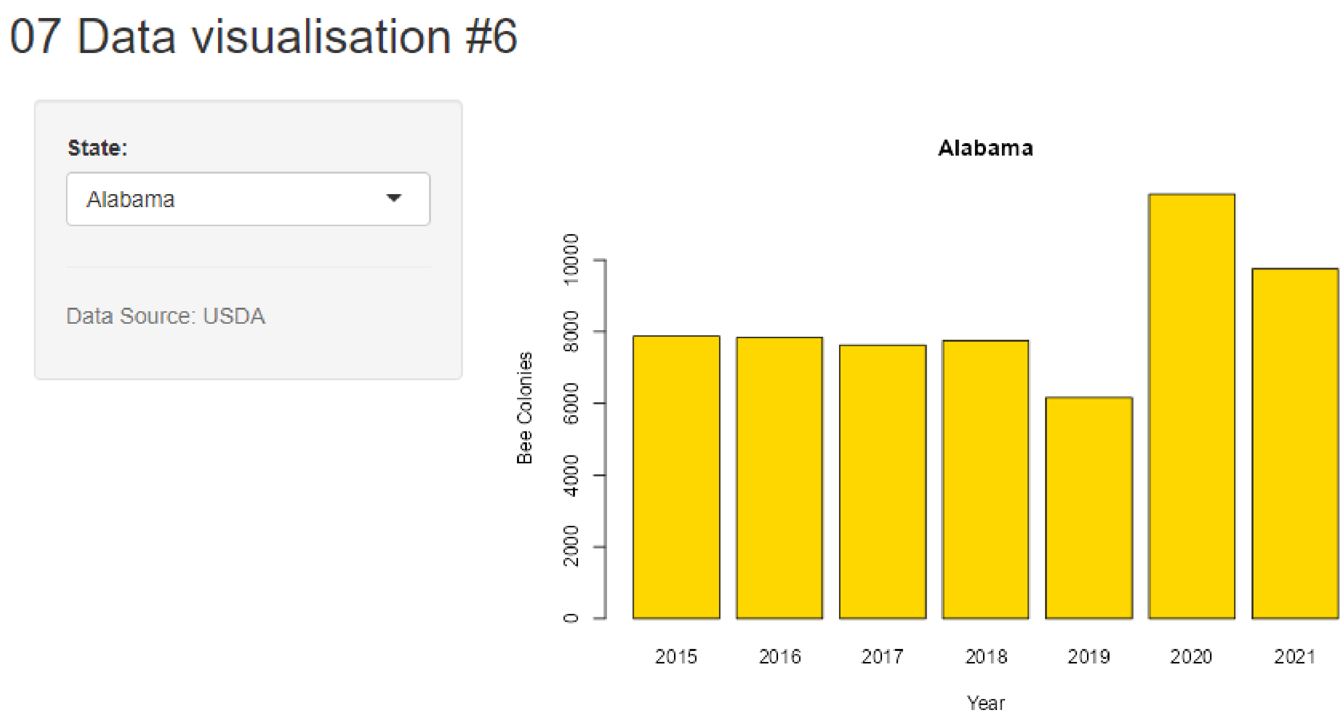

Figure 6: This interactive data visualisation expands on the previous graph and allows the user to choose a state and see the maximum number of colonies for each year from 2015 to 2021 on the barplot. This is an excellent way to display a large dataset containing 7 years and 50 states, which would be too much data to show individually on a graph and would become illegible. The app uses Shiny to create an interactive interface.

08 Graph Critique

Good Graph Figure 7: Please see annotated graph in the Appendix. The graph shows an example of a good data visualisation which demonstrates a set of data in a manner that is easy to read and follow what the author is trying to convey. The bars in the barplot allow the reader to understand the proportions that are being expressed without having to guess. The use of a rotated barplot allows easy reading of the land use titles. The figures are expressed as percentages and actual land areas which makes the data easier to interpret. Some of the boxes that are not of interest or are not the goal of demonstration are grouped as a grey colour. This is effective as it draws the reader's eye to the coloured boxes which are relevant here, and the author has not tried to over-complicate the graphic by using too much colour coding. The colours towards the bottom are coded to show green as crop production and red for livestock production, but the green on the third line from the top, is only a slightly different green while it represents all agriculture and not crop production, the similarities between the greens could be misleading. The graphic is explained with a descriptive title and subtitle. A good addition would be if the 71% ocean box could be to scale with the 29% of land use. This would need to be done in a way not to make the last two lines become very small and illegible, but having this box at full scale against the land would give a strong impression of the percentage of agricultural land use against all the surface of the earth.

Bad Graph Figure 8: Please see annotated graph in the Appendix. This barplot is an example of a bad data visualisation which needs alterations to make the information more easily available to the reader. The three dimensional aspect of the graph can be removed as it adds no positive effect. The shading adds confusion as the reader is not sure if they should be interpreting the line between the white and blue as a value in itself or not, which in this case is not showing any extra data and should be removed to become one solid colour. The y-axis is properly labelled, but it is difficult to compare the bars against the scale and obtain an actual value. Removing the 3D aspect will help this problem, but grid lines could be added at the major intervals to assist the reader in placing a bar's height against the scale. The maximum point of the axis could be brought down slightly as there is some space at the top of the graph that is wasted. If the bars could be zoomed in on, perhaps better differentiation of the first three bars would be possible. The labels on the x-axis are sufficient, and only writing every other year avoids clutter on the x-axis. The graphic has no title or subtitle, so while one can tell it is related to publications, it cannot be known publications of which subject. The use of a barplot could also be questioned, as the trend of increasing could be shown with a line graph using points and connecting lines.

09 References

Kudr, J., Skalickova, S., Nejdl, L., Moulick, A., Ruttkay-Nedecky, B., Adam, V., Kizek, R. (2015). Fabrication of solid-state nanopores and its perspectives. ELECTROPHORESIS, 36(19), 2367–2379. https://doi.org/10.1002/elps.201400612

Ritchie, H. and Roser, M. (2013) Land use, Our World in Data. Available at: https://ourworldindata.org/land-use (Accessed: March 9, 2023).

Wilke, C. O. (2020). Fundamentals of data visualization a primer on making informative and compelling figures. O'Reilly. Retrieved from https://clauswilke.com/dataviz/.

10 Appendix

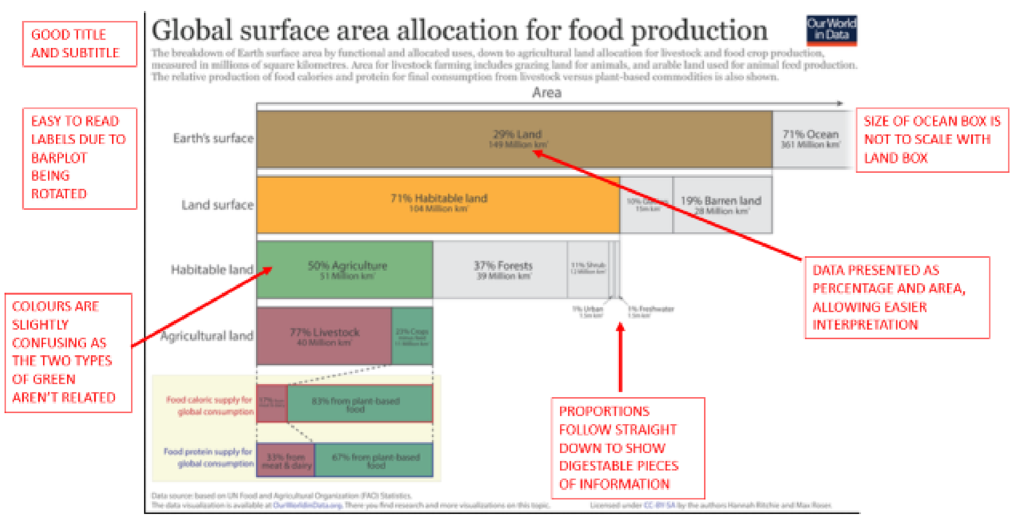

Figure 7

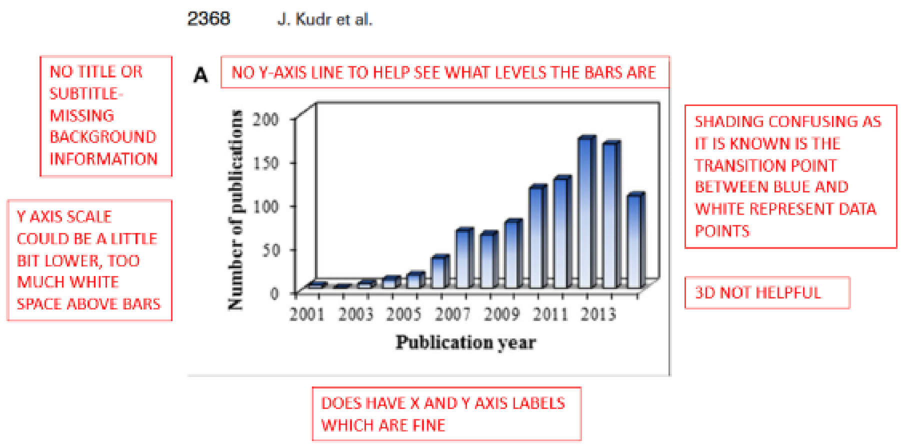
2368 J. Kudr et al.
A
NO TITLE OR SUBTITLE- MISSING BACKGROUND INFORMATION
NO Y-AXIS LINE TO HELP SEE WHAT LEVELS THE BARS ARE
SHADING CONFUSING AS IT IS KNOWN IS THE TRANSITION POINT BETWEEN BLUE AND WHITE REPRESENT DATA POINTS
Y AXIS SCALE COULD BE A LITTLE BIT LOWER, TOO MUCH WHITE SPACE ABOVE BARS
3D NOT HELPFUL
Number of publications
0
50
100
150
200
2001 2003 2005 2007 2009 2011 2013
Publication year
DOES HAVE X AND Y AXIS LABELS WHICH ARE FINE

Figure 8