Midterm Project: Honeybee Colony Loss Analysis

The USDA records data on bee colony additions, losses, collapses, and renovations each quarter. The various bee colony stressors are also recorded. The visualizations below explore the effect that stressors have on bee populations over time.
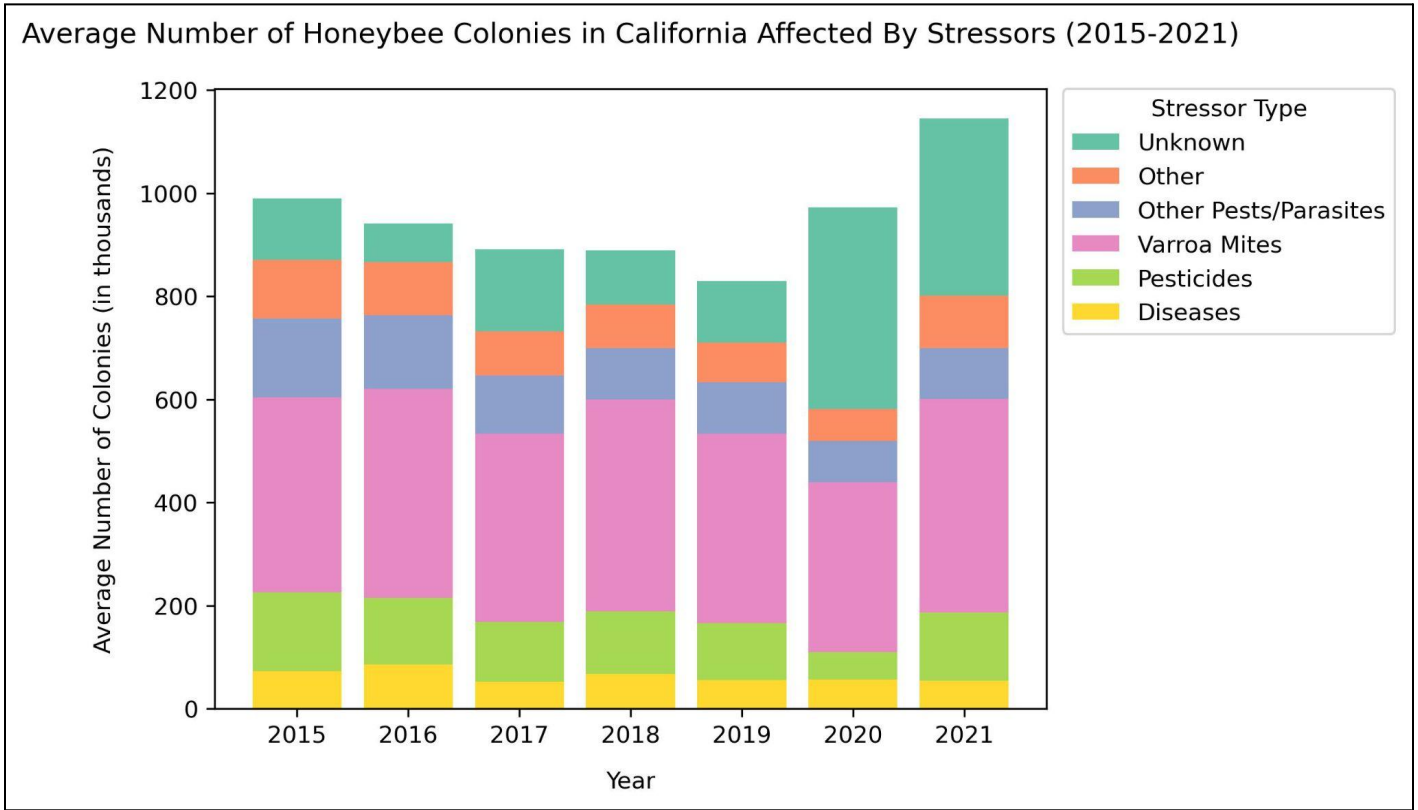
Honest Visualization:

Figure 1: Stacked bar chart showing the average number of colonies affected by each type of stressor per year in California from 2015 to 2021. It is shown here that the number of colonies in California varied year-to-year, but the proportion of colonies affected by each stressor did not change significantly. This data comes from the Bee Colonies USDA dataset, published in January 2022, and was collected quarterly, so the values shown in this graph are averages taken across each year.
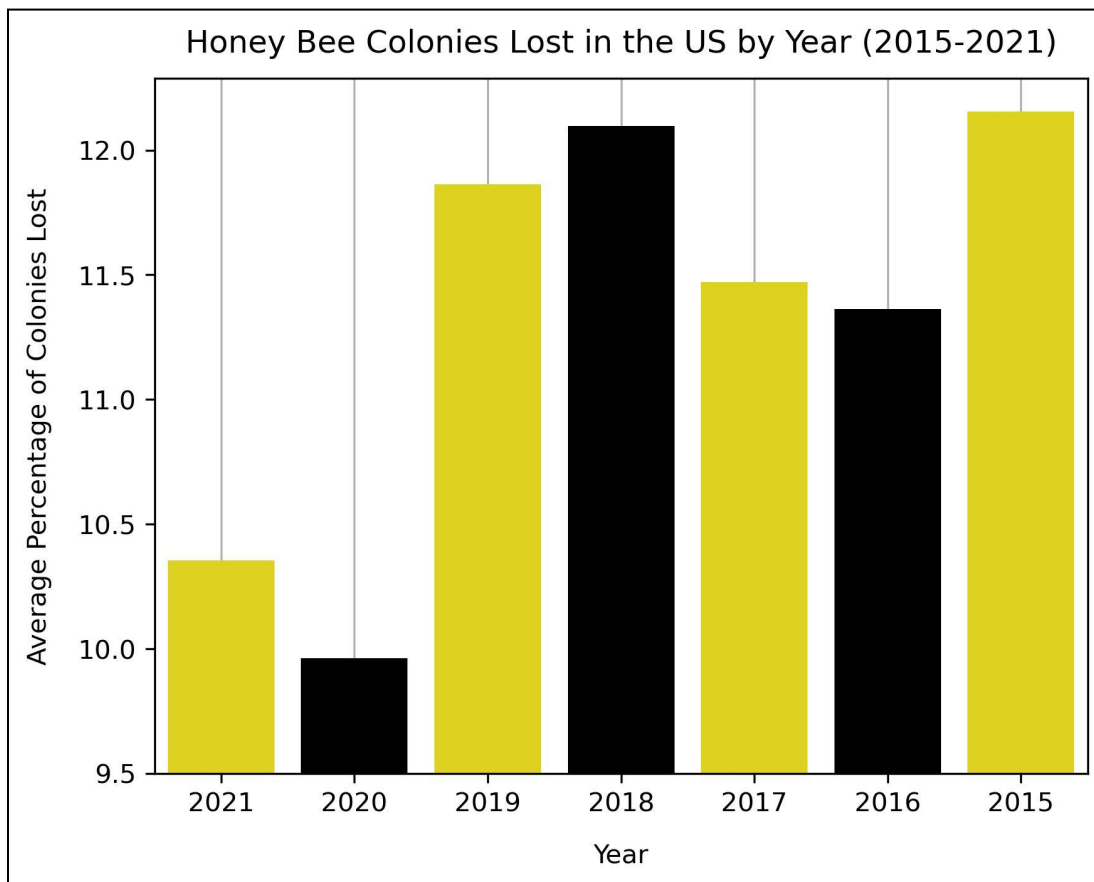
Deceptive Visualization:

Figure 2: Bar graph representing the average percentage of honeybee colony losses from 2015 to 2021 in the United States. Percentages are calculated as colonies collapsed out of the total colonies, representing the reduction in colony number each year. The percentage of losses was lowest in 2020 and highest in 2015, with an overall trend of decreasing colony losses. This data comes from the Bee Colonies USDA dataset, published in January 2022.

Deception: Truncation, reversed axis, misleading data.
While this is a valid graph with correct titles/axis labels made with numbers in the bee dataset, there are three primary misleading/deceptive techniques used here. (1) A truncated y-axis exaggerates the year-on-year difference in percentages. The minimum percentage on the graph is about 10%, while the maximum is about 12%, but the truncated y-axis makes this 2% difference seem much more significant than it actually is. The percentages actually stay relatively consistent from 2015-2021. (2) The reversed x-axis makes it look like the general trend is for bee colony losses to increase year-on-year, but the opposite is actually true. Logically, the years should be labeled 2015-2021, but the x-axis has them labeled from 2021-2015 in decreasing order. (3) The most misleading part of all is the percentages; this bar chart counts the wrong thing entirely. The values were calculated by taking an average of the colony loss percentages recorded each season, meaning that the actual percentage of annual

losses is actually higher; the bee colony loss percentages are weighted differently each quarter because the total colony counts change as more are added/lost. The values in the bar chart are still technically averages, but they are ultimately meaningless. Lastly, there are several unnecessary details in this graph. The grid lines in the background do nothing to help the viewer understand the data, and the colors are purely aesthetic (but are bee-colored!)

Citations:
Author/Publisher: USDA (United States Department of Agriculture)
Date Published: last updated January 11, 2022; published January 10, 2022
Dataset Title: Bee Colonies
URL:
https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-01-11/readme.md