

## Homework 2: Improving Effectiveness Rubric

**Your Name:** Pavan Arram

**Name of your classmate:** Devi Sowjanya

<b>Exploration creativity</b>	Full Marks (2.5 pts)	2.5 / 2.5 pts
	No Marks (0 pts)	

Comment: You explored an interesting question — linking sunshine hours to Airbnb occupancy — which demonstrates originality and practical relevance. The integration of external Airbnb data and a custom-joined dataset shows initiative.

<b>Plot expressiveness</b>	Full Marks (2.5 pts)	2.5 / 2.5 pts
	No Marks (0 pts)	

Comment: Graphs are very clear with suitable color pallets and there is no overlapping of datapoints which makes it easy to understand and express the insights of data.

<b>Plot effectiveness</b>	Full Marks (2.5 pts)	2.0 / 2.5 pts
	No Marks (0 pts)	

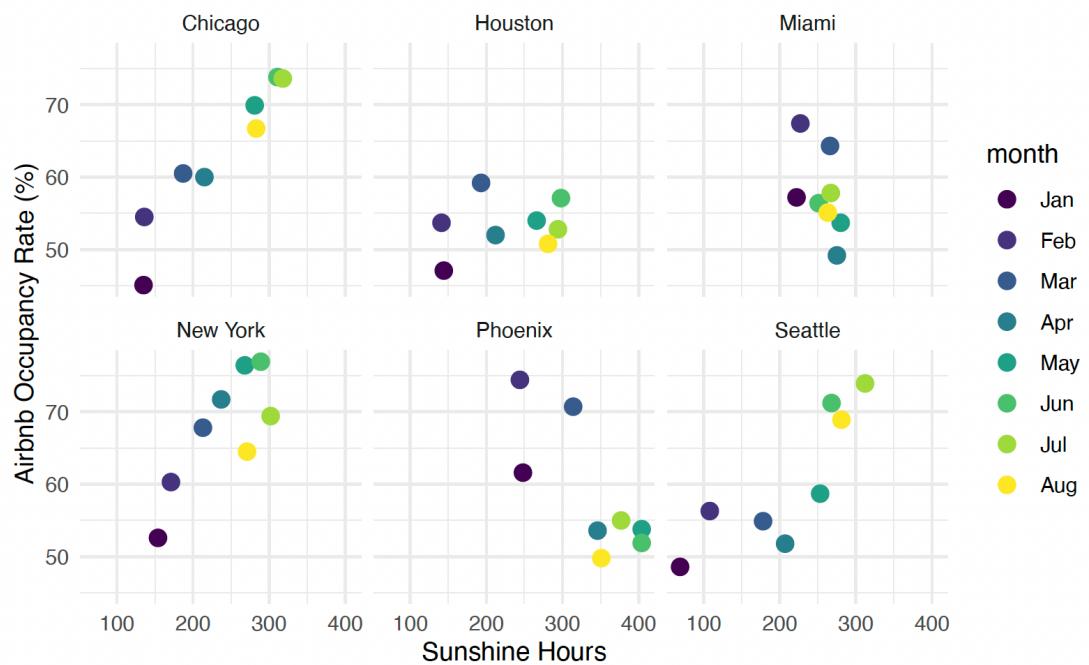
Comment: The plot shows relationships per city but doesn't strongly communicate "how" sunshine affects occupancy — readers must infer trends themselves. Add a smoother or fitted line per city, or a summary comparison plot .

<b>Plot description</b>	Full Marks (2.5 pts)	2.0 / 2.5 pts
	No Marks (0 pts)	

Comment: The write-up is descriptive and well-structured, but too focused on explaining ggplot syntax. Shift some focus to interpretation — what story does the data tell? What's the key takeaway for each city? Add 1–2 sentences about what surprised you or what might cause those differences .

## Plot before changes

How do sunshine hours influence Airbnb occupancy rates across different US cities from January to August?

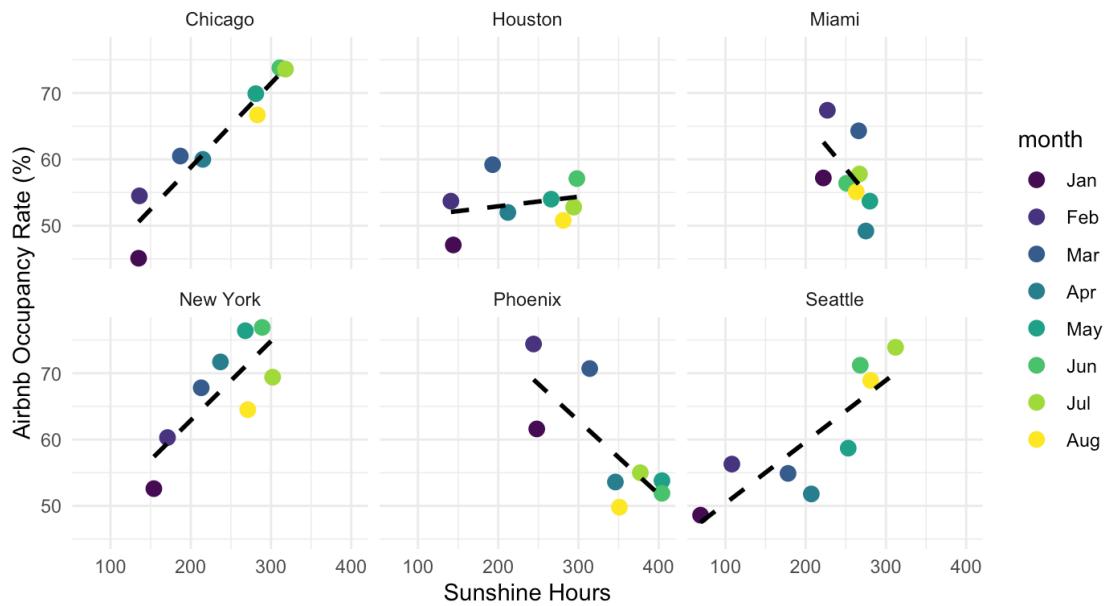


I initially created a faceted scatter plot using `geom_point()` to display the raw monthly data points for each city. The motivation for choosing the above plot was to visually explore how sunshine hours related to Airbnb occupancy rates across six U.S. cities from January to August, while preserving the individuality of each observation. I intentionally chose a faceted layout (`facet_wrap(~city)`) to separate cities into small multiples allowing localized patterns to be seen clearly without overlap while maintaining consistent x- and y-scales across all panels to support fair comparison between cities.

However, I received a feedback for the above plot that it although it displays the data clearly, but it is hard to interpret the trend / direction to reveal the effect between sunshine hours and Airbnb occupancy rates.

However, I received feedback that although the plot displayed the data clearly, it was difficult to interpret the overall trend or direction, making it harder to see how sunshine hours actually affected occupancy. To address this, I added a smoother layer (`geom_smooth(method = "lm")`) to each facet. This introduced a linear model fit that visually emphasizes the direction and magnitude of the trend within each city and the new plot is shown below.

How do sunshine hours influence Airbnb occupancy rates across different US cities from January to August?



The black dashed line in each panel represents this fitted regression line, which not only helps visualize points but also how sunshine hours affect occupancy. For example, in Seattle and Chicago, there is a strong positive association as sunshine hours increase from winter (January) to summer (August), occupancy rates rise sharply, indicating higher booking activity during sunnier months. In contrast, Miami shows slightly negative slope, with relatively higher occupancy during winter months, likely because its climate remains warm year-round and attracts visitors escaping colder regions. Similarly, Phoenix shows a negative trend, suggesting a drop in occupancy during extremely hot, sunny months.