Visual story telling part 2: Capital Metro data

2024-08-14

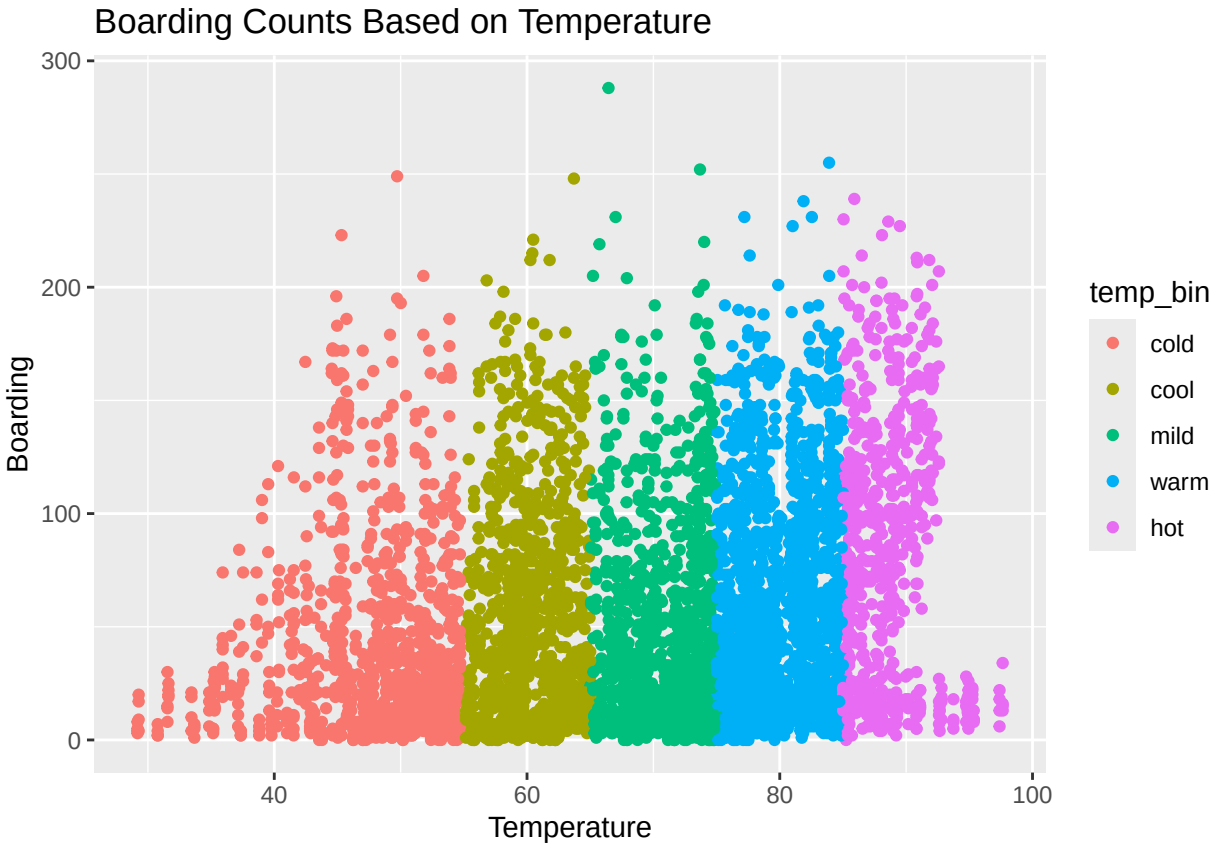
The file capmetro_UT.csv contains data from Austin's own Capital Metro bus network, including shuttles to, from, and around the UT campus. These data track ridership on buses in the UT area. Ridership is measured by an optical scanner that counts how many people embark and alight the bus at each stop. Each row in the data set corresponds to a 15-minute period between the hours of 6 AM and 10 PM, each and every day, from September through November 2018. The variables are:

timestamp: the beginning of the 15-minute window for that row of data
boarding: how many people got on board any Capital Metro bus on the UT campus in the specific 15 minute window
alighting: how many people got off ("alit") any Capital Metro bus on the UT campus in the specific 15 minute window
day_of_week and weekend: Monday, Tuesday, etc, as well as an indicator for whether it's a weekend.
temperature: temperature at that time in degrees F
hour_of_day: on 24-hour time, so 6 for 6 AM, 13 for 1 PM, 14 for 2 PM, etc.
month: July through December

Objective: Your task is to create a figure, or set of related figures, that tell an interesting story about Capital Metro ridership patterns around the UT-Austin campus during the semester in question. Provide a clear annotation/caption for each figure, but the figure(s) should be more or less stand-alone, in that you shouldn't need many, many paragraphs to convey its meaning. Rather, the figure together with a concise caption should speak for itself as far as possible.

Note: You have broad freedom to look at any variables you'd like here – try to find that sweet spot where you're showing genuinely interesting relationships among more than just two variables, but where the resulting figure or set of figures doesn't become overwhelming/confusing. (Faceting/panel plots might be especially useful here.)

First, we plotted the relationship between boarding rates and temperature, with temperature wrangled into bins (cold, cool, mild, warm, hot).

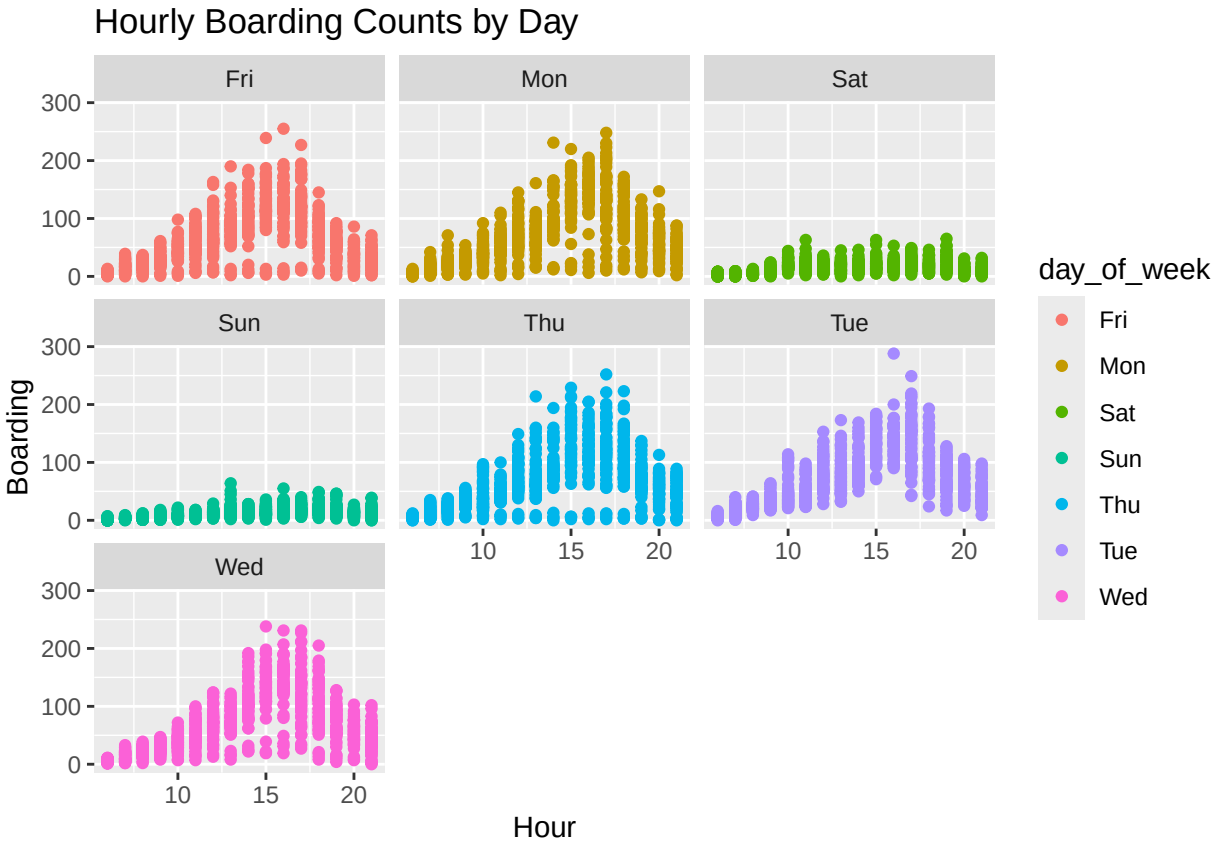


It looks like boarding rates are especially low at below 40 and above 90 degrees, perhaps due to extreme weather conditions. Also, these cold temperatures are often observed during Thanksgiving break, which is around November 20 - 25 at UT. Students and professors are not on campus as much during this time.

Boarding rates seem to increase slightly as the temperature gets warmer. We see that when the temperature reaches the 90s (likely in September), boarding increases probably because people cannot stand the heat and opt for air conditioned public transport.

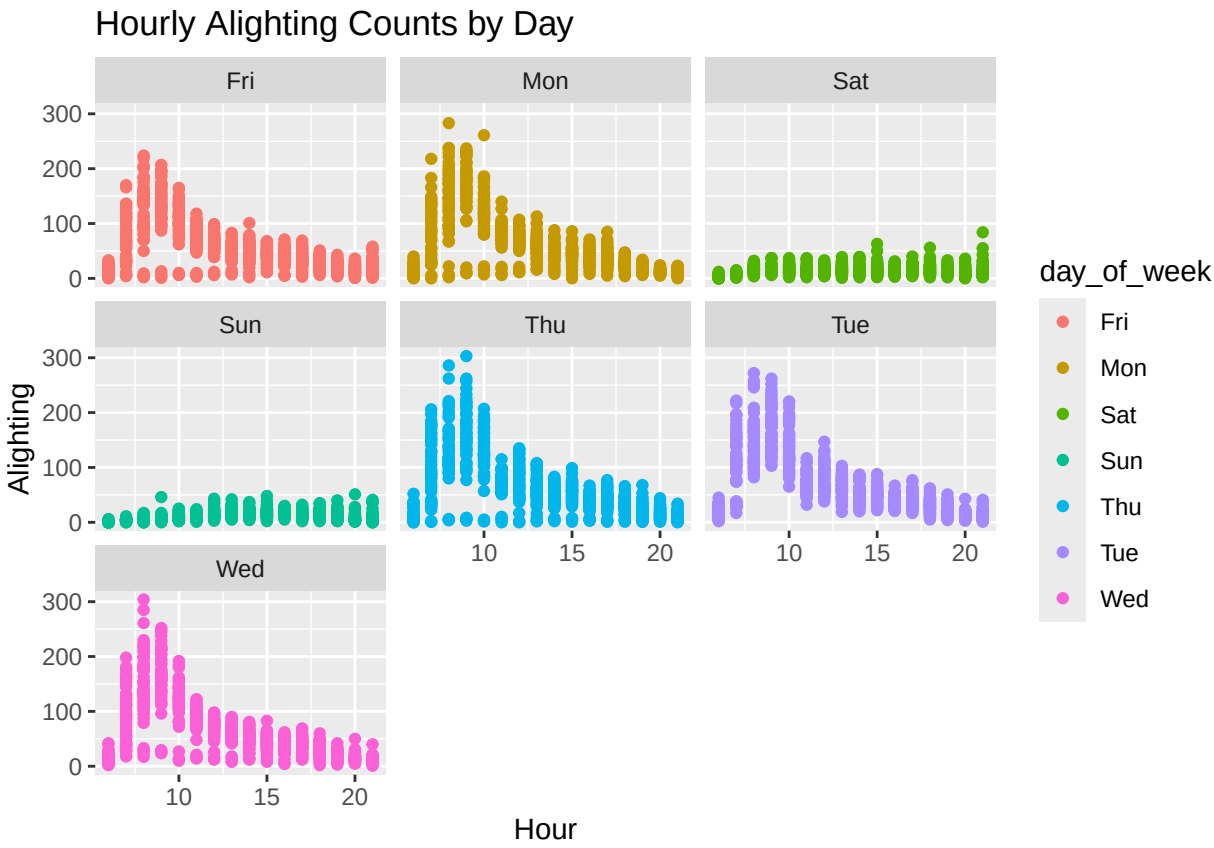
Therefore, the relationship between boarding and temperature tells us not only the effect of weather on ridership, but also season, and thus, campus traffic.

Next, let's look at the relationship between hourly boarding rates and the day of the week.



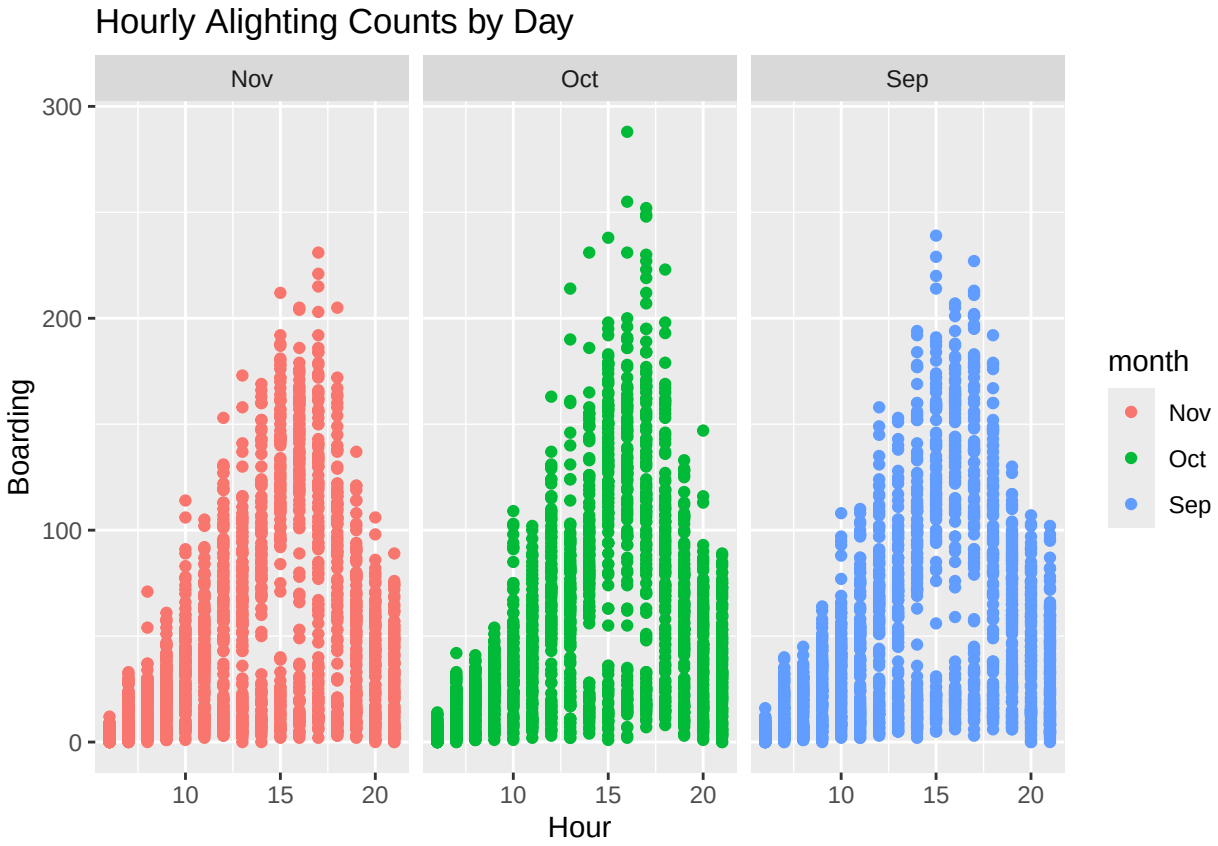
This collection of scatterplots tells us that boarding count is low on Satudays and Sundays, and follows a very similar trend from Monday through Friday, with a peak in boarding from hours 15-17, or 3-5 PM. This tends to be rush hour time for workers and likely the time of day that students finish their classes, so people are boarding to go home.

We applied the same process to alighting rates.



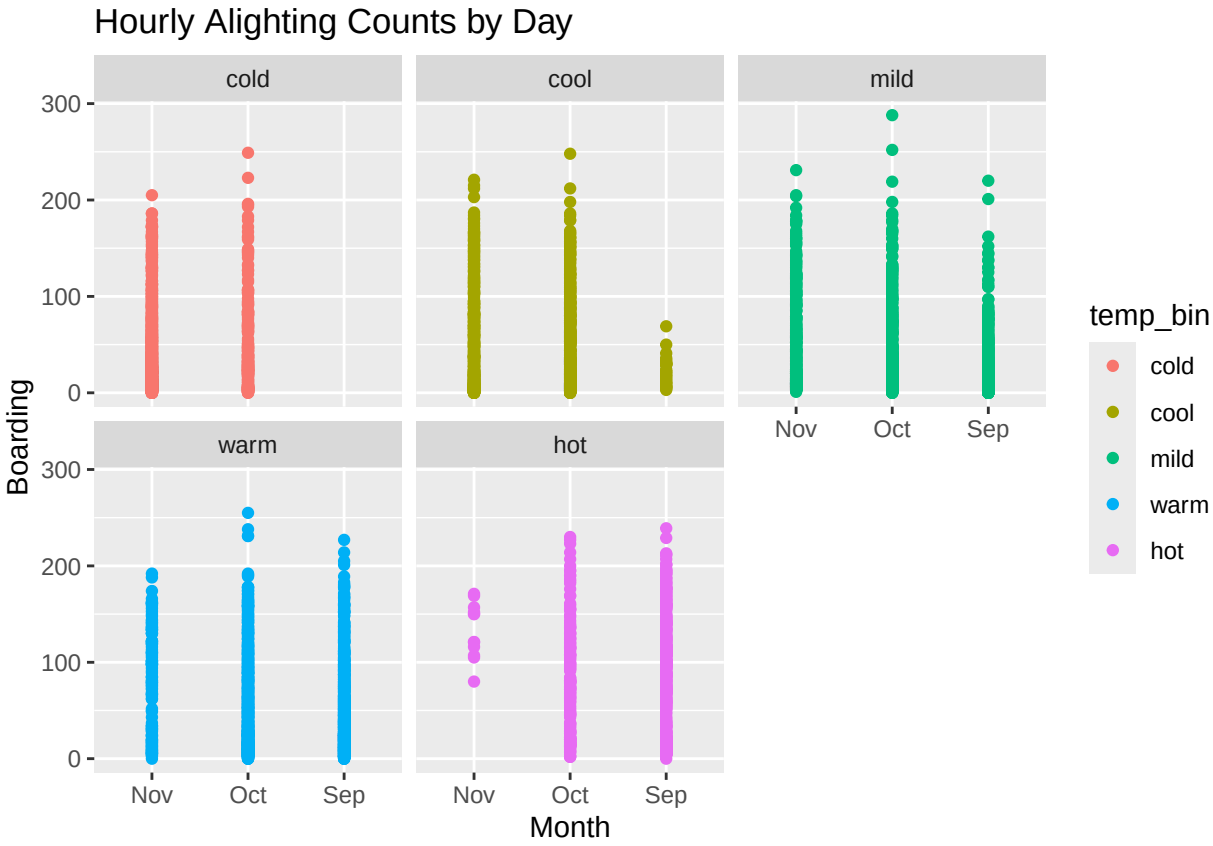
In contrast to boarding, alighting peaks at around 5-10 AM, indictaing the rush of staff and students to work and school. Again, Saturday and Sunday sees significantly lower ridership. We also notice that Friday has a lower peak at around. 200 instead of 300, perhaps due to hybrid workers who stay home that day and the fact that many students don't have class on Fridays.

Let's see if boarding times differ month to month:



Boarding seems fairly even across all three months, but the month with the highest peak in ridership is October.

Then, we combined month, hour, and temperature on a "facet wrapped" plot for analysis.



This plot confirms prior obervations. We see an overall peak in ridership in October when the weather is mild, and a slight uptick in boarding counts when the weather goes from cold to hot. In general, however, ridership is pretty consistent, with little observation of extreme dips in boarding (except for extreme weather conditions).

Conclusion

Given all of this information, CapMetro can better understand the seasonality and hourly patterns in ridership, catering their services to meet demand during those times and improve customer service.