Clayton Nagle

UCSD Extension Data Science Bootcamp

20 March 2021

Homework 6; 3 trends in weather data

1. The first observable trend is the very clear shift in temperature with change in latitude. There’s a very clear hyperbolic relationship between temperature and latitude, with only a handful of errant points lying outside of the curve of best fit. Also interesting about this temperature curve is that it appears to have two ‘peaks’ right around 20N and 20S, where the hottest points are. I would have suspected based on past knowledge and the general fit curve that the hottest places would be right at the equator, but for whatever reason (maybe just this specific day) that isn’t the case. Maybe this has something to do with certain hemisphere trends? I.e. air circulations that don’t cross the equator.
2. When looking at the humidity graph, there’s a noticeable blank spot, again roughly between 20N and 20S, all the way up to 60% humidity. This implies that places in this range are always at 60% humidity or above. This could be an artifact of the data sampling process, or again due to random chance on the day, but it’s very strange to see humidity values ranging from practically - all the way to 100 at most other occupied latitudes, then nothing in this middle belt from 0% to 60% humidity.
3. This was noted in the linear regression analysis, but there is something strange about the way cloud data is classified. I’m not sure exactly why things are defined this way, but there are very clear bands that show the overwhelming majority of data points at 0%, 20%, 40%, 75%, 90%, and 100% cloudiness. I’m not sure the exact reason why (perhaps the OWM defines cloudiness in some kind of rounding way?) but these distinct, discrete delineations of cloud data are not found in any other data relationship.