Progress report

So far we have had a couple meetings as a team, a meeting with the instructor and a meeting with the TA which was really helpful in terms of having a direction of how to use the data and what results we should expect under certain circumstances.

After looking at the data, we have had a few ideas of how we should use the data. First, under "collection date" there are many ####### useless data, we could clean the dataset first by dropping all the ######### symbols. Then we can visualize the data by creating a US map that contains colors, which would show how many people were getting affected in each area. Each area will have a certain percentage of transparency, from light to dark. Next step, we need to correlate the sequence with the map, by showing how they are related in which state, what kind of sequence happened the most in which area as well as how the temperature might have some effects etc..

Also, we want to find out how sequence and population density are related, and hopefully we could come to a conclusion about how those sequences are correlated with the environment. We might use python to try it out as well if we have enough time since we just picked up R this semester.