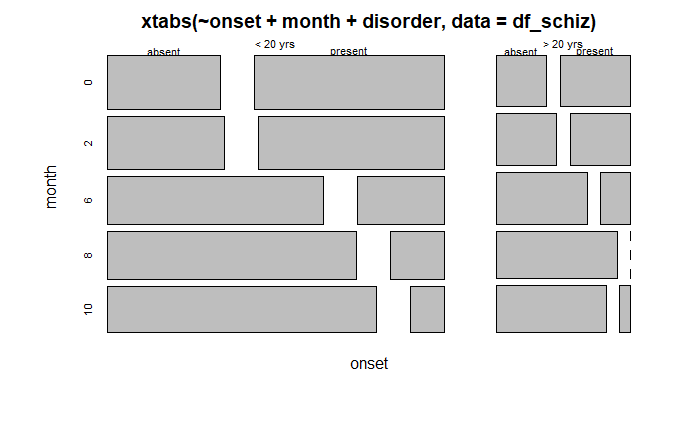
Final Project- Part 2

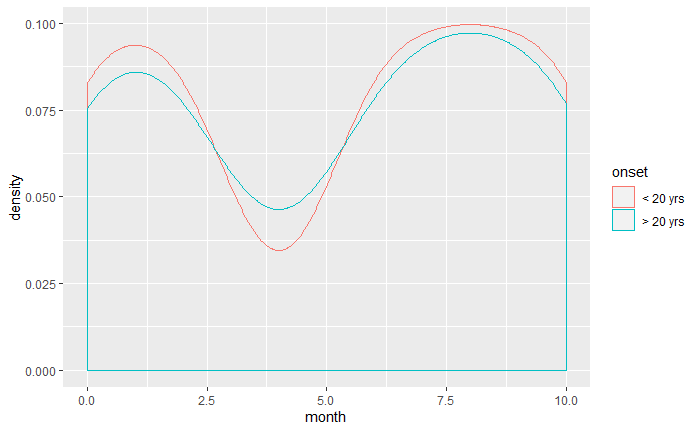
1. **Initial Analysis:**

In the initial analysis of the “schizophrena.csv” dataset, I gathered some basic statistics on the onset variable. I discovered that the mean (5.2,5.2) and standard deviation (3.72,3.74) for each of the two onset variable groups (<20 yrs, >20yrs) in relation to the numbers of months after hospitalization is about the same. Considering this, I felt that additional analysis was needed to verify whether the course of the illness differed between patients with early onset (< 20 yrs) as opposed to late onset (> 20 yrs).

1. **Model/Plot Section:**

Mosaic plot: This mosaic plot below displays how onset (early and late) relates to the number of months a patient has been hospitalized. If appears that the fewer months a patient has been hospitalized, the higher the number of patients with the illness present. Conversely, the more months spent in the hospital, the more occurrences of the illness being absent for both onset groups. I also created a density curve to further confirm the occurrences of both onset groups were comparable.





1. **Final Conclusion:** Based on my analysis, I have come to the conclusion that there is little difference between the course of the illness for patients with early onset versus late onset. The distinguishing factor appears to be the amount of time the patient has been hospitalized.