PS6 McGuire

Joel McGuire

March 2018

1   Cleaning and Transformation Steps

1. Discretization of Continuous variables

2. Recode NA as factor level

3. Flatten Lists and extract values

4. Create date object, then create duration variable

5. Remove unnecessary variables

6. Transform distributions to normal

2   Visualizations

Figure 1 shows that the most goals are lost relatively soon after they are created. This suggests that if the duration of a goal passes a certain threshold it will be more likely to succeed than fail.

Figure 2 shows that the distribution of goals that are won and loss appear to be different when sorted by month created. That there is not a uniform distribution across all types of time horizons (present term, near term, and long term) suggests a strong seasonal component.

Figure 3 displays the relationship between how much someone has paid for going off track of their goals, and the density of goal duration. Goal duration is exponentially distributed by becomes less skewed to the left as users pay more. This also suggests that the distributions of the duration of winning and losing goals are roughly similar.
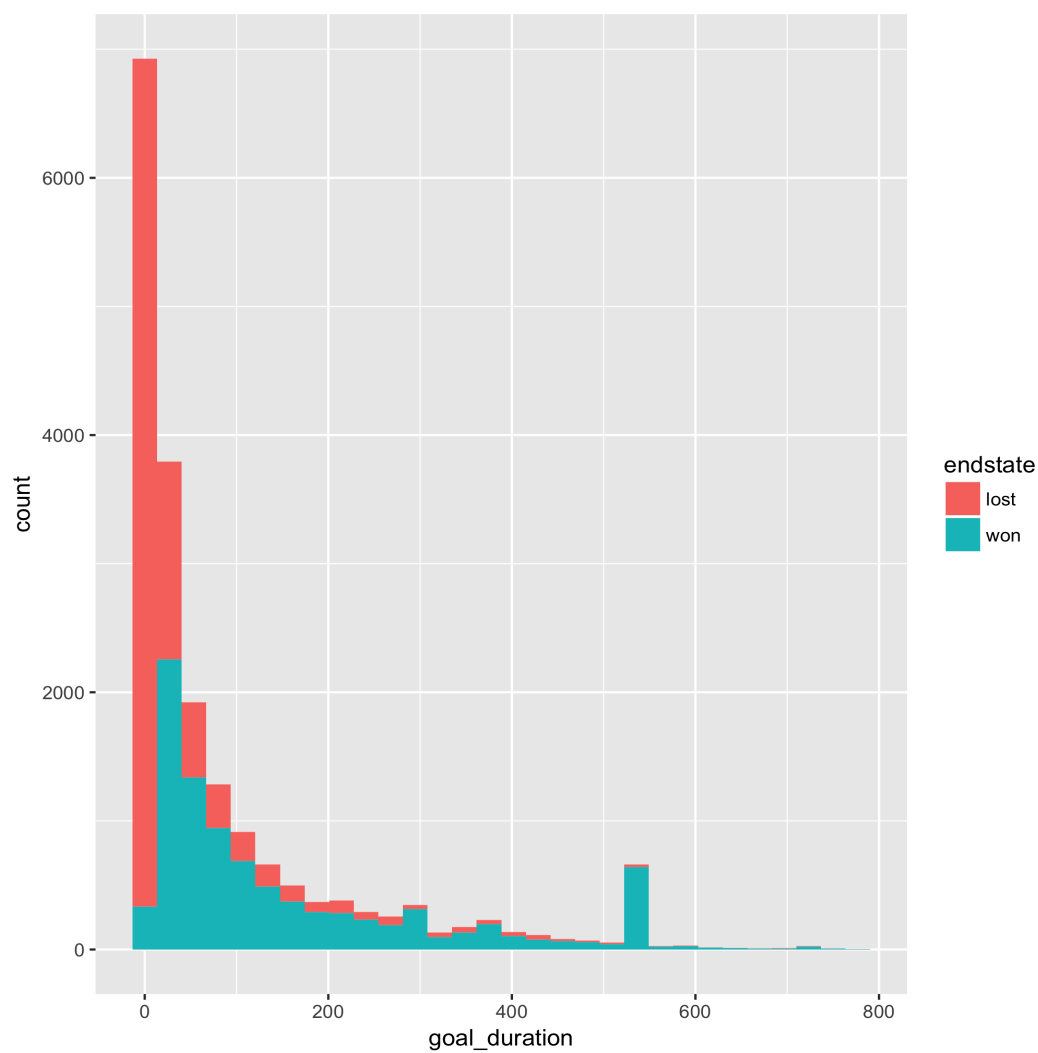
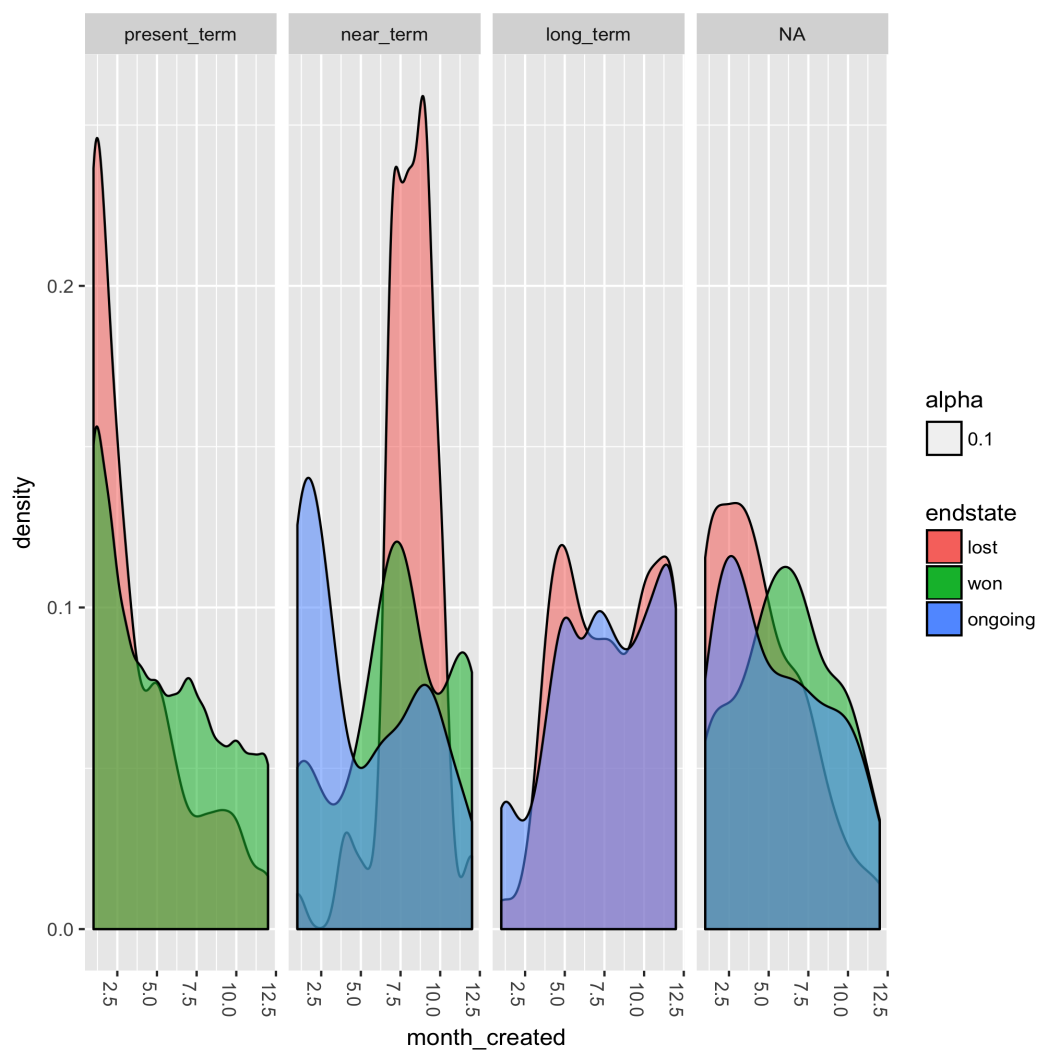Figure 1: Histogram of Duration and Endstate

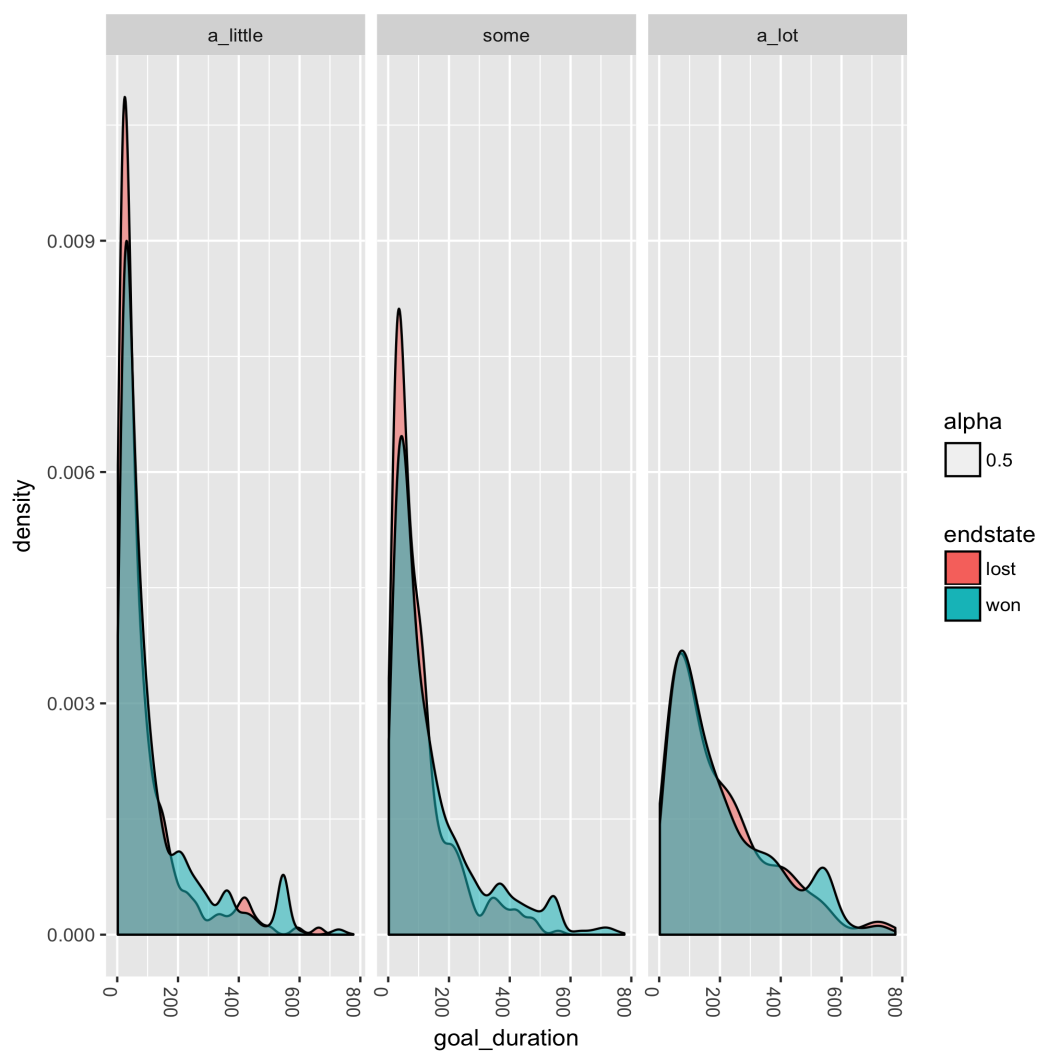Figure 2: Density of Goals Created by Month of Year, Bracketed by Time Horizon

Figure 3: Density of Goal Duration Bracketed By Amount Paid