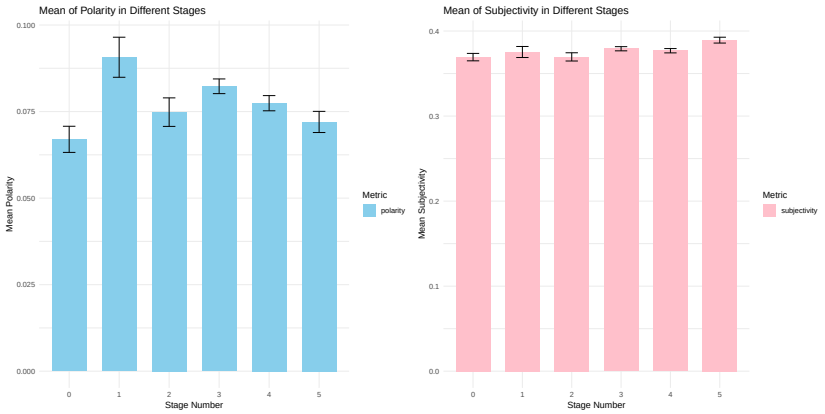
1 Mean



Here are the mean of polarity and subkectivity in different stages. As noted in the github, the higher the polarity score is, the more positive the text is. The higher the subjectivity score is, the more subjective the text is.

2 Linear Regression

From the plot above, we assume there is a correlation between the stages Austin is in and people's sentiment. So we performed a linear regression. Our hypothesis model is:

$$\text{polarity} = \beta \times \text{stage number} + \mu$$

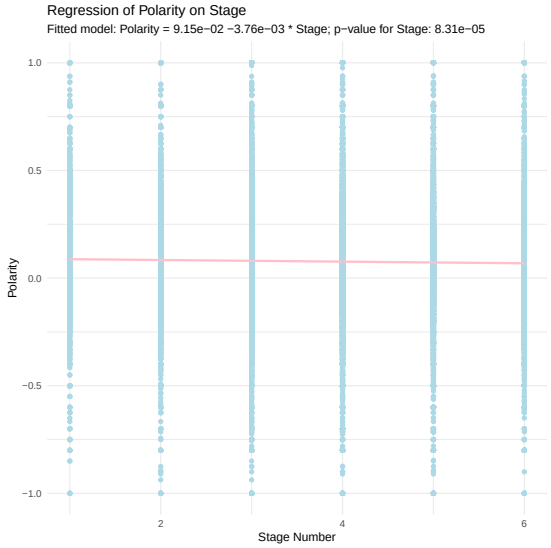
Here is our regression analysis: Due to the scaling problem, the fitted model seems not so obvious. But the p value for the model supports our hypothesis model. (p value $= 8.31 \times 10^{-5} < 0.05$) We have evidence saying there is a linear correlation between the stage Austin is in and people's sentiment.

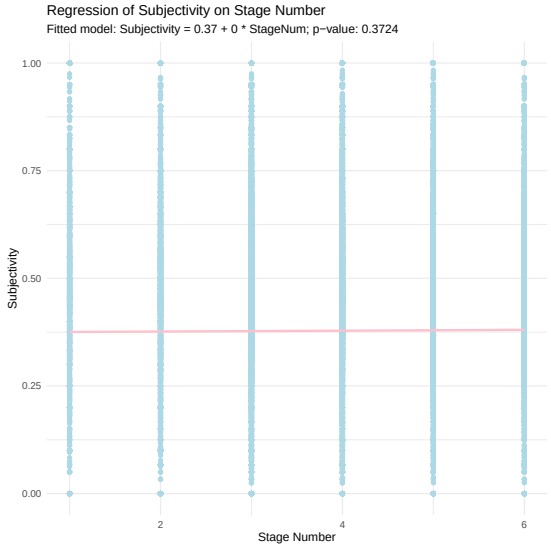
Besides, we regressed subkectivity against stage numbers as well. The mean plot above is flat. We want to check if there is no relation between how subkective people are and the stages Austin is in.
The hypothesis model is similar to the model for polarity.

$$\text{subjectivity} = \beta \times \text{stage number} + \mu$$

The p value suggests(p value $= 0.3724 > 0.05$) there is not enough evidence suggesting there is a relation between subkectivity and the stage Austin is in.

In this section, we discovered the higher the stage Austin is in, the more negative sentiment people express in their reddit. But people are as subjective (or objective) under different stages.

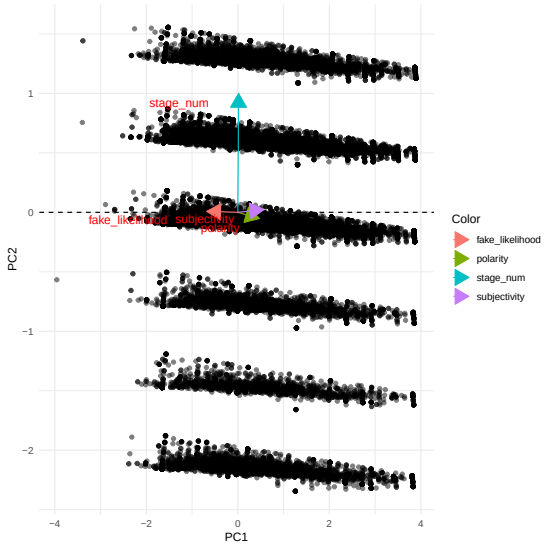

Regression of Polarity on Stage
Fitted model: Polarity = 9.15e−02 −3.76e−03 * Stage; p−value for Stage: 8.31e−05
Polarity
1.0
0.5
0.0
−0.5
−1.0
2
4
6
Stage Number


Regression of Subjectivity on Stage Number
Fitted model: Subjectivity = 0.37 + 0 * StageNum; p−value: 0.3724
Subjectivity
1.00
0.75
0.50
0.25
0.00
2
4
6
Stage Number

3 Principal Component Analysis

Besides, we performed principal component analysis on stage, polarity, subjectivity and fake likelihood to delver deeper into the problem.

The points are separated in 6 clusters which stand for six different stages. Fale



likelihood is vertical to the stage: they are independent. The stage num and the polarity score are on the opposite direction(not obvious though): the higher the stage num(more serious stage) the more negative people think. The subjectivity and polarity are in the same direction: the more positive people are, the more they use subjective words.

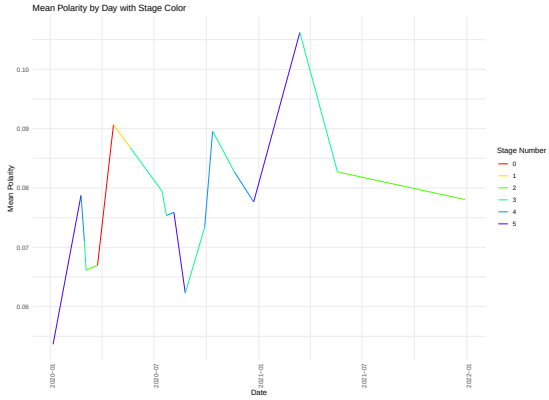
4 Regression in each stage

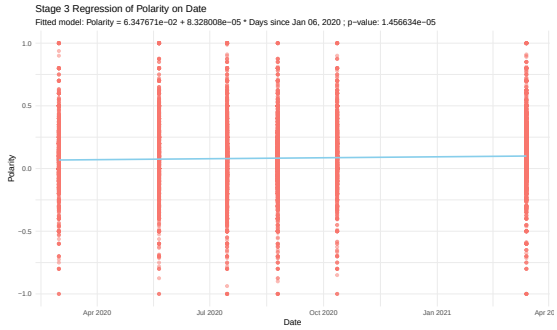
We are curious if people's polarity changes in different dates of the same stage. The plot for polarity by day might seem a little messy. So we extract different stages and fit a linear model on each stages.

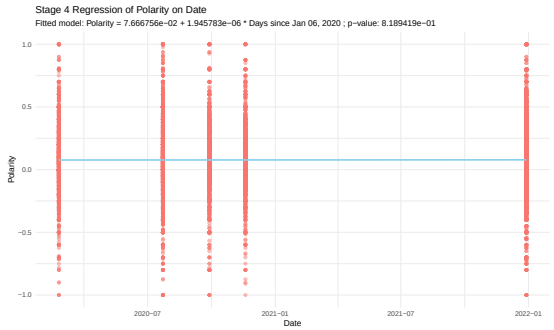
As there is only one period of time Austin is under stage 1 and stage 2. We just regressed for stage 3, 4 and 5.

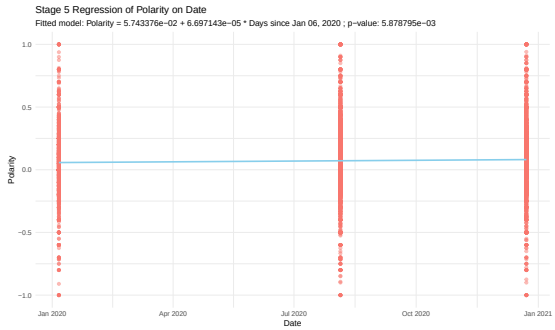
According to the p value, we have strong evidence that there is a linear relationship between people's polarity and the time in the same stage. One thing noteworthy is that people are actually expressing in a more positive way as time fly by.

Though, according to our previous analysis, people are expressing more negatively in more severe stages, people are saying more positively in the same status.


Mean Polarity by Day with Stage Color
Mean Polarity
0.10
0.09
0.08
0.07
0.06
2020-01
2020-07
2021-01
2021-07
2022-01
Date
Stage Number
0
1
2
3
4
5


Stage 3 Regression of Polarity on Date
Fitted model: Polarity = 6.347671e−02 + 8.328008e−05 * Days since Jan 06, 2020 ; p−value: 1.456634e−05
Polarity
1.0
0.5
0.0
−0.5
−1.0
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date


Stage 4 Regression of Polarity on Date
Fitted model: Polarity = 7.666756e−02 + 1.945783e−06 * Days since Jan 06, 2020 ; p−value: 8.189419e−01
Polarity
1.0
0.5
0.0
−0.5
−1.0
2020−07
2021−01
2021−07
2022−01
Date


Stage 5 Regression of Polarity on Date
Fitted model: Polarity = 5.743376e−02 + 6.697143e−05 * Days since Jan 06, 2020 ; p−value: 5.878795e−03
Polarity
1.0
0.5
0.0
−0.5
−1.0
Jan 2020
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date

5 Future Work

We are interested in if there is a pattern about how each user is behaving under different stages. We might track the posts from some active users and examine how positively or negatively they express under different stages or in different time.

We might identify various groups of users who have their expressing patterns.