Task 3 Sentence-Level Analysis

Zening Ye

12/6/2021

TNUM Analysis

For the first step in this task, I will connect to the tnum database and upload the book I chose from the task I, The Martian, and then use the function from class to upload the book into the tnum database under "test 2". After downloading the data from tnum, I tidy the data frame and make it more presentable. The data frame should look like Table 1 and Table 2 in Appendix.

Plot With TNUM Database

Plotting

By using the "sentimentr" package, there are several graphs that illustrate the emotion levels. In Figure 1, I created two different levels, sentiment and emotional. On the sentiment level, negative sentiment increased in Section 2/Chapter 2, however, this is different from the results I obtained in Task 2. In addition, the graph shows that there are a lot of emotions located at a neutral level, which means neither positive nor negative. If you look at the first graph I made from Task 2, the emotion levels are completely different. Next, by using the emotion levels, we can show that the whole emotion level stays at a low level, which means that there are no big emotional fluctuations throughout the book, in other words, the emotions in this book are mostly sad and flat.

I also made some graphs to illustrate the distribution in different type of word on sentiment and emotion level. For Figure 2, the left side is sentiment level. I used 3 levels to show the distribution of sentiment words in different sections/chapters. On the right side, it indicates the emotion type in different sections/chapters, in other words, it is a complementary plot for sentiment level. For Figure 3, same as Figure 2 to illustrate the sentiment level of the entire book, the emotion level plot I put in the Appendix.

```
# only run one time and it will save as HTML file
# sentence$string.value %>% get_sentences() %>% sentiment_by() %>% highlight()
```

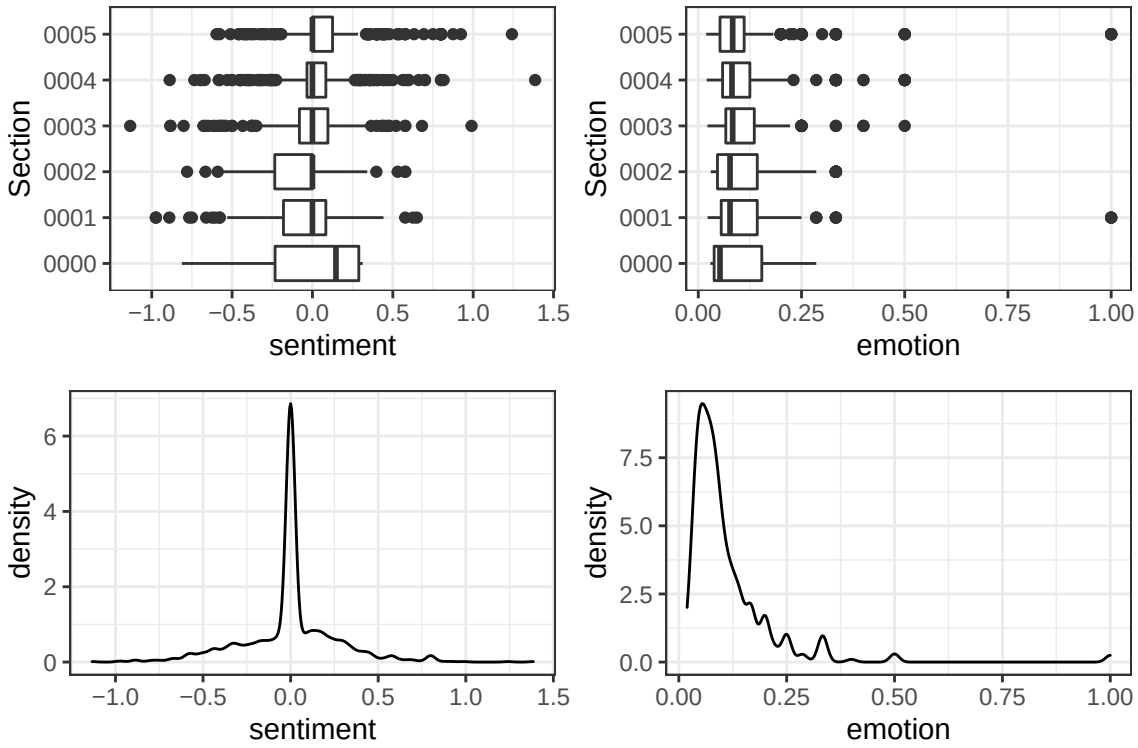


Figure 1: Sentiment and Emotion Level

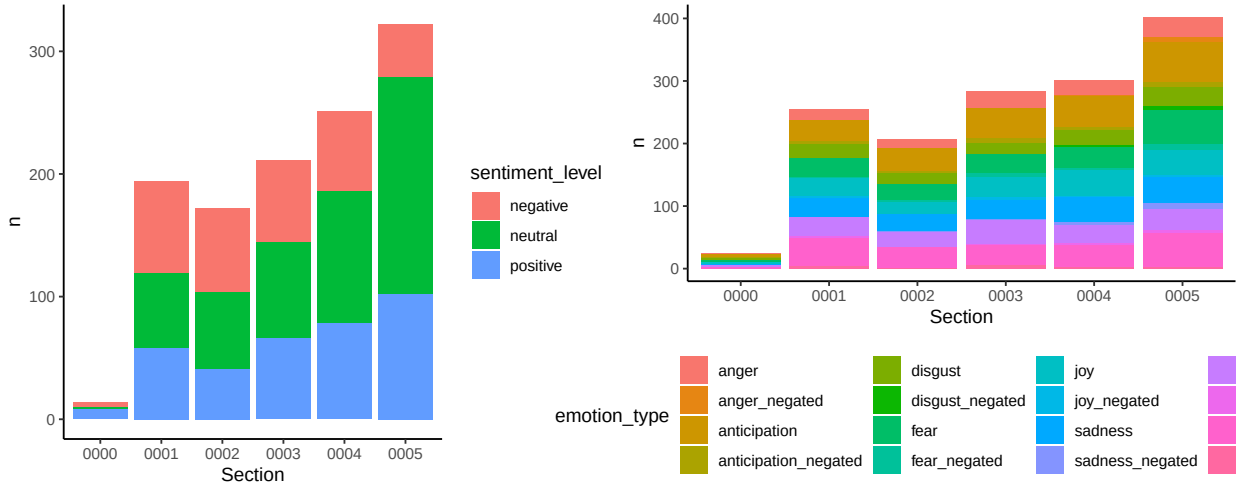


Figure 2: Distribution of Each Type

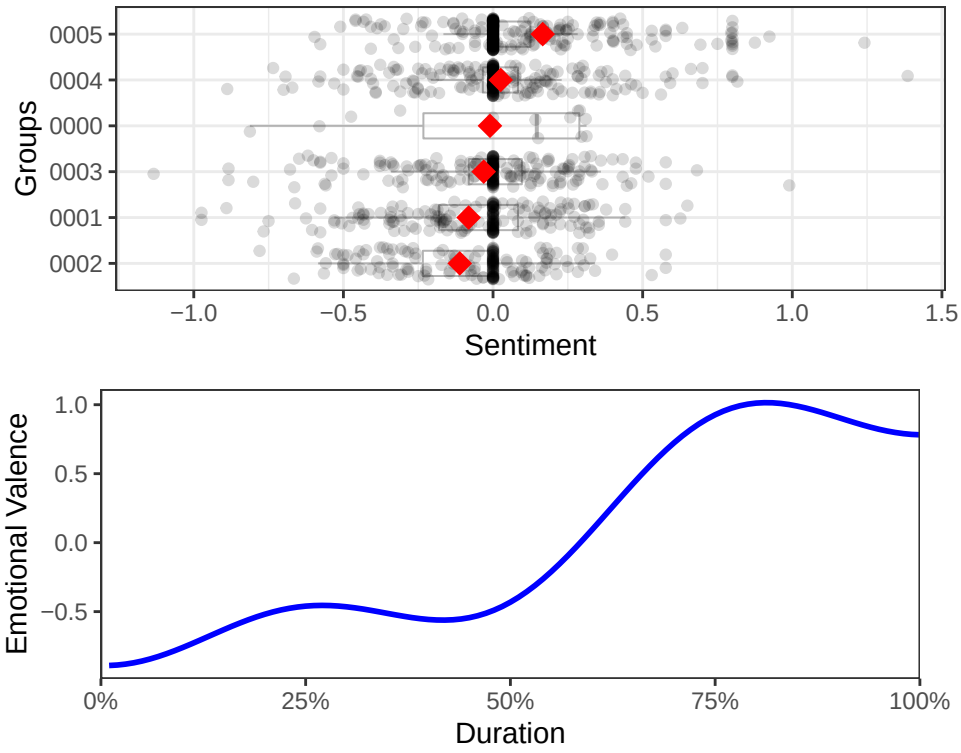
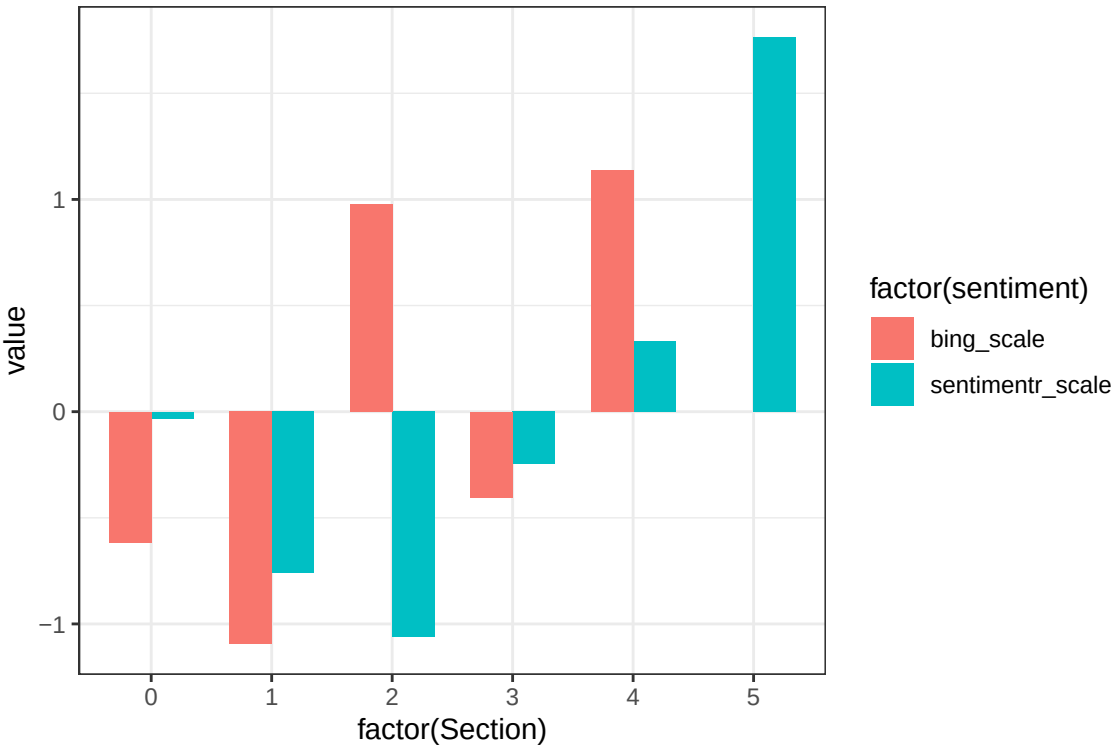


Figure 3: Sentiment Level Analysis



Conclusion

Even though I used the same database to do the sentiment analysis, it still has a lot of differences especially with the word that different packages define. For the "sentimentr" package, it is easy for me to extract the sentiment either by sentence or group. Furthermore, the "sentimentr" can analyze the text with the emotion level, it will be very useful to combine sentiment level and emotion level together to do the analysis.

Extra Credit

The book did not come up with a name of character, the most frequency word to describe the character is "He", so I will use "He" as a tag to query from the database. The first thing is I will add the tag for the word "He" and upload into the database. After that I read the data directly from the database to do the analysis.

As the graphs illustrate, the sentiment level of the character in this book is pretty neutral. Both graphs indicate that the sentiment level at 0 has higher frequency than other values, which means based on the analysis I did, the character did not have a lot of sentiment change in the entire book.

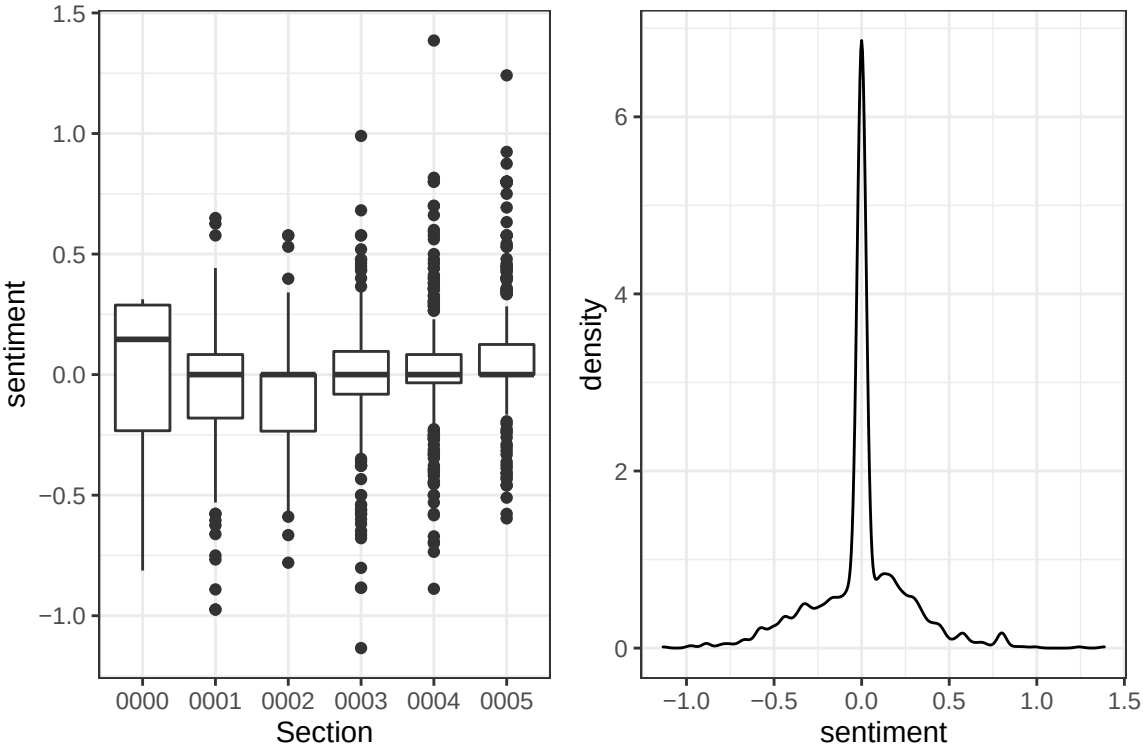


Figure 4: Sentiment Analysis with Specifit Character

TNUM data query and add tag for specific word

```
# get "fear" from database
# fear <- tnum.query("*the_martian_1* has * = REGEXP(\"fear\")", max= 100) %>%
#   tnum.objectsToDf()

# get "soft" from database
```

```
# soft <- tnum.query("*the_martian_1* has * = REGEXP(\"soft\")", max = 100) %>%
#  tnum.objectsToDf()

# get "dark" from database
# dark <- tnum.query("*the_martian_1* has * = REGEXP(\"dark\")", max = 100) %>%
#  tnum.objectsToDf()

#-----------------------------------------------------------------------------------------------
# Add tag into the database
#tnum.tagByQuery("*the_martian_1* has * = REGEXP(\"fear\")", adds=("ref:fear_the_martian_1"))
#list(modifiedCount = 16, tagged = 16, removed = 0)

#tnum.tagByQuery("*the_martian_1* has * = REGEXP(\"soft\")", adds=("ref:soft_the_martian_1"))
#list(modifiedCount = 10, tagged = 10, removed = 0)

# tnum.tagByQuery("*the_martian_1* has * = REGEXP(\"dark\")", adds=("ref:dark_the_martian_1"))
# list(modifiedCount = 17, tagged = 17, removed = 0)
```

Specific word query by tag from tnum

```
# fear_tag <- tnum.query("@ref:fear_the_martian_1", max= 100) %>% tnum.objectsToDf()
# soft_tag <- tnum.query("@ref:soft_the_martian_1", max= 100) %>% tnum.objectsToDf()
# dark_tag <- tnum.query("@ref:dark_the_martian_1", max= 100) %>% tnum.objectsToDf()


# separate the section(chapter) then plot with x = chapter y - sentiment
# sentiment(fear_tag$string.value) %>% ggplot(aes(element_id, sentiment)) + geom_bar(stat = "identity")
# sentiment(soft_tag$string.value) %>% ggplot(aes(element_id, sentiment)) + geom_bar(stat = "identity")
# sentiment(dark_tag$string.value) %>% ggplot(aes(element_id, sentiment)) + geom_bar(stat = "identity")
# tnum.graphPathList(tnum.query(query = "glasser/the_martian_1# has *", max = 20))
```

Citation

Technical support:

Yuli Jin: Code for comparison visualization Line 100 - 138

Runqi Zhao: tnum Visualization advice

Ranfei Xu: tnum Visualization advice

Tyler Rinker (2014): *sentimentr,* [online] https://github.com/trinker/sentimentr

AbdulMajedRaja RS (2016), *Sentiment Analysis in R with {sentimentr} that handles Negation (Valence Shifters),* [online] https://www.r-bloggers.com/2020/04/sentiment-analysis-in-r-with-sentimentr-that-handles-negation-valence-shifters

Table 1: Original Data Frame

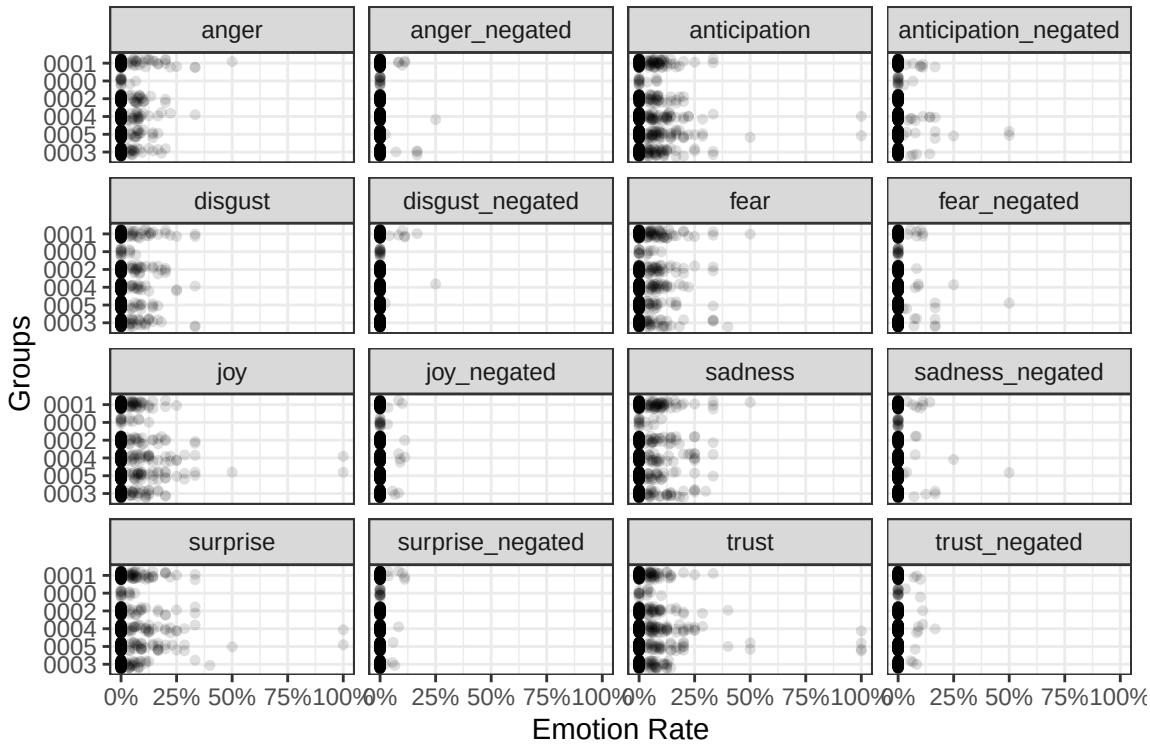
Author	Name	Section	Paragraph	Sentence	property
glasser	the_martian_1	section:0000	paragraph:0002	sentence:0001	text
glasser	the_martian_1	section:0000	paragraph:0002	sentence:0002	text
glasser	the_martian_1	section:0000	paragraph:0003	sentence:0001	text
glasser	the_martian_1	section:0000	paragraph:0003	sentence:0002	text
glasser	the_martian_1	section:0000	paragraph:0004	sentence:0001	text
glasser	the_martian_1	section:0000	paragraph:0007	sentence:0001	text
glasser	the_martian_1	section:0000	paragraph:0008	sentence:0001	text
glasser	the_martian_1	section:0000	paragraph:0008	sentence:0002	text
glasser	the_martian_1	section:0000	paragraph:0009	sentence:0001	text
glasser	the_martian_1	section:0000	paragraph:0009	sentence:0002	text

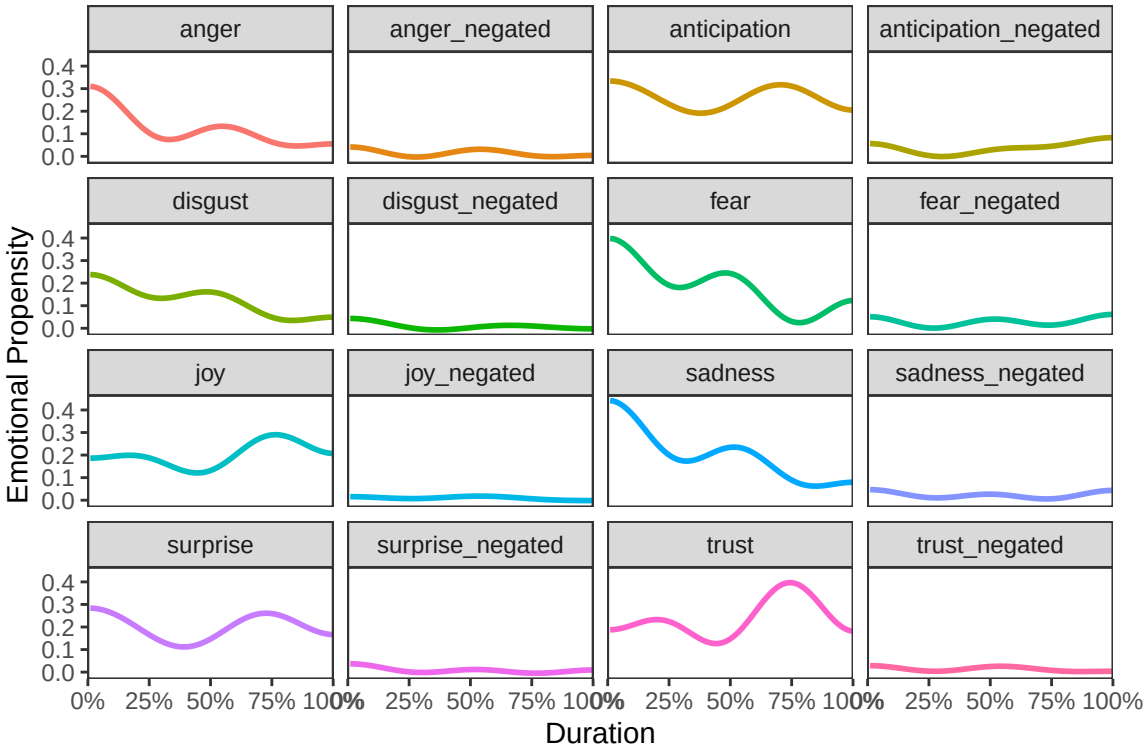
Table 2: Tidy Data Frame

Section	Sentence	property	Paragraph	string.value	numeric.value	tags
0000	0003	text	0009	"And if a stranger from another world, dazed by new conditions and unable to make his wants known, were to fall into their hands his fate might not be happy."	NA	ref:he_the_martian_1
0000	0001	text	0010	"We have read no story that pictures with such clarity and insight the experiences of a man on another world than his own, than does this present story."	NA	ref:he_the_martian_1
0000	0002	text	0010	"With the basis of a splendid plot Mr Hilliard has worked up a simply marvelous story."	NA	ref:he_the_martian_1

Appendix

Emotion Level Analysis


anger
anger_negated
anticipation
anticipation_negated
disgust
disgust_negated
fear
fear_negated
joy
joy_negated
sadness
sadness_negated
surprise
surprise_negated
trust
trust_negated
Groups
0001
0000
0002
0004
0005
0003
0% 25% 50% 75% 100%
Emotion Rate


anger
anger_negated
anticipation
anticipation_negated
disgust
disgust_negated
fear
fear_negated
joy
joy_negated
sadness
sadness_negated
surprise
surprise_negated
trust
trust_negated
Emotional Propensity
0.0
0.1
0.2
0.3
0.4
0% 25% 50% 75% 100%
Duration