134 Project

2025-05-15

```
comments <- read.csv("labeled_comments.csv")

str(comments)
```

```
## 'data.frame':    14056 obs. of  9 variables:
##  $ id           : chr  "mn6rd1i" "mna0z8a" "mn6uha4" "mn8hteo" ...
##  $ author       : chr  "Haze_Shadez" "Humblerbee" "Ok_Mouse_3791" "myNameBurnsGold" ...
##  $ body         : chr  "I want Roy to represent us, man was always clutch" "Two Cronin quotes from hi
##  $ score        : int  10 9 8 8 9 8 6 5 4 3 ...
##  $ comment_karma: num  4456 81257 10945 35259 16969 ...
##  $ created_utc  : num  1.74e+09 1.74e+09 1.74e+09 1.74e+09 1.74e+09 ...
##  $ subreddit    : chr  "ripcity" "ripcity" "ripcity" "ripcity" ...
##  $ vader_score  : num  0.0772 0.9796 0.4019 0 0 ...
##  $ sentiment    : chr  "positive" "positive" "positive" "neutral" ...
```

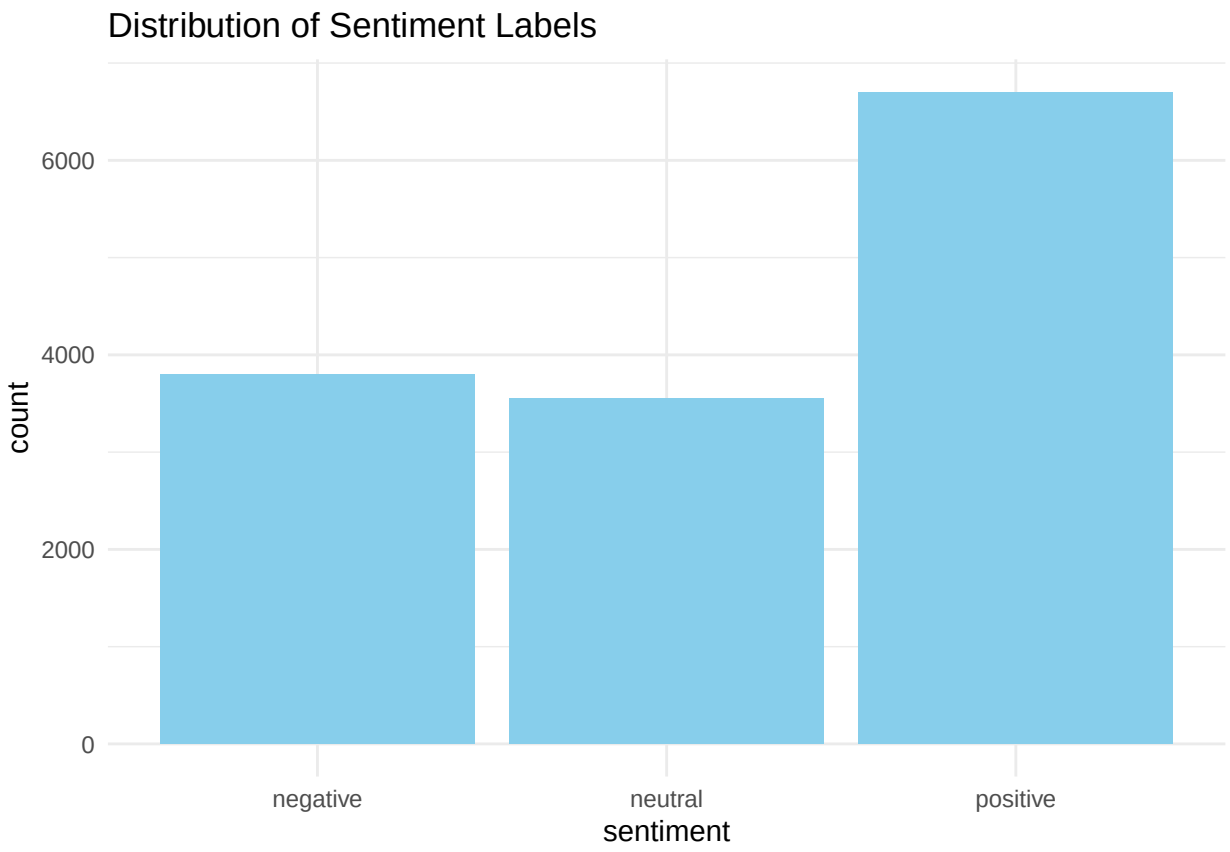
```
summary(comments)
```

```
##       id                author             body               score         
##  Length:14056       Length:14056       Length:14056       Min.   : -95.000  
##  Class :character   Class :character   Class :character   1st Qu.:   1.000  
##  Mode  :character   Mode  :character   Mode  :character   Median :   2.000  
##                                                           Mean   :   8.102  
##                                                           3rd Qu.:   6.000  
##                                                           Max.   :1061.000  
##                                                                             
##  comment_karma       created_utc          subreddit           vader_score      
##  Min.   :   -100   Min.   :1.742e+09   Length:14056       Min.   :-0.9949  
##  1st Qu.:   2841   1st Qu.:1.747e+09   Class :character   1st Qu.:-0.1280  
##  Median :  14769   Median :1.747e+09   Mode  :character   Median : 0.0000  
##  Mean   :  60726   Mean   :1.747e+09                      Mean   : 0.1376  
##  3rd Qu.:  57284   3rd Qu.:1.747e+09                      3rd Qu.: 0.5423  
##  Max.   :3141592   Max.   :1.747e+09                      Max.   : 0.9992  
##  NA's   :73                                                                
##   sentiment        
##  Length:14056       
##  Class :character   
##  Mode  :character   
##                     
##                     
##                     
##
```

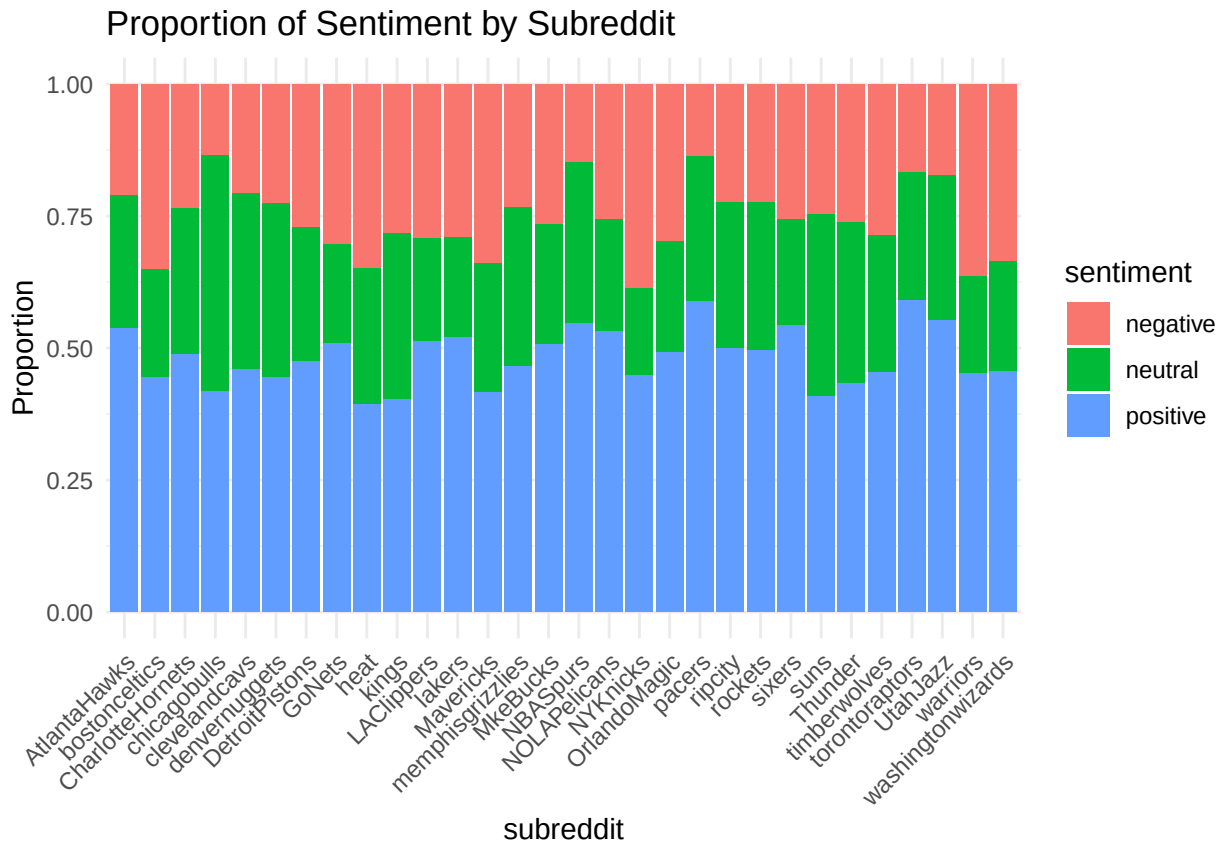
```r
table(comments$sentiment)
```

```
##
## negative  neutral positive
##     3801     3552     6703
```

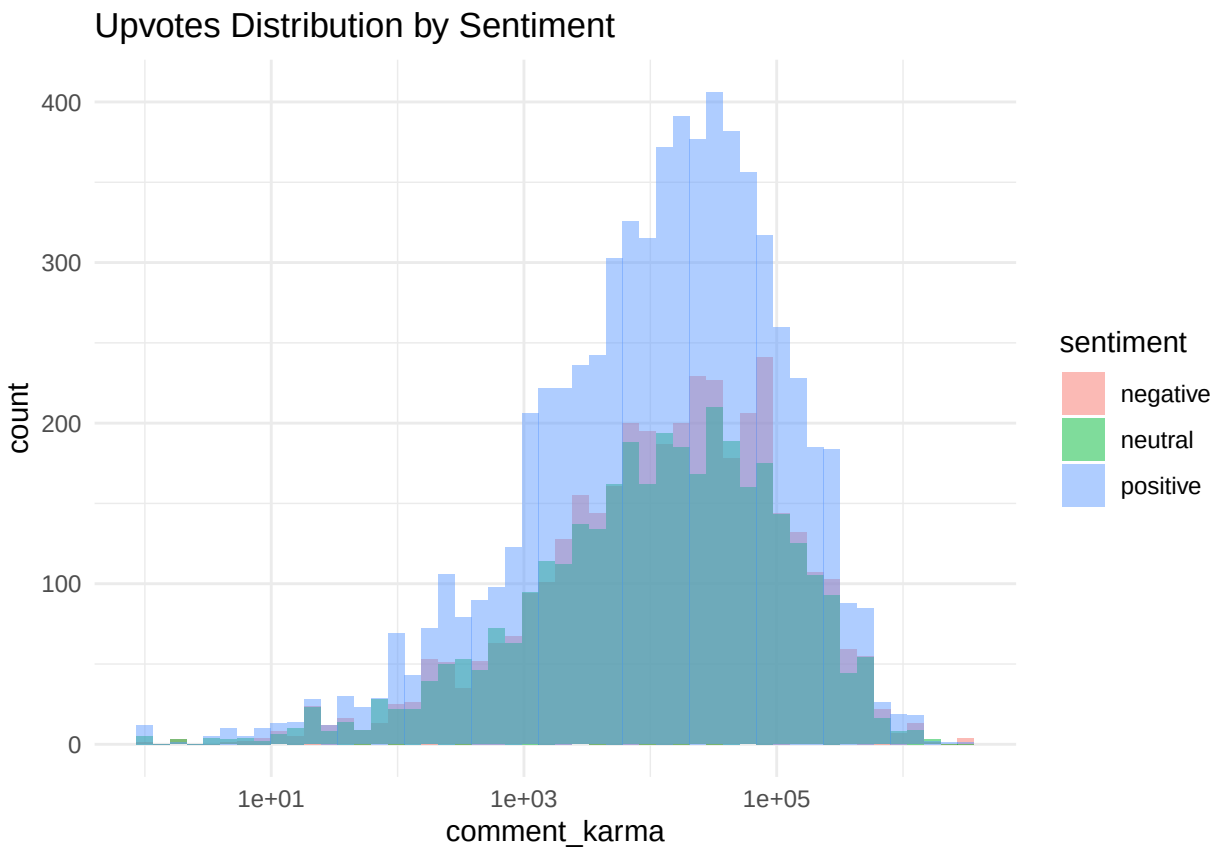
```r
library(ggplot2)

#overall sentiment distribution
ggplot(comments, aes(x = sentiment)) +
  geom_bar(fill = "skyblue") +
  theme_minimal() +
  labs(title = "Distribution of Sentiment Labels")
```



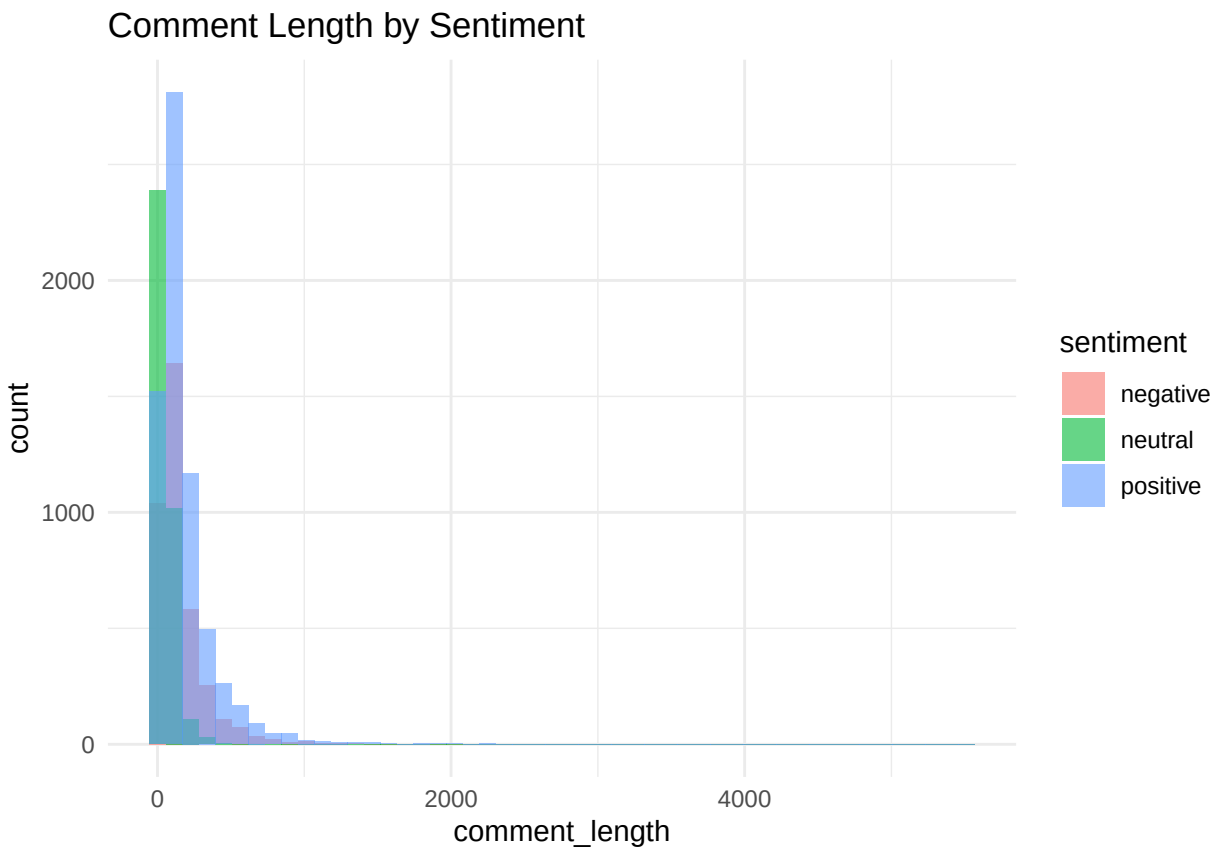

```r
#sentiment by subreddit
ggplot(comments, aes(x = subreddit, fill = sentiment)) +
  geom_bar(position = "fill") +
  theme_minimal() +
  labs(title = "Proportion of Sentiment by Subreddit", y = "Proportion") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



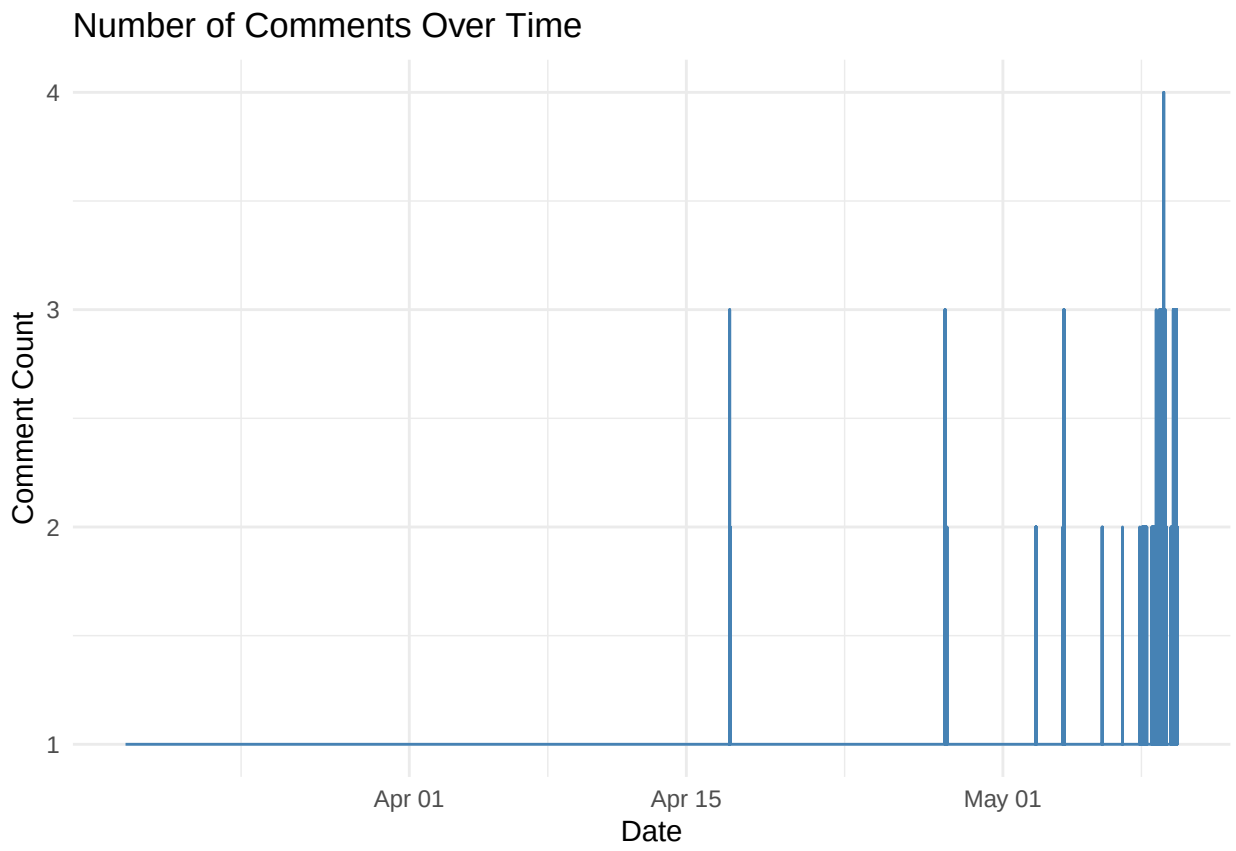

```
#upvotes distribution by sentiment
ggplot(comments, aes(x = comment_karma, fill = sentiment)) +
  geom_histogram(position = "identity", alpha = 0.5, bins = 50) +
  theme_minimal() +
  scale_x_log10() +
  labs(title = "Upvotes Distribution by Sentiment")
```



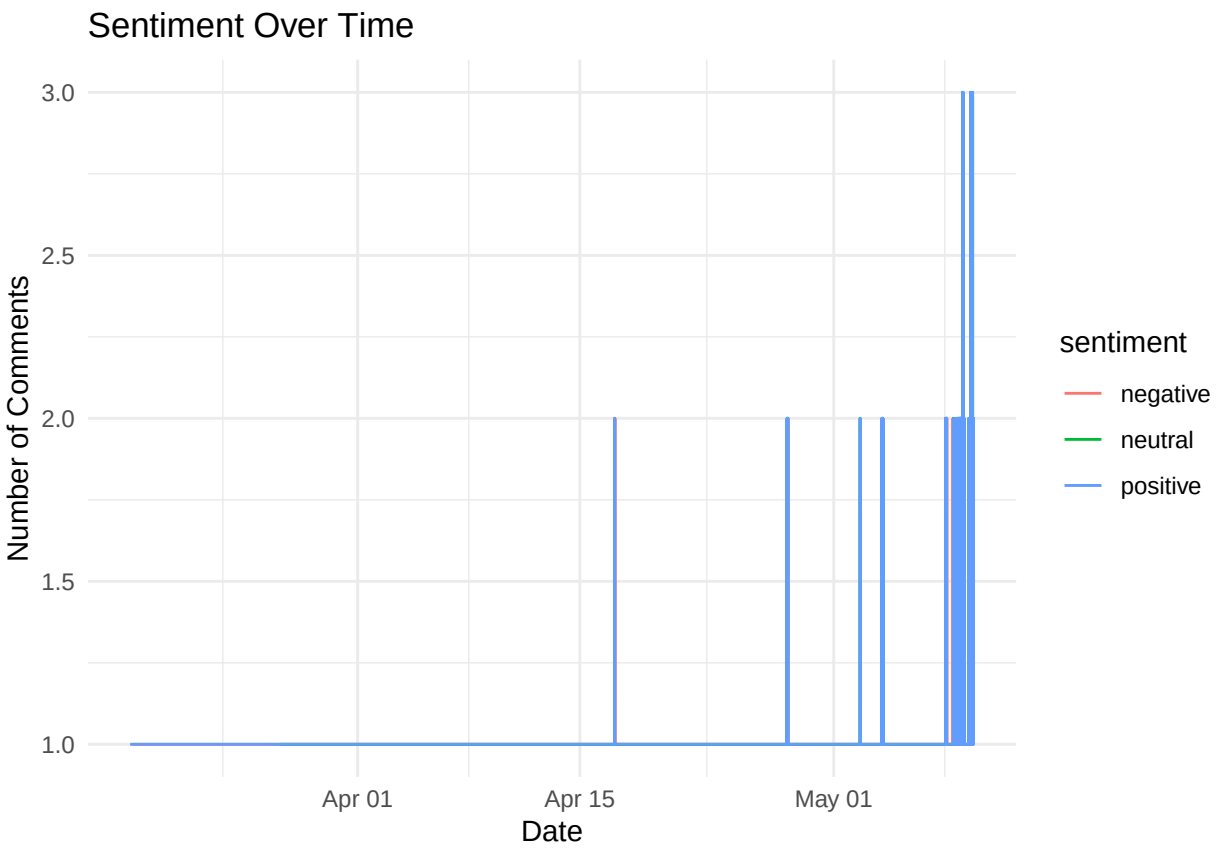

```
#comment length by sentiment
comments$comment_length <- nchar(comments$body)

ggplot(comments, aes(x = comment_length, fill = sentiment)) +
  geom_histogram(position = "identity", alpha = 0.6, bins = 50) +
  theme_minimal() +
  labs(title = "Comment Length by Sentiment")
```




```
#comment sentiment by time
comments$created_datetime <- as.POSIXct(comments$created_utc, origin = "1970-01-01", tz = "UTC")

library(dplyr)

comments %>%
  count(created_datetime) %>%
  ggplot(aes(x = created_datetime, y = n)) +
  geom_line(color = "steelblue") +
  theme_minimal() +
  labs(title = "Number of Comments Over Time", x = "Date", y = "Comment Count")
```




```
comments %>%
  count(created_datetime, sentiment) %>%
  ggplot(aes(x = created_datetime, y = n, color = sentiment)) +
  geom_line() +
  theme_minimal() +
  labs(title = "Sentiment Over Time", x = "Date", y = "Number of Comments")
```


Sentiment Over Time
Number of Comments
3.0
2.5
2.0
1.5
1.0
Apr 01
Apr 15
May 01
Date
sentiment
negative
neutral
positive