

# note 記事アクセス分析レポート

Masahito

2025-06-29

## Contents

1 はじめに	1
2 データ内容の確認	1
3 データの前処理	2
4 データ解析（閲覧数）	4
5 データ解析（エンゲージメント：スキ数）	9
6 データ解析（エンゲージメント：スキ率）	13
7 参考情報	18

## 1 はじめに

このレポートでは、仮想 note 記事データをもとに、どんな条件の記事がより多く読まれるのか分析・可視化し、傾向を把握します。さらに、閲覧された記事のエンゲージメント（ユーザーの反応：スキ）について詳しく分析します。

## 2 データ内容の確認

```
## データの読み込み
note_data <- read_csv("step01_note_virtual_data.csv")
```

### ## データの構造確認

```
glimpse_text <- capture.output(glimpse(note_data))
cat(paste(glimpse_text, collapse = "\n"))
```

Rows: 15 Columns: 6 \$ title “AI 時代に必要なスキルとは”, “30 代でキャリアに悩んだときに読んだ本”, “SNS 疲れから抜け出す方法”, “副業を始め ~ \$ length 2051, 1716, 2455, 2186, 1881, 1938, 1345, 2368, 1785, 2210, 197~ \$ time 10:37:00, 18:11:00, 23:50:00, 05:20:00, 14:27:00, 13:14:00, 20~ \$ tag ” キャリア, 働き方”, “読書, ライフハック”, “SNS, メンタル”, “副業, マネー”, “生活習慣, 朝活”, “健康, ラ ~ \$ pv 1233, 1289, 833, 636, 1122, 1171, 664, 789, 1142, 1256, 351, 76~ \$ likes 79, 123, 47, 61, 72, 123, 53, 143, 185, 98, 48, 139, 64, 196, 49

### ## 各数値の要約

```
summary(note_data)
```

```
title          length          time          tag

Length:15 Min. :1274 Min. :01:31:00 Length:15
Class :character 1st Qu.:1750 1st Qu.:07:26:00 Class :character
Mode :character Median :1948 Median :10:37:00 Mode :character
Mean :1897 Mean :12:33:40
3rd Qu.:2118 3rd Qu.:19:17:00
Max. :2455 Max. :23:50:00
pv likes
Min. : 351.0 Min. : 47.00
1st Qu.: 716.0 1st Qu.: 57.00
Median :1122.0 Median : 79.00
Mean : 949.3 Mean : 98.67
3rd Qu.:1207.5 3rd Qu.:131.00
Max. :1369.0 Max. :196.00
```

### ## 欠損値の有無

```
colSums(is.na(note_data))
```

```
title length time tag pv likes 0 0 0 0 0 0
```

---

## 3 データの前処理

### ## タイトル文字数

```
note_data <- note_data %>%
```

```

mutate(title_num = nchar(title))
## スキ率列の追加
note_data <- note_data %>%
  mutate(like_rate = likes / pv)
## タグの分解
note_data_long <- note_data %>%
  separate_rows(tag, sep = ",\\s*") %>%
  mutate(tag = str_trim(tag))
## タグの重複チェック
note_data_long %>%
  count(tag) %>%
  arrange(desc(n))

```

```

## # A tibble: 29 x 2
##   tag          n
##   <chr>      <int>
## 1 読書          2
## 2 AI            1
## 3 SNS            1
## 4 Z世代          1
## 5 note           1
## 6 アクセス解析  1
## 7 キャリア        1
## 8 デザイン        1
## 9 ブログ          1
## 10 マネー         1
## # i 19 more rows

```

```

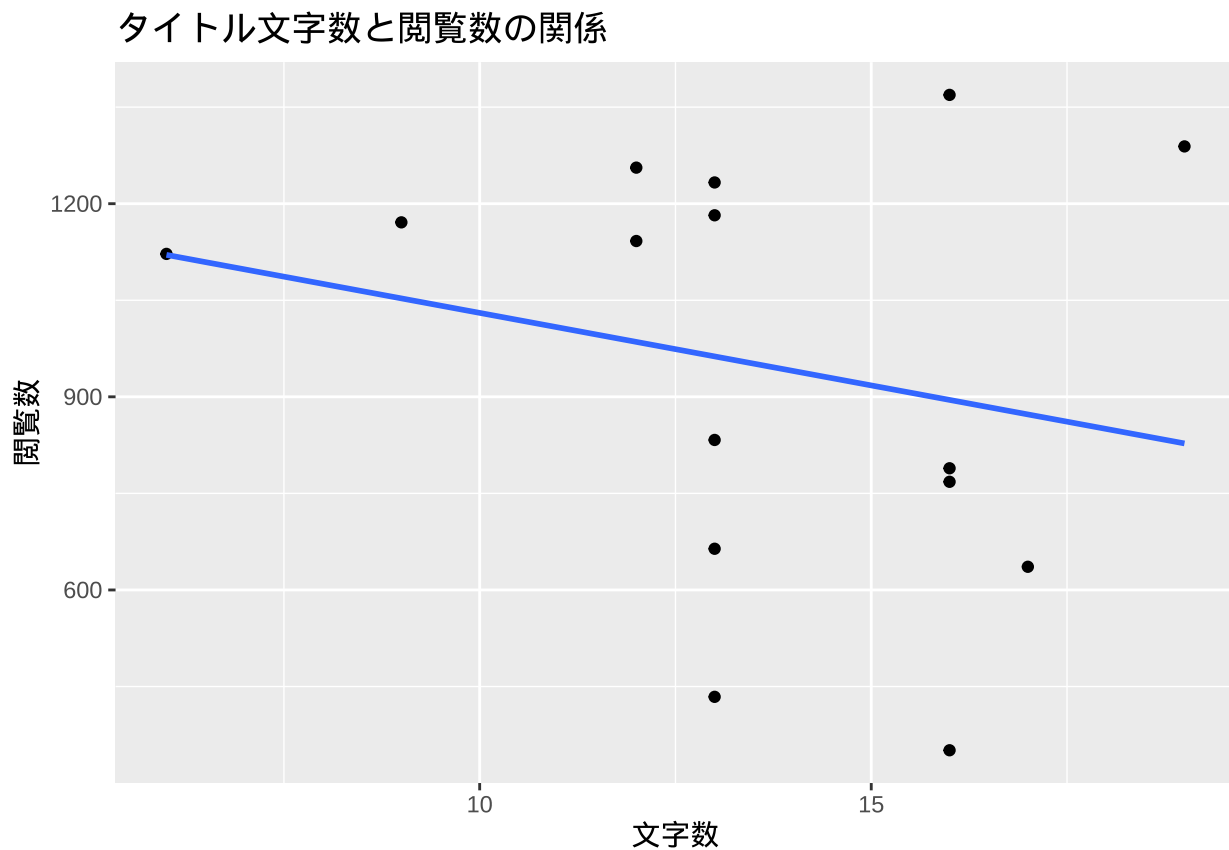
## 時間帯カテゴリの追加
note_data <- note_data %>%
  mutate(
    time_hms = hms::as_hms(paste0(time, ":00")),
    time_hours = hour(time_hms) + minute(time_hms)/60
  )
note_data <- note_data %>%
  mutate(time_category = case_when(
    time_hours < 12 ~ "朝",
    time_hours < 18 ~ "昼",
    TRUE ~ "夜"
  ))
## 文字数カテゴリの追加
note_data <- note_data %>%
  mutate(length_category = case_when(
    length < 800 ~ "短文",

```

```
length < 1500 ~ " 中文",
TRUE ~ " 長文"
))
## 文字数カテゴリの可視化
showtext_auto()
```

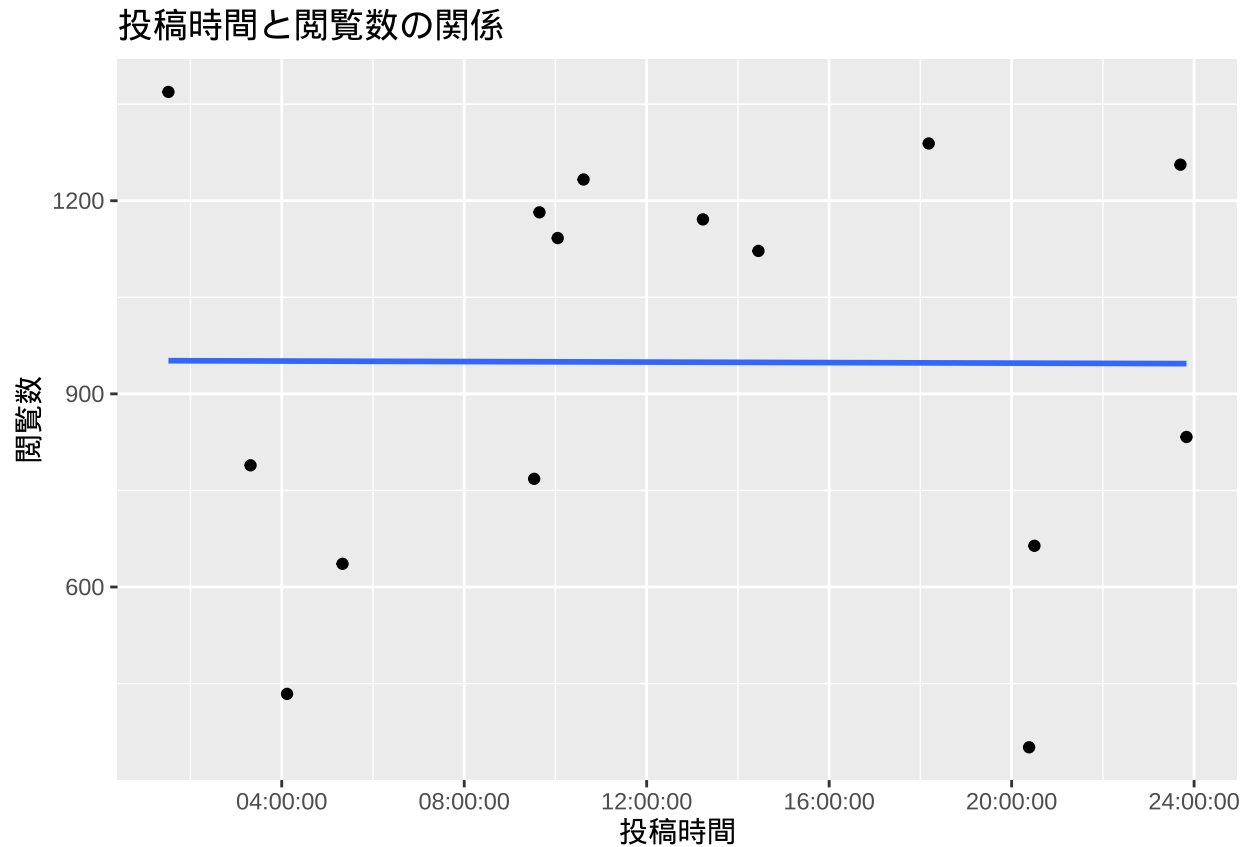
## 4 データ解析（閲覧数）

```
## タイトルの文字数と閲覧数の関係
ggplot(note_data, aes(x = title_num, y = pv)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = " タイトル文字数と閲覧数の関係", x = " 文字数", y = " 閲覧数")
```



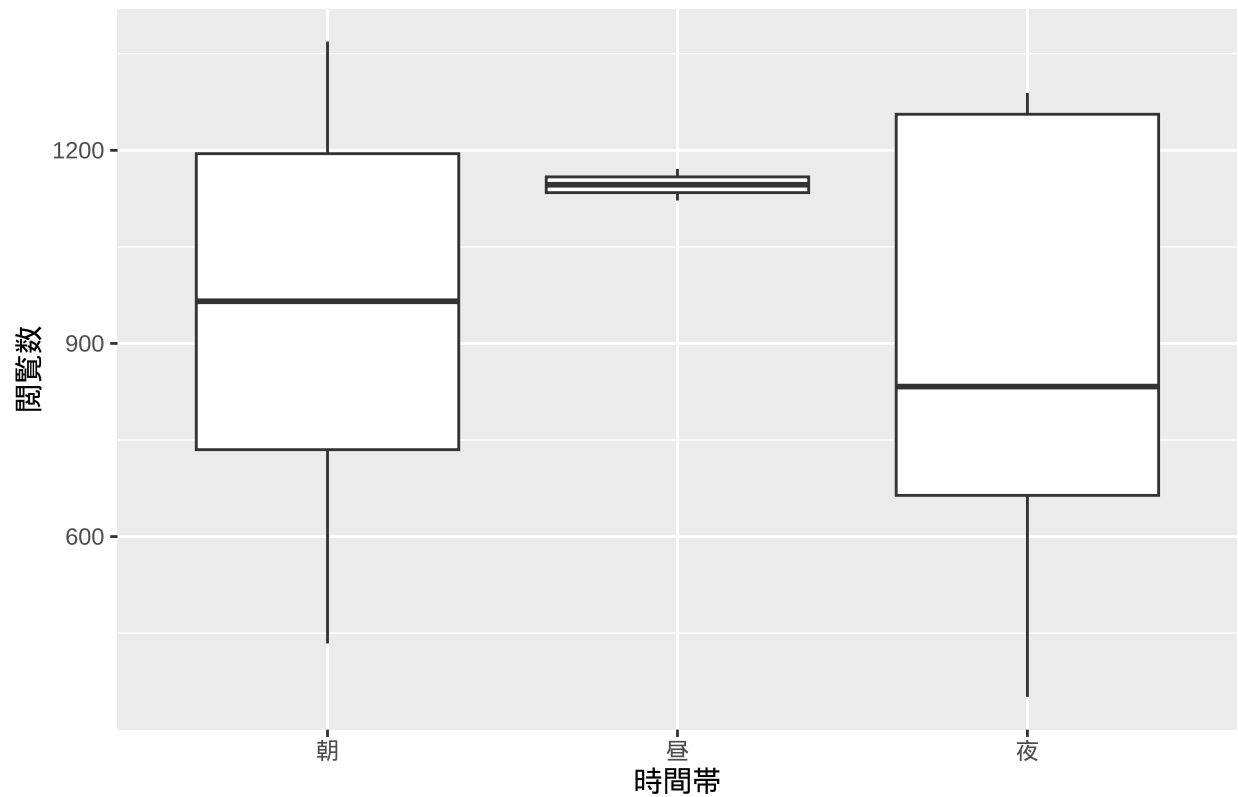
```
## 投稿時間と閲覧数の関係
ggplot(note_data, aes(x = time, y = pv)) +
```

```
geom_point() +
geom_smooth(method = "lm", se = FALSE) +
labs(title = " 投稿時間と閲覧数の関係", x = " 投稿時間", y = " 閲覧数")
```



```
## 投稿時間帯と閲覧数の関係
note_data <- note_data %>%
  mutate(time_category = factor(time_category, levels = c(" 朝", " 昼", " 夜")))
ggplot(note_data, aes(x = time_category, y = pv)) +
  geom_boxplot() +
  labs(title = " 時間帯別の閲覧数 (PV)", x = " 時間帯", y = " 閲覧数")
```

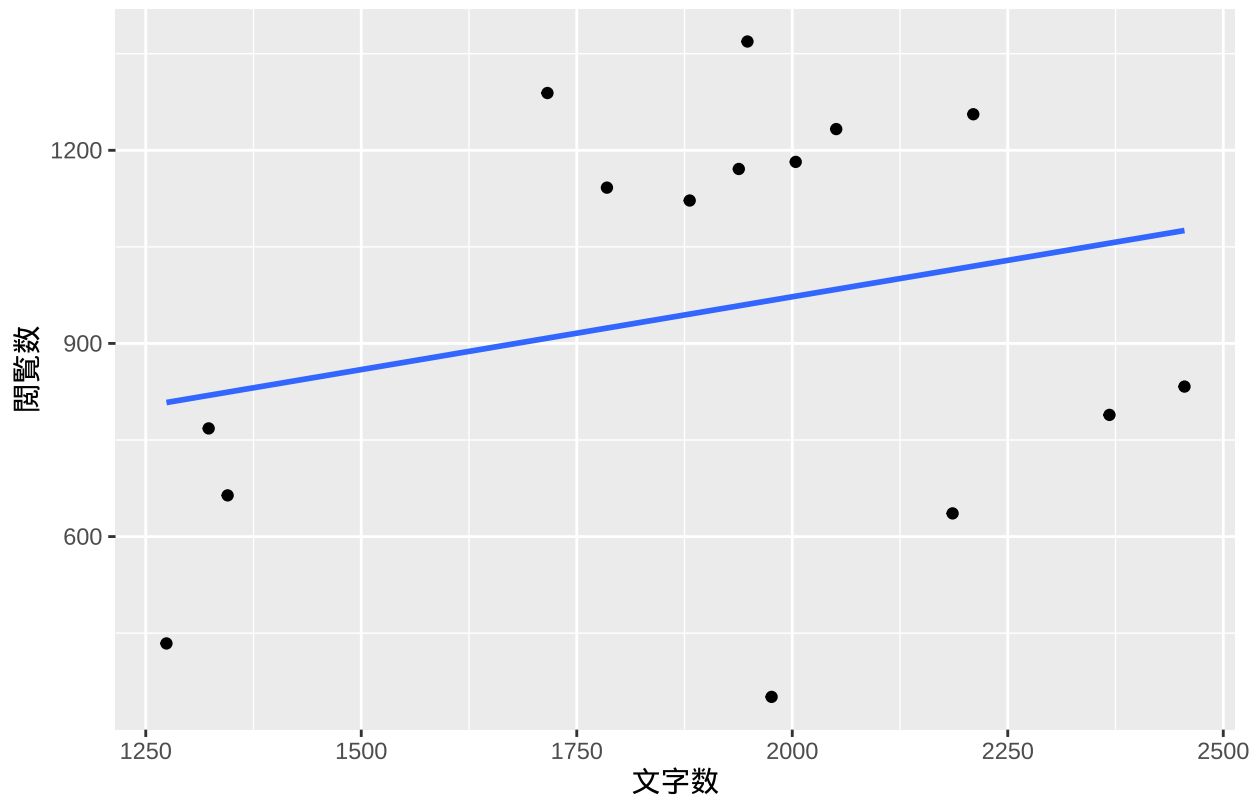
時間帯別の閲覧数(PV)



## 文字数と閲覧数の関係

```
ggplot(note_data, aes(x = length, y = pv)) +  
  geom_point() +  
  geom_smooth(method = "lm", se = FALSE) +  
  labs(title = " 文字数と閲覧数の関係", x = " 文字数", y = " 閲覧数")
```

## 文字数と閲覧数の関係



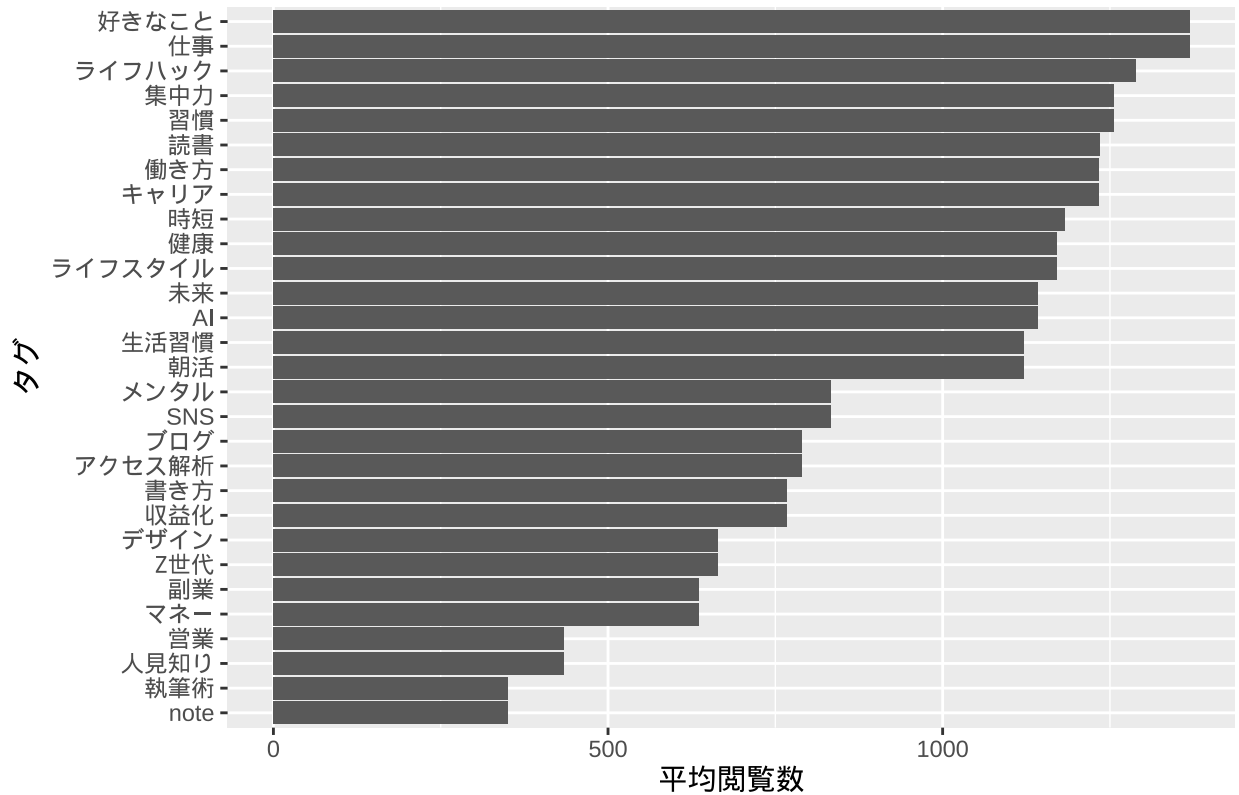
```
## タグ別の平均閲覧数
```

```
note_data %>%  
  summarise(  
    avg_pv = mean(pv)  
  )
```

```
## # A tibble: 1 x 1  
##   avg_pv  
##   <dbl>  
## 1   949.
```

```
tag_summary <- note_data_long %>%  
  group_by(tag) %>%  
  summarise(mean_pv = mean(pv), .groups = "drop")  
ggplot(tag_summary, aes(x = reorder(tag, mean_pv), y = mean_pv)) +  
  geom_col() +  
  coord_flip() +  
  labs(title = "タグ別平均閲覧数", x = "タグ", y = "平均閲覧数")
```

## タグ別平均閲覧数



### ## 閲覧数が高い記事の特徴を見る

```
note_data %>%
  arrange(desc(pv)) %>%
  select(title, title_num, time, length, pv, likes, like_rate) %>%
  head(5)
```

```
## # A tibble: 5 x 7
##   title                                title_num time  length    pv likes like_rate
##   <chr>                                <int> <tim>  <dbl> <dbl> <dbl>    <dbl>
## 1 「好き」を仕事にするための第一歩         16 01:31   1948  1369   196    0.143
## 2 30代でキャリアに悩んだときに読んだ本~~         19 18:11   1716  1289   123    0.0954
## 3 集中力を高める3つの習慣                   12 23:42   2210  1256    98    0.0780
## 4 AI時代に必要なスキルとは                 13 10:37   2051  1233    79    0.0641
## 5 忙しい人のための時短読書術                 13 09:39   2004  1182    64    0.0541
```



## 5 データ解析（エンゲージメント：スキ数）

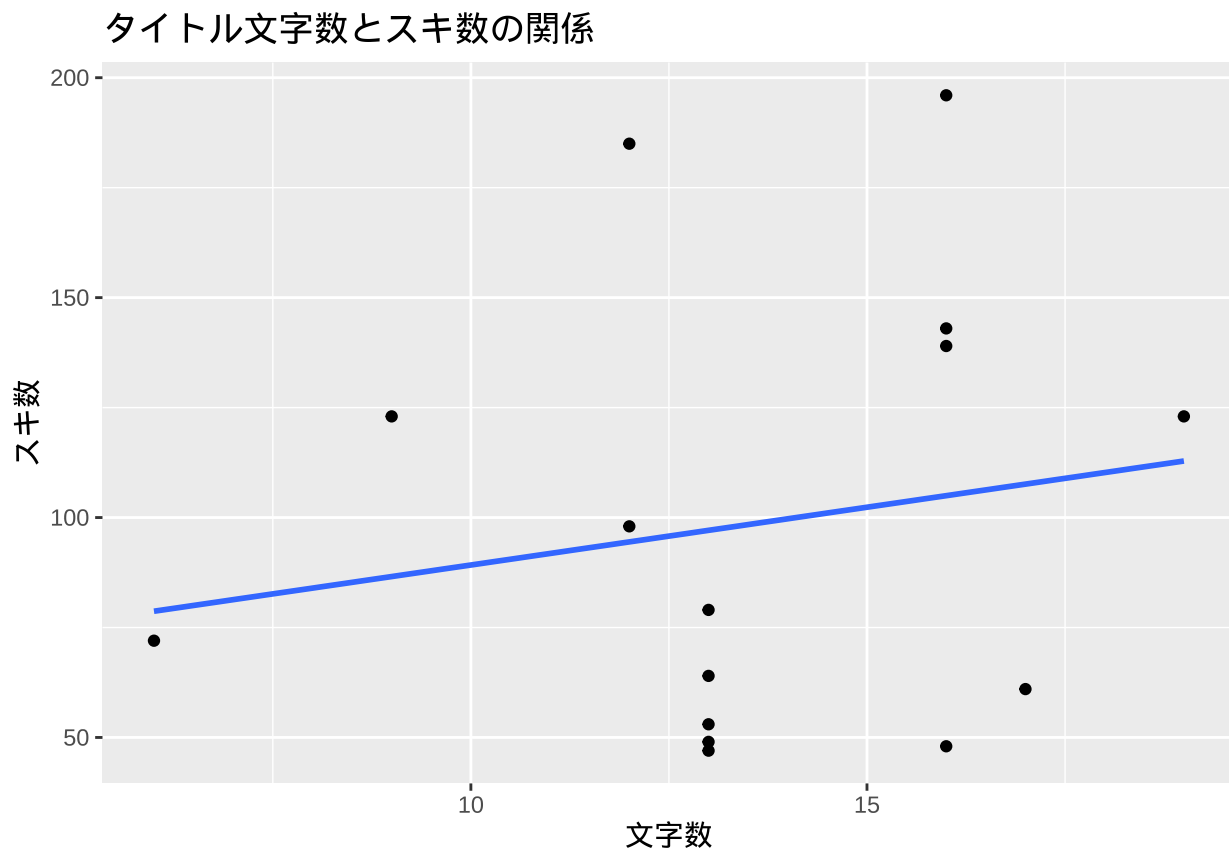
```
## 閲覧数とスキ数との相関関係
```

```
cor(note_data$pv, note_data$likes)
```

```
## [1] 0.5419124
```

```
## タイトルの文字数とスキ数の関係
```

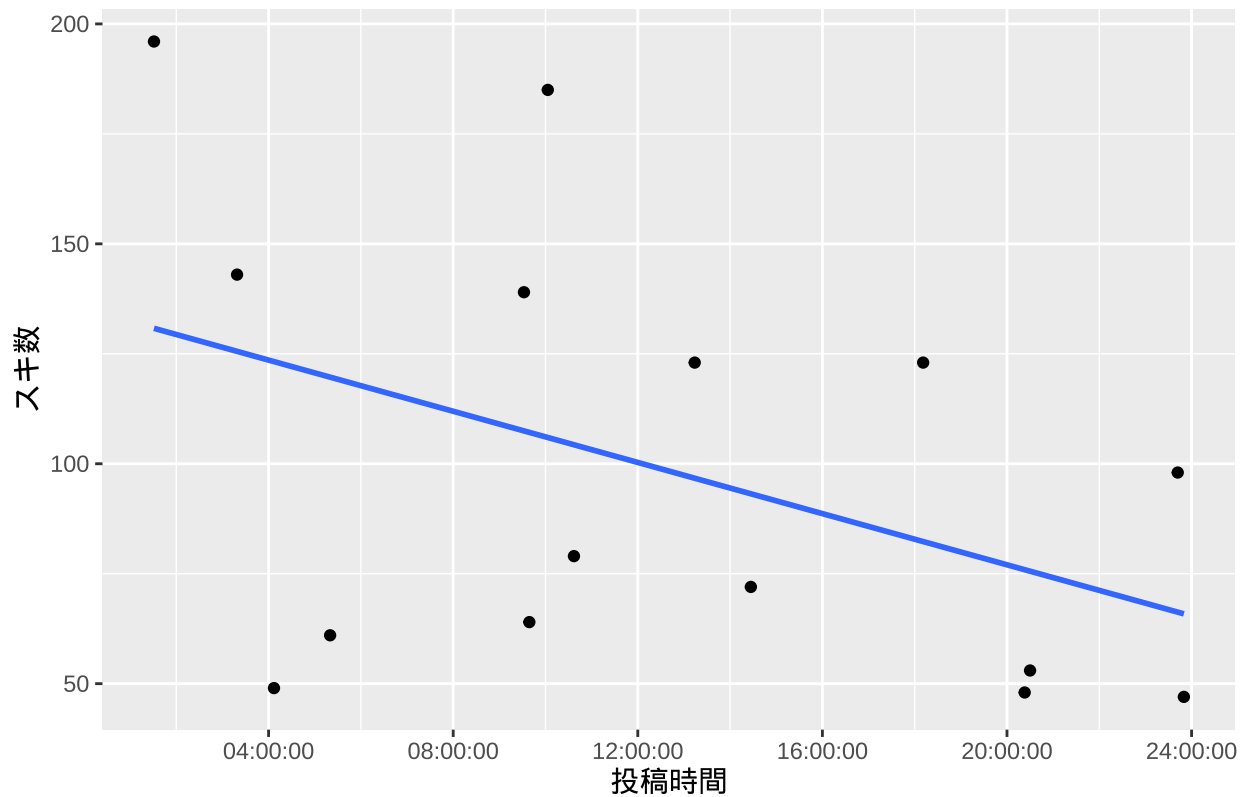
```
ggplot(note_data, aes(x = title_num, y = likes)) +  
  geom_point() +  
  geom_smooth(method = "lm", se = FALSE) +  
  labs(title = " タイトル文字数とスキ数の関係", x = " 文字数", y = " スキ数")
```



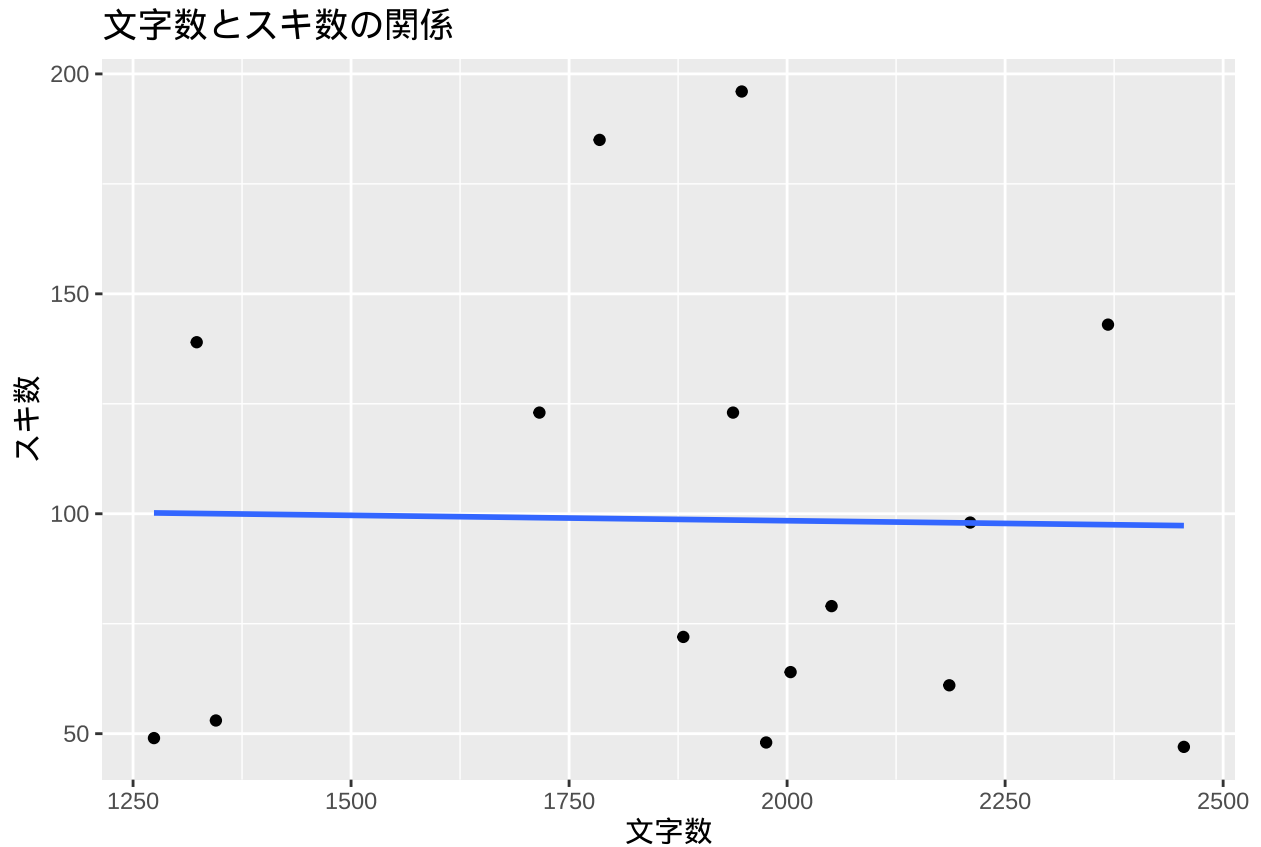
```
## 投稿時間とスキ数の関係
```

```
ggplot(note_data, aes(x = time, y = likes)) +  
  geom_point() +  
  geom_smooth(method = "lm", se = FALSE) +  
  labs(title = " 投稿時間とスキ数の関係", x = " 投稿時間", y = " スキ数")
```

投稿時間とスキ数の関係

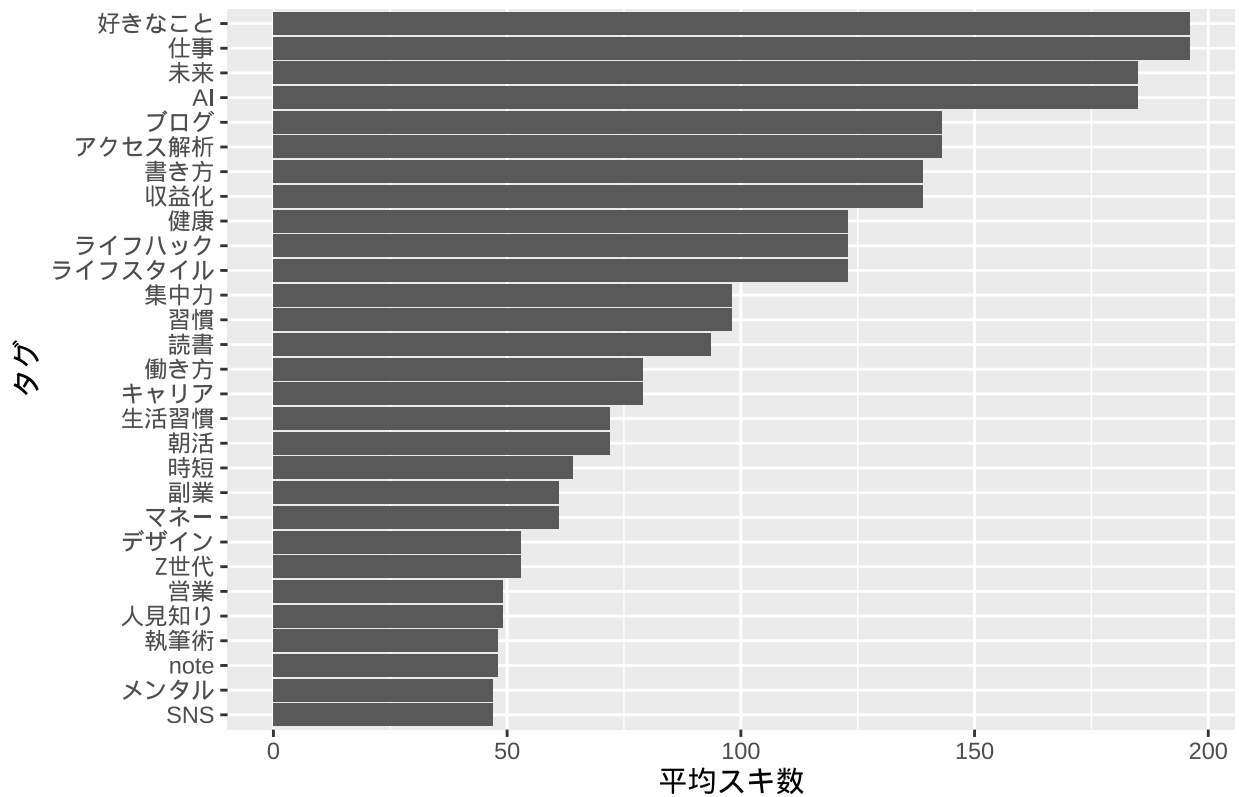


```
## 文字数とスキ数の関係
ggplot(note_data, aes(x = length, y = likes)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = " 文字数とスキ数の関係", x = " 文字数", y = " スキ数")
```



```
## タグ別の平均スキ数
tag_summary <- note_data_long %>%
  group_by(tag) %>%
  summarise(mean_like = mean(likes), .groups = "drop")
ggplot(tag_summary, aes(x = reorder(tag, mean_like), y = mean_like)) +
  geom_col() +
  coord_flip() +
  labs(title = "タグ別平均スキ数", x = "タグ", y = "平均スキ数")
```

## タグ別平均スキ数



### ## スキ数が高い記事の特徴を見る

```
note_data %>%
  arrange(desc(likes)) %>%
  select(title, title_num, time, length, pv, likes, like_rate) %>%
  head(5)
```

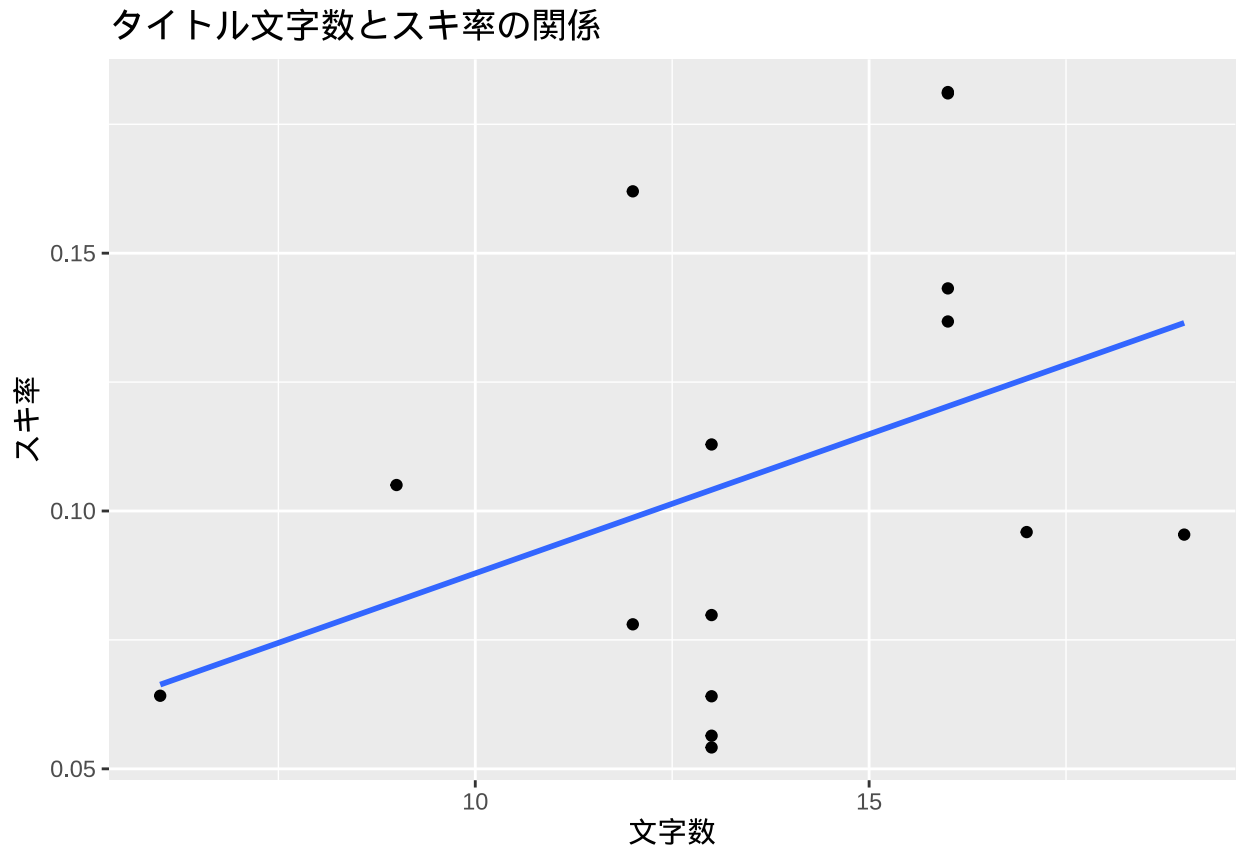
```
## # A tibble: 5 x 7
```

	title	title_num	time	length	pv	likes	like_rate
	<chr>	<int>	<tim>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	「好き」を仕事にするための第一歩	16	01:31	1948	1369	196	0.143
## 2	生成AIとの共存を考える	12	10:03	1785	1142	185	0.162
## 3	ブログアクセスが3倍になった理由	16	03:19	2368	789	143	0.181
## 4	noteを使って収益化できるか？	16	09:32	1323	768	139	0.181
## 5	30代でキャリアに悩んだときに読んだ本~~	19	18:11	1716	1289	123	0.0954

## 6 データ解析（エンゲージメント：スキ率）

## タイトルの文字数とスキ率の関係

```
ggplot(note_data, aes(x = title_num, y = like_rate)) +  
  geom_point() +  
  geom_smooth(method = "lm", se = FALSE) +  
  labs(title = " タイトル文字数とスキ率の関係", x = " 文字数", y = " スキ率")
```



## タイトルの文字数とスキ率との相関関係

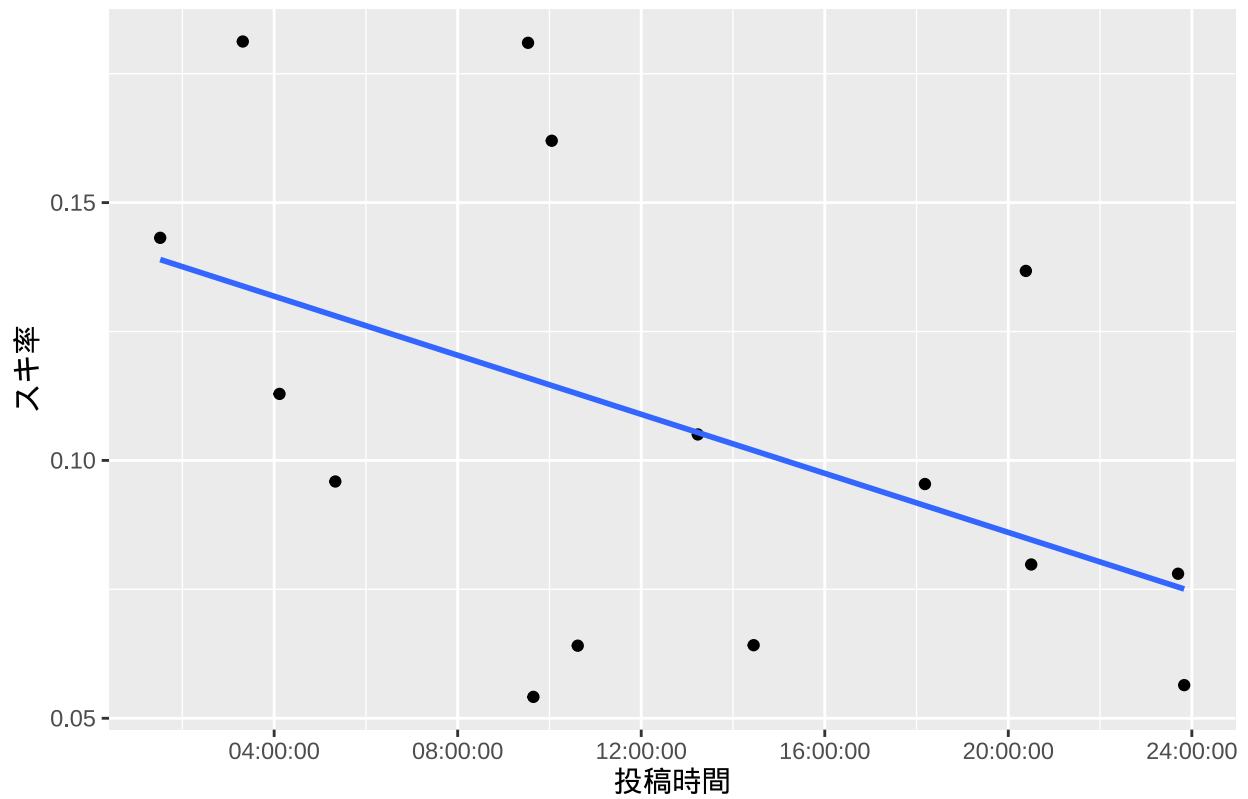
```
cor(note_data$title_num, note_data$like_rate)
```

```
## [1] 0.4008658
```

## 投稿時間とスキ率の関係

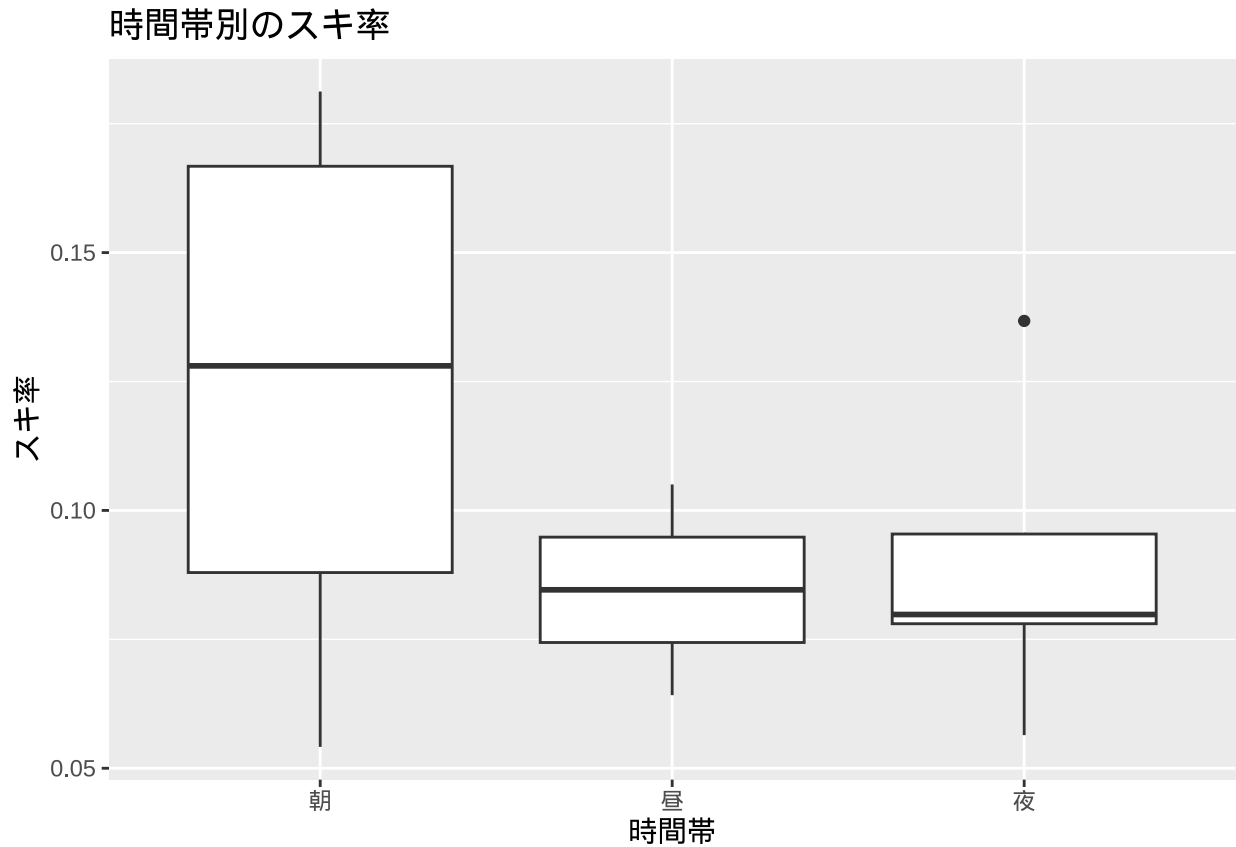
```
ggplot(note_data, aes(x = time, y = like_rate)) +  
  geom_point() +  
  geom_smooth(method = "lm", se = FALSE) +  
  labs(title = " 投稿時間とスキ率の関係", x = " 投稿時間", y = " スキ率")
```

投稿時間とスキ率の関係



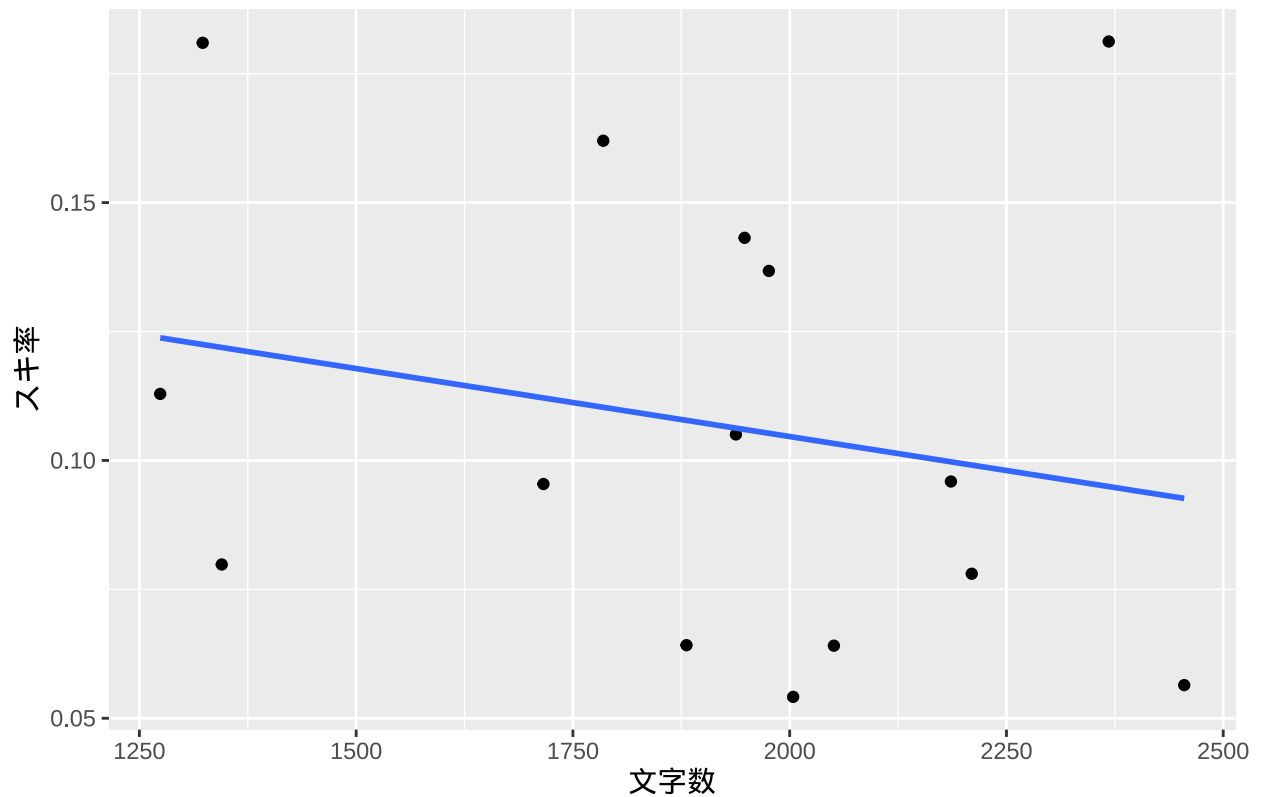
## 投稿時間帯とスキ率の関係

```
ggplot(note_data, aes(x = time_category, y = like_rate)) +  
  geom_boxplot() +  
  labs(title = "時間帯別のスキ率", x = "時間帯", y = "スキ率")
```



```
## 文字数とスキ率の関係
ggplot(note_data, aes(x = length, y = like_rate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = " 文字数とスキ率の関係", x = " 文字数", y = " スキ率")
```

文字数とスキ率の関係

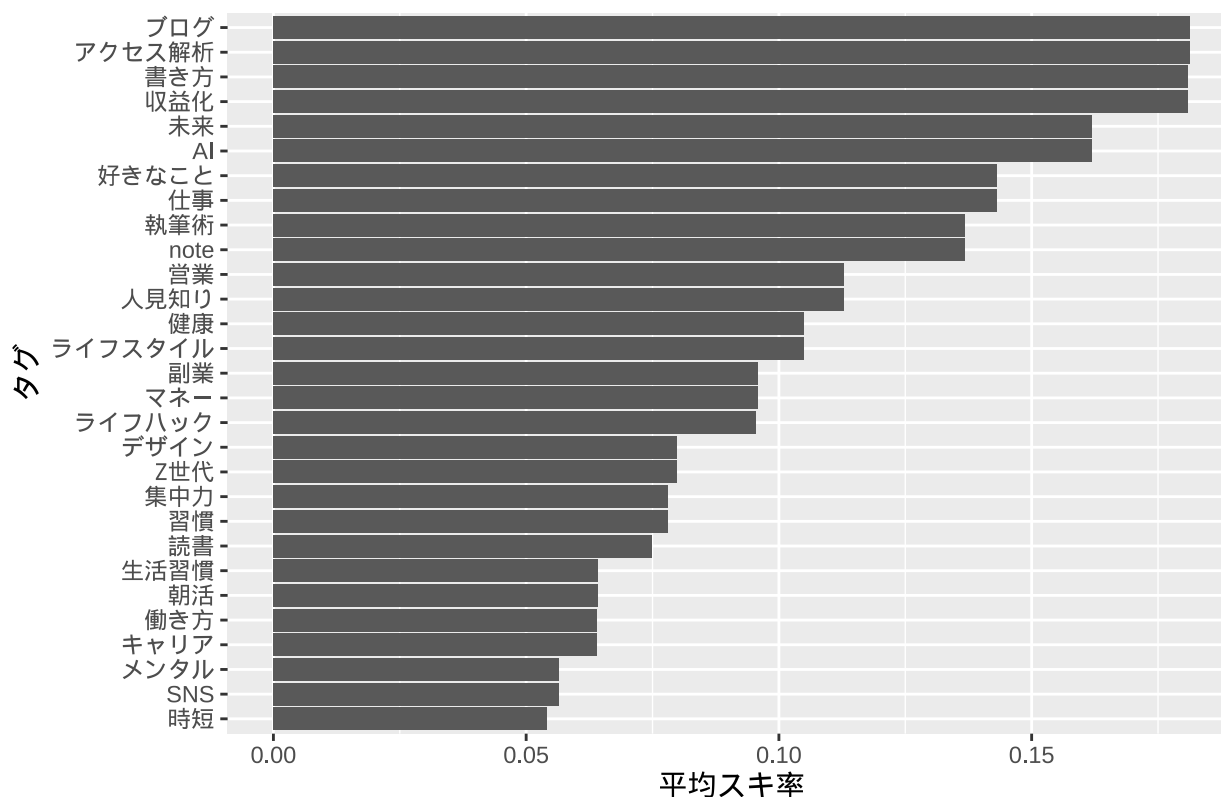


#### ## タグ別の平均スキ率

```
tag_summary <- note_data_long %>%
  group_by(tag) %>%
  summarise(mean_like_rate = mean(like_rate), .groups = "drop")
ggplot(tag_summary, aes(x = reorder(tag, mean_like_rate), y = mean_like_rate)) +
  geom_col() +
  coord_flip() +
  labs(title = "タグ別平均スキ率", x = "タグ", y = "平均スキ率")
```



## タグ別平均スキ率



### ## スキ率が高い記事の特徴を見る

```
note_data %>%
  arrange(desc(like_rate)) %>%
  select(title, title_num, time, length, pv, likes, like_rate) %>%
  head(5)
```

```
## # A tibble: 5 x 7
```

	title	title_num	time	length	pv	likes	like_rate
	<chr>	<int>	<time>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	ブログアクセスが3倍になった理由	16	03:19	2368	789	143	0.181
## 2	noteを使って収益化できるか？	16	09:32	1323	768	139	0.181
## 3	生成AIとの共存を考える	12	10:03	1785	1142	185	0.162
## 4	「好き」を仕事にするための第一歩	16	01:31	1948	1369	196	0.143
## 5	書くことがない日のnote活用術	16	20:23	1976	351	48	0.137

---

```
# 考察・まとめ
```

---

- 昼に投稿すると安定してよく閲覧されることがわかった。- 「好きなこと」「仕事」などのタグが高閲覧数に貢献していた。- 閲覧数が多いとスキと好評価をもらいやすい傾向にあった。- 昼の投稿は閲覧されやすいが、好評価はされにくいことがわかった。- 「好きなこと」「仕事」などのタグが閲覧数と同じようにスキに貢献していたが、閲覧されば好評価を受けやすいのは「ブログ」「アクセス解析」「書き方」「収益化」など実益に繋がりそうなタグがついた記事であった。

今後、投稿内容やタグに合わせて投稿時間を工夫することで、さらに多くのユーザーに読まれる可能性がある。

---

## 7 参考情報

- 仮想データ：step01\_note\_virtual\_data.csv
- 使用言語：R (tidyverse, readr, knitr, kableExtra, dplyr, tidyr, stringr, showtext, gridExtra, grid)