Replication: Estimating the Ideology of Political YouTube Videos

Political Analysis (2024), **0**, 1–16
doi:10.1017/pan.2023.42

PA

ARTICLE

Estimating the Ideology of Political YouTube Videos

Angela Lai[1,2], Megan A. Brown[1,6], James Bisbee[1], Joshua A. Tucker[1,3], Jonathan Nagler[1,3] and Richard Bonneau[1,2,4,5]

[1]Center for Social Media and Politics, New York University, New York, NY, USA; [2]Center for Data Science, New York University, New York, NY, USA; [3]Politics Department, New York University, New York, NY, USA; [4]Computer Science Department, New York University, New York, NY, USA; [5]Department of Biology, New York University, New York, NY, USA; [6]School of Information, University of Michigan, Ann Arbor, MI, USA

Corresponding author: Angela Lai; Email: csmap@nyu.edu

Junjie LIU

Background and Purpose

1. Increasing Influence of Online Media
2. Previous Work on Ideology Estimation
    1. Political actor
    2. Social Media
3. Gaps in Existing Research
    1. Rely on static Human-Labeled dataset

1. Develop a Novel Ideological Estimation Method
2. Application: Evaluate the presence of echo chambers

Methodology: Video-Subreddit Matrix

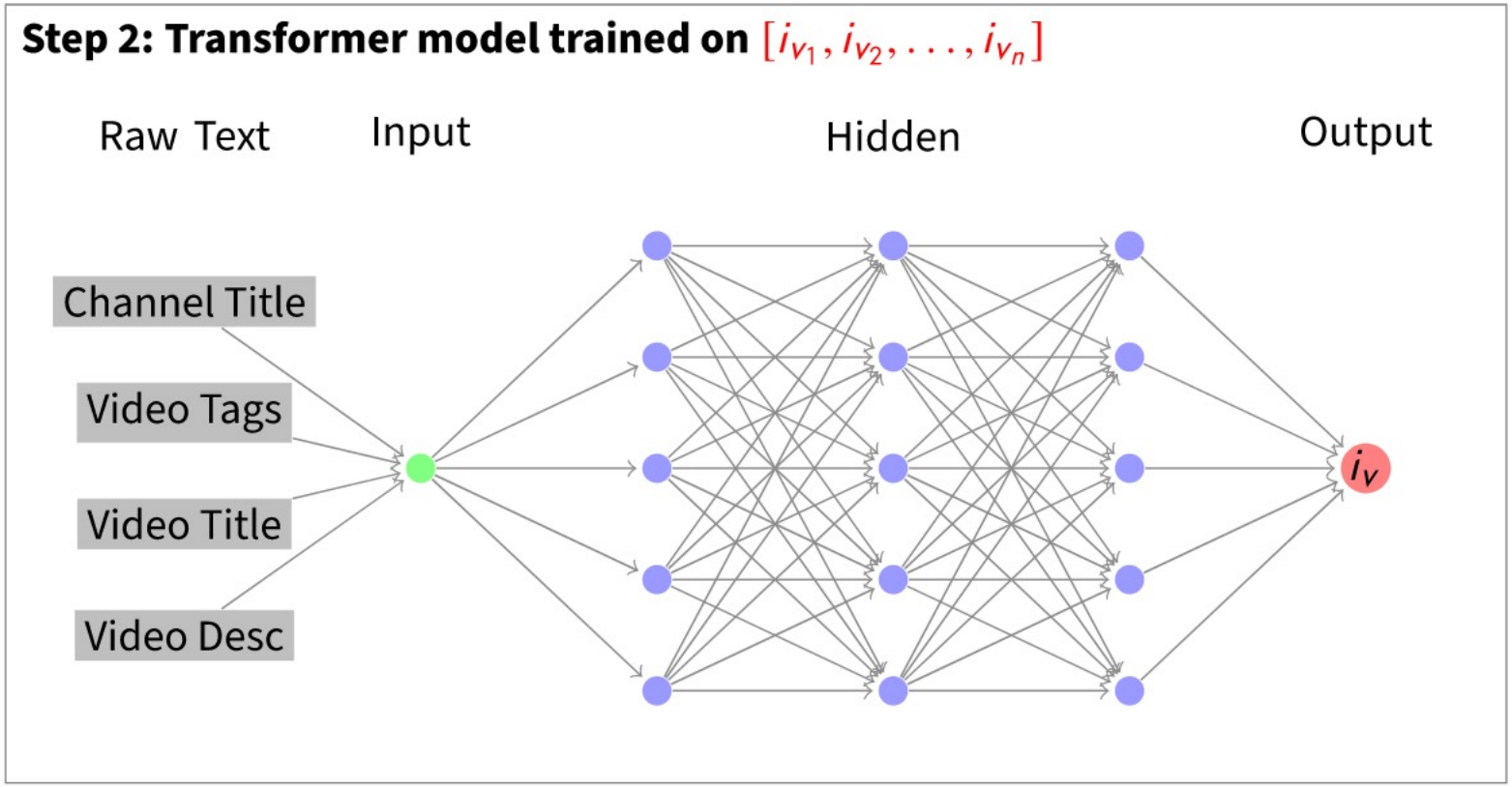
Identifying Political Subreddits

Seed Subreddits + Community Detection $\rightarrow [s_1, s_2, \ldots, s_m]$

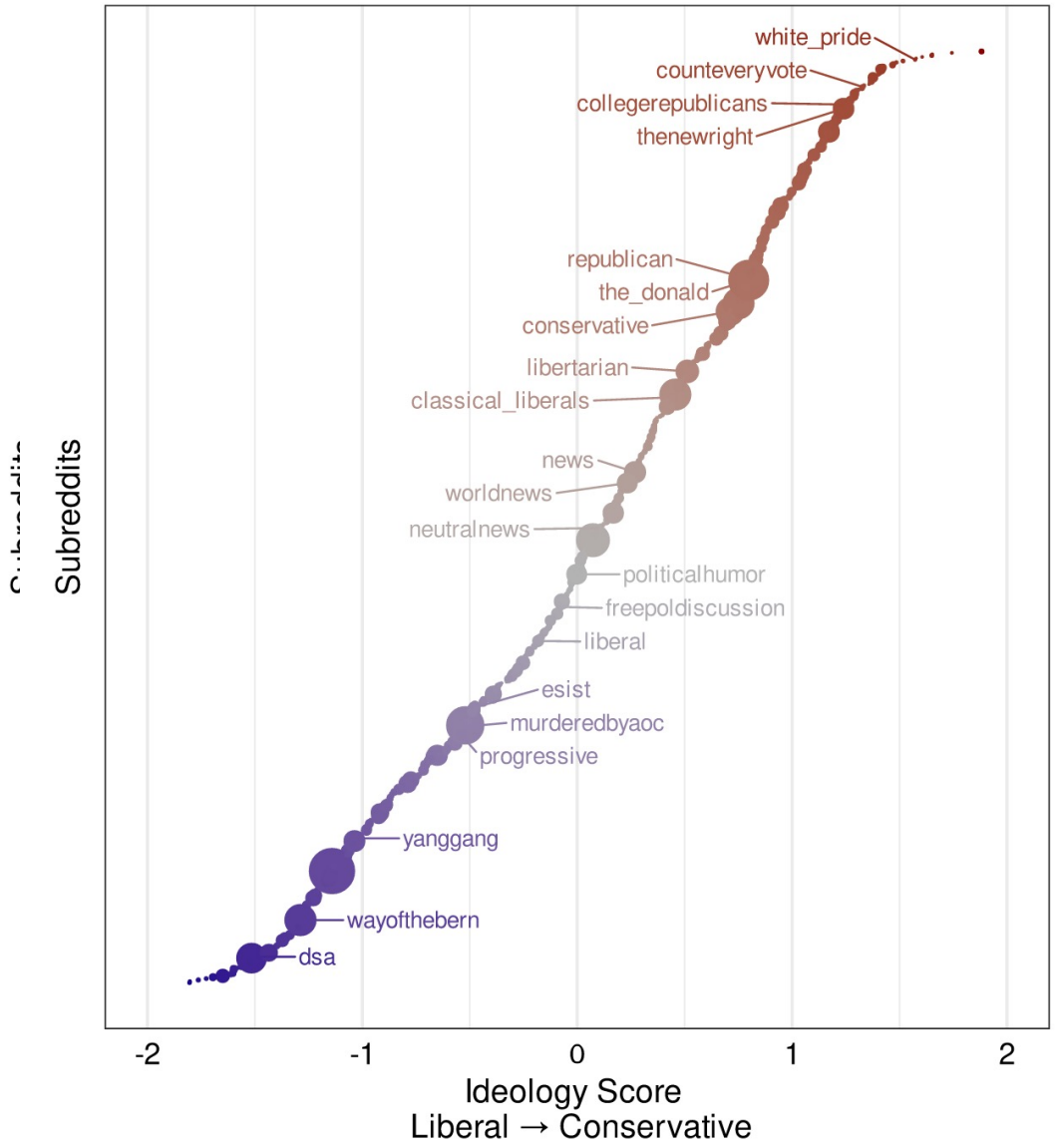
Step 1: Correspondence Analysis on video-subreddit matrix

$$\begin{array}{c} \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{array} \begin{array}{c} \begin{array}{cccc} s_1 & s_2 & \ldots & s_m \end{array} \\ \begin{bmatrix} 3 & 8 & \cdots & 4 \\ 8 & 0 & \cdots & 4 \\ \vdots & \vdots & \ddots & \vdots \\ 4 & 9 & \cdots & 0 \end{bmatrix} \end{array} \rightarrow \begin{array}{c} \\ v_1 \\ v_2 \\ \\ v_n \end{array} \begin{array}{c} \begin{array}{cc} CA_1 & CA_2 \end{array} \\ \begin{bmatrix} .79 & -.11 \\ -.67 & -1.21 \\ \vdots & \vdots \\ .02 & .47 \end{bmatrix} \end{array} \rightarrow [i_{v_1}, i_{v_2}, \ldots, i_{v_n}]$$

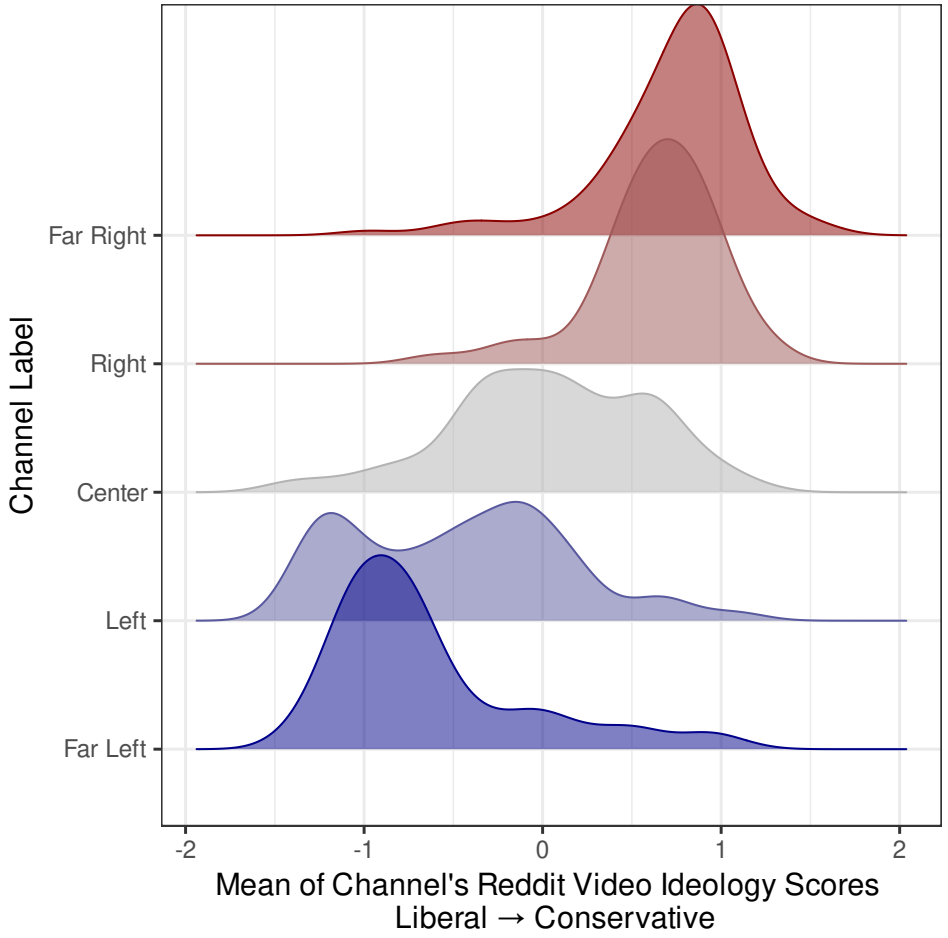
Methodology: Deep Learning Model for Video Ideology Training

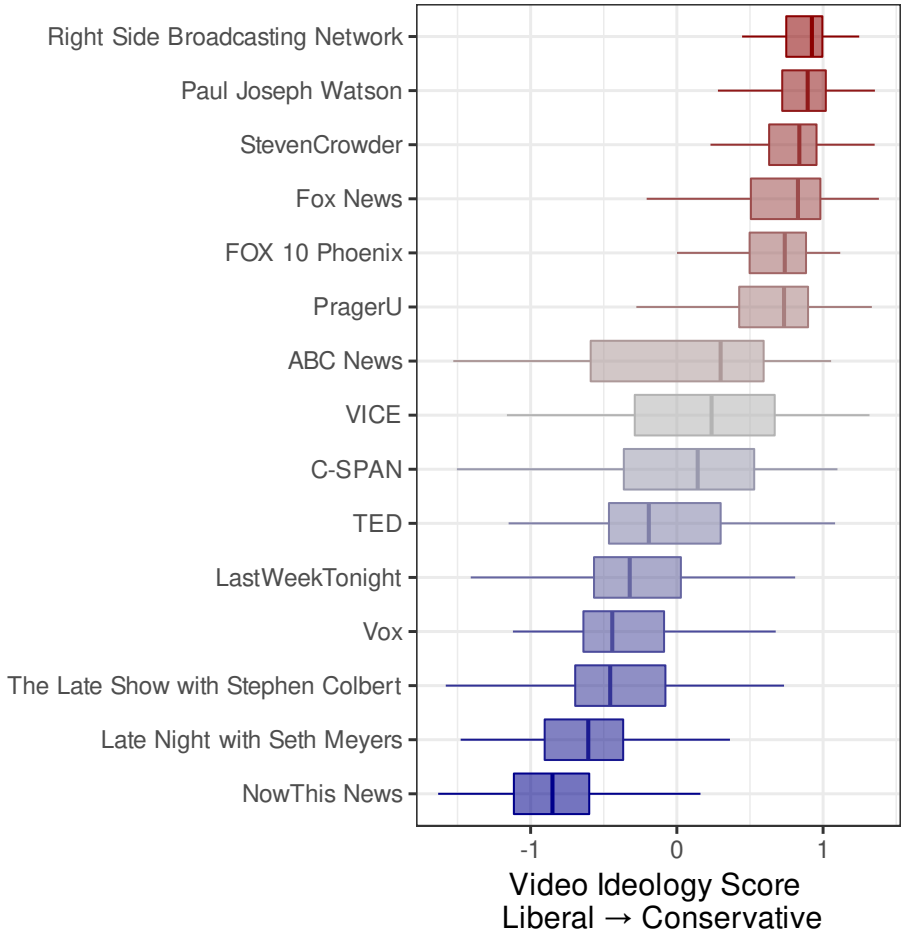


Result: Ideology Score based on the Subreddit Matrix



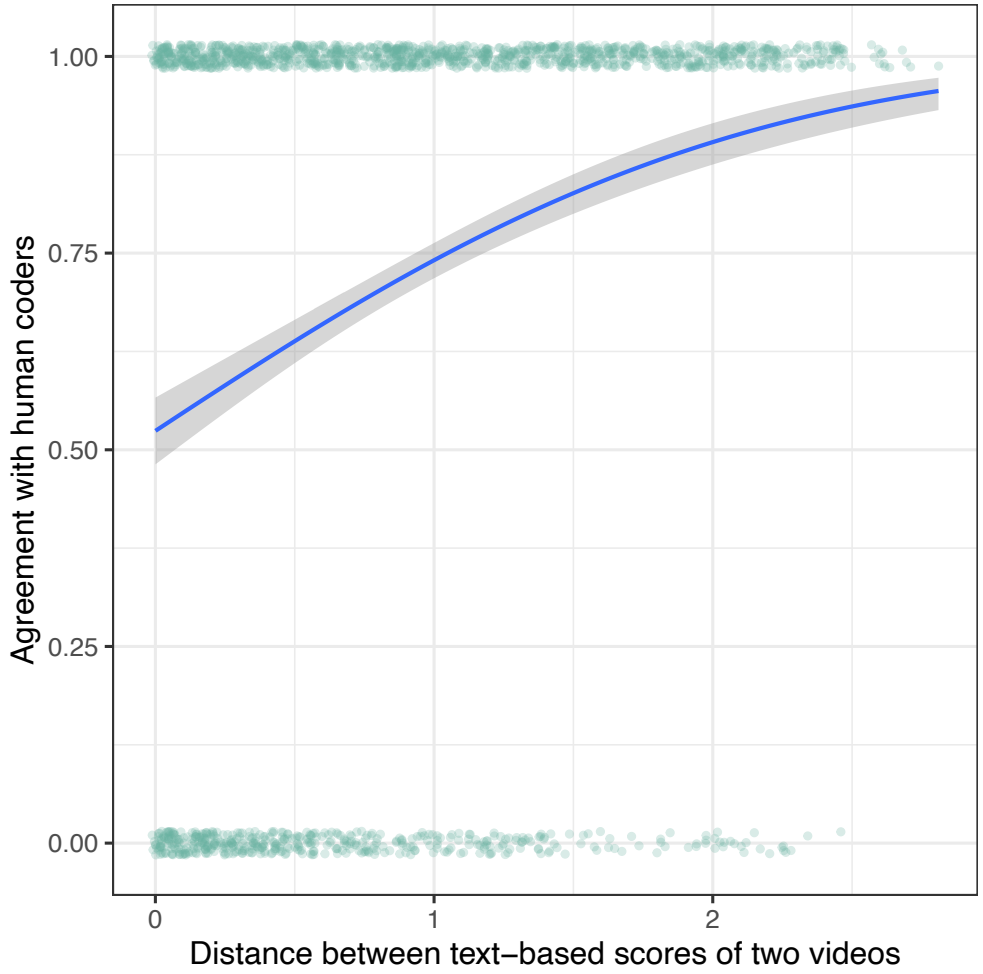
Result: Ideology Distributions



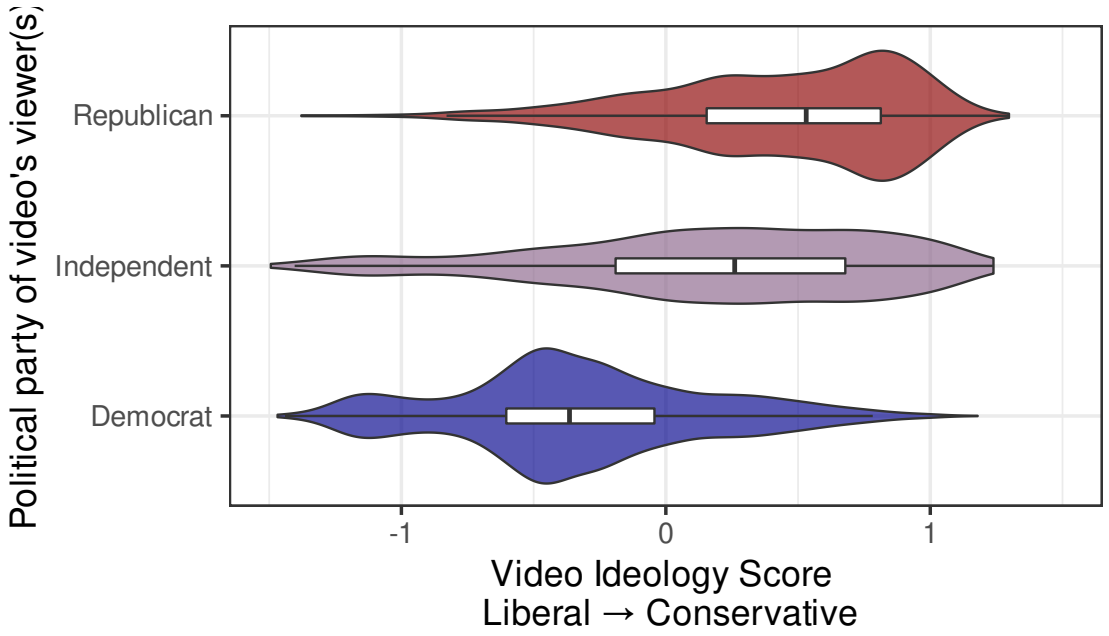


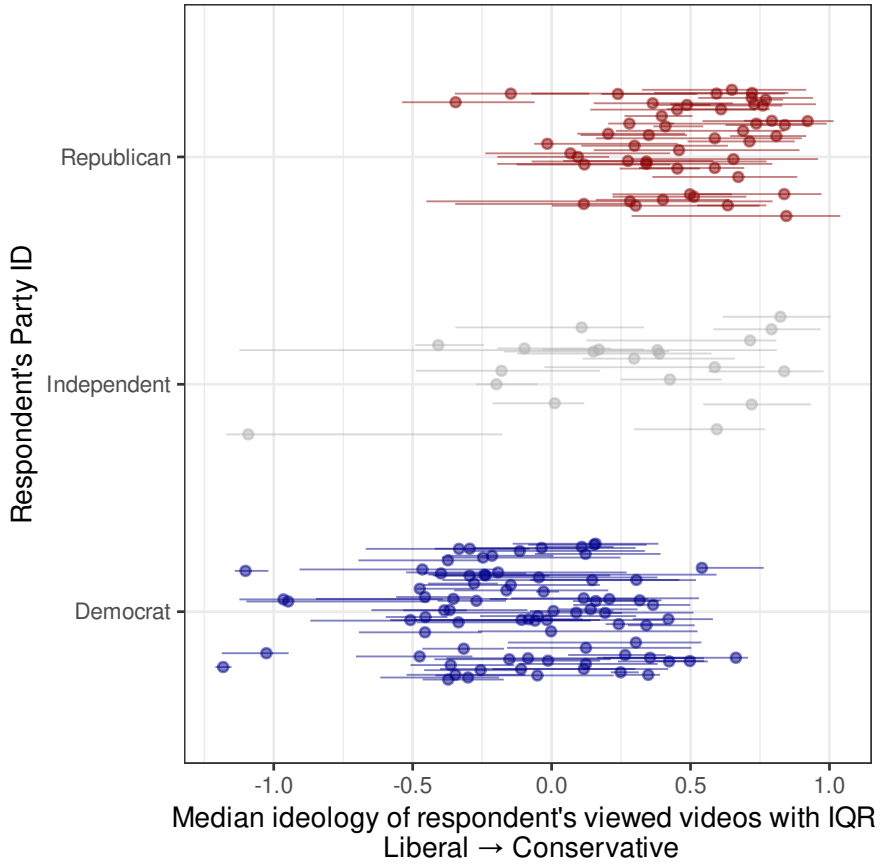
Result: Consistency of Score and Human Coders

Score distance	% Agreement
(0.0,0.25]	54.88
(0.25,0.5]	58.68
(0.5, 1.0]	71.02
(1.0, 1.5]	77.21
(1.5,2.0]	88.94
(2.0,3.0]	88.70



Application: Media Diets and Echo Chambers





My Contribution: Sentiment & Ideology

Sentiment and Ideology Score:

1. Sentiment (using Vader)
2. Ideology (using proposed Matrix)
3. Use LM to estimate the relation between them:

```
> m1 = lm(dat$ca_score~dat$senti)
> summary(m1)

Call:
lm(formula = dat$ca_score ~ dat$senti)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8868 -0.6959  0.1139  0.7273  1.8013

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02870    0.03507   0.818   0.4135
dat$senti   -0.06789    0.03149  -2.156   0.0314 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8497 on 683 degrees of freedom
Multiple R-squared:  0.006759,  Adjusted R-squared:  0.005305
F-statistic: 4.648 on 1 and 683 DF,  p-value: 0.03144
```