TBD*

TBD

Yingying Zhou, Yang Wu, Xinyi Xu, Hong Pan

06 March 2021

Abstract

First sentence. Second sentence. Third sentence. Fourth sentence.

Contents

---

*Code and data are available at: https://github.com/yangg1224/Political_Deepfake_Videos.git.

1 Introduction

2 Data

2.1 EDA
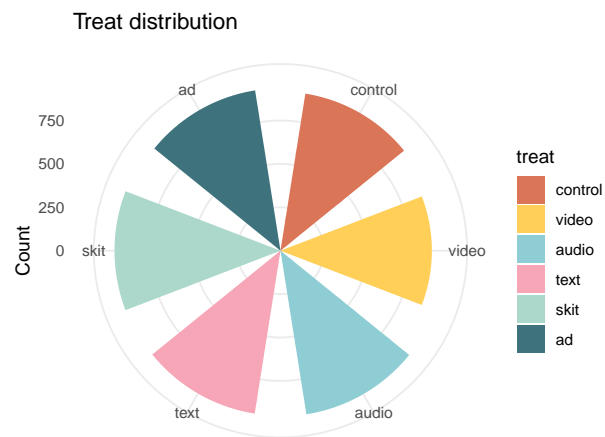
2.1.1 treat distribution

Figure 1: Employee numbers distribution

2.1.2 education level distribution by PID

2.1.3 sexism by education level

2.2 internet usage frequency by education level

```
## Warning: Use of `EDA$educ` is discouraged. Use `educ` instead.

## Warning: Use of `EDA$meta_OS` is discouraged. Use `meta_OS` instead.

## Warning: Use of `EDA$age_65` is discouraged. Use `age_65` instead.
```

2.2.1 post favor by treat

2.2.2 Average deception level by treat
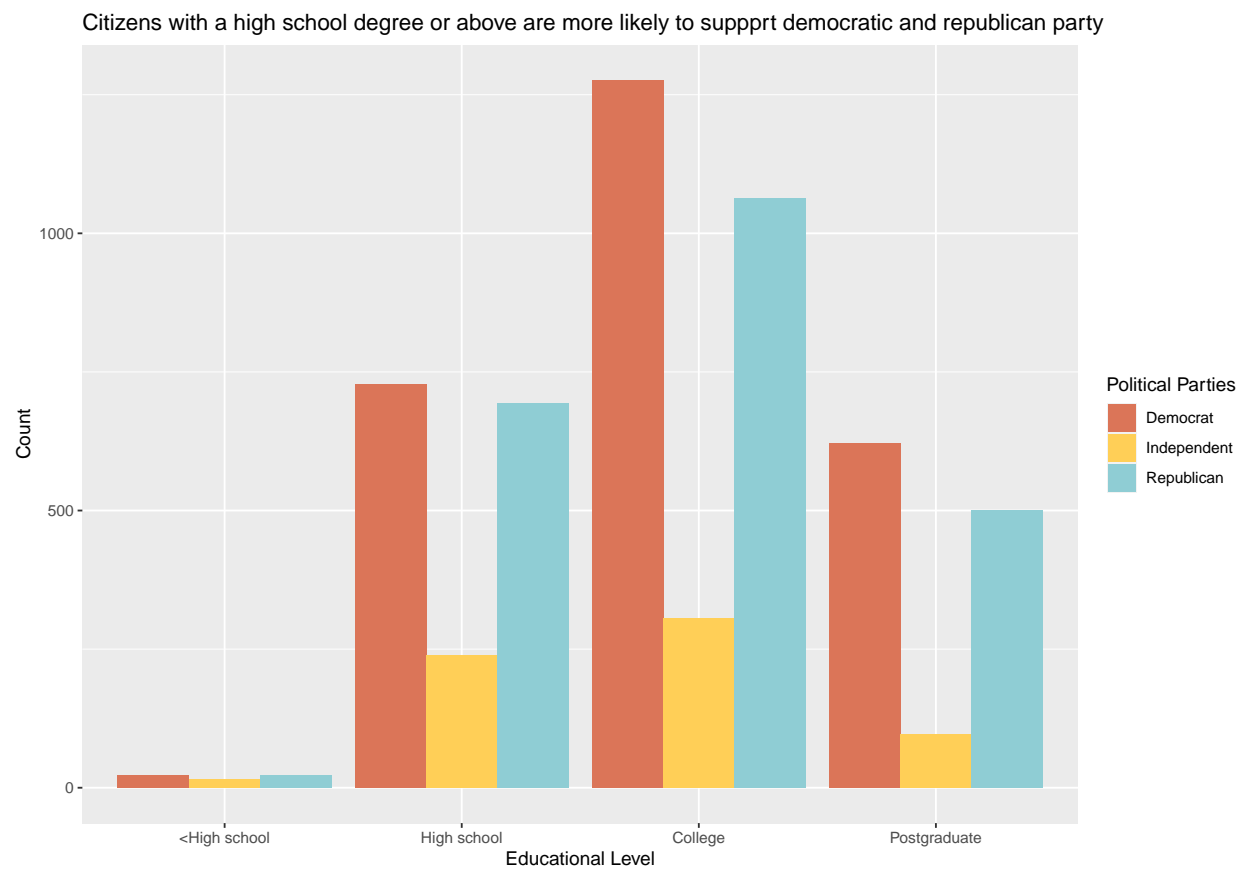
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Figure 2: Educational level by PID

Table 1: Average deception level of each media format

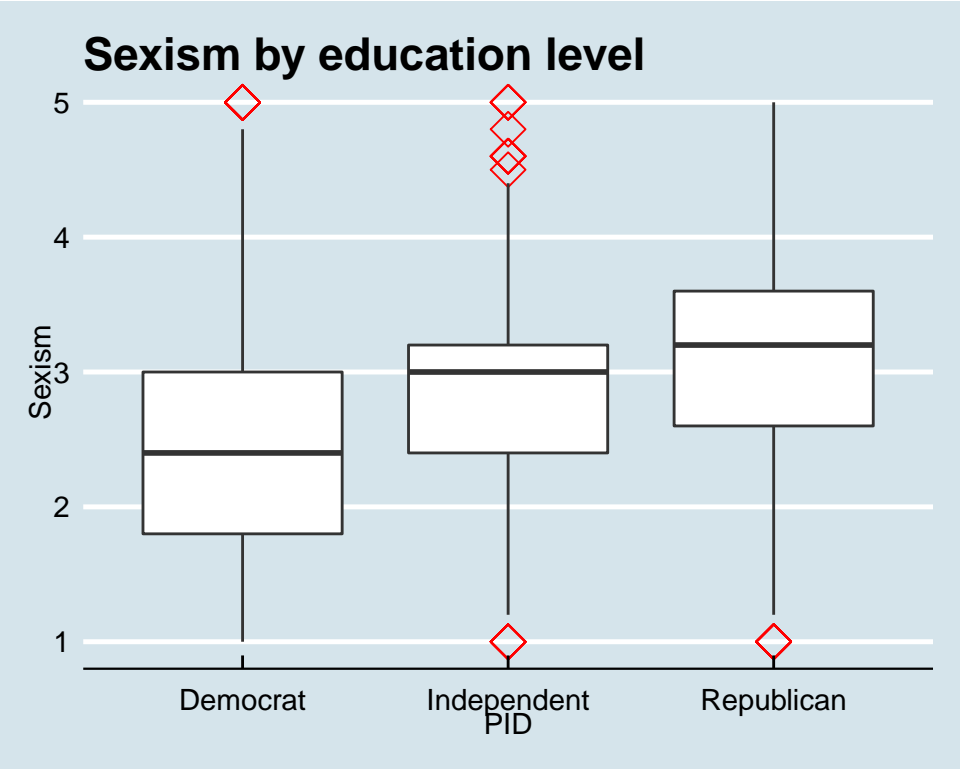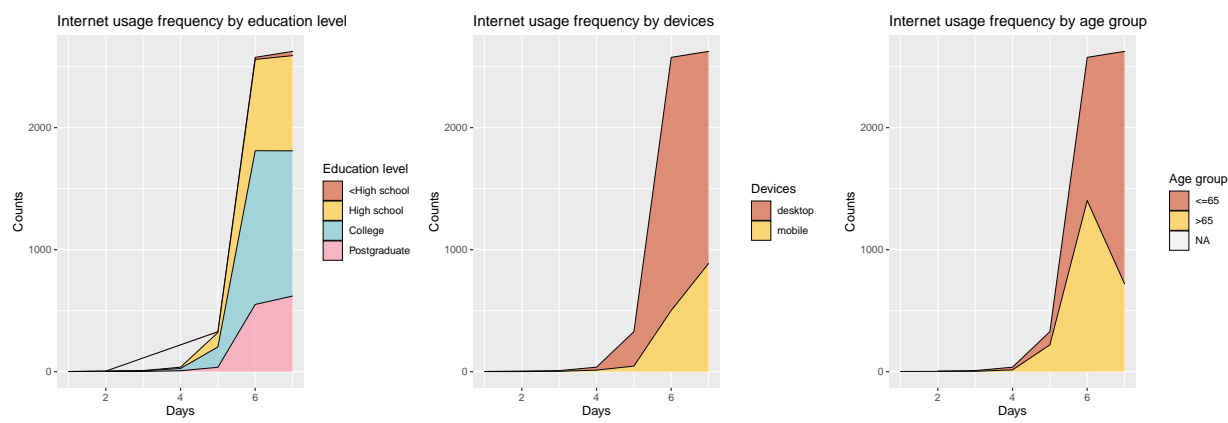| treat	Average Deception Level
video	3.227538
audio	3.351178
text	3.304442
skit	2.568519
ad	2.991228

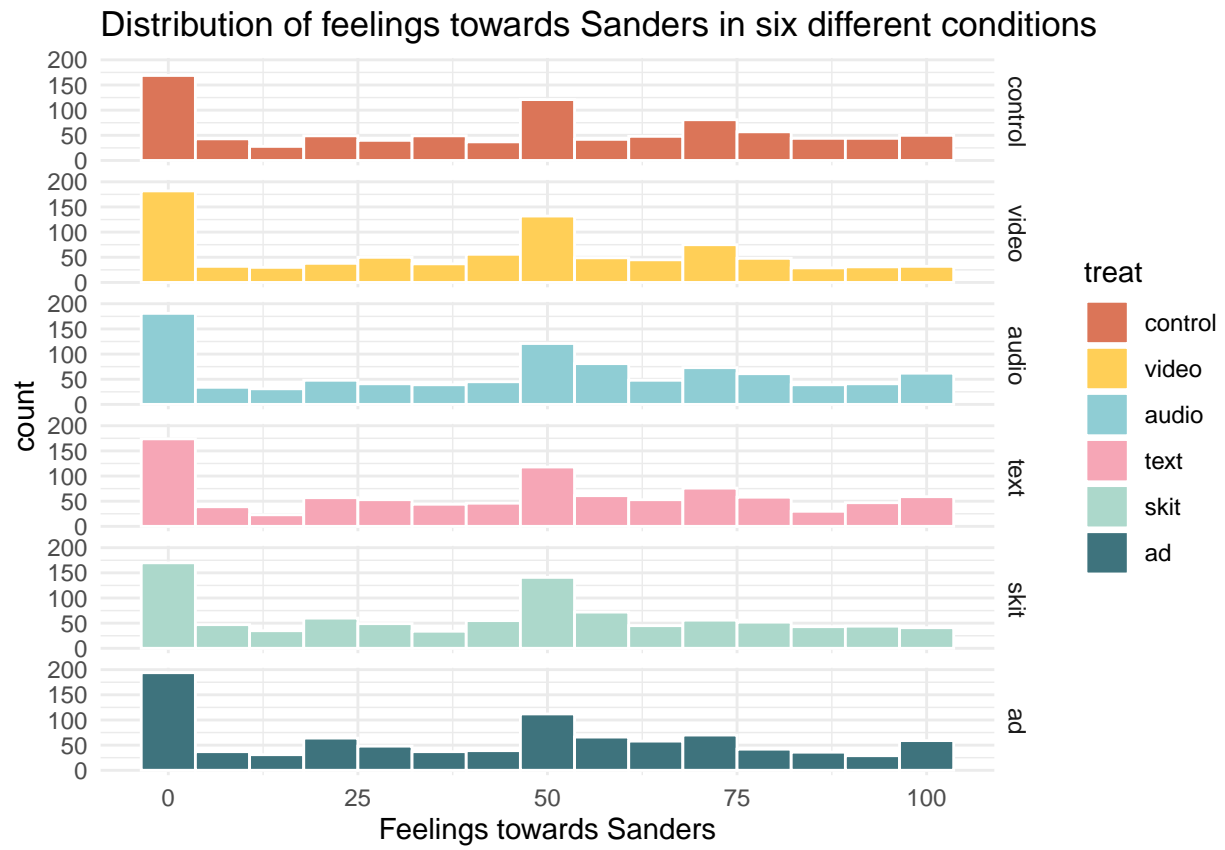Figure 3: sexism by education level

Figure 4: internet usages

Figure 5: Distribution of feelings towards Sanders in six different situations

Table 2: T test: Deception level of video vs audio

AVG__deception__video	AVG__deception__audio	p.value	conf.low	conf.high	method	alternative
3.228438	3.348243	0.0538155	-0.2415774	0.0019682	Welch Two Sample t-test	two.sided

3 Model

3.1 T test1

3.2 T test2

Table 3: T test: Deception level of video vs text

AVG__deception__video	AVG__deception__text	p.value	conf.low	conf.high	method	alternative
3.228438	3.304207	0.2244956	-0.1980699	0.0465321	Welch Two Sample t-test	two.sided

3.3 T test3

Table 4: T test: Deception level of video vs skit

AVG__deception__video	AVG__deception__skit	p.value	conf.low	conf.high	method	alternative
3.228438	2.574586	0	0.5024785	0.8052267	Welch Two Sample t-test	two.sided

3.4 Feature selection

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(plyr)
```

```
## --------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)


## ------------------------------------------------------------------------------


##
## Attaching package: 'plyr'


## The following object is masked from 'package:here':
##
##      here


## The following object is masked from 'package:ggpubr':
##
##      mutate


## The following objects are masked from 'package:dplyr':
##
##      arrange, count, desc, failwith, id, mutate, rename, summarise,
##      summarize


## The following object is masked from 'package:purrr':
##
##      compact
```

```r
# deal with missing value
FS<- dat%>%
select(-c(PID_presurvey,comments,PID_leaners,script,quality_pretreat_duration_tooquick,quality_pretreat
na.omit()

set.seed(123)
FS_tree<-randomForest(post_favor_Sanders~., data=FS, importance=TRUE,proximity=TRUE)

FS_tree<-as.data.frame(FS_tree$importanceSD)
```

```r
FS_tree[order(-FS_tree[,1]),,drop=FALSE]%>%
  head(20)%>%
  kableExtra::kbl(caption="Feature importance")%>%
  kableExtra::kable_styling(latex_options = "scale_down")%>% # use scale_down option to make the font
  kableExtra::kable_styling(latex_options = c("hold_position"))
```

$$Pr(\theta|y) = \frac{Pr(y|\theta)Pr(\theta)}{Pr(y)} \tag{1}$$

Equation (1) seems useful, eh?

Here's a dumb example of how to use some references: In paper we run our analysis in R (R Core Team 2020). We also use the `tidyverse` which was written by Wickham et al. (2019) If we were interested in baseball data then Friendly et al. (2020) could be useful.

We can use maths by including latex between dollar signs, for instance θ.

Table 5: Feature importance

	FS_tree$importanceSD
PID	6.5440027
PID_main	6.3462560
age	2.2151690
ambivalent_sexism	1.8626093
agegroup	1.6713382
believed_true	1.4076484
EndDate	1.3557475
StartDate	1.3546429
polknow	1.2407895
HHI	1.2094671
Zip	1.1898840
duration_secs	1.0828439
educ	1.0762351
age_65	1.0254143
post_dig_lit	0.9188808
treat	0.9073697
meta_screenres	0.8755030
meta_resolution	0.8685490
polknow_warren	0.7383741
Region	0.6179277

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

A Appendix

A.1 missing value

```
## Loading required package: colorspace

## Loading required package: grid

## VIM is ready to use.

## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues

##
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':
##
##     sleep

## Warning in plot.aggr(res, ...): not enough vertical space to display frequencies
## (too many combinations)
```
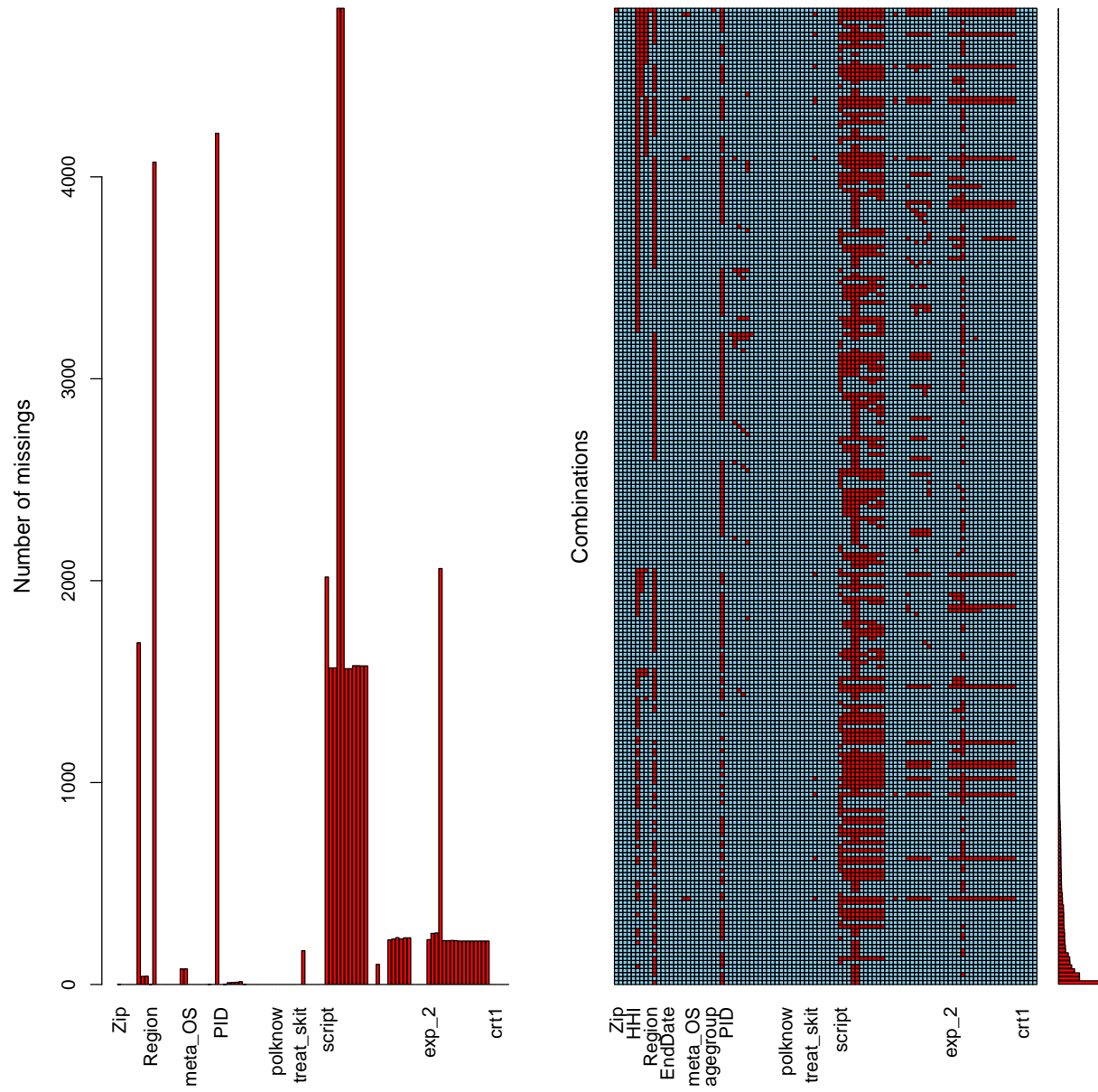
Figure 6: Missing value Visualization

References

Friendly, Michael, Chris Dalzell, Martin Monkman, and Dennis Murphy. 2020. *Lahman: Sean "Lahman" Baseball Database.* https://CRAN.R-project.org/package=Lahman.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.