TBD*

TBD

Yingying Zhou, Yang Wu, Xinyi Xu, Hong Pan

16 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

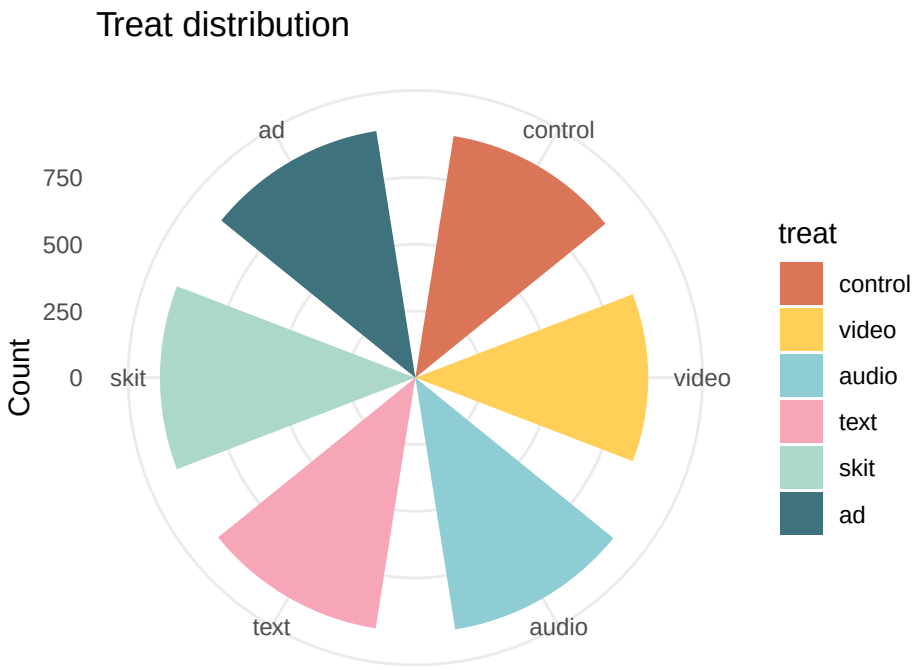


Figure 1: Employee numbers distribution

2.1.2 education level distribution by PID

2.1.3 sexism by education level

2.2 internet usage frequency by education level

2.2.1 post favor by treat

2.2.2 Average deception level by treat

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

3 Model

3.1 Deception Level

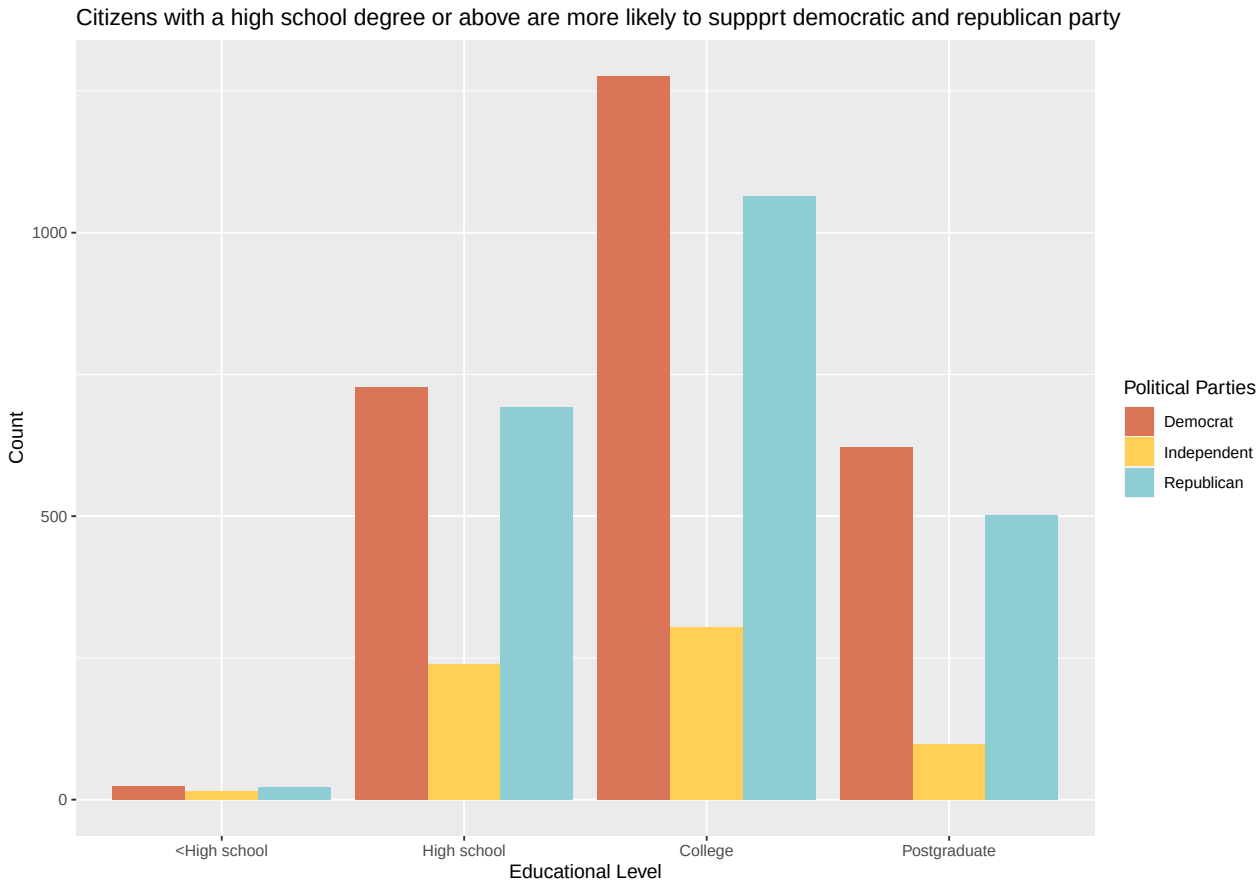


Figure 2: Educational level by PID

Table 1: Average deception level of each media format

treat	Average Deception Level
video	3.228438
audio	3.351178
text	3.305946
skit	2.569316
ad	2.989059

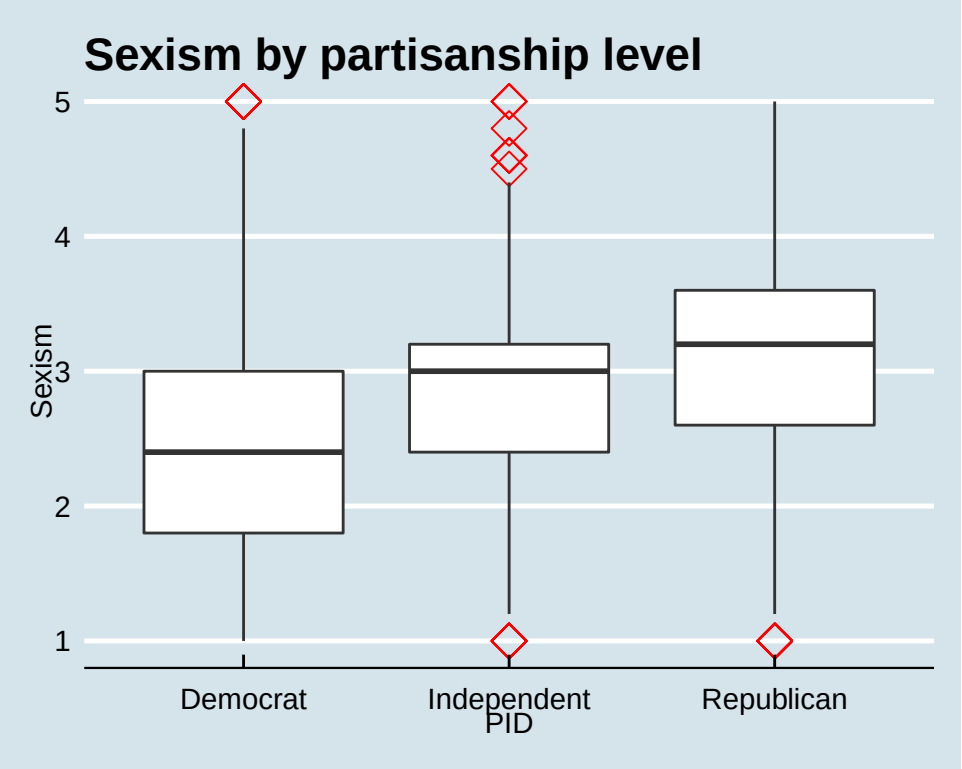


Figure 3: sexism by education level

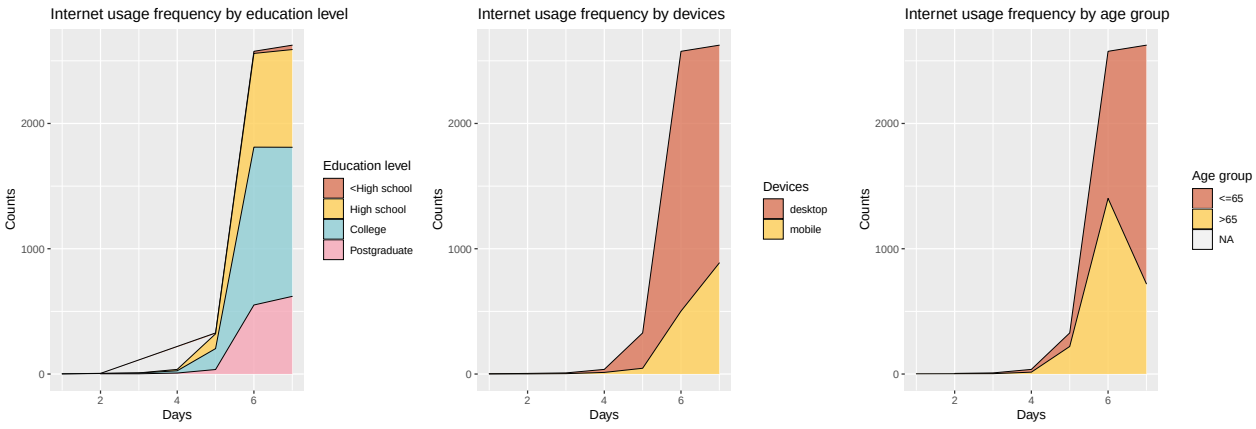


Figure 4: internet usages

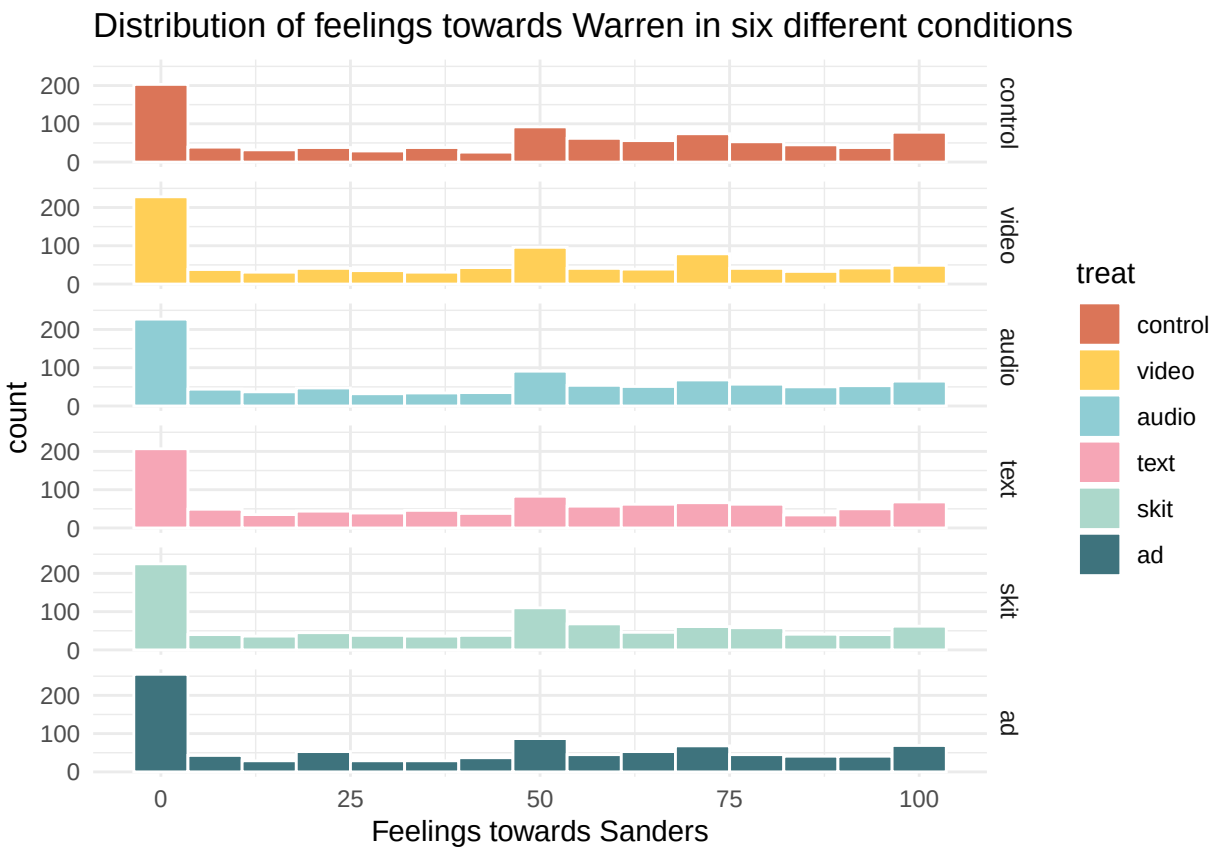


Figure 5: Distribution of feelings towards Sanders in six different situations

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

3.1.1 T test1

3.1.2 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

3.1.3 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

3.2 Affect Level

3.2.1 T test 1

Table 5: T test: Affect level of video vs control

AVG_affect_video	AVG_affect_control	p.value	conf.low	conf.high	method	alternative
41.27797	45.81395	0.005278	-7.721219	-1.350748	Welch Two Sample t-test	two.sided

3.2.2 T test2

Table 6: T test: Affect level of video vs text

AVG_affect_video	AVG_affect_text	p.value	conf.low	conf.high	method	alternative
41.27797	44.2234	0.0652461	-6.077025	0.1861569	Welch Two Sample t-test	two.sided

3.2.3 T test3

Table 7: T test: Affect level of video vs audio

AVG_affect_video	AVG_affect_audio	p.value	conf.low	conf.high	method	alternative
41.27797	43.92593	0.0997404	-5.80127	0.5053586	Welch Two Sample t-test	two.sided

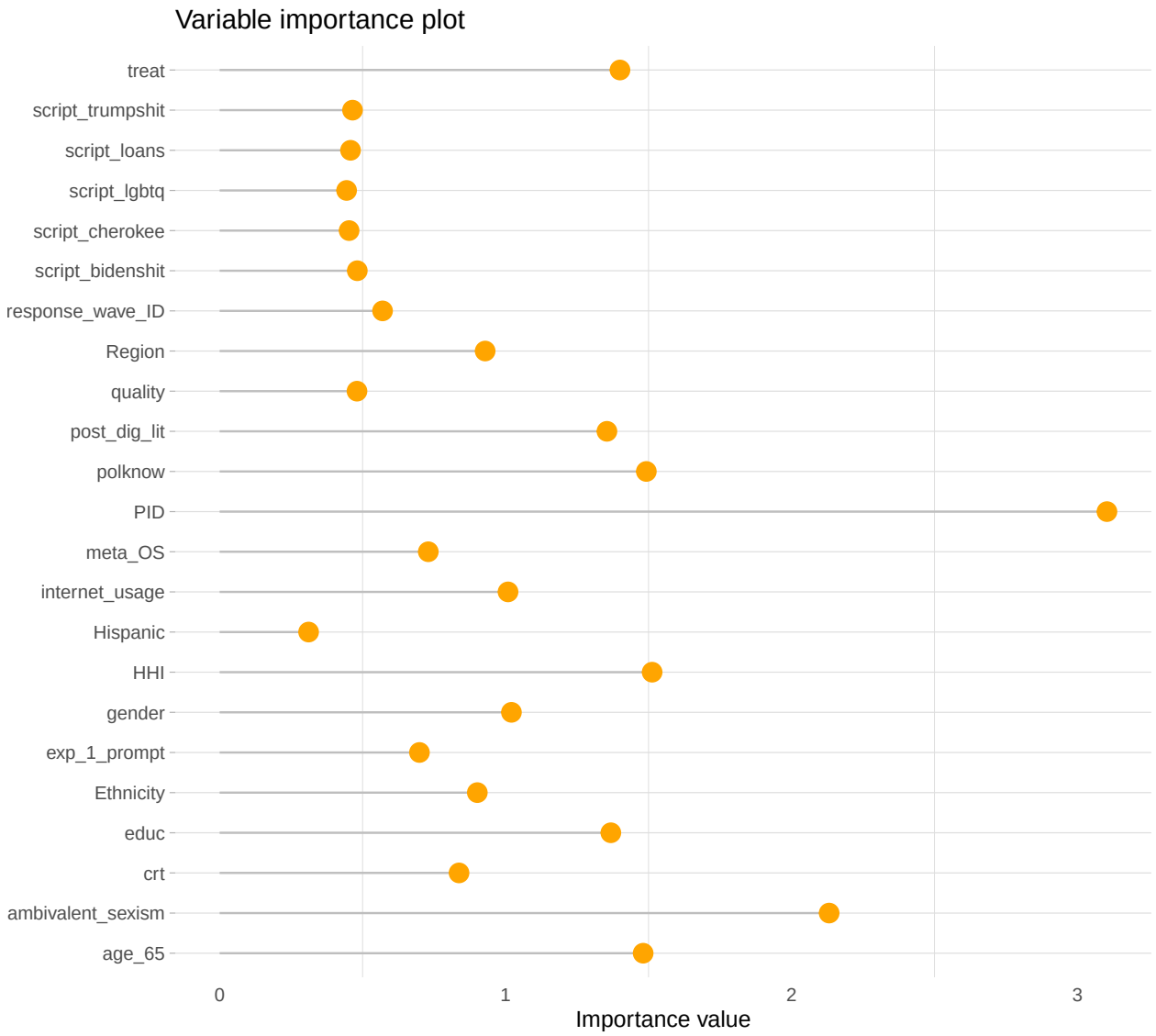
Table 8: T test: Affect level of video vs skit

AVG_affect_video	AVG_affect_skit	p.value	conf.low	conf.high	method	alternative
41.27797	43	0.2772717	-4.829684	1.385625	Welch Two Sample t-test	two.sided

3.2.4 T test 4

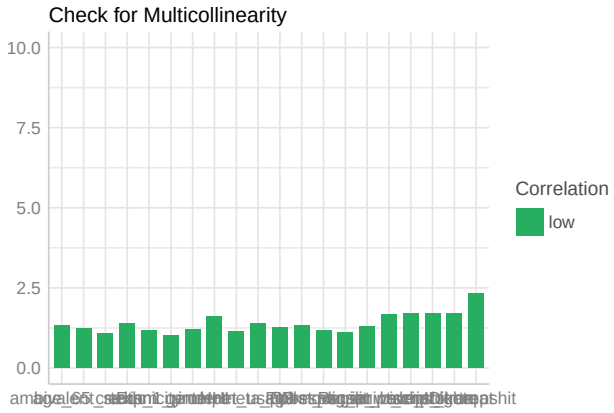
3.3 Feature selection

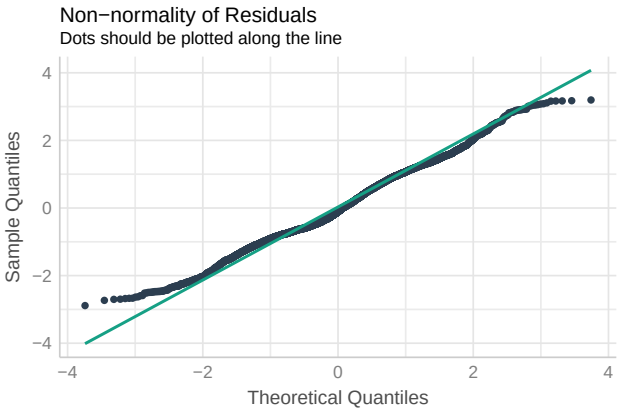
```
## randomForest 4.6-14

## 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

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

## 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
```

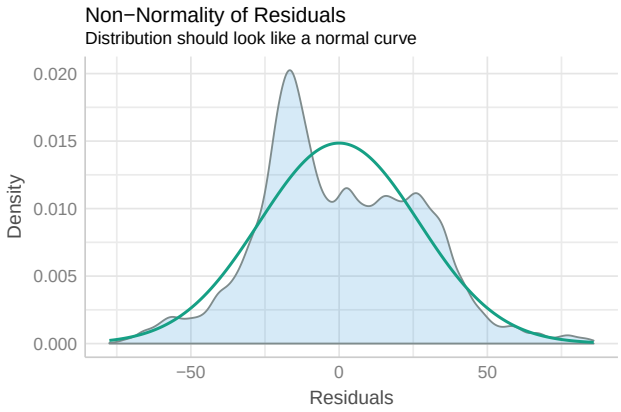


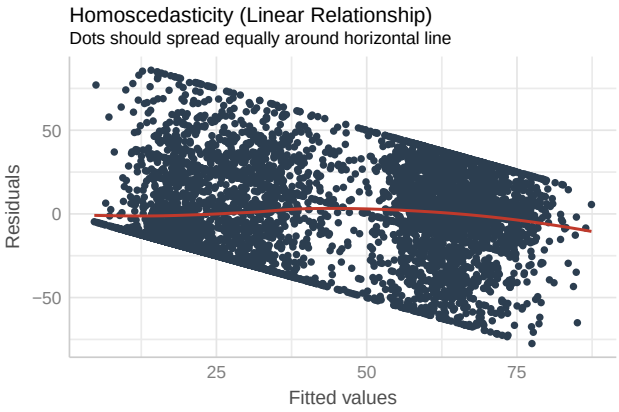
3.4 regression model

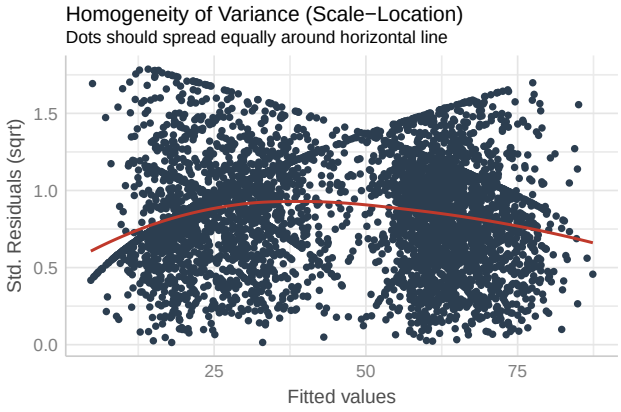
```
## Loading required namespace: qqplotr

## For confidence bands, please install ‘qqplotr‘.

## ‘geom_smooth()‘ using formula ’y ~ x’
## ‘geom_smooth()‘ using formula ’y ~ x’

## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```


Check for Multicollinearity
10.0
7.5
5.0
2.5
0.0
Correlation
low


Non−normality of Residuals
Dots should be plotted along the line
Sample Quantiles
Theoretical Quantiles
4
2
0
−2
−4


Non−Normality of Residuals
Distribution should look like a normal curve
Density
Residuals
0.020
0.015
0.010
0.005
0.000
−50
0
50


Homoscedasticity (Linear Relationship)
Dots should spread equally around horizontal line
Residuals
Fitted values
50
0
−50
25
50
75


Homogeneity of Variance (Scale−Location)
Dots should spread equally around horizontal line
Std. Residuals (sqrt)
Fitted values
1.5
1.0
0.5
0.0
25
50
75

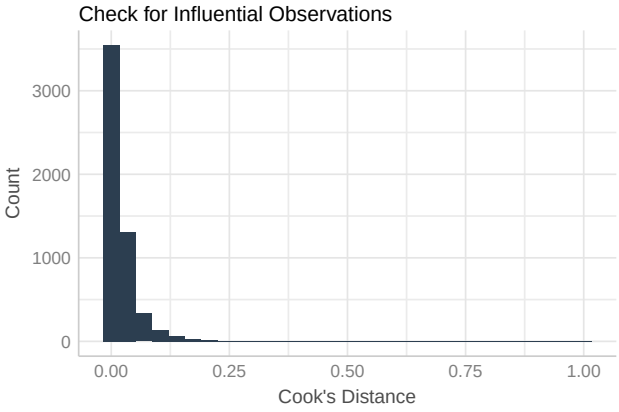

Check for Influential Observations
Count
Cook's Distance
3000
2000
1000
0
0.00
0.25
0.50
0.75
1.00

Table 9: Regression results

	Model 1
(Intercept)	69.872***
	(6.448)
gender.L	-1.560
	(4.824)
gender.Q	-0.969
	(2.806)
educHigh school	-1.207
	(3.606)
educCollege	0.709
	(3.605)
educPostgraduate	9.317**
	(3.693)
HHI$100k-$150k	3.904**
	(1.757)
HHI>$150k	4.007**
	(1.939)
HHI$25k-$49k	-0.919
	(1.041)
HHI$50k-$74k	-0.208
	(1.172)
HHI$75k-$99k	0.057
	(1.151)
HHIN/A	-4.947*
	(2.920)
EthnicityBlack	3.113
	(2.459)
EthnicityOther	-1.064
	(2.643)
EthnicityWhite	0.563
	(1.921)
RegionNortheast	3.891***
	(1.106)
RegionSouth	0.453
	(0.988)
RegionWest	0.923
	(1.164)
response_wave_IDSV_eyxdeXOuISXzakt	-1.306
	(0.945)
meta_OSmobile	-1.269
	(0.987)
age_65>65	-4.357***
	(0.847)
PIDIndependent	-26.689***
	(1.222)
PIDRepublican	-39.521***
	(0.856)
ambivalent_sexism	-4.064***
	(0.477)
polknow	1.333
	(1.724)
treatvideo	-2.531
	(1.551)
treataudio	-2.817*
	(1.530)
treattext	-1.725
	(1.555)
treatskit	-2.850*
	(1.527)
treatad	-3.838***
	(1.274)
script_bidenshit	-1.680
	(1.406)
script_trumpshit	-2.132
	(1.394)
script_cherokee	-1.145
	(1.401)
script_lgbtq	0.647
	(1.401)
exp_1_promptinfo	0.763
	(0.731)
post_dig_lit	-3.950
	(2.775)
internet_usage	1.120*
	(0.582)
crt	-0.893
	(1.639)
Num.Obs.	5468
R2	0.386
R2 Adj.	0.382
AIC	51582.2
BIC	51839.8
Log.Lik.	-25752.077
F	92.263

* p < 0.1, ** p < 0.05, *** p < 0.01

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

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.

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)
```

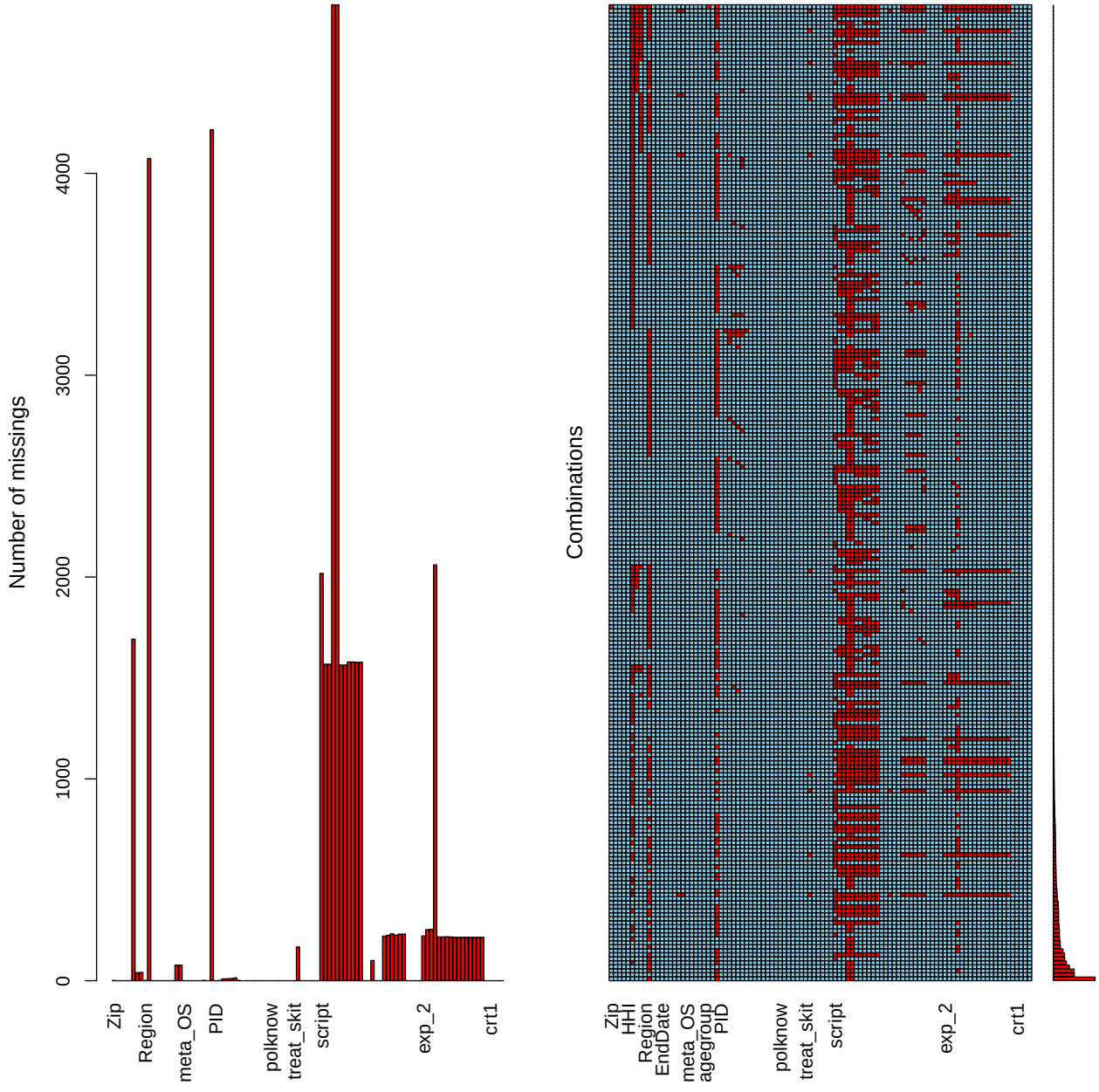


Figure 6: Missing value Visualization

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.