

TBD*
TBD

Yingying Zhou, Yang Wu, Xinyi Xu, Hong Pan

11 March 2021

Abstract

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

Contents

1	Introduction	2
2	Data	2
2.1	EDA	2
2.2	internet usage frequency by education level	2
3	Model	6
3.1	Deception Level	6
3.2	Affect Level	6
3.3	Feature selection	8
3.4	regression model	10
4	Results	12
5	Discussion	12
5.1	First discussion point	12
5.2	Second discussion point	12
5.3	Third discussion point	12
5.4	Weaknesses and next steps	12
A	Appendix	13
A.1	missing value	13
	References	15

*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

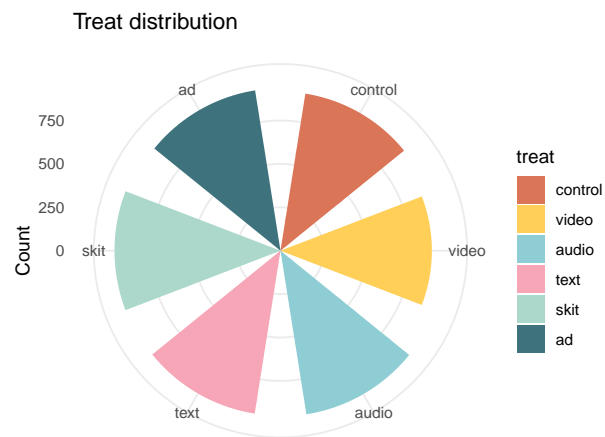


Figure 1: Employee numbers distribution

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)
```

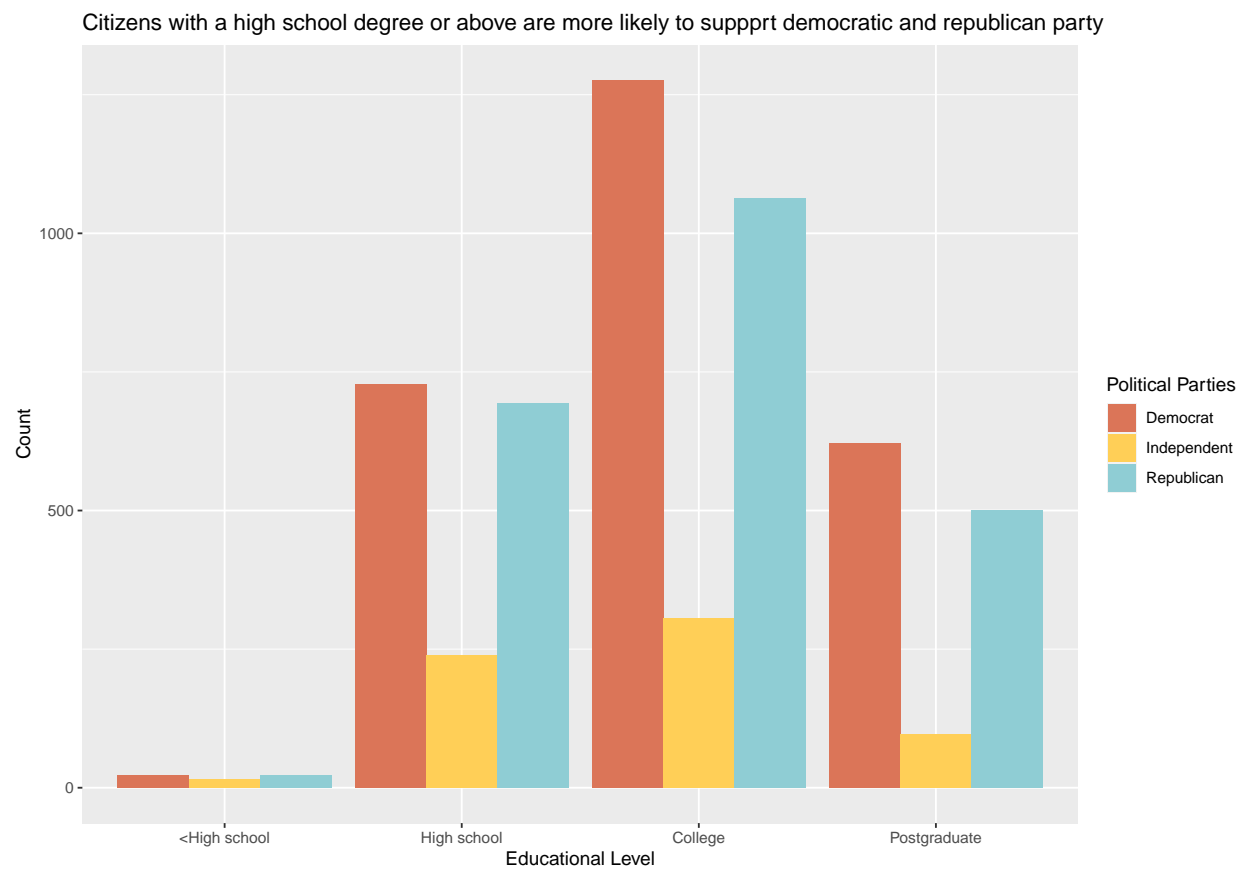


Figure 2: Educational level by PID

Table 1: Average deception level of each media format

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

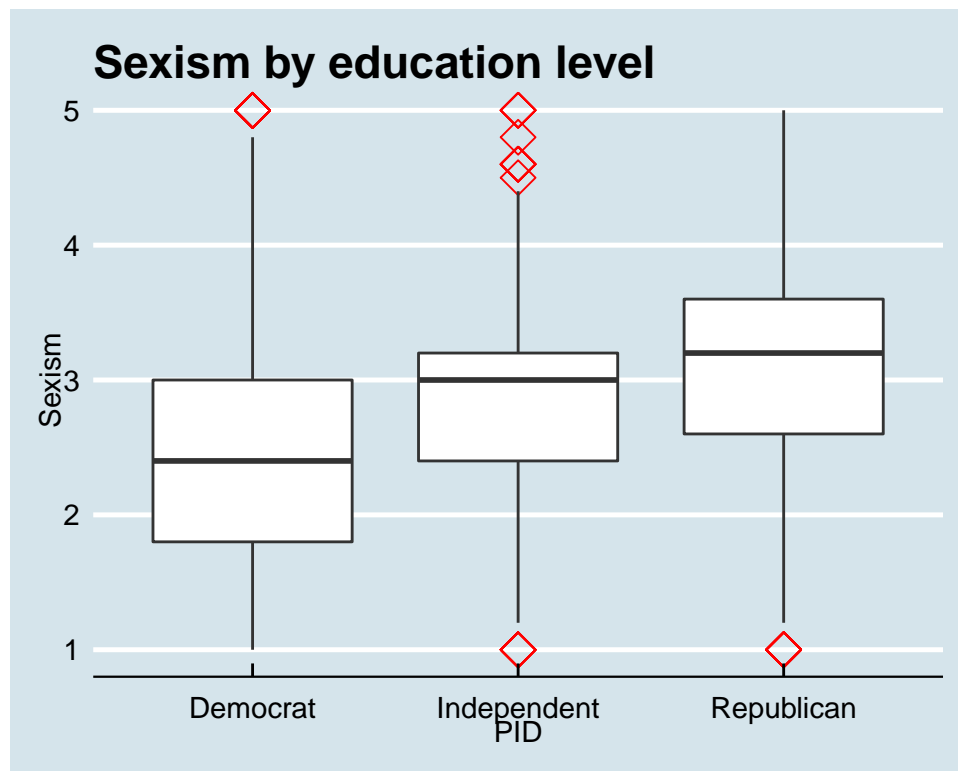


Figure 3: sexism by education level

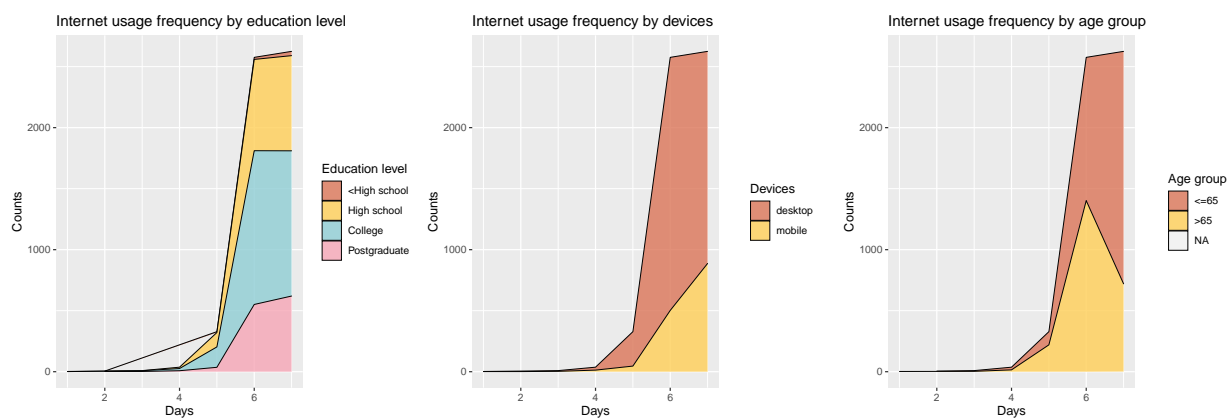


Figure 4: internet usages

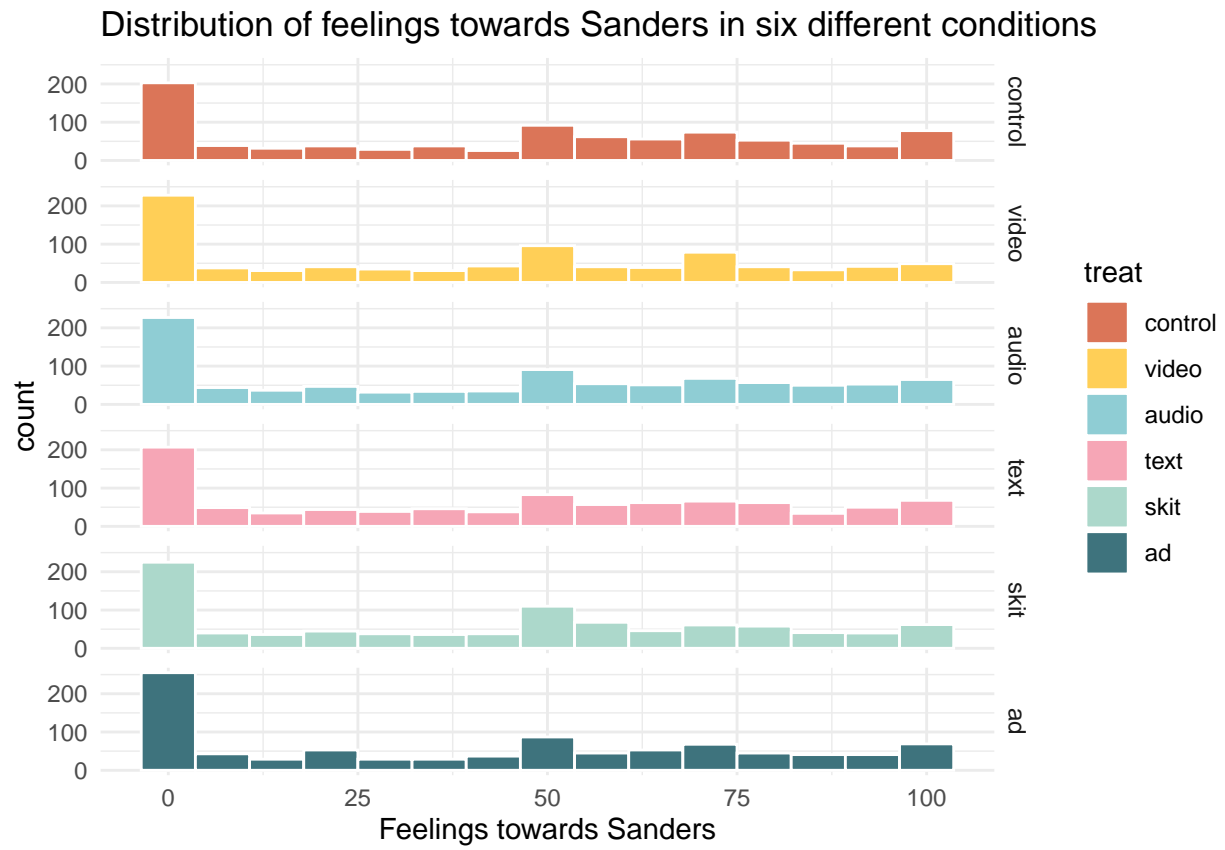


Figure 5: Distribution of feelings towards Sanders in six different situations

3 Model

3.1 Deception Level

3.1.1 T test1

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

3.1.2 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

```
#video vs control
t1_a<-t.test(na.omit(dat$post_favor_Warren[dat$treat_fake_video == 1]),
             na.omit(dat$post_favor_Warren[dat$treat_control == 1]))

m1_a<-tidy(t1_a,digits=1)

m1_a%>%
  select(-c(estimate,statistic,parameter))%>%
  rename(
    AVG_affect_video=estimate1,
    AVG_affect_control=estimate2
  )%>%
  kableExtra::kbl(caption = "T test: Affect level of video vs control")%>%
  kableExtra::kable_styling(latex_options = "scale_down")%>%# use scale_down option to make the font
  kableExtra::kable_styling(latex_options = c("hold_position"))
```

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

```
#video vs text
t2_a<-t.test(na.omit(dat$post_favor_Warren[dat$treat_fake_video == 1]),
             na.omit(dat$post_favor_Warren[dat$treat_fake_text == 1]))

m2_a<-tidy(t2_a,digits=1)

m2_a%>%
  select(-c(estimate,statistic,parameter))%>%
  rename(
    AVG_affect_video=estimate1,
    AVG_affect_text=estimate2
  )%>%
  kableExtra::kbl(caption = "T test: Affect level of video vs text")%>%
  kableExtra::kable_styling(latex_options = "scale_down")%>%# use scale_down option to make the font
  kableExtra::kable_styling(latex_options = c("hold_position"))
```

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

```
#video vs audio
t3_a<-t.test(na.omit(dat$post_favor_Warren[dat$treat_fake_video == 1]),
             na.omit(dat$post_favor_Warren[dat$treat_fake_audio == 1]))

m3_a<-tidy(t3_a,digits=1)

m3_a%>%
  select(-c(estimate,statistic,parameter))%>%
  rename(
    AVG_affect_video=estimate1,
    AVG_affect_audio=estimate2
  )%>%
  kableExtra::kbl(caption = "T test: Affect level of video vs audio")%>%
  kableExtra::kable_styling(latex_options = "scale_down")%>%# use scale_down option to make the font
  kableExtra::kable_styling(latex_options = c("hold_position"))
```

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

3.2.4 T test 4

```
#video vs skit
t4_a<-t.test(na.omit(dat$post_favor_Warren[dat$treat_fake_video == 1]),
             na.omit(dat$post_favor_Warren[dat$treat_skit == 1]))

m4_a<-tidy(t4_a,digits=1)

m4_a%>%
  select(-c(estimate,statistic,parameter))%>%
  rename(
    AVG_affect_video=estimate1,
    AVG_affect_skit=estimate2
  )%>%
  kableExtra::kbl(caption = "T test: Affect level of video vs skit")%>%
  kableExtra::kable_styling(latex_options = "scale_down")%>%# use scale_down option to make the font
  kableExtra::kable_styling(latex_options = c("hold_position"))
```

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

3.3 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

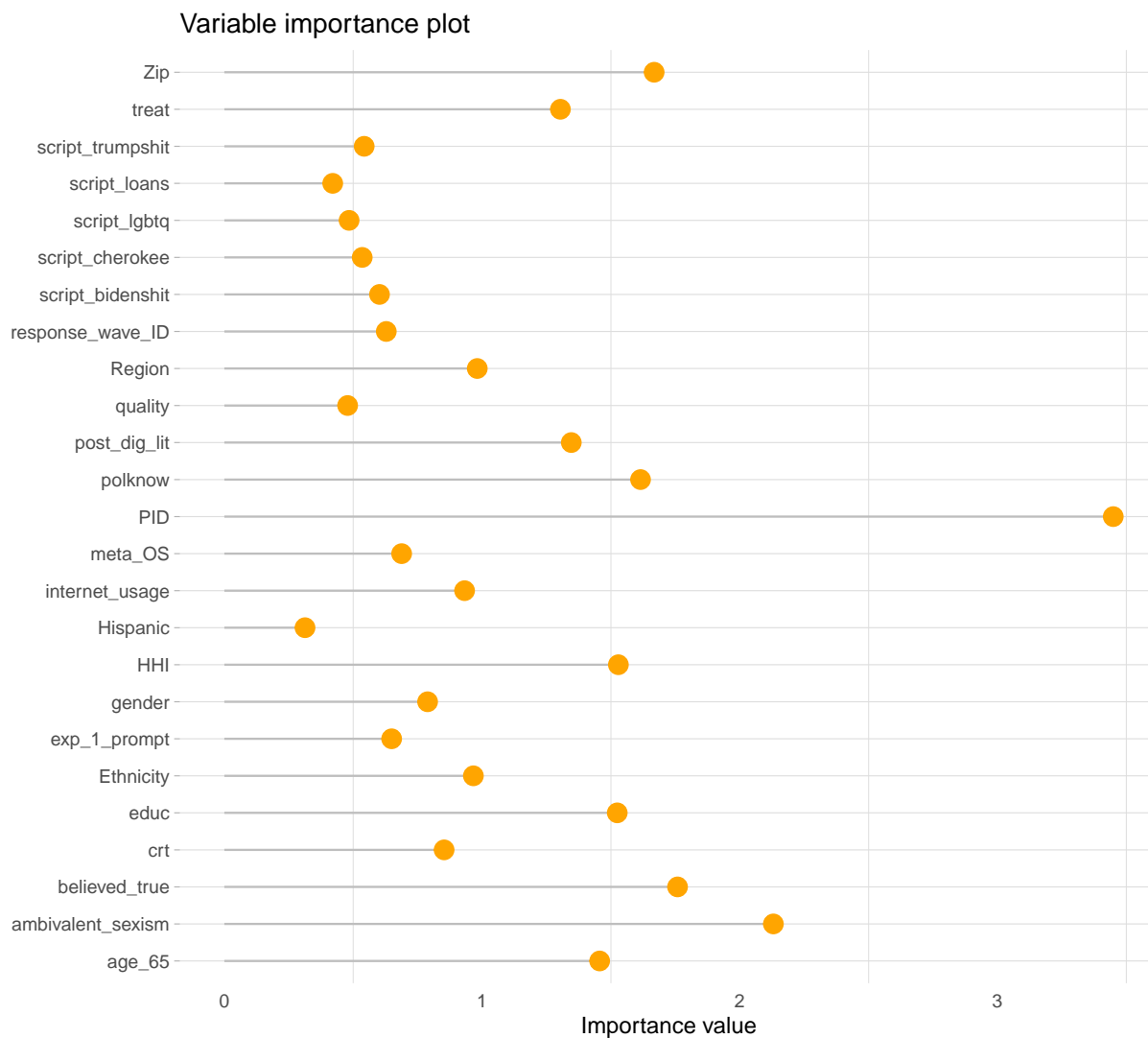
## The following object is masked from 'package:ggplot2':
##
##   margin

## -----

## 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

```
##
## Call:
## lm(formula = post_favor_Warren ~ ., data = FS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -75.156 -18.255  -2.615   19.784   89.818
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.849e+01  8.255e+00   8.297  < 2e-16 ***
## Zip           7.439e-05  3.109e-05   2.393  0.01676 *
## gender.L      -7.315e+00  6.818e+00  -1.073  0.28339
## gender.Q      -4.712e+00  3.963e+00  -1.189  0.23444
## educHigh school -2.905e+00  4.301e+00  -0.675  0.49950
## educCollege   -1.266e+00  4.305e+00  -0.294  0.76867
```

```

## educPostgraduate          6.278e+00  4.404e+00  1.426  0.15408
## HHI$100k-$150k           4.664e+00  2.024e+00  2.304  0.02125 *
## HHI>$150k                 3.488e+00  2.243e+00  1.555  0.12003
## HHI$25k-$49k             -1.720e+00  1.189e+00 -1.447  0.14790
## HHI$50k-$74k             -6.516e-01  1.342e+00 -0.485  0.62745
## HHI$75k-$99k             -4.493e-01  1.313e+00 -0.342  0.73231
## HHIN/A                   -5.883e+00  3.361e+00 -1.750  0.08018 .
## EthnicityBlack           5.574e+00  2.791e+00  1.997  0.04588 *
## EthnicityOther           7.123e-02  3.160e+00  0.023  0.98202
## EthnicityWhite           1.307e+00  2.166e+00  0.603  0.54627
## HispanicNot Hispanic     6.639e-01  2.083e+00  0.319  0.74993
## RegionNortheast          7.647e+00  1.841e+00  4.154  3.34e-05 ***
## RegionSouth              1.290e+00  1.180e+00  1.093  0.27429
## RegionWest               -2.061e+00  1.741e+00 -1.183  0.23672
## response_wave_IDSV_eyxdeX0uISXzakt -1.593e+00  1.086e+00 -1.467  0.14249
## qualitypassed all quality screens -3.558e+00  1.365e+00 -2.606  0.00919 **
## meta_0Smobile            -8.608e-01  1.120e+00 -0.769  0.44201
## age_65>65                -3.940e+00  9.753e-01 -4.040  5.45e-05 ***
## PIDIndependent           -2.628e+01  1.406e+00 -18.695 < 2e-16 ***
## PIDRepublican            -3.808e+01  9.968e-01 -38.208 < 2e-16 ***
## ambivalent_sexism        -3.378e+00  5.464e-01 -6.182  6.93e-10 ***
## polknow                  1.353e+00  1.973e+00  0.686  0.49293
## treataudio               3.738e-01  1.280e+00  0.292  0.77036
## treattext                1.250e+00  1.283e+00  0.975  0.32960
## treatskit                -2.700e+00  1.552e+00 -1.740  0.08190 .
## treatad                  -1.188e+00  1.632e+00 -0.728  0.46682
## script_bidenshit         -1.819e+00  1.574e+00 -1.156  0.24781
## script_trumpshit         -1.299e+00  1.554e+00 -0.836  0.40317
## script_cherokee          6.823e-01  1.561e+00  0.437  0.66209
## script_lgbtq             2.196e+00  1.641e+00  1.338  0.18090
## script_loans              NA          NA          NA          NA
## believed_true             -2.872e+00  3.301e-01 -8.699 < 2e-16 ***
## exp_1_promptinfo          7.495e-01  8.425e-01  0.890  0.37372
## post_dig_lit              -1.190e+00  3.158e+00 -0.377  0.70633
## internet_usage            1.004e+00  6.711e-01  1.496  0.13472
## crt                       -8.351e-01  1.861e+00 -0.449  0.65365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.8 on 4090 degrees of freedom
## Multiple R-squared:  0.3974, Adjusted R-squared:  0.3915
## F-statistic: 67.44 on 40 and 4090 DF,  p-value: < 2.2e-16

```

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

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)
```

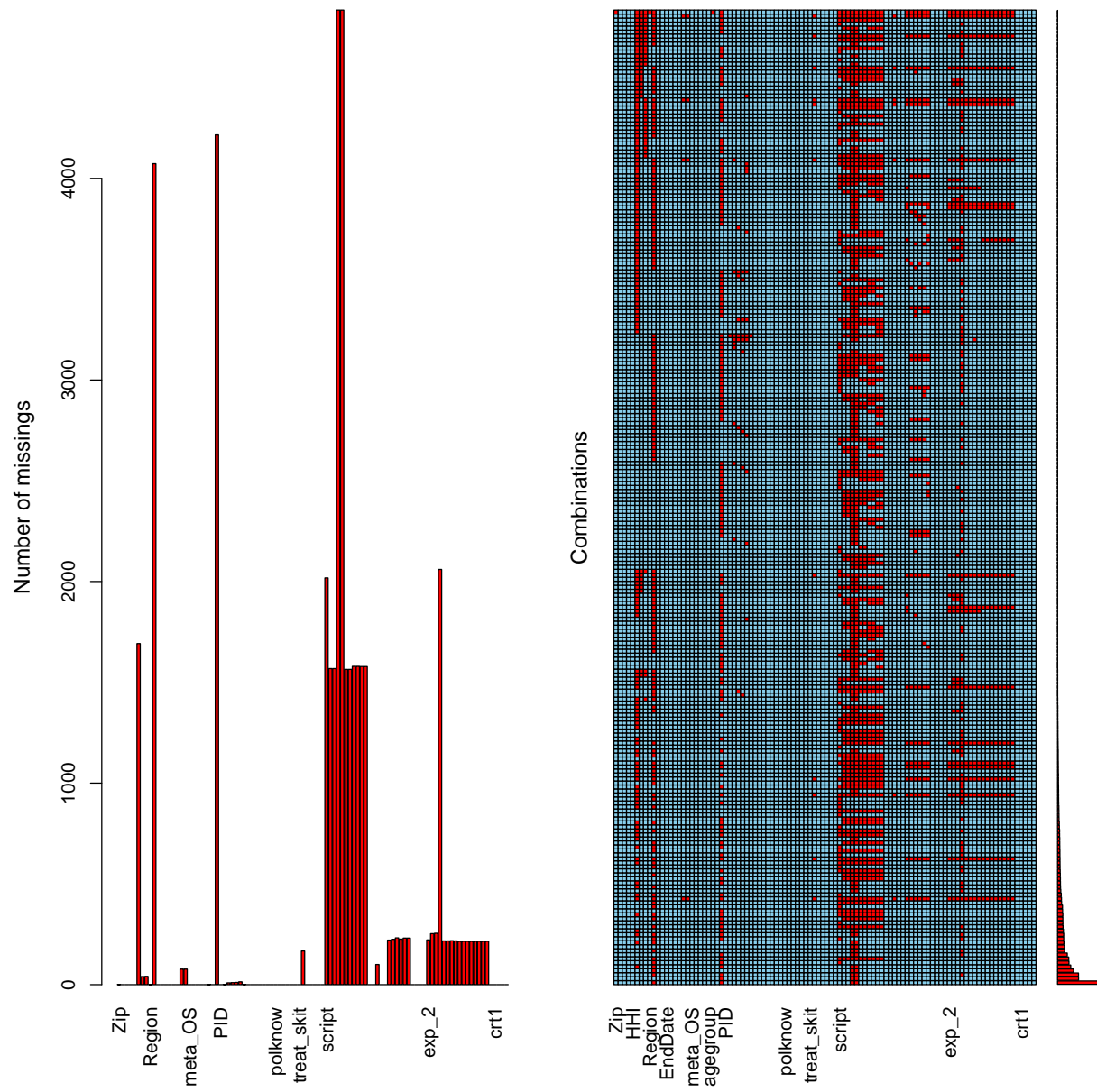


Figure 6: Missing value Visualization

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 Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.