Who Trusts the News? An Analysis of the Determinants of Trust in the News Media

Marina Wyss

11/9/2019

Abstract: *Contemporary perceptions of the media are shaped by concerns over fake news, "alternative facts," and deep fakes, and deciphering whether or not media portrayals of events are "true" is becoming more challenging than ever, resulting in far-reaching consequences for compromise and effective governance. This paper analyzes data from a 2016 Pew Research Center study on Americans' attitudes towards the national news in an effort to shed light on which socio-demographic factors contribute to a higher or lower association with trust in the news media. The investigatory methodology employed is binary logistic regression, and the primary findings are that trust in the news varies greatly by gender, racial affiliation, age, income, level of education, social media use, and particularly by political party affiliation, with the more liberal and more privileged (those with a more "socially included" status) having a higher likelihood of trusting the news.*

Introduction

Around the world, trust in the news has been on a steady decline for the past several years. The 2019 Reuters Institute Digital News Report found that,

> "Across all countries, the average level of trust in the news in general is down 2 percentage points to 42% and less than half (49%) agree that they trust the news media they themselves use." (Reuters, 2019, p. 10)

This trend is driven by anxiety over misinformation, political polarization, and ever-growing fragmentation of the media. Individuals are becoming their own information curators, selectively exposing themselves to tailored media, often without the knowledge necessary to determine whether or not a source is reputable. This is a critical development, because it fundamentally changes individuals' ability to arrive at informed opinions, which lay the groundwork for effective public and personal discourse.

In this context, the primary research question for this paper is: Who trusts the news? Are there clear differences that might provide an avenue for programs or policy changes to address the issue of lack of trust in the media? To explore these questions, this inquiry relies on data from a 2016 Pew Research Center study on Americans' perceptions of the media.

This paper is structured as follows: First, prior studies on the subject are reviewed, resulting in hypotheses about the potential associations that may be uncovered regarding socio-economic variables and trust in the news media. A description of the data set follows, as well as an overview of the methodology employed. The paper concludes with a discussion of the results, as well as the limitations to the current analysis.

Prior Studies

Many studies have found that there are clear differences in the types of people that tend to trust the news media. In general, those that have higher levels of "social inclusion" - those who are wealthier, more educated, and white - are more likely to trust the news media. In the US there is also a noticeable partisan divide, with Democrats being significantly more likely to trust the news than Republicans - a diverging trend that has grown more extreme since around 2015, with Democrats increasingly trusting the news and Republicans becoming ever more mistrustful (Reuters, 2019).

In recent years another variable has come into play: the role of social media and the abundance of news available online. Interestingly, social media use has previously been found to be associated with lower trust in the news (Reuters, 2019), perhaps because of exposure to and awareness of inaccurate or unreliable sources. This is consistent with the findings of Tsfati (2010) that exposure to non-mainstream online media was associated with media skepticism in Israel.

This association between exposure to social media and distrust in the news may help to explain the finding that younger people in the United States tend to trust the news at lower rates than older Americans, with only about one-third of those under 45 considering the news media to be trustworthy according to a 2018 study by the Media Insight Project (www.americanpressinstitute.org, 2018).

There is some evidence to suggest that people from communities that are not well-represented in the news media may also be less likely to trust the news. For example, those in rural communities often find that their local issues are not covered as comprehensively as those in urban environments (www.pewresearch.org, 2019) Similarly, a lack of racial diversity in journalism has led to bias in the kinds of stories and coverage, which may alienate communities of color (Cobb, 2018).

Against this background, three primary hypotheses are explored in this paper:

1. Democrats are more likely than Republicans to trust the news.
2. Older people are more likely than younger people to trust the news.
3. Those with a higher level of "social inclusion" (the wealthier, more highly educated, etc.) are more likely to trust the news than those from disadvantaged (or "excluded") communities.

Data

Data for this analysis come from the Pew Research Center *Information Engaged and Wary Survey*, which was conducted between September 29th and November 6th, 2016. The survey respondents included a nationally-representative sample of 3,015 adults aged 18 and over from around the country. After removing respondents without an answer on the key question of interest, *"How much do you trust [information from national news organizations] when it comes to making decisions?"*, 3,005 respondents remained in the analytical sample. This dependent variable was converted to a binary response (trust/don't trust).

The analysis began with a comprehensive list of potentially-relevant independent variables:

- Whether the respondent lives in an **urban or rural** environment.
- Whether the respondent uses the **internet** or **social media**.
- The respondent's **age, gender**, and **race**.
- Whether the respondent is a **parent** to any children under 18 years old and their **marital status**.
- Their highest level of **education**.
- Adjusted **household income**.
- Their **employment status** (working part-time, full-time, or not at all).
- **Party affiliation** (Democrat, Independent, Republican, or other).

Missing values for the independent variables were imputed using MICE (Azur et al., 2011). To check the robustness of the imputation, the regressions were performed on both the imputed dataset as well as one with only "complete cases" (using listwise deletion of any respondents with missing values on any variable), and found that the results were not significantly different.

The measure of household income was standardized to account for the number of individuals (adults and

children) living in the home, according to OECD Equivalence Scales (OECD, n.d.).[1]

Finally, to aid in the ease of interpretation, continuous variables such as age and adjusted household income were transformed into meaningful categorical groupings.

Binary logistic regression was employed to model the relationships, with results presented in terms of odds ratios. The empirical analysis consisted of two primary steps: First, quantitatively evaluating which independent variables would be the most meaningful to include using a variety of feature selection techniques, and then composing three models for comparison.

Descriptive Results & Feature Selection

As an initial investigation, a few crosstabs were created to evaluate relationships between trust in the news and likely independent variables of interest, including party affiliation, social media use, level of education, and age:

Table 1: Cross-Tabs of Trust in News - Political Preference, Social Media Use, Education, and Age

	Party Affiliation				Social Media User		Highest Level of Education			Age			
	Democrat	Independent	Republican	Other Party	No	Yes	High School	College	Advanced	Young	Middle-Aged	Old	Retired
0-dontTrustNews	0.2	0.38	0.45	0.37	0.41	0.31	0.38	0.34	0.25	0.31	0.33	0.34	0.4
1-trustNews	0.8	0.62	0.55	0.63	0.59	0.69	0.62	0.66	0.75	0.69	0.67	0.66	0.6

Note:
Column Percent

Two of these variables appear as expected: party affiliation and education: Democrats do appear to be much more likely to trust the news than Independents or Republicans (80% compared to 62% and 55%, respectively), and those with higher levels of education do in fact have higher levels of trust in the news (75% for those with advanced degrees in comparison to 64% for those with only a high school education or below).

In contrast, social media use does not seem to be associated with lower trust in the news in this data set, unlike what was found in prior studies. Indeed, 69% of social media users trust the news, compared to 59% of those who are not on social media. This may be because younger generations also appear to have higher rates of trust in news relative to older generations in this dataset, which may partially explain the incongruence with previous research that controlled for age. The cross-tab shows that 69% of those under 30 trust the news, 67% of those 31 to 50, around the same at 66% of those between 51 and 70, and only 60% of individuals 71 and older. This finding runs counter to Hypothesis 2, which presumed that younger individuals would be less likely to trust the news than older generations.

Next, a chi-squared test was performed for each feature against the target variable of "trust in the news."

[1]The OECD defines a child as a person 14 years of age and younger. This was not possible to differentiate in this dataset, as they asked about individuals under the age of 18. Therefore, a small discrepancy may be evident.

Table 2: Chi-Squared Results for Trust in News Explanatory Variables		
Variable	ChiSquaredOutput	Significance
urbanRural	0.0005	***
internetUser	0.0005	***
socialUser	0.0005	***
sex	0.0005	***
married	0.6737	
parent	0.8586	
age	0.0175	**
education	0.0005	***
adjustedIncome	0.0040	***
employmentStatus	0.0010	***
race	0.0005	***
party	0.0005	***
region	0.0955	*

Note:
* = 0.1, ** = 0.05, *** = 0.01

A statistically significant relationship was observed between each feature and the target variable, with the exception of marital and parental status. This is not hugely surprising, given that prior research has not highlighted these family status variables as relevant for this topic.

The findings from the chi-squared test were further supported by two other feature selection techniques: Step AIC (Zhang, 2016) and Recursive Feature Elimination (RFE) (scikit-yb developers, 2016) using the `caret` package in R.

Regression Results

Based on the findings from the chi-squared test, Step AIC, and RFE analyses, in conjunction with the empirical results of prior studies, two sets of independent variables and three models (one including interaction effects) were constructed:

Model 1 is the most limited model, relying on age, gender, social media use, income, and party affiliation.

In this model, similar trends appear to those evidenced in the cross-tab:

- First, trust in news declines with age, though the only statistically significant finding on this variable is for the very old "retired" age group (over 70), who have about 36% lower odds of trusting the news than the reference category of those 18 to 30 years old.
- Also as alluded to in the cross-tab, social media users have 34% higher odds of trusting the news than non-social media users.
- Independents have about 60% lower odds than Democrats of trusting the news, and Republicans have 69% lower odds. Those who indicate their party preference as "other" have similar findings to Independents.
- In terms of the new variables, men have approximately 31% lower odds of trusting the news than women, and those with "middle" levels of adjusted household income (40,000 - 75,000 USD per year) have 34% higher odds of trusting the news than their less affluent counterparts. Interestingly, trust in news declines with incomes above this range relative to the middle income bracket (with high-earners having only 26% higher odds than low earners), but this finding is not statistically significant.

Model 2 includes additional variables for race, internet use, education, and employment status.

Table 3: Odds-ratio Models of Trust in News

	Dependent variable:		
	trustNews		
	(1)	(2)	(3)
age2-middle	0.859 (0.669, 1.099)	0.902 (0.695, 1.168)	
age3-old	0.836 (0.652, 1.071)	1.011 (0.775, 1.316)	
age4-retired	0.644*** (0.479, 0.866)	0.868 (0.621, 1.213)	
sex1-male	0.691*** (0.588, 0.811)	0.682*** (0.578, 0.805)	0.676*** (0.573, 0.797)
race2-hispanic		1.426* (1.000, 2.061)	1.417* (0.992, 2.052)
race3-black		0.956 (0.726, 1.266)	0.955 (0.725, 1.266)
race4-asian		2.041*** (1.216, 3.602)	2.117*** (1.256, 3.753)
race5-other		0.585* (0.316, 1.081)	0.598 (0.322, 1.110)
socialUser1-yes	1.340*** (1.119, 1.604)	1.245** (1.034, 1.498)	
internetUser1-yes		1.317** (1.012, 1.712)	1.338** (1.027, 1.740)
adjustedIncome2-middle	1.336*** (1.133, 1.578)	1.174* (0.982, 1.404)	1.177* (0.984, 1.408)
adjustedIncome3-high	1.261 (0.804, 2.020)	1.071 (0.674, 1.736)	1.053 (0.662, 1.708)
education2-college		1.035 (0.855, 1.253)	1.040 (0.858, 1.259)
education3-advanced		1.376** (1.059, 1.792)	1.378** (1.060, 1.796)
employmentStatus2-partTime		1.413** (1.045, 1.920)	1.439** (1.063, 1.959)
employmentStatus3-fullTime		1.125 (0.924, 1.371)	1.125 (0.923, 1.370)
age1-young:socialUser0-no			0.647 (0.348, 1.200)
age2-middle:socialUser0-no			0.658* (0.415, 1.041)
age3-old:socialUser0-no			1.210 (0.834, 1.754)
age4-retired:socialUser0-no			0.948 (0.636, 1.411)
age1-young:socialUser1-yes			1.350 (0.905, 2.013)
age2-middle:socialUser1-yes			1.221 (0.846, 1.756)
age3-old:socialUser1-yes			1.159 (0.822, 1.628)
age4-retired:socialUser1-yes			
party2-independent	0.405*** (0.329, 0.496)	0.412*** (0.333, 0.508)	0.413*** (0.334, 0.509)
party3-republican	0.308*** (0.249, 0.382)	0.333*** (0.266, 0.416)	0.332*** (0.265, 0.415)
party4-other	0.422*** (0.309, 0.578)	0.446*** (0.325, 0.615)	0.452*** (0.328, 0.623)
urbanRural1-urban		1.517*** (1.287, 1.790)	1.515*** (1.284, 1.788)
Constant	4.146*** (3.060, 5.647)	2.152*** (1.435, 3.238)	2.106*** (1.435, 3.109)
Observations	3,005	3,005	3,005
Log Likelihood	−1,817.662	−1,785.619	−1,779.600
Akaike Inf. Crit.	3,657.325	3,613.238	3,607.199

Note: *p<0.1; **p<0.05; ***p<0.01

The direction of effects remains the same as in Model 1 for all common variables, though the magnitude is diminished for several:

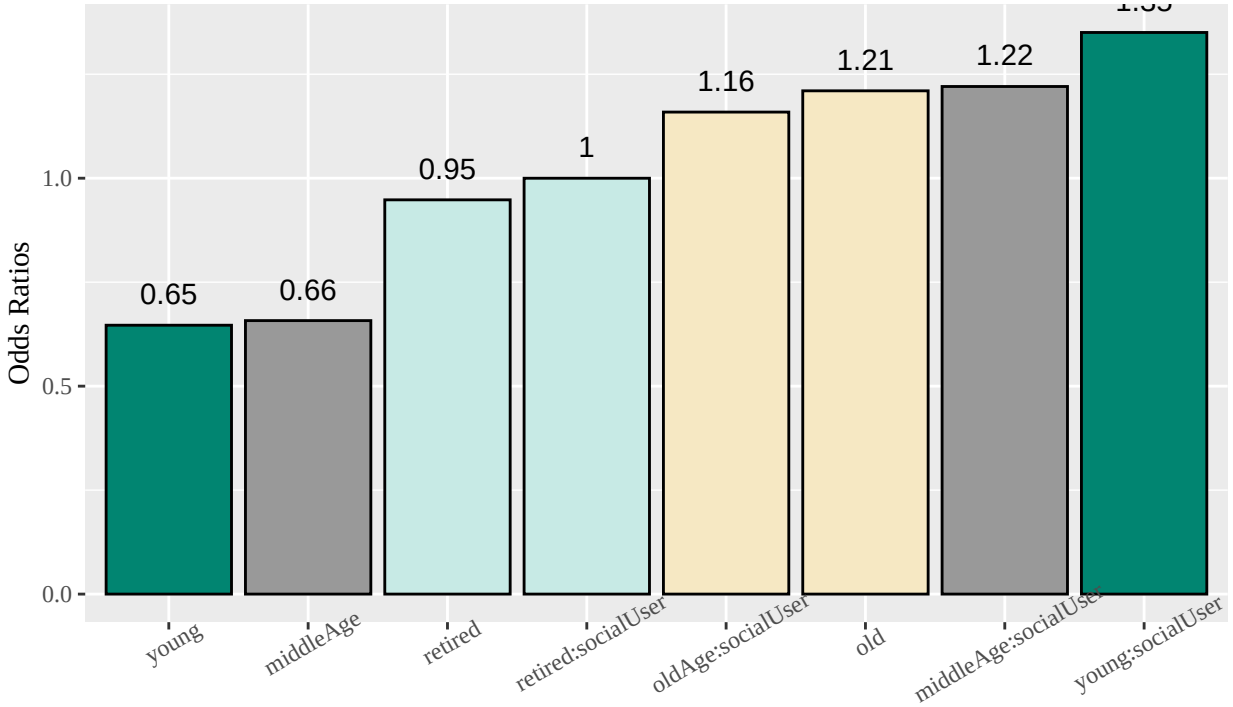
- Increasing age is no longer a statistically significant finding once controlling for other variables, though the pattern of decreasing trust with age generally holds.
- The introduction of race as a control variable illustrates notable differences, with Hispanic Americans having 43% higher odds of trusting the news than white Americans, and Asian Americans having 104% higher odds. Findings for black Americans show no statistically significant difference from white Americans, while those from other minority races are much less likely to trust the news than white Americans (at 41% lower odds).
- Internet and social media users both have higher odds of trusting the news than their counterparts who are not online.
- Having an advanced degree is associated with higher odds of trusting the news as well - about 38% higher.
- Interestingly, those who are employed part time have higher odds of trusting the news than both those who are unemployed and employed full time.
- Effects of party identification appear essentially the same as in the more constricted model.
- Urban Americans are much more likely than rural Americans to trust the news, with 52% higher odds.

Model 3 includes all of the variables from *Model 2*, with the addition of an interaction effect between age and social media use. This was included to try to explore the counter-intuitive finding that younger people and social media users are in fact more likely to trust the news than their older, non-social media-user counterparts.

This inclusion did not have a major effect on any other coefficients, but did make a meaningful change to the analysis of the effects of age and social media use.

- Examining the plot below, we can see that very little difference exists between retired individuals (70+) on and off social media and their trust in the news - though retired people on social media have about 5% higher odds of trusting the news than those who are not.
- The pattern is similar for the older generation (between 51 and 70 years of age), though the pattern is reversed - those who are on social media have about 4% lower odds.
- The really interesting pattern emerges for the middle-aged and young individuals. Middle-aged people (between 30 and 50 years old) who are on social media have 85% higher odds of trusting the news than their counterparts who are not, and young social media users (between 18 and 30 years old) have 108% higher odds of trusting the news than young people who are not on social media.
- One can also look at the effect of age among non-social users, which shows that trust in news actually is much lower for younger and middle-aged people than for older and retired people. Indeed, using the retired who are not social media users as the reference category, young people who are not on social media have 32% lower odds of trusting the news

Fig. 1: Odds ratios − Interaction of Age and Social Media Use on Trust in the News



Overall, *Model 3* including the interaction effect appears to make the most meaningful contribution towards understanding the relationship between demographics and trust in the news.

Conclusions

Results from the analysis indicate the following about the three hypotheses:

- Hypothesis 1: Democrats are indeed notably more likely than Republicans or Independents to trust the news.
- Hypothesis 2: While young people as a whole are actually *more* likely to trust the news than their older counterparts, this effect may be impacted by social media use. Young and middle-aged social media users tend to be much more trusting in the news than older social media user and younger non-social media users. However, among non-social media users, the young are actually less likely to trust the news than the older generations.
- Hypothesis 3: Those with a higher level of "social inclusion" (the wealthier, more highly educated, etc.) do appear to generally have higher odds of trusting the news. This is evidenced by higher odds for white and Asian Americans, those with advanced degrees, and those with higher rather than lower earnings. An interesting exception may be the lower rates of trust in the media by men than women, and higher rates for the part-time employed rather than full-time, however.

Finally, it is important to note some limitations to this analysis. First, the wording of the question for the dependent variable is ambiguous: The survey references trust in the "National News" - it is not clarified to the respondent whether or not this refers to state-run media, or simply major news outlets. It also does not clarify where they are accessing this news, given that much news is available on social media. Also, even the wording "trust" could be better clarified, as studies have shown that different people and even socio-economic groups define the term in distinct ways: For example, women place greater emphasis on the way news is delivered,

not just its content, when deciding whether it is trustworthy or not (www.americanpressinstitute.org, 2016). It is also important to note that no causal link can be established from this research, and that it is merely a descriptive look into the kinds of factors that tend to be associated with higher or lower rates of trust in the news.

References

American Press Institute. (2016). *How trust differs across generation, socioeconomics, race and ethnicity, and gender*. Retrieved from https://www.americanpressinstitute.org/publications/reports/survey-research/news-trust-across-age-class-race-gender/

American Press Institute. (2018). *How younger and older Americans understand and interact with news*. Retrieved from https://www.americanpressinstitute.org/publications/reports/survey-research/ages-understand-news/

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). *Multiple imputation by chained equations: what is it and how does it work?*. International journal of methods in psychiatric research, 20(1), 40-49.

Cobb, Jelani. (2018). *When newsrooms are dominated by white people, they miss crucial facts*. The Guardian. Retrieved from https://www.theguardian.com/world/commentisfree/2018/nov/05/newsroom-diversity-media-race-journalism

Greico, Elizabeth. (2019). *For many rural residents in U.S., local news media mostly don't cover the area where they live*. Pew Research Center. Retrieved from https://www.pewresearch.org/fact-tank/2019/04/12/for-many-rural-residents-in-u-s-local-news-media-mostly-dont-cover-the-area-where-they-live/

OECD. (n.d.) *What are Equivalence Scales?*. Retrieved from http://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf

Recursive Feature Elimination. (2019). Retrieved from https://www.scikit-yb.org/en/latest/api/model_selection/rfecv.html

Reuters. (2019). *Reuters Institute Digital News Report*.

Tsfati, Y. (2010). *Online news exposure and trust in the mainstream media: Exploring possible associations*. American Behavioral Scientist, 54(1), 22-42.

Zhang, Z. (2016). *Variable selection with stepwise and best subset approaches*. Annals of translational medicine, 4(7).

Appendix

```
library(tidyverse)
library(ggthemes)
library(mice)
library(caret)
library(stargazer)
library(kableExtra)
library(lmtest)
library(MASS)

# loading data and renaming
data <- read.csv("Data/September 29 - November 6, 2016 - Information engaged wary - CSV.csv")

data <- data %>%
  dplyr::select(q6a, live1, eminuse, snsint2, sex, age, marital,
         par, educ2, emplnw, race3m1, inc, party,
         cregion, hh1, hh3) %>%
  rename(
    trustNews = q6a,
    urbanRural = live1,
    internetUser = eminuse,
    socialUser = snsint2,
    married = marital,
    parent = par,
    education = educ2,
    employmentStatus = emplnw,
    race = race3m1,
    income = inc,
    region = cregion,
    liveInHouse = hh1,
    adultsInHouse = hh3
  )

# dropping NAs on dependent variable
data <- data %>%
  filter(trustNews != 8 & trustNews != 9)

# data prep
data <- data %>%
  mutate(education = ifelse(education == 8, 7, education),
         income = ifelse(income == 8, 7, income),
         employmentStatus = ifelse(employmentStatus == 8, 7, employmentStatus))

data <- data %>%
  mutate_each(funs(replace(., . == 8, NA))) %>%
  mutate_each(funs(replace(., . == 9, NA))) %>%
  mutate_each(funs(replace(., . == 98, NA))) %>%
  mutate_each(funs(replace(., . == 99, NA)))

# adjusting household income for number of people
data <- data %>%
  mutate(adjustedIncome = income / (1 +
                                      (0.5 *(adultsInHouse - 1)) +
```

```
                                    (0.3 * (liveInHouse - adultsInHouse)) )
  ) %>%
  mutate(adjustedIncome = ifelse(liveInHouse > 1, adjustedIncome, income))

# make categorical and impute missings
dataBinary <- data %>%
  mutate(trustNews = factor(case_when(trustNews %in% c(1, 2) ~ "1-trustNews",
                                      trustNews %in% c(3, 4) ~ "0-dontTrustNews"),
                            levels = c("0-dontTrustNews", "1-trustNews")),
         urbanRural = factor(case_when(urbanRural %in% c(1, 2) ~ "1-urban",
                                       urbanRural %in% c(3, 4) ~ "0-rural"),
                             levels = c("0-rural", "1-urban")),
         internetUser = factor(case_when(internetUser == 1 ~ "1-yes",
                                         internetUser == 2 ~ "0-no"),
                               levels = c("0-no", "1-yes")),
         socialUser = factor(case_when(socialUser == 1 ~ "1-yes",
                                       socialUser == 2 ~ "0-no"),
                             levels = c("0-no", "1-yes")),
         sex = factor(case_when(sex == 1 ~ "1-male",
                                sex == 2 ~ "0-female"),
                      levels = c("0-female", "1-male")),
         married = factor(case_when(married == 1 ~ "1-married",
                                    married %in% c(2, 3, 4, 5, 6) ~ "0-notMarried"),
                             levels = c("0-notMarried", "1-married")),
         parent = factor(case_when(parent == 1 ~ "1-parent",
                                   parent == 2 ~ "0-notParent"),
                         levels = c("0-notParent", "1-parent"))) %>%
  dplyr::select(trustNews, urbanRural, internetUser, socialUser, sex, married, parent)

dataOrdinal <- data %>%
  mutate(age = factor(case_when(age < 30 ~ "1-young",
                                age >= 30 & age < 50 ~ "2-middle",
                                age >= 50 & age < 70 ~ "3-old",
                                age >= 70 & age < 98 ~ "4-retired"),
                      levels = c("1-young", "2-middle", "3-old", "4-retired")),
         education = factor(case_when(education %in% c(1, 2, 3) ~ "1-highSchool",
                                      education %in% c(4, 5, 6) ~ "2-college",
                                      education == 7 ~ "3-advanced"),
                            levels = c("1-highSchool", "2-college", "3-advanced")),
         adjustedIncome = factor(case_when(adjustedIncome <= 3 ~ "1-low",
                                   adjustedIncome > 3 & adjustedIncome <= 6 ~ "2-middle",
                                   adjustedIncome > 6 & adjustedIncome <= 9 ~ "3-high"),
                         levels = c("1-low", "2-middle", "3-high"))) %>%
  dplyr::select(age, education, adjustedIncome)

dataCategorical <- data %>%
  mutate(employmentStatus = factor(case_when(employmentStatus %in% c(1, 5) ~ "3-fullTime",
                                             employmentStatus == 2 ~ "2-partTime",
                                             employmentStatus %in% c(3, 4, 6, 7) ~ "1-notEmployed"),
                                   levels = c("1-notEmployed", "2-partTime", "3-fullTime")),
         race = factor(case_when(race == 1 ~ "1-white",
                                 race == 2 ~ "3-black",
                                 race == 3 ~ "4-asian",
```

```
                                race %in% c(5, 6) ~ "5-other",
                                race == 7 ~ "2-hispanic"),
                         levels = c("1-white", "2-hispanic", "3-black", "4-asian", "5-other")),
         party = factor(case_when(party == 1 ~ "3-republican",
                                  party == 2 ~ "1-democrat",
                                  party == 3 ~ "2-independent",
                                  TRUE ~ "4-other"),
                        levels = c("1-democrat", "2-independent", "3-republican", "4-other")),
         region = factor(case_when(region == 1 ~ "1-northEast",
                                   region == 2 ~ "2-midWest",
                                   region == 3 ~ "3-south",
                                   region == 4 ~ "4-west"))) %>%
  dplyr::select(employmentStatus, race, party, region)

## imputation using MICE by variable type
set.seed(123)
require(mice)
imputedBinary <- mice(dataBinary, method = "logreg")
dataBinary <- complete(imputedBinary)

imputedOrdinal <- mice(dataOrdinal, method = "polr")
dataOrdinal <- complete(imputedOrdinal)

imputedCategorical <- mice(dataCategorical, method = "polyreg")
dataCategorical <- complete(imputedCategorical)

## rejoining data and removing intermediate steps
data <- cbind(dataBinary, dataOrdinal, dataCategorical)
rm(dataBinary, dataOrdinal, dataCategorical,
   imputedBinary, imputedOrdinal, imputedCategorical)

# cross tables
partyCross <- table(data$trustNews, data$party) %>%
  prop.table(2) %>%
  round(2)
partyCross <- data.frame(partyCross)
partyCross <- reshape(partyCross, idvar = "Var1", timevar = "Var2", direction = "wide")

socialCross <- table(data$trustNews, data$socialUser) %>%
  prop.table(2) %>%
  round(2)
socialCross <- data.frame(socialCross)
socialCross <- reshape(socialCross, idvar = "Var1", timevar = "Var2", direction = "wide")

educationCross <- table(data$trustNews, data$education) %>%
  prop.table(2) %>%
  round(2)
educationCross <- data.frame(educationCross)
educationCross <- reshape(educationCross, idvar = "Var1", timevar = "Var2", direction = "wide")

ageCross <- table(data$trustNews, data$age) %>%
  prop.table(2) %>%
  round(2)
```

```r
ageCross <- data.frame(ageCross)
ageCross <- reshape(ageCross, idvar = "Var1", timevar = "Var2", direction = "wide")

allCross <- left_join(partyCross, socialCross, by = "Var1")
allCross <- left_join(allCross, educationCross, by = "Var1")
allCross <- left_join(allCross, ageCross, by = "Var1")


names(allCross) <- c("", "Democrat", "Independent", "Republican", "Other Party",
                     "No", "Yes",
                     "High School", "College", "Advanced",
                     "Young", "Middle-Aged", "Old", "Retired")

allCrossTable <- allCross %>%
  kable() %>%
  kable_styling(latex_options = "scale_down") %>%
  column_spec(1, bold = T, border_right = T) %>%
  column_spec(5, border_right = T) %>%
  column_spec(7, border_right = T) %>%
  column_spec(10, border_right = T) %>%
  add_header_above(c("", "Party Affiliation" = 4, "Social Media User" = 2,
                     "Highest Level of Education" = 3, "Age" = 4)) %>%
  add_header_above(c("", "Table 1: Cross-Tabs of Trust in News - Political Preference, Social Media Use
  footnote("Column Percent")

# feature selection
## chi squared
x2 <- data.frame(lapply(data[,-1],
                        function(x) chisq.test(table(x, data$trustNews),
                                               simulate.p.value = TRUE)$p.value))
x2Table <- x2 %>%
  type_convert %>%
  mutate_if(is.numeric, round, digits = 4) %>%
  t %>%
  as.data.frame() %>%
  mutate(Variable = names(x2)) %>%
  mutate(Significance = case_when(V1 <= 0.1 & V1 > 0.05 ~ "*",
                                  V1 <= 0.05 & V1 > 0.01 ~ "**",
                                  V1 <= 0.01 ~ "***",
                                  TRUE ~ " ")) %>%
  rename(ChiSquaredOutput = V1) %>%
  dplyr::select(Variable, ChiSquaredOutput, Significance) %>%
  kable() %>%
  kable_styling() %>%
  footnote(general = "* = 0.1, ** = 0.05, *** = 0.01") %>%
  add_header_above(c("Table 2: Chi-Squared Results for Trust in News Explanatory Variables" = 3))

## rfe
control <- rfeControl(functions = rfFuncs, method = "cv", number = 10)
results <- rfe(data[ ,2:14], data[ ,1], sizes = c(1:14), rfeControl = control)
print(results)
predictors(results)
plot(results, type = c("g", "o"))
```

```r
## stepwise
stepModelFull <- glm(trustNews ~.,
              family = binomial,
              data = data)

stepModelFull <- stepModelFull %>% stepAIC(trace = FALSE)

# models
model1 <- glm(trustNews ~
                age + sex + socialUser + adjustedIncome + party,
              family = binomial,
              data = data)

model2 <- glm(trustNews ~
                age + sex + race + socialUser + internetUser + party +
                urbanRural + adjustedIncome + education + employmentStatus,
              family = binomial,
              data = data)

model3 <- glm(trustNews ~ age:socialUser + sex + race + internetUser +
                party + urbanRural + adjustedIncome + education + employmentStatus,
              family = binomial,
              data = data)

# stargazer
model1or <- exp(coef(model1))
model1conf <- exp(confint(model1))
model1p <- list(summary(model1)$coefficients[ ,4])

model2or <- exp(coef(model2))
model2conf <- exp(confint(model2))
model2p <- list(summary(model2)$coefficients[ ,4])

model3or <- exp(coef(model3))
model3conf <- exp(confint(model3))
model3p <- list(summary(model3)$coefficients[ ,4])

stargazer(model1, model2, model3,
          coef = list(model1or, model2or, model3or),
          ci = T,
          ci.custom = list(model1conf, model2conf, model3conf),
          p = c(model1p, model2p, model3p),
          header = FALSE,
          title = "Table 3: Odds-ratio Models of Trust in News",
          type = "text")

# interaction effects plots
model3coef <- as.data.frame(summary(model3)$coefficients)
plotData <- rownames_to_column(model3coef)[18:24, 1:2]

plotOR <- as.data.frame(model3or)
plotOR <- rownames_to_column(plotOR)[18:24, 1:2]
```

```r
plotCI <- as.data.frame(model3conf)
plotCI <- rownames_to_column(plotCI)[18:24, 1:3]

plotData <- left_join(plotData, plotOR, by = "rowname")
plotData <- left_join(plotData, plotCI, by = "rowname")
plotData[nrow(plotData) + 1, ] <- 1

plotData$rowname <- c("young", "middleAge", "old", "retired",
                      "young:socialUser", "middleAge:socialUser", "oldAge:socialUser",
                      "retired:socialUser")

palette <- c("#999999", "#999999", "#f6e8c3", "#f6e8c3", "#c7eae5",
             "#c7eae5", "#018571", "#018571")

plotInteraction <- ggplot(plotData, aes(x = reorder(rowname, model3or), y = model3or)) +
  geom_bar(stat = "identity", aes(fill = rowname), color = "black") +
  geom_text(aes(label = round(model3or, 2)), vjust = -0.5) +
  theme(legend.position = "none",
        text = element_text(family = "Times")) +
  scale_fill_manual(values = palette) +
  labs(title = "Fig. 1: Odds ratios - Interaction of Age and Social Media Use on Trust in the News",
       x = "Age and Social Media Use",
       y = "Odds Ratios")

plotInteraction
```