# STP598 Final Project: Predicting 2016 Election Vote Choices

*Xin Liu (alphabetical)*
*Fan Hong*
*Dantong Huang*
*Hao Wang*
*Arizona State University*

*April 29, 2017*

# 1   Variable Selection

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
mydata <- dplyr::select(x,
                        CC16_401, #whether voted or not
                        CC16_410a, #vote for
                        gender, #gender
                        educ, #education
                        race, #race
                        marstat, #marriage
                        inputstate, #state
                        religpew, #religion
                        industryclass, #industry class
                        ownhome, #home owner
                        immstat, #immigrant
                        faminc, #family income
                        employ, #employment status
                        pid7, #party id
                        CC16_302, #national econ better: past
                        CC16_303, #homehold income better: past
                        CC16_304, # national econ better: next
                        CC16_307, # feel safe about police
                        CC16_321a,
                        CC16_321b,
                        CC16_320a,
                        CC16_320b,
                        CC16_331_1,
                        CC16_331_2,
```

```
                          CC16_331_3,
                          CC16_331_7,
                          CC16_332a,
                          CC16_332b,
                          CC16_332c,
                          CC16_332f,
                          CC16_334c,
                          CC16_334d,
                          CC16_335,
                          CC16_337_1,
                          CC16_337_2,
                          CC16_337_3)

mydata$race <- factor(mydata$race, ordered = FALSE)
mydata$inputstate <- factor(mydata$inputstate, ordered = FALSE)
mydata$religpew <- factor(mydata$religpew, ordered = FALSE)
mydata$faminc[mydata$faminc == "Prefer not to say"] <- NA



mydata <- na.omit(mydata)

mydata2 <-  dplyr::select(x,
                          CC16_401, #whether voted or not
                          CC16_410a, #vote for
                          gender, #gender
                          educ, #education
                          race, #race
                          marstat, #marriage
                          inputstate, #state
                          religpew, #religion
                          industryclass, #industry class
                          ownhome, #home owner
                          immstat, #immigrant
                          faminc, #family income
                          employ) #employment status)
mydata2 <- na.omit(mydata2)

dim(mydata)
```

```
## [1] 34224     36
```

```
dim(mydata2)
```

```
## [1] 39864     13
```