Vote for Us!: Examining Demographics of Individuals with Weak Party Ties

QMSS5069 Team 1: Svyatoslav Andriyishen, Shilpa Sure, Meredith Meeks, Kai Huo, Hannah Hao

April 26, 2017

if (!require("dplyr")) {  
 install.packages("dplyr", dependencies = TRUE)  
 library(dplyr) }  
if (!require("tidyr")) {  
 install.packages("tidyr", dependencies = TRUE)  
 library(tidyr) }  
if (!require("readr")) {  
 install.packages("readr", dependencies = TRUE)  
 library(readr) }  
if (!require("readxl")) {  
 install.packages("readxl", dependencies = TRUE)  
 library(readxl) }  
if (!require("haven")) {  
 install.packages("haven", dependencies = TRUE)  
 library(haven) }  
if (!require("ggplot2")) {  
 install.packages("ggplot2", dependencies = TRUE)  
 library(ggplot2) }

## Warning: package 'ggplot2' was built under R version 3.3.2

if (!require("ggthemes")) {  
 install.packages("ggthemes", dependencies = TRUE)  
 library(ggthemes) }

## Warning: package 'ggthemes' was built under R version 3.3.2

### Clean and Subset Data

# pew16 <- read\_sav("Data/March16 public.sav") # used for prelim analysis  
Pew\_Yearly\_Trends <- read\_excel("Data/Pew\_Yearly\_Trends(hand\_coded).xlsx",   
 col\_types = c("numeric", "numeric", "numeric",   
 "numeric", "numeric", "numeric"),  
 na = "NA")   
# hand coded into Excel from data in PDF attached to Pew March 2016 data  
  
# Create GSS subset to fit GitHub file size requirements  
# GSS 1972-2016  
# gss7216 <- read\_sav("GSS7216.sav")  
# gss\_sub <- select(gss7216, year, partyid, sex, race, relig, born,   
# age, degree, reg16, prestg80, marital, adults, childs,  
# rincome, rincom06)  
# save(gss\_sub, file = "gss\_sub.rda")  
  
load("Data/gss\_sub.rda")  
  
# attach party names to partyid factor variable for ease of reference  
partynames <- c("Strong Democrat", "Not Strong Democrat", "Independent near Democrat",   
 "Independent", "Independent near Republican", "Not Strong Republican",   
 "Strong Republican", "Other")  
gss\_sub$partyid <- as.factor(gss\_sub$partyid)  
levels(gss\_sub$partyid) <- partynames  
  
gss <- read.csv("Data/GSS.2006.csv") # team\_1\_project/Data  
  
gss$partyid <- as.factor(gss$partyid)  
levels(gss$partyid) <- partynames  
  
# other recodings  
gss$child1 <- ifelse(gss$childs==1, 1, 0)  
gss$child2 <- ifelse(gss$childs==2, 1, 0)  
gss$child3more <- ifelse(gss$childs>=3, 1, 0)  
gss$region <- as.factor(gss$region)  
gss <- within(gss, region <- relevel(region, ref = 3))  
gss$relig <- as.factor(gss$relig)  
gss <- within(gss, relig <- relevel(relig, ref = 4))  
gss$sex <- as.factor(gss$sex)  
gss$race <- as.factor(gss$race)

### Yearly Trends (source: Pew Research Center)

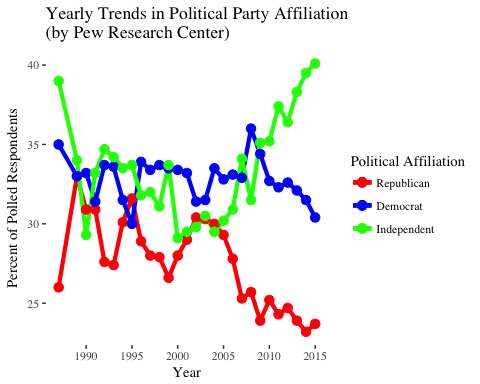
According to Pew:

"The recent changes in partisan identification serve as a reminder that affiliation with a party is an attitude, one which can and does change. Previous research has shown that few people switch immediately from Republican to Democratic identification or vice versa. **Most of the movement is from independents who assume a party label or from partisans who no longer identify with their former party.**

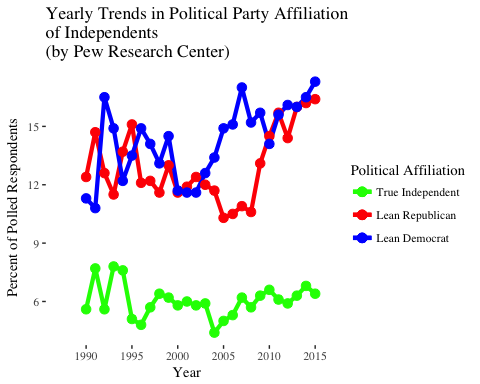
In fact, most of the survey's independent and non-partisan voters report changing their party affiliation - 72% say they have thought of themselves as a Republican, a Democrat or both at some point in their life and 54% considered themselves a partisan in the past five years. About one-in-five independent voters (22%) say they considered themselves a Republican (but not a Democrat) during this period, while 23% of independents considered themselves a Democrat (but not a Republican)." Source: <http://www.people-press.org/2010/09/23/section-3-trends-in-party-affiliation/>

Therefore, determining the population distributions of Independents with particular party leanings might help canvassers rally more votes in elections through concentrated efforts directed at individuals that will need the least amount of convincing to vote for the canvassers party, but might not vote otherwise.

ggplot(Pew\_Yearly\_Trends, aes(x = Year)) +  
 geom\_line(aes(y = Republican, col = "1"), lwd = 1.5) +  
 geom\_line(aes(y = Democrat, col = "2"), lwd = 1.5) +  
 geom\_line(aes(y = Independent, col = "3"), lwd = 1.5) +  
 geom\_point(aes(y = Republican, col = "1"), size = 3) +  
 geom\_point(aes(y = Democrat, col = "2"), size = 3) +  
 geom\_point(aes(y = Independent, col = "3"), size = 3) +  
 scale\_y\_continuous("Percent of Polled Respondents") +  
 scale\_colour\_manual("Political Affiliation", values = c("red", "blue", "green"),  
 labels = c("Republican", "Democrat", "Independent")) +  
 ggtitle("Yearly Trends in Political Party Affiliation\n(by Pew Research Center)") +  
 theme(legend.position = "right") +  
 ggthemes::theme\_tufte()



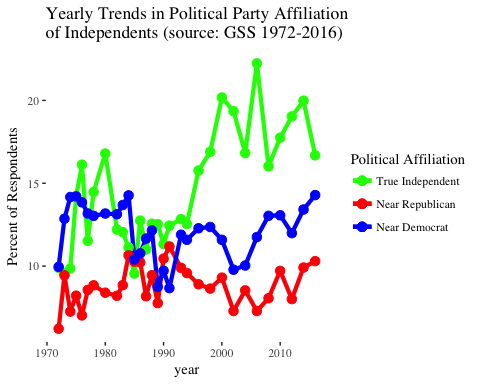
ggplot(Pew\_Yearly\_Trends, aes(x = Year)) +  
 geom\_line(aes(y = Independent - (`Lean Republican` + `Lean Democrat`), col = "1"), lwd = 1.5) +  
 geom\_line(aes(y = `Lean Republican`, col = "2"), lwd = 1.5) +  
 geom\_line(aes(y = `Lean Democrat`, col = "3"), lwd = 1.5) +  
 geom\_point(aes(y = Independent - (`Lean Republican` + `Lean Democrat`), col = "1"), size = 3) +  
 geom\_point(aes(y = `Lean Republican`, col = "2"), size = 3) +  
 geom\_point(aes(y = `Lean Democrat`, col = "3"), size = 3) +  
 scale\_y\_continuous("Percent of Polled Respondents") +  
 scale\_colour\_manual("Political Affiliation", values = c("green", "red", "blue"),  
 labels = c("True Independent", "Lean Republican", "Lean Democrat")) +  
 ggtitle("Yearly Trends in Political Party Affiliation\nof Independents\n(by Pew Research Center)") +  
 theme(legend.position = "right") +  
 ggthemes::theme\_tufte()



As seen from the charts above, the percentage of people identifying themselves as Independents is on the rise, according to Pew. Therefore, the fight for votes may very well be concentrated increasingly more on swaying Independents to one political camp or another, assuming that a third party will not be created.

### Yearly Trends (source: General Social Survey 1972-2016)

yrsum <- gss\_sub %>% group\_by(year) %>%  
 summarise(partyid\_sum = sum(!is.na(partyid)),  
 Indy = sum(partyid == "Independent", na.rm = TRUE),  
 Indy\_rep = sum(partyid == "Independent near Republican", na.rm = TRUE),  
 Indy\_dem = sum(partyid == "Independent near Democrat", na.rm = TRUE),  
 Rep = sum(partyid == "Not Strong Republican" | partyid == "Strong Republican", na.rm = TRUE),  
 Dem = sum(partyid == "Not Strong Democrat" | partyid == "Strong Democrat", na.rm = TRUE))  
  
ggplot(yrsum, aes(x = year)) +  
 geom\_line(aes(y = 100\*Indy/partyid\_sum, col = "1"), lwd = 1.5) +  
 geom\_line(aes(y = 100\*Indy\_rep/partyid\_sum, col = "2"), lwd = 1.5) +  
 geom\_line(aes(y = 100\*Indy\_dem/partyid\_sum, col = "3"), lwd = 1.5) +  
 geom\_point(aes(y = 100\*Indy/partyid\_sum, col = "1"), size = 3) +  
 geom\_point(aes(y = 100\*Indy\_rep/partyid\_sum, col = "2"), size = 3) +  
 geom\_point(aes(y = 100\*Indy\_dem/partyid\_sum, col = "3"), size = 3) +  
 scale\_y\_continuous("Percent of Respondents") +  
 scale\_colour\_manual("Political Affiliation", values = c("green", "red", "blue"),  
 labels = c("True Independent", "Near Republican", "Near Democrat")) +  
 ggtitle("Yearly Trends in Political Party Affiliation\nof Independents (source: GSS 1972-2016)") +  
 theme(legend.position = "right") +  
 ggthemes::theme\_tufte()



GSS data largely conforms to the trends found by Pew. Although the trend is smaller, the percentage of Independents is on the rise, with many of them leaning towards one party or another and potentially holding the key to elections.

# average percentage of Independent near Republicans  
mean(100\*yrsum$Indy\_rep/yrsum$partyid\_sum)

## [1] 8.859647

# average percentage of Independent near Democrats  
mean(100\*yrsum$Indy\_dem/yrsum$partyid\_sum)

## [1] 12.08739

Averaged out over the 31 years of GSS, Independent near Republicans provide a potential ~9% of population that might vote Republican, and Independent near Democrats, 12% that might vote Democrat, with recent numbers being higher, and not counting true Independents that could also potentially be swayed. Since these votes are not guaranteed, indicated by the respondents' choice to identify as Independents instead of belonging to a party, securing them may be a step towards election victory.

### Data Analysis

Below are the models attempting to draw out the background characteristics of Independents to help canvassers concentrate their efforts on individuals with similar characteristics. The presentation file for this project contains a summary cheatsheet of what characteristics to concentrate on and what characteristics to avoid. Intermediary models were excluded and only final models are presented here.

We have chosen to run logit models, since they are more appropriate for classification problems of this kind, at the cost of interpretability. Direct coefficients give us a change in the log odds of the outcome given a one unit increase in that predictor variable, holding all other predictors constant. For Independents, an increase of 1 in an individual's occupational prestige score decreases their log odds of being an Independent by 0.022. Exponentiating coefficients allows us to interpret them as odds ratios. For Independents, exponentiated coefficient on child1 tells us that someone who has 1 child is 1.6 times more likely to be an Independent than someone who does not. For Independents, exponentiated coefficient on occupational prestige tells us that an increase of one point in an occupational prestige scores means you have 0.977 the odds of being an independent (a slight decrease in odds).

# Independent near Democrat, no subset  
summary(glm((partyid == "Independent near Democrat") ~ as.factor(race) + (relig == 4) + (born == 2 | reg16 == 0) + sex + age + rincom06 + sex:rincom06 + prestg80 + degree + degree:as.factor(race) + marital + adults + child1 + child2 + child3more + union, family = binomial(link = "logit"), data = gss))

##   
## Call:  
## glm(formula = (partyid == "Independent near Democrat") ~ as.factor(race) +   
## (relig == 4) + (born == 2 | reg16 == 0) + sex + age + rincom06 +   
## sex:rincom06 + prestg80 + degree + degree:as.factor(race) +   
## marital + adults + child1 + child2 + child3more + union,   
## family = binomial(link = "logit"), data = gss)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.9828 -0.5536 -0.4688 -0.3816 2.4209   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.602099 0.803054 -1.995 0.0460 \*  
## as.factor(race)2 1.127289 0.447318 2.520 0.0117 \*  
## as.factor(race)3 0.479402 0.386390 1.241 0.2147   
## relig == 4TRUE 0.201526 0.225983 0.892 0.3725   
## born == 2 | reg16 == 0TRUE 0.223339 0.247695 0.902 0.3672   
## sex2 -0.360023 0.477890 -0.753 0.4512   
## age -0.012830 0.008028 -1.598 0.1100   
## rincom06 -0.043105 0.024289 -1.775 0.0760 .  
## prestg80 0.008722 0.007791 1.119 0.2630   
## degree 0.225999 0.104621 2.160 0.0308 \*  
## marital 0.029798 0.062806 0.474 0.6352   
## adults -0.062769 0.113440 -0.553 0.5800   
## child1 -0.027015 0.290009 -0.093 0.9258   
## child2 0.283859 0.262262 1.082 0.2791   
## child3more -0.106853 0.289286 -0.369 0.7119   
## union -0.044795 0.089409 -0.501 0.6164   
## sex2:rincom06 0.020789 0.031421 0.662 0.5082   
## as.factor(race)2:degree -0.658127 0.277085 -2.375 0.0175 \*  
## as.factor(race)3:degree -0.154032 0.176703 -0.872 0.3834   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 917.14 on 1233 degrees of freedom  
## Residual deviance: 891.54 on 1215 degrees of freedom  
## (3276 observations deleted due to missingness)  
## AIC: 929.54  
##   
## Number of Fisher Scoring iterations: 5

# Independent near Democrat, other Democrats removed  
summary(glm((partyid == "Independent near Democrat") ~ as.factor(race) + (relig == 4) + (born == 2 | reg16 == 0) + sex + age + rincom06 + sex:rincom06 + prestg80 + degree + degree:as.factor(race) + marital + adults + child1 + child2 + child3more + union, family = binomial(link = "logit"), data = gss, subset = !(gss$partyid %in% c("Strong Democrat", "Not Strong Democrat"))))

##   
## Call:  
## glm(formula = (partyid == "Independent near Democrat") ~ as.factor(race) +   
## (relig == 4) + (born == 2 | reg16 == 0) + sex + age + rincom06 +   
## sex:rincom06 + prestg80 + degree + degree:as.factor(race) +   
## marital + adults + child1 + child2 + child3more + union,   
## family = binomial(link = "logit"), data = gss, subset = !(gss$partyid %in%   
## c("Strong Democrat", "Not Strong Democrat")))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3365 -0.6530 -0.5190 -0.4034 2.3655   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.531603 0.871718 -1.757 0.078919 .   
## as.factor(race)2 1.837763 0.501076 3.668 0.000245 \*\*\*  
## as.factor(race)3 0.617901 0.405226 1.525 0.127301   
## relig == 4TRUE 0.393738 0.241110 1.633 0.102464   
## born == 2 | reg16 == 0TRUE 0.197109 0.262604 0.751 0.452896   
## sex2 -0.558558 0.503706 -1.109 0.267475   
## age -0.004164 0.008473 -0.491 0.623096   
## rincom06 -0.051927 0.025628 -2.026 0.042745 \*   
## prestg80 0.013213 0.008252 1.601 0.109331   
## degree 0.246834 0.108937 2.266 0.023461 \*   
## marital 0.046046 0.068270 0.674 0.500008   
## adults -0.023053 0.120901 -0.191 0.848776   
## child1 -0.338598 0.304039 -1.114 0.265423   
## child2 0.055086 0.277180 0.199 0.842467   
## child3more -0.379709 0.303136 -1.253 0.210350   
## union -0.117015 0.097168 -1.204 0.228493   
## sex2:rincom06 0.046744 0.033088 1.413 0.157736   
## as.factor(race)2:degree -0.614646 0.297361 -2.067 0.038734 \*   
## as.factor(race)3:degree -0.050392 0.189721 -0.266 0.790537   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 801.47 on 865 degrees of freedom  
## Residual deviance: 757.98 on 847 degrees of freedom  
## (2208 observations deleted due to missingness)  
## AIC: 795.98  
##   
## Number of Fisher Scoring iterations: 4

# Independent near Republican, no subset  
summary(republican <- glm((partyid == "Independent near Republican") ~ (race == 2) + (relig == 1 | relig == 2) + (born == 2 | reg16 == 0) + sex + age + rincom06 + prestg80 + degree + (relig == 1 | relig == 2):(race == 2) + (marital == 5) + adults + child1 + child2 + child3more + union, family = binomial(link = "logit"), data = gss))

##   
## Call:  
## glm(formula = (partyid == "Independent near Republican") ~ (race ==   
## 2) + (relig == 1 | relig == 2) + (born == 2 | reg16 == 0) +   
## sex + age + rincom06 + prestg80 + degree + (relig == 1 |   
## relig == 2):(race == 2) + (marital == 5) + adults + child1 +   
## child2 + child3more + union, family = binomial(link = "logit"),   
## data = gss)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.6775 -0.4040 -0.3381 -0.2750 2.7764   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -0.457703 0.951811 -0.481  
## race == 2TRUE -0.701320 1.063626 -0.659  
## relig == 1 | relig == 2TRUE 0.048401 0.281591 0.172  
## born == 2 | reg16 == 0TRUE -0.530313 0.330029 -1.607  
## sex2 -0.602890 0.255266 -2.362  
## age -0.015796 0.010748 -1.470  
## rincom06 -0.039289 0.023449 -1.675  
## prestg80 -0.002792 0.010380 -0.269  
## degree 0.053037 0.125625 0.422  
## marital == 5TRUE 0.016806 0.331196 0.051  
## adults -0.323318 0.161808 -1.998  
## child1 0.291274 0.368346 0.791  
## child2 0.067197 0.380083 0.177  
## child3more 0.243097 0.381959 0.636  
## union -0.027120 0.123947 -0.219  
## race == 2TRUE:relig == 1 | relig == 2TRUE -0.215197 1.163586 -0.185  
## Pr(>|z|)   
## (Intercept) 0.6306   
## race == 2TRUE 0.5097   
## relig == 1 | relig == 2TRUE 0.8635   
## born == 2 | reg16 == 0TRUE 0.1081   
## sex2 0.0182 \*  
## age 0.1416   
## rincom06 0.0938 .  
## prestg80 0.7879   
## degree 0.6729   
## marital == 5TRUE 0.9595   
## adults 0.0457 \*  
## child1 0.4291   
## child2 0.8597   
## child3more 0.5245   
## union 0.8268   
## race == 2TRUE:relig == 1 | relig == 2TRUE 0.8533   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 587.10 on 1233 degrees of freedom  
## Residual deviance: 567.87 on 1218 degrees of freedom  
## (3276 observations deleted due to missingness)  
## AIC: 599.87  
##   
## Number of Fisher Scoring iterations: 6

# Independent near Republican, other Republicans removed  
summary(republican <- glm((partyid == "Independent near Republican") ~ (race == 2) + (relig == 1 | relig == 2) + (born == 2 | reg16 == 0) + sex + age + rincom06 + prestg80 + degree + (relig == 1 | relig == 2):(race == 2) + (marital == 5) + adults + child1 + child2 + child3more + union, family = binomial(link = "logit"), data = gss, subset = !(gss$partyid %in% c("Strong Republican", "Not Strong Republican"))))

##   
## Call:  
## glm(formula = (partyid == "Independent near Republican") ~ (race ==   
## 2) + (relig == 1 | relig == 2) + (born == 2 | reg16 == 0) +   
## sex + age + rincom06 + prestg80 + degree + (relig == 1 |   
## relig == 2):(race == 2) + (marital == 5) + adults + child1 +   
## child2 + child3more + union, family = binomial(link = "logit"),   
## data = gss, subset = !(gss$partyid %in% c("Strong Republican",   
## "Not Strong Republican")))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.8152 -0.4786 -0.3816 -0.2800 2.7169   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) -5.864e-01 9.617e-01 -0.610  
## race == 2TRUE -8.245e-01 1.064e+00 -0.775  
## relig == 1 | relig == 2TRUE 4.265e-01 2.882e-01 1.480  
## born == 2 | reg16 == 0TRUE -8.592e-01 3.344e-01 -2.569  
## sex2 -6.674e-01 2.593e-01 -2.574  
## age -2.141e-02 1.126e-02 -1.901  
## rincom06 -2.595e-02 2.492e-02 -1.041  
## prestg80 5.643e-05 1.047e-02 0.005  
## degree 9.642e-02 1.285e-01 0.750  
## marital == 5TRUE -8.606e-02 3.298e-01 -0.261  
## adults -3.092e-01 1.613e-01 -1.917  
## child1 4.415e-01 3.778e-01 1.169  
## child2 1.929e-01 3.906e-01 0.494  
## child3more 4.335e-01 3.913e-01 1.108  
## union 1.498e-03 1.266e-01 0.012  
## race == 2TRUE:relig == 1 | relig == 2TRUE -5.212e-01 1.167e+00 -0.447  
## Pr(>|z|)   
## (Intercept) 0.5420   
## race == 2TRUE 0.4382   
## relig == 1 | relig == 2TRUE 0.1390   
## born == 2 | reg16 == 0TRUE 0.0102 \*  
## sex2 0.0101 \*  
## age 0.0573 .  
## rincom06 0.2977   
## prestg80 0.9957   
## degree 0.4531   
## marital == 5TRUE 0.7941   
## adults 0.0553 .  
## child1 0.2425   
## child2 0.6215   
## child3more 0.2679   
## union 0.9906   
## race == 2TRUE:relig == 1 | relig == 2TRUE 0.6552   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 536.36 on 905 degrees of freedom  
## Residual deviance: 508.57 on 890 degrees of freedom  
## (2472 observations deleted due to missingness)  
## AIC: 540.57  
##   
## Number of Fisher Scoring iterations: 6

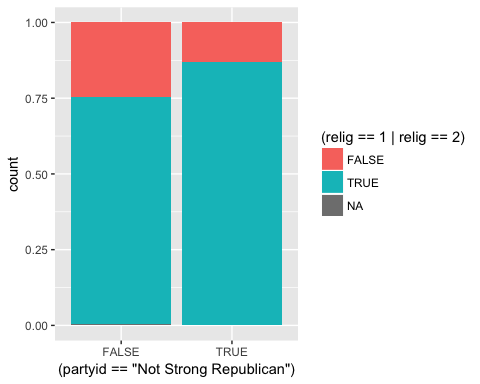
# Independent model full  
summary(independent <- glm((partyid == "Independent") ~ as.factor(race) + (relig == 4) + (born == 2 | reg16 == 0) + sex + age + rincom06 + sex:rincom06 + prestg80 + degree + degree:as.factor(race) + marital + adults + child1 + child2 + child3more + union, family = binomial(link = "logit"), data = gss))

##   
## Call:  
## glm(formula = (partyid == "Independent") ~ as.factor(race) +   
## (relig == 4) + (born == 2 | reg16 == 0) + sex + age + rincom06 +   
## sex:rincom06 + prestg80 + degree + degree:as.factor(race) +   
## marital + adults + child1 + child2 + child3more + union,   
## family = binomial(link = "logit"), data = gss)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6656 -0.7419 -0.5548 -0.3096 2.4122   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.077120 0.672703 -0.115 0.908729   
## as.factor(race)2 -0.808836 0.374019 -2.163 0.030575 \*   
## as.factor(race)3 -0.097765 0.286388 -0.341 0.732823   
## relig == 4TRUE 0.332002 0.186365 1.781 0.074838 .   
## born == 2 | reg16 == 0TRUE 0.831117 0.194450 4.274 1.92e-05 \*\*\*  
## sex2 -0.729471 0.392016 -1.861 0.062770 .   
## age -0.007056 0.006423 -1.099 0.271984   
## rincom06 -0.044351 0.020563 -2.157 0.031018 \*   
## prestg80 -0.022692 0.006614 -3.431 0.000602 \*\*\*  
## degree -0.254623 0.094468 -2.695 0.007032 \*\*   
## marital 0.003772 0.050976 0.074 0.941014   
## adults 0.020528 0.087287 0.235 0.814067   
## child1 0.490810 0.226421 2.168 0.030182 \*   
## child2 0.343103 0.226208 1.517 0.129328   
## child3more 0.258794 0.229163 1.129 0.258772   
## union 0.183538 0.084561 2.170 0.029971 \*   
## sex2:rincom06 0.049991 0.026707 1.872 0.061228 .   
## as.factor(race)2:degree 0.403271 0.209937 1.921 0.054742 .   
## as.factor(race)3:degree -0.010849 0.162515 -0.067 0.946776   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1331.4 on 1233 degrees of freedom  
## Residual deviance: 1208.8 on 1215 degrees of freedom  
## (3276 observations deleted due to missingness)  
## AIC: 1246.8  
##   
## Number of Fisher Scoring iterations: 4

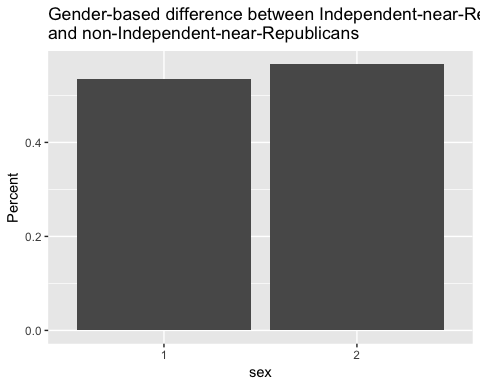
### Project Book

Exploratory visualizations used in exploring relevant characteristics for the models:

ggplot(gss[gss$partyid %in% partynames[c(1:7)], ],  
 aes(x = (partyid == "Not Strong Republican"), fill = (relig == 1 | relig == 2))) +  
 geom\_bar(position = "fill")

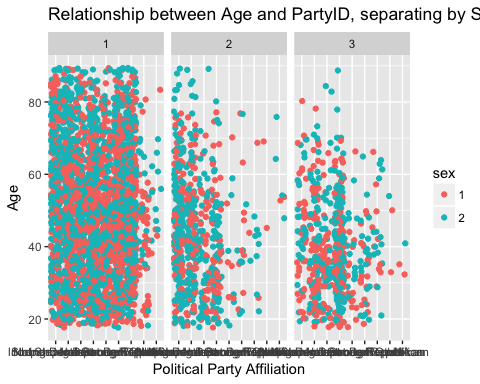


ggplot() +  
 geom\_bar(data = gss[gss$partyid %in% partynames[c(1:4,6,7)], ],   
 aes(x = sex, y = (..count..)/sum(..count..))) +  
 geom\_bar(data = gss[gss$partyid %in% partynames[5], ],  
 aes(x = sex, y = (..count..)/sum(..count..))) +  
 scale\_y\_continuous("Percent") +  
 ggtitle("Gender-based difference between Independent-near-Republicans\nand non-Independent-near-Republicans")

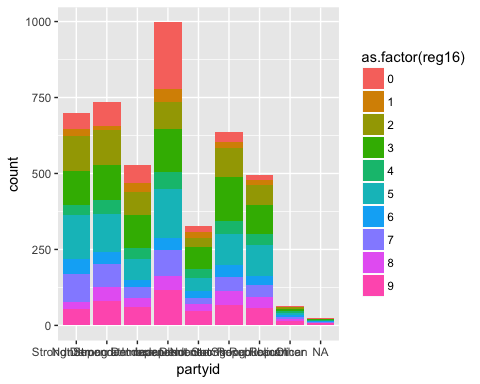


ggplot(gss, aes(x = partyid, y = age)) +   
 geom\_jitter((aes(color=sex))) +   
 facet\_wrap(~ race) +   
 ggtitle("Relationship between Age and PartyID, separating by Sex and Race") +   
 labs(x="Political Party Affiliation", y="Age")

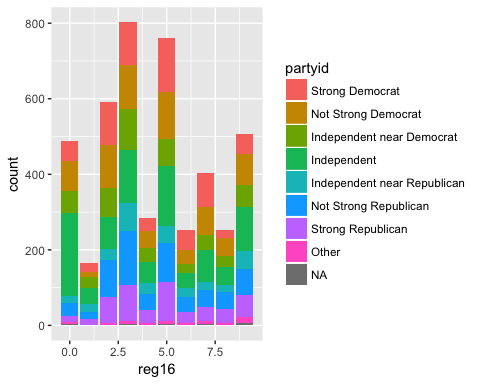
## Warning: Removed 18 rows containing missing values (geom\_point).



ggplot(gss, aes(x = partyid, fill = as.factor(reg16))) +  
 geom\_bar(stat = "count")



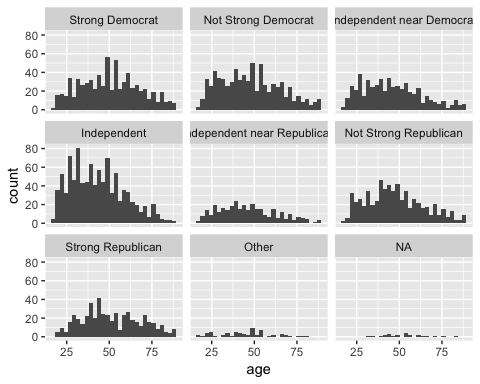
ggplot(gss, aes(x = reg16, fill = partyid)) +  
 geom\_bar(stat = "count")



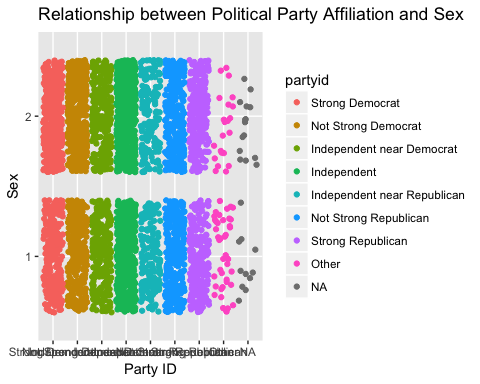
ggplot(gss, aes(x = age)) +  
 geom\_histogram() +   
 facet\_wrap(~ partyid)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

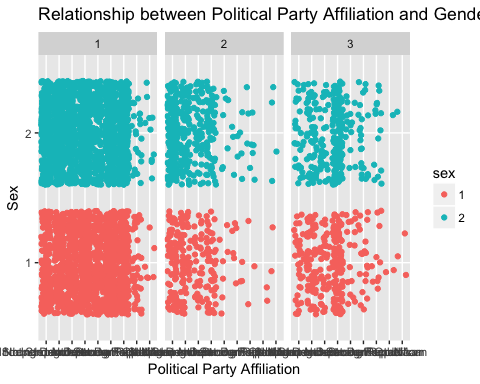
## Warning: Removed 18 rows containing non-finite values (stat\_bin).



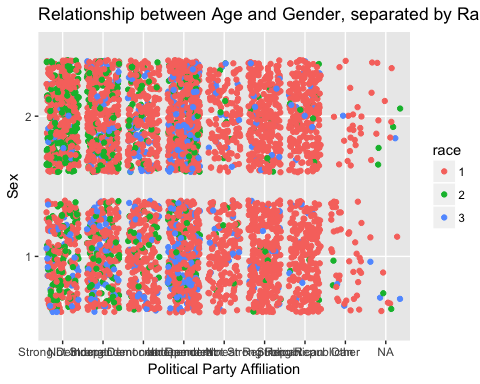
ggplot(gss, aes(x = partyid, y = sex)) +   
 geom\_jitter((aes(color=partyid))) +   
 ggtitle("Relationship between Political Party Affiliation and Sex") +   
 labs(x="Party ID", y="Sex")



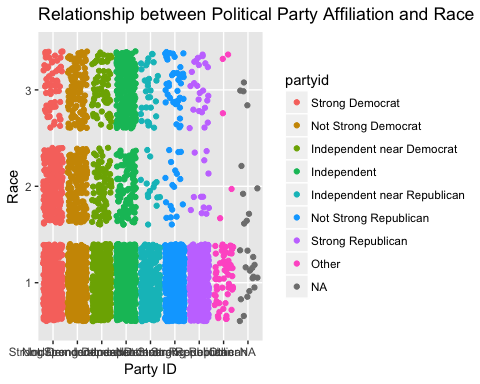
ggplot(gss, aes(x = partyid, y = sex)) +   
 geom\_jitter((aes(color=sex))) +   
 facet\_wrap(~ race) +   
 ggtitle("Relationship between Political Party Affiliation and Gender, separated by Race") +   
 labs(x="Political Party Affiliation", y="Sex")



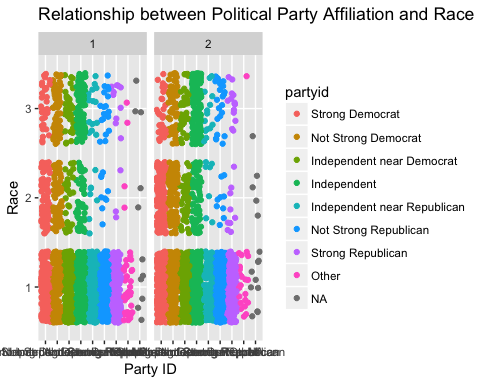
ggplot(gss, aes(x = partyid, y = sex)) +   
 geom\_jitter((aes(color=race))) +   
 ggtitle("Relationship between Age and Gender, separated by Race") +   
 labs(x="Political Party Affiliation", y="Sex")



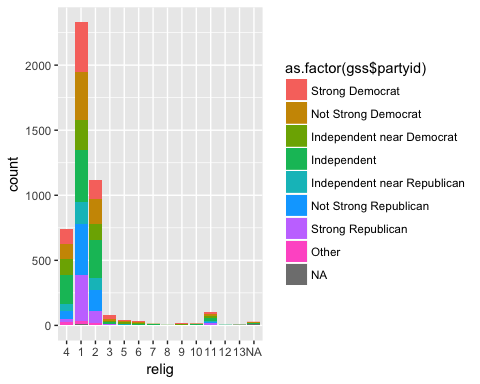
ggplot(gss, aes(x = partyid, y = race)) +   
 geom\_jitter((aes(color=partyid))) +   
 ggtitle("Relationship between Political Party Affiliation and Race") +   
 labs(x="Party ID", y="Race")



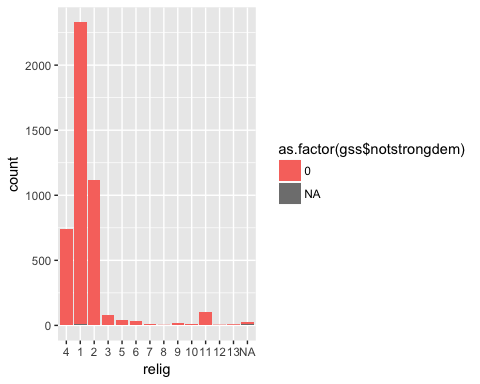
ggplot(gss, aes(x = partyid, y = race)) +   
 geom\_jitter((aes(color=partyid))) +   
 ggtitle("Relationship between Political Party Affiliation and Race") +   
 labs(x="Party ID", y="Race") + facet\_wrap(~sex)



gss$notstrongdem <- ifelse(gss$partyid==1, 1, 0)  
   
ggplot(gss, aes(relig)) +   
 geom\_bar(aes(fill=as.factor(gss$partyid)))

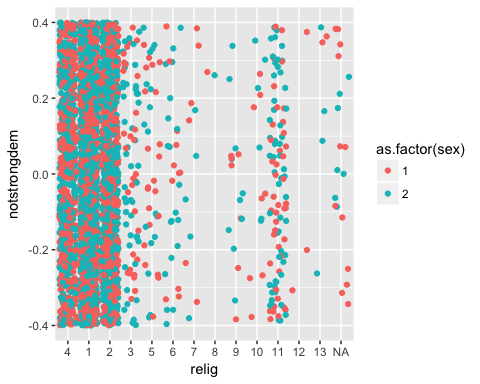


#blue - weak democrat  
ggplot(gss, aes(relig)) +   
 geom\_bar(aes(fill=as.factor(gss$notstrongdem)))



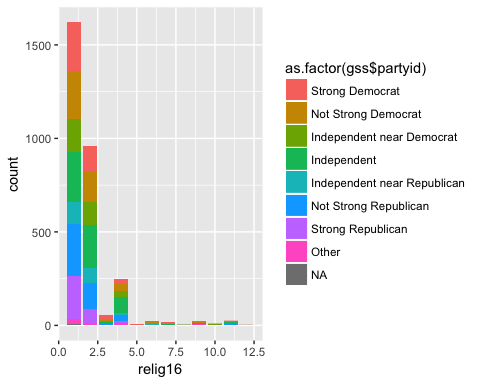
ggplot(gss, aes(x=relig, y=notstrongdem, color=as.factor(sex))) +   
 geom\_jitter()

## Warning: Removed 26 rows containing missing values (geom\_point).



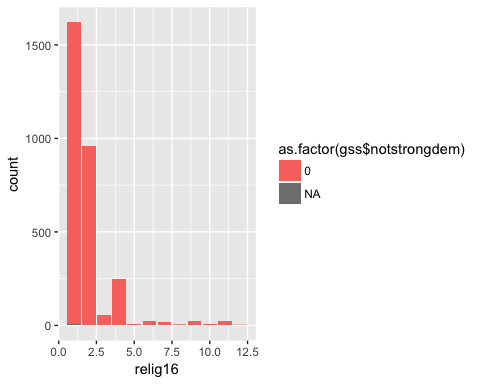
#relig16  
ggplot(gss, aes(relig16)) +   
 geom\_bar(aes(fill=as.factor(gss$partyid)))

## Warning: Removed 1519 rows containing non-finite values (stat\_count).



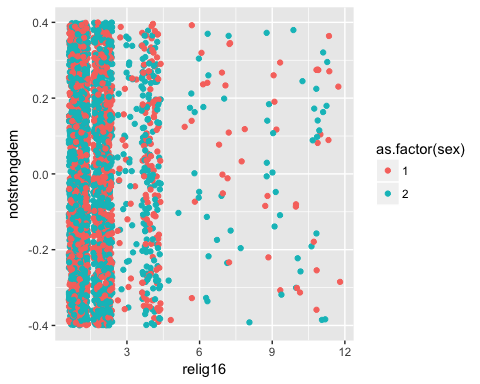
#blue - weak democrat  
ggplot(gss, aes(relig16)) +   
 geom\_bar(aes(fill=as.factor(gss$notstrongdem)))

## Warning: Removed 1519 rows containing non-finite values (stat\_count).

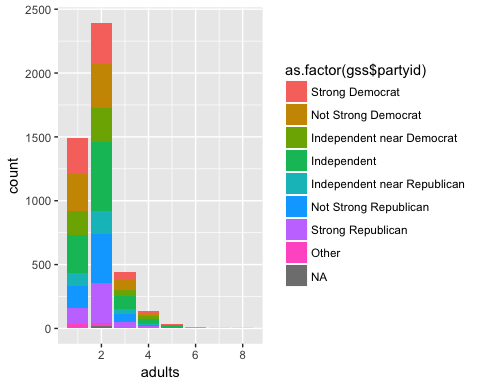


ggplot(gss, aes(x=relig16, y=notstrongdem, color=as.factor(sex))) +   
 geom\_jitter()

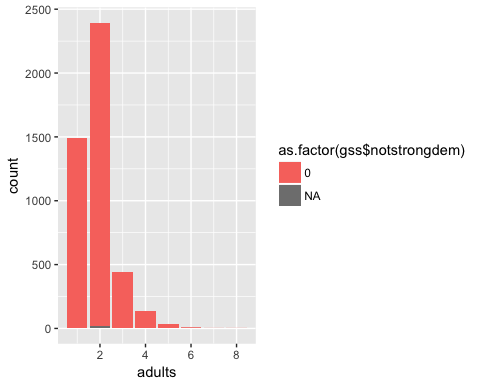
## Warning: Removed 1533 rows containing missing values (geom\_point).



#adults  
ggplot(gss, aes(adults)) +   
 geom\_bar(aes(fill=as.factor(gss$partyid)))

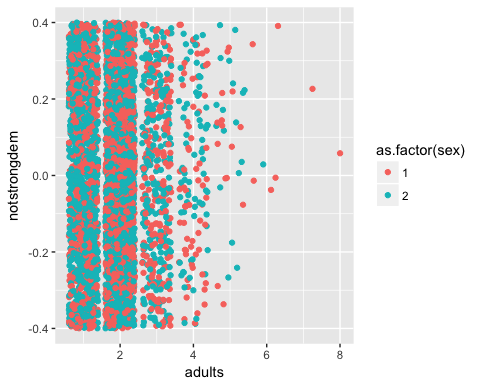


#blue - weak democrat  
ggplot(gss, aes(adults)) +   
 geom\_bar(aes(fill=as.factor(gss$notstrongdem)))



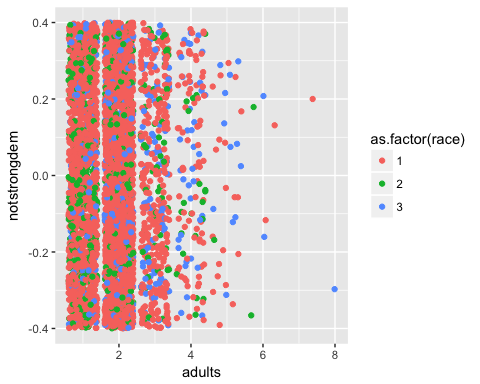
ggplot(gss, aes(x=adults, y=notstrongdem, color=as.factor(sex))) +   
 geom\_jitter()

## Warning: Removed 26 rows containing missing values (geom\_point).



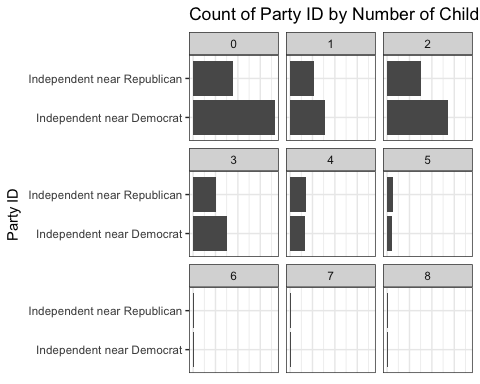
ggplot(gss, aes(x=adults, y=notstrongdem, color=as.factor(race))) +   
 geom\_jitter()

## Warning: Removed 26 rows containing missing values (geom\_point).

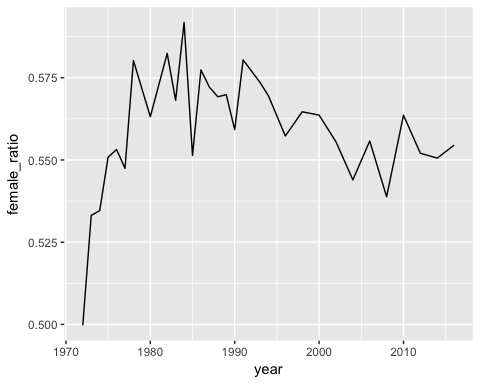


gss\_subset <- filter(gss, partyid=="Independent near Democrat" | partyid=="Independent near Republican", childs!="NA")  
  
ggplot(gss\_subset, aes(x = partyid)) +   
 geom\_histogram(stat="count") +   
 facet\_wrap(~ childs) +   
 coord\_flip() +   
 theme\_bw() +   
 xlab("Party ID") + ylab(NULL) +   
 ggtitle("Count of Party ID by Number of Children") +   
 theme(axis.title.x=element\_blank(), axis.text.x=element\_blank(),   
 axis.ticks.x=element\_blank())

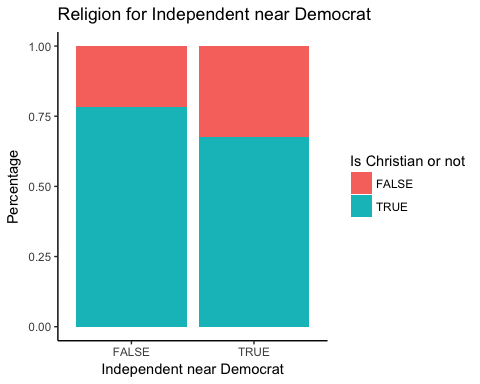
## Warning: Ignoring unknown parameters: binwidth, bins, pad



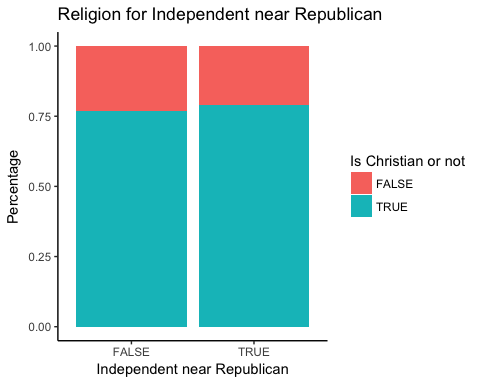
gss\_nona <- gss\_sub[!is.na(gss\_sub$partyid),]  
  
gss\_nona %>% group\_by(year) %>%  
 summarise(male = sum(sex == 1),  
 female = sum(sex == 2),  
 total = sum(sex == 1 | sex == 2),  
 male\_ratio = male/total,  
 female\_ratio = female/total) %>%  
ggplot(aes(x = year)) +  
 geom\_line(aes(y = female\_ratio))



ggplot(gss[(is.na(gss$relig) == F & gss$partyid %in% partynames[c(1:7)]), ],   
 aes(x = (partyid == "Independent near Democrat"), fill = (relig == 1 | relig == 2))) +   
 geom\_bar(position = "fill") + ggtitle("Religion for Independent near Democrat") +   
 ylab("Percentage") + xlab("Independent near Democrat") + theme\_classic() +   
 scale\_fill\_discrete(name = "Is Christian or not")



ggplot(gss[(is.na(gss$relig) == F & gss$partyid %in% partynames[c(1:7)]), ],   
 aes(x = (partyid == "Independent near Republican"), fill = (relig == 1 | relig == 2))) +   
 geom\_bar(position = "fill") + ggtitle("Religion for Independent near Republican") +   
 ylab("Percentage") + xlab("Independent near Republican") + theme\_classic() +   
 scale\_fill\_discrete(name = "Is Christian or not")



gss$republican <- (gss$partyid == "Not Strong Republican" | gss$partyid == "Strong Republican")  
ggplot(gss[(is.na(gss$relig) == F & gss$partyid %in% partynames[c(1:7)]), ],   
 aes(x = (republican == 1), fill = (relig == 1 | relig == 2))) +   
 geom\_bar(position = "fill") + ggtitle("Religion for Republican") +   
 ylab("Percentage") + xlab("Republican") + theme\_classic() +   
 scale\_fill\_discrete(name = "Is Christian or not")

