TURNOUT EXAMPLE Exploring the ANES Cumulative File (1948–2012)

A Person

Read in the cumulative file

Read in the ANES Cumulative File (1948-2012) and create a survey object with weights.

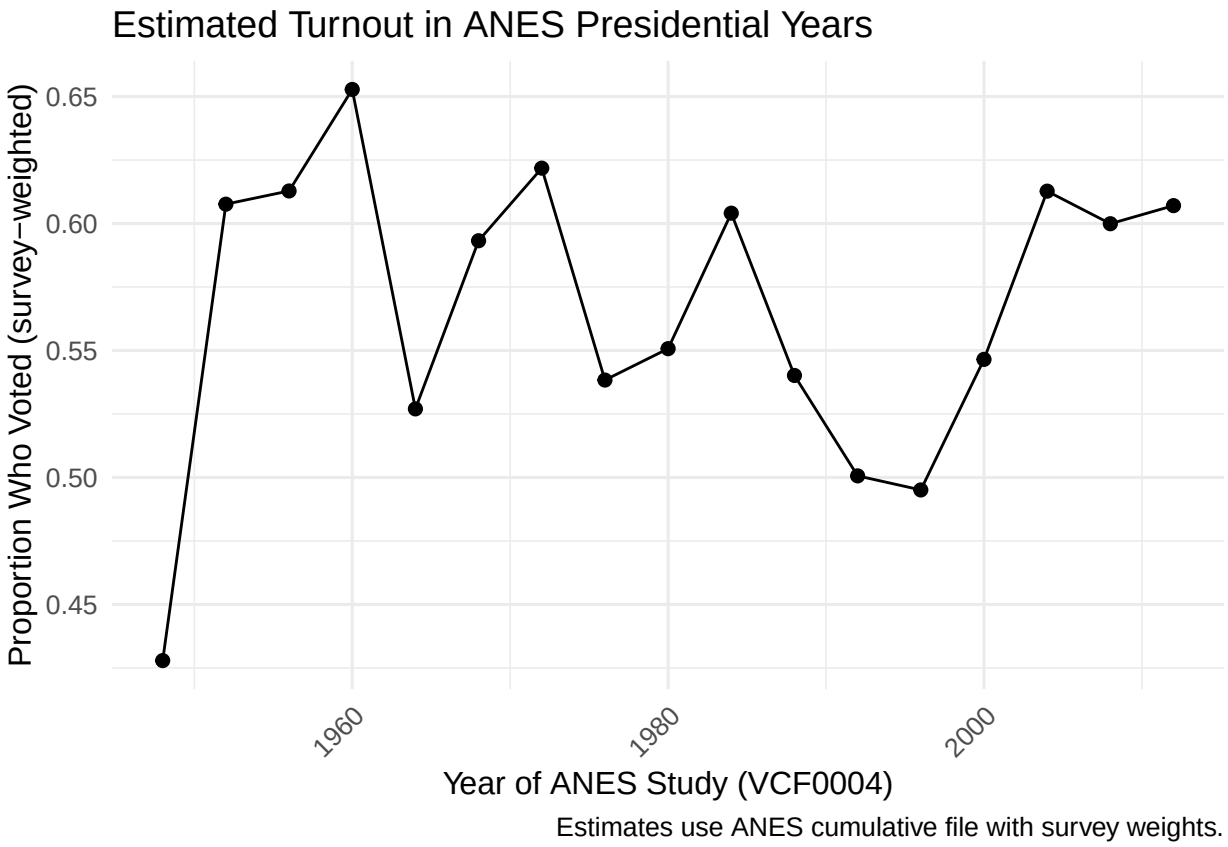
Create recoded variables

Create new variables indicating whether the respondent voted, whether they voted for the Democratic candidate, and create a "democrat incumbent flipper" that you can multiply by some issue positions to code them in a "democratic incumbent higher scores liberal, republican incumbent higher scores conservative" way.

Turnout Over Time

Simple weighted tables and line chart showing the proportion of respondents who voted in each election year

Year	Proportion Who Voted	prop_voted_se
1948	42.7885%	0.0242585
1952	60.7627%	0.0142146
1956	61.2825%	0.0138779
1960	65.2803%	0.0226195
1964	52.7007%	0.0190764
1968	59.3220%	0.0170924
1972	62.1803%	0.0119675
1976	53.8358%	0.0149017
1980	55.0725%	0.0166086
1984	60.4072%	0.0134303
1988	54.0171%	0.0145706
1992	50.0628%	0.0149970
1996	49.5089%	0.0199400
2000	54.6499%	0.0186651
2004	61.2749%	0.0216553
2008	59.9920%	0.0182610
2012	60.7052%	0.0121510



Leighley Nagler

Compare the impact of race on turnout in 1988, 2000, and 2016

```
# Restrict survey design to individual years
anes1988 <- subset(cumulative_nes_weight_recode, VCF0004 == 1988) %>%
  mutate(age = as.numeric(VCF0101))   # Need to do this for the margins command to work at the end
anes2000 <- subset(cumulative_nes_weight_recode, VCF0004 == 2000)
anes1016 <- subset(cumulative_nes_weight_recode, VCF0004 == 2016)


# OLS / Linear Probability Model for 1988
ols_1988 <-
  svyglm(
    voted ~ pol377black + pol377inc2 + pol377inc3 + pol377inc4 + pol377inc5 + age,
    design = anes1988,
    family = gaussian()
  )

# Logistic Regression (Logit) for 1988
logit_1988 <-
  svyglm(
    voted  ~ pol377black + pol377inc2 + pol377inc3 + pol377inc4 + pol377inc5 + age,
    design = anes1988,
    family = quasibinomial()
```

```
  )

# Display model summaries
summary(ols_1988)
```

```
##
## Call:
## svyglm(formula = voted ~ pol377black + pol377inc2 + pol377inc3 +
##     pol377inc4 + pol377inc5 + age, design = anes1988, family = gaussian())
##
## Survey design:
## subset(cumulative_nes_weight_recode, VCF0004 == 1988)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0086580  0.0471870   0.183   0.8545
## pol377black -0.2992416  0.0406846  -7.355 3.78e-13 ***
## pol377inc2   0.0908452  0.0471115   1.928   0.0541 .
## pol377inc3   0.2761618  0.0386249   7.150 1.60e-12 ***
## pol377inc4   0.4396874  0.0394498  11.145  < 2e-16 ***
## pol377inc5   0.6200194  0.0537302  11.539  < 2e-16 ***
## age          0.0067628  0.0007928   8.530  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1934985)
##
## Number of Fisher Scoring iterations: 2
```

```
margins(ols_1988)
```

```
## Note: Estimating marginal effects without survey weights. Specify 'design' to adjust for weighting.

## Average marginal effects

## svyglm(formula = voted ~ pol377black + pol377inc2 + pol377inc3 +     pol377inc4 + pol377inc5 + age, d

##  pol377black pol377inc2 pol377inc3 pol377inc4 pol377inc5      age
##      -0.2992    0.09085     0.2762     0.4397       0.62 0.006763
```

```
summary(logit_1988)
```

```
##
## Call:
## svyglm(formula = voted ~ pol377black + pol377inc2 + pol377inc3 +
##     pol377inc4 + pol377inc5 + age, design = anes1988, family = quasibinomial())
##
## Survey design:
## subset(cumulative_nes_weight_recode, VCF0004 == 1988)
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.518979   0.299793  -8.402  < 2e-16 ***
## pol377black -1.985747   0.387570  -5.124 3.55e-07 ***
## pol377inc2   0.521812   0.250754   2.081   0.0377 *
## pol377inc3   1.411066   0.217720   6.481 1.38e-10 ***
## pol377inc4   2.153577   0.228401   9.429  < 2e-16 ***
## pol377inc5   3.791459   0.773666   4.901 1.10e-06 ***
## age          0.034353   0.004482   7.665 4.00e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.076129)
##
## Number of Fisher Scoring iterations: 5
```

```
margins(logit_1988)
```

```
## Note: Estimating marginal effects without survey weights. Specify 'design' to adjust for weighting.

## Average marginal effects

## svyglm(formula = voted ~ pol377black + pol377inc2 + pol377inc3 +     pol377inc4 + pol377inc5 + age, d

##  pol377black pol377inc2 pol377inc3 pol377inc4 pol377inc5      age
##      -0.3813     0.1002     0.2709     0.4135      0.728 0.006596
```

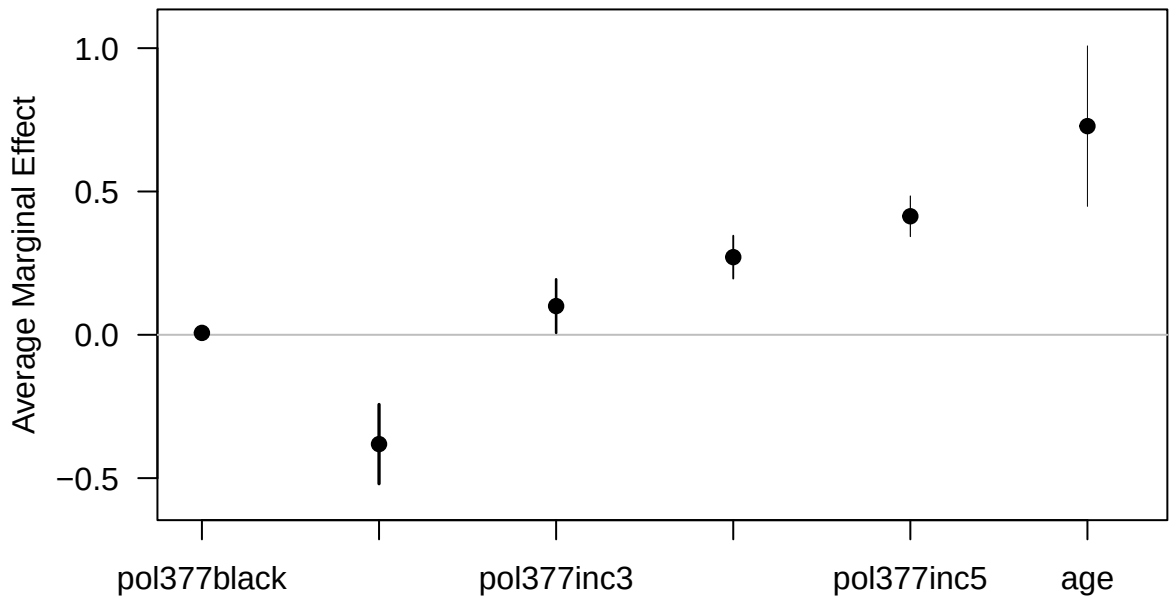
```
m <- margins(logit_1988)
```

```
## Note: Estimating marginal effects without survey weights. Specify 'design' to adjust for weighting.
```

```
summary(m)
```

```
##       factor     AME     SE       z      p   lower   upper
##          age  0.0066 0.0007  8.8108 0.0000  0.0051  0.0081
##  pol377black -0.3813 0.0705 -5.4056 0.0000 -0.5195 -0.2430
##   pol377inc2  0.1002 0.0477  2.0992 0.0358  0.0066  0.1937
##   pol377inc3  0.2709 0.0382  7.0968 0.0000  0.1961  0.3458
##   pol377inc4  0.4135 0.0361 11.4560 0.0000  0.3428  0.4842
##   pol377inc5  0.7280 0.1429  5.0957 0.0000  0.4480  1.0080
```

```
plot(m)
```



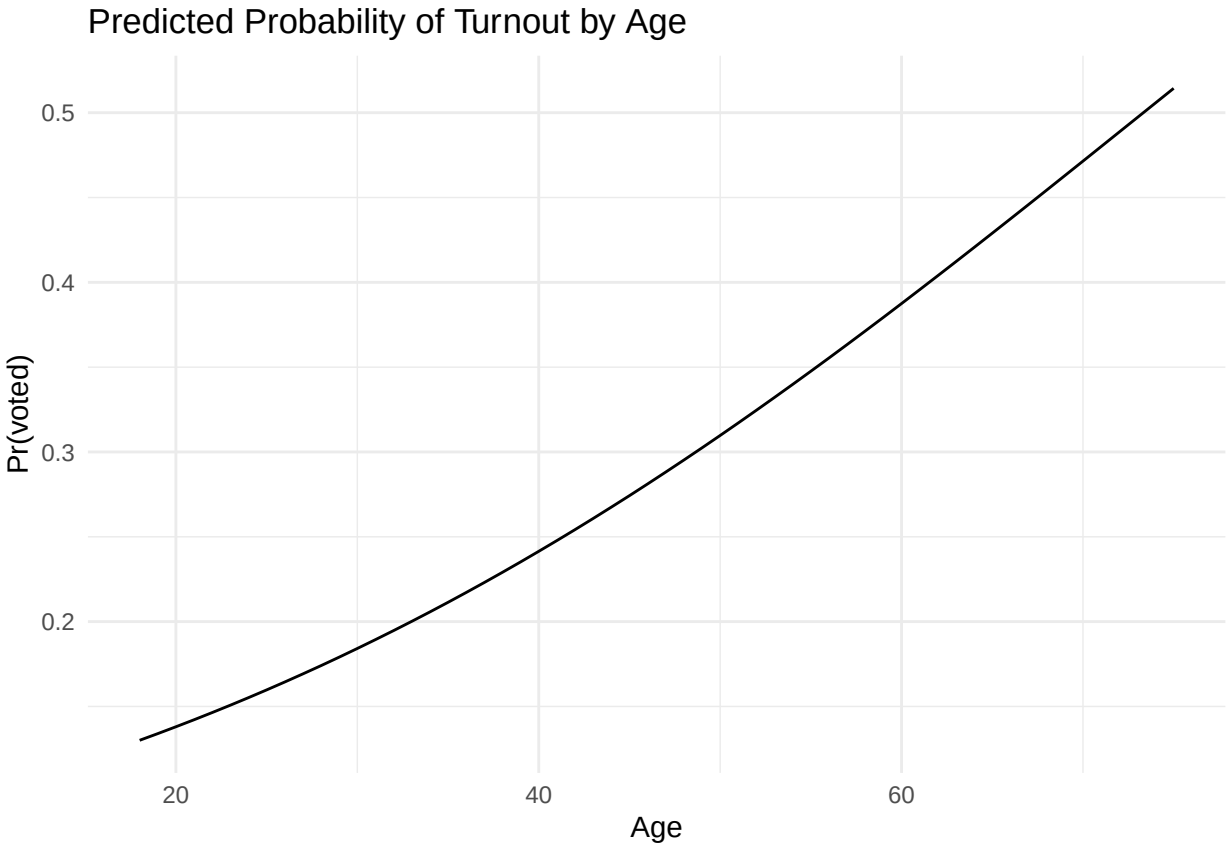

```r
# New data for age 18-65, holding other vars at typical values
newdat <- expand.grid(
  pol377black = 0,
  pol377inc2  = 0,
  pol377inc3  = 0,
  pol377inc4  = 0,
  pol377inc5 = 0,
  age         = seq(18, 75, by = 1)
)

# Predicted probabilities from the logit model
newdat$pred_p <- predict(logit_1988, newdata = newdat, type = "response")

# Plot

ggplot(newdat, aes(x = age, y = pred_p)) +
  geom_line() +
  labs(
    x = "Age",
    y = "Pr(voted)",
    title = "Predicted Probability of Turnout by Age"
  ) +
  theme_minimal()
```

```
## Don't know how to automatically pick scale for object of type <svystat>.
## Defaulting to continuous.
```


Predicted Probability of Turnout by Age
0.5
0.4
0.3
0.2
Pr(voted)
20
40
60
Age