R Notebook

AP VoteCast is a survey of the American electorate conducted in all 50 states by NORC at the University of Chicago for The Associated Press and Fox News. The survey is funded by AP. The survey of 138,929 registered voters was conducted October 29 to November 6, 2018, concluding as polls closed on Election Day. Interviews were conducted via phone and web, with 11,059 completing by phone and 127,870 completing by web.

```
library(readr)
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
```

```
library(ggplot2)
library(fastDummies)
knitr::opts_chunk$set(echo = TRUE)
def.chunk.hook  <- knitr::knit_hooks$get("chunk")
knitr::knit_hooks$set(chunk = function(x, options) {
  x <- def.chunk.hook(x, options)
  ifelse(options$size != "normalsize", paste0("\\", options$size,"\n\n", x, "\n\n \\normalsize"), x)
})
require(runjags)
```

```
## Loading required package: runjags
```

```
require(coda)
```

```
## Loading required package: coda
```

```
data <- read.csv("voting.csv")
```

```
data$RACETH5 <- recode(data$RACETH5, "(1) White" = 1, "(2) African American or Black" = 2, "(3) Latino
data$RACETH <- NULL
data$EDUC <- recode(data$EDUC, "(1) High school or less" = 1, "(2) Some college/assoc. degree" = 2, "(3)
data$INCOME <- recode(data$INCOME, "(1) Under $25,000" = 1, "(2) $25,000-$49,999" = 2, "(3) $50,000-$74
data$SEX <- recode(data$SEX, "(1) Men" = 1, "(2) Women" =2, "(99) DON'T KNOW/SKIPPED/REFUSED (VOL)" = 9
data$AGE <- recode(data$AGE65, "(1) 18-24" = 1, "(2) 25-29" = 2, "(3) 30-39" = 3, "(4) 40-49" = 4, "(5)
data$AGE65 <- NULL
data$RELIG <- recode(data$RELIG, "(1) Protestant" = 1, "(2) Catholic" = 2, "(3) Mormon" = 3, "(4) Other
data$PARTY <- NULL
data$PARTY <- recode(data$PARTYFULL, "(1) Democrat/Lean Dem" = 1, "(2) Republican/Lean Rep" =2, "(3) In
data$PARTYFULL <- NULL
data$IDEO <- recode(data$IDEO, "(1) Very liberal" = 1, "(2) Somewhat liberal" = 2, "(3) Moderate" = 3,
data$RELIG <- recode(data$RELIG4, "(1) Protestant/Other Christian" = 1, "(2) Catholic"= 2, "(3) Other"
data$RELIG4 <- NULL
data$FORMFLAG <- NULL
```

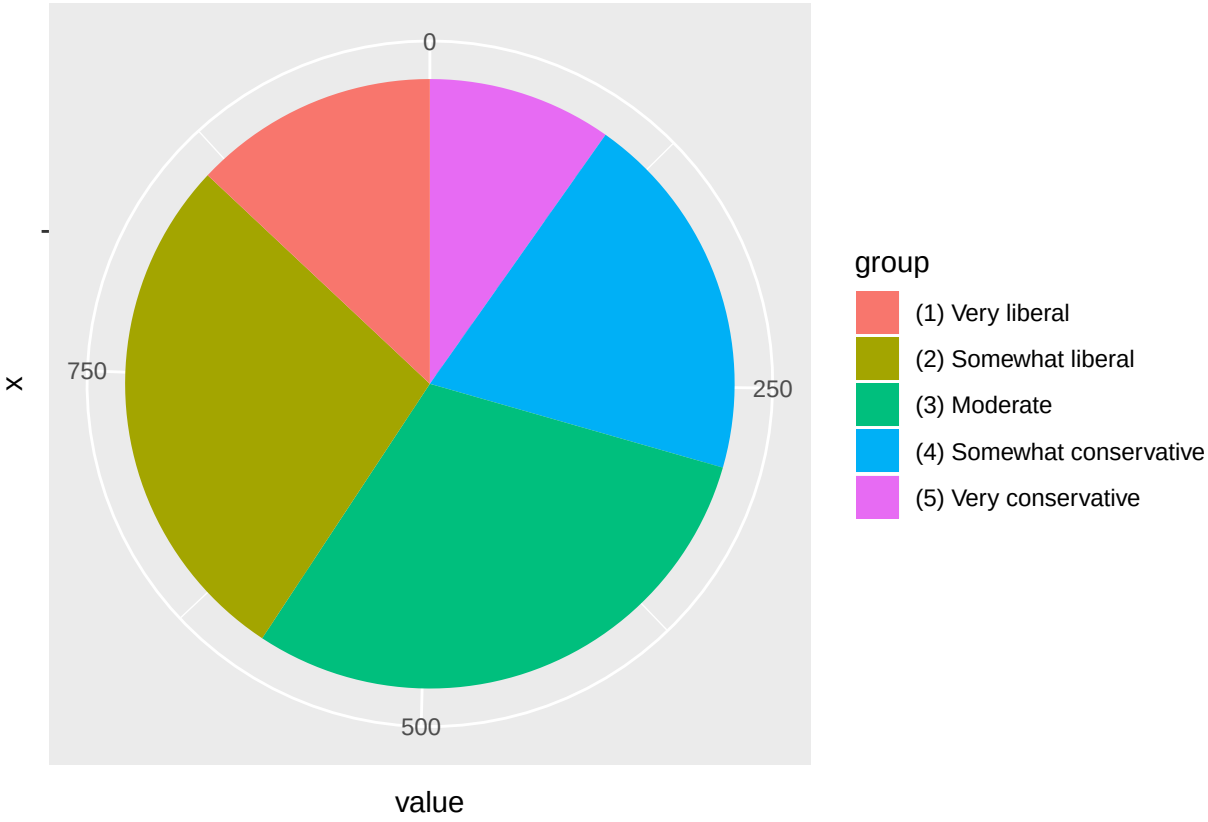
```
data[1:5,]
```

```
##      SU_ID POLLCLOSE_STATE_WEIGHT POLLCLOSE_NATIONAL_WEIGHT
## 1 20000058              86.625560                 83.562187
## 2 20000102             231.679211                223.486250
## 3 20000106            7639.065341               7368.922117
## 4 20000165              34.959933                 33.723631
## 5 20000191               4.811467                  4.641317
##   FINALVOTE_STATE_WEIGHT FINALVOTE_NATIONAL_WEIGHT            P_STATE
## 1              57.404245                  55.73196       (MT) Montana
## 2             156.736562                 178.51506        (KS) Kansas
## 3             502.237088                 484.47628       (FL) Florida
## 4              37.813306                  36.95838       (MT) Montana
## 5               3.857918                   3.67032 (NH) New Hampshire
##             STATENUM                       LVA                      LVB
## 1       (26) Montana (1) Extremely interested      (5) I already voted
## 2        (16) Kansas (1) Extremely interested (1) Definitely will vote
## 3        (9) Florida (1) Extremely interested (1) Definitely will vote
## 4       (26) Montana (1) Extremely interested      (5) I already voted
## 5 (30) New Hampshire (1) Extremely interested (1) Definitely will vote
##                         LV      LIKELYVOTER           TIMEVOTE
## 1        (12) Already voted (1) Likely voter (2) Early/Absentee
## 2 (10) 10-Certain will vote (1) Likely voter (2) Early/Absentee
## 3 (10) 10-Certain will vote (1) Likely voter   (1) Election Day
## 4        (12) Already voted (1) Likely voter (2) Early/Absentee
## 5 (10) 10-Certain will vote (1) Likely voter   (1) Election Day
##                                TRUMPFACTOR3          FAVSENDEM
## 1 (2) Voted to express opposition for Trump (1) Very favorable
## 2                (3) Trump was not a factor
## 3                (3) Trump was not a factor (1) Very favorable
## 4 (2) Voted to express opposition for Trump (1) Very favorable
## 5    (1) Voted to express support for Trump
##                                FAVSENREP           PARTYCONTROL
## 1 (99) DON'T KNOW/SKIPPED/REFUSED (VOL)  (3) Not too important
## 2
## 3                   (4) Very unfavorable (2) Somewhat important
## 4                   (4) Very unfavorable     (1) Very important
## 5
##             GETAHEAD TRADENATIONALECON TRADELOCALECON
## 1 (2) Holding steady          (2) Hurt       (2) Hurt
## 2  (1) Getting ahead          (1) Help       (2) Hurt
## 3 (3) Falling behind (3) No difference       (1) Help
## 4 (3) Falling behind          (2) Hurt       (2) Hurt
## 5  (1) Getting ahead          (1) Help       (1) Help
##                      HEALTHLAW
## 1          (4) Expand the law
## 2 (1) Repeal the law entirely
## 3 (2) Repeal parts of the law
## 4          (4) Expand the law
## 5 (2) Repeal parts of the law
##                                       IMMDEPORT
## 1 (1) Offered a chance to apply for legal status
## 2 (1) Offered a chance to apply for legal status
## 3 (1) Offered a chance to apply for legal status
```

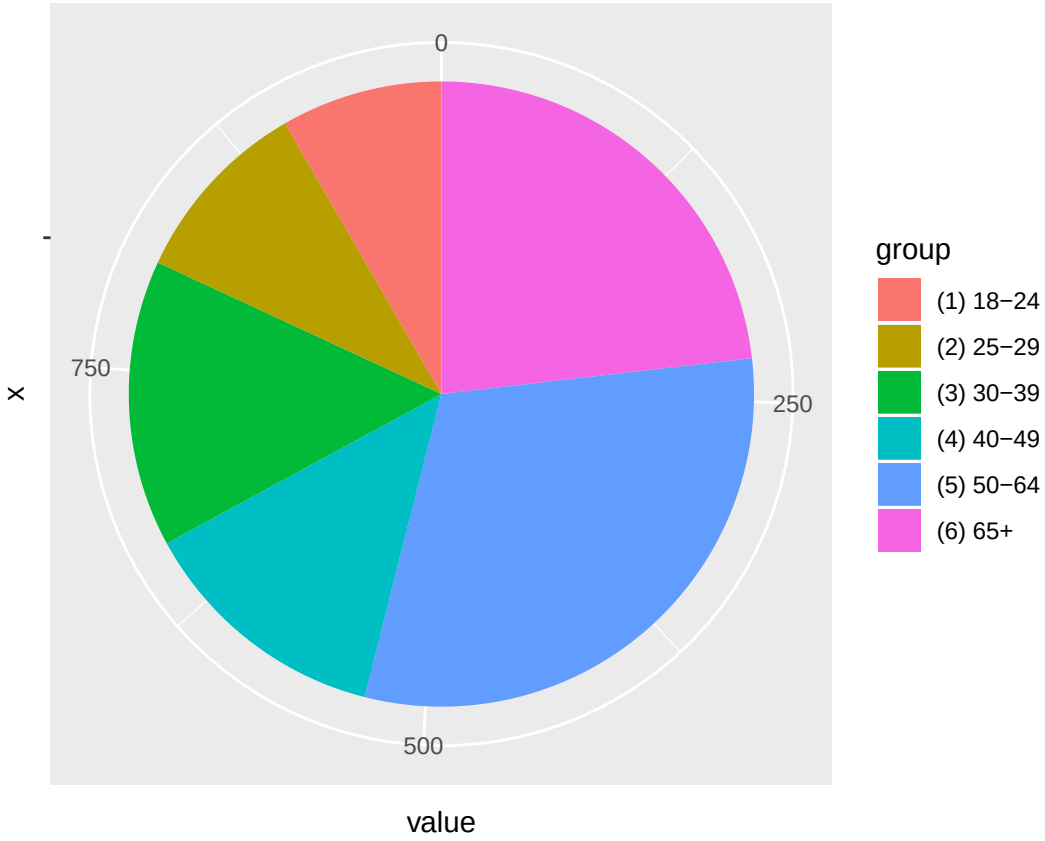
```
## 4 (1) Offered a chance to apply for legal status
## 5 (1) Offered a chance to apply for legal status
##                                RUSSIA
## 1                             (1) Yes
## 2                              (2) No
## 3 (99) DON'T KNOW/SKIPPED/REFUSED (VOL)
## 4                             (1) Yes
## 5                              (2) No
##                                                                QPVVOTE
## 1                                                    (4) I'm sure I voted
## 2                                                    (4) I'm sure I voted
## 3 (3) I usually vote, but I didn't in the 2016 presidential election
## 4                                                    (4) I'm sure I voted
## 5                                                    (4) I'm sure I voted
##   FIRSTTIME
## 1   (2) Not
## 2   (2) Not
## 3   (2) Not
## 4   (2) Not
## 5   (2) Not
##                                                                   QPVVOTE3
## 1                                                       (4) I'm sure I voted
## 2                                                       (4) I'm sure I voted
## 3 (2) I thought about voting in the 2014 election for Congress, but didn't
## 4                                                       (4) I'm sure I voted
## 5                                                       (4) I'm sure I voted
##              BREAKA            BREAKB
## 1 (2) No, it is not    (1) Yes, it is
## 2 (2) No, it is not (2) No, it is not
## 3 (2) No, it is not    (1) Yes, it is
## 4 (2) No, it is not    (1) Yes, it is
## 5    (1) Yes, it is (2) No, it is not
##                                        BREAKC SEX RACETH5 EDUC INCOME
## 1                       (1) I've known all along   2       1    2      2
## 2 (2) I decided over the course of the campaign   1       1    4      4
## 3                       (1) I've known all along   1       2    1      5
## 4                       (1) I've known all along   1       1    4      3
## 5 (2) I decided over the course of the campaign   1       3    4      5
##   IDEO RELIG      SIZEPLACE AGE PARTY
## 1    2     1      (4) Rural   6     3
## 2    4     1 (3) Small town   4     2
## 3   99     3 (3) Small town   5     1
## 4    1     2 (3) Small town   6     1
## 5    4     1   (2) Suburban   5     2
```

```r
data <- data %>% filter(IDEO<99)
IDEO_data <- c()
for (i in 1:5){
  IDEO_data[i] <- sum(data$IDEO==i)
}
lbls <- c("(1) Very liberal", "(2) Somewhat liberal", "(3) Moderate", "(4) Somewhat conservative", "(5)
df <- data.frame(
  group = lbls,
  value = IDEO_data
```

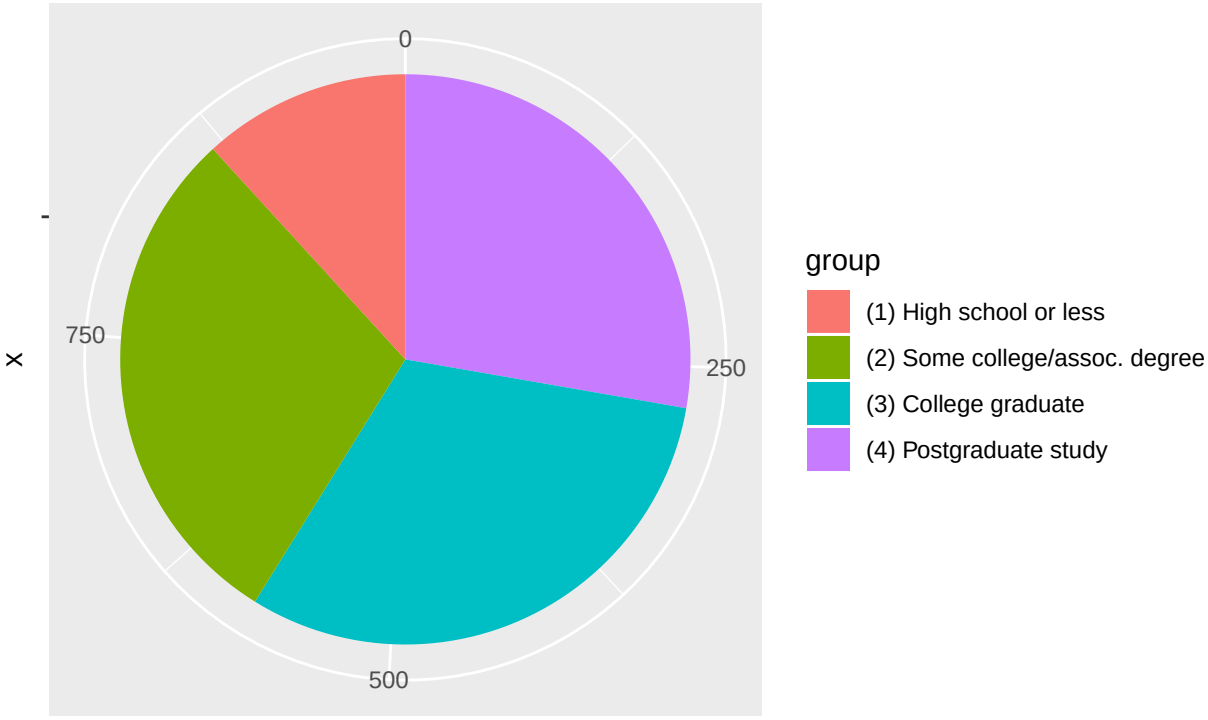
```
  )
ggplot(df, aes(x="", y=value, fill=group))+ geom_bar(width = 1, stat = "identity")+ coord_polar("y", sta
```



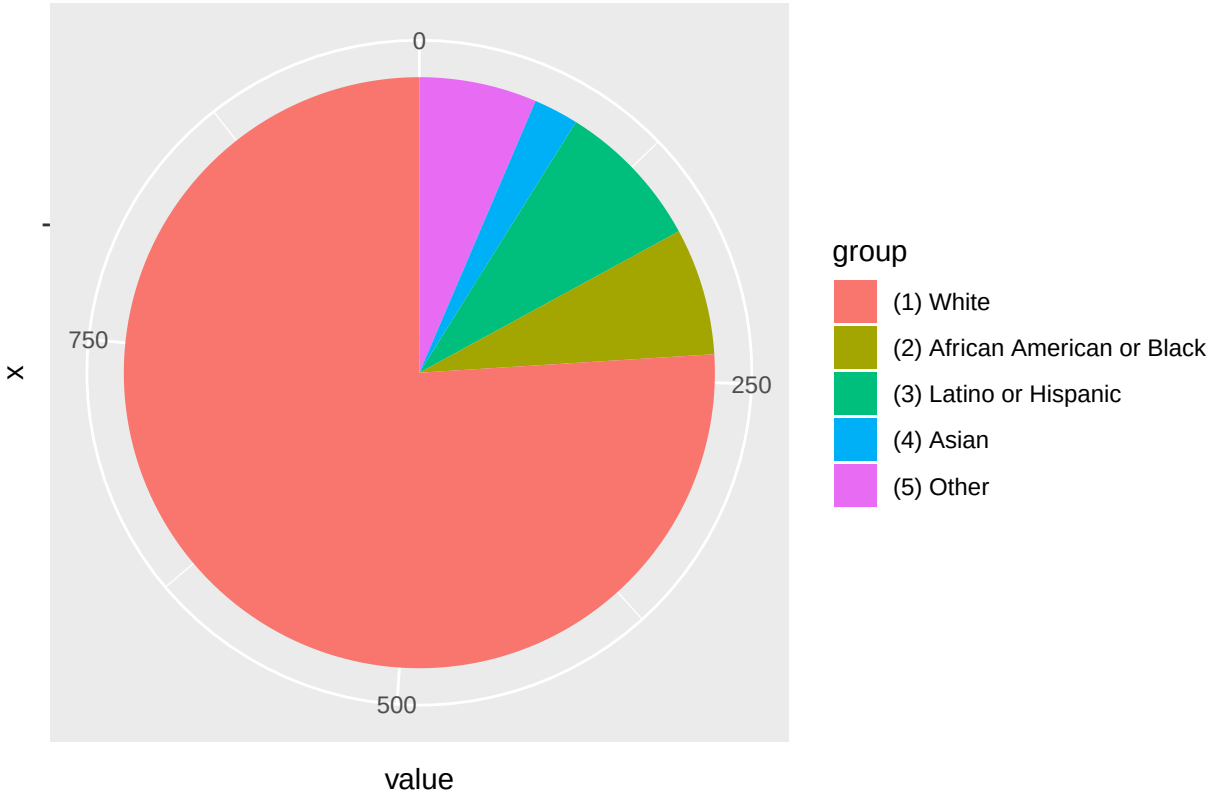

```
data <- data %>% filter(AGE<99)
AGE_data <- c()
for (i in 1:6){
  AGE_data[i] <- sum(data$AGE==i)
}
lbls <- c("(1) 18-24",
"(2) 25-29",
"(3) 30-39",
"(4) 40-49",
"(5) 50-64",
"(6) 65+")
df <- data.frame(
  group = lbls,
  value = AGE_data
  )
ggplot(df, aes(x="", y=value, fill=group))+ geom_bar(width = 1, stat = "identity")+ coord_polar("y", sta
```



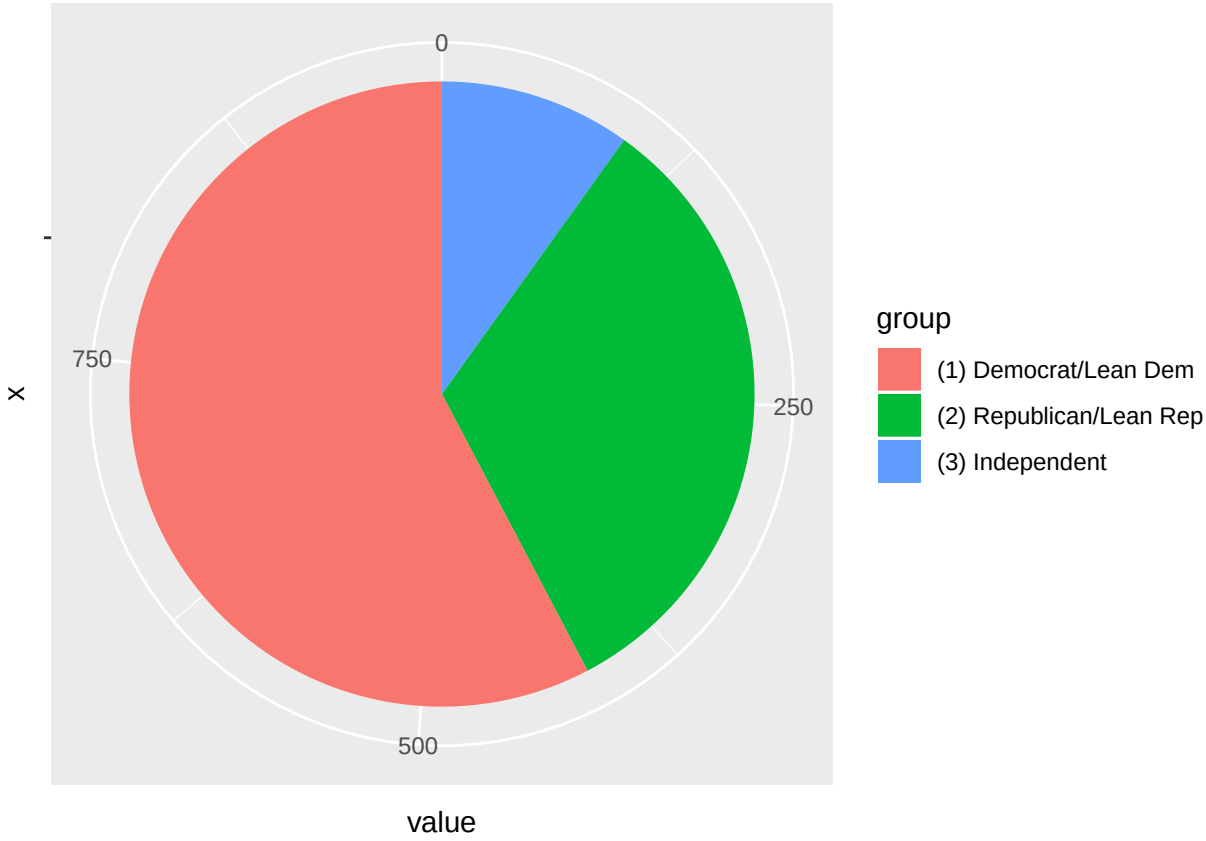

```
data <- data %>% filter(EDUC<88)
EDUC_data <- c()
for (i in 1:4){
  EDUC_data[i] <- sum(data$EDUC==i)
}
lbls <- c("(1) High school or less",
"(2) Some college/assoc. degree",
"(3) College graduate",
"(4) Postgraduate study")
df <- data.frame(
  group = lbls,
  value = EDUC_data
  )
ggplot(df, aes(x="", y=value, fill=group))+ geom_bar(width = 1, stat = "identity")+ coord_polar("y", st
```



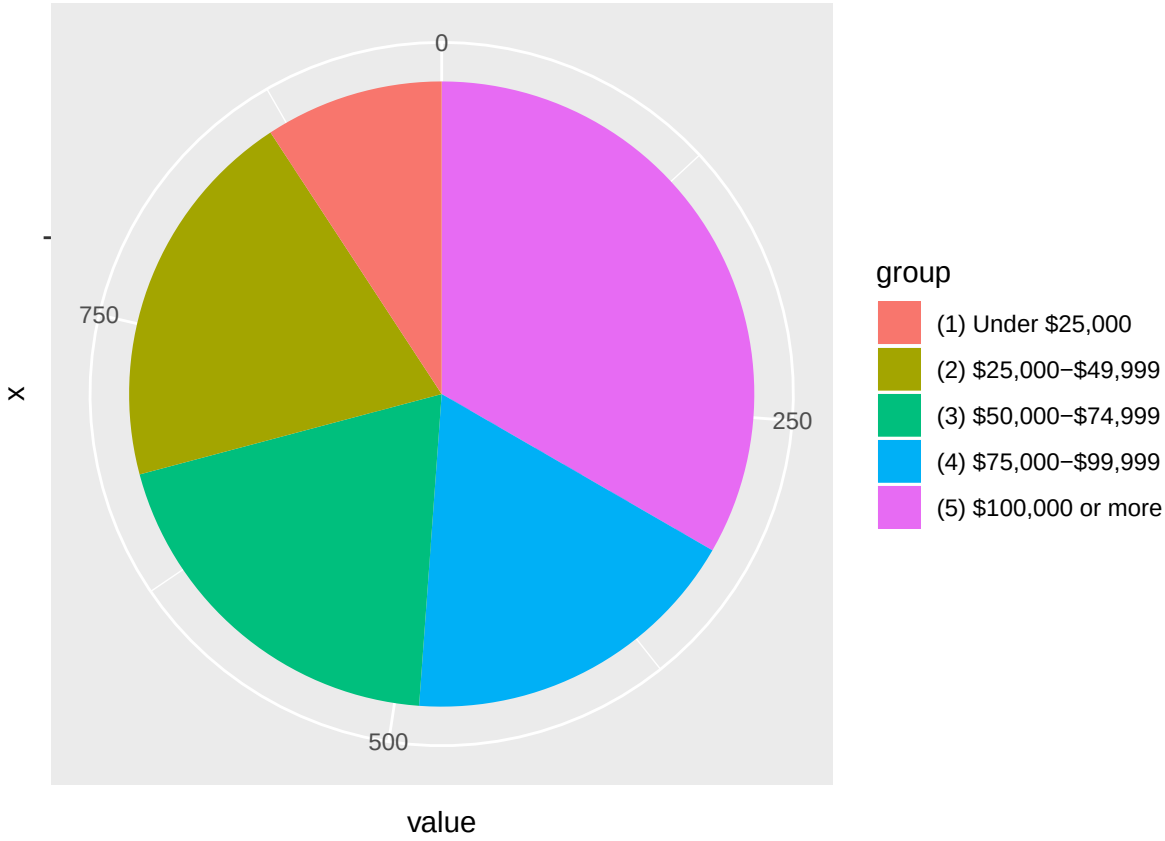

```
data <- data %>% filter(RACETH5<88)
RACE_data <- c()
for (i in 1:5){
  RACE_data[i] <- sum(data$RACETH5==i)
}
lbls <- c("(1) White",
"(2) African American or Black",
"(3) Latino or Hispanic",
"(4) Asian",
"(5) Other")
df <- data.frame(
  group = lbls,
  value = RACE_data
  )
ggplot(df, aes(x="", y=value, fill=group))+ geom_bar(width = 1, stat = "identity")+ coord_polar("y", sta
```



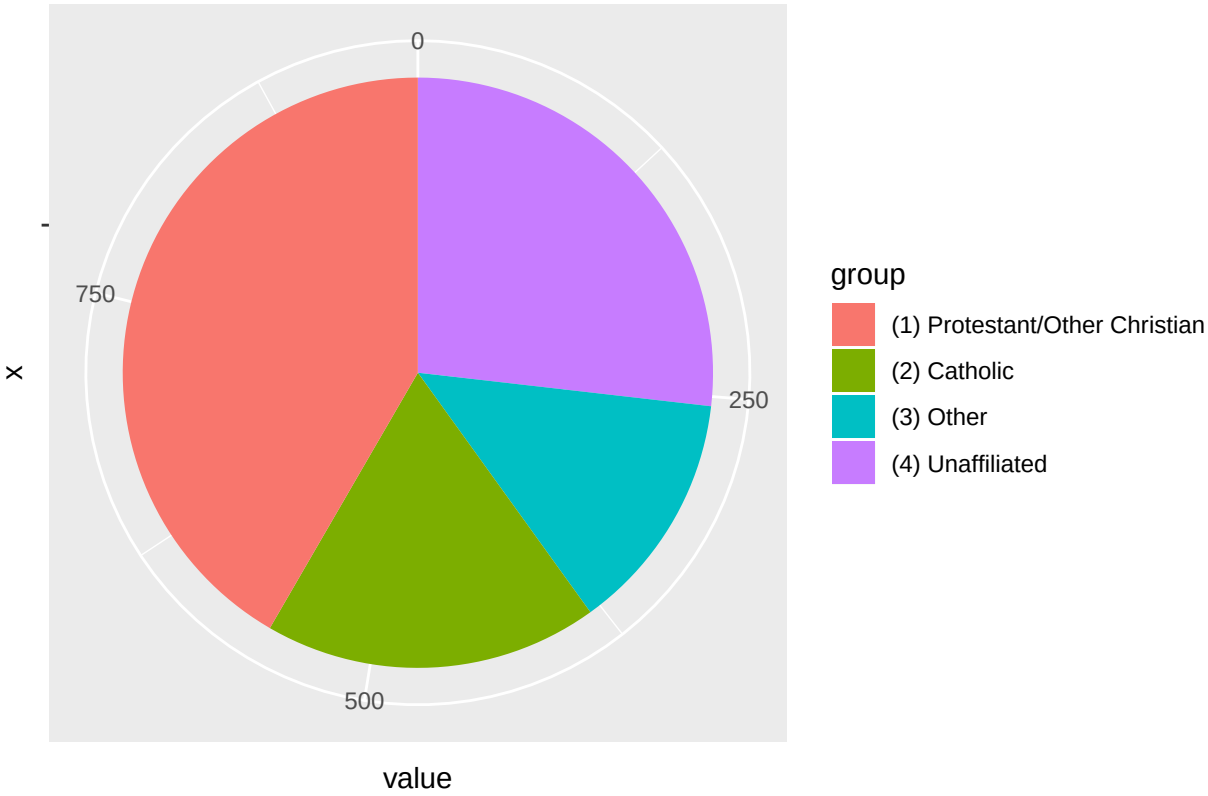

```
data <- data %>% filter(PARTY<88)
PAR_data <- c()
for (i in 1:3){
  PAR_data[i] <- sum(data$PARTY==i)
}
lbls <- c("(1) Democrat/Lean Dem", "(2) Republican/Lean Rep", "(3) Independent")
df <- data.frame(
  group = lbls,
  value = PAR_data
  )
ggplot(df, aes(x="", y=value, fill=group))+ geom_bar(width = 1, stat = "identity")+ coord_polar("y", sta
```




```
data <- data %>% filter(INCOME<88)
INC_data <- c()
for (i in 1:5){
  INC_data[i] <- sum(data$INCOME==i)
}
lbls <- c("(1) Under $25,000",
"(2) $25,000-$49,999",
"(3) $50,000-$74,999",
"(4) $75,000-$99,999",
"(5) $100,000 or more")
df <- data.frame(
  group = lbls,
  value = INC_data
  )
ggplot(df, aes(x="", y=value, fill=group))+ geom_bar(width = 1, stat = "identity")+ coord_polar("y", sta
```




```
data <- data %>% filter(RELIG<88)
REL_data <- c()
for (i in 1:4){
  REL_data[i] <- sum(data$RELIG==i)
}
lbls <- c("(1) Protestant/Other Christian",
"(2) Catholic",
"(3) Other",
"(4) Unaffiliated")
df <- data.frame(
  group = lbls,
  value = REL_data
  )
ggplot(df, aes(x="", y=value, fill=group))+ geom_bar(width = 1, stat = "identity")+ coord_polar("y", st
```




```
table(data$IDEO, data$RACETH5)
```

```
##
##       1   2   3   4   5
##   1 100   6  10   1   9
##   2 199  23  17   9  16
##   3 198  31  28   9  21
##   4 149   4  15   4  11
##   5  78   3   7   0   3
```

```
table(data$IDEO, data$AGE)
```

```
##
##      1  2  3  4  5  6
##   1 15 18 25  8 32 28
##   2 23 38 36 34 70 63
##   3 19 25 53 35 97 58
##   4 16 10 20 33 66 38
##   5  6  3  9 12 29 32
```

```
table(data$IDEO, data$EDUC)
```

```
##
##      1  2  3  4
##   1 11 28 47 40
##   2 19 60 89 96
##   3 46 97 71 73
##   4 21 57 60 45
##   5 17 34 24 16
```

```
table(data$IDEO, data$INCOME)
```

```
##
##      1  2  3  4  5
##   1 13 25 26 16 46
##   2 23 49 65 46 81
##   3 32 63 43 55 94
##   4 13 32 31 35 72
##   5  7 20 22 17 25
```

```
table(data$IDEO, data$RELIG)
```

```
##
##       1   2   3   4
##   1  23  12  25  66
##   2  95  42  40  87
##   3 111  67  39  70
##   4 109  36  15  23
##   5  58  17   7   9
```

The group with the most conservative ideology is exposed to the greatest disclosure risk, across differenet demographic variables.

```
vcons <- data %>% filter(IDEO==5)
table(vcons$EDUC, vcons$RELIG)
```

```
##
##      1  2  3  4
##   1 11  2  0  4
##   2 21  6  4  3
##   3 15  6  2  1
##   4 11  3  1  1
```

```
table(vcons$INCOME, vcons$SEX)
```

```
##
##      1  2
##   1  4  3
##   2 12  8
##   3 13  9
##   4 12  5
##   5 17  8
```

```
table(data$PARTY, data$IDEO)
```

```
##
##       1   2   3   4   5
##   1 122 252 155  18   4
##   2   2   4  69 149  81
##   3   2   8  63  16   6
```

```
modelString <-"
model {
## sampling
for (i in 1:N){
   y[i] ~ dmulti(p[i,1:C],1)
   for (c in 1:C){
     p[i,c] <- q[i,c]/sum(q[i,])
```

```
    log(q[i,c]) <- beta0[c] + beta1[c]*x[i]
   }
}
## priors
beta0[1] <- 0
beta1[1] <- 0
for (c in 2:C){
  beta0[c] ~ dnorm(0, 0.00001)
  beta1[c] ~ dnorm(0, 0.00001)
}
}
"
```

```
y = as.vector(data$IDEO)
x = as.vector(data$INCOME)
N = length(y)
C = 5
```

```
the_data <- list("y" = y,
                 "x" = x,
                 "N" = N,
                 "C" = C)
```

```
initsfunction <- function(chain){
  .RNG.seed <- c(1,2)[chain]
  .RNG.name <- c("base::Super-Duper",
                 "base::Wichmann-Hill")[chain]
  return(list(.RNG.seed=.RNG.seed,
              .RNG.name=.RNG.name))
}
```

The model I want to use resembles the sequential model we learnt in the paper: Age group $f(y_1)$, and follows a Dirichlet-multinomial prior $(p_{i1}, p_{i2}, p_{i3}, p_{i4}, p_{i5} \sim Dirichlet(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ Education level $f(y_2|y_1)$, and follows a multinomial logistic model with multiple explanatory variables (age groups) $log(\frac{p_{ic}}{p_{i1}}) = \beta_{0c} + \beta_{1c}X_{age1} + \cdots + \beta_{5c}X_{age5}$ Income group $f(y_3|y_2, y_1)$, and is an ordered categorical variable Ideology group $f(y_4|y_3, y_2, y_1)$, and follows a multinomial logistic model with multiple explanatory variables