hw2

Part 1

Question 1

```
samples <- sample(1:nrow(conf),
                  nrow(conf)*0.8,
                  replace = FALSE)
train <- conf[samples, ]
test <- conf[-samples, ]
```

Question 2

```
logit <- glm(vote ~ EuclDist2 + qual + strngprs + sameprty,
               data = train,
               family = binomial); summary(logit)
```

```
##
## Call:
## glm(formula = vote ~ EuclDist2 + qual + strngprs + sameprty,
##     family = binomial, data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.13524   0.09643   0.21153   0.41680   2.13229
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.9017     0.2069  -4.358 1.31e-05 ***
## EuclDist2    -4.3780     0.3089 -14.173  < 2e-16 ***
## qual          4.0593     0.2493  16.281  < 2e-16 ***
## strngprs      1.2094     0.1402   8.628  < 2e-16 ***
## sameprty      1.5135     0.1658   9.126  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2380.7  on 3046  degrees of freedom
## Residual deviance: 1474.1  on 3042  degrees of freedom
## AIC: 1484.1
##
## Number of Fisher Scoring iterations: 6
```

```
vote <- test$vote

logit.probs1 <- predict(logit, newdata=test, type="response")

logit.pred1 <- ifelse(logit.probs1 > 0.5, 1, 0)
```

```r
table(logit.pred1, vote)
```

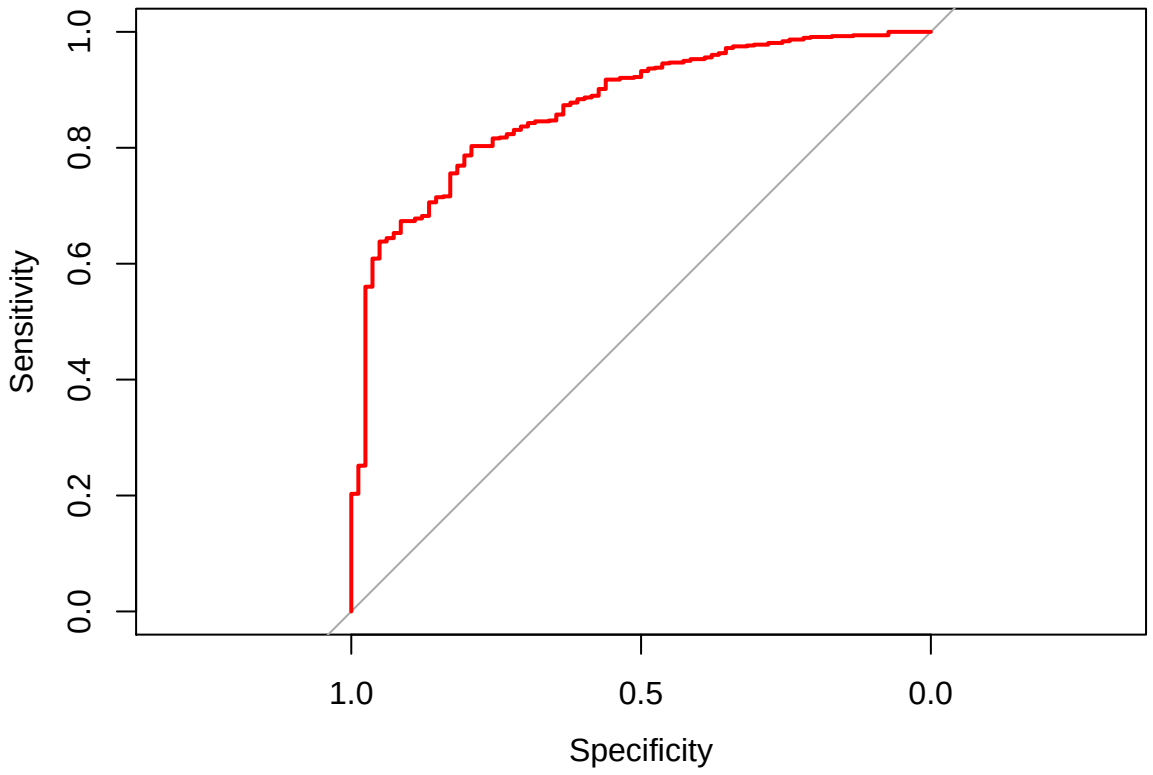
```
##            vote
## logit.pred1   0   1
##           0  29  21
##           1  53 659
```

```r
mean(logit.pred1 == vote)
```

```
## [1] 0.9028871
```

```r
y_logit_roc <- test$vote
pred_logit_roc <- predict(logit,
                  newdata = test,
                  type="response")

plot.roc(y_logit_roc, pred_logit_roc, col = "red")
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```



Question 3

```r
## Linear Discriminant Analysis
library(MASS) # for LDA
```

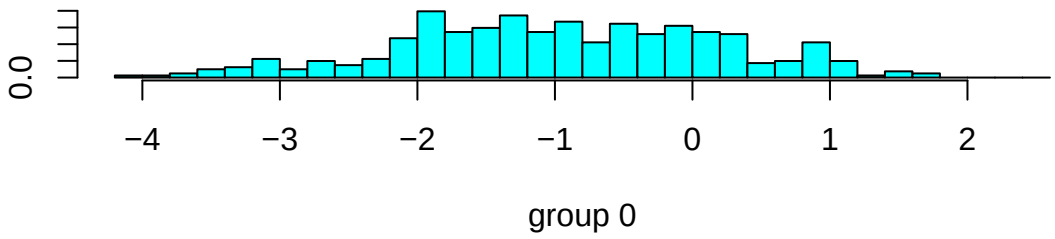
```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

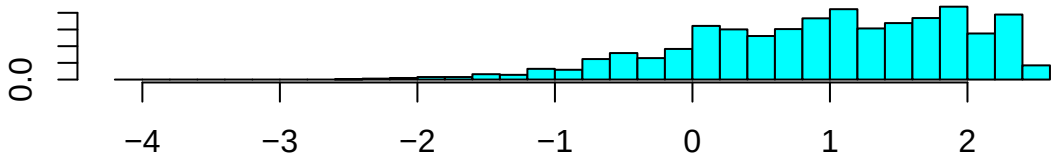
```
lda <- lda(vote ~ EuclDist2 + qual + strngprs + sameprty,
               data=train)

# inspect the model (don't use summary here)
lda
```

```
## Call:
## lda(vote ~ EuclDist2 + qual + strngprs + sameprty, data = train)
##
## Prior probabilities of groups:
##         0         1
## 0.1322612 0.8677388
##
## Group means:
##   EuclDist2      qual  strngprs  sameprty
## 0 0.3995576 0.5641191 0.3101737 0.1662531
## 1 0.1512391 0.8094402 0.6232980 0.6127080
##
## Coefficients of linear discriminants:
##                  LD1
## EuclDist2 -3.2668503
## qual       2.6055455
## strngprs   0.5889113
## sameprty   0.6342449
```

```
plot(lda)
```



group 0



group 1

```
lda.pred <- predict(lda, newdata=test)

data.frame(lda.pred)[1:5,]
```

```
##   class posterior.0 posterior.1        LD1
## 1     1   0.1840942   0.8159058 -0.9098328
## 2     1   0.1553313   0.8446687 -0.8031931
## 3     0   0.5888485   0.4111515 -1.8733775
## 4     1   0.1469283   0.8530717 -0.7690347
## 5     0   0.8616454   0.1383546 -2.6397135
```

```
# predicting are different than posterior becaus we have the distribution for every point of the origin
# confusion matrix
table_lda <- table(lda.pred$class, vote)
table_lda
```

```
##    vote
##       0   1
##   0  33  28
##   1  49 652
```

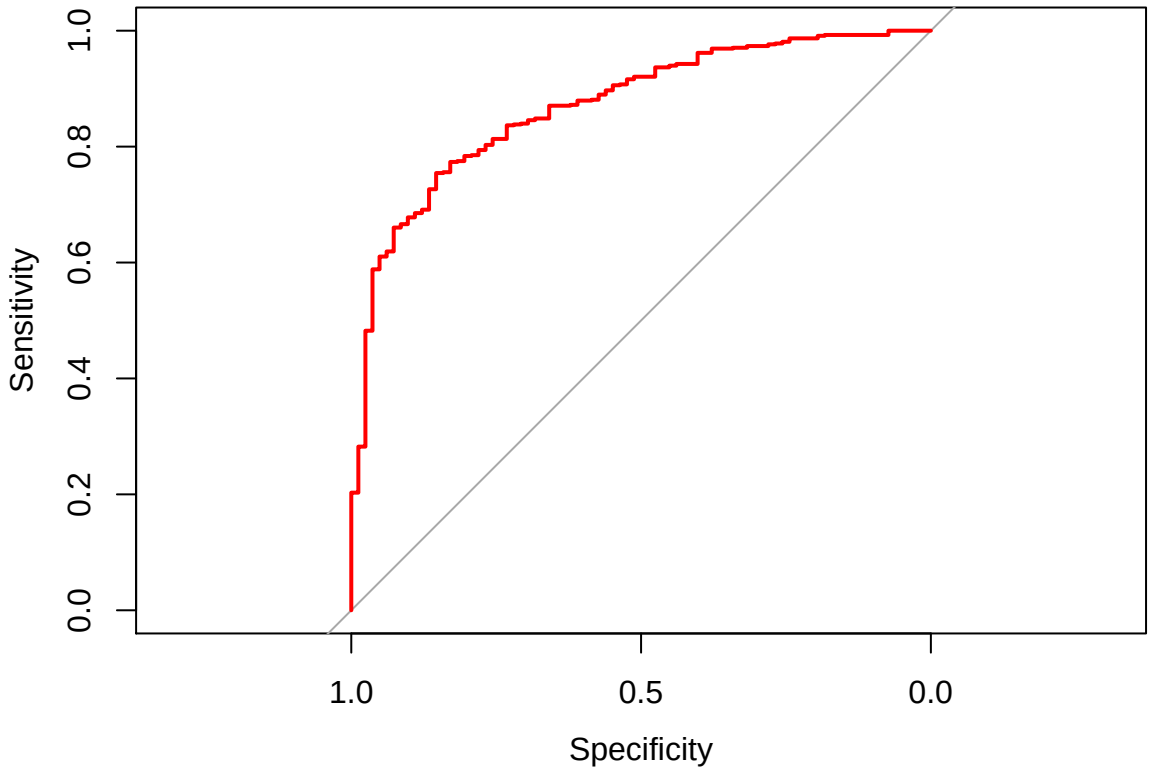
```
# check the classification rate
mean(lda.pred$class ==vote)
```

```
## [1] 0.8989501
```

```
y_lda_roc <- test$vote
pred_lda_roc <- predict(lda,
                 newdata = test,
                 type="response")

plot.roc(y_lda_roc ,pred_lda_roc$posterior[,2], col = "red")
```

```
## Setting levels: control = 0, case = 1
```

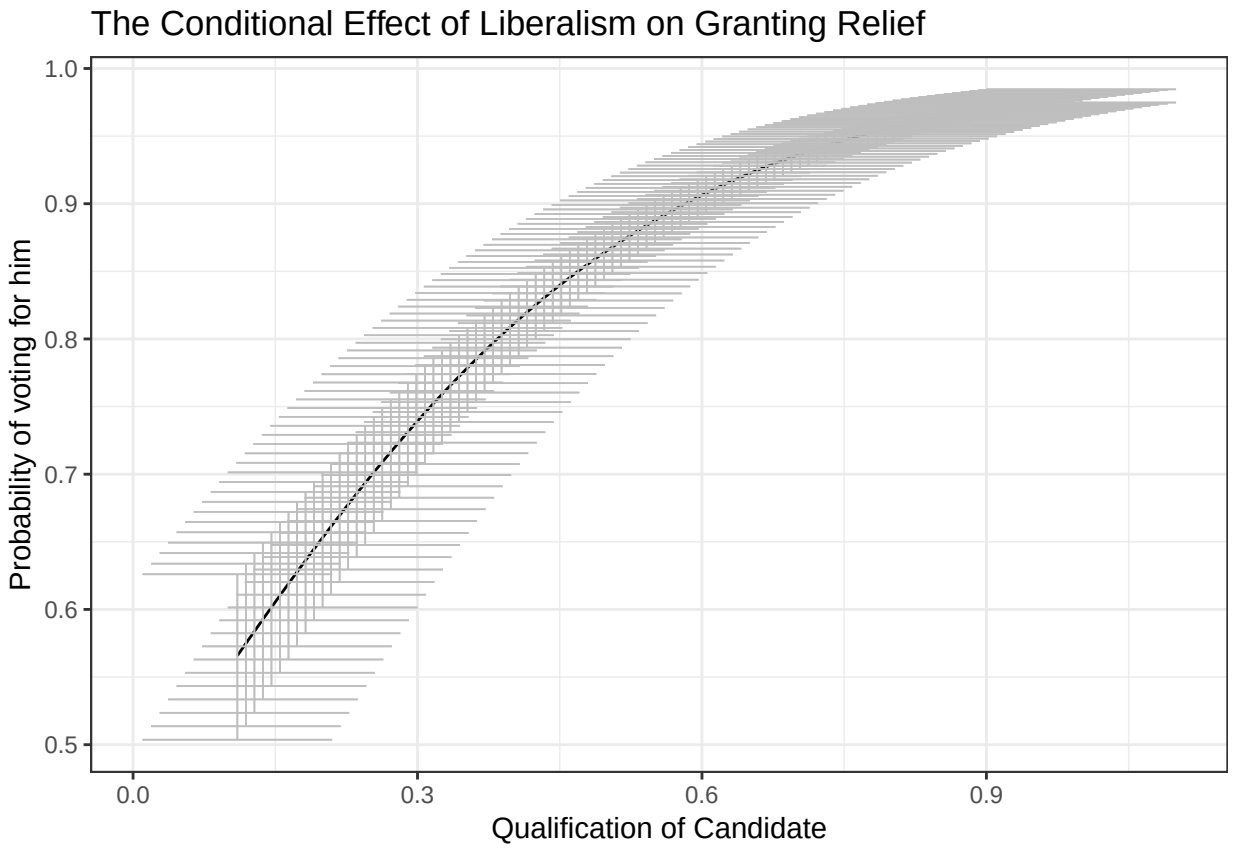
```
## Setting direction: controls < cases
```



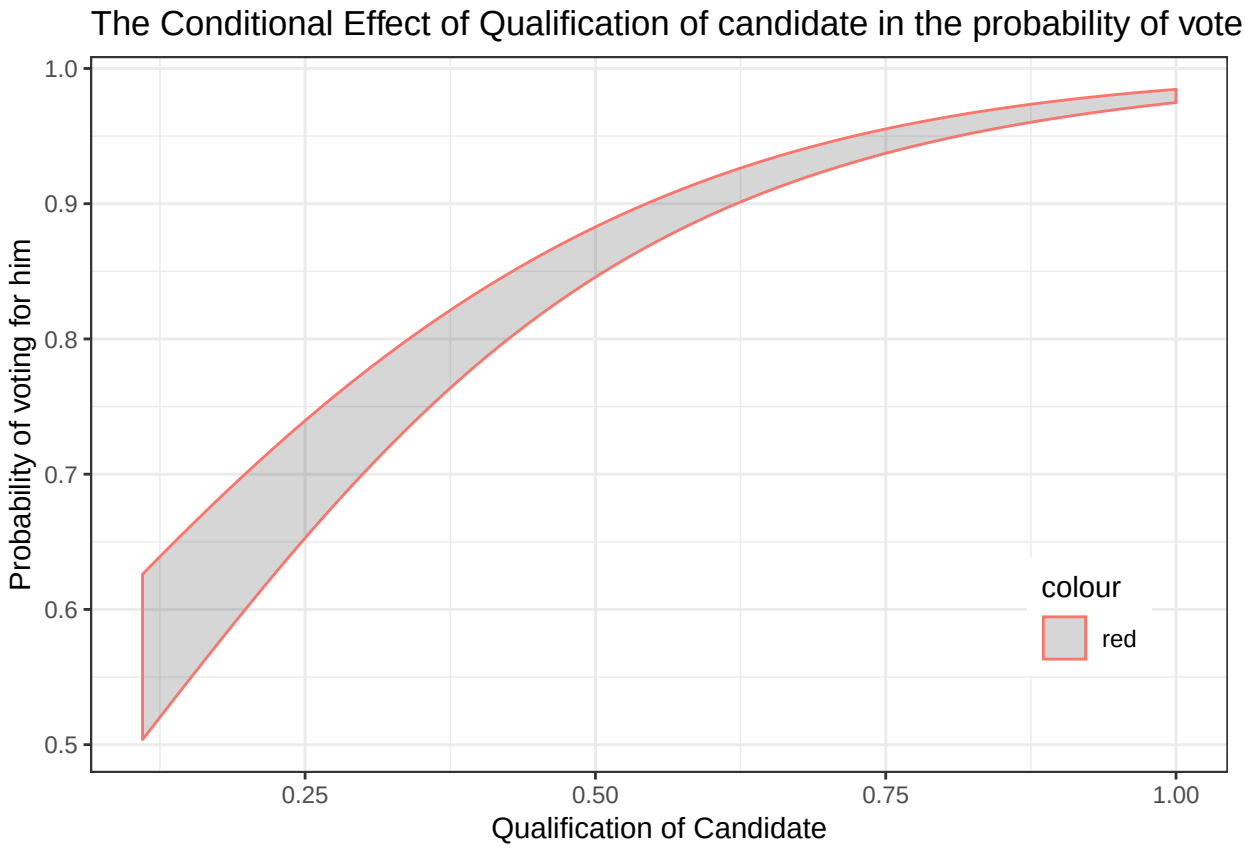
Question 4

```
# CIs for predicted probabilities

#propetit ~ ineffcou + multpet + usparty + liberal2
#vote ~ EuclDist2 + qual + lackqual + strngprs + sameprty

logit2 <- glm(vote ~ EuclDist2 + qual  + strngprs + sameprty,
              data = conf,
              family = binomial(link=logit))


newdata2 <- with(conf, data.frame(qual = rep(seq(from = min(conf$qual) , to = max(conf$qual), length.ou
                                                 2),
                                  EuclDist2 = mean(EuclDist2),
                                  strngprs = mean(strngprs),
                                  sameprty = mean(sameprty)))

newdata3 <- cbind( newdata2,predict(logit2, newdata = newdata2, type = "link",se = TRUE))

# Add CIs
newdata3 <- within(newdata3, {
  PredictedProb <- plogis(fit)
  LL <- plogis(fit - (1.96 * se.fit))
  UL <- plogis(fit + (1.96 * se.fit))
})

# Recode usparty as a factor
#newdata3$usparty <- factor(newdata3$usparty, labels=c("No", "Yes"))

# Plot predictions with CIs
ggplot(newdata3, aes(x = qual, y = PredictedProb)) +
  geom_line() +
  geom_errorbar(aes(ymin = LL, ymax = UL),
                color="gray",
                size=.3,
                width=.2,
                position = position_dodge(.9)) +
  labs(x = "Qualification of Candidate",
       y = "Probability of voting for him"
       )  +
  ggtitle("The Conditional Effect of Liberalism on Granting Relief") +
  theme_bw() +
  theme(legend.justification = c(.7,1),
        legend.position = c(.9,.3))
```

```
## Warning: position_dodge requires non-overlapping x intervals
```




```
ggplot(newdata3, aes(x = qual, y = PredictedProb, color='red')) + geom_ribbon(aes(ymin = LL,
    ymax = UL), alpha = 0.2) +
  labs(x = "Qualification of Candidate",
       y = "Probability of voting for him"
       )  +
  ggtitle("The Conditional Effect of Qualification of candidate in the probability of vote for him") +
  theme_bw() +
  theme(legend.justification = c(.7,1),
        legend.position = c(.9,.3))
```



Question 5

Question 6

Part 2

Question 1

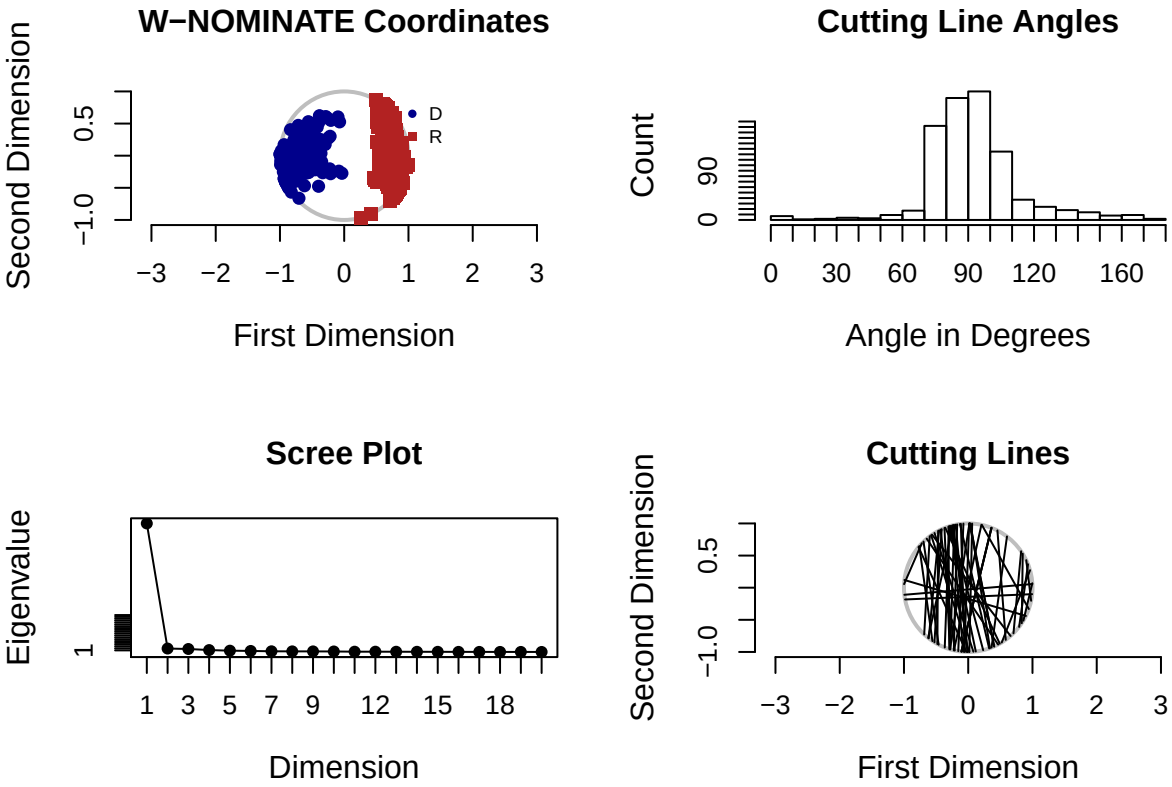
```
## Loading required package: pscl

## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis

##
## ## W-NOMINATE Ideal Point Package

## ## Copyright 2006 -2019

## ## Keith Poole, Jeffrey Lewis, James Lo, and Royce Carroll

## ## Support provided by the U.S. National Science Foundation

## ## NSF Grant SES-0611974
```

```
## Attempting to read file in Keith Poole/Howard Rosenthal (KH) format.
## Attempting to create roll call object
## 113 th_ House _ Roll _ Call _ Data
## 445 legislators and 1202 roll calls
## Frequency counts for vote types:
## rollCallMatrix
##      0      1      6      7      9
##  14576 295753 202943    290  21328
##
## Preparing to run W-NOMINATE...
##
##  Checking data...
##
##      ... 1 of 445 total members dropped.
##
##      Votes dropped:
##      ... 181 of 1202 total votes dropped.
##
##  Running W-NOMINATE...
##
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##
##
## W-NOMINATE estimation completed successfully.
## W-NOMINATE took 173.351 seconds to execute.
##
## Preparing to run W-NOMINATE...
##
##  Checking data...
##
##      ... 1 of 445 total members dropped.
##
##      Votes dropped:
##      ... 181 of 1202 total votes dropped.
##
##  Running W-NOMINATE...
##
##      Getting bill parameters...
```

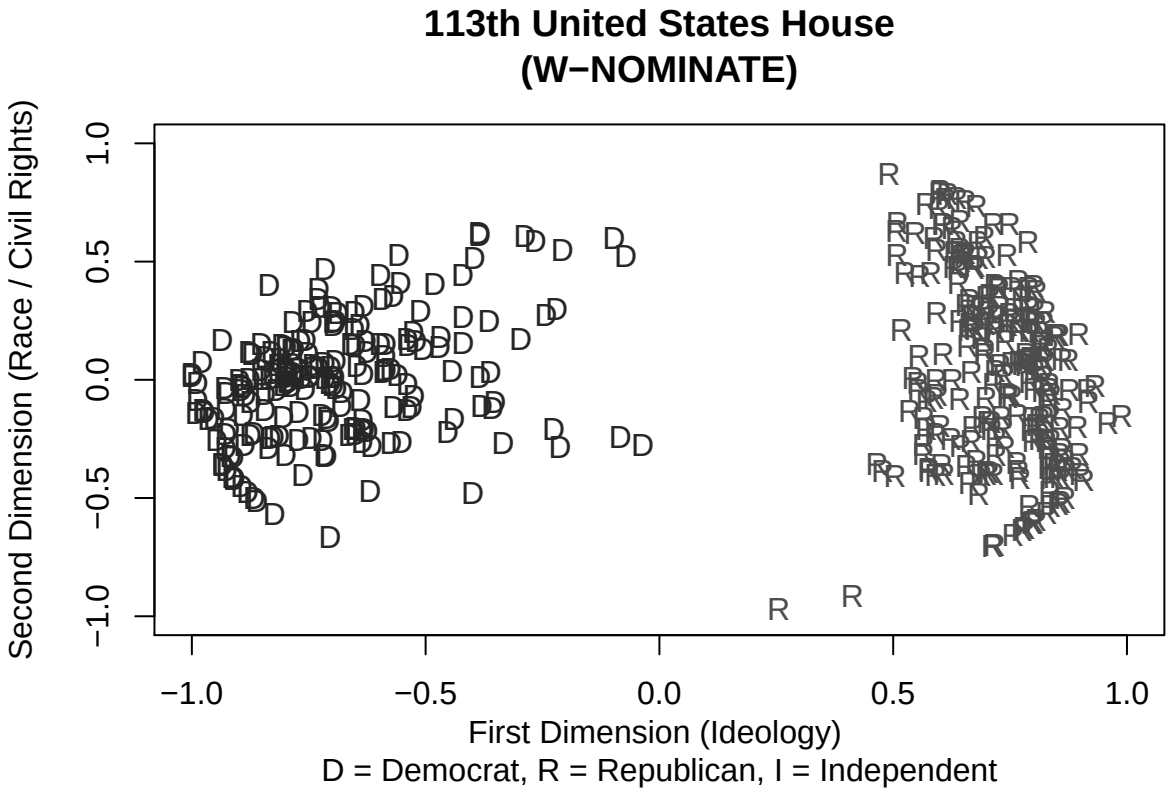
```
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##
##
## W-NOMINATE estimation completed successfully.
## W-NOMINATE took 250.812 seconds to execute.

##
## Preparing to run W-NOMINATE...
##
##  Checking data...
##
##      ... 1 of 445 total members dropped.
##
##      Votes dropped:
##      ... 181 of 1202 total votes dropped.
##
##  Running W-NOMINATE...
##
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
```

```
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##
##
## W-NOMINATE estimation completed successfully.
## W-NOMINATE took 346.128 seconds to execute.
##
## Preparing to run W-NOMINATE...
##
##  Checking data...
##
##      ... 1 of 445 total members dropped.
##
##      Votes dropped:
##      ... 181 of 1202 total votes dropped.
##
##  Running W-NOMINATE...
##
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Starting estimation of Beta...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Getting bill parameters...
##      Getting legislator coordinates...
##      Estimating weights...
```

```
##       Getting bill parameters...
##       Getting legislator coordinates...
##       Estimating weights...
##       Getting bill parameters...
##       Getting legislator coordinates...
##       Getting bill parameters...
##       Getting legislator coordinates...
##       Estimating weights...
##       Getting bill parameters...
##       Getting legislator coordinates...
##       Estimating weights...
##       Getting bill parameters...
##       Getting legislator coordinates...
##       Getting bill parameters...
##       Getting legislator coordinates...
##       Estimating weights...
##       Getting bill parameters...
##       Getting legislator coordinates...
##       Estimating weights...
##       Getting bill parameters...
##       Getting legislator coordinates...
##
##
## W-NOMINATE estimation completed successfully.
## W-NOMINATE took 429.291 seconds to execute.
```




```
## NULL
```



Question 2

Discuss the dimensionality of the space. You can present and inspect fit via the aggregate proportion reduction in errors (APRE), the geometric mean prediction (GMP) rate, scree plots, or any other diagnostic tool (visual or numeric) to inspect the overall fit of the algorithm.

For viewing the dimmensionality of the space to fit the data, we can see the eigen values of the fitting of the data.Examining the eigen values we see that the first four are over one. Meaningn that maybe we can see that we can explain this data with four dimmension. Nervertheless, we can see that the first value is 69, meaning that most of the variance is explianed by

Eigenvalues

```
## [1] 69.2931290  1.9777323  1.7397527  1.1627728  0.8384978  0.6917516
```

Question 3