Intra-Party Factions with DW-NOMINATE (119th Congress)

Group 1

2025-11-29

```r
library(tidyverse)
library(cluster)
library(factoextra)   # fviz_nbclust helpers
library(ggrepel)
set.seed(params$seed)
knitr::opts_chunk$set(comment = NA)
```

Introduction

This project analyzes ideological patterns in the 119th U.S. Congress (2025–2027) using NOMINATE data from Voteview. By examining how members are positioned along key ideological dimensions, we aim to better understand the structure of both major parties.

Overall, our goal is to evaluate how congressional members cluster ideologically, how demographic factors such as age relate to their positions, and how consistently they align with their party in voting behavior. These analyses help reveal both internal party dynamics and broader patterns of polarization.

Data Description

Our dataset comes from Voteview, which provides ideology estimates for every voting member of Congress based on their roll-call behavior. We focus on the 119th Congress (2025–2027) and use each member's NOMINATE scores, a widely used measure of legislative ideology. The first dimension (nominate_dim1) places members on a liberal–conservative scale, while the second dimension (nominate_dim2) captures additional ideological variation often interpreted as establishment vs. populist tendencies.

In addition to ideology scores, the dataset includes each member's party affiliation, state and district, birth year, and identifying information such as name and ICPSR ID. These variables allow us to examine ideological clustering across parties, study demographic patterns such as age, and measure partisan voting behavior. Overall, the dataset provides a structured and quantitative foundation for analyzing congressional ideology and internal party dynamics.

Methodology

Data Analysis

Intra-Party Factions

This report identifies **intra-party factions** within the 119th Congress using **DW-NOMINATE** coordinates. We cluster members **within each party** using k-means on (`Nominate_dim1, Nominate_dim2`) and validate with elbow and silhouette diagnostics.

- `Nominate_dim1`: liberal $(-)$ conservative $(+)$

- `Nominate_dim2`: often captures establishment $(+)$ vs. populist $(-)$ tendencies (modern era)

Load & Prepare Data

```r
infile <- params$infile
which_chamber <- params$which_chamber
k_grid <- params$k_min:params$k_max

df_raw <- readr::read_csv(infile, show_col_types = FALSE)

# Basic sanity check for expected columns
needed <- c("chamber","party_code","bioname","nominate_dim1","nominate_dim2")
missing <- setdiff(needed, names(df_raw))
if (length(missing) > 0) {
  stop(paste("Missing required columns:", paste(missing, collapse=", ")))
}

df <- df_raw %>%
  filter(chamber == which_chamber) %>%
  select(bioname, party_code, chamber, nominate_dim1, nominate_dim2) %>%
  filter(!is.na(nominate_dim1), !is.na(nominate_dim2))

df_dem <- df %>% filter(party_code == 100)   # Democrats
df_gop <- df %>% filter(party_code == 200)   # Republicans

tibble(
  party = c("Democrats","Republicans"),
  N = c(nrow(df_dem), nrow(df_gop))
)
```

```
# A tibble: 2 x 2
  party          N
  <chr>      <int>
1 Democrats    219
2 Republicans  224
```

Method: Within-Party Clustering

We choose k using **elbow (WSS)** and **average silhouette** over `k_grid`. Then we fit k-means and visualize a 2D cluster map.

```r
fit_party_clusters <- function(df_party, party_label, k_grid = 2:5) {
  if (nrow(df_party) < min(k_grid)) {
    warning(paste0("Too few members in ", party_label, " for k>=2. Skipping."))
    return(NULL)
  }
  X <- df_party %>% select(nominate_dim1, nominate_dim2)

  # Elbow
  wss <- sapply(k_grid, function(k){
    km <- kmeans(X, centers = k, nstart = 50)
    km$tot.withinss
  })

  # Silhouette
  sil_avgs <- sapply(k_grid, function(k) {
    km <- kmeans(X, centers = k, nstart = 50)
    sil <- silhouette(km$cluster, dist(X))
    mean(sil[, "sil_width"])
  })

  #k_best <- k_grid[which.max(sil_avgs)]
  k_best <- 3
  km <- kmeans(X, centers = k_best, nstart = 200)
  df_out <- df_party %>% mutate(cluster = factor(km$cluster))

  centers <- as_tibble(km$centers) %>% mutate(cluster = factor(1:n()))

  p_elbow <- tibble(k = k_grid, wss = wss) %>%
    ggplot(aes(k, wss)) +
    geom_line() + geom_point() +
    labs(title = paste0(party_label, " - Elbow (WSS)"),
         x = "k", y = "Total within-cluster SS") +
    theme_minimal()

  p_sil <- tibble(k = k_grid, silhouette = sil_avgs) %>%
    ggplot(aes(k, silhouette)) +
    geom_line() + geom_point() +
    labs(title = paste0(party_label, " - Average Silhouette by k"),
         x = "k", y = "Average silhouette width") +
    theme_minimal()

  p_map <- df_out %>%
  ggplot(aes(nominate_dim1, nominate_dim2, color = cluster)) +
  geom_point(size = 3, alpha = 0.9) +  # larger, clear points
  # Mark cluster centers with an X
  geom_point(data = centers, aes(nominate_dim1, nominate_dim2, color = cluster),
             size = 6, shape = 4, stroke = 1.5, inherit.aes = FALSE) +
  labs(
    title = paste0(party_label, " - NOMINATE Cluster Map (k=", k_best, ")"),
```

```r
      subtitle = "Each point = Member of Congress; Color = Faction cluster center (X)",
      x = "NOMINATE dim1 (Liberal - Conservative)",
      y = "NOMINATE dim2 (Populist - Establishment)",
      color = "Faction Cluster"
    ) +
    theme_minimal(base_size = 13) +
    theme(
      plot.title = element_text(face = "bold"),
      legend.position = "right",
      legend.title = element_text(size = 12),
      legend.text = element_text(size = 10)
    )


  list(
    df = df_out,
    centers = centers,
    k_best = k_best,
    silhouette_table = tibble(k = k_grid, avg_silhouette = sil_avgs),
    elbow_plot = p_elbow,
    silhouette_plot = p_sil,
    map_plot = p_map
  )
}


# Heuristic labeling rules - tweak after inspecting centers
label_dem_cluster <- function(center_dim1, center_dim2) {
  if (center_dim1 <= -0.55) return("Progressive")
  if (center_dim1 >= -0.30 && center_dim2 >= 0.05) return("Blue Dog / Moderate-Establishment")
  return("Mainline Moderate")
}


label_gop_cluster <- function(center_dim1, center_dim2) {
  if (center_dim1 >= 0.60 && center_dim2 <= 0.00) return("Trump-aligned / Populist")
  if (center_dim1 <= 0.35 && center_dim2 >= 0.05) return("Moderate / Establishment")
  return("Traditional Conservative")
}


apply_labels <- function(res, party = c("D","R")) {
  party <- match.arg(party)
  if (is.null(res)) return(NULL)
  centers_labeled <- res$centers %>%
    rowwise() %>%
    mutate(
      faction = if (party == "D")
        label_dem_cluster(nominate_dim1, nominate_dim2)
      else
        label_gop_cluster(nominate_dim1, nominate_dim2)
    ) %>%
    ungroup()
  df_labeled <- res$df %>%
    left_join(centers_labeled %>% select(cluster, faction), by = "cluster")
  list(members = df_labeled, centers = centers_labeled)
```

```
}
```
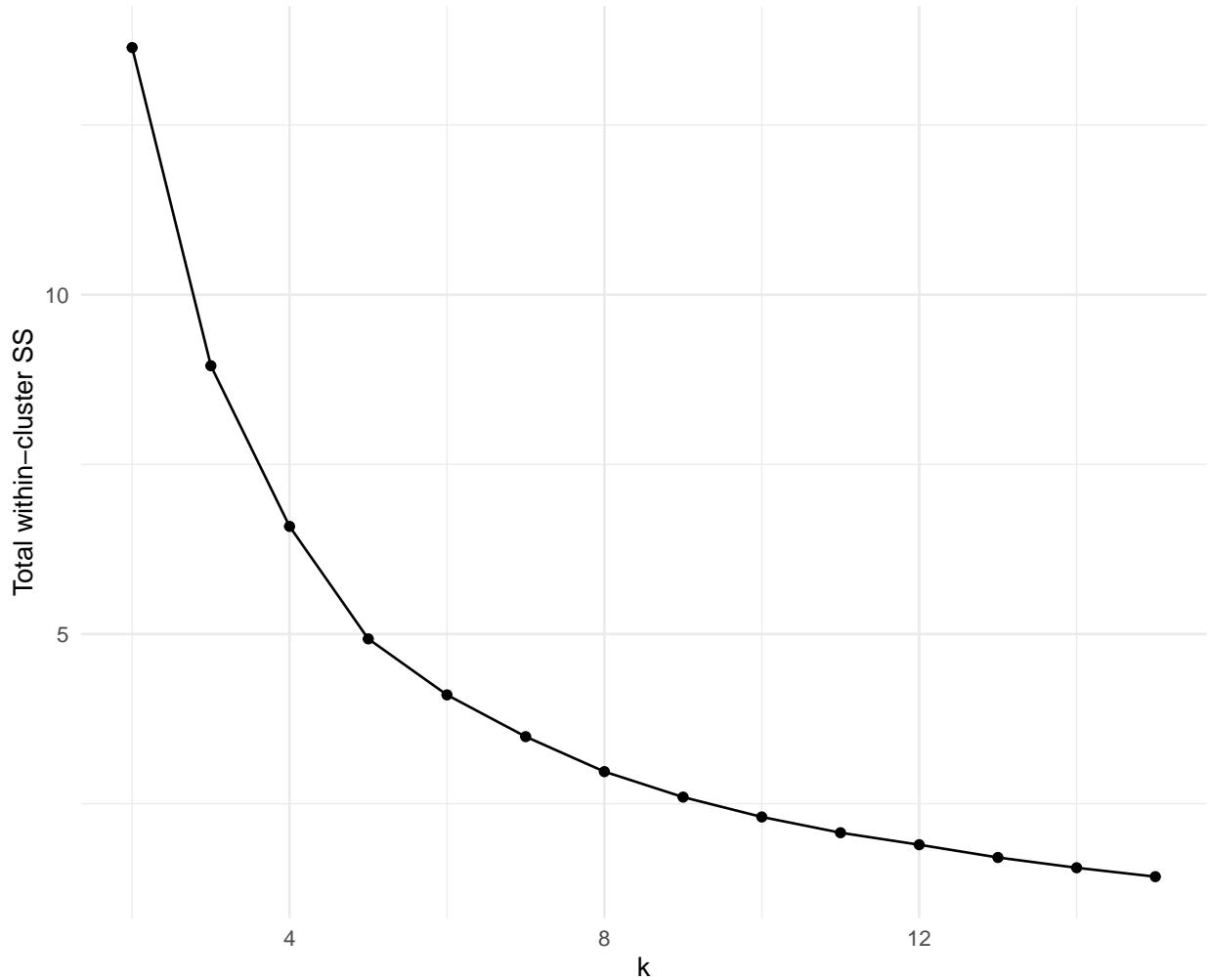
Democrats

```
res_dem <- fit_party_clusters(df_dem, "Democrats", k_grid)
if (!is.null(res_dem)) {
  print(res_dem$elbow_plot)
}
```
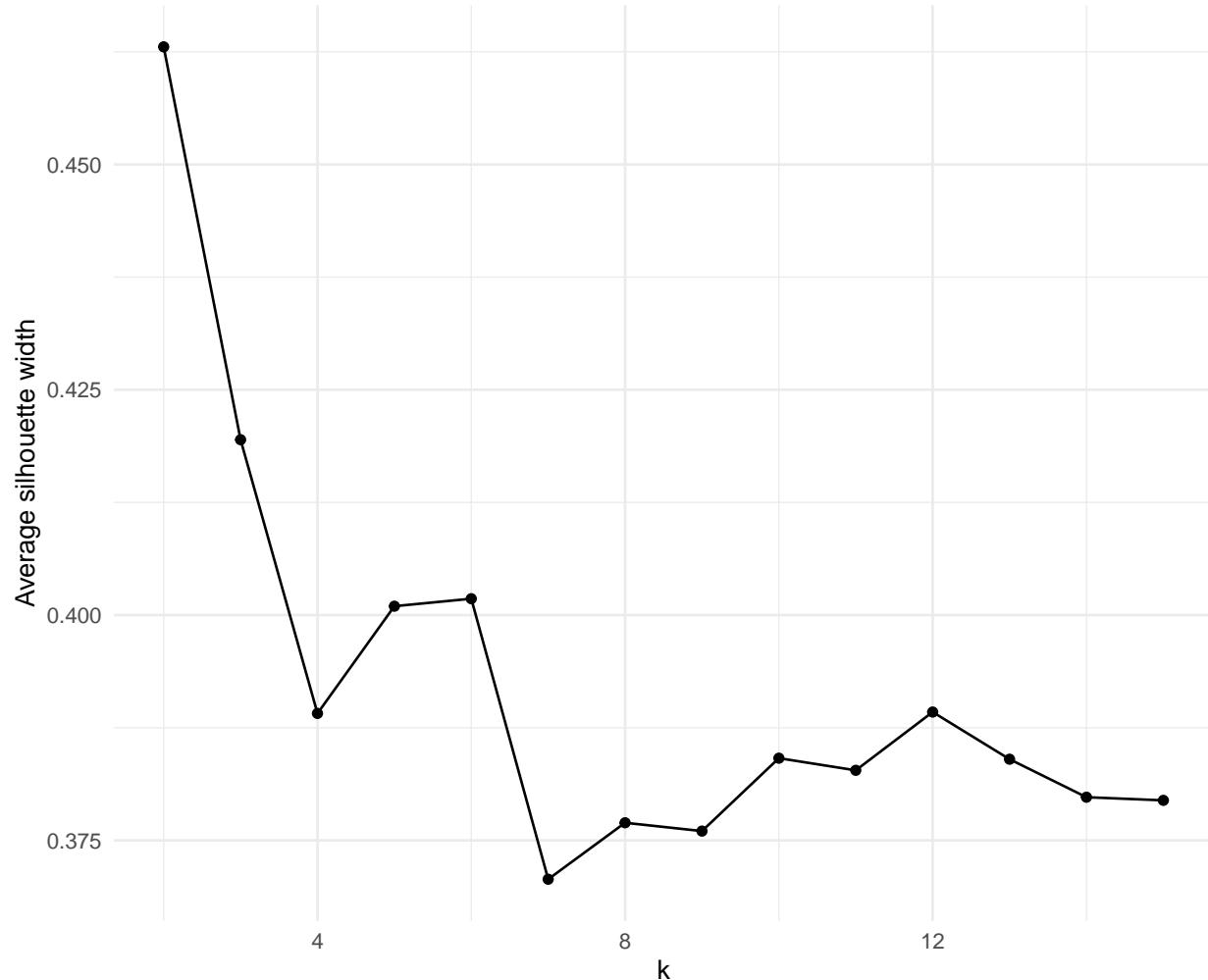
Democrats – Elbow (WSS)

Since we start at k = 2, the WSS drops a lot when moving from 2 to 3 clusters(), drops a bit more at 3 to 4, and after that the curve flattens out — adding more clusters doesn't really help much.

```
if (!is.null(res_dem)) {
  print(res_dem$silhouette_plot)
}
```

Democrats – Average Silhouette by k
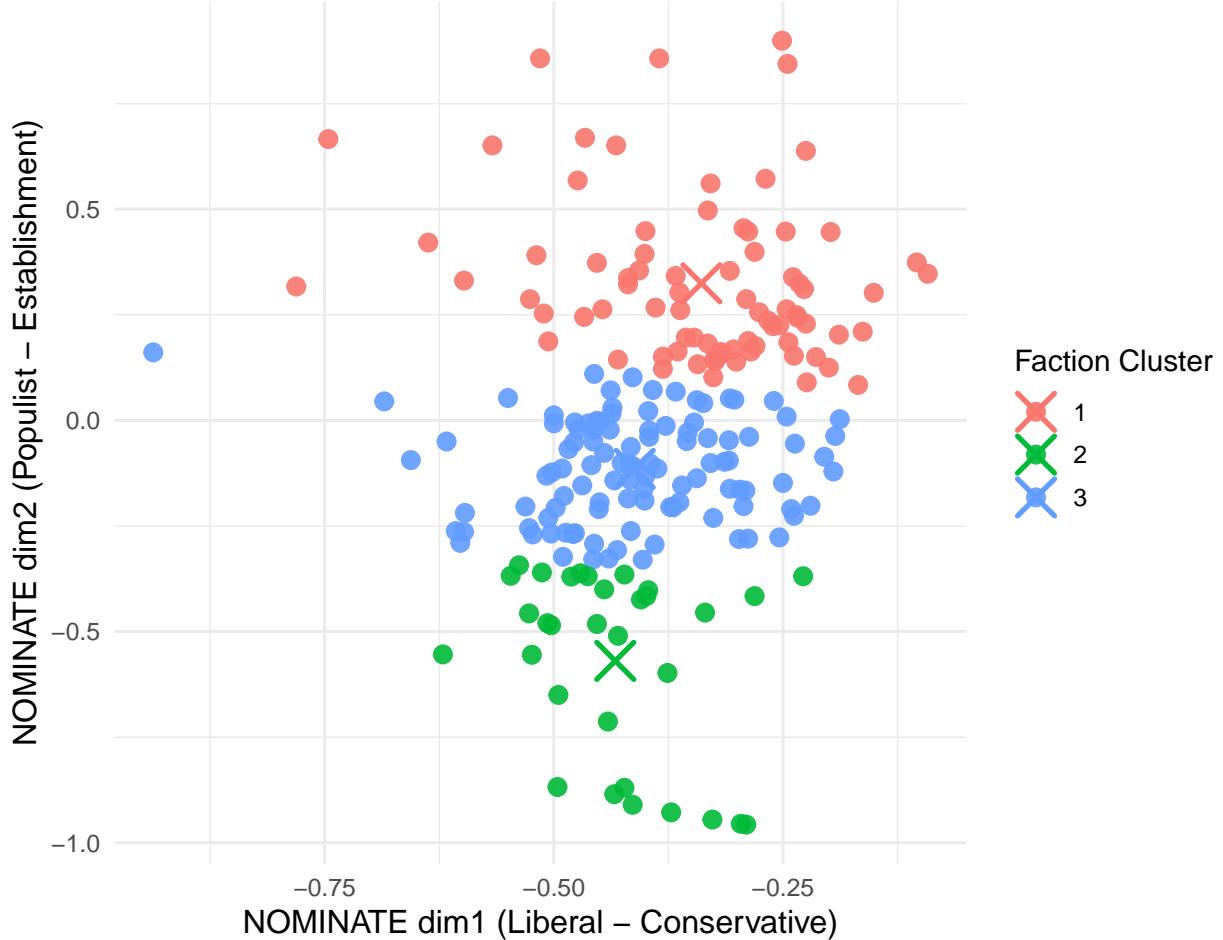

The silhouette score is highest at k = 2, then it drops at k = 3 and 4, and after that it stays pretty flat with only small ups and downs. This basically means the clusters don't get much clearer beyond 3–4 groups — adding more clusters doesn't really improve separation.

```r
if (!is.null(res_dem)) {
  print(res_dem$map_plot)
}
```

Democrats – NOMINATE Cluster Map (k=3)

Each point = Member of Congress; Color = Faction cluster center (X)

Although the Democrats were split into three clusters in the k-means map (k = 3), this choice was made mainly to stay consistent with the research question. In fact, the silhouette plot shows that k = 2 has the highest average silhouette score, meaning two clusters provide the clearest separation according to the underlying NOMINATE distances.

When we examine the numerical cluster centers, the three Democratic clusters are extremely close to each other. All three centers fall within a very narrow range.

As a result, the labeling rules classify all three Democratic clusters as "Mainline Moderate," indicating that the algorithm's three-way split does not correspond to three meaningfully distinct ideological factions within the caucus.

```
if (!is.null(res_dem)) {
  out_dem <- apply_labels(res_dem, "D")

  cat("## Cluster Centers (Democrats)
")
  print(out_dem$centers %>% arrange(faction))
  "~/Documents/GitHub/Stat_380_Final_Proj/HS119_members.csv"
  cat("## Counts by Faction (Democrats)
")
```

```
  print(out_dem$members %>% count(faction, sort = TRUE))
}
```

```
## Cluster Centers (Democrats)
# A tibble: 3 x 4
  nominate_dim1 nominate_dim2 cluster faction
          <dbl>         <dbl> <fct>   <chr>
1        -0.339         0.325 1       Mainline Moderate
2        -0.433        -0.569 2       Mainline Moderate
3        -0.412        -0.115 3       Mainline Moderate
## Counts by Faction (Democrats)
# A tibble: 1 x 2
  faction                n
  <chr>              <int>
1 Mainline Moderate    219
```
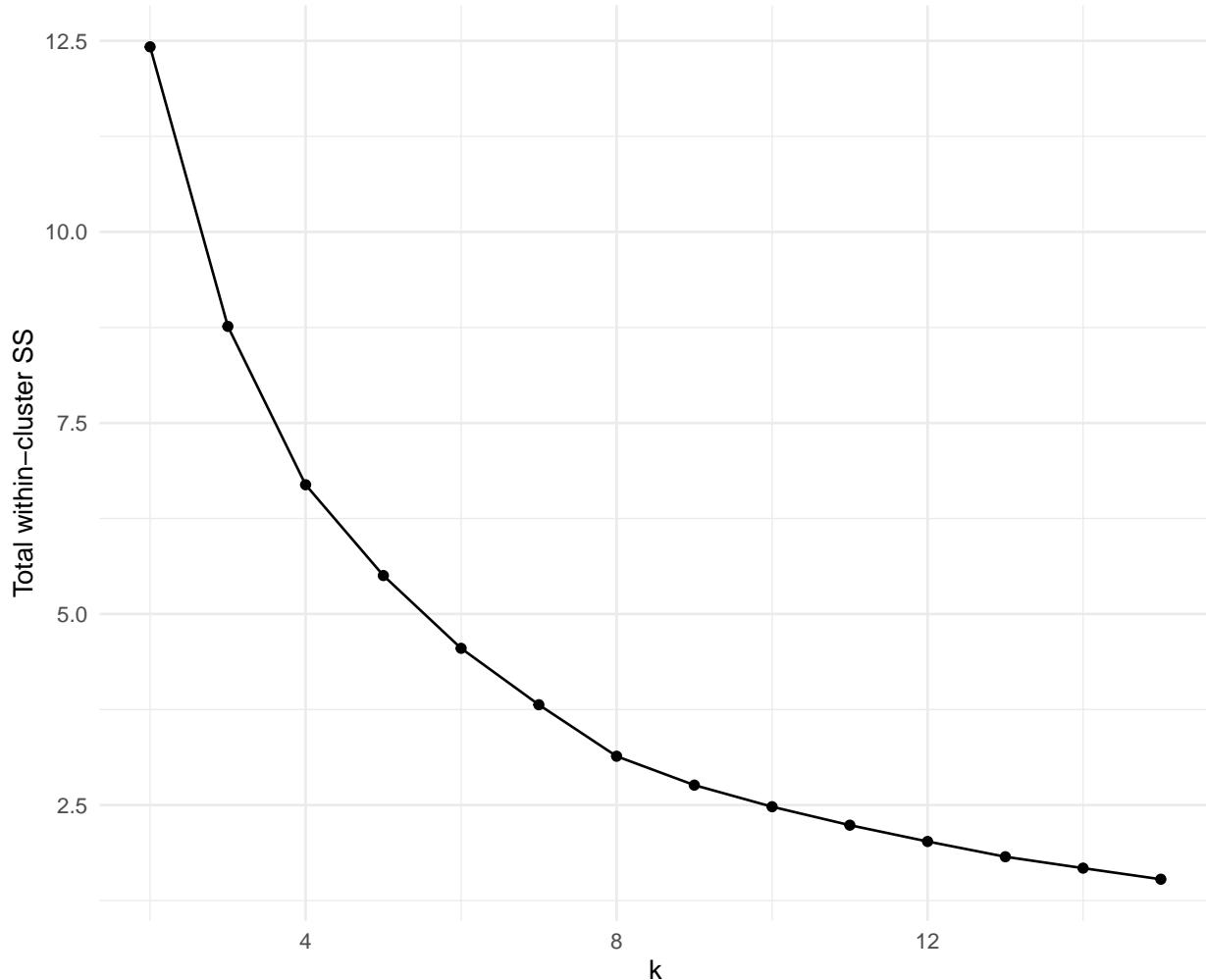
Republicans

```
res_gop <- fit_party_clusters(df_gop, "Republicans", k_grid)
if (!is.null(res_gop)) {
  print(res_gop$elbow_plot)
}
```
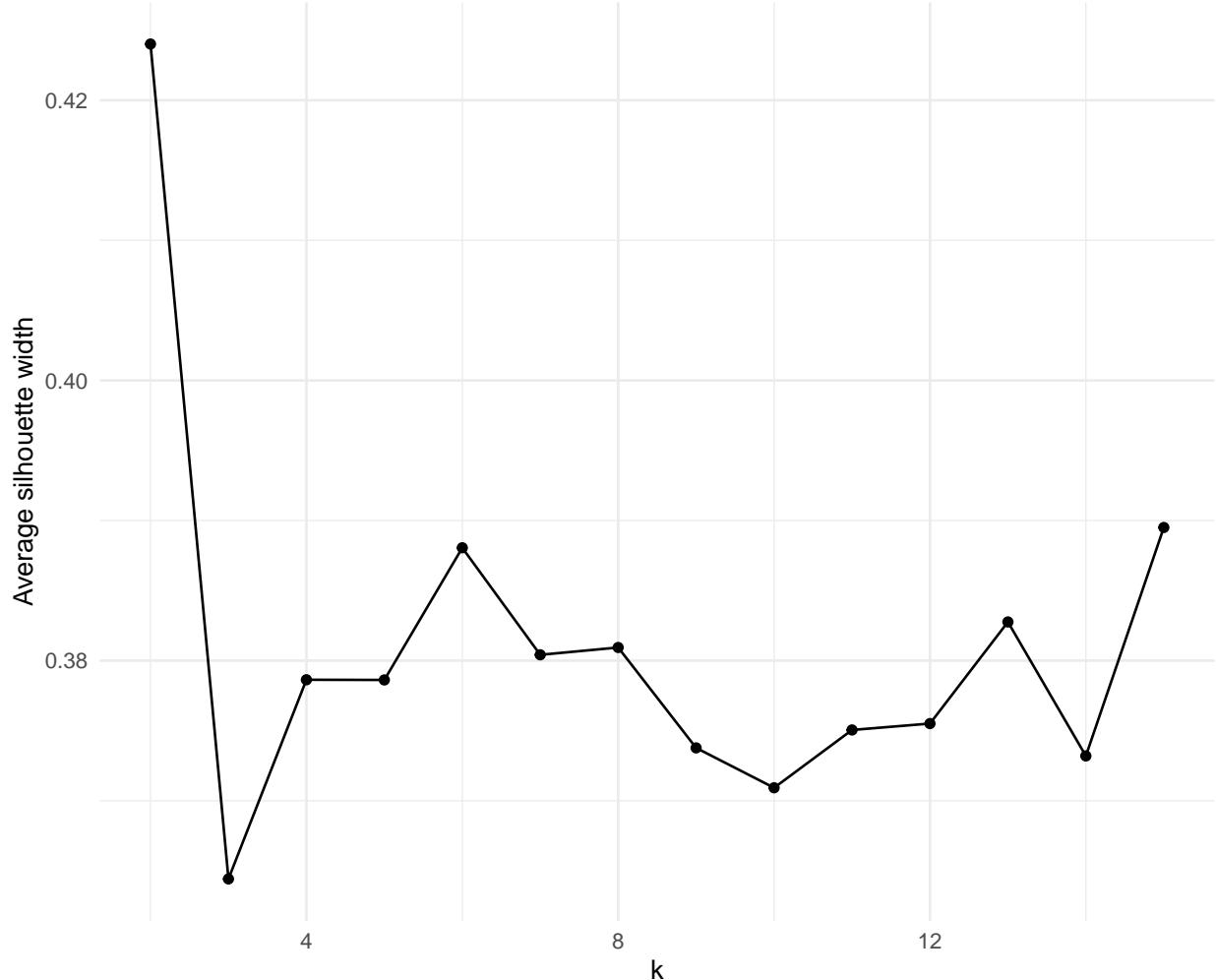
Republicans – Elbow (WSS)

The WSS curve shows its steepest drop from k = 2 to k = 3, meaning that's where most of the improvement happens. After that, the decrease becomes more gradual, with a softer decline from k = 4 to 8, and then it flattens even further from k = 8 to 15, indicating diminishing returns as more clusters are added.

```r
if (!is.null(res_gop)) {
  print(res_gop$silhouette_plot)
}
```

Republicans – Average Silhouette by k
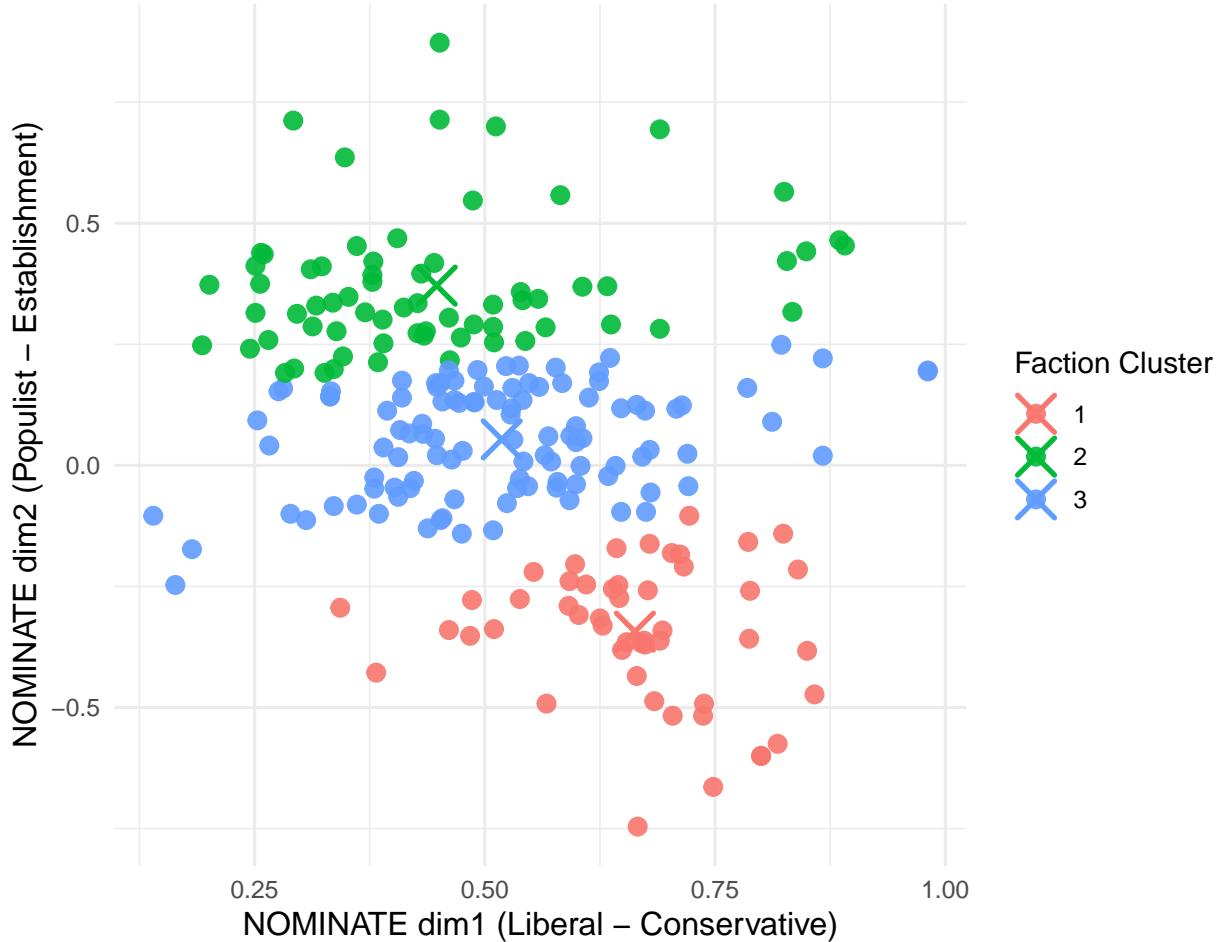
The silhouette score peaks at k = 2, then drops sharply at k = 3 and stays noticeably lower for all larger k. From k = 4 through about k = 12, the silhouette values move within a tight band with only small ups and downs, showing no clear improvement as more clusters are added. The slight rise at k = 15 doesn't change the overall pattern: the best separation occurs at k = 2, and adding more clusters mostly weakens the clustering structure rather than strengthening it.

```
if (!is.null(res_gop)) {
  print(res_gop$map_plot)
}
```

Republicans – NOMINATE Cluster Map (k=3)

Each point = Member of Congress; Color = Faction cluster center (X)

With k = 3, the Republican NOMINATE map visually splits into three clusters, but the numerical results show that two of the clusters sit extremely close to each other in the ideological space and both represent the same mainstream "Traditional Conservative" group. The third cluster—located lower on the second dimension—is the only one that clearly separates, corresponding to a "Trump-aligned / Populist" faction. So even though the algorithm technically identifies three groups, the substantive interpretation is that Republicans largely divide into one major establishment bloc and a smaller Trump-aligned faction, with the two establishment clusters reflecting minor within-group variation rather than distinct ideological camps.

```
if (!is.null(res_gop)) {
  out_gop <- apply_labels(res_gop, "R")

  cat("## Cluster Centers (Republicans)
")
  print(out_gop$centers %>% arrange(faction))

  cat("## Counts by Faction (Republicans)
")
  print(out_gop$members %>% count(faction, sort = TRUE))
}
```

```
## Cluster Centers (Republicans)
# A tibble: 3 x 4
  nominate_dim1 nominate_dim2 cluster faction
          <dbl>         <dbl> <fct>   <chr>
1         0.448        0.371  2       Traditional Conservative
2         0.518        0.0537 3       Traditional Conservative
3         0.663       -0.343  1       Trump-aligned / Populist
## Counts by Faction (Republicans)
# A tibble: 2 x 2
  faction                     n
  <chr>                   <int>
1 Traditional Conservative  174
2 Trump-aligned / Populist   50
```

Members Closest to Each Faction Center

```
closest_to_centers <- function(df_members, centers_tbl, top_n = 8) {
  if (is.null(df_members)) return(invisible(NULL))
  out <- list()
  for (cl in centers_tbl$cluster) {
    cen <- centers_tbl %>% filter(cluster == cl) %>%
      select(nominate_dim1, nominate_dim2) %>% as.numeric()
    tmp <- df_members %>%
      filter(cluster == cl) %>%
      mutate(dist_to_center = sqrt((nominate_dim1 - cen[1])^2 + (nominate_dim2 - cen[2])^2)) %>%
      arrange(dist_to_center) %>%
      slice_head(n = top_n) %>%
      select(bioname, nominate_dim1, nominate_dim2, faction, dist_to_center)
    out[[as.character(cl)]] <- tmp
  }
  out
}

if (exists("out_dem") && !is.null(out_dem)) {
  cat("## Democrats - Examples Closest to Centers
")
  print(closest_to_centers(out_dem$members, out_dem$centers, 10))
}
```

```
## Democrats - Examples Closest to Centers
$`1`
# A tibble: 10 x 5
   bioname              nominate_dim1 nominate_dim2 faction     dist_to_center
   <chr>                        <dbl>         <dbl> <chr>                <dbl>
 1 McBATH, Lucy                -0.363         0.303 Mainline M~         0.0322
 2 PETTERSEN, Brittany         -0.367         0.342 Mainline M~         0.0330
 3 MRVAN, Frank J.             -0.308         0.354 Mainline M~         0.0428
 4 MCCLAIN DELANEY, April      -0.29          0.287 Mainline M~         0.0617
 5 KENNEDY, Timothy M.         -0.362         0.261 Mainline M~         0.0675
 6 VEASEY, Marc                -0.407         0.355 Mainline M~         0.0745
 7 STEVENS, Haley              -0.389         0.267 Mainline M~         0.0762
 8 LATIMER, George             -0.419         0.322 Mainline M~         0.0800
```

```
 9 NORCROSS, Donald              -0.419         0.337 Mainline M~        0.0810
10 STANTON, Greg                 -0.276         0.256 Mainline M~        0.0931
```

$`2`
```
# A tibble: 10 x 5
   bioname              nominate_dim1 nominate_dim2 faction    dist_to_center
   <chr>                        <dbl>         <dbl> <chr>               <dbl>
 1 HUFFMAN, Jared               -0.43         -0.51  Mainline Mo~       0.0595
 2 GARCIA, Robert               -0.376        -0.598 Mainline Mo~       0.0637
 3 BALINT, Becca                -0.453        -0.482 Mainline Mo~       0.0897
 4 POCAN, Mark                  -0.524        -0.555 Mainline Mo~       0.0922
 5 DEXTER, Maxine               -0.495        -0.65  Mainline Mo~       0.102
 6 WALKINSHAW, James R.         -0.503        -0.485 Mainline Mo~       0.110
 7 NADLER, Jerrold Lewis        -0.507        -0.48  Mainline Mo~       0.116
 8 FROST, Maxwell               -0.441        -0.713 Mainline Mo~       0.144
 9 KAMLAGER-DOVE, Sydney        -0.527        -0.457 Mainline Mo~       0.147
10 JACOBS, Sara                 -0.405        -0.424 Mainline Mo~       0.148
```

$`3`
```
# A tibble: 10 x 5
   bioname                    nominate_dim1 nominate_dim2 faction dist_to_center
   <chr>                              <dbl>         <dbl> <chr>            <dbl>
 1 DeLAURO, Rosa L.                  -0.414        -0.108 Mainli~        0.00717
 2 NEAL, Richard Edmund              -0.417        -0.106 Mainli~        0.0103
 3 NEGUSE, Joseph                    -0.426        -0.101 Mainli~        0.0198
 4 CASTRO, Joaquin                   -0.4          -0.131 Mainli~        0.0200
 5 BEYER, Donald Sternoff Jr.        -0.395        -0.102 Mainli~        0.0211
 6 FRIEDMAN, Laura                   -0.387        -0.114 Mainli~        0.0248
 7 TRAHAN, Lori                      -0.417        -0.14  Mainli~        0.0257
 8 MATSUI, Doris                     -0.434        -0.142 Mainli~        0.0351
 9 BROWN, Shontel  M.                -0.459        -0.106 Mainli~        0.0481
10 PALLONE, Frank, Jr.               -0.402        -0.162 Mainli~        0.0482
```

```r
if (exists("out_gop") && !is.null(out_gop)) {
  cat("## Republicans - Examples Closest to Centers
")
  print(closest_to_centers(out_gop$members, out_gop$centers, 10))
}
```

```
## Republicans - Examples Closest to Centers
$`1`
# A tibble: 10 x 5
   bioname              nominate_dim1 nominate_dim2 faction      dist_to_center
   <chr>                        <dbl>         <dbl> <chr>                 <dbl>
 1 MCGUIRE, John J., III        0.673        -0.362 Trump-align~         0.0212
 2 DAVIDSON, Warren             0.654        -0.365 Trump-align~         0.0235
 3 McCLINTOCK, Tom              0.67         -0.367 Trump-align~         0.0247
 4 HARRIS, Andy                 0.674        -0.37  Trump-align~         0.0289
 5 STEUBE, William              0.693        -0.341 Trump-align~         0.0301
 6 CLOUD, Michael               0.69         -0.362 Trump-align~         0.0329
 7 CRANK, Jeff                  0.628        -0.331 Trump-align~         0.0371
 8 HAGEMAN, Harriet             0.649        -0.381 Trump-align~         0.0402
 9 HIGGINS, Clay                0.625        -0.316 Trump-align~         0.0468
10 MILLS, Cory                  0.602        -0.309 Trump-align~         0.0700
```

```
$`2`
# A tibble: 10 x 5
   bioname                  nominate_dim1 nominate_dim2 faction dist_to_center
   <chr>                            <dbl>         <dbl> <chr>            <dbl>
 1 FLEISCHMANN, Chuck               0.431         0.396 Tradit~         0.0299
 2 LETLOW, Julia                    0.427         0.335 Tradit~         0.0418
 3 BAIRD, James                     0.445         0.418 Tradit~         0.0468
 4 BICE, Stephanie I.               0.412         0.326 Tradit~         0.0577
 5 CRAWFORD, Rick                   0.461         0.305 Tradit~         0.0676
 6 ROGERS, Mike Dennis              0.378         0.379 Tradit~         0.0702
 7 DUNN, Neal Patrick               0.509         0.332 Tradit~         0.0728
 8 KELLY, Mike                      0.378         0.393 Tradit~         0.0731
 9 EDWARDS, Chuck                   0.379         0.421 Tradit~         0.0849
10 BARR, Garland H. (Andy) IV       0.488         0.291 Tradit~         0.0898

$`3`
# A tibble: 10 x 5
   bioname              nominate_dim1 nominate_dim2 faction         dist_to_center
   <chr>                        <dbl>         <dbl> <chr>                    <dbl>
 1 HUIZENGA, Bill               0.531         0.053 Traditional Co~         0.0127
 2 ISSA, Darrell                0.476         0.03  Traditional Co~         0.0485
 3 ALFORD, Mark                 0.569         0.06  Traditional Co~         0.0510
 4 LaMALFA, Doug                0.542         0.008 Traditional Co~         0.0515
 5 MILLER, Carol                0.528         0.105 Traditional Co~         0.0522
 6 JOHNSON, Mike                0.565         0.021 Traditional Co~         0.0570
 7 GUEST, Michael               0.529         0.118 Traditional Co~         0.0652
 8 TENNEY, Claudia              0.464         0.012 Traditional Co~         0.0685
 9 MOORE, Tim                   0.572         0.008 Traditional Co~         0.0705
10 RESCHENTHALER, Guy           0.446         0.055 Traditional Co~         0.0724
```

Summary

There is no clear point in either WSS plot where the total within-cluster sum of squares levels off; instead, both show a gradually flattening pattern as k increases. Silhouette scores above 0.7 indicate strong clustering and scores above 0.5 indicate reasonable clustering, yet for both Republicans and Democrats, the highest silhouette value occurs at k = 2 and remains below 0.5. This suggests that neither party exhibits a meaningful clustering structure. Although k = 3 was chosen to match the requirements of the research question, the more appropriate choice based on the metrics would be k = 2. Furthermore, the resulting NOMINATE dimension analysis shows that the ideological separation is not very distinct: while we expected three ideological levels for both parties, the actual clustering reveals fewer than three clearly differentiated ideological groups.