

# Project 2: SVD Analysis of Partisanship in U.S. Congress

Weiqiao Shen

2/17/2021

## Introductory

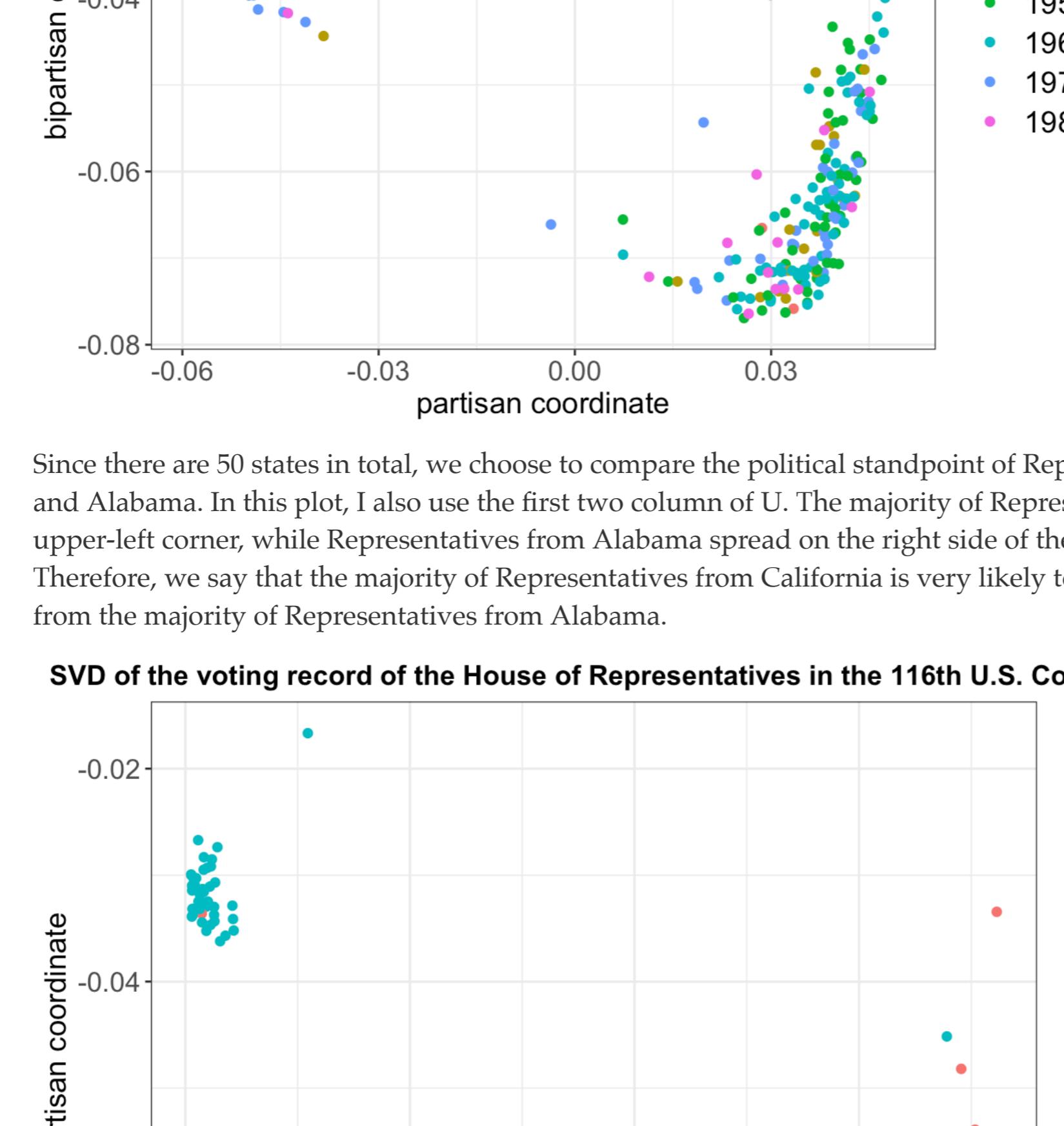
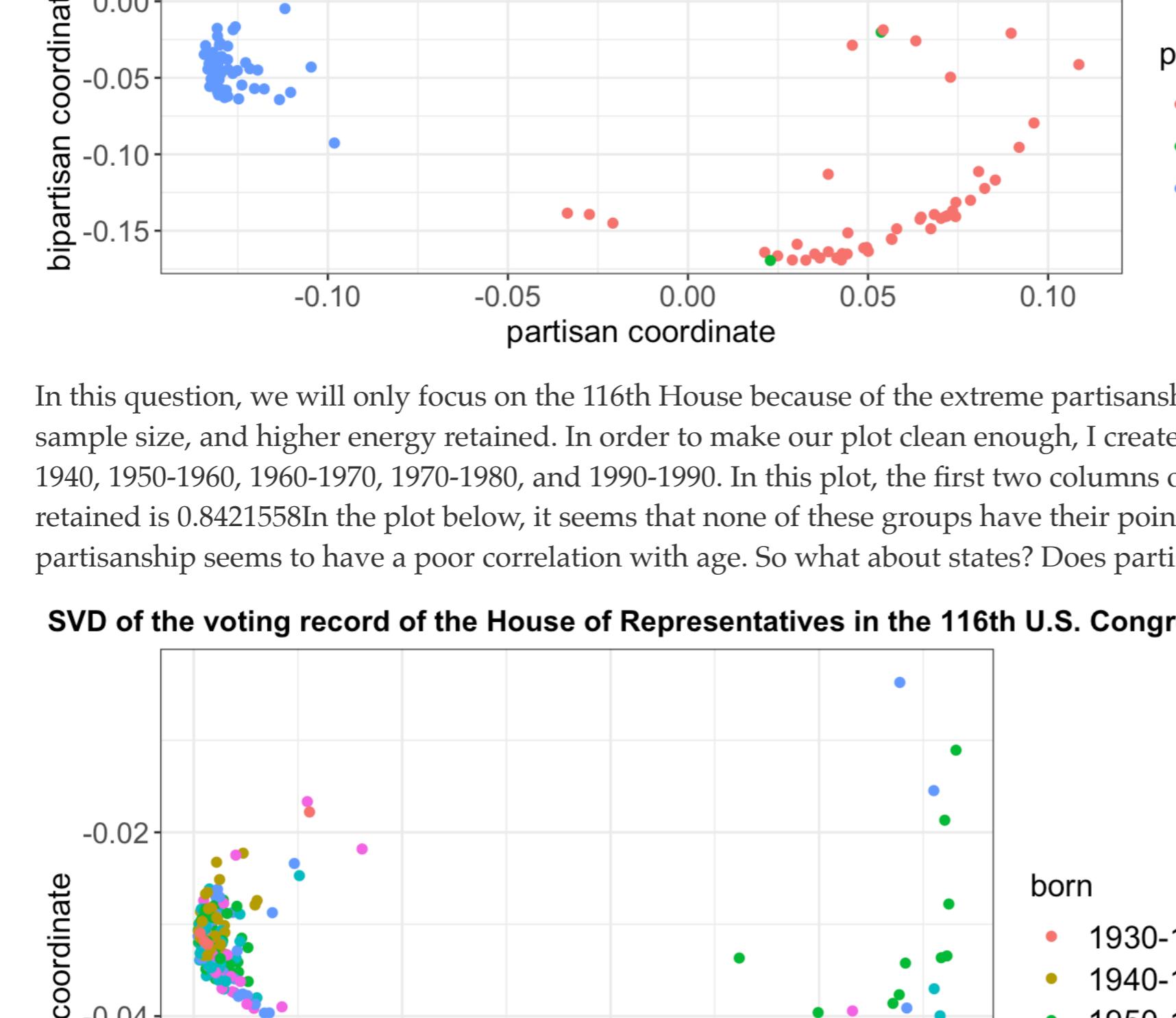
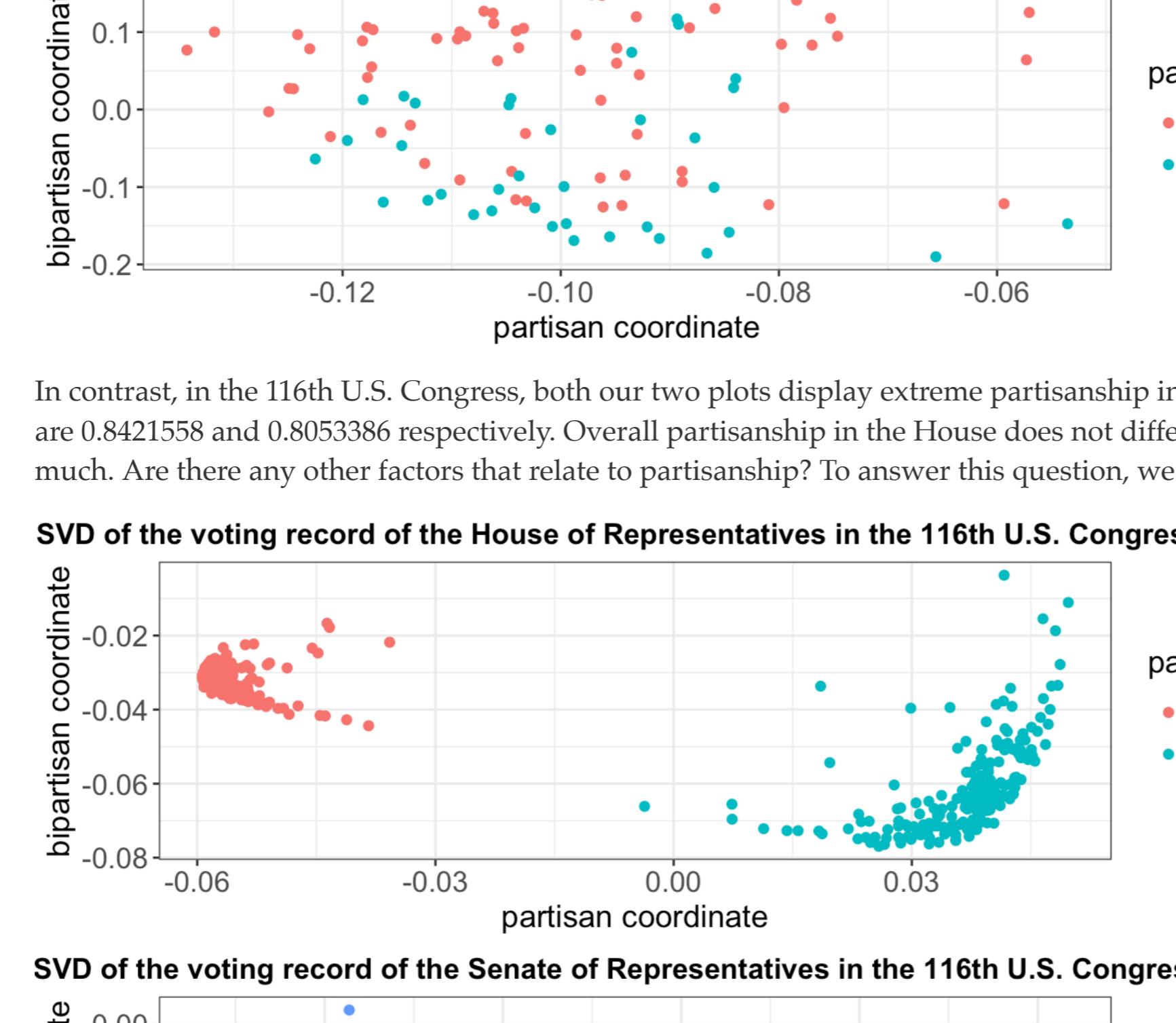
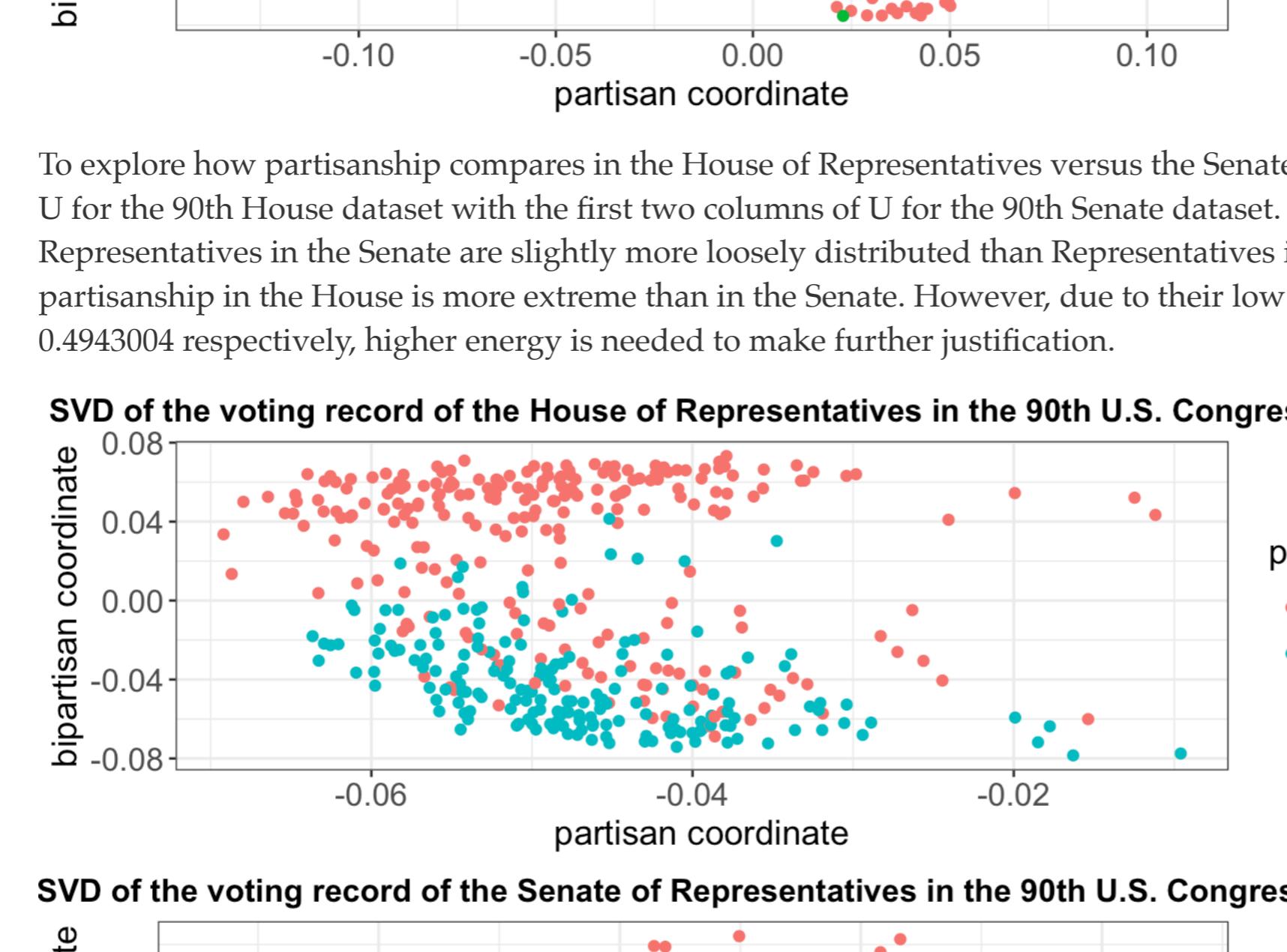
In this report, we will explore the 90th and 116th U.S. Congress voting data to investigate the partisanship in the U.S. congress. Our analysis is simply based on roll-call votes using singular value decomposition (SVD) without any other political information. To make our data workable and avoid extra bias, I removed rows with missing data.

## Methodology

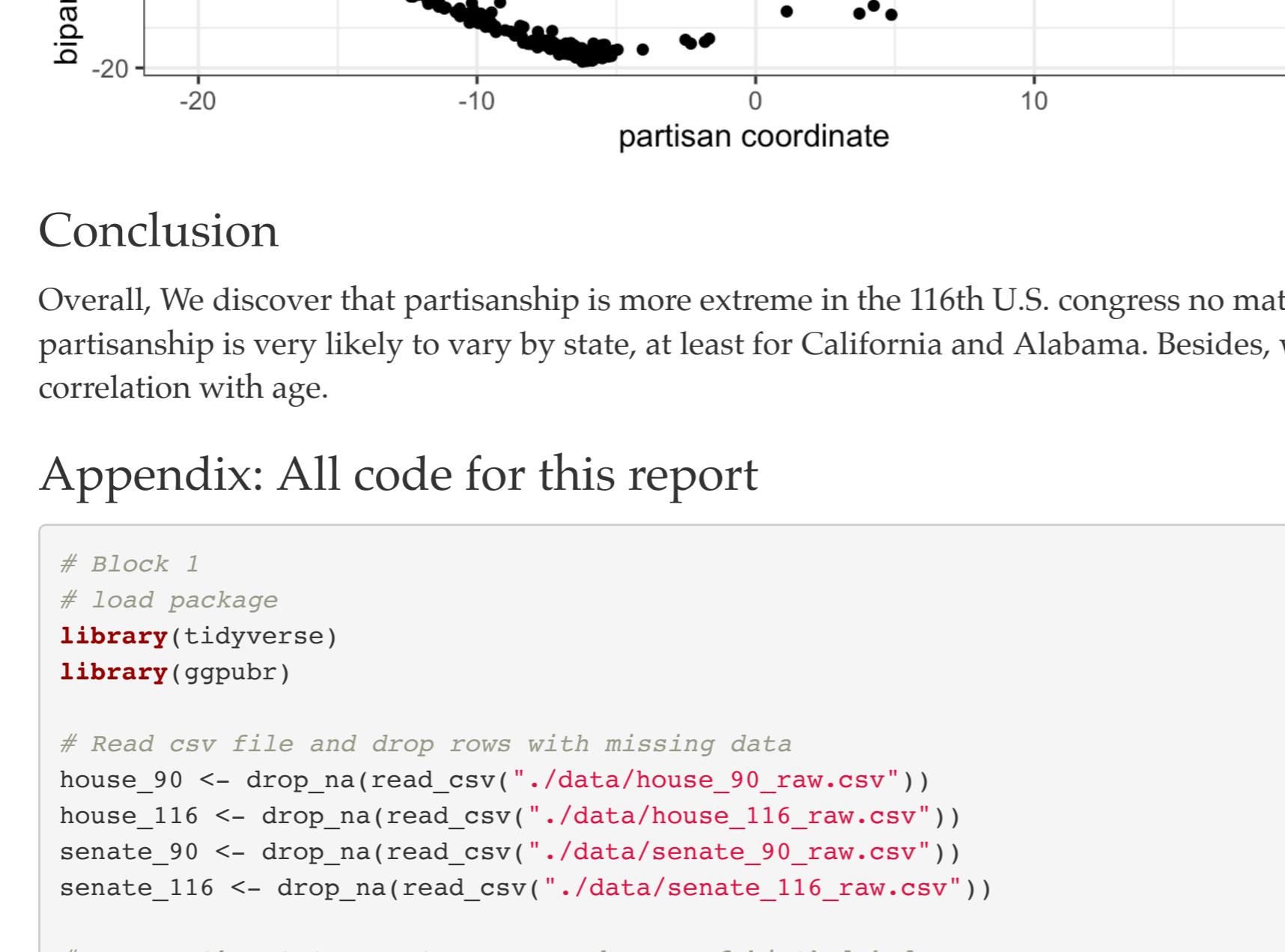
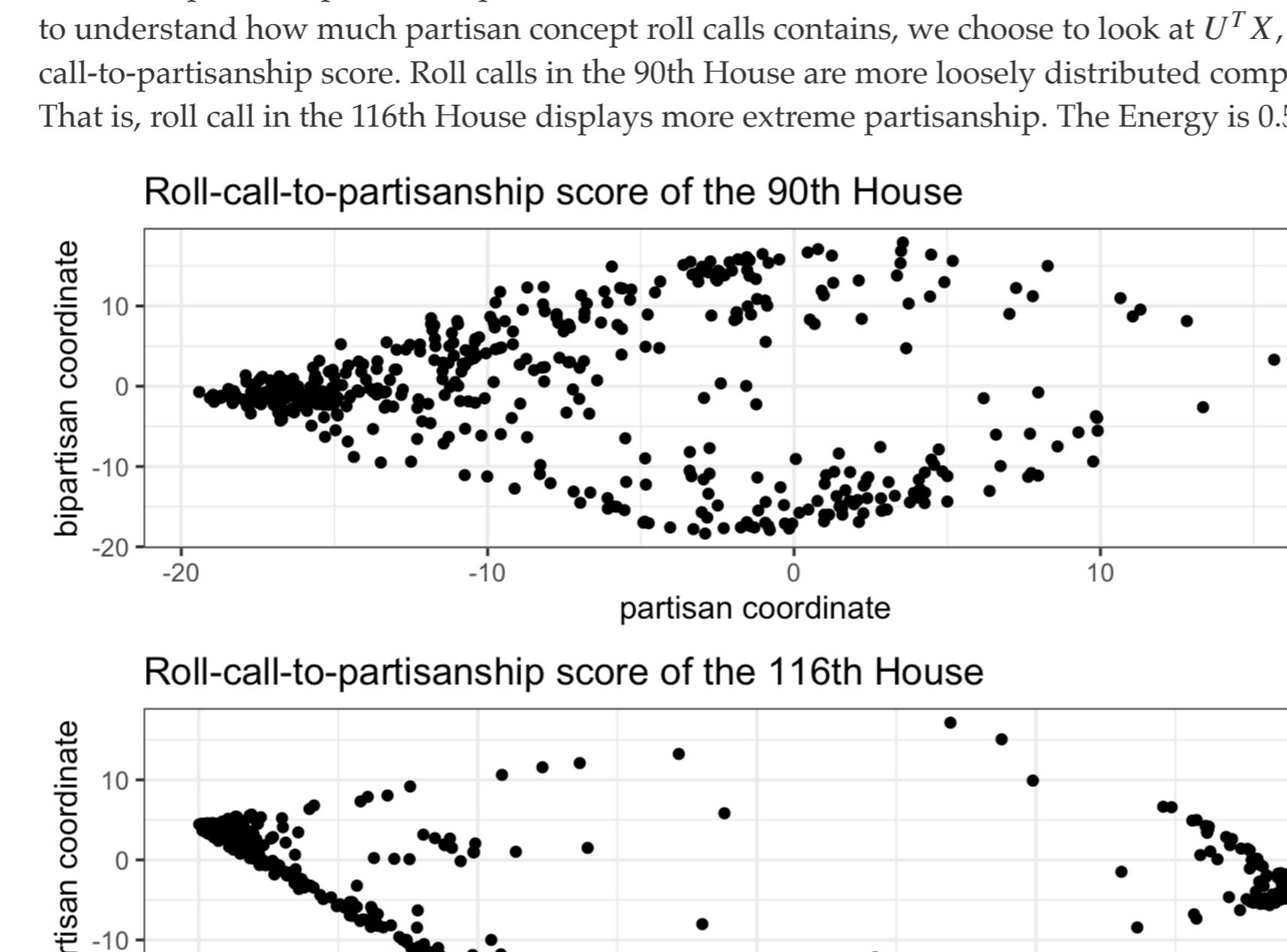
We define four  $n \times p$  voting matrices with one row for each of the  $n$  Representatives in the House or Senate and one column for each of the  $p$  votes taken during the session. The element in  $[i, j]$  is 1 if Representative  $i$  voted "yea" on roll-call vote  $j$ ; -1 if he or she voted "nay", and 0 if he or she is absent or not a member of the chamber during roll call. We will decompose our matrices respectively into the matrix multiplication based on  $X = UDV^T$ . In the following analysis, we will use matrices  $U$ ,  $D$ , and  $V$  to get an insight into partisanship in the 90th and 116th U.S. Congress. Our analysis below is based on rank-2 approximation, and the energy of our approximation will be included together with analysis.

## Findings

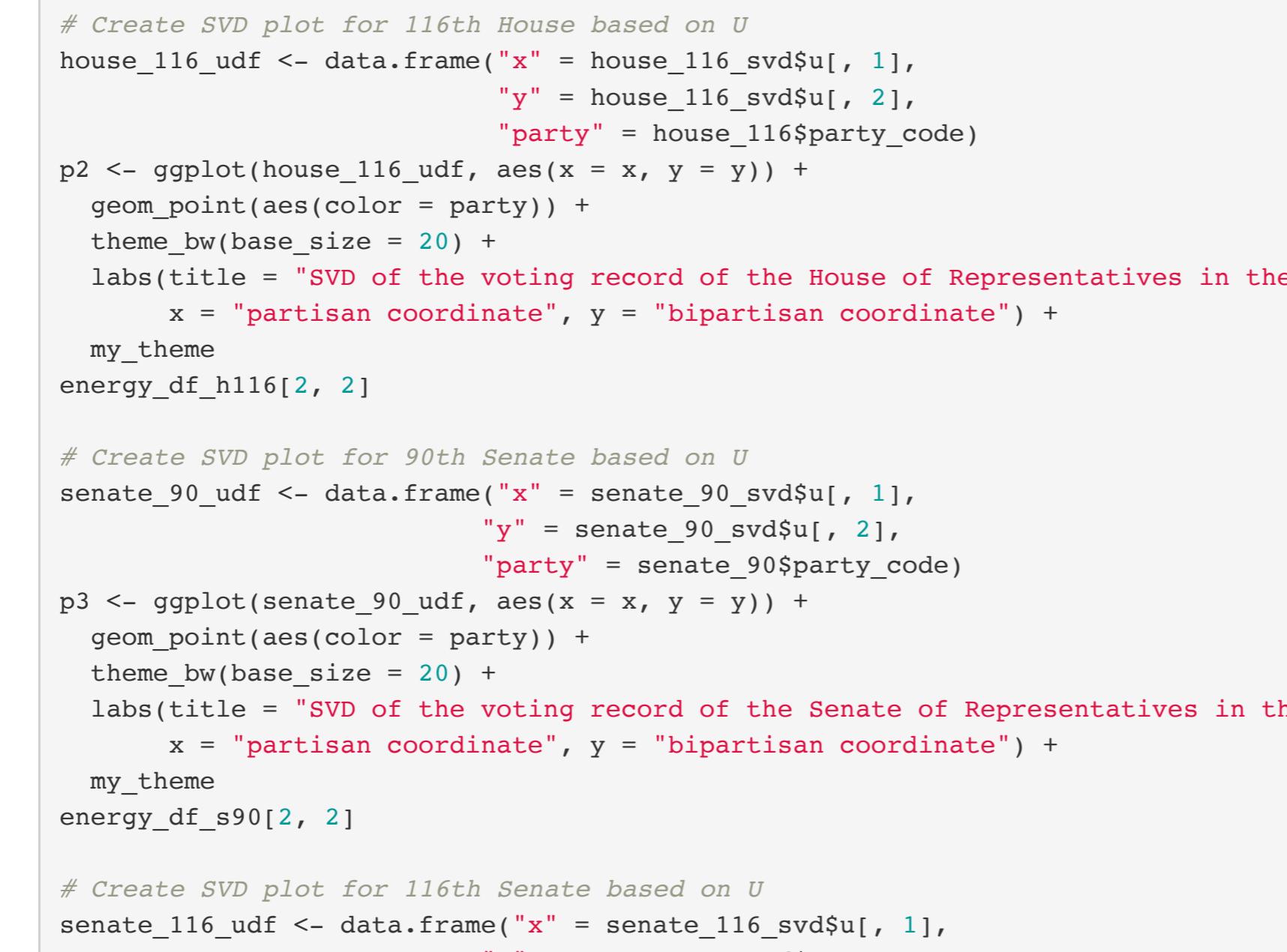
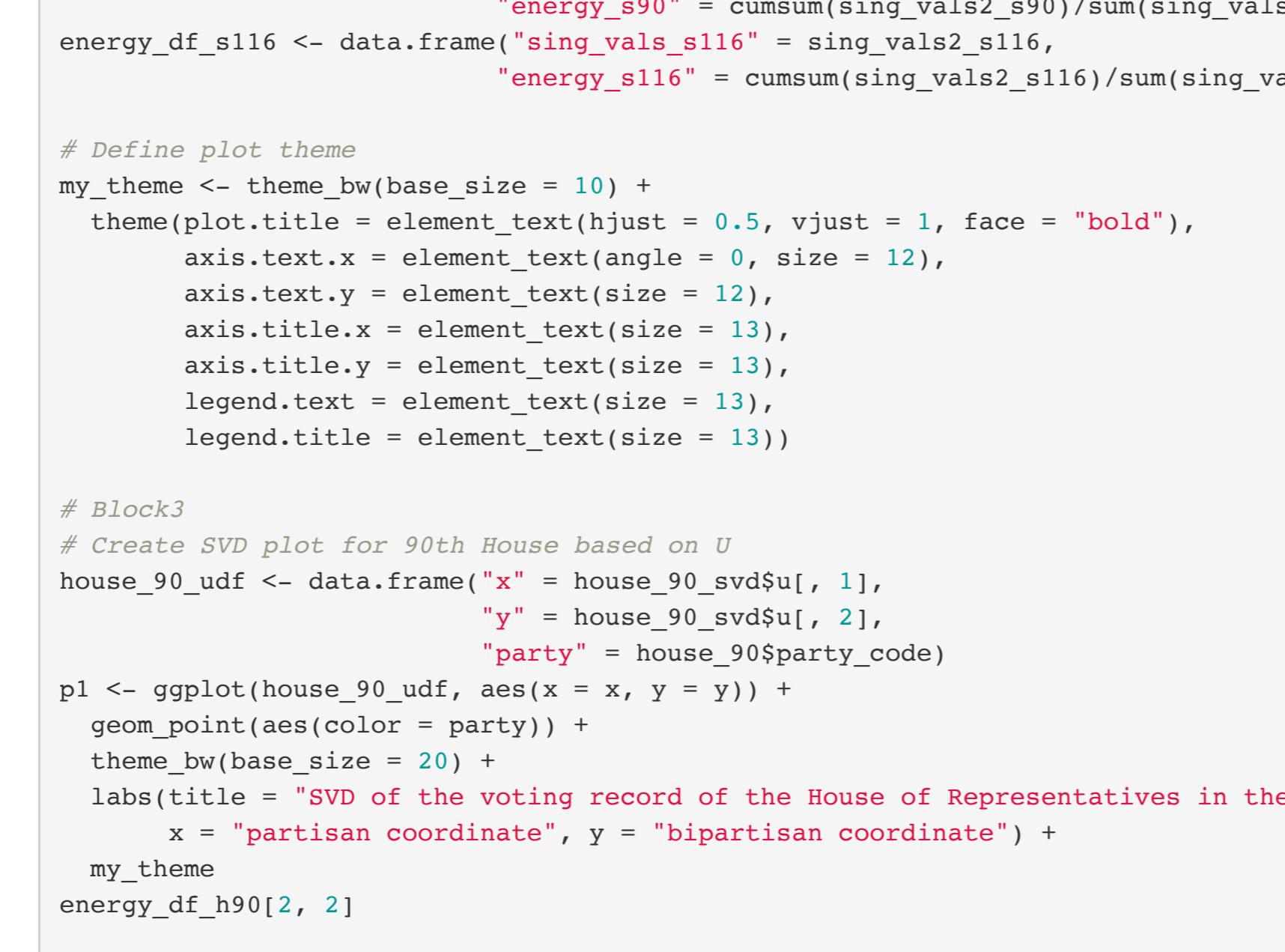
The first thing we are interested in is how different partisanship in the 90th U.S. Congress is, compared to partisanship in the 116th U.S. Congress. Therefore, we take the first two columns from  $U$  as our  $x$  and  $y$  respectively.  $U$  represents how partisanship of a Representative is. In our plots, each point represents a projection of a Representative's votes onto eigenvectors corresponding to the leading two singular values. The two axes are denoted "partisan" and "bipartisan". Representatives who score highly on it (either positively or negatively) tend often to vote with members of their own party. The energy is 0.5645974 and 0.8421558 for the 90th House voting approximation and the 116th House voting approximation respectively. As a result, the approximation to the 116th house vote may be more close to the actual voting matrix. As we can see, Representatives' votes vary by their party that the majority of Democrats are above the line  $y = 0$ , while Republicans are shown below the line  $y = 0$ . But points are loosely distributed that there is no line to separate Representatives perfectly by the political party. However, in the second plot, the approximation to the 116th House voting matrix, Democrats and Republicans form distinct clusters that we can easily draw a line to separate Representatives by the political affiliation. Therefore, partisanship in the 116th House is more extreme than partisanship in the 90th House, regardless of energy.



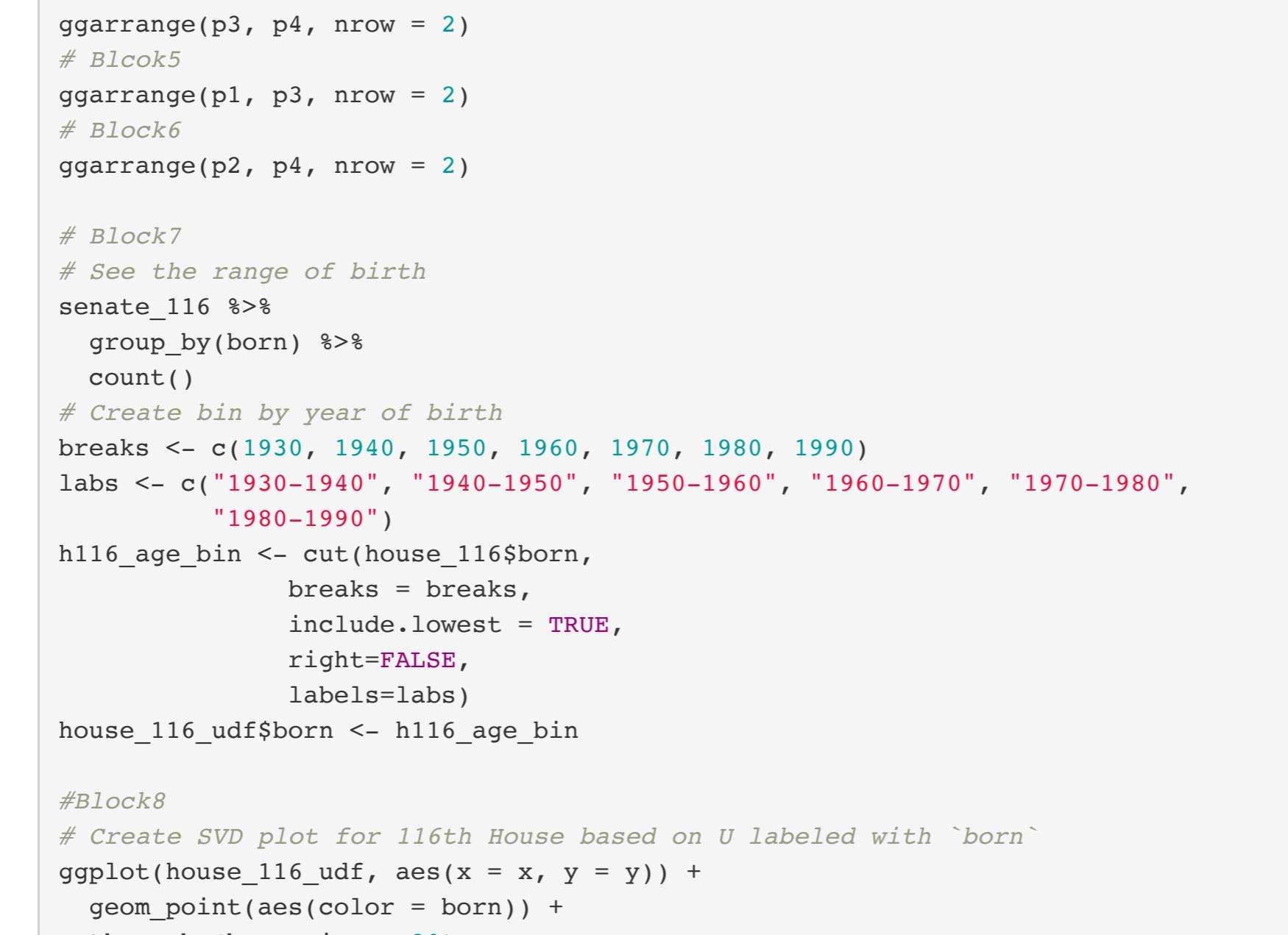
To explore how partisanship compares in the House of Representatives versus the Senate, we compare the first two columns of  $U$  for the 90th House dataset with the first two columns of  $U$  for the 90th Senate dataset. In the 90th U.S. Congress, we find that Representatives in the Senate are slightly more loosely distributed than Representatives in the House. Therefore, we say that partisanship in the House is more extreme than in the Senate. However, due to their low energy, which is 0.5645974 and 0.4943004 respectively, higher energy is needed to make further justification.



In contrast, in the 116th U.S. Congress, both our two plots display extreme partisanship in the House and the Senate. The energy are 0.8421558 and 0.8053386 respectively. Overall partisanship in the House does not differ from partisanship in the Senate too much. Are there any other factors that relate to partisanship? To answer this question, we turn our attention to age.



In this question, we will only focus on the 116th House because of the extreme partisanship in the 116th U.S. congress, larger sample size, and higher energy retained. In order to make our plot clean enough, I create 6 birth year groups, which are 1930-1940, 1950-1960, 1960-1970, 1970-1980, and 1990-1990. In this plot, the first two columns of  $U$  are used as well. The energy retained is 0.8421558 in the plot below, it seems that none of these groups have their points only appear in one cluster. Instead, partisanship seems to have a poor correlation with age. So what about states? Does partisanship vary by state?



Since there are 50 states in total, we choose to compare the political standpoint of Representatives from two states, California and Alabama. In this plot, I also use the first two column of  $U$ . The majority of Representatives from California cluster at the upper-left corner, while Representatives from Alabama spread on the right side of the plot, with energy being 0.8421558. Therefore, we say that the majority of Representatives from California is very likely to belong to a different political affiliation from the majority of Representatives from Alabama.

## Appendix: All code for this report

```
# Block 1
# load package
library(tidyverse)
library(ggpubr)

# Read csv file and drop rows with missing data
house_90 <- drop_na(read_csv("./data/house_90_raw.csv"))
house_116 <- drop_na(read_csv("./data/house_116_raw.csv"))
senate_90 <- drop_na(read_csv("./data/senate_90_raw.csv"))
senate_116 <- drop_na(read_csv("./data/senate_116_raw.csv"))

# remove state, party, name and year of birth labels
house_90_svd <- svd(house_90[, -c(1:4)])
house_116_svd <- svd(house_116[, -c(1:4)])
senate_90_svd <- svd(senate_90[, -c(1:4)])
senate_116_svd <- svd(senate_116[, -c(1:4)])

# Calculate Energy maintained
sing_vals2_h90 <- house_90_svd$svd^2
sing_vals2_h116 <- house_116_svd$svd^2
sing_vals2_s90 <- senate_90_svd$svd^2
sing_vals2_s116 <- senate_116_svd$svd^2
energy_df_h90 <- data.frame("sing_vals_h90" = sing_vals2_h90,
                             "energy_h90" = cumsum(sing_vals2_h90)/sum(sing_vals2_h90))
energy_df_h116 <- data.frame("sing_vals_h116" = sing_vals2_h116,
                             "energy_h116" = cumsum(sing_vals2_h116)/sum(sing_vals2_h116))
energy_df_s90 <- data.frame("sing_vals_s90" = sing_vals2_s90,
                             "energy_s90" = cumsum(sing_vals2_s90)/sum(sing_vals2_s90))
energy_df_s116 <- data.frame("sing_vals_s116" = sing_vals2_s116,
                             "energy_s116" = cumsum(sing_vals2_s116)/sum(sing_vals2_s116))

# Define plot theme
my_theme <- theme_bw(base_size = 10) +
  theme(plot.title = element_text(hjust = 0.5, vjust = 1, face = "bold"),
        axis.text.x = element_text(angle = 0, size = 12),
        axis.text.y = element_text(size = 12),
        axis.title.x = element_text(size = 13),
        axis.title.y = element_text(size = 13),
        legend.text = element_text(size = 13),
        legend.title = element_text(size = 13))

# Block2
# Create SVD plot for 90th House based on U
house_90_udf <- data.frame("x" = house_90_svd$u[, 1],
                           "y" = house_90_svd$u[, 2],
                           "party" = house_90$party_code)
p1 <- ggplot(house_90_udf, aes(x = x, y = y)) +
  geom_point(aes(color = party)) +
  theme_bw(base_size = 20) +
  labs(title = "SVD of the voting record of the House of Representatives in the 90th U.S. Congresses",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme
energy_df_h90[2, 2]

# Create SVD plot for 116th House based on U
house_116_udf <- data.frame("x" = house_116_svd$u[, 1],
                           "y" = house_116_svd$u[, 2],
                           "party" = house_116$party_code)
p2 <- ggplot(house_116_udf, aes(x = x, y = y)) +
  geom_point(aes(color = party)) +
  theme_bw(base_size = 20) +
  labs(title = "SVD of the voting record of the House of Representatives in the 116th U.S. Congresses",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme
energy_df_h116[2, 2]

# Create SVD plot for 90th Senate based on U
senate_90_udf <- data.frame("x" = senate_90_svd$u[, 1],
                           "y" = senate_90_svd$u[, 2],
                           "party" = senate_90$party_code)
p3 <- ggplot(senate_90_udf, aes(x = x, y = y)) +
  geom_point(aes(color = party)) +
  theme_bw(base_size = 20) +
  labs(title = "SVD of the voting record of the Senate of Representatives in the 90th U.S. Congresses",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme
energy_df_s90[2, 2]

# Create SVD plot for 116th Senate based on U
senate_116_udf <- data.frame("x" = senate_116_svd$u[, 1],
                           "y" = senate_116_svd$u[, 2],
                           "party" = senate_116$party_code)
p4 <- ggplot(senate_116_udf, aes(x = x, y = y)) +
  geom_point(aes(color = party)) +
  theme_bw(base_size = 20) +
  labs(title = "SVD of the voting record of the Senate of Representatives in the 116th U.S. Congresses",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme
energy_df_s116[2, 2]

# Combine ggplots
ggarrange(p1, p2, nrow = 2)
# Block4
ggarrange(p3, p4, nrow = 2)
# Block5
ggarrange(p1, p3, nrow = 2)
# Block6
ggarrange(p2, p4, nrow = 2)

# Block7
# See the range of birth
senate_116 %>%
  group_by(born) %>%
  count()
# Create bin by year of birth
breaks <- c("1930-1940", "1940-1950", "1950-1960",
           "1960-1970", "1970-1980", "1980-1990")
labz <- c("1930-1940", "1940-1950", "1950-1960",
         "1960-1970", "1970-1980", "1980-1990",
         "born")
h116_age_bin <- cut(house_116$born,
                     breaks = breaks,
                     include.lowest = TRUE,
                     right=FALSE,
                     labels=labz)
house_116_udf$born <- h116_age_bin

# Block8
# Create SVD plot for 116th House based on U labeled with `born`
ggplot(house_116_udf, aes(x = x, y = y)) +
  geom_point(aes(color = born)) +
  theme_bw(base_size = 20) +
  labs(title = "SVD of the voting record of the House of Representatives in the 116th U.S. Congresses",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme

# Block9
# Create SVD plot for 116th House based on U labeled with `state`
house_116_udf$state <- house_116$state_abbrev
house_116_udf %>%
  filter(state %in% c("CA", "AL")) %>%
  ggplot(aes(x = x, y = y)) +
  geom_point(aes(color = state)) +
  theme_bw(base_size = 10) +
  labs(title = "SVD of the voting record of the House of Representatives in the 116th U.S. Congresses",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme

# Block10
# Create Roll-call-to-partisanship score plot for 90th House
house_90_score <- t(house_90_svd$u)[,-c(1:4)]
house_90_sdf <- data.frame("x" = house_90_score[, 1],
                           "y" = house_90_score[, 2],
                           "rc" = colnames(house_90)[-c(1:4)])
house_90_sdf %>%
  mutate(rc = house_90_svd$u[, -c(1:4)]) +
  ggplot(aes(x = x, y = y, label = rc)) +
  geom_point() +
  theme_bw(base_size = 12) +
  labs(title = "Roll-call-to-partisanship score of the 90th House",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme

# Block11
# Create Roll-call-to-partisanship score plot for 116th House
p5 <- ggplot(house_116_svd, aes(x = x, y = y, label = rc)) +
  geom_point() +
  theme_bw(base_size = 12) +
  labs(title = "Roll-call-to-partisanship score of the 90th House",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme

p6 <- ggplot(house_116_svd, aes(x = x, y = y, label = rc)) +
  geom_point() +
  theme_bw(base_size = 12) +
  labs(title = "Roll-call-to-partisanship score of the 116th House",
       x = "partisan coordinate", y = "bipartisan coordinate") +
  my_theme

#Combine ggplots
ggarrange(p5, p6, nrow = 2)
```

## Bibliography

Mason A. Porter (2005, April 5). A network analysis of committees in the U.S. House of Representatives.

[https://bryandmartin.github.io/STAT302/docs/Projects/project2\\_svd/HouseOfRepresentatives.pdf](https://bryandmartin.github.io/STAT302/docs/Projects/project2_svd/HouseOfRepresentatives.pdf)