Lecture 18 Notes

2024-04-02

Loading data

```
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

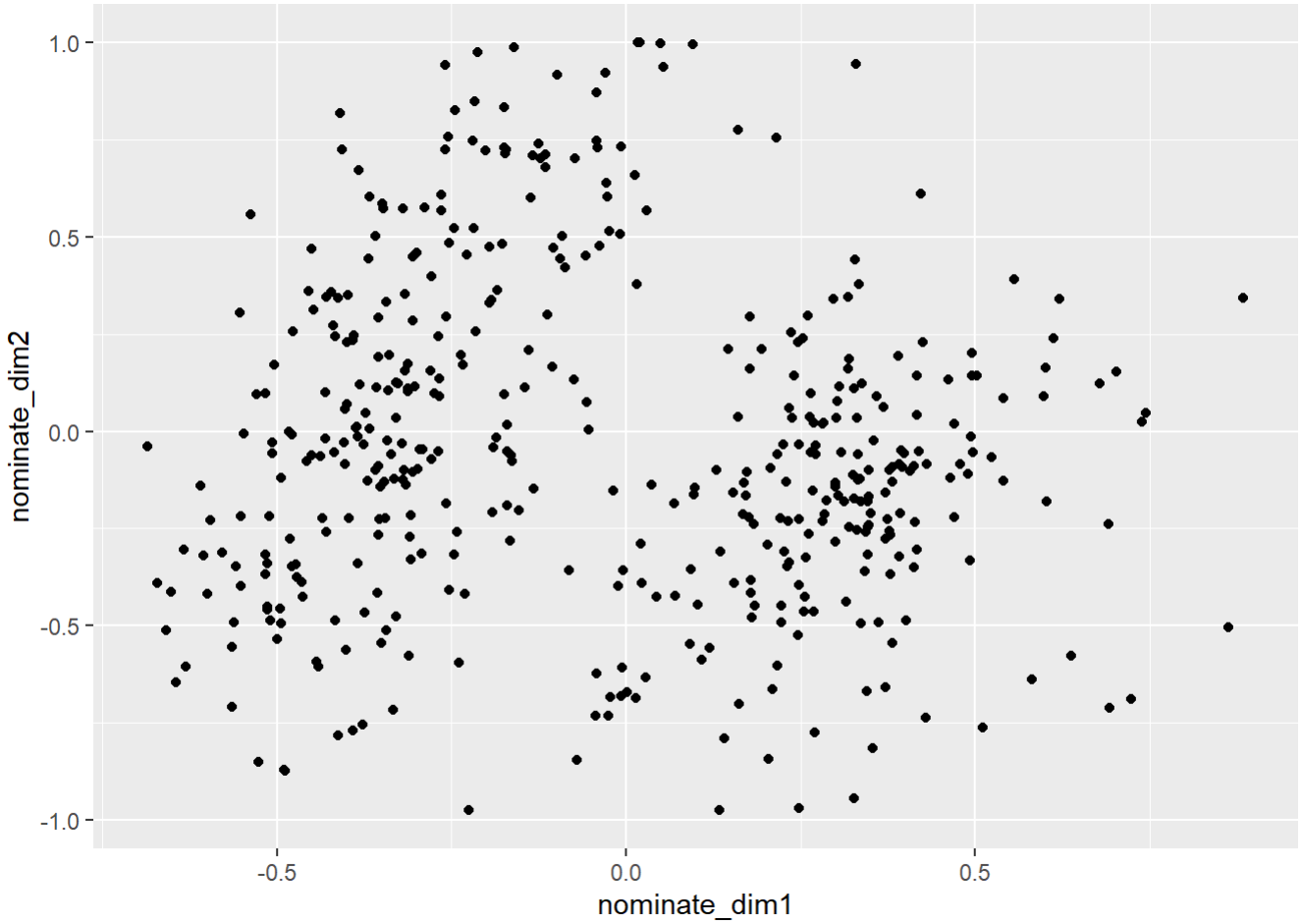
```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts t
o become errors
```

```
dat <- read_csv("https://raw.githubusercontent.com/jbisbee1/DS1000_S2024/main/data/H097_
members.csv")
```

```
## Rows: 445 Columns: 22
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (4): chamber, state_abbrev, bioname, bioguide_id
## dbl (17): congress, icpsr, state_icpsr, district_code, party_code, occupancy...
## lgl  (1): conditional
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
dat %>%
  ggplot(aes(x = nominate_dim1,
             y = nominate_dim2)) +
  geom_point()
```




```
# kmeans() function
forAnalysis <- dat %>%
  select(nominate_dim1,
         nominate_dim2) %>%
  drop_na()
# set.seed(123)
kRes <- kmeans(x = forAnalysis,
               centers = 2,
               nstart = 20)

kRes
```

```
## K-means clustering with 2 clusters of sizes 279, 166
##
## Cluster means:
##   nominate_dim1 nominate_dim2
## 1    0.07833333    -0.2810108
## 2   -0.21563253     0.3701205
##
## Clustering vector:
##   [1] 1 1 2 2 2 2 2 1 1 1 2 2 2 2 2 2 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 2
##  [38] 1 1 1 1 1 2 1 2 2 1 1 1 2 1 2 1 1 1 2 1 1 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1
##  [75] 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 1 1 2 2 1 1 1 2
## [112] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 1 1
## [149] 1 1 2 2 2 1 2 1 1 2 2 2 1 2 1 2 1 1 1 2 1 1 1 1 1 2 2 2 1 1 2 1 2 1 1 1 2
## [186] 2 1 1 1 2 1 2 1 2 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 2 1 1 2
## [223] 2 1 2 1 2 2 2 2 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2
## [260] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1
## [297] 2 1 2 2 1 2 2 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1
## [334] 2 2 1 2 2 2 1 1 1 1 1 2 1 2 1 2 1 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 2 2 1 2 1 1
## [371] 1 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 1 2 2 2 1 1 2
## [408] 1 2 2 2 1 1 1 2 2 1 2 1 2 1 1 1 1 2 1 2 2 1 2 1 2 2 1 1 1 1 2 2 1 1 1 1 1
## [445] 1
##
## Within cluster sum of squares by cluster:
## [1] 57.98885 23.66725
##  (between_SS / total_SS =  39.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
toplot <- forAnalysis %>%
  mutate(cluster = kRes$cluster)

summary(lm(nominate_dim1 ~ factor(cluster),toplot))
```

```
##
## Call:
## lm(formula = nominate_dim1 ~ factor(cluster), data = toplot)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74933 -0.21337  0.04063  0.24067  0.80567
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.07833    0.01941   4.036 6.41e-05 ***
## factor(cluster)2 -0.29397    0.03178  -9.251  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3242 on 443 degrees of freedom
## Multiple R-squared:  0.1619, Adjusted R-squared:   0.16
## F-statistic: 85.57 on 1 and 443 DF,  p-value: < 2.2e-16
```

```
toplot %>%
  ggplot(aes(x = nominate_dim1,
             y = nominate_dim2,
             color = factor(cluster))) +
  geom_point()
```