AE-10

Brandon Leslie

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.4
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(pls)
```

```

Attaching package: 'pls'

The following object is masked from 'package:stats':

    loadings
```

```
library(tidymodels)
```

```
-- Attaching packages --------------------------------------- tidymodels 1.3.0 --
v broom        1.0.7     v rsample      1.2.1
v dials        1.4.0     v tune         1.3.0
v infer        1.0.7     v workflows    1.2.0
v modeldata    1.4.0     v workflowsets 1.1.0
v parsnip      1.3.0     v yardstick    1.3.2
```

```
v recipes      1.1.1
-- Conflicts ---------------------------------------- tidymodels_conflicts() --
x scales::discard() masks purrr::discard()
x dplyr::filter()   masks stats::filter()
x recipes::fixed()  masks stringr::fixed()
x dplyr::lag()      masks stats::lag()
x yardstick::spec() masks readr::spec()
x recipes::step()   masks stats::step()
```

```
USArrests <- USArrests |> mutate(state = rownames(USArrests))

arrests_pca <- prcomp(USArrests %>% select( -state),
                      center = TRUE, scale. = TRUE)
head(USArrests)
```

```
           Murder Assault UrbanPop Rape       state
Alabama      13.2     236       58 21.2     Alabama
Alaska       10.0     263       48 44.5      Alaska
Arizona       8.1     294       80 31.0     Arizona
Arkansas      8.8     190       50 19.5    Arkansas
California    9.0     276       91 40.6  California
Colorado      7.9     204       78 38.7    Colorado
```

```
arrests_pca
```

```
Standard deviations (1, .., p=4):
[1] 1.5748783 0.9948694 0.5971291 0.4164494

Rotation (n x k) = (4 x 4):
                PC1        PC2        PC3         PC4
Murder   -0.5358995 -0.4181809  0.3412327  0.64922780
Assault  -0.5831836 -0.1879856  0.2681484 -0.74340748
UrbanPop -0.2781909  0.8728062  0.3780158  0.13387773
Rape     -0.5434321  0.1673186 -0.8177779  0.08902432
```

```
arrests_pca$x |> head()
```

```
                 PC1        PC2         PC3          PC4
Alabama    -0.9756604 -1.1220012  0.43980366  0.154696581
Alaska     -1.9305379 -1.0624269 -2.01950027 -0.434175454
```

```
Arizona    -1.7454429  0.7384595 -0.05423025 -0.826264240
Arkansas    0.1399989 -1.1085423 -0.11342217 -0.180973554
California -2.4986128  1.5274267 -0.59254100 -0.338559240
Colorado   -1.4993407  0.9776297 -1.08400162  0.001450164
```

```
pca_rec <- recipe(state ~., data = USArrests) %>%
  step_normalize(all_predictors()) %>%
  step_pca(all_predictors())

pca_prep <- prep(pca_rec)
```

```
tidied_pca <- tidy(pca_prep, 2)
tidied_pca
```

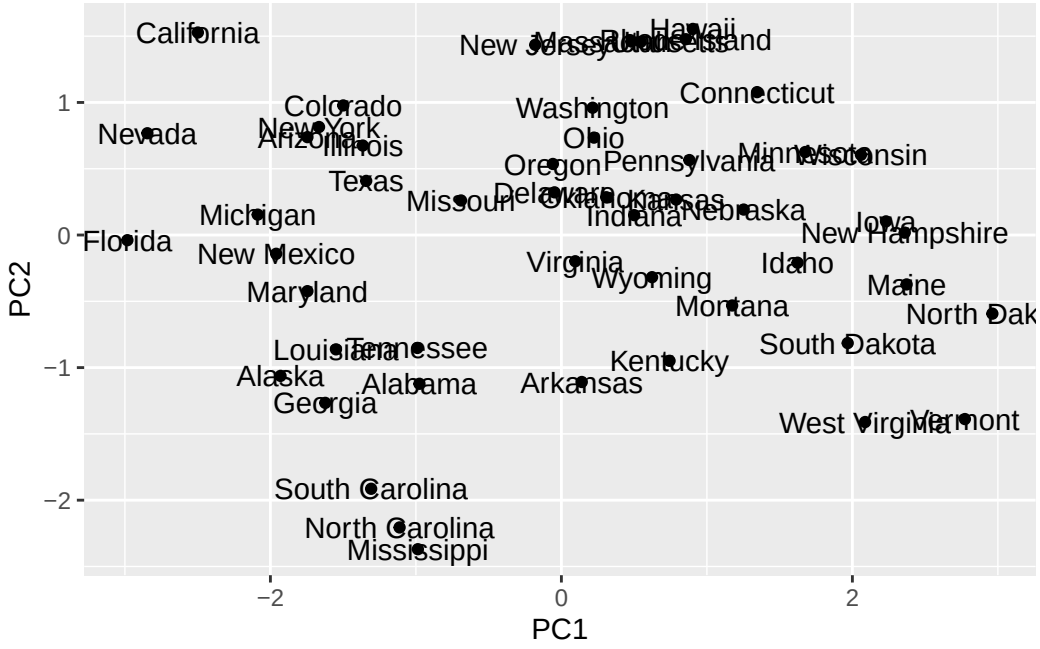
```
# A tibble: 16 x 4
   terms      value component id
   <chr>      <dbl> <chr>     <chr>
 1 Murder   -0.536  PC1       pca_xe1GY
 2 Assault  -0.583  PC1       pca_xe1GY
 3 UrbanPop -0.278  PC1       pca_xe1GY
 4 Rape     -0.543  PC1       pca_xe1GY
 5 Murder   -0.418  PC2       pca_xe1GY
 6 Assault  -0.188  PC2       pca_xe1GY
 7 UrbanPop  0.873  PC2       pca_xe1GY
 8 Rape      0.167  PC2       pca_xe1GY
 9 Murder    0.341  PC3       pca_xe1GY
10 Assault   0.268  PC3       pca_xe1GY
11 UrbanPop  0.378  PC3       pca_xe1GY
12 Rape     -0.818  PC3       pca_xe1GY
13 Murder    0.649  PC4       pca_xe1GY
14 Assault  -0.743  PC4       pca_xe1GY
15 UrbanPop  0.134  PC4       pca_xe1GY
16 Rape      0.0890 PC4       pca_xe1GY
```

```
arrests_prep <- pca_rec |> prep() %>% bake(USArrests)
arrests_prep |> head()
```

```
# A tibble: 6 x 5
  state         PC1    PC2     PC3      PC4
  <fct>       <dbl>  <dbl>   <dbl>    <dbl>
1 Alabama    -0.976 -1.12   0.440   0.155
```

```
2 Alaska     -1.93  -1.06  -2.02   -0.434
3 Arizona    -1.75   0.738 -0.0542 -0.826
4 Arkansas    0.140 -1.11  -0.113  -0.181
5 California -2.50   1.53  -0.593  -0.339
6 Colorado   -1.50   0.978 -1.08    0.00145
```

```
arrests_prep %>%
  ggplot(aes(PC1, PC2)) +
  geom_text(aes(label = state)) +
  geom_point()
```

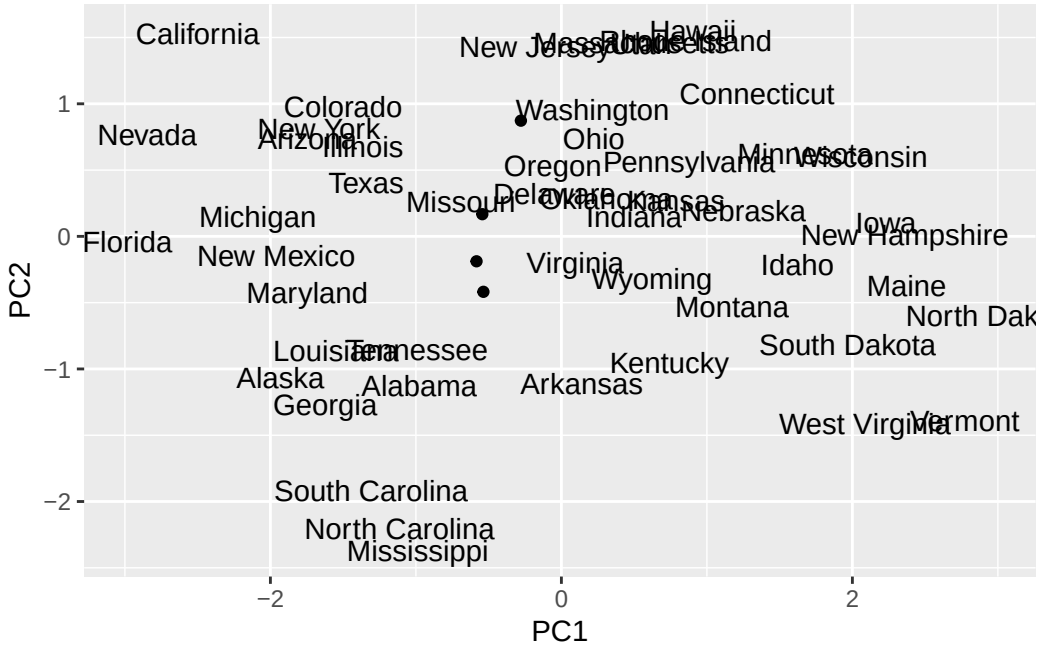



```
tidied_pca %>%
  pivot_wider(names_from = component, values_from = value) %>%
  select(-id)
```

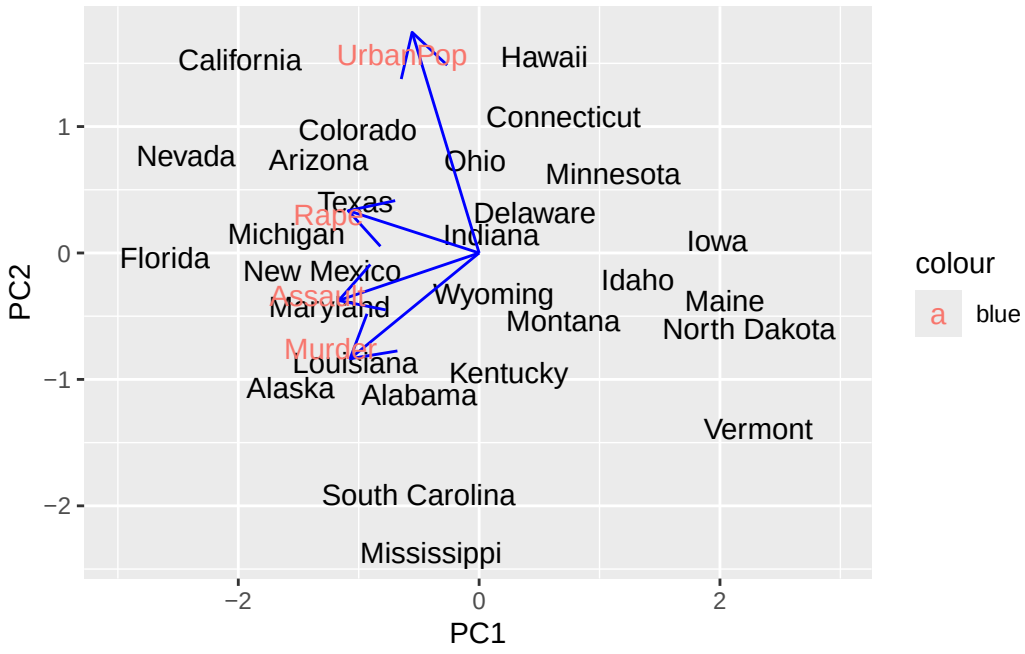
```
# A tibble: 4 x 5
  terms       PC1    PC2    PC3     PC4
  <chr>     <dbl>  <dbl>  <dbl>   <dbl>
1 Murder   -0.536 -0.418  0.341  0.649
2 Assault  -0.583 -0.188  0.268 -0.743
3 UrbanPop -0.278  0.873  0.378  0.134
4 Rape     -0.543  0.167 -0.818  0.0890
```

```
tidy_pca_wide <- tidied_pca %>%
  pivot_wider(names_from = component, values_from = value) %>%
  select(-id)
```

```
arrests_prep %>%
  ggplot(aes(PC1, PC2)) +
  geom_text(aes(label = state)) +
  geom_point(data = tidy_pca_wide)
```




```
arrests_prep %>%
  ggplot(aes(PC1, PC2)) +
  geom_text(aes(label = state), check_overlap = TRUE, hjust = "inward") +
  geom_segment(data = tidy_pca_wide, aes(x = 0, y = 0, xend = 2*PC1, yend = 2 *PC2),
               arrow = arrow(), color = "blue") +
  geom_text(data = tidy_pca_wide, aes(x = 2.3 *PC1, y = 1.8*PC2, label = terms, color = "blue
```


California
UrbanPop
Hawaii
Connecticut
Colorado
Nevada
Arizona
Ohio
Minnesota
Texas
Rape
Delaware
Michigan
Indiana
Iowa
Florida
New Mexico
Idaho
Wyoming
Assault
Maryland
Maine
Montana
North Dakota
Murder
Louisiana
Kentucky
Alaska
Alabama
Vermont
South Carolina
Mississippi
PC2
PC1
1
0
−1
−2
−2
0
2
colour
a
blue