tidy\_pca

friendly

2023-09-22

## PCA, tidy style

This example from: <https://clauswilke.com/blog/2020/09/07/pca-tidyverse-style/> see also: <https://broom.tidymodels.org/reference/tidy.prcomp.html>

library(tidyverse)  
library(broom) # devtools::install\_github("tidymodels/broom")  
library(cowplot)

## Read the data

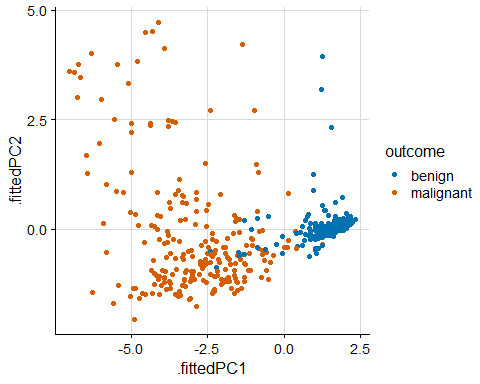
biopsy <- read\_csv("https://wilkelab.org/classes/SDS348/data\_sets/biopsy.csv")

## Do the PCA

pca\_fit <- biopsy |>   
 select(where(is.numeric)) |> # retain only numeric columns  
 scale() |> # scale data  
 prcomp() # do PCA

## plot points

pca\_fit |>  
 augment(biopsy) |> # add original dataset back in  
 ggplot(aes(.fittedPC1, .fittedPC2, color = outcome)) +   
 geom\_point(size = 1.5) +  
 scale\_color\_manual(  
 values = c(malignant = "#D55E00", benign = "#0072B2")  
 ) +  
 theme\_half\_open(12) + background\_grid()



## extract rotation matrix (loadings)

pca\_fit |>  
 tidy(matrix = "rotation")

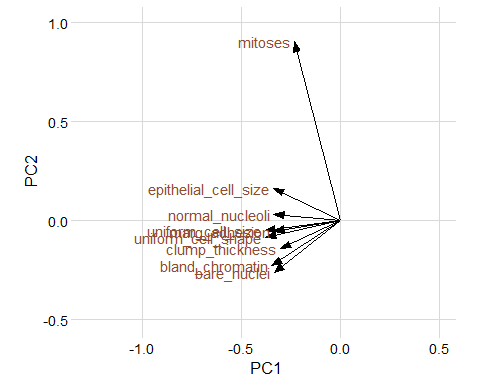
## # A tibble: 81 × 3  
## column PC value  
## <chr> <dbl> <dbl>  
## 1 clump\_thickness 1 -0.302   
## 2 clump\_thickness 2 -0.141   
## 3 clump\_thickness 3 0.866   
## 4 clump\_thickness 4 0.108   
## 5 clump\_thickness 5 -0.0803   
## 6 clump\_thickness 6 -0.243   
## 7 clump\_thickness 7 -0.00852  
## 8 clump\_thickness 8 -0.248   
## 9 clump\_thickness 9 0.00275  
## 10 uniform\_cell\_size 1 -0.381   
## # ℹ 71 more rows

define arrow style for plotting

arrow\_style <- arrow(  
 angle = 20, ends = "first", type = "closed", length = grid::unit(8, "pt")  
)

## plot rotation matrix

pca\_fit |>  
 tidy(matrix = "rotation") |>  
 pivot\_wider(names\_from = "PC", names\_prefix = "PC", values\_from = "value") |>  
 ggplot(aes(PC1, PC2)) +  
 geom\_segment(xend = 0, yend = 0, arrow = arrow\_style) +  
 geom\_text(  
 aes(label = column),  
 hjust = 1, nudge\_x = -0.02,   
 color = "#904C2F"  
 ) +  
 xlim(-1.25, .5) + ylim(-.5, 1) +  
 coord\_fixed() + # fix aspect ratio to 1:1  
 theme\_minimal\_grid(12)



# scree plot  
pca\_fit |>  
 tidy(matrix = "eigenvalues") |>  
 ggplot(aes(PC, percent)) +  
 geom\_col(fill = "#56B4E9", alpha = 0.8) +  
 scale\_x\_continuous(breaks = 1:9) +  
 scale\_y\_continuous(  
 labels = scales::percent\_format(),  
 expand = expansion(mult = c(0, 0.01))  
 ) +  
 theme\_minimal\_hgrid(12)

