title: "Clustering" author: "Franchesca Johnson" date: "2023-08-11" output: pdf_document editor_options:
markdown: wrap: 72

```r
summary(clustering_data)
```

Error in eval(expr, envir, enclos): object 'clustering_data' not found

```r
ggplot(clustering_data, aes(x=x, y=y)) + geom_point()
```

Error in ggplot(clustering_data, aes(x = x, y = y)): could not find function "ggplot"

```r
library(ggplot2)
```

Warning: package 'ggplot2' was built under R version 4.3.1

```r
library(tidyverse)
```

Warning: package 'tidyverse' was built under R version 4.3.1

Warning: package 'tibble' was built under R version 4.3.1

Warning: package 'tidyr' was built under R version 4.3.1

Warning: package 'readr' was built under R version 4.3.1

Warning: package 'purrr' was built under R version 4.3.1

Warning: package 'dplyr' was built under R version 4.3.1

Warning: package 'stringr' was built under R version 4.3.1

Warning: package 'forcats' was built under R version 4.3.1

Warning: package 'lubridate' was built under R version 4.3.1

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## -- 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 become errors
```

```r
library(tidyverse)  # data manipulation
library(cluster)    # clustering algorithms

df <- clustering_data
```

Error in eval(expr, envir, enclos): object 'clustering_data' not found

```r
set.seed(123)
km.out <- kmeans(df, centers = 3, nstart = 25)
```

```
## Error in as.vector(x, mode): cannot coerce type 'closure' to vector of type 'any'
```

```r
km.out
```

```
## Error in eval(expr, envir, enclos): object 'km.out' not found
```

```r
k.values <- 2:12
k_values <- as.numeric(k.values)

n_clusters <- 11
wss1 <- numeric(n_clusters)

for (i in k.values) {
  km.out <- kmeans(df, centers = i, nstart = 25)
  wss1[i] <- km.out$tot.withinss
}
```

```
## Error in as.vector(x, mode): cannot coerce type 'closure' to vector of type 'any'
```

```r
wss_values <- map_dbl(k.values, wss)
```

```
## Error in eval(expr, envir, enclos): object 'wss' not found
```

```r
wss_df <- tibble(cluster = k.values, wss = wss_values)
```

```
## Error: object 'wss_values' not found
```

```r
ggplot(wss_df, aes(x=cluster, y=wss)) + geom_point()
```

```
## Error in eval(expr, envir, enclos): object 'wss_df' not found
```

```r
df <- clustering_data
```

```
## Error in eval(expr, envir, enclos): object 'clustering_data' not found
```

```r
df <- na.omit(df)
df <- scale(df)
```

```
## Error in as.vector(x, mode): cannot coerce type 'closure' to vector of type 'any'
```

```r
head(df)
```

```
##
## 1 function (x, df1, df2, ncp, log = FALSE)
## 2 {
## 3     if (missing(ncp))
## 4         .Call(C_df, x, df1, df2, log)
## 5     else .Call(C_dnf, x, df1, df2, ncp, log)
## 6 }
```

```r
kmeans(df, centers = 4, iter.max = 10, nstart = 25)
```

```
## Error in as.vector(x, mode): cannot coerce type 'closure' to vector of type 'any'
```

```r
set.seed(123)

# function to compute total within-cluster sum of square
wss <- function(k) {
  kmeans(df, k, nstart = 10)$tot.withinss
}

# Compute and plot wss for k = 2 to k = 12
k.values <- 2:12


# extract wss for 2-12 clusters
wss_values <- map_dbl(k.values, wss)
```

```
## Error in `map_dbl()`:
## i In index: 1.
## Caused by error in `as.vector()`:
## ! cannot coerce type 'closure' to vector of type 'any'
```

```r
plot(k.values, wss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")
```

```
## Error in eval(expr, envir, enclos): object 'wss_values' not found
```

6. Elbow point is 5