Klasterisasi ICD-10 Data Klaim FKRTL BPJS Kesehatan Menggunakan Metode Density-Based Spatial Clustering of Application with Noise (DBSCAN)

Misbahul Huda

May 22, 2024

Pre-Processing Data

Input Library

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.5.1     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1

## -- 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
```

Import Dataset

```r
## data utama
datamentah <- read_dta('C:/Users/Yurnalis/Documents/Misbah/Buat Kerja/Portofolio/clustering-icd10-data-
peserta <- read_dta('C:/Users/Yurnalis/Documents/Misbah/Buat Kerja/Portofolio/clustering-icd10-data-fkr

## data tambahan
icd10 <- read.csv('C:/Users/Yurnalis/Documents/Misbah/Buat Kerja/Portofolio/clustering-icd10-data-fkrtl
provinsi <- read.csv('C:/Users/Yurnalis/Documents/Misbah/Buat Kerja/Portofolio/clustering-icd10-data-fk

## cuplikan data
head(datamentah)
```

```
## # A tibble: 6 x 55
##      PSTV01  PSTV02 PSTV15 FKP02 FKL02 FKL03      FKL04      FKL05    FKL06
##       <dbl>   <dbl>  <dbl> <chr> <chr> <date>     <date>     <dbl+lb> <dbl+lbl>
## 1  74539301  1.40e8   1.05 ""    1000~ 2019-01-07 2019-01-07 94 [PAP~ 9403 [JAY~
## 2  68573044  1.35e8  21.3  ""    1000~ 2019-01-07 2019-01-15 94 [PAP~ 9403 [JAY~
## 3  45797532  1.53e8   1.79 ""    1000~ 2019-01-08 2019-01-08 94 [PAP~ 9403 [JAY~
## 4 135497918  7.97e7  10.5  ""    1000~ 2019-01-10 2019-01-10 94 [PAP~ 9403 [JAY~
## 5  14459017  1.36e8  50.8  ""    1000~ 2019-01-08 2019-01-08 94 [PAP~ 9403 [JAY~
## 6  19251321  1.93e7   1.05 "290~ 1000~ 2019-01-11 2019-01-11 94 [PAP~ 9403 [JAY~
## # i 46 more variables: FKL07 <dbl+lbl>, FKL08 <dbl+lbl>, FKL09 <dbl+lbl>,
## #   FKL10 <dbl+lbl>, FKL11 <dbl+lbl>, FKL12 <dbl+lbl>, FKL13 <dbl+lbl>,
## #   FKL14 <dbl+lbl>, FKL15 <dbl+lbl>, FKL15A <chr>, FKL16 <chr>, FKL16A <chr>,
## #   FKL17 <dbl+lbl>, FKL17A <chr>, FKL18 <chr>, FKL18A <chr>, FKL19 <chr>,
## #   FKL19A <chr>, FKL20 <dbl+lbl>, FKL21 <dbl+lbl>, FKL22 <dbl>,
## #   FKL23 <dbl+lbl>, FKL25 <dbl+lbl>, FKL26 <dbl+lbl>, FKL27 <dbl+lbl>,
## #   FKL28 <dbl+lbl>, FKL29 <dbl+lbl>, FKL30 <chr>, FKL31 <dbl>, ...
```

```r
head(peserta)
```

```
## # A tibble: 6 x 18
##   PSTV01    PSTV02 PSTV03     PSTV04    PSTV05  PSTV06  PSTV07  PSTV08  PSTV09
##    <dbl>     <dbl> <date>     <dbl+lbl> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <dbl+lb>
## 1     15        15 1944-03-01 1 [PESER~ 2 [PER~ 9 [TID~ 3 [KEL~ 2 [PBI~ 72 [SUL~
## 2     64        64 1971-12-10 1 [PESER~ 2 [PER~ 2 [KAW~ 3 [KEL~ 3 [PBI~ 76 [SUL~
## 3    101       101 1967-12-31 1 [PESER~ 1 [LAK~ 2 [KAW~ 2 [KEL~ 5 [PPU] 12 [SUM~
## 4    218       218 1961-01-30 1 [PESER~ 2 [PER~ 3 [CER~ 3 [KEL~ 2 [PBI~ 18 [LAM~
## 5    340  70225684 1991-05-31 3 [ISTRI] 2 [PER~ 2 [KAW~ 2 [KEL~ 5 [PPU] 33 [JAW~
## 6    349 148483215 1995-07-01 4 [ANAK]  2 [PER~ 9 [TID~ 3 [KEL~ 2 [PBI~ 62 [KAL~
## # i 9 more variables: PSTV10 <dbl+lbl>, PSTV11 <dbl+lbl>, PSTV12 <dbl+lbl>,
## #   PSTV13 <dbl+lbl>, PSTV14 <dbl+lbl>, PSTV15 <dbl>, PSTV16 <dbl>,
## #   PSTV17 <dbl+lbl>, PSTV18 <dbl>
```

```r
head(icd10)
```

```
##   FKL15A                                                    ICD10_Text
## 1    A00                                                   A00 Cholera
## 2    A01                          A01 Typhoid and paratyphoid fevers
## 3    A02                             A02 Other salmonella infections
## 4    A03                                                 A03 Shigellosis
## 5    A04                       A04 Other bacterial intestinal infections
```

```
## 6    A05 A05 Other bacterial foodborne intoxications, not elsewhere classified
##           Kategori
## 1 Penyakit Lainnya
## 2 Penyakit Lainnya
## 3 Penyakit Lainnya
## 4 Penyakit Lainnya
## 5 Penyakit Lainnya
## 6 Penyakit Lainnya
```

```
head(provinsi)
```

```
##   PSTV13 FKL05         Provinsi
## 1     11    11             Aceh
## 2     12    12   Sumatera Utara
## 3     13    13   Sumatera Barat
## 4     14    14             Riau
## 5     15    15            Jambi
## 6     16    16 Sumatera Selatan
```

Memilah dan menggabungkan data yang akan digunakan

```
## data klaim dari tahun 2019 hingga 2020
data1 = datamentah %>%
  select(PSTV01,PSTV15,FKL03,FKL05,FKL15A)
data1$FKL03 = as.Date(data1$FKL03)
data2 = with(data1, data1[(FKL03 >= "2019-01-01" & FKL03 <= "2020-12-31"),])

## merge data klaim dan kategori yang akan menjadi variabel model clustering
data3 = merge(data2,icd10,by="FKL15A")
data4 = merge(data3, provinsi, by="FKL05")
head(data4)
```

```
##   FKL05 FKL15A   PSTV01     PSTV15      FKL03
## 1    11    H17 38061656   9.664808 2020-09-08
## 2    11    A01 27701658 149.909576 2020-12-03
## 3    11    K30 18167452 112.510979 2019-08-16
## 4    11    I15 74055526   3.361672 2020-01-13
## 5    11    A88 11764293   6.933449 2019-10-02
## 6    11    Z09  2136959 471.474579 2019-03-13
##                                                                                 ICD10_Text
## 1                                                                H17 Corneal scars and opacities
## 2                                                              A01 Typhoid and paratyphoid fevers
## 3                                                                        K30 Functional dyspepsia
## 4                                                                     I15 Secondary hypertension
## 5                    A88 Other viral infections of central nervous system, not elsewhere classified
## 6 Z09 Follow-up examination after treatment for conditions other than malignant neoplasms
##           Kategori PSTV13 Provinsi
## 1 Penyakit Lainnya     11     Aceh
## 2 Penyakit Lainnya     11     Aceh
## 3 Penyakit Lainnya     11     Aceh
## 4      Katastropik     11     Aceh
```

```
## 5 Penyakit Lainnya     11     Aceh
## 6 Penyakit Lainnya     11     Aceh
```

```
## select data kepesertaan dan merge provinsi
peserta1 = peserta %>%
  select(PSTV01,PSTV13,PSTV15,PSTV16)
peserta2 = filter(peserta1, PSTV16 >= "2019" & PSTV16 <= "2020")
peserta3 = merge(peserta2,provinsi,by="PSTV13")
head(peserta3)
```

```
##   PSTV13    PSTV01     PSTV15 PSTV16 FKL05 Provinsi
## 1     11 426927845   2.962019   2019    11     Aceh
## 2     11 290031595  38.169888   2019    11     Aceh
## 3     11 306397726  44.507668   2019    11     Aceh
## 4     11 313565482 604.251282   2019    11     Aceh
## 5     11 313651411   7.447783   2019    11     Aceh
## 6     11 315530206   0.933933   2019    11     Aceh
```

Checking blank row data klaim

```
cat("Summary data yang digunakan:\n\n")
```

```
## Summary data yang digunakan:
```

```
summary(data4)
```

```
##      FKL05          FKL15A              PSTV01              PSTV15
##  Min.   :11.00   Length:1761625     Min.   :     1086   Min.   :    0.548
##  1st Qu.:31.00   Class :character   1st Qu.: 36008749   1st Qu.:   12.447
##  Median :33.00   Mode  :character   Median : 72743886   Median :   39.605
##  Mean   :36.93                      Mean   :110134417   Mean   :   84.426
##  3rd Qu.:36.00                      3rd Qu.:131235705   3rd Qu.:  102.006
##  Max.   :94.00                      Max.   :455387636   Max.   :18235.559
##      FKL03             ICD10_Text          Kategori            PSTV13
##  Min.   :2019-01-01   Length:1761625     Length:1761625     Min.   :11.00
##  1st Qu.:2019-06-21   Class :character   Class :character   1st Qu.:31.00
##  Median :2019-11-27   Mode  :character   Mode  :character   Median :33.00
##  Mean   :2019-12-10                                        Mean   :36.93
##  3rd Qu.:2020-06-03                                        3rd Qu.:36.00
##  Max.   :2020-12-31                                        Max.   :94.00
##    Provinsi
##  Length:1761625
##  Class :character
##  Mode  :character
##
##
##
```

```r
cat("Informasi data:\n\n")
```

```
## Informasi data:
```

```r
str(data4)
```

```
## 'data.frame':    1761625 obs. of  9 variables:
##  $ FKL05     : dbl+lbl [1:1761625] 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11...
##    ..@ label       : chr "Provinsi FKRTL"
##    ..@ format.stata: chr "%26.0g"
##    ..@ labels      : Named num  11 12 13 14 15 16 17 18 19 21 ...
##    .. ..- attr(*, "names")= chr [1:35] "ACEH" "SUMATERA UTARA" "SUMATERA BARAT" "RIAU" ...
##  $ FKL15A    : chr  "H17" "A01" "K30" "I15" ...
##  $ PSTV01    : num  38061656 27701658 18167452 74055526 11764293 ...
##  $ PSTV15    : num  9.66 149.91 112.51 3.36 6.93 ...
##  $ FKL03     : Date, format: "2020-09-08" "2020-12-03" ...
##  $ ICD10_Text: chr  "H17 Corneal scars and opacities" "A01 Typhoid and paratyphoid fevers" "K30 Funct
##  $ Kategori  : chr  "Penyakit Lainnya" "Penyakit Lainnya" "Penyakit Lainnya" "Katastropik" ...
##  $ PSTV13    : int  11 11 11 11 11 11 11 11 11 11 ...
##  $ Provinsi  : chr  "Aceh" "Aceh" "Aceh" "Aceh" ...
```

```r
cat("total NA values:", sum(is.na(data4)))
```

```
## total NA values: 0
```

Checking blank row data kepesertaan

```r
cat("Summary data yang digunakan:\n\n")
```

```
## Summary data yang digunakan:
```

```r
summary(peserta3)
```

```
##      PSTV13          PSTV01              PSTV15              PSTV16
##  Min.   :11.00   Min.   :271321573   Min.   :    0.195   Min.   :2019
##  1st Qu.:19.00   1st Qu.:317167480   1st Qu.:    6.200   1st Qu.:2019
##  Median :33.00   Median :363273800   Median :   20.376   Median :2019
##  Mean   :38.05   Mean   :363291928   Mean   :  126.519   Mean   :2019
##  3rd Qu.:52.00   3rd Qu.:409334251   3rd Qu.:   63.190   3rd Qu.:2020
##  Max.   :94.00   Max.   :455388519   Max.   :15281.054   Max.   :2020
##      FKL05         Provinsi
##  Min.   :11.00   Length:229216
##  1st Qu.:19.00   Class :character
##  Median :33.00   Mode  :character
##  Mean   :38.05
##  3rd Qu.:52.00
##  Max.   :94.00
```

```r
cat("Informasi data:\n\n")
```

```
## Informasi data:
```

```r
str(peserta3)
```

```
## 'data.frame':    229216 obs. of  6 variables:
##  $ PSTV13  : dbl+lbl [1:229216] 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11,...
##    ..@ label       : chr "Provinsi Fasilitas Kesehatan Peserta Terdaftar"
##    ..@ format.stata: chr "%26.0g"
##    ..@ labels      : Named num  11 12 13 14 15 16 17 18 19 21 ...
##    .. ..- attr(*, "names")= chr [1:35] "ACEH" "SUMATERA UTARA" "SUMATERA BARAT" "RIAU" ...
##  $ PSTV01  : num  4.27e+08 2.90e+08 3.06e+08 3.14e+08 3.14e+08 ...
##  $ PSTV15  : num  2.96 38.17 44.51 604.25 7.45 ...
##  $ PSTV16  : num  2019 2019 2019 2019 2019 ...
##  $ FKL05   : int  11 11 11 11 11 11 11 11 11 11 ...
##  $ Provinsi: chr  "Aceh" "Aceh" "Aceh" "Aceh" ...
```

```r
cat("total NA values:", sum(is.na(peserta3)))
```

```
## total NA values: 0
```

Pengelompokkan data berdasarkan provinsi dan menghitung jumlah klaim tiap kategori per provinsi

```r
## variabel kepesertaan merupakan nilai sampel peserta bpjs kesehatan yang digunakan untuk transformasi
## nilai sampel(kepesertaan) = jumlah bobot peserta per provinsi X 1%
kepesertaan = peserta3 %>%
  group_by(Provinsi) %>%
  summarise(bobot = sum(PSTV15)*0.01)
head(kepesertaan)
```

```
## # A tibble: 6 x 2
##   Provinsi       bobot
##   <chr>          <dbl>
## 1 Aceh           2106.
## 2 Bali           3010.
## 3 Banten        11328.
## 4 Bengkulu       2244.
## 5 DI Yogyakarta  2352.
## 6 DKI Jakarta    4150.
```

Pengelompokkan data berdasarkan provinsi dan menghitung jumlah klaim tiap kategori per provinsi

```r
data5 = data4 %>% group_by(Provinsi) %>%
                  summarise(persalinan = sum(Kategori == 'Persalinan'),
                            kecelakaan = sum(Kategori == 'Kecelakaan'),
                            katastropik = sum(Kategori == 'Katastropik'),
                            penyakit_lainnya = sum(Kategori == 'Penyakit Lainnya'))
head(data5)
```

```
## # A tibble: 6 x 5
##   Provinsi      persalinan kecelakaan katastropik penyakit_lainnya
##   <chr>              <int>      <int>       <int>            <int>
## 1 Aceh                 557         20        5975            25842
## 2 Bali                 761          1       21555            49092
## 3 Banten               840          3       13045            42059
## 4 Bengkulu             473          1        3517            11720
## 5 DI Yogyakarta        154          0       17353            28916
## 6 DKI Jakarta          869          3       29605            62454
```

Transformasi Data

```r
## transformasi data dilakukan dengan membagi jumlah klaim dengan nilai sampel peserta dikali 100%
## Transformasi(datafinal) = Jumlah Klaim(data5) / Nilai Sampel(kepesertaan) X 100%
transformasi = merge(data5,kepesertaan,by="Provinsi")
head(transformasi)
```

```
##        Provinsi persalinan kecelakaan katastropik penyakit_lainnya     bobot
## 1          Aceh        557         20        5975            25842  2106.044
## 2          Bali        761          1       21555            49092  3010.275
## 3        Banten        840          3       13045            42059 11327.568
## 4      Bengkulu        473          1        3517            11720  2243.762
## 5 DI Yogyakarta        154          0       17353            28916  2352.397
## 6   DKI Jakarta        869          3       29605            62454  4150.243
```

```r
datafinal = transformasi %>% group_by(Provinsi) %>%
                             summarise(Klaim_Persalinan = (persalinan/bobot),
                                       Klaim_Kecelakaan = (kecelakaan/bobot),
                                       Klaim_Katastropik = (katastropik/bobot),
                                       Klaim_Penyakit_Lainnya = (penyakit_lainnya/bobot))
head(as.data.frame(datafinal))
```

```
##        Provinsi Klaim_Persalinan Klaim_Kecelakaan Klaim_Katastropik
## 1          Aceh       0.26447696     0.0094964797          2.837073
## 2          Bali       0.25280082     0.0003321956          7.160475
## 3        Banten       0.07415537     0.0002648406          1.151615
## 4      Bengkulu       0.21080669     0.0004456801          1.567457
## 5 DI Yogyakarta       0.06546513     0.0000000000          7.376730
## 6   DKI Jakarta       0.20938530     0.0007228492          7.133316
##   Klaim_Penyakit_Lainnya
## 1              12.270401
## 2              16.308144
## 3               3.712977
```

```
## 4               5.223371
## 5              12.292140
## 6              15.048274
```

Normalisasi Data

```
## normalisasi data dilakukan agar skala antar variabel tidak mendominasi variabel lain dan mengganggu
## normalisasi data yang dilakukan penulis menggunakan fungsi scale
normdata = as.data.frame(as.data.frame(datafinal[,-c(1,1)]))
row.names(normdata) = paste(datafinal$Provinsi)
datanormal = scale(normdata)
cat("Summary Data yang telah dinormalisasi:\n\n")
```

```
## Summary Data yang telah dinormalisasi:
```

```
summary(datanormal)
```

```
##  Klaim_Persalinan  Klaim_Kecelakaan  Klaim_Katastropik  Klaim_Penyakit_Lainnya
##  Min.   :-1.0507   Min.   :-0.4107   Min.   :-0.79770   Min.   :-0.9429
##  1st Qu.:-0.6915   1st Qu.:-0.4107   1st Qu.:-0.58358   1st Qu.:-0.6619
##  Median :-0.4167   Median :-0.2973   Median :-0.33568   Median :-0.4450
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.4840   3rd Qu.:-0.1969   3rd Qu.: 0.04229   3rd Qu.: 0.2196
##  Max.   : 3.7095   Max.   : 3.9073   Max.   : 2.94952   Max.   : 2.5935
```

```
cat("Informasi data:\n\n")
```

```
## Informasi data:
```

```
str(datanormal)
```

```
##  num [1:34, 1:4] 1.389 1.261 -0.695 0.801 -0.79 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:34] "Aceh" "Bali" "Banten" "Bengkulu" ...
##   ..$ : chr [1:4] "Klaim_Persalinan" "Klaim_Kecelakaan" "Klaim_Katastropik" "Klaim_Penyakit_Lainnya"
##  - attr(*, "scaled:center")= Named num [1:4] 0.137602 0.000968 1.772274 5.723547
##   ..- attr(*, "names")= chr [1:4] "Klaim_Persalinan" "Klaim_Kecelakaan" "Klaim_Katastropik" "Klaim_P
##  - attr(*, "scaled:scale")= Named num [1:4] 0.09135 0.00236 1.90013 4.08124
##   ..- attr(*, "names")= chr [1:4] "Klaim_Persalinan" "Klaim_Kecelakaan" "Klaim_Katastropik" "Klaim_P
```

```
cat("total NA values:", sum(is.na(datanormal)))
```

```
## total NA values: 0
```

```
cat("Cuplikan data yang telah dinormalisasi dan siap untuk diolah:\n\n")
```

```
## Cuplikan data yang telah dinormalisasi dan siap untuk diolah:
```

```
head(datanormal)
```

```
##               Klaim_Persalinan Klaim_Kecelakaan Klaim_Katastropik
## Aceh                 1.3888766        3.6183443         0.5603833
## Bali                 1.2610598       -0.2697947         2.8357067
## Banten              -0.6945428       -0.2983714        -0.3266409
## Bengkulu             0.8013569       -0.2216465        -0.1077914
## DI Yogyakarta       -0.7896735       -0.4107356         2.9495172
## DKI Jakarta          0.7857973       -0.1040517         2.8214136
##               Klaim_Penyakit_Lainnya
## Aceh                       1.6041334
## Bali                       2.5934754
## Banten                    -0.4926369
## Bengkulu                  -0.1225548
## DI Yogyakarta              1.6094600
## DKI Jakarta                2.2847775
```

Data Visualization and Modelling Data

Input Library

```
library(ggplot2)
library(scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(dbscan)
```

```
##
## Attaching package: 'dbscan'

## The following object is masked from 'package:stats':
##
##     as.dendrogram
```

```r
library(clusterSim)
```

```
## Loading required package: cluster
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(fpc)
```

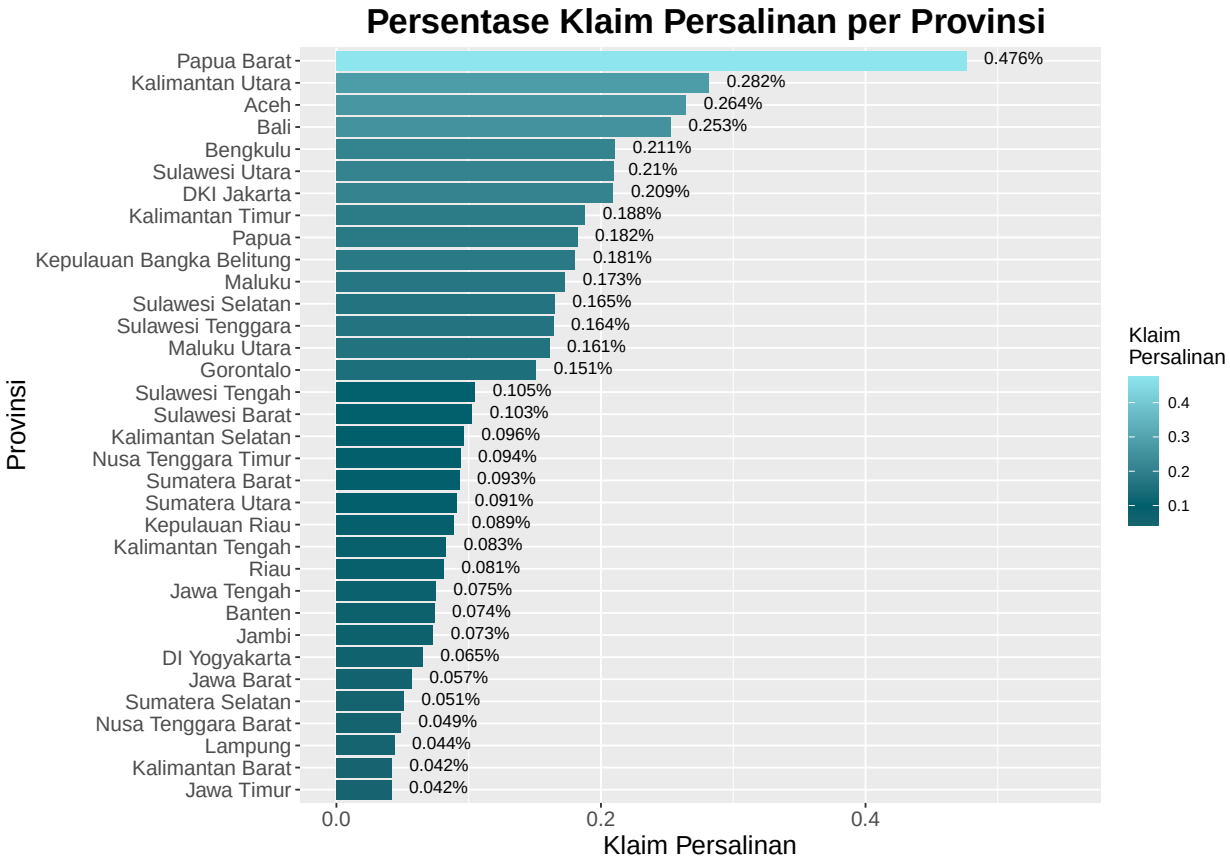
```
##
## Attaching package: 'fpc'
```

```
## The following object is masked from 'package:dbscan':
##
##     dbscan
```

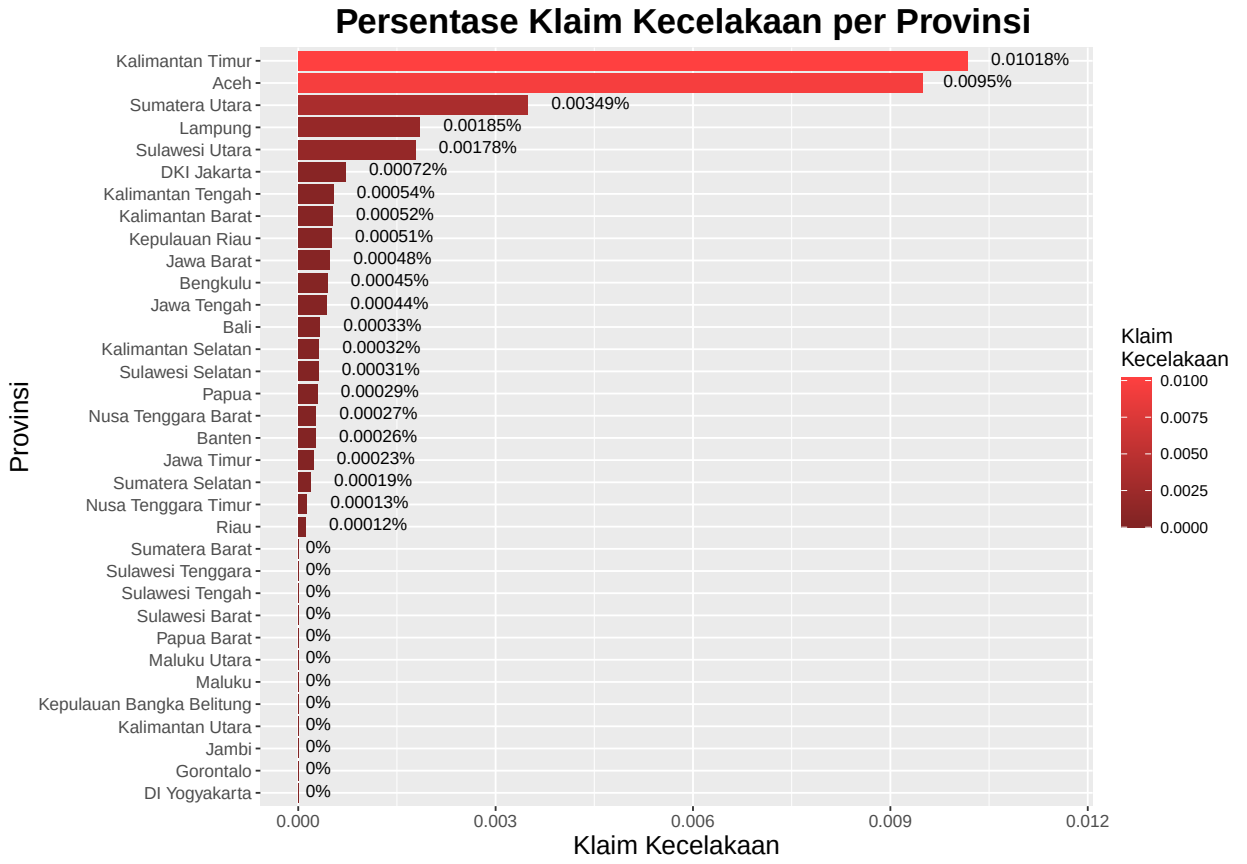
Statistika Daskriptif

```r
## filter data per variabel dari datafinal
persalinan1 = datafinal %>% dplyr::select(Provinsi, Klaim_Persalinan)
kecelakaan1 = datafinal %>% dplyr::select(Provinsi, Klaim_Kecelakaan)
katastropik1 = datafinal %>% dplyr::select(Provinsi, Klaim_Katastropik)
lainnya1 = datafinal %>% dplyr::select(Provinsi, Klaim_Penyakit_Lainnya)
```

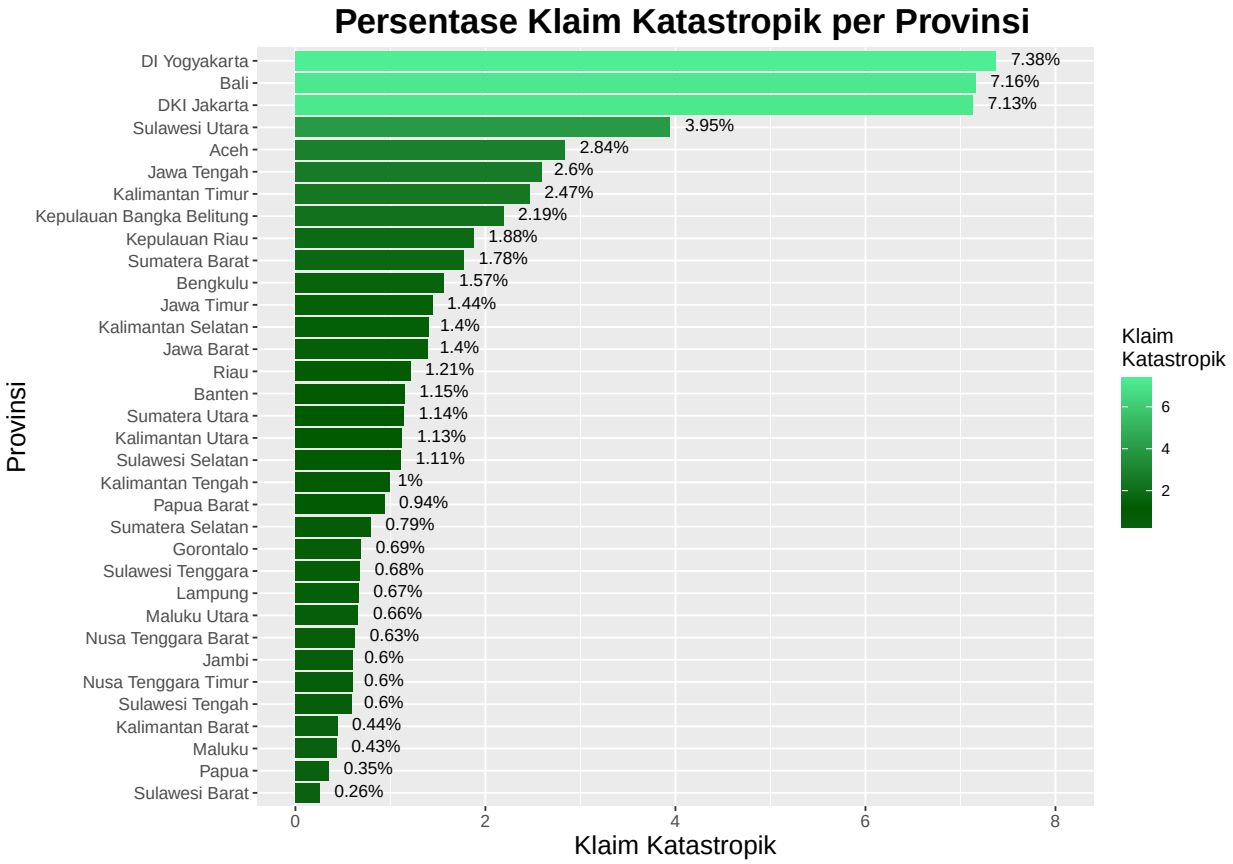
```r
## Barplot Klaim Persalinan
persalinan1$Provinsi <- factor(persalinan1$Provinsi, levels = persalinan1$Provinsi[order(persalinan1$Kl
Persalinan <- ggplot(persalinan1, aes(Provinsi, Klaim_Persalinan, fill = Klaim_Persalinan)) +
  labs (title = "Persentase Klaim Persalinan per Provinsi", x = "Provinsi", y = "Klaim Persalinan", fil
  coord_flip() +
  geom_bar(stat = "identity") +
  scale_fill_gradient2(position = "right", low = "cadetblue4", mid = muted("cadetblue3"), high = "cadetb
  geom_text(aes(label = paste0(round(Klaim_Persalinan,3),"%"), vjust = 0.3 ,hjust = -0.3), size = 3.5) +
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.55, vjust = 0.7),
        axis.title = element_text(size = 15, color = "Black"),
        axis.text = element_text(size = 12),
        legend.title = element_text(size = 12)) +
  ylim(0,0.55)
Persalinan
```



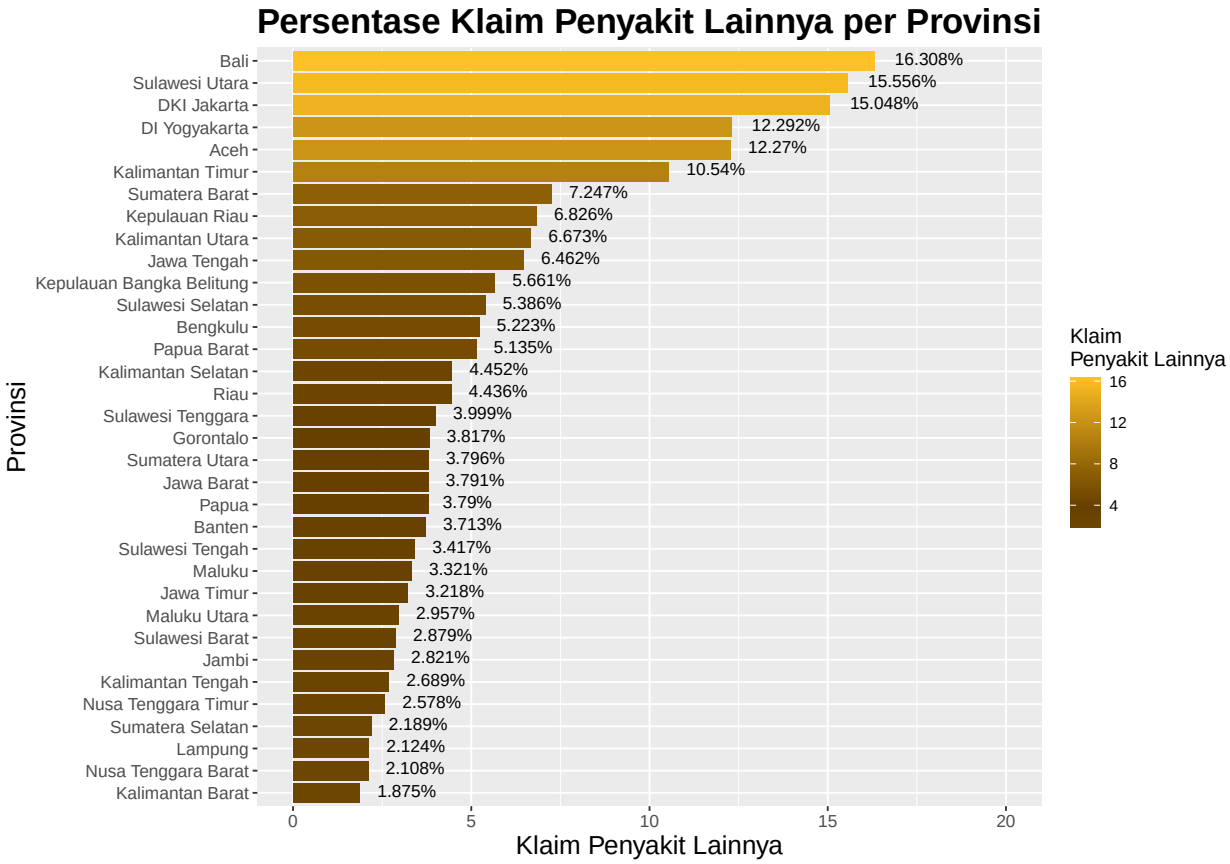

```r
## Barplot Klaim Kecelakaan
kecelakaan1$Provinsi <- factor(kecelakaan1$Provinsi, levels = kecelakaan1$Provinsi[order(kecelakaan1$Kl
Kecelakaan <- ggplot(kecelakaan1, aes(Provinsi, Klaim_Kecelakaan, fill = Klaim_Kecelakaan)) +
  labs (title = "Persentase Klaim Kecelakaan per Provinsi", x = "Provinsi", y = "Klaim Kecelakaan", fil
  coord_flip() +
  geom_bar(stat = "identity") +
  scale_fill_gradient2(position = "right", low = "brown4", mid = muted("brown3"), high = "brown1", midp
  geom_text(aes(label = paste0(round(Klaim_Kecelakaan,5),"%"), vjust = 0.3 ,hjust = -0.3), size = 3.5)
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.55, vjust = 0.7)
        axis.title = element_text(size = 15, color = "Black"),
        axis.text = element_text(size = 10),
        legend.title = element_text(size = 12)) +
  ylim(0,0.0115)
Kecelakaan
```




```
## Barplot Klaim Katastropik
katastropik1$Provinsi <- factor(katastropik1$Provinsi, levels = katastropik1$Provinsi[order(katastropik
Katastropik <- ggplot(katastropik1, aes(Provinsi, Klaim_Katastropik, fill = Klaim_Katastropik)) +
  labs (title = "Persentase Klaim Katastropik per Provinsi", x = "Provinsi", y = "Klaim Katastropik", fi
  coord_flip() +
  geom_bar(stat = "identity") +
  scale_fill_gradient2(position = "right", low = "seagreen4", mid = muted("seagreen3"), high = "seagreen
  geom_text(aes(label = paste0(round(Klaim_Katastropik,2),"%"), vjust = 0.3 ,hjust = -0.3), size = 3.5)
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.55, vjust = 0.7)
        axis.title = element_text(size = 15, color = "Black"),
        axis.text = element_text(size = 10),
        legend.title = element_text(size = 12)) +
  ylim(0,8)
Katastropik
```



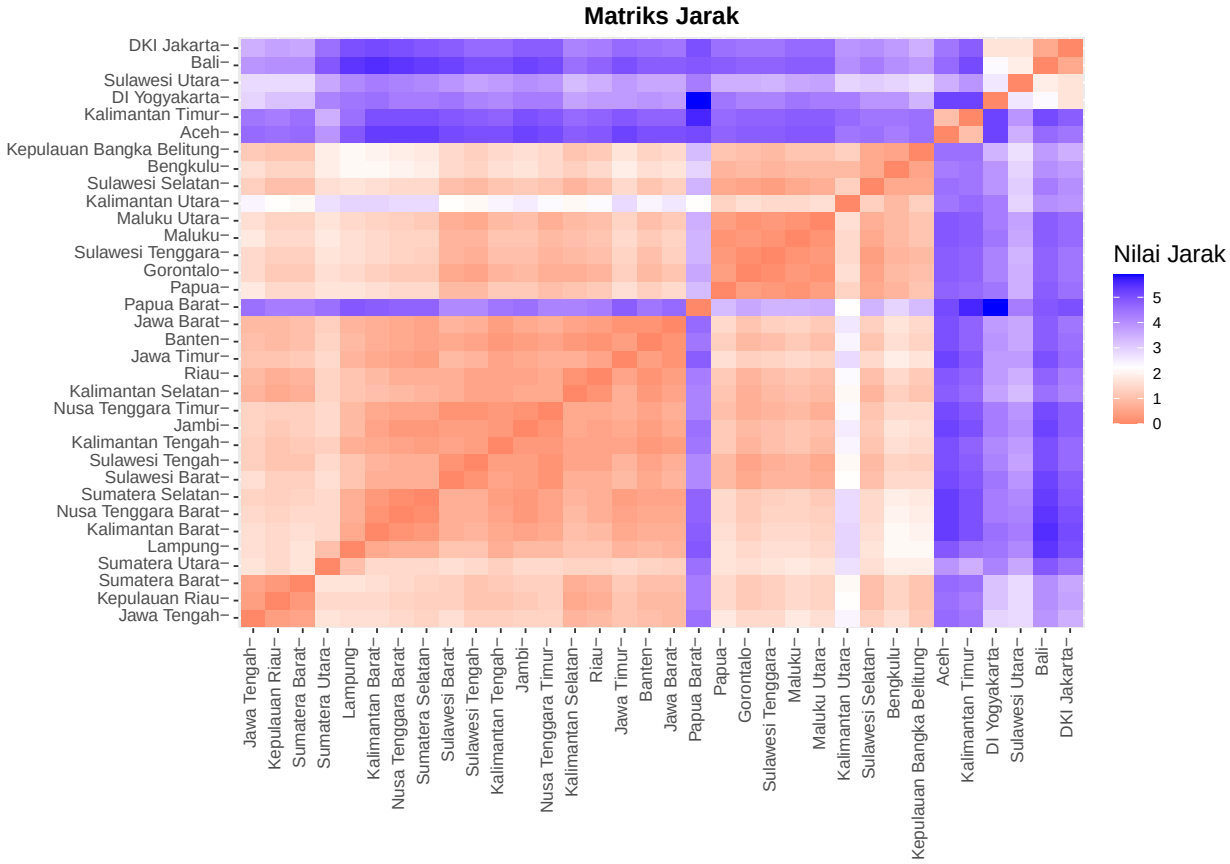

```
## Barplot Klaim Penyakit Lainnya
lainnya1$Provinsi <- factor(lainnya1$Provinsi, levels = lainnya1$Provinsi[order(lainnya1$Klaim_Penyakit_
Lainnya <- ggplot(lainnya1, aes(Provinsi, Klaim_Penyakit_Lainnya, fill = Klaim_Penyakit_Lainnya)) +
  labs (title = "Persentase Klaim Penyakit Lainnya per Provinsi", x = "Provinsi", y = "Klaim Penyakit La
  coord_flip() +
  geom_bar(stat = "identity") +
  scale_fill_gradient2(position = "right", low = "goldenrod4", mid = muted("goldenrod3"), high = "golde
  geom_text(aes(label = paste0(round(Klaim_Penyakit_Lainnya,3),"%"), vjust = 0.3 ,hjust = -0.3), size =
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.55, vjust = 0.7)
        axis.title = element_text(size = 15, color = "Black"),
        axis.text = element_text(size = 10),
        legend.title = element_text(size = 12)) +
  ylim(0,20)
Lainnya
```



Plot Matriks Jarak

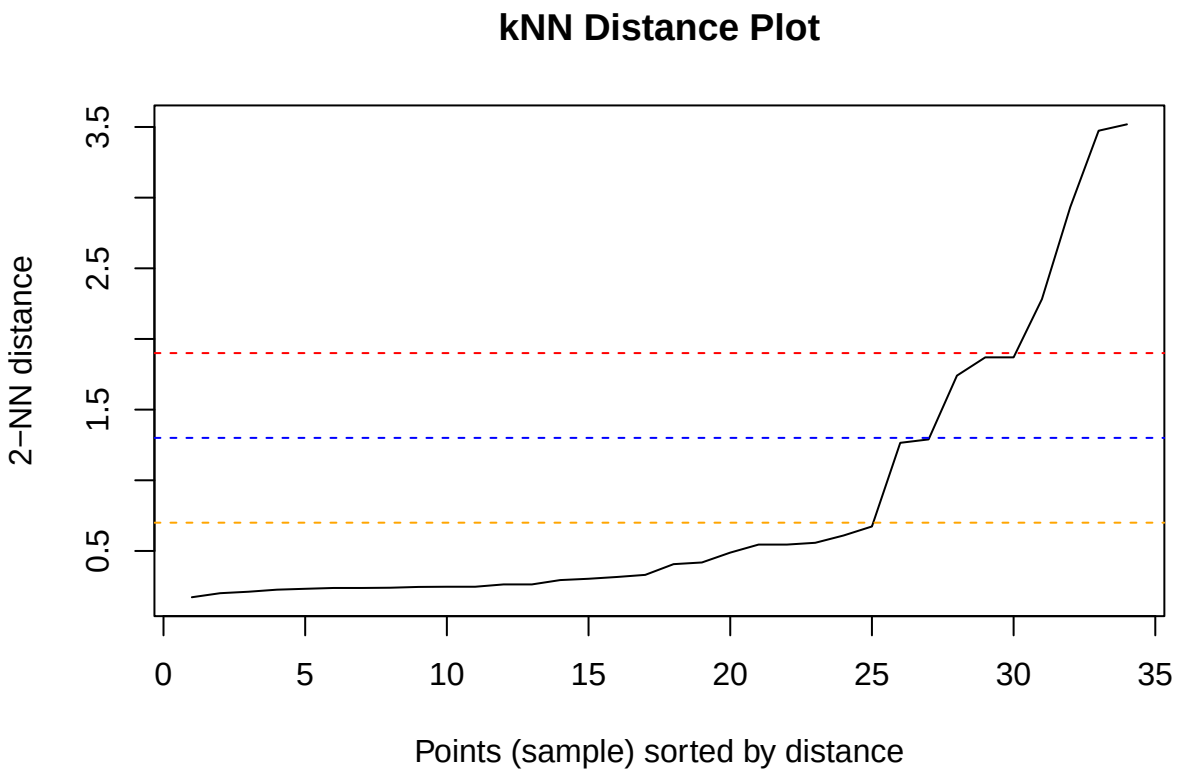
```
## menghitung jarak antar provinsi menggunakan metode euclidean distance
hitungjarak = round(dist(datanormal, method = "euclidean"),1)

## plot matriks jarak
matriksjarak = fviz_dist(hitungjarak) +
  labs(title = "Matriks Jarak", fill = "Nilai Jarak") +
  theme(plot.title = element_text(size = 14, color = "Black", face = "bold", hjust = 0.5, vjust = 0.8),
        axis.text.x = element_text(size = 10, vjust = 0.2, angle = 90),
        axis.text.y = element_text(size = 10, vjust = 0.25),
        legend.title = element_text(size = 14))
matriksjarak
```



kNNdisplot

```
## kNN displot digunakan untuk menentukan parameter epsilon yang nantinya digunakan untuk membentuk mod
kNNdistplot(datanormal, k = 2)
abline(h=1.9, col="red", lty=2)
abline(h=1.3, col="blue", lty=2)
abline(h=0.7, col="orange", lty=2)
title("kNN Distance Plot")
```

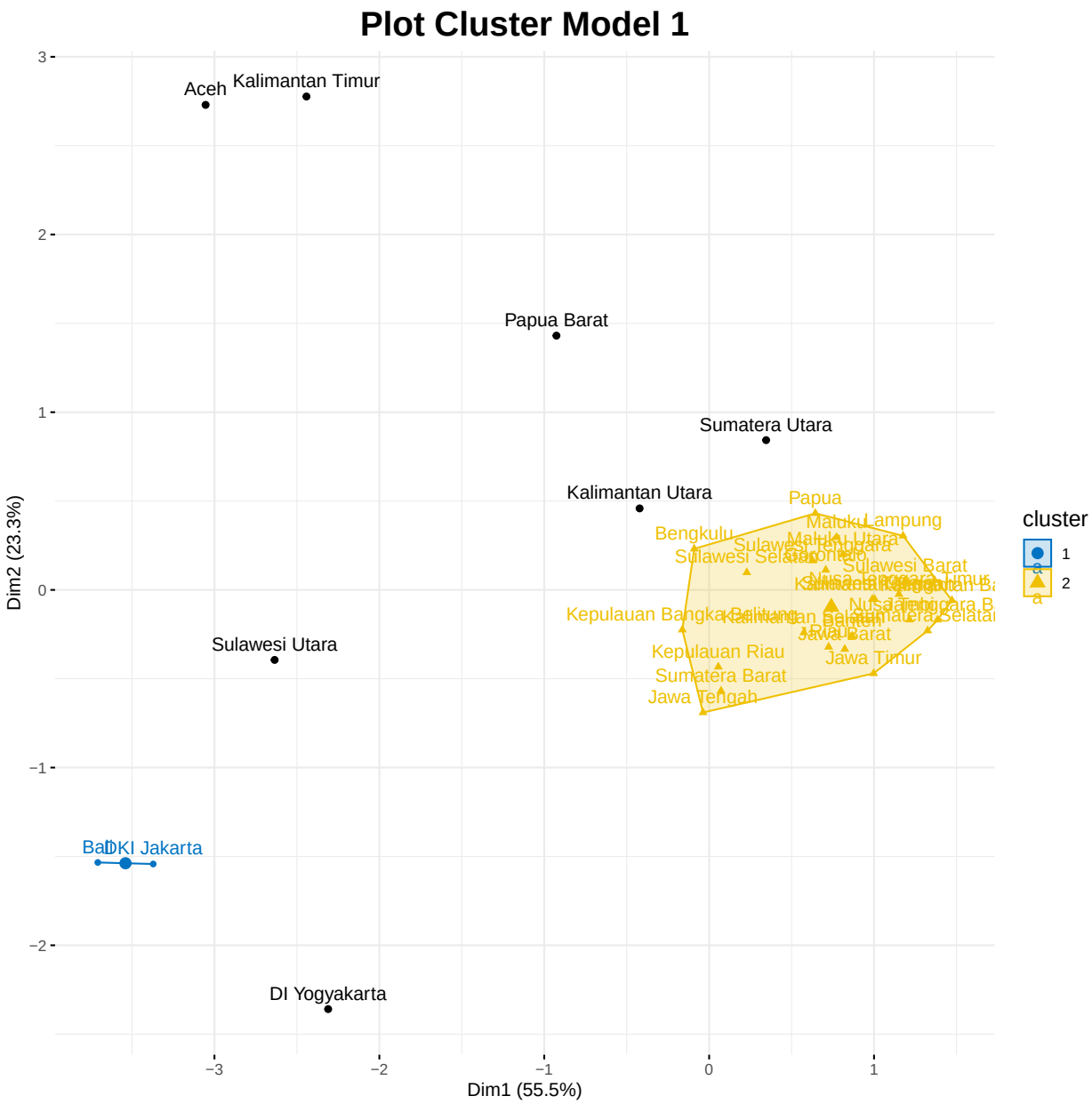



```
cat("Cara membaca kNNdisplot ialah dengan melihat lengkungan curam yang ada pada plot\nDidapati dari pl
```

```
## Cara membaca kNNdisplot ialah dengan melihat lengkungan curam yang ada pada plot
## Didapati dari plot diatas bahwa terdapat tiga lengkungan curam, yakni pada h1 = 0.7, h2 = 1.3, h3 =
## Nilai h yang didapat dari kNNdisplot akan digunakan sebagai parameter epsilon pada model DBSCAN dan
## terdapat tiga model, yakni:
## Model 1 (MinPts = 2, Eps = 0.7)
## Model 2 (MinPts = 2, Eps = 1.3)
## Model 3 (MinPts = 2, Eps = 1.9)
```

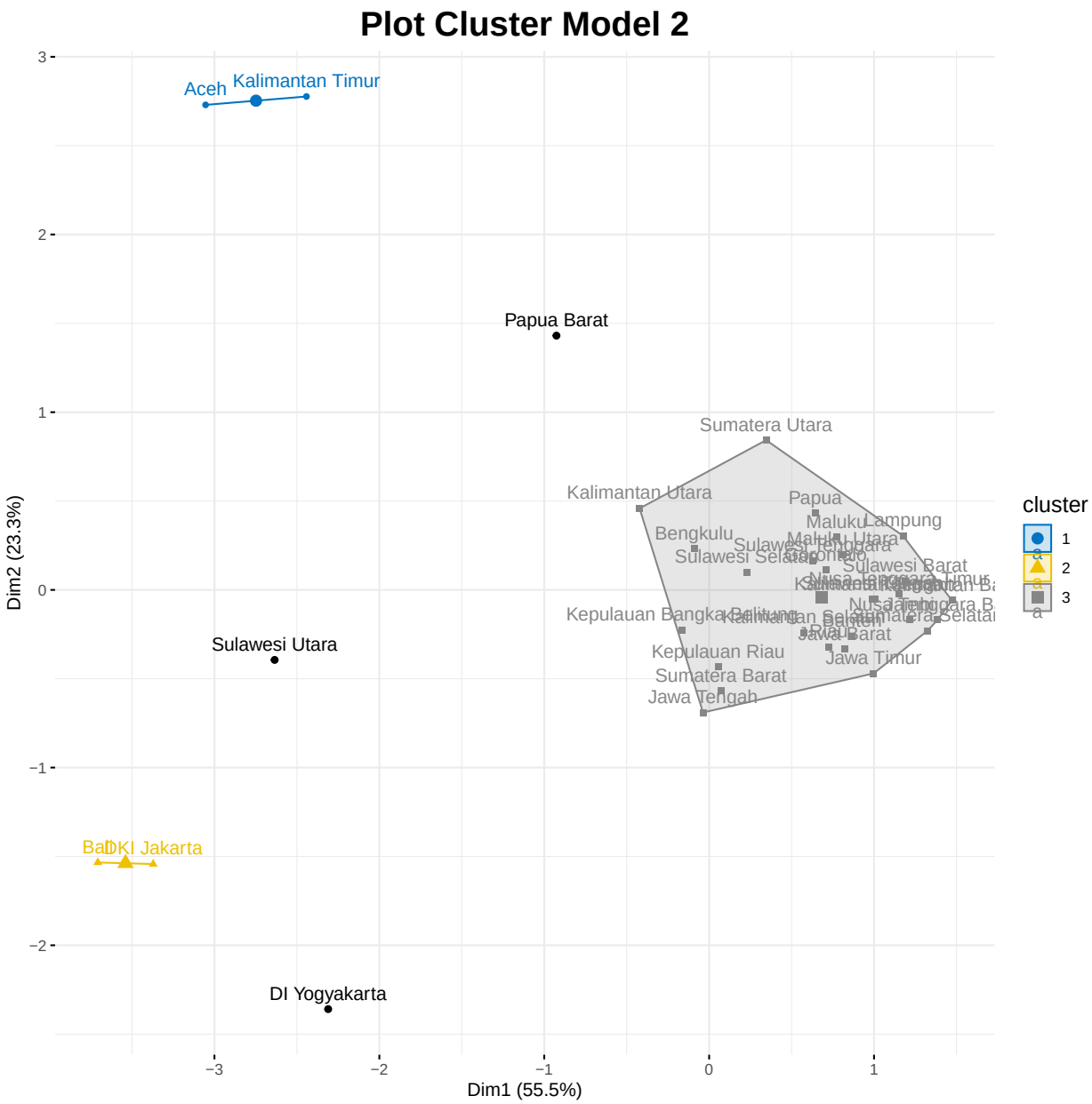
DBSCAN

```
## Plot Model 1
db_clust1 = dbscan(datanormal, eps = 0.7, MinPts = 2)
fviz_cluster(db_clust1, datanormal, palette="jco", ggtheme = theme_minimal()) +
labs(title = "Plot Cluster Model 1") +
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.5, vjust = 0.8),
        legend.title = element_text(size = 13))
```



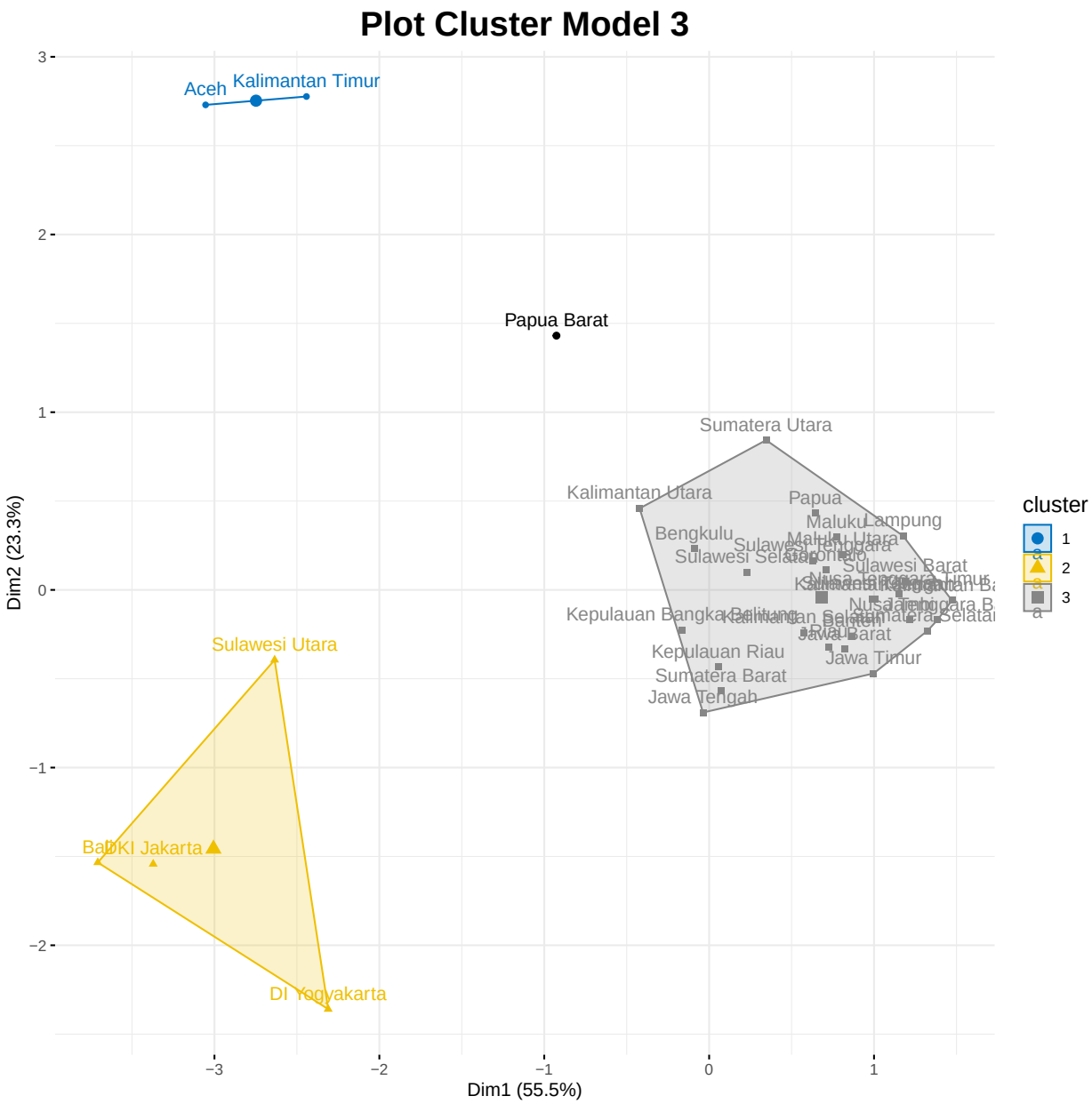

```
## Dunn Index, DB Index, Average Silhouette Index
db_index1 = index.DB(datanormal, as.integer(db_clust1$cluster) + 1)
clustering_indices1 = cluster.stats(dist(datanormal),as.integer(db_clust1$cluster) + 1)
dunn_index1 = clustering_indices1$dunn
sil1 = silhouette(as.integer(db_clust1$cluster) + 1, dist(datanormal))
avg_silhouette1 = mean(sil1[, "sil_width"])
```

```
## Plot Model 2
db_clust2 = dbscan(datanormal, eps = 1.3, MinPts = 2)
fviz_cluster(db_clust2, datanormal, palette="jco", ggtheme = theme_minimal()) +
labs(title = "Plot Cluster Model 2") +
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.5, vjust = 0.8),
        legend.title = element_text(size = 13))
```




```
## Dunn Index, DB Index, Average Silhouette Index
db_index2 = index.DB(datanormal, as.integer(db_clust2$cluster) + 1)
clustering_indices2 = cluster.stats(dist(datanormal),as.integer(db_clust2$cluster) + 1)
dunn_index2 = clustering_indices2$dunn
sil2 = silhouette(as.integer(db_clust2$cluster) + 1, dist(datanormal))
avg_silhouette2 = mean(sil2[, "sil_width"])
```

```
## Plot Model 3
db_clust3 = dbscan(datanormal, eps = 1.9, MinPts = 2)
fviz_cluster(db_clust3, datanormal, palette="jco", ggtheme = theme_minimal()) +
labs(title = "Plot Cluster Model 3") +
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.5, vjust = 0.8),
        legend.title = element_text(size = 13))
```




```
## Dunn Index, DB Index, Average Silhouette Index
db_index3 = index.DB(datanormal, as.integer(db_clust3$cluster) + 1)
clustering_indices3 = cluster.stats(dist(datanormal),as.integer(db_clust3$cluster) + 1)
dunn_index3 = clustering_indices3$dunn
sil3 = silhouette(as.integer(db_clust3$cluster) + 1, dist(datanormal))
avg_silhouette3 = mean(sil3[, "sil_width"])
```

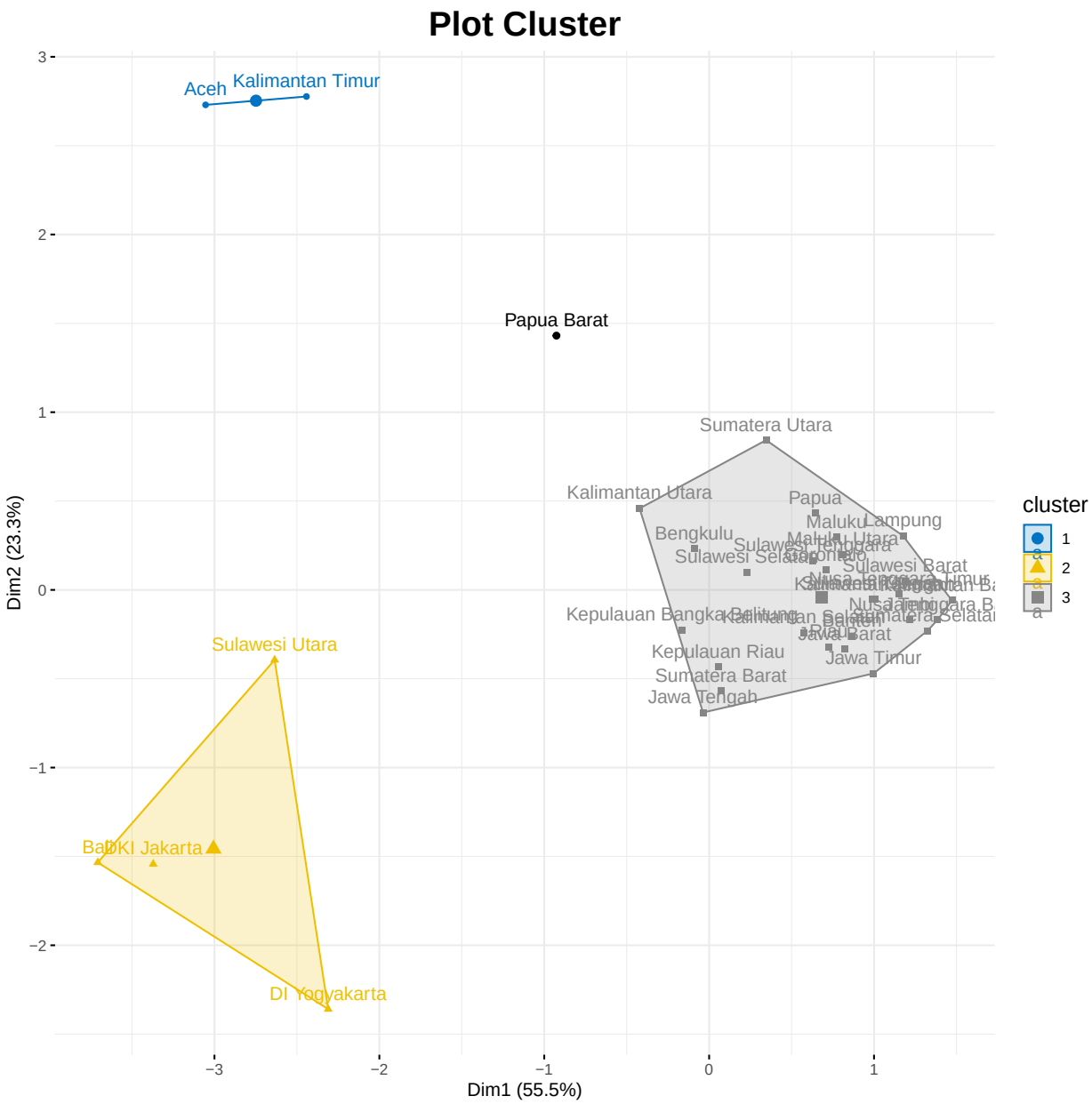
```
## Menentukan model terbaik
cat("Cara untuk menentukan model terbaik untuk metode DBSCAN ialah dengan melihat nilai DB Index, Dunn I
```

```
## Cara untuk menentukan model terbaik untuk metode DBSCAN ialah dengan melihat nilai DB Index, Dunn In
##
## Parameter DB Index: Semakin rendah nilai DB Index, semakin baik pula pengelompokkannya.
```

```
## Parameter Dunn Index: Semakin tinggi nilai Dunn Index, semakin baik pula pengelompokkannya.
## Parameter Average Silhouette Width: Semakin nilai Silhouette mendekati 1, maka semakin kuat model Clu
##
## Berikut ringkasan nilai DB Index, Dunn Index, dan Average Silhouette width pada masing-masing model y
##
## Model 1 :
## DB Index: 1.234361
## Dunn Index: 0.1534671
## Average Silhouette Width: 0.5257819
##
## Model 2 :
## DB Index: 1.222499
## Dunn Index: 0.2950617
## Average Silhouette Width: 0.6160388
##
## Model 3 :
## DB Index: 0.4113953
## Dunn Index: 0.7347957
## Average Silhouette Width: 0.6596688
##
## Berdasarkan nilai DB Index, Dunn Index, dan Average Silhouette Width pada masing-masing model, dapat
## Model 3
```

Hasil Cluster

```
## plot Cluster
fviz_cluster(db_clust3, datanormal, palette="jco", ggtheme = theme_minimal()) +
labs(title = "Plot Cluster") +
  theme(plot.title = element_text(size = 20, color = "Black", face = "bold", hjust = 0.5, vjust = 0.8),
        legend.title = element_text(size = 13))
```




```
## Hasil
Cluster = db_clust3$cluster
hasilklaster = datafinal %>% mutate(Cluster)
klaster0 = hasilklaster %>% filter(Cluster == 0)
klaster1 = hasilklaster %>% filter(Cluster == 1)
klaster2 = hasilklaster %>% filter(Cluster == 2)
klaster3 = hasilklaster %>% filter(Cluster == 3)
klaster00 = paste(klaster0$Provinsi)
klaster11 = paste(klaster1$Provinsi)
klaster22 = paste(klaster2$Provinsi)
klaster33 = paste(klaster3$Provinsi)
cat("Anggota Cluster 1:\n")
```

```
## Anggota Cluster 1:
```

```
klaster11
```

```
## [1] "Aceh"             "Kalimantan Timur"
```

```
cat("Anggota Cluster 2:\n")
```

```
## Anggota Cluster 2:
```

```
klaster22
```

```
## [1] "Bali"           "DI Yogyakarta"  "DKI Jakarta"    "Sulawesi Utara"
```

```
cat("Anggota Cluster 3:\n")
```

```
## Anggota Cluster 3:
```

```
klaster33
```

```
##  [1] "Banten"                   "Bengkulu"
##  [3] "Gorontalo"                "Jambi"
##  [5] "Jawa Barat"               "Jawa Tengah"
##  [7] "Jawa Timur"               "Kalimantan Barat"
##  [9] "Kalimantan Selatan"       "Kalimantan Tengah"
## [11] "Kalimantan Utara"         "Kepulauan Bangka Belitung"
## [13] "Kepulauan Riau"           "Lampung"
## [15] "Maluku"                   "Maluku Utara"
## [17] "Nusa Tenggara Barat"      "Nusa Tenggara Timur"
## [19] "Papua"                    "Riau"
## [21] "Sulawesi Barat"           "Sulawesi Selatan"
## [23] "Sulawesi Tengah"          "Sulawesi Tenggara"
## [25] "Sumatera Barat"           "Sumatera Selatan"
## [27] "Sumatera Utara"
```

```
cat("Anggota Cluster Noise: \n")
```

```
## Anggota Cluster Noise:
```

```
klaster00
```

```
## [1] "Papua Barat"
```