Ocean Data Analysis with R Programming for Early Career Ocean Professionals (ECOPs) (Asia)

Mohamad Lukman Aidid bin Mohd Yusoff

2023-11-05

Assignment. Lesson 2: MCA

1. Use the MCA() function to perform MCA on the data. What is the percentage of total variance explained by the first two principal components?
2. Use the factoextra package to create a variable correlation plot of the first two principal components. What variables are most important in explaining the first two components?

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- 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
```

```
library(stats)
library(FactoMineR)
```

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

```
library(factoextra)
```

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

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

```
setwd('C:/Users/Administrator/Desktop/R/')
```

Question 1:

```r
obis_malaysia <- read.csv("C:/Users/Administrator/Desktop/R/obis_malaysia/Occurrence.csv")
head(obis_malaysia)
```

```
##                                     id                              dataset_id
## 1 b268a4cb-4594-486e-947a-8d4aa1669b0c 2ed3cb38-e033-491d-95a7-5ae37f1f1904
## 2 15bce48e-3176-4025-b7c2-daea5f5624c4 2ed3cb38-e033-491d-95a7-5ae37f1f1904
## 3 dcc5cf4d-5426-47b8-8305-6dc591c5d4ae 36b58f28-0a03-4447-b688-e4eed56afa3d
## 4 9dd63c0c-054a-4e27-9da7-fa70728652b1 36b58f28-0a03-4447-b688-e4eed56afa3d
## 5 55c2a71b-f590-453f-ac41-25429f26cee3 7005b764-2feb-436a-8f17-0b41c6cd8435
## 6 11523180-c456-42b1-8a60-3f3217618f53 7005b764-2feb-436a-8f17-0b41c6cd8435
##   decimallongitude decimallatitude   date_start     date_mid     date_end
## 1         -74.0000         39.0000 1.057622e+12 1.057622e+12 1.057622e+12
## 2         -74.0000         39.0000 1.057622e+12 1.057622e+12 1.057622e+12
## 3         -66.7833         41.4667 8.983872e+11 8.983872e+11 8.983872e+11
## 4         -66.7833         41.4667 8.983872e+11 8.983872e+11 8.983872e+11
## 5          69.6830         22.4000 2.481408e+11 2.481408e+11 2.481408e+11
## 6          73.0000         10.0000 4.844448e+11 4.844448e+11 4.844448e+11
##   date_year        scientificname originalscientificname minimumdepthinmeters
## 1      2003 Merluccius bilinearis  Merluccius bilinearis                   NA
## 2      2003 Merluccius bilinearis  Merluccius bilinearis                   NA
## 3      1998  Calanus finmarchicus   Calanus finmarchicus                    0
## 4      1998  Calanus finmarchicus   Calanus finmarchicus                    0
## 5      1977            Ulva linza     Enteromorpha linza                   NA
## 6      1985            Ulva linza     Enteromorpha linza                   NA
##   maximumdepthinmeters depth coordinateuncertaintyinmeters      flags dropped
## 1                   NA    NA                           NA {NO_DEPTH}       0
## 2                   NA    NA                           NA {NO_DEPTH}       0
## 3                   50    25                           NA         {}       0
## 4                   50    25                           NA         {}       0
## 5                   NA    NA                           NA                  0
## 6                   NA    NA                           NA                  0
##   absence shoredistance bathymetry   sst   sss marine brackish freshwater
## 1       0         53558         37 15.17 32.96      1        0          0
## 2       0         53558         37 15.17 32.96      1        0          0
## 3       0        222884         70 11.12 32.42      1        0          0
## 4       0        222884         70 11.12 32.42      1        0          0
## 5       0           837         -1 26.73 35.05      1       NA         NA
## 6       0         69042       2012 28.89 34.95      1       NA         NA
##   terrestrial taxonrank aphiaid redlist_category superdomain domain  kingdom
## 1           0   Species  158962               NT       Biota     NA Animalia
## 2           0   Species  158962               NT       Biota     NA Animalia
## 3           0   Species  104464                        Biota     NA Animalia
## 4           0   Species  104464                        Biota     NA Animalia
## 5          NA   Species  234474                        Biota     NA  Plantae
## 6          NA   Species  234474                        Biota     NA  Plantae
##      subkingdom infrakingdom     phylum phylum_division subphylum_subdivision
## 1                            Chordata
## 2                            Chordata
## 3                           Arthropoda
## 4                           Arthropoda
## 5 Viridiplantae                                Chlorophyta         Chlorophytina
## 6 Viridiplantae                                Chlorophyta         Chlorophytina
##    subphylum   infraphylum   parvphylum      gigaclass megaclass     superclass
```

```
## 1 Vertebrata Gnathostomata Osteichthyes Actinopterygii              Actinopteri
## 2 Vertebrata Gnathostomata Osteichthyes Actinopterygii              Actinopteri
## 3  Crustacea                                                     Multicrustacea
## 4  Crustacea                                                     Multicrustacea
## 5
## 6
##         class  subclass  infraclass subterclass superorder      order suborder
## 1   Teleostei Teleostei                                     Gadiformes
## 2   Teleostei Teleostei                                     Gadiformes
## 3    Copepoda           Neocopepoda              Gymnoplea  Calanoida
## 4    Copepoda           Neocopepoda              Gymnoplea  Calanoida
## 5 Ulvophyceae                                                  Ulvales
## 6 Ulvophyceae                                                  Ulvales
##   infraorder parvorder superfamily       family    subfamily supertribe tribe
## 1                                  Merlucciidae Merlucciinae         NA
## 2                                  Merlucciidae Merlucciinae         NA
## 3                                     Calanidae                      NA
## 4                                     Calanidae                      NA
## 5                                      Ulvaceae                      NA
## 6                                      Ulvaceae                      NA
##   subtribe      genus subgenus section subsection series              species
## 1       NA Merluccius                                  NA Merluccius bilinearis
## 2       NA Merluccius                                  NA Merluccius bilinearis
## 3       NA    Calanus                                  NA  Calanus finmarchicus
## 4       NA    Calanus                                  NA  Calanus finmarchicus
## 5       NA       Ulva                                  NA            Ulva linza
## 6       NA       Ulva                                  NA            Ulva linza
##   subspecies natio variety subvariety forma subforma type modified language
## 1               NA                 NA             NA    NA                NA
## 2               NA                 NA             NA    NA                NA
## 3               NA                 NA             NA    NA                NA
## 4               NA                 NA             NA    NA                NA
## 5               NA                 NA             NA    NA                NA
## 6               NA                 NA             NA    NA                NA
##   license rightsholder accessrights bibliographiccitation references
## 1      NA           NA           NA                                NA
## 2      NA           NA           NA                                NA
## 3      NA           NA           NA                                NA
## 4      NA           NA           NA                                NA
## 5      NA           NA           NA                                NA
## 6      NA           NA           NA                                NA
##   institutionid collectionid datasetid institutioncode collectioncode
## 1            NA           NA        NA           RUMFS   OTTE-03-0037
## 2            NA           NA        NA           RUMFS   OTTE-03-0038
## 3            NA           NA        NA         Zoogene
## 4            NA           NA        NA         Zoogene
## 5            NA           NA        NA             NIO            168
## 6            NA           NA        NA             NIO            152
##   datasetname ownerinstitutioncode    basisofrecord informationwithheld
## 1                              NA HumanObservation                  NA
## 2                              NA HumanObservation                  NA
## 3                              NA HumanObservation                  NA
## 4                              NA HumanObservation                  NA
## 5                              NA HumanObservation                  NA
```

```
## 6                                NA HumanObservation                  NA
##   datageneralizations          dynamicproperties materialsampleid occurrenceid
## 1                  NA observedindividualcount=1;               NA
## 2                  NA observedindividualcount=1;               NA
## 3                  NA                                          NA
## 4                  NA                                          NA
## 5                  NA                                          NA
## 6                  NA                                          NA
##   catalognumber
## 1
## 2
## 3
## 4
## 5        NIO186
## 6        NIO167
##
## 1 location=OT 2;gear=OTTE;mesh=6;length=4;areasampled=1;heading=1;towdir=;sudo=7.8;bodo=6.78;susal=3
## 2       location=STA15;gear=OTTE;mesh=6;length=4;areasampled=1;heading=3;towdir=UP CRK;sudo=6.54;bo
## 3
## 4
## 5
## 6
##   recordnumber             recordedby recordedbyid individualcount
## 1                                            NA              NA
## 2                                            NA              NA
## 3                                            NA              NA
## 4                                            NA              NA
## 5            6 Dr. Arvind G. Untawale        NA              NA
## 6            0                Unknown        NA              NA
##   organismquantity organismquantitytype  sex lifestage reproductivecondition
## 1               NA                   NA             NA                    NA
## 2               NA                   NA             NA                    NA
## 3               NA                   NA             NA                    NA
## 4               NA                   NA             NA                    NA
## 5               NA                   NA Male        NA                    NA
## 6               NA                   NA Male        NA                    NA
##   behavior establishmentmeans occurrencestatus preparations disposition
## 1       NA                 NA
## 2       NA                 NA
## 3       NA                 NA
## 4       NA                 NA
## 5       NA                 NA
## 6       NA                 NA
##   othercatalognumbers associatedmedia associatedreferences associatedsequences
## 1                  NA              NA                                      NA
## 2                  NA              NA                                      NA
## 3                  NA              NA                                      NA
## 4                  NA              NA                                      NA
## 5                  NA              NA                                      NA
## 6                  NA              NA                                      NA
##   associatedtaxa organismid organismname organismscope associatedoccurrences
## 1             NA         NA           NA             NA                    NA
## 2             NA         NA           NA             NA                    NA
## 3             NA         NA           NA             NA                    NA
```

```
## 4             NA         NA           NA            NA                    NA
## 5             NA         NA           NA            NA                    NA
## 6             NA         NA           NA            NA                    NA
##   associatedorganisms previousidentifications organismremarks eventid
## 1                  NA                     NA
## 2                  NA                     NA
## 3                  NA                     NA
## 4                  NA                     NA
## 5                  NA                     NA
## 6                  NA                     NA
##   parenteventid samplingprotocol samplesizevalue samplesizeunit samplingeffort
## 1            NA                              NA             NA             NA
## 2            NA                              NA             NA             NA
## 3            NA                              NA             NA             NA
## 4            NA                              NA             NA             NA
## 5            NA                              NA             NA             NA
## 6            NA                              NA             NA             NA
##              eventdate eventtime startdayofyear enddayofyear year month day
## 1 2003-07-08T12:57:00Z        NA             NA           NA 2003     7   8
## 2 2003-07-08T12:57:00Z        NA             NA           NA 2003     7   8
## 3 1998-06-21T12:00:00Z        NA             NA           NA 1998     6  21
## 4 1998-06-21T12:00:00Z        NA             NA           NA 1998     6  21
## 5 1977-11-12T12:00:00Z        NA             NA           NA 1977    11  12
## 6 1985-05-09T12:00:00Z        NA             NA           NA 1985     5   9
##   verbatimeventdate habitat fieldnumber fieldnotes eventremarks locationid
## 1                                   NA         NA           NA         NA
## 2                                   NA         NA           NA         NA
## 3                                   NA         NA           NA         NA
## 4                                   NA         NA           NA         NA
## 5                                  155         NA           NA         NA
## 6                                  139         NA           NA         NA
##   highergeographyid highergeography continent waterbody islandgroup island
## 1                NA              NA        NA                    NA     NA
## 2                NA              NA        NA                    NA     NA
## 3                NA              NA        NA                    NA     NA
## 4                NA              NA        NA                    NA     NA
## 5                NA              NA        NA                    NA     NA
## 6                NA              NA        NA                    NA     NA
##   country countrycode stateprovince county municipality    locality
## 1              NA                                   NA
## 2              NA                                   NA
## 3              NA                                   NA
## 4              NA                                   NA
## 5   India          NA       Gujarat                 NA     Pirotan
## 6   India          NA   Lakshadweep                 NA Lakshadweep
##   verbatimlocality verbatimelevation minimumelevationinmeters
## 1               NA                NA                      NA
## 2               NA                NA                      NA
## 3               NA                NA                      NA
## 4               NA                NA                      NA
## 5               NA                NA                      NA
## 6               NA                NA                      NA
##   maximumelevationinmeters verbatimdepth minimumdistanceabovesurfaceinmeters
## 1                       NA            NA                                 NA
```

```
## 2                       NA            NA                                  NA
## 3                       NA            NA                                  NA
## 4                       NA            NA                                  NA
## 5                       NA            NA                                  NA
## 6                       NA            NA                                  NA
##   maximumdistanceabovesurfaceinmeters locationaccordingto locationremarks
## 1                                  NA                  NA              NA
## 2                                  NA                  NA              NA
## 3                                  NA                  NA              NA
## 4                                  NA                  NA              NA
## 5                                  NA                  NA              NA
## 6                                  NA                  NA              NA
##   verbatimcoordinates verbatimlatitude verbatimlongitude
## 1                  NA               NA                NA
## 2                  NA               NA                NA
## 3                  NA               NA                NA
## 4                  NA               NA                NA
## 5                  NA               NA                NA
## 6                  NA               NA                NA
##   verbatimcoordinatesystem verbatimsrs geodeticdatum coordinateprecision
## 1                       NA          NA                                NA
## 2                       NA          NA                                NA
## 3                       NA          NA                                NA
## 4                       NA          NA                                NA
## 5                       NA          NA                                NA
## 6                       NA          NA                                NA
##   pointradiusspatialfit footprintwkt footprintsrs footprintspatialfit
## 1                    NA                       NA                  NA
## 2                    NA                       NA                  NA
## 3                    NA                       NA                  NA
## 4                    NA                       NA                  NA
## 5                    NA                       NA                  NA
## 6                    NA                       NA                  NA
##   georeferencedby georeferenceddate georeferenceprotocol georeferencesources
## 1              NA                NA                   NA                   NA
## 2              NA                NA                   NA                   NA
## 3              NA                NA                   NA                   NA
## 4              NA                NA                   NA                   NA
## 5              NA                NA                   NA                   NA
## 6              NA                NA                   NA                   NA
##   georeferenceverificationstatus georeferenceremarks geologicalcontextid
## 1                             NA                  NA                  NA
## 2                             NA                  NA                  NA
## 3                             NA                  NA                  NA
## 4                             NA                  NA                  NA
## 5                             NA                  NA                  NA
## 6                             NA                  NA                  NA
##   earliesteonorlowesteonothem latesteonorhighesteonothem
## 1                          NA                         NA
## 2                          NA                         NA
## 3                          NA                         NA
## 4                          NA                         NA
## 5                          NA                         NA
## 6                          NA                         NA
```

```
##   earliesteraorlowesterathem latesteraorhighesterathem
## 1                         NA                        NA
## 2                         NA                        NA
## 3                         NA                        NA
## 4                         NA                        NA
## 5                         NA                        NA
## 6                         NA                        NA
##   earliestperiodorlowestsystem latestperiodorhighestsystem
## 1                           NA                          NA
## 2                           NA                          NA
## 3                           NA                          NA
## 4                           NA                          NA
## 5                           NA                          NA
## 6                           NA                          NA
##   earliestepochorlowestseries latestepochorhighestseries
## 1                          NA                         NA
## 2                          NA                         NA
## 3                          NA                         NA
## 4                          NA                         NA
## 5                          NA                         NA
## 6                          NA                         NA
##   earliestageorloweststage latestageorhigheststage lowestbiostratigraphiczone
## 1                       NA                      NA                         NA
## 2                       NA                      NA                         NA
## 3                       NA                      NA                         NA
## 4                       NA                      NA                         NA
## 5                       NA                      NA                         NA
## 6                       NA                      NA                         NA
##   highestbiostratigraphiczone lithostratigraphicterms group formation member
## 1                          NA                      NA    NA        NA     NA
## 2                          NA                      NA    NA        NA     NA
## 3                          NA                      NA    NA        NA     NA
## 4                          NA                      NA    NA        NA     NA
## 5                          NA                      NA    NA        NA     NA
## 6                          NA                      NA    NA        NA     NA
##   bed identificationid identifiedby identifiedbyid dateidentified
## 1  NA               NA                         NA
## 2  NA               NA                         NA
## 3  NA               NA                         NA
## 4  NA               NA                         NA
## 5  NA               NA      Unknown            NA
## 6  NA               NA      Unknown            NA
##   identificationreferences identificationremarks identificationqualifier
## 1                       NA                     NA                       NA
## 2                       NA                     NA                       NA
## 3                       NA                     NA                       NA
## 4                       NA                     NA                       NA
## 5                       NA                     NA                       NA
## 6                       NA                     NA                       NA
##   identificationverificationstatus typestatus taxonid
## 1                               NA           NA
## 2                               NA           NA
## 3                               NA           NA
## 4                               NA           NA
```

```
## 5                              NA                NA
## 6                              NA                NA
##                          scientificnameid acceptednameusageid
## 1 urn:lsid:marinespecies.org:taxname:158962                 NA
## 2 urn:lsid:marinespecies.org:taxname:158962                 NA
## 3 urn:lsid:marinespecies.org:taxname:104464                 NA
## 4 urn:lsid:marinespecies.org:taxname:104464                 NA
## 5 urn:lsid:marinespecies.org:taxname:145967                 NA
## 6 urn:lsid:marinespecies.org:taxname:145967                 NA
##   parentnameusageid originalnameusageid nameaccordingtoid namepublishedinid
## 1                NA                  NA                NA                NA
## 2                NA                  NA                NA                NA
## 3                NA                  NA                NA                NA
## 4                NA                  NA                NA                NA
## 5                NA                  NA                NA                NA
## 6                NA                  NA                NA                NA
##   taxonconceptid acceptednameusage parentnameusage originalnameusage
## 1             NA                NA               NA                NA
## 2             NA                NA               NA                NA
## 3             NA                NA               NA                NA
## 4             NA                NA               NA                NA
## 5             NA                NA               NA                NA
## 6             NA                NA               NA                NA
##   nameaccordingto namepublishedin namepublishedinyear higherclassification
## 1              NA              NA                   NA                    NA
## 2              NA              NA                   NA                    NA
## 3              NA              NA                   NA                    NA
## 4              NA              NA                   NA                    NA
## 5              NA              NA                   NA                    NA
## 6              NA              NA                   NA                    NA
##   specificepithet infraspecificepithet verbatimtaxonrank
## 1                                                    NA
## 2                                                    NA
## 3                                                    NA
## 4                                                    NA
## 5                                                    NA
## 6                                                    NA
##   scientificnameauthorship vernacularname nomenclaturalcode taxonomicstatus
## 1                                                       NA              NA
## 2                                                       NA              NA
## 3                                                       NA              NA
## 4                                                       NA              NA
## 5                J. Agardh                              NA              NA
## 6                J. Agardh                              NA              NA
##   nomenclaturalstatus taxonremarks
## 1                  NA           NA
## 2                  NA           NA
## 3                  NA           NA
## 4                  NA           NA
## 5                  NA           NA
## 6                  NA           NA
```

```
str(obis_malaysia)
```

```
## 'data.frame':    51974 obs. of  226 variables:
##  $ id                                : chr  "b268a4cb-4594-486e-947a-8d4aa1669b0c" "15bce48e-3176-4
##  $ dataset_id                        : chr  "2ed3cb38-e033-491d-95a7-5ae37f1f1904" "2ed3cb38-e033-4
##  $ decimallongitude                  : num  -74 -74 -66.8 -66.8 69.7 ...
##  $ decimallatitude                   : num  39 39 41.5 41.5 22.4 ...
##  $ date_start                        : num  1.06e+12 1.06e+12 8.98e+11 8.98e+11 2.48e+11 ...
##  $ date_mid                          : num  1.06e+12 1.06e+12 8.98e+11 8.98e+11 2.48e+11 ...
##  $ date_end                          : num  1.06e+12 1.06e+12 8.98e+11 8.98e+11 2.48e+11 ...
##  $ date_year                         : int  2003 2003 1998 1998 1977 1985 1985 2006 NA 2007 ...
##  $ scientificname                    : chr  "Merluccius bilinearis" "Merluccius bilinearis" "Calanus
##  $ originalscientificname            : chr  "Merluccius bilinearis" "Merluccius bilinearis" "Calanus
##  $ minimumdepthinmeters              : num  NA NA 0 0 NA NA NA NA NA NA ...
##  $ maximumdepthinmeters              : num  NA NA 50 50 NA NA NA NA NA NA ...
##  $ depth                             : num  NA NA 25 25 NA NA NA NA NA NA ...
##  $ coordinateuncertaintyinmeters     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ flags                             : chr  "{NO_DEPTH}" "{NO_DEPTH}" "{}" "{}" ...
##  $ dropped                           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ absence                           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ shoredistance                     : int  53558 53558 222884 222884 837 69042 69042 -1328 83 -555
##  $ bathymetry                        : num  37 37 70 70 -1 ...
##  $ sst                               : num  15.2 15.2 11.1 11.1 26.7 ...
##  $ sss                               : num  33 33 32.4 32.4 35 ...
##  $ marine                            : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ brackish                          : int  0 0 0 0 NA NA NA 0 NA NA ...
##  $ freshwater                        : int  0 0 0 0 NA NA NA 0 NA NA ...
##  $ terrestrial                       : int  0 0 0 0 NA NA NA 0 NA NA ...
##  $ taxonrank                         : chr  "Species" "Species" "Species" "Species" ...
##  $ aphiaid                           : int  158962 158962 104464 104464 234474 234474 234474 510565
##  $ redlist_category                  : chr  "NT" "NT" "" "" ...
##  $ superdomain                       : chr  "Biota" "Biota" "Biota" "Biota" ...
##  $ domain                            : logi  NA NA NA NA NA NA ...
##  $ kingdom                           : chr  "Animalia" "Animalia" "Animalia" "Animalia" ...
##  $ subkingdom                        : chr  "" "" "" "" ...
##  $ infrakingdom                      : chr  "" "" "" "" ...
##  $ phylum                            : chr  "Chordata" "Chordata" "Arthropoda" "Arthropoda" ...
##  $ phylum_division                   : chr  "" "" "" "" ...
##  $ subphylum_subdivision             : chr  "" "" "" "" ...
##  $ subphylum                         : chr  "Vertebrata" "Vertebrata" "Crustacea" "Crustacea" ...
##  $ infraphylum                       : chr  "Gnathostomata" "Gnathostomata" "" "" ...
##  $ parvphylum                        : chr  "Osteichthyes" "Osteichthyes" "" "" ...
##  $ gigaclass                         : chr  "Actinopterygii" "Actinopterygii" "" "" ...
##  $ megaclass                         : chr  "" "" "" "" ...
##  $ superclass                        : chr  "Actinopteri" "Actinopteri" "Multicrustacea" "Multicrus
##  $ class                             : chr  "Teleostei" "Teleostei" "Copepoda" "Copepoda" ...
##  $ subclass                          : chr  "Teleostei" "Teleostei" "" "" ...
##  $ infraclass                        : chr  "" "" "Neocopepoda" "Neocopepoda" ...
##  $ subterclass                       : chr  "" "" "" "" ...
##  $ superorder                        : chr  "" "" "Gymnoplea" "Gymnoplea" ...
##  $ order                             : chr  "Gadiformes" "Gadiformes" "Calanoida" "Calanoida" ...
##  $ suborder                          : chr  "" "" "" "" ...
##  $ infraorder                        : chr  "" "" "" "" ...
##  $ parvorder                         : chr  "" "" "" "" ...
##  $ superfamily                       : chr  "" "" "" "" ...
##  $ family                            : chr  "Merlucciidae" "Merlucciidae" "Calanidae" "Calanidae" .
```

```
##  $ subfamily                         : chr  "Merlucciinae" "Merlucciinae" "" "" ...
##  $ supertribe                        : logi  NA NA NA NA NA NA ...
##  $ tribe                             : chr  "" "" "" "" ...
##  $ subtribe                          : logi  NA NA NA NA NA NA ...
##  $ genus                             : chr  "Merluccius" "Merluccius" "Calanus" "Calanus" ...
##  $ subgenus                          : chr  "" "" "" "" ...
##  $ section                           : chr  "" "" "" "" ...
##  $ subsection                        : chr  "" "" "" "" ...
##  $ series                            : logi  NA NA NA NA NA NA ...
##  $ species                           : chr  "Merluccius bilinearis" "Merluccius bilinearis" "Calanus
##  $ subspecies                        : chr  "" "" "" "" ...
##  $ natio                             : logi  NA NA NA NA NA NA ...
##  $ variety                           : chr  "" "" "" "" ...
##  $ subvariety                        : logi  NA NA NA NA NA NA ...
##  $ forma                             : chr  "" "" "" "" ...
##  $ subforma                          : logi  NA NA NA NA NA NA ...
##  $ type                              : logi  NA NA NA NA NA NA ...
##  $ modified                          : chr  "" "" "" "" ...
##  $ language                          : logi  NA NA NA NA NA NA ...
##  $ license                           : logi  NA NA NA NA NA NA ...
##  $ rightsholder                      : logi  NA NA NA NA NA NA ...
##  $ accessrights                      : logi  NA NA NA NA NA NA ...
##  $ bibliographiccitation             : chr  "" "" "" "" ...
##  $ references                        : logi  NA NA NA NA NA NA ...
##  $ institutionid                     : logi  NA NA NA NA NA NA ...
##  $ collectionid                      : logi  NA NA NA NA NA NA ...
##  $ datasetid                         : logi  NA NA NA NA NA NA ...
##  $ institutioncode                   : chr  "RUMFS" "RUMFS" "Zoogene" "Zoogene" ...
##  $ collectioncode                    : chr  "OTTE-03-0037" "OTTE-03-0038" "" "" ...
##  $ datasetname                       : chr  "" "" "" "" ...
##  $ ownerinstitutioncode              : logi  NA NA NA NA NA NA ...
##  $ basisofrecord                     : chr  "HumanObservation" "HumanObservation" "HumanObservation
##  $ informationwithheld               : logi  NA NA NA NA NA NA ...
##  $ datageneralizations               : logi  NA NA NA NA NA NA ...
##  $ dynamicproperties                 : chr  "observedindividualcount=1;" "observedindividualcount=1
##  $ materialsampleid                  : logi  NA NA NA NA NA NA ...
##  $ occurrenceid                      : chr  "" "" "" "" ...
##  $ catalognumber                     : chr  "" "" "" "" ...
##  $ occurrenceremarks                 : chr  "location=OT 2;gear=OTTE;mesh=6;length=4;areasampled=1;
##  $ recordnumber                      : chr  "" "" "" "" ...
##  $ recordedby                        : chr  "" "" "" "" ...
##  $ recordedbyid                      : logi  NA NA NA NA NA NA ...
##  $ individualcount                   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ organismquantity                  : logi  NA NA NA NA NA NA ...
##  $ organismquantitytype              : logi  NA NA NA NA NA NA ...
##  $ sex                               : chr  "" "" "" "" ...
##   [list output truncated]
```

```
colnames(obis_malaysia)
```

```
##   [1] "id"
##   [2] "dataset_id"
##   [3] "decimallongitude"
##   [4] "decimallatitude"
```

```
##   [5] "date_start"
##   [6] "date_mid"
##   [7] "date_end"
##   [8] "date_year"
##   [9] "scientificname"
##  [10] "originalscientificname"
##  [11] "minimumdepthinmeters"
##  [12] "maximumdepthinmeters"
##  [13] "depth"
##  [14] "coordinateuncertaintyinmeters"
##  [15] "flags"
##  [16] "dropped"
##  [17] "absence"
##  [18] "shoredistance"
##  [19] "bathymetry"
##  [20] "sst"
##  [21] "sss"
##  [22] "marine"
##  [23] "brackish"
##  [24] "freshwater"
##  [25] "terrestrial"
##  [26] "taxonrank"
##  [27] "aphiaid"
##  [28] "redlist_category"
##  [29] "superdomain"
##  [30] "domain"
##  [31] "kingdom"
##  [32] "subkingdom"
##  [33] "infrakingdom"
##  [34] "phylum"
##  [35] "phylum_division"
##  [36] "subphylum_subdivision"
##  [37] "subphylum"
##  [38] "infraphylum"
##  [39] "parvphylum"
##  [40] "gigaclass"
##  [41] "megaclass"
##  [42] "superclass"
##  [43] "class"
##  [44] "subclass"
##  [45] "infraclass"
##  [46] "subterclass"
##  [47] "superorder"
##  [48] "order"
##  [49] "suborder"
##  [50] "infraorder"
##  [51] "parvorder"
##  [52] "superfamily"
##  [53] "family"
##  [54] "subfamily"
##  [55] "supertribe"
##  [56] "tribe"
##  [57] "subtribe"
##  [58] "genus"
```

```
##  [59] "subgenus"
##  [60] "section"
##  [61] "subsection"
##  [62] "series"
##  [63] "species"
##  [64] "subspecies"
##  [65] "natio"
##  [66] "variety"
##  [67] "subvariety"
##  [68] "forma"
##  [69] "subforma"
##  [70] "type"
##  [71] "modified"
##  [72] "language"
##  [73] "license"
##  [74] "rightsholder"
##  [75] "accessrights"
##  [76] "bibliographiccitation"
##  [77] "references"
##  [78] "institutionid"
##  [79] "collectionid"
##  [80] "datasetid"
##  [81] "institutioncode"
##  [82] "collectioncode"
##  [83] "datasetname"
##  [84] "ownerinstitutioncode"
##  [85] "basisofrecord"
##  [86] "informationwithheld"
##  [87] "datageneralizations"
##  [88] "dynamicproperties"
##  [89] "materialsampleid"
##  [90] "occurrenceid"
##  [91] "catalognumber"
##  [92] "occurrenceremarks"
##  [93] "recordnumber"
##  [94] "recordedby"
##  [95] "recordedbyid"
##  [96] "individualcount"
##  [97] "organismquantity"
##  [98] "organismquantitytype"
##  [99] "sex"
## [100] "lifestage"
## [101] "reproductivecondition"
## [102] "behavior"
## [103] "establishmentmeans"
## [104] "occurrencestatus"
## [105] "preparations"
## [106] "disposition"
## [107] "othercatalognumbers"
## [108] "associatedmedia"
## [109] "associatedreferences"
## [110] "associatedsequences"
## [111] "associatedtaxa"
## [112] "organismid"
```

```
## [113] "organismname"
## [114] "organismscope"
## [115] "associatedoccurrences"
## [116] "associatedorganisms"
## [117] "previousidentifications"
## [118] "organismremarks"
## [119] "eventid"
## [120] "parenteventid"
## [121] "samplingprotocol"
## [122] "samplesizevalue"
## [123] "samplesizeunit"
## [124] "samplingeffort"
## [125] "eventdate"
## [126] "eventtime"
## [127] "startdayofyear"
## [128] "enddayofyear"
## [129] "year"
## [130] "month"
## [131] "day"
## [132] "verbatimeventdate"
## [133] "habitat"
## [134] "fieldnumber"
## [135] "fieldnotes"
## [136] "eventremarks"
## [137] "locationid"
## [138] "highergeographyid"
## [139] "highergeography"
## [140] "continent"
## [141] "waterbody"
## [142] "islandgroup"
## [143] "island"
## [144] "country"
## [145] "countrycode"
## [146] "stateprovince"
## [147] "county"
## [148] "municipality"
## [149] "locality"
## [150] "verbatimlocality"
## [151] "verbatimelevation"
## [152] "minimumelevationinmeters"
## [153] "maximumelevationinmeters"
## [154] "verbatimdepth"
## [155] "minimumdistanceabovesurfaceinmeters"
## [156] "maximumdistanceabovesurfaceinmeters"
## [157] "locationaccordingto"
## [158] "locationremarks"
## [159] "verbatimcoordinates"
## [160] "verbatimlatitude"
## [161] "verbatimlongitude"
## [162] "verbatimcoordinatesystem"
## [163] "verbatimsrs"
## [164] "geodeticdatum"
## [165] "coordinateprecision"
## [166] "pointradiusspatialfit"
```

```
## [167] "footprintwkt"
## [168] "footprintsrs"
## [169] "footprintspatialfit"
## [170] "georeferencedby"
## [171] "georeferenceddate"
## [172] "georeferenceprotocol"
## [173] "georeferencesources"
## [174] "georeferenceverificationstatus"
## [175] "georeferenceremarks"
## [176] "geologicalcontextid"
## [177] "earliesteonorlowesteonothem"
## [178] "latesteonorhighesteonothem"
## [179] "earliesteraorlowesterathem"
## [180] "latesteraorhighesterathem"
## [181] "earliestperiodorlowestsystem"
## [182] "latestperiodorhighestsystem"
## [183] "earliestepochorlowestseries"
## [184] "latestepochorhighestseries"
## [185] "earliestageorloweststage"
## [186] "latestageorhighststage"
## [187] "lowestbiostratigraphiczone"
## [188] "highestbiostratigraphiczone"
## [189] "lithostratigraphicterms"
## [190] "group"
## [191] "formation"
## [192] "member"
## [193] "bed"
## [194] "identificationid"
## [195] "identifiedby"
## [196] "identifiedbyid"
## [197] "dateidentified"
## [198] "identificationreferences"
## [199] "identificationremarks"
## [200] "identificationqualifier"
## [201] "identificationverificationstatus"
## [202] "typestatus"
## [203] "taxonid"
## [204] "scientificnameid"
## [205] "acceptednameusageid"
## [206] "parentnameusageid"
## [207] "originalnameusageid"
## [208] "nameaccordingtoid"
## [209] "namepublishedinid"
## [210] "taxonconceptid"
## [211] "acceptednameusage"
## [212] "parentnameusage"
## [213] "originalnameusage"
## [214] "nameaccordingto"
## [215] "namepublishedin"
## [216] "namepublishedinyear"
## [217] "higherclassification"
## [218] "specificepithet"
## [219] "infraspecificepithet"
## [220] "verbatimtaxonrank"
```

```
## [221] "scientificnameauthorship"
## [222] "vernacularname"
## [223] "nomenclaturalcode"
## [224] "taxonomicstatus"
## [225] "nomenclaturalstatus"
## [226] "taxonremarks"
```

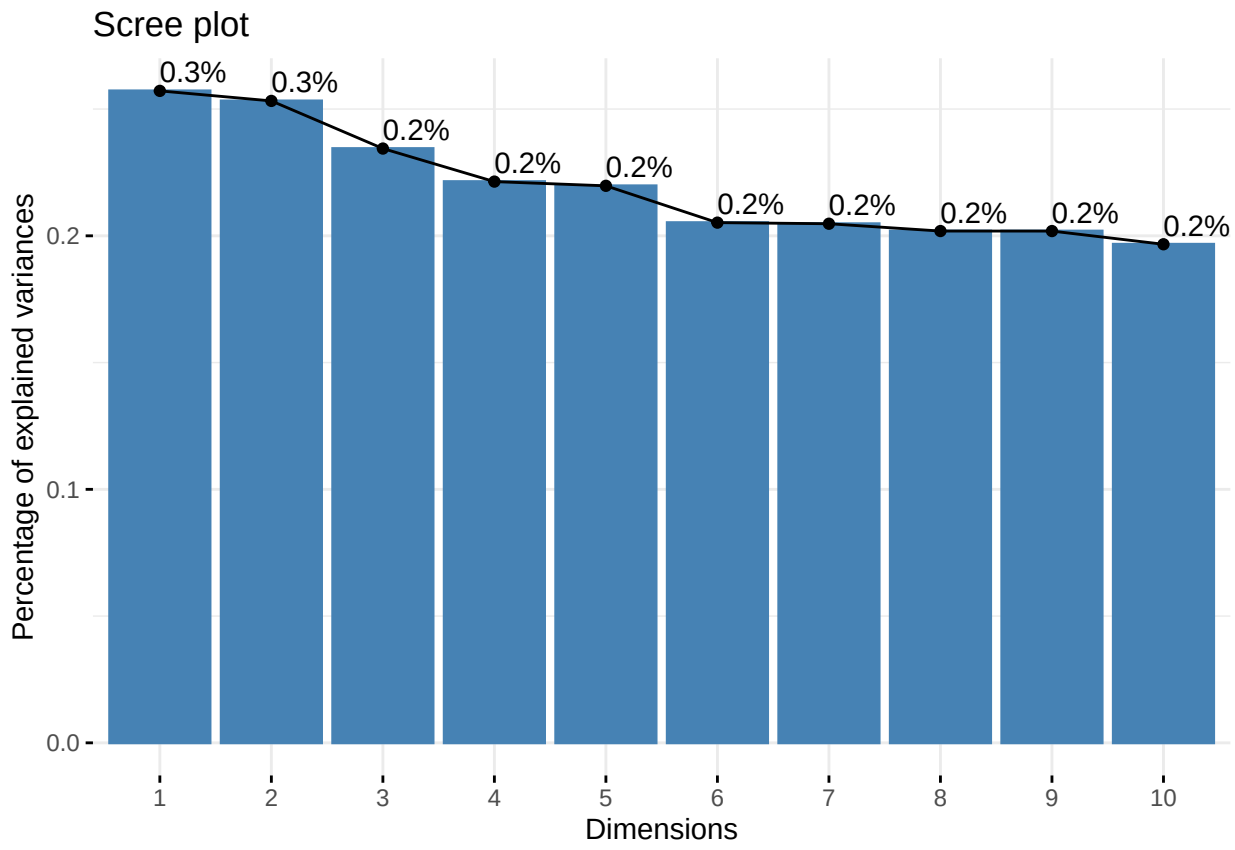
Question 1:

Data set has to be sampled to lower the computation time for MCA.

```
df <- obis_malaysia %>% sample_n(2000) %>% mutate(date_year_char = as.character(date_year)) %>% select(s
```

```
mca_results <- MCA(df, graph = F)
```

```
fviz_eig(mca_results, addlabels = TRUE)
```

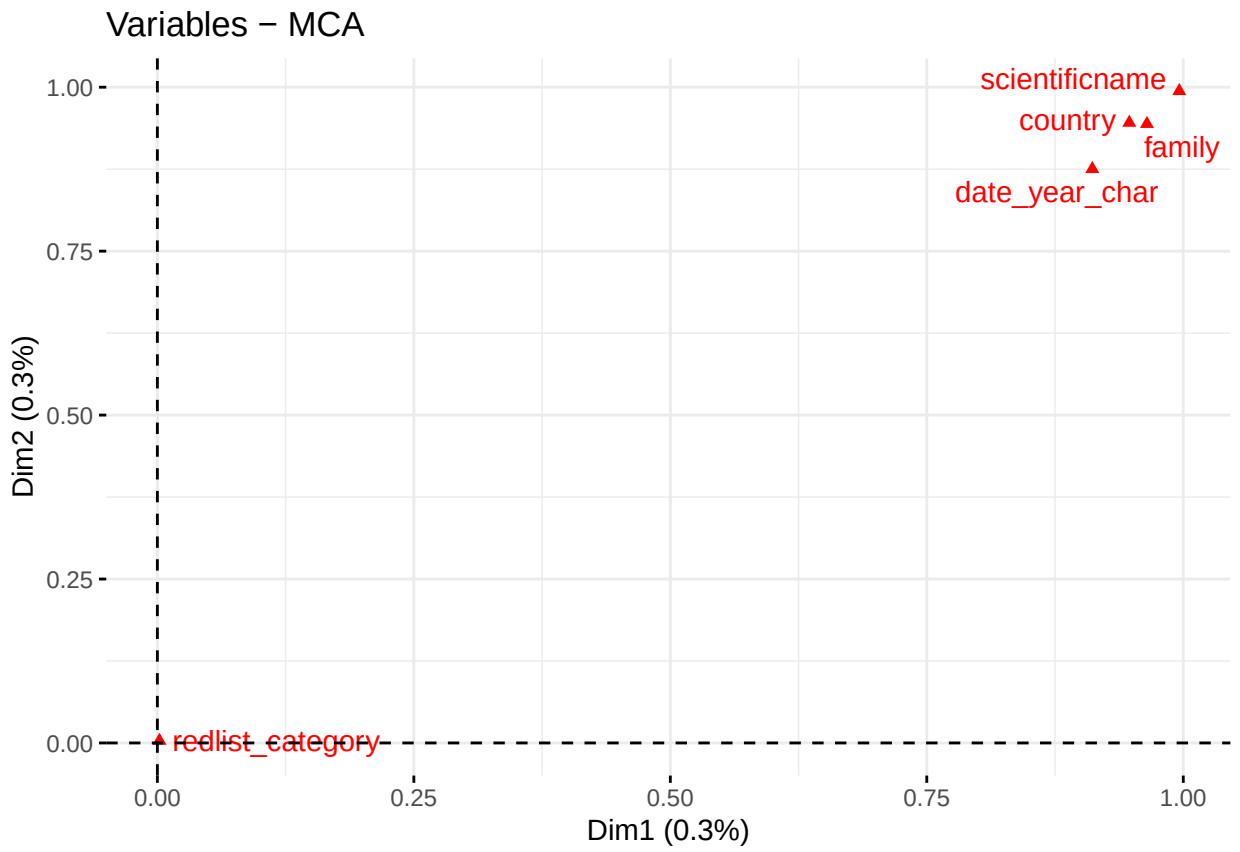


For MCA, the variables considered in this analysis is scientificname, country, family, date_year_char, and redlist_category.

The percentage of total variance explained by the first two principal components is 0.3% + 0.3% = 0.6%.

Question 2:

```
fviz_mca_var(mca_results, choice = "mca.cor",
             repel = TRUE, # Avoid text overlapping (slow)
             ggtheme = theme_minimal())
```



The most important variables out of the five are scientificname, country, family, and date_year_char, with all of them being postively correlated with both DIm1 and Dim2.

Only redlist_category is not correlated with both Dim1 and Dim2.