R-programında Synbreed Paketini Kullanarak Genomik Seleksiyon Uygulaması

Prof. Dr. Kadir KIZILKAYA

Aydın Adnan Menderes Üniversitesi

Ziraat Fakültesi Zootekni Bölümü

Biyometri ve Genetik Anabilim Dalı

kkizilkaya@adu.edu.tr

http://synbreed.r-forge.r-project.org/

R-Forge

Welcome to synbreed project!

Collection of statistical and genetic methods developed for the synbreed project for synergistic plant and animal breeding.

Description:

- The R-package `synbreed` provides a framework for the analysis of genomic prediction data (Genomic Selection, GWAS, QTL-mapping) within an open source software.
- Features
  - Data processing
    - Combining raw data sources to a gpData object
    - Conversion from and to class cross in package qtl
    - Coding marker data into number of copies of the minor allele
    - Preselection of markers according to MAF, % missing values and LD
    - Imputation of missing genotypes by marginal allele distribution, family structure for fully homozygous inbred individuals or flanking markers using Beagle
  - Data visualization and analysis
    - Summary method for classes gpData, pedigree and relationshipMatrix
    - Marker map representation for low and high density maps
    - LD computation as r2 and LD decay visualization as as scatterplot or stacked histogram
    - Pedigree tree and kinship visualization of relatedness between individuals
  - Statistical models
    - Estimation of pedigree based relationship (additive and dominance)
    - Marker based relationship
    - Cross-validation for BLUP, Ridge Regression and Bayesian methods
- The package comes with ABSOLUTELY NO WARRANTY; for details see http://www.gnu.org/copyleft/gpl.html (GPL).

Download:

To install the latest development version of package `synbreed` from R-Forge (if you are running a recent R version), use

```
install.packages("synbreed",repos="http://r-forge.r-project.org")
```

You can also manually download the source code from
https://r-forge.r-project.org/R/?group_id=710

A stable release is available from CRAN

http://synbreed.r-forge.r-project.org/

Documentation:

- Publication in Bioinformatics
- Citation information:
  `citation(package="synbreed")`
- pdf version of the manual
- package vignette with detailed background information and examples:
  `library(synbreed)`
  `vignette("IntroSyn")`
- overview over the functions
- Poster presented at the 4th International Conference on Quantitative Genetics, June 2012, Edinburgh

Workshop: Introduction to the synbreed R package, November 2012, TUM:

- Course slides (2012 - 11 - 08)
- Code of the slides (2012 - 11 - 08)

Developers:

- Valentin Wimmer , Chair of Plant Breeding, Technische Universität München
- Theresa Albrecht , Chair of Plant Breeding, Technische Universität München
- Hans-Juergen Auinger , Chair of Plant Breeding, Technische Universität München

Financial support:

The development of the package was financially supported by the German Federal Ministry of Education and Research (BMBF) within the AgroClustEr "Synbreed Synergistic plant and animal breeding" (FKZ 0315528A)

Synbreed Paketinden Önce Yüklenmesi Gereken Paketler

- lattice
  - install.packages("lattice")
- igraph
  - install.packages("igraph")
- MASS
  - install.packages("MASS")
- LDheatmap
  - install.packages("LDheatmap")
- qtl
  - install.packages("qtl")
- doBy
  - install.packages("doBy")
- BLR
  - install.packages("BLR")
- Regress
  - install.packages("regress")

install.packages("synbreed",repos="http://r-forge.r-project.org")

Synbreed Paketi – Veri Setleri

- Synbreed paketinin aktifleştirilmesi
  - library("synbreed")

- Synbreed paketinde mevcut veri setleri

```
# Maize (Mısır) veri seti
  data(maize)
  str(maize)

# Mice (Fare) veri seti
  data(mice)
  str(mice)

# Cattle (Sigir) veri seti
  data(cattle)
  str(cattle)
```

Synbreed – Cattle Veri Seti

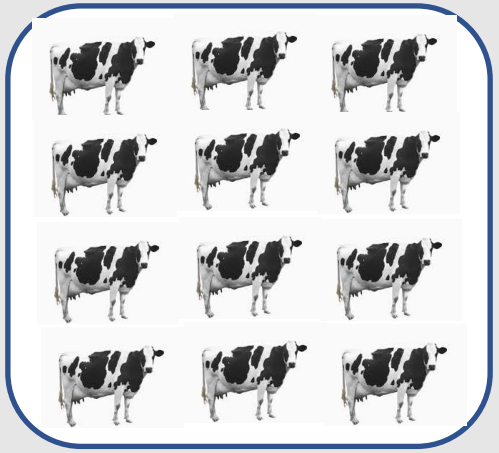
```
# Cattle (Sigir) veri seti
  data(cattle)
  str(cattle)
List of 7
 $ covar      :'data.frame':  1929 obs. of  3 variables:
  ..$ id         : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004"
...
  ..$ phenotyped: logi [1:1929] FALSE FALSE FALSE FALSE FALSE FALSE ...
  ..$ genotyped : logi [1:1929] FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ pheno      : num [1:500, 1:2, 1] -23.4 15.5 -19.2 -10.4 -14.1 ...
  ..- attr(*, "dimnames")=List of 3
  .. ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  .. ..$ : chr [1:2] "Phenotype1" "Phenotype2"
  .. ..$ : chr "1"
```

Synbreed – Cattle Veri Seti

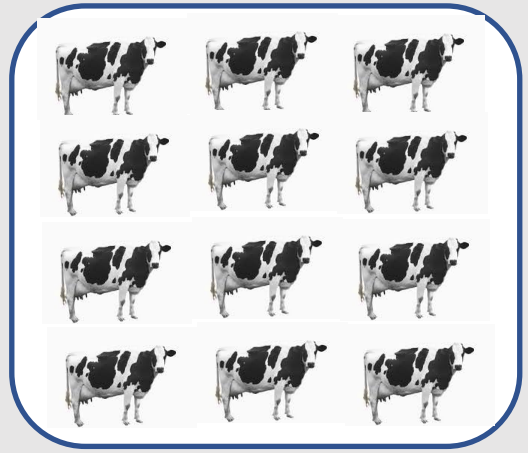
```
$ geno       : chr [1:500, 1:7250] "AB" "BB" "BB" "BB" ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  .. ..$ : chr [1:7250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 $ map        :Classes 'GenMap' and 'data.frame':   7250 obs. of  2 variables:
  ..$ chr: int [1:7250] 1 1 1 1 1 1 1 1 1 1 ...
  ..$ pos: num [1:7250] 0.000001 0.20258 0.279809 0.668932 0.763913 ...
 $ pedigree   :Classes 'pedigree' and 'data.frame': 1929 obs. of  4 variables:
  ..$ ID   : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004" ...
  ..$ Par1 : chr [1:1929] "0" "0" "0" "0" ...
  ..$ Par2 : chr [1:1929] "0" "0" "0" "0" ...
  ..$ gener: num [1:1929] 0 0 0 0 0 0 0 0 0 0 ...
 $ phenoCovars: NULL
 $ info       :List of 2
  ..$ map.unit: chr "Mb"
  ..$ codeGeno: logi FALSE
 - attr(*, "class")= chr "gpData"
```

Synbreed – Genomik Seleksiyon

Uygulama Popülasyonu (Veriseti)
Training – Referance Population



Seleksiyon Uygulanacak Adaylar (Veriseti)
Validation Population



Veriseti

- **Özellikler**:
  - Fenotipik değerler (Süt veya et verimi, canlı ağırlık)
- **Kovaryet**:
  - Sabit veya şansa bağlı faktör (Sürü, Yıl, Mevsim, Laktasyon sırası vb)
- **Pedigri**:
  - Soy kütüğü (baba, ana ve doğum yılı kayıtları)
- **Genotip**:
  - SNP markır verileri (700 K, 50 K veya 3 K)
- **SNP markır harita kayıtları**

Model: GBLUP, BayesA, B, C veya LASSO

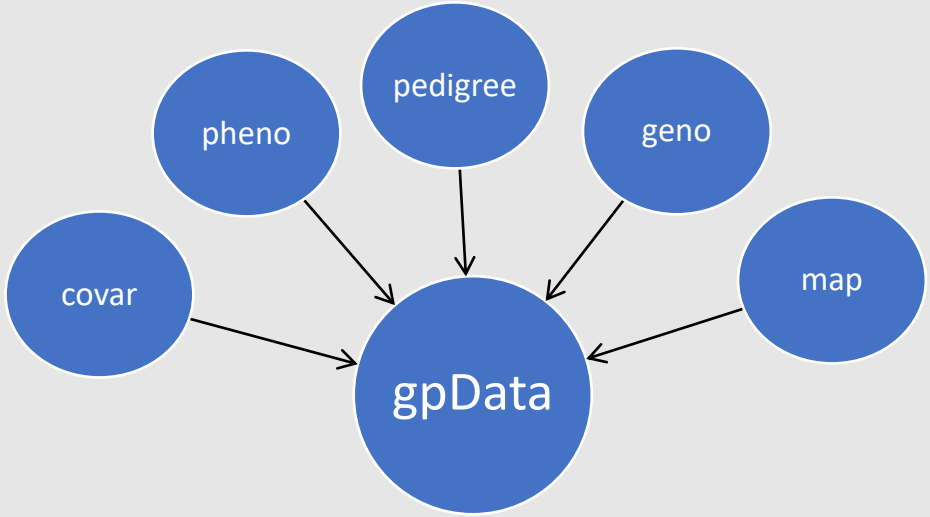
SNP markır etkileri tahmin edilir

Veriseti

- **Genotip**:
  - SNP markır verileri (700 K, 50 K veya 3 K)

Adayların genomik damızlık değer tahmininde ve buna bağlı olarak seçiminde kullanılır

Synbreed -gpData Veri Yapısının Oluşturulması



gpData:

covar, phone, pedigree, geno, map nesnelerinin birey ID'lerine ve SNP markır ID'lerine göre birleştirilmesiyle oluşturulur

```
create.gpData(pheno=NULL, geno=NULL, map=NULL, pedigree=NULL, family=NULL,
              covar=NULL, reorderMap=TRUE, map.unit="cM", repeated=NULL,
              modCovar=NULL, na.string="NA", cores=1)
```

Synbreed - Veri Yapısı

- covar:

Veri analizinde yer alacak kesikli-sürekli kovaryet değişkenlere ait değerleri içerir

- phone:

Analiz edilecek değişkenlere (özelliklere) ait değerleri içerir

- pedigree:

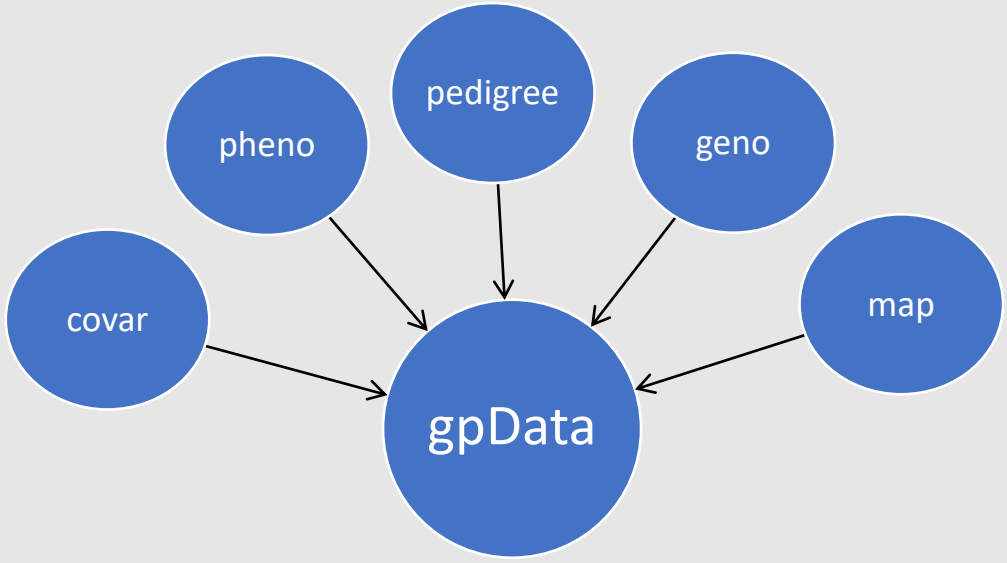
Analizde yer alan bireylerin pedigri (soy kütüğü) kayıtlarını içerir

- geno:

Analizde yer alacak SNP markır değerlerini içerir

- map:

Analizde yer alacak SNP markırları haritalamak için gerekli değerleri içerir



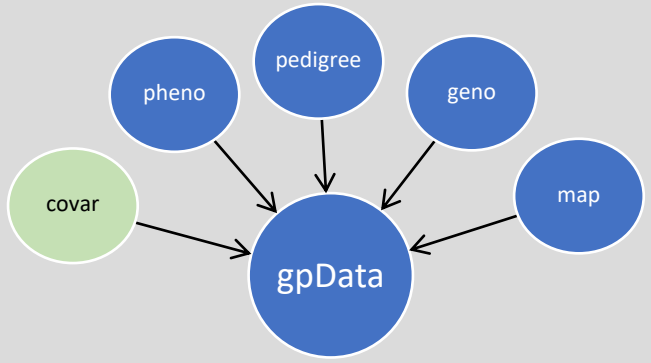
Sığır (cattle) Veri Seti
covar – data.frame

Kesikli-Sürekli Kovaryet Değişkenler

Analizde yer alacak kovaryet değişkenler mevcut ise data.frame veri yapısında R'a yüklenir.

Kovaryet değişkenler yüklenirken hayvan numaraları rownames olarak aktarılmalıdır.

Sığır veri setinde kovaryet değişkenler yoktur

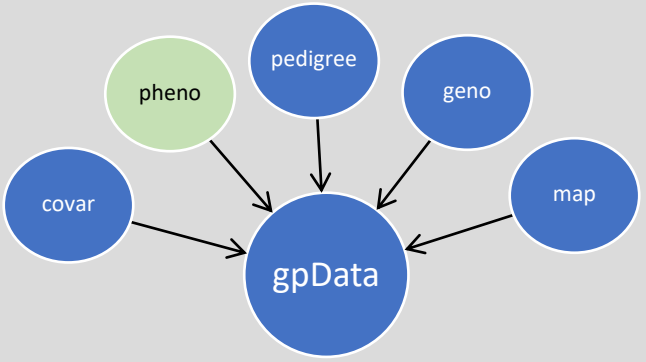


Sığır Veri Seti
pheno – data.frame

Fenotipik Değerler – Özelliklere Ait Gözlem Değerleri

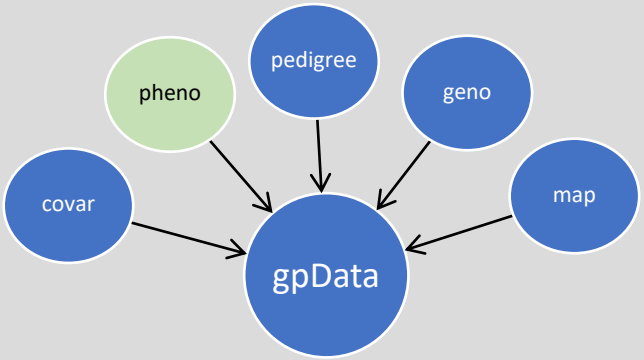
Dosya adı: cattle_pheno.txt
 Sütun-1: Hayvanların numarası, ID
 Sütun-2: Özellik 1 için gözlem değerleri, Trait1
 Sütun-3: Özellik 2 için gözlem değerleri, Trait2

```
ID Trait1 Trait2
ID11430 -23.43 263.9
ID11431 15.48 182.22
ID11432 -19.18 6.64
ID11433 -10.43 42.55
ID11434 -14.07 326.41
```



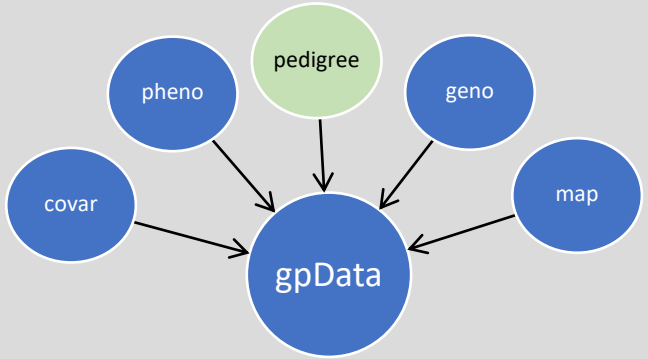
Sığır Veri Seti
pheno – data.frame

```
# Fenotipik degerler – Ozelliklere ait degerler

pheno = read.table(file = "cattle_pheno.txt", header = TRUE, row.names = 1)

str(pheno)
'data.frame':     500 obs. of  2 variables:
 $ Trait1: num  -23.4 15.5 -19.2 -10.4 -14.1 ...
 $ Trait2: num  263.9 182.22 6.64 42.55 326.41 ...

head(pheno)
        Trait1 Trait2
ID11430 -23.43 263.90
ID11431  15.48 182.22
ID11432 -19.18   6.64
ID11433 -10.43  42.55
ID11434 -14.07 326.41
ID11435  -6.62 -34.01
```



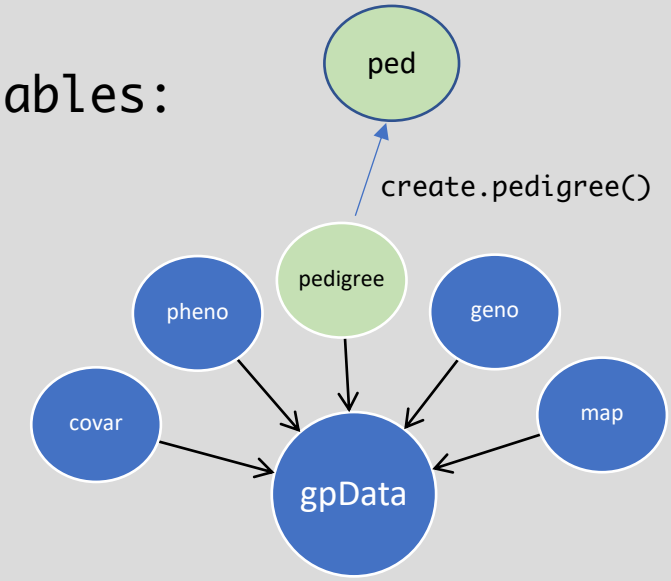
Sığır Veri Seti
pedigree – data.frame

```
# Pedigri degerleri

pedigree = read.table(file = "cattle_pedigree.txt", header = TRUE)

str(pedigree)
'data.frame':     1929 obs. of  4 variables:
 $ ID   : chr  "ID10001" "ID10002" "ID10003" "ID10004" ...
 $ Par1 : chr  "0" "0" "0" "0" ...
 $ Par2 : chr  "0" "0" "0" "0" ...
 $ gener: int  0 0 0 0 0 0 0 0 0 0 ...

head(pedigree)
        ID Par1 Par2 gener
1 ID10001    0    0     0
2 ID10002    0    0     0
3 ID10003    0    0     0
4 ID10004    0    0     0
```



Sığır Veri Seti
pedigree – data.frame - create.pedigree

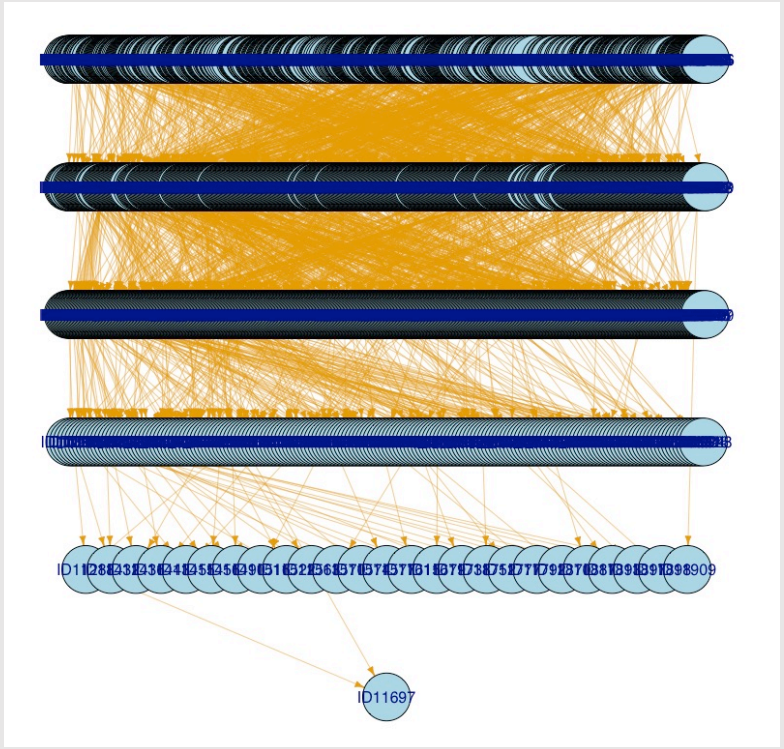
```
create.pedigree(ID, Par1, Par2, gener=NULL,sex=NULL,add.ancestors=FALSE,unknown=0)


# create.pedigree() fonksiyonu ile pedigri formatını olustur

  ped = create.pedigree(pedigree$ID, pedigree$Par1, pedigree$Par2, pedigree$gener)

  str(ped)

Classes ‘pedigree’ and 'data.frame':    1929 obs. of  4 variables:
 $ ID   : chr  "ID10001" "ID10002" "ID10003" "ID10004" ...
 $ Par1 : chr  "0" "0" "0" "0" ...
 $ Par2 : chr  "0" "0" "0" "0" ...
 $ gener: int  0 0 0 0 0 0 0 0 0 0 ...
```

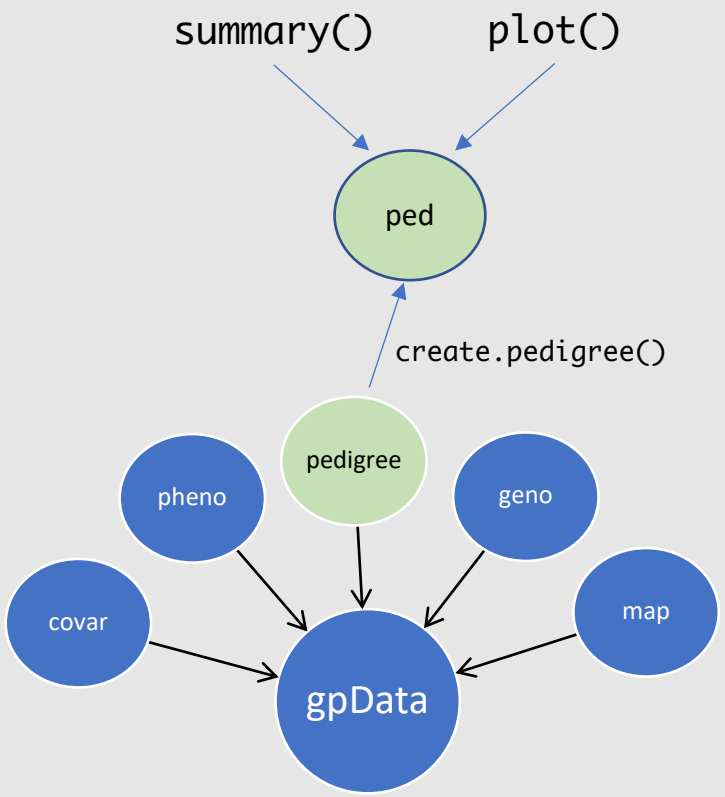


Sığır Veri Seti
pedigree – data.frame - create.pedigree

```
summary(ped)
$nID
[1] 1929
$nPar1
[1] 376
$nPar2
[1] 1053
$nGener
[1] 6
$nUnknownParents
[1] 1512
attr(,"class")
[1] "summary.pedigree"
```

```
plot(ped)
```

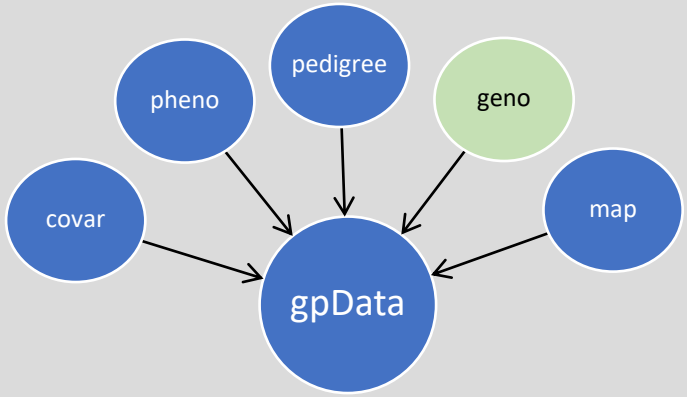




Sığır Veri Seti
geno – matrix

SNP markir degerleri

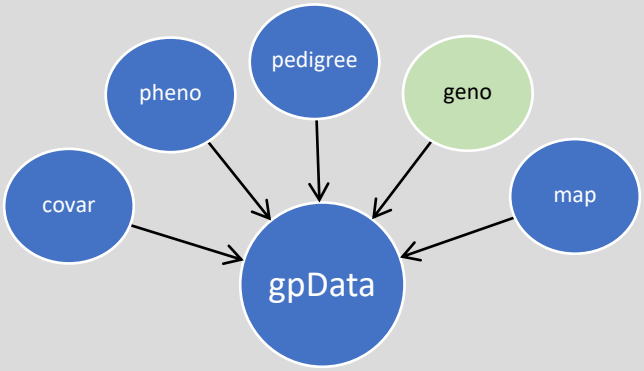
Dosya adı: cattle_geno012.txt
 Sütun-1: Hayvanların numarası, ID
 Sütun-2: SNP 1 için gözlem değerleri, SNP_1
 Sütun-3: SNP 2 için gözlem değerleri, SNP_2
    :
 Sütun-7250: SNP 7250 için gözlem değerleri, SNP_7250



ID	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5
ID11430	1	2	0	2	0
ID11431	2	1	0	0	0
ID11432	2	2	0	2	0
ID11433	2	1	0	1	0
ID11434	2	2	0	2	1

Sığır Veri Seti
geno – matrix

```
# SNP markir degerleri

geno = read.table(file = "cattle_geno012.txt", header = TRUE, row.names = 1)
geno = as.matrix(geno)

str(geno)

 int [1:500, 1:7250] 1 2 2 2 2 2 2 2 2 2 ...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  ..$ : chr [1:7250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...

geno[1:5,1:10]

        SNP_1 SNP_2 SNP_3 SNP_4 SNP_5 SNP_6 SNP_7 SNP_8 SNP_9 SNP_10
ID11430     1     2     0     2     0     2     2     0     2      1
ID11431     2     1     0     0     0     2     1     1     2      0
ID11432     2     2     0     2     0     2     2     0     2      2
ID11433     2     1     0     1     0     2     2     0     2      1
ID11434     2     2     0     2     1     2     2     2     2      0
```

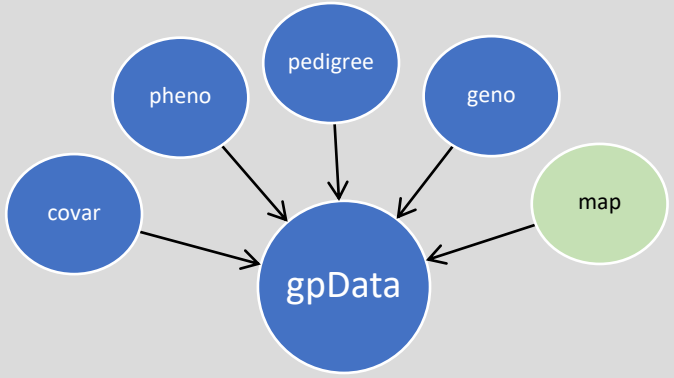


Sığır Veri Seti
map – data.frame

SNP markır haritası

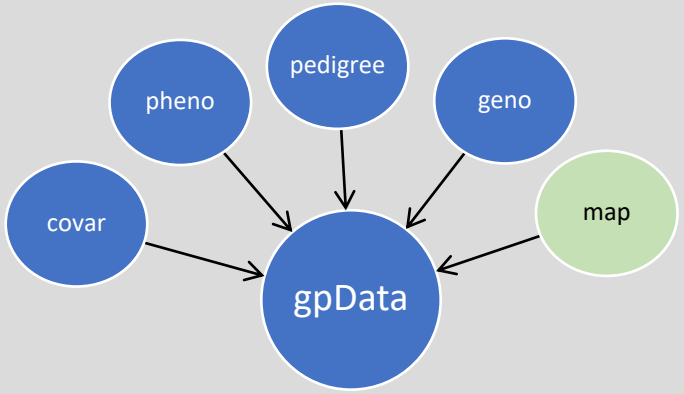
Dosya adı: cattle_map.txt
 Sütun-1: SNP markır numarası, SNP_ID
 Sütun-2: Kromozom numarası, chr
 Sütun-3: SNP markırın kromozom üzerindeki yeri, pos

SNP_ID	chr	pos
SNP_1	1	1e-06
SNP_2	1	0.20258
SNP_3	1	0.279809
SNP_4	1	0.668932
SNP_5	1	0.763913



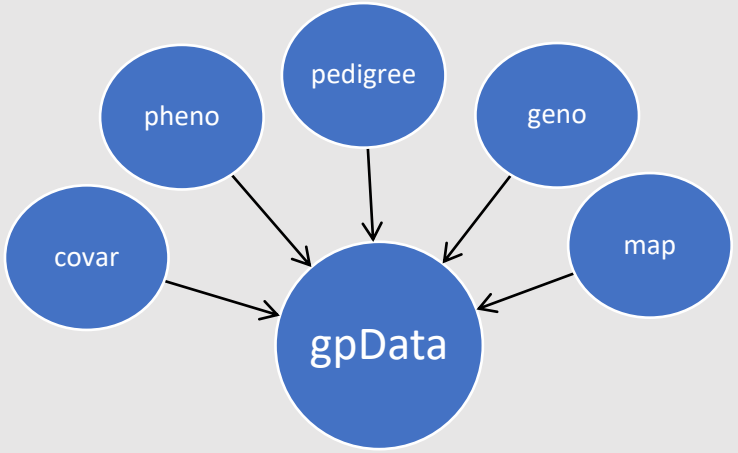
Sığır Veri Seti
map – data.frame

```
# Map degerleri

map = read.table(file = "cattle_map.txt", header = TRUE, row.names = 1)

str(map)
'data.frame':     7250 obs. of  2 variables:
 $ chr: int  1 1 1 1 1 1 1 1 1 1 ...
 $ pos: num  0.000001 0.20258 0.279809 0.668932 0.763913 ...

head(map)
       chr       pos
SNP_1    1 0.000001
SNP_2    1 0.202580
SNP_3    1 0.279809
SNP_4    1 0.668932
SNP_5    1 0.763913
SNP_6    1 0.828652
```



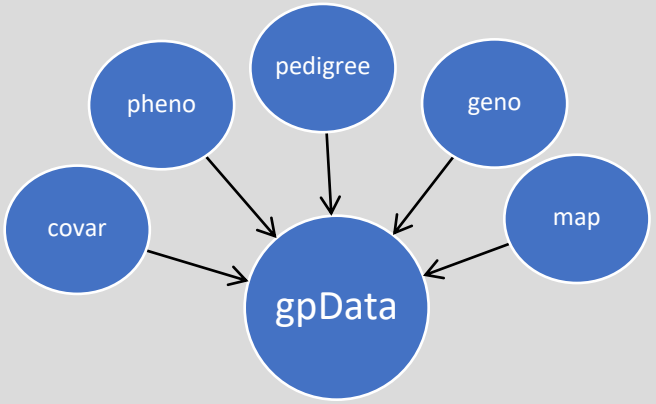
Sığır Veri Seti - gpData - list

```
# gpData olustur
  gpData <- create.gpData(pheno=pheno, geno=geno,
                          map=map, pedigree=ped)
```



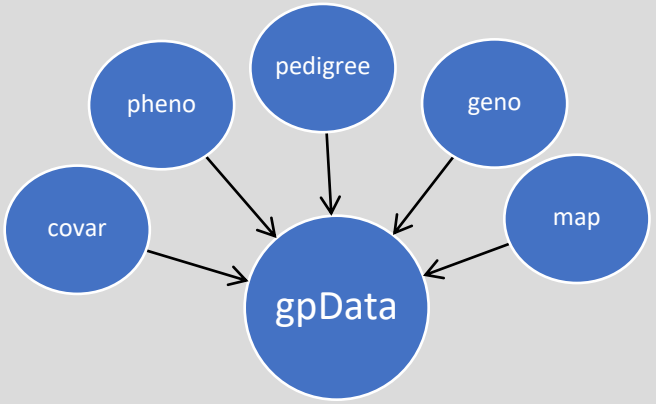
Sığır Veri Seti - gpData - list

```
# gpData olustur
  gpData <- create.gpData(pheno = pheno, geno = geno, map = map,  pedigree = ped)
  str(gpData)
List of 7
 $ covar      :'data.frame':  1929 obs. of  4 variables:
  ..$ id         : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004" ...
  ..$ phenotyped: logi [1:1929] FALSE FALSE FALSE FALSE FALSE FALSE ...
  ..$ genotyped : logi [1:1929] FALSE FALSE FALSE FALSE FALSE FALSE ...
  ..$ family     : logi [1:1929] NA NA NA NA NA NA ...
 $ pheno      : num [1:500, 1:2, 1] -23.4 15.5 -19.2 -10.4 -14.1 ...
  ..- attr(*, "dimnames")=List of 3
  .. ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  .. ..$ : chr [1:2] "Trait1" "Trait2"
  .. ..$ : chr "1"
 $ geno       : int [1:500, 1:7250] 1 2 2 2 2 2 2 2 2 2 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  .. ..$ : chr [1:7250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 $ map        :Classes 'GenMap' and 'data.frame': 7250 obs. of  2 variables:
  ..$ chr: int [1:7250] 1 1 1 1 1 1 1 1 1 1 ...
  ..$ pos: num [1:7250] 0.000001 0.20258 0.279809 0.668932 0.763913 ...
 $ pedigree   :Classes 'pedigree' and 'data.frame':        1929 obs. of  4 variables:
  ..$ ID   : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004" ...
  ..$ Par1 : chr [1:1929] "0" "0" "0" "0" ...
  ..$ Par2 : chr [1:1929] "0" "0" "0" "0" ...
  ..$ gener: int [1:1929] 0 0 0 0 0 0 0 0 0 0 ...
 $ phenoCovars: NULL
 $ info       :List of 4
  ..$ map.unit: chr "cM"
  ..$ codeGeno: logi FALSE
  ..$ version : chr "gpData object was created by synbreed version 0.12-14"
  ..$ Call    : language create.gpData(pheno = pheno, geno = geno, map = map, pedigree = ped)
 - attr(*, "class")= chr "gpData"
```



Sığır Veri Seti - gpData - list

```
# gpData olustur
  gpData <- create.gpData(pheno = pheno, geno = geno, map = map,  pedigree = ped)
  str(gpData)
List of 7
 $ covar      :'data.frame':  1929 obs. of  4 variables:
  ..$ id         : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004" ...
  ..$ phenotyped: logi [1:1929] FALSE FALSE FALSE FALSE FALSE FALSE ...
  ..$ genotyped : logi [1:1929] FALSE FALSE FALSE FALSE FALSE FALSE ...
  ..$ family     : logi [1:1929] NA NA NA NA NA NA ...
 $ pheno      : num [1:500, 1:2, 1] -23.4 15.5 -19.2 -10.4 -14.1 ...
  ..- attr(*, "dimnames")=List of 3
  .. ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  .. ..$ : chr [1:2] "Trait1" "Trait2"
  .. ..$ : chr "1"
 $ geno       : int [1:500, 1:7250] 1 2 2 2 2 2 2 2 2 2 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  .. ..$ : chr [1:7250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 $ map        :Classes 'GenMap' and 'data.frame': 7250 obs. of  2 variables:
  ..$ chr: int [1:7250] 1 1 1 1 1 1 1 1 1 1 ...
  ..$ pos: num [1:7250] 0.000001 0.20258 0.279809 0.668932 0.763913 ...
 $ pedigree   :Classes 'pedigree' and 'data.frame':        1929 obs. of  4 variables:
  ..$ ID   : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004" ...
  ..$ Par1 : chr [1:1929] "0" "0" "0" "0" ...
  ..$ Par2 : chr [1:1929] "0" "0" "0" "0" ...
  ..$ gener: int [1:1929] 0 0 0 0 0 0 0 0 0 0 ...
 $ phenoCovars: NULL
 $ info       :List of 4
  ..$ map.unit: chr "cM"
  ..$ codeGeno: logi FALSE
  ..$ version : chr "gpData object was created by synbreed version 0.12-14"
  ..$ Call    : language create.gpData(pheno = pheno, geno = geno, map = map, pedigree = ped)
 - attr(*, "class")= chr "gpData"
```



Sığır Veri Seti - gpData - list

```
# Map birimi olarak "Mb" tanımla
gpData$info$map.unit <- "Mb"
str(gpData)
```

```
$ info       :List of 4
  ..$ map.unit: chr "Mb"
  ..$ codeGeno: logi FALSE
  ..$ version : chr "gpData object was created by synbreed version 0.12-14"
  ..$ Call    : language create.gpData(pheno = pheno, geno = geno, map = map, pedigree = ped)
 - attr(*, "class")= chr "gpData"
```

Sığır Veri Seti - gpData - list summary(gpData)

```
# gpData icin Ozet bilgiler
  summary(gpData)

$covar
$covar$n
[1] 1929

$covar$nphenotyped
[1] 500

$covar$ngenotyped
[1] 500
```

```
$pheno
     Trait1              Trait2
 Min.   :-50.07000   Min.   :-424.43
 1st Qu.:-10.25750   1st Qu.: -94.40
 Median :  0.13500   Median :  -0.84
 Mean   : -0.00104   Mean   :   0.05
 3rd Qu.: 10.35250   3rd Qu.:  96.99
 Max.   : 49.41000   Max.   : 422.63
```

Sığır Veri Seti - gpData – list summary(gpData)

```
$geno
$geno$nMarkers
[1] 7250

$geno$genotypes
geno
        0         1         2
0.2889283 0.3483592 0.3627126

$geno$nNA
[1] 0

$geno$markerChr
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21
250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250
 22  23  24  25  26  27  28  29
250 250 250 250 250 250 250 250

$geno$mappedMarkers
[1] 7250
```

Sığır Veri Seti - gpData – list summary(gpData)

```
$pedigree

$nID
[1] 1929

$nPar1
[1] 376

$nPar2
[1] 1053

$nGener
[1] 6

$nUnknownParents
[1] 1512
```

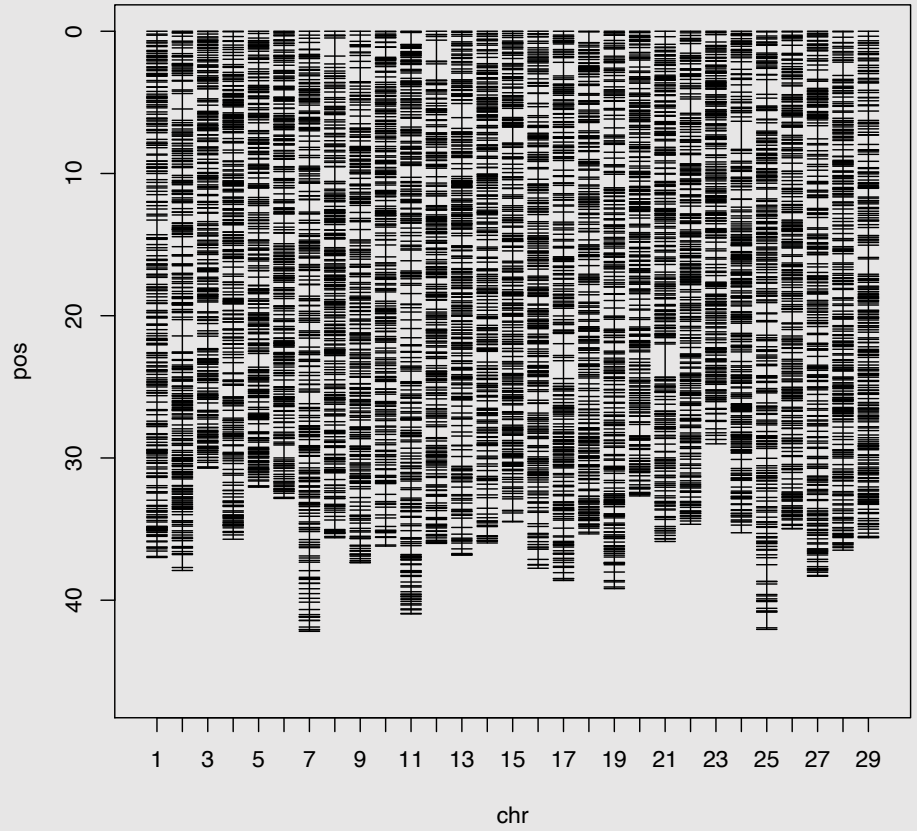
```
$nUnknownParents
[1] 1512

attr(,"class")
[1] "summary.pedigree"

attr(,"class")
[1] "summary.gpData"
```

SNP Markır Haritası

```
plotGenMap(map,
           dense=FALSE,
           nMarker=TRUE,
           bw=1,
           centr=NULL,
           file=NULL,
           fileFormat="pdf",...)
```

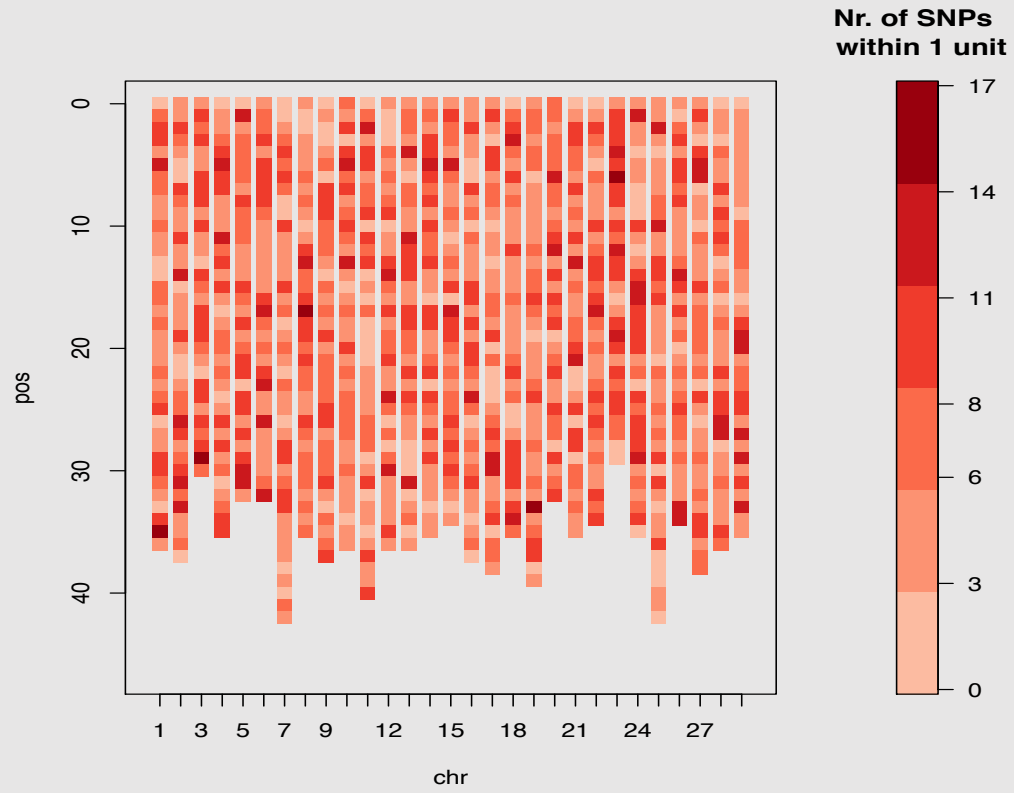
SNP Markır Haritası

```
plotGenMap(gpData$map)  # plot ekranda oluşturulur

plotGenMap(gpData$map, file="CattleMap1") # plot dosyaya kaydedilir
```



SNP Markır Haritası

```
# plot ekranda oluşturulur
plotGenMap(gpData$map, dense=TRUE, nMarker=FALSE)

# plot dosyaya kaydedilir
plotGenMap(gpData$map, dense=TRUE, nMarker=FALSE, file="CattleMap2")
```



SNP Markır – Kodlama - Eksik Markır Tahmini (Imputation) - Markır Ayıklama

```
codeGeno(gpData,
         impute=FALSE,
         impute.type=c("random","family","beagle","beagleAfterFamily","beagleNoRand",
                       "beagleAfterFamilyNoRand","fix"),
         replace.value=NULL,
         maf=NULL,
         nmiss=NULL,
         label.heter="alleleCoding",
         reference.allele="minor",
         keep.list=NULL,
         keep.identical=TRUE,
         verbose=FALSE,
         minFam=5,
         showBeagleOutput=FALSE,
         tester=NULL,
         print.report=FALSE,
         check=FALSE,
         ploidy=2,
         cores=1)
```

SNP Markır – Kodlama - Eksik Markır Tahmini (Imputation) - Markır Ayıklama (Sığır SNP markır)

```
# SNP kodlama, impute ve kalite kontrol

  gpCode <- codeGeno(gpData, label.heter="1", impute=TRUE, impute.type="random",
                     verbose=TRUE)

   step 1  : No markers removed due to fraction of missing values
   step 2  : Recoding alleles
   step 4  : No markers discarded due to minor allele frequency
   step 7  : Imputing of missing values
   step 7d : Random imputing of missing values
   step 8  : No recoding of alleles necessary after imputation
   step 9  : No markers discarded due to minor allele frequency
   step 10 : No duplicated markers removed
   End     : 7250 marker(s) remain after the check

   Summary of imputation
   total number of missing values                  : 0
   number of random imputations                    : 0
```

Bağlantı Dengesizliği - Linkage Disequilibrium (LD)

$$r^2 = \frac{D_{AB}^2}{p_A\, p_B\, p_a\, p_b}$$

$$D_{AB} = p_{AB} - p_A\, p_B$$

p_{AB}: AB halotipin frekansı
$p_A = 1 - p_a$: bir lokustaki A alelin frekansı
$p_B = 1 - p_b$: diğer lokustaki B alelin frekansı

```
pairwiseLD(gpCode, chr=NULL, type=c("data.frame","matrix"), use.plink=FALSE,
           ld.threshold=0, ld.window=99999, rm.unmapped=TRUE, cores=1)
```

Hill WG, Robertson A (1968). Linkage Disequilibrium in Finite Populations. Theoretical and Applied Genetics, 6(38), 226 - 231.

Bağlantı Dengesizliği - Linkage Disequilibrium (LD)

```
# Kromozom-1 için LD hesaplama
chr1 <- pairwiseLD(gpCode, chr=1, type="data.frame")
```

```
> str(chr1)
List of 1
 $ chr_1:'data.frame': 31125 obs. of  5 variables:
  ..$ marker1: chr [1:31125] "SNP_1" "SNP_1" "SNP_1" "SNP_1" ...
  ..$ marker2: chr [1:31125] "SNP_2" "SNP_3" "SNP_4" "SNP_5" ...
  ..$ r      : num [1:31125] -0.08286 0.19335 0.00712 -0.12762 -0.42093 ...
  ..$ r2     : num [1:31125] 6.87e-03 3.74e-02 5.07e-05 1.63e-02 1.77e-01 ...
  ..$ dist   : num [1:31125] 0.203 0.28 0.669 0.764 0.829 ...
 - attr(*, "class")= chr "LDdf"
```

```
> chr1
$chr_1
    marker1 marker2           r           r2      dist
1     SNP_1   SNP_2 -0.0828573035 6.865333e-03  0.202579
2     SNP_1   SNP_3  0.1933493688 3.738398e-02  0.279808
3     SNP_1   SNP_4  0.0071181940 5.066869e-05  0.668931
4     SNP_1   SNP_5 -0.1276181342 1.628639e-02  0.763912
5     SNP_1   SNP_6 -0.4209317562 1.771835e-01  0.828651
```

Bağlantı Dengesizliği - Linkage Disequilibrium (LD)

```
# Kromozom-1 için LD hesaplama
chr1mat <- pairwiseLD(gpCode, chr=1, type="matrix")
```

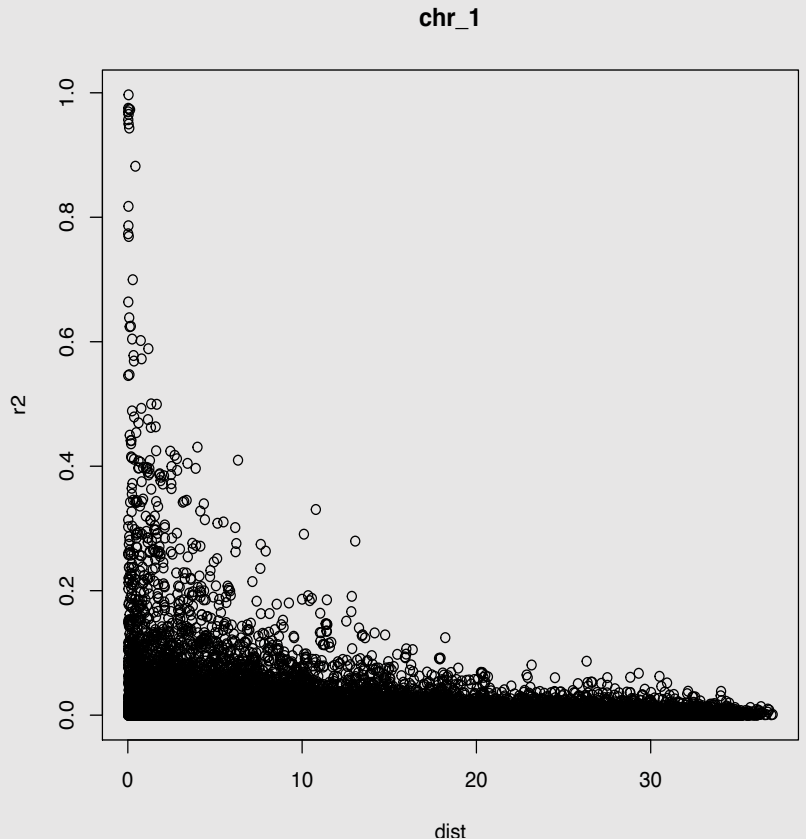
```
> str(chr1mat)
List of 3
 $ LD      :List of 1
  ..$ chr_1: num [1:250, 1:250] 1.00 6.87e-03 3.74e-02 5.07e-05 1.63e-02 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
  .. .. ..$ : chr [1:250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 $ distance:List of 1
  ..$ chr_1: num [1:250, 1:250] 0 0.203 0.28 0.669 0.764 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
  .. .. ..$ : chr [1:250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 $ LDcor   :List of 1
  ..$ : num [1:250, 1:250] 1 -0.08286 0.19335 0.00712 -0.12762 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
  .. .. ..$ : chr [1:250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 - attr(*, "class")= chr "LDmat"
```

Bağlantı Dengesizliği - Linkage Disequilibrium (LD)

```
# Kromozom-1 için LD değerlerinin SNP markırlar arası mesafeye göre dağılımı
# plot.LDdf

plot(chr1, gpCode, plotType="dist", chr=1, file="Chr1_LD", fileFormat="pdf")
```

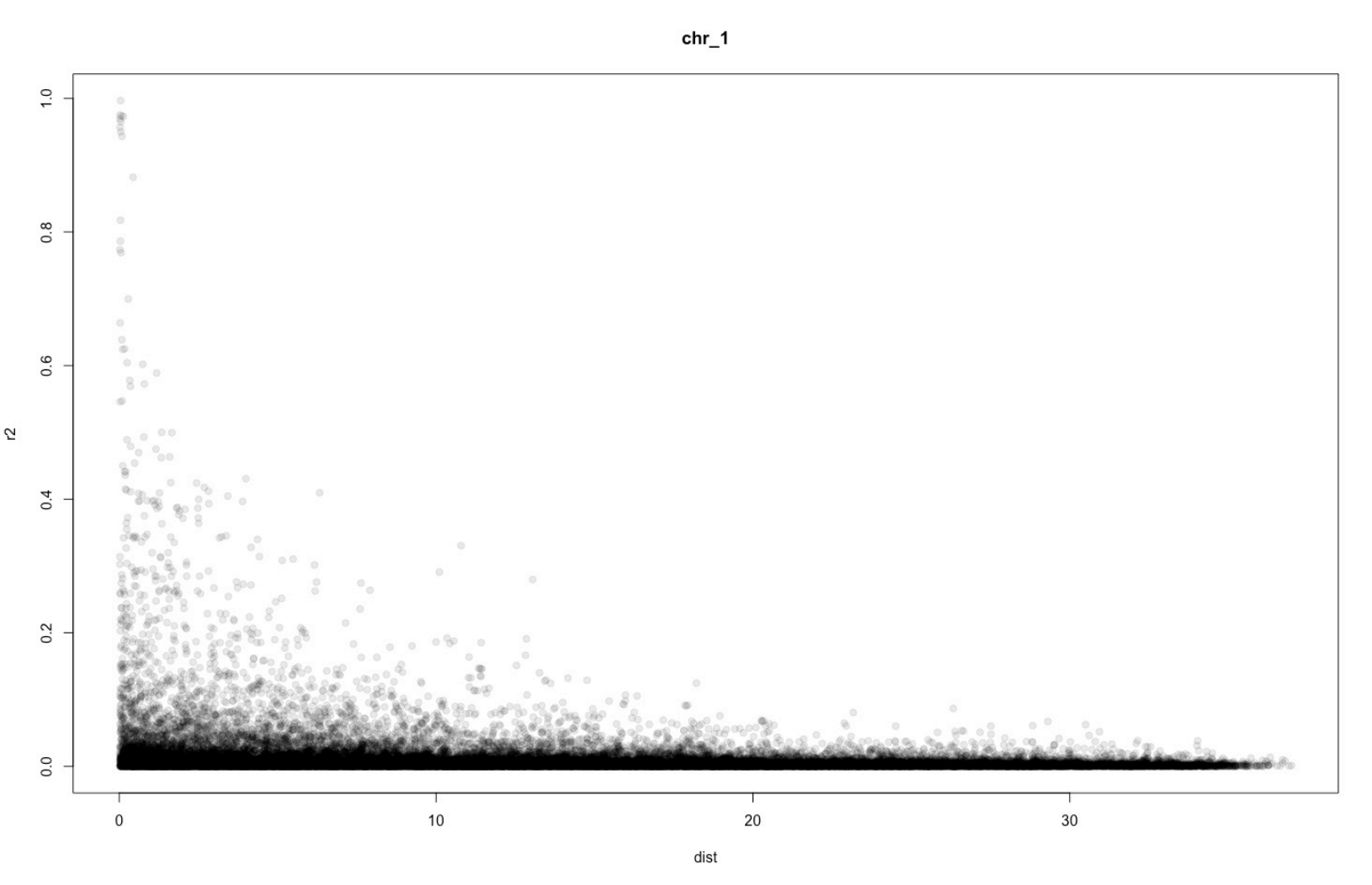
```
> chr1
$chr_1
    marker1 marker2             r            r2      dist
1     SNP_1   SNP_2 -0.0828573035 6.865333e-03  0.202579
2     SNP_1   SNP_3  0.1933493688 3.738398e-02  0.279808
3     SNP_1   SNP_4  0.0071181940 5.066869e-05  0.668931
4     SNP_1   SNP_5 -0.1276181342 1.628639e-02  0.763912
5     SNP_1   SNP_6 -0.4209317562 1.771835e-01  0.828651
```

```
>  summary(chr1)

      noM     avgr2 minr2     maxr2        Pr02  averDist  maxDist
chr_1 250 0.0124188     0 0.9967143 0.007453815 0.0124188 36.99836
```



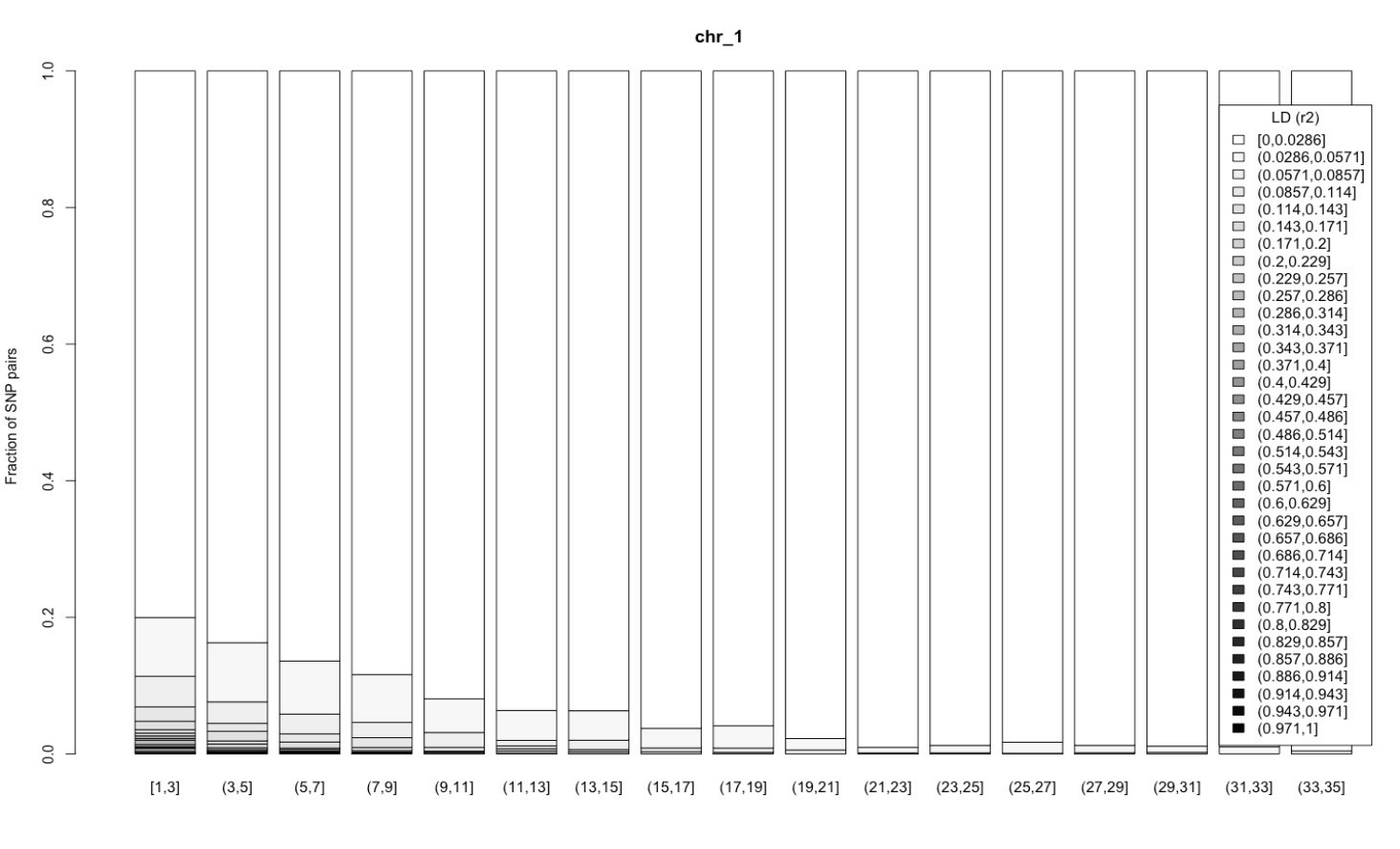
Bağlantı Dengesizliği - Linkage Disequilibrium (LD)

```
LDDist(chr1,type="p",pch=19,colD=hsv(alpha=0.1,v=0))
```



Bağlantı Dengesizliği - Linkage Disequilibrium (LD)

```
LDDist(chr1,
       type="bars",
       breaks=list(dist=seq(1,36,by=2),r2=seq(1,0,length.out=36)))
```



Akrabalık İlişkisi – Pedigri - SNP Markır

```
kin(gpCode,
    ret=c("add",
          "kin",
          "dom",
          "gam",
          "realized",
          "realizedAB"
          "sm",
          "sm-smin",
          "gaussian"),
     DH=NULL,
     maf=NULL,
     selfing=NULL,
     lambda=1,
     P=NULL,
     cores=1)
```

Pedigriye göre akrabalık

SNP markırlara göre akrabalık

Akrabalık İlişkisi – Pedigri - SNP Markır

Pedigriye göre akrabalığın hesaplanmasında Pedigri Kayıtlarının bilinmesini gerektirir

ret=c("add"): Ekleme genlerden ileri gelen akrabalık (A matrisi)

ret=c("kin"): Ekleme genlerden ileri gelen akrabalığın yarısı

ret=c("dom"): Dominans ilişkiye göre akrabalık

ret=c("gam"): Gametik ilişkiye göre akrabalık

Akrabalık İlişkisi – Pedigri (A Matrisi)

ret=c("add"): Ekleme genlerden ileri gelen akrabalık (***A*** matrisi)

```
A <- kin(gpCode, ret="add")

str(A)

 'relationshipMatrix' num [1:1929, 1:1929] 1 0 0 0 0 0 0 0 0 0
...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004" ...
  ..$ : chr [1:1929] "ID10001" "ID10002" "ID10003" "ID10004" ...
 - attr(*, "info")= chr "This relationshipMatrix was calculated
by synbreed version 0.12-14"
 - attr(*, "type")= chr "add"
```

Akrabalık İlişkisi – Pedigri (A Matrisi)

```
summary(A)

$dim
nrow ncol
1929 1929

$rank
[1] 1929

$range.off.diagonal
  min   max
0.000 0.625

$mean.diag
[1] 1.000194

$mean.off.diag
[1] 0.002926408

$diag.val
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.000   1.000   1.000   1.000   1.125

$empty
[1] 0

attr(,"class")
[1] "summary.relationshipMatrix"
```

Akrabalık İlişkisi – SNP Markır (*G*)

SNP markırlara göre akrabalık

```
ret=c("realized",
      "realizedAB",
      "sm",
      "sm-smin",
      "gaussian")
```

ret=c("realized"), Habier et al. (2007) veya vanRaden (2008)'e göre genomik akrabalık matrisini ($\boldsymbol{G}$) oluşturu.

$$\boldsymbol{G} = \frac{\boldsymbol{Z}\boldsymbol{Z}'}{2\sum p_i(1-p_i)} = \frac{(\boldsymbol{W}-\boldsymbol{P})(\boldsymbol{W}-\boldsymbol{P})'}{2\sum p_i(1-p_i)}$$

$\boldsymbol{W}$: SNP markır değerlerini içeren matris,
$\boldsymbol{P}$: 2 ile çarpılmış allel frekans değerlerini içeren matris
p_i: SNP markır i nin allel frekansı

Akrabalık İlişkisi – SNP Markır (*G*)

ret=c("realized) SNP markırlara göre akrabalık

```
# SNP markırlara göre Akrabalık (G) Matrisini olustur

  G <- kin(gpCode, ret="realized")

  str(G)

 'relationshipMatrix' num [1:500, 1:500] 24.345 4.546 0.857 0.388 -1.234 ...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
  ..$ : chr [1:500] "ID11430" "ID11431" "ID11432" "ID11433" ...
 - attr(*, "alleleFrequencies")= Named num [1:7250] 0.132 0.376 -0.986 0.934 -0.584 0.88
0.866 -0.602 0.9 0.008 ...
  ..- attr(*, "names")= chr [1:7250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 - attr(*, "expectedMAX")= num 168
 - attr(*, "SNPs")= chr [1:7250] "SNP_1" "SNP_2" "SNP_3" "SNP_4" ...
 - attr(*, "info")= chr "This relationshipMatrix was calculated by synbreed version 0.12-14"
 - attr(*, "type")= chr "realized"
```

Akrabalık İlişkisi – SNP Markır (*G*)

```
ret=c("realized) SNP markırlara göre akrabalık
```

```
summary(G)
$dim
nrow ncol
 500  500

$rank
[1] 499

$range.off.diagonal
      min       max
-1.980661  7.483978

$mean.diag
[1] 20.29519
```

```
$mean.off.diag
[1] -0.04067172

$diag.val
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  17.41   19.91   20.23   20.30   20.60   24.35

$empty
[1] 0

attr(,"class")
[1] "summary.relationshipMatrix"
```

Tahmin Modeli

Pedigri-BLUP (PBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{e}$$

$\boldsymbol{y}$: gözlem değerleri vektörü

$\boldsymbol{\beta}$: sabit etkiler vektörü

$\boldsymbol{a}$: şans balığı etkiler (damızlık değer) vektörü $\boldsymbol{a} \sim N(\boldsymbol{0}, \boldsymbol{A}\sigma_a^2)$

$\boldsymbol{A}$: eklemeli akrabalık matrisi

σ_a^2: eklemeli genetik varyans

$\boldsymbol{e}$: şansa bağlı hata vektörü $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_e^2)$

$\boldsymbol{I}$: Birim matrisi

σ_e^2: hata varyansı

$\boldsymbol{X}$: sabit etkilere ait desen matrisi

$\boldsymbol{Z}$: şansa bağlı etkilere ait desen matrisi

Genomik-BLUP (GBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

$\boldsymbol{y}$: gözlem değerleri vektörü

$\boldsymbol{\beta}$: sabit etkiler vektörü

$\boldsymbol{u}$: şans balığı etkiler (genomik damızlık değer) vektörü $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{U}\sigma_u^2)$

$\boldsymbol{U}$: genomik akrabalık matrisi

σ_u^2: genomik varyans

$\boldsymbol{e}$: şansa bağlı hata vektörü $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_e^2)$

$\boldsymbol{I}$: Birim matrisi

σ_e^2: hata varyansı

$\boldsymbol{X}$: sabit etkilere ait desen matrisi

$\boldsymbol{Z}$: şansa bağlı etkilere ait desen matrisi

Tahmin Modeli

Pedigri-BLUP (PBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{e}$$

$$E\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

$$Var\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}\sigma_a^2 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\sigma_e^2\end{bmatrix}$$

$$E[\boldsymbol{y}] = \boldsymbol{X\beta}$$

$$Var[\boldsymbol{y}] = \boldsymbol{V} = \boldsymbol{ZGZ'} + \boldsymbol{R}$$

Karışık Model Eşitlikleri

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{a}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z}\\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z} + \boldsymbol{G^{-1}}\end{bmatrix}^{-}\begin{bmatrix}\boldsymbol{X'R^{-1}y}\\ \boldsymbol{Z'R^{-1}y}\end{bmatrix}$$

Genomik-BLUP (GBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

$$E\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

$$Var\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{bmatrix} = \begin{bmatrix}\boldsymbol{U}\sigma_a^2 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\sigma_e^2\end{bmatrix}$$

$$E[\boldsymbol{y}] = \boldsymbol{X\beta}$$

$$Var[\boldsymbol{y}] = \boldsymbol{V} = \boldsymbol{ZGZ'} + \boldsymbol{R}$$

Karışık Model Eşitlikleri

$$\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{u}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z}\\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z} + \boldsymbol{G^{-1}}\end{bmatrix}^{-}\begin{bmatrix}\boldsymbol{X'R^{-1}y}\\ \boldsymbol{Z'R^{-1}y}\end{bmatrix}$$

Tahmin Modeli (Pedigri-BLUP - PBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{e}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}} \end{bmatrix}^{-} \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix}$$

```
PBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=A, trait="Trait1")
```

```
summary(PBLUP_Trait1)
             Length  Class              Mode
fit              24 regress            list
model             1 -none-             character
y               500 -none-             numeric  ⇒ Gözlem değerleri vektörü
g               500 -none-             numeric  ⇒ Damızlık değer tahminleri (â)
prediction        0 -none-             NULL
markerEffects     0 -none-             NULL
kin         3721041 relationshipMatrix numeric
```

Tahmin Modeli (Pedigri-BLUP - PBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{e}$$

$$\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}}\end{bmatrix}^{-} \begin{bmatrix}\boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'(I\sigma_e^2)^{-1}X} & \boldsymbol{X'(I\sigma_e^2)^{-1}Z} \\ \boldsymbol{Z'(I\sigma_e^2)^{-1}X} & \boldsymbol{Z'(I\sigma_e^2)^{-1}Z + (A\sigma_a^2)^{-1}}\end{bmatrix}^{-} \begin{bmatrix}\boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y}\end{bmatrix}$$

```
PBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=A, trait="Trait1")
```

```
summary(PBLUP_Trait1$fit)
Likelihood kernel: K = (Intercept)
Maximized log likelihood with kernel K is  -1613.348
Linear Coefficients:
             Estimate Std. Error
 (Intercept)    0.526      1.009
Variance Coefficients:
             Estimate Std. Error
      kinTS   99.552     40.605
      In     142.800     37.399
```

$\hat{\sigma}_a^2 = 99.552$
$\hat{\sigma}_e^2 = 142.800$

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} = \frac{99.552}{99.552 + 142.800} = 0.41$$

Tahmin Modeli (Pedigri-BLUP - PBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{e}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}} \end{bmatrix}^{-} \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix}$$

```
PBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=A, trait="Trait1")
summary(PBLUP_Trait1)
              Length  Class              Mode
fit               24 regress           list
model              1 -none-            character
y                500 -none-            numeric  ⇒ Gözlem değerleri vektörü
g                500 -none-            numeric  ⇒ Damızlıkdeğere tahminleri (â)
```

Korelasyon $r(y, \hat{a}) = 0.9560781$

```
>  cor(PBLUP_Trait1$y, PBLUP_Trait1$g)
[1] 0.9560781
```

Tahmin Modeli (Pedigri-BLUP - PBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{e}$$

```
PBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=A, trait="Trait1")
```

```
>  data.frame(PBLUP_Trait1$y, PBLUP_Trait1$g)
```

	PBLUP_Trait1.y	PBLUP_Trait1.g
ID11430	-23.43	-11.774060986
ID11431	15.48	0.264133336
ID11432	-19.18	-6.861402741
ID11433	-10.43	-4.943506739
ID11434	-14.07	-3.295750040

Korelasyon $r(y, \hat{a}) = 0.9560781$

```
>  cor(PBLUP_Trait1$y, PBLUP_Trait1$g)
[1] 0.9560781
```

Tahmin Modeli (Genomik-BLUP - GBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}} \end{bmatrix}^{-} \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix}$$

```
GBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=G, trait="Trait1")
```

```
summary(GBLUP_Trait1)
             Length  Class      Mode
fit              24 regress    list
model             1 -none-     character
y               500 -none-     numeric  ⇒ Gözlem değerleri vektörü
g               500 -none-     numeric  ⇒ Genomik damızlık değere tahminleri (û)
prediction        0 -none-     NULL
markerEffects     0 -none-     NULL
kin          250000 relationshipMatrix numeric
```

Tahmin Modeli (Genomik-BLUP - GBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

$$\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}}\end{bmatrix}^{-} \begin{bmatrix}\boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'(I\sigma_e^2)^{-1}X} & \boldsymbol{X'(I\sigma_e^2)^{-1}Z} \\ \boldsymbol{Z'(I\sigma_e^2)^{-1}X} & \boldsymbol{Z'(I\sigma_e^2)^{-1}Z + (U\sigma_a^2)^{-1}}\end{bmatrix}^{-} \begin{bmatrix}\boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y}\end{bmatrix}$$

```
GBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=G, trait="Trait1")
```

```
summary(GBLUP_Trait1$fit)
 Likelihood kernel: K = (Intercept)
Maximized log likelihood with kernel K is  -1615.298
Linear Coefficients:
             Estimate Std. Error
 (Intercept)   -0.001      0.585
 Variance Coefficients:
             Estimate Std. Error
       kinTS    3.503      1.474
       In     171.030     28.834
```

$\hat{\sigma}_a^2 = 3.503$

$\hat{\sigma}_e^2 = 171.030$

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \frac{3.503}{3.503 + 171.030} = 0.02$$

Tahmin Modeli (Genomik-BLUP - GBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}} \end{bmatrix}^{-} \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix}$$

```
GBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=G, trait="Trait1")
summary(GBLUP_Trait1)
             Length  Class      Mode
fit              24 regress    list
model             1 -none-     character
y               500 -none-     numeric  ⇒ Gözlem değerleri vektörü
g               500 -none-     numeric  ⇒ Genomik damızlık değere tahminleri (û)
```

Korelasyon $r(y, \hat{u}) = 0.9127231$

```
>  cor(GBLUP_Trait1$y, GBLUP_Trait1$g)
[1] 0.9127231
```

Tahmin Modeli (Genomik-BLUP - GBLUP)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

```
GBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=G, trait="Trait1")
```

```
>  data.frame(GBLUP_Trait1$y, GBLUP_Trait1$g)
```

	GBLUP_Trait1.y	GBLUP_Trait1.g
ID11430	-23.43	-8.52562829
ID11431	15.48	0.21847379
ID11432	-19.18	-5.04440346
ID11433	-10.43	-2.01580621
ID11434	-14.07	0.53287980

Korelasyon $r(y, \hat{u}) = 0.9127231$

```
>  cor(GBLUP_Trait1$y, GBLUP_Trait1$g)
[1] 0.9127231
```

Çapraz-Doğrulama (Genomik-BLUP - GBLUP)

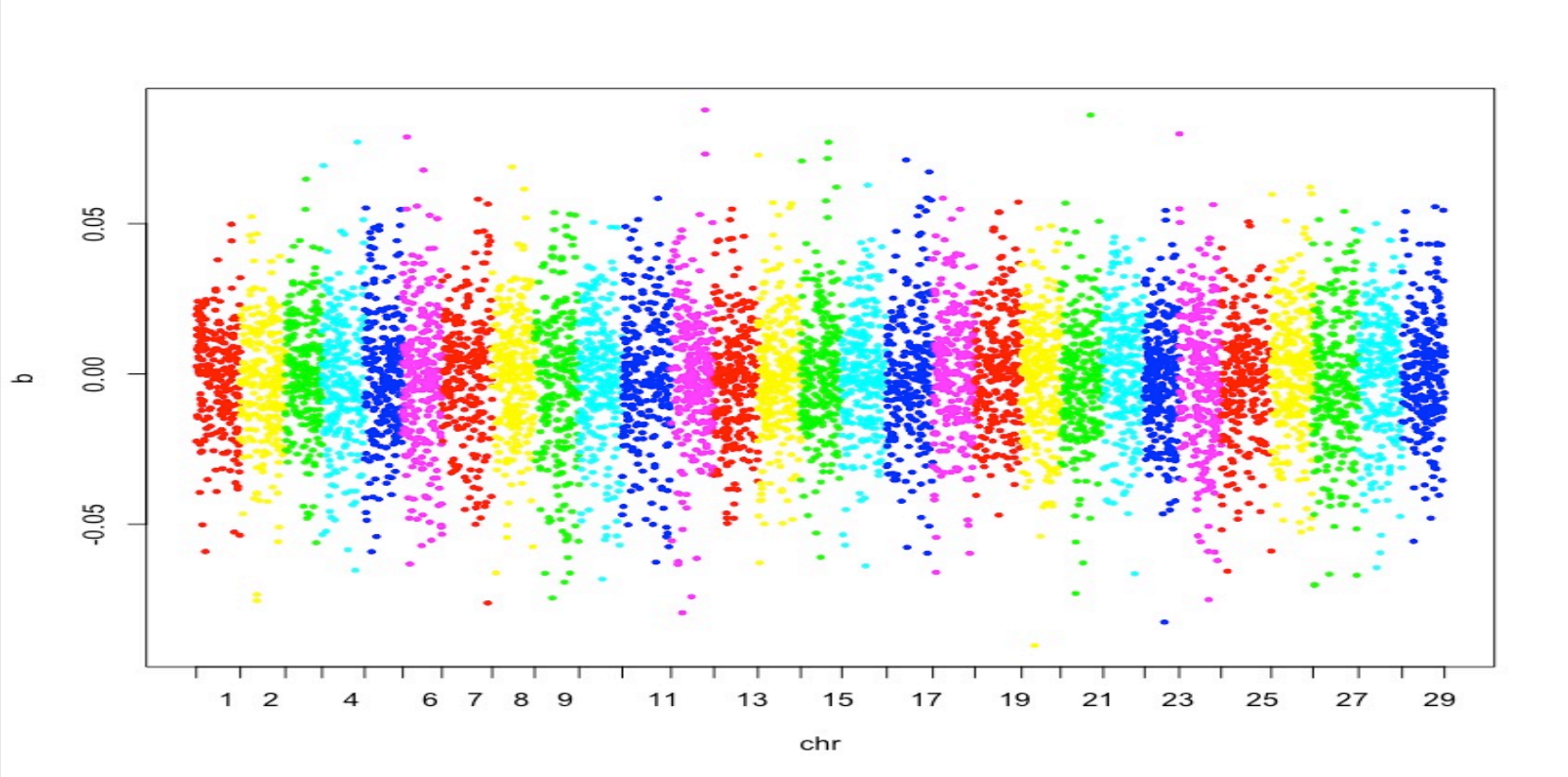
```
# Son 50 birey Çapraz Doğrulama için belirlendi
 son50 <- rownames(gpCode$pheno)[451:500]

# Son 50 birey veri setinden çıkartıldı
# İlk 450 birey Uygulama (Training) veri seti olarak oluşturuldu
 gpCode2 <- discard.individuals(gpCode, son50)

# GBLUP ile Uygulama veri setinin analizi
 GBLUP450 <- gpMod(gpCode2, model="BLUP", kin=G)

# Son 50 bireyin genomik damızlık degerlerinin tahmini
 g50 <- predict(GBLUP450, son50)

 cor(data.frame(g50, gpCode$pheno[451:500]))
                             g50 gpCode.pheno.451.500.
g50                    1.0000000             0.2081486
gpCode.pheno.451.500. 0.2081486             1.0000000
```

Çapraz-Doğrulama (Genomik-BLUP - GBLUP) – crossVal()

```
# Çapraz doğrulama Replication:1 ve Validasyon:2
  cv.GBLUP_Trait1 <- crossVal(gpCode, cov.matrix=list(G), k=2, Rep=1,
                              Seed=123, sampling="random",
                              varComp=GBLUP_Trait1$fit$sigma,
                              VC.est="commit")

Validayon gruplarına ait korelasyonlar
  cv.GBLUP_Trait1$PredAbi
         rep1
fold1 -0.0564
fold2 -0.0658

Validayon gruplarına ait hata kareler ortalaması
  cv.GBLUP_Trait1$mse
          rep1
fold1 213.5414
fold2 273.1165
```

Genomik-BLUP – RRBLUP

```
RRBLUP_Trait1 <- gpMod(gpCode, model="BLUP", kin=G, trait="Trait1",
                       markerEffects=TRUE)

manhattanPlot(RRBLUP_Trait1, gpCode, colored=TRUE, pch=19, cex=0.5)
```



Teşekkürler