Missing Data Analysis-ALL

This section presents an exploratory analysis of missing data in the ALL dataset.

```
#Load data set
df.all <-  read.csv("data_lla.csv",
                    header = TRUE, stringsAsFactors = FALSE)
#See first 3 rows
head(df.all,3)
```

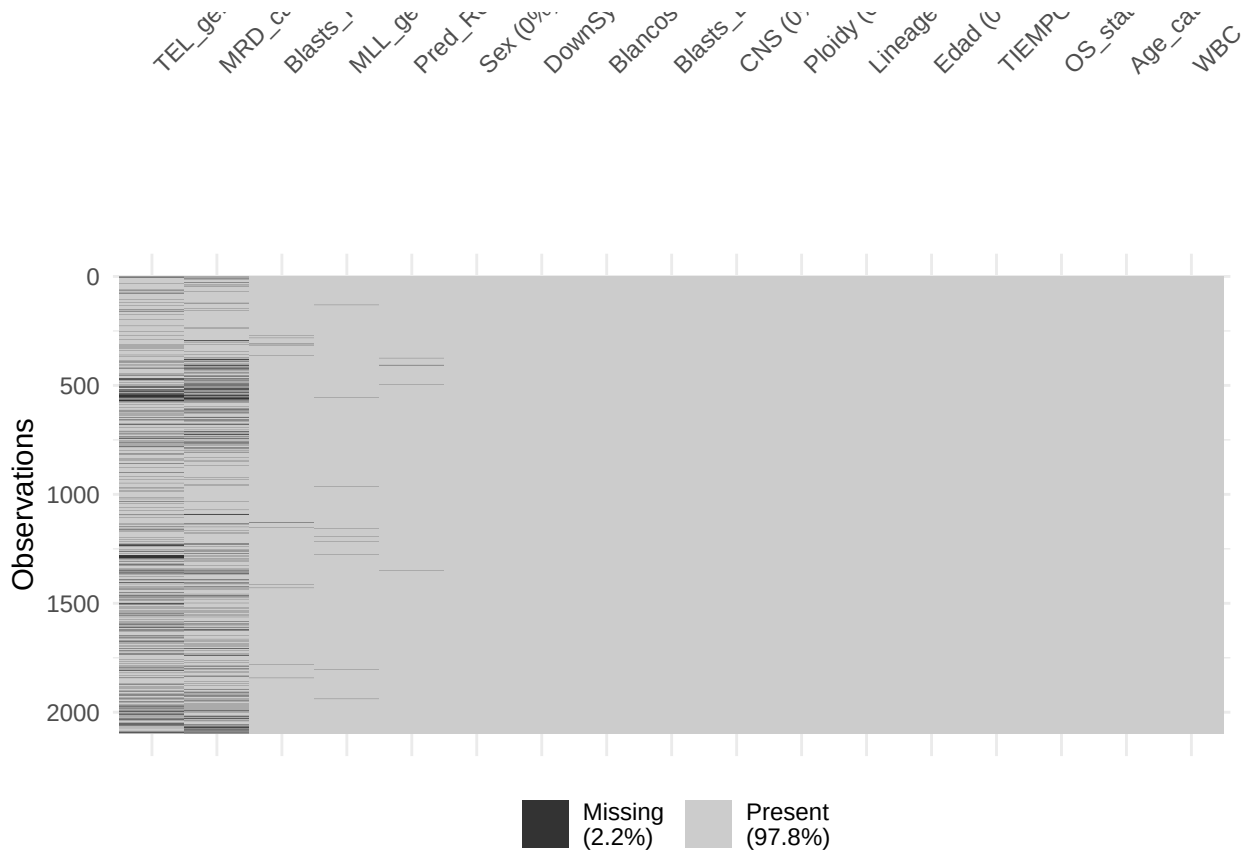
```
##   Sex DownSyndrome Blancos Blasts_PB Blasts_BM CNS Ploidy Lineage Pred_Response
## 1   1            0       3         2        90   1      1       B             1
## 2   1            0      36        48        50   1      1       T             0
## 3   0            0      10        95        98   1      2       B             0
##   MRD_cat Edad MLL_gene TIEMPOSG OS_status Age_cat WBC_cat TEL_gene
## 1      NA    2        0       81         0       1       1        0
## 2       3   14        0       16         1       2       2       NA
## 3       3    4        0       67         0       1       1        0
```

```
# Load required packages
library(dplyr)
library(ggplot2)
library(naniar)
library(visdat)
```

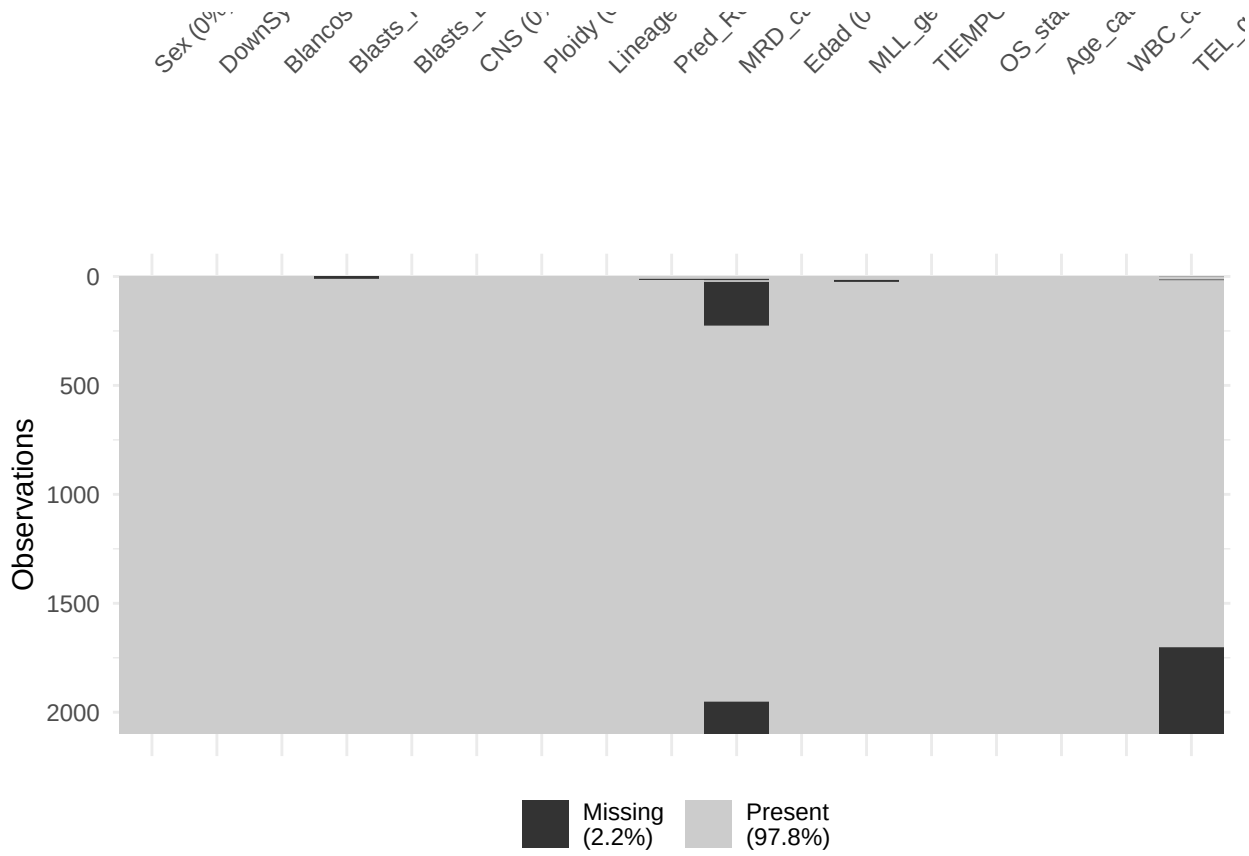
```
# Summary of missing values per variable
colSums(is.na(df.all))
```

```
##           Sex  DownSyndrome       Blancos     Blasts_PB     Blasts_BM
##             0             0             0            13             0
##           CNS        Ploidy       Lineage Pred_Response       MRD_cat
##             0             0             0             5           354
##         Edad      MLL_gene     TIEMPOSG     OS_status       Age_cat
##             0             9             0             0             0
##       WBC_cat      TEL_gene
##             0           401
```

```
# Visualize missing data pattern
vis_miss(df.all, sort_miss = TRUE)
```




```
# Clustered visualization of missingness
vis_miss(df.all, cluster = TRUE)
```




```
# Visualize data types and missing values
vis_dat(df.all)
```


Lineage
Sex
DownSy
Blancos
Blasts_
Blasts_
CNS
Ploidy
Pred_R
MRD_c
Edad
MLL_ge
TIEMPO
OS_sta
Age_ca
WBC_c
TEL_ge
Observations
0
500
1000
1500
2000
Type
character
integer
NA