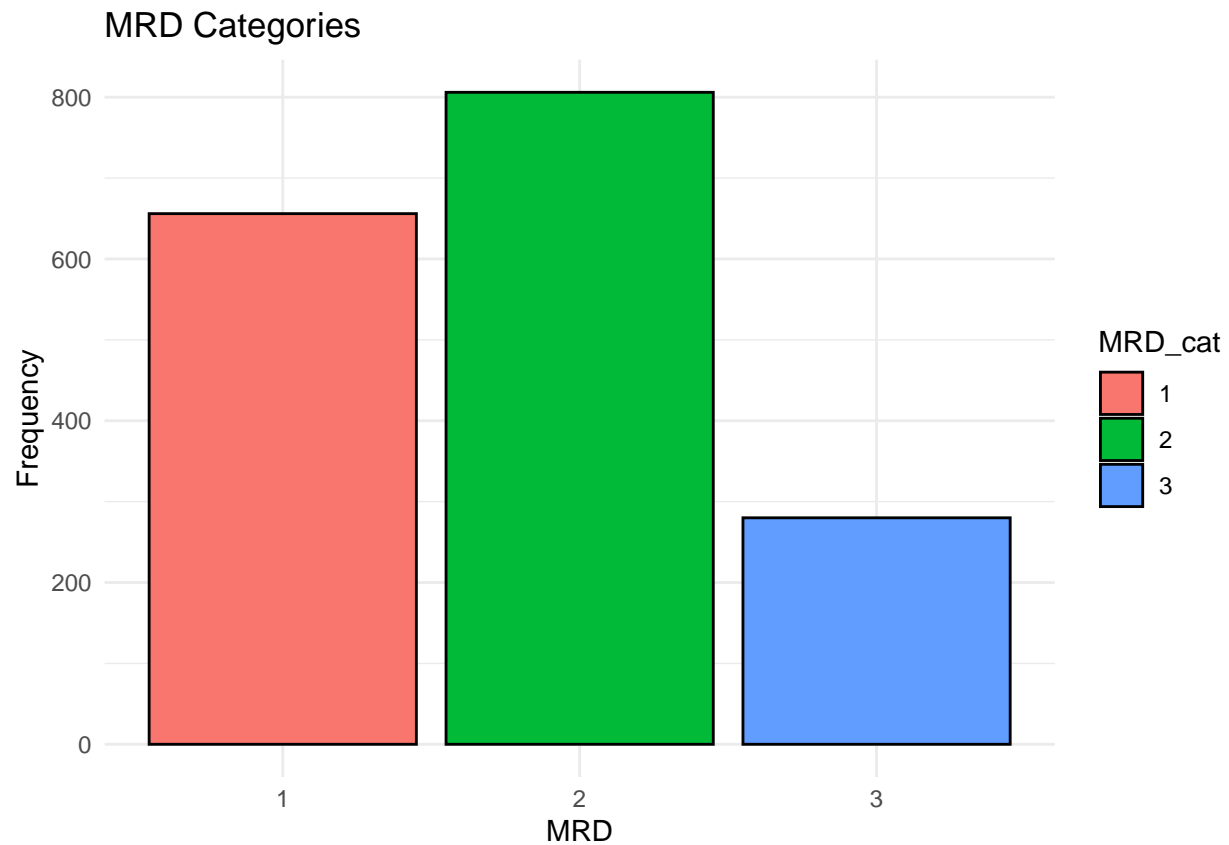# Data Visualization- ALL

```r
#Load data set
df.all <-  read.csv("data_lla.csv",
                    header = TRUE, stringsAsFactors = FALSE)
```

```r
# Load required packages
library(dplyr)
library(ggplot2)
library(highcharter)
```

```r
#Prepare variables
df.all <- df.all %>%
  mutate(
    MRD_cat = factor(MRD_cat),
    Sex = factor(Sex, levels = c(0,1), labels = c("Female","Male")),
    DownSyndrome = factor(DownSyndrome, levels = c(0,1), labels = c("No","Yes")),
    OS_status = factor(OS_status, levels = c(0,1), labels = c("Alive","Dead")),
    Lineage = factor(Lineage),
    Pred_Response = factor(Pred_Response, levels = c(0,1), labels = c("Poor","Good")),
    CNS = factor(CNS),
    Ploidy = factor(Ploidy),
    Age_cat = factor(Age_cat),
    WBC_cat = factor(WBC_cat),
    TEL_gene = factor(TEL_gene, levels = c(0,1), labels = c("Negative","Positive")),
    MLL_gene = factor(MLL_gene, levels = c(0,1), labels = c("Negative","Positive"))
  )
```
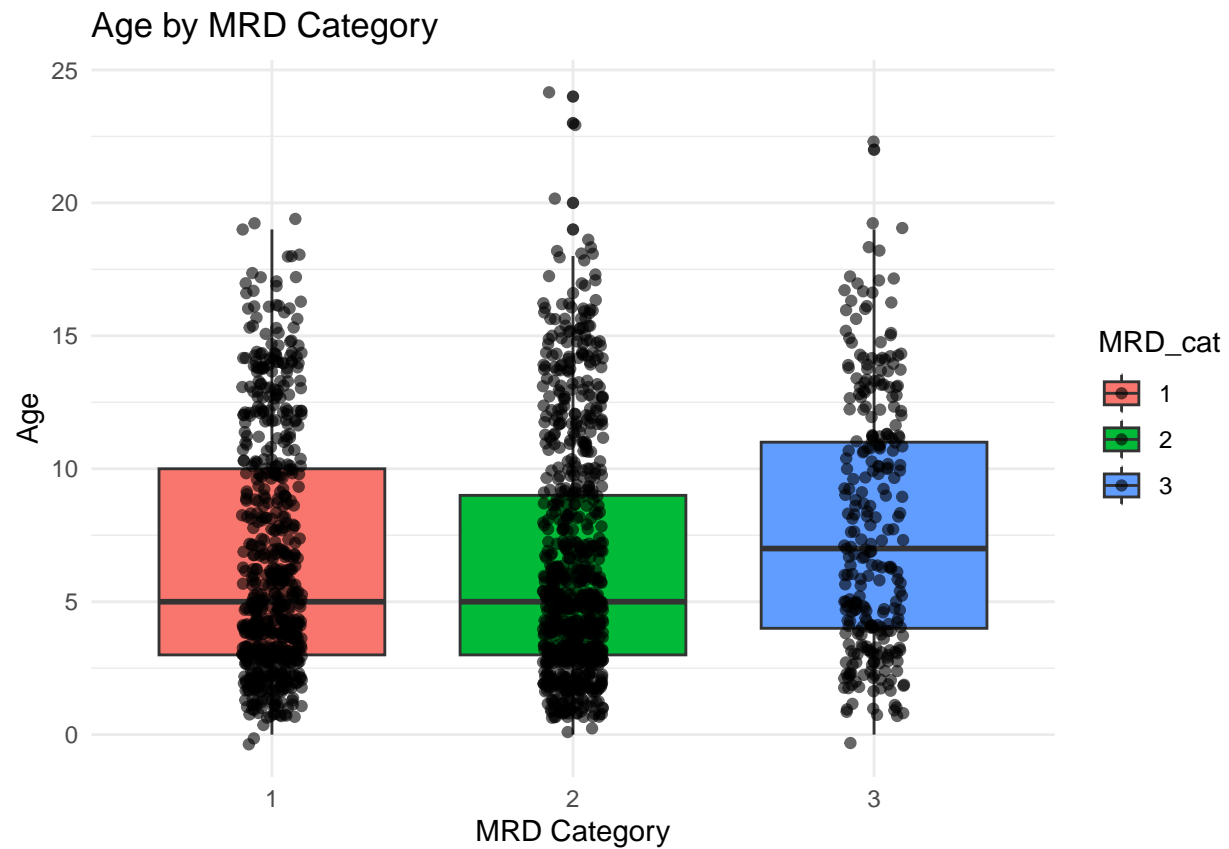
MRD Categories - Bar Plot

```r
ggplot(df.all %>% filter(!is.na(MRD_cat)), aes(x = MRD_cat, fill = MRD_cat)) +
  geom_bar(color = "black") +
  labs(title = "MRD Categories",
       x = "MRD",
       y = "Frequency") +
  theme_minimal()
```
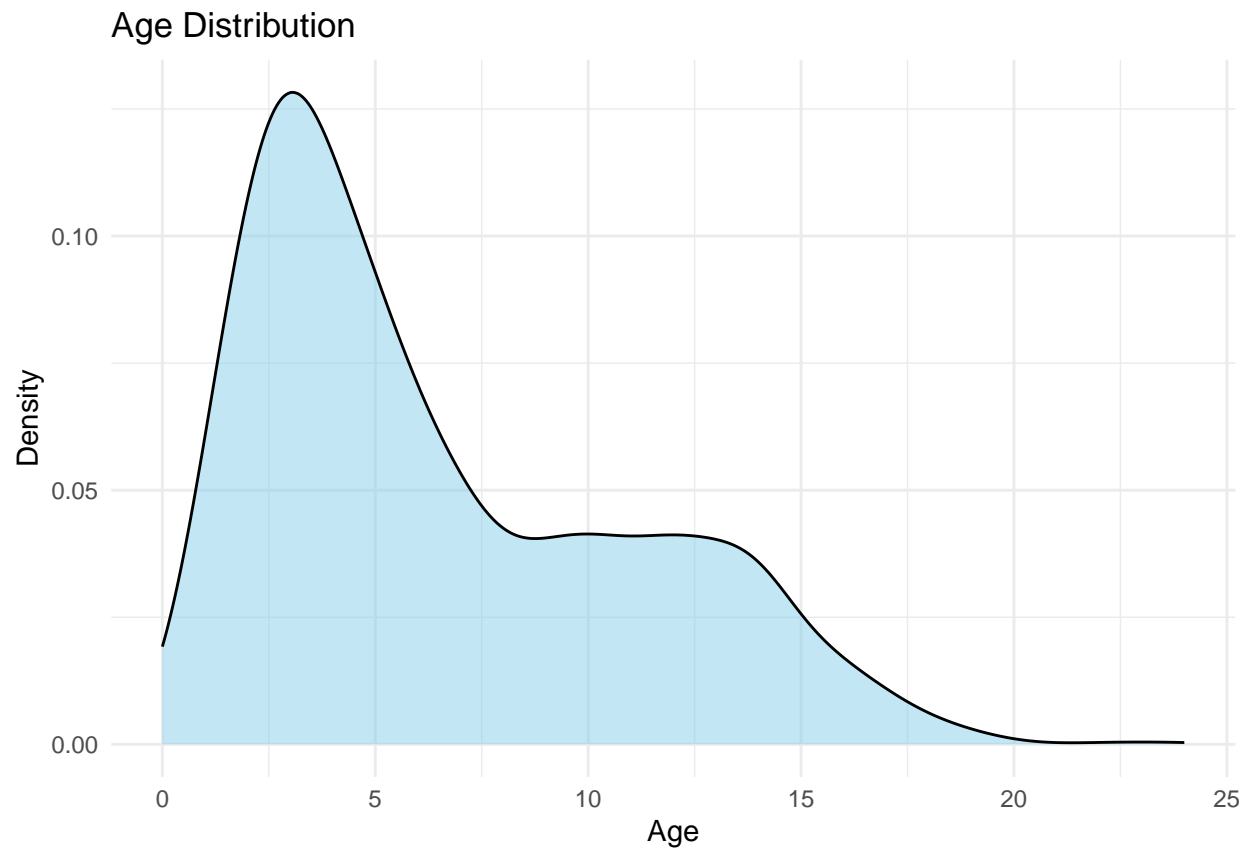
MRD Categories

Age

Age by MRD Category

```r
ggplot(df.all %>% filter(!is.na(MRD_cat) & !is.na(Edad)),
       aes(x = MRD_cat, y = Edad, fill = MRD_cat)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.6) +
  labs(title = "Age by MRD Category",
       x = "MRD Category",
       y = "Age") +
  theme_minimal()
```

Age by MRD Category
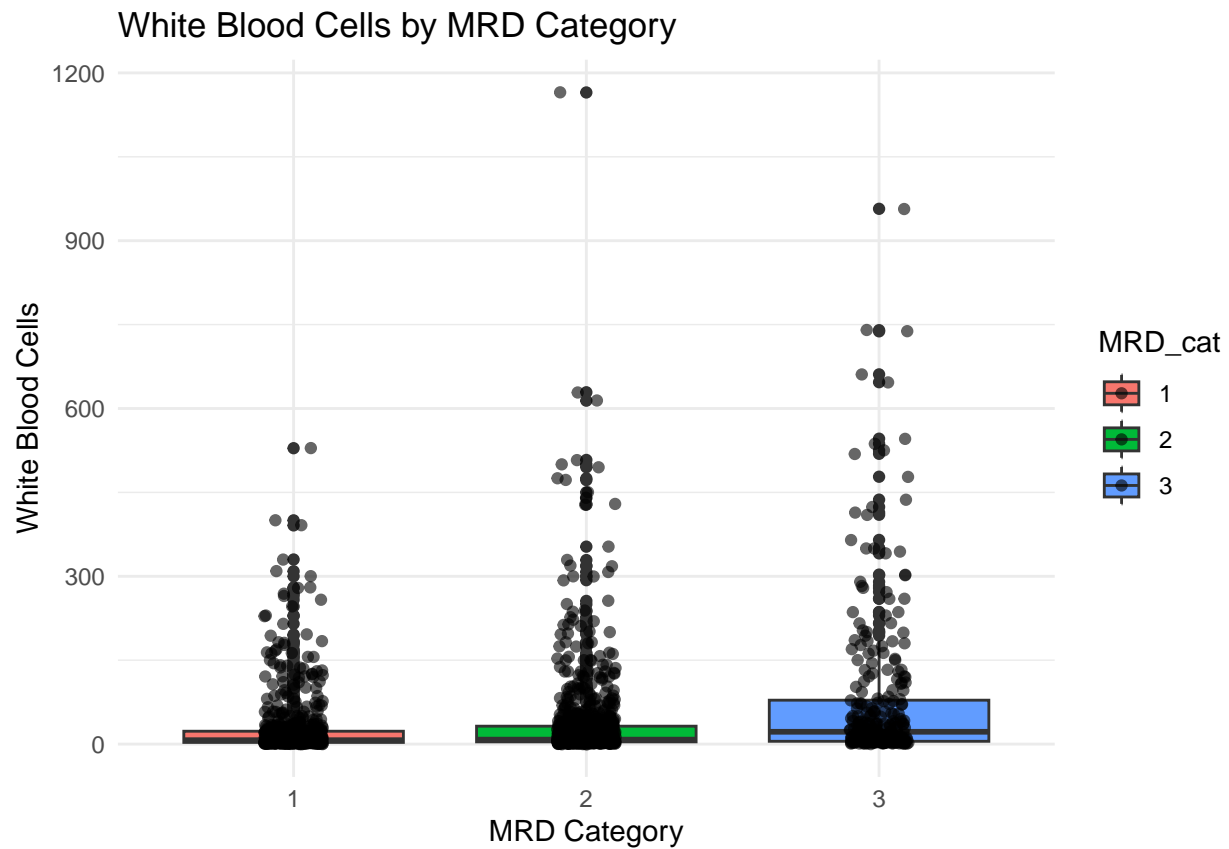

Age Density

```
ggplot(df.all %>% filter(!is.na(Edad)), aes(x = Edad)) +
  geom_density(fill = "skyblue", alpha = 0.5) +
  labs(title = "Age Distribution",
       x = "Age",
       y = "Density") +
  theme_minimal()
```

Age Distribution

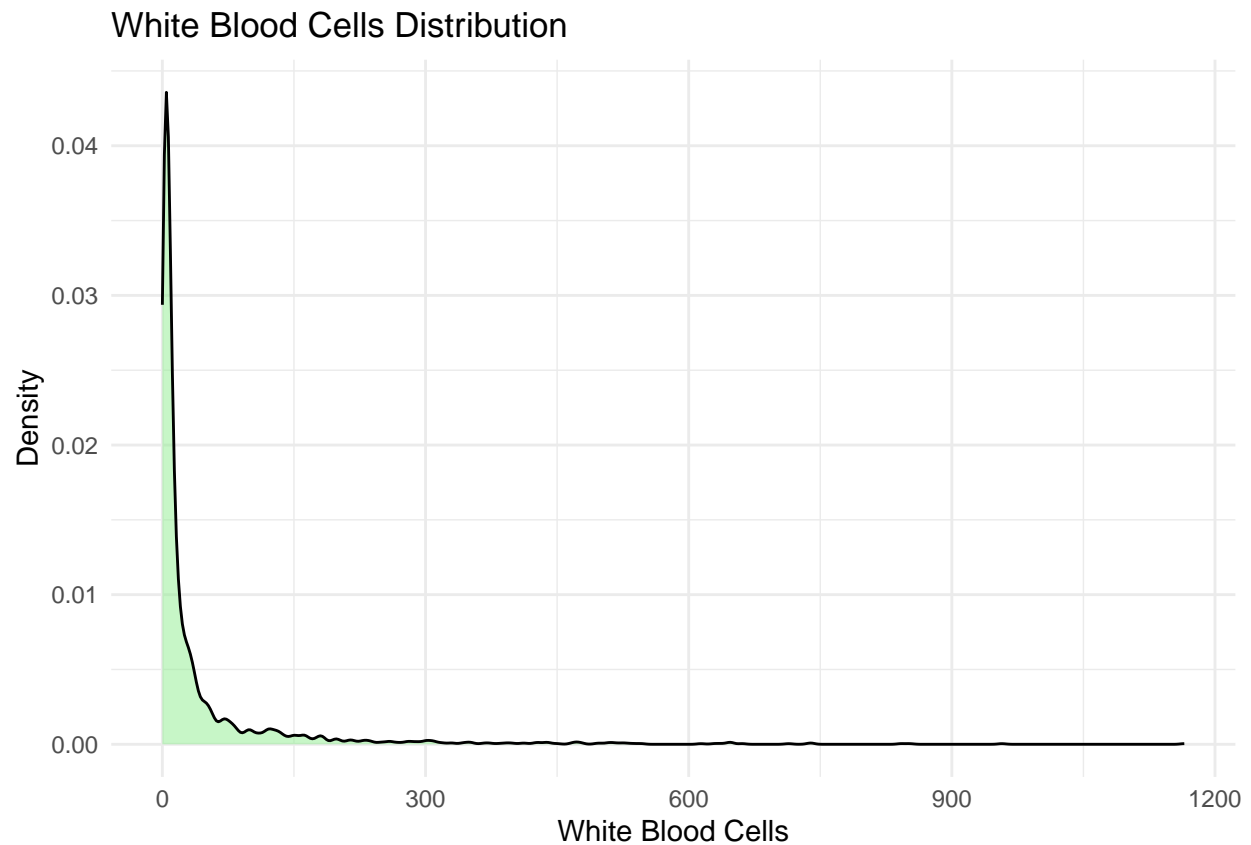

White Blood Cell Count (Blancos)

WBC by MRD Category

```
ggplot(df.all %>% filter(!is.na(MRD_cat) & !is.na(Blancos)),
       aes(x = MRD_cat, y = Blancos, fill = MRD_cat)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.6) +
  labs(title = "White Blood Cells by MRD Category",
       x = "MRD Category",
       y = "White Blood Cells") +
  theme_minimal()
```

White Blood Cells by MRD Category
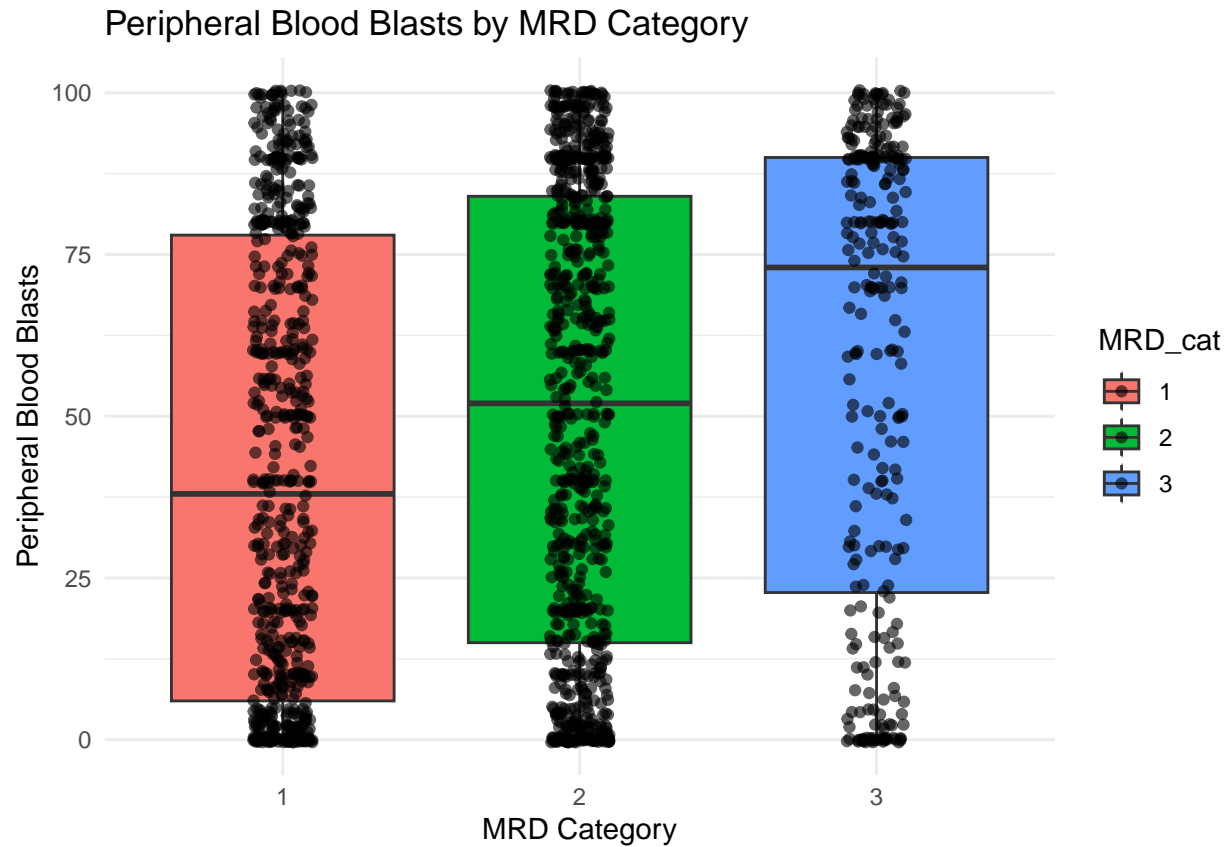
WBC Density

```r
ggplot(df.all %>% filter(!is.na(Blancos)), aes(x = Blancos)) +
  geom_density(fill = "lightgreen", alpha = 0.5) +
  labs(title = "White Blood Cells Distribution",
       x = "White Blood Cells",
       y = "Density") +
  theme_minimal()
```

White Blood Cells Distribution
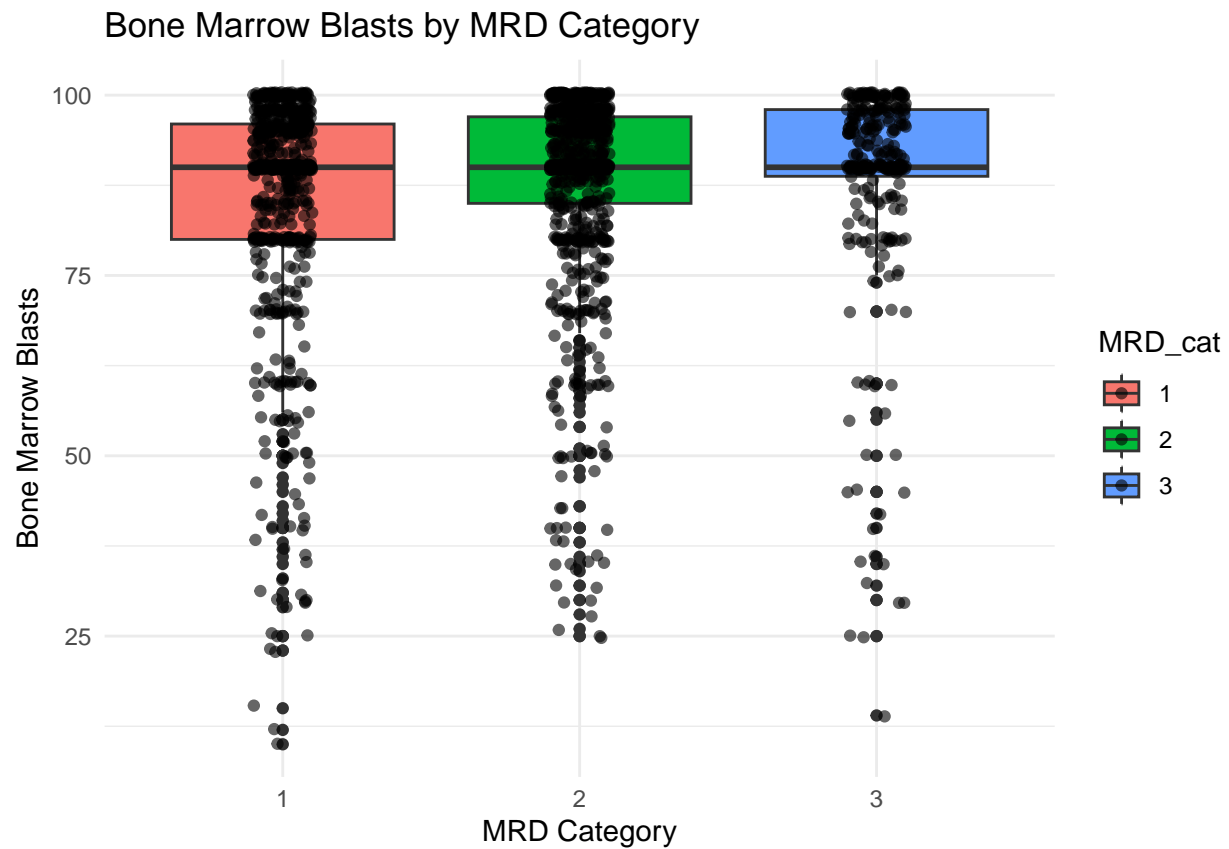

Blasts

Peripheral Blood

```
ggplot(df.all %>% filter(!is.na(MRD_cat) & !is.na(Blasts_PB)),
       aes(x = MRD_cat, y = Blasts_PB, fill = MRD_cat)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.6) +
  labs(title = "Peripheral Blood Blasts by MRD Category",
       x = "MRD Category",
       y = "Peripheral Blood Blasts") +
  theme_minimal()
```
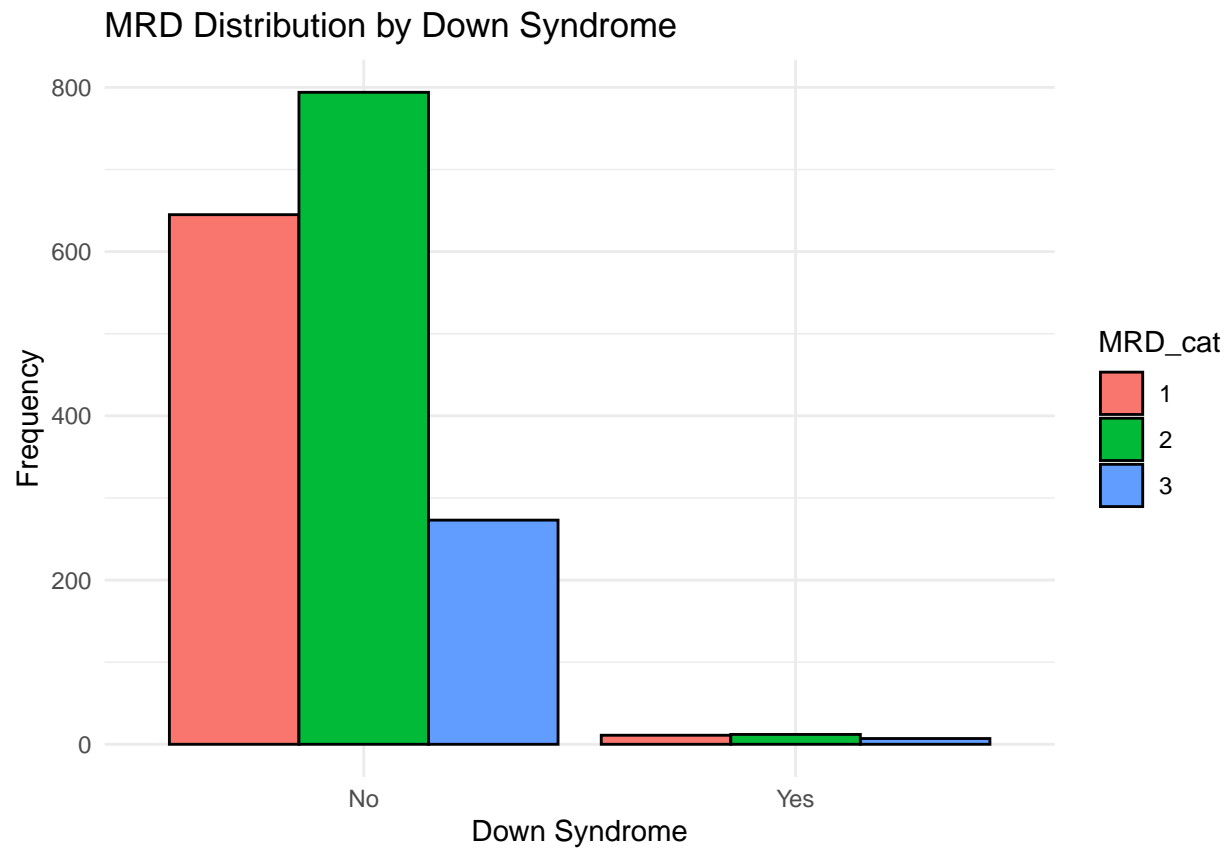
Bone Marrow

```r
ggplot(df.all %>% filter(!is.na(MRD_cat) & !is.na(Blasts_BM)),
       aes(x = MRD_cat, y = Blasts_BM, fill = MRD_cat)) +
  geom_boxplot() +
  geom_jitter(width = 0.1, alpha = 0.6) +
  labs(title = "Bone Marrow Blasts by MRD Category",
       x = "MRD Category",
       y = "Bone Marrow Blasts") +
  theme_minimal()
```
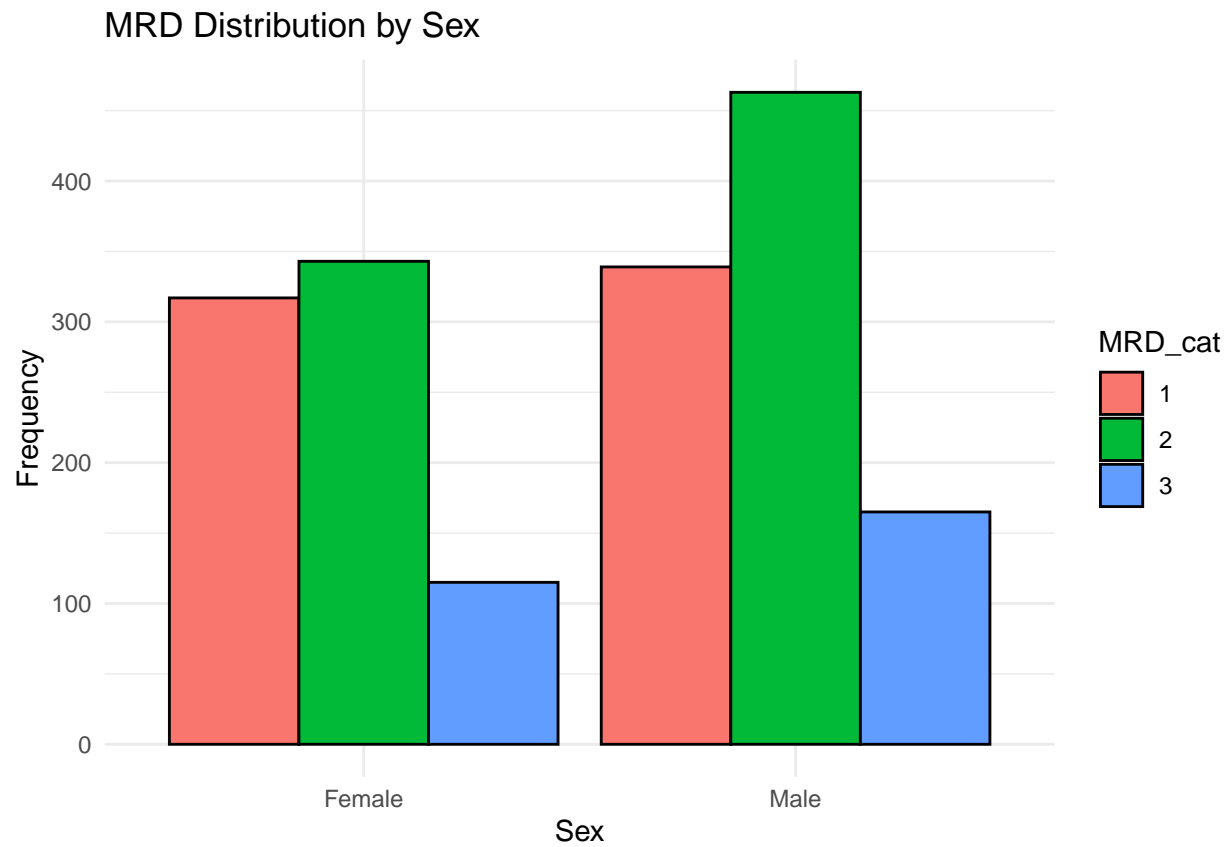
Bone Marrow Blasts by MRD Category

Down Syndrome

```r
ggplot(df.all %>% filter(!is.na(DownSyndrome) & !is.na(MRD_cat)),
       aes(x = DownSyndrome, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by Down Syndrome",
       x = "Down Syndrome",
       y = "Frequency") +
  theme_minimal()
```

MRD Distribution by Down Syndrome
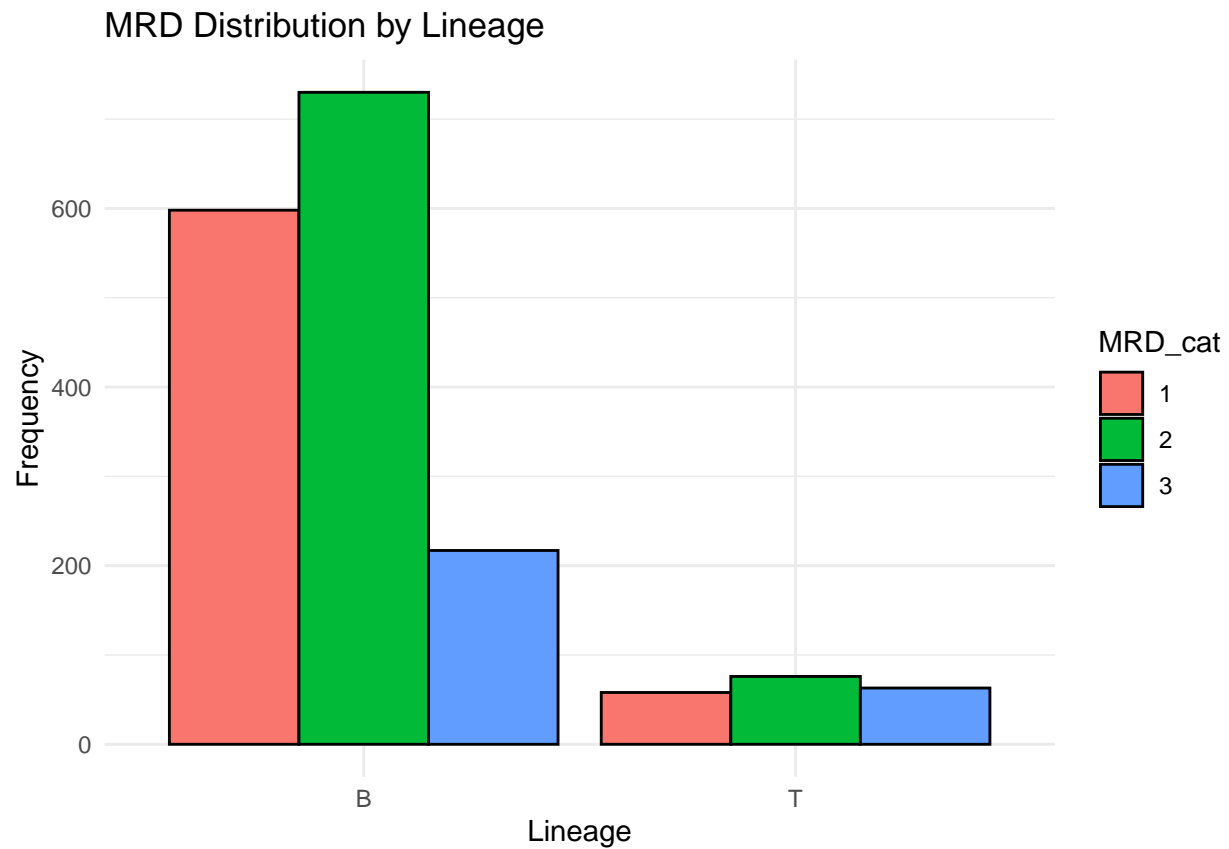

Sex

```
ggplot(df.all %>% filter(!is.na(Sex) & !is.na(MRD_cat)),
       aes(x = Sex, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by Sex",
       x = "Sex",
       y = "Frequency") +
  theme_minimal()
```
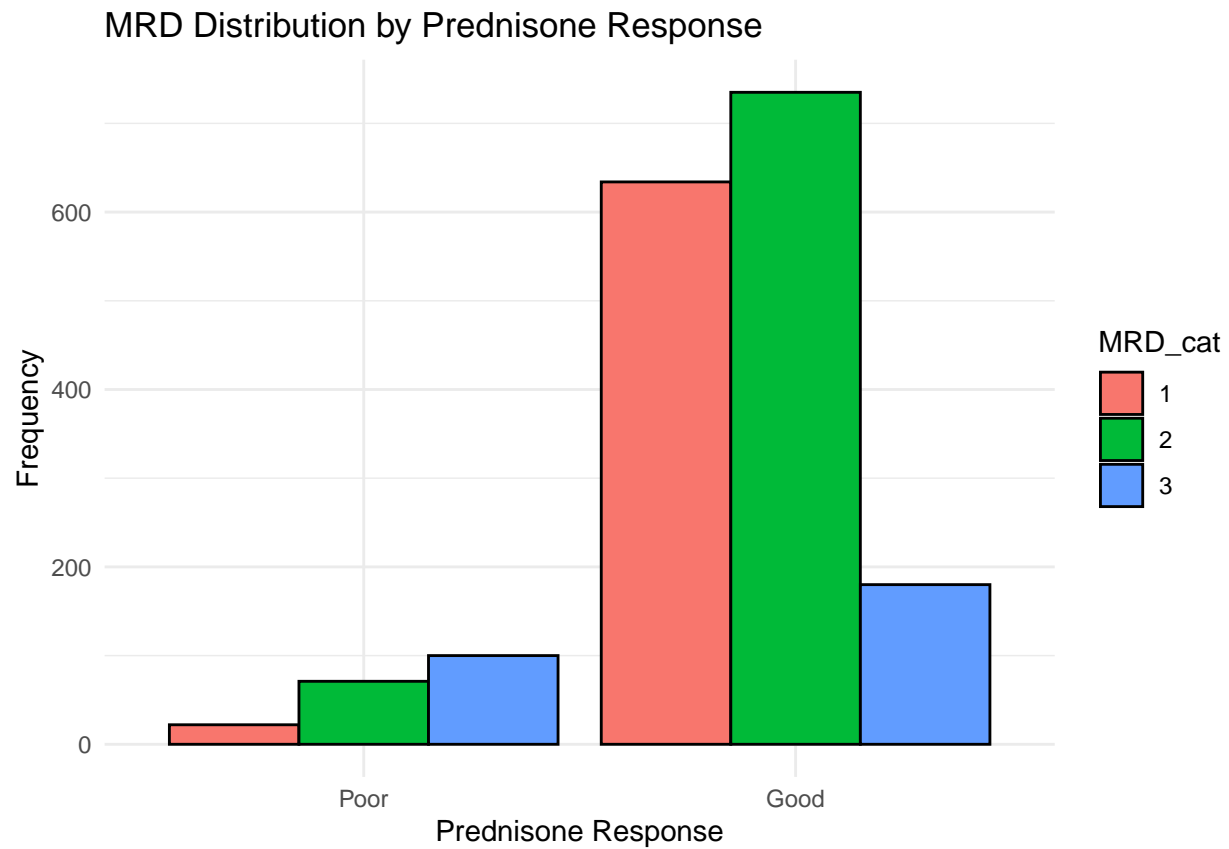
MRD Distribution by Sex

Lineage

```r
ggplot(df.all %>% filter(!is.na(Lineage) & !is.na(MRD_cat)),
       aes(x = Lineage, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by Lineage",
       x = "Lineage",
       y = "Frequency") +
  theme_minimal()
```

MRD Distribution by Lineage
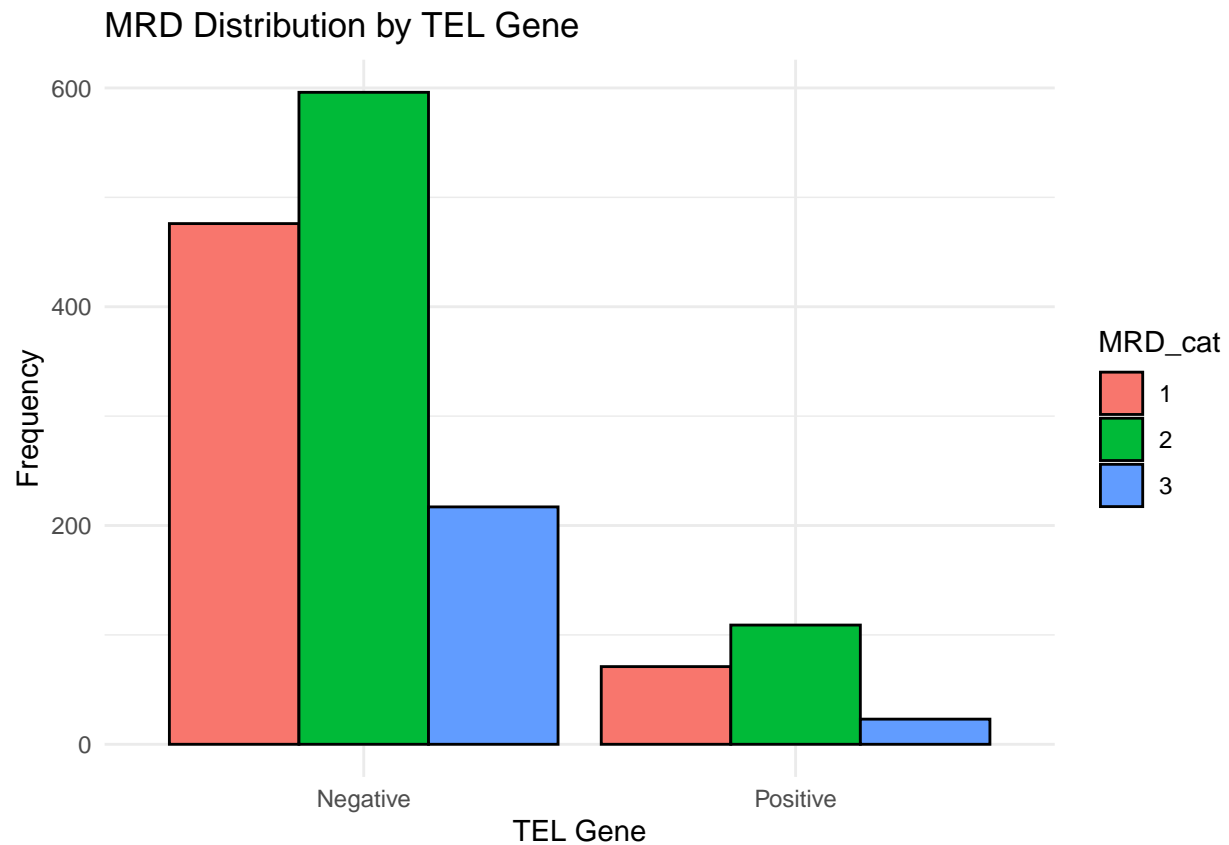

Prednisone Response

```
ggplot(df.all %>% filter(!is.na(Pred_Response) & !is.na(MRD_cat)),
       aes(x = Pred_Response, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by Prednisone Response",
       x = "Prednisone Response",
       y = "Frequency") +
  theme_minimal()
```

MRD Distribution by Prednisone Response

Genetic Markers

TEL Gene

```
ggplot(df.all %>% filter(!is.na(TEL_gene) & !is.na(MRD_cat)),
       aes(x = TEL_gene, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by TEL Gene",
       x = "TEL Gene",
       y = "Frequency") +
  theme_minimal()
```
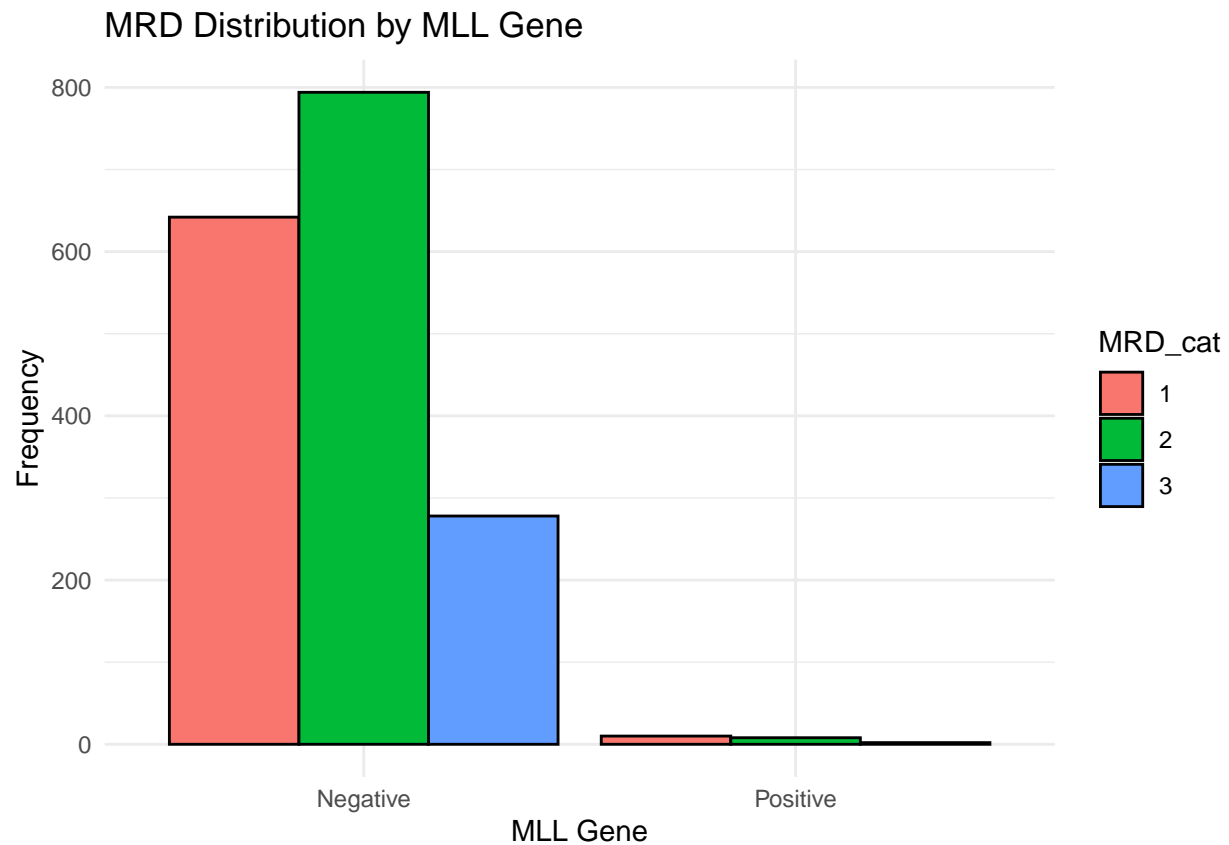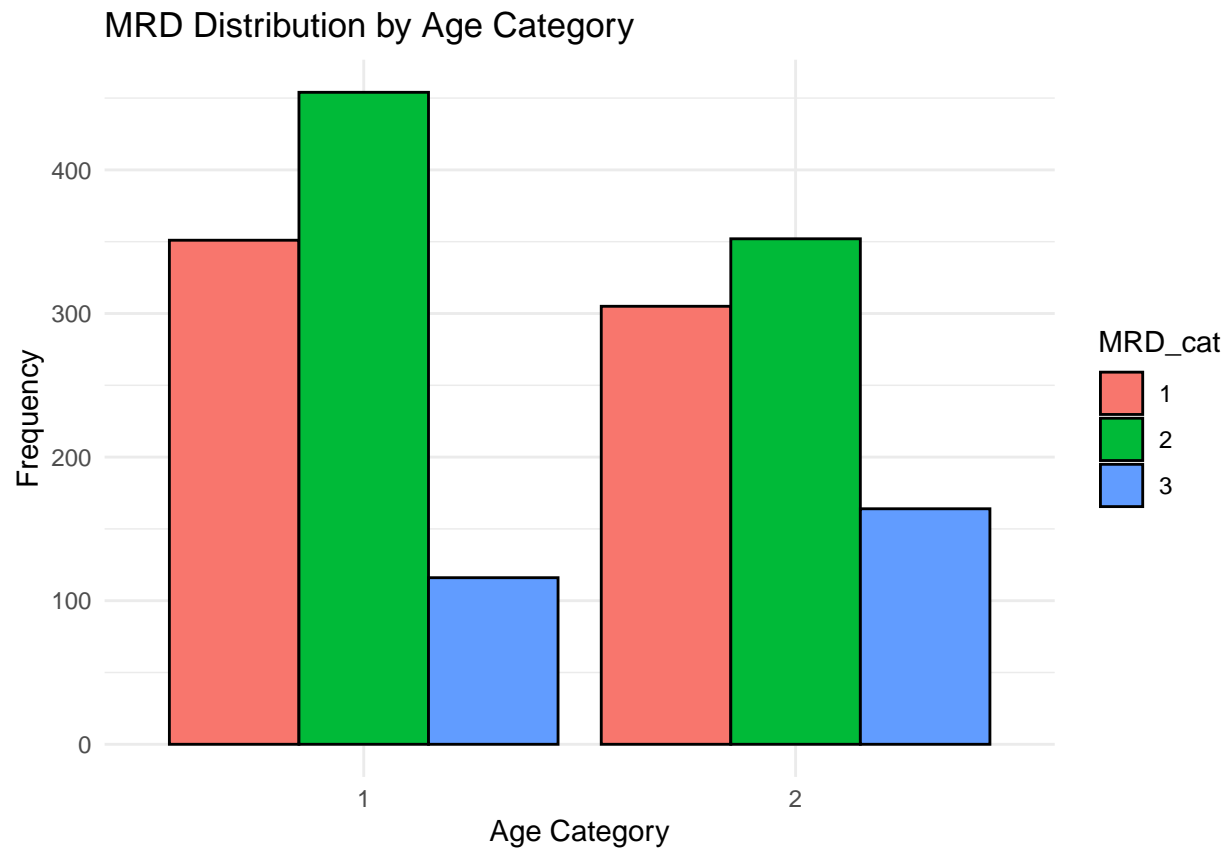
MRD Distribution by TEL Gene

MLL Gene

```
ggplot(df.all %>% filter(!is.na(MLL_gene) & !is.na(MRD_cat)),
       aes(x = MLL_gene, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by MLL Gene",
       x = "MLL Gene",
       y = "Frequency") +
  theme_minimal()
```
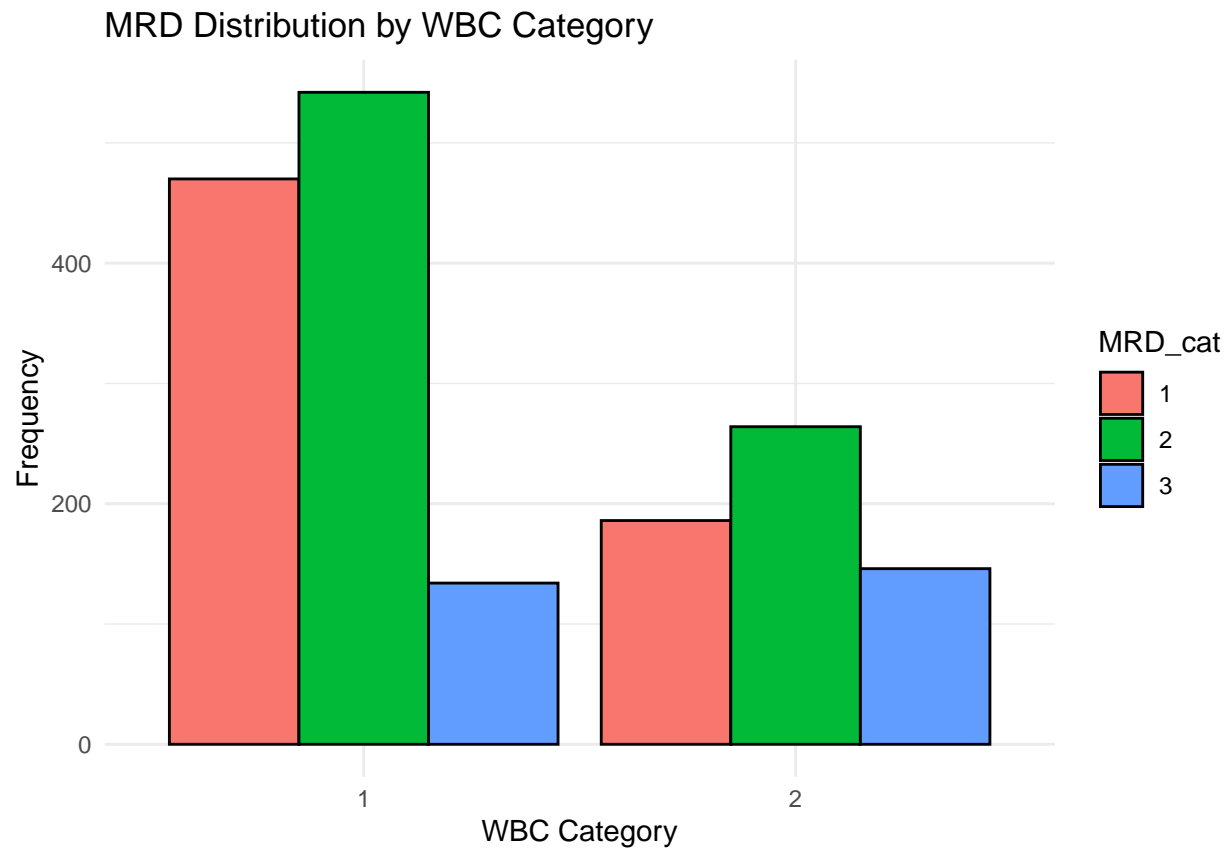
MRD Distribution by MLL Gene

Age Category

```r
ggplot(df.all %>% filter(!is.na(Age_cat) & !is.na(MRD_cat)),
       aes(x = Age_cat, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by Age Category",
       x = "Age Category",
       y = "Frequency") +
  theme_minimal()
```

MRD Distribution by Age Category

WBC Category

```
ggplot(df.all %>% filter(!is.na(WBC_cat) & !is.na(MRD_cat)),
       aes(x = WBC_cat, fill = MRD_cat)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "MRD Distribution by WBC Category",
       x = "WBC Category",
       y = "Frequency") +
  theme_minimal()
```

MRD Distribution by WBC Category

Survival Sankey Plot

```
df.all$Survival <- df.all$OS_status
# Filter out NAs in MRD_cat and Survival
df.clean <- df.all %>%
  filter(!is.na(MRD_cat), !is.na(Survival)) %>%
  select(MRD_cat, Survival)

# Create the Sankey plot
hchart(
  data_to_sankey(df.clean),
  "sankey",
  name = "Survival by MRD Category"
)
```

```
#Save sankey plot HTML
library(htmlwidgets)
```

```
sankey_widget <- hchart(data_to_sankey(df.clean), "sankey", name = "Survival by MRD Category")
saveWidget(sankey_widget, "sankey_temp.html", selfcontained = TRUE)

# Convert to png formAT
webshot2::webshot("sankey_temp.html", "sankey_plot.png", vwidth = 1000, vheight = 800)
```