# EPIDEMIOLOGY OF MULTI-DRUG RESISTANT TB USING ISOLATES SENT TO THE NATIONAL TUBERCULOSIS REFERENCE LABORARY (2015-2016)

METHODS

```r
library('ggflowchart')
flow_chart <- tibble::tibble(from = c("Total TB+ Entries recieved
  8308", "Total TB+ Entries recieved
  8308", "Entries after removing null values
  2621", "Entries after removing null values
  2621", "Null entries removed
  2621"),
                                to= c("Entries after removing null values
  2621","Entries removed
  5682", "Entries after removing Non-MDR
  143
  ", "Null entries removed
  2621", "MDR entries analyzed
  143"))
ggflowchart(flow_chart, colour = 'cyan', family = 'serif', x_nudge = 0.45)
```

```
Total TB+ Entries recieved
8308
```

```
Entries after removing null values          Entries removed
2621                                        5682
```

```
Null entries removed          Entries after removing Non−MDR
2621                          143
```

```
MDR entries analyzed
143
```
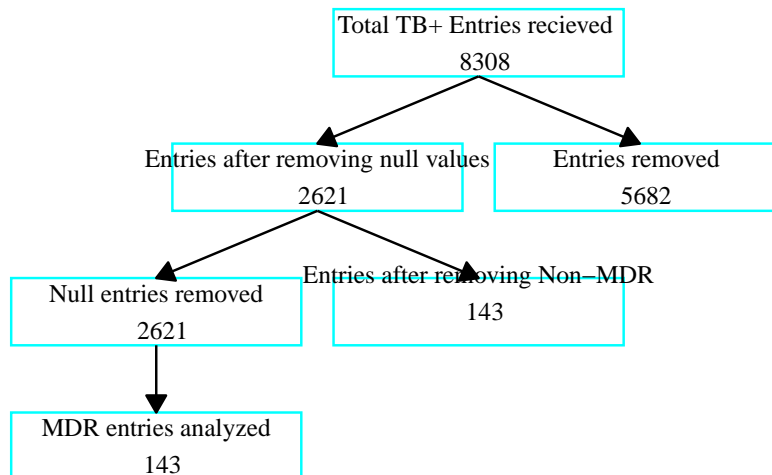
R Packages used

```r
#\ label: load-packages
suppressPackageStartupMessages(library('tidyverse'))

#install.packages('plotly')

#install.packages('ggmap')

#install.packages('forcats')
suppressPackageStartupMessages(library('ggmap'))
suppressPackageStartupMessages(library('plotly'))
suppressPackageStartupMessages(library('ggdist'))
suppressPackageStartupMessages(library('ggmap'))
```

FUNCTIONS

Fisher's function

```r
Fishersfn <- function(df, var1, var2){

  contingency_table <- table(df[[var1]], df[[var2]])

  fisherExact <- fisher.test(contingency_table)

  return (fisherExact)

}
```

Getting percentages

```r
pcts <- function(df, var_to_group_by) {

  results <- df %>%

    group_by_at(vars({{ var_to_group_by }})) %>%

    summarise(counts = n()) %>%

    mutate(Percentages = counts/sum(counts)*100)


  return(results)

}
```

Reading the file

```r
mdr_tb_df <- read.csv('mdr_tbv4.csv')
```

Creating a new column to define MDR_TYPE

```r
mdr_tb_df$MDR_TYPE <- NA
```

Here I have defined Rpob + katg as MDRKR (MDR KatG and RpoB), as well
as those that are rpob + katg + inha

```r
mdr_tb_dfv1 <- mdr_tb_df$MDR_TYPE[mdr_tb_df$RpoB == 'Resistant' & mdr_tb_df$KatG == 'Resi

mdr_tb_df$MDR_TYPE[mdr_tb_df$RpoB == 'Resistant' & mdr_tb_df$KatG != 'Resistant' & mdr_tb

mdr_tb_dfv1 <- mdr_tb_df %>% select(-RpoB, -KatG, -inhA)
```

4

VISUALIZATIONS

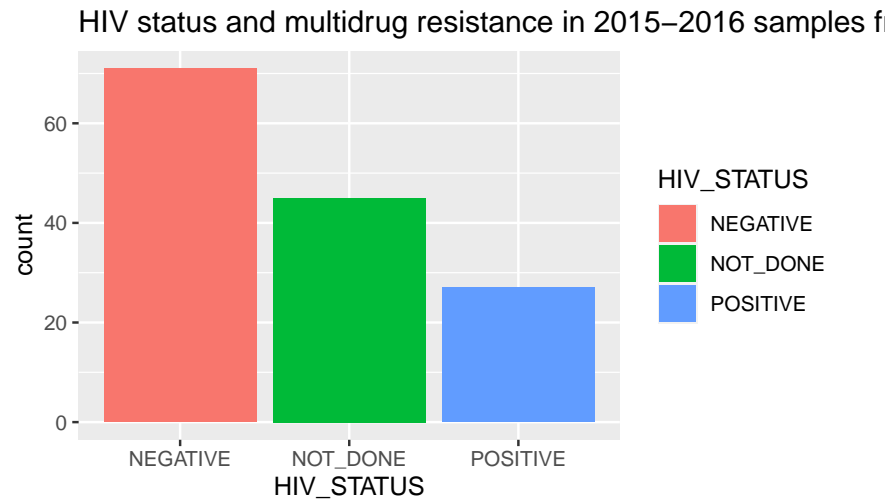Setting theme

```
theme_set(theme_gray())

theme_update(

  plot.margin = margin(rep(20, 4))

)
```

HIV BAR PLOT

```
hivbar <- ggplot(mdr_tb_dfv1, mapping = aes(x=HIV_STATUS)) + geom_bar(mapping = aes(fill
hivbar
```
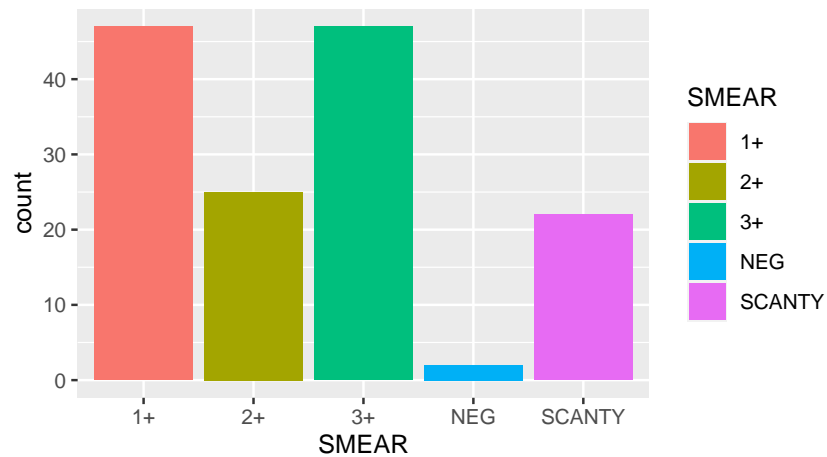
HIV status and multidrug resistance in 2015–2016 samples f[...]

TB TYPE BAR PLOT

```
tb_type_bar <- ggplot(mdr_tb_dfv1, mapping = aes(x=TB_TYPE)) + geom_bar(mapping = aes(fil
```

SMEAR BAR PLOT

```
smearBar <- ggplot(mdr_tb_dfv1, mapping = aes(x=SMEAR)) + geom_bar(mapping = aes(fill = S
smearBar
```

Smear results and multidrug resistance in 2015–2016 sample

Age groups based on WHO typical age divisions

```r
mdr_tb_dfv3 <- mdr_tb_dfv1 %>%   mutate(
  # Create categories
  age_group = dplyr::case_when(
    AGE <= 4              ~ "0-4",
    AGE > 4 & AGE <= 14 ~ "5-14",
    AGE > 14 & AGE<= 24 ~ "15-24",
    AGE > 24 & AGE <= 34  ~ "25-34",
    AGE > 34 & AGE <= 44 ~ "35-44",
    AGE > 44 & AGE <= 54 ~ "45-54",
    AGE > 54 & AGE <= 64 ~ "65+",
    AGE >= 64 ~ "35-44"
  ),
  # Convert to factor
  age_group = factor(
    age_group,
    level = c("0-4", "5-14","15-24","25-34","35-44","45-54", "65+")
  )
)
```

LOLLIPOP PLOT FOR COUNTY/GEOGRAPHICAL DISTRIBUTION

```
geoGrpsPcts <- pcts(mdr_tb_dfv1, 'Province')

GeoPops <- geoGrpsPcts %>%

  arrange(desc(Percentages)) %>%

  mutate( Province = forcats::fct_reorder(Province, Percentages)) %>%

  ggplot(aes(x=Province, y=Percentages)) +

  geom_point(color = 'orange') +

  geom_segment(aes(x=Province, y=Percentages, xend=Province, yend=0, color = 'red')) +

  theme(axis.text.x = element_text(angle = 90), panel.background = element_rect(fill = 'l

  labs(title = 'MDR TB: Percentage of MDR TB cases by region/county reported to the NTRL'

  guides(color = 'none')
GeoPops
```
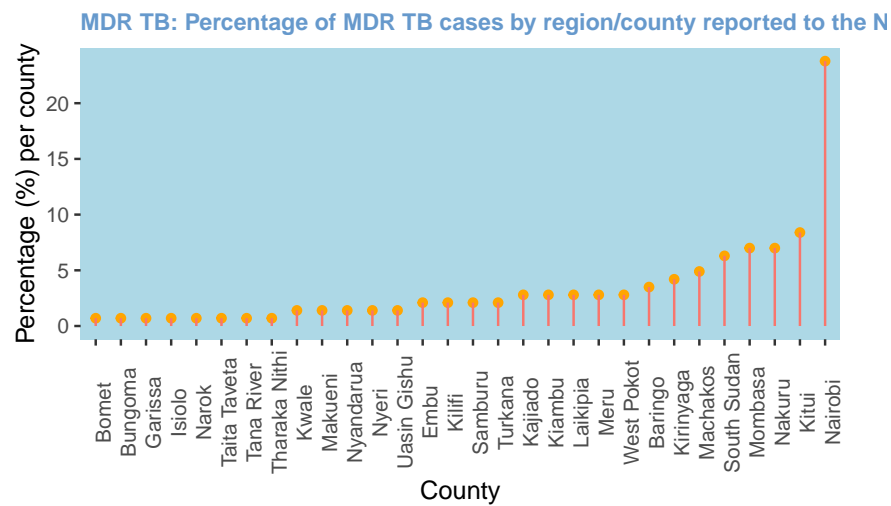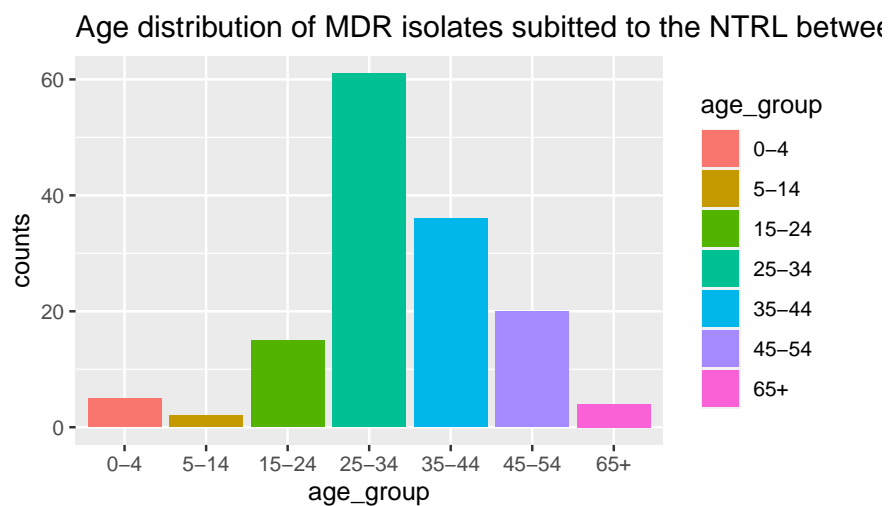

8

AGE GROUP BAR PLOT

```
age_groupBars <- mdr_tb_dfv3 %>%

  group_by(age_group) %>%

  summarise(counts = n())

ggplot() +

  geom_bar(data = age_groupBars, aes(x = age_group, y = counts, fill = age_group), stat =

  labs(title = 'Age distribution of MDR isolates subitted to the NTRL between 2015-2015')
```



```
  age_groupBars
```

```
# A tibble: 7 x 2
  age_group counts
  <fct>      <int>
1 0-4            5
2 5-14           2
3 15-24         15
4 25-34         61
5 35-44         36
```

```
6  45-54         20
7  65+            4
```

INFERENTIAL STATISTICS

Fishers exact test for HIV status

```
p_value <- Fishersfn(mdr_tb_dfv3, 'HIV_STATUS', 'MDR_TYPE')
p_value
```

    Fisher's Exact Test for Count Data

data: contingency_table
p-value = 0.2897
alternative hypothesis: two.sided

Fisher's exact test for age

```
p_value <- Fishersfn(mdr_tb_dfv3, 'age_group', 'MDR_TYPE')
p_value
```

    Fisher's Exact Test for Count Data

data: contingency_table
p-value = 0.175
alternative hypothesis: two.sided

Fisher's exact test for COUNTY

```r
p_value <- Fishersfn(mdr_tb_dfv1, 'Province', 'MDR_TYPE')
p_value
```

```
    Fisher's Exact Test for Count Data

data:  contingency_table
p-value = 0.06311
alternative hypothesis: two.sided
```

Fisher's exact test for TB type

```r
p_value <- Fishersfn(mdr_tb_dfv1, 'TB_TYPE', 'MDR_TYPE')
p_value
```

```
    Fisher's Exact Test for Count Data

data:  contingency_table
p-value = 0.9482
alternative hypothesis: two.sided
```

Fisher's exact test for GENDER

```r
p_value <- Fishersfn(mdr_tb_dfv1, 'GENDER', 'MDR_TYPE')
p_value
```

```
    Fisher's Exact Test for Count Data

data:  contingency_table
p-value = 0.3319
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.5015292 9.1443367
sample estimates:
odds ratio
  2.088592
```

Fisher's exact test for SMEARS

```r
p_value <- Fishersfn(mdr_tb_dfv1, 'SMEAR', 'MDR_TYPE')
p_value
```

```
	Fisher's Exact Test for Count Data

data:  contingency_table
p-value = 0.7377
alternative hypothesis: two.sided
```

More characterizations

```r
mdr_tb_dfv5 <- mdr_tb_df$MDR_TYPE[mdr_tb_df$RpoB == 'Resistant' & mdr_tb_df$KatG == 'Resi
```

Percentages of the MDR types

```
pcts(mdr_tb_df, 'MDR_TYPE')
```

```
# A tibble: 3 x 3
  MDR_TYPE counts Percentages
  <chr>     <int>       <dbl>
1 MDR+          7        4.90
2 MDRIR        11        7.69
3 MDRKR       125       87.4
```

TABLES OF DEMOGRAPHIC DISTRIBUTION OF MDR TYPES

```
table(mdr_tb_df$Province, mdr_tb_df$MDR_TYPE)
```

	MDR+	MDRIR	MDRKR
Baringo	1	0	4
Bomet	0	1	0
Bungoma	0	0	1
Embu	0	0	3
Garissa	1	0	0
Isiolo	0	0	1
Kajiado	1	0	3
Kiambu	0	0	4
Kilifi	0	0	3
Kirinyaga	0	1	5
Kitui	0	1	11
Kwale	0	0	2
Laikipia	0	0	4
Machakos	0	0	7
Makueni	0	0	2

```
Meru                1    1     2
Mombasa             0    0    10
Nairobi             2    1    31
Nakuru              0    0    10
Narok               0    0     1
Nyandarua           0    0     2
Nyeri               0    0     2
Samburu             0    1     2
South Sudan         1    2     6
Taita Taveta        0    0     1
Tana River          0    1     0
Tharaka Nithi       0    0     1
Turkana             0    0     3
Uasin Gishu         0    1     1
West Pokot          0    1     3
```

```
table(mdr_tb_df$HIV_STATUS, mdr_tb_df$MDR_TYPE)
```

```
          MDR+ MDRIR MDRKR
NEGATIVE     2     3    66
NOT_DONE     3     5    37
POSITIVE     2     3    22
```

```
table(mdr_tb_df$SMEAR, mdr_tb_df$MDR_TYPE)
```

```
          MDR+ MDRIR MDRKR
1+           2     5    40
2+           2     2    21
3+           1     2    44
NEG          1     0     1
SCANTY       1     2    19
```

```
table(mdr_tb_df$TB_TYPE, mdr_tb_df$MDR_TYPE)
```

```
          MDR+ MDRIR MDRKR
FA_1STL_TR   1     3    31
FA_RE_TREA   0     0     3
MDR_CONTA    0     0     3
MDR_F        1     1    22
NEW          1     4    26
RE_AF_DEFA   0     1    13
SP_SM_NR     0     0     3
```

```
SP_SM_PR       4      2     23
UNKNOWN        0      0      1
```