ADS503_Team3

Jonathan Yoon, Luke Awino, Michael Nguyen

1. Importing data as data frames

There are 9 different worksheets in the original data. Each worksheet is imported as separate data frames.

```
library(readxl)

df1 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 1)

df2 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 2)

df3 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 3)

df4 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 4)

df5 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 5)

df6 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 6)

df7 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 7)

df8 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 8)

df9 <- read_excel('/Users/yhjnthn/Documents/USD_MS-ADS/ADS503/Database-on-education-for-children-with-disabilities.xlsx', 9)

head(df1)
```

```
## # A tibble: 6 × 18
##   `Countries and are…` `ISO Code` Region `Sub-region` `Development r…` Indicator
##   <chr>                <chr>      <chr>  <chr>        <chr>            <chr>
## 1 <NA>                 <NA>       <NA>   <NA>         <NA>             <NA>
## 2 Bangladesh           BGD        SA     SA           Least Developed  ANAR Pri…
## 3 Bangladesh           BGD        SA     SA           Least Developed  ANAR Pri…
## 4 Bangladesh           BGD        SA     SA           Least Developed  ANAR Pri…
## 5 Bangladesh           BGD        SA     SA           Least Developed  ANAR Pri…
## 6 Bangladesh           BGD        SA     SA           Least Developed  ANAR Pri…
## # … with 12 more variables: Category <chr>, Total <chr>, ...9 <chr>,
## #   ...10 <chr>, `Children without functional difficulties` <chr>, ...12 <chr>,
## #   ...13 <chr>, `Children with functional difficulties` <chr>, ...15 <chr>,
## #   ...16 <chr>, `Data source` <chr>, `Time period` <chr>
```

Column names for upper & lower limits for Total, Children without functional disabilities, and Children with functional disabilities do not have the correct names. The names of these columns are assigned by following:

```
colnames(df1)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df1)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_without_functional_difficulties_lower_limit")
```

```
colnames(df1)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df2)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df2)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df2)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df3)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df3)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df3)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df4)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df4)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df4)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df5)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df5)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df5)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df6)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df6)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df6)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df7)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df7)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df7)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df8)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df8)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df8)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
colnames(df9)[9:10] <- c("Total_Upper_Limit", "Total_Lower_Limit")
colnames(df9)[12:13] <- c("Children_without_functional_difficulties_upper_limit", "Children_withou
t_functional_difficulties_lower_limit")
colnames(df9)[15:16] <- c("Children_with_functional_difficulties_upper_limit", "Children_with_func
tional_difficulties_lower_limit")
```

Removal of the unnecessary subcategories (row 1)

```
df1_clean <- as.data.frame(df1[-1,])
df2_clean <- as.data.frame(df2[-1,])
df3_clean <- as.data.frame(df3[-1,])
df4_clean <- as.data.frame(df4[-1,])
df5_clean <- as.data.frame(df5[-1,])
df6_clean <- as.data.frame(df6[-1,])
df7_clean <- as.data.frame(df7[-1,])
df8_clean <- as.data.frame(df8[-1,])
df9_clean <- as.data.frame(df9[-1,])
```

After removing the unneeded subcategory row, the columns are combined into one for convenience.

```
df_comb <- cbind(df1_clean[,1:16], df2_clean[,6:16], df3_clean[,6:16],
                 df4_clean[,6:16], df5_clean[,6:16], df6_clean[,6:16],
                 df7_clean[,6:16], df8_clean[,6:16], df9_clean[,6:18])
dim(df_comb)
```

```
## [1] 160 106
```

The dataset are now combined as a single dataframe df_comb that has 160 rows and 106 columns.

```
dfc <- df_comb[,c(5,8,11,14,19,22,25,30,33,36,41,44,47,52,55,58,63,66,69,74,77,80,85,88,91,96,99,1
02)] # Excluding explanatory and limit values
total <- dfc[c(1,6,11,16,21,26,31,36,41,46,51,56,61,66,71,76,81,86,91,96,
               101,106,111,116,121,126,131,136,141,146,151,156),] # subsetting total values
male <- dfc[c(2,7,12,17,22,27,32,37,42,47,52,57,62,67,72,77,82,87,92,97,
              102,107,112,117,122,127,132,137,142,147,152,157),] # subsetting male values
female <- dfc[c(3,8,13,18,23,28,33,38,43,48,53,58,63,68,73,78,83,88,93,98,
                103,108,113,118,123,128,133,138,143,148,153,158),] # subsetting female values
urban <- dfc[c(4,9,14,19,24,29,34,39,44,49,54,59,64,69,74,79,84,89,94,99,
               104,109,114,119,124,129,134,139,144,149,154,159),] # subsetting urban values
rural <- dfc[c(5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,95,
               100,105,110,115,120,125,130,135,140,145,150,155,160),] # subsetting rural values
```

The dataset has not been separated into 5 different subsets.

```
colnames(total) <- c("level","primaryANAR_total","primaryANAR_nodiff_total","primaryANAR_diff_total",
                     "lowsecondaryANAR_total","lowsecondaryANAR_nodiff_total","lowsecondaryANAR_diff_total",
                     "uppsecondaryANAR_total","uppsecondaryANAR_nodiff_total","uppsecondaryANAR_diff_total",
                     "primaryOOS_total","primaryOOS_nodiff_total","primaryOOS_diff_total",
                     "lowsecondaryOOS_total","lowsecondaryOOS_nodiff_total","lowsecondaryOOS_diff_total",
                     "uppsecondaryOOS_total","uppsecondaryOOS_nodiff_total","uppsecondaryOOS_diff_total",
                     "primarycomp_total","primarycomp_nodiff_total","primarycomp_diff_total",
                     "reading_total","reading_nodiff_total","reading_diff_total",
                     "numeric_total","numeric_nodiff_total","numeric_diff_total")
colnames(male) <- c("level","primaryANAR_male","primaryANAR_nodiff_male","primaryANAR_diff_male",
                    "lowsecondaryANAR_male","lowsecondaryANAR_nodiff_male","lowsecondaryANAR_diff_male",
                    "uppsecondaryANAR_male","uppsecondaryANAR_nodiff_male","uppsecondaryANAR_diff_male",
                    "primaryOOS_male","primaryOOS_nodiff_male","primaryOOS_diff_male",
                    "lowsecondaryOOS_male","lowsecondaryOOS_nodiff_male","lowsecondaryOOS_diff_male",
                    "uppsecondaryOOS_male","uppsecondaryOOS_nodiff_male","uppsecondaryOOS_diff_male",
                    "primarycomp_male","primarycomp_nodiff_male","primarycomp_diff_male",
                    "reading_male","reading_nodiff_male","reading_diff_male",
                    "numeric_male","numeric_nodiff_male","numeric_diff_male")
colnames(female) <- c("level","primaryANAR_female","primaryANAR_nodiff_female","primaryANAR_diff_female",
                      "lowsecondaryANAR_female","lowsecondaryANAR_nodiff_female","lowsecondaryANAR_diff_female",
                      "uppsecondaryANAR_female","uppsecondaryANAR_nodiff_female","uppsecondaryANAR_diff_female",
                      "primaryOOS_female","primaryOOS_nodiff_female","primaryOOS_diff_female",
                      "lowsecondaryOOS_female","lowsecondaryOOS_nodiff_female","lowsecondaryOOS_diff_female",
                      "uppsecondaryOOS_female","uppsecondaryOOS_nodiff_female","uppsecondaryOOS_diff_female",
                      "primarycomp_female","primarycomp_nodiff_female","primarycomp_diff_female",
                      "reading_female","reading_nodiff_female","reading_diff_female",
                      "numeric_female","numeric_nodiff_female","numeric_diff_female")
colnames(urban) <- c("level","primaryANAR_urban","primaryANAR_nodiff_urban","primaryANAR_diff_urban",
                     "lowsecondaryANAR_urban","lowsecondaryANAR_nodiff_urban","lowsecondaryANAR_diff_urban",
                     "uppsecondaryANAR_urban","uppsecondaryANAR_nodiff_urban","uppsecondaryANAR_diff_urban",
                     "primaryOOS_urban","primaryOOS_nodiff_urban","primaryOOS_diff_urban",
                     "lowsecondaryOOS_urban","lowsecondaryOOS_nodiff_urban","lowsecondaryOOS_diff_urban",
                     "uppsecondaryOOS_urban","uppsecondaryOOS_nodiff_urban","uppsecondaryOOS_diff_urban",
                     "primarycomp_urban","primarycomp_nodiff_urban","primarycomp_diff_urban",
                     "reading_urban","reading_nodiff_urban","reading_diff_urban",
                     "numeric_urban","numeric_nodiff_urban","numeric_diff_urban")
colnames(rural) <- c("level","primaryANAR_rural","primaryANAR_nodiff_rural","primaryANAR_diff_rural",
                     "lowsecondaryANAR_rural","lowsecondaryANAR_nodiff_rural","lowsecondaryANAR_diff_rural",
                     "uppsecondaryANAR_rural","uppsecondaryANAR_nodiff_rural","uppsecondaryANAR_diff_rural",
                     "primaryOOS_rural","primaryOOS_nodiff_rural","primaryOOS_diff_rural",
                     "lowsecondaryOOS_rural","lowsecondaryOOS_nodiff_rural","lowsecondaryOOS_diff_rural",
                     "uppsecondaryOOS_rural","uppsecondaryOOS_nodiff_rural","uppsecondaryOOS_diff_rural",
                     "primarycomp_rural","primarycomp_nodiff_rural","primarycomp_diff_rural",
                     "reading_rural","reading_nodiff_rural","reading_diff_rural",
                     "numeric_rural","numeric_nodiff_rural","numeric_diff_rural")
```

```
df_c <- cbind(total,male[,2:28],female[,2:28],urban[,2:28],rural[,2:28])

dim(df_c)

## [1]  32 136
```

Now, the columns have been re-named, and they have been combined so each column represents different categories (Total, Male, Female, Urban, Rural).

```
df_c[,1]

##  [1] "Least Developed" "More Developed"  "Least Developed" "Least Developed"
##  [5] "Least Developed" "Least Developed" "Less Developed"  "Least Developed"
##  [9] "Less Developed"  "Less Developed"  "Least Developed" NA
## [13] "Less Developed"  "Least Developed" "Least Developed" "Less Developed"
## [17] "Least Developed" "More Developed"  "Less Developed"  "Less Developed"
## [21] "Less Developed"  "Least Developed" "Least Developed" "Less Developed"
## [25] "Less Developed"  "Not Classified"  "Least Developed" "Less Developed"
## [29] "Less Developed"  "Less Developed"  "Least Developed" "Less Developed"

# Among the 32 values of the Development Levels, one NA value and on Not Specified (26th row) valu
e were noted.
library(zoo)

df_c[,1] <- na.fill(df_c[,1], "Less Developed")
df_c[26,1] <- "More Developed"
summary(df_c)

##     level           primaryANAR_total  primaryANAR_nodiff_total
##  Length:32          Length:32          Length:32
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  primaryANAR_diff_total lowsecondaryANAR_total lowsecondaryANAR_nodiff_total
##  Length:32              Length:32              Length:32
##  Class :character       Class :character       Class :character
##  Mode  :character       Mode  :character       Mode  :character
##  lowsecondaryANAR_diff_total uppsecondaryANAR_total
##  Length:32                   Length:32
##  Class :character            Class :character
##  Mode  :character            Mode  :character
##  uppsecondaryANAR_nodiff_total uppsecondaryANAR_diff_total primaryOOS_total
##  Length:32                     Length:32                   Length:32
##  Class :character              Class :character            Class :character
##  Mode  :character              Mode  :character            Mode  :character
##  primaryOOS_nodiff_total primaryOOS_diff_total lowsecondaryOOS_total
##  Length:32               Length:32             Length:32
##  Class :character        Class :character      Class :character
##  Mode  :character        Mode  :character      Mode  :character
##  lowsecondaryOOS_nodiff_total lowsecondaryOOS_diff_total uppsecondaryOOS_total
##  Length:32                    Length:32                  Length:32
##  Class :character             Class :character           Class :character
##  Mode  :character             Mode  :character           Mode  :character
##  uppsecondaryOOS_nodiff_total uppsecondaryOOS_diff_total primarycomp_total
##  Length:32                    Length:32                  Length:32
##  Class :character             Class :character           Class :character
##  Mode  :character             Mode  :character           Mode  :character
##  primarycomp_nodiff_total primarycomp_diff_total reading_total
##  Length:32                Length:32              Length:32
##  Class :character         Class :character       Class :character
##  Mode  :character         Mode  :character       Mode  :character
```

```
##  reading_nodiff_total reading_diff_total numeric_total
##  Length:32            Length:32          Length:32
##  Class :character     Class :character   Class :character
##  Mode  :character     Mode  :character   Mode  :character
##  numeric_nodiff_total numeric_diff_total primaryANAR_male
##  Length:32            Length:32          Length:32
##  Class :character     Class :character   Class :character
##  Mode  :character     Mode  :character   Mode  :character
##  primaryANAR_nodiff_male primaryANAR_diff_male lowsecondaryANAR_male
##  Length:32               Length:32             Length:32
##  Class :character        Class :character      Class :character
##  Mode  :character        Mode  :character      Mode  :character
##  lowsecondaryANAR_nodiff_male lowsecondaryANAR_diff_male uppsecondaryANAR_male
##  Length:32                    Length:32                  Length:32
##  Class :character             Class :character           Class :character
##  Mode  :character             Mode  :character           Mode  :character
##  uppsecondaryANAR_nodiff_male uppsecondaryANAR_diff_male primaryOOS_male
##  Length:32                    Length:32                  Length:32
##  Class :character             Class :character           Class :character
##  Mode  :character             Mode  :character           Mode  :character
##  primaryOOS_nodiff_male primaryOOS_diff_male lowsecondaryOOS_male
##  Length:32              Length:32            Length:32
##  Class :character       Class :character     Class :character
##  Mode  :character       Mode  :character     Mode  :character
##  lowsecondaryOOS_nodiff_male lowsecondaryOOS_diff_male uppsecondaryOOS_male
##  Length:32                   Length:32                 Length:32
##  Class :character            Class :character          Class :character
##  Mode  :character            Mode  :character          Mode  :character
##  uppsecondaryOOS_nodiff_male uppsecondaryOOS_diff_male primarycomp_male
##  Length:32                   Length:32                 Length:32
##  Class :character            Class :character          Class :character
##  Mode  :character            Mode  :character          Mode  :character
##  primarycomp_nodiff_male primarycomp_diff_male reading_male
##  Length:32               Length:32             Length:32
##  Class :character        Class :character      Class :character
##  Mode  :character        Mode  :character      Mode  :character
##  reading_nodiff_male reading_diff_male  numeric_male       numeric_nodiff_male
##  Length:32           Length:32          Length:32          Length:32
##  Class :character    Class :character   Class :character   Class :character
##  Mode  :character    Mode  :character   Mode  :character   Mode  :character
##  numeric_diff_male  primaryANAR_female primaryANAR_nodiff_female
##  Length:32          Length:32          Length:32
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  primaryANAR_diff_female lowsecondaryANAR_female lowsecondaryANAR_nodiff_female
##  Length:32               Length:32               Length:32
##  Class :character        Class :character        Class :character
##  Mode  :character        Mode  :character        Mode  :character
##  lowsecondaryANAR_diff_female uppsecondaryANAR_female
##  Length:32                    Length:32
##  Class :character             Class :character
##  Mode  :character             Mode  :character
##  uppsecondaryANAR_nodiff_female uppsecondaryANAR_diff_female primaryOOS_female
##  Length:32                        Length:32                   Length:32
##  Class :character                 Class :character            Class :character
##  Mode  :character                 Mode  :character            Mode  :character
##  primaryOOS_nodiff_female primaryOOS_diff_female lowsecondaryOOS_female
##  Length:32                Length:32              Length:32
##  Class :character         Class :character       Class :character
##  Mode  :character         Mode  :character       Mode  :character
```

```
##  lowsecondaryOOS_nodiff_female lowsecondaryOOS_diff_female
##  Length:32                     Length:32
##  Class :character              Class :character
##  Mode  :character              Mode  :character
##  uppsecondaryOOS_female uppsecondaryOOS_nodiff_female
##  Length:32              Length:32
##  Class :character       Class :character
##  Mode  :character       Mode  :character
##  uppsecondaryOOS_diff_female primarycomp_female primarycomp_nodiff_female
##  Length:32                   Length:32          Length:32
##  Class :character            Class :character   Class :character
##  Mode  :character            Mode  :character   Mode  :character
##  primarycomp_diff_female reading_female     reading_nodiff_female
##  Length:32               Length:32          Length:32
##  Class :character        Class :character   Class :character
##  Mode  :character        Mode  :character   Mode  :character
##  reading_diff_female numeric_female     numeric_nodiff_female
##  Length:32           Length:32          Length:32
##  Class :character    Class :character   Class :character
##  Mode  :character    Mode  :character   Mode  :character
##  numeric_diff_female primaryANAR_urban  primaryANAR_nodiff_urban
##  Length:32           Length:32          Length:32
##  Class :character    Class :character   Class :character
##  Mode  :character    Mode  :character   Mode  :character
##  primaryANAR_diff_urban lowsecondaryANAR_urban lowsecondaryANAR_nodiff_urban
##  Length:32              Length:32              Length:32
##  Class :character       Class :character       Class :character
##  Mode  :character       Mode  :character       Mode  :character
##  lowsecondaryANAR_diff_urban uppsecondaryANAR_urban
##  Length:32                   Length:32
##  Class :character            Class :character
##  Mode  :character            Mode  :character
##  uppsecondaryANAR_nodiff_urban uppsecondaryANAR_diff_urban primaryOOS_urban
##  Length:32                     Length:32                  Length:32
##  Class :character              Class :character           Class :character
##  Mode  :character              Mode  :character           Mode  :character
##  primaryOOS_nodiff_urban primaryOOS_diff_urban lowsecondaryOOS_urban
##  Length:32               Length:32             Length:32
##  Class :character        Class :character      Class :character
##  Mode  :character        Mode  :character      Mode  :character
##  lowsecondaryOOS_nodiff_urban lowsecondaryOOS_diff_urban uppsecondaryOOS_urban
##  Length:32                    Length:32                  Length:32
##  Class :character             Class :character           Class :character
##  Mode  :character             Mode  :character           Mode  :character
##  uppsecondaryOOS_nodiff_urban uppsecondaryOOS_diff_urban primarycomp_urban
##  Length:32                    Length:32                  Length:32
##  Class :character             Class :character           Class :character
##  Mode  :character             Mode  :character           Mode  :character
##  primarycomp_nodiff_urban primarycomp_diff_urban reading_urban
##  Length:32                Length:32              Length:32
##  Class :character         Class :character       Class :character
##  Mode  :character         Mode  :character       Mode  :character
##  reading_nodiff_urban reading_diff_urban numeric_urban
##  Length:32            Length:32          Length:32
##  Class :character     Class :character   Class :character
##  Mode  :character     Mode  :character   Mode  :character
##  numeric_nodiff_urban numeric_diff_urban primaryANAR_rural
##  Length:32            Length:32          Length:32
##  Class :character     Class :character   Class :character
##  Mode  :character     Mode  :character   Mode  :character
```

```
##  primaryANAR_nodiff_rural primaryANAR_diff_rural lowsecondaryANAR_rural
##  Length:32                Length:32              Length:32
##  Class :character         Class :character       Class :character
##  Mode  :character         Mode  :character       Mode  :character
##  lowsecondaryANAR_nodiff_rural lowsecondaryANAR_diff_rural
##  Length:32                     Length:32
##  Class :character              Class :character
##  Mode  :character              Mode  :character
##  uppsecondaryANAR_rural uppsecondaryANAR_nodiff_rural
##  Length:32              Length:32
##  Class :character       Class :character
##  Mode  :character       Mode  :character
##  uppsecondaryANAR_diff_rural primaryOOS_rural   primaryOOS_nodiff_rural
##  Length:32                   Length:32          Length:32
##  Class :character            Class :character   Class :character
##  Mode  :character            Mode  :character   Mode  :character
##  primaryOOS_diff_rural lowsecondaryOOS_rural lowsecondaryOOS_nodiff_rural
##  Length:32             Length:32             Length:32
##  Class :character      Class :character      Class :character
##  Mode  :character      Mode  :character      Mode  :character
##  lowsecondaryOOS_diff_rural uppsecondaryOOS_rural uppsecondaryOOS_nodiff_rural
##  Length:32                  Length:32             Length:32
##  Class :character           Class :character      Class :character
##  Mode  :character           Mode  :character      Mode  :character
##  uppsecondaryOOS_diff_rural primarycomp_rural  primarycomp_nodiff_rural
##  Length:32                  Length:32          Length:32
##  Class :character           Class :character   Class :character
##  Mode  :character           Mode  :character   Mode  :character
##  primarycomp_diff_rural reading_rural      reading_nodiff_rural
##  Length:32              Length:32          Length:32
##  Class :character       Class :character   Class :character
##  Mode  :character       Mode  :character   Mode  :character
##  reading_diff_rural numeric_rural      numeric_nodiff_rural numeric_diff_rural
##  Length:32          Length:32          Length:32            Length:32
##  Class :character   Class :character   Class :character     Class :character
##  Mode  :character   Mode  :character   Mode  :character     Mode  :character
```

The levels have now been filled in, but the values are all in characters, not numeric.

```
df_c[,2:136] <- lapply(df_c[,2:136], FUN = function(y){as.numeric(y)})
summary(df_c)
```

```
##     level           primaryANAR_total primaryANAR_nodiff_total
##  Length:32          Min.   :36.20     Min.   :36.50
##  Class :character   1st Qu.:77.05     1st Qu.:77.50
##  Mode  :character   Median :89.95     Median :89.90
##                     Mean   :84.12     Mean   :84.55
##                     3rd Qu.:96.30     3rd Qu.:96.40
##                     Max.   :98.70     Max.   :99.30
##
##  primaryANAR_diff_total lowsecondaryANAR_total lowsecondaryANAR_nodiff_total
##  Min.   :34.90          Min.   : 1.10          Min.   : 1.00
##  1st Qu.:72.40          1st Qu.:34.98          1st Qu.:36.05
##  Median :84.20          Median :56.30          Median :59.15
##  Mean   :80.62          Mean   :57.80          Mean   :58.60
##  3rd Qu.:93.80          3rd Qu.:91.25          3rd Qu.:91.35
##  Max.   :97.30          Max.   :99.20          Max.   :99.40
##  NA's   :3
##  lowsecondaryANAR_diff_total uppsecondaryANAR_total
##  Min.   : 3.80               Min.   : 0.10
```

```
##  1st Qu.:23.65              1st Qu.:18.88
##  Median :41.35              Median :41.95
##  Mean   :45.51              Mean   :45.95
##  3rd Qu.:68.42              3rd Qu.:76.78
##  Max.   :95.40              Max.   :95.80
##  NA's   :6
##  uppsecondaryANAR_nodiff_total uppsecondaryANAR_diff_total primaryOOS_total
##  Min.   : 0.10                 Min.   : 0.00               Min.   : 0.90
##  1st Qu.:19.80                 1st Qu.: 9.85               1st Qu.: 3.75
##  Median :37.60                 Median :21.30               Median : 8.85
##  Mean   :45.71                 Mean   :26.57               Mean   :14.79
##  3rd Qu.:77.35                 3rd Qu.:40.20               3rd Qu.:21.38
##  Max.   :96.60                 Max.   :73.80               Max.   :63.60
##  NA's   :1                     NA's   :13
##  primaryOOS_nodiff_total primaryOOS_diff_total lowsecondaryOOS_total
##  Min.   : 0.800          Min.   : 1.20         Min.   : 0.000
##  1st Qu.: 3.725          1st Qu.: 6.40         1st Qu.: 4.025
##  Median : 8.300          Median :10.60         Median : 6.650
##  Mean   :14.350          Mean   :17.77         Mean   :11.759
##  3rd Qu.:20.625          3rd Qu.:25.90         3rd Qu.:16.725
##  Max.   :62.100          Max.   :67.40         Max.   :52.600
##                          NA's   :3
##  lowsecondaryOOS_nodiff_total lowsecondaryOOS_diff_total uppsecondaryOOS_total
##  Min.   : 0.000               Min.   : 1.00              Min.   : 0.00
##  1st Qu.: 3.225               1st Qu.: 6.55              1st Qu.:12.38
##  Median : 7.050               Median :13.40              Median :17.45
##  Mean   :11.266               Mean   :17.13              Mean   :20.52
##  3rd Qu.:15.625               3rd Qu.:23.20              3rd Qu.:28.12
##  Max.   :50.800               Max.   :57.50              Max.   :54.50
##                               NA's   :6
##  uppsecondaryOOS_nodiff_total uppsecondaryOOS_diff_total primarycomp_total
##  Min.   : 0.00                Min.   :13.60              Min.   : 25.20
##  1st Qu.:11.40                1st Qu.:20.05              1st Qu.: 66.03
##  Median :16.90                Median :24.90              Median : 84.30
##  Mean   :19.93                Mean   :30.36              Mean   : 79.17
##  3rd Qu.:27.65                3rd Qu.:40.20              3rd Qu.: 98.10
##  Max.   :54.10                Max.   :61.30              Max.   :100.00
##  NA's   :1                    NA's   :13
##  primarycomp_nodiff_total primarycomp_diff_total reading_total
##  Min.   : 25.90           Min.   :18.80          Min.   : 4.40
##  1st Qu.: 67.67           1st Qu.:60.23          1st Qu.:19.60
##  Median : 85.10           Median :75.35          Median :43.80
##  Mean   : 80.03           Mean   :68.82          Mean   :41.63
##  3rd Qu.: 98.40           3rd Qu.:84.05          3rd Qu.:59.70
##  Max.   :100.00           Max.   :98.60          Max.   :82.40
##                           NA's   :8              NA's   :1
##  reading_nodiff_total reading_diff_total numeric_total   numeric_nodiff_total
##  Min.   : 4.40        Min.   : 3.50      Min.   : 0.50   Min.   : 0.60
##  1st Qu.:20.90        1st Qu.:14.70      1st Qu.:10.25   1st Qu.:10.70
##  Median :44.40        Median :33.10      Median :25.00   Median :26.10
##  Mean   :42.67        Mean   :32.66      Mean   :29.15   Mean   :29.74
##  3rd Qu.:61.00        3rd Qu.:47.10      3rd Qu.:43.95   3rd Qu.:44.70
##  Max.   :82.70        Max.   :77.00      Max.   :72.50   Max.   :72.70
##  NA's   :1            NA's   :3          NA's   :1       NA's   :1
##  numeric_diff_total primaryANAR_male primaryANAR_nodiff_male
##  Min.   : 0.20      Min.   :37.90    Min.   :38.00
##  1st Qu.: 7.90      1st Qu.:79.47    1st Qu.:79.40
##  Median :21.40      Median :89.55    Median :88.90
##  Mean   :23.99      Mean   :84.08    Mean   :84.70
##  3rd Qu.:37.20      3rd Qu.:96.10    3rd Qu.:96.33
```

```
##  Max.   :66.10      Max.   :98.90    Max.   :99.90
##  NA's   :3
##  primaryANAR_diff_male lowsecondaryANAR_male lowsecondaryANAR_nodiff_male
##  Min.   :37.40         Min.   : 1.3          Min.   : 1.30
##  1st Qu.:70.70         1st Qu.:33.9          1st Qu.:32.23
##  Median :79.20         Median :55.1          Median :56.60
##  Mean   :78.61         Mean   :56.4          Mean   :55.81
##  3rd Qu.:93.00         3rd Qu.:89.1          3rd Qu.:89.10
##  Max.   :98.40         Max.   :98.5          Max.   :98.80
##  NA's   :7                                   NA's   :2
##  lowsecondaryANAR_diff_male uppsecondaryANAR_male uppsecondaryANAR_nodiff_male
##  Min.   : 0.00              Min.   : 0.00         Min.   : 0.00
##  1st Qu.:21.30              1st Qu.:18.20         1st Qu.:19.30
##  Median :33.90              Median :39.90         Median :35.75
##  Mean   :38.68              Mean   :43.75         Mean   :42.83
##  3rd Qu.:49.00              3rd Qu.:69.60         3rd Qu.:68.25
##  Max.   :94.10              Max.   :97.00         Max.   :97.00
##  NA's   :11                                       NA's   :2
##  uppsecondaryANAR_diff_male primaryOOS_male primaryOOS_nodiff_male
##  Min.   : 3.40              Min.   : 1.20   Min.   : 1.000
##  1st Qu.:16.80              1st Qu.: 4.05   1st Qu.: 3.575
##  Median :22.30              Median : 8.50   Median : 7.950
##  Mean   :28.52              Mean   :14.52   Mean   :13.906
##  3rd Qu.:37.00              3rd Qu.:20.30   3rd Qu.:17.950
##  Max.   :68.90              Max.   :60.30   Max.   :58.900
##  NA's   :19
##  primaryOOS_diff_male lowsecondaryOOS_male lowsecondaryOOS_nodiff_male
##  Min.   : 1.20        Min.   : 0.00        Min.   : 0.000
##  1st Qu.: 7.80        1st Qu.: 3.10        1st Qu.: 3.025
##  Median :14.30        Median : 8.20        Median : 7.800
##  Mean   :19.98        Mean   :11.36        Mean   :11.123
##  3rd Qu.:31.60        3rd Qu.:14.85        3rd Qu.:14.225
##  Max.   :63.60        Max.   :48.60        Max.   :47.200
##  NA's   :7                                 NA's   :2
##  lowsecondaryOOS_diff_male uppsecondaryOOS_male uppsecondaryOOS_nodiff_male
##  Min.   : 3.70             Min.   : 0.00        Min.   : 0.00
##  1st Qu.: 9.20             1st Qu.:13.30        1st Qu.:12.72
##  Median :18.30             Median :20.90        Median :20.55
##  Mean   :18.17             Mean   :21.24        Mean   :20.68
##  3rd Qu.:22.80             3rd Qu.:29.27        3rd Qu.:28.02
##  Max.   :52.30             Max.   :54.80        Max.   :54.30
##  NA's   :11                                     NA's   :2
##  uppsecondaryOOS_diff_male primarycomp_male primarycomp_nodiff_male
##  Min.   :14.70             Min.   : 24.20   Min.   : 24.80
##  1st Qu.:21.90             1st Qu.: 66.85   1st Qu.: 68.35
##  Median :34.40             Median : 81.50   Median : 82.35
##  Mean   :32.48             Mean   : 78.09   Mean   : 78.05
##  3rd Qu.:38.10             3rd Qu.: 97.25   3rd Qu.: 96.83
##  Max.   :59.50             Max.   :100.00   Max.   :100.00
##  NA's   :19                                 NA's   :2
##  primarycomp_diff_male  reading_male   reading_nodiff_male reading_diff_male
##  Min.   :17.30         Min.   : 4.80   Min.   : 4.60       Min.   : 3.60
##  1st Qu.:47.40         1st Qu.:19.75   1st Qu.:20.25       1st Qu.:14.55
##  Median :63.30         Median :40.50   Median :42.10       Median :27.65
##  Mean   :60.63         Mean   :39.54   Mean   :40.44       Mean   :28.01
##  3rd Qu.:77.75         3rd Qu.:56.25   3rd Qu.:58.10       3rd Qu.:34.58
##  Max.   :99.30         Max.   :81.60   Max.   :82.00       Max.   :67.10
##  NA's   :14            NA's   :1       NA's   :1           NA's   :6
##   numeric_male   numeric_nodiff_male numeric_diff_male primaryANAR_female
##  Min.   : 0.50   Min.   : 0.60       Min.   : 0.00     Min.   :34.40
```

```
##  1st Qu.:10.05   1st Qu.:10.45       1st Qu.: 7.80     1st Qu.:75.55
##  Median :24.20   Median :25.40       Median :15.80     Median :90.40
##  Mean   :28.65   Mean   :29.31       Mean   :19.66     Mean   :84.13
##  3rd Qu.:45.65   3rd Qu.:46.00       3rd Qu.:25.85     3rd Qu.:96.00
##  Max.   :72.20   Max.   :72.50       Max.   :65.20     Max.   :98.80
##  NA's   :1       NA's   :1           NA's   :6
##  primaryANAR_nodiff_female primaryANAR_diff_female lowsecondaryANAR_female
##  Min.   :34.90             Min.   :32.10           Min.   :  1.00
##  1st Qu.:76.08             1st Qu.:71.80           1st Qu.: 35.88
##  Median :90.65             Median :82.90           Median : 58.85
##  Mean   :84.41             Mean   :78.69           Mean   : 59.23
##  3rd Qu.:96.30             3rd Qu.:92.38           3rd Qu.: 90.92
##  Max.   :98.80             Max.   :97.40           Max.   :100.00
##                            NA's   :8
##  lowsecondaryANAR_nodiff_female lowsecondaryANAR_diff_female
##  Min.   :  0.60                 Min.   : 6.40
##  1st Qu.: 32.92                 1st Qu.:23.57
##  Median : 57.85                 Median :43.10
##  Mean   : 57.93                 Mean   :45.68
##  3rd Qu.: 91.85                 3rd Qu.:61.70
##  Max.   :100.00                 Max.   :91.60
##  NA's   :2                      NA's   :12
##  uppsecondaryANAR_female uppsecondaryANAR_nodiff_female
##  Min.   : 0.20           Min.   : 0.20
##  1st Qu.:17.12           1st Qu.:18.65
##  Median :44.40           Median :37.15
##  Mean   :48.08           Mean   :46.55
##  3rd Qu.:80.20           3rd Qu.:78.88
##  Max.   :94.60           Max.   :96.10
##                          NA's   :2
##  uppsecondaryANAR_diff_female primaryOOS_female primaryOOS_nodiff_female
##  Min.   : 0.00                Min.   : 0.400    Min.   : 0.400
##  1st Qu.:10.30                1st Qu.: 3.475    1st Qu.: 3.675
##  Median :26.80                Median :10.400    Median : 9.950
##  Mean   :29.75                Mean   :15.106    Mean   :14.834
##  3rd Qu.:42.92                3rd Qu.:21.050    3rd Qu.:21.425
##  Max.   :78.40                Max.   :67.000    Max.   :65.300
##  NA's   :16
##  primaryOOS_diff_female lowsecondaryOOS_female lowsecondaryOOS_nodiff_female
##  Min.   : 0.000         Min.   : 0.000         Min.   : 0.00
##  1st Qu.: 5.175         1st Qu.: 3.175         1st Qu.: 2.75
##  Median :15.250         Median : 5.400         Median : 4.90
##  Mean   :19.554         Mean   :12.322         Mean   :11.91
##  3rd Qu.:27.800         3rd Qu.:16.525         3rd Qu.:16.68
##  Max.   :71.100         Max.   :57.000         Max.   :54.70
##  NA's   :8                                     NA's   :2
##  lowsecondaryOOS_diff_female uppsecondaryOOS_female
##  Min.   : 0.80               Min.   : 0.00
##  1st Qu.: 6.85               1st Qu.: 8.15
##  Median :16.10               Median :14.50
##  Mean   :19.68               Mean   :20.21
##  3rd Qu.:26.25               3rd Qu.:28.25
##  Max.   :63.50               Max.   :64.80
##  NA's   :12
##  uppsecondaryOOS_nodiff_female uppsecondaryOOS_diff_female primarycomp_female
##  Min.   : 0.00                 Min.   : 9.70               Min.   : 24.30
##  1st Qu.:10.07                 1st Qu.:18.55               1st Qu.: 68.33
##  Median :16.70                 Median :27.10               Median : 89.95
##  Mean   :20.33                 Mean   :34.02               Mean   : 80.19
##  3rd Qu.:29.35                 3rd Qu.:50.88               3rd Qu.: 99.33
```

```
##  Max.   :62.60                Max.   :73.40              Max.   :100.00
##  NA's   :2                    NA's   :16
##  primarycomp_nodiff_female primarycomp_diff_female reading_female
##  Min.   : 25.20            Min.   : 20.40          Min.   : 3.9
##  1st Qu.: 67.08            1st Qu.: 55.55          1st Qu.:20.3
##  Median : 88.10            Median : 74.20          Median :47.1
##  Mean   : 79.64            Mean   : 66.56          Mean   :43.8
##  3rd Qu.: 99.12            3rd Qu.: 83.40          3rd Qu.:61.9
##  Max.   :100.00            Max.   :100.00          Max.   :85.0
##  NA's   :2                 NA's   :13              NA's   :1
##  reading_nodiff_female reading_diff_female numeric_female
##  Min.   : 4.30         Min.   : 2.30       Min.   : 0.50
##  1st Qu.:21.70         1st Qu.:13.10       1st Qu.:10.45
##  Median :49.00         Median :33.50       Median :26.80
##  Mean   :44.89         Mean   :29.56       Mean   :29.69
##  3rd Qu.:62.10         3rd Qu.:41.30       3rd Qu.:45.70
##  Max.   :84.70         Max.   :58.50       Max.   :72.80
##  NA's   :1             NA's   :7           NA's   :1
##  numeric_nodiff_female numeric_diff_female primaryANAR_urban
##  Min.   : 0.50         Min.   : 0.20       Min.   :48.30
##  1st Qu.:10.90         1st Qu.: 7.20       1st Qu.:84.28
##  Median :27.50         Median :18.30       Median :91.65
##  Mean   :30.16         Mean   :19.77       Mean   :88.24
##  3rd Qu.:47.55         3rd Qu.:29.60       3rd Qu.:96.80
##  Max.   :72.90         Max.   :55.70       Max.   :98.50
##  NA's   :1             NA's   :7
##  primaryANAR_nodiff_urban primaryANAR_diff_urban lowsecondaryANAR_urban
##  Min.   :48.80            Min.   :46.20          Min.   : 1.70
##  1st Qu.:84.72            1st Qu.:76.15          1st Qu.:45.42
##  Median :92.60            Median :85.90          Median :64.70
##  Mean   :88.73            Mean   :83.07          Mean   :63.57
##  3rd Qu.:97.05            3rd Qu.:92.60          3rd Qu.:91.90
##  Max.   :99.30            Max.   :97.50          Max.   :99.60
##                           NA's   :9
##  lowsecondaryANAR_nodiff_urban lowsecondaryANAR_diff_urban
##  Min.   :  1.50                Min.   : 7.00
##  1st Qu.: 45.90                1st Qu.:32.08
##  Median : 64.10                Median :42.70
##  Mean   : 63.97                Mean   :44.19
##  3rd Qu.: 92.40                3rd Qu.:53.35
##  Max.   :100.00                Max.   :84.60
##  NA's   :1                     NA's   :14
##  uppsecondaryANAR_urban uppsecondaryANAR_nodiff_urban
##  Min.   : 0.20          Min.   : 0.2
##  1st Qu.:26.45          1st Qu.:26.7
##  Median :52.00          Median :42.5
##  Mean   :51.90          Mean   :49.6
##  3rd Qu.:79.22          3rd Qu.:81.4
##  Max.   :97.00          Max.   :97.1
##                         NA's   :3
##  uppsecondaryANAR_diff_urban primaryOOS_urban primaryOOS_nodiff_urban
##  Min.   : 9.10               Min.   : 0.800   Min.   : 1.000
##  1st Qu.:20.30               1st Qu.: 2.750   1st Qu.: 2.575
##  Median :32.90               Median : 7.550   Median : 6.650
##  Mean   :33.13               Mean   : 9.753   Mean   : 9.322
##  3rd Qu.:43.40               3rd Qu.:14.225   3rd Qu.:14.375
##  Max.   :62.70               Max.   :40.600   Max.   :38.200
##  NA's   :19
##  primaryOOS_diff_urban lowsecondaryOOS_urban lowsecondaryOOS_nodiff_urban
##  Min.   : 0.40         Min.   : 0.00         Min.   : 0.000
```

```
##  1st Qu.: 4.60         1st Qu.: 3.20         1st Qu.: 2.250
##  Median :11.50         Median : 5.65         Median : 5.400
##  Mean   :13.66         Mean   : 8.10         Mean   : 7.539
##  3rd Qu.:19.95         3rd Qu.:10.03         3rd Qu.: 8.900
##  Max.   :46.70         Max.   :34.50         Max.   :31.800
##  NA's   :9                                   NA's   :1
##  lowsecondaryOOS_diff_urban uppsecondaryOOS_urban uppsecondaryOOS_nodiff_urban
##  Min.   : 3.70              Min.   : 0.000        Min.   : 0.00
##  1st Qu.: 8.15              1st Qu.: 8.775        1st Qu.:10.40
##  Median :11.05              Median :12.950        Median :12.90
##  Mean   :14.70              Mean   :14.953        Mean   :15.13
##  3rd Qu.:19.73              3rd Qu.:20.675        3rd Qu.:19.80
##  Max.   :42.30              Max.   :37.600        Max.   :36.00
##  NA's   :14                                       NA's   :3
##  uppsecondaryOOS_diff_urban primarycomp_urban primarycomp_nodiff_urban
##  Min.   :12.5               Min.   : 46.90    Min.   : 48.7
##  1st Qu.:15.2               1st Qu.: 79.25    1st Qu.: 80.0
##  Median :23.9               Median : 87.20    Median : 87.8
##  Mean   :27.5               Mean   : 85.64    Mean   : 86.0
##  3rd Qu.:35.0               3rd Qu.: 98.45    3rd Qu.: 98.8
##  Max.   :50.4               Max.   :100.00    Max.   :100.0
##  NA's   :19                                   NA's   :1
##  primarycomp_diff_urban reading_urban   reading_nodiff_urban reading_diff_urban
##  Min.   :45.70          Min.   : 9.20   Min.   : 9.80        Min.   : 7.80
##  1st Qu.:65.72          1st Qu.:32.05   1st Qu.:33.40        1st Qu.:20.98
##  Median :74.45          Median :51.00   Median :52.70        Median :35.45
##  Mean   :73.92          Mean   :47.56   Mean   :48.86        Mean   :34.80
##  3rd Qu.:81.88          3rd Qu.:64.70   3rd Qu.:66.55        3rd Qu.:42.38
##  Max.   :99.30          Max.   :83.80   Max.   :84.10        Max.   :77.30
##  NA's   :16             NA's   :1       NA's   :1            NA's   :6
##  numeric_urban  numeric_nodiff_urban numeric_diff_urban primaryANAR_rural
##  Min.   : 0.9   Min.   : 1.10        Min.   : 0.20      Min.   :25.20
##  1st Qu.:16.0   1st Qu.:16.15        1st Qu.:11.45      1st Qu.:71.55
##  Median :32.0   Median :33.60        Median :20.70      Median :89.35
##  Mean   :32.8   Mean   :33.38        Mean   :23.87      Mean   :81.55
##  3rd Qu.:44.7   3rd Qu.:45.40        3rd Qu.:37.45      3rd Qu.:96.53
##  Max.   :74.6   Max.   :74.90        Max.   :68.80      Max.   :99.30
##  NA's   :1      NA's   :1            NA's   :6
##  primaryANAR_nodiff_rural primaryANAR_diff_rural lowsecondaryANAR_rural
##  Min.   :25.90            Min.   :21.20          Min.   :  0.10
##  1st Qu.:71.35            1st Qu.:64.62          1st Qu.: 20.73
##  Median :87.90            Median :79.75          Median : 48.80
##  Mean   :81.38            Mean   :76.58          Mean   : 53.09
##  3rd Qu.:96.85            3rd Qu.:92.50          3rd Qu.: 90.38
##  Max.   :99.30            Max.   :98.00          Max.   :100.00
##  NA's   :1                NA's   :6
##  lowsecondaryANAR_nodiff_rural lowsecondaryANAR_diff_rural
##  Min.   : 0.20                 Min.   : 0.00
##  1st Qu.:20.30                 1st Qu.:13.50
##  Median :47.50                 Median :35.20
##  Mean   :51.82                 Mean   :39.78
##  3rd Qu.:90.90                 3rd Qu.:55.25
##  Max.   :98.70                 Max.   :95.30
##  NA's   :2                     NA's   :9
##  uppsecondaryANAR_rural uppsecondaryANAR_nodiff_rural
##  Min.   : 0.000         Min.   : 0.000
##  1st Qu.: 9.125         1st Qu.: 8.625
##  Median :34.850         Median :25.800
##  Mean   :40.631         Mean   :39.350
##  3rd Qu.:74.050         3rd Qu.:72.975
```

```
##  Max.   :95.100         Max.   :96.200
##                         NA's   :2
##  uppsecondaryANAR_diff_rural primaryOOS_rural primaryOOS_nodiff_rural
##  Min.   : 0.00               Min.   : 0.90    Min.   : 0.50
##  1st Qu.: 4.75               1st Qu.: 4.15    1st Qu.: 3.95
##  Median :16.80               Median :12.05    Median :12.70
##  Mean   :25.39               Mean   :18.13    Mean   :18.15
##  3rd Qu.:38.83               3rd Qu.:28.25    3rd Qu.:28.75
##  Max.   :75.60               Max.   :68.50    Max.   :67.30
##  NA's   :16                                   NA's   :1
##  primaryOOS_diff_rural lowsecondaryOOS_rural lowsecondaryOOS_nodiff_rural
##  Min.   : 0.20         Min.   : 0.000        Min.   : 0.000
##  1st Qu.: 7.85         1st Qu.: 2.725        1st Qu.: 2.925
##  Median :16.20         Median : 8.600        Median : 9.250
##  Mean   :23.08         Mean   :14.325        Mean   :14.430
##  3rd Qu.:35.23         3rd Qu.:22.150        3rd Qu.:21.300
##  Max.   :72.20         Max.   :57.800        Max.   :57.400
##  NA's   :6                                   NA's   :2
##  lowsecondaryOOS_diff_rural uppsecondaryOOS_rural uppsecondaryOOS_nodiff_rural
##  Min.   : 0.0               Min.   : 0.00         Min.   : 0.00
##  1st Qu.: 7.7               1st Qu.:13.97         1st Qu.:12.45
##  Median :16.8               Median :23.20         Median :22.60
##  Mean   :20.6               Mean   :25.62         Mean   :25.16
##  3rd Qu.:31.6               3rd Qu.:37.92         3rd Qu.:36.73
##  Max.   :61.8               Max.   :66.00         Max.   :66.00
##  NA's   :9                                        NA's   :2
##  uppsecondaryOOS_diff_rural primarycomp_rural primarycomp_nodiff_rural
##  Min.   :13.70              Min.   : 10.40    Min.   : 11.90
##  1st Qu.:22.25              1st Qu.: 57.62    1st Qu.: 55.40
##  Median :34.90              Median : 85.30    Median : 83.25
##  Mean   :37.19              Mean   : 74.20    Mean   : 73.48
##  3rd Qu.:49.80              3rd Qu.: 99.00    3rd Qu.: 97.90
##  Max.   :68.90              Max.   :100.00    Max.   :100.00
##  NA's   :16                                   NA's   :2
##  primarycomp_diff_rural reading_rural   reading_nodiff_rural reading_diff_rural
##  Min.   : 3.80          Min.   : 2.10   Min.   : 2.40        Min.   : 1.30
##  1st Qu.:38.25          1st Qu.:11.55   1st Qu.:12.70        1st Qu.: 7.30
##  Median :60.70          Median :37.30   Median :39.00        Median :28.90
##  Mean   :55.28          Mean   :37.35   Mean   :38.18        Mean   :27.65
##  3rd Qu.:75.33          3rd Qu.:55.75   3rd Qu.:56.60        3rd Qu.:42.50
##  Max.   :97.20          Max.   :83.50   Max.   :83.80        Max.   :66.00
##  NA's   :14             NA's   :1       NA's   :1            NA's   :5
##  numeric_rural   numeric_nodiff_rural numeric_diff_rural
##  Min.   : 0.20   Min.   : 0.10        Min.   : 0.30
##  1st Qu.: 6.55   1st Qu.: 7.30        1st Qu.: 4.00
##  Median :21.60   Median :21.80        Median :14.10
##  Mean   :26.54   Mean   :27.15        Mean   :18.53
##  3rd Qu.:44.50   3rd Qu.:45.30        3rd Qu.:28.60
##  Max.   :69.40   Max.   :69.60        Max.   :60.10
##  NA's   :1       NA's   :1            NA's   :5
```

The values are now all numeric, except the levels column, and there are NA values in these numeric variables.

```
library(caret)

df_imp <- preProcess(df_c[,2:136], "knnImpute")
df_imputed <- predict(df_imp, df_c)
summary(df_imputed) # Now, the NA are filled in with values computed with KNN.
```

```
tooHigh <- findCorrelation(cor(df_imputed[,2:136]), 0.8)
df_filtered <- df_imputed[,-tooHigh]
level_final <- as.factor(make.names(df_c[,1]))
dim(df_filtered)

## [1] 32 13 # Now, the columns with highly correlated variables (>0.8) have been removed.

# There are now 32 rows and 13 columns.

summary(level_final)

## Least.Developed  Less.Developed  More.Developed
##              14              15               3

df_final <- cbind(level_final, df_filtered)

# Dataset splitting into train and test sets.

df_split <- sort(sample(nrow(df_final), nrow(df_final)*0.8))
df_train <- df_final[df_split,]
df_test <- df_final[-df_split,]

# Building Naïve Bayes model.
library(naivebayes)

df_nb <- naive_bayes(df_train[,2:13], df_train[,1])
df_nb_pred <- predict(df_nb, df_test[,2:13])
nb_result <- confusionMatrix(df_nb_pred, df_test[,1])
nb_result

## Confusion Matrix and Statistics
##
##                  Reference
## Prediction        Least.Developed Less.Developed More.Developed
##   Least.Developed               2              1              0
##   Less.Developed                1              1              0
##   More.Developed                0              2              0
##
## Overall Statistics
##
##                Accuracy : 0.4286
##                  95% CI : (0.099, 0.8159)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 0.8734
##
##                   Kappa : 0.125
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Least.Developed Class: Less.Developed
## Sensitivity                          0.6667                0.2500
## Specificity                          0.7500                0.6667
## Pos Pred Value                       0.6667                0.5000
## Neg Pred Value                       0.7500                0.4000
## Prevalence                           0.4286                0.5714
## Detection Rate                       0.2857                0.1429
## Detection Prevalence                 0.4286                0.2857
## Balanced Accuracy                    0.7083                0.4583
##                      Class: More.Developed
```

```
## Sensitivity                             NA
## Specificity                         0.7143
## Pos Pred Value                          NA
## Neg Pred Value                          NA
## Prevalence                          0.0000
## Detection Rate                      0.0000
## Detection Prevalence                0.2857
## Balanced Accuracy                       NA
```

```
# Naïve Bayes model gives a prediction accuracy of 0.4286.

df_knn <- train(x = df_train[,2:13], y = df_train[,1],
                method = "knn", tuneLength = 20)

df_knn_pred <- predict(df_knn, df_test[,2:13])
varImp(df_knn) # The table below shows the importance of each variable in the preprocess dataset.
```

```
## ROC curve variable importance
##
##   variables are sorted by maximum importance across the classes
##                                Least.Developed Less.Developed More.Developed
## numeric_nodiff_urban                    100.00         52.381       100.0000
## reading_nodiff_urban                    100.00         80.952       100.0000
## lowsecondaryANAR_nodiff_urban            92.06         52.381        92.0635
## lowsecondaryANAR_urban                   88.10         52.381        88.0952
## numeric_nodiff_rural                     84.13         42.857        84.1270
## primarycomp_nodiff_urban                 80.16          7.143        80.1587
## uppsecondaryANAR_nodiff_urban            64.29         47.619        64.2857
## uppsecondaryOOS_urban                    40.48         23.810        40.4762
## primaryANAR_nodiff_urban                 36.51         14.286        36.5079
## lowsecondaryOOS_nodiff_urban             36.51         28.571        36.5079
## lowsecondaryOOS_diff_male                28.57          0.000        28.5714
## numeric_diff_rural                       20.63          4.762        20.6349
## uppsecondaryOOS_nodiff_urban              1.19          1.190         0.7937
```

```
df_knn_result <- confusionMatrix(df_knn_pred, df_test$level_final)
df_knn_result
```

```
## Confusion Matrix and Statistics
##
##                  Reference
## Prediction        Least.Developed Less.Developed More.Developed
##   Least.Developed               3              2              0
##   Less.Developed                0              2              0
##   More.Developed                0              0              0
##
## Overall Statistics
##
##                Accuracy : 0.7143
##                  95% CI : (0.2904, 0.9633)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 0.3593
##
##                   Kappa : 0.4615
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Least.Developed Class: Less.Developed
## Sensitivity                          1.0000                0.5000
```

```
## Specificity                          0.5000               1.0000
## Pos Pred Value                       0.6000               1.0000
## Neg Pred Value                       1.0000               0.6000
## Prevalence                           0.4286               0.5714
## Detection Rate                       0.4286               0.2857
## Detection Prevalence                 0.7143               0.2857
## Balanced Accuracy                    0.7500               0.7500
##                       Class: More.Developed
## Sensitivity                              NA
## Specificity                               1
## Pos Pred Value                           NA
## Neg Pred Value                           NA
## Prevalence                                0
## Detection Rate                            0
## Detection Prevalence                      0
## Balanced Accuracy                        NA

# K-nearest neighbor model gives a prediction accuracy of 0.7143.

library(nnet)
df_nnet <- nnet(level_final~., df_train, size=5, decay=0.1)

varImp(df_nnet) # The table below shows importance of each variable in the preprocessed dataset.

##                                Overall Least.Developed Less.Developed
## lowsecondaryOOS_diff_male     6.714648        6.714648       6.714648
## primaryANAR_nodiff_urban     11.513821       11.513821      11.513821
## lowsecondaryANAR_urban        5.355194        5.355194       5.355194
## lowsecondaryANAR_nodiff_urban 6.347611        6.347611       6.347611
## uppsecondaryANAR_nodiff_urban 6.420607        6.420607       6.420607
## lowsecondaryOOS_nodiff_urban  8.494920        8.494920       8.494920
## uppsecondaryOOS_urban         5.910350        5.910350       5.910350
## uppsecondaryOOS_nodiff_urban  6.929178        6.929178       6.929178
## primarycomp_nodiff_urban      3.749887        3.749887       3.749887
## reading_nodiff_urban         14.945962       14.945962      14.945962
## numeric_nodiff_urban         10.198255       10.198255      10.198255
## numeric_nodiff_rural          5.996767        5.996767       5.996767
## numeric_diff_rural            7.422799        7.422799       7.422799
##                               More.Developed
## lowsecondaryOOS_diff_male           6.714648
## primaryANAR_nodiff_urban           11.513821
## lowsecondaryANAR_urban              5.355194
## lowsecondaryANAR_nodiff_urban       6.347611
## uppsecondaryANAR_nodiff_urban       6.420607
## lowsecondaryOOS_nodiff_urban        8.494920
## uppsecondaryOOS_urban               5.910350
## uppsecondaryOOS_nodiff_urban        6.929178
## primarycomp_nodiff_urban            3.749887
## reading_nodiff_urban               14.945962
## numeric_nodiff_urban               10.198255
## numeric_nodiff_rural                5.996767
## numeric_diff_rural                  7.422799

df_nnet_pred <- predict(df_nnet, df_test[,2:13], type = "class")
df_nnet_pred <- as.factor(df_nnet_pred)
df_nnet_result <- confusionMatrix(df_nnet_pred, df_test[,1])
```

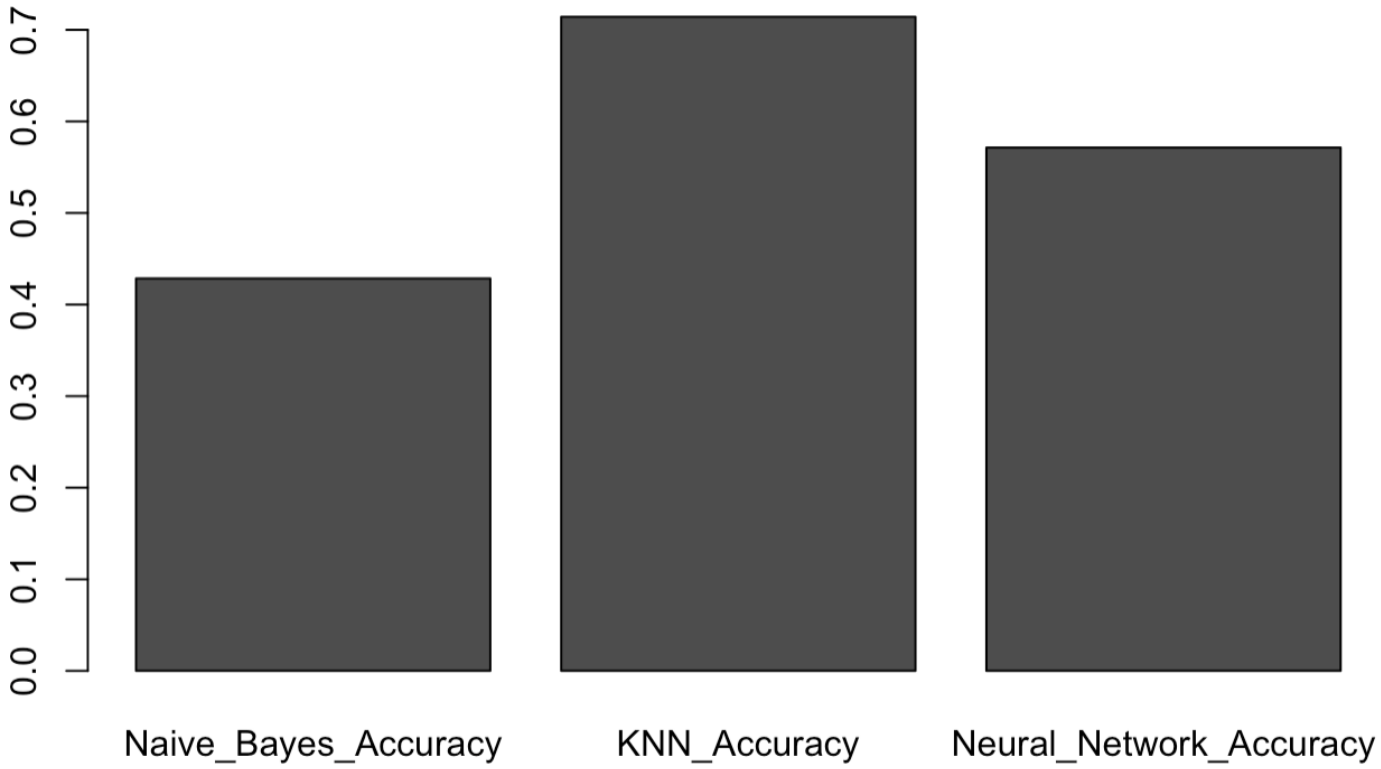
```
df_nnet_result
```

```
## Confusion Matrix and Statistics
##
##                  Reference
## Prediction        Least.Developed Less.Developed More.Developed
##   Least.Developed               1              1              0
##   Less.Developed                2              3              0
##   More.Developed                0              0              0
##
## Overall Statistics
##
##                Accuracy : 0.5714
##                  95% CI : (0.1841, 0.901)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 0.6531
##
##                   Kappa : 0.087
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: Least.Developed Class: Less.Developed
## Sensitivity                          0.3333                0.7500
## Specificity                          0.7500                0.3333
## Pos Pred Value                       0.5000                0.6000
## Neg Pred Value                       0.6000                0.5000
## Prevalence                           0.4286                0.5714
## Detection Rate                       0.1429                0.4286
## Detection Prevalence                 0.2857                0.7143
## Balanced Accuracy                    0.5417                0.5417
##                      Class: More.Developed
## Sensitivity                             NA
## Specificity                              1
## Pos Pred Value                          NA
## Neg Pred Value                          NA
## Prevalence                               0
## Detection Rate                           0
## Detection Prevalence                     0
## Balanced Accuracy                       NA
```

```
# Neural Network model gives a prediction accuracy of 0.5714.

# Below compares the accuracies of the three different models.

nb_overall <- as.data.frame(nb_result$overall)
Naive_Bayes_Accuracy <- nb_overall[1,]
knn_overall <- as.data.frame(df_knn_result$overall)
KNN_Accuracy <- knn_overall[1,]
nnet_overall <- as.data.frame(df_nnet_result$overall)
Neural_Network_Accuracy <- nnet_overall[1,]
xyz <- cbind(Naive_Bayes_Accuracy, KNN_Accuracy, Neural_Network_Accuracy)
barplot(xyz)
```


0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Naive_Bayes_Accuracy
KNN_Accuracy
Neural_Network_Accuracy