

Step-by-Step Guide to Data Importing and Presentation in R

StatCoding Ethiopia

Step 1: Install and Load Necessary Packages

```
1  install.packages("readxl")      # For reading Excel files
2  install.packages("DescTools")   # For enhanced frequency tables
3
4  library(readxl)
5  library(DescTools)
6
```

Step 2: Import and Explore the Dataset

```
1  hhs <- read_excel("dhhhu.xlsx")
2
3  head(hhs)          # Preview first few rows
4  str(hhs)           # Structure of the dataset
5  names(hhs)         # Column names
6  summary(hhs)       # Summary statistics
7
```

Step 3: Frequency Table for Marital Status

```
1  marital_freq <- table(hhs$marital)
2  marital_prop <- prop.table(marital_freq)
3
4  marital_detailed <- cbind(
5    Frequency = marital_freq,
6    Percentage = round(marital_prop * 100, 2),
7    Cumulative = cumsum(marital_freq)
8  )
9
10 print(marital_detailed)
11 Freq(hhs$marital) # From DescTools
12
```

Step 4: Bar Chart for Marital Status

```
1 barplot(marital_freq,  
2 main = "Bar Chart of Marital Status",  
3 xlab = "Marital Status",  
4 ylab = "Frequency",  
5 col = "skyblue")  
6
```

Step 5: Pie Chart for Marital Status

```
1 marital_pct <- round(100 * marital_freq / sum(marital_freq), 1)  
2 marital_labels <- paste(names(marital_freq), marital_pct, "%")  
3  
4 pie(marital_freq,  
5 labels = marital_labels,  
6 main = "Marital Status Pie Chart",  
7 col = rainbow(length(marital_freq)))  
8
```

Step 6: Ungrouped Frequency Table for Family Size

```
1 familysize_freq <- table(hhs$familysize)  
2 familysize_detailed <- data.frame(  
3 FamilySize = as.numeric(names(familysize_freq)),  
4 Frequency = as.numeric(familysize_freq)  
5 )  
6 familysize_detailed <- familysize_detailed[order(familysize_detailed$  
7 FamilySize), ]  
8 familysize_detailed$Cumulative <- cumsum(familysize_detailed$Frequency  
9 )  
10 print(familysize_detailed)
```

Step 7: Bar and Pie Charts for Family Size

```
1 barplot(familysize_freq,  
2 main = "Distribution of Family Size",  
3 xlab = "Family Size",  
4 ylab = "Frequency",  
5 col = "lightgreen",  
6 border = "darkgreen",  
7 space = 0.5)  
8  
9 pie(familysize_freq,  
10 main = "Family Size Pie Chart",
```

```

11   col = heat.colors(length(familysize_freq)))
12

```

Step 8: Grouped Frequency Distribution for Age

8.1 Determine Class Parameters

```

1   age_range <- range(hhs$age, na.rm = TRUE)
2   n <- length(na.omit(hhs$age))
3   k <- ceiling(1 + 3.322 * log10(n)) # Sturges' Rule
4   class_width <- ceiling(diff(age_range) / k)
5

```

8.2 Create Class Intervals and Frequency Table

```

1   lower_limit <- floor(age_range[1])
2   upper_limit <- lower_limit + class_width * k
3   age_breaks <- seq(lower_limit, upper_limit, by = class_width)
4
5   age_groups <- cut(hhs$age,
6   breaks = age_breaks,
7   right = FALSE,
8   include.lowest = TRUE)
9   age_freq <- table(age_groups)
10

```

8.3 Compute Class Marks and Percentages

```

1   lower_bounds <- age_breaks[-length(age_breaks)]
2   upper_bounds <- age_breaks[-1]
3   class_marks <- (lower_bounds + upper_bounds) / 2
4
5   rel_percent <- round(prop.table(age_freq) * 100, 2)
6   cum_freq <- cumsum(age_freq)
7   cum_percent <- round(cumsum(rel_percent), 2)
8

```

8.4 Final Grouped Frequency Table

```

1   age_detailed <- data.frame(
2   AgeGroup      = names(age_freq),
3   LowerBound    = lower_bounds,
4   UpperBound    = upper_bounds,
5   ClassMark     = round(class_marks, 1),
6   Frequency     = as.numeric(age_freq),
7   RelativePercent = rel_percent,
8   CumulativeFreq = as.numeric(cum_freq),

```

```
9     CumulativePercent= cum_percent
10   )
11
12   print(age_detailed)
13
```