USAGE

Boniface Kalong

2023-06-09

library(readxl)

## Warning: package 'readxl' was built under R version 4.2.3

USAGE <- read\_excel("USAGE.xlsx")

## New names:  
## \* `` -> `...13`

# Assign the column names  
colnames(USAGE) <- c(  
 "Timestamp",  
 "Gender",  
 "Education",  
 "Employment",  
 "AgeRange",  
 "NHIS\_Subscription\_Duration",  
 "Previous\_Healthcare\_Access",  
 "Healthcare\_Delivery\_Perception",  
 "NHIS\_Mobile\_Renewal\_Awareness",  
 "Previous\_Mobile\_Renewal\_Use",  
 "Mobile\_Renewal\_Rating",  
 "Mobile\_Renewal\_Continuation\_Opinion",  
 "...13",  
 "Mobile\_Renewal\_Speed\_Efficiency",  
 "Mobile\_Renewal\_Time\_Resource\_Savings",  
 "Mobile\_Renewal\_Effectiveness",  
 "Mobile\_Renewal\_Productivity",  
 "Mobile\_Renewal\_Ease\_of\_Use",  
 "Mobile\_Renewal\_Learning\_Ease",  
 "Mobile\_Renewal\_Mental\_Effort",  
 "Mobile\_Renewal\_Interactions\_Clear",  
 "Mobile\_Renewal\_Intention\_to\_Use",  
 "Mobile\_Renewal\_Benefits\_Awareness",  
 "Mobile\_Renewal\_Mobile\_Money\_Access",  
 "Mobile\_Renewal\_Service\_Usefulness",  
 "Mobile\_Renewal\_Service\_Quality",  
 "Mobile\_Renewal\_Service\_Cost",  
 "The\_service\_charge\_for\_Mobile\_Renewal",  
 "Mobile\_Renewal\_Financial\_Barriers",  
 "Mobile\_Renewal\_Service\_Accessibility",  
 "Mobile\_Renewal\_Service\_Disruptions",  
 "Mobile\_Renewal\_Service\_Availability",  
 "Mobile\_Renewal\_Service\_Reliability",  
 "Mobile\_Renewal\_Service\_Authenticity",  
 "Trust\_Telecom\_Operators",  
 "Mobile\_Renewal\_Service\_Data\_Security",  
 "Mobile\_Renewal\_Service\_Satisfaction"  
)

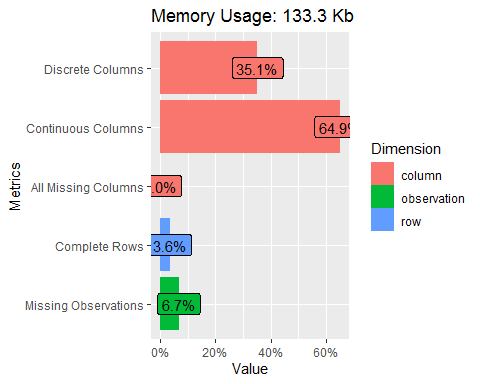
library(DataExplorer)  
introduce(USAGE)

## # A tibble: 1 x 9  
## rows columns discrete\_columns continuous\_columns all\_missing\_columns  
## <int> <int> <int> <int> <int>  
## 1 418 37 13 24 0  
## # i 4 more variables: total\_missing\_values <int>, complete\_rows <int>,  
## # total\_observations <int>, memory\_usage <dbl>

The summary table above indicates the following information about the dataset:

* Rows: 418
* Columns: 37
* Discrete columns: 13
* Continuous columns: 24
* All missing columns: 0 (no columns with missing values)
* Total missing values: This information is not provided in the summary table.
* Complete rows: This information is not provided in the summary table.
* Total observations: This information is not provided in the summary table.
* Memory usage: This information is not provided in the summary table.

plot\_intro(USAGE)



Function to abbreviate columns by using the first letter of each word in the column.

# source("functions.R")  
# # Call the function to abbreviate columns  
# USAGE <- abbreviate\_df\_columns(USAGE)  
#   
# # Print the abbreviated column names  
# print(colnames(USAGE))  
# # Call the function to abbreviate columns  
# column\_names <- abbreviate\_columns(USAGE\_xlsx)  
#   
# # Print the data frame with previous and abbreviated column names  
# # View(column\_names)  
# # colnames(USAGE)

library(skimr)  
skim(USAGE)

Data summary

|  |  |
| --- | --- |
| Name | USAGE |
| Number of rows | 418 |
| Number of columns | 37 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 12 |
| numeric | 24 |
| POSIXct | 1 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gender | 2 | 1.00 | 4 | 6 | 0 | 2 | 0 |
| Education | 6 | 0.99 | 3 | 24 | 0 | 6 | 0 |
| Employment | 5 | 0.99 | 10 | 14 | 0 | 3 | 0 |
| AgeRange | 4 | 0.99 | 11 | 18 | 0 | 4 | 0 |
| NHIS\_Subscription\_Duration | 26 | 0.94 | 16 | 23 | 0 | 3 | 0 |
| Previous\_Healthcare\_Access | 9 | 0.98 | 2 | 3 | 0 | 2 | 0 |
| Healthcare\_Delivery\_Perception | 7 | 0.98 | 4 | 9 | 0 | 5 | 0 |
| NHIS\_Mobile\_Renewal\_Awareness | 9 | 0.98 | 2 | 3 | 0 | 2 | 0 |
| Previous\_Mobile\_Renewal\_Use | 10 | 0.98 | 2 | 3 | 0 | 2 | 0 |
| Mobile\_Renewal\_Rating | 48 | 0.89 | 4 | 9 | 0 | 5 | 0 |
| Mobile\_Renewal\_Continuation\_Opinion | 12 | 0.97 | 9 | 14 | 0 | 3 | 0 |
| …13 | 393 | 0.06 | 8 | 8 | 0 | 1 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mobile\_Renewal\_Speed\_Efficiency | 11 | 0.97 | 1.85 | 1.30 | 1 | 1.00 | 1 | 2 | 5 | ▇▂▁▁▁ |
| Mobile\_Renewal\_Time\_Resource\_Savings | 12 | 0.97 | 1.81 | 1.25 | 1 | 1.00 | 1 | 2 | 5 | ▇▂▁▁▁ |
| Mobile\_Renewal\_Effectiveness | 14 | 0.97 | 2.03 | 1.22 | 1 | 1.00 | 2 | 3 | 5 | ▇▅▂▁▁ |
| Mobile\_Renewal\_Productivity | 14 | 0.97 | 2.50 | 1.26 | 1 | 1.75 | 2 | 3 | 5 | ▆▇▅▃▂ |
| Mobile\_Renewal\_Ease\_of\_Use | 13 | 0.97 | 2.01 | 1.27 | 1 | 1.00 | 2 | 3 | 5 | ▇▃▂▁▁ |
| Mobile\_Renewal\_Learning\_Ease | 17 | 0.96 | 2.08 | 1.23 | 1 | 1.00 | 2 | 3 | 5 | ▇▆▂▂▁ |
| Mobile\_Renewal\_Mental\_Effort | 17 | 0.96 | 3.21 | 1.30 | 1 | 2.00 | 3 | 4 | 5 | ▃▅▇▇▆ |
| Mobile\_Renewal\_Interactions\_Clear | 14 | 0.97 | 2.09 | 1.15 | 1 | 1.00 | 2 | 3 | 5 | ▇▇▃▁▂ |
| Mobile\_Renewal\_Intention\_to\_Use | 13 | 0.97 | 1.82 | 1.22 | 1 | 1.00 | 1 | 2 | 5 | ▇▃▁▁▁ |
| Mobile\_Renewal\_Benefits\_Awareness | 16 | 0.96 | 2.05 | 1.18 | 1 | 1.00 | 2 | 3 | 5 | ▇▆▃▁▁ |
| Mobile\_Renewal\_Mobile\_Money\_Access | 15 | 0.96 | 1.84 | 1.26 | 1 | 1.00 | 1 | 2 | 5 | ▇▃▁▁▁ |
| Mobile\_Renewal\_Service\_Usefulness | 16 | 0.96 | 2.10 | 1.16 | 1 | 1.00 | 2 | 3 | 5 | ▇▇▃▂▁ |
| Mobile\_Renewal\_Service\_Quality | 40 | 0.90 | 1.98 | 1.11 | 1 | 1.00 | 2 | 2 | 5 | ▇▆▂▂▁ |
| Mobile\_Renewal\_Service\_Cost | 22 | 0.95 | 3.53 | 1.35 | 1 | 3.00 | 4 | 5 | 5 | ▃▃▅▇▇ |
| The\_service\_charge\_for\_Mobile\_Renewal | 21 | 0.95 | 3.40 | 1.36 | 1 | 2.00 | 4 | 5 | 5 | ▃▃▅▇▇ |
| Mobile\_Renewal\_Financial\_Barriers | 23 | 0.94 | 3.37 | 1.32 | 1 | 2.00 | 4 | 4 | 5 | ▃▃▅▇▆ |
| Mobile\_Renewal\_Service\_Accessibility | 20 | 0.95 | 2.40 | 1.18 | 1 | 1.00 | 2 | 3 | 5 | ▇▇▆▃▂ |
| Mobile\_Renewal\_Service\_Disruptions | 24 | 0.94 | 3.15 | 1.22 | 1 | 2.00 | 3 | 4 | 5 | ▃▅▇▇▅ |
| Mobile\_Renewal\_Service\_Availability | 20 | 0.95 | 2.29 | 1.33 | 1 | 1.00 | 2 | 3 | 5 | ▇▅▅▂▂ |
| Mobile\_Renewal\_Service\_Reliability | 32 | 0.92 | 2.08 | 1.12 | 1 | 1.00 | 2 | 3 | 5 | ▇▇▃▂▁ |
| Mobile\_Renewal\_Service\_Authenticity | 31 | 0.93 | 2.18 | 1.10 | 1 | 1.00 | 2 | 3 | 5 | ▇▇▅▂▁ |
| Trust\_Telecom\_Operators | 32 | 0.92 | 2.27 | 1.10 | 1 | 1.00 | 2 | 3 | 5 | ▆▇▆▂▁ |
| Mobile\_Renewal\_Service\_Data\_Security | 33 | 0.92 | 2.38 | 1.25 | 1 | 1.00 | 2 | 3 | 5 | ▇▇▆▃▂ |
| Mobile\_Renewal\_Service\_Satisfaction | 33 | 0.92 | 2.06 | 1.21 | 1 | 1.00 | 2 | 3 | 5 | ▇▆▃▁▁ |

**Variable type: POSIXct**

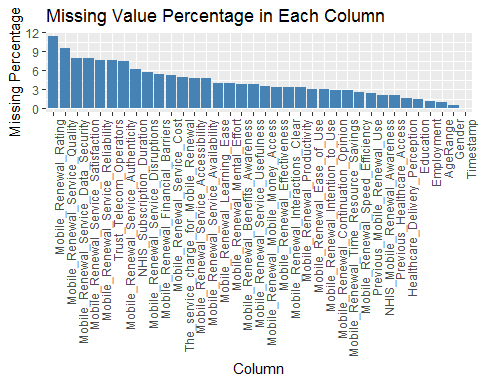
| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| Timestamp | 0 | 1 | 2021-07-17 14:09:20 | 2021-07-23 16:15:58 | 2021-07-19 17:06:52 | 418 |

Based on the data summary, we can see that there are 418 rows and 37 columns in the dataset. The data types of the columns are as follows:

Character: 12 columns Numeric: 24 columns POSIXct: 1 column There are no group variables in the dataset.

# Calculate the missing value percentages for each column  
missing\_percentages <- colSums(is.na(USAGE)) / nrow(USAGE) \* 100  
  
# Identify columns with missing value percentages above 50%  
columns\_to\_drop <- names(missing\_percentages[missing\_percentages > 50])  
  
# Drop the columns from the data frame  
USAGE <- USAGE[, !(names(USAGE) %in% columns\_to\_drop)]

library(ggplot2)  
# Create a data frame with the column names and their corresponding missing value percentages  
missing\_data <- data.frame(  
 Column = colnames(USAGE),  
 Missing\_Percentage = colSums(is.na(USAGE)) / nrow(USAGE) \* 100  
)  
  
# Plot the missing value percentages  
ggplot(missing\_data, aes(x = reorder(Column, -Missing\_Percentage), y = Missing\_Percentage)) +  
 geom\_bar(stat = "identity", fill = "steelblue") +  
 labs(x = "Column", y = "Missing Percentage") +  
 ggtitle("Missing Value Percentage in Each Column") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))



# library(summarytools)  
#   
# dfSummary(  
# USAGE,  
# varnumbers = FALSE,  
# round.digits = 2,  
# plain.ascii = FALSE,  
# style = "grid",  
# graph.magnif = .33,  
# valid.col = FALSE,  
# tmp.img.dir = "img"  
# )

library(tidyverse)  
library(dplyr)  
source('functions.R')  
filled\_data <- fillMissingValues(USAGE)  
Data <- filled\_data%>%  
 mutate\_if(is.numeric, round)

# Calculate the percentage for each category  
Individual\_ratinng <- Data %>%  
 group\_by(Mobile\_Renewal\_Rating) %>%  
 tally() %>%  
 mutate(Percentage = n / nrow(filled\_data) \* 100,  
 ymax = cumsum(Percentage),  
 ymin = ifelse(is.na(lag(ymax)), 0, lag(ymax)))

selected\_columns <- Data%>%  
 select(Gender, Education, Employment, AgeRange, Mobile\_Renewal\_Ease\_of\_Use)

The selected\_columns data frame provided contains four columns: Piyg, Hloe, Waye, and Wiyar. Each column represents different attributes or categories. Here’s a brief description of each column:

**Piyg**: This column represents the gender of the individuals. It contains values such as “Male” and “Female”.

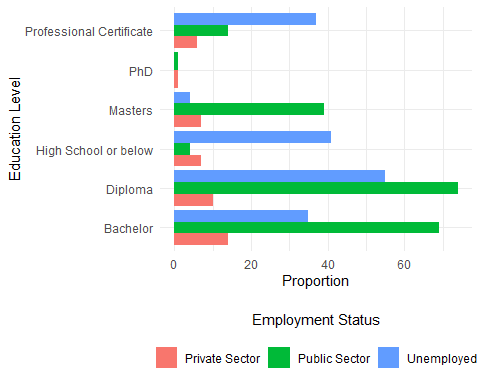
**Hloe**: This column represents the highest level of education of the individuals. It includes categories like “High School or below”, “Bachelor”, and “Diploma”.

**Waye**: This column represents the employment status of the individuals. It includes categories such as “Unemployed” and “Public Sector”.

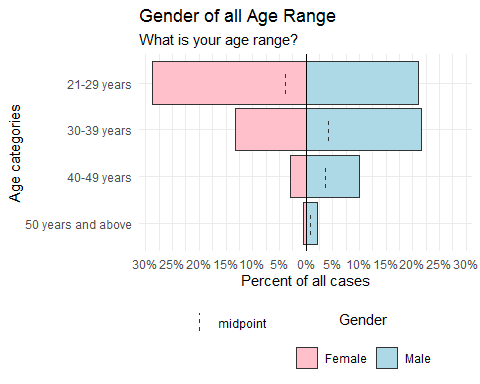
**Wiyar**: This column represents the age range of the individuals. It includes categories like “40-49 years”, “50 years and above”, “21-29 years”, and so on.

**TMRSietu:** This column represents the individual rating if the Mobile Renewal Service is easy to use

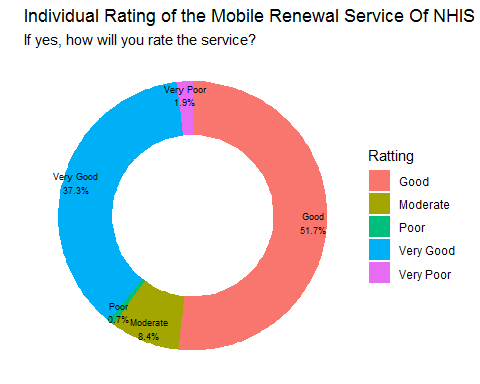
ggplot(selected\_columns, aes(x = Education, fill = Employment)) +  
 geom\_bar(position = position\_dodge()) +  
 coord\_flip() +  
 labs(x = "Education Level", y = "Proportion", fill = "Employment Status") +  
 # scale\_fill\_manual(values = c("Female" = "pink", "Male" = "lightblue")) +  
 theme\_minimal()+  
 theme(  
 legend.title = element\_text(hjust = 0.5, margin = margin(b = 10)),  
 legend.position = "bottom",  
 legend.direction = "horizontal"  
 ) +  
 guides(fill = guide\_legend(title.position = "top"))



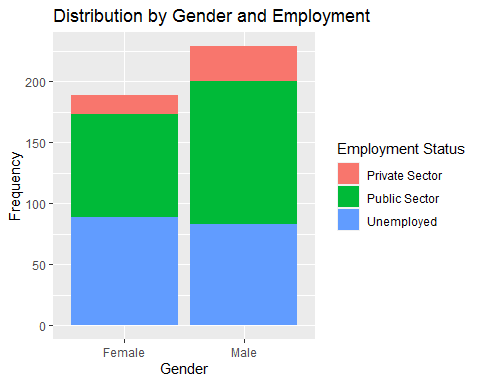
library(apyramid)  
pyramid <- selected\_columns %>%  
 ## ensure that age and gender are factors  
 mutate(Age\_Group = factor(AgeRange, levels = c("50 years and above",  
 "40-49 years",  
 "30-39 years",  
 "21-29 years")),  
 Gender = factor(Gender)) %>%  
 group\_by(Age\_Group, Gender, .groups = 'drop') %>%  
 ## add the counts for each age group and gender together  
 summarise(population = sum(n())) %>%  
 ## remove the grouping so we can calculate the overall proportion  
 ungroup() %>%  
 mutate(proportion = population / sum(population))  
  
## plot pyramid  
pyramid %>%  
 age\_pyramid(  
 age\_group = "Age\_Group",  
 split\_by = "Gender",  
 count = proportion,  
 proportional = TRUE) +  
 scale\_fill\_manual( # specify colors AND labels  
 values = c("pink", "lightblue"),   
 labels = c("m" = "Male", "f" = "Female"))+  
 labs(y = "Percent of all cases", # note x and y labs are switched  
 x = "Age categories",   
 fill = "Gender",   
 # caption = "",  
 title = "Gender of all Age Range",  
 subtitle = "What is your age range?")+  
 theme\_minimal()+  
 theme(  
 legend.title = element\_text(hjust = 0.5, margin = margin(b = 10)),  
 legend.position = "bottom",  
 legend.direction = "horizontal"  
 ) +  
 guides(fill = guide\_legend(title.position = "top"))



ggplot(data = Individual\_ratinng )+  
 geom\_rect(aes(ymin = ymin, ymax = ymax,fill = Mobile\_Renewal\_Rating,  
 xmin= 2, xmax = 4))+  
 coord\_polar(theta="y")+  
 theme\_void()+  
 xlim(c(-1, 4))+  
 geom\_text(aes(x = 3.5, y = Percentage ,label = paste0(Mobile\_Renewal\_Rating,"\n",round(Percentage,1), "%")),  
 position = position\_stack(vjust = 0.5), size = 2.5)+  
 labs(title = "Individual Rating of the Mobile Renewal Service Of NHIS",  
 fill = "Ratting")+  
 labs(subtitle = "If yes, how will you rate the service?")



# Calculate the frequency of Piyg and Waye combinations  
frequency <- selected\_columns %>%  
 group\_by(Gender, Employment) %>%  
 tally()  
  
# Bar plot  
bar\_plot <- ggplot(frequency, aes(x = Gender, y = n, fill = Employment)) +  
 geom\_bar(stat = "identity", position = "stack") +  
 labs(title = "Distribution by Gender and Employment",  
 x = "Gender",  
 y = "Frequency",  
 fill = "Employment Status")  
  
  
  
# Display the plots  
bar\_plot



Convert all columns except the first to factors

To assign survey questions to corresponding latent variables as indicators and determine appropriate measurement models, we need more information about the nature of the survey questions. The type of measurement model (binary, ordinal, or continuous) depends on the scaling and response options of each question.

# Convert all columns except the first to factors  
USAGE[, -1] <- lapply(USAGE[, -1], as.factor)

## Likert Scale Analysis

1. Define latent variables:
   1. Gender: Create a latent variable representing gender, which can be indicated by the responses to question 2.
   2. Education Level: Create a latent variable representing the highest level of education, based on question 3.
   3. Employment Status: Create a latent variable representing employment status, using the responses from question 4.
   4. Age Range: Create a latent variable representing age range, based on the responses to question 5.
   5. Subscriber Duration: Create a latent variable representing the duration of subscription to the National Health Insurance Scheme, using the responses from question 6.
   6. Access Healthcare: Create a latent variable representing previous access to healthcare under the National Health Insurance Scheme, based on question 7.
   7. Healthcare Delivery: Create a latent variable representing the perception of healthcare delivery under the National Health Insurance Scheme, using the responses from question 8.
   8. Awareness of Mobile Renewal Service: Create a latent variable representing awareness of the NHIS Mobile Renewal Service, based on the responses to question 9.
   9. Mobile Renewal Usage: Create a latent variable representing the usage of the Mobile Renewal Service, using the responses from question 10.
   10. Service Rating: Create a latent variable representing the rating of the Mobile Renewal Service, based on the responses to question 11.
   11. Service Continuation: Create a latent variable representing the agreement on the continuation or suspension of the Mobile Renewal Service, using the responses from question 12.
   12. Perceived Benefits: Create a latent variable representing the perceived benefits of using the Mobile Renewal Service, based on the responses to questions 14-17.
   13. Usability: Create a latent variable representing the usability of the Mobile Renewal Service, using the responses from questions 18-21.
   14. Trust: Create a latent variable representing trust in the Mobile Renewal Service, based on the responses to questions 24-27.
   15. Accessibility: Create a latent variable representing the accessibility of the Mobile Renewal Service, using the responses from questions 30-32.
   16. Reliability: Create a latent variable representing the reliability of the Mobile Renewal Service, based on the responses to questions 33-34.
   17. Data Security: Create a latent variable representing the perception of data security in using the Mobile Renewal Service, using the responses from questions 35-36.
   18. Satisfaction: Create a latent variable representing satisfaction with the overall performance of the Mobile Renewal Service, based on the responses to question 37.

## Modelling

1. Define measurement model:
   * Assign survey questions to corresponding latent variables as indicators. For example, question 2 will be an indicator of the Gender latent variable, question 3 will be an indicator of the Education Level latent variable, and so on. Assign appropriate measurement models (e.g., binary, ordinal, or continuous) based on the nature of the survey questions.
2. Define structural model:
   * Specify the relationships between latent variables based on theoretical expectations. For example, we hypothesize that Age Range, Gender,Education Level ,Employment Status affects or influences awareness of the Mobile Renewal Service.
3. Estimate the model:
   * Use SEM software (such as lavaan, AMOS, or Mplus) to estimate the model parameters. The software will provide estimates for factor loadings, regression coefficients, and latent variable variances and covariances.

# install.packages("lavaan")  
library(lavaan)  
  
model <- '  
 # Measurment   
 # Regression  
 # Residual Correlation'

![Run Current Chunk](data:image/png;base64;base64,iVBORw0KGgoAAAANSUhEUgAAACgAAAAaCAYAAADFTB7LAAAAa0lEQVR42u3OywnAIBBAwcXSUoCW5D11xDoNCBGNv0MOecJOBSOi1OZMsJ4dvFxEJ1OQnMxBarIKEpNNkJbsBknJYZCSnAYJyVVQziNig7/nZkFEbhTE5HpBVO4dxOXKIDL3BLG5BJ1T6rsbMfep2CaMN00AAAAASUVORK5CYII=)

## Model Evaluation

1. Assess model fit:
   * Evaluate the goodness-of-fit of the model using fit indices such as the chi-square test, Comparative Fit