FML-Ass 1

ANANDAGANIREDDY\_N

2024-02-04

**1.Importing the Dataset:**

# Downloading the Behavioral Risk Factor dataset from Kaggle.  
url <- "https://www.kaggle.com/datasets/gennadiimanzhos/behavioral-risk-factor-hrqol?resource=download"  
BRF\_data <- read.csv("C://Users//dell//Dropbox//My PC (DESKTOP-LE4HNP6)//Desktop//FML-Spring//BRF\_Data.csv")

# First few rows of the dataset  
head(BRF\_data)

## Year LocationAbbr LocationDesc Category Topic  
## 1 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 2 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 3 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 4 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 5 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 6 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## Question DataSource Data\_Value\_Unit  
## 1 Mean days of activity limitation HRQOL   
## 2 Mean days of activity limitation HRQOL   
## 3 Mean days of activity limitation HRQOL   
## 4 Mean days of activity limitation HRQOL   
## 5 Mean days of activity limitation HRQOL   
## 6 Mean days of activity limitation HRQOL   
## Data\_Value\_Type Data\_Value Data\_Value\_Footnote\_Symbol  
## 1 Average number of days 0.9   
## 2 Average number of days 2.0   
## 3 Average number of days 2.3   
## 4 Average number of days 4.2   
## 5 Average number of days 4.0   
## 6 Average number of days 3.4   
## Data\_Value\_Footnote Data\_Value\_Std\_Err Low\_Confidence\_Limit  
## 1 NA 0.3  
## 2 NA 1.3  
## 3 NA 1.8  
## 4 NA 3.6  
## 5 NA 3.4  
## 6 NA 2.7  
## High\_Confidence\_Limit Sample\_Size Break\_Out Break\_Out\_Category  
## 1 1.5 202 18-24 years old Age Group  
## 2 2.7 597 25-34 years old Age Group  
## 3 2.9 892 35-44 years old Age Group  
## 4 4.9 1352 45-54 years old Age Group  
## 5 4.6 1621 55-64 years old Age Group  
## 6 4.0 1477 65-74 years old Age Group  
## GeoLocation CategoryId TopicId QuestionId  
## 1 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 2 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 3 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 4 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 5 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 6 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## LocationId BreakOutId BreakOutCategoryid  
## 1 1 Age1 GPAGE  
## 2 1 Age2 GPAGE  
## 3 1 Age3 GPAGE  
## 4 1 Age4 GPAGE  
## 5 1 Age5 GPAGE  
## 6 1 Age6 GPAGE

**2.Printing the Descriptive Statistics:**

# Summary statistics for quantitative variables  
summary(BRF\_data$Data\_Value\_Unit)

## Length Class Mode   
## 126464 character character

# Frequency table for categorical variable  
table(BRF\_data$Topic)

##   
## Activity Limitation General Health Mental Health Physical Health   
## 31616 31616 31616 31616

**3. Transforming the Variable:**

library("dplyr")

## Warning: package 'dplyr' was built under R version 4.3.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

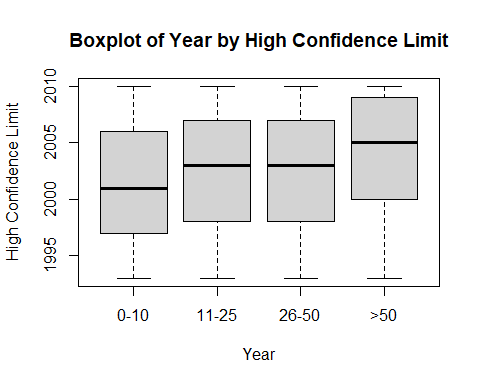
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

BRF\_data <- BRF\_data %>%  
 mutate(Data\_Value = sqrt(Data\_Value))  
head(BRF\_data)

## Year LocationAbbr LocationDesc Category Topic  
## 1 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 2 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 3 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 4 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 5 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## 6 2010 AL Alabama Health Status/Healthy Days Activity Limitation  
## Question DataSource Data\_Value\_Unit  
## 1 Mean days of activity limitation HRQOL   
## 2 Mean days of activity limitation HRQOL   
## 3 Mean days of activity limitation HRQOL   
## 4 Mean days of activity limitation HRQOL   
## 5 Mean days of activity limitation HRQOL   
## 6 Mean days of activity limitation HRQOL   
## Data\_Value\_Type Data\_Value Data\_Value\_Footnote\_Symbol  
## 1 Average number of days 0.9486833   
## 2 Average number of days 1.4142136   
## 3 Average number of days 1.5165751   
## 4 Average number of days 2.0493902   
## 5 Average number of days 2.0000000   
## 6 Average number of days 1.8439089   
## Data\_Value\_Footnote Data\_Value\_Std\_Err Low\_Confidence\_Limit  
## 1 NA 0.3  
## 2 NA 1.3  
## 3 NA 1.8  
## 4 NA 3.6  
## 5 NA 3.4  
## 6 NA 2.7  
## High\_Confidence\_Limit Sample\_Size Break\_Out Break\_Out\_Category  
## 1 1.5 202 18-24 years old Age Group  
## 2 2.7 597 25-34 years old Age Group  
## 3 2.9 892 35-44 years old Age Group  
## 4 4.9 1352 45-54 years old Age Group  
## 5 4.6 1621 55-64 years old Age Group  
## 6 4.0 1477 65-74 years old Age Group  
## GeoLocation CategoryId TopicId QuestionId  
## 1 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 2 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 3 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 4 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 5 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## 6 (32.84057112200048, -86.63186076199969) HLT001 POORHLTH AL002  
## LocationId BreakOutId BreakOutCategoryid  
## 1 1 Age1 GPAGE  
## 2 1 Age2 GPAGE  
## 3 1 Age3 GPAGE  
## 4 1 Age4 GPAGE  
## 5 1 Age5 GPAGE  
## 6 1 Age6 GPAGE

**4. Plot Quantitative and Scatter plot:**

# Creating categories for High\_Confidence\_Limit  
BRF\_data$High\_Confidence\_Category <- cut(BRF\_data$High\_Confidence\_Limit,  
 breaks = c(0, 10, 25, 50, Inf),  
 labels = c("0-10", "11-25", "26-50", ">50"),  
 include.lowest = TRUE)  
  
# Boxplot for quantitative variable across different categories  
boxplot(Year ~ High\_Confidence\_Category, data = BRF\_data,  
 main = "Boxplot of Year by High Confidence Limit",  
 xlab = "Year", ylab = "High Confidence Limit")



# Scatterplot for two quantitative variables  
plot(BRF\_data$Low\_Confidence\_Limit, BRF\_data$High\_Confidence\_Limit, main = "Scatterplot of Low Confidence Limit vs High Confidence Limit", xlab = "Low Confidence Limit", ylab = "High Confidence Limit", col="violet")

