

# Dementia\_Dataset\_Kaggle

Trinity disAbility Service - Mark Likeman

2024-11-01

## Load data

```
dementia_data <- read.csv("/Users/marklikeman/documents/Dementia_Kaggle/dementia_dataset.csv", header =
```

## About Longitudinal research

### Package Loading Function

```
pkgTest <- function(pkg){  
  new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]  
  if (length(new.pkg)) install.packages(new.pkg, dependencies = TRUE)  
  sapply(pkg, require, character.only = TRUE)  
}  
lapply(c("tidyverse"), pkgTest)
```

```
## [[1]]  
## tidyverse  
##      TRUE
```

## Data Summary

```
# View the first few rows  
head(dementia_data)
```

```
##   Subject.ID      MRI.ID      Group Visit MR.Delay M.F Hand Age EDUC SES  
## 1 OAS2_0001 OAS2_0001_MR1 Nondemented    1         0  M   R  87  14   2  
## 2 OAS2_0001 OAS2_0001_MR2 Nondemented    2        457  M   R  88  14   2  
## 3 OAS2_0002 OAS2_0002_MR1   Demented    1         0  M   R  75  12  NA  
## 4 OAS2_0002 OAS2_0002_MR2   Demented    2        560  M   R  76  12  NA  
## 5 OAS2_0002 OAS2_0002_MR3   Demented    3       1895  M   R  80  12  NA  
## 6 OAS2_0004 OAS2_0004_MR1 Nondemented    1         0  F   R  88  18   3  
##   MMSE CDR eTIV  nWBV  ASF  
## 1   27  0.0 1987 0.696 0.883  
## 2   30  0.0 2004 0.681 0.876  
## 3   23  0.5 1678 0.736 1.046  
## 4   28  0.5 1738 0.713 1.010  
## 5   22  0.5 1698 0.701 1.034  
## 6   28  0.0 1215 0.710 1.444
```

```
# View data structure  
str(dementia_data)
```

```
## 'data.frame':   373 obs. of  15 variables:
```

```
## $ Subject.ID: chr "OAS2_0001" "OAS2_0001" "OAS2_0002" "OAS2_0002" ...
## $ MRI.ID : chr "OAS2_0001_MR1" "OAS2_0001_MR2" "OAS2_0002_MR1" "OAS2_0002_MR2" ...
## $ Group : chr "Nondemented" "Nondemented" "Demented" "Demented" ...
## $ Visit : int 1 2 1 2 3 1 2 1 2 3 ...
## $ MR.Delay : int 0 457 0 560 1895 0 538 0 1010 1603 ...
## $ M.F : chr "M" "M" "M" "M" ...
## $ Hand : chr "R" "R" "R" "R" ...
## $ Age : int 87 88 75 76 80 88 90 80 83 85 ...
## $ EDUC : int 14 14 12 12 12 18 18 12 12 12 ...
## $ SES : int 2 2 NA NA NA 3 3 4 4 4 ...
## $ MMSE : int 27 30 23 28 22 28 27 28 29 30 ...
## $ CDR : num 0 0 0.5 0.5 0.5 0 0 0 0.5 0 ...
## $ eTIV : int 1987 2004 1678 1738 1698 1215 1200 1689 1701 1699 ...
## $ nWBV : num 0.696 0.681 0.736 0.713 0.701 0.71 0.718 0.712 0.711 0.705 ...
## $ ASF : num 0.883 0.876 1.046 1.01 1.034 ...
```

```
# Summary statistics
summary(dementia_data)
```

```
## Subject.ID      MRI.ID      Group      Visit
## Length:373      Length:373      Length:373      Min. :1.000
## Class :character Class :character Class :character 1st Qu.:1.000
## Mode :character Mode :character Mode :character Median :2.000
##                                     Mean :1.882
##                                     3rd Qu.:2.000
##                                     Max. :5.000
##
## MR.Delay      M.F      Hand      Age
## Min. : 0.0      Length:373      Length:373      Min. :60.00
## 1st Qu.: 0.0      Class :character      Class :character 1st Qu.:71.00
## Median : 552.0      Mode :character      Mode :character Median :77.00
## Mean : 595.1                                     Mean :77.01
## 3rd Qu.: 873.0                                     3rd Qu.:82.00
## Max. :2639.0                                     Max. :98.00
##
## EDUC      SES      MMSE      CDR      eTIV
## Min. : 6.0      Min. :1.00      Min. : 4.00      Min. :0.0000      Min. :1106
## 1st Qu.:12.0      1st Qu.:2.00      1st Qu.:27.00      1st Qu.:0.0000      1st Qu.:1357
## Median :15.0      Median :2.00      Median :29.00      Median :0.0000      Median :1470
## Mean :14.6      Mean :2.46      Mean :27.34      Mean :0.2909      Mean :1488
## 3rd Qu.:16.0      3rd Qu.:3.00      3rd Qu.:30.00      3rd Qu.:0.5000      3rd Qu.:1597
## Max. :23.0      Max. :5.00      Max. :30.00      Max. :2.0000      Max. :2004
## NA's :19      NA's :2
## nWBV      ASF
## Min. :0.6440      Min. :0.876
## 1st Qu.:0.7000      1st Qu.:1.099
## Median :0.7290      Median :1.194
## Mean :0.7296      Mean :1.195
## 3rd Qu.:0.7560      3rd Qu.:1.293
## Max. :0.8370      Max. :1.587
##
```

```
# Count missing values
colSums(is.na(dementia_data))
```

```
## Subject.ID      MRI.ID      Group      Visit      MR.Delay      M.F      Hand
```

```
##      0      0      0      0      0      0      0
##    Age    EDUC    SES    MMSE    CDR    eTIV    nWBV
##      0      0     19      2      0      0      0
##    ASF
##      0
```

```
# Mean, median, and standard deviation of Age, MMSE, and eTIV
```

```
mean(dementia_data$Age, na.rm = TRUE)
```

```
## [1] 77.0134
```

```
median(dementia_data$MMSE, na.rm = TRUE)
```

```
## [1] 29
```

```
sd(dementia_data$eTIV, na.rm = TRUE)
```

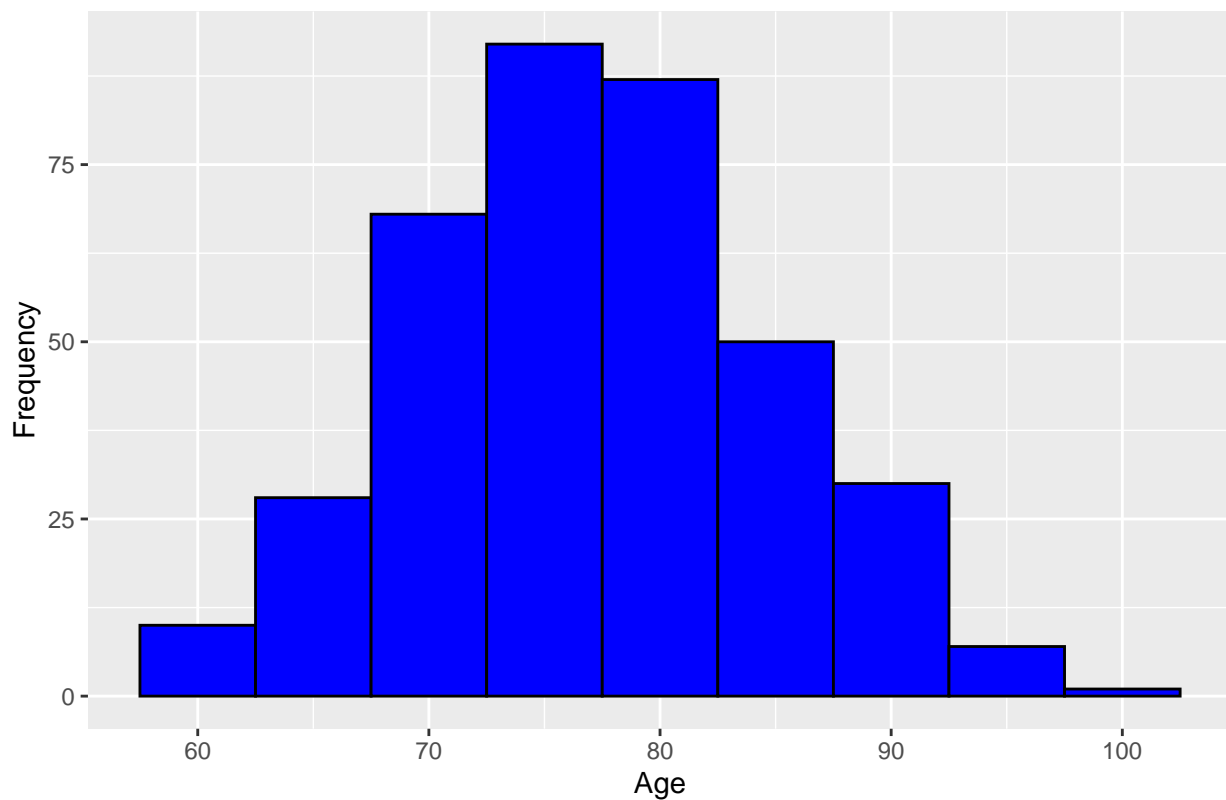
```
## [1] 176.1393
```

## Visualizations

```
# Age histogram using ggplot2
```

```
ggplot(dementia_data, aes(x = Age)) +  
  geom_histogram(binwidth = 5, fill = "blue", color = "black") +  
  labs(title = "Histogram of Age", x = "Age", y = "Frequency")
```

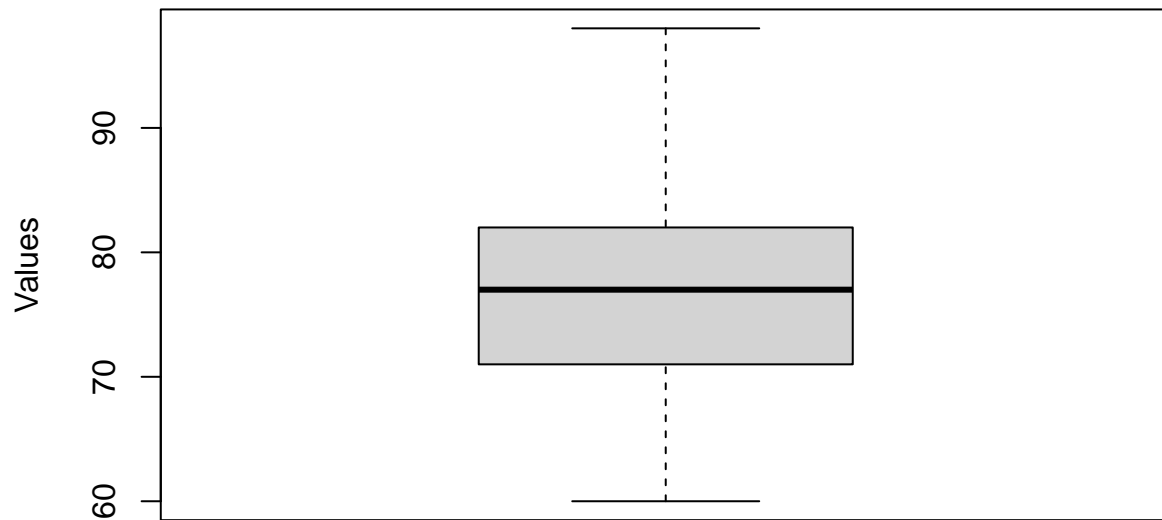
Histogram of Age



```
# Boxplot for Age
```

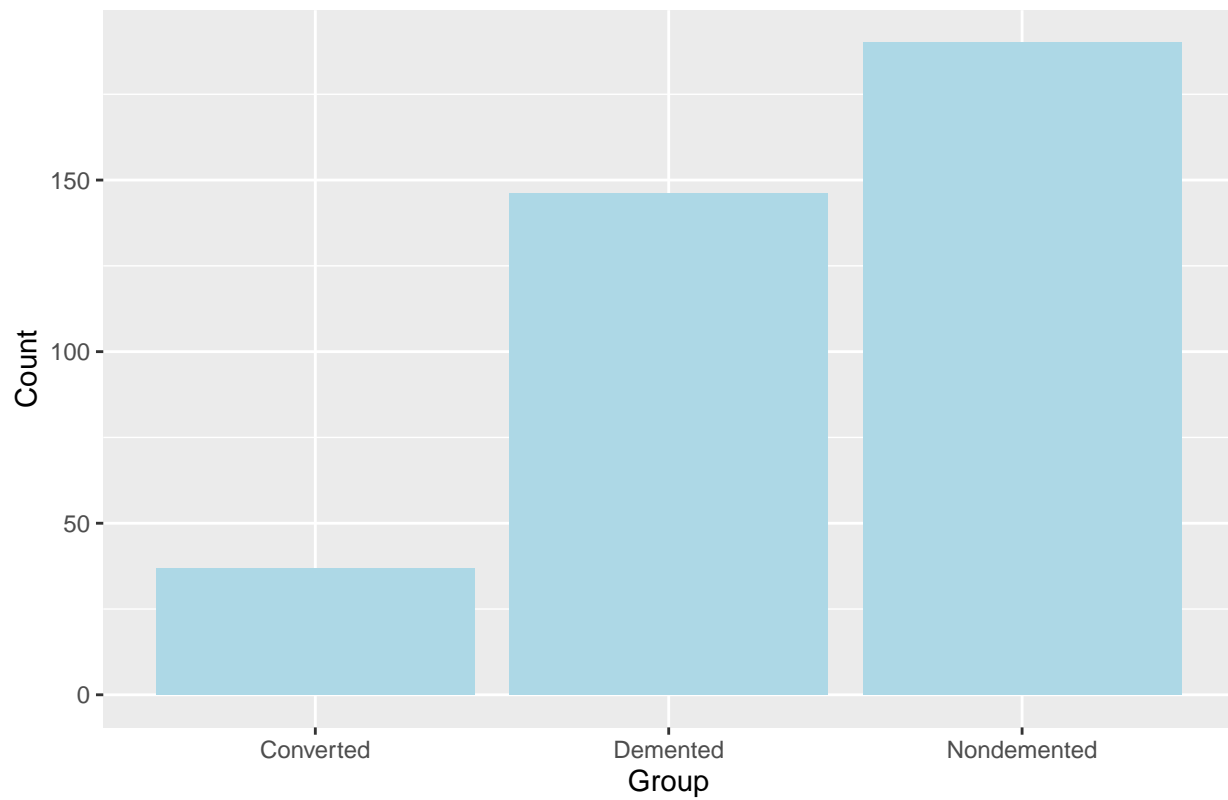
```
boxplot(dementia_data$Age, main = "Boxplot of Age", ylab = "Values")
```

Boxplot of Age



```
# Bar plot for categorical variable 'Group'
ggplot(dementia_data, aes(x = Group)) +
  geom_bar(fill = "lightblue") +
  labs(title = "Group Distribution", x = "Group", y = "Count")
```

Group Distribution



## Group Summarization

```
# Summary statistics by Group
dementia_data %>%
  group_by(Group) %>%
  summarize(
    count = n(),
    mean_age = mean(Age, na.rm = TRUE),
    sd_age = sd(Age, na.rm = TRUE)
  )
```

```
## # A tibble: 3 x 4
##   Group      count mean_age sd_age
##   <chr>    <int>   <dbl> <dbl>
## 1 Converted      37    79.8   7.43
## 2 Demented     146    76.3   6.94
## 3 Nondemented  190    77.1   8.10
```