**Universidade Cruzeiro do Sul**

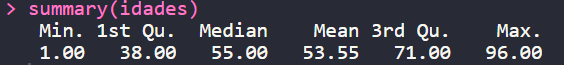
**Nome:** José Cleison de Lima

**RGM:** 25177591

**Atividade II – Projeto integrador transdisciplinar em ciência de dados II**

Chart, bar chart

Description automatically generated



**Observações:**

Os dados fornecidos estava em um documento word e com separação por vírgulas. Dessa forma precisei fazer uma limpeza dos dados utilizando o excel colocando como uma tabela simples, e por fim salvando com CSV.

Com esse processo ficou mais fácil a manipulação dos dados utilizando o Rstudio.

Foi necessário transformar os dados para numérico dentro da ferramenta. Não houve dados núlos com isso entre outros problemas e poucas falhas devido aos ajustes antecipados fora do ambiente.

Realizei apenas o comando Sumary (idade) no qual me revelou os dados estatísticos necessários para essa análise.

No gráfico acima conseguimos observar que não existem dados discrepantes mas existe uma assimetria notável a esquerda, assim revelando o tempo de falha.

O valor Mínimo é 1, 1º quartil 38, a mediana 55, a média 55.55, 3º quartil 71 e por fim o valor máximo é 96.

Com isso da para saber os valores que mais se apresentam no conjunto de dados.

O dados se concentram entre 38 e 71 tendo maior intensidade no centro entre a média e a mediana.

A mediana se encontra bem no centro da caixa.

Utilizei o reaproveitamento de códigos anteriores para adiantar o processo.

**Segue o código:**

# Instalação dos pacotes

install.packages("hrbrthemes")

# Execução dos pacontes

library(hrbrthemes)

library(tidyverse)

library(ggridges)

library(ggthemes)

library(cowplot)

library(viridis)

library(GGally)

library(dplyr)

library(ggplot2)

# Importação dos dados

dados *<-* read.csv("idades.csv", *sep* *=* ";")

# Visualização dos dados

View(dados)

# Verificando os tipos de dados

str(dados)

# Transformação dos dados para numéricos

*idades* *=* as.numeric(dados*$*idades)

str(idades)

any(is.na(idades)) # verifica sem tem qualquer valor núlo

summary(idades) # Verificação estatística dos dados

# Tema da visualização

*tema* *=* theme(*plot.background=*element\_rect(*fill=*"#E0FFFF"),

*plot.title* *=* element\_text(*size=*25, *hjust=*.5, *vjust=*1),

*axis.title.x* *=* element\_text(*size=*22),

*axis.title.y* *=* element\_text(*size=*22, *vjust=*2),

*axis.text.x* *=* element\_text(*size=*20),

*axis.text.y* *=* element\_text(*size=*20, *hjust=*2))

# Criação do boxplot

box2 *<-* ggplot(*idade* *=* idades, *mapping* *=* aes(*x* *=* idades)) *+*

              geom\_boxplot(*fill* *=* "#8B008B", *color* *=* "black", *size* *=* 1.2, *alpha* *=* .7) *+*

              stat\_boxplot(*geom=*"errorbar") *+*

              theme\_economist() *+*

              xlab("Idades") *+*

              ggtitle("Análise exploratória de Idades") *+*

              tema

# Grafico

plot\_grid(box2)