

LABORATÓRIO 1: Relação da concentração de glicose no plasma com características diversas

Fernando Bispo, Jeff Caponero

Sumário

Introdução	1
Objetivos	2
Sumarização dos dados	2
Histogramas das variáveis	4
Relação da variável resposta (diabetes) com as demais variáveis	6
Comparação do teste com as variáveis quantitativas	8
Conclusão	10
Apêndice	10
Código em linguagem R	10

Introdução

Este relatório visa a descrição da análise realizada nos dados oriundos do Instituto Nacional de Diabetes e de Doenças Digestivas e Renais dos EUA que conduziram um estudo com 768 mulheres da tribo Pina, que residem próximo a Phoenix e coletaram as seguintes características das participantes do estudo:

- Número de gestações [**pregnat**];
- Concentração de glicose no plasma (obtido duas horas depois da realização de um teste de tolerância a glicose) [**glucose**]; pressão sanguínea diastólica (mmHg) [**diastolic**];
- Largura do tríceps (mm) [**triceps**];
- Nível de insulina ($\mu\text{U}/\text{ml}$) [**insulin**];
- Índice de massa corpórea (kg/m^2) [**bmi**];
- Nível de função diabética [**diabetes**];
- Idade em anos [**age**];
- Teste para avaliação de sinais de diabetes (0 = negativo e 1 = positivo) [**teste**].

Ao analisar o arquivo contendo as medidas coletadas identificaram-se valores iguais a zero em situações em que o valor *zero* não poderia existir, não correspondendo à realidade da característica medida. Para avaliar a influência destes valores no comportamento geral dos dados serão realizadas duas análises, sendo estas sem a remoção dos valores *zero* e com a remoção destes e para uma maior

Table 1: Medidas Resumo dos dados sem a remoção dos valores zero

	Min	Q1	Med	Média	Q3	Max	D.Padrão	CV
Glicose	0,00	99,00	117,00	120,89	140,50	199,00	31,97	0,26
Idade	21,00	24,00	29,00	33,24	41,00	81,00	11,76	0,35
IMC	0,00	27,30	32,00	31,99	36,60	67,10	7,88	0,25
Largura Triceps	0,00	0,00	23,00	20,54	32,00	99,00	15,95	0,78
Nº de Gestações	0,00	1,00	3,00	3,85	6,00	17,00	3,37	0,88
Nível Diabético	0,08	0,24	0,37	0,47	0,63	2,42	0,33	0,70
Nível Insulina	0,00	0,00	30,50	79,80	127,50	846,00	115,24	1,44
P. Diastólica	0,00	62,00	72,00	69,11	80,00	122,00	19,36	0,28

Note:

Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais - EUA

percepção desta diferença (caso haja) serão construídas as tabelas e gráficos de análise lado a lado. Com a remoção dos dados não usuais (*zero*) houve uma redução de cerca de 51% no quantitativo de dados, resultando em 392 observações.

Objetivos

1. Sumarização dos dados do estudo.
2. Verificação de observações não usuais.
3. Relação da variável resposta (diabetes) com as demais variáveis.
4. Comparação do teste com as variáveis quantitativas e os níveis de glicose e de insulina, a pressão diastólica, o tríceps, o bmi, a idade e o nível de função diabética entre aqueles que apresentaram resultados do teste positivo e negativo.

Sumarização dos dados

Seguem as tabelas 1 e 2 com medidas de tendência central e variabilidade dos dados sem a remoção dos valores *zero* e com a remoção dos valores respectivamente, a fim de avaliar o impacto dos valores nas medidas calculadas.

Com o intuito de identificar a parcela dos participantes que foram identificados com sinais de diabetes, foi construída a Figura 1 sendo esta subdividida em Sub. Fig. A e B representando os dados sem a remoção dos valores *zero* e com a remoção dos valores respectivamente.

Table 2: Medidas Resumo dos dados com a remoção dos valores zero

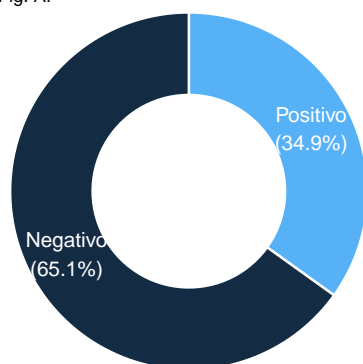
	Min	Q1	Med	Média	Q3	Max	D. Padrão	CV
Glicose	56,00	99,00	119,00	122,63	143,00	198,00	30,86	0,25
Idade	21,00	23,00	27,00	30,86	36,00	81,00	10,20	0,33
IMC	18,20	28,40	33,20	33,09	37,10	67,10	7,03	0,21
Largura Triceps	7,00	21,00	29,00	29,15	37,00	63,00	10,52	0,36
Nº de Gestações	0,00	1,00	2,00	3,30	5,00	17,00	3,21	0,97
Nível Diabético	0,09	0,27	0,45	0,52	0,69	2,42	0,35	0,66
Nível Insulina	14,00	76,50	125,50	156,06	190,00	846,00	118,84	0,76
P. Diastólica	24,00	62,00	70,00	70,66	78,00	110,00	12,50	0,18

Note:

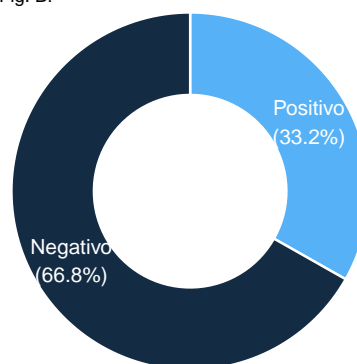
Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais - EUA

Figura 1: Resultado dos testes que avaliam sinais de diabetes realizados nas mulheres da Tribo Pina

Sud Fig. A:



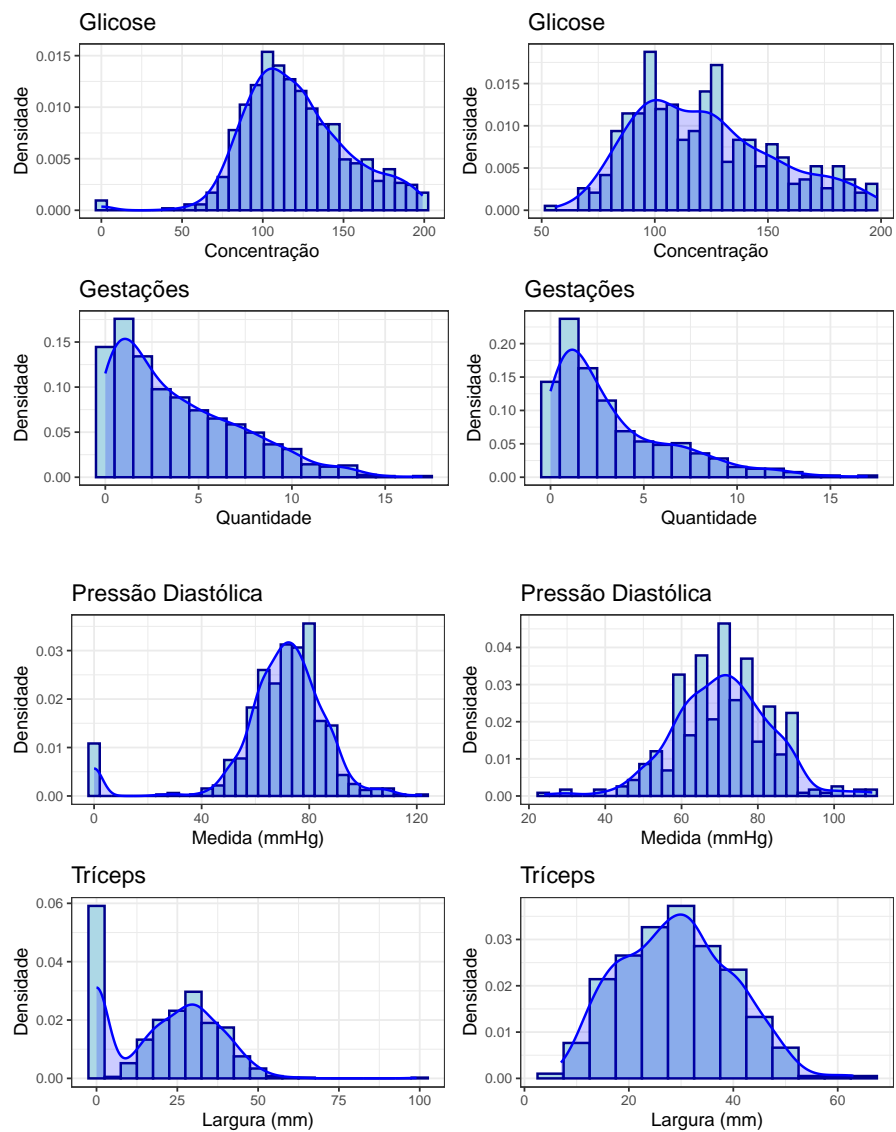
Sud Fig. B:

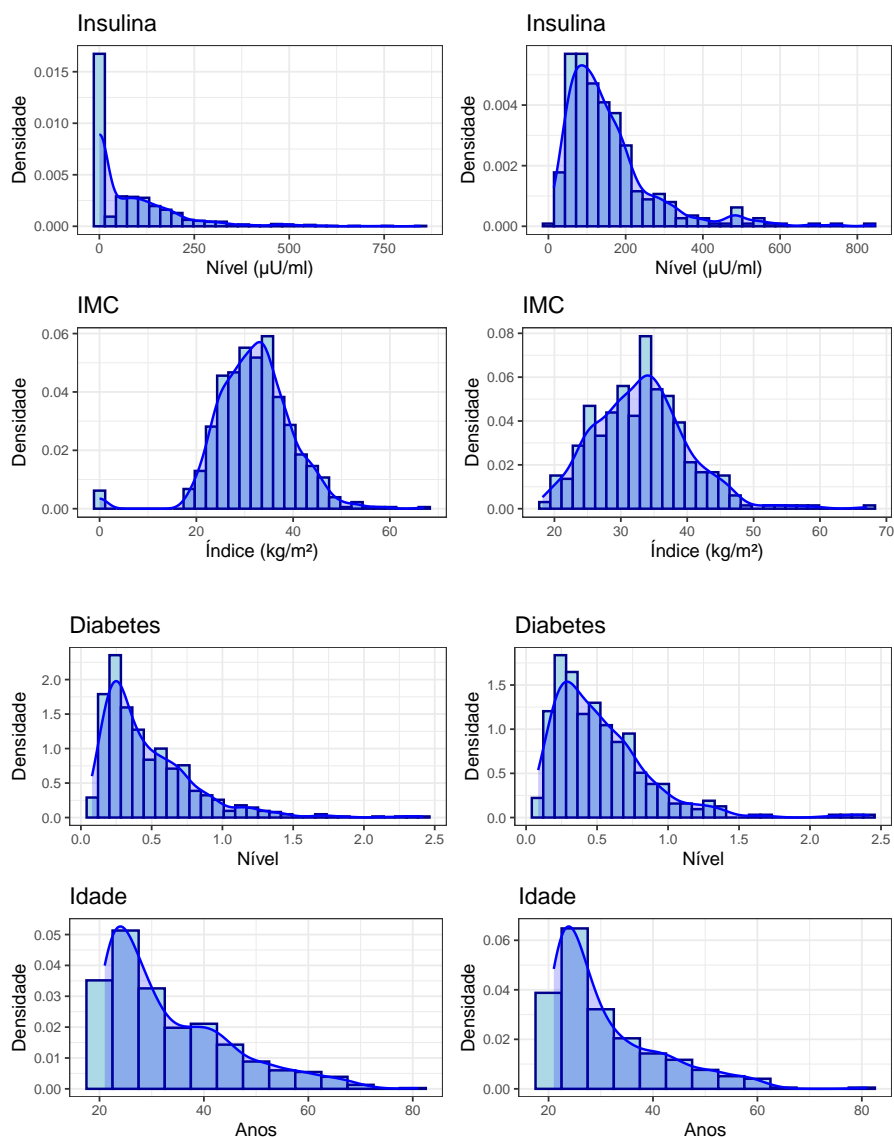


Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais – EUA

Histogramas das variáveis

Figura 2: Histogramas das variáveis em análise.





Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais – EUA

Observa-se que a presença de dados não usuais (*zeros*) prejudicou o comportamento de algumas variáveis, no que diz respeito a assimetria e curtose, como é o caso do nível de glicose no plasma que percebe-se um achatamento da distribuição (comportamento platocúrtico), a pressão diastólica, a largura do tríceps, a concentração de insulina (comportamento assimétrico a direita) e o índice de massa corpórea (comportamento assimétrico a direita). Fazendo-se necessário um tratamento dos dados analisados.

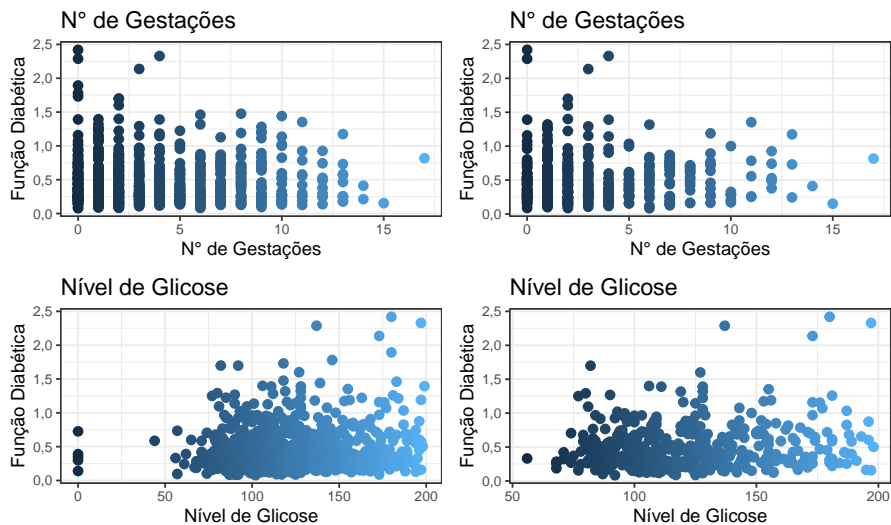
Embora a eliminação dos registros com dados faltantes tenha melhorado a visualização do comportamento de algumas variáveis, como é o caso do nível de glicose no plasma, a pressão diastólica, a largura do tríceps, a concentração de insulina e o índice de massa corpórea; nota-se uma perda de informação para as variáveis que estavam corretamente descritas anteriormente como é o caso do número de gestações. Nota-se que o número médio foi reduzido de 3,85 para 3,30. Semelhantemente, há perda de informação para as variáveis nível de função diabética, idade e teste de sinais de diabetes.

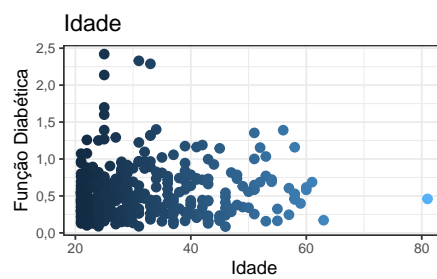
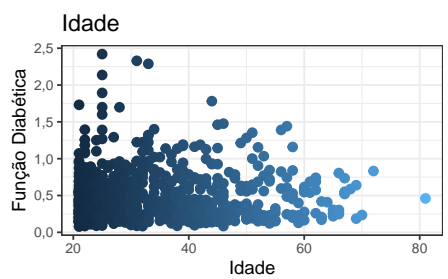
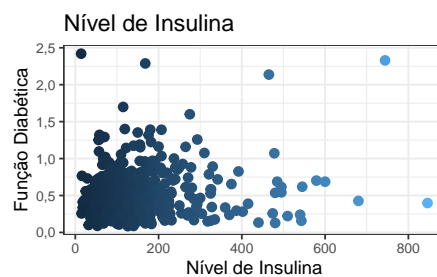
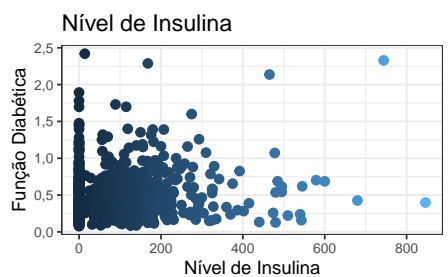
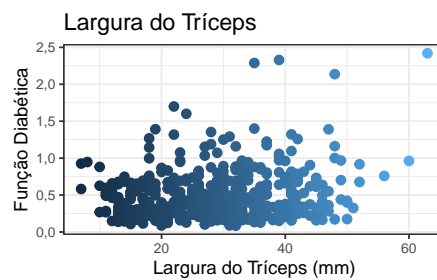
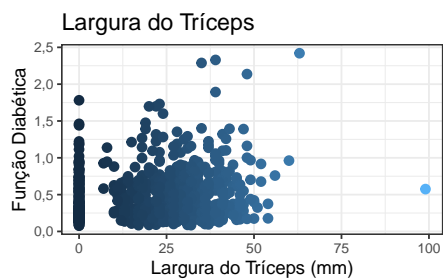
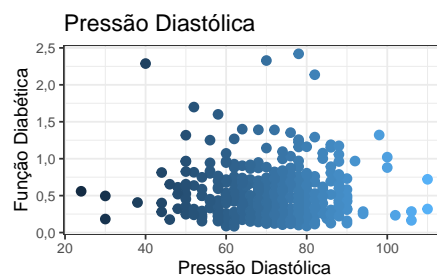
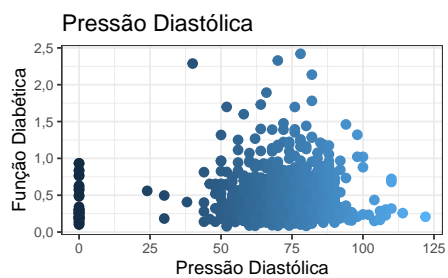
Observou-se ainda que algumas variáveis têm valores que parecem irrealistas, como o caso de largura de tríceps de 7mm ou nível de insulina de 849 $\mu\text{U}/\text{ml}$. Entretanto esta avaliação só pode ser feita por um especialista, isto é, alguém capaz de definir intervalo valores possíveis para cada variável.

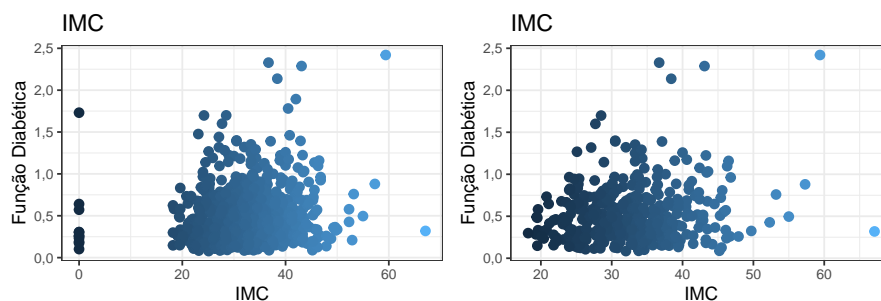
Relação da variável resposta (diabetes) com as demais variáveis

A fim de avaliar a correlação entre a variável de interesse (Nível de Diabetes) com as demais variáveis, foram construídos gráficos de dispersão.

Figura 3: Relação entre a Função Diabética e demais medições







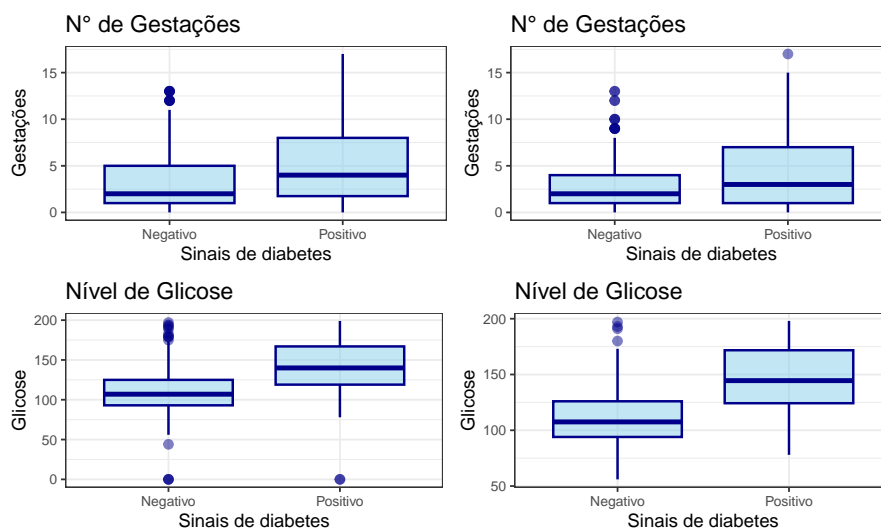
Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais – EUA

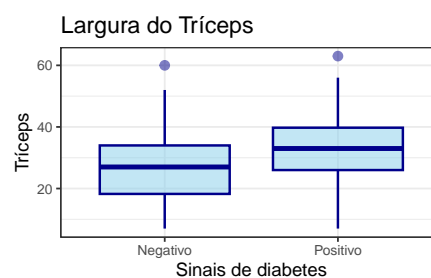
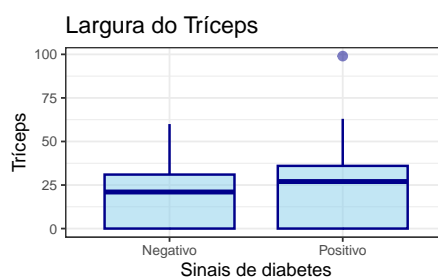
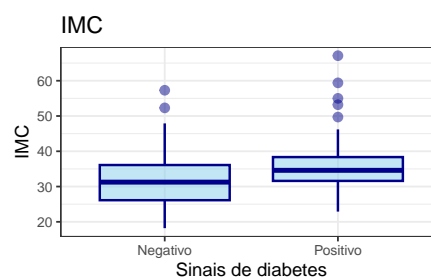
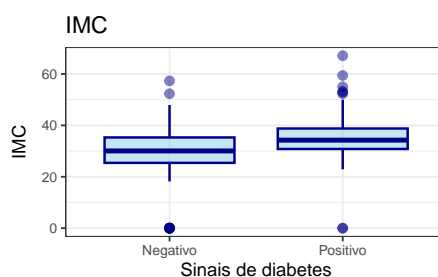
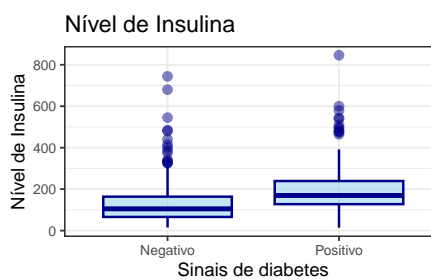
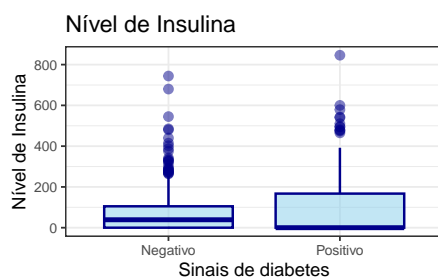
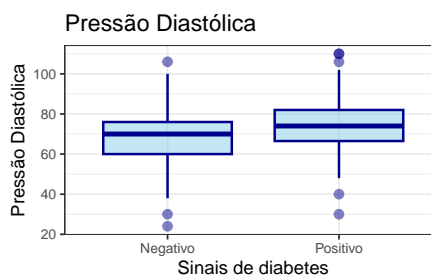
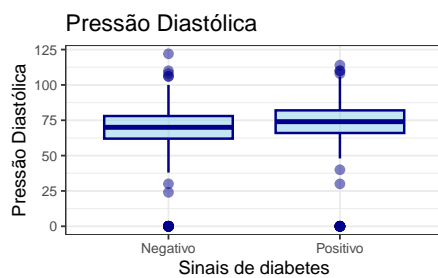
Após análise visual se constata que nenhuma das características analisadas indica ter qualquer correlação com o nível de função diabética.

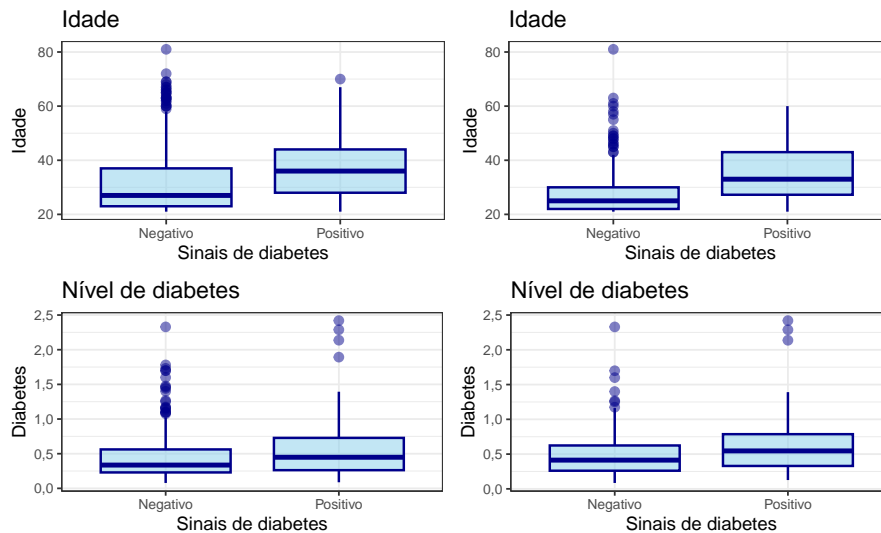
Comparação do teste com as variáveis quantitativas

É possível constatar, conforme exposto na Tabela 2 e gráficos a seguir que a retirada dos dados discrepantes e faltantes, representados por *zero*, influenciaram o comportamento dos dados.

Figura 3: BoxPlots entre o teste de nível diabético e demais variáveis em análise.







Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais – EUA

A análise dos BoxPlots permite uma análise um pouco mais informativa, uma vez que agora é possível verificar certa correlação entre o teste de sinais de diabetes e as demais variáveis. Nota-se que esta correlação não se apresenta em grau superlativo, mas em níveis mais fracos. Destacam-se os níveis de glicose no plasma e o nível de função diabética com uma maior correlação, mas ainda assim sujeita a grande variabilidade.

Conclusão

Com base nos resultados analisados, verifica-se que não parece ser suficiente a análise laboratorial na definição do estado de diabetes das mulheres da Tribo Pina. Infere-se que a análise clínica pode se valer das informações coletadas, mas seu julgamento final deve contar com outras técnicas diagnósticas.

Apêndice

Código em linguagem R

```
if (!require(pacman))
  install.packages("pacman")
library(pacman)

pacman::p_load(tidyverse, janitor, kableExtra, summarytools,
               moments, ggthemes, patchwork, glue, ggpubr)
```

```

# DADOS ----
dados <- read.csv2("Dados_Lab01.csv")

dados <- dados |>
  mutate(
    diabetes = as.numeric(diabetes),
    bmi = as.numeric(bmi)
  )

#####
###          Sumarização dos dados          ###
#####

dados|>
  select(-test)|>
  rename(
    "Nº de Gestações" = pregnant, "Glicose" = glucose, "Idade" = age,
    "P. Diastólica" = diastolic, "Largura Triceps" = triceps,
    "Nível Insulina" = insulin, "IMC" = bmi, "Nível Diabético" = diabetes)|>
  summarytools::descr(
    stats = c("min", "q1", "med", "mean", "q3", "max", "sd", "cv"),
    justify = "c",
    style = "grid",
    transpose = T
  ) |>
  kbl(
    caption = "Medidas Resumo dos dados sem a remoção dos valores zero",
    digits = 2,
    format.args=list(big.mark=".", decimal.mark=","),
    align = "c",
    row.names = T,
    col.names =
      c("Min", "Q1", "Med", "Média", "Q3", "Max", "D.Padrão", "CV")
  )|>
  footnote(general = "Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e R

# -----

p1 <- dados |>
  count(test) |>
  mutate(
    pct = round(prop.table(n)*100, 1),
    tipo = case_when(
      test == 0 ~ "Negativo",

```

```

      test == 1 ~ "Positivo"),
    labs = glue::glue('{tipo}\n({pct}%)') %>%
ggpubr::ggdonutchart(., "pct",
                      label = "labs", lab.pos = "in",
                      lab.font = c(3, "plain", "white"),
                      fill = "test", color = "white")

p2 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
         triceps>0, insulin>0, bmi>0)|>
  count(test) |>
  mutate(
    pct = round(prop.table(n)*100, 1),
    tipo = case_when(
      test == 0 ~ "Negativo",
      test == 1 ~ "Positivo"),
    labs = glue::glue('{tipo}\n({pct}%)') %>%
ggpubr::ggdonutchart(., "pct",
                      label = "labs", lab.pos = "in",
                      lab.font = c(3, "plain", "white"),
                      fill = "test", color = "white")

(p1+p2) + plot_annotation(
  title = "Figura 1: Resultado dos testes que avaliam sinais de diabetes realizados \nnas
  caption = "Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais - EUA
  tag_levels = c("A", "1"), tag_prefix = "Sud Fig. ", tag_sep = ".",
  tag_suffix = ":"
) & theme_void(base_size = 8) &
  theme(
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 3),
    legend.position = "none"
  )

# -----

g1 <- dados|>
  ggplot() +
    aes(x = glucose) +
  geom_histogram(
    aes(y = ..density..),
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(

```

```

    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
labs(
  title = "Glicose",
  x = "Concentração",
  y = "Densidade"
)

g1.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot() +
  aes(x = glucose) +
  geom_histogram(
    aes(y = after_stat(density)),
    # binwidth = 5,
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Glicose",
    x = "Concentração",
    y = "Densidade"
  )

g2 <- dados|>
  ggplot() +
  aes(x = pregnant) +
  geom_histogram(
    aes(y = ..density..),
    binwidth = 1,
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Gestações",

```

```

      x = "Quantidade",
      y = "Densidade"
    )

g2.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
         triceps>0, insulin>0, bmi>0)|>
  ggplot() +
    aes(x = pregnant) +
    geom_histogram(
      aes(y = ..density..),
      binwidth = 1,
      fill = "lightblue",
      colour = "darkblue") +
    geom_density(
      alpha = 0.2,
      fill = "blue",
      colour = "blue") +
    labs(
      title = "Gestações",
      x = "Quantidade",
      y = "Densidade"
    )

g3 <- dados|>
  ggplot() +
    aes(x = diastolic) +
    geom_histogram(
      aes(y = ..density..),
      fill = "lightblue",
      colour = "darkblue") +
    geom_density(
      alpha = 0.2,
      fill = "blue",
      colour = "blue") +
    labs(
      title = "Pressão Diastólica",
      x = "Medida (mmHg)",
      y = "Densidade"
    )

g3.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
         triceps>0, insulin>0, bmi>0)|>

```

```

ggplot() +
  aes(x = diastolic) +
  geom_histogram(
    aes(y = ..density..),
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Pressão Diastólica",
    x = "Medida (mmHg)",
    y = "Densidade"
  )

g4 <- dados|>
ggplot() +
  aes(x = triceps) +
  geom_histogram(
    aes(y = ..density..),
    binwidth = 5,
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Tríceps",
    x = "Largura (mm)",
    y = "Densidade"
  )

g4.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
ggplot() +
  aes(x = triceps) +
  geom_histogram(
    aes(y = ..density..),
    binwidth = 5,
    fill = "lightblue",
    colour = "darkblue") +

```



```

geom_density(
  alpha = 0.2,
  fill = "blue",
  colour = "blue") +
labs(
  title = "Tríceps",
  x = "Largura (mm)",
  y = "Densidade"
)

g5 <- dados|>
ggplot() +
aes(x = insulin) +
geom_histogram(
  aes(y = ..density..),
  fill = "lightblue",
  colour = "darkblue") +
geom_density(
  alpha = 0.2,
  fill = "blue",
  colour = "blue") +
labs(
  title = "Insulina",
  x = "Nível (µU/ml)",
  y = "Densidade"
)

g5.1 <- dados|>
filter(glucose>0, diabetes>0, diastolic>0,
  triceps>0, insulin>0, bmi>0)|>
ggplot() +
aes(x = insulin) +
geom_histogram(
  aes(y = ..density..),
  fill = "lightblue",
  colour = "darkblue") +
geom_density(
  alpha = 0.2,
  fill = "blue",
  colour = "blue") +
labs(
  title = "Insulina",
  x = "Nível (µU/ml)",
  y = "Densidade"
)

```

```

    )

g6 <- dados|>
  ggplot() +
  aes(x = bmi) +
  geom_histogram(
    aes(y = ..density..),
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "IMC",
    x = "Índice (kg/m²)",
    y = "Densidade"
  )

g6.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot() +
  aes(x = bmi) +
  geom_histogram(
    aes(y = ..density..),
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "IMC",
    x = "Índice (kg/m²)",
    y = "Densidade"
  )

g7 <- dados|>
  ggplot() +
  aes(x = diabetes) +
  geom_histogram(
    aes(y = ..density..),
    fill = "lightblue",

```

```

    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Diabetes",
    x = "Nível",
    y = "Densidade"
  )

g7.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot() +
  aes(x = diabetes) +
  geom_histogram(
    aes(y = ..density..),
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Diabetes",
    x = "Nível",
    y = "Densidade"
  )

g8 <- dados|>
  ggplot() +
  aes(x = age) +
  geom_histogram(
    aes(y = ..density..),
    binwidth = 5,
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Idade",

```

```

      x = "Anos",
      y = "Densidade"
    )

g8.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
         triceps>0, insulin>0, bmi>0)|>
  ggplot() +
  aes(x = age) +
  geom_histogram(
    aes(y = ..density..),
    binwidth = 5,
    fill = "lightblue",
    colour = "darkblue") +
  geom_density(
    alpha = 0.2,
    fill = "blue",
    colour = "blue") +
  labs(
    title = "Idade",
    x = "Anos",
    y = "Densidade"
  )

(g1+g1.1)/(g2+g2.1) + plot_annotation(
  title = "Figura 2: Histogramas das variáveis em análise.") &
  theme_bw(base_size = 8) &
  theme(
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

(g3+g3.1)/(g4+g4.1) + plot_annotation() &
  theme_bw(base_size = 8) &
  theme(
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

(g5+g5.1)/(g6+g6.1) + plot_annotation() &
  theme_bw(base_size = 8) &
  theme(
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

```

```

)

(g7+g7.1)/(g8+g8.1) + plot_annotation(
  caption = "Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais - EUA",
  theme_bw(base_size = 8) &
  theme(
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

#####
###      Diagramas de Dispersão      ###
#####

d1 <- dados |>
  ggplot(aes(
    y = diabetes,
    x = pregnant, color = pregnant)) +
  geom_point()+
  labs(
    title = 'Nº de Gestações',
    y = 'Função Diabética',
    x = 'Nº de Gestações'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d1.1 <- dados |>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot(aes(
    y = diabetes,
    x = pregnant, color = pregnant)) +
  geom_point()+
  labs(
    title = 'Nº de Gestações',
    y = 'Função Diabética',
    x = 'Nº de Gestações'
  )+
  scale_y_continuous(
    labels = scales::number_format(

```

```

        big.mark = ".",
        decimal.mark = ",",
    ))

d2 <- dados |>
  ggplot(aes(
    y = diabetes,
    x = glucose, color = glucose)) +
  geom_point()+
  labs(
    title = 'Nível de Glicose',
    x = 'Nível de Glicose',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    ))

d2.1 <- dados |>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot(aes(
    y = diabetes,
    x = glucose, color = glucose)) +
  geom_point()+
  labs(
    title = 'Nível de Glicose',
    x = 'Nível de Glicose',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    ))

d3 <- dados |>
  ggplot(aes(
    y = diabetes,
    x = diastolic, color = diastolic)) +
  geom_point()+
  labs(

```

```

    title = 'Pressão Diastólica',
    x = 'Pressão Diastólica',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d3.1 <- dados |>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot(aes(
    y = diabetes,
    x = diastolic, color = diastolic)) +
  geom_point()+
  labs(
    title = 'Pressão Diastólica',
    x = 'Pressão Diastólica',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d4 <- dados |>
  ggplot(aes(
    y = diabetes,
    x = triceps, color = triceps)) +
  geom_point()+
  labs(
    title = 'Largura do Tríceps',
    x = 'Largura do Tríceps (mm)',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

```

```

d4.1 <- dados |>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot(aes(
    y = diabetes,
    x = triceps, color = triceps)) +
  geom_point()+
  labs(
    title = 'Largura do Tríceps',
    x = 'Largura do Tríceps (mm)',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d5 <- dados |>
  ggplot(aes(
    y = diabetes,
    x = insulin, color = insulin)) +
  geom_point()+
  labs(
    title = 'Nível de Insulina',
    x = 'Nível de Insulina',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d5.1 <- dados |>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot(aes(
    y = diabetes,
    x = insulin, color = insulin)) +
  geom_point()+
  labs(
    title = 'Nível de Insulina',
    x = 'Nível de Insulina',

```



```

    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d6 <- dados |>
  ggplot(aes(
    y = diabetes,
    x = age, color = age)) +
  geom_point()+
  labs(
    title = 'Idade',
    x = 'Idade',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d6.1 <- dados |>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot(aes(
    y = diabetes,
    x = age, color = age)) +
  geom_point()+
  labs(
    title = 'Idade',
    x = 'Idade',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d7 <- dados |>
  ggplot(aes(

```

```

    y = diabetes,
    x = bmi, color = bmi)) +
  geom_point()+
  labs(
    title = 'IMC',
    x = 'IMC',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

d7.1 <- dados |>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  ggplot(aes(
    y = diabetes,
    x = bmi, color = bmi)) +
  geom_point()+
  labs(
    title = 'IMC',
    x = 'IMC',
    y = 'Função Diabética'
  )+
  scale_y_continuous(
    labels = scales::number_format(
      big.mark = ".",
      decimal.mark = ",",
    )
  )

(d1+d1.1)/(d2+d2.1) + plot_annotation(
  title = "Figura 3: Relação entre a Função Diabética e demais medições") &
  theme_bw(base_size = 8) &
  theme(
    legend.position = "none",
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

(d3+d3.1)/(d4+d4.1) + plot_annotation() &
  theme_bw(base_size = 8) &

```

```

theme(
  legend.position = "none",
  plot.tag.position = c(0, 1),
  plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
)

(d5+d5.1)/(d6+d6.1) + plot_annotation() &
  theme_bw(base_size = 8) &
  theme(
    legend.position = "none",
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

(d7+d7.1)/(plot_spacer() + plot_spacer()) +
  # plot_layout(nrow = 2, ncol = 2)
  plot_annotation(
    caption = "Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais - EUA",
    theme_bw(base_size = 8) &
    theme(
      legend.position = "none",
      plot.tag.position = c(0, 1),
      plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
    )
  )

#####
###          Boxplot dos Testes          ###
#####

b1 <- dados|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = pregnant)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'N° de Gestações',
    x = "Sinais de diabetes",
    y = "Gestações"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",

```

```

    decimal.mark = ","))

b1.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
         triceps>0, insulin>0, bmi>0)|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = pregnant)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'Nº de Gestações',
    x = "Sinais de diabetes",
    y = "Gestações"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ","))

b2 <- dados|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = glucose)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'Nível de Glicose',
    x = "Sinais de diabetes",
    y = "Glicose"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ","))

b2.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
         triceps>0, insulin>0, bmi>0)|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )

```

```

)|>
ggplot(aes(x = test, y = glucose)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(
  title = 'Nível de Glicose',
  x = "Sinais de diabetes",
  y = "Glicose"
) +
scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ",")
)

b3 <- dados|>
mutate(
  test = as_factor(test),
  test = lvls_revalue(test, c("Negativo", "Positivo"))
)|>
ggplot(aes(x = test, y = diastolic)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(
  title = 'Pressão Diastólica',
  x = "Sinais de diabetes",
  y = "Pressão Diastólica"
) +
scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ",")
)

b3.1 <- dados|>
filter(glucose>0, diabetes>0, diastolic>0,
  triceps>0, insulin>0, bmi>0)|>
mutate(
  test = as_factor(test),
  test = lvls_revalue(test, c("Negativo", "Positivo"))
)|>
ggplot(aes(x = test, y = diastolic)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(
  title = 'Pressão Diastólica',
  x = "Sinais de diabetes",
  y = "Pressão Diastólica"
) +

```

```

scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ","))

b4 <- dados|>
mutate(
  test = as_factor(test),
  test = lvls_revalue(test, c("Negativo", "Positivo"))
)|>
ggplot(aes(x = test, y = insulin)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(
  title = 'Nível de Insulina',
  x = "Sinais de diabetes",
  y = "Nível de Insulina"
) +
scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ","))

b4.1 <- dados|>
filter(glucose>0, diabetes>0, diastolic>0,
  triceps>0, insulin>0, bmi>0)|>
mutate(
  test = as_factor(test),
  test = lvls_revalue(test, c("Negativo", "Positivo"))
)|>
ggplot(aes(x = test, y = insulin)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(
  title = 'Nível de Insulina',
  x = "Sinais de diabetes",
  y = "Nível de Insulina"
) +
scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ","))

b5 <- dados|>
mutate(
  test = as_factor(test),

```

```

    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = bmi)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'IMC',
    x = "Sinais de diabetes",
    y = "IMC"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")
  )

b5.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = bmi)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'IMC',
    x = "Sinais de diabetes",
    y = "IMC"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")
  )

b6 <- dados|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = triceps)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'Largura do Triceps',
    x = "Sinais de diabetes",
    y = "Triceps"
  )

```

```

) +
scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ",")
)

b6.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,
    triceps>0, insulin>0, bmi>0)|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = triceps)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'Largura do Tríceps',
    x = "Sinais de diabetes",
    y = "Tríceps"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")
  )

b7 <- dados|>
  mutate(
    test = as_factor(test),
    test = lvls_revalue(test, c("Negativo", "Positivo"))
  )|>
  ggplot(aes(x = test, y = age)) +
  geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
  labs(
    title = 'Idade',
    x = "Sinais de diabetes",
    y = "Idade"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")
  )

b7.1 <- dados|>
  filter(glucose>0, diabetes>0, diastolic>0,

```



```

        triceps>0, insulin>0, bmi>0)|>
mutate(
  test = as_factor(test),
  test = lvls_revalue(test, c("Negativo", "Positivo"))
)|>
ggplot(aes(x = test, y = age)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(
  title = 'Idade',
  x = "Sinais de diabetes",
  y = "Idade"
) +
scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ",")
)

b8 <- dados|>
mutate(
  test = as_factor(test),
  test = lvls_revalue(test, c("Negativo", "Positivo"))
)|>
ggplot(aes(x = test, y = diabetes)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(
  title = 'Nível de diabetes',
  x = "Sinais de diabetes",
  y = "Diabetes"
) +
scale_y_continuous(
  labels = scales::number_format(
    dig.mark = ".",
    decimal.mark = ",")
)

b8.1 <- dados|>
filter(glucose>0, diabetes>0, diastolic>0,
        triceps>0, insulin>0, bmi>0)|>
mutate(
  test = as_factor(test),
  test = lvls_revalue(test, c("Negativo", "Positivo"))
)|>
ggplot(aes(x = test, y = diabetes)) +
geom_boxplot(col="darkblue", fill="skyblue", alpha = 0.5)+
labs(

```

```

    title = 'Nível de diabetes',
    x = "Sinais de diabetes",
    y = "Diabetes"
  ) +
  scale_y_continuous(
    labels = scales::number_format(
      dig.mark = ".",
      decimal.mark = ",")
  )

(b1+b1.1)/(b2+b2.1) + plot_annotation(
  title = "Figura 3: BoxPlots entre o teste de nível diabético e demais variáveis em anál
  theme_bw(base_size = 8) &
  theme(
    legend.position = "none",
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

(b3+b3.1)/(b4+b4.1) + plot_annotation() &
  theme_bw(base_size = 8) &
  theme(
    legend.position = "none",
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

(b5+b5.1)/(b6+b6.1) + plot_annotation() &
  theme_bw(base_size = 8) &
  theme(
    legend.position = "none",
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0)
  )

(b7+b7.1)/(b8+b8.1) + plot_annotation(
  caption = "Fonte: Instituto Nacional de Diabetes e de Doenças Digestivas e Renais - EUA
  theme_bw(base_size = 8) &
  theme(
    legend.position = "none",
    plot.tag.position = c(0, 1),
    plot.tag = element_text(size = 8, hjust = 0, vjust = 0))
# -----

```