



## ROTEIRO 4

### Aula 04 – Tipos de dados, tipos de variáveis, características

#### 1. Objetivo

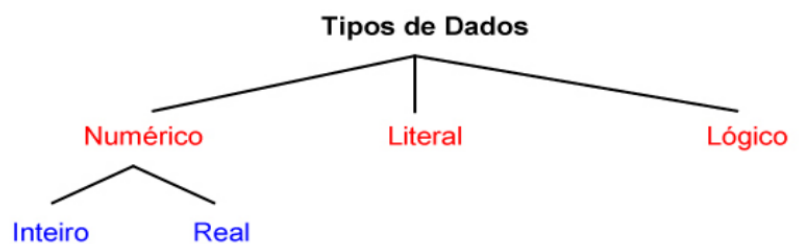
Trabalhar com a base de dados **airquality** no RStudio, de forma a consolidar conceitos relacionados aos tipos de dados, tipos de variáveis e suas características.

#### 2. Expectativa

Apresentar os passos básicos para a análise da base de dados **airquality** no R/Rstudio, para que o aluno possa construir conceitos acerca de tipos de dados, tipos de variáveis e suas características, de modo a ser possível sua aplicação em qualquer contexto.

#### 3. Teoria

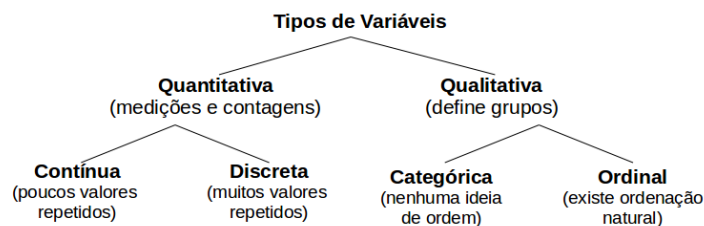
##### 1. Tipos de Dados



Responda:

O que cada tipo de dados armazena? Dê um exemplo para cada tipo de dados apresentado.

##### 2. Tipos de Variáveis



Responda:

Aproveitando a breve descrição que acompanha cada tipo de variável, dê um exemplo para cada uma delas. Como você acha que seria interessante resumir cada um deles?

### **3. Apresentação/ Representação de dados**

- O papel dos números
  - Transmitir informação
  - Tornar o dado disponível ao leitor para verificação detalhada, análise, ou ambos.
- Tabelas – Tipos de tabelas
  - Séries de tempo
  - Tabelas cruzadas  $n \times n$  – contagens
  - Tabelas de frequências – intervalar
- Gráficos
  - Gráficos de pontos
  - Gráficos de Ramo-e-Folhas
  - Gráficos de linhas
  - Gráficos de setores em círculos (pizza)
  - Gráficos de barras
  - Gráficos de área/superfícies
  - Histogramas
  - Boxplot
  - Gráficos multivariados (faces de chernoff, diagrama de pareto, múltiplos grupos, etc.)
- Resumos numéricos de variáveis contínuas e distribuições
  - Média
  - Mediana
  - Moda
  - Resumo de cinco números (Min,  $Q_1$ , Mediana,  $Q_3$ , Max)
  - Medidas de dispersão (Amplitude, amplitude interquartil, desvio padrão, coeficiente de variação, etc.)
  - Medidas de assimetria e curtose

#### Tarefa:

Aproveitando a breve descrição que acompanha cada elemento, faça uma busca simples na internet para a ilustração de cada um destes pontos.

### **4. Como resumir variáveis qualitativas e quantitativas em grandes bases de dados?**

- Histórico não muito longínquo
- Big Data
- Screening
- Estratégias para análise de grandes bases de dados
- Pictogramas gráficos
- Dashboards
- Storytelling

#### Tarefa:

Aproveitando o tempo em aula, faça uma busca simples na internet para a ilustração de cada um destes pontos.

#### 4. Passos

Execute os comandos abaixo em R ou RStudio (ou o equivalente em outro software de sua preferência) e observe os seus resultados. Aproveite a sequência obtida para criar uma análise resumida desta base de dados.

```
require(ggplot2) # ← verificar o carregamento deste package
require(moments) # ← verificar o carregamento deste package
aq <- data.frame(airquality)
colnames(aq)
head(aq)
colSums(is.na(aq))
aq$Day <- factor(aq$Day, levels=c(1:31), ordered=TRUE)
aq$Month <- factor(aq$Month, levels=5:9, labels=month.abb[5:9], ordered=TRUE)
summary(aq)
summary(aq$Ozone)
ggplot(data=aq) + geom_histogram(mapping=aes(Ozone))
skewness(aq$Ozone, na.rm = T)
kurtosis(aq$Ozone, na.rm=T)
aq$Ozone.sqrt <- sqrt(aq$Ozone)
ggplot(data=aq) + geom_histogram(mapping=aes(Ozone.sqrt))
skewness(aq$Ozone.sqrt, na.rm=T)
kurtosis(aq$Ozone.sqrt, na.rm=T)
qplot(Month, Ozone, data=aq, geom="boxplot", color=Month)
summary(aq$Solar.R)
ggplot(data=aq) + geom_histogram(mapping=aes(Solar.R))
qplot(Month, Solar.R, data=aq, geom="boxplot", color=Month)
summary(aq$Wind)
ggplot(data=aq) + geom_histogram(mapping=aes(Wind))
skewness(aq$Wind)
kurtosis(aq$Wind)
qplot(Month, Wind, data=aq, geom="boxplot", color=Month)
summary(aq$Temp)
qplot(Month, Temp, data=aq, geom="boxplot", color=Month)
table(aq$Month)
table(aq$Day)
columns <- c('Ozone.sqrt', 'Solar.R', 'Wind', 'Temp')
rows <- rowSums(is.na(aq)) == 0
round(cor(aq[rows, columns]), 2)
qplot(Solar.R, Wind, data=aq, geom="point", color=Month)
qplot(Temp, Solar.R, data=aq, geom="point", color=Month)
qplot(Solar.R, Ozone.sqrt, data=aq, color=Month, geom="point")
qplot(Temp, Wind, data=aq, color=Month, geom="point")
qplot(Wind, Ozone.sqrt, data=aq, color=Month, geom='point')
qplot(Temp, Ozone.sqrt, data=aq, color=Month, geom='point')
```

**Sugestão:** Copie estes comandos em um arquivo texto e abra este arquivo dentro do RStudio para executar passo a passo. Observe os comandos que por ventura dêem erro em sua execução.

#### 5. Referências

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983). *Graphical Methods for Data Analysis*. Belmont, CA: Wadsworth.