



Vector



- 1 column or row of data
- 1 type (numeric or text)

Matrix



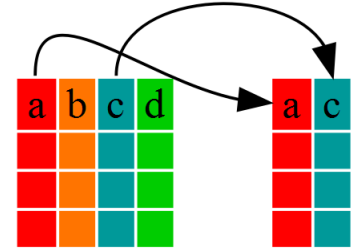
- multiple columns and/or rows of data
- 1 type (numeric or text)

Data Frame

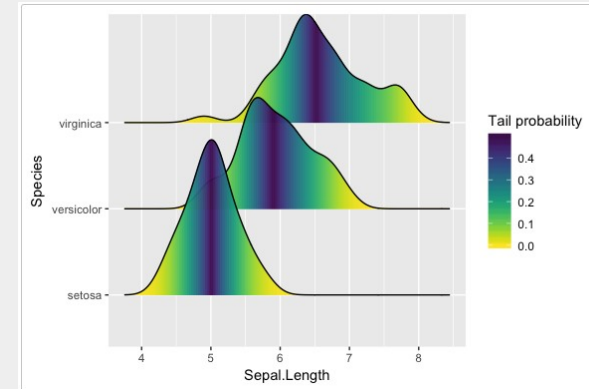
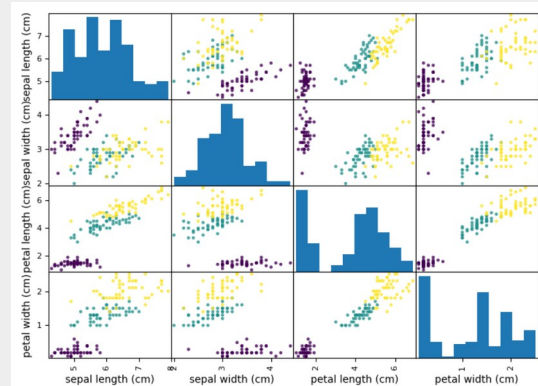
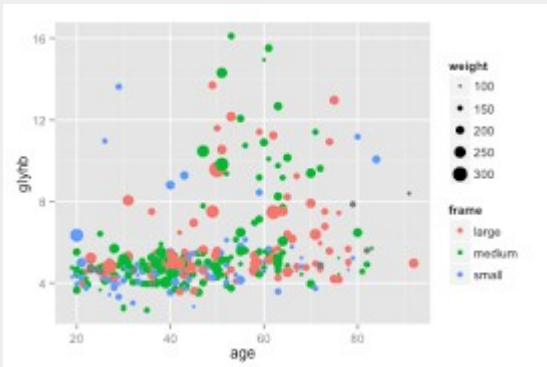


- multiple columns and/or rows of data
- multiple types

`select(data.frame,a,c)`



Um guia para iniciantes em R



Um guia para iniciantes em R



Análise básicas

```
sexo <- c("Masculino", "Feminino", "Feminino", "Masculino", "Feminino")
```

```
idade <- c(40, 37, 17, 13, 10)
```

```
estado.civil <- c("casado", "casado", "solteiro", "solteiro", "solteiro")
```

```
b <- data.frame(sexo, idade, estado.civil)
```

```
summary(b$estado.civil)
```

Um guia para iniciantes em R

Normalizando strings – Converter para maiúsculo ou minúsculo

```
df      <- read.csv2("r_iniciantes_csv_05.csv", sep=";", stringsAsFactors=FALSE)
```

```
summary(df)
```

```
df$uf      <- tolower(df$uf)
```

```
df$candidato <- tolower(df$candidato)
```

```
df$partido  <- tolower(df$partido)
```

```
df$sexo     <- tolower(df$sexo)
```

```
df$resultado <- df$resultado(df$resultado)
```



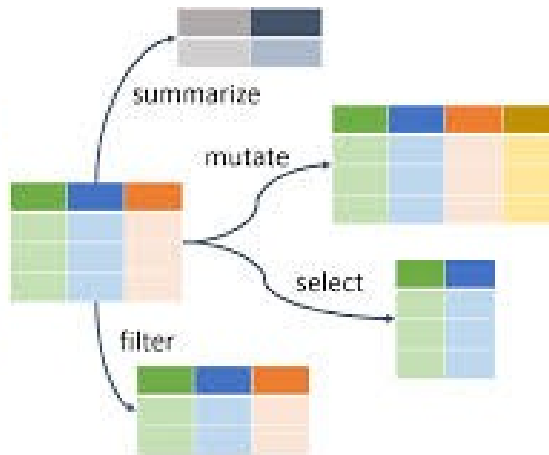
Um guia para iniciantes em R

Normalizando strings – Eliminando os espaços desnecessários

```
df$uf      <- trimws(df$uf)
df$candidato <- trimws(df$candidato)
df$partido  <- trimws(df$partido)
df$sexo     <- trimws(df$sexo)
df$resultado <- trimws(df$resultado)
```



Um guia para iniciantes em R



Um guia para iniciantes em R

Normalizando strings – dplyr e stringr

```
df <- df %>% mutate_if(is.character, tolower)
```

```
df <- df %>% mutate_if(is.character, str_trim)
```



Um guia para iniciantes em R

Normalizando strings – dplyr e stringr

```
df <- df %>% mutate_if(is.character, tolower)
```

```
df <- df %>% mutate_if(is.character, str_trim)
```



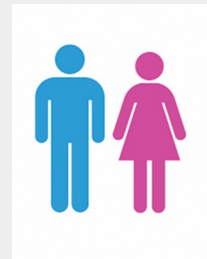
Um guia para iniciantes em R

Variáveis Qualitativas (categóricas)

- *nominais*
- *ordinais*

Variáveis Quantitativas

- *discretas*
- *contínuas*



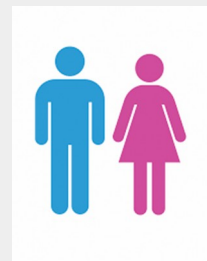
Um guia para iniciantes em R

```
df$partido <- as.factor(df$partido)
```

```
df$sexo <- as.factor(df$sexo)
```

```
df$resultado <- as.factor(df$resultado)
```

```
summary(df)
```



Um guia para iniciantes em R

Análise descritiva - descrição univariada

```
tab_resultado_abs <- table(df$resultado)
tab_resultado_abs
pie(tab_resultado_abs,col=blues9)
barplot(tab_resultado_abs, beside=T, leg=T , col=blues9)
```

```
tab_resultado_rel <- prop.table(tab_resultado_abs)
tab_resultado_rel
pie(tab_resultado_rel)
barplot(tab_resultado_rel, beside=T, leg=T)
```



Um guia para iniciantes em R

Análise descritiva - descrição univariada

```
tab_resultado_cruzado <- table(df$partido, df$sexo)  
tab_resultado_cruzado
```

```
tab_resultado_cruzado_rel <- prop.table(tab_resultado_cruzado)  
tab_resultado_cruzado_rel
```

```
barplot(tab_resultado_cruzado_rel , beside=T, leg=T, col=blues9)
```



Um guia para iniciantes em R

Análise descritiva – variáveis numéricas

Em algumas situações, como quando estiver trabalhando com variáveis numéricas, será necessário contar o número de observações que estão dentro de um intervalo.

Para isso, terá que categorizar a variável em questão e depois colocar os valores em uma tabela de frequência.

Para categorizar uma variável contínua no R, você pode usar o comando `cut`, os argumentos dessa função são as observações, um argumento chamado `breaks`, em que se armazenam os limites dos intervalos e o argumento `labels`, que são os nomes que os elementos receberão.



Um guia para iniciantes em R

Tabelas de Frequência para variáveis contínuas

```
factors_logvotos <- cut(df$logvotos, breaks = c(10,12,14), labels = c("10-12", "12-14"))
```

```
table_freq_votos <- table(factors_logvotos)  
table_freq_votos  
barplot(table_freq_votos,col="powderblue")
```

```
table_freq_votos <- table (factors_logvotos, df$resultado)  
table_freq_votos
```



Um guia para iniciantes em R

Estatísticas descritivas

```
summary(df$logvotos)
```

```
mean(df$logvotos)
```

```
sd(df$logvotos)
```

```
median(df$logvotos)
```

```
boxplot(df$logvotos, main = "Log vostos", col = "darksalmon")
```

```
hist(df$logvotos , main = "Log vostos", ylab = "Frequência", xlab = "logvostos", col = "beige")
```



Um guia para iniciantes em R

Média, Desvio Padrão e Mediana

```
factors_logvotos <- cut(df$logvotos, breaks = c(10,12,14), labels = c("10-12","12-14"))
```

```
table_freq_votos <- table(factors_logvotos)
table_freq_votos
barplot(table_freq_votos,col="powderblue")
```

```
table_freq_votos <- table (factors_logvotos, df$resultado)
table_freq_votos
```

