R Notebook

### Carregar todas os pacotes que serão utilizadas:

library(dplyr)  
library(skimr)  
library(base)  
library(utils)  
library(readr)

### Carregar todos os arquivos de cada ano separado e concatená-los em um único Data Frame:

disponibilidade <- list.files(path = "C:\\Users\\Sanderson\\Documents\\RStudio\\Projeto\\disponibilidade",pattern = "\*.csv", full.names = TRUE ) %>%   
 lapply(read.csv2) %>%   
 bind\_rows()

### Analisar tipo do dado de cada coluna:

str(disponibilidade)

## 'data.frame': 135 obs. of 4 variables:  
## $ dat\_referencia: chr "2011-01-01" "2011-02-01" "2011-03-01" "2011-04-01" ...  
## $ val\_dispf : chr "91.7105" "89.9401" "89.3349" "89.9785" ...  
## $ val\_indisppf : chr "7.0246" "8.2315" "8.4003" "8.374" ...  
## $ val\_indispff : chr "1.2647" "1.8282" "2.2647" "1.6474" ...

### Converter tipo de dado das colunas de acordo com o é fornecido pelo dicionário:

* **dat\_referencia** - char **para** date
* **val\_dispf** - char **para** numeric
* **val\_indisppf** - char **para** numeric
* **val\_indispff** - char **para** numeric

disponibilidade$dat\_referencia <- as.Date(disponibilidade$dat\_referencia)  
disponibilidade$val\_dispf <- as.numeric(disponibilidade$val\_dispf)  
disponibilidade$val\_indisppf <- as.numeric(disponibilidade$val\_indisppf)  
disponibilidade$val\_indispff <- as.numeric(disponibilidade$val\_indispff)

### Analisar o comportamento dos dados e verificar a existência valores em branco:

skim\_without\_charts(disponibilidade)

Data summary

|  |  |
| --- | --- |
| Name | disponibilidade |
| Number of rows | 135 |
| Number of columns | 4 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| Date | 1 |
| numeric | 3 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: Date**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| dat\_referencia | 0 | 1 | 2011-01-01 | 2022-03-01 | 2016-08-01 | 135 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| val\_dispf | 0 | 1 | 86.95 | 2.68 | 78.58 | 84.89 | 87.62 | 89.06 | 91.82 |
| val\_indisppf | 0 | 1 | 9.56 | 2.18 | 5.10 | 8.02 | 9.28 | 10.56 | 17.42 |
| val\_indispff | 0 | 1 | 3.48 | 1.20 | 1.26 | 2.51 | 3.21 | 4.57 | 6.27 |

### Exportar o CSV:

write\_csv2(disponibilidade, file = "disponibilidade.csv")