

PEC4: Final assessment

Marco Russo - Silvia Gamundi Sumando

Enero, 2025

Contents

1	1. Contexto y objetivo del estudio. Datos (1 punto)	3
2	2. Prospección y preparación de los datos (2 puntos)	3
2.1	2.1 Descripción de los datos (1 punto)	3
2.2	2.2 Preguntas “objetivo” (1 punto)	34
3	3. Análisis exploratorio de los datos (2,5 puntos)	34
3.1	3.1 Análisis descriptivo y gráfico (1 punto)	34
3.2	3.2. Ejercicios de inferencia y simulación (1,5 puntos)	34
4	4. Modelos de aprendizaje automático (2,5 puntos)	34
5	5. Visualización (1,5 puntos)	34
6	6. Conclusiones (0,5 puntos)	34
7	Bibliografía	35

Información del Estudiante

Nombre	Marco Russo
Email	mrussorb@uoc.edu
GitHub	https://github.com/marcusRB/uoc-ub-scientific-programming
LinkedIn	https://www.linkedin.com/in/marcusrb/
Fecha	January 6, 2026

Nombre	Silvia Gamundi Sumando
Email	sgamundis@uoc.edu
GitHub	https://github.com/
LinkedIn	https://www.linkedin.com/in/
Fecha	January 6, 2026

1 1. Contexto y objetivo del estudio. Datos (1 punto)

El dataset elegido es **Bacteremia** del autor Heinze, G. (2023). Bacteremia [Data set]. In PLoS One (Version S2, Vol. 9, Number 9, p. e106765). Zenodo. <https://doi.org/10.5281/zenodo.7554815> [?]

El resto de informaciones extraído del sitio oficial:

The data set consists of 14,691 observations from different patients with the clinical suspicion to suffer from bacteremia, for whom a blood culture analysis was performed at the Vienna General Hospital, Austria, between January 2006 and December 2010. It contains the results of the blood culture analysis for bacteremia and the values of 51 potential predictors of bacteremia. To protect data privacy our version of this data was slightly modified compared to the original version, and this modified version was cleared by the Medical University of Vienna for public use (DC 2019-0054). Details on the meaning of the variables can be found in the data dictionary. The original version of the data set was used by Ratzinger et al (2014) to develop a model for screening bacteremic patients based on highly standardizable laboratory variables. This public version has been used by Gregorich et al (2021).

Basada en la descripción oficial del mismo, se indican que existen 14,691 observaciones de diferentes pacientes que podrían ser afectados de **bacteriemia**. De la información disponible en Wikipedia:Bacteriemia, la bacteriemia es la presencia de bacterias en la sangre. La sangre es normalmente un medio estéril, por lo tanto la detección de bacterias es indicativa de infección.

Es importante entender este punto respecto al **diagnóstico**, muchas personas se recuperan completamente de la bacteriemia. Sin embargo, la bacteriemia es grave y puede provocar sepsis. Cuando tiene sepsis, el daño a los órganos principales puede ser irreversible.

Entre las **causas**, la entrada de bacterias en el torrente sanguíneo puede ser producto de una infección localizada (ej: neumonía, absceso en piel o mucosas), o por interrupción de la piel como barrera defensiva. Se destacan las intervenciones quirúrgicas, utilización de dispositivos invasivos (catéteres, sondas, asistencia mecánica respiratoria), heridas accidentales, o quemaduras.

La infección suele empezar en los pulmones, el tracto genitourinario, gastrointestinal o los tejidos blandos, entre ellos la piel de pacientes con úlceras. También puede ser secundaria a una intervención dental en pacientes de alto riesgo, especialmente en los que tienen prótesis intravasculares.

Respecto a las **consecuencias**, dependen del tipo de bacteria y el estado del paciente. La respuesta inmunológica a la infección puede causar sepsis y devenir en shock séptico. También puede ocurrir que la sangre transporte las bacterias a otros tejidos, que podrán ser infectados. Ejemplos incluyen endocarditis, osteomielitis, y meningitis. El tratamiento es fundamental para erradicar a las bacterias y requiere el uso de antibióticos por vía intravenosa.

Tenemos un pequeño diccionario disponible del significado de cada característica del dataset, lo descargaremos y visualizaremos para entender mejor el contexto.

2 2. Prospección y preparación de los datos (2 puntos)

2.1 2.1 Descripción de los datos (1 punto)

Preparamos el entorno cargando el resto de librerías que serán útiles para realizar un análisis exploratorio de los datos.

Descargamos el dataset en un formato dataframe y comprobaremos su estructura.

```
# Define the URL and file path
url <- "https://zenodo.org/records/7554815/files/Bacteremia_public_S2.csv?download=1"
filename <- "data/bacteremia_dataset.csv"

# Create directory if it doesn't exist
if (!dir.exists("data")) {
  dir.create("data", recursive = TRUE)

  # Download the file
  download.file(url, destfile = filename, mode = "wb")
} else {
  sprintf("The dataset has been downloaded and is available in 'data' path")
}
```

```
## [1] "The dataset has been downloaded and is available in 'data' path"
```

```
# Read the CSV file
bacteremia_df <- read.csv(file = filename)

# Preview the data
dplyr::glimpse(bacteremia_df)
```

```
## Rows: 14,691
## Columns: 53
## $ ID      <int> 1, 3, 5, 7, 9, 10, 11, 12, 13, 19, 21, 22, 23, 25, 26, 27~
## $ SEX     <int> 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 1, 2, 2, 2, 1, 2, 2, 2, ~
## $ AGE     <int> 62, 72, 46, 84, 38, 68, 55, 55, 67, 52, 47, 29, 59, 51, 4~
## $ MCV     <dbl> 99.3, 85.1, 96.3, 91.3, 85.1, 104.5, 99.3, 77.0, 95.3, 83~
## $ HGB     <dbl> 11.5, 10.3, 7.4, 10.3, 13.7, 15.7, 14.6, 10.8, 10.9, 10.3~
## $ HCT     <dbl> 35.9, 34.7, 22.8, 31.1, 38.7, 46.9, 43.5, 34.8, 30.4, 30.~
## $ PLT     <int> 307, 182, 64, 309, 183, 144, 242, 38, 88, 105, 216, 188, ~
## $ MCH     <dbl> 31.5, 26.0, 31.2, 30.4, 30.2, 34.8, 33.1, 23.8, 33.6, 28.~
## $ MCHC    <dbl> 31.8, 30.6, 32.4, 33.3, 35.3, 33.5, 33.4, 30.5, 35.3, 34.~
## $ RDW     <dbl> 19.5, 15.0, 19.7, 13.8, 12.6, 13.9, 13.1, 16.8, 13.3, 13.~
## $ MPV     <dbl> 10.8, 9.7, 11.1, 8.5, 10.0, 10.9, 10.3, NA, 10.7, 11.3, 1~
## $ LYM     <dbl> 0.4, 0.4, 1.5, 1.3, 0.8, 2.2, 2.1, 0.4, 0.4, 0.9, 0.7, 1.~
## $ MONO    <dbl> 1.7, 0.2, 1.2, 0.8, 0.4, 0.9, 1.6, 0.1, 0.2, 0.9, 0.6, 0.~
## $ EOS     <dbl> 0.0, 0.1, 0.1, 0.0, 0.0, 0.1, 0.3, 0.1, 0.0, 0.3, 0.0, 0.~
## $ BASO    <dbl> 0.1, 0.0, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.1, 0.1, 0.~
## $ NT      <int> 86, 90, 58, 67, 95, 61, NA, 93, 57, 69, 108, 86, 93, 83, ~
## $ APTT    <dbl> 28.8, 29.8, 36.3, 38.2, 33.1, 41.8, NA, 36.3, 33.8, 28.1,~
## $ FIB     <int> 578, NA, 313, 487, 490, 400, NA, 413, 431, 407, 604, 476,~
## $ SODIUM   <int> 137, 141, 147, 141, 137, 141, 139, 142, 143, 136, 131, 13~
## $ POTASS   <dbl> 3.88, NA, 4.61, 4.71, NA, 4.41, 3.69, 4.67, 2.35, 3.80, 5~
## $ CA      <dbl> 2.29, 2.21, 1.92, 2.05, 2.34, 2.08, NA, 2.31, 2.10, 1.92,~
## $ PHOS    <dbl> 1.20, 0.58, 1.51, 2.17, 0.97, 0.99, NA, 1.16, 0.51, 0.72,~
## $ MG      <dbl> 0.66, NA, 1.03, 0.83, 0.74, 0.56, NA, 0.87, 0.36, 0.53, 0~
## $ CREA    <dbl> 0.65, 0.76, 1.25, 2.78, 0.65, 0.82, 1.21, 1.77, 1.00, 0.5~
## $ BUN     <dbl> 5.7, 19.9, 50.6, 47.5, 8.5, 15.3, 13.0, 29.8, 15.0, 14.0,~
## $ HS      <dbl> 5.3, NA, NA, 9.7, 3.0, 5.5, NA, 6.2, 4.7, 4.0, 4.0, 4.1, ~
## $ GBIL    <dbl> 0.59, 0.48, 8.42, 0.35, 0.42, 2.40, 1.13, 0.45, 1.21, 2.4~
## $ TP      <dbl> 67.0, 65.3, 40.5, 61.2, 78.4, 57.5, NA, 70.8, 67.4, 53.8,~
## $ ALB     <dbl> 36.7, 37.4, 22.1, 33.2, 43.8, 30.1, NA, 43.6, 35.4, 24.8,~
```

```
## $ AMY      <int> 30, NA, 146, 92, 84, 95, 117, 177, NA, 35, 79, 16, 25, 32~
## $ PAMY     <int> 16, NA, NA, 28, 50, 57, NA, 43, NA, 35, 63, 14, 15, 20, 3~
## $ LIP      <int> 10, NA, 89, 18, 50, 25, 73, 30, NA, 38, 52, 19, 14, 26, 5~
## $ CHE      <dbl> 5.12, 5.61, 2.52, 4.10, 6.91, 6.79, NA, 7.40, NA, 2.64, 2~
## $ AP       <int> 85, 80, 119, 94, 108, 68, 51, 153, 239, 146, 180, 64, 74, ~
## $ ASAT     <int> 22, 28, 124, 774, 35, 32, 29, 26, 91, 97, 24, 13, 25, 31, ~
## $ ALAT     <int> 14, 25, 135, 72, 22, 11, 20, 32, 57, 156, 63, 23, 27, 53, ~
## $ GGT      <int> 48, 61, 134, 23, 72, 68, 138, 96, 446, 192, 266, 19, 66, ~
## $ LDH      <int> 284, NA, 696, 1787, NA, 263, 303, 181, 183, 277, 221, 299~
## $ CK       <int> 23, 36, 40, 2422, 79, 75, 230, 87, 53, 87, 30, 118, 17, 1~
## $ GLU      <int> 107, 84, 107, 105, 93, 89, 91, 96, 86, 104, 104, 102, 161~
## $ TRIG     <int> 105, NA, NA, 134, 152, 85, NA, 129, 62, 207, 292, 221, 12~
## $ CHOL     <int> 175, NA, NA, 141, 167, 144, NA, 156, 118, 123, 194, 151, ~
## $ CRP      <dbl> 3.94, 1.42, 12.09, 3.78, 11.17, 5.89, 17.84, 1.29, 1.36, ~
## $ BASOR    <dbl> 0.4132231, 0.0000000, 0.5681818, 0.0000000, 0.0000000, 0.~
## $ EOSR     <dbl> 0.0000000, 0.8264463, 0.5681818, 0.0000000, 0.0000000, 1.~
## $ LYMR     <dbl> 1.652893, 3.305785, 8.522727, 11.016949, 8.333333, 22.000~
## $ MONOR    <dbl> 7.024793, 1.652893, 6.818182, 6.779661, 4.166667, 9.00000~
## $ NEU      <dbl> 22.0, 11.4, 14.7, 9.7, 8.4, 6.8, 8.9, 1.2, NA, 3.8, 8.2, ~
## $ NEUR     <dbl> 90.90909, 94.21488, 83.52273, 82.20339, 87.50000, 68.0000~
## $ PDW      <dbl> 10.6, 11.4, 14.1, 8.7, 12.2, 12.9, 12.5, NA, NA, 13.2, 12~
## $ RBC      <dbl> 3.7, 3.9, 2.5, 3.5, 4.4, 4.3, 4.5, 4.7, NA, 3.5, 3.3, 2.5~
## $ WBC      <dbl> 24.10, 12.17, 17.45, 11.58, 9.86, 9.94, 13.06, 1.78, NA, ~
## $ BloodCulture <chr> "no", "no", "no", "no", "no", "no", "no", "no", "yes", "n~
```

Finalmente mostramos los primeros datos y la naturaleza de las características.

```
# Mostramos los primeros datos con head()
head(bacteremia_df, 10)
```

```
##      ID SEX AGE  MCV HGB  HCT PLT  MCH MCHC  RDW  MPV LYM MONO EOS BASO NT APTT
## 1     1   2  62  99.3 11.5 35.9 307 31.5 31.8 19.5 10.8 0.4  1.7 0.0  0.1 86 28.8
## 2     3   1  72  85.1 10.3 34.7 182 26.0 30.6 15.0  9.7 0.4  0.2 0.1  0.0 90 29.8
## 3     5   1  46  96.3  7.4 22.8  64 31.2 32.4 19.7 11.1 1.5  1.2 0.1  0.1 58 36.3
## 4     7   1  84  91.3 10.3 31.1 309 30.4 33.3 13.8  8.5 1.3  0.8 0.0  0.0 67 38.2
## 5     9   2  38  85.1 13.7 38.7 183 30.2 35.3 12.6 10.0 0.8  0.4 0.0  0.0 95 33.1
## 6    10   1  68 104.5 15.7 46.9 144 34.8 33.5 13.9 10.9 2.2  0.9 0.1  0.0 61 41.8
## 7    11   1  55  99.3 14.6 43.5 242 33.1 33.4 13.1 10.3 2.1  1.6 0.3  0.0 NA  NA
## 8    12   1  55  77.0 10.8 34.8  38 23.8 30.5 16.8  NA 0.4  0.1 0.1  0.0 93 36.3
## 9    13   1  67  95.3 10.9 30.4  88 33.6 35.3 13.3 10.7 0.4  0.2 0.0  0.0 57 33.8
## 10   19   2  52  83.0 10.3 30.1 105 28.6 34.1 13.2 11.3 0.9  0.9 0.3  0.1 69 28.1
##      FIB SODIUM POTASS  CA PHOS  MG CREA  BUN  HS GBIL  TP  ALB AMY PAMY LIP
## 1   578    137    3.88 2.29 1.20 0.66 0.65  5.7 5.3 0.59 67.0 36.7  30  16  10
## 2    NA    141    NA 2.21 0.58  NA 0.76 19.9  NA 0.48 65.3 37.4  NA  NA  NA
## 3   313    147    4.61 1.92 1.51 1.03 1.25 50.6  NA 8.42 40.5 22.1 146  NA  89
## 4   487    141    4.71 2.05 2.17 0.83 2.78 47.5 9.7 0.35 61.2 33.2  92  28  18
## 5   490    137    NA 2.34 0.97 0.74 0.65  8.5 3.0 0.42 78.4 43.8  84  50  50
## 6   400    141    4.41 2.08 0.99 0.56 0.82 15.3 5.5 2.40 57.5 30.1  95  57  25
## 7    NA    139    3.69  NA  NA  NA 1.21 13.0  NA 1.13  NA  NA 117  NA  73
## 8   413    142    4.67 2.31 1.16 0.87 1.77 29.8 6.2 0.45 70.8 43.6 177  43  30
## 9   431    143    2.35 2.10 0.51 0.36 1.00 15.0 4.7 1.21 67.4 35.4  NA  NA  NA
## 10  407    136    3.80 1.92 0.72 0.53 0.58 14.0 4.0 2.46 53.8 24.8  35  35  38
##      CHE  AP ASAT ALAT GGT  LDH  CK GLU TRIG CHOL  CRP  BASOR  EOSR
## 1   5.12  85  22  14  48  284  23 107  105  175  3.94 0.4132231 0.0000000
```

```
## 2 5.61 80 28 25 61 NA 36 84 NA NA 1.42 0.0000000 0.8264463
## 3 2.52 119 124 135 134 696 40 107 NA NA 12.09 0.5681818 0.5681818
## 4 4.10 94 774 72 23 1787 2422 105 134 141 3.78 0.0000000 0.0000000
## 5 6.91 108 35 22 72 NA 79 93 152 167 11.17 0.0000000 0.0000000
## 6 6.79 68 32 11 68 263 75 89 85 144 5.89 0.0000000 1.0000000
## 7 NA 51 29 20 138 303 230 91 NA NA 17.84 0.0000000 2.3255814
## 8 7.40 153 26 32 96 181 87 96 129 156 1.29 0.0000000 5.5555556
## 9 NA 239 91 57 446 183 53 86 62 118 1.36 NA NA
## 10 2.64 146 97 156 192 277 87 104 207 123 3.80 1.6666667 5.0000000
## LYMR MONOR NEU NEUR PDW RBC WBC BloodCulture
## 1 1.652893 7.024793 22.0 90.90909 10.6 3.7 24.10 no
## 2 3.305785 1.652893 11.4 94.21488 11.4 3.9 12.17 no
## 3 8.522727 6.818182 14.7 83.52273 14.1 2.5 17.45 no
## 4 11.016949 6.779661 9.7 82.20339 8.7 3.5 11.58 no
## 5 8.333333 4.166667 8.4 87.50000 12.2 4.4 9.86 no
## 6 22.000000 9.000000 6.8 68.00000 12.9 4.3 9.94 no
## 7 16.279070 12.403101 8.9 68.99225 12.5 4.5 13.06 no
## 8 22.222222 5.555556 1.2 66.66667 NA 4.7 1.78 no
## 9 NA NA NA NA NA NA NA yes
## 10 15.000000 15.000000 3.8 63.33333 13.2 3.5 5.98 no
```

```
# Mostramos los últimos datos también con tail()
tail(bacteremia_df, 10)
```

```
## ID SEX AGE MCV HGB HCT PLT MCH MCHC RDW MPV LYM MONO EOS BASO NT
## 14682 62410 1 72 91.5 11.5 33.7 126 31.3 34.1 14.0 10.1 0.7 1.1 0.1 0.0 61
## 14683 62411 1 30 81.9 14.7 41.2 255 29.4 35.9 12.3 9.5 1.9 1.1 0.0 0.0 84
## 14684 62417 1 63 89.7 13.0 38.0 157 30.8 34.4 13.7 10.2 0.3 0.8 0.0 0.0 60
## 14685 62420 1 28 87.5 10.4 29.8 217 31.0 35.3 12.4 9.7 1.4 2.0 0.1 0.0 95
## 14686 62432 1 37 82.0 9.0 25.9 145 28.6 34.6 16.4 9.9 0.0 0.3 0.0 0.0 44
## 14687 62436 2 44 97.4 7.9 22.5 248 33.6 34.4 13.3 9.8 1.3 0.4 0.1 0.0 NA
## 14688 62438 1 23 67.4 11.5 38.2 58 20.1 29.8 18.4 NA 0.6 0.5 0.0 0.0 NA
## 14689 62446 1 79 86.3 9.7 31.4 345 26.8 30.8 14.4 8.9 0.7 1.3 0.1 0.0 75
## 14690 62454 1 81 88.6 11.5 32.6 262 31.0 35.0 16.4 11.6 1.3 1.0 0.2 0.0 99
## 14691 62455 2 21 79.2 12.3 35.6 192 27.5 34.7 12.2 9.5 1.3 0.6 0.0 0.1 68
## APTT FIB SODIUM POTASS CA PHOS MG CREA BUN HS GBIL TP ALB AMY
## 14682 45.4 483 136 4.10 2.09 0.84 0.93 0.83 10.3 2.9 1.41 55.2 30.6 143
## 14683 35.5 523 137 3.87 2.44 0.81 0.89 0.93 15.6 5.8 1.98 83.5 45.7 64
## 14684 34.9 667 135 3.90 1.91 0.91 0.79 1.20 27.2 6.0 0.93 57.5 30.0 44
## 14685 33.2 674 139 3.61 2.21 1.30 0.80 0.75 12.1 3.9 1.12 65.8 33.3 57
## 14686 44.1 457 138 5.59 2.17 2.76 1.03 4.10 63.1 5.8 3.06 58.0 32.1 535
## 14687 NA NA 134 3.98 2.20 0.91 0.62 0.56 5.7 8.1 0.70 53.8 26.9 46
## 14688 NA NA 137 NA 2.24 0.97 NA 0.64 12.6 NA NA 66.3 42.8 31
## 14689 32.7 NA 137 4.02 2.32 0.81 0.92 1.65 23.4 5.3 0.78 72.2 32.9 62
## 14690 34.3 551 143 3.82 2.08 0.80 0.80 0.96 10.1 3.3 1.43 56.2 28.6 50
## 14691 43.0 347 135 3.49 2.22 1.02 0.78 0.69 8.4 5.1 1.76 72.9 38.4 32
## PAMY LIP CHE AP ASAT ALAT GGT LDH CK GLU TRIG CHOL CRP BASOR
## 14682 NA 35 4.91 37 48 21 34 288 219 101 61 73 18.56 0.000000
## 14683 18 14 7.47 90 34 50 26 349 216 99 46 159 7.88 0.000000
## 14684 NA 12 3.59 36 19 15 28 156 144 160 123 96 21.05 0.000000
## 14685 35 50 3.59 111 38 36 187 370 379 77 162 162 11.85 0.000000
## 14686 14 NA 2.56 87 62 28 102 346 150 113 70 89 15.12 0.000000
## 14687 34 146 2.60 51 28 22 144 564 61 76 298 NA 26.80 0.000000
## 14688 NA 14 NA 155 60 72 36 NA 28 85 NA NA 4.68 0.000000
```

```
## 14689 NA 15 4.32 71 56 14 34 272 274 113 NA NA 12.76 0.000000
## 14690 32 55 2.89 68 28 24 88 304 24 88 117 115 11.18 0.000000
## 14691 18 33 5.22 352 2751 2487 332 2080 42 95 53 134 1.00 2.941176
##          EOSR      LYMR      MONOR NEU      NEUR PDW RBC  WBC BloodCulture
## 14682 0.8403361 5.882353 9.243697 10.0 84.03361 11.2 3.8 11.95      no
## 14683 0.0000000 13.013699 7.534247 11.6 79.45205 11.1 5.0 14.63      no
## 14684 0.0000000 7.894737 21.052632 2.7 71.05263 11.5 4.2 4.00      no
## 14685 1.0638298 14.893617 21.276596 5.9 62.76596 10.5 3.3 9.41      no
## 14686 0.0000000 0.000000 1.477833 20.0 98.52217 11.6 3.4 20.34      no
## 14687 2.0833333 27.083333 8.333333 3.0 62.50000 11.0 2.4 4.89      no
## 14688 0.0000000 10.714286 8.928571 4.5 80.35714 NA 5.7 5.62      no
## 14689 0.8771930 6.140351 11.403509 9.3 81.57895 9.3 3.8 11.39      no
## 14690 2.1276596 13.829787 10.638298 6.9 73.40426 14.1 3.7 9.30      no
## 14691 0.0000000 38.235294 17.647059 1.4 41.17647 10.8 4.4 3.46      no
```

Realizaremos un exploratorio genérico del dataset. Mostrando información básica del dataset, para pasar luego a los estadísticos básico y comenzaremos a interactuar con las características luego.

Verificamos la estructura del juego de datos principal. Vemos el número de columnas que tenemos y ejemplos de los contenidos de las filas.

```
str(bacteremia_df)
```

```
## 'data.frame': 14691 obs. of 53 variables:
## $ ID : int 1 3 5 7 9 10 11 12 13 19 ...
## $ SEX : int 2 1 1 1 2 1 1 1 1 2 ...
## $ AGE : int 62 72 46 84 38 68 55 55 67 52 ...
## $ MCV : num 99.3 85.1 96.3 91.3 85.1 ...
## $ HGB : num 11.5 10.3 7.4 10.3 13.7 15.7 14.6 10.8 10.9 10.3 ...
## $ HCT : num 35.9 34.7 22.8 31.1 38.7 46.9 43.5 34.8 30.4 30.1 ...
## $ PLT : int 307 182 64 309 183 144 242 38 88 105 ...
## $ MCH : num 31.5 26 31.2 30.4 30.2 34.8 33.1 23.8 33.6 28.6 ...
## $ MCHC : num 31.8 30.6 32.4 33.3 35.3 33.5 33.4 30.5 35.3 34.1 ...
## $ RDW : num 19.5 15 19.7 13.8 12.6 13.9 13.1 16.8 13.3 13.2 ...
## $ MPV : num 10.8 9.7 11.1 8.5 10 10.9 10.3 NA 10.7 11.3 ...
## $ LYM : num 0.4 0.4 1.5 1.3 0.8 2.2 2.1 0.4 0.4 0.9 ...
## $ MONO : num 1.7 0.2 1.2 0.8 0.4 0.9 1.6 0.1 0.2 0.9 ...
## $ EOS : num 0 0.1 0.1 0 0 0.1 0.3 0.1 0 0.3 ...
## $ BASO : num 0.1 0 0.1 0 0 0 0 0 0 0.1 ...
## $ NT : int 86 90 58 67 95 61 NA 93 57 69 ...
## $ APTT : num 28.8 29.8 36.3 38.2 33.1 41.8 NA 36.3 33.8 28.1 ...
## $ FIB : int 578 NA 313 487 490 400 NA 413 431 407 ...
## $ SODIUM : int 137 141 147 141 137 141 139 142 143 136 ...
## $ POTASS : num 3.88 NA 4.61 4.71 NA 4.41 3.69 4.67 2.35 3.8 ...
## $ CA : num 2.29 2.21 1.92 2.05 2.34 2.08 NA 2.31 2.1 1.92 ...
## $ PHOS : num 1.2 0.58 1.51 2.17 0.97 0.99 NA 1.16 0.51 0.72 ...
## $ MG : num 0.66 NA 1.03 0.83 0.74 0.56 NA 0.87 0.36 0.53 ...
## $ CREA : num 0.65 0.76 1.25 2.78 0.65 0.82 1.21 1.77 1 0.58 ...
## $ BUN : num 5.7 19.9 50.6 47.5 8.5 15.3 13 29.8 15 14 ...
## $ HS : num 5.3 NA NA 9.7 3 5.5 NA 6.2 4.7 4 ...
## $ GBIL : num 0.59 0.48 8.42 0.35 0.42 2.4 1.13 0.45 1.21 2.46 ...
## $ TP : num 67 65.3 40.5 61.2 78.4 57.5 NA 70.8 67.4 53.8 ...
## $ ALB : num 36.7 37.4 22.1 33.2 43.8 30.1 NA 43.6 35.4 24.8 ...
## $ AMY : int 30 NA 146 92 84 95 117 177 NA 35 ...
```

```
## $ PAMY      : int  16 NA NA 28 50 57 NA 43 NA 35 ...
## $ LIP       : int  10 NA 89 18 50 25 73 30 NA 38 ...
## $ CHE       : num  5.12 5.61 2.52 4.1 6.91 6.79 NA 7.4 NA 2.64 ...
## $ AP        : int  85 80 119 94 108 68 51 153 239 146 ...
## $ ASAT      : int  22 28 124 774 35 32 29 26 91 97 ...
## $ ALAT      : int  14 25 135 72 22 11 20 32 57 156 ...
## $ GGT       : int  48 61 134 23 72 68 138 96 446 192 ...
## $ LDH       : int  284 NA 696 1787 NA 263 303 181 183 277 ...
## $ CK        : int  23 36 40 2422 79 75 230 87 53 87 ...
## $ GLU       : int  107 84 107 105 93 89 91 96 86 104 ...
## $ TRIG      : int  105 NA NA 134 152 85 NA 129 62 207 ...
## $ CHOL      : int  175 NA NA 141 167 144 NA 156 118 123 ...
## $ CRP       : num  3.94 1.42 12.09 3.78 11.17 ...
## $ BASOR     : num  0.413 0 0.568 0 0 ...
## $ EOSR      : num  0 0.826 0.568 0 0 ...
## $ LYMR      : num  1.65 3.31 8.52 11.02 8.33 ...
## $ MONOR     : num  7.02 1.65 6.82 6.78 4.17 ...
## $ NEU       : num  22 11.4 14.7 9.7 8.4 6.8 8.9 1.2 NA 3.8 ...
## $ NEUR      : num  90.9 94.2 83.5 82.2 87.5 ...
## $ PDW       : num  10.6 11.4 14.1 8.7 12.2 12.9 12.5 NA NA 13.2 ...
## $ RBC       : num  3.7 3.9 2.5 3.5 4.4 4.3 4.5 4.7 NA 3.5 ...
## $ WBC       : num  24.1 12.17 17.45 11.58 9.86 ...
## $ BloodCulture: chr  "no" "no" "no" "no" ...
```

```
# Observamos su composición
dim(bacteremia_df)
```

```
## [1] 14691    53
```

Vemos que tenemos **53** características y **14691** observaciones. Deberíamos profundizar ahora respecto al tipo de dato de cada variable y realizar un exploratorio básico para ver inconsistencias.

```
# Obtenemos el datatype de cada característica
sapply(bacteremia_df, class)
```

```
##      ID      SEX      AGE      MCV      HGB      HCT
## "integer" "integer" "integer" "numeric" "numeric" "numeric"
##      PLT      MCH      MCHC      RDW      MPV      LYM
## "integer" "numeric" "numeric" "numeric" "numeric" "numeric"
##      MONO      EOS      BASO      NT      APTT      FIB
## "numeric" "numeric" "numeric" "integer" "numeric" "integer"
##      SODIUM      POTASS      CA      PHOS      MG      CREA
## "integer" "numeric" "numeric" "numeric" "numeric" "numeric"
##      BUN      HS      GBIL      TP      ALB      AMY
## "numeric" "numeric" "numeric" "numeric" "numeric" "integer"
##      PAMY      LIP      CHE      AP      ASAT      ALAT
## "integer" "integer" "numeric" "integer" "integer" "integer"
##      GGT      LDH      CK      GLU      TRIG      CHOL
## "integer" "integer" "integer" "integer" "integer" "integer"
##      CRP      BASOR      EOSR      LYMR      MONOR      NEU
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##      NEUR      PDW      RBC      WBC      BloodCulture
## "numeric" "numeric" "numeric" "numeric" "character"
```



```
# Install and load necessary packages
# install.packages(c("summarytools", "skimr", "DataExplorer"))
library(summarytools)
library(skimr)
library(DataExplorer)

# Try a simpler version first
dfSummary(bacteremia_df)
```

```
## Data Frame Summary
## bacteremia_df
## Dimensions: 14691 x 53
## Duplicates: 0
##
## -----
## No   Variable      Stats / Values      Freqs (% of Valid)  Graph      Va
## ----
## 1    ID            Mean (sd) : 29352.5 (18178.7)  14691 distinct values  : . . . . .      14
##      [integer]    min < med < max:          : : : : : . .      (1
##      1 < 28755 < 62455          : : : : : : : : :
##      IQR (CV) : 31083.5 (0.6)  : : : : : : : : :
##                                     : : : : : : : : :
##
## 2    SEX          Min   : 1              1 : 8536 (58.1%)      IIIIIIIIIII      14
##      [integer]    Mean   : 1.4          2 : 6155 (41.9%)      IIIIIIII         (1
##      Max   : 2                                     : : : : :
##
## 3    AGE          Mean (sd) : 56.2 (18.2)      85 distinct values      :              14
##      [integer]    min < med < max:          . : .          (1
##      16 < 58 < 101          . : : : : :
##      IQR (CV) : 27 (0.3)      . : : : : : :
##                                     : : : : : : : : :
##
## 4    MCV          Mean (sd) : 88.4 (6.5)      506 distinct values      :              14
##      [numeric]    min < med < max:          :              (9
##      51 < 88.3 < 128.7          : :              :
##      IQR (CV) : 7.3 (0.1)      : :              :
##                                     : : : : .
##
## 5    HGB          Mean (sd) : 11.6 (2.3)      157 distinct values      . :              14
##      [numeric]    min < med < max:          : : :          (9
##      3 < 11.4 < 21          : : : : .
##      IQR (CV) : 3.3 (0.2)      : : : : :
##                                     : : : : : .
##
## 6    HCT          Mean (sd) : 34.5 (6.5)      404 distinct values      : :              14
##      [numeric]    min < med < max:          : :              (9
##      0 < 34.3 < 66.6          : : : .
##      IQR (CV) : 9.3 (0.2)      . : : : :
##                                     : : : : : .
##
## 7    PLT          Mean (sd) : 220 (122.8)      718 distinct values      :              14
##      [integer]    min < med < max:          : :              (9
##      0 < 204 < 2092          : :              : :
```

```
##          IQR (CV) : 137 (0.6)          : :
##          : : .
##
## 8      MCH          Mean (sd) : 29.6 (2.5)          232 distinct values          :          14
##      [numeric]      min < med < max:          :          (9
##          14.9 < 29.7 < 47.4          :
##          IQR (CV) : 2.6 (0.1)          . : .
##          . : : :
##
## 9      MCHC          Mean (sd) : 33.5 (1.4)          124 distinct values          : .          14
##      [numeric]      min < med < max:          : :          (9
##          23.7 < 33.5 < 43.5          : :
##          IQR (CV) : 1.8 (0)          : :
##          : : : .
##
## 10     RDW          Mean (sd) : 15 (2.3)          173 distinct values          :          14
##      [numeric]      min < med < max:          :          (9
##          10.6 < 14.5 < 31.8          : :
##          IQR (CV) : 2.6 (0.2)          : :
##          : : : : .
##
## 11     MPV          Mean (sd) : 10.4 (1)          71 distinct values          : :          13
##      [numeric]      min < med < max:          . : :          (9
##          7.3 < 10.3 < 15          : : : .
##          IQR (CV) : 1.3 (0.1)          : : : :
##          : : : : : .
##
## 12     LYM          Mean (sd) : 1.4 (7.5)          114 distinct values          :          14
##      [numeric]      min < med < max:          :          (9
##          0 < 1 < 578.1          :
##          IQR (CV) : 0.9 (5.5)          :
##          :
##
## 13     MONO          Mean (sd) : 0.9 (0.7)          67 distinct values          :          14
##      [numeric]      min < med < max:          :          (9
##          0 < 0.8 < 20.4          :
##          IQR (CV) : 0.6 (0.8)          :
##          :
##
## 14     EOS          Mean (sd) : 0.1 (0.3)          36 distinct values          :          14
##      [numeric]      min < med < max:          :          (9
##          0 < 0.1 < 15.8          :
##          IQR (CV) : 0.1 (2.3)          :
##          :
##
## 15     BASO          Mean (sd) : 0 (0.1)          18 distinct values          :          14
##      [numeric]      min < med < max:          :          (9
##          0 < 0 < 6.5          :
##          IQR (CV) : 0 (4.7)          :
##          :
##
## 16     NT          Mean (sd) : 83.2 (27.2)          149 distinct values          : :          12
##      [integer]      min < med < max:          : : :          (8
##          4 < 83 < 152          . : : : .
```

```
##          IQR (CV) : 34 (0.3)          . : : : : :
##          . . : : : : : : : .
##
## 17  APTT          Mean (sd) : 40.1 (11)          631 distinct values  . :          12
##      [numeric]    min < med < max:              : :          (8
##                  21.4 < 37.7 < 176.1            : :
##                  IQR (CV) : 8.6 (0.3)            : :
##                  : : .
##
## 18  FIB           Mean (sd) : 547.4 (208.1)        1084 distinct values  : .          12
##      [integer]    min < med < max:              : :          (8
##                  55 < 529 < 1593                : : : :
##                  IQR (CV) : 277 (0.4)            : : : :
##                  . : : : : : .
##
## 19  SODIUM        Mean (sd) : 137.2 (4.7)          58 distinct values      :          13
##      [integer]    min < med < max:              : .          (9
##                  106 < 137 < 170                : :
##                  IQR (CV) : 5 (0)                : :
##                  : : : .
##
## 20  POTASS        Mean (sd) : 4 (0.6)              408 distinct values    :          12
##      [numeric]    min < med < max:              :          (8
##                  1.9 < 4 < 36.6                 :
##                  IQR (CV) : 0.6 (0.2)            :
##                  :
##
## 21  CA            Mean (sd) : 2.2 (0.2)            185 distinct values    :          13
##      [numeric]    min < med < max:              :          (9
##                  1 < 2.2 < 4.4                  :
##                  IQR (CV) : 0.3 (0.1)            . : .
##                  : : :
##
## 22  PHOS          Mean (sd) : 1 (0.4)              306 distinct values    :          13
##      [numeric]    min < med < max:              . :          (9
##                  0.3 < 1 < 6.2                  : :
##                  IQR (CV) : 0.4 (0.4)            : :
##                  : : .
##
## 23  MG            Mean (sd) : 0.8 (0.2)            146 distinct values    : :          12
##      [numeric]    min < med < max:              : :          (8
##                  0.2 < 0.8 < 2.2                : :
##                  IQR (CV) : 0.2 (0.2)            : :
##                  . : : :
##
## 24  CREA          Mean (sd) : 1.3 (1.2)            674 distinct values    :          14
##      [numeric]    min < med < max:              :          (9
##                  0.3 < 1 < 20.8                 :
##                  IQR (CV) : 0.5 (0.9)            :
##                  : .
##
## 25  BUN           Mean (sd) : 22.7 (18.1)          947 distinct values    :          14
##      [numeric]    min < med < max:              :          (9
##                  2.5 < 16.6 < 184.8             :
```

```
##          IQR (CV) : 15.3 (0.8)          : :
##          : : . .
##
## 26  HS          Mean (sd) : 5.4 (2.5)          169 distinct values          :          11
##      [numeric]  min < med < max:          : .          (7
##          1.3 < 5 < 22.7          . : :
##          IQR (CV) : 2.9 (0.5)          : : :
##          : : : .
##
## 27  GBIL        Mean (sd) : 1.4 (2.8)          885 distinct values          :          13
##      [numeric]  min < med < max:          :          (9
##          0.1 < 0.8 < 51.8          :
##          IQR (CV) : 0.7 (2)          :
##          :
##
## 28  TP          Mean (sd) : 64.9 (11.5)          649 distinct values          . :          13
##      [numeric]  min < med < max:          : :          (8
##          29.9 < 65.7 < 120.9          : : : .
##          IQR (CV) : 16.4 (0.2)          : : : :
##          . : : : : .
##
## 29  ALB          Mean (sd) : 33.4 (7.4)          401 distinct values          . : .          13
##      [numeric]  min < med < max:          : : : :          (8
##          10 < 33.6 < 55.7          : : : : .
##          IQR (CV) : 11.2 (0.2)          : : : : :
##          . : : : : : .
##
## 30  AMY          Mean (sd) : 90.8 (805.2)          488 distinct values          :          10
##      [integer]  min < med < max:          :          (7
##          8 < 49 < 56146          :
##          IQR (CV) : 43 (8.9)          :
##          :
##
## 31  PAMY          Mean (sd) : 41.7 (448)          280 distinct values          :          75
##      [integer]  min < med < max:          :          (5
##          1 < 22 < 38369          :
##          IQR (CV) : 22 (10.8)          :
##          :
##
## 32  LIP          Mean (sd) : 63.8 (603.9)          444 distinct values          :          10
##      [integer]  min < med < max:          :          (7
##          0 < 23 < 45991          :
##          IQR (CV) : 26 (9.5)          :
##          :
##
## 33  CHE          Mean (sd) : 4.8 (2.1)          997 distinct values          : : .          12
##      [numeric]  min < med < max:          : : : .          (8
##          1 < 4.6 < 13.9          . : : : :
##          IQR (CV) : 3.1 (0.4)          : : : : :
##          : : : : : .
##
## 34  AP          Mean (sd) : 118.8 (132.9)          672 distinct values          :          13
##      [integer]  min < med < max:          :          (9
##          11 < 84 < 2995          :
```

```
##          IQR (CV) : 60 (1.1)          :
##                                          :
##
## 35  ASAT          Mean (sd) : 86.9 (404.7)      650 distinct values :
##      [integer]    min < med < max:             :
##                  3 < 31 < 13991                 :
##                  IQR (CV) : 34 (4.7)           :
##                                          :
##
## 36  ALAT          Mean (sd) : 67.7 (311)        578 distinct values :
##      [integer]    min < med < max:             :
##                  0 < 26 < 15059                 :
##                  IQR (CV) : 32 (4.6)           :
##                                          :
##
## 37  GGT           Mean (sd) : 115.1 (209)        858 distinct values :
##      [integer]    min < med < max:             :
##                  3 < 49 < 5171                  :
##                  IQR (CV) : 92 (1.8)           :
##                                          :
##
## 38  LDH           Mean (sd) : 331.2 (475.2)      1137 distinct values :
##      [integer]    min < med < max:             :
##                  39 < 239 < 13906               :
##                  IQR (CV) : 145 (1.4)          :
##                                          :
##
## 39  CK            Mean (sd) : 385 (2241.2)       1506 distinct values :
##      [integer]    min < med < max:             :
##                  8 < 80 < 98801                 :
##                  IQR (CV) : 142 (5.8)          :
##                                          :
##
## 40  GLU           Mean (sd) : 126.4 (56.9)       389 distinct values :
##      [integer]    min < med < max:             :
##                  19 < 113 < 1403                :
##                  IQR (CV) : 41 (0.5)           :
##                                          :
##
## 41  TRIG          Mean (sd) : 141.7 (120.7)      538 distinct values :
##      [integer]    min < med < max:             :
##                  14 < 115 < 5440                :
##                  IQR (CV) : 82 (0.9)           :
##                                          :
##
## 42  CHOL          Mean (sd) : 150.8 (55.5)       339 distinct values :
##      [integer]    min < med < max:             . :
##                  25 < 145 < 1104                 : :
##                  IQR (CV) : 69 (0.4)           : :
##                                          : : .
##
## 43  CRP           Mean (sd) : 10.9 (9.6)        3328 distinct values :
##      [numeric]    min < med < max:             :
##                  0 < 8.6 < 76.3                 : :
```

```
##          IQR (CV) : 13.6 (0.9)          : : .
##          : : : : .
##
## 44  BASOR      Mean (sd) : 0.1 (0.6)      419 distinct values :          13
##      [numeric] min < med < max:          :          (9
##              0 < 0 < 23.7              :
##              IQR (CV) : 0 (4)           :
##              :
##
## 45  EOSR      Mean (sd) : 1.3 (2.4)      927 distinct values :          13
##      [numeric] min < med < max:          :          (9
##              0 < 0.6 < 73.5            :
##              IQR (CV) : 1.8 (1.8)       :
##              :
##
## 46  LYMR      Mean (sd) : 14.6 (12.7)     3121 distinct values :          13
##      [numeric] min < med < max:          : :          (9
##              0 < 11.3 < 100            : :
##              IQR (CV) : 11.4 (0.9)       : : .
##              : : : : .
##
## 47  MONOR     Mean (sd) : 8.8 (5.8)      2334 distinct values :          13
##      [numeric] min < med < max:          :          (9
##              0 < 8 < 100              :
##              IQR (CV) : 5.2 (0.7)       : :
##              : :
##
## 48  NEU       Mean (sd) : 8.4 (5.6)      374 distinct values :          13
##      [numeric] min < med < max:          :          (9
##              0 < 7.3 < 83.8            : :
##              IQR (CV) : 6.2 (0.7)       : :
##              : : .
##
## 49  NEUR      Mean (sd) : 75.2 (15.5)     3850 distinct values . :          13
##      [numeric] min < med < max:          : :          (9
##              0 < 78.3 < 100            : :
##              IQR (CV) : 16.1 (0.2)       : : : : .
##              . : : : : :
##
## 50  PDW       Mean (sd) : 12.3 (2.2)      167 distinct values :          13
##      [numeric] min < med < max:          : .          (9
##              6.6 < 12 < 25.3            : :
##              IQR (CV) : 2.6 (0.2)       : : : .
##              : : : : .
##
## 51  RBC       Mean (sd) : 3.9 (0.8)      65 distinct values . :          14
##      [numeric] min < med < max:          : :          (9
##              1 < 3.9 < 8.2              : :
##              IQR (CV) : 1.1 (0.2)       : : : :
##              : : : : .
##
## 52  WBC       Mean (sd) : 11.2 (12.9)     2710 distinct values :          14
##      [numeric] min < med < max:          :          (9
##              0 < 9.6 < 604.5            :
```

```
##          IQR (CV) : 6.9 (1.2)          :
##                                          :
##
## 53   BloodCulture   1. no                13511 (92.0%)      IHHHHHHHHHHHHHHHHH      14
##      [character]   2. yes                1180 ( 8.0%)      I                      (1
## -----
```

If that works, try with minimal parameters

```
dfSummary(bacteremia_df,
  style = "grid",
  plain.ascii = FALSE)
```

```
## ### Data Frame Summary
```

```
## ##### bacteremia_df
```

```
## **Dimensions:** 14691 x 53
```

```
## **Duplicates:** 0
```

```
##
```

```
## +---+-----+-----+-----+-----+
## | No | Variable      | Stats / Values                | Freqs (% of Valid) | Graph |
## +---+-----+-----+-----+-----+
## | 1  | ID\           | Mean (sd) : 29352.5 (18178.7)\ | 14691 distinct values | : . . \ \ . .\ |
## |   | [integer]     | min < med < max:\           |                       | : : : : : . \ \ . |
## |   |               | 1 < 28755 < 62455\          |                       | : : : : : : : : : \ |
## |   |               | IQR (CV) : 31083.5 (0.6)     |                       | : : : : : : : : : \ |
## |   |               |                               |                       | : : : : : : : : : |
## +---+-----+-----+-----+-----+
## | 2  | SEX\          | Min : 1\                     | 1 : 8536 (58.1%\    | IHHHHHHHHH \ |
## |   | [integer]     | Mean : 1.4\                  | 2 : 6155 (41.9%)    | IHHHHHHH |
## |   |               | Max : 2                      |                       | |
## +---+-----+-----+-----+-----+
## | 3  | AGE\          | Mean (sd) : 56.2 (18.2)\     | 85 distinct values  | \ \ \ \ \ \ \ \ \ \ |
## |   | [integer]     | min < med < max:\           |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               | 16 < 58 < 101\              |                       | \ \ \ \ \ . : : : : \ |
## |   |               | IQR (CV) : 27 (0.3)         |                       | . : : : : : : : : \ |
## |   |               |                               |                       | : : : : : : : : : |
## +---+-----+-----+-----+-----+
## | 4  | MCV\          | Mean (sd) : 88.4 (6.5)\     | 506 distinct values | \ \ \ \ \ \ \ \ \ \ |
## |   | [numeric]     | min < med < max:\           |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               | 51 < 88.3 < 128.7\          |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               | IQR (CV) : 7.3 (0.1)        |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               |                               |                       | \ \ \ \ \ \ : : : . |
## +---+-----+-----+-----+-----+
## | 5  | HGB\          | Mean (sd) : 11.6 (2.3)\     | 157 distinct values | \ \ \ \ \ \ \ \ \ \ |
## |   | [numeric]     | min < med < max:\           |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               | 3 < 11.4 < 21\              |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               | IQR (CV) : 3.3 (0.2)        |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               |                               |                       | \ \ \ \ \ : : : : . |
## +---+-----+-----+-----+-----+
## | 6  | HCT\          | Mean (sd) : 34.5 (6.5)\     | 404 distinct values | \ \ \ \ \ \ \ \ \ \ |
## |   | [numeric]     | min < med < max:\           |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               | 0 < 34.3 < 66.6\           |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               | IQR (CV) : 9.3 (0.2)        |                       | \ \ \ \ \ \ \ \ \ \ |
## |   |               |                               |                       | \ \ \ \ \ \ : : : . |
## +---+-----+-----+-----+-----+
```

##	7	PLT\	Mean (sd) : 220 (122.8)\	718 distinct values	:\
##		[integer]	min < med < max:\		: :\
##			0 < 204 < 2092\		: :\
##			IQR (CV) : 137 (0.6)		: :\
##					: : .
##	8	MCH\	Mean (sd) : 29.6 (2.5)\	232 distinct values	\ \ \ \ \ \ \ \ :\
##		[numeric]	min < med < max:\		\ \ \ \ \ \ \ \ :\
##			14.9 < 29.7 < 47.4\		\ \ \ \ \ \ \ \ :\
##			IQR (CV) : 2.6 (0.1)		\ \ \ \ \ \ \ . : :\
##					\ \ \ \ \ . : : :
##	9	MCHC\	Mean (sd) : 33.5 (1.4)\	124 distinct values	\ \ \ \ \ \ \ \ : :\
##		[numeric]	min < med < max:\		\ \ \ \ \ \ \ \ : :\
##			23.7 < 33.5 < 43.5\		\ \ \ \ \ \ \ \ : :\
##			IQR (CV) : 1.8 (0)		\ \ \ \ \ \ \ \ : :\
##					\ \ \ \ \ \ \ : : : .
##	10	RDW\	Mean (sd) : 15 (2.3)\	173 distinct values	\ \ \ :\
##		[numeric]	min < med < max:\		\ \ \ :\
##			10.6 < 14.5 < 31.8\		\ \ \ : :\
##			IQR (CV) : 2.6 (0.2)		\ \ \ : :\
##					: : : : .
##	11	MPV\	Mean (sd) : 10.4 (1)\	71 distinct values	\ \ \ \ \ \ \ : :\
##		[numeric]	min < med < max:\		\ \ \ \ \ . : :\
##			7.3 < 10.3 < 15\		\ \ \ \ \ : : : :\
##			IQR (CV) : 1.3 (0.1)		\ \ \ \ \ : : : :\
##					\ \ \ : : : : : .
##	12	LYM\	Mean (sd) : 1.4 (7.5)\	114 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 1 < 578.1\		:\
##			IQR (CV) : 0.9 (5.5)		:\
##					:
##	13	MONO\	Mean (sd) : 0.9 (0.7)\	67 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 0.8 < 20.4\		:\
##			IQR (CV) : 0.6 (0.8)		:\
##					:
##	14	EOS\	Mean (sd) : 0.1 (0.3)\	36 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 0.1 < 15.8\		:\
##			IQR (CV) : 0.1 (2.3)		:\
##					:
##	15	BASO\	Mean (sd) : 0 (0.1)\	18 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 0 < 6.5\		:\
##			IQR (CV) : 0 (4.7)		:\
##					:

##	16	NT\	Mean (sd) : 83.2 (27.2)\	149 distinct values	\ \ \ \ \ \ \ \ : \
##		[integer]	min < med < max:\		\ \ \ \ \ \ \ \ : :
##			4 < 83 < 152\		\ \ \ \ \ \ \ \ . : : :
##			IQR (CV) : 34 (0.3)		\ \ \ \ \ . : : : : \
##					. . : : : : : .
##	17	APTT\	Mean (sd) : 40.1 (11)\	631 distinct values	. :\
##		[numeric]	min < med < max:\		: :\
##			21.4 < 37.7 < 176.1\		: :\
##			IQR (CV) : 8.6 (0.3)		: :\
##					: : .
##	18	FIB\	Mean (sd) : 547.4 (208.1)\	1084 distinct values	\ \ \ \ \ : .\
##		[integer]	min < med < max:\		\ \ \ \ \ : :\
##			55 < 529 < 1593\		\ \ \ : : : :\
##			IQR (CV) : 277 (0.4)		\ \ \ : : : :\
##					. : : : : .
##	19	SODIUM\	Mean (sd) : 137.2 (4.7)\	58 distinct values	\ \ \ \ \ \ \ \ \ :\
##		[integer]	min < med < max:\		\ \ \ \ \ \ \ \ \ : .\
##			106 < 137 < 170\		\ \ \ \ \ \ \ \ \ : :\
##			IQR (CV) : 5 (0)		\ \ \ \ \ \ \ \ \ : :\
##					\ \ \ \ \ \ \ \ : : : .
##	20	POTASS\	Mean (sd) : 4 (0.6)\	408 distinct values	: \
##		[numeric]	min < med < max:\		: \
##			1.9 < 4 < 36.6\		: \
##			IQR (CV) : 0.6 (0.2)		: \
##					: :
##	21	CA\	Mean (sd) : 2.2 (0.2)\	185 distinct values	\ \ \ \ \ \ \ :\
##		[numeric]	min < med < max:\		\ \ \ \ \ \ \ :\
##			1 < 2.2 < 4.4\		\ \ \ \ \ \ \ :\
##			IQR (CV) : 0.3 (0.1)		\ \ \ \ \ . : .\
##					\ \ \ \ \ : : :
##	22	PHOS\	Mean (sd) : 1 (0.4)\	306 distinct values	\ \ \ :\
##		[numeric]	min < med < max:\		. :\
##			0.3 < 1 < 6.2\		: :\
##			IQR (CV) : 0.4 (0.4)		: :\
##					: : .
##	23	MG\	Mean (sd) : 0.8 (0.2)\	146 distinct values	\ \ \ \ \ : :\
##		[numeric]	min < med < max:\		\ \ \ \ \ : :\
##			0.2 < 0.8 < 2.2\		\ \ \ \ \ : :\
##			IQR (CV) : 0.2 (0.2)		\ \ \ \ \ : :\
##					\ \ \ . : : :
##	24	CREA\	Mean (sd) : 1.3 (1.2)\	674 distinct values	: \
##		[numeric]	min < med < max:\		: \
##			0.3 < 1 < 20.8\		: \
##			IQR (CV) : 0.5 (0.9)		: \
##					: .

##	25	BUN\	Mean (sd) : 22.7 (18.1)\	947 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			2.5 < 16.6 < 184.8\		:\
##			IQR (CV) : 15.3 (0.8)		: :\
##					: : . .
##	26	HS\	Mean (sd) : 5.4 (2.5)\	169 distinct values	\ \ :\
##		[numeric]	min < med < max:\		\ \ : .\
##			1.3 < 5 < 22.7\		. : :\
##			IQR (CV) : 2.9 (0.5)		: : :\
##					: : : .
##	27	GBIL\	Mean (sd) : 1.4 (2.8)\	885 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0.1 < 0.8 < 51.8\		:\
##			IQR (CV) : 0.7 (2)		:\
##					:
##	28	TP\	Mean (sd) : 64.9 (11.5)\	649 distinct values	\ \ \ \ \ \ . :\
##		[numeric]	min < med < max:\		\ \ \ \ \ \ : :\
##			29.9 < 65.7 < 120.9\		\ \ \ \ \ : : : .\
##			IQR (CV) : 16.4 (0.2)		\ \ \ \ \ : : : :\
##					. : : : : .
##	29	ALB\	Mean (sd) : 33.4 (7.4)\	401 distinct values	\ \ \ \ \ \ \ \ . :
##		[numeric]	min < med < max:\		\ \ \ \ \ \ : : : :\
##			10 < 33.6 < 55.7\		\ \ \ \ \ \ : : : : :
##			IQR (CV) : 11.2 (0.2)		\ \ \ \ \ : : : : : :\
##					\ \ \ . : : : : : .
##	30	AMY\	Mean (sd) : 90.8 (805.2)\	488 distinct values	:\
##		[integer]	min < med < max:\		:\
##			8 < 49 < 56146\		:\
##			IQR (CV) : 43 (8.9)		:\
##					:
##	31	PAMY\	Mean (sd) : 41.7 (448)\	280 distinct values	:\
##		[integer]	min < med < max:\		:\
##			1 < 22 < 38369\		:\
##			IQR (CV) : 22 (10.8)		:\
##					:
##	32	LIP\	Mean (sd) : 63.8 (603.9)\	444 distinct values	:\
##		[integer]	min < med < max:\		:\
##			0 < 23 < 45991\		:\
##			IQR (CV) : 26 (9.5)		:\
##					:
##	33	CHE\	Mean (sd) : 4.8 (2.1)\	997 distinct values	\ \ : : .\
##		[numeric]	min < med < max:\		\ \ : : : .\
##			1 < 4.6 < 13.9\		. : : : : \
##			IQR (CV) : 3.1 (0.4)		: : : : : :\
##					: : : : : .

##	34	AP\	Mean (sd) : 118.8 (132.9)\	672 distinct values	:\
##		[integer]	min < med < max:\		:\
##			11 < 84 < 2995\		:\
##			IQR (CV) : 60 (1.1)		:\
##					:
##	35	ASAT\	Mean (sd) : 86.9 (404.7)\	650 distinct values	:\
##		[integer]	min < med < max:\		:\
##			3 < 31 < 13991\		:\
##			IQR (CV) : 34 (4.7)		:\
##					:
##	36	ALAT\	Mean (sd) : 67.7 (311)\	578 distinct values	:\
##		[integer]	min < med < max:\		:\
##			0 < 26 < 15059\		:\
##			IQR (CV) : 32 (4.6)		:\
##					:
##	37	GGT\	Mean (sd) : 115.1 (209)\	858 distinct values	:\
##		[integer]	min < med < max:\		:\
##			3 < 49 < 5171\		:\
##			IQR (CV) : 92 (1.8)		:\
##					:
##	38	LDH\	Mean (sd) : 331.2 (475.2)\	1137 distinct values	:\
##		[integer]	min < med < max:\		:\
##			39 < 239 < 13906\		:\
##			IQR (CV) : 145 (1.4)		:\
##					:
##	39	CK\	Mean (sd) : 385 (2241.2)\	1506 distinct values	:\
##		[integer]	min < med < max:\		:\
##			8 < 80 < 98801\		:\
##			IQR (CV) : 142 (5.8)		:\
##					:
##	40	GLU\	Mean (sd) : 126.4 (56.9)\	389 distinct values	:\
##		[integer]	min < med < max:\		:\
##			19 < 113 < 1403\		:\
##			IQR (CV) : 41 (0.5)		:\
##					:
##	41	TRIG\	Mean (sd) : 141.7 (120.7)\	538 distinct values	:\
##		[integer]	min < med < max:\		:\
##			14 < 115 < 5440\		:\
##			IQR (CV) : 82 (0.9)		:\
##					:
##	42	CHOL\	Mean (sd) : 150.8 (55.5)\	339 distinct values	:\ \ \
##		[integer]	min < med < max:\		:\
##			25 < 145 < 1104\		:\
##			IQR (CV) : 69 (0.4)		:\
##					:

##	43	CRP\	Mean (sd) : 10.9 (9.6)\	3328 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 8.6 < 76.3\		: :\
##			IQR (CV) : 13.6 (0.9)		: : :\
##					: : : .
##	44	BASOR\	Mean (sd) : 0.1 (0.6)\	419 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 0 < 23.7\		:\
##			IQR (CV) : 0 (4)		:\
##					:
##	45	EOSR\	Mean (sd) : 1.3 (2.4)\	927 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 0.6 < 73.5\		:\
##			IQR (CV) : 1.8 (1.8)		:\
##					:
##	46	LYMR\	Mean (sd) : 14.6 (12.7)\	3121 distinct values	:\
##		[numeric]	min < med < max:\		: :\
##			0 < 11.3 < 100\		: :\
##			IQR (CV) : 11.4 (0.9)		: : :\
##					: : : .
##	47	MONOR\	Mean (sd) : 8.8 (5.8)\	2334 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 8 < 100\		:\
##			IQR (CV) : 5.2 (0.7)		: :\
##					: :
##	48	NEU\	Mean (sd) : 8.4 (5.6)\	374 distinct values	:\
##		[numeric]	min < med < max:\		:\
##			0 < 7.3 < 83.8\		: :\
##			IQR (CV) : 6.2 (0.7)		: :\
##					: : .
##	49	NEUR\	Mean (sd) : 75.2 (15.5)\	3850 distinct values	\\ \\ \\ \\ \\ \\ \\ \\ \\ \\
##		[numeric]	min < med < max:\		\\ \\ \\ \\ \\ \\ \\ \\ \\ \\
##			0 < 78.3 < 100\		\\ \\ \\ \\ \\ \\ \\ \\ \\ \\
##			IQR (CV) : 16.1 (0.2)		\\ \\ \\ \\ \\ \\ \\ \\ \\ \\
##					\\ \\ \\ \\ \\ \\ \\ . : :
##	50	PDW\	Mean (sd) : 12.3 (2.2)\	167 distinct values	\\ \\ \\ \\ :\
##		[numeric]	min < med < max:\		\\ \\ \\ \\ : .\
##			6.6 < 12 < 25.3\		\\ \\ \\ \\ : :\
##			IQR (CV) : 2.6 (0.2)		\\ \\ : : : .\
##					\\ \\ : : : : .
##	51	RBC\	Mean (sd) : 3.9 (0.8)\	65 distinct values	\\ \\ \\ \\ \\ . :\
##		[numeric]	min < med < max:\		\\ \\ \\ \\ \\ : :\
##			1 < 3.9 < 8.2\		\\ \\ \\ \\ \\ : :\
##			IQR (CV) : 1.1 (0.2)		\\ \\ \\ \\ : : : :\
##					\\ \\ \\ \\ : : : : .

```
## | 52 | WBC\          | Mean (sd) : 11.2 (12.9)\      | 2710 distinct values | :\
## |    | [numeric]     | min < med < max:\          |                       | :\
## |    |               | 0 < 9.6 < 604.5\          |                       | :\
## |    |               | IQR (CV) : 6.9 (1.2)       |                       | :\
## |    |               |                           |                       | :
## +---+-----+-----+-----+-----+-----+
## | 53 | BloodCulture\ | 1\. no\                    | 13511 (92.0%)\       | IHHHHHHHHHHHHHHHH \
## |    | [character]   | 2\. yes                    | 1180 ( 8.0%)         | I
## +---+-----+-----+-----+-----+-----+-----+
```

Alternative: Use temporary directory explicitly

```
tmp_dir <- "data"
dfSummary(bacteremia_df,
  style = "grid",
  tmp.img.dir = tmp_dir)
```

```
## Data Frame Summary
## bacteremia_df
## Dimensions: 14691 x 53
## Duplicates: 0
##
```

```
## +---+-----+-----+-----+-----+-----+
## | No | Variable      | Stats / Values              | Freqs (% of Valid)    | Graph                  |
## +---+-----+-----+-----+-----+-----+
## | 1  | ID            | Mean (sd) : 29352.5 (18178.7) | 14691 distinct values | : . . . . .          |
## |    | [integer]     | min < med < max:            |                       | : : : : : . . . . . |
## |    |               | 1 < 28755 < 62455          |                       | : : : : : : : : : : |
## |    |               | IQR (CV) : 31083.5 (0.6)    |                       | : : : : : : : : : : |
## |    |               |                           |                       | : : : : : : : : : : |
## +---+-----+-----+-----+-----+-----+
## | 2  | SEX           | Min : 1                    | 1 : 8536 (58.1%)      | IHHHHHHHHH           |
## |    | [integer]     | Mean : 1.4                 | 2 : 6155 (41.9%)      | IHHHHHHH             |
## |    |               | Max : 2                    |                       |                       |
## +---+-----+-----+-----+-----+-----+
## | 3  | AGE           | Mean (sd) : 56.2 (18.2)     | 85 distinct values    | :                     |
## |    | [integer]     | min < med < max:            |                       | : : :                 |
## |    |               | 16 < 58 < 101              |                       | : : : : : : : : : : |
## |    |               | IQR (CV) : 27 (0.3)        |                       | : : : : : : : : : : |
## |    |               |                           |                       | : : : : : : : : : : |
## +---+-----+-----+-----+-----+-----+
## | 4  | MCV           | Mean (sd) : 88.4 (6.5)      | 506 distinct values   | :                     |
## |    | [numeric]     | min < med < max:            |                       | :                     |
## |    |               | 51 < 88.3 < 128.7          |                       | : :                 |
## |    |               | IQR (CV) : 7.3 (0.1)       |                       | : :                 |
## |    |               |                           |                       | : : : .             |
## +---+-----+-----+-----+-----+-----+
## | 5  | HGB           | Mean (sd) : 11.6 (2.3)      | 157 distinct values   | . :                   |
## |    | [numeric]     | min < med < max:            |                       | : : :                 |
## |    |               | 3 < 11.4 < 21              |                       | : : : .             |
## |    |               | IQR (CV) : 3.3 (0.2)       |                       | : : : :             |
## |    |               |                           |                       | : : : : .           |
## +---+-----+-----+-----+-----+-----+
## | 6  | HCT           | Mean (sd) : 34.5 (6.5)      | 404 distinct values   | : :                   |
## |    | [numeric]     | min < med < max:            |                       | : :                   |
```

##			0 < 34.3 < 66.6		:	:	.
##			IQR (CV) : 9.3 (0.2)		.	:	:
##					:	:	.
##	7	PLT	Mean (sd) : 220 (122.8)	718 distinct values	:		
##		[integer]	min < med < max:		:	:	
##			0 < 204 < 2092		:	:	
##			IQR (CV) : 137 (0.6)		:	:	
##					:	:	.
##	8	MCH	Mean (sd) : 29.6 (2.5)	232 distinct values	:		
##		[numeric]	min < med < max:		:		
##			14.9 < 29.7 < 47.4		:		
##			IQR (CV) : 2.6 (0.1)		.	:	.
##					.	:	:
##	9	MCHC	Mean (sd) : 33.5 (1.4)	124 distinct values	:	.	
##		[numeric]	min < med < max:		:	:	
##			23.7 < 33.5 < 43.5		:	:	
##			IQR (CV) : 1.8 (0)		:	:	
##					:	:	.
##	10	RDW	Mean (sd) : 15 (2.3)	173 distinct values	:		
##		[numeric]	min < med < max:		:		
##			10.6 < 14.5 < 31.8		:	:	
##			IQR (CV) : 2.6 (0.2)		:	:	
##					:	:	.
##	11	MPV	Mean (sd) : 10.4 (1)	71 distinct values	:	:	
##		[numeric]	min < med < max:		.	:	:
##			7.3 < 10.3 < 15		:	:	.
##			IQR (CV) : 1.3 (0.1)		:	:	:
##					:	:	.
##	12	LYM	Mean (sd) : 1.4 (7.5)	114 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 1 < 578.1		:		
##			IQR (CV) : 0.9 (5.5)		:		
##					:		
##	13	MONO	Mean (sd) : 0.9 (0.7)	67 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 0.8 < 20.4		:		
##			IQR (CV) : 0.6 (0.8)		:		
##					:		
##	14	EOS	Mean (sd) : 0.1 (0.3)	36 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 0.1 < 15.8		:		
##			IQR (CV) : 0.1 (2.3)		:		
##					:		
##	15	BASO	Mean (sd) : 0 (0.1)	18 distinct values	:		
##		[numeric]	min < med < max:		:		

##			0 < 0 < 6.5		:	
##			IQR (CV) : 0 (4.7)		:	
##					:	
##	16	NT	Mean (sd) : 83.2 (27.2)	149 distinct values	:	:
##		[integer]	min < med < max:		:	:
##			4 < 83 < 152		:	:
##			IQR (CV) : 34 (0.3)		:	:
##					:	:
##	17	APTT	Mean (sd) : 40.1 (11)	631 distinct values	:	:
##		[numeric]	min < med < max:		:	:
##			21.4 < 37.7 < 176.1		:	:
##			IQR (CV) : 8.6 (0.3)		:	:
##					:	:
##	18	FIB	Mean (sd) : 547.4 (208.1)	1084 distinct values	:	:
##		[integer]	min < med < max:		:	:
##			55 < 529 < 1593		:	:
##			IQR (CV) : 277 (0.4)		:	:
##					:	:
##	19	SODIUM	Mean (sd) : 137.2 (4.7)	58 distinct values	:	:
##		[integer]	min < med < max:		:	:
##			106 < 137 < 170		:	:
##			IQR (CV) : 5 (0)		:	:
##					:	:
##	20	POTASS	Mean (sd) : 4 (0.6)	408 distinct values	:	:
##		[numeric]	min < med < max:		:	:
##			1.9 < 4 < 36.6		:	:
##			IQR (CV) : 0.6 (0.2)		:	:
##					:	:
##	21	CA	Mean (sd) : 2.2 (0.2)	185 distinct values	:	:
##		[numeric]	min < med < max:		:	:
##			1 < 2.2 < 4.4		:	:
##			IQR (CV) : 0.3 (0.1)		:	:
##					:	:
##	22	PHOS	Mean (sd) : 1 (0.4)	306 distinct values	:	:
##		[numeric]	min < med < max:		:	:
##			0.3 < 1 < 6.2		:	:
##			IQR (CV) : 0.4 (0.4)		:	:
##					:	:
##	23	MG	Mean (sd) : 0.8 (0.2)	146 distinct values	:	:
##		[numeric]	min < med < max:		:	:
##			0.2 < 0.8 < 2.2		:	:
##			IQR (CV) : 0.2 (0.2)		:	:
##					:	:
##	24	CREA	Mean (sd) : 1.3 (1.2)	674 distinct values	:	:
##		[numeric]	min < med < max:		:	:

##			0.3 < 1 < 20.8		:	
##			IQR (CV) : 0.5 (0.9)		:	
##					:	.
##	25	BUN	Mean (sd) : 22.7 (18.1)	947 distinct values	:	
##		[numeric]	min < med < max:		:	
##			2.5 < 16.6 < 184.8		:	
##			IQR (CV) : 15.3 (0.8)		:	:
##					:	:
##	26	HS	Mean (sd) : 5.4 (2.5)	169 distinct values	:	
##		[numeric]	min < med < max:		:	.
##			1.3 < 5 < 22.7		:	:
##			IQR (CV) : 2.9 (0.5)		:	:
##					:	:
##	27	GBIL	Mean (sd) : 1.4 (2.8)	885 distinct values	:	
##		[numeric]	min < med < max:		:	
##			0.1 < 0.8 < 51.8		:	
##			IQR (CV) : 0.7 (2)		:	
##					:	
##	28	TP	Mean (sd) : 64.9 (11.5)	649 distinct values	:	.
##		[numeric]	min < med < max:		:	:
##			29.9 < 65.7 < 120.9		:	:
##			IQR (CV) : 16.4 (0.2)		:	:
##					:	:
##	29	ALB	Mean (sd) : 33.4 (7.4)	401 distinct values	:	.
##		[numeric]	min < med < max:		:	:
##			10 < 33.6 < 55.7		:	:
##			IQR (CV) : 11.2 (0.2)		:	:
##					:	:
##	30	AMY	Mean (sd) : 90.8 (805.2)	488 distinct values	:	
##		[integer]	min < med < max:		:	
##			8 < 49 < 56146		:	
##			IQR (CV) : 43 (8.9)		:	
##					:	
##	31	PAMY	Mean (sd) : 41.7 (448)	280 distinct values	:	
##		[integer]	min < med < max:		:	
##			1 < 22 < 38369		:	
##			IQR (CV) : 22 (10.8)		:	
##					:	
##	32	LIP	Mean (sd) : 63.8 (603.9)	444 distinct values	:	
##		[integer]	min < med < max:		:	
##			0 < 23 < 45991		:	
##			IQR (CV) : 26 (9.5)		:	
##					:	
##	33	CHE	Mean (sd) : 4.8 (2.1)	997 distinct values	:	.
##		[numeric]	min < med < max:		:	.

##			1 < 4.6 < 13.9		. : : : :	
##			IQR (CV) : 3.1 (0.4)		: : : : :	
##					: : : : : .	
##	34	AP	Mean (sd) : 118.8 (132.9)	672 distinct values	:	
##		[integer]	min < med < max:		:	
##			11 < 84 < 2995		:	
##			IQR (CV) : 60 (1.1)		:	
##					:	
##	35	ASAT	Mean (sd) : 86.9 (404.7)	650 distinct values	:	
##		[integer]	min < med < max:		:	
##			3 < 31 < 13991		:	
##			IQR (CV) : 34 (4.7)		:	
##					:	
##	36	ALAT	Mean (sd) : 67.7 (311)	578 distinct values	:	
##		[integer]	min < med < max:		:	
##			0 < 26 < 15059		:	
##			IQR (CV) : 32 (4.6)		:	
##					:	
##	37	GGT	Mean (sd) : 115.1 (209)	858 distinct values	:	
##		[integer]	min < med < max:		:	
##			3 < 49 < 5171		:	
##			IQR (CV) : 92 (1.8)		:	
##					:	
##	38	LDH	Mean (sd) : 331.2 (475.2)	1137 distinct values	:	
##		[integer]	min < med < max:		:	
##			39 < 239 < 13906		:	
##			IQR (CV) : 145 (1.4)		:	
##					:	
##	39	CK	Mean (sd) : 385 (2241.2)	1506 distinct values	:	
##		[integer]	min < med < max:		:	
##			8 < 80 < 98801		:	
##			IQR (CV) : 142 (5.8)		:	
##					:	
##	40	GLU	Mean (sd) : 126.4 (56.9)	389 distinct values	:	
##		[integer]	min < med < max:		:	
##			19 < 113 < 1403		:	
##			IQR (CV) : 41 (0.5)		:	
##					: :	
##	41	TRIG	Mean (sd) : 141.7 (120.7)	538 distinct values	:	
##		[integer]	min < med < max:		:	
##			14 < 115 < 5440		:	
##			IQR (CV) : 82 (0.9)		:	
##					:	
##	42	CHOL	Mean (sd) : 150.8 (55.5)	339 distinct values	:	
##		[integer]	min < med < max:		. :	

##			25 < 145 < 1104		:	:	
##			IQR (CV) : 69 (0.4)		:	:	
##					:	:	.
##	43	CRP	Mean (sd) : 10.9 (9.6)	3328 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 8.6 < 76.3		:	:	
##			IQR (CV) : 13.6 (0.9)		:	:	.
##					:	:	.
##	44	BASOR	Mean (sd) : 0.1 (0.6)	419 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 0 < 23.7		:		
##			IQR (CV) : 0 (4)		:		
##					:		
##	45	EOSR	Mean (sd) : 1.3 (2.4)	927 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 0.6 < 73.5		:		
##			IQR (CV) : 1.8 (1.8)		:		
##					:		
##	46	LYMR	Mean (sd) : 14.6 (12.7)	3121 distinct values	:		
##		[numeric]	min < med < max:		:	:	
##			0 < 11.3 < 100		:	:	
##			IQR (CV) : 11.4 (0.9)		:	:	.
##					:	:	.
##	47	MONOR	Mean (sd) : 8.8 (5.8)	2334 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 8 < 100		:		
##			IQR (CV) : 5.2 (0.7)		:	:	
##					:	:	
##	48	NEU	Mean (sd) : 8.4 (5.6)	374 distinct values	:		
##		[numeric]	min < med < max:		:		
##			0 < 7.3 < 83.8		:	:	
##			IQR (CV) : 6.2 (0.7)		:	:	
##					:	:	.
##	49	NEUR	Mean (sd) : 75.2 (15.5)	3850 distinct values		.	:
##		[numeric]	min < med < max:			:	:
##			0 < 78.3 < 100			:	:
##			IQR (CV) : 16.1 (0.2)			:	:
##						.	:
##	50	PDW	Mean (sd) : 12.3 (2.2)	167 distinct values	:		
##		[numeric]	min < med < max:		:	.	
##			6.6 < 12 < 25.3		:	:	
##			IQR (CV) : 2.6 (0.2)		:	:	.
##					:	:	.
##	51	RBC	Mean (sd) : 3.9 (0.8)	65 distinct values		.	:
##		[numeric]	min < med < max:			:	:

```
# Or try with the "multiline" style (doesn't need images)
dfSummary(bacteremia_df, style = "multiline")
```

```
## Data Frame Summary
## bacteremia_df
## Dimensions: 14691 x 53
## Duplicates: 0
##
## -----
## No      Variable      Stats / Values      Freqs (% of Valid)      Graph      V
## -----
## 1      ID      Mean (sd) : 29352.5 (18178.7)      14691 distinct values      : . . . . .      1
##      [integer]      min < med < max:      : : : : : . .      (
##      1 < 28755 < 62455      : : : : : : : : :
##      IQR (CV) : 31083.5 (0.6)      : : : : : : : : :
##      : : : : : : : : :
##
## 2      SEX      Min : 1      1 : 8536 (58.1%)      IIIIIIIIIII      1
##      [integer]      Mean : 1.4      2 : 6155 (41.9%)      IIIIIIII      (
##      Max : 2
##
## 3      AGE      Mean (sd) : 56.2 (18.2)      85 distinct values      :      1
##      [integer]      min < med < max:      . : .      (
##      16 < 58 < 101      . : : : : :
##      IQR (CV) : 27 (0.3)      . : : : : : :
##      : : : : : : : : :
##
## 4      MCV      Mean (sd) : 88.4 (6.5)      506 distinct values      :      1
##      [numeric]      min < med < max:      :      (
##      51 < 88.3 < 128.7      : :
##      IQR (CV) : 7.3 (0.1)      : :
##      : : : .
##
## 5      HGB      Mean (sd) : 11.6 (2.3)      157 distinct values      . :      1
##      [numeric]      min < med < max:      : : :      (
##      3 < 11.4 < 21      : : : .
##      IQR (CV) : 3.3 (0.2)      : : : :
##      : : : : : .
##
## 6      HCT      Mean (sd) : 34.5 (6.5)      404 distinct values      : :      1
```

```
##      [numeric]      min < med < max:      : :      (9
##      0 < 34.3 < 66.6      : : .
##      IQR (CV) : 9.3 (0.2)      . : : :
##      : : : : .
##
## 7    PLT      Mean (sd) : 220 (122.8)      718 distinct values      :      14
##      [integer]      min < med < max:      : :      (9
##      0 < 204 < 2092      : :
##      IQR (CV) : 137 (0.6)      : :
##      : : .
##
## 8    MCH      Mean (sd) : 29.6 (2.5)      232 distinct values      :      14
##      [numeric]      min < med < max:      :      (9
##      14.9 < 29.7 < 47.4      :
##      IQR (CV) : 2.6 (0.1)      . : .
##      . : : :
##
## 9    MCHC      Mean (sd) : 33.5 (1.4)      124 distinct values      : .      14
##      [numeric]      min < med < max:      : :      (9
##      23.7 < 33.5 < 43.5      : :
##      IQR (CV) : 1.8 (0)      : :
##      : : : .
##
## 10   RDW      Mean (sd) : 15 (2.3)      173 distinct values      :      14
##      [numeric]      min < med < max:      :      (9
##      10.6 < 14.5 < 31.8      : :
##      IQR (CV) : 2.6 (0.2)      : :
##      : : : : .
##
## 11   MPV      Mean (sd) : 10.4 (1)      71 distinct values      : :      13
##      [numeric]      min < med < max:      . : :      (9
##      7.3 < 10.3 < 15      : : : .
##      IQR (CV) : 1.3 (0.1)      : : : :
##      : : : : : .
##
## 12   LYM      Mean (sd) : 1.4 (7.5)      114 distinct values      :      14
##      [numeric]      min < med < max:      :      (9
##      0 < 1 < 578.1      :
##      IQR (CV) : 0.9 (5.5)      :
##      :
##
## 13   MONO      Mean (sd) : 0.9 (0.7)      67 distinct values      :      14
##      [numeric]      min < med < max:      :      (9
##      0 < 0.8 < 20.4      :
##      IQR (CV) : 0.6 (0.8)      :
##      :
##
## 14   EOS      Mean (sd) : 0.1 (0.3)      36 distinct values      :      14
##      [numeric]      min < med < max:      :      (9
##      0 < 0.1 < 15.8      :
##      IQR (CV) : 0.1 (2.3)      :
##      :
##
## 15   BASO      Mean (sd) : 0 (0.1)      18 distinct values      :      14
```

```
##      [numeric]      min < med < max:      :      (9)
##      0 < 0 < 6.5      :
##      IQR (CV) : 0 (4.7)      :
##      :
##
## 16  NT      Mean (sd) : 83.2 (27.2)      149 distinct values      : :      12
##      [integer]      min < med < max:      : : :      (8)
##      4 < 83 < 152      . : : : .
##      IQR (CV) : 34 (0.3)      . : : : :
##      . . : : : : : .
##
## 17  APTT      Mean (sd) : 40.1 (11)      631 distinct values      . :      12
##      [numeric]      min < med < max:      : :      (8)
##      21.4 < 37.7 < 176.1      : :
##      IQR (CV) : 8.6 (0.3)      : :
##      : : .
##
## 18  FIB      Mean (sd) : 547.4 (208.1)      1084 distinct values      : .      12
##      [integer]      min < med < max:      : :      (8)
##      55 < 529 < 1593      : : : :
##      IQR (CV) : 277 (0.4)      : : : :
##      . : : : : .
##
## 19  SODIUM      Mean (sd) : 137.2 (4.7)      58 distinct values      :      13
##      [integer]      min < med < max:      : .      (9)
##      106 < 137 < 170      : :
##      IQR (CV) : 5 (0)      : :
##      : : : .
##
## 20  POTASS      Mean (sd) : 4 (0.6)      408 distinct values      :      12
##      [numeric]      min < med < max:      :      (8)
##      1.9 < 4 < 36.6      :
##      IQR (CV) : 0.6 (0.2)      :
##      :
##
## 21  CA      Mean (sd) : 2.2 (0.2)      185 distinct values      :      13
##      [numeric]      min < med < max:      :      (9)
##      1 < 2.2 < 4.4      :
##      IQR (CV) : 0.3 (0.1)      . : .
##      : : :
##
## 22  PHOS      Mean (sd) : 1 (0.4)      306 distinct values      :      13
##      [numeric]      min < med < max:      . :      (9)
##      0.3 < 1 < 6.2      : :
##      IQR (CV) : 0.4 (0.4)      : :
##      : : .
##
## 23  MG      Mean (sd) : 0.8 (0.2)      146 distinct values      : :      12
##      [numeric]      min < med < max:      : :      (8)
##      0.2 < 0.8 < 2.2      : :
##      IQR (CV) : 0.2 (0.2)      : :
##      . : : :
##
## 24  CREA      Mean (sd) : 1.3 (1.2)      674 distinct values      :      14
```

```
##      [numeric]      min < med < max:      :      (9
##      0.3 < 1 < 20.8      :      (8
##      IQR (CV) : 0.5 (0.9)      :      (7
##      : .      :      (6
##
## 25  BUN      Mean (sd) : 22.7 (18.1)      947 distinct values      :      14
##      [numeric]      min < med < max:      :      (9
##      2.5 < 16.6 < 184.8      :      (8
##      IQR (CV) : 15.3 (0.8)      : :      (7
##      : : . .      :      (6
##
## 26  HS      Mean (sd) : 5.4 (2.5)      169 distinct values      :      11
##      [numeric]      min < med < max:      : .      (7
##      1.3 < 5 < 22.7      . : :      (6
##      IQR (CV) : 2.9 (0.5)      : : :      (5
##      : : : : .      :      (4
##
## 27  GBIL      Mean (sd) : 1.4 (2.8)      885 distinct values      :      13
##      [numeric]      min < med < max:      :      (9
##      0.1 < 0.8 < 51.8      :      (8
##      IQR (CV) : 0.7 (2)      :      (7
##      :      :      (6
##
## 28  TP      Mean (sd) : 64.9 (11.5)      649 distinct values      . :      13
##      [numeric]      min < med < max:      : :      (8
##      29.9 < 65.7 < 120.9      : : : .      (7
##      IQR (CV) : 16.4 (0.2)      : : : :      (6
##      . : : : : .      :      (5
##
## 29  ALB      Mean (sd) : 33.4 (7.4)      401 distinct values      . : .      13
##      [numeric]      min < med < max:      : : : :      (8
##      10 < 33.6 < 55.7      : : : : .      (7
##      IQR (CV) : 11.2 (0.2)      : : : : :      (6
##      . : : : : : .      :      (5
##
## 30  AMY      Mean (sd) : 90.8 (805.2)      488 distinct values      :      10
##      [integer]      min < med < max:      :      (7
##      8 < 49 < 56146      :      (6
##      IQR (CV) : 43 (8.9)      :      (5
##      :      :      (4
##
## 31  PAMY      Mean (sd) : 41.7 (448)      280 distinct values      :      75
##      [integer]      min < med < max:      :      (5
##      1 < 22 < 38369      :      (4
##      IQR (CV) : 22 (10.8)      :      (3
##      :      :      (2
##
## 32  LIP      Mean (sd) : 63.8 (603.9)      444 distinct values      :      10
##      [integer]      min < med < max:      :      (7
##      0 < 23 < 45991      :      (6
##      IQR (CV) : 26 (9.5)      :      (5
##      :      :      (4
##
## 33  CHE      Mean (sd) : 4.8 (2.1)      997 distinct values      : : .      12
```

```
##      [numeric]      min < med < max:      : : : .      (8
##      1 < 4.6 < 13.9      . : : : :
##      IQR (CV) : 3.1 (0.4)      : : : : :
##      : : : : : .
##
## 34  AP      Mean (sd) : 118.8 (132.9)      672 distinct values      :      13
##      [integer]      min < med < max:      :      (9
##      11 < 84 < 2995      :
##      IQR (CV) : 60 (1.1)      :
##      :
##
## 35  ASAT      Mean (sd) : 86.9 (404.7)      650 distinct values      :      13
##      [integer]      min < med < max:      :      (9
##      3 < 31 < 13991      :
##      IQR (CV) : 34 (4.7)      :
##      :
##
## 36  ALAT      Mean (sd) : 67.7 (311)      578 distinct values      :      13
##      [integer]      min < med < max:      :      (9
##      0 < 26 < 15059      :
##      IQR (CV) : 32 (4.6)      :
##      :
##
## 37  GGT      Mean (sd) : 115.1 (209)      858 distinct values      :      13
##      [integer]      min < med < max:      :      (9
##      3 < 49 < 5171      :
##      IQR (CV) : 92 (1.8)      :
##      :
##
## 38  LDH      Mean (sd) : 331.2 (475.2)      1137 distinct values      :      12
##      [integer]      min < med < max:      :      (8
##      39 < 239 < 13906      :
##      IQR (CV) : 145 (1.4)      :
##      :
##
## 39  CK      Mean (sd) : 385 (2241.2)      1506 distinct values      :      12
##      [integer]      min < med < max:      :      (8
##      8 < 80 < 98801      :
##      IQR (CV) : 142 (5.8)      :
##      :
##
## 40  GLU      Mean (sd) : 126.4 (56.9)      389 distinct values      :      10
##      [integer]      min < med < max:      :      (7
##      19 < 113 < 1403      :
##      IQR (CV) : 41 (0.5)      :
##      : :
##
## 41  TRIG      Mean (sd) : 141.7 (120.7)      538 distinct values      :      9
##      [integer]      min < med < max:      :      (6
##      14 < 115 < 5440      :
##      IQR (CV) : 82 (0.9)      :
##      :
##
## 42  CHOL      Mean (sd) : 150.8 (55.5)      339 distinct values      :      9
```

##	[integer]	min < med < max:	. :	(6
##		25 < 145 < 1104	: :	
##		IQR (CV) : 69 (0.4)	: :	
##			: : .	
##				
## 43	CRP	Mean (sd) : 10.9 (9.6)	3328 distinct values	: 14
##	[numeric]	min < med < max:	:	(9
##		0 < 8.6 < 76.3	: :	
##		IQR (CV) : 13.6 (0.9)	: : .	
##			: : : .	
##				
## 44	BASOR	Mean (sd) : 0.1 (0.6)	419 distinct values	: 13
##	[numeric]	min < med < max:	:	(9
##		0 < 0 < 23.7	:	
##		IQR (CV) : 0 (4)	:	
##			:	
##				
## 45	EOSR	Mean (sd) : 1.3 (2.4)	927 distinct values	: 13
##	[numeric]	min < med < max:	:	(9
##		0 < 0.6 < 73.5	:	
##		IQR (CV) : 1.8 (1.8)	:	
##			:	
##				
## 46	LYMR	Mean (sd) : 14.6 (12.7)	3121 distinct values	: 13
##	[numeric]	min < med < max:	: :	(9
##		0 < 11.3 < 100	: :	
##		IQR (CV) : 11.4 (0.9)	: : .	
##			: : : .	
##				
## 47	MONOR	Mean (sd) : 8.8 (5.8)	2334 distinct values	: 13
##	[numeric]	min < med < max:	:	(9
##		0 < 8 < 100	:	
##		IQR (CV) : 5.2 (0.7)	: :	
##			: :	
##				
## 48	NEU	Mean (sd) : 8.4 (5.6)	374 distinct values	: 13
##	[numeric]	min < med < max:	:	(9
##		0 < 7.3 < 83.8	: :	
##		IQR (CV) : 6.2 (0.7)	: :	
##			: : .	
##				
## 49	NEUR	Mean (sd) : 75.2 (15.5)	3850 distinct values	. : 13
##	[numeric]	min < med < max:	: :	(9
##		0 < 78.3 < 100	: :	
##		IQR (CV) : 16.1 (0.2)	: : : .	
##			. : : : : :	
##				
## 50	PDW	Mean (sd) : 12.3 (2.2)	167 distinct values	: 13
##	[numeric]	min < med < max:	: .	(9
##		6.6 < 12 < 25.3	: :	
##		IQR (CV) : 2.6 (0.2)	: : : .	
##			: : : : .	
##				
## 51	RBC	Mean (sd) : 3.9 (0.8)	65 distinct values	. : 14


```
##      [numeric]      min < med < max:      : :      (S
##      1 < 3.9 < 8.2      : :
##      IQR (CV) : 1.1 (0.2)      : : : :
##      : : : : .
##
## 52   WBC      Mean (sd) : 11.2 (12.9)      2710 distinct values :      14
##      [numeric]      min < med < max:      :      (S
##      0 < 9.6 < 604.5      :
##      IQR (CV) : 6.9 (1.2)      :
##      :
##
## 53   BloodCulture  1. no      13511 (92.0%)      IIIIIIIIIIIIIIIIIIIII 14
##      [character]  2. yes      1180 ( 8.0%)      I      (1
## -----
```

De las características observamos que existen 3 variables del tipo “integer”, del tipo double, y el resto, 10 variables del tipo objeto o character.

A simple vista las que podrían ser variables categoróricas, en realidad son características binarias (TRUE o FALSE), como podríamos interpolar por 1,0 si deseamos.

Podemos revisar la descripción de las variables contenidas en el fichero y si los tipos de variables se corresponden con las que hemos cargado. Las organizamos lógicamente para darles sentido y construimos un pequeño diccionario de datos utilizando la documentación auxiliar.

Respecto al significado de cada una de las features utilizaremos el fichero DataDictionary - <https://zenodo.org/records/7554815/files/bacteremia-DataDictionary.csv> disponible en la misma página oficial para interpretar los valores.

```
# Read directly from URL
dictionary_df <- read.csv("https://zenodo.org/records/7554815/files/bacteremia-DataDictionary.csv?download=1")

## # Display as a nice formatted table
# library(DT)
# datatable(dictionary_df,
#             options = list(pageLength = 20,
#                             scrollX = TRUE,
#                             searchHighlight = TRUE),
#             rownames = FALSE)

# Verificación de valores perdidos
na_count <- sapply(bacteremia_df, function(x) sum(is.na(x)))
print(na_count[na_count > 0])
```

##	MCV	HGB	HCT	PLT	MCH	MCHC	RDW	MPV	LYM	MONO	EOS
##	42	41	42	42	42	42	56	702	262	246	135
##	BASO	NT	APTT	FIB	SODIUM	POTASS	CA	PHOS	MG	CREA	BUN
##	146	2467	2549	2567	1282	2008	1276	1242	1869	159	172
##	HS	GBIL	TP	ALB	AMY	PAMY	LIP	CHE	AP	ASAT	ALAT
##	3061	1441	1583	1676	3913	7114	3699	2447	1400	1154	987
##	GGT	LDH	CK	GLU	TRIG	CHOL	CRP	BASOR	EOSR	LYMR	MONOR
##	1262	1714	2080	4192	5061	5045	155	732	732	732	732
##	NEU	NEUR	PDW	RBC	WBC						
##	728	732	1102	461	462						

2.2 2.2 Preguntas “objetivo” (1 punto)

ver LAB3 y PEC1

3 3. Análisis exploratorio de los datos (2,5 puntos)

3.1 3.1 Análisis descriptivo y gráfico (1 punto)

El primer dato que tenemos es la proporción que hay entre las dos clases:

```
class_ratio <- prop.table(table(bacteremia_df$BloodCulture))
sprintf("La proporción de la variable dependiente es: %.2f
        para los valores %s %.2f y para los valores %s.", class_ratio[1]*100, names(class_ratio)[1], cla
        names(class_ratio)[2])
```

```
## [1] "La proporción de la variable dependiente es: 91.97 \n          para los valores no 8.03 y para los
```

3.2 3.2. Ejercicios de inferencia y simulación (1,5 puntos)

ver LAB3, LAB4, PEC2

4 4. Modelos de aprendizaje automático (2,5 puntos)

ver LAB5 y PEC3

5 5. Visualización (1,5 puntos)

ver LAB6 y PEC3

6 6. Conclusiones (0,5 puntos)

7 Bibliografía

References

1. Wolberg, W.H., and Mangasarian, O.L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In *Proceedings of the National Academy of Sciences*, 87, 9193-9196.

- Size of data set: only 369 instances (at that point in time)
- Collected classification results: 1 trial only
- Two pairs of parallel hyperplanes were found to be consistent with 50% of the data
- Accuracy on remaining 50% of dataset: 93.5%
- Three pairs of parallel hyperplanes were found to be consistent with 67% of data
- Accuracy on remaining 33% of dataset: 95.9%

2. Zhang, J. (1992). Selecting typical instances in instance-based learning. In *Proceedings of the Ninth International Machine Learning Conference* (pp. 470-479). Aberdeen, Scotland: Morgan Kaufmann.

- Size of data set: only 369 instances (at that point in time)
- Applied 4 instance-based learning algorithms
- Collected classification results averaged over 10 trials
- Best accuracy result:
- 1-nearest neighbor: 93.7%
- trained on 200 instances, tested on the other 169
- Also of interest:
- Using only typical instances: 92.2% (storing only 23.1 instances)
- trained on 200 instances, tested on the other 169

Blake, C.L. & Merz, C.J. (1998). *UCI Repository of Machine Learning Databases*. Irvine, CA: University of California, Irvine, Department of Information and Computer Science. Formerly available from <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

- <https://shiny.posit.co/py/templates/>
-