

# Analysis of the age variable

José Carlos Soares Junior

2023

## Contents

<b>R packages</b>	<b>2</b>
<b>Data preparation</b>	<b>3</b>
<b>Quantitative analysis</b>	<b>3</b>
Analysis of age(days) . . . . .	3
Analysis of age(months) . . . . .	6
<b>Grouped age</b>	<b>8</b>
First case . . . . .	8
Second case . . . . .	8
<b>Analysis by evolution(death = yes)</b>	<b>9</b>
Quantitative analysis . . . . .	9
Age(days) . . . . .	9
Age(months) . . . . .	11
Grouped age . . . . .	13
First case . . . . .	13
Second case . . . . .	14

## R packages

```
# Load packages
loadlibrary <- function(x) {
  if (!require(x, character.only = TRUE)) {
    install.packages(x, dependencies = T)
  }
  if (!require(x, character.only = TRUE))
    stop("Package not found")
}

packages <- c(
  "readr",
  "readxl",
  "janitor",
  "dplyr",
  "forcats",
  "stringr",
  "lubridate",
  "summarytools",
  "magrittr",
  "questionr",
  "knitr",
  "data.table",
  "writexl",
  "modelsummary",
  "abjData",
  "awek",
  "ggplot2",
  "viridis"
)

lapply(packages, loadlibrary)

# Global changes for the summarytools packages
st_options(freq.cumul = FALSE,
            display.labels = TRUE,
            lang = "pt",
            bootstrap.css = FALSE,
            plain.ascii = FALSE,
            dfSummary.silent = TRUE,
            footnote = NA,
            subtitle.emphasis = FALSE,
            headings = FALSE)

st_css()
```

## Data preparation

```
##### Loading the data #####  
data_age <- readRDS("data_paper_1000d.rds")
```

```
##### age variable #####
```

```
##### First case #####
```

```
# <= 30 days (1 month);  
# 31 days |- 180 days (between 1 and 6 months);  
# 181 days |- 365 days (between 6 and 12 months);  
# >= 366 days (more than 12 months).
```

```
data_age1 <- data_age %>%  
  mutate(age_months_firstcase = case_when(  
    idade_dias <= 30 ~ "1 month",  
    idade_dias > 30 & idade_dias <= 180 ~ "between 1 and 6 months",  
    idade_dias > 180 & idade_dias <= 365 ~ "between 6 and 12 months",  
    idade_dias > 365 ~ "more than 12 months")  
  )
```

```
##### Second case #####
```

```
# <= 30 days (1 month);  
# 31 days |- 180 days (between 1 and 6 months);  
# 181 days |- 365 days (between 6 and 12 months);  
# 366 days |- 545 days (between 12 and 18 months);  
# >= 546 days (more than 18 months).
```

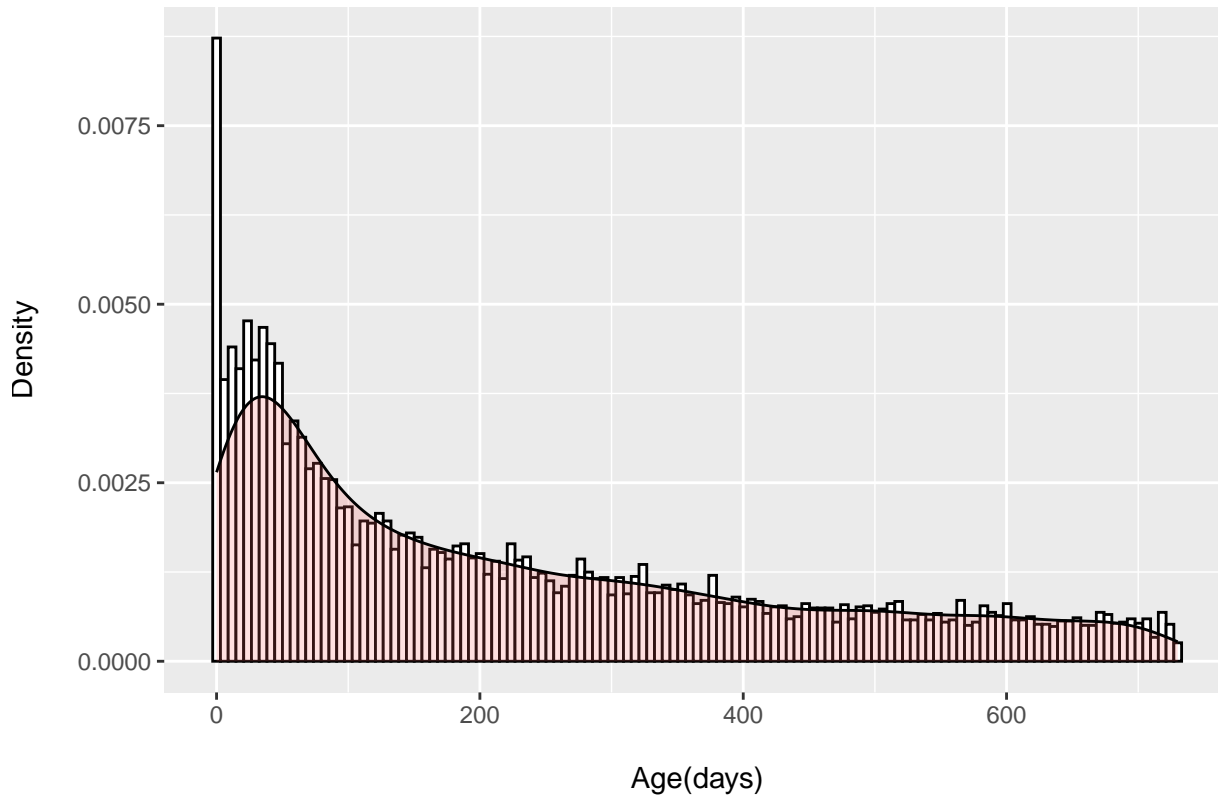
```
data_age2 <- data_age1 %>%  
  mutate(age_months_secondcase = case_when(  
    idade_dias <= 30 ~ "1 month",  
    idade_dias > 30 & idade_dias <= 180 ~ "between 1 and 6 months",  
    idade_dias > 180 & idade_dias <= 365 ~ "between 6 and 12 months",  
    idade_dias > 365 & idade_dias <= 545 ~ "between 12 and 18 months",  
    idade_dias > 545 ~ "more than 18 months")  
  )
```

## Quantitative analysis

### Analysis of age(days)

```
ggplot(data_age2, aes(x = idade_dias)) +  
  geom_histogram(aes(y = after_stat(density)),  
    bins = 125,  
    colour = "black",  
    fill = "white") +  
  geom_density(alpha = .2, fill = "#FF6666") +  
  labs(x = "\nAge(days)",  
    y = "Density\n",  
    title = "Histogram of age(days)")
```

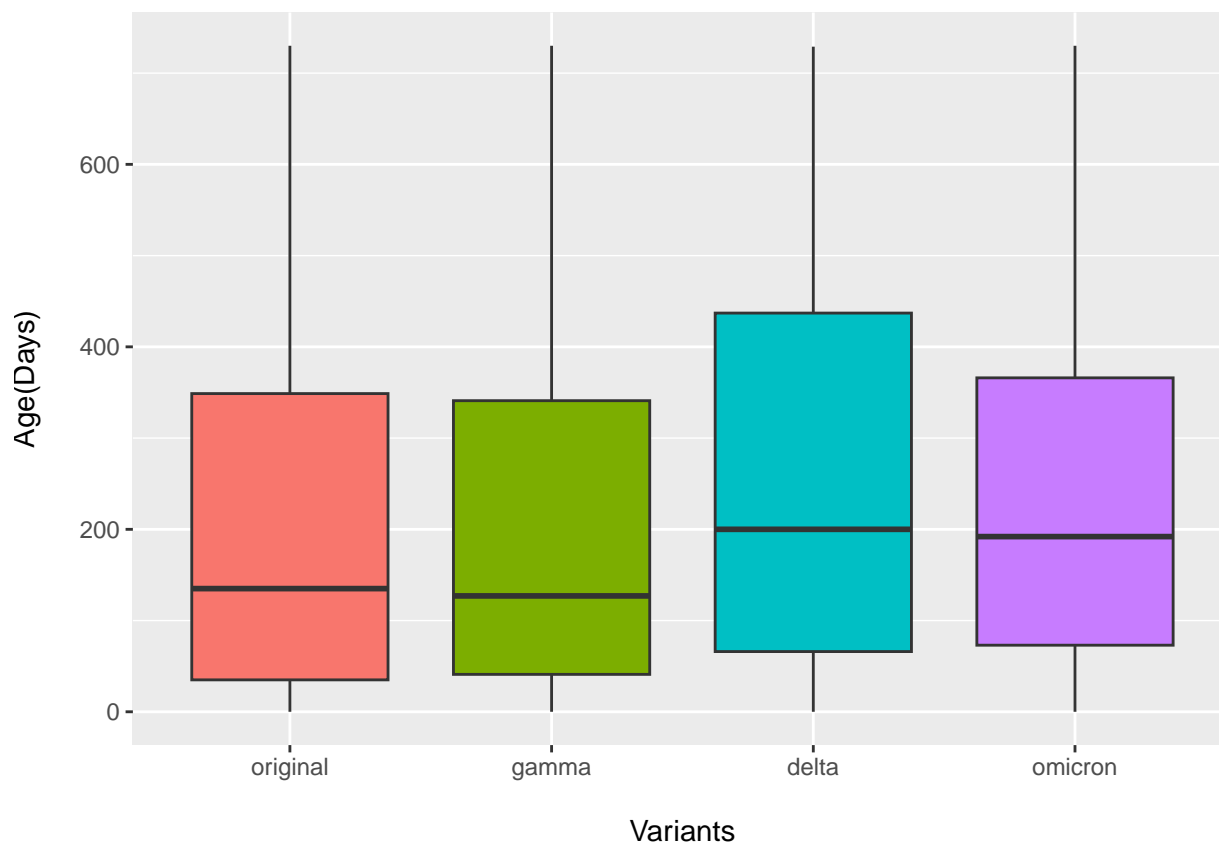
Histogram of age(days)



**Figure 1:** Histogram of age in unit of days (bars) and density (curve with red area).

```
data_age2$variants <- factor(data_age2$variants,
                             levels = c("original",
                                           "gamma",
                                           "delta",
                                           "omicron"))

data_age2 %>%
  ggplot(aes(x = variants,
             y = idade_dias,
             fill = variants)) +
  geom_boxplot() +
  xlab("\nVariants") +
  ylab("Age(Days)\n") +
  theme(legend.position = "none")
```



**Figure 2:** Boxplot of variants by age in a unit of days.

```
# Summary measurement functions
media <- function(x)
  mean(x, na.rm = TRUE)
mediana <- function(x)
  median(x, na.rm = TRUE)
DP <- function(x)
  sd(x, na.rm = TRUE)
minimo <- function(x)
  base::min(x, na.rm = TRUE)
maximo <- function(x)
  base::max(x, na.rm = TRUE)
q25 <- function(x)
  stats::quantile(x, p = 0.25, na.rm = TRUE)
q75 <- function(x)
  stats::quantile(x, p = 0.75, na.rm = TRUE)
IQR <- function(x)
  round(q75(x) - q25(x), 2)
n <- function(x)
  sum(!is.na(x))
```

```
datasummary(variants ~ idade_dias*(n + media + DP + mediana + minimo +
                                     maximo + q25 + q75 + IQR),
            data = data_age2, output = 'markdown')
```

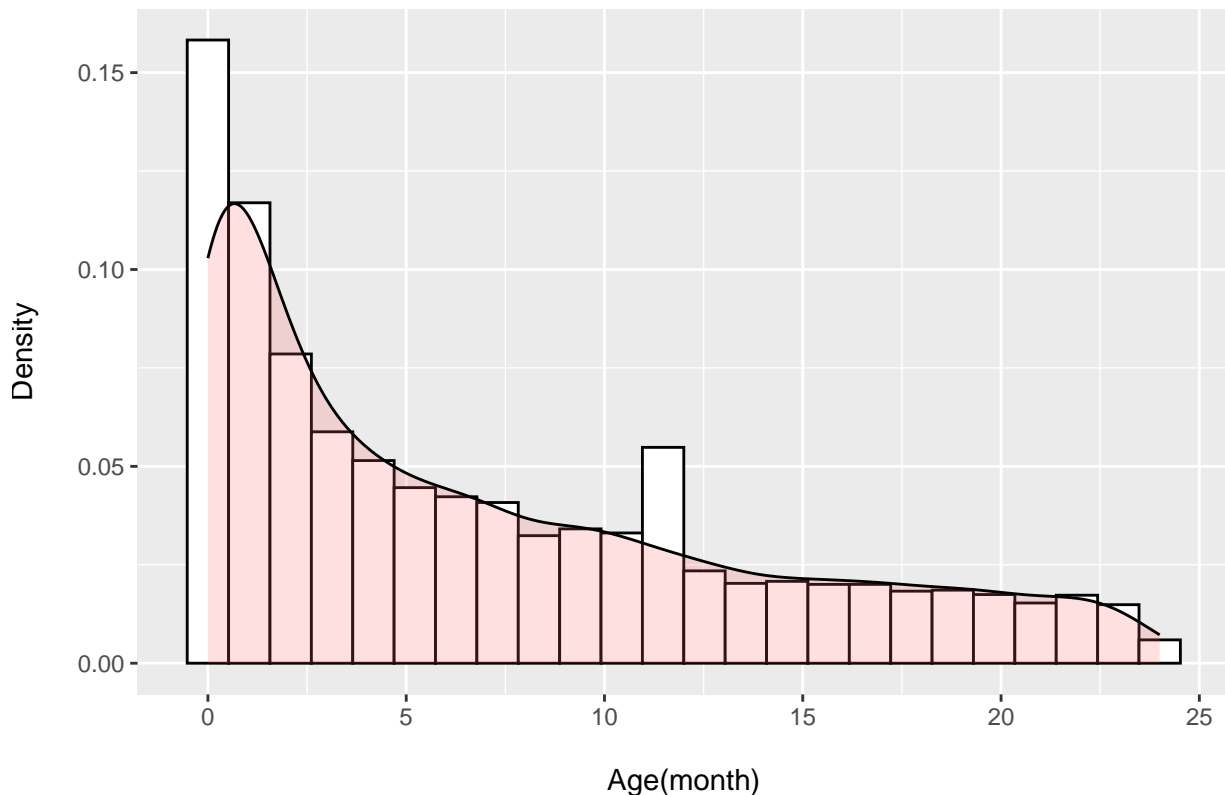
variants	n	media	DP	mediana	minimo	maximo	q25	q75	IQR
original	3706	210.63	207.17	135.00	0.00	730.00	35.00	348.75	313.75
gamma	2941	208.77	204.84	127.00	0.00	730.00	41.00	341.00	300.00
delta	795	259.04	215.66	200.00	0.00	729.00	66.00	437.00	371.00
omicron	3711	242.08	196.30	192.00	0.00	730.00	73.00	366.00	293.00

**Table 1:** Descriptive metrics of age (days) by variants.

## Analysis of age(months)

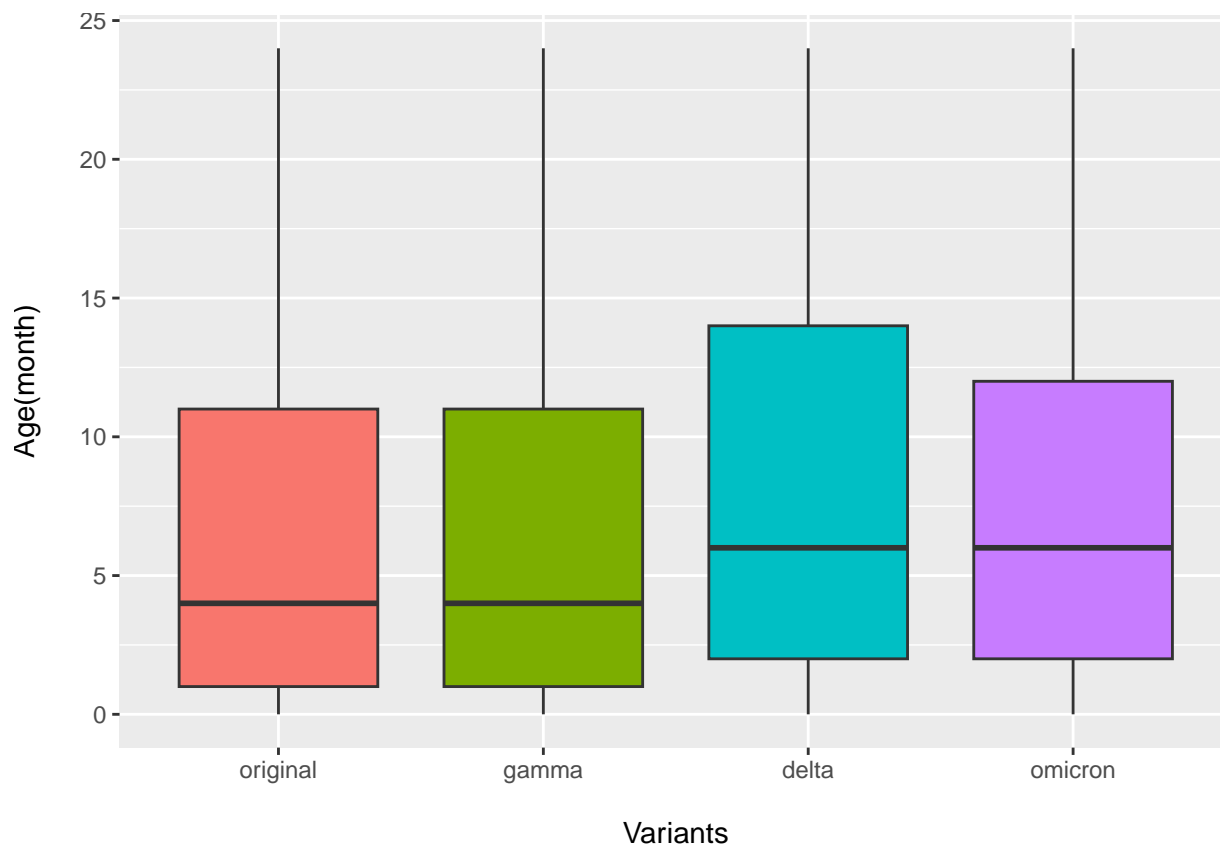
```
ggplot(data_age2, aes(x = age_month)) +
  geom_histogram(aes(y = after_stat(density)),
                bins = 24,
                colour = "black",
                fill = "white") +
  geom_density(alpha = .2, fill = "#FF6666") +
  labs(x = "\nAge(month)",
       y = "Density\n",
       title = "Histogram of age(month)")
```

**Histogram of age(month)**



**Figure 3:** Histogram of age in unit of months (bars) and density (curve with red area).

```
data_age2 %>%
  ggplot(aes(x = variants,
             y = age_month,
             fill = variants)) +
  geom_boxplot() +
  xlab("\nVariants") +
  ylab("Age(month)\n") +
  theme(legend.position = "none")
```



**Figure 4:** Boxplot of variants by age in a unit of months.

```
datasummary(variants ~ age_month*(n + media + DP + mediana + minimo +
                                maximo + q25 + q75 + IQR),
            data = data_age2, output = 'markdown')
```

variants	n	media	DP	mediana	minimo	maximo	q25	q75	IQR
original	3706	6.59	6.87	4.00	0.00	24.00	1.00	11.00	10.00
gamma	2941	6.51	6.80	4.00	0.00	24.00	1.00	11.00	10.00
delta	795	8.18	7.18	6.00	0.00	24.00	2.00	14.00	12.00
omicron	3711	7.59	6.54	6.00	0.00	24.00	2.00	12.00	10.00

**Table 2:** Descriptive metrics of age (months) by variants.

## Grouped age

### First case

```
##### First case #####
data_age2$age_months_firstcase <-
  as.factor(data_age2$age_months_firstcase)

summarytools::cTable(
  data_age2$age_months_firstcase,
  data_age2$variants,
  prop = "c",
  headings = st_options("headings"),
  display.labels = st_options("display.labels"),
  useNA = "ifany",
  OR = TRUE,
  chisq = TRUE
)
```

	variants	original	gamma	delta	omicron	Total
age_months_firstcase						
1 month		842 ( 22.7%)	586 ( 19.9%)	102 ( 12.8%)	356 ( 9.6%)	1886 ( 16.9%)
between 1 and 6 months		1248 ( 33.7%)	1103 ( 37.5%)	269 ( 33.8%)	1430 ( 38.5%)	4050 ( 36.3%)
between 6 and 12 months		743 ( 20.0%)	585 ( 19.9%)	179 ( 22.5%)	996 ( 26.8%)	2503 ( 22.4%)
more than 12 months		873 ( 23.6%)	667 ( 22.7%)	245 ( 30.8%)	929 ( 25.0%)	2714 ( 24.3%)
Total		3706 (100.0%)	2941 (100.0%)	795 (100.0%)	3711 (100.0%)	11153 (100.0%)

Chi.squared	df	p.value
298.4065	9	0

**Table 3:** Age(months) versus variants, where age is grouped as:  $\leq 30$  days (1 month); 31 days  $\mid$ - 180 days (between 1 and 6 months); 181 days  $\mid$ - 365 days (between 6 and 12 months);  $\geq 366$  days (more than 12 months).

### Second case

```
##### Second case #####
data_age2$age_months_secondcase <-
  factor(data_age2$age_months_secondcase,
    levels = c("1 month",
               "between 1 and 6 months",
               "between 6 and 12 months",
               "between 12 and 18 months",
               "more than 18 months")
  )

summarytools::cTable(
  data_age2$age_months_secondcase,
  data_age2$variants,
  prop = "c",
  headings = st_options("headings"),
  display.labels = st_options("display.labels"),
  useNA = "ifany",
  OR = TRUE,
)
```



```
chisq = TRUE
)
```

	variants	original	gamma	delta	omicron	Total
age_months_secondcase						
1 month		842 ( 22.7%)	586 ( 19.9%)	102 ( 12.8%)	356 ( 9.6%)	1886 ( 16.9%)
between 1 and 6 months		1248 ( 33.7%)	1103 ( 37.5%)	269 ( 33.8%)	1430 ( 38.5%)	4050 ( 36.3%)
between 6 and 12 months		743 ( 20.0%)	585 ( 19.9%)	179 ( 22.5%)	996 ( 26.8%)	2503 ( 22.4%)
between 12 and 18 months		493 ( 13.3%)	366 ( 12.4%)	129 ( 16.2%)	514 ( 13.9%)	1502 ( 13.5%)
more than 18 months		380 ( 10.3%)	301 ( 10.2%)	116 ( 14.6%)	415 ( 11.2%)	1212 ( 10.9%)
Total		3706 (100.0%)	2941 (100.0%)	795 (100.0%)	3711 (100.0%)	11153 (100.0%)

Chi.squared	df	p.value
299.8064	12	0

**Table 4:** Age(months) versus variants, where age is grouped as:  $\leq 30$  days (1 month); 31 days  $\mid$ - 180 days (between 1 and 6 months); 181 days  $\mid$ - 365 days (between 6 and 12 months); 366 days  $\mid$ - 545 days (between 12 and 18 months);  $\geq 546$  days (more than 18 months).

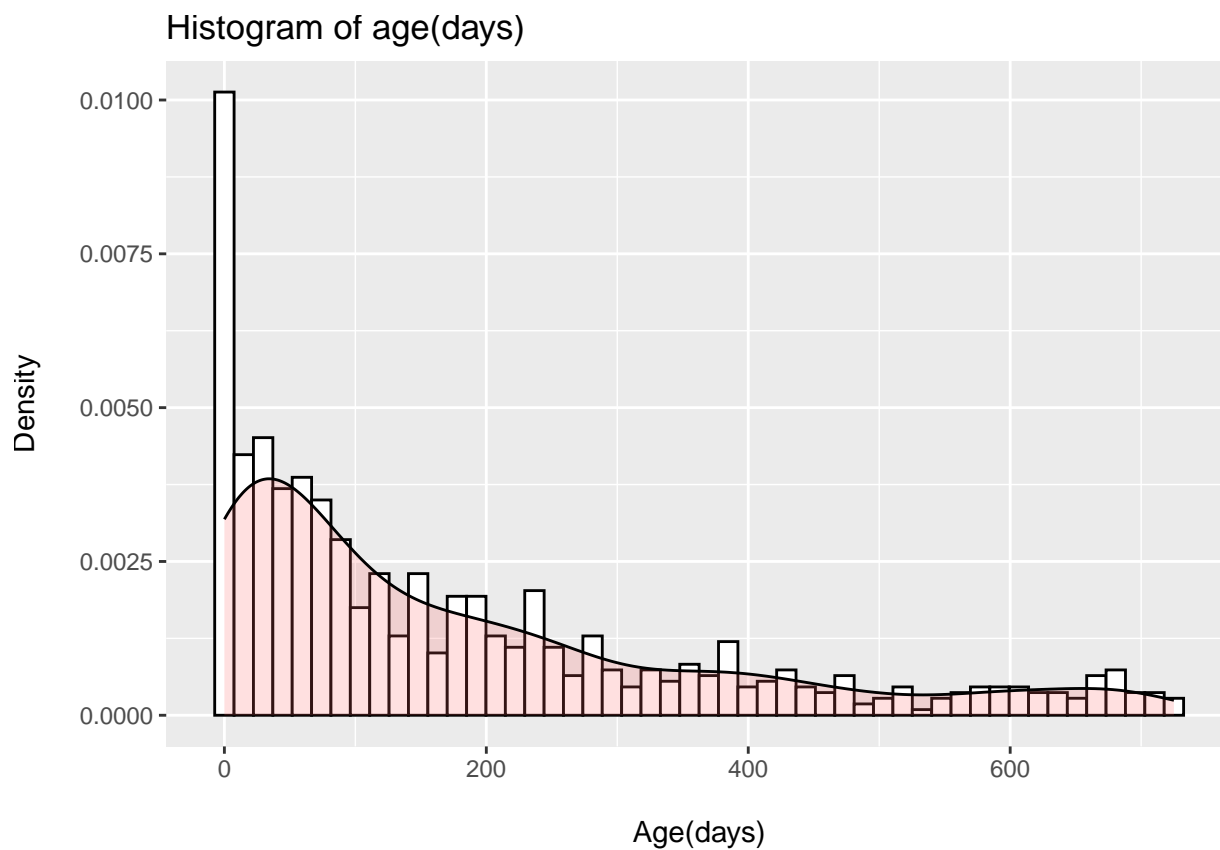
## Analysis by evolution(death = yes)

### Quantitative analysis

#### Age(days)

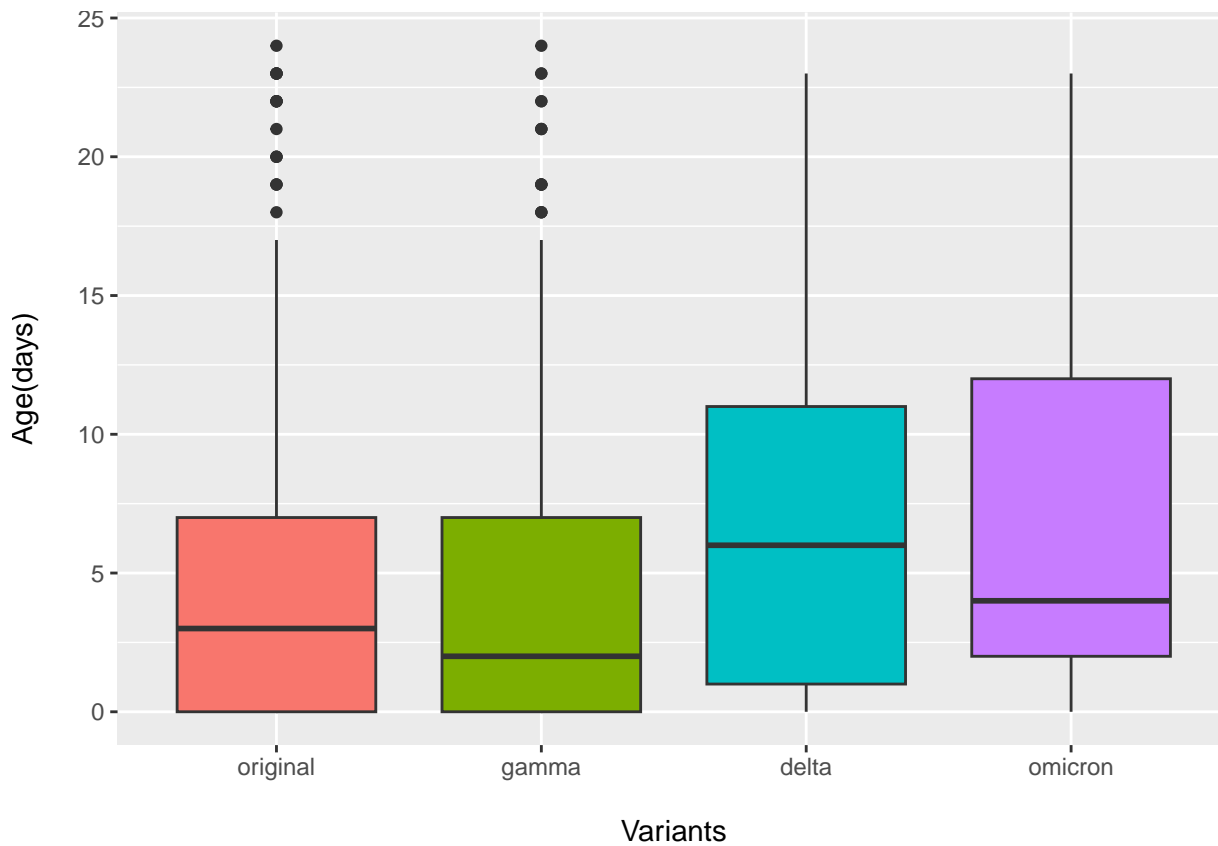
```
data_age2_death <- filter(data_age2, death == "death")

ggplot(data_age2_death, aes(x = idade_dias)) +
  geom_histogram(aes(y = after_stat(density)),
    bins = 50,
    colour = "black",
    fill = "white") +
  geom_density(alpha = .2, fill = "#FF6666") +
  labs(x = "\nAge(days)",
    y = "Density\n",
    title = "Histogram of age(days)")
```



**Figure 5:** Histogram of age in unit of days (bars) and density (curve with red area). Here, only cases that evolved into death are considered.

```
data_age2_death %>%
  ggplot(aes(x = variants,
             y = age_month,
             fill = variants)) +
  geom_boxplot() +
  xlab("\nVariants") +
  ylab("Age(days)\n") +
  theme(legend.position = "none")
```



**Figure 6:** Boxplot of variants by age in a unit of days. Here, only cases that evolved into death are considered.

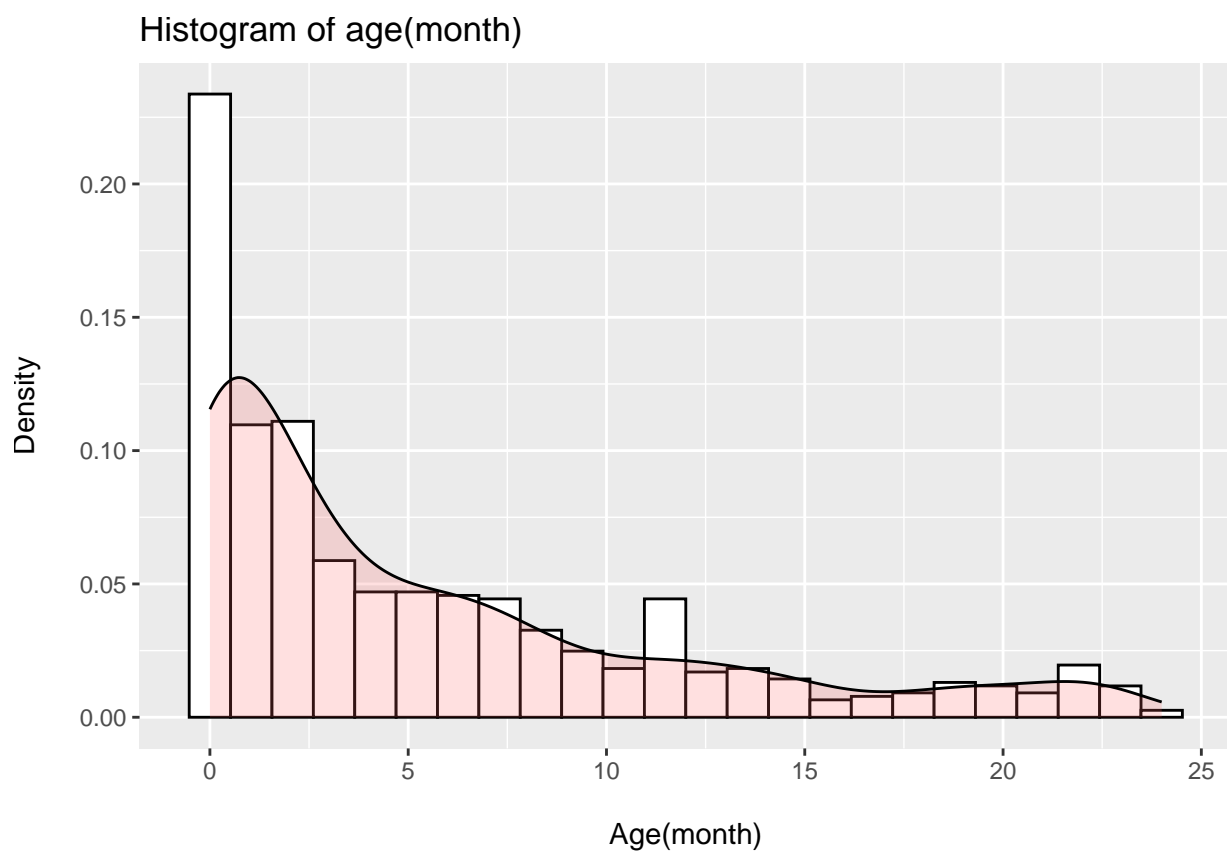
```
datasummary(variants ~ idade_dias*(n + media + DP + mediana + minimo +
  maximo + q25 + q75 + IQR),
  data = data_age2_death, output = 'markdown')
```

variants	n	media	DP	mediana	minimo	maximo	q25	q75	IQR
original	335	156.70	177.34	92.00	0.00	725.00	19.00	222.50	203.50
gamma	217	159.51	183.56	75.00	0.00	722.00	24.00	236.00	212.00
delta	41	233.07	219.76	187.00	0.00	714.00	58.00	331.00	273.00
omicron	141	233.50	202.08	144.00	0.00	692.00	70.00	369.00	299.00

**Table 5:** Descriptive metrics of age (days) by variants. Here, only cases that evolved into death are considered.

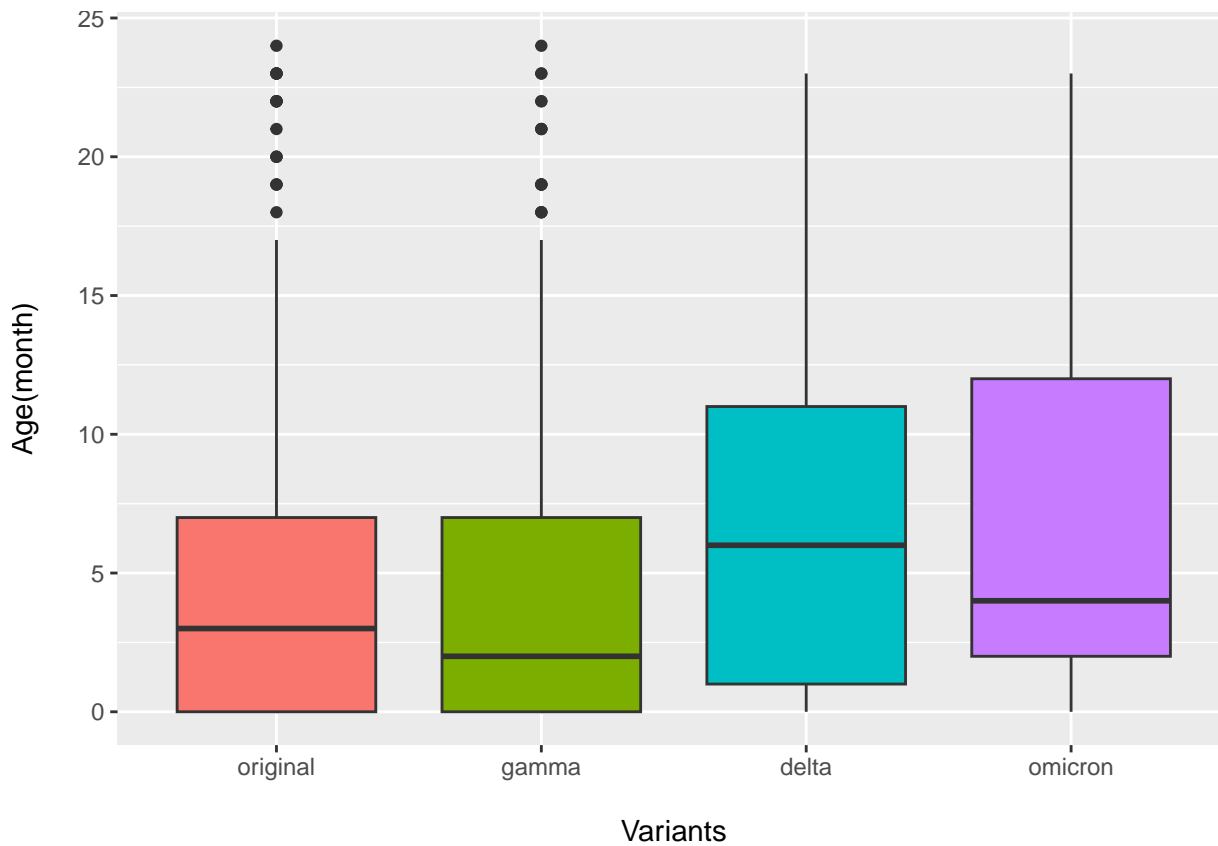
### Age(months)

```
ggplot(data_age2_death, aes(x = age_month)) +
  geom_histogram(aes(y = after_stat(density)),
    bins = 24,
    colour = "black",
    fill = "white") +
  geom_density(alpha = .2, fill = "#FF6666") +
  labs(x = "\nAge(month)",
    y = "Density\n",
    title = "Histogram of age(month)")
```



**Figure 7:** Histogram of age in unit of months (bars) and density (curve with red area). Here, only cases that evolved into death are considered.

```
data_age2_death %>%
  ggplot(aes(x = variants,
             y = age_month,
             fill = variants)) +
  geom_boxplot() +
  xlab("\nVariants") +
  ylab("Age(month)\n") +
  theme(legend.position = "none")
```



**Figure 8:** Boxplot of variants by age in a unit of months. Here, only cases that evolved into death are considered.

```
datasummary(variants ~ age_month*(n + media + DP + mediana + minimo +
                                maximo + q25 + q75 + IQR),
            data = data_age2_death, output = 'markdown')
```

variants	n	media	DP	mediana	minimo	maximo	q25	q75	IQR
original	335	4.81	5.86	3.00	0.00	24.00	0.00	7.00	7.00
gamma	217	4.89	6.04	2.00	0.00	24.00	0.00	7.00	7.00
delta	41	7.41	7.30	6.00	0.00	23.00	1.00	11.00	10.00
omicron	141	7.28	6.72	4.00	0.00	23.00	2.00	12.00	10.00

**Table 6:** Descriptive metrics of age (months) by variants. Here, only cases that evolved into death are considered.

## Grouped age

### First case

```
##### First case #####
summarytools::ctable(
  data_age2_death$age_months_firstcase,
  data_age2_death$variants,
  prop = "c",
  headings = st_options("headings"),
  display.labels = st_options("display.labels"),
  useNA = "ifany",
  OR = TRUE,
```

```

    chisq = TRUE
  )

```

	variants	original	gamma	delta	omicron	Total
age_months_firstcase						
1 month		105 ( 31.3%)	61 ( 28.1%)	5 ( 12.2%)	11 ( 7.8%)	182 ( 24.8%)
between 1 and 6 months		121 ( 36.1%)	81 ( 37.3%)	15 ( 36.6%)	66 ( 46.8%)	283 ( 38.6%)
between 6 and 12 months		63 ( 18.8%)	43 ( 19.8%)	12 ( 29.3%)	28 ( 19.9%)	146 ( 19.9%)
more than 12 months		46 ( 13.7%)	32 ( 14.7%)	9 ( 22.0%)	36 ( 25.5%)	123 ( 16.8%)
Total		335 (100.0%)	217 (100.0%)	41 (100.0%)	141 (100.0%)	734 (100.0%)

Chi.squared	df	p.value
40.4386	9	0

**Table 7:** Age(months) versus variants, where age is grouped as:  $\leq 30$  days (1 month); 31 days  $\mid$ - 180 days (between 1 and 6 months); 181 days  $\mid$ - 365 days (between 6 and 12 months);  $\geq 366$  days (more than 12 months). Here, only cases that evolved into death are considered.

## Second case

```

##### Second case #####
summarytools::ctable(
  data_age2_death$age_months_secondcase,
  data_age2_death$variants,
  prop = "c",
  headings = st_options("headings"),
  display.labels = st_options("display.labels"),
  useNA = "ifany",
  OR = TRUE,
  chisq = TRUE
)

```

	variants	original	gamma	delta	omicron	Total
age_months_secondcase						
1 month		105 ( 31.3%)	61 ( 28.1%)	5 ( 12.2%)	11 ( 7.8%)	182 ( 24.8%)
between 1 and 6 months		121 ( 36.1%)	81 ( 37.3%)	15 ( 36.6%)	66 ( 46.8%)	283 ( 38.6%)
between 6 and 12 months		63 ( 18.8%)	43 ( 19.8%)	12 ( 29.3%)	28 ( 19.9%)	146 ( 19.9%)
between 12 and 18 months		27 ( 8.1%)	18 ( 8.3%)	2 ( 4.9%)	18 ( 12.8%)	65 ( 8.9%)
more than 18 months		19 ( 5.7%)	14 ( 6.5%)	7 ( 17.1%)	18 ( 12.8%)	58 ( 7.9%)
Total		335 (100.0%)	217 (100.0%)	41 (100.0%)	141 (100.0%)	734 (100.0%)

Chi.squared	df	p.value
45.7022	12	0

**Table 8:** Age(months) versus variants, where age is grouped as:  $\leq 30$  days (1 month); 31 days  $\mid$ - 180 days (between 1 and 6 months); 181 days  $\mid$ - 365 days (between 6 and 12 months); 366 days  $\mid$ - 545 days (between 12 and 18 months);  $\geq 546$  days (more than 18 months). Here, only cases that evolved into death are considered.