

analysis of COVID-19 cases in Mexico

bartłomiej brzostek

it_depends



agenda

1. introduction

- 1.1. project objective
- 1.2. basic country information

2. data analysis

- 2.1. distribution of the number of patients over time range
- 2.2. distribution of the number of patients over age group
- 2.3. correlation study

3. summary

- 3.1. observations

introduction

objective:

To analyse data on individuals who **contracted COVID-19** in Mexico. Factors influencing the course of the disease including **death** are sought.

Cases of hospitalised covid-19 positive patients will be analysed, as well as the relationship between covid-19 incidence and/or death and the gender, age and comorbidities of the patient.

stakeholder:

A medical concern that wants to learn the lessons from the pandemic and, based on these, know which groups may be at risk in the future.

<https://www.kaggle.com/datasets/meirnizri/covid19-dataset>
(source updated on: 2021.02.28)



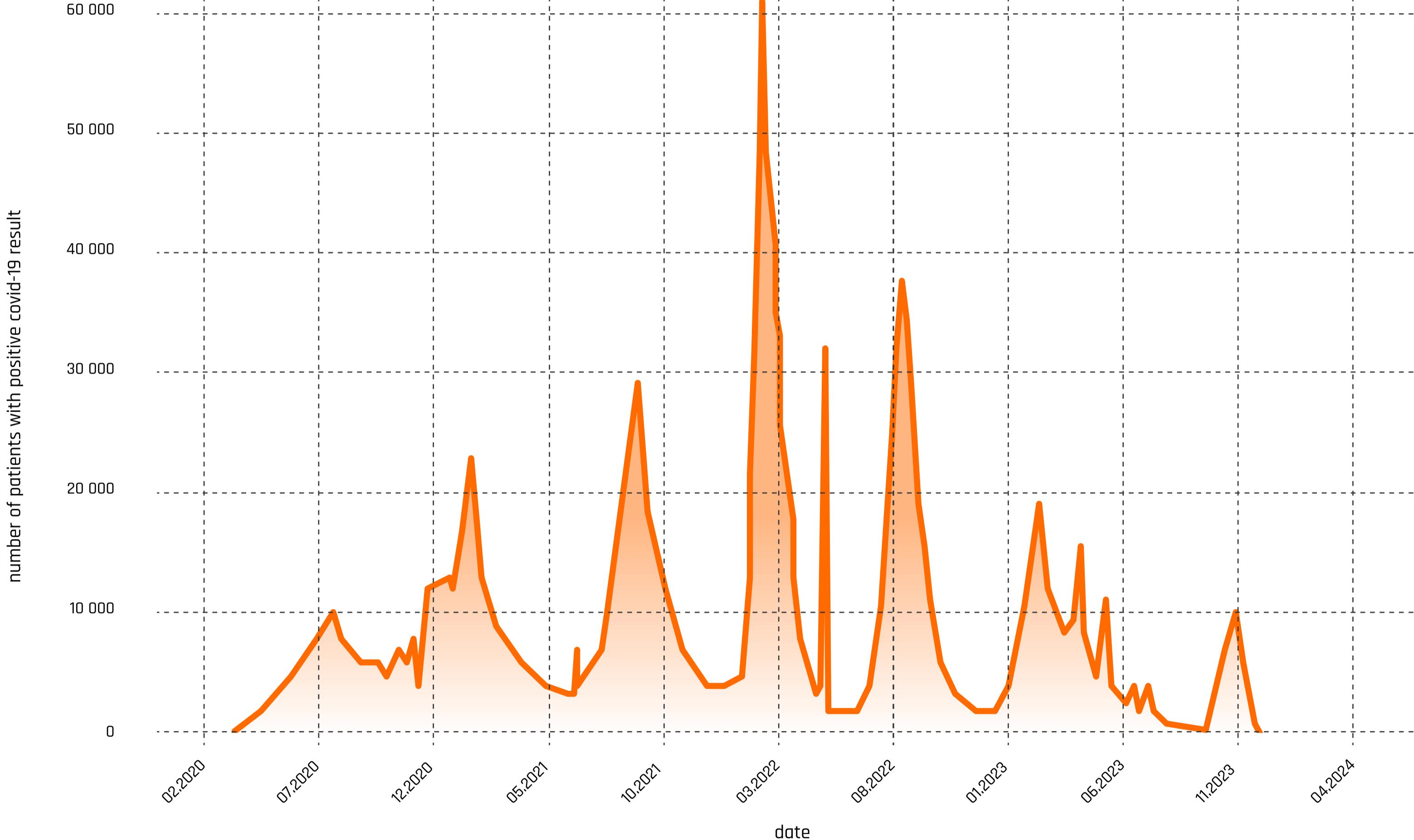


country: **Mexico**
area: **1 972 550 km²**
population: **129 875 529**

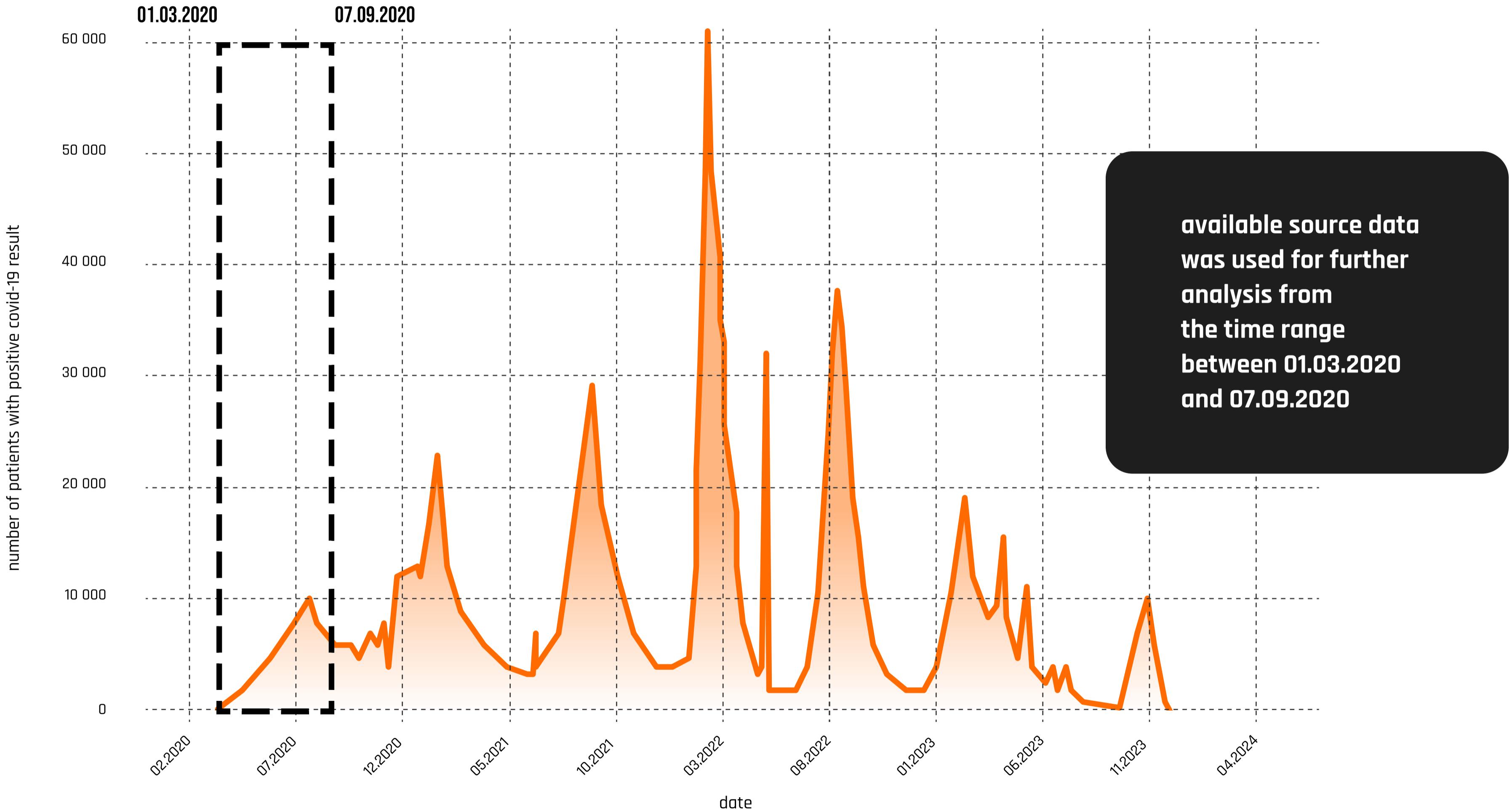
[13th largest country]
[10th most populous country]

data analysis

number of covid-19 positive patients over time in Mexico - cases from the begining of covid-19 until 01.2024



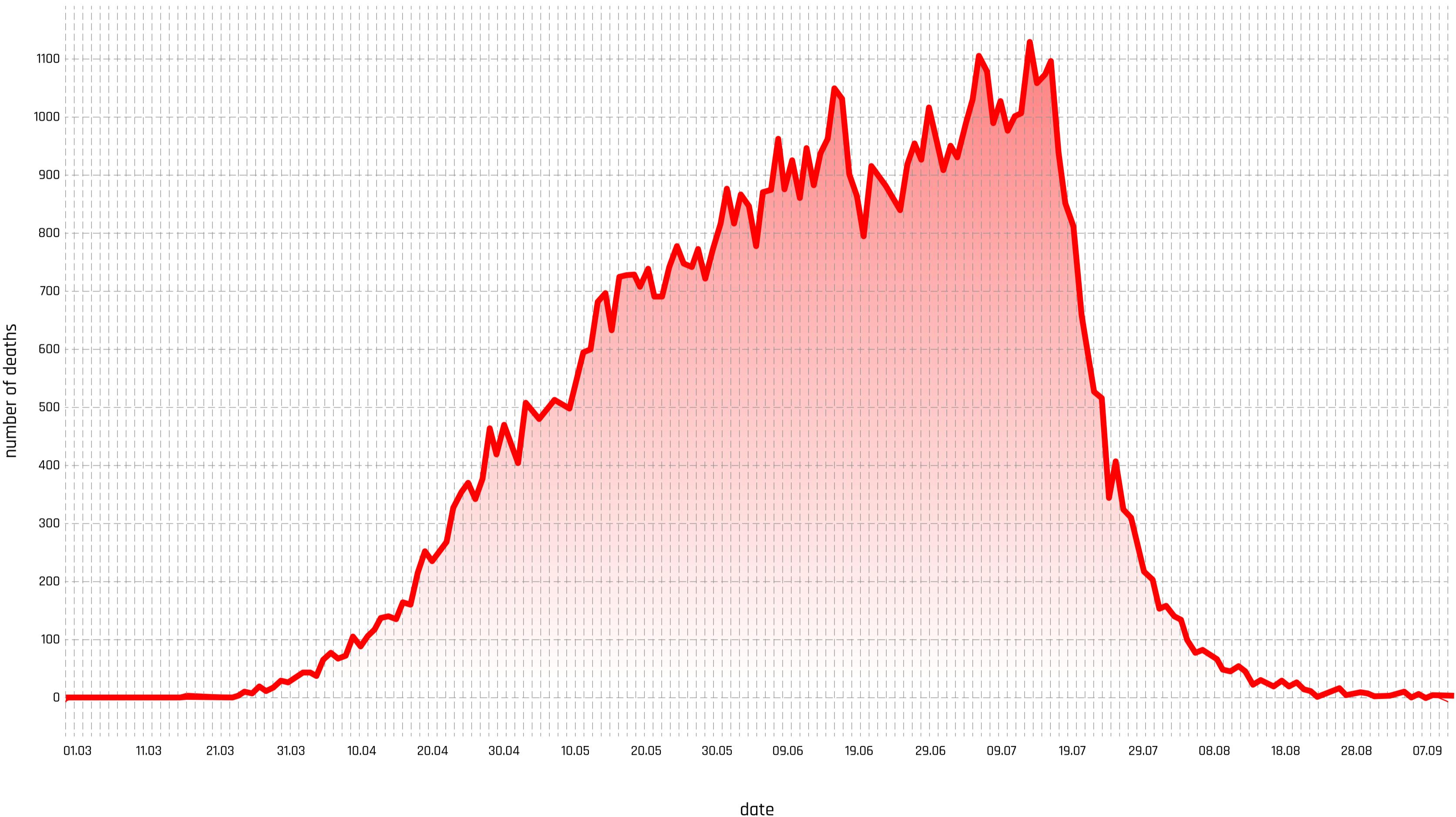
number of covid-19 positive patients over time in Mexico - cases from the begining of covid-19 until 01.2024



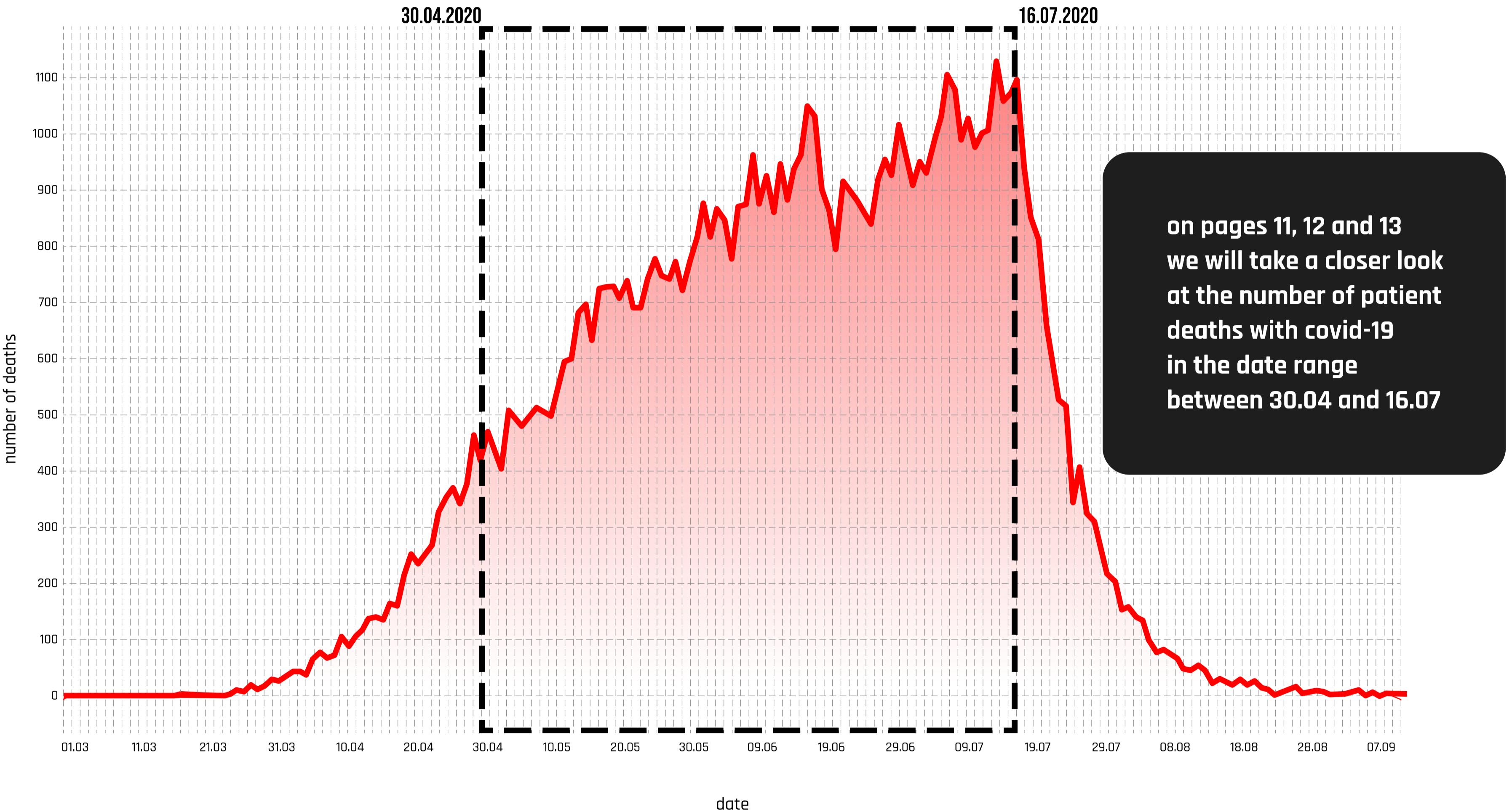
2.1. distribution of the number of patients over time range

it_depends

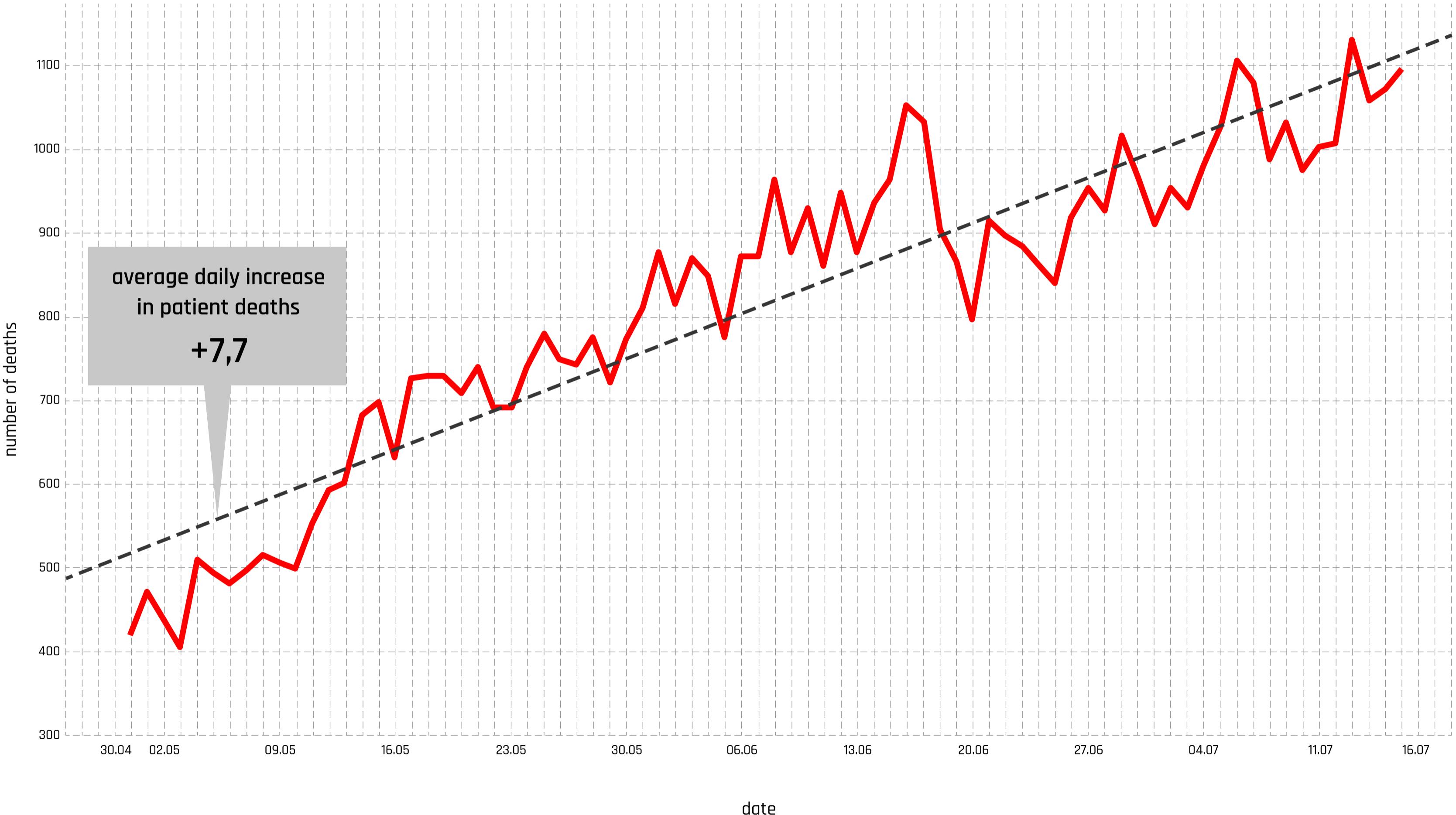
**number of daily dead patients with a positive covid-19 test result - cases from the time range between
01.03.2020 and 07.09.2020**



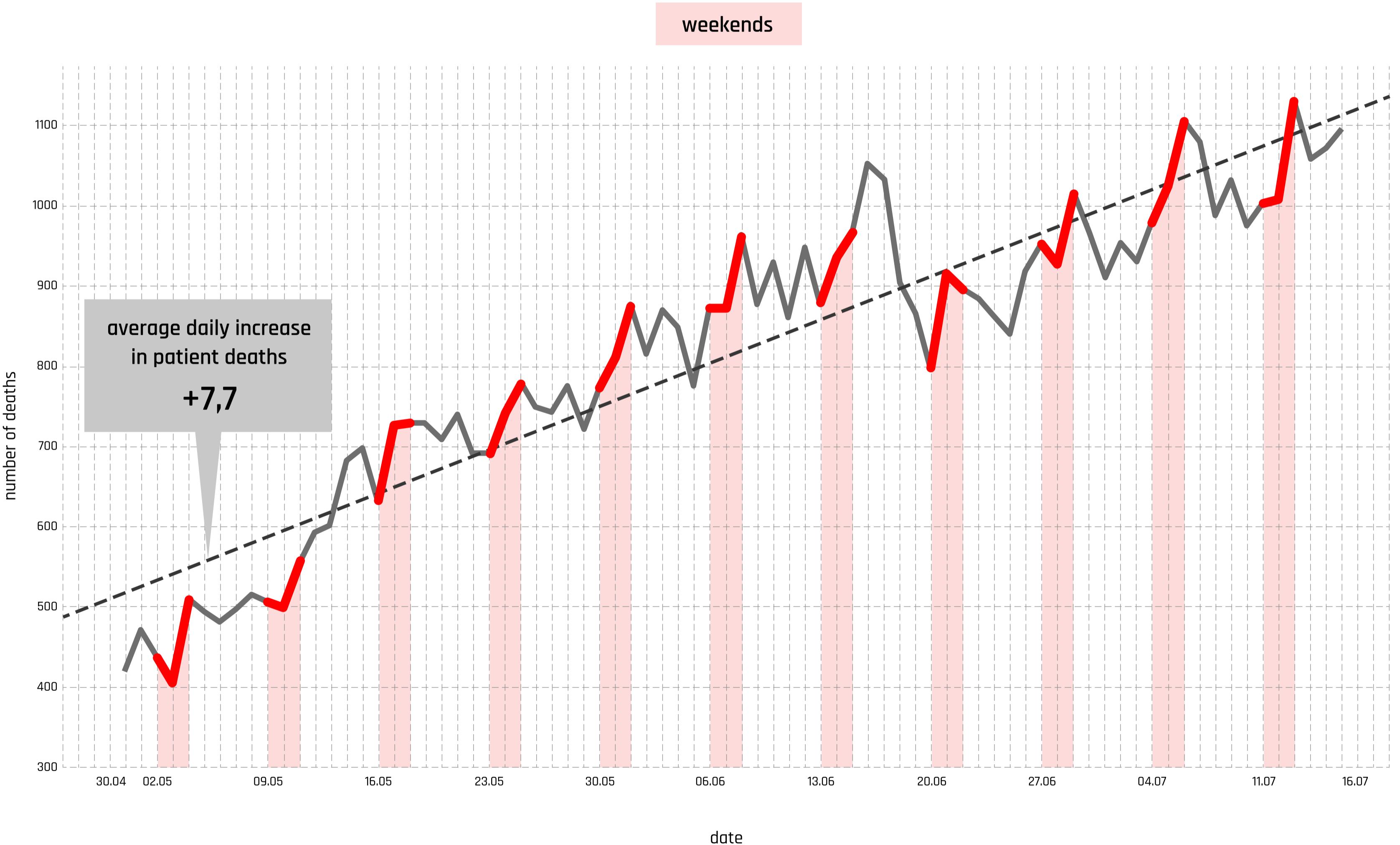
**number of daily dead patients with a positive covid-19 test result - cases from the time range between
01.03.2020 and 07.09.2020**



number of daily dead patients with a positive covid-19 test result - cases from the time range between 30.04 and 16.07



number of daily dead patients with a positive covid-19 test result - cases from the time range between 30.04 and 16.07



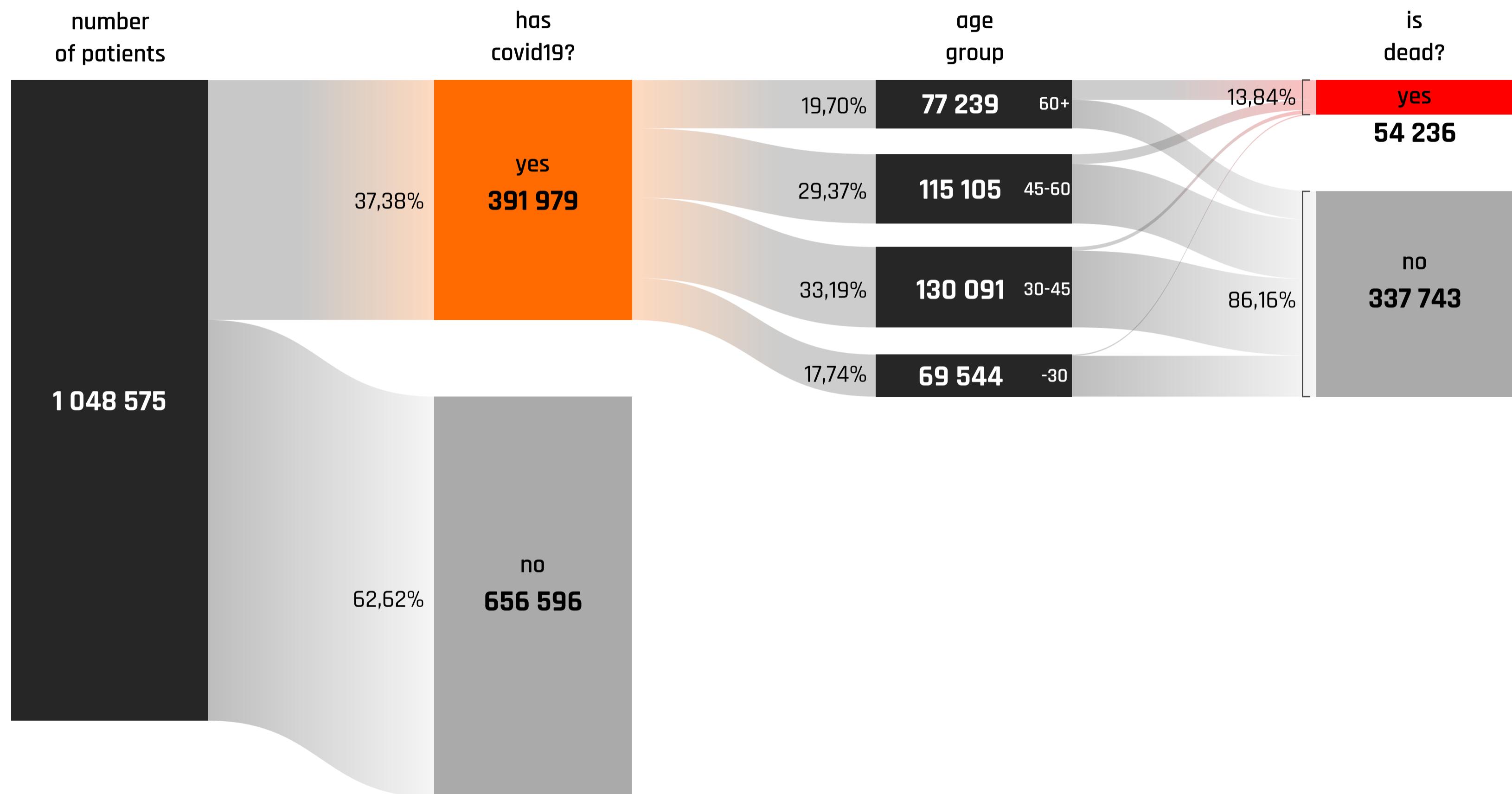
actual number of deaths at weekends vs. number of deaths calculated according to the regression line

weekends	02-04.05	09-11.05	16-18.05	23-25.05	30.05-01.06	06-08.06	13-15.06	20-22.06	27-29.06	04-06.07	11-13.07	sum
[A] number of deaths [actual]	913	1 051	1 453	1 517	1 686	1 835	1 898	1 809	1 942	2 130	2 135	18 369
[B] number of deaths [regression line]	899	1 035	1 287	1 403	1 569	1 765	1 783	1 613	1 933	1 981	2 027	17 296
[A-B] difference	14	16	166	114	117	70	115	196	9	149	108	1 073

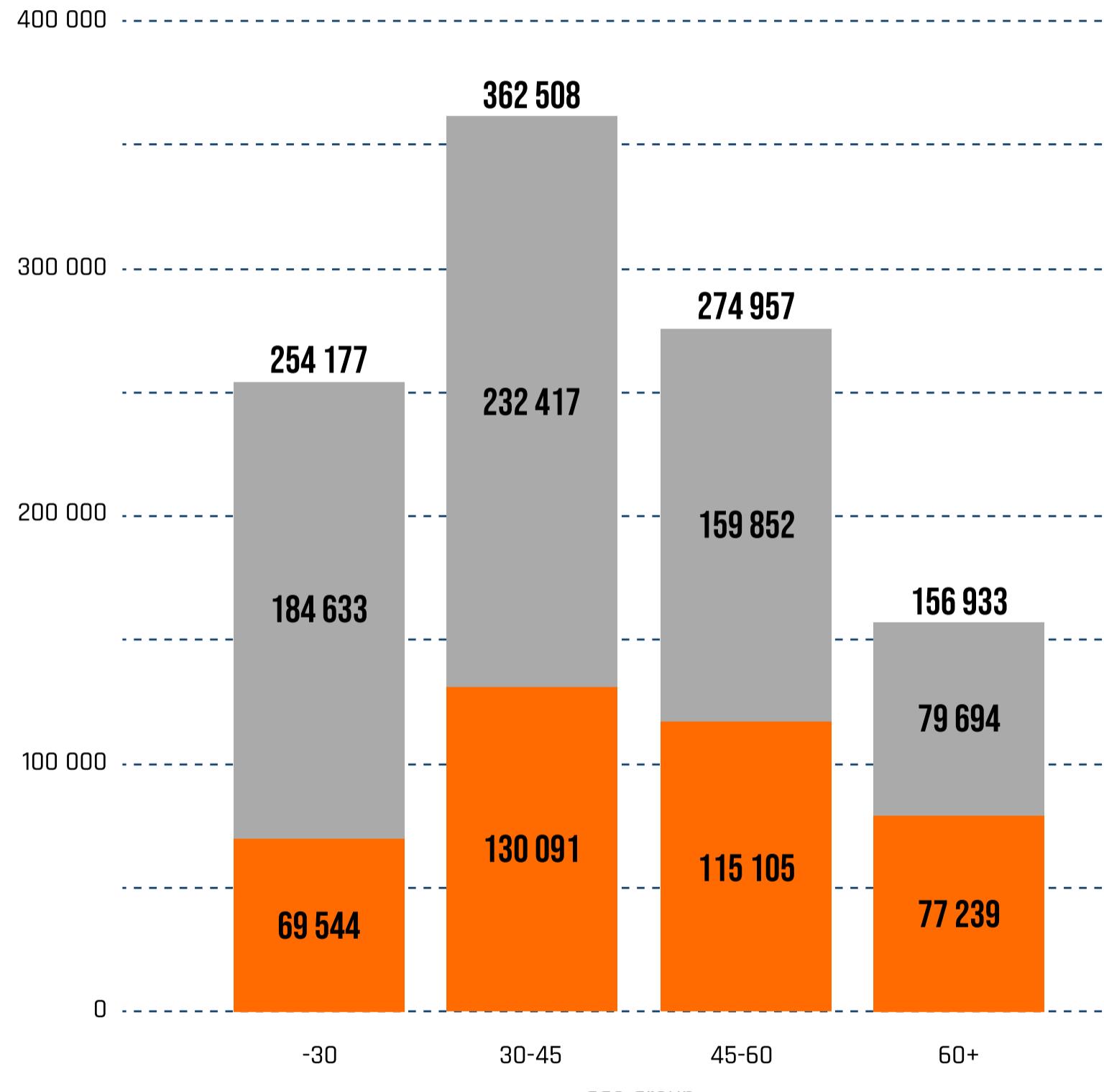
actual number of deaths at weekends vs. number of deaths calculated according to the regression line

**in case if the number of deaths in weekends
had been rising according to the linear regression,
1 073 patients more would potentially survived**

Number of patients studied and its breakdown by covid-19 test result, age group and whether the patient died



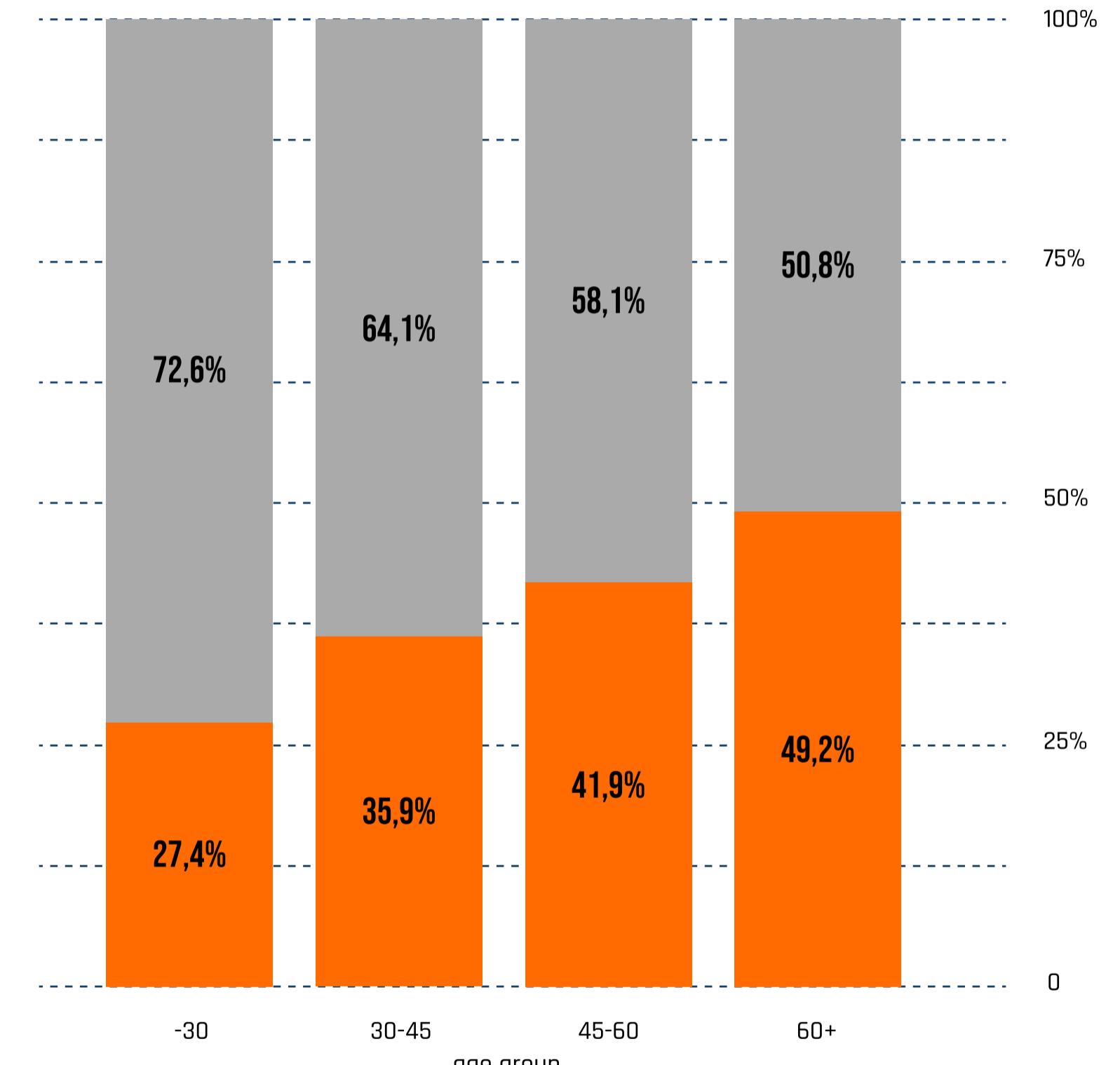
Number of patients studied and its distribution by age group and covid-19 test result



has patient a positive covid-19 test result? [no.]

YES

NO

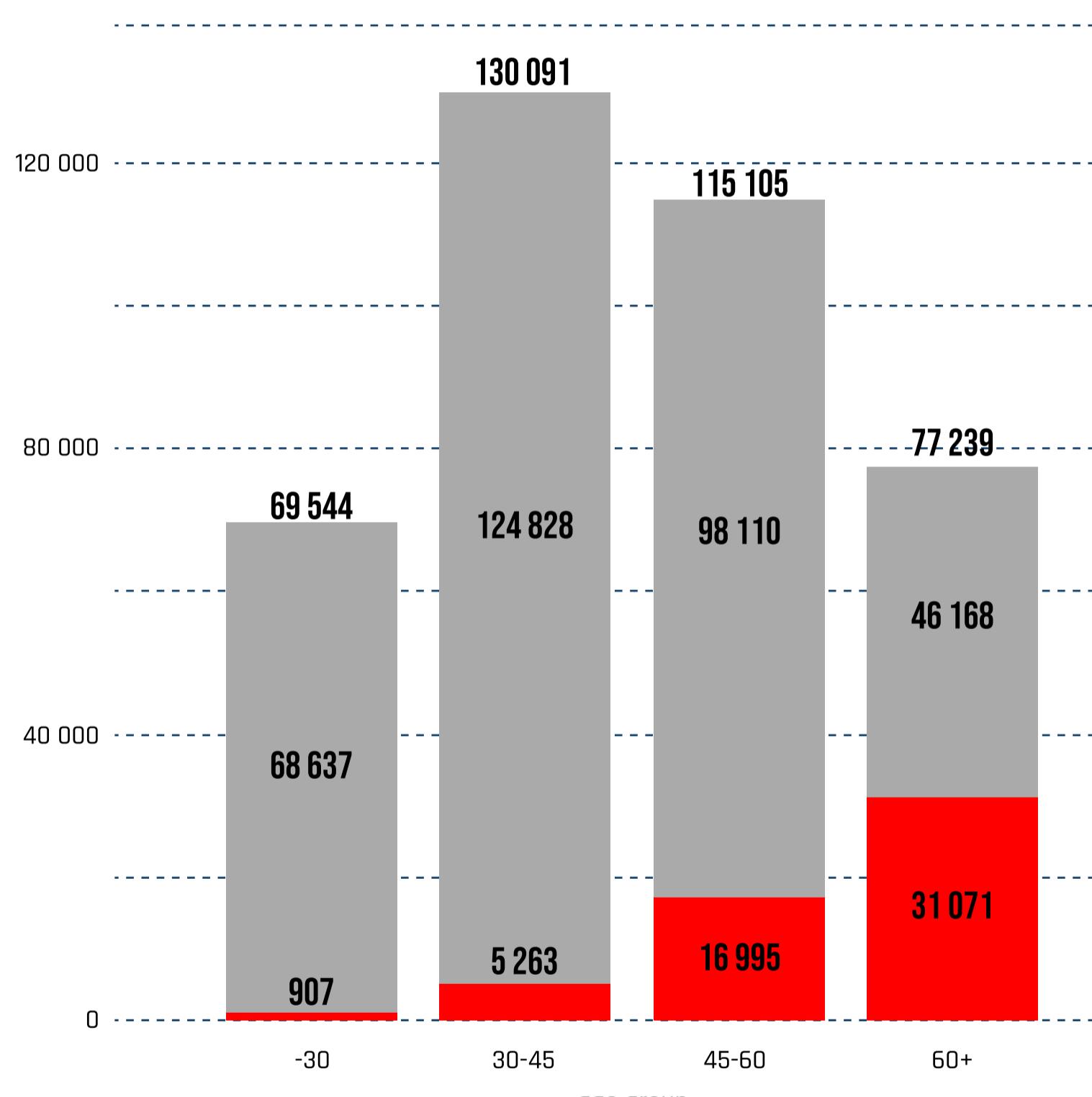


has patient a positive covid-19 test result? [%]

YES

NO

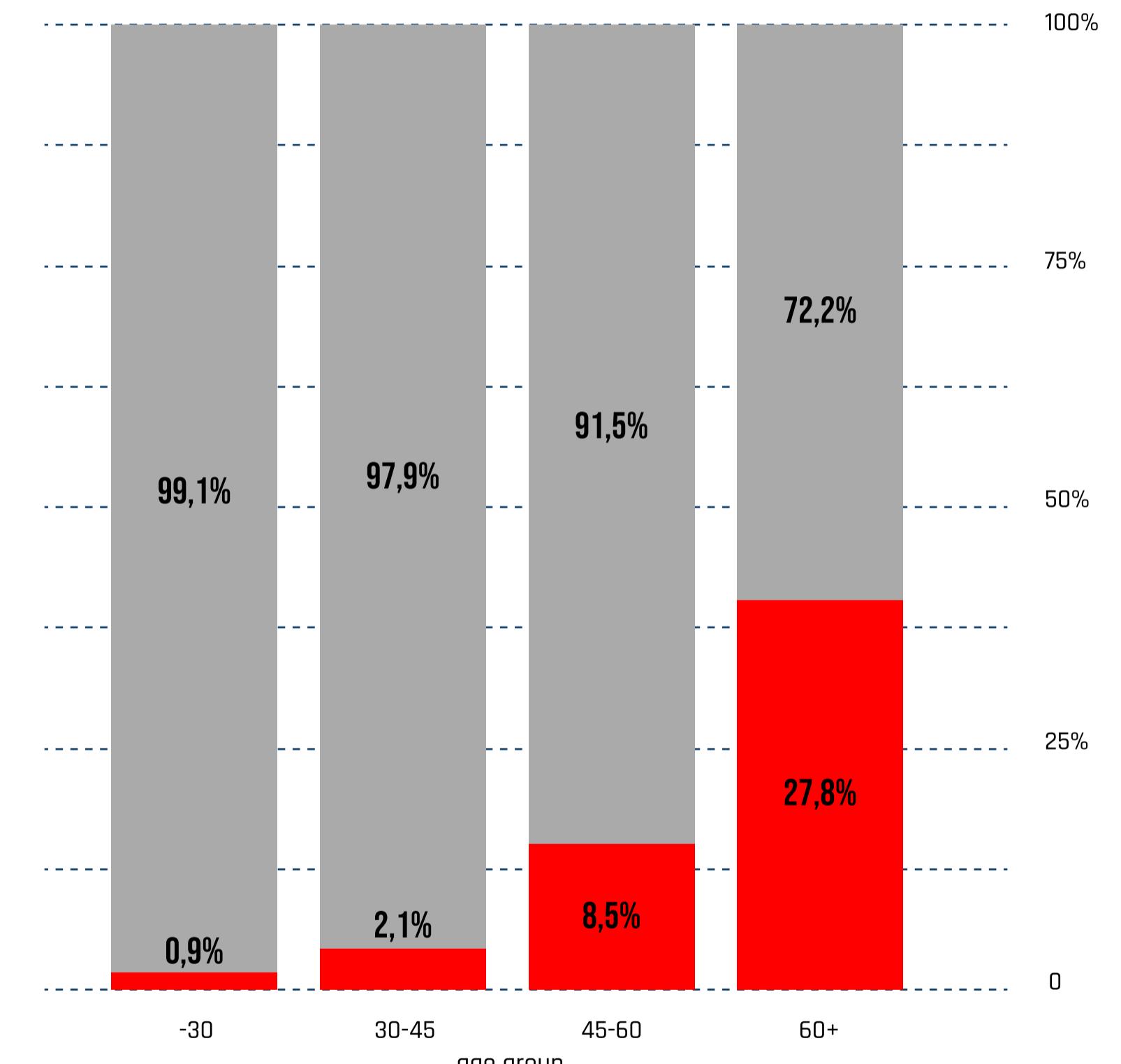
Number of patients with positive covid-19 test result and its distribution by age group and the fact if patient died



did covid-19 positive patient die? [no.]

YES

NO

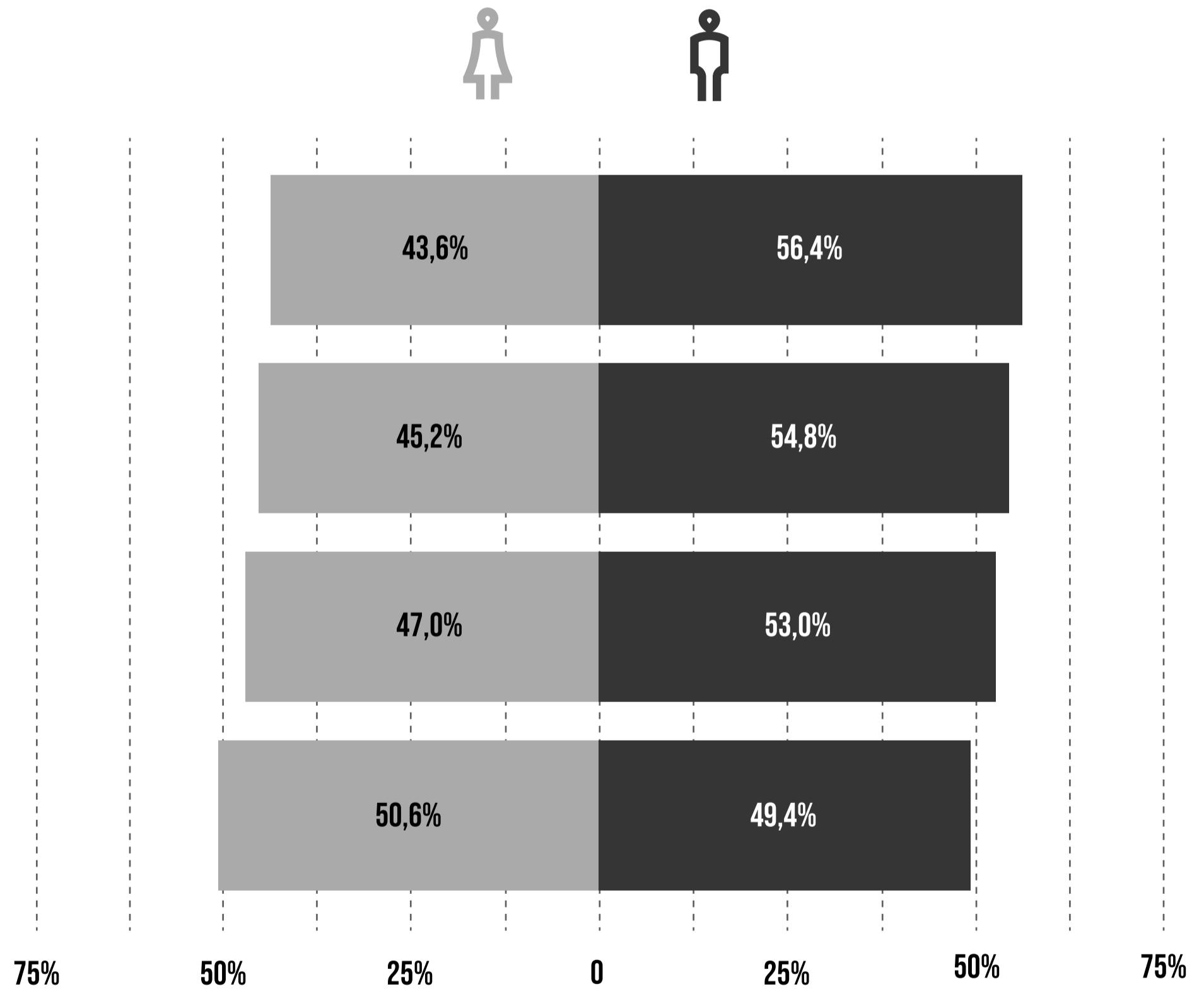


did covid-19 positive patient die? [%]

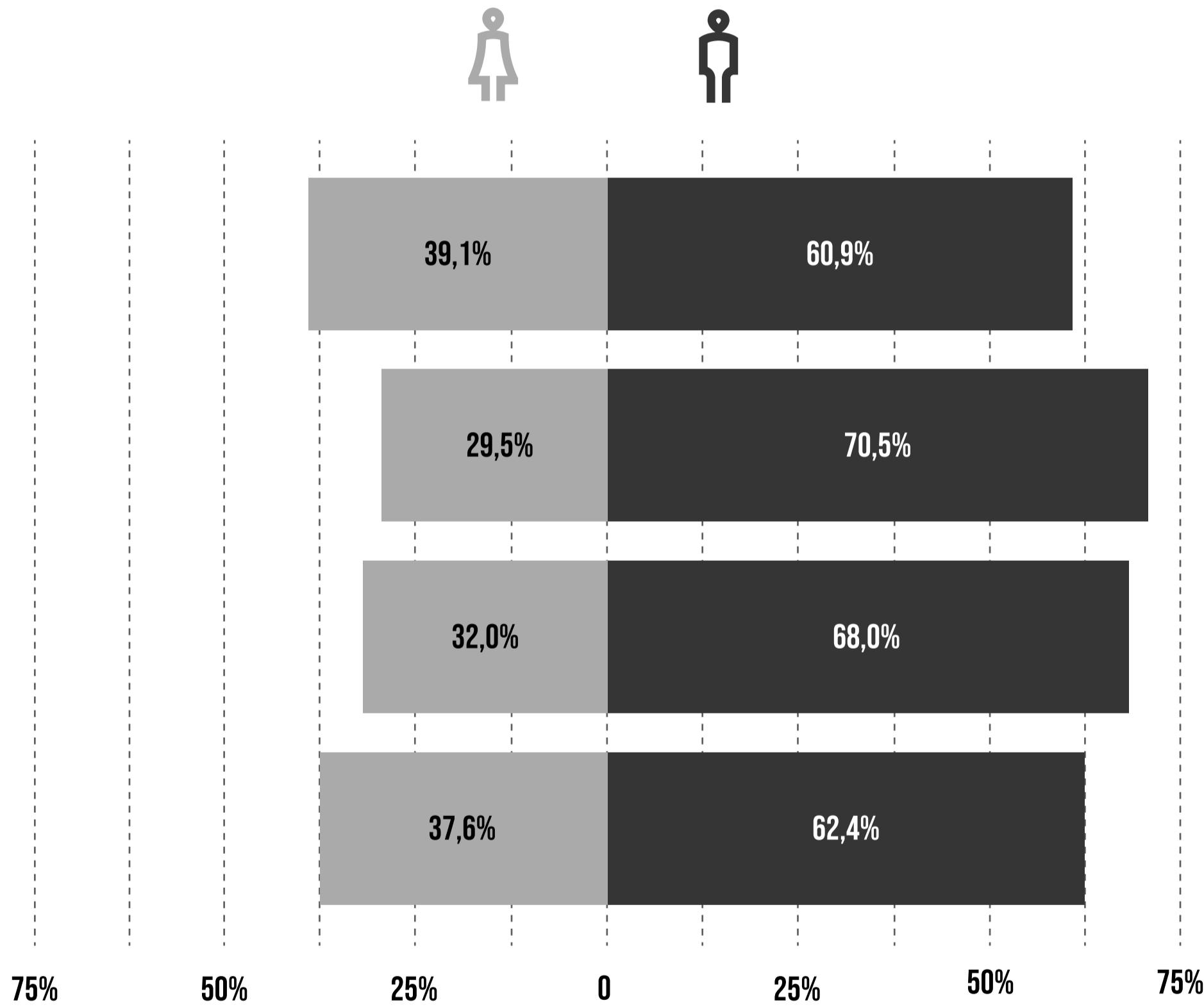
YES

NO

percentage distribution of covid-19 positive patients by gender and age group



as men get older,
covid-19 is more
common

percentage distribution of death cases of covid-19 positive patients by gender and age group

**in each age group
there are significantly
more deaths among
men**

tetrachoric correlation definition and formula explenation

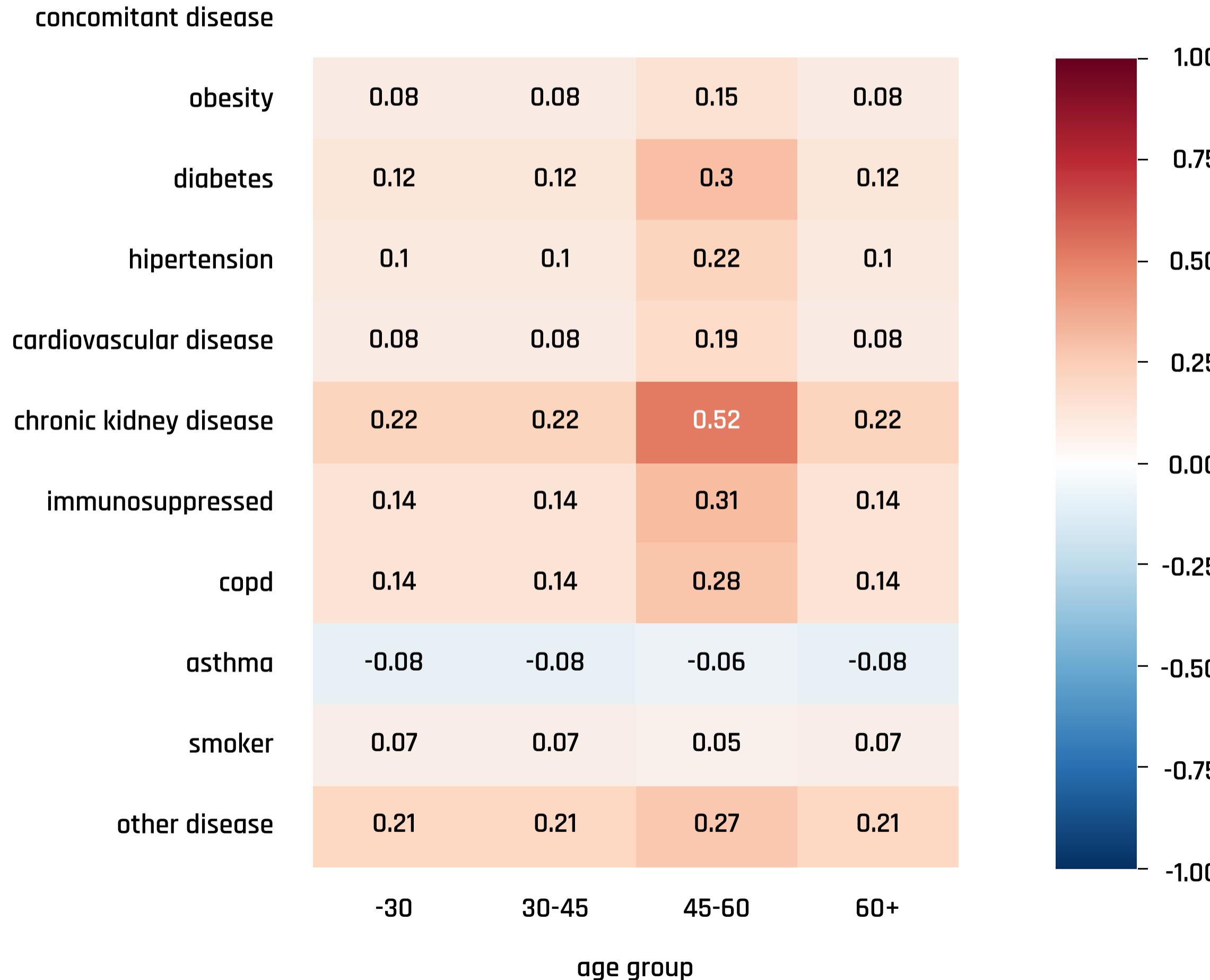
Tetrachoric correlation is used to measure rater agreement for binary data; Binary data is data with two possible answers—usually right or wrong. The tetrachoric correlation estimates what the correlation would be if measured on a continuous scale. This type of correlation has the advantage that it's not affected by the number of rating levels, or the marginal proportions for rating levels.

The term "tetrachoric correlation" comes from the tetrachoric series, a numerical method used before the advent of computers. While it's more common to estimate correlations with methods like maximum likelihood estimation, there is a basic formula you can use:

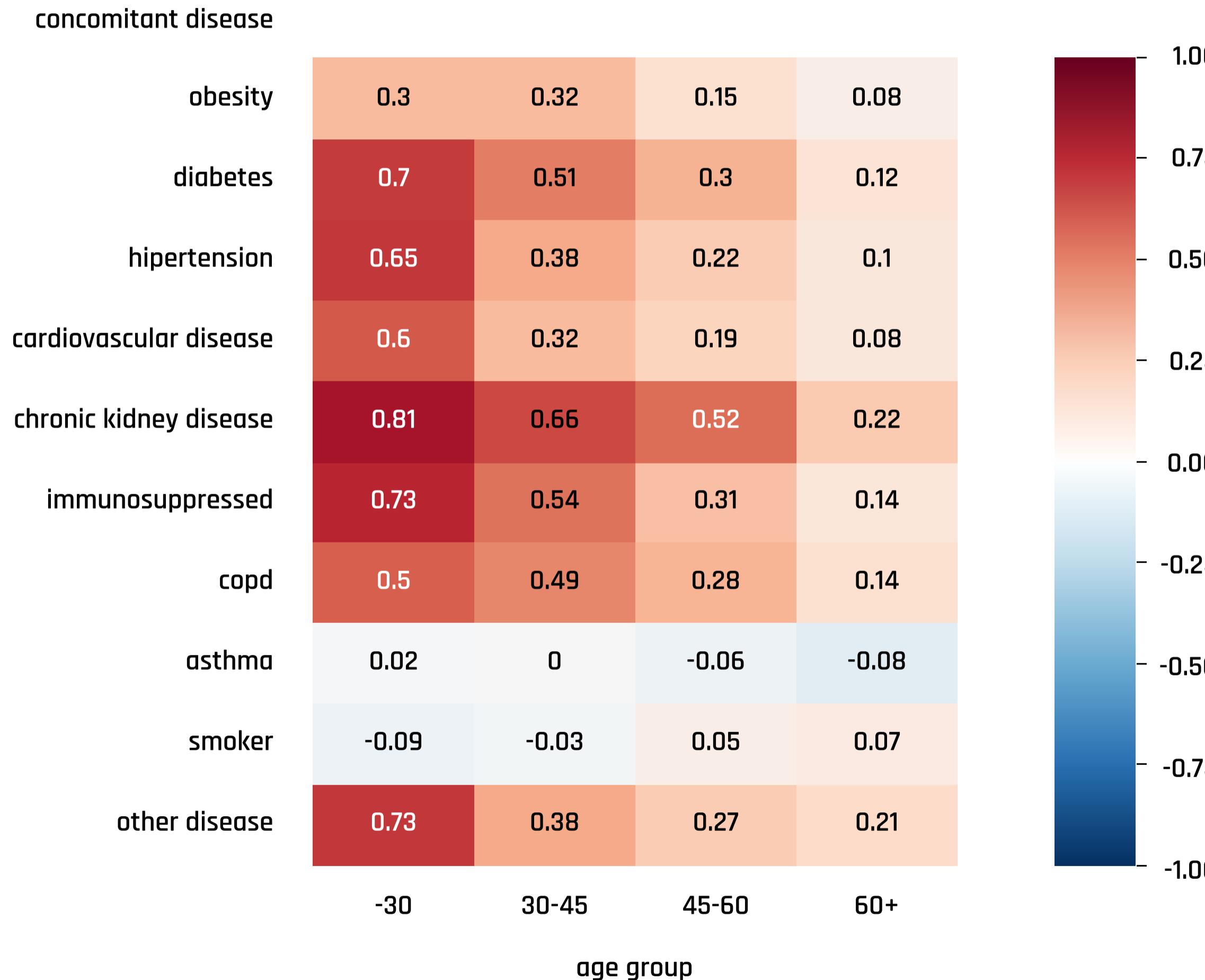
		feature 1 (binar)	
		0	1
feature 2 (binar)	0	A	B
	1	C	D

```
tet_corr = math.cos( math.pi / ( 1 + math.sqrt( (A/C) / (B/D) ) ) )
```

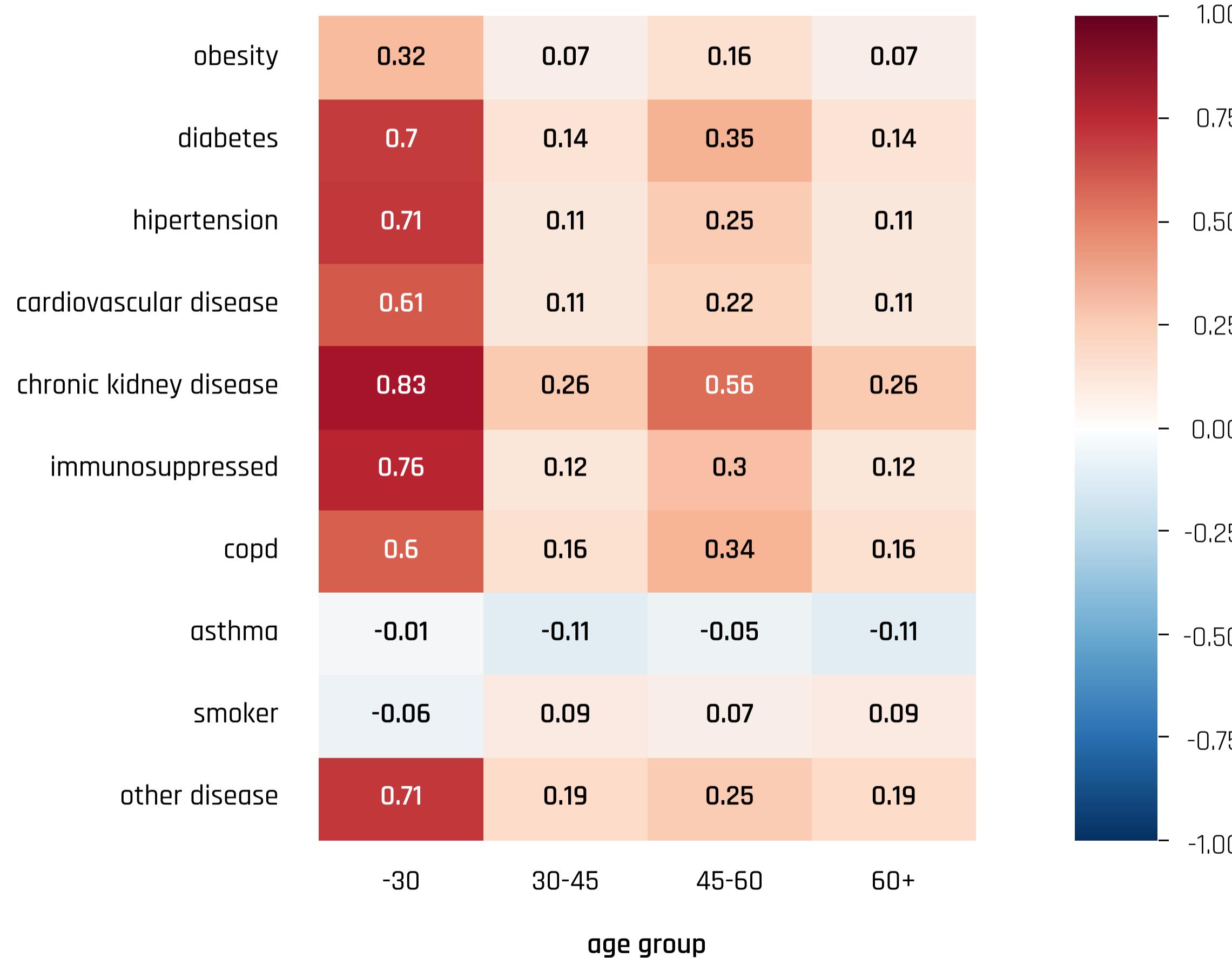
correlations between mortality and comorbidities among patients of different age groups [women]



correlations between mortality and comorbidities among patients of different age groups [men]



correlations between mortality and comorbidities among patients of different age groups [women + men]

concomitant disease

Comorbidities have a greater impact on mortality in patients under 30 years. Patients with chronic kidney disease had the highest risk of death.

summary

The data included 1,048,575 patients, 37% tested positive for COVID-19 and of these, 14% died (of which 57% were patients over 60 years of age).

The risk of death related to COVID-19 increases in patients over 45 years of age.

Women accounted for 46.5% of covid-19 positive patients.

Comorbidities have a greater impact on mortality in patients under 30 years of age.

Patients with chronic kidney disease had the highest risk of death.

When several comorbidities were present, the combination of chronic kidney disease with diabetes and/or hypertension was the most serious.



tools and libraries utilized in the presentation:

Python programming language
NumPy, Pandas libraries for data manipulation and analysis
Matplotlib library for statistical data visualization

graphs and pictures:

Pages 5, 7-22: Created internally
Pages 1, 4, 26: Sourced from www.pexels.com



analysis of COVID-19 cases in Mexico

bartłomiej brzostek

it_depends

