

FACTORS INFLUENCING THE MORTALITY RATE AMONG COVID-19 PATIENS ANALYSIS

Ky Anh Le

Content Outline

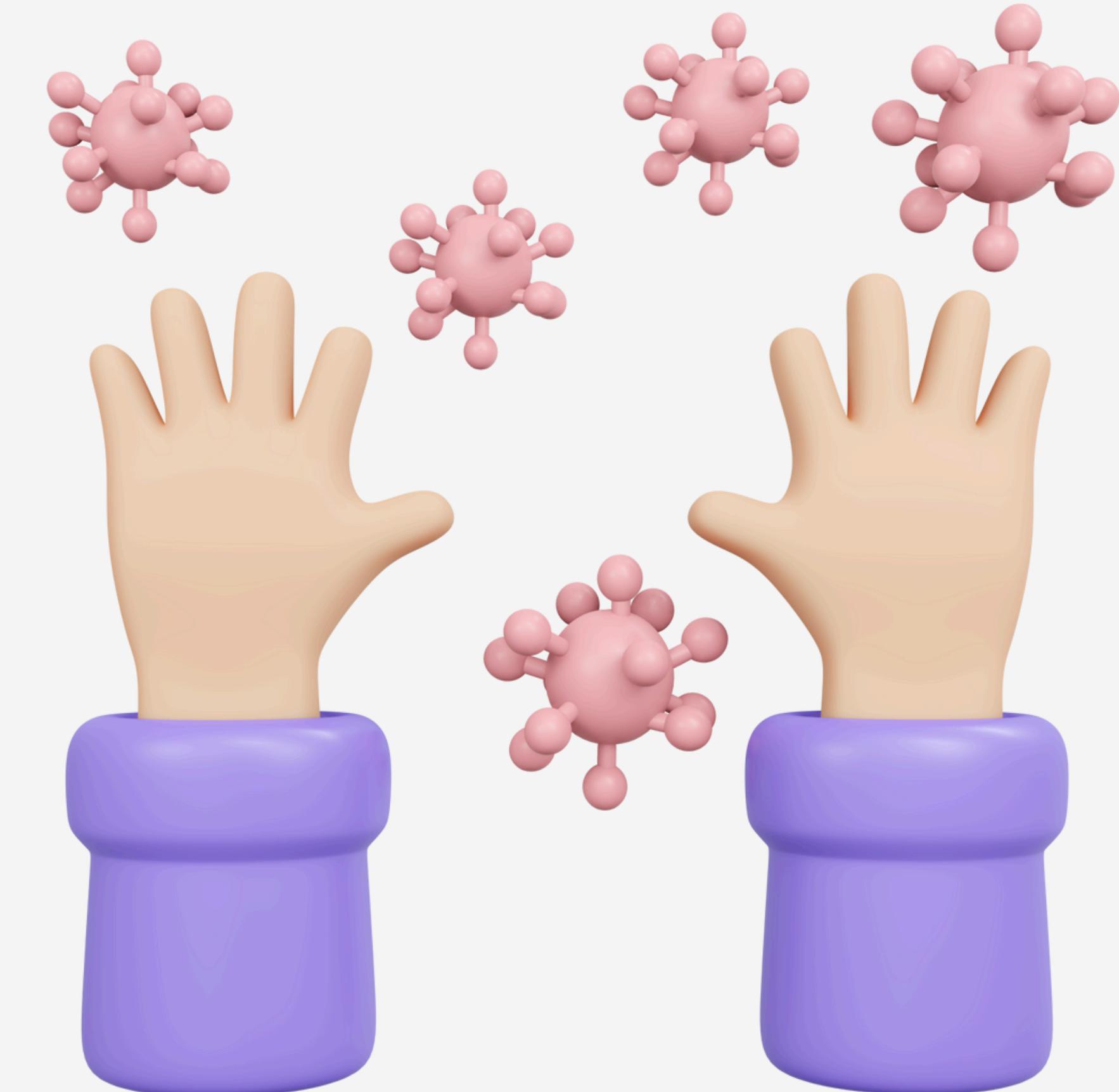
Topics for discussion

01 PROJECT OBJECTIVE

02 DATASET OVERVIEW

03 DATA PROCESSING

04 DATA ANALYSIS



Topic Jusitfication & Project objective



Data Overview

Source

The dataset was provided by the Mexican government and downloaded from the Kaggle platform

Data summary

The dataset contains 1048575 rows, 21 columns

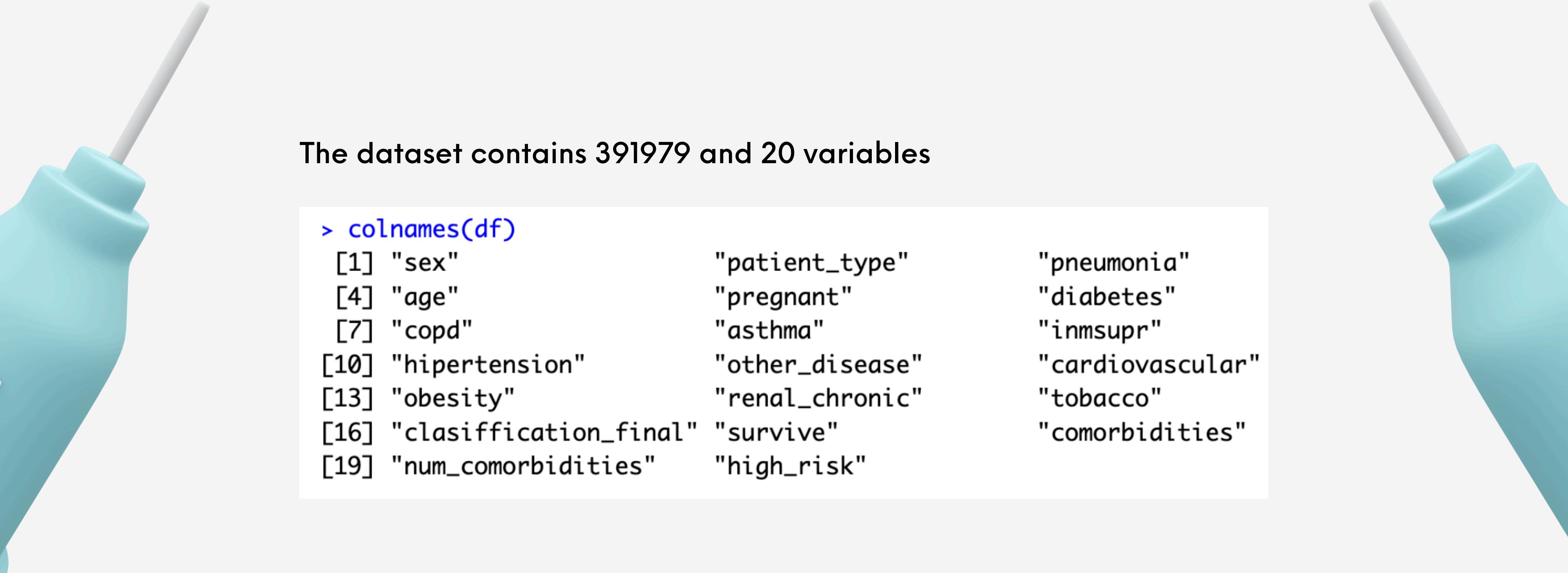


USMER	MEDICAL_UNIT	SEX	PATIENT_TYPE	DATE_DIED
Min. :1.000	Min. : 1.000	Min. :1.000	Min. :1.000	Length:1048575
1st Qu.:1.000	1st Qu.: 4.000	1st Qu.:1.000	1st Qu.:1.000	Class :character
Median :2.000	Median :12.000	Median :1.000	Median :1.000	Mode :character
Mean :1.632	Mean : 8.981	Mean :1.499	Mean :1.191	
3rd Qu.:2.000	3rd Qu.:12.000	3rd Qu.:2.000	3rd Qu.:1.000	
Max. :2.000	Max. :13.000	Max. :2.000	Max. :2.000	
INTUBED	PNEUMONIA	AGE	PREGNANT	DIABETES
Min. : 1.00	Min. : 1.000	Min. : 0.00	Min. : 1.00	Min. : 1.000
1st Qu.:97.00	1st Qu.: 2.000	1st Qu.: 30.00	1st Qu.: 2.00	1st Qu.: 2.000
Median :97.00	Median : 2.000	Median : 40.00	Median :97.00	Median : 2.000
Mean :79.52	Mean : 3.347	Mean : 41.79	Mean :49.77	Mean : 2.186
3rd Qu.:97.00	3rd Qu.: 2.000	3rd Qu.: 53.00	3rd Qu.:97.00	3rd Qu.: 2.000
Max. :99.00	Max. :99.000	Max. :121.00	Max. :98.00	Max. :98.000
COPD	ASTHMA	INMSUPR	HIPERTENSION	OTHER_DISEASE
Min. : 1.000				
1st Qu.: 2.000				
Median : 2.000				
Mean : 2.261	Mean : 2.243	Mean : 2.298	Mean : 2.129	Mean : 2.435
3rd Qu.: 2.000				
Max. :98.000				
CARDIOVASCULAR	OBESITY	RENAL_CHRONIC	TOBACCO	CLASIFICATION_FINAL
Min. : 1.000	Min. : 1.000	Min. : 1.000	Min. : 1.000	Min. :1.000
1st Qu.: 2.000	1st Qu.: 2.000	1st Qu.: 2.000	1st Qu.: 2.000	1st Qu.:3.000
Median : 2.000	Median : 2.000	Median : 2.000	Median : 2.000	Median :6.000
Mean : 2.262	Mean : 2.125	Mean : 2.257	Mean : 2.214	Mean : 5.306
3rd Qu.: 2.000	3rd Qu.: 2.000	3rd Qu.: 2.000	3rd Qu.: 2.000	3rd Qu.:7.000
Max. :98.000	Max. :98.000	Max. :98.000	Max. :98.000	Max. :7.000
ICU				
Min. : 1.00				
1st Qu.:97.00				
Median :97.00				
Mean :79.55				
3rd Qu.:97.00				
Max. :99.00				

Data Processing

- Convert data type of variables
- Convert “sex” variable from 1, 2 to Woman, Man
- Convert “patient_type” variable from 1, 2 to “returned home”, “hospitalization”
- Convert “classification_final” variable from number to “Positive”, “Negative”
- Convert boolean variables from 1, 2 to Yes, No
- Convert missing data from 97, 99 to NA
- Impute missing data
- Drop “intubed”, “icu” variables as they contains more than 85% missing data
- Create “survive” column from date_died column: if the date-died is 9999-99-99 then “Yes”, else “No”
- Create “comorbidities” column from disease columns
- Create “num_comorbidities” column
- Create “high-risk” column: if the patient is older than 60 or has more than 2 comorbidities then “Yes”, else “No”
- Drop unnecessary columns: Some columns such as “usmer”, “medical_unit”, “date_died” are not related to the project’s objective
- Filter dataset with classification_final is “Positive”





The dataset contains 391979 and 20 variables

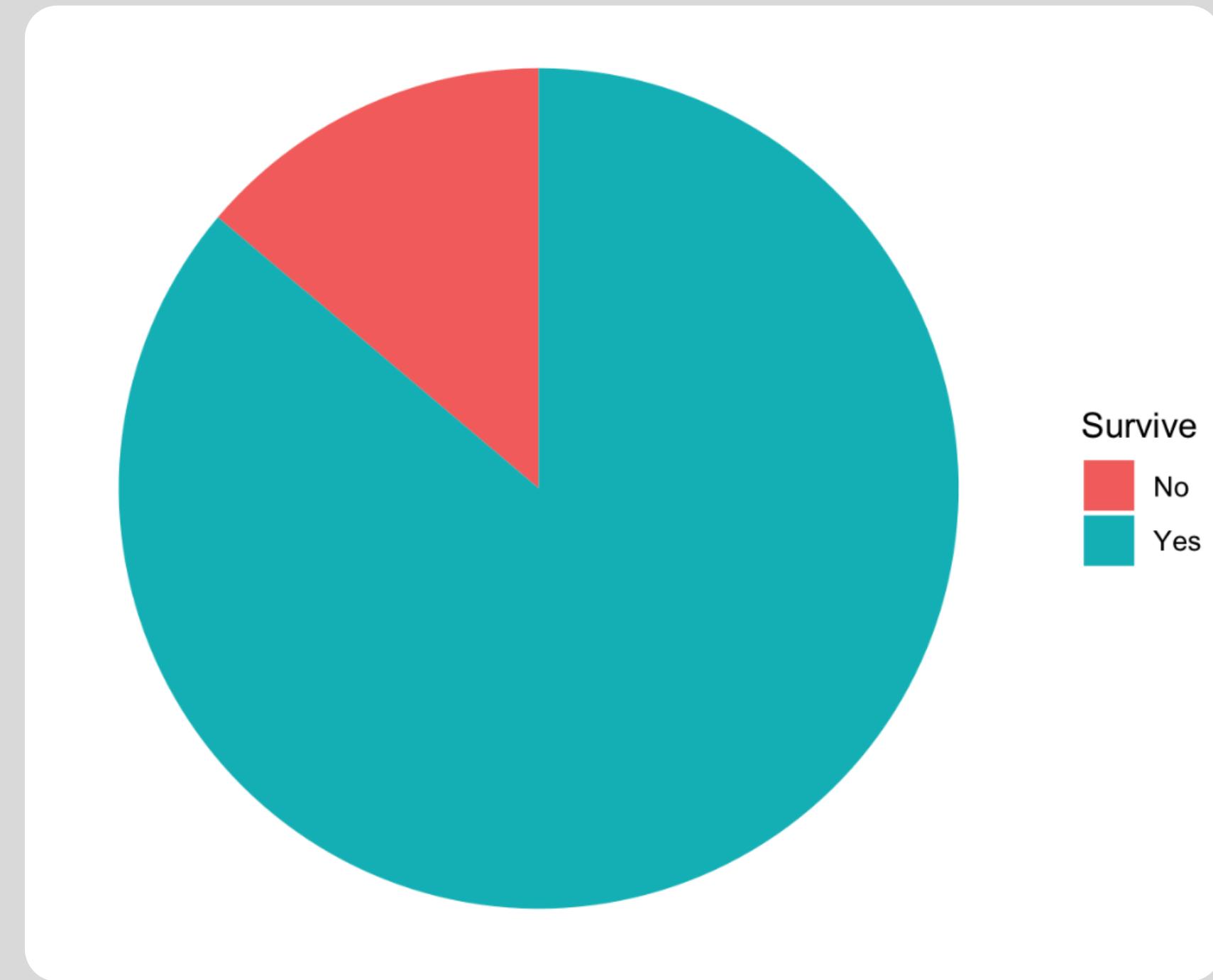
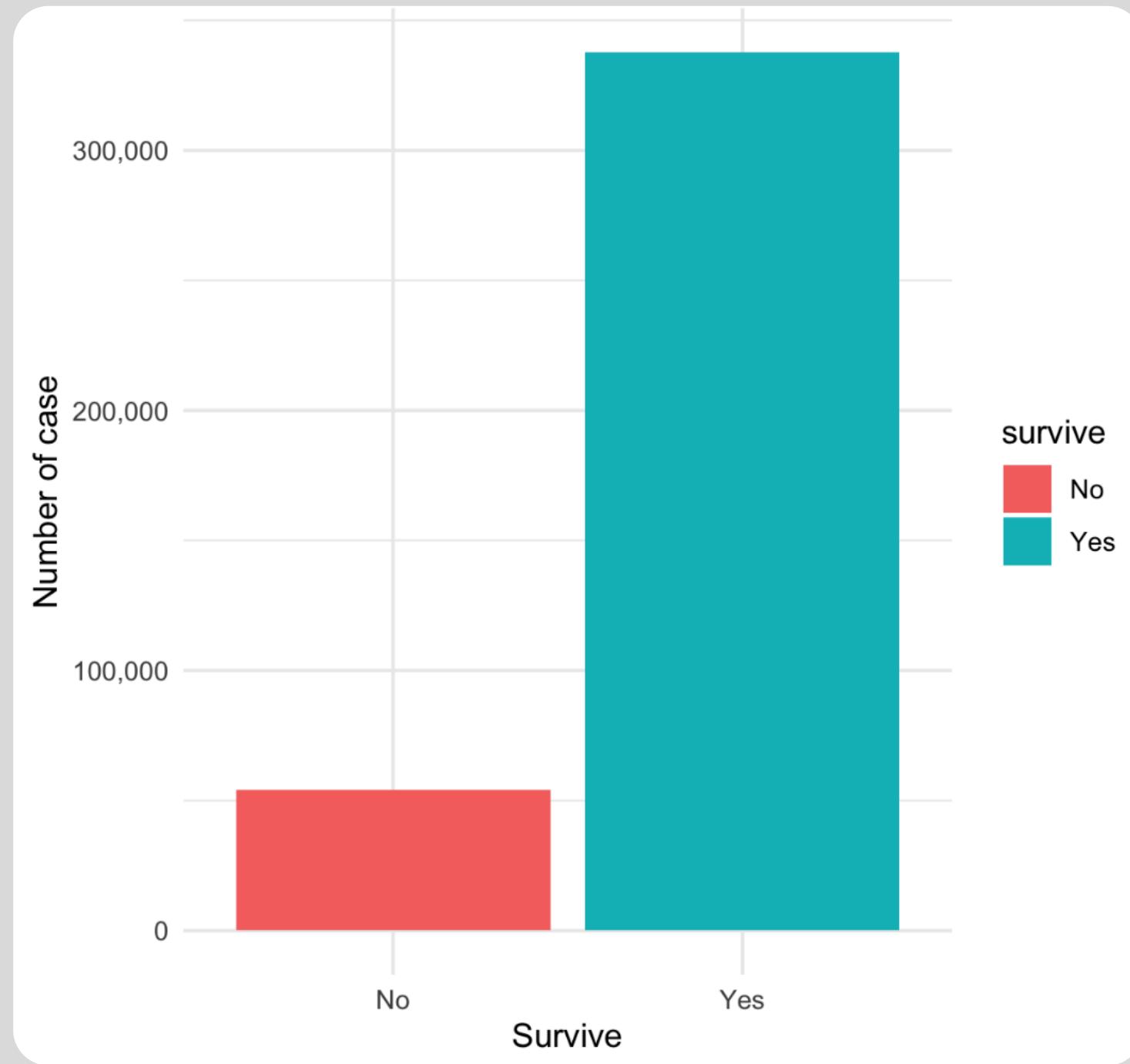
```
> colnames(df)
[1] "sex"                      "patient_type"          "pneumonia"
[4] "age"                       "pregnant"              "diabetes"
[7] "copd"                      "asthma"                "inmsupr"
[10] "hipertension"             "other_disease"        "cardiovascular"
[13] "obesity"                  "renal_chronic"        "tobacco"
[16] "clasification_final"     "survive"               "comorbidities"
[19] "num_comorbidities"        "high_risk"
```



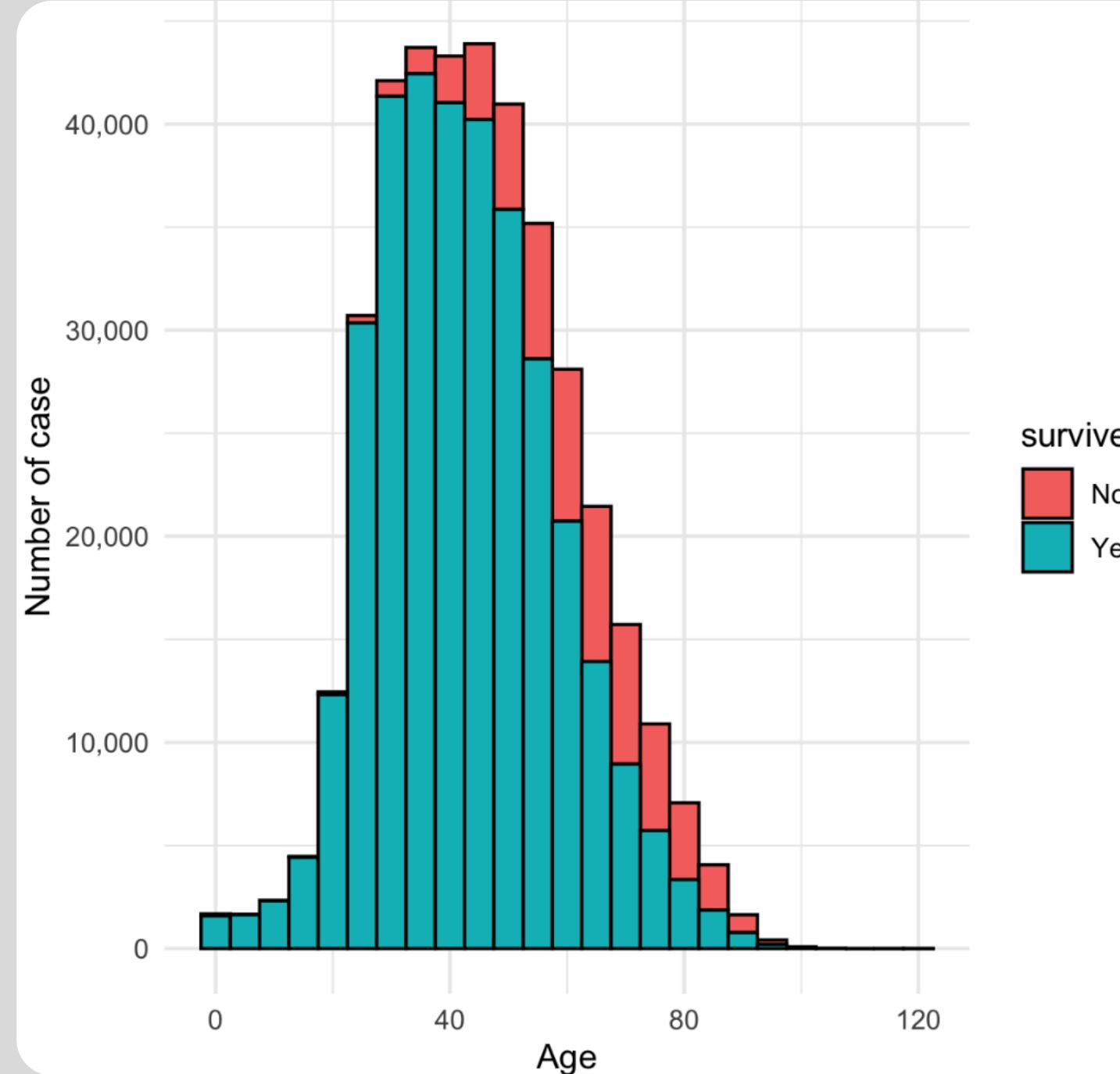
Data Analysis

- What is the death rate of COVID-19 patients?
- Does the death rate correlation with the increasing of human age?
- Does gender affect the death rate?
- Which comorbidity is the most dangerous to COVID-19 patients?
- Are there different between death rate by patient type (home treatment - hospital treatment patients)?
- What is the death rate of COVID-19 patients who are at high risk?
- Are pregnant patients in danger if they are positive to COVID-19

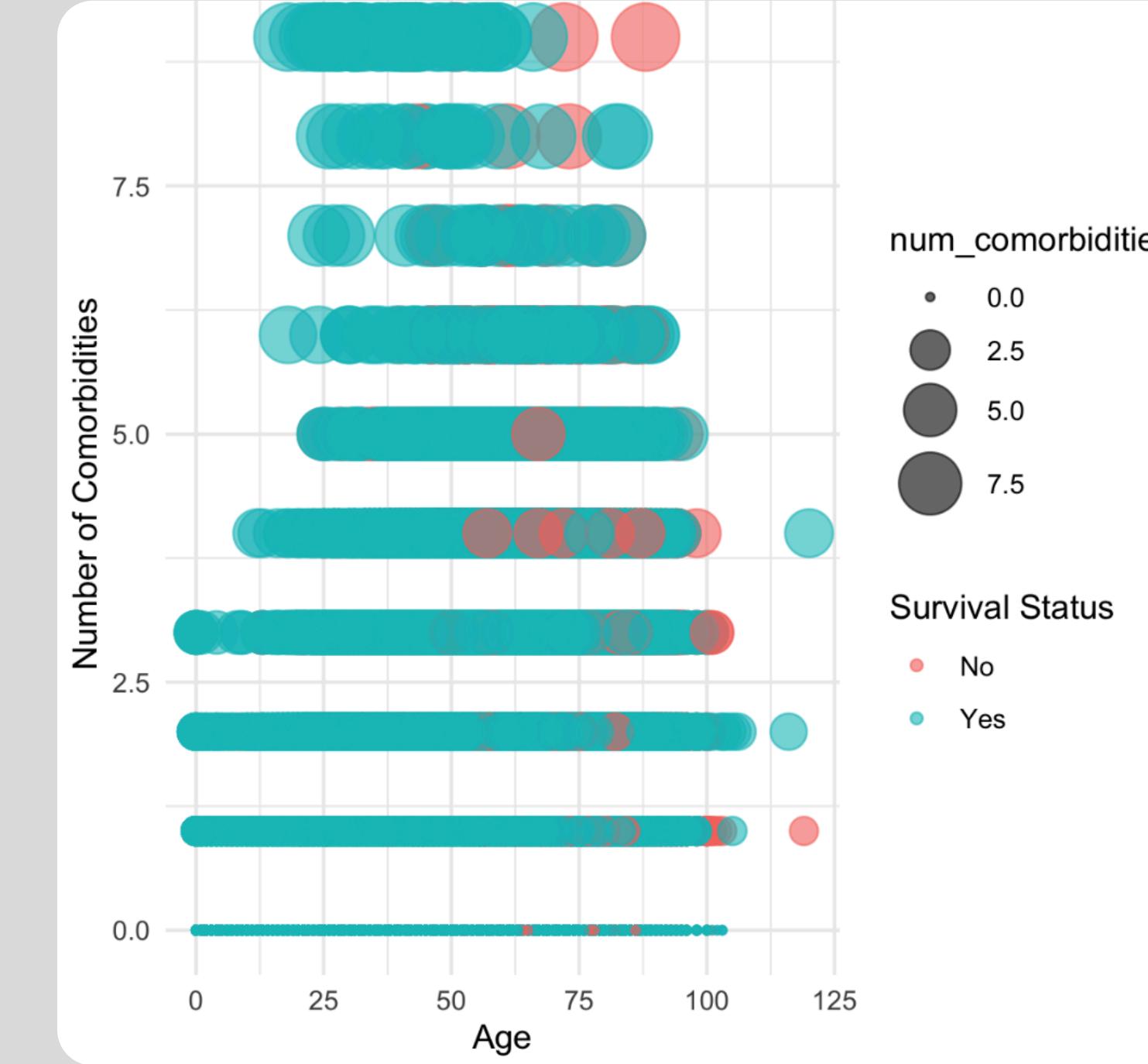
Death rate of COVID-19 patients



Death rate by patients' ages

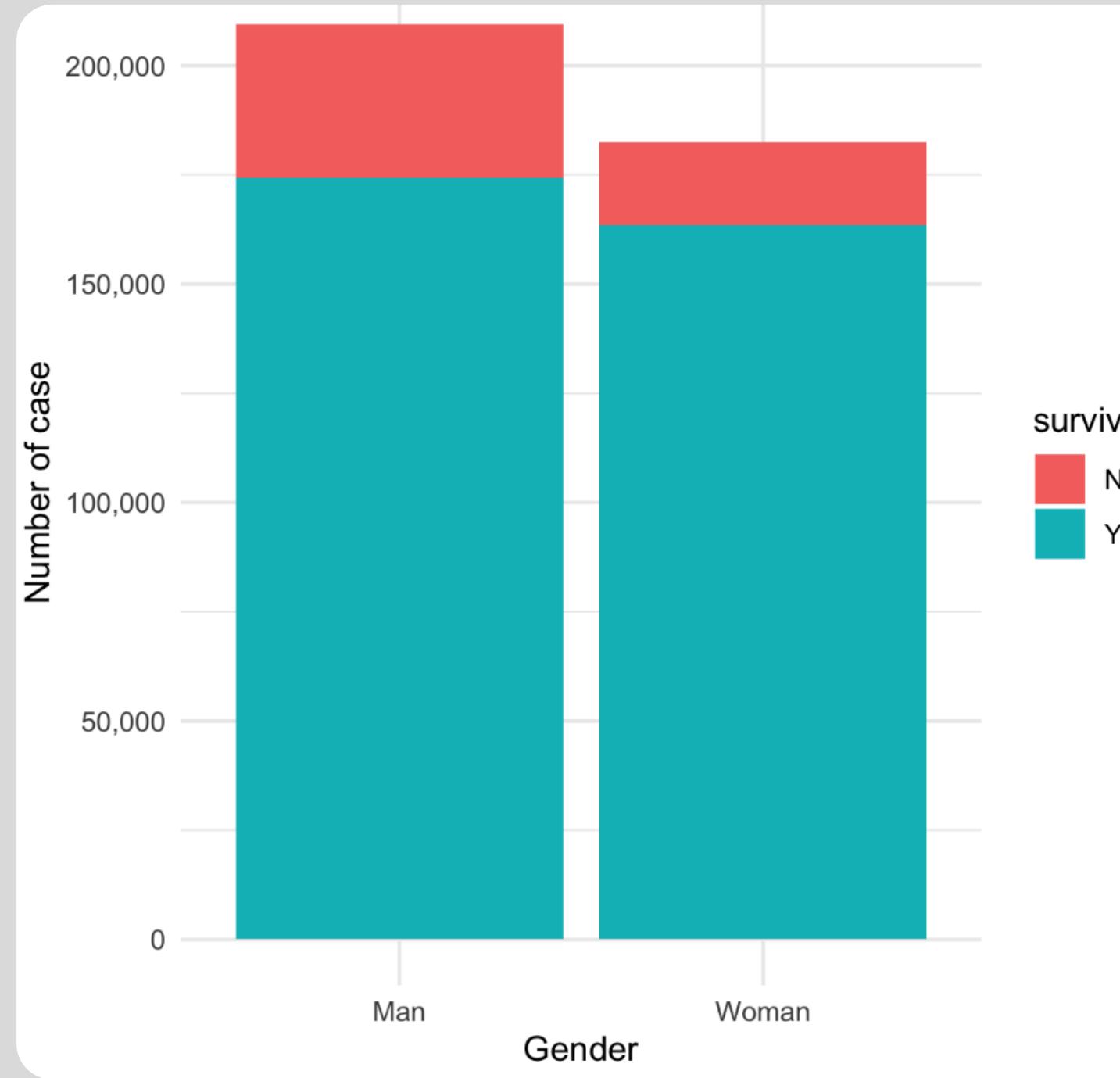


Death rate by age group

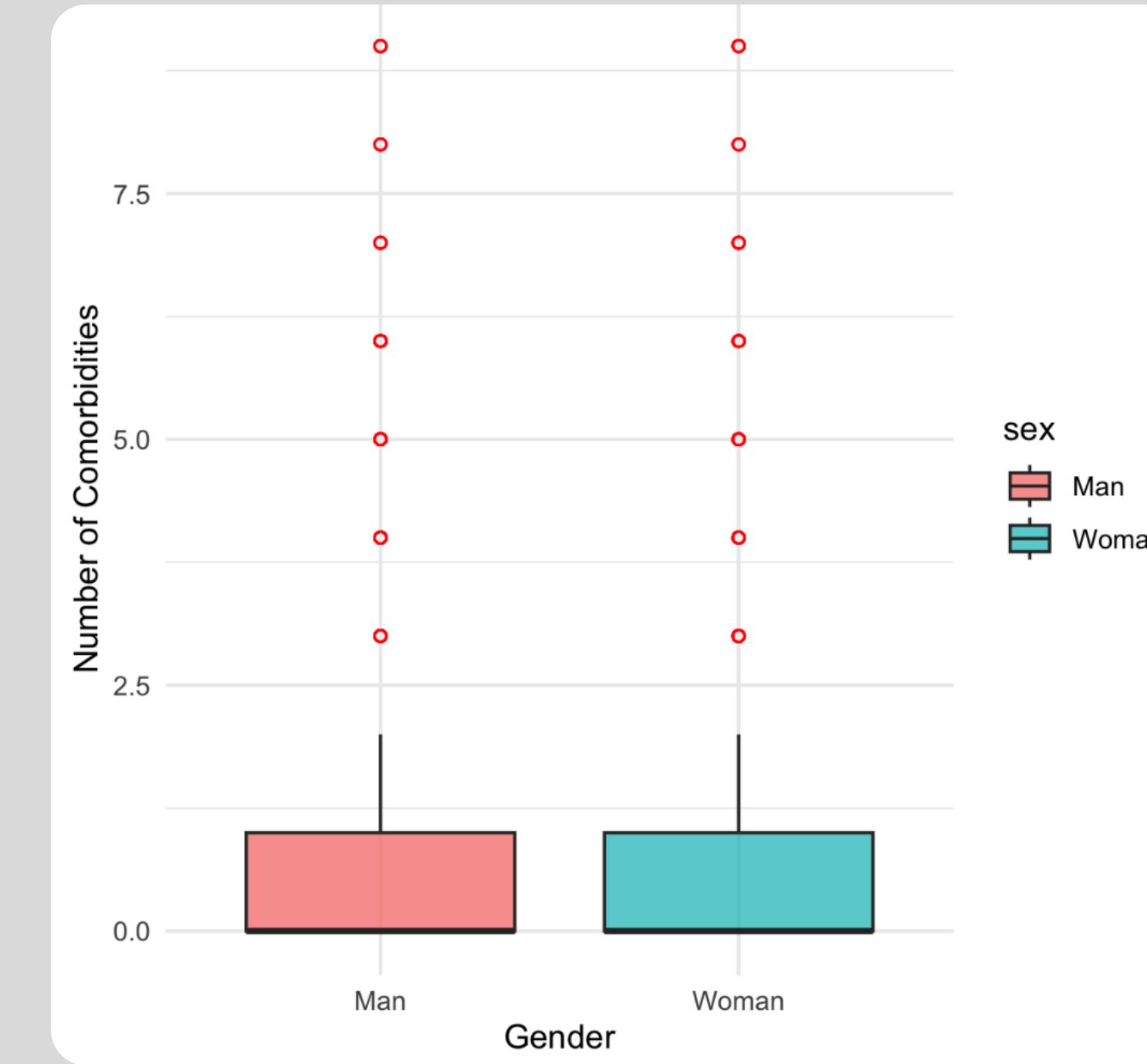


Death rate by number of comorbidities and age group

Death rate by patients' gender

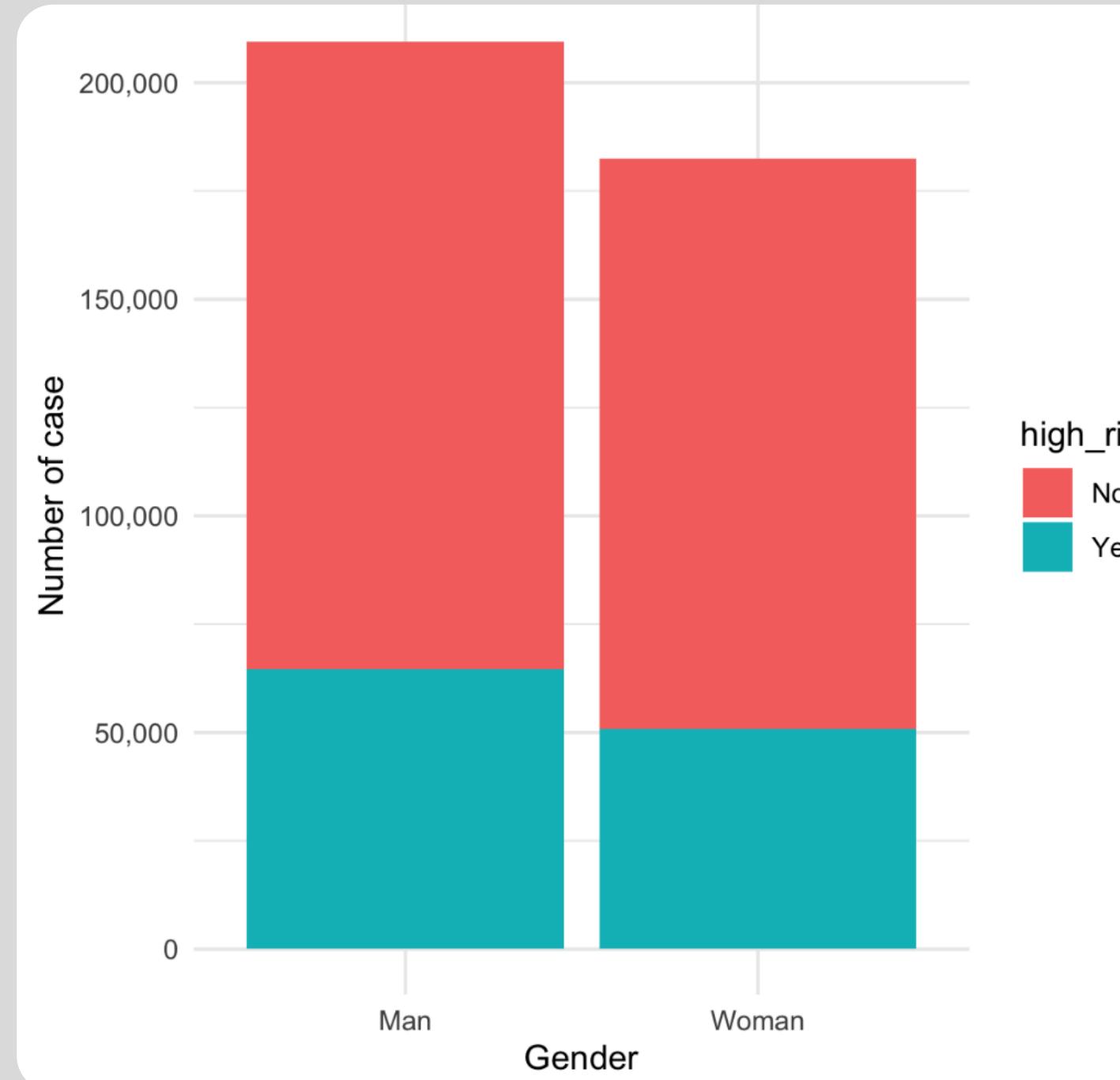


Death rate by gender

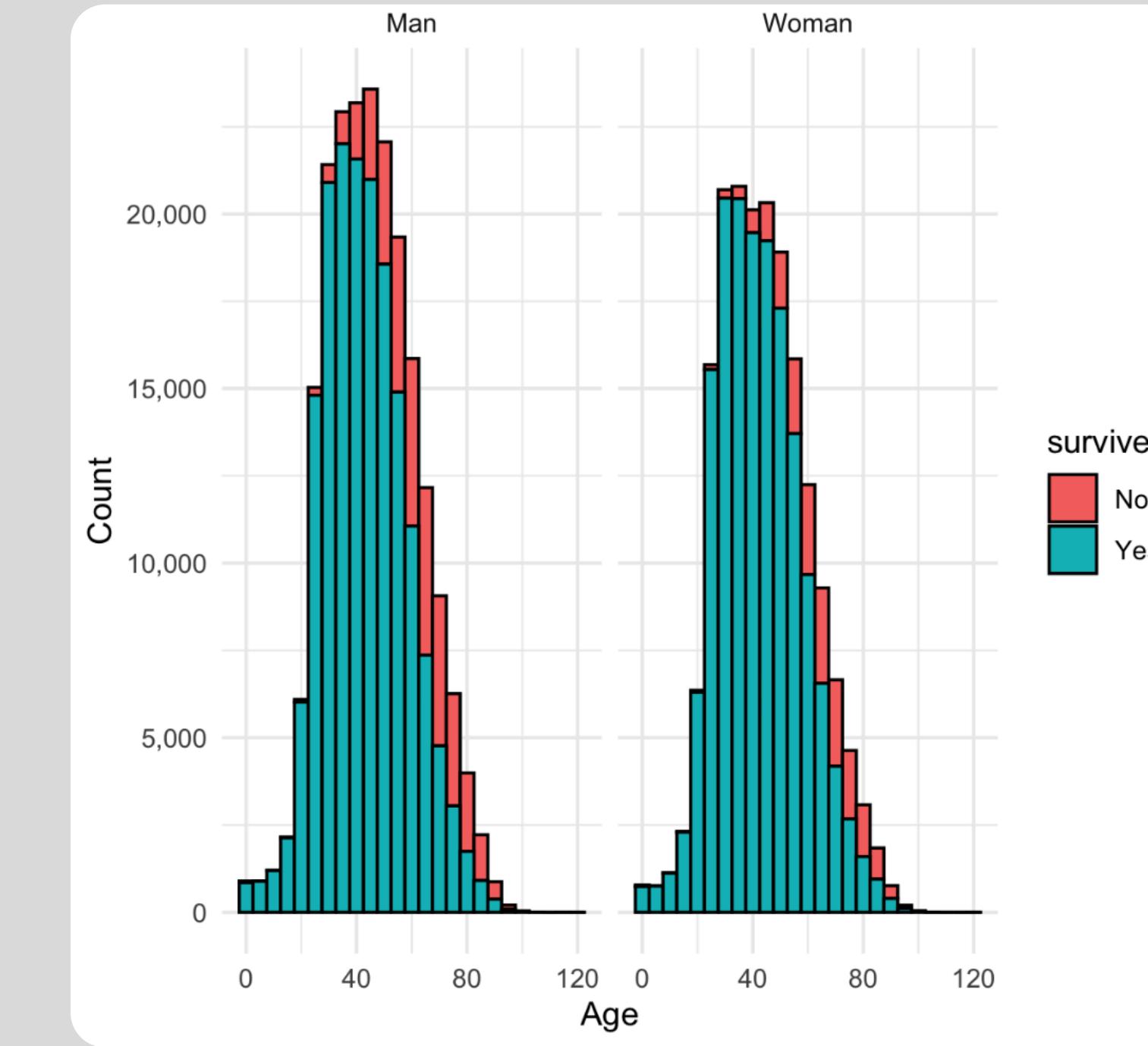


Death rate by number of comorbidities and gender

Death rate by patients' gender

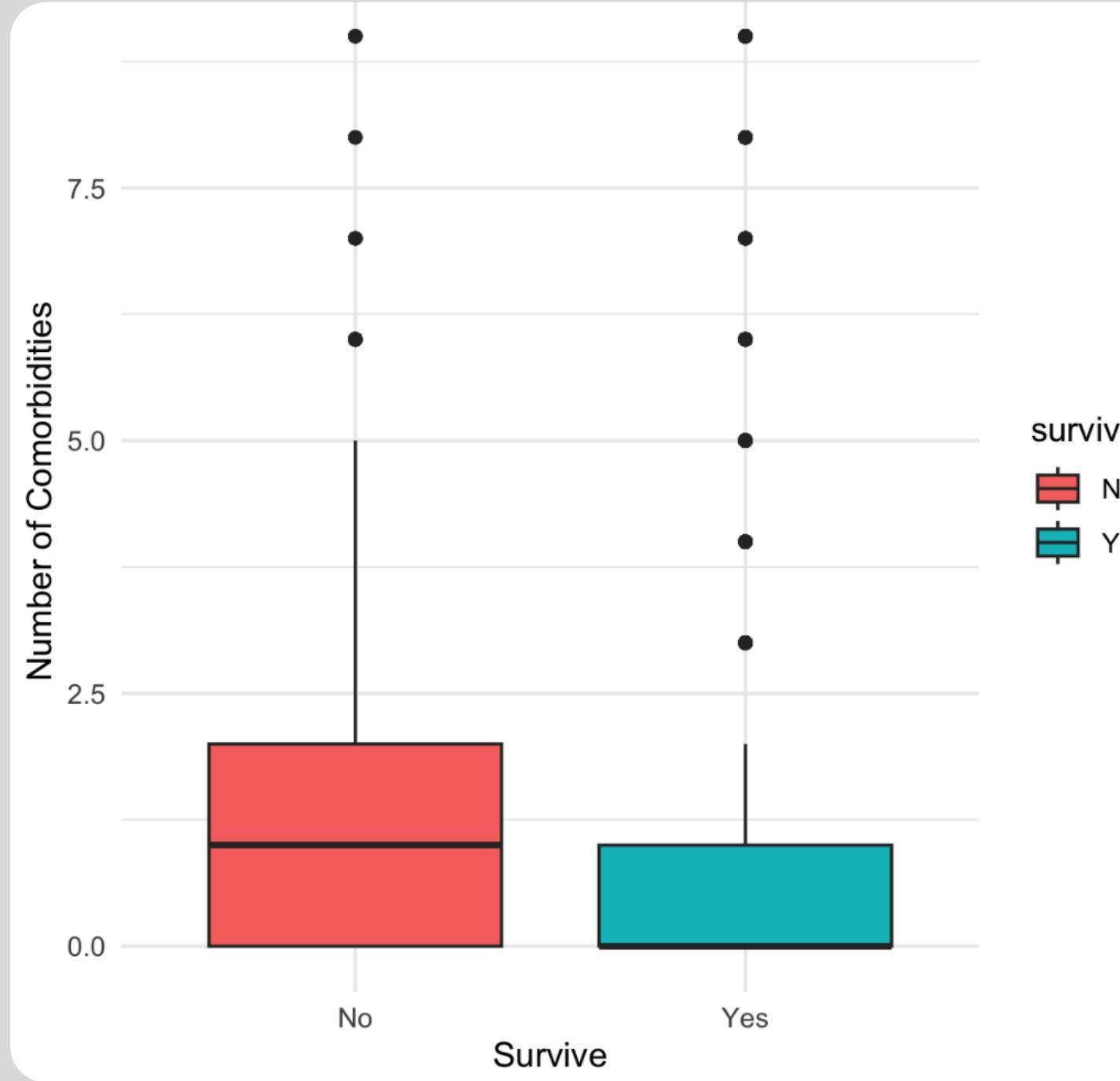


High risk cases by gender

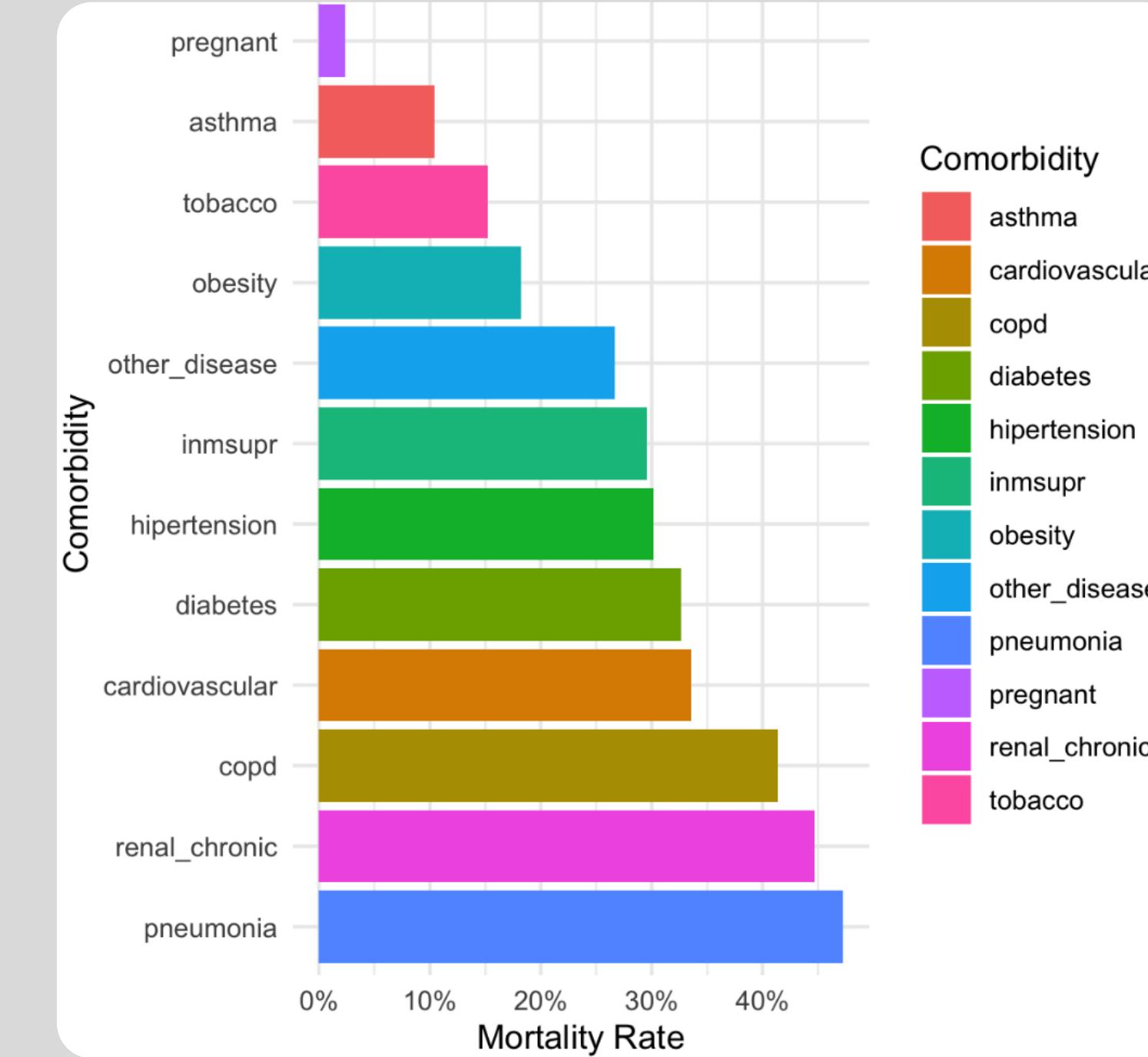


Death rate by gender
and age group

Death rate by comorbidity



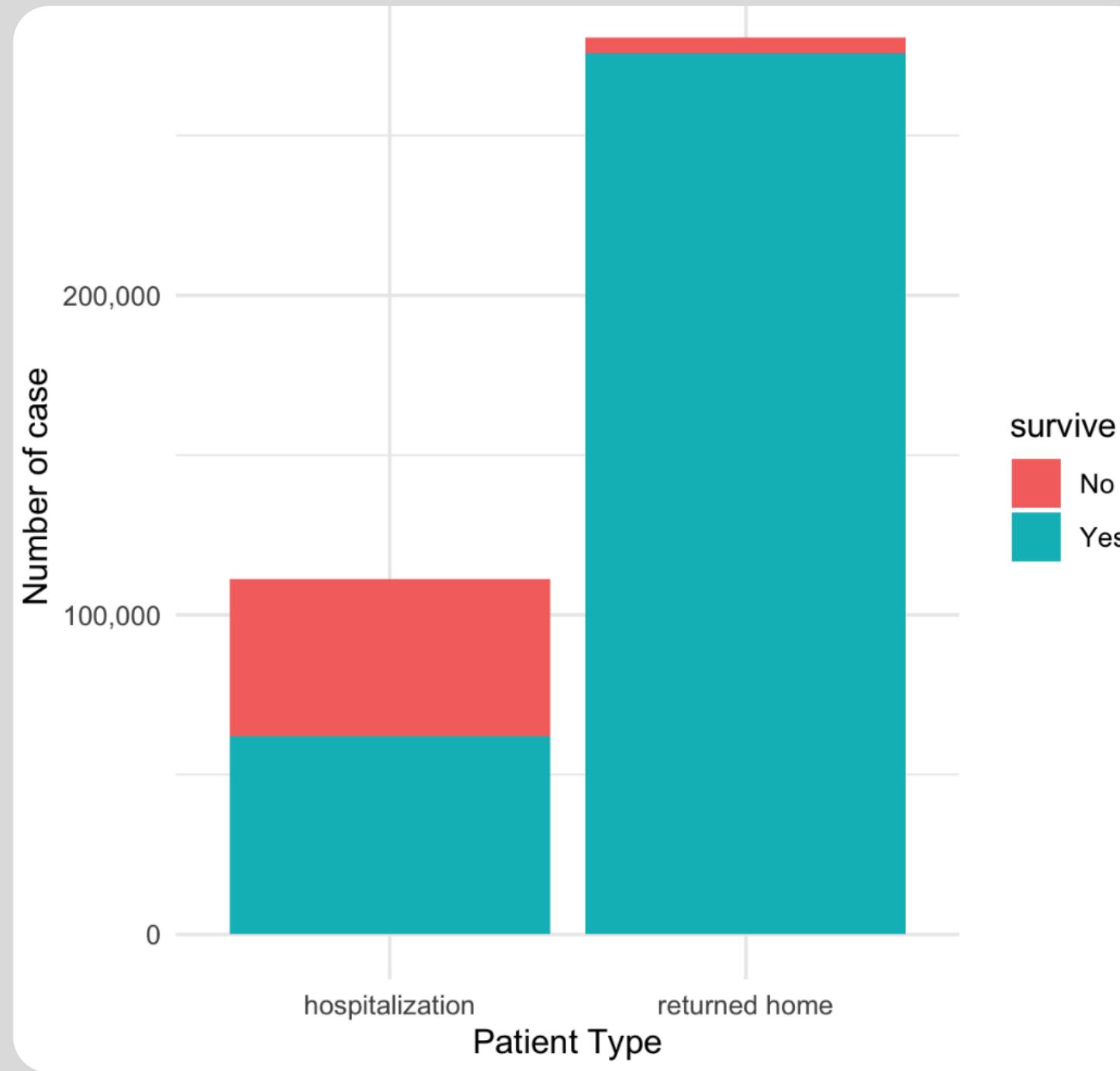
Death rate by number of comorbidity



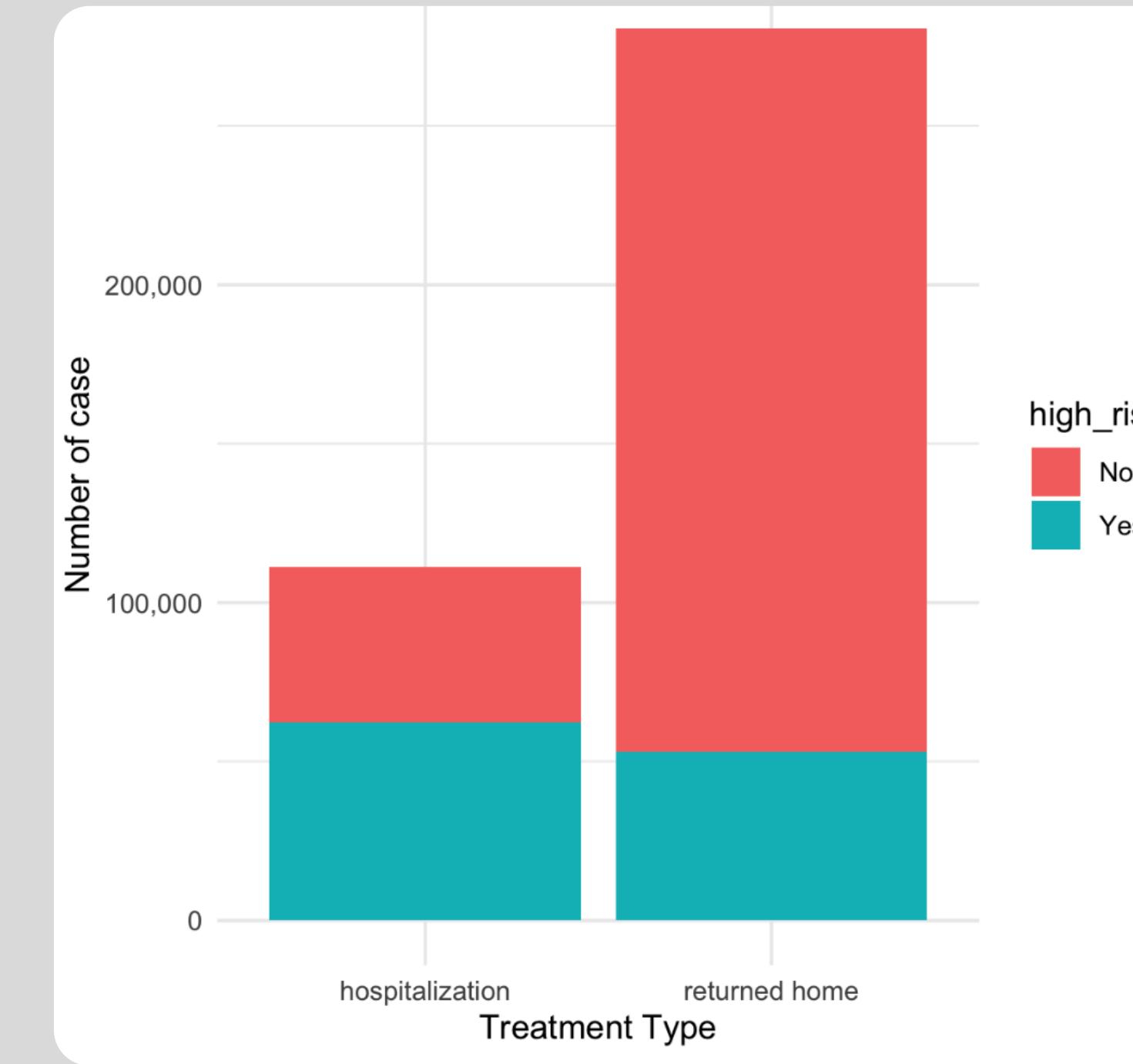
Death rate by comorbidity

Death rate by patient type

(home treatment - hospitalize)



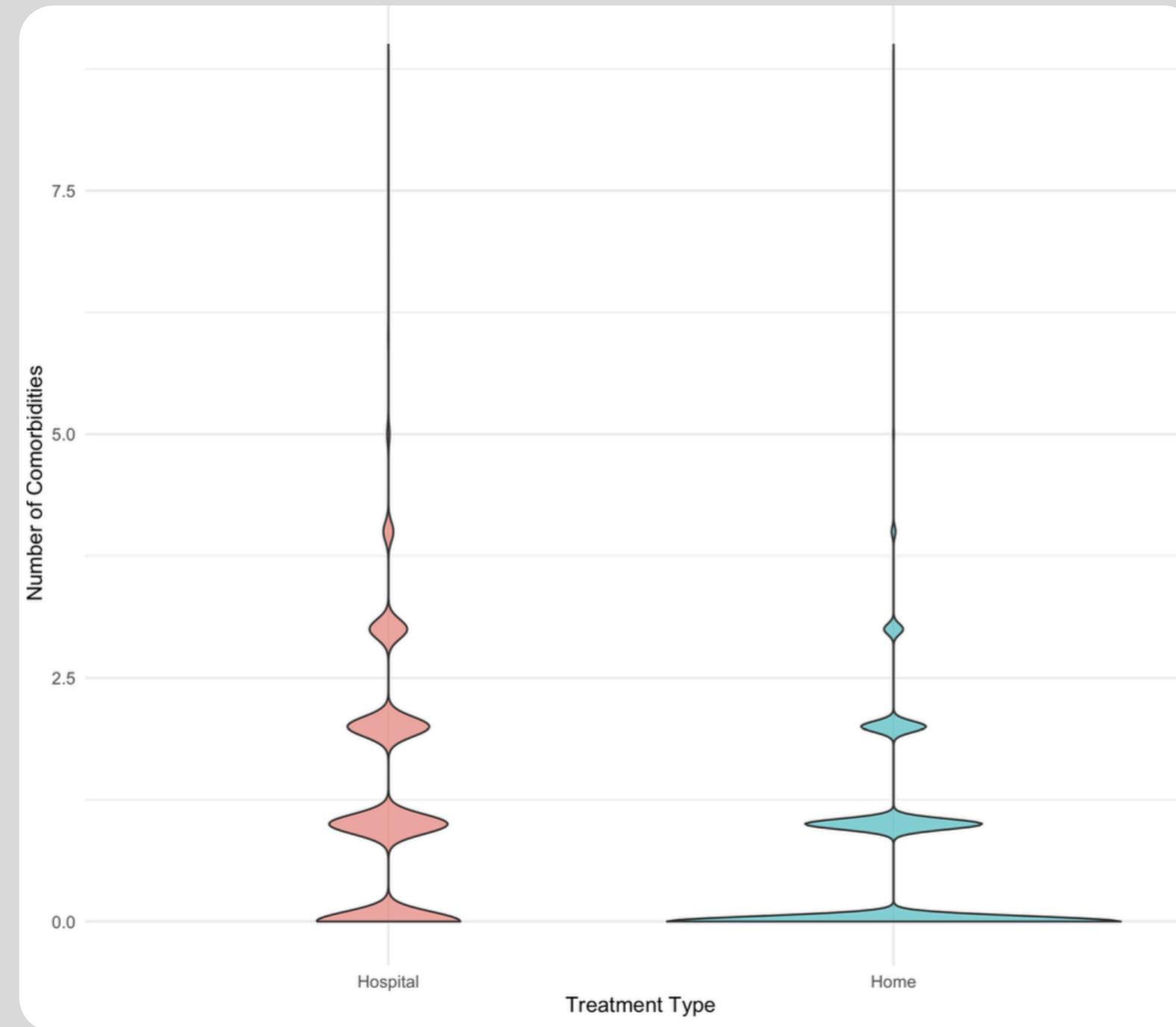
Death rate by patient type



High risk case by patient type

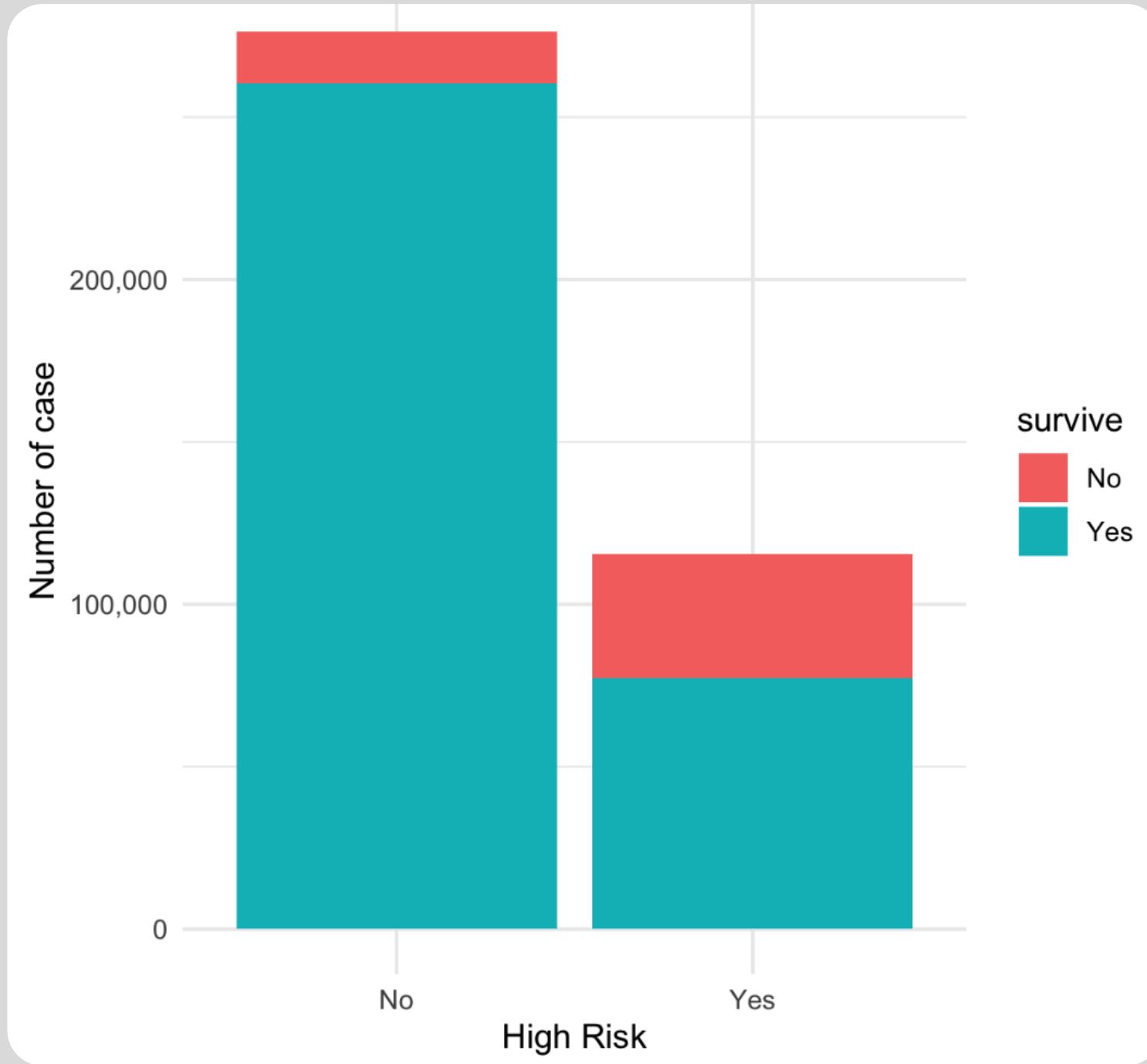
Death rate by patient type

(home treatment - hospitalize)



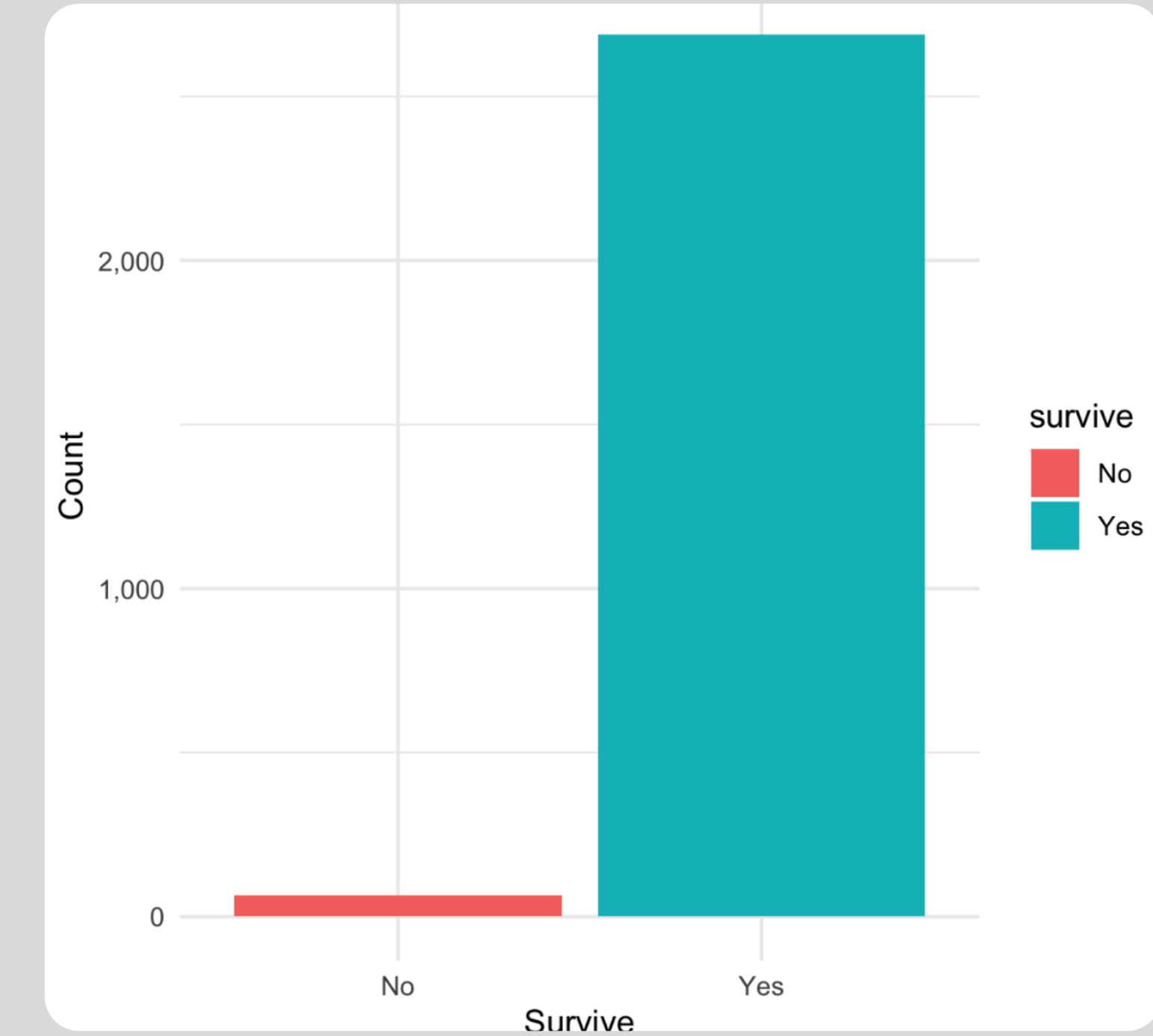
Number of comobidity by patient type

Death rate by high risk



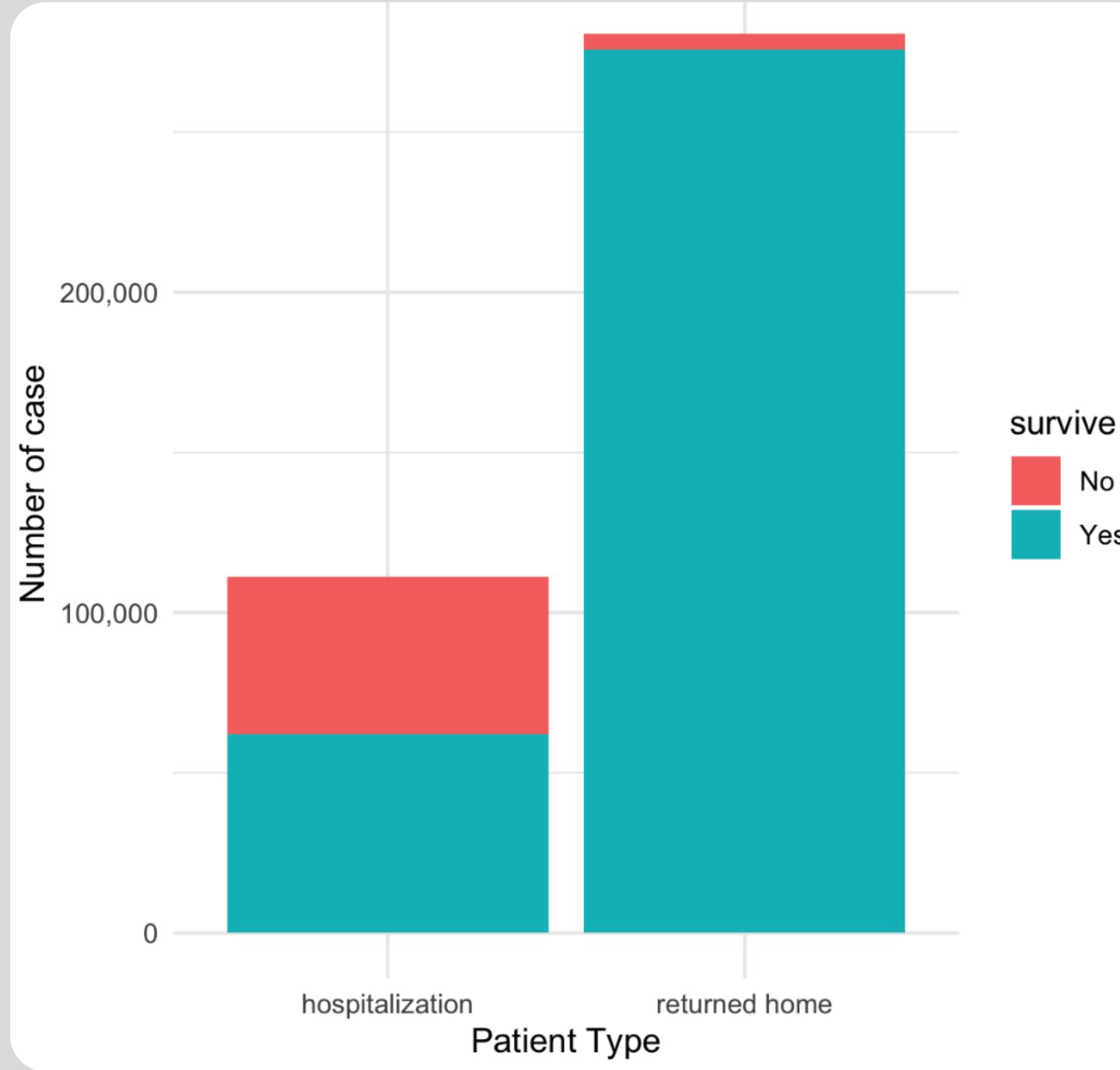
Death rate by high risk

Death rate for pregnant

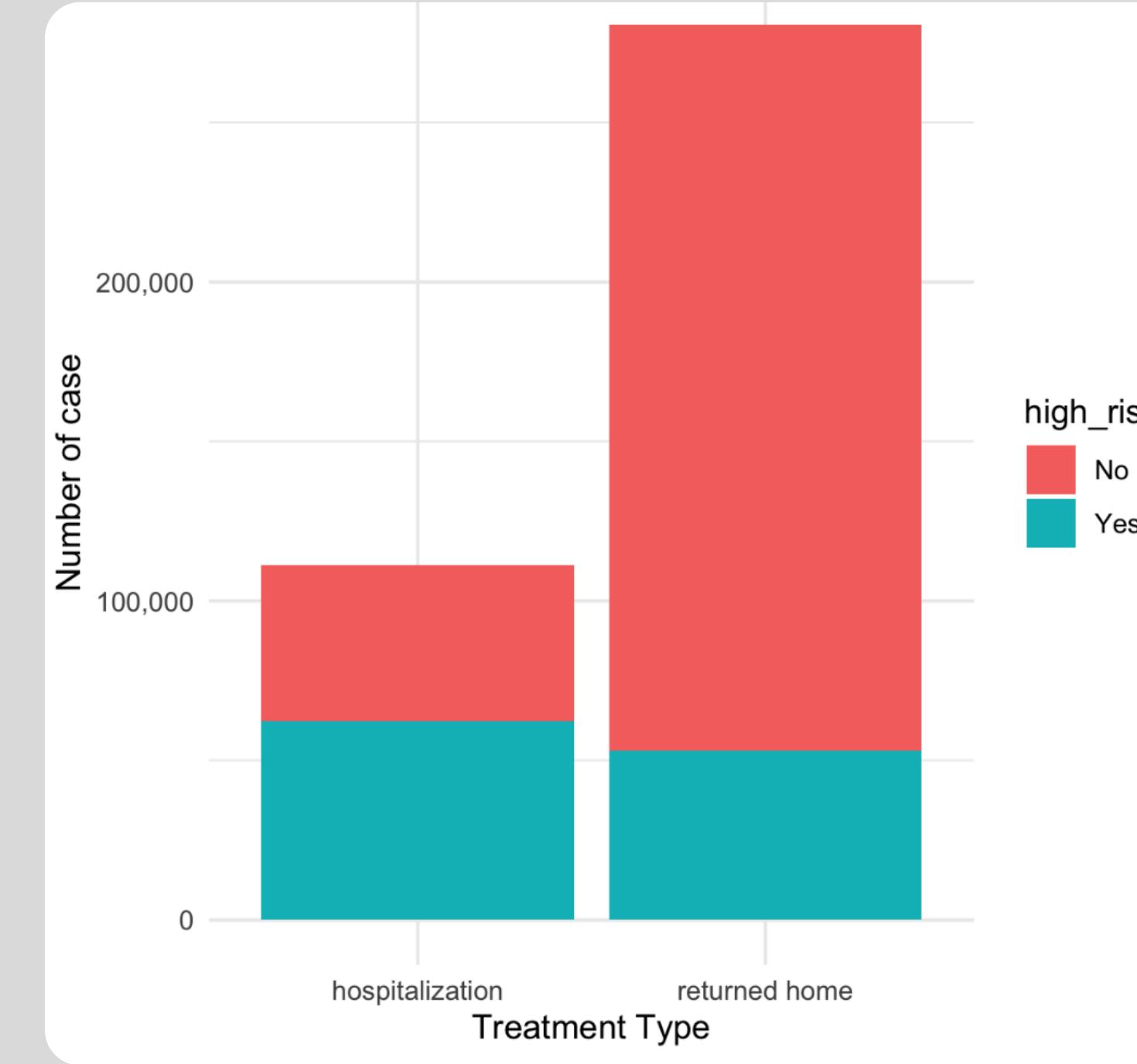


Death rate by pregnancy

Death rate by high risk



Death rate by patient type



High risk case by patient type

Conclusion

- Gradual increase in death rates with age, particularly among the elderly
- Higher severity and mortality rates in men compared to women
- Underlying conditions such as pneumonia, chronic kidney disease, and COPD contributing to the highest mortality rates when combined with COVID-19
- High death rate (~40%) among patients treated in hospitals and a significant mortality rate (~27%) in high-risk cases
- Patients with multiple underlying conditions were often prioritized for hospital treatment

