Data visualization for Scientist Report

Nicolò Romandini - 0001023921 - XXXVII Cycle

---

1 Mortality of ICU Patients

This report presents various visualizations related to numerous patients admitted to the intensive care unit (ICU), with the objective of analyzing specific parameters and their impact on the admission outcome.

1.1 Dataset Description

The dataset was created specifically for the PhysioNet/Computing in Cardiology Challenge 2012 [1], which was a competition and research effort focused on advancing computational cardiology. It involved collaboration between PhysioNet, an online repository of physiological data, and the Computing in Cardiology (CinC) conference. This dataset includes records from 12,000 ICU admissions for adult patients in cardiac, medical, surgical, and trauma ICUs. Within the first 48 hours after ICU admission, up to 42 variables were recorded, but not all variables are available for every case. Six are general descriptors collected upon admission, while the rest are time series data, meaning there could be multiple observations for each variable. Each dataset entry has a timestamp showing how long after ICU admission it was recorded in hours and minutes (e.g., 35:19 indicates 35 hours and 19 minutes after admission).

1.1.1 General Descriptors

These six descriptors are collected at the time the patient is admitted to the ICU:

- RecordID (a unique integer for each ICU stay)
- Age (years)
- Gender (0: female, or 1: male)
- Height (cm)
- ICUType (1: Coronary Care Unit, 2: Cardiac Surgery Recovery Unit, 3: Medical ICU, or 4: Surgical ICU)
- Weight (kg)

1.1.2 Time Series Variables

These 37 variables may be observed once, more than once, or not at all in some cases. They includes various medical measurements and parameters, such as levels of different substances in the blood (like albumin and bilirubin), blood pressure, oxygen levels, heart rate, and more. It helps analyze and understand patients' health conditions. Given that these variables were not utilized in the visualizations, there was no need for additional descriptions.

1.1.3 Outcome-related Descriptors

Each line of the outcomes file contains these descriptors:

- RecordID (defined as above)
- SAPS-I score

- SOFA score
- Length of stay (days)
- In-hospital death (0: survivor, or 1: died in-hospital)

The Length of Stay represents the duration in days from a patient's admission to the ICU until the end of their hospitalization, which includes any post-ICU hospitalization period. In cases where the patient's death was documented (whether inside or outside the hospital), Survival measures the number of days between ICU admission and death. If no death record exists, Survival is set to -1. It's important to note that due to the exclusion of patients with less than 48 hours of ICU stay, neither Length of Stay nor Survival in the challenge datasets ever have values of 0 or 1.

2 Tools and Visualizations

In all the visualizations, two Python libraries were utilized: matplotlib [2] for the first three graphs, and seaborn [3] for the last one. In addition, for all the charts, Arial font and a color palette suitable for colorblind individuals were set.

The first bar chart (Figure 1) illustrates the count of patients who survived and those who passed away during their hospitalization, revealing a distinct advantage for the survivors. This bar chart offers an immediate overview for evaluating the most numerous category. Additionally, hatch patterns have been incorporated to improve readability, particularly for black and white printing.

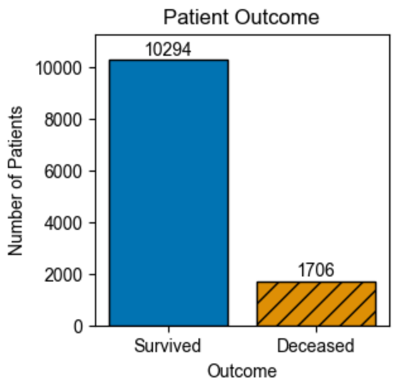

Figure 1: Outcome of patients.

The second bar graph (Figure 2) illustrates the distribution of patients' survival and mortality across different age categories. The age categories, ranging from '0-18' to '80+', are defined on the x-axis. The graph employs a grouped bar chart format, with each age category

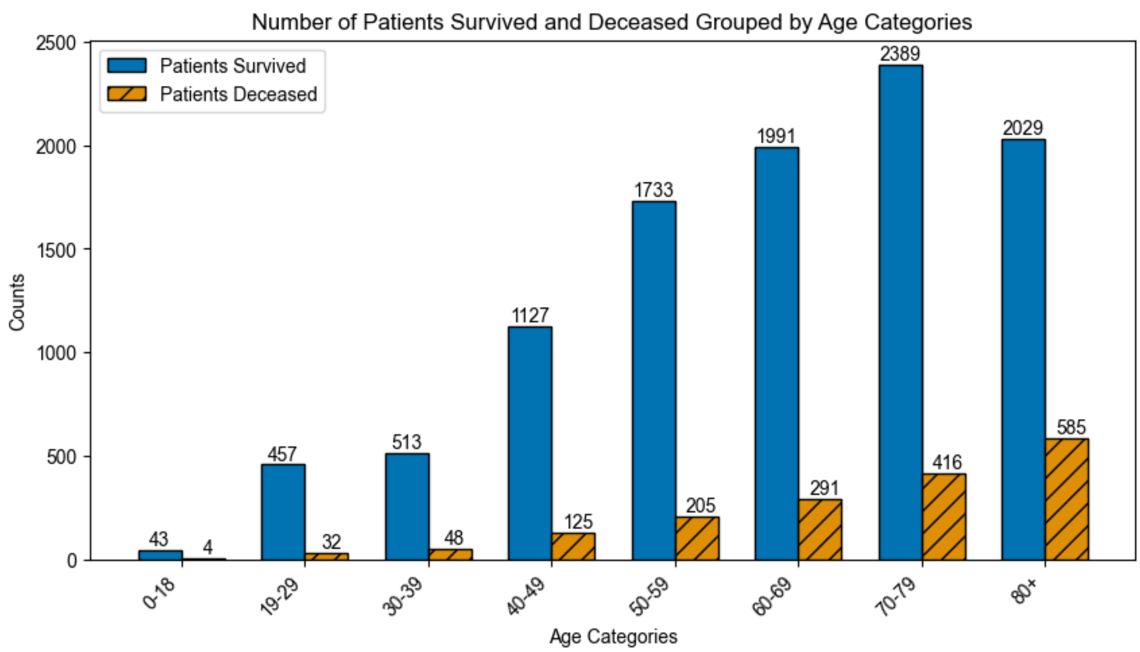

Figure 2: Patient distribution and relative outcome according to age category.

displaying two bars: one representing the count of patients who survived and the other indicating the count of patients who deceased. The y-axis denotes the counts, and labels displaying the exact counts are positioned just above each bar for clarity. The bars representing patient mortality are distinguished with diagonal hatching. This visualization effectively communicates how patient outcomes vary across different age groups, offering insights into the impact of age on survival and mortality rates within the given dataset.

The third graph (Figure 3) is a pie chart depicting the distribution of deceased patients among different types of ICUs. At first glance, it's evident that the majority of deceased patients were admitted to the Medical ICU. This pie chart provides a rapid and visual means to assess the percentage of deceased patients based on their admission location. Additionally, hatch patterns have been incorporated into this visualization to enhance readability. The exclusion of the first two colors from the palette was deliberate, as they had already been utilized in the previous graphs. This approach ensures a distinct and consistent color scheme throughout the visualizations.

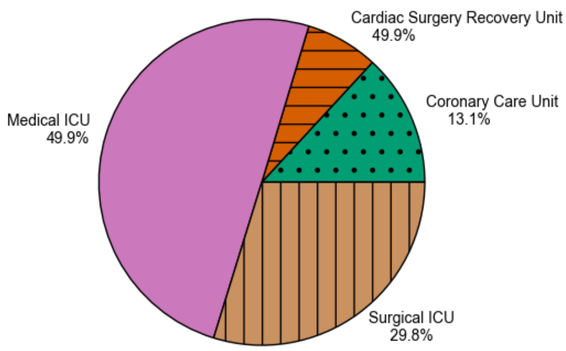

Figure 3: Distribution of patients according to which ICU they were admitted to.

The last graph (Figure 4) focuses specifically on deceased patients within the dataset. It utilizes a horizontal violin plot to visualize the age distribution for deceased patients, further categorized by gender. The horizontal orientation of the plot allows for a clear comparison of age distributions between male and female deceased patients. The inner quartiles of the violins provide insights into the distribution's central tendency and spread. The y-axis label is intentionally left empty for a more compact and focused visualization. In this case as well, colors from the palette that hadn't been previously used were selected. This graph highlights that, on average, deceased women had a higher age compared to men.

Figure 4: Age distribution for deceased patients by gender.

References

[1] I. Silva, G. Moody, D. J. Scott, L. A. Celi, and R. G. Mark, "Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012," *2012 Computing in Cardiology*, pp. 245–248, 2012.

[2] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[3] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.