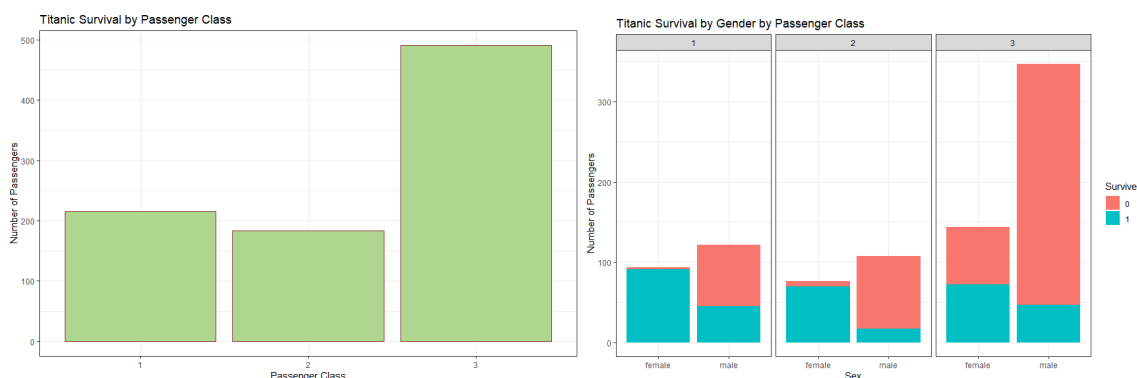


Analiza statystyczna - PROJEKT w R

Przyjąłem dane z pliku .csv (dostępnego w plikach projektu) na temat pasażerów „Titanica”. Pasażerowie są podzieleni na płeć, klasę przedziału oraz czy przeżyli. Dodatkowo jest informacja na temat wieku. Z reszty danych w pliku nie korzystamy.

Prostym poleceniem w R pakiet ggplot2 umożliwia wyświetlenie ilości pasażerów z podziałem na klasy (wykres lewy), a w następnym kroku z dodatkowym podziałem na płeć oraz czy przeżyli (wykres prawy):



(na tym diagramie możemy zaobserwować, że kobiety, w szczególności z klasy pierwszej, miały pierwszeństwo w ewakuacji – przewaga koloru niebieskiego nad czerwonym)

Używając opcji histogramu możemy do poprzedniego wykresu dodać jeszcze podział na wiek:



(na tym diagramie dodatkowo możemy zaobserwować, że dzieci, w szczególności te z klasy pierwszej, również miały pierwszeństwo w ewakuacji – przewaga koloru niebieskiego nad czerwonym)

Następujące wykresy przedstawiają rozkłady wieku osób, w formie boxplot'u (który wskazuje nam kwantyle, medianę oraz wartości odstające), oraz violin'u który nam pozwala lepiej zobrazować rozkład:

