

Analiza objaśniająca szanse przeżycia katastrofy Titanica

Adam Komosiński

Wstęp

W raporcie przedstawiona została analiza szans na przeżycie pasażerów statku Titanica. Spośród atrybutów w zbiorze danych Titanica oczyszczonego z brakujących i niespójnych danych wybrano te, które mogły mieć decydujący wpływ na przeżywalność pasażerów.

##	PassengerId	Survived	Pclass	Name	Sex	Age
##	Min. : 1.0	Mode :logical	Min. :1.000	Length:679	Length:679	Min. : 0.9
##	1st Qu.: 226.5	FALSE:403	1st Qu.:1.000	Class :character	Class :character	1st Qu.:20.0
##	Median : 446.0	TRUE :276	Median :2.000	Mode :character	Mode :character	Median :28.0
##	Mean : 450.4		Mean :2.231			Mean :29.7
##	3rd Qu.: 675.0		3rd Qu.:3.000			3rd Qu.:38.0
##	Max. :1000.0		Max. :3.000			Max. :80.0
##	Parch	Ticket	Fare	Cabin	Embarked	
##	Min. :0.0000	Length:679	Min. : 0.00	Length:679	Length:679	
##	1st Qu.:0.0000	Class :character	1st Qu.: 8.05	Class :character	Class :character	
##	Median :0.0000	Mode :character	Median : 15.74	Mode :character	Mode :character	
##	Mean :0.4153		Mean : 33.14			
##	3rd Qu.:1.0000		3rd Qu.: 33.76			
##	Max. :5.0000		Max. :263.00			

Wybrane atrybuty:

1. Płeć

- najpierw ratowane są kobiety i dzieci, później ratuje się mężczyzn więc płeć ma decydujące znaczenie w kwestii przetrwania

2. Wiek

- większe prawdopodobieństwo przetrwania najmłodszych

3. Klasa biletu

- procentowo większy udział uratowanych będzie mieć klasa 1 biletów (najwyższa) niż klasa 3cia (najniższa)

4. Liczba rodziców/dzieci

- pierwsze są ratowane kobiety i dzieci więc osoby z dziećmi/rodzicami mają większy współczynnik przeżywalności niż osoby bezdzietne

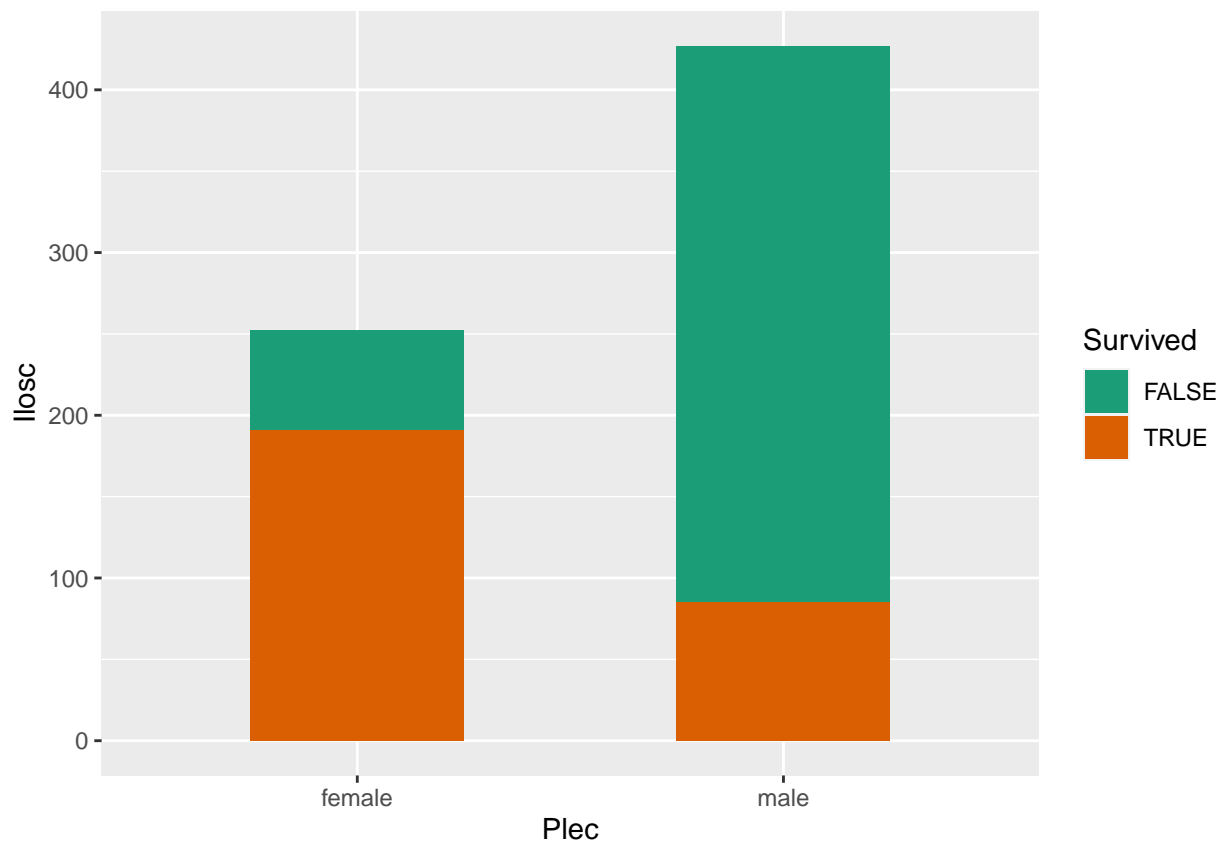
5. Liczba rodzeństwa/małżonków

- warto sprawdzić czy nie istnieje jakaś korelacja między tym atrybutem a przeżywalnością, a jeśli tak to z czego wynika (czy np. osoby z większą ilością rodzeństwa, małżonków wybierały wyższą klasę biletu)

6. Port odprawy

- atrybut raczej nie wpływający na przeżywalność ale być może z nią skorelowany, a jeśli tak to najprawdopodobniej będzie miało znaczenie powiązanie z którymś z bardziej znaczących atrybutów

Płeć

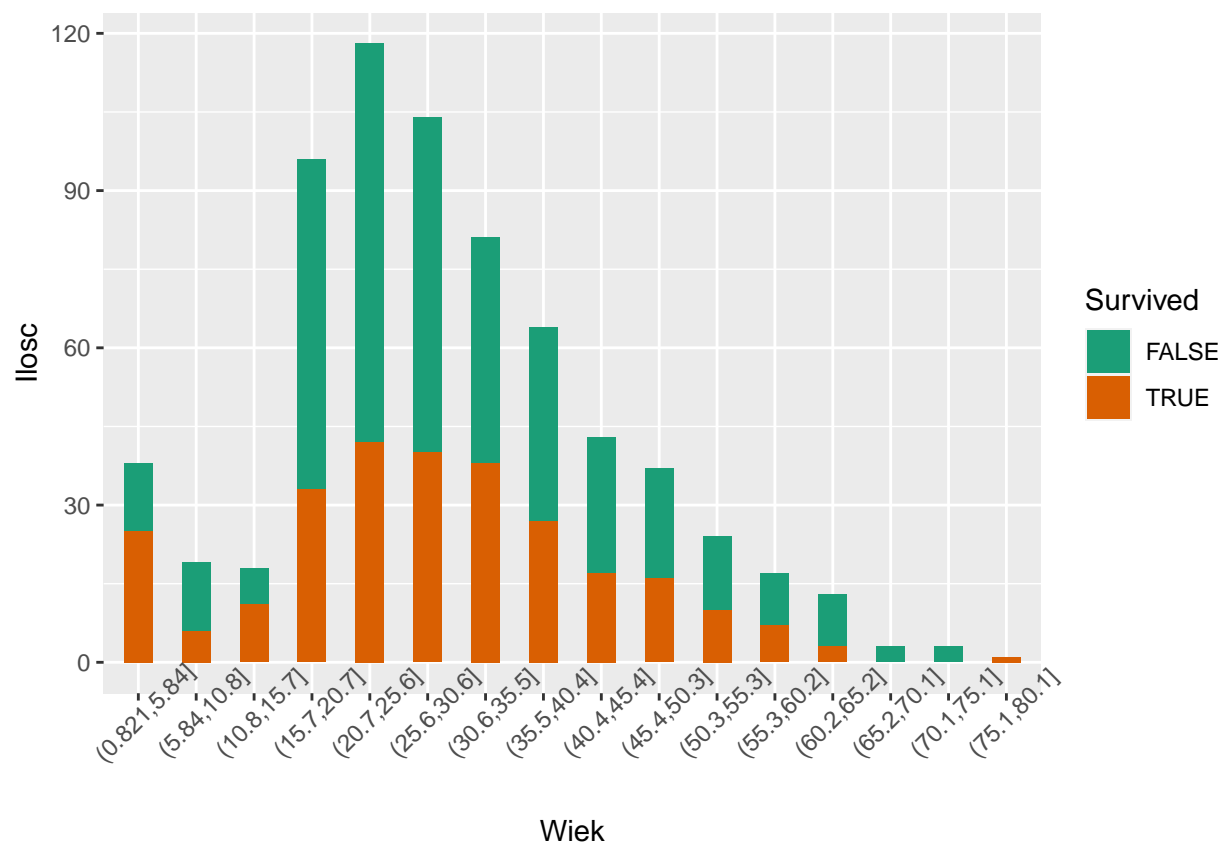


Przeżywalność a płeć

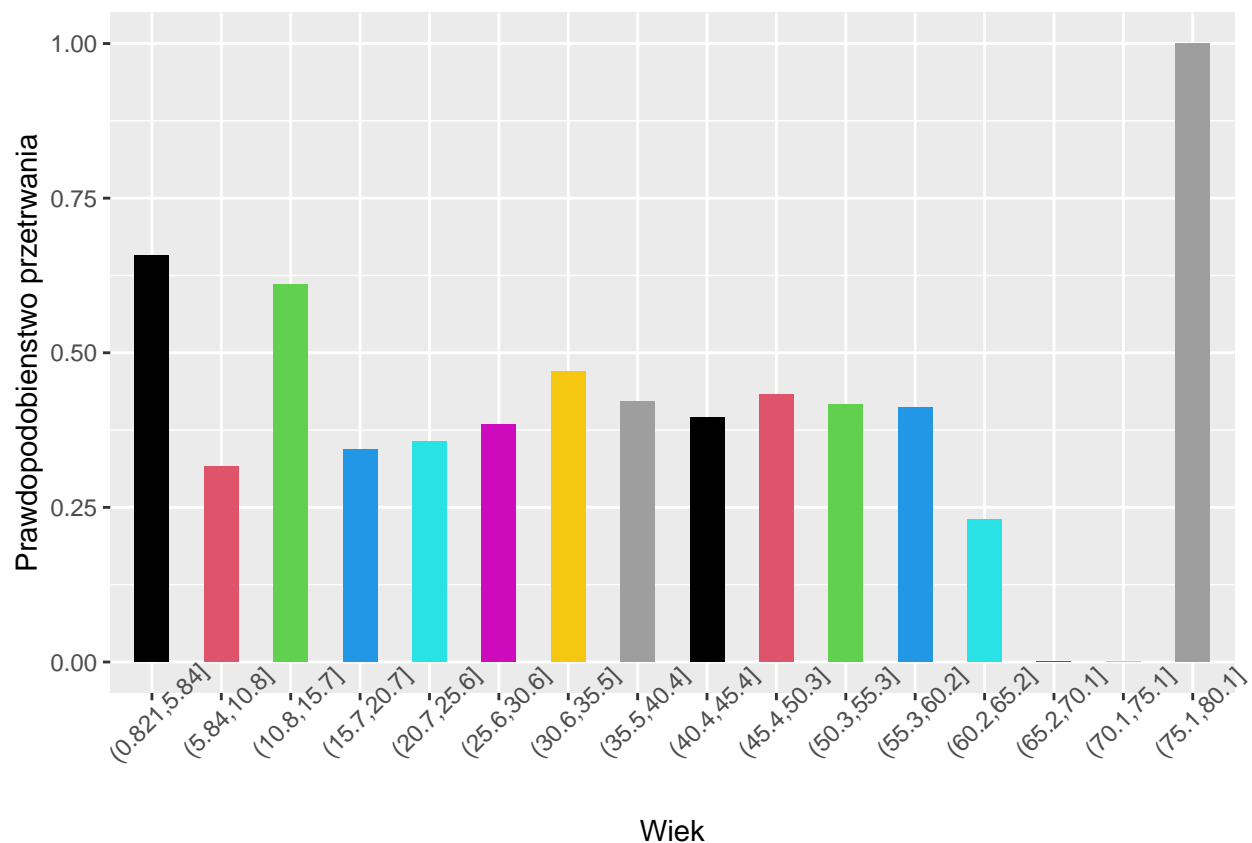
- 191 kobiet przeżyło katastrofę Titanica. Wszystkich kobiet było 252 czyli przeżyło 75.7936508%
- 85 mężczyzn przeżyło katastrofę Titanica. Wszystkich mężczyzn było 427 czyli przeżyło 19.9063232%
- szansa przeżycia dla kobiet była 3.8075163 razy większa niż mężczyzn.

Wiek

Wiek pasażera, a przeżywalność



Prawdopodobieństwo przetrwania według kategorii wiekowej



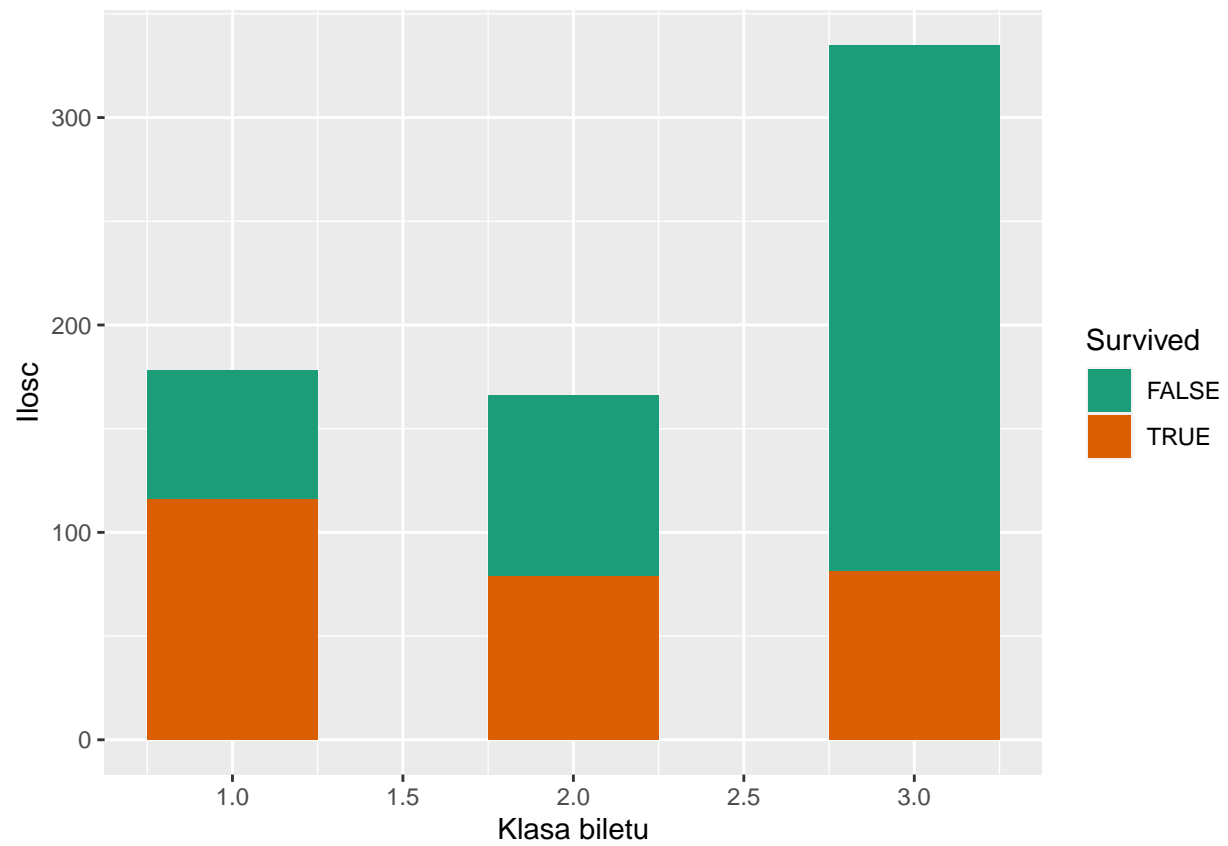
```
##      ageIntervals  survRate
## 1 (0.821,5.84] 0.6578947
## 2 (5.84,10.8] 0.3157895
## 3 (10.8,15.7] 0.6111111
## 4 (15.7,20.7] 0.3437500
## 5 (20.7,25.6] 0.3559322
## 6 (25.6,30.6] 0.3846154
## 7 (30.6,35.5] 0.4691358
## 8 (35.5,40.4] 0.4218750
## 9 (40.4,45.4] 0.3953488
## 10 (45.4,50.3] 0.4324324
## 11 (50.3,55.3] 0.4166667
## 12 (55.3,60.2] 0.4117647
## 13 (60.2,65.2] 0.2307692
## 14 (65.2,70.1] 0.0000000
## 15 (70.1,75.1] 0.0000000
## 16 (75.1,80.1] 1.0000000
```

Przeżywalność a wiek

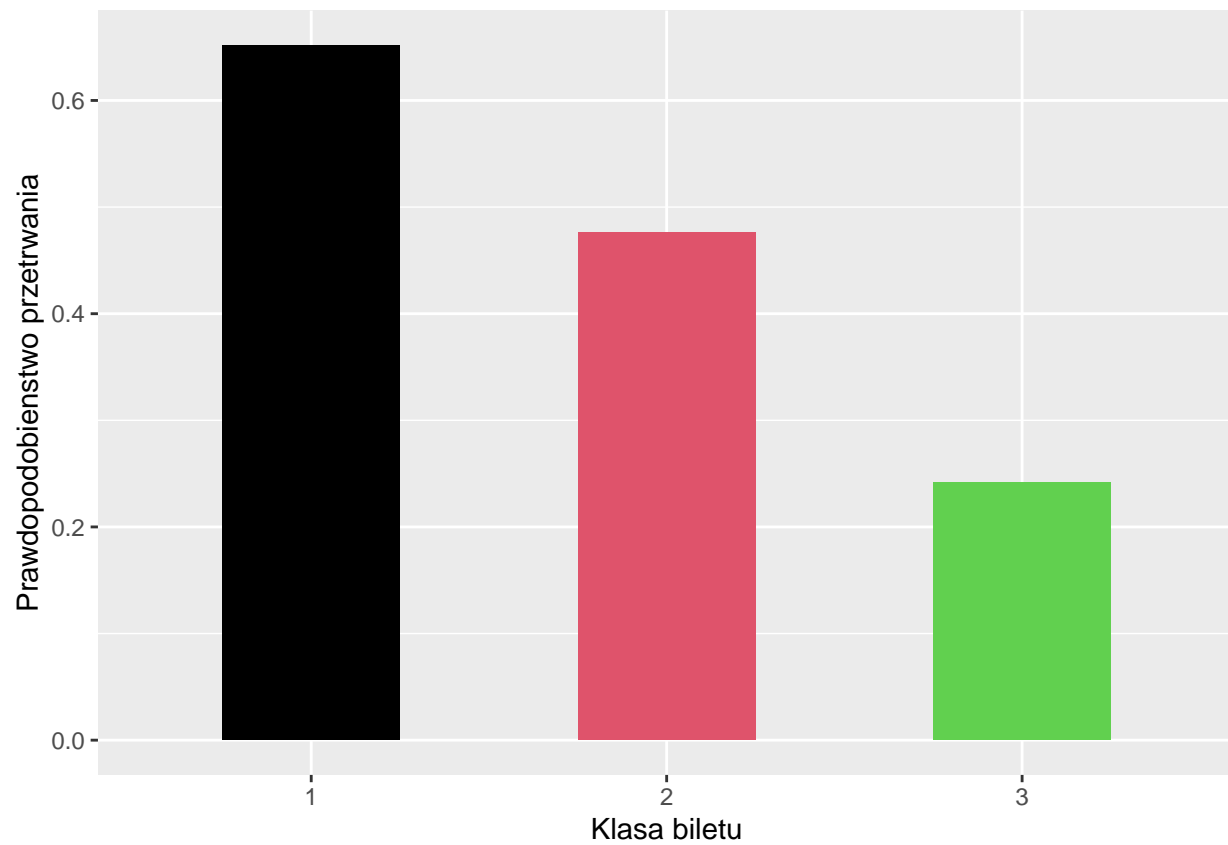
- osoba w wieku 80 lat miała największe prawdopodobieństwo przetrwania ale dane są przekłamane ponieważ był tylko jeden taki przypadek
- największe szanse przetrwania (nie licząc poprzedniego przypadku) mają osoby w najmłodsze
- wśród dzieci w przedziale wiekowym (5.84,10.8] z jakiegoś powodu przeżywalność była wyraźnie mniejsza. Mogło to mieć związek z przypadkami losowymi lub przekłamaniami danych z powodu

usunięcia niekompletnych przypadków ze zbioru.

Klasa biletu



Prawdopodobieństwo przetrwania według klasy biletu



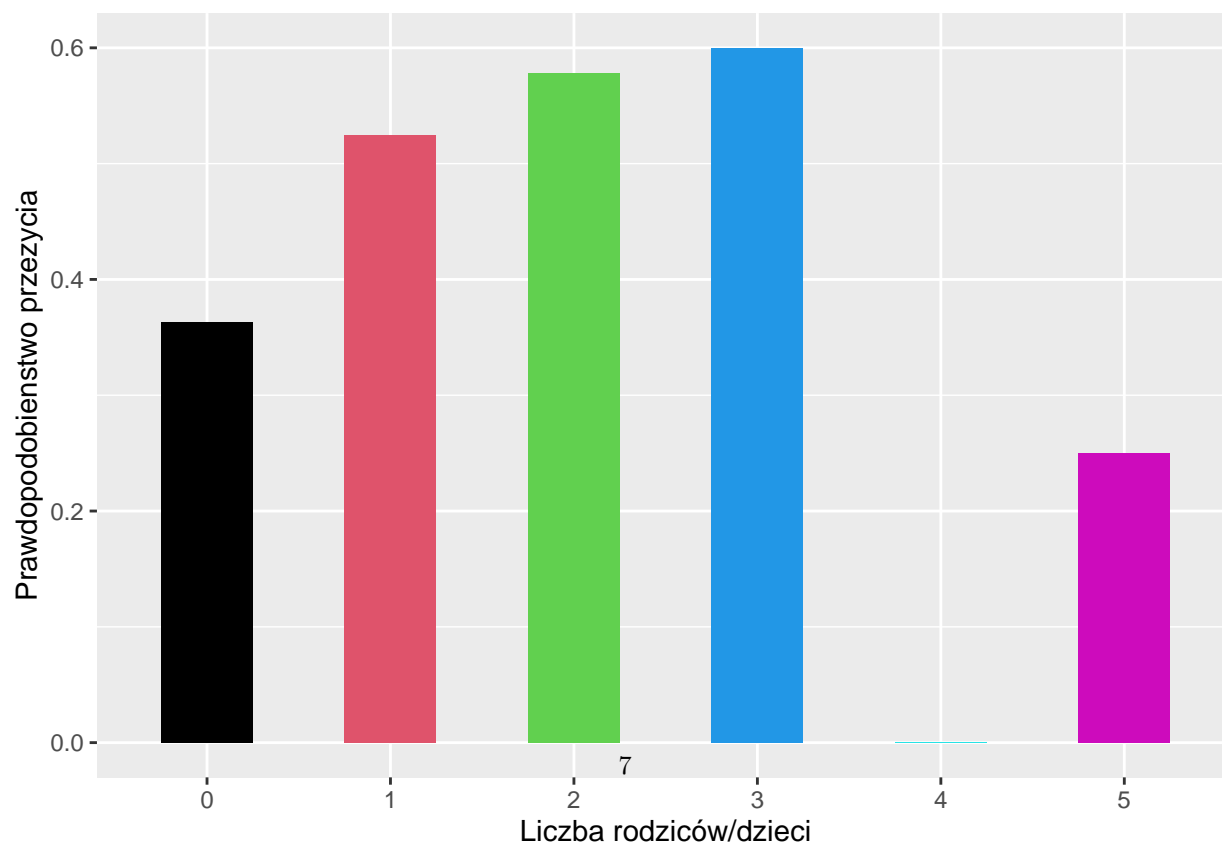
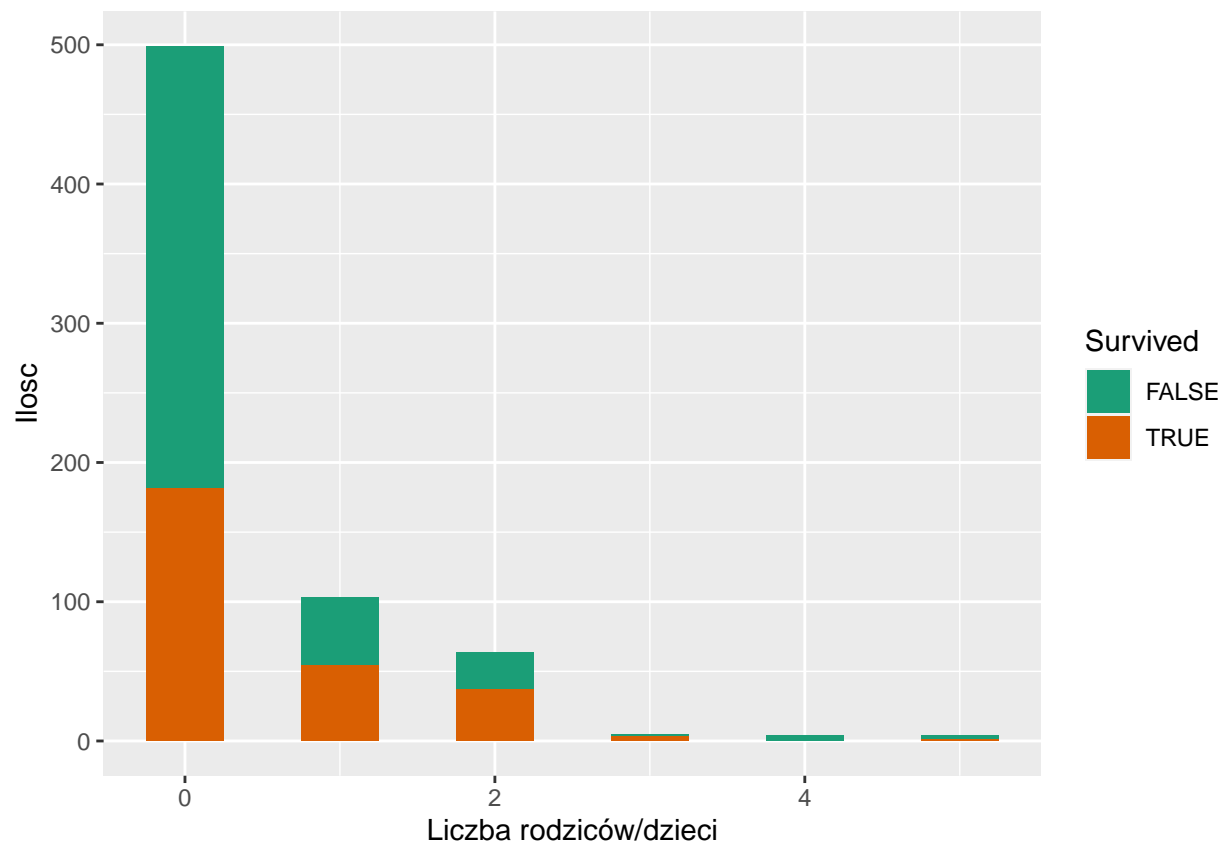
```
## classes survRate
## 1      1 0.6516854
## 2      2 0.4759036
## 3      3 0.2417910
```

Przeżywalność a klasa biletu

- największą szansę na przeżycie katastrofy miały osoby posiadające bilet w najwyższej (pierwszej) klasie

Liczba rodziców/dzieci oraz rodzeństwa/małżonków a prawdopodobieństwo przetrwania

Liczba rodziców/dzieci



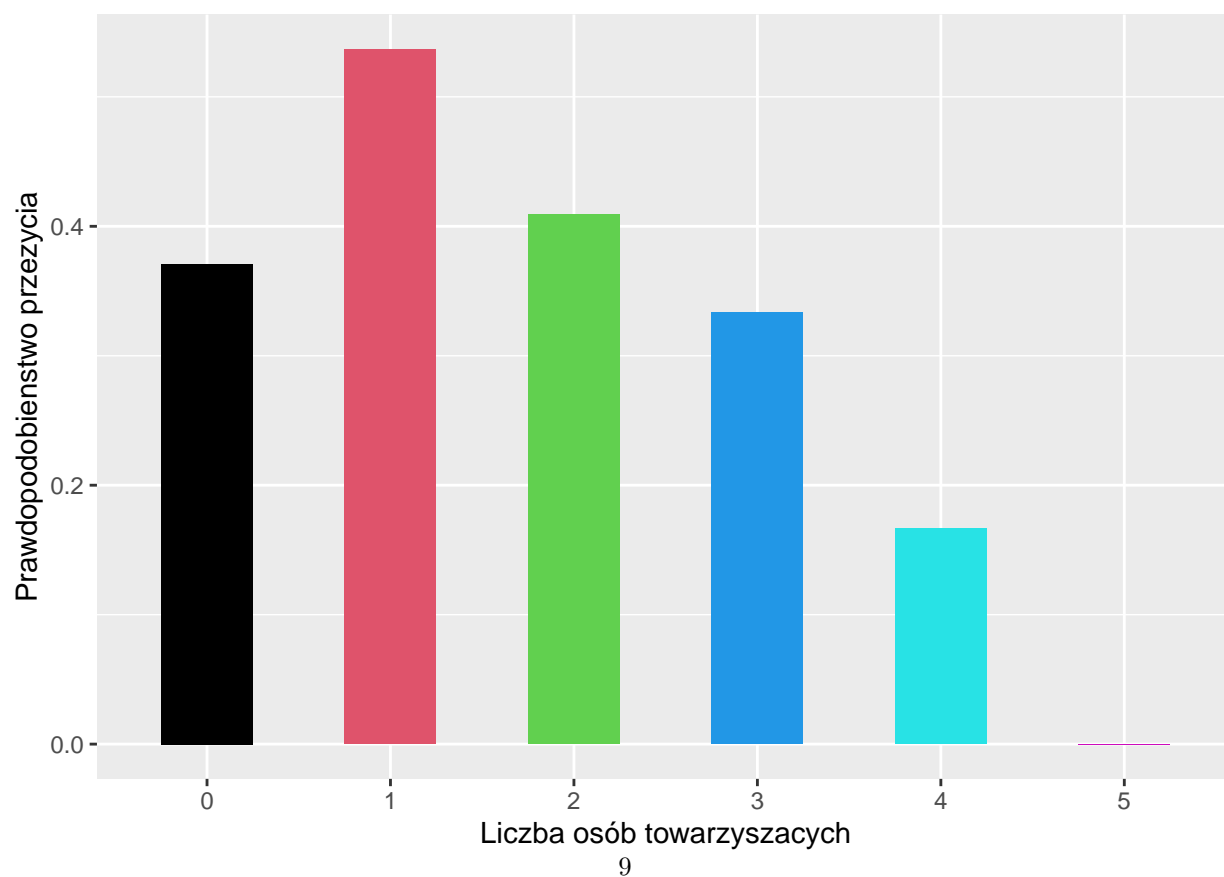
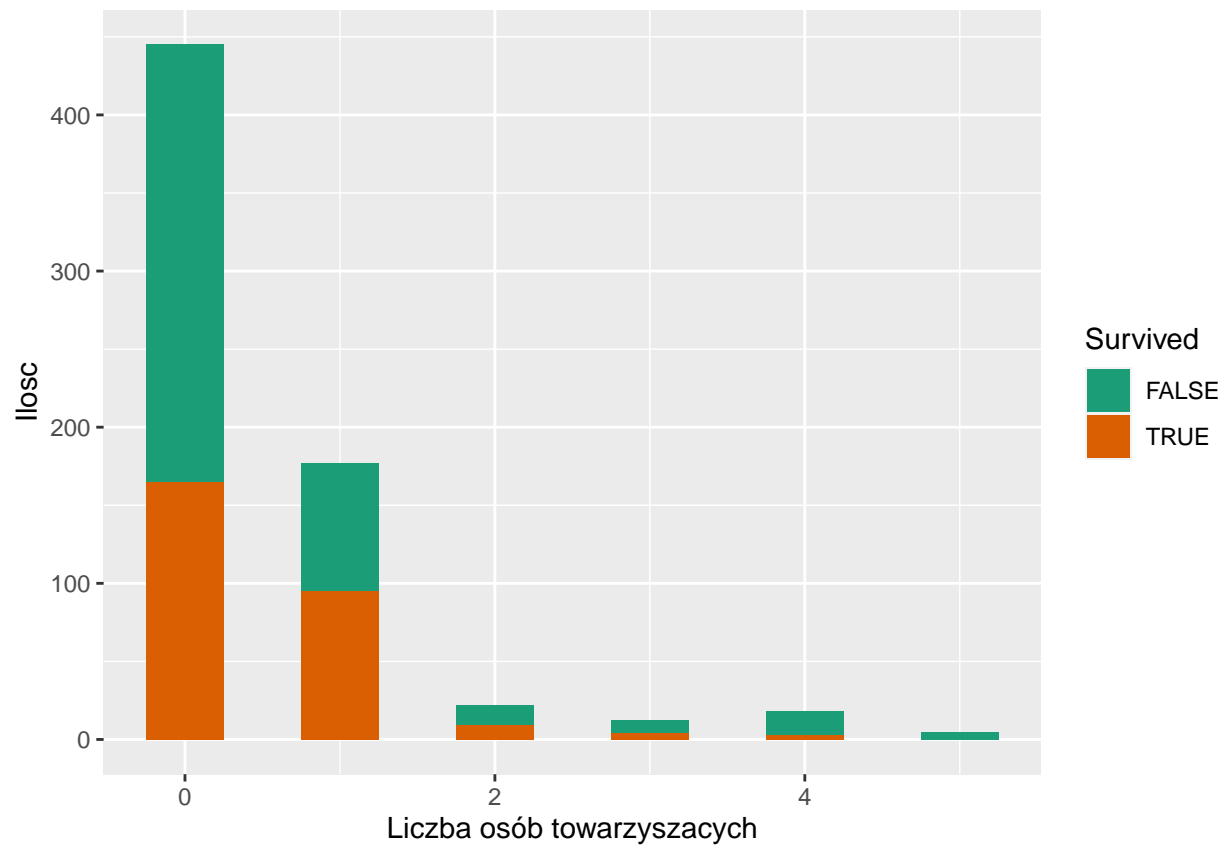
```
##   parch   survRate
## 1     0 0.3627255
## 2     1 0.5242718
## 3     2 0.5781250
## 4     3 0.6000000
## 5     4 0.0000000
## 6     5 0.2500000
```

Przeżywalność a rodzice/dzieci

```
##   PassengerId Survived Pclass                               Name    Sex Age SibSp Par
## 124          168   FALSE      3  Skoog, Mrs. William (Anna Bernhardina Karlsson) female  45    1
## 271          361   FALSE      3                               Skoog, Mr. Wilhelm   male  40    1
## 334          439   FALSE      1                               Fortune, Mr. Mark   male  64    1
## 428          568   FALSE      3  Palsson, Mrs. Nils (Alma Cornelia Berglund) female  29    0
##   Cabin Embarked
## 124              S
## 271              S
## 334 C23 C25 C27   S
## 428              S
```

- po podejrzeniu tych wpisów widać, że większość z tych osób miała wykupione bilety w najniższej (trzeciej) klasie co zmniejszało szansę na przeżycie

Liczba rodzeństwa/małżonków



```
##      sibsp  survRate
## 1         0 0.3707865
## 2         1 0.5367232
## 3         2 0.4090909
## 4         3 0.3333333
## 5         4 0.1666667
## 6         5 0.0000000
```

Przeżywalność a rodzeństwo/małżonkowie

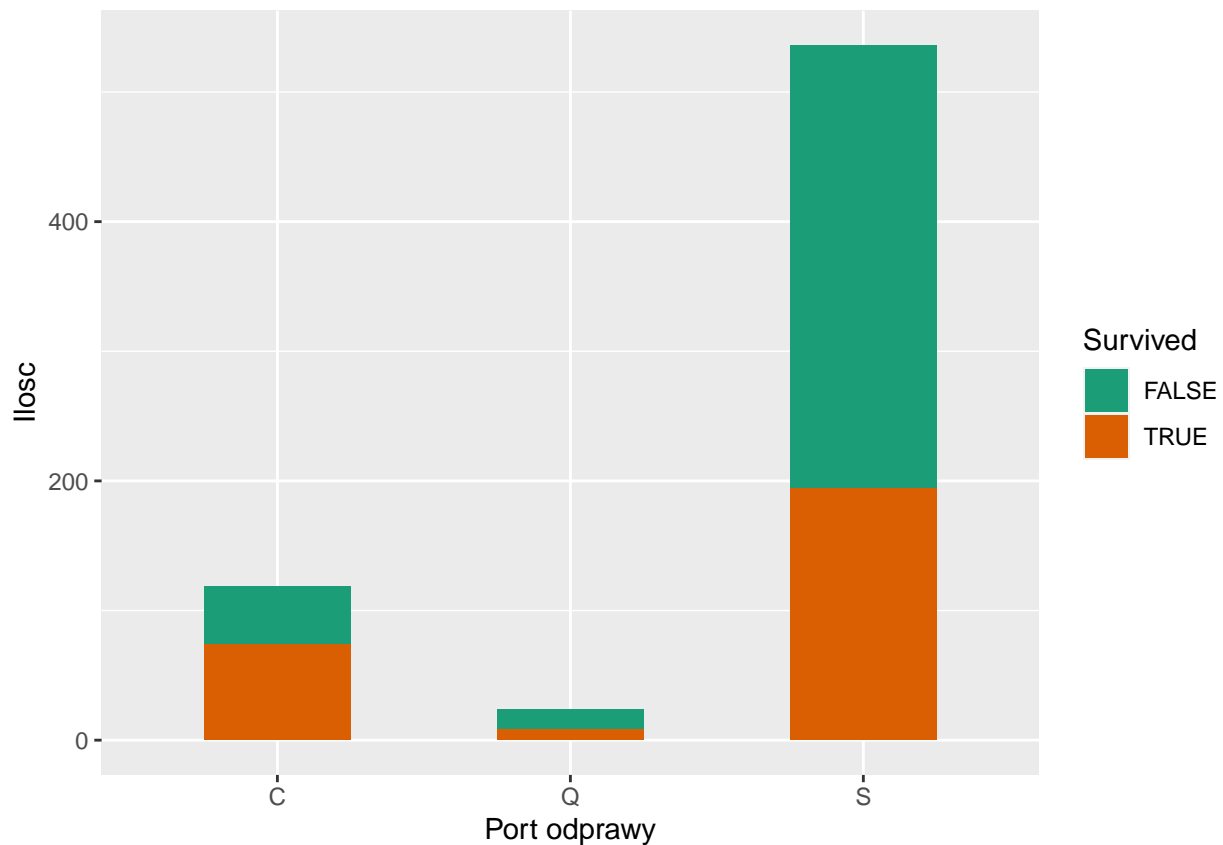
- największą szansę na przeżycie mieli ludzie którzy wybrali się z rodzeństwem (1 osoba) lub małżonkiem

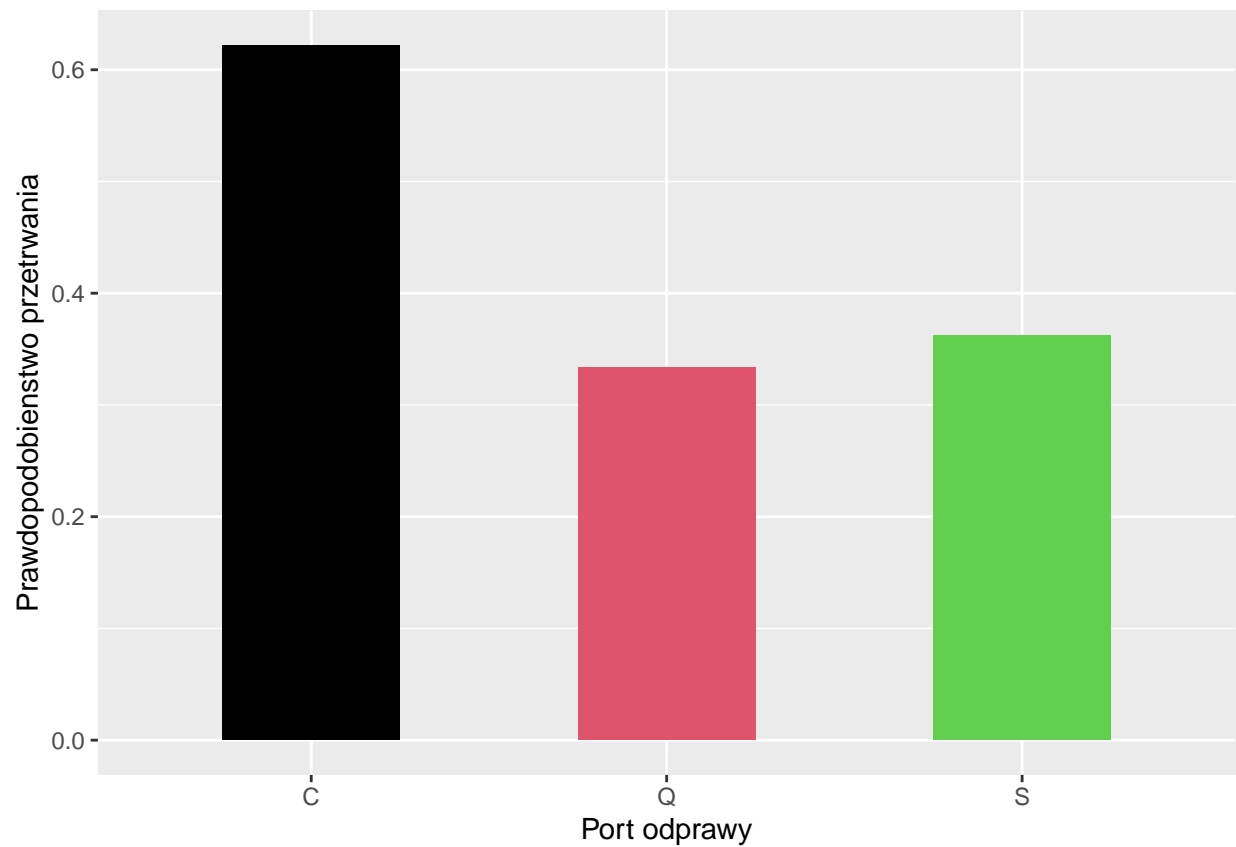
```
##      PassengerId Survived Pclass      Name Sex Age SibSp Parch Ticket Fa
## 43             60   FALSE      3 Goodwin, Master. William Frederick  male  11     5     2  CA  2144 46
## 52             72   FALSE      3      Goodwin, Miss. Lillian Amy  female  16     5     2  CA  2144 46
## 292            387   FALSE      3 Goodwin, Master. Sidney Leonard  male    1     5     2  CA  2144 46
## 365            481   FALSE      3      Goodwin, Master. Harold Victor  male    9     5     2  CA  2144 46
## 514            684   FALSE      3      Goodwin, Mr. Charles Edward  male   14     5     2  CA  2144 46
```

- wszystkie te osoby miały zakupiony 1 bilet w trzeciej klasie i byli w większości mężczyznami co zmniejszało szansę na ich przeżycie

Port odprawy

Ostatnim elementem zależnym jest port odprawy.

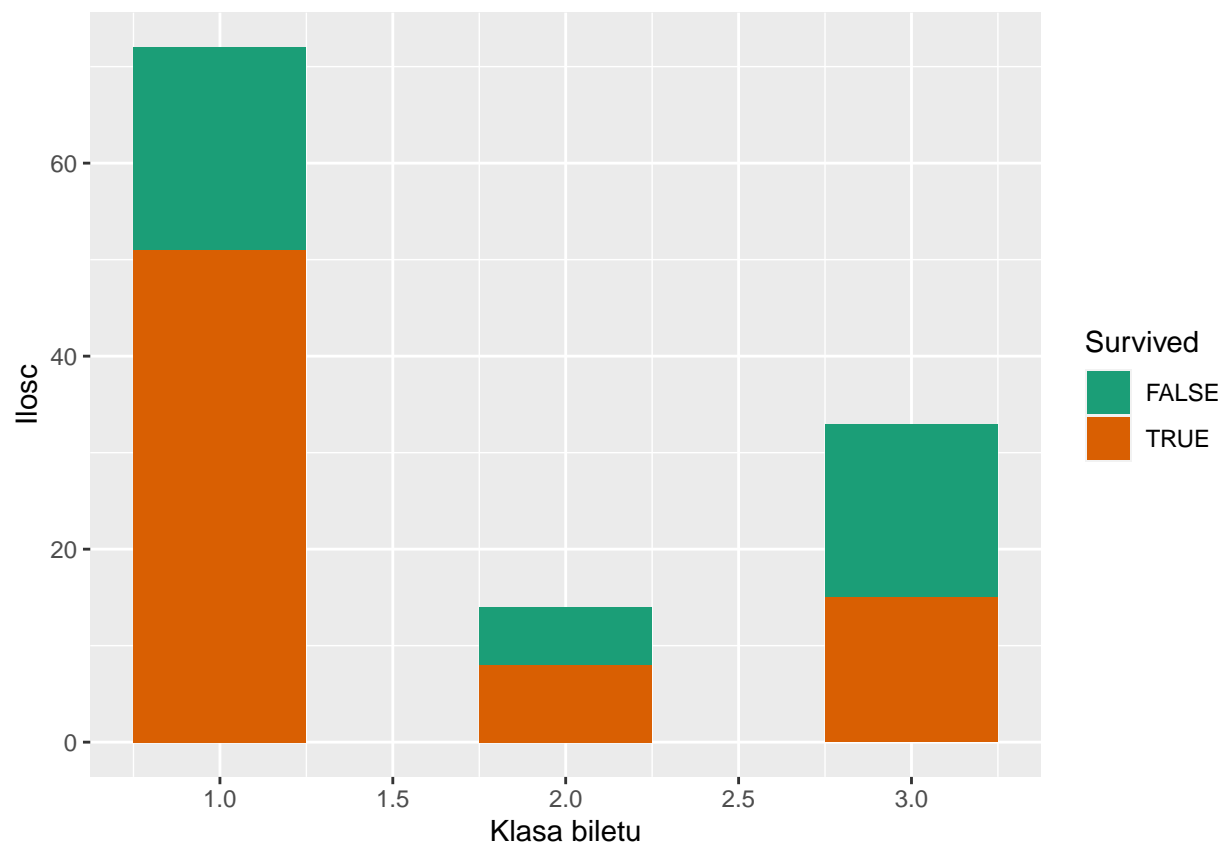




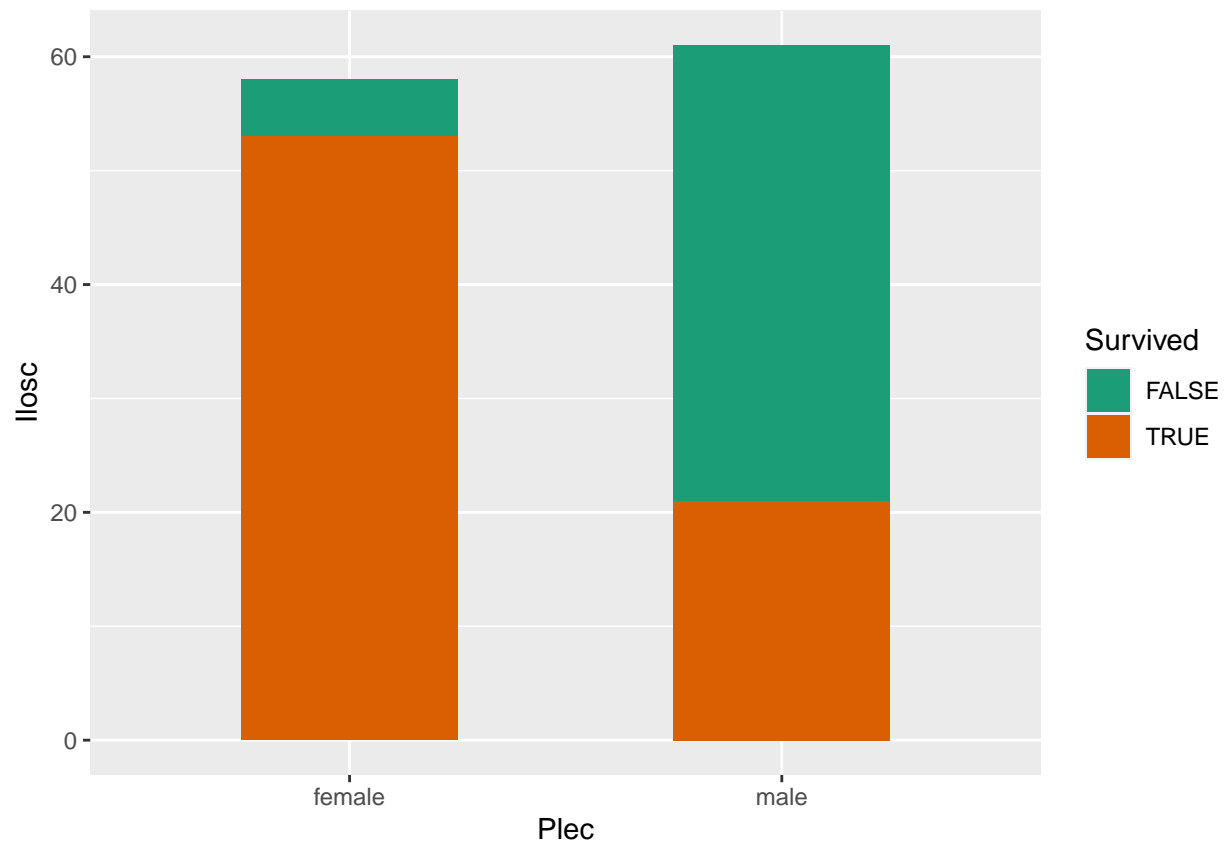
```
## embarked survRate
## 1      C 0.6218487
## 2      Q 0.3333333
## 3      S 0.3619403
```

Przeżywalność a port odprawy

- na wykresie widać, że mimo iż mało osób odpływało z portu w Cherbourg to miały one największe prawdopodobieństwo przetrwania



- widać dlaczego największe prawdopodobieństwo przetrwania było przy wypływaniu z tego portu, ale można sprawdzić rozkład płci odpływających z tego miejsca



Podsumowanie

- analiza wskazuje na to, że głównymi czynnikami dla przetrwania były płeć i klasa biletu.
- następnym czynnikiem był wiek ,a reszta czynników była skorelowana
- największą współczynnik przeżycia katastrofy Titanica miały kobiety i dzieci pływące pierwszą klasą