

Szanse i czynniki przeżycia katastrofy Titanica

Paweł Lurka

18/01/2022

Przedstawienie problemu

W niniejszym raporcie przedstawiono szansę jaką na przeżycie mieli pasażerowie statku pasażerskiego Titanic oraz od jakich czynników ona zależała. Spośród wszystkich atrybutów zbioru Titanic oczyszczonego z brakujących i niespójnych danych w poprzednim zadaniu wybrano te, które mogły mieć decydujący wpływ na przeżywalność pasażerów.

```
## PassengerId    Survived    Pclass    Name
## Min.      : 1.0    Mode :logical Min.      :1.000    Length:679
## 1st Qu.   : 226.5  FALSE:403    1st Qu.  :1.000    Class :character
## Median    : 446.0  TRUE :276    Median   :2.000    Mode  :character
## Mean      : 450.4                      Mean     :2.231
## 3rd Qu.   : 675.0                      3rd Qu.  :3.000
## Max.      :1000.0                      Max.     :3.000
## Sex                Age                SibSp                Parch
## Length:679        Min.      : 0.90    Min.      :0.0000    Min.      :0.0000
## Class :character  1st Qu.:20.00    1st Qu.:0.0000    1st Qu.:0.0000
## Mode :character  Median :28.00    Median :0.0000    Median :0.0000
##                      Mean      :29.72    Mean      :0.5214    Mean      :0.4153
##                      3rd Qu.:38.00    3rd Qu.:1.0000    3rd Qu.:1.0000
##                      Max.      :80.00    Max.      :5.0000    Max.      :5.0000
## Ticket            Fare                Cabin                Embarked
## Length:679        Min.      : 0.00    Length:679        Length:679
## Class :character  1st Qu.: 8.05    Class :character  Class :character
## Mode :character  Median :15.74    Mode :character  Mode :character
##                      Mean      :33.14
##                      3rd Qu.:33.76
##                      Max.      :263.00
```

Wybrane atrybuty:

1. Płeć (Sex)

- w sytuacjach kryzysowych najpierw ratowane są kobiety i dzieci, dopiero później ratuje się mężczyzn więc płeć może mieć decydujące znaczenie w kwestii przetrwania

2. Wiek (Age)

- jak wyżej, większe prawdopodobieństwo przetrwania najmłodszych

3. Klasa biletu (Pclass)

- zasobność portfela i poziom obsługi zależny od klasy zakupionego biletu też może mieć spore znaczenie w wypadku katastrofy, przypuszczalnie, procentowo większy udział uratowanych będzie mieć klasa 1 biletów (najwyższa) niż klasa 3cia (najniższa)

4. Liczba rodziców/dzieci (Parch)

- jeśli pierwsze są ratowane kobiety i dzieci, być może osoby z dziećmi/rodzicami będą miały większy współczynnik przeżywalności niż te które dzieci nie mają

5. Liczba rodzeństwa/małżonków (SibSp)

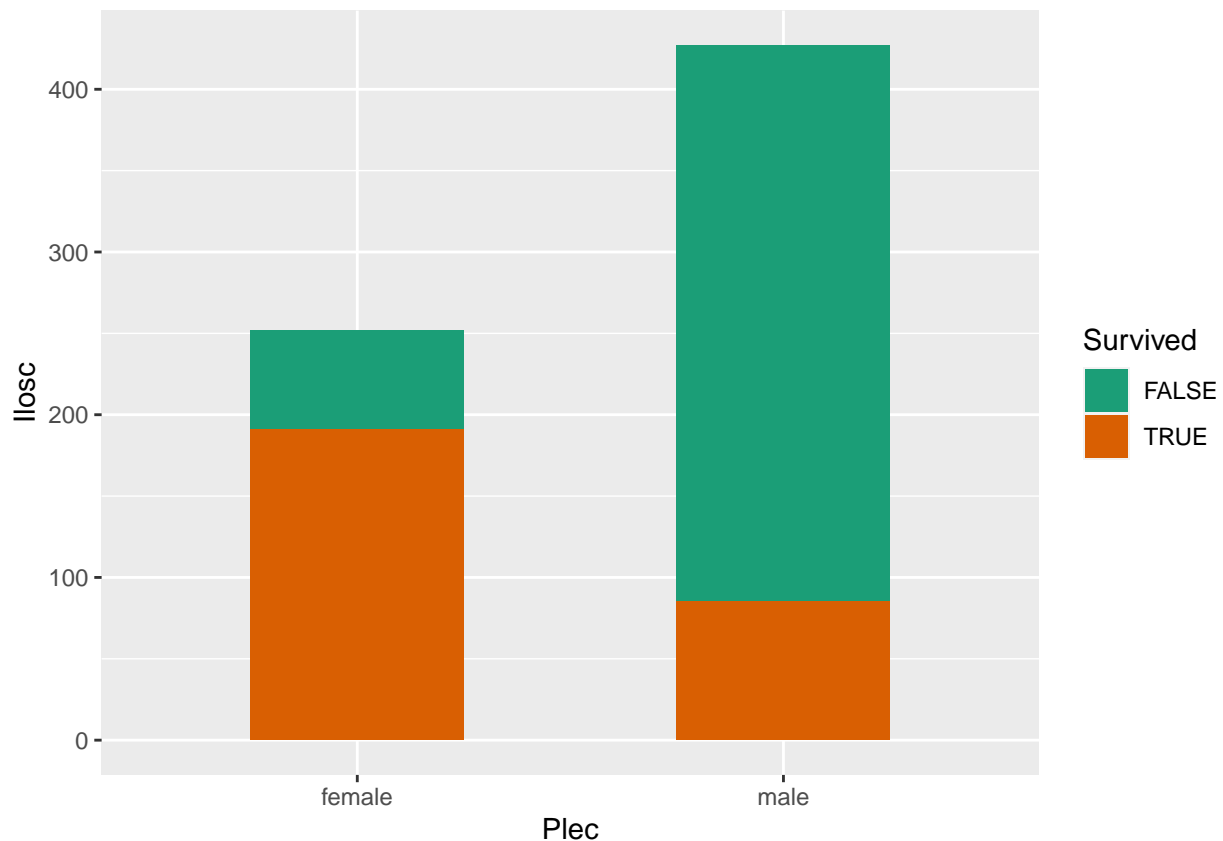
- na pierwszy rzut oka nie powinno mieć to większego wpływu na przetrwanie ale warto sprawdzić czy nie istnieje jakaś korelacja między tym atrybutem a przeżywalnością, a jeśli tak to z czego wynika (czy np. osoby z większą ilością rodzeństwa, małżonków wybierały wyższą klasę biletu)

6. Port odprawy (Embarked)

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

Płeć

Na początek sprawdzimy zależność przetrwania od płci pasażera.

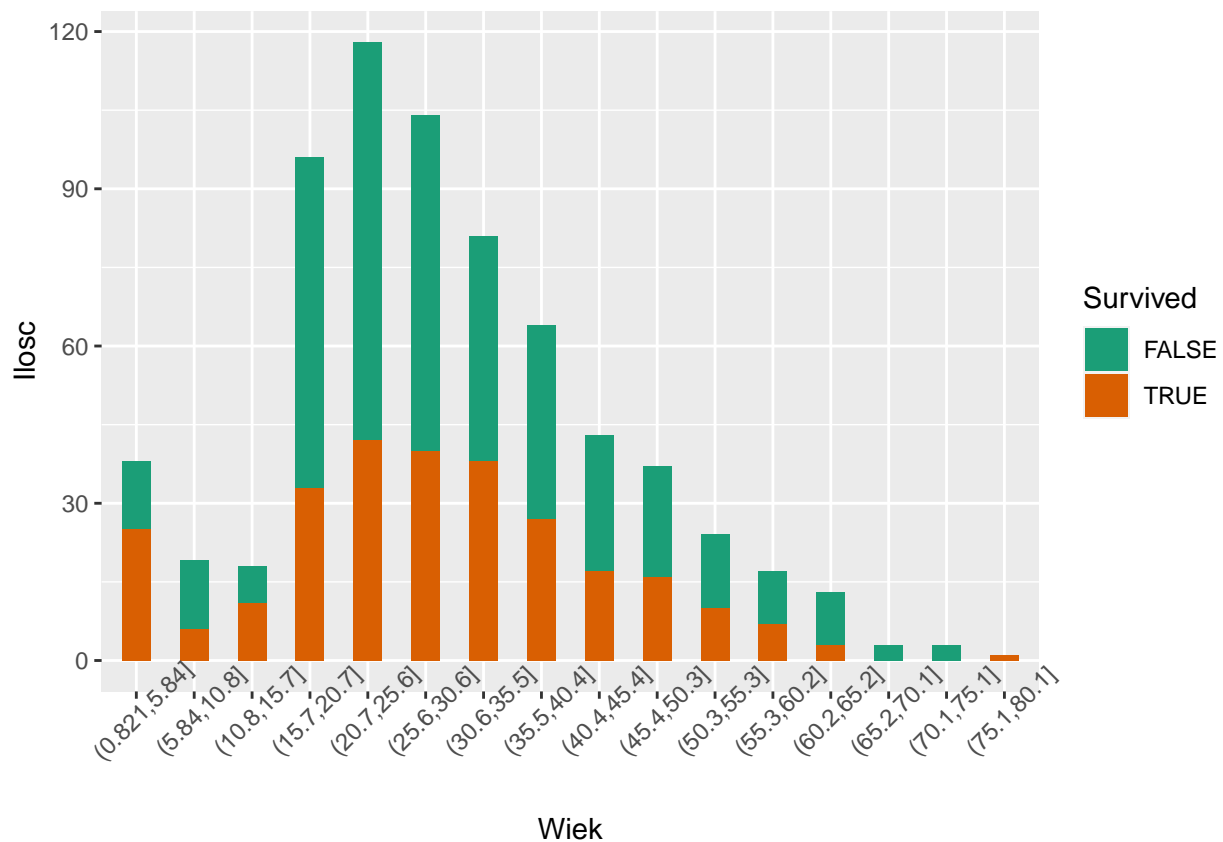


Przeżywalność a płeć

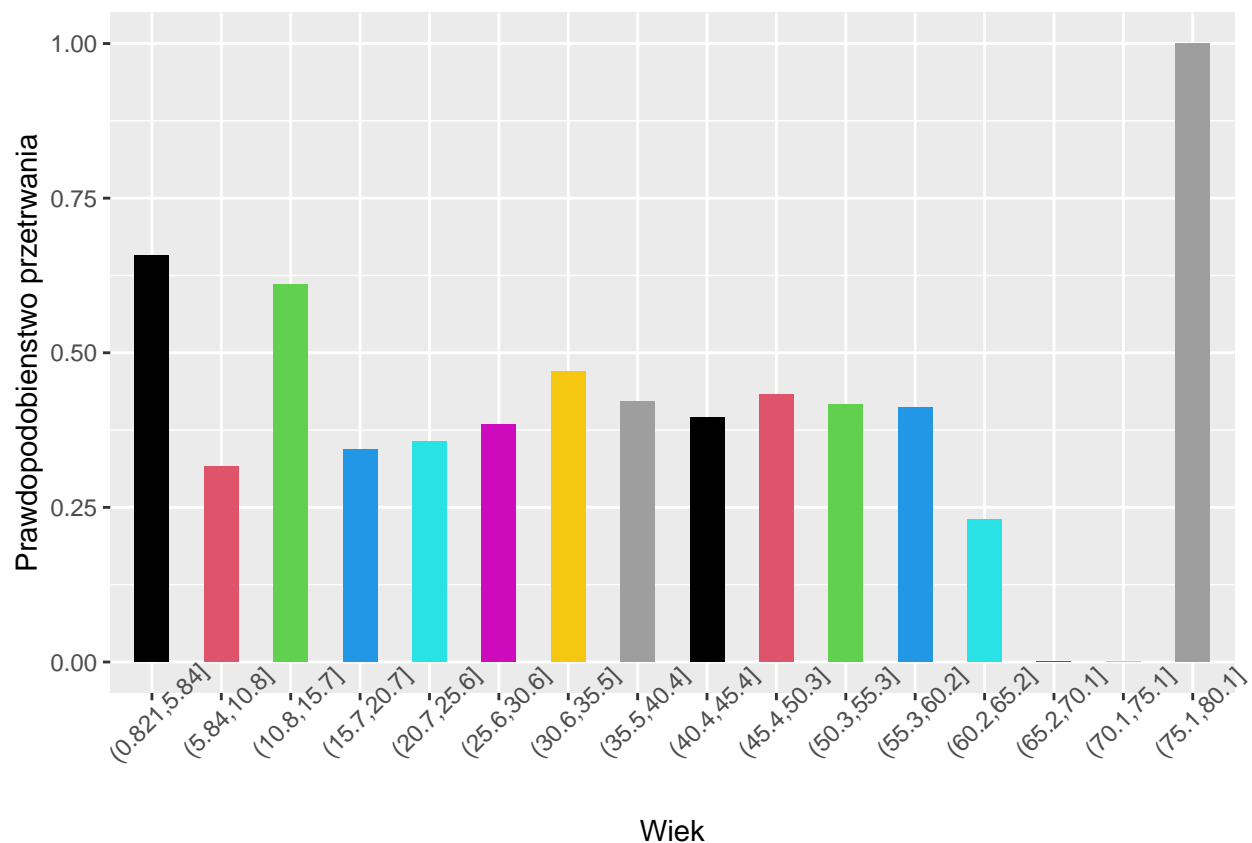
- Liczba kobiet które przeżyły katastrofę to 191. Wszystkich kobiet było na statku 252 co oznacza, że przeżyło 75.7936508% wszystkich kobiet.
- Liczba mężczyzn którzy przeżyli katastrofę to 85. Wszystkich mężczyzn było na statku 427 co oznacza, że przeżyło 19.9063232% wszystkich mężczyzn.
- Jak widać, początkowe przypuszczenia potwierdziły się ponieważ szansa przeżycia dla kobiet była 3.8075163 razy większa niż dla mężczyzn.

Wiek

Kolejną rzeczą którą można sprawdzić jest zależność przetrwania od wieku pasażera



Spójrzmy na prawdopodobieństwo przetrwania dla każdej kategorii wiekowej



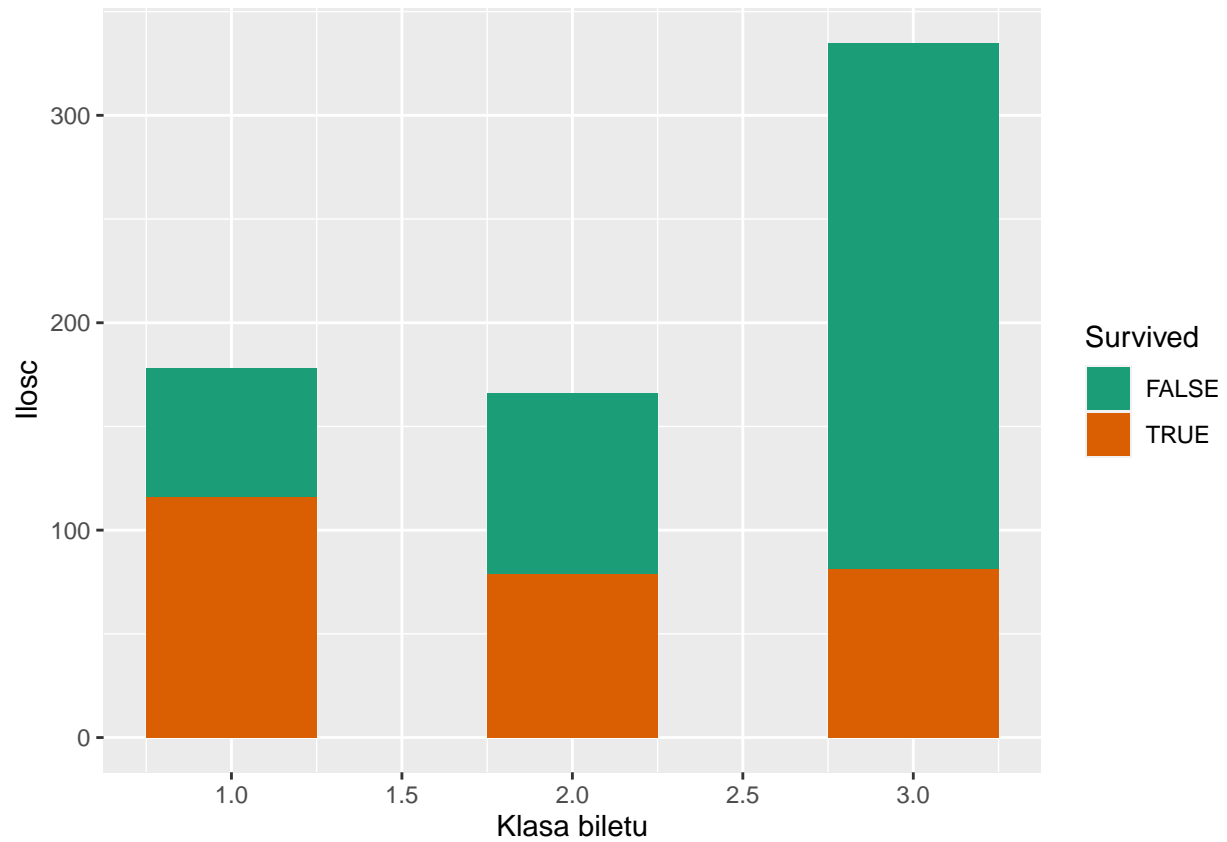
```
##      ageIntervals  survRate
## 1 (0.821,5.84] 0.6578947
## 2 (5.84,10.8] 0.3157895
## 3 (10.8,15.7] 0.6111111
## 4 (15.7,20.7] 0.3437500
## 5 (20.7,25.6] 0.3559322
## 6 (25.6,30.6] 0.3846154
## 7 (30.6,35.5] 0.4691358
## 8 (35.5,40.4] 0.4218750
## 9 (40.4,45.4] 0.3953488
## 10 (45.4,50.3] 0.4324324
## 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

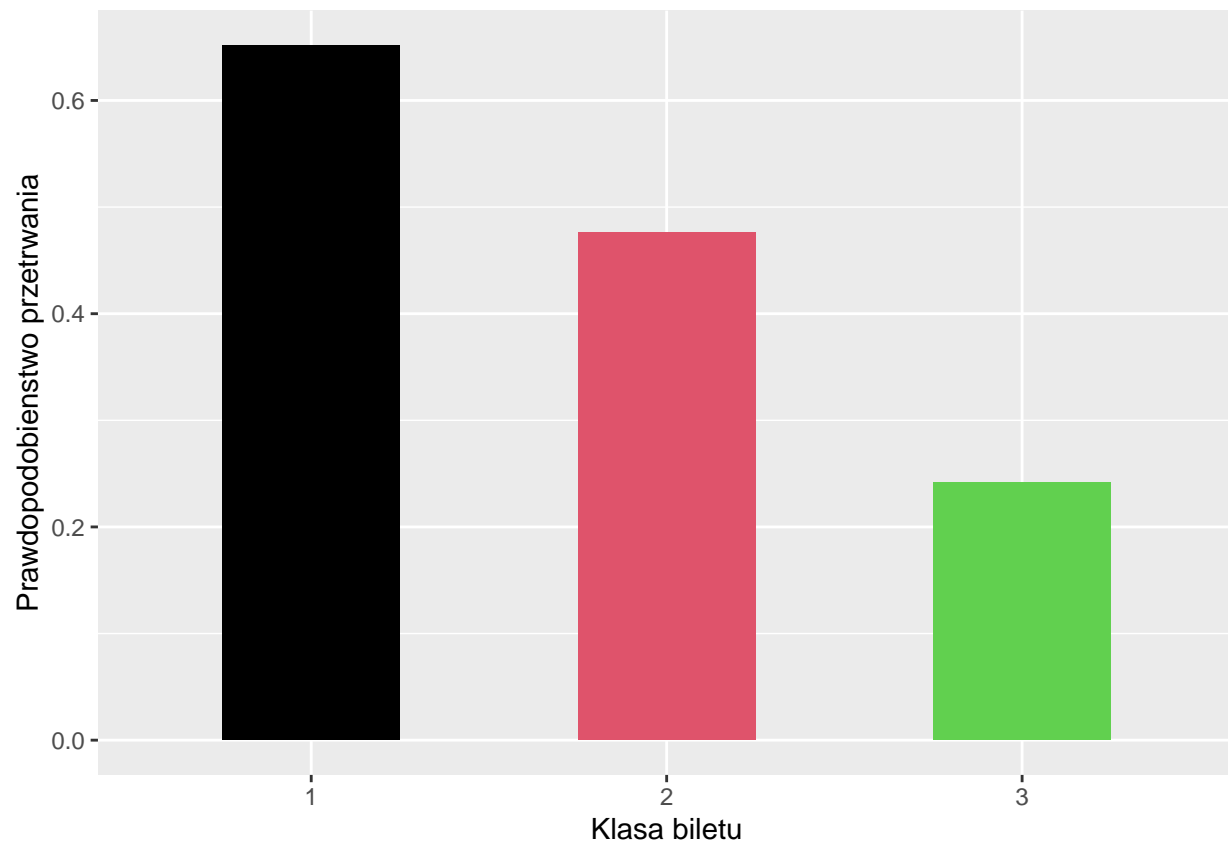
- Widać tutaj że osoba w wieku 80 lat miała największe prawdopodobieństwo przetrwania ale dane są przekłamanie ponieważ był to jeden odosobniony przypadek.
- Poza nim największe szanse przetrwania mają osoby w najmłodszym wieku czyli zgodnie z oczekiwaniami.

- Wśród dzieci w przedziale wiekowym $(5.84, 10.8]$ z jakiegoś powodu przeżywalność była wyraźnie mniejsza. Mogło to mieć związek z przypadkami losowymi lub przekłamaniem danych z powodu usunięcia niekompletnych przypadków ze zbioru.

Klasa biletu



Spójrzmy na prawdopodobieństwo przetrwania dla każdej klasy biletu



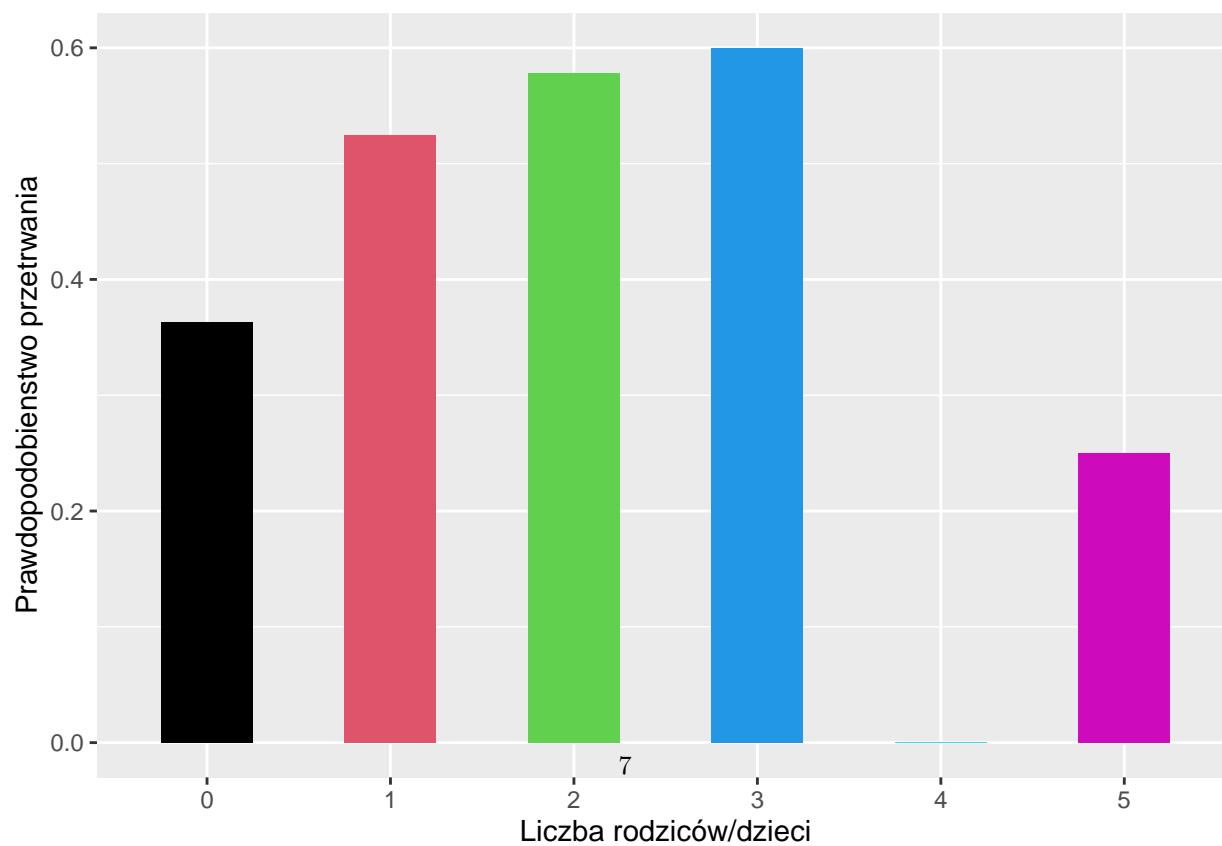
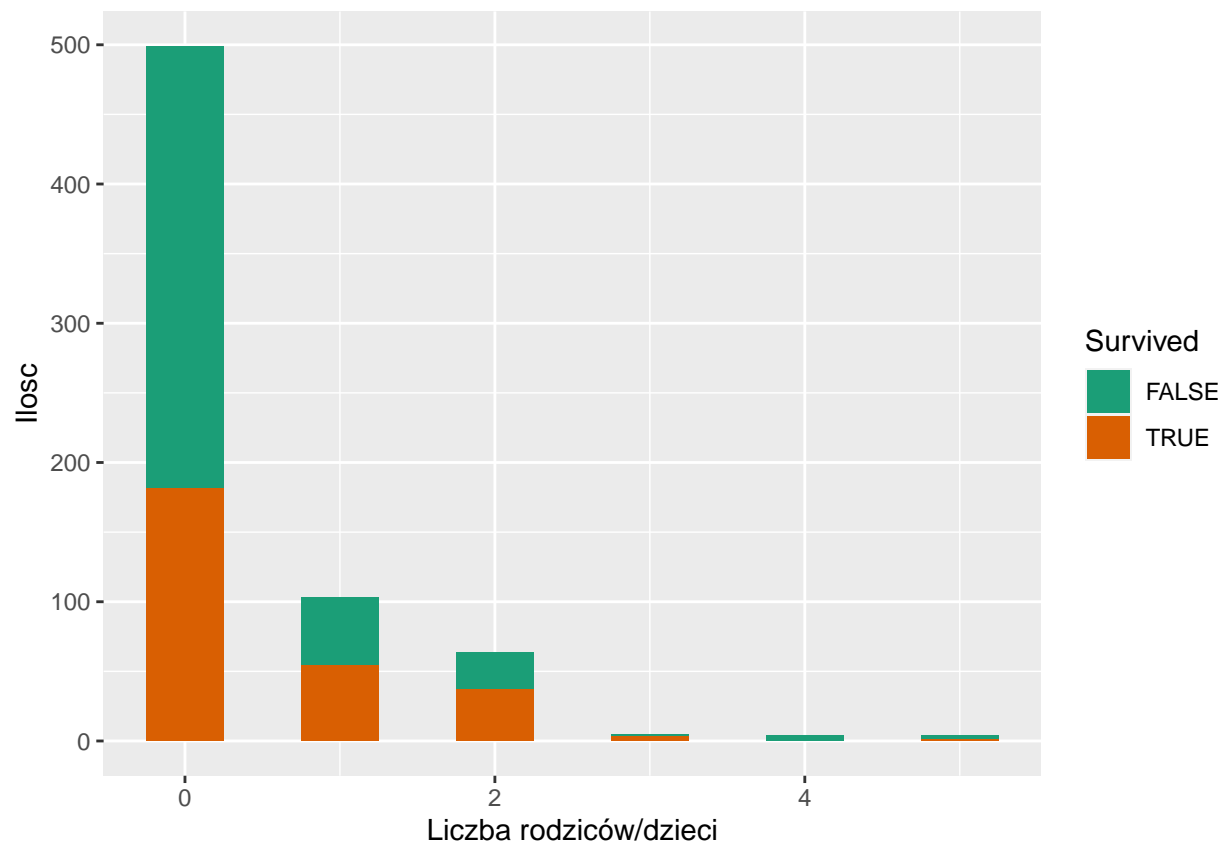
```
## classes survRate
## 1      1 0.6516854
## 2      2 0.4759036
## 3      3 0.2417910
```

Przeżywalność a klasa biletu

- zgodnie z oczekiwaniami wykres pokazuje, że największą szansę na przeżycie katastrofy miały osoby posiadające bilet w najwyższej 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
```

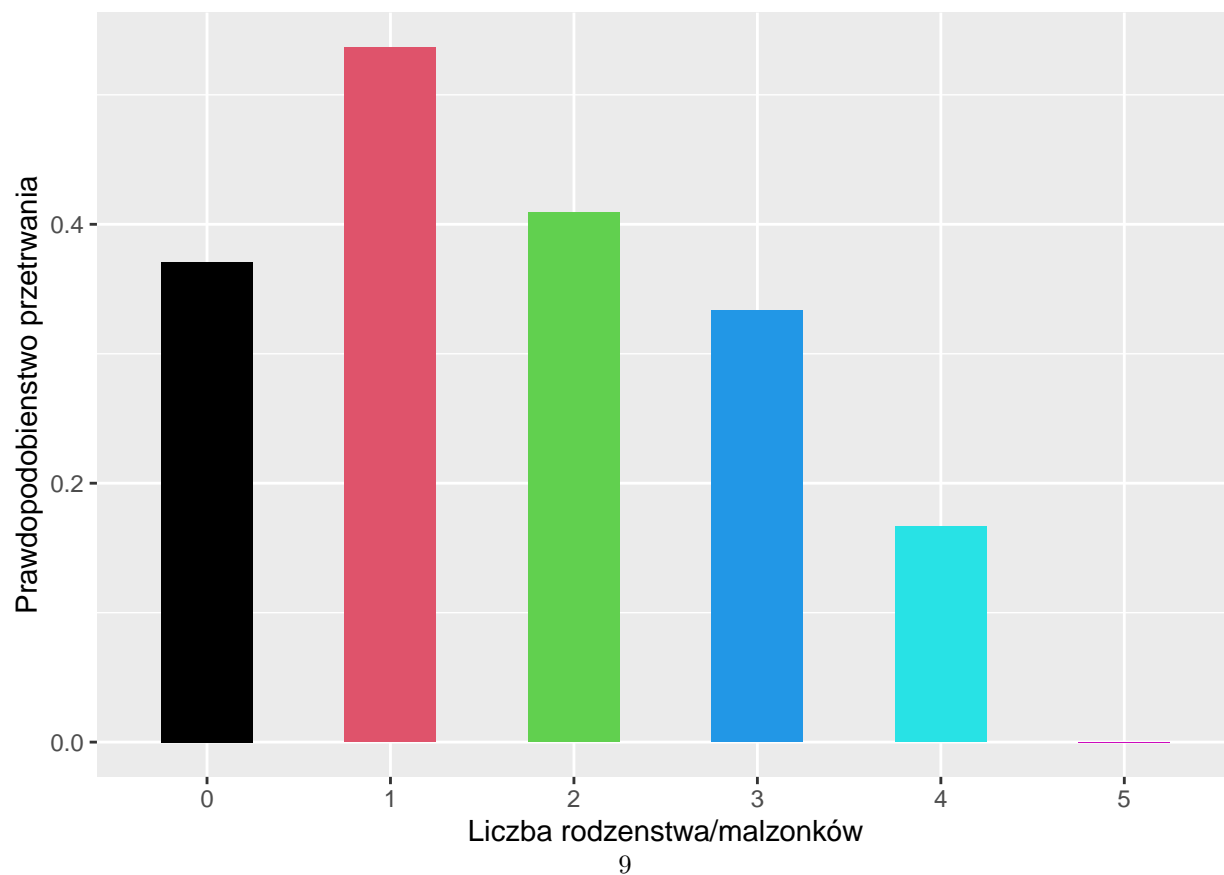
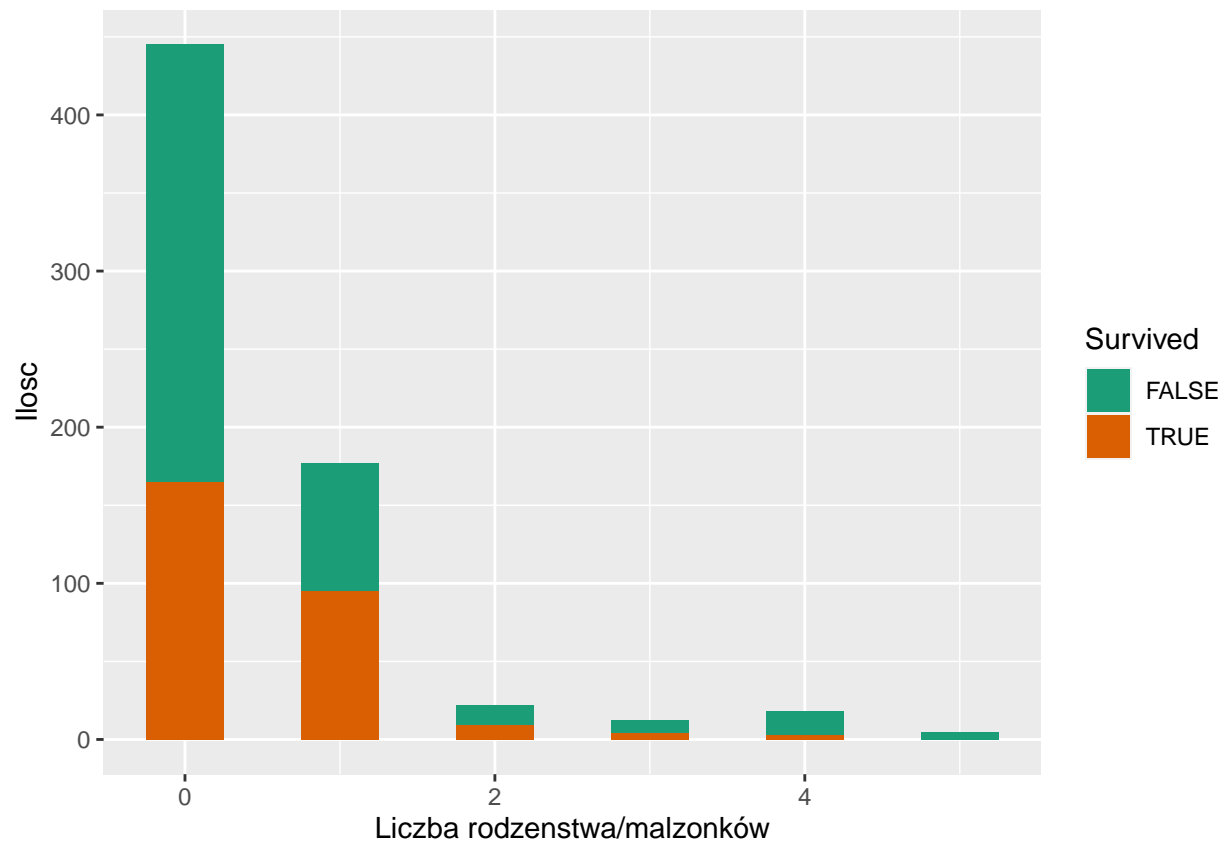
Przeżywalność a rodzice/dzieci

- Jak widać przypuszczenia odnośnie większej szansy przetrwania przy większej liczbie rodziców/dzieci sprawdziły się tylko do liczby 3. Potem przeżywalność znowu maleje.
- Jak widać przeżywalność dla 4 sztuk dzieci/rodziców była zerowa ale może mieć to związek z innymi czynnikami.

```
##   PassengerId Survived Pclass                               Name
## 124         168   FALSE     3  Skoog, Mrs. William (Anna Bernhardina Karlsson)
## 271         361   FALSE     3                               Skoog, Mr. Wilhelm
## 334         439   FALSE     1                               Fortune, Mr. Mark
## 428         568   FALSE     3  Palsson, Mrs. Nils (Alma Cornelia Berglund)
##      Sex Age SibSp Parch Ticket   Fare      Cabin Embarked
## 124 female  45     1     4 347088  27.900              S
## 271  male  40     1     4 347088  27.900              S
## 334  male  64     1     4 19950 263.000 C23 C25 C27      S
## 428 female  29     0     4 349909  21.075              S
```

- po podejrzeniu tych wpisów widać, że większość z tych osób miała wykupione biletyw 3ciej klasie co już samo drastycznie zmniejszało szansę na przetrwanie

Liczba rodzeństwa/malżonków



```
##      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

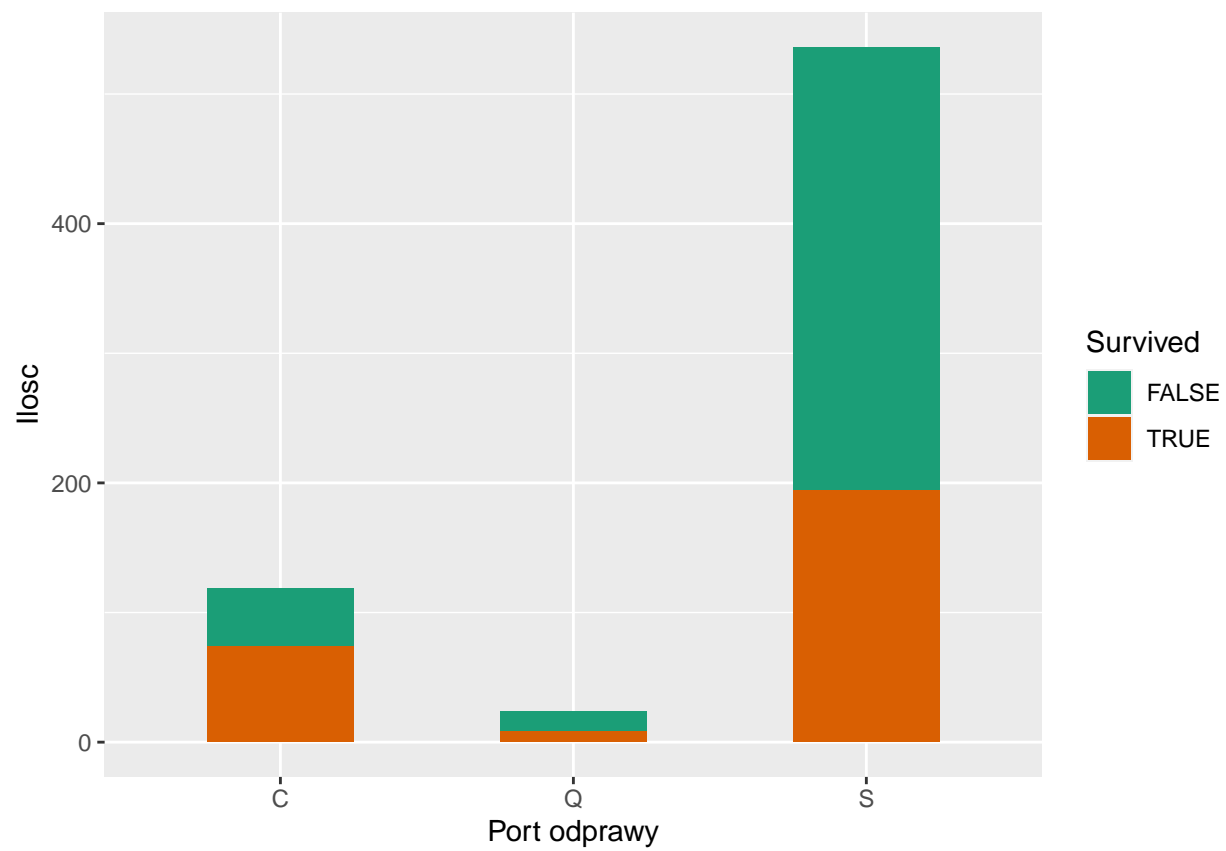
- widać, że największą szansę na przeżycie mieli ludzie którzy wybrali się z jednym bratem/siostrą lub małżonkiem
- można sprawdzić czym się charakteryzowały pozycje posiadające 5 spokrewnionych osób

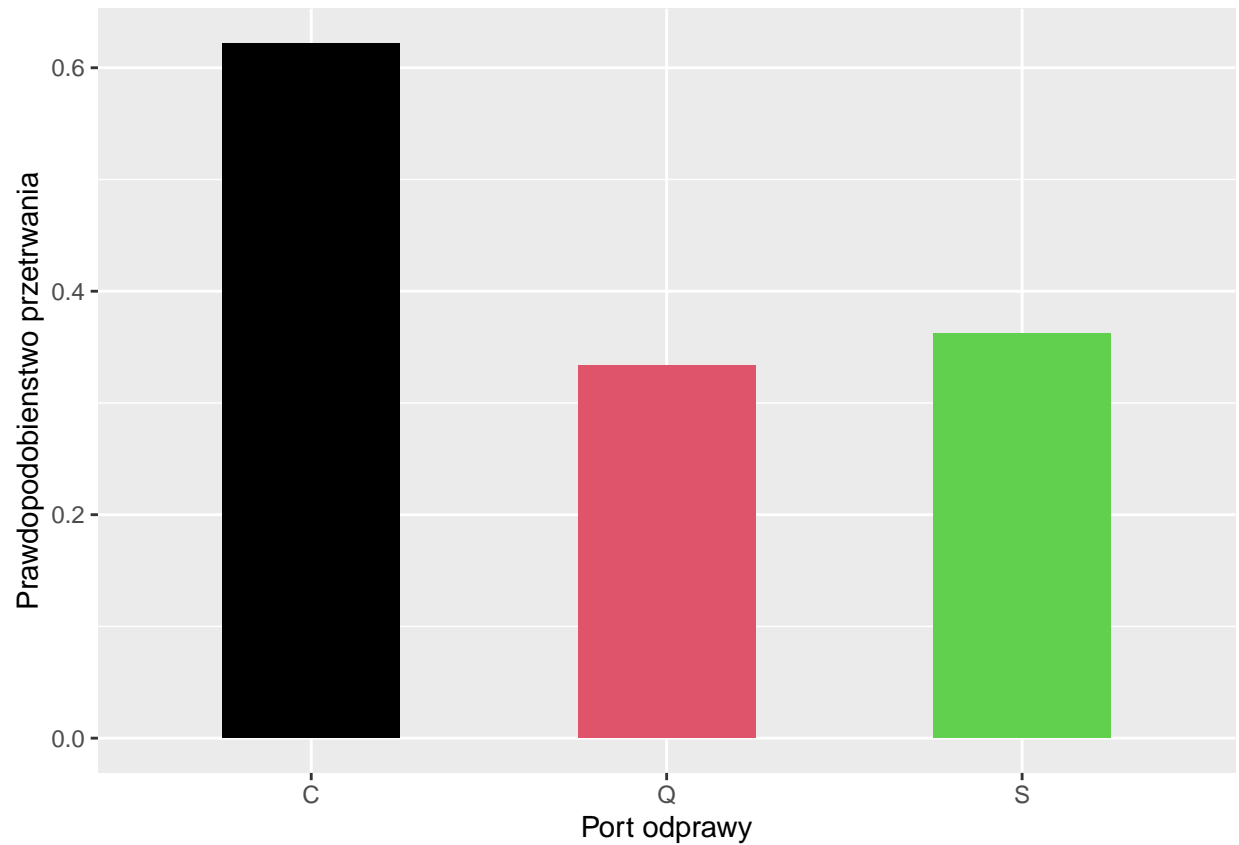
```
##      PassengerId Survived Pclass                      Name      Sex Age
## 43             60    FALSE      3 Goodwin, Master. William Frederick  male  11
## 52             72    FALSE      3      Goodwin, Miss. Lillian Amy female  16
## 292            387    FALSE      3   Goodwin, Master. Sidney Leonard  male   1
## 365            481    FALSE      3   Goodwin, Master. Harold Victor   male   9
## 514            684    FALSE      3   Goodwin, Mr. Charles Edward    male  14
##      SibSp Parch  Ticket Fare Cabin Embarked
## 43      5      2 CA  2144 46.9          S
## 52      5      2 CA  2144 46.9          S
## 292     5      2 CA  2144 46.9          S
## 365     5      2 CA  2144 46.9          S
## 514     5      2 CA  2144 46.9          S
```

- widać, że wszystkie te osoby miały zakupiony 1 bilet w 3ciej klasie i byli to w większości mężczyźni więc to mogło spowodować ich niską przeżywalność

Port odprawy

Ostatnim elementem który może pokazać ciekawe wyniki jest port odprawy.

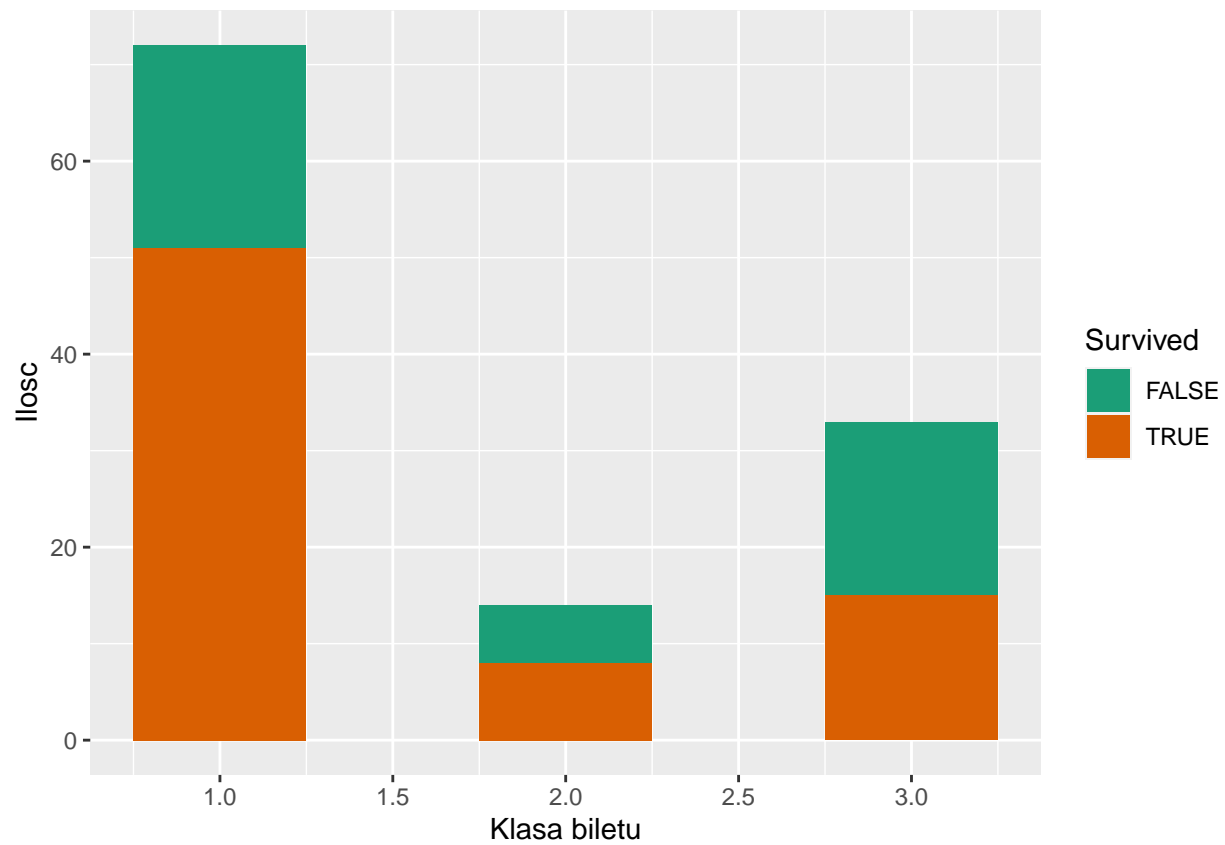




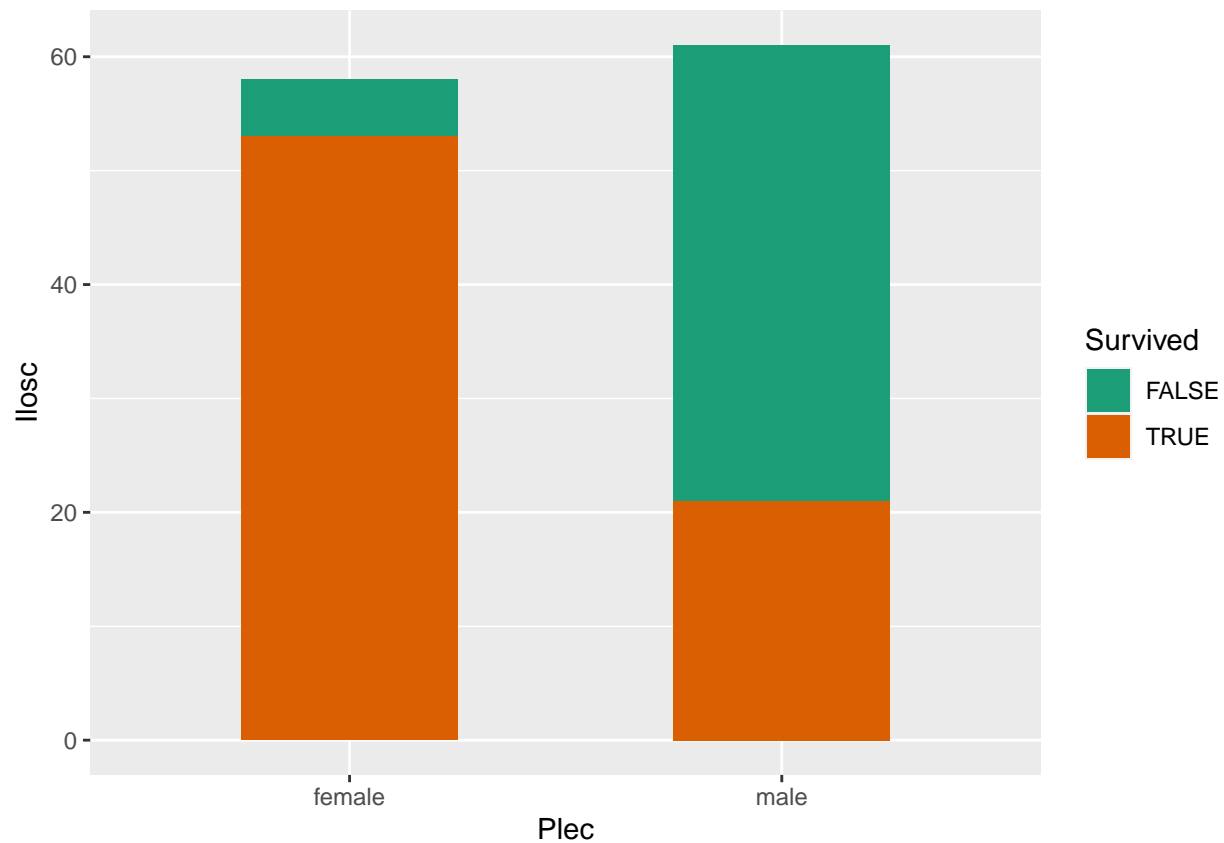
```
## embarked survRate
## 1      C 0.6218487
## 2      Q 0.3333333
## 3      S 0.3619403
```

Przeżywalność a port odprawy

- na wykresie widać, że mimo iż stosunkowo niewiele osób odpływało z portu w Cherbourg to miały one wyraźnie największe prawdopodobieństwo przetrwania
- sprawdźmy rozkład klas biletów odpływających z tego portu



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



Podsumowanie

- Po analizie danych pod wieloma kątami wygląda na to, że głównymi czynnikami dla przetrwania były płeć i klasa biletu, zaraz potem wiek a reszta czynników była skorelowana z przetrwaniem ale niekoniecznie wpływała na nie
- największą szansę przetrwania na Titanicu miały kobiety i dzieci płynące pierwszą klasą, zgodnie z przewidywaniami