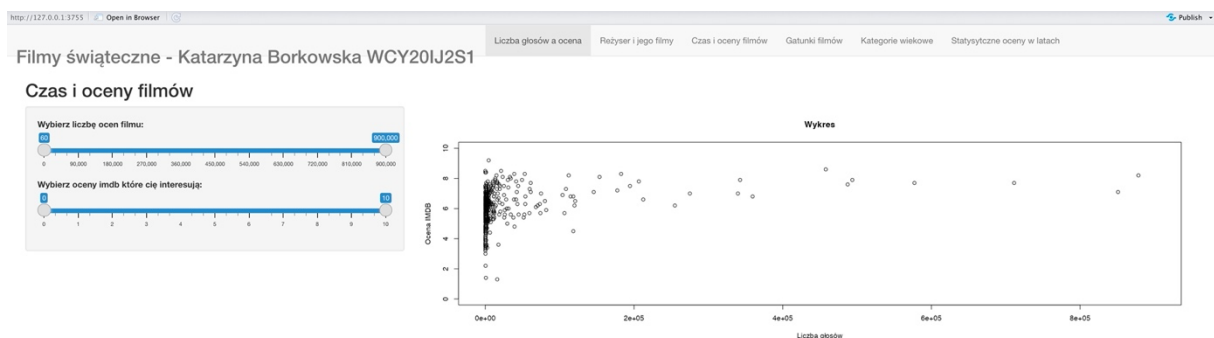
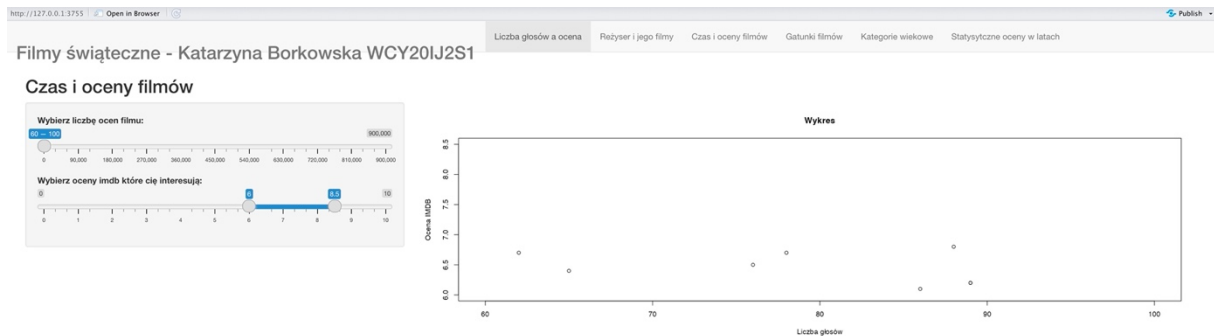


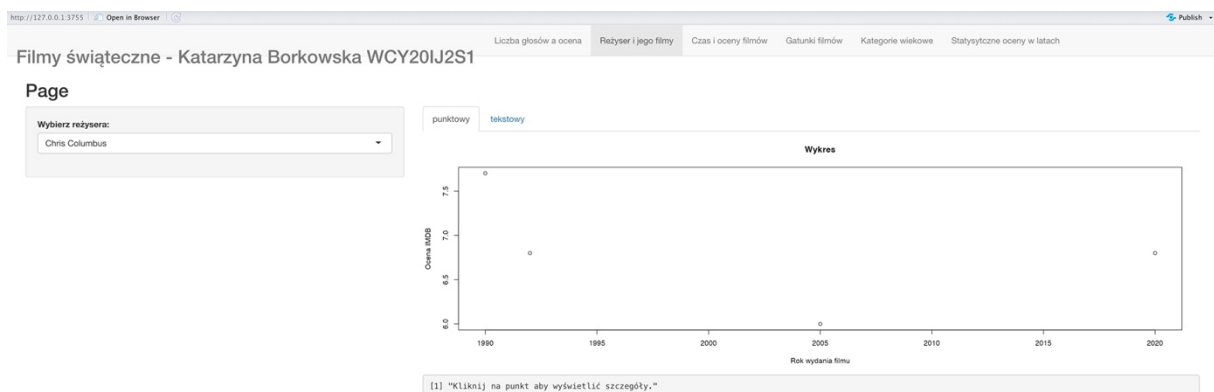
Projekt

W moim projekcie analizuję dane zawarte w pliku `christmas_movies.csv` w którym znajdują się dane o filmach świątecznych oparte o informacje ze strony `imdb.com`.

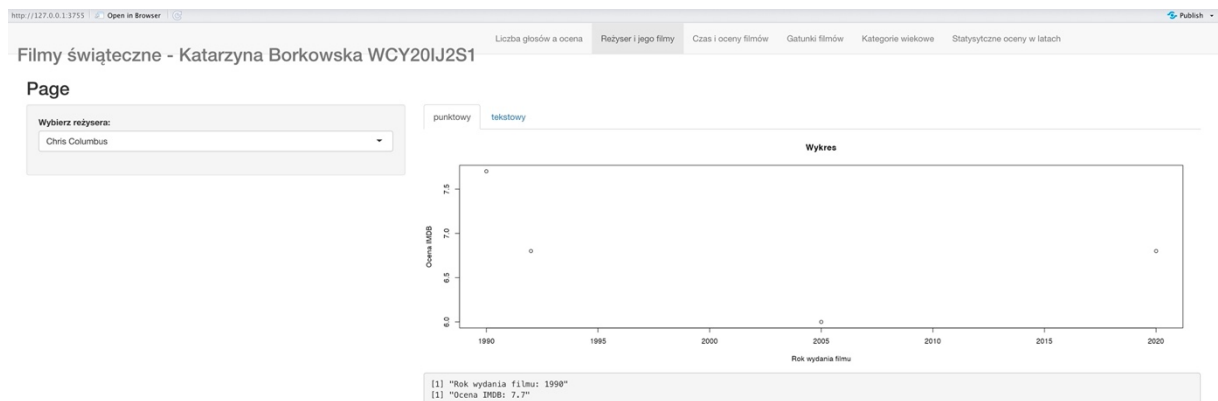
W pierwszym kroku musiałam dostosować dane do możliwości analizy w języku R. To znaczy usunęłam duplikaty oraz kolumny niepotrzebne mi w analizie. A także zamieniłam typy zmiennych z `char` na `numeric` aby móc je swobodnie używać w analizie.



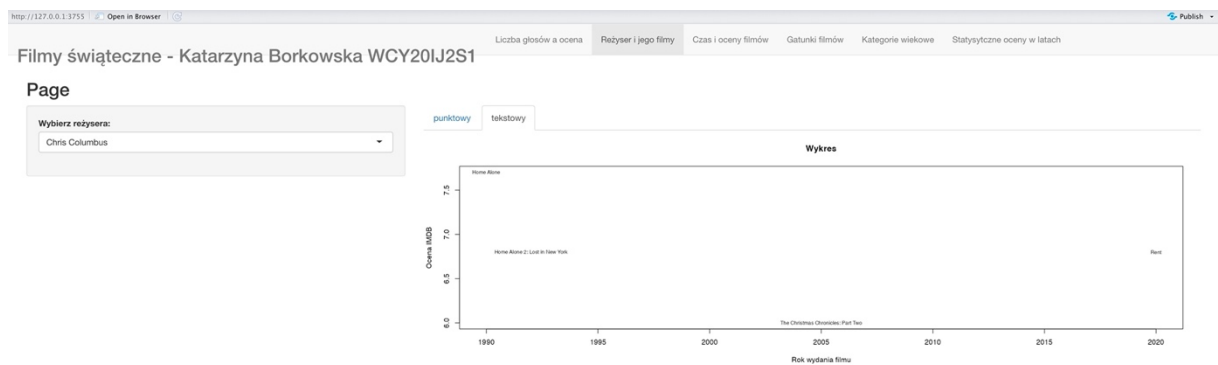
W pierwszym wykresie analizujemy jakie oceny dostały filmy w zależności od ilości głosów. Możemy modyfikować wykres zarówno na osi x jak i y przy pomocy suwaków. Wykres na maksymalnych zakresach ukazuje nam, że przy niskiej ilości ocen filmu ma to wpływu na jego ocenę, jednak przy dużej ilości ocen możemy się spodziewać oceny filmu na poziomie 6-8 gwiazdek. Wniosek jest taki, że jeżeli film ma niskie oceny nie jest oglądany tak jak filmy z wysokimi ocenami.



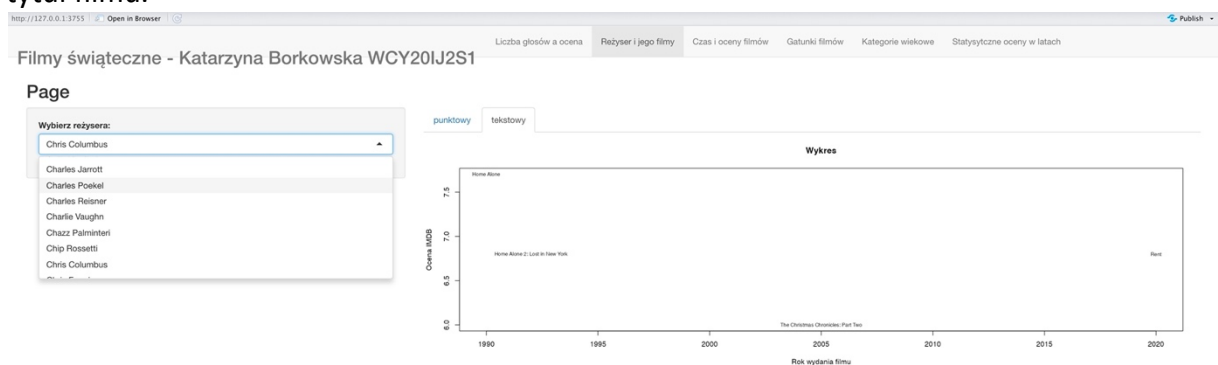
Na tym wykresie możemy zobaczyć zależność oceny filmu w latach danego reżysera. Jest tu umieszczona także opcja w której możemy po naciśnięciu na wykresie punktu wyświetlić jego wartości z osi x i y.



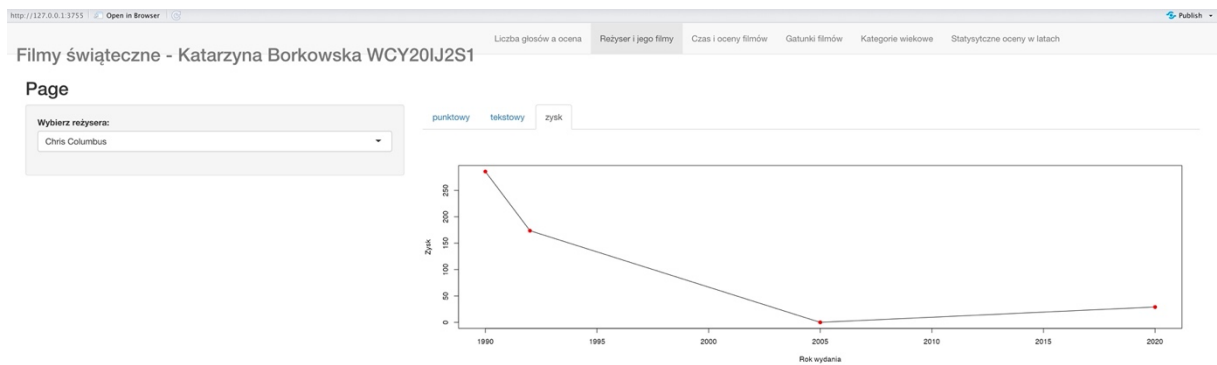
Jest tu przykład po naciśnięciu punktu najbardziej na lewo.



Jest też dostępna opcja wykresu tekstowego w którym zamiast punktów wyświetlany jest tytuł filmu.

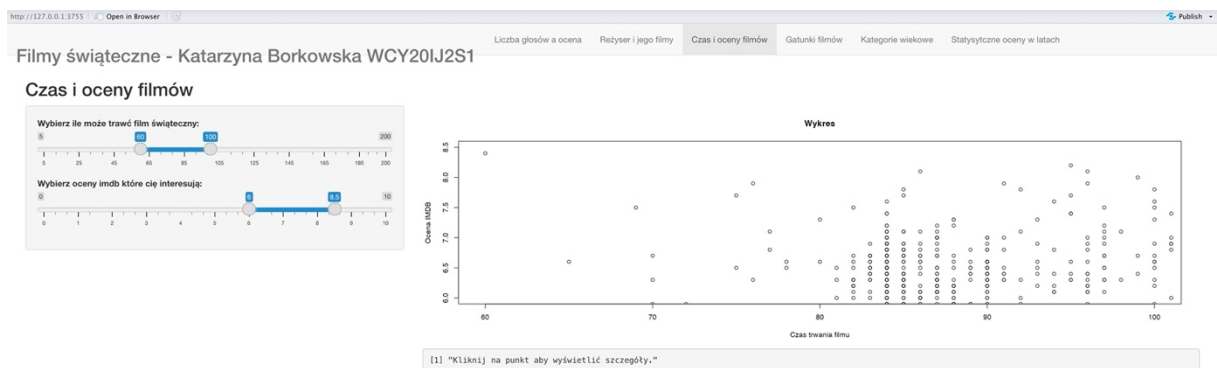


Reżysera wybieramy z rozwijanej listy z lewej strony.

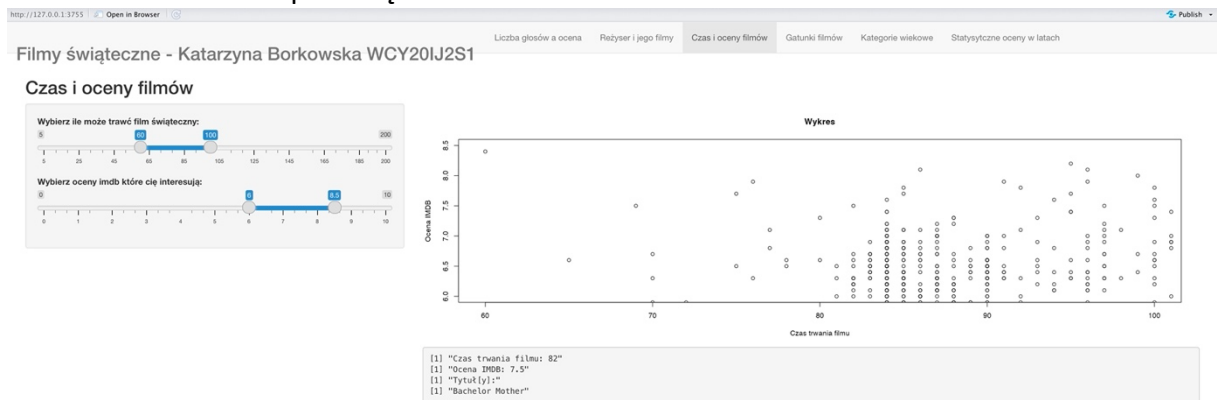


Tutaj mamy też zyski filmów danego reżysera według lat.

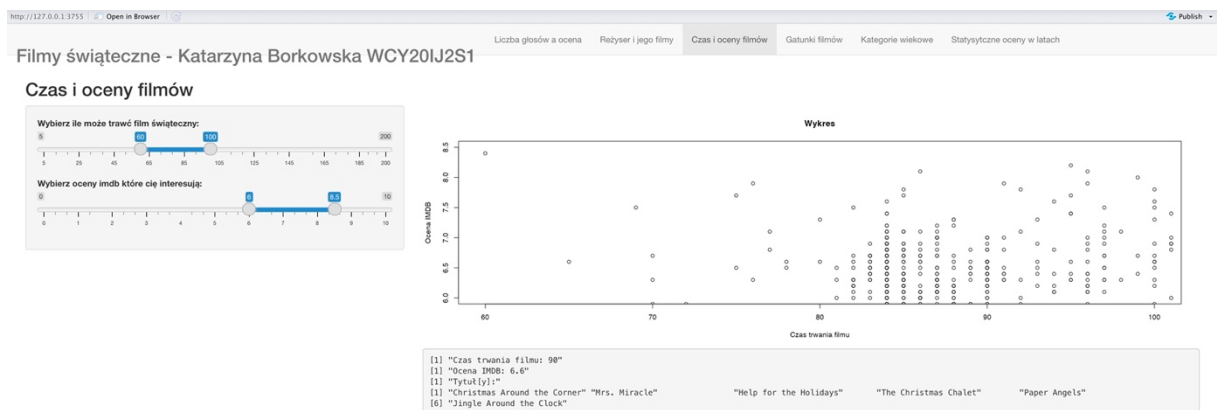
Z wykresu o reżyserze Chris Columbus możemy zobaczyć, że mimo hitu jakim był Kevin sam w domu (Home Alone) w kolejnych filmach nie odniósł takiego sukcesu. Widzimy też, że ocena filmu i jego zysk są powiązane. Wysoka ocena to też wysoki zysk filmu, a przy niskiej ocenie jest niski zysk.



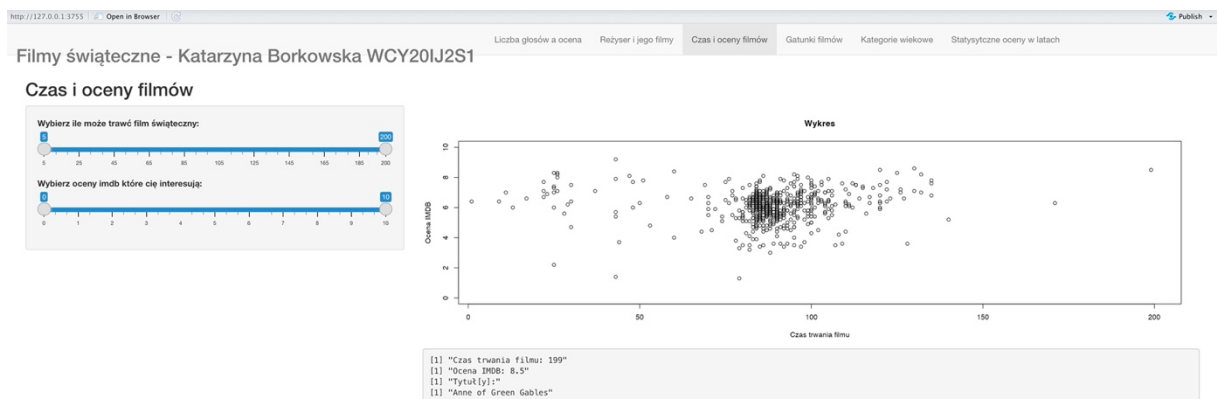
Kolejny wykres przedstawia zależność między czasem trwania filmu a jego oceną. Wartości te można ustawiać za pomocą suwaków.



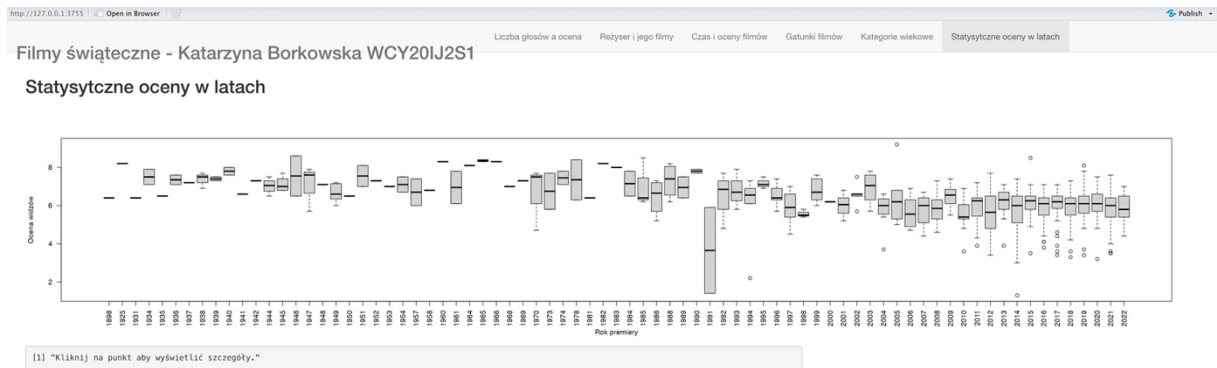
Po naciśnięciu okręgu zostaje nam wyświetlony tytuł filmu oraz szczegóły takie jak czas trwania filmu oraz ocena IMBD.



Po naciśnięciu pogrubionego okręgu przedstawiane nam są tytuły kilku filmów.



Na wykresie możemy odczytać, że najwięcej jest filmów które trwają między 80 a 100 minutą. Natomiast ocena filmu nie ma związku z czasem jego trwania.



Ten wykres to wykres pudełkowy (ang. box plot). Przedstawia on rozkład danych w postaci kilku podstawowych statystyk, takich jak minimum, pierwszy kwartył, median, trzeci kwartył i maksimum. Wykres pudełkowy pozwala łatwo porównywać rozkłady danych w różnych grupach. W tym przypadku dane są podzielone według roku premiery filmu (os X) i dla każdej grupy przedstawiona jest ocena IMDB widzów (os Y). Możemy tu także tutaj klikając na dany punkt w wykresie odczytać ocenę widzów co ułatwia odczytywanie wartości z osi Y. Wykres ukazuje, że po roku 2000 wartości utrzymują się na podobnym poziomie. Warto się też zainteresować co się stało w roku 1991, że jest taki spadek ocen w tym roku.

Wnioski:

Udało mi się wykonać założenia projektu. Sprawił mi on małe problemy, ale dzięki temu udało mi się zgłębić tajniki języka R oraz pakietu shiny. Rysując wykresy starałam się korzystać z różnych technik i bibliotek, aby ukazać jak najwięcej rzeczy których nauczyłam się na zajęciach z tego przedmiotu.