

Wyniki uczniów, a ich środowisko

405799, Szymon Żychowicz, poniedziałek 16:15

AGH, Wydział Informatyki, Elektroniki i Telekomunikacji

Rachunek Prawdopodobieństwa i statystyka 2021/2022

Kraków, 27 stycznia 2022

Ja, niżej podpisany własnoręcznym podpisem deklaruję, że przygotowałem przedstawiony do oceny projekt samodzielnie i żadna jego część nie jest kopią pracy innej osoby.

Szymon Żychowicz

Streszczenie raportu

Raport powstał w oparciu o analizę danych, dotyczących ocen uczniów i informacji o środowisku, w którym żyją.

Opis danych

Dane do projektu pochodzą ze strony:

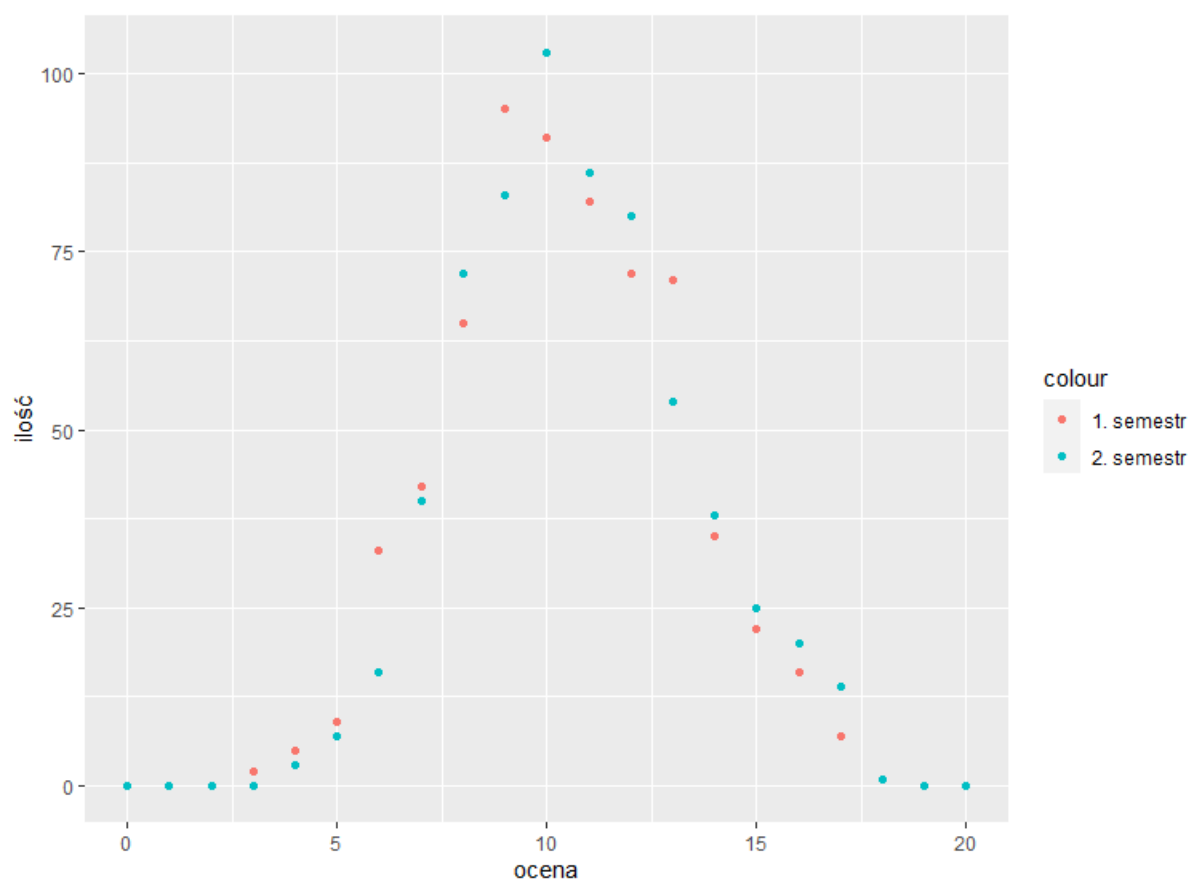
<https://www.kaggle.com/impapan/student-performance-data-set>

Plik w formacie .csv przechowuje 649 wierszy po 33 kolumny danych. Oto znaczenie kolumn:

1. school – szkoła, do której uczeń uczęszcza,
2. sex – płeć metryczna ucznia,
3. age – wiek ucznia,
4. address – informacja czy uczeń mieszka w mieście czy poza nim,
5. famsize – wielkość rodziny,
6. Pstatus – informacja czy rodzice ucznia mieszkają razem,
7. Medu – wykształcenie matki,
8. Fedu – wykształcenie ojca,
9. Mjob – praca matki,
10. Fjob – praca ojca,
11. reason – powód wybrania szkoły,
12. guardian – opiekun ucznia,
- 13. traveltime - czas podróży do szkoły,**
- 14. studytime – tygodniowy czas nauki,**
15. failures – ilość wcześniej niezaliczonych przedmiotów,
16. schoolsup – beneficjent dodatkowych pomocy naukowych od szkoły,
17. famsup – informacja czy rodzina pomaga w nauce,
18. paid – „korki”,
19. activities – zajęcia nadobowiązkowe,
20. nursery – uczęszczał do przedszkola,
21. higher – chce iść na studia,
22. internet – ma dostęp do internetu,
23. romantic – jest w związku,
- 24. famrel – stosunki w rodzinie,**
- 25. freetime – ilość czasu wolnego,**
- 26. goout – częstotliwość wychodzenia ze znajomymi,**
- 27. Dalc – spożycie alkoholu w tygodniu roboczym,**
- 28. Walc – spożycie alkoholu w weekend,**
- 29. health – stan zdrowia,**
- 30. absences – ilość nieobecności,**
- 31. G1 – ocena z pierwszego semestru,**
- 32. G2 – ocena z drugiego semestru,**
33. G3 – ocena końcowa;

Analiza danych

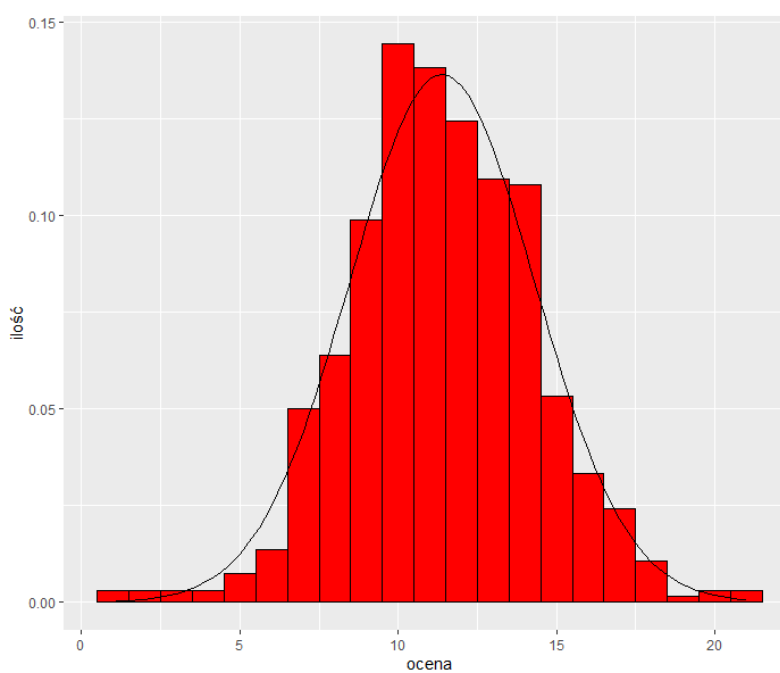
Na początek przyjrzyjmy się samym ocenom uczniów.



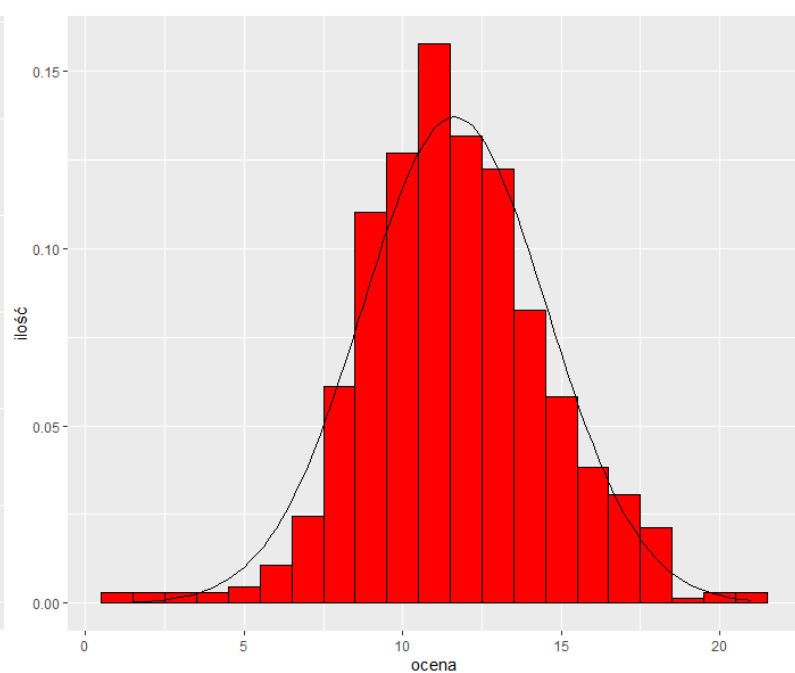
Na wykresie przedstawiona jest ilość uczniów (oś Y) z daną oceną (oś X; uwaga: skala ocen jest inna niż w Polsce: 0-20, zamiast 1-6 lub 2-5) w odpowiednio pierwszym i drugim semestrze. Kształt dla obu zbiorów punktów zdaje się być znajomy. Można dopasować do nich rozkład normalny. Jego parametry to odpowiednio:

1. Semestr:
mean = 11.386018 ,
sd = 2.922492;
2. Semestr:
mean = 11.637615 ,
sd = 2.912883;

Widać, że odchylenie standardowe jest bardzo podobne, ale średnia różni się o około 0,25. Wynika z tego, że oceny na drugim semestrze poprawiły się średnio o 0,25 stopnia.



Pierwszy semestr

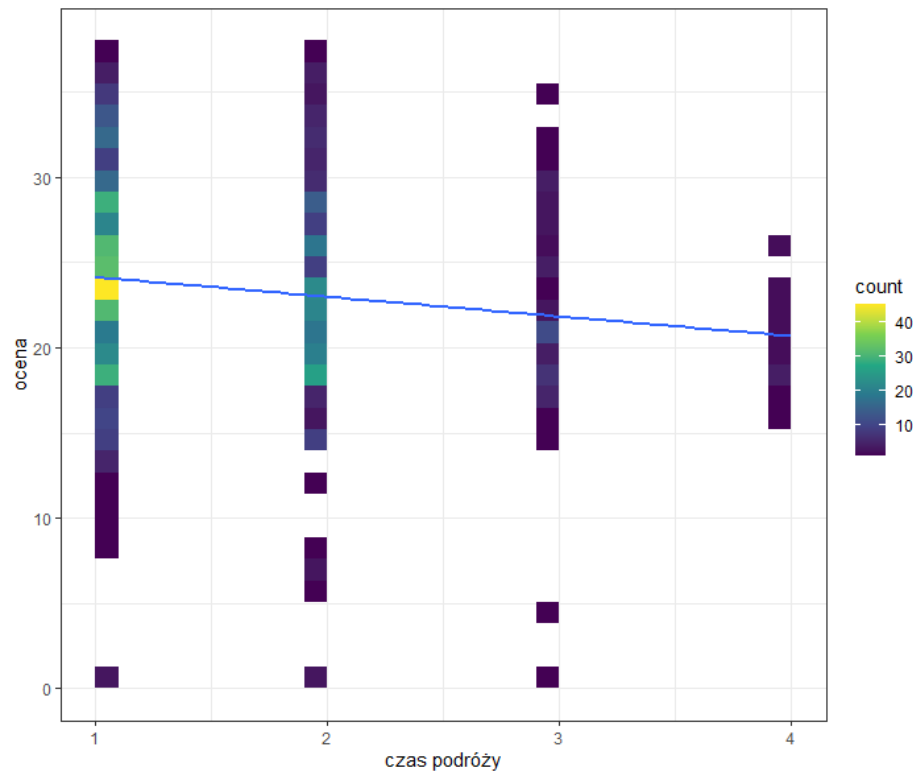


Drugi semestr

Spodziewamy się, że oceny uczniów mogą zależeć od:

- 1) czasu podróży do szkoły,
- 2) czasu nauki,
- 3) jakości rodzinnych relacji,
- 4) jakości zdrowia,
- 5) ilości nieobecności,
- 6) ilości spożywanego alkoholu,
- 7) częstości wyjść ze znajomymi,
- 8) ilości czasu wolnego;

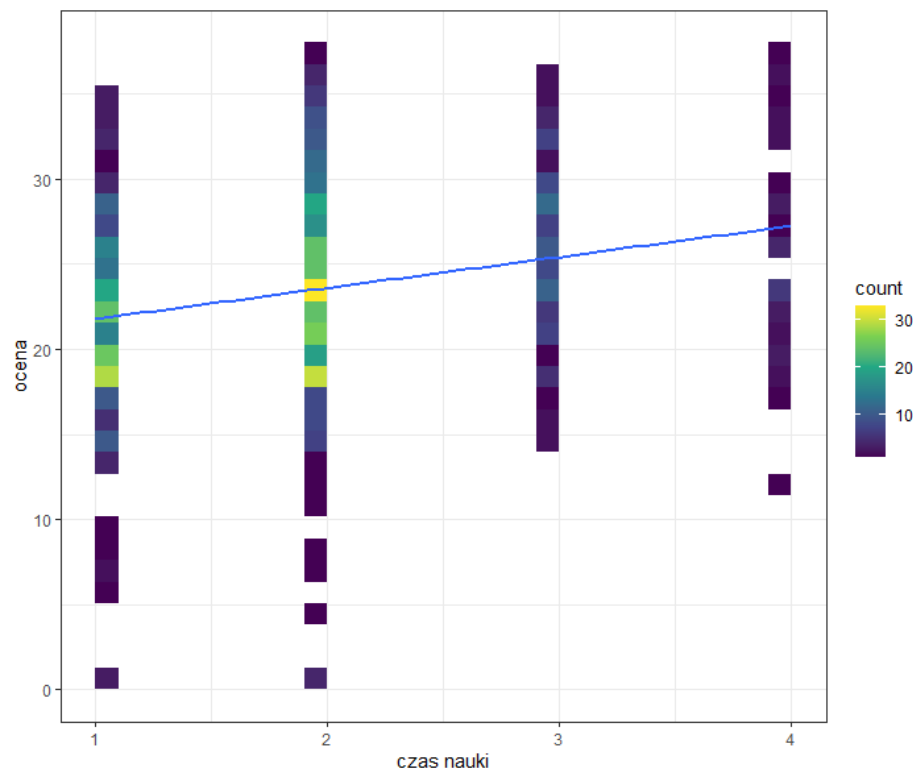
Czas podróży do szkoły



Na wykresie przedstawiono rozkład ocen uczniów w zależności od czasu jaki zajmuje im na dojazd do szkoły. Widzimy, że wraz z wzrostem czasu dojazdu poza spadkiem liczby uczniów, spada także średnia ocen.

x	czas podróży
1	mniej niż 15 min
2	15 do 30 min
3	30 min do 1 h
4	ponad 1 h

Czas nauki

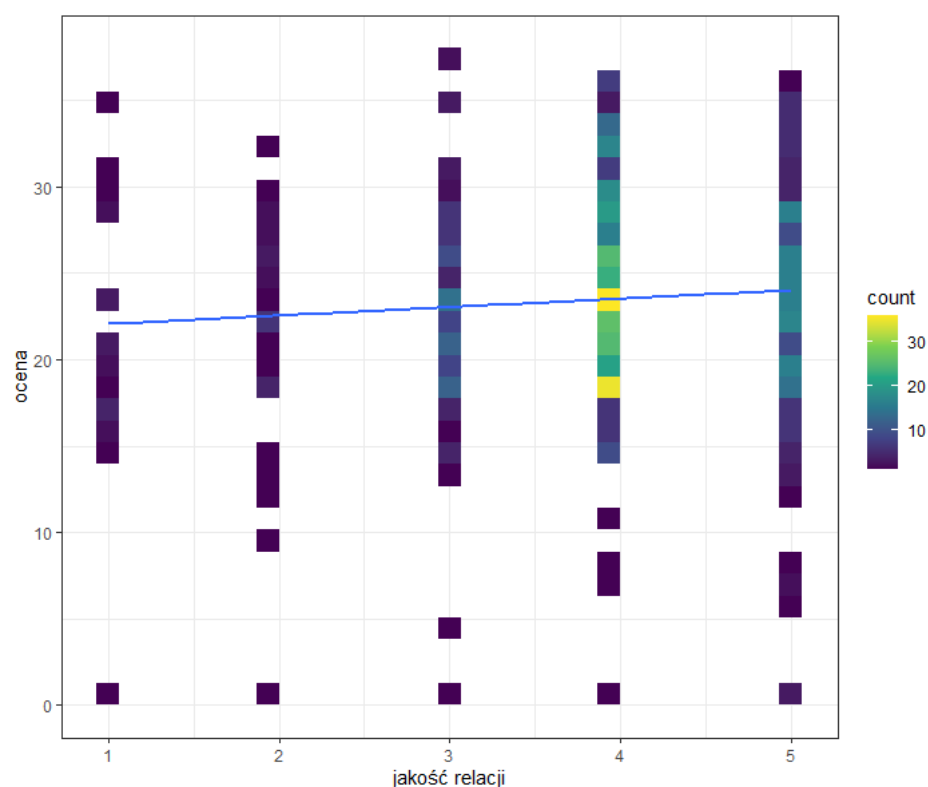


Jak można się spodziewać czas nauki bardzo wpływa na oceny. Aby to pokazać:

x	tygodniowy czas nauki
1	mniej niż 2 h
2	2 do 5 h
3	5 do 10 h
4	ponad 10 h

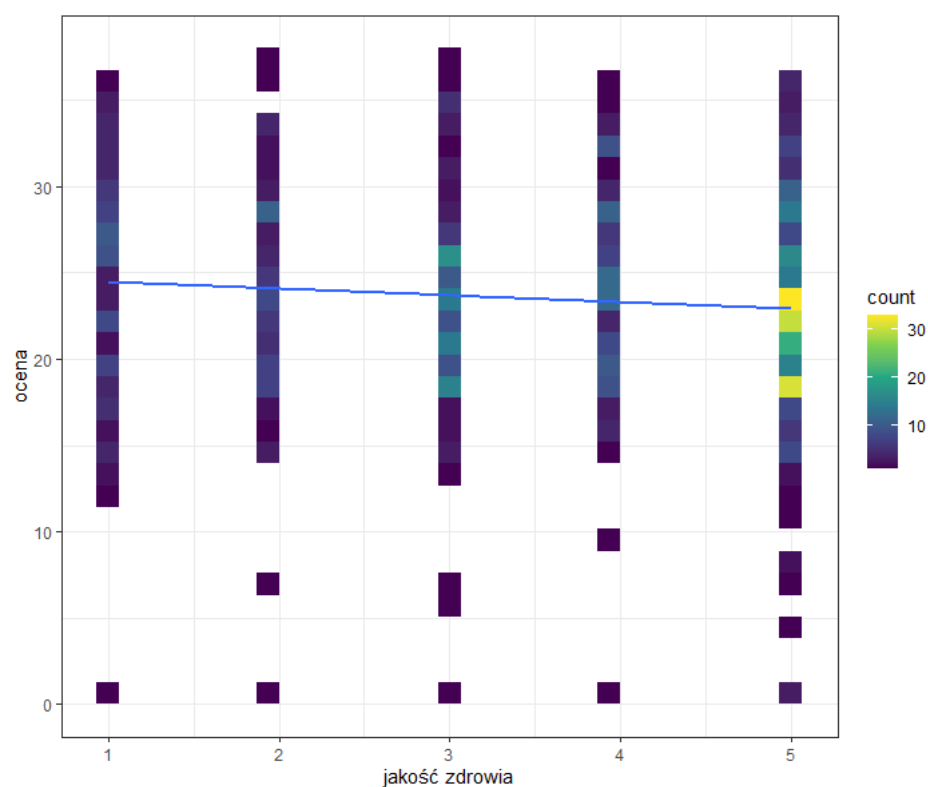
Średnia ocena osoby uczącej się <2h wynosi ~21.55, a uczącej się <5h wynosi już ~23.75, czyli o ponad 2 stopnie więcej.

Jakość rodzinnych relacji



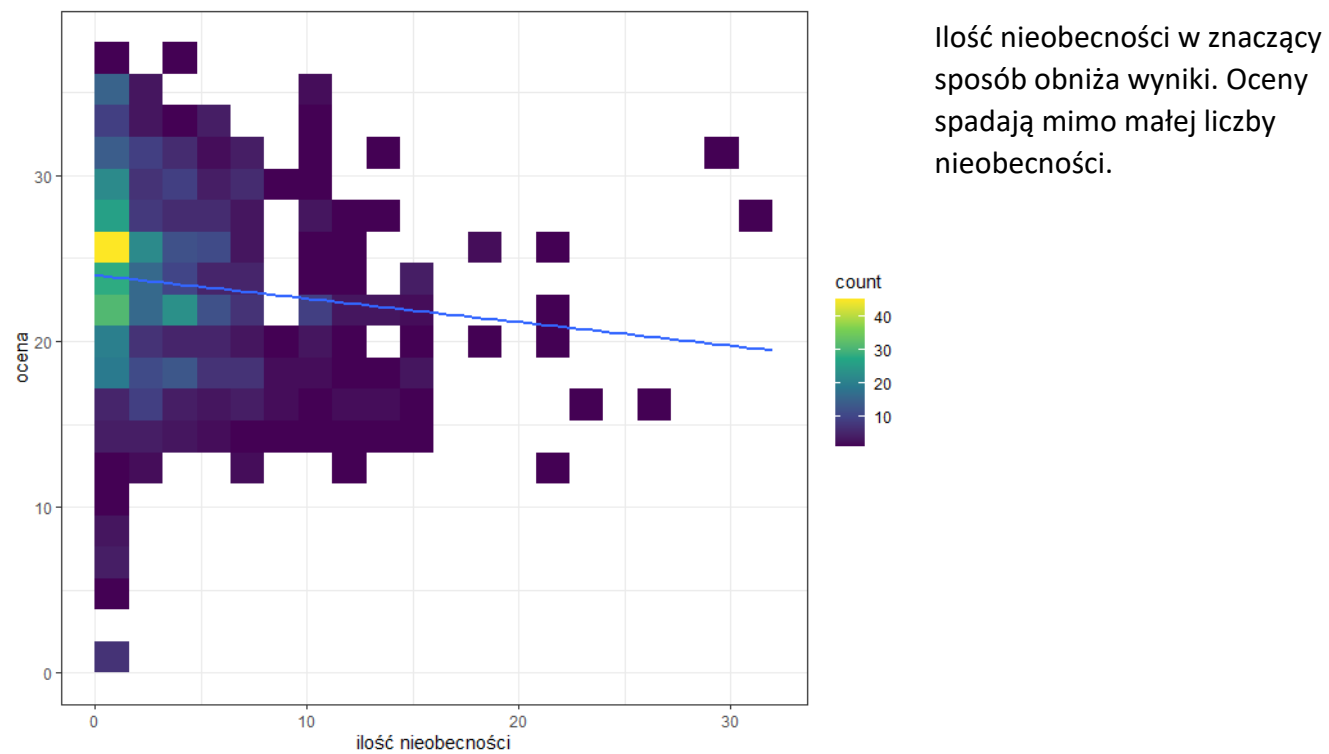
Dobre relacje rodzinne wpływają pozytywnie na oceny, jednak gdy nie ma się z nimi większych problemów to wpływ jest bardzo nieznaczny.

Jakość zdrowia



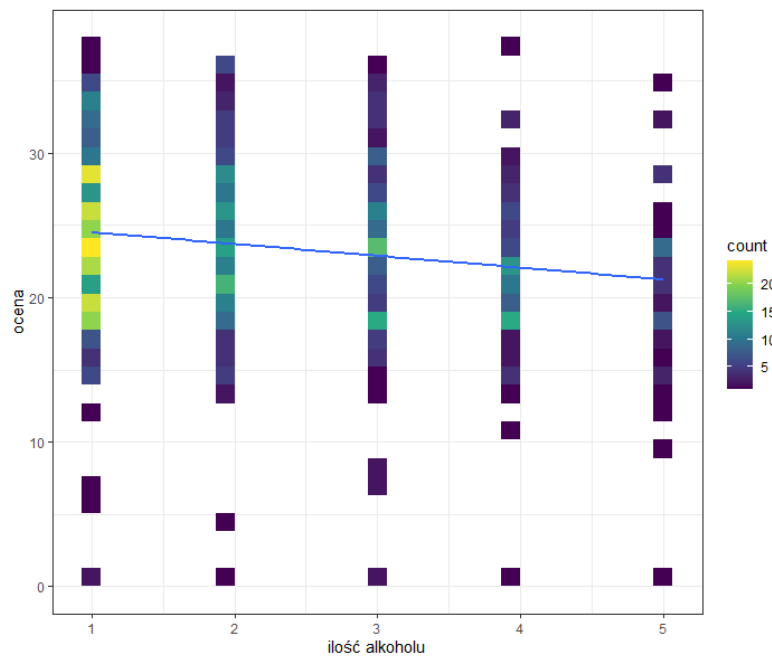
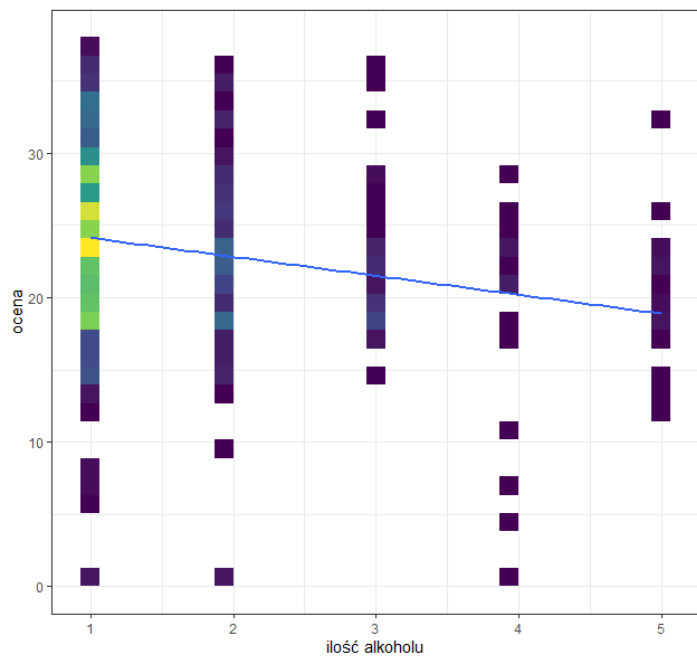
Wykres wskazuje na to, że stan zdrowia działa negatywnie na wyniki w nauce, jednak licznosci dysproporcje w licznosciach badanych grup są nie do pominięcia. Wykres opisuje także aktualny stan zdrowia a nie ogólne jego odczucie.

Ilość nieobecności



Ilość spożywanego alkoholu

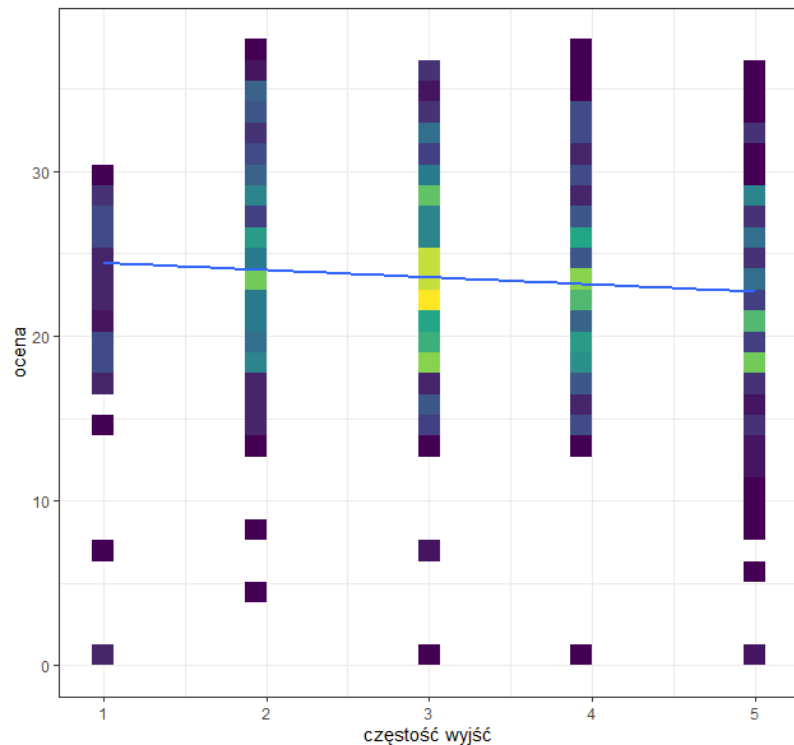
Dostępne dane pozwalają nam sprawdzić wpływ spożywanego alkoholu na oceny z podziałem na spożycie w tygodniu roboczym i w weekendy.



Na wykresie pierwszym widać, że wraz ze wzrostem ilości spożywanego alkoholu w tygodniu średnie oceny drastycznie spadają. Należy jednak zauważyć, że osób spożywających alkohol w tygodniu jest dużo mniej, niż tych z grup 1 i 2, więc dla większej populacji wynik może być inny.

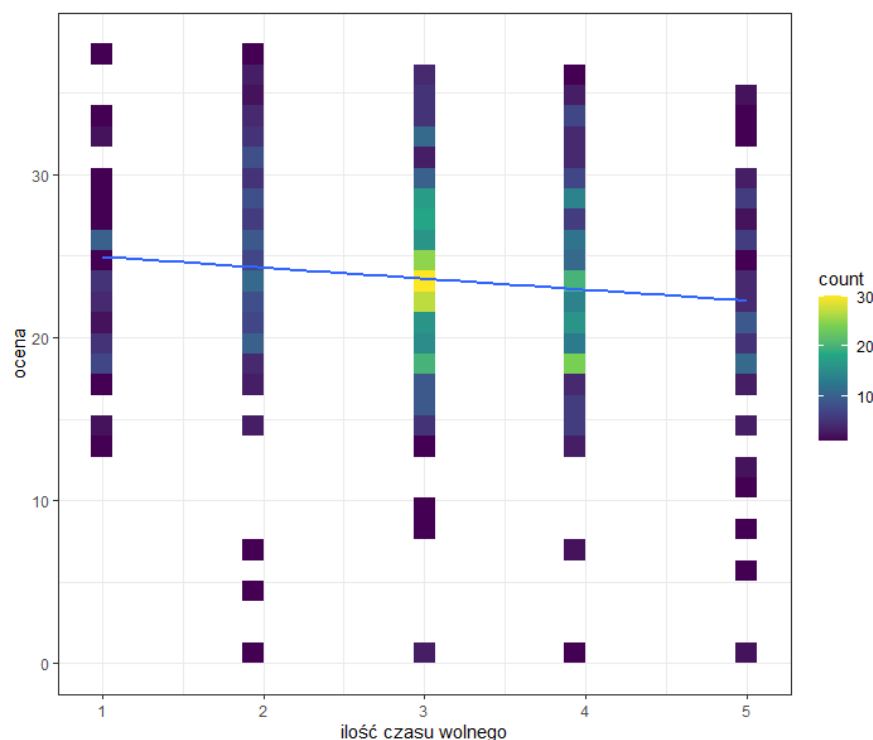
Inaczej sytuacja wygląda na drugim wykresie, gdzie ilości uczniów są bardziej równomiernie rozłożone we wszystkich grupach. Wraz z wzrostem ilości spożycia alkoholu oceny spadają, ale mniej niż w poprzednim przypadku.

Częstość wyjść ze znajomymi



Zgodnie z intuicją większa ilość wyjść ze znajomymi powoduje spadek ocen. Pocieszające jest, że spadek jest bardzo mały, a wyniki licznych grup 2, 3 i 4 są bardzo zbliżone.

Ilość czasu wolnego



Najliczniejsze grupy (3 i 4) mają podobne wyniki, a w grupach 2 i 5 widać znaczny rozrzut wyników. Może to oznaczać, że wyniki zależą od powodów, dla których ktoś ma mało lub dużo czasu (nie nadąża z nauką, problemy pozaszkolne, problem z organizacją czasu).

Wnioski

Przeanalizowane wskaźniki pokazują, że środowisko w jakim obraca się uczeń ma znaczący wpływ na jego wyniki, jednak na każdym wykresie widać, że dla pojedynczego czynnika wpływ jest mały. Można zauważyć, że wśród tych czynników, żaden nie determinuje oceny ucznia (słupki są zawsze wysokie).

Podsumowując:

Pozytywny wpływ	Negatywny wpływ
Jakość rodzinnych relacji	Czas podróży do szkoły
Czas nauki	Ilość spożytego alkoholu
	Ilość nieobecności
	Ilość czasu wolnego