

Poprawna analiza wykresów

Paweł Stefańczyk

1 Cel pracy

Celem tej pracy domowej jest przygotowanie wybranych wykresów, przedstawiających ten sam zbiór danych, ale w różnych formach. Należy sprawdzić, który rodzaj wykresu jest najwygodniejszy i który niesie najbardziej rzetelne informacje. W tym celu przeprowadziłem ankietę w gronie 15 osób, dzięki którym będzie możliwa wiarygodna ocena eksperymentu.

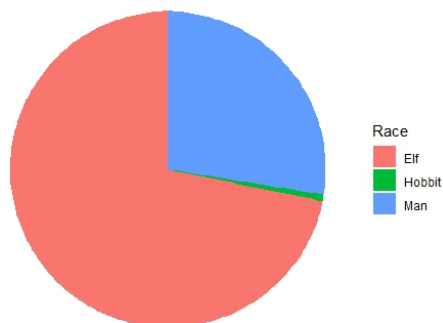
2 Dane

Korzystamy z następującej ramki danych:

	Race	Gender	Words
1	Elf	Female	1743
2	Elf	Male	1994
3	Hobbit	Female	16
4	Hobbit	Male	8780
5	Man	Female	669
6	Man	Male	8043

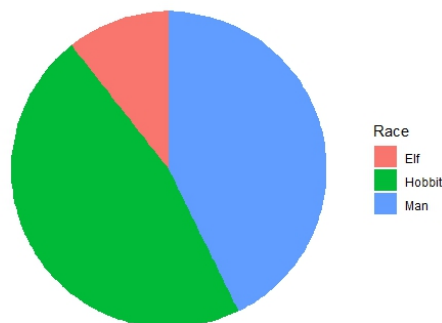
Jest to zbiór przedstawiający liczbę wyrazów, które zostały użyte przez przedstawicieli poszczególnych ras z podziałem na płeć z filmów "Władca Pierścieni". W celu wizualizacji przygotowałem następujące wykresy:

Liczba słów z filmów dla żeńskich postaci

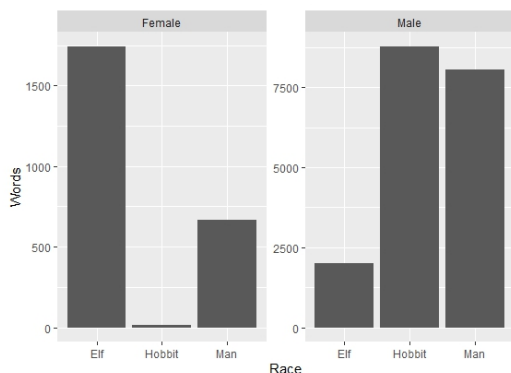


Rysunek 1: Wykres 1

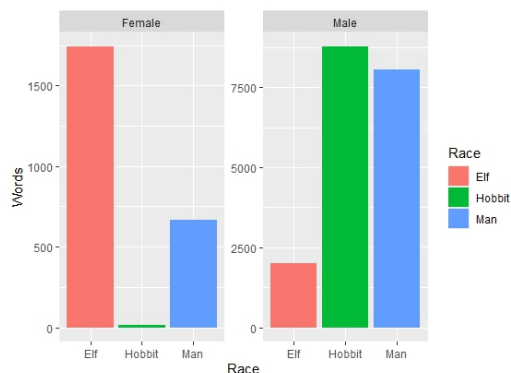
Liczba słów z filmów dla męskich postaci



Rysunek 2: Wykres 2



Rysunek 3: Wykres 3



Rysunek 4: Wykres 4

3 Wyniki ankiety

Wyniki eksperymentu na temat, który wykres jest lepszy przedstawia tabela:

Osoba	Wykres kołowy	Wykres słupkowy	Oś i legenda
1	✓	✗	✓
2	✗	✓	✗
3	✗	✓	✓
4	✓	✗	✗
5	✓	✗	✓
6	✓	✗	✗
7	✗	✓	✗
8	✗	✓	✗
9	✓	✗	✗
10	✗	✓	✓
11	✗	✓	✗
12	✗	✓	✗
13	✗	✓	✓
14	✓	✗	✗
15	✗	✓	✗

Badani, którzy wybrali wykres kołowy kierowali się głównie estetyką tego typu wykresu, jednak przyznali, że jeżeli różnica słów między danymi rasami jest mała to trudno jest skonstatować, który kąt jest większy. Ankietowani, którzy zdecydowali się na wykres słupkowy byli za czytelnością tego typu wykresu i jego prostotą. Jeżeli chodzi o jednoczesne zamieszczenie legendy i etykiet na osi poziomej, badani mieli różne zdanie na ten temat. Większość przyznała, że nie powinno się powielać informacji, ponieważ może wprowadzić to zamieszanie. Pozostali zaś powiedzieli, że wykres słupkowy z większą liczbą kolorów przykuwa większą uwagę.

4 Wnioski

- Wykresy kołowe mimo swojej estetyki nie zawsze dają dokładne wyniki
- Wykresy słupkowe cieszą się popularnością dzięki swojej prostocie i widocznym subtelnym różnicom między danymi
- Podwajanie tej samej informacji nie jest dobrym pomysłem nawet kosztem straty estetyki wykresu

Źródła

Dane pochodzą ze strony: https://github.com/jennybc/lotr-tidy/blob/master/data/lotr_tidy.csv