

Praca domowa 3

Monika Bednarek, Jakub Konieczny

2023-03-27

1. Wprowadzenie

Naszym zadaniem było sprawdzenie, jak dobre praktyki wykorzystywane w wizualizacji danych wpływają na czytelność i generalny odbiór oraz interpretację wykresów. W przeprowadzonych przez nas eksperymentach skupiliśmy się na zbadaniu tego, jak odbierane są wykresy kołowe oraz wykresy 3D.

2. Opis badania

Do naszych eksperymentów wykorzystaliśmy dwa zestawy danych i dla każdego z nich utworzyliśmy po dwa wykresy. Następnie w Google Forms stworzyliśmy ankietę, w której zadaliśmy pytania dotyczące wizualizacji. W ankiecie wzięło udział 68 osób.

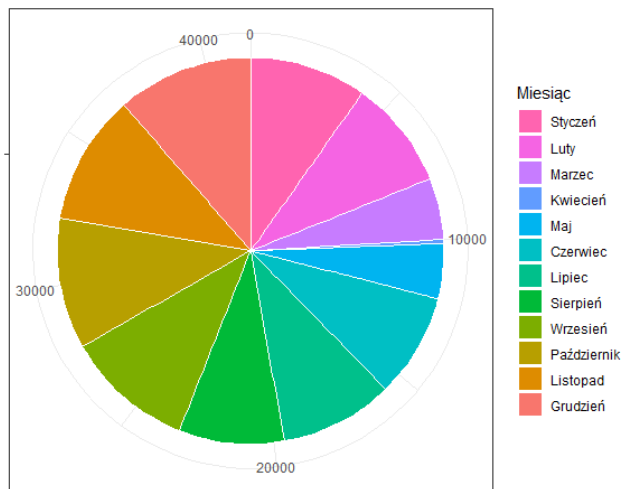
3. Eksperyment 1

W pierwszym eksperymencie wykorzystaliśmy dane dotyczące liczby wydanych w województwie mazowieckim dokumentów prawa jazdy w roku 2020. Dane te pochodzą ze strony Centralnej Ewidencji Pojazdów i Kierowców:

<http://www.cepik.gov.pl/documents/76251/76577/Prawa+jazdy+wydane+w+roku+2020+%28pdf%29/c547d611-b77f-4d56-a93c-47b87196b999>

Na początku ankietowani otrzymali poniższy wykres:

Wydane dokumenty prawa jazdy w podziale na miesiące w 2020 roku



Na jego podstawie poprosiliśmy o udzielenie odpowiedzi na dwa pytania.

Pytanie 1: W którym z poniższych miesięcy wydano najwięcej dokumentów?

- luty
- czerwiec
- lipiec

Liczba głosów oddanych na poszczególne miesiące:

Luty	Czerwiec	Lipiec
20	20	28

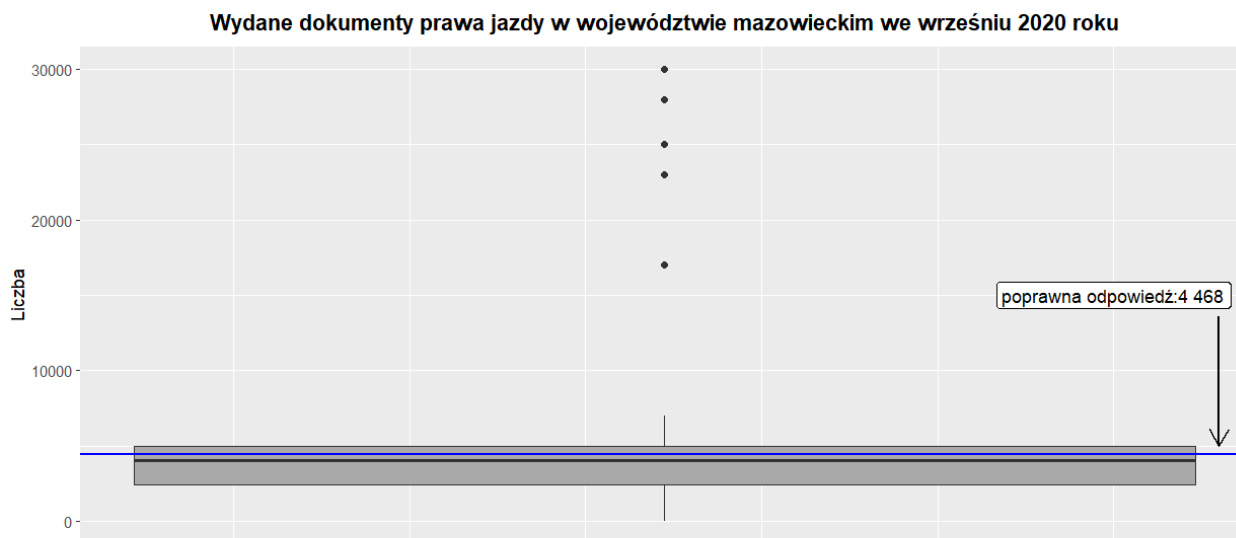
Poprawną odpowiedź na to pytanie był **Lipiec**.

Wnioski:

Głosy ankietowanych są podzielone niemalże po równo między trzema możliwymi odpowiedziami. Wyniki te pokazują zatem, że w przypadku wykresu kołowego trudno porównać wycinki odpowiadające podobnym wartościom.

Pytanie 2: Ile dokumentów prawa jazdy wydano we wrześniu? Wymagaliśmy odpowiedzi w postaci jednej liczby. Prawidłowa wartość to **4 468**.

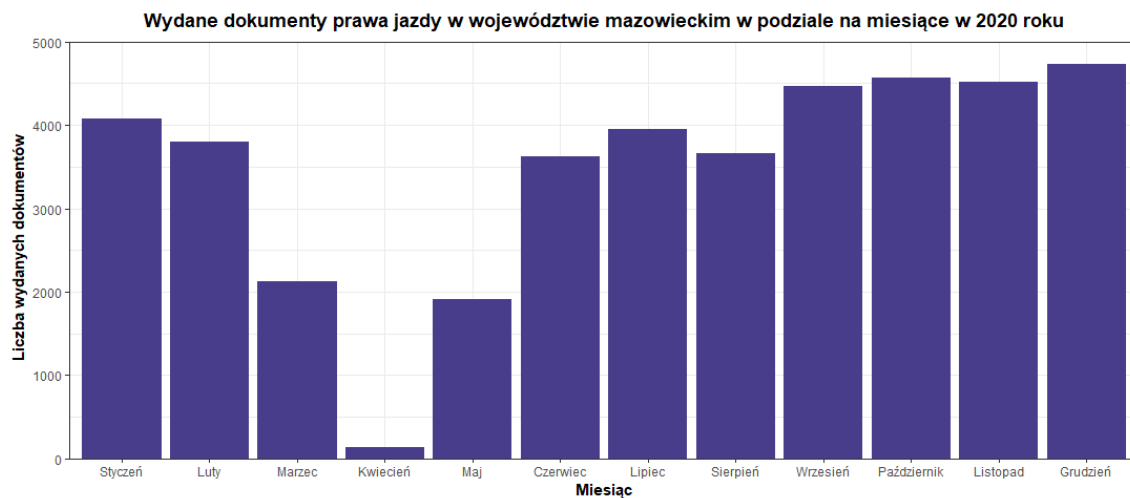
Wyniki ankiety przedstawiliśmy na wykresie skrzynkowym:



Obserwacje:

Rozstęp międzykwartyłowy jest dość duży i wynosi 2500. Oznacza to, że ankietowani mieli trudność ze wskazaniem poprawnej odpowiedzi.

Następnie prezentowany był drugi wykres:



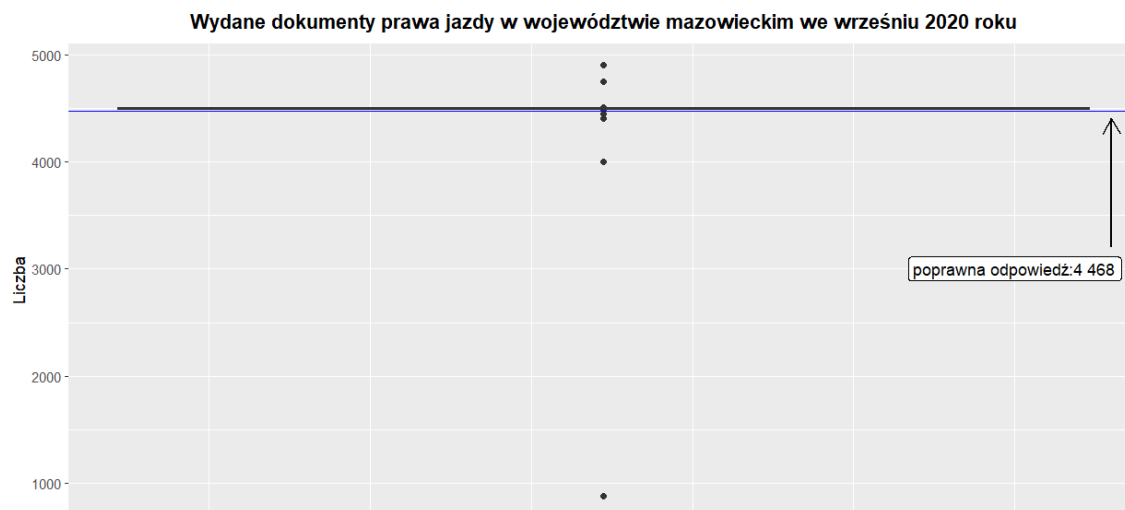
Ankietowani zostali poproszeni o ponowne odpowiedzenie na wcześniejsze pytania, lecz na podstawie drugiej wizualizacji.

1. Tym razem odpowiedzi wyglądały następująco:

Luty	Czerwiec	Lipiec
4	4	60

2. Ponownie należało wskazać, ile dokumentów prawa jazdy wydano we wrześniu.

Wyniki ankiety przedstawione na wykresie skrzynkowym:



Obserwacje:

W przypadku obu pytań znaczna większość odpowiedziała poprawnie lub była blisko poprawnej odpowiedzi.

Na końcu poprosiliśmy o wskazanie wykresu, który był czytelniejszy i łatwiejszy w interpretacji.

Odpowiedzi:

Wykres kołowy	Wykres słupkowy
2	66

Wnioski:

Wykres słupkowy sprawdził się dużo lepiej niż wykres kołowy. Dane prezentowane na nim są bardziej czytelne i łatwiej odczytać z niego wartości.

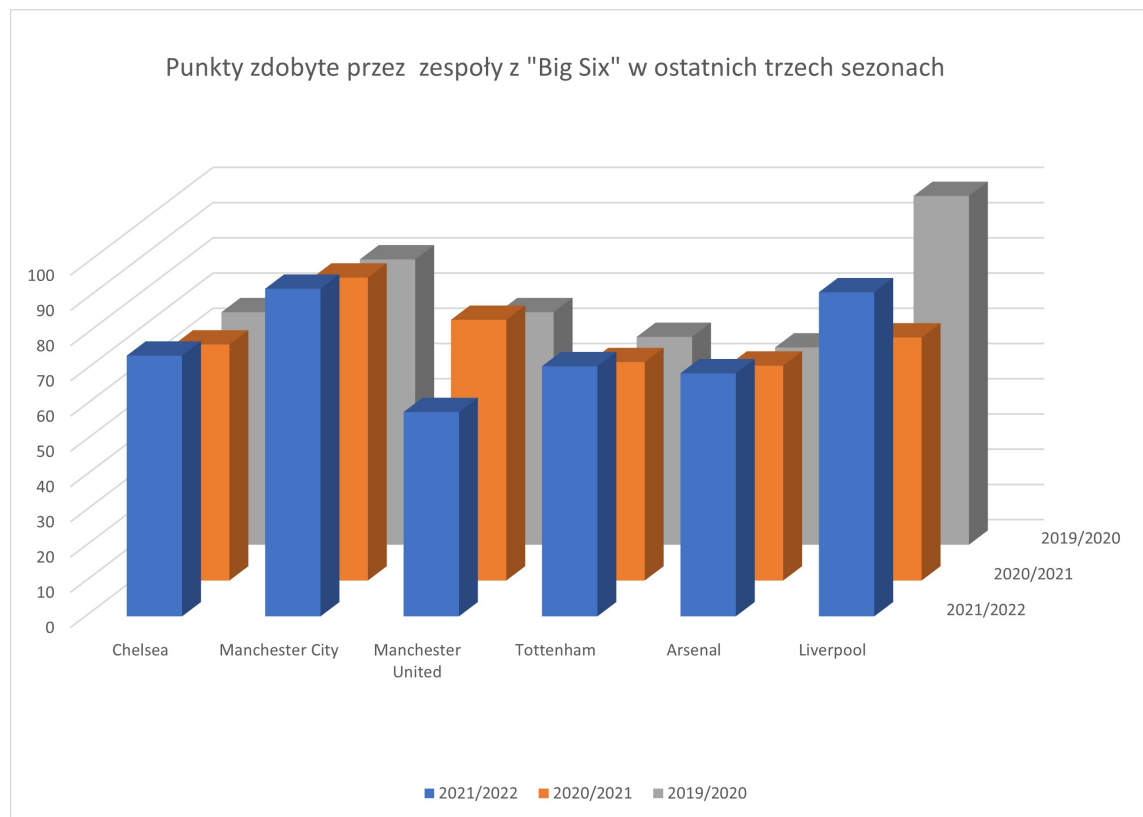
4. Eksperyment 2

W drugim eksperymencie wykorzystaliśmy dane dotyczące punktów zdobytych w Premier League przez zespoły z “Big Six” w ostatnich trzech sezonach. Pochodzą one ze strony Transfermarkt.

<https://www.transfermarkt.com/>

To badanie miało na celu sprawdzić jakie trudności można napotkać podczas odczytywania informacji z wykresów 3D.

Pierwszy wykres, który otrzymali ankietowani:



Na jego podstawie poprosiliśmy o udzielenie odpowiedzi na następujące pytania:

Pytanie 1: W którym z analizowanych sezonów, Chelsea zdobyła najmniej punktów?

- 2019/2020
- 2020/2021
- 2021/2022

Zebrane odpowiedzi wyglądają następująco:

2019/2020	2020/2021	2021/2022
15	43	10

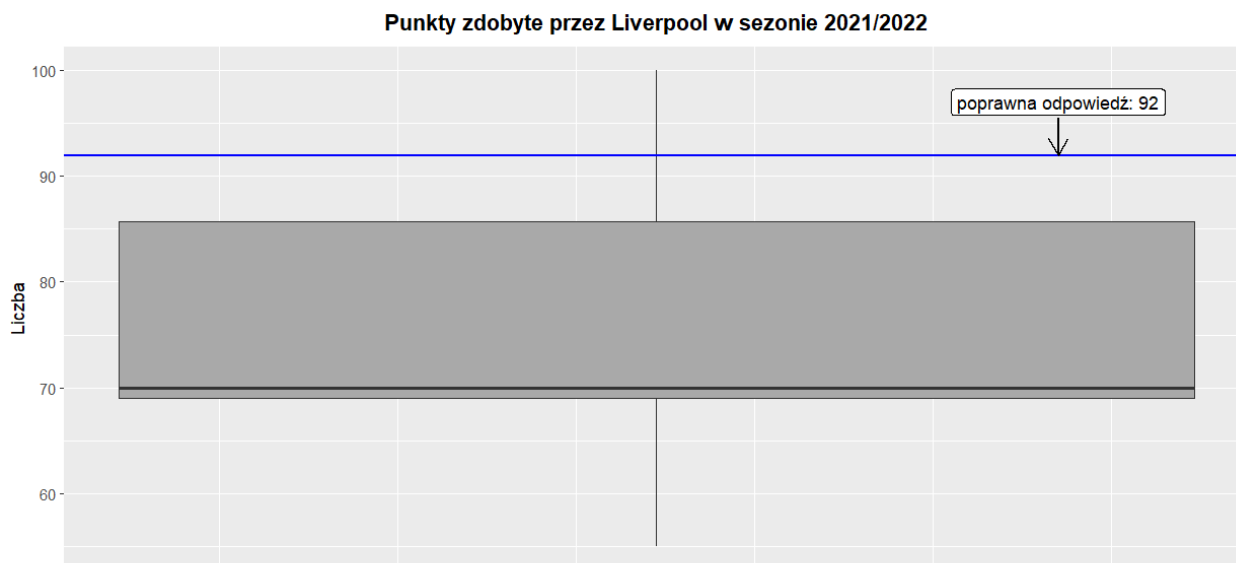
Poprawną odpowiedzią był sezon **2019/2020**.

Obserwacje:

Aż 63% badanych wskazało sezon 2020/2021 jako poprawną odpowiedź. Pokazuje to, że wykres 3D może nie tylko uniemożliwiać poprawne odczytanie danych, ale nawet wprowadzać odbiorcę w błąd.

Pytanie 2: Ile punktów zdobył Liverpool w sezonie 2021/2022? Wymagaliśmy odpowiedzi w postaci jednej liczby. Poprawną odpowiedzią było **92**.

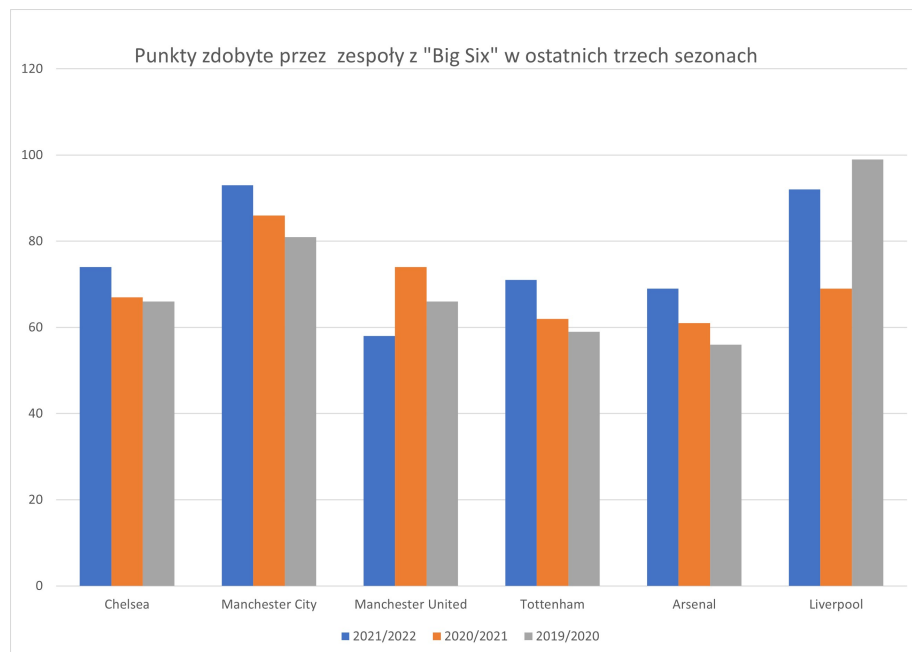
Wyniki ankiety przedstawione na wykresie skrzynkowym:



Obserwacje:

Rozstęp międzykwartyłowy wynosi 4. Pojawia się jednak sporo obserwacji odstających, co sugeruje, że odczytanie danej wartości z wykresu nie było proste.

Następnie, prezentowany był drugi wykres:

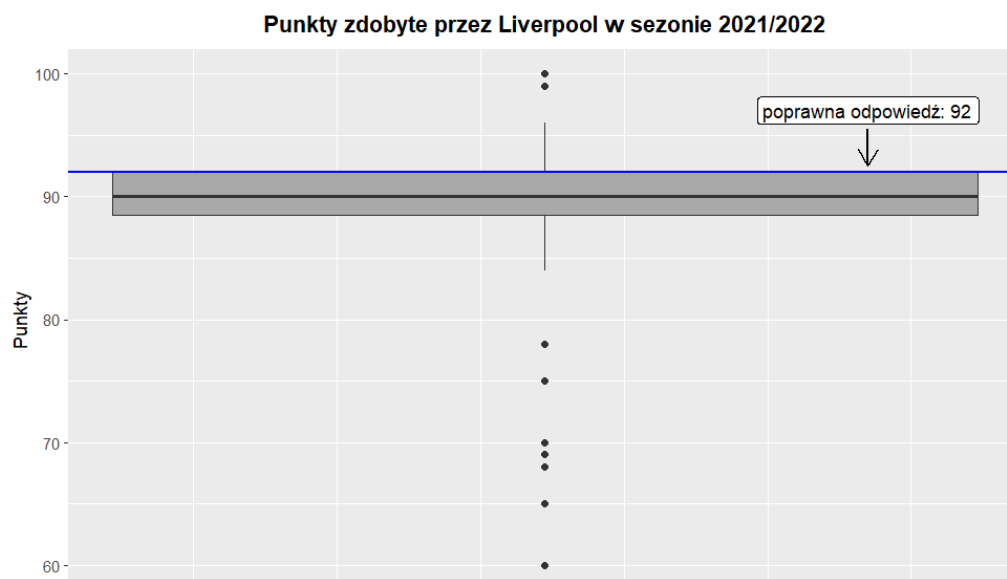


Ankietowani znów zostali poproszeni o odpowiedzi na wcześniejsze pytania.

1. Tym razem odpowiedzi wyglądały następująco:

2019/2020	2020/2021	2021/2022
59	1	8

2. Ponownie należało wskazać, ile dokumentów prawa jazdy wydano we wrześniu.
Wyniki ankiety przedstawione na wykresie skrzynkowym:



Obserwacje:

W przypadku obu pytań znacznie poprawiła się skuteczność ankietowanych, a odpowiedzi w większości były prawidłowe lub zbliżone do prawidłowych.

Na końcu poprosiliśmy o ocenę przejrzystości i czytelności obu wykresów.

4.1 Wyniki ankiety

Wykres 3D	Wykres 2D
2	66

Wykres 2D jest oceniany jako **czytelniejszy** przez 97% osób.

5. Podsumowanie i wnioski

Przeprowadzone eksperymenty pokazują jak duże znaczenie ma dobieranie dobrego typu wykresu do prezentowanych danych.

Zgodnie z wynikami ankiety widzimy, że w przypadku wykresu kołowego, trudno jest dostrzec różnicę między zbliżonymi do siebie wielkością wycinkami. Problem ten nie istnieje w przypadku wykresu słupkowego.

Wykresy 3D przez kąt widzenia oraz nienaturalną perspektywę w której są ukazane słupki, prowadzą do nieprawidłowej oceny ich wysokości i w konsekwencji błędnego odczytania danych. Dużo lepiej sprawdzają się wykresy 2D, które w dużej mierze eliminują wyżej opisane problemy.