

Techniki Wizualizacji Danych

Praca domowa 3

Mikołaj Piórczyński

24 listopada 2021

1 Wprowadzenie

Niniejszy raport został przygotowany na potrzeby pracy domowej nr. 3 z przedmiotu Techniki Wizualizacji Danych na kierunku Inżynieria i Analiza Danych w semestrze zimowym 2021-22 na Politechnice Warszawskiej, której celem było przeprowadzenie eksperymentu sprawdzającego czy określone problemy z czytaniem danych wciąż występują. Raport przedstawia wyniki przygotowanego eksperymentu.

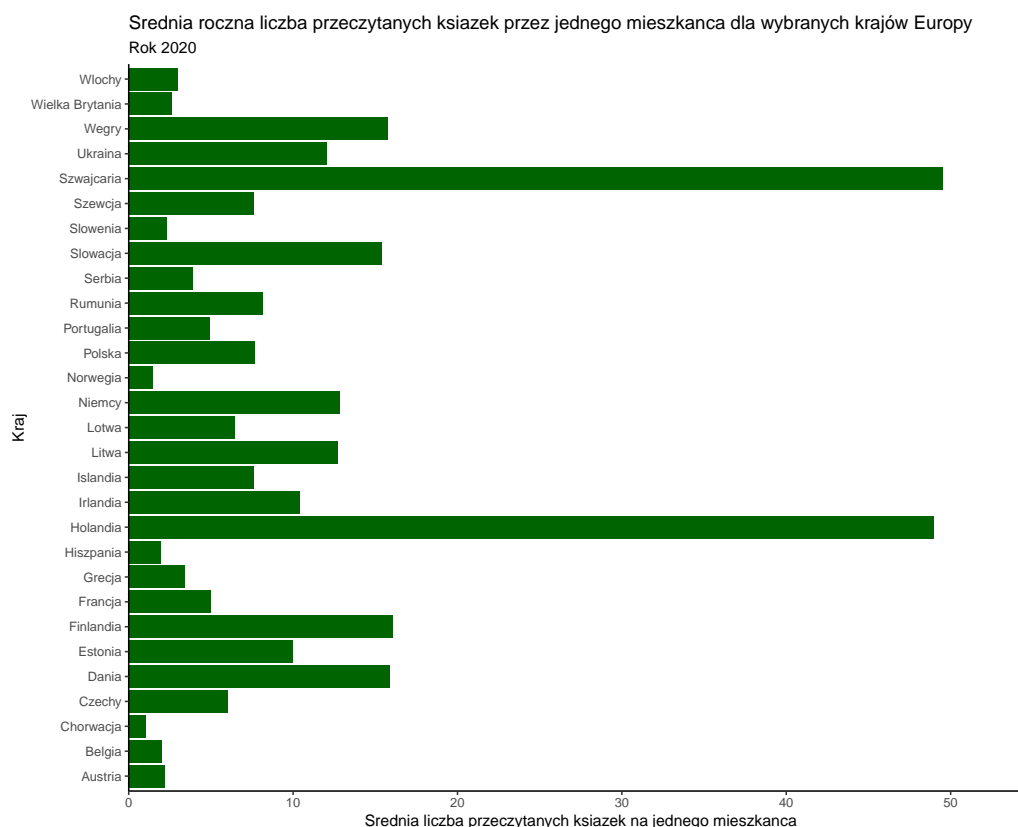
2 Eksperyment

2.1 Opis

Problem: Sortowania kolejności słupków względem ich długości pozytywnie wpływa na poprawne odczytywanie wartości z wykresu w przypadku wykresów słupkowych.

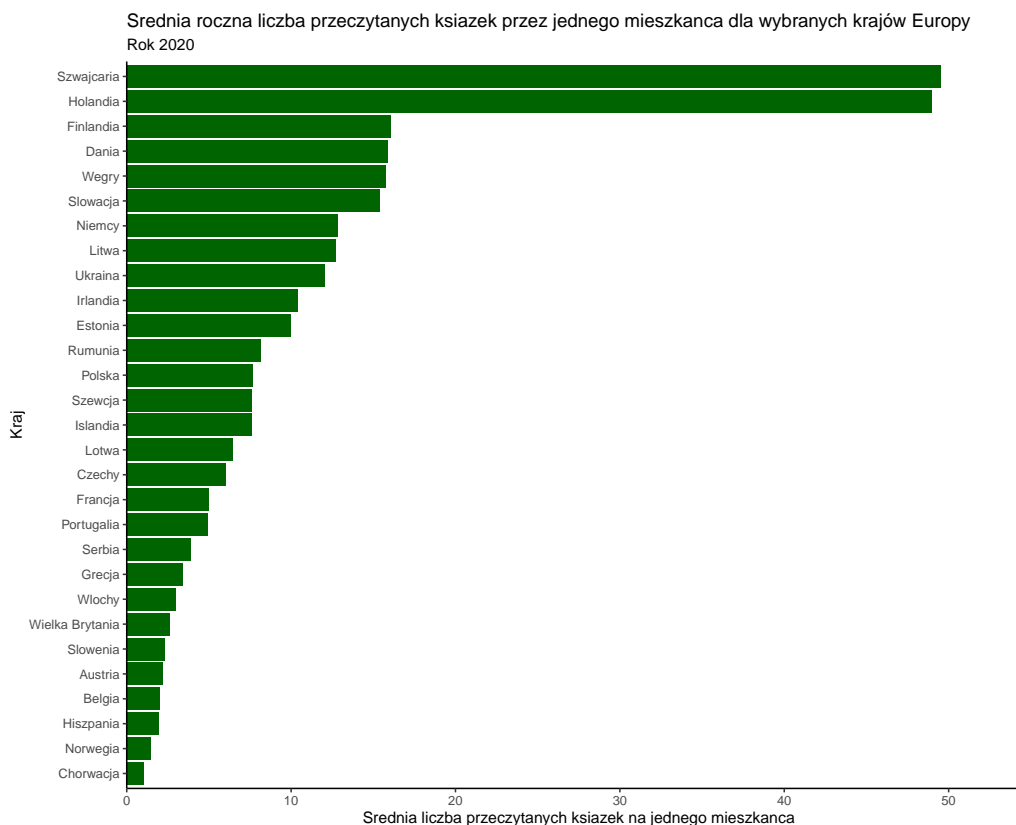
Celem eksperymentu było sprawdzenie wpływu sortowania kolejności słupków względem ich długości na odczytywanie wartości z wykresu w przypadku wykresów słupkowych. Specjalnie na potrzeby eksperymentu zostały stworzone sztuczne dane dotyczące średniej rocznej liczby przeczytanych książek przez jednego mieszkańca dla wybranych krajów Europy w roku 2020, przedstawione w tabeli 1.

Na ich podstawie następnie zostały przygotowane dwa wykresy słupkowe. Pierwszy – nie posiadający posortowanych słupków względem wartości, którą przedstawiają – grupa badana (rys. 1) oraz drugi – posiadający posortowane słupki – grupa kontrolna (rys. 2).



Rysunek 1: Grupa badana - wykres słupkowy nieposortowany

Oba wykresy przedstawiały te same dane oraz miały tę samą estetykę. W celu stworzenia realnej trudności przy odczytywaniu wartości z wykresów dane zawierały dużą liczbę kategorii (słupków) oraz poszczególne słupki potrzebne do udzielenia odpowiedzi na pytania przyjmowały nieznacznie różniące się wartości. Dodatkowo dla utrudnienia nie umieszczono na wykresach linii pomocniczych.



Rysunek 2: Grupa kontrolna - wykres słupkowy posortowany

W celu przeprowadzenia eksperymentu została przygotowana ankieta w postaci formularza Google, która została rozesłana do badanych drogą elektroniczną¹. Ankieta składała się z następujących pytań:

1. W którym kraju średnia roczna liczba książek przeczytanych na jednego mieszkańca była najwyższa w roku 2020?
2. W którym kraju średnia roczna liczba książek przeczytanych na jednego mieszkańca była najniższa w roku 2020?
3. Który kraj uplasował się na trzecim miejscu pod względem średniej rocznej liczby książek przeczytanych na jednego mieszkańca w roku 2020?

¹<https://forms.gle/j5DQrPqfNWdeLb2R6>

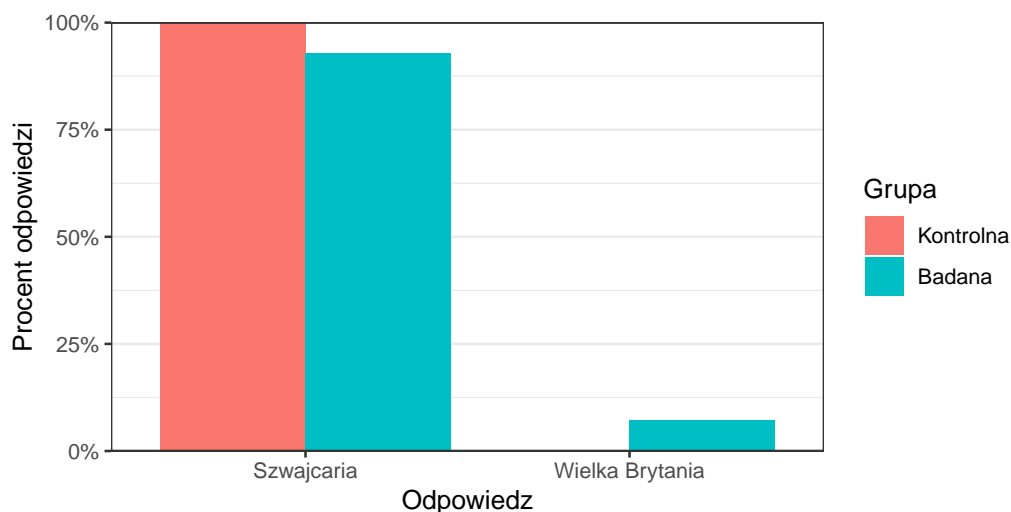
4. Porównaj, w którym kraju średnia roczna liczba przeczytanych książek na jednego mieszkańca była wyższa w 2020 roku - w Polsce czy na Islandii? A może te wartości były równe?
5. Jak oceniasz poziom trudności odczytywania danych z powyższego wykresu?
 - Łatwy
 - Średni
 - Trudny

Ankietowani losowali wykres, z którego będą odczytywać dane, tym samym przypisując się do grupy kontrolnej lub badanej, dzięki temu żaden z respondentów nie widział tych samych danych dwukrotnie. Przyjęta metodyka miała zapobiec ewentualnym wpływom odczytywania danych z jednego wykresu na odpowiedzi w przypadku drugiego. Eksperyment został przeprowadzony na grupie 27 osób, przy czym w grupie badanej znalazło się 14 osób, natomiast w grupie kontrolnej 13.

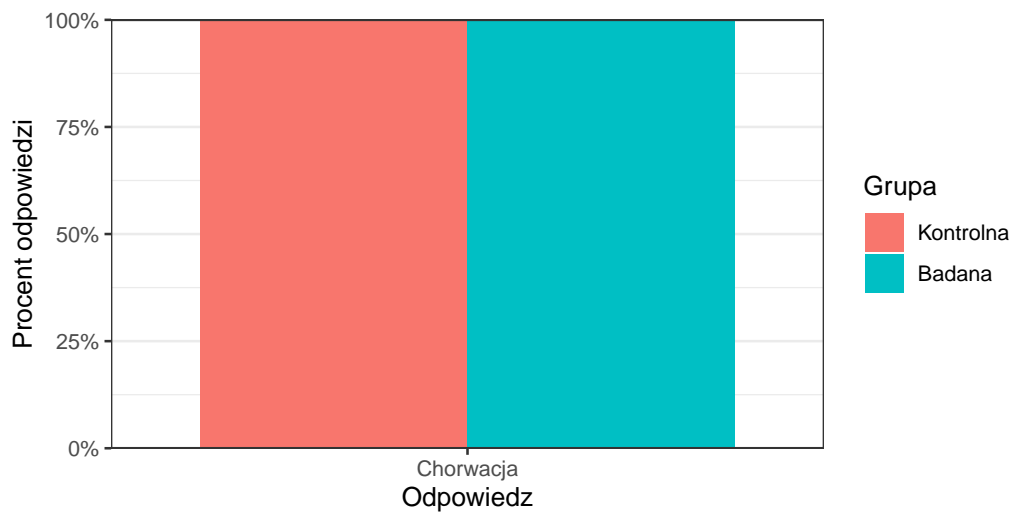
2.2 Wyniki

Uzyskane w eksperymencie dane zostały oparacowane w postaci wykresów przedstawionych na rys. 3, 4, 5, 6 oraz 7. Ze względu na dużą liczbę możliwych odpowiedzi w pytaniach 1-3 na wykresach zostały przedstawione tylko kraje, które zostały wskazane przez co najmniej 1 osobę.

Na pytania 1-3 w grupie kontrolnej otrzymano 100% poprawnych odpowiedzi. W przypadku grupy badanej otrzymano nieznacznie większy rozrzut, ponieważ w pytaniach 1 i 3 pojedynczy ankietowani wskazali kraje nie będące poprawną odpowiedzią. Zauważmy jednak, że zarówno w pytaniu 1 jak i w pytaniu 3 odpowiednio Wielka Brytania i Hiszpania nie były nawet blisko wartości odpowiadających najdłuższemu i trzeciemu najdłuższemu słupkowi. Może to więc sugerować udzielanie losowej odpowiedzi lub złe zrozumienie pytania, zwłaszcza w przypadku pytania 3, ponieważ Hiszpania była trzecim najkrótszym słupkiem. W przypadku ankietowanych wskazujących w grupie kontrolnej Węgry w pytaniu 3, odpowiedź taka jest całkowicie zrozumiała, ponieważ wartość odpowiadająca Węgrom różniła się zaledwie o 0.28. od poprawnej odpowiedzi.

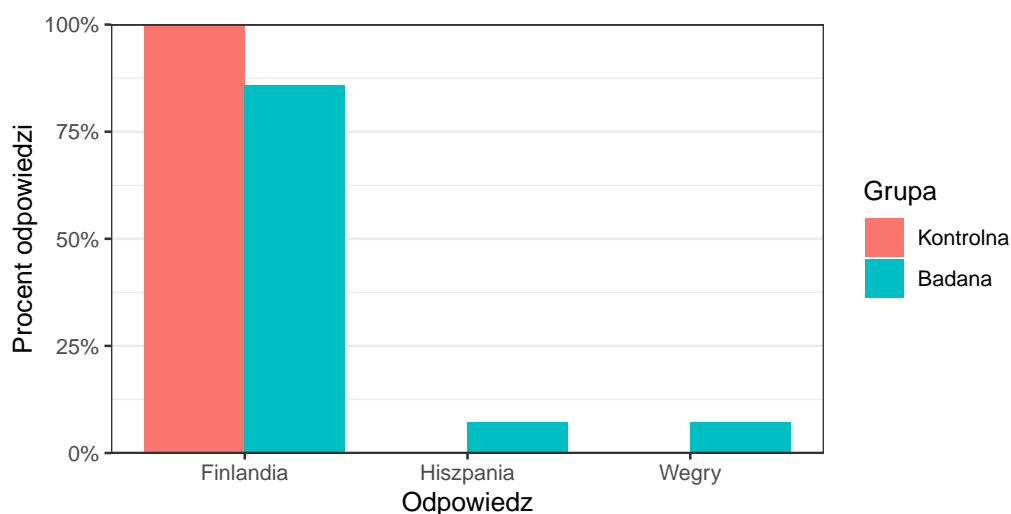


Rysunek 3: Wyniki dla pytania 1. **Prawidłowa odpowiedź:** Szwajcaria.

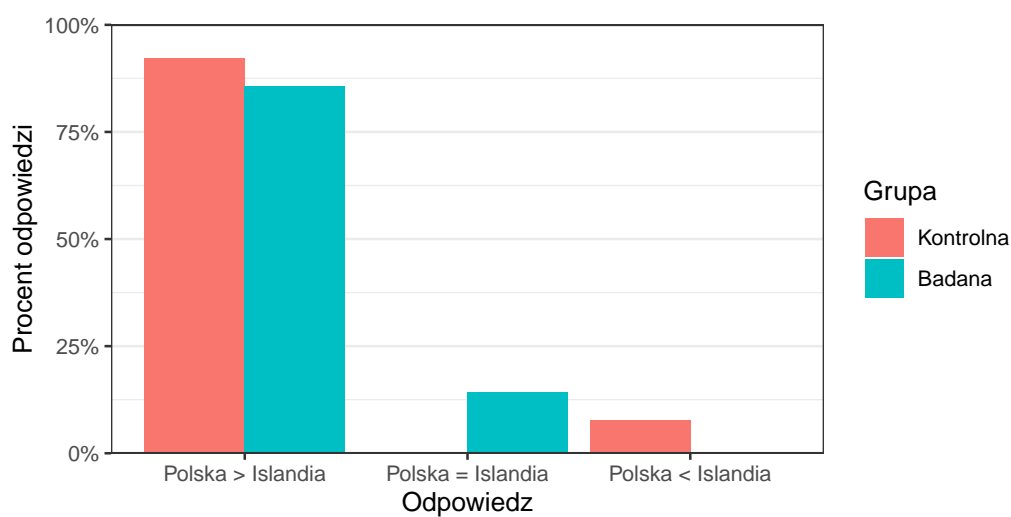


Rysunek 4: Wyniki dla pytania 2. **Prawidłowa odpowiedź:** Chorwacja.

Zauważmy, że nieznaczna część badanych z grupy kontrolnej udzieliła w pytaniu 4 odpowiedzi przeciwnej względem prawidłowej. Mogło to być spowodowane pomyleniem zwrotu zanku mniejszości lub nie zrozumieniem pytania. W przypadku grupy badanej kilka osób wskazało równość długości słupków mimo nieznacznej różnicy ich długości, co było wynikiem oczekiwanym.

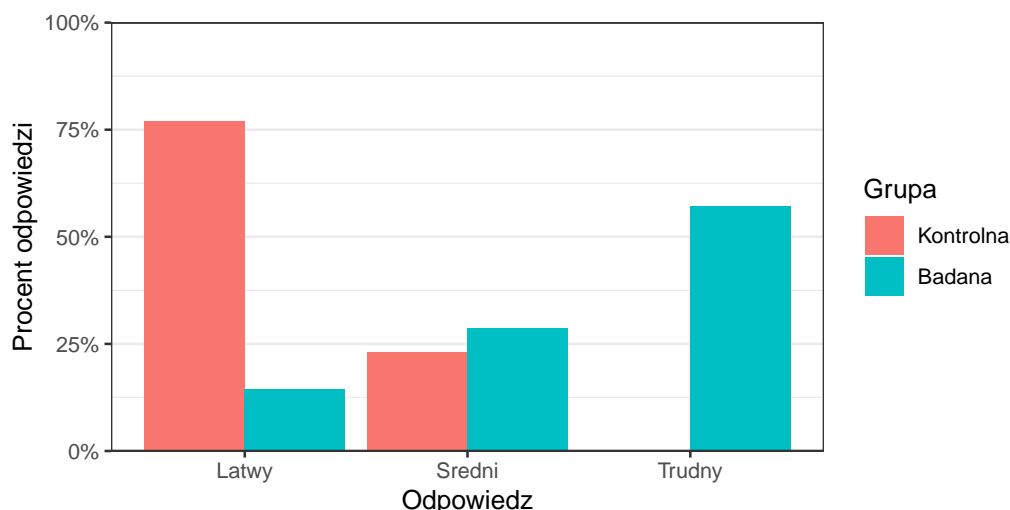


Rysunek 5: Wyniki dla pytania 3. **Prawidłowa odpowiedź:** Finlandia



Rysunek 6: Wyniki dla pytania 4. **Prawidłowa odpowiedź:** Polska > Islandia

W pytaniu 5 ponad 50% ankietowanych w grupie badanej odpowiedziało, że odczytywanie danych z wykresu nieposortowanego było trudne, trochę poniżej 25%, że było to średnio trudne oraz pozostała najmniejsza część, że było to łatwe. W przypadku grupy kontrolnej ponad 75% respondentów



Rysunek 7: Wyniki dla pytania 5

wskazywało, że odczytywanie danych z wykresu posortowanego było łatwe, a pozostała część, że było to średnio trudne. Widać zatem, że dla większości osób odczytywanie danych w przypadku posortowanych słupków było łatwe, natomiast nieposortowanych trudne.

3 Podsumowanie

Podsumowując, przeprowadzone badania nie wykazały znaczącego wpływu sortowania kolejności słupków względem ich długości na odczytywanie danych z wykresu. Ankietowani zarówno dobrze poradzili sobie z porównywaniem zbliżonych długości słupków sąsiadujących ze sobą jak i nie sąsiadujących ze sobą. W przypadku obu grup procent poprawnych odpowiedzi wynosił ponad 80. Przyczyn takiego wyniku można doszukiwać się m. in. w tym, że eksperyment nie był ściśle nadzorowany, przez co możliwe były takie sytuacje jak np. w przypadku wypełniania ankiety na smartfonie dopasowywanie końca wybranego słupka do krawędzi ekranu i porównywanie z końcem innego czy nawet mierzenie długości słupków linijką.

To co, natomiast wykazały badania to znacząco większą łatwość odczytywania danych i porównywania wartości dla wykresu posortowanego niż niepo-

sortowanego. Umacnia to przyjętą "dobrą praktykę" dokonywania sortowania słupków (o ile słupki nie posiadają jakiegoś innego wewnętrznego porządku jak np. lata). Działanie to zwiększa czytelność wykresu i ułatwia jego zrozumienie.

	name	value
1	Austria	2.20
2	Belgia	2.00
3	Chorwacja	1.05
4	Czechy	6.05
5	Dania	15.85
6	Estonia	10.00
7	Finlandia	16.05
8	Francja	5.01
9	Grecja	3.39
10	Hiszpania	1.95
11	Holandia	49.00
12	Irlandia	10.41
13	Islandia	7.58
14	Litwa	12.71
15	Łotwa	6.46
16	Niemcy	12.84
17	Norwegia	1.47
18	Polska	7.69
19	Portugalia	4.92
20	Rumunia	8.15
21	Serbia	3.91
22	Słowacja	15.39
23	Słowenia	2.32
24	Szwajcaria	49.50
25	Szwecja	7.60
26	Ukraina	12.04
27	Węgry	15.77
28	Wielka Brytania	2.64
29	Włochy	3.01

Tabela 1: Dane wykorzystane w eksperymencie