

# Praca domowa nr 3 + 4

Adrian Stańdo

24.10.2020

## Opis problemu

Wiele wykresów, które spotykamy na co dzień w mediach, mogą wprowadzać nas w błąd. Jest to spowodowane różnymi zabiegami, jednak jedną z najbardziej popularnych metod jest dobieranie takiego rodzaju wykresu, że trudno jest porównać przedstawione dane.

W pracy domowej zbadałem wpływ rodzaju wykresu na jakość odbieranych danych. Przebrałem trzy wykresy: wykres skumulowany, kołowy oraz słupkowy.

## Opis i źródło danych

Dane zawierają informacje o najbardziej popularnych kierunkach podróży wakacyjnych Polaków w 2019 r. Dane użyte do wykresów (z lekkimi modyfikacjami) oparłem na raporcie Polskiej Izby Turystyki. Ankietę przeprowadziłem przy pomocy narzędzia Microsoft Forms na znajomych, którzy nie studiują na IAD-dzie. Z tego powodu wyniki ankiety powinny najlepiej oddawać rzeczywistość, ponieważ Ci znajomi nie uczestniczyli w zajęciach z przedmiotu Techniki Wizualizacji Danych ani innych podobnych.

Link do ankiety: [kliknij tutaj](#).

Link do raportu Polskiej Izby Turystyki: [kliknij tutaj](#).

## Tabela danych

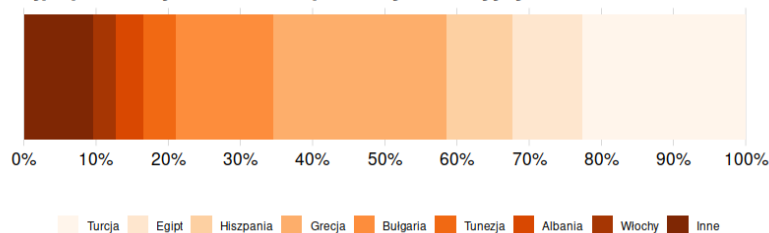
Kierunek	Procent wszystkich wyjazdów
Grecja	24.0
Turcja	22.6
Bułgaria	13.5
Egipt	9.7
Hiszpania	9.1
Tunezja	4.5
Albania	3.9
Włochy	3.1
Inne	9.6

## Zbudowane wykresy

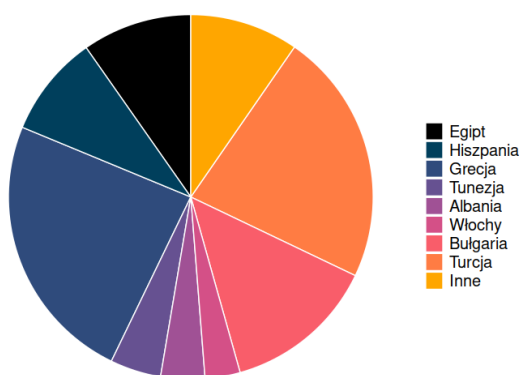
Poniżej zamieszczam zbudowane wykresy potrzebne do przeprowadzenia ankiety.

UWAGA: Wykresy w tytułach mają różne lata (choć są to te same dane) - był to zabieg celowy, by osoba rozwiązująca ankietę nie sugerowała się poprzednimi odpowiedziami.

**Najpopularniejsze kierunki podróży wakacyjnych Polaków w 2017 r.**



**Najpopularniejsze kierunki podróży wakacyjnych Polaków w 2018 r.**



**Najpopularniejsze kierunki podróży wakacyjnych Polaków w 2019 r.**

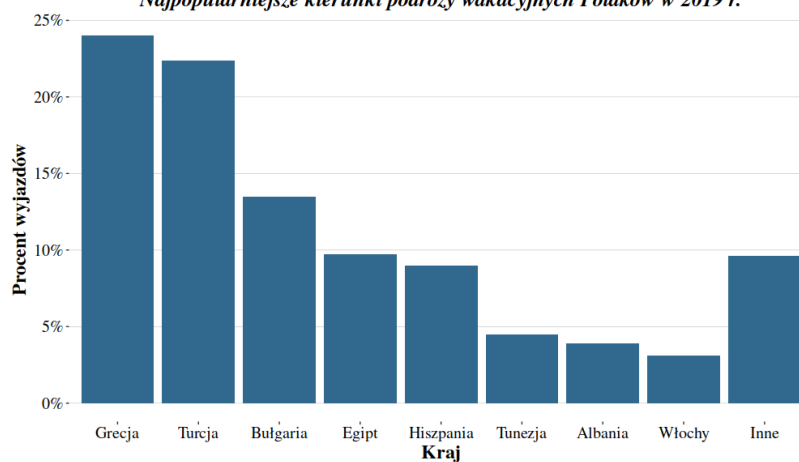


Figure 1: Wykresy użyte do ankiety

## Oceny wykresów

Pierwszym zadaniem w ankiecie było ocenienie każdego z wykresów w skali od 1 (wykres bardzo nieprzejrzysty) do 5 (wykres przejrzysty, najbardziej odpowiedni do przedstawienia danych). Odpowiedzi otrzymane w tym zadaniu przedstawiłem na wykresie poniżej.

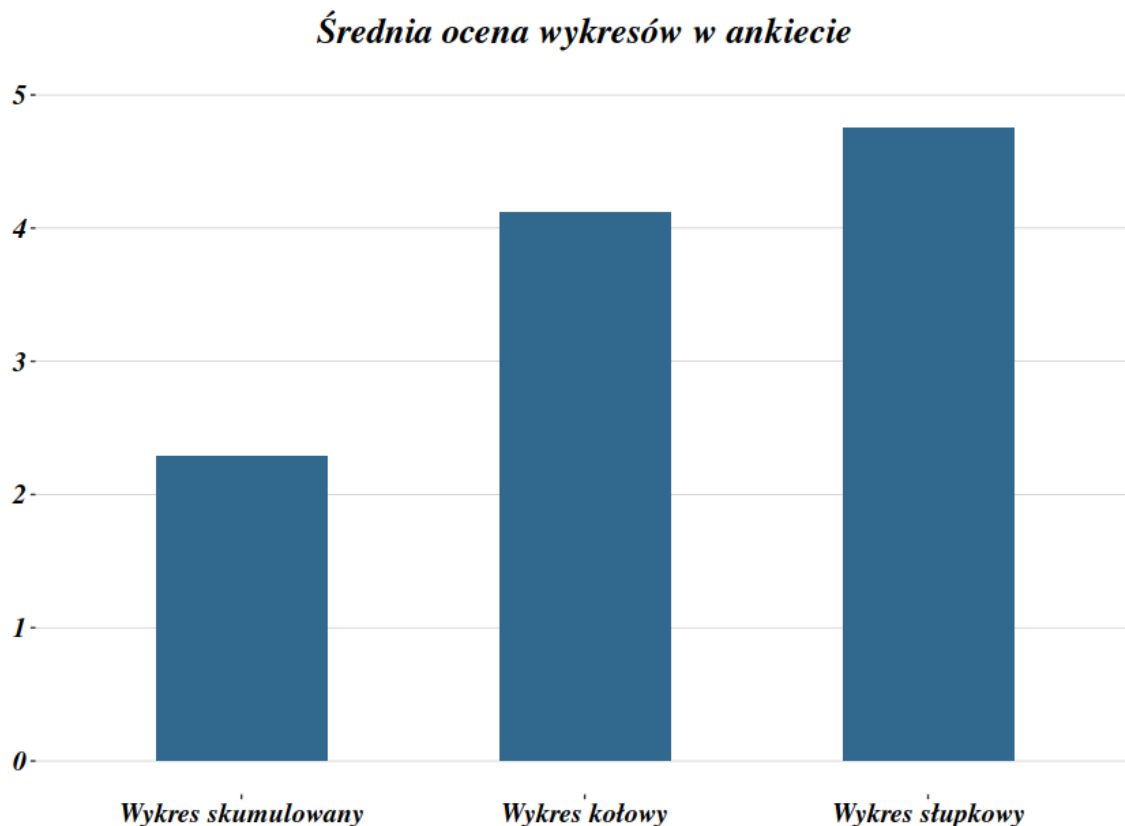


Figure 2: Ocena wykresów użytych do ankiety

Jak widać, najlepiej został oceniony wykres słupkowy (średnia ocena 4,76), niewiele gorzej wykres kołowy (średnia ocena 4,12) i najgorzej wypadł wykres skumulowany (średnia ocena 2.29).

## Odpowiedzi na zadane pytania

Kolejne zadania polegały na wskazaniu na każdym z wykresów państwa najczęściej wybieranego przez turystów oraz kraju, który znalazł się na czwartej pozycji. Wyniki tych pytań przedstawiłem na wykresie poniżej.

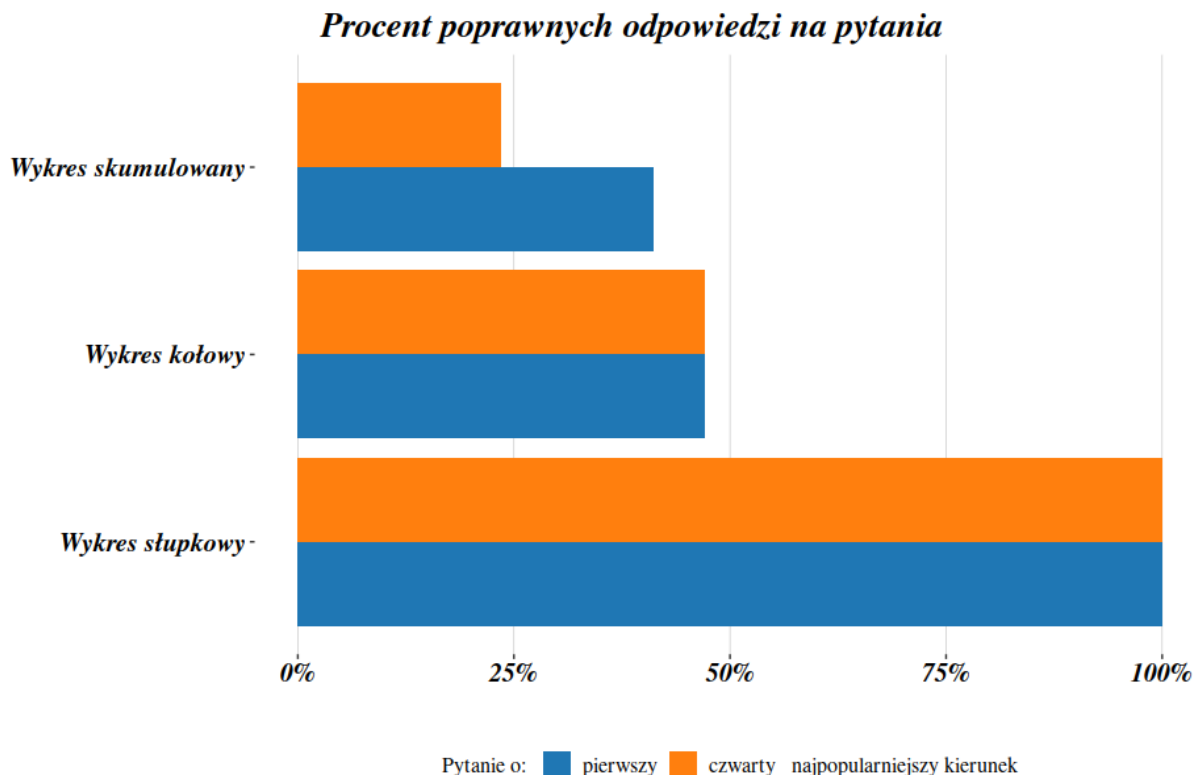


Figure 3: Odpowiedzi na pytania w ankiecie

Jak widać na wykresie, odpowiedzi na pytania zadane w przypadku wykresu słupkowego były w 100% poprawne. W przypadku wykresu kołowego - było to trochę mniej niż 50%. Co ciekawe, najczęstszymi niepoprawnymi odpowiedziami były te, których wartość była bliska poprawnej wartości. Dla pierwszego miejsca najpopularniejsze odpowiedzi to Grecja (pierwsza) i Turcja (druga), natomiast dla czwartego Egipt (czwarty) i pole Inne (piąte).

Najgorzej wypadł wykres skumulowany. O ile największy wynik był wskazywany poprawnie w około 40%, to czwarty wynik już tylko w około 20% odpowiedzi. Ciekawą obserwacją jest to, że czwarta wartość (Egipt) w przypadku wykresu kołowego była często mylona z polem Inne (piąte) i tylko raz podano odpowiedź Hiszpania (szósta). Natomiast w przypadku wykresu skumulowanego, na drugie pytanie najczęściej odpowiadano Hiszpania (szósta) oraz Egipt (czwarty). Tylko dwie z siedemnastu odpowiedzi były polem Inne (piąte). Najprawdopodobniej wynika to z tego, że pole Inne umieściłem na początku, z dala od bliskiego co do wartości Egiptu i Hiszpanii. Nawet kolor pola Innego, który był dużo bardziej wyrazisty, nie sprawił, że uczestnicy ankiety zwrócili na nie uwagę.

## Wnioski i podsumowanie

Z przeprowadzonej ankiety wynika, że najlepszym sposobem wizualizacji danych jest wykres słupkowy. Nie dosyć, że jest on wskazywany jako najbardziej przejrzysty, to dodatkowo jako jedyny zapewnił dużą (a nawet stuprocentową) poprawność odczytywanych danych.

Wykres kołowy - tak bardzo lubiany przez większość ludzi - dostał bardzo wysoką ocenę. Według mnie, wyróżniał się on ładnie skomponowanymi barwami. Jednak poprawność odczytywanych danych wyniosła tylko około 50% - co jest słabym wynikiem. Dodatkowo, najczęściej mylono bliskie sobie wartości. Z tego powodu wnioskuję, że wykres kołowy nie nadaje się do przedstawienia bliskich sobie wartości.

Najgorzej wypadł wykres skumulowany, na którym nie tylko strasznie trudno jest odczytać związki między wielkościami, to jeszcze nie zauważamy oddalonych danych - mimo podobnej wartości. Z tego powodu ten rodzaj wykresu jest bardzo rzadko używany - jednakże od czasu do czasu spotykamy się z nim.