

# Wstęp do eksploracji danych

## Praca domowa 3

Oliwia Trzcińska, Jakub Niemyjski

Celem tego raportu jest opis przebiegu i wyników badania sprawdzającego, czy i w jaki sposób niestosowanie się do „dobrych praktyk” tworzenia wykresów wpływa na ich odbiór oraz powoduje problemy z czytaniem danych.

Grupę badanych stanowiło 17 osób z grupy wiekowej 19 – 60. Każdej z osób zostały zadane 3 pytania:

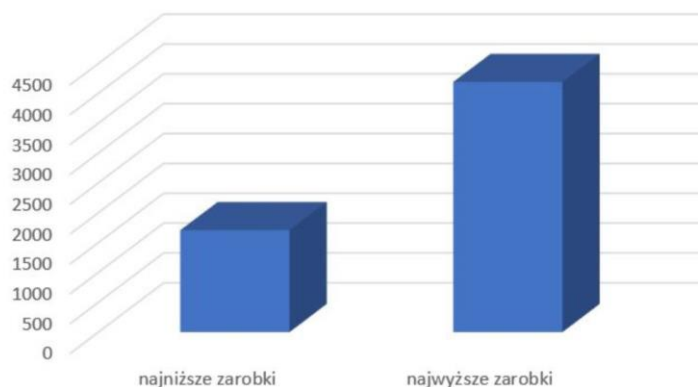
1. Czy odpowiedź większości ankietowanych na zadane pytanie była pozytywna? Jaki był stosunek ilości odpowiedzi pozytywnych do negatywnych?



<http://smarterpoland.pl/index.php/2020/12/plebiscyt-na-najgorszy-wykres-roku-2020/>

2. Co przedstawia wykres? Ile wynosiły najniższe i najwyższe zarobki?

Wykres 2. Różnicowanie dochodów najniższych i najwyższych wśród ankietowanych



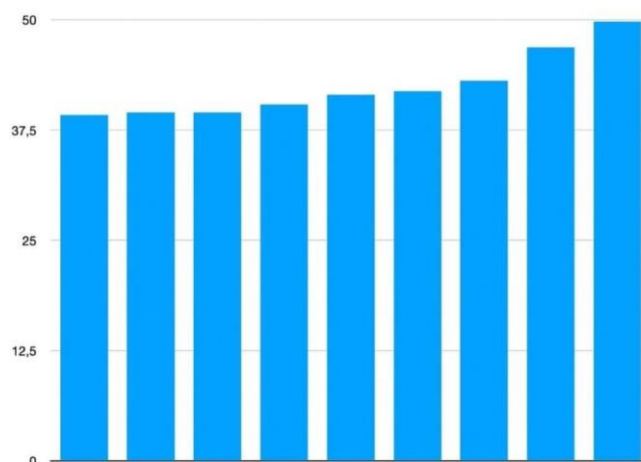
<http://smarterpoland.pl/index.php/2018/12/najgorszy-wykres-2018/>

3. Jakie jest tempo wzrostu kwoty części oświatowej subwencji ogólnej w latach 2012 – 2020? Czy różnice między wartościami w kolejnych latach są znacząco duże?



[https://www.facebook.com/janinadaily/photos/a.1524649467770881/2836063543296127/?type=3&source=48&paipv=0&eav=Afbdpn4GUvv\\_gURf35Rb\\_H4F5ujXh4aEZP2RMo\\_BqipXBG3IDug-EUny90iSTG5oJmg&\\_rdr](https://www.facebook.com/janinadaily/photos/a.1524649467770881/2836063543296127/?type=3&source=48&paipv=0&eav=Afbdpn4GUvv_gURf35Rb_H4F5ujXh4aEZP2RMo_BqipXBG3IDug-EUny90iSTG5oJmg&_rdr)

Dla porównania, oto słupki prezentujące te same dane, ale przeskalowane w poprawny sposób:



[https://www.facebook.com/janinadaily/photos/a.1524649467770881/2836063543296127/?type=3&source=48&paipv=0&eav=Afbdpn4GUvv\\_gURf35Rb\\_H4F5ujXh4aEZP2RMo\\_BqipXBG3IDug-EUny90iSTG5oJmg&\\_rdr](https://www.facebook.com/janinadaily/photos/a.1524649467770881/2836063543296127/?type=3&source=48&paipv=0&eav=Afbdpn4GUvv_gURf35Rb_H4F5ujXh4aEZP2RMo_BqipXBG3IDug-EUny90iSTG5oJmg&_rdr)

### Podsumowanie wyników przeprowadzonego eksperymentu:

- Wykres 1

Czas patrzenia na wykres został ograniczony. W związku z tym odpowiedzi na pytanie, która odpowiedź (pozytywna czy negatywna) występowała częściej kształtowały się następująco:

ODPOWIEDŹ	LICZBA OSÓB
Pozytywna	5
Negatywna	12

Okazało się, że tylko 29,4% osób dokładnie przeczytało legendę i tytuł wykresu. Pozostali polegali na swojej intuicji dotyczącej tego, jak odbierają kolory czerwony oraz zielony i przez to zostali wprowadzeni w błąd.

Na pytanie o stosunek liczby jednych odpowiedzi do drugich zdecydowanie przeważała teza, że odpowiedzi „czerwonych” było około 3 razy więcej niż „zielonych”. Poprawna odpowiedź na to pytanie to 1,67, czyli około 2 razy więcej. Przez przedstawienie danych na wykresie 3D zostało stworzone wrażenie iluzji, co utrudniło poprawne odczytanie wartości.

- Wykres 2

Niemal połowa, bo aż 7 osób miało problem z określeniem, czego dotyczy wykres. Okazuje się, że słupki pokazują tylko tyle, że najniższe zarobki są niższe od najwyższych. Przedstawianie tych danych na wykresie nie miało sensu.

Pytanie o wartości, jakie są przedstawione na wizualizacji również nie było proste. Ich odczytanie zajęło każdemu zdecydowanie więcej czasu niż powinno, ponieważ wymagało dużego skupienia uwagi. Ponadto nawet mimo tego 4 osoby podały błędne przedziały liczbowe jako odpowiedź. To kolejny przykład pokazujący, że przedstawianie danych na wykresie 3D jest kłopotliwe dla odbiorców, ponieważ występują problemy z poprawnym odczytywaniem kątów.

- Wykres 3

Wszyscy pytani odpowiedzieli, że w kolejnych latach wartości wzrastają. I to była prawda. Jednak tylko 3 osoby zauważyły, że skala na osi OY nie zaczyna się od zera, przez co wielkości różnic są w rzeczywistości znacznie mniejsze. Pozostali uważali, że wzrost w stosunku do poprzedniego roku jest dość szybki, szczególnie jeśli porównujemy lata po 2015, czyli pomarańczowe słupki. Po pokazaniu drugiej wizualizacji wszyscy zgodnie stwierdzili, że wzrost wartości nie jest tak spektakularny jak mógł się wydawać na początku. Na przykładzie tej wizualizacji można wysnuć wniosek, że złe skalowanie wykresu przynosi oczekiwane przez jego twórców skutki. Znacząco zaburza odczytywanie danych oraz nakierowuje odbiorców na wysnuwanie tez niekoniecznie zgodnych z prawdą, ale mających pokazać to, co chcą oni zobaczyć.

## **Wnioski:**

Niestosowanie się do znanych zasad poprawnego tworzenia wykresów faktycznie powoduje liczne problemy z odczytywaniem danych. Aby ich uniknąć oraz być w pełni uczciwym w stosunku do odbiorcy, należy stosować się do ustalonych reguł.