

Raport - Praca domowa nr 3 i 4

Patryk Świątek

1 Opis i cel eksperymentu

Celem mojego eksperymentu jest:

- * pokazanie wpływu zmiany skali wykresu na jego interpretację przez czytelnika,
- * sprawdzenie, który typ skalowania osi, odbiorcy postrzegają jako lepszy,
- * porównanie wyboru bardziej efektywnego skalowania w poszczególnych wykresach przy wyciąganiu na ich podstawie wniosków przez ankietowanych, do wyboru ekspertów w dziedzinie analizy danych i autorów importowanych przeze mnie wykresów.

2 Ankiety

2.1 Opis ankiet

Przeprowadziłem ankietę składającą się z siedmiu pytań (poszczególne pytania przytoczone zostaną przy omówieniu wyników).

- Trzy z nich dotyczą ogólnej wiedzy i poglądów na temat iluzji i związanej z nią manipulacji w wizualizacji danych.
- Pozostałe pytania dotyczą porównania konkretnych wykresów i ich interpretacji.

Na każde z pytań odpowiedziało 28 osób. Żadna z tych osób nie ma zawodowego doświadczenia związanego z analizą i wizualizacją danych.

Link do ankiety: [Link](#)

2.2 Wyniki ankiet

- **Pytanie 1**

Treść pytania: Czy uważasz, że możemy manipulować faktycznymi danymi w taki sposób, żeby czytelnik wyciągnął z nich pożądane przez nas wnioski?

Możliwe odpowiedzi: Tak / Nie

Wyniki: Na to pytanie 100% ankietowanych udzieliło odpowiedzi pozytywnej.

- **Pytanie 2**

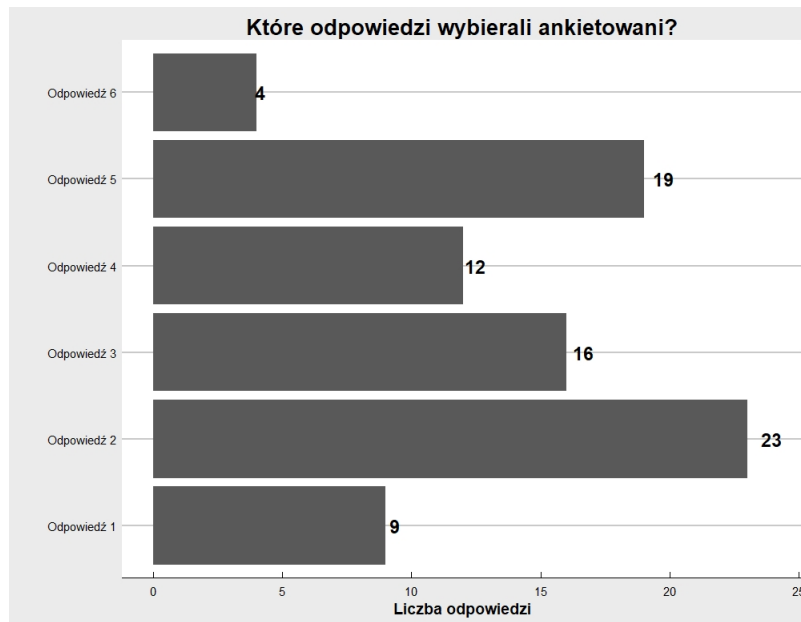
Treść pytania: W jaki sposób możemy nimi manipulować? (Wybierz 3 najlepsze odpowiedzi)

Możliwe odpowiedzi:

1. Zmieniając dane (np. zaokrąglając dane liczbowe w odpowiedni sposób).
2. Pokazując tylko te dane, które pasują do pożądanego przez nas wyniku.
3. Zmieniając typ wykresu na taki, żeby czytelnik nie mógł z niego zbyt wiele wyczytać.

4. Pokazując fragment tabeli wyłącznie z najbardziej pożądanymi przez nas wynikami.
5. Zmieniać skalę wykresu tak, aby zatrzeć lub sztucznie pokazać na nim pewne różnice.
6. Dodać niepotrzebne dane tak, aby czytelnik nie zwrócił uwagi na wyniki, które są przez nas niepożądane.

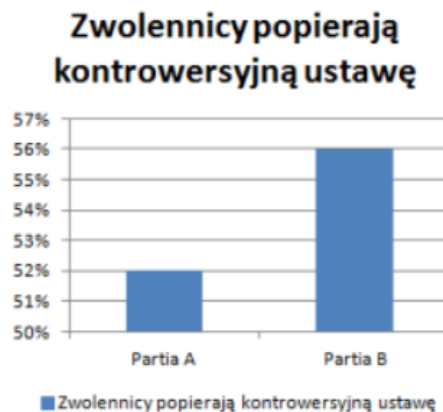
Wyniki:



Na wykresie widzimy, że najwięcej ankietowanych uważa, że możemy manipulować wykresami pokazując na nich tylko pożądaną przez nas dane, odpowiednio zmieniając ich skalę lub modyfikując ich typy.

• Pytanie 3

Treść pytania: Popatrz na wykres. Jakbyś porównał poparcie kontrowersyjnej ustawy w obu partiach?



Możliwe odpowiedzi:

- W partii B jest znacząco większe niż w partii A.
- W partii B jest nieznacznie większe niż w partii A.
- Różnica w obu partiach jest praktycznie nierozróżnialna.
- W partii A jest większe niż w partii B.

Wyniki



• Pytanie 4

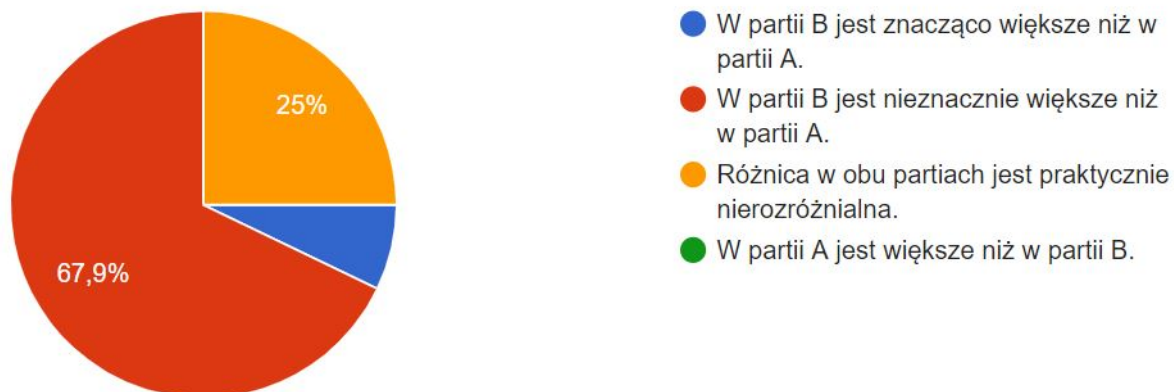
Treść pytania: Przyjrzyj się teraz następującemu wykresowi i odpowiedz jeszcze raz na to samo pytanie.



Możliwe odpowiedzi:

- W partii B jest znacząco większe niż w partii A.
- W partii B jest nieznacznie większe niż w partii A.
- Różnica w obu partiach jest praktycznie nierozróżnialna.
- W partii A jest większe niż w partii B.

Wyniki



Źródło wykresów z pytania 3 i 4: [Link](#)

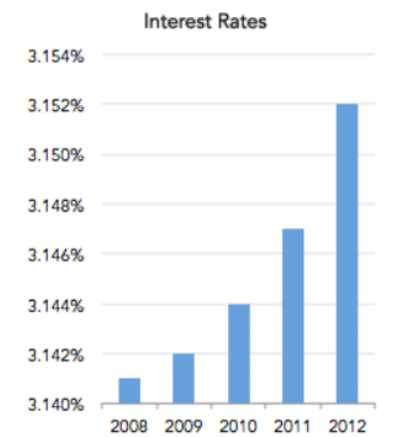
Wnioski na podstawie pytań 3 i 4:

- Widzimy, że na podstawie obu wykresów ankietowani w większości odpowiadali, że poparcie w partii B jest nieznacznie większe niż w partii A.
- Jednakże (mimo, że oba wykresy opierały się na tych samych danych) na podstawie pierwszego wykresu 25% ankietowanych stwierdziło, że w partii B poparcie jest znacząco większe niż w partii A. Z kolei, analizując drugi wykres, taka sama część stwierdziła, że różnica poparcia jest praktycznie niezauważalna.

• Pytanie 5

Treść pytania: Przyjrzyj się poniższym wykresom. Każdy z nich przedstawia identyczne dane dotyczące wielkości stopy procentowej w poszczególnych latach. Z którego z nich jesteśmy w stanie odczytać więcej informacji, wyciągnąć więcej wniosków?

Możliwe odpowiedzi:



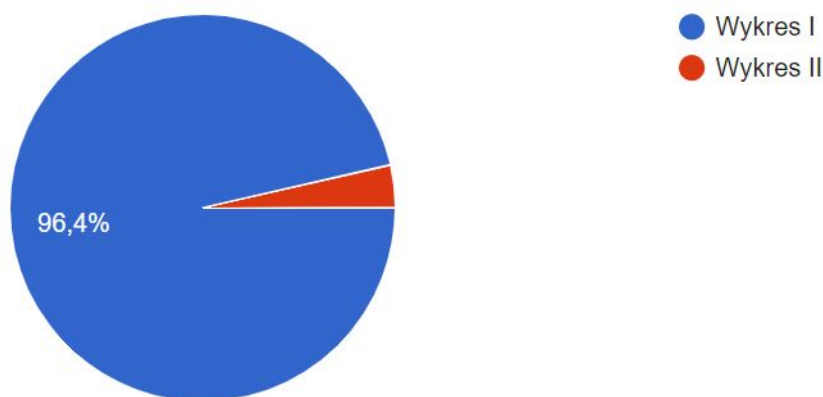
(a) Wykres I



(b) Wykres II

Źródło wykresów: [Link](#)

Wyniki



Wniosek i porównanie ze zdaniem ekspertów

Zdecydowana większość ankietowanych wybrała pierwszy wykres, czyli ten ze skalą na osi pionowej, która pozwala uwydatnić różnice w wartościach na wykresie. Według ekspertów i autorów tych wykresów są one jednak pokazane sztucznie i prowadzą do błędnych wniosków. Różnice pomiędzy wartościami zaznaczonymi na osi pionowej są nieznaczne, więc taki sposób tworzenia wykresów jest w tym przypadku zbędny.

• Pytanie 6

Treść pytania: Na podstawie obejrzanych wykresów i własnych przemyśleń, która zasada jest najbardziej trafna w poprawnej wizualizacji danych?

Możliwe odpowiedzi:

- Skala na osi pionowej w wykresie zawsze powinna zaczynać się od zera.
- Skala na osi pionowej w wykresie powinna zaczynać się od zera, jednakże są od tego wyjątki.
- Dobór skali na osi pionowej nie wpływa na interpretację wykresów przez czytelnika.
- Skala na osi pionowej zazwyczaj powinna zaczynać się od najmniejszej wartości tak, aby uwydatnić różnice w wartościach na wykresie.

Wyniki



Wniosek i porównanie ze zdaniem ekspertów

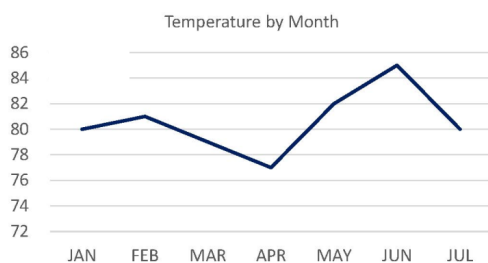
Zdaniem prawie 60% ankietowanych skala na osi pionowej zazwyczaj powinna zaczynać się od najmniejszej wartości tak, aby uwidocznić różnice w wartościach na wykresie. Zdanie to potwierdza wybór wykresu z pytania szóstego.

Jednakże ponad 30% osób odpowiedziało zgodnie ze zdaniem ekspertów w poprawnej wizualizacji danych, którzy uważają, że w większości przypadków, skala na osiach powinna zaczynać się od zera (poza nielicznymi wyjątkami).

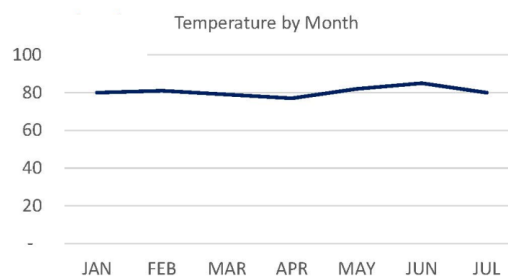
• Pytanie 7

Treść pytania: Na koniec przyjrzyj się poniższym wykresom średniej temperatury w zależności od miesiąca. Który z nich pozwala Ci odczytać więcej informacji?

Możliwe odpowiedzi:



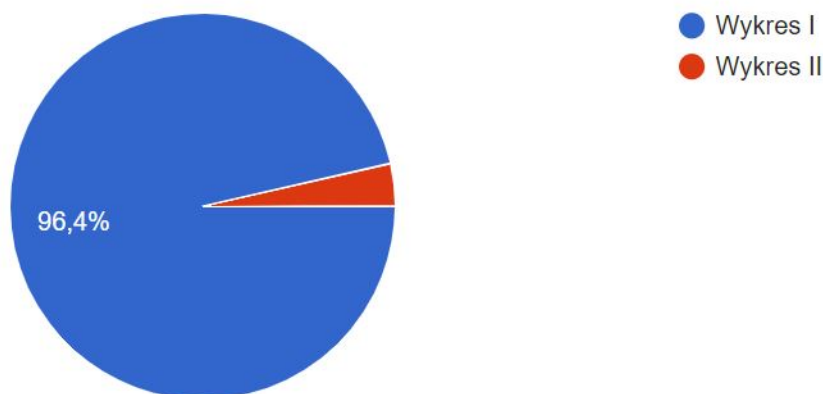
(a) Wykres I



(b) Wykres II

Źródło wykresów: [Link](#)

Wyniki



Wniosek i porównanie ze zdaniem ekspertów

Ankietowani odpowiedzieli zgodnie ze zdaniem zawodowych analityków, ponieważ w tym przypadku mamy wyjątek od poprawnej reguły z pytania nr 6. Przy wizualizacji tych danych ważne jest ukazanie różnic temperatur w ciągu roku, nawet jeśli są one niewielkie.

3 Podsumowanie i ogólne wnioski

- W większości ankietowani uważają, że pożądane jest pokazanie różnic w wykresach, co nie zawsze jest zgodne z powszechnie stosowanymi regułami wizualizacji danych.
- Większość ankietowanych uważa, że lepszym sposobem jest skalowanie w taki sposób, żeby wartości na osiach zaczynały się od najmniejszej wartości na wykresie, co również nie jest poprawnym postępowaniem w większości przypadków.
- Wnioski czytelników na podstawie przedstawiania tych samych danych na wykresach o innych skalowaniach różnią się, zatem wpływ doboru skali na wykresie jest znaczący.