

# Techniki Wizualizacji Danych - PD 3+4

Patryk Słowakiewicz, Dominik Pawlak

## Eksperyment 1

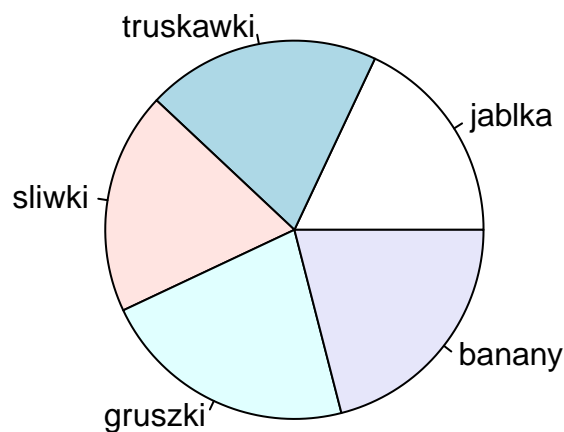
Oba eksperymenty zostały przeprowadzone na 38 osobach za pomocą ankiety, link do ankiety: <https://docs.google.com/forms/d/1PE3nCkCb8JMartvVs9vjaB5JYrI0P8YgM8mO6OZOHA/edit?usp=sharing>.

### Wstęp

W pierwszym eksperymencie postanowiliśmy zbadać jak czytelnik odczytuje dane z wykresów kołowych, kiedy wartości są do siebie zbliżone. Chcieliśmy sprawdzić umiejętność porównania wartości, nie odczytanie dokładnych wartości, dlatego nie umieszczaliśmy żadnych danych liczbowych. W tym celu stworzyliśmy wykres kołowy pokazujący ilość kupionych owoców. (Dane fikcyjne)

### Wykres kołowy

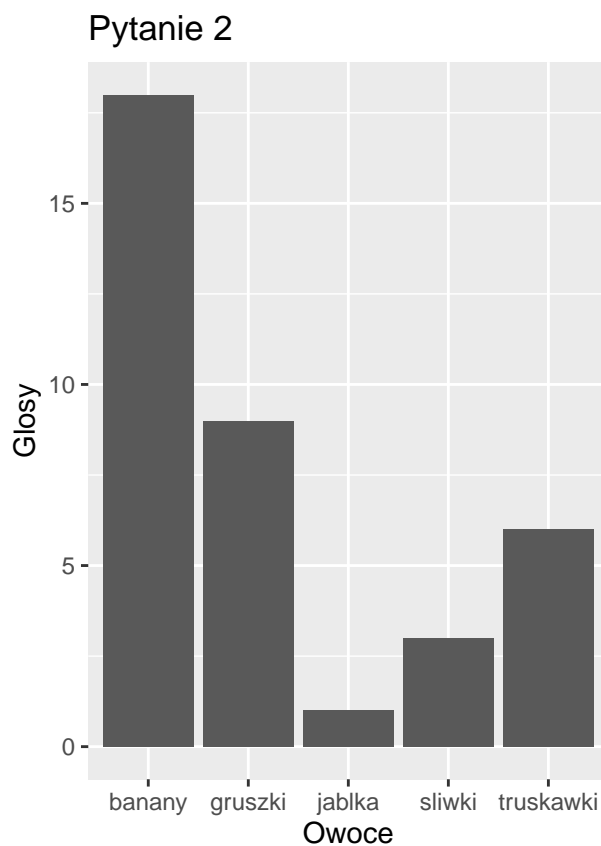
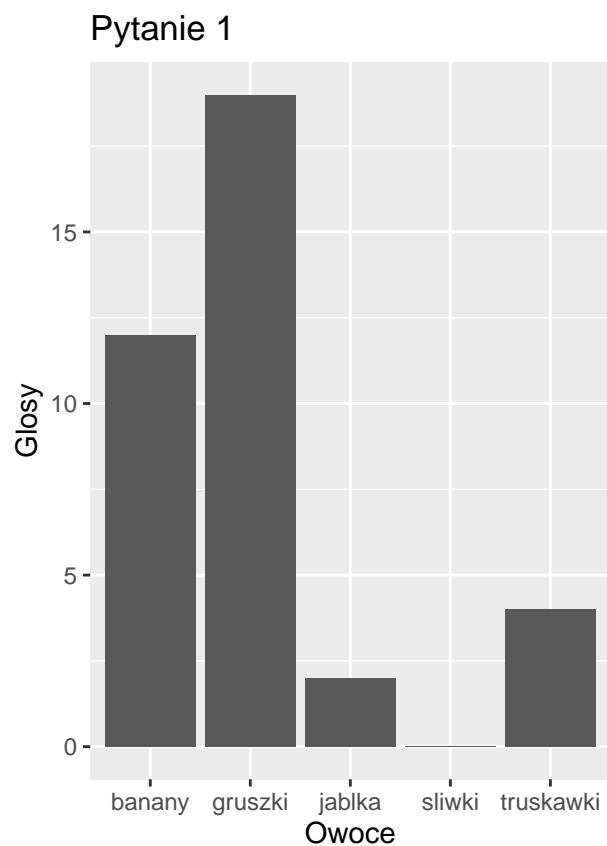
#### Ilość kupionych owoców



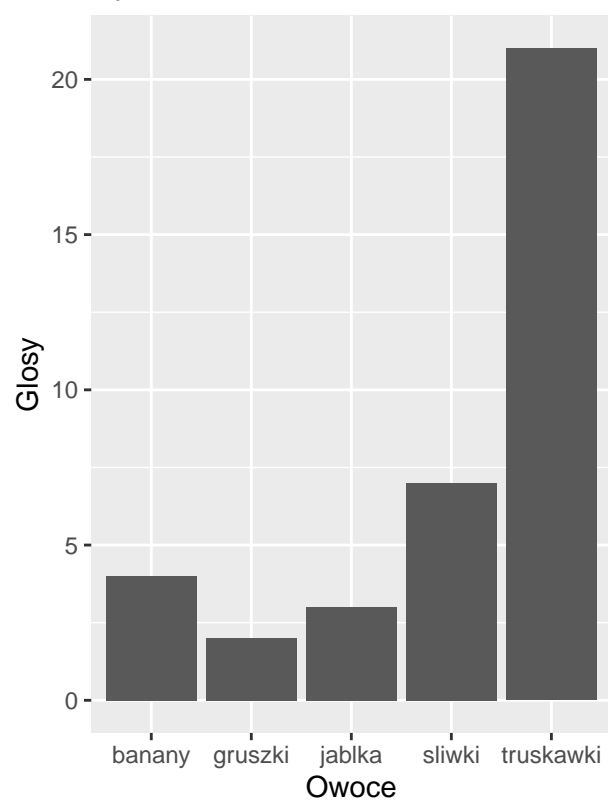
##	jablka	truskawki	sliwki	gruszki	banany
##	18	20	19	22	21

Zadalismy piec pytan brzmiacych “Który z owocow znajduje sie na i-tym miejscu pod wzgledem liczebnosci”. Wyniki nie byly zaskakujace.

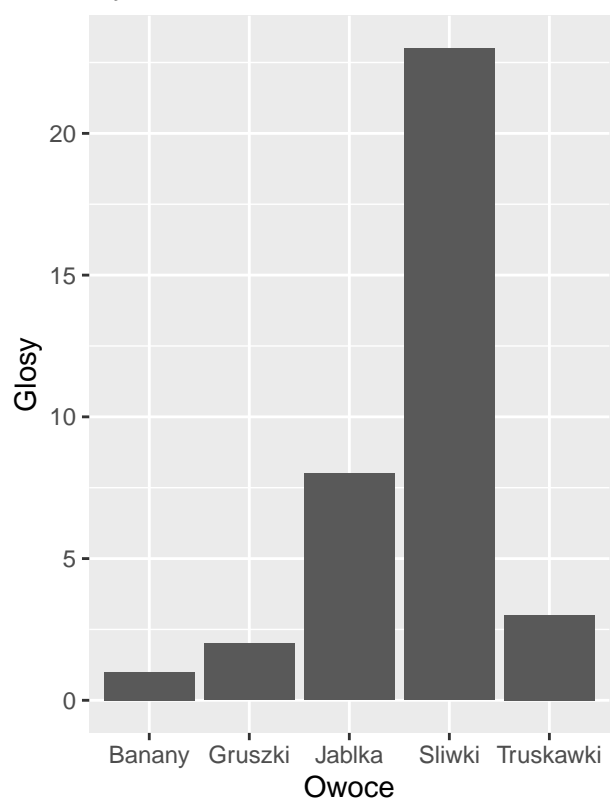
1. Na 1 pytanie 19 z 37 osob odpowiedzialo poprawnie. Byly tez 2 osoby, ktore wybraly najmniejsza wartosc jako najwieksza.
2. Na 2 pytanie 18 z 37 osob odpowiedzialo poprawnie.
3. Na 3 pytanie 21 z 37 osob odpowiedzialo poprawnie.
4. Na 4 pytanie 23 z 37 osob odpowiedzialy poprawnie.
5. Na 5 pytanie 26 z 37 osob odpowiedzialy poprawnie. Byly tez 3 osoby, ktory wybraly najwieksza wartosc jako najmniejsza.



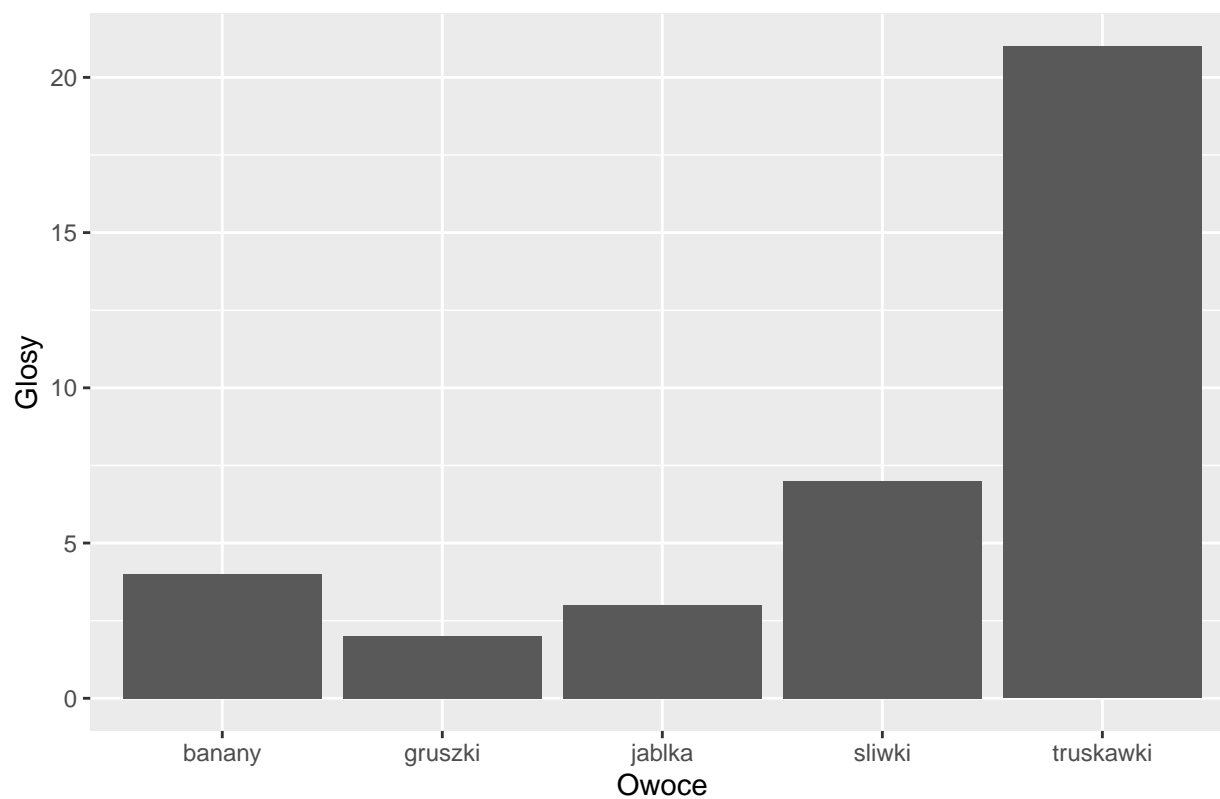
Pytanie 3



Pytanie 4

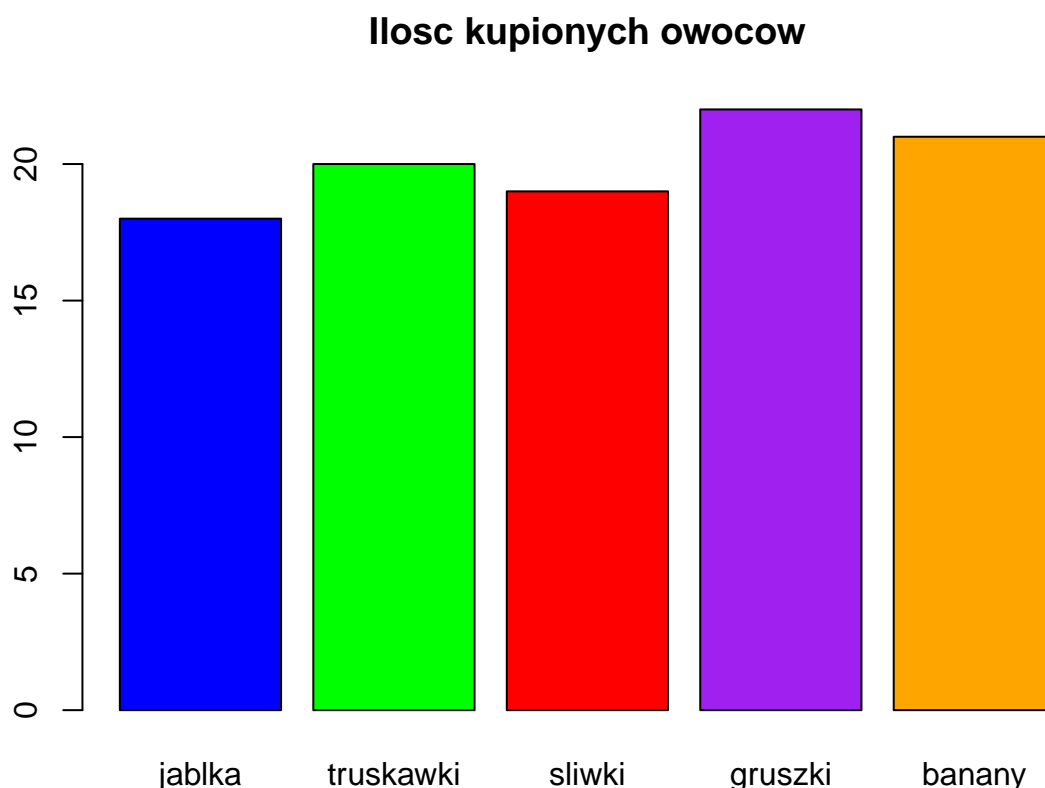


### Pytanie 5



### Wykres słupkowy

Następnie te same dane pokazaliśmy na wykresie słupkowym.



```
## numeric(0)
```

Poprosilismy ankietowanych, aby odczytali najwyzsza wartosc z powyzzszego wykresu. Na 37 osob, tylko 1 osoba nie dokonala poprawnego odczytu.

## Wnioski

Eksperyment pokazal, ze przy podawaniu danych o zblizonych wartosciach duzo lepiej sie sprawdzaja wykresy slupkowe. Zauwazylismy rowniez, ze latwiej odczytuje sie roznic w malych katach (70% poprawnych odpowiedzi w pytaniu o najmniejsza wartosc, 50% w pytaniu o najwieksza). Rowniez odbiorcy wykresow uwazaja wykresy slupkowe za wygodniejsze. Takiej odpowiedzi udzielilo 100% ankietowanych.

Pytalismy rowniez o to jak bardzo ankietowani sa pewni swoich odpowiedzi. Odpowiedz 0 oznaczala, ze nie sa pewni, 5 ze sa pewni.

W przypadku wykresow kolowych srednia wyniosla: 1.78.

W przypadku wykresow slupkowych: 4.64.

Wyniki tych pytan ostatecznie pokazaly przewage wykresow slupkowych nad kolowymi.

## Eksperyment 2

### Wstęp

Zadaniem tego eksperymentu było zbadanie jaka jest możliwość wprowadzania odbiorców wizualizacji w błąd przy użyciu zmian w osi Y na wykresach. W tym celu skorzystałem z danych “countries” pochodzących z biblioteki SmarterPoland. Dane zostały zagregowane według kontynentów.

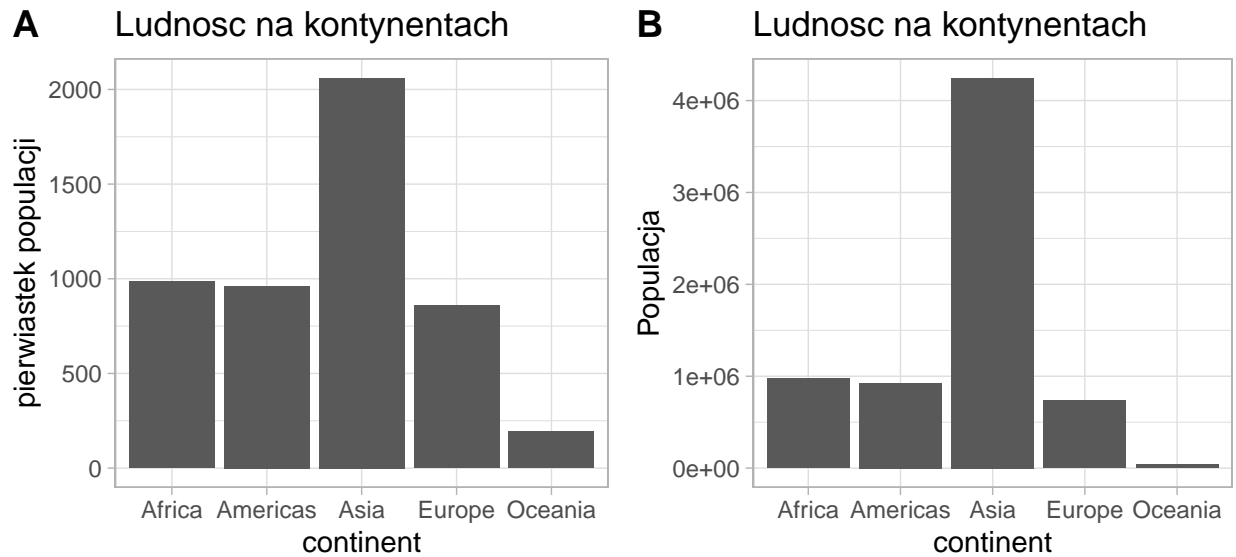
```
## # A tibble: 5 x 4
##   continent b.rate d.rate population
##   <chr>      <dbl>  <dbl>      <dbl>
## 1 Africa      33.7    9.25    971885
## 2 Americas    18.0    6.36    925417
## 3 Asia        19.9    5.89   4246175
## 4 Europe      10.7    9.99    736492
## 5 Oceania     22.3    6.19     37378
```

## Wykres 1

Pierwszą sposobem zmiany osi Y była zmiana skalowania wartości. Dane najpierw zostały spierwiastkowane, a następnie stworzony został wykres. Oś została stosownie podpisana.

A - wykres zmieniony

B - Wykres poprawny

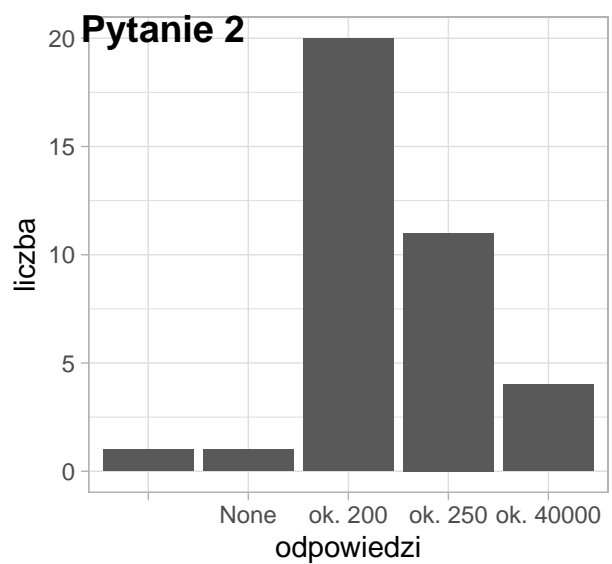
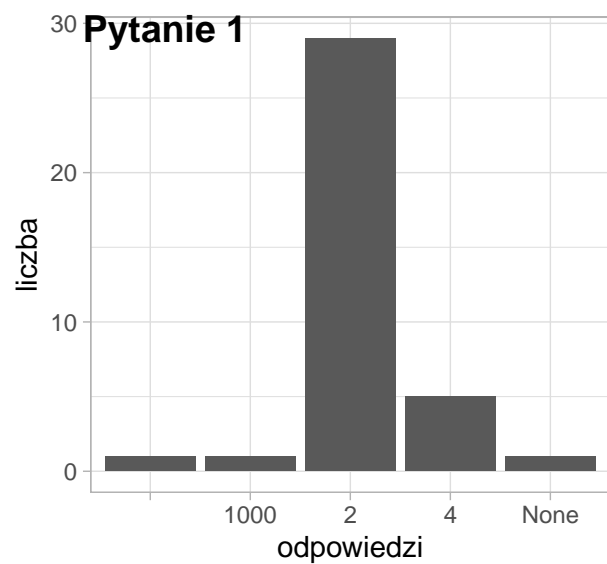


### Pytanie 1

Ile razy więcej liczy populacja Azji od Afryki? Zgodnie z przypuszczeniami przeważały odpowiedzi (2). Poprawnej odpowiedzi (4) udzieliło pięć osób. Mimo nieintuicyjnych wartości populacji całych kontynentów wiele osób nie zwróciło uwagi na podpis osi Y.

### Pytanie 2

Jaka jest populacja Oceanii? Także tutaj odpowiedzi (200) oraz (250) były najczęstsze (w sumie 31 razy). Poprawnej odpowiedzi (40000) udzieliło cztery osoby.

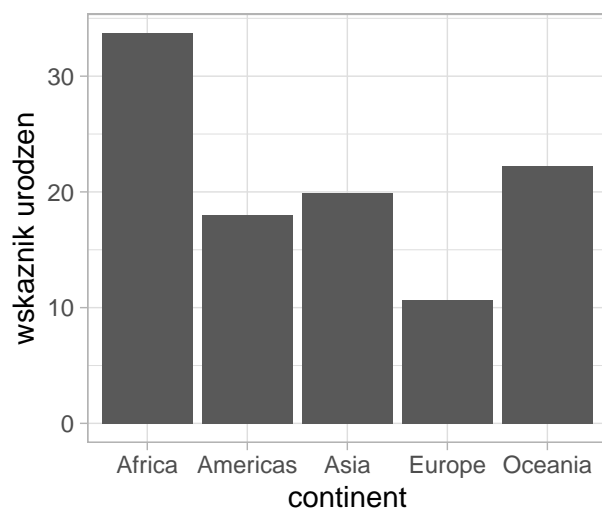




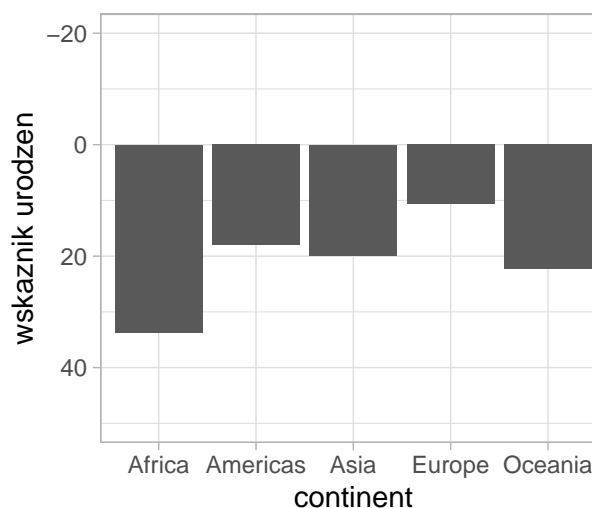
## Wykres 2

Na kolejnym wykresie odwróciłem oś Y tak aby większe wartości kończyły się niżej. Dzięki temu miało to sprawiać wrażenie, że to wartości największe są najmniejsze.

**A** Urodzenia na kontynentach



**B** Urodzenia na kontynentach

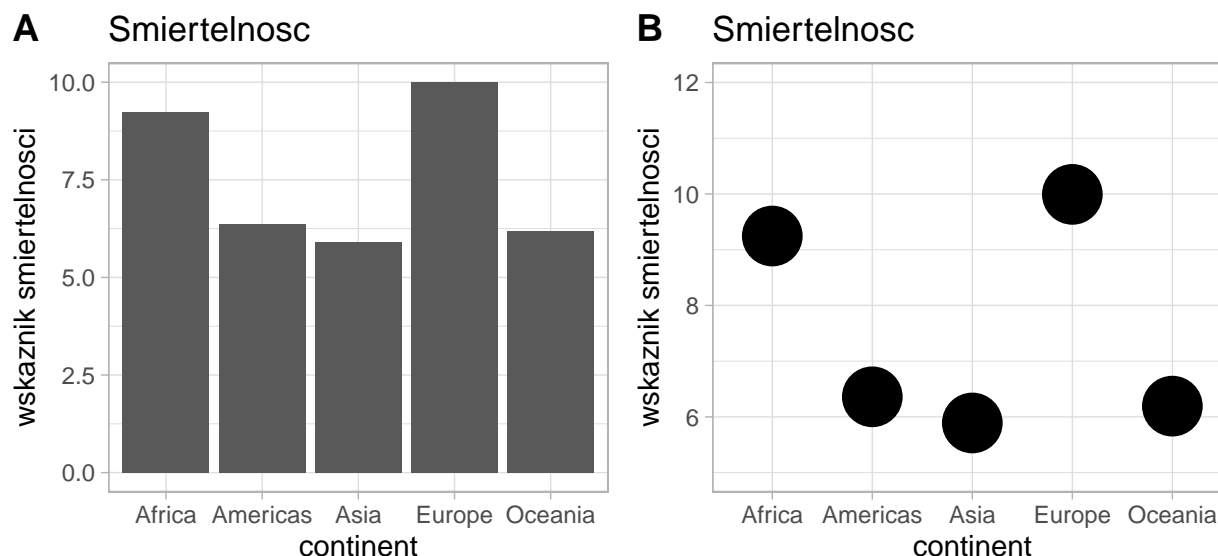


## Pytanie 3

W którym kontynencie jest najmniejszy wskaźnik urodzeń ? Przeciwnie do spodziewanego wyniku jedynie jedna osoba odpowiedziała błędnie (Afryka), z kolei 35 osobą udało się prawidłowo odczytać poprawną odpowiedź (Europa).

### Wykres 3

Na trzecim wykresie zastosowałem punktowy sposób prezentacji danych. Oraz zmieniłem oś Y tak aby zaczynała się od wartości 5. Dzięki temu mogłem uzyskać wrażenie dużej rozbieżności.

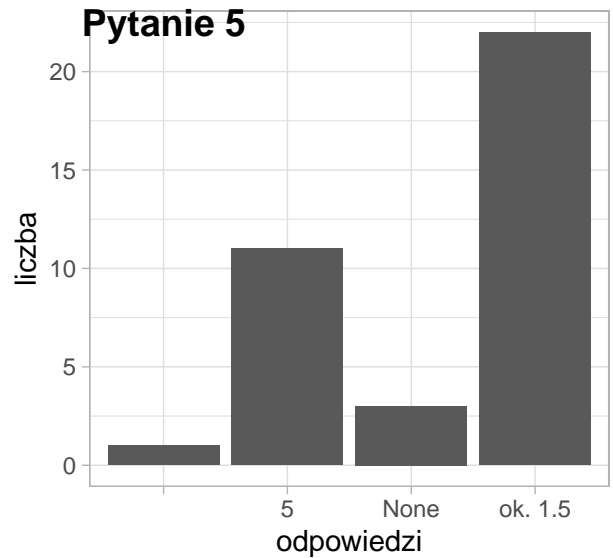
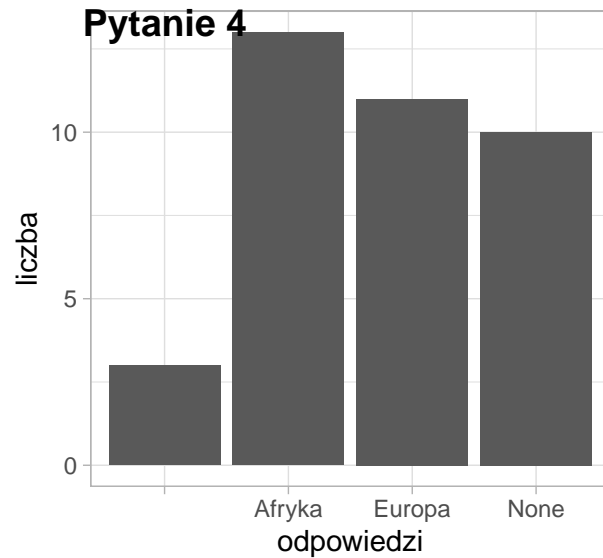


#### Pytanie 4

W którym kontynencie wskaźnik zgonów jest najwyższy? Pytanie to zostało umieszczone powyżej wykresów obrazujących wskaźnik śmiertelności, więc nieuważnie można było odczytać najmniejszą wartość z wykresu powyżej, który przedstawiał wskaźnik urodzeń. Jak widać 13 osób udzieliło odpowiedzi (Afryka) na podstawie wykresu urodzeń lub błędnie odczytało wykres zgonów. Kolejne 11 osób wskazało prawidłową odpowiedź (Europa). Natomiast 10 osób zaznaczyło odpowiedź (None = nie da się odczytać).

#### Pytanie 5

Ile razy większa jest śmiertelność w Europie niż w Azji? W ostatnim pytaniu poprawność odpowiedzi zależała od sposobu czytania wykresu. Jedenastu ankietowanych zaznaczyło odpowiedź (5) która wynikała z graficznej odległości punktów od siebie. Jednak większość, dwadzieścia dwie osoby zwróciło uwagę na przyjętą skalę dzięki czemu zaznaczyła poprawną odpowiedź (ok. 1.5). Były również 3 osoby które zaznaczyły (None = nie da się porównać).



## Wnioski

Z przedstawionych wyżej wyników eksperymentu można wyciągnąć następujące wnioski:

**skalowanie wartości** - jest w stanie wprowadzić w błąd znaczną część odbiorców

**odwrócenie osi** - nie jest to skuteczny sposób do wprowadzania odbiorców w błąd, prawdopodobnie ponieważ taka wizualizacja nie jest tak często używana wymaga większej uwagi

**zamiana kolejności informacji** - wprowadziła w błąd ok. 1/3 badanych co przy podobnych danych może być jeszcze skuteczniejsze, natomiast kolejna 1/3 zauważyła problem co może zmniejszyć wiarygodność wizualizacji

**zmiana wartości początkowej** - przy niewielkiej uwadze odbiorcy metoda ta może wprowadzać w błąd, natomiast zadając pytanie wymagające skupienia uwagi na wykresie minimalizuje błędy w czytaniu wykresu