

Wartości bliskie

Magdalena Szypulska

Streszczenie

W poniższej pracy została zbadana zależność pomiędzy różnymi czynnikami jak wiek, płeć czy preferencje wizualizacji w zależności od skuteczności odczytywania wartości bliskich z wykresu kołowego oraz wykresu słupkowego.

Spis treści

1. Doświadczenie.
2. Opis próbki i motywacja. Metryczka.
3. Ankieta. Użyte dane oraz pozyskanie danych.
4. Wizualizacja i omówienie wyników.
5. Podsumowanie badania.
6. Dalsze wnioski. Posłowie.

Doświadczenie

Eksperyment polegał na wypełnieniu ankiety stworzonej za pomocą Microsoft Forms, na którą składały się pytania, polegające na rozróżnieniu, która z prezentowanych wielkości ma większą/mniejszą wartość czy też czy grupa danych przekracza daną wartość. Wykresy do badania zostały stworzone za pomocą pakietu ggplot2 oraz bazowego R. Badani zostali poproszeni o podanie swojej płci oraz wieku. Na koniec ankiety zostało umieszczone zapytanie na temat preferencji w odczytywaniu danych.

Opis próbki i motywacja. Metryczka.

Opis próbki i motywacja

Ankiety wypełniło 25 osób w dniach 21-22.03.2021. W sondażu brali udział głównie prawnicy biurowi powyżej 36+ oraz studenci między innymi takich kierunków jak: architektura, prawo, elektrotechnika czy filologia angielska.

Podany sondaż może nam dostarczyć spojrzenia jak wygląda odczytywanie danych u osób, które podczas swojej pracy zawodowej bądź naukowej korzystają z wykresów. Innymi słowy: celem badania będzie odpowiedź na pytanie: 'Czy dla osób, które na codzien od wielu lat spotykają się z wieloma wizualizacjami nie ma znaczenia, czy odczytują dane z wykresu słupkowego czy kołowego?'.

Metryczka

Poniżej w tabelach znajdują się dane na temat płci oraz wieku badanych.
wiek oraz płeć badanych

Płeć	
kobieta	16
mężczyzna	9

Wiek	
18 - 26 lat	15
27 - 35 lat	1
powyżej 36 lat	9

Ankieta. Użyte dane oraz pozyskanie danych

Użyte dane oraz pozyskanie danych

Dane do ankiety zostały wygenerowane w sposób sztuczny tj. nie pochodzą z żadnej bazy ani celowo nie odzwierciedalają żadnego konkretnego zjawiska. Wykres słupkowy zostały stworzony przy pomocy pakietu ggplot2 a wykres kołowy za pomocą funkcji 'piechart'.

Dane ankietowanych zostały pobrane z Microsoft Forms w formie pliku z rozszerzeniem .xlsx (format programu Excel) oraz wczytane za pomocą pakietu "readxl" w RStudio.

Poniżej zostały umieszczone dane użyte w badaniu:

```
## # A tibble: 7 x 2
##   nazwy wartosc_slupkowy
##   <chr>             <dbl>
## 1 A                 6
## 2 B                6.2
## 3 C                7.4
## 4 D               12.1
## 5 E               17.9
## 6 F               18.1
## 7 G               32.3
```

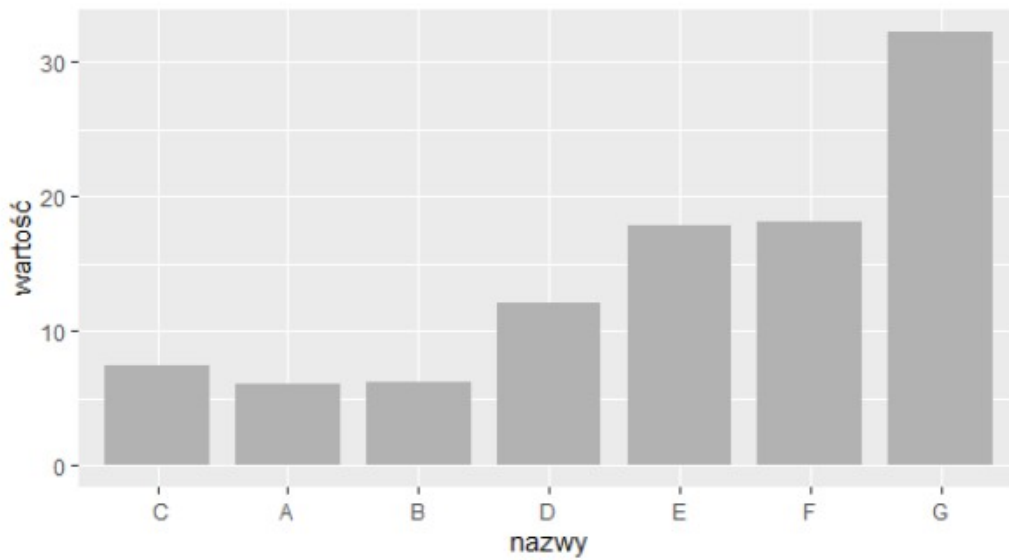
```
## # A tibble: 5 x 2
##   nazwy_kolowy procenty
##   <chr>             <dbl>
## 1 "A"              15
## 2 "B"              29
## 3 " C"             17
## 4 "D"              8
## 5 "E"             31
```

Ankieta

Ankietowani zostali spytani o relacje, w jakiej znajdują się dwie wartości na wykresie słupkowym oraz kołowym tj. która z nich jest większa bądź mniejsza, a także przy każdym wykresie zostało dodane pytanie dodatkowe.

Ze względu na trudność i złe dobranie danych (wymagamy zbyt dużej dokładności) w pytaniu 5. (dotyczącym wykresu słupkowego) nie będzie one brane pod uwagę w dalszych wizualizacjach wyników ankiety. Zostało jednak one umieszczone poniżej w celach informacyjnych.

Pytania 3 - 5. będą dotyczyły poniższego wykresu.



3

Która z wielkości ma niższą wartość? *

☐ E

☐ F

4

Która z wielkości ma wyższą wartość? *

☐ A

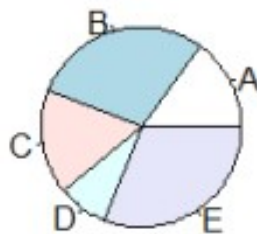
☐ B

5

Wartość D należy do przedziału: *

- ☐ [0,10)
- ☐ [10,12)
- ☐ [12,15)
- ☐ [15,20)

Pytania 6 - 8. będą dotyczyły się poniższego wykresu.



6

Która z danych ma większą wartość? *

☐ B

☐ E

7

Która z danych ma mniejszą wartość? *

☐ A

☐ C

8

Czy suma B, C oraz D to więcej niż 50 procent? *

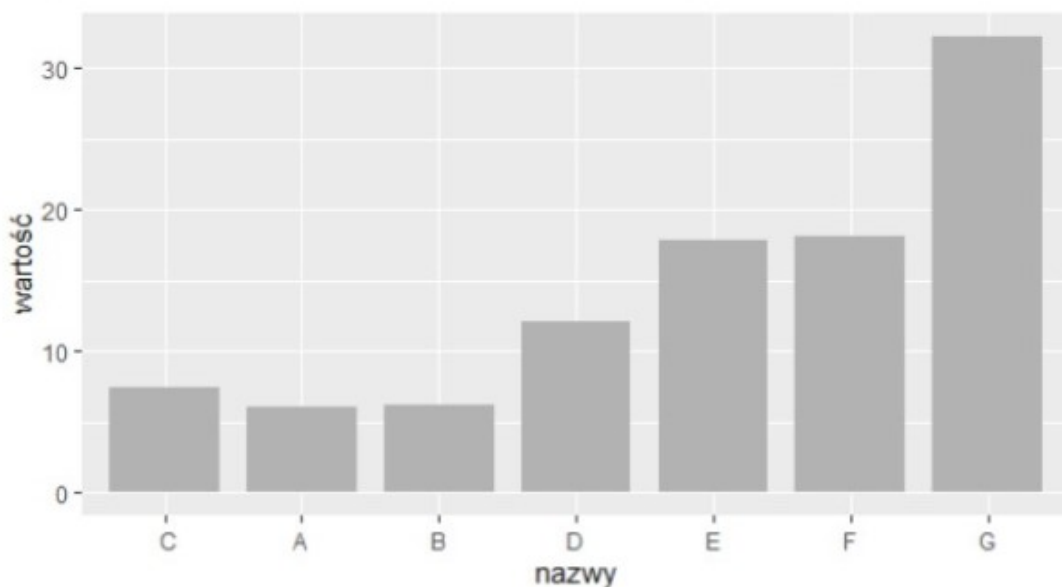
☐ Tak

☐ Nie

Wizualizacja i omówienie wyników

Na początku przedstawimy statystyki dotyczące każdego pytania, a następnie sprawdzimy jak z badaniem poradziły sobie dane grupy wiekowe, płcie oraz sprawdzimy, czy preferencja wykresu ma związek z ilością poprawnych odpowiedzi. Tak jak było wspomniane to w paragrafie poprzedzającym, pytanie nr 5 nie było brane pod uwagę w statystykach.

Pytania 3 - 5. będą dotyczyły poniższego wykresu.



3

Która z wielkości ma niższą wartość? *

☐ E

☐ F

Poprawną odpowiedzią jest odpowiedź E, 24 z 25 badanych odpowiedziało poprawnie.

4

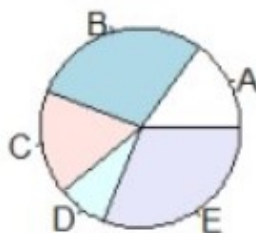
Która z wielkości ma wyższą wartość? *

☐ A

☐ B

Poprawną odpowiedzią jest odpowiedź B, 23 z 25 badanych odpowiedziało poprawnie.

Pytania 6 - 8. będą dotyczyły się poniższego wykresu.



6

Która z danych ma większą wartość? *

☐ B

☐ E

Poprawną odpowiedzią jest odpowiedź E, 23 z 25 badanych odpowiedziało poprawnie.

7

Która z danych ma mniejszą wartość? *

☐ A

☐ C

Poprawną odpowiedzią jest odpowiedź A, 23 z 25 badanych odpowiedziało poprawnie.

8

Czy suma B, C oraz D to więcej niż 50 procent? *

☐ Tak

☐ Nie

Wszyscy badani odpowiedzieli w sposób poprawny.

9

Dane podczas badania łatwiej odczytywało się Pani/Panu z wykresu:

☐ kołowego

☐ słupkowego

☐ z obu podobnie

Taka sama liczba badanych preferowała wykres kołowy i słupkowy (tj. liczność obu grup wynosi 9), niewiele mniej badanych wybrało odpowiedź bez wskazania konkretnego wykresu - 7.

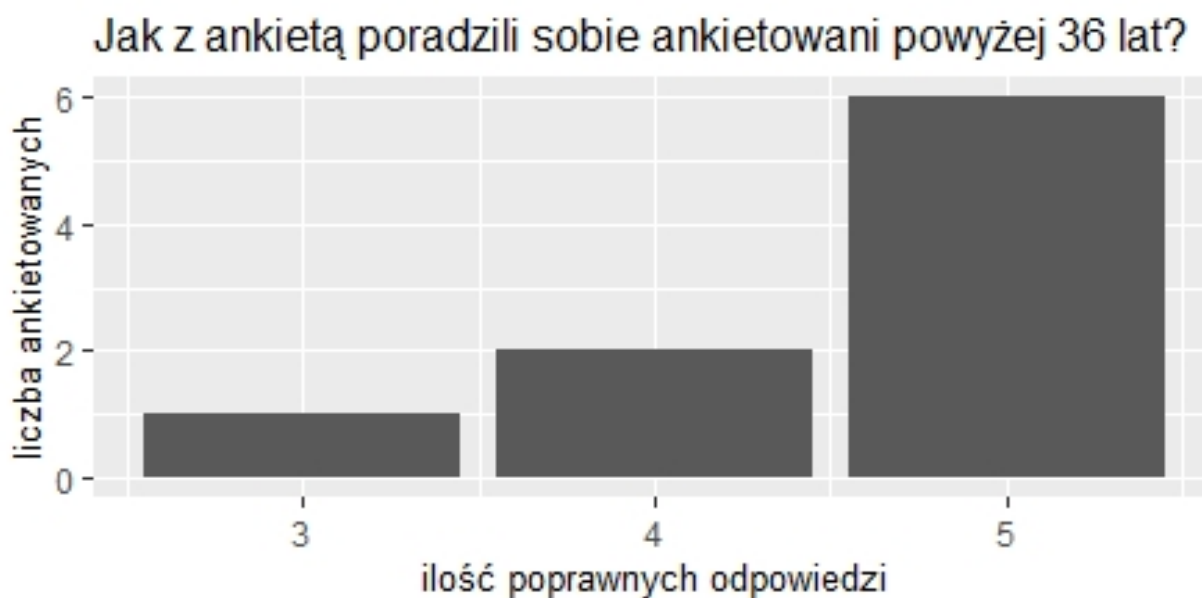
Średnia ilość poprawnych odpowiedzi w badaniu wyniosła 4.72, co stanowi dosyć wysoki wskaźnik, liczność grup została zwizualizowana na wykresie słupkowym poniżej:



W zależności od wieku średnia arytmetyczna uzyskanych poprawnych odpowiedzi różniła się:

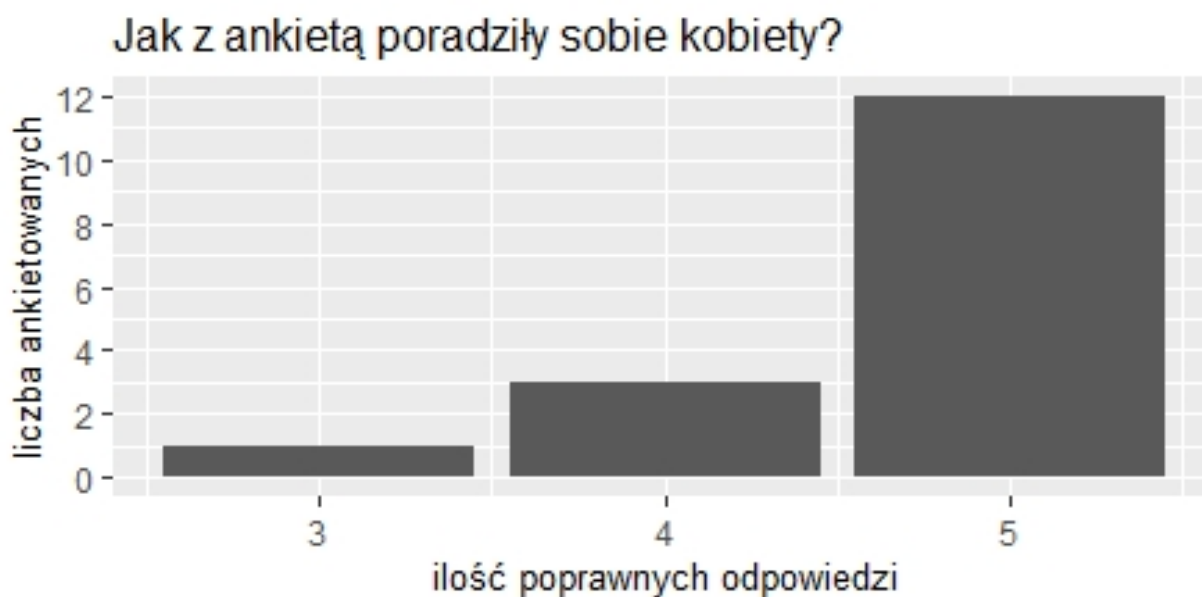
- jedyny przedstawiciel grup wiekowej pomiędzy 27 a 35 rokiem życia uzyskał 5 poprawnych odpowiedzi,
- grupa 36+ uzyskała średnio 4.56,
- osoby pomiędzy 18 a 26 rokiem życia uzyskały średnio 4.8 poprawnych odpowiedzi.





W zależności od płci średnia arytmetyczna uzyskanych odpowiedzi różniła się:

- kobiety uzyskały średnio 4.69 poprawnych odpowiedzi,
- mężczyźni - 4.78.

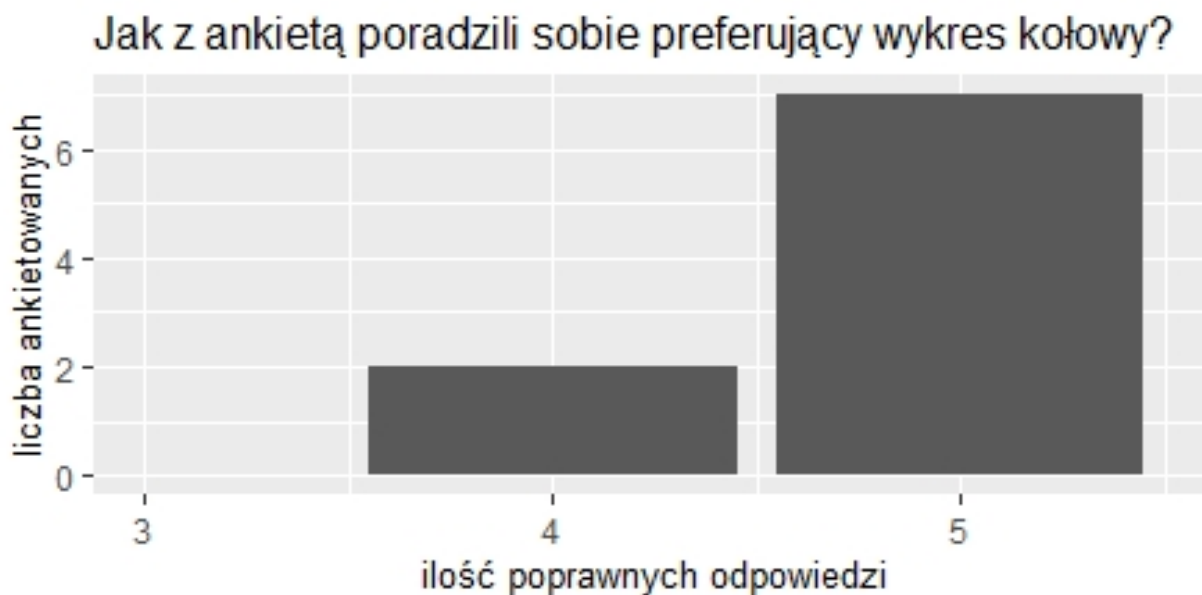
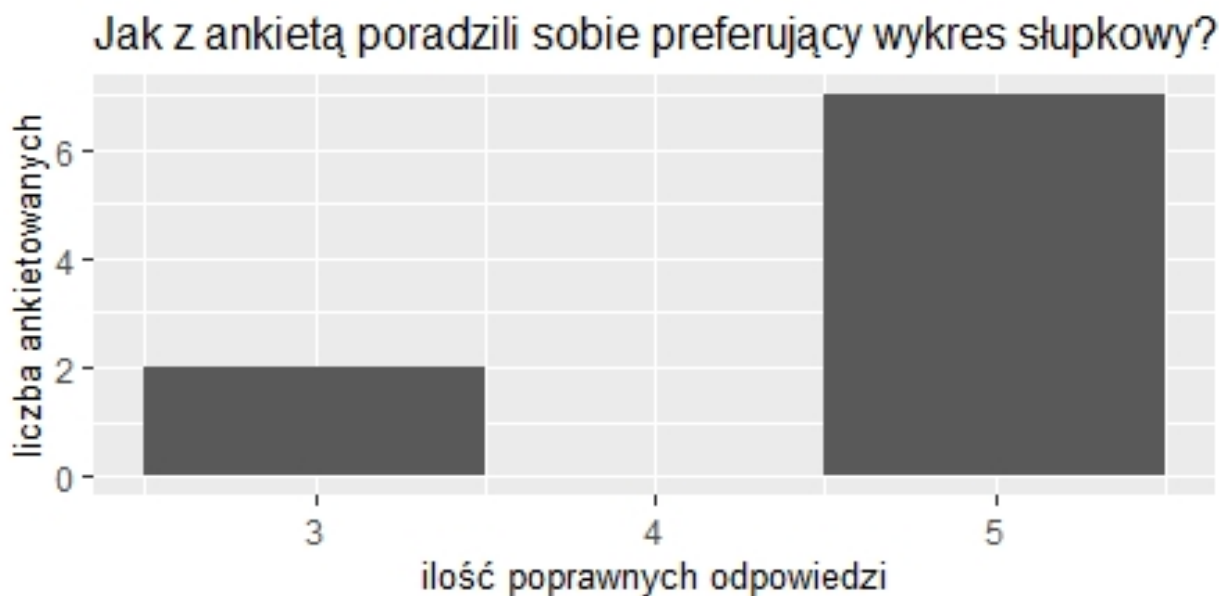




W zależności od preferencji wykresu podczas badania średnia arytmetyczna poprawnych odpowiedzi wyniosła:

- brak preferencji wykresu - 4.86,
- słupkowy - 4.56,
- kołowy - 4.78.





Podsumowanie badania

Większość ankietowanych poradziła sobie bardzo dobrze z postawionym zadaniem, o czym świadczy wysoki wskaźnik prawidłowych odpowiedzi. Skłania to do wniosku, że osoby, które pracują na co dzień z danymi radzą sobie bardzo dobrze z odczytywaniem bliskich wielkości niezależnie od typu wizualizacji. Porównując grupy wiekowe pomiędzy 18 a 26 rokiem życia a 36+, możemy zauważyć, że młodszy badani poradziła sobie stosunkowo lepiej. Może być to związane z ekspozycją na dane od najmłodszych lat czy też cechami fizycznymi jak np. lepszy wzrok. Zgodny z intuicją wydaje się rezultat, że najlepiej w badaniu poradziły sobie osoby, dla których nie miał znaczenia rodzaj wykresu. Zastanawiający może być wynik, że osoby preferujące wykres kołowy poradziły sobie lepiej od osób, które wolały wykres słupkowy, jednak ze względu na małą grupę badanych i inne czynniki pozwolę sobie nie wyciągać daleko idących wniosków. Kobiety poradziły sobie stosunkowo gorzej od mężczyzn, należy wziąć pod uwagę jednak to, że wśród ankietujących przeważały

kobiety, co mogło być przyczyną takiego wyniku.

Dalsze wnioski. Pośłowie

Po przeprowadzaniu ankiety i dokonaniu wizualizacji ciężko obyć się bez wniosków na temat jakości pytań czy też wykorzystanych wizualizacji.

Zauważmy, że zwyczajowo na wykresach słupkowych umieszcza się dane posortowane, zatem sytuacja z badania mogła się okazać nienaturalna dla części ankietowanych. Problem dla mniej uważnych ankietowanych mogły stanowić pytania “o zmienniającym się charakterze” tj. raz zostali spytani o to, która z wielkości jest większa, a raz która z nich jest mniejsza. Być może na przyszłość należałoby używać pytań o “jednym charakterze”.

W pytaniu, które zostało odrzucone powinny zostać dobrane lepiej przedziały (lub ewentualnie na wykresie powinny znaleźć się dodatkowe podziały), w których znajduje się dana. Obecny kształt pytania mógł zadecydować o pewnej losowości w przypadku wybierania odpowiedzi.

W celu wyciągnięcia lepszych wniosków należałoby zadbać, żeby ilość badanych w danych grupach bardziej odzwierciedlała populację, co prowadziło do udoskonalenia charakterystyki umiejętności odczytywania informacji wśród grup wiekowych.

W dalszym badaniu hipotez na temat odczytywania wartości bliskich warto byłoby rozszerzyć ankietę na osoby, które nie mają doświadczenia z danymi na co dzień - czyli z jednostkami, które prawdopodobnie mogą być najbardziej narażone na manipulacje danymi.