

# Raport - praca domowa 3

Aleks Kapich

13 listopada 2022

## 1 Wstęp

Celem przeprowadzonego w ramach pracy domowej eksperymentu było zbadanie, jaki wpływ na odczytywanie danych mają osie wykresów.

Eksperyment polegał na zaprezentowaniu dwóm różnym grupom badawczym zestawienia dwóch wykresów słupkowych. Badani proszeni byli o odczytanie informacji z wykresów i udzielenie odpowiedzi na pytanie dotyczące przedstawionych im danych.

Pierwszej grupie przedstawione zostały wykresy stworzone w poprawny sposób. Druga grupa z kolei została poproszona o udzielenie odpowiedzi patrząc na wykresy, gdzie osie były zmanipulowane - poprzez dobranie nieodpowiedniej skali lub rozpoczęcie osi w ustalonym punkcie różnym od zera.

## 2 Przygotowanie eksperymentu

### 2.1 Opis danych

Dane użyte w eksperymencie dotyczą statystyk trzech zawodników angielskiej Premier League osiągniętych w obecnym sezonie. Zostały one zwizualizowane na parach wykresów słupkowych, gdzie pierwszy wykres z pary dotyczył "bramek oczekiwanych" (Expected Goals), czyli dorobku bramkowego zawodnika oszacowanego z użyciem modelu matematycznego. Drugi wykres w parze przedstawiał realną liczbę goli zdobytą przez danego gracza. Źródłem danych jest firma Opta, udostępnione zostały poprzez witrynę [FBref](#).

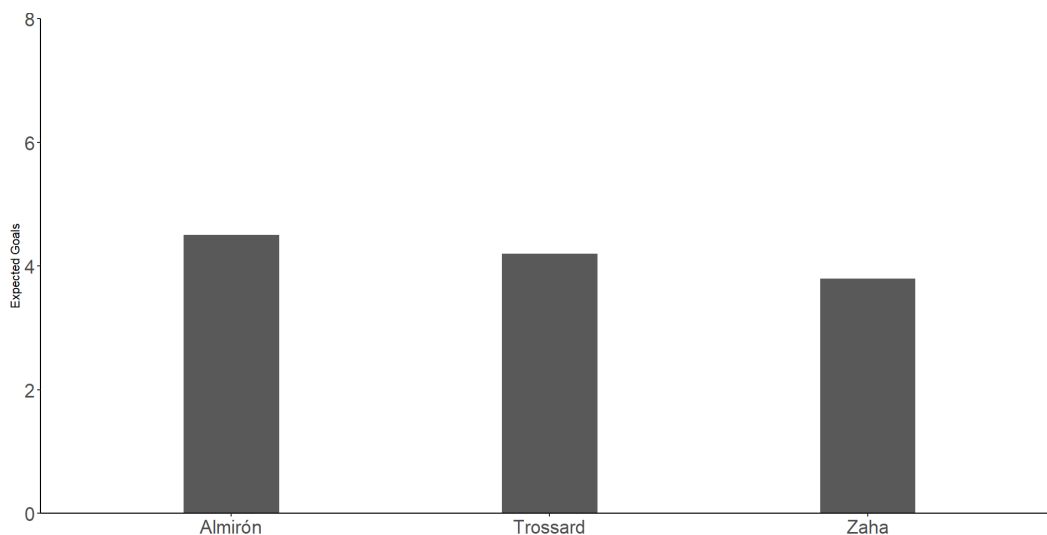
Ponadto, użyte dane zostały wybrane w taki sposób, aby nie dotyczyły zawodników cieszących się dużą popularnością. Dzięki temu minimalizowane jest prawdopodobieństwo zajęcia sytuacji, w której odpowiadający zamiast odczytać informacje z wykresu, kieruje się sympatią podczas udzielania odpowiedzi.

Gracz	Bramki	Oczekiwane bramki
M.Almirón	8	4.5
L.Trossard	7	4.2
W.Zaha	6	3.8

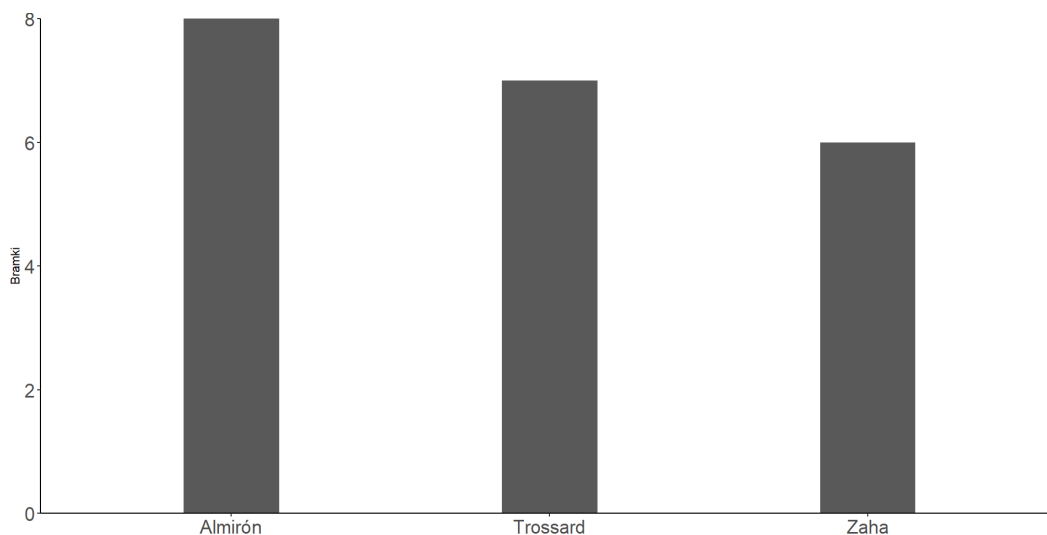
Tabela 1: Statystyki przedstawione na wykresach.

### 2.2 Wykresy prawidłowe

W poprawnym zestawie wykresów, na obydwu z nich zastosowane zostały te same granice osi, aby nie utrudniać odczytywania danych. W wypadku wizualizacji przedstawionej poniżej na rysunku 1 decyzja ta może wzbudzać kontrowersje, bowiem prowadzi do spłaszczenia wykresu, jednakże w obliczu konieczności porównywania dwóch wykresów uznałem, że zabieg ten ułatwi zrozumienie danych. Ponadto, osie na wykresach zaczynają się od wartości 0, aby nie zniekształcać proporcji między poszczególnymi słupkami.



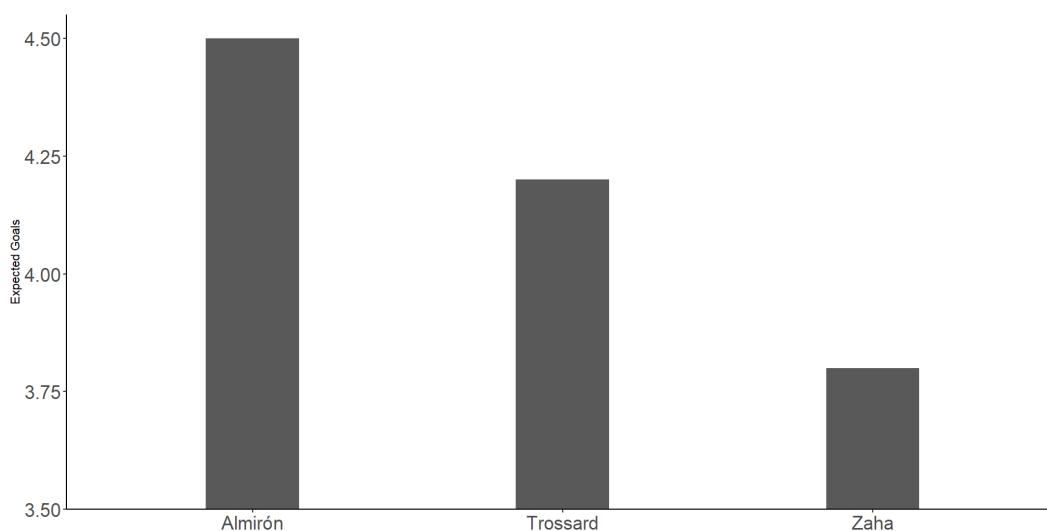
Rysunek 1: Prawidłowy wykres przedstawiający oczekiwane bramki.



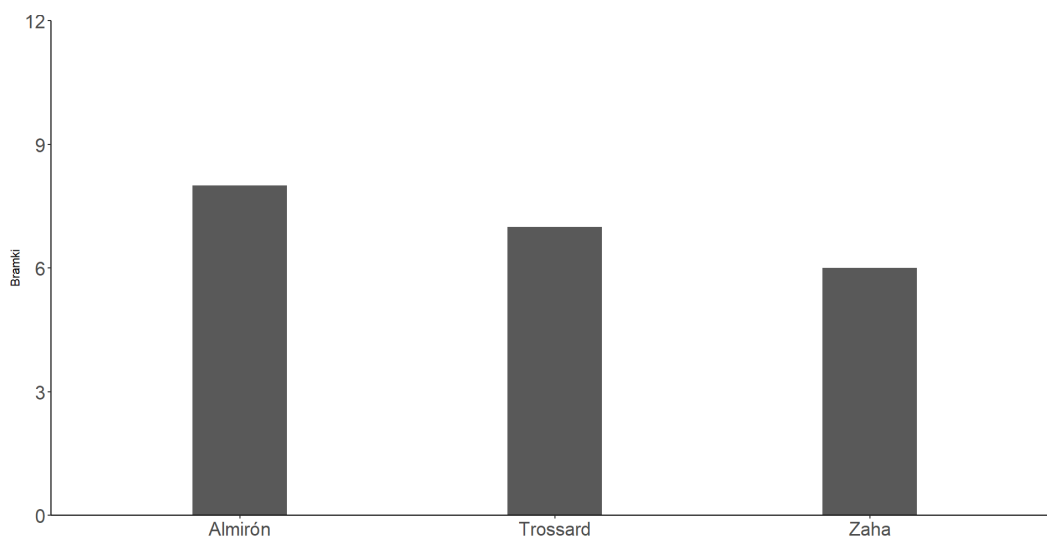
Rysunek 2: Prawidłowy wykres przedstawiający zdobyte bramki.

## 2.3 Wykresy nieprawidłowe

Na wykresach nieprawidłowych osie dobrane są tak, by zmienić postrzeganie przedstawionych danych. Wykres "Expected Goals" zaczyna się od punktu 3.5, przez co wydawać się może, że różnica w wartościach bramek oczekiwanych dla poszczególnych zawodników jest relatywnie wysoka. Na wykresie przedstawiającym bramki, oś Y bez wyraźnej potrzeby kończy się na wartości 1.5 razy większej niż maksymalna wartość z wykresu, co daje wrażenie "spłaszczenia" danych - wartości wydają się być bardziej zbliżone do siebie niż w rzeczywistości.



Rysunek 3: Nieprawidłowy wykres przedstawiający oczekiwane bramki.



Rysunek 4: Nieprawidłowy wykres przedstawiający zdobyte bramki.

### 3 Przeprowadzenie eksperymentu

Osoby partycypujące w eksperymencie nie zostały poinformowane dokładnie, jaki jest jego cel, aby w sposób sztuczny nie zwrócić ich uwagi na manipulację wykresami. Partycypanci po zapoznaniu się z wizualizacjami oraz ich dokładnymi opisami zostali poproszeni o odpowiedź na jedno nieskomplikowane pytanie:

**Na podstawie wykresów odczytaj, który zawodnik prezentuje się najlepiej względem oczekiwań?**

Obydwie części badania przeprowadzane były do momentu uzyskania 30 odpowiedzi.

### 4 Wyniki

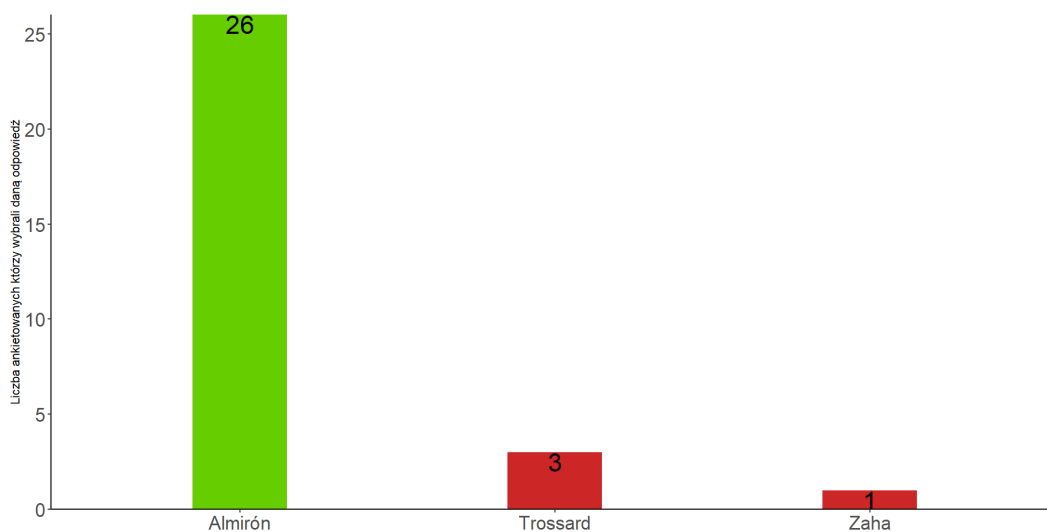
Różnicę między zdobytymi bramkami a ich oczekiwaną liczbą obrazuje poniższa tabela:

Gracz	Różnica
M.Almirón	+3.5
L.Trossard	+2.8
W.Zaha	+2.2

Tabela 2: Który zawodnik prezentuje się najlepiej?

Poprawna odpowiedź na zadane w badaniu pytanie to M.Almirón, bowiem strzelił on najwięcej bramek ponad oczekiwania spośród piłkarzy w zestawieniu.

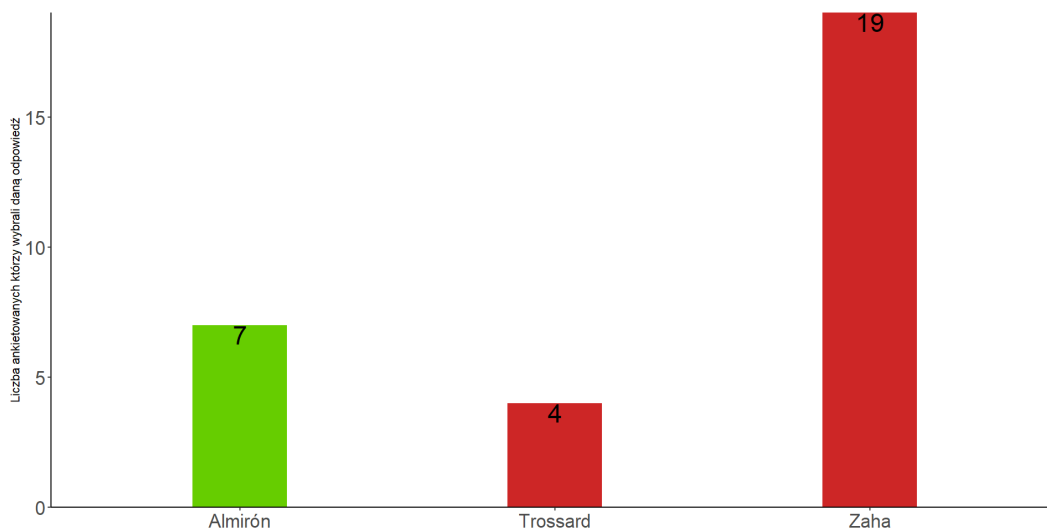
W grupie I, zdecydowana większość odczytała informacje z wykresu w sposób poprawny:



Rysunek 5: Odpowiedzi badanych z I grupy.

26 osób z 30 nie miało problemów z wskazaniem prawidłowej odpowiedzi, jest to aż 87% pytanych. Powodem pomyłki u pozostałych 13% być może okazał się brak zrozumienia pytania (szczególnie u osoby, która wskazała Wilfrieda Zahe) bądź też pobieżne odczytanie danych z wykresów (u osób, które wskazały opcję środkową, Leandro Trossarda).

W II grupie znaczna liczba ankietowanych dała się nabrać na manipulację z użyciem osi i udzieliła odpowiedzi patrząc jedynie na wysokość słupków.



Rysunek 6: Odpowiedzi badanych z II grupy.

Z 30 osób jedynie 7 (23% respondentów) poprawnie odczytało, że najlepsze wyniki osiąga obecnie M.Almirón. Spośród pozostałych 23 osób aż 19 wskazało Wilfrieda Zahe, najprawdopodobniej ponieważ w jego wypadku różnica wysokości słupków była najbardziej wyraźna. W przypadku pozostałych 4 osób, tak jak w I grupie, prawdopodobnym jest brak zrozumienia pytania bądź niedbała odpowiedź na zadane pytanie.

## 5 Podsumowanie

Rezultat eksperymentu nie stanowi dużego zaskoczenia - zgodnie z przypuszczeniami, manipulowanie osiami wykresu negatywnie wpływa na łatwość odczytywania danych. Większość ludzi natrafiając na wizualizacje danych w przestrzeni internetowej bądź innych mediach nie poświęca na ich analizę więcej niż kilka sekund, co często prowadzić może do wyciągania złych wniosków z wykresów. Proste zabiegi, takie jak zmiana granic osi, diametralnie mogą zmienić odbiór wykresu, tak jak miało to miejsce wśród respondentów z grupy II.

Pewne wątpliwości odnośnie tego, czy pytanie w eksperymencie zostało postawione klarownie, wzbudzają osoby, które wskazały Leandro Trossarda (odpowiednio 3 i 4 w I oraz II grupie). Zarówno podejście do zadania w sposób adekwatny, czyli odczytanie danych z wykresu z użyciem osi, jak też podejście do zadania w sposób "wzrokowy", czyli porównanie wielkości słupków, nie powinny wiązać się z udzieleniem tejże odpowiedzi.

Głównym wnioskiem płynącym z eksperymentu jest istotność dbałości o osie wykresów. Tworząc diagramy kolumnowe należy zadbać o spójność skali na powiązanych ze sobą wykresach oraz starać się zachowywać odpowiednie proporcje między słupkami, ustalając wartość 0 jako granicę osi. Wtedy znacząco zwiększymy szanse na to, że nawet nieuważne oko poprawnie odczyta dane z wykresu.