

Praca domowa 3

Jakub Seliga

Wstęp do przeprowadzonego eksperymentu

Eksperymentem, jaki w ramach tej pracy domowej wybrałem, jest próba zbadania tego, jak skala kolorystyczna wpływa na dokładność odczytania danych z mapy oraz porównanie metody kolorowania mapy z wykresem słupkowym. Skala kolorów w wizualizacji poprzez kolorowanie mapy to obszerny temat, dlatego ograniczyłem się do badania jedynie sekwencyjnej, niegrupowanej skali kolorów (dane typu mało-dużo, bez dzielenia na przedziały)

Wg klasycznie przyjętych dobrych praktyk, tworząc wizualizację poprzez takie kolorowanie mapy, przy wyborze kolorów powinniśmy trzymać się następujących punktów:

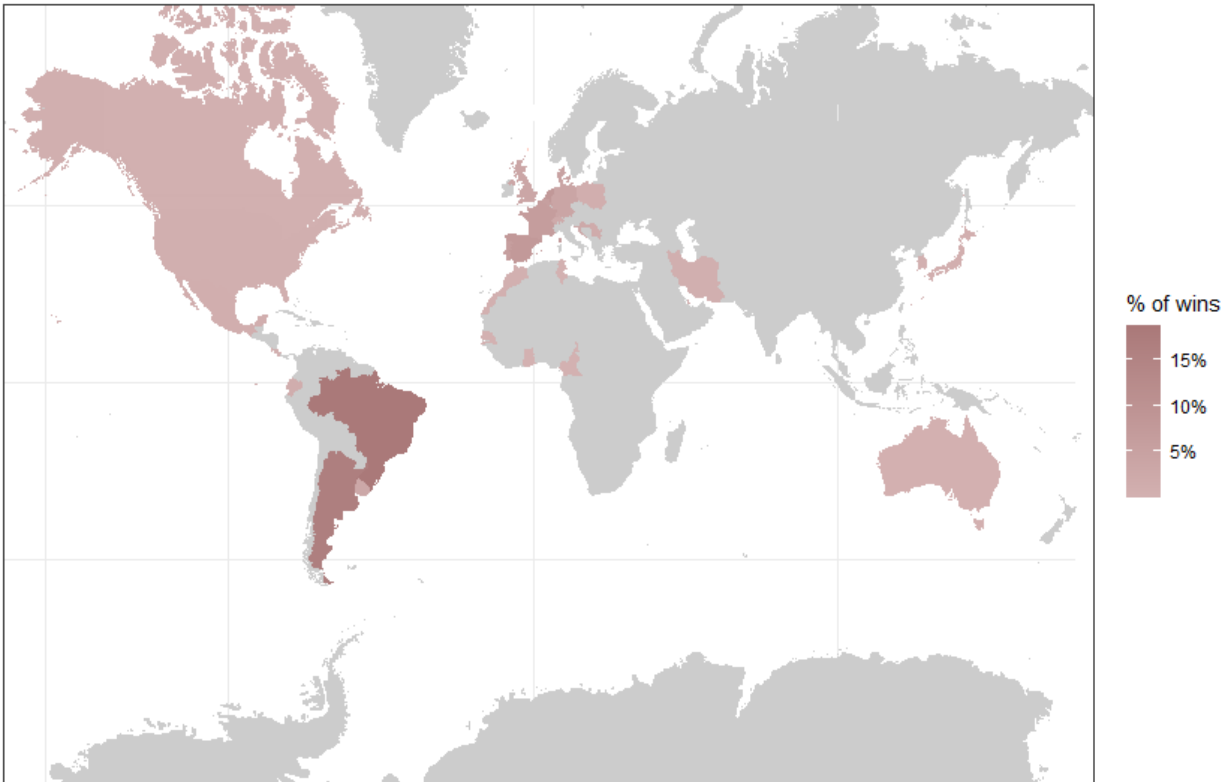
1. Dobrze jest zmieniać tylko jedną wartość HSV koloru w ramach skali
2. Im większy kontrast, tym większa widoczność

Dlatego w ramach eksperymentu, sprawdzę czy nagięcie tych zasad faktycznie pogorszy komfort i skuteczność w odczytywaniu danych.

Konstruowanie eksperymentu

Danymi, na jakich podstawie przeprowadziłem eksperyment, jest odsetek zwycięstw danego państwa w symulacjach Mistrzostw Świata w piłce nożnej 2022 przeprowadzonych na podstawie rankingu ELO. Do 31 państw świata (Anglia i Walia zostały zsumowane w Zjednoczone Królestwo) przyporządkowane zostały wartości procentowe zwycięstw, do innych państw wartości *NA*. Na tej podstawie utworzone zostały dwie mapy:

Gorsza mapa:

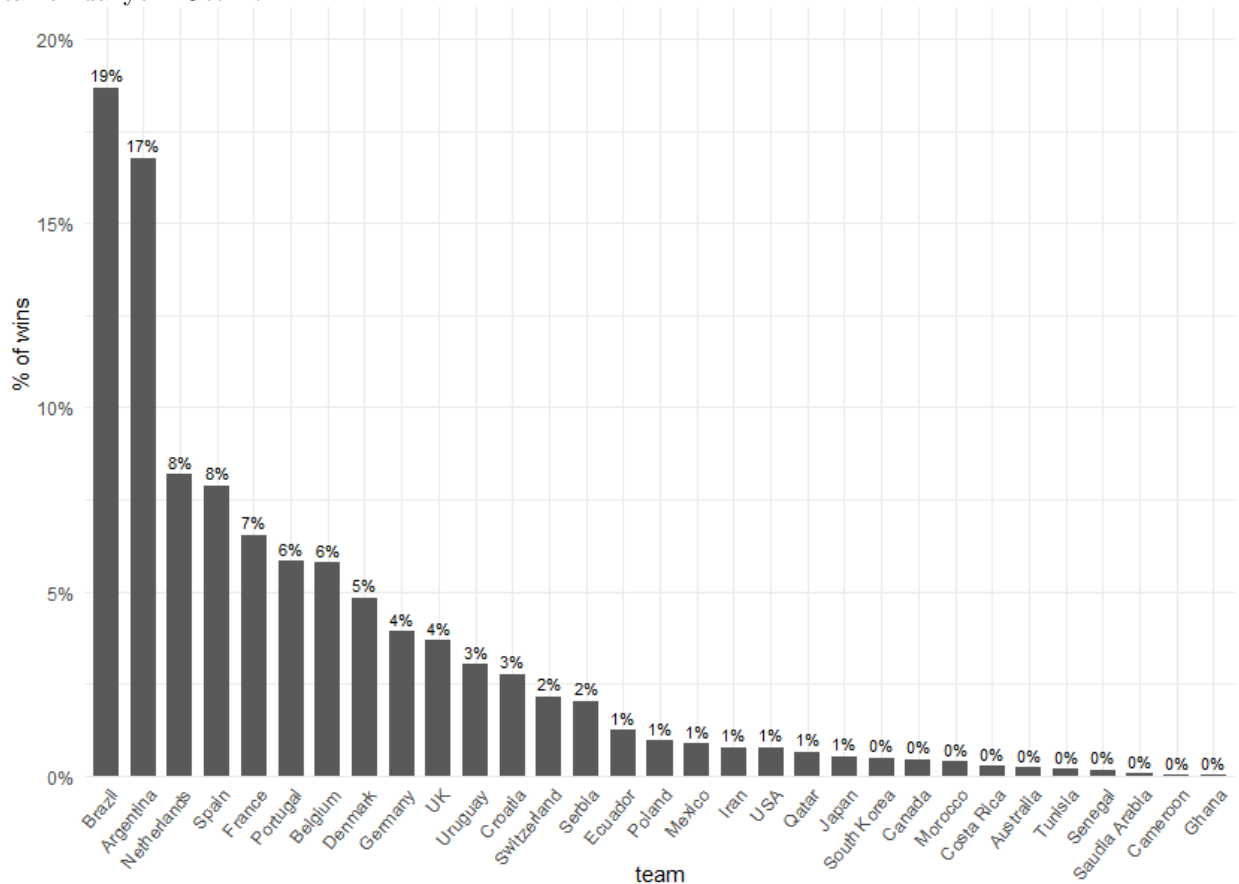


Widzimy, że skala kolorów jest bardzo blada, ma mały kontrast, w jej ramach zmieniają się jednocześnie wartości *Saturation* i *Value*
Lepsza mapa:



Na tej mapie zastosowana skala kolorów jest ‘szersza’, kontrastowa, zmienia się w niej jedynie saturacja odcienia.

Aby mieć próbkę, czy respondenci wiedzą co się dzieje i zaznaczają wg nich prawidłowe odpowiedzi, do eksperymentu został dołączony wykres słupkowy, który powinien być o wiele prostszy w odczytaniu akurat takich danych. Oto i on:



Do każdego z trzech wykresów zadane zostały te same 3 pytania, mające na celu sprawdzić, z którego z nich występują najmniejsze/największe błędy i trudności w odczycie danych:

1. Które europejskie państwo osiągnęło największy % zwycięstw wg danych z powyższego wykresu?

(do wyboru: Holandia [prawidłowa odp.], Niemcy, Francja, Hiszpania, Belgia, Portugalia, Nie mogę stwierdzić)

2. Kanada czy USA ma wyższy % zwycięstw?

(do wyboru: Kanada, USA [prawidłowa odp.], Nie mogę stwierdzić)

3. Brazylia czy Argentyna ma wyższy % zwycięstw?

(do wyboru: Brazylia [prawidłowa odp.], Argentyna, Nie mogę stwierdzić)

Na końcu respondentom zostały zadane 2 dodatkowe pytania o ich odczucia:

- Na której mapie łatwiej porównywało się wartości?

(Mapa 1/Mapa 2)

- Czy wykres słupkowy był dużym ułatwieniem w odczytywaniu wartości?

(Tak, dużym/Tak, lekkim/Nie, mapy wystarczyły)

Analiza odpowiedzi

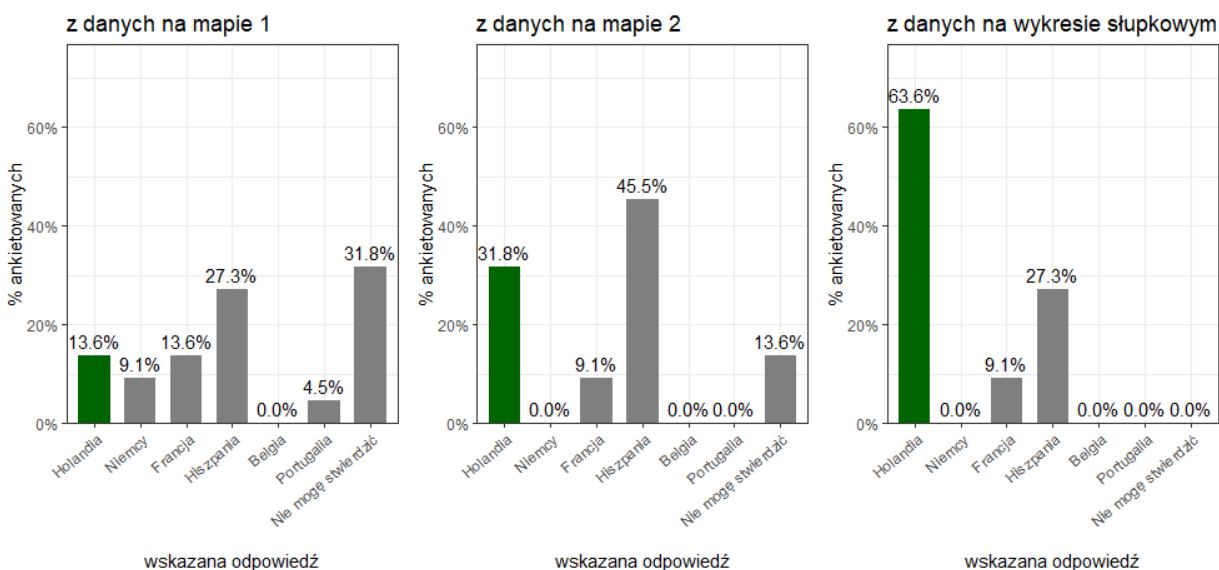
Na ankietę odpowiedziało 22 niezależnych respondentów. Dla poszczególnych typów wykresów, wskazywali oni prawidłowe odpowiedzi następującą ilość razy:

rodzaj wykresu	% poprawnych odpowiedzi na pytanie nr 1	% poprawnych odpowiedzi na pytanie nr 2	% poprawnych odpowiedzi na pytanie nr 3
mapa nr 1	13,6%	18,2%	27,3%
mapa nr 2	31,8%	59,1%	68,2%
wykres słupkowy	63,6%	95,5%	100%

Widzimy, że dla każdego z 3 pytań zachodziła ta sama zależność: Zdecydowanie najwięcej poprawnych odpowiedzi ankietowani oddawali przy wykresie słupkowym, mniej przy mapie nr 2, a najmniej przy mapie nr 1. Przyjrzyjmy się teraz rozkładowi odpowiedzi dla każdego pytania:

Pytanie nr 1

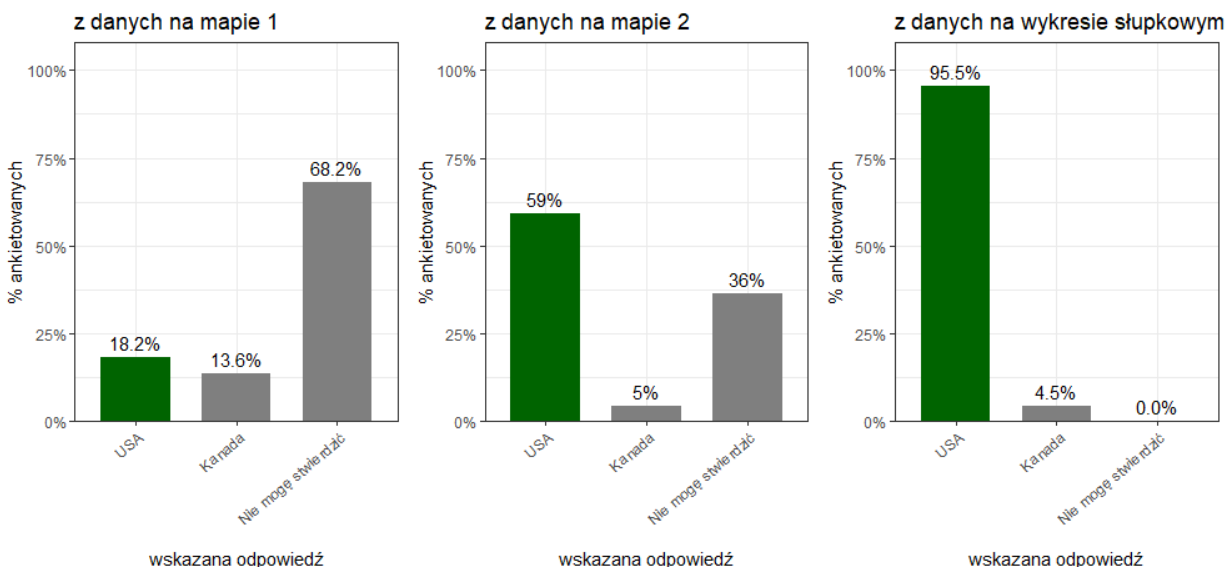
Który europejski kraj ma największy % zwycięstw?



Z mapy nr 2, poprawną odpowiedź odczytywano ponad 2 razy częściej niż z mapy nr 1. Warto także zauważyć, że 'tylko' 63% respondentów podało poprawną odpowiedź po zobaczeniu posortowanego wykresu słupkowego. Wynika to najprawdopodobniej ze zwykłej nieuwagi ankietowanych, a także ich przekonania po zobaczeniu mapy nr 2, że największą wartość posiada Hiszpania (najwięcej wskazań z odczytu mapy 2). Ilość odpowiedzi 'Nie mogę stwierdzić' była ponad 2 razy większa dla mapy nr 1, w porównaniu do drugiej.

Pytanie nr 2

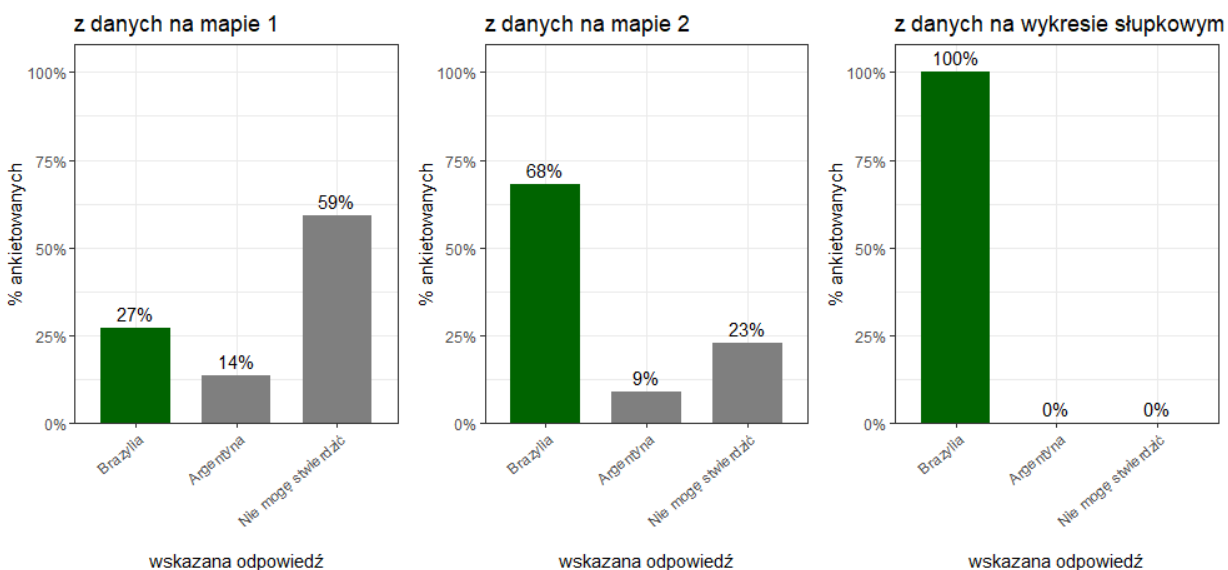
Kanada czy USA ma większy % zwycięstw?



Drugie i trzecie pytanie okazało się być prostsze niż pierwsze, co wynika zapewne z tego, że tym razem wystarczyło porównać tylko 2 wartości, i to dla państw o dużo większej powierzchni niż te europejskie. Poprawną odpowiedź - USA - wskazywano ponad 3 razy częściej po zobaczeniu mapy 2. Warto zwrócić uwagę na stosunek ilości odpowiedzi z wyłączeniem opcji 'Nie mogę stwierdzić' - dla mapy nr 1, wyniki odpowiedzi USA i Kanada były porównywalne, dla mapy 2 ta różnica była bardzo duża. Po wykresie słupkowym, prawie wszyscy ankietowani odpowiedzieli poprawnie.

Pytanie nr 3

Brazylia czy Argentyna ma większy % zwycięstw?



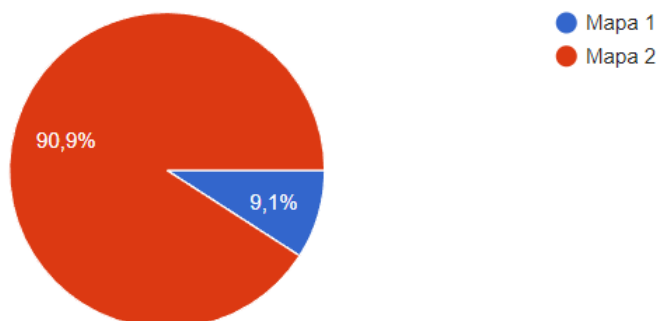
Trzecie pytanie miało bardzo podobną konstrukcję do drugiego, toteż otrzymaliśmy bardzo podobny rozkład odpowiedzi ankietowanych. Od pytania nr 2 różni je przede wszystkim to, że tam badani porównywali zbliżone do siebie niskie wartości, a w tym pytaniu, dwie najwyższe na całej mapie. Poprawną odpowiedź

wskazywano ponad 2.5 raza częściej dla mapy 2, a stosunek ilości odpowiedzi ‘Brazylia’ do ‘Argentyna’ zwiększył się prawie 4-krotnie. Po wykresie słupkowym, wszyscy odpowiedzieli prawidłowo.

Pytania dodatkowe

Na której mapie łatwiej porównywało się wartości?

22 odpowiedzi



Pierwsze pytanie dodatkowe miało zwrócić uwagę na porównanie nie tylko samej poprawności odczytu, ale też komfortu odczytu dla obu map. Spośród 22 ankietowanych, tylko 2 z nich wskazało na mapę nr 1.

Czy wykres słupkowy był dużym ułatwieniem w odczytywaniu wartości?

22 odpowiedzi



To pytanie miało dać odpowiedź na to, jak zastosowanie wykresu słupkowego ułatwia odczytanie takiego typu danych. Dla większości respondentów, wykres słupkowy przyniósł bardzo duże ułatwienie, a odpowiedź, że kolorowane mapy były wystarczające do porównania tych wartości wskazano zaledwie 2 razy.

Potencjalne źródła błędów eksperymentu

Przeprowadzony eksperyment mógł posiadać następujące słabości, z których mogły wyniknąć niereprezentatywne wyniki:

1. Ankietowany mógł nie znać geograficznego położenia państw, o które były pytania - jednakże, uwzględniane państwa są powszechnie znane, a grupa na której przeprowadzone było badanie, wg mojego założenia posiada podstawową wiedzę geograficzną. Ankietowani mieli możliwość zgłosić mi ten problem, jednak tego nie zrobili.
2. Ankietowany mógł nie znać angielskich nazw państw z wykresu słupkowego - z tego co wiem, badana grupa posiada wystarczającą znajomość angielskiego, aby to nie stanowiło problemu. Nie zgłoszono tego kłopotu.
3. Ankietowany mógł kierować się swoją wiedzą na temat piłki nożnej i być stronniczy wobec odpowiedzi, które wydawały mu się możliwe - w tym aspekcie trzeba liczyć na szczerść w odpowiedziach respondentów, sam temat piłki nożnej został tylko pobieżnie wspomniany w rzadko czytany opis ankiety, a nie w tytułach wykresów czy pytaniach, co powinno zmniejszyć wpływ tego błędu.
4. Ankietowany mógł odczytać prawidłową wartość już z mapy nr 1, potem powielić ją dla mapy nr 2, mimo że gdyby widział tylko mapę nr 2, nie wskazałby prawidłowej odpowiedzi - to problem, który można wyeliminować poprzez przeprowadzenie równoległego eksperymentu z zamienioną kolejnością map. Zważając jednak na ilość odpowiedzi 'Nie mogę stwierdzić' dla obu map, widzimy zależność, że jeśli ktoś nie znał odpowiedzi po pierwszej mapie, była duża szansa że pozna ją po mapie nr 2. Ręcznie przeglądając odpowiedzi, dowiemy się, że większość ankietowanych poprawiała swoją odpowiedź po ujrzaniu mapy nr 2.

Wnioski

Wnioski eksperymentu potwierdzają postawioną hipotezę - powszechnie uważane za poprawne skale kolorów na mapie, tzn. jedno-odcieniowe i kontrastowe, są bardziej przystępne w odczytywaniu z nich wartości. Skala kolorów z mapy nr 1 przysporzyła ankietowanym wyraźnych kłopotów w odczycie wartości, o które były wszystkie 3 pytania.

Ciekawym zagadnieniem jest porównanie kolorowania map do wykresu słupkowego. Oczywiście, dla takiego typu danych, różnicę między wartościami łatwiej odczytuje się z wykresu słupkowego, co pokazał eksperyment. Nie znaczy to jednak, że wykresy słupkowe powinny zastąpić kolorowane mapy dla tych danych - z map możemy odczytać inne zależności, o które akurat nie pytano w badaniu, takie jak położenie państw o najwyższych wartościach, czy średnią wartość dla każdego z kontynentów. Eksperyment dowiódł jednak, że jeżeli już przedstawiamy dane na kolorowanej mapie, powinniśmy zastosować skalę kolorów taką, jak z według znanych i dobrych praktyk.