Praca domowa 3 - raport

Mariusz Dwużnik

Kwiecień 2021

1 Abstrakt

Obiektem tego raportu jest analiza wyników ankiety przeprowadzonej w dniach 01-04.04.2021. Tematem ankiety była kwestia oceny poprawności i czytelności wykresów przez grupę złożoną z 20 osób w wieku 18-25 lat. Ankieta skłdała się z 10 pytań. Sześć z nich wymagało od ankietowanych oceny wykresu w skali 1-5, pozostałe cztery polegały na wybraniu wykresu, który ankietowany uważa za czytelniejszy. W dalszej części raportu znajduję się dokładniejsza charakterystyka uczestników badania oraz szczegółowe przedstawienie poszczególnych pytań, wyników ankiety oraz płynących z nich wniosków. Uważam, że eksperyment w znaczący sposób pokazał, czy stosowanie się do dobrych praktyk przy tworzeniu wykresów ma wpływ na ich odbiór.

2 Cel badawczy

Głównym celem eksperymentu było sprawdzenie czy wykresy zbudowane w oparciu o 'dobre praktyki', z którymi zapoznaliśmy się na wykładzie są odbierane korzystniej niż takie, które popełniają ewidentne błędy, bądź zakłamują obraz danych, które przedstawiają. Moglibyśmy się spodziewać, że wykresy wykonane w sposób poprawny powinny zebrać wyższe oceny oraz być częściej wybierane niż te, którym w kwestii poprawności można wiele zarzucić.

3 Omówienie eksperymentu

Pochodzenie danych

Dane wykorzystane do analizy zostały zebrane w formie anonimowej ankiety internetowej przez narzędzie google forms. Średni czas jej wypełnienia wyniósł cztery minuty.

Charakterystyka badanej grupy

Próbkę badawczą stanowiły ankiety wypełnione przez 20 osób. Nie była ona reprezentatywna jeśli chodzi o przekrój społeczny, co można szybko stwierdzić chociażby patrząc na rozkład wieku ankietowanych. Od siebie dodam jednak, że nie było to również puste badanie pośród osób uczęszczjących na ten przedmiot, a co za tym idzie zapoznanych z zasadami tworzenia wykresów, co wpłynęłoby negatywnie na wiarygodność badania. Pośród ankietowanych jedynie 4 osoby studiują na naszym wydziale.

W celu lepszego zobrazowania charakterystyki grupy wspomogę się wykresami wygenerowanymi automatycznie przez narzędzie google forms. Nie są one co prawda książkowym przykładem idealnych wykresów, ale w celu pomocniczym i w połączeniu z czytelną legendą wydają się wystarczające.



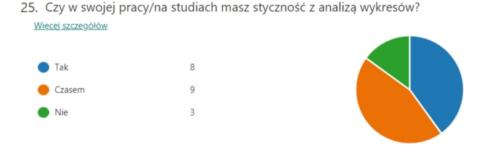
Rysunek 1: Rozkład wieku ankietowanych

Tak jak powiedziałem wcześniej, wszyscy ankietowani okazali się należeć do grupy wiekowej 18-25 lat. Najwięcej spośród nich było studentów - moich znajomych z liceum, ale wypowiedziało się również kilkoro tegorocznych maturzystów.



Rysunek 2: Rozkład płci ankietowanych

Jak widzimy na powyższym wykresie 65 procent ankietowanych to kobiety, a 35 - mężczyźni.



Rysunek 3: Obycie ankietowanych z wykresami

Powyższy wykres przedstawia ciekawą dla nas informację. Tylko 3 ankietowanych uważa, że w swojej pracy bądź na studiach nie analizuje wykresów. Ta stosunkowo niewielka liczba może wskazywać na dwie rzeczy. Po pierwsze, trzeba wspomnieć o tym, że nie doprecyzowałem co ma znaczyć określenie 'analizuje'. Praca z wykresami niekoniecznie musi oznaczać ich analizę matematyczną czy tworzenie. Każda osóba, która ma styczność z informacjami zawartymi na wykresach i wykorzystująca je w swojej codziennej pracy mogła uznać, że to wystarczy aby odpowiedzieć na pytanie twierdząco. Gdy to już wyjaśniliśmy, możemy pomyśleć nad najważniejszym wnioskiem. Najczęściej wybieraną opcją była odpowiedź 'czasem', gdy połączymy ją z niemal równie wysokim udziałem odpowiedzi 'tak', otrzymamy 85 procent osób stwierdziło, że ma w swojej pracy bądź na studiach kontakt z analizą wykresów. Pokazuje to, jak ważne są wykresy w codziennym życiu młodych ludzi w naszych czasach oraz daje silne podstawy do tego, żebyśmy mogli zbadać czy wraz z tak częstym obcowaniem z wykresami pojawia się intuicja jak poprawny wykres powinien wyglądać.

Typy zadanych pytań

Ankiete podzieliłem na dwie części.

W pierwszej z nich osoby ankietowane miały za zadanie ocenić w skali od 1 do 5 pięć wykresów i jedną grafikę(źródła dla wszystkich wykresów podam w sekcji na końcu). Cztery z wykresów były przedstawione na naszym wykładzie jako przykład wykresów z błędami. Istotnym dla mnie było, że były to zarówno błędy wpływające na czytelność wizualizacji, jak i błędy rzeczowe przy samym konstruowaniu wykresu. Chciałem sprawdzić, czy typ błędu będzie miał wpływ na odpowiednio niższą ocenę. Kontrolnie umieściłem wśród wykresów swój poprawiony wykres z pracy domowej 2, aby sprawdzić czy faktycznie przejdzie próbę bycia wystawionym pod szerokie głosowanie i zostanie oceniony lepiej niż pozostałe.

Druga część polegała na wyborze tej z dwóch przedstawionych wizualizacji tych samych danych, którą osoba ankietowana uznała za czytelniejszą. Ten rodzaj pytania miał wprost pokazać, który typ wykresu bardziej odpowiada odbiorcom. Wykresy zostały dobrane w ten sposób, aby jeden z nich był 'nietypowy', co zazwyczaj znaczyło wykres niezważający na sugerowane dobre praktyki, a drugi bardziej tradycyjny, bliższy zaleceniom. Pytania te miały zatem odpowiedź 'lepszą', której wyboru się spodziewałem, a celem eksperymentu było sprawdzenie, czy faktycznie głosy na tę odpowiedź będą przeważały.

4 Analiza poszczególnych pytań - wyniki

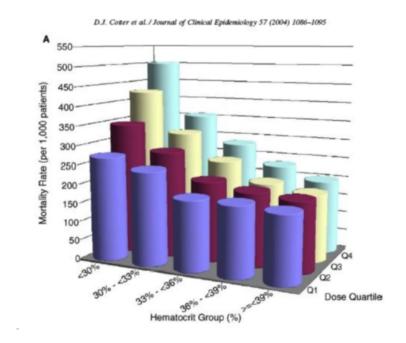
Przejdźmy zatem do analizy wyników eksperymentu. Konwencja tej części będzie następująca: Na pierwszej grafice będzie zawsze przestawiony wykres (lub wykresy), które widzieli ankietowani. Poniżej będzie się znajdował komentarz i mój zamysł dotyczący umieszczenia danego wykresu w badaniu. Kolejna grafika to wizualizacja wyników, również z komentarzem. //Aby uniknąć znacznego wpływu outlierów przy stosunkowo nielicznej grupie badanych, gdy będę mówił o średniej, będę posługiwał się średnią winsorowską o ucięciu równym 10 procentom (dwóm obserwacjom).

Pytania typu 'oceń wykres'

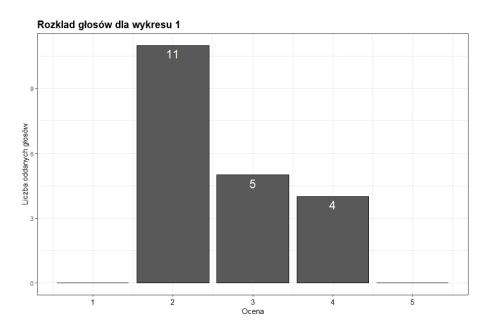
Przy pierwszym typie pytań polecenie brzmiało następująco:

"W pierwszej części Twoim zadaniem będzie subiektywna ocena wykresu (jak Ci się podoba, czy jest czytelny, czy dobrze spełnia swoją rolę) w skali 1-5 gwiazdek, gdzie: 1 - bardzo słaby, 2 - słaby, 3 - średni, 4 - dobry, 5 - bardzo dobry"

Pytanie pierwsze



Wykres ten widzieliśmy już na wykładzie. Popełnia on wiele błędów. Jest to wykres 3D, na który naniesiono cztery grupy danych. Mnogość słupków oraz zaburzona perspektywa sprawia, że nie jesteśmy w stanie odczytać poprawnie niemal żadnej wartości. Ciężko nawet określić wzajemną relację poszczególnych słupków. Z całą pewnością można stwierdzić, że nie jest to wykres szczególnie udany.

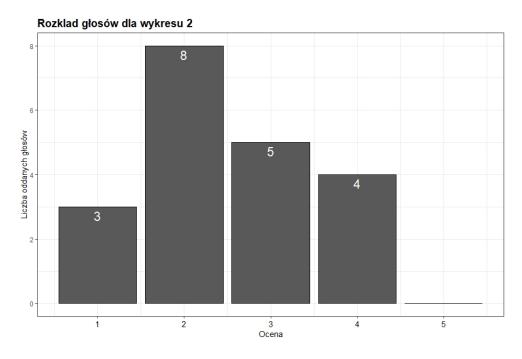


Jak widzimy znajduje to odniesienie w głosach osób ankietowanych. Aż 11 osób oceniło go na 2 punkty, co odpowiada ocenie 'słaby'. Średnia ocena wyniosła 2.65, co jak zobaczymy poniżej jest raczej niską wartością. Wykres zgodnie z przewidywaniami został oceniony dość surowo. Trzeba mu jednak oddać, że nikt nie zdecydował się ocenić go na 1. Zgodnie z prawdą, wykres nie manipulował danymi, a jedynie był nieczytelny.

Pytanie drugie



O ile przy poprzednim pytaniu mieliśmy do czynienia ze słabym wykresem, to ten wykres jawnie wprowadza w błąd. Skala nie zaczyna się od zera, kolor ma działać na rzecz tezy, za którą wykres ma przemawiać (wykres nie powinien mieć tezy!), a na dodatek na wykresie zostały umieszczone zarówno dane z poprzednich lat jak i prognozy na lata przyszłe, bez żadnego rozróżnienia jednych od drugich. Same trójwymiarowe słupki nie są dobrą praktyką, ale to najmniejszy problem tego wykresu.

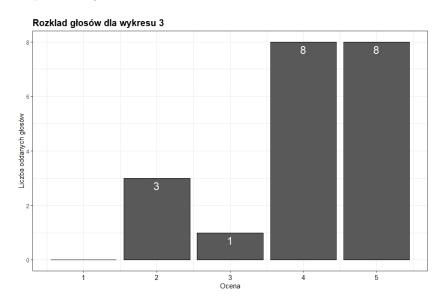


Podobnego zdania było większość ankietowanych. Widzimy że 11 osób uznało wykres za słaby lub bardzo słaby. Średnia ocena wyniosła 2.5, czyli niżej niż w przypadku poprzedniego wykresu, ale nie aż tak nisko na ile wykres zasługiwał. Aż 4 osoby oceniły wykres jako dobry. Moja interpretacja tych wyników jest taka, że te osoby prawdopodobnie nie zauważyły tego w jaki sposób wykres operuje danymi, bądź uznały, że to nie rzetelność jest obiektem badania, a sama czytelność wykresu i pod tym kątem go oceniły.

Pytanie trzecie

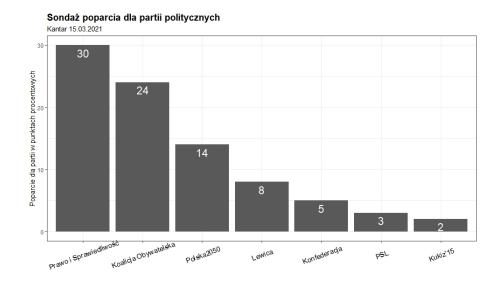


Wykres ten, mimo że na pierwszy rzut oka całkiem poprawny i czytelny, również znalazł się wśród wykresów z obecnością błędów, głównie poprzez niedopasowanie wartości odpowiadających poszczególnym słupkom i skali wykresu. Czytając wartości ze skali możemy zobserwować różnicę około 500 jednostek w porównaniu z liczbą nad słupkiem. Może to powodować niepewność co do tego, która z wartości jest poprawna. Intuicyjnie będziemy się kierować raczej liczbą nad słupkiem niż skalą, ale tego typu sytuacja nie powinna występować. Pewnym drobiazgiem, choć istotnym, jest też to, że czerwona krzywa wyznaczająca linię trendu nie była w tym wypadku potrzebna, jej dodanie też można uznać za pewien błąd.

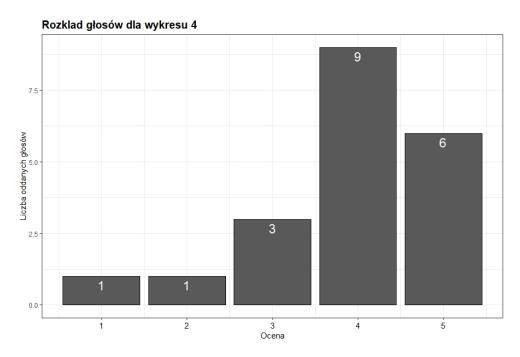


Wyniki ankiety są bardzo ciekawe. Aż 16 osób uznało wykres za dobry lub bardzo dobry. Jestem skłonny zaryzykować tezę, że najsurowsze oceny mogły należeć do osób, które ten wykres widziały wcześniej i usłyszały dlaczego jest nienajlepszy. (Jak wspomniałem wcześniej, ankietę wypełniały 4 osoby, które uczestniczyły w wykładzie). Dla wiekszości osób wykres ten sprawiał wrażenie niemalże wzorowego. Faktycznie,gdyby nie błąd ze skalą, możnaby było śmiało stwierdzić, że wykres jest czytelny i stworzony w myśl dobrych praktyk (zastosowanie dwuwymiarowych słupków, podpisanie wartości, tytuł i podtytuł), stąd gdy samego błędu nie zauważymy, wysokie oceny nie powinny dziwić. Na marginesie dodam, że mam hipotezę, że na wysoki wynik mógło też wpłynąć pochodzenie wykresu. Zauważmy, że poprzedni wykres, mocno zmanipulowany, pochodzi z programu telewizji polskiej, która nie cieszy się ostatnio dobrą renomą. Wykres trzeci pochodzi z konkurencyjnej stacji TVN24, dlatego niektórzy ankietowani mogli zadziałać kierując się zasadą dychotomii. Skoro przed chwilą mieliśmy do czyniania z naprawdę złym wykresem z TVP, a kolejny wykres wydaje się dobry i pochodzi ze stacji kojarzonej z większą rzetelnością, to można śmiało przyznać mu skrajnie wysoką notę. Trzeba jednak pamiętać, że z pewnością główną przesłanką przy ocenie wykresu był sam jego wygląd. Średnia ocena była bardzo wysoka i wyniosła 4.05.

Pytanie czwarte

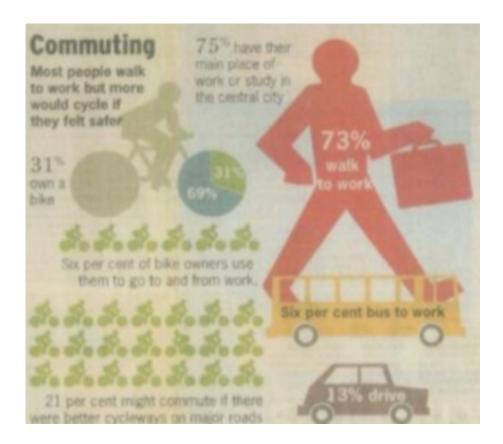


Powyżej widzimy wykres stworzony przeze mnie w ramach pracy domowej nr.2. Tworząc go i opisująć starałem kierować się dobrymi praktykami. Umieszczając go tutaj chciałem sprawdzić, czy faktycznie udało mi się stworzyć wykres, który będzie powszechnie uznany za czytelny. Aby nie podważyć wiarygodności badania nigdzie nie zamieściłem oczywiście informacji o tym, że wykres został stworzony przeze mnie.

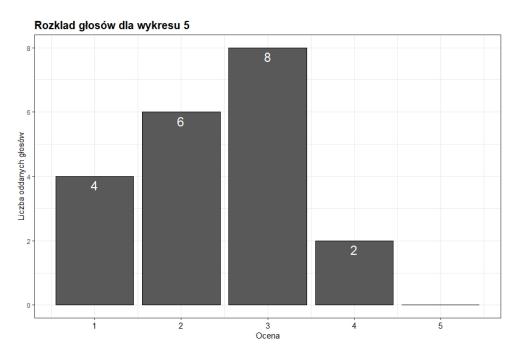


Z zadowoleniem mogę przyjąć, że wykres faktycznie został oceniony w przeważającej większości pozytywnie. Średnia ocena wyniosła 4.05, co na równi z poprzednim wykresem jest najwyższym wynikiem ze wszystkich. Myślę, że jego minimalistyczna estetyka była tym, co powstrzymało więcej osób przed przyznaniem mu oceny 5 (bogatsza jeśli chodzi o temat grafika Tvn24 otrzymała aż 8 ocen 5). Muszę jednak przyznać, że wynik, w którym 75 procent badanych uznaje wykres za dobry lub bardzo dobry jest dla mnie jasnym sygnałem, że po pierwsze, stosowanie się do dobrych praktyk faktycznie pozytywnie wpływa na odbiór wykresu, a po drugie, że dobrze wywiązałem się z poprzedniej pracy domowej (uff).

Pytanie piąte

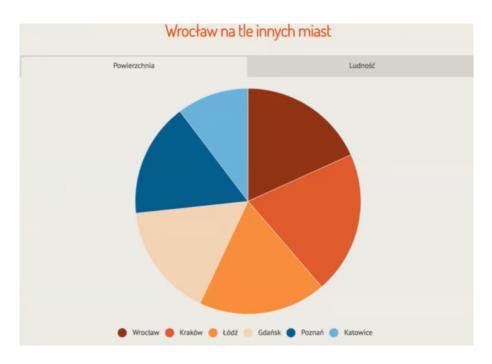


Pytanie piąte było trochę nietypowe, ponieważ tym razem oceniana była grafika, a nie wykres. Jak sami widzimy, dużo się na niej dzieje, rozrzuconych jest wiele informacji przedstawionych w oryginalne sposoby. Każdy może mieć swoja opinię, ale dla mnie jest ona raczej nieczytelna i raczej utrudnia niż ułatwia zrozumienie przekazywanych danych.

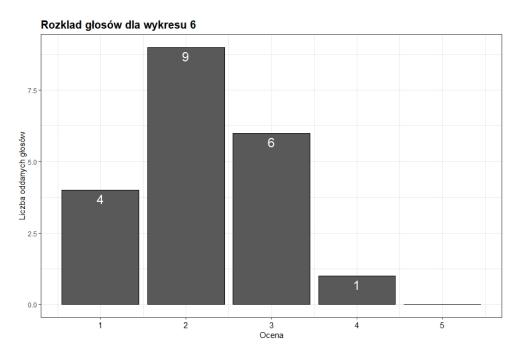


Podobnie grafikę tę ocenili ankietowani. Tak wiele ocen 3 może wskazywać na to, że ciężko było ją ocenić w jednoznaczy sposób. Widzimy jednak, że aż połowa osób uznała ją za słabą bądź bardzo słabą. Średnia ocena 2.3 jest najniższą ze wszystkich do tej pory.

Pytanie szóste



Ostatni z wykresów jest według mnie najgorzej przedstawiony ze wszystkich do tej pory. Widzimy próbę naniesienia na wykres kołowy danych o powierzchni 6 miast. Podstawowym i niewybaczalnym błędem jest to, że ten typ danych w ogóle nie powinien być przedstawiany na wykresie kołowym. Jeśli już używamy wykresu kołowego, to wtedy, gdy dane sumują się do jakiejś całości. W tym wypadku jedyne do czego sumuja się te dane, to suma powierzchni, co nie ma żadnej wartości interpretacyjnej w przypadku tego wykresu. Powierzchnie miast bardzo trudno jest porównywać, nie zostały także w żaden sposób opisane liczbami.



Z zadowoleniem przyjmuję, że na twarzy większości badanych także prawdopodobnie pojawił się grymas niezadowolenia, gdy zobaczyli ten wykres, na co wskazuje najniższa średnia ocena ze wszystkich, wynosząca 2.15. Wykres został uznany za słaby lub bardzo słaby przez aż 13 spośród 20 ankietowanych, a osoba, która uznała go za dobry z pewnością zrobiła to z zamkniętymi oczami, stojąc tyłem do komputera. Mówiąc trochę poważniej, tak niski wynik absolutnie nie dziwi i jest moim zdaniem uzasadniony.

Pytania typu 'wybierz lepszy wykres'

W drugim typie pytań zadaniem ankietowanych było wybranie wykresu, który ich zdaniem jest czytelniejszy. Aby uniknąc nieporozumień pozwoliłem sobie również wytłumaczyć co rozumiem poprzez określenie 'czytelniejszy'. Ankietowani mogli zobaczyć przed pytaniem taką krótką notkę:

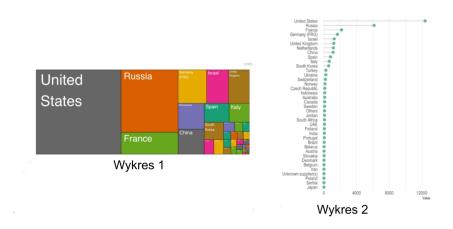
'Poprzez czytelność rozumiem łatwość odczytania danych z wykresu i zorientowania się w tym, co ma przedstawiać.'.

Tak jak powiedziałem wczesniej, w każdym pytaniu znalazł się jeden wykres bardziej egzotyczny i jeden bardziej tradycyjny (z małym wyjątkiem pytania drugiego, gdzie porównywałem odbiór wykresu 3D i 2D). Spodziewałem się, że wizualizacje konwencjonalne, a więc te, w których w większym stopniu kierowaliśmy się dobrymi praktykami, powinny zostać wybrane częściej.

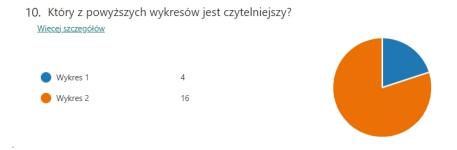
W celu wizualizacji wyników eksperymentu polegającego na wyborze jednej z dwóch opcji nie widziałem sensu w tworzeniu kolejnych wykresów słupkowych, dlatego posłużyłem się, podobnie jak przy metryczce, wykresami generowanymi automatycznie przez google forms. Moją intencją było skorzystanie z widocznej po lewej stornie legendy z liczbą oddanych głosów na poszczególne opcje niczym z tabelki.

Pytania 7,9 i 10 dotyczą wizualizacji tych samych danych.

Pytanie siódme

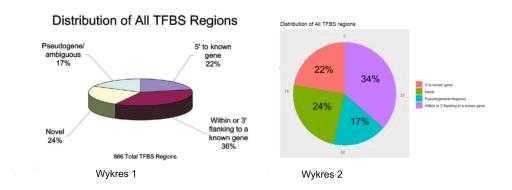


Pierwszy wykres, to wykres o zamyśle podobnym do kołowego, ale przedstawiony w inny sposób. Suma wszystkich danych przedstawiona jest jako pole dużego prostokąta, a poszczególne udział poszczególnych krajów obrazowany jest za pomocą pól coraz to mniejszych kolorowych prostokątów. Prowadzi to do tego, że realnie czytelne jest tylko kilka największych państw, nie jesteśmy za to w stanie powiedzieć jakie mniejsze państwa były brane pod uwagę. Drugi wykres przypomina z zamysłu wykres kolumnowy. Wszystkie kraje zostały wypisane po lewej stronie i przyporządkowane zostały im odcinki o odpowiedniej dla nich długości. Da sie ję ze sobą stosunkowo łatwo porównać. Wykres drugi wypada moim zdaniem lepiej niż pierwszy.

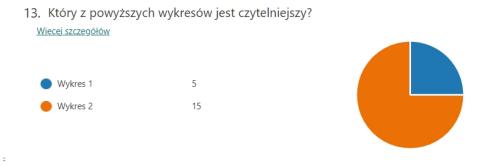


Ankietowani również byli tego zdania. Jak widzimy 80 procent z nich opowiedziało się za drugim wykresem. Wykres pierwszy zdobył jedynie czterech zwolenników. Warto te wyniki zapamiętać i porównać z tym, co będziemy mogli zaobserwować przy pytaniu dziewiątym.

Pytanie ósme

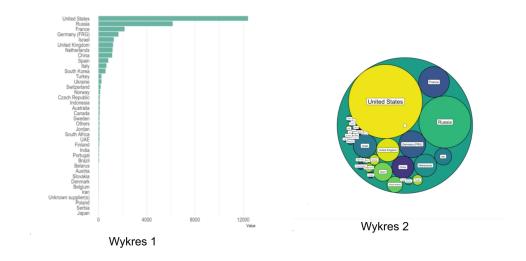


W tym pytaniu widzimy dwa podobne wykresy kołowe przedstawiające te same dane. Wykres pierwszy został znaleziony w internecie i jest wykresem 3D, a wykres drugi stworzyłem na jego podstawie za pomocą pakietu ggplot2 i jest wykresem 2D. Mimo skromnej szaty graficznej, drugi wykres powinien być czytelniejszy w myśl zasady, że 2D jest lepsze niż 3D. Przekonajmy się jak do tego podeszli ankietowani.

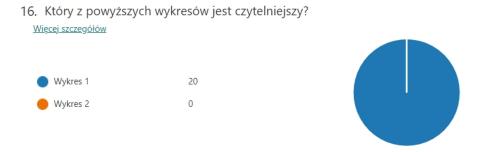


Wyniki te również mnie ucieszyły, ponieważ aż 75 procent osób wskazało mój wykres drugi jako czytelniejszy. Nie byłem przekonany że tak właśnie się stanie, ponieważ patrząc z zewnątrz wykres pierwszy może robić bardziej profesjonalne wrażenie. Ankietowani jednak nie dali się zwieść i znacząco przychylili się ku wykresowi drugiemu i to właśnie on został oceniony jako czytelniejszy.

Pytanie dziewiąte

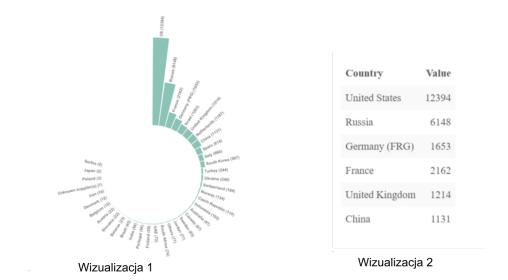


Pytanie dziewiąte jest moim zdaniem bardzo podobne do siódmego. Zarówno wykres tradycyjny w obu przypadkach opiera się na podobnym założeniu - wykresie słupkowym, który uwzględnia wartości przypisane dla każdego państwa, jak i wykres niekonwencjonalny opierał się na dużej figurze, w którą wpisane były mniejsze o odpowiendnio niższych wartościach.

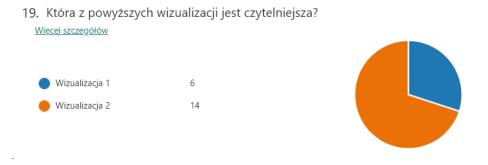


Wyniki jednak okazały się zaskakująco jednostronne. Każdy ankietowany zaznaczył, że wykres pierwszy jest dla niego lepszy niż drugi. Oczywiście w żaden sposób nas to nie dziwi, ponieważ wykres pierwszy jest faktycznie dużo czytelniejszy, zaskakująca jest jednak zgodność wśród ankietowanych, gdy w przypadku pytania siódmego aż 25 procent wybrało wykres prostokątny. Myślę, że można to wytłumaczyć tym, że ten wykres złożony z okręgów okazał się jednak czymś o krok za daleko i coś, co mogło wcześniej uchodzić za kreatywne podejście do wykresu, teraz stało stało się po prostu niepotrzebną komplikacją w zestawieniu z dobrze znanym, klasycznym wykresem słupkowym.

Pytanie dziesiąte



Ostatnie zestawienie również nie było do końca typowe. Zestawiłem ze sobą dośc rzadko spotykany typ wykresu, który możemy nazwać naniesieniem wykresu kolumnowego na okrąg oraz prostą tabelkę. Warto zauważyć, że tabelka daje możliwość zapoznania się z informacją, której pierwszy wykres nie daje, czyli z wartością przypisaną do danego kraju. Niepokój mogło jednak budzić to, że w tabelce znajdowało się jedynie kilka największych państw. Obiektem pytania była czytelność danej formy wizualizacji, a nie ilość obserwacji w niej zawartych, ale mogło to wpłynąć na ocenę przez ankietowanych.



Na koniec naszego eksperymentu widzimy zwycięstwo prostoty w postaci tabelki, nad niecodziennym i niekoniecznie czytelnym wykresem naniesionym na okrąg. 70 procent ankietowanych opowiedziało się za tabelką.

5 Podsumowanie i wnioski

Po zapoznaniu się z poszczególnymi pytaniami, możemy zająć się zebraniem wniosków i odniesieniu ich do naszej hipotezy.

Najważniejszym, pozytywnym wnioskiem, który towarzyszył mi przy całej analizie wyników ankiety, jest to, że wybory grupy, która brała udział w badaniu są zaskakująco bliskie mojej ocenie poszczególnych wykresów. Oczywiście taka zbieżność może mieć dwie strony, nie musi wcale świadczyć o słuszności naszych wyborów, bo może wskazywać na to, że zarówno ja jak i ankietowani w podoby sposób jesteśmy w błędzie. Biorac jednak pod uwage cel, który przyświecał mi przy tworzeniu ankiety i fakt, że dobierając wykresy opierałem się na założeniach wspomnianych już wielokrotnie w tym raporcie 'dobrych praktyk' mogę postawić teze, że ankietowani w przeważającej wiekszości trafnie ocenili poszczególne wykresy z cześci pierwszej oraz dokonali dobrych wyborów w cześci drugiej. Ankietowani faktycznie najwyżej ocenili wykresy, o których można powiedzieć, że zostały zbudowane w najbardziej poprawny sposób. Co więcej, dobrze ten fakt podkreśla bardzo duża różnica w średniej ocenie pomiędzy dwoma pierwszymi pod względem średniej oceny wykresami, a kolejnym. Dla tych pierwszych średnia ocena wyniosła 4.05, pozostałe średnie zaś mieściły się w przedziale 2.15-2.65. Zastrzeżenie może budzić jedynie tak wysoka średnia dla wykresu z Tvn24, jednak powody tego, dlaczego tak się stało, opisałem w części poświeconej tamtemu wykresowi. Streszczając - mamy wszelkie powody aby twierdzić, że dwa najwyżej ocenione wykresy faktycznie sa najlepsze i najczytelniejsze, co moim zdaniem powinno być przyjęte z dużą dozą optymizmu jeśli chodzi o intuicję osób wypełniających ankietę.

Niegorzej jest w przypadku drugiej części. W każdym przypadku wybrana została odpowiedź, której mogliśmy się spodziewać. Przy pytaniu dziewiątym nawet jednogłośnie! Jest to kolejny sygnał w kierunku tezy, że dobre praktyki faktycznie mają znaczenie.

Podsumowując, wyniki ankiety oceniam bardzo pozytywnie. Oceny były różnorodne, ale tak jak opisałem powyżej, w znaczący sposób wskazywały, że wykresy tworzone w myśl dobrych praktyk, faktycznie mogą cieszyć się bardziej pozytywnym odbiorem.

6 Źródła

Poniżej zamieszczam linki do stron z wykresami, którymi posiłkowałem się przy swoim badaniu.

https://github.com/mini-pw/2021L-ExploratoryDataAnalysis/tree/main

https://iase-web.org/islp/apps/gov_stats_graphing/GoodBad/GoodBadGraphs.pdf

https://www.data-to-viz.com/story/OneNumOneCat.html

https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/