

Wpływ wybranych statystyk na poziom szczęścia społeczeństwa

Joanna Kołaczek, Adam Wrzesiński

2022-12-18

Spis treści

Wstęp	3
Przedstawienie danych	3
Cel analizy	4
Analiza i wizualizacja danych	5
Podsumowanie	9

Wstęp

Niniejsze sprawozdanie zostało sporządzone w ramach laboratorium Pakietów Statystycznych, prowadzonych przez dr inż. Daniela Kucharczyka¹, do wykładu dr Andrzeja Giniewicza². Naszym zadaniem jest analiza wybranych danych rzeczywistych, która posłuży do rozwiązania ustalonego problemu badawczego. Do wykonania naszego raportu wykorzystamy pakiet R wraz z jego wybranymi bibliotekami, których krótki opis zawarliśmy w Tabeli 1.

Tabela 1: Biblioteki używane w raporcie

nazwa	opis
corrplot	Wizualizacja korelacji między zmiennymi w postaci macierzy korelacji lub wykresu słupkowego.
cowplot	Łatwiejsze komponowanie wielu wykresów na jednym rysunku.
dplyr	Wykonywanie prostych i wydajnych operacji na tabelach danych.
ggplot2	Tworzenie estetycznych i profesjonalnych wykresów i wizualizacji danych.
rnaturalearth	Pobiera i przetwarza dane o kształcie i granicach krajów oraz innych obszarów geograficznych na całym świecie.
rnaturalearthdata	Dostarcza dodatkowo wiele innych danych geograficznych.

Przedstawienie danych

Wykorzystamy dane zestawiające estymowany wpływ poszczególnych zmiennych na Poziom Szczęścia społeczeństwa w poszczególnych krajach w latach 2015-2020 zebranych przez instytut Gallupa³, wraz z postrzeganym poziomem korupcji CPI (z ang. *corruption perceptions index*) pozyskanymi przez pozarządową organizację Transparency International⁴ w 2015 roku. Konkretny zestaw pobraliśmy z platformy Kaggle, gdzie jeden z użytkowników połączył raporty w jeden zbiór⁵ składający się z następujących zmiennych:

- **Poziom szczęścia** - średnia odpowiedzi na pytanie dotyczące oceny życia,
- **Państwo** - nazwa państwa do którego odnoszą się pozostałe dane,
- **Kontynent** - kontynent na którym znajduje się dane państwo,
- **Rok** - rok z którego pochodzą zebrane dane,
- **PKB na osobę** - zakres, w jakim PKB na mieszkańca przyczynia się do obliczenia poziomu szczęścia. Jest to stosunek wartości PKB w cenach bieżących do liczby ludności danego obszaru,
- **Rodzina** -stopień, w jakim rodzina przyczynia się do obliczenia poziomu szczęścia. Jest to średnia odpowiedzi w danym kraju, na pytanie GWP „Czy będąc w potrzebie, możesz zawsze liczyć na krewnych bądź przyjaciół?”,
- **Zdrowie** -stopień, w jakim oczekiwana długość życia przyczynia się do obliczenia poziomu szczęścia. Estymowana długość życia, obliczana na podstawie danych Światowej Organizacji Zdrowia,
- **Wolność** -stopień, w jakim wolność przyczynia się do obliczenia poziomu szczęścia. Uśredniona odpowiedź na pytanie GWP „Czy jesteś usatysfakcjonowany z poziomu wolności wyboru, do tego co chcesz robić w swoim życiu?”,
- **Szczodrość** -stopień, w jakim szczodrość przyczynia się do obliczenia poziomu szczęścia. Równa residuum z regresji średniej krajowej odpowiedzi na pytanie „Czy w ostatnim miesiącu przekazałeś pieniądze na cele charytatywne?”. Przeliczony w stosunku do PKB na osobę,
- **Zaufanie do rządu** - stopień, w jakim zaufanie do rządu przyczynia się do obliczenia poziomu szczęścia. Wnioskowany na podstawie odpowiedzi na pytania „Czy korupcja jest rozpowszechniona

¹: <http://prac.im.pwr.edu.pl/giniew/doku.php>

²: <https://dkucharc.github.io/academic/>

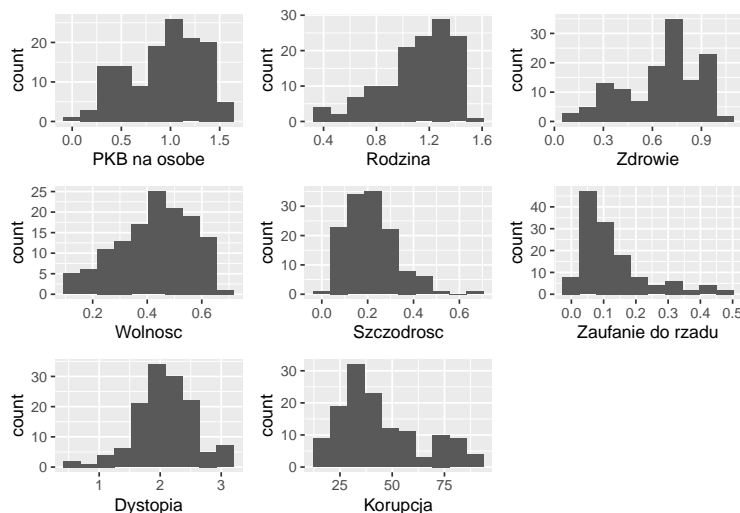
³: <https://worldhappiness.report/archive/>

⁴: <https://www.transparency.org/en/cpi/2021>

⁵: <https://www.kaggle.com/datasets/eliasturk/world-happiness-based-on-cpi-20152020>

w rządzie?” oraz „Czy korupcja jest rozpowszechniona w przedsiębiorstwach?”. W przypadku braku danych dotyczących korupcji w rządzie, jako ogólną miarę postrzegania korupcji stosuje się percepcję korupcji w biznesie.

- **Korupcja** - Postrzegany poziom korupcji w państwie. Im wyższa punktacja, tym niższa korupcja.



Wykres 1: Poziom szczęścia na świecie

Cel analizy

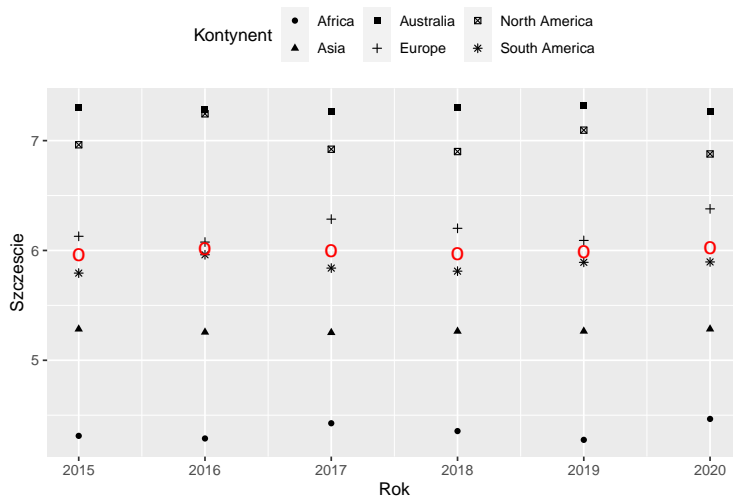
Naszym pierwszorzędnym priorytetem będzie znalezienie korelacji między *poziomą korupcją* a *poziomym szczęściem*. Oczekujemy, że w państwach, gdzie żyje się gorzej, panuje wysoka korupcja. W tym celu zbadamy globalny *poziom szczęścia* i znajdziemy kraje, w których przyjmuje on skrajne wartości. Następnie przyjrzymy się jak poszczególne statystyki wpływają na *szczęście* dla tych państw. Dodatkowo sprawdzimy, jak zmieniał się estymowany wpływ korupcji na *szczęście* w czasie. Na podstawie tych informacji wyciągniemy wnioski.

Tabela 2: Pierwsze rzędy zestawu danych

Państwo	Szczęście	PKB na osobę	Rodzina	Zdrowie	Wolność	Szczodrość	Zaufanie do rządu	Dystopia	Kontynent	Rok	Korupcja
Norway	7.537	1.616463	1.533524	0.7966665	0.6354226	0.3620122	0.3159638	2.277027	Europe	2015	88
Denmark	7.522	1.482383	1.551122	0.7925655	0.6260067	0.3552805	0.4007701	2.313707	Europe	2015	91
Iceland	7.504	1.480633	1.610574	0.8335521	0.6271626	0.4755402	0.1535266	2.322715	Europe	2015	79
Switzerland	7.494	1.564980	1.516912	0.8581313	0.6200706	0.2905493	0.3670073	2.276716	Europe	2015	86
Finland	7.469	1.443572	1.540247	0.8091577	0.6179509	0.2454828	0.3826115	2.430182	Europe	2015	90
Netherlands	7.377	1.503945	1.428939	0.8106961	0.5853845	0.4704898	0.2826618	2.294804	Europe	2015	84

Analiza i wizualizacja danych

Najpierw zobaczymy jak wyglądają dane. Wnioskujemy z opisu danych (zrobić odniesienie), że kolumny *Rodzina* i *Wsparcie społeczne* mówią o tym samym i uzupełniają się wzajemnie. W związku z tym wartości z *wsparcia* dopisujemy do *rodziny* i usuwamy niepotrzebną kolumnę. Braki danych w zestawie są opisane jako zera. Do znalezienia ich użyjemy funkcji *which* użytej do każdej kolumny naszego zbioru. Zwróci ona pozycję zer. Gdziekolwiek te braki danych mogą nieść pewną dodatkową informację. Przykładowo, jeśli napotkamy zera w kolumnie *Wolność* może oznaczać brak możliwości umieszczenia konkretnego pytania w ankiecie z powodów politycznych. Zamieniamy te wartości na *NA*, by nie przeszkadzały w liczeniu średnich wartości w dalszej części sprawozdania. Kilka pierwszych rzędów przekształconego w ten sposób zestawu umieszczamy w tabeli 2. Zauważmy, że estymowane statystyki sumują się do *Szczęścia*. Po użyciu funkcji *unique* dla kolumn Państw i kontynentów dowiadujemy się, że mamy do dyspozycji dane pochodzące ze 132 państw, rozmieszczonych na wszystkich kontynentach zebrane w latach 2015-2020. Zobaczymy jak na przestrzeni czasu wyglądało *Szczęście* dla każdego z kontynentów 2. Kolorem czerwonym została oznaczona średnia z całego świata.

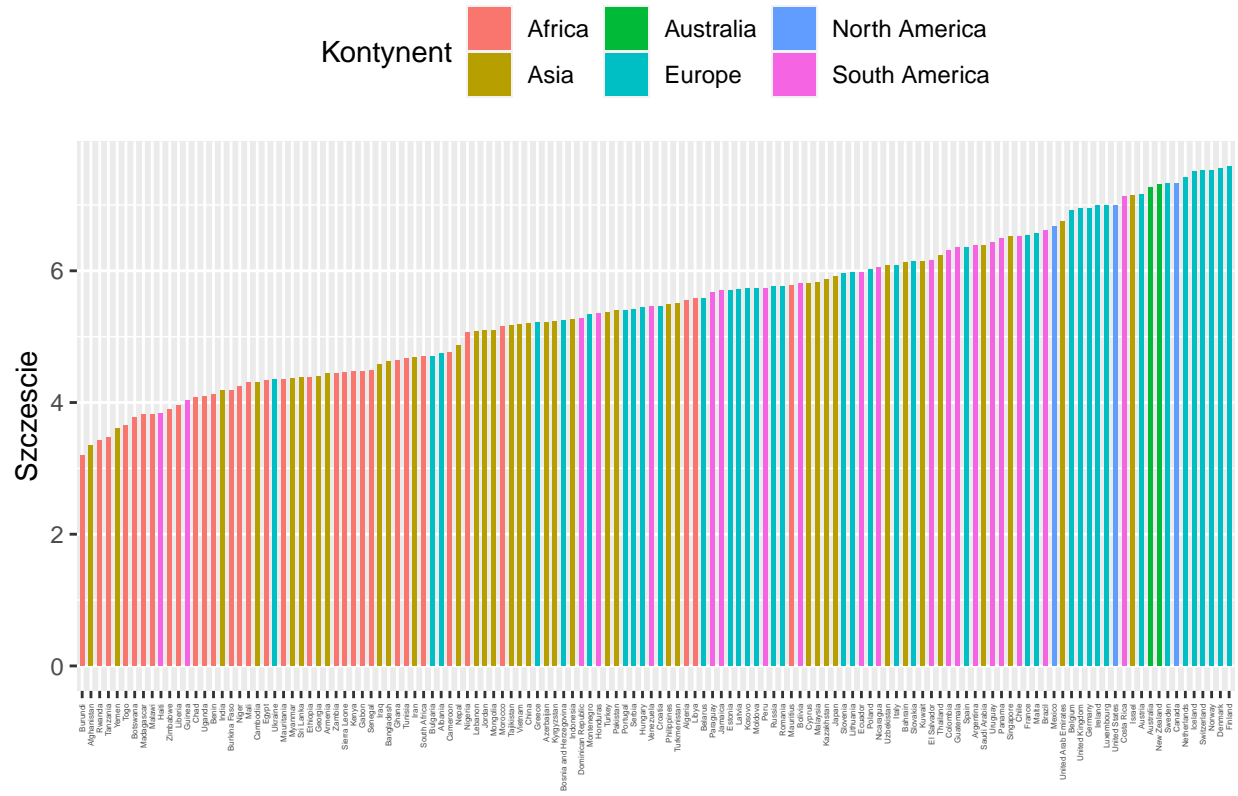


Wykres 2: Średnie szczęście na przestrzeni lat

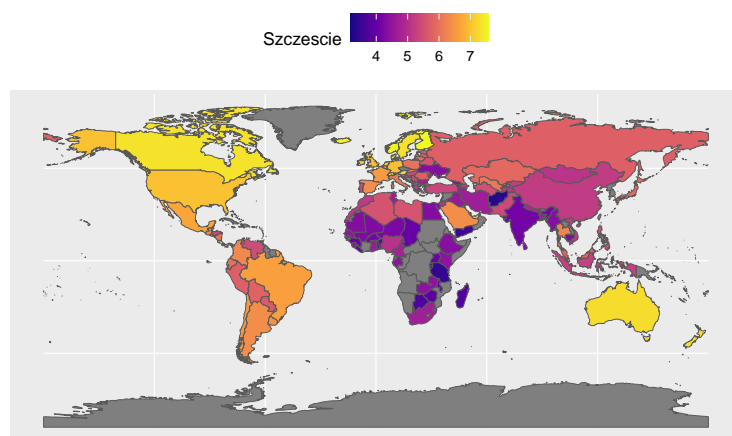
Widzimy, że dane nieznacznie fluktuują zatem uśrednimy je w celu bardziej przejrzystej analizy. Średni *poziom szczęścia* każdego państwa umieszczamy na wykresie 3, oznaczając kolorem kontynent w którym się ono znajduje. Zauważamy, że w niższej strefie dominują kraje z Afryki i Azji, natomiast w wyższej - Australii i Ameryki Północnej. Nanieśmy więc średni *poziom szczęścia* na mapę świata (rys. 4). Nasze przypuszczenia się potwierdzają. Jest stosunkowo dużo braków danych w Afryce. Biorąc pod uwagę jednak to, że dosyć podobnie rozłożony jest tam średni *poziom szczęścia*, nie traktujemy tego jako problem. Na uwagę zasługuje Afganistan, który stanowi pewnego rodzaju ciemną wyspę na tle Azji. Podobnie Ukraina na tle Europy.

Tabela 3: Średnie wartości statystyk na kontynentach

Kontynent	Szczęście	PKB na osobę	Rodzina	Zdrowie	Wolność	Szczodrość	Zaufanie do rządu	Dystopia	Korupcja
Africa	4.344436	0.5476496	0.8394981	0.3614624	0.3621062	0.2013179	0.1038954	1.928492	34.46774
Asia	5.294185	0.9513222	1.0013320	0.6464492	0.4397155	0.2550675	0.1366757	1.863819	38.51724
Australia	7.292175	1.3539607	1.3661302	0.9016853	0.6155996	0.4383893	0.3761506	2.240290	83.50000
Europe	6.217848	1.2345014	1.2379599	0.8010944	0.4444672	0.2260501	0.1579771	2.151797	60.53793
North America	7.045225	1.3069333	1.2101162	0.8253695	0.5296117	0.2990196	0.2075012	2.666684	60.50000
South America	5.885511	0.8749319	1.1200259	0.6583502	0.4613355	0.2006185	0.1010448	2.469186	36.35000



Wykres 3: Poziom szczęścia dla danych krajów

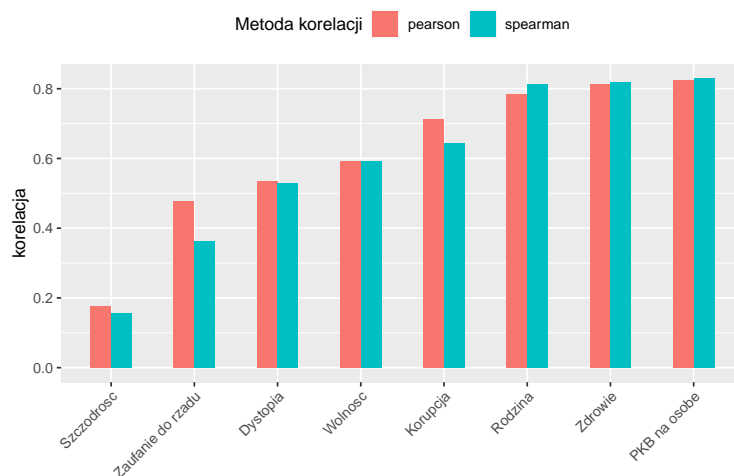


Wykres 4: Poziom szczęścia na świecie

Przejdźmy teraz do analizy korelacji między poziomem *szczęścia* a wybranymi czynnikami. Przypomnijmy po krótku na czym będą polegały użyte w raporcie metody.

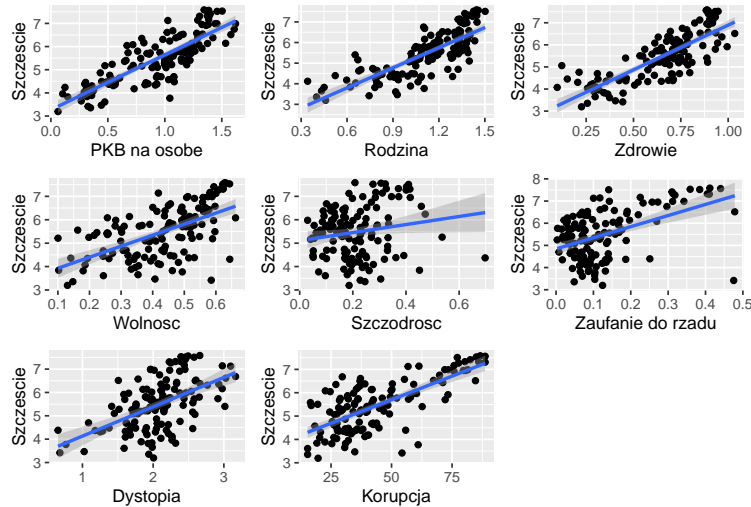
- Korelacja Pearsona mierzy stopień zależności liniowej pomiędzy dwoma zmiennymi.
- Korelacja Spearmana mierzy stopień korelacji pomiędzy dwoma zmiennymi na podstawie ich pozycji w rankingu. Może być obliczana dla danych, które nie są dobrze opisane przez prostą, ale istnieje między nimi jakiś rodzaj zależności.

Obie wymienione wyżej korelacje przyjmują wartości z przedziału od -1 do 1. Im bliżej zera, tym mniej skorelowane są dane, natomiast im bliżej wartości brzegowych, tym korelacja jest silniejsza (odpowiednio ujemna i dodatnia).



Wykres 5: Korelacja poziomem szczęścia a wybranymi statystykami

Wykres 5 przedstawia wielkość korelacji Pearsona i Spearmana między poziomem *szczęścia* a czynnikami z naszego zbioru danych. Oba dają podobne wyniki, zatem możemy się spodziewać zależności liniowej. Aby lepiej przyjrzeć się występującym zależnościom, przedstawimy badane dane na wykresach punktowych.

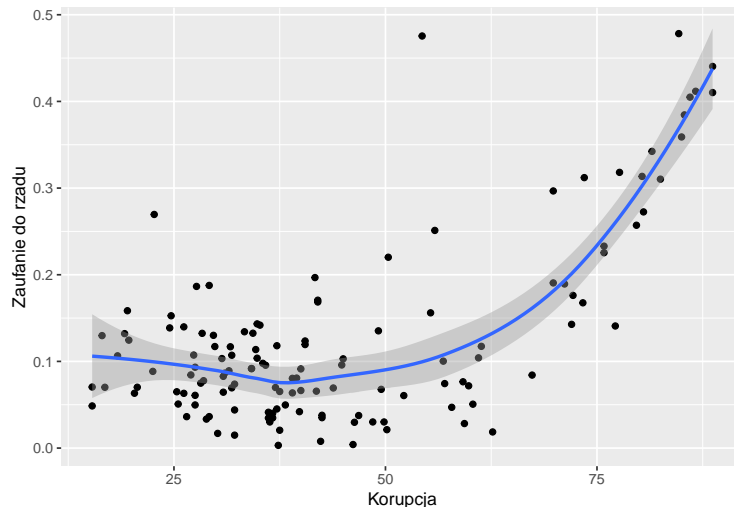


Wykres 6: Poziom szczęścia na świecie

Na wykresach 6 traktujemy poszczególne czynniki jako zmienne opisujące, natomiast *szczęście* jako zmienną opisywaną. Dodatkowo prosta jest wynikiem zastosowania regresji liniowej. Patrząc na otrzymane wykresy oraz korelacje widzimy, że największą zależność ze *szczęściem* mają *rodzina*, *zdrowie* i *PKB na osobę*, czyli powszechnie uznawane wartości. Jeżeli te czynniki będą utrzymane na wysokim poziomie, spodziewamy się że *szczęście* również takie będzie. Najgorzej w zestawieniu wypada *szczodrość* - nie możemy wnioskować czy jeżeli ta będzie rosła, rosła będzie również nasza zmienna opisywana, regresja liniowa w tym przypadku nie jest uzasadniona. Uwagę może zwrócić również fakt, że korelacja *szczęścia* z *korupcją* jest istotnie wyższa niż korelacja *szczęścia* z *zaufaniem do rządu*, gdzie oba czynniki mierzą poziom korupcji, przy czym to drugie bada jej wpływ na *szczęście*. Pokazuje nam to, że nie zawsze korelacja oznacza przyczynowość.

##

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Wykres 7: Zestawienie statystyk dotyczących korupcji

Aby zobaczyć dokładniej jak ma się *korupcja* do *zaufania do rządu*, przyjrzymy się wykresowi 7. Widzimy,

że gdy korupcja jest wysoka (tzn. przyjmuje niskie wartości na osi), jej wpływ na szczęście jest nikły, ale nie występuje tu oczywista zależność, natomiast gdy zaczyna ona spadać (dla wartości większych od 50), jej wpływ na szczęście staje się bardziej istotny. W tym przypadku możemy próbować doszukiwać się zależności innej niż liniowa, chociażby oznaczonej.

Podsumowanie