Wstęp do eksploracji danych

Matematyka i Analiza Danych, II rok

Anna Kozak

Data Scientist | Data Visualization | Responsible Machine Learning

Data Scientist (5 years)
Research and teaching assistant - MI2 Lab (3 years)
Associate Teacher - WUT (1 year)

Warsaw RUG Meetup - Spotkania Entuzjastów R
R-Ladies Warsaw
Koło Naukowe Data Science PW

💬 MS Teams
anna.kozak@mini.pw.edu.pl

PhD student | Computer Science

AutoML | Meta-learning | RML

Research & Business Experience

Data Scientist (5 years)

Research assistant - MI2 Lab (3 years)

Coordinator of Case Studies 2020/2021

MS Teams
katarzyna.woznica.dokt@pw.edu.pl

Wstęp do eksploracji danych składa się z:

- wykładu
- zajęć laboratoryjnych

Wykłady - wtorki 12:15

Laboratoria - wtorki 14:15, 16:15

Konsultacje - prośba o kontakt

Wykład

Na wykładzie będą przedstawione teoretyczne aspekty pracy z danymi, jak i praktyczne.

Materiały

1. https://github.com/MI2-Education/2022L-ExploratoryDataAnalysis
2. Biecek Przemysław, Odkrywać! Ujawniać! Objaśniać! Zbiór esejów o sztuce prezentowania danych, http://www.biecek.pl/Eseje/index.html
3. Rosling Hans, Factfulness
4. Tufte E.R., The Visual Display of Quantitative Information
5. Biecek Przemysław, Przewodnik po pakiecie R

Ocena

- prace domowe (40 pkt)
- projekty (44 pkt)
- wejściówki (6 pkt)

Suma 90 pkt

Ocena	3	3.5	4	4.5	5
Punkty	(50, 60]	(60, 70]	(70, 80]	(80, 90]	(90, ∞)

Z każdego projektu należy uzyskać ponad 50% możliwych punktów.

Laboratorium

- praca w R i Python (głównie R)
- powtórzenie operacji na danych (R: dplyr, tidyr, forcats; Python: numpy, pandas)
- wstęp do narzędzi pozwalających na estetyczne prezentowanie danych (R: ggplot2 + rozszerzenia, Python: matplotlib, seaborn)
- przygotowywanie interaktywnych wizualizacji i raportów (R: Rmarkdown, plotly, shiny)
- różne sposoby oceny zmiennych, danych, wizualizacji

Cel zajęć projektowych

- wykorzystanie i utrwalenie zdobytej wiedzy z wykładu oraz laboratoriów
- praktyczna praca z danymi
- ćwiczenie sposobu prezentacji wyników

Zasady

- 2 projekty w ciągu semestru
- zespoły 3 osobowe, różne podczas 1 i 2 projektu
- projekt trwa 5-6 tygodni
- 24 pkt i 20 pkt (w tym 5 pkt za pracę na zajęciach projektowych)

Projekt 1

Zadanie: Przygotowanie plakatu na zadany temat.

Rezultat: Plakat w formacie A2 w wersji pdf.

Zajęcia:

- wspólne dyskusje
- prezentacje kolejnych etapów

CENTRUM
NAUKI
KOPERNIK

Za tydzień

- podział na zespoły 3 osobowe w obrębie grup laboratoryjnych

Ocena

Za projekt można otrzymać od 0 do 24 punktów, z czego:

- 5p (1 x 1p, 2 x 2p) uzyskuje się za przedstawienie postępu prac w danym tygodniu
- 5p uzyskuje się za przygotowanie estetycznych wykresów (dwa lub więcej)
- 5p uzyskuje się, jeżeli przygotowane wykresy mają wszystkie niezbędne elementy do poprawnego odczytania danych (tytuł, podtytuł, adnotacje na osiach, legenda, jednostki, opis jak czytać wykres)
- 5p uzyskuje się za estetykę i pomysłowość aranżacji wykresów i opisów w jedną całość
- 4p uzyskuje się za prezentację projektu

Projekt 2

Zadanie: Przygotowanie interaktywnego raportu .

Temat: Spojrzenie na klimat i środowisko.

Rezultat: Raport lub aplikacja Shiny.

Zajęcia:

- wspólne dyskusje
- prezentacje kolejnych etapów

Ocena

Za projekt można otrzymać od 0 do 20 punktów, z czego:

- 5p (1 x 1p, 2 x 2p) uzyskuje się za przedstawienie postępu prac w danym tygodniu
- 7p uzyskuje się za postawienie pytań badawczych oraz znalezienie adekwatnych danych
- 6p uzyskuje się za jakość wizualizacji, tabel, opisów, element interaktywności (wykresy mają wszystkie niezbędne elementy do poprawnego odczytania danych (tytuł, podtytuł, adnotacje na osiach, legenda, jednostki, opis jak czytać wykres))
- 2p uzyskuje się prezentację projektu

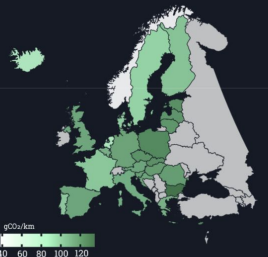
Plakaty, które zmieniają spojrzenie na klimat i środowisko

https://www.atmoterm.pl/ochrona-srodowiska-jako-problem-globalny/

BRIGHT RIDE

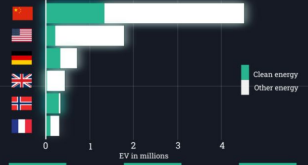
THE LIGHT AT THE END OF THE TUNNEL FOR OUR ENVIRONMENT

Average gCO_2/km emissions by cars registered in 2020



Why does Norway occupy the first position in terms of emitting CO_2 by newly registered cars in Europe with 38.2 gCO_2/km emissions? Not only are there electric cars relatively cheap due to lowering taxes in EVs but drivers also benefit from reduced road and ferry tolls, as well as discounted parking. In contrary to Norwegian pro-ecological attitude bulgaria is among the countries of the EU with the smallest number of charging stations for electric vehicles per 100 kms. of roads that leads to having the biggest average emission of carbon dioxide per km which indicates 134.4 gCO_2/km.

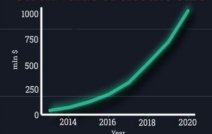
Number of EV that drive on clean energy



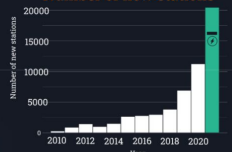
Some say that the answer to the emission problems are electric vehicles. The concept of clean energy makes us believe that we actually save our environment by choosing it over other energy sources. The energy used to charging electric cars mostly comes from conventional energy sources. In fact, only Norway produces the proper amout of EV to serve electric cars in green way.

Due to the new legislation that focuses on limiting CO_2 emissions and technology breakthroughs, electric cars are peaking on popularity. More and more motor companies invest in this branch meanwhile limiting their gasoline and diesel car production, but is the market ready for this sudden changes?

Stock value of electric cars



Number of new stations



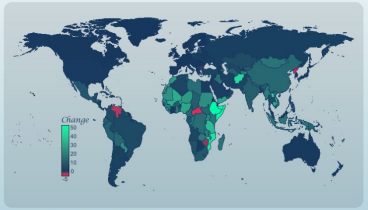
Critics of eletric vehicles say that there is a problem with insufficient number of charging stations. But is it still a problem? The chart above presents number of new charging stations in the United States each year. As one may see, that number grows almost exponentially. Due to the fact that data was collected only to September 2021, the number of new charging stations in this year may significantly increase.

Wojciech Kosiuk | Szymon Matuszewski | Michał Mazuryk

Source: eea.europa.eu/ | afdc.energy.gov | iea.org

Sea the Change

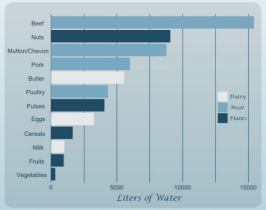
Change In Access To Improved Water
since 2000 (in percentage points)



https://ourworldindata.org/explorers/water-and-sanitation

With just a few exceptions, the access to improved drinking water is on the rise. The improvement is especially visible in Africa and south-east Asia. Despite these changes, the access in those regions is still not on par with the access seen in developed countries where it is almost universal.

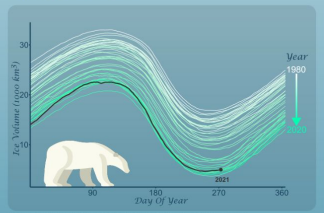
Water Used In Food Production
per kilogram of product



https://waterfootprint.org/media/downloads/Report-48-WaterFootprint-AnimalProducts-Vol1_1.pdf

Agriculture is the largest water user worldwide, accounting for 70 percent of total freshwater withdrawals, but these amounts can reach as much as 95 percent in some developing countries. It is 3496 litres of water 'eaten' per person, per day.

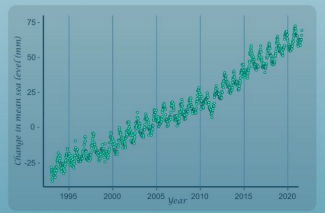
Arctic Ice Volume
between 1980 and 2020



http://psc.apl.uw.edu/research/projects/arctic-sea-ice-volume-anomaly/data/

Polar ice caps are melting as global warming causes climate change. We lose Arctic sea ice at a rate of almost 13% per decade. If emissions continue to rise unchecked, the Arctic could be ice-free in the summer by 2040. But what happens in the Arctic does not stay in the Arctic. Sea ice loss has far-reaching effects around the world.

Global Mean Sea Level
between 1992 and 2021



https://sealevel.colorado.edu/index.php/data/2021rel2-0

The rising water level is mostly due to a combination of melt water from glaciers and ice sheets as well as thermal expansion of seawater as it warms. In urban settings along coastlines around the world, rising seas threaten infrastructure necessary for local jobs and regional industries.

Jan Krezel Maja Andrzejczuk Nikola Miszalska

Calamintha nepeta

Bee aware!

Number of beehives worldwide

The world map shows which countries have the highest ratio of beehives per square km of their area, with South Korea as the leader.

Number of beehives per square km

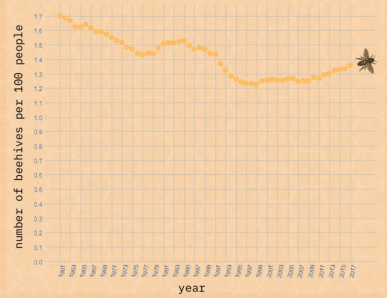
15
10
5

Origano

What can we do to help our little friends?
There are fewer and fewer green areas in our neighborhood. An easy way to help bees to survive is planting in our gardens some flowers they could visit. Before that, you need to learn which plants our insects are fond of. In the diagram on the right, you can find their favourite ones. The choice is wide so you will surely find the one that will also suit your taste.

Bees' favourite flowers

Population of bees in the years 1961-2017



Borago officinalis *Cirsium arvense (spinosissimum)* *Borago*

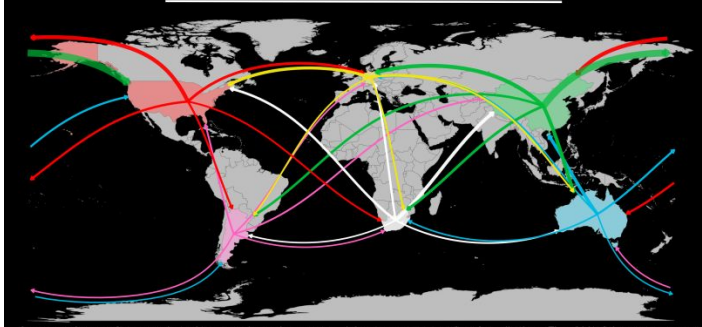
How does the situation of bees present overtime?
Looking into data, the number of beehives is steadly increasing. Sadly it doesn't suffice the growth in number of people worldwide. The amount of beehives per person is still smaller than 50 years ago.

Sources: fao.org/faostat/en/#search/beehive
link.springer.com/article/10.1007/s10841-019-00177-3/tables/1
worldometers.info/world-population/world-population-by-year

Authors: Agata Kopyt, Zuzanna Kotlińska, Jan Kruszewski

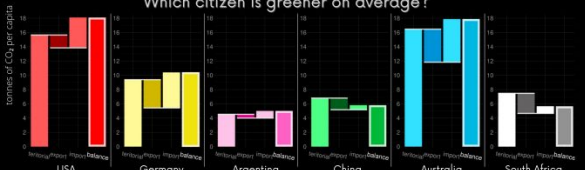
The hidden emissions

CO_2 in production and trade



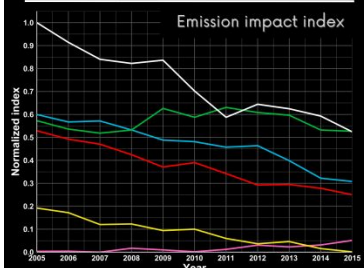
Arrows coming out of a country have the same color as the country and they represent export from it to all others. The width of the arrow represents the amount of CO_2 exported in 2015. From this chart we can clearly see that some countries, like China, export much more than they import, while other countries, like Australia, do the opposite. Those exports and imports are usually omitted when calculating countries' emissions. This leads to significant changes in total emissions.

Which citizen is greener on average?



The other factor that usually is not taken into account is the number of people living in the country. Although the map would imply that China is the biggest polluter in the world, it does not necessarily reflect the whole truth. The chart above presents carbon dioxide values per capita in 2015. From the per capita perspective there seems to be a discrepancy between what general public reckons and what the numbers say. As one can observe, it would appear that USA CO_2 demand surpasses the Chinese by a whooping 10 tonnes disparity.

Emission impact index



Previously mentioned data depicts the current situation but it does not take into consideration the standard of living. Our custom index aims to correct this by including GDP as a measure of a country's wealth.
The chart on the left represent the value of our custom index in the years 2005 - 2015. The index is calculated by dividing the CO_2 balance of a given country by it's GDP. The index was then normalized for improved readability using the following formula:

$$I_{normalized} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad \text{where } I_{min} = 0.235,\ I_{max} = 0.668$$

The index represents how impactful are the emissions of citizens of a given country. This is important, as the citizens of a wealthier and more developed country are capable of purchasing more goods, while at the same time, well developed countries often introduce a plethora of different regulations which aim to decrease CO_2 pollution. This is clearly visible in the case of Germany.
Additionally the chart shows that, were the residents of China as wealthy as Germans, they would produce much more CO_2 per capita. To summarize, China produces vastly more CO_2 for each 1$ of income.

Faculty of Mathematics and Information Science
Patryk Rakus Kacper Trębacz Malwina Wojewoda

Data sources: ourworldindata.org stats.oecd.org

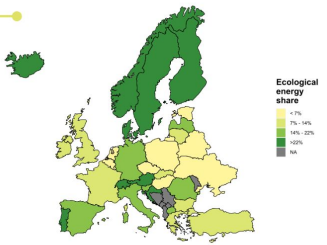
ECOLOGICAL ENERGY AND LIFE OF EUROPEANS

Due to the global warming, being ecological is becoming more and more important. Countries are constantly aiming to reduce carbon emissions by utilizing renewable sources of energy. How does it affect their residents? To answer that, we'll be taking a closer look at European countries.

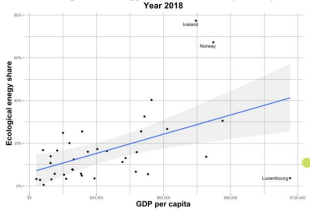
HOW ECOLOGICAL IS EUROPE?

There are some countries that care a lot about being ecological, while others have a very different approach. In Europe, we've got both ends of this spectrum, starting with Iceland, which gets around 77% of its energy from renewable sources, and ending with Belraus which uses such sources to obtain less than a percent of its energy.

Ecological energy share in European countries
Year 2018



Ecological energy share vs GDP per capita
Year 2018



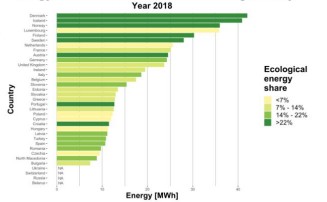
RICHER = MORE ECOLOGICAL?

We can see the main trend in data: The richer the country is, the more ecological energy it uses. Iceland and Norway are perfect examples of this trend. They both have relatively high GDP as well as ecological energy share. Luxembourg, however, is an exception. It has only about 4 percent of energy that comes from renewables sources, despite having the highest GDP per capita in Europe.

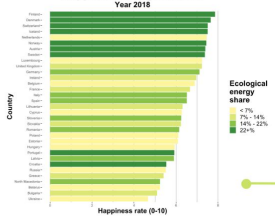
HOW MUCH ENERGY CAN EUROPEANS AFFORD?

In 2018 the price of energy was relatively the cheapest in Denmark which uses a fair amount of renewable energy sources. Does the method of producing energy impact its prices? We can see that the top of the plot corresponding to the best prices is filled mostly with countries that gain energy from more eco-friendly sources. There are some exceptions but overall, it seems like the eco-friendlier your country is, the more energy you can afford.

Energy in MWh available for an average salary
Year 2018



Happiness and greenness
Year 2018



WHAT ABOUT HAPPINESS?

In 2018 seven out of nine top ecological European countries were also leading in happiness for the region. The farther down the happiness rate, the less of a distinction can be made considering the greenness percentage. All in all, it is probably the rich that can afford to be both happy and ecological.

Faculty of Mathematics and Information Science
WARSAW UNIVERSITY OF TECHNOLOGY

AUTHORS: Kaznowska Julia, Piórczyński Mikołaj, Sokołowski Jędrzej

SOURCES: https://data.worldbank.org, https://happyplanetindex.org, https://ec.europa.eu/eurostat, https://github.com/owid/energy-data

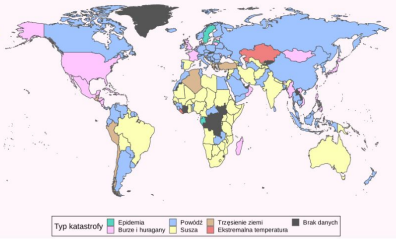
Katastrofy naturalne na mapach świata

Natura objawia swoją potęgę w wielu różnych zjawiskach. Niektóre z nich pokazują jej piękno, inne budzą grozę. My zainteresowaliśmy się katastrofami naturalnymi, które mogą być niezwykle niszczycielskie i śmiercionośne. Postanowiliśmy zbadać, które kraje są najbardziej dotykane przez katastrofy naturalne i które kataklizmy występują w konkretnych częściach świata.

Najbardziej dotkliwe katastrofy naturalne w poszczególnych krajach

Na przestrzeni lat 1971-2008

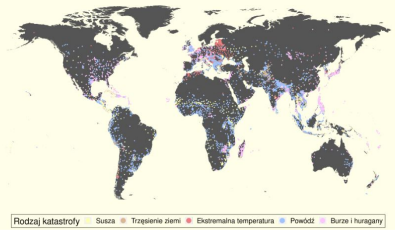
Która z katastrof naturalnych dotyka najwięcej obywateli danego państwa? Okazuje się, że wyniki są zgodne z naszymi przewidywaniami. W większości krajów europejskich najwięcej osób nękanych jest przez powodzie, natomiast w Afryce dominują susze. W USA są to burze i huragany, a w Chile trzęsienia ziemi.



Występowanie katastrof naturalnych

Na przestrzeni lat 2010-2018

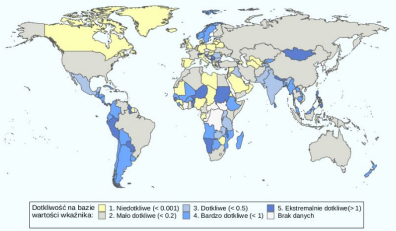
Gdzie występują konkretne katastrofy? Po naniesieniu pojedynczych wystąpień na mapę widać, że w dużej mierze pokrywa się to z oczekiwaniami które mogłibyśmy mieć po przeanalizowaniu poprzedniej mapki. Niestety w niektórych częściach świata zbieranych jest znacznie mniej danych niż w innych. Dlatego na przykład w Europie jest znacznie więcej zaznaczonych wydarzeń niż w Brazylii czy Rosji.



Dotkliwość katastrof naturalnych

Na przestrzeni lat 2010-2018

W którym kraju katastrofy naturalne są najbardziej dotkliwe dla człowieka? Aby to sprawdzić, dla każdego z nich policzyliśmy nasz własny wskaźnik. Jest to liczba osób dotkniętych w wyniku kataklizmów przez populację danego kraju, uśredniona na przestrzeni lat. Następnie pogrupowaliśmy ten wskaźnik na 5 stopni dotkliwości i umieściliśmy na mapie. Okazuje się, że w najgroźniejszych krajach, takich jak Filipiny, Niger czy Peru , wskaźnik osiąga wartości powyżej 1.



Dane:
https://data.unglobalpulse.org/disaster-risk-mitigation-and-adaptation-measures
https://data.unglobalpulse.org/drought-risk-mitigation-and-adaptation-measures
https://www.emdat.be/publications
https://www.gaponline.de/projekte/global-risk-index/
https://www.gaponline.de/projekte/world-risk-report/
https://www.gaponline.de/projekte/world-risk-report/publikationen/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2016/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2017/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2018/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2019/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2020/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2021/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2022/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2023/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2024/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2025/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2026/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2027/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2028/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2029/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2030/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2031/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2032/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2033/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2034/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2035/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2036/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2037/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2038/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2039/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2040/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2041/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2042/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2043/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2044/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2045/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2046/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2047/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2048/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2049/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2050/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2051/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2052/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2053/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2054/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2055/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2056/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2057/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2058/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2059/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2060/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2061/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2062/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2063/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2064/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2065/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2066/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2067/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2068/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2069/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2070/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2071/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2072/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2073/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2074/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2075/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2076/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2077/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2078/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2079/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2080/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2081/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2082/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2083/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2084/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2085/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2086/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2087/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2088/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2089/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2090/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2091/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2092/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2093/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2094/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2095/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2096/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2097/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2098/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2099/
https://www.gaponline.de/projekte/world-risk-report/publikationen/publikationen-2100/

Jakub Piwko
Krzysztof Wodnicki
Łukasz Tomaszewski

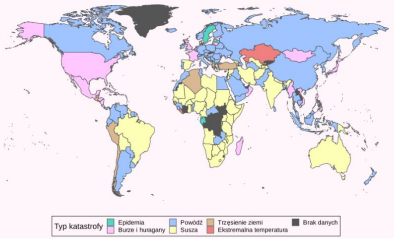
Katastrofy naturalne na mapach świata

Natura objawia swoją potęgę w wielu różnych zjawiskach. Niektóre z nich pokazują jej piękno, inne budzą grozę. My zainteresowaliśmy się katastrofami naturalnymi, które mogą być niezwykle niszczycielskie i śmiercionośne. Postanowiliśmy zbadać, które kraje są najbardziej dotykane przez katastrofy naturalne i które kataklizmy występują w konkretnych częściach świata.

Najbardziej dotkliwe katastrofy naturalne w poszczególnych krajach

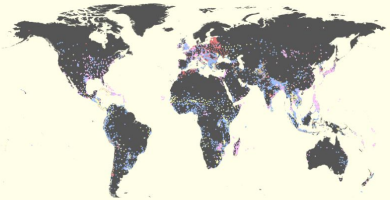
Na przestrzeni lat 1971-2008

Która z katastrof naturalnych dotyka najwięcej obywateli danego państwa? Okazuje się, że wyniki są zgodne z naszymi przewidywaniami. W większości krajów europejskich najwięcej osób nękanych jest przez powodzie, natomiast w Afryce dominują susze. W USA są to burze i huragany, a w Chile trzęsienia ziemi.



Występowanie katastrof naturalnych

Na przestrzeni lat 2010-2018

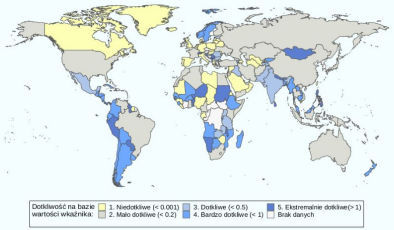


Gdzie występują konkretne katastrofy? Po naniesieniu pojedynczych wystąpień na mapę widać, że w dużej mierze pokrywa się to z oczekiwaniami które moglibyśmy mieć po przeanalizowaniu poprzedniej mapki. Niestety w niektórych częściach świata zbieranych jest znacznie mniej danych niż w innych. Dlatego na przykład w Europie jest znacznie więcej zaznaczonych wydarzeń niż w Brazylii czy Rosji.

Dotkliwość katastrof naturalnych

Na przestrzeni lat 2010-2018

W którym kraju katastrofy naturalne są najbardziej dotkliwe dla człowieka? Aby to sprawdzić, dla każdego z nich policzyliśmy nasz własny wskaźnik . Jest to liczba osób dotkniętych w wyniku kataklizmów przez populację danego kraju, uśredniona na przestrzeni lat. Następnie pogrupowaliśmy ten wskaźnik na 5 stopni dotkliwości i umieściliśmy na mapie. Okazuje się, że w najgroźniejszych krajach, takich jak Filipiny, Niger czy Peru , wskaźnik osiąga wartości powyżej 1.



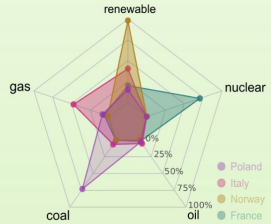
Dane:
https://data.gov.pl/zestaw/zestawienie-wskaznikow-dla-wojewodztw-i-powiatow-2016-2017/
https://data.gov.pl/zestaw/dane-o-ofiarach-smiertelnych-wypadkow-drogowych-w-latach-2007-2016/
https://ourworldindata.org/natural-disasters
https://www.gapminder.org/data/

Jakub Piwko
Krzysztof Wodnicki
Łukasz Tomaszewski

Are Electric cars that Eco?

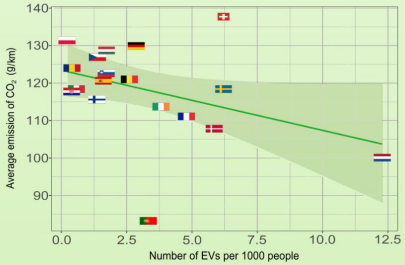
PL

Study on in which countries electric vehicles really reduce CO_2 emissions



Average CO_2 emissions of Renault Zoe by country

To check how different methods of energy production impact emissions from electric cars we chose 4 countries with different primary way of producing energy and calculated how many grams of CO_2 are produced by driving 1km.



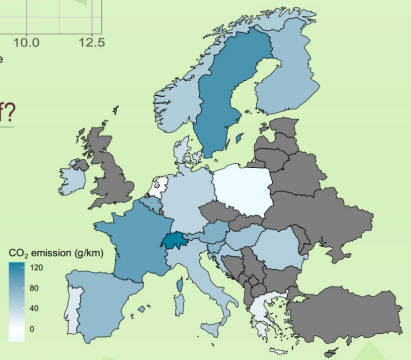
Average CO_2 emissions of passanger cars by country

To further examine beneficiality of electric vehicles we compared average CO_2 emission of all types of cars and number of electric vehicles in each country. It's easy to observe that generally the more electric cars there are in a country, the less CO_2 is produced by all vehicles on average.

Does an electric car pay off?

The difference in CO_2 production of electric and internal combustion cars

In previous visualizations we showed that in some countries it's more beneficial to own an electric vehicle than in others. We calculated the difference of CO_2 production of an electric and combustion cars in each country. The bigger the difference, the bigger is the positive influence on environment of buying an electric car.



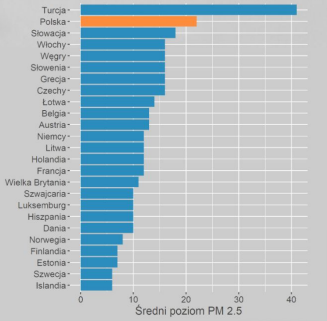
Authors:
Filip Szympliński
Michał Tomczyk
Piotr Wilczyński

Data sources:
https://ourworldindata.org
https://ev-database.org/cheatsheet/energy-consumption-electric-car
https://ec.europa.eu/eurostat/web/transport/data/database

ZANIECZYSZCZENIE POWIETRZA W POLSCE

Polskie powietrze jest jednym z najbardziej zanieczyszczonych w EU, a mimo tego, nie są prowadzone adekwatne działania zapobiegające temu problemowi. Największy problem występuje z wielopierścieniowymi węglowodorami aromatycznymi (WWA) (m.in. z BaP), pyłem zawieszonym PM_{10} oraz drobniejszą frakcją pyłu $PM_{2.5}$. Ich normy w licznych polskich miejscowościach są przekraczane kilku a nawet kilkunastokrotnie.

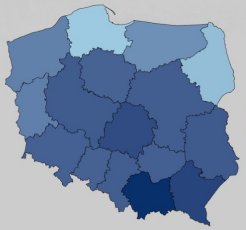
Średnie zanieczyszczenie państw Europy pyłem $PM_{2.5}$



PYŁ ZAWIESZONY $PM_{2.5}$

Nasz niepokój może wzbudzać poziom $PM_{2.5}$. Ze względu na mały rozmiar, dociera do pęcherzyków płucnych, skąd przenika do krwiobiegu. Jego stężenie roczne nie powinno być wyższe niż 10 $\mu g/m^3$, a według mapy zanieczyszczenia $PM_{2.5}$, Polska (szczególnie południowa) boryka się z jego wysokimi średniorocznymi stężeniami.

Średni pomiar $PM_{2.5}$ w poszczególnych województwach w 2019 roku



WWA I INNE SKŁADNIKI SMOGU

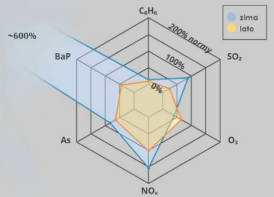
Najlepiej przebadanym wielopierścieniowym węglowodorem aromatycznym jest benzopiren. W 1987 roku Międzynarodowa Agencja Badań nad Rakiem (IARC) uznała BaP za główny ludzki kancerogen. W naszych badaniach porównaliśmy uśrednione pomiary składników smogu z ich normami prawnymi. Jak widać, tak toksyczna substancja jak BaP mocno je przekracza. Tak samo poziomy SO_2 i tlenków azotu (NO_x) mogą wzbudzać nasz niepokój.

WPŁYW LOCKDOWNU NA STAN POWIETRZA

Czy siedząc w domu poprawiliśmy jakość naszego powietrza? Celem ostatniego badania było porównanie wpływu lockdownu na poziom wytwarzanego zanieczyszczenia. Jak się okazuje, wiele się nie poprawiło. Być może, w Polsce to nie komunikacja samochodami jest głównym źródłem tego problemu, a jest nim ogrzewanie gospodarstw domowych.

Stan powietrza w naszym kraju jest alarmujący - brakuje regulacji ku tej kwestii: ogrzewanie piecami, kotłami czy kominkami nie powinno być na porządku dziennym. Musimy zacząć zwracać uwagę na to, czym palimy i co palimy.

Uśredniony pomiar emisji poszczególnych związków chemicznych w Polsce - porównanie lata i zimy



Porównanie poziomu $PM_{2.5}$ w 2019 i 2020 roku

Źródła:
http://powietrze.gios.gov.pl/pjp/archives
https://www.greenmatch.co.uk/blog/2018/11/mapped-europes-most-and-least-polluted-countries

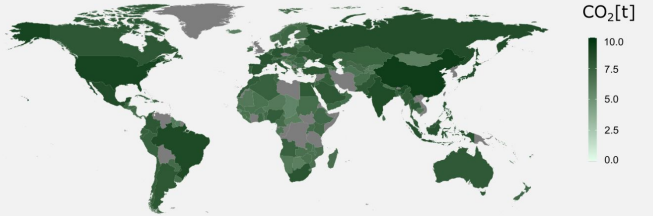
Autorzy:
Jakub Grunas, Laura Hoang, Joanna Kajka

Produkcja żywności a ekologia

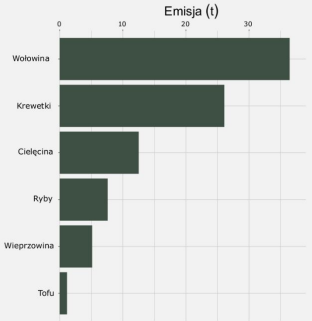
Dawid Płudowski
Antoni Zajko
Grzegorz Kiersnowski

Źródła: FAOSTAT, Kaggle, OWID

Emisja CO2 *per capita* w roku 2013 w wyniku produkcji żywności

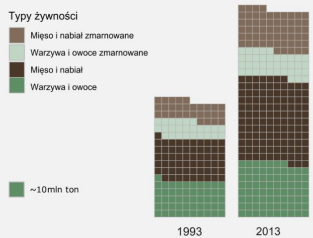


Emisja CO2 w wyniku produkcji żywności w przeliczeniu na 1000 kcal



Badania sugerują, że 20% dwutlenku węgla, który wytwarzamy, pochodzi z produkcji żywności. Marnowanie jedzenia również przyczynia się do transmisji nadmiarowego dwutlenku węgla do atmosfery. Najbardziej odpowiedzialne za ten stan rzeczy są kraje rozwinięte.

Produkcja i marnowanie żywności W latach 1993 i 2013



Spożycie wybranych produktów w 2013 roku

W kilogramach *per capita*

Pytania?

Eksploracja danych

Dane

Mogą być generowane przez:

- ?

Dane

Mogą być generowane przez:

- banki,
- ubezpieczenia,
- portale społecznościowe,
- firmy telekomunikacyjne,
- szpitale,
- dane eksperymentalne,
- tekst,
- mapy,
- sklepy internetowe,
- ...

Eksploracja danych - czym jest?

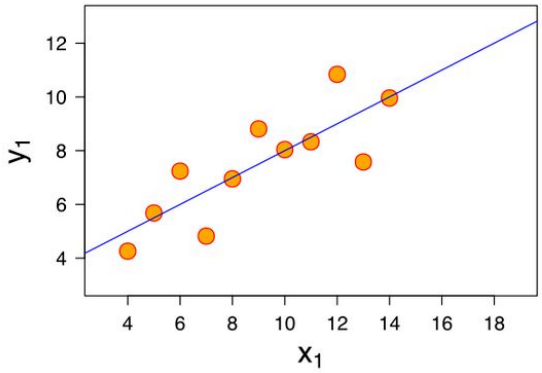
"proces odkrywania nietrywialnych, dotychczas nieznanych, potencjalnie użytecznych reguł, zależności, trendów"

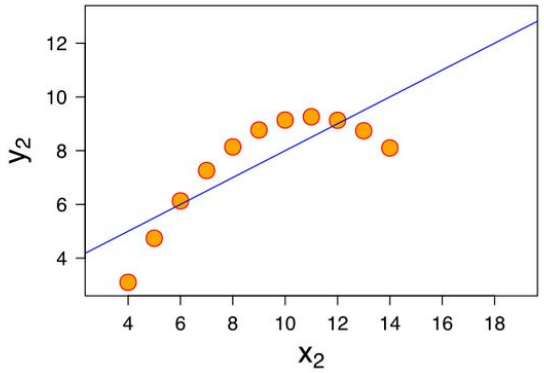
Cel: analiza danych w celu lepszego ich zrozumienia

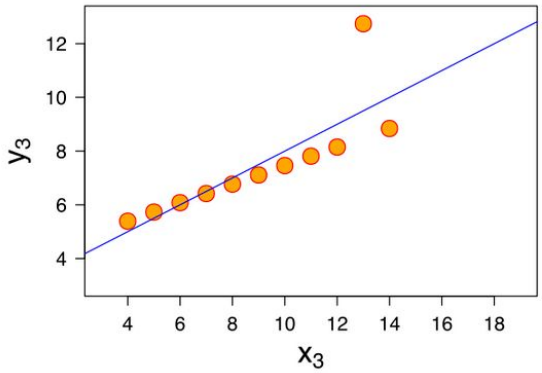
Eksploracja danych - czym jest?

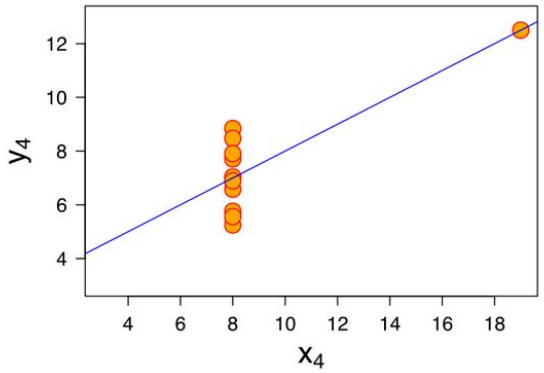
Na eksplorację danych składa się wiele dyscyplin, między innymi:

- bazy danych
- statystyka
- uczenie maszynowe
- wizualizacja danych
- wyszukiwanie informacji

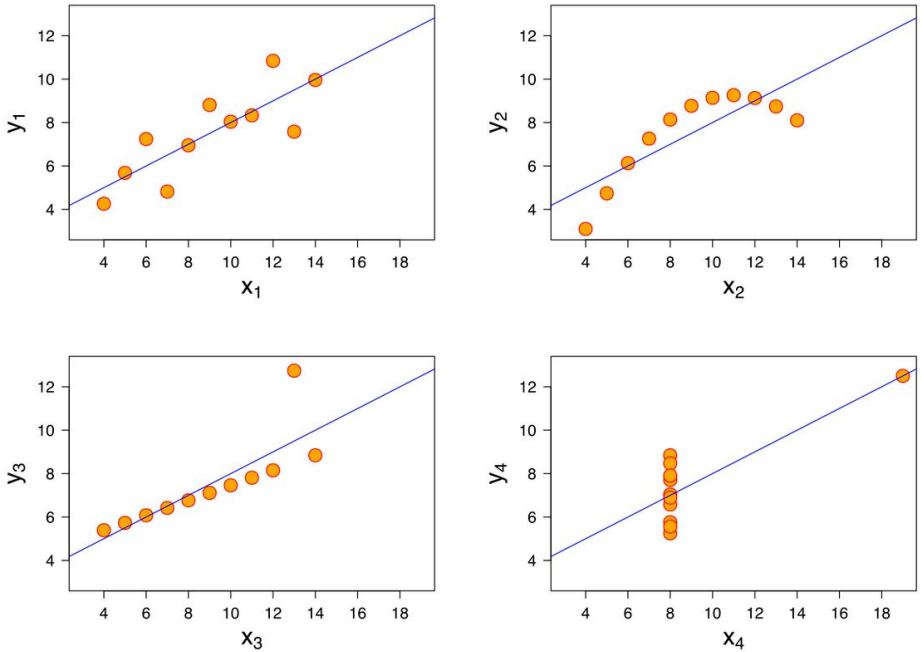








Kwartet Anscombe'a



Cecha	Wartość
Średnia arytmetyczna zmiennej x	9
Wariancja zmiennej x	11
Średnia arytmetyczna zmiennej y	7.50 (identyczna do dwóch cyfr po przecinku)
Wariancja zmiennej y	4.122 lub 4.127 (identyczna do trzech cyfr po przecinku)
Współczynnik korelacji pomiędzy zmiennymi	0.816 (identyczny do trzech cyfr po przecinku)

Jak rozpoznać rodzaj zmiennej?

“dane liczbowe to nie tylko liczby”

Typy danych

Zmienne jakościowe (nazywane również wyliczeniowymi, czynnikowymi lub kategorycznymi), to zmienne przyjmujące określoną liczbę wartości (najczęściej nie liczbowych). Zmienne te można dalej podzielić na:

- *binarne* (nazywane również dwumianowymi, dychotomicznymi) np. płeć (poziomy: kobieta/mężczyzna),
- *nominalne* (nazywane również zmiennymi jakościowymi nieuporządkowanymi) np. marka samochodu,
- *uporządkowane*, np. wykształcenie (poziomy: podstawowe/średnie/wyższe), ocena z przedmiotu.

Typy danych

Zmienne ilościowe, z których można dodatkowo wyróżnić:

- *zliczenia* (liczba wystąpień pewnego zjawiska, opisywana liczbą całkowitą), np. liczba lat nauki, liczba wypadków,
- *ilorazowe*, czyli zmienne mierzone w skali, w której można dzielić wartości (ilorazy mają sens). Np. długość w metrach (coś jest 2 razy dłuższe, 10 razy krótsze itp.),
- *przedziałowe* (nazywane też interwałowymi), mierzone w skali, w której można odejmować wartości (wyznaczać długość przedziału).

Struktura zbioru danych

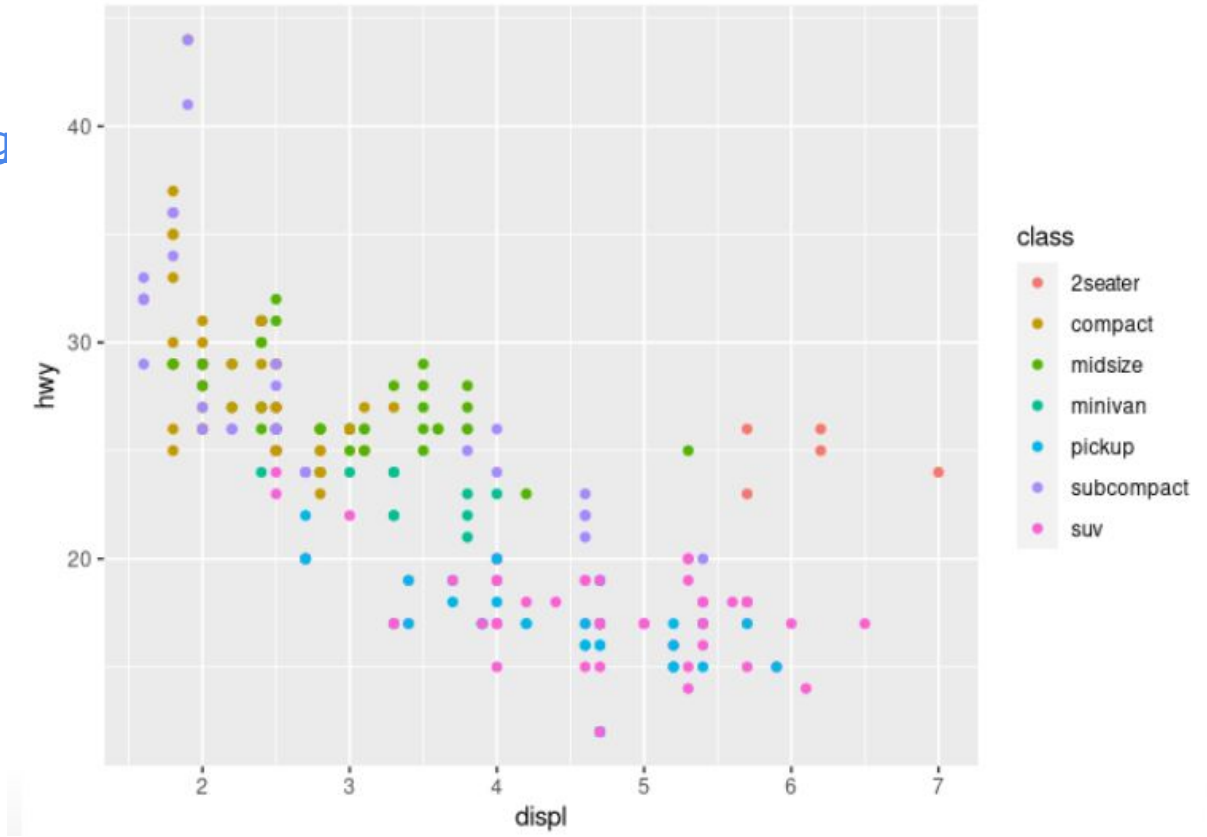
ID	PŁEĆ	ZAWÓD	WZROST	DATA URODZENIA
ID_23	K	INFORMATYK	158	1978-03-12
ID_45	K	PRAWNIK	178	1989-05-29
ID_46	M	MATEMATYK	183	1991-01-19
ID_89	M	INFORMATYK	167	1982-02-20
ID_101	K	LEKARZ	163	1973-02-23

Narzędzia do wizualizacji danych

- programistyczne (R, Python, JavaScript)
- programy graficzne (Inkscape)
- programy dedykowane do wizualizacji danych (Tableau)

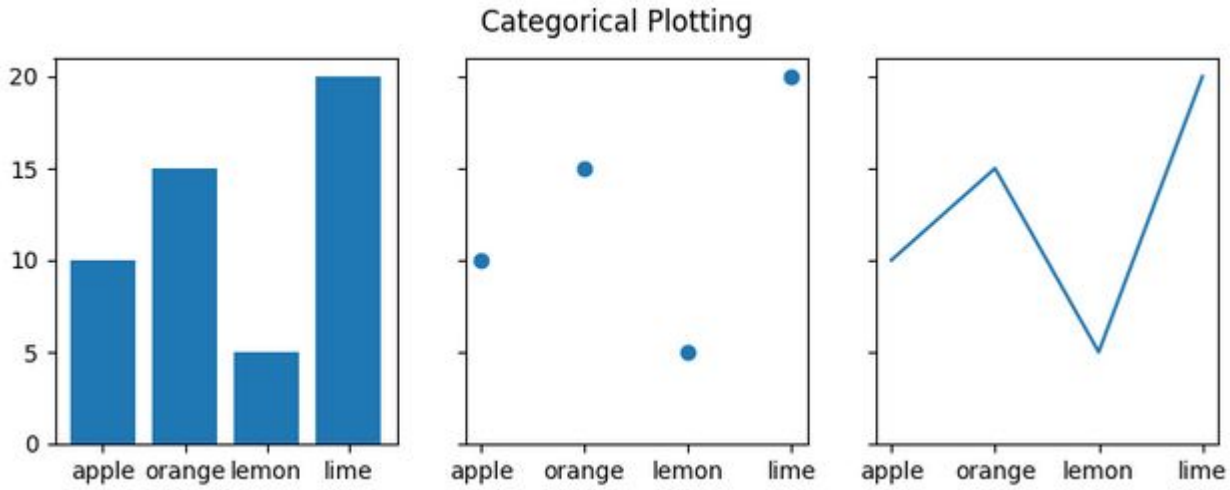
ggplot2 (R)

https://ggplot2.tidyverse.org



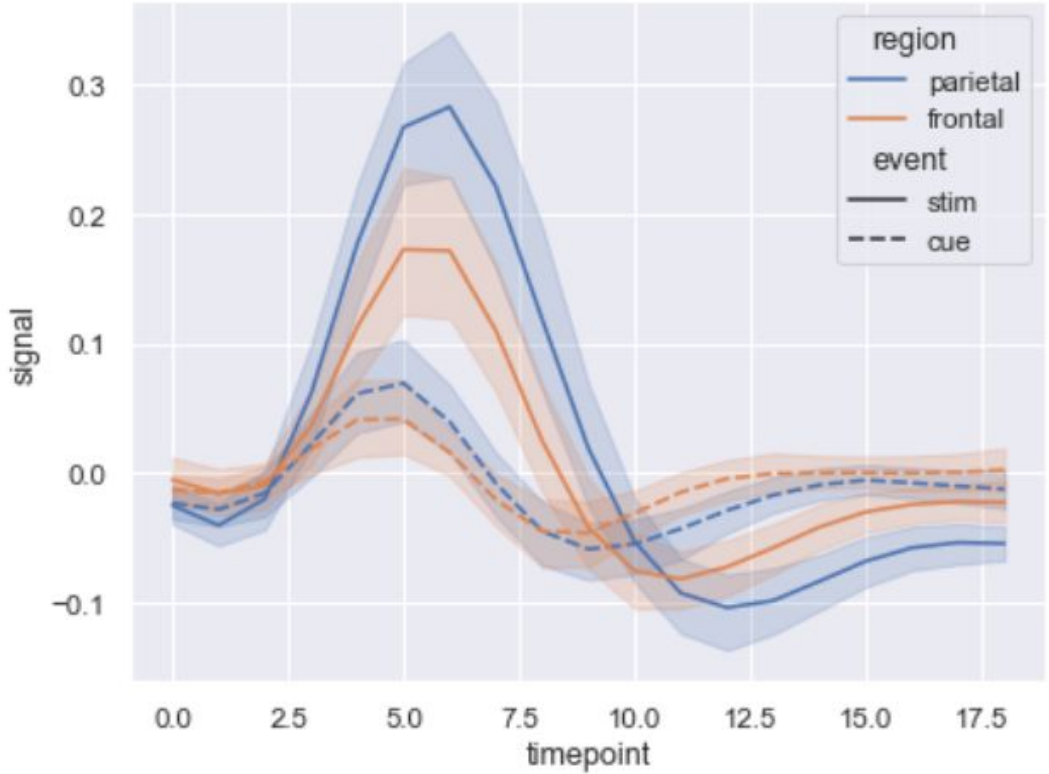
matplotlib (Python)

https://matplotlib.org/



seaborn (Python)

https://seaborn.pydata.org/



plot.ly

Interaktywne wizualizacje w Javascript z interfejsem w Python i R.

https://plotly.com/python/line-and-scatter/

plotly.js: https://github.com/plotly/plotly.js

plotly.py: https://github.com/plotly/plotly.py

plotly.R: https://github.com/ropensci/plotly