

Interreg



Kofinanziert von
der Europäischen Union
Spolufinancováno
Evropskou unií

Sachsen – Tschechien | Česko – Sasko

**Interreg Sachsen – Tschechien
2021-2027**

**Interreg Česko – Sasko
2021-2027**



Von Excel zur Analyse Vratislav Žabka

Od Excelu k analýze Vratislav Žabka



Data gapminder

Jsou načtena data o HDP, populaci a očekávané době dožití většiny zemí světa. AI (gpt-4o-mini) má informace o jejich struktuře.

Zkuste napsat příkaz pro AI, podle kterého bude vygenerován graf.

Napište příkaz pro tvorbu grafu

Ukázka: Vytvoř bodový graf závislosti HDP na očekávané době dožití v jednotlivých zemích. Body obarvi podle kontinentu.

Proveď příkaz

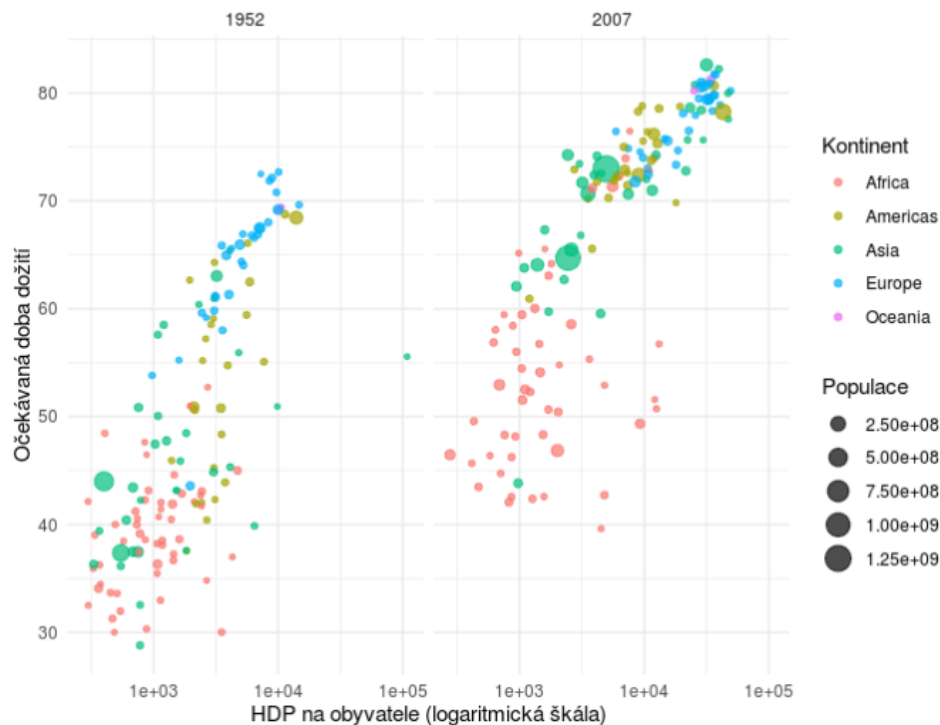
Graf

Kód grafu

Interaktivní graf

Testované prompty

Závislost HDP na očekávané době dožití



Obsah prezentace

- **Příprava dat**
 - Zdroj dat (INCIEN)
 - Původní struktura dat
 - Motivace pro transformaci dat
- **Zpracování dat a základní statistika**
 - Přehled analýz
 - Mapy, korelace, PCA, koláče
 - AI na závěr

Stránky projektu s prezentací: www.reni-netw.org

Zkrácené URL: 1url.cz/dJzPZ

Fyzická analýza dat INCIEN

INCIEN - institut cirkulární ekonomiky (<https://incien.org/>)

- **Odpady jako zdroj informací o chování obyvatel**
- **Sledování časových trendů**
- **Identifikace možných vztahů mezi složkami odpadu**
- **Porovnání chování různých obcí a městských částí**
- **Přesná čísla místo obecných doporučení**

Konkrétní data mohou mít větší přesvědčovací sílu než obecná kampaň.

Fyzická analýza odpadu INCIEN



Původní podoba dat 2016

Fyzická analýza odpadu v roce 2016				
Název obce	Lhenice		Lhenice	
Datum analýzy	#####		#####	
Počet obyvatel	1989		1989	
Druh odpadu	množství	Podíl v	množství	Podíl v
	váha [kg]	váha [kg]	váha [kg]	váha [kg]
ZAHRADNÍ ZELEŇ	28	3%	28,8	4%
GASTROODPAD	85	10%	131,1	17%
BRKO	113	14%	159,9	21%
PLAST MĚKKÝ	36	4%	34,35	4%
FOLIE MIX	55	7%	38,7	5%
PET LAHEV MIX	7	1%	0	0%
PET LAHEV SVĚTLÁ	0	0%	24	3%
PET LAHEV BAREVNÁ	0	0%	15,45	2%
HDP TVRDÉ PLASTY	10	1%	9,45	1%
PLAST	108	13%	121,95	16%
PAPÍR MIX	12	1%	0	0%
LEPENKA, KARTON	0	0%	12,9	2%
TISKOVINY	0	0%	37,5	5%
PAPÍR	12	1%	50,4	7%
CELKEM	824,5	100%	770,85	100%
▶ 2015 2016 2017 2018 2019				

- Popis dat
- Problémy pro automatizované zpracování:
 - Nejednotná struktura
 - Ručně upravované sloupce a formátování
 - Názvy odpadu nejsou jednoznačné
 - Nepřehledné pro automatizované zpracování

Původní podoba dat 2024

	A	B	E	F
1		Obec:	Napajedla	
2		Termín:	21.10. 2024	
3		Celková hmotnost vzorku [kg]:	106,46	
4		Poznámka:	2VZ - RD	
5	249,24		106,46	100,00
6		Celkem	106,46	100,00
7		Kategorie odpadu	Váha [kg]	Podíl [%]
8	PAPÍR	lepenka a karton	2,11	1,98
9		ostatní obalový	0,87	0,82
10		kancelářský papír a	0,44	0,41
11		Papír	3,42	3,21
12	PLAST	fólie	2,54	2,39
13		tvrdé plasty	0,49	0,46
14		PET láhve	0,57	0,54
15		PS	0,05	0,05
16		ostatní měkké plasty	3,65	3,43
17		Plast	7,30	6,86
18	BIO	(ovoce, zelenina)	32,89	30,89
19		nekompostovatelný (prošlé	17,90	16,81
20		ze zahrad a parků	5,58	5,24
21		BIO	56,37	52,95
22	DŘEVO - upravené	procesované dřevo	0,01	0,01

- Popis dat
- Problémy pro automatizované zpracování:
 - Nejednotná struktura
 - Ručně upravované sloupce a formátování
 - Názvy odpadu nejsou jednoznačné
 - Nepřehledné pro automatizované zpracování

Cíle a podmínky nové struktury dat

- **Automatizovatelné načítání do nástrojů jako R**
- **Jednoduchá aktualizace vkládáním nových analýz**
- **Možnost přímé vizualizace v Excelu (pomocí implementovaných dashboardů)**
- **Jasná vazba mezi odpady a analýzami**

Nová datová struktura - kompromis

Kompromis mezi optimální databázovou strukturou a jednoduchým používáním v excelu

Analýzy

Data

Ukázka části tabulky s analýzami

an_id	obec	kod_obce	rok	mesic	den	obyvatel	typ_zastavby	KOD_NUTS3
an_1	Praha Malešice		2015	11	15	42381	nespecifikováno	CZ010
an_9	Horní Suchá	552739	2016	8	23	4537	bytová	CZ080
an_10	Horní Suchá	552739	2016	6	21	4537	rodinné domy	CZ080
an_29	Choceň	580350	2018	4	25	8666	nespecifikováno	CZ053
an_71	Choceň	580350	2022	5	26	8477	nespecifikováno	CZ053

Nová datová struktura - kompromis

Kompromis mezi optimální databázovou strukturou a jednoduchým používáním v excelu

Analýzy	Data
---------	------

Ukázka části tabulky s daty

nazev_odpadu	kategorie_odpadu	odpad_id	an_1	an_2
zahradní zeleň	BIO	1	0	0
plast měkký	PLAST	4	34.8	85.4
HDP tvrdé plasty	PLAST	7	7.5	11.1
papír, tiskoviny	PAPIR	8	40.9	162.8
lepenka, karton	PAPIR	9	79.4	64.2
sklo	SKLO	10	42.2	71.7

Nová datová struktura - R

Analýzy

Data

Kategorie odpadů

Ukázka části tabulky s analýzami

an_id	muni	muni_code	date	pop	settlement_type	nuts3
1	Praha Malešice		2015-11-15	42381	nespecifikováno	CZ010
9	Horní Suchá	552739	2016-08-23	4537	bytová	CZ080
10	Horní Suchá	552739	2016-06-21	4537	rodinné domy	CZ080
29	Choceň	580350	2018-04-25	8666	nespecifikováno	CZ053
71	Choceň	580350	2022-05-26	8477	nespecifikováno	CZ053

Nová datová struktura - R

Analýzy

Data

Kategorie odpadů

Ukázka části tabulky s data

an_id ↕	waste_id ↕	val ↕
4	1	28
5	1	28.8
21	1	49.725
22	1	161.8
23	1	29.25
24	1	109.5
25	1	39.4
26	1	110.8

Nová datová struktura - R

Analýzy	Data	Kategorie odpadů
---------	------	------------------

Tabulka kategorií odpadu

waste_id	waste_name	waste_cathegory
All	All	All
1	zahradní zeleň	BIO
2	kuchyňský odpad kompostovatelný	BIO
3	kuchyňský odpad nekompostovatelný	BIO
4	plast měkký	PLAST
5	plastové folie	PLAST
6	PET lahve	PLAST
7	HDP tvrdé plasty	PLAST
8	papír, tiskoviny	PAPIR

Showing 1 to 8 of 58 entries

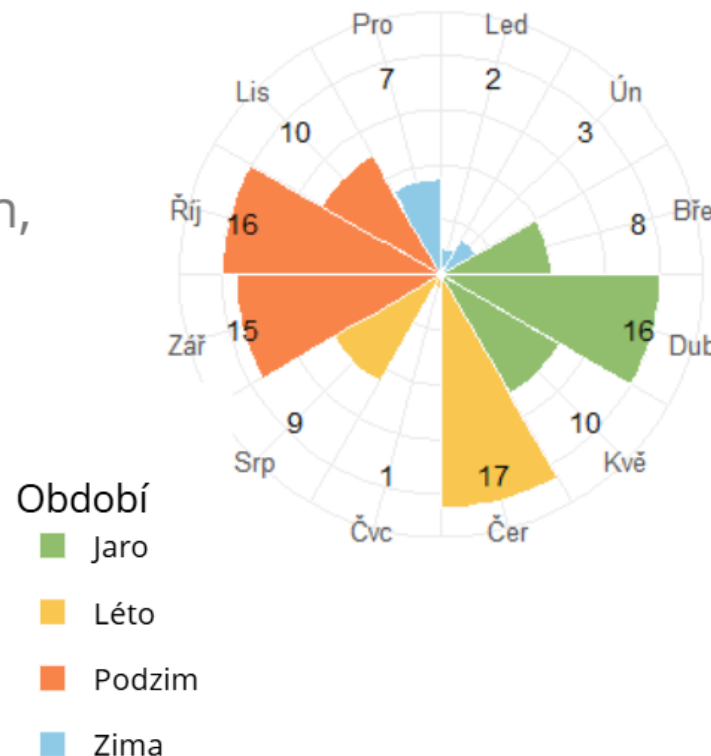
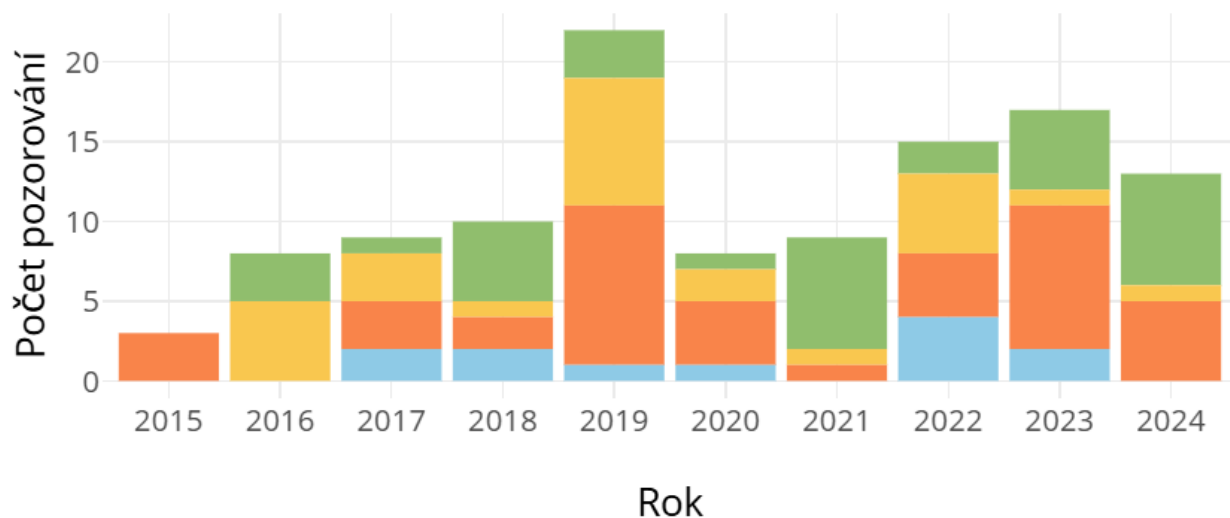
Previous 1 2 3 4 5 ... 8 Next

Ukázka různých vizualizací v R

- **Časový přehled analýz**
- **Počty podle krajů, mapy**
- **Koláče- Analýza trendů**
- **PCA analýza**
- **Korelační matice složek**
- **Grafy pomocí AI**

Časový přehled analýz

- Nejvíce analýz provedeno v roce 2019
- Pro analýzy byly nejméně oblíbené měsíce leden, únor a červenec



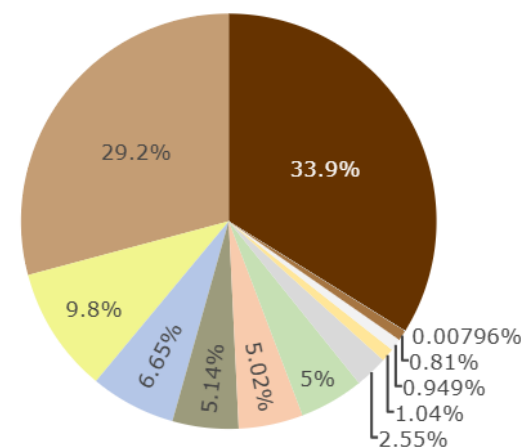
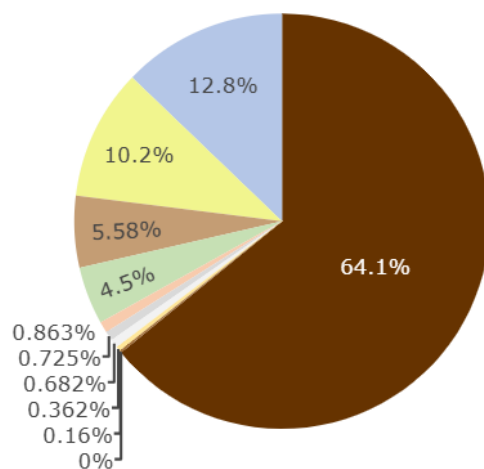
Interaktivní koláčové grafy

Koláče

Tabulka analýz

an_1 - Praha Malešice - 2015

Souhrn všech analýz



Výběr druhů odpadu

BATERIE/ AKUMULÁTORY BIO DŘEVO - upravené ELEKTROZAŘÍZENÍ KOV NÁPOJOVÉ KARTONY Ostatní ODPADY PAPIR PLAST
SKLO SKO TEXTIL

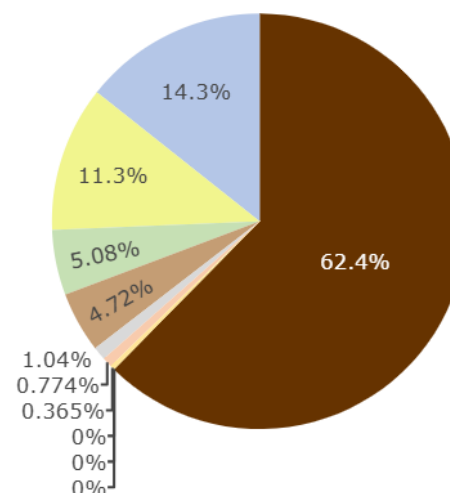
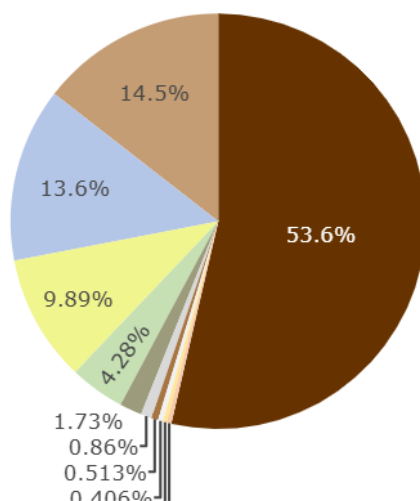
Interaktivní koláčové grafy

Koláče

[Tabulka analýz](#)

an_2 - Praha Malešice - 2015

an_3 - Praha Malešice - 2015



Výběr druhů odpadu

BATERIE/ AKUMULÁTORY BIO DŘEVO - upravené ELEKTROZAŘÍZENÍ KOV NÁPOJOVÉ KARTONY Ostatní ODPADY PAPIR PLAST
SKLO SKO TEXTIL

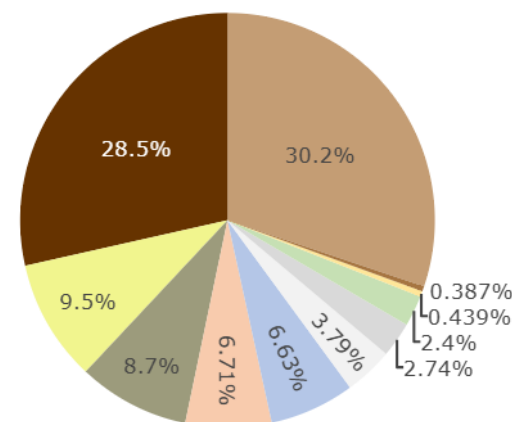
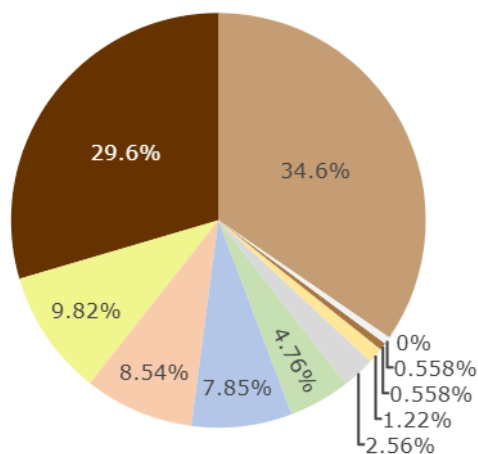
Interaktivní koláčové grafy

Koláče

[Tabulka analýz](#)

an_29 - Choceň - 2018

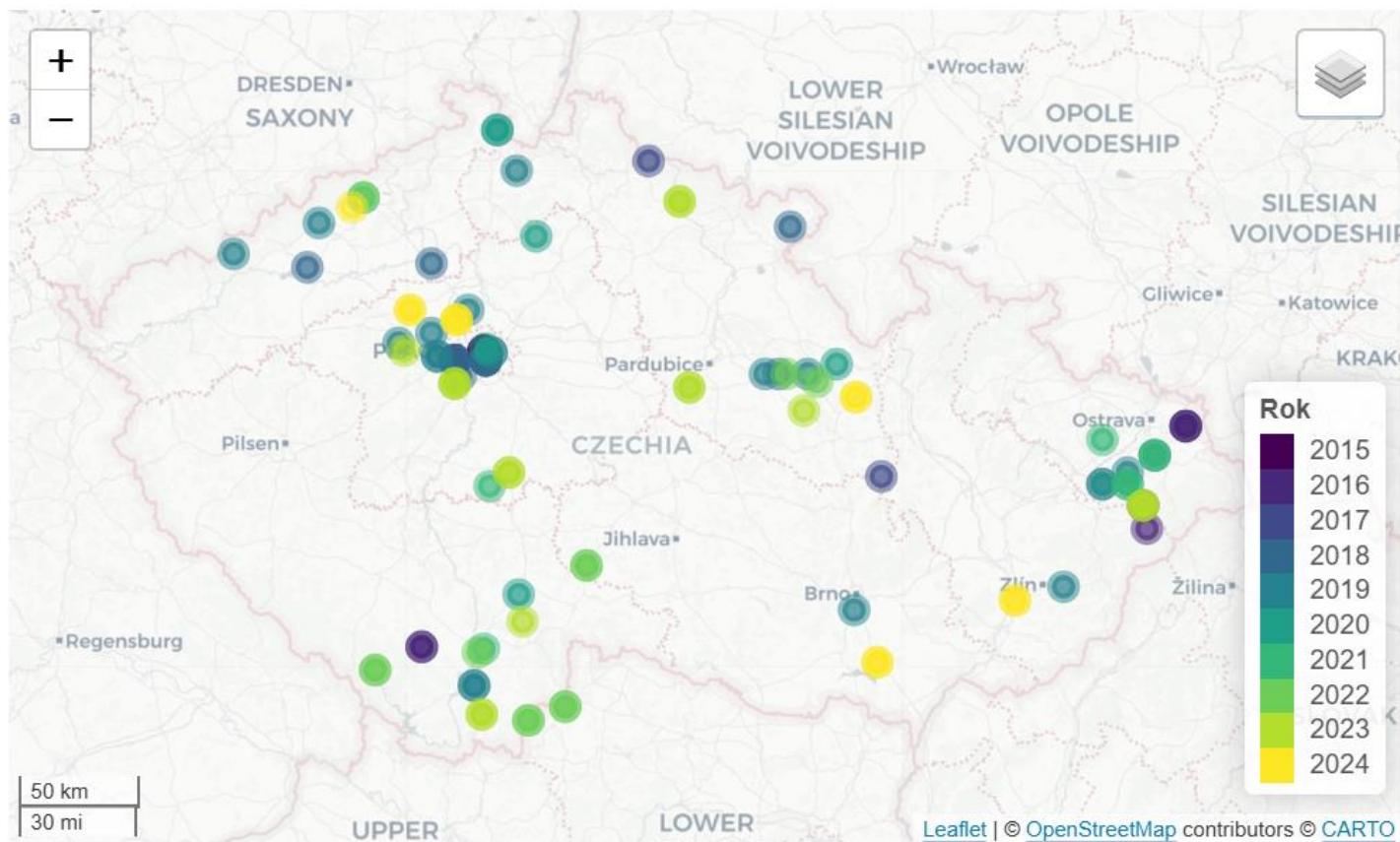
an_71 - Choceň - 2022



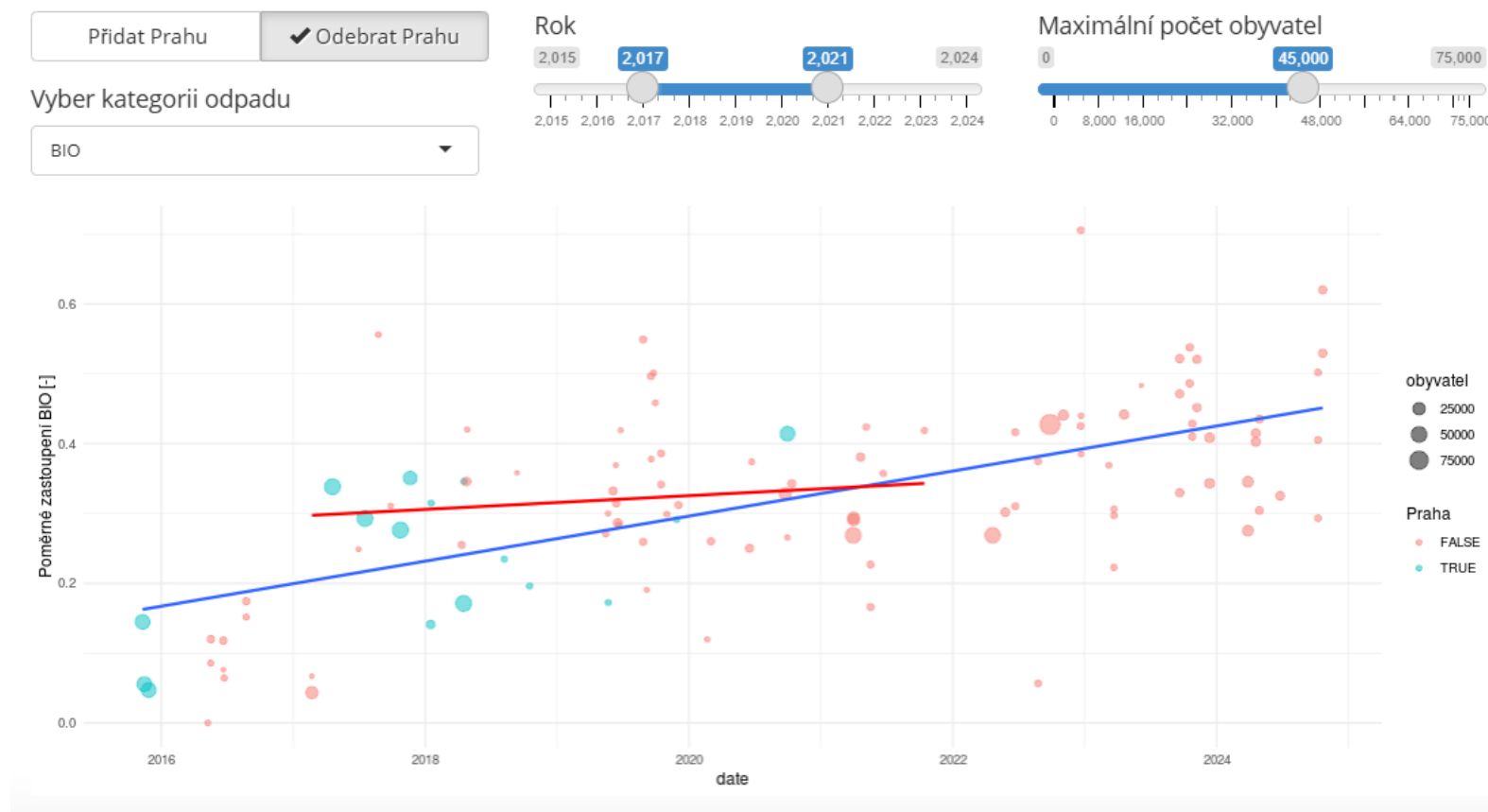
Výběr druhů odpadu

BATERIE/ AKUMULÁTORY BIO DŘEVO - upravené ELEKTROZAŘÍZENÍ KOV NÁPOJOVÉ KARTONY Ostatní ODPADY PAPIR PLAST
SKLO SKO TEXTIL

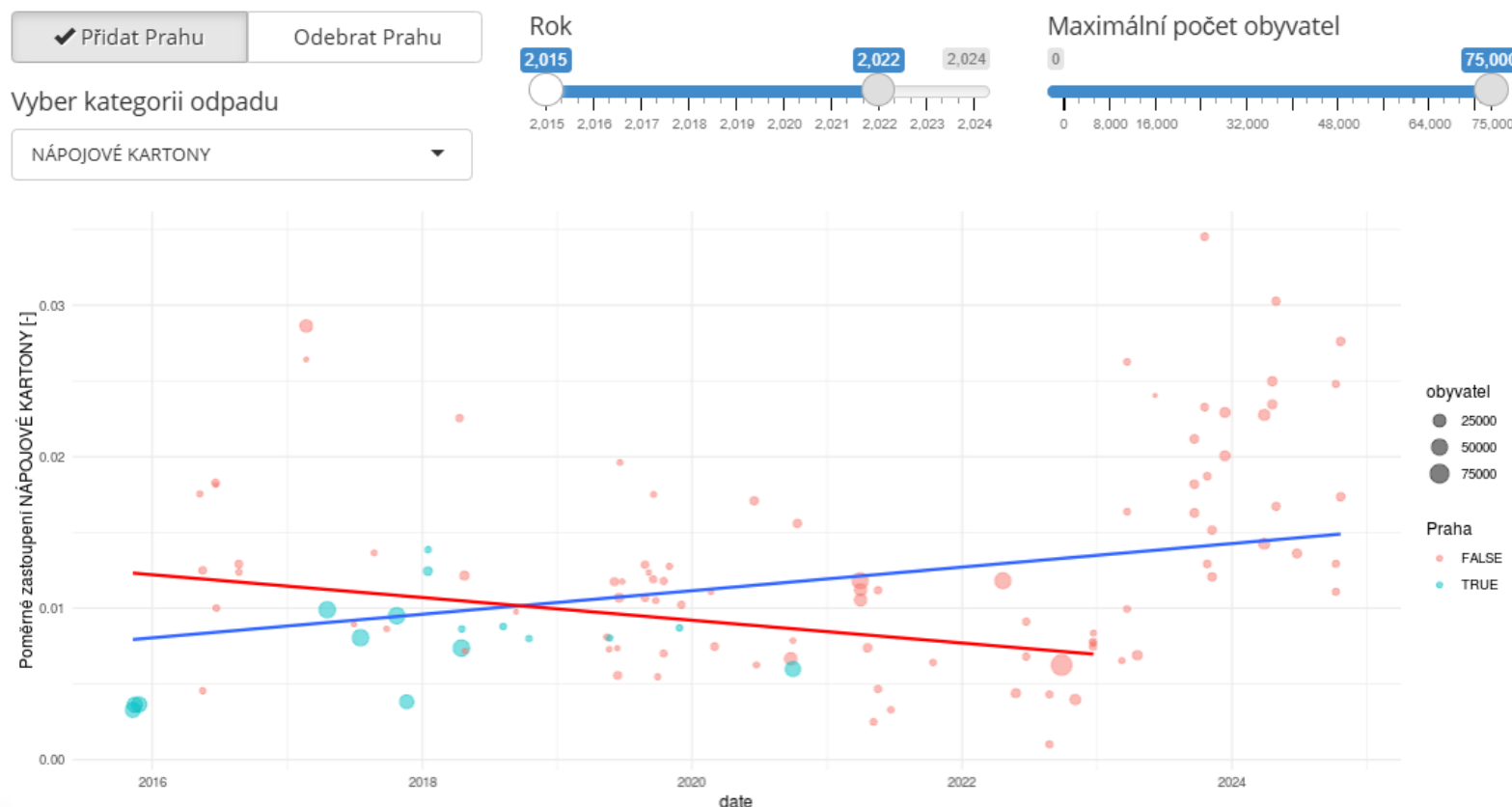
Interaktivní mapa analýz podle roků



Srovnávání trendů zastoupení složek odpadů



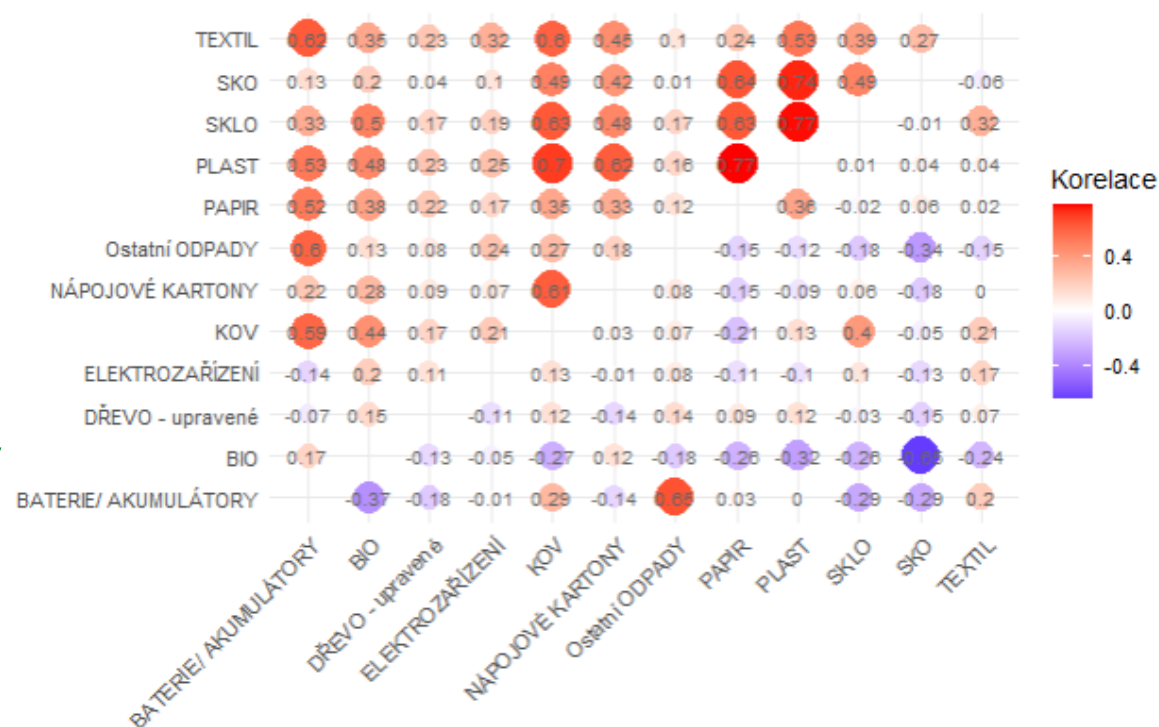
Srovnávání trendů zastoupení složek odpadů



Korelace dvojic odpadů

Na co si dát pozor:

- **Povaha dat**
 - absolutní [kg]
 - relativní [podíl]
- **Odlehlé hodnoty**
- **Lineární vs. nelineární vztahy**
- **Vzájemná závislost kategorií**
- **Počet pozorování (n)**
- **Interpretace**



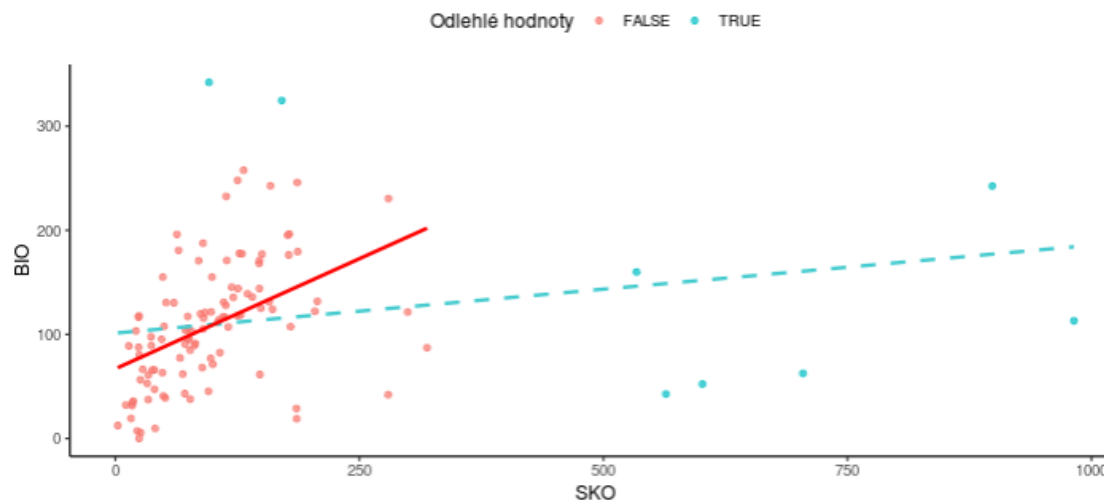
Korelace dvojic

Data FAO

SKO ▼

BIO ▼

☒ Absolutní ☐ Relativní



dataset	correlation	p_value	conf_interval
Všechna data	0.20	3.20e-02	(0.02, 0.37)
Bez odlehlých hodnot	0.47	4.40e-07	(0.30, 0.60)

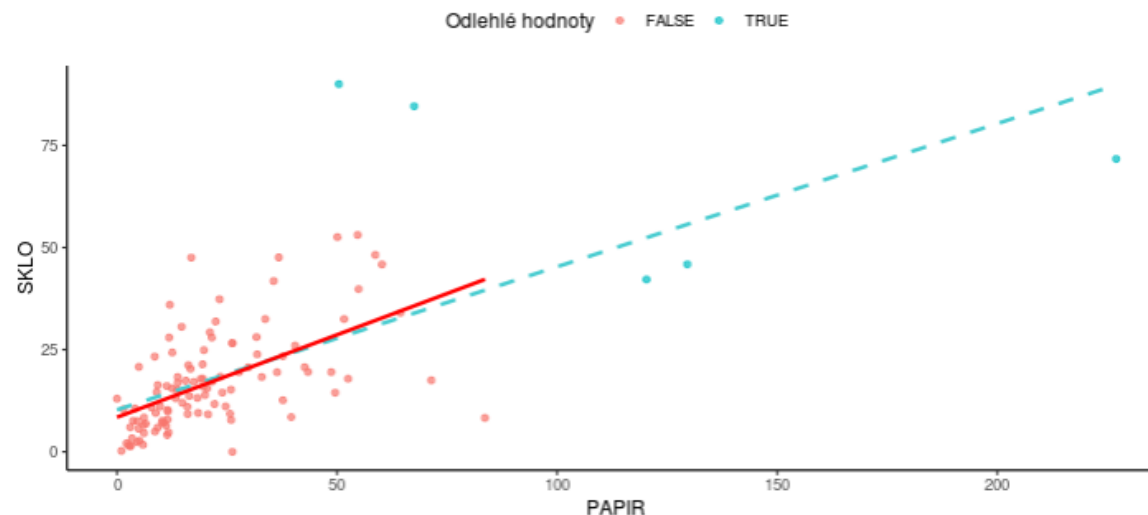
Korelace dvojic

Data FAO

PAPIR ▼

SKLO ▼

☒ Absolutní ☐ Relativní



dataset	correlation	p_value	conf_interval
Všechna data	0.63	6.06e-14	(0.50, 0.73)
Bez odlehlých hodnot	0.56	1.62e-10	(0.42, 0.68)

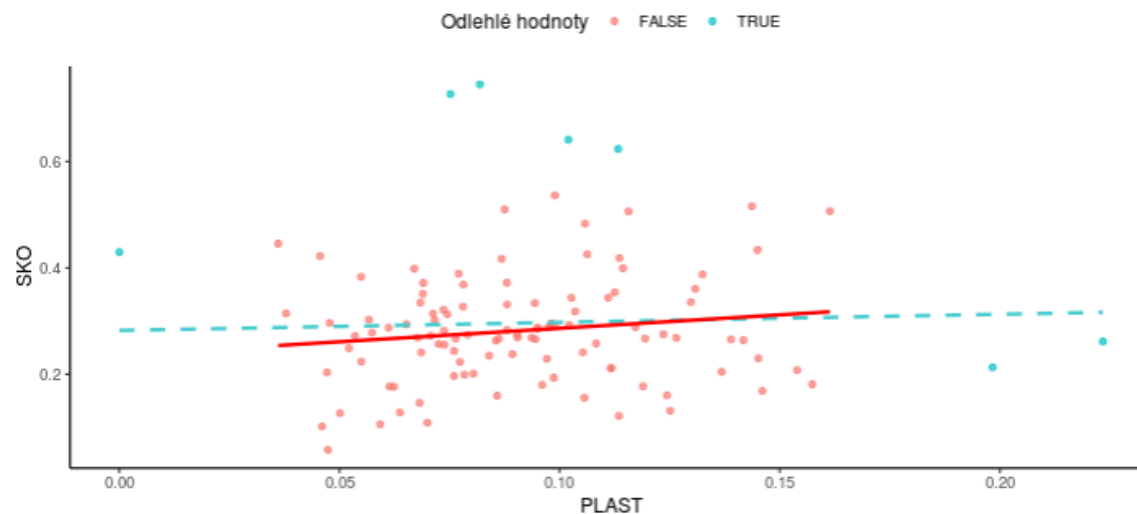
Korelace dvojic

Data FAO

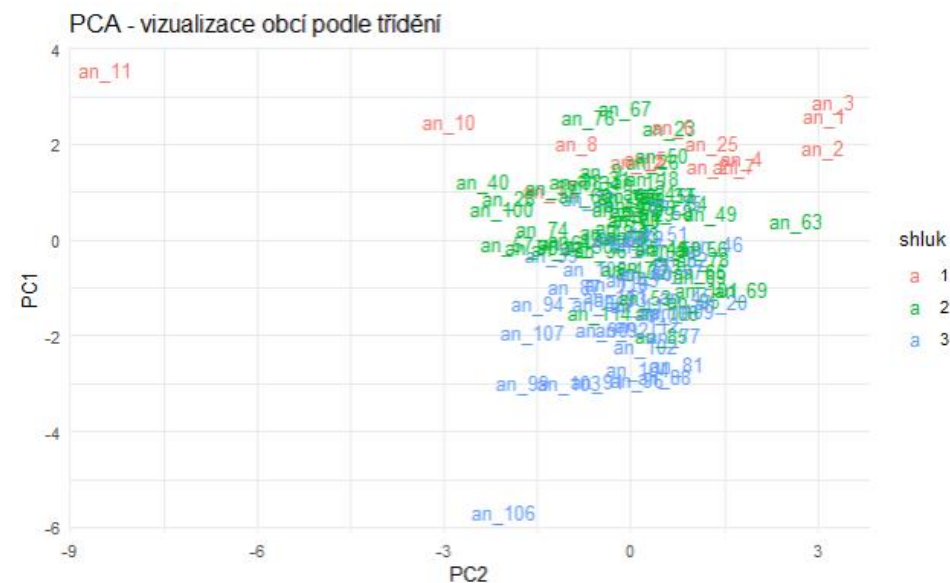
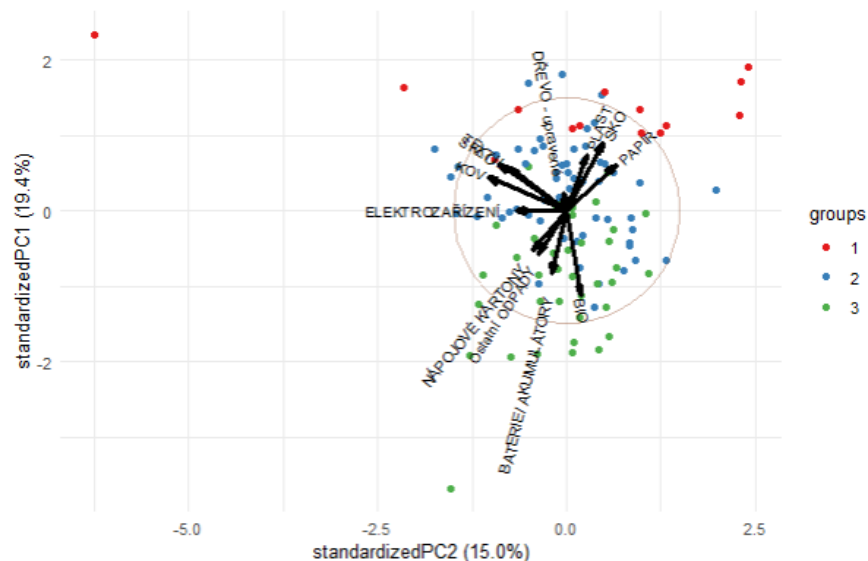
PLAST ▼

SKO ▼

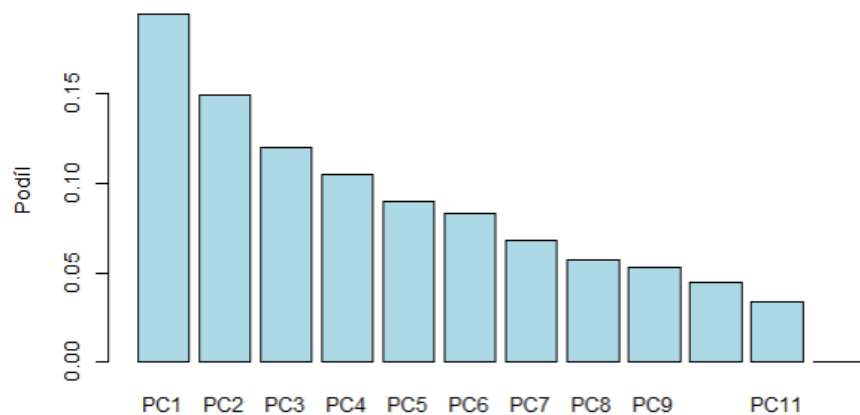
Absolutní ☒ Relativní



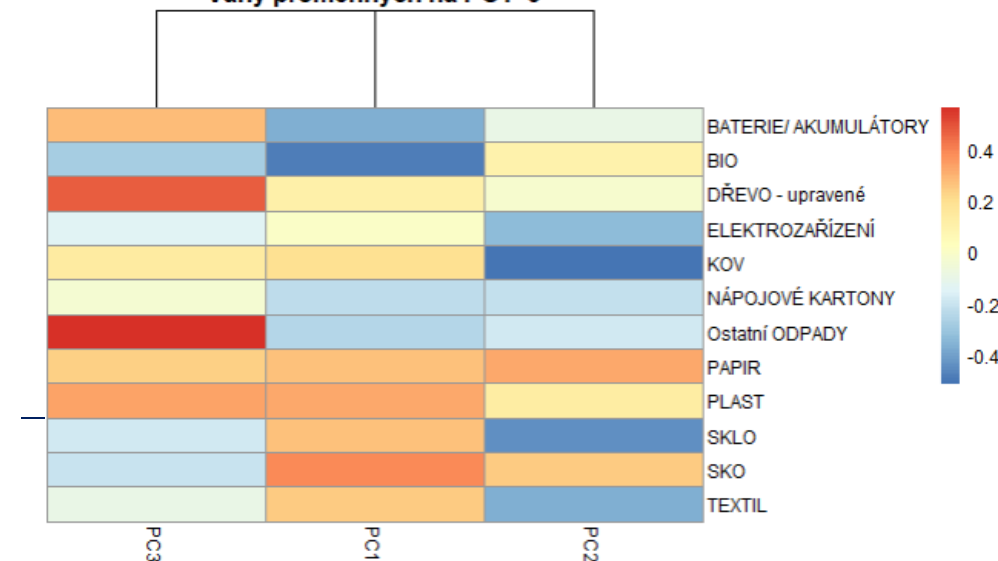
dataset	correlation	p_value	conf_interval
Všechna data	0.04	6.59e-01	(-0.14, 0.22)
Bez odlehlých hodnot	0.15	1.26e-01	(-0.04, 0.33)



Podíl vysvětlené variance



Váhy proměnných na PC1-3



Data FAO

Jsou načtena data FAO a AI zná jejich strukturu.

Zkuste napsat příkaz pro AI, podle kterého bude vygenerován graf.

Napište příkaz pro tvorbu grafu

Vykresli časový trend jednotlivých složek odpadu. Pro každý rok spočítej medián z poměrů. Graf obarvi podle klasických barev odpadu (papír modrý, sklo žluté..)

Proveď příkaz

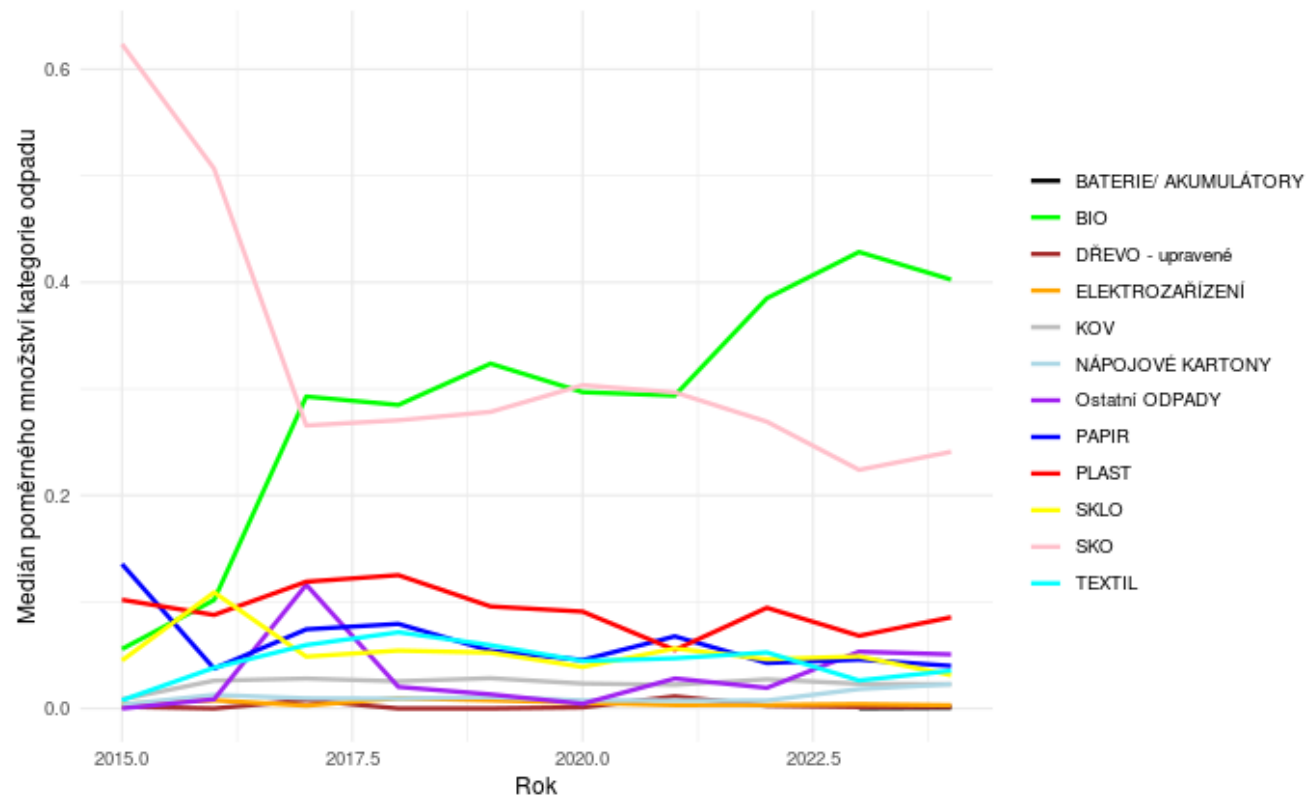
Graf

Kód grafu

Interaktivní graf

Testované prompty

Časový trend jednotlivých složek odpadu



Data FAO

Jsou načtena data FAO a AI zná jejich strukturu.

Zkuste napsat příkaz pro AI, podle kterého bude vygenerován graf.

Napište příkaz pro tvorbu grafu

Vytvor krabicové grafy v jednotlivých letech pro poměrné množství SKO a BIO a rozdel je pomocí grid do jednoho sloupce.

Proveď příkaz

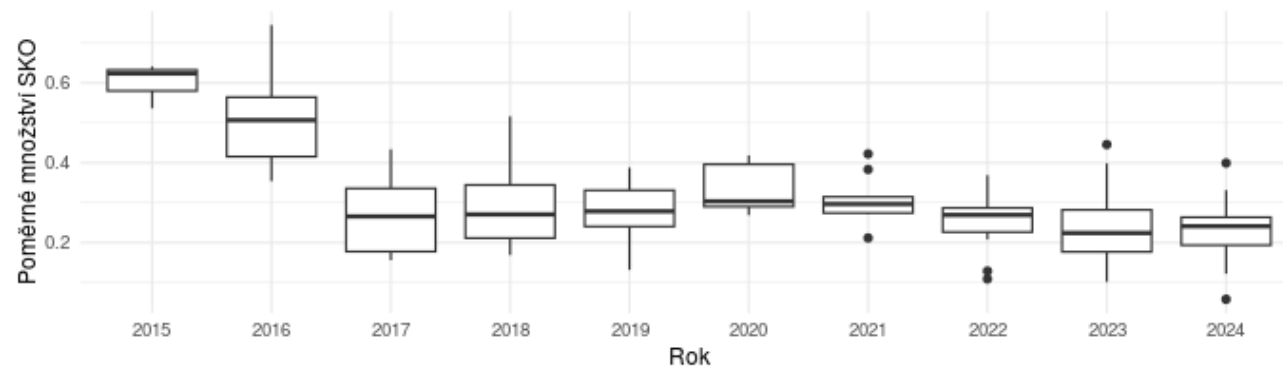
Graf

Kód grafu

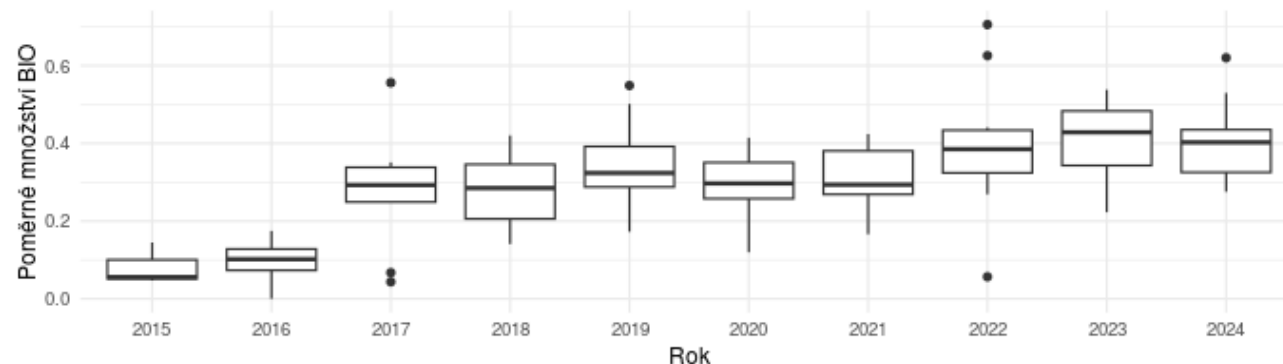
Interaktivní graf

Testované prompty

Krabicový graf poměrného množství SKO



Krabicový graf poměrného množství BIO



Děkuji za pozornost

