



## Zadání bakalářské práce

<b>Název:</b>	Hledání a analýza proměnných souvisejících s TFR
<b>Student:</b>	Jan Bartoš
<b>Vedoucí:</b>	PhDr. Ing. Tomáš Evan, Ph.D.
<b>Studijní program:</b>	Informatika
<b>Obor / specializace:</b>	Informační systémy a management
<b>Katedra:</b>	Katedra softwarového inženýrství
<b>Platnost zadání:</b>	do konce letního semestru 2022/2023

### Pokyny pro vypracování

Nastudujte problematiku total fertility rate (TFR) a získejte vhodné datové zdroje pro vytvoření analýzy v Pythonu, která bude hledat souvislost TFR s četností vyhledávání různých socio-ekonomických pojmů v různých státech a jazycích v čase. Reflektujte veličiny, u kterých bylo prokázáno že mají na tuto hodnotu vliv. Proveďte analýzu několika online zdrojů (Google Trends, Google Ngram atd.) a nalezněte veličiny s významnou korelací s TFR.



Bakalářská práce

# HLEDÁNÍ A ANALÝZA PROMĚNNÝCH SOUVISEJÍCÍCH S TFR

**Jan Bartoš**

Fakulta informačních technologií  
Katedra softwarového inženýrství  
Vedoucí: PhDr. Ing. Tomáš Evan, Ph.D  
4. ledna 2023

České vysoké učení technické v Praze  
Fakulta informačních technologií

© 2022 Jan Bartoš. Všechna práva vyhrazena.

*Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení, je nezbytný souhlas autora.*

Odkaz na tuto práci: Bartoš Jan. *Hledání a analýza proměnných souvisejících s TFR*. Bakalářská práce. České vysoké učení technické v Praze, Fakulta informačních technologií, 2022.

## Obsah

Poděkování	viii
Prohlášení	ix
Abstrakt	x
Seznam zkratek	xi
Úvod	1
<b>1 Analýza</b>	<b>3</b>
1.1 Úhrnná plodnost	3
1.1.1 Výběr sledovaných zemí	5
1.1.2 Úhrnná plodnost v Brazílii	5
1.1.3 Úhrnná plodnost v České republice	7
1.1.4 Úhrnná plodnost ve Francii	7
1.1.5 Úhrnná plodnost v Německu	9
1.1.6 Úhrnná plodnost v Nizozemí	10
1.1.7 Úhrnná plodnost ve Spojených státech amerických	11
1.1.8 Úhrnná plodnost ve Španělsku	11
1.2 Faktory úhrnné plodnosti	12
1.2.1 Zbraně a zločiny	13
1.2.2 Antikoncepce	13
1.2.3 Sport	13
1.2.4 Bydlení	13
1.2.5 Drogy	13
1.2.6 Vysoké vzdělání	14
1.2.7 Pracovní zátěž	14
1.2.8 Chudoba a bohatství	15
1.2.9 Vliv zábavního obsahu	15
1.2.10 Sociální mobilita	15
1.2.11 Náboženství	15
1.2.12 Zdravotnictví	15
1.2.13 Steroidy	16
1.3 Google Trends a Google Ngram	16
1.3.1 Google Trends	16
1.3.2 Google Ngram a národní korpusy	16
1.4 Analýza existujících řešení	16
1.4.1 Index budoucí orientace	16
1.4.2 Využívání Google Trends s plodností	16
1.5 Analýza uživatelských požadavků	17
1.5.1 Funkční požadavky	17
1.5.2 Nefunkční požadavky	17
1.6 Případy užití	17

1.6.1	Pokrytí funkčních požadavků . . . . .	18
1.7	Analýza technologií . . . . .	18
1.7.1	Python . . . . .	18
1.8	Statistické nástroje . . . . .	18
1.8.1	Pearsonův korelační koeficient . . . . .	18
1.8.2	Spearmanův korelační koeficient . . . . .	19
1.8.3	p-hodnota . . . . .	19
1.8.4	Kovariance . . . . .	19
1.8.5	Lineární regrese . . . . .	19
1.8.6	Kvadratická regrese . . . . .	19
<b>2</b>	<b>Návrh</b> . . . . .	<b>21</b>
2.1	Další technologie . . . . .	21
2.1.1	Git . . . . .	21
2.1.2	Ukládání dat . . . . .	21
2.2	Data . . . . .	21
2.2.1	Získání základních klíčových slov . . . . .	21
2.2.2	Překlad klíčových pojmů . . . . .	21
2.2.3	Zpracování a struktura dat . . . . .	22
2.3	Aplikace . . . . .	23
2.3.1	Jednotlivé veličiny . . . . .	23
2.3.2	Kategorie . . . . .	23
2.3.3	Zajímavé objevy . . . . .	25
<b>3</b>	<b>Implementace</b> . . . . .	<b>27</b>
3.1	Získání dat . . . . .	27
3.1.1	Vytvoření datasetu . . . . .	27
3.1.2	Překlad datasetu . . . . .	27
3.1.3	Google Trends . . . . .	28
3.1.4	Google Ngram . . . . .	28
3.1.5	Problémy s automatizací . . . . .	28
3.2	Zpracování dat . . . . .	28
3.2.1	Spojení dat . . . . .	28
3.2.2	Korelace pro vyhledávání . . . . .	28
3.2.3	Seskupení do kategorií . . . . .	29
3.3	Vytvoření aplikace . . . . .	30
3.3.1	Zobrazení statistických veličin . . . . .	30
3.3.2	Dvouosý graf pro veličiny a TFR . . . . .	30
3.3.3	Graf na zobrazení regresí . . . . .	30
3.3.4	Přístup a instalace . . . . .	32
<b>4</b>	<b>Testování</b> . . . . .	<b>33</b>
4.1	Testovací scénáře . . . . .	33
<b>5</b>	<b>Zhodnocení veličin</b> . . . . .	<b>35</b>
5.1	Google Ngram . . . . .	35
5.2	Google Trends . . . . .	35
5.2.1	Antidepresiva . . . . .	35
5.2.2	Auta . . . . .	35
5.2.3	Charity . . . . .	36
5.2.4	Konspirační teorie . . . . .	37
5.2.5	Konzumerismus . . . . .	37
5.2.6	Antikoncepce . . . . .	37

5.2.7	Zločiny . . . . .	38
5.2.8	Vzdělání . . . . .	38
5.2.9	Zábava . . . . .	39
5.2.10	Environmentalismus . . . . .	39
5.2.11	Gambling . . . . .	40
5.2.12	Hry . . . . .	40
5.2.13	Zbraně . . . . .	40
5.2.14	Zdravotní péče . . . . .	41
5.2.15	Bydlení . . . . .	41
5.2.16	Ideologie . . . . .	41
5.2.17	Nelegální drogy . . . . .	42
5.2.18	Investování . . . . .	42
5.2.19	Práce . . . . .	43
5.2.20	Legální drogy . . . . .	44
5.2.21	Nízký příjem . . . . .	44
5.2.22	Manželství . . . . .	44
5.2.23	Zprávy . . . . .	45
5.2.24	Náboženství . . . . .	45
5.2.25	S&P 500 . . . . .	45
5.2.26	Sporty . . . . .	46
5.2.27	Steroidy . . . . .	47
5.2.28	Turismus . . . . .	47
5.2.29	Univerzity . . . . .	47
5.2.30	Emigrace . . . . .	48
5.2.31	Statistiky . . . . .	48
5.2.32	Jiné . . . . .	48
<b>6</b>	<b>Závěr</b>	<b>51</b>
	<b>Obsah příloženého média</b>	<b>59</b>

## Seznam obrázků

1.1	Úhrnná plodnost ve světě z dat World Bank[4] . . . . .	5
1.2	Úhrnná plodnost v Brazílii z dat OECD[3] . . . . .	5
1.3	Úhrnná plodnost v České republice z dat OECD[3] . . . . .	7
1.4	Úhrnná plodnost ve Francii z dat OECD[3] . . . . .	8
1.5	Rozdíly úhrnné plodnosti mezi západním a východním Německem [21] . . . . .	10
1.6	Úhrnná plodnost v Německu z dat OECD[3] . . . . .	10
1.7	Úhrnná plodnost v Nizozemsku z dat OECD[3] . . . . .	11
1.8	Úhrnná plodnost ve Spojených státech amerických z dat OECD[3] . . . . .	12
1.9	Úhrnná plodnost ve Španělsku z dat OECD[3] . . . . .	12
2.1	Diagram rozložení dat pro data z Google Ngram . . . . .	22
2.2	Diagram rozložení dat pro data z Google Trends . . . . .	23
2.3	Vizualizace stránky pro jednotlivé veličiny . . . . .	24
2.4	Vizualizace stránky pro zajímavé objevy . . . . .	25
3.1	Regrese pro Ikea z dat Google Trends v České republice . . . . .	31

## Seznam tabulek

1.1	Pokrytí funkčních požadavků . . . . .	18
5.1	Statistika antidepresiv . . . . .	36
5.2	Statistika aut . . . . .	36
5.3	Statistika charit . . . . .	36
5.4	Statistika konspiračních teorií . . . . .	37
5.5	Statistika konzumerismu . . . . .	37
5.6	Statistika antikoncepcí . . . . .	38
5.7	Statistika zločinu . . . . .	38
5.8	Statistika vzdělání . . . . .	38
5.9	Statistika zábavy . . . . .	39
5.10	Statistika environmentalismu . . . . .	39
5.11	Statistika gamblingu . . . . .	40
5.12	Statistika her . . . . .	40
5.13	Statistika zbraní . . . . .	41
5.14	Statistika zdravotní péče . . . . .	41
5.15	Statistika bydlení . . . . .	42
5.16	Statistika ideologií . . . . .	42



5.17	Statistika nelegálních drog . . . . .	43
5.18	Statistika investování . . . . .	43
5.19	Statistika práce . . . . .	43
5.20	Statistika legálních drog . . . . .	44
5.21	Statistika nízkých příjmů . . . . .	44
5.22	Statistika manželství . . . . .	45
5.23	Statistika zpráv . . . . .	45
5.24	Statistika náboženství . . . . .	46
5.25	Statistika S&P 500 . . . . .	46
5.26	Statistika sportů . . . . .	46
5.27	Statistika steroidů . . . . .	47
5.28	Statistika turismu . . . . .	47
5.29	Statistika univerzit . . . . .	48
5.30	Statistika emigrací . . . . .	48
5.31	Statistika statistik . . . . .	48
5.32	Statistika jiných dat . . . . .	49

## Seznam výpisů kódu

3.1	Kód k překladu z DeepL . . . . .	27
3.2	Spojení a komprese dat . . . . .	29
3.3	Vypočítání korelace dat . . . . .	29
3.4	Shlukování dat . . . . .	29
3.5	Kód pro sloupcové widgety . . . . .	30
3.6	Kód dvouosého grafu pro zobrazení TFR a dat . . . . .	31
3.7	Kód pro zobrazení regresí . . . . .	32
4.1	Kód ukazující opravu chyby . . . . .	34

*Děkuji vedoucímu práce PhDr. Ing. Tomášovi Evanovi, Ph.D za ochotu, pomoc a rady při tvorbě této práce. Dále bych chtěl poděkovat rodině a přátelům za podporu při studiu.*

## Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací. Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 citovaného zákona.

V Praze dne 4. ledna 2023

.....

## Abstrakt

Cílem této práce je analyzovat porodnost a její závislost na různých trendech ve vyhledávání na internetu nebo v literatuře. K analýze trendů je využíván nástroj Google Trends a Google Ngram. V práci zkoumám klíčová slova, která jsem sestavil na základě faktorů, které byly již identifikovány odbornou veřejností i těch, které literatura neoznačila za pravděpodobné. Každé klíčové slovo je zařazeno do kategorie, která bude sledována s větší pozorností.

K analýze patří aplikace napsaná v Pythonu za pomoci knihovny Streamlit, která z ní dělá aplikaci, na kterou lze přistoupit z internetu. Výsledkem práce je interaktivní a veřejně přístupná aplikace, která ukazuje vývoj porodnosti v čase v různých státech světa s podobností s vyhledáváním na internetu a literatuře se statistikami, které mohou uživateli pomoci s odhalením, zda má vyhledávání klíčových slov vztah s porodností.

**Klíčová slova** TFR, Porodnost, Google Trends, Google Ngram, aplikace, Streamlit

## Abstract

The aim of this thesis is to analyze the fertility rate and its dependence on various trends in Internet or literature searches. Google Trends and Google Ngram are used to analyze the trends. In this thesis, I examine the keywords that I have compiled based on factors that have already been identified by the literature and those that have not been identified as likely by the literature. Each keyword is placed in a category that will be monitored with more attention.

The analysis includes an application written in Python using the Streamlit library, which makes it an application that can be accessed on the Internet. The result of the work is an interactive and publicly accessible application that shows the evolution of fertility rates over time in different countries of the world with similarities to internet and literature searches with statistics that can help the user to discover if keyword searches are in relation to fertility rates.

**Keywords** TFR, Fertility rate, Google Trends, Google Ngram, application, Streamlit

## Seznam zkratek

API	Application Programming Interface
CSV	Comma-separated values
cURL	Client for URLs
JSON	JavaScript Object Notation
NAP	Non-aggression principle
OECD	Organisation for Economic Co-operation and Development
OPEC	Organization of the Petroleum Exporting Countries
OR	Odds ratio
TFR	Total Fertility Rate



# Úvod

Plodnost je jedním z největších ekonomických témat dnešní doby. Plodnost v téměř všech státech dlouhodobě klesá, což do budoucna vytváří problémy udržitelnosti sociálního státu tak jak jej zná většina Evropy. [1] Plodnost nad míru udržitelnosti (2,1) nám zajišťuje rostoucí nebo žádný přírůstek obyvatelstva (více lidí se rodí než umírá), který je dostatečný pro udržení ekonomické úrovně. V současnosti v Evropě není stát, který má TFR nad 2,1. [2]

Výsledek práce je určen široké veřejnosti, která se chce informovat o porodnosti v různých státech světa a jejich vlivů na ně, protože je důležité, aby se informace o klesající porodnosti dostala do povědomí občanů.

Cílem práce je analýza tohoto problému a webová aplikace zabývající se trendy na internetu a literatuře v souvislosti s porodností. Aplikace nechá uživatele vybrat si z mnoha klíčových pojmů, ze sedmi různých zemí: Brazílie, České republiky, Francie, Německa, Nizozemska, Spojených států amerických a Španělska. Tyto pojmy pak aplikace zobrazí v grafu s TFR daného státu. K těmto datům budou k dispozici i různá statistická data jako je Pearsonova a Spearmanova korelace, lineární regrese a kvadratická regrese.

V první kapitole se zabývám pojmy úhrnné porodnosti. Na tu se poté podívám ve zkoumaných státech, tam v některých případech rozeberu i zajímavé lokální skutečnosti, poté analyzuji možné faktory úhrnné plodnosti jako například zbraně, drogy, sportovní výsledky, bydlení nebo pracovní zátěž. Následně se budu věnovat rešerší prací, které byly provedeny nad Google Trends a Google Ngram či jiných národních korpusů. Závěrem kapitoly se budu věnovat analýze funkčních požadavků v aplikaci. V druhé kapitole se zabývám návrhem aplikace, jak budou vypadat data, jak je budu zpracovávat a rozvržením stran ve webové aplikaci. Ve třetí kapitole se zaměřím na implementaci, na získání dat, zpracování dat a implementaci webové aplikace. Ve čtvrté kapitole se zaměřím na testování aplikace. V páté kapitole analyzuju zkoumané faktory porodnosti zjištěné v první kapitole a rozeberu je.





# Kapitola 1

## Analýza

*V této kapitole si rozeberu, co je to míra plodnosti. Dále se zaměřím na porodnost ve zkoumaných státech. To je Brazílie, Česká republika, Francie, Německo, Nizozemí, Spojené státy americké, a Španělsko. Dále prozkoumám vybrané kategorie, ze kterých poté budou vycházet klíčová slova v praktické části. Tuto kapitolu zakončím rešerší studií spojených s Google Trends a Google Ngram (nebo národních korpusů).*

### 1.1 Úhrnná plodnost

OECD definuje úhrnnou plodnost v určitém roce jako celkový počet novorozenat, která by se mohla narodit každé ženě, kdyby se dožila konce svého plodného věku a porodila děti v souladu s převládajícími mírami plodnosti podle věku. Je vypočítána součtem věkově specifických měr porodnosti definovaných v pětiletých intervalech. Aby se uchovala stabilní populace, je třeba, aby na každou ženu vycházela úhrnná plodnost 2,1. Aby tato hodnota byla validní, je třeba počítat, že rozdíl imigrace a emigrace je nulový, a také s tím, že míra úmrtí je stabilně stejná. Mezi příčiny dramatického poklesu porodnosti patří odkládání zakládání rodin, rození dětí a snižování velikosti rodin.[3]

Encyklopedie Britannica pak téma rozvíjí o další údaje ohledně takzvané míry reprodukce (replacement rate), což je již udávaná hodnota úhrnné plodnosti 2,1. Pokud je úhrnná plodnost větší než 2,1 pak se v dané oblasti populace zvýší, je-li tomu naopak, populace se časem sníží. Do toho se pak mohou vložit již zmíněné faktory jako vyšší úmrtnost nebo transfery lidí mezi oblastmi. Přesněji řečeno, pokud je v dané oblasti mnoho žen v plodném věku a malá populace starších osob, pak bude úmrtnost nízká, takže i když bude úhrnná plodnost menší než 2,1, tak stále může zůstat populace na stejné úrovni, ne-li se i zvýšit. Avšak tento trend má dočasný trvání a může přetrvávat maximálně desítky let.

Sledování úhrnné porodnosti může státům pomoci s efektivnějším a prospěšnějším plánováním a s alokací zdrojů. Pokud země zažívá dlouhodobě vysokou míru úhrnné porodnosti, může se na nové občany připravit a vybudovat infrastrukturu potřebnou k jejich vzdělání nebo zajistit sociální programy na podporu rodin.[2]

Po druhé světové válce ve Spojených státech nastalo období zvané Baby boom, ve kterém se zvýšil počet svateb a zvýšila se úhrnná plodnost. Tento nárůst plodnosti je historicky vyjimečný a krátkodobý a nemůžeme očekávat, že se plodnost dlouhodobě zvýší nad 2,1. V Československu v 70. letech nastal podobný jev, během kterého se zrodila generace Husákových dětí. Baby boom éra pozastavila dlouhotrvající pokles plodnosti, který započal již v devatenáctém století. Tento růst byl nečekaný a nenastal pouze ve Spojených státech, ale napříč celým západním industrializovaným světem. Demografové konec éry datují do konce 60. nebo začátku 70. let. Vzhledem k tomu, že lidé, narození během ranného Baby boomu jsou již v důchodovém věku,

představují zvýšenou zátěž pro zdravotnictví a penzijní systém. [1]

Ačkoliv míra reprodukce zůstává v mnohých částech světa vysoko nad úrovní 2,1. Například v afrických státech jako Niger, Somálsko nebo Demokratická republika Kongo dostahuje úhrnná plodnost hodnoty přes 5,7, celosvětová míra reprodukce se od poloviny 60. let snižuje. V té době činila míra reprodukce 5,1, v roce 2020 to bylo jen 2,4. Již zmíněné země se také potýkají s pádem úhrnné plodnosti, například Demokratická republika Kongo měla největší plodnost za posledních 80 let v polovině devadesátých let a to 6,8. Na druhém konci spektra jsou dvě asijské země a to Korejská republika (Jižní Korea) a Hong Kong s plodností 0,8 respektive 0,9. Plodnost obou zemí klesá celou dobu měření těchto dat Světovou bankou, to je od 60. let minulého století. Pod míru reprodukce klesly již v 80. letech. [4]

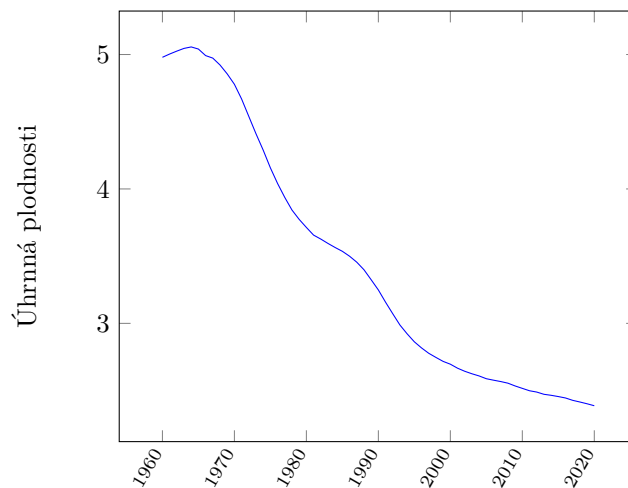
Jak již bylo zmíněno, velká část vyspělého světa po druhé světové válce zaznamenala zvýšený počet úhrnné plodnosti. Strach z dalšího růstu, který by zapříčinil zhoršení dalšího vývoje rozvojových zemí, dal za vznik iniciativě ke snížení populačního růstu v rozvojových zemích. Z počátku za touto iniciativou stály elity ze západního světa a Indie. Za podpory soukromých nadací, vládních organizací a neziskových organizací začal stoupat tlak na zavedení nástrojů pro snížení produkce populace. V roce 1974 proběhla Světová populační konference v Bukurešti, která se zabývala akčním plánem světové populace. Po 2 letech 40 zemí světa, které představovaly 58 procent světové populace, zavedly programy na snížení plodnosti. [5] [6]

Na druhou stranu, počet zemí OECD, které se snažily počet své populace zvyšovat nebo zanechat stejný, byl pouze 7 a to v letech od 1976 do 1993. Do roku 2003 se tento počet zvětšil na 13 zemí. Důvodem, proč většina zemí prudký pokles ignorovala, byla hypotéza demografů, že snížení plodnosti je dočasným fenoménem, zapříčiněný opožděným vstupem do manželství a rodičovství. V některých zemích se zdálo, že hypotéza je pravdivá, například ve Spojených státech, kde korekce přišla během osmdesátých let. Ale země jako například Itálie či Rakousko se své korekce nedočkaly. Tato dlouhodobě nízká plodnost zanechala šrám na věkové struktuře obyvatel těchto států. Další důvod, proč se pokles ignoroval, vycházel z přesvědčení, že proporodní politiky, jsou často velmi drahé a setkávají se s malým úspěchem. [5]

Toto přesvědčení mohlo pramenit z práce doktora Glasse, který se zaměřoval na efektivitu pronatalních politik během třicátých let. Rozděluje politiky na pozitivní a negativní. Mezi negativní politiku řadil například zákazy potratů. Pozitivní politiky rozděloval na tři. Přidávky na děti, rodinné půjčky a speciální zacházení s rodinami, jako například nárok na bydlení nebo zlevněné jízdenky na vlak. Žádné z těchto politik neměly vliv na zvýšení plodnosti s výjimkou nacistického Německa, kde se zavedly půjčky pro rodiny, tam ale Glass argumentuje tvrzením, že se jednalo o odložení plození dětí z důvodu Velké hospodářské krize.[7]

Naopak zjištění z pozdějších studií ukazuje, že vlády mohou svými politikami ovlivňovat plodnost. Studie tvrdí, že zrušení politik, které se věnovaly podpoře rodin v Polsku, Východním Německu a Španělsku přispělo k pádu plodnosti v těchto zemích. Studie ukazují slabý pozitivní vztah mezi reprodukčním chováním a různými finančními benefity a daňovými úlevami. [5]

Vliv na úhrnnou porodnost mohou také mít nálady ve společnosti. Tohle se stalo očividným po pádu železné opony, kdy se po sjednocení Spolkové republiky Německo ukázaly velké rozdíly v úhrnné porodnosti mezi západním a východním Německem. Pád komunistického režimu způsobil velkou ekonomickou nestabilitu. Zatímco porodnost v západním Německu nevykazovala statisticky významnou odchylku vůči normálu, ve východním Německu se porodnost významně snížila. Podobný pád úhrnné porodnosti nastal i v Polsku, které po úspěšných demonstracích a vyjednáváních odborového hnutí Solidarność přešlo na volnotržní ekonomiku. Oběma těmito státy se již podařilo plodnost vyrovnat na úroveň západoevropských zemí. Ale i tyto státy se potýkají s dlouhodobým poklesem plodnosti. [2]



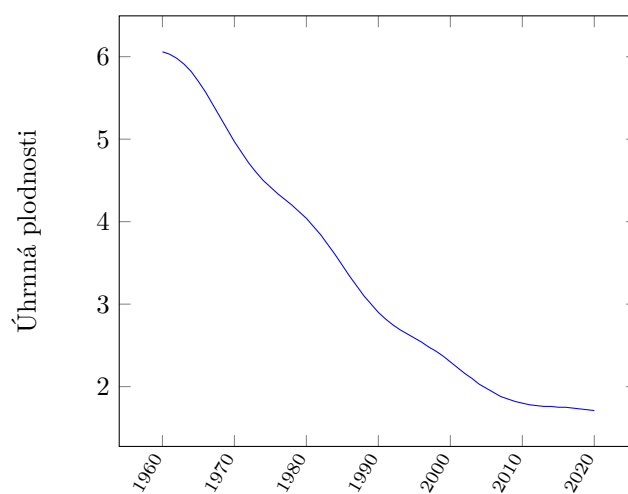
■ **Obrázek 1.1** Úhrnná plodnost ve světě z dat World Bank[4]

### 1.1.1 Výběr sledovaných zemí

Zkoumané země jsem vybral na základě těchto kritérií:

- Země s větším počtem obyvatel
  - Větší počet zdrojů informací
  - Obsáhlejší data z Google Trends - více uživatelů = více dat
- Evropské jazyky
  - Překlady mezi evropskými jazyky a neevropskými bývají do dnes stále nepřesné [8]

### 1.1.2 Úhrnná plodnost v Brazílii



■ **Obrázek 1.2** Úhrnná plodnost v Brazílii z dat OECD[3]

Strmý pád porodnosti v Brazílii započal během padesátých let minulého století, kdy se země potýkala s velkými sociálními změnami, tyto změny šly ruku k ruce se zvýšeným ekonomickým růstem. Modernizace země poté změnila charakter ekonomických aktivit a zrychlila migraci obyvatel z venkova do měst. Společnost se postupně transformovala do konzumní společnosti. Tyto faktory poté zvýšily náklady s pořízením dítěte a také pokřivila představu o tom, jak velká má být rodina.

Modernizace země měla však i další dopad na porodnost a to byla lepší dostupnost antikoncepce. Antikoncepce avšak nebyla jediným faktorem, dalším byla i lepší dostupnost interrupce a sterilizace. Brazilská vláda se nikdy nepokoušela plodnost snižovat, ale snahy o modernizaci sociální, politické a ekonomické sféry země akcelerovaly sociální vývoj a pomohly k snížení porodnosti.

Brazilský případ ukazuje, jak dopadají nepřímé důsledky hospodářských reforem na porodnost a také, že k propadu porodnosti není třeba silných podpor plánovaného rodičovství. Mezi faktory, které mohly snížení způsobit, patří především změny ve vztazích mezi muži a ženami, mateřské vzory chování, mezigenerační změny v transféru zdrojů, zapojení žen do pracovního procesu, složení domácností a urbanizace. Tyto faktory mohou mít vliv na význam dětí v životě žen a na schopnost výchovy dětí a tím pádem i na porodnost.[9]

### 1.1.2.1 Vliv telenovel na porodnost

V roce 2012 vyšel ve vědeckém časopisu *American Economic Journal: Applied Economics* článek od trojice Eliana La Ferrara, Alberto Chong a Suzanne Duryea, který se zaměřuje na vliv televizního vysílání na porodnost žen. Specializují se na telenovely televizního kanálu Rede Globo, který měl do ranných devadesátých let dominanci na trhu s telenovelami. 72 % žen, které se v telenovelách objevily a měly méně než 50 let, byly bezdětné a 21 % mělo pouze jedno dítě. Zkoumala se data mezi lety 1979 až 1991, kdy byla plodnost v Brazílii výrazně vyšší, než bylo zobrazeno v těchto pořadech, konkrétně v roce 1979 byla plodnost 4,1 a v roce 1991 2,8. Zjistili, že pokud bylo možné v lokalitě naladit kanál Rede Globo, výrazně to snížilo porodnost v regionu. Tento efekt se nejsilněji projevoval u žen z nižších sociálně-ekonomických vrstev a také u žen, které byly ve střední či pozdní fázi svého plodného věku, což naznačuje, že nebylo prvorodiček, ale spíše televize přispěla ke snížení porodnosti žen, které již dítě měly.

Toto zjištění může mít velké dopady na společnost v rozvojových zemích, které mají často malou gramotnost, takže audiovizuální média mají často největší vliv na informovanost. Jejich práce naznačuje, že programy zaměřené na kulturu místní populace mají velký potenciál dosahu za relativně nízkou cenu. Tyto programy by pak mohly být využity politiky k propagaci významných sociálních či ekonomických agend. [10]

### 1.1.2.2 Faktory ovlivňující porodnost místních a migrantů v urbanizovaných částech Brazílie

Článek zaměřující se na faktory porodnosti v městech od trojice S. Iutaka, E. W. Bock a W. G. Varnes se specializuje primárně na tyto faktory: věk, věk v době svatby, velikost města, migrační status, barva kůže, vzdělání, sociální status, tchánův sociální status a otcův sociální status.

Migranti mají tendenci přijmout porodní vzor od domácích obyvatel, kteří mají většinou nižší porodnost. Změny v porodnosti migrantů jsou ve vazbě s městem, kam emigrovali, sociálním zázemím a časem, který stráví v urbanistickém centru města. Dalším faktorem pro migranty je i věk, ve kterém migrují a také sociální třída, do které se po emigraci dostanou. Důležitost tchánova sociálního statutu v porovnání s otcovým sociálním statutem je menší, což indikuje, že ženy mají tendenci si uchovat své rodičovské hodnoty a nepřijmout manželovy. Počet faktorů, které predikují porodnostní vzory je početnější a statisticky signifikantnější u migrantů než u domácích obyvatel, to značí, že migranti tvoří z hlediska porodnosti heterogeničtější skupinu obyvatel. Míra

porodnosti migrantů se pak dramaticky snižuje v první generaci a v druhé generaci se již blíží porodnosti místních městských obyvatel. [11]

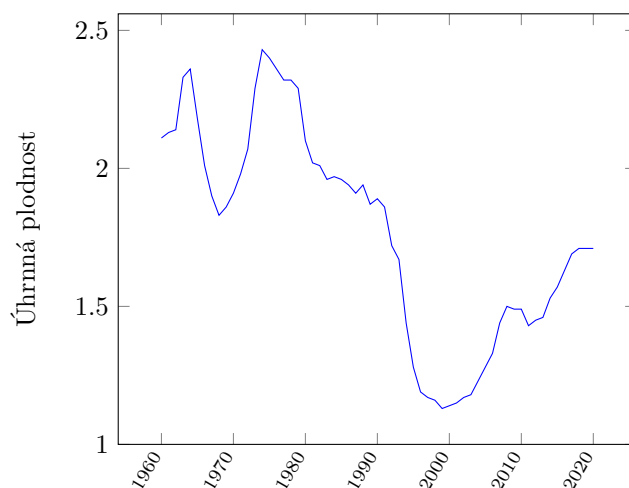
### 1.1.3 Úhrnná plodnost v České republice

Po pádu socialistického režimu se v Česku po roce 1989 změnilы vzory chování k rodině a rození dětí. Páry často odkládaly zakládání rodin, což zavinilo výrazný propad plodnosti, tato situace nejčastěji nastávala u mladých párů. Zvýšil se i počet nemanželských soužití, kde byly svatby často odsouvány nebo úplně ztraceny. Tyto jevy byly zapříčiněny změnou režimu, jak již bylo zmíněno i u Polska či východního Německa. Rychlost těchto změn může být přičítána několika současně nastávajícím jevům, například rozšíření vyššího vzdělání, výskyt nových pracovních příležitostí, které upozadovaly rodinný život, zvýšená konkurenceschopnost na trhu práce a zvyšující se ekonomická nejistota. [12]

Podobnému tématu se věnovali i Stephan Klasen a Andrey Launov, kteří dříve přišli s podobnými závěry. K identifikaci přechodově specifických rysů poklesu úhrnné plodnosti vyjádřili Heckmanův model za užití dat z roku 1998. Domnívají se, že pokles porodnosti je ovlivněn zvýšeným vlivem vyššího vzdělání, a omezené možnosti kombinovat pracovní a rodinný život a nedostatek školek a škol. Také zjišťují vysoký vliv antikoncepce, jejíž užití je motivováno ekonomickými i demografickými důvody. [13]

#### 1.1.3.1 Vztah mezi vírou v Boha a touhou po plodnosti

Stephen Cranney se ve svém článku této tématice věnoval ve dvou zemích postkomunistické části Evropy: Česku a Slovinsku. Využíval k tomu 2251 respondentek z České republiky a 951 respondentek ze Slovinska. Tato data byla nasbírána v průběhu devadesátých let. Ženy, které věřily v Boha chtěly přibližně o 0,2 více dětí než nevěřící a ty, které věřily v nadpřirozeno chtěly o 0,1 více dětí než nevěřící. [14]



■ **Obrázek 1.3** Úhrnná plodnost v České republice z dat OECD[3]

### 1.1.4 Úhrnná plodnost ve Francii

Porodnost ve Francii je relativně vysoká v porovnání s ostatními evropskými státy, navzdory faktorům, které ovlivňují porodnost jinde v Evropě, jako jsou odkládání rodičovství, rostoucí nepopularita manželství, užívání antikoncepce a vysoký počet interrupcí. Hlavním důvodem,

proč je porodnost ve Francii vyšší než jinde v Evropě (ale stále hluboce pod mírou reprodukce) je porodnost u žen ve věku 30-40 let, kdy je většina žen v pracovním procesu. Článek došel k několika možným vlivům na tento jev. Jedním z nich je velice proaktivní politika podpory rodičovství započatá již od konce druhé světové války. Tyto politické kroky mají podporu napříč francouzským politickým spektrem i společností. Tyto politiky mají pomáhat ženám kombinovat pracovní a rodinný život. [15] Jiný zdroj ale tvrdí že, Francií zavedená dotace 50 % na jesle v roce 2004 zvedla počet účastnic na trhu práce jen o jedno procento. Studie zmíněné v práci níže zjišťují, že finanční dotace pro hlídání dětí neovlivní zapojení do pracovního procesu v zemích, kde hlídání dětí je levnější, například v Norsku než v zemích, kde je hlídání dražší například ve Spojených státech nebo Kanadě. Studie dále tvrdí, že neexistuje žádný konsensus ohledně vlivu mateřských dávek a mateřské dovolené na trh práce. [16]

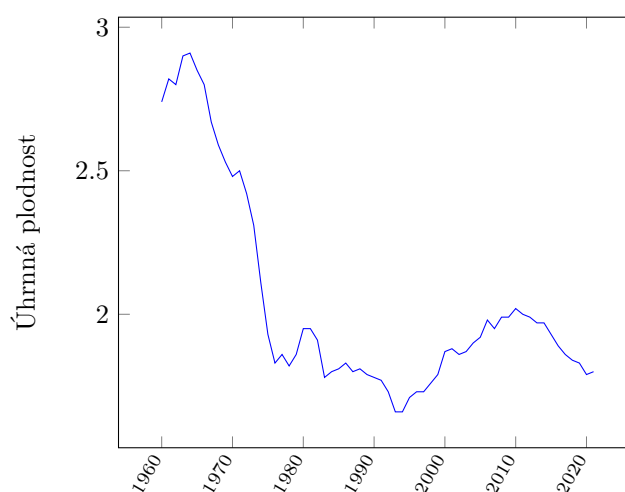
#### 1.1.4.1 Vzorec chování druhé generace migrantů k plodnosti

Článek Ariane Pailthé se zabývá vlivem druhé generace migrantů na vzorec plodnosti. Zabývá se čtyřmi množinami žen, jejichž rodiče jsou původem z Magrebu, Turecka, subsaharské Afriky a jihovýchodní Asie. Zjistila, že ženy s předky v jihovýchodní Asii mají nižší porodnost a více oddalují mateřství než ostatní zkoumané ženy. Dokonce vstupují do mateřství později a mají nižší porodnost prvorozených než rodilé Francouzky. Jedním z faktorů, proč se tak děje je jejich vysoké vzdělání a nižší náchylnost k vytváření svazků.

Oproti tomu dcery lidí tureckého původu vstupují do mateřství dříve než Francouzky, tohle platí i pro druhé dítě, které mají také dříve než Francouzky. Tento jev je zdůvodňován jejich kulturním pozadím a jejich nižším lidským kapitálem. Jako hlavní faktory bere stav, kdy ženy mluví doma pouze turecky, a jejich náboženství.

Potomci subsaharských migrantů typicky odkládají porod prvního i druhého dítěte, částečně i kvůli tomu, že odkládají vstup do svazku a také kvůli tomu, že si adaptovali francouzskou formu mateřství.

Děti migrantů z Magrebu vykazují podobnou porodnost (1.9) jako francouzské ženy (1.8) ve věku 40 let. Ale mají větší poměr rodin s třemi dětmi - 39 % oproti 26 %. [17]

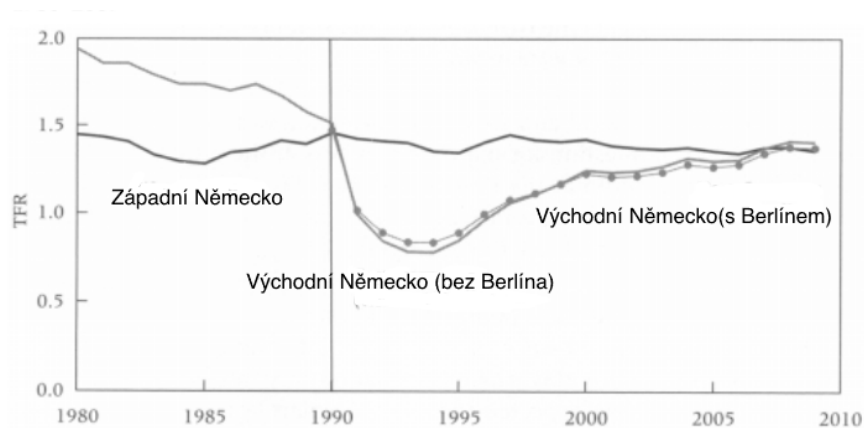


■ **Obrázek 1.4** Úhrnná plodnost ve Francii z dat OECD[3]

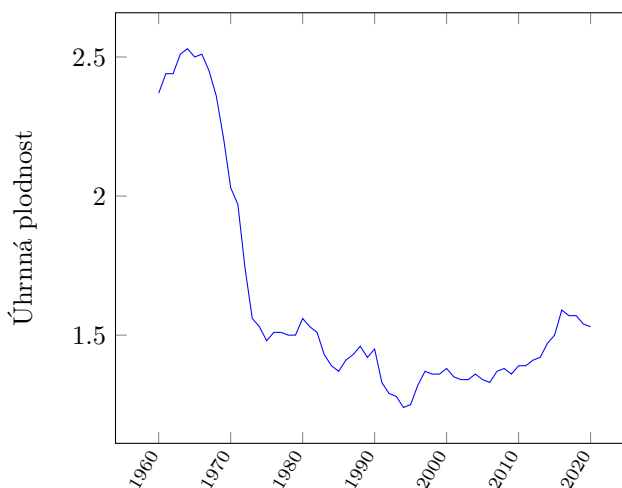
### 1.1.5 Úhrnná plodnost v Německu

Plodností se Německo od poloviny 20. století řadilo k těm nejhorším v Evropě. Dobritz (2008) [18] charakterizuje devět příčin nízké plodnosti v Německu.

- Individualizace - lidé se často nechtějí připravit o svůj volný čas, zájmy, kariéru a finanční stabilitu. Předsudky vůči lidem, kteří mají děti, jsou výsledkem individualizovaného stylu života, kdy děti a jejich náklady jsou brány jako soukromá záležitost.
- Bezdětnost - té je často přičítána nutnost žen si vybrat mezi rodinným a pracovním životem. Většinou založení rodiny znamená ztrátu pracovní pozice pro ženu. Proto hlavně ženy, které mají vyšší vzdělání, mají větší predispozici k odmítnutí založení rodiny. Chtěná bezdětnost se již stala životním stylem.
- Rodinná politika - ta je brána jako neúspěšná vzhledem k porodnosti. Německý model je postaven na finanční pomoci, která nahraňuje zažitým sociálním rolím, které berou muže jako chleboďárce a matku jako ženu v domácnosti. Model postaví ženu před rozhodnutí, zda chce pokračovat v práci nebo zplodit potomka a z práce odejít. Což mnoho žen utvrdí v jejich názoru a z práce neodejdou.
- Podmínky pro nově narozené dítě také v Německu mohou ovlivňovat porodnost. Dostupnost zařízení péče o děti je jedno z nejhorších v EU. Dle dat Eurostatu z roku 2020 80 % dětí, které mají méně než tři roky jsou hlídány rodiči. Toto číslo je největší v Evropské Unii.[19]
- Změna vzoru chování mladých párů - tradičně byl život ženy koncipován takto: škola, práce, sňatek, děti, odchod z práce. Tento model je stále nejčastější, ale nová data ukazují, že se často páry vyhýbají sňatkům a řeší tyto věci dohodami.
- Ekonomická situace párů - německé domácnosti zažívají výkyvy financí v čase, kdy se stanou rodiči. Na začátku mateřství mají peněz dostatek díky benefitům ze zdravotního pojištění, ale v momentě, kdy rodiče začnou přijímat podporu na dítě, která již není spojena s jejich předešlými příjmy, se jejich finanční situace výrazně zhorší. Mít dítě je v Německu bráno jako jedno z rizik chudoby. Jedno z deseti dětí je ohrožené chudobou. Novější zdroje (2017) dokonce zmiňují, že až 19,7 % dětí pod 18 let je ohrožených chudobou. Největší podíl na této hodnotě mají rodiče samoživitelé a imigranti. [20]
- Situaci ve východním Německu jsem již zmiňoval, zde je uvedena také. Porodnost byla v době psaní článku (2008) již na stejné úrovni jako v bývalém západním Německu. 1.5 Východní Německo je považováno za problematický region, kvůli horší ekonomické situaci, při které mají východní Němci nižší mzdu a cena, která stojí vychovat dítě, je stejná jako v západním Německu. Vysoká nezaměstnanost, ztráta mnoha lidí, ať už stárnutím obyvatelstva, nebo i emigrací, jsou důvody, které porodnosti nepomáhají. [21]
- Genderový problém se často řešil v severských částech Evropy, kde se snažili rodinný model chleboďárce a žena v domácnosti změnit. Což se v Německu nestalo. Model ženy v domácnosti a chleboďárce stále zůstává a to i poté, co žena po mateřství nastupuje do práce. Sociální role jí však zůstane a tak mimo pracovní závazek se stará také o dítě a domácnost, kdežto pozice otce dítěte se nemění a charakter práce v domácnosti je spíše ve formě výpomoci. Takže podíl práce v páru je asymetrický. Jiná studie [22] našla vztah mezi genderovou rovností a plodností a to ve tvaru písmene U, to znamená, že při nízké rovnosti, je porodnost vysoká, při zvýšení rovnosti se plodnost sníží a poté začne narůstat.
- Nízká žádoucí plodnost - mnoho let byl rozdíl mezi žádoucí plodností a reálnými čísly velký, ale v poslední době rozdíl mezi těmito veličinami prakticky vymizely. Nemůžeme předpokládat, že souhrnná porodnost může být větší než žádoucí plodnost.



■ **Obrázek 1.5** Rozdíly úhrnné plodnosti mezi západním a východním Německem [21]



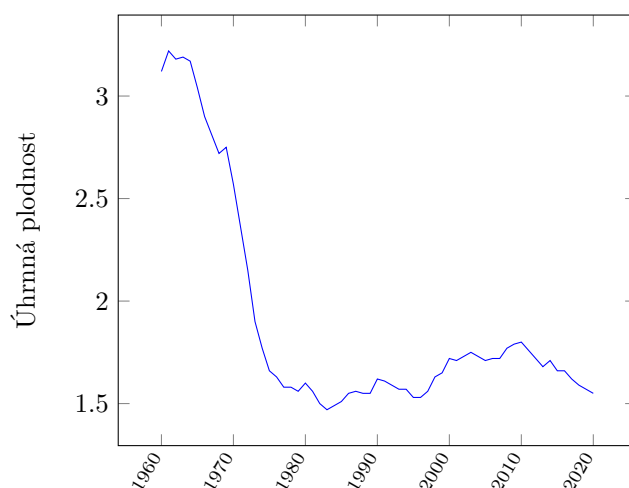
■ **Obrázek 1.6** Úhrnná plodnost v Německu z dat OECD[3]

### 1.1.6 Úhrnná plodnost v Nizozemí

Ani Nizozemsku se nevyhl výrazný pokles porodnosti. Z hodnot úhrnné plodnosti 3 v 60. letech minulého století na hodnotu 1,5 v letech 80. Od 80. let se situace stabilizovala a úhrnná plodnost od té doby osciluje kolem hodnoty 1,6. Nizozemská vláda se kromě monetární pomoci zaměřila hlavně na implementaci rodinných politik, které by měly pomoci rodičům kombinovat rodinný život s pracovním. Rozvoj prací se zkráceným úvazkem a flexibilní pracovní dobou pomohly ženám s dětmi zapojit se do pracovního procesu. Zvýšená porodnost a zvýšený počet žen na trhu práce ukazují, že politiky, které umožňují ženám jednoduchý návrat do pracovního procesu asi dosáhly ve své době účinku, ale po roce 2010 opět plodnost klesá. [23]

V důsledku toho má Nizozemsko největší podíl žen, které pracují na částečný úvazek na světě. Avšak nedávná finanční krize zpochybnila udržitelnost tohoto systému a signalizovala konec štědrých dávek v kombinaci se zvýšeným voláním po tom, aby se ženy staly méně závislými na svých manželech a byly více ekonomicky nezávislé. Což jim práce v se zkráceným úvazkem často neumožňuje. [24]





■ **Obrázek 1.7** Úhrnná plodnost v Nizozemsku z dat OECD[3]

### 1.1.7 Úhrnná plodnost ve Spojených státech amerických

Plodnost ve Spojených státech amerických klesala pomalejším tempem než vyspělé státy v Evropě. Například Německo se pod úroveň reprodukce dostalo už na začátku 70. let. Spojené státy spadly pod úroveň reprodukce také v 70. letech, ale díky již zmíněné korekci se opět plodnost dostala nad hladinu reprodukce. Až v 10. letech 21. století se pod tuto úroveň zase dostaly.

Faktory jako industrializace, urbanizace, zvyšování rodinných příjmů, zlepšení zdravotní dostupnosti a růst ženských mezd snížily benefity a zvětšily relativní cenu, kterou stojí přivést dítě na svět. Snížení dětské úmrtnosti a zlepšení dostupnosti antikoncepce změnilo vzory chování s ohledem na plození dětí. [25]

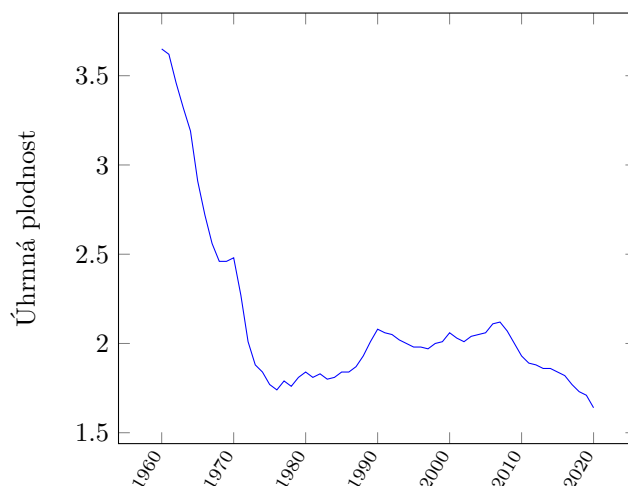
S. Philip Morgan a Heather Rackin zkoumali podobnost mezi záměry těhotenství a chováním ve Spojených státech. Zjistili, že ženy a muži, kteří rodičovství odkládali, měli ve výsledku méně dětí než předpokládali. Domnívají se, že faktory, které za tohle mohou, jsou snížená plodnost, opakované odkládání rodičovství kvůli jiným záležitostem a nenalezení vhodného partnera do manželství. [26]

Osobní ideály plodnosti u žen jsou pro většinu (69 %) stabilní, ale pro zbývající třetinu je tento parametr střídavě měnitelný ve vlnách. Zhruba polovina žen, u kterých se ideál mění, se jim ideál zvětšuje a polovině zmenšuje. Multinomiální logistická regrese naznačuje, že ženy s vyšším ideálem plodnosti v první vlně měly vyšší šanci na zvětšení jejich ideálu plodnosti a nižší šanci na snížení jejich ideálu plodnosti. Vyšší vzdělání u žen je spojováno s nižší šancí na zvětšení jejich ideálu plodnosti, naopak zaměstnání na plný úvazek mělo za následek větší šanci na snížení jejich ideálu plodnosti. [27]

### 1.1.8 Úhrnná plodnost ve Španělsku

Zhruba v polovině 70. let se výrazně začala propadat úhrnná porodnost. V té době došlo k několika událostem, které na to mohly mít vliv. První z nich je ropná krize z roku 1973, která byla zapříčiněna sedmdesátiprocentním zvýšením ceny ropy organizací OPEC v listopadu. Další zvýšení cen následovalo již v prosinci 1973 a to o dalších 130 %. Postupně se z 1,3 USD za barel v roce 1970, dostala cena ropy v roce 1981 na cenu 32 amerických dolarů za barel. Tohle zvýšení zapříčinilo vysokou inflaci hlavně v západní Evropě a severní Americe. [28] [29]

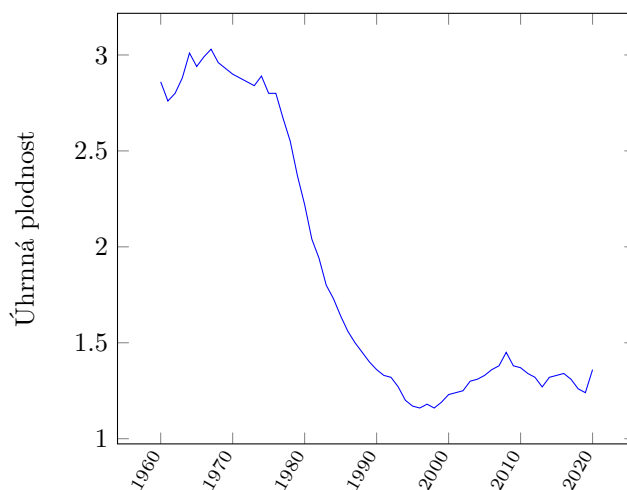
Další možnou událostí, která mohla ovlivnit porodnost ve Španělsku, byla smrt hlavy státu - nacionalisty Francisca Franca. Po jeho smrti se španělský král Juan Carlos chopil moci a rozložil



■ **Obrázek 1.8** Úhrnná plodnost ve Spojených státech amerických z dat OECD[3]

autoritářské instituce a podpořil vytvoření nových politických stran. [30]

Španělsko si od 70. let prošlo velkými sociálními, kulturními a politickými změnami. S nimi se i výrazně snížila porodnost. Tento pokles se od poloviny devadesátých let zastavil a začal střídě růst. Stále je ale Španělsko jedna ze zemí s nejnižší plodností v EU. Výzkumy dávají za vinu kromě sociálních a politických změn i absenci španělských vlád v řešení porodnosti. Ačkoli se od nového milénia zavedly nové zákony na podporu porodnosti, nejsou určitě jedinými strůjci zvýšené porodnosti, jelikož tyto politiky nemohou ovlivnit hluboce zakořeněné problémy španělské ekonomiky. Mezi tyto problémy se řadí vysoká nezaměstnanost, neflexibilní trh práce, zvýšené ceny nemovitostí, sekularizace společnosti a pozdější vstupy do manželství. [31]



■ **Obrázek 1.9** Úhrnná plodnost ve Španělsku z dat OECD[3]

## 1.2 Faktory úhrnné plodnosti

### 1.2.1 Zbraně a zločiny

Studie zaměřující se na čerstvě sezdané páry v Nepálu objevila, že ženy mohou být oběťmi domácího násilí a to ze strany partnerů a dalších lidí. Jejich objev podporuje hypotézu, že šance na spáchání násilí partnera na partnerce je vyšší za stavu, když je žena těhotná. Studie přispívá k další literatuře tím, že přidává důkazy o tom, že násilí v nových manželstvích je pozitivně spojováno s těhotenstvím a porodem. [32]

V Chicagu byla provedena studie na vzorku 175 065 dětí. Tyto děti byly rozděleny na třetiny podle lokalit s největší zbraňovou kriminalitou. V nejhorší třetině mělo 10,6 % dětí nižší porodní váhu, 10,6 % předčasný porod (před 37. týdnem těhotenství) a 9,1 % mělo tělesnou hmotnost menší než 90 % dětí. Za užití bílých žen jako referenčního bodu, poměr šancí nízké tělesné váhy byl u dětí od afroameričanek mezi 1,9 - 2,1 napříč všemi třetinami a 0,8 až 1,2 mezi dětmi hispánských žen. Poměr šancí pro předčasně narozené děti afroameričanek byl 1,6 až 1,7 a 1,0 až 1,2 pro ženy hispánského původu. Poměr šancí pro afroamerické ženy, které porodily dítě s nižší tělesnou hmotností menší než 90 % dětí, je 1,6 až 1,7 a pro hispánské ženy 0,9 až 1,0.

V Chicagu byla rasa a etnicita asociována s výsledky těhotenství bez zjevných spojení se zbraňovou kriminalitou.[33]

### 1.2.2 Antikoncepce

Studie Davida Antonia Sánchez-Páeze a Josého Antonia Ortegy se zaměřila na užívání antikoncepce u dospívajících a jejich vliv na plodnost. Přišli na to, že s častějším používáním antikoncepce se snížila plodnost dospívajících o 6,8 % v Latinské Americe a 4,1 % v subsaharské Africe. Uspokojení celkové poptávky po antikoncepčních přípravcích svobodných dospívajících by vedlo k dalšímu poklesu plodnosti o 8,9 % v Latinské Americe a 17,4 % v subsaharské Africe.[34] Existují zdroje, které tvrdí, že ženy používající antikoncepci tento fakt neodrazuje od dostažení svého požadovaného počtu dětí. Tedy, že antikoncepce nemá vliv na rozhodování, kolik dětí daná žena chce, ale je pouze nástrojem pro plánování. [35]

### 1.2.3 Sport

V květnu 2009 FC Barcelona vyhrála 3 soutěže (La Ligu, Ligu mistrů a Copa del Rey) o devět měsíců později místní média oznámila 45% nárůst porodnosti v Katalánsku. Tento jev, známý jako „Barcelonský babyboom“, poprvé pozorovali lékaři a zdravotníci, kteří zaznamenali náhlý nárůst počtu porodů v místních nemocnicích. Studie objevila významný nárůst porodnosti. Jejich návrh studie však neumožňuje potvrdit kauzální vztah mezi fotbalovým úspěchem a porodností. [36]

### 1.2.4 Bydlení

Studie zabývající se rozdíly v porodnosti v závislosti na typu bydlení přišla na to, že porodnost byla největší v jednogeneračních domech a nejnižší v bytech. Variace porodnosti se snížila, ale stále zůstala významnou po zohlednění demografických a socioekonomických rozdílů žen. Dále bylo zjištěno, že páry, které se přestěhovaly dosahovaly vyšší plodnosti, což naznačuje, že velká část již zmíněných porodností je dána spíše stěhováním než bydlením samotným. [37]

### 1.2.5 Drogy

### 1.2.5.1 Konopí

Konopí ovlivňuje počet spermií. Na vzorku 1 215 zdravých mužů studie odhalila spojení mezi kouřením konopí a sníženou tvorbou spermií o 29 %, za předpokladu, že muž kouří jednou týdně po dobu tří měsíců. Motilita a morfologie nebyla ovlivněna. Na vzorku 201 žen byla ovulace opožděna v průměru o 1,7 až 3,5 dne u 29 žen, které kouřily konopí alespoň jednou za poslední 3 měsíce ve srovnání s kontrolním vzorkem. 43 % potvrzených anovulačních cyklů se vyskytlo u kuřáku konopí, ačkoliv reprezentovaly pouze 15 % vzorku. Americká data z celostátního průzkumu naznačují, že kouření konopí významně neovlivňuje dobu otěhotnění, ale pro páry, které mají s početím problém, mohou změny v ovulačním cyklu a počtu spermií spojené s kouřením konopí přispívat k jejich potížím s otěhotněním. [38]

### 1.2.5.2 Alkohol a cigarety

Studie objevila, že pro ženy, které nikdy nekouřily, byla doba oplodnění 4,3 měsíců. Pro ženy, které kouřily v minulosti, byla doba oplodnění 4,6 měsíců a 5,1 měsíců pro ženy, které kouří. Navíc k tomu jsou kuřáci vystaveni vyššímu riziku neplodnosti.

Užití alkoholu neovlivnilo průměrnou délku početí.[39]

### 1.2.5.3 Kokain, metamfetaminy a opiáty

Konopí, kokain, metamfetaminy a opiáty negativně ovlivňují mužskou plodnost a byly hlášeny nežádoucí účinky na HPG osu, funkci spermií a strukturu varlat. [40]

## 1.2.6 Vysoké vzdělání

Výsledky studií byly již zmíněny u některých států. Jmenovitě České republice [12] [13], Francii [17], Německu [18] a Spojených státech amerických. [27]

Všechny se shodly, že vyšší vzdělání u žen potenciálně zvyšuje věk prvního početí nebo jej eliminuje.

## 1.2.7 Pracovní zátěž

Existuje studie zaměřující se na porodnost mezi pracovníky v IT sektoru v Indii. Práce v IT sektoru v Indii je dlouhodobě jedna z nejžádanějších hlavně díky vysokým platům. S těmito pracemi ale přichází i mnoho negativních faktorů jako vysoká pracovní zátěž, nebo strach ze ztráty zaměstnání kvůli vysoké konkurenci na trhu práce. Bylo zjištěno, že většina IT specialistů našla svého partnera také v IT. Mezi partnery pak vznikají partnerské problémy kvůli vysokým pracovním nárokům jako jsou přísné denní limity a stresové situace. Ty mohou poté přerůst i do jejich sexuálního života. Dalším faktorem může být i prevalence používání notebooků (laptopů), které jsou využívány doslova na jejich klíně, což může negativně ovlivnit počty spermií. [41]

Článek zaměřující se na dánské páry (n=399), které se snažily otěhotnět poprvé, je sledoval po dobu šesti menstruačních cyklů. Všem mužům byly odebrány vzorky spermatu a u obou vzorek krve. Pracovní zátěž byla zjišťována prostřednictvím úvodního dotazníku. Poté při každém odběru nahlašovali změny v jejich pracovní zátěži během posledních 30 dnů. V analýze nebyly nalezeny žádné souvislosti mezi vlastnostmi spermií nebo sexuálními hormony a zvýšenou psychickou pracovní zátěží. Pravděpodobnost otěhotnění není spojováno s psychickou pracovní zátěží. [42]

### 1.2.8 Chudoba a bohatství

Studie zabývající se daty z El Salvadoru, Guatemaly, Uruguaye, Mexika a Kostariky se zaměřila na to, zda porodnost ovlivňuje reálný příjem na hlavu nebo naopak. Jejich výzkum přišel na to, že existuje kauzální propojení mezi zvýšením reálného příjmu na hlavu a zvýšením porodnosti v El Salvadoru, Guatemale a Uruguayi. Na druhou stranu nenašli žádný důkaz pro hypotézu, že porodnost ovlivňuje reálný příjem na hlavu. Příkládají vinu primárně nedostatečnému vzorku dat reálného příjmu na hlavu ve zkoumaných zemích. [43]

### 1.2.9 Vliv zábavního obsahu

Již byl zmíněn případ brazilských telenovel, které snižovaly porodnost v regionech, kde kanál Rede Globo vysílal. Efekt těchto telenovel dopadal hlavně na ženy, které byly již ve střední či pozdní fázi svého plodného věku a tudíž výrazně neovlivňoval prvorodičky. [10]

Televizní stanice MTV od roku 2009 vysílá pořad s názvem 16 a těhotná, který se zaobírá těhotnými náctiletými dívkami a následky těhotenství na jejich život. O tomto pořadu napsali článek Melissa S. Kearney a Phillip B. Levine, kteří se zaměřili na změny porodnosti u náctiletých dívek a reakce na internetu a sociálních sítí pomocí služeb Google Trends a Twitter. Výsledky ukazují na snížení porodnosti dospívajících o 4,3 procenta. Data z Google Trends a Twitteru zase ukázala zvýšený zájem o antikoncepci a potraty. [44]

### 1.2.10 Sociální mobilita

Sociální mobilita je pohyb jednotlivců, rodin nebo skupin napříč ekonomickými vrstvami. [45] Mezigenerační sociální mobilita hrála v minulosti významnou roli při pokusech o vysvětlení vzorů chování ohledně plodnosti a pokračuje v ní dodnes. Názory vědců zabývajících se společenskými vědami z první poloviny 20. století tvrdí že. Vnitrogenerační sociální mobilita byla posílena osobními aspiracemi na blahobyt a pracovní kariérou žen. Úzkost rodičů v závislosti na jejich sociální statut snižuje porodnost na velkém území rozvinutého světa. Hlavně na území jižní Evropy a východní Asie. [46]

### 1.2.11 Náboženství

Článek Kevina McQuilana [47] se zaměřil na shrnutí dosavadních prací o vztahu mezi náboženstvím a porodností. Se zvýšenou pozorností se zaměřil na případy, kdy náboženství bylo identifikováno jako důležitý faktor vzoru chování k porodnosti. McQuilane dochází k závěru, že náboženství může hrát vlivnou roli na porodnost jsou-li splněny tři podmínky:

1. náboženství formuluje vzory chování s ohledem na vzory chování v oblasti porodnictví
2. náboženství má prostředky ke sdělování těchto hodnot a prostředky k vymáhání náboženských hodnot
3. náboženství tvoří hlavní komponentu sociální identity jejich stoupců

### 1.2.12 Zdravotnictví

Studie zaměřující se na souvislosti mezi vládními výdaji na zdravotní péči a úmrtností matek mezi lety 1981 až 2010 se zabývala 24 zeměmi Evropské unie. Zjistili, že každé roční snížení vládních výdajů na zdravotní péči o 1 % zvýší úmrtnost o 89 matek, což je meziroční nárůst o 10,6 %. Nicméně po započítání jen těch porodů, kterých se účastnil kvalifikovaný personál, se význam mezi vládními výdaji a úmrtností matek prakticky vytratil.

Snížení vládních výdajů na zdravotní péči je spojeno se zvýšenou úmrtností matek, k níž může dojít prostřednictvím změn v poskytování zdravotních služeb kvalifikovaných zdravotníků. Úsporná opatření a snížení rozpočtu pravděpodobně zhorší úmrtnost matek. [48]

### 1.2.13 Steroidy

Anabolické steroidy jsou často zneužívány atleti pro vybudování svalové hmoty. Článek zabývající se zneužitím anabolických steroidů a jejich vliv na mužskou plodnost tvrdí, že steroidy mohou přinést spoustu zdravotních problémů, ale často záleží hlavně na dávkování, dietě a životnímu stylu, který může přispět k některým zdravotním problémům. Autoři shledali, že nízká plodnost je jeden z nejméně ohlášených zdravotních problémů v souvislosti s plodností. [49]

## 1.3 Google Trends a Google Ngram

### 1.3.1 Google Trends

Google Trends je analytický nástroj zabývající se vyhledáváním na Googlu. Tyto požadavky jsou anonymizovány - nelze k jednotlivému vyhledávání přiřadit konkrétního člověka. Data jsou shromažďována od roku 2004. V Googlu pak data normalizují, například pro každé klíčové slovo dělí počet jeho vyhledávání počtem celkových vyhledávání ve zkoumané oblasti. Tuto hodnotu pak přeškálují na číslo v intervalu 0 až 100. To má za následek fakt, že pokud bude ve dvou oblastech kvantitativně stejné vyhledávání, může být celkový počet vyhledávání všech pojmů jiný a proto nemusí hodnoty Google Trends být stejné. [50]

### 1.3.2 Google Ngram a národní korpusy

Google Ngram je analytický nástroj zabývající se tím, jak často se fráze vyskytovaly v korpusech. Název Ngram znamená, jak velká fráze je. Například 2-gram znamená, že je možné vyhledávat v dvouslovných frázích. Výsledek je udáván v procentech, ty udávají, jak velký podíl měla daná fráze v literatuře, která je obsažena v daném korpusu. [51] Korpus je sbírka textů, obvykle uložená v elektronické podobě za účelem výzkumu jazyka. [52]

## 1.4 Analýza existujících řešení

### 1.4.1 Index budoucí orientace

V roce 2012 představila skupina vědců Index budoucí orientace (future orientation index), který se zabývá poměrem internetového vyhledávání frází spojených s budoucností ku internetovému vyhledávání frází spojených s minulostí. Analyzovala data Googlu a našli pozoruhodnou korelaci s HDP zkoumané země a predispozicí obyvatel těšit se na budoucnost. [53]

### 1.4.2 Využívání Google Trends s plodností

Článek několika vědců se zabývá daty z Google Correlate a Google Trends ze Spojených států s daty z průzkumu z American Community Survey, aby získali nové poznatky o plodnosti a jejím významu. Přišli na to, že data Googlu jdou použít k ilustrování socioekonomických rozdílů v oblasti těhotenství a porodu. Například fráze létání během těhotenství jsou spojeny s těhotenstvím žen z vyšších sociálních vrstev a test otcovství je spojen s plodností nemanželských párů. [54]

## 1.5 Analýza uživatelských požadavků

### 1.5.1 Funkční požadavky

- F1 Uživatel si bude moci vybrat klíčové heslo napsáním do příslušného řádku
- F2 Uživatel si bude moci vybrat klíčové heslo vyhledáním v seznamu
- F3 Uživatel si bude moci vybrat, zda chce nabídku klíčových slov z Google Trends nebo z Google Ngram
- F4 Uživatel si bude moci vybrat interval v jakém se nachází korelace klíčových slov s TFR
- F5 Uživatel si bude moci vybrat jazyk klíčového slova
- F6 Po výběru zadaného klíčového slova se zobrazí graf, ukazující křivku hledání klíčového slova a TFR v daném státě společně se statistickými údaji, lineární regresí a kvadratickou nelineární regresí
- F7 Uživatel si bude moci vybrat z kategorií napsáním žádané kategorie do příslušného řádku
- F8 Uživatel si bude moci vybrat z kategorií vyhledáním v příslušném seznamu
- F9 Uživatel se bude moci podívat na výtažek zajímavavých veličin
- F10 Po výběru kategorie se zobrazí graf ukazující křivku zkombinovaných klíčových slov v kategorii a TFR v daném státě společně se statistickými údaji, velikostí vzorku, lineární regresí a kvadratickou regresí

### 1.5.2 Nefunkční požadavky

Nefunkční požadavky neboli obecné požadavky jsou takové, které určují omezení kladená na systém.

- N1 Dostupnost přes webové rozhraní
- N2 Jednoduché ovládání
- N3 Aplikace bude v anglickém jazyce
- N4 Aplikace se bude přizpůsobovat velikosti obrazovky

## 1.6 Případy užití

- UC1 Uživatel si chce zobrazit statistiky konkrétního klíčového slova v konkrétním státu z dat Google Trends
- UC2 Uživatel si chce zobrazit statistiky nespecifikovaného klíčového slova v konkrétním státu z dat Google Ngram
- UC3 Uživatel si chce zobrazit nespecifikované klíčové slovo ze seznamu v konkrétním státu
- UC4 Uživatel si chce zobrazit klíčová slova v konkrétním státu, které mají Pearsonovu korelaci ve vybraném intervalu
- UC5 Uživatel si chce zobrazit konkrétní kategorii klíčových slov v konkrétním státu
- UC6 Uživatel se chce podívat na zajímavé výsledky a data

	UC1	UC2	UC3	UC4	UC5	UC6
F1	+					
F2	+		+	+		
F3	+	+	+	+		
F4		+	+	+		
F5	+	+	+			
F6	+	+	+			
F7					+	
F8					+	
F9						+
F10					+	

■ **Tabulka 1.1** Pokrytí funkčních požadavků

### 1.6.1 Pokrytí funkčních požadavků

Vytvořil jsem tabulku, která ukazuje pokrytí případů užití funkčními požadavky.

## 1.7 Analýza technologií

Můj projekt je zaměřen na datovou analýzu. Na tento obor se zaměřuje mnoho programovacích jazyků jako například Python, JavaScript, Scala nebo R. Každý z těchto jazyků má své výhody i nevýhody. Pro Python je velká výhoda vysoká popularita mezi komunitou datových vědců, která znamená velkou podporu a nesčetné množství naučných materiálů pro uživatele. Popularitu nabyl i díky široké škále nástrojů a knihoven pro strojové učení a vizualizaci. Největší nevýhodou, na kterou si uživatelé stěžují, je jeho rychlost. Python dominuje i na žebříčcích popularity jako TIOBE Index nebo PYPL Index. Na výběru Pythonu jako nejlepšího programovacího jazyka pro datové vědy se shodne mnoho zdrojů. [55] [56] [57]

### 1.7.1 Python

Python je objektově orientovaný, interpretovaný a interaktivní jazyk. [58] Kód jazyka Python je efektivní, často vyžaduje kratší zápis kódu než alternativní programovací jazyky a díky Python syntaxi má čitelný kód. Lidé využívají Python pro mnoho činností od tvorby her k budování webových aplikací nebo k vývoji interních nástrojů. Python je také velmi využíván na poli vědy pro akademický výzkum. [59]

#### 1.7.1.1 knihovna Streamlit

Streamlit je open-source knihovna pro Python, usnadňující vytváření webových aplikací pro účely strojového učení a datových věd. Streamlit využívá jazyky jako Python, TypeScript nebo JavaScript. Za pomoci chytrých widgetů lze sestavit aplikace pro vizualizace dat. [60]

## 1.8 Statistické nástroje

### 1.8.1 Pearsonův korelační koeficient

Pearsonův korelační koeficient je typ korelačního koeficientu, který představuje vztah mezi dvěma proměnnými, které jsou měřeny na stejném intervalu. Pearsonův koeficient je měřítkem síly



asociace mezi dvěma spojitými proměnnými. Pearsonova korelace měří pouze lineární vztahy. [61] [62] [63] Pearsonova korelace je určena vzorcem

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y}$$

kde:

$\rho$  = Pearsonova korelace

$\text{cov}(X, Y)$  = kovariance proměnných X a Y

$\sigma_x$  = směrodatná odchylka x

$\sigma_y$  = směrodatná odchylka y

## 1.8.2 Spearmanův korelační koeficient

Spearmanův korelační koeficient je typ korelačního koeficientu, který představuje vztah mezi dvěma proměnnými, které jsou měřeny na stejném intervalu. Spearmanova pořadová korelace představuje alternativu, která zmírňuje působení odlehých hodnot a sešikmených rozdělení. [61] Spearmanova korelace je určena vzorcem

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

## 1.8.3 p-hodnota

P-hodnota poskytuje vhodný základ pro vyvozování závěru při testování hypotéz. Čím nižší p-hodnota je, tím větší je statistická významnost pozorovaných veličin. Za statisticky významnou hodnotu se považuje p-hodnota menší nebo rovna 0,05. [64]

## 1.8.4 Kovariance

Kovariance vyjadřuje míru tendence dvou proměnných ke společné variabilitě. [61]

$$\text{cov}_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

## 1.8.5 Lineární regrese

Lineární regrese je proložení dat přímkou za pomoci metody nejmenších čtverců. [61]

## 1.8.6 Kvadratická regrese

Kvadratická regrese je proložení dat polynomem druhého řádu, většinou také za pomoci metody nejmenších čtverců. [65]



## Kapitola 2

# Návrh

*V této kapitole rozberu další používané technologie, návrh zpracování dat, design a rozložení stran.*

### 2.1 Další technologie

Při tvoreni aplikace je potreba si ujasnit par veci, a to hlavne pokud vim, ze budu vyuzivat Python s knihovnou Streamlit, jak verzovat svou aplikaci a jak ukladat data.

#### 2.1.1 Git

Git je distribuovany system spravy verzí, který je bezplatny a ma otevreny zdrojovy kod. Dle jejich webové stránky je určen pro malé i velké projekty. [66] U porovnání podobných softwarů končí na vrchních příčkách. [67] [68] V doprovodu s velkou dokumentací se mi git jeví jako správná volba pro tuto bakalářskou práci.

#### 2.1.2 Ukládání dat

Pro ukládání dat aplikace se dá využít mnoho způsobů ukládání dat. Ať už využití databází jako je SQL či MySQL nebo data uložit lokálně například přes soubory CSV. Vzhledem k velikosti projektu a potřeby relativně málo dat, se mi zdá ukládání souboru lokálně jako lepší řešení.

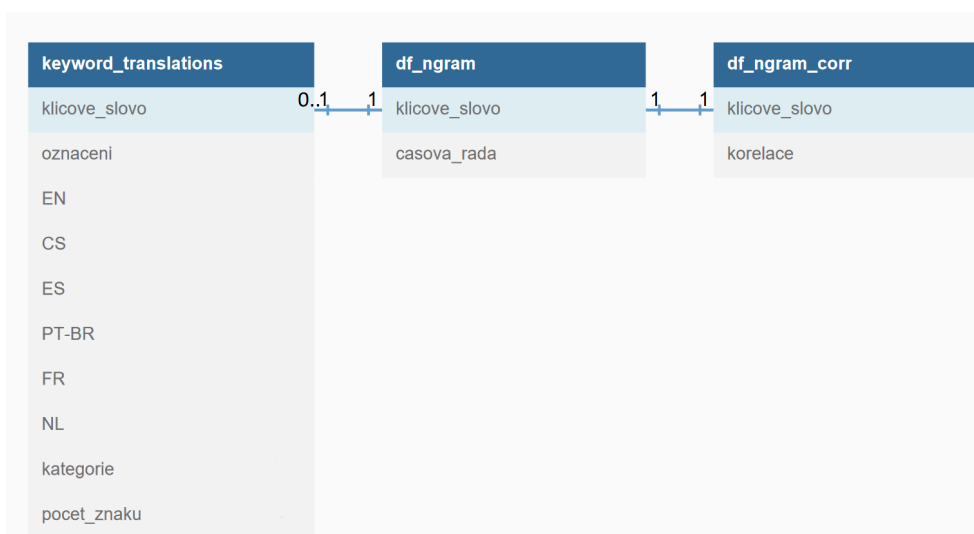
### 2.2 Data

#### 2.2.1 Získání základních klíčových slov

Využiji znalosti z analytické části práce a k nim vypracuji dostatečný dataset, který bude obsahovat klíčová slova z témat jako jsou drogy, steroidy, hledání zaměstnání, univerzity nebo zdravotnictví.

#### 2.2.2 Překlad klíčových pojmů

Aby se dataset dal využít i v jiných jazycích použiji překladač pro převedení klíčových slov a frází i do ostatních jazyků, přesněji do češtiny, brazilské portugalštiny, němčiny, francouzštiny, nizozemštiny a španělštiny.



■ **Obrazek 2.1** Diagram rozložení dat pro data z Google Ngram

### 2.2.2.1 Google Translate

Google Translate je překládací nástroj od Googlu za využití neuronových sítí. Umí automaticky přeložit text do více než sto jazyků. Nabízí rozhraní pro automatizované překládání větších textů. Všechny požadované jazyky podporuje.[69]

### 2.2.2.2 DeepL

DeepL je překladač od společnosti DeepL SE. Také využívá umělé neuronové sítě k překladač. Aktuálně podporuje 29 jazyků, všechny požadované jazyky podporuje. Také nabízí rozhraní pro automatické překládání textů.[70]

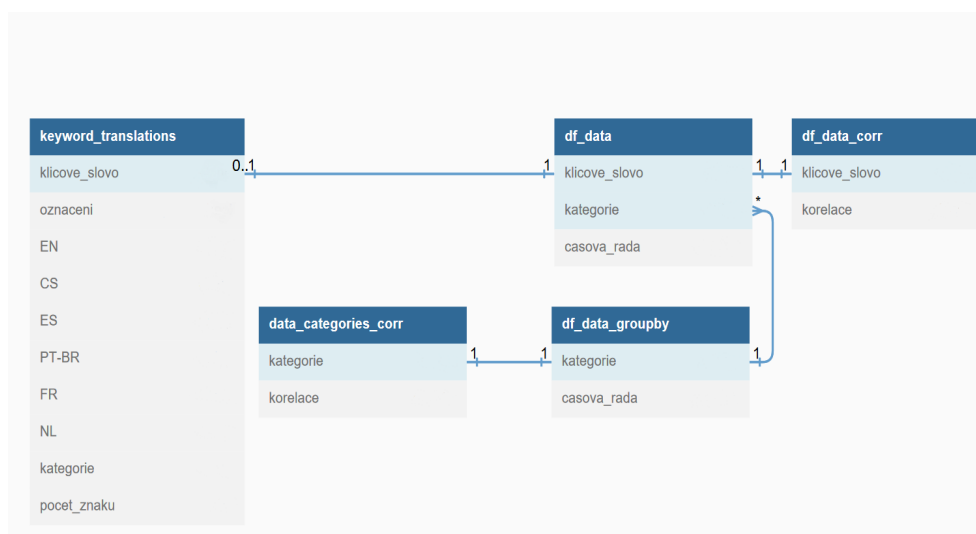
### 2.2.2.3 Porovnání

Dle porovnání kvality překladač, DeepL vítězí. Působí přirozeněji a nejvíce vyniká právě při překladač mezi evropskými jazyky. Prvních 500 000 znaků překladač je pro obě služby zdarma. Nad tento rámec vychází DeepL draž, ale 500 000 znaků by mělo být více než dostatečné množství při překladač klíčových slov.[71]

## 2.2.3 Zpracování a struktura dat

Pro každý jazyk budou shromážděna data tak, že po řádcích budou uložena klíčová slova. Ve sloupcích se budou nacházet jednotlivé roky od 2004 do 2021. Z těchto dat, poté bude vypočtena korelace mezi daty a úhrnnou plodností daného státu. Ze základních dat poté budou vypočteny data kategorií za využití průměru všech veličin v dané kategorii. Vytvořil jsem vizualizaci pro představu jak data budou vypadat. 2.1 2.2

V tabulce keyword\_translations se nachází několik položek, první je dané klíčové slovo a označení, které určuje, zda je nutno dané klíčové slovo překládat, poté následují kódy pro označení jazyků, které využívám. Dále se zde nachází kategorie klíčového slova a počet znaků klíčového slova. V tabulce df\_ngram se nachází dvě položky, klíčové slovo a časová řada, která obsahuje data získaná ze stránky Google Ngram. K této tabulce také patří i tabulka s korelacemi záznamů z Ngramu. Pro data z Google Trends existuje tabulka df\_data, která obsahuje klíčové slovo a časovou řadu, obsahující právě získaná data z Google Trends. K této tabulce patří i tabulka



■ **Obrázek 2.2** Diagram rozložení dat pro data z Google Trends

s korelacemi. Další tabulka se zaměřuje na kategorie klíčových slov u dat z Google Trends, ta pouze shlukuje data z tabulky `df_data` pro jednotlivé kategorie, pro tuto tabulku poté existuje i tabulka s korelacemi jednotlivých kategorií.

## 2.3 Aplikace

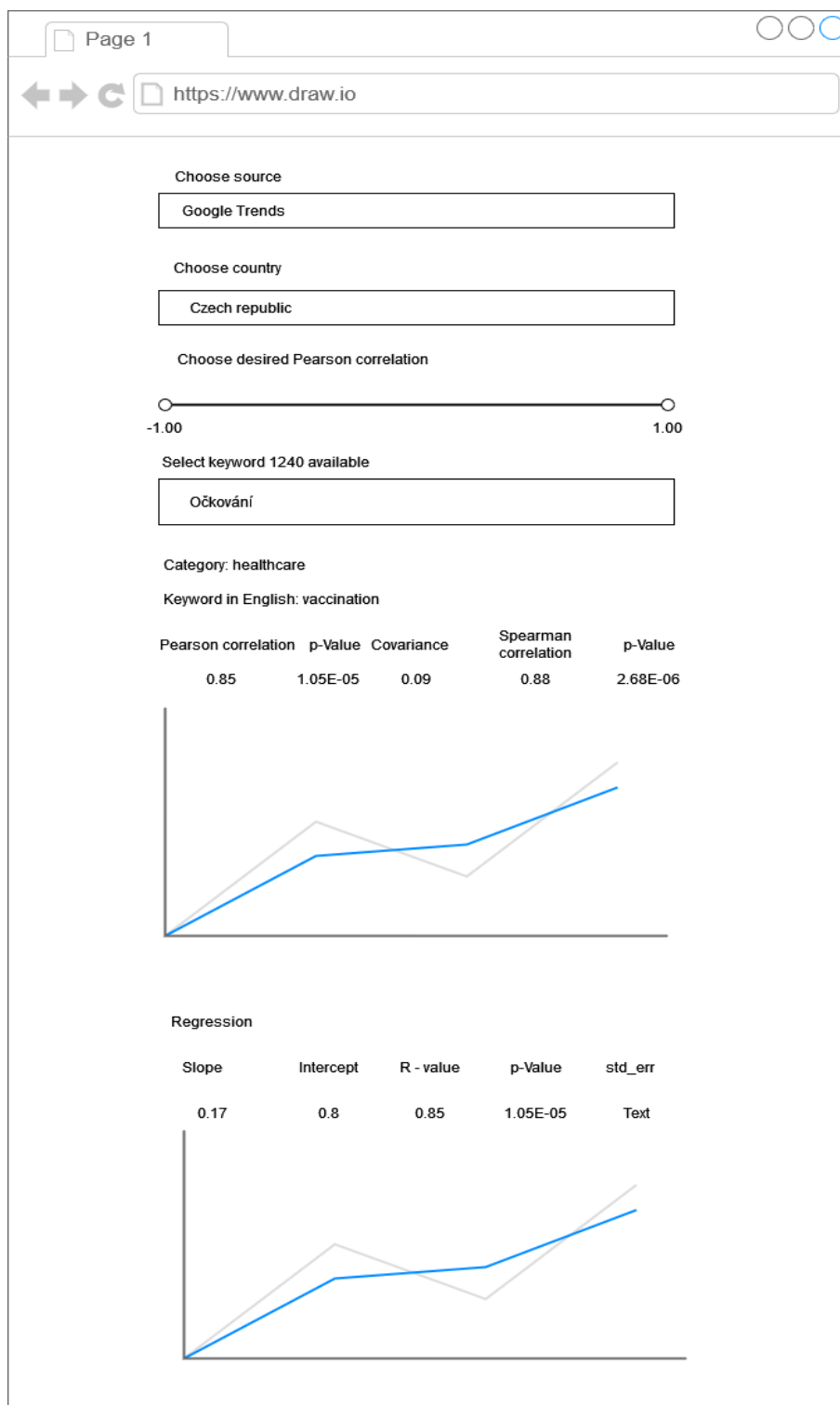
Aplikace bude formou webové stránky, která bude složena ze tří stránek. Jedna se bude zaměřovat na jednotlivá klíčová slova, druhá se zaměří na statistiky kategorií a třetí bude obsahovat podstránky se veličinami, které obsahují klíčová slova a kategorie, které mohou být pro uživatele zajímavé.

### 2.3.1 Jednotlivé veličiny

Na této stránce by měla být možnost si vybrat zdroj dat (Google Trends nebo Google Ngram) a místo sběru informací (Brazílie, Česká republika, Francie, Německo, Nizozemsko, Spojené státy americké nebo Španělsko). K výběru požadované korelace klíčového slova bude k dispozici dvoubodový posuvník. Pod posuvníkem se bude nacházet rozbalovací lišta, která uživateli dovolí si vybrat konkrétní klíčové slovo. Lišta bude obsahovat jen ta klíčová slova, která svými statistikami budou odpovídat korelaci dané na posuvníku. Pro představu jsem vytvořil vizualizaci návrhu.

### 2.3.2 Kategorie

Tato stránka by měla vypadat přibližně stejně jako stránka s jednotlivými veličinami 2.3 jen bez posuvníku, protože kategorií je výrazně méně než klíčových slov díky tomu nebude mít uživatel potíže se na všechna data podívat. Navíc vůči stránce s jednotlivými veličinami bude informace o velikosti vzorku dané kategorie.



■ Obrázek 2.3 Vizualizace stránky pro jednotlivé veličiny



■ **Obrázek 2.4** Vizualizace stránky pro zajímavé objevy

### 2.3.3 Zajímavé objevy

Na této straně budou uloženy zajímavé objevy z práce. Stránka bude rozložena do několika podstránek, do kterých se bude moci uživatel dostat přes horní menu. Pro představu je vytvořena vizualizace. 2.4





# Implementace

*V této kapitole se budu zabývat implementací webové aplikace, získáním dat a zpracováním dat.*

### 3.1 Získání dat

V této části se budu zabývat vytvořením a přeložením klíčových slov. Poté se zaměřím na získání dat z Google Trends a Google Ngram.

#### 3.1.1 Vytvoření datasetu

Za využití znalostí z analytické části jsem vytvořil množinu o velikosti 2 000 klíčových slov. Ty jsem postupně rozdělil do 32 kategorií. Ke každému klíčovému slovu jsem přiřadil značku, zda slovo překládat. Například u kategorie S&P 500 zabývající se názvy pětistý největšími společnostmi ve Spojených státech není potřeba překlad.

#### 3.1.2 Překlad datasetu

Jak jsem již v návrhu zmiňoval, je potřeba přeložit 2000 klíčových pojmů do 6 jazyků. Na to jsem zvolil na základě zjištění nástroj DeepL. Tento nástroj disponuje rozhraním pro Python, díky kódu níže jsem poté přeložil dataset do požadovaných jazyků. Kód detekuje, zda má dané klíčové slovo značku, která určí zda slovo překládat. Pokud je slovo označeno, že je určeno k překladu, je slovo předáno rozhraní DeepL, které si převezme klíčové slovo a jazyk, do kterého chci termín přeložit. 3.1

#### ■ Výpis kódu 3.1 Kód k překladu z DeepL

```
for index, row in df.iterrows():
    if row['Flag'] == 0:
        for i in range(len(languages)):
            translations[i].append(row['Keyword'])
    elif row['Flag'] == 1:
        for i in range(len(languages)):
            translations[i].append(translator.translate_text(row['Keyword'],
                                                             target_lang=languages[i]))
```

### 3.1.3 Google Trends

Data na stránce <https://trends.google.com/> jdou stáhnout poté, co si vyberu klíčové slovo a zadám ho do vyhledávacího řádku. Stisknutím tlačítka signalizující stažení stáhnou CSV soubor obsahující data obsahující výhledávání zadané fráze. Tato data jsou řazena po týdnech od roku 2004 do současnosti.

### 3.1.4 Google Ngram

Pro získání dat z Google Ngram využiji možnosti získat data ze stránky ve formátu JSON. Toho docílím tak, že v adresním řádku webového prohlížeče zaměním slovo graph za slovo json. Takže změna vypadá takto: z webové adresy

```
https://books.google.com/ngrams/graph?content=fertility&year_start=1800&year_end=2019&corpus=26&smoothing=3
```

 se stane webová adresa

```
https://books.google.com/ngrams/json?content=fertility&year_start=1800&year_end=2019&corpus=26&smoothing=3.
```

 Data v tomto formátu mohou poté sloučit manuálně. Tímto krokem získám soubor ve formátu JSON a data jsou připravena ke zpracování.

### 3.1.5 Problémy s automatizací

Vzhledem k podmínkám užívání Google služeb je nemožné proces získávání dat automatizovat. Pokud by to bylo však legálně možné, dala by se využít pro Google Trends knihovna `pytrends` nebo specializované těžiče dat. V případě `pytrends` ale uživatelé mohou narazit na několik problémů spojených s aktivním potlačováním těžení těchto dat. Jedním z nich je odhalování botů na základě hlaviček cookie. Tyto hlavičky se dokáží měnit na základě javascriptového kódu běžícího na stránce Google Trends. Tyto specifické hlavičky lze překonat zkopírováním cURL GET metody požadavku, poté zkonvertovat cURL příkaz na příslušný programovací jazyk.

Ovšem to není jediný problém, který může nastat, Google detekuje boty i na základě opakujících se požadavků z jedné IP adresy. Tato potíž by se dala obejít díky rotating proxy. To je metoda, kdy se při každém připojení uživatel připojí k serveru z jiné IP adresy náležející nějakému proxy serveru.

Tyto metody jsou ale v rozporu s podmínkami užití Google služeb, a proto jsem je nepoužil a data z Google Trends získal legálně.

## 3.2 Zpracování dat

### 3.2.1 Spojení dat

Data jsem spojil kódem, který si zpracuje data ze souboru CSV, která jsem získal ze stránky Google Trends. Nejdříve byla potřeba získat klíčové slovo z původního souboru, poté zjistit kategorii daného klíčového slova, kterou jsem zjistil pomocí souboru zaobírající se překladem dat, kde byla mimo jiné uvede právě hledaná kategorie. Poté jsem měsíční data zprůměroval tak, abych měl data pro jednotlivé roky od 2004 do 2021. Pro ukázkou jsem přiložil kód 3.2.

### 3.2.2 Korelace pro vyhledávání

Pro jednodušší vyhledávání klíčových slov s vysokou korelací jsem si musel připravit korelace jednotlivých klíčových slov. K výpočtu jsem využil původní soubor s daty z vyhledávání či frekvencemi v literatuře, do kterého jsem přidal i sloupec s hodnotami plodnosti. Poté jsem na celou tabulku zavolał funkci `corr()`, která vypočítala korelace všech sloupců se všemi řádky. Tento

**■ Výpis kódu 3.2** Spojení a komprese dat

```
for f in files:
    data = pd.read_csv(f)
    keyword = data.iloc[0].values[0].split(":")[0]
    category = df.loc[df['EN'] == keyword].Category.values[0]
    data = data.iloc[1:, :]
    data.rename(columns = {'Kategorie: Vsechny kategorie': 'Values'},
                inplace = True)
    data['Values'] = data['Values'].astype(int)
    row = [keyword]
    for i in range(2004, 2022):
        row.append(
            mean(data[ str(i) + "-01": str(i) + "-12"].Values.values))
    row.append(category)
    df_data_us.loc[len(df_data_us.index)] = row
```

**■ Výpis kódu 3.3** Vypočítání korelace dat

```
for country in countries_data:
    df = pd.read_csv("df_data_" + country + ".csv")
    df.drop(columns=['Category'])
    df1 = pd.DataFrame(columns=df['keyword'].values)
    for i in range(2004, 2021):
        df1.loc[len(df1.index)] = df[str(i)].values
    df1['fertility'] = fertility[countries_data.index(country)]
    df2 = df1.corr()['fertility']
    df2.to_csv('df_data_' + country + '_corr.csv',
              encoding='utf-8', index=False)
```

výpočet jsem poté omezil na korelace, které měly vztah s plodností. Ty jsem si uložil do dalšího souboru. Pro ukázkou přikládám kód 3.3.

### 3.2.3 Seskupení do kategorií

Abych mohl prezentovat i data kategorií, bylo zapotřebí data sloučit. K tomuto jsem využil funkci `groupby`, která za pomoci parametru `by` určí, podle čeho má data sloučit. Já zvolil sloučení dat vyhledávání za pomoci kategorie. Tyto data jsem ještě zprůměroval, aby data byla normalizovaná jako v původních datech od Google Trends. Pro ukázkou přikládám kód 3.4.

**■ Výpis kódu 3.4** Shlukování dat

```
for key in langs:
    df_lang = pd.read_csv("df_data_" + key + ".csv")

    df_lang = df_lang.groupby(by = "Category").mean()

    df_lang.to_csv("df_data_groupby_" + key + ".csv",
                  encoding='utf-8', index = False)
```

### 3.3 Vytvoření aplikace

Tvoření aplikace díky Streamlitu je velmi přímočaré a pohodlné, aplikace se staví za pomoci prvků, které mi umožnily vytvořit celou aplikaci. Využil jsem několik prvků pro uživatelský vstup a to rozbalovací seznam a dvoubodový posuvník. Tyto vstupy jsou již ošetřeny a snižují riziko problémů v dalších částech aplikace. Pro výstupy jsem využil hlavně objekt, na zobrazení dat v několika sloupcích a prvek umožňující zobrazení grafů.

#### 3.3.1 Zobrazení statistických veličin

K zobrazení statistických veličin jsem využil prvek s názvem `columns`, který normálně uživatelé využívají pro zobrazení dat jako je teplota, aktuální hodnota akcie nebo čas. Já jsem do jednotlivých sloupců zakomponoval pět statistických veličin a to Pearsonovu korelaci a její p-hodnotu, kovarianci, Spearmanovu korelaci a její p-hodnotu. Pro lepší přehlednost jsem modifikoval výstup p-hodnot, tak ať se zobrazují ve vědeckém zápisu čísel, protože nastávala situace, kdy zaokrouhlení na 5 desetinných míst pořád v aplikaci ukazovalo hodnotu 0, což pro uživatele nemá žádnou vypovídající hodnotu. Příkládám i kus kódu pro představu, jak se s `columns` pracuje 3.5 .

#### 3.3.2 Dvouosý graf pro veličiny a TFR

Pro vizualizaci dat úhrnné plodnosti a dat z hledání na internetu nebo četnosti v literatuře byla potřeba vzhledem k různorodosti dat (data úhrnné plodnosti bývají v nižších jednotkách dětí na ženu, data z hledání na internetu jsou z intervalu 0 až 100 a data z četnosti v literatuře jsou z intervalu 0 až 1) využít dvou os zobrazení. Toho jsem docílil díky grafům z knihovny `Matplotlib`, které podporuje i `Streamlit`. Pro ukázkou příkládám kód. 3.6

#### 3.3.3 Graf na zobrazení regresí

K zobrazení chování dat vůči sobě jsem využil grafu, který zobrazí na ose x data klíčových slov a na ose y plodnost ve vybraném státu. Tento graf jsem také využil k zobrazení regresí a to lineární, pro kterou jsou její parametry zobrazeny v sloupcích nad grafem a kvadratickou, jejíž předpis je zobrazen v legendě grafu. Tento graf může uživateli více přiblížit vztah dat. Například pokud si zvolíme klíčové slovo `IKEA` v datech z České republiky můžeme vidět 3.1 několik zajímavostí, první je, že data krásně oscilují okolo obou regresí a také, že čím více se na českém internetu vyhledává klíčové slovo `IKEA`, tak s tím úměrně roste i úhrnná plodnost. Příkládám i kus kódu pro zobrazení celého grafu. 3.7

#### ■ Výpis kódu 3.5 Kód pro sloupcové widgety

```
col1, col2, col3, col4, col5 = st.columns(5)
pearson = stats.pearsonr(df_stats["Data"].values,
df_stats["FTR"].values)
col1.metric("Pearson correlation", round(pearson[0], 4))
col2.metric("p-Value", '%.2E' % pearson[1])
col3.metric("Covariance",
round(df_stats.cov()["Data"].values[1], 4))
spearman = stats.spearmanr(
df_stats["Data"].values, df_stats["FTR"].values)
col4.metric("Spearman correlation", round(spearman[0], 4))
col5.metric("p-Value", '%.2E' % spearman[1])
```

**■ Výpis kódu 3.6** Kód dvousého grafu pro zobrazení TFR a dat

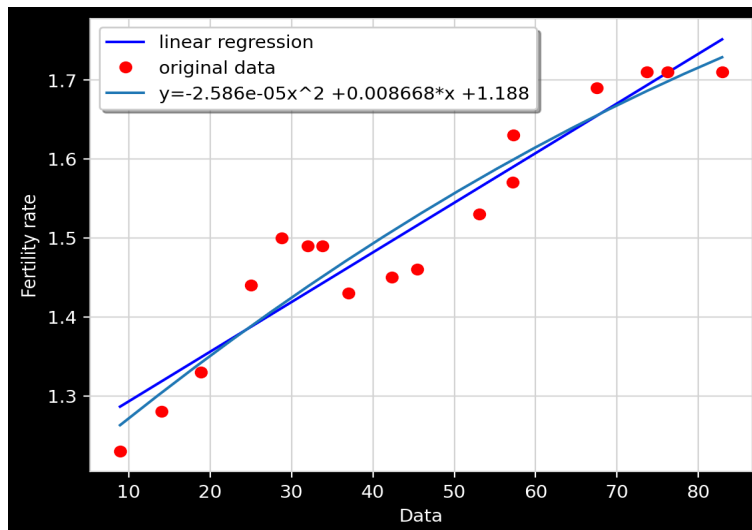
```
time = np.arange(2004, 2021)
ftr = df_stats["FTR"].values
key_data = df_stats["Data"].values

fig = plt.figure()
ax = fig.add_subplot(111)

lns1 = ax.plot(time, ftr, '-', label='TFR in ' + country)
ax2 = ax.twinx()
lns2 = ax2.plot(time, key_data, '-r', label=keyword)

lns = lns1 + lns2
labs = [l.get_label() for l in lns]
ax.legend(lns, labs, loc=0)

ax.grid()
ax.set_xlabel("Years")
ax.set_ylabel(r"Fertility rate")
ax2.set_ylabel(r"Data")
st.pyplot(fig)
```

**■ Obrázek 3.1** Regrese pro Ikea z dat Google Trends v České republice

**■ Výpis kódu 3.7** Kód pro zobrazení regresí

```
y = ftr
x = key_data

polyline = np.linspace(min(x), max(x), 17)

model = np.poly1d(np.polyfit(x, y, 2))

abline_values = [slope * i + intercept for i in key_data]

ax3.plot(key_data, abline_values, 'b' , label = "linear regression")
ax3.plot(key_data, ftr, 'ro', label = 'original data')
ax3.plot(polyline, model(polyline),
         label="y=%sx^2 %s*x %s" % (fmt_float(model[2]),
         fmt_float1(model[1]), fmt_float1(model[0])))
ax3.legend(["TFR in " + country, str(keyword)])
ax3.legend(loc='best', fancybox=True, shadow=True)
ax3.grid()
ax3.set_xlabel(r"Data")
ax3.set_ylabel(r"Fertility rate")

st.pyplot(fig2)
```

### 3.3.4 Přístup a instalace

Aplikace může být dostupná jak z internetu, v mém případě na stránce <https://janbartos-bp-main-page-e18je7.streamlit.app/>, nebo v případě potřeby i lokálně. Lokálně se instalace provede v prostředí, kde je nainstalován Python a několik knihoven.

K instalaci je potřeba nainstalovat všechny knihovny, které jsou uvedeny v souboru requirements.txt a knihovnu Streamlit. Instalace knihoven se provádí pomocí příkazu `pip install knihovna`, kde se místo slova knihovna napíše název knihovny, která je potřeba nainstalovat. Poté, co jsou všechny knihovny nainstalovány, stačí spustit aplikaci příkazem `streamlit run Main_Page.py`.

*V této kapitole se zaměřím na testování aplikace. Využiji k tomu předem připravené testovací scénáře, které vycházejí z případů užití. Díky robustnosti knihovny Streamlit byla většina chyb související s uživatelským vstupem ošetřena základními prvky stránky - výběrovým seznamem a posuvníkem.*

### 4.1 Testovací scénáře

První testovací scénář se zaměřuje na zobrazení statistik konkrétního klíčového slova z konkrétního státu. Může se stát, že se klíčové slovo uživateli nepodaří zvolit, pokud má posuvníky korelací nastavené tak, že vybrané klíčové slovo nemá ve zvoleném rozmezí danou korelaci. Tato vlastnost funguje tak, jak jsem ji zamýšlel a není to chyba. Chybou se, ale v tomto případě stal výběr Ngramu jako zdroje dat. Data národních korpusů z Google Ngram byly pouze do roku 2019, ačkoliv data plodnosti byly až do roku 2020, což mělo za výsledek různě velká pole, kvůli kterým nešlo vypočítat korelace, zobrazit křivku plodnosti a dat z Googlu či zobrazit regrese. Tuto chybu jsem ošetřil jednoduchou podmínkou, která zmenšuje pole dat na základě potřeby, Google Trends do roku 2020 a Google Ngram do roku 2019.

Druhý a třetí testovací scénář se zaměřuje na zvolení klíčového slova na základě výběru ze seznamu klíčových slov. Zde se může pouze stát, že posuvník je uživatelem nastaven tak, že v daném rozmezí korelací se nanacházejí žádná data a proto se uživateli nezobrazí žádné klíčové slovo k výběru.

Čtvrtý testovací scénář se zaměřuje na zvolení si klíčového slova na základě omezeného výběru díky posuvníku s korelacemi. U posuvníku nastával problém takový, že pokud v daném rozmezí neexistovalo klíčové slovo, celý následující blok aplikace detekoval chybu. Tuto chybu jsem vyřešil jednoduchou podmínkou, která ověřuje, že klíčové slovo, pro které má aplikace zobrazit statistiky, je vybráno, ať už automaticky aplikací nebo uživatelem. 4.1

Pátý testovací scénář využívá stránku zabývající se kategoriemi, na této stránce nemůže vybrat korelaci kategorií kvůli absenci posuvníku, který toto na stránce s klíčovými slovy umožňoval omezit výběr klíčových slov, takže odpadá již zmíněný problém s prázdným klíčovým slovem. Uživatel díky tomu může vybrat jen ta data, která jsou mu nabídnuta v rozbalovacích seznamech.

Šestý testovací scénář využívá stránku se zajímavými daty. Tato stránka je vytvořena tak, že uživatel se může podívat pouze na mnou zvolená data. To znamená, že uživatel nemá na této stránce žádnou možnost výběru, pouze změnu karet v horní části obrazovky.

**■ Výpis kódu 4.1** Kód ukazující opravu chyby

```
keyword = ""
selectBool = False

if len(df_corr[df_corr['fertility'].between(corr[0], corr[1],
inclusive="neither")]) != 0:
    keyword = st.selectbox(
        'Select keyword ' + str(len(df_corr[df_corr['fertility'].
        between(corr[0], corr[1],
        inclusive="neither")])) + " available" ,
        df_corr[df_corr['fertility'].between(corr[0], corr[1],
        inclusive="neither")]
    )
    selectBool = True

if keyword != "":
```



# Zhodnocení veličin

*V této kapitole se zaměřím na výsledky analýzy, jejich statistiky a jejich porovnání s informacemi, která jsem získal v analýze. Dále se zaměřím i na nějaké výrazné pojmy u všech států.*

### 5.1 Google Ngram

Data z Google Ngram byly obtížně dostupné k získání ve větším počtu. Vzhledem k malému vzorku nelze dospět ke statisticky zajímavým výsledkům.

### 5.2 Google Trends

Data z Google Trends jsou řádově početnější než z Google Ngram. Každý jazyk obsahoval 2000 vstupních hodnot. Každý jazyk měl pak ve finále různě velkou velikost dat. Od 1074 záznamů z českého Googlu po 1815 v americkém Googlu. Tyto slova byla rozřazena do různých kategorií. V této kategorii pak byla vypočtena řada z průměru všech hodnot v této kategorii. Tuto řadu jsem následně analyzoval.

#### 5.2.1 Antidepresiva

Tato kategorie obsahovala názvy antidepresiv. Velikost původního vzorku byla 81. Mezi klíčová slova této kategorie patřila názvy antidepresiv jako Xanax, Prozac nebo Zoloft. Nejzajímavější výsledkem jsou bez výhrad data z Francie, se Spearmanovým korelačním koeficientem  $-0,8322$  a  $p$ -hodnotou  $3,42E-05$  na vzorku velikosti 42. Data z Francie 1.4 ukazují, že TFR se propadá s menším nárůstem od poloviny 90. let do desátých let nového tisíciletí. Vzhledem k faktu, že data od Google Trends se začala sbírat od roku 2004, kdy TFR pořád rostlo, takže graf TFR ve Francii má konkávní tvar, který má vrchol okolo roku 2008, což je čas, kdy začala ekonomická krize. Je možné že nárůst hledání antidepresiv je ovlivněn více ekonomickou krizí než TFR.

#### 5.2.2 Auta

Tato kategorie obsahovala názvy aut a klíčová slova spojená s automobily. Například obsahovala ceny paliv, financování automobilů nebo velké automobilové značky v Evropě či Americe. Velikost původního vzorku byla 23 klíčových slov. Absolutně nejzajímavějším výsledkem je výsledek ve Spojených státech amerických, kde jsem získal celý vzorek. Spearmanův korelační koeficient činil  $-0,9920$  s  $p$ -hodnotou  $6,54E-15$ . Ostatní státy ovšem nezůstaly pozadu a některé jako Německo,

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	56	0.1584	5.44E-01	0.0488	-0.2798	2.77E-01
cz	26	-0.7217	1.07E-03	-0.4230	-0.5126	3.54E-02
us	74	-0.6340	6.28E-03	-0.2732	-0.6781	2.77E-03
de	45	0.5900	1.27E-02	0.1184	0.4936	4.41E-02
es	49	-0.4513	6.90E-02	-0.0540	-0.5117	3.58E-02
nl	39	-0.6224	7.62E-03	-0.1068	-0.6605	3.90E-03
fr	42	-0.8313	3.55E-05	-0.1347	-0.8322	3.42E-05

■ **Tabulka 5.1** Statistika antidepresiv

Nizozemsko či Francie měly také statisticky významné výsledky. Z dat USA 1.8, můžeme vidět, že se plodnost snižuje a to i od roku 2004, kdy se data Google Trends zaznamenávají. Což může znamenat, že se například jen výrazně zvýšil zájem o automobilismus ve vyspělém světě a nemá s plodností nic společného.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	21	0.7498	5.28E-04	0.1402	0.5672	1.76E-02
cz	20	0.5488	2.25E-02	0.2400	0.5495	2.23E-02
us	23	-0.9552	2.48E-09	-0.9507	-0.9920	6.54E-15
de	22	0.9124	3.32E-07	0.4595	0.8766	3.90E-06
es	23	-0.5088	3.70E-02	-0.1357	-0.5879	1.31E-02
nl	21	-0.7382	7.16E-04	-0.2259	-0.7159	1.23E-03
fr	22	-0.8367	2.83E-05	-0.2061	-0.8052	9.64E-05

■ **Tabulka 5.2** Statistika aut

### 5.2.3 Charity

Tato kategorie se zaměřila na názvy charit. Obsaženy byly charity jako Unicef, Goodwill nebo i YMCA. Velikost vzorku byla 46 klíčových slov. Nejzajímavější výsledky jsou z Brazílie, která je na obyvatele s výrazným nárůstem nejchudší ze všech ostatních, s Pearsonovým korelačním koeficientem <sup>1</sup> 0,9737 a p-hodnotou 4,84E-11. Z dat můžeme vidět, že TFR dlouhodobě v Brazílii klesá 1.2, to znamená, že klesá i zájem o charity. Dalším zajímavým zjištěním je Spearmanova korelace -0,9109 s p-hodnotou 3,77E-07 v České republice. Při pohledu na data 1.3 TFR v České republice vidíme, že plodnost od nového milénia rostla, můžeme očekávat, že na vině sníženého zájmu o charitativní organizace může být ekonomická stránka.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	25	0.9737	4.84E-11	0.3785	0.9583	1.46E-09
cz	21	-0.8840	2.51E-06	-0.4475	-0.9109	3.77E-07
us	36	0.6650	3.58E-03	0.3028	0.7603	3.97E-04
de	25	-0.7022	1.68E-03	-0.2179	-0.8132	7.22E-05
es	25	-0.0128	9.61E-01	-0.0021	0.1132	6.65E-01
nl	25	0.6127	8.92E-03	0.1812	0.7835	1.99E-04
fr	30	-0.1109	6.72E-01	-0.0232	-0.2680	2.98E-01

■ **Tabulka 5.3** Statistika charit

<sup>1</sup>Pearsonův korelační koeficient byl použit protože vztahy mezi Google Trends daty a TFR v Brazílii jsou prakticky lineární. V této situaci je lepší použít Pearsonův k.k.

## 5.2.4 Konspirační teorie

Tato kategorie se zabývá konspiračními teoriemi. Obsaženy byly názvy konspiračních teorií jako například falešné přistání na měsíci, HAARP, či QAnon. Velikost vzorku byla 70 klíčových slov. Nejzajímavější výsledek je z České republiky, kde Spearmanův korelační koeficient je -0,9133 s p-hodnotou 3,08E-07. Je to jev, který si osobně nedokážu vysvětlit a nenašel jsem k němu žádné zdroje.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	50	0.5132	3.51E-02	0.1280	-0.0902	7.31E-01
cz	43	-0.8397	2.48E-05	-0.4052	-0.9133	3.08E-07
us	65	-0.4249	8.91E-02	-0.0875	-0.3961	1.16E-01
de	45	-0.0174	9.47E-01	-0.0031	0.0430	8.70E-01
es	52	-0.4532	6.77E-02	-0.0405	-0.5621	1.88E-02
nl	45	0.0348	8.95E-01	0.0055	-0.0910	7.28E-01
fr	47	-0.2854	2.67E-01	-0.0431	-0.4499	7.00E-02

■ **Tabulka 5.4** Statistika konspiračních teorií

## 5.2.5 Konzumerismus

Kategorie konzumerismus se zabývá konzumním životem a známými značkami. Například značky jako IKEA, Walmart, Lidl, iPhone. Vzorek měl velikost 22 klíčových slov. Nejzajímavější data přinesly Česká republika se Spearmanovým korelačním koeficientem 0,9060 a p-hodnotou 3,77E-07 a Spojené státy americké s korelačním koeficientem -0,9430 s p-hodnotou 1,45E-08. Data Google Trends ohledně konzumerismu ukazují stálý nárůst zájmu o tato klíčová slova, avšak trend TFR v České republice 1.3 a Spojených státech 1.8 je odlišný. Můžeme se tedy domnívat, že konzumerismus buď nemá vliv na plodnost nebo specifické nároky lidí na konzumní život v různých státech mohou ovlivňovat úhrnnou plodnost.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	22	-0.6912	2.12E-03	-0.2959	-0.5963	1.15E-02
cz	20	0.9205	1.64E-07	1.0272	0.9060	5.55E-07
us	22	-0.9496	5.87E-09	-1.2041	-0.9430	1.45E-08
de	21	0.7157	1.24E-03	0.3325	0.7281	9.21E-04
es	20	-0.3243	2.04E-01	-0.1097	-0.2718	2.91E-01
nl	20	-0.5972	1.14E-02	-0.3202	-0.7159	1.23E-03
fr	18	-0.7997	1.17E-04	-0.3256	-0.6773	2.82E-03

■ **Tabulka 5.5** Statistika konzumerismu

## 5.2.6 Antikoncepce

Tato kategorie se zaměřuje na antikoncepci a medikamentózní interrupcí<sup>2</sup>. V kategorii byly zařazeny klíčová slova jako Postinor, kondomy nebo vasektomie. Velikost vzorku byla 10 klíčových slov. Je jasné, že antikoncepce bude mít vliv na plodnost. Data nedosáhly výrazných statistik, snad kromě Brazílie s Pearsonovou korelací 0,6730 s hodnotou 3,07E-03. Dávám to za vinu hlavně malému vstupnímu vzorku, kde bylo obsaženo málo možností antikoncepce.

<sup>2</sup>Interrupce, která je vyvolána léky. Například Plan B nebo Postinor.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	9	0.6730	3.07E-03	0.2214	-0.0245	9.26E-01
cz	8	-0.6178	8.23E-03	-0.2702	-0.4438	7.44E-02
us	10	-0.6874	2.30E-03	-0.5923	-0.5395	2.54E-02
de	8	0.4933	4.42E-02	0.1439	0.5623	1.88E-02
es	9	-0.6412	5.54E-03	-0.1169	-0.6212	7.78E-03
nl	8	-0.3985	1.13E-01	-0.0810	-0.3063	2.32E-01
fr	9	-0.2047	4.31E-01	-0.0292	-0.1512	5.62E-01

■ **Tabulka 5.6** Statistika antikoncepcí

## 5.2.7 Zločiny

Tato kategorie se zabývá zločinem. Nacházely se v ní klíčová slova související s domácím násilím či násilím ve školách. Velikost vzorku byla 5 klíčových slov. Data z Brazílie měla vysoký Pearsonův korelační koeficient 0,8387 s p-hodnotou 2,60E-05 na vzorku velikosti 3. Tato korelace je vysoká, ale vzhledem k velikosti vzorku není moc směrodatná. Data z analýzy ukázala, že domácí násilí může mít vliv na plodnost, ale násilí páchané v okolí zbraní na ně vliv nemá.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	3	0.8387	2.60E-05	0.4007	0.8417	2.27E-05
cz	1	-0.5205	3.22E-02	-0.3564	-0.3854	1.27E-01
us	5	0.5422	2.45E-02	0.2694	0.5739	1.60E-02
de	2	-0.4218	9.17E-02	-0.1124	-0.4457	7.30E-02
nl	1	0.5628	1.87E-02	0.1428	0.6463	5.06E-03
fr	2	-0.0536	8.38E-01	-0.0153	-0.2645	3.05E-01

■ **Tabulka 5.7** Statistika zločinu

## 5.2.8 Vzdelání

Tato kategorie se zaměřuje na vzdělání. Bere v potaz školky, školy i univerzity, kurzy jazyků a školní předměty. Vstupní vzorek činil 30 klíčových slov. Nejzajímavěji dopadly statistiky Brazílie a České republiky. Brazílie s 0,9173 Pearsonovy korelace a p-hodnotou 9,23E-11 a Česko s -0,9035 Spearmanovy korelace a p-hodnotou 6,69E-07. Oba výsledky vlastně říkají to samé, zájem o vzdělání se snižuje. Jenže v Česku při tom roste TFR a v Brazílii se snižuje. V analytické části zmiňuji v několika případech, že vzdělání žen má vliv na TFR, přičemž má dopad jak na hodnotu TFR, tak i na dobu prvního početí.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	22	0.9713	9.23E-11	0.8977	0.7509	5.13E-04
cz	20	-0.9070	5.13E-07	-0.5112	-0.9035	6.69E-07
us	28	0.7145	1.27E-03	0.5398	0.6941	2.00E-03
de	21	-0.1481	5.70E-01	-0.0342	0.0503	8.48E-01
es	20	0.1476	5.72E-01	0.0513	0.5228	3.13E-02
nl	19	0.4596	6.34E-02	0.1596	0.5966	1.15E-02
fr	18	-0.1357	6.04E-01	-0.0484	-0.4093	1.03E-01

■ **Tabulka 5.8** Statistika vzdělání

## 5.2.9 Zábava

Tato kategorie se zabývá zábavou, nachází se v ní klíčová slova jako Netflix, nebo velká sbírka filmů. Vstupní vzorek má velikost 308 klíčových slov. Předem bych chtěl zmínit, že ovlivnit výsledky překladu jsem mohl jen u jazyků které znám, to jsou český jazyk a anglický jazyk. Další překlady mohou i nemusí být správně přeloženy, ale vzhledem k jak velké intervenci jsem musel zakročit v případě češtiny, se dá říct, že data z ostatních zemí jsou nepoužitelná. Navíc filmy byly vybrány na základě hodnocení na IMDB a nejsou ohraničeny časem. Takže film, který byl uveden až později než v roce 2004 má s největší pravděpodobností trendovost v předchozích letech na nule. To by ale mohlo být vyrovnáno, pokud by byla zajištěna rovnoměrná distribuce vydání filmů ve vzorku. Ovšem toto vyrovnání nekontroluji.

V České republice byl Spearmanův korelační koeficient  $-0,7744$  a  $p$ -hodnota  $2,63E-04$ . To na základě dat 1.3 značí, že se postupně snižuje zájem o klíčová slova z oblasti zábavy. Osobně bych očekával spíše opačný trend.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	209	0.1196	6.48E-01	0.0139	-0.0233	9.29E-01
cz	279	-0.7560	4.46E-04	-0.2444	-0.7744	2.63E-04
us	297	-0.2911	2.57E-01	-0.0483	-0.3397	1.82E-01
de	220	0.5260	3.01E-02	0.0646	0.4948	4.35E-02
es	254	-0.4351	8.09E-02	-0.0256	-0.5818	1.43E-02
nl	253	0.1490	5.68E-01	0.0132	0.0185	9.44E-01
fr	242	-0.2016	4.38E-01	-0.0189	-0.2200	3.96E-01

■ **Tabulka 5.9** Statistika zábavy

## 5.2.10 Environmentalismus

Kategorie se zabývá tématy ochrany přírodního prostředí. Klíčová slova se týkají alternativních získávání elektřiny, frakování nebo kvality vody. Velikost vstupního vzorku je 22 klíčových slov. Na této kategorii jde i vidět důvod, proč je lepší používat v případě Brazílie Pearsonův korelační koeficient. Je to taky jediná statisticky zajímavá hodnota, s korelačním koeficientem  $0,7467$  a  $p$ -hodnotou  $5,74E-04$ . Vzhledem ke konvexnosti environmentální křivky a neměnnému trendu TFR, si nemyslím, že tento soubor hodnot má vliv na TFR. Je možné, že některé hodnoty zabývající se konkrétními místními ekologickými problémy, pak mohou mít vysokou korelaci nebo i kauzální vztah.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	12	0.7467	5.74E-04	0.2981	0.2859	2.66E-01
cz	10	0.3004	2.41E-01	0.0619	0.2876	2.63E-01
us	14	0.3975	1.14E-01	0.2047	0.4427	7.52E-02
de	13	-0.0505	8.47E-01	-0.0090	-0.1461	5.76E-01
es	12	-0.0508	8.47E-01	-0.0108	-0.0831	7.51E-01
nl	10	-0.6722	3.12E-03	-0.1017	-0.5609	1.92E-02
fr	12	0.0735	7.79E-01	0.0188	0.1328	6.11E-01

■ **Tabulka 5.10** Statistika environmentalismu

### 5.2.11 Gambling

Tato kategorie se zabývá různými sázkovými kanceláři a loteriemi. Například sázkovými kanceláři jako je Bet365, Betway nebo loterie Powerball. Vstupní vzorek činil 19 klíčových slov. Nejslibnější výsledek se vyskytl v Nizozemsku, kde Spearmanův korelační koeficient je -0,8696 s p-hodnotou 5,77E-06. V praxi to znamená, že Nizozemci plodí méně dětí a zvyšuje se jim zájem o sázkové kanceláře, kasína a loterie. Domnívám se, že tento jev nemá výrazný vliv na plodnost. Nanejvýše u párů, které jsou na sázení závislé.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	16	-0.1503	5.65E-01	-0.0415	-0.1485	5.70E-01
cz	15	0.7716	2.86E-04	0.4859	0.7535	4.78E-04
us	17	-0.3057	2.33E-01	-0.1268	-0.3262	2.01E-01
de	14	0.7878	1.74E-04	0.4189	0.7551	4.58E-04
es	16	-0.5568	2.02E-02	-0.1347	-0.5486	2.26E-02
nl	15	-0.9066	5.29E-07	-0.4731	-0.8696	5.77E-06
fr	17	-0.4953	4.32E-02	-0.0738	-0.4302	8.47E-02

■ **Tabulka 5.11** Statistika gamblingu

### 5.2.12 Hry

V této kategorii se zabývám videohrami. V této kategorii se nachází hry jako Grand Theft Auto, Minecraft, League of Legends nebo Dota. Vstupní vzorek činil 15 klíčových slov. Má idea za touto kategorií byla, že více dětí bude znamenat více hráčů her. Na druhou stranu by to také mohlo znamenat zvýšený počet hráčů na druhé straně věkové škály a to u lidí, kteří jsou ve věku, ve kterém jejich rodiče už měli děti. Statistiky ale neprokázaly nějaké výrazné korelace, kromě Brazílie s -0,5790 a p-hodnotou 1,49E-02 při velikosti vzorku 10. Tento výsledek mi nepřijde statisticky významný, vzhledem k jeho nízké absolutní hodnotě.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	10	-0.5790	1.49E-02	-0.4182	-0.5914	1.24E-02
cz	15	0.0816	7.56E-01	0.0899	0.0430	8.70E-01
us	15	-0.3236	2.05E-01	-0.3665	-0.3765	1.36E-01
de	11	0.0967	7.12E-01	0.0704	0.2861	2.66E-01
es	14	-0.1269	6.27E-01	-0.0472	0.0037	9.89E-01
nl	14	0.4561	6.58E-02	0.1775	0.3998	1.12E-01
fr	15	0.4371	7.94E-02	0.2267	0.4683	5.79E-02

■ **Tabulka 5.12** Statistika her

### 5.2.13 Zbraně

Tato kategorie se zaměřuje na zbraně. Zahrnuty jsou například názvy zbraní jako je CZ-75, G46 nebo AR-15. Vstupní vzorek měl velikost 17 klíčových slov. Nejzajímavěji dopadly statistiky ze Spojených států a Francie. Spojené státy mají Spearmanův korelační koeficient na hodnotě -0,7811 a p-hodnotě 2,14E-04 a Francie na hodnotě -0,7707 a p-hodnotu 2,93E-04. V analytické části byla zmíněna studie zabývající se porodností a násilím spáchaným zbraněmi. Studie násilí spáchané zbraní neshledala jako faktor, který by mohl ovlivnit snížení plodnosti nebo ovlivnit vývoj plodu. Zbraně a zájem o ně může mít vliv na životní styl jedinců, ať už v oblasti například myslivectví či lovu, což může následně ovlivnit i TFR.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	16	-0.0271	9.18E-01	-0.0068	-0.3939	1.18E-01
cz	15	-0.6526	4.51E-03	-0.2418	-0.4401	7.71E-02
us	16	-0.8101	8.08E-05	-0.7198	-0.7811	2.14E-04
de	17	0.6048	1.01E-02	0.1498	0.6028	1.04E-02
es	16	-0.3753	1.38E-01	-0.0427	-0.5141	3.47E-02
nl	13	0.1116	6.70E-01	0.0152	-0.0541	8.37E-01
fr	16	-0.6858	2.37E-03	-0.1273	-0.7707	2.93E-04

■ **Tabulka 5.13** Statistika zbraní

### 5.2.14 Zdravotní péče

Tato kapitola se zabývá zdravotní péčí. Nachází se zde klíčová slova zabývající se redukcí kouření, vlivu slunce na kůži a životosprávou. Velikost vstupního vzorku je 55 klíčových slov. Data ze Spojených států dopadly asi nejméně výrazněji se Spearmanovou korelací -0,9430 a p-hodnotou 1,45E-08. V analytické části jsem zmiňoval studii zabývající se náklady na zdravotní systém ze strany státu a jejich vliv na úmrtnost matek z důvodu využívání většího množství nekvalifikovaných zdravotníků. Je možné, že data ze Spojených států mají vyšší korelaci z důvodu absence univerzální zdravotní péče, kterou známe z jiných částí světa, například Evropy. Proto místo toho, aby své zdravotní záležitosti konzultovali s lékařem, dotazují se zpočátku hlavně internetu.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	34	0.3146	2.19E-01	0.1196	-0.0699	7.90E-01
cz	29	-0.3013	2.40E-01	-0.0714	0.0283	9.14E-01
us	47	-0.9265	9.33E-08	-0.5512	-0.9430	1.45E-08
de	33	0.6245	7.36E-03	0.1750	0.6262	7.17E-03
es	32	-0.4027	1.09E-01	-0.0610	-0.5191	3.27E-02
nl	33	-0.0482	8.54E-01	-0.0081	-0.4760	5.34E-02
fr	30	-0.8023	1.07E-04	-0.2842	-0.6921	2.08E-03

■ **Tabulka 5.14** Statistika zdravotní péče

### 5.2.15 Bydlení

V této kategorii se zabývám bydlením. Ve vzorku se nachází klíčová slova jako byt k pronájmu, nebo spojení byt k pronájmu a velké město ze zkoumaných států, například Berlín nebo Los Angeles. Velikost vstupního vzorku činila 28 klíčových slov. Data z Brazílie, České republiky, Spojených států, Německa a Nizozemí vykazují podobný trend a to vyšší zájem o bydlení. V analytické části jsem se zmínil o článku, který psal o faktorech bydlení na plodnost. Došel k závěru, že změna bydlení může mít vliv na plodnost. Ale reálné zvýšení plodnosti nastává jen v České republice.

### 5.2.16 Ideologie

Tato kategorie se zabývá klíčovými slovy zabývající se ideologiemi. Příkladem takových slov je decentralizace, komunistický manifest, Marx nebo NAP<sup>3</sup>. Velikost vstupního vzorku byla 37 klíčových slov. Nejzajímavěji dopadla Brazílie s Pearsonovým korelačním koeficientem 0,8642 a p-hodnotou 7,72E-06. Například výrazy jako Marx, Che Guevara nebo odbory měly korelace přes

<sup>3</sup>Non-aggression principle, jedna z hlavních myšlenek anarchokapitalismu

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	13	-0.5839	1.39E-02	-0.1633	-0.7288	9.04E-04
cz	8	0.8041	1.00E-04	0.4351	0.7634	3.63E-04
us	17	-0.7603	3.96E-04	-0.3465	-0.7836	1.98E-04
de	12	-0.8983	9.76E-07	-0.4653	-0.8034	1.02E-04
es	16	-0.2890	2.61E-01	-0.0523	-0.3370	1.86E-01
nl	10	-0.6393	5.73E-03	-0.1368	-0.6937	2.01E-03
fr	10	0.3599	1.56E-01	0.1254	0.4708	5.65E-02

■ **Tabulka 5.15** Statistika bydlení

0,97. Je možné, že čím více lidí opouští levicové ideologie, tím méně motivací pak mají k plození dětí. To by ale mělo implikovat i klesající porodnost v České republice, ale to se neděje 1.3. Komunistická strana Čech a Moravy a Česká strana sociálně demokratická dlouhodobě ztrácí voliče a plodnost stále roste. Ale možná jde jen o transformaci české levice do catch-all-party<sup>4</sup>, jako je v Česku například ANO. Myslím, si že by mohlo mít význam se zaměřit na ženy levicového smýšlení v porovnání s těmi pravicovými.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	30	0.8642	7.72E-06	0.3947	0.4479	7.14E-02
cz	24	-0.7970	1.28E-04	-0.4126	-0.7880	1.72E-04
us	33	-0.0686	7.94E-01	-0.0325	-0.3004	2.41E-01
de	29	-0.2212	3.94E-01	-0.0687	-0.2014	4.38E-01
es	26	0.1195	6.48E-01	0.0249	0.3715	1.42E-01
nl	25	0.2986	2.44E-01	0.0860	0.5474	2.30E-02
fr	29	-0.3013	2.40E-01	-0.0955	-0.6466	5.03E-03

■ **Tabulka 5.16** Statistika ideologií

### 5.2.17 Nelegální drogy

Tato kategorie se zabývá drogami. V této kategorii se nachází státem regulované drogy. V některých státech jsou některé tyto látky povoleny. V této sadě se nachází látky jako heroin, extáze, mdma, ketamin nebo fentanyl. Velikost vstupního vzorku je 29 položek. V analytické části zmiňuji několik drog, které mohou mít na plodnost jedinců vliv, například konopí, které má vliv na obě pohlaví. Výsledky dopadly zajímavě v celé západní Evropě a Spojených státech. V Nizozemí 1.7 je Spearmanova korelace zdaleka největší s korelačním koeficientem -0,9619 a p-hodnotou 7,51E-10. Vzhledem k tomu, že v Nizozemsku plodnost klesá, by se dalo říci, že drogy mohou mít vliv na plodnost, ale naprosto opačná situace se děje v Německu, kde plodnost stoupá a s ní stoupají i vyhledávání na Googlu ohledně nelegálních drog. Z těchto dat nelze vyvodit jednoznačný závěr.

### 5.2.18 Investování

Tato kategorie se zabývá investováním, V této kategorii se nachází klíčová slova jako prodejní opce, kryptoměna, Bitcoin, Ethereum, nákup kryptoměn nebo pasivní příjem. Velikost tohoto vzorku je 17 klíčových slov. Nejzajímavěji vycházejí data z Nizozemska, Francie, ale i ty mají pouze -0,6138 a -0,6331 Spearmanův korelační koeficient. Výsledky vyhledávání na Google

<sup>4</sup>český překlad je všelidové strany



Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	25	0.3794	1.33E-01	0.1577	-0.0810	7.57E-01
cz	22	-0.2716	2.92E-01	-0.1146	-0.0111	9.66E-01
us	29	-0.8394	2.52E-05	-0.8145	-0.9013	7.88E-07
de	28	0.9179	2.08E-07	0.7218	0.8410	2.34E-05
es	28	-0.6667	3.47E-03	-0.1906	-0.6691	3.31E-03
nl	25	-0.9662	3.11E-10	-0.3850	-0.9619	7.51E-10
fr	26	-0.9081	4.70E-07	-0.3563	-0.8961	1.14E-06

■ **Tabulka 5.17** Statistika nelegálních drog

ohledně investic nemají významné korelace ani příliš velký vzorek na to, aby se staly statisticky významné.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	16	0.3356	1.88E-01	0.1130	0.0761	7.72E-01
cz	11	-0.0502	8.48E-01	-0.0231	0.0037	9.89E-01
us	17	-0.2856	2.66E-01	-0.1759	-0.1754	5.01E-01
de	17	0.6983	1.82E-03	0.2540	0.5719	1.65E-02
es	16	-0.1814	4.86E-01	-0.0478	-0.3530	1.65E-01
nl	17	-0.7628	3.69E-04	-0.2076	-0.6138	8.77E-03
fr	16	-0.5351	2.69E-02	-0.1454	-0.6331	6.38E-03

■ **Tabulka 5.18** Statistika investování

## 5.2.19 Práce

Tato kategorie se zabývá prací. Nachází se v ní klíčová slova jako nabídky pracovních míst, nová práce, jak vytvořit životopis, motivační dopis. Velikost vzorku byla 17 klíčových slov. Data ukazují velkou Pearsonovu korelaci 0,9207 s p-hodnotou 1,62E-07 v Brazílii. To znamená, že zájem o změnu zaměstnání klesá. Pokles zájmu o nová pracovní místa se děje i v České republice s -0,8298 Spearmanovou korelací a p-hodnotou 3,78E-05. Naopak ve Spojených státech se zájem o změnu pracovní pozice zvyšuje, zatímco plodnost klesá. Vypadá to, že práce není faktor plodnosti, což podporuje i článek rozebraný v analýze zaměřující se na náročnost zaměstnání a plodnosti pracovníků.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	11	0.9207	1.62E-07	0.2648	0.9399	2.14E-08
cz	6	-0.8572	1.10E-05	-0.3394	-0.8298	3.78E-05
us	12	-0.8810	3.01E-06	-0.5529	-0.8853	2.31E-06
de	10	-0.2012	4.39E-01	-0.0635	-0.1215	6.42E-01
es	11	-0.2502	3.33E-01	-0.0410	-0.4945	4.36E-02
nl	11	-0.3879	1.24E-01	-0.1066	-0.4539	6.72E-02
fr	8	0.0379	8.85E-01	0.0077	0.0578	8.26E-01

■ **Tabulka 5.19** Statistika práce

### 5.2.20 Legální drogy

Tato kategorie se zabývá drogami, které jsou legální. Mezi tyto drogy patří alkohol, káva nebo tabákové výrobky. Velikost vstupního vzorku je 70 klíčových slov. Ve Spojených státech jsou data nejpřesvědčivější. Spearmanův korelační koeficient má -0,9552 s p-hodnotou 2,45E-09. Což znamená, že se zvyšuje zájem o legální drogy a zároveň klesá plodnost. Zvýšený zájem o legální drogy se nachází ve všech sledovaných státech. Zdá se, že nemají vliv na TFR. Z analytické části ale vím, že tabákové výrobky mohou ovlivnit dobu oplodnění.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	28	0.1411	5.89E-01	0.0358	-0.3337	1.90E-01
cz	29	0.6649	3.59E-03	0.4019	0.6921	2.08E-03
us	32	-0.9778	1.37E-11	-0.8353	-0.9552	2.45E-09
de	30	0.8881	1.94E-06	0.3883	0.8226	5.04E-05
es	30	-0.6821	2.56E-03	-0.1678	-0.6999	1.76E-03
nl	30	-0.9196	1.79E-07	-0.4480	-0.8352	3.02E-05
fr	28	-0.8701	5.62E-06	-0.2434	-0.8740	4.53E-06

■ **Tabulka 5.20** Statistika legálních drog

### 5.2.21 Nízký příjem

Tato kategorie se zabývá termíny zaměřující se na lidi s nízkými příjmy. V této kategorii se nachází klíčová slova jako potravinové známky nebo podpora v nezaměstnanosti. Vzorek má velikost 14 klíčových slov. Zajímavý Spearmanův korelační koeficient vyšel ve Francii s -0,8740 a p-hodnotou 4,53E-06, ale tato data vycházejí z malého vzorku. Data z Německa ukazují, že zájem o slova z této kategorie klesal do roku 2008 a poté narůstal. V analytické části zmiňují, že Německé děti jsou jedny z nejvíce ohrožených chudobou v Evropě, což dle mého názoru poměrně odpovídá výsledkům. Určitě by stálo za to se na tuto kategorii více zaměřit.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	9	0.7640	3.56E-04	0.2285	0.4748	5.41E-02
cz	10	0.1857	4.75E-01	0.1379	0.1979	4.46E-01
us	13	-0.2938	2.52E-01	-0.2807	-0.1950	4.53E-01
de	11	0.7613	3.85E-04	0.1908	0.8189	5.81E-05
es	12	-0.6436	5.30E-03	-0.1333	-0.6175	8.27E-03
nl	9	0.4743	5.44E-02	0.0891	0.3948	1.17E-01
fr	8	-0.7650	3.47E-04	-0.2411	-0.8740	4.53E-06

■ **Tabulka 5.21** Statistika nízkých příjmů

### 5.2.22 Manželství

Tato kategorie se zabývá zánikem manželství. Součástí této kategorie jsou výrazy jako rozvodový právník nebo manželské poradenství. Velikost datasetu je 8 klíčových slov. Nejzajímavější jsou data z Brazílie, kde byla Pearsonova korelace 0,8826 a p-hodnotou 2,73E-06 na vzorku o velikosti pouze 5. Ale přesto může rozvodovost mít za následek menší TFR kvůli menšímu počtu párů, které spolu žijí.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	5	0.8826	2.73E-06	0.3473	0.5595	1.95E-02
cz	2	-0.1553	5.52E-01	-0.0609	-0.1636	5.30E-01
us	6	-0.7056	1.55E-03	-0.1691	-0.6769	2.84E-03
de	5	0.0500	8.49E-01	0.0172	0.1105	6.73E-01
es	4	-0.2613	3.11E-01	-0.0407	-0.4123	1.00E-01
nl	3	0.6141	8.73E-03	0.2894	0.7848	1.91E-04
fr	3	-0.1223	6.40E-01	-0.0265	-0.0860	7.43E-01

■ **Tabulka 5.22** Statistika manželství

### 5.2.23 Zprávy

Tato kategorie se zabývá televizními zprávami a tiskem. Mezi tyto zprávy a noviny se řadí například BBC, USA today, ČT, Die Welt, The Guardian. Nachází se v ní i noviny z jiných států jako je Čína nebo Indie. Velikost vzorku je 77 klíčových slov. Nejzajímavější jsou data z Brazílie, kde je Pearsonova korelace 0,9319 při p-hodnotě 5,32E-08. To znamená, že zájem o tisk upadá, to samé nastává i v České republice, kde je Spearmanova korelace -0,7621 a p-hodnota 3,76E-04. Takže v Brazíli s úpadkem zájmu o noviny upadá i TFR a v Česku s úpadkem zájmu o tisk roste TFR. Tato tvrzení se navzájem vyvrací. Informovanost může mít vliv na rozhodnutí jedinců, ale pravděpodobně nemění TFR států.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	56	0.9319	5.32E-08	0.2398	0.9141	2.88E-07
cz	52	-0.8717	5.15E-06	-0.2094	-0.7621	3.76E-04
us	65	0.4684	5.79E-02	0.2067	0.4464	7.25E-02
de	65	-0.0185	9.44E-01	-0.0034	0.0479	8.55E-01
es	60	-0.4983	4.18E-02	-0.0393	-0.6310	6.60E-03
nl	54	-0.0022	9.93E-01	-0.0003	-0.1242	6.35E-01
fr	59	0.1213	6.43E-01	0.0272	0.2213	3.93E-01

■ **Tabulka 5.23** Statistika zpráv

### 5.2.24 Náboženství

V této kategorii se nacházejí klíčová slova související s abrahámovskými náboženstvími. Tento soubor obsahuje pojmy jako mešita, halal, šabat, Jidiš, antisemitismus, nebo názvy různých církví jako Luteráni nebo Metodisté. Velikost vzorku byla 30 klíčových slov. V analytické části zmiňují data z Česka, která říkají že nábožensky založené ženy měly větší plodnost než ty, které jsou nevěřící. Podobná data známe i z Francie. V České republice mají data negativní korelaci. To znamená, že s rostoucí TFR se propadá zájem o vyhledávání informací o náboženstvích. Ačkoliv se to do německých dat nepropsalo, některé pojmy týkající se muslimské komunity mají vysokou korelaci s rostoucí plodností v Německu. Jeden příklad za všechny: slovo Moschee - mešita, má Spearmanovu korelaci 0,9515 s p-hodnotou 4,43E-09.

### 5.2.25 S&P 500

Tato kategorie se zabývá akciovým indexem Standard & Poor's 500. Nachází se v ní 500 největších firem, zalistovaných na americkém akciovém trhu. Velikost vzorku je 505, jelikož některé firmy jsou z pohledu akciového trhu rozděleny na menší celky. Z pohledu na data lze vidět, že zájem o

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	35	0.7272	9.40E-04	0.1979	0.1166	6.56E-01
cz	34	-0.8802	3.16E-06	-0.3560	-0.8924	1.46E-06
us	37	0.0192	9.42E-01	0.0057	-0.0822	7.54E-01
de	35	0.4345	8.14E-02	0.0818	0.3941	1.18E-01
es	36	-0.4275	8.69E-02	-0.0469	-0.5621	1.88E-02
nl	33	-0.5685	1.73E-02	-0.0688	-0.6277	6.98E-03
fr	35	-0.7250	9.91E-04	-0.1922	-0.7425	6.41E-04

■ **Tabulka 5.24** Statistika náboženství

tyto firmy klesá všude kromě Německa a Spojených Států. Dle mého názoru, je všeobecný trend snížení popularity zájmu o tyto společnosti. Proto data z Brazílie nejsou směrodatná. Data ze Spojených států zase ukazují i velkou fluktuaci dat. Nemyslím si, že z této kategorie lze vývést spolehlivé závěry.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	324	0.9154	2.59E-07	0.2416	0.9178	2.10E-07
cz	215	-0.7917	1.53E-04	-0.3521	-0.7658	3.38E-04
us	488	0.1335	6.09E-01	0.0305	0.0785	7.65E-01
de	436	-0.2965	2.48E-01	-0.0650	-0.4297	8.52E-02
es	345	-0.0280	9.15E-01	-0.0036	0.1513	5.62E-01
nl	313	0.1890	4.68E-01	0.0310	0.2964	2.48E-01
fr	356	-0.2276	3.80E-01	-0.0481	-0.6417	5.49E-03

■ **Tabulka 5.25** Statistika S&P 500

## 5.2.26 Sporty

Tato kategorie se zabývá sporty. Příklady klíčových slov z této kategorie jsou fotbal, basketbal, hokej nebo Sebastian Vettel. Velikost vstupního vzorku činila 15 klíčových slov. V analytické části zmiňují, že plodnost v Katalánsku vzrosla po úspěšné fotbalové sezóně. Nejzajímavější data pocházejí z Německa a Spojených států. V Německu plodnost stoupá, ve Spojených státech klesá. Zájem o sporty v obou zemích má podobnou křivku. Od 2004 rostly a v posledních pěti letech buď fluktuují nebo klesají. Čímž se vlastně tato data vyvracejí.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	14	0.6136	8.80E-03	0.1430	0.2945	2.51E-01
cz	14	0.4496	7.02E-02	0.1150	0.4561	6.58E-02
us	14	-0.7895	1.64E-04	-0.4532	-0.7958	1.33E-04
de	14	0.8085	8.56E-05	0.1768	0.8250	4.57E-05
es	14	-0.6494	4.79E-03	-0.0624	-0.6138	8.77E-03
nl	14	-0.0421	8.73E-01	-0.0048	-0.1931	4.58E-01
fr	14	-0.2984	2.45E-01	-0.0372	-0.3503	1.68E-01

■ **Tabulka 5.26** Statistika sportů

### 5.2.27 Steroidy

Tato kategorie se zabývá steroidy. Nachází se zde látky jako kortikosteroidy, danazol nebo anabolika. Velikost vzorku je 25 klíčových slov. V analytické části zmiňují studii zabírající se steroidy, které s velkou pravděpodobností nemají vliv na plodnost mužů. Jediný významný výsledek přineslo Nizozemsko se Spearmanovou korelací -0,8057 a p-hodnotou 9,47E-05. V této situaci je nutné kontrolovat specifické steroidy a hledat jejich možný vliv na potenci. Myslím si, že pokud by byl vliv steroidů na plodnost výrazný, pořad by se nepropsal do dat TFR, kvůli relativně malému počtu lidí, kteří je využívají.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	16	-0.3600	1.56E-01	-0.0784	-0.4217	9.18E-02
cz	10	-0.7144	1.27E-03	-0.2629	-0.6724	3.10E-03
us	19	0.4234	9.04E-02	0.1409	0.5064	3.80E-02
de	14	-0.6968	1.88E-03	-0.1525	-0.6335	6.33E-03
es	11	-0.0577	8.26E-01	-0.0094	0.0246	9.25E-01
nl	10	-0.7654	3.43E-04	-0.1856	-0.8057	9.47E-05
fr	9	-0.4645	6.03E-02	-0.0785	-0.6494	4.78E-03

■ **Tabulka 5.27** Statistika steroidů

### 5.2.28 Turismus

Tato kapitola se zabývá turismem. Nachází se zde klíčová slova jako dovolená v Itálii, tipy na dovolenou, Plitvická jezera nebo festival Coachella. Soubor dat měl velikost 115 klíčových slov. V datasetu se nacházejí i klíčová slova zabývající se národními parky a zoologickými zahradami. Nejvyšší korelací mají data z Brazílie s Pearsonovým kvocientem 0,6170 a p-hodnotou 8,33E-03. Očekával jsem v této kategorii lepší úspěšnost dat, jelikož jsem předpokládal spojení dětí, zoo a dovolených. To se nevyplnilo a nevypadá to, že by data nějak ovlivňovaly porodnost.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	53	0.6170	8.33E-03	0.1550	0.0994	7.04E-01
cz	62	-0.2907	2.58E-01	-0.0738	-0.1082	6.79E-01
us	83	0.2383	3.57E-01	0.0976	0.0748	7.75E-01
de	67	0.0938	7.20E-01	0.0223	0.1608	5.37E-01
es	49	-0.4480	7.13E-02	-0.0639	-0.4600	6.32E-02
nl	61	0.0912	7.28E-01	0.0132	-0.2718	2.91E-01
fr	6	-0.3595	1.56E-01	-0.0630	-0.4438	7.44E-02

■ **Tabulka 5.28** Statistika turismu

### 5.2.29 Univerzity

Tato sekce se zabývá univerzitami. V datasetu se nachází 300 světových univerzit. Nachází se zde i Harvardova univerzita, Caltech nebo Berkeley. Od této kategorie jsem měl největší očekávání. Mnoho článků zmiňovalo vzdělání jako jeden z největších negativních faktorů plodnosti u mladých žen. Brazílie má Pearsonovu korelací 0,9469 a p-hodnotu 8,66E-09, což přesvědčuje spíše o opaku předpokladu. Zájem o světové vysoké školy v Brazíli upadá a s ním i TFR. Podobnému závěru lze dospět i ve Spojených státech. V Německu TFR roste a zájem o univerzity klesá, což odpovídá mému předpokladu. Ale pokud vidíme, že data z Brazílie a Spojených států ukazují podobný pád zájmu o univerzity, pak můžeme dojít k závěru, že univerzity nemusí být faktorem TFR.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	109	0.9469	8.66E-09	0.3248	0.9460	9.74E-09
cz	47	-0.7077	1.48E-03	-0.3623	-0.5077	3.75E-02
us	274	0.8672	6.58E-06	1.1105	0.9503	5.27E-09
de	101	-0.7232	1.04E-03	-0.2521	-0.9171	2.23E-07
es	91	0.1358	6.03E-01	0.0206	0.4600	6.32E-02
nl	87	0.4857	4.81E-02	0.1240	0.7405	6.75E-04
fr	72	-0.0456	8.62E-01	-0.0126	0.0774	7.68E-01

■ **Tabulka 5.29** Statistika univerzit

### 5.2.30 Emigrace

Tato kategorie se zabývala emigrací, vzorek měl dvě klíčová slova a to jak se přestěhovat do Evropy a jak se přestěhovat do UK. Vzhledem k malému vzorku a datům pouze z USA nelze tyto data brát za statisticky směrodatná.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
us	2	-0.8037	1.02E-04	-0.4591	-0.8204	5.50E-05

■ **Tabulka 5.30** Statistika emigrací

### 5.2.31 Statistiky

Tato kategorie je velice malá a zabývá se statistikami. Vzorek obsahuje pouze dvě klíčová slova. Například zlepšené životní standardy a HDP na hlavu. Vzorek je malý, nelze z něj dělat velké závěry.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	2	0.8656	7.18E-06	0.2491	0.5169	3.36E-02
cz	1	-0.6216	7.73E-03	-0.1656	-0.5709	1.67E-02
us	1	-0.9361	3.35E-08	-1.3729	-0.9571	1.81E-09
es	1	0.3082	2.29E-01	0.0727	0.1477	5.72E-01
nl	1	-0.0615	8.15E-01	-0.0167	-0.3313	1.94E-01
fr	1	-0.7296	8.86E-04	-0.4191	-0.6908	2.14E-03

■ **Tabulka 5.31** Statistika statistik

### 5.2.32 Jiné

Tato kategorie obsahuje slova, které nebyly zařazeny do jiných kategorií. Patří tu klíčová slova jako volební hlasování nebo bestialita. Statistiky jsou uvedeny jen pro kompletnost.

Země	Vzorek	Pearsonův k.k.	p-hodnota	Kovariance	Spearmanův k.k.	p-hodnota
br	4	0.9730	5.77E-11	0.5688	0.9840	1.16E-12
cz	3	-0.1820	4.85E-01	-0.1279	-0.2778	2.80E-01
us	4	0.7730	2.74E-04	0.8508	0.7762	2.49E-04
de	3	-0.0481	8.55E-01	-0.0237	0.0037	9.89E-01
es	4	-0.4536	6.74E-02	-0.1464	-0.4244	8.96E-02
nl	4	0.0903	7.30E-01	0.0441	-0.0234	9.29E-01
fr	2	-0.5530	2.13E-02	-0.1331	-0.6478	4.92E-03

■ **Tabulka 5.32** Statistika jiných dat





## Kapitola 6

# Závěr

Má bakalářská práce si dala za cíl analyzovat zdroje z Google Trends a Google Ngram s hodnotami úhrnné plodnosti v různých zemích. Pro tento účel bylo potřeba získat data z již zmíněných nástrojů od Googlu. Získání dat bylo nejnáročnější částí práce vzhledem k velikosti dat získaných z Google Trends. Data z Google Ngram byly časově mnohem náročnější na získání, a proto jich je mnohem méně. K analýze byla potřeba implementace aplikace, která by uživateli ukázala zajímavé veličiny, které mohou mít vztah s úhrnnou plodností.

Zkoumání veličin odhalilo pár zajímavých hodnot, které určitě stojí za zmínku. Například levicové termíny jako jsou Che Guevara, Marx nebo odbory korelují s úpadkem plodnosti v Brazílii nebo kategorie Vzdělání v České republice, která má korelaci  $-0,9035$  a obsahuje klíčová slova jako chemie, zeměpis, vysoká škola nebo Univerzita. Překvapivou veličinou poté byla kategorie konspirační teorie, která v České republice měla korelaci  $-0,9133$ . Další kategorií, která stojí za zmínku je kategorie nízký příjem, která dosahuje korelaci  $0,8189$ , což může potvrzovat tvrzení z analytické části o vysokém riziku chudoby malých dětí v Německu.

Tato práce by mohla být rozšířena za pomoci přesnějších zkoumaných veličin. Bylo by možné rozšířit zkoumaná data na měsíční data úhrnné plodnosti a měsíční data Google Trends. Ty by díky vyšší hustotě dat zpřesnily statistické výstupy dvanáctkrát.



# Bibliografie

1. The Baby Boom and Its Causes: What We Know and What We Need to Know. *Population and Development Review* [online]. 2013, roč. 39, č. 2, s. 257–288 [cit. 2022-12-23]. ISSN 00987921, ISSN 17284457. Dostupné z: <http://www.jstor.org/stable/41857595>.
2. SMOAK, Natalie. fertility rate. In: *Encyclopedia Britannica* [online]. [B.r.] [cit. 2022-12-04]. Dostupné z: <https://www.britannica.com/topic/fertility-rate>.
3. OECD. Fertility rates [online]. 2014 [cit. 2022-12-04]. Dostupné z DOI: <https://doi.org/10.1787/8272fb01-en>.
4. THE WORLD BANK. *Fertility rate, total (births per woman)* [online]. 2021. [cit. 2022-12-05]. Dostupné z: <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?end=2020&start=1960&view=chart>.
5. SILVA, Tiloka de; TENREYRO, Silvana. Population Control Policies and Fertility Convergence. *The Journal of Economic Perspectives* [online]. 2017, roč. 31, č. 4, s. 205–228 [cit. 2022-12-23]. ISSN 08953309. Dostupné z: <http://www.jstor.org/stable/44425388>.
6. FINKLE, Jason L.; CRANE, Barbara B. The Politics of Bucharest: Population, Development, and the New International Economic Order. *Population and Development Review* [online]. 1975, roč. 1, č. 1, s. 87–114 [cit. 2022-12-23]. ISSN 00987921, ISSN 17284457. Dostupné z: <http://www.jstor.org/stable/1972272>.
7. GREBENIK, E. *The Economic Journal* [online]. 1940, roč. 50, č. 200, s. 488–490 [cit. 2022-12-23]. ISSN 00130133, ISSN 14680297. Dostupné z: <http://www.jstor.org/stable/2226211>.
8. *How accurate is Google Translate?: Google Translate vs DeepL*. Riga: Lokalise, 2022. Dostupné také z: <https://lokalise.com/blog/google-translate-accuracy/>.
9. MARTINE, George. Brazil's Fertility Decline, 1965-95: A Fresh Look at Key Factors. *Population and Development Review* [online]. 1996, roč. 22, č. 1, s. 47–75 [cit. 2022-12-05]. ISSN 00987921, ISSN 17284457. Dostupné z: <http://www.jstor.org/stable/2137686>.
10. FERRARA, Eliana La; CHONG, Alberto; DURYEA, Suzanne. Soap Operas and Fertility: Evidence from Brazil. *American Economic Journal: Applied Economics* [online]. 2012, roč. 4, č. 4, s. 1–31 [cit. 2022-12-05]. ISSN 19457782, ISSN 19457790. Dostupné z: <http://www.jstor.org/stable/23269740>.
11. IUTAKA, S.; BOCK, E. W.; VARNES, W. G. Factors Affecting Fertility of Natives and Migrants in Urban Brazil. *Population Studies* [online]. 1971, roč. 25, č. 1, s. 55–62 [cit. 2022-12-05]. ISSN 00324728. Dostupné z: <http://www.jstor.org/stable/2172747>.

12. SOBOTKA, Tomáš; ŠTASTNÁ, Anna; ZEMAN, Kryštof; HAMPLOVÁ, Dana; KANTOROVÁ, Vladimíra. Czech Republic: A rapid transformation of fertility and family behaviour after the collapse of state socialism. *Demographic Research* [online]. 2008, roč. 19, s. 403–454 [cit. 2022-12-05]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26349255>.
13. KLASEN, Stephan; LAUNOV, Andrey. Analysis of the Determinants of Fertility Decline in the Czech Republic. *Journal of Population Economics* [online]. 2006, roč. 19, č. 1, s. 25–54 [cit. 2022-12-05]. ISSN 09331433, ISSN 14321475. Dostupné z: <http://www.jstor.org/stable/20007996>.
14. CRANNEY, Stephen. The Association Between Belief in God and Fertility Desires in Slovenia and the Czech Republic. *Perspectives on Sexual and Reproductive Health* [online]. 2015, roč. 47, č. 2, s. 83–89 [cit. 2022-12-05]. ISSN 15386341, ISSN 19312393. Dostupné z: <https://www.jstor.org/stable/48576721>.
15. TOULEMON, Laurent; PAILHÉ, Ariane; ROSSIER, Clémentine. France: High and stable fertility. *Demographic Research* [online]. 2008, roč. 19, s. 503–556 [cit. 2022-12-05]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26349257>.
16. OLIVETTI, Claudia; PETRONGOLO, Barbara. The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries. *The Journal of Economic Perspectives* [online]. 2017, roč. 31, č. 1, s. 205–230 [cit. 2022-12-25]. ISSN 08953309. Dostupné z: <http://www.jstor.org/stable/44133957>.
17. PAILHÉ, Ariane. The convergence of second-generation immigrants' fertility patterns in France: The role of sociocultural distance between parents' and host country. *Demographic Research* [online]. 2017, roč. 36, s. 1361–1398 [cit. 2022-12-05]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26332168>.
18. DORBRITZ, Jürgen. Germany: Family diversity with low actual and desired fertility. *Demographic Research* [online]. 2008, roč. 19, s. 557–598 [cit. 2022-12-06]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26349258>.
19. *Childcare arrangements in the EU*. Lucemburk: Eurostat, 2022. Dostupné také z: <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220504-2>.
20. KNIGHT, Ben. *Child poverty still rising in Germany*. Bonn: Deutsche Welle, 2017. Dostupné také z: <https://www.dw.com/en/child-poverty-still-rising-in-germany-official-report-shows/a-40532068>.
21. GOLDSTEIN, Joshua R.; KREYENFELD, Michaela. Has East Germany Overtaken West Germany? Recent Trends in Order-Specific Fertility. *Population and Development Review* [online]. 2011, roč. 37, č. 3, s. 453–472 [cit. 2022-12-25]. ISSN 00987921, ISSN 17284457. Dostupné z: <http://www.jstor.org/stable/23036051>.
22. ARPINO, Bruno; ESPING-ANDERSEN, Gøsta; PESSIN, Léa. How Do Changes in Gender Role Attitudes Towards Female Employment Influence Fertility? A Macro-Level Analysis. *European Sociological Review*. 2015, roč. 31, č. 3, s. 370–382. ISSN 0266-7215. Dostupné z DOI: 10.1093/esr/jcv002.
23. FOKKEMA, Tineke; VALK, Helga de; BEER, Joop de; DUIN, Coen van. The Netherlands: Childbearing within the context of a “Poldermodel” society. *Demographic Research* [online]. 2008, roč. 19, s. 743–794 [cit. 2022-12-06]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26349262>.
24. MILLS, Melinda C. The Dutch Fertility Paradox: How the Netherlands Has Managed to Sustain Near-Replacement Fertility. In: *Low and Lower Fertility: Variations across Developed Countries*. Ed. RINDFUSS, Ronald R.; CHOE, Minja Kim. Cham: Springer International Publishing, 2015, s. 161–188. ISBN 978-3-319-21482-5. Dostupné z DOI: 10.1007/978-3-319-21482-5\_9.

25. BAILEY, Martha J; HERSHBEIN, Brad J. US Fertility Rates and Childbearing in American Economic History, 1800–2010. *The Oxford Handbook of American Economic History*. 2018, roč. 1, s. 75–100. ISBN 9780190882617. Dostupné z DOI: [10.1093/oxfordhb/9780190882617.013.21](https://doi.org/10.1093/oxfordhb/9780190882617.013.21).
26. MORGAN, S. Philip; RACKIN, Heather. The Correspondence Between Fertility Intentions and Behavior in the United States. *Population and Development Review* [online]. 2010, roč. 36, č. 1, s. 91–118 [cit. 2022-12-06]. ISSN 00987921, ISSN 17284457. Dostupné z: <http://www.jstor.org/stable/25699038>.
27. RAY, Colleen M.; HARCEY, Sela R.; GREIL, Arthur L.; TIEMEYER, Stacy; MCQUILLAN, Julia. Stability and change in personal fertility ideals among US women in heterosexual relationships. *Demographic Research* [online]. 2018, roč. 39, s. 459–486 [cit. 2022-12-06]. ISSN 14359871, ISSN 23637064. Dostupné z: <https://www.jstor.org/stable/26585337>.
28. DANIELSEN, Albert L. OPEC. In: *Encyclopedia Britannica* [online]. [B.r.] [cit. 2022-12-08]. Dostupné z: <https://www.britannica.com/topic/OPEC>.
29. KETTELL, Steven. oil crisis. In: *Encyclopedia Britannica* [online]. [B.r.] [cit. 2022-12-08]. Dostupné z: <https://www.britannica.com/topic/oil-crisis>.
30. PAYNE, Stanley G. Francisco Franco. In: *Encyclopedia Britannica* [online]. [B.r.] [cit. 2022-12-08]. Dostupné z: <https://www.britannica.com/biography/Francisco-Franco>.
31. HOORENS, Stijn; CLIFT, Jack; STAETSKY, Laura; JANTA, Barbara; DIEPEVEEN, Stephanie; JONES, Molly Morgan; GRANT, Jonathan. Case study: Spain. In: *Low fertility in Europe: Is there still reason to worry?* [online]. RAND Corporation, 2011, s. 45–52 [cit. 2022-12-08]. Dostupné z: <http://www.jstor.org/stable/10.7249/mg1080re.13>.
32. RAIFMAN, Sarah; PURI, Mahesh; ARCARA, Jennet; DIAMOND-SMITH, Nadia. Is there an association between fertility and domestic violence in Nepal? *AJOG Global Reports*. 2021, roč. 1, č. 2, s. 100011. ISSN 2666-5778. Dostupné z DOI: <https://doi.org/10.1016/j.xagr.2021.100011>.
33. MATOBA, Nana; REINA, Margarita; PRACHAND, Nikhil; DAVIS, Matthew M; COLLINS, James W. Neighborhood gun violence and birth outcomes in Chicago. *Maternal and child health journal*. 2019, roč. 23, č. 9, s. 1251–1259. Dostupné z DOI: <https://doi.org/10.1007/s10995-019-02765-w>.
34. SÁNCHEZ-PÁEZ, David Antonio; ORTEGA, José Antonio. Adolescent contraceptive use and its effects on fertility. *Demographic Research* [online]. 2018, roč. 38, s. 1359–1388 [cit. 2022-12-08]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26457079>.
35. FRANK, Odile. The Demand for Fertility Control in sub-Saharan Africa. *Studies in Family Planning* [online]. 1987, roč. 18, č. 4, s. 181–201 [cit. 2022-12-27]. ISSN 00393665, ISSN 17284465. Dostupné z: <http://www.jstor.org/stable/1966870>.
36. MONTESINOS, Jesus. Does sporting success affect birth rate? *BMJ: British Medical Journal* [online]. 2013, roč. 347, č. 7938, s. 8–10 [cit. 2022-12-09]. ISSN 09598138, ISSN 17561833. Dostupné z: <http://www.jstor.org/stable/43507755>.
37. KULU, Hill; VIKAT, Andres. Fertility differences by housing type: The effect of housing conditions or of selective moves? *Demographic Research* [online]. 2007, roč. 17, s. 775–802 [cit. 2022-12-09]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26347971>.
38. ILNITSKY, Sara; VAN UUM, Stan. Marijuana and fertility. *CMAJ*. 2019, roč. 191, č. 23, E638–E638. Dostupné z DOI: <https://doi.org/10.1503/cmaj.181577>.

39. JOESOEUF, M Riduan; BERAL, Valerie; ARAL, Sevgi O; ROLFS, Robert T; CRAMER, Daniel W. Fertility and use of cigarettes, alcohol, marijuana, and cocaine. *Annals of Epidemiology*. 1993, roč. 3, č. 6, s. 592–594. Dostupné z DOI: [https://doi.org/10.1016/1047-2797\(93\)90080-n](https://doi.org/10.1016/1047-2797(93)90080-n).
40. FRONCZAK, Carolyn M; KIM, Edward D; BARQAWI, Al B. The insults of illicit drug use on male fertility. *Journal of andrology*. 2012, roč. 33, č. 4, s. 515–528. Dostupné z DOI: <https://doi.org/10.2164/jandrol.110.011874>.
41. GHATULE, Arjun P; GHATULE, Archana A; MUNDHE, Shivaji D. Effect of Heavy Workload on Fertility of IT-Professionals: a Study with Respect to Indian IT-Sector. In: *2nd International Conference on Management and Artificial Intelligence IPEDR*. 2012, sv. 35. Dostupné také z: <http://www.ipedr.com/vol135/010-ICMAI2012-E00028.pdf>.
42. HJOLLUND, Niels Henrik I.; BONDE, Jens Peter E.; HENRIKSEN, Tine Brink; GIWERCMAN, Aleksander; OLSEN, Jørn; TEAM, The Danish First Pregnancy Planner Study. Job Strain and Male Fertility. *Epidemiology* [online]. 2004, roč. 15, č. 1, s. 114–117 [cit. 2022-12-10]. ISSN 10443983. Dostupné z: <http://www.jstor.org/stable/20485848>.
43. BONITSIS, Theologos Homer; GEITHMAN, David T. Does Income Affect Fertility or Does Fertility Affect Income? *Eastern Economic Journal* [online]. 1987, roč. 13, č. 4, s. 447–451 [cit. 2022-12-10]. ISSN 00945056, ISSN 19394632. Dostupné z: <http://www.jstor.org/stable/40325160>.
44. KEARNEY, Melissa S.; LEVINE, Phillip B. Media Influences on Social Outcomes: The Impact of MTV's "16 and Pregnant" on Teen Childbearing. *The American Economic Review* [online]. 2015, roč. 105, č. 12, s. 3597–3632 [cit. 2022-12-10]. ISSN 00028282. Dostupné z: <http://www.jstor.org/stable/43821387>.
45. ENCYCLOPAEDIA BRITANNICA, T. Editors of. social mobility. In: *Encyclopedia Britannica* [online]. [B.r.] [cit. 2022-12-10]. Dostupné z: <https://www.britannica.com/topic/social-mobility>.
46. ZUANNA, Gianpiero Dalla. Social mobility and fertility. *Demographic Research* [online]. 2007, roč. 17, s. 441–464 [cit. 2022-12-10]. ISSN 14359871, ISSN 23637064. Dostupné z: <http://www.jstor.org/stable/26347960>.
47. MCQUILLAN, Kevin. When Does Religion Influence Fertility? *Population and Development Review* [online]. 2004, roč. 30, č. 1, s. 25–56 [cit. 2022-12-10]. ISSN 00987921, ISSN 17284457. Dostupné z: <http://www.jstor.org/stable/3401497>.
48. MARUTHAPPU, M; NG, KYB; WILLIAMS, C; ATUN, R; AGRAWAL, P; ZELTNER, T. The association between government healthcare spending and maternal mortality in the European Union, 1981–2010: a retrospective study. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2015, roč. 122, č. 9, s. 1216–1224. Dostupné z DOI: <https://doi.org/10.1111/1471-0528.13205>.
49. EL OSTA, Rabih; ALMONT, Thierry; DILIGENT, Catherine; HUBERT, Nicolas; ESCHWÈGE, Pascal; HUBERT, Jacques. Anabolic steroids abuse and male infertility. *Basic and clinical andrology* [online]. 2016, roč. 26, č. 1, s. 1–8 [cit. 2022-12-17]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4744441/>.
50. *Časté dotazy ohledně dat Trendů Google*. Mountain View: Google, 2022. Dostupné také z: [https://support.google.com/trends/answer/4365533?hl=cs%5C&ref\\_topic=6248052](https://support.google.com/trends/answer/4365533?hl=cs%5C&ref_topic=6248052).
51. *Google Ngram Viewer*. Mountain View: Google, 2022. Dostupné také z: <https://books.google.com/ngrams/info%5C#>.
52. MCARTHUR, Tom. *Corpus*. Oxford University Press, 2003. ISBN 9780191727023. Dostupné z DOI: 10.1093/acref/9780192800619.013.0319.

53. PREIS, Tobias; MOAT, Helen Susannah; STANLEY, H Eugene; BISHOP, Steven R. Quantifying the advantage of looking forward. *Scientific reports* [online]. 2012, roč. 2, č. 1, s. 1–2 [cit. 2022-12-18]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3320057/>.
54. OJALA, Jussi; ZAGHENI, Emilio; BILLARI, Francesco; WEBER, Ingmar. Fertility and Its Meaning: Evidence from Search Behavior. *Proceedings of the International AAAI Conference on Web and Social Media* [online]. 2017, roč. 11, č. 1, s. 640–643 [cit. 2022-12-18]. Dostupné z DOI: 10.1609/icwsm.v11i1.14915.
55. LUNA, Javier Canales. *Top programming languages for data scientists in 2022*. New York: Datacamp, 2022. Dostupné také z: <https://www.datacamp.com/blog/top-programming-languages-for-data-scientists-in-2022>.
56. GALLINELLI, Nicholas. *10 Best Data Science Programming Languages*. Denver: Flatiron School, 2021. Dostupné také z: <https://flatironschool.com/blog/data-science-programming-languages/>.
57. SHARMA, Rohit. *Top 6 Data Science Programming Languages 2023 [Hand-Picked]*. Bangalore: Upgrad, 2022. Dostupné také z: <https://www.upgrad.com/blog/data-science-programming-languages/>.
58. *The Python Wiki*. Wilmington: Python Software Foundation, 2018. Dostupné také z: <https://wiki.python.org/moin/FrontPage>.
59. MATTHES, Eric. *Python crash course: a hands-on, project-based introduction to programming*. 2nd edition. San Francisco: No Starch Press, 2019. ISBN 978-1-59327-928-8.
60. *Streamlit documentation*. San Francisco: Streamlit, 2022. Dostupné také z: <https://docs.streamlit.io/>.
61. DOWNEY, A.B. *Think stats: pravděpodobnost a statistika pro programátory: verze 1.6.0*. Flow, 2014. ISBN 9788090548046. Dostupné také z: <https://books.google.cz/books?id=ZXOGrgEACAAJ>.
62. KENTON, Will. *What Is the Pearson Coefficient? Definition, Benefits, and History*. New York: Investopedia, 2022. Dostupné také z: <https://www.investopedia.com/terms/p/pearsoncoefficient.asp>.
63. FERNANDO, Jason. *The Correlation Coefficient: What It Is, What It Tells Investors*. New York: Investopedia, 2021. Dostupné také z: <https://www.investopedia.com/terms/c/correlationcoefficient.asp>.
64. BEERS, Brian. *P-Value: What It Is, How to Calculate It, and Why It Matters*. New York City: Investopedia, 2022. Dostupné také z: <https://www.investopedia.com/terms/p/p-value.asp>.
65. SINGH, Abhigyan. *Understanding Polynomial Regression!!!* San Francisco: Medium, 2020. Dostupné také z: <https://medium.com/analytics-vidhya/understanding-polynomial-regression-5ac25b970e18>.
66. *Git*. Brooklyn: Git, 2022. Dostupné také z: <https://git-scm.com/>.
67. *Best Version Control Systems*. Chicago: G2.com, c2022. Dostupné také z: <https://www.g2.com/categories/version-control-systems>.
68. *15 BEST Version Control Software (Source Code Management Tools)*. India: Software Testing Help, 2022. Dostupné také z: <https://www.softwaretestinghelp.com/version-control-software/>.
69. ADEWUSI, David. *Everything You Need to Know about Google Translate*. Toronto: Scientific Editing, 2021. Dostupné také z: <https://www.scientific-editing.info/blog/everything-you-need-to-know-about-google-translate>.

70. *DeepL*. Cologne: DeepL, 2022. Dostupné také z: <https://www.deepl.com/whydeepl/>.
71. NEWCOMER, Colin. *DeepL vs Google Translate: Which Is Better? Plus How to Use Them (2022)*. Timisoara: Translate press, 2022. Dostupné také z: <https://translatepress.com/deepl-vs-google-translate-comparison/>.



# Obsah přiloženého média

	readme.txt .....	stručný popis obsahu média
	src	
	impl .....	zdrojové kódy implementace
	thesis .....	zdrojová forma práce ve formátu $\text{\LaTeX}$
	text .....	text práce
	thesis.pdf .....	text práce ve formátu PDF