

Angewandte Generative KI in den (digitalen) Geisteswissenschaften

Datenvisualisierung und Data Stories: Analyse biographischer Daten von ins Exil geflüchteten Übersetzer*innen mit Python und Jupyter Notebook

Workshop-Materialien: <http://chpollin.github.io/GM-DH>

Christopher Pollin

<https://chpollin.github.io>

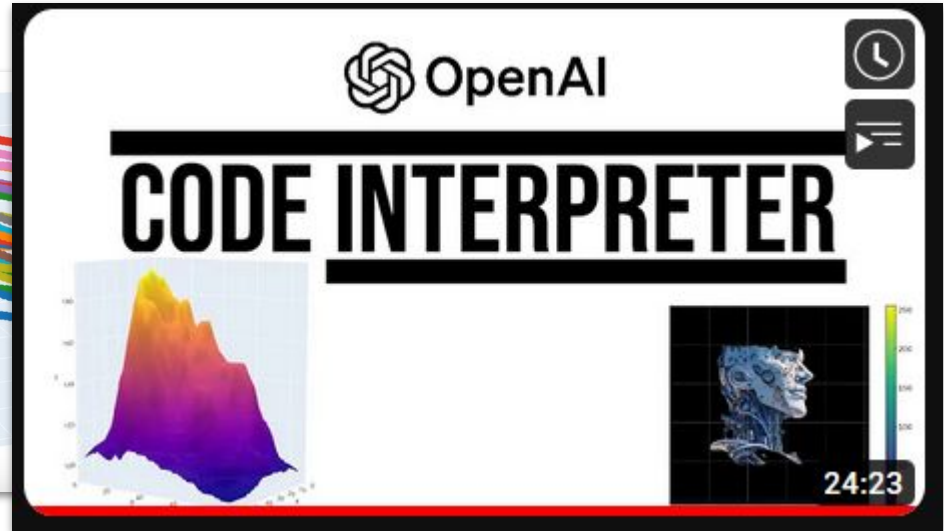
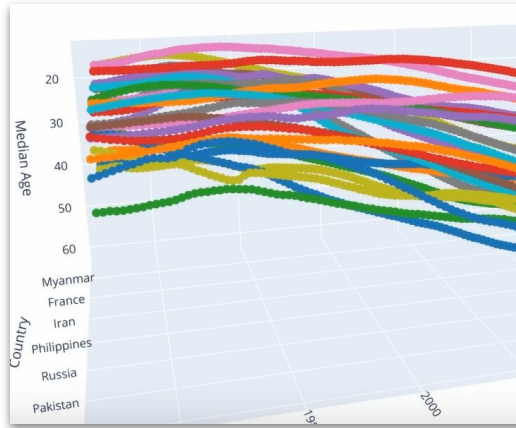
Zentrum für Informationsmodellierung, Graz

<https://informationsmodellierung.uni-graz.at/de/institut>

Digital Humanities Craft OG

www.dhcraft.org

News



AI Explained. GPT-4 Got Upgraded - Code Interpreter (ft. Image Editing, MP4s, 3D Plots, Data Analytics and more!)

https://www.youtube.com/watch?v=O8GUH0_hRM

Workshop begleitende YouTube-Playlist: Generative Models (AI) and Digital Humanities

<https://youtube.com/playlist?list=PLaHADNRco7n3GKVUD8mAc36pXQ5pnJQVL>

Find 5 unexpected, non-obvious insights from this data and offer plausible explanations for them. For the most interesting observation provide a compelling and clear visualisation.

Alright, let's start by performing an exploratory data analysis to find unusual or interesting insights. Here are the steps we're going to take:

1. Analyze the overall trends in the data.
2. Look for countries that have seen the most significant changes in median age over the years.
3. Identify countries that have consistently had the highest or lowest median ages.
4. Search for any abrupt changes in the median age which could indicate a significant historical event.
5. Examine the projected future trends and see if any countries are expected to see a significant change in median age.

After these analyses, I'll be able to suggest possible explanations for the insights found. Let's get started!

6. Data Analysis

4:59 / 24:22 • Data Analysis >

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT May 4/23 version

GPT 4 Got Upgraded - Code Interpreter (ft. Image Editing, MP4s, 3D Plots, Data Analytics and more!)

AI Explained. GPT 4 Got Upgraded - Code Interpreter (ft. Image Editing, MP4s, 3D Plots, Data Analytics and more!) <https://youtu.be/O8GUH0 htRM?t=299>

News

Seit 29.06:

DeepLearning.AI. **Generative AI with Large Language Models.**

<https://www.coursera.org/learn/generative-ai-with-llms>

- Generative AI use cases, project lifecycle, and model pre-training
- Fine-tuning and evaluating large language models
- Reinforcement learning and LLM-powered applications

Ressourcen

Conversations:

- **Exil:Trans Data Visualization 1. GPT-4, ChatGPT May 24 Version. 20.05.2023.**
<https://chat.openai.com/share/3721faec-d840-42ac-bd63-7d80f8a5c6e1>
- **Exil:Trans Data Visualization 2. GPT-4, ChatGPT May 24 Version. 23.06.2023.**
<https://chat.openai.com/share/4a9f07e1-3fab-4ecb-b59a-6d1faa5d0a62>.
- **Exil:Trans Data Visualization 3. GPT-4, ChatGPT May 24 Version. 23.06.2023.**
<https://chat.openai.com/share/2efe0d28-9e15-4225-991b-e356cdba865c>.
- **Exil:Trans Data Visualization 4. GPT-4, ChatGPT May 24 Version. 23.06.2023.**
<https://chat.openai.com/share/6f109a9b-615d-459c-a970-9f2e61c8fad5>.

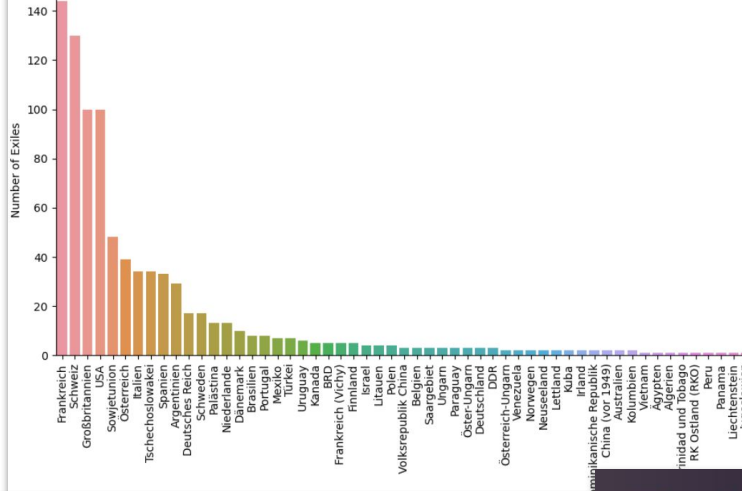
YouTube und weitere Tutorials

- AI Explained, <https://www.youtube.com/@ai-explained->
- Matthew Berman, https://www.youtube.com/@matthew_berman
- Yannic Kilcher, <https://www.youtube.com/@YannicKilcher>
- ChatGPT Prompt Engineering for Developers, <https://learn.deeplearning.ai/chatgpt-prompt-eng>
- Hötter, Johannes and Warmuth, Christian. OpenHPI. ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? <https://open.hpi.de/courses/kizukunft2023/items/3jMwelGWPssDCNQw63XXsn>
- **Chad Skelton. How to do data analysis using AI with ChatGPT and the Noteable plugin. 2023.**
<https://www.youtube.com/watch?v=A1ualvzqJoo>. Blogpost:
<http://www.chadskelton.com/2023/05/using-chatgpt-and-noteable-plugin-for.html>

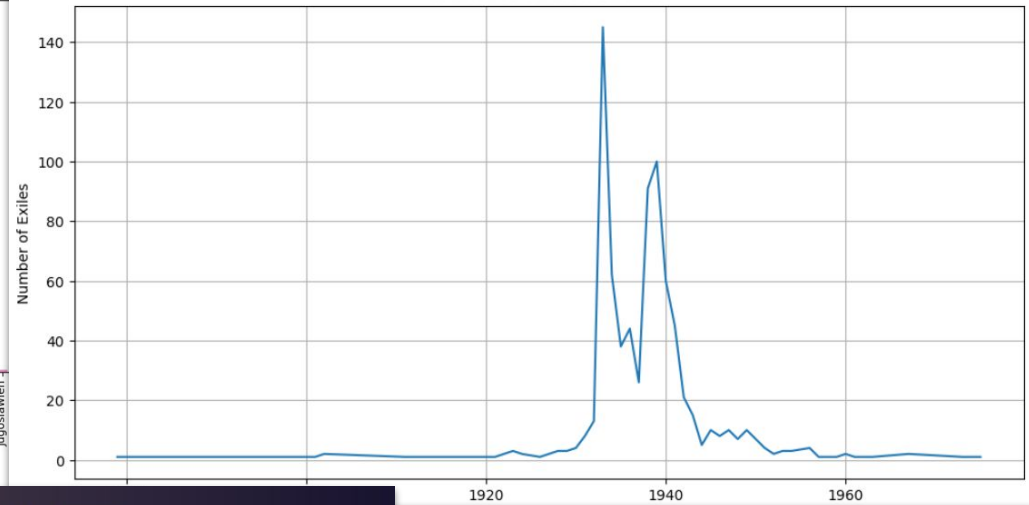
Ressourcen

- Baktash, Jawid Ahmad, and Mursal Dawodi. 2023. 'Gpt-4: A Review on Advancements and Opportunities in Natural Language Processing'. arXiv. <https://doi.org/10.48550/arXiv.2305.03195>.
- Møller, Anders Giovanni, Jacob Aarup Dalsgaard, Arianna Pera, and Luca Maria Aiello. 2023. 'Is a Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks'. arXiv. <https://doi.org/10.48550/arXiv.2304.13861>.
- Mukherjee, Subhabrata, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. 2023. 'Orca: Progressive Learning from Complex Explanation Traces of GPT-4'. arXiv. <https://doi.org/10.48550/arXiv.2306.02707>.
- Roberts, Jonathan, Timo Lüddecke, Sowmen Das, Kai Han, and Samuel Albanie. 2023. 'GPT4GEO: How a Language Model Sees the World's Geography'. arXiv. <http://arxiv.org/abs/2306.00020>.
- Watkins, Ryan. 'Guidance for Researchers and Peer-Reviewers on the Ethical Use of Large Language Models (LLM) in Scientific Research Workflows'. OSF Preprints, 21 April 2023. <https://doi.org/10.31219/osf.io/6uh8p>.
- Yao, Shunyu, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 'Tree of Thoughts: Deliberate Problem Solving with Large Language Models'. arXiv, 17 May 2023. <https://arxiv.org/abs/2305.10601>.
- Gou, Zhibin, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Nan Duan, and Weizhu Chen. 'CRITIC: Large Language Models Can Self-Correct with Tool-Interactive Critiquing'. arXiv, 19 May 2023. <https://doi.org/10.48550/arXiv.2305.11738>.
- Zhang, Muru, Ofir Press, William Merrill, Alisa Liu, and Noah A. Smith. 'How Language Model Hallucinations Can Snowball'. arXiv, 22 May 2023. <https://doi.org/10.48550/arXiv.2305.13534>.
- Why AI Is Incredibly Smart and Shockingly Stupid | Yejin Choi | TED. <https://www.youtube.com/watch?v=SvBR0OGT5VI>.
- Orca: The Model Few Saw Coming. AI Explained. https://www.youtube.com/watch?v=Dt_UNg7Mchg
- Mukherjee, Subhabrata, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. 'Orca: Progressive Learning from Complex Explanation Traces of GPT-4'. arXiv, 5 June 2023. <https://doi.org/10.48550/arXiv.2306.02707>.

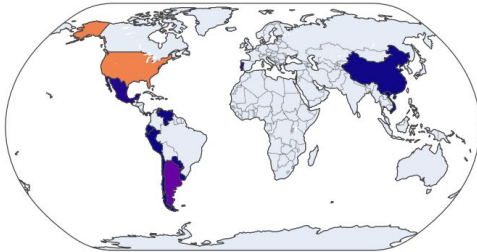
Frequency of Exile Destinations



Number of Exiles Over Time



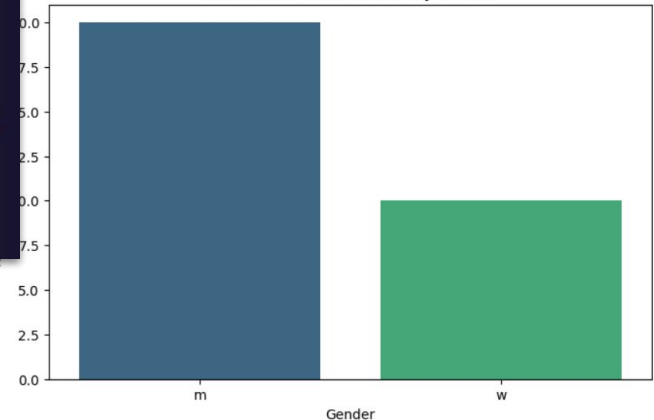
Number of Exiles per Country



Exil:Trans

Forschungsprojekt und Datenbank
zu Leben und Arbeit verfolgter
Übersetzer und Übersetzerinnen

Distribution of Individuals by Gender



Ein Beispiel: Hermann Adler



Translatorisches Profil

Genres: Literatur

Zielsprachen: Deutsch

Ausgangssprachen:

- Hebräisch
- Jiddisch

Tätigkeitsfelder:

- Literatur und Verlagswesen
- Publizistik und Medien
- Universität, Wissenschaft, Bildung
- Übersetzer/in

Tätigkeiten:

- Psychologe/Psychologin
- Publizist/in
- Schriftsteller/in
- Übersetzer/in



Lebensdaten

Nachname: Adler

Vorname: Hermann

Geschlecht: männlich

Geburtsdatum: 02.10.1911

Sterbedatum: 18.02.2001

Geburtsort: Diószeg

Sterbeort: Basel



Migration und Exil

Letzter Wirkungsort vor dem Exil:

Unbekannt

Exil:

Jahr: 1934

Ort: Prag

Land: Tschechoslowakei

Jahr: 1939

Ort: Krakau

Land: Polen

Jahr: 1940

Ort: Vilnius

Land: Litauen

Jahr: 1943

Ort: Warschau

Land: Polen

Remigration:

Jahr: 1945

Ort: Basel

Land: Schweiz

Namen und Lebensdaten			Varianten werden hintereinander geschrieben, durch Komma getrennt.	Dropdown	TT.MM.JJJJ bzw. ??.??.????	Dropdown	TT.MM.JJJJ bzw. ??.??.????	Dropdown		
name_id	nachname	vorname	pseudonyme / namensvarianten	geschlecht	geburtsdatum	geburtsort	geburtsland	sterbedatum	sterbeort	sterbeland
1	Walden	Ellen	Ellen Bork	w	??.??.1912	Berlin	Deutsches Reich	??.??.????	München	BRD
2	Adler-Herzmark	Jenny	Jenny Adler (Ehename), Jenny	w	??.??.1877	Riga	Russisches Kaiser	??.??. 1950	?	USA
3	Biefang	Johann	Hans Biefang, Julius Biefang	m	02.08.1893	Moers	Deutsches Reich	??.??.1938	?	Spanien
4	Feld	Friedrich	Friedrich Rosenfeld	m	12/5/1902	Wien	Öster-Ungarn	12/27/1987	Bexhill	Großbritannien
5	Frisch	Justinian	friliko	m	19.07.1879	Kritzendorf	Öster-Ungarn	??.??.1949	Cambridge	Großbritannien
6	Waldinger	Ernst		m	16.10.1896	Wien-Neulerchenfel	Öster-Ungarn	2/1/1970	New York	USA
7	Thelen	Albert Vigoleis	Leopold Fabrizio (Pseudonym),	m	9/28/1903	Süchteln	Deutsches Reich	4/9/1989	Dülken	BRD
8	Adler	Raissa	Raissa Timofejewna Epstein	w	09.11.1873	Moskau	Russisches Kaiser	4/21/1962	New York	USA
9	Lohan	Robert		m	02.04.1884	Bielitz	Öster-Ungarn	6/18/1953	Oneonta	USA

```
names.csv
1 name_id,nachname,vorname,pseudonyme / namensvarianten,geschlecht,geburtsdatum,geburtsort,geburtsland,sterbedatum,sterbeort,sterbeland
2 1,Walden,Ellen,Ellen Bork,w,1912,Berlin,Deutsches Reich,,München,BRD
3 2,Adler-Herzmark,Jenny,"Jenny Adler (Ehename), Jenny Herzmark (Geburtsname)",w,1877,Riga,Russisches Kaiserreich,1950,,USA
4 3,Biefang,Johann,"Hans Biefang, Julius Biefang",m,02.08.1893,Moers,Deutsches Reich,1938,,Spanien
5 4,Rosenfeld,Friedrich,Friedrich Feld,m,05.12.1902,Wien,Österreich-Ungarn,12/27/1987,Bexhill,Großbritannien
6 5,Frisch,Justinian,friliko,m,19.07.1879,Kritzendorf,Österreich-Ungarn,1949,Cambridge,Großbritannien
7 6,Waldinger,Ernst,,m,16.10.1896,Wien-Neulerchenfeld,Österreich-Ungarn,2/1/1970,New York,USA
8 7,Thelen,Albert Vigoleis,"Leopold Fabrizio (Pseudonym), Albert Thelen, Albert Vigoleis Thelen-Bruckner",m,28.09.1903,Süchteln,Deutsches Reich
9 8,Adler,Raissa Timofejewna,Raissa Timofejewna Epstein (Geburtsname),w,09.11.1873,Moskau,Russisches Kaiserreich,4/21/1962,New York,USA
10 9,Lohan,Robert,,m,02.04.1884,Bielitz,Österreich-Ungarn,6/18/1953,Oneonta,USA
11 10,Zwillinger,Frank Gerhard,,m,29.11.1909,Wien,Österreich-Ungarn,11/22/1989,Garches,Frankreich
12 11,Zweig,Stefan,,m,28.11.1881,Wien,Österreich-Ungarn,2/23/1942,Petropolis,Brasilien
13 12,Zweig,Friderike Maria,"Friederike Maria Zweig, Friderike Maria Burger (Geburtsname), Friderike von Winternitz (in erster Ehe), Friederike
14 13,zur Mühlen,Hermynia,"Maria Berg, Franziska Marisa Rautenberg, Lawrence H. Desberry, Traugott Lehmann (Pseudonyme)",w,12.12.1883,Wien,Öst
```

Exil- und Lebensstationen

			Dropdown			Dropdown			Dropdown			Dropdown	
name_id	nachname	vorname	exilort1	exilland1	jahr1	exilort2	exilland2	jahr2	exilort3	exilland3	jahr3	exilort4	exilland4
1	Walden	Ellen	Moskau	Sowjetunion	1932								
2	Adler-Herzmark	Jenny		Frankreich	1939		USA	1942					
3	Biefang	Johann	Donbass	Sowjetunion	1930	Moskau	Sowjetunion	1932	?	Spanien	1937		
4	Feld	Friedrich	Wien	Öster-Ungarn	1902	Prag	Tschechoslow	1934	Großbritanni		1939		
5	Frisch	Justinian	Wien	Österreich	1879	Stockholm	Schweden	1939	Cambridge	Großbritannien	1948		
6	Waldinger	Ernst	Wien	Österreich	1938	Paris	Frankreich	1938	London	Großbritannien	1938	New York	USA
7	Thelen	Albert Vigoleis	Mallorca	Spanien	1931	Marseille	Frankreich	1936	Auressio	Schweiz	1936		Frankreich
8	Adler	Raissa	New York	USA	1935	Locarno	Schweiz	?	New York	USA	1940		

exil.csv

	name_id	nachname	vorname	exilort1	exilland1	jahr1	exilort2	exilland2	jahr2	exilort3	exilland3	jahr3	exilort4	exilland4	jahr4	exilort5	exilland5	jahr5
1	1	Walden	Ellen	Moskau	Sowjetunion	1932										Berlin	Deutsches Reich	1941
2	2	Adler-Herzmark	Jenny		Frankreich	1939		USA	1942									
3	3	Biefang	Johann	Donbass	Sowjetunion	1930	Moskau	Sowjetunion	1932		Spanien	1937						
4	4	Rosenfeld	Friedrich	Prag	Tschechoslowakei	1934		Großbritannien	1939									Wien, Österreich
5	5	Frisch	Justinian	Stockholm	Schweden	1939	Cambridge	Großbritannien	1948									Wien, Österreich
6	6	Waldinger	Ernst	Paris	Frankreich	1938	London	Großbritannien	1938	New York	USA	1938						
7	7	Thelen	Albert	Vigoleis	Marseille	Frankreich	1936	Auressio	Schweiz	1936		Frankreich	1939	Amarante	Portugal	1939		
8	8	Adler	Raissa	Timofejewna	New York	USA	1935	Locarno	Schweiz		New York	USA	1940					Wien, Österreich
9	9	Lohan	Robert	Reading	Großbritannien	1938	New York	USA	1939									Wien, Österreich
10	10	Zwillinger	Frank	Gerhard	Rom	Italien	1938	Saigon	Vietnam	1939	Paris	Frankreich	1946					

Berufe und Tätigkeiten

name_id	nachname	vorname	Dropdown				
			beruf1 / tätigkeit1	beruf2 / tätigkeit2	beruf3 / tätigkeit3	beruf4 / tätigkeit4	beruf5 / tätigkeit5
1	Walden	Ellen	Sekretär	Übersetzer			
2	Adler-Herzmark	Jenny	Arzt	Übersetzer	Vereinsfunktionär		
3	Biefang	Johann	Bergarbeiter	Parteifunktionär	Verlagsredakteur	Übersetzer	
4	Feld	Friedrich	Schriftsteller	Zeitungsmitarbeite	Journalist	Redakteur	Übersetzer
5	Frisch	Justinian	Buchdrucker	Unternehmer	Verleger	Grafiker	Übersetzer
6	Waldinger	Ernst	Schriftsteller	Übersetzer	Verlagsredakteur	Zeitungsmitarbeiter	Universitätsprofessor
7	Thelen	Albert Vigoleis	Schriftsteller	Übersetzer	Lyriker	Arbeiter	Sekretär
8	Adler	Raissa	Schriftsteller	Aktivist, politischer	Psychologe	Übersetzer	
9	Lohan	Robert	Dramaturg	Schriftsteller	Verleger	Universitätsprofessor	Übersetzer
10	Zwillinger	Frank Gerhard	Schriftsteller	Geschäftsmann	Übersetzer		

occupation.csv

```

1 name_id,nachname,vorname,beruf1 / tätigkeit1,beruf2 / tätigkeit2,beruf3 / tätigkeit3,beruf4 / tätigkeit4,beruf5 / tätigkeit5,beruf6 / tätig
2 1,Walden,Ellen,Sekretär/in,Übersetzer/in,,,,,,,,,
3 2,Adler-Herzmark,Jenny,Arzt/Ärztin,Übersetzer/in,Vereinsfunktionär/in,,,,,,,,,
4 3,Biefang,Johann,Bergarbeiter/in,Parteifunktionär/in,Verlagsmitarbeiter/in,Übersetzer/in,,,,,,,,,
5 4,Rosenfeld,Friedrich,Schriftsteller/in,Zeitungsmitarbeiter/in,Journalist/in,Redakteur/in,Übersetzer/in,Lektor/in,,,
6 5,Frisch,Justinian,Buchdrucker/in,Unternehmer/in,Verleger/in,Grafiker/in,Übersetzer/in,,,,,
7 6,Waldinger,Ernst,Schriftsteller/in,Übersetzer/in,Verlagsmitarbeiter/in,Zeitungsmitarbeiter/in,Universitätsprofessor/in,Bibliothekar/in,,,
8 7,Thelen,Albert Vigoleis,Schriftsteller/in,Übersetzer/in,Lyriker/in,Arbeiter/in,Sekretär/in,Fremdsprachenkorrespondent/in,,,,
9 8,Adler,Raissa Timofejewna,Schriftsteller/in,"Aktivist/in, politische/r",Psychologe/Psychologin,Übersetzer/in,,,,,
10 9,Lohan,Robert,Dramaturg/in,Schriftsteller/in,Verleger/in,Universitätsprofessor/in,Übersetzer/in,Philologe/Philologin,,,,
11 10,Zwillinger,Frank Gerhard,Schriftsteller/in,Unternehmer/in,Übersetzer/in,,,,,,,,,

```

Prompt Template: Hinzufügen größerer Dateien zu den Gesprächen mit GPT-4: <https://chatgpt-prompt-splitter.jldiaz.dev>

In a first step I will give you the data as csv. The total length of the content that I want to send you is too large to send in only one piece.

For sending you that content, I will follow this rule:

```
[START CSV "titleOfCSV.csv" 1/4]
```

...

```
[END CSV "titleOfCSV.csv" 1/4]
```

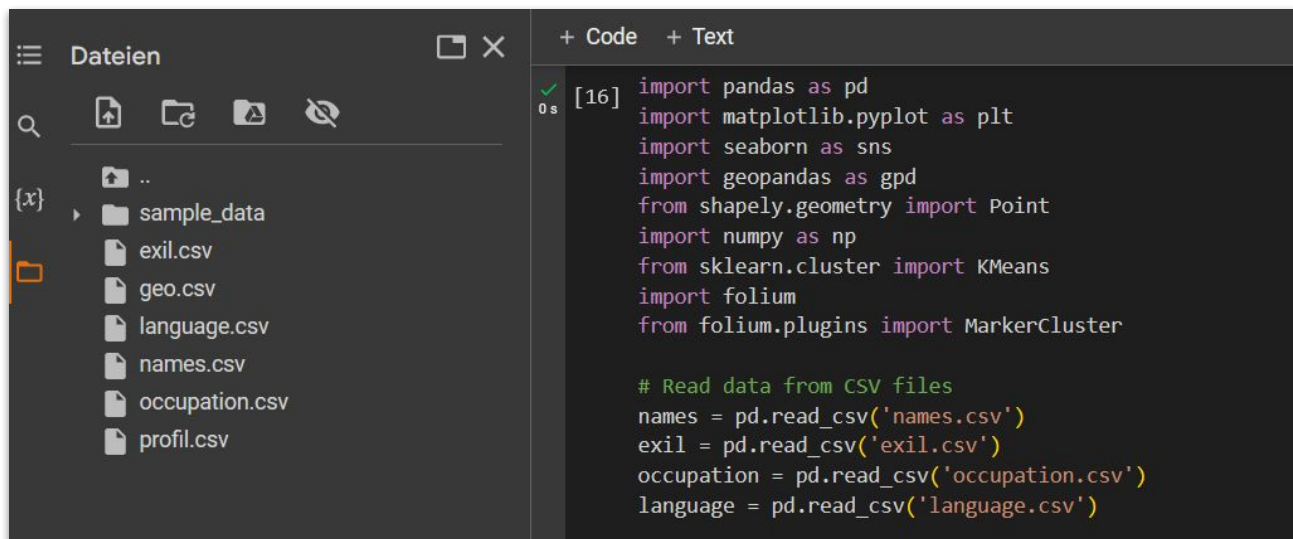
Then you just answer: "Received part 1/4"

And when I tell you "ALL PARTS SENT", then you can continue.

Oder einfach mit einer kleinen Stichprobe von Daten arbeiten und nicht mit allen: Stichwort Aufmerksamkeit von GPT-4!

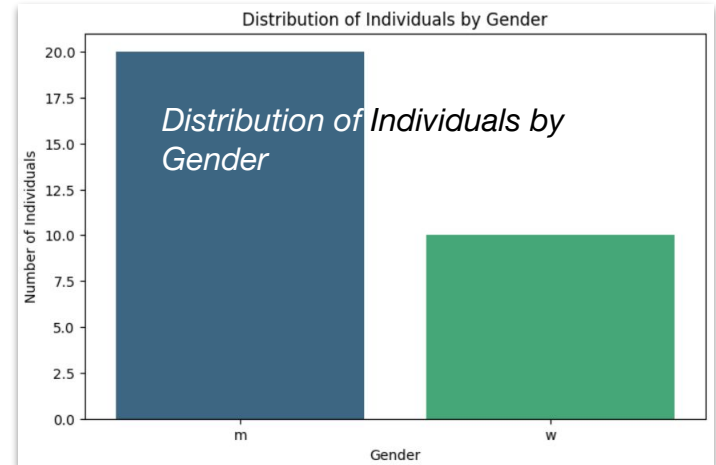
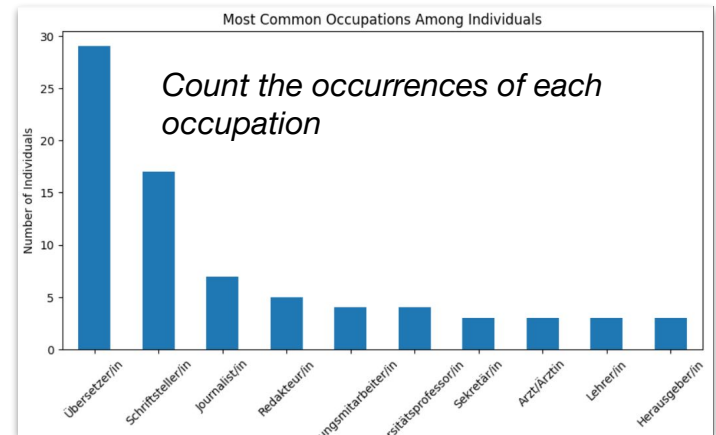
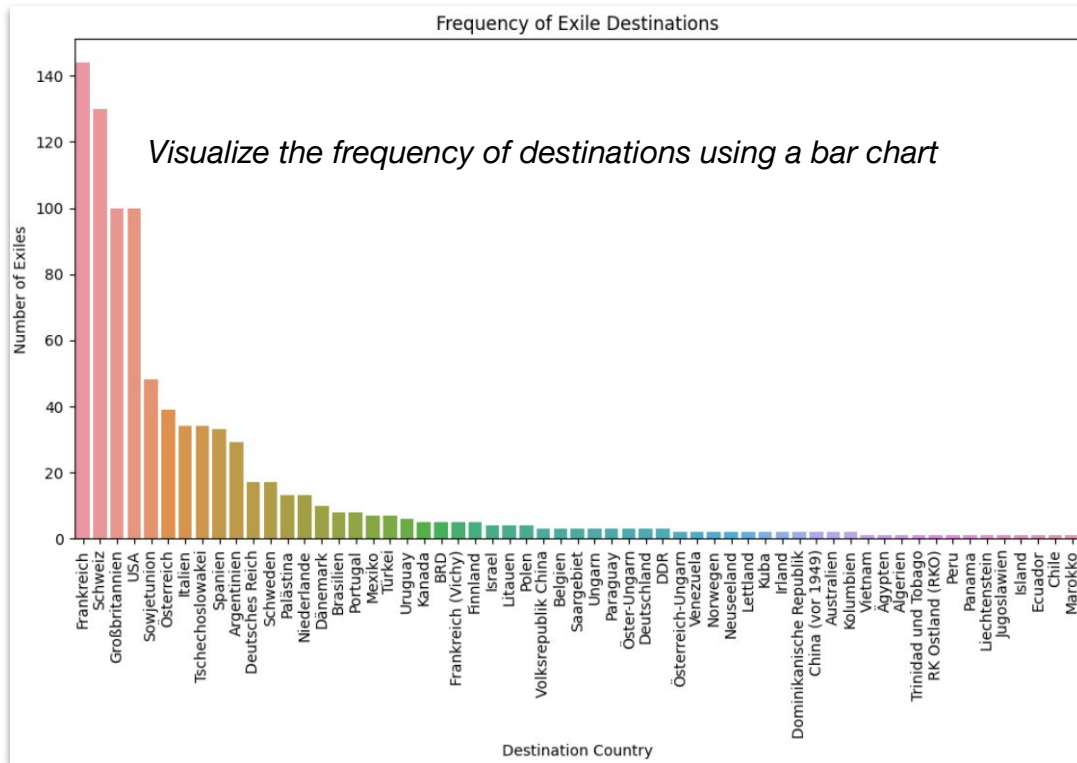
GPT-unterstützte Umsetzung eines Jupyter Notebook (Google Colab)

- [GPT-PromptEngineering-DataStory-ExilTrans-1.ipynb](#)
- [GPT-PromptEngineering-DataStory-ExilTrans-2.ipynb](#)
- [CSV Data snippets](#)



The screenshot displays a Jupyter Notebook environment. On the left, a file explorer titled 'Dateien' shows a directory structure with a folder named 'sample_data' and several CSV files: 'exil.csv', 'geo.csv', 'language.csv', 'names.csv', 'occupation.csv', and 'profil.csv'. The 'names.csv' file is selected. On the right, the code editor shows a Python code cell [16] that imports various libraries and reads data from the CSV files.

```
+ Code + Text  
✓ [16] 0 s  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
import geopandas as gpd  
from shapely.geometry import Point  
import numpy as np  
from sklearn.cluster import KMeans  
import folium  
from folium.plugins import MarkerCluster  
  
# Read data from CSV files  
names = pd.read_csv('names.csv')  
exil = pd.read_csv('exil.csv')  
occupation = pd.read_csv('occupation.csv')  
language = pd.read_csv('language.csv')
```



**Einfache Visualisierungen
sind leicht zu realisieren!**

Die beiden Beispiele rechts wurden mit einer Stichprobe von 30 Einträgen erstellt. Das linke Beispiel wurde mit dem gesamten Datensatz erstellt.

Initial Prompt

You are always in the role of a data visualisation expert. I want to create data visualisations and data stories using a Jupyter notebook and Python.

First, I will give you some context about the Exil:Trans project and the 100 records in the dataset, and then I want you to ask me as many questions as you need to better understand the domain.

I always want you to do tasks step by step.

About the project:

Exil:Trans- Forschungsprojekt und Datenbank zu Leben und Arbeit verfolgter Übersetzer und Übersetzerinnen

{... und Textpassagen aus der Webseite zusammenkopiert.}



Thank you for providing the context about the Exil:Trans project. I understand that Exil:Trans is a research project and database dedicated to the lives and work of translators who had to flee into exile due to Nazi threats or persecution. The database contains biographical research data, including life data, translation profiles, and exile paths of these translators. The project aims to invite in-depth investigations into individual translators and their networks.

To better understand the domain and the dataset, I have a few questions:

1. What kind of data fields are available in the dataset? For example, does it include names, birth dates, countries of origin, countries of exile, translation works, etc.?
2. What is the time period covered by the dataset? Does it cover World War II, or does it span multiple time periods?
3. Are there any specific geographic regions that the dataset primarily cover European countries, or is it more global?
4. Does the dataset include information on the translators' translated to/from, the genres they focused on, or the impact with?
5. Are there any specific types of visualizations or analyses? For example, are you looking to analyze trends over time, geographical distribution of translators, etc.?
6. Is the dataset structured (e.g., in a tabular format like CSV) or in the form of text documents?
7. Do you have any specific goals or questions that you want to visualize and analysis?

Aufbau von Kontext: Frage - Antwort



1. I will give you snippets of the csv data. you will see all the relevant data fields.
2. The exile and life stages shown in the data begin around 1930 and end around 1958 with the remigration of individuals.
3. the focus is on europe, as individuals fled from europe, mainly the german-speaking area. however, north america, south america and asia are also among the countries of exile.
4. yes. the data includes information on exile locations, translator profile, languages, profession and activity.
5. Visualisations that show the movement of people in different ways to answer research questions are very interesting.
6. it is CSV. because it is humanities data, it is sometimes not complete and not easy to normalise.
7. I want to develop this together with you.
8. data is free to use and already published. no sensitive data.
9. i dont think so
- 10 static visualisation like images created in theJupyter notebook.



Thank you for the additional information. It's great to know that the dataset is in CSV format and contains information on exile locations, translator profiles, languages, professions, and activities. The time frame from around 1930 to 1958 and the focus on Europe with inclusion of North America, South America, and Asia is also helpful to know.

Before we proceed with creating visualizations, let's take the following steps:

1. **Inspect the Dataset:** I would like to take a look at a snippet of the CSV data to understand the structure and fields available. This will help in planning the visualizations.
2. **Identify Research Questions:** We should identify specific research questions that we want to answer through the visualizations. For example, we might want to understand the patterns of movement among the exiled translators, or analyze the distribution of translators across different countries.
3. **Create Visualizations:** Based on the research questions, we will create static visualizations using Python in a Jupyter Notebook. We can use libraries like Matplotlib, Seaborn, or Plotly for creating the visualizations.
4. **Interpret and Communicate Results:** Once the visualizations are created, we can interpret the results and communicate the insights.

Let's start with step 1. Could you please provide a snippet of the CSV data? You can either paste a few rows here or, if the file is available online, provide a link to it.

“Step by Step”- Struktur aufbauen



there a multiple csv: names.csv, exil.csv, occupation.csv, profile.csv, language.csv. all of them

names.csv:

```
name_id,nachname,vorname,pseudonyme / namensvarianten,geschlecht,geburtsdatum,geb
1,Walden,Ellen,Ellen Bork,w,1912,Berlin,Deutsches Reich,,München,BRD
2,Adler-Herzmark,Jenny,"Jenny Adler (Ehename), Jenny Herzmark (Geburtsname)",w,1877,Ri
3,Biefang,Johann,"Hans Biefang, Julius Biefang",m,02.08.1893,Moers,Deutsches Reich,1938,,S
4,Rosenfeld,Friedrich,Friedrich Feld,m,05.10.1900,Wien,Österreich,1938,10/1974,1997,Berlin
```

5,Frisch,Justinian,frili occupation.csv:

6,Waldinger,Ernst,,m,

7,Thelen,Albert Vigole

8,Adler,Raissa Timofe

9,Lohan,Robert,,m,02

10,Zwillinger,Frank Ge

11,Zweig,Stefan,,m,28

12,Zweig,Friderike Ma

13,zur Mühlen,Hermyn

14,Zuckmayer,Eduarc

15,Zuckerandl-Szep

16,Zernik-Bern,Cleme

17,Zorn,Fritz,,m,09.06

18,Zohn,Harry,,m,21.11

19,Zerner,Fritz,"Friedr

20,Zerner,Elisabeth F

21,Zehetner,Alois,,m,1

22,Zahorsky,Anton M

23,Zuckerandl,Marie

24,Wyant,Gordon Michael

25,Wronkow,George,Georg Wronkow,m,17.02.1905,Berlin,Deutsches Reich,12/15/1989,New Yo

name_id,nachname,vorname,beruf1 / tätigkeit1,beruf2 / tätigkeit2,beruf3 / tätigkeit3,beruf4 /

1,Walden,Ellen,Sekretär/in,Übersetzer/in,,,,,,,,

2,Adler-Herzmark,Jenny,Arzt/Ärztin,Übersetzer/in,Vereinsfunktionär/in,,,,,,,,

3,Biefang,Johann,Bergarbeiter/in,Parteifunktionär/in,Verlagsmitarbeiter/in,Übersetzer/in

4,Rosenfeld,Friedrich,Schriftsteller/in,Zeitungsmitarbeiter/in,Journalist/

5,Frisch,Justinian,Buchdrucker/in,Unternehmer/in,Verleger/in,Grafiker/in

6,Waldinger,Ernst,Schriftsteller/in,Übersetzer/in,Verlagsmitarbeiter/in,Z

7,Thelen,Albert Vigoleis,Schriftsteller/in,Übersetzer/in,Lyriker/in,Arbeite

8,Adler,Raissa Timofejewna,Schriftsteller/in,"Aktivist/in, politische/r",Ps

9,Lohan,Robert,Dramaturg/in,Schriftsteller/in,Verleger/in,Universitätspr

10,Zwillinger,Frank Gerhard,Schriftsteller/in,Unternehmer/in,Übersetzer/

11,Zweig,Stefan,Schriftsteller/in,Übersetzer/in,,,,,,,,

12,Zweig,Friderike Maria,Schriftsteller/in,Übersetzer/in,Lehrer/in,Journa

13,zur Mühlen,Hermynia,Schriftsteller/in,Übersetzer/in,,,,,,,,

14,Zuckmayer,Eduard,Lehrer/in,Musiker/in,Übersetzer/in,Komponist/in,,

15,Zuckerandl-Szeps,Bertha,Schriftsteller/in,Journalist/in,Übersetzer/in

16,Zernik-Bern,Clementine,Rechtsanwalt/-anwältin,Bibliothekar/in,Über

17,Zorn,Fritz,Schäftler/in,Übersetzer/in,Soldat/in,Journalist/in

exil.csv:

name_id,nachname,vorname,exilort1,exiland1,jahr1,exilort2,exiland2,jahr2,exilort3,exiland3,j

1,Walden,Ellen,Moskau,Sowjetunion,1932,,,,,,,,,,,,,Berlin,Deutsches Reic

2,Adler-Herzmark,Jenny,,Frankreich,1939,USA,1942,,,,,,,,,,,,,

3,Biefang,Johann,Donbass,Sowjetunion,1930,Moskau,Sowjetunion,1932,,Spanien,1937,,,,,,,,

4,Rosenfeld,Friedrich,Prag,Tschechoslowakei,1934,,Großbritannien,1939,,,,,,,,,,,,,

5,Frisch,Justinian,Stockholm,Schweden,1939,Cambridge,Großbritannien,1948,,,,,,,,,,,,,

6,Waldinger,Ernst,Paris,Frankreich,1938,London,Großbritannien,1938,New York,USA,1938,,,,,,,,

7,Thelen,Albert Vigoleis,Marseille,Frankreich,1936,Auessio,Schweiz,1936,,Frankreich,1939,An

8,Adler,Raissa Timofejewna,New York,USA,1935,Locarno,Schweiz,,New York,USA,1940,,,,,,,,

9,Lohan,Robert,Reading,Großbritannien,1938,New York,USA,1939,,,,,,,,,,,,,

10,Zwillinger,Frank Gerhard,Rom,Italien,1938,Saigon,Vietnam,1939,Paris,Frankreich,1946,,,,,,,,

11,Zweig,Stefan,London,Großbritannien,1934,New York,USA,,Argentinien,,Paraguay,,Brasilie

12,Zweig,Friderike Maria,Paris,Frankreich,1938,New York,USA,1940,,,,,,,,,,,,,

13,zur Mühlen,Hermynia,Wien,Österreich,1933,Bratislava,Tschechoslowakei,1938,London,Gro

14,Zuckmayer,Eduard,Ankara,Türkei,1936,,,,,,,,,,,,,Ober-Hambach,De

15,Zuckerandl-Szeps,Bertha,Paris,Frankreich,1938,Algier,Algerien,1940,Paris,Frankreich,194

16,Zernik-Bern,Clementine,New York,USA,1938,London,Großbritannien,1943,New York,USA,19



Thank you for providing the snippets of the CSV data. From the snippets, it is clear that the dataset is divided into multiple CSV files, each containing different aspects of information about individuals. Here is a brief summary of the structure of each CSV file:

1. **names.csv:** Contains information about individuals' names, gender, birth date, birthplace, death date, and place of death.
2. **occupation.csv:** Contains information about individuals' occupations and activities.
3. **exil.csv:** Contains information about the places exiled, the year of exile, and places of remigration.
4. **language.csv:** Contains information about the languages spoken.

Given this structure, we can create visualizations to analyze the data. Some potential research questions could be:

1. What is the distribution of individuals by gender?
2. What are the most common occupations among individuals?
3. What are the trends in the movement of individuals over time?
4. What are the most common countries of exile?
5. What languages were most commonly spoken?
6. Are there any patterns in the remigration of individuals?

Please let me know if you have any specific research questions, and I can proceed with any of the questions listed above. Once you provide more detail, I can start creating visualizations in a Jupyter Notebook.



You are an expert in data visualisation and the migration and exile of translators. Go into more detail for each research question.



Certainly! Let's delve into more detail for each research question and discuss the types of visualizations that can be used to analyze the data effectively:

1. **Distribution of Individuals by Gender:**
 - **Visualization Type:** Bar Chart
 - **Description:** We can create a simple bar chart to show the number of male and female individuals in the dataset. This will give us an understanding of the gender distribution among the translators who experienced exile.
2. **Most Common Occupations Among Individuals:**
 - **Visualization Type:** Horizontal Bar Chart
 - **Description:** We can create a horizontal bar chart to display the various occupations of the individuals, sorted by frequency. This will help us understand which professions were most common among the exiled translators.
3. **Trends in the Movement of Individuals Over Time (1930 to 1958):**
 - **Visualization Type:** Line Chart or Streamgraph
 - **Description:** We can plot a line chart with years on the x-axis and the number of individuals on the y-axis. This will show us how the movement of individuals changed over time. Alternatively, a streamgraph can be used to visualize the ebb and flow of groups of individuals moving to different countries over time.

Visualisierungsmöglichkeiten aus dem Kontext und den Beispieldaten generieren lassen!



Continue with the most common occupations of individuals. We use the Python for Google Colab notebook.



Certainly! I apologize for the interruption. Let's continue with the code to visualize the most common occupations among individuals. Below is the complete code to create a bar plot of the most common occupations. Please add this code to a new cell in your Google Colab notebook and execute it.

```
python 📄 Copy code  
  
import matplotlib.pyplot as plt  
  
# Read the occupation.csv file  
occupation_df = pd.read_csv('occupation.csv')  
  
# Extract the occupation columns  
occupation_columns = [col for col in occupation_df.columns if 'beruf' in col]  
  
# Count the occurrences of each occupation  
occupation_counts = {}  
for col in occupation_columns:  
    for occupation in occupation_df[col]:  
        if pd.isna(occupation):  
            continue  
        if occupation not in occupation_counts:  
            occupation_counts[occupation] = 1  
        else:  
            occupation_counts[occupation] += 1
```

Mit dem Outline unserer "Step-by-Step"-Struktur ("Anker setzen") wird nun der Python-Code für das Jupiter-Notebook in Google Colab generiert.

```

import matplotlib.pyplot as plt

# Read the Occupation.csv file
Occupation_df = pd.read_csv('Occupation.csv')

# Extract the Occupation columns
Occupation_columns = [col for col in Occupation_df.columns if 'Beruf' in col]

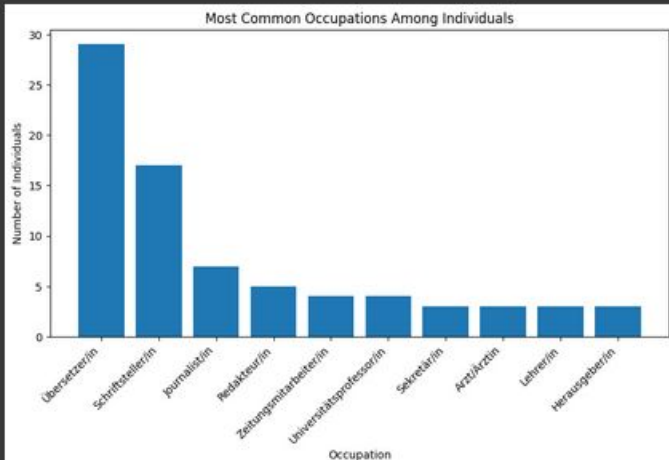
# Count the occurrences of each Occupation
Occupation_counts = {}
for col in Occupation_columns:
    for Occupation in Occupation_df[col]:
        if pd.isna(Occupation):
            continue
        if Occupation not in Occupation_counts:
            Occupation_counts[Occupation] = 1
        else:
            Occupation_counts[Occupation] += 1

# Sort Occupations by count
sorted_occupations = sorted(Occupation_counts.items(), key=lambda x: x[1], reverse=True)

# Extract the top 10 Occupations
top_occupations = sorted_occupations[:10]

# Plot the data
Occupations, counts = zip(*top_occupations)
plt.figure(figsize=(10, 5))
plt.bar(Occupations, counts)
plt.xlabel('Occupation')
plt.ylabel('Number of Individuals')
plt.title('Most Common Occupations Among Individuals')
plt.xticks(rotation=45, ha='right')
plt.show()

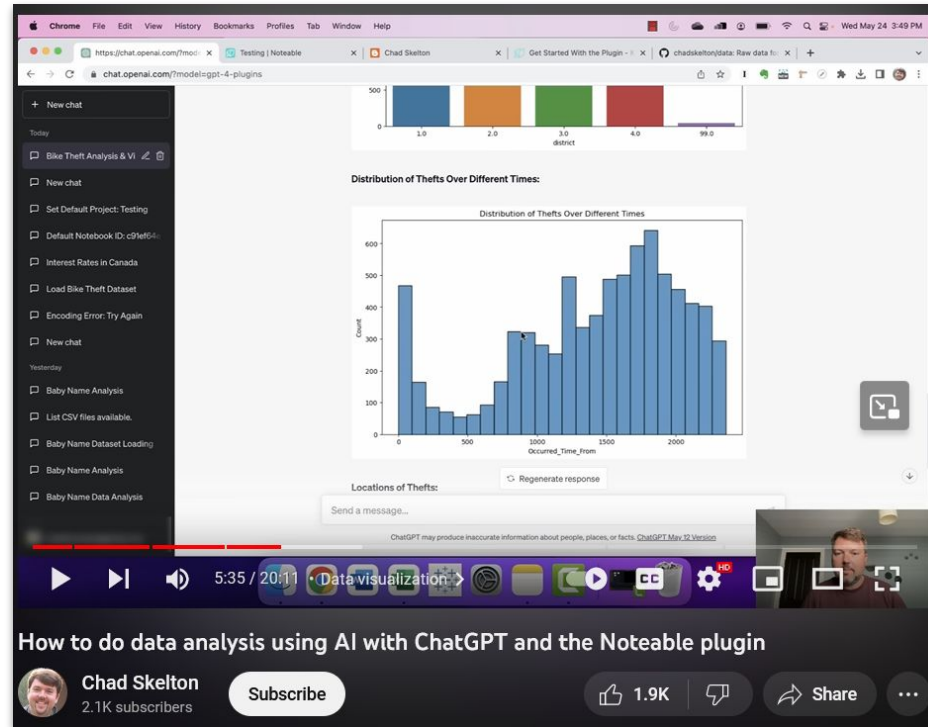
```



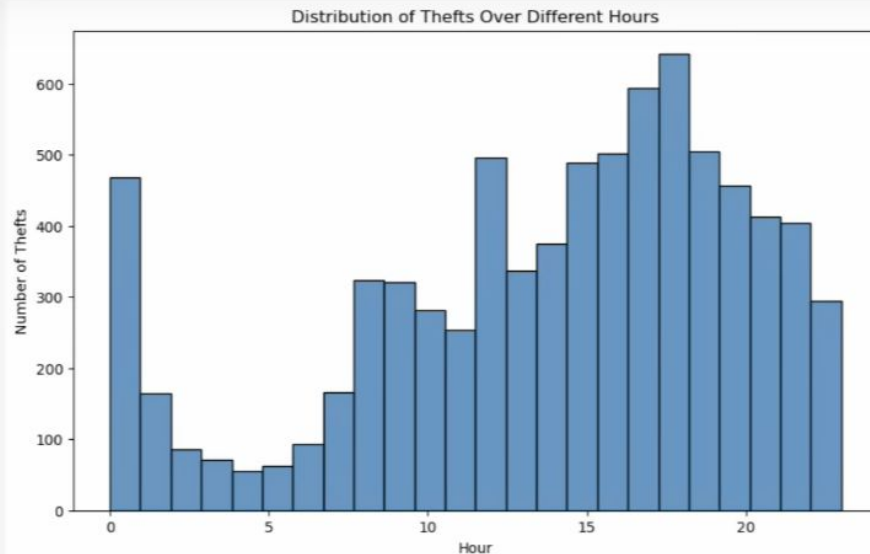
Code kopieren und ins Notizbuch einfügen! Ggf. Code selbst korrigieren oder einfach die Fehlermeldung kopieren und GPT-4 die Verbesserung durchführen lassen.

Jetzt mit allen Daten testen, sofern die Daten einheitlich strukturiert sind.

GPT-4 Plugin “Noteable”



Chad Skelton. How to do data analysis using AI with ChatGPT and the Noteable plugin. 2023. <https://www.youtube.com/watch?v=A1ualvzqJoo>
Blogpos: <http://www.chadskelton.com/2023/05/using-chatgpt-and-noteable-plugin-for.html>



From the plot, we can see that thefts tend to occur more frequently in the afternoon and evening hours, with a peak around 18:00 (6 PM). The number of thefts decreases significantly during the late night and early morning hours.



Why are there so many bike thefts right after midnight?



The spike in bike thefts right after midnight could be due to several reasons:

1. **Less Surveillance:** There are usually fewer people around after midnight, which might give thieves more opportunities to steal bikes without being noticed.
2. **More Bikes Left Unattended:** People often leave their bikes locked up outside for the night, making them easy targets for thieves.
3. **Reporting Bias:** Sometimes, if the exact time of the theft is not known, it might be reported as having occurred at the start of the day, which is midnight.
4. **Data Entry Errors:** If the time of the theft was not recorded, it might be entered as 00:00 by default, leading to an overrepresentation of thefts at midnight.

It's important to note that these are just possible explanations. To understand the exact reasons, we would need more detailed data and possibly additional information that might not be ■

GPT-4 erzeugt nicht nur Visualisierungen, sondern ermöglicht auch das Reden darüber.

Spannend: GPT-4 erkennt “Data Entry Errors”!



To interpret the data, I will start by loading each CSV file and displaying the first few rows.

This will give us an overview of the contents and structure of each file. Let's start with the first file, `exil.csv`, and take a look at its contents.

Used Noteable

The `exil.csv` file contains information about individuals who were in exile. It has the following columns in this file:

- * `index`: A unique identifier for each record.
- * `name_id`: An identifier for the name of the individual.
- * `nachname`: The last name of the individual.
- * `vorname`: The first name of the individual.
- * `exilort1`, `exilort2`, ...: The places where the individual was in exile.
- * `exilland1`, `exilland2`, ...: The countries corresponding to the places where the individual was in exile.
- * `jahr1`, `jahr2`, ...: The years corresponding to when the individual was in exile.
- * `remigration_ort1`, `remigration_ort2`, ...: The places where the individual returned after exile.
- * `remigration_land1`, `remigration_land2`, ...: The countries corresponding to the places where the individual returned after exile.
- * `remigration_jahr1`, `remigration_jahr2`, ...: The years corresponding to when the individual returned to each place after exile.
- * `wirkungsort_ort`: The place of effect or influence.
- * `wirkungsort_land`: The country of effect or influence.
- * `wirkungsort_jahr`: The year of effect or influence.

This data seems to be historical and could be related to individuals who were in exile for various reasons, possibly political or related to conflicts.

Let's continue by loading the next file, `language.csv`, to understand its contents.

Used Noteable

The `language.csv` file contains information about individuals and the languages they are associated with. Here are the columns in this file:

- * `index`: A unique identifier for each record.
- * `name_id`: An identifier for the name of the individual.
- * `nachname`: The last name of the individual.
- * `vorname`: The first name of the individual.
- * `z-sprache 1`, `z-sprache 2`, ...: The languages associated with the individual (possibly native languages).
- * `a-sprache 1`, `a-sprache 2`, ...: Additional languages associated with the individual.

This data seems to be related to the languages spoken or associated with individuals who might have been in exile.

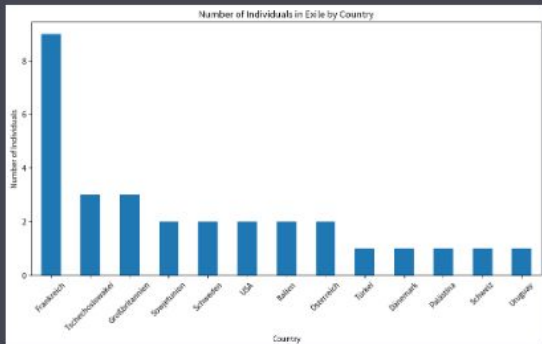
In summary, the data in these CSV files seems to be historical information about individuals who were in exile, including their names, places of exile, languages, occupations, and literary categories associated with them. This data could be used for historical research or analysis of individuals in exile.

Daten werden sinnvoll (ohne weiteren Kontext) interpretiert!

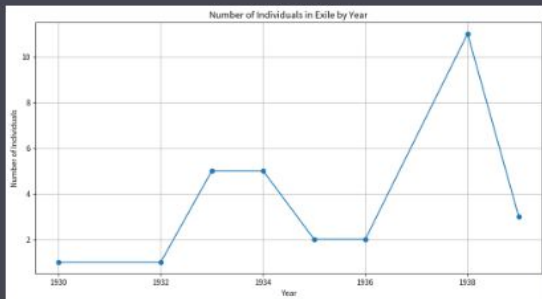


Certainly! Here are the visualizations I created based on the data in the "exil.csv" file:

1. **Number of Individuals in Exile by Country:** This bar chart shows the distribution of individuals in exile across different countries.



1. **Number of Individuals in Exile by Year:** This line chart shows the number of individuals in exile over time, based on the first year of exile for each individual.



These visualizations were created using the data in the "exil.csv" file and can help in understanding the distribution of individuals in exile by country and the trend over time.

If you have any further questions or if there's something specific you would like to analyze or visualize, please let me know!

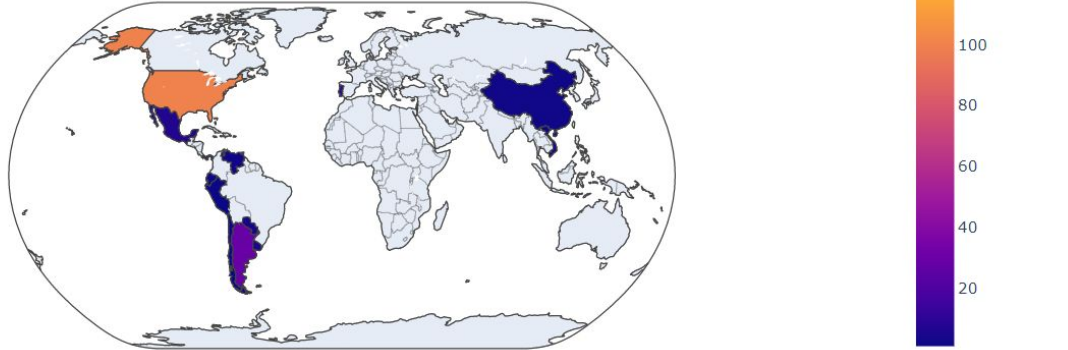
Flow of Individuals from France to the USA



I have the following question: Which people moved to the USA via France? Create a data visualisation to show me that

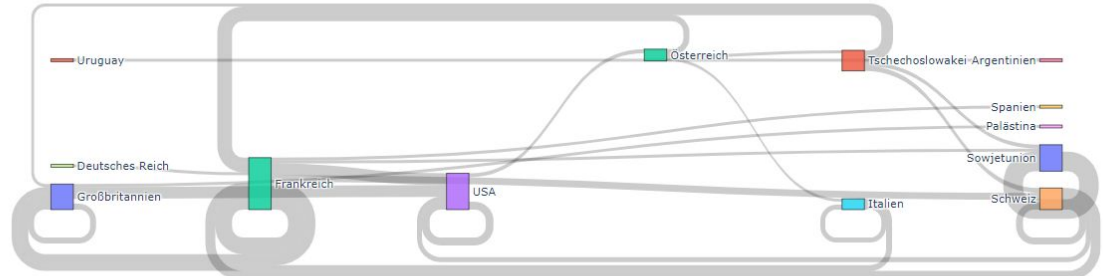
Number of Exiles per Country

Create a world map showing the number of exiles per country, highlighting the countries with the highest number of exiles.

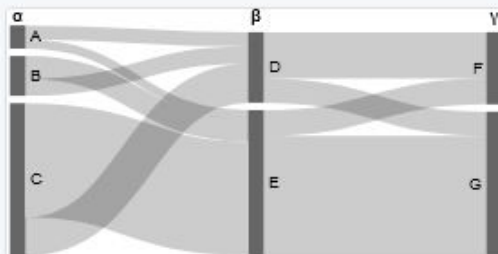


Mit GPT-4 komplexere Visualisierungen und Interpretationen oder “Mini-Data-Stories” erzeugen.

Exile Routes (Cardinality > 2)



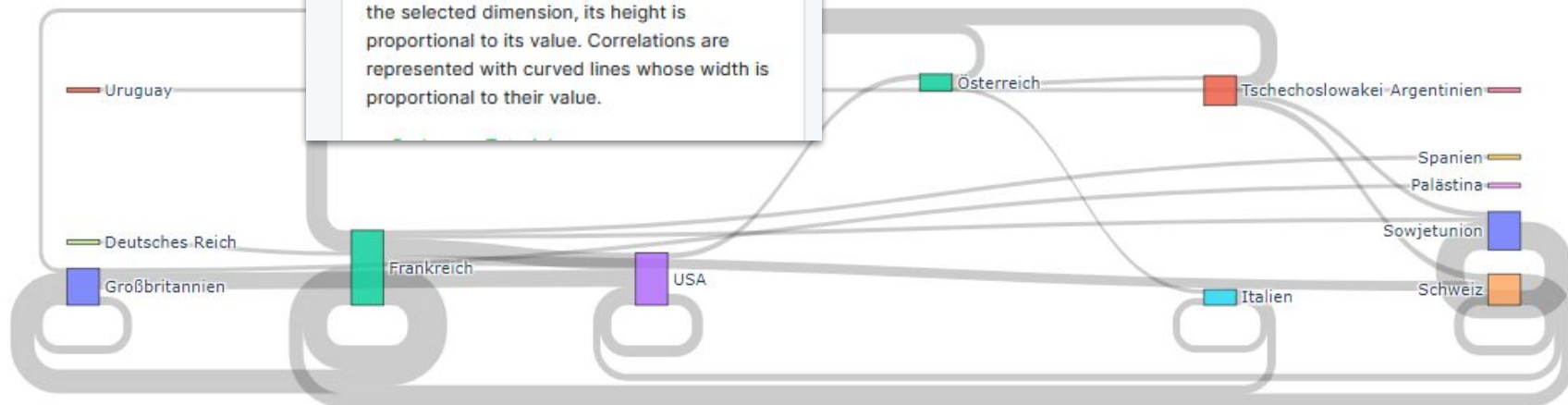
Examine the key routes and paths taken by exiled translators



Alluvial Diagram

It shows correlations between categorical dimensions representing them as flows, visually linking categories with shared items. Each rectangle represents a unique value in the selected dimension, its height is proportional to its value. Correlations are represented with curved lines whose width is proportional to their value.

Exile Routes (Cardinality > 2)



Examine the key routes and paths taken by exiled translators**

```
def extract_routes(row):
    route = []
    for i in range(1, 21):
        exilort = row[f'exilort{i}']
        exilland = row[f'exilland{i}']
        jahr = row[f'jahr{i}']
        if pd.notna(exilort) and pd.notna(exilland) and pd.notna(jahr):
            route.append((exilort, exilland, jahr))
    return route

exilStations['routes'] = exilStations.apply(extract_routes, axis=1)
```

This will generate a list of the most common paths taken by the exiled translators in the dataset, sorted by frequency.

*This code filters out connections with a cardinality of 2 and creates a **Sankey diagram** with the remaining data. The resulting visualization should be less cluttered and easier to interpret.*

For example, let's consider the path taken by Albert Ehrenstein. He initially fled to Brissago, Switzerland, in 1932, before moving to New York, USA, in 1941. This is one of the common routes observed in the dataset, where an individual first sought refuge in a European country, such as Switzerland, and later moved to the United States.

```
[ ] from collections import Counter

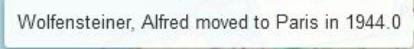
route_counter = Counter()
for routes in exilStations['routes']:
    for i in range(len(routes) - 1):
        route_counter[(routes[i][1], routes[i + 1][1])] += 1

route_counter.most_common()
```

By analyzing these routes, we can gain a better understanding of the decisions and challenges faced by translators during their exile journey, as well as the preferred destinations and paths chosen by the majority of exiled translators.

```
[(('Sowjetunion', 'Sowjetunion'), 20),  
 (('Frankreich', 'Frankreich'), 17),
```

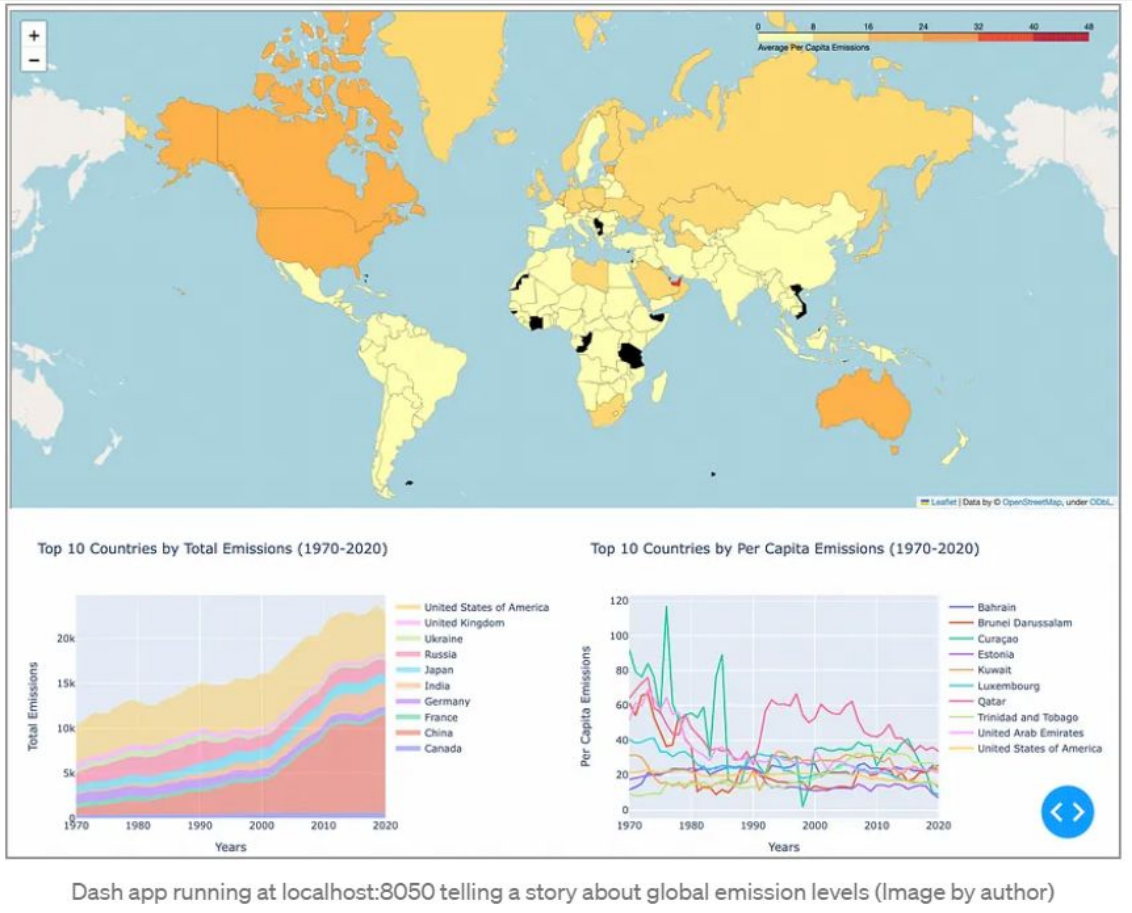
For example, in the visualization, you will see a relatively thick connection between "Frankreich" (France) and "USA," as well as "Großbritannien" (Great Britain) and "USA." This indicates that many exiled translators moved from France or Great Britain to the United States, which is supported by the provided data.



Python Folium: Create Web Maps From Your Data

Data Storytelling

- CSV-Daten
- Python + Jupyter Notebook
- Biographische Daten (Wikipedia?) als Kontext
- Happy Prompting! ;)



Insanely Fast Data Storytelling With ChatGPT and Python. Prompt engineering ChatGPT for Python data visualizations. 2023. <https://artificialcorner.com/insanely-fast-data-storytelling-with-chatgpt-and-python-1bddae3976f3>