

Proposal: [Your project name]

DATA 450 Capstone

[Your Name]

February 7, 2024

1 Introduction

Migration seems to be ingrained in the blueprint of our existence. From the dawn of our species, humans have migrated from Africa to Asia to Eurasia to America. Early ancestors moved from places to places in search of better habitats and food. The pattern of migration not only ensured the survival of our species with through suitable climate and food but also changed the behaviors and lifestyle of our species as a whole. The hunter-gatherers tribe settled into agricultural societies, and there are several revolutionary transformations to the structures of our society that can be ascribed to migration—rise and fall of kingdoms, industrial revolution, displacement of indigenous population, changes in the population characteristics (gender, religion, age groups), to mention some. While the nature of migration and its consequences have changed along with times, it is important that the patterns of migration are scrutinized to understand the socio-economic drivers of migration. This project will study the mobility of human from different countries across two different decades, 2000-2010 and 2010-2020. As stated before, humans have always been mobile species, but the motives have evolved over time. So it will also be equally important to shed the driving force of migration, rather than bluntly mentioned the statistics. In order to achieve this aim, this project will look migration patterns through the lenses of economy, war, epidemics, and human rights.

2 Dataset

The datasets that will be used to study the drivers and pattern of migration will be obtained from “Organization of Economic Cooperation and Development (OECD)” and “The Global Knowledge Partnership on Migration and Development (KNOMAD)”. The datasets from OECD will contain the measurements of migrations in different countries across 2002-2022 where as KNOMAD’s datasets will contain the capital outflows and inflows from a specific country.

The brief overlook of all the datasets will be presented below:

Migration Datasets

/opt/anaconda3/lib/python3.9/site-packages/IPython/core/formatters.py:343: FutureWarning:

In future versions `DataFrame.to_latex` is expected to utilise the base implementation of `S

| | CO2 | Country of birth/nationality | VAR | Variable | GEN | Gender |
|---|-----|------------------------------|-----|--|-----|--------|
| 0 | AFG | Afghanistan | B11 | Inflows of foreign population by nationality | TOT | Total |
| 1 | AFG | Afghanistan | B11 | Inflows of foreign population by nationality | TOT | Total |
| 2 | AFG | Afghanistan | B11 | Inflows of foreign population by nationality | TOT | Total |
| 3 | AFG | Afghanistan | B11 | Inflows of foreign population by nationality | TOT | Total |
| 4 | AFG | Afghanistan | B11 | Inflows of foreign population by nationality | TOT | Total |

Table 1

| CO2 | Country of birth/nationality | VAR | Variable | GEN | Gender | COU | Country |
|--|------------------------------|-----|----------|-----|--------|-----|---------|
| Loading... (need help?) | | | | | | | |

The other data that will be used for the purpose of study is remittance database. The remittance data are in two categories: i. capital money outflowing a country into a foreign country and ii. capital gains of a country from abroad.

/opt/anaconda3/lib/python3.9/site-packages/IPython/core/formatters.py:343: FutureWarning:

In future versions `DataFrame.to_latex` is expected to utilise the base implementation of `S

| | Remittance inflows (US\$ million) | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|-----------------------------------|-------|-------|--------|------------|-------------|-------------|
| 0 | Afghanistan | 0.0 | 0.0 | 0.00 | 0.000000 | 0.000000 | 0.000000 |
| 1 | Albania | 597.8 | 699.3 | 733.57 | 888.748582 | 1160.672105 | 1289.704316 |
| 2 | Algeria | 0.0 | 0.0 | 0.00 | 0.000000 | 0.000000 | 170.000000 |
| 3 | American Samoa | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | Andorra | 0.0 | 0.0 | 0.00 | 0.000000 | 0.000000 | 0.000000 |

[In this section, describe the dataset(s). This includes things like where you obtained the dataset. Include a full citation, as specified [here](#). Describe how the data was obtained by the data owner/curator, as best as you can. List the variables that you plan to use in your analysis, for example:

- weight: The patient's weight (kg)
- sex: The patient's sex, male or female
- age: The patient's age (months)

]

3 Data Acquisition and Processing

[In this section, if applicable, describe how you will obtain the data (if it's anything more complicated than a simple download). Discuss what data processing steps will be needed, such as recoding variables, data cleaning, data tidying, imputing missing values, etc. See sections 1c, 1d, 1e in the "Good Enough Practices" paper.]

4 Research Questions and Methodology

[In this section, list each of the questions you will explore. Following each question, provide a detailed and specific plan for how you plan to answer the question. Include the specific steps you will take, what form the answer will take (a number? table? visualization? model? Give all the specifics), and estimate how many hours each question will take to complete.]

1. Is smoking correlated with diabetes? To answer this, I will create a filled bar plot, with the left bar representing non-smokers, the middle bar representing people who smoke moderately, and the right bar representing heavy smokers. The bars will be the same height, and each bar will be colored two colors based on the proportion of patients in the group who do or do not have diabetes.
2. Question 2? Plan for question 2.
3. Question 3? Plan for question 3.
4. etc.

5 Work plan

[Fill in the list below with a plan for what you will do each week, starting 2/12. You should have around 7 hours worth of work each week. Writing work counts. Several tasks have already been filled in for you.]

Week 4 (2/12 - 2/18): [Just an example:

- Data tidying and recoding (4 hours)

- Question 2 (4 hours).]

Week 5 (2/19 - 2/25):

Week 6 (2/26 - 3/3):

Week 7 (3/4 - 3/10):

- Presentation prep and practice (4 hours)

Week 8 (3/11 - 3/17): *Presentations given on Wed-Thu 3/13-3/14. Poster Draft due Friday 3/15 (optional extension till 3/17).*

- Poster prep (4 hours)
- Presentation peer review (1.5 hours)

Week 9 (3/25 - 3/31): *Final Poster due Sunday 3/31.*

- Peer feedback (3.5 hours)
- Poster revisions (3.5 hours)
- [Do not schedule any other tasks for this week.]

Week 10 (4/1 - 4/7):

Week 11 (4/8 - 4/14):

Week 12 (4/15 - 4/21):

Week 13 (4/22 - 4/28): *Blog post draft 1 due Sunday night 4/28.* [All project work should be done by the end of this week. The remaining time will be used for writing up and presenting your results.]

- Draft blog post (4 hours).

Week 14 (4/29 - 5/5):

- Peer feedback (3 hours)
- Blog post revisions (4 hours)
- [Do not schedule any other tasks for this week.]

Week 15 (5/6 - 5/12): *Final blog post due Weds 5/8. Blog post read-throughs during final exam slot, Thursday May 9th, 8:00-11:20am.*

- Blog post revisions (2 hours)
- Peer feedback (2 hours)
- [Do not schedule any other tasks for this week.]

5.1 Citations

5.2 Some cool Quarto stuff

[You can delete this section from your proposal.]

For your reference, here’s an example of a Python code cell in Quarto, along with a figure that gets generated, along with a caption and a label so that it can be referred to automatically as “Figure 1” (or whatever) in the writeup.

For a demonstration of a line plot on a polar axis, see Figure 1.

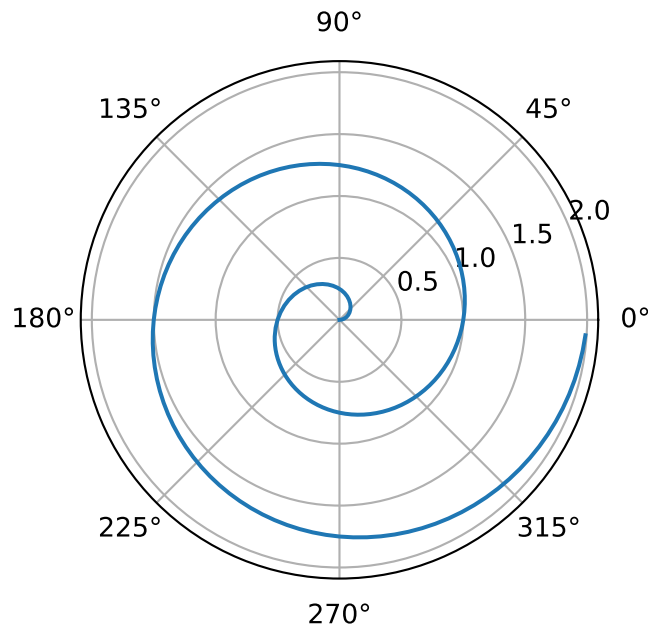


Figure 1: A line plot on a polar axis

Here’s an example of citing a source (see Phillips 1999, 33–35). Be sure the source information is entered in “BibTeX” form in the `references.bib` file.

6 References

[The bibliography will automatically get generated. Any sources you cite in the document will be included. Other entries in the `.bib` file will not be included.]

Phillips, T. P. 1999. “Possible Influence of the Magnetosphere on American History.” *J. Oddball Res.* 98: 1000–1003.