**FINAL EXERCISE**

**Due Thursday, December 1, 2022, 10:00 am**

**Objective**: The objective of the final exercise is to give you an opportunity to apply the analytical tools you have acquired in API-209 to a real policy setting, and to demonstrate how much you have learned in this course. The cases have been designed to engage you in the exercise by providing you with micro and macro options, which are linked to the rest of the MPA/ID core curriculum.

While a clear substantive understanding of health policies and policies aimed at fostering productivity will certainly be an advantage, it is neither necessary nor sufficient for doing well in this exercise. Regardless of the option you choose, the focus of the assignment should be the empirical analysis you conduct using the tools from API-209, and not the economic issues and broad policy recommendations, which are extremely important but not for the purpose of this exercise, unless explicitly stated.

**Grading**: The final exercise will make up 15% of your overall course grade. You will be graded on the level of sophistication of your analysis and your ability to present your analysis in words and visuals effectively, so do not simply submit R output like you might do for a problem set. In some cases, you might realize that you do not need the statistical methods we have covered in this course. The assessment of the level of sophistication of your analysis is not determined by the complexity of your analysis, but by your ability to use the appropriate statistical tools to approach the issue at hand.

**Instructions**: This is a group assignment. Each group should comprise 3 to 6 members.

**Final Work Products:** You are asked to submit two final products:

(1) **A memo** with the following format requirements: 10 single-spaced pages, with 1-inch margins, using font size 11 or higher. The memo should comprise a one-page executive summary, a four-page description of your key findings (i.e., the text of the memo), and a five-page technical appendix. The executive summary and text of the memo should be directed to a policymaker who has some familiarity with the policy issue but no extensive statistical training. The technical appendix is meant to document the technical aspects of your work and is directed at a more technical audience.

(2) **A PowerPoint presentation** with the following format requirements: 10 slides plus a cover slide. Two or more of the groups will be asked to make a short presentation in class on the day the final exercise is due. Two key components:

* Slides: What you would present (i.e., make sure your slides are effective visual aids for your presentation rather than the place where you try to write everything you did in final exercise). They should be addressed to a policymaker who has some familiarity with the policy issue but no extensive statistical training.
* Notes: What your team would say if you were asked to present (this can contain more details than the slides). These notes should be embedded in the notes section of the PowerPoint slides.

**Submission**: Your group should submit an electronic version of the following three files:

1. Your memo (executive summary, text and technical appendix) as a PDF
2. Your PowerPoint presentation
3. Your PowerPoint presentation as a PDF, showing both the slides and notes.

Please make one submission per group.

Please select **one** of the two following options:

1. **Identifying Labor Productivity Gaps in Mongolia:** You are asked to help the Ministry of Economic Development in Mongolia to identify and map the sources of low productivity of Mongolia’s labor force in the agriculture, service, and mining sectors, especially in the aftermath of the COVID-19 pandemic. The Mongolian Government is worried about an increasing mismatch between the skills required on the job and the current skill levels of Mongolian workers in three specific sectors (agriculture, services, and the mining sector), and wants to determine what’s at the source of this mismatch and how to improve it. To do this, you will adapt the growth diagnostic approach pioneered here at the Harvard Kennedy School (Hausmann, Rodrik, and Velasco) to analyze what the Ministry considers to be one of the most binding constraints to high productivity levels - human capital. You are welcome to look at other possibilities such as macroeconomic fundamentals and lack of sectoral capabilities if you wish to, but since the Mongolian government is particularly concerned about human capital, your focus should be to delve deeply into this sector to understand the magnitude and nature of the constraints involved there. I will select the most outstanding memo(s) and send them to two high-level officials in the Mongolian Government, the Minister of the Economic Development Ministry and the Director of the New Recovery Policy unit within the Prime Minister’s office. They, in turn, have agreed to set up a video conference with the team(s) that produce the best memo(s), where you’ll get the chance to present your work to them and their teams. They are eager to learn about your findings.
2. **Contributing to Health Care Expansion Strategy in Brazil**: You are asked to devise a strategy for expansion of health care infrastructure (physical and human) in the country. To do this, you will first characterize health outcomes in Brazil by describing the burden of disease in the country. You will then develop a machine learning algorithm that will predict disease burden for each municipality in the country. Then you will apply this algorithm and data on existing medical infrastructure to rank municipalities in terms of greatest need of additional infrastructure. The Health Ministry will then apply your algorithm to the Census data that is being collected at this moment. I will select the most outstanding memo(s) and send them to Arthur Aguillar, former MPA/ID and now Public Policies Director at the Public Health Studies Institute, who has worked several years on helping shape health public policies in the country. He, in turn, has agreed to set up a video conference with the team producing the best memo(s) where you will have a chance to present your work to him and his team. They are eager to learn about your findings so they can affect the future of health policies in Brazil.

**Data**: Although some data and documentation have been provided on the course website, it is your responsibility to ensure that you have the data and documentation required for the analysis. This includes searching for data and/or information that are not provided on the course website.

**Support**: The API-209 teaching team is available for consultation throughout the project. You are encouraged to consult them at any point in which guidance would be beneficial to you. Please refer to the course website for information on office hours.

**Schedule of Deadlines**: To encourage steady progress on your Final Exercise, you must submit both interim and final work product on a regular schedule over the coming weeks (as part of your problem set submissions). Your deadlines are as follows:

* **Tuesday, Nov 22 (by 10 AM)**: Submit draft version of your memo.
* **Tuesday Nov 29 (by 10 AM)**: Submit draft version of your presentation (two files; see details above).
* **Thursday Dec 1 (by 10 AM)**: Submit final set of deliverables

Specific submission instructions will be announced later. Late submission of the final memo and/or PowerPoint presentation will be penalized by 3 points plus an additional 3 points per day late.

Good luck!

### OPTION 1 – IMPROVING LABOR PRODUCTIVITY IN MONGOLIA[[1]](#footnote-1)

The Mongolian Government is worried about the increasing mismatch between the workers’ skills and the labor force demand required to economically develop the country. For a country that consistently ranks high in educational attainment and literacy rates, it is a conundrum that not the amount of the labor force, but the productivity of the labor force has been identified as a challenge. In the context of the Ukraine war and Mongolia’s increasing reliance on its neighboring countries, China and Russia, as a consequence of the COVID-19 pandemic as a landlocked import-reliant country,you are asked by Mongolia’s Ministry of Economic Development to **prepare a memo analyzing human capital as a binding constraint, including education, skills, and skill mismatch, and proposing best bets on how to lift the constraints you have identified to improve the labor force productivity in Mongolia.** You will focus on human capital as a whole and across the three sectors mentioned above (Agriculture, Manufacturing, and Services) and aim to compare Mongolia’s growth with peer benchmark countries of your own choosing.

One of the most widely used metrics to measure productivity is GDP per hour worked. However, this suffers majorly from various statistical issues such as unavailability of data on hours worked, and the lack of regard to the differences of the worker’s education level, sector, and experience. To that degree, the Mongolian Government and the Ministry of Economic Development use GDP per employee/worker as a proxy for productivity measurement. You may do the same when measuring labor productivity or suggest other ways to measure it - the Ministry is open to new ideas. The Mongolian National Statistics Office has published the national Labor Force Survey yearly since 2006, which is free and available to the public, containing information on individual level data on levels of education, income, and industry the subject is working in. You will also find the [2019 Report of Barometer Survey on Labor Market Demand](https://mlsp.gov.mn/uploads/files/Final_report_Barometr_survey_2019_ENG.pdf) and the [Dynamics of Unemployment 2020](https://1212.mn/BookLibraryDownload.ashx?url=Dynamics_of_unemployment_2020.pdf&ln=En), prepared by the World Bank and the Ministry of Labor and Social Protection, to be particularly helpful.

The original Growth Diagnostic (GD) framework has been adapted by Santos and Hani in their book “Diagnosing Human Capital as a Binding Constraint to Growth: Tests, Symptoms and Prescriptions”. One of the most important steps for identifying a binding constraint within the GD framework to analyze whether the four Principles of Differential Diagnosis apply to it:

|  |  |
| --- | --- |
| **The four diagnostic signals** | |
| **Diagnostic signal:**  if human capital is a binding constraint to growth | **Description:** |
| The shadow price of human capital should be high; there should be high-wage premiums for skilled workers | * High actual or implied market prices (wages) * A high shadow rice implies that relieving the constraint would have a large impact on private investment |
| Changes in the stock of human capital should be associated to changes in private investment and growth | * If a human capital is a binding constraint to investment or growth, relaxing the constraint should be associated with incremental private investments and growth |
| Agents attempting to overcome or bypass skills shortages | * Agents in the economy are likely responding to the constraint through various interventions or investments to circumvent human capital shortages. |
| Camels and hippos: Agents less intensive in human capital are more likely to thrive (and vice versa) | * Sectors that rely more intensely on human capital should be less prevalent or have a relatively lower contribution to exports, value added, or employment than those that do not |

*Source: Santos and Hani (2022)*

The final product should be a ten-page document, comprising a one-page executive summary, a four-page memo and a five-page technical appendix. While the team working in the Ministry of Economic Development have technical backgrounds, write the memo in a way that is suitable to a broad array of policymakers that will be engaged in the discussion of this issue, and leave the technical analysis to the appendix. The appendix should contain the details of the statistical analyses you conducted, including a description of the exact cause you propose is at the root of the low productivity, and the process you employed to arrive at this conclusion.

The exercise is purposefully open-ended as most assignments are in the real world. We will provide you with extensive support during the course of this assignment. You have been assigned mentors: Tyler Simko, Marco Brancher, Nomko Baatar, Vaishnavi Prathap and Dan Levy, all of whom will be available during their usual consultation hours.

**We wish you the best of luck in identifying Mongolia’s most binding human capital constraints in this project!**

**Appendix to Option 1: Description of Datasets and Resources**

Below are some data sources to begin the analysis. It is your responsibility to ensure that you have the data and documentation required for the analysis. To download the microdata from the [NSO](http://web.nso.mn/nada/index.php/catalog/central/about), you have to create an account and request to download the file (you can download the data directly from the website, you might not get a confirmation email). Once you agree to the terms and conditions of the data usage, you will be able to download the file in .sav format (please see screenshot in appendix).[[2]](#footnote-2) You will also find the full data dictionary as well as the survey/questionnaires used within the website.

**You might find the following resources useful:**

**General:**

* [World Bank Mongolia overview](https://www.worldbank.org/en/country/mongolia/overview): Useful to familiarize yourself with the context, especially the in-depth materials and open data. Feel free to approach Nomko for additional materials if needed.
* [Atlas of Economic Complexity](https://atlas.cid.harvard.edu/countries/145): Useful for understanding what the main drivers of GDP are in the country.
* “Diagnosing Human Capital as a Binding Constraint to Growth: Tests, Symptoms and Prescriptions” ([excerpts](https://growthlab.cid.harvard.edu/publications/diagnosing-human-capital-binding-constraint-growth-tests-symptoms-and)) by Miguel Santos and Farah Hani. The book is about how to assess when human capital is a constraint. Miguel worked for eight years at the Growth Lab as Senior Research Fellow and Director for Applied Research, and helped the teaching team to design this exercise. He recently moved to London to get the Growth Lab started at the London School of Economics.
  + Some ideas from the Growth Lab to get you started - you can test for years of education (Mincer regressions), quality of education, experience, discrimination (gender and disabilities), sector (urban vs rural, formal vs informal, industry, and occupation), accessibility to jobs, and migrant status.

**Main data:**

* [Labor force survey (2002-2021)](http://web.nso.mn/nada/index.php/catalog/LFS) – 46,933 observations with 212 variables for 2020
  + Unit of analysis: household, household member, weighted
  + Disability information, educational attainment, industry code by ISIC-04 and job responsibilities by ISCO-08, wage information, hours worked, and more
* [Time use study (2011, 2015, 2019)](http://web.nso.mn/nada/index.php/catalog/TUS/dataset) – [questionnaires](http://web.nso.mn/nada/index.php/catalog/124/related_materials) and data dictionary available, 13,499 observations with 77 variables
  + Unit of analysis: Household, household members aged 12 years and above, and mother/caretakers of children under 5 years, weighted
  + Leisure time spent on which activities, employment and wage information, time spent on unpaid labor
* [Household socio-economic survey (2002-2021)](http://web.nso.mn/nada/index.php/catalog/HSES/dataset) - 11,197 households with 100+ variables for 2019, [data dictionary](http://web.nso.mn/nada/index.php/catalog/121/data_dictionary) available
* [2020 population census, individual and household level](http://web.nso.mn/nadamn/index.php/catalog/PHC/dataset) – 333,402 observations with 65 variables, data dictionary attached
  + In Mongolian, ask Nomko for help if needed

**Education specific data**

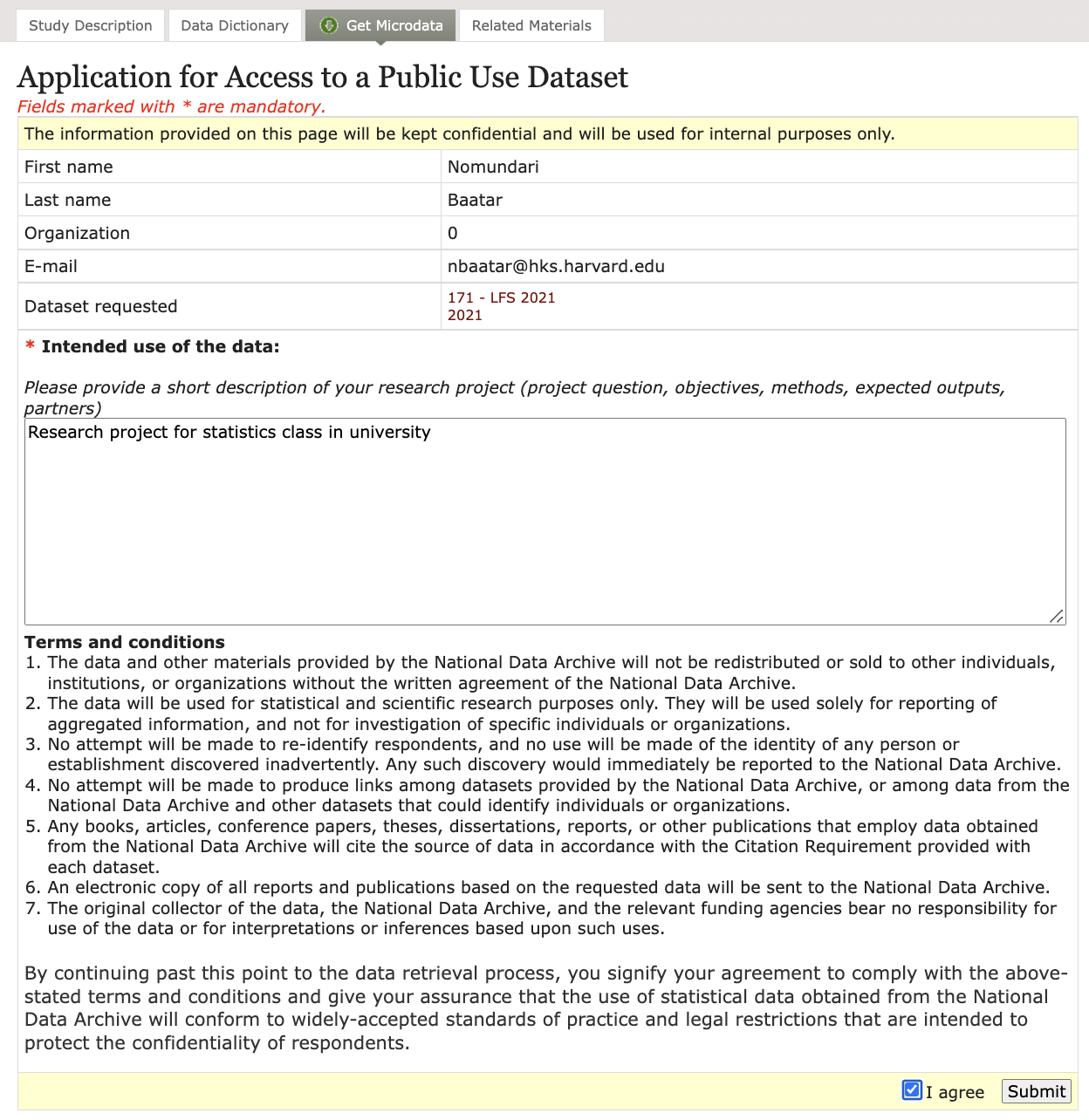
* [NSO Data on education](https://www.1212.mn/Stat.aspx?LIST_ID=976_L20_1&type=description) 
  + [Educational overview report from the Ministry of Education](https://1212.mn/BookLibraryDownload.ashx?url=8._Education_2021-2022_mn.pdf&ln=Mn)
* [Education Sector Mid-term Development Plan](https://cdn.greensoft.mn/uploads/users/2649/files/PDF%20files/Development%20plan%202030%20En.pdf)
* [Ministry of Education Comprehensive Statistics](https://www.meds.gov.mn/statistics-industry?category_id=14932)

**You might also find the following databases helpful for additional data on Mongolia and peer countries:**

* [All other microdata published by the NSO](http://web.nso.mn/nada/index.php/catalog/central/about)
* [Mongolian Statistical Information Service](http://www.1212.mn/)
* WB’s World Development Indicators
* Penn World Tables
* WEF’s Global Competitiveness Index
* WB’s World Governance Indicators
* IMF’s World Economic Outlook
* CID’s Atlas of Economic Complexity
* Reinhart and Rogoff’s Foreign Exchange System data
* Barro-Lee’s Education Indicators
* WB’s Enterprise Surveys
* WB’s Doing Business Indicators (now defunct)
* Economist Intelligence Unit data
* Quality of Government Indicators

### 

**Application for Access to Dataset**



### 

### OPTION 2 – HEALTH REFORM IN BRAZIL[[3]](#footnote-3)

After the elections held in Brazil in October, one of your MPA/ID classmates was nominated as Health Minister. By having experienced the program with you, the Minister is very confident that with the tools learned at the Kennedy School and with the help of other MPA/IDs, she will contribute to make significant changes on the Brazilian Health landscape. You have now been invited to be part of the team that aims to change the future of health policies in Brazil.

The Brazilian 1988 Constitution granted every citizen the right to access free-of-charge health care services, but universal health care coverage has not been achieved. Despite many improvements in health access and health outcomes over the past two decades, gaps remain. Therefore, the newly elected President is committed to producing changes and has asked the Ministry to focus on two tasks:

* **Task #1 - Characterize health care in Brazil**: This task consists of analyzing health indicators in Brazil, especially describing health care access and health outcomes, focusing on the burden of diseases. This analysis needs to be done for the overall population, and will also seek to characterize differences across regions, socio-demographic sub-groups, and other sub-groups of interest. The information you gather will be fundamental for the new government as it will help the government understand the health status of the country. Your work will therefore serve as the basis for every health public policy discussion in the government.
* **Task #2 - Strategy for expanding health care infrastructure:** The Minister believes that having a good diagnostic is a necessary condition to transform health policies in Brazil, but wants to present to the President a new strategy for expansion of health care infrastructure (physical and human) in the country. To do this, you are asked to develop a machine learning algorithm that will predict disease burden for each municipality. Then you will apply this algorithm and data on existing medical infrastructure to rank municipalities in terms of greatest need of additional infrastructure.

The final product should be a ten-page document, comprising a one-page executive summary, a four-page memo and a five-page technical appendix. While people in the high levels of the Brazilian Government have technical backgrounds, write the memo in a way that is suitable to a broad array of policymakers that will be engaged in the discussion of this issue, and leave the technical analysis to the appendix. The appendix should contain the details of the statistical analyses you conducted, including a description of the exact design you propose, and the process you employed to arrive at this design. Remember that your work will be used to potentially determine who will have access to the public health system and who does not, so the stakes are very high.

We will provide you with extensive support during the course of this assignment. You have been assigned mentors: Tyler Simko, Marco Brancher, Nomko Baatar, Vaishnavi Prathap and Dan Levy, all of whom will be available during their usual consultation hours. We wish you the best of luck in this crucial assignment for the Brazilian people!

**Appendix to Option 2: Description of Datasets and Resources**

* **PNS – National Health Survey**: The PNS is a probabilistic sampling nationwide household survey conducted by IBGE. It aims to ‘evaluate health conditions, health service access’ as well as perform ‘surveillance of noncommunicable diseases and their social determinants. The PNS targets individuals aged 15 and over. In the PNS of 2019, a total of 279,382 households and 94,114 respondents to the individual questionnaire were included. Individual respondents were randomly selected from all household members aged 15 and above. The survey collected individual socio-demographic and contained modules addressing general health status, lifestyle, communicable and NCDs and health service usage, among others. Weighting of responses are adjusted for likelihood of selection and rate of non-response by sex and age category (available [**here**](https://drive.google.com/file/d/11o2HemkcSd0BRTBPF-dW9a5jMge2PpPy/view?usp=sharing)).
* **SIM – Mortality Information System**: Data consolidated by municipality and cause of death by ICD-10 chapter ([**ICD-10**](https://www.cms.gov/Medicare/Coding/ICD10/Downloads/2018-ICD-10-CM-Coding-Guidelines.pdf) is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems, a medical classification list by the World Health Organization). Note that you are going to be working with ICD-10 chapters, and not with specific diseases. For example, you are not going to find Covid-19, but Diseases of the Respiratory System (chapter 10).
  + The data set for 2020 is available [**here**](https://drive.google.com/open?id=1Ho0rEUxus7hc0blvKKkwSsshdrody_bW&authuser=marcobrancher%40gmail.com&usp=drive_fs).
  + The data set for 2010 is available [**here**](https://drive.google.com/open?id=1Jup5oE_kGSv3lxUg0SqbqnjphdU_ufk9&authuser=marcobrancher%40gmail.com&usp=drive_fs).
* **SIH – Hospitalizations Information System**: Data consolidated by municipality and disease that caused the hospitalization (by ICD-10 chapter). This data covers all hospital admissions covered by SUS, both in public facilities and in accredited private hospitals and are recorded by the patient'’ municipality of origin, and not the municipality where she was hospitalized.
  + The data set for 2020 is available [**here**](https://drive.google.com/file/d/1HvozKTR7Jy5Cdk2P10sXh7QjzsZzc-Db/view?usp=sharing).
  + The data set for 2010 is available [**here**](https://drive.google.com/open?id=1Jup5oE_kGSv3lxUg0SqbqnjphdU_ufk9&authuser=marcobrancher%40gmail.com&usp=drive_fs).
* **CNES – Health Infrastructure Database**:  Administrative dataset that contains information on every health facility in Brazil. The data include the location, type of services provided, type of establishment, and the human resources available in each facility. Auxiliary microdata also includes an individual identification number for each professional, which allows to identify the number of different professionals enrolled in each facility as well as their average number of hours worked per week. We have already consolidated the data at the municipality level and we provide you three databases:
  + one that provides the number of health professionals (2020 is available [**here**](https://drive.google.com/open?id=1JtL3sPhWNzXkEv25Oq6VyLJ2epC4zfhg&authuser=marcobrancher%40gmail.com&usp=drive_fs) and 2010 is available [**here**](https://drive.google.com/open?id=1JoQcY9pNZ9FVpkEqwEHmtKrXzwgw8pF8&authuser=marcobrancher%40gmail.com&usp=drive_fs));
  + one that provides the number of health facilities (2020 is available [**here**](https://drive.google.com/open?id=1Ic5-bRrk3EdR4uyRNOZ9g3bRkN8RT8iS&authuser=marcobrancher%40gmail.com&usp=drive_fs) and 2010 is available [**here**](https://drive.google.com/open?id=1K4OWJ8NiT7YmRzi9COwno_iuAy_XNbSI&authuser=marcobrancher%40gmail.com&usp=drive_fs)); and
  + one that provides the number of health equipment (2020 is available [**here**](https://drive.google.com/open?id=1ISt45yBLZ0fAzDVZpefFaoJHS_RpJPBt&authuser=marcobrancher%40gmail.com&usp=drive_fs) and 2010 is available [**here**](https://drive.google.com/open?id=1KE08GkA6ig-ChhAu8_9D4_yFKy3Szm5x&authuser=marcobrancher%40gmail.com&usp=drive_fs)).
* **2010 Brazilian Census**: Survey that investigates life conditions of the population in all the municipalities of Brazil and in their internal territorial divisions, having the resident person in the housing unit in the National Territory in the reference date as the data collection unit. We provide you two datasets both consolidated at the municipality level. The first one contains information of [**individuals**](https://drive.google.com/file/d/1IQ4tgy7XEhv9SPEWOlYDjeBYC99M3_n0/view?usp=sharing), and the second, provides you information on [**households**](https://drive.google.com/file/d/1IPItIMxU_yInEeCZH4vJZrxYJ62uO4-e/view?usp=sharing).
* Finally, you can find [**here**](https://docs.google.com/spreadsheets/d/1KGcRTULG_p5q1D4_EiqshlMAqcYpLZTW?rtpof=true&authuser=marcobrancher%40gmail.com&usp=drive_fs) a table with all the municipality codes and their respective states.

To assist with your analysis, please note the following:

* For PNS: given the variation in sampling of households, you will want to use weights in your analysis. You will notice a variable named “*V00281*” in the dataset. For your analysis, you should use this as your weighting variable. Note that this is a large health survey and we are providing it without any data cleaning so you can have some extra fun! The codebook in Portuguese can be found [**here**](https://docs.google.com/spreadsheets/d/1DS3-ET-DGILcfHqTuDPQOyYMMyiheBEs/edit#gid=1078377405) and the one in English is [**here**](https://docs.google.com/spreadsheets/d/1DXlBHwqxYAGQIPRH2OI6Cs9COJ_IcIK4/edit#gid=1819249255).
* You might want to to use other data sources, some of them available [**here**](https://iepsdata.org.br) [in portuguese, but you can find a codebook with the translation for all the variables [**here**](https://docs.google.com/spreadsheets/d/1uMGfkpJjs_O5OXrXE-lDir9mtVc8BYFc?rtpof=true&authuser=marcobrancher%40gmail.com&usp=drive_fs) 😀]
* To help you with your tasks, please consider the following steps:
  + Use the **SIM** and **SIH** datasets to understand the burden of diseases in Brazil.
  + Then, use the **PNS** (which covers a random sample of municipalities) to develop an algorithm that can predict the disease burden of each municipality on the basis of the existing human and health infrastructure (for the infrastructure you can use the **CNES** database).
  + Apply this algorithm to the **2010 Brazilian Census** (which contains all municipalities in Brazil) so that on the basis of the existing infrastructure, you can predict disease burden for that municipality.
  + Make recommendations on where Brazil should prioritize the expansion of infrastructure on the basis of disease burden and your findings.

#### 

1. I am very grateful to current MPA/ID student and API-209 Course Assistant Nomko Baatar for her help conceptualizing, designing, and producing this final exercise, and to Miguel Santos for his guidance in producing it in a manner that builds on the work the Growth Lab has been doing in several of their engagements in the world. [↑](#footnote-ref-1)
2. You will be able to read a .sav file into R by using the “haven” package and the following function: read\_sav("data.sav") [↑](#footnote-ref-2)
3. I am immensely grateful to former MPA/ID student Arthur Aguillar and to his colleagues at IEPS, especially Manuel Farias and Marcela Camargo for their help in designing this final exercise in a way that involves tackling a real challenge faced by a development practitioner, and to current MPA/ID student and API-209 Course Assistant Marco Brancher for his help conceptualizing, designing, and producing this final exercise. [↑](#footnote-ref-3)