

# Data Sources

Prof. Tzu-Ting Yang

楊子霆

Institute of Economics, Academia Sinica

中央研究院經濟研究所

September 3, 2024

# Data Sources

- Survey data
- Administrative data
- Firm-level data
- Country-level data
- Government Open data
- Geographical data
- Journal data archives
- Other data source

## Survey Data: Cross-sectional Data

# Survey Data

- General information for Taiwan survey data
  - Survey Research Data Archive (Academia Sinica): <https://srda.sinica.edu.tw/>
- General information for US survey data
  - IPUMS (University of Minnesota): <https://www.ipums.org/>

# Survey Data

- Taiwan (cross-sectional data)
  - Population and Housing Census (戶口及住宅普查):  
[https://srda.sinica.edu.tw/datasearch\\_detail.php?id=205](https://srda.sinica.edu.tw/datasearch_detail.php?id=205)
  - Firm Census (工業及服務業普查):  
[https://srda.sinica.edu.tw/browsingbydatatype\\_result.php?category=surveymethod&type=4&typeb=007&csid=53](https://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=4&typeb=007&csid=53)
  - Agricultural, Forestry, Fishery and Husbandry Census (農林漁牧業普查):  
[https://srda.sinica.edu.tw/browsingbydatatype\\_result.php?category=surveymethod&type=4&typeb=007&csid=51](https://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=4&typeb=007&csid=51)

# Survey Data

- Taiwan (cross-sectional data)
  - Manpower Utilization Survey: [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2826](https://srda.sinica.edu.tw/datasearch_detail.php?id=2826)
  - Survey of Family Income and Expenditure: [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2850](https://srda.sinica.edu.tw/datasearch_detail.php?id=2850)
  - Taiwan Social Change Survey: [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2822](https://srda.sinica.edu.tw/datasearch_detail.php?id=2822)

# Population and Housing Census

戶口及住宅普查

- **Website:**

[https://srda.sinica.edu.tw/datasearch\\_detail.php?id=205](https://srda.sinica.edu.tw/datasearch_detail.php?id=205)

- **Observation unit:** individual

- **Frequency:** 1960-2010, every 10 year

- **Key variables:** age, occupation, living township, marital status, disability status, housing size

- It does not have English version.

- Only faculty member at Academia Sinica can download it. Let me know if you want to use it

# Firm Census

工業及服務業普查

- **Website:**  
[https://srda.sinica.edu.tw/browsingbydatatype\\_result.php?category=surveymethod&type=4&typeb=007&csid=53](https://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=4&typeb=007&csid=53)
- **Observation unit:** firm / village
- **Frequency:** 1981-2016, every 5 year
- **Key variables:** Number of employees, Payroll expenses, Land area, and Total production value
- It does not have English version.
- Only faculty member at Academia Sinica can download it. Let me know if you want to use it



# Agricultural, Forestry, Fishery and Husbandry Census

農林漁牧業普查

- **Website:**

[https://srda.sinica.edu.tw/browsingbydatatype\\_result.php?category=surveymethod&type=4&typeb=007&csid=51](https://srda.sinica.edu.tw/browsingbydatatype_result.php?category=surveymethod&type=4&typeb=007&csid=51)

- **Observation unit:** individual / village
- **Frequency:** 1970-2020, every 5 year
- **Key variables:** gender of agricultural and livestock household managers, their age, educational level, the number of days engaged in personal farming and livestock activities,
- It does not have English version.
- Only faculty member at Academia Sinica can download it. Let me know if you want to use it

# Manpower Utilization Survey

- **Website:**  
[https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2826](https://srda.sinica.edu.tw/datasearch_detail.php?id=2826)
- **Observation unit:** individual (above age 15)
- **Frequency:** 1976-now, every year (May)
- **Key variables:** age, occupation, industry, monthly wage, working hour, part-time job, firm size, number of children by children's age, employment status, unemployment reason
- Labor force survey
- It has English version.

# Survey of Family Income and Expenditure

- **Website:** [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2850](https://srda.sinica.edu.tw/datasearch_detail.php?id=2850)
- **Observation unit:** household
- **Frequency:** 1976-now, every year
- **Key variables:** household income, household income by family members, household expenditure
- It has English version.

# Taiwan Social Change Survey

- **Website:** [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2822](https://srda.sinica.edu.tw/datasearch_detail.php?id=2822)
- **Observation unit:** individual
- **Frequency:** 1980-now, every year
- **Key variables:** marital status, occupation, industry, monthly wage, birth year, birth month, many questions about a respondent's opinion on politics, social issue, religion
- It has English version.

# Survey Data

- United States (cross-sectional data)
  - US Census Data: <https://usa.ipums.org/usa>
  - Current Population Survey (CPS): <https://cps.ipums.org/cps>
  - American Time Use Survey (ATUS): <https://www.atusdata.org/atus>  
<https://www.ahtusdata.org/ahtus>
  - National Health Interview Survey (NHIS): <https://ihis.ipums.org/ihis>
  - IPUMS Higher Education Survey: <https://highered.ipums.org/highered>

# Current Population Survey (CPS)

- **Website:**  
<https://cps.ipums.org/cps>
- **Observation unit:** individual and household
- **Frequency:** 1962-now, every year
- **Key variables:** age, occupation, marital status, migration, wage, disability, tax, health insurance, education

# American Time Use Survey (ATUS)

- **Website:**  
<https://www.atusdata.org/atus>  
<https://www.ahtusdata.org/ahtus>
- **Observation unit:** activity, individual and household
- **Frequency:** 2003-now, every year
- **Key variables:** time use on each activity
- Select one family member in CPS household and ask them about their time use (24 hour) on specific date
- This survey can link with CPS to get more information

# National Health Interview Survey (NHIS)

- **Website:**  
<https://ihis.ipums.org/ihis>
- **Observation unit:** individual and household
- **Frequency:** 1960-now, every year
- **Key variables:** health condition, mortality, health behavior, healthcare use



# IPUMS Higher Education Survey

- **Website:**  
<https://ihis.ipums.org/ihis>
- **Observation unit:** individual and household
- **Frequency:** 1993-now, every year
- **Key variables:** education, employment, income, work history

# Survey Data

## Developed Countries

- Other countries (cross-sectional data)
  - Census data from other countries: <https://international.ipums.org/international/>
  - North Atlantic Population Project: <https://www.nappdata.org/napp/>
    - Census microdata from Canada, Denmark, Great Britain, Germany, Iceland, Norway, Sweden, and the United States from 1703 to 1911
  - Multinational Time Use Study: <https://www.mtusdata.org/mtus/>

# Survey Data

## Developed Countries

- Other countries (cross-sectional data)
  - Luxembourg Income Study Database (LIS)  
<http://www.lisdatacenter.org/our-data/lis-database/>
    - The largest available income database of harmonised microdata collected from about 50 countries spanning five decades.
  - Codebook:  
<http://www.lisdatacenter.org/wp-content/uploads/files/data-lis-variables.pdf>
  - Labour income, capital income, pensions, public social benefits (excl. pensions) and private transfers, as well as taxes and contributions, demography, employment, and expenditures

# Survey Data

## Developed Countries

- Other countries (cross-sectional data)
  - Luxembourg Wealth Study Database (LWS)  
<http://www.lisdatacenter.org/our-data/lws-database/>
    - The first cross-national wealth database in existence. Harmonized into a common framework
  - Codebook:  
<http://www.lisdatacenter.org/wp-content/uploads/files/data-lws-variables.pdf>
  - Assets and debt, market and government income, household characteristics, labour market outcomes and, in some datasets, expenditures and behavioural indicators

# Survey Data

## Developed Countries

- Other countries (cross-sectional data)
  - The International Social Survey Programme (ISSP)  
<https://issp.org/>
    - The ISSP is a cross-national collaboration programme conducting annual surveys on diverse topics relevant to social sciences
  - Codebook:  
<https://issp.org/data-download/by-topic/>
  - People's opinions about Family and Changing Gender Roles, Social Inequality, Environment, National Identity, and Social Networks

# Survey Data

## Developing Countries

- Other countries (cross-sectional data)
  - World Bank Survey database:  
<http://microdata.worldbank.org/index.php/home>
  - Demographic and Health Survey (DHS):  
<https://dhsprogram.com/data/>
    - DHS collect and disseminate accurate, nationally representative data on health and population in developing countries.

# Survey Data

## Standardized Test Data

- Other countries (cross-sectional data)
  - PISA test database:  
<http://www.oecd.org/pisa/data/>
  - TIMSS test database:  
<https://timssandpirls.bc.edu/>
    - PISA and TIMSS are databases for international comparative assessments of student achievement

## Survey Data: Panel Data



# Survey Data

- Taiwan (panel data)
  - Panel Survey of Family Dynamics (PSFD): [http://psfd.sinica.edu.tw/web/data\\_en.htm](http://psfd.sinica.edu.tw/web/data_en.htm)
  - Taiwan Education Panel Study (TEPS): [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2755](https://srda.sinica.edu.tw/datasearch_detail.php?id=2755)

# Panel Survey of Family Dynamics (PSFD)

- **Website:**  
[http://psfd.sinica.edu.tw/web/data\\_en.htm](http://psfd.sinica.edu.tw/web/data_en.htm)
- **Observation unit:** follow the same households over time and follow their offspring's households over time
- **Frequency:** 1999-now, every year
- **Key variables:** household income, education, working experience, marriage, family value and attitudes, household expenditure, child care
- It has English version.

# Taiwan Education Panel Study (TEPS)

- **Website:** [https://srda.sinica.edu.tw/datasearch\\_detail.php?id=2755](https://srda.sinica.edu.tw/datasearch_detail.php?id=2755)
- **Observation unit:** follow the same students from their junior/senior high school and also have questions for their parents, teachers, and schools
- **Frequency:** 2001-2007 and 2010-2014
- **Key variables:** test score, friends, part-time work, learning attitude, extracurricular activities, wage
- It has English version.

# Survey Data

- United States (panel data)
  - Panel of Income Dynamics (PSID): <http://psidonline.isr.umich.edu>
  - Health and Retirement Study (HRS): <https://hrs.isr.umich.edu/data-products>
  - National Longitudinal Surveys (NLS): <https://www.nlsinfo.org/content/getting-started/accessing-data>
  - Survey of Income and Program Participation (SIPP): <http://www.nber.org/data/survey-of-income-and-program-participation.html>
  - Medical Expenditure Panel Survey (MEPS): <https://www.meps.ahrq.gov/mepsweb/>

- Other countries (panel data)
  - Cross-National Equivalent File (CNEF): <https://www.cnefdata.org/about/about-cnef>
    - The Cross-National Equivalent File is a user package of equivalently defined variables for 9 general population household panel surveys.
  - Codebook:  
<https://www.cnefdata.org/documentation/codebooks>

# Administrative Data

# Administrative Data

- House transaction data (Taiwan): <http://lvr.land.moi.gov.tw/homePage.action>
  - I have bought this dataset. Let me know if you want to use this data for your term paper
- Data on mortgages (US): <https://www.fhfa.gov/DataTools/Downloads/Pages/Public-Use-Databases.aspx>

# Administrative Data

- National Health Insurance Research Database (Taiwan): <http://nhird.nhri.org.tw/en/index.html>
  - It is a confidential dataset. Let me know if you want to use this data for your term paper
- Tax Return Data (Taiwan): <http://www.econ.sinica.edu.tw/FIA-Data>
  - It is a confidential dataset. I do not recommend you to use this data for your term paper because you need to spend a lot of time applying for it.



- Sports Reference (US): <https://www.sports-reference.com/>
  - This website contains many statistics about sports

# Administrative Data

- Yelp Open Dataset (US): <https://www.yelp.com/dataset>
  - Yelp is a company, which publishes crowd-sourced reviews about businesses
  - A subset of data on Yelp's businesses, reviews, and users

- Google Mobility Data: <https://www.google.com/covid19/mobility/>
  - The data shows how visits to places, such as grocery stores and parks, are changing in each geographic region during COVID-19 pandemic

- Apple Mobility Data: <https://covid19.apple.com/mobility>
  - Similar to Google Mobility Data

## Firm-Level Data

# Firm-Level Data

- Taiwan Economic Journal (TEJ): <http://www.finasia.biz/ensite/>
  - TEJ is a database of financial and market information on **listed company at stock exchange market** in Taiwan, Hong Kong or China
  - Business school should have bought this database
- Compustat (US): <https://marketintelligence.spglobal.com/client-solutions/>
  - Compustat is a database of financial and market information on active and inactive global companies throughout the world
  - Business school should have bought this database

## Country-level data

# Country-level data

- Penn World Table version 9.0:  
<https://www.rug.nl/ggdc/productivity/pwt/>
- World Wealth & Income Database  
<http://wid.world/>
- Work Bank Databank:  
<https://databank.worldbank.org/data/home.aspx>
- OECD data  
<https://data.oecd.org/>
- eurostat data  
<https://ec.europa.eu/eurostat/data/database>
- FRED Economic data  
<https://fred.stlouisfed.org/tags/series>



# Government open data

# Government open data

- US open data:  
<https://www.data.gov/local>
- Taiwan open data:  
<https://data.gov.tw/>
- Taiwan Agri open data:  
<https://data.coa.gov.tw/index.aspx>
- UK open data:  
<http://www.data.gov.uk>
- Japan open data:  
<https://www.data.go.jp/>

# Government open data

- 社會經濟資料服務平台：  
<https://segis.moi.gov.tw/STATCloud/Profile>
  - Provides socio-economic variables for various administrative regions in Taiwan

# Geographical data

# Geographical data

- GADM

<https://gadm.org/index.html>

- It provides map (shapefile) for administrative areas of all countries, at all levels of sub-division

- Geographically based Economic data

<https://gecon.yale.edu/g-econ-project-yale-university->

- It develops a geophysically scaled economic data set

# Geographical data

- Google Earth Engine  
<https://earthengine.google.com/>
  - It collects settlight image data from the following data sources
- Landsat (土地利用)  
<http://landsat.usgs.gov/>
  - Settlight image for land use, land cover, and land condition since 1972
- MODIS (大氣狀況)  
<https://modis.gsfc.nasa.gov/>
  - Settlight image for atmosphere or ocean since 1999

# Geographical data

- DMSP-OLS (夜間光照)

<https://eogdata.mines.edu/products/dmsp/>

- Settlight image for nightlight from 1992-2013 (annual data)

- VIIRS (夜間光照)

<https://eogdata.mines.edu/products/vnl/>

- Settlight image for nightlight since 2012 (monthly data)

# Geographical data

- Dynamic World V1 (土地利用)  
[https://developers.google.com/earth-engine/datasets/catalog/GOOGLE\\_DYNAMICWORLD\\_V1](https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_DYNAMICWORLD_V1)
  - A 10m near-real-time Land Use/Land Cover dataset that includes class probabilities and label information for nine classes
- ERA5-Land Hourly - ECMWF Climate Reanalysis (温度)  
[https://developers.google.com/earth-engine/datasets/catalog/ECMWF\\_ERA5\\_LAND\\_HOURLY](https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_HOURLY)
  - Providing an accurate description of the climate of the past



# Geographical data

- GPM: Global Precipitation Measurement (GPM) v6 (雨量)  
[https://developers.google.com/earth-engine/datasets/catalog/NASA\\_GPM\\_L3\\_IMERG\\_V06](https://developers.google.com/earth-engine/datasets/catalog/NASA_GPM_L3_IMERG_V06)
  - Provide observations of rain and snow worldwide every three hours
- MERIT DEM: Multi-Error-Removed Improved-Terrain DEM (地形坡度)  
[https://developers.google.com/earth-engine/datasets/catalog/MERIT\\_DEM\\_v1\\_0\\_3](https://developers.google.com/earth-engine/datasets/catalog/MERIT_DEM_v1_0_3)
  - A high accuracy map of global terrain elevations

# Geographical data

- Sentinel-5P NRTI AER (空氣品質)  
<https://developers.google.com/earth-engine/datasets/tags/air-quality>
- Provide near real-time high-resolution imagery of air quality index

- Startup Cartography Project (創業地理分布資料)  
<https://www.startupcartography.com/home>
  - Utilizes population-level business registration records
  - create innovative measures of entrepreneurial quantity, quality, and performance at various geographic and temporal levels

# Geographical data

- Social and Economic Statistics (台灣社會經濟統計地理資訊網)  
<https://moisagis.moi.gov.tw/moiap/gis2010/index.cfm>
- Provides various socio-economic variables for different regions in Taiwan (townships, smallest statistical areas)

- 臺灣氣候變遷推估資料

[https://tccip.ncdr.nat.gov.tw/ds\\_intro.aspx](https://tccip.ncdr.nat.gov.tw/ds_intro.aspx)

- Constructs long-term climate data for various regions in Taiwan using various observational data and estimation methods

# Journal Data Archives

- American Economic Review:  
<https://www.aeaweb.org/journals/aer/issues>
- AEJ Economic Policy:  
<https://www.aeaweb.org/journals/pol/issues>
- AEJ Applied Economics:  
<https://www.aeaweb.org/journals/app/issues>

# Other data source

- Google Trend:  
<https://trends.google.com/trends/>
- Kaggel:  
<https://www.kaggle.com/datasets>
- Jerry Smith dataset collection:  
<https://datascientistinsights.com/2013/02/02/data-monet>
- Our World in Data:  
<https://ourworldindata.org/>