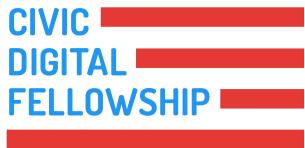


MODERNIZING THE INTERNATIONAL DATABASE (IDB)

U.S. Census Bureau (Population Division)

Supervisor: John T (Tom) Fitzwater — Chief, Demographic & Economic Studies



BELLA CHANG
UC Berkeley
B.A., Data Science

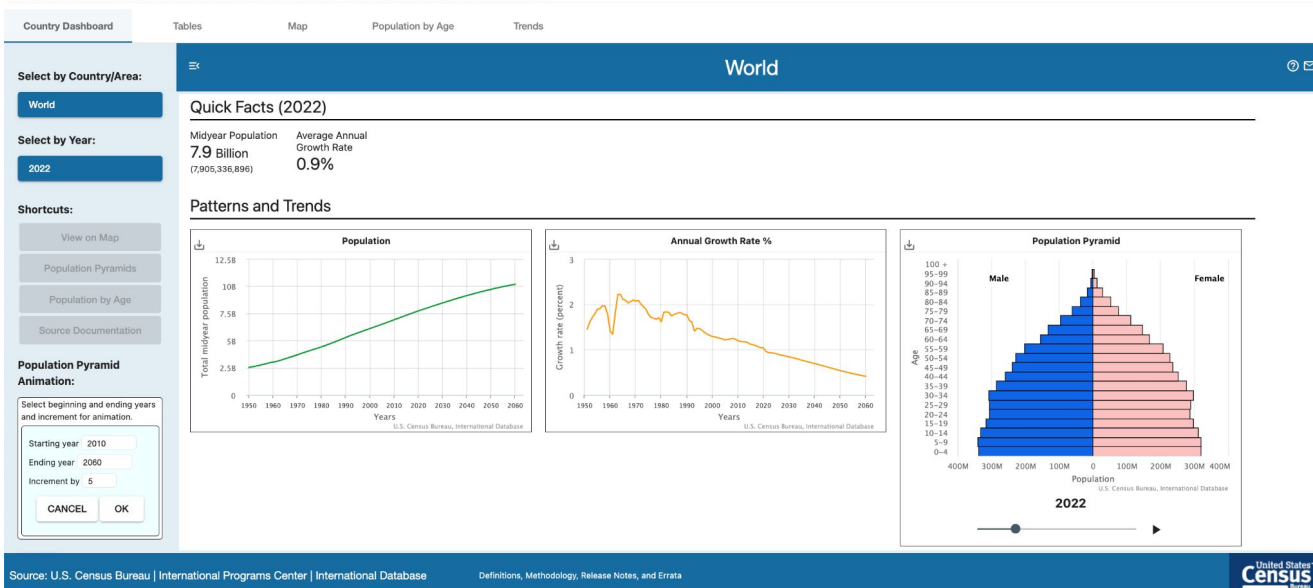
BACKGROUND

- **International Data Base (IDB)** = “U.S. government’s [source of population estimates and projections \(E&Ps\)](#) for over 200 countries and areas of the world to the year 2100”
 - **E & P** = estimates and projections (estimates refers to calculation of population in the [present/past](#), projections refers to calculations of [future](#))
- Data for E & P collected from **DAPPS** = Demographic Analysis & Population Projection System Software

BACKGROUND

International Database (IDB)

Population estimates and projections for 227 countries and areas. Visit the [IDB Help page](#) for more information.



IDB WEB TOOL

BACKGROUND

International Database (IDB)

Population estimates and projections for 227 countries and areas. Visit the [IDB Help page](#) for more information.

Country Dashboard **Tables** Map Population by Age Trends

Report: Demographic Overview

Custom Report Columns: Select columns...

Select by Country/Area: --All--

Select by Year: Select years and/or year range.
☒ Use specified year(s)
Years (comma separated): 2022
☐ Use specified range
1950 2100
Increment by 1 year(s)
CANCEL OK
Reset

Notes: "...-" means that data are not available for that year. Click the country/area name to see data sources.

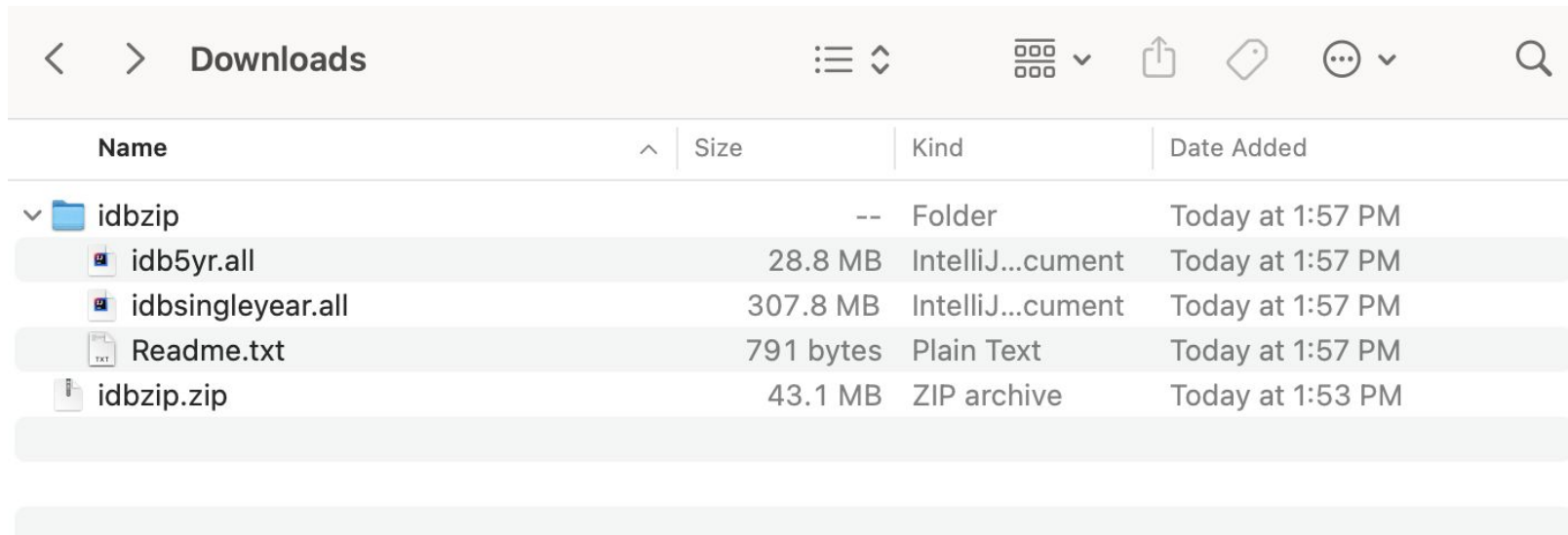
Demographic Overview

Row	GENC	Country/Area Name	Year	Population	Annual Growth Rate %	Area (sq km)	Density (per sq km)	Total Fertility Rate	Life Expectancy at Birth	Under-5 Mortality Rate
1	AF	Afghanistan	2022	38,346,720	2.30	652,230	58.8	4.62	53.7	154.1
2	AL	Albania	2022	3,095,344	0.22	27,398	113.0	1.54	79.5	12.7
3	DZ	Algeria	2022	44,179,884	1.34	2,381,740	18.5	2.51	78.0	22.9
4	AS	American Samoa	2022	45,443	-1.92	198	229.5	2.21	75.3	12.7
5	AD	Andorra	2022	85,560	-0.10	468	182.8	1.45	83.4	4.1
6	AO	Angola	2022	34,795,287	3.36	1,246,700	27.9	5.83	62.1	90.0
7	AI	Anguilla	2022	18,741	1.80	91	205.9	1.72	82.2	4.5
8	AG	Antigua and Barb...	2022	100,335	1.15	443	226.5	1.95	77.8	17.3
9	AR	Argentina	2022	46,245,668	0.82	2,736,690	16.9	2.18	78.3	11.0
10	AM	Armenia	2022	3,000,756	-0.38	28,203	106.4	1.65	76.1	13.8
11	AW	Aruba	2022	122,320	1.14	180	679.6	1.83	78.0	14.0
12	AU	Australia	2022	26,141,369	1.25	7,682,300	3.4	1.73	83.1	3.6
13	AT	Austria	2022	8,913,088	0.32	82,445	108.1	1.51	82.3	3.9
14	AZ	Azerbaijan	2022	10,353,296	0.67	82,629	125.3	1.86	74.2	26.1
15	BS	Bahamas, The	2022	355,808	0.82	10,010	35.5	1.98	76.1	15.3
16	BH	Bahrain	2022	1,540,558	0.88	760	2,027.0	1.67	79.9	12.1
17	BD	Bangladesh	2022	165,650,475	0.93	130,170	1,272.6	2.09	74.7	37.4
18	BB	Barbados	2022	302,674	0.26	430	703.9	1.70	78.6	12.1
19	BY	Belarus	2022	9,413,505	-0.31	202,900	46.4	1.51	74.3	5.2
20	BE	Belgium	2022	11,847,338	0.57	30,278	391.3	1.77	81.9	3.9
21	BZ	Belize	2022	412,387	1.64	22,806	18.1	2.62	75.8	12.9
22	BJ	Benin	2022	13,754,688	3.34	110,622	124.3	5.43	62.2	84.3
23	BM	Bermuda	2022	72,337	0.34	54	1,339.6	1.90	82.0	
24	BT	Bhutan	2022	867,775	0.97	38,394	22.6	1.79	72.3	

Is this page helpful? ☒ Yes ☐ No

IDB WEB TOOL

BACKGROUND



Downloads				
Name	Size	Kind	Date Added	
idbzip	--	Folder	Today at 1:57 PM	
idb5yr.all	28.8 MB	IntelliJ...cument	Today at 1:57 PM	
idbsingleyear.all	307.8 MB	IntelliJ...cument	Today at 1:57 PM	
Readme.txt	791 bytes	Plain Text	Today at 1:57 PM	
idbzip.zip	43.1 MB	ZIP archive	Today at 1:53 PM	

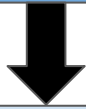
IDB DOWNLOADABLE FILES (PUBLIC)

CURRENT ISSUES

- Uses **old programming language and software** to pull inputs to database
 - => *Little to no familiarity, difficult to alter*
- File **format needs updating**, to be capable of holding all data and future data
 - => *Hard to query*
- **Many sources** of data/inputs
 - => *Difficult to keep organized*
- **Difficult to visualize differences** between versions of IDB, IDB and UN

OBJECTIVES

Create an efficient workflow from various input files to one database



Design and build schema
(‘blueprint’) to
organize all data
into one container



Write R script
based off schema



Create
visualization
dashboard to show
differences
between data
sources

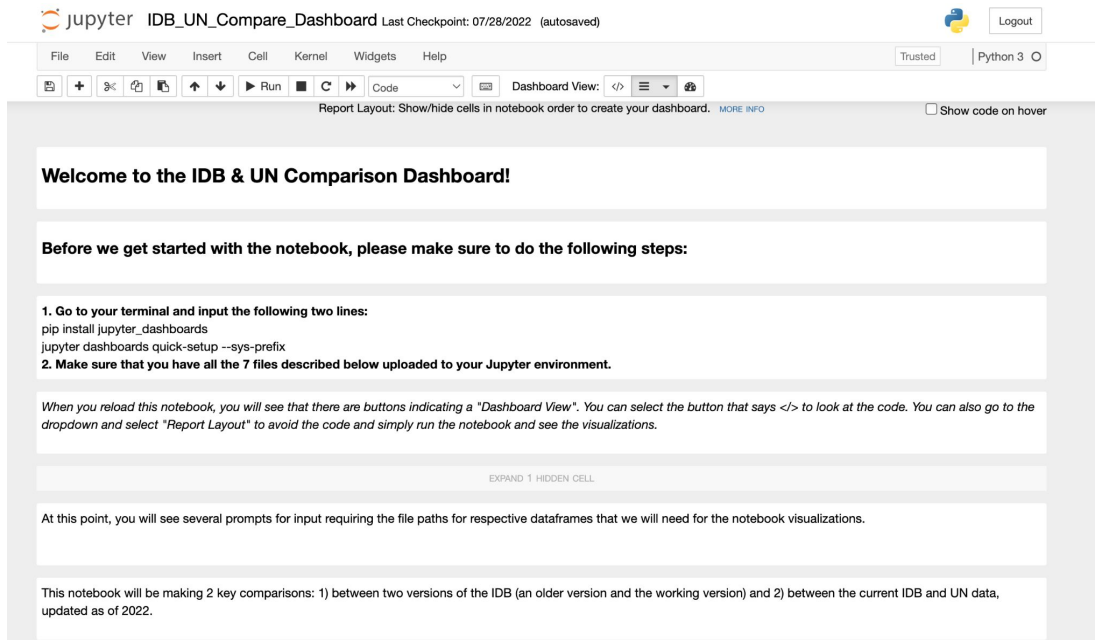
IMPACT

- Accomplishment of objectives will create:
 - **Faster analysis on the IDB** for future census review
 - **Better organization** of database
 - **Data inputs and process conducted in R**, creating a seamless tradeoff between the steps
 - **Code is much more intuitive** and is **easier to query**, making it easier for future programmers
- Main stakeholders: IPC analysts & programmers

DELIVERABLE (DASHBOARD)

- Jupyter Notebook dashboard
- Compare certain key features observed at census review
 - Total population, true fertility rate (TFR), life expectancy, etc.
- 2 comparisons: 1) between IDB releases and 2) between IDB recent release and UN recent release
- Options to look at/alter code and run notebook without referencing code

DELIVERABLE (DASHBOARD)



The screenshot shows a Jupyter Notebook titled "IDB_UN_Compare_Dashboard" with a last checkpoint of "07/28/2022 (autosaved)". The interface includes a top menu bar with options like File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. Below the menu is a toolbar with icons for file operations, running, and viewing. The main content area displays a series of text blocks: a welcome message, instructions on how to get started, a list of two steps (installing the dashboard and uploading files), a note about reloading the notebook, an "EXPAND 1 HIDDEN CELL" button, and a final note about the comparisons being made.

Welcome to the IDB & UN Comparison Dashboard!

Before we get started with the notebook, please make sure to do the following steps:

- 1. Go to your terminal and input the following two lines:**
`pip install jupyter_dashboards`
`jupyter dashboards quick-setup --sys-prefix`
- 2. Make sure that you have all the 7 files described below uploaded to your Jupyter environment.**

When you reload this notebook, you will see that there are buttons indicating a "Dashboard View". You can select the button that says </> to look at the code. You can also go to the dropdown and select "Report Layout" to avoid the code and simply run the notebook and see the visualizations.

EXPAND 1 HIDDEN CELL

At this point, you will see several prompts for input requiring the file paths for respective dataframes that we will need for the notebook visualizations.

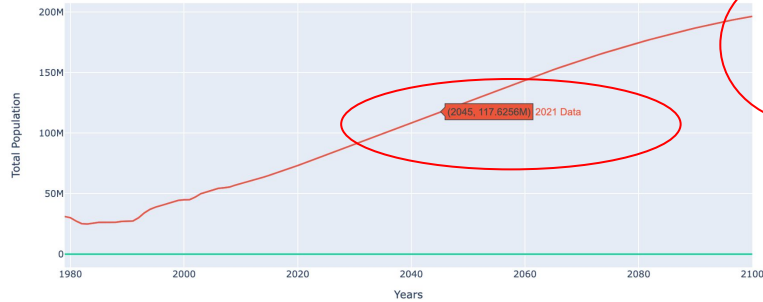
This notebook will be making 2 key comparisons: 1) between two versions of the IDB (an older version and the working version) and 2) between the current IDB and UN data, updated as of 2022.

DELIVERABLE (DASHBOARD)

TOTAL POPULATION

Country: Afghanistan

Total Population for Afghanistan from 1979 to 2100

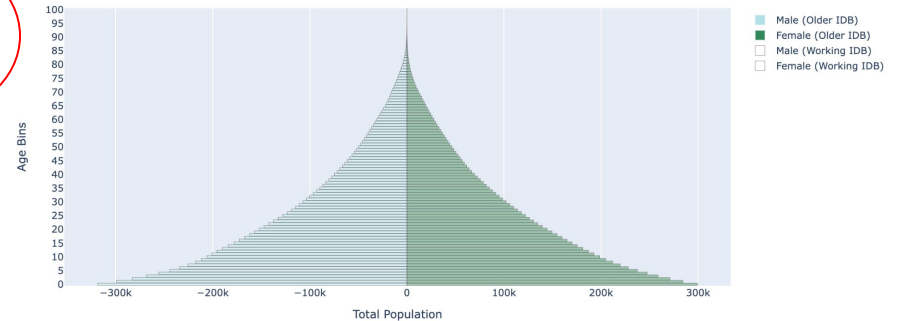


POPULATION PYRAMID

Country: Afghanistan

Year: 1979

Population Pyramid for Afghanistan in 1979



DELIVERABLE (DASHBOARD)

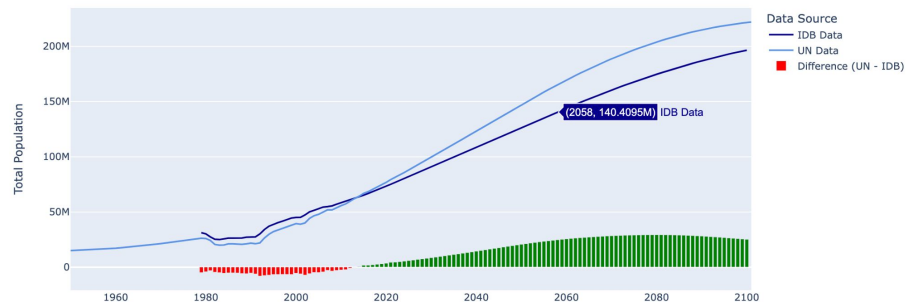
2) Visualizations between IDB & UN (updated 2022)

This section compares between the newest working version of the IDB and the newest working version of UN data, which are both indexed by single years and both updated as of July 2022. Again, these visualizations will look at total population, life expectancy, and TFR, and will also map some of the patterns of population distribution through population pyramids.

TOTAL POPULATION

Country

Total Population for Afghanistan

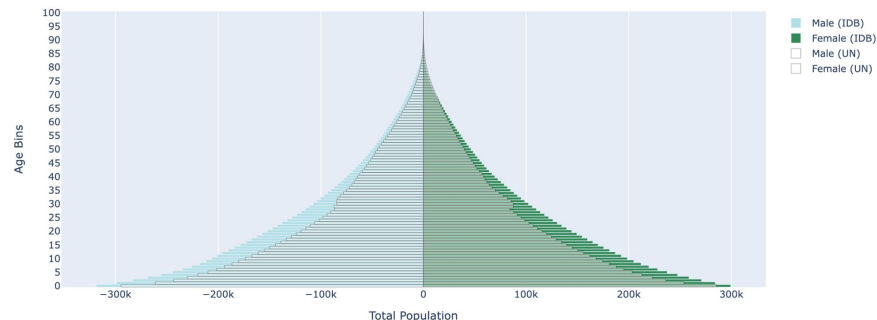


POPULATION PYRAMID

Country

Year

Population Pyramid for Afghanistan in 1979



DELIVERABLE (DASHBOARD)

TABLES

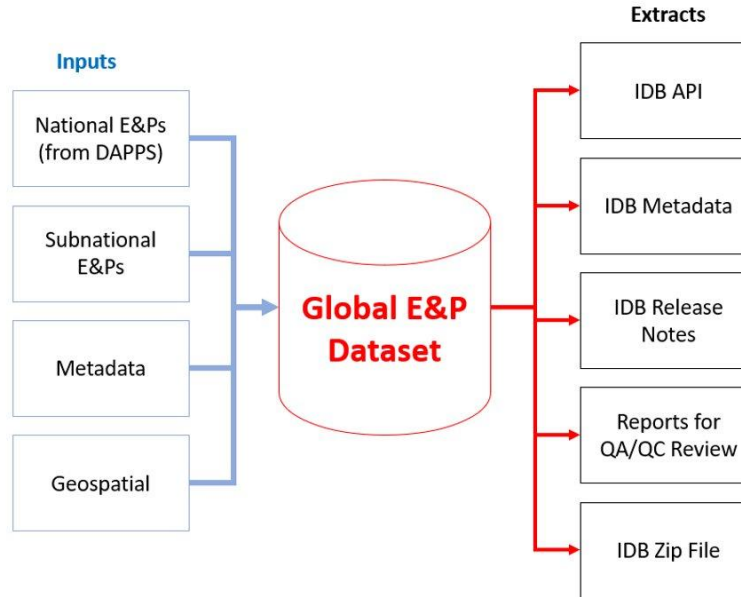
year

2094

	GENC	NAME	TIME	POP_IDB	AREA_KM2	E0_IDB	POP_DENS_IDB	TFR_IDB	GR_IDB	POP_UN	POP_DENS_UN	TFR_UN	GR_UN	E0_UN	POP_DIFF	POP_PERCENT
0	AD	Andorra	2094	53260	47268	92.66	113.8	1.7000	-0.625	123578	131.1851	1.5040	-0.428	92.2616	-70318	-79.528156
1	AE	United Arab Emirates	2094	14958245	8443600	90.99	178.9	1.6799	0.227	27300306	192.6687	1.5826	0.496	90.6476	-12342061	-58.412135
2	AF	Afghanistan	2094	95492798	65875230	76.97	146.4	2.0900	0.521	217962684	168.1099	1.9005	0.315	76.6453	-122469886	-78.141805
3	AG	Antigua and Barbuda	2094	136476	44743	90.31	308.1	1.7000	0.015	159564	180.7273	1.6288	-0.669	89.2675	-23088	-15.597892
4	AI	Anguilla	2094	38094	9191	92.20	418.6	1.7000	0.564	20192	114.2670	1.5376	-0.806	88.1188	17902	61.428130
...
186	WS	Samoa	2094	276516	284921	89.08	98.0	1.7373	-0.117	871051	154.0986	2.0771	0.262	83.4368	-594535	-103.616608
187	YE	Yemen	2094	55130039	53324768	85.02	104.4	1.7408	-0.051	147085481	139.4162	1.8474	0.176	77.9062	-91955442	-90.947957
188	ZA	South Africa	2094	71130815	122661470	82.05	58.6	1.7168	-0.077	150753826	61.6965	1.7336	-0.151	75.5937	-79623011	-71.769736
189	ZM	Zambia	2094	68074472	75083198	84.45	91.6	2.1256	0.800	122639640	82.7399	2.0792	0.612	73.2319	-54565168	-57.221951
190	ZW	Zimbabwe	2094	37828904	39071547	83.93	97.8	1.8560	0.358	70656819	91.4022	1.9155	0.173	70.9026	-32827915	-60.520245

191 rows x 16 columns

DELIVERABLE (RESTRUCTURING)



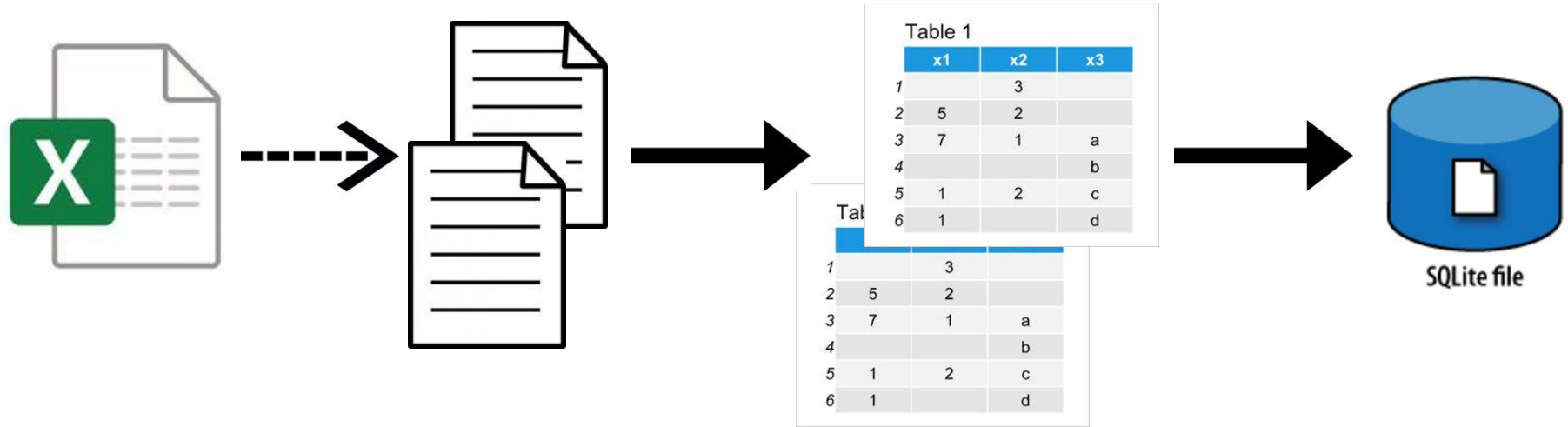
VISUAL OF PRODUCTION WORKFLOW

DELIVERABLE (RESTRUCTURING)

- Multiple input scripts converted to dataframes
- New SQLite file format for database
- New Excel schema that lays out structure of table shells for all data
- R script:
 - **Pulls data in** from input scripts
 - Uses Excel schema to **format SQLite database**
 - **Inputs data** into table shells



DELIVERABLE (RESTRUCTURING)



FUTURE STEPS

- Database work
 - Work on pulling various extracts from database
 - Continue finding strategies to minimize space within SQLite database
- Visualization dashboard work
 - Development on the front-end
 - Creating static reports of the dashboard

REFLECTION

- Academic experience
 - Highly informative about the details (statistics and data aggregation/collection) behind Census' work
 - Was able to see a more specific area of data science
 - Gained insight on data contextualization
- Personal experience
 - Educational to learn alongside two other talented CDF's
 - Fun to work in a new workplace environment outside of school :-)