MODERNIZING THE INTERNATIONAL DATABASE (IDB)

U.S. Census Bureau (Population Division)

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

CIVIC DIGITAL FELLOWSHIP

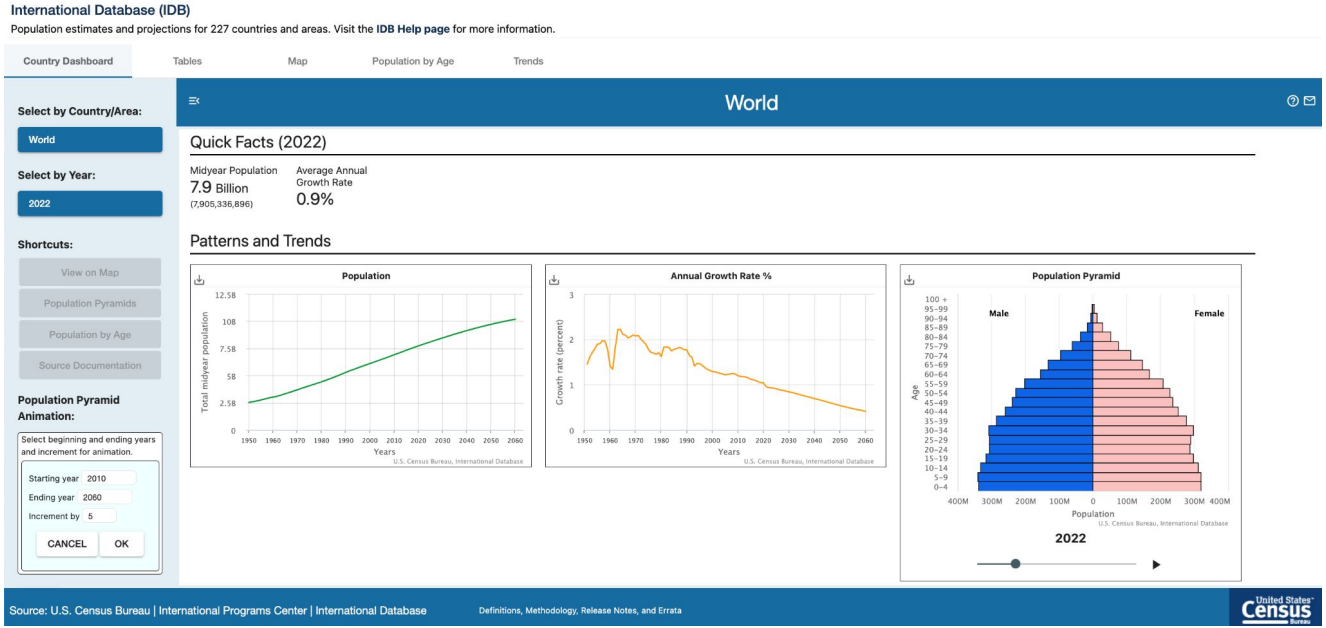
United States™ Census Bureau

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

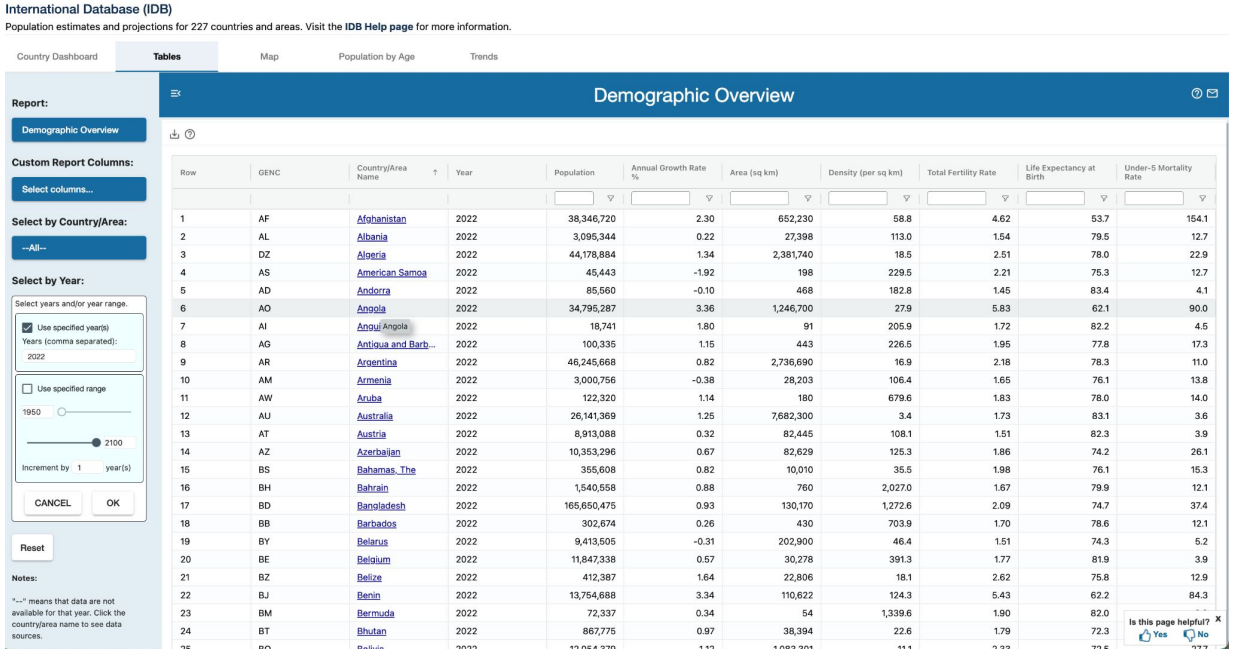
BACKGROUND



IDB WEB TOOL

CIVIC DIGITAL FELLOWSHIP

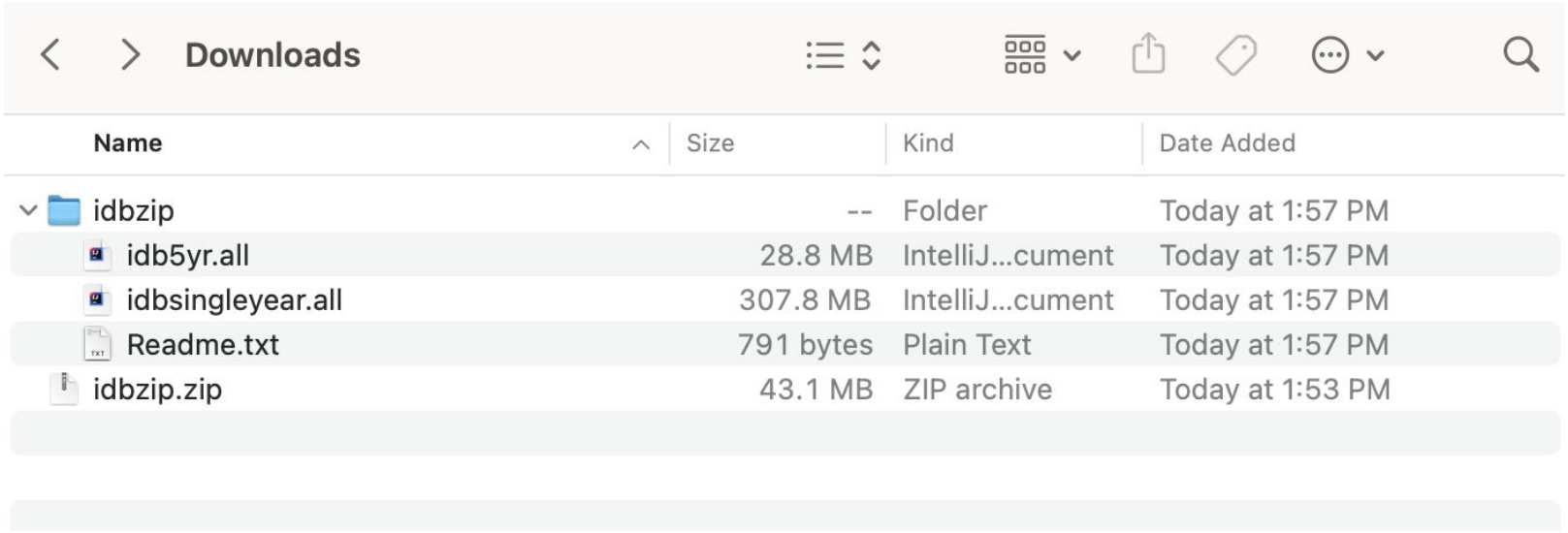
BACKGROUND



IDB WEB TOOL

CIVIC DIGITAL FELLOWSHIP

BACKGROUND



IDB DOWNLOADABLE FILES (PUBLIC)

CIVIC DIGITAL FELLOWSHIP

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

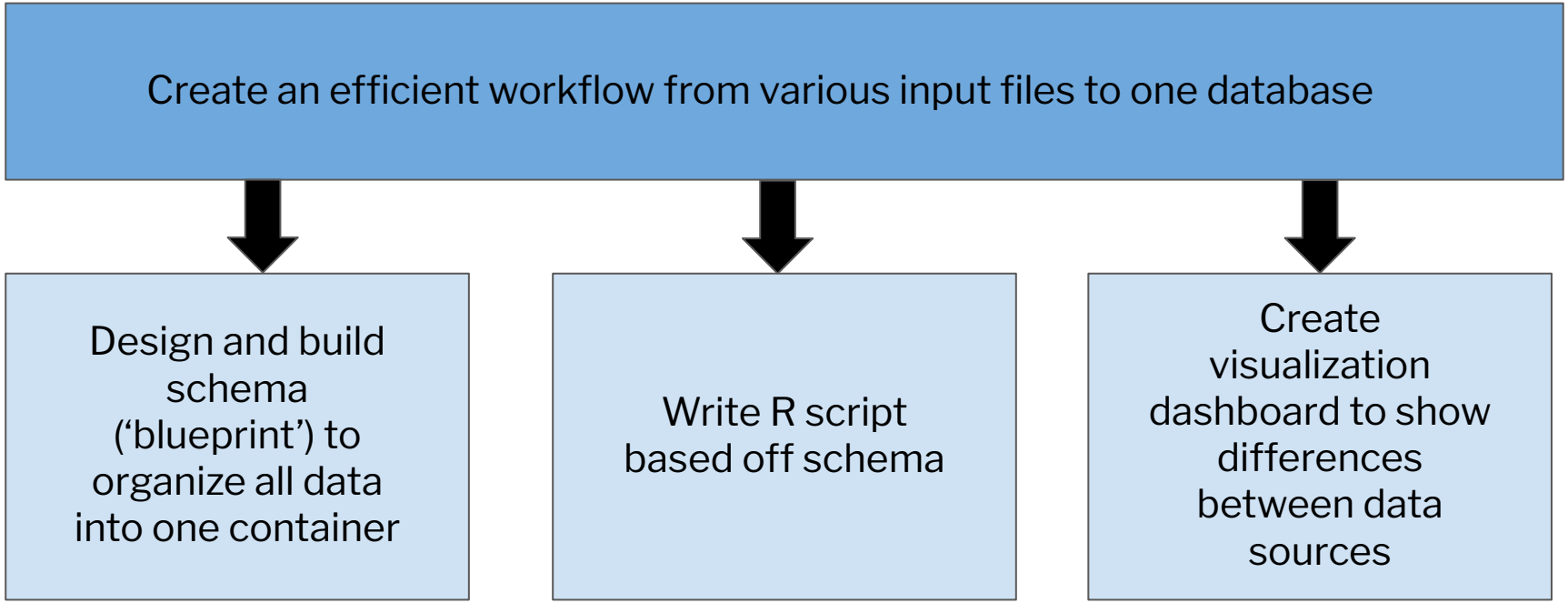
  => *Hard to query*

- **Many sources** of data/inputs

  => *Difficult to keep organized*

- **Difficult to visualize differences** between versions of IDB, IDB and UN

OBJECTIVES



CIVIC DIGITAL FELLOWSHIP

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

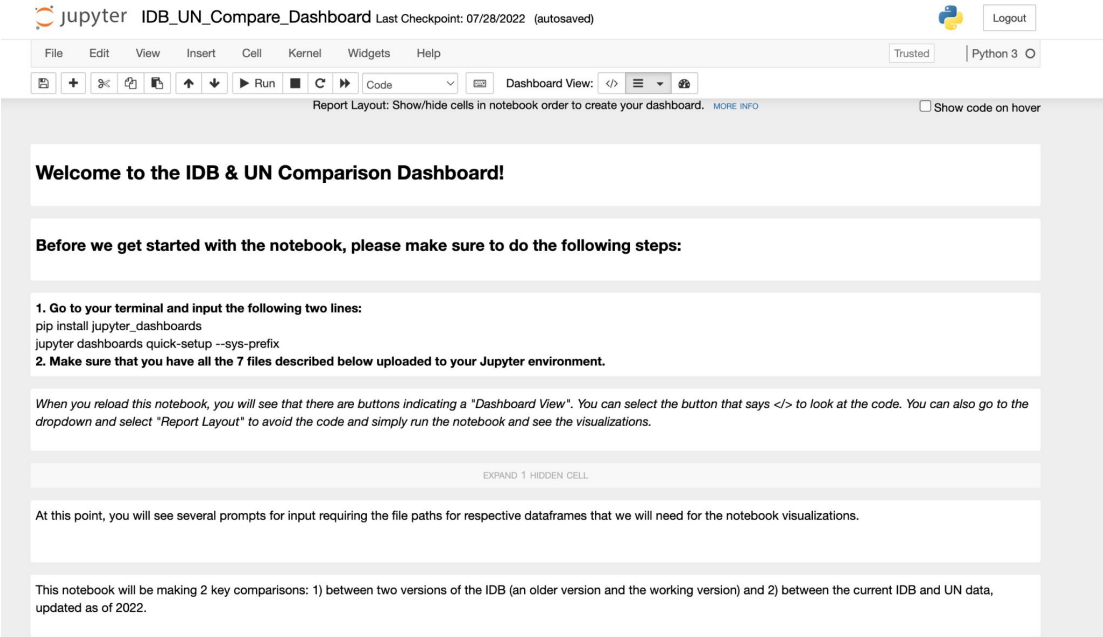
CIVIC DIGITAL FELLOWSHIP

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

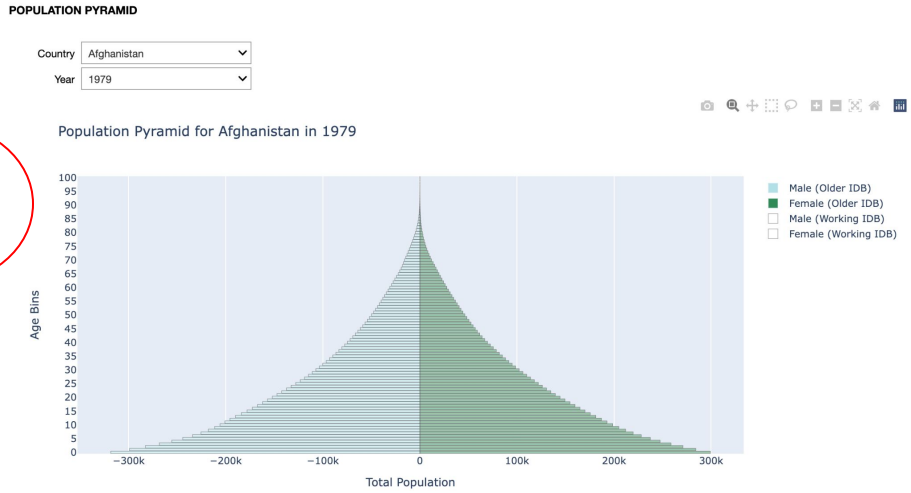
CIVIC DIGITAL FELLOWSHIP

DELIVERABLE (DASHBOARD)



CIVIC DIGITAL FELLOWSHIP

DELIVERABLE (DASHBOARD)



CIVIC DIGITAL FELLOWSHIP

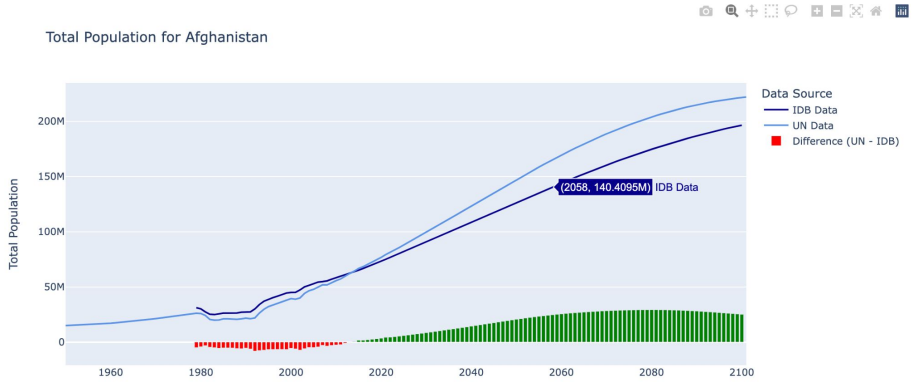
DELIVERABLE (DASHBOARD)

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

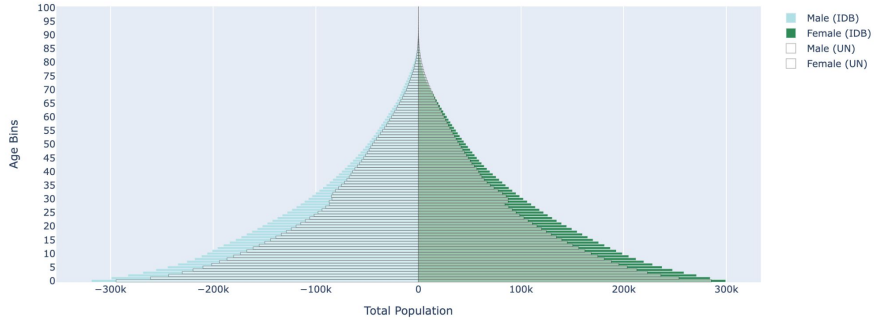
Country: Afghanistan



POPULATION PYRAMID

Country: Afghanistan

Year: 1979



CIVIC DIGITAL FELLOWSHIP

DELIVERABLE (DASHBOARD)

TABLES

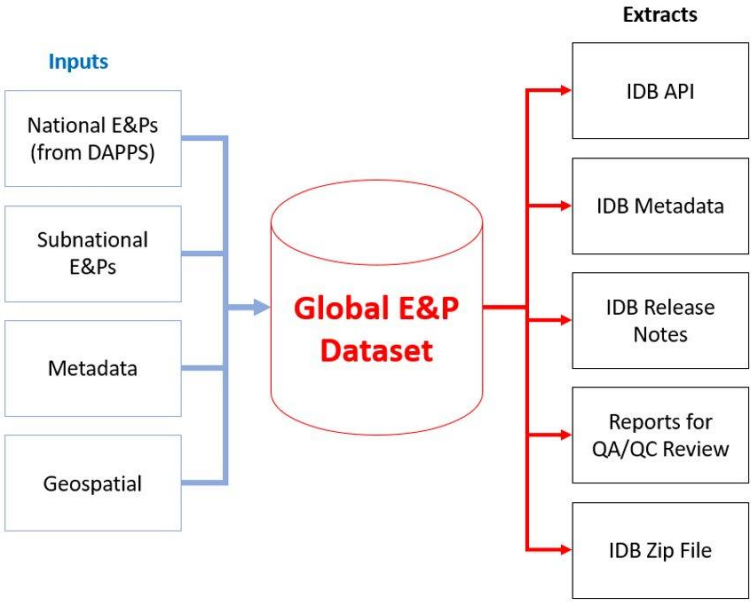
year 2094

	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	-60.520249

191 rows × 16 columns

CIVIC DIGITAL FELLOWSHIP

DELIVERABLE (RESTRUCTURING)



VISUAL OF PRODUCTION WORKFLOW

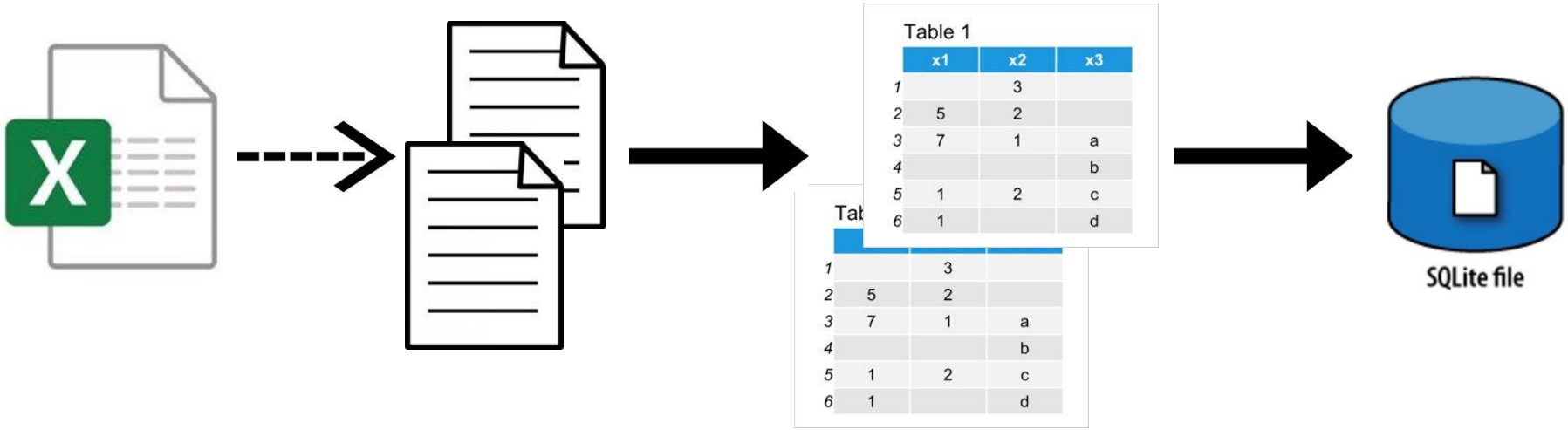
CIVIC DIGITAL FELLOWSHIP

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

CIVIC DIGITAL FELLOWSHIP

DELIVERABLE (RESTRUCTURING)



CIVIC DIGITAL FELLOWSHIP

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 :-)

CIVIC DIGITAL FELLOWSHIP