naphsis | ANNUAL MEETING

DALLAS, TX • MAY 12-14

220 Years of NYC Vital Statistics: How Race Data Changed with Policy and Demographic Shifts

May 2025 | Muriel Silin, MPH, New York City Department of Health and Mental Hygiene, Bureau of Vital Statistics

Historical Significance of Vital Events

- Registration of births and deaths is a human right. Vital event registration is critical to create and document a person's legal identity.
  - UN Sustainable Development Goal Targets 16.9 and 17.19 call for 100% registration of all births in the world and 80% registration of all deaths.
- NYC has consistently collected vital events data since 1804; the current NYC Health Department was founded in 1805[1]
- Race data has been collected for NYC births and deaths since 1854
- U.S. Census started collecting race data in 1790
- How did race collection practices change over time? And why?

1. Protecting Public Health in New York City: 200 Years of Leadership. New York City Department of Health and Mental Hygiene. April 2005. https://www.nyc.gov/assets/doh/downloads/pdf/bicentennial/historical-booklet.pdf. Accessed December 4th, 2024.
Home — UN Legal Identity Agenda

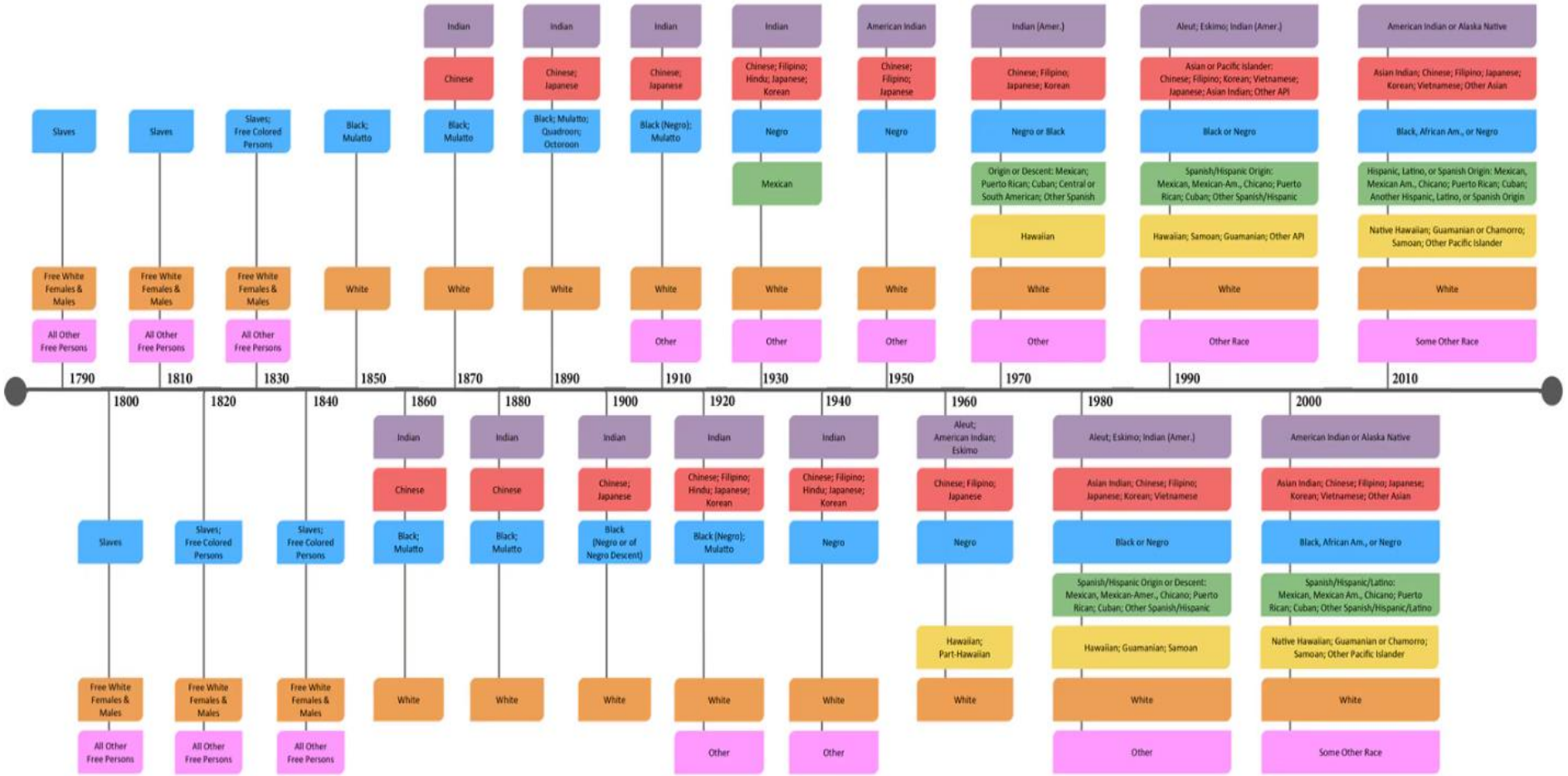
How Did Race Data Take on New Meaning?

- The collection of race and ethnicity data changed over time
- U.S. Census (which originally housed national vital statistics compilation efforts) changed race and ethnicity categories as policies and political changes took place which affected the data collection across the U.S.
- Over time, race and ethnicity data changed to:
  - Reinforce policies dominant at the time, including racist ideas[1,2]
  - Address current public health issues through interventions
- How did national categories for race & ethnicity change over time?
- Were NYC's categories different?

1. Pearson SJ. Birth Registration and the Administration of White Supremacy. Modern American History. 2022;5(2):117-141. doi:10.1017/mah.2022.13
2. The Washington Post. *Secret use of census info helped send Japanese Americans to internment camps in WWII.* The Washington Post. April 6th, 2018. https://www.washingtonpost.com/news/retropolis/wp/2018/04/03/secret-use-of-census-info-helped-send-japanese-americans-to-internment-camps-in-wwii/?noredirect=on. Accessed December 4th, 2024.

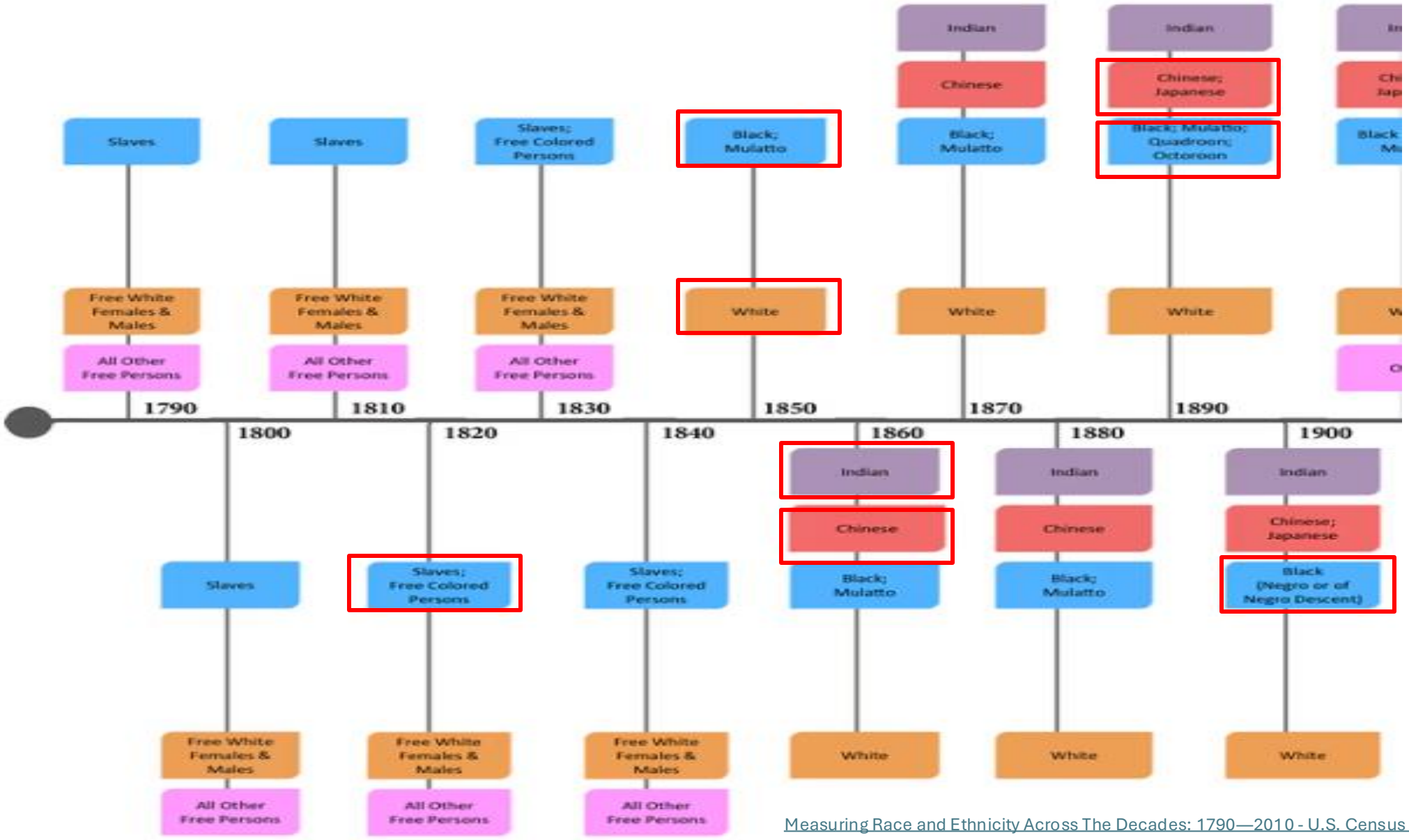
Measuring Race and Ethnicity Across the Decades: 1790–2010

Mapped to 1997 U.S. Office of Management and Budget Classification Standards



Gibson, Campbell, and Kay Jung. 2002. "Historical Census Statistics on Population By Race, 1790 to 1990, and By Hispanic Origin, 1790 to 1990, For The United States, Regions, Divisions, and States."
Humes, Karen, and Howard Hogan. 2009. "Measurement of Race and Ethnicity in a Changing, Multicultural America."
Humes, Karen R., Nicholas A. Jones, and Roberto R. Ramirez. 2011. "Overview of Race and Hispanic Origin: 2010."
Office of Management and Budget. 1978. "Statistical directive no. 15: Race and ethnic standards for federal agencies and administrative reporting."
Office of Management and Budget. 1997. "Revisions to the standards for the classification of federal data on race and ethnicity."
U.S. Census Bureau History Questionnaires. (2014, March 31).

US Census Race and Ethnicity, 1790 to 1900



Measuring Race and Ethnicity Across The Decades: 1790—2010 - U.S. Census Bureau

1790-1890 Foundation of Race Data

- Race data were collected for taxation, slavery, and representation
- 1790 Census categories:
  - Slaves, Free White Persons, and All Other Free Persons
- Race recorded based on Enumerator observation
- 1820 – New Census race category: ‘Free Colored Persons’
- 1840 – Last census with ‘All Other Free Persons’
- 1850 - Census only had two race categories: ‘Black/Mulatto’ and ‘White’
- 1857 – Dred Scott v. Sandford decision stated that African Americans could not be citizens
- 1860 - First use of **national origin** for U.S. Census for ‘Chinese’
  - ‘Chinese’ was added to California Census only.
  - U.S. Census added ‘Indian’ but only enumerated those who paid taxes

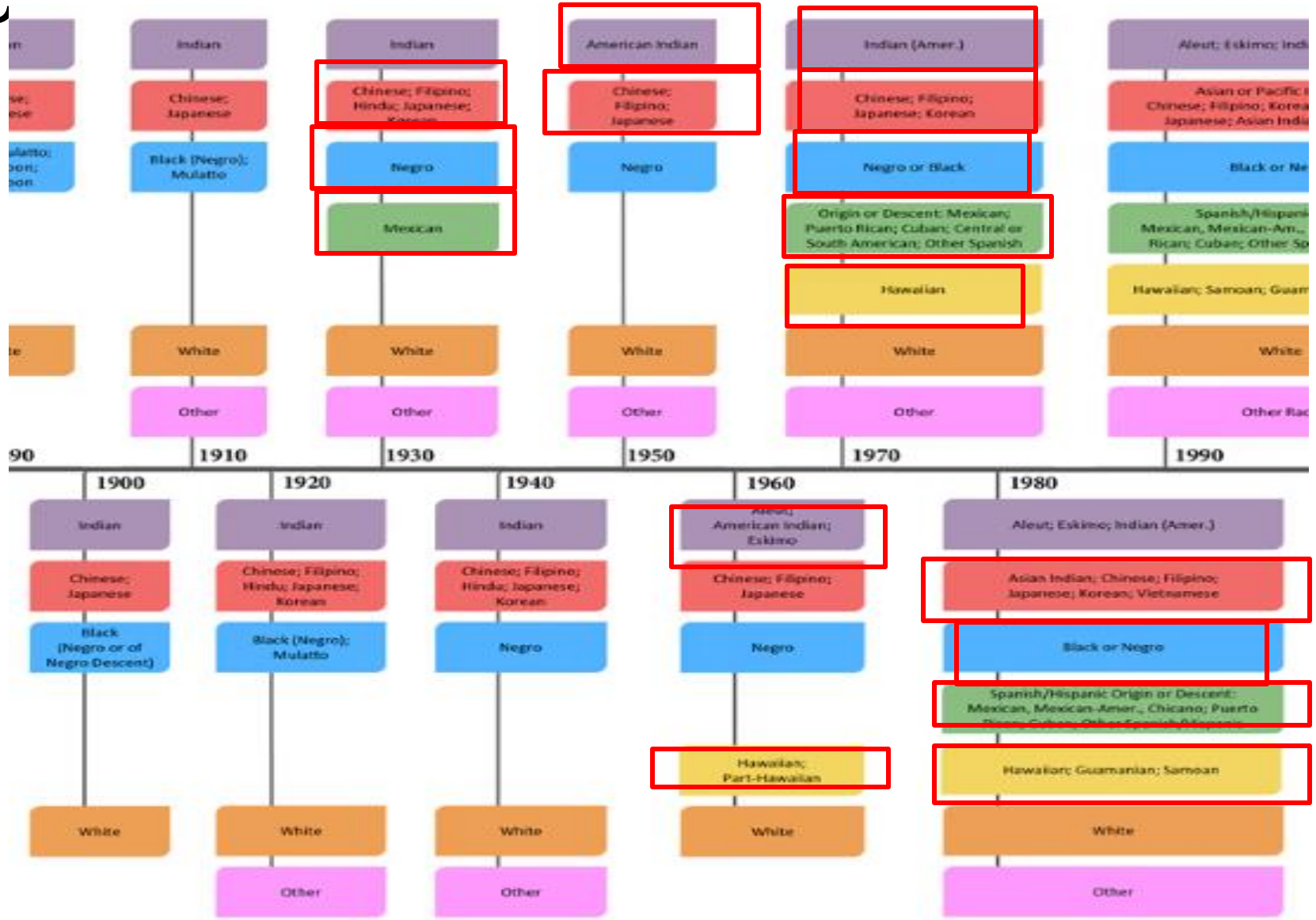
14th Amendment: Citizenship | U.S. Constitution.net
https://www.census.gov/newsroom/blogs/random-samplings/2020/08/measuring-race-and-ethnicity.html (2020a)
https://yalebooks.yale.edu/book/9780300218503/the-american-census/ (2015)
Measuring Race and Ethnicity Across the Decades: 1790—2010 - U.S. Census Bureau

Important Milestones for Race Data

- 1868 – 14th Amendment ratified – anyone born in US is citizen – based on a case of Wong Kim Ark, a Chinese American born in the US
- 1870-1880's: Increase in Chinese migrants
- 1890 – Census
  - First attempt to count all American Indians
  - Increased use of 'race science theories'
  - Congress mandated more categories to collect 'Black blood'; introduced 'quadroon', 'octoroon'
  - High demand for labor led to increased migration from Japan; 'Japanese' was added

14th Amendment & Birthright Citizenship – U.S. Constitution.net
Measuring Race and Ethnicity Across The Decades: 1790—2010 - U.S. Census Bureau

US C



Measuring Race and Ethnicity Across The Decades: 1790—2010 - U.S. Census Bureau

1900–1940 – Racial Thinking and Category Expansion

- Racial purity concepts influenced race categories
- More emphasis on race categories to support and continue segregation policies
- 1900 – ‘Negro’ used for the first time in the U.S. Census
- 1910 – ‘Other’ was first used for the Census
- 1920 – New categories
  - Korean, Filipino, and Asian Indian
  - First religious group for ‘Hindu’
- 1930 – ‘Mexican’ added as a category
  - Mexicans were previously included as ‘White’, but this category was dropped in the 1940 Census.
  - Mexican Americans and Mexico lobbied successfully to remove this category.

https://yalebooks.yale.edu/book/9780300219630/the-american-census (2015)
https://nyupress.org/9780814775466/changing-race/ (2000)
https://www.ucpress.edu/book/9780520269673/the-nature-of-race (2011)

NYC Summary of Vital Statistics, 1930, 1940

DEATHS ACCORDING TO AGE AND COLOR—YEAR 1930

	Manhattan	The Bronx	Brooklyn	Queens	Richmond	City of New York
Total, All Ages	26,608	10,865	25,886	9,701	1,828	74,888
Under One Year	2,227	997	2,740	915	151	7,030
One Year	285	139	403	123	26	976
Two Years	174	75	211	63	9	532
Three Years	99	52	164	58	5	378
Four Years	84	46	119	46	13	308
Total Under Five Years	2,869	1,309	3,637	1,205	204	9,224
5 to 9 Years	311	160	450	175	29	1,125
10 to 14 Years	228	122	300	111	27	788
15 to 19 Years	457	219	539	163	25	1,403
20 to 24 Years	846	309	723	243	39	2,160
25 to 29 Years	986	371	792	268	50	2,467
30 to 34 Years	1,186	417	903	346	60	2,912
35 to 39 Years	1,609	499	1,032	424	76	3,640
40 to 44 Years	1,761	561	1,407	495	97	4,321
45 to 49 Years	2,069	767	1,712	608	85	5,241
50 to 54 Years	2,384	915	2,018	779	146	6,242
55 to 59 Years	2,392	991	2,276	878	155	6,692
60 to 64 Years	2,615	1,140	2,667	976	166	7,564
65 to 69 Years	2,334	1,066	2,386	974	179	6,939
70 to 74 Years	1,944	908	2,042	841	174	5,909
75 to 79 Years	1,310	566	1,410	631	153	4,070
80 to 84 Years	786	331	904	346	90	2,457
85 Years and Over	521	214	688	238	73	1,734
Negroes	3,634	230	1,103	255	42	5,264
Chinese	140	4	21	3	1	169

U.S 1930 Census categories: Indian, Chinese, Filipino, Hindu, Japanese, Korean, Negro, Mexican, White, Other

Age Distribution of the Population, By Color and Sex, for New York City (Preliminary*): 1940

	Total	White	Non-white
Total, all ages	7,454,995	[illegible]	[illegible]
0-4	441,984	410,137	31,847
5-9	475,387	440,163	35,224
10-13	445,252	415,440	29,812
14-19	721,902	679,050	42,852
20-24	648,570	605,478	43,092
25-44	2,689,094	2,483,844	205,250
45-64	1,620,845	1,543,756	77,089
65 and over	411,961	398,781	13,180**
Males, all ages	3,686,034	3,462,587	223,447
0-4	225,440	209,766	15,674**
5-9	242,589	224,399	18,190**
10-13	223,266	208,850	14,416**
14-19	358,723	339,087	19,636**
20-24	301,156	285,498	15,658**
25-44	1,313,584	1,218,507	95,077
45-64	834,395	795,199	39,196
65 and over	186,881	181,281	5,600**
Females, all ages	3,768,961	3,514,062	254,899
0-4	216,544	200,371	16,173**
5-9	232,798	215,764	17,034**
10-13	221,986	206,590	15,396**
14-19	363,179	339,963	23,216
20-24	347,414	319,980	27,434
25-44	1,375,510	1,265,337	110,173
45-64	786,450	748,557	37,893
65 and over	225,080	217,500	7,580**

* Based upon a 5 per cent sample.
** Numbers less than 20,000 are based on relatively small samples and are subject to error.

Deaths, New York City, 1940, By Age and Color

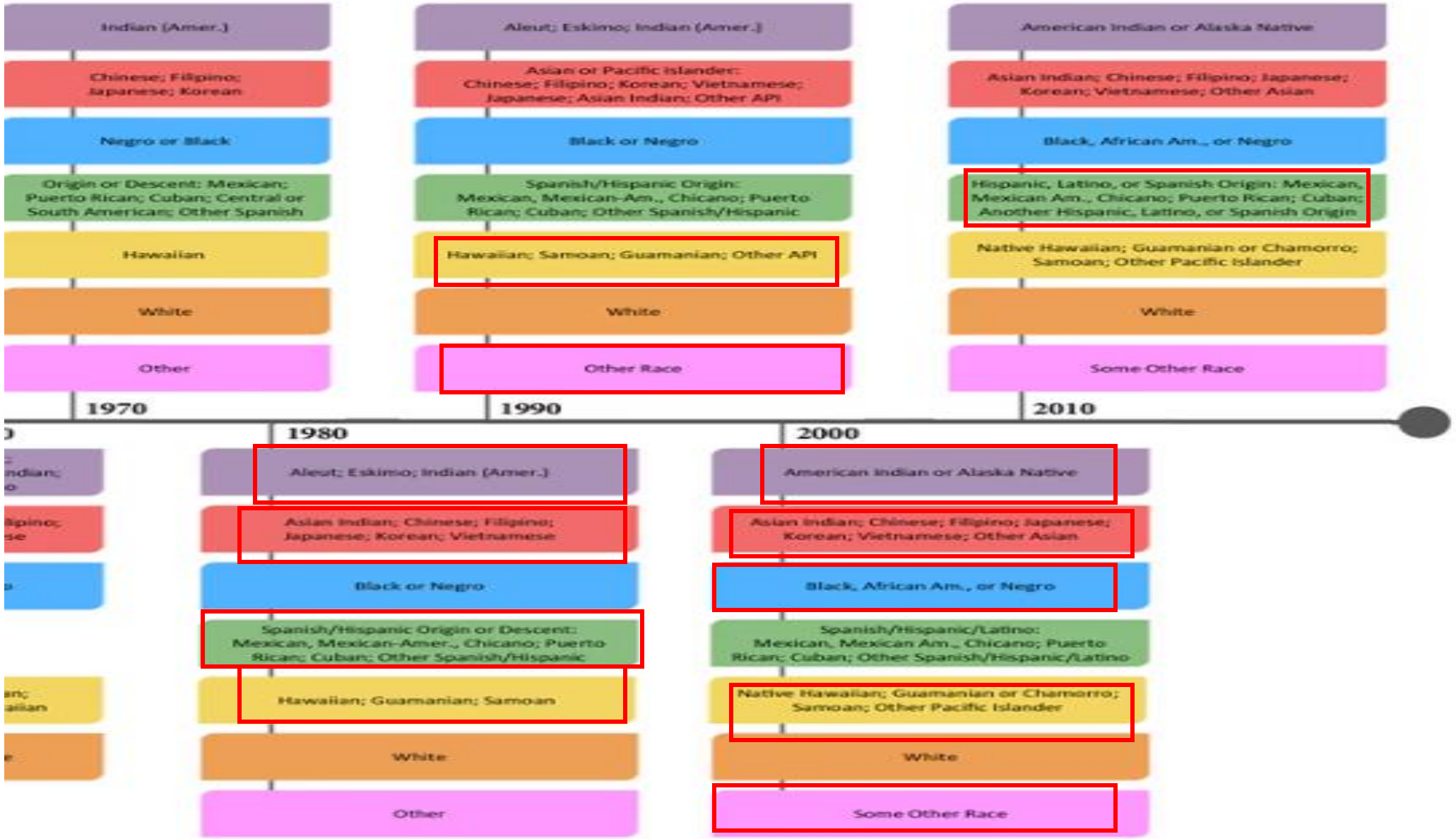
Age (in Years)	Total	White	Negro	Other Colored
All Ages	76,008	70,177	5,616	215
Under 5	4,402	3,873	514	15
5-9	401	359	42	....
10-14	465	394	67	4
15-19	716	566	149	1
20-24	1,104	849	253	2
25-29	1,429	1,115	307	7
30-34	1,910	1,483	422	5
35-39	2,536	2,071	452	13
40-44	3,723	3,163	533	27
45-49	5,135	4,539	569	27
50-54	6,695	6,088	577	30
55-59	7,575	7,105	442	28
60-64	8,867	8,444	396	27
65-69	9,091	8,741	338	12
70-74	8,631	8,385	237	9
75-79	6,551	6,392	153	6
80-84	4,151	4,052	98	1
85-89	1,897	1,859	37	1
90-94	567	550	17	....
95-99	135	127	8	....
100 and Over	27	22	5	....
Unknown	...	....	....	....

U.S 1940 Census categories: Indian, Chinese, Filipino, Hindu, Japanese, Korean, Negro, Mexican, White, Other

1950s–1970s – Policy Reform and Civil Rights

- 1950 – ‘Korean’ and ‘Hindu’ removed for Census; ‘Indian’ changed to ‘American Indian’
- 1960 – Self response used instead enumerator’s observation; first Census with ‘Alaska Native and Pacific Islander’
- Civil Rights Acts (1964, 1965) required better race data to address discrimination
- 1970 – Census changes
  - New Hispanic question: Mexican, Cuban, Puerto Rican, Central or South American, Other Spanish, None of these
  - ‘Korean’ added back because of increased migration
  - ‘Negro or Black’ used – Influenced by the civil rights movement
- OMB Directive No. 15 (1977): Introduced standard categories for race/ethnicity
  - White, Black, American Indian/Alaska Native, Asian/Pacific Islander
  - Hispanic should be collected as **ethnicity**, not race

https://obamawhitehouse.archives.gov/omb/fedreg_directive_15/(1977)
https://www.census.gov/newsroom/blogs/random-samplings/2020/08/measuring-race-and-ethnicity.html (2015)



[Measuring Race and Ethnicity Across The Decades: 1790—2010 - U.S. Census Bureau]

1980s–1990s – Immigration & Identity Movements

- Increased immigration from Latin America, Asia, and Africa
- Activism by multiracial individuals led to change
- Advocacy for more flexible and inclusive race options
- 1990 - Hispanic origin question was added with a list of options (Colombian, Dominican, etc.)
- 1997 - OMB revision for a separate 'Asian and Pacific Islander' category

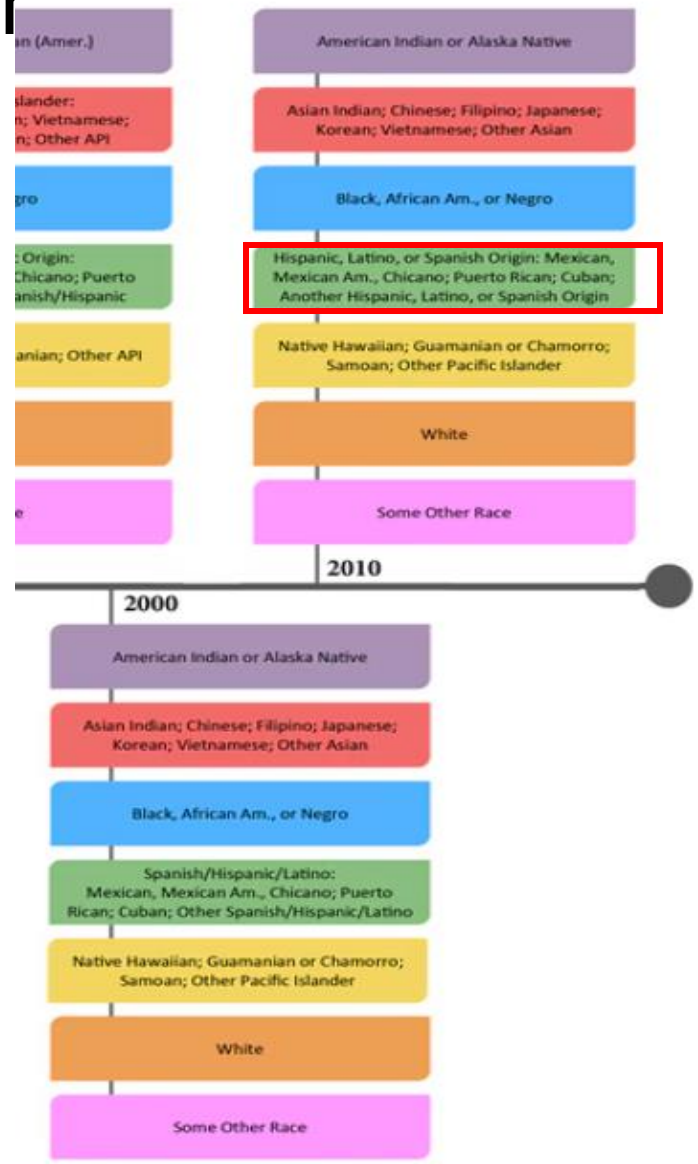
Table 16. Live Births by Mother's Descent or Origin and Borough Of Residence New York City, 1980

Mother's Ethnic Origin	Total	Manhattan	Bronx	Brooklyn	Queens	Richmond	Non-Residents	Residence Unknown
Mexican	446	115	89	185	44	2	11	–
Puerto Rican	17,838	2,634	6,959	6,442	1,443	189	171	–
Cuban	499	131	79	68	156	11	54	–
Other Spanish	18,444	4,049	2,394	7,073	4,398	142	388	–
American	16,851	2,832	1,184	4,458	6,373	858	3,146	–
European	17,429	1,535	1,637	5,143	4,524	2,367	2,223	–
Asian	8,327	1,226	348	3,398	2,528	231	596	–
African	25,490	3,670	5,504	10,581	4,832	524	379	–
Not Stated	1,742	357	487	231	276	236	151	4
Total	107,066	16,549	18,681	37,579	22,574	4,560	7,119	4

https://nyupress.org/9780814775466/changing-race/ (2000)
https://www.pewresearch.org/social-trends/2015/06/11/multiracial-in-america/ (2015)
https://www.npr.org/2020/10/04/919927641/the-u-s-census-has-changed-a-lot-since-1790-heres-how (2020)

2000–2010 – Multiracial & Detailed Respor

- 2000 Census
  - First time respondents could select multiple races
  - Two separate question for race and ethnicity
  - First use of ‘Latino’
- 2010 Instruction added that ‘Hispanic origins are not races’
- 2020: Enhanced write-in options for clearer race/ethnicity data
  - Detailed origin questions (Chinese, Jamaican, etc.)
- Testing for combined questions (race + ethnicity) and inclusion of a MENA (Middle Eastern/North African) category



https://www.census.gov/prod/2001pubs/c2kbr01-6.pdf (2001)
https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html (2021)
https://www.pewresearch.org/social-trends/2015/06/11/multiracial-in-america/ (2015)
Measuring Race and Ethnicity Across The Decades: 1790—2010 - U.S. Census Bureau

Modern Collection of Race/Ethnicity for Vital Events

- Addressing disparities through:
  - Fair and equitable public health policies
  - Creating public health interventions towards racial justice
  - Analysis of drivers of public health issues, i.e., disparities during or after COVID-19 with different rates of infection and mortality
- Root of disparities
  - Stem from historically racist policies that still affect health outcomes to this day

1. Pearson SJ. Birth Registration and the Administration of White Supremacy. Modern American History. 2022;5(2):117-141. doi:10.1017/mah.2022.13
2. The Washington Post. *Secret use of census info helped send Japanese Americans to internment camps in WWII.* The Washington Post. April 6th, 2018. https://www.washingtonpost.com/news/retropolis/wp/2018/04/03/secret-use-of-census-info-helped-send-japanese-americans-to-internment-camps-in-wwii/?noredirect=on. Accessed December 4th, 2024.

Historical Events for NYC Race & Ethnicity Data

- 1854-1952: Race data presented mostly as ‘White’ vs. ‘Non-White
  - Collected on forms as ‘Race or Color’ or ‘Color’
- Starting in 1953: Data also presented by ethnic groups
  - Show statistics for the Puerto Rican population and other groups using nativity
- ‘Race or Color’ – printed on legal NYC certificates until 1960
- Race switched to the confidential medical report

Was NYC Similar or Different on Race & Ethnicity Data Collection?

- NYC mostly collected data in parallel to Census categories
- Census inclusion of race categories align with national policies for:
  - Citizenship, Slavery , Migration
- NYC started reporting on Puerto Rican origin earlier due to migration patterns

Changes or Shifts Have Meaning

- Race and ethnicity data shape:
  - Policy, funding, and representation
- Rigid categories can affect identities and lead to some marginalized groups being undercounted
- NYC's categories were at times more detailed, reflecting local diversity better
- Recent efforts show progress in modern thinking
  - Increasing collection of and reporting on additional race and ethnicity categories modernize national standards (e.g., combined race/ethnicity question, MENA category).

https://www.ucpress.edu/book/9780520269673/the-nature-of-race (2011)
https://www.npr.org/2020/10/04/919927641/the-u-s-census-has-changed-a-lot-since-1790-heres-how (2020)
https://nyupress.org/9780814775466/changing-race/ (2000)

Race and Ethnicity Data – Why do they matter?

- Because they represent us as citizens
- Data are used for policy, to address disparities and inequities
- Used to design interventions and provide services
- Accurate race & ethnicity data -> lead to better and more equitable policies

Thank you!

msilin@health.nyc.gov

Washington State Department of HEALTH

SUPPLEMENTATION AND IMPUTATION OF DEMOGRAPHIC DATA THROUGH LINKAGES

Missing Data

DOB	Name	Sex	Race	Ethnicity	Address
NA	Robert Ross	M	White	Non-Hispanic	5495 14th Street
08/06/1945	Stacy Ruiz	F	White	Hispanic	NA
04/05/1984	Jimmy Hernandez-Soto	NA	Multiracial	NA	8795 85th Road
10/04/1985	David R	NA	NA	NA	NA

Missing Data

DOB	Name	Sex	Race	Ethnicity	Address
NA	Robert Ross	M	White	Non-Hispanic	5495 14th Street
08/06/1945	Stacy Ruiz	F	White	Hispanic	NA
04/05/1984	Jimmy Hernandez-Soto	NA	Multiracial	NA	8795 85th Road
10/04/1985	David R	NA	NA	NA	NA

Imputation and Supplementation

- Imputation
  - Filling in NAs with derived estimates or implied values

ID	x1	x2
Person 1	AIAN	Male

ID	x1	x2
Person 1	NA	NA

- Supplementation
  - Providing additional confirmation or alternatives to filled or missing data values

ID	x1	x2
Person 1	AIAN	Male

ID	x1	x2
Person 2	White	Other

Imputation and Supplementation

Familial/relational networks

DOB	Name	Sex	Race	Ethnicity	Relationship
04/05/1984	Jimmy Hernandez-Soto	NA	Multiracial	Hispanic	Target
04/05/1984	Martina Johnson	F	Multiracial	Hispanic	Fraternal Twin of Target
08/19/1956	Martina Soto	F	Black	Hispanic	Mother of Target

Predictive models

DOB	Name	Sex	Race	Ethnicity	Address
04/05/1984	Jimmy Hernandez-Soto	M	Multiracial	Hispanic	8795 85th Road

Linkage to other datasets

Linkage

ID	x1	x2
Person 1	AIAN	Male

ID	x1	x2
Person 1	NA	Male

ID	x1	x2
Person 1	AIAN	Other

ID	x1	x2
Person 1	White	Male

ID	x1	x2
Person 1	AIAN(2); White	Male; Other

- Supplementation can provide weighted values to inform true value or reveal a more multifaceted truth
- Data source nuances need to be taken into account
  - Data quality
  - Limitations of the data system

Pros

- Clearer demographics allow for more robust disaggregation and representation
- Helps avoid overwhelming “Missing” or “Other” buckets of data
  - These buckets are not equal in subgroup representation
- Assists in more targeted surveillance, analysis, interventions, and public health action

Cons

- Imputation assumes the risk of error inherently
- Supplementation can “muddy” the truth with inferior data inclusion
- Linkage to derive demographics and/or geographical variables obviously involves… linkage
    - Linkage as a practice can do more harm than good…
        - If done suboptimally
        - If done without robust procedures to ensure privacy and security
        - If expanded beyond the targeted scope (intentionally or not)

If Done Suboptimally…

In many cases, linkages need to go beyond out-of-the-box probabilistic solutions

- We use robust machine learning strategies to find more links, with less errors, while prioritizing equity in representation
- If equity is not prioritized, the overrepresentation of underrepresented subgroups is compounded in Missing/Other disaggregations

If Done Without Robust Procedures to Ensure Privacy and Security…

- If privacy and security are not prioritized, deliberate/intentional omissions in data (resulting in NAs) will be violated. All linkages should be targeted, narrow as possible, and agreed on by all parties
  - This is difficult in low-trust situations

If Done Without Robust Procedures to Ensure Privacy and Security…

Low-trust example:

- Have to restrict data to those in tribal roster (doesn't want to share ALL data)
- Unable to share tribal affiliation for members (doesn't want state government to have roster data)

Planned a shared encoded and blind model with encryption, but problems remain:

1. Someone has to hold the key, a trusted third party, or both entities
2. At point of delivery, roster is known without deliberate obscurity and error inclusion
3. A collaborative trust-building process prioritizing privacy and security is the only option

If Expanded Beyond Targeted Scope

- Inappropriate disclosure of unnecessary information can arise through linkage if not careful, and some federal/state statutes may be violated
  - Unintentionally and indirectly disclosing variables and patterns of variables
    - Surrogacy, adoptions, name changes, sex changes
    - SUDs, abortion related data, gender affirming care data, SOGI

Takeaways

- Linkage can be leveraged to supplement and impute demographic variables into analytical datasets. However, this does not come without caveats and warrants consideration.
  - Use a robust linkage that is as accurate and representative as possible
  - Keep the scope narrow, abide by data minimization standards, and prioritize privacy and security
  - Know the nuances of data sources and communicate with partners
  - Realize that imputation and supplementation can introduce novel errors, so use them appropriately