Analyzing the Demographic and Socioeconomic Makeup of Biomedical Research Workers in the US

A NIGMS - GW Collaboration

Contents

Team Introduction

- Abdul Haleem Abdul Salam
- Sai Nityamani Sahith Matsa
- Yiying Niu (Judy)

Client Introduction

National Institute of General Medical Sciences (NIGMS)

- Vital component of the National Institutes of Health (NIH).
- Role:
  - Advancing biomedical research and scientific discovery.
- Mission:
  - Foster a vibrant and diverse biomedical research workforce.
  - Promote scientific discovery, advancement, and achievement.

- Division: Division of Data Integration, Modeling, and Analytics (DIMA).
  - Responsible for data-driven discussions, decisions, and actions.
  - Tasked with conducting strategic analyses and evaluations.

Problem Understanding

Estimating descriptive demographic and socioeconomic statistics of the US biomedical research workforce.

Better understanding the makeup of workers conducting biomedical research by industry, state and other necessary varaibles.

Descriptive Analytics and Regression Analysis.

Project Rationale: Addressing Data Gaps

- Lack of Detailed Information:
  NIGMS lacks comprehensive data on the demographic and socioeconomic composition of the U.S. biomedical research workforce.

- Incomplete Understanding:
  Current data gaps hinder NIGMS' ability to evaluate program effectiveness and plan for future initiatives.

- Strategic Necessity:
  Detailed insights into workforce demographics, industry distribution, and geographic representation are crucial for informed decision-making and resource allocation.

Methodology

01 Data Selection and Understanding.

02 Data Wrangling, Descriptive and Regression Analysis.

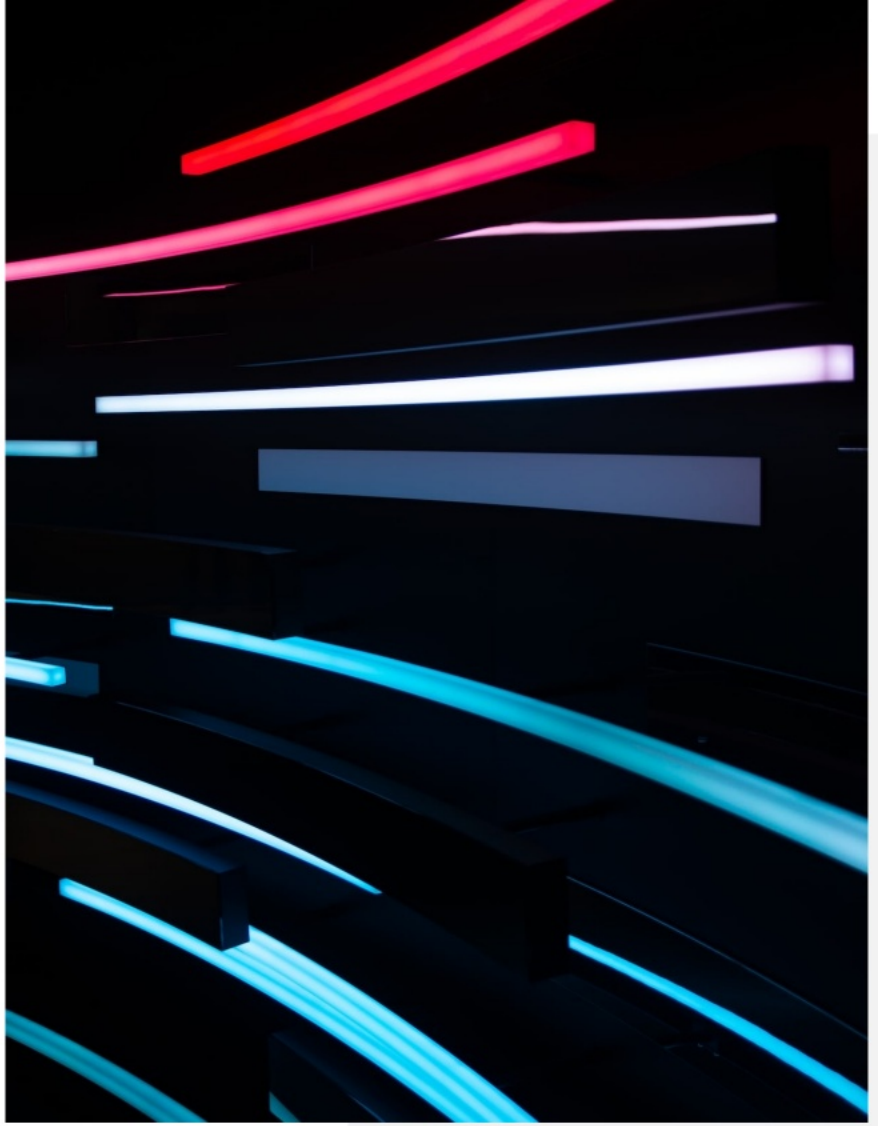
03 Visualization & Results (Graphs, table shells and Tableau Dashboard).

Data Overview

- The dataset is sourced from the American Community Survey (ACS) Public Use Microdata Sample (PUMS) for the year 2022.
- The 2022 5-year data files were obtained, consisting of five distinct files: psam_pusa, psam_pusb, psam_pusc, psam_pusd, and psam_puse.
- These files were meticulously appended to compile a comprehensive dataset for subsequent analysis.
- Five key occupations within the biomedical research domain are identified and isolated from the dataset:
  - 1610: Biological Scientists (All Others)
  - 1650: Medical Scientists (Except Epidemiologists)
  - 1720: Chemists
  - 1910: Biological Technicians
  - 1920: Chemical Technicians

Data Wrangling

- Added 'Year' column derived from serial number.
- Meticulously prepared dataset post-occupation delineation.
- Facilitated temporal analysis and trend identification.
- Incorporated person weights to rectify sampling biases.
- Each observation weighted based on demographic characteristics.

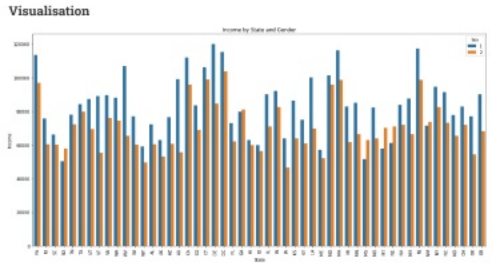


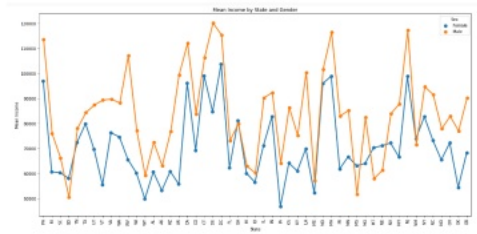
Data Wrangling

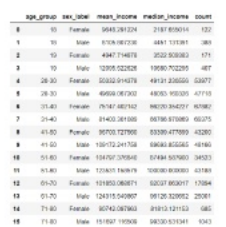
- Adjusted wage data to mitigate inflationary impacts.
- Created 'income adjusted' column for accurate income portrayal.
- Meticulously selected variables essential for analysis.
- Discarded redundant or extraneous variables post-variable refinement.

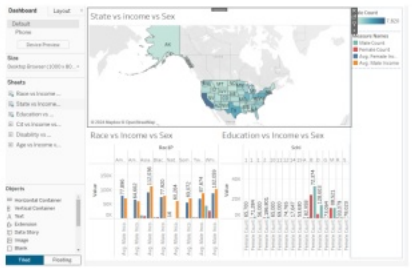
Descriptive Analysis

- Stratified analysis by gender, income level, and educational status.
- Examined each state's workforce composition within occupations.
- Analyzed gender disparities in income across states.
- Explored educational attainment's impact on income.
- Provided granularity by dissecting workforce attributes.

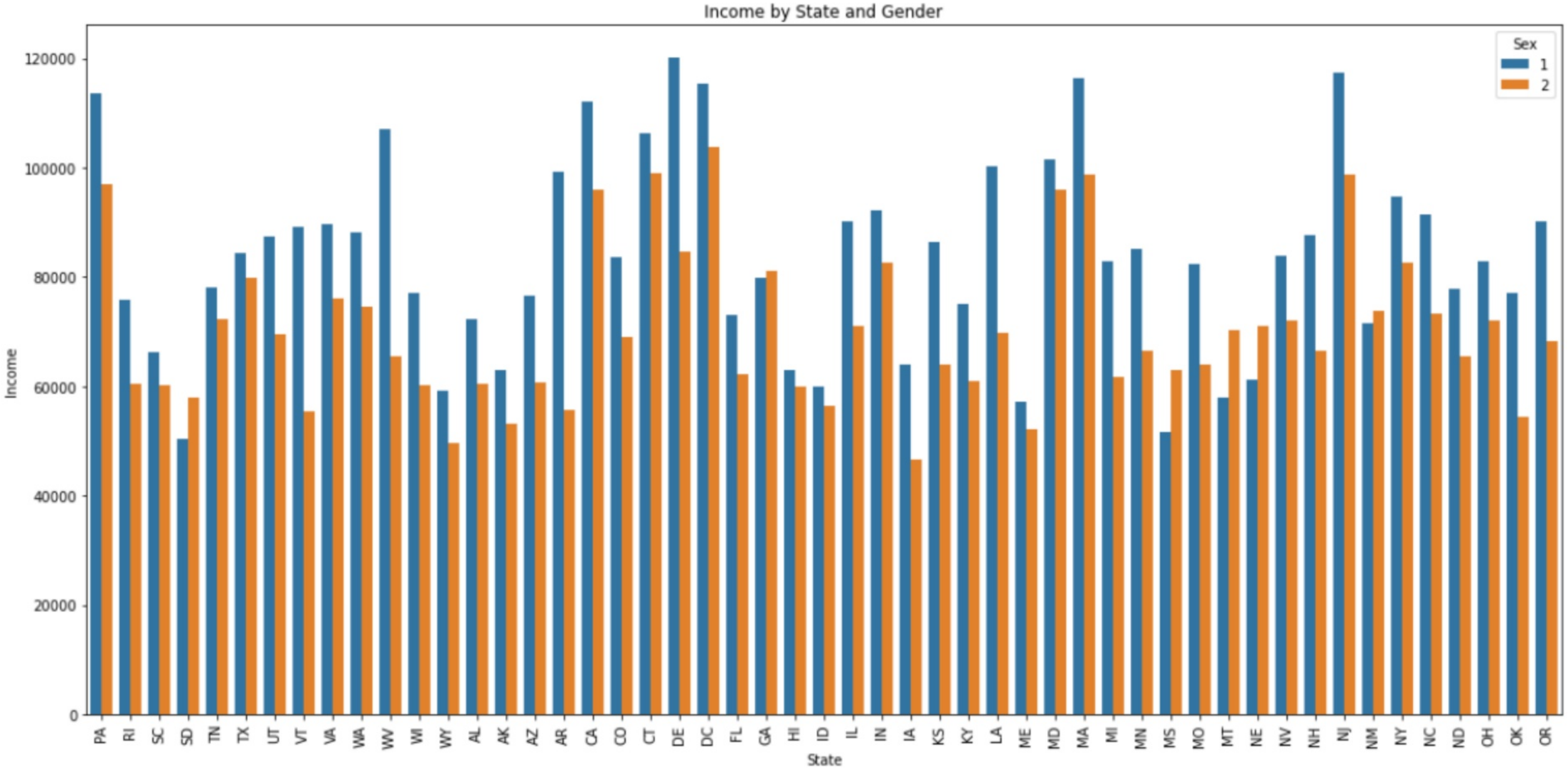


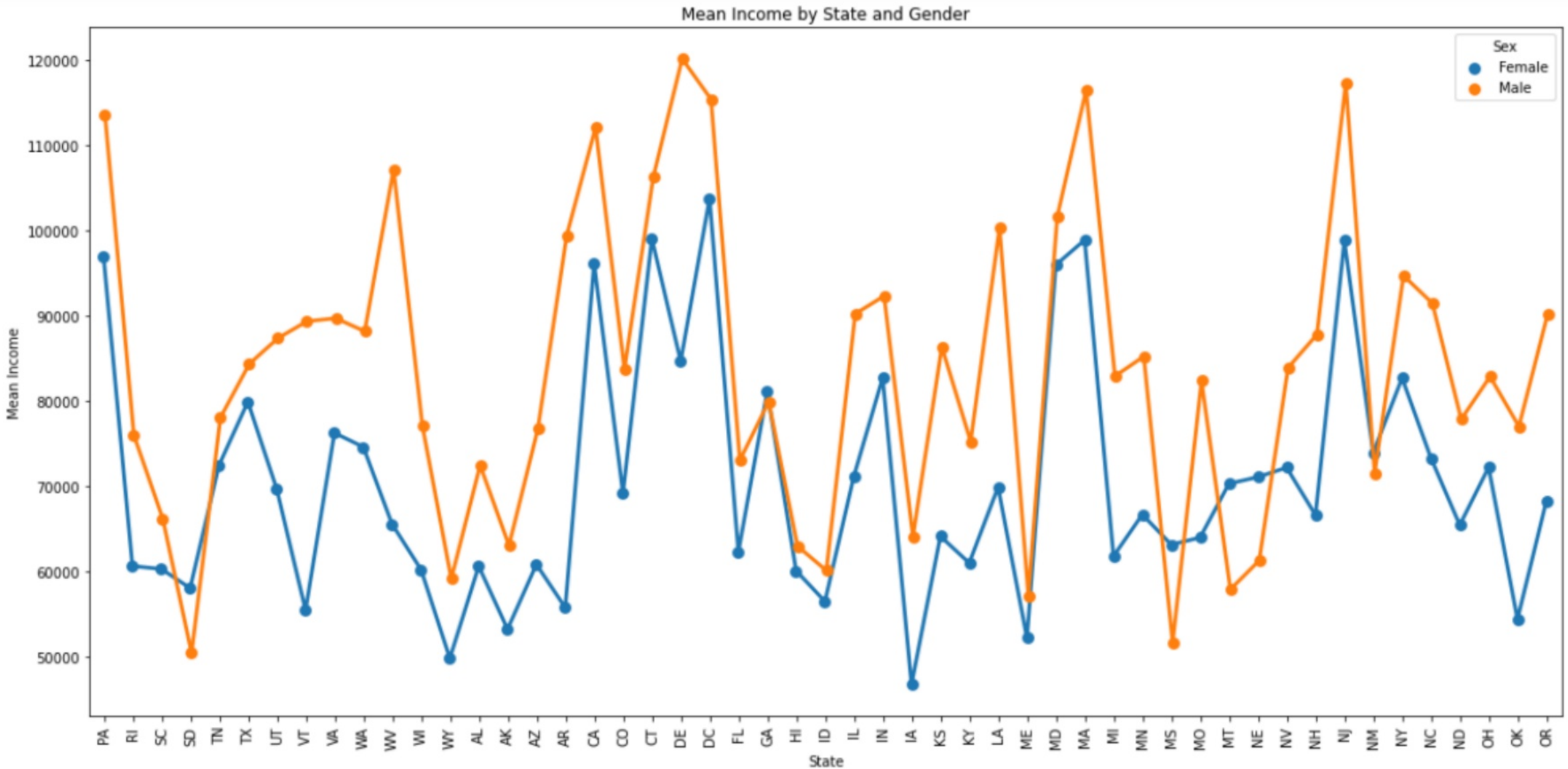






Visualisation




Mean Income by State and Gender
Sex
Female
Male
Mean Income
120000
110000
100000
90000
80000
70000
60000
50000
State
PA RI SC SD TN TX UT VT VA WA WV WI WY AL AK AZ AR CA CO CT DE DC FL GA HI ID IL IN IA KS KY LA ME MD MA MI MN MS MO MT NE NV NH NJ NM NY NC ND OH OK OR

	age_group	sex_label	mean_income	median_income	count
0	18	Female	9648.291224	2187.655014	122
1	18	Male	8105.807230	4451.131391	388
2	19	Female	4947.714878	3522.509383	171
3	19	Male	12095.522626	10680.702295	407
4	20-30	Female	50332.914378	49131.230556	53977
5	20-30	Male	49699.067202	48063.160326	47718
6	31-40	Female	75147.402142	66220.354227	67882
7	31-40	Male	81400.301089	66766.970869	69375
8	41-50	Female	96700.727560	83309.477899	43200
9	41-50	Male	109172.241758	89693.855565	48166
10	51-60	Female	104797.376640	87494.587900	34523
11	51-60	Male	123531.158579	100000.000000	43188
12	61-70	Female	101850.068671	82037.063017	17894
13	61-70	Male	124315.549867	96126.320652	25001
14	71-80	Female	90742.097963	81813.121153	685
15	71-80	Male	151697.116509	99330.531341	1043

	sex_label	schl	occp	mean_income
0	Female	21	1610.0	60530.842884
1	Female	22	1610.0	75929.500406
2	Female	23	1610.0	90882.197235
3	Female	24	1610.0	103820.189948
4	Male	21	1610.0	71331.825438
5	Male	22	1610.0	80207.035049
6	Male	23	1610.0	101043.334115
7	Male	24	1610.0	116800.478797

Dashboard | Layout

Default
Phone

Device Preview

Size

Desktop Browser (1000 x 80...

Sheets

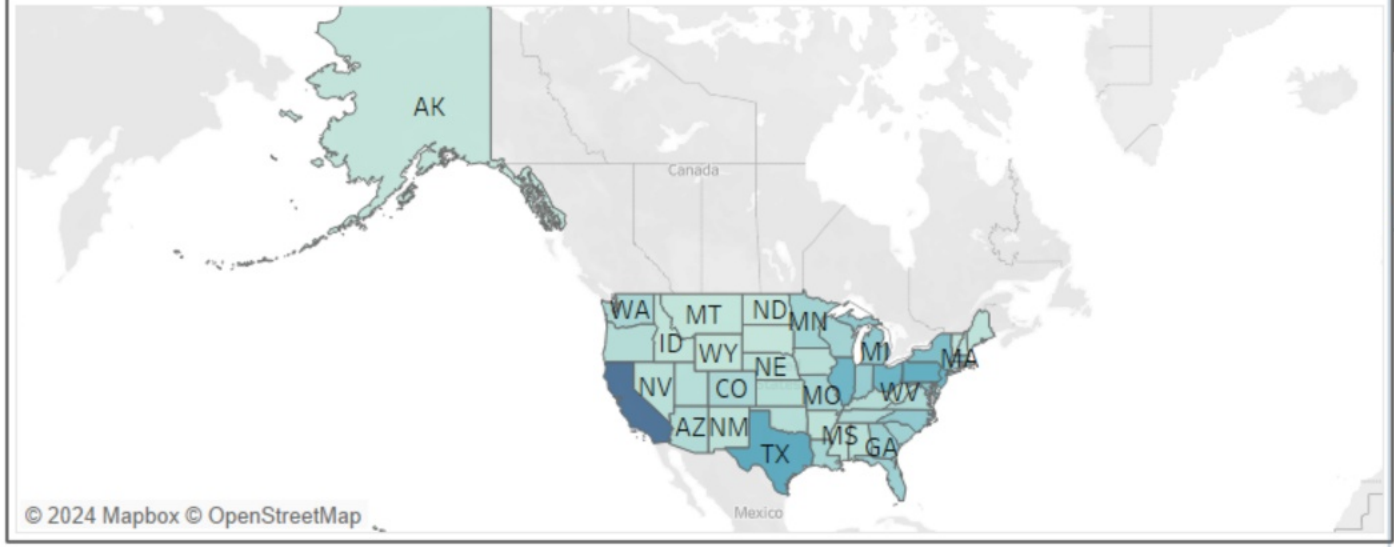
- Race vs Income ...
- State vs Income...
- Education vs ...
- Cit vs Income vs...
- Disability vs ...
- Age vs Income v...

Objects

- Horizontal Container
- Vertical Container
- Text
- Extension
- Data Story
- Image
- Blank

Tiled | Floating

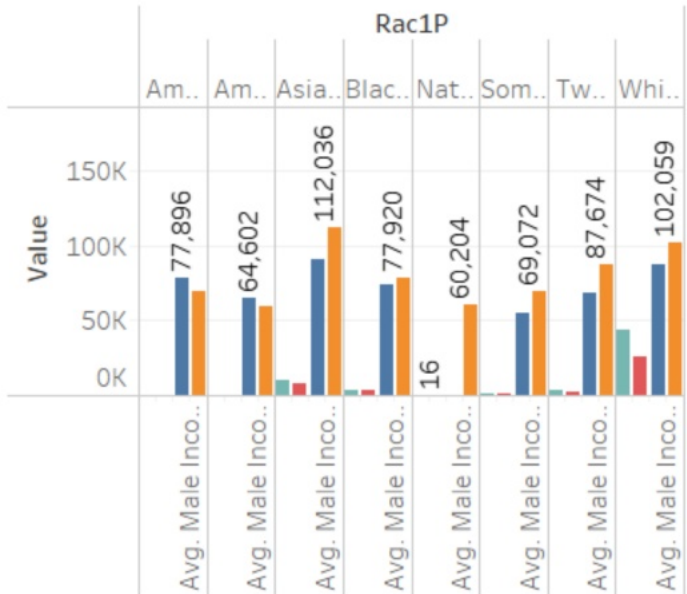
State vs income vs Sex



ale Count
7,820

Measure Names
- Male Count
- Female Count
- Avg. Female Inc..
- Avg. Male Income

Race vs Income vs Sex



Education vs Income vs Sex

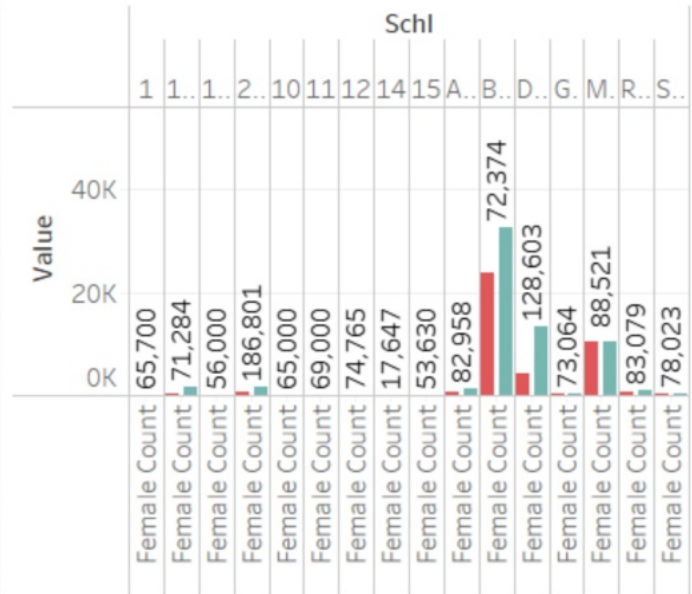


Table Shell

Race	Female_Income	Male_Income	Male_Income_Biolo	Female_Income_Biolo	Male_Income_Medic	Female_Income_Medi	Male_Income_Other	Female_Income_Other	Male_Income_Cher	Female_Income_Cher	Male_Income_Biological	Female_Income_Biological_Technicians
White alone	78455	93790	82354	68558	126435	91374	102059	87174	65311	58036	56356	50233
Black or African American alone	73192	73617	70533	76566	85616	85402	77920	73499	62801	52014	70455	65839
American Indian alone	65452	57515	48024	55022	83390	94254	70021	77896	66377	16254	39699	27363
Alaska Native alone	115819	23770	0	97194	0	119311	0	0	23770	0	0	0
American Indian and Alaska Native tribes	57868	44384	87338	50894	0	0	59139	64602	36517	15487	7954	38284
Asian alone	93488	105300	110059	93987	109034	98670	112036	90269	58232	62152	78269	69236
Native Hawaiian and Other Pacific Islander alc	82922	65497	0	0	62354	90421	60204	0	48861	0	92613	67682
Some Other Race alone	56896	66650	52712	61278	99342	63464	69072	55095	60674	44314	49569	57216
Two or More Races	75912	80758	75097	79817	100879	88155	87674	68357	53619	55657	51554	45389

	Education	Male_income	Female_Income	Male_Income_Biolo	Female_Income_Biolo	Male_Income_Medic	Female_Income_Medi	Male_Income_Other	Female_Income_Other	Male_Income_Cher	Female_Income_Cher	Male_Income_Biological	Female_Income_Biological_Technicians
1	No schooling completed	59212	63737	0	0	54893	67596	77221	65700	50154	48869	65989	57502
2	Nursery school, preschool	0	94587	0	0	0	0	0	0	0	0	0	94587
3	Kindergarten	33383	0	0	0	0	0	0	0	33383	0	0	0
4	Grade 1	0	0	0	0	0	0	0	0	0	0	0	0
5	Grade 2	0	0	0	0	0	0	0	0	0	0	0	0
6	Grade 3	39519	18000	0	0	0	0	0	0	39519	18000	0	0
7	Grade 4	0	0	0	0	0	0	0	0	0	0	0	0
8	Grade 5	54691	0	0	0	0	0	0	0	54691	0	0	0
9	Grade 6	48901	37978	0	0	56815	55540	0	0	46487	69951	0	8181
10	Grade 7	37579	40661	0	0	0	48072	0	65000	37579	27800	0	0
11	Grade 8	88332	52394	0	0	338578	54691	67817	69000	80412	39167	44511	62269
12	Grade 9	53892	31962	0	0	225328	71099	63294	74765	41462	28345	60140	19774
13	Grade 10	58495	52850	0	0	0	0	70520	56000	54740	43981	63780	79742
14	Grade 11	81281	44649	0	0	0	68266	77907	17647	82003	27010	81018	61619
15	12th grade – no diploma	58092	57815	0	0	181589	78113	66095	53630	53025	48602	47710	50551
16	Regular high school diploma	57421	52860	0	0	48961	90859	67293	83079	59149	48807	42145	42885
17	GED or alternative credential	56952	48876	0	0	69069	72721	78051	73064	54829	49424	42646	40025
18	Some college, but less than 1 year	67021	51534	0	0	148214	69662	82087	78023	64399	48737	61080	42524
19	1 or more years of college credit, no degree	60775	47713	0	0	115269	87292	84266	71284	60069	55631	42350	29286
20	Associate's degree	68249	59935	0	0	77219	69092	88085	82958	66979	59578	57288	49738
21	Bachelor's degree	78890	70469	71332	60531	99327	82551	83068	72374	66345	56392	56553	61335
22	Master's degree	95095	81467	80207	75930	104505	81907	105439	88521	70359	77955	69675	82065
23	Professional degree beyond a bachelor's degr	143638	117535	101043	90882	153185	116580	122196	186801	62626	60374	257910	120611
24	Doctorate degree	128369	109175	116800	103820	125106	106384	144299	128603	104849	108820	203925	222639

Analyzing the Demographic and Socioeconomic Makeup of Biomedical Research Workers in the US

A NIGMS - GW Collaboration

Regression Analysis

- Utilized regression models to identify statistically significant trends and changes in descriptive statistics over time.
- Investigated relationships between demographic and socioeconomic variables to uncover underlying patterns.
- Identified key predictors influencing income disparities and workforce dynamics.

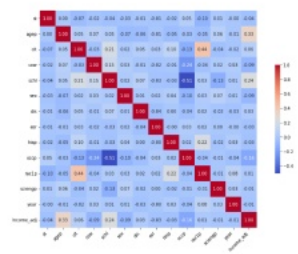
Visualisation

OLS Regression Results

Dep. Variable:	Income_adj	**R-squared:**	0.177
Model:	OLS	**Adj. R-squared:**	0.177
Method:	Least Squares	**F-statistic:**	4986.
Date:	Mon, 29 Apr 2024	**Prob (F-statistic):**	0.00
Time:	15:32:52	**Log-Likelihood:**	-1.1377e+06
No. Observations:	92696	**AIC:**	2.275e+06
Df Residuals:	92693	**BIC:**	2.275e+06
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.136e+04	592.826	19.157	0.000	1.02e+04	1.25e+04
C(schl)[T.22]	6490.9954	400.106	16.223	0.000	5706.791	7275.200
C(schl)[T.23]	2.222e+04	1045.067	21.264	0.000	2.02e+04	2.43e+04
C(schl)[T.24]	3.745e+04	473.343	79.119	0.000	3.65e+04	3.84e+04
agep	1425.5997	14.140	100.821	0.000	1397.885	1453.314

Omnibus:	76790.807	**Durbin-Watson:**	0.104
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4275949.119
Skew:	3.809	**Prob(JB):**	0.00
Kurtosis:	35.389	**Cond. No.**	262.



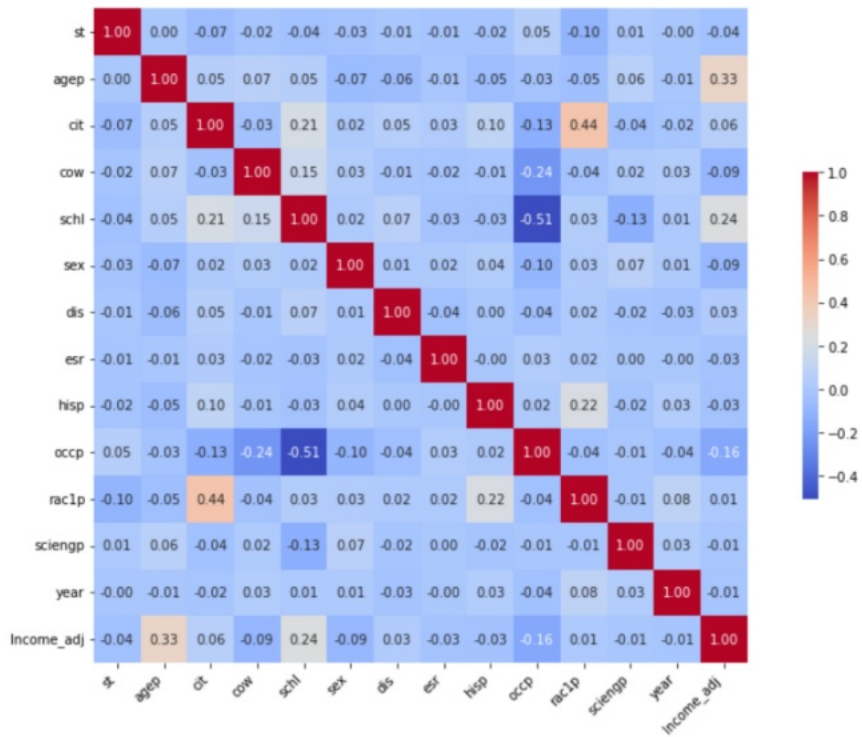
Visualisation

OLS Regression Results

Dep. Variable:	Income_adj	**R-squared:**	0.177
Model:	OLS	**Adj. R-squared:**	0.177
Method:	Least Squares	**F-statistic:**	4986.
Date:	Mon, 29 Apr 2024	**Prob (F-statistic):**	0.00
Time:	15:32:52	**Log-Likelihood:**	-1.1377e+06
No. Observations:	92698	**AIC:**	2.275e+06
Df Residuals:	92693	**BIC:**	2.275e+06
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.136e+04	592.826	19.157	0.000	1.02e+04	1.25e+04
C(schl)[T.22]	6490.9954	400.106	16.223	0.000	5706.791	7275.200
C(schl)[T.23]	2.222e+04	1045.067	21.264	0.000	2.02e+04	2.43e+04
C(schl)[T.24]	3.745e+04	473.343	79.119	0.000	3.65e+04	3.84e+04
agep	1425.5997	14.140	100.821	0.000	1397.885	1453.314

Omnibus:	78790.807	**Durbin-Watson:**	0.104
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4275949.119
Skew:	3.809	**Prob(JB):**	0.00
Kurtosis:	35.389	**Cond. No.**	262.



Observations - Gender and Income Disparities by State

- Females exhibit higher mean income than males in Georgia, Mississippi, Maryland, Nebraska, New Mexico, and South Dakota.
- Female technicians generally earn more than males across most states, except in biochemical sciences where males outperform females.

Occupational Analysis

- Occupation 1720 (Chemists) presents an anomaly with zero male Washington D.C., resulting in higher female incomes in states like Kentucky, and Nebraska.
- Females in Florida receive comparable incomes to males in occupation 1910 (Biological Technicians), despite lower overall counts.
- In occupation 1920 (Chemical Technicians), despite male outnumbering, females earn higher incomes in most states due to their experience in the field.
- In occupation 1610 (Biological Scientists), males tend to dominate with experience, leading to higher incomes, particularly after age 41.

Race and Employment Dynamics

- Among Alaska Native individuals, females are more prevalent in employment and often hold scientist roles.

- Asians exhibit higher incomes compared to other racial groups, followed by White individuals, while American Indians and Native Alaska tribes have relatively lower performance.

Educational Attainment and Income Disparities

- Majority of individuals, both males and females, hold bachelor's degrees or higher across all occupations, indicating a high educational attainment level.
- Exceptions exist in occupations other than 1610 (Biological Scientist), where lower educational attainment correlates with lower income, albeit with smaller sample sizes.
- Individuals with professional degrees earn higher incomes than those with other degrees, including doctorates, attributed to their experience in the field.
- Higher age correlates with greater experience, leading to higher incomes, especially among individuals with professional degrees.

Source: BillionPhotos/stock.adobe.com

Links

Type of ML Use		Sample Mean[1]	Effect Estimate[2]	*p*-value
Resource Use				
	Recommended	XXX	XXX (XXX to XXX)	0.XXX
	None	XXX		
	Recommended	XXX	XXX (XXX to XXX)	0.XXX
	Under-use	XXX		
	Recommended	XXX	XXX (XXX to XXX)	0.XXX
	Over-use	XXX	XXX (XXX to XXX)	
	Any	XXX	XXX (XXX to XXX)	0.XXX
	None	XXX		
Time-to-completion				
	Recommended	XXX	XXX (XXX to XXX)	0.XXX
	None	XXX		
	Recommended	XXX	XXX (XXX to XXX)	0.XXX
	Under-use	XXX		
	Recommended	XXX	XXX (XXX to XXX)	0.XXX
	Over-use	XXX	XXX (XXX to XXX)	
	Any	XXX	XXX (XXX to XXX)	0.XXX
	None	XXX		

ta are mean number of person-hours or weeks to completion. Sample mean resource use may be underestimated due to right-
nsoring of ongoing projects.

- Tableau Dashboard.
- Table Shell.

Data Verfication

- Employed Census Bureau data to verify the accuracy of our estimates.
- Conducted rigorous comparisons between our analysis results and data provided by the Census Bureau.
- Observed that discrepancies between our estimates and Census Bureau data were within a margin of error of 5%.
- Demonstrated the reliability and validity of our analysis methodology through robust data verification processes.

United States® Census Bureau

Conclusion

01 Our analysis reveals significant gender disparities in income distribution and anomalies within occupational demographics.

02 Race-based employment dynamics underscore the need for diversity and inclusion initiatives within the biomedical research workforce.

03 Educational attainment correlations highlight the importance of fostering equity and diversity to promote innovation and inclusivity in the field.

We extend our gratitude to NIGMS for the opportunity to contribute to this impactful research project.

We appreciate the guidance and support provided by our instructors, mentors, and colleagues throughout this endeavor.

Analyzing the Demographic and Socioeconomic Makeup of Biomedical Research Workers in the US

A NIGMS - GW Collaboration

Analyzing the Demographic and Socioeconomic Makeup of Biomedical Research Workers in the US

A NIGMS - GW Collaboration

Analyzing the Demographic and Socioeconomic Makeup of Biomedical Research Workers in the US

A NIGMS - GW Collaboration