Project 3: UMD

Yangjianchen Xu

Overview

- Background
- Purpose
- Related questions
- Data processing
- Answers for questions
- Conclusions
- Future work

Background

- The programs of **Urban Ministries of Durham (UMD)** end homelessness by providing neighbors with emergency shelter and case management to help them overcome barriers such as unemployment, medical and mental health problems, past criminal convictions and addiction.
- The data from UMD contains the demographic (gender, age, race and ethnicity), income, disability and other information of the clients.

Purpose

- Overall purpose: We aim to provide better service for the clients by analyzing these data.

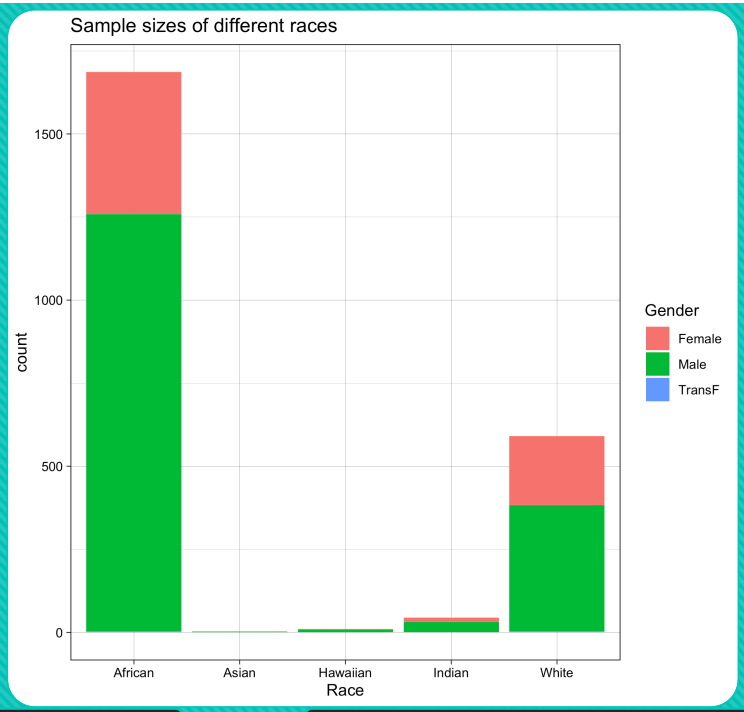
Related questions

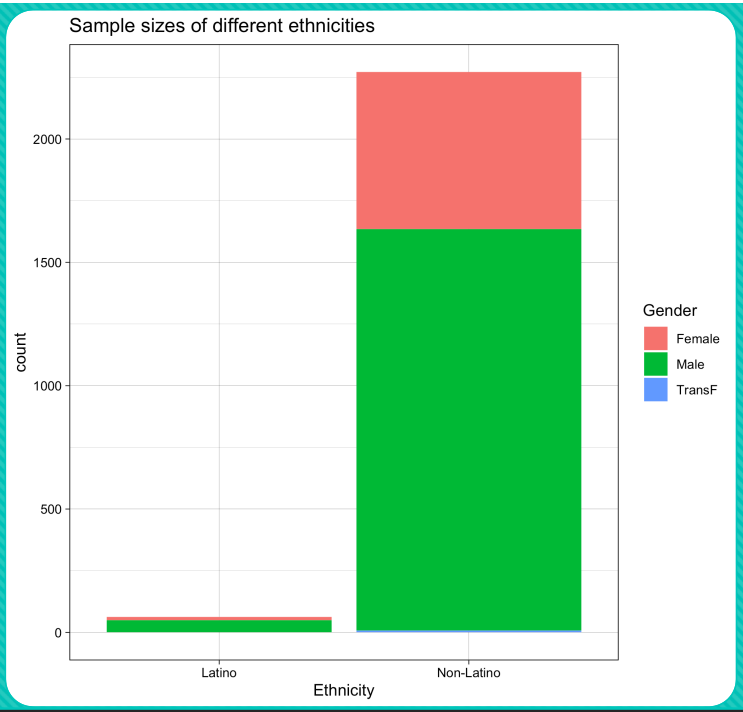
- What are the demographic characteristics of the clients?
- How do the demographics influence the income of the clients?
- How do the demographics influence the disability of the clients?

Data processing

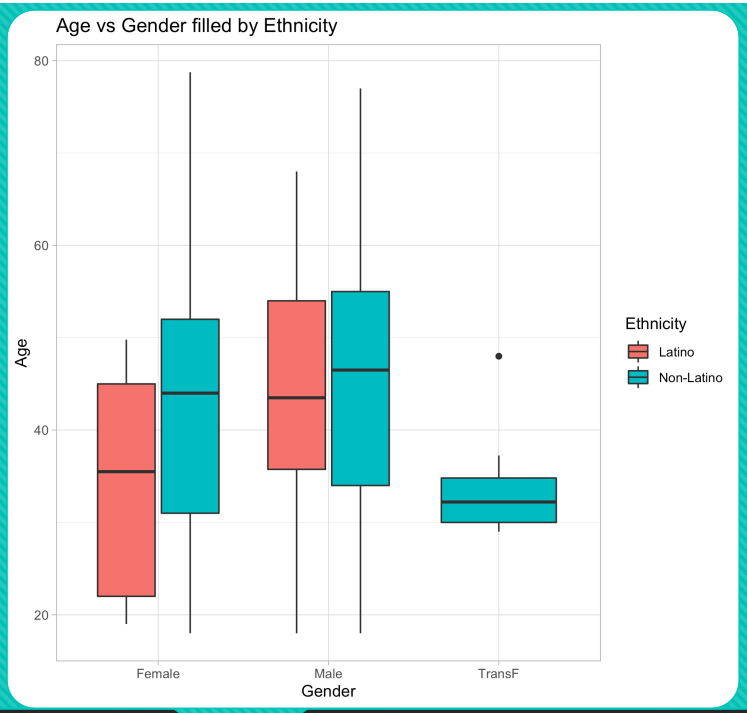
- Select variables: ID, Age, Gender, Race, Ethnicity, Income and whether disabled or not.
- Discard observations with NA in these variables.
- Group the data by Client ID.
- Merge different datasets by Client ID.

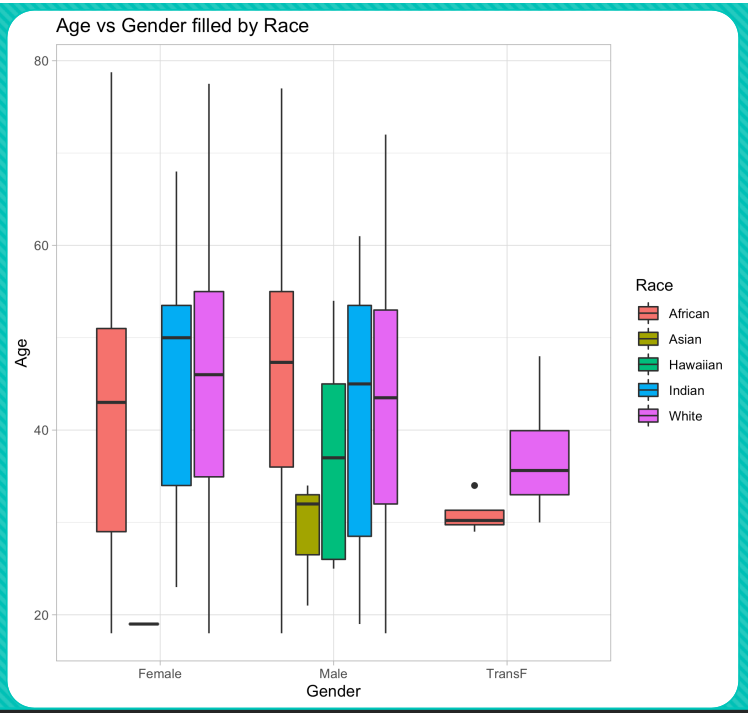
	ID	Gender	Race	Ethnicity	Age	Income	Disability
0	1096	Male	Black or African American (HUD)	Non-Hispanic/Non-Latino (HUD)	60.666667	1015.000000	No (HUD)
1	1097	Male	Black or African American (HUD)	Non-Hispanic/Non-Latino (HUD)	58.000000	748.000000	No (HUD)
2	2054	Female	White (HUD)	Non-Hispanic/Non-Latino (HUD)	63.318182	524.666667	No (HUD)
3	2142	Male	Black or African American (HUD)	Non-Hispanic/Non-Latino (HUD)	62.333333	998.000000	No (HUD)
4	2416	Male	Black or African American (HUD)	Non-Hispanic/Non-Latino (HUD)	48.333333	1200.000000	No (HUD)





Demographic characteristics





Demographic characteristics

Logistic regression

- How do the demographics influence the disability of the clients?
- Since disability of clients is a binary variable, we can use logistic regression to fit the data.
- Logistic regression assumes:
- [Distributional assumption]: $y_i|x_i \sim B\big(n_i, \pi(x_i)\big)\ for\ i = 1,2,\dots,N$
- [Structural assumption]: $\pi(x_i)$ is related to x_i by $g_i\big(\pi(x_i)\big) = \log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = x_i^T\beta$

Logistic regression results

Variable	Estimate	StdErr	Z-value	P-value
Intercept	-4.012	1.471	0.008	0.006384 **
Gender_Male	-0.770	0.216	-3.566	0.000363 ***
Gender_TransF	-13.729	901.387	-0.015	0.987848
Race_Asian	3.327	1.600	2.080	0.037524 *
Race_African	0.708	1.022	0.693	0.488258
Race_Hawaiian	-12.586	753.695	-0.017	0.986676
Race_White	0.067	1.047	0.064	0.948648
Ethnicity_Non-Latino	0.571	1.024	0.558	0.576533
Age	0.007	0.008	0.812	0.416732

Inverse Gaussian regression

- How does the demographics influence the income of the clients?
- Since the income of clients is a continuous variable, we can use logistic regression to fit the data.
- Inverse Gaussian regression assumes:
- [Distributional assumption]: $y_i \sim IG(\mu_i, \lambda)\ for\ i = 1,2, \dots, N$
- [Structural assumption]: μ_i is related to x_i by $g(\mu_i) = x_i^T \beta$

Inverse Gaussian regression results

Variable	Estimate	StdErr	t-value	P-value
Intercept	8.890e-07	3.254e-07	2.732	0.00638 **
Gender_Male	-2.653e-07	9.196e-08	-2.885	0.00398 **
Gender_TransF	1.990e-06	1.296e-06	1.536	0.12493
Race_African	3.451e-07	2.366e-07	1.459	0.14491
Race_Hawaiian	1.644e-06	1.262e-06	1.303	0.19272
Race_White	3.809e-07	2.461e-07	1.548	0.12199
Ethnicity_Non-Latino	2.900e-07	2.044e-07	1.419	0.15619
Age	-3.655e-09	2.897e-09	-1.262	0.20738

Conclusion

- Most of the clients are African American and white.
- Most of the clients are non-Latino.
- Asian and American Indian clients are mostly male.
- Non-Latino women are older than Latinos.
- Male are less likely to be disabled than female.
- Asian are more likely to be disabled than American Indian.
- **UMD can use these results to adjust their policy.**

Future work

- Find a better model to fit the income data.