
DIABETES
130

The Diabetes 130 Dataset

The data set is the result of deciphering the early readmission of the patient within 30 days of discharge. It reports across 10 years of clinical care at 130 different US hospitals.

Our data is filtered into categories describing what self-identifying characteristics each patient possess. This includes race, gender, and age. The dataset is arranged with 101767 rows and 50 columns, each representing a singular patient.

Problem Statement

The Diabetes 130 Dataset contains sensitive information, such as the age, gender, and the race of patients.

How does this infringe upon the autonomy of diabetic patients in their self-identification and selfhood?

We found that race and gender fields are lacking variety in what ways patients identify themselves. The age field also goes in increments of 10, rather than going numerically in order. Limiting these options could cause the data to be skewed and possibly excluding or misrepresenting different groups of people.

Data Analyzed

Feature Name	Type	Description and Values	% Missing
Race	Nominal	Values: Caucasian, Asian, African American, Hispanic, and other	2%
Gender	Nominal	Values: Male, Female, and Unknown/Invalid	0%
Age	Nominal	Grouped in 10-year intervals [0, 10), [10, 20), …, [90, 100)	0%

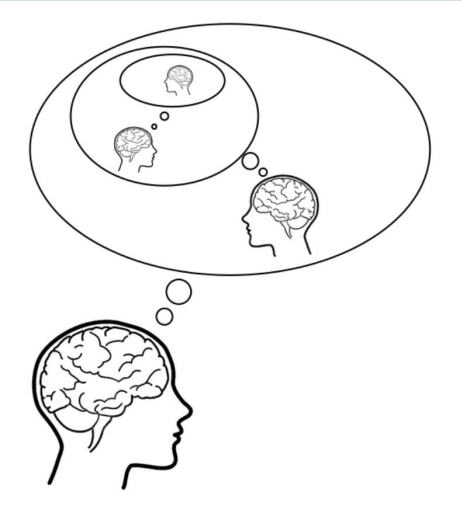
Connotative and Deconstructive Research Methods

“Reading Datasets” – Lindsay Poirier

Connotative Research:

- How has meaning changed over time, according to what social interests, and within what contexts?
  - Requires more creative searching of connected materials

Using this element will help us to inquire into the socio-political context behind the confining data labels, which may not account for every individual accurately.



Deconstructive Research:

- Who or what is rendered absent or externalized due to the particular focus of the dataset definitions
  - Involves analyzing definitions, thinking deeply about what is inside the dataset, and how what is inside neglects what is left outside

Format Anatomies: Micro Method

“Studying Data Conditions: Anatomies of Formats” – Colin Koopman

Research at the <u>micro</u> level:

- Data inputs and their configurations
  - Look at data dictionaries as well as specific data points for permissible variables, data types, and present and missing data

Using this element will aid in making visible and intelligible that which is obscure. We will dive into the information already present, specifically those which are descriptors of individuals, as opposed to looking at the context behind this data.

Race Allowed Values

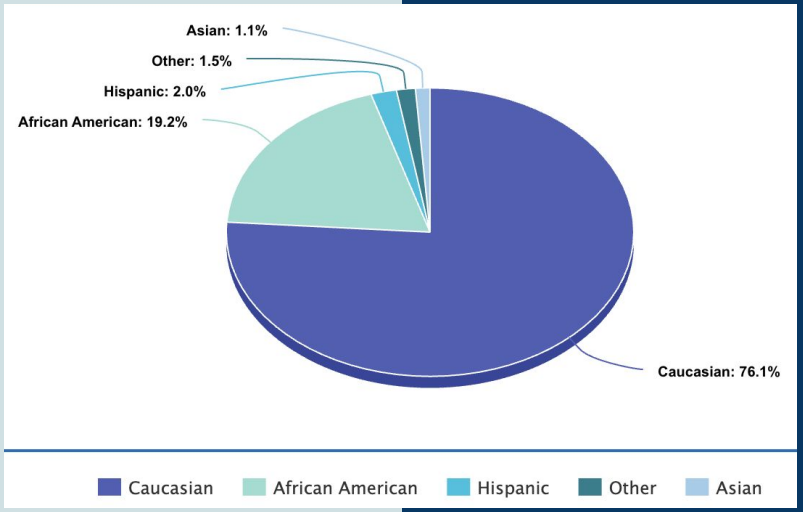
Caucasian

Asian

African American

Hispanic

Other



- No option for mixed races
- Only 5 set options to choose from and one of them is “other”
    - This could lead to mixed people only choosing one race because the other is not included

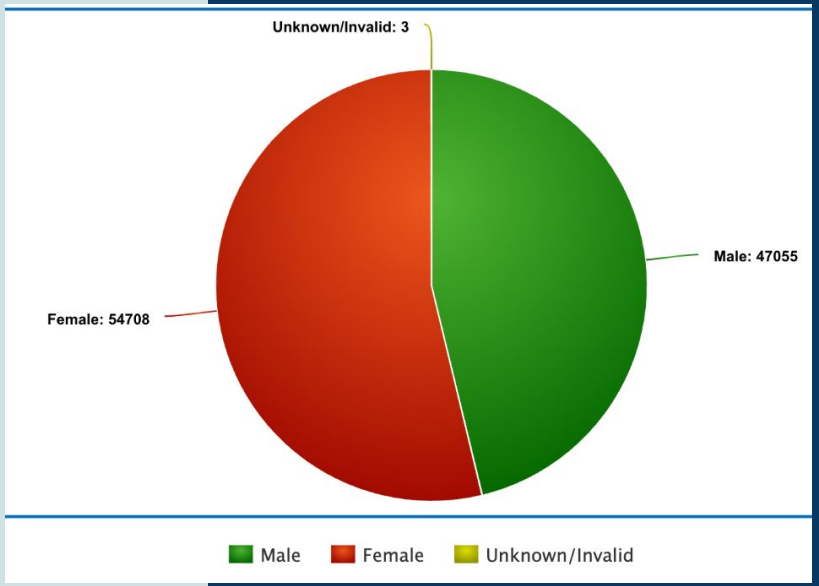
Gender Allowed Values

Male

Female

Unknown/Invalid

- Not labeled “sex,” yet the only options are either male, female or “unknown/invalid”



Age Allowed Values

0-10

10-20

20-30

30-40

40-50

50-60

60-70

70-80

80-90

90-100

- Not numeric, is very limiting because it goes in increments of 10
- Big difference between 1 yr old and 9 yr old
- Not specific enough