# Diabetes U.S. Hospitals from 1999-2008

By: Evelyn Borst, Emma Nussbaum, Karson Brown, & Sophia Nguyen

Dataset Description

The age range of the participants in this dataset is from ages 0 to 100. There are 101,767 rows across 50 columns, and was developed to describe if patients were readmitted within 30 days using race, gender, age, the various types of mediation used, and the readmission of patients.
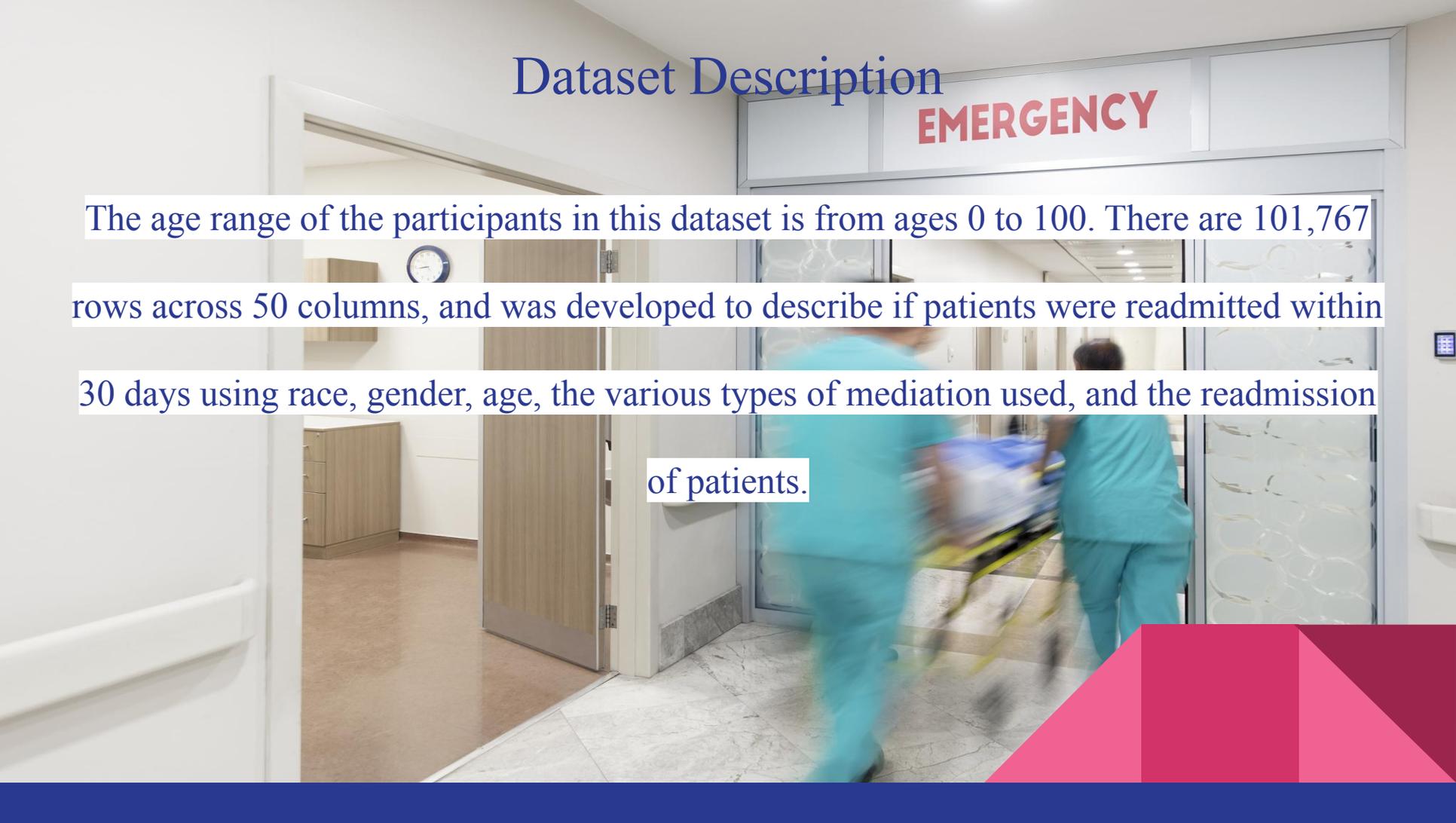
Reading Datasets Elements

To analyze Reading Datasets by Poirier we used these three main elements for the reading.

1. Denotative
   a. The literal reading of the data set
2. Connotative
   a. The interpretive readings
      i. What variables are saying
3. Deconstructive
   a. Investigation in what is not in the data set

Using these elements within the reading will help the reader comprehend the data, learning what it actually means externally and implicitly.

Format Anatomies Element

To analyze *Format Anatomies* by Colin Koopman we used the "How to Look at it Method".

❖ Interpretive inquiry between a dataset's literal meaning, cultural context, and politicized processes of signification

❖ Helps us to understand the literal context as to why the patient was readmitted to the hospital within 30 days

Problem Statement

The Diabetes 130 Dataset focuses on patients, ranging from ages 0-100 years, who were recently discharged from the hospital and whether they were readmitted within 30 days of their discharge date.

❖ Is there enough variables present to determine whether the patient was readmitted because of diabetes?

Different Variables

Encounter ID

Patient Number

Race

Gender

Age

Weight

Admission Type

Discharge Disposition ID

Admission Source ID

Time In Hospital

Payer Code

Medical Specialty

Number of Lab Procedures

Number of Procedures

Number of Medication

Number of Outpatients

Number of Emergencies

Number of Inpatients

Diagnosis 1, 2, & 3

Different Variables

Number of Diagnosis

Medications Prescribed: Insulin, Glipizide, Glyburide, etc.

Change

Diabetes Medication

Readmitted

Missing Variables

Weight

Medical Specialty

Payer Code

Race
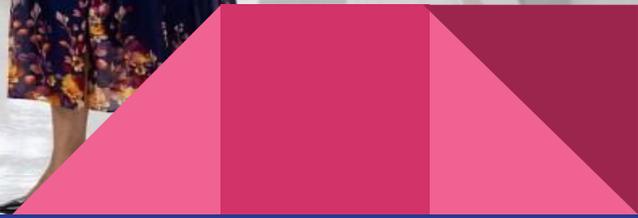
Guides for Certain Variables

RACE

- ❖ Caucasian
- ❖ Asian
- ❖ African American
- ❖ Hispanic
- ❖ Other

AGE

- ❖ Split up by tenths
  - ➢ Ex. [1-10)
- ❖ Goes from (1-100)

TIME IN HOSPITAL

- ❖ Number represents how many days the patient has been admitted to discharged
- ❖ (1-14)

WEIGHT

- ❖ Weight in pounds
- ❖
- ❖ Missing variables

Denotative Reading

❖ Units of observation
 ➢ the hospital stay including variables of day, diabetes diagnosis, medications they're on
 ➢ data is 101,766 hospitalizations NOT patients
❖ If a patient is hospitalized multiple times, they will show up as separate data rows
 ➢ this means that we aren't counting patients, and it doesn't matter what happens to the patient, as long as they entered and exited the hospital they count as a data point
❖ Data Dictionary
 ➢ diagnosis first 3 digits of ICD-9 codes (insurance billing codes)
 ➢ their diagnosis was categorized in terms for money and profit, not the actual complexity of the patient condition
❖ Limits to Variables
 ➢ stays are from 1-14 days, 14+ are excluded, less than 24 hours excluded
 ➢ to be in dataset patient must be prescribed a medication from the medication variables list

Connotative Readings

- ❖ Social contexts
  - ➤ 1999-2008
  - ➤ rise of electronic health records
    - ■ Structure of dataset & software limitations of time period
- ❖ Focus on Readmission Rates
  - ➤ pay-for-performance for hospitals
  - ➤ Hospitals penalized for patients returning too fast & often
  - ➤ dataset reflects interests of insurance companies and hospital admin to cut costs rather than take care of the patient better to keep them out of the hospital
- ❖ How dataset view diabetes
  - ➤ is diabetes a primary diagnosis or secondary

Deconstructive Reading

❖ Erasure of patients who stayed less than 24 hours and over 15 days
  ➢ it gets rid of patients who maybe be older and more ill, with multiple complications, and simplifies diabetes to a more manageable condition than what may be reality

❖ patients who are not on medication are excluded
  ➢ Those who manage with lifestyle changes like diet and exercise are not included
  ➢ medicine and health focused on pharmaceutical view rather than the actual condition

❖ weight is missing for the majority of the data points
  ➢ failed to record weight
  ➢ weight is a primary factor in diabetes, but may not be required for insurance billings
  ➢ it reduces identity of the patients body to lab results

❖ A lot of variables are labeled as "unknown" or "other"
  ➢ so many other types of patients (biracial pts, trans pts)

How to Look at it Method

❖ Framework of Argument
  ➢ Why do diabetes patient need to be readmitted so often? Why is their weight no applicable within this dataset?

❖ How to Look At It
  ➢ The data emphasizes certain variables that show that it is focuses on hospital management and the insurance side of hospitals, rather than the patient side.

Conclusion

This dataset is not representative of the patient or provider experience with diabetes and patients readmission into the hospital or what causes it, but instead the data set is a limited insurance/administrative dataset from the time during a push for more profitable medicine in America. Diabetes is limited to medicated, and shorter inpatient stays, which reflects an institutional press for efficiency in hospitals. By not including weight, and making a lot of the variables limited in their categories or binary, researchers generalize and clean the data and ignore the complex reality of the disease.