

EDA Exercise 3

P. McDonald

August 30, 2016

University of California Irvine maintains an interesting collection of data sets for those interested in machine learning

<https://archive.ics.uci.edu/ml/datasets.html>

Navigate to <https://archive.ics.uci.edu/ml/machine-learning-databases/00296/>

and download the zipped diabetes data set.

Prepare an R Markdown document which documents the use of R tools/code to address the following

1. Read the data into memory.
2. Document any missing values in the data.
3. What percentage of patients are admitted from the emergency room? Given a patient is admitted from the emergency room, what is the probability that their discharge status will be “expired”?
4. What is the most frequent admission status? What is the most frequent discharge status? For the most frequent admission status, what is the most frequent discharge status?
5. Characterize the distribution of admission type.