

Stat 184 Final Project

Soren Epple, Lucas Sadoulet, Tjoet Lakmana

2025-05-01

Introduction

Data Provenance

FAIR Principles

The dataset we used for this project aligns with what are known as the FAIR Principles. These principles help guide how data should be handled to make it more usable and shareable across different settings.

Findable

You can find this dataset easily on Kaggle, which is a popular platform for sharing and discovering datasets. However, this specific dataset does not come with detailed metadata or column explanations, so some interpretation and exploration are needed to fully understand its structure. Still, it is clearly labeled and easy to locate on the platform.

Accessible

Getting access to this data is straightforward. There are no paywalls or complicated steps, you just need a free Kaggle account. That means anyone interested in exploring the dataset, whether for school or personal projects, can do so easily.

Interoperable

The data is in CSV format, which is a standard that works with all sorts of tools like R, Python, Excel, and more. This makes it convenient to load into whatever program you're most comfortable using.

Reusable

This dataset is shared in a way that encourages others to use it. Because it's clearly structured and documented, it's great for reuse in class projects, personal analysis, or even more advanced research.

CARE Principles

The CARE Principles are all about making sure data involving Indigenous communities is handled ethically and respectfully. In this case, the dataset doesn't include any cultural or community-specific information, so the CARE guidelines don't really apply here.