

**WASHINGTON STATE DEATH
CERTIFICATE DATA 2003-17**

DATA CLEANING AND STANDARDIZATION

1. Limit data to decedents who were WA residents who died in WA.
2. Standardize attribute names and values
3. Limit data set to relevant attributes (variables)
4. Derive new attributes e.g. age group, leading causes of death

n = 745,935

**CREATE SUBSET OF KING COUNTY DEATHS
n = 200,692**

**CREATE SUBSET OF KING COUNTY DECEDENTS
WITH PERMANENT HOMES AT TIME OF DEATH
n = 174,292**

EXPLORATORY DATA ANALYSIS & MACHINE LEARNING

**KING COUNTY MEDICAL EXAMINER
HOMELESS DEATH DATA 2003-17**

**DATA CLEANING AND
STANDARDIZATION**

Standardize
attribute names
and values

n = 1,131

**LINK HOMELESS DEATHS TO
THEIR DEATH CERTIFICATES
n = 1,104**

EXPLORATORY DATA ANALYSIS & MACHINE LEARNING