
INTRODUCTION TO DATA CLEANING

WITH OPENREFINE & WORD CLOUDS

Meg Miller, GIS & Data Visualization Librarian
Natalie Vielfaure, Digital Curation Archivist
University of Manitoba Libraries



SESSION BREAKDOWN



Software Overview



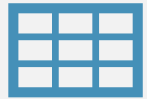
Data Cleaning: What is it?



Benefits



Hands-on Exercises



OpenRefine

What is it?
Uses



Voyant Tools

What is it?
Word Clouds

SOFTWARE OVERVIEW

A word cloud where the words are arranged to form the outline of the map of Canada. The word 'Ontario' is the largest and most prominent, centered in the middle. Other large words include 'Alberta', 'British Columbia', and 'Quebec'. Smaller words include 'Manitoba', 'Saskatchewan', 'Nova Scotia', 'Newfoundland', 'Yukon', 'Labrador', 'Toronto', 'Oshawa', 'Alberqrrta', 'Nb', 'Ab', 'NI', 'Ns', 'Sk', 'Qc', 'On', 'Qc', 'Quebec', 'Canada', 'Noyb', 'Saskatchewan', and 'Ont'. The colors of the words are varied, including shades of green, blue, orange, and purple.

Labrador Yukon
Toronto Scotia
Newfoundland
Columbia Alberta And
Québec Manitoba NI Ab
On Ontario Ns
Qc Quenec Sk
Oshawa Alberqrrta Nb British
Canada Quebec
Noyb Bc Ont
Saskatchewan

VOYANT-TOOLS.ORG/

EXERCISE I

DATA CLEANING: WHAT IS IT?

- Improving the overall quality of your data.

Manitoba

MB

Man.

Manitoba

Manitoab

mb

RESOLVING INCONSISTENCIES

- Variant spelling
- Inconsistent case
- Duplication

FORMATTING



UNIT OF MEASUREMENTS



DATES, TEXT, NUMBERS



COLUMNS / ROWS



FORMAT MIGRATION

NULL VALUES

-
- No value
 - 0 value
 - Errors
 - Examples:
 - =MODE.MULT(A2:G2) → #NULL!
 - =SUM(A2:A5) → #NAME?

DATA CLEANING: BENEFITS

**Efficient
research**

**Easier
conversion**

**Secondary
use**



bit.ly/orf_2023

HANDS ON EXERCISES