Ujjval Rajput

22 October, 2023

1. Identification and labeling of each field in the dataset:

REF_DATE - Name: Reference Date, Type: Date

GEO - Name: Geography, Type: Categorical (String)

DGUID - Name: Data Guide, Type: Categorical (String)

Building type - Name: Building Type, Type: Categorical (String)

Construction type - Name: Construction Type, Type: Categorical (String)

UOM - Name: Unit of Measurement, Type: Categorical (String)

UOM_ID - Name: Unit of Measurement ID, Type: Numeric (Integer)

SCALAR_FACTOR - Name: Scalar Factor, Type: Categorical (String)

SCALAR_ID - Name: Scalar ID, Type: Numeric (Integer)

VECTOR - Name: Vector, Type: Categorical (String)

COORDINATE - Name: Coordinate, Type: Categorical (String)

VALUE - Name: Value, Type: Numeric (Float and Integer, depending on the nature of the values)

STATUS - Name: Status, Type: Categorical (String)

SYMBOL - Name: Symbol, Type: Categorical (String)

TERMINATED - Name: Terminated, Type: Categorical (String)

DECIMALS - Name: Decimals, Type: Numeric (Integer)

2. Visualization strategy for the data set by referring to the representation of the types based on The Mackinlay ranking of perceptual tasks:

a. Temporal (Time-Series) Data (REF_DATE):

Use line charts to show trends in the data over time.

Consider using a time slider to allow users to interactively explore data across different time periods.

    b.   Categorical Data (GEO, Building type, Construction type, etc.):

For single-category comparisons, use bar charts or pie charts.

For multiple-category comparisons, consider stacked bar charts or grouped bar charts.

Use color coding for different categories to aid visual interpretation.

    c.   Numeric Data (VALUE):

Use bar charts for comparing numerical values across different categories.

Scatter plots can be useful to show the relationship between two numeric variables (if applicable).

Consider using heatmaps for visualizing patterns in a matrix of values.

    d.   Ordinal Data (STATUS, SYMBOL, etc.):

Use bar charts or dot plots to show the distribution of ordinal data.

Consider using different shapes or colors to represent different ordinal categories.

    e.   Categorical-Numeric Combination (UOM, SCALAR_FACTOR, etc.):

Consider using bar charts to show the distribution of categorical variables.

For combinations of categorical and numeric data, stacked bar charts or grouped bar charts can be effective.

    f.   Spatial Data (GEO, DGUID):

If your data has a spatial component, consider using maps to show geographic distributions.

Choropleth maps can be effective for visualizing variations in data across different regions.

    g.   Magnitude Data (VALUE, UOM_ID, etc.):

Use bar charts or dot plots to represent magnitudes.

Consider log scales for better representation of wide-ranging magnitudes.