

Teen Health Risk Index: A County-Level SDOH Composite Score for Maryland

Tools Used: R, Excel, Tableau Public

Data: Synthetic dataset modeled after common public SDOH indicators

Skills Demonstrated: Data cleaning, z-score standardization, composite scoring, geographic visualization, public health analytics, dashboard design

1. Project Overview

Adolescent health outcomes vary significantly across Maryland due to differences in underlying **Social Determinants of Health (SDOH)** including poverty, education, and community safety. To support decision-making and highlight regional disparities, I recreated a county-level composite index reflecting **teen health risk** using fully synthetic data modeled after real-world public health indicators.

The final deliverable included:

- A clean county-level dataset
- Standardized z-score transformations of health and SDOH indicators
- A composite **Teen Health Risk Index**
- **Quintile classification** of all 24 Maryland jurisdictions
- A **Tableau choropleth map** visualizing risk levels across the state

This project mimics a real evaluation workflow I completed in a professional public health context, but uses non-confidential synthetic data for sharing.

2. Live Interactive Map

View the full interactive choropleth map on Tableau Public:

👉 Live Dashboard:
<https://public.tableau.com/app/profile/victoria.de.los.santos/viz/MarylandTeenHealthRiskIndexSyntheticData/Dashboard1?publish=yes>

This map allows users to explore the teen risk index by county, view tooltips with underlying indicators, and compare regions across Maryland.

3. Data & Indicators

The synthetic dataset includes the 24 Maryland counties plus Baltimore City, with the following variables:

- **Teen Birth Rate**
- **Poverty Rate**
- **High School Dropout Rate**
- **Juvenile Arrests**
- **Unemployment Rate**

These indicators were selected because they are commonly associated with teen risk and structural disadvantage in public health research.

4. Methods & Analytical Approach

Step 1 — Data Cleaning (R)

- Imported the dataset from Excel
- Ensured county names matched Tableau's built-in geography format
- Verified data types for all indicators
- Removed formatting inconsistencies (apostrophes, spacing, etc.)

Step 2 — Standardization Using Z-Scores (R)

Each SDOH indicator was converted into a **z-score**:

$$z = \frac{x - \mu}{\sigma}$$

This puts all variables on the same scale, allowing them to be combined fairly.

Step 3 — Composite Index Calculation

All z-scored variables were averaged to create a **Teen Health Risk Index** for each county:

$$\text{Risk Index} = \text{mean of all } z - \text{scores}$$

Step 4 — Quintile Classification

Counties were divided into **five quintiles**, where:

- **1 = Lowest risk**

- **5 = Highest risk**

Step 5 — Geographic Visualization in Tableau

Using Tableau Public:

- County boundaries were automatically recognized using geographic roles
- Quintiles were used as a discrete color category
- A sequential stepped color palette (5 steps) was applied
- Tooltip included county name, quintile, index score, and raw indicators

The final map highlights geographic disparities and allows viewers to quickly identify higher-risk counties.

5. Key Findings (Synthetic Data Interpretation)

Although the dataset is synthetic and not tied to real Maryland health outcomes, the model demonstrates:

- **Clear regional clustering** of higher-risk synthetic counties
- Higher composite scores consistently aligned with elevated poverty, unemployment, and dropout indicators
- Lower-risk counties tended to have stronger socioeconomic indicators

This showcases how public-health teams can visualize SDOH-driven patterns to guide resource allocation.

6. What This Project Demonstrates

Technical Skills

- R-based data wrangling
- Using `dplyr` for transformations
- Standardizing variables via `scale()`
- Composite index development
- Exporting clean data for visualization
- Geographic analysis and map creation in Tableau
- Tooltip design & color theory in choropleths

Public Health Evaluation Skills

- SDOH indicator selection
 - Interpretation of standardized composite scores
 - Communicating risk through quantitative indices
 - Use of synthetic data to preserve privacy while demonstrating skill
-

7. Deliverables

- **R Script:** Cleaning, z-scores, composite index, quintiles, CSV export
 - **CSV Dataset:** Fully synthetic Maryland SDOH indicators
 - **Tableau Map:** Quintile-based county-level visualization
 - **GitHub Repository:** Full project files + documentation
-

8. Final Reflection

This project highlights my ability to combine **data science, public health evaluation, and geospatial visualization** to create actionable insights. By reconstructing my real-world work with synthetic data, I can safely demonstrate the full workflow — from raw data to a polished, interactive map suitable for stakeholders and program staff.