OVERVIEW OF DATASETS

GitHub repository:
https://github.com/maya-papaya/ads1-cervical-cancer-analysis/tree/main

CERVICAL CANCER RATES

- GBD Results (cervical_cancer_USA.csv)

  - **Usage:** This dataset will be the main source of information on cervical cancer incidence rates and mortality estimates from 1980 to 2021 for states within the United States.

  - **Number of Files / Rows:** 1 file, 7549 rows

  - **Relevant Columns:**

    - measure

    - location

    - sex

    - age

    - cause

    - metric

    - year

    - val

    - upper

    - lower

HPV VACCINES

- Cervical Cancer & HPV Vaccines (in hpv_vaccines folder)

  - **Usage:** These datasets will be used to provide information on HPV vaccinations and the number of cervical cancer cases prevented due to vaccination, as well as the cost efficiencies of the HPV vaccination program. The files range from 2010 to 2020 and include 176 countries.

- ○ **Number of Files / Rows:** 11 files, 1936 rows (176 each)

- ○ **Relevant Columns:**

  - ■ country

  - ■ cohort_size = total number of girls who are in the cohort that should be getting vaccinated

  - ■ current_cov = coverage in X year

  - ■ curr_vacc_cohort_size = number vaccinated in that cohort

  - ■ future_cov = 90% coverage rate

  - ■ future_vacc_cohort_size = number vaccinated in cohort under new coverage rate

  - ■ curr_cc_prev = cervical cancer cases prevented under the current coverage

  - ■ curr_mort_prev = deaths from cervical cancer prevented under the current coverage

  - ■ curr_cost = current cost of the program

  - ■ curr_cost_prev = current costs prevented by reducing cervical cancer

  - ■ proj_cc_prev / proj_mort_prev / proj_cost / proj_cost_prev = same as above but under the "projected" 90% coverage rate the WHO says is necessary for elimination of cervical cancer

- ● https://healthdata.gov/dataset/Vaccination-Coverage-among-Adolescents-13-17-Years/47pk-jpce/about_data (adolescent_vaccine_coverage.csv)

  - ○ **Usage:** This dataset will be used to gather information on HPV vaccine coverage and sociodemographic factors among adolescents between the ages of 13 and 17 years. This data is specifically about the United States.

  - ○ **Description from site:** "Vaccination Coverage among Adolescents (13-17 Years). Data on adolescent vaccination coverage and selected sociodemographic characteristics by State, HHS Region, and the United States from the National Immunization Survey-Teen (NIS-Teen)."

  - ○ **Number of Files / Rows:** 1 file, 27566 rows

- ○ **Relevant Columns:**

    - ■ Vaccine/Sample = name of the vaccine or survey sample for which estimates are calculated

    - ■ Dose = (this is only applicable for vaccination coverage estimates) when populated, it shows dose of vaccine (≥1, ≥2, ≥3, ≥4) for which coverage is estimated

    - ■ Geography

    - ■ Survey Year

    - ■ Dimension Type = the classification of the sociodemographic category for which estimates are calculated. For vaccination coverage in a single survey year, the only dimension type is age. For five-survey year estimates, estimates are calculated overall and by insurance coverage, urbanicity, race/ethnicity, and poverty.

    - ■ Dimension = sociodemographic group for which estimates are calculated

    - ■ Estimate (%) = estimated vaccination coverage or percent of sample represented by each survey sociodemographic characteristic

    - ■ 95% CI (%)

PAP SMEAR TESTS

- ● https://data.cdc.gov/500-Cities-Places/500-Cities-Census-Tract-level-Data-GIS-Friendly-Fo/kucs-wizg/about_data (pap_smear_2014.csv)

    - ○ **Usage:** This dataset will be used to estimate the number of Pap smear tests that were issued in the United States in 2014.

    - ○ **Description from site:** "2015, 2014. Data were provided by the Centers for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology and Surveillance Branch. The project was funded by the Robert Wood Johnson Foundation (RWJF) in conjunction with the CDC Foundation. 500 cities project census tract-level data in GIS-friendly format can be joined with census tract spatial data (https://chronicdata.cdc.gov/500-Cities/500-Cities-Census-Tract-Boundarie

s/x7zy-2xmx) in a geographic information system (GIS) to produce maps of 27 measures at the census tract level. Because some questions are only asked every other year in the BRFSS, there are 7 measures in this 2017 release from the 2014 BRFSS that were the same as the 2016 release."

- ○ **Number of Files / Rows:** 1 file, 27211 rows

- ○ **Relevant Columns:**

  - ■ StateAbbr = state abbreviation

  - ■ PlaceName = city name

  - ■ population_count = 2010 Census population count

  - ■ PAPTEST_CrudePrev = model-based estimate for crude prevalence of papanicolaou smear use among adult women aged 21–65 years, 2014

  - ■ PAPTEST_Crude95CI = estimated confidence interval for crude prevalence of papanicolaou smear use among adult women aged 21–65 years

- ● https://data.cdc.gov/500-Cities-Places/500-Cities-Census-Tract-level-Data-GIS-Friendly-Fo/k86t-wghb/about_data (pap_smear_2016.csv)

  - ○ **Usage:** This dataset will be used to estimate the number of Pap smear tests that were issued in the United States in 2016.

  - ○ **Description from site:** "2017, 2016. Data were provided by the Centers for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology and Surveillance Branch. The project was funded by the Robert Wood Johnson Foundation (RWJF) in conjunction with the CDC Foundation. 500 cities project census tract-level data in GIS-friendly format can be joined with census tract spatial data (https://chronicdata.cdc.gov/500-Cities/500-Cities-Census-Tract-Boundaries/x7zy-2xmx) in a geographic information system (GIS) to produce maps of 27 measures at the census tract level. There are 7 measures (all teeth lost, dental visits, mammograms, Pap tests, colorectal cancer screening, core preventive services among older adults, and sleep less than 7 hours) in this 2019 release from the 2016 BRFSS that were the same as the 2018 release."

  - ○ **Number of Files / Rows:** 1 file, 27211 rows

○ **Relevant Columns:**

- StateAbbr = state abbreviation

- PlaceName = city name

- Population2010 = 2010 Census population count

- PAPTEST_CrudePrev = model-based estimate for crude prevalence of papanicolaou smear use among adult women aged 21–65 years, 2016

- PAPTEST_Crude95CI = estimated confidence interval for crude prevalence of papanicolaou smear use among adult women aged 21–65 years