|  |  |
| --- | --- |
| **NIH SciEd 2024**  **AI Data Activity** |  |

**EXAMPLES AND LINKS FROM SEPA “*DATA DETECTIVES – USING BIG DATA TO SOLVE REAL PROBLEMS IN REAL COMMUNITIES*”**

Theresa Gillespie, PhD & Adam Marcus, PhD, M-PI

**AI Section of Data Detectives Curriculum**

1. Set up Google Gemini to ask questions

* Go to [gemini.google.com](http://gemini.google.com/) (Note: You will need a gmail or Google account for this)
  1. As a group, or in pairs, pick specific cancers and counties to test Google Gemini’s responses:
     1. What is **incidence** of different cancers (e.g., breast, lung) in Cherokee County & state of GA?
     2. What is **mortality** of different cancers (e.g., prostate, lung) in Cherokee County & state of GA?
     3. What are important **risk factors** (smoking, obesity, physical inactivity) that may lead to cancers in Cherokee County & state of GA?
     4. Record responses provided by Google Gemini in the table on the next page.
  2. Use links to Cancer Big Data to validate information provided by Google Gemini
     1. Links for Cancer Big Data:
        1. **State Cancer Profiles** (CDC, NCI) – incidence, mortality, demographics, risk factors by tables, maps <https://statecancerprofiles.cancer.gov/>
* Also includes 2020 BRFSS screening & risk factor survey
* 2019 American Community Survey Data
* Cancer Screening and risk factor data
* Smoking statistics

2. **US Cancer Statistics – Data Visualizations (CDC)**

<https://gis.cdc.gov/Cancer/USCS/#/AtAGlance/>

* Specific cancers, incidence, mortality, by sex, race, ethnicity
* Stage at diagnosis, screening & risk factors, prevalence, trends
* CDC - Chronic Disease Indicators - <https://www.cdc.gov/cdi/>
* State level data for chronic diseases, including cancer, and risk factors
* Includes screening data (mammography, Pap test, colorectal screening)

3. **Behavioral Risk Factor Surveillance System (BRFSS)**

* 2020 data: <https://www.cdc.gov/brfss/>
* Includes phone survey data for states related to risk behaviors, e.g., fruit and vegetable consumption, physical activity, use of screening and prevention

4. Ask the same cancer data questions you asked using Google Gemini but this time, use the Cancer Big Data.

5. Record data derived from Cancer Big Data sources & compare to Google Gemini data.

**Discussion Questions about Using AI and Cancer Big Data:**

* What are discrepancies noted?
* How large are the discrepancies?
* How important might these discrepancies be if you were planning an intervention to address some of the cancer data findings, such as a plan to reduce cancer mortality from colon cancer?
* Why do you think these differences in the data occur?
* How accurate and valid do you think cancer data from Google Gemini are compared to the national/state data from the Cancer Big Data sites we used?
* How do these findings influence your view of using ChatGPT or Google Gemini or other AI tools to write an essay or do your homework?

**Using AI and Cancer Big Data Table**

| **Research Question** | **Google Gemini** | **State Cancer Profiles** | **US Cancer Statistics** | **BRFSS** |
| --- | --- | --- | --- | --- |
| 1. What is **incidence** of different cancers (e.g., breast, lung) in Cherokee County & state of GA? |  |  |  |  |
| 1. What is **mortality** of different cancers (e.g., prostate, lung) in Cherokee County & state of GA? |  |  |  |  |
| 1. What are important **risk factors** (smoking, obesity, physical inactivity) that may lead to cancers in Cherokee County & state of GA? |  |  |  |  |