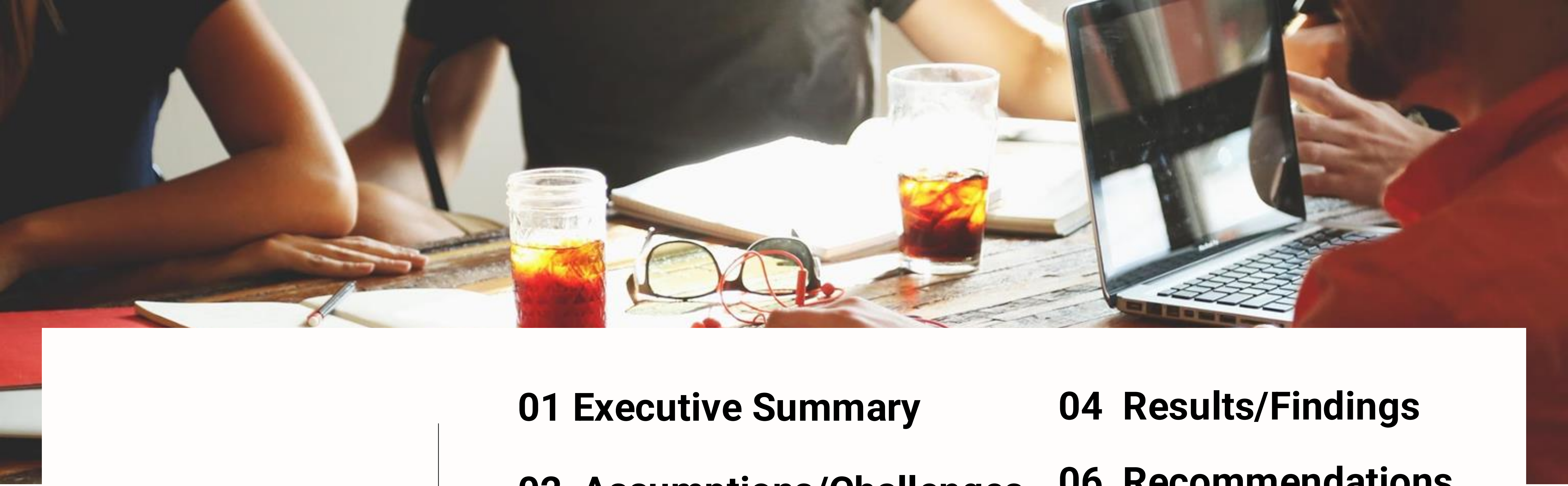
Unveiling Iowa's Cancer Landscape: A Data-Driven Capstone Investigation

AArete 3-2

The University of Iowa



Agenda

Executive Summary

Problem Description:

Iowa is experiencing a troubling rise in cancer incidence across all major demographic groups, despite national trends showing an overall decline.

Objective:

We aimed to analyze and identify 5-10 factors significantly associated with the rising Iowa cancer rates contributing factors by analyzing publicly available federal and state health data for Iowa’s four most prevalent cancer types: breast, lung, melanoma, and prostate.

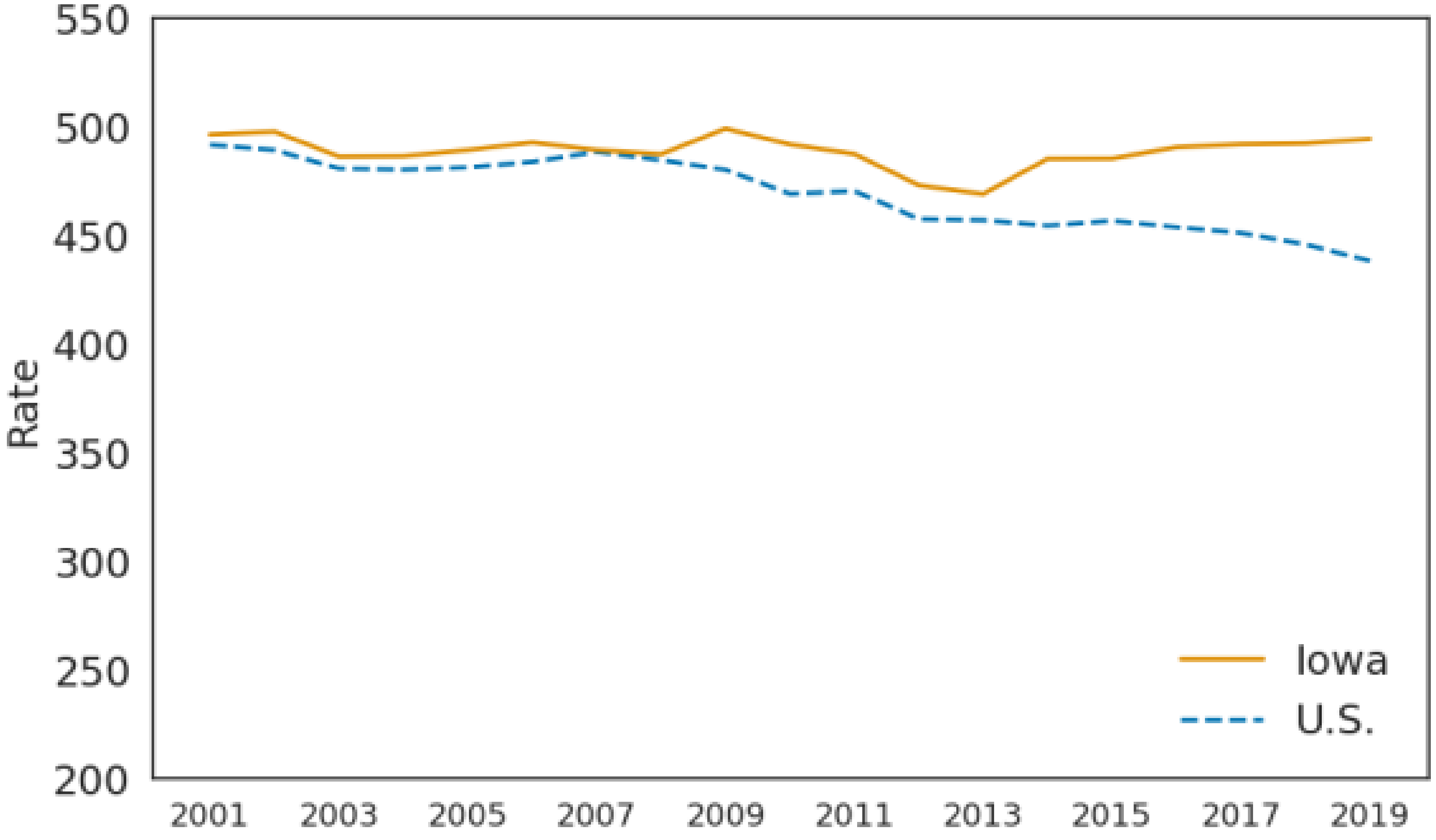
Our Purpose:

To uncover statistically and conceptually significant drivers behind Iowa’s rising cancer rates, using data-driven modeling to inform public health strategies and improve future health outcomes.

Current Iowa Challenges

- Iowa is the second highest new cancer incidence rate state in the US
- The current US rate is declining
- Spans across all major demographic groups

Age-Adjusted Incidence Trends for All Cancer Types in Iowa compared to Other States (per 100,000 population)



(University of Iowa, College of Public Health, 2024)

What Iowa is Currently Doing

Some of Iowa's Cancer Plan Priorities:

1. Reduce exposure to tobacco and secondhand smoke
2. Decrease alcohol consumption
3. Reduce Exposure to environmental carcinogens
4. Decrease exposure to radon
5. Reduce ultraviolet (UV) radiation exposure

(Iowa Cancer Consortium, 2023)

Assumptions and Challenges

Data Assumptions for Models

- **US Breast, Lung, Melanoma, and Prostate Cancer Rates** – 5-year average (2017-2021)
- **Behavioral Factors –** Our marijuana, Cocaine, Meth, Pain Med, Binge Drinking, Tobacco, Cigarette, and Substance Use Disorder Rates by state are from 2018, 2019, 2022, and 2023.
- **US Air Quality and Pesticide Usage –**The only public US dataset for these were found over a single-year span (2018-19).
- **Obesity Rate –** The only US public dataset was found from 2023
- **US Radon Data –** Our US radon data found is a rolling average forecast by state for 2024

Challenges

- Overall limited access to publicly-available data over the last 5-year timeframe
- Final OLS (Ordinary Least Squares) models **excluded state identity variables** to focus on general risk factors
- Orange Software to generate predictive models

Data Gathering / Preparation

Original Data

- **CDC, National Cancer Institute, Kaiser Family Foundation (KFF), Substance Abuse and Mental Health Services Administration (SAMHSA)**, **US Census Bureau**

Data Sources Covered

- Cancer Incidence Rates (**4**)
- Environmental exposure factors (**4**)
- Behavioral risk factors (**8**)
- Socio-demographic metrics (**3**)

Data Cleaning/Preprocessing

- Utilized RStudio IDAS (~540 lines of code)
- Standardized all datasets at the **state level**
- Removed 20 rows with null values
- Converted State variable to categorical

Data Selection

- **Public health relevance** (risk factors)
- **Modeling diagnostics** (removal of multicollinearity)
- **Data completeness** across 50 states

Final Dataset

- **180 rows, 35 columns with no missing data**
- 18 merged dataframes
- **Target Variables:** Age-adjusted Breast, Lung, Melanoma, and Prostate cancer incidence rates (cases per 100,000)

Modeling Methodologies

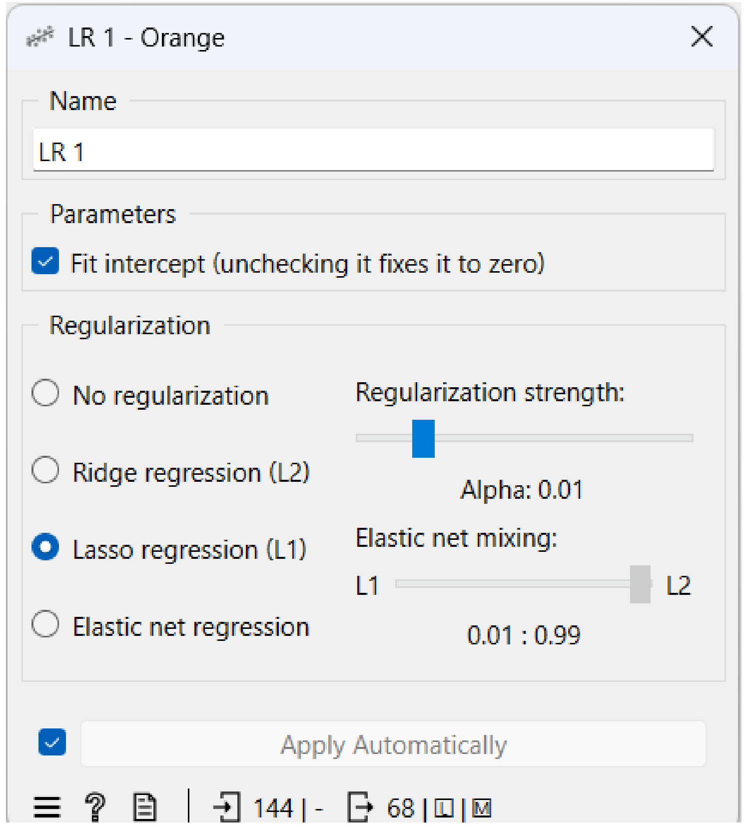
Modeling Techniques

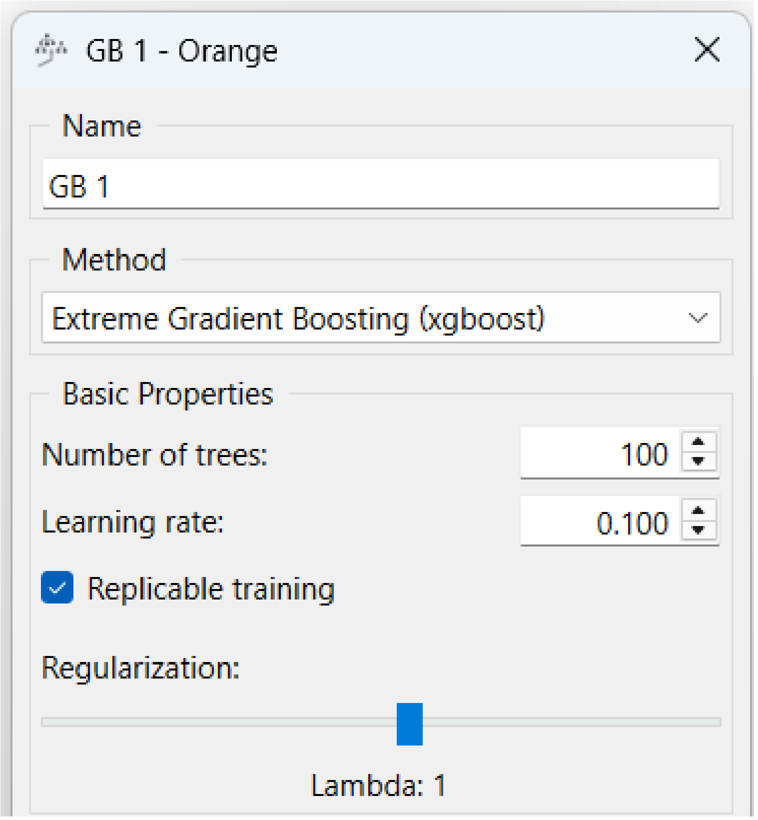
- Linear Regression **(Orange)**
- Gradient Boosting using scikit-learn **(Orange)**
- Ordinary Least Squares **(OLS) (Python)**

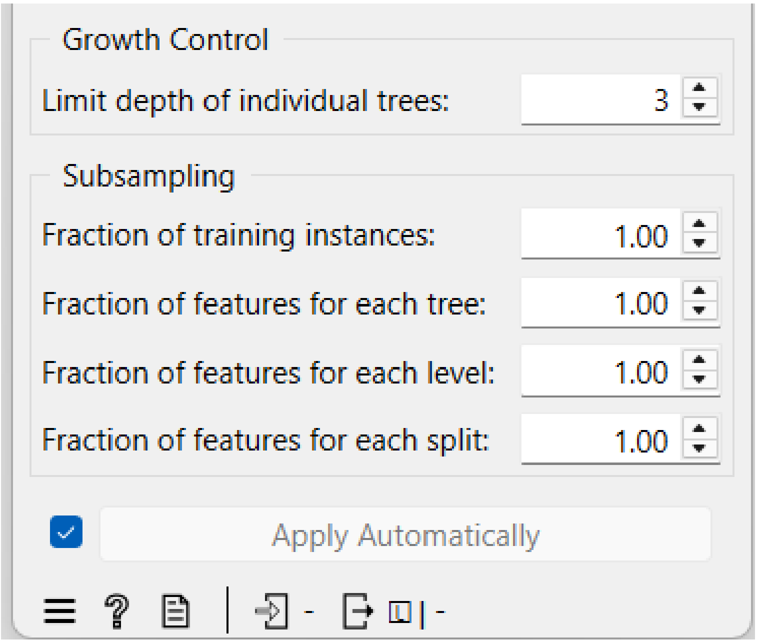
Evaluation Criteria

- **R^2:** measures the proportion of variance in cancer incidence rates explained by the predictors
- **RMSE:** measures the average prediction error in cancer incidence rates (per 100,000)
- **Cross-validation**
- **Post-model Multicollinearity Diagnoses**

Predictive Model Parameters



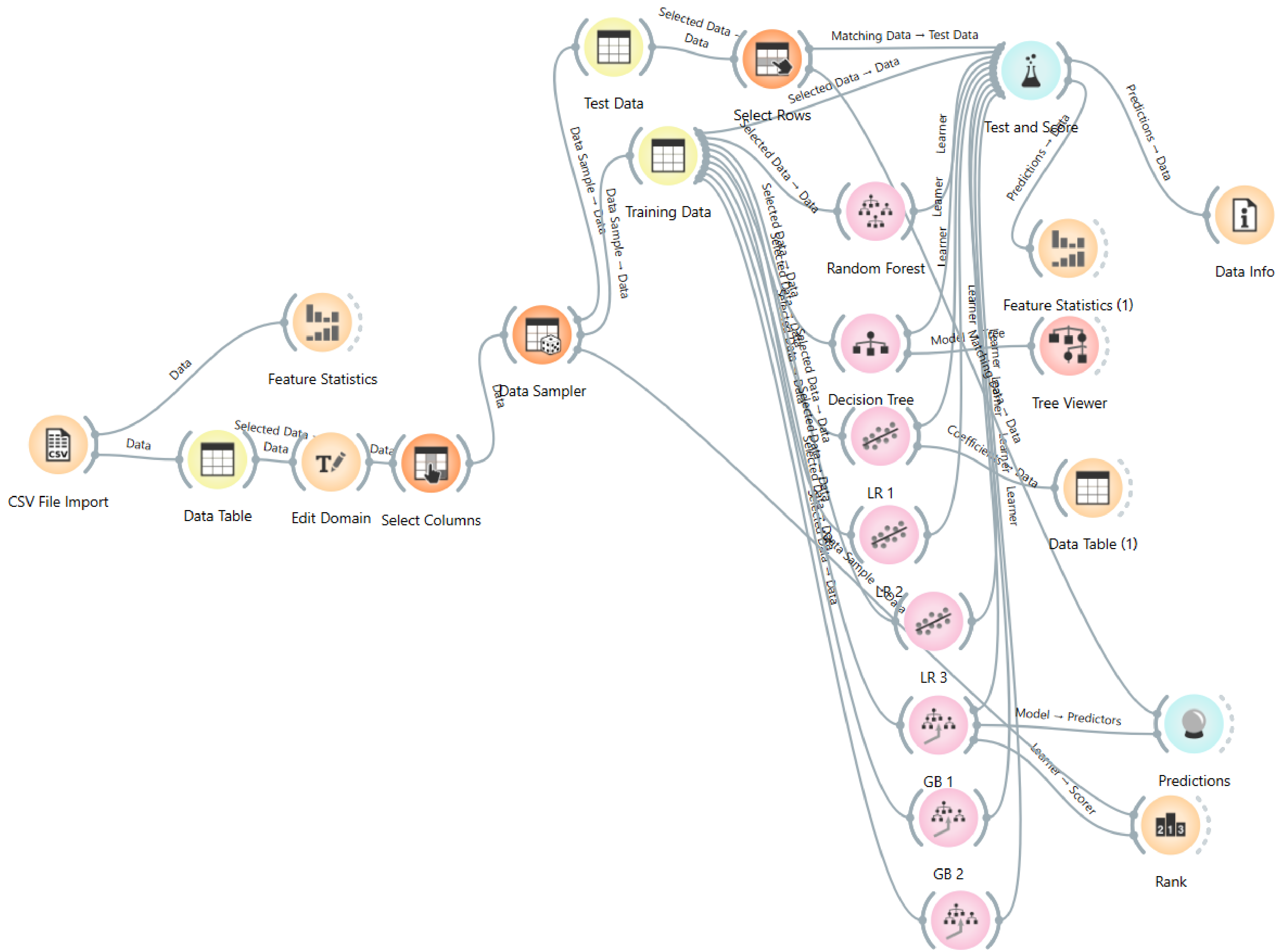




- **LR 1:** Best of the 3 LR models
- **Parameters:** Lasso Regularization for accuracy and overfitting prevention and eliminating unrelated factors

- **GB 1:** Best of the 3 GB models
- **Parameters:** Default learning rate, limited number of trees for overfitting prevention, L2 Regularization

Predictive Model Environments and Validation



- 8 models checked for 4 age-adjusted cancer incidence rates per 100,000 cases
- Ordinary Least Squares (OLS, Python) used to validate statistically significant predictors across cancers

Combined Key Cancer Drivers

Cancer Type	Key Drivers (α = 0.05)
Breast	Meth Rate 18+, PM2.5, SUD Rate 18+, Drinking Rate 18+
Lung	UV Exposure, Tobacco Rate 18+
Melanoma	UV Exposure, PM2.5
Prostate	Meth Rate 18+, PM2.5, SUD Rate 18+

Methamphetamine Usage and Cancer Risk In Iowa

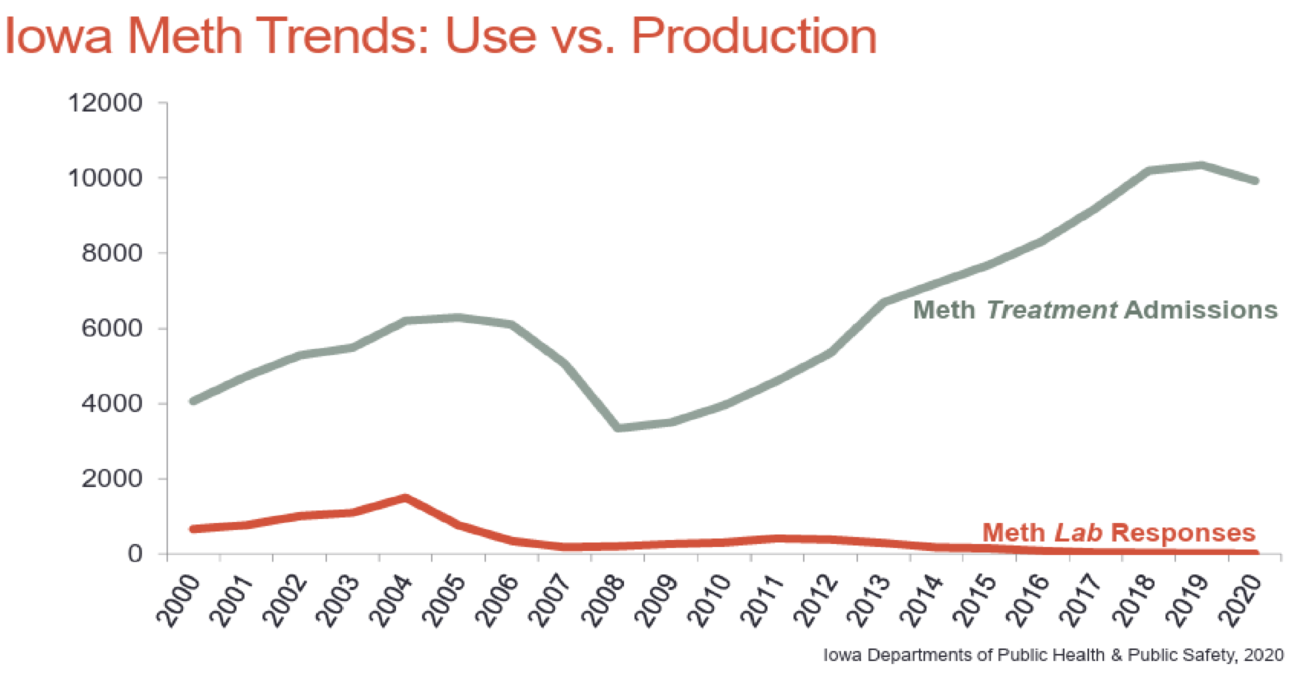
Model Findings

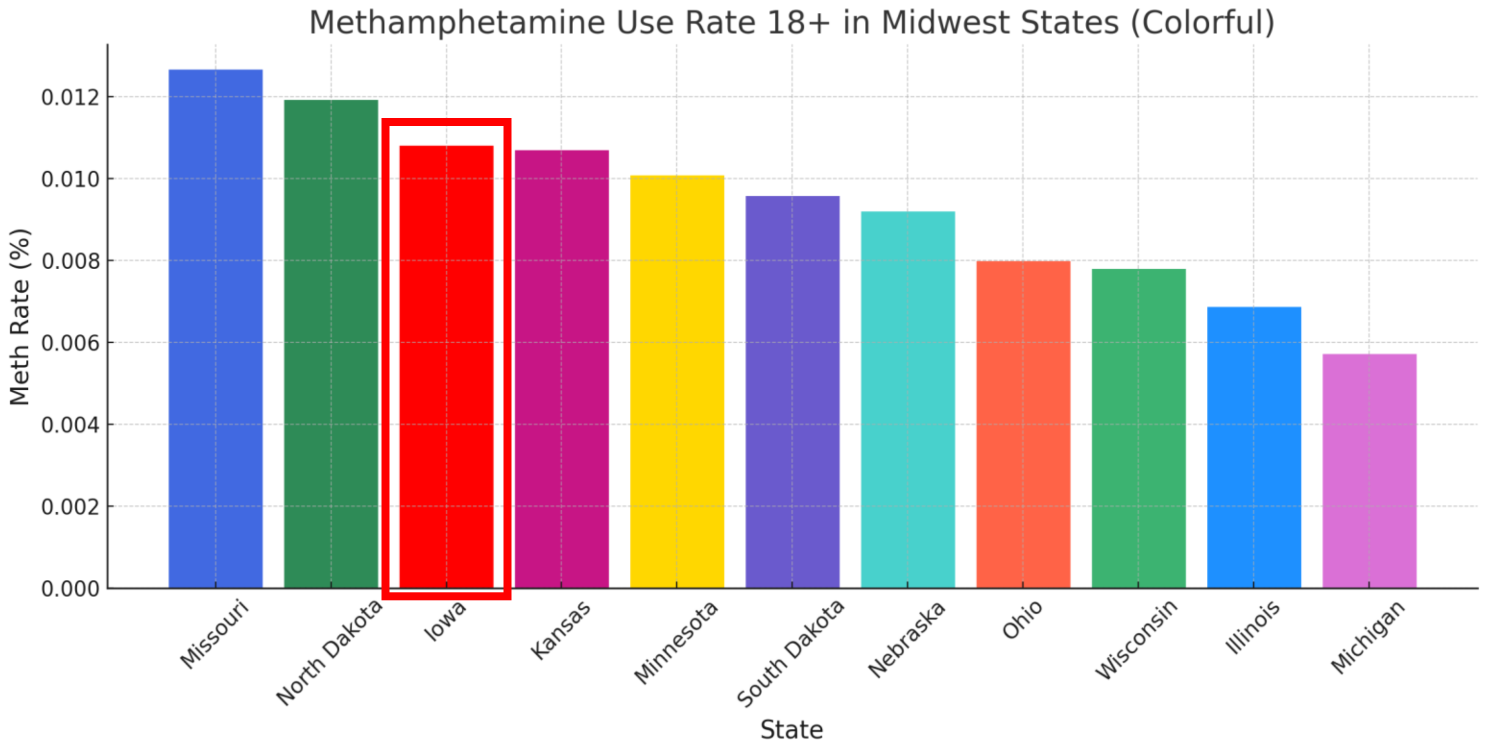
- **P-values:**
  - Breast Cancer: **p = 0.034** (significant)
  - Lung Cancer: **p = 0.02** (significant)
- **Iowa has the 8th highest rate of meth use disorder treatment admissions in the US (SAMHSA, 2020)**

Supplemental Research

- Smoking methamphetamine **damages lung tissue**, leading to conditions like pneumonia and acute respiratory distress syndrome (National Library of Medicine, 2022)

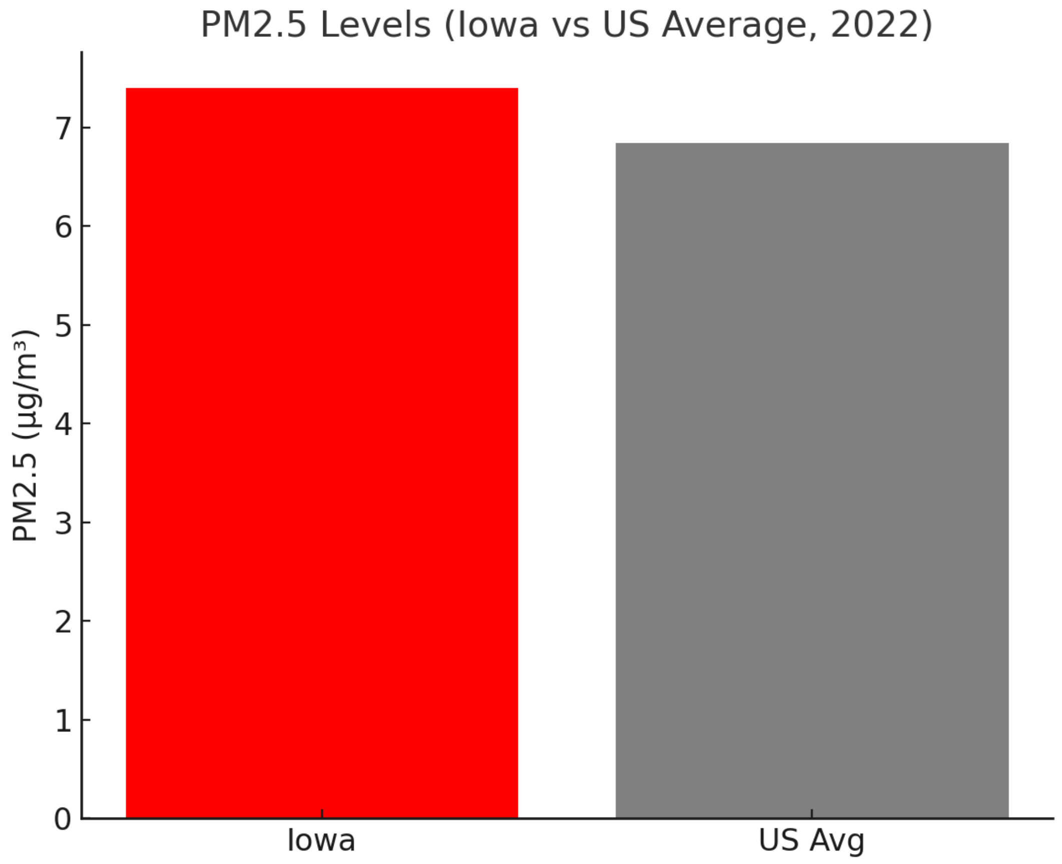
- Methamphetamine **induces oxidative stress** and **DNA damage**, processes implicated in **cancer development** (National Library of Medicine, 2022)

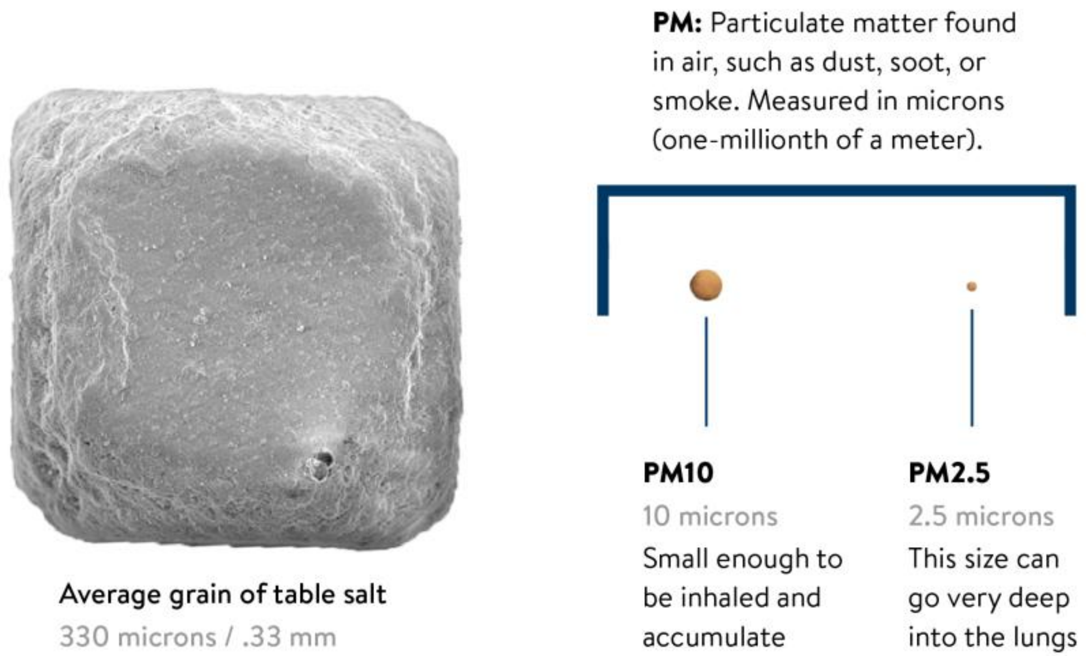




Particulate Matter: The Air We Breathe

Higher PM2.5 Levels in Iowa May Contribute to Elevated Cancer Rates





Source: Minnesota Pollution Control Agency, 2025

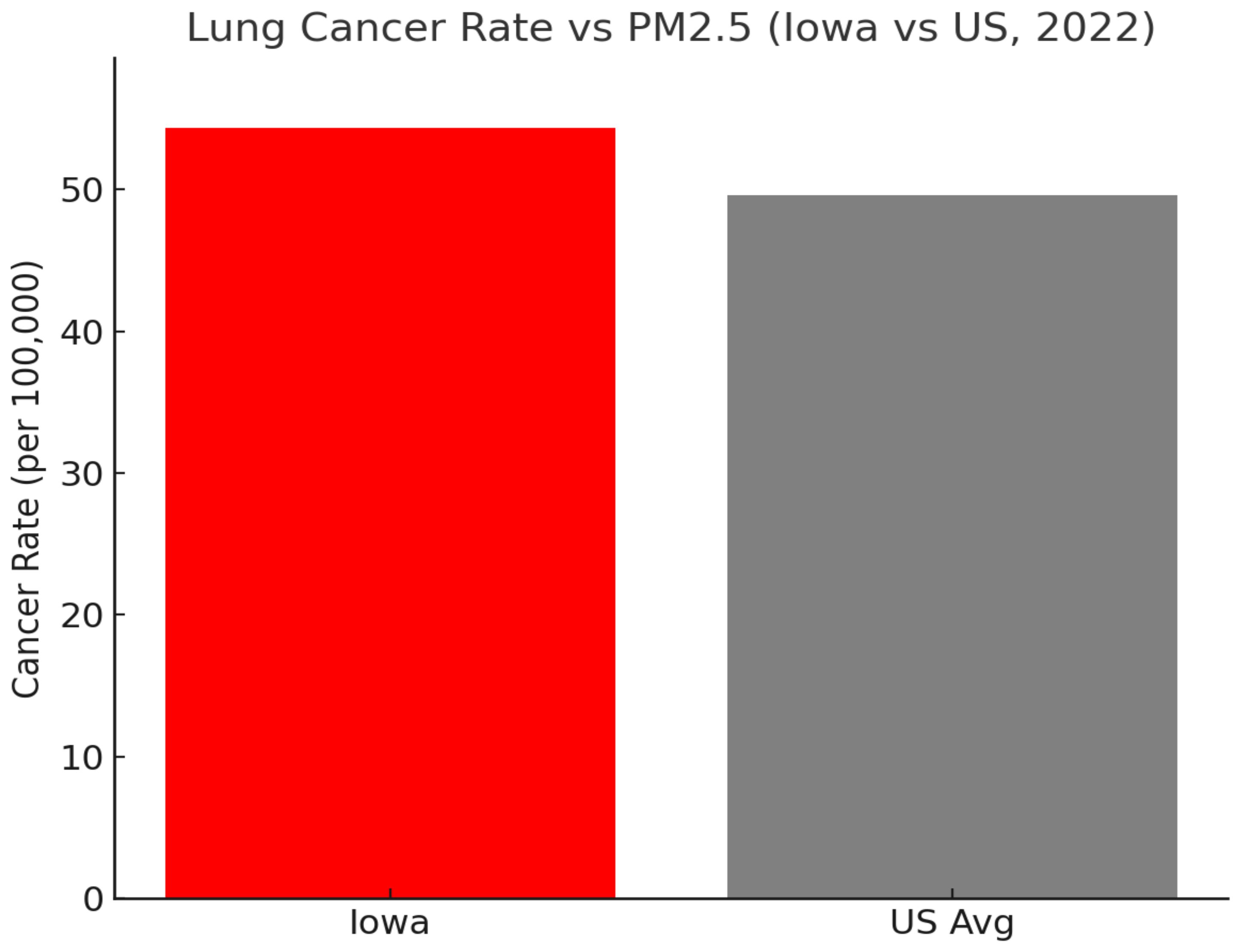
Model Findings:

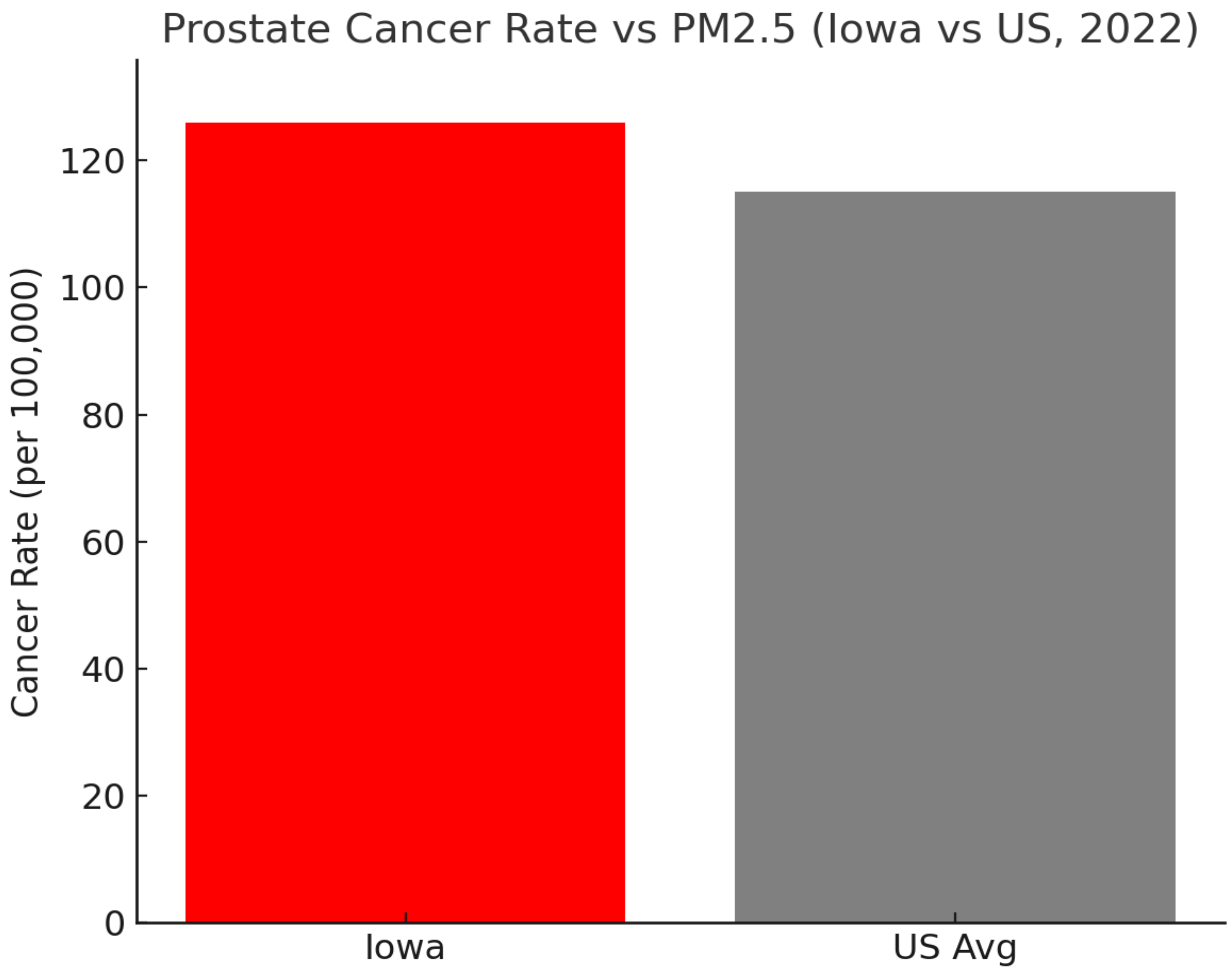
- **P-values:**
  - Lung Cancer**: $p < 0.01$** (significant)
  - Prostate: **$p < 0.05$** (significant)

Interpretation:

Particulate matter may contribute substantially to Iowa's cancer rates through airborne exposure from farming and industry.

PM2.5 vs Cancer Findings: Iowa vs US Average





UV Exposure: An Inverse Finding

Model Findings

- **P-values:**
  - Highly significant ($p < 0.001$) for breast, lung, melanoma, and prostate cancers

- Iowa's UV exposure is comparable to other states with the highest cancer rates

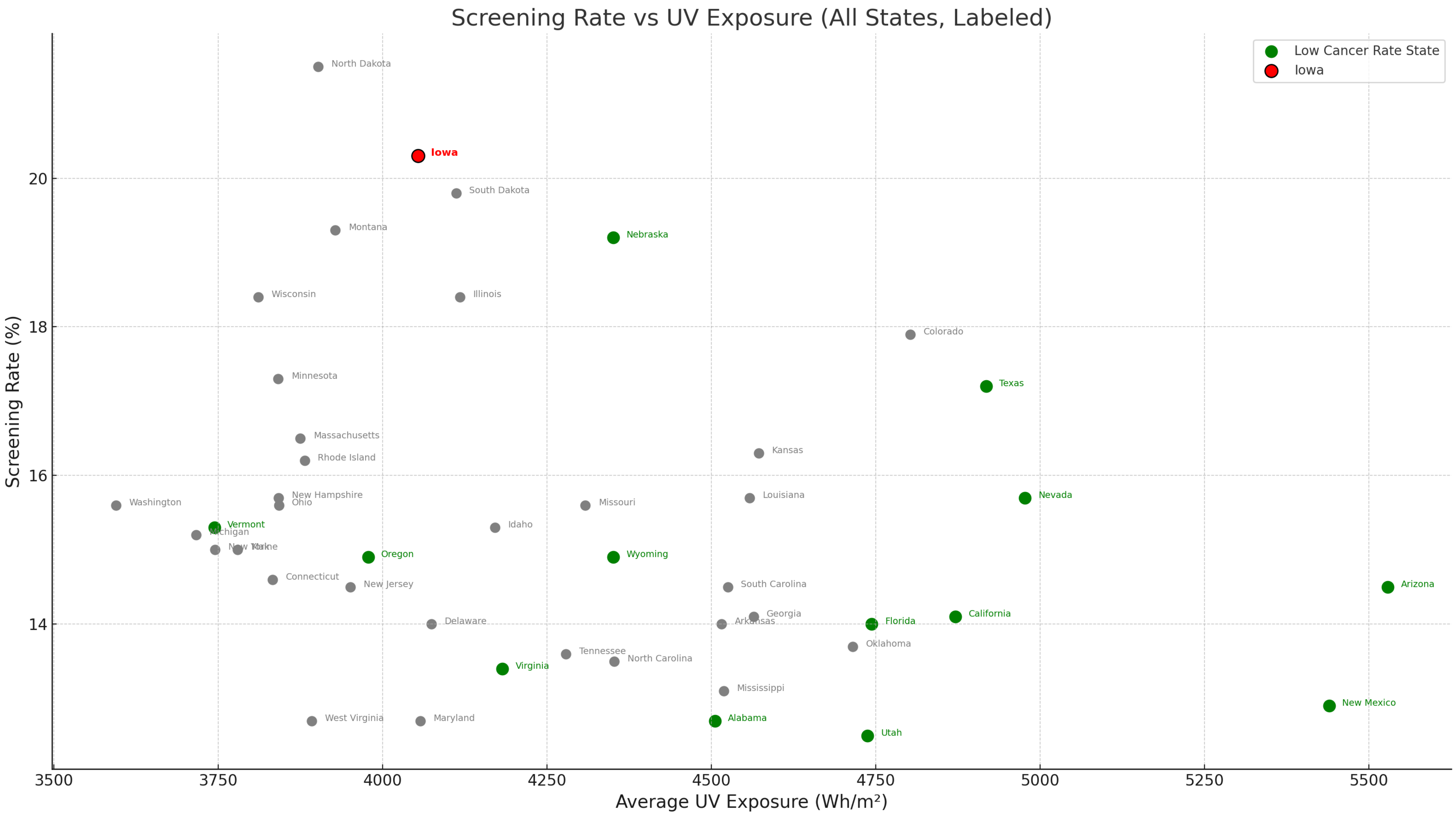
Model Validation

- OLS model validation showed a negative coefficient with UV and cancer rates

Supplemental Research

- “But overexposure to UV radiation can weaken the immune system, reducing the skin’s ability to protect against [cancers and infections]” (EPA, 2025).

Potential Explanation for the Inverse Finding



Radon: A Pervasive Risk Factor

Model Findings

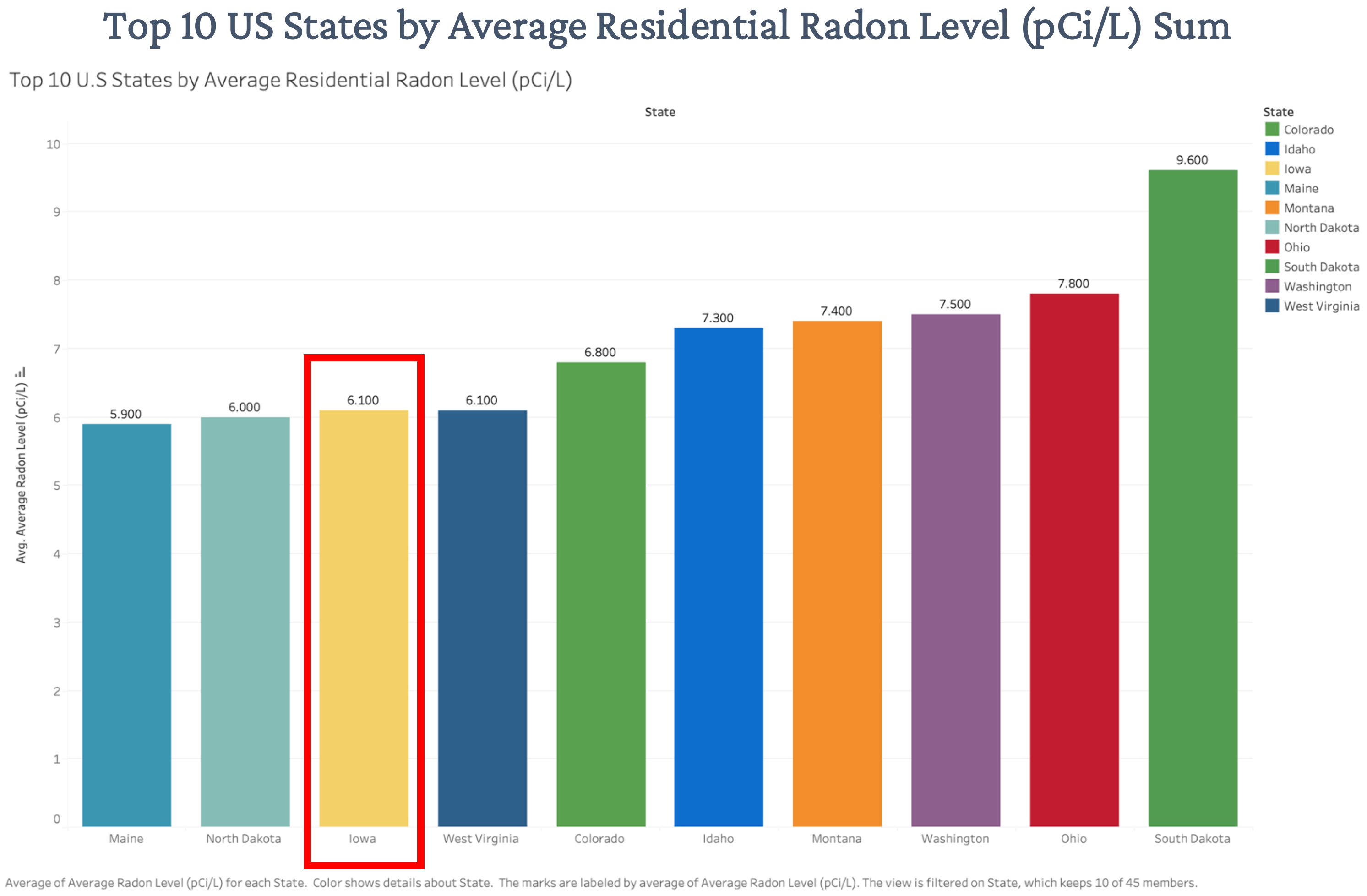
- **P-values:**
  - Borderline insignificant (p = ~ 0.07) for breast, lung, melanoma, and prostate cancers

- However, Iowa's radon levels are among the highest nationally (6.1 pCi/L)

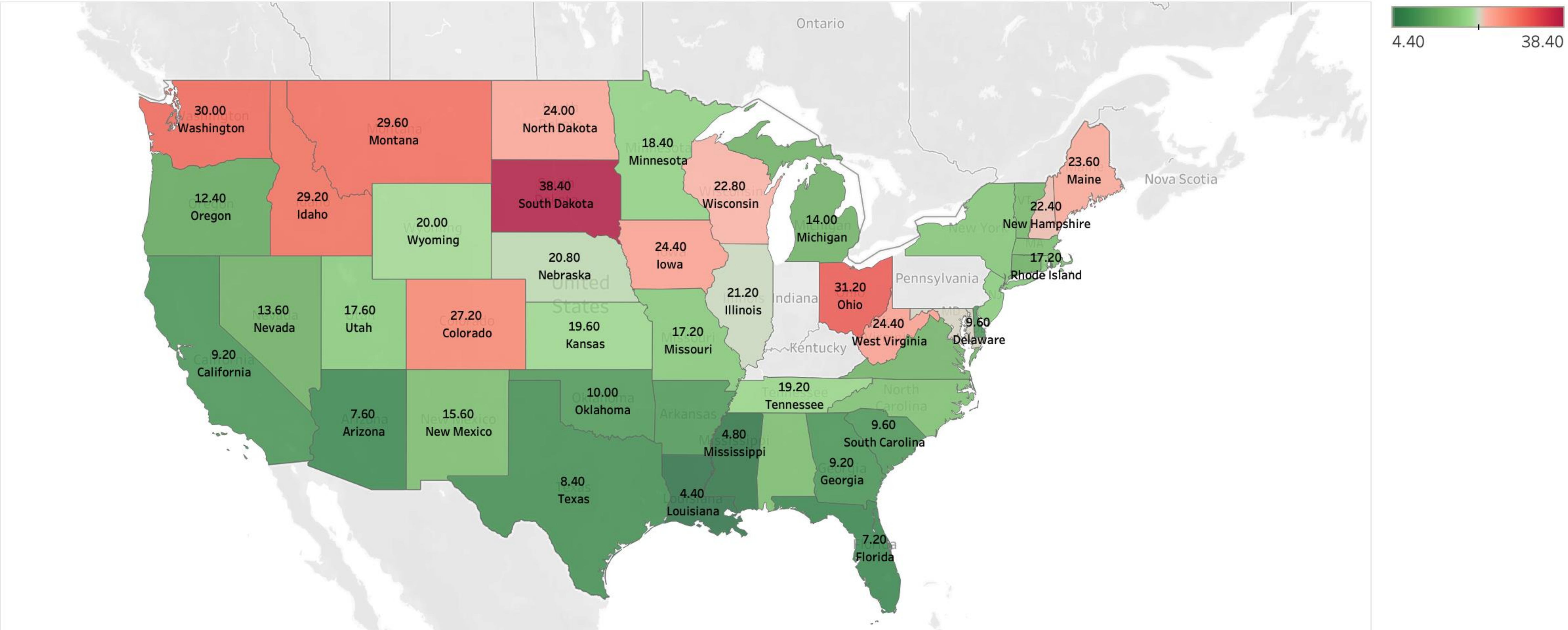
Supplemental Research

- Radon is second leading cause of lung cancer in U.S (EPA, 2024)

- Iowa has the highest average residential radon levels in the country due to its soil composition (EPA, 2024)



Average Radon Levels Heat Map by State



Iowa ranks among the highest states for average radon level sum over 4 years at 24.4 pCi/L, highlighting environmental risk factors linked to elevated incidence rates of lung, breast, prostate, and melanoma cancers

Agricultural Employment: An Occupational Amplifier

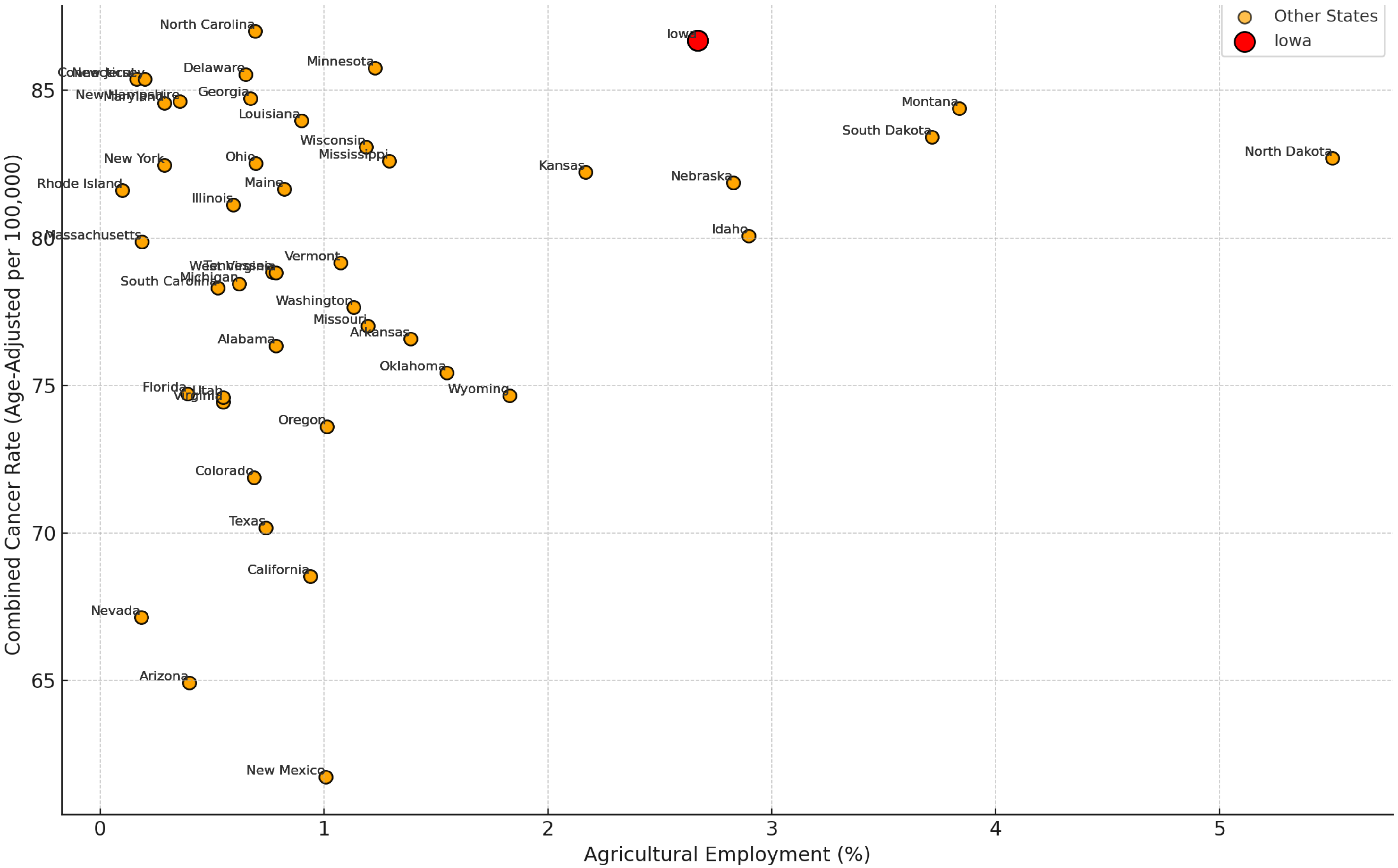
Model Findings

- **P-values:**
  - Statistically insignificant ($p > 0.05$) for breast, lung, melanoma, and prostate cancers
- Agricultural employment rates showed a **positive association** with cancer incidence

Supplemental Research

- **19% of Iowans** are employed in agriculture, totaling 385,332 jobs (Iowa Farm Bureau, 2024)
- **83.9% of Iowa**'s land is dedicated to **farms** (Iowa Census, 2022)

Agricultural Employment vs Combined Cancer Rate by State



Pesticide Usage: Another Nonlinear Link

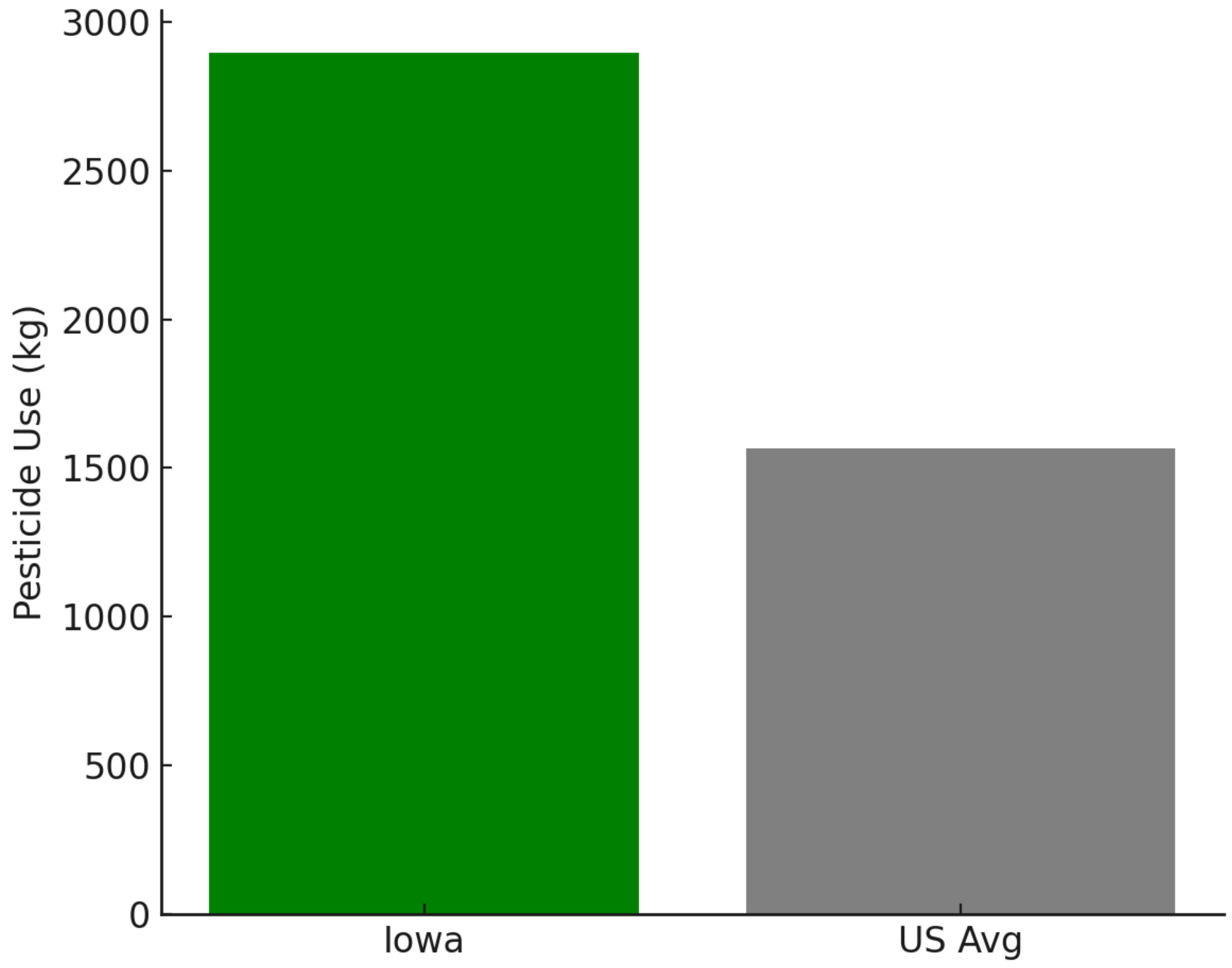
Model Findings

- **P-values:**
  - Statistically insignificant ($p > 0.05$) for breast, lung, melanoma, and prostate cancers
- Pesticide exposure showed a small **positive association** with cancer incidence

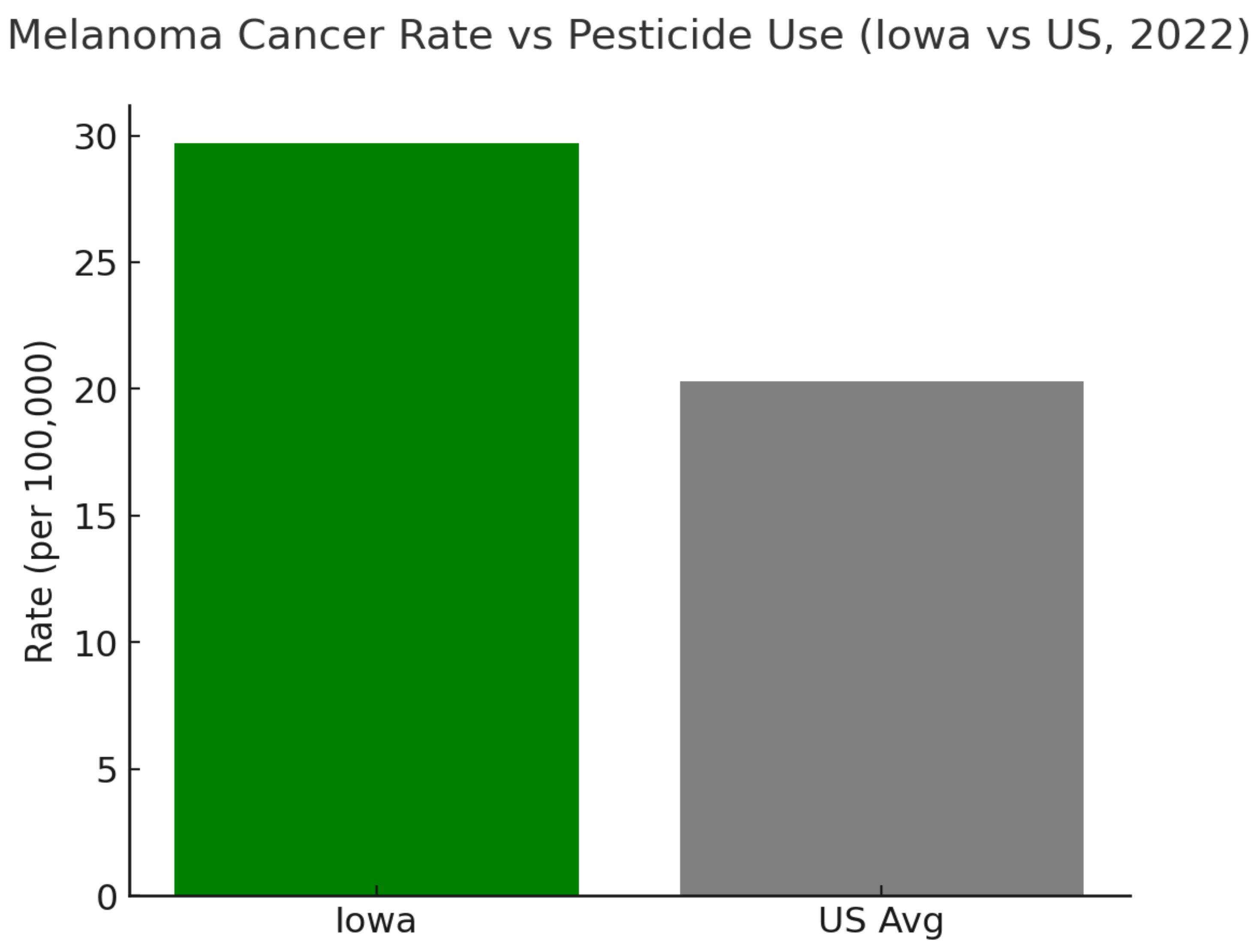
Supplemental Research

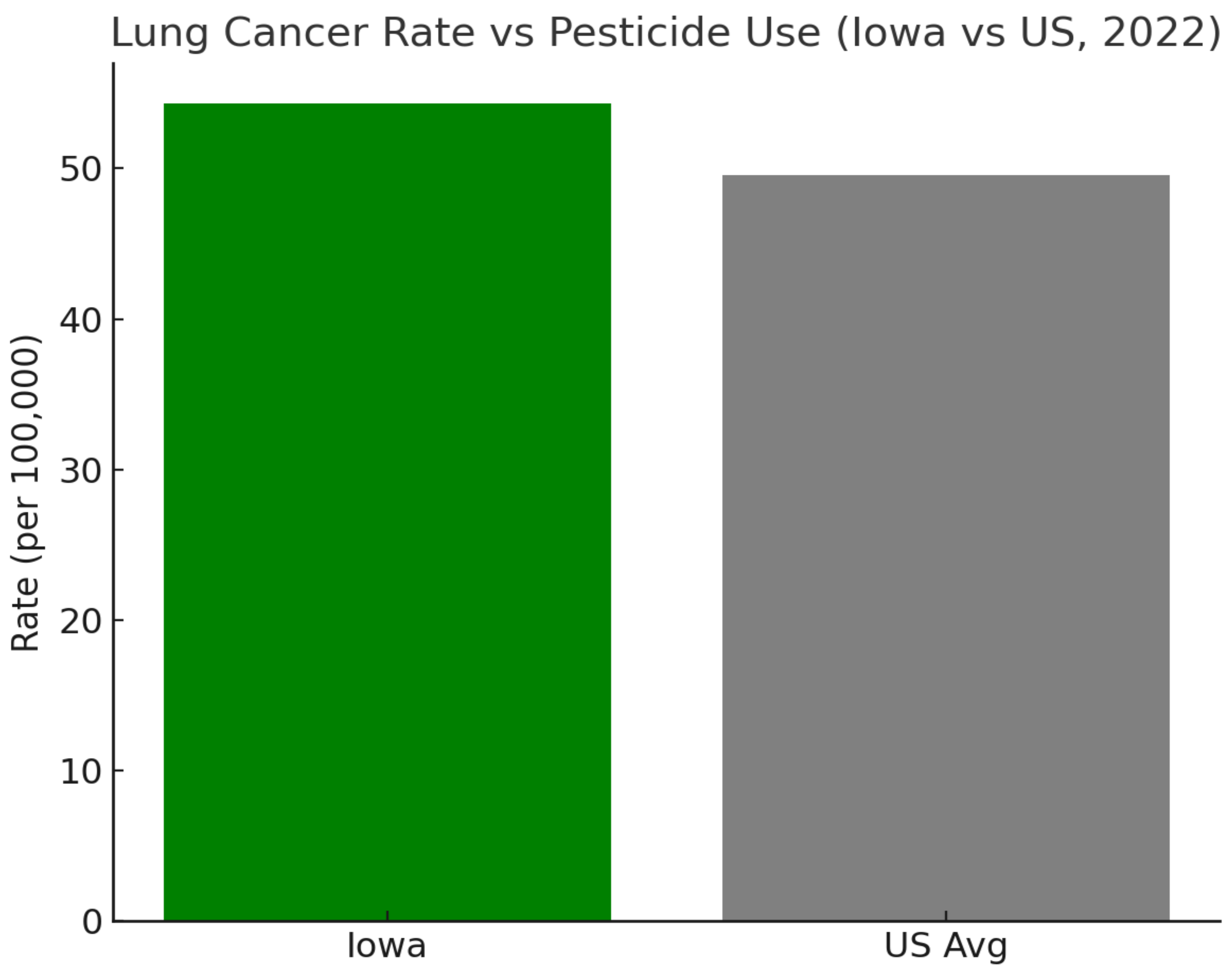
- "Many studies showed positive associations between pesticide exposure and solid tumours… found for brain and prostate cancer" (Bassil et al., 2007).

Pesticide Use (kg) vs Pesticide Use (Iowa vs US, 2022)



Notable Cancers Indirectly Related to Pesticide Exposure





Alcohol Use: A Breast Cancer Risk Factor for Iowa

Model Findings

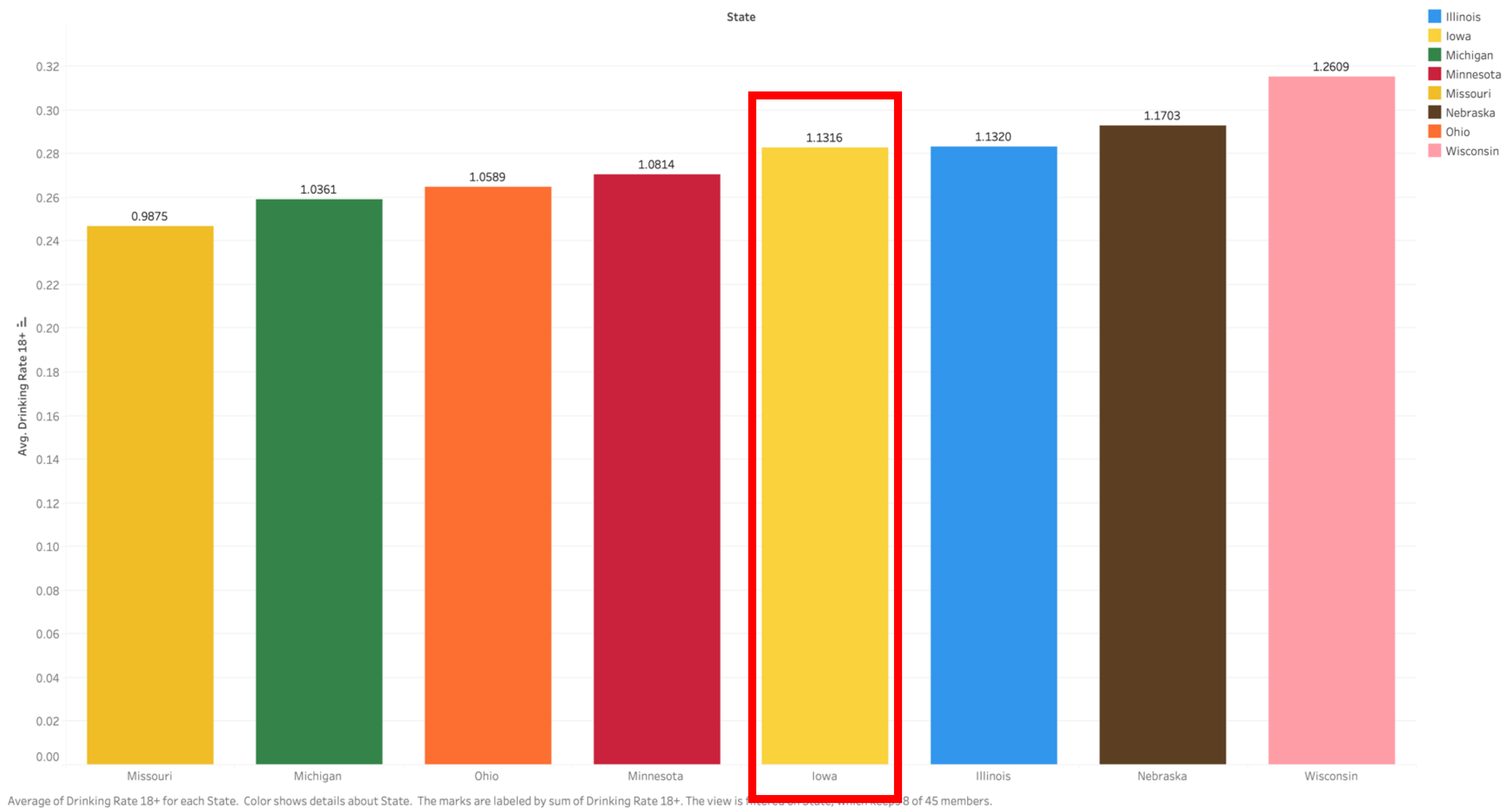
- **P-values:**
  - Breast Cancer: **p = 0.021** (significant)
  - Statistically insignificant (**p > 0.05**) for lung, melanoma, and prostate cancers

  - Iowa has higher than average alcohol use rates compared to the national average

Supplemental Research

- Regular **alcohol use is linked to** a higher risk of developing **breast cancer** (Iowa Cancer Consortium (2022)
- Approximately 1 in 6 breast cancer deaths is attributed to alcohol consumption (CDC, 2022)

Binge Drinking Rates in Midwestern States

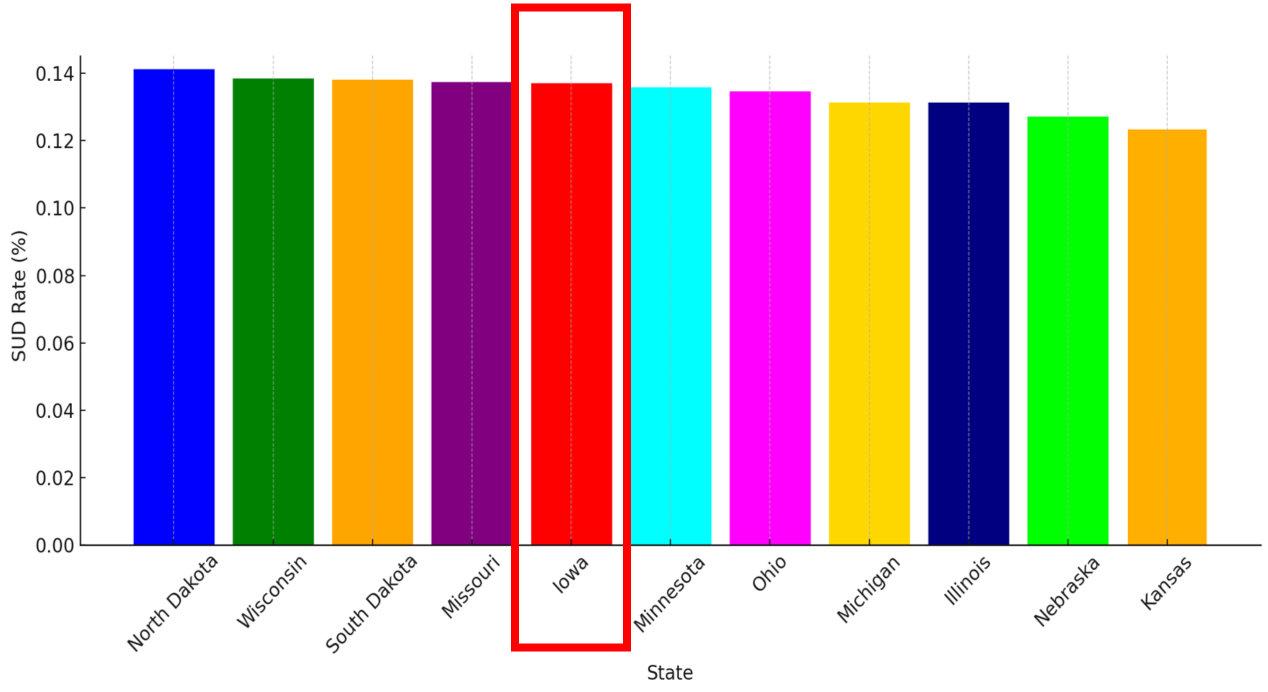


Substance Use Disorder: Breast and Prostate Cancer Links

Model Findings

- **P-values:**
  - Breast Cancer: **p = 0.008** (significant)
  - Prostate Cancer: **p = ~0.01** (significant)
  - Statistically insignificant (**p > 0.05**) for lung and melanoma cancers
  - Iowa is higher than the national average for SUD rate

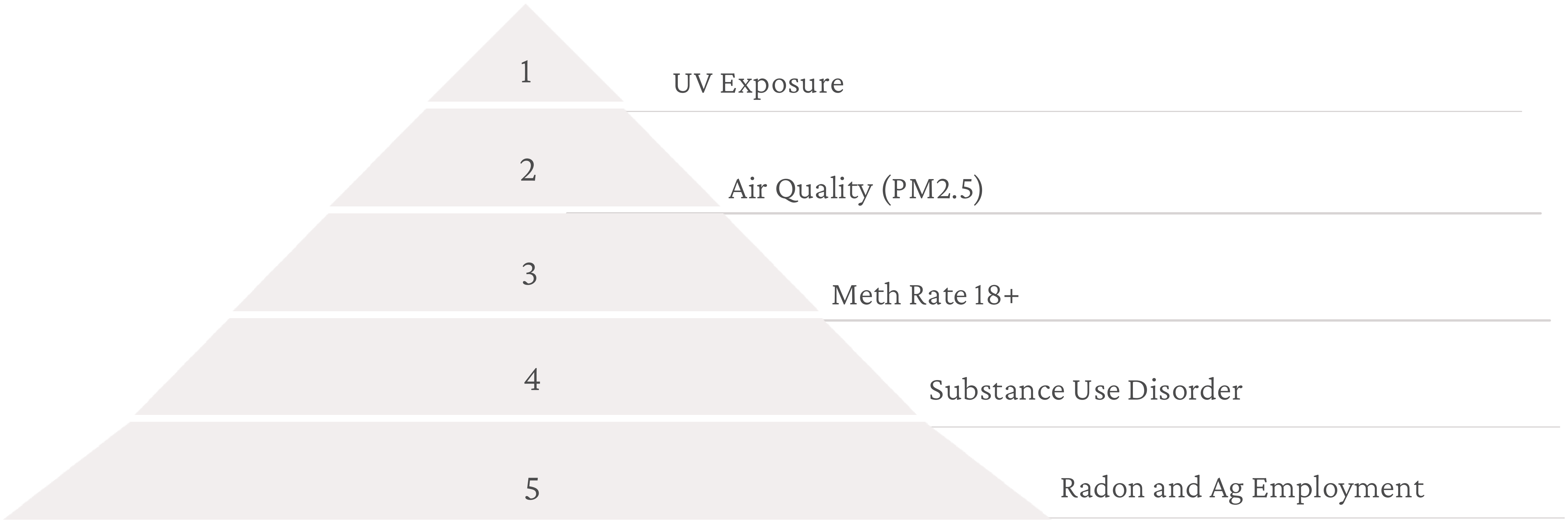
Substance Use Disorder (SUD) Rate in Midwestern States



Supplemental Research

- A Swedish national study found that women with SUD had higher rates of breast cancer incidence, mortality, and were more likely to be diagnosed at a later stage compared to the general population. (National Library of Medicine, 2020)
- A study found that men with SUD had a slightly higher risk of developing prostate cancer and a significantly higher risk of dying from it. (Spring Nature, 2021)

Hierarchy of Compounding Risk Factors



These risk factors are likely not acting in isolation. The combination of UV exposure, particulate matter, substance uses, radon exposure, and agricultural employment creates a unique cancer risk profile in Iowa. Further investigation is needed to understand these interactions.

Lung Cancer Treatment Costs

Scenario	Estimated Cost to insurance
• Early-stage	• $15,000 - $40,000
• Advanced cases (1st year)	• $150,000 - $500,000+
• Long-term advanced cases	• $250,000 - $1,000,000

Melanoma Cancer Treatment Costs

Scenario	Estimated Cost to insurance
Early-stage removal	$1,000 - $5,000
Advanced treatment	$100,000 - $250,000+
One immunotherapy	$150,000/year

Prostate Cancer Treatment Costs

Scenario	Estimated Cost to insurance
Active Surveillance	$3,000 - $10,000/year
Surgery/Radiation (one-time)	$12,000 - $40,000
Advanced treatment	$100,000 - $250,000+

Breast Cancer Treatment Costs

Scenario	Estimated Cost to insurance
Early-stage	$15,000 - $80,000
Advanced/Metastatic	$100,000-$300,000+
Annual targeted therapy	$70,000 - $150,000+/Year

Screening Costs

Cancer Type	Typical Screening Test	Cost per Screening (Billed)	Insurance Payment (Negotiated)
Melanoma (skin cancer)	Full-body skin exam (dermatologist)	~$150–$300	~$75–$200
Lung Cancer	Low-dose CT scan (for high-risk people)	~$300–$500	~$250–$400
Breast Cancer	Mammogram	~$250–$450	~$100–$250
Prostate Cancer	PSA blood test	~$50–$100	~$30–$80

Midwest Farmers Screening Rates

Screening Type	How Often Midwest Farmers Actually Get Screened (on average)
Skin Cancer (Melanoma)	Very low rates — only about 20–30% get regular skin checks. Most don't get checked unless a problem appears.
Lung Cancer	Very low — less than 5–10% of eligible high-risk farmers get annual CT scans.
Breast Cancer (Female farmers)	About 50–65% get mammograms every 1–2 years (lower than urban populations).
Prostate Cancer (Male)	Around 40–60% of farmers over 50 have had a PSA test in the past 2 years — again a little lower than urban men

Recommendations

Validate Findings

Conduct further research to validate model findings. Investigate interactions between risk factors.

Targeted Interventions

Develop targeted interventions to mitigate specific risk factors. Focus on high-risk populations and regions.

Public Awareness

Increase public awareness of cancer risk factors. Promote preventive measures and early detection.

Iowa's cancer rates demand a comprehensive, multi-faceted response. By understanding the key drivers, we can develop effective strategies to protect public health.

Outcome: New Total Cost Equation (Lung Cancer)

Using this equation: New Total Cost = (More Early-Stage × Cost Early) + (Fewer Late-Stage × Cost Late)
(Forecasting can be done for pricing)

Example: Lung Cancer (simple numbers)
Suppose you have 1,000 lung cancer cases:

•**Without screening:**
- 200 early-stage cases × $50,000 = $10M
- 800 late-stage cases × $400,000 = $320M
- **Total = $330M**

•**With better screening:**
- 500 early-stage cases × $50,000 = $25M
- 500 late-stage cases × $400,000 = $200M
- **Total = $225M**

•**Estimated savings:**
$330M − $225M = **$105M saved** just from catching cancer earlier.

Outcome: New Total Cost Equation (Breast Cancer)

Using this equation: New Total Cost = (More Early-Stage × Cost Early) + (Fewer Late-Stage × Cost Late)
(Forecasting can be done for pricing)

Example: Breast Cancer (simple numbers)
Suppose you have 1,000 breast cancer cases:
Without better screening:

- 600 early-stage cases × $40,000 = $24M
- 400 late-stage cases × $200,000 = $80M
- **Total = $104M**

With better screening:

- 800 early-stage cases × $40,000 = $32M
- 200 late-stage cases × $200,000 = $40M
- **Total = $72M**

Estimated savings: $104M − $72M = **$32M saved** just from catching cancer earlier.

.

Outcome: New Total Cost Equation (Melanoma Cancer)

Using this equation: New Total Cost = (More Early-Stage × Cost Early) + (Fewer Late-Stage × Cost Late)
(Forecasting can be done for pricing)

Example: Melanoma (simple numbers)
Suppose you have 1,000 melanoma cases:
Without better screening:

- 850 early-stage cases × $8,000 = $6.8M
- 150 late-stage cases × $150,000 = $22.5M
- **Total = $29.3M**

With better screening:

- 950 early-stage cases × $8,000 = $7.6M
- 50 late-stage cases × $150,000 = $7.5M
- **Total = $15.1M**

Estimated savings: $29.3M − $15.1M = **$14.2M saved** just from catching cancer earlier

Outcome: New Total Cost Equation (Prostate Cancer)

Using this equation: New Total Cost = (More Early-Stage × Cost Early) + (Fewer Late-Stage × Cost Late)
(Forecasting can be done for pricing)

Example: Prostate Cancer (simple numbers)
Suppose you have 1,000 prostate cancer cases:
Without better screening:

- 700 early-stage cases × $30,000 = $21M
- 300 late-stage cases × $120,000 = $36M
- **Total = $57M**

With better screening:

- 850 early-stage cases × $30,000 = $25.5M
- 150 late-stage cases × $120,000 = $18M
- **Total = $43.5M**

Estimated savings: $57M − $43.5M = **$13.5M saved** just from catching cancer earlier

Conclusion To Iowa's Rising Cancer Rates

- Iowa faces one of the highest new cancer incidence rates nationally despite U.S. declines.
- Five compounding risk factors drive Iowa's cancer landscape: UV exposure, Air Quality (PM2.5), Methamphetamine Use, Substance Use Disorder, and Radon/Agricultural Employment.
- Predictive models validated strong statistical associations for key cancers (breast, lung, melanoma, prostate).
- Screening rates among high-risk groups like Midwest farmers remain significantly lower than needed, reducing early detection opportunities.
- Improving screening and addressing environmental and behavioral risk factors could:
  - Save millions annually in insurance costs.
  - Substantially reduce cancer mortality and incidence.

Thank You

Q & A

Iowa Cancer Incidence Contributing Factors Analysis

Group AArete 3-2
Alex, Nick, Patrick, Kellen

1. Objective

We aim to identify and quantify key environmental, behavioral, and socio-economic drivers of elevated cancer incidence rates in Iowa. Specifically, we aim to create predictive models that impact the contributing factors, such as UV exposure, air quality, substance use rates, and occupational exposures across the four main cancers in Iowa: breast, lung, melanoma, and prostate. This objective will be fulfilled by gathering publicly available state and federal data over a 5-year span.

2. Assumptions and Challenges

Assumptions for Models

- Cancer incidence rates from 2016–2020 are assumed to accurately reflect environmental and behavioral exposures from previous years.
- Behavioral, environmental, and socio-demographic factors collected at the state level represent true population-level exposures.
- Linear relationships are assumed between predictors and cancer rates for Ordinary Least Squares (OLS) modeling.

Challenges

- Some datasets (e.g., radon, pesticide usage) had limited temporal granularity or coverage.
- Certain risk factors (agricultural employment, radon) showed positive correlations but did not reach statistical significance.
- Modeling multiple cancer types simultaneously posed risks of multicollinearity among predictors.

3. Methodology

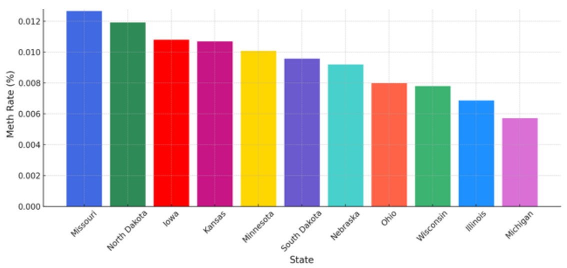
R-studio is chosen to do data cleaning and selection of 18 unique datasets from the CDC, NCI, SAMHSA, KFF, and U.S. Census Bureau sources into a unified, state-level panel. Orange software is used to produce the predictive models, and Python is used to validate the Linear Regression models' independent variables. We modeled cancer incidence rates using: Ordinary Least Squares (OLS) regression (Python statsmodels) and Linear Regression and Gradient Boosting models (Orange scikit-learn). These models were also evaluated along R^2, RMSE, and cross-validation.

4. Data Description

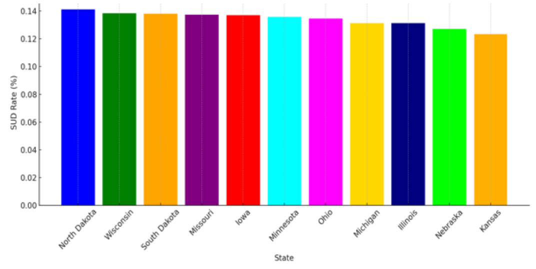
Our final dataset consisted of 180 state-level observations across 35 variables, with no missing data. The target variables included age-adjusted incidence rates for breast, lung, melanoma, and prostate cancers, each measured per 100,000 population. Predictor variables encompassed 8 behavioral risk factors (such as methamphetamine use, substance use disorder rates, tobacco use, and drinking rates), 4 environmental exposure factors (including UV exposure, PM2.5 air pollution, pesticide use, and radon levels), and 3 socio-demographic metrics (average age, obesity rate, and agricultural employment percentage).

5. Model Deliverables

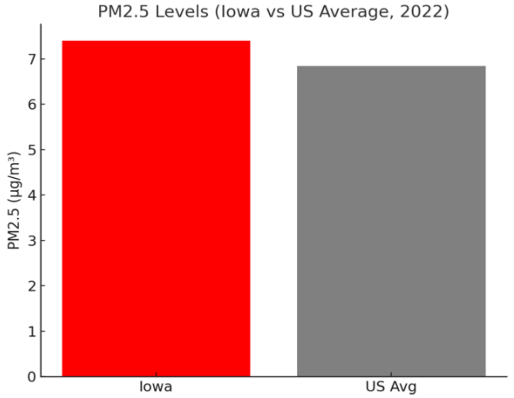
Methamphetamine Use Rate in Midwestern States



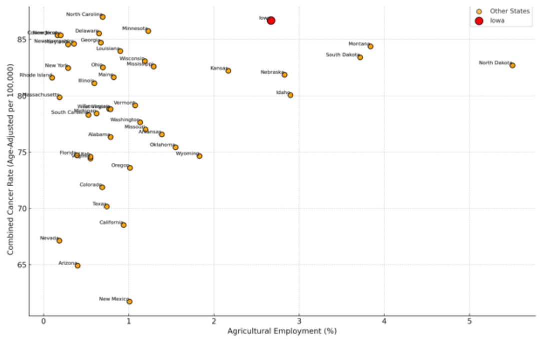
Substance Use Disorder (SUD) Rate in Midwestern States



Higher PM2.5 Levels in Iowa May Contribute to Elevated Cancer Rates



Agricultural Employment vs Combined Cancer Rate by State

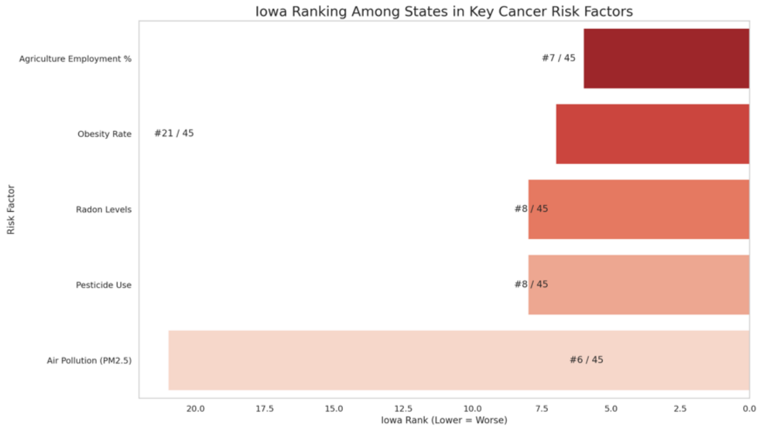


6. Business Recommendation

Based on our model results, Iowa should prioritize reducing environmental exposure risks by targeting air quality improvements, specifically lowering PM2.5 pollution levels. Public health campaigns should also be enhanced to address substance misuse, particularly methamphetamine and alcohol consumption, while increasing awareness of UV exposure risks and strengthening early skin cancer screening initiatives. Although radon levels and agricultural exposures were not statistically significant predictors, ongoing monitoring is recommended given their biological plausibility. Targeting UV exposure, PM2.5 pollution, and substance use rates could significantly reduce cancer incidence across multiple cancer types, leading to broader public health benefits.

Additional Information:

- Environmental exposures such as UV radiation and fine particulate matter (PM2.5) are established carcinogens that can damage DNA and weaken immune defenses.
- Behavioral risk factors, including substance use disorders and alcohol consumption, contribute to chronic inflammation and cellular damage, elevating cancer risk.
- Radon exposure and agricultural occupational factors, while not statistically significant in our models, are supported by biological research as potential cancer risks.
- Future public health efforts should address the compounding effects of environmental, behavioral, and socio-demographic factors on cancer incidence rates.



References

Bassil, K. L., Vakil, C., Sanborn, M., Cole, D. C., Kaur, J. S., & Kerr, K. J. (2007). Cancer health effects of pesticides: Systematic review. *Canadian Family Physician, 53*(10), 1704–1711. https://pmc.ncbi.nlm.nih.gov/articles/PMC2231435/

Beckland, B. (n.d.). UV radiation suppresses experimental autoimmune encephalomyelitis independent of vitamin D production. *Proceedings of the National Academy of Sciences (PNAS)*. https://www.pnas.org/doi/10.1073/pnas.2221311120

Centers for Disease Control and Prevention. (2024, February 15). *Alcohol and cancer*. U.S. Department of Health and Human Services. https://www.cdc.gov/cancer/risk-factors/alcohol.html

González Maglio, D. H., Paz, M. L., & Leoni, J. (2016). Sunlight effects on immune system: Is there something else in addition to UV-induced immunosuppression? *BioMed Research International*. https://pmc.ncbi.nlm.nih.gov/articles/PMC5187459/

Iowa Farm Bureau. (2025, April 14). *Top 10 farm and agriculture facts – Iowa*. https://www.iowafarmbureau.com/Article/Top-10-farm-and-agriculture-facts-Iowa

IASTATE. (n.d.). *Estimated costs of crop production in Iowa – 2024*. Iowa State University Extension and Outreach. https://www.extension.iastate.edu/agdm/crops/pdf/a1-21.pdf

Massey, R. (2022, May 1). *Days suitable for fieldwork in Missouri*. University of Missouri Extension. https://extension.missouri.edu/publications/g362

Neale, R. E., Lucas, R. M., Byrne, S. N., Hollestein, L., Rhodes, L. E., Yazar, S., Young, A. R., Berwick, M., Ireland, R. A., & Olsen, C. M. (2023, May). The effects of exposure to solar radiation on human health. *Photochemical & Photobiological Sciences: Official Journal of the European Photochemistry Association and the European Society for Photobiology*. https://pmc.ncbi.nlm.nih.gov/articles/PMC9976694/

Rohlfing, N. (2024, October 9). Here's how harvest stacks up among the top three corn growing states. *Successful Farming*. https://www.agriculture.com/here-s-how-harvest-stacks-up-among-the-top-three-corn-growing-states-8725973

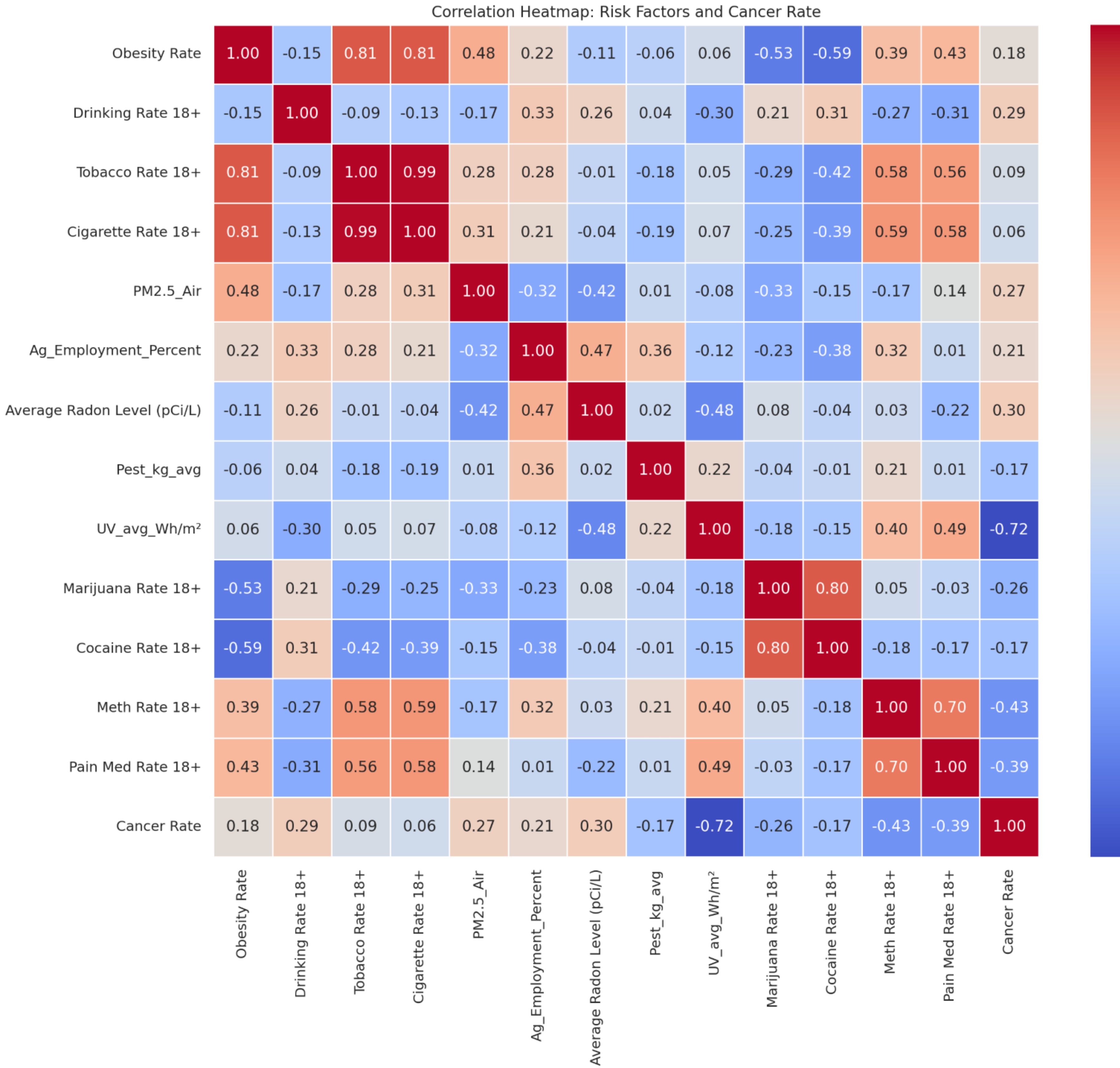
Substance Abuse and Mental Health Services Administration. (2025, February 13). *2022 National Survey on Drug Use and Health (NSDUH) releases*. U.S. Department of Health and Human Services. https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/national-releases/2022

U.S. Environmental Protection Agency. (2025, February 4). *Health effects of UV radiation*. https://www.epa.gov/sunsafety/health-effects-uv-radiation

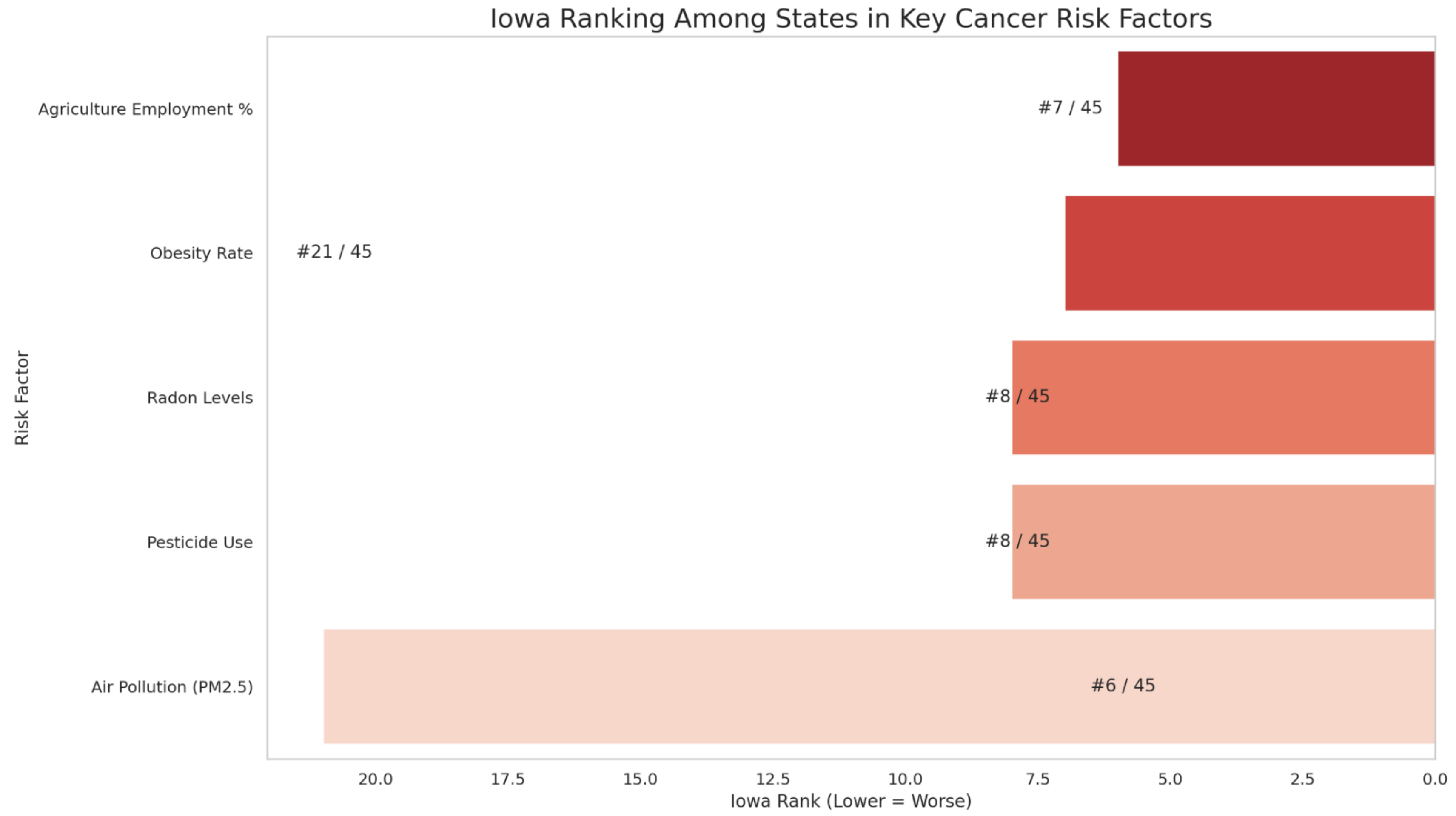
Appendix

Correlation Matrix

- **UV exposure** shows a strong **negative correlation** with cancer rate (r = -0.72), suggesting potential lower cancer incidence in states with more UV radiation.
  - Top 3 highest UV avg states' cancer incidence rates: Arizona (47th), New Mexico (50th), and Nevada (49th)
- **Obesity rate** and **air pollution (PM2.5)** also show noticeable positive correlations.
- Some drug use rates (like **meth** and **cocaine**) have weaker or even slightly negative correlations, meaning they might not be major drivers here.



Iowa Compared To Other States



Risk Factors For Iowa Compared To Other States