HAPs, Micromaps and GPL - Visualization of Geographically Referenced Statistical Summaries on the World Wide Web

J. Symanzik*, D. A. Axelrad, D. B. Carr, J. Wang, D. Wong, T. J. Woodruff

* Jürgen Symanzik
George Mason University
Center for Computational Statistics 4A7
Fairfax, VA 22030
e-mail: symanzik@galaxy.gmu.edu

Abstract

In this paper we report on an approach to visualize the U.S. Environmental Protection Agency (EPA) hazardous air pollutant (HAP) data on the World Wide Web (WWW). Long-term cumulative concentrations for 148 HAPs have been estimated as a part of EPA's Cumulative Exposure Project (URL:CumulativeExposureProject) for each of the 60,803 census tracts in the 48 contiguous states. Confidence bounds are available as well for this data set.

Our WWW approach is based on the Graphics Production Library (GPL) (Carr, Valliant, and Rope, 1996) and extends it with micromaps (Carr and Pierson, 1996; Carr, Olsen, Courbois, Pierson, and Carr, 1998). The GPL is a JAVA-based application, developed and maintained within the Bureau of Labor Statistics (BLS), that allows to display statistical data as row-labeled plots and time series plots on the WWW. This data visualization approach closely follows recent recommendations on statistical graphics by Carr (1994) and Cleveland (1993, 1994). Micromaps are a series of small generalized maps that highlight the geographical region associated with statistical features in accompanying plots.

When added to the GPL, micromaps serve two purposes. In addition to providing the geographical component of a linked statistical summary, they serve as a navigation tool for drilling down through a hierarchy of maps. A WWW user of our application can click on a state and advance to the underlying county-within-state micromap display. Selection of a county leads to information at the census tract level. This hierarchy of clickable maps and statistical displays allows the WWW user to look at estimates and distributional summaries at the desired geographic resolution. Atypical values in the statistical summaries may lead the investigator to points of interest at a higher spatial resolution level.

Additional interaction through the Web browser is possible. The user can easily toggle between individual HAPs, look at formatted tabular data, and display and download the raw unformatted data.

1. Introduction

This paper reports on an approach to access and visualize the U.S. Environmental Protection Agency (EPA) hazardous air pollutant (HAP - also called air toxics) data on the World Wide Web (WWW). The Cumulative Exposure Project (URL:Cumulative ExposureProject) Web page (Figure 1) has been designed and implemented to provide insight into the scope and the underlying modeling process of the project. A major goal of the Web page is to provide access to the modeled 1990 HAP data at different spatial resolutions.

The Cumulative Exposure Project Web site has scheduled updates. A first version of the Cumulative Exposure Project Web page went online in March 1998 to provide background information on the project. An updated version that provides access to the modeled 1990 HAP data at different spatial resolutions is scheduled for release in December 1998. Additional graphical components will be added to the page in January 1999.

In the Cumulative Exposure Project, EPA has developed modeled concentration estimates of 148 air toxics for every census tract in the continental United States. Air toxics are pollutants known or suspected to cause cancer and other serious human health effects. The modeled concentrations are annual averages for the year 1990, and uncertainty bounds have been developed for each estimate. Development of the modeled concentrations for the 60,803 census tracts in the continental U.S. is described in Rosenbaum, Axelrad, Woodruff, Wei, Ligocki, and Cohen (1999), and analysis of the modeled concentrations is presented in Woodruff, Axelrad, Caldwell, Morello-Frosch, and Rosenbaum (1998) and Caldwell, Woodruff, Morello-Frosch, and Axelrad (1998).

The purpose of the web page is to provide easy, fast, and understandable access to the model results. Since the results are numerous, involving 148 x 60,803 x 3 numeric values, we have taken a hierarchical spatial approach in their display. The WWW user starts at the top (US) level with data at the state level being displayed. Selecting a state causes its county level estimates to be shown. Selecting a county causes its underlying census tract level to be displayed. It is possible to toggle between individual HAPs at any stage. Mouse-clickable maps, menus, and geographical names in the Web document allow the user to easily move up and down in this hierarchy of maps and displays. Thus, the user can look at data and summary statistics at any desired spatial resolution.

In Section 2 of this paper we introduce micromaps and the Graphics Production Library (GPL) and explain their joint use in the Cumulative Exposure Project Web page. Section 3 focuses on the tabular displays of the Web page. Section 4 concludes with a discussion of achievements so far and mentions possible extensions. Additional details on the Cumulative Exposure Project and the related Web page can be found in Symanzik, Wong, Wang, Carr, Woodruff, and Axelrad (1999).

2. Linked Micromap Plots and the GPL

Linked micromap (LM) plots, often simply called micromaps, provide a new way of viewing spatially indexed estimates and summaries statistics. Carr and Pierson (1996) and Carr, Olsen, Courbois, Pierson, and Carr (1998) provide the basic descriptions, several examples, and cite the connections of LM plots to other statistical graphics. A LM plot consists of parallel sequences of micromap panels, label panels, and statistical summary panels. The micromap panels are typically map generalizations or caricatures. The caricatures preserve region neighbors and enlarge very small regions so their color is visible. The label panels provide regions names. The statistical summary panels represent estimates, confidence bounds, and related information using the most perceptually accurate representation, i.e., the position along a scale. Individual panels take familiar forms such as dot plots, bar plots, and box plots. The representation typically uses the encoding that has the highest perceptual accuracy of extraction, i.e., the position along a scale extraction (Cleveland and McGill, 1984). The sequence of panels results from sorting and logical or perceptual grouping. Color and position link corresponding elements within the parallel sequences.

The design of LM plots contrasts with the design of classed choropleth maps. In LM plots, the maps are caricatures but sufficiently convey the location or regions associated with the statistical summary. The statistical summaries have high perceptual accuracy of extraction. Classed choropleth maps use the best representation, i.e., the position along a scale, and most of the space to represent political boundaries. The classed choropleth map design typically discards confidence bounds on statistical estimates, degrades the estimates into a few classes, and represents the order classes with a poor encoding, usually color. Thus the LM plots representation places more emphasis on the estimates and their quality. Knowing the census tract, the county, or state name and general location is usually good enough for an interpretation of the statistical data.

The Graphics Production Library (GPL) is a set of JAVA class libraries for interactive statistical graphics (Carr, Valliant, and Rope, 1996). The GPL was initially intended to add interactivity, such as drag and drop comparisons, panel reordering and rescaling; and pan and zoom to the row-labeled plots of Carr (1994). It also follows recent recommendations on statistical graphics as given in Cleveland (1993, 1994). The design of the GPL also addressed the display of times series and provided for incorporation of metadata, such as warning flags on the time series and links to articles on the time series adjustments. The library was developed to facilitate web distribution of statistical summaries from the Bureau of Labor Statistics (BLS). The GPL, as maintained by the BLS, provided a reasonable starting point for incorporating the micromap capabilities and producing LM plots on the WWW.

Even though LM plots will not be fully available on the Cumulative Exposure Web Page before January 1999, we can provide a first idea how these LM plots will look like. Figure 2 shows a micromap at the top (US) level with states sorted in alphabetical order. The total modeled 1990 HAP concentration of all 148 HAPs, aggregated over the census tracts in each state, is displayed. In our final version on the Web, this display will allow the user to Mouse-click on a state and move to next underlying level. As an example, Figure 3 shows a micromap for Michigan. Again, the total modeled 1990 concentration of all 148 HAPs, aggregated over the census tracts in each county, is displayed. However, although we used the total modeled concentration as a summary for all HAPs in these two figures, this information will not be provided on the Web page. Obviously, the total is quite meaningless with respect to the toxicity of a particular HAP. A relatively small quantity of one air toxic may be far more poisonous than a much larger quantity of another air toxic. The total might easily lead to wrong conclusions when publicly being displayed on the Cumulative Exposure Project Web page.

3. Tabular Displays

Similar to the graphical display based on LM plots and the GPL, we can access our data through formatted data tables. These tables are available at the same spatial resolutions as the LM plots. We start at the top (US) level where data is being displayed for one selected HAP, here benzene, for all 48 states and the District of Columbia (Figure 4). Summary statistics with respect to the selected HAP are being displayed.

We can now select a new state either by changing the menu option for States on top of the table or by Mouse-clicking on the desired state. In our case, we decided to look at data for Rhode Island. The resulting table is displayed in Figure 5. Obviously, the selection menu on top reflects this state selection, allowing us to select one of Rhode Island's five states. The same summary statistics as in the previous figure with respect to the selected HAP are being displayed, however this time at the state level.

Finally, we can now select a single county either by changing the menu option for Counties on top of the table or by Mouse-clicking on the desired county. In our case, we decided to look at data for Rhode Island's Bristol County. The resulting table is displayed in Figure 6. Obviously, the selection menu on top reflects this state/county selection. Now the modeled 1990 concentration of benzene and a corresponding 90% confidence interval are being displayed in the table.

At any time, the user can toggle to a different HAP, a different representation (micromaps or raw data), or move up and down in the spatial hierarchy.

The final display type, the raw data representation, gives an unrestricted insight into the raw data as it comes out of the modeling process. In particular, no rounding takes place. This representation can best be used to download data for one or multiple regions and perform additional calculations or display the data in a different format. The data is comma-delimited and can be transferred by copy/paste from the Web browser window into any application window. At the next stage of this Web page, it is planned to provide access to entire data files at a higher spatial resolution to ease further usage of the data.

4. Discussion and Outlook

When finalized in January 1999, the Cumulative Exposure Project Web page will provide an easy-to-use and up-to-date interface for access and display of a large environmental data set at different spatial resolutions. The combination of LM plot design and the GPL provides the basis of a hierarchical clickable approach to the display of spatially-indexed estimates and summaries. The hierarchical spatial approach is intuitive and has been easily understood by WWW users with little statistical experience. The approach provides a visual query language that can provide access to summary tables and raw data. Future work will add more user preferences and allow for an additional analysis of the data.

This work promotes the display of confidence (or uncertainty) bounds. Such bounds are regularly reported with professional political polls. The presence of such bounds suggests the use of appropriate statistical methodology and the absence of bounds raises the question, do the reported estimates have any statistical validity? The modeling of HAPs results in wide uncertainty bounds. The Web page provides background on these modeled bounds. While methods for calculating bounds are subject to debates and refinements, publishing wide bounds reduces the chances the reader will be badly lead astray by poor estimates.

There are several possibilities for additions to this Web page after the completion of the micromaps display in January 1999. Currently, it is not possible to access the modeled 1990 concentration for all 148 HAPs for a particular location (either state, county, or census tract) at the same time. Instead, one has to toggle through 148 different Web pages to learn about all the HAPs at one location. One goal is to make this data available at one glance, which requires to display data of very different dimensionalities.

Also, currently there is no easily accessible site on the Web that allows the mapping of a particular ZIP code to a census tract number which are used throughout the Cumulative Exposure Project. We plan to add such a mechanism to our Web page.

At this stage, only the modeling of the 1990 air toxics is finished. Once available, we plan to dissiminate the modeled data for exposure levels for chemical contaminants found in public and private drinking water supplies and exposures to contaminants in foods in a similar way through this Web page.

Acknowledgments

EPA funded the majority of the work behind this paper under contract No. 098272 and cooperative agreement No. CR825564-01-0. Additional federal agencies, BLS and NCHS, supported some facets of this work. The article has not been subject to review by any of these agencies so does not necessarily reflect the view of the agencies, and no official endorsement should be inferred. The conclusions and opinions are solely those of the authors and are not necessarily the views of the agencies.

References

Caldwell, J. C., Woodruff, T. J., Morello-Frosch, R., and Axelrad, D. A. (1998), "Application of Health Information to Hazardous Air Pollutants Modeled in EPA's Cumulative Exposure Project", *Toxicology and Industrial Health*, Vol. 14, No. 3, pp. 429-454.

Carr, D. B. (1994), "Converting Tables to Plots", Technical Report 101, Center for Computational Statistics, George Mason University, Fairfax, VA.

Carr, D. B., Olsen, A. R., Courbois, J. P., Pierson, S. M., and Carr, D. A. (1998), "Linked Micromap Plots: Named and Described", *Statistical Computing and Statistical Graphics Newsletter*, Vol. 9, No. 1, pp. 24-32.

Carr, D. B., and Pierson, S. M. (1996), "Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps", *Statistical Computing and Statistical Graphics Newsletter*, Vol. 7, No. 3, pp. 16-23.

Carr, D. B., Valliant, R., and Rope, D. (1996), "Plot Interpretation and Information Webs: A Time-Series Example from the Bureau of Labor Statistics", *Statistical Computing and Statistical Graphics Newsletter*, Vol. 7, No. 2, pp. 19-26.

Cleveland, W. S., and McGill, R. (1984), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods", *Journal of the American Statistical Association*, Vol. 79, pp. 531-554.

Cleveland, W. S. (1993), *Visualizing Data*, Hobart Press, Summit, NJ.

Cleveland, W. S. (1994), *The Elements of Graphing Data*, Hobart Press, Summit, NJ.

Rosenbaum, A. S., Axelrad, D. A., Woodruff, T. J., Wei, Y.-H., Ligocki, M. P., and Cohen, J. P. (1999), "National Estimates of Outdoor Air Toxics Concentrations", *Journal of the Air & Waste Management Association*, (In Press).

Symanzik, J., Wong, D., Wang, J., Carr, D. B., Woodruff, T., and Axelrad, D. (1999), "WWW-based Access and Visualization of Hazardous Air Pollutants", *Proceedings GIS in Public Health Conference, August 17-20, 1998, San Diego, California* (In Preparation).

Woodruff, T. J., Axelrad, D. A., Caldwell, J. C., Morello-Frosch, R., and Rosenbaum, A. S. (1998), "Public Health Implications of 1990 Air Toxics Concentrations Across the United States", *Environmental Health Perspectives*, Vol. 106, No. 5, pp. 245-251.

URL:CumulativeExposureProject http://www.epa.gov/CumulativeExposure/

Figures

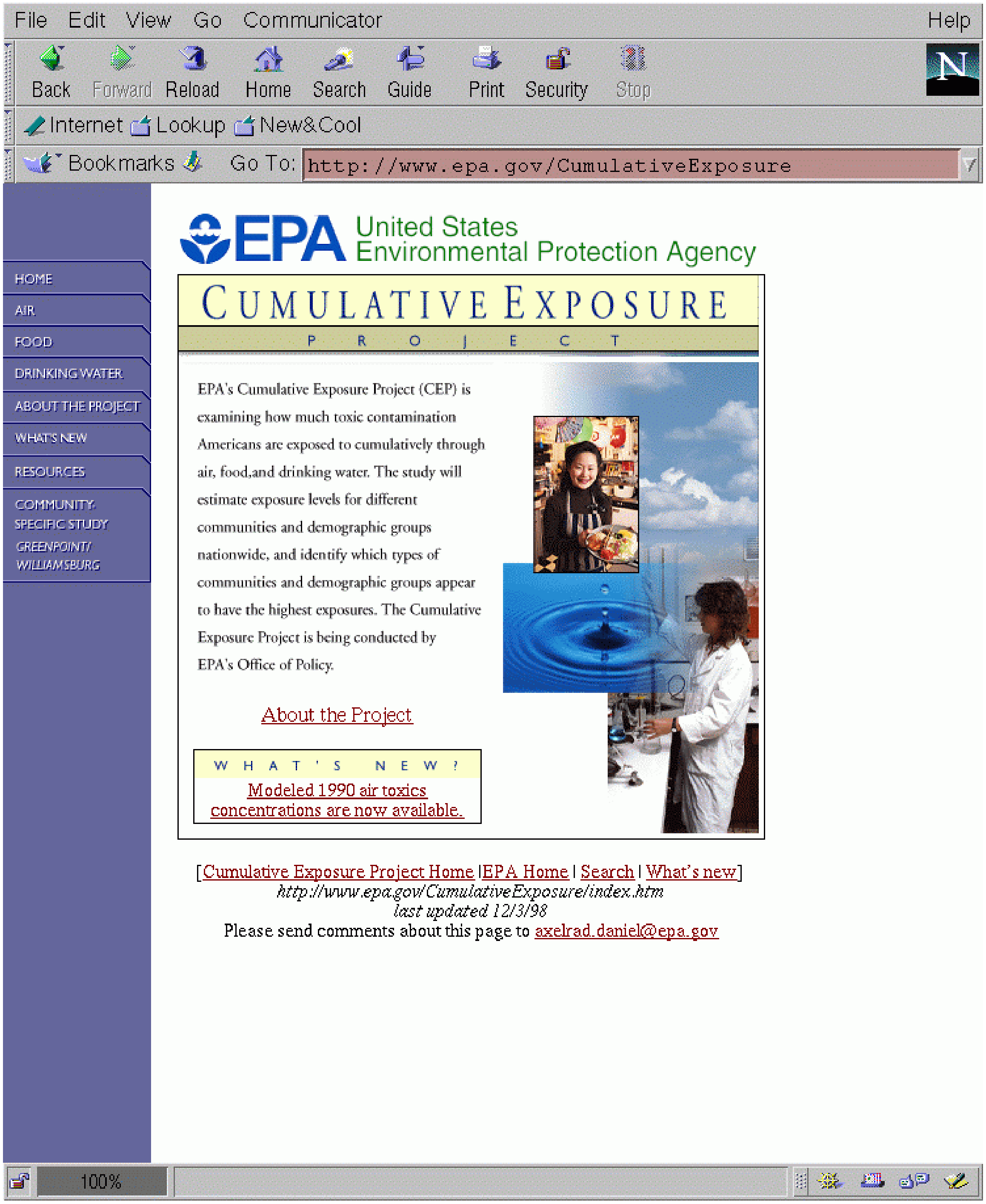


Figure 1: The starting page of the Cumulative Exposure Project Web page, accessible at http://www.epa.gov/CumulativeExposure/.

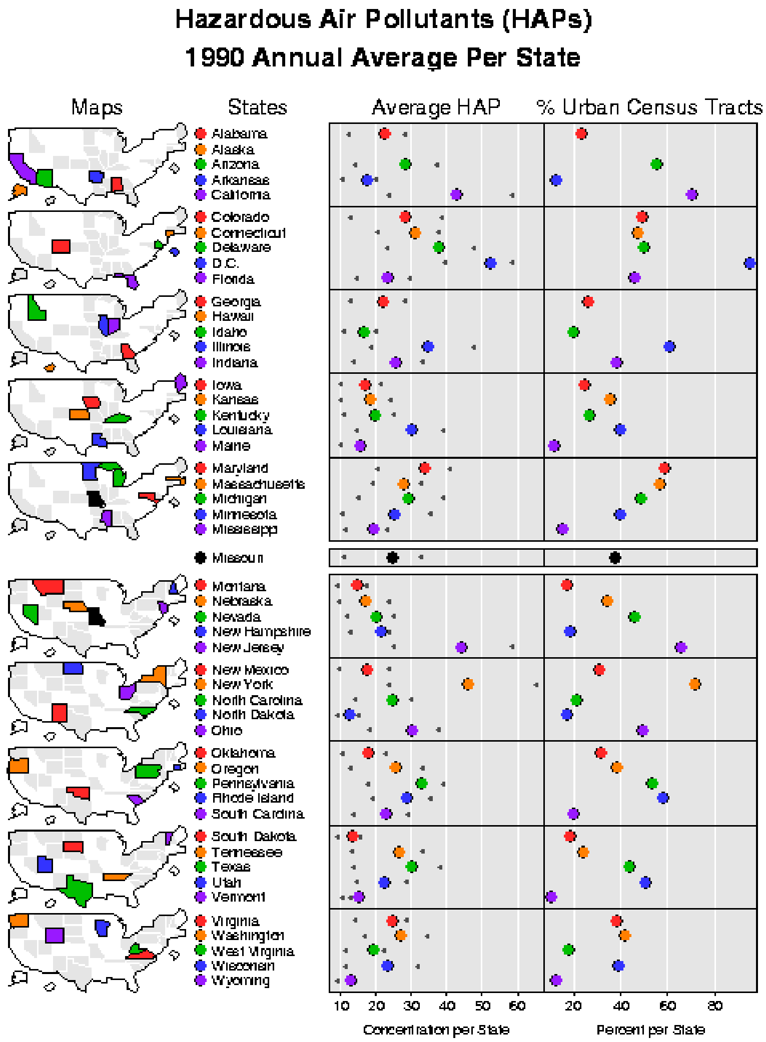


Figure 2: Micromap display at the top (US) level for all 50 states and the District of Columbia. The total modeled 1990 HAP concentration in micrograms per cubic meter for all 148 HAPs has been averaged over the number of census tracts in each state. The 25th and 75th percentiles have been displayed as well. For some states, e.g., Vermont, the distribution of HAP concentrations is extremely skewed which becomes obvious when the displayed arithmetic mean falls above (or below) the 75th (or 25th) percentile. The accompanying second data column (which will not be available on the Web page) indicates the percentage of urban census tracts. A strong correlation between the two displayed variables can be visually detected.

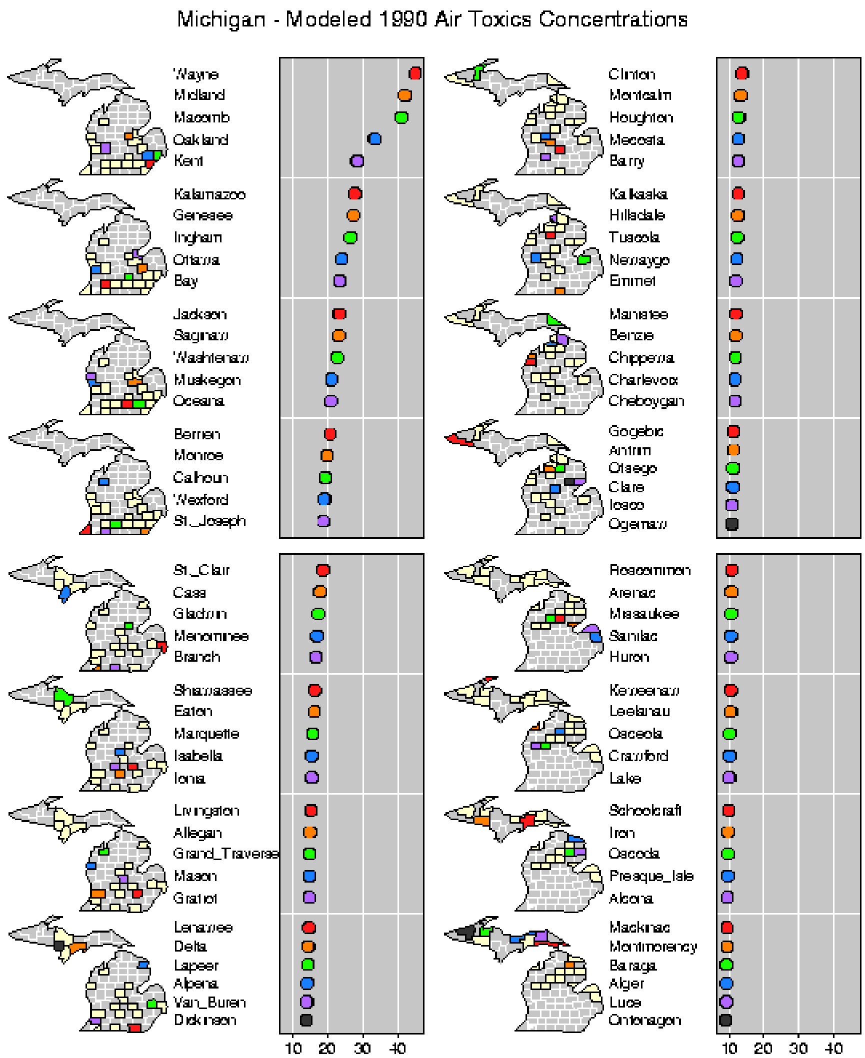
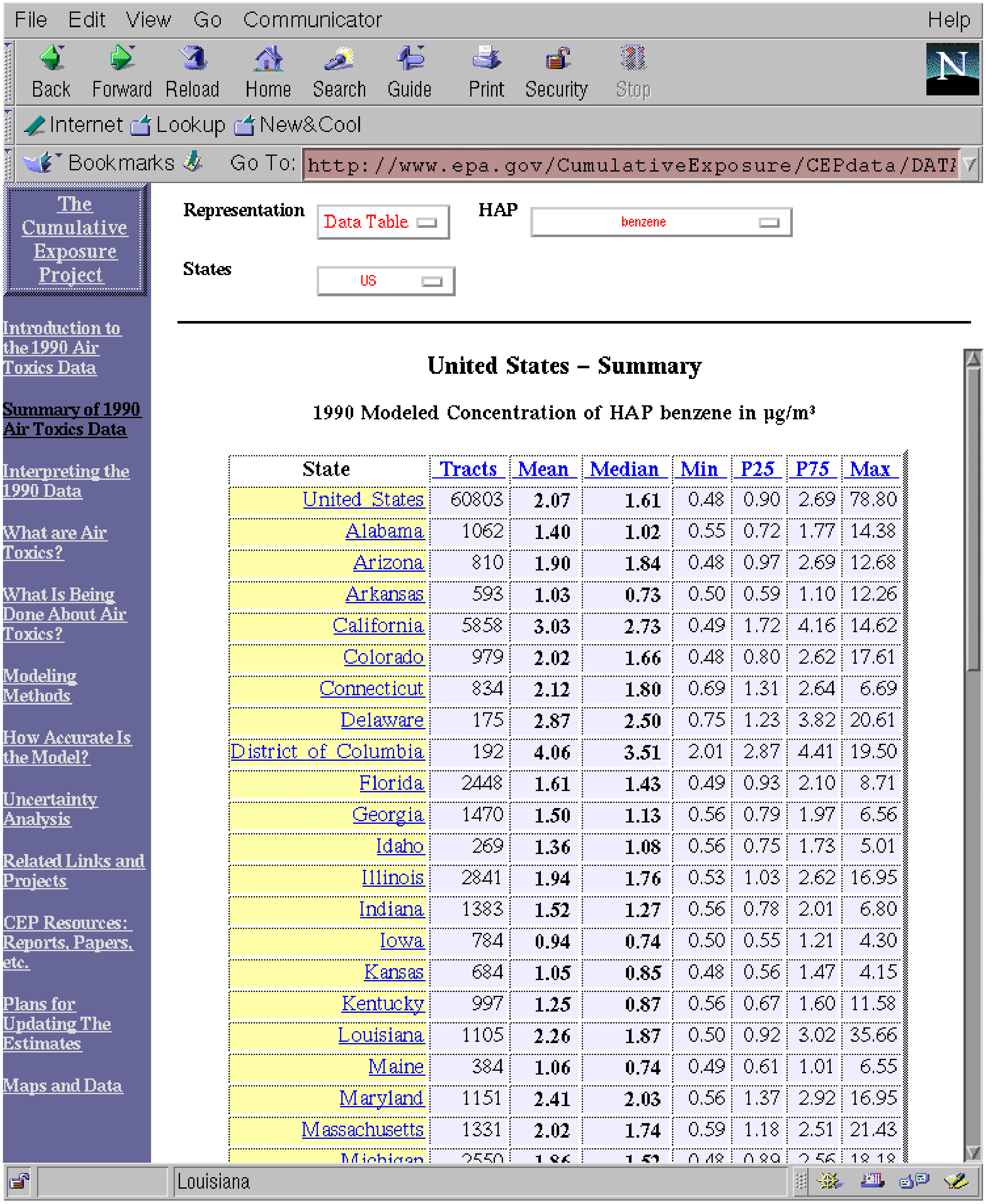


Figure 3: Micromap display at the state level for Michigan for all of its 83 counties. The total modeled 1990 HAP concentration in micrograms per cubic meter for all 148 HAPs has been averaged over the number of census tracts in each county. Counties are ordered from highest to lowest with respect to the total modeled 1990 HAP concentration.

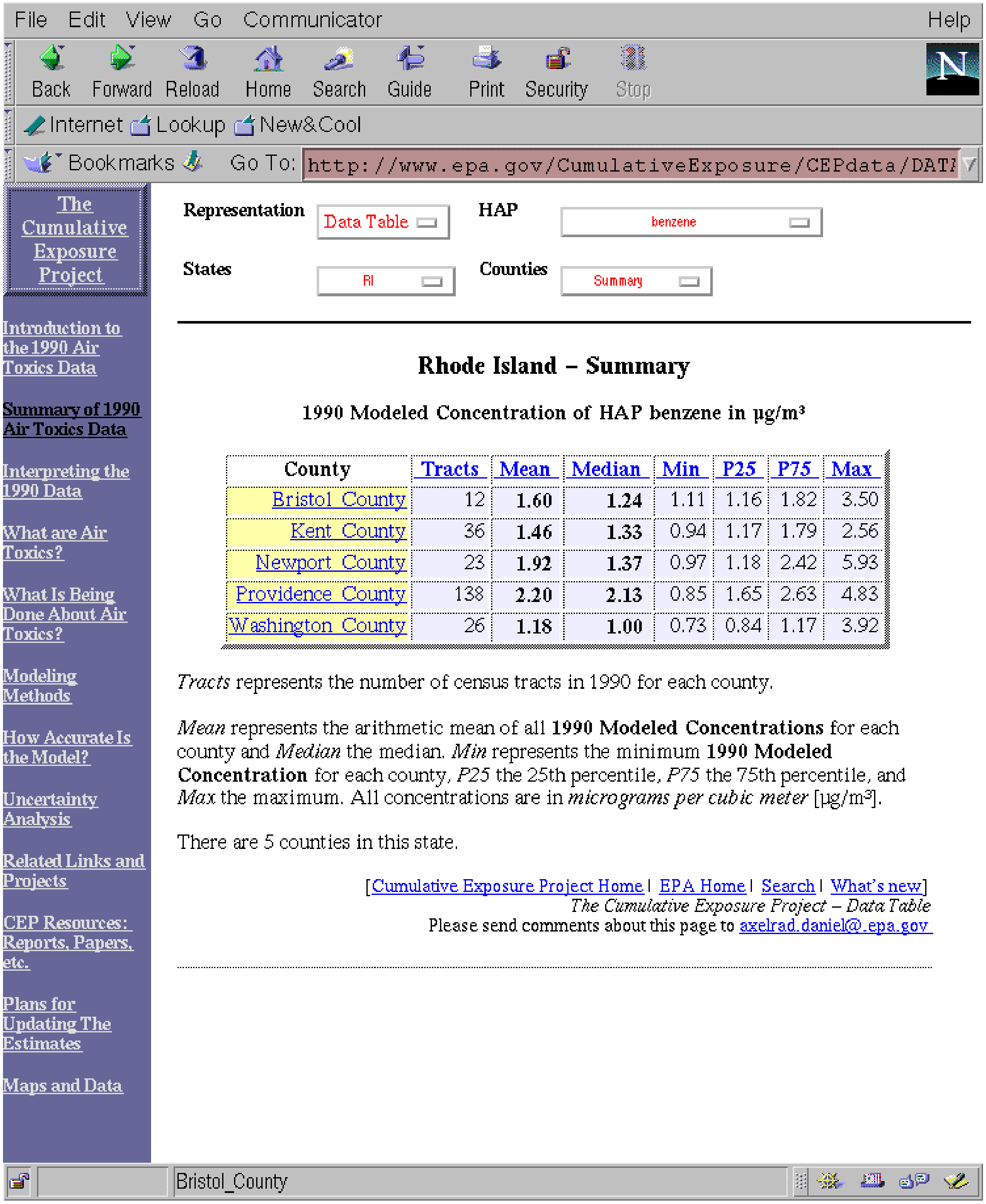


United States – Summary

1990 Modeled Concentration of HAP benzene in µg/m³

State	Tracts	Mean	Median	Min	P25	P75	Max
United States	60803	**2.07**	**1.61**	0.48	0.90	2.69	78.80
Alabama	1062	**1.40**	**1.02**	0.55	0.72	1.77	14.38
Arizona	810	**1.90**	**1.84**	0.48	0.97	2.69	12.68
Arkansas	593	**1.03**	**0.73**	0.50	0.59	1.10	12.26
California	5858	**3.03**	**2.73**	0.49	1.72	4.16	14.62
Colorado	979	**2.02**	**1.66**	0.48	0.80	2.62	17.61
Connecticut	834	**2.12**	**1.80**	0.69	1.31	2.64	6.69
Delaware	175	**2.87**	**2.50**	0.75	1.23	3.82	20.61
District of Columbia	192	**4.06**	**3.51**	2.01	2.87	4.41	19.50
Florida	2448	**1.61**	**1.43**	0.49	0.93	2.10	8.71
Georgia	1470	**1.50**	**1.13**	0.56	0.79	1.97	6.56
Idaho	269	**1.36**	**1.08**	0.56	0.75	1.73	5.01
Illinois	2841	**1.94**	**1.76**	0.53	1.03	2.62	16.95
Indiana	1383	**1.52**	**1.27**	0.56	0.78	2.01	6.80
Iowa	784	**0.94**	**0.74**	0.50	0.55	1.21	4.30
Kansas	684	**1.05**	**0.85**	0.48	0.56	1.47	4.15
Kentucky	997	**1.25**	**0.87**	0.56	0.67	1.60	11.58
Louisiana	1105	**2.26**	**1.87**	0.50	0.92	3.02	35.66
Maine	384	**1.06**	**0.74**	0.49	0.61	1.01	6.55
Maryland	1151	**2.41**	**2.03**	0.56	1.37	2.92	16.95
Massachusetts	1331	**2.02**	**1.74**	0.59	1.18	2.51	21.43
Michigan	2550	**1.86**	**1.52**	0.48	0.89	2.56	18.18

Figure 4: Tabular display at the op (US) level for all 48 states and the District of Columbia. The currently invisible states can be displayed by scrolling down the slider on the right hand side of the table. Displayed is the modeled 1990 concentration for benzene in micrograms per cubic meter. Summary statistics include number of census tracts, mean, median, minimum, 25th percentile, 75h percentile, and maximum with respect to the underlying census tracts for each state.



File Edit View Go Communicator Help

Back Forward Reload Home Search Guide Print Security Stop

Internet Lookup New&Cool

Bookmarks Go To: http://www.epa.gov/CumulativeExposure/CEPdata/DAT

The Cumulative Exposure Project

Introduction to the 1990 Air Toxics Data

Summary of 1990 Air Toxics Data

Interpreting the 1990 Data

What are Air Toxics?

What Is Being Done About Air Toxics?

Modeling Methods

How Accurate Is the Model?

Uncertainty Analysis

Related Links and Projects

CEP Resources: Reports, Papers, etc.

Plans for Updating The Estimates

Maps and Data

Representation: Data Table   HAP: benzene

States: RI   Counties: Summary

Rhode Island – Summary

1990 Modeled Concentration of HAP benzene in µg/m³

County	Tracts	Mean	Median	Min	P25	P75	Max
Bristol County	12	**1.60**	**1.24**	1.11	1.16	1.82	3.50
Kent County	36	**1.46**	**1.33**	0.94	1.17	1.79	2.56
Newport County	23	**1.92**	**1.37**	0.97	1.18	2.42	5.93
Providence County	138	**2.20**	**2.13**	0.85	1.65	2.63	4.83
Washington County	26	**1.18**	**1.00**	0.73	0.84	1.17	3.92

Tracts represents the number of census tracts in 1990 for each county.

Mean represents the arithmetic mean of all **1990 Modeled Concentrations** for each county and *Median* the median. *Min* represents the minimum **1990 Modeled Concentration** for each county, *P25* the 25th percentile, *P75* the 75th percentile, and *Max* the maximum. All concentrations are in *micrograms per cubic meter* [µg/m³].

There are 5 counties in this state.

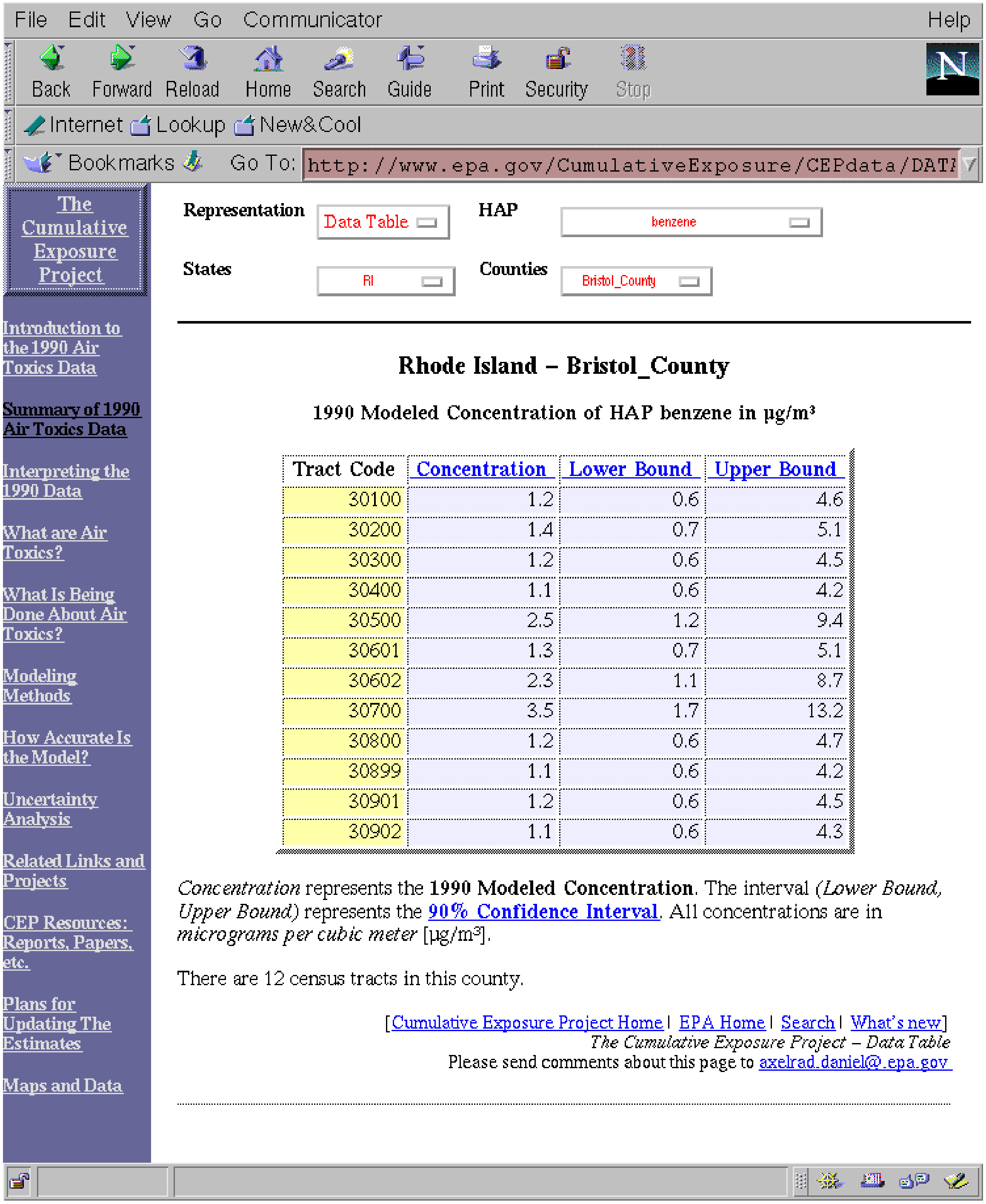
[Cumulative Exposure Project Home | EPA Home | Search | What's new]

The Cumulative Exposure Project – Data Table

Please send comments about this page to axelrad.daniel@epa.gov

Bristol_County

Figure 5: Tabular display at the state level for Rhode Island for all of its 5 counties. Displayed is the modeled 1990 concentration for benzene in micrograms per cubic meter. Summary statistics include number of census tracts, mean, median, minimum, 25th percentile, 75h percentile, and maximum with respect to the underlying census tracts for each county.



File Edit View Go Communicator Help

Back Forward Reload Home Search Guide Print Security Stop

Internet Lookup New&Cool

Bookmarks Go To: http://www.epa.gov/CumulativeExposure/CEPdata/DAT/

The Cumulative Exposure Project

Introduction to the 1990 Air Toxics Data

Summary of 1990 Air Toxics Data

Interpreting the 1990 Data

What are Air Toxics?

What Is Being Done About Air Toxics?

Modeling Methods

How Accurate Is the Model?

Uncertainty Analysis

Related Links and Projects

CEP Resources: Reports, Papers, etc.

Plans for Updating The Estimates

Maps and Data

Representation Data Table **HAP** benzene

States RI **Counties** Bristol_County

Rhode Island – Bristol_County

1990 Modeled Concentration of HAP benzene in µg/m³

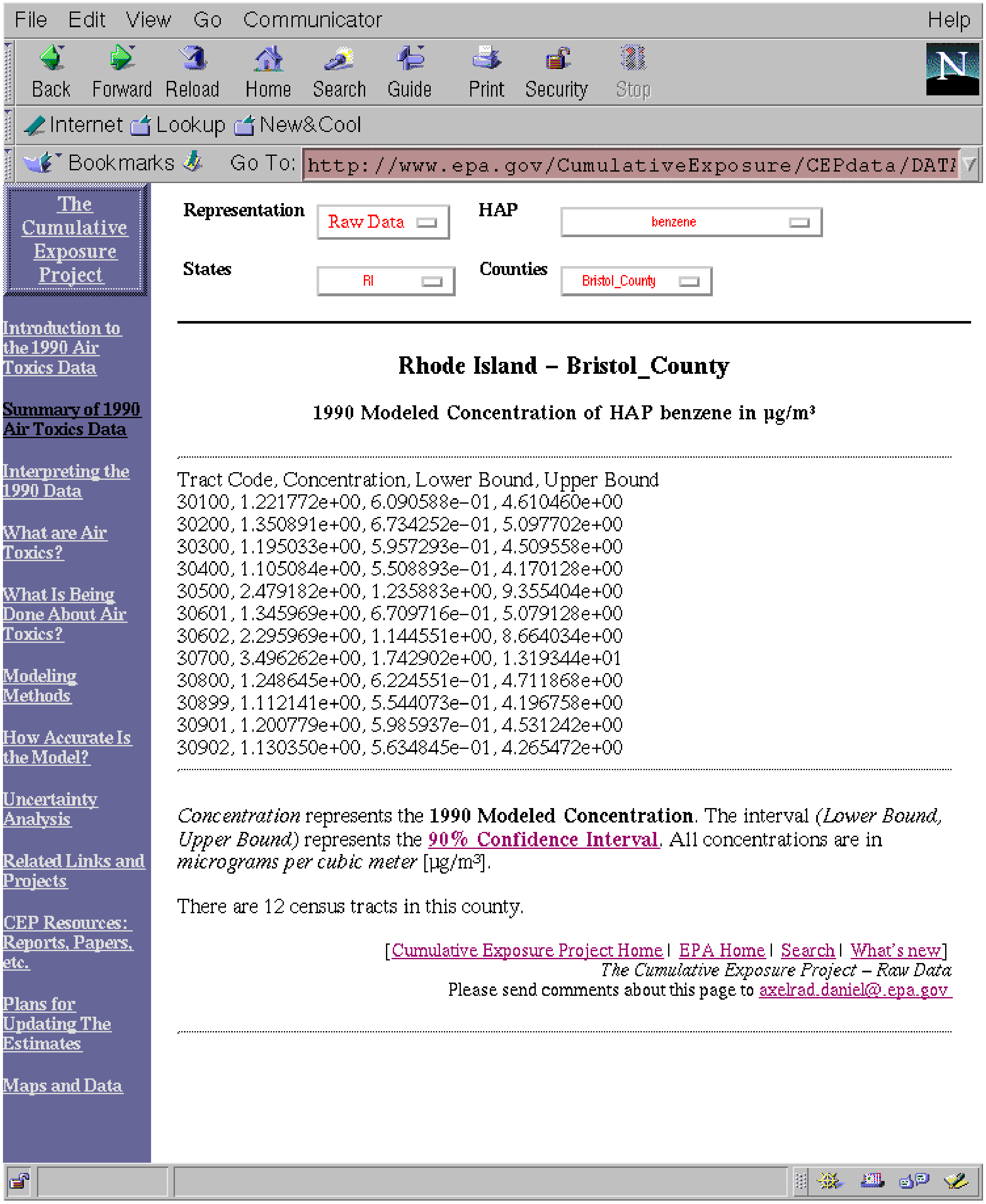
Tract Code	Concentration	Lower Bound	Upper Bound
30100	1.2	0.6	4.6
30200	1.4	0.7	5.1
30300	1.2	0.6	4.5
30400	1.1	0.6	4.2
30500	2.5	1.2	9.4
30601	1.3	0.7	5.1
30602	2.3	1.1	8.7
30700	3.5	1.7	13.2
30800	1.2	0.6	4.7
30899	1.1	0.6	4.2
30901	1.2	0.6	4.5
30902	1.1	0.6	4.3

Concentration represents the **1990 Modeled Concentration**. The interval *(Lower Bound, Upper Bound)* represents the **90% Confidence Interval**. All concentrations are in *micrograms per cubic meter* [µg/m³].

There are 12 census tracts in this county.

[Cumulative Exposure Project Home | EPA Home | Search | What's new]
The Cumulative Exposure Project – Data Table
Please send comments about this page to axelrad.daniel@epa.gov

Figure 6: Tabular display at the county level for Rhode Island's Bristol County for all of its 12 census tracts. Displayed is the modeled 1990 concentration for benzene in micrograms per cubic meter. In addition, a 90% confidence interval is displayed.



Rhode Island – Bristol_County

1990 Modeled Concentration of HAP benzene in µg/m³

Tract Code, Concentration, Lower Bound, Upper Bound
30100, 1.221772e+00, 6.090588e-01, 4.610460e+00
30200, 1.350891e+00, 6.734252e-01, 5.097702e+00
30300, 1.195033e+00, 5.957293e-01, 4.509558e+00
30400, 1.105084e+00, 5.508893e-01, 4.170128e+00
30500, 2.479182e+00, 1.235883e+00, 9.355404e+00
30601, 1.345969e+00, 6.709716e-01, 5.079128e+00
30602, 2.295969e+00, 1.144551e+00, 8.664034e+00
30700, 3.496262e+00, 1.742902e+00, 1.319344e+01
30800, 1.248645e+00, 6.224551e-01, 4.711868e+00
30899, 1.112141e+00, 5.544073e-01, 4.196758e+00
30901, 1.200779e+00, 5.985937e-01, 4.531242e+00
30902, 1.130350e+00, 5.634845e-01, 4.265472e+00

Concentration represents the **1990 Modeled Concentration**. The interval *(Lower Bound, Upper Bound)* represents the **90% Confidence Interval**. All concentrations are in *micrograms per cubic meter* [µg/m³].

There are 12 census tracts in this county.

[Cumulative Exposure Project Home | EPA Home | Search | What's new]
The Cumulative Exposure Project – Raw Data
Please send comments about this page to axelrad.daniel@epa.gov

Figure 7: Raw data display at the county level for Rhode Island's Bristol County for all of its 12 census tracts. Displayed is the modeled 1990 concentration for benzene in micrograms per cubic meter. In addition, a 90% confidence interval is displayed.