Design of Information Graphics for Causal Arguments

Nancy Green
Department of Mathematical Sciences
University of North Carolina at Greensboro
Greensboro, NC 27402-6170 USA
nlgreen@uncg.edu

Abstract

In this paper we analyze some information graphic design techniques used to help present causal arguments in human-authored articles on science and technology intended for an audience that may include non-scientists. In particular, our focus is on design techniques used to argue for or against causal relationships. We represent the causal claims of the arguments in the notation of qualitative probabilistic networks, which provides a concise computation-oriented representation of claims involving causation. The main goal of our analysis is to acquire knowledge to enable automatic synthesis of information graphics in an intelligent argument generation system.

1 Introduction

In quantitative fields, presentations often include information graphics, i.e., non-pictorial graphics such as line graphs and bar charts. The fields of statistics and information visualization have provided numerous design guidelines based on properties of the data and human perception (e.g., showing part-whole relationships with pie charts), convention (e.g., showing independent variables on the horizontal axis) and task analysis (e.g., facilitating lookup or comparison tasks). In addition, it has been noted often that, given the same data to present in a graphic, the design of the graphic may be varied for persuasive purposes [Cleveland, 1994].

In this paper we analyze some information graphic design techniques used to help present causal arguments in human-authored articles on science and technology intended for an audience that may include non-scientists. In particular, our focus is on design techniques used to argue for or against causal relationships. We represent the causal claims of the arguments in the notation of qualitative probabilistic networks [Wellman, 1990; Druzdzel and Henrion, 1993]. That notation provides a concise computation-oriented representation of claims involving causation. The main goal of our analysis is to acquire knowledge to enable automatic synthesis of information graphics in an intelligent argument generation system. In the following section, we demonstrate the use of this notation to represent causal claims of five classes of argument. In section 3, we discuss related research. In section 4, we describe how knowledge gained from the analyses could be operationalized for use by an information graphics component of an argument generation system, and other future work.

CMNA 5. 5th Workshop on Computational Models of Natural Argument at International Joint Conference on Artificial Intelligence 2005.

2 Analysis of Examples

2.1 Influence of a One-Time Event

Figure 1 appeared in an article on traffic safety claiming that "Changes in public policy have affected driver behavior and led to large reductions in traffic deaths" (Evans, p. 249). At one level of analysis, each graph in Fig. 1 might be summarized by the statement that the value of some variable (fatalities, etc.) tended to decrease over time. However, the design of these graphs also reflects the author's causal claim. In each, the time periods before and after a purportedly relevant change in public policy occurred are made salient. In the left and middle panels, color and annotations are used to make the periods salient. In the panel on the right, the contrast is made salient by presenting aggregate data for the periods before and after the change of state.

In summary, Figure 1 graphs the observed values (or an aggregate of the values) of a dependent variable (y-axis) over time (x-axis). The independent variable is a boolean-valued variable whose state changed once during the time period represented on the x-axis, and the time of the state change is made salient in the graphic. A related convention used in news reports about the stock market is to plot time series data, e.g., the Nasdaq composite index, with certain data points annotated. The annotations may refer to events (such as negative news reports) purported to account for significant changes in the data (such as a decline in stock price). (It should be noted that this technique is not employed only for causal argumentation. For example, annotations are sometime used to provide historical context without implying causation.)

Our interpretation of the author's causal explanation for the data presented in the bar chart on the left side of the right panel in Figure 1 is shown in Figure 2 as a qualitative probabilistic network (or QPN) [Wellman, 1990; Druzdzel and Henrion, 1993]. A QPN is an abstraction of a Bayesian belief network in which numeric probabilities are replaced by qualitative constraints such as positive/negative qualitative influence and additive synergy (defined below). The boolean-valued variable labeled *Law?* represents whether the seat belt law was in effect, and can be inferred from the date of an observation of *# fatalities*. The nodes labeled *# wearing seat belts* and *# fatalities* are discrete variables counting the number of car occupants who wore seat belts and the number who died in automobile accidents, respectively.

While the latter is an observable variable, the former was hypothesized in the article to account for the change in the number of fatalities. A relation of positive qualitative influence (denoted as S^+) holds between *Law?* and *# wearing seat belts*, and a relation of negative influence (denoted as S^-) holds between *# wearing seat belts* and *# fatalities*. Informally, $S^+(A,B)$ expresses that a higher value of A makes higher values of B more likely [Druzdzel and van der Gaag, 1995]. In other words, when *Law?* is *true*, *# wearing seat belts* increases, while *# fatalities* decreases. Similar networks can be used to represent the causal claims of the other graphics in Fig. 1.

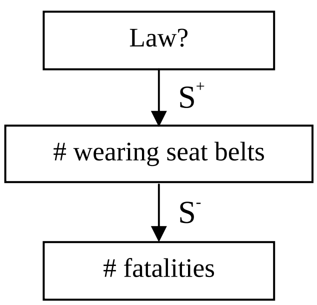

Figure 2. QPN for Fig. 1

2.2 Refuting Influence of a One-Time Event

Figure 3 appeared in an article that claimed, “Although medicine was credited with defeating [tuberculosis], more than 90 percent of the decline in mortality rates had taken place before a vaccine became available, suggesting that social and economic change had done most of the work” (Hertzman p. 540). At one level of analysis, the graph in Fig. 3 can be summarized by the statement that the value of some variable tended to decrease over time. However, the design of the graph also reflects the claim about an absence of a causal relationship. In Figure 3, an annotation shows when vaccination became available, which is near the end of the time series.

In summary, Figure 3 graphs the observed values of a dependent variable (y-axis) over time (x-axis). The independent variable is a boolean-valued variable whose state changed once during the time period represented on the x-axis, and the time of the state change is made salient in the graphic. In addition, the graphic makes salient that the behavior of the dependent variable before and after the state change remained constant (i.e., in this case, there was a continued decreasing trend).

Our interpretation of the author's causal explanation for the data presented in Figure 3 is shown as a QPN in Figure 4. The lack of significant negative (or positive) causal influence of vaccination availability on reducing mortality from tuberculosis is indicated by the S^0 relation on the link from *Vaccination?* to *# TB deaths*. The author’s alternative hypothesis (not reflected in Figure 3 at all) is represented by the variable *Socio-econ*, whose purported causal influence is indicated by the S^- on the link from *Socio-econ* to *# TB deaths*.

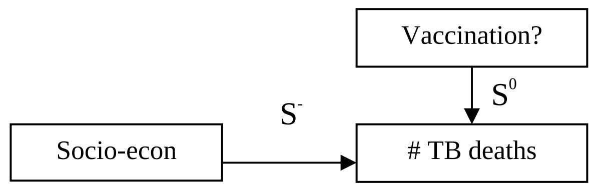

Figure 4. QPN for Fig. 3

2.3 Common Cause

The left panel of Figure 5 can be summarized loosely by the author’s statement that “traffic safety is largely a problem of young male drivers” (Evans, p. 244). Figure 5 shows the association between age and automobile crashes (or fatalities, in the inset) for males and females in individual connected scatter plots. Then, by presenting the two plots in the same graph, the design enables a viewer to compare the plots and see that crashes (or fatalities) are indeed highest for young male drivers.

However, association alone does not imply causation. The author proposes that the higher accident/fatality rate in this demographic is due to the tendency of young males to engage in reckless behavior. The author argues for this by providing the graph shown in the right panel of Figure 5, and pointing out the "strikingly similar profiles [in terms of age and gender] of those arrested for crimes unrelated to driving" (p. 250). In summary, this article’s causal claim is expressed in part by providing the graphs showing association, and also in the juxtaposition of the two figures for different dependent variables (i.e., crashes and arrests).

In Figure 6, we represent the author's causal explanation. As indicated by the S^+ on the arc from *Gender* to *Safe Behavior?*, a higher value of *Gender* (i.e., *female*) increases the likelihood of a higher value of *Safe Behavior?* (i.e., *true*), where *Safe Behavior?* subsumes unsafe driving and other inappropriate behaviors that may lead to arrest. (Of course, *Gender* is actually a categorical variable; the ordering used in our model was chosen to enable the arc from *Gender* to *Safe Behavior?* to be marked with a positive sign, which this author finds more "natural".)

In Fig. 6, we have extended the notation of qualitative causal graphs with ordered lists of influences; the list of causal influences (S^-, S^+) on the arc from *Age* to *Safe Behavior?* expresses that the direction of qualitative influence is negative (*S*-) up to a certain age range, and positive (*S+*) afterwards. Also, a relationship of negative additive synergy [Druzdzel and Henrion, 1993] holds between *Gender* and *Age* (annotated as Y^-). Informally, positive (negative) additive synergy, annotated as Y^+ (Y^-), between A and B with respect to C expresses that the joint influence of A and B is greater (less) than the sum of their individual influences (Druzdzel and van der Gaag 1995). That is, the negative effect of lower values of *Age* on *Safe Behavior?* is mitigated by the effect of a higher value of *Gender*.

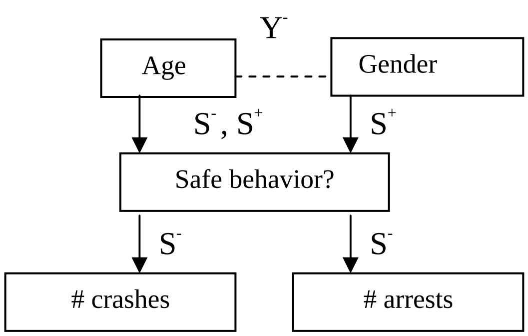

Figure 6. QPN for Fig. 5.

2.4 Cumulative Effect

The graph in Figure 7 can be summarized by the statement that anthropogenic emissions have risen concurrently with a rise in atmospheric concentrations of greenhouse gases. However, the graph was presented in an on-line document claiming that the increase in atmospheric concentrations is due to "human (anthropogenic) activity" (U.S. Dept. of Energy). Figure 7 graphs the observed values of the dependent variable atmospheric concentrations over time, in juxtaposition with a graph of the observed values of the independent variable, anthropogenic emissions, over time. (Note that in order to juxtapose the two graphs, the designer used two different measurement scales. The vertical axis on the left shows CO_2 atmospheric concentrations in parts per million, while the one on the right shows anthropogenic CO_2 emissions in million metric tons.) However, temporal association alone does not imply causation. To strengthen the causal argument, the article included a diagram of the global carbon cycle. That diagram depicts how natural processes can remove the greenhouse gases arising from natural causes from the atmosphere, and how the excess arising from anthropogenic activity remains in the atmosphere.

Our interpretation of the author's causal explanation for the data presented in Fig. 7 is shown as a time slice of a QPN in Fig. 8. The variables labeled *Anthro*(T_i) and *Anthro*(T_{i+1}) represent anthropogenic emissions, and *GG*(T_i) and *GG*(T_{i+1}) atmospheric greenhouse gases, at times T_i and T_{i+1}, respectively. *Absorb*(T_i) represents the amount of greenhouse gases absorbed by natural processes at time T_i . The list of influences (S^+, S^0) on the arc from *GG*(T_i) to *Absorb*(T_i) represents the claim that as atmospheric greenhouse gases increase the amount absorbed increases, but only up to a point, after which they are not. The QPN also shows that *GG*(T_{i+1}) increases directly due both to *Anthro*(T_{i+1}) and *GG*(T_i), i.e., the atmospheric greenhouse gases that were not absorbed.

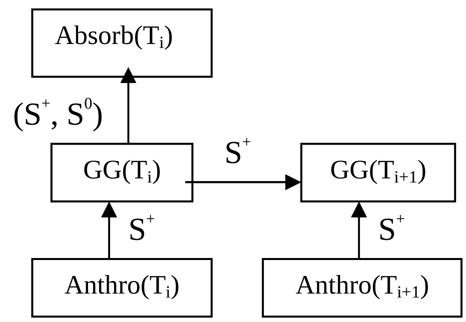

Figure 8. QPN for Fig. 7.

2.5 Spatial Coordinates of Causal Variables

Another design technique for expressing causal relationships is the presentation of quantitative data with a map. Tufte [1997] reprints a street map designed by John Snow (On the Mode of Communication of Cholera, 1855) that shows the locations of water pumps and the number of deaths from cholera; the display of a cluster of deaths close to one of the pumps was used to support the argument that the pump had played a role in spreading the disease. Tufte [1983] reprints a more complex, six-variable graphic designed by Charles Joseph Minard to display data about Napolean's army's Russian campaign in 1812. The army's route to and from Moscow is shown as an upper and lower band, respectively, whose width represents the size of the army at the corresponding location on the map. In addition, the lower (return route) band is correlated with a reversed time series graph at the bottom of the graphic showing the temperature at points on the return route. The design is intended to show the effects of a river crossing and freezing weather on the army's size.

Our interpretation of Snow's argument could be represented in a QPN similar to that of Fig. 2, i.e., by treating spatial data like time-series data. The Minard graphic could be represented in a QPN with two causal variables, one representing a river crossing and the other the temperature, with causal influences on a third variable representing losses to Napolean's army.

3 Related Work

Probabilistic belief network formalisms were developed in Artificial Intelligence (AI) for performing automated reasoning under uncertainty, especially for problems involving causal relationships. These formalisms have been used to implement domain models and user models in argument generation systems [e.g., Zukerman et al., 2000; Carofiglio and de Rosis, 2003]. However, probabilistic belief networks have not been used in AI to represent causal claims conveyed through information graphics.

Previous AI research on automatic generation or interpretation of information graphics has not addressed causal argument. The AutoBrief project addressed the automatic generation of information graphics to express communicative goals of quantification (e.g., *Most x are y*) and comparison (e.g., *There is twice as much x as y*) [Green et al., 2004]. The Graph project addressed the automatic interpretation and summarization of trends and comparisons in bar charts [Elzer et al., 2005]. In earlier work, we analyzed the role of information graphics in multimedia arguments but did not attempt to provide a computation-oriented representation of the causal claims of an argument [Green, 2001].

In philosophy and science education, several computer-supported argument mapping tools have been developed to enable students to construct diagrams representing their analysis of arguments. For example, Reason!Able [van Gelder, 2002] and Belvedere [Suthers et al., 1995] provide computerized diagramming tools for representing an argument as a network, where nodes contain textual descriptions of claims, and arcs represent evidential relationships between claims. Although causal arguments can be represented, these notations do not encode causal models.

In business management, software tools have been developed to enable users to construct causal diagrams representing the analysis of causal relationships in business systems [Kirkwood, 1998]. As in a QPN, nodes represent event or state variables and arcs represent positive or negative qualitative causal influence. Kirkwood [1998] provides examples illustrating causal models underlying common patterns in business data graphics. Since we were not aware of that work until quite recently, we have not yet had a chance to incorporate those insights into our collection of graphic design techniques found in human-authored articles.

Argument mapping and business systems analysis notation is intended as a cognitive aid; the diagrams using that notation are created by and for human agents. In contrast, as discussed in the next section, the function of the QPN representation in our work is to provide an internal representation of causal claims for use in an argument generation system.

4 Future Work

We plan to continue analyzing uses of information graphics in human-authored causal arguments. Our goal is to acquire a repertoire of information graphic design techniques that can be used to convey different types of causal claims. Then, the knowledge could be employed by an intelligent argument generation system. Given a communicative goal, such a system would use relevant data, a domain model including qualitative constraints illustrated in this paper, knowledge of argumentation techniques, and knowledge of information graphic design techniques to produce an argument in text and information graphics. We shall outline a design for such a system now.

The proposed design uses AI planning as in some previous systems for intelligent multimedia generation [e.g., Green et al., 2004] and generation of arguments presented in text [e.g., Carenini, 2001; Reed and Long, 1998; Grasso et al., 2000]. (These systems did not, however, address generation of information graphics for causal arguments.) In this approach, strategies for achieving communicative goals are encoded as abstract plan operators; an AI planner selects, combines and instantiates plan operators to construct a plan to achieve a given communicative goal.

For example, given the *goal to persuade an audience A to support a proposed law SBL in country C to require automobile occupants to wear seat belts*, our proposed system might construct a plan paraphrased as follows: given that A believes that reducing future traffic fatalities in C is worthwhile, achieve the *subgoal G1 that A believe that if SBL is passed, then most automobile occupants in C will wear seat belts; and if most automobile occupants in C will wear seat belts, then traffic fatalities in C will decrease*. The part of the plan to achieve subgoal G1 is paraphrased as: achieve the *subgoal G2 that A know that there is a seat belt law SBL2 in effect in the past in a country C2* [presupposing similar traffic conditions, compliance with traffic laws, etc.] *such that SBL2 is responsible for a decrease in traffic fatalities in C2 after SBL2 went into effect*.

The part of the plan to achieve subgoal G2 is paraphrased as: perform the *action of presenting A with quantitative data as evidence for the claim of G2*. We assume that, although this part of the plan does not specify whether the subgoal should be achieved by presenting the data in text or graphics, the system could apply knowledge to select an appropriate medium, e.g., as in [Green et al., 2004]. Suppose that the medium of information graphics has been selected. Also, for simplicity of exposition, let us assume that information graphic design techniques for conveying causal claims have been encoded by the system developers as plan operators. An operator to generate a graphic like one of the bar charts in Figure 1 could be specified as follows:

- Constraint 1: There exists a set of time-series data with attributes F [e.g., the number of traffic fatalities] and T, where T is the temporal attribute
- Constraint 2: There exists a domain model including a boolean variable E representing an event [e.g., a seat belt law going in effect] such that the time of E, t(E), is within the range of values of T occurring in the data set, and a variable P [e.g., the number of people wearing seat belts] such that $S^+(E,P)$ and $S^-(P, F)$
- Constraint 3: Let Before be the sequence of time points up to and including t(E), and After be the sequence after t(E). Let Sum_{before} and Sum_{after} be the sum of the values of F over the time intervals Before and After, respectively. $Sum_{before} > Sum_{after}$.
- Goal: Viewer believe that there exists a variable X such that $S^+(E,X)$ and $S^-(X, F)$
- Decomposition: Create a vertical bar chart with Before and After mapped to the x-axis and Sum_{before} and Sum_{after} mapped to the y-axis.

Thus, the part of the plan to achieve G2 could be created by instantiating this operator. Then, an automated graphics generator, such as that used by AutoBrief [Green et al., 2004], would transform this part of the plan into an information graphic.

It is beyond the scope of our research to enable a system to derive a causal domain model automatically from a data set. However, a possible offshoot of our research is use of AI methods to infer causal domain models conveyed by information graphics designed by humans. The goal of the Graph project is to make information graphics in popular media accessible to individuals with sight impairments [Elzer et al., 2005]. Using probabilistic plan inference techniques to infer the communicative goals of an information graphic, the current Graph system is limited to recognizing non-causal claims conveyed by simple bar charts. However, using knowledge such as that encoded in the above plan operator, it may be possible to extend the Graph system's approach to interpretation of information graphics intended to convey causal claims. A system's interpretation of the causal model underlying a graphic then could be communicated to the user. In addition to providing a spoken summary for users with limited sight, a system could generate causal influence diagrams to assist sighted users in comprehending an information graphic used in causal argument.

5 Conclusions

We showed a variety of techniques used by human designers for expressing causality (or absence thereof) in information graphics presenting time series data, spatial data, and associations between variables. We represented the causal claims of the examples in terms of QPN notation, which provides a computation-oriented representation of the causal claims of an argument. We outlined how knowledge gained from the analyses might be used to generate information graphics in an argument generation system. Also, we suggest that it might be used in systems that perform automated interpretation of information graphics.

Acknowledgments

This material is based upon work supported by the National Science Foundation under CAREER Award No. 0132821.

Graphics Sources

(1) Figure 1 in Energy Information Administration, U.S. Department of Energy. *Greenhouse Gases, Global Climate Change, and Energy*. Retrieved May 3, 2005 from http://www.eia.doe.gov/oiaf/1605/ggccebro/chapter1.html.

(2) Figures 8 and 9 in Evans, L. 2002. Traffic Crashes. *American Scientist* 90, May-June 2002, 244-253. Reprinted by permission of *American Scientist*, magazine of Sigma Xi, The Scientific Research Society.

(3) Figure 2 in Hertzman, C. 2001. Health and Human Society. *American Scientist* 89, November-December 2001, 538-545.

References

Carafiglio, V. and de Rosis, F. 2003. Combining Logical with Emotional Reasoning in Natural Language Argumentation. *Proc. of Worshop on Affective User Modeling at User Modeling 2003*, 9-15.

Carenini, G. 2001. GEA: A Complete, Modular System for Generating Evaluative Arguments. *Proc. of the Int. Conf. on Computational Systems (ICCS2001), Workshop on Computational Models of Natural Language Arguments (CMNA 01), May 2001*. Springer Lecture Notes in Computer Science 2073, 959-968.

Cleveland, W. S. 1994. *The Elements of Graphing Data. Revised Edition.* Summit, NJ: Hobart Press.

Druzdzel, M.J. and van der Gaag, L.C. 1995. Elicitation of Probabilities for Belief Networks: Combining Qualitative and Quantitative Information. *Proc. of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*, pages 141-148, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1995.

Elzer, S., Carberry, S., Zukerman, I., Chester, D., Green, N., and Demir, S. 2005. A Probabilistic Framework for Recognizing Intention in Information Graphics. *Proc. of the Int. Joint Conf. on Artificial Intelligence.* To appear.

Grasso, F., Cawsey, A., and Jones, R. 2000. Dialectical Argumentation to Solve Conflicts in Advice Giving: A Case Study in the Promotion of Healthy Nutrition. *Int .J. Human-Computer Studies* 53, 1077-1115.

Green, N. 2001. An Empirical Study of Multimedia Argumentation. *Proc. of the Int. Conf. on Computational Systems (ICCS2001), Workshop on Computational Models of Natural Language Arguments (CMNA 01), May 2001*. Springer Lecture Notes in Computer Science 2073, 1009-18.

Green, N.L., Carenini, G., Kerpedjiev, S., Mattis, J., Moore, J.D., and Roth, S.F. 2004. AutoBrief: an Experimental System for the Automatic Generation of Briefings in Integrated Text and Information Graphics. *Int. J. of Human-Computer Studies,* 61(1):32-70.

Kirkwood, C.W. 1998. *System Dynamics Methods: A Quick Introduction.* Retrieved May 3, 2005 from http://www.public.asu.edu/~kirkwood/sysdyn/SDRes.htm

Reed, C.A. and Long, D.P. 1998. Generating the structure of argument. *Proc. 17th Int. Conf. on Computational Linguistics and 36th Annual Meeting of the Assoc. for Computational Linguistics (COLING-ACL'98),* 1091-1097.

Suthers, D., Weiner, A., Connelly, J., Paolucci, M. 1995. Belvedere: Engaging Students in Critical Discussion of Science and Public Policy Issues. *Proc. 7th World Conf. on AI in Education (AI-ED 95)*.

Tufte, E.R. 1983. *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.

Tufte, E.R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, Connecticut: Graphics Press.

van Gelder, T. 2002 Argument Mapping with Reason!Able. *APA Newsletter*, Fall 2002, Volume 02, Number 1, p. 85-90.

Wellman, M.P. 1990. Fundamental Concepts of Qualitative Probabilistic Networks. *Artificial Intelligence*, 44(3):257-303, August 1990.

Zukerman, I., McConachy, R., and Korb, K. Using Argumentation Strategies in Automated Argument Generation. *Proc. of INLG2000,* 55-62.

Appendix: Figures from sources.

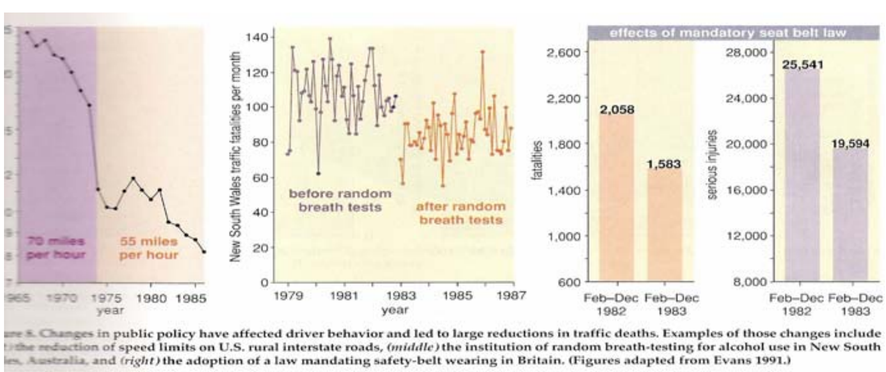

Figure 1. Figure 8 from Evans.

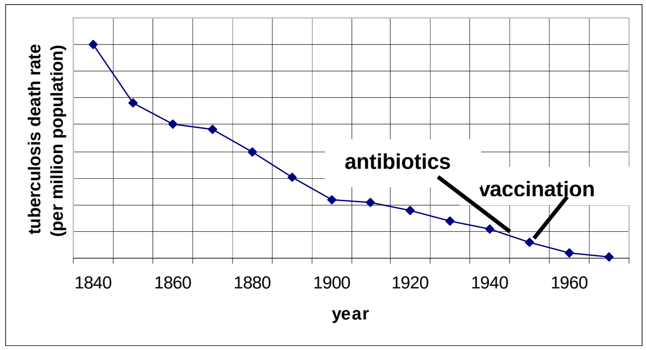

Figure 3. Line graph similar to Figure 2 in Hertzman. (Data in Figure 3 not the same as data in Hertzman's figure.)

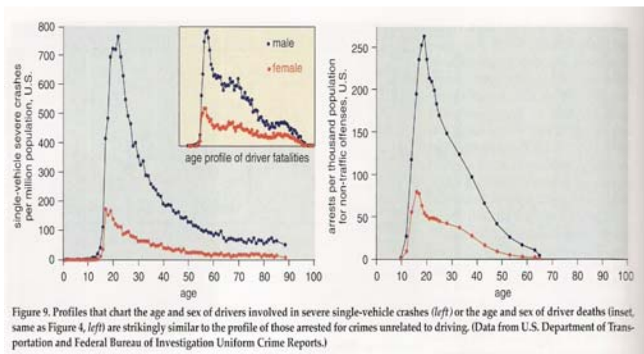

Figure 5. Figure 9 from Evans.

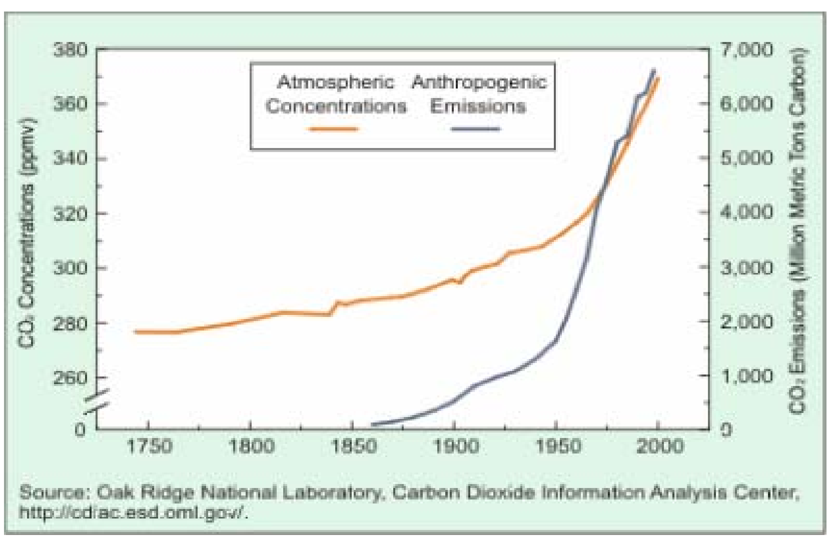

Figure 7. Figure 1 from U. S. Dept. of Energy.