**DSO 424: HW 2**

**Due 11:59 PM September 30, 2021**

**Submission instructions:**

* I suggest working in groups of 1-3 students.
* Use word file template to answer questions and use R code template to write your code.
* There are 2 submission links: to submit word and R files with solutions. In word file provide answer the questions in R file include code that is needed to be run to justify your answers.
* If you are working in a team, **only one** team member needs to submit both word and R files. Since you list all team members’ names at the top of this file I will know who worked on the assignment.

**Learning Objectives and Outcomes:**

* Explain in your own words explain the key ideas behind AB-testing and regression for intervention analysis;
* Identify new opportunities to use these techniques across multidisciplinary domains to guide decision making;
* Confidently apply these techniques to novel problems using analytical software;
* Formulate and communicate actionable business recommendations based upon your analysis, including its limitations;
* Critically assess the validity of analytics-based recommendations in the context of specific business decisions.

**Case 1**

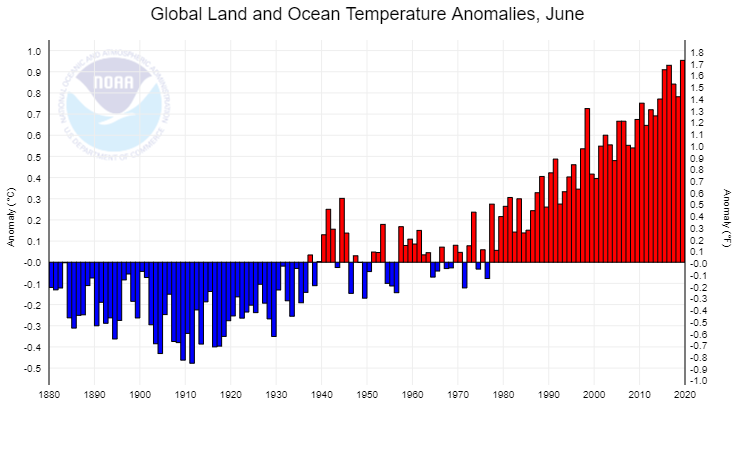
**Background (the info in this section is gathered from internet)**

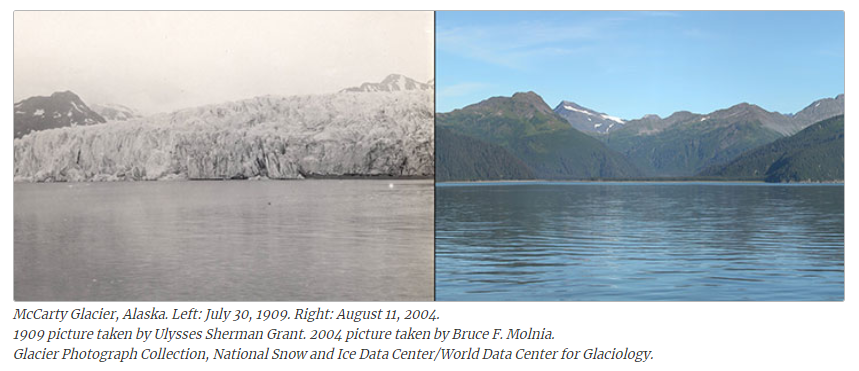
Remark: problem background description contains the parts of article

*My 1975 'Cooling World' Story Doesn't Make Today's Climate Scientists Wrong*

That was published in journal Inside Science

My 1975 'Cooling World' Story Doesn't Make Today's Climate Scientists Wrong – Peter Gwynne reflects on his publication in Science (the peer-reviewed academic journal of the American Association for the Advancement of Science (AAAS) and one of the world's top academic journals) about climate science that appeared almost 40 years ago.







Titled "The Cooling World," it was remarkably popular; in fact, it might be the only decades-old magazine story about science ever carried onto the set of a late-night TV talk show. Now, as the author of that story, after decades of scientific advances, let me say this: while the hypotheses described in that original story seemed right at the time, climate scientists now know that they were seriously incomplete. Nevertheless, certain websites and individuals that dispute, disparage and deny the science that shows that humans are causing the Earth to warm continue to quote my article. Their message: how can we believe climatologists who tell us that the Earth's atmosphere is warming when their colleagues asserted that it's actually cooling? Well, yes, we should trust them, despite the views of detractors such as comedian Dennis Miller, who brought my story to The Tonight Show in 2006. Several atmospheric scientists did indeed believe in global cooling, as I reported in the April 28, 1975 issue of Newsweek. But that was then. What is new over the last decade is that we know with increasing certainty that climate change is happening now," it states. "While scientists continue to refine projections of the future, observations unequivocally show that climate is changing and that the warming of the past 50 years is primarily due to human-induced emissions of heat-trapping gases. The certainty that our atmosphere is indeed warming stems from a series of rigorous observations and theoretical concepts that fit into computer models and an overall framework outlining the nature of Earth's climate. These capabilities were primitive or non-existent in 1975. In fact my report reflected a real strand of climatological thinking back then. I was far from the only science writer to cover the possibility of global cooling. Time, Science News, and the New York Times, among other media outlets, wrote about it, because some climate scientists had genuine reasons to believe that the global climate might be cooling and had published scholarly papers on the matter. Speaking personally, though, I accept that I didn't tell the full story back then. Indeed, the issue raises questions about the relationship between science writers and scientists as well as the attitudes toward science of individuals with political agendas.

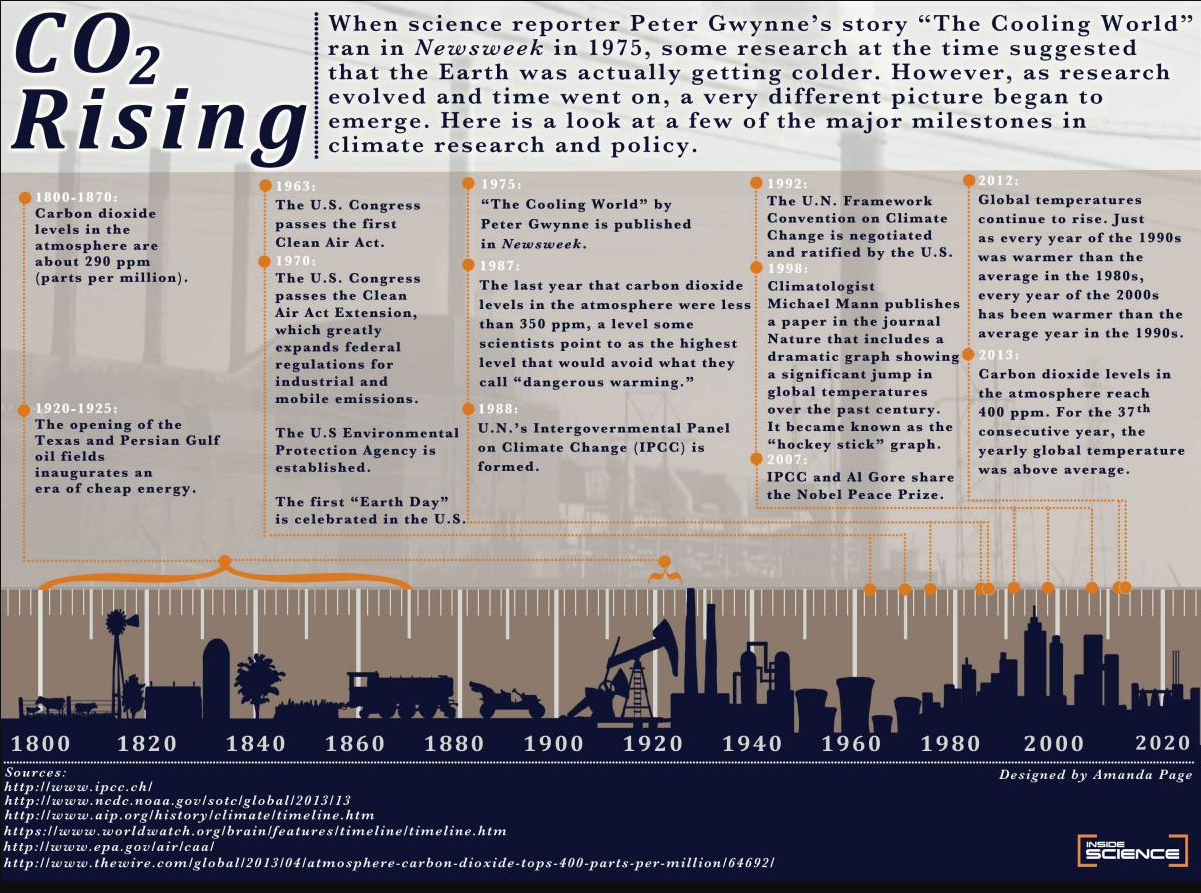
"There's no serious dispute any more about whether the globe is warming, whether humans are responsible, and whether we will see large and dangerous changes in the future – in the words of the National Academy of Sciences – which we didn't know in the 1970s," said Michael Mann, a climatologist at Pennsylvania State University in University Park. "The science was sort of speculative [in 1975]," Mann recalled.

Ironically, efforts to clean up the atmosphere made it possible to resolve the scientific mystery and convince climatologists that human activity is warming the planet. Policy actions such as the Clean Air Act of 1970 in the United States and similar initiatives in other countries aimed to reduce the amount of sulfate aerosols in the atmosphere. Since those compounds primarily reflect heat, their reduction effectively gave carbon dioxide and other greenhouse gases more control over the Earth's temperature. NASA scientist James Hansen was the first to sound the alarm. In 1988, he pointed out that a sort of Faustian bargain had cleaned up the atmosphere but at the cost of worsening the greenhouse problem. (Faustaian bargain, best exemplified by the legend of Faust and the figure of Mephistopheles, is a metaphor that represents a deal with the “devil” sacrificing anything to satisfy a limitless desire for money, knowledge, power.)

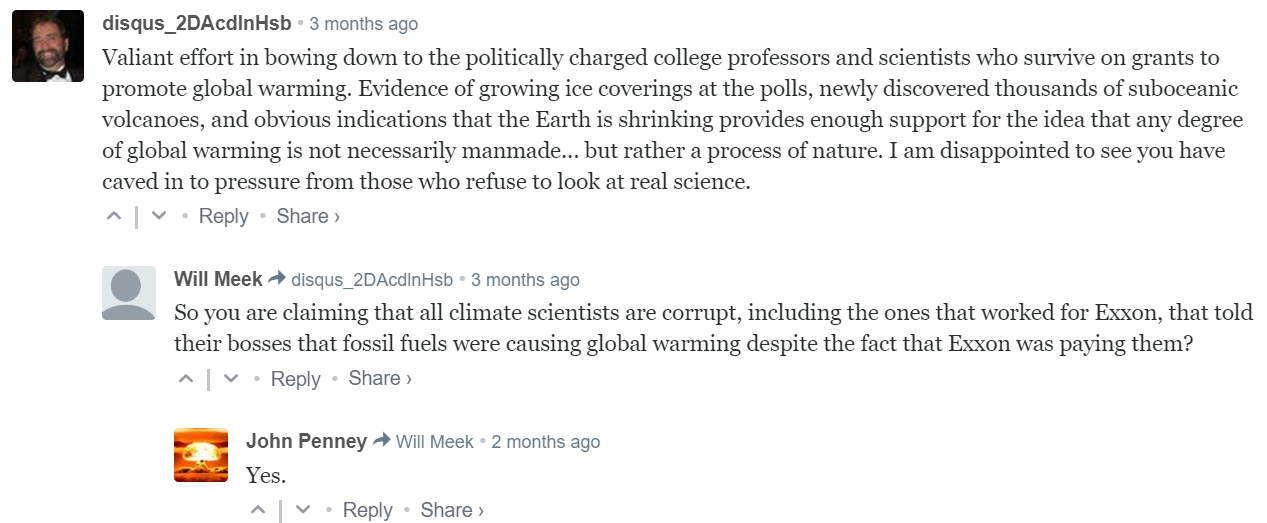
"There are many lines of observational evidence that the world is warming, including globally rising air and ocean temperatures, retreating glaciers worldwide, increasing sea level, decreasing Arctic Sea ice extent, and mass loss on the ice sheets of Greenland and Antarctica," Richard Somerville, a climate scientist at the Scripps Institution of Oceanography and the University of California, San Diego, summarized the findings. "In addition, an entire new body of climate science called 'detection and attribution' convincingly shows that the observed climate changes have distinctive space-time patterns that are consistent with causes due to human activities."

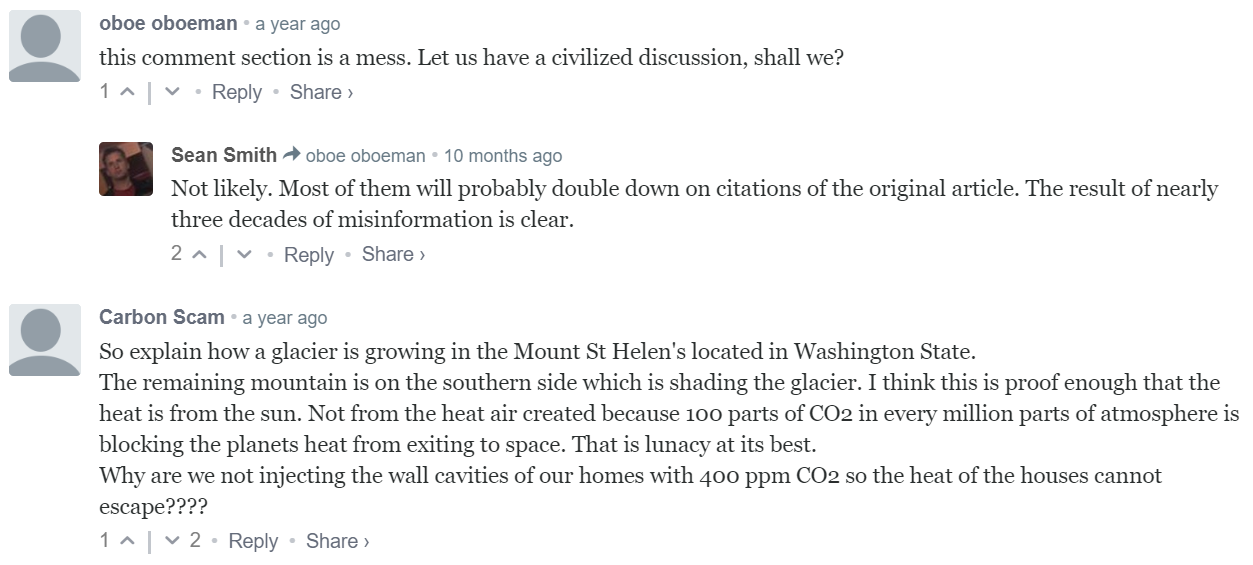
In its 2001 report on global climate, the Intergovernmental Panel on Climate Change of the United Nations prominently featured the “Hockey Stick,” a chart showing global temperature data over the past one thousand years. The Hockey Stick demonstrated that temperature had risen with the increase in industrialization and use of fossil fuels. The inescapable conclusion was that worldwide human activity since the industrial age had raised CO2 levels, trapping greenhouse gases in the atmosphere and warming the planet.

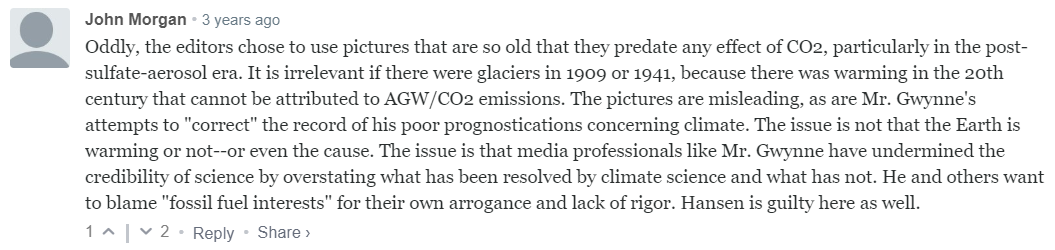
The Hockey Stick became a central icon in the “climate wars,” and well-funded science deniers immediately attacked the chart and the scientists responsible for it. Yet the controversy has had little to do with the depicted temperature rise and much more with the perceived threat the graph posed to those who oppose governmental regulation and other restraints to protect our environment and planet. Michael E. Mann, lead author of the original paper in which the Hockey Stick first appeared, shares the real story of the science and politics behind this controversy. He introduces key figures in the oil and energy industries, and the media front groups who do their bidding in sometimes slick, bare-knuckled ways to cast doubt on the science. As the scientific evidence becomes clearer and the threat becomes clearer, it takes yet more disinformation and propaganda to obscure the truth. Hundreds of millions of dollars have been spent by fossil fuel interests seeking to muddy the waters. That has, in turn, provided cover for politicians doing their bidding in opposing any attempts to regulate carbon emissions. Mann concludes with an account of the “Climategate” scandal, the 2009 hacking of climate scientists’ emails. Throughout, Mann reveals the role of science deniers, abetted by an uninformed media, in once again diverting attention away from one of the central scientific and policy issues of our time.

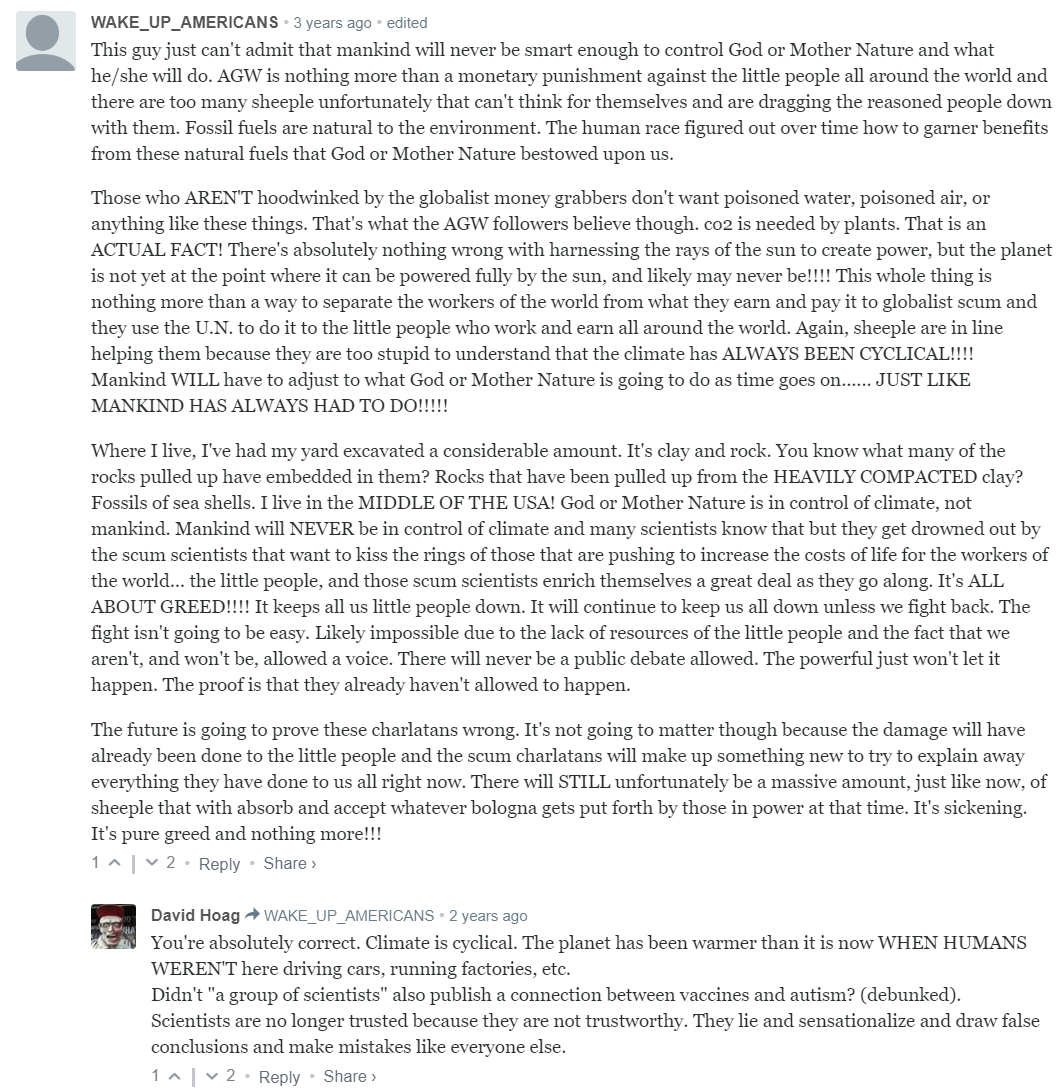


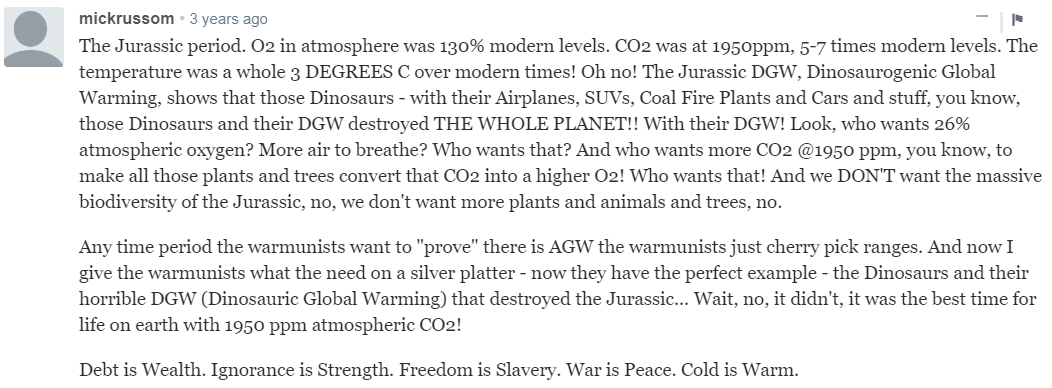
A few comments below the published article that I included below are rather controversial. Those who are curious please glance over. Hearing different opinions and getting familiar with different perspectives help us develop and use our critical and analytical thinking skills.











The question of global warming has been the topic of critical importance in academic research. Below are a few examples that illustrate the negative impact of global warming:

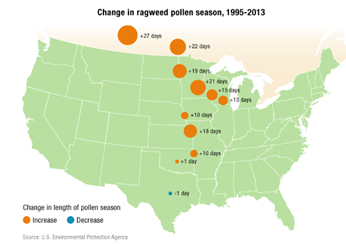
**Hardiness zones move north:**

Changing temperatures have an impact on what farmers and gardeners can grow, of course. U.S. Department of Agriculture measurements show, for example, that zones delineating what plants might survive the coldest temperatures in a given area have moved north. Plants that once worked no farther north than Iowa are now successful in southern Minnesota.



**Allergy season lengthens:**

Allergies have become annoying for a longer period each year, one of a number of health impacts that health experts worry about. The ragweed pollen season is three weeks longer in the Twin Cities than it was 20 years ago. The growth in the ragweed season has been even greater farther north. Warmer temperatures are a factor, but researchers also think some plants are producing more pollen because there is more carbon in the atmosphere.



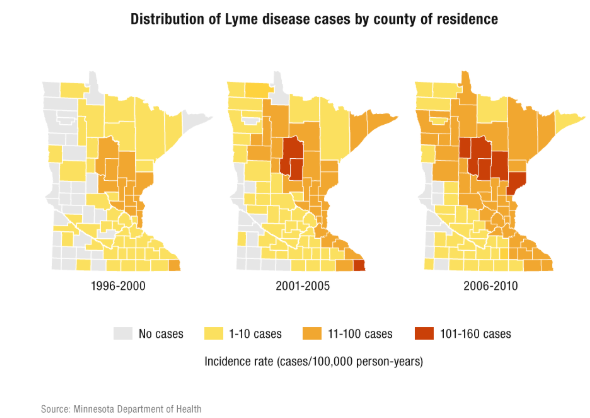
**Home, infrastructure damages rise:**

The insurance industry has pointed to increasing claims for hail and other storm damage in Minnesota. While this seems consistent with greater precipitation severity, some scientists express skepticism that it truly can be attributed to climate change.



**Lyme disease spreads:**

In the natural world, the changing climate translates into making some regions more or less hospitable to wild plants and animals. The area where Lyme disease has been reported has gotten bigger, for example. Lyme disease is a health problem known as vector-borne. The "vector" is the blacklegged (deer) tick, and it is showing up in wooded regions where it was not previously found.



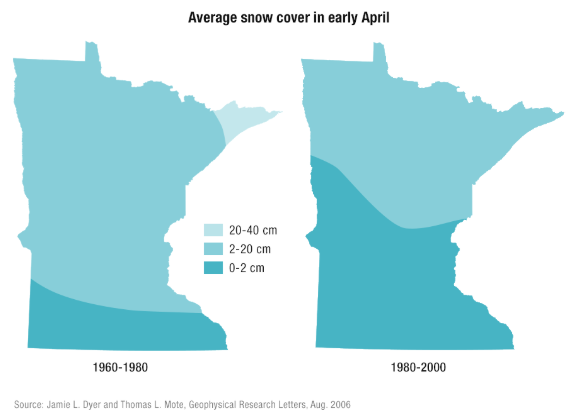
**Aspen, tamarack die:**

Aspen is the most widely distributed tree species in North America, but scientists have found a decline in aspen in this century in Minnesota and elsewhere. This apparently is the result of changing precipitation patterns, increased temperatures and insect defoliation. Also prompting concern is a decline in tamarack in northern forests. As with aspen, changing climate patterns have made tamaracks vulnerable. Those changes also have encouraged populations of the eastern larch beetle, which attacks the tamaracks.



**Snow season ends sooner:**

Snow tends to disappear earlier in the spring. Between 1960 and 1980, on the left, about three-quarters of the state was covered with two to 20 centimeters on average in early April. On the right, for the years 1980 to 2000, only about half the state had that much snow. One implication has been an apparent decline in Minnesota of lynx, which need deep snow to compete, and a subsequent increase in bobcats. Another is that land barren of snow is darker and tends to heat more quickly.



Many more examples of the negative impact of global warming on different real life phenomena can be found by performing a simple google search on internet. We are going to look at the issue of global warming from a different perspective. We will evaluate the global warming hypothesis using *Nenena Ice Classic Contest* data. The Nenana Ice Classic is an annual ice pool contest and one of the Alaska’s favorite guessing games held in Nenana, Alaska (<https://en.wikipedia.org/wiki/Nenana_Ice_Classic>) It is a fundraising event in which individuals attempt to guess the exact time the Tanana River ice will break up at Nenana. In 1917 workers assembled to build a bridge across the Tanana River near Nenana Alaska, however work could not begin because the river was frozen over. To kill time, the workers began betting on exactly when the ice would break up. A large tripod was erected on the ice and it was decided that when the ice began to break up so that the tripod moved a certain distance downstream, that would be the breakup time and the winner would be the one with the closest time. This occurred at 11:30 a.m. April 30, 1917, but this was only the beginning. Every year, this same routine is followed and the betting has become more intense. The tripod is hooked by a rope to an official clock which is tripped to shut off when the rope is extended by the tripod floating downstream. There is even a Nenana ice classic web site: [**https://www.nenanaakiceclassic.com/**](https://www.nenanaakiceclassic.com/)

**Task:**

The data have some ecological value. One wonders if this unofficial start of spring, as the residents interpret it, is coming earlier in the year and if so, could it be some indicator of global warming? It is believed that 1960 is the year when global warming began.

**Data:**

The data in **Nenana1.txt** and **Nenana2.txt** (I downloaded the data for you from the NSIDC website: [**https://nsidc.org/data/nsidc-0064**](https://nsidc.org/data/nsidc-0064) **)** were popularized in the October 26, 2001 issue of ***Science*** in an article by Stanford professors Raphael Sagarin and Fiorenza Micheli. (If in case you are not familiar: ***Science*** publishes only the very best peer-reviewed and the most cited science research articles in the world)

**Questions:**

1. The data in file **Nenana1.txt** is stored in the following format:

Text

Description automatically generated

The current data format is not ready to be analyzed. Data cleaning and some preparatory steps need to be performed first. Working with date/time variables is not most pleasant task either as there are inconsistencies across different software. Since this is only our 2nd HW, I want help you with the data preparation steps that need to be performed before we start modeling. Your duty is to go over every each line of code and implement in Rstudio and try to understand. You will need to type and execute the same code I shared with you in the screenshots below. Note that I included a lot of explanations (comments after # sign in the code) as some of the functions are new to you. You don’t need to include them in your code submission if you do not want to type them, only commands will be graded.

Text, letter

Description automatically generated

Example that shows how to create date and time variables in R:

Text

Description automatically generated

Text, letter

Description automatically generated

1. Let’s define a KPI that represents the amount of time in days since January 1 and the time ice breaks each ear.  
   **Remark:** Working on a similar problem outside of class, you would need to come up with a meaningful KPI on your own. Understanding the questions and business objectives should help you to create a KPI. For global warming you may use actual temperature data and not this data set. Nevertheless, this is an interesting data set and more importantly it illustrates how one can be creative about solving a problem using data science tools.

Text

Description automatically generated

1. Create a graph of your KPI defined in the previous question and describe the pattern(s) you see in data.
2. Meteorologist who study weather patterns, hired your consulting firm to help them quantitatively assess the hypotheses of global warming. It had been known since 1960 (judgmental call) that humans were increasing the amount of heat-trapping greenhouse gases in the atmosphere that led to the questions: Does this human activity warm the climate noticeably? Judgmentally1960 has been selected to be a year of a long term temperature change(You may ask your parents or grandparents why 1960). Does the data provide statistically significant evidence of global warming? Answer this question by assessing the significance in the abrupt change in KPI.
3. Answer the previous question by assessing the significance of the gradual change in KPI.
4. Report and creatively interpret the estimates of both abrupt and gradual change in climate. Which one do you think it is more appropriate and meaningful?
5. You recall from intro to business statistics course that statistical significance ≠ practical importance. Suggest and implement a way(s) to quantify the evidence of global warming from a practical perspective. If you use terms such as “result is statistically significant or in/nonsignificant”, many people won’t have a clue what they mean and will ask you to put it Layman terms.
6. Compare your approach with the approach we used to assess the effectiveness of the batmobile program. Have you observed any differences? Briefly comment.
7. Read the original article published by two Stanford professors:   
   [**https://science.sciencemag.org/content/sci/294/5543/811.full.pdf**](https://science.sciencemag.org/content/sci/294/5543/811.full.pdf)

And also a Critique of the article and the comments below the critique:   
[**https://www.john-daly.com/nenana.htm**](https://www.john-daly.com/nenana.htm)

**Write a commentary on the above articles. (One paragraph is enough. But don’t limit yourself!)**

1. This summer I received an email from NASA National Snow and Ice Data Center, that the data set was update by including more data. The data set that contains the most recent information is in **Nenana2.txt**. Does the data set that includes more recent info provide statistically significant evidence of global warming? Answer this question by assessing the significance of both abrupt and gradual change in KPI. Does more recent that change your conclusions compared to the previous conclusions that you obtained using data in **Nenana1.txt**?

**Case 2**

**Background:**

Back in the 50s and 60s, people in Los Angeles breathed the dirtiest air in the world. It has been known to be plagued by a special air pollution problem. The problem comes from substances produced by chemical reactions in sunlight among some primary pollutants such as oxides of nitrogen and reactive hydrocarbons. The products of these chemical reactions are responsible for the notorious Los Angeles smog, which causes such health hazards as irritation and lung damage. One measured product that is indicative of the degree of this photochemical pollution is ozone, often denoted as O3. Ozone is one of many elements found in the stratosphere, a region of the atmosphere located between 10 and 50 km above the earth's surface. Stratospheric ozone plays an important role in the life cycle on earth due mainly to its ability to absorb the harmful ultraviolet radiation from the sun and to prevent most of it from reaching the surface. To ease the air pollution problem, different methods, interventions, were instituted, including the diversion of traffic in early 1960 by opening of the Golden State Freeway and the inception of a new law (Rule 63) that reduced the allowable proportion of reactive hydrocarbons in the gasoline sold locally. Also after 1966 special regulations were implemented to require engine design changes in new cars in order to reduce the production of O3.

An interesting side note: It was through the study of the effects of the above these events on the pollution problem that Box and Tiao introduced the intervention analysis in 1975. I attached the article **Intervention Analysis with Applications to Economic and Environmental Problems.pdf** in case some of you want to glance over. The article is published in the journal of American Statistical Association, the Top #1 statistics journal in the world. With the help of one more professor, same authors published the same results in a different journal, Journal of the Air Pollution Control Association. I attached this article too **Analysis of Los Angeles Photochemical Smog Data A Statistical Overview.pdf**. I don’t expect you to read these articles. However they will provide you with much deeper knowledge than a 3 unit course.

**Task:**

Assess the significance of the interventions on controlling and reducing pollution.

**Data:**

File **dtwnLAozone.txt** contains monthly average of hourly readings of O3 in parts per hundred million (pphm) in downtown Los Angeles from 1955 to 1972.

**Questions:**

1. Import data.

I am going to help you with this answer:

Text, letter

Description automatically generated

1. Graph the data and describe the patterns you see in the data set?
2. Does the data provide statistically significant evidence that the opening of the Golden State Freeway and the implementation of Rule 63 in 1960 reduced the pollution statistically significantly? Evaluate both the abrupt and the gradual changes? Report and interpret the estimates of both the abrupt and the gradual changes in ozone. Which one do you think it is more appropriate and meaningful?
3. Does the data provide statistically significant evidence that the regulations implemented in 1966 requiring engine changes in new cars reduced the pollution statistically significantly? Evaluate both the abrupt and the gradual changes? Report and interpret the estimates of both the abrupt and the gradual changes in ozone. Which one do you think it is more appropriate and meaningful?
4. Can you capture both interventions, the opening of the Golden State Freeway and the implementation of Rule 63 in 1960, and the regulations for new car engines implemented in 1966 in one model? Can you model both interventions and both abrupt and gradual changes? What insights do you gain from such model? Summarize the results.