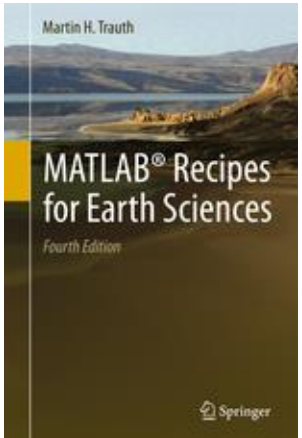
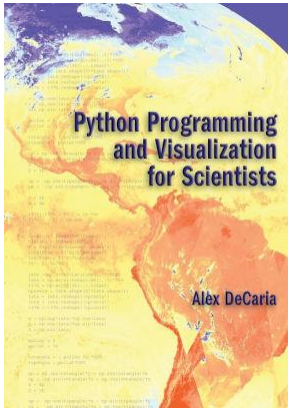
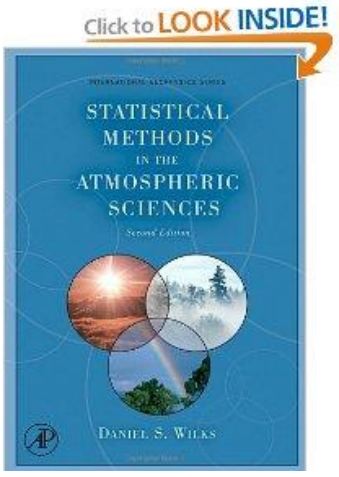
ATMOSPHERIC SCIENCES 5040/6040- Environmental Statistics

- Instructors: John Horel, Court Strong
- Grader: Chris Foster
- Required Text: *Matlab Recipes for Earth Sciences.*
- 4th edition is available as a 3-hour reserve through the library
- 3rd edition is available as a e-book (ok for undergrads)
- Recommended if interested in python: *Python Programming*
- Recommended text for 6040: Statistical *Methods in the Atmospheric Sciences*







Steps for Effective Research

- distill a general interest in a subject into a specific question/hypothesis that can be evaluated
- organize the data
- find relationship(s) among the data
- examine the significance of your results
- review thoroughly what you have done and document your analysis and results
- submit your results and study for independent evaluation

What you should have been doing...

- Read Chapter 1 Notes
- Read Chapter 1 in text
- Complete Assignment 2 based on Chapter 1 notes and text
  - Due last night

What you will be doing

- Class Wed and MW next week in MLIB 1110
- Read chapter 2 in text
- Read chapter 2 Notes
- In class Assignment 3 on Wednesday
- Complete Assignment 4 as really basic introduction to Matlab
  - Due 22nd

I'm taking this course because:

I don't want to but it is required.	2 respondents	7 %
I'm required to take it and I suspect it might be useful in my future career.	11 respondents	39 %
I expect it to be very useful for my future career.	20 respondents	71 %
I have no clue why I am here.		0 %

How confident are you with your level of understanding of basic Matlab commands?

I really know nothing about Matlab	2 respondents	7 %
I took a CS course but don't remember a thing	5 respondents	18 %
I might not know them well, but I know how to use matlab help to get more info	11 respondents	39 %
I use Matlab all the time and feel very comfortable with it	2 respondents	7 %
I took ATMOS 5020, but still don't feel very confident about using Matlab	14 respondents	50 %
I took ATMOS 5020 and now feel confident using Matlab	2 respondents	7 %

How confident are you with your level of understanding of basic Python commands?

I really know nothing about Python	6 respondents	**21 %**	
I took a CS course but don't remember a thing	3 respondents	11 %	
I might not know them well, but I know how to use the web to get more info on python	2 respondents	7 %	
I use python all the time and feel very comfortable with it	5 respondents	18 %	
I took ATMOS 5020, but still don't feel very confident about using python	10 respondents	36 %	
I took ATMOS 5020 and now feel confident using python	2 respondents	7 %	

How comfortable are you working with computers to solve problems?

It scares me a lot to even try		**0 %**
I think I'll be ok if I understand what is being asked	15 respondents	54 %
I just want to get the assignment over with and will look for the fastest way to do so.	3 respondents	11 %
I enjoy the challenge of working through and solving tough problems	13 respondents	46 %

What level of statistics have you been exposed to?

What's statistics?	6 respondents	21 %	
I like to compute on base percentages before going to bed	1 respondents	4 %	
I had a statistics class in high school	13 respondents	46 %	
I have taken one or more statistics classes in college	11 respondents	39 %	

My perception of statistics is:

something really boring, really really boring	4 respondents	**14 %**
likely boring but so are most of the other classes I have to take	1 respondents	4 %
going to be somewhat useful for my career	7 respondents	25 %
going to be very useful for my career	23 respondents	82 %

What level of exposure have you had to analyze environmental data?

Never, only canned text book stuff		**0 %** ✓
Some in a really cool class- ATMOS 5020	14 respondents	50 %
Frequently as an undergrad	13 respondents	46 %
Frequently as a grad student	9 respondents	32 %

My perception of analyzing environmental data is

really really boring	1 respondents	4 %
something I might have to do but I won't like it		0 %
I expect I better learn how to do it	16 respondents	57 %
I really want to learn so I can get a good job in my field	25 respondents	89 %

How do you perceive the use of statistics in society? For example, how much attention do you pay regarding statistical studies on health and diet?

Ignore all that stuff	3 respondents	**11 %**	
Read headlines but don't do much more than that	15 respondents	54 %	
Worry about it but don't change my behavior	8 respondents	29 %	
I take it very seriously and adjust my behavior	7 respondents	25 %	

How do you perceive the use of statistics in relation to the climate change "debate"?

what debate?	8 respondents	**29 %**	
there are two sides and everyone can make up their own mind	6 respondents	21 %	
statistics really doesn't play much of a role as far as I can tell- it's all about climate models	4 respondents	14 %	
people on both sides of the debate really abuse statistics	13 respondents	46 %	

What do you want to get out of this course?

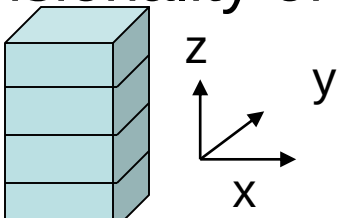
A good grade	23 respondents	**82 %**
some understanding of statistical tools	23 respondents	82 %
how to analyze environmental data	26 respondents	93 %
determine statistically a cure for cancer	4 respondents	14 %

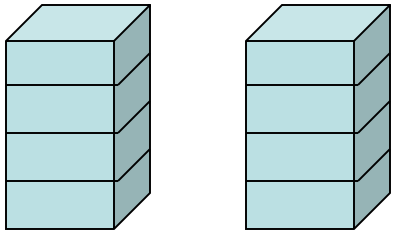
Common Goal is to Synthesize and Reduce Dimensionality

- Statistical analysis of environmental data typically involves reducing the dimensionality of the data to a manageable size.
- Which variable(s) do we need to consider?
- Can we consider one variable (univariate analysis) or must we consider multiple variables (multivariate analysis)?
- What time scales are we interested in? Hours, days, months, years? And, what region (local, regional, globally) or level in the vertical (surface, subsurface, upper air)?
- Are the data available on a spatial grid or at specific points?

Large Dimensionality of Geophysical Data Sets

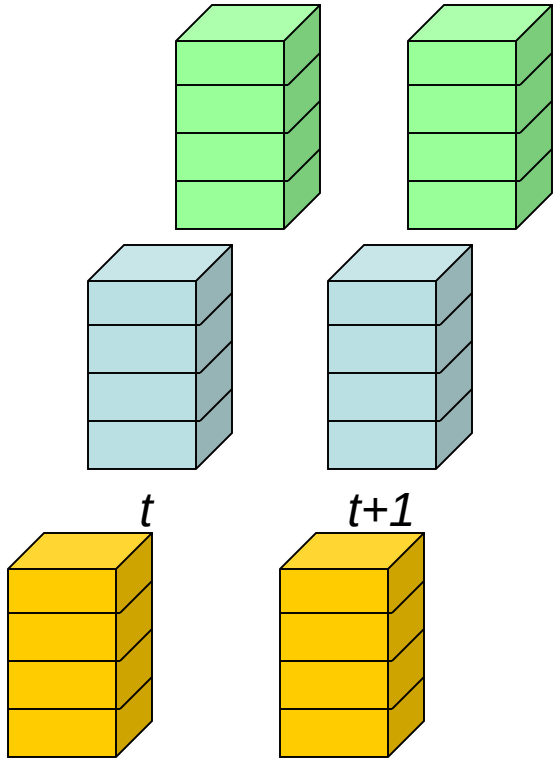
•Space: x, y, z





•Time: time (t) and forecast time (t_f)

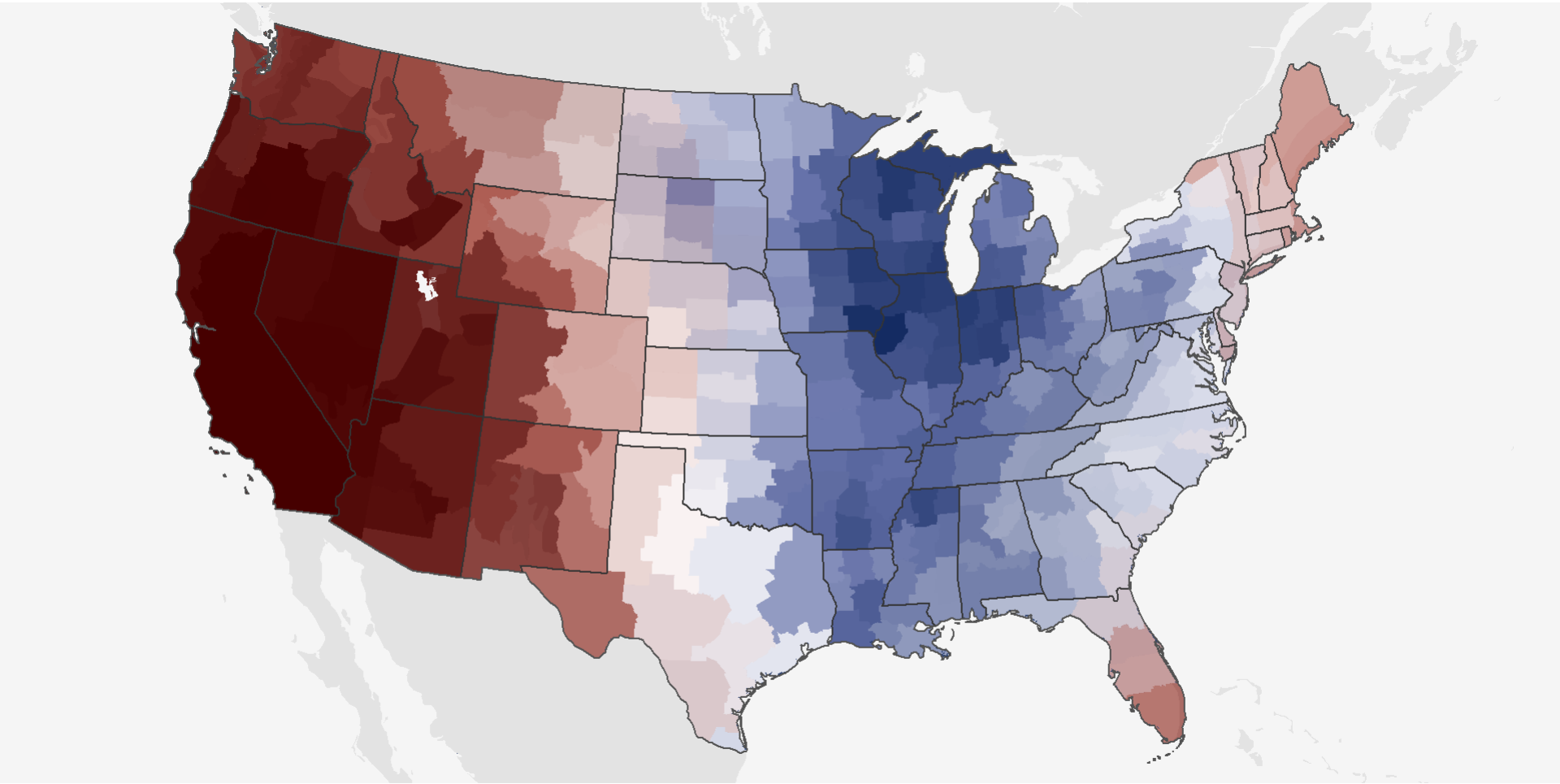
•Parameter & Source (temperature, winds, different models, measuring systems, perturbations)



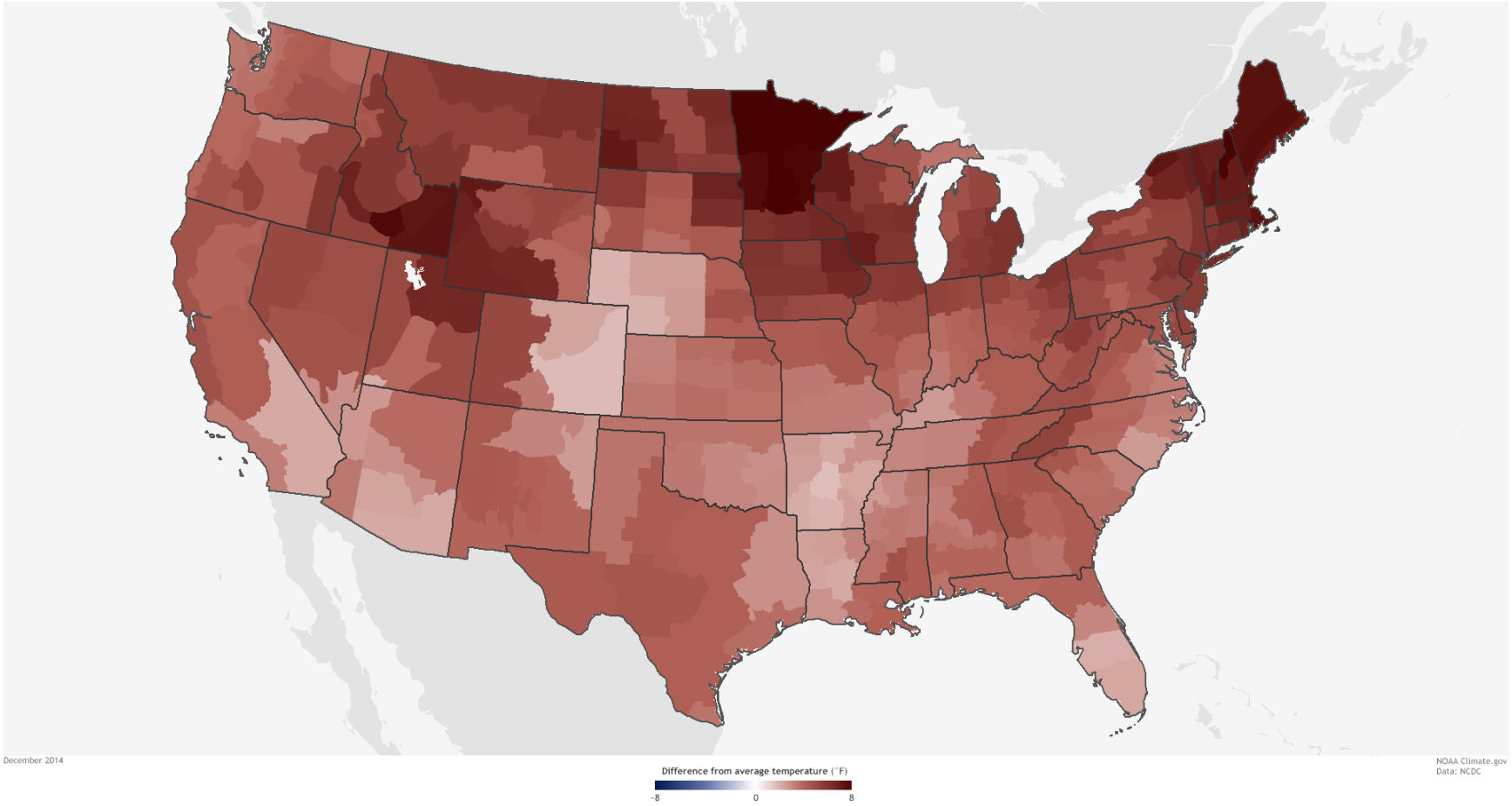
Climate

- Statistics often misinterpreted as bookkeeping.
  - What is the warmest temperature on record at Salt Lake City?
  - What is the biggest snow storm at Alta?
- weather and climate:
  - weather- state of the environment
  - climate- aggregate summary of the environment
  - Climate normal: arbitrarily defined reference state: 1981-2010

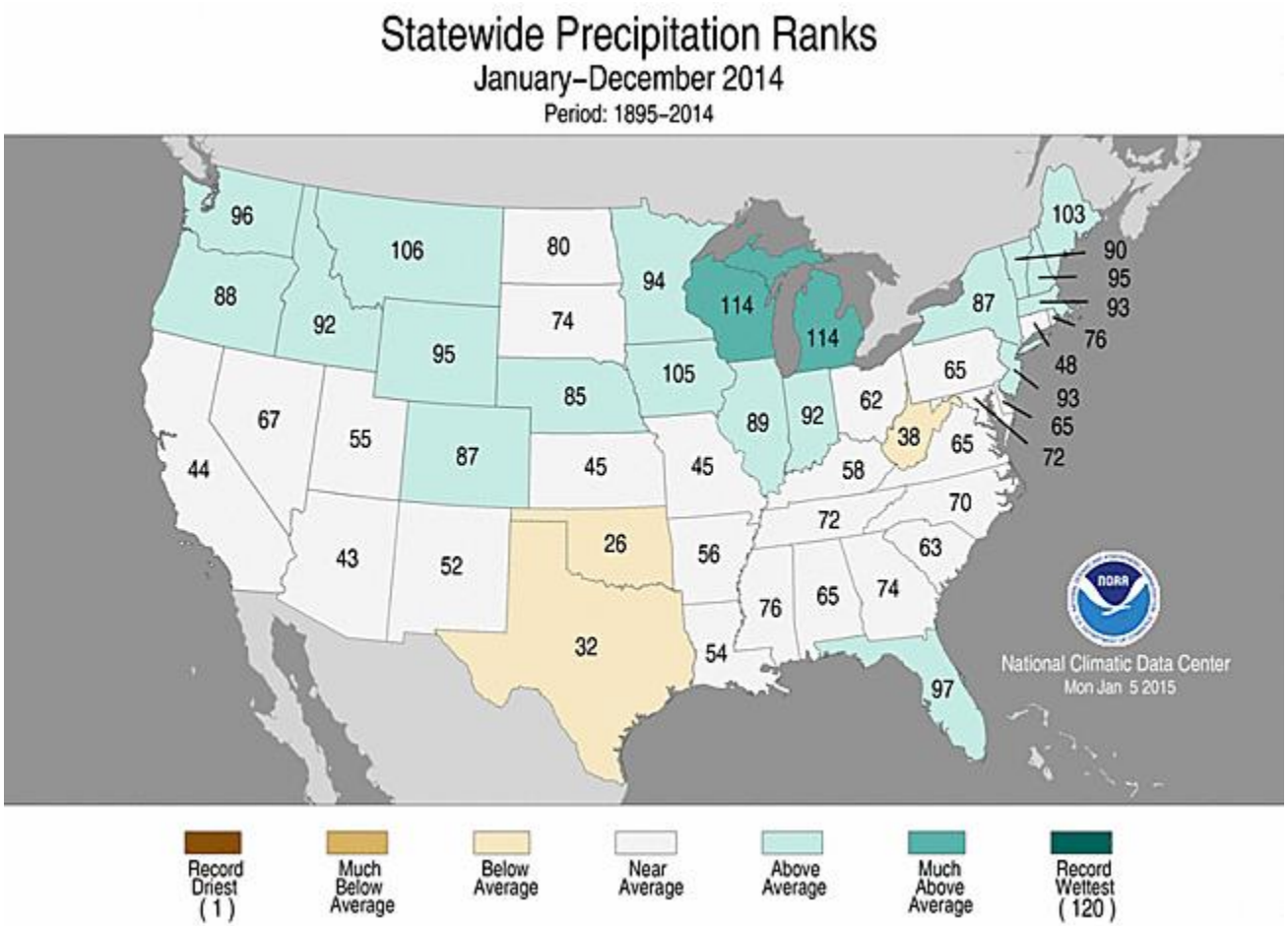
Climate.gov: 2014 temperature relative to climate normal



December 2014 temperature anomaly



Comparing 2014 to 1895-2013

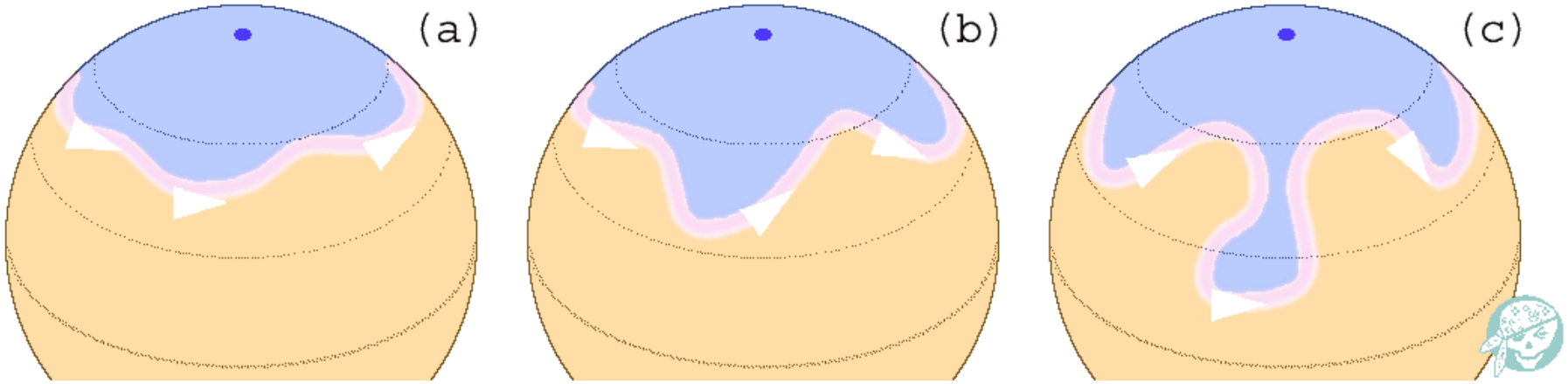


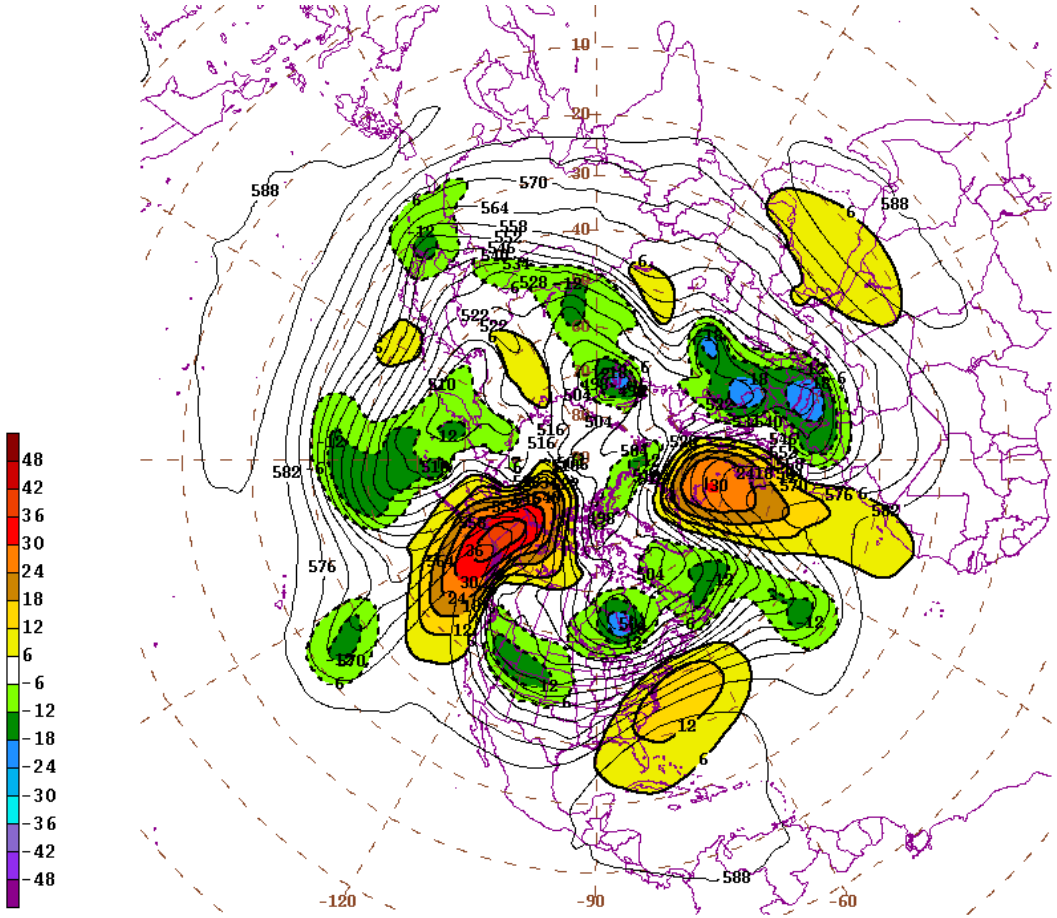
Why use statistics to describe the environment?

- Environment controlled by innumerable factors, which we hope to segregate into a few critical factors from the rest that, for the most part, simply contribute to background noise
- the characteristics of the system include linearly unstable processes such as baroclinic waves that cause growth of small features into larger ones
- the characteristics of the system (dynamics, thermodynamics) are nonlinear and include discrete step functions (i.e., rain/no rain) that can lead to the amplification of small errors into large ones
- the system is dissipative, which guarantees “stationarity”, i.e., the climate system will remain stable and not run away from the current state

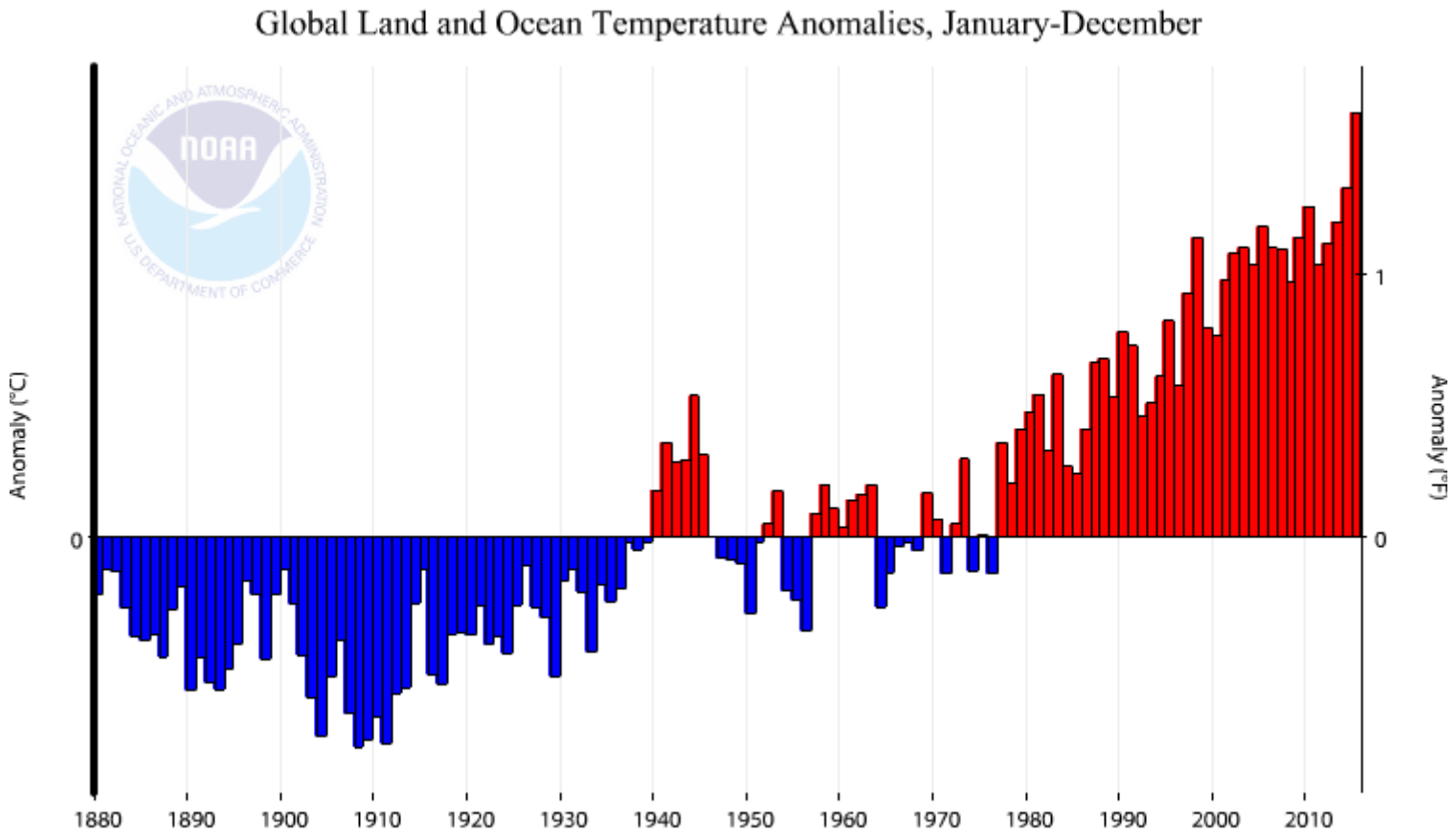
Baroclininc Wave Growth

- Physical processes in atmosphere can be unstable at times but environmental system overall is “stationary”

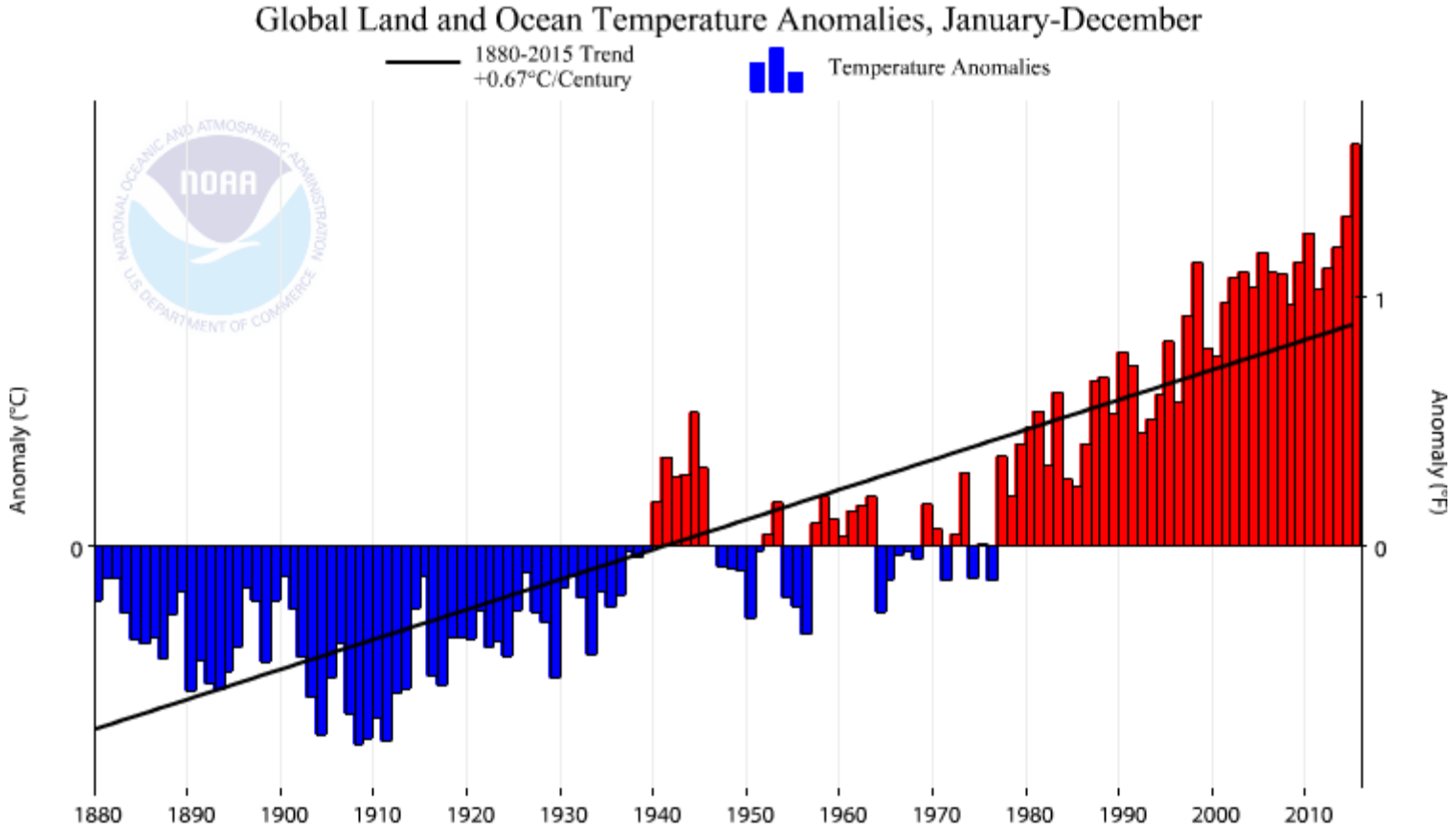


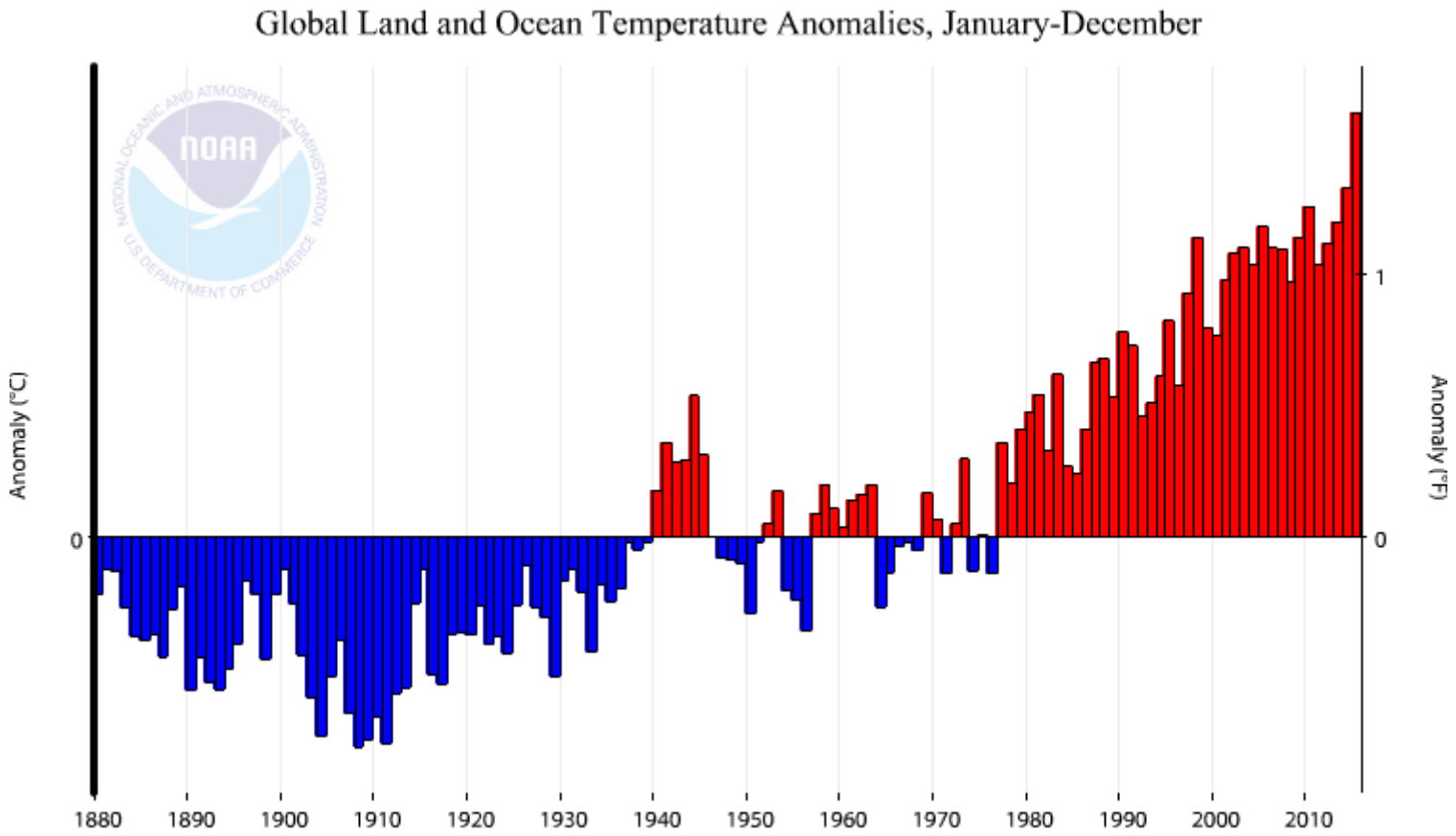


141229/0000F000 500 hPa Height and Height Anomaly (dam), from GFS



https://www.ncdc.noaa.gov/cag/time-series/global/globe/land_ocean/ytd/12/1880-2016

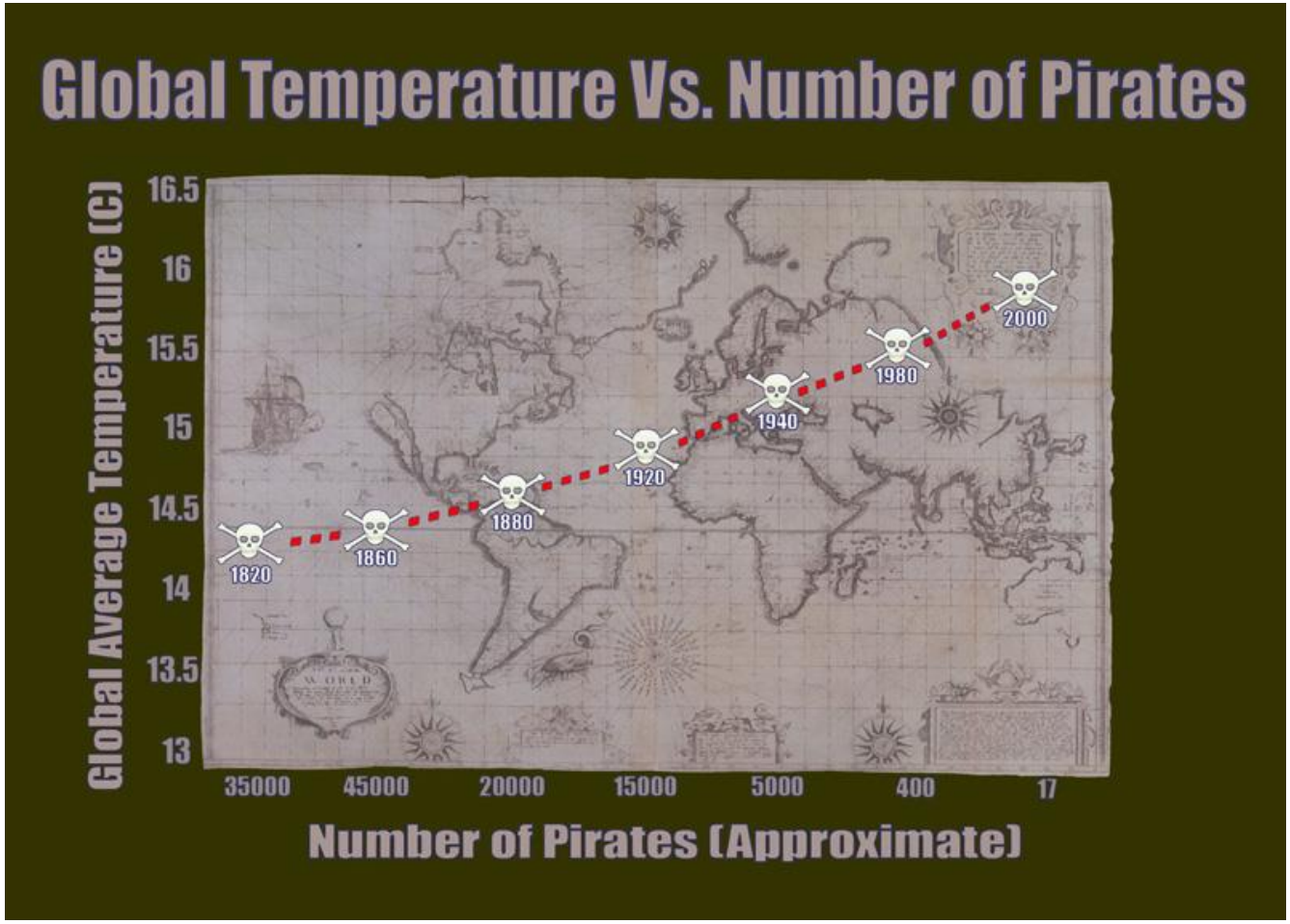

Global Land and Ocean Temperature Anomalies, January-December
1880-2015 Trend
+0.67°C/Century
Temperature Anomalies
Anomaly (°C)
Anomaly (°F)
0
1
0
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010



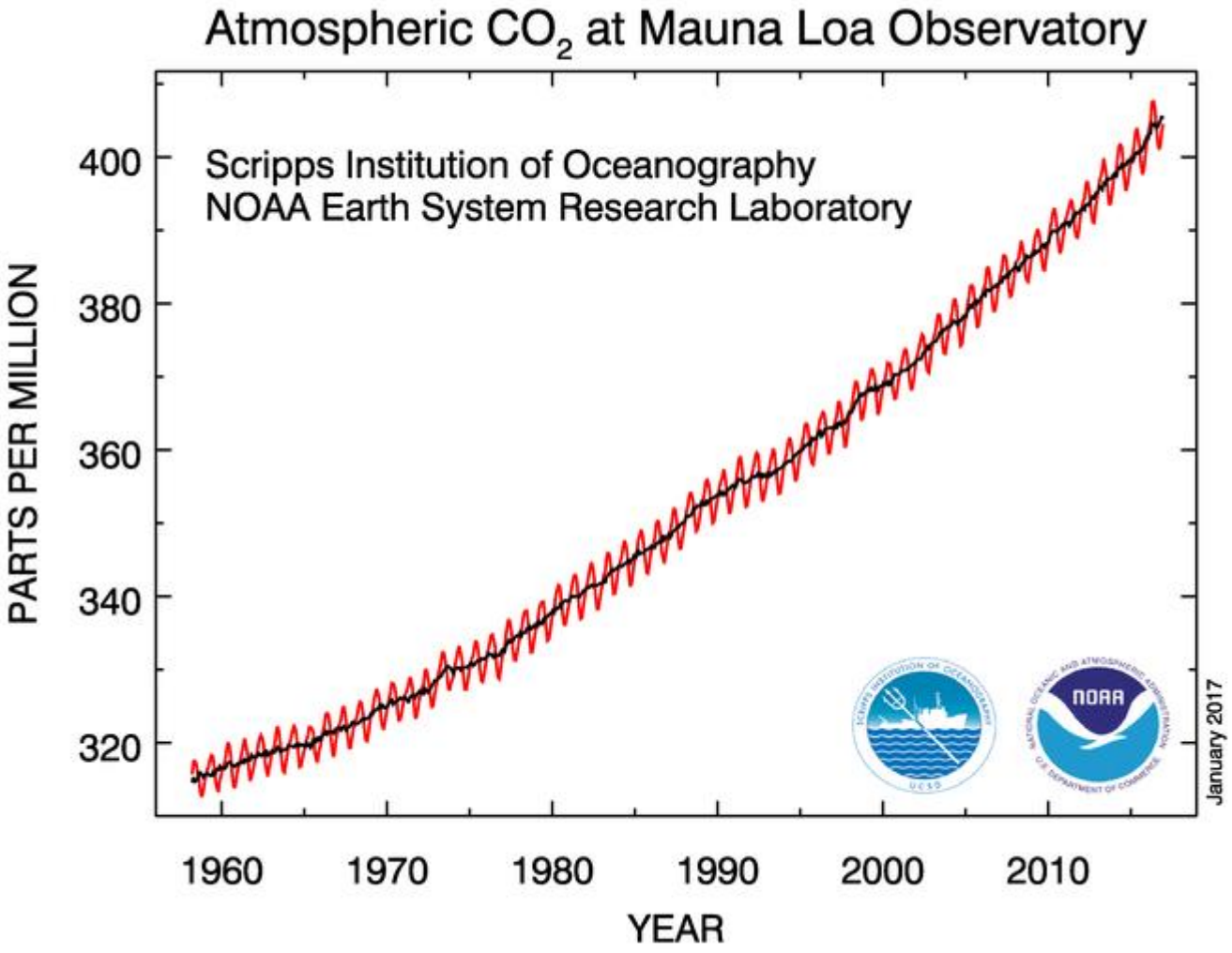
https://www.ncdc.noaa.gov/cag/time-series/global/globe/land_ocean/ytd/12/1880-2016


HOAX!

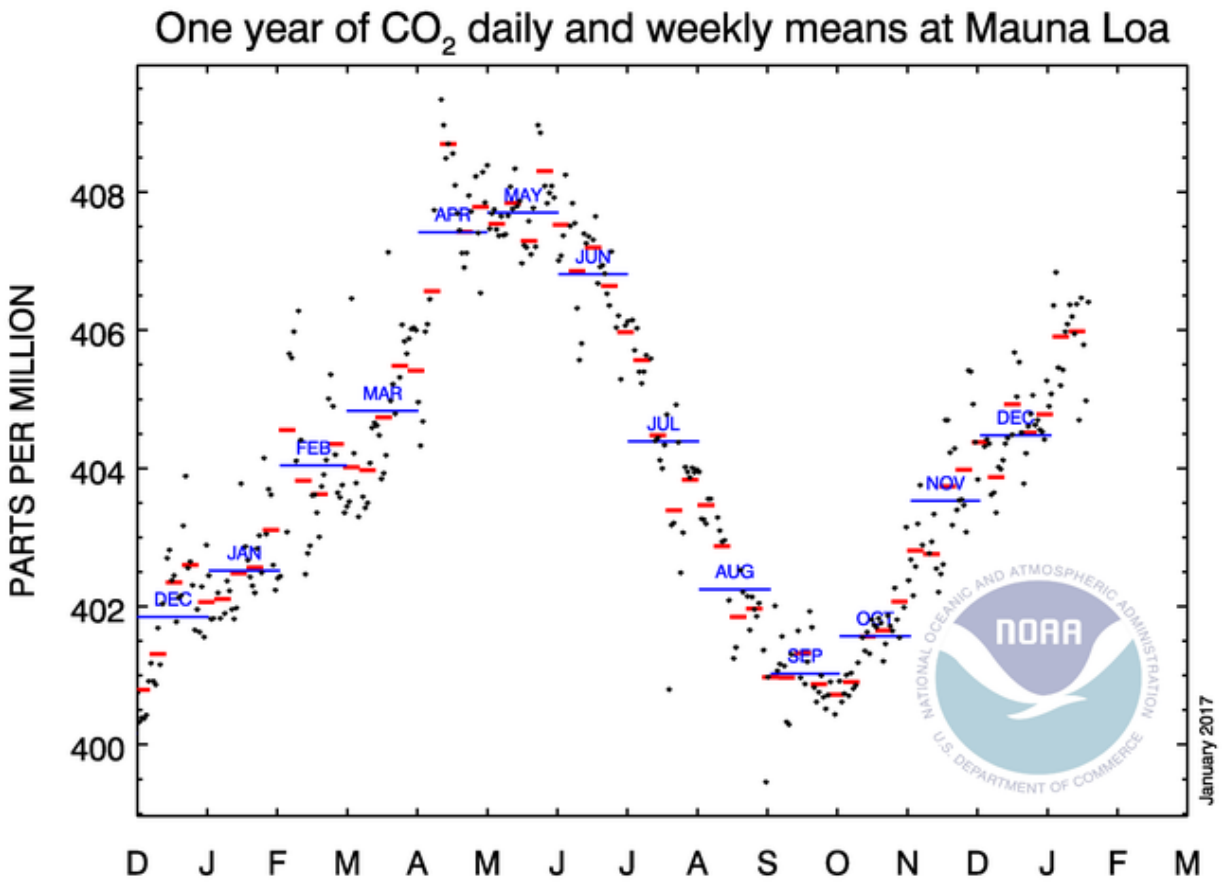
Causality vs. Simultaneity?



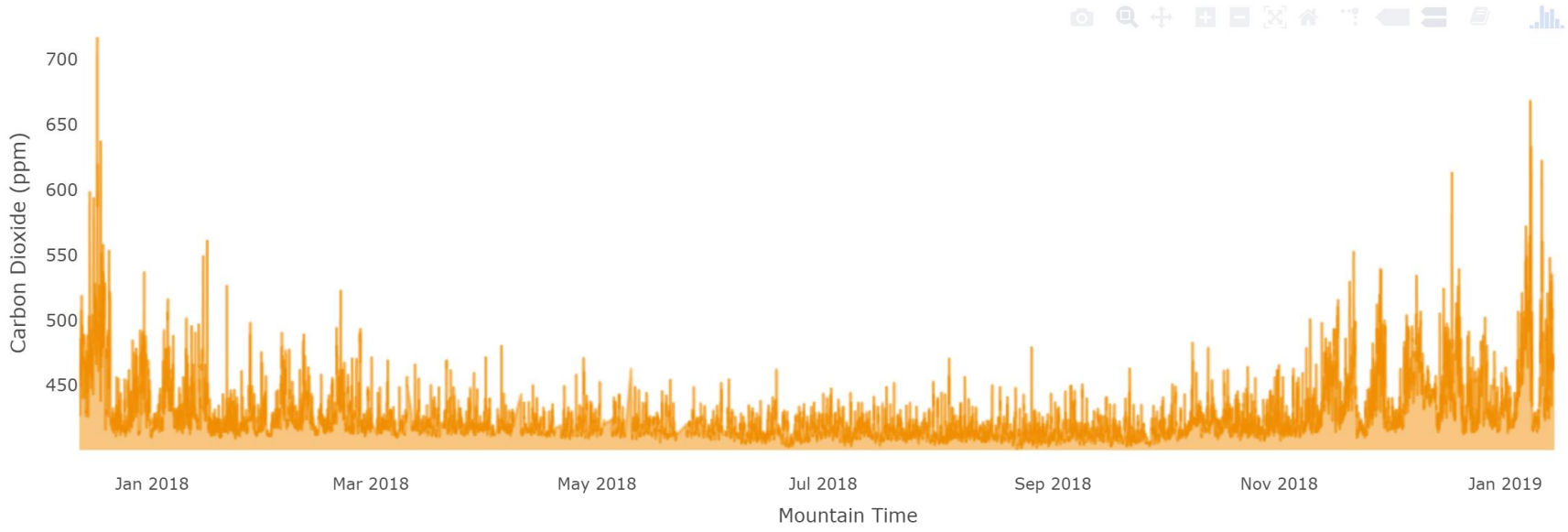
http://www.usnews.com/news/blogs/washington-whispers/2013/03/15/apollo-7-astronaut-uses-pastafarian-chart-on-pirates-and-global-temperatures-to-argue-climate-change-isnt-real



https://www.esrl.noaa.gov/gmd/ccgg/trends/full.html


One year of CO_2 daily and weekly means at Mauna Loa
PARTS PER MILLION
408
406
404
402
400
DEC
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
D J F M A M J J A S O N D J F M
NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE
January 2017

CO_2 on campus



http://co2.utah.edu/

Is consensus irrelevant?

[T]he work of science has nothing whatever to do with consensus. Consensus is the business of politics. Science, on the contrary, requires only one investigator who happens to be right, which means that he or she has results that are verifiable by reference to the real world. In science consensus is irrelevant. What is relevant is reproducible results.

—Michael Crichton, from “Aliens Cause Global Warming,” a lecture given at California Institute of Technology, Pasadena, California, January 17, 2003

Reproducibility

- The extent to which measurements or observations agree when performed by different individuals defines this important tenet of the scientific method.

Editorials

Irreproducible Experimental Results

Causes, (Mis)interpretations, and Consequences

Joseph Loscalzo, MD, PhD

+ Author Affiliations

Reproducible-Science-Curriculum

Irreproducible Examples

tnabtaf edited this page on Dec 11, 2014 · 18 revisions

Notes on examples of irreproducible research that has caused significant issues. Some of these will become case studies during the course. Initially, these are all from RetractionWatch.

Papers

- Editorial Expression of Concern: Non-adaptive origins of interactome complexity
- Retraction Watch Univ.: No misconduct, but “poor research practice” in mgt prof’s work now subject to 7 retractions
    - "The investigation was severely hindered by the near total unavailability of raw data files and statistical output files for the seven papers under study."
- Retraction Watch Data questions prompt retraction of PLOS ONE cardiovascular paper
    - The authors have indicated that the raw data for Figures 3, 5 and 6 are not

Reproducibility vs. Replicability

- Replicability is assessed by performing an experiment under exactly the same conditions at different times
- reproducibility is assessed by performing similar, but not identical, experiments at different times, in different locations, and under somewhat different experimental conditions.
- replicability reflects the technical stringency or precision of a specific experiment
- reproducibility reflects the fundamental accuracy of an experimental observation
- a precisely conducted experiment can be inaccurate, and an accurate experiment may be performed imprecisely—especially in biomedicine where many factors can account for irreproducible results

Statistics and Irreproducibility

- errors in design, data analysis, and data interpretation.
- Bias clearly plays an important role in promoting false-positive results.
- Reasons for bias:
  - lack of experimental balance leading to an impassioned belief in one particular experimental outcome clouding objectivity;
  - perceived pressure to publish for academic advancement or to enhance the likelihood of competing successfully for grant funding[7]; and the lack of appeal of negative (or neutral) studies in most high-impact journals.
- study illustrating the inverse relationship between the scientific hierarchy (physical sciences at the top, social sciences at the bottom, and biological sciences in between) and the publication frequency of “positive” results
- Random and systematic errors contribute as well

Impact of Rarity on Irreproducibility

Black swan (Taleb)

- outcome is theoretically reproducible, but so rare that a sufficient number of experiments needed to determine its frequency exceeds feasibility
- Owing to its inherent rarity, it is effectively impossible to explain away an irreproducible result on the basis of a rare, singular event for which no supportive evidence can be marshaled
- Examples of irreproducible events in environmental fields?

What a mess we're in…

- Over the past decade, the number of journal articles worldwide doubled from 1.1 to 1.9 million.
- Of these, 0.5 million articles are published in the field of biomedicine, for an output of 1400 articles per day.
- Given this volume alone, is it any wonder that an increasing number of articles are being published that contain irreproducible results?
- Further-more, with the growth in search technologies, a greater number of articles is being retracted over time: ~40/year in the late 1990s, ~300 in 2010, and ~400 in 2011.
- The reasons for these retractions include plagiarism, data manipulation (especially in figures), and proven data falsification;
- however, irreproducibility resulting from innocent causes may also be included in this pool of retracted publications without being recognized as such.
- With as many as 50% of all articles listed in PubMed never cited at all, one can conclude either that the work is of minimal significance and not worthy of further pursuit or that it has been pursued and could not be reproduced.
- The extent to which these two explanations account for this statistic has not been (nor cannot easily be) determined.

Taking Steps

- Nature
- Reproducibility, rigour, transparency and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not make it right, and just because it is not reproducible does not make it wrong.
- The guidelines recommend that journals include in their information for authors their policies for statistical analysis and how they review the statistical accuracy of work under consideration.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 515 > Issue 7525 > Editorial > Article

NATURE | EDITORIAL

Journals unite for reproducibility

Consensus on reporting principles aims to improve quality control in biomedical research and encourage public trust in science.

05 November 2014

PDF | Rights & Permissions

Reproducibility, rigour, transparency and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not make it right, and just because it is not reproducible does not make it wrong. A transparent and rigorous approach, however, will almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come from refutations and the objective examination of the resulting data.

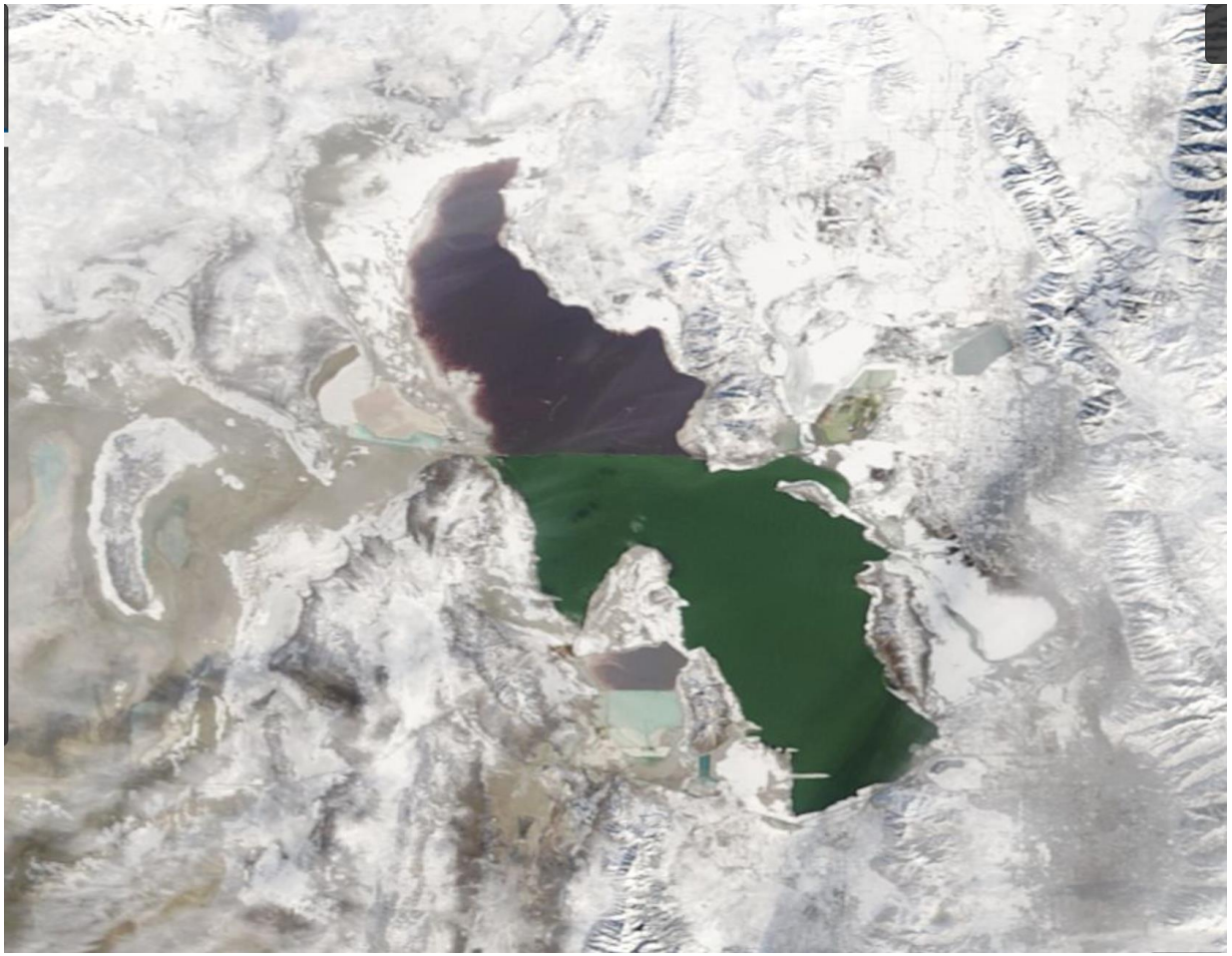
It was with the goal of strengthening such approaches in the biomedical sciences that a group of editors representing more than 30 major journals; representatives from funding agencies; and scientific leaders assembled at the American Association for the Advancement of Science's headquarters in June 2014 to discuss principles and guidelines for preclinical biomedical research. The gathering was convened by the US National Institutes of Health, *Nature* and *Science* (see *Science* **346,** 679; 2014).

Related stories

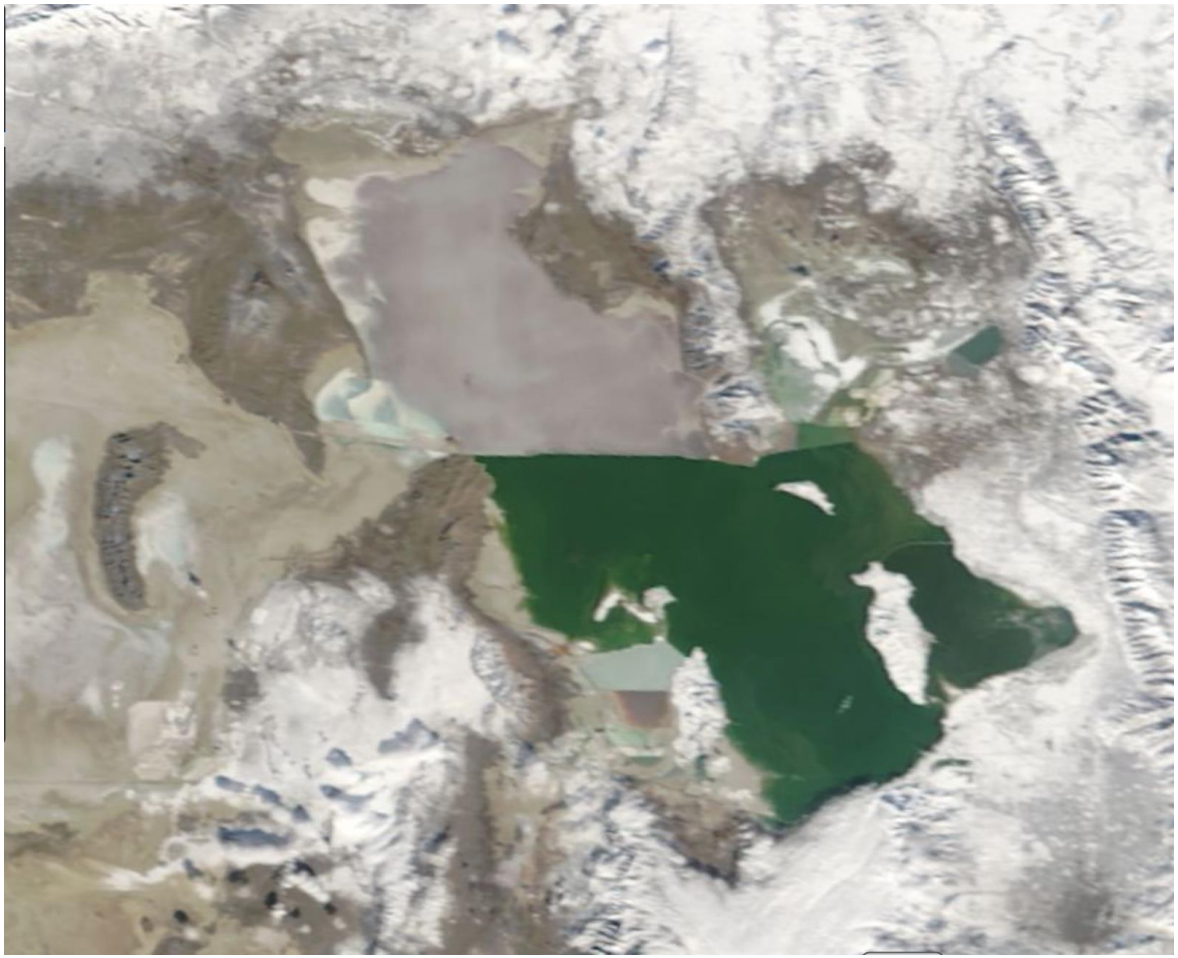
- Metascience could rescue the 'replication crisis'
- A blueprint to boost reproducibility of results
- Code share

More related stories

Great Salt Lake
January 9, 2019

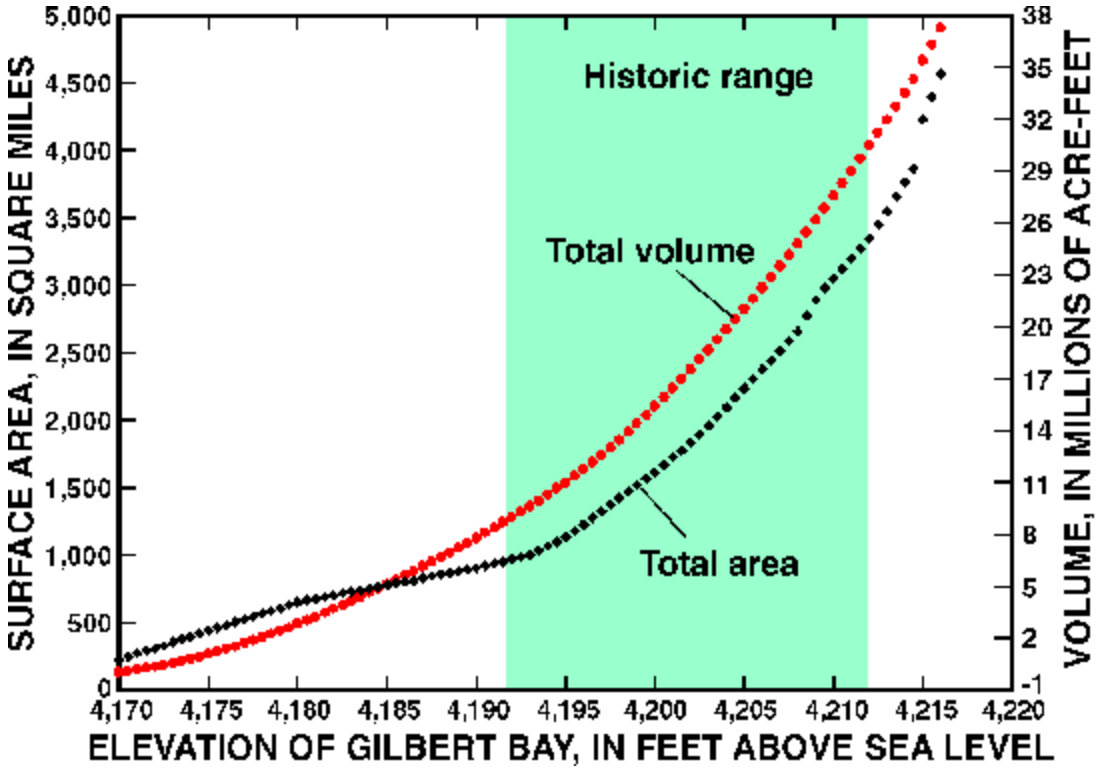


Great Salt Lake
December 25, 2000



Great Salt Lake

- http://ut.water.usgs.gov/greatsaltlake/elevations/



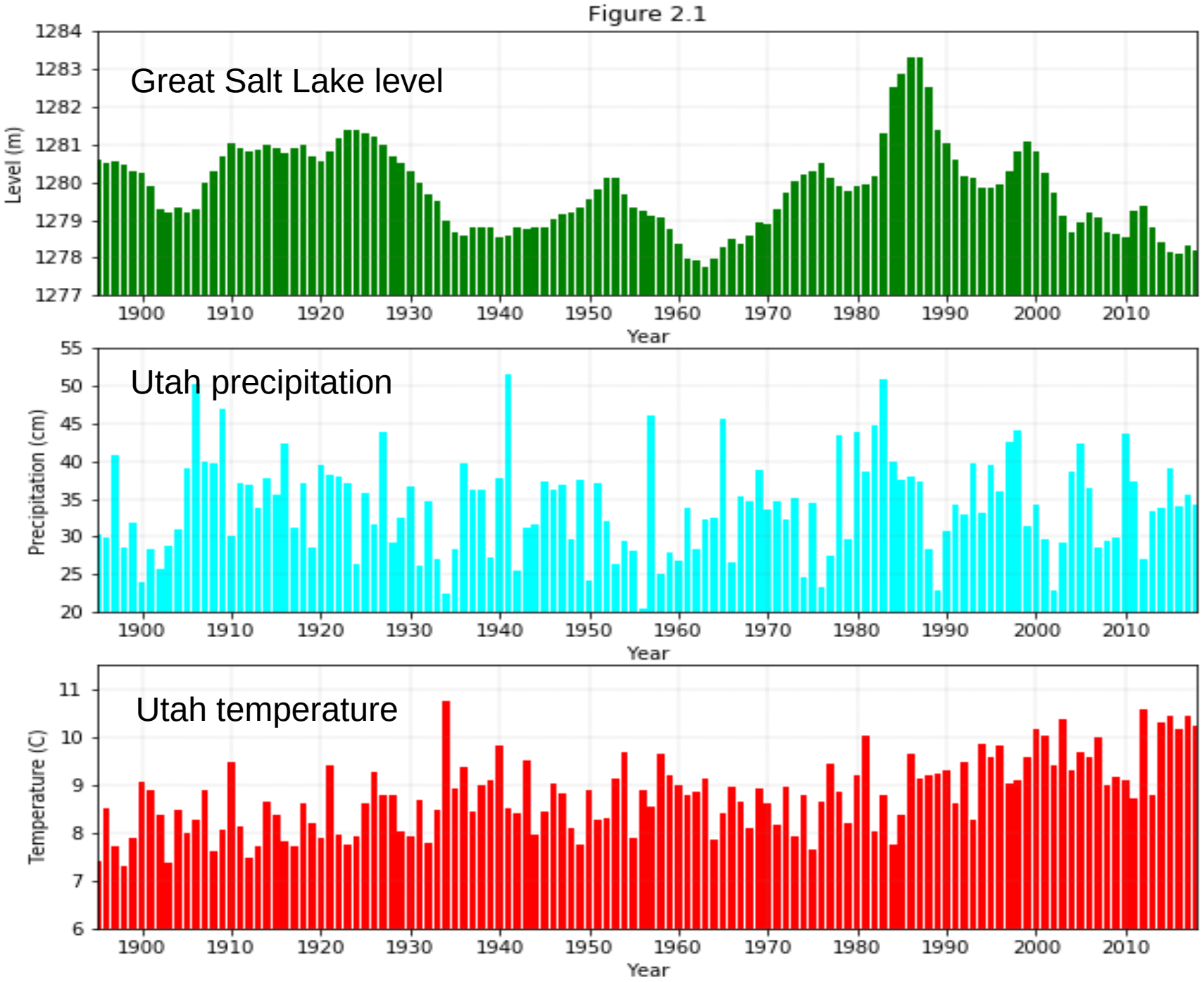
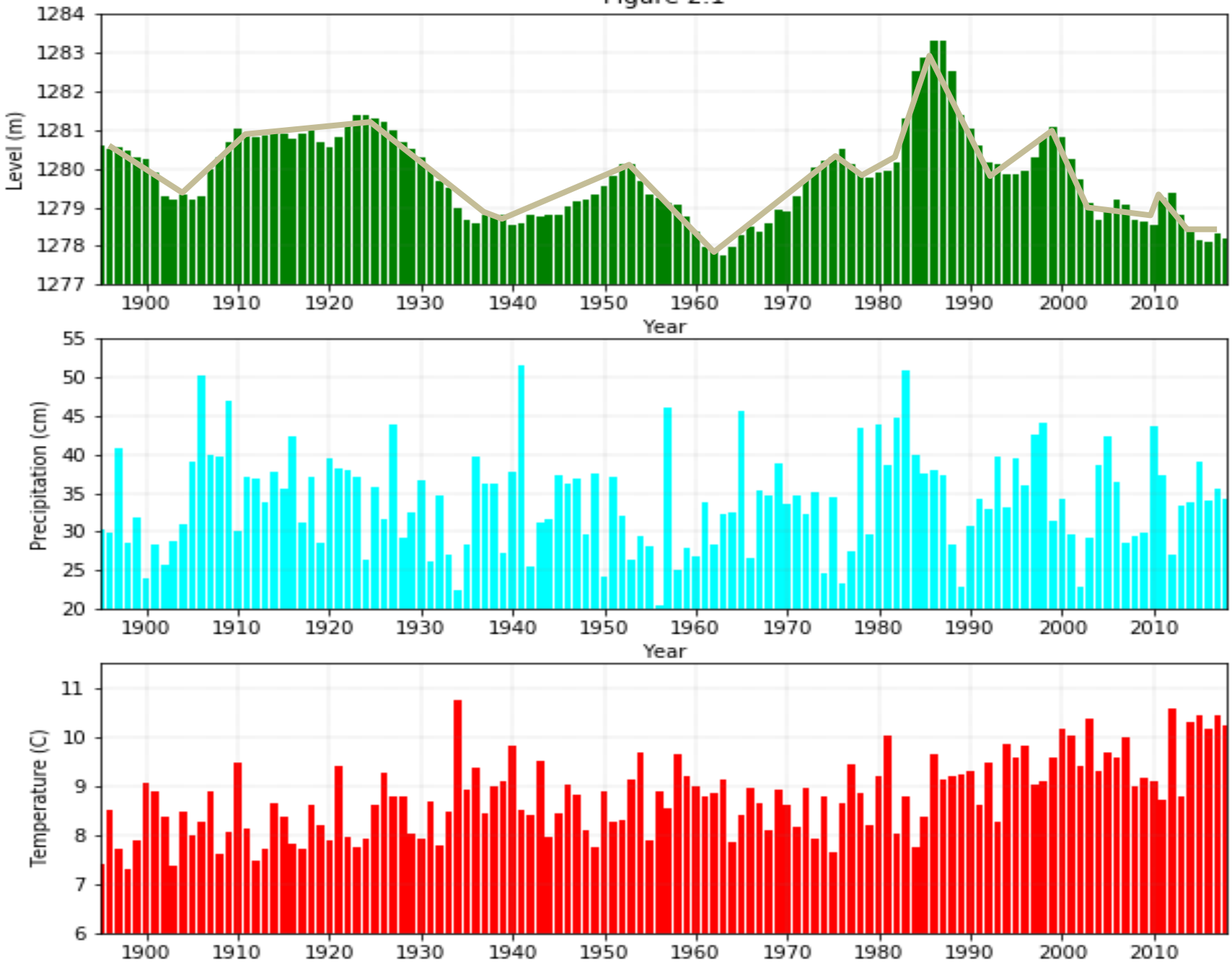


Figure 2.1
Great Salt Lake level
Level (m)
1284
1283
1282
1281
1280
1279
1278
1277
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year
Utah precipitation
Precipitation (cm)
55
50
45
40
35
30
25
20
Year
Utah temperature
Temperature (C)
11
10
9
8
7
6
Year

Figure 2.1


Level (m)
1284
1283
1282
1281
1280
1279
1278
1277
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year
Precipitation (cm)
55
50
45
40
35
30
25
20
Year
Temperature (C)
11
10
9
8
7
6

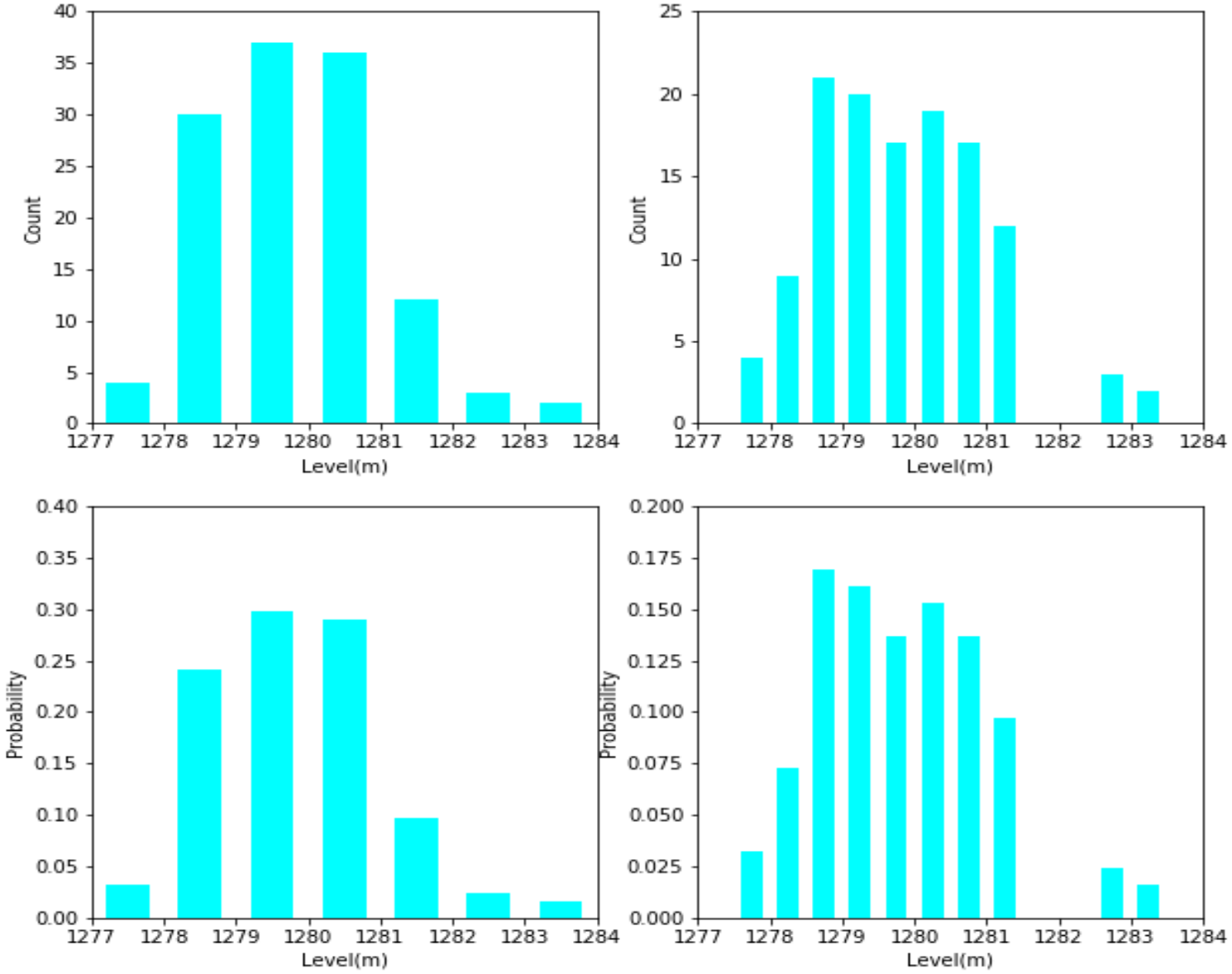


Figure 2.2. Histogram of lake level using 1m (upper left) and .5 m (upper right) intervals. Empirical probabilities that lake level falls within 1 m intervals (lower left) and within 0.5 m interval (lower right).

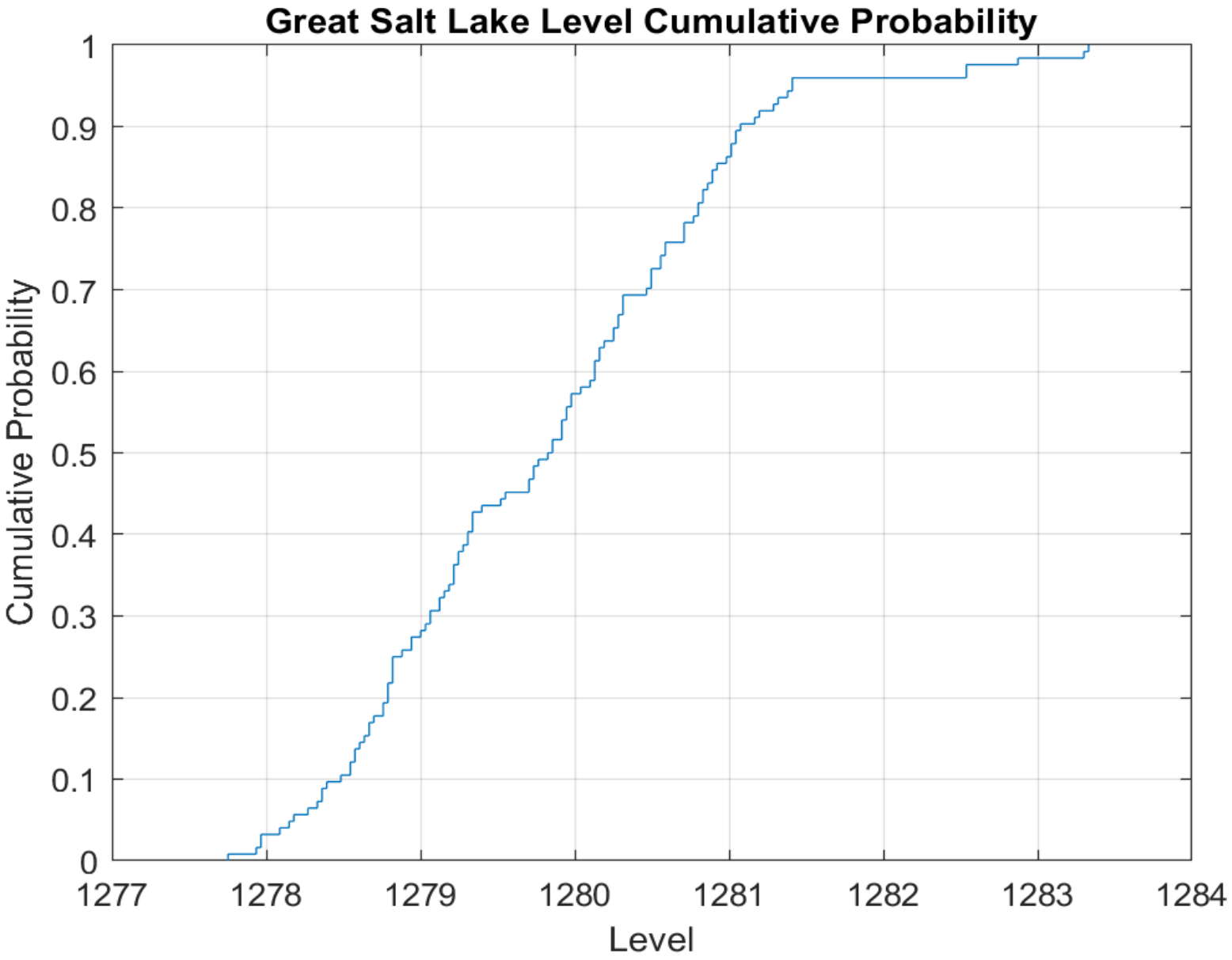


Figure 2.3. Cumulative frequency distribution of lake level.

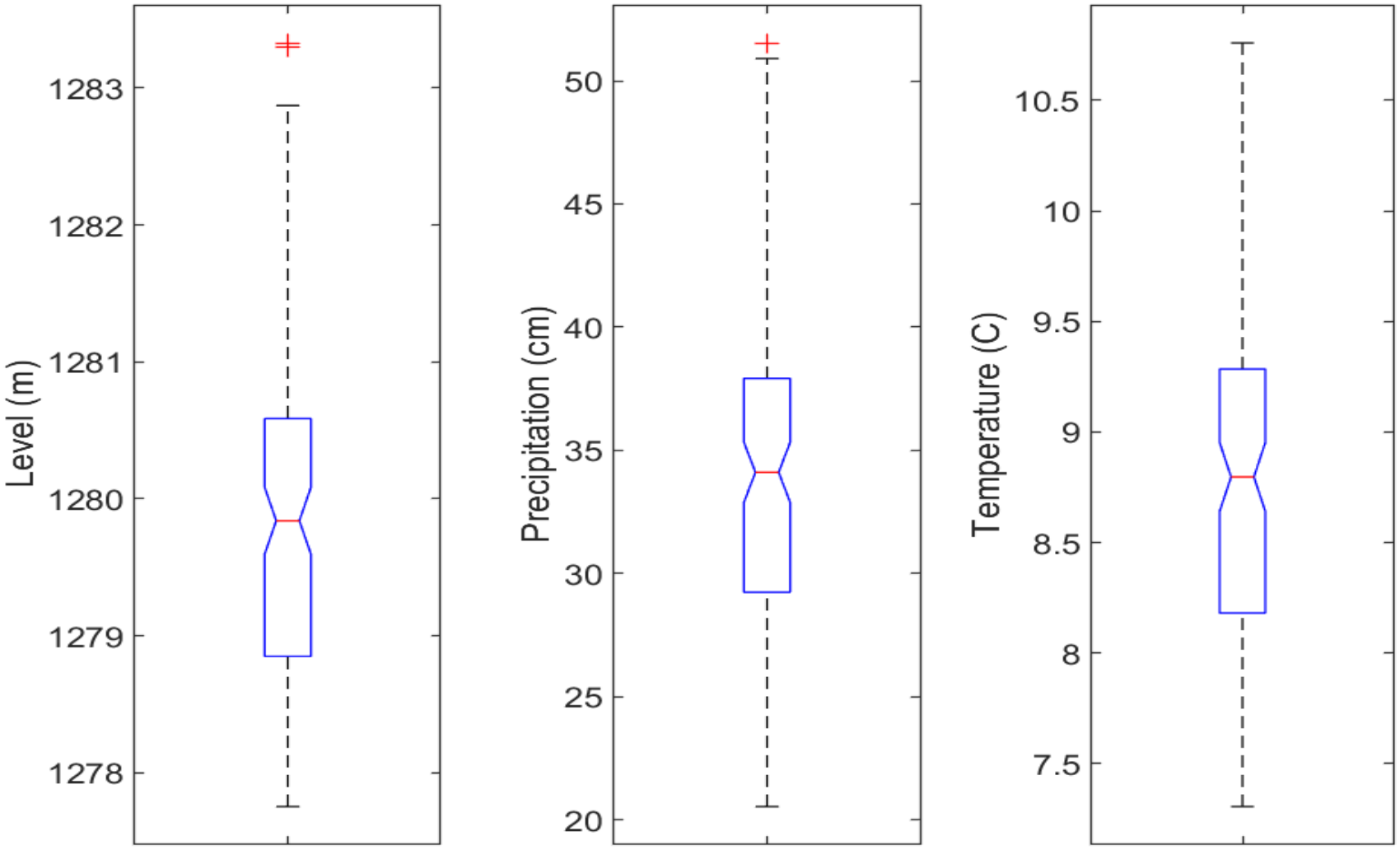


Figure 2.4. Boxplots of lake level (left), Utah precipitation (center), and Utah temperature (right).

What you will be doing

- Class Wed and MW in MLIB 1110
- Read chapter 2 in text
- Read chapter 2 Notes
- In class Assignment 3 on Wednesday
- Complete Assignment 4 as really basic introduction to Matlab
  - Due 22nd