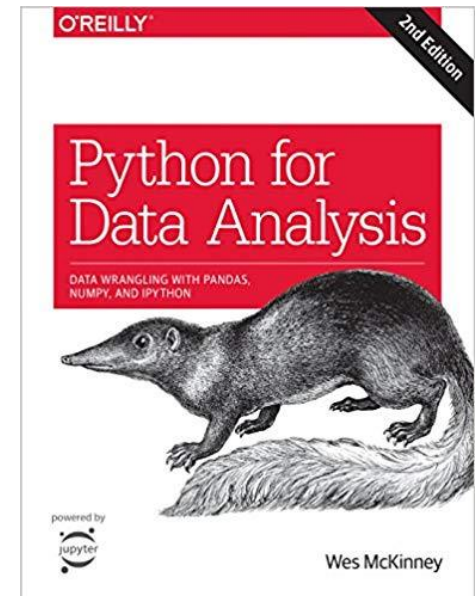
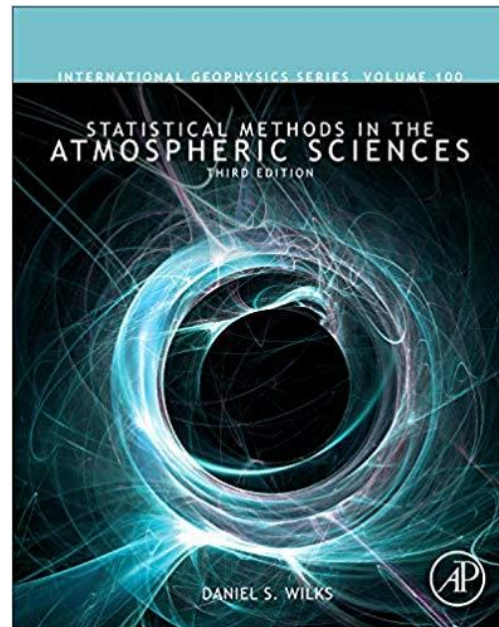
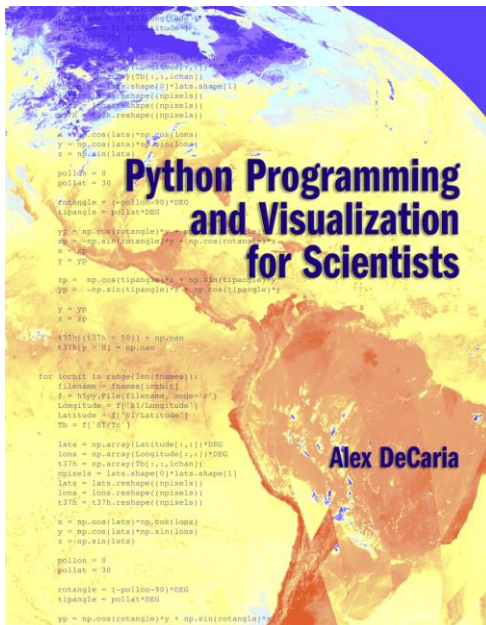


# ATMOSPHERIC SCIENCES

## 5040/6040- Environmental Statistics

- Instructors: John Horel, TA Aaron Meyer
- Required text for everyone: Python Programming ... A. deCaria
- Recommended text for graduate students: Statistical *Methods in the Atmospheric Sciences*



# Official Syllabus

- In Canvas
- Direct link to course:
- <https://utah.instructure.com/courses/573776/>
- Github
- [https://github.com/johnhorel/atmos\\_5040\\_2019/](https://github.com/johnhorel/atmos_5040_2019/)

# When you have questions...

- email: [john.horel@utah.edu](mailto:john.horel@utah.edu)
- Use the class Teams
  - You should have received an invite or use [the following link](#)
- Office hours: by appointment
  - Send email, a message in Teams or phone

# 5040 Tentative Course Outline

- Week 1. Oct. 15, 17. Basic concepts.
- Week 2. Oct. 22, 24. Python review.
- Week 3. Oct. 29, 31. Exploratory Univariate Data Analysis.
- Week 4. Nov. 5, 7. Exploratory Univariate Data Analysis.
- Week 5. Nov. 12, 14. Transforming Data.
- Week 6. Nov. 19, 21. Probability.
- Week 7 Nov. 26, Exploratory Multivariate Data Analysis.
- Week 8. Dec. 3, 5. Exploratory Multivariate Data Analysis.

# Format

- Meet in 703 WBB to start- we'll shift to 711 WBB at times
- You must read the assigned text and class notes prior to the corresponding lecture
- Assignments completed late will receive at most partial credit
- Take Home Exam due- December 11

# Class Policies and Grading

- 5040:
  - (1) final (33%)
  - (2) class assignments (66%)

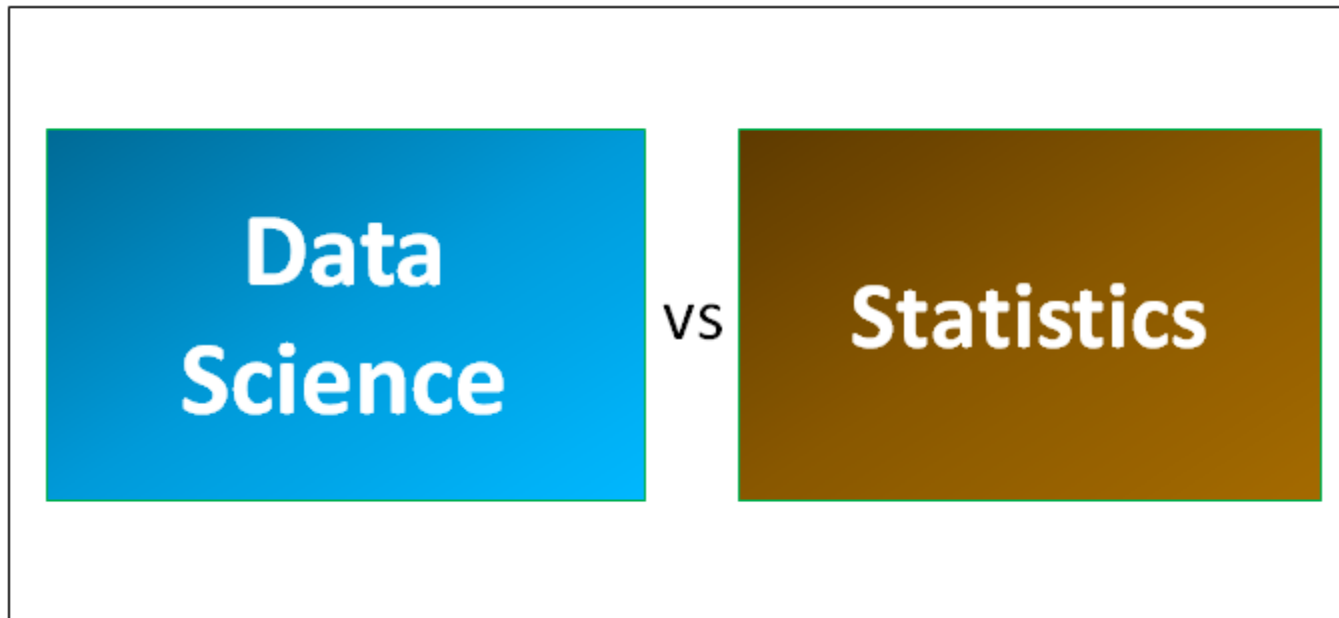
# Course Learning Objectives

- At the end of the short course students will be able to:
- State and use basic statistical metrics to analyze environmental information
- Develop proficiency to program using a high-level programming language, Python, to analyze environmental data sets
- State and demonstrate the characteristics of effective research: organize, quality control, and find relationship(s) among data

# What you should be doing

- Assignment 1- Survey- Due Friday
- Read Chapter 1 & 2a-2b Notes for Thursday
- Refamiliarize with using Mac lab and Python
- Assignment 2 due next week





- *Data Science* encompasses computer programming, statistics, and many other subjects.
- Involves collecting, preparing, analyzing, managing, visualizing, and storing large volumes of information
- *Statistics provides* methods and approaches to analyze information and draw conclusions from that information

# Programming Languages

- Programs written in a specific language are just variations on ways to pass instructions to a computer
- Each language has its own syntax (form) and semantics (meaning)
  - **Syntax:** Sequences of text including words, numbers, and punctuation using rules like written languages (grammar)
  - **Semantics:** The meaning given to the syntax
    - a sequence of words that makes sense to a computer

# Popularity of programming languages

<https://www.tiobe.com/tiobe-index/>

Oct 2019	Oct 2018	Change	Programming Language	Ratings	Change
1	1		Java	16.884%	-0.92%
2	2		C	16.180%	+0.80%
3	4	⬆	Python	9.089%	+1.93%
4	3	⬇	C++	6.229%	-1.36%
5	6	⬆	C#	3.860%	+0.37%
15	14	⬇	R	1.261%	+0.05%
16	20	⬆	Visual Basic	1.234%	+0.58%
17	12	⬇	Go	1.100%	-0.15%
18	17	⬇	Delphi/Object Pascal	1.046%	-0.11%
19	16	⬇	Perl	1.023%	-0.14%
20	11	⬇	MATLAB	0.924%	-0.39%



**vs.**

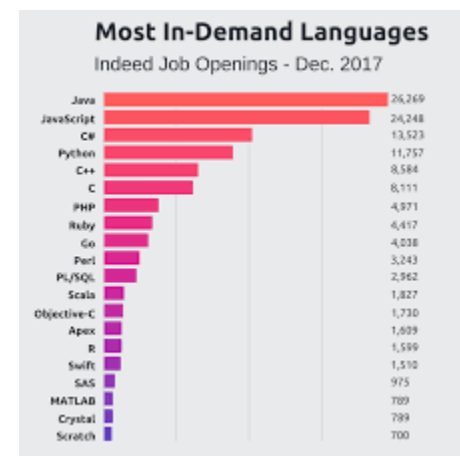


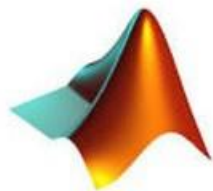
## Matlab

- Proprietary
- Inexpensive for educational uses, expensive otherwise
- Flexible graphical user interface
- Many toolboxes
- Matrix oriented

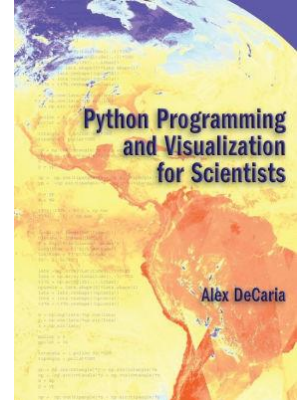
## Python

- Open source
- Free, although enterprise releases for commercial applications
- Notebooks are a convenient way to learn
- Many modules
- Object oriented





vs.



## In this course:

### Matlab

- Legacy codes available in Matlab
- Matlab text is useful

### Python

- All codes available in python
- Required that you do assignments in Python
- Intro text recommended

- Everything in one github repository (notes, codes, data):  
[https://github.com/johnhorel/atmos\\_5040\\_2019](https://github.com/johnhorel/atmos_5040_2019)
- Programming language reviews from ATMOS 5020 in repository
- Notes independent of language
- Use the class time to not feel pressured about completing assignments on time

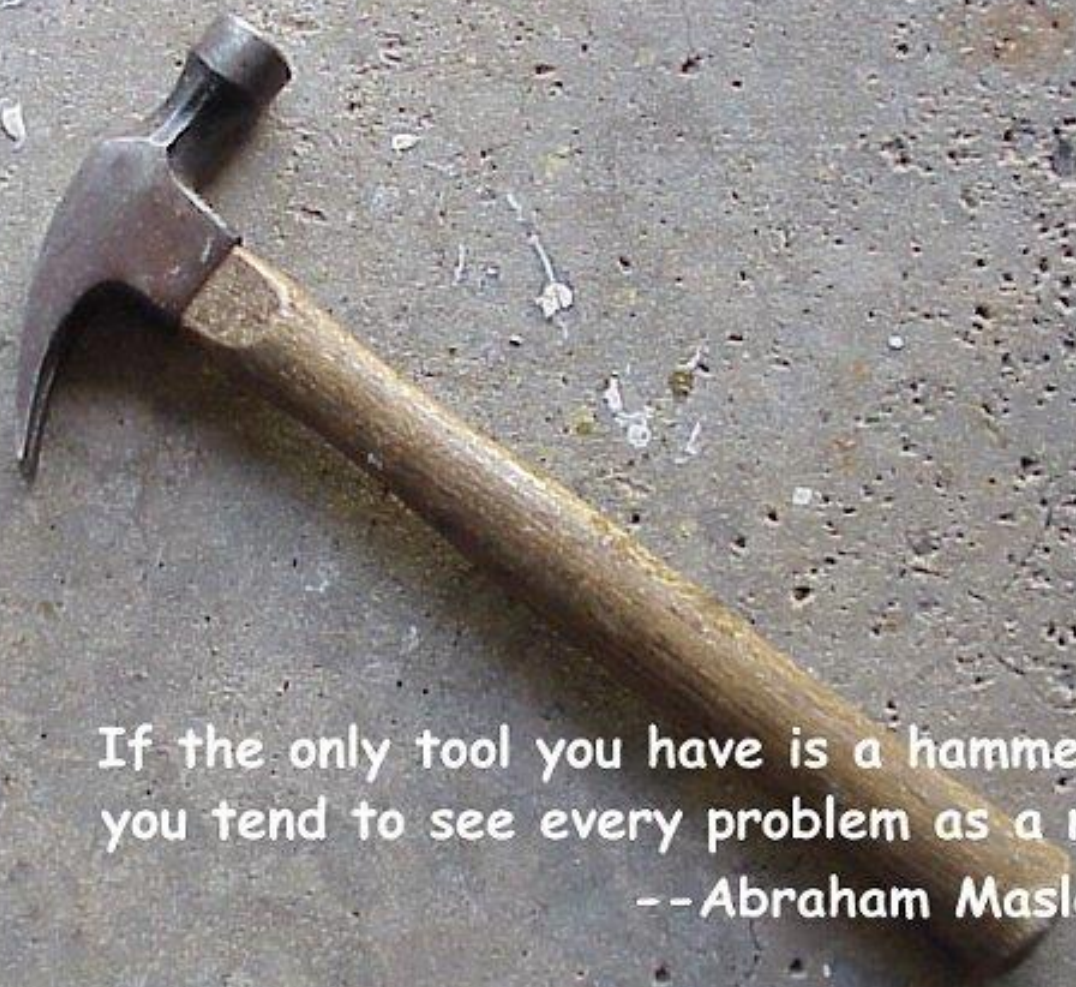
# Review Material

- [https://github.com/johnhorel/ATMOS\\_5020\\_2019/blob/master/Aug\\_20\\_2019.md](https://github.com/johnhorel/ATMOS_5020_2019/blob/master/Aug_20_2019.md)
- [https://github.com/johnhorel/ATMOS\\_5020\\_2019/blob/master/Sept\\_03\\_2019.md](https://github.com/johnhorel/ATMOS_5020_2019/blob/master/Sept_03_2019.md)
- [https://github.com/johnhorel/ATMOS\\_5020\\_2019/blob/master/Sept\\_05\\_2019.md](https://github.com/johnhorel/ATMOS_5020_2019/blob/master/Sept_05_2019.md)

# Problems with statistics?

- Oriented towards confirming preconceived ideas?
- Start with a technique and look for a data set to apply it to?
- There's always two sides to every issue?
- What do you think? Other examples of poor statistics?





If the only tool you have is a hammer,  
you tend to see every problem as a nail.  
-- Abraham Maslow



## The Health Risks of Being Left-Handed

*Lefties Face Chance Of ADHD, Other Disorders; Brain Wiring Holds Clues*

- <http://www.wsj.com/articles/SB10001424052970204083204577080562692452538>
- Modern lefty lore says left-handers are smarter, more creative and have an advantage over righties.
- About 10% of people are left-handed, Six of the last 12 U.S. presidents, including Barack Obama and George H. W. Bush, have been lefties.
- Babies born to older mothers or at a lower birth weight are more likely to be lefties
- On average there is no difference in intelligence between right-and left-handed people.
- Left-handed people earn on average 10% lower salaries than righties
- lefties aren't more accident prone than right-handed people and don't tend to die at a younger age.
- Left-handedness appears to be associated with a greater risk for a number of psychiatric and developmental disorders.
  - About 20% of people with schizophrenia are lefties even though 10% of people are lefties
  - 1% of general population has schizophrenia
  - Ooh! 80% of schizophrenics are righties! and 0.2% of all people are lefties with schizophrenia

# Questioning traditional statistical approaches

- **Nassim Nicholas Taleb**
- <http://www.foolledbyrandomness.com/>
- Black Swan: extreme impact of rare and unpredictable events and human tendency to search for simplistic explanations for these events retrospectively
- Antifragile: Some things benefit from shocks and thrive when exposed to volatility, randomness, disorder, stressors, and uncertainty

# Oddball statistics?

<https://www.statnews.com/2016/05/09/john-oliver-bad-science/>

Warning: some inappropriate language

John Oliver rips apart bad science on  
'Last Week Tonight'

By MEGAN THIELKING @meggophone / MAY 9, 2016



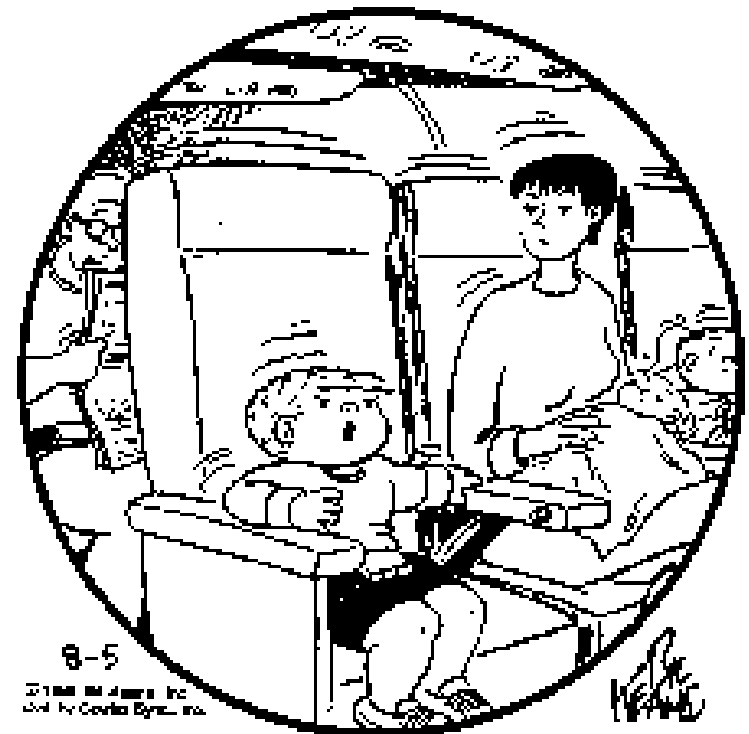
# Scientific Explanations are:

- based on empirical observations or experiments
- tentative
- historical
- probabilistic
- assume cause-effect relationships
- limited
- made public
- influenced by individuals and culture

## What's the Goal??

- Exploratory or descriptive statistics:
  - Organize and interpret volumes of data
- Inferential statistics:
  - Assess the underlying physical processes that generate environmental data

### THE FAMILY CIRCUS



"I wish they didn't turn on that seatbelt sign so much! Every time they do, it gets bumpy."

# Observations and Truth

- True value- value of a quantity sought through measurement, but unknown usually in the field
- Truth depends on application
- Assumption: **average** of many **unbiased** observations should be same as **expected value** of truth
- However, accurate observations may be biased or **unrepresentative** due to siting or other factors

<https://www.youtube.com/watch?v=9FnO3igOkOk>



# Rudy Giuliani says 'truth isn't truth'



By [Caroline Kenny](#), CNN

Updated 4:50 PM ET, Sun August 19, 2018

NBC News  
This morning

WASHINGTON, DC

NEW YORK

MEET THE PRESS

@RELIABLESOURCES

WHO'S "PANICKING?"

CNN

11:04 AM ET

0:42 / 1:30

MOI



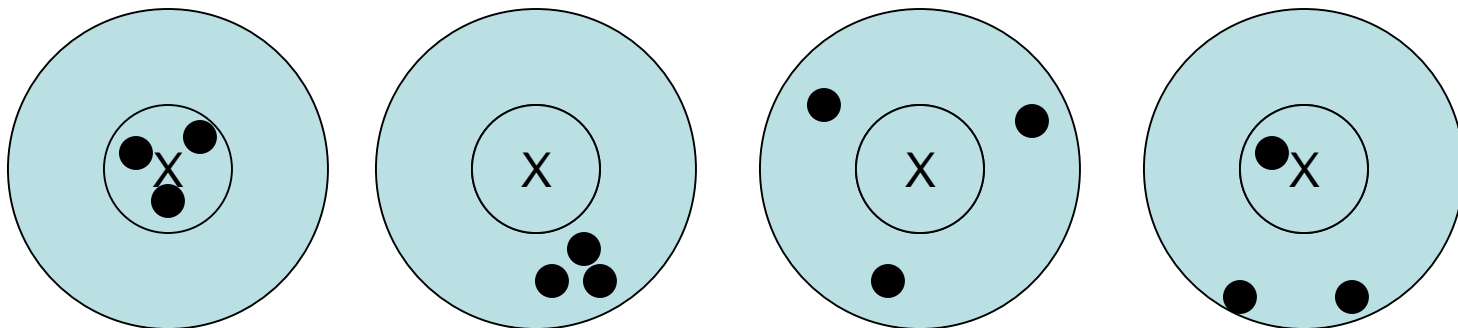
# Causes of Uncertainty

- 1. we can never measure the environment with complete accuracy and precision
- 2. the environment is a chaotic system, which is a maddening combination of randomness and order arising from the characteristics of a complex nonlinear system,
- 3. our understanding of the environmental system is imperfect, so physical (and certainly statistical) models do not capture the complete behavior of the system.



# Gauging Uncertainty

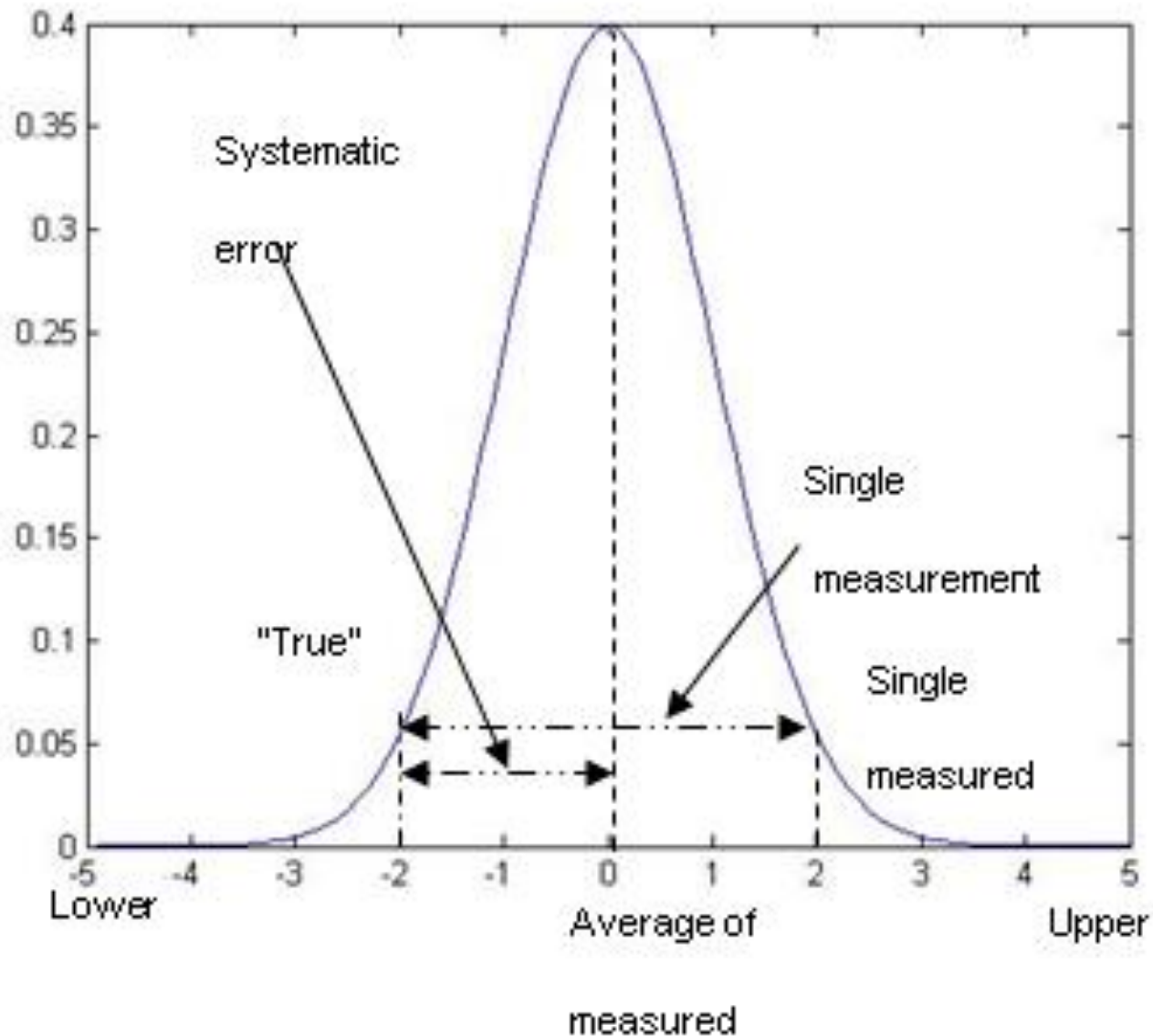
- Accuracy- difference in response between a standard and instrument in varying environmental conditions a measure of how close a measurement is to the “true” value
- Precision- how well repeated measurements of some quantity agree with each other. A precise instrument can be inaccurate



# Systematic vs. Random Errors

- Random- that which is not precisely predictable or determinable
- Systematic- errors arising from a consistent response of a measuring device to environmental conditions or faulty characteristics of instrumentation that occurs frequently

# Truth vs. single and large sample of observations



# Population vs. Sample

- we never know the entire population of true values as the environmental conditions change in time or space.
- We hope that we choose a sample of observations for analysis such that each element in the population has an equal chance to be selected.
- Sampling issues
  - Trends
  - serial dependence of environmental data
  - model sample tend to be less variable than observed samples

# Selecting a Sample

- Selecting the sample for analysis is a critical aspect of organizing the data and depends on the question to be addressed by the study
- rule of thumb: sample should be large enough to capture the phenomenon of interest many times
- “Degrees of freedom”: number of independent elements in the sample;
  - usually much smaller than the total number of members in the sample in environmental data sets
- Keeping your powder dry- saving data for an independent sample to evaluate and confirm your results.
- Tendency to assume sample is drawn randomly from the population, when sample grossly underestimates the variability inherent in the population

# Selfies

“It’s not surprising that men who post a lot of selfies and spend more time editing them are more narcissistic, but this is the first time it has actually been confirmed in a study,”

## Study Links Selfies To Narcissism And Psychopathy

The Huffington Post | By Carolyn Gregoire



Posted: 01/12/2015 8:15 am EST | Updated: 01/12/2015 8:59 am EST



The sample included 800 men from age 18 to 40 who completed an online survey asking about their photo posting behavior on social media. The participants also completed standard questionnaires for anti-social behaviors and for self-objectification. (This study doesn’t include women because the dataset, which Fox received from a magazine, did not have comparable data for women.):

**Facebook updates may stave off loneliness, even if no one 'likes' you, study finds**



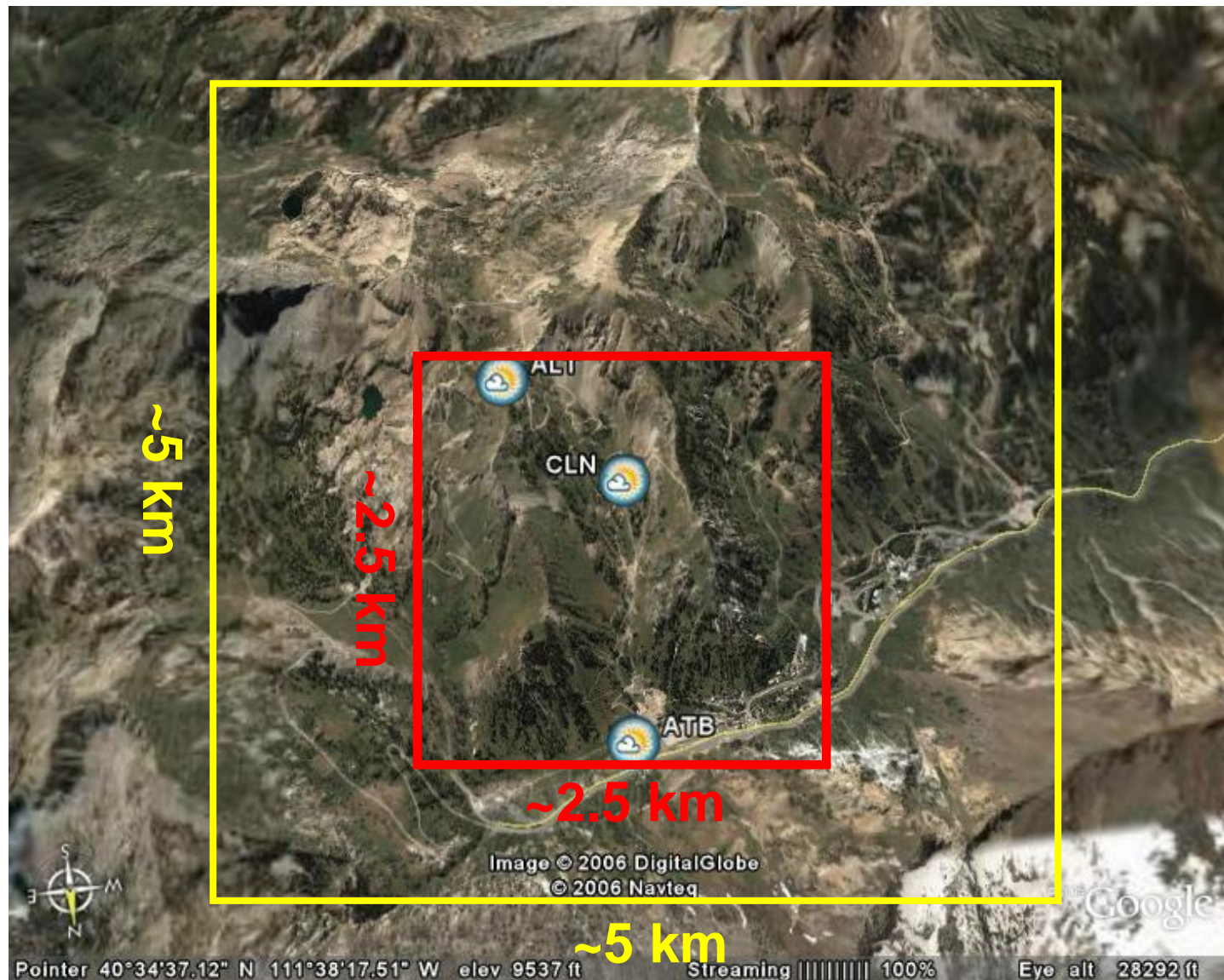
- Compared with other students, those who had been urged to go on a status-writing blitz felt less lonely
- happiness and depression levels unchanged, “suggesting that the effect is specific to experienced loneliness,” Simply thinking about their friends can have a “social snacking” effect.
- “Similar to snack temporarily reducing hunger until next meal, social snacking may help tolerate lack of ‘real’ social interaction for a certain amount of time”
- Scientists have found clues to what compels people to constantly update their Facebook status. College students who posted more status updates than they normally did felt less lonely over course of a week, even if no one “liked” or commented on their posts
- researchers at Free University Berlin recruited 100 undergraduates at University of Arizona; paper published last month in [Social Psychological and Personality Science](#).
- Participants filled out surveys to measure their levels of loneliness, [happiness](#) and depression, and gave researchers access to their Facebook
- students were sent an analysis of their average weekly status updates; some were told to post more updates than usual over next seven days. During that week, all completed a short online questionnaire at the end of each day about their mood and level of social connection.

# Observations

- Observations are not perfect...
  - Gross errors
  - Local siting errors
  - Instrument errors
  - Representativeness errors



# Representative errors to be expected in mountains Alta Ski Area

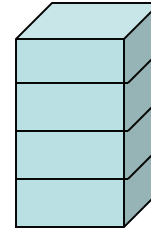
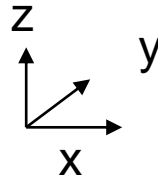
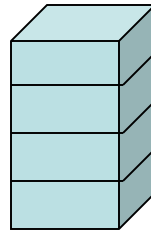


# Common Goal is to Synthesize and Reduce Dimensionality

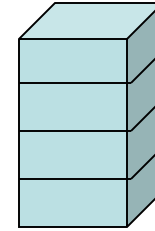
- Statistical analysis of environmental data typically involves reducing the dimensionality of the data to a manageable size.
- Which variable(s) do we need to consider?
- Can we consider one variable (univariate analysis) or must we consider multiple variables (multivariate analysis)?
- What time scales 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:  $x, y, z$



$t$

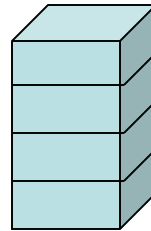
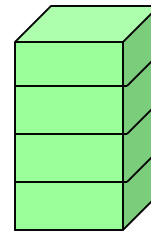
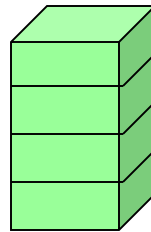


$t+1$

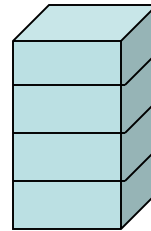
- Time: time ( $t$ ) and forecast time ( $t_f$ )

- Parameter &  
Source

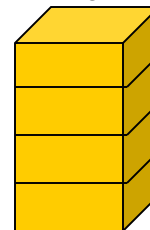
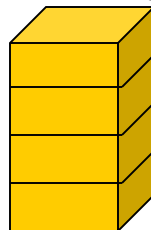
(temperature, winds,  
different models,  
measuring systems,  
perturbations)



$t$



$t+1$



# Review Material

- [https://github.com/johnhorel/ATMOS\\_5020\\_2019/blob/master/Aug\\_20\\_2019.md](https://github.com/johnhorel/ATMOS_5020_2019/blob/master/Aug_20_2019.md)
- [https://github.com/johnhorel/ATMOS\\_5020\\_2019/blob/master/Sept\\_03\\_2019.md](https://github.com/johnhorel/ATMOS_5020_2019/blob/master/Sept_03_2019.md)
- [https://github.com/johnhorel/ATMOS\\_5020\\_2019/blob/master/Sept\\_05\\_2019.md](https://github.com/johnhorel/ATMOS_5020_2019/blob/master/Sept_05_2019.md)

# Climate

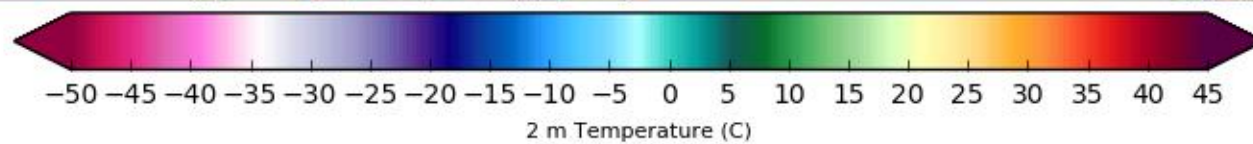
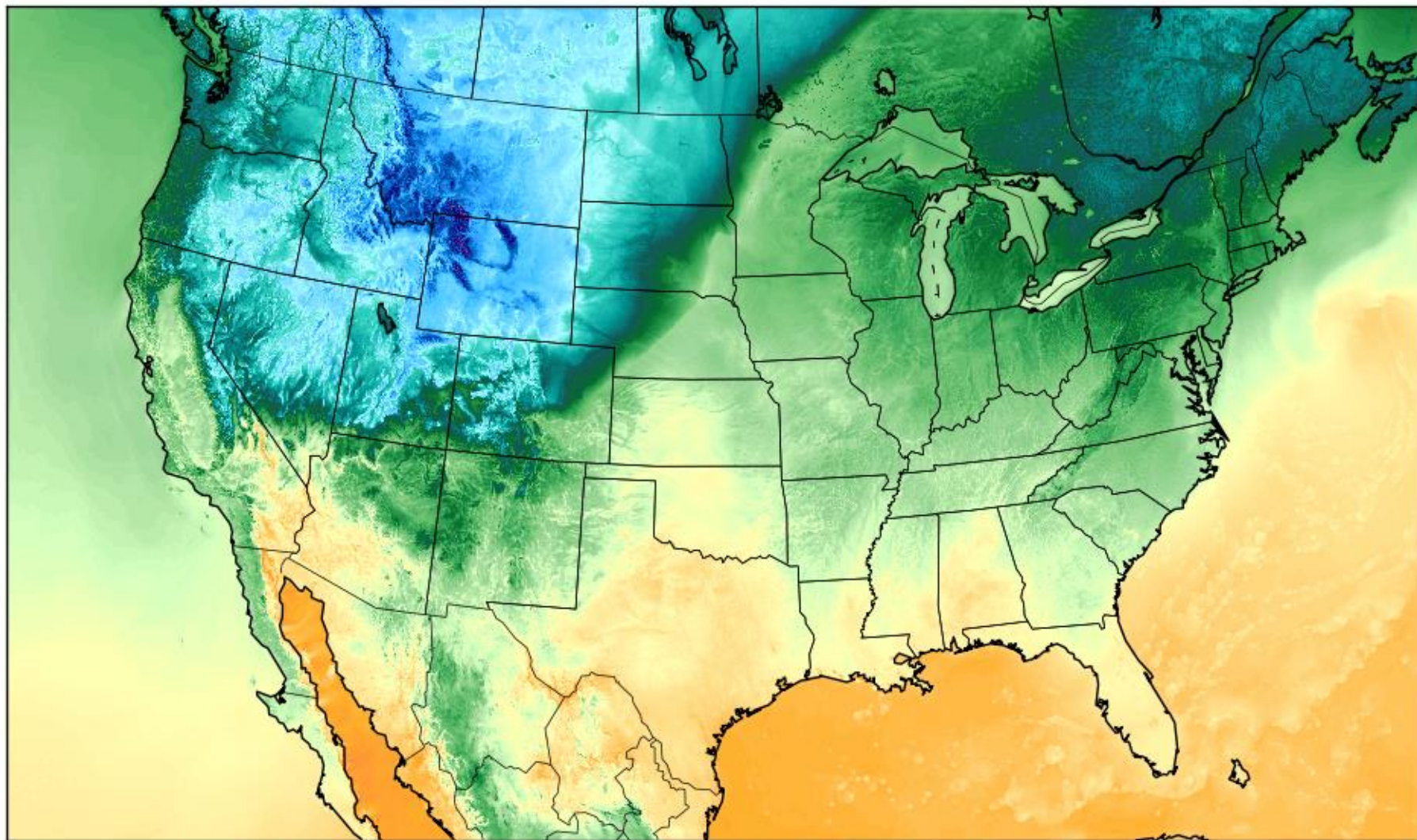
- Statistics often misinterpreted as bookkeeping.
  - What is the warmest temperature on record at Salt Lake City?
  - What is the biggest snow storm at Alta?
- weather and climate:
  - weather- state of the environment
  - climate- aggregate summary of the environment
  - Climate normal: arbitrarily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



Run: 2019-10-09 12:00 UTC F18

**HRRR**

Valid: 2019-10-10 06:00 UTC

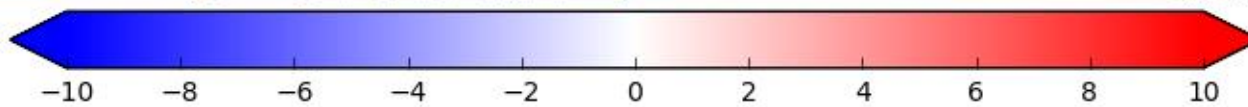
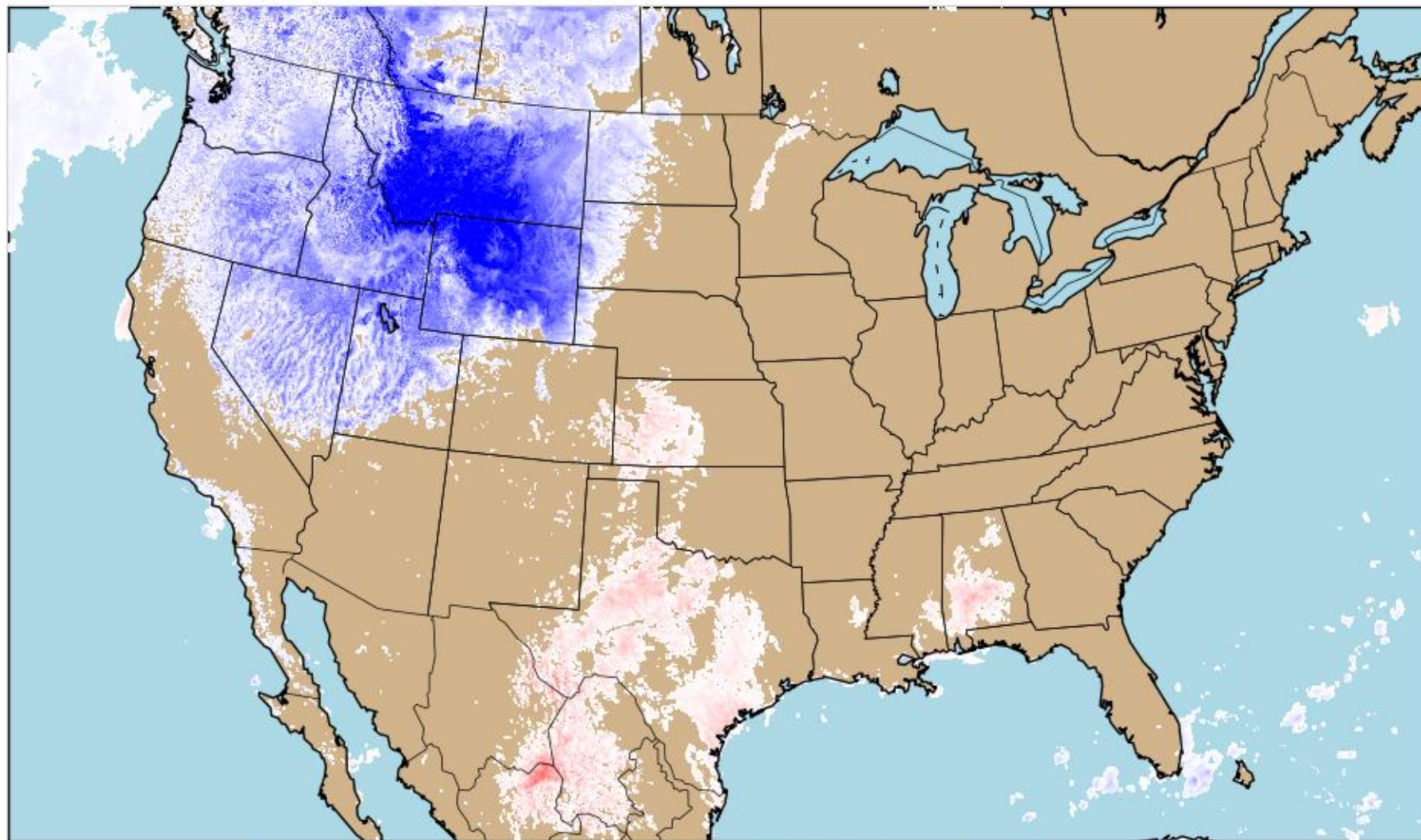




Run: 2019-10-09 12:00 UTC F18

**HRRR**

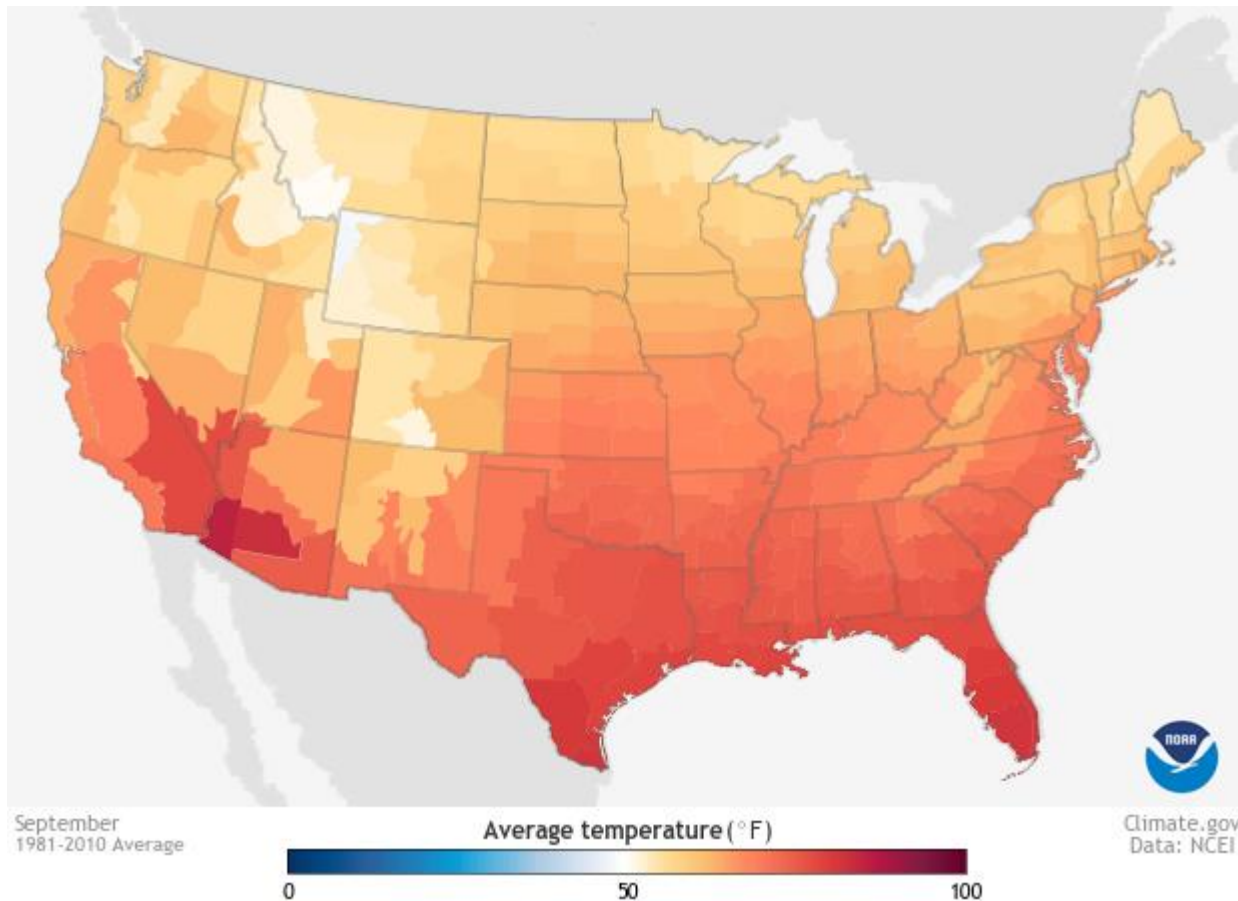
Valid: 2019-10-10 06:00 UTC



5<sup>th</sup>/95<sup>th</sup> percentile Depression/Exceedance 2 m Temperature (C)

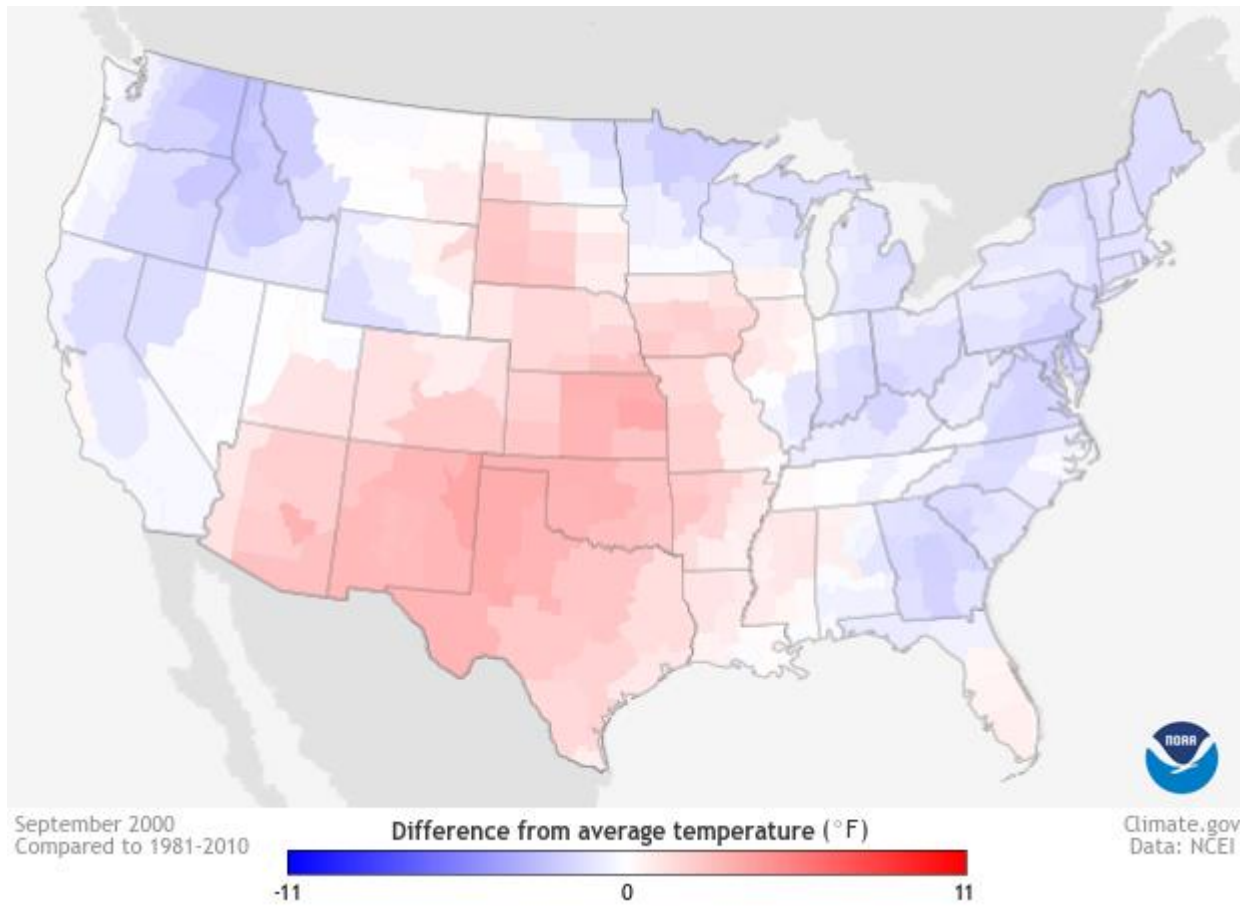
# <https://www.climate.gov/maps-data>

September 30 year average





# September 2019 Departure from Normal



# <https://www.ncdc.noaa.gov/sotc/global/201908>

## Global Climate Report - August 2019

Climate Monitoring

State of the Climate

Temp, Precip, and Drought

Climate at a Glance

Extremes

Societal Impacts

Snow and Ice

Teleconnections

Monitoring References

[State of the Climate Reports](#) | [Summary Information](#) | [Monthly Climate Briefings](#) | [RSS Feed](#) [XML](#)

Report:  Year:  Month:

[Global Climate Report](#)  
[« July 2019](#)

[Introduction](#) [Temperature](#) [Precipitation](#) [References](#)

### Maps and Time Series

#### Temperature and Precipitation Maps

[August 2019](#) | [June - August 2019](#)

#### Temperature Anomalies Time Series

[August](#) | [June - August](#) | [Year-to-Date](#)

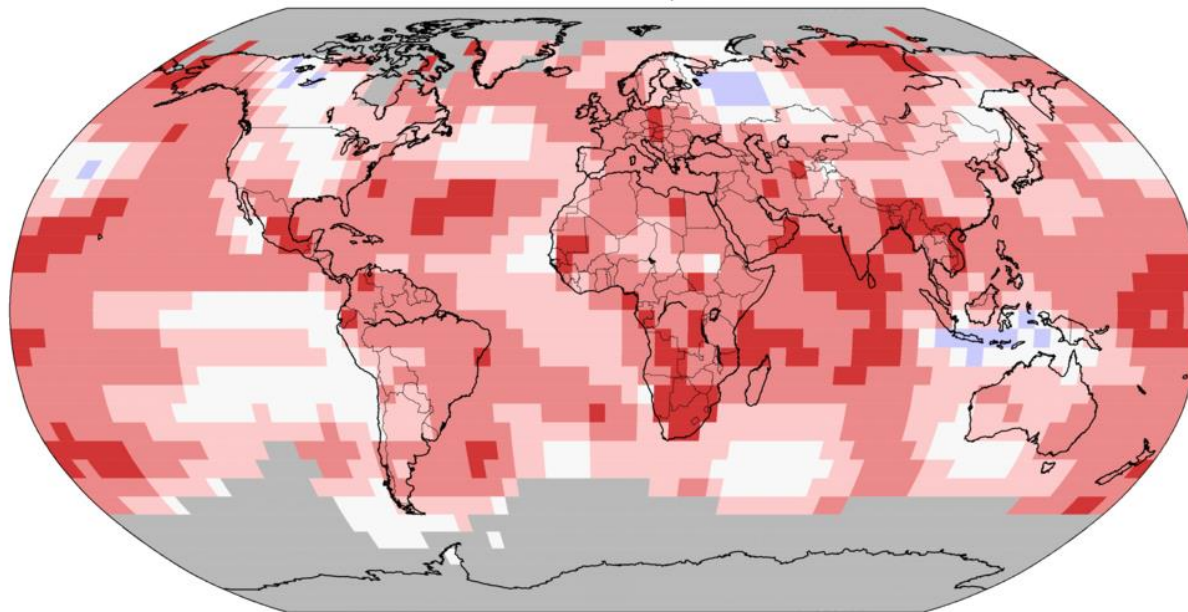
### Supplemental Material

- [2019 year-to-date temperatures versus previous years](#)
- [Global Annual Temperature Rankings Outlook](#)
- [Monthly temperature anomalies versus El Niño](#)
- [Mean Monthly Temperature Records Across the Globe](#)

## Land & Ocean Temperature Percentiles Jun 2019–Aug 2019

NOAA's National Centers for Environmental Information


Data Source: NOAAGlobalTemp v5.0.0–20190908




  
Record  
Coldest

  
Much  
Cooler than  
Average

  
Cooler than  
Average

  
Near  
Average

  
Warmer than  
Average

  
Much  
Warmer than  
Average

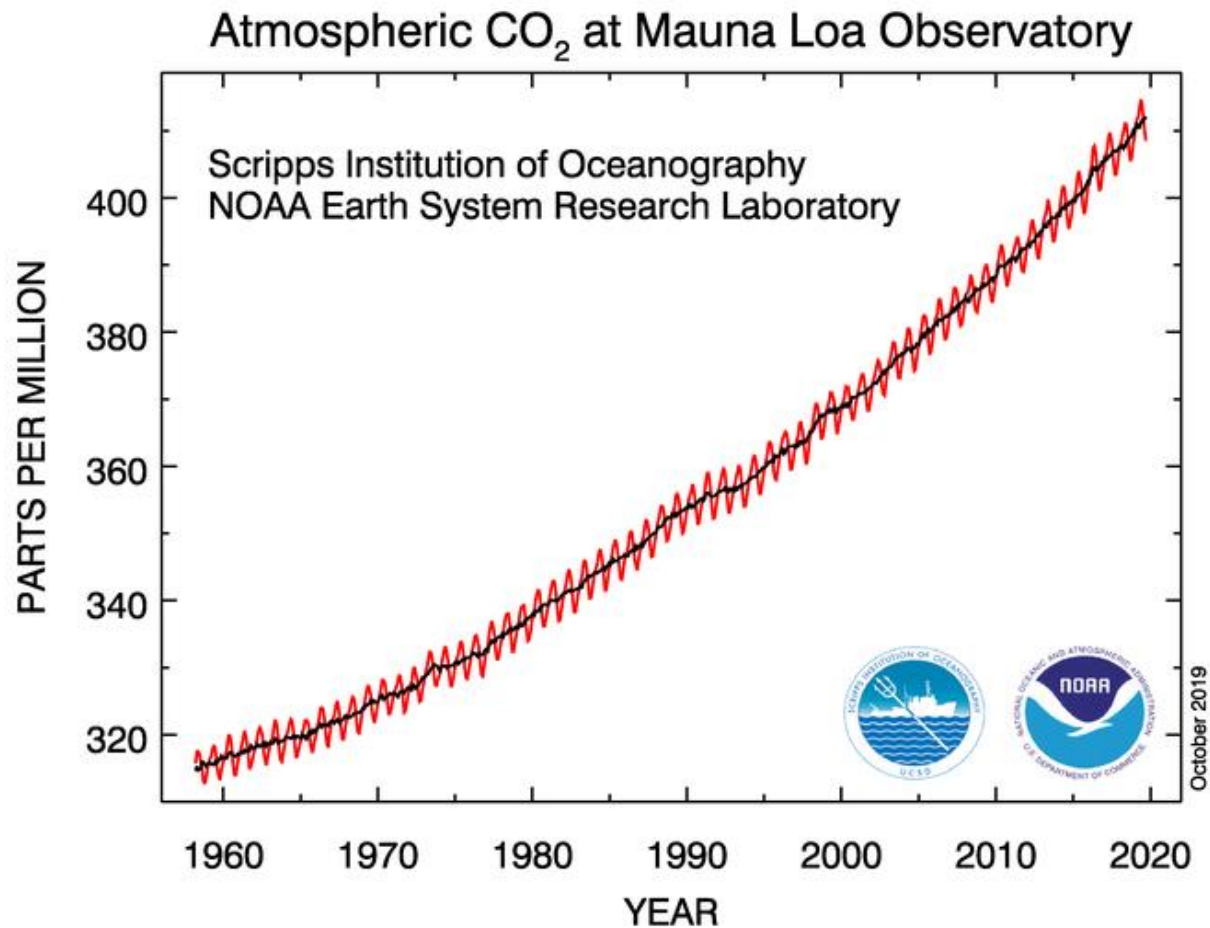
  
Record  
Warmest



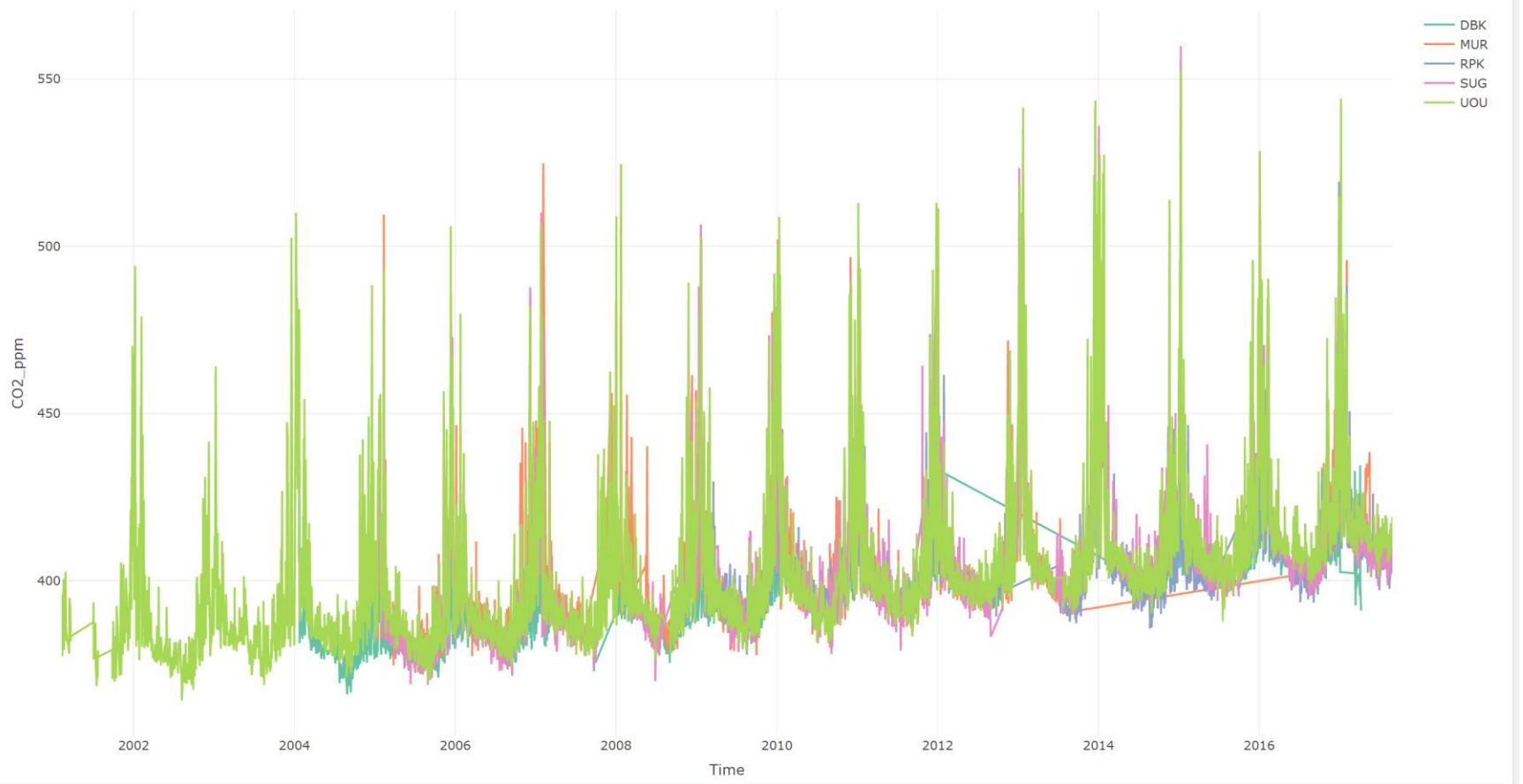
GHCNM v4.0.1.20190906.qfe



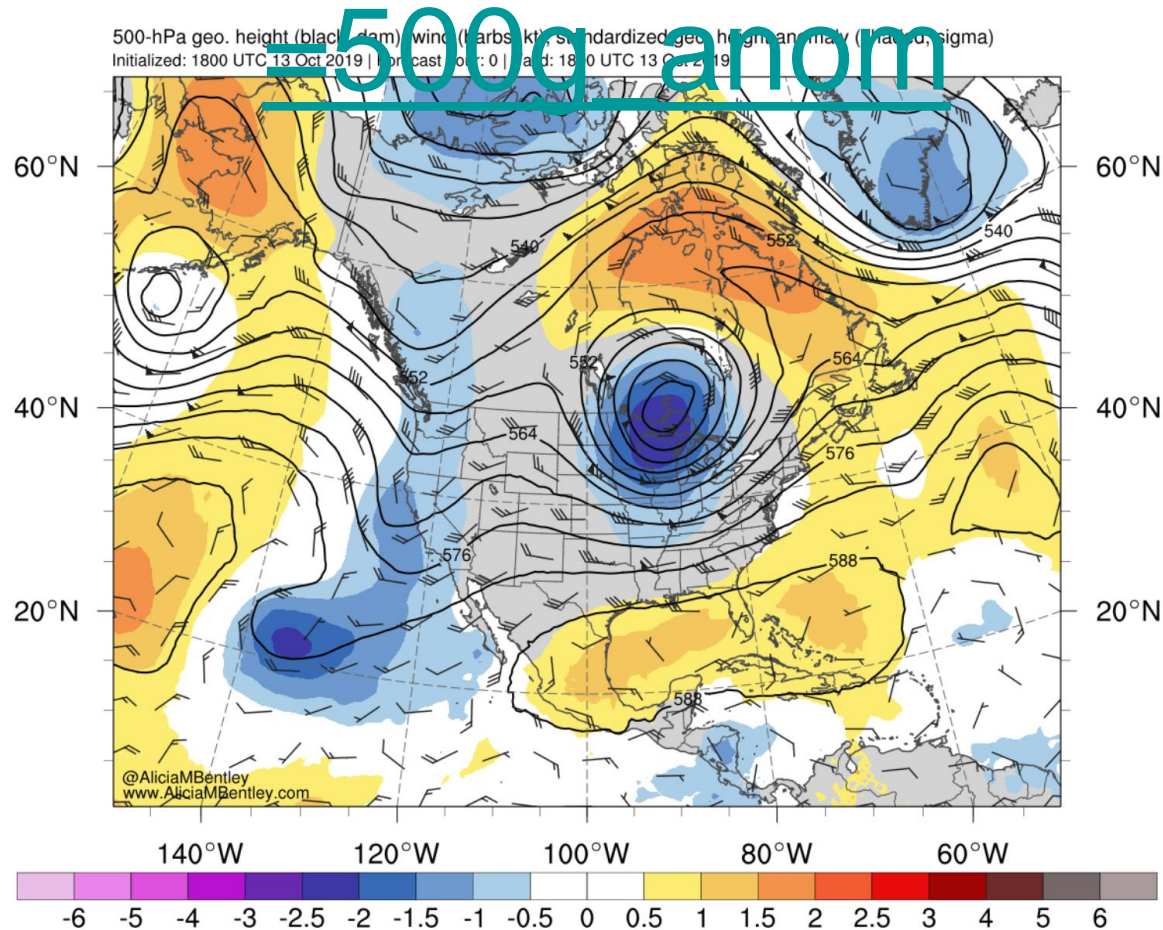
<https://www.esrl.noaa.gov/gmd/obop/mlo/>



# CO2 on Campus



udent/abentley/realtime/anom.p  
hp?domain=northamer&variable





The screenshot shows the USGS Earthquake Map application. The main map area displays a world map with earthquake epicenters. The epicenters are color-coded based on their age: red for less than 1 day, yellow for 1-7 days, and white for 8-30 days. The map shows a high density of earthquakes along the Mid-Atlantic Ridge and the western Pacific. The interface includes a search bar at the top, a sidebar with settings for earthquakes, list sort order, map layers, and time zone, and a taskbar at the bottom.

**USGS**

Search:  Zoom to:

**MY SETTINGS**  
Bookmark to return to map/list with same settings.

**Earthquakes**

- ☒ Auto Update
- ☐ 1 Day, Magnitude 2.5+ Worldwide
- ☐ 1 Day, All Magnitudes Worldwide
- ☐ 7 Days, Magnitude 4.5+ Worldwide
- ☐ 7 Days, Magnitude 2.5+ Worldwide
- ☐ 7 Days, All Magnitudes Worldwide
- ☐ 30 Days, Significant Worldwide
- ☒ 30 Days, Magnitude 4.5+ Worldwide
- ☐ 30 Days, Magnitude 2.5+ Worldwide

[Create new or edit existing search.](#)  
[Search Earthquake Archives](#)

**List Sort Order**

- ☒ Newest first
- ☐ Oldest first
- ☐ Largest magnitude first
- ☐ Smallest magnitude first

☒ Only List Earthquakes Shown on Map

**Map Layers**

- ☒ Grayscale
- ☐ Terrain
- ☐ Street
- ☐ Satellite
- ☒ Plate Boundaries
- ☐ U.S. Faults
- ☐ U.S. Hazard

**Time Zone**

- ☒ Local System Time (UTC-07:00)
- ☐ Event Epicenter Time
- ☐ UTC

1000 km 2000 mi

Earthquake Age  
Red Day Week Year

71.746°N 01.758°W

AH\_1.jpg

Show all downloads...

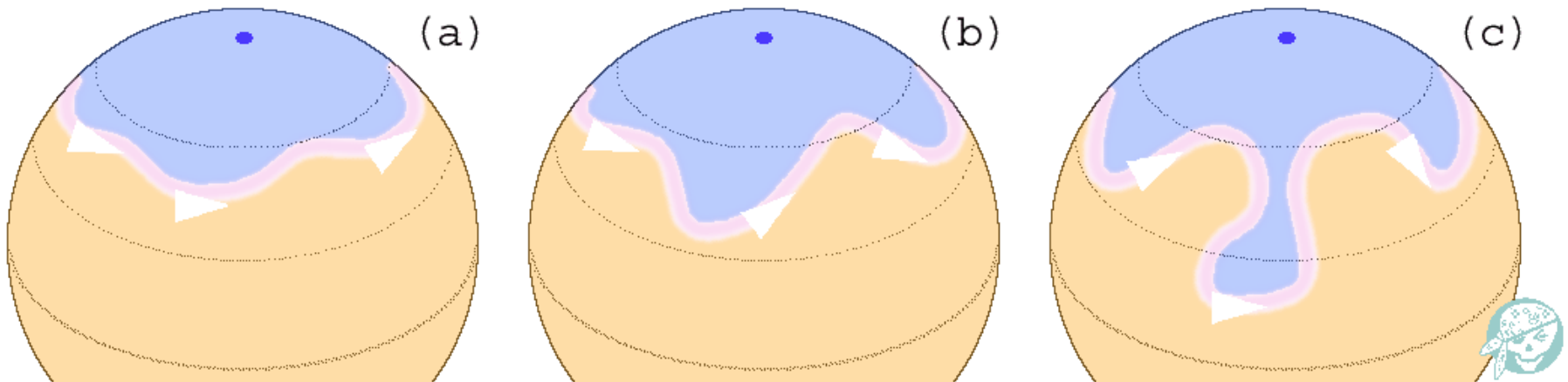
# Why use statistics to describe the environment?

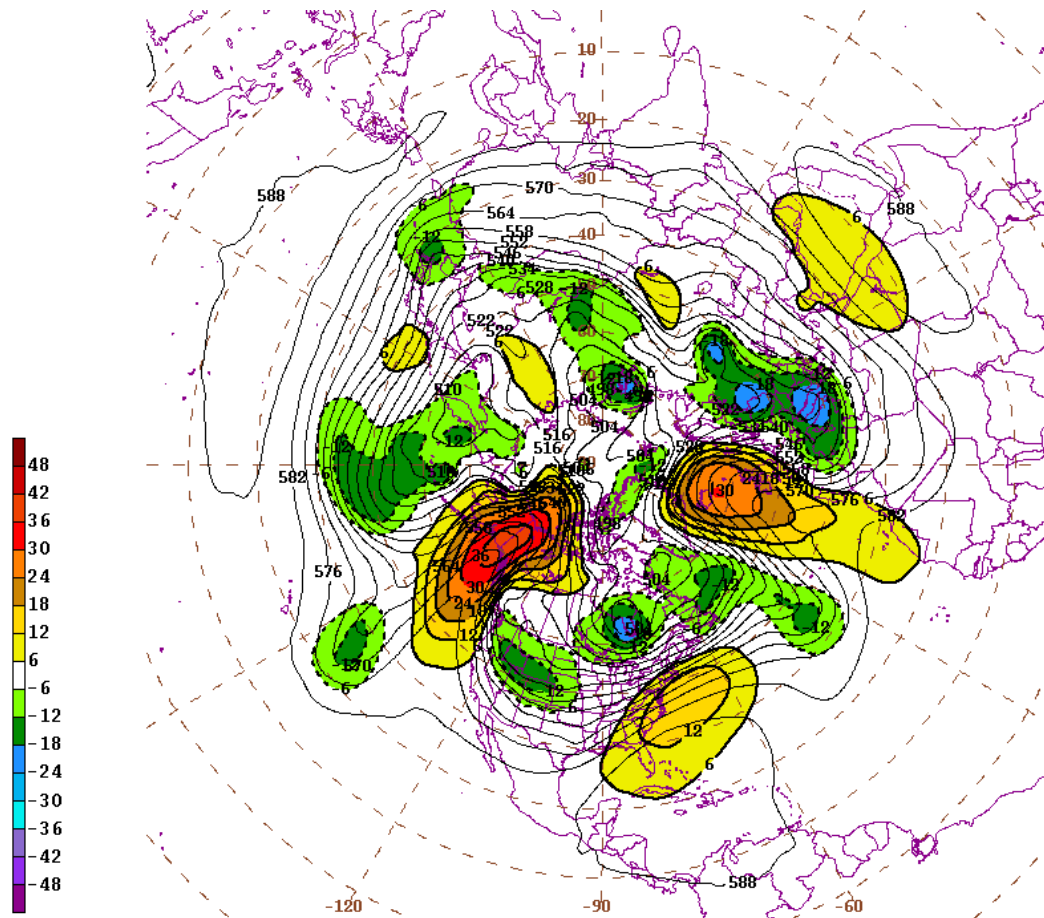
- Environment controlled by innumerable factors, which we hope to segregate into a few critical factors from the rest that, for the most part, simply contribute to background noise
- the characteristics of the system include linearly unstable processes such as baroclinic waves that cause growth of small 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



# Baroclinic 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

# 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 be doing

- Assignment 1- Survey- Due Friday
- Read Chapter 1 & 2a-2b Notes for Thursday
- Refamiliarize with using Mac lab and Python
- Assignment 2 due next week