Machine Learning for Climate Change

DS-UA 301 “Advanced Topics”

Prof. Grace Lindsay

Plan for the day

Introductions

Course logistics

Climate Change Overview

Break

Lecture on Energy Efficiency and Regression Techniques

Paper assignment

Introductions

About this course

In the coming years, the world will need to change with a speed and scale unseen before in order to tackle climate change and machine learning will play an important role in making this transformation possible.

Students should walk away from this course with an understanding of the complexities involved in tackling climate change and the landscape of applications for machine learning in this domain.

Are you comfortable with...

python

jupyter notebook

pandas

scikit-learn

matplotlib/seaborn

reading academic ML papers

climate change topics

Course Logistics

Syllabus - on **Brightspace**!

Homeworks will be posted and submitted on **Brightspace**

Lecture videos will be posted on **Brightspace**

Lecture slides will be posted on my **website** (linked to in **Brightspace**)

Syllabus

DS-UA 301 Advanced Topics in Data Science

Machine Learning for Climate Change

Center for Data Science, New York University
Spring 2024
Note: This syllabus is subject to change as announced in class.

Instructor

Prof. Grace Lindsay
Office: CDS, Room 601
Email: grace.lindsay@nyu.edu

TA/Grader

Rijul Dahiya
rd3629@nyu.edu

Grader

Ohm Patel, odp2008@nyu.edu

Meeting schedule

Class time: 8 AM to 10:30 AM on Thursdays
Class Location: Kimmel (60 Washington Square S) Rm 808
Labs:
Fr 2:45PM - 3:35PM, 60 Fifth Ave (CDS) Room 110Fr 3:45PM - 4:35PM, 70 Washington Sq S (Bobst) Room LL139

Office hours

With Rijul:
Fridays, 1:30 PM to 2:30 PM. CDS 244. Or by appointment.
With Grace:
By appointment (email me), either in-person or on Zoom.

Links to materials

Lecture slides:
https://lindsay-lab.github.io/teaching/#machine-learning-for-climate-change-spring-2025
Class Slack:
https://join.slack.com/t/coursediscuss-tcx6125/shared_invite/zt-2xhlkuycx-gBMr04Y6qA4YO27mwTSveQ (join the ml4cc channel)

Overview

Syllabus

Course Structure

This course will introduce you to both modern machine learning techniques as well as their domain-specific application toward climate change problems. Lectures will be organized around published research papers. Unless otherwise noted, in the second half of each class, there will be a lecture that provides the necessary background needed to read the assigned paper. In the first half of the next class, we will have a group discussion on the assigned paper.

Lecture order

Reminder of assignments
Climate change in the news
Recap of previous class
Discussion of assigned paper
Break
Lecture on background for next paper
(incl. climate and ML topic)

Syllabus

Evaluation

Your grade will be calculated according to following proportions:

Exams - 20% (10% each exam)
Project - 35%
Homeworks - 25%
Participation - 20%

Syllabus

Exams: Two in-class exams will be administered that evaluate understanding of climate change concepts and the papers read in class. Make-up exams will only be offered in the case of medical emergencies and must be arranged before the start of the exam. They must be completed before the next class.
Project: Students will work in groups on an advanced research project of their design. Grades will be the same for all members of the group and will be based on both the project and its presentation.
Homework: Three out of four homework assignments will be coding assignments related to the themes of the class. The fourth will be a project plan. Assignments must be turned in before the start of class on the due date unless otherwise noted. Students will have 3 "grace days" they can use toward the first three assignments that will allow them to turn in assignments late. They can be used altogether (allowing a single assignment to be 3 days late) or separately. *Once the grace days are used, late assignments will not be accepted and will be treated as a 0.* Grace days cannot be applied to the project plan or final project.
Participation: Students will be required to submit a PMIRO+Q before and after in-class discussions of assigned papers. Participation grades will also be based on attendance, which will be taken by discussion groups (if you need to participate virtually you must set this up with your discussion group). You can miss one discussion without penalty.

Extra Credit - Students who demonstrate a sincere desire to correct poor performance before the end of the course can request an additional assignment.
Requests for extra credit ***cannot*** be made after April 30.

Syllabus

<u>Academic integrity</u>

Academic Integrity, Plagiarism, and Cheating (adapted from [the website of the College of Arts & Science]): Academic integrity means that the work you submit is original. Obviously, bringing answers into an examination or copying all or part of a paper/code straight from a book, the Internet, or a fellow student is a violation of this principle. But there are other forms of cheating or plagiarizing which are just as serious — for example, presenting an oral report drawn without attribution from other sources (oral or written); writing a sentence or paragraph which, despite being in different words, expresses someone else's idea(s) without a reference to the source of the idea(s); or submitting essentially the same paper in two different courses (unless both instructors have given their permission in advance). Receiving or giving help on a take-home paper, examination, or quiz is also cheating, unless expressly permitted by the instructor (as in collaborative projects).

ChatGPT

Admissions cycle	Acceptance rate	Number of applicants
Class of 2028	**8.0%**	**118,000**
Class of 2027	8.0%	120,000
Class of 2026	12.2%	105,000
Class of 2025	12.8%	100,131
Class of 2024	15.0%	85,000
Class of 2023	16.0%	84,481
Class of 2022	19.0%	75,000
Class of 2021	27.0%	67,000

General Cost of Attendance 2024-2025

*Undergraduate Students (two semesters, full-time)**

Category	On/Off Campus Student	Commuter Student
Tuition	$62,796	$62,796
Food and Housing	$24,652	$7,554
Estimated Total Direct Costs	**$87,448**	**$70,350**
Books and Supplies	$1,442	$1,442
Transportation	$2,342	$1,188
Personal Expenses	$1,952	$1,952
Estimated Total Indirect Costs	**$5,736**	**$4,582**
Total	**$93,184**	**$74,932**

Climate Change

NPR

2023 was the hottest year ever recorded, scientists say

Global temperatures soared past previous records in 2023, according to new data from the European Union. Nations must slash fossil fuel...

The New York Times

2023 Was Hottest Year on Record by a Lot

Month after month global temperatures didn't just break records, they surpassed them by far. This year could be even warmer.

Carbon Brief

State of the Climate: 2023 smashes records for surface temperature and ocean heat

Carbon Brief examines the latest data across the oceans, atmosphere, cryosphere and surface temperature of the planet.

CNN

2023 will officially be the hottest year on record, scientists report

This year will be the hottest on record — by a significant margin. Between January and November 2023, the world's average temperature reached...

NPR

2023 was the hottest year ever recorded, scientists say

Global temperatures soared past previous records in 2023, according to new data from the European Union. Nations must slash fossil fuel...

The New York Times

2023 Was Hottest Year on Record by a Lot

Month after month global temperatures didn't just break records, they surpassed them by far. This year could be even warmer.

Carbon Brief

State of the Climate: 2023 smashes records for surface temperature and ocean heat

Carbon Brief examines the latest data across the oceans, atmosphere, cryosphere and surface temperature of the planet.

CNN

2023 will officially be the hottest year on record, scientists report

This year will be the hottest on record — by a significant margin. Between January and November 2023, the world's average temperature reached...

National Oceanic and Atmospheric Administration (.gov)

2024 was the world’s warmest year on record

It's official: 2024 was the planet's warmest year on record, according to an analysis by scientists from NOAA's National Centers for...

World Meteorological Organization WMO

WMO confirms 2024 as warmest year on record at about 1.55°C above pre-industrial level

The World Meteorological Organization (WMO) has confirmed that 2024 is the warmest year on record, based on six international datasets.

New Scientist

2024 may have been the rainiest – as well as hottest – year on record

Global average precipitation in 2024 may have broken the previous record set in 1998, as rising temperatures boosted the amount of moisture in the...

Axios

Hundreds of U.S. locations had their hottest year on record

It may be frigid in parts of the U.S. now, but communities from coast to coast saw record warmth during 2024, with many cities crushing...

AP News

Earth breaks yearly heat record and lurches past dangerous warming threshold

Earth recorded its hottest year ever in 2024, with such a big jump that the planet temporarily passed a major climate threshold,...

2 weeks ago

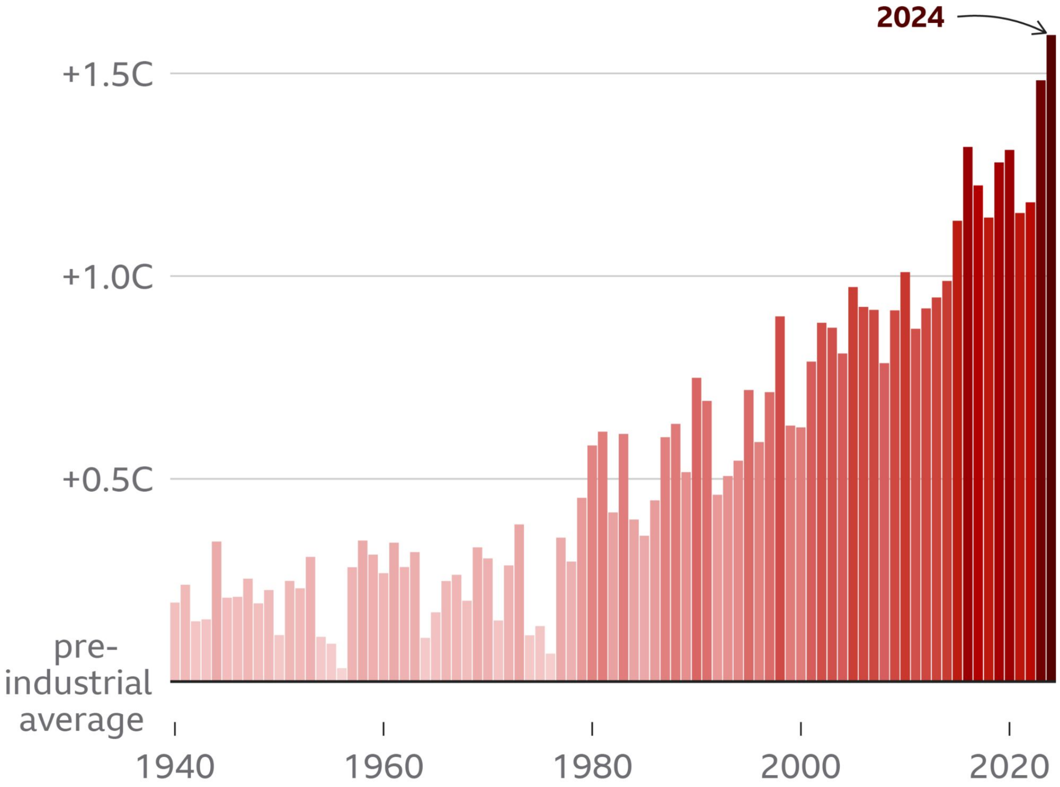
NBC News

Last year was hottest in recorded history, NASA says

NASA found that 2024 was hotter than any other year since at least 1880. Previously, 2023 had been considered the warmest on record.

2024 was the hottest year on record

Global average temperature by year, compared with the pre-industrial average, 1850-1900



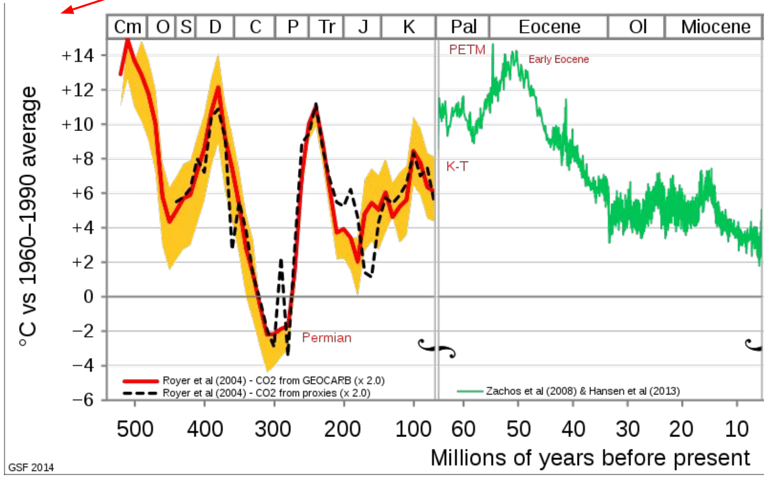
Source: ERA5, C3S/ECMWF

BBC

Historical Climate

Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

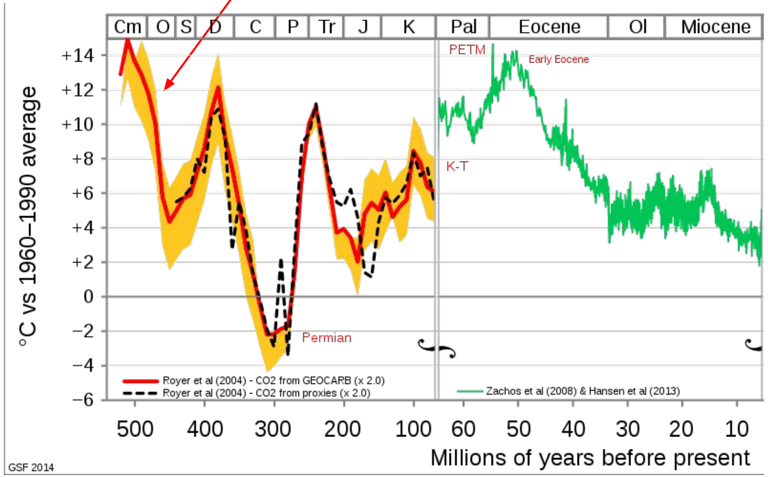
Snowball Earth ~650mya



Historical Climate

Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

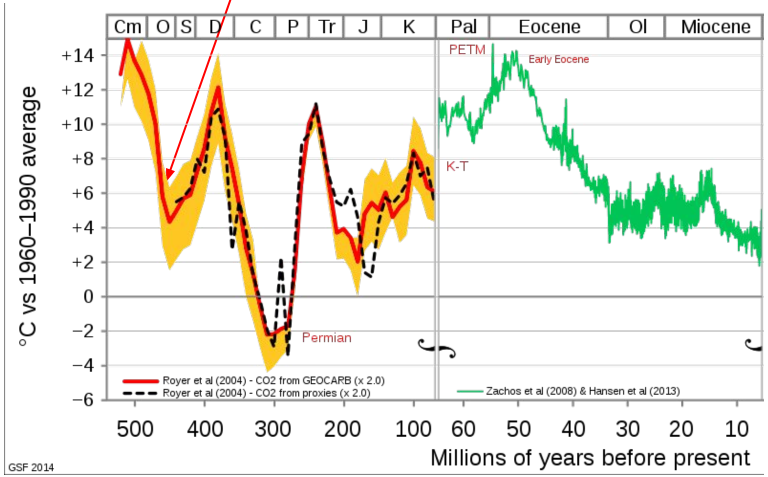
Warm and wet without distinct seasons



Historical Climate

Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

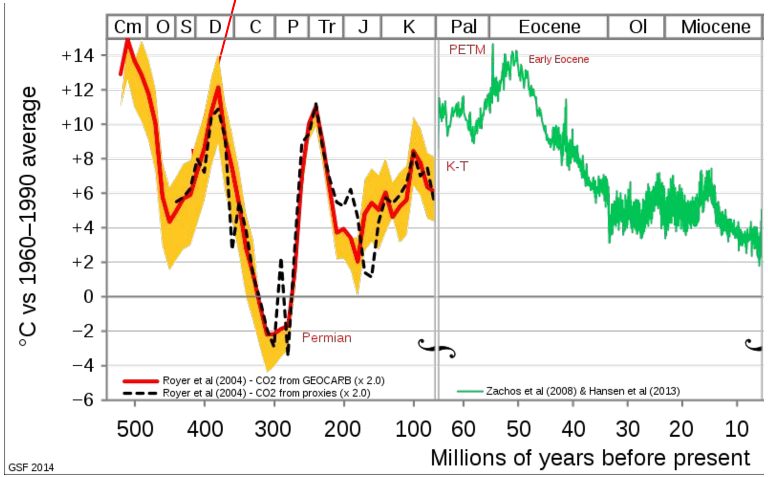
Ice age



Historical Climate

Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

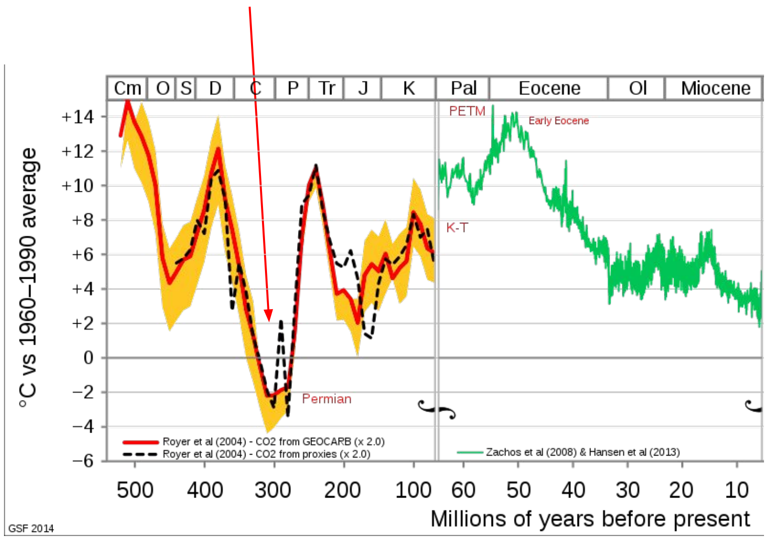
Glaciers melt, sea levels rise, cold south pole and hot at the equator



Historical Climate

Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

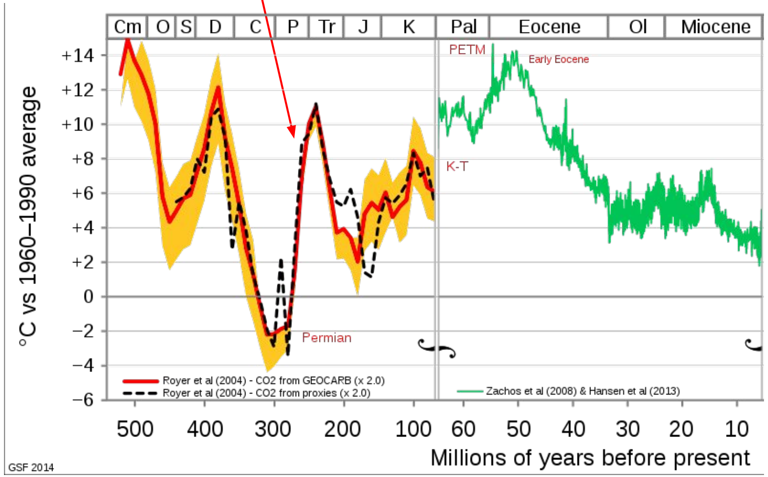
Swings between ice ages and ice melts, equator still tropical



Historical Climate

Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

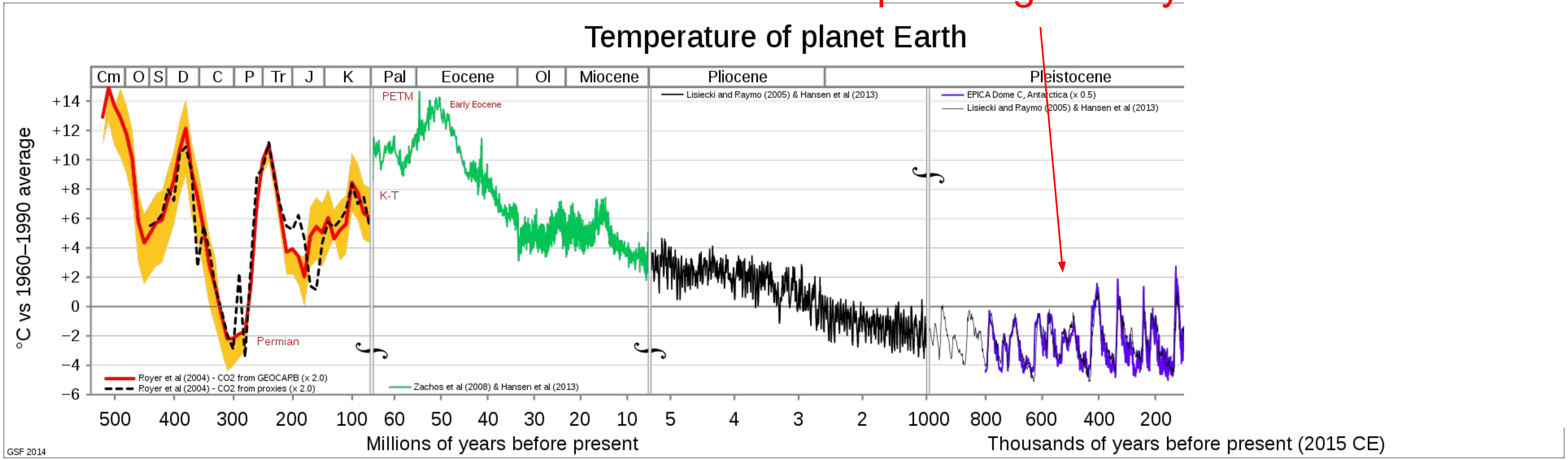
Pangea formation led to deserts. Huge warming led to extinction of 95% of species



Historical Climate

Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

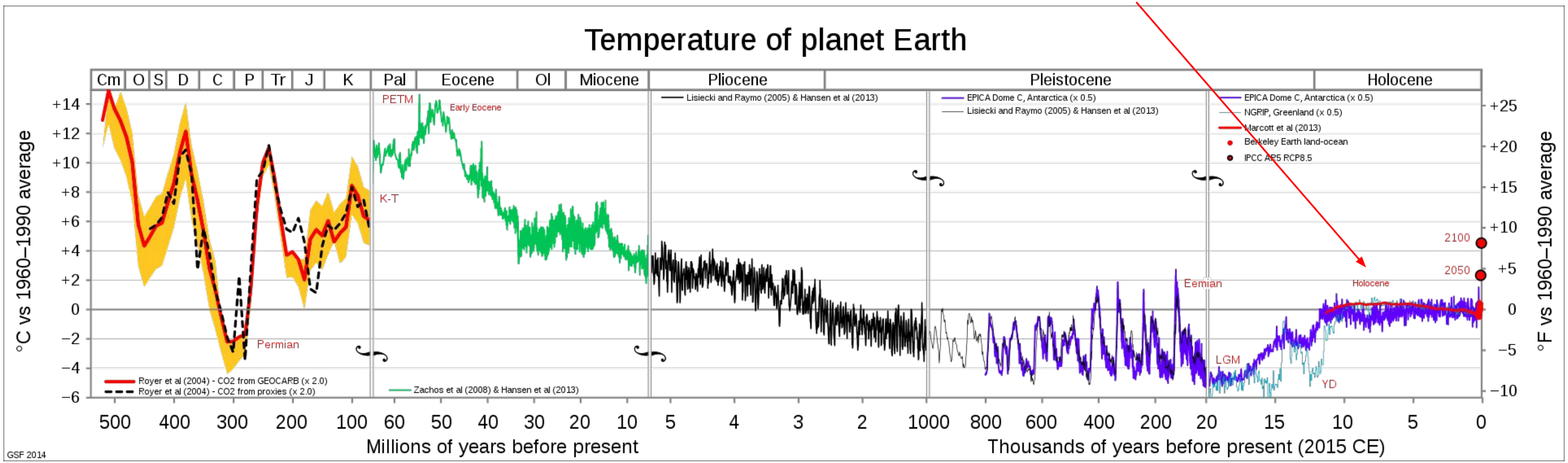
Repeated glacial cycles



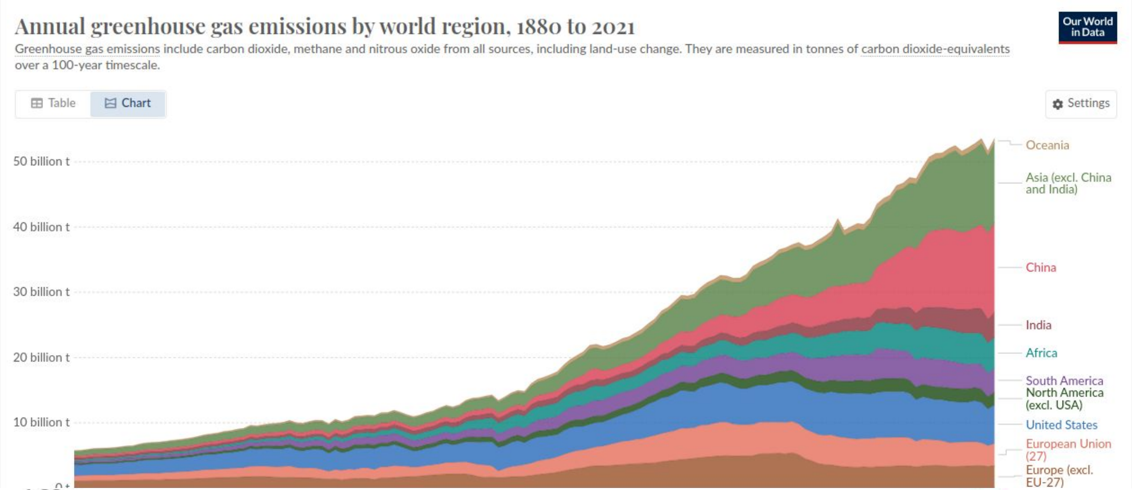
Historical Climate

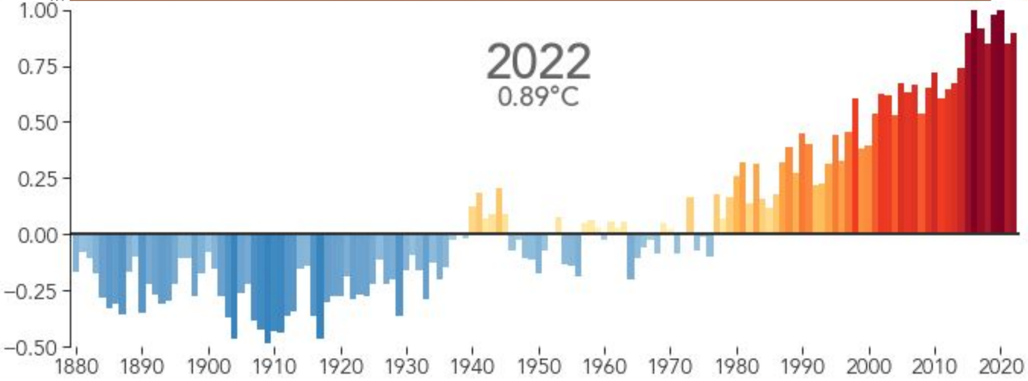
Paleoclimatology uses a variety of techniques to reconstruct past climate trends to understand how Earth systems work.

Holocene era aligns with expansion of human civilization



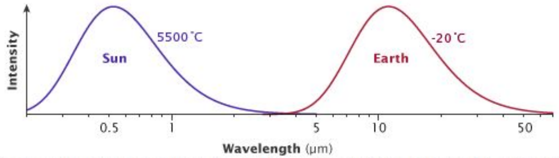
What’s causing the change?



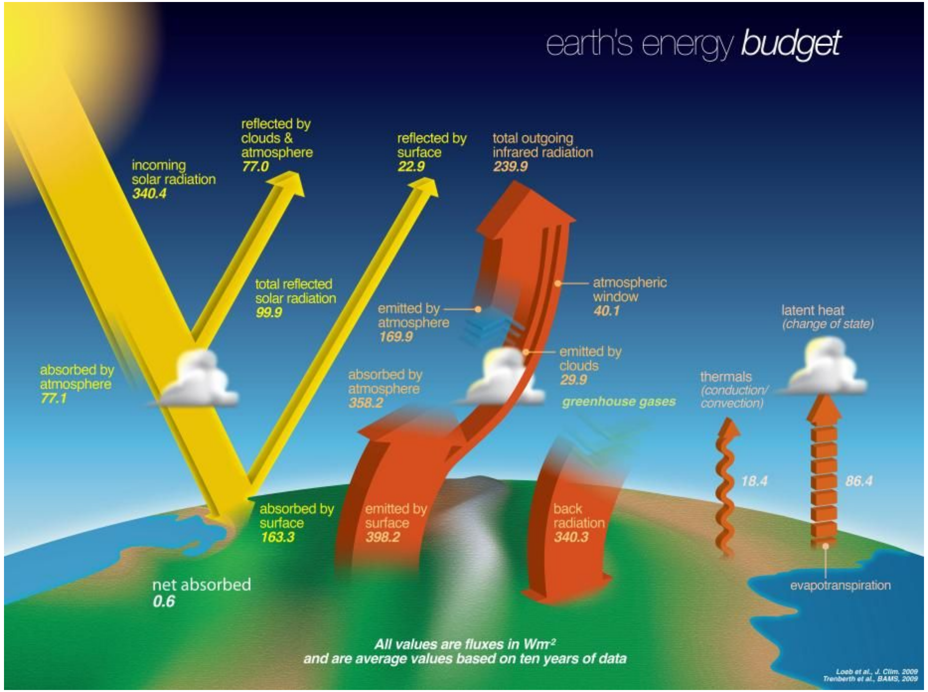


Earth’s energy “budget”

The earth receives energy from the sun, but also radiates energy back into space. The radiated energy is of a longer wavelength.



The Sun’s surface temperature is 5,500° C, and its peak radiation is in visible wavelengths of light. Earth’s effective temperature—the temperature it appears when viewed from space—is -20° C, and it radiates energy that peaks in thermal infrared wavelengths. (Illustration adapted from Robert Rohde.)

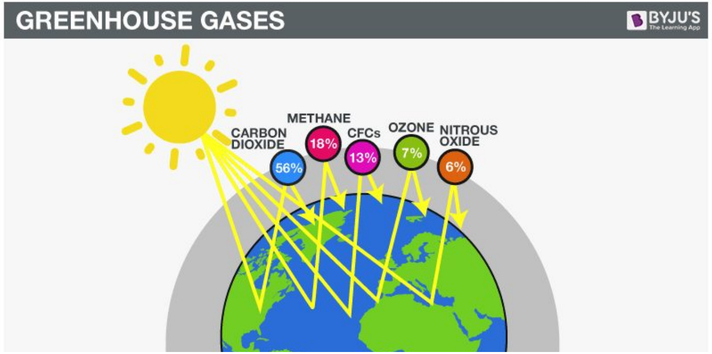


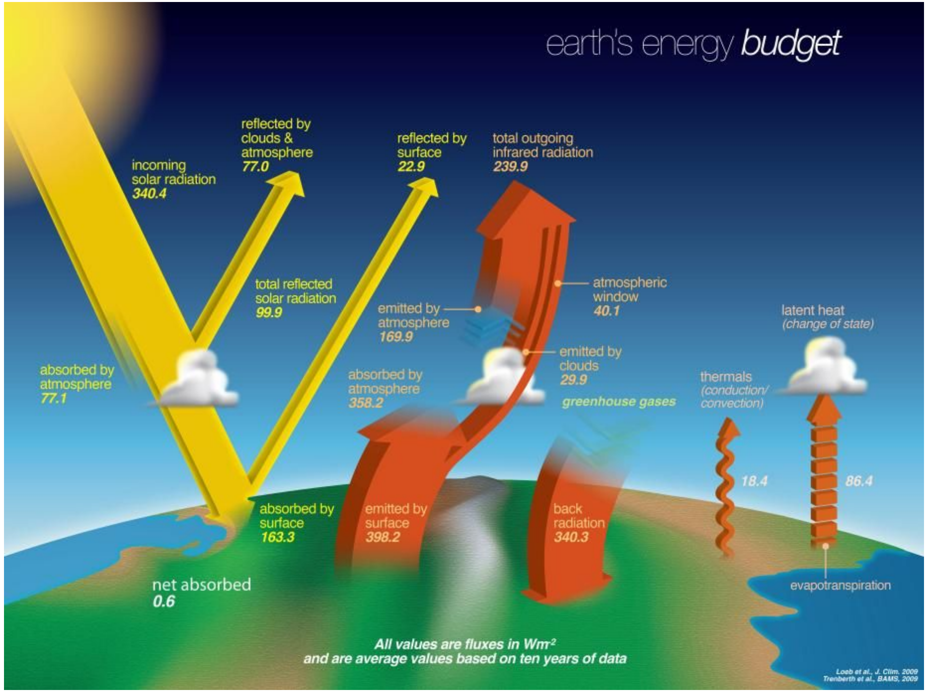
NASA

Earth’s energy “budget”

The greenhouse effect refers to the fact that heat emitted by the Earth’s surfaces radiates back to the Earth.

The composition of the atmosphere determines the amount of back-radiation.



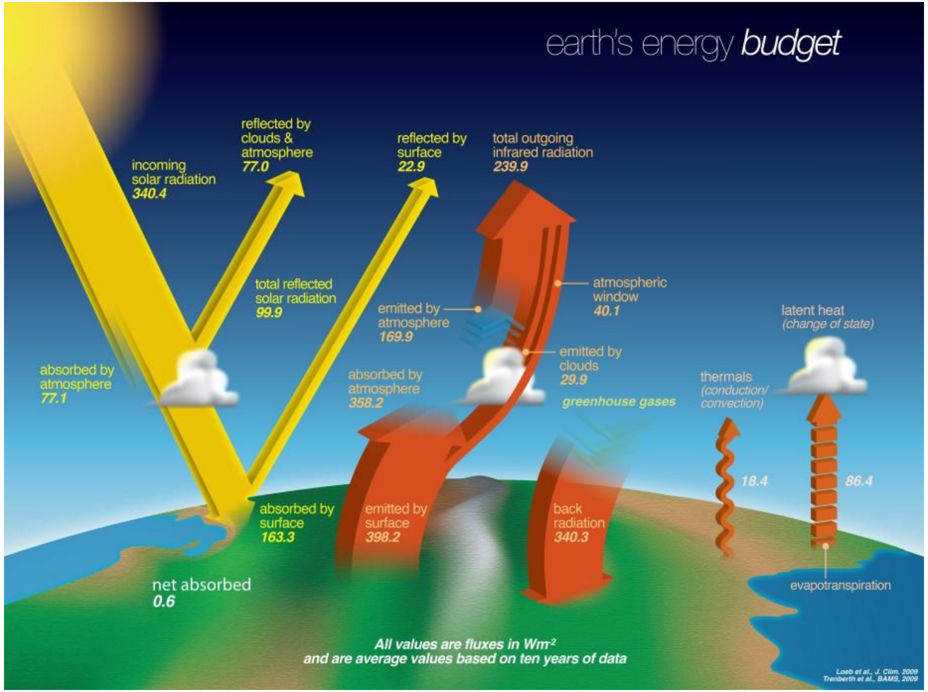


UCAR SciEd/NASA

Earth’s energy “budget”

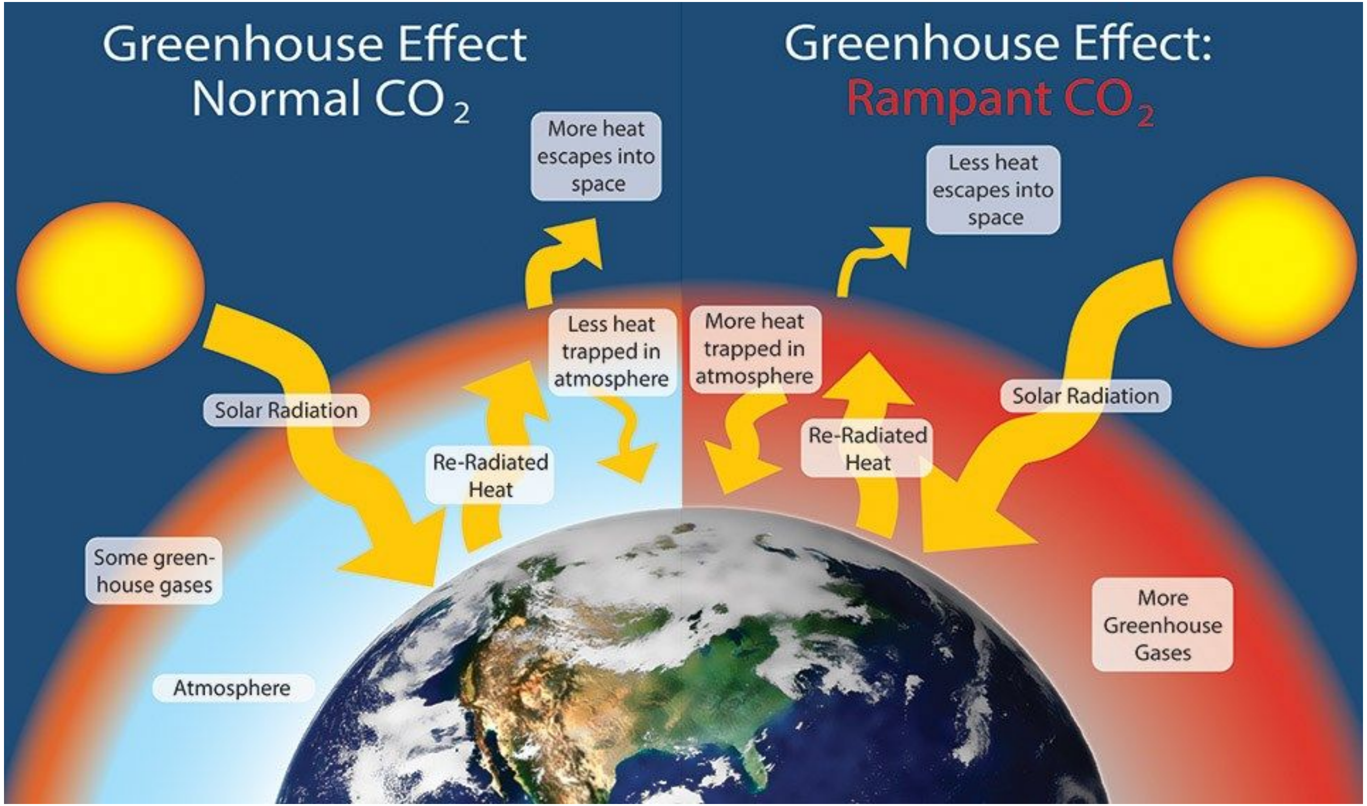
The greenhouse effect refers to the fact that heat emitted by the Earth’s surfaces radiates back down to the Earth.

The imbalance between the amount of energy that enters the earth’s atmosphere and that which leaves is known as “radiative forcing”.



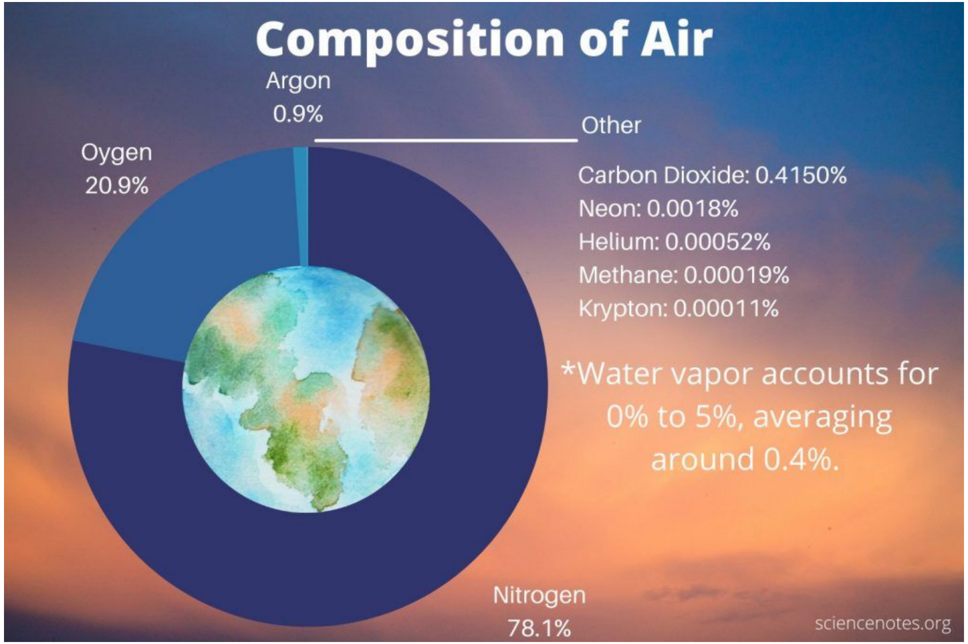
UCAR SciEd/NASA

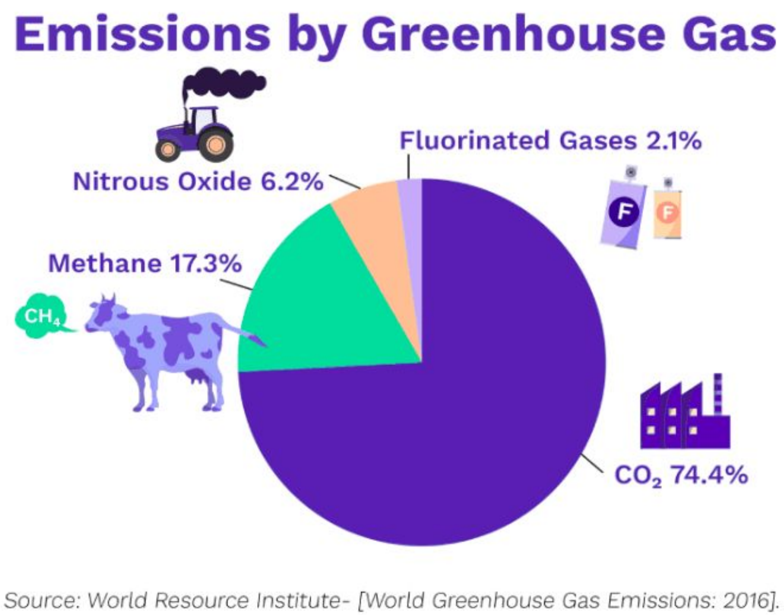
Greenhouse effect



nps.org

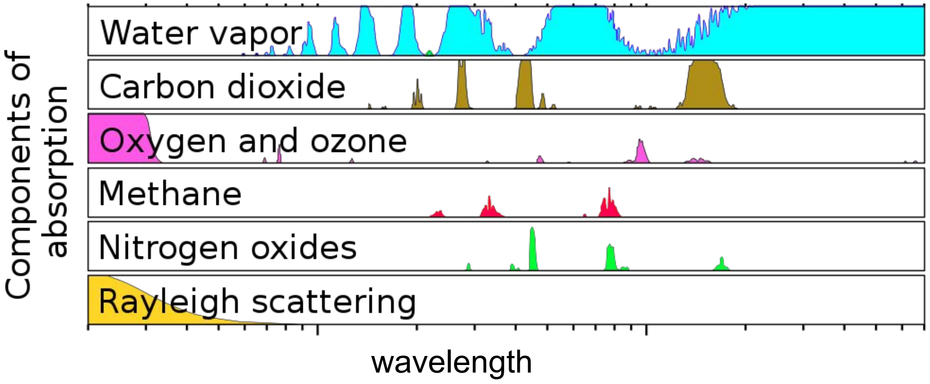
What are greenhouse gases?



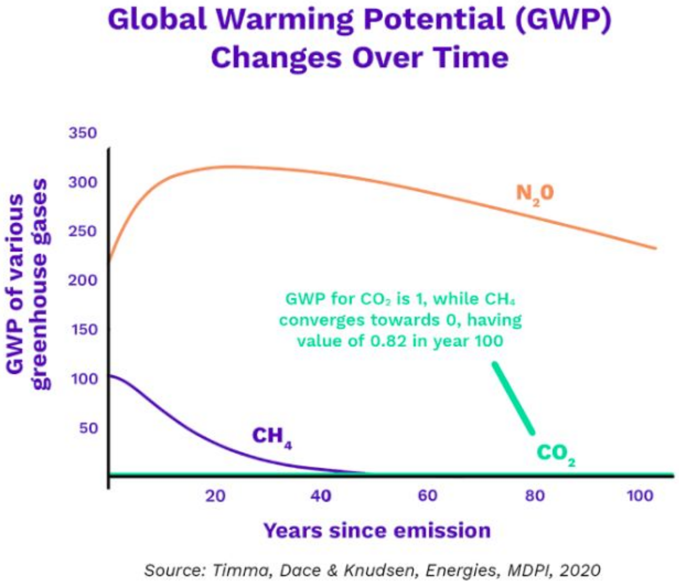


Non-water vapor greenhouse gases make up a tiny (>1%) fraction of the atmosphere

Not all greenhouse gases work the same way



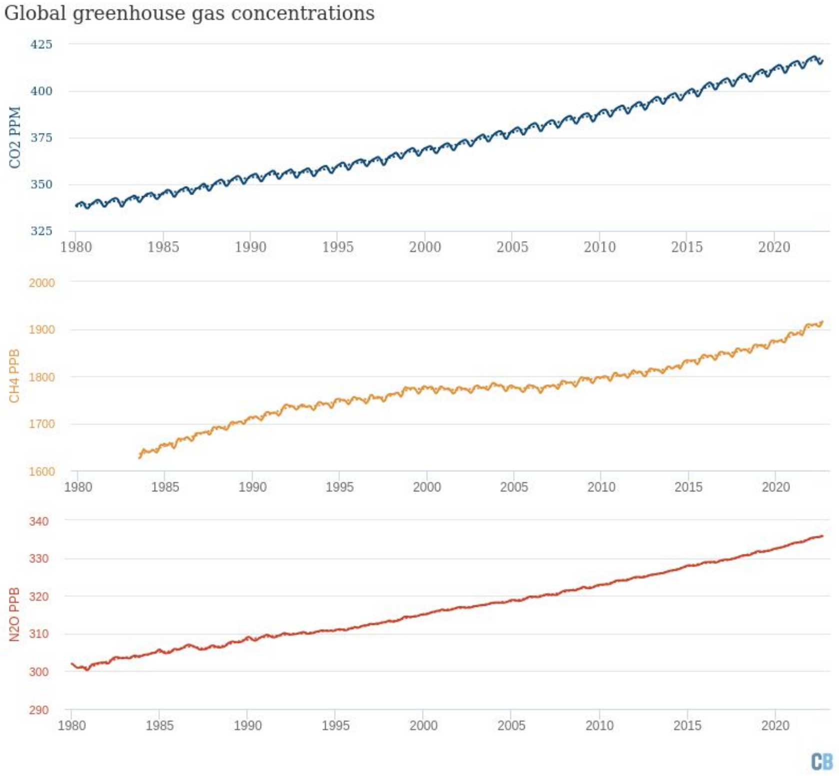
Chemical structure determines the absorption properties of gases

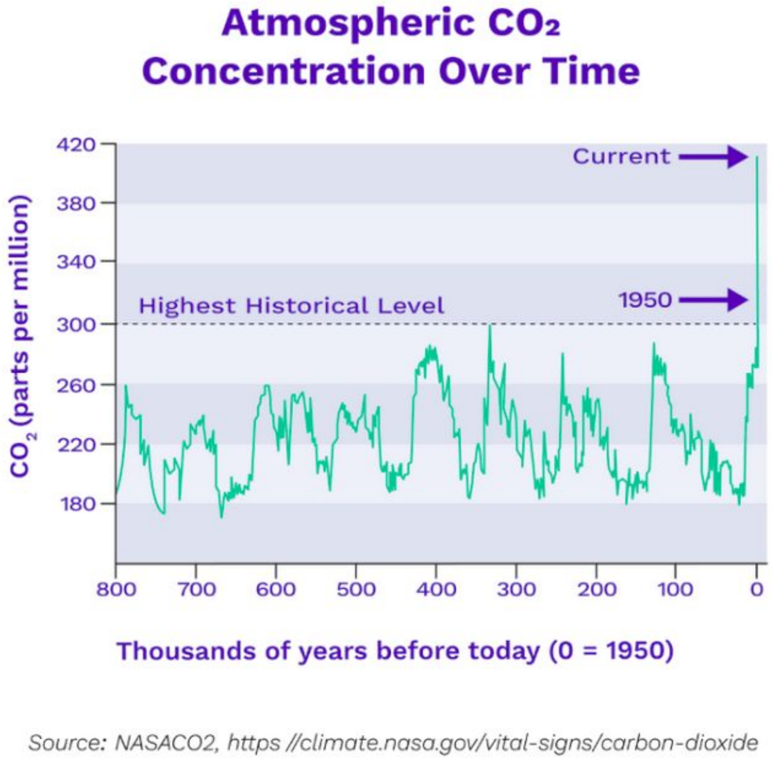


*water vapor has a timescale of only 9 days

The GWP measures the relative warming impact of one tonne of a greenhouse gas compared to one tonne of CO_2 over a given period of time. Different gases have different effects on warming

Humans are rapidly increasing the amount of greenhouse gases in the atmosphere

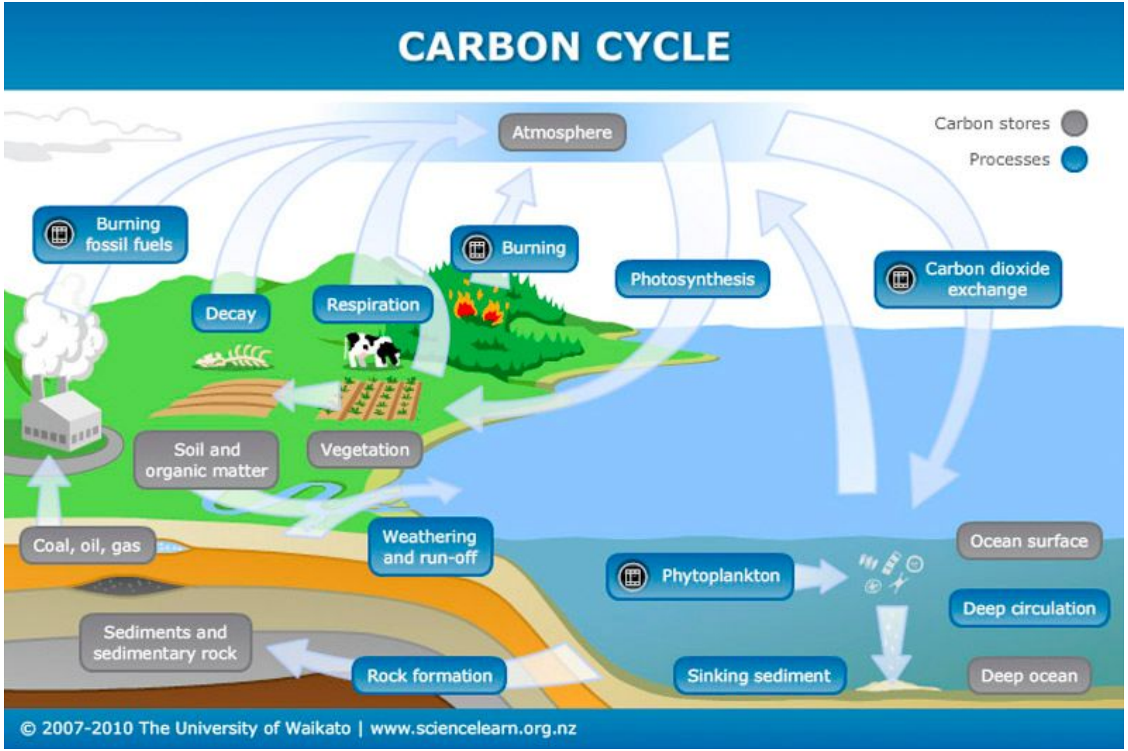




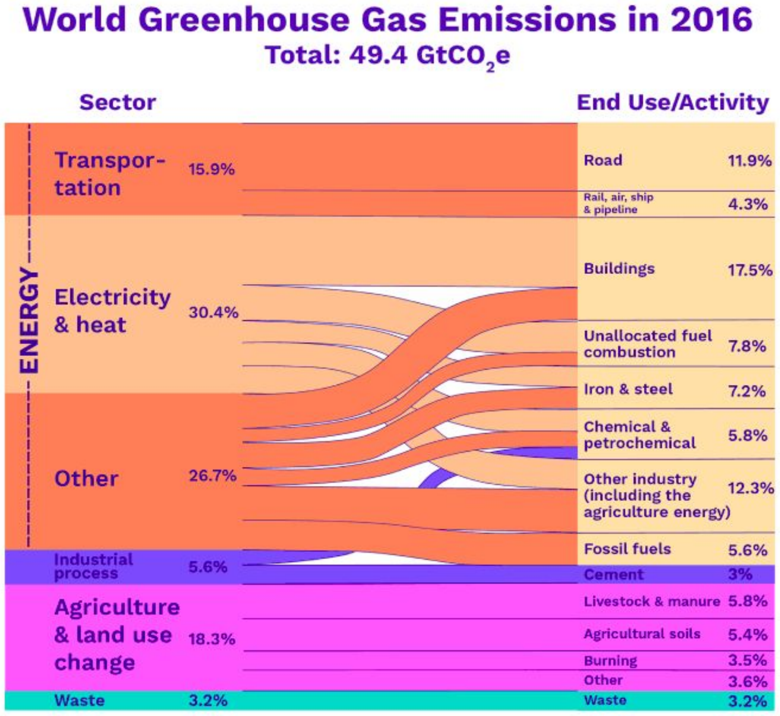
Carbon cycle

Carbon can be stored in the land, ocean, and atmosphere.

Both biological and abiological processes can convert carbon between different stores.

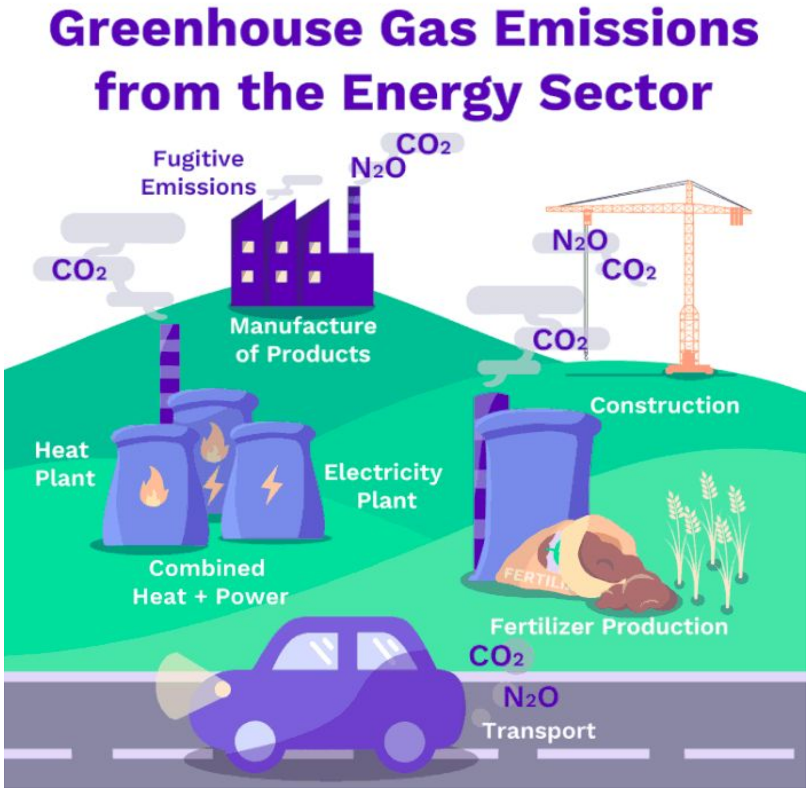


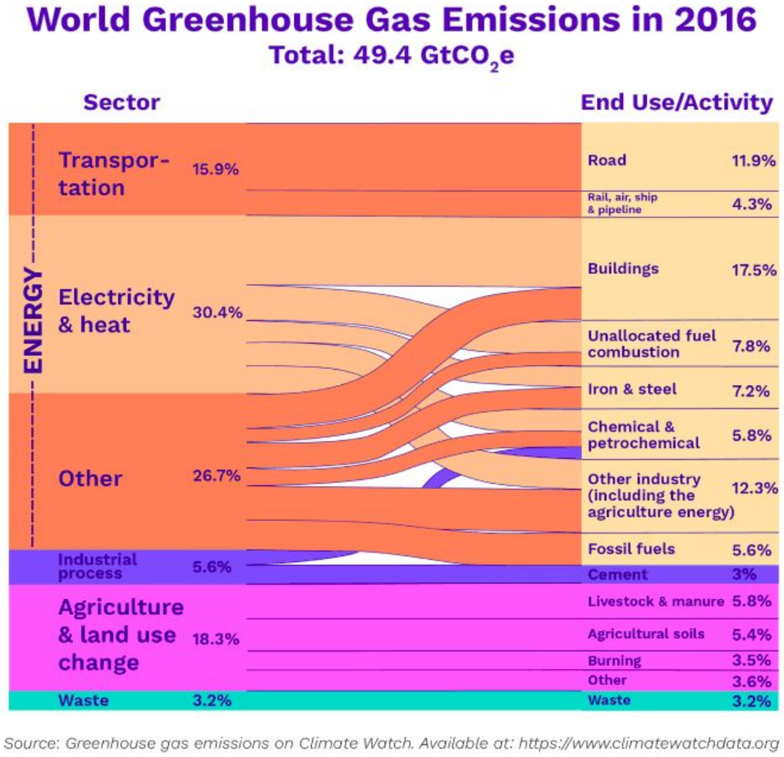
Which activities contribute to GHG emissions?



Source: Greenhouse gas emissions on Climate Watch. Available at: https://www.climatewatchdata.org

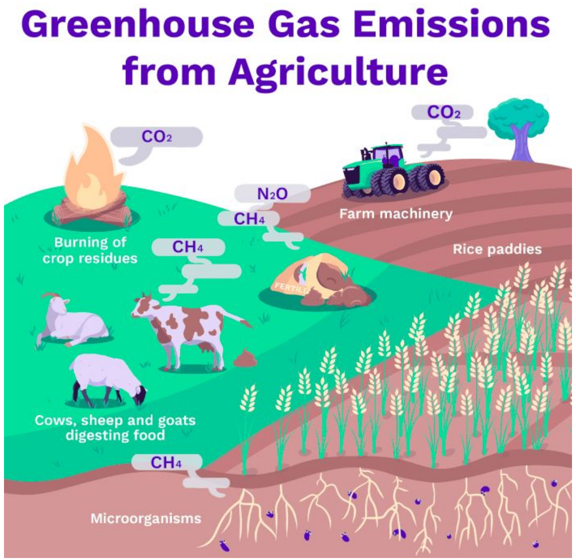
Which activities contribute to GHG emissions?



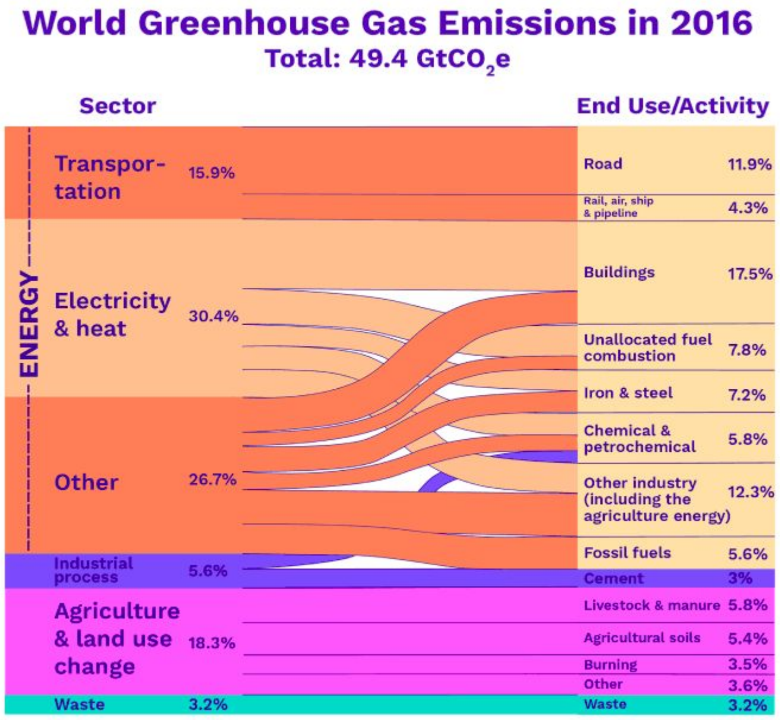


climatescience.org

Which activities contribute to GHG emissions?

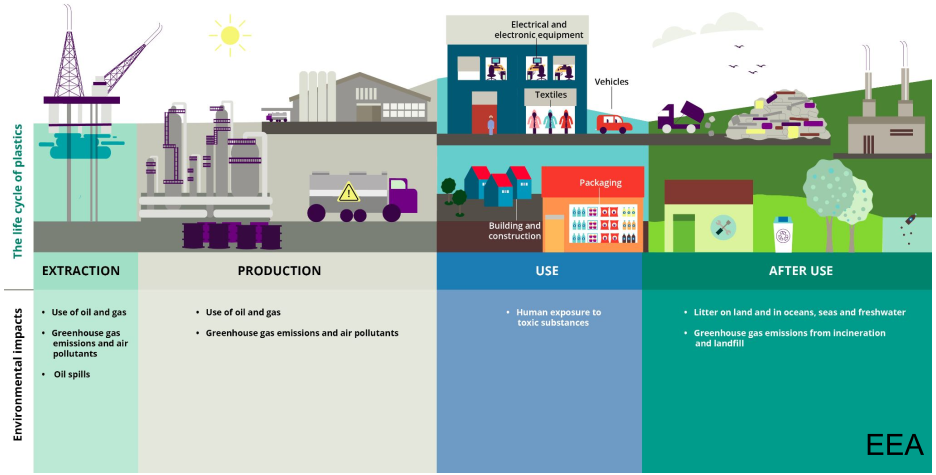


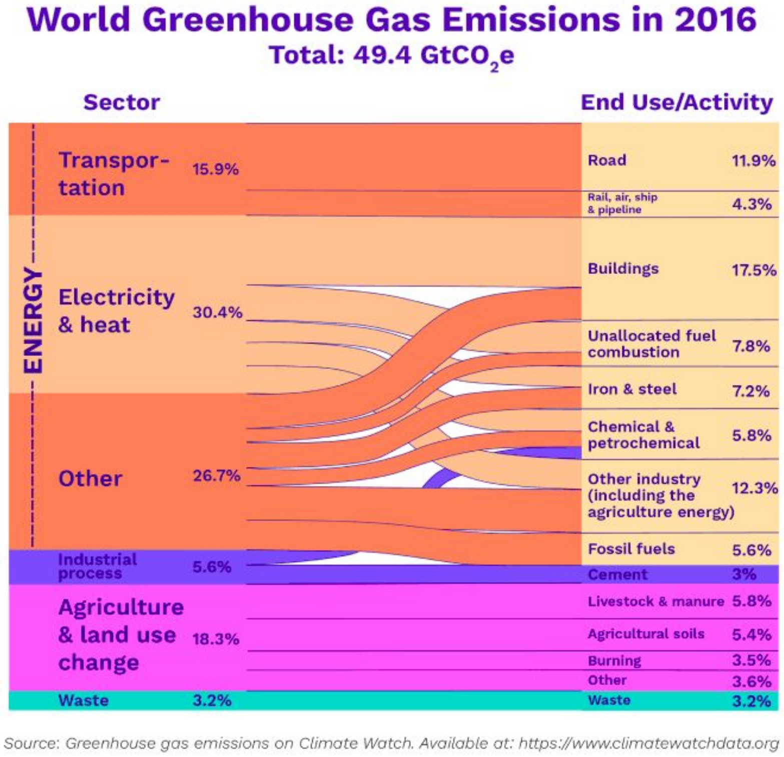
Flooding of rice fields encourages bacteria that decomposes organic matter, leading to up to **12% of global methane emissions**



Source: Greenhouse gas emissions on Climate Watch. Available at: https://www.climatewatchdata.org

Which activities contribute to GHG emissions?





Industrial processes includes emissions from producing cement, chemicals, and various other materials (like plastics, rubber and human-made fabric).This is one of the fastest growing sources of greenhouse emissions and has grown by 203% since 1990.

Which activities contribute to GHG emissions?

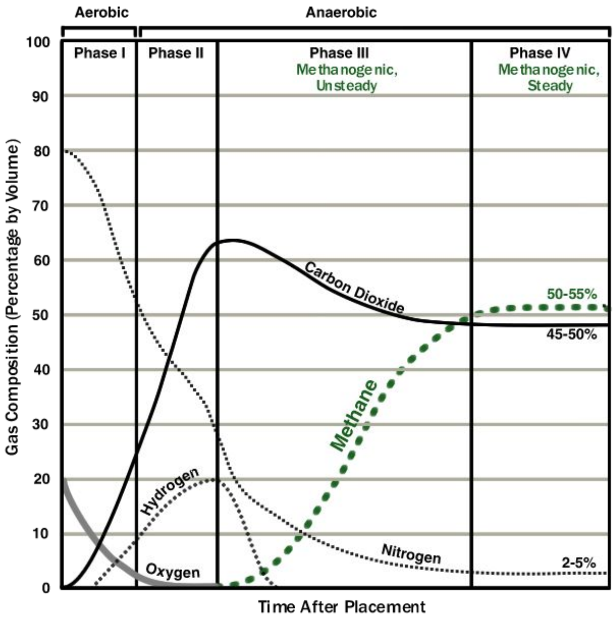
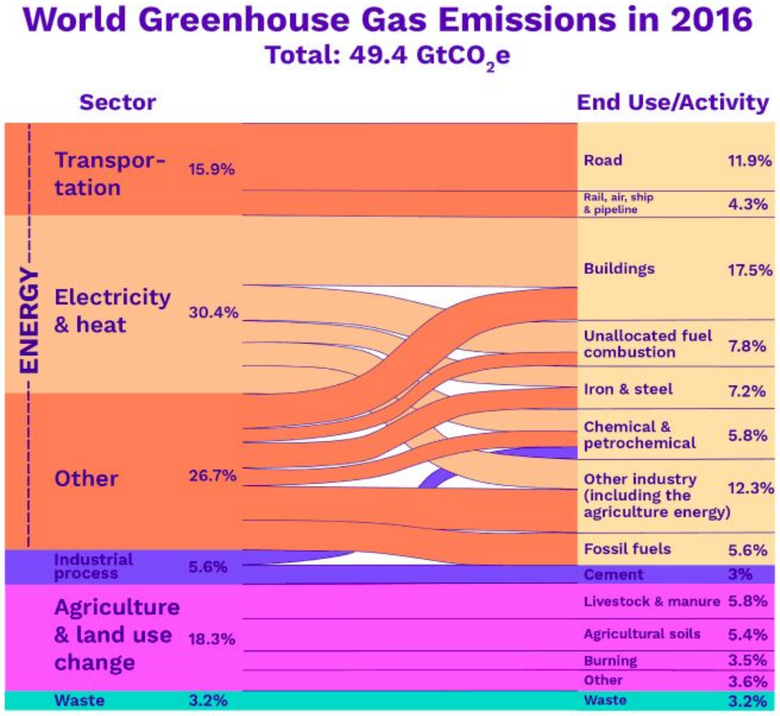


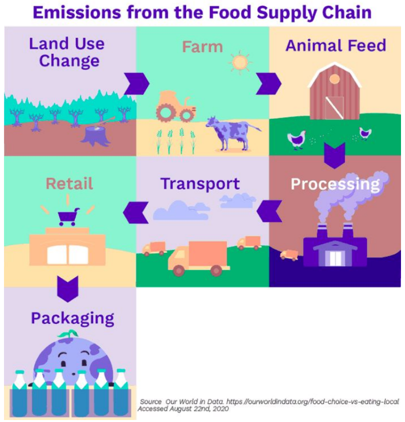
Figure adapted from ATSDR 2008. Chapter 2: Landfill Gas Basics.

Decomposing waste releases a variety of greenhouse gases. Sorting wastewater and treating human sewage also contributes.

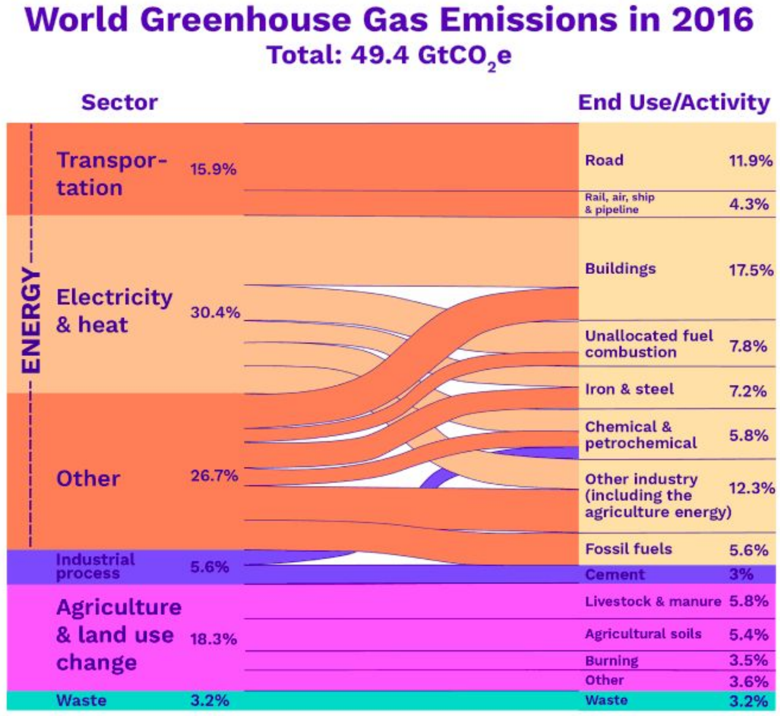


Source: Greenhouse gas emissions on Climate Watch. Available at: https://www.climatewatchdata.org

Which activities contribute to GHG emissions?



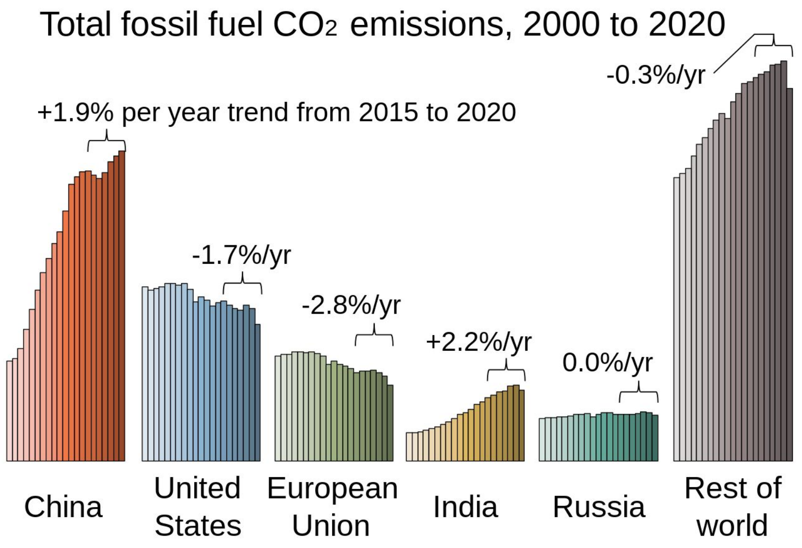
Source: European Parliament.
https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/633143/EPRS_BRI(2019)633143_EN.pdf
Accessed 20 October 2020.

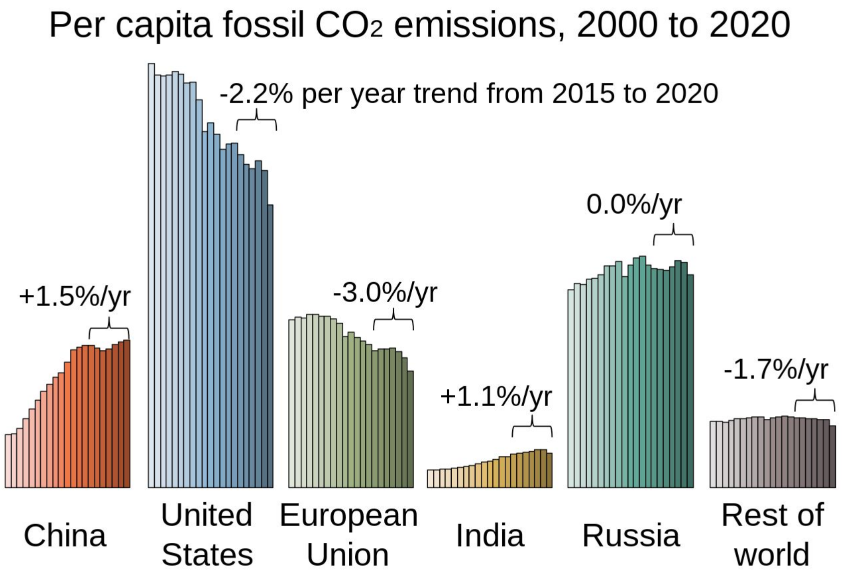


Source: Greenhouse gas emissions on Climate Watch. Available at: https://www.climatewatchdata.org

Calculating GHGs in terms of end product can be more effective for encouraging individual change.

Where are GHGs coming from?

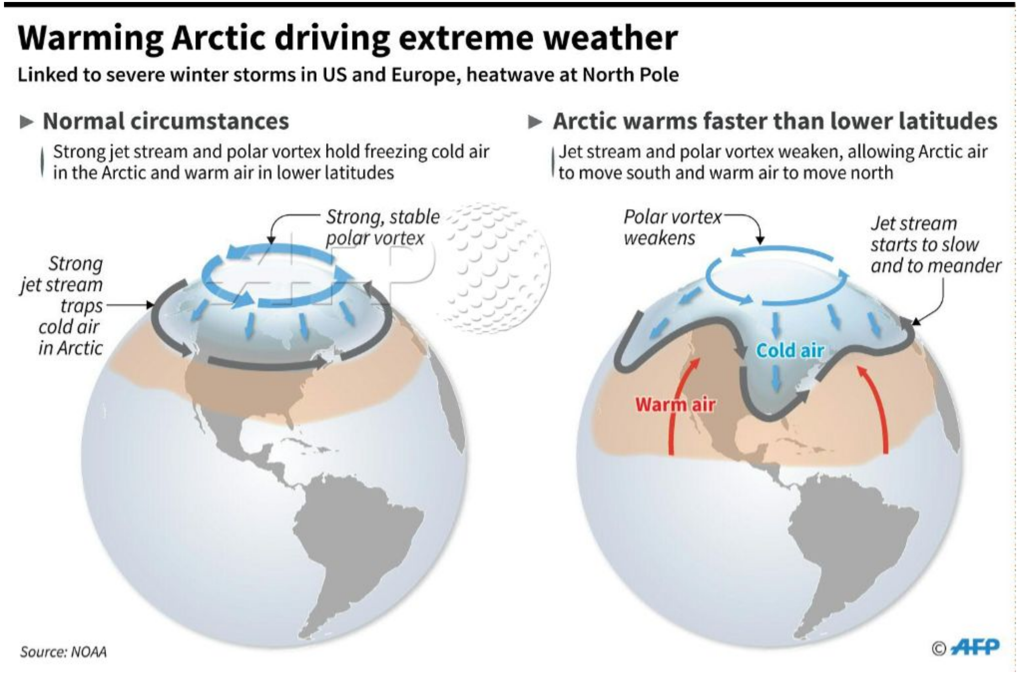




global carbon product

What does warming do?

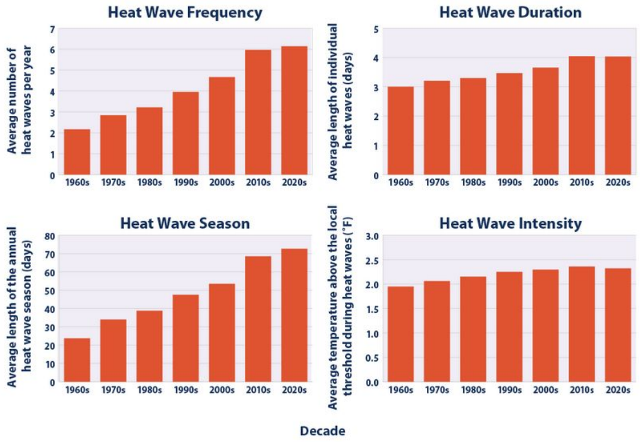
Changing the amount of heat in the atmosphere has impacts on global climate patterns and precipitation



EPA

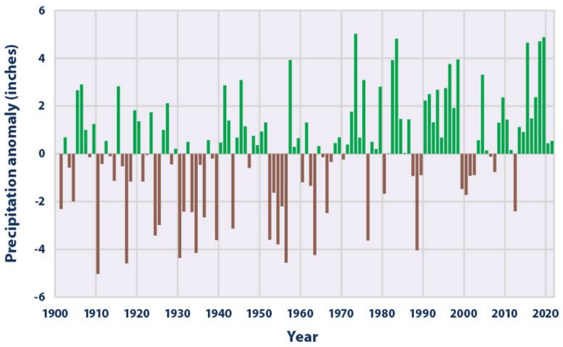
Climate change has already increased extreme events

Figure 1. Heat Wave Characteristics in the United States by Decade, 1961–2021



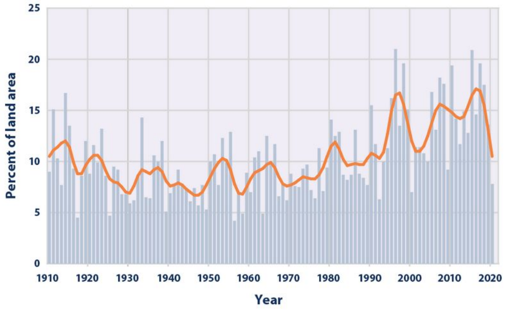
This indicator describes trends in average precipitation for the United States and the world.

Figure 1. Precipitation in the Contiguous 48 States, 1901–2021



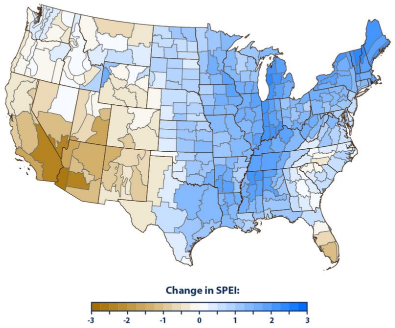
This indicator tracks the frequency of heavy precipitation events in the United States.

Figure 1. Extreme One-Day Precipitation Events in the Contiguous 48 States, 1910–2020



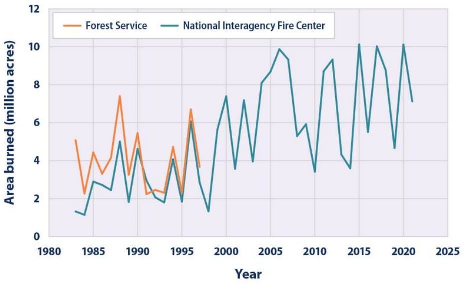
This indicator measures drought conditions of U.S. lands.

Figure 3. Average Change in Drought (Five-Year SPEI) in the Contiguous 48 States, 1900–2020

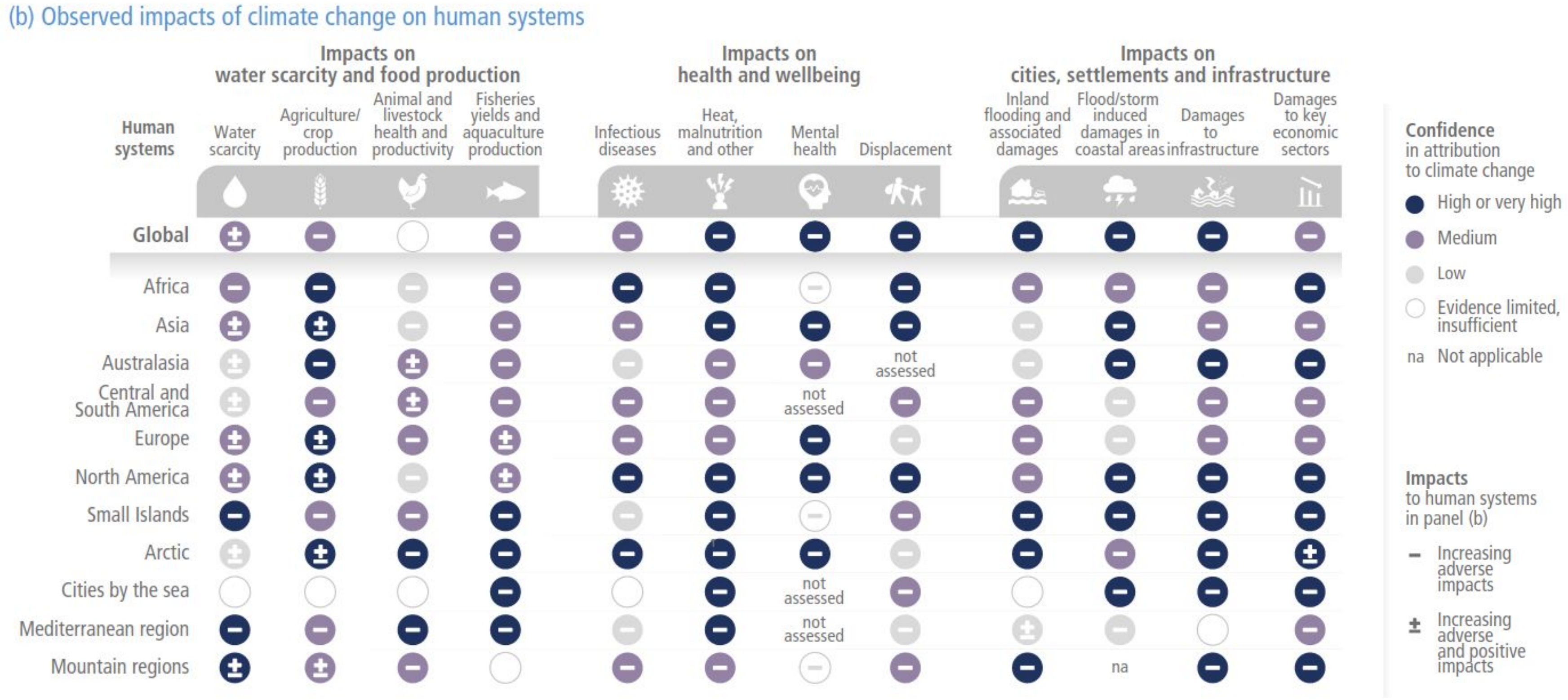


This indicator tracks the frequency, extent, and severity of wildfires in the United States.

Figure 2. Wildfire Extent in the United States, 1983–2021



Climate change has already impacted human systems



What’s in store for the future?

How 2 degree warming will play out in different locations

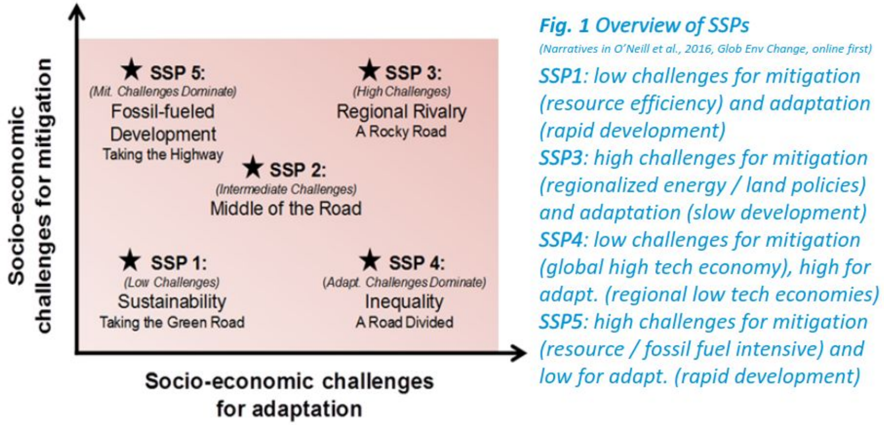
1:33-5:50

Paris Agreement

- International treaty on climate change adopted in 2015
- Aims are to keep the rise in mean global temperature to well below 2 °C above pre-industrial levels (preferably to 1.5 °C), to increase the ability to adapt to climate change, and to encourage finance flows to support emissions reductions.
- Different scenarios have been devised to understand how to achieve these goals.

Shared Socio-Economic Pathways (SSPs)

- 5 narrative scenarios, representing different approaches and resulting challenges
- SSPs include social, economic, and governmental forces and challenges
- Provide a framework for making predictions of possible futures



The Intergovernmental Panel on Climate Change

The Intergovernmental Panel on Climate Change (IPCC) is the United Nations body for assessing the science related to climate change.

Reports

The IPCC prepares comprehensive Assessment Reports about the state of scientific, technical and socio-economic knowledge on climate change, its impacts and future risks, and options for reducing the rate at which climate change is taking place. It also produces Special Reports on topics agreed to by its member governments, as well as Methodology Reports that provide guidelines for the preparation of greenhouse gas inventories. The IPCC is working on the Sixth Assessment Report which consists of three Working Group contributions and a Synthesis Report. The Working Group I contribution was finalized in August 2021, the Working Group II contribution in February 2022 and the Working Group III contribution in April 2022.

Working Group 1
The Physical Science Basis

This report focuses on how and why the world's climate has changed in the past, and how it is projected to change in the future.

Working Group 2
Impacts, Adaptation, and Vulnerability

This report focuses on how climate change affects people, our built systems, and the natural world. It also addresses how we can adapt and become more resilient to climate change.

Working Group 3
Mitigation of Climate Change

This report focuses on the actions we can take to reduce future climate change and prevent it from becoming too extreme.

Scenarios considered in recent IPCC reports

SSP1-1.9 for the 1.5°C Paris Agreement goal

SSP1-2.6 for sustainable pathways

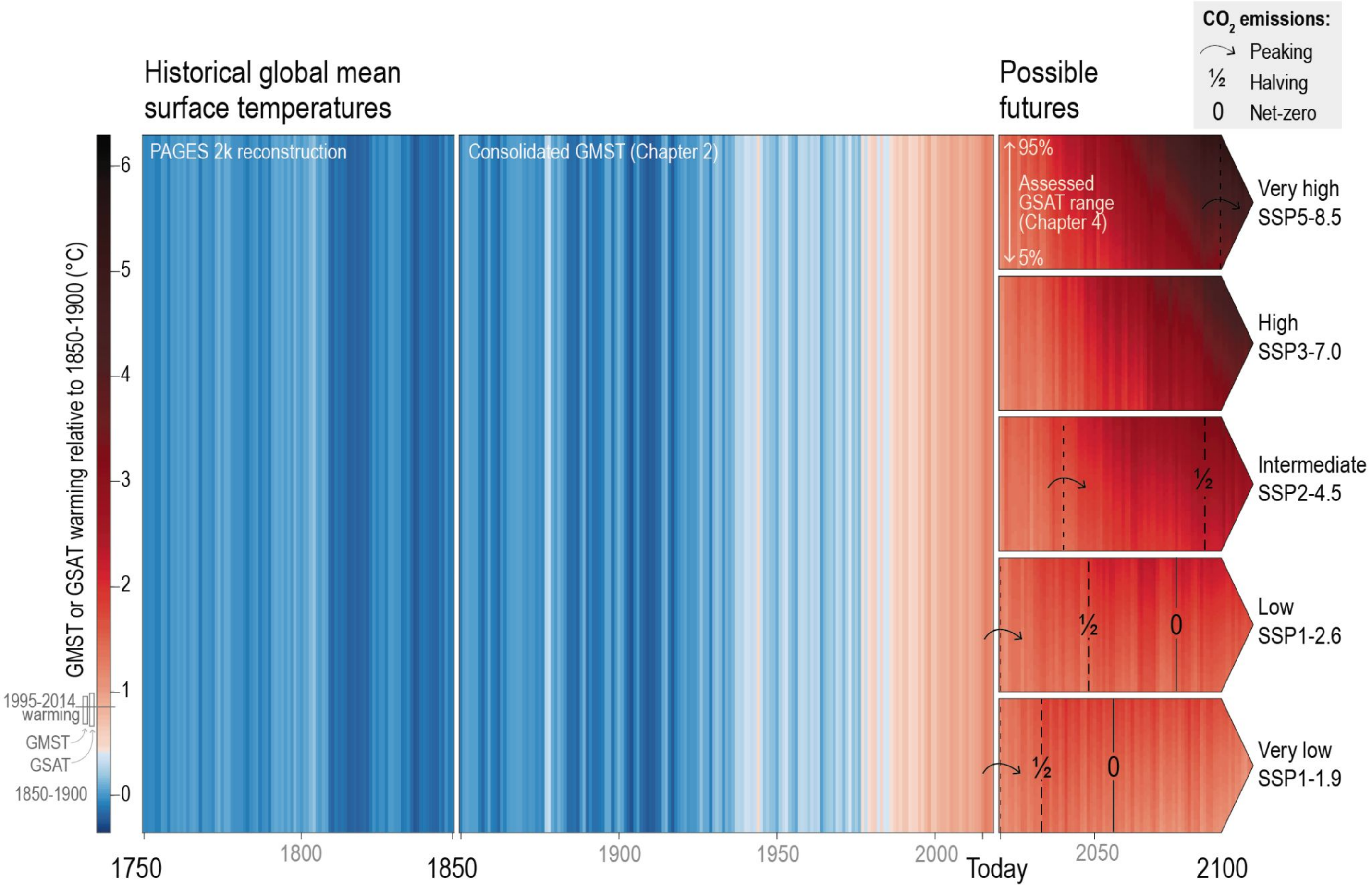
SSP2-4.5 for middle-of-the-road

SSP3-7.0 for regional rivalry

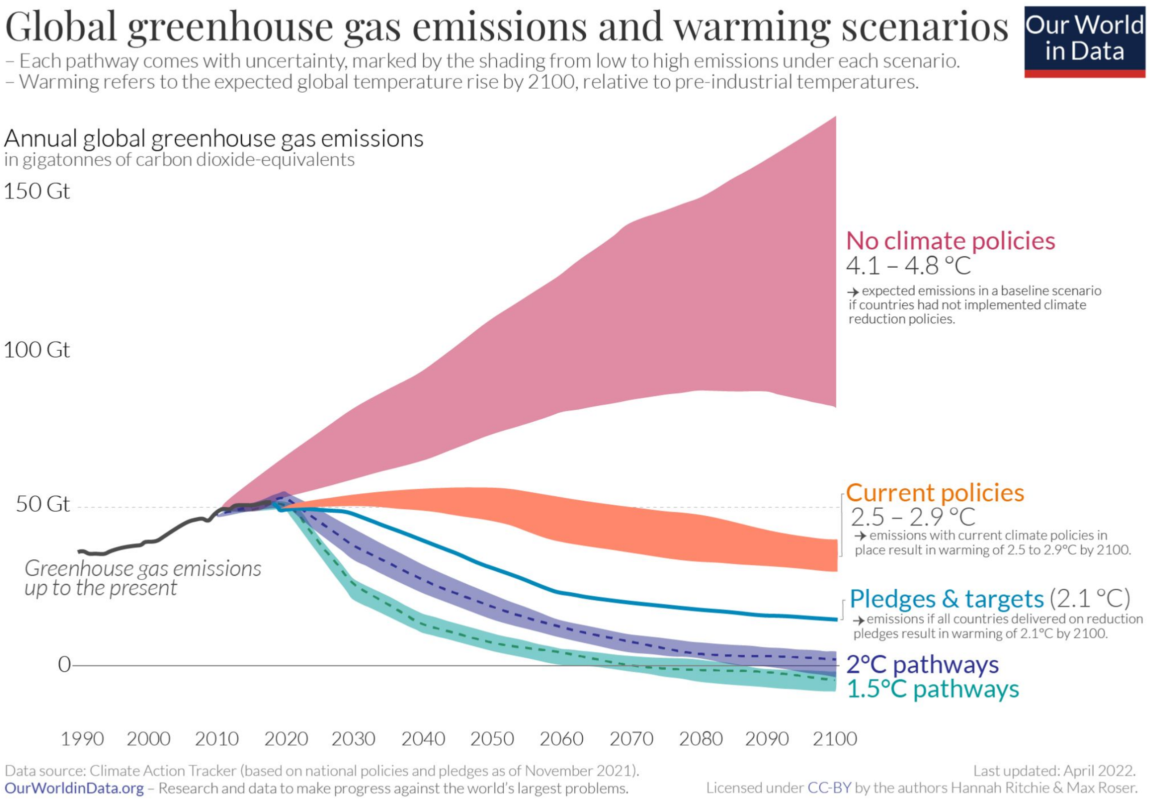
SSP5-8.5 for fossil fuel-rich development.

(Labels indicate SSP type and amount of radiative forcing)

Possible futures



Possible futures



Discussions of temperature increases are usually in degrees celsius relative to pre-industrial averages. As of 2022, we were at an average of +1.2C.

2024 had an average of +1.55C.

What do we need to do?

Mitigation - attempts to reduce climate change (mainly by reducing GHG emissions)

Adaptation - efforts to prepare humans and human infrastructure for a changing climate

Recommendations from IPCC reports

More laws, activism, and policy:

Following current climate pledges to 2030 would make it “impossible” to limit warming to 1.5C with “no or limited overshoot” – and “strongly increas[e] the challenge” for 2C. (most likely path under current policy leads to 2.5-3C by 2100)

Although at least 90% of global GHG emissions are covered by climate targets, only 53% are covered by “direct” climate laws.

Accelerated climate action is “critical” to achieving sustainable development.

Carbon Brief

Recommendations from IPCC reports

No more fossil fuels:

All scenarios limiting warming to 2C or below include “greatly reduced” fossil fuel use, with unabated coal being “completely” phased out by 2050.

The world can emit just 460 gigatonnes more of carbon dioxide, measured from the start of 2020, if we want at least a 50 percent chance of staying below 1.5 degrees. In recent years, the world has emitted about 36.4 gigatonnes annually. **If we continue at that pace, we will blow our entire carbon budget in about a decade.**

Carbon Brief

Recommendations from IPCC reports

In real terms:

“[The report is a] file of shame, cataloguing the empty pledges that put us firmly on track towards an unlivable world”.

“Climate activists are sometimes depicted as dangerous radicals. But the truly dangerous radicals are the countries that are increasing the production of fossil fuels…[The report] sets out viable, financially sound options [for cutting emissions] in every sector that can keep the possibility of limiting warming to 1.5C alive.”

-UN secretary-general António Guterres

Furthermore, the global economic benefit of limiting warming to 2C is reported to exceed the cost of mitigation in most of the assessed literature.

Carbon Brief

What you’ll learn in this course...

is a little bit of everything about how the world works.

15 min break

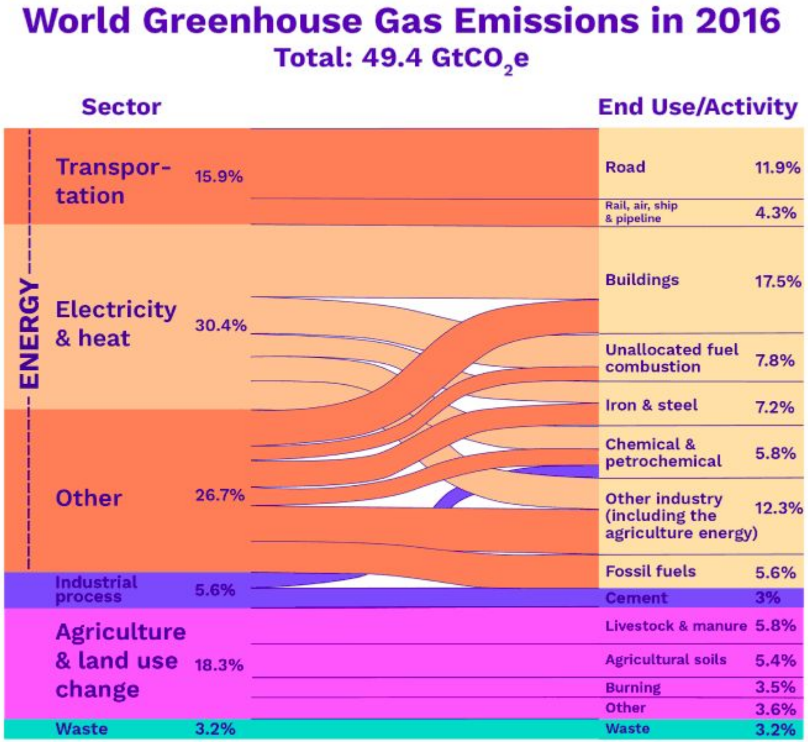

Vox

Today:

Climate change domain: energy use and efficiency

ML technique: Regression techniques

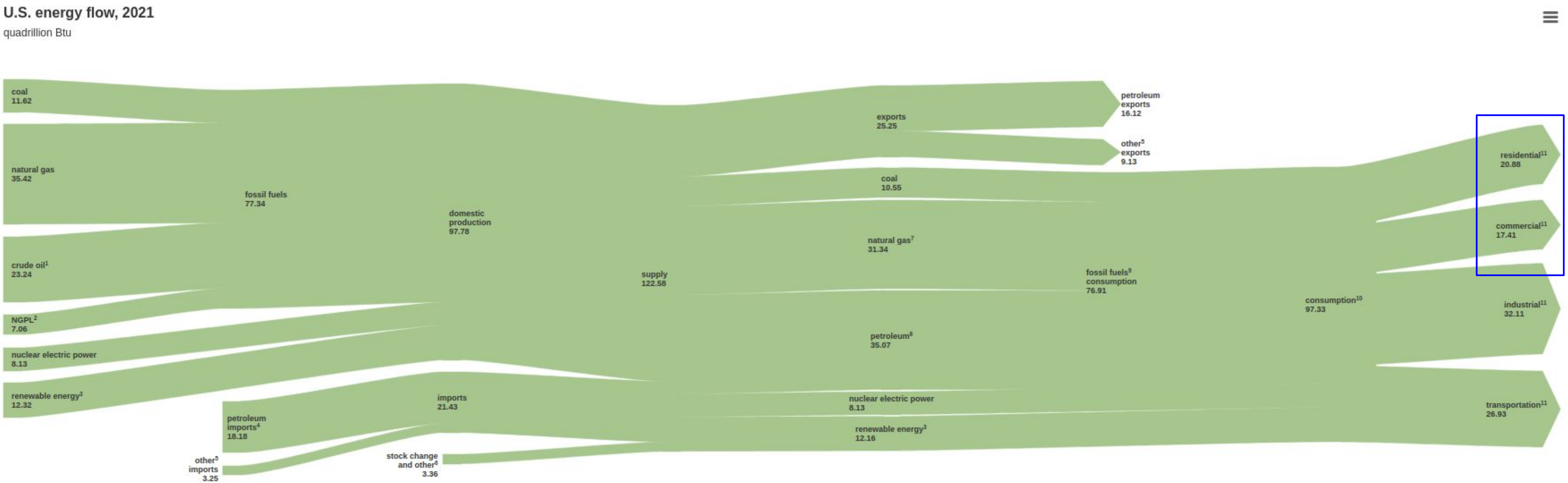
Energy efficiency as a way to reduce GHGs



Source: Greenhouse gas emissions on Climate Watch. Available at: https://www.climatewatchdata.org

Buildings and Energy

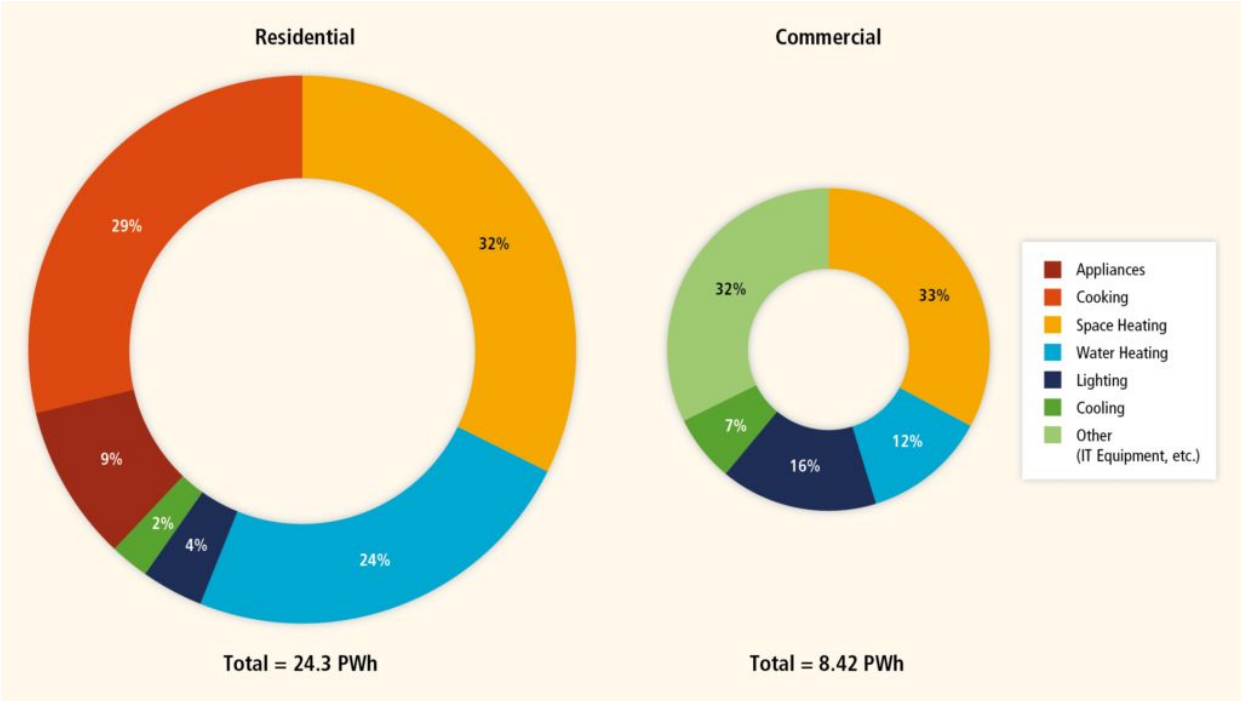
EIA



Residential and commercial building energy consumption makes up 40% of the total energy consumption in the United States.

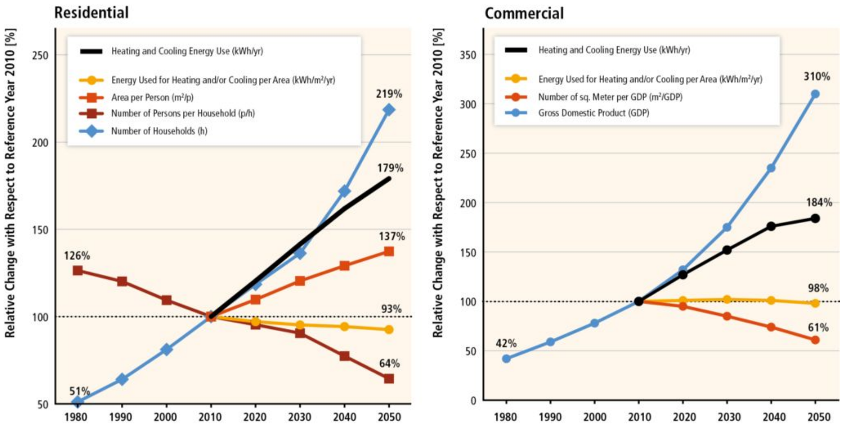
Energy is measured in British thermal unit, defined as the amount of heat required to raise the temperature of one pound of water by one degree Fahrenheit

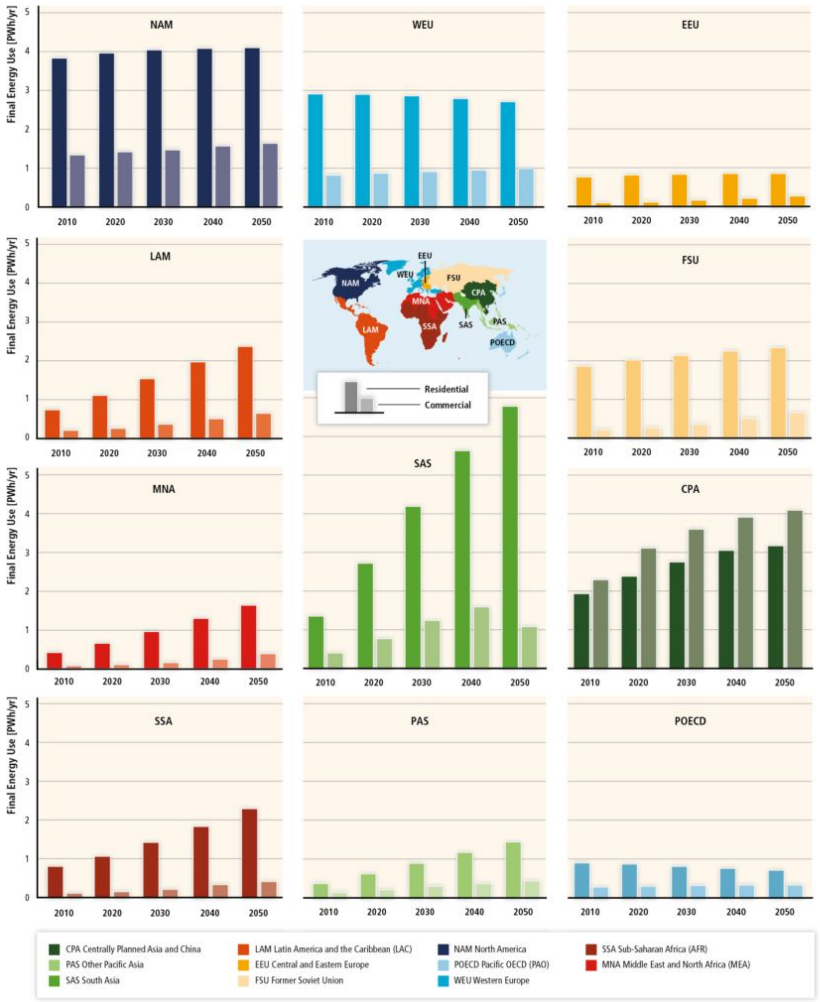
How is that energy being used?



Future Projections

IPCC





How to increase building efficiency

post-war gas low-rise

Energy Conservation Measure	Anytime/ Anywhere	Midcycle Retrofit	Refinancing/ Substantial Retrofit	Tenant Turnover	Payback (years)	Cost per SF	Energy Savings per SF
			Building Touchpoint				
Install Exhaust Fan Timers	•	•	•		5.0	$	
Install Submetering		•	•		2.0	$$$	
Install Solar/Photovoltaic			•		17.0	$$$$	
Upgrade Motors		•	•		5.5	$$	
Upgrade Lights		•	•		2.5	$	
Install Lighting Sensors		•	•		4.0	$	
Upgrade Burner			•		6.5	$$	
Upgrade Boiler			•		>20	$$$$	
Install TRVs and Zone Control		•	•		6.5	$$$	
Install Heating Controls and Thermostats		•	•		2.5	$$	
Insulate Condensate Tank	•	•	•		2.5	$	
Replace or Repair Steam Traps	•	•	•		3.5	$$	
Insulate Pipes	•	•	•		2.0	$	
Install or Upgrade Master Venting		•	•		3.0	$$	
Replace Windows and Glazing			•		>20	$$$$	
Increase Wall Insulation			•		>20	$$$$	
Increase Roof Insulation			•		>20	$$$	
Complete Air sealing	•	•	•	•	6.0	$$	
Separate DHW from Heating			•		6.5	$$$	
Install Low-Flow Showerheads	•	•	•	•	1.0	$$	
Install DHW Controls	•	•	•		0.5	$	
Install Low Flow Aerators	•	•	•	•	1.5	$$	
Insulate Pipes and Tank	•	•	•		6.0	$	

Energy Conservation Measure

- Ventilation & Cooling
- Other
- Lighting
- Heating Equipment
- Heating Distribution
- Envelope
- Domestic Hot Water

Cost per Square Foot

$	<$.05
$$	$0.05-$0.25
$$$	$0.26-$1.00
$$$$	>$1.00

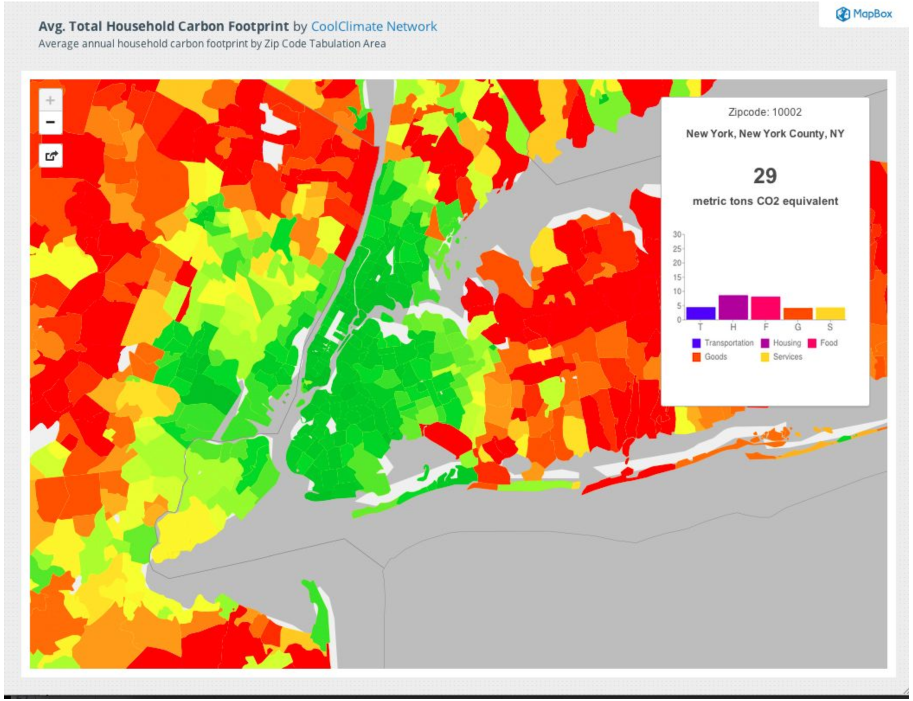
Energy Savings per SF (kBtu)

0–3
3.1–8
8.1–12
>12

Notes

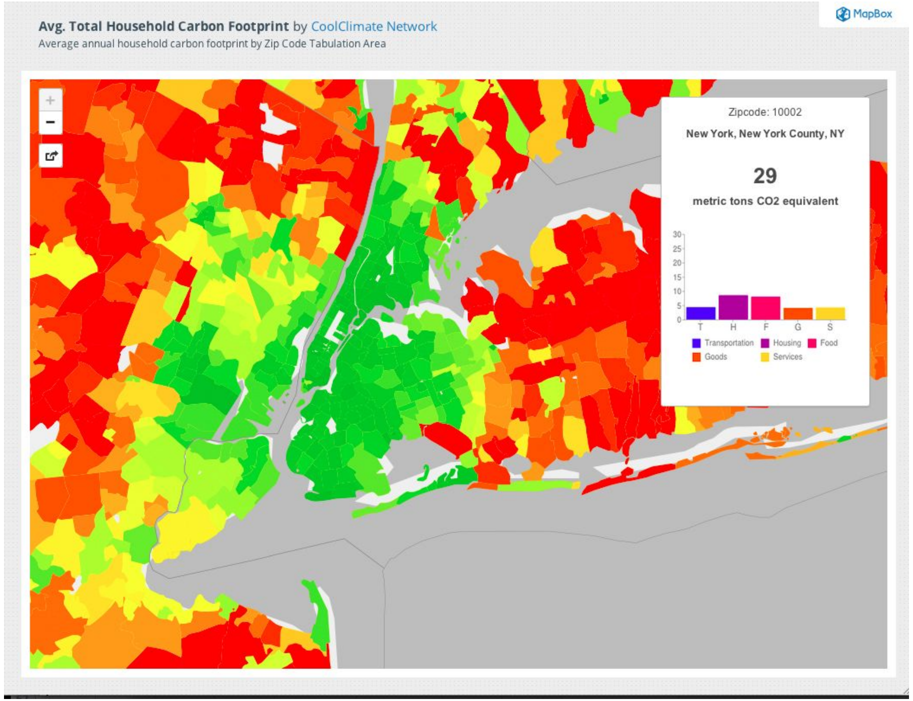
This list of Energy Conservation Measures (ECM) is based on LL87 audit data and therefore may be incomplete. Suggested ECMs for each Building Touchpoint are representative, and not necessarily applicable to every building. Variety in specific building systems and condition of equipment must be considered in determining the appropriate packages of ECMs for individual buildings. The first step of any upgrade should be to work with a qualified service provider to develop a scope of work appropriate for your building.

Urban vs Suburban Energy Consumption



Low density suburban development is 2.0–2.5× as energy and GHG emissions intensive as high-density urban core development per capita.

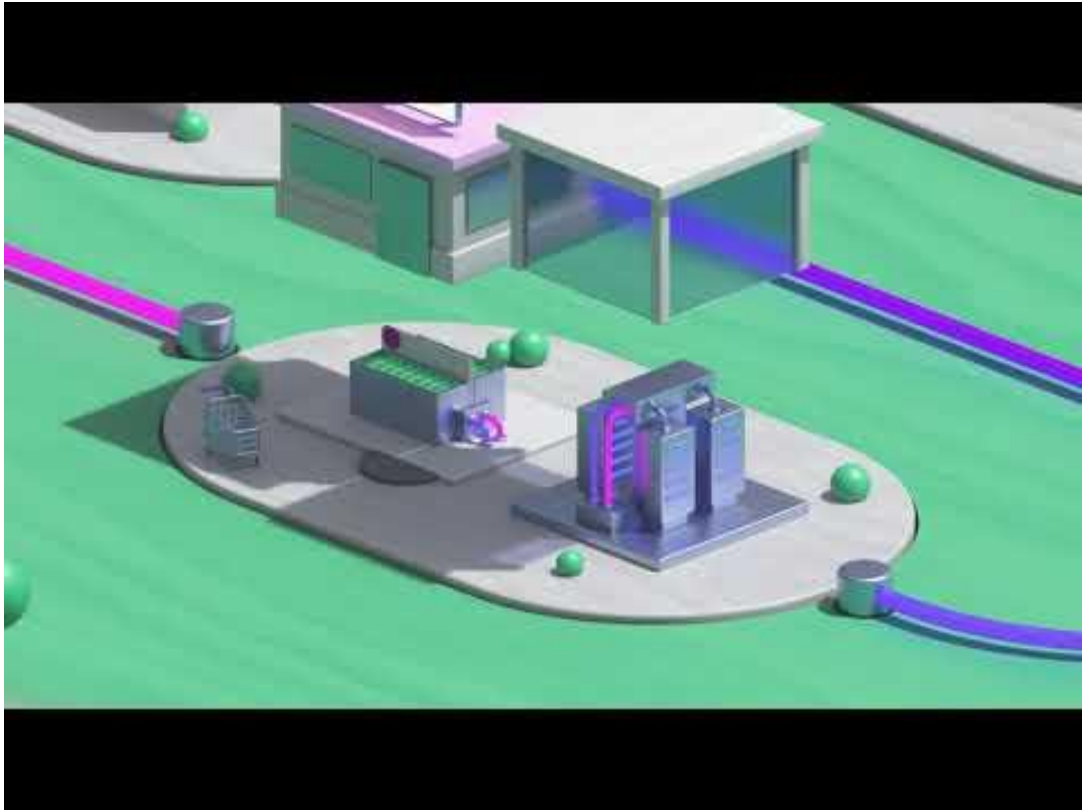
Urban vs Suburban Energy Consumption



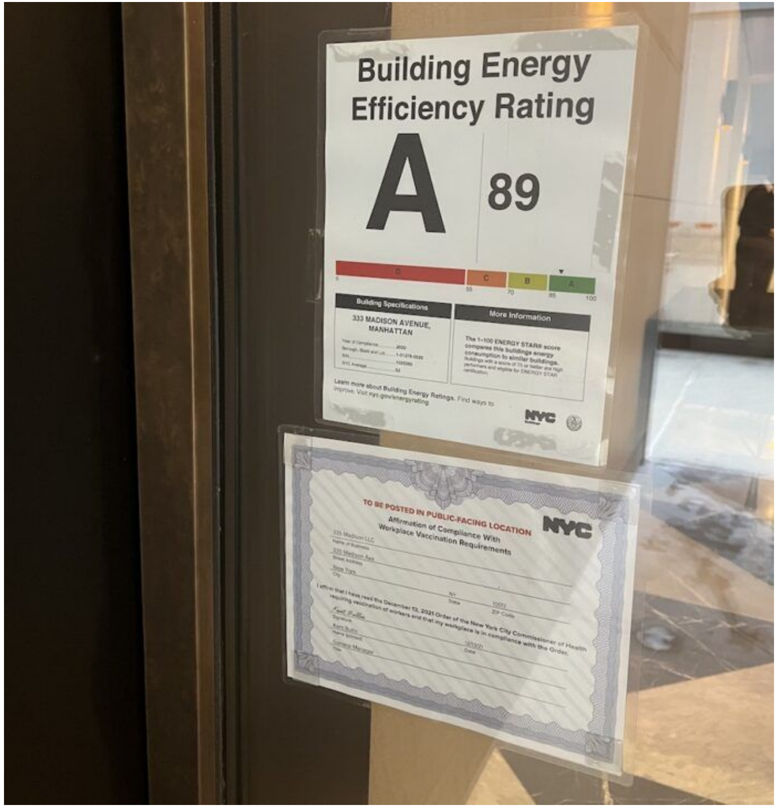
With shared resources, shared walls and generally smaller square footage, households in buildings with five or more units consume only 38 percent of the energy of households in single-family homes (Brown et al., 2005).

At a suburban density of four homes per acre, carbon dioxide emissions per household were found to be 25 percent higher than in an urban neighborhood with 20 homes per acre (Mazza, 2004).

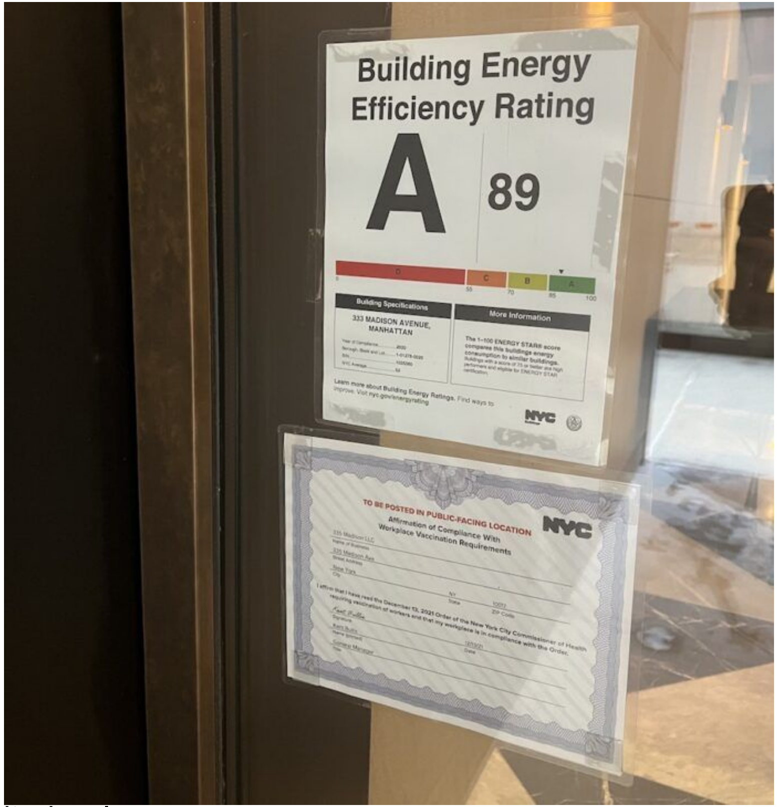
Benefit of urban density: district heating



NYC Building Efficiency Grades



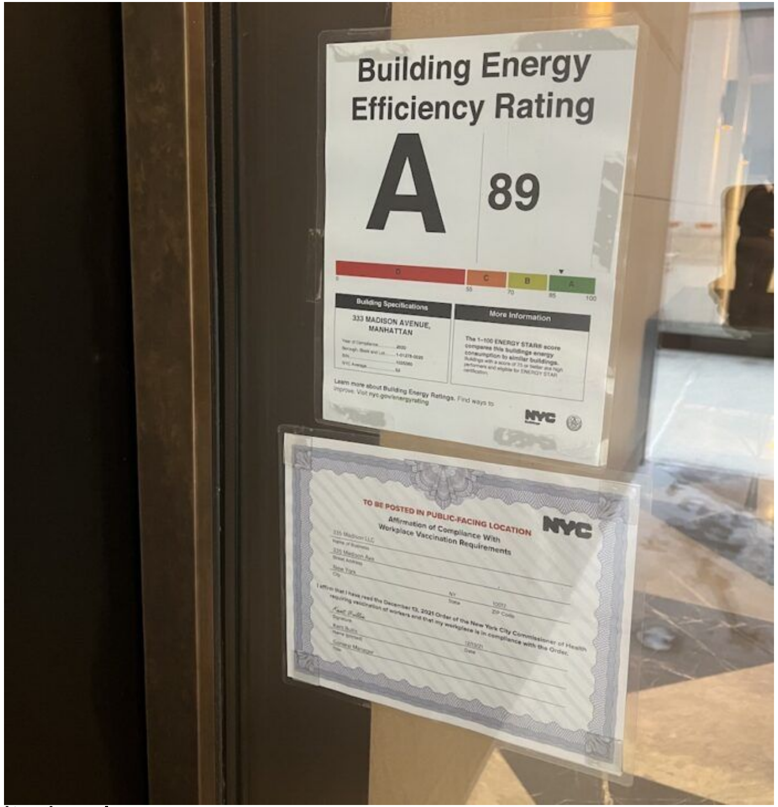
NYC Building Efficiency Grades



www.citysignal.com

Buildings in New York City are responsible for 70% of emissions

NYC Building Efficiency Grades



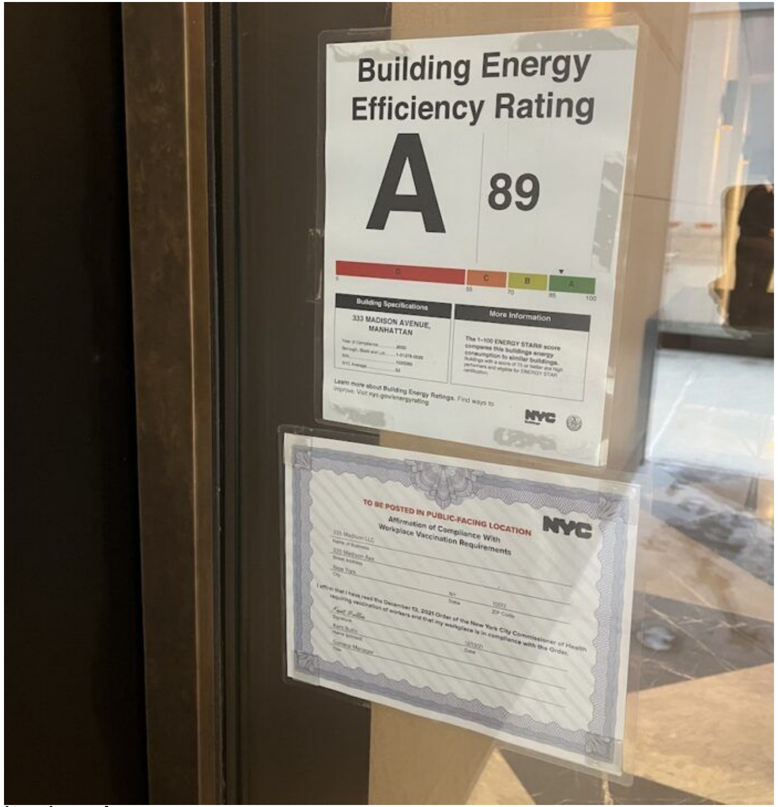
www.citysignal.com

Local Law No. 95 requires all buildings to have an efficiency rating based on Energy Star and EPA guidelines. In October 2020, an additional rule was added that larger buildings had to have these scores posted on their main entrances. That's around 50,000 buildings and nearly two-thirds of the building area in the city.

The rankings come from Energy Star and are based on energy consumption, water consumption, and greenhouse gas emissions.

Buildings must submit energy information to the government for a 12 month period of time, and a list of fuels burned and converted on site. These criteria add together for a total score of anywhere between 1-100.

NYC Building Efficiency Grades



www.citysignal.com

Of the approximately 40,000 buildings that submitted reports when the law was first implemented, **about half of them received a D**. Thousands submitted nothing, receiving F’s. Buildings like The New York Stock Exchange and Trump Tower received some of the lowest scores, while the Flatiron Building and the Empire State Building did rather well.

In 2024 Local Law No. 97 went into effect, which **fines buildings with lower scores**. Depending on how low the scores are, the buildings could receive fines as high as hundreds of thousands of dollars. This has incentivized many buildings to change their energy consumption and distribution.

How can we know how much energy buildings consume?

(this is the focus of the paper you will read)

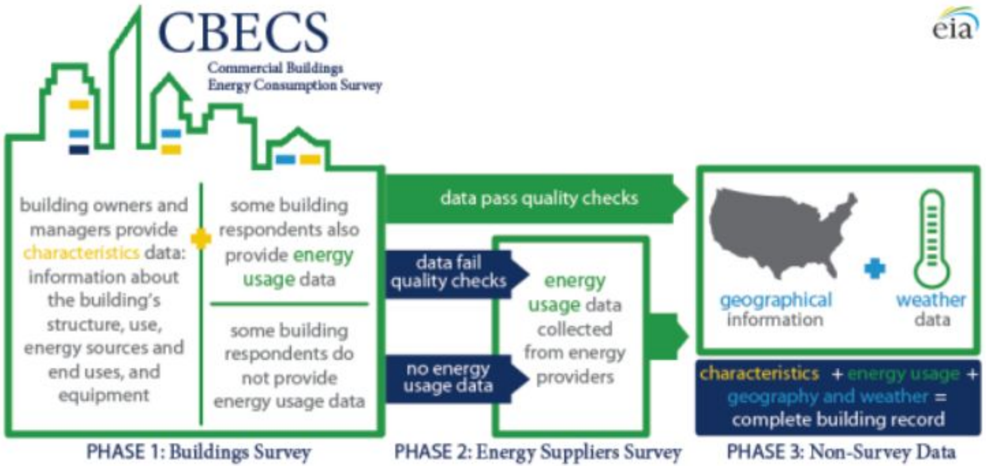
NYC Local Law 84 Dataset

The **New York City Benchmarking Law, known as Local Law 84** (LL84), requires buildings that are over 50,000 square feet, or lots with buildings with over 100,000 square feet combined, to ***report their annual energy and water consumption values*** in a standardized manner using the EPA's portfolio manager database.

This consumption data, along with some of the building characteristics (such as: total square feet, year built, primary building activity, and energy use intensity), have been released annually since 2011.

What about outside NYC?

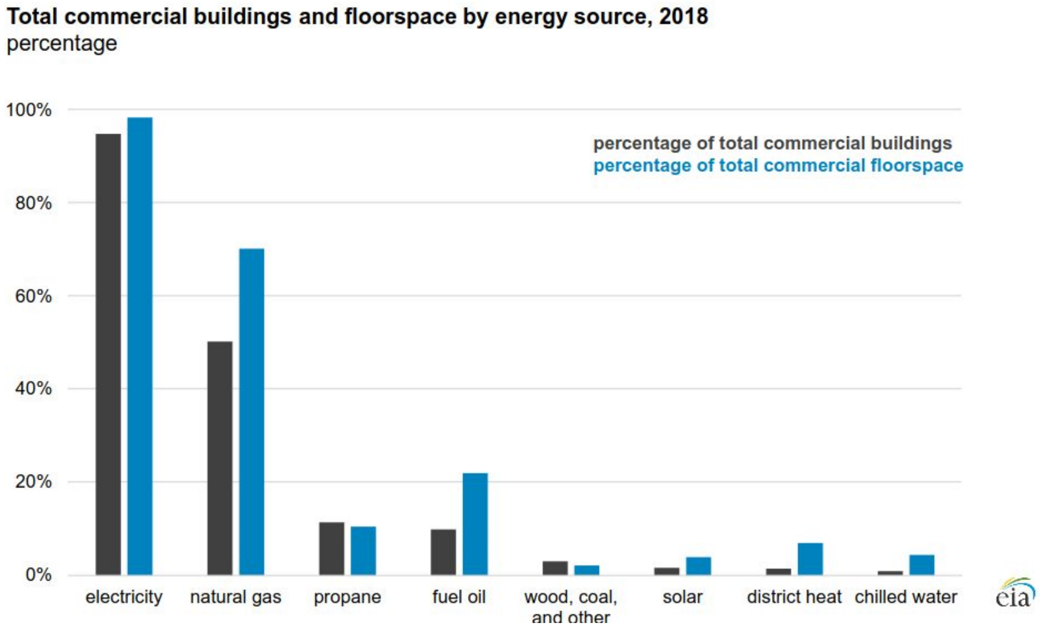
Commercial Buildings Energy Consumption Survey



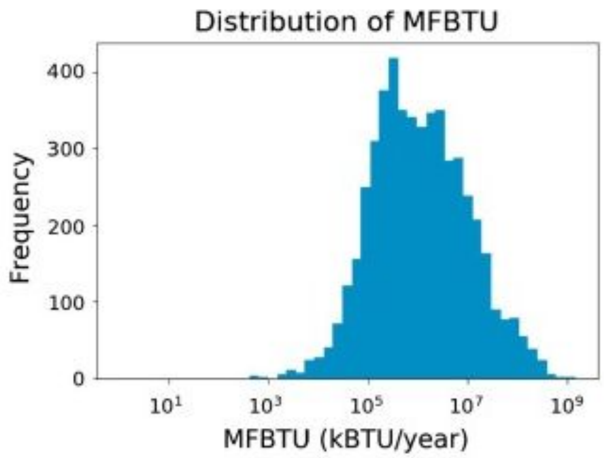
- The only independent, statistically representative source of national-level data on the characteristics and energy use of commercial buildings
- Building characteristics collected through an in-person or web survey of managers of 6,436 buildings, representing 5.9 million buildings in the United States.
- Energy usage data collected from suppliers of electricity, natural gas, fuel oil, and district heat

Also: number of workers, hours of operation, year of construction, etc.

Energy information



-
-



“Heating” and “Cooling” Degree Days

Heating and Cooling Degree Days

Degree days are based on the assumption that when the outside temperature is 65°F, we don't need heating or cooling to be comfortable. Degree days are the difference between the daily temperature mean, (high temperature plus low temperature divided by two) and 65°F. If the temperature mean is above 65°F, we subtract 65 from the mean and the result is ***Cooling Degree Days***. If the temperature mean is below 65°F, we subtract the mean from 65 and the result is ***Heating Degree Days***.

Example 1: The high temperature for a particular day was 90°F and the low temperature was 66°F. The temperature mean for that day was:

(90°F + 66°F) / 2 = 78°F

Because the result is *above* 65°F:

78°F - 65°F = 13 ***Cooling Degree Days***

Example 2: The high temperature for a particular day was 33°F and the low temperature was 25°F. The temperature mean for that day was:

(33°F + 25°F) / 2 = 29°F

Because the result is *below* 65°F:

65°F - 29°F = 36 ***Heating Degree Days***.

What if you have a building that is outside New York and not sampled by the CBECS?

Use a statistical model to make an informed guess!

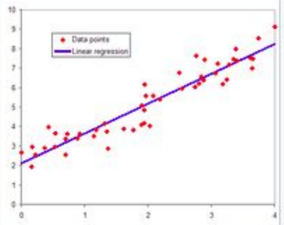
Regression analysis

50 languages

Article Talk Read Edit View history Tools

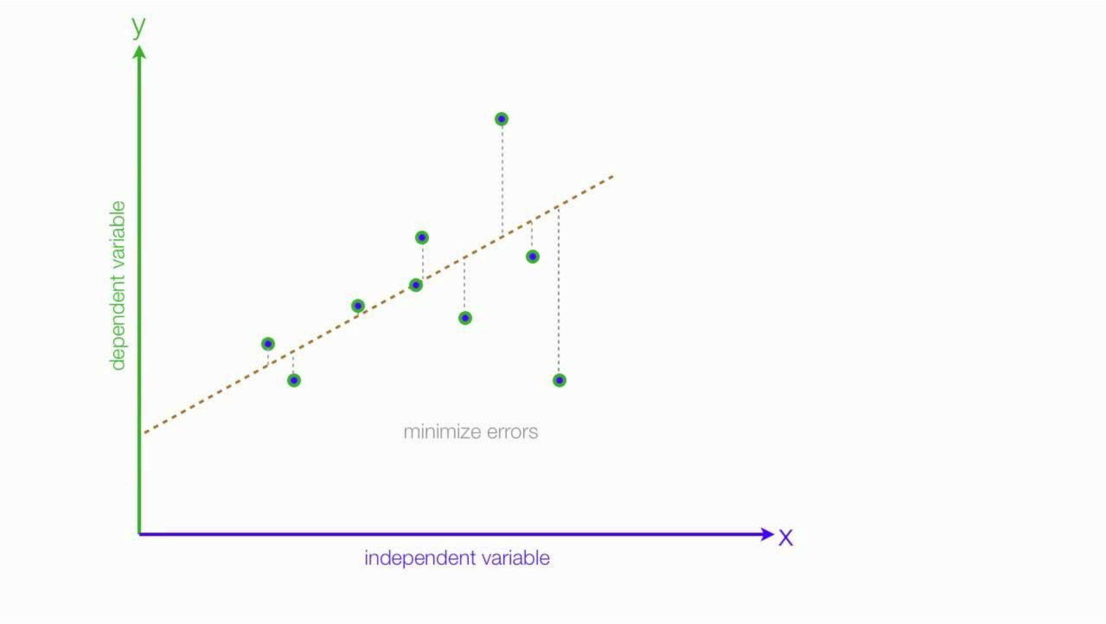
From Wikipedia, the free encyclopedia

In statistical modeling, **regression analysis** is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome' or 'response' variable, or a 'label' in machine learning parlance) and one or more independent variables (often called 'predictors', 'covariates', 'explanatory variables' or 'features'). The most common form of regression analysis is linear regression, in which one finds the line (or a more complex linear combination) that most closely fits the data according to a specific mathematical criterion. For example, the method of ordinary least squares computes the unique line (or hyperplane) that minimizes the sum of squared differences between the true data and that line (or hyperplane). For specific mathematical reasons (see linear regression), this allows the researcher to estimate the conditional expectation (or population average value) of the dependent variable when the independent variables take on a given set of values. Less common forms of regression use slightly different procedures to estimate alternative location parameters (e.g., quantile regression or Necessary Condition Analysis[1]) or estimate the conditional expectation across a broader collection of non-linear models (e.g., nonparametric regression).

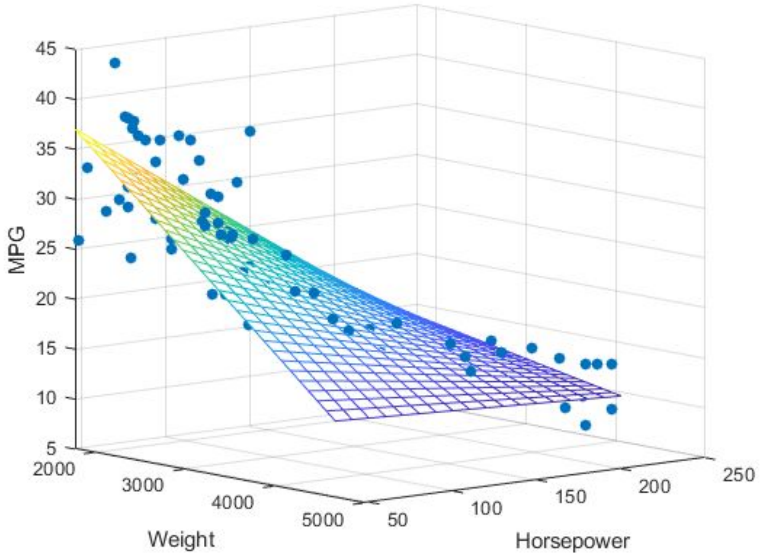


Regression line for 50 random points in a Gaussian distribution around the line $y=1.5x+2$ (not shown)

Linear Regression



Linear Regression



$$\hat{\beta} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{y}$$

$$\hat{\mathbf{y}} = \mathbf{X}\,\hat{\beta}$$

$\hat{\beta}$ = ordinary least squares estimator

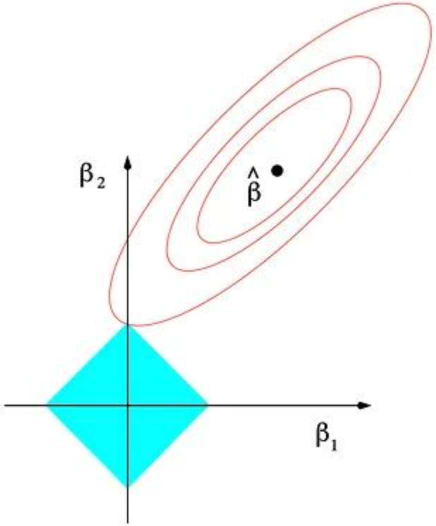
X = matrix regressor variable X

$\top$ = matrix transpose

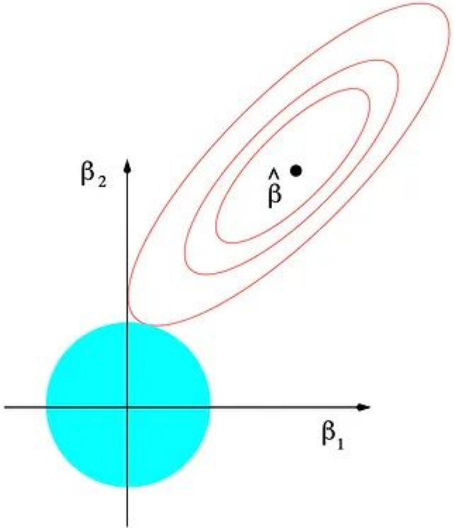
y = vector of the value of the response variable

$\hat{y}$ = predicted values

Linear Regression



Lasso



Ridge

Lasso

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq s \tag{6.8}$$

and

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 \leq s, \tag{6.9}$$

Ridge

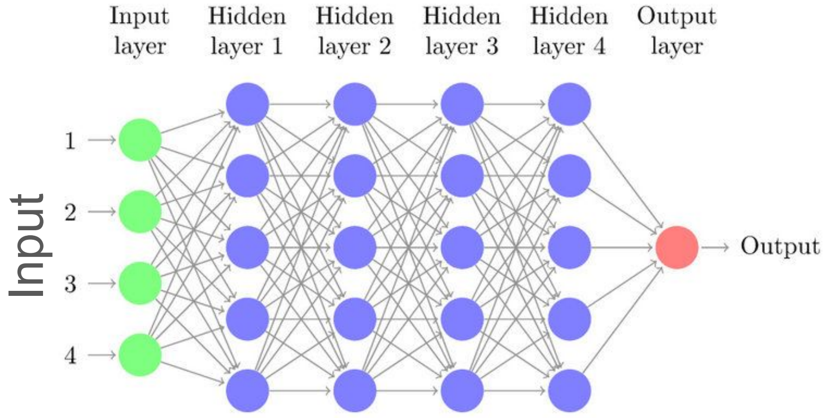
Lasso + Ridge = “ElasticNet”

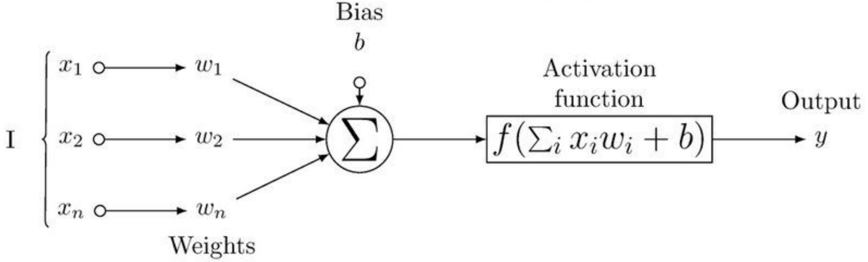
Multi-layer Perceptron (MLP)

Simple artificial neural network

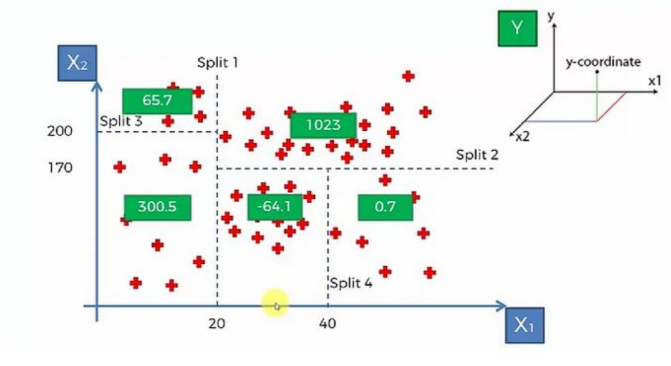
Functions like a stack of individual regression nodes

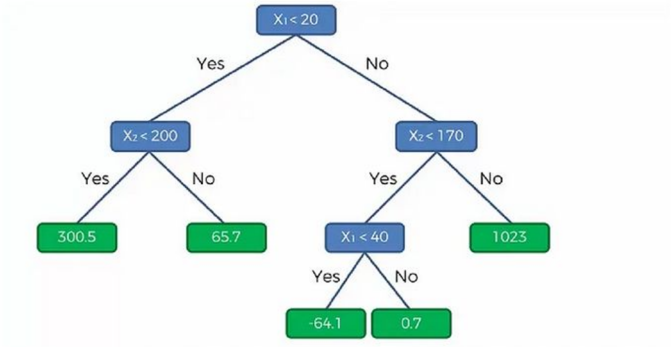
Needs to be trained with backpropagation and gradient descent

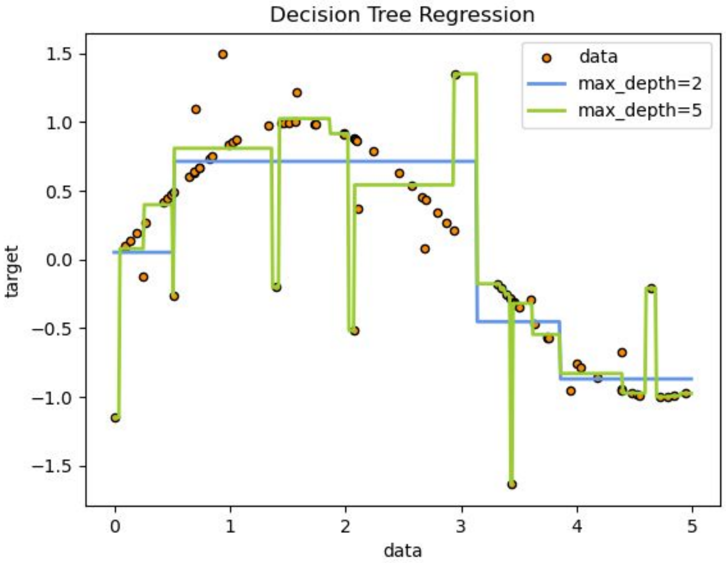




Decision Tree Regression







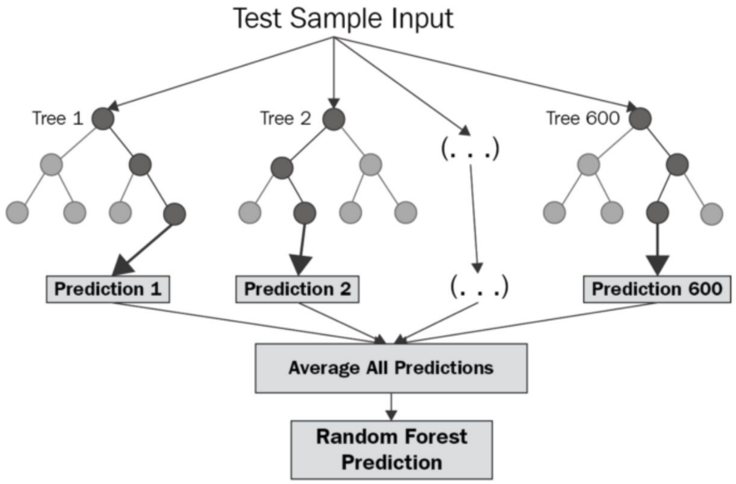
For each input region, the value is the average of all samples that fit the logical criteria.

Quality of a split is determined by how much it reduces Mean Squared Error

Ensemble methods

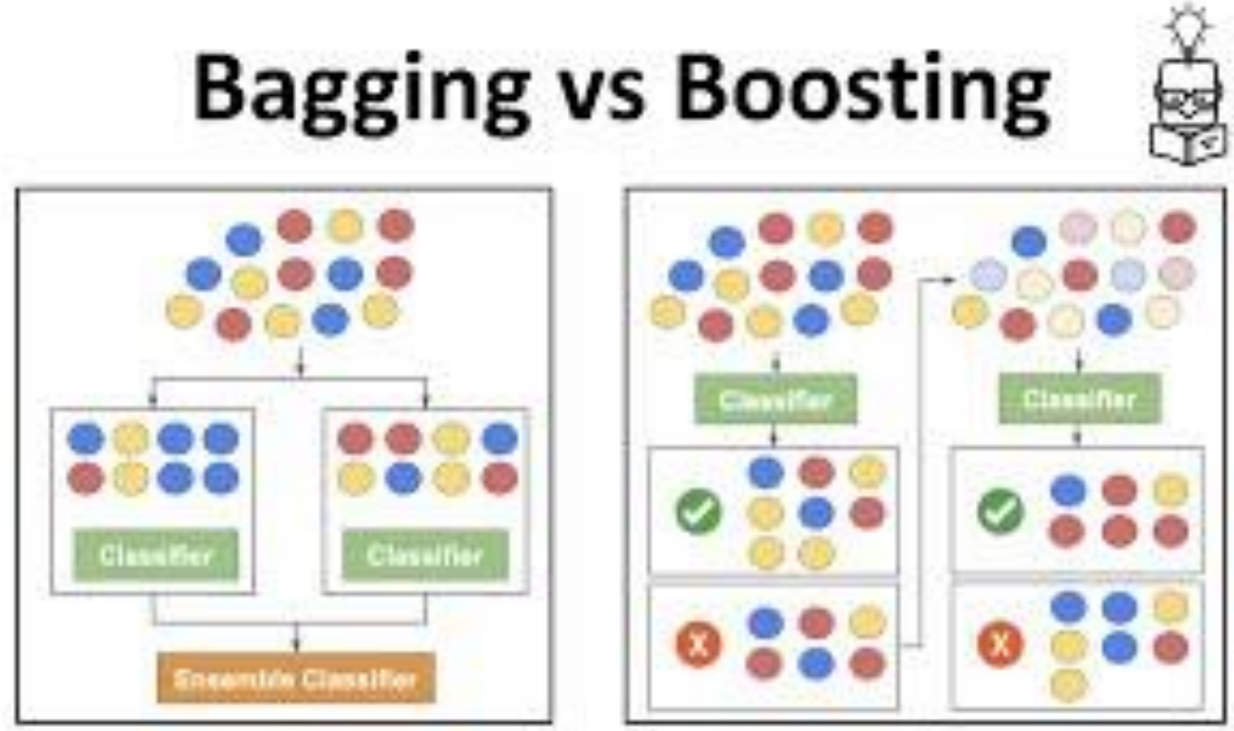
Output is a combination of multiple individual models.

e.g. Random Forest regression is an ensemble of many decision trees



Keboola

Bagging and Boosting

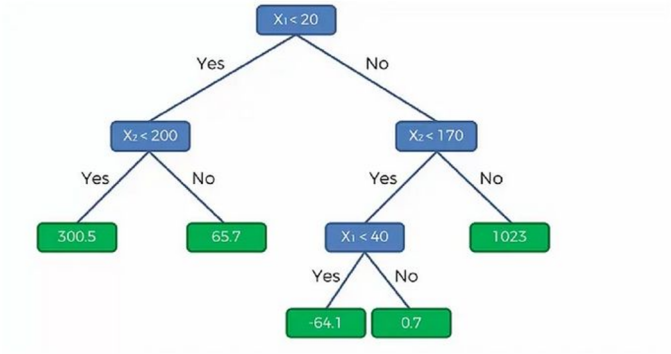


Feature Importance

Determine which inputs are most important for accurate predictions

Many methods exist for different regression models

When working with tree-based models, the splitting process provides a built-in metric of importance



Quality of a split is determined by how much it reduces Mean Squared Error

Paper assignments

Each week you will be assigned a paper to read (pdf posted in Brightspace).

You will need to read that paper and submit your answers to a set list of questions before next class, to prepare you for the discussion.

You will not be judged on the accuracy of your answers, but just that you made an honest attempt.

Paper Questions

How to Read a Scientific Article: The QDAFI Method of Structured Relevant Gist

Pascal Wallisch

Abstract

Gainfully reading a scientific paper can be challenging, particularly for beginners. This chapter introduces a principled method to extract relevant information from scientific articles to produce a concise summary—the "QDAFI" method. To be as efficient as possible, this method takes into account both the conventional structure of scientific papers as well as our understanding of cognition.

Q - Question
D - Do
A - rAtionale
F - Findings
I - Interpretation

But this is really more for science than engineering

Paper Questions

P - Problem

What problem are the authors trying to solve? Who needs this tool and for what?

M - Method

What did they use to solve the problem? E.g., what type of model and/or training technique

I - Innovation

How is what they did different from how people previously solved the problem? Did they come up with any clever new idea or explore a new problem?

R - Results

How well did their method(s) do, particularly in comparison to other methods?

O - Open issues

What problems remain? In what way are there still constraints or issues that prevent fully solving the original problem

Paper Assignment

PMIRO: Keep it short, sentence or two for each section.

You will also need to submit **at least one question** regarding something you were confused about in the paper.

Submit your PMIRO+Q via the Brightspace assignment before the start of class.

We will spend the first half of next lecture discussing the paper in groups.

Remember: you are not expected to fully understand the paper!