First Day of Class

EES 4891/5891

Probability & Statistics for Geosciences

Jonathan Gilligan

Class #1: Tuesday, January 07 2025

Introductions

Introduce Yourselves

1. Name
2. What are your Major or Program, and research interests?
3. What's your year or how long you've been in your graduate program?
4. What previous math or stats have you taken?
5. What do you hope to get from this class?
6. Something interesting about yourself
7. (Optional) Ask me a question about some aspect of probability or statistics

What Are Probability & Statistics?

What Is Statistics?

- Define “statistics”
- What do you use statistics for?
- What kinds of questions do you expect statistics to answer?
- Any questions for me?

What Is Proability?

- What is the probability the sun will rise tomorrow?
- What is the probability that if I roll a fair die, it will come up 5
- How would you know whether an answer to one of these questions is correct?

Applications

Climate Change

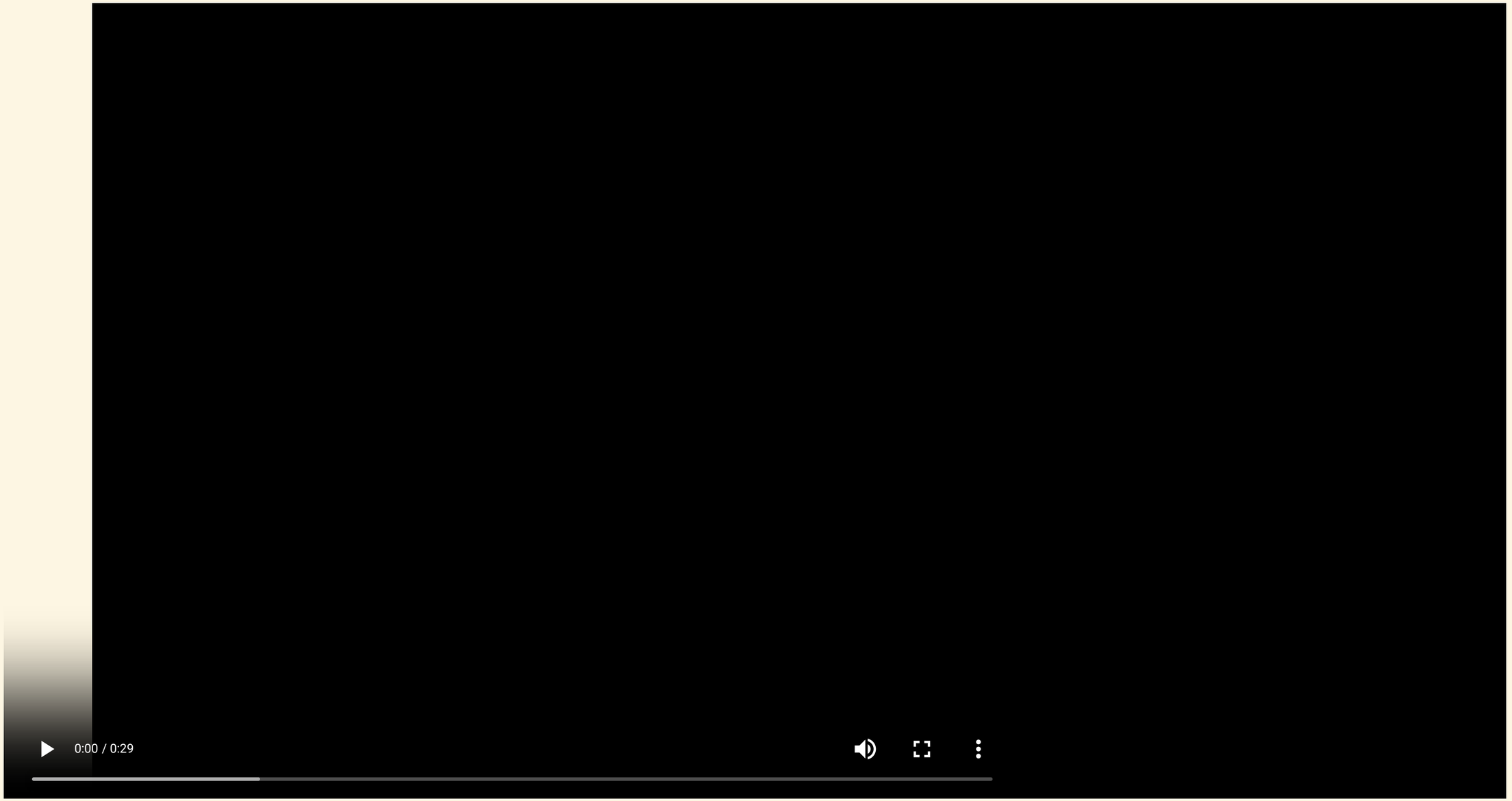
- What do scientists mean when they say the earth is getting hotter?
- How do we define the temperature of the earth?
- How do we *measure* the temperature of the earth?
- How do paleoclimatologists figure out what the earth’s temperature used to 1000 years ago?

Research Design

- You are studying a species of snail that lives on two islands, about 100 miles apart.
- On each island, there is considerable variation in the size of the snails' shells
- You want to determine whether snails on island A are larger than snails on island B, and by how much.
- How would you do this?
- How many snails would you need to collect and measure from each island?

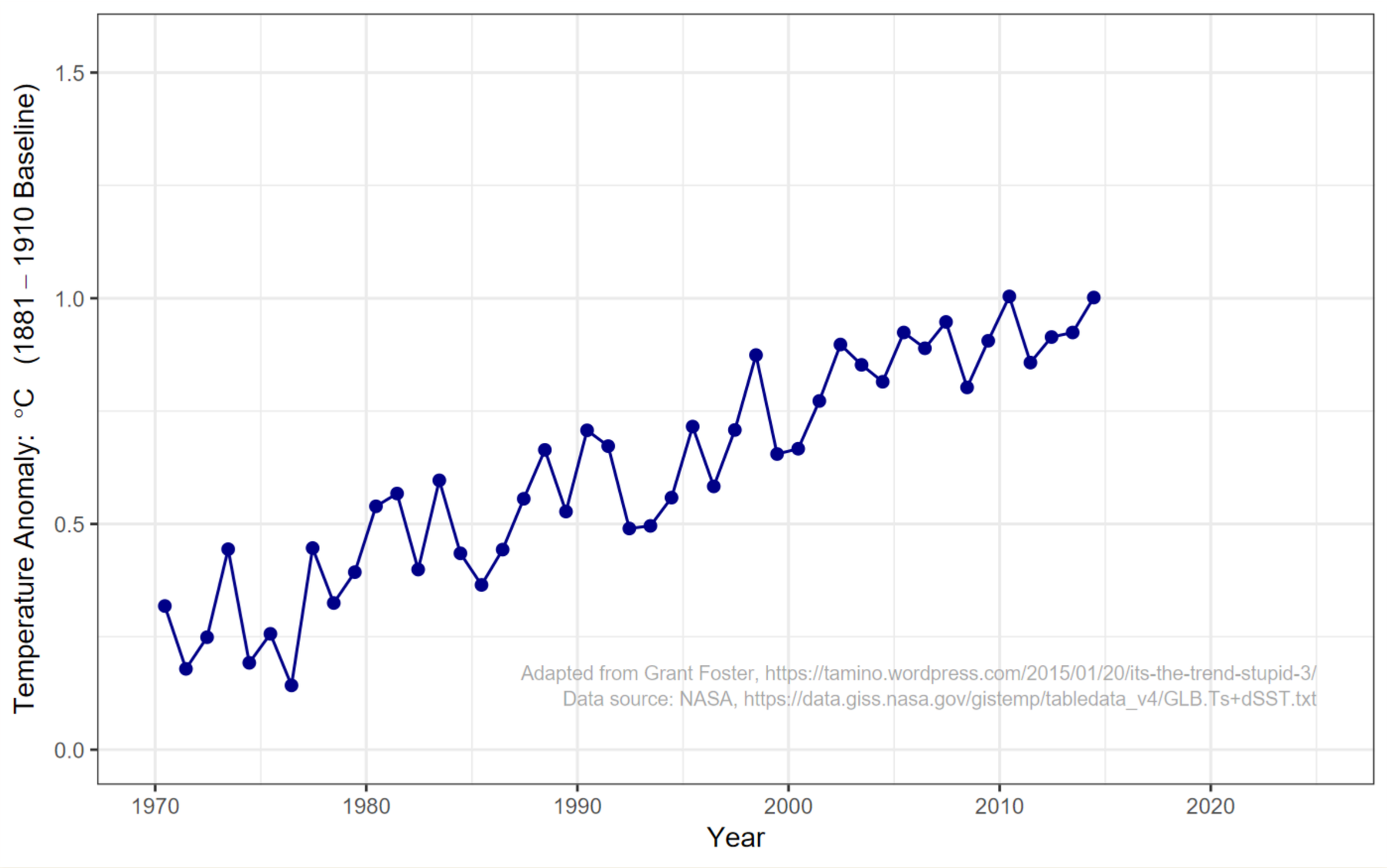
A Statistical Conundrum

Marsha Blackburn

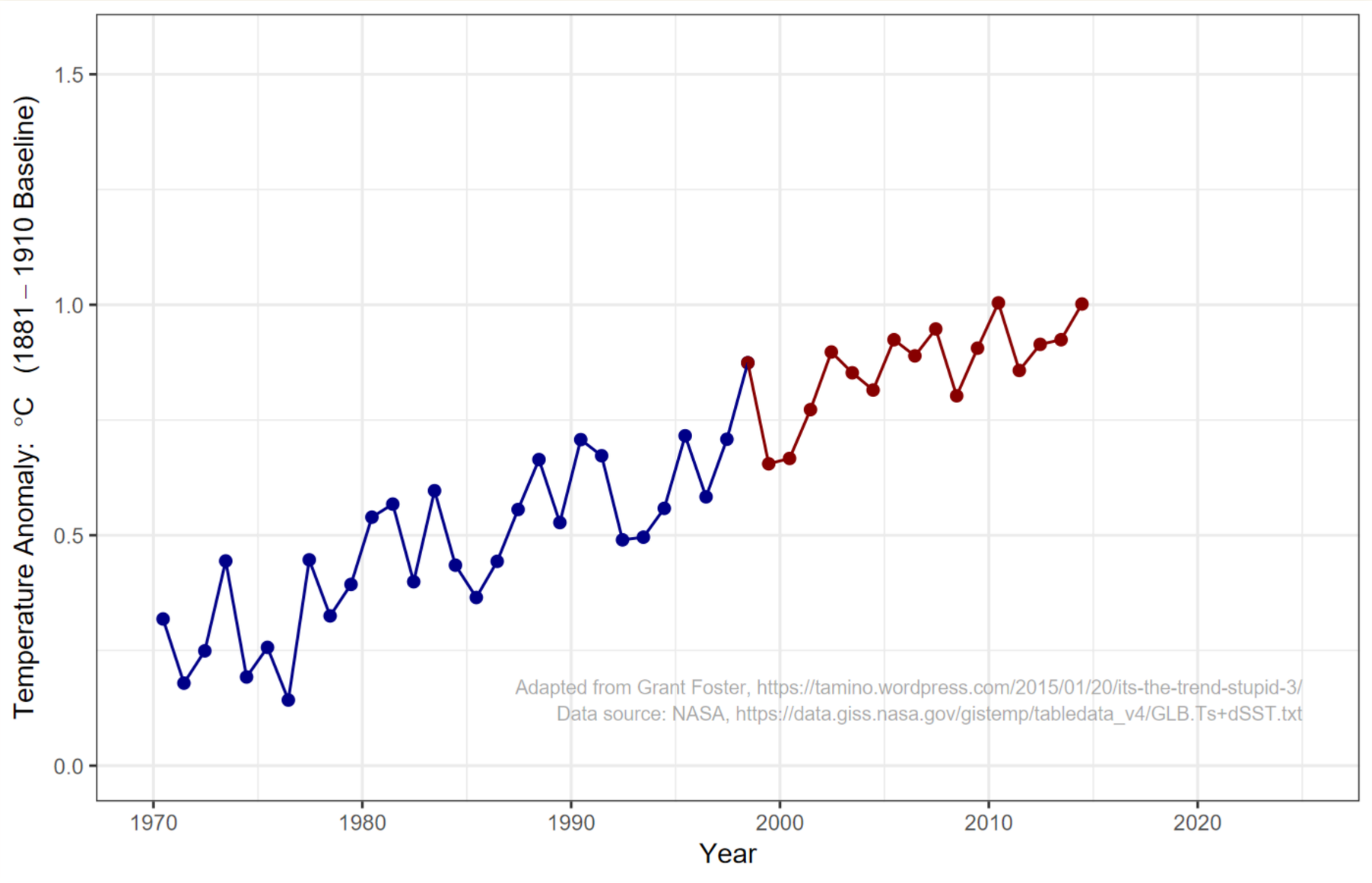


Did temperatures stop rising 18 years ago?

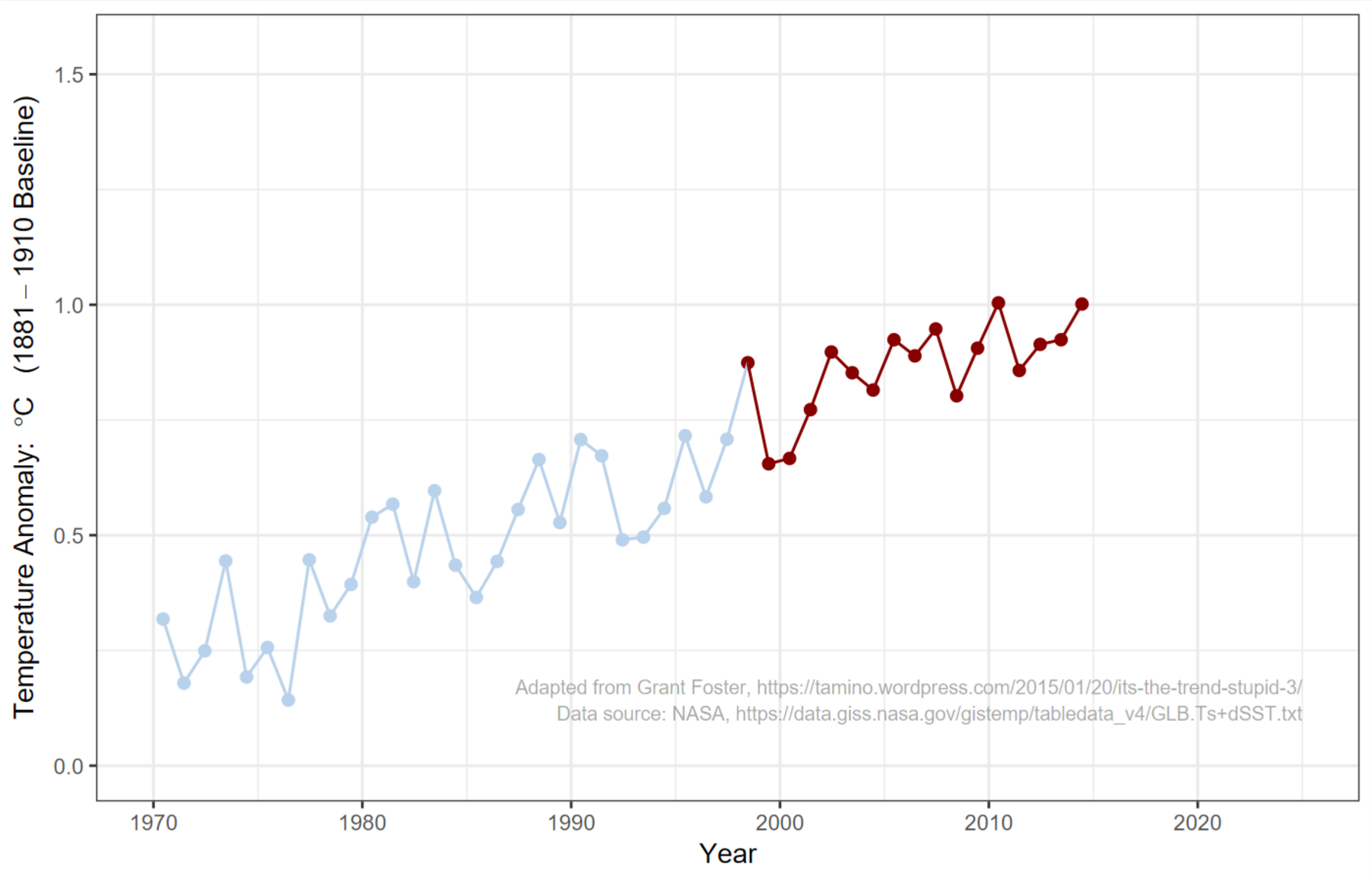
Look at 1970–2014



Did temperatures stop rising?



Did temperatures stop rising?



Scientists worried

AGU PUBLICATIONS

Earth's Future

RESEARCH ARTICLE
10.1002/2013EF000165

An apparent hiatus in global warming?

Kevin E. Trenberth[1] and John T. Fasullo[1]

[1]National Center for Atmospheric Research, Boulder, Colorado, USA

Key Points:
- There is a hiatus in the rise in global mean surface temperatures over the past decade
- Global warming continues but manifested in different ways
- Natural variability is playing the major role in the hiatus, through the PDO

Corresponding author:
K. Trenberth, trenbert@ucar.edu

Citation:
Trenberth, K. E., and J. T. Fasullo (2013), An apparent hiatus in global warming? *Earth's Future*, *1*, 19–32, doi:10.1002/2013EF000165.

Received 23 AUG 2013
Accepted 26 OCT 2013
Published online 5 DEC 2013

Abstract Global warming first became evident beyond the bounds of natural variability in the 1970s, but increases in global mean surface temperatures have stalled in the 2000s. Increases in atmospheric greenhouse gases, notably carbon dioxide, create an energy imbalance at the top-of-atmosphere (TOA) even as the planet warms to adjust to this imbalance, which is estimated to be 0.5–1 W m^{-2} over the 2000s. Annual global fluctuations in TOA energy of up to 0.2 W m^{-2} occur from natural variations in clouds, aerosols, and changes in the Sun. At times of major volcanic eruptions the effects can be much larger. Yet global mean surface temperatures fluctuate much more than these can account for. An energy imbalance is manifested not just as surface atmospheric or ground warming but also as melting sea and land ice, and heating of the oceans. More than 90% of the heat goes into the oceans and, with melting land ice, causes sea level to rise. For the past decade, more than 30% of the heat has apparently penetrated below 700 m depth that is traceable to changes in surface winds mainly over the Pacific in association with a switch to a negative phase of the Pacific Decadal Oscillation (PDO) in 1999. Surface warming was much more in evidence during the 1976–1998 positive phase of the PDO, suggesting that natural decadal variability modulates the rate of change of global surface temperatures while sea-level rise is more relentless. Global warming has not stopped; it is merely manifested in different ways.

Trenberth, K.E., & Fasullo, J.T. (2013). An apparent hiatus in global warming? *Earth's Future*, **1**, 19. doi: 10.1002/2013EF000165

CLIMATE SCIENCE

The cause of the pause

A global climate model that factors in the observed temperature of the surface ocean in the eastern equatorial Pacific offers an explanation for the recent hiatus in global warming. **SEE LETTER P.403**

ISAAC M. HELD

After a rise of 0.5 °C in the 25 years starting in the mid-1970s, the change in Earth's global mean surface temperature has been close to zero since the turn of the century (Fig. 1). This hiatus in global warming has occurred despite retreating Arctic sea ice and rising sea levels. On page 403 of this issue, Kosaka and Xie[1] make a strong case for the hypothesis that this hiatus is driven by the equatorial Pacific.

Although the rise in carbon dioxide and other greenhouse gases explains many aspects of the overall warming trend over the past century (including the heat uptake by the oceans and the spatial and seasonal patterns of the warming), it cannot explain the multi-decadal fluctuations superimposed on this trend (Fig. 1). Forcing agents such as anthropogenic and volcanic aerosols and variations in the Sun's energy output are often called on

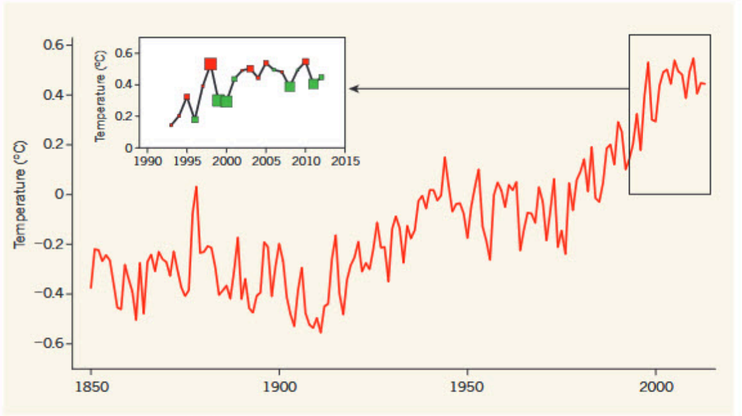
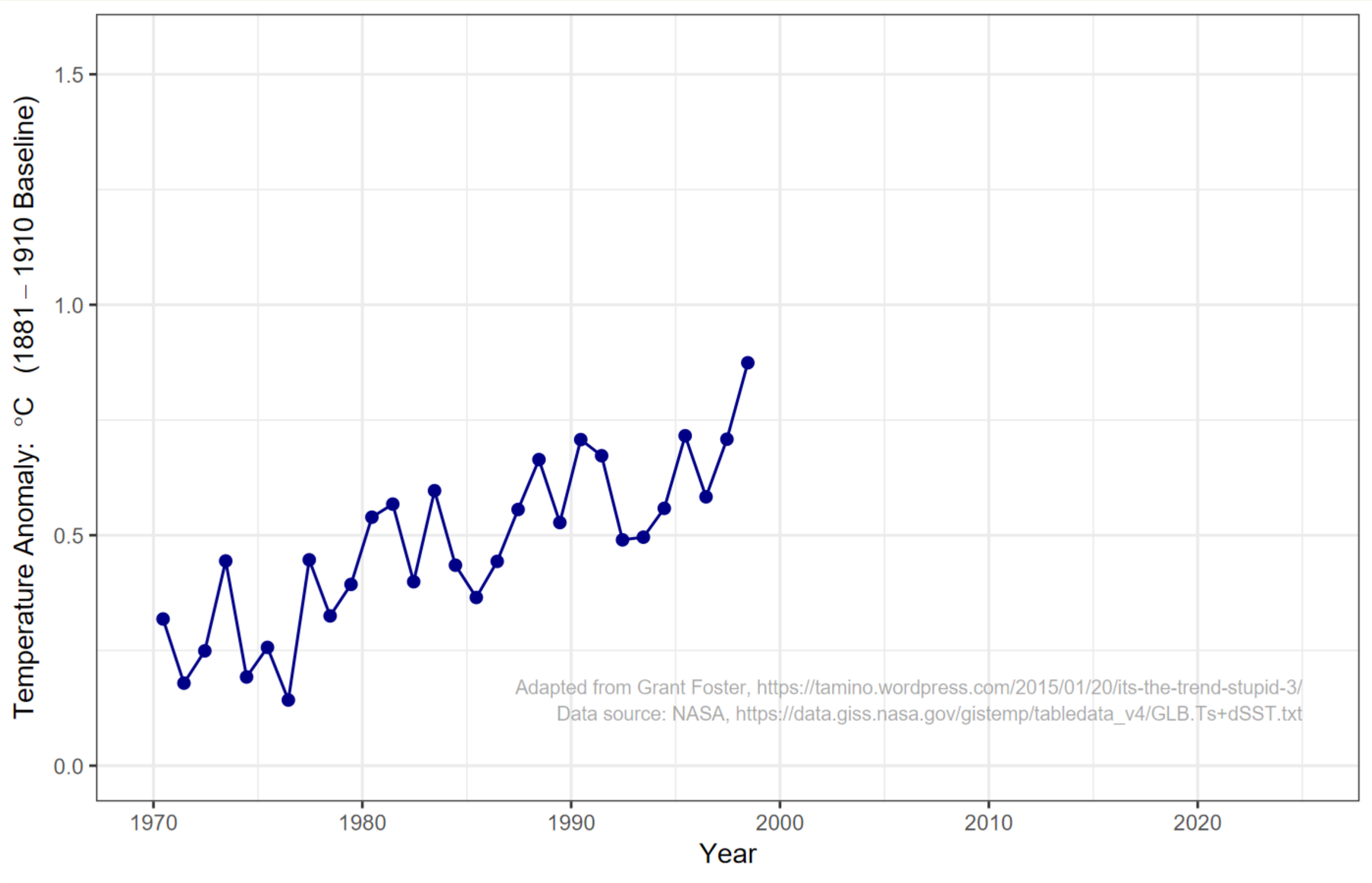


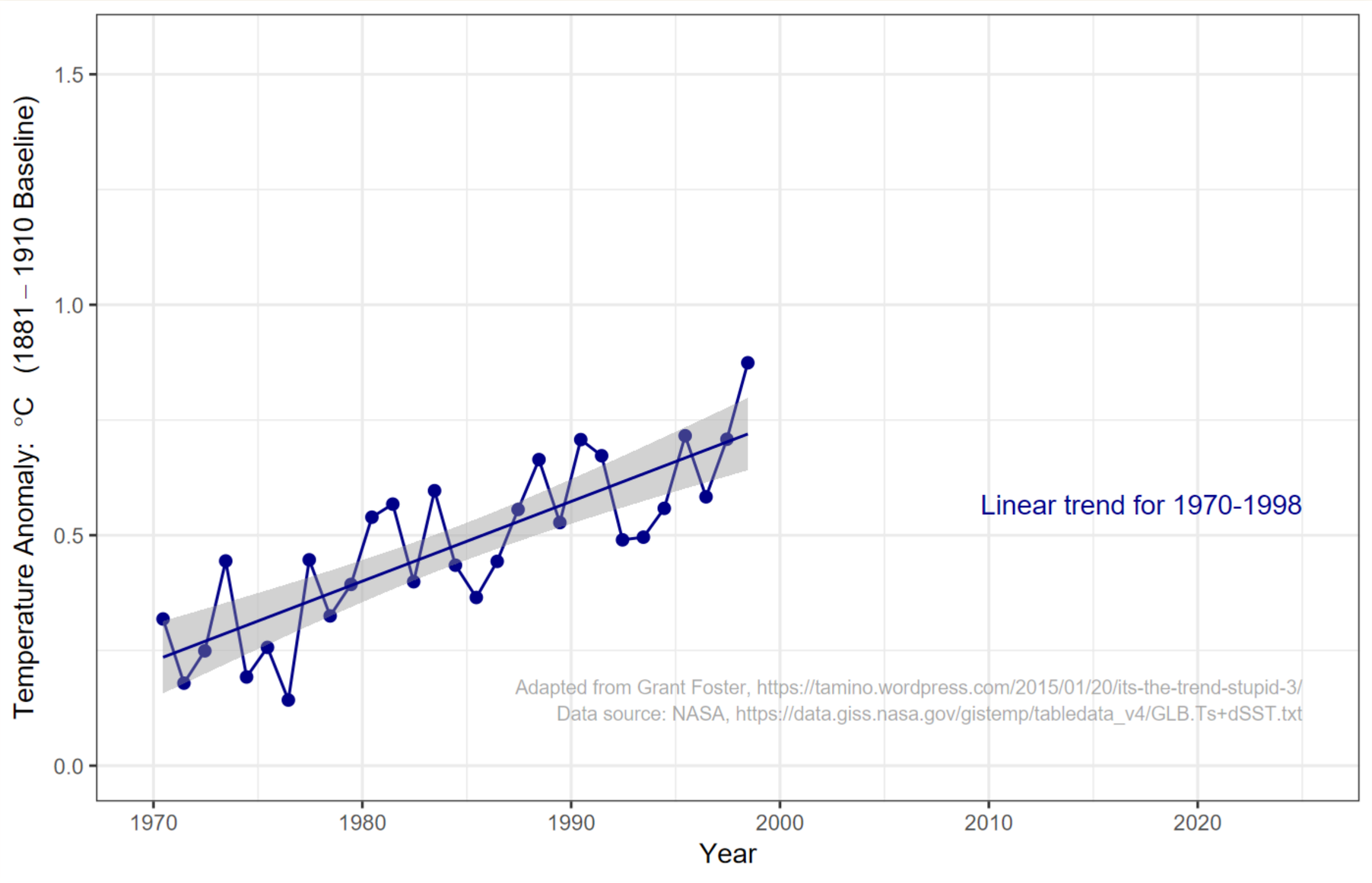
Figure 1 | Temperature evolution. The graph shows the global mean surface temperature relative to the 1961–90 mean, based on the HadCRUT4.2.0.0 data set[16]. The inset shows the 1993–2012 time span, with green denoting La Niña years and red, El Niño years; the size of the symbol indicates the strength of

Held, I.M. (2013). The cause of the pause. *Nature*, **501**, 318. doi: doi.org/10.1038/501318a

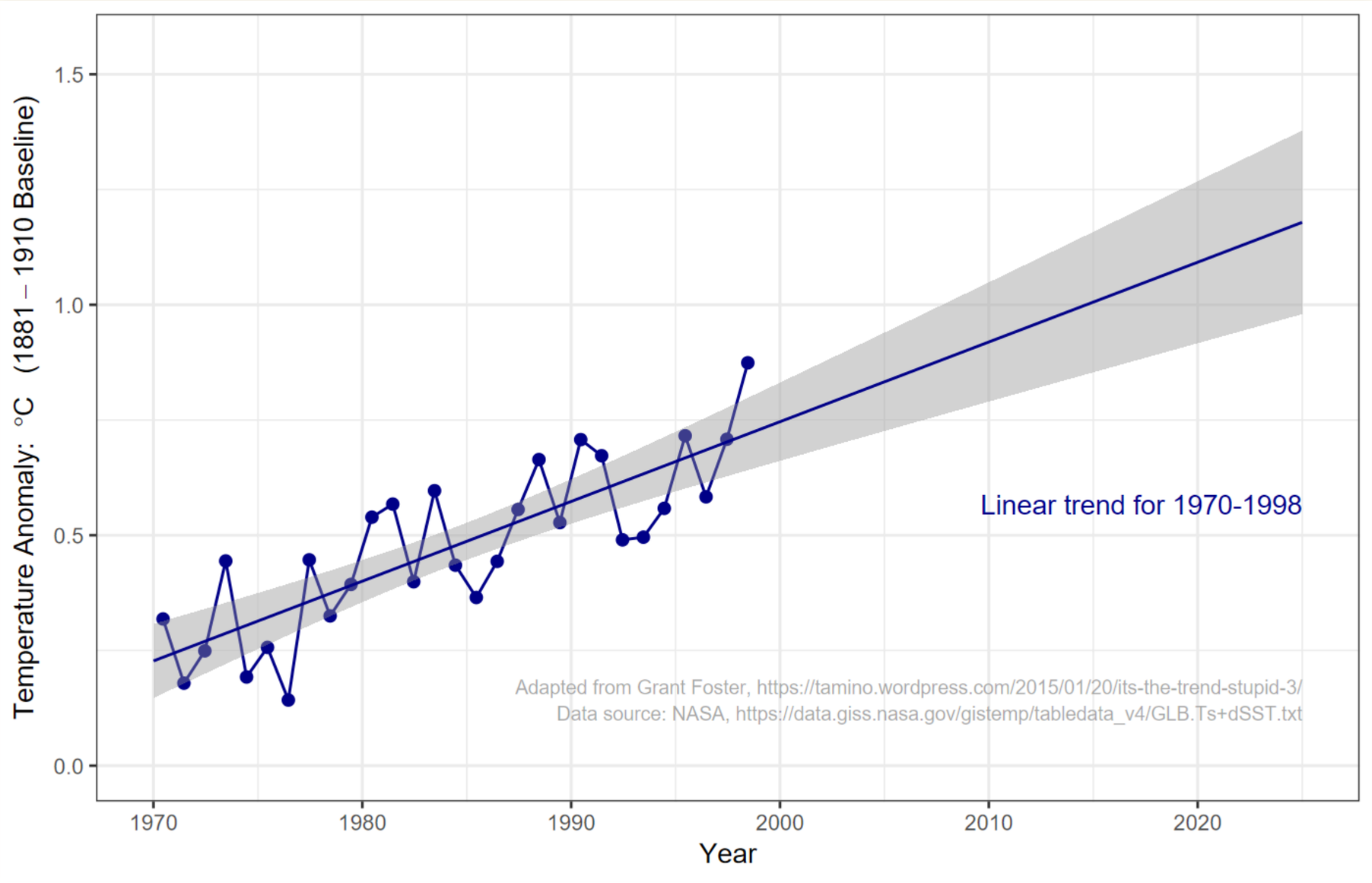
Did temperatures really stop rising?



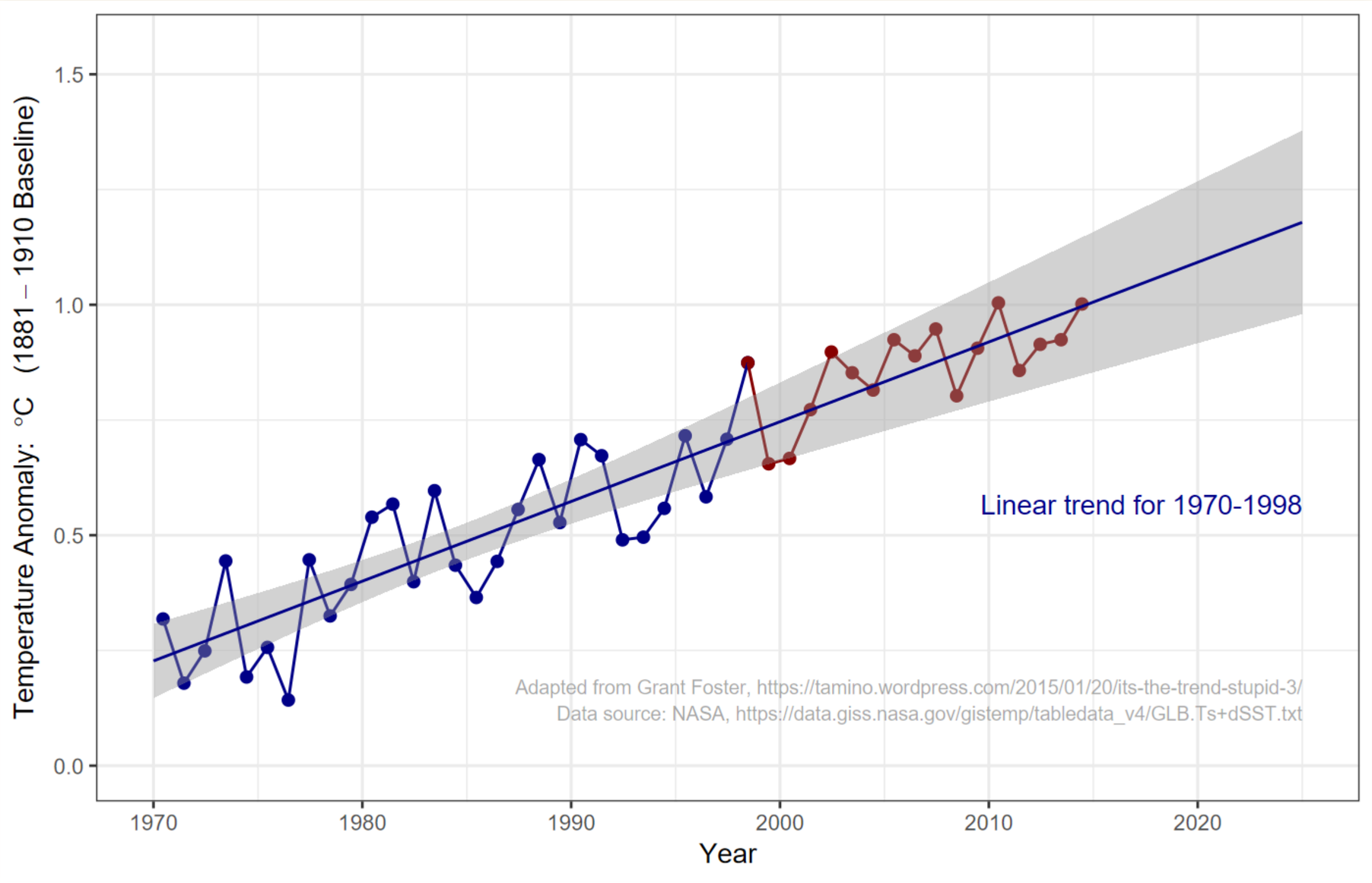
Did temperatures really stop rising?



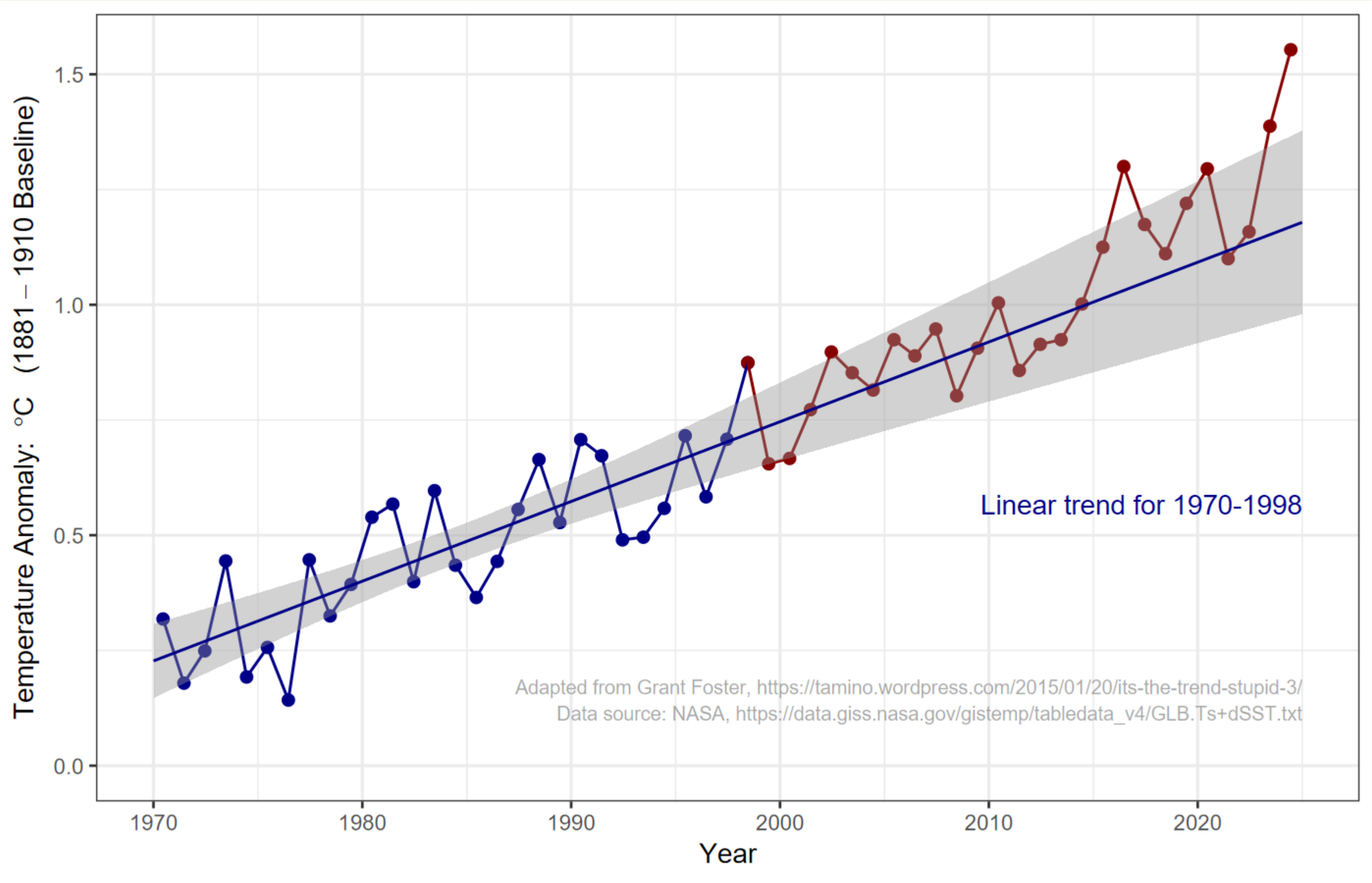
Did temperatures really stop rising?



Did temperatures really stop rising?



Did temperatures really stop rising?



Lessons

- Many leading scientists underestimated how many years of data they needed to tell whether an apparent slowdown in a treand was real, or just a temporary statistical fluke.

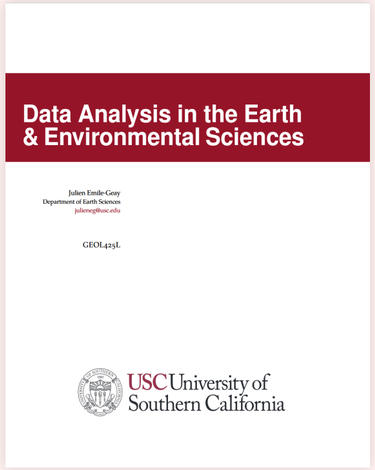
Organization of the Course

Organization of the Course

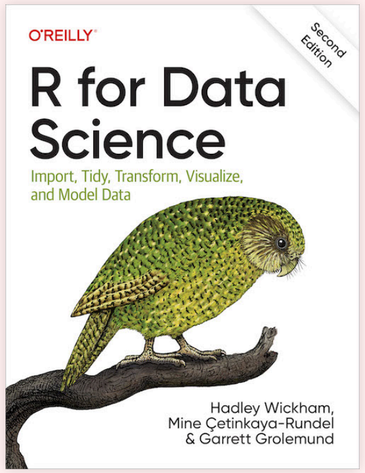
- The main course website is https://ees5891.jgilligan.org
  - Information:
    - Syllabus
    - Reading assignments
    - Homework assignments
    - Slides from class
    - Instructions for installing coourse software on your computer
  - Links to helpful resources
  - Slides:
    - Title slide has a QR code with link to online version
    - I also post PDF versions
    - Slides have two-dimensional navigation
      - Press “?” in your browser for help

Textbooks

Textbooks



Julien Emile-Geay, *Data Analysis for Earth & Environmental Science*,
5th edition (USC, 2023)
Required (Free to download PDF)



Hadley Wickham, Mina Çetinkaya-Rundel & Garrett Grolemund, *R for Data Science*,
2nd edition (O'Reilly, 2023)
Required (Free web version online)

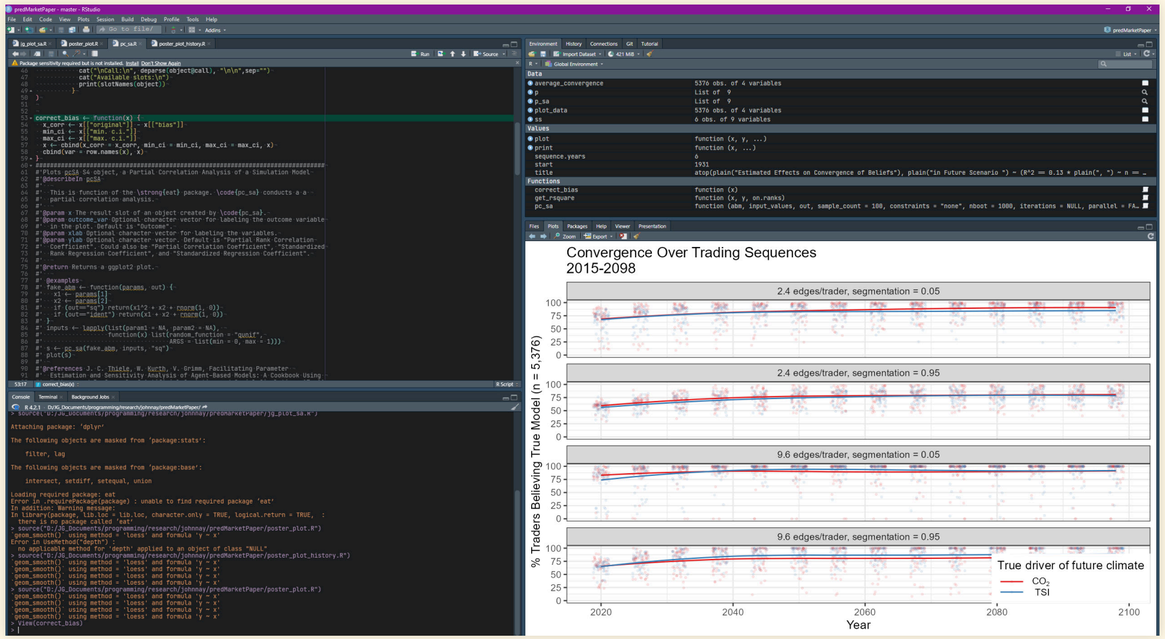
Computational Resources

R

- We will use the R for all the statistical analysis in this course
  - It is free and open source
  - It is widely used and powerful
  - It is written specifically for statistical analysis
  - There is an extensive library of free packages people have written to extend it.
  - You can find instructions for installing R and the other software for this course in the “Tools” page of the class web site and in the homework assignment for next Tuesday.

RStudio

- We will use the RStudio integrated development environment for working with R.
  - RStudio combines an editor for writing code and an environment for running R scripts and also using R interactively, displaying graphics, etc.
  - Like R, RStudio is free and open-source.



Reproducible Research

Spreadsheets for Data Analysis

- Statistical analysis with spreadsheets
  - What happens if someone raises questions several years after you publish your research?
    - Can you go back to your spreadsheet and understand everything you did in the analysis?
    - If you give the spreadsheet to someone else, can *they* understand your analysis by reading it?
  - If you make an error in a formula in one of the cells of a large spreadsheet, how hard would it be to find and correct the error?
  - You have just finished a long analysis and find that some of the data was entereed incorrectly.

    How much work would it take to correct the error and update your report?

Henrico Dolfing Case Studies, https://www.henricodolfing.com/2024/07/case-study-jp-morgan-chase-london-whale.html

Scientific Research with Spreadsheets

ELSEVIER

Appetite

Volume 71, 1 December 2013, Pages 126-136

RETRACTED: Relation of parenting styles, feeding styles and feeding practices to child overweight and obesity. Direct and moderated effects

Laura Hubbs-Tait [a], Katherine L. Dickin [b], Madeleine Sigman-Grant [c], Lisa Jahns [d], Amy R. Mobley [e f]

L. Hubbs-Tait et al. (2013) *Appetite*, **71**, 126. doi: 10.1016/j.appet.2013.08.004

I had found an error in the data for the manuscript and had tracked down the source of the error to a column switching mistake in copying data from one spreadsheet to another.... The error was difficult to detect because the coefficients for internal consistency of all measures were acceptable as were the descriptive statistics for all measures.

BBC

Home News Sport Business Innovation Culture Arts Travel Earth Video Live

Excel: Why using Microsoft's tool caused Covid-19 results to be lost

5 October 2020

Share Save

Leo Kelion
Technology desk editor

The badly thought-out use of Microsoft's Excel software was the reason nearly 16,000 coronavirus cases went unreported in England.

Reproducible Research

- All analysis is performed with computer scripts.
  - Plain text makes it easy to read, understand, and check for errors.
- Revision control tracks all edits
  - Old versions can be recovered
- Automate generating figures & tables, inserting numbers into text.
  - If data changes, or errors are corrected in analysis scripts, press a button to automatically regenerate the report or manuscript.

Tools:

- `git` for managing files, tracking edits and revisions.
  - `GitHub` service stores project in the cloud. Everything can be recovered if your computer dies.
- `Quarto` or `RMarkdown` to automatically generate manuscripts, reports, websites, blog posts, and presentation slides from your data analysis.