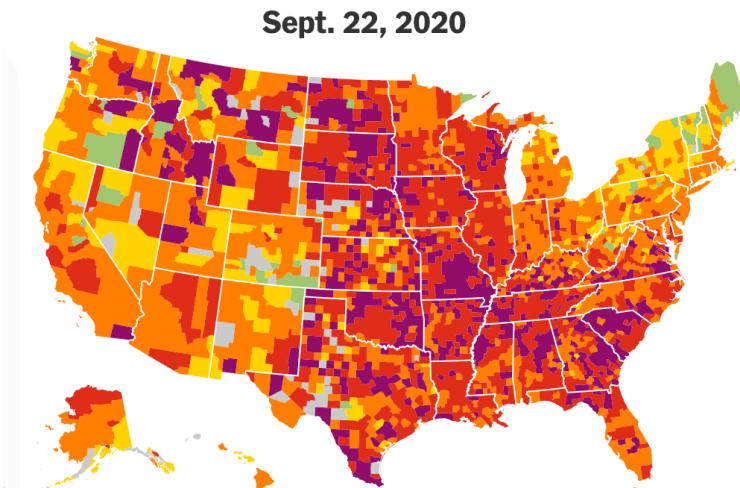


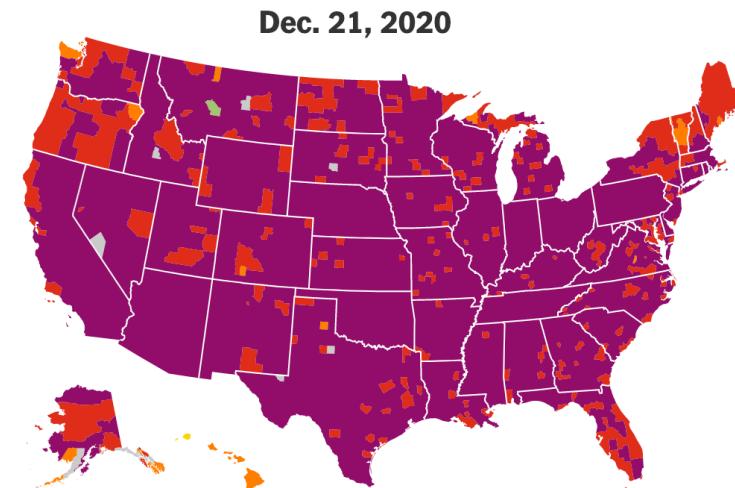
We are in a golden age of data



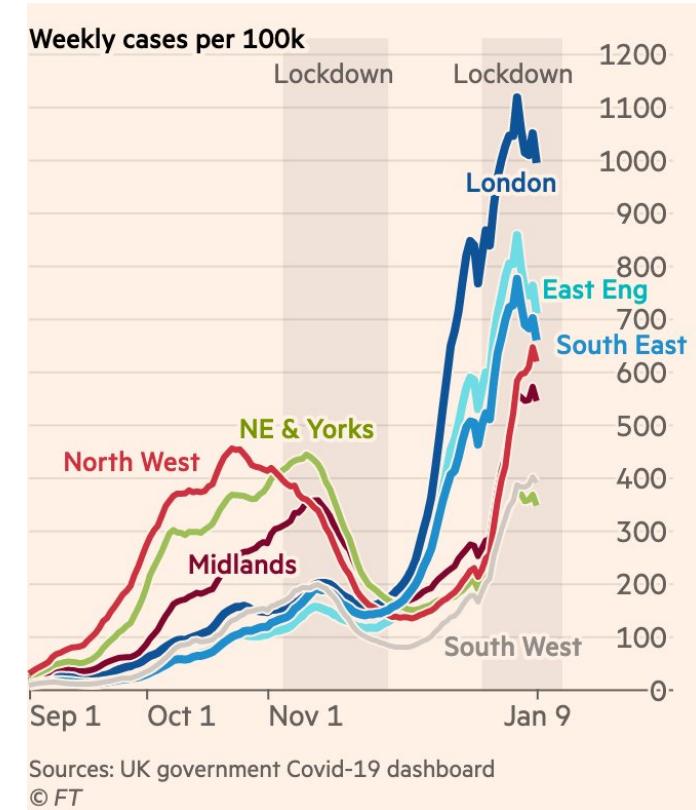
Epidemiological data



Sept. 22, 2020



Dec. 21, 2020



COVID-19 progression in the United States, from NYT

COVID-19 progression in the UK, from FT



Social media data

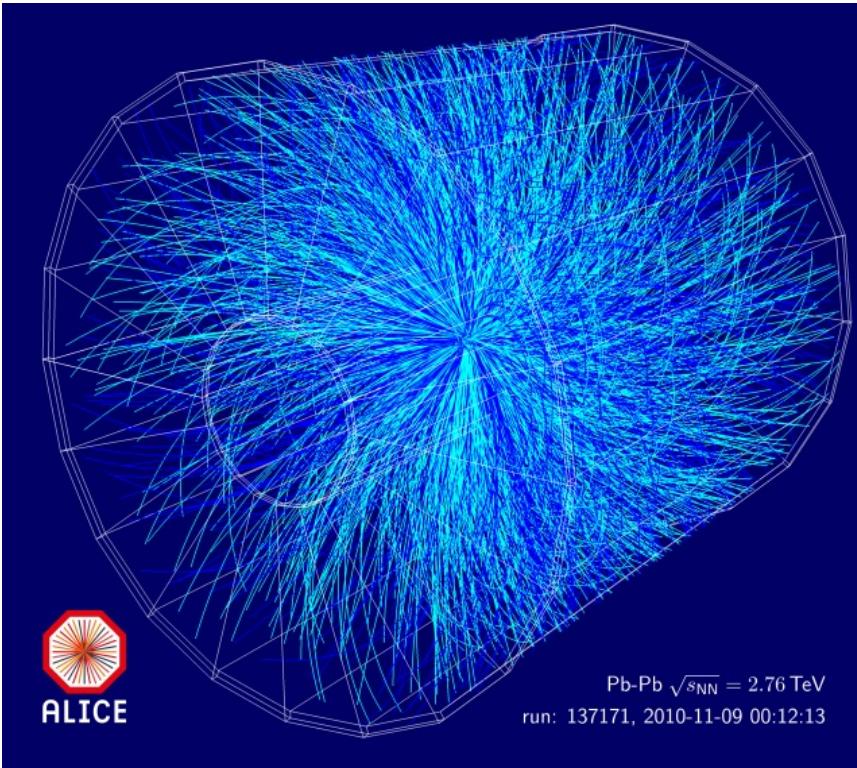
facebook

December 2010

Precision agriculture



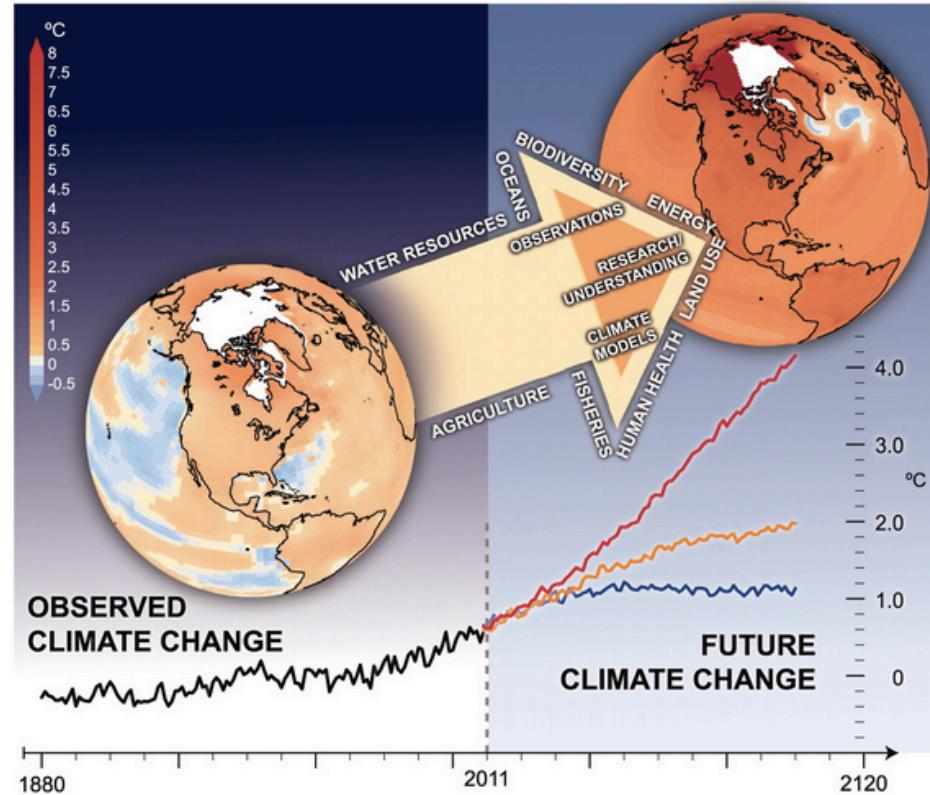
Big science data



Large Hadron Collider:

- 27 km circumference
- 15 petabytes of data per year
- (that's 10^{16} bytes)
- ((that's about 50 million episodes of The Wire per year))

Global change data – observed and predicted



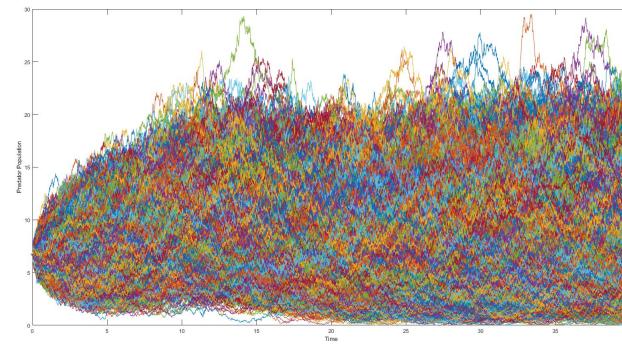
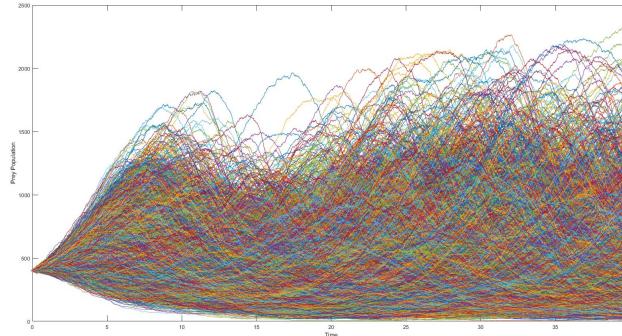
Remote sensors



Microsensors



Computer simulations

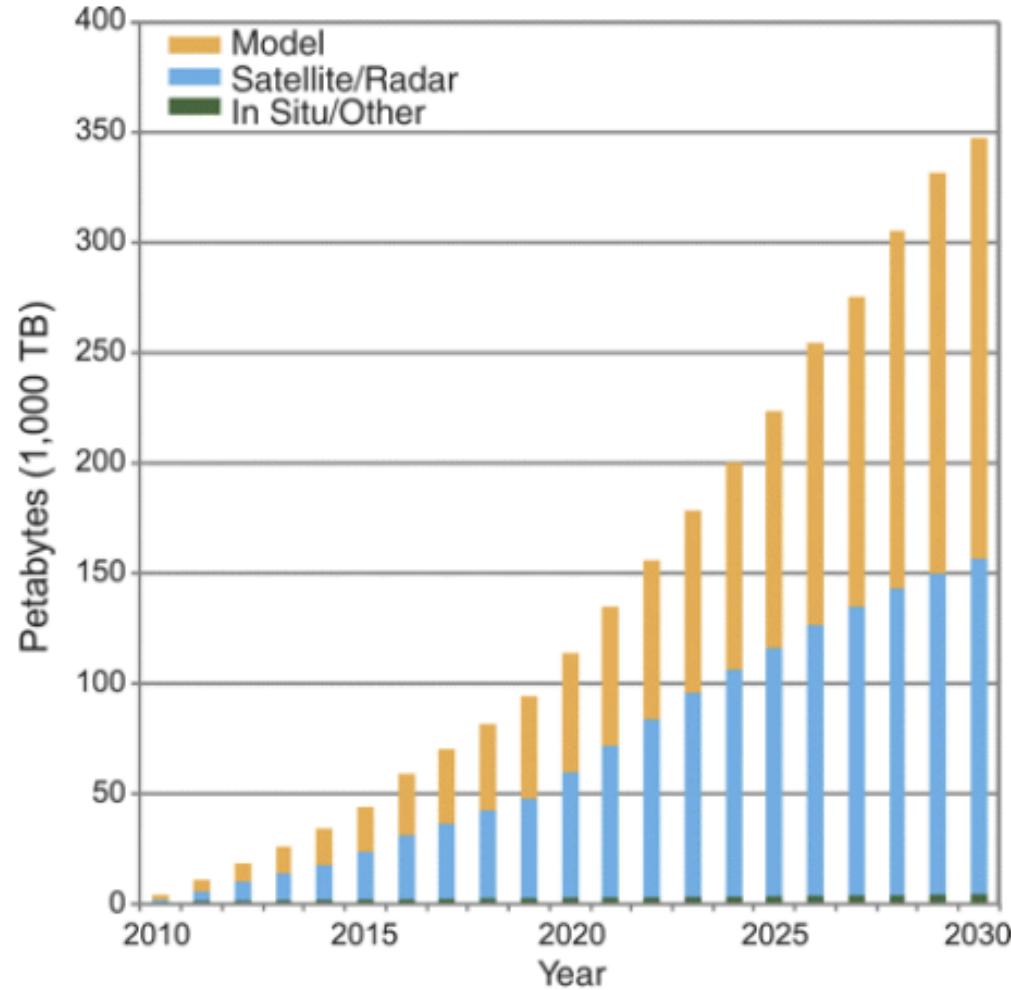


4

Field data

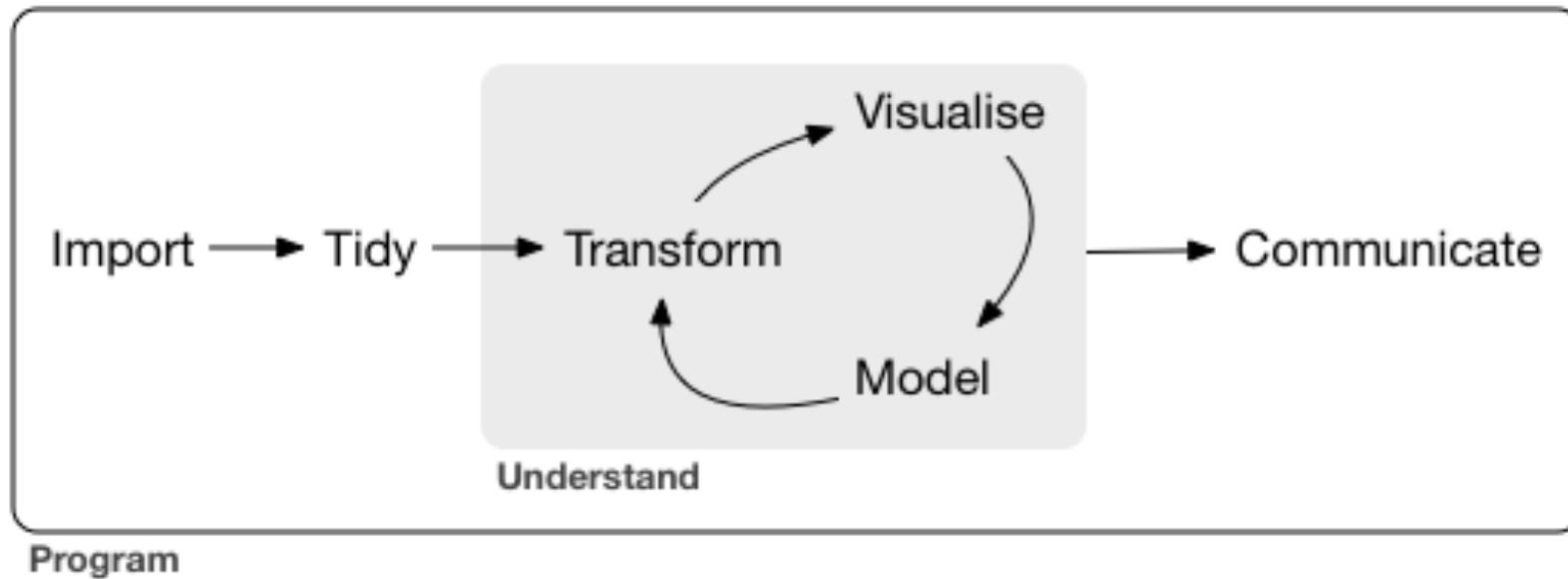


Growth of climate data by type



Overpeck+ (2011) doi:[10.1126/science.1197869](https://doi.org/10.1126/science.1197869)

Data life cycle



Wickham and Grolemund, R for Data Science

Step 1: Data cleaning



TECHNOLOGY

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

By STEVE LOHR AUG. 17, 2014



Monica Rogati, Jawbone's vice president for data science, with Brian Wilt, a senior data scientist.
Peter DaSilva for The New York Times

Step 2: Data analysis and modelling

- Formally the 80%
- Computational developments have made this much easier
- And led to an explosion in new methods and tools

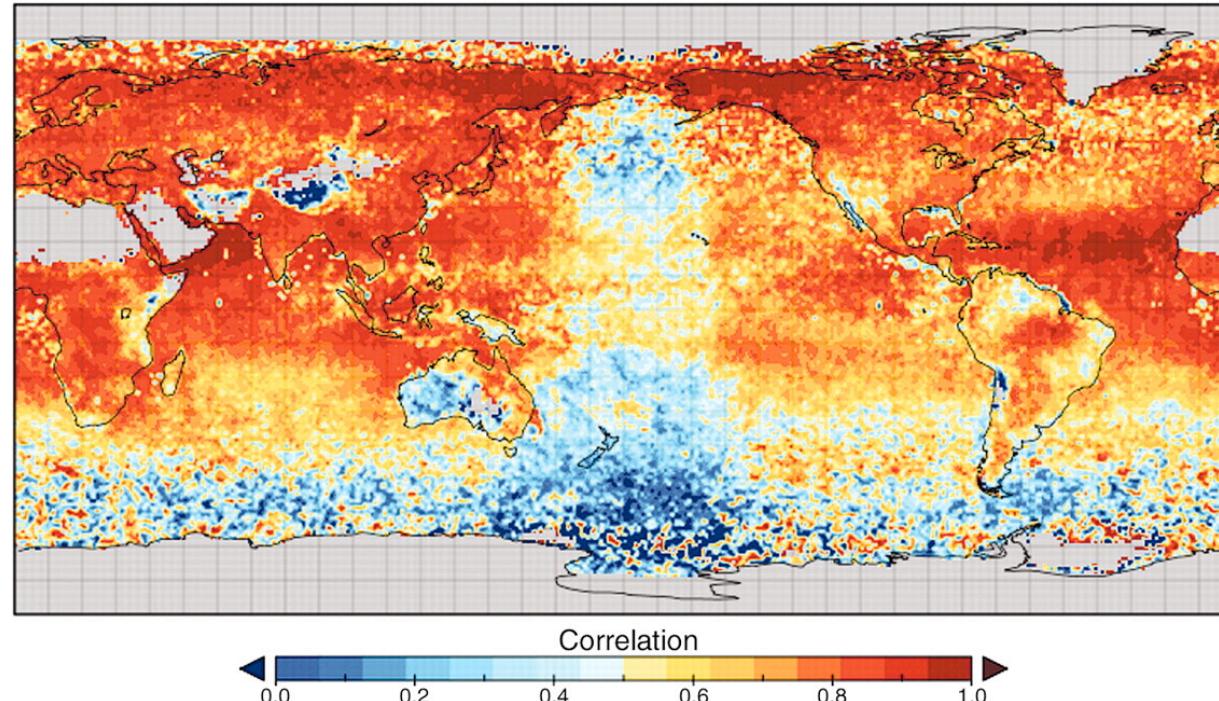
Step 3: Data visualization

- Today the visualization component has become a bottleneck
- Too often visualization becomes only an end-product rather than an exploration tool
- A problem because visualization is a quick way to spot errors

Data visualization for exploration

Correlation between two satellite images – measuring where they agree

Figure “...depicts Earth observation results from two satellites. It becomes immediately clear from this visualization that something odd is happening in the middle...”

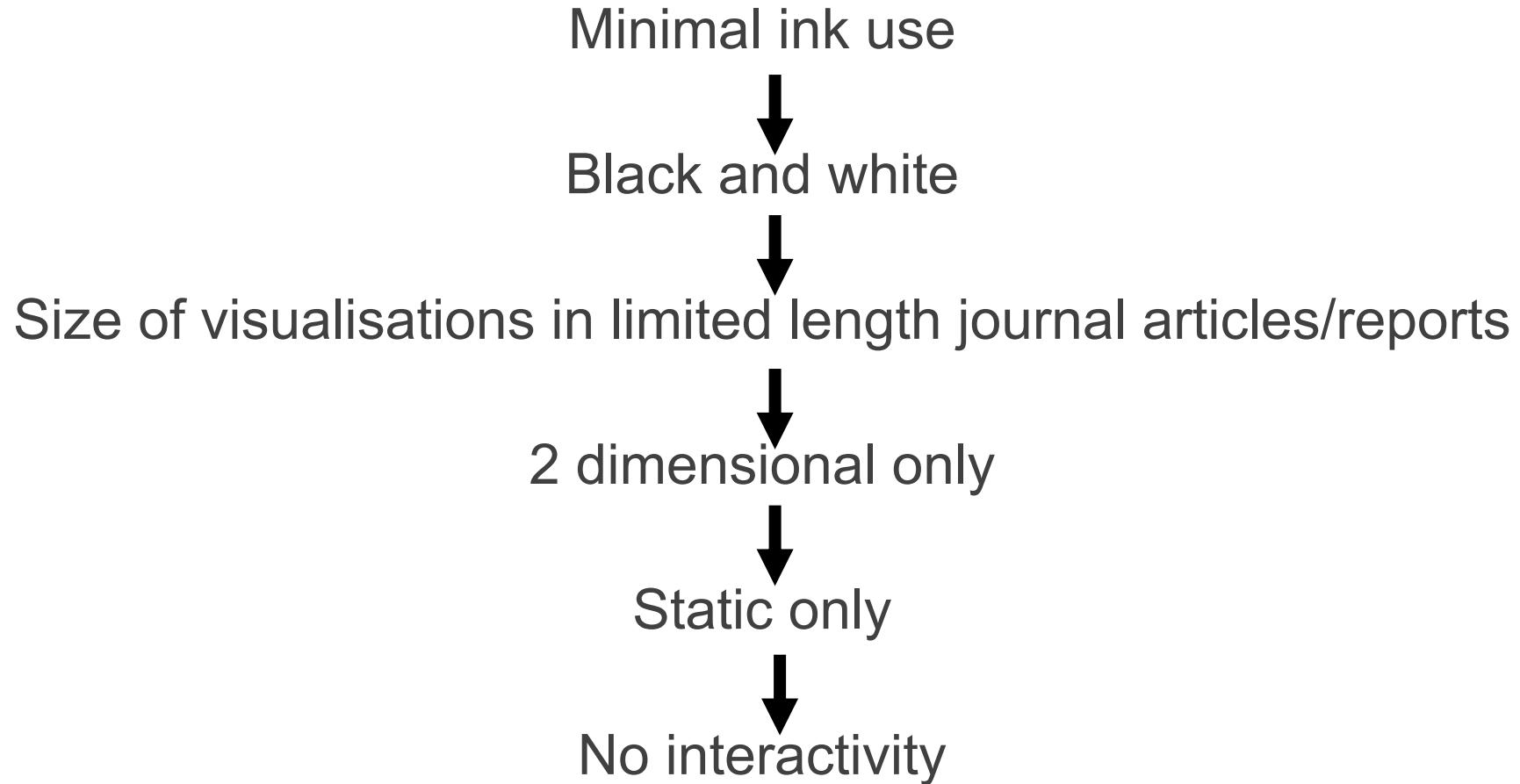


“When known, these time differences can be accounted for and a new and corrected visualization created.”

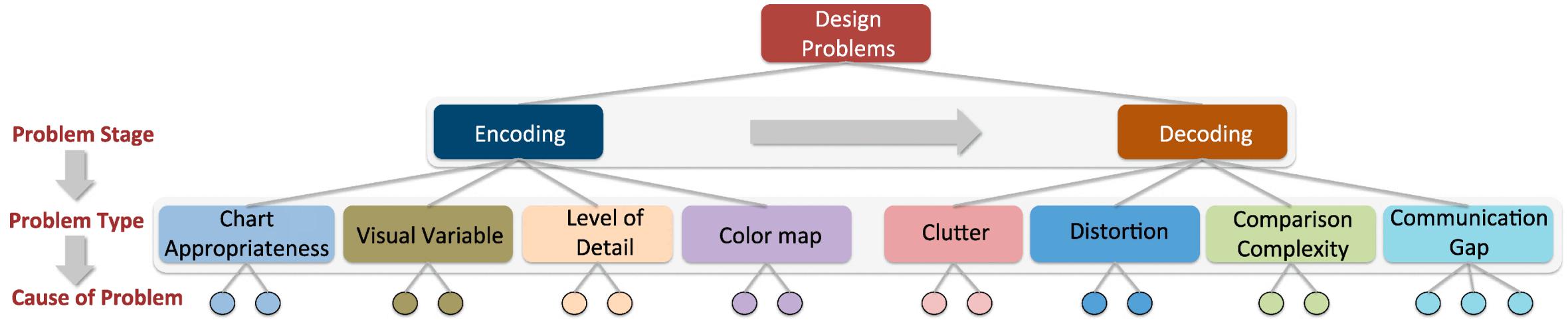
Fox & Hendler (2011): doi:[10.1126/science.1197654](https://doi.org/10.1126/science.1197654)

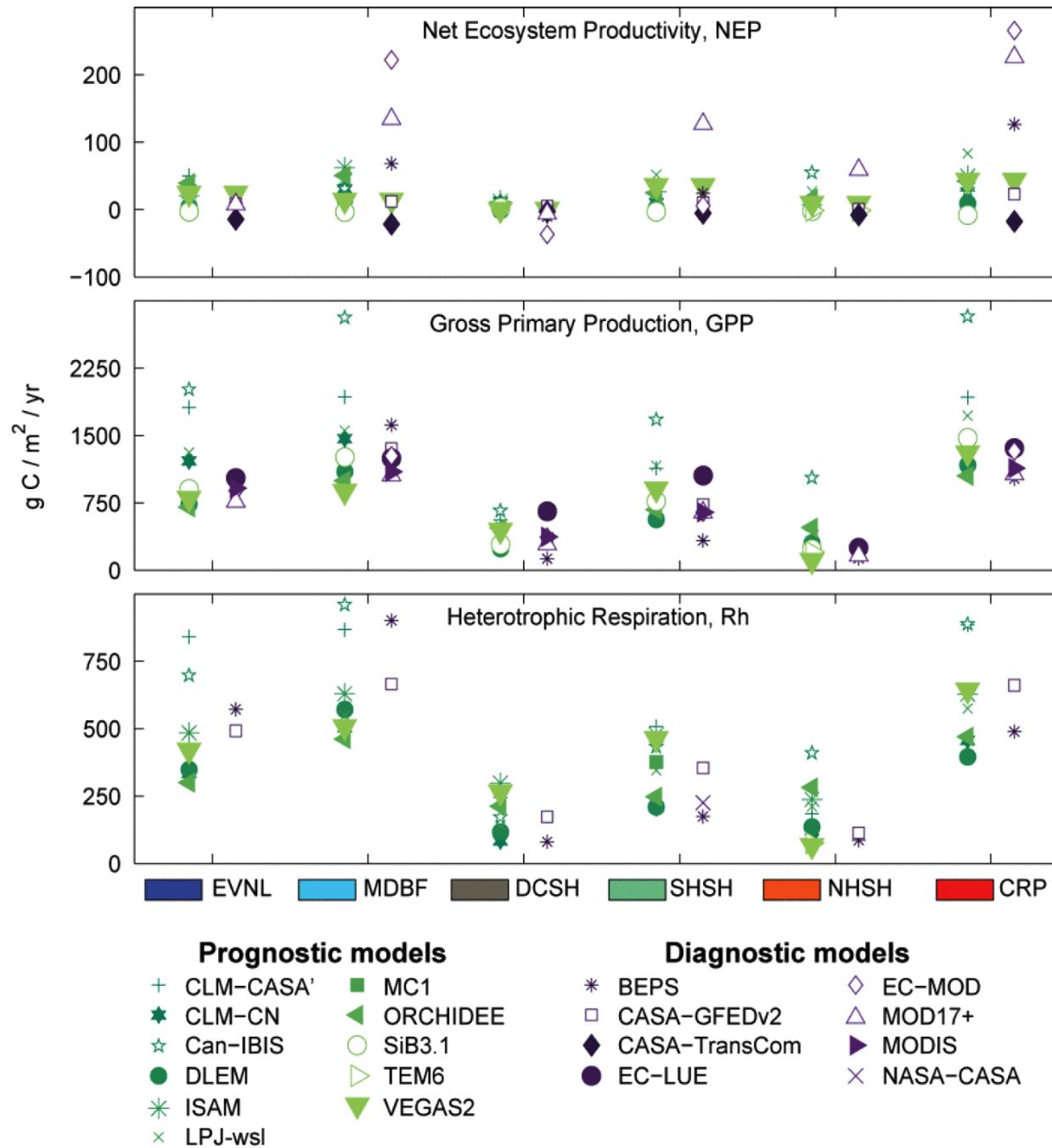
Data visualization as an end product

Continually overcoming restrictions:



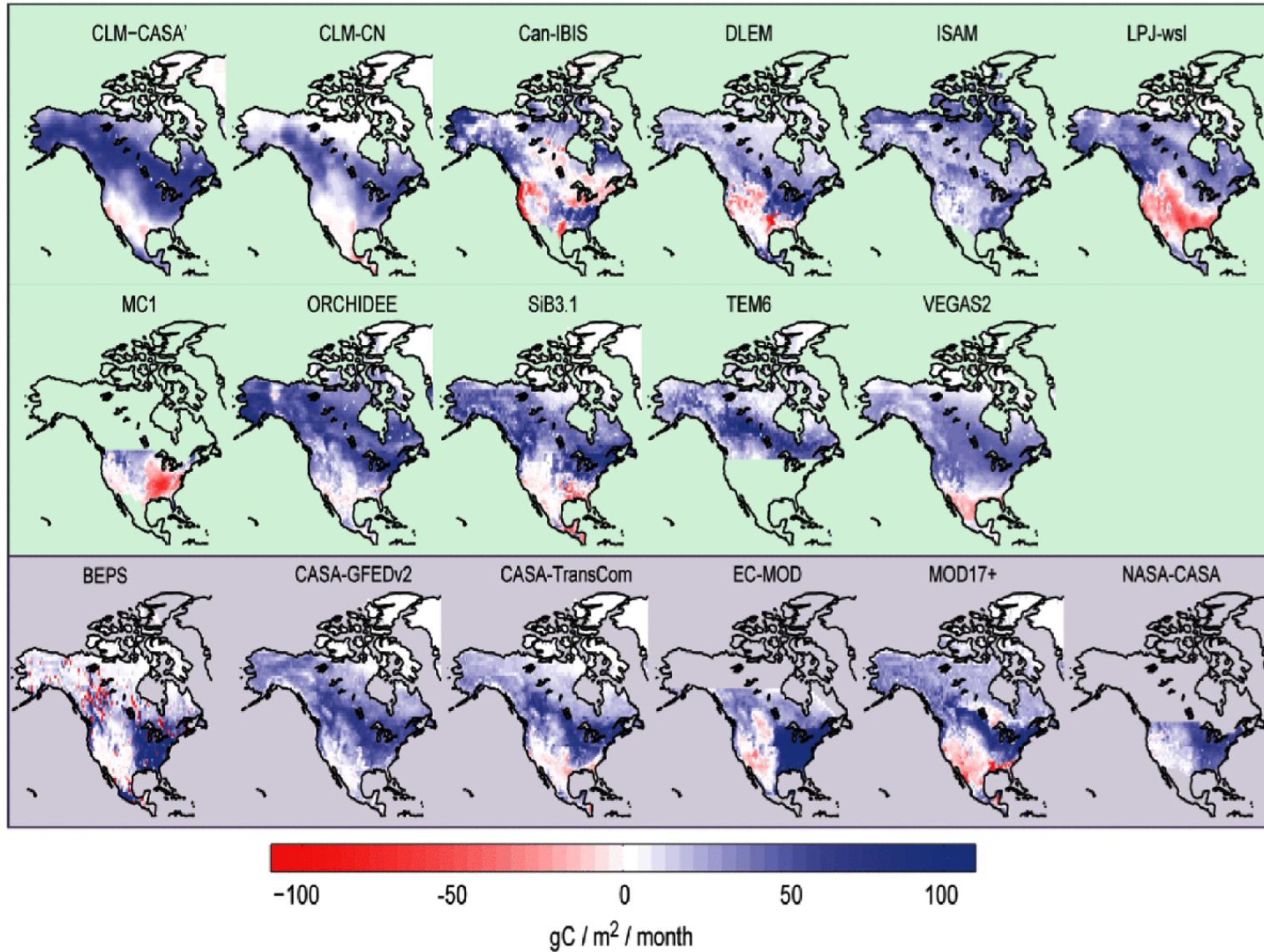
Visual communication problems





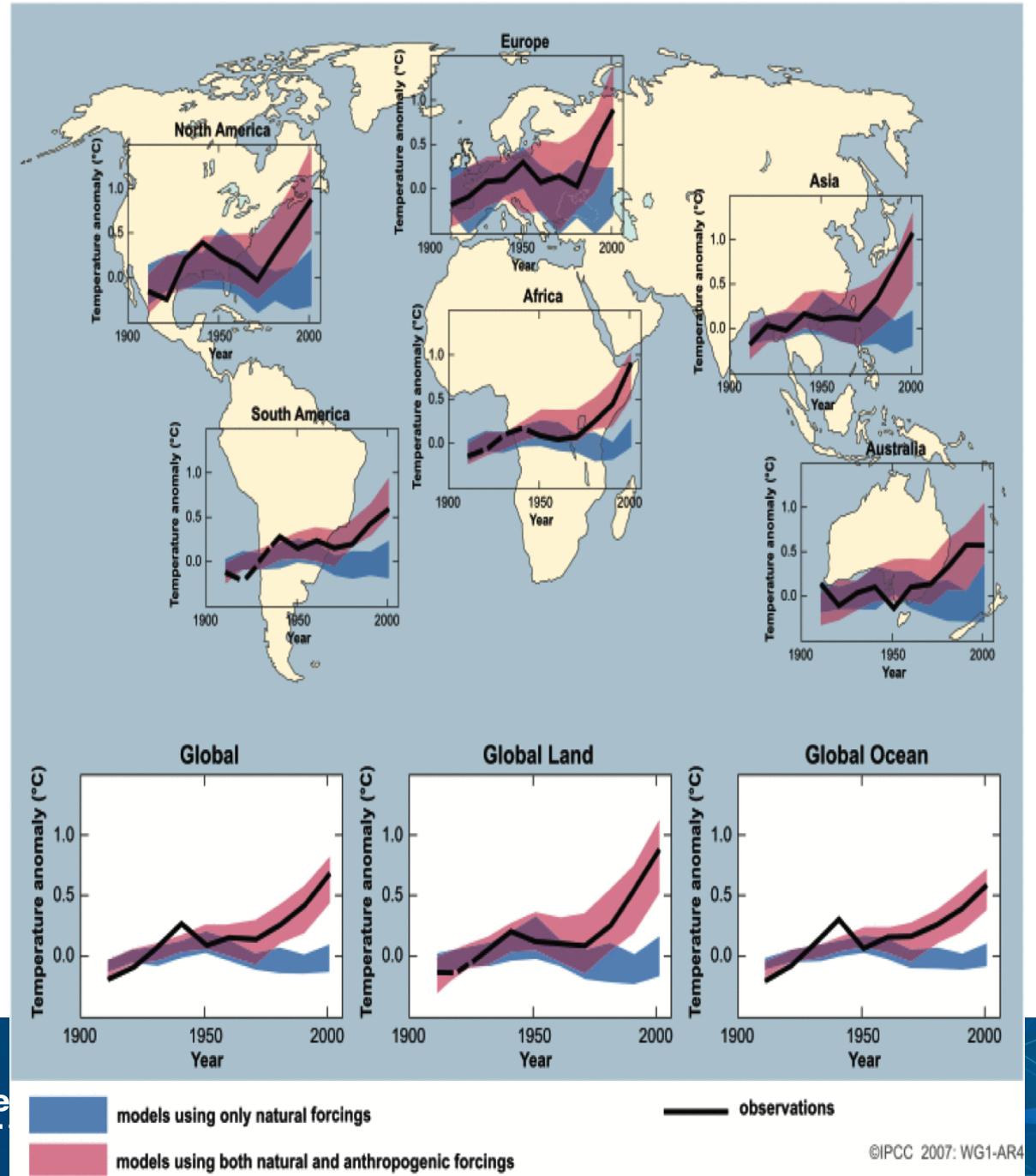
Scatter plot comparing output variables for multiple climate models

From Dasgupta et al.
2015



multiple maps showing
similarity of climate models over
different spatial regions

From Dasgupta et al.
2015



Example of optimal visualization design

- intent effectively captured and communicated
- small multiple line charts are used in conjunction with maps for showing region-wise temperature variation of two classes of climate models.

From Dasgupta et al.
2015