MEASURES AND METRICS

PLAN FOR TODAY

- 1. Fundamentals of Measurement
- 2. Metrics - Definition and Critical Analysis
- 3. Case Study - Metrics and Climate Change
- 4. Hands-On: Using R Functions to Create Metrics from Measures

LEARNING OBJECTIVES

- Become familiar with:
  - the definition of a measure
  - some types of measures
  - some issues with measures
  - multiple definitions of metric
  - recent history of the use of metrics
- Engage in critical analysis of metrics
- Exposure to several case studies of developing metrics (climate change context)
- Preliminary introduction to combining measures to create metrics in R.

FUNDAMENTALS OF MEASUREMENTS

MOTIVATING EXAMPLE

Is this 60 degree piece of metal twice as hot as this 30 degree piece of metal?

MEASUREMENT CONTEXT

- We need to keep the previous example in mind if we are creating models of the system using our measures.
- It's still fine to say that the ***measurement*** is twice as large, but it's the interpretation of this that is important!
- What does it mean to attach a number to a quality or a quantity? This is called 'measurement theory'
- (also – metrology: "the science of measurement and its application")

MEASUREMENT: SOME BASIC DEFINITIONS

- **Helmholtz** (1887: 17) the procedure by which one finds the denominate number that expresses the value of a magnitude, where a "denominate number" is a number together with a unit, e.g., 5 meters, and a magnitude is a quality of objects that is amenable to ordering from smaller to greater, e.g., length.
- **Bertrand Russell**: measurement is any method by which a unique and reciprocal correspondence is established between all or some of the magnitudes of a kind and all or some of the numbers, integral, rational or real. (1903: 176)
- **Norman Campbell**: "the process of assigning numbers to represent qualities", where a quality is a property that admits of non-arbitrary ordering (1920: 267).
- (from Measurement in Science: https://plato.stanford.edu/entries/measurement-science/)

SOME TYPES AND WAYS OF THINKING ABOUT MEASUREMENT

- Levels of measurement:
  - nominal,
  - ordinal
  - interval
  - ratio
- intensive vs extensive quantities
- fundamental vs derived quantities (measurement)

WHAT IS IT POSSIBLE TO MEASURE?

"Those who believe that what you cannot quantify does not exist also believe that what you can quantify, does."

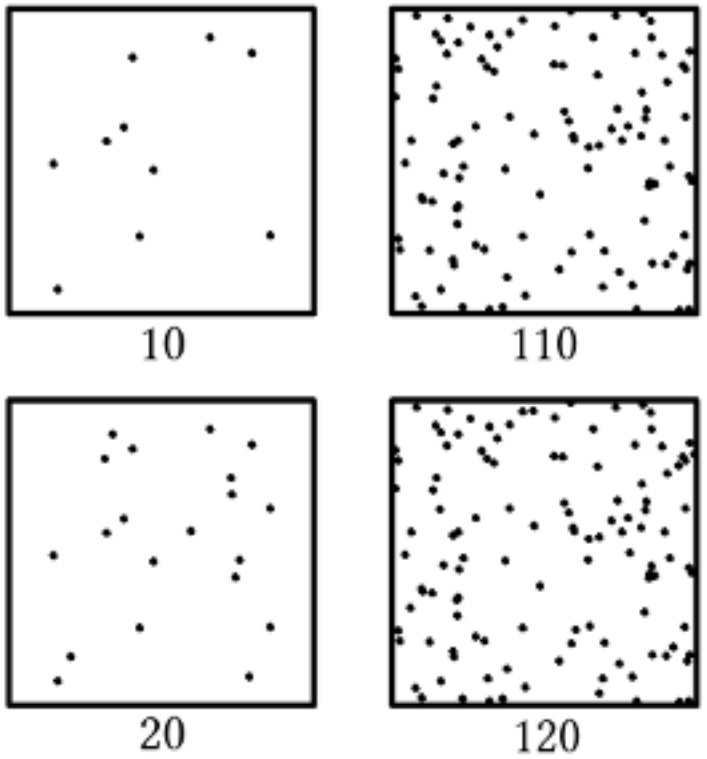
(Aaron Haspel, as quoted in Muller, The Tyranny of Metrics)

CAN WE MEASURE EXPERIENCE?

- Subjective measures vs instrumental (objective) measures
- What can we use to make measurements of the subjective?
- Self reporting?

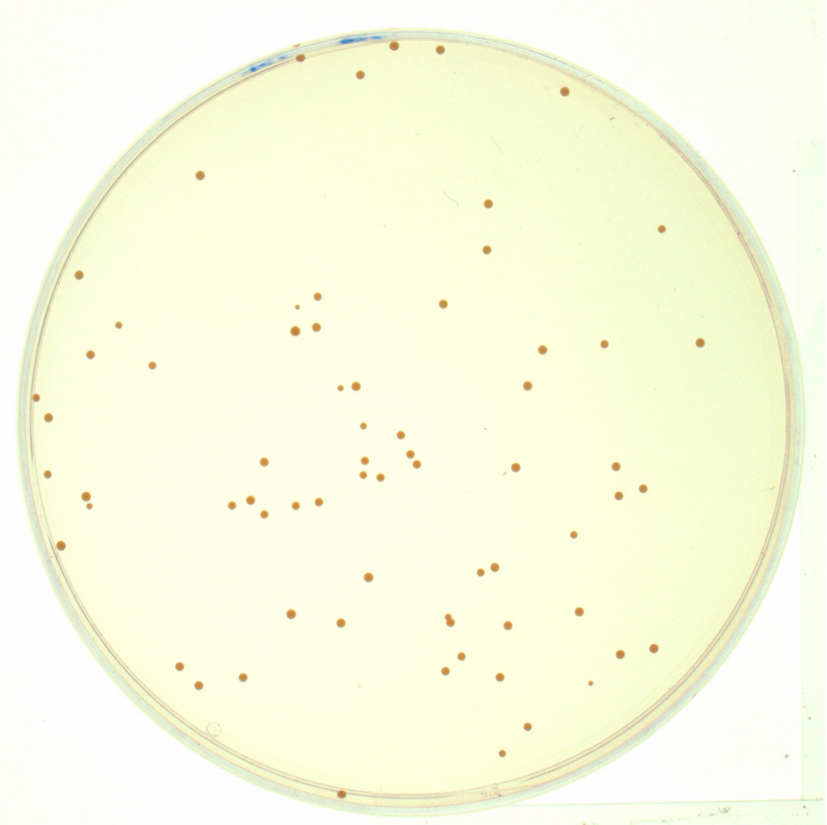
PSYCHOPHYSICS: WEBER-FECHNER LAW

A gloss: the relationship between stimulus and perception in humans is more logarithmic than linear.



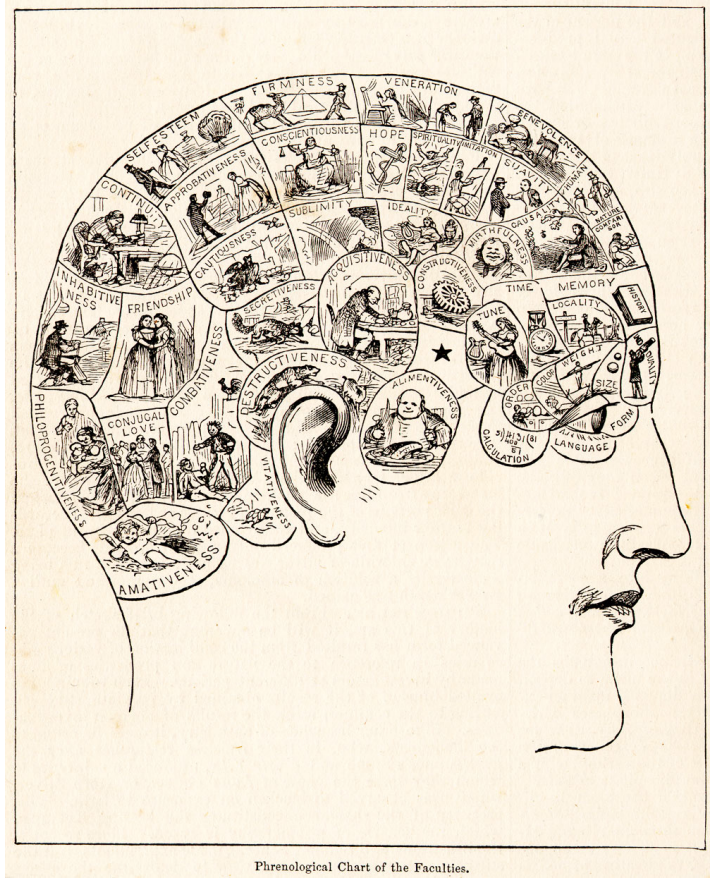
SOME ADDITIONAL RELEVANT CONCEPTS

- Proxy measures
- **Population measure vs individual measure**
- Measurement accuracy and precision
- Measurement error:
  - Random error
  - Systematic error



EXAMPLE OF POOR MEASUREMENT USE

- Phrenology and Craniometry
- Pseudo-science!
- The measurements are perhaps real, the conclusions drawn from the measurements – not so much!
- We will discuss more of this type of issue when we get into metrics.



Phrenological Chart of the Faculties.

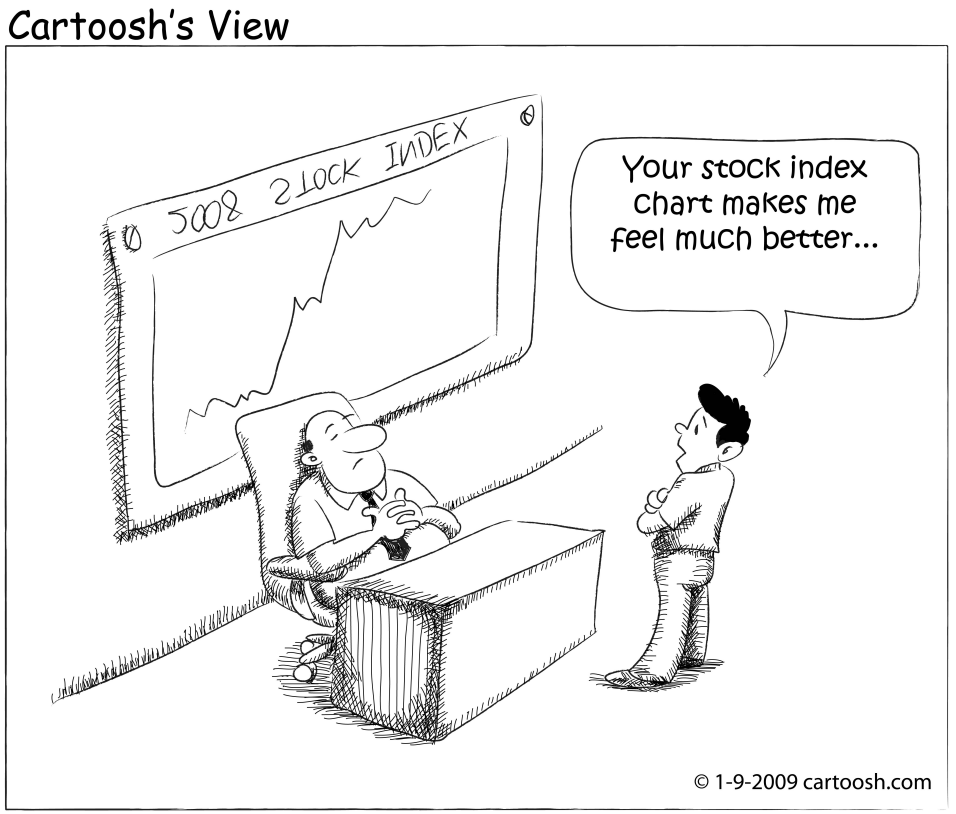
METRICS - DEFINITION AND CRITICAL ANALYSIS

DEFINITION OF METRICS

- A measure of a more abstract concept created from a a **combination** of (more basic) measures.
- A measure created *relative to a particular goal* (e.g. a performance metric).
- (not talking about a mathematical metric)

METRICS – SOME RELATED CONCEPTS

- **Index**: An observed (measured) value relevant to a concept of interest. Compare proxy measure.
- **Indicator**: An aggregation of multiple indicators. Compare metric.
- **Example**: Human Development Index.



(https://commons.wikimedia.org/wiki/File:20090109_stock_index-01.jpg)

CONCEPTUAL BUILDING BLOCKS

- Metrics are conceptually useful - they seemingly provide a useful high level summary of some element of an otherwise complex system.
- It can be an interesting intellectual challenge to take a collection of of raw measures and combine them to better understand something more conceptually abstract.
- In this respect, we can almost view a metric as a higher level concept that we are building out of more fundamental concepts.

NOT SO FAST?

- Hoping to provide solutions to real problems **BUT** the cure should not be as bad as the disease - or at least, not create even worse complications!
- What is measured is attended to, **BUT** what is not measured is neglected?
- Does this just mean we need more measures? More data?
- Well - this is great if you want to be a data practitioner.... **BUT** is this feasible or desirable?

REPURPOSING SCIENCE (AGAIN)

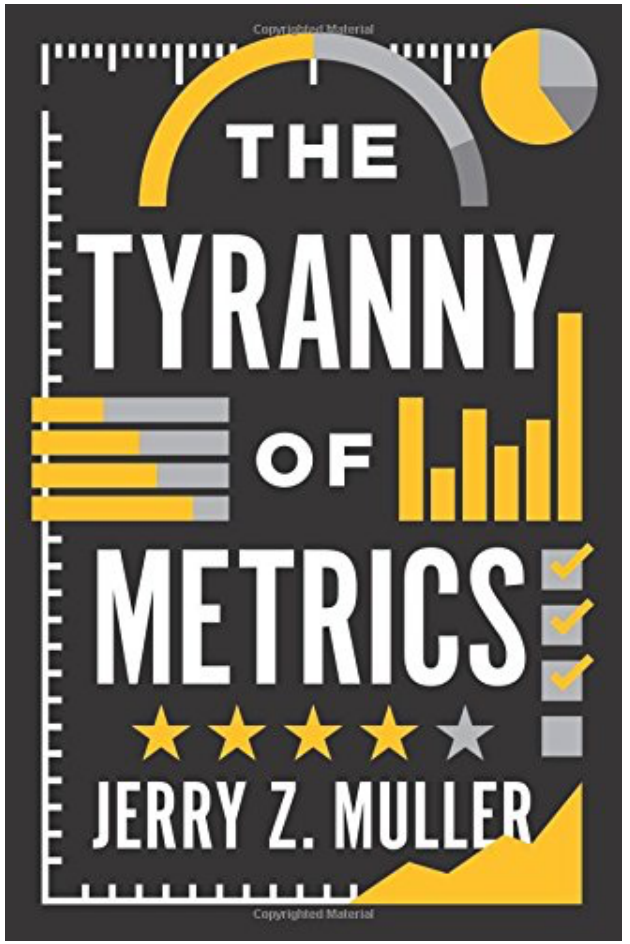
- In a sense, we are trying to move scientific behaviors - in this case measurement - out into the non-scientific realm - "Scientific Management" - there can be many pitfalls!
- Example- the issue of edge cases - our metric or proxy measure might work very well for the 'middle' cases, but very poorly for the edge cases.
- We want everything to be 'data-driven' - this presupposes that we can get data or evidence on everything.

FINDING A BALANCE

data cynicism **data realism** **data optimism**

THE TYRANNY OF METRICS

- The Tyranny of Metrics by Jerry Z. Muller discusses:
  - Recurring flaws seen in metrics construction
  - Origin of "Metric Fixation'
  - Case Studies
  - Issues with and strategies for avoiding ‘Metric Fixation'



ACCOUNTABILITY

- Accountable - two meanings:
  - To be responsible
  - To be countable
- Responsibility through counting?

SOCIAL-HISTORIC ORIGINS OF METRICS

- Have already seen that record keeping more generally has been around as long as 'civilization' has been around (Against the Grain, James C. Scott)
- 1860's - England – Finance - school performance metrics
- 1900s – Taylor – Engineering - Scientific Management
- 1950's -Robert McNamara the concept of the 'general manager' - objective numbers could help with this!
- 1980's - the rise of IT, and with it an increasing ability to track and collect data to generate metrics

WHAT HAPPENED NEXT?

"Since then, the growing opportunities to collect data, and the declining cost of doing so, contribute to the meme that data is the answer, for which organizations have to come up with the questions.

There is an often unexamined faith that amassing data and sharing it widely within the organization will result in improvements of some sort - even if much information has to be denuded of nuances and context to turn it into easily transferred 'data'. ”

(Muller, The Tyranny of Metrics)

ARE METRICS REDEEMABLE?

- "The problem is not measurement, but excessive measurement and inappropriate measurement... while they are a potentially valuable tool, the virtues of accountability metrics have been oversold and their costs are often under appreciated."
- "...[T]here are many situations where decision making based on standardized measurement is superior to judgment based upon personal experience and expertise... Used judiciously, then, measurements of the previously unmeasurable can provide real benefits."
- So when are metrics good and when are they problematic? Muller spends the rest of his book spelling this out.

GOODHART'S + CAMPBELL'S LAW

- "When a measure becomes a target, it ceases to be a good measure."

  (Goodhart's Law)

- "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."

  (Campbell's Law)

METRIC FIXATION

- ‘Metric Fixation’ is a term coined by Muller to describe:
  - The belief that it is possible and desirable to replace judgment, acquired by personal experience and talent, with numerical indicators of comparative performance based upon standardized data (metrics);
  - The belief that making such metrics public (transparent) assures that institutions are actually carrying out their purposes (accountability)
  - The belief that the best way to motivate people within these organizations is by attaching rewards and penalties to their measured performance, rewards that are either monetary (pay-for-performance) or reputational (rankings).

FUNCTIONAL ISSUES WITH METRICS (I)

- **Distortion of information**:
  - Measuring the most easily measurable
  - Measuring the simple when the desired outcome is complex
  - Measuring inputs rather than outcomes (e.g. measuring resources provided to a project rather than the outcome of the project)
  - Degrading information quality through standardization

FUNCTIONAL ISSUES WITH METRICS (II)

- **Gaming the metrics**:
  - Gaming through creaming
  - Improving numbers by lowering standards
  - Improving numbers through omission or distortion of data
  - Cheating

OUT WITH ALL METRICS?

- Muller's main point: **Don't use metrics for reward or punishment**.
- He sites a number of examples (e.g. in education, medicine, policing) where metrics were indeed effective in improving a situation that needed improving.

CONSIDERATIONS WHEN DESIGNING METRICS (I)

- **Mueller's checklist of questions and points to consider**:
  - What kind of information are you thinking of measuring?
  - How useful is the information?
  - How useful are more measurements?
  - What are the costs of not relying on standardized measurements?
  - To what purpose will the measurement be put/to whom will information be made transparent?

CONSIDERATIONS WHEN DESIGNING METRICS (II)

- **Mueller's checklist of questions and points to consider (cont.)**:
  - What are the costs of acquiring the metric?
  - Why are people demanding this metric?
  - How and by whom are the measures of performance being developed?
  - Remember that even the best measures are subject to corruption and goal diversion.
  - Recognizing the limits of the possible is the beginning of wisdom.

CASE STUDY - METRICS AND CLIMATE CHANGE

CLIMATE CHANGE / GLOBAL WARMING

- "Global warming is a long-term rise in the average temperature of the Earth's climate system, an aspect of climate change shown by temperature measurements and by multiple effects of the warming.[2][3]
- The term commonly refers to the mainly human-caused observed warming since pre-industrial times and its projected continuation,[4] though there were also much earlier periods of global warming.[5]
- In the modern context the terms global warming and climate change are commonly used interchangeably,[6] but climate change includes both global warming and its effects, such as changes to precipitation and impacts that differ by region.[7][8]"

METRICS FOR CLIMATE CHANGE

Climate change is an area that is both complex and far-reaching - it would seem measures and metrics could be very beneficial in tackling the issue.

Some sub-disciplines within this area - what metrics might you develop to evaluate:

- **Scientific research projects that address or are relevant to climate change issues**
- **The behavior of public and private sector banks as it relates to climate change**
- **The impact and contribution of federal climate change policies of countries relative to other countries**
- **Climate Change Impacts on Agriculture and Agricultural ability to adapt to climate change**

What data would you need to collect? How would you then work with this?

EXAMPLE 1: METRICS FOR CLIMATE CHANGE RESEARCH

- Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program (2005)
- Looking to build metrics that will measure the progress of the Climate Change Science Program (CCSP) - and projects within this program.
- In this case, they define metrics as "**simple qualitative or quantitative measures of performance with respect to a stated goal**" (p. 11)
- A very nice example of a carefully thought out approach to developing metrics.

CCSP OVERARCHING GOALS (I)

- Improve knowledge of the Earth's past and present climate and environment, including its natural variability, and improve understanding of the causes of observed variability and change.
- Improve quantification of the forces bringing about changes in the Earth's climate and related systems.
- Reduce uncertainty in projections of how the Earth's climate and related systems may change in the future.

CCSP OVERARCHING GOALS (II)

- Understand the sensitivity and adaptability of different natural and managed ecosystems and human systems to climate and related global changes.
- Explore the uses and identify the limits of evolving knowledge to manage risks and opportunities related to climate variability and change.

CCSP GOALS OF METRIC DEVELOPING PROJECT

- Provide a general assessment of how well CCSP objectives lend themselves to quantitative metrics.
- Identify three to five areas of climate change and global change research that can and should be evaluated through quantitative performance measures.
- For these areas, recommend specific metrics for documenting progress, measuring future performance (such as skill scores, correspondence across models, correspondence with observations), and communicating levels of performance.
- Discuss possible limitations of quantitative performance measures for other areas of climate change and global change research.

END RESULT

- Fairly generic metrics (what they call 'general metrics') relating to, essentially scientific project management:
  - 5 process metrics
  - 5 input metrics
  - 5 output metrics
  - 6 outcome metrics
  - 3 impact metrics

EXAMPLE 2: PUBLIC AND PRIVATE BANKING BEHAVIOUR

- PORTFOLIO CARBON INITIATIVE - EXPLORING METRICS TO MEASURE THE CLIMATE PROGRESS OF BANKS (2018)
- "This paper informs the ongoing debate about how public- and private-sector banks should assess and report on their contribution to the transition toward a low-carbon economy."

Table ES-1 | Climate Progress Metrics

CATEGORIES OF METRICS	SPECIFIC TYPES OF METRICS	DESCRIPTION
Greenhouse Gas Accounting	Corporate accounting	Corporate-level tracking of annual GHG emissions related to a company's operations
	Project accounting	Estimating net GHG emissions or emission reductions from projects relative to a baseline scenario
	Financed emissions	(Generally) portfolio level aggregation of GHG emissions associated with a portfolio's underlying entities or projects, allocated proportionally, based on financial stake in the underlying entity or project
Green/Brown Metrics	Exposure-based	Metrics that measure climate progress of a project, activity, or asset class in terms of exposure in financial terms such as $ invested in green energy, counts such as number of energy star buildings in a real estate portfolio, or percentages such as % car loans to hybrids. Metrics could also be ratios such as $ invested in hybrids or total $ invested in cars
Sector-Specific Energy and Carbon Metrics	Physical unit-based (e.g., kWh, ft^2, km, etc.)	Metrics that are specific to a sector and expressed in absolute units (e.g., kWh generated) or intensity units (kWh/ft^2). Metrics can also be expressed in ratios such as KWh from green energy or total Kwh generated from power generation

Source: Authors.

EXAMPLE 3: COMPARISON OF CLIMATE CHANGE POLICIES

- Metrics for Evaluating Policy Commitments in a Fragmented World: The Challenges of Equity and Integrity (2008)
- Considers a number of alternative approaches to measuring climate policy contributions, including measures of
  - emissions performance
  - emission reductions
  - [total or marginal] costs
- Each can provide some valuable information, but none is terribly satisfying as a reliable measure of effort or equity.

A USEFUL PROBLEM STATEMENT ANALOGY

- The challenges of moving ahead in a world of diverse policies are illustrated by a simple story:
- Two individuals are approached on the street by a sympathetic homeless person seeking assistance. Person A, an established professional with a relatively high income, proposes to rent the homeless individual an apartment for six months. Person B, a younger, struggling academic, offers to donate $100. Regrettably, neither one can make good on their offers immediately. However, they both agree to return to the same location at an appointed time the following week to complete the transactions.

A NOTE ON THE CHALLENGES OF METRICS

- “There are no ‘neutral’ metrics: different metrics will show different countries in a good (or less good) light” (OECD, p. 6). Furthermore, subjective views may change: whatever seems fair at one time may well be perceived differently in the future.”

EXAMPLE 4: METRICS FOR AGRICULTURE

- Developing Climate Change Impacts and Adaptation Metrics for Agriculture (Cynthia Rosenzweig and Francesco N. Tubiello)
- Their definition of metrics:
  - "A set of metrics can be defined as a system of measurement, one that can be used in an objective, transparent, and reproducible manner to describe the characteristics and transformations of observable systems. **Metrics carry with them – implicitly or explicitly – a definition of the system being measured, as well as the set of measurement units to be used**." (emphasis mine)

GOAL OF PROJECT AND METRICS

- “Within the context of this work, criteria for developing agricultural metrics were investigated, in order to define and characterize the status of given agricultural production systems against the changing climate of the coming decades, with a focus on both short-term (20-30 years) and long-term (80-100 years) horizons.”
- “The underlying idea in this exercise is that a set of such metrics can be used by decision-makers to provide an easy-to-understand “health report,” or a snapshot of an agricultural system, with regard to the likely risks of climate change impacts in coming decades.”

FINAL (DRAFT) METRICS (I)

- The metrics proposed are, in a sense, less abstract than some other examples but nonetheless difficult to interpret by non-experts:
  - **Crop Yield**: Ton/ha
  - **Yield Variability**: CV Long-term standard deviation from mean over mean yield (%)
  - **Production Level**: At local to regional and national levels (Ton/yr)
  - **Economic Value at Risk Net**: production value at local to regional level. Agricultural GDP at national level ($) .

FINAL (DRAFT) METRICS (II)

- Final metrics (cont.)
  - **Land Value at Risk**: Land value of areas most affected ($)
  - **Changes in Event Frequency**: Impacts of increased frequency of droughts/floods on damage (Ton and/or $)
  - **Nutrition Index**: Food demand over supply (sum of internal production and trade)
  - **Water Requirements/Withdrawals**: Irrigation water requirements over available resources.
- Although we might imagine how to calculate these, we don't necessarily immediately understand how they would be relevant to the question at hand.

HANDS-ON: USING R FUNCTIONS TO CREATE METRICS FROM MEASURES