

# A Big Storm's A-Comin

Big data, informal education, & cultural collisions

Science on a Sphere Working Group – Kansas City  
11/28/2018







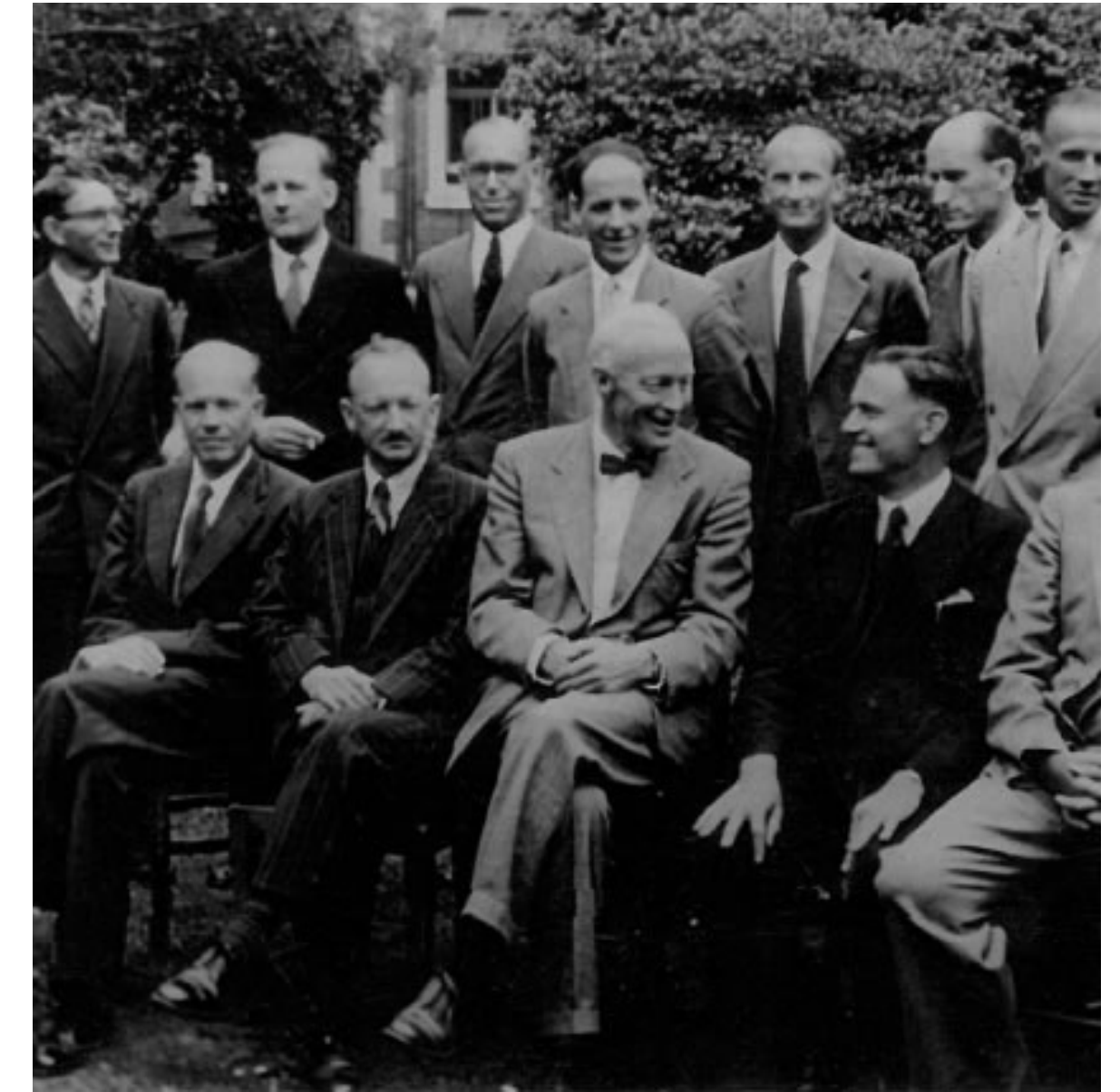
Type to enter a caption.



Type to enter a caption.

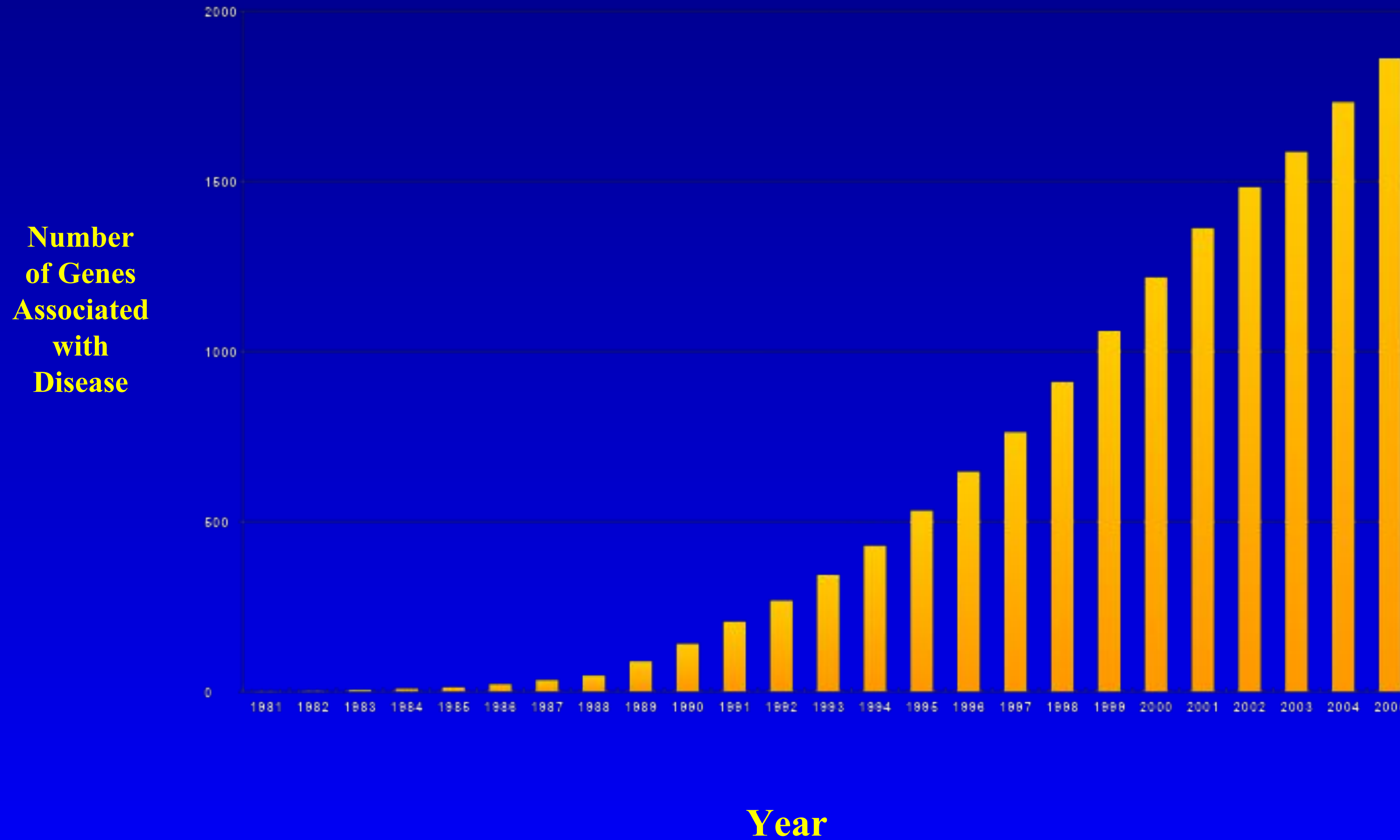
# Fancy Schmancy

So smart, much brainy, wow

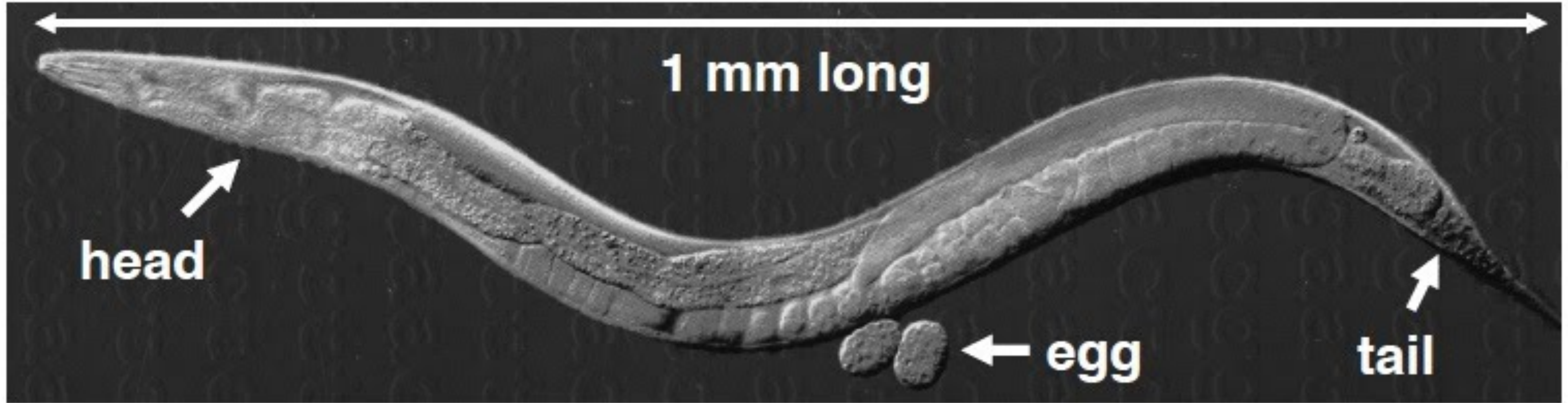




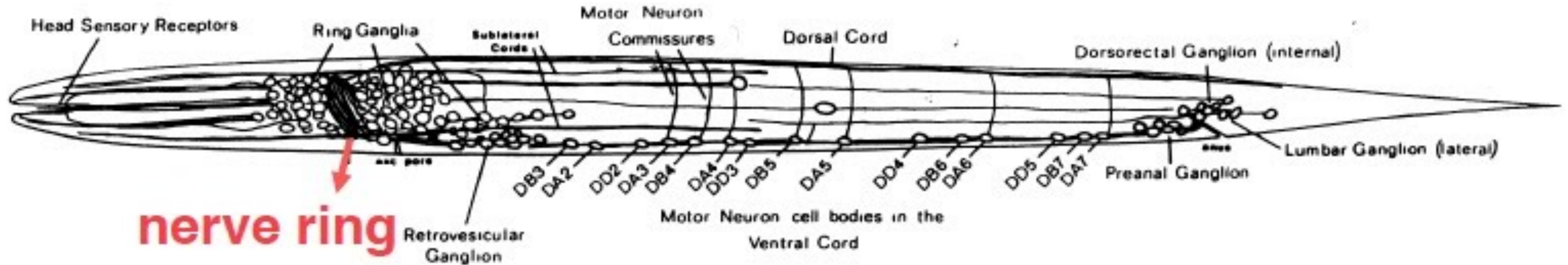
# Cumulative Pace of Disease Gene Discovery 1981-2005



Source: Online Mendelian Inheritance in Man



## *C. elegans* hermaphrodite adult





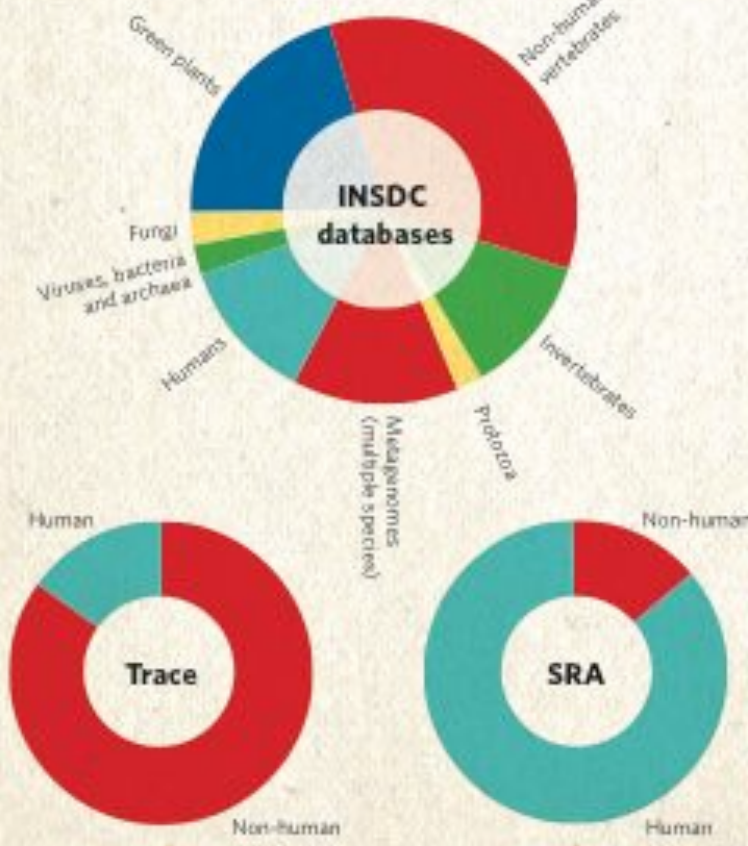


# THE SEQUENCE EXPLOSION

At the time of the announcement of the first drafts of the human genome in 2000, there were 8 billion base pairs of sequence in the three main databases for 'finished' sequence: GenBank, run by the US National Center for Biotechnology Information; the DNA Databank of Japan; and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database. The databases share their data regularly as part of the International Nucleotide Sequence Database Collaboration (INSDC). In the subsequent first post-genome decade, they have added another 270 billion bases to the collection of finished sequence, doubling the size of the database roughly every 18 months. But this number is dwarfed by the amount of raw sequence that has been created and stored by researchers around the world in the Trace archive and Sequence Read Archive (SRA). See Editorial, page 649, and human genome special at [www.nature.com/humangenome](http://www.nature.com/humangenome)

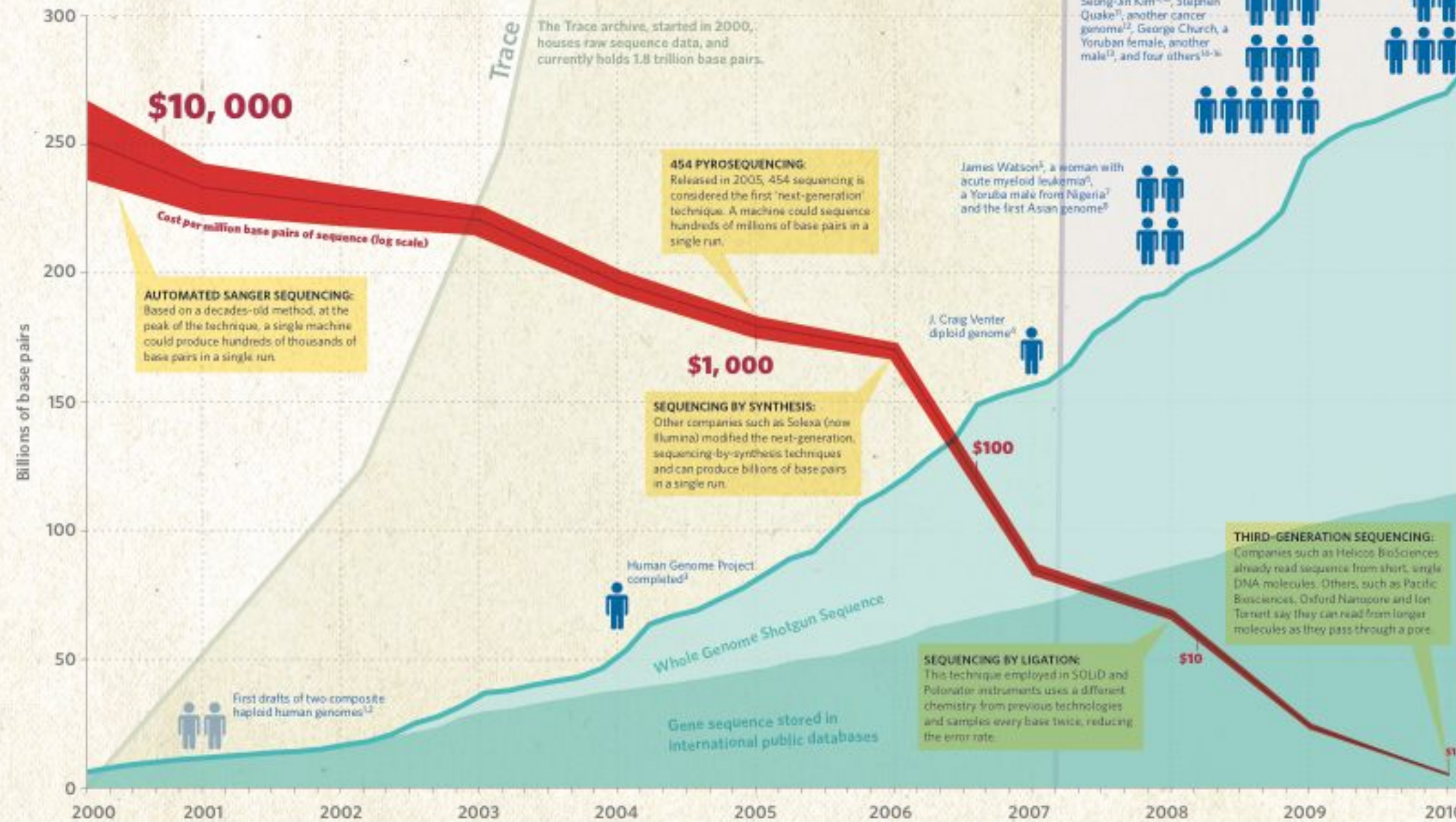
## DNA SEQUENCES BY TAXONOMY

**International Nucleotide Sequence Database Collaboration:** The main repositories of 'finished' sequence span a wide range of organisms, representing the many priorities of scientists worldwide.



**Trace Archive:** Developed to house the raw output of high-throughput sequencers built in the late 1990s, the trace archive spans a wide range of taxa.

**Sequence Read Archive:** Houses raw data from next-generation sequencing. Dominated by human sequence, including multiple coverage for more than 170 people.



The Sequence Read Archive (SRA) houses raw data from next-generation sequencing and has grown to 25 trillion base pairs. If this chart were to accommodate it, it would stretch to more than 12 metres — twice the height of an average giraffe.

A glioma cell line<sup>17</sup>, Imuk<sup>18</sup>, Gubi and Archbishop Desmond Tutu<sup>19</sup>, James Lupski<sup>20</sup>, and a family of four<sup>21</sup>

Two Korean males including Seong-In Kim<sup>13,10</sup>, Stephan Quake<sup>11</sup>, another cancer genome<sup>12</sup>, George Church, a Yoruban female, another male<sup>13</sup>, and four others<sup>14-16</sup>

James Watson<sup>5</sup>, a woman with acute myeloid leukemia<sup>6</sup>, a Yoruba male from Nigeria<sup>7</sup> and the first Asian genome<sup>8</sup>

J. Craig Venter diploid genome<sup>9</sup>

THIRD-GENERATION SEQUENCING: Companies such as Helicos BioSciences already read sequence from short, single DNA molecules. Others, such as Pacific Biosciences, Oxford Nanopore and Ion Torrent say they can read from longer molecules as they pass through a pore.

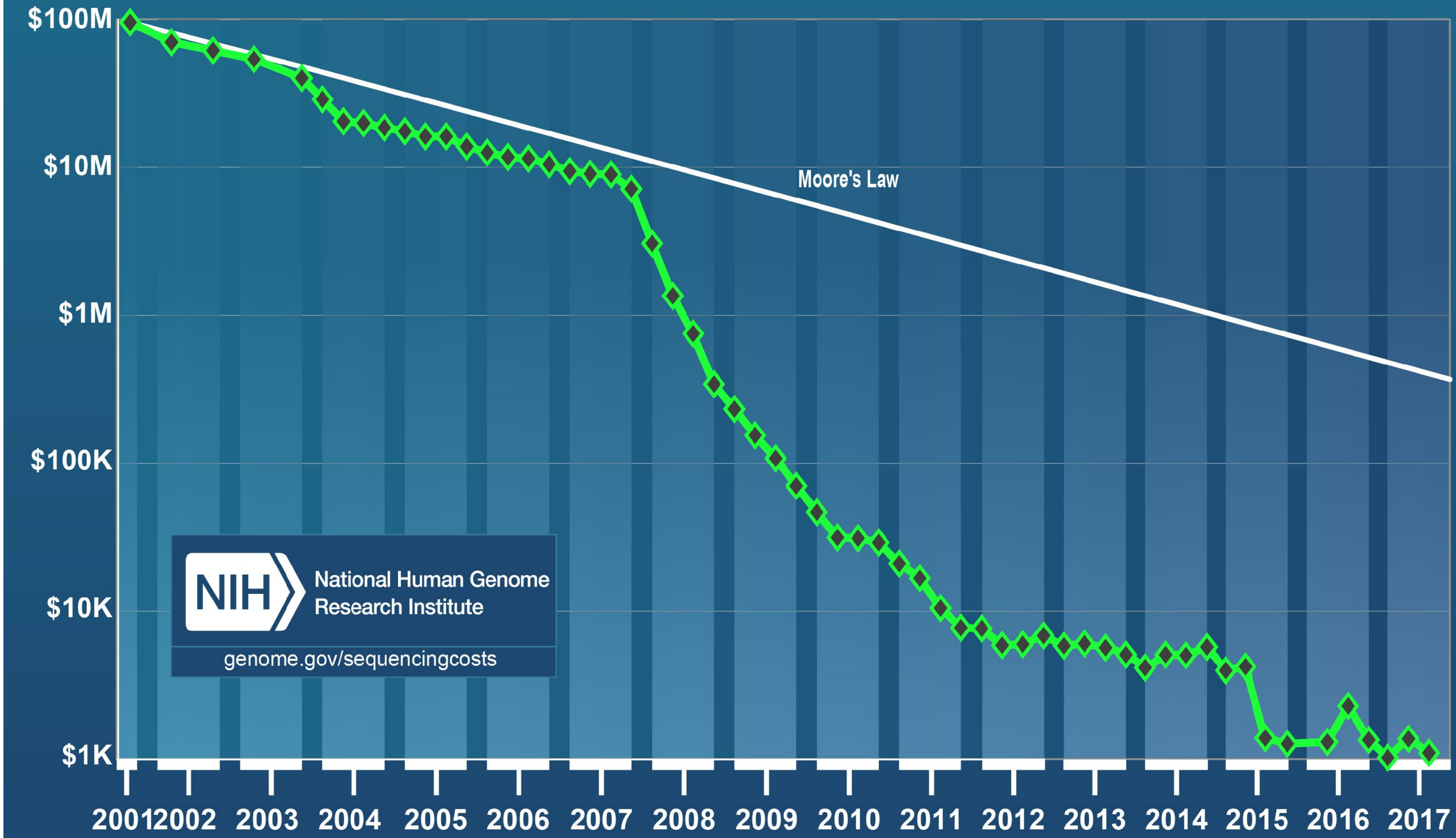
## HOW MANY HUMAN GENOMES?

The graphic shows all published, fully sequenced human genomes since 2000, including nine from the first quarter of 2010. Some are resequencing efforts on the same person and the list does not include unpublished completed genomes.

- Venter, J. C. et al. *Science* **291**, 1304-1351 (2001).
- International Human Genome Sequencing Consortium *Nature* **409**, 860-921 (2001).
- International Human Genome Sequencing Consortium *Nature* **431**, 931-945 (2004).
- Lavy, S. et al. *PLoS Biol.* **5**, e254 (2007).
- Wheeler, D. A. et al. *Nature* **452**, 872-876 (2008).
- Lay, T. J. et al. *Nature* **456**, 66-72 (2008).
- Bentley, D. R. et al. *Nature* **456**, 53-59 (2008).
- Wang, J. et al. *Nature* **456**, 60-65 (2008).
- Ali, S.-M. et al. *Genome Res.* **19**, 1622-1629 (2009).
- Kim, J.-I. et al. *Nature* **460**, 1011-1015 (2009).
- Pushkarev, D., Nair, N. F. & Quake, S. R. *Nature Biotechnol.* **27**, 847-850 (2009).
- Mardis, E. R. et al. *N. Engl. J. Med.* **10**, 1058-1066 (2009).
- Drmanac, R. et al. *Science* **327**, 78-81 (2009).
- McKernan, K. J. et al. *Genome Res.* **19**, 1527-1541 (2009).
- Pleasanton, E. D. et al. *Nature* **463**, 191-196 (2010).
- Pleasanton, E. D. et al. *Nature* **463**, 184-190 (2010).
- Clark, M. J. et al. *PLoS Genet.* **6**, e1000832 (2010).
- Rasmussen, M. et al. *Nature* **463**, 757-762 (2010).
- Schuster, S. C. et al. *Nature* **463**, 943-947 (2010).
- Lupski, J. R. et al. *N. Engl. J. Med.* doi:10.1056/NEJMo0908094 (2010).
- Roach, J. C. et al. *Science* doi:10.1126/science.1186802 (2010).

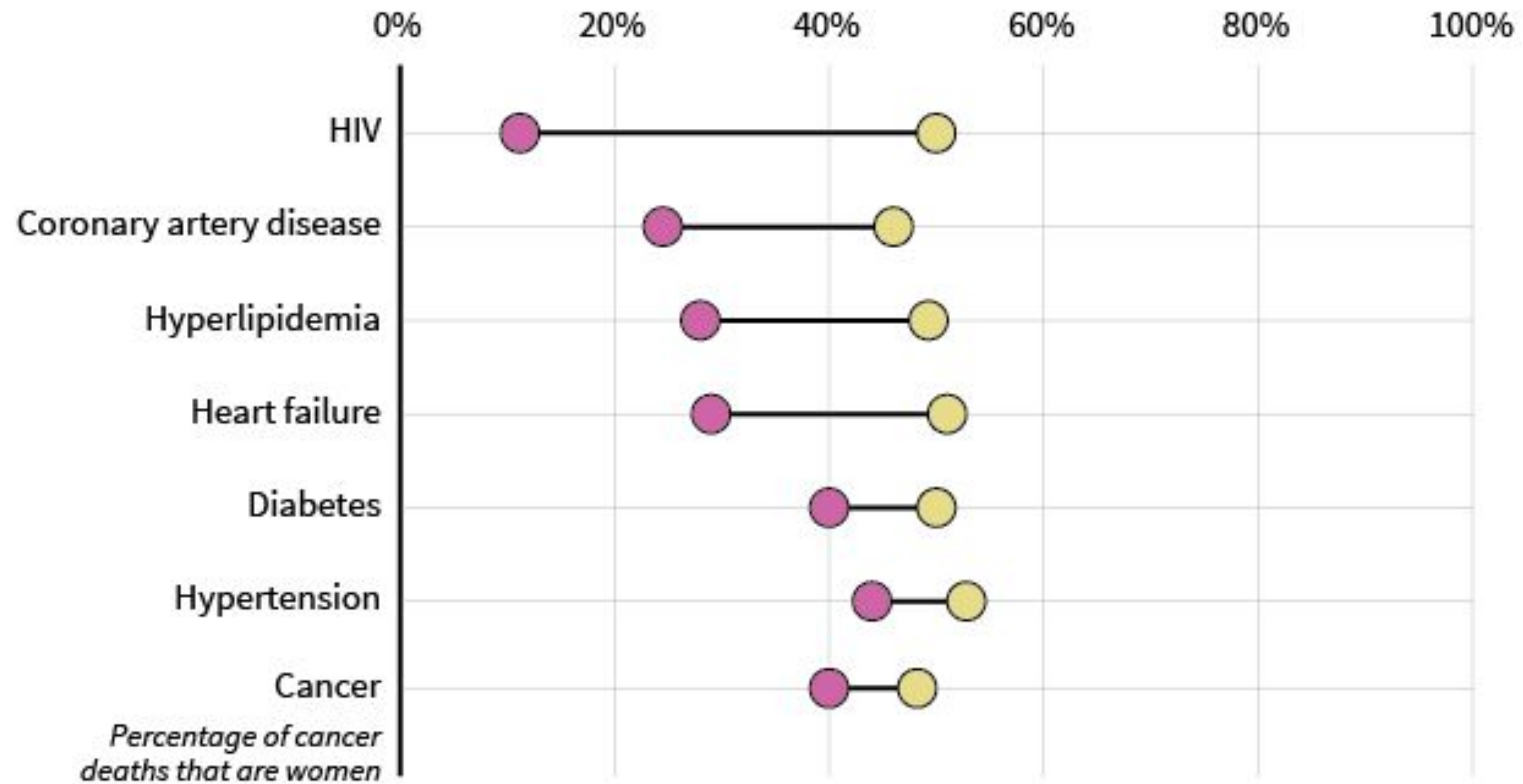
Page size by comparison

# Cost per Genome



# Women Are Underrepresented In Clinical Trials

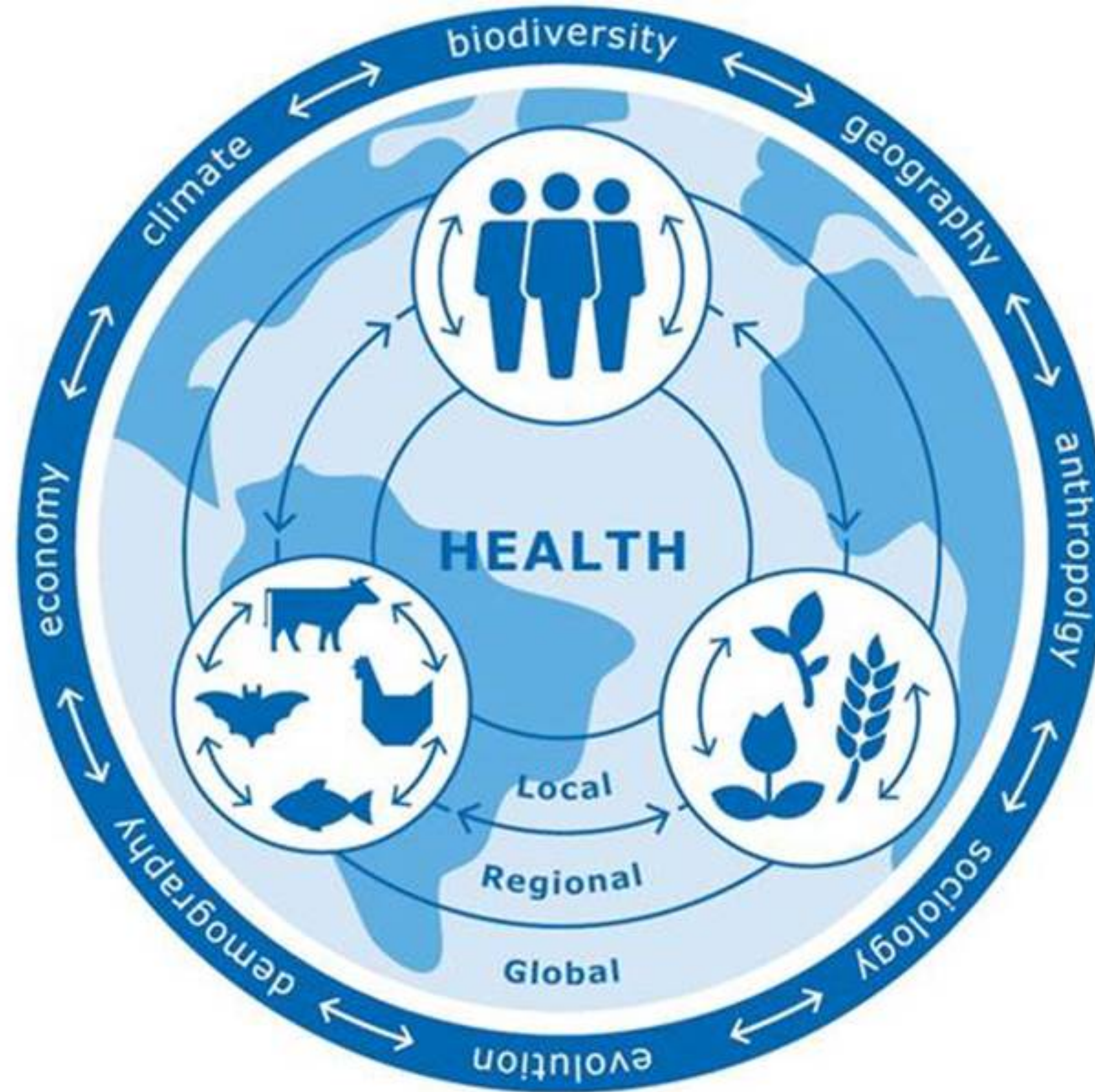
- Percent of clinical trial participants that are women
- Percent of cases that are women



Source: BMC Women's Health, Cardiovascular Quality and Outcomes

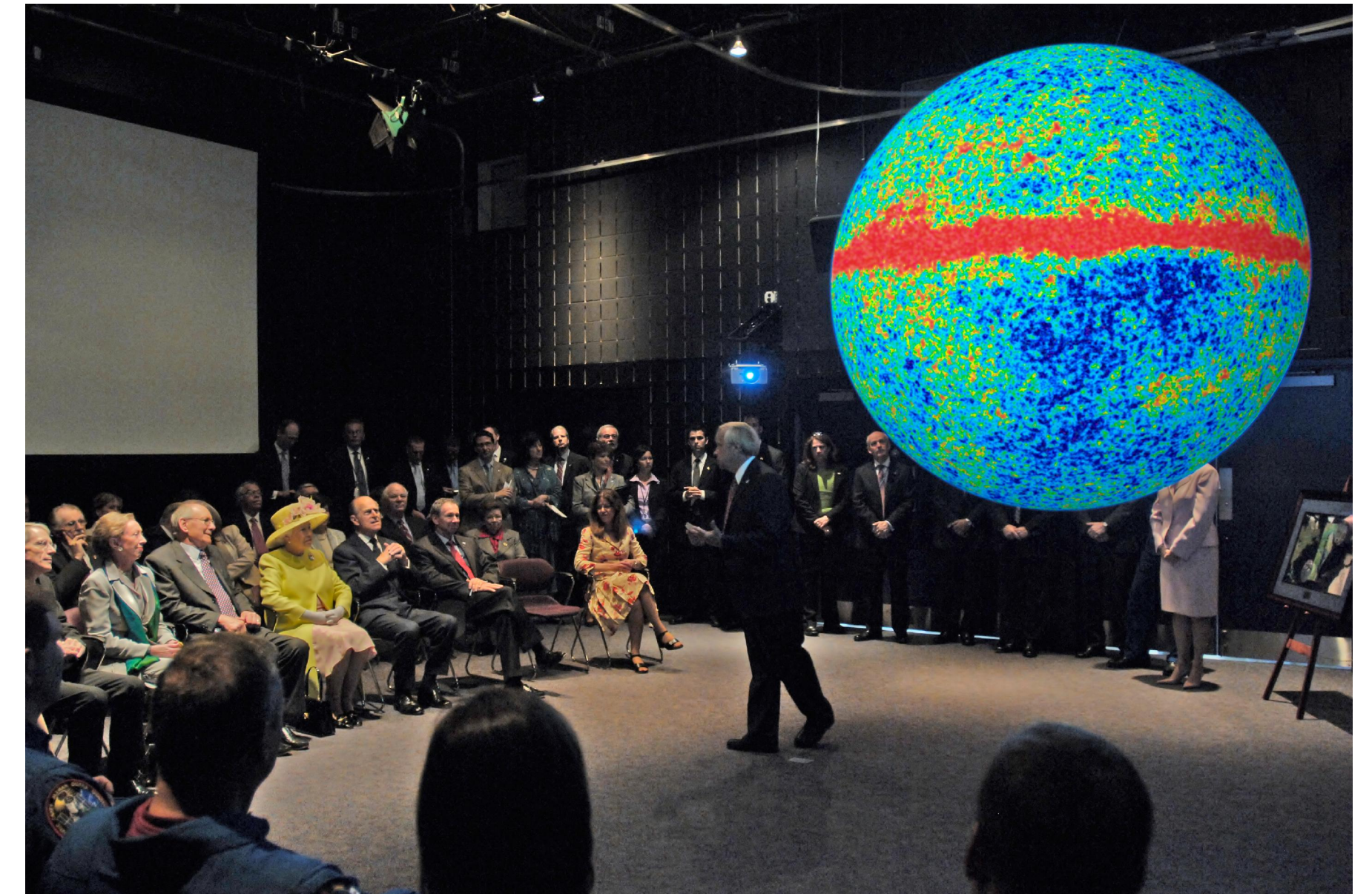
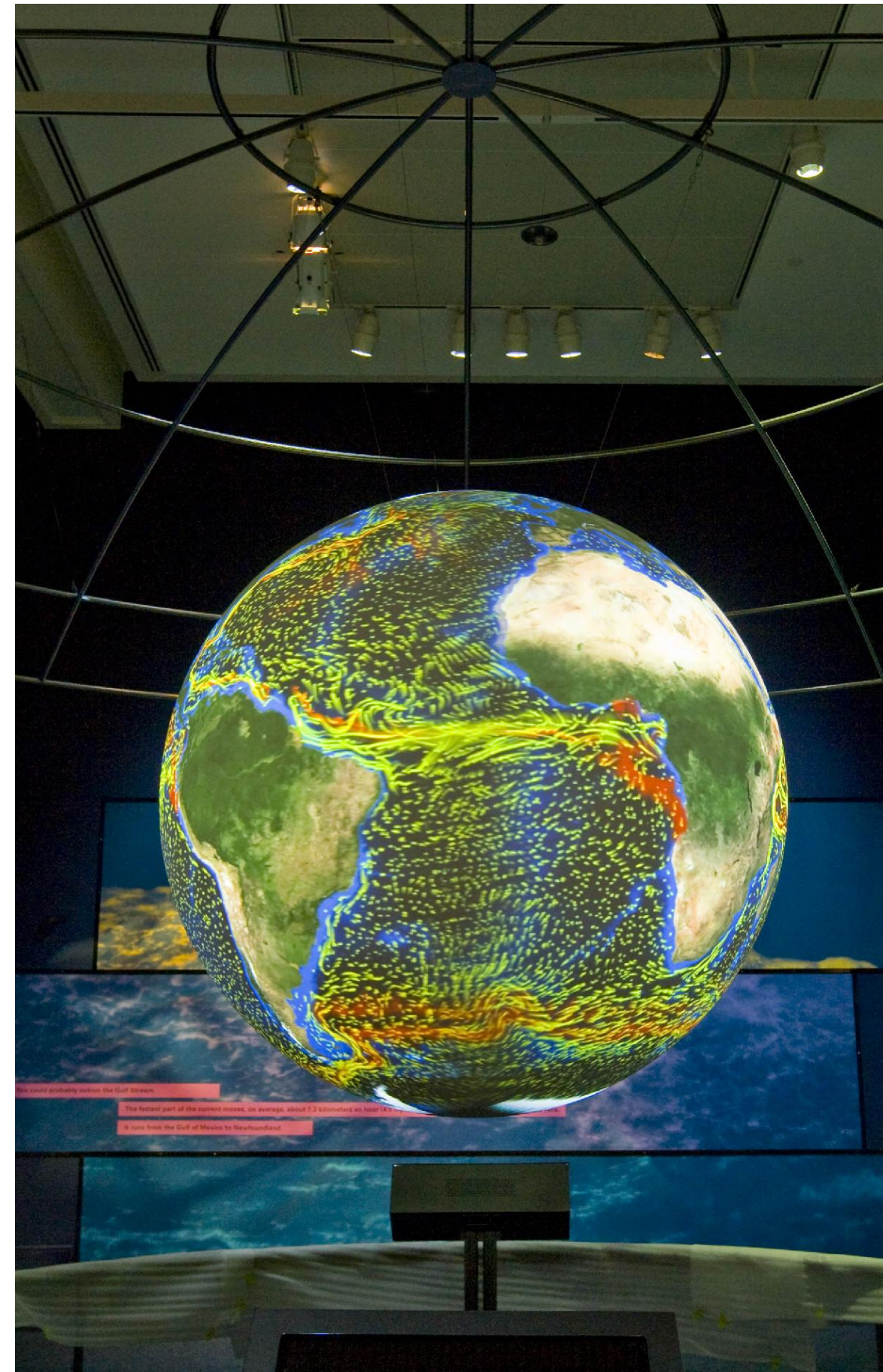
THE HUFFINGTON POST









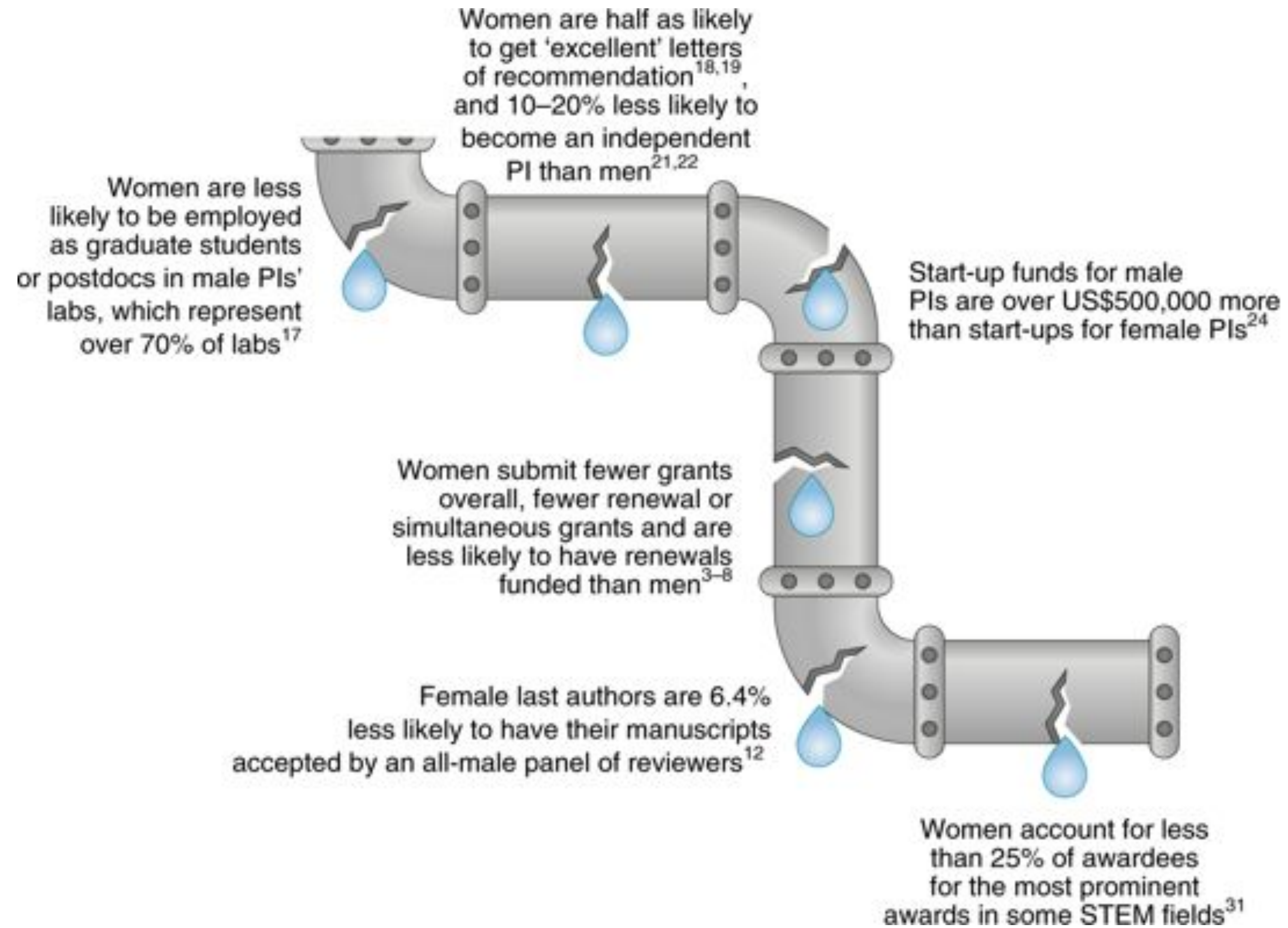


my stock photo



**MUSEUMS  
ARE NOT  
NEUTRAL**

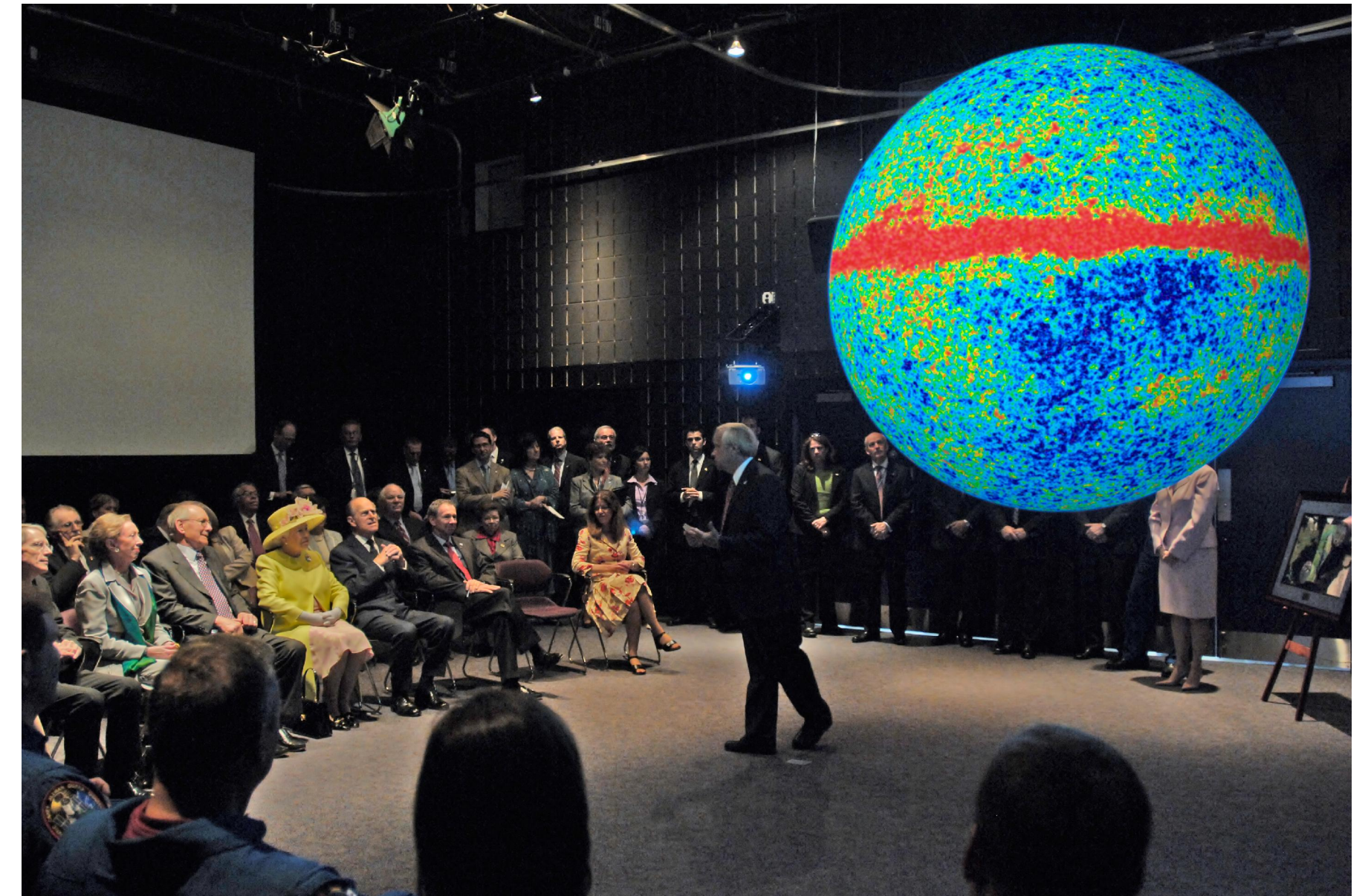
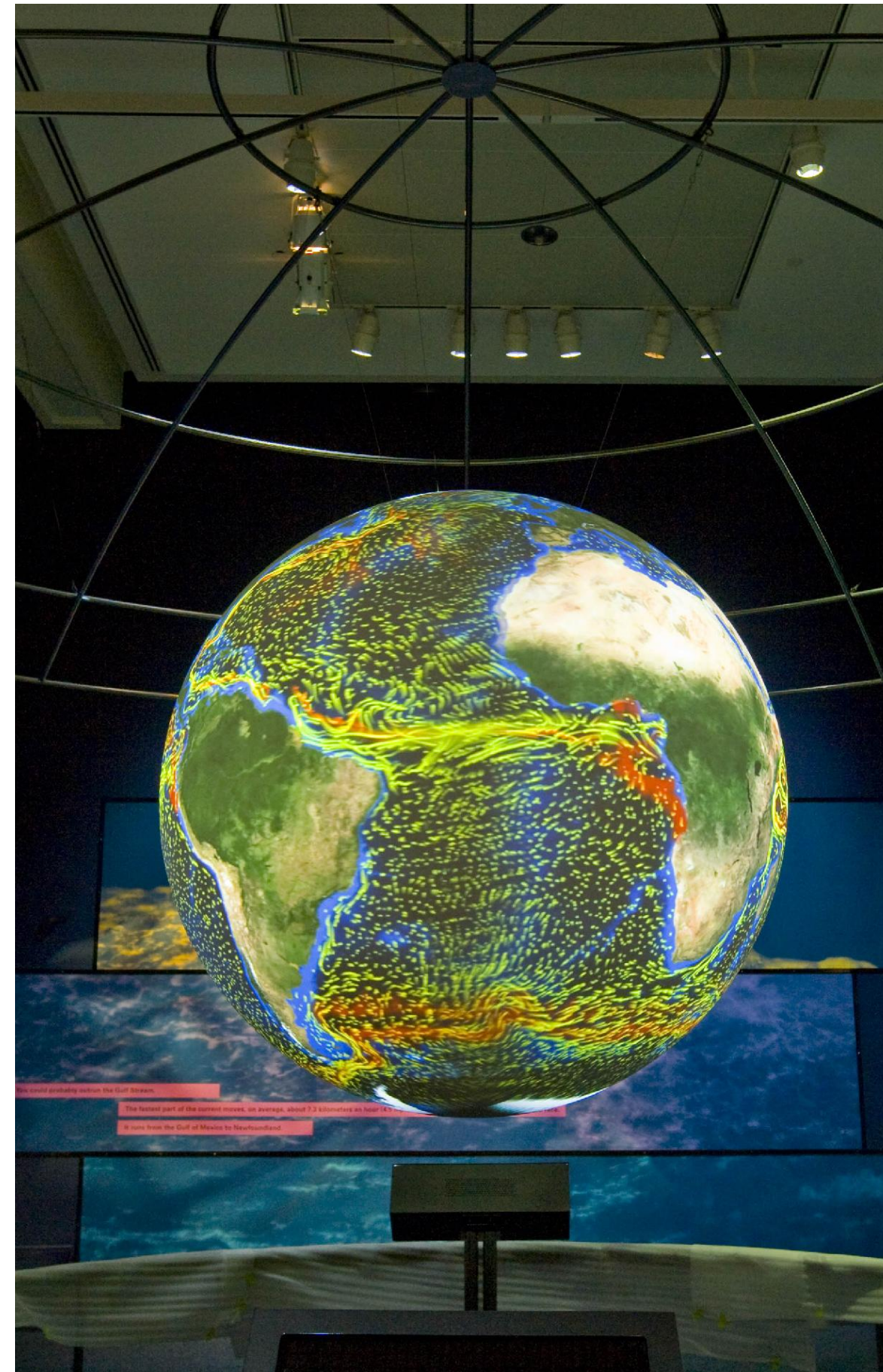




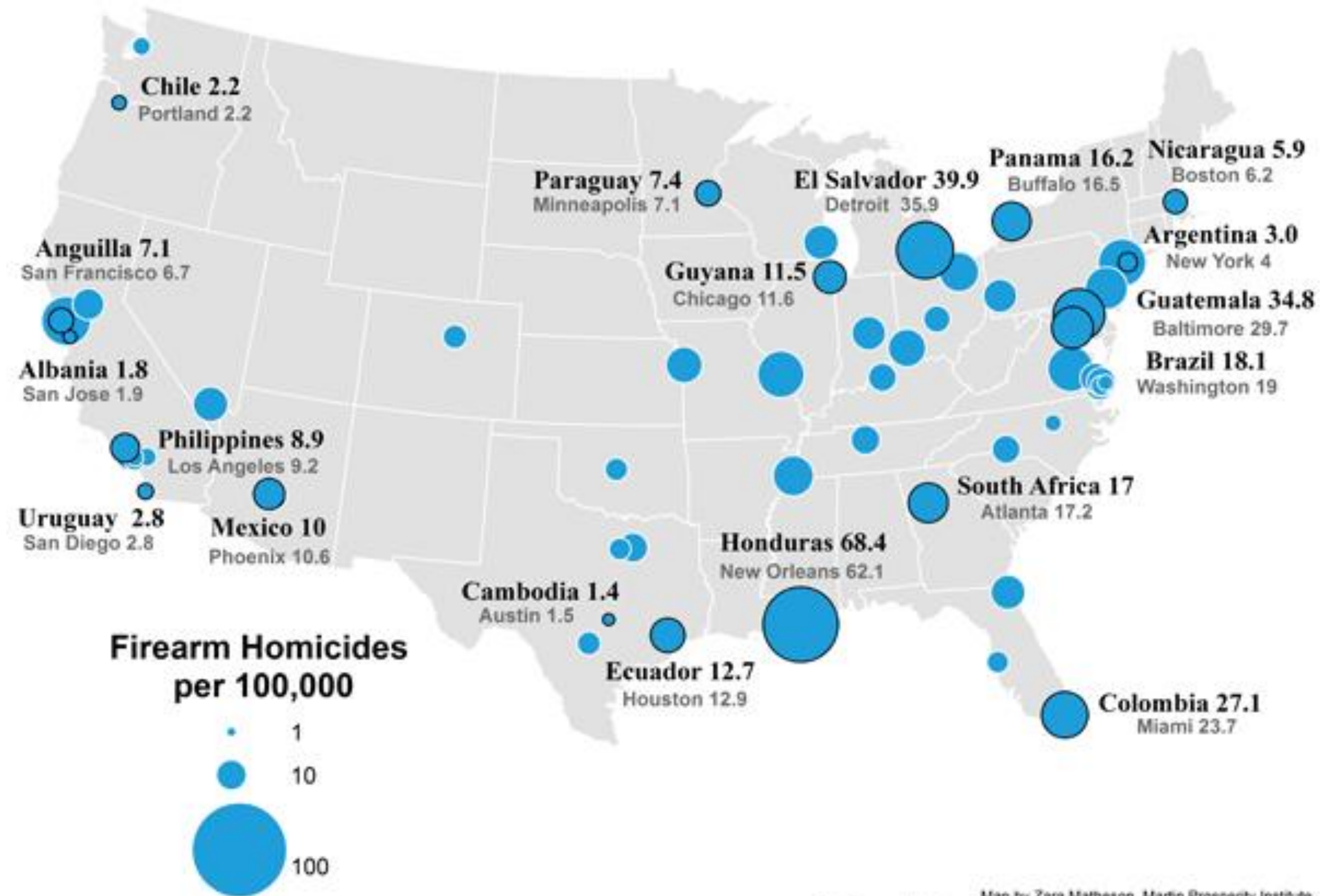
How can the entire STEM community address this gender bias problem?



- Collect and publish data on gender representation in all areas of STEM visibility and success
- Set clear and specific goals and guidelines for gender and minority representation in your area of STEM
- Identify and offer effective training in unconscious bias and effective hiring and retention strategies



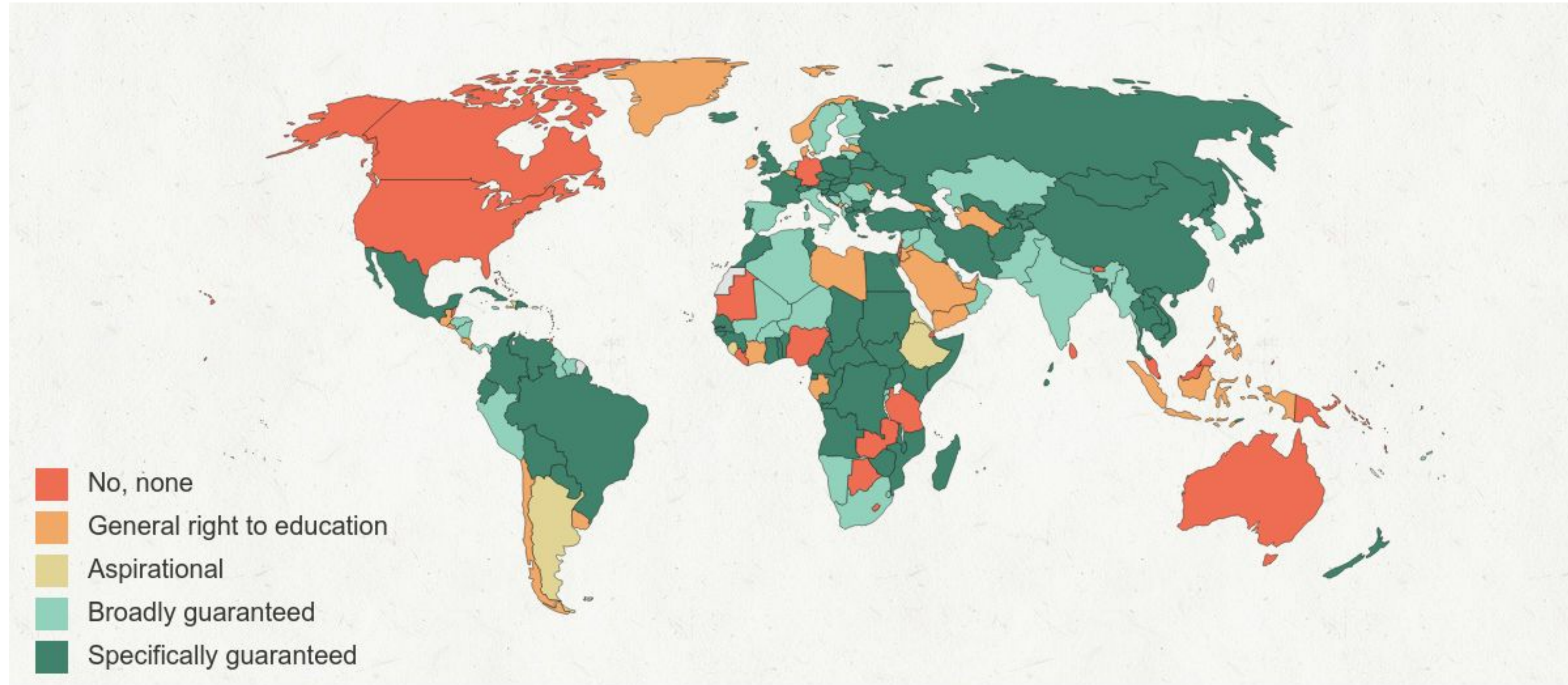
my stock photo



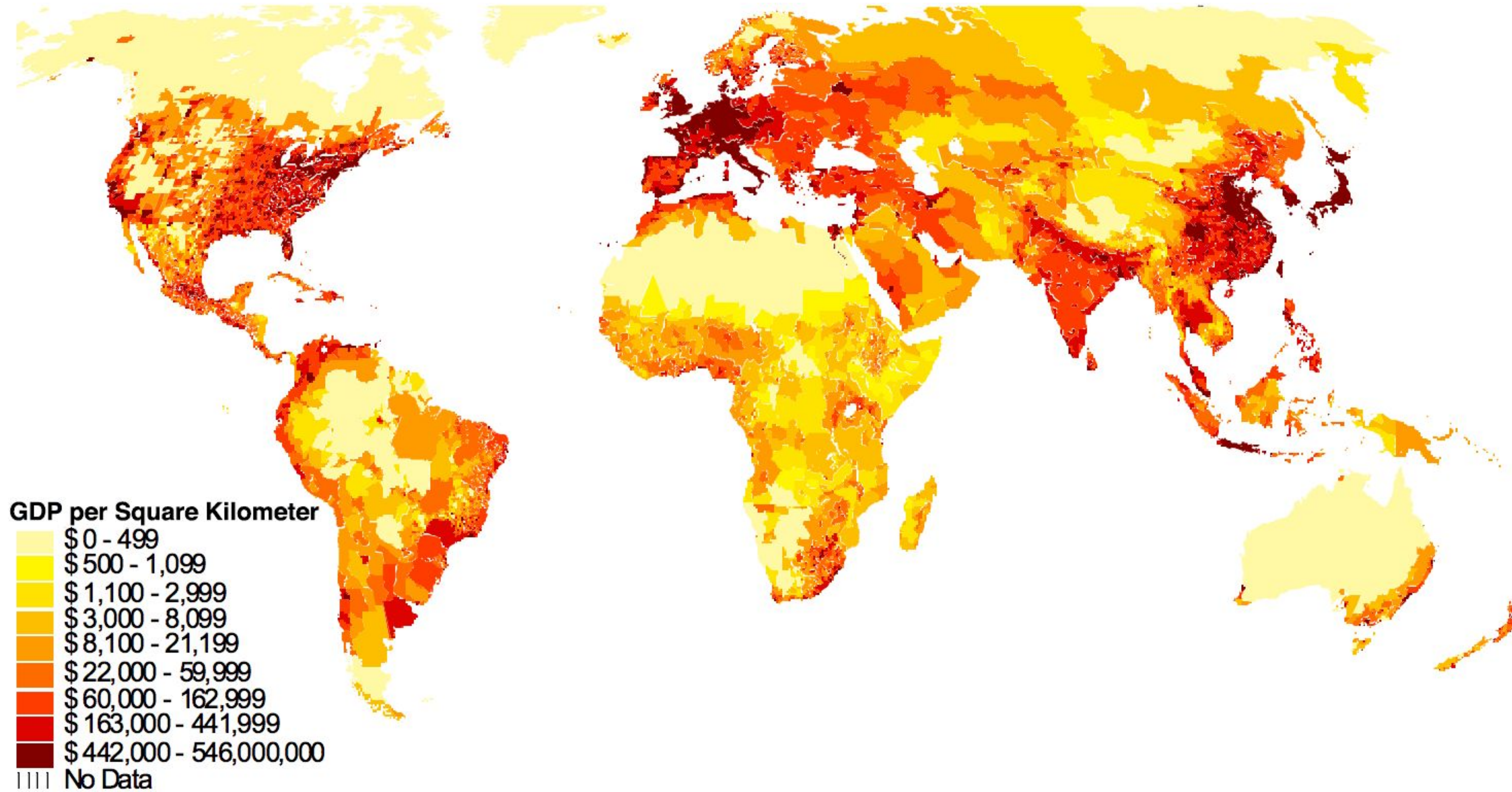
Map by Zara Matheson, Martin Prosperity Institute  
Data Source: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6018a1.htm>  
<http://www.guardian.co.uk/news/datablog/2012/jul/22/gun-homicides-ownership-world-list>



## Does the constitution protect the right to education regardless of gender?



[www.worldpolicycenter.org](http://www.worldpolicycenter.org) © 2016 WORLD Policy Analysis Center



Percentage of population using an unimproved drinking water source

**<1**   **1-10**   **11-20**   **>20**   **Insufficient data or not applicable**





# Thanks

For more information, please  
feel free to contact me at

[csharpe@thinkwellgroup.com](mailto:csharpe@thinkwellgroup.com)

Go out. And do amazing  
things.

