



# Supplanting Institutionalized Science of Science Data Collection with Big Data Methods ‘in the Wild’

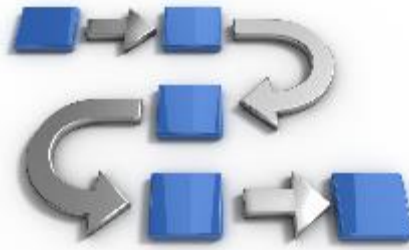
---

**Ioannis Pavlidis**  
**Mohammed Emtiaz Ahmed**  
Computational Physiology Lab  
University of Houston





**Introduction**



**Methods**



**Results**



**Conclusion**



## UNESCO SCIENCE REPORT



### Contents

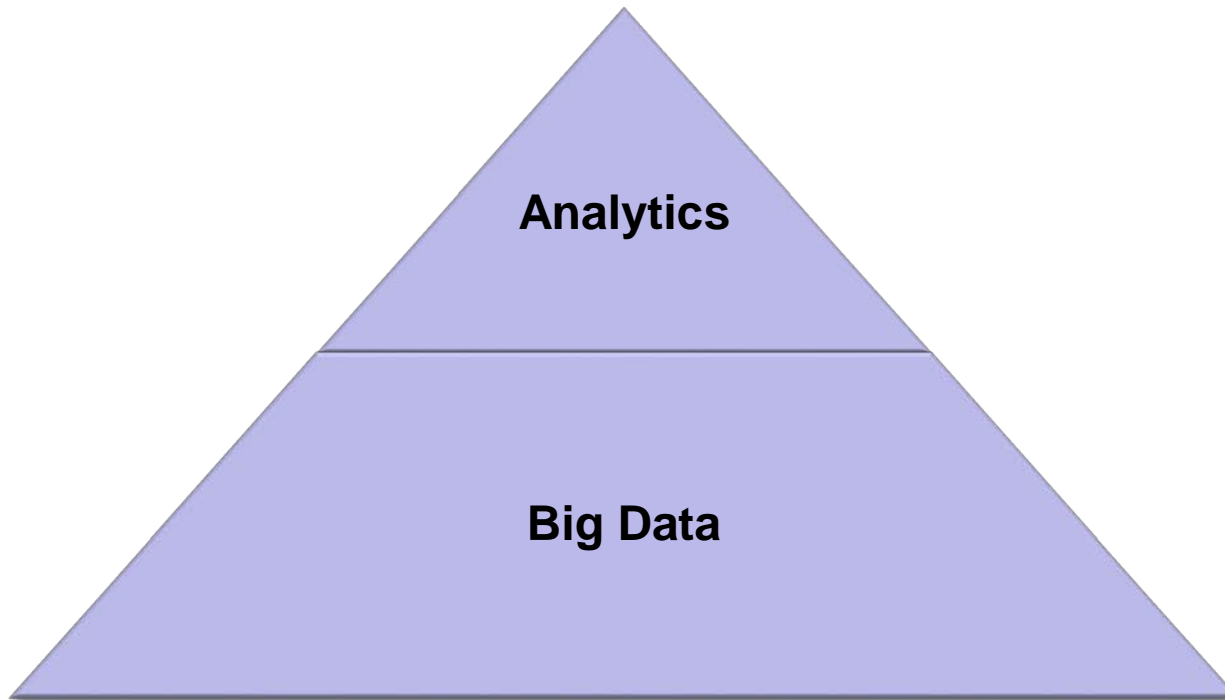
- # of researchers
- Total publications
- Continent based analysis
- Country based analysis

### Requires

- 5 years
- 50+ specialists
- Govt bureaucracy

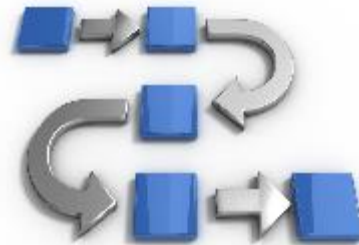


Create an **UNESCO** type report using **data science**





**Introduction**



**Methods**



**Results**



**Conclusion**



## UNESCO

### Data collection

- Governmental bureaucracy

### Analytics

- 50+ statisticians

### Time

- 5 years

## Our Method

### Data collection

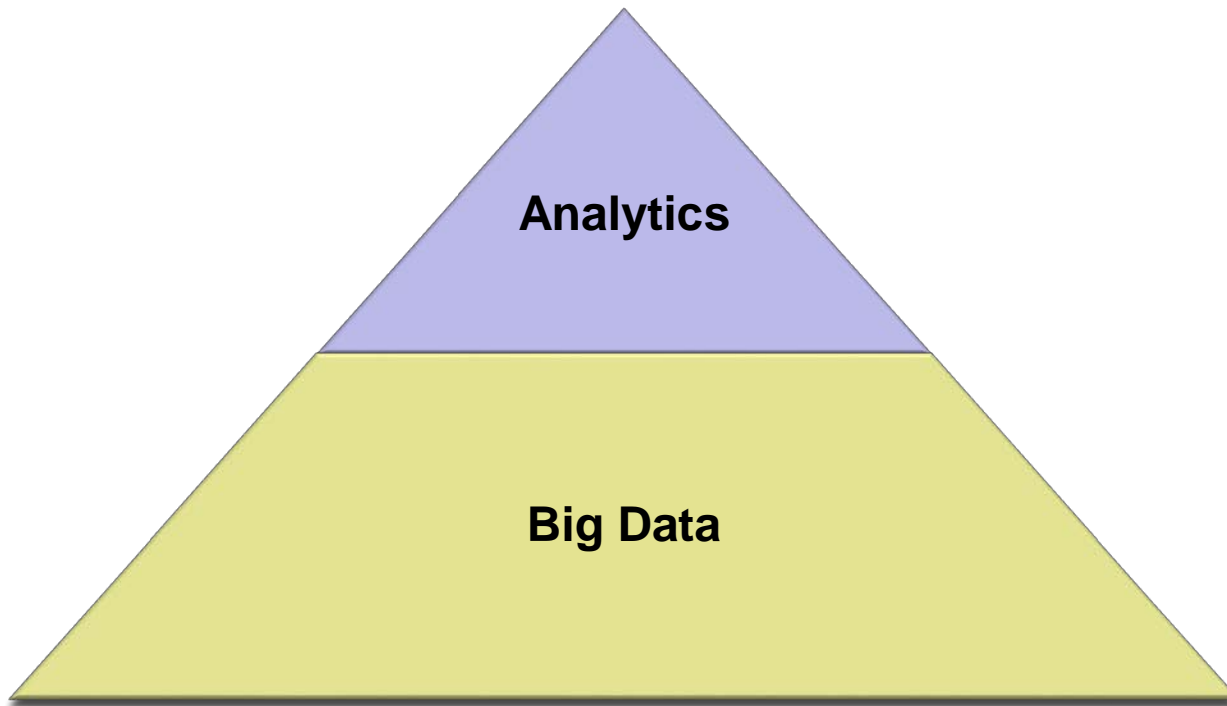
- Scripting

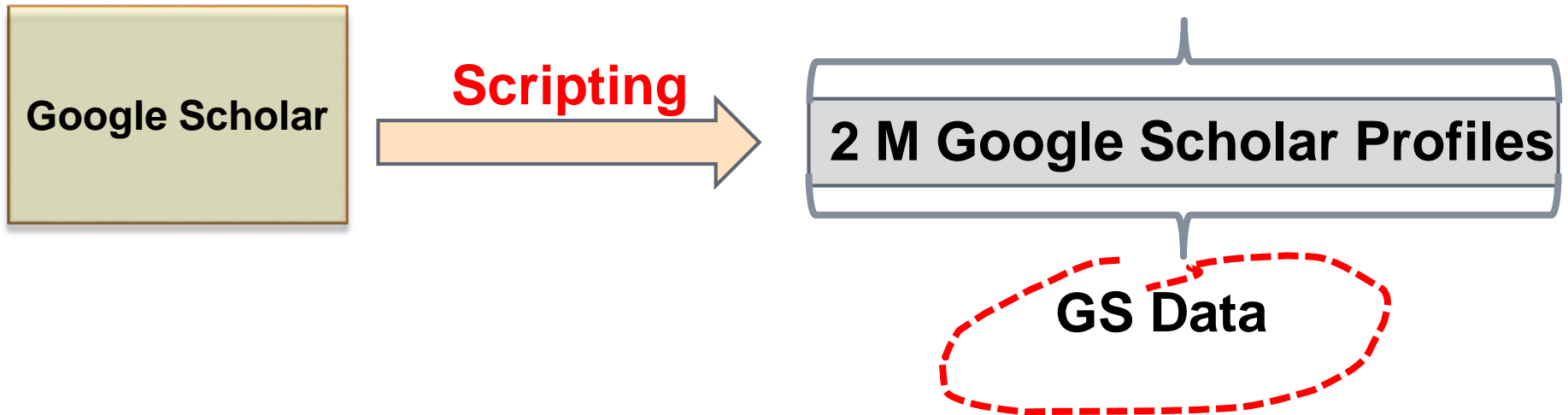
### Analytics

- 1 statistician/1 computer sci.

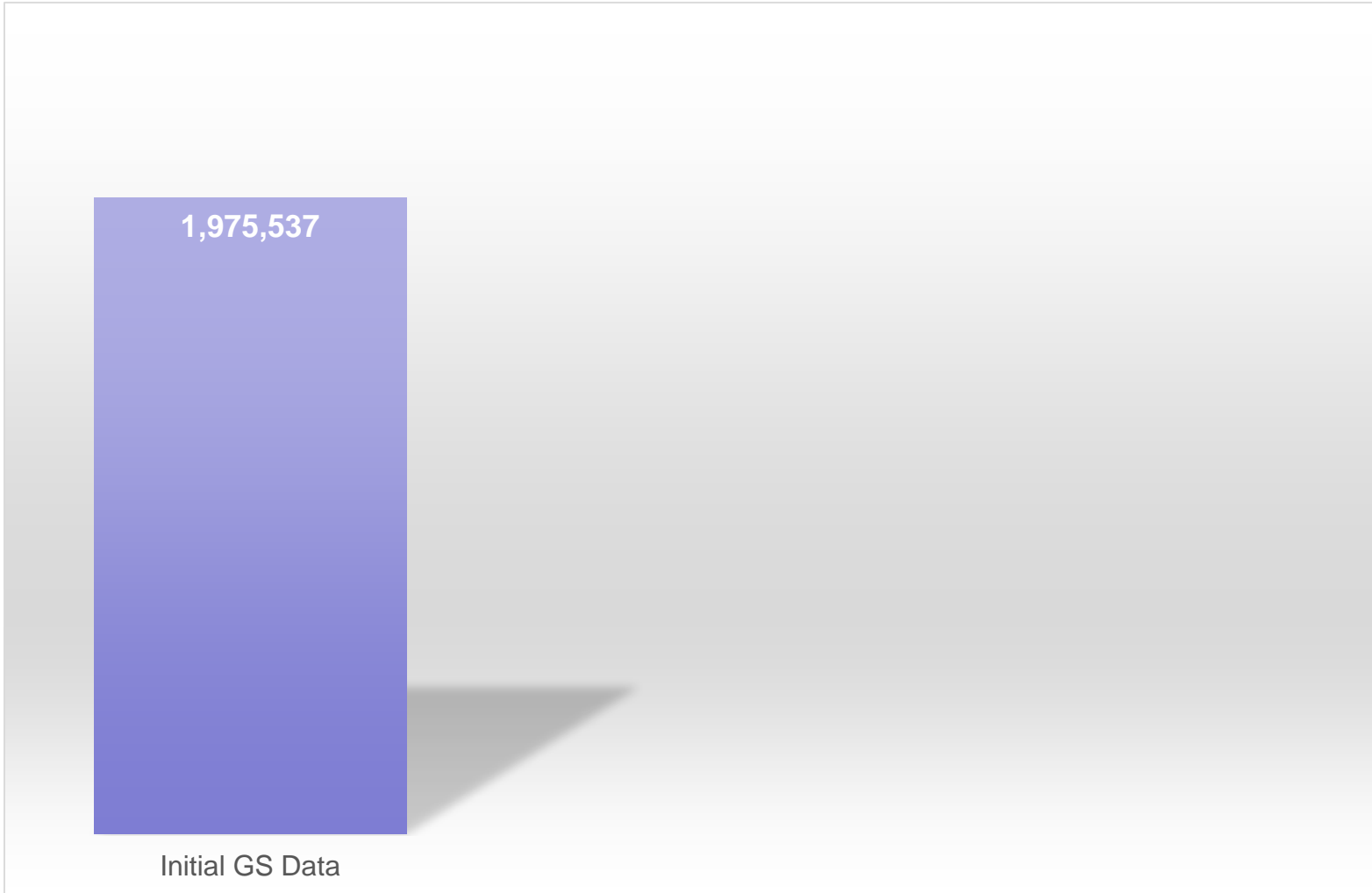
### Time

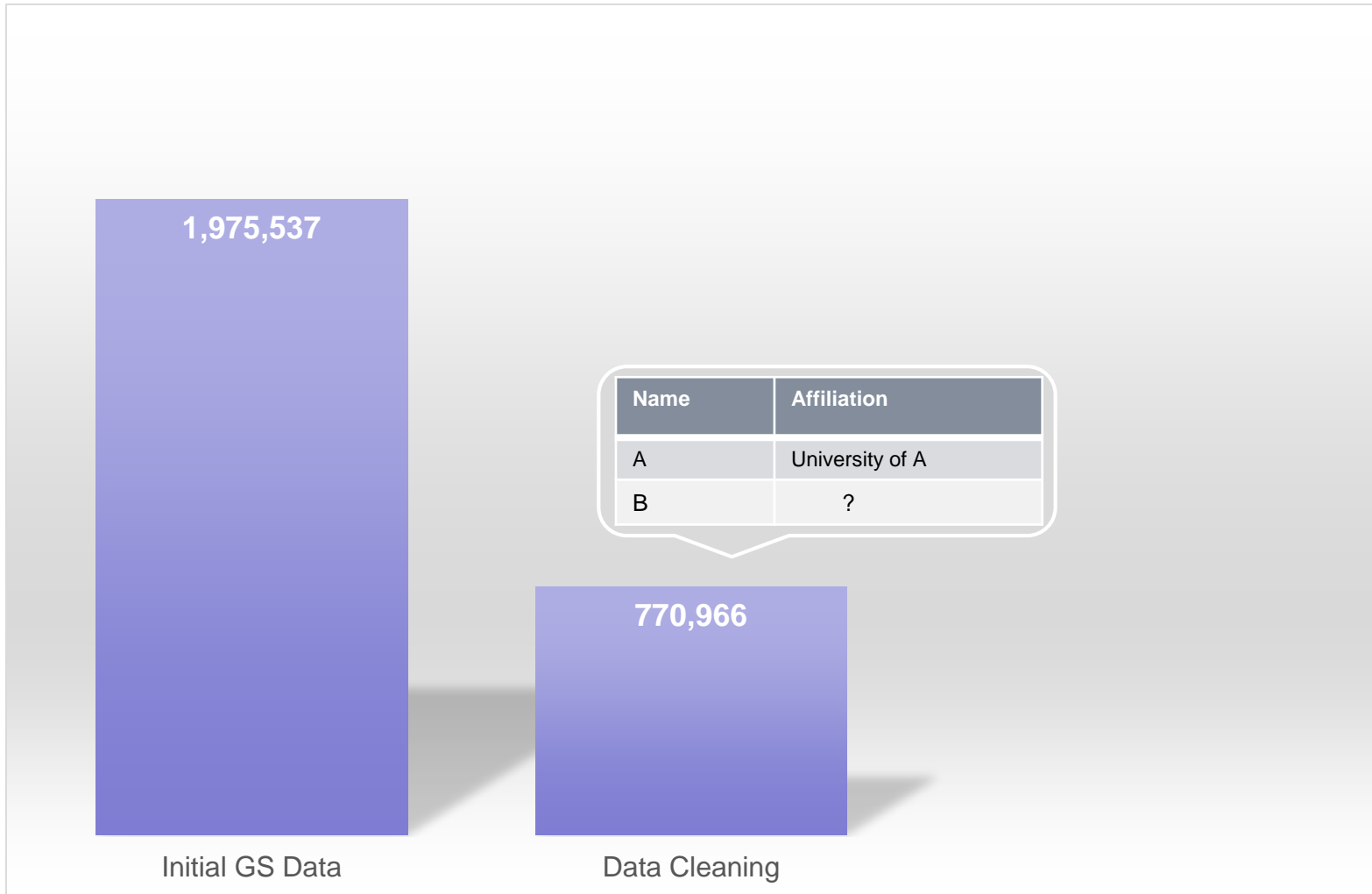
- 5 weeks

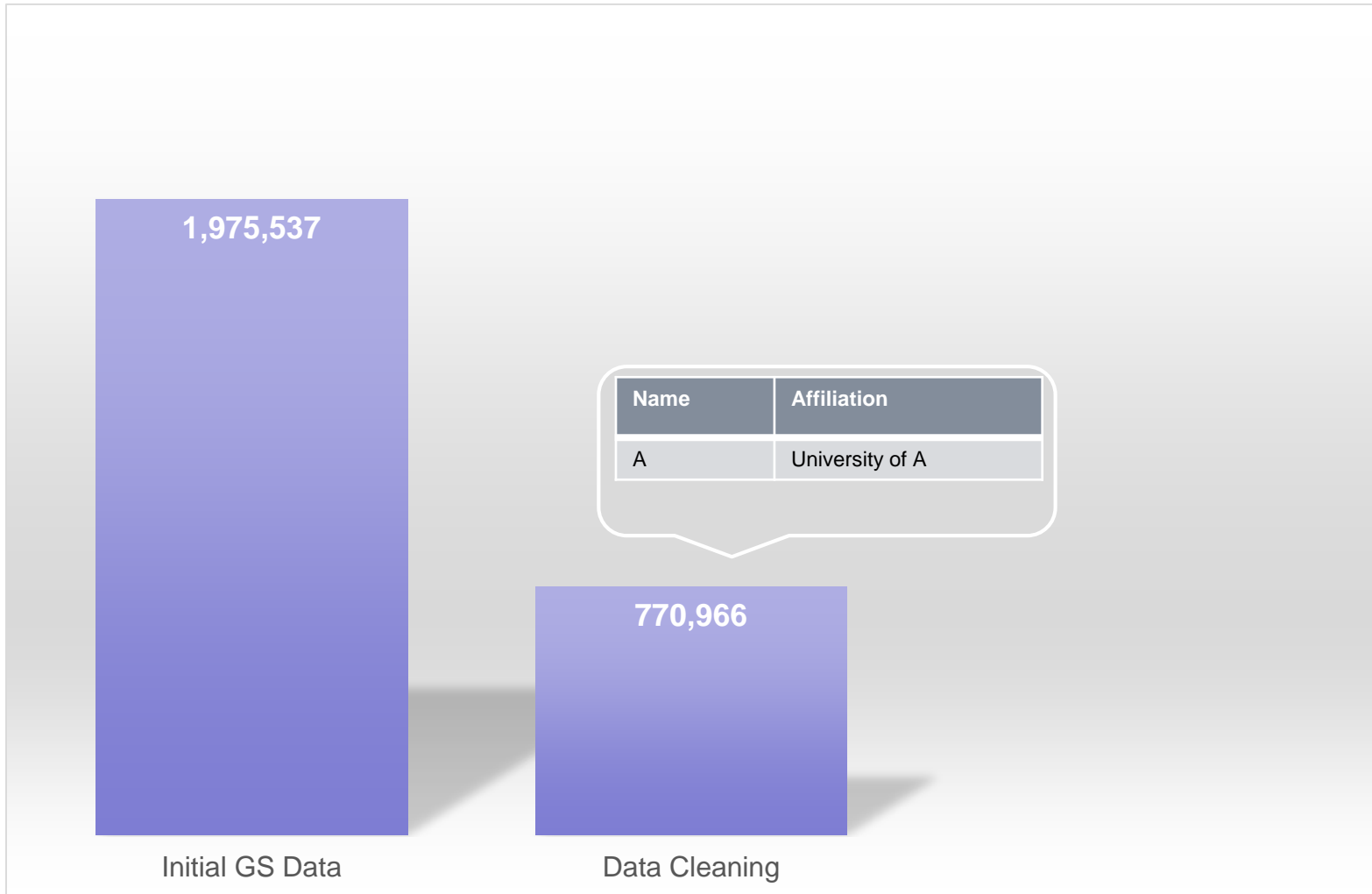


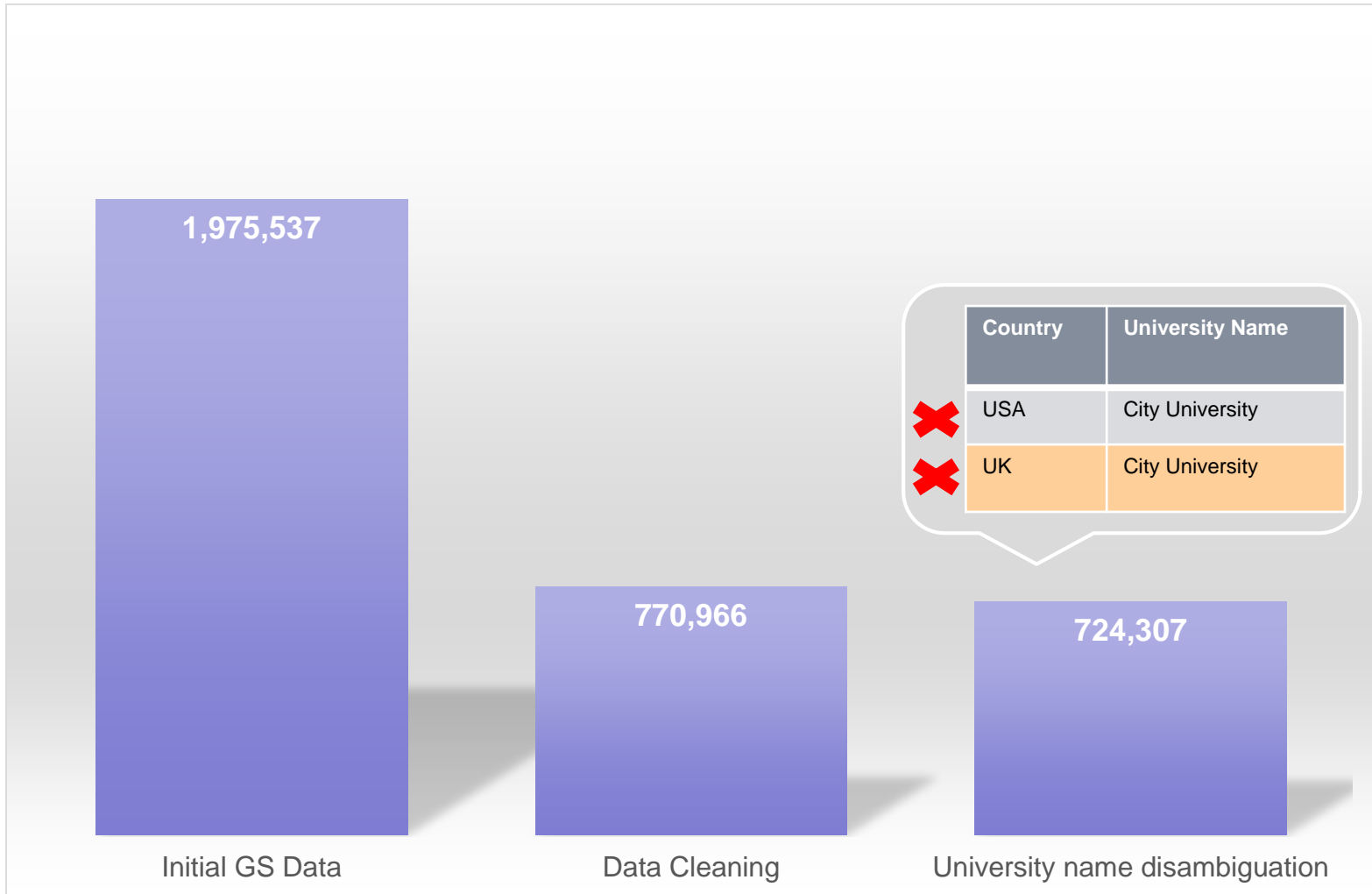








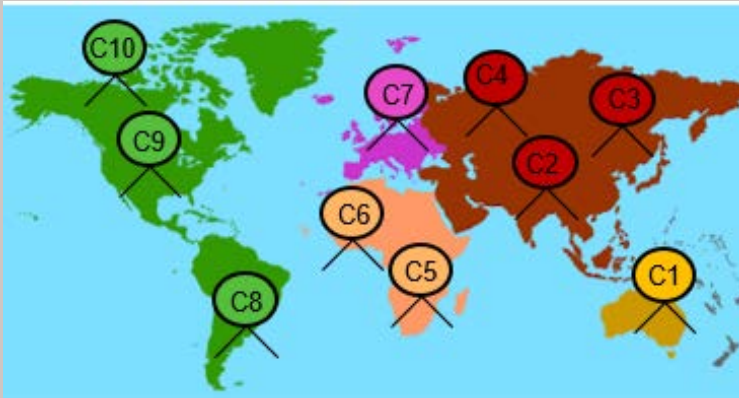







## UNESCO data

Country & Continent

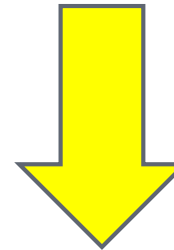
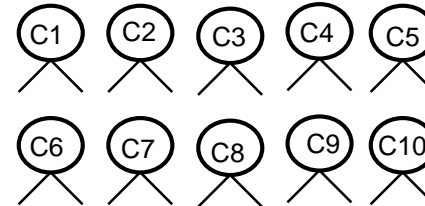
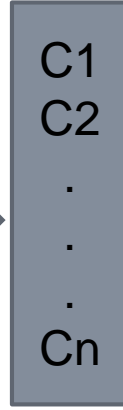
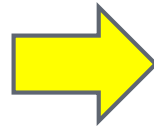
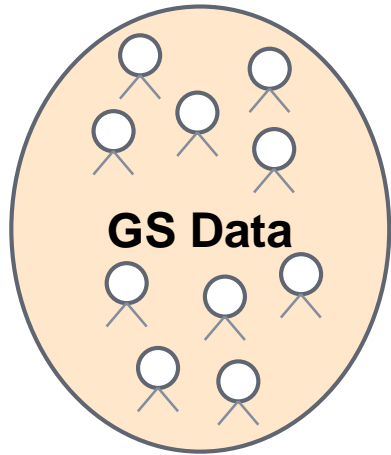


## GS driven data

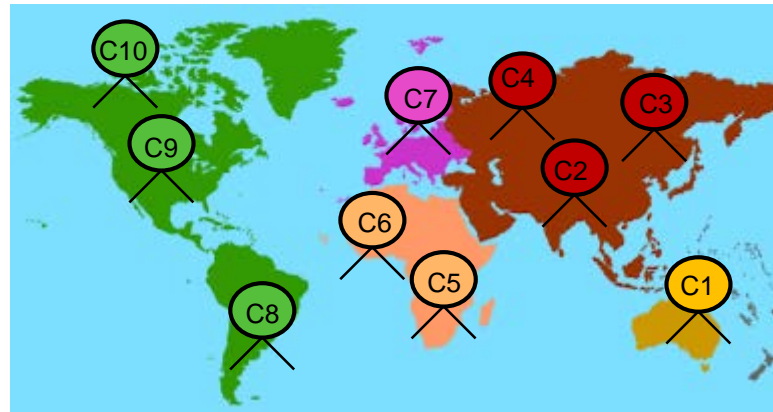
- 
- 
- Full scholar name
  - University affiliation
  - # publications
  - # citations



## Univ. to Country Mapping

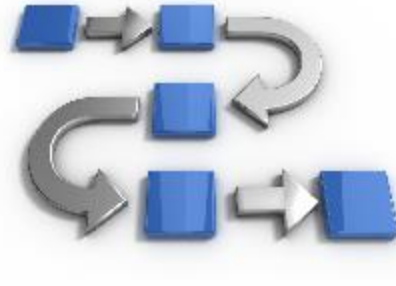


## Country to Continent Map





**Introduction**



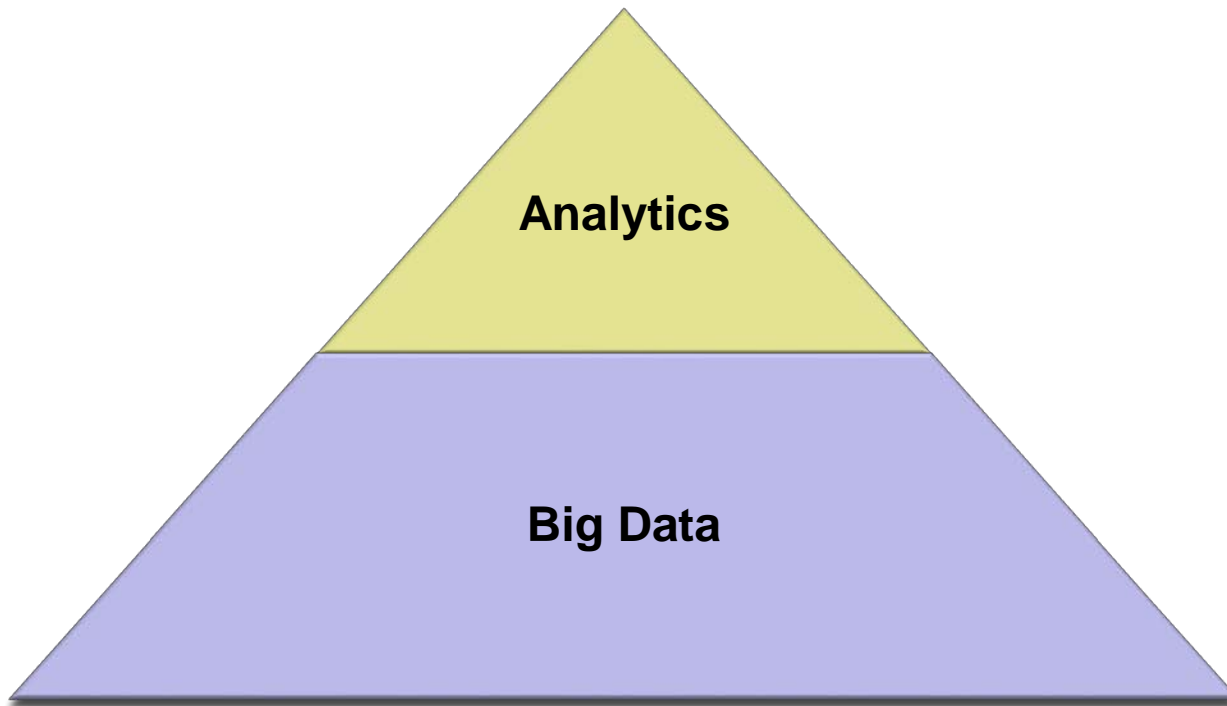
**Methods**



**Results**



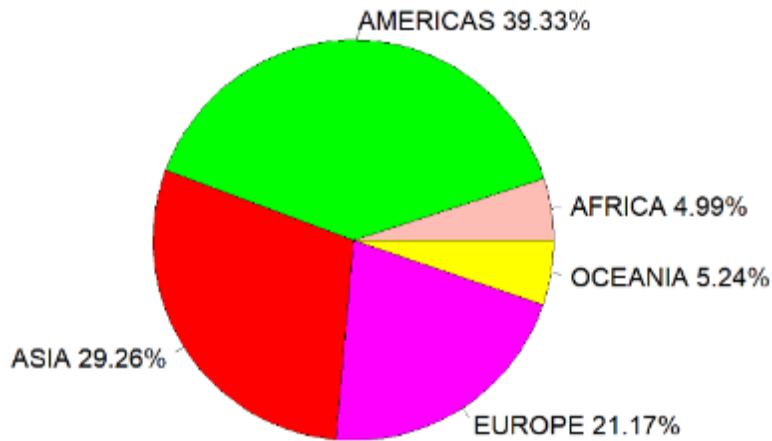
**Conclusion**



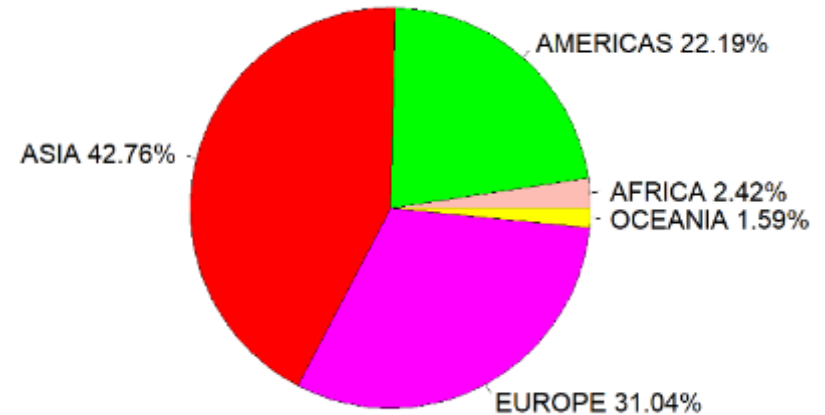


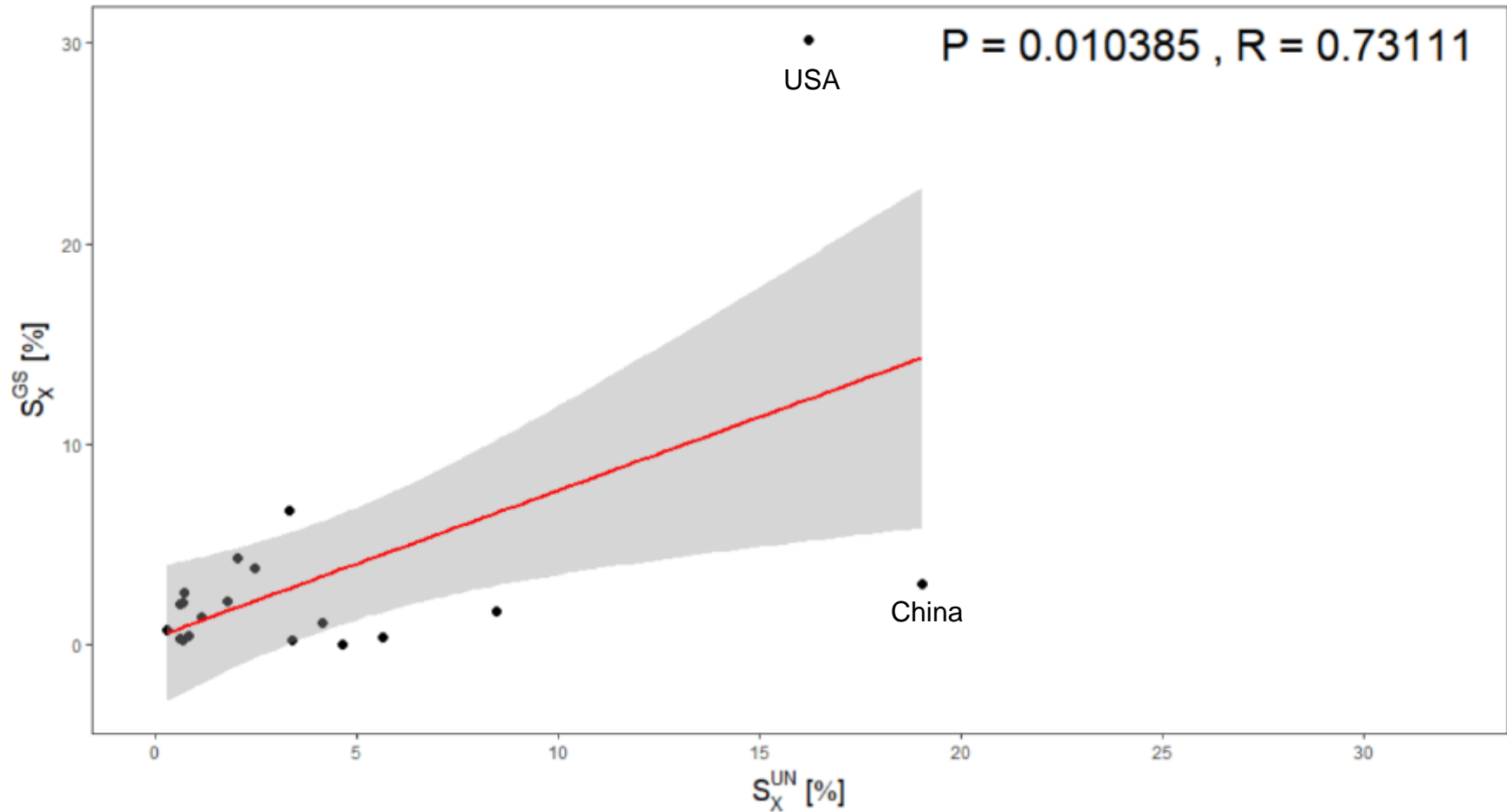


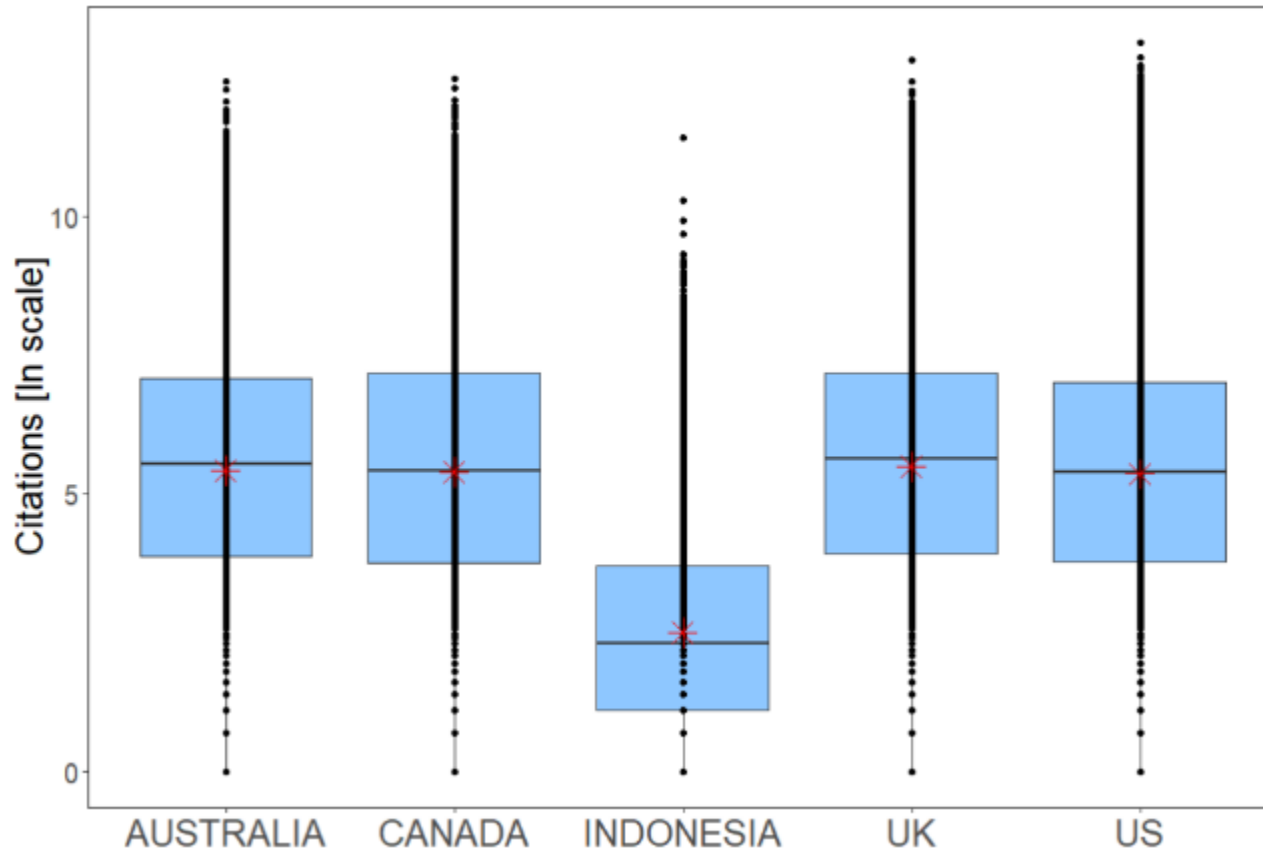
GS data: Scholars > 0 citations



UNESCO data: Scholars









**Introduction**



**Methods**



**Results**



**Conclusion**



## UNESCO Report

- # of researchers
- Continent analysis
- Country analysis
  - select countries
- Production efficiency
  - 5 years
- Manpower
  - 50+ specialists

## GS-driven Report

- # of researchers
- # of citations
- Continent analysis
- Country analysis
  - 207 countries
- Production efficiency
  - 5 weeks
- Manpower
  - 2 researchers



- ❖ GS data is a subset of UNESCO data
  - **Not accurate total numbers**
- ❖ GS data correlate with UNESCO data
  - **Accurate country proportions!**
- ❖ **THINK:** Google Scholar was not even designed to carry out science of science census
  - The potential of replacing conventional science policy processes with big data methods is huge!



**Thank You!**