

# Data Science for Global Studies

Gender Inequality in Asia



# What is Data Science?

Drawing useful conclusions from data using  
computation and statistics

AND

Communicating findings clearly



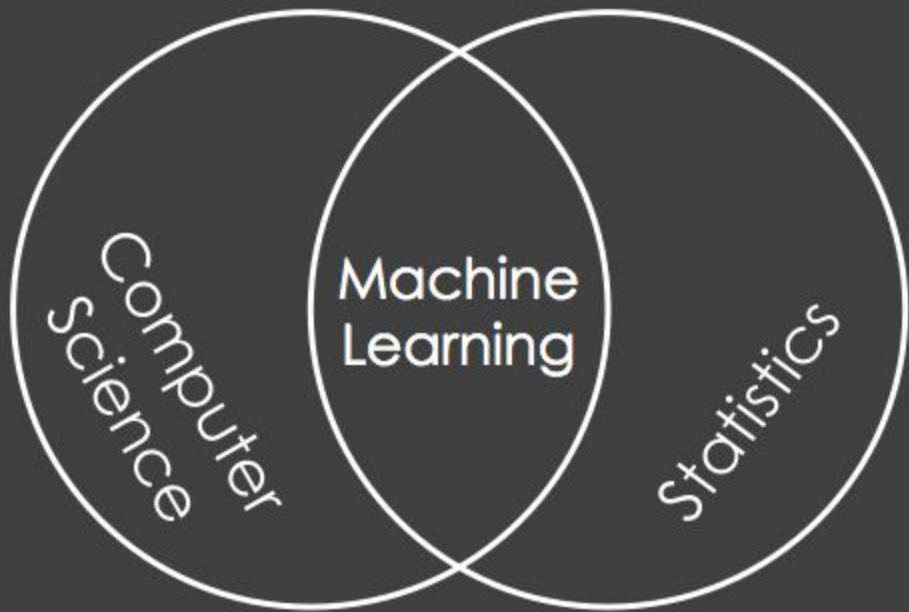
# Skills of Data Science



Computer  
Science

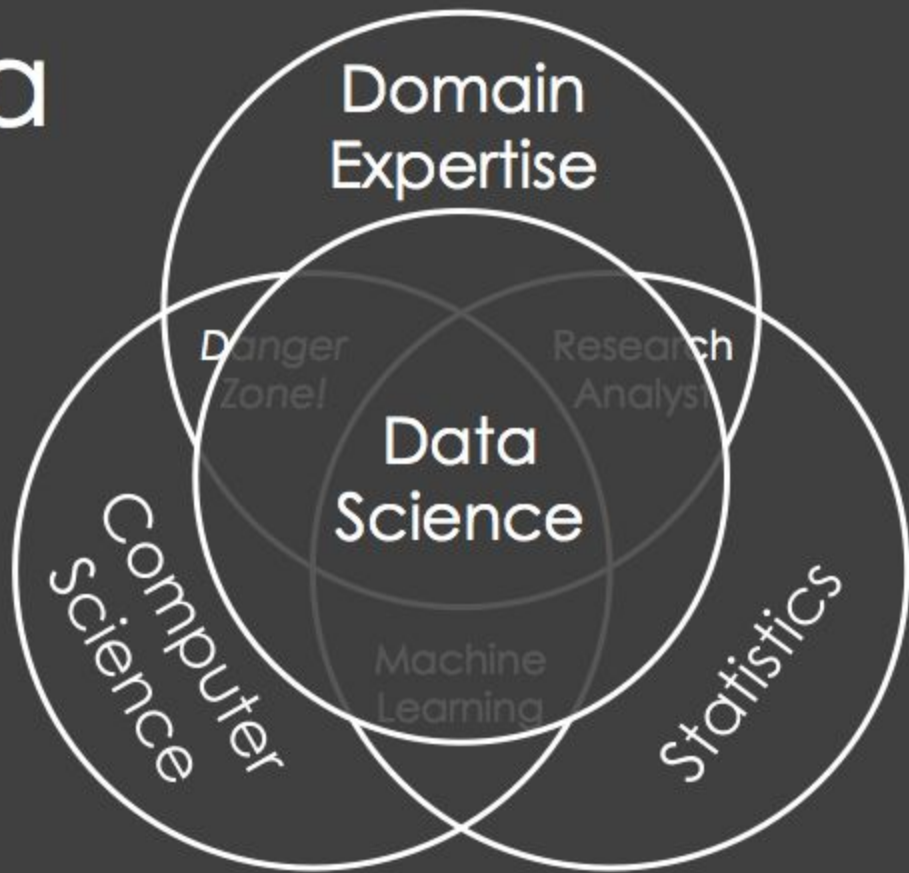
Statistics

# Skills of Data Science



# Skills of Data Science

More ***Union***  
than *Intersection*

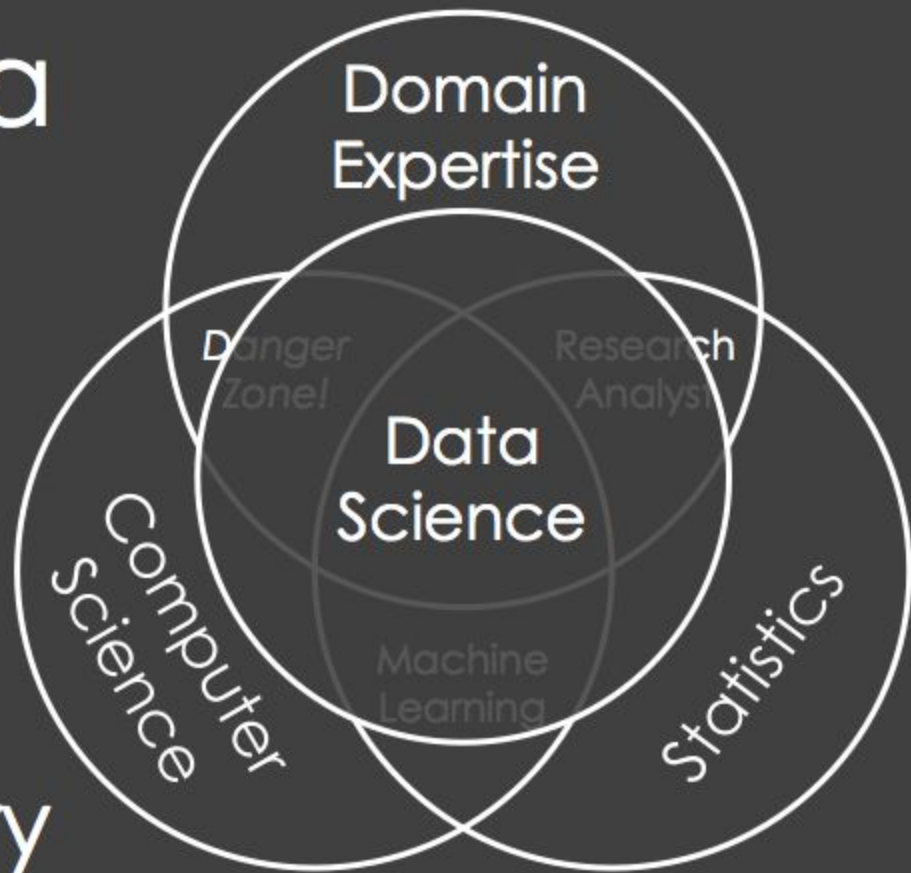


Drew Conway's Venn Diagram of Data Science

# Skills of Data Science

More **Union**  
than *Intersection*

Data Science is  
interdisciplinary



Drew Conway's Venn Diagram of Data Science

# How can it help me?

Build marketable, quantitative skillset

Look at world in a new way

Stop guessing!



# Our Dataset

## United Nations Gender Inequality Index (GII)

- Calculation is complex!!
- <http://hdr.undp.org/en/composite/GII>





# Coding?! Geometric averaging?!

A **function** is a bit of code that does something. They often have inputs and outputs, and in the notebook, you'll see a lot of them that do the work of calculating the GII.

**Geometric mean:** The central number in a geometric progression (e.g. 9 in 3, 9, 27), also calculable as the  $n$ th root of a product of  $n$  numbers

**Harmonic mean:** To find the harmonic mean of  $n$  numbers, add the reciprocals of the numbers in the set, divide the sum by  $n$ , then take the reciprocal of the result.



# *Python* - our programming language

- A great language for beginners! Used in Data 8, CS61A, and all over industry. Popular for data science work, too.
- Variables - can (and should) give everything a descriptive name
- Use # to “comment out” the rest of that line - the computer won’t execute it, and you can write notes to yourself, or “erase” without permanently deleting



# Questions to keep in mind:

- What sorts of data are valuable and/or valid when constructing this sort of ranking system?
- What does a country look like (culturally, demographically, geographically) when it has high/low values for some data points, and how does this affect its ranking?
- If you wanted to change these numbers (in real life), how would you do it? What would you prioritize?



# Sources:

UN Gender Inequality Index info and graph:

<http://hdr.undp.org/en/content/gender-inequality-index-gii>

UN HDR Technical Notes:

[http://dev-hdr.pantheonsite.io/sites/default/files/hdr2016\\_technical\\_notes\\_0.pdf](http://dev-hdr.pantheonsite.io/sites/default/files/hdr2016_technical_notes_0.pdf)

- These notes are what we used to write all those long Python calculation functions



# Where to go from here?

- Data 8 - Spring 2018 and beyond!
- Connector courses
  - Cog Sci, demography, economics, text analysis
- Research + career opportunities

