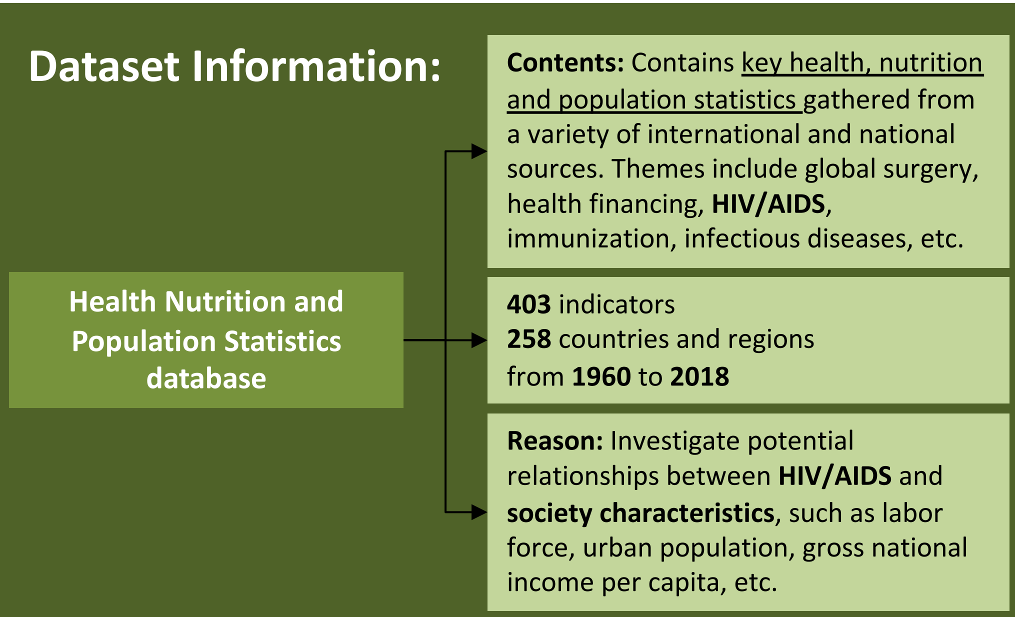
VOLGENAU SCHOOL OF ENGINEERING

GEORGE MASON UNIVERSITY

AIT-580

AIT-580 Spring 2019 Project Summary
The regression analysis between HIV and society characteristics
Long Zhang

Dataset Information:

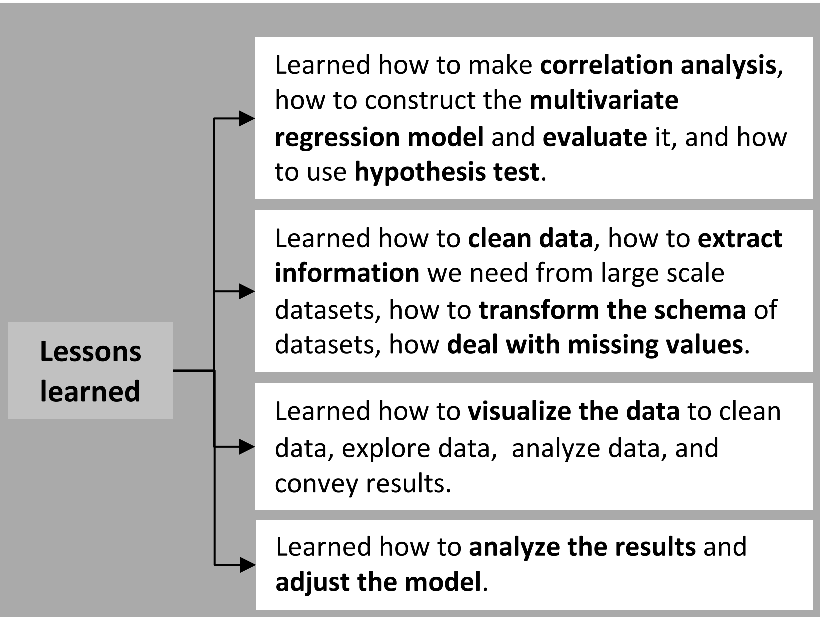


Health Nutrition and Population Statistics database

Contents: Contains key health, nutrition and population statistics gathered from a variety of international and national sources. Themes include global surgery, health financing, **HIV/AIDS**, immunization, infectious diseases, etc.

403 indicators
258 countries and regions
from **1960** to **2018**

Reason: Investigate potential relationships between **HIV/AIDS** and **society characteristics**, such as labor force, urban population, gross national income per capita, etc.



Lessons learned

Learned how to make **correlation analysis**, how to construct the **multivariate regression model** and **evaluate** it, and how to use **hypothesis test**.

Learned how to **clean data**, how to **extract information** we need from large scale datasets, how to **transform the schema** of datasets, how **deal with missing values**.

Learned how to **visualize the data** to clean data, explore data, analyze data, and convey results.

Learned how to **analyze the results** and **adjust the model**.

Findings:

Based on the analyzation between HIV indicator and society indicators, **construct a multivariate regression model** to

predict women's share of population ages 15+ living with HIV **by**

coefficient		Single Correlation
4.4343	Urban population growth (annual %)	0.53
0.8632	Labor force, female (% of total labor force)	0.48
-0.0006	GNI(gross national income) per capita	-0.44

statistically Significant (hypothesis test)

R-squared: 0.950
Adj. R-squared: 0.948

*Based on **the cleaned dataset**, we can analyze the relationships between other HIV indicators and society indicators to extract more information via the same process in the project.*