Data Source: Kaggle

The [World University Rankings dataset](https://www.kaggle.com/datasets/mylesoneill/world-university-rankings) consists of five Excel files were downloaded from Kaggle on World University Rankings. Three of the files contain global university rankings from free sources: Times Higher Education World University Ranking, Academic Ranking of World Universities, and the Center for World University Rankings. Each of these sources has their set of potential biases.

The [Times Higher Education World University Ranking](https://www.timeshighereducation.com/world-university-rankings/latest/world-ranking) was founded in the United Kingdom in 2010. While it is highly regarded, criticism has been levied that it is too commercialized and undermines non-English-speaking institutions.

The [Academic Ranking of World Universities](https://www.shanghairanking.com/rankings) was founded in China in 2003 and is also highly regarded. It has been criticized for focusing on research and devaluing humanities and instruction quality.

The final source is the [Center for World University Rankings](https://cwur.org/), founded in 2012. It comes from Saudia Arabia and is less well known than the other two.

Additionally, there are three supplemental files on educational attainment, national educational expenditures, and national socioeconomic status. The educational attainment data comes from the [World Bank](https://data.worldbank.org/indicator/SE.PRM.CUAT.ZS) and comprises information from the UNESCO Institute for Statistics and the Barro-Lee Dataset. The second file compares expenditures on public and private education nationally and comes from the [National Center for Education Statistics](https://www.aecf.org/). The final dataset was found on [Kaggle](https://www.kaggle.com/datasets/sdorius/globses) for the time series analysis. An individual compiled it from seven different sources. While it is interesting, it may not be as reliable since it is difficult to know if the seven sources used the same criteria.

Data Description

Each files uses different metrics to determine their final score. These metrics will be important when figuring out what values each institution considered important to their analysis. However, they are not comparable across the database.

Times Higher Education

* Consists of 2,603 rows and 14 columns
* Scored based on
  + learning environment
  + International outlook (staff, students, and research)
  + Research (based on volume, income, and reputation)
  + Income (Industry income and knowledge transfer)
* Criticism: Too commercialized and biased against non-English speaking institutions.
* Hypothesis: Institutions from English speaking countries will appear at a higher percentage and have higher scores in the Times Higher Education dataset than in the other two.

Academic Ranking of World Universities

* Consists of 4,897 rows and 11 columns
* Scored based on
  + Alumni winning nobel prizes or field medals
  + Number of staff wining nobel prizes in physics, chemistry, medicine, and economics and field medals in math.
  + Highly cited researchers
  + Papers published in nature and science
  + Papers indexed in Science Citation Index and Social Science Citation Index
* Criticism: Values research over humanities and instruction quality.
* Hypothesis: Institutions that score high on the Times Higher Education learning environment and on the Center for World University Rankings quality of education scores will be ranked lower by the Academic Ranking of World University than the other two datasets.

Center for World University Rankings

* Consists of 2,200 rows and 14 columns
* Scored based on
  + Quality of education
  + Alumni employment
  + Quality of faculty
  + Publications
  + Influence
  + Citations
  + Broad impact
  + Patents
* Criticism: Newer ranking system that may produce unreliable results.
* Hypothesis: The Center for World University Rankings will contain a significant number of institutions not included in the other two datasets.

Correlation heatmap for each source

ARWU

Correlation for world rank

* ARWU\_alumni\_prizes = -.605
* ARWU\_awards = -.688
* ARWU\_highly\_cited = -.769
* ARWU\_published\_papers = -.817
* ARWU\_papers\_indexed = -.587

Correlation for Total Score

* ARWU\_alumni\_prizes = .798
* ARWU\_awards = .837
* ARWU\_highly\_cited = .869
* ARWU\_published\_papers = .928
* ARWU\_papers\_indexed = .611

Other Correlations

* ARWU\_alumni\_prizes and ARWU\_awards = .765
* ARWU\_highly\_cited and ARWU\_published\_papers = .877

CWU

Correlation for world rank

* Quality\_education = .4
* Alumni\_employment = .247
* Quality of faculty = .543
* Publications = .315
* Influence = .460
* Citations=.291
* Broad impact = .307
* Patents = .273

Correlation for total score

* Quality\_education = -.354
* Alumni\_employment = -.226
* Quality of faculty = -.466
* Publications = -.307
* Influence = -.405
* Citations = -.279
* Broad impact = -.368
* Patents = -.207

Other correlations

* Publications and Broad Impact = .818
* Publications and influence = .744
* Publications and citations = .756
* Influence and citations = .728
* Influence and broad impact = .914
* Citations and broad impact = .845

THE

Correlation for world rank

* Teaching = -.887
* International=-.177
* Research = -.879
* Citations = -.440
* Income=-.149

Correlation for total score

* Teaching = .923
* International=.181
* Research = .913
* Citations = .490
* Income= .156

Other correlations

* Teaching and Research = .882