UberMegaMaster Codebook

Variables.
Each variable has (1) a description, (2) a range of values, and (3) commentary on any misleading or missing data in the dataset.

==Year:== (1) This denotes the academic year for the given entries (note that this may differ from the actual year of enrollment):
- For Fall semesters, the academic year is the next calendar year (e.g. a Fall semester in 2005 would be grouped in the 2006 academic year).
- For Spring semesters, the academic year is the same calendar year (so the Spring semester in 2006 would also be grouped in the 2006 academic year).

(2) Year ranges from a minimum of 1970 to a maximum of 2020 in the Ubermega dataset.

(3) Missing data consists of the 1999 fall semester and the 2020 Spring Semester. Also note that a semester in XXXX year does NOT necessarily have an academic year XXXX, as described above.

==Semester:== (1) This denotes the semester (AKA term) of a given entry.

(2) Semester spans the following: Fall, Spring.

(3) Semester may be misleading because students can enroll in multiple semesters each academic year. This means they should usually not be added to prevent double counting.

==Origin:== (1) This denotes the origin state of a student from the United States, and the origin country or territory of a student from outside the United States.

(2) Origin spans hundreds of countries and all states in the US.

(3) Given historical changes in territories and countries, origin demographics have changed tremendously over the past 50 years. To the best of our ability, we have tried to preserve these historical origins in this dataset given their value to research.
 In some situations, it may make sense to map historical origins (e.g. Canal Zone) to modern-day countries (in this case Panama). For countries that no longer exist (e.g. Yugoslavia or the Soviet Union), it may also make sense to map any data to the most populous country that exists in the same area today. With this explicit dataset, however, we have not done anything of the sort.

==Sex:== (1) This denotes the biological (legal) sex of a student, as indicated to the university upon enrollment.

(2) Sex spans Male, Female, Unknown, and All.

(3) "Unknown" does NOT necessarily mean that the student is non-cisgender. "Unknown" indicates that the student entered an unidentifiable record in the "Sex" field of their profile form. For example, the student may have left the field blank, entered only a space, or entered something not identifiable as either male or female.

==Trinity:== (1) This denotes the number of students enrolled in Trinity College in the corresponding academic year and semester, with the corresponding sex and from the corresponding origin. For example, if a given row in Ubermega had 1975 for Year, "Fall" for Semester, "Female" for Sex, "Antigua and Barbuda" for Origin, and 5 for Trinity, this would indicate that in the 1975 fall semester, 5 female students were enrolled in Trinity College hailing from Antigua and Barbuda.

(2) Trinity ranges from 0 to 5305, the high point of which occurred from all students in the United States in the year 2000.

(3) While the vast majority of our dataset reflects accurate Trinity numbers, it should be noted that the Duke Women's College (an undergraduate college exclusively for female students) existed from 1930-1971. The Total and Female categories for Trinity in 1971 and 1970 thus reflect numbers from the Women's College for continuity's sake.

==Nursing:== (1) This denotes the number of students enrolled in the undergraduate school of nursing.

(2) Nursing enrollment ranges from 0 to 221.

(3) The "Nursing" program refers broadly to all students who have pursued Nursing in a given year, a field which has changed rapidly over time at the University. Prior to 1984, nursing was offered as an undergraduate program in addition to a graduate program, although this changed in the mid-80s. Nursing today remains as a graduate program and an accelerated undergraduate program. Additionally, the homes of male nursing students from 1970-1984 (when the program temporarily closed) are not documented, so these students are not represented in the data set for the mapping tool OR data table. These two factors explain the large spans of zero-values in this variable.

==Graduate:== (1) This denotes the number of students enrolled in the graduate school, i.e. any student pursuing a Masters or PhD in a given field.

(2) Graduate enrollment ranges from 0 to 2331.

(3) This does not include professional schools/other specific graduate programs offered at Duke (Divinity, Law, Business, Medicine, etc); to get a grand total of these variables, see All_Grad variable.

==Divinity:== (1) This denotes the number of students enrolled in the Divinity school.

(2) Divinity school enrollment ranges from 0 to 678.

Law: (1) This denotes the number of students enrolled in the law school.

(2) Law school enrollment ranges from 0 to 660.

Business: (1) This denotes the number of students enrolled in the graduate school of business.

(2) Business school enrollment ranges from 0 to 1029.

Engineering: (1) This denotes the number of students enrolled in the undergraduate school of engineering/Pratt.

(2) Undergraduate engineering enrollment ranges from 0 to 1127.

(3) Graduate engineering is NOT represented by this variable — this is solely undergraduate. We have not created a variable for graduate engineering at this time.

Environment: (1) This denotes the number of students enrolled in the school of the Environment, presently Nicholas School of the Environment but previously the school of forestry.

(2) School of the environment enrollment ranges from 0 to 319.

Medicine: (1) This denotes the number of students enrolled in the medical school.

(2) Medical school enrollment ranges from 0 to 627.

All_Schools: (1) This denotes the grand total of undergraduate and graduate students.

(2) Total for all university enrollment ranges from 0 to 13654.

(3) This zero-value indicates that a certain origin (or multiple regions) sent no students of a certain sex to any program in a certain semester. This does *not* indicate that the region *never* sent students, nor does it indicate that there were no students in a program at this time.

Region: (1) Notes the continent/overall region a student is from according to the home state or country listed.

(2) Regions of origin include Africa, Asia, Atlantic Ocean, Australia, Caribbean Sea, Central America, Eurasia, Europe, Indian Ocean, North America, Pacific Ocean, South America, and United States.

(3) States will be included within the United States region (although the U.S. is a part of North America) so that they can be isolated within the domestic map.

All_UG: (1) Notes the total of undergraduate students, which includes Trinity, engineering (Pratt), and Nursing.

(2) Undergraduate enrollment ranges from 0 to 5527.

(3) This zero-value indicates that a certain origin (or multiple regions) sent no students of a certain sex to any undergraduate program in a certain semester. This does *not* indicate that the region *never* sent students, nor does it indicate that there were no students in a program at this time.

All_Grad: (1) Notes the total of graduate and professional students, which includes every non-undergraduate program.

(2) Graduate enrollment ranges from 0 to 8318.

(3) This zero-value indicates that a certain origin (or multiple regions) sent no students of a certain sex to any graduate program in a certain semester. This does *not* indicate that the region *never* sent students, nor does it indicate that there were no students in a program at this time.

Graduate.Nursing: (1) This denotes the number of students enrolled in graduate nursing, which is NOT included in nursing or All_UG.

(2) Enrollment in the graduate nursing program(s) ranges from 0 to 939.

(3) In 1984, the Graduate Nursing program temporarily closed (along with undergraduate), but reopened in 1985. In 2002, the accelerated BSN program replaced the previous undergraduate program, so this is NOT considered a part of graduate, and is represented by the Nursing variable. In 2006, the PhD program opened, and in 2008, we began offering a Doctorate of Nursing. These numbers are often low, so it is important to note that this is NOT BINARY — every variable is a count of students in a given region, sex, year, and semester.

Allied.Health: (1) This denotes the number of students enrolled in the Allied Health program, which is NOT included in the medical school or Trinity totals.

(2) Enrollment in the Allied Health program ranges from 0 to 529.

(3) These numbers are often low, so it is important to note that this is NOT BINARY — every variable is a count of students in a given region, sex, year, and semester.