UC Data Warehouse Student Ethnicity Training Overview

9/12/2017

How is Ethnicity Collected?

- Undergraduates:
  - Universitywide application, 2 separate sections, one for IPEDS/Census and one for UC Ethnicity
  - IPEDS is a 2-part question. 1) Are you Hispanic/Latino? 2) What is your racial background?
  - UC Ethnicity currently has 75 categories. Disaggregated Asian categories were introduced in 2010 and SWANA (Southwest Asian/North African) introduced in 2014.
- Graduates:
  - We recommend the same procedure as undergraduates but campus implementations vary.
- Students can change race/ethnicity identification; it's reported to us on enrollment and degree files
- For more information, go here: https://www.universityofcalifornia.edu/infocenter/disaggregated-data to explore the data and also visit the link at the bottom.

How is Ethnicity Reported?

- International overrides everything else
- For IPEDS, Hispanic/Latino overrides all other categories; Two or More if you are more than one of the other broad categories
- For UC, the "Accountability Report" aggregation applies a hierarchy by which we report
- In addition to the link on the previous slide, reference business rules here: http://data.ucop.edu/support-training/support-resources-files/Business%20Rules%20-%20UC%20and%20IPEDS%20Ethnicities%20v2.0.pdf

Hierarchy

Undergraduates	Graduate Students
1. Black	1. Black
2. Mexican/Mexican American/Chicano	2. Mexican/Mexican American/Chicano
3. Other Spanish American/Latino	3. Other Spanish American/Latino
4. American Indian/Alaskan Native	4. American Indian/Alaskan Native
5. Filipino	**5. Pacific Islander**
6. Vietnamese	6. Vietnamese
7. Chinese	7. Filipino
8. East Indian/Pakistani	8. Japanese
9. Japanese	9. Chinese
10. Korean	10. Korean
11. Pacific Islander	11. East Indian/Pakistani
12. Other Asian	12. Other Asian
13. White/Caucasian	13. White/Caucasian
14. Unknown	14. Unknown

These reporting orders are in UC_ETN_CD_LVL_1_UGRAD_RPTG_ORD_NUM and UC_ETN_CD_LVL_1_GRAD_RPTG_ORD_NUM. There is also a Level 2 reporting order, but we don't use that in practice.

How was/is Ethnicity Stored?

- Enrollment:
  - Hierarchy was applied and we only stored a single race/ethnicity prior to 2010 (with a few rare exceptions)
- Undergraduate Admissions:
  - Hierarchy was applied and we only stored a single/race ethnicity prior to 2016 (because of historical copying)
- Graduate Admissions:
  - We only have “Level 1” and no “Level 2”

UC Level 1 and Level 2

- Level 1: 14 categories
- Level 2: 75 categories
- All Level 2's roll up into Level 1
- Level 1 rolls up into IPEDS for deriving IPEDS when students did not respond to the IPEDS question
- Look at **ETHNIC_CODE_V** for the rollup

The Basic Joins

- Start with ENROLLMENT_HEAD_COUNT_M or DEGREE_HEAD_COUNT_M. These are materialized query tables that have one record per student per (sub)term.
- Don't use the fact (F) tables; these are duplicated for multiple majors/degrees, unless you really know what you are doing
- Join STUDENT_ETHNIC_CODE_D (this is the "Bridge Table" on STUDENT_KEY
- Join the term dimension (important!)

Filtering on Effective Date

- Do not forget this step!
- **WHERE ACAD_(SUB_)T_CYCLE_BEG_DT BETWEEN STUD_ETN_CD_BEG_EFF_DT AND STUD_ETN_CD_END_EFF_DT**
- The normal star process of relying on the surrogate key (STUDENT_KEY) to deal with time dimensions does not work for ethnicity.
- Example: if something changes in STUDENT_D, such as a student becomes a CA resident, they will get a new entry in STUDENT_D, with a corresponding new STUDENT_KEY to go with it.
- But if a student's ethnicity changes from one term to the next, (and nothing else in STUDENT_D changes), the student will get a new entry in STUDENT_ETHNIC_CODE_D with the **SAME STUDENT_KEY**, but the previous record will have an end date corresponding to the last term, and the new record will have a begin date matching the current term. That's why you need to use the filter above, or you will get duplicate counts.

Examining STUDENT_ETHNIC_CODE_D

STUD_KEY	IPEDS_ETN_CD_KEY	UC_ETN_CD_LVL_1_KEY	UC_ETN_CD_LVL_2_KEY	STUD_ETN_CD_PRIM_FL
XXXXX	17	0	0	Y
XXXXX	0	13	0	Y
XXXXX	0	13	21	N

- Each student has **at least** 3 records in STUDENT_ETHNIC_CODE_D. One for IPEDS, one for level 1, and one for level 2.
- This is the simplest case: One Level 1 value, One Level 2 value, and one IPEDS value. The student has two "primary" records. One for IPEDS and one for UC Level 1. There's no primary Level 2.
- The 0's are "placeholder" values. So if you only want IPEDS, you can filter on IPEDS_ETN_CD_KEY<>0, and likewise for level 1 or level 2.
- There are 3 MQT's that have those things already filtered out: STUDENT_IPEDS_ETHNIC_CODE_M, STUDENT_UC_ETHNIC_CODE_LEVEL_1_M, and STUDENT_UC_ETHNIC_CODE_LEVEL_2_M

Getting the Actual Ethnicity Names

- Join UC_ETHNIC_LEVEL_1_D, UC_ETHNIC_CODE_LEVEL_2_D, and/or IPEDS_ETHNIC_CODE_D
- Note that the “V” versions automatically filter out the placeholder values (doing the exact same thing as using the MQT’s STUDENT_IPEDS_ETHNIC_CODE_M, STUDENT_UC_ETHNIC_CODE_LEVEL_1_M, and STUDENT_UC_ETHNIC_CODE_LEVEL_2_M

Multiple Ethnicities and the Primary Flag

- This student has two level 2 values (Japanese, 32 and White, 21) that roll up to two level 1s (Japanese, 8, and White, 13). Since Japanese is higher on the hierarchy list than White, that one gets the “primary” flag.
- This student also only said they were White (17) for IPEDS, so they only have one primary IPEDS.

STUD_KEY	IPEDS_ETN_CD_KEY	UC_ETN_CD_LVL_1_KEY	UC_ETN_CD_LVL_2_KEY	STUD_ETN_CD_PRIM_FL
XXXX	17	0	0	Y
XXXX	0	8	0	Y
XXXX	0	13	0	N
XXXX	0	8	32	N
XXXX	0	13	21	N

Another Example

- This student is similar to the last student (both Japanese and White), but has an IPEDS entry for White (17), Asian (16), and Two or More (14), with Two or More getting the primary flag

STUD_KEY	IPEDS_ETN_CD_KEY	UC_ETN_CD_LVL_1_KEY	UC_ETN_CD_LVL_2_KEY	STUD_ETN_CD_PRIM_FL
XXXX	16	0	0	N
XXXX	14	0	0	Y
XXXX	17	0	0	N
XXXX	0	8	0	Y
XXXX	0	13	0	N
XXXX	0	8	32	N
XXXX	0	13	21	N

International Status for IPEDS vs UC

- IPEDS ethnicity stores the derived US nonresident field as a separate code value
- If you are looking at UC ethnicity, you have to use STUDENT_D.STUD_DMSTC_FGN_CZ_STAT_CD to derive international status

Excerpt from IPEDS_ETHNIC_CODE_D

IPEDS_ETN_CD_KEY	IPEDS_ETN_CD	IPEDS_ETN_CD_NAM
0	N/A	PLACEHOLDER - DO NOT USE
10	001	HISPANIC/LATINO
11	003	AMERICAN INDIAN/ALASKA NATIVE
12	005	NATIVE HAWAIIAN/OTHER PACIFIC ISLANDER
13	007	UNKNOWN
14	009	TWO OR MORE RACES
15	002	AFRICAN AMERICAN/BLACK
16	004	ASIAN
17	006	WHITE
18	**008**	**NON-RESIDENT ALIEN**

More oddities to be aware of

- IPEDS derivation: if we derive "two or more", then the original responses are also in the BI as "non-primary" records, but if we derive "US nonresident", the original responses are not in BI (only in stage)
- Students faced different questions over time
  - For example, Hawaiian/Other Pacific Islander was broken into two categories
- Some slight differences between UG and Grad
  - UG has “Asian Indian” and “Pakistani”, Grad has “Asian Indian” and “East Indian/Pakistani”
  - Grad has an “other unknown”, UG doesn’t
- UC Ethnicity response is used to derive IPEDS when unknown, but not the other way around (though we are working to change this)
- Remember the different reporting hierarchy for Grad and Undergrad
- If a student has both a degree record and an enrollment EOT record for the same term, the race/ethnicity reported in the EOT file overrides the race/ethnicity reported in the degree file.

References

- Sample queries: https://sp.ucop.edu/sites/its/apptech/businessint/dss/ETL/How%20to%20Guides/How%20to%20Use%20Ethnicity%20Tables_20121206.docx
- Business Rules: http://data.ucop.edu/support-training/support-resources-files/Business%20Rules%20-%20UC%20and%20IPEDS%20Ethnicities%20v2.0.pdf
- Disaggregated data: https://www.universityofcalifornia.edu/infocenter/disaggregated-data