Dan O'Brien • Research Director • Boston Area Research Initiative
Radcliffe Institute for Advanced Study • Harvard University • 10 Garden Street, Cambridge, MA 02138
Phone: 617-496-9975 • E-Mail: daniel_obrien@radcliffe.harvard.edu • Web: www.bostonarearesearchinitiative.net

Documentation of Boston Historic Ward Map Database

Overview

This document describes the structure and organization of the Boston Historic Ward Map Database, which contains information on occupation, school attendance, nativity, and race of the Boston population, by ward, for the years 1880, 1900, and 1930. This database includes both datasets describing the wards at each time point and corresponding shape files (for GIS use).

The database was constructed by Cristina Groeger (Dept. of History, Harvard University) as part of a project on the profound changes in the economic, educational, and ethnic composition of Boston between 1880 and 1930. It can be used to visualize, among other things, the great expansion of secondary school enrollment, the decline of youth participation in the work force, the growth of white collar jobs, the decline of unskilled labor, the distribution of the Boston Irish, and the sudden influx and dense concentration of Italian immigrants. This contextual knowledge is useful for historians researching this time period, and useful to non-historians by depicting the origins of fundamental changes whose legacy is still present in Boston today.

The database is generated from the Integrated Public Use Microdata Series provided by the Minnesota Population Center at the University of Minnesota. (Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010.)

The database's contents include:

1. *Census Data 1880_1930.xlsx* contains the original microdata IPUMS Census samples for years 1880, 1900, and 1930, with newly added variables.

2. *WardData_1880.xlsx, WardData_1900.xlsx, and WardData_1930.xlsx*; each contain the relevant variables aggregated to the Ward level for their respective year. Each also has a corresponding shape file (*.shp*).

Table of Contents

1. ORIGINAL CENSUS SAMPLES (*CENSUSORIGINAL1880_1930.CSV)* 3
1.1 DESCRIPTION OF CONTENTS 3
1.2 DESCRIPTION OF VARIABLES 3
1.2.1. IPUMS Variables 3
1.2.2. Newly Calculated Variables 4

2. BOSTON WARDS (WARDDATA_1880, _1900, _1930.XLSX) 5
2.1 DESCRIPTION OF CONTENTS 5
2.2 DESCRIPTION OF VARIABLES 6

APPENDIX 1: CODES FOR CALCULATED VARIABLES 8
A1.1. CLASS 8
A1.2. ATWORK 9
A1.3. IMMB 9
A1.4. IMM 11

1. Original Census Samples (*Census Data 1880_1930.xlsx)*

1.1 Description of Contents

This file contains the original data sample from the Integrated Public Use Microdata Series (IPUMS) project. The Minnesota Population Center has generated samples of census individual-level microdata from the U.S. Federal manuscript censuses back to 1790, and users can freely download any set of samples and variables.

These data are at the individual level, attributing them to households that are nested within wards. This dataset includes the 1880, 1900 and 1930 samples for the city of Boston, which are 10%, 5%, and 5% samples. There are 104,572 individuals in this dataset, 37,210 individuals from 1880, 28,455 individuals from 1900, and 38,905 individuals from 1930.

1.2 Description of Variables

The variables are divided into two groups. The first set includes those present in the publicly-available IPUMS file. The second set includes contains a series of added demographic variables. In addition, the original 1880 sample does not contain numeric data for the *ward* variable. This information was imputed by examining enumeration districts (*enumdist*) in conjunction with digitized manuscript censuses on ancestry.com.

1.2.1. IPUMS Variables

- *year* of census
- *serial* is a unique identifier for the household in that census.
- *City* of household (Boston, MA for all cases in this database)
- *Ward* indicates the ward in which the household is located.
- *enumdist* represents the enumeration districts, several of which made up a ward.
- *Gq* indicates the household type (whether group quarters or not).
- *Pernum* designates the individual's position within the household.
- *Age* of the individual.
- *Sex* of the individual, male or female.

- *Race* of the individual
    - *Note:* String variable of racial category, categorized in 1880 and 1900 as White, Black, or Chinese, and in 1930 as White, Black, American Indian, Chinese, Japanese, or Other Asian.
- *Raced* is a more detailed version of *race* with the above categories subdivided.
- *Bpl* indicates the birthplace of the individual.
- *Bpld* is a more detailed version of bpl.
- *Mbpl* indicates the birthplace of the individual's mother.
- *Mbpld* is a more detailed version of *mbpl*, following the logic of *bpld*.
- *Fbpl* indicates the birthplace of the individual's mother.
- *Fbpld* is a more detailed version of *Fbpl*, following the logic of *bpld*.
- *School* indicates whether the individual was attending school at the time of the census.
- *occ* is a numerical code for one's occupation at the time of the census.
- *occ1950* is a reclassification of *occ* using the 1950 classification scheme, making the variable consistent across time.

1.2.2. Newly Calculated Variables

Each of these variables was calculated based on existing variables in the IPUMS sample. The logic for each calculation is described here, but more detailed code (from STATA) is available in Appendix 1.

- *Class* is a numeric variable that groups individual occupations from *occ1950* into a 11-category classification scheme. See Appendix 1 for the classification logic.
    - *Note*: Possible values include: 0=Professional, 1=High White Collar, 2=Low White Collar, 3=Blue Collar Skilled, 4=Blue Collar Semi Skilled, 5=Personal Service, 6=Laborer, 7=Keeping House, 8=At School, 9= Other not working, 10=Missing, N/A

- *Atwork* describes the primary activity of an individual, using the *school* and *class* variables.
    - *Note*: 0=Recorded as having an occupation, minus those attending school, 1=attending school, 2=keeping house, 3=other not working, 4= missing/NA. The small subset of individuals listed as attending school and working (presumably part time) are designated attending school for the purpose of this database.
- *Nativity* describes the nativity of the individual, based on groupings of the birthplace variable, *bpl*.
    - *Note*: 0=MA, 1=Northeast (including NY, minus MA), 2=U.S. (minus Northeast), 3=Canada, 4=England/Scotland, 5=Ireland, 6=Italy, 7=Germany, 8=Poland, 9=Austrian, 10=Lithuanian, 11=Russian, 12=Scandinavia, 13= Western Europe, 14=Greece, 15=Portugal, 16=Middle East, 17= Asia.
- *Heritage* describes the heritage of an individual based on their mother or their father's nativity (*mbpl* & *fbpl*). The parent whose nativity was numerically higher according to *immB* is the value taken here.

2. Boston Wards (WardData_1880, _1900, _1930.xlsx)

2.1 Description of Contents

These datasets describe the demographics of the wards of Boston for the years 1880, 1900 and 1930. Each is available in both standard spreadsheet and shape file (*.shp* for GIS) form. There were 25 wards in 1880 and 1900, and 22 wards in 1930. All three data sets have the same variables.

The historic ward boundary shape files for the city of Boston were accessed from the Historical Urban Ecological GIS Data Portal (http://hue.uadata.org/). Demographic variables calculated from the individual-level census data were merged to these to create the final files. Because ward boundaries changed in Boston about every 10-15 years, there is a different file for each of the three time points. *Ward numbers are not compatible across years and should not be used to merge data longitudinally*.

2.2 Description of Variables

- *ward_num* is the unique identifier for the ward number.
- *class_* is a prefix for variables describing the percentage of working individuals residing in a ward that were of a specific class (see *class* in Appendix 1), as indicated by the suffix.
    - *Note:* Individuals classified as non-occupational (*class* = {7, 8, 9}) were excluded from this calculation.
    - *Suffixes*: *prof* = Professional; *hwhite* = High White Collar; *lwhite* = Low White Collar; *sblue* = Blue Collar Skilled Workers; *ssblue* = Blue Collar Semi Skilled; *serv* = Personal Service; *unskill* = Unskilled Laborers.
- *Atwork_* is a prefix for variables that describe the percentage of 13-17 year olds residing in a ward who have each primary activity (see *atwork* in Appendix 1), as indicated by the suffix.
    - *Note*: "Missing/NA" were not excluded when aggregated, since this was a common designation for a young child neither working nor at school.
    - *Suffixes*: *employ* = employed; *school* = attending school; *home* = listed as "at home" or "keeping house"; *other* = listed as other non-working response; *miss* = listed as missing (not included in WardData_1880.xlsx, which has no missing cases).
- *Native_* is a prefix for variables that describe the percentage of residents in a ward with each nativity, as indicated by the suffix.
    - *Suffixes: MA* = Born in Massachusetts; *NE* = Born in the Northeast (including NY, minus MA); *US* = Born in the US; *Can* = Born in Canada; *Eng* = Born in England; *Ire* = Born in Ireland; *Ita* = Born in Italy; *Ger* = Born in Germany; *Pol* = Born in Poland; *Aus* = Born in Austria; *Lith* = Born in Lithuania (not included in *WardData_1880.xlsx* or *WardData_1900.xslx*, which has no matching cases); *Rus* = Born in Russia; *Sca* = Born in Scandanavia; *WE* = Born in West Europe; *Gre* = Born in Greece (not included in *WardData_1880.xlsx*, which has no matching cases); *Por* = Born in Portugal; *MidE* = Born in the Middle East; *Asia* = Born in Asia
- *Hrtge_* is a prefix for variables that describe the percentage of residents whose parents have each nativity, as indicated by the suffix (see *imm__* in Appendix 1 for more detail on how the individual-level variables this is derived from was calculated).
    - *Note*: Suffixes the same as for immB.

- *Race_* is a prefix for variables that describe the percentage of residents in a ward ascribed to each race (see *race* in Appendix 1), as indicated by the suffix.
    - *Suffixes*: *White, Black, Chinese, American Indian* (for 1930 only), *Japanese* (for 1930 only), *Other Asian* (for 1930 only)

Appendix 1: Codes for Calculated Variables

A1.1. class

Coded based on the Occ1950 variable. For complete the list of IPUMS Occ1950 codes, see: https://usa.ipums.org/usa-action/variables/OCC1950/#codes_section

```
/*High White Collar- Prof Service*/

gen class = .

replace class=0 if occ1950 >000 & occ1950<100

/*High White Collar--Managers, Off., Proprietors */

replace class=1 if occ1950 >99 & occ1950<291

/*Low White Collar--Clerical and Sales*/

replace class=2 if occ1950 >299 & occ1950<491

/*Blue Collar Skilled-- Craftsmen*/

replace class=3 if occ1950 >499 & occ1950<596

/*Blue Collar Semi Skilled-- Operatives*/

replace class=4 if occ1950 >599 & occ1950<691

/*Personal Service Workers*/

replace class=5 if occ1950 >699 & occ1950<791

/*Laborers*/

replace class=6 if occ1950 >809 & occ1950<971

/*Keeping House*/
```

```
replace class=7 if occ1950==980

replace class=7 if occ1950==981

/*At School*/

replace class=8 if occ1950==983

/*Other Non Working*/

replace class=9 if occ1950 >983 & occ1950<997

replace class=9 if occ1950==982

/*Missing, N/A*/

replace class=10 if occ1950 >=997
```

A1.2. atwork

Code based on *school* and *class* variables.

```
replace atwork = .

replace atwork=0 if class >=0 & class <=6 & school !=2

replace atwork=1 if school==2 | class==8

replace atwork=2 if class ==7 & school != 2

replace atwork=3 if class==9 & school != 2

replace atwork=4 if class==10 & school != 2
```

A1.3. Nativity

Nativity is based on groupings of the birthplace variable: *bpl*. For complete list of IPUMS *bpl* codes, see: https://usa.ipums.org/usa-action/variables/BPL/#codes_section.

0=MA: 025

1=Northeast (including NY, minus MA): 025, 023, 033, 044, 050, 36

2=U.S. (minus Northeast): 001-099

3=Canada:150

4=England/Scotland: 410, 411

5=Ireland: 414

6=Italy:434

7=Germany: 453

8=Poland: 455

9=Austrian: 450

10=Lithuanian: 462

11=Russian: 465

12=Scandanavia: 400-405

13= Western Europe: 420-426

14=Greece=433

15=Portugal: 436

16=Middle East: 530-547

17= Asia: 500-524

gen Nativity = .

replace Nativity = 0 if bpl ==025

replace Nativity = 1 if bpl ==023|bpl ==033|bpl ==044|bpl ==050|bpl ==36

replace Nativity = 2 if bpl >=001 & bpl <=099 & bpl !=025 & bpl !=023 & bpl !=033 & bpl !=044 & bpl !=050 & bpl !=36

replace Nativity = 3 if bpl==150

replace Nativity = 4 if bpl ==410| bpl==411

replace Nativity =5 if bpl ==414

replace Nativity =6 if bpl ==434

replace Nativity =7 if bpl ==453

replace Nativity=8 if bpl ==455

replace Nativity=9 if bpl ==450

replace Nativity=10 if bpl ==462

replace Nativity=11 if bpl ==465

replace Nativity=12 if bpl >=400 & bpl <=405

replace Nativity=13 if bpl >=420 & bpl <=426

replace Nativity=14 if bpl ==433

replace Nativity=15 if bpl ==436

replace Nativity=16 if bpl >=530 & bpl <=547

replace Nativity=17 if bpl >=500 & bpl <=524

A1.4. Heritage

Heritage is based on groupings of the mother's birthplace an father's birthplace variables: *mbpl & fbpl*. The codes for *mbpl* and *fbpl* are the same as *bpl*, listed above.

```
gen Heritage = .

replace Heritage = 0 if fbpl ==025 | mbpl ==025

replace Heritage = 1 if fbpl ==023|fbpl ==033|fbpl ==044|fbpl ==050|fbpl ==36

replace Heritage = 2 if fbpl >=001 & fbpl <=099 & fbpl !=025 & fbpl !=023 & fbpl
!=033 & fbpl !=044 & fbpl !=050 & fbpl !=36

replace Heritage = 3 if fbpl==150 |mbpl ==150

replace Heritage = 4 if fbpl ==410| fbpl==411 |mbpl ==410 |mbpl==411

replace Heritage =5 if fbpl ==414 |mbpl ==414

replace Heritage =6 if fbpl ==434|mbpl ==434

replace Heritage =7 if fbpl ==453| mbpl ==453

replace Heritage =8 if fbpl ==455| mbpl ==455

replace Heritage =9 if fbpl ==450| mbpl ==450

replace Heritage =10 if fbpl ==462| mbpl ==462

replace Heritage =11 if fbpl ==465| mbpl ==465

replace Heritage =12 if fbpl >= 400 & fbpl <=405 | mbpl >=400 & mbpl <=405

replace Heritage =13 if fbpl >= 420 & fbpl <=426 | mbpl >=420 & mbpl <=426

replace Heritage =14 if fbpl==433| mbpl ==433

replace Heritage =15 if fbpl==436| mbpl ==436

replace Heritage =16 if fbpl >=530  & fbpl <=547 | mbpl >=530 & mbpl <=547
```

replace Heritage =17 if fbpl >=500 & fbpl <=524 | mbpl >=500 & mbpl <=524