



IPUMS Training and Development: Acquire and Use NHGIS Data



IPUMS NHGIS Exercise 2

OBJECTIVE: Gain an understanding of how NHGIS datasets are organized and how they can be used to explore your research interests. This exercise will use an NHGIS dataset to explore **changes in the number of college graduates living in Minnesota cities.**

IPUMS NHGIS – Training and Development

Research Question

Which cities in Minnesota saw the greatest change in the number of college-educated residents since 1990?

Objectives

- Find and download NHGIS time series data
- Analyze the data using Microsoft Excel
- Validate work using answer key

Step 1 Log in to NHGIS

- Go to <http://www.nhgis.org> and click on LOGIN in the top right.
- If you have already registered on any IPUMS website...
 - If you remember your password, log in now. Otherwise, click the "Forgot your password?" link on the right and follow the instructions.
- If you have not already registered...
 - Click on the "Create an account" link on the right, fill in the required information, and submit your registration.
- You will then enter the NHGIS Data Finder...

...

Step 2 Investigate the Scope of Relevant Data

A common first step is to look into the range of data available on the topic of interest...

- Click the **TOPICS** filter button, then select 'Educational Attainment', and submit the selection.

1) How many source tables are available for this topic? ____

2) From what year is the oldest table that gives *population counts* by educational attainment?

Step 3

- With the topic already selected, click the **YEARS** filter button,

Find Data for the Period of Interest

then select '1990', and submit the selection.

The SELECT DATA grid now lists all the tables related to the topic of 'Educational Attainment' with data from 1990. One way to proceed would be to select one of the 'source tables' and then look for another more recent table to compare with it. However, the categories, terms, and universes used by census tables often change over time, which can make it difficult to pull together comparable data.

For many topics (including this one, conveniently!), NHGIS provides a simpler alternative, "time series tables," which link together comparable data from multiple years in one table.

- Click on the TIME SERIES TABLES tab (just to the right of the SOURCE TABLES tab) at the top of the SELECT DATA grid.
- Locate the following time series table and answer the questions that follow:

Persons 25 Years and Over by Educational Attainment [7]

• • •

Step 4 Learn About the Table in the Data Finder

- Click the table name to see additional information.
- 3) How many time series does this table contain? _____
 - 4) Which 3 source tables are used to create this time series table?

 - 5) Can you think of an advantage to using this table rather than the table of "Persons 18 Years and Over by Educational Attainment [7]"?

 - 6) What type of "geographic integration" does this table use?

- Back in the SELECT DATA grid, click on "Nominal" in the GEOGRAPHIC INTEGRATION column.
- 7) With this type of integration, what should we keep in mind as we compare data across time?

• • •

Step 5 Create a Data Extract

To download and view NHGIS data, we first have to select a set of data to “extract” from the NHGIS database...

- Click the plus sign to the left of the table name to add it to your Data Cart.
- Click the green CONTINUE button in your Data Cart (at top right of page).
- On the DATA OPTIONS page, select the “Place” geographic level.
 - In U.S. Census terminology, cities, villages, and town centers are all “places.”
- Click the green CONTINUE button in your Data Cart.
- On the REVIEW AND SUBMIT page, check the box to “Include additional descriptive header row (best for spreadsheets)”.
- Add an extract description if you wish. (NHGIS saves your extract request details, so you can come back anytime to revise or resubmit a request. Adding descriptions can help you identify your older requests.)
- Click SUBMIT.

• • •

Step 6 Download the Data Extract

From the EXTRACT HISTORY page, you will be able to download your data extract once it has finished processing, typically within a minute or two. For longer requests, you may leave this page and return once you have received the email alerting you that the extract is ready.

- When the extract status is “complete” (as listed in the “STATUS” column of the EXTRACTS HISTORY table), click on the “tables” link for this extract. This will download a zip file to your computer.
- Open the downloaded file...
 - Depending on which browser you are using, you may see a pop-up window asking if you would like to open or save the file. If so, choose to open the file. Otherwise, you may see a link to the file at the bottom of your browser. If so, click on it to open the file.

Within the zip archive, you should see a folder named nhgis####_csv (where “####” is the number of this extract).

- Open the folder and confirm that it contains two files: a comma

• • •

Step 7

Analyze the Table in Microsoft Excel

separated values (.csv) data file, which typically has a Microsoft Excel icon, and a text (.txt) "codebook" file.

- Double-click on the "nhgis####_ts_nominal_place" data file to open it in Excel.
- 8)** How many places are included in this table? _____
- 9)** Why do you think some places have missing values for some years?

- In the original worksheet, find and copy all of the Minnesota records along with the 2nd row, containing descriptions.
 - Tip: Using Excel's Filter tool on the STATE column is a quick way to isolate the Minnesota records.
- Create a new worksheet in the Excel document, and paste the copied records into the new sheet.
- 10)** How many place records are there for Minnesota? _____
- To see all the field descriptions, select the top row and "Wrap Text".

Aiming to compare counts of college graduates from 1990 and 2008-2012, it will be helpful first to highlight the columns of interest.

11) Defining "college graduates" as anyone with a bachelor's degree or higher, which columns should we highlight? *Note: The 2008-2012 data include both estimates and margins of error columns. For now, we're only interested in the estimates.*

- Change the font color for these columns to highlight them.
- Create 2 new fields called "CollegeGrad90" and "CollegeGrado812", and fill them with sums of the appropriate counts to create totals for all places.
- 12)** How many college graduates were living in White Bear Lake in 1990? _____
- Create a new field called "ChangeCollegeGrad", and compute in it the change (difference) in college graduates between 1990 and 2008-2012 for all places.
- 13)** Which city had the largest increase? How much was it?

We would expect that cities with great increases also had high overall

population growth and vice versa. Continue working through the next set of questions if you'd like to find out which cities had the greatest increases in the proportion of the population with bachelor's degrees.

Optional:

- Create 2 new fields called Totalgo and Totalo812, and sum the appropriate counts to get the total of all persons 25 years and over for 1990 and for 2008-2012.

14) What was the total population of age 25+ in St. Paul in 2008-2012?

- Create 2 more new fields called %Collegego and %Collegeo812. Multiply 100 times each CollegeGrad variable divided by each Total variable to calculate the percentage of the 25+ population with college degrees.

15) Which city had the highest percentage of college grads in 2008-2012?

- Create a final variable called Change%College and calculate the differences between the %College values between 1990 and 2008-2012.

16) Which city had the largest increase in its proportion of college graduates?

• • •

Complete!

ANSWERS

You have finished Exercise 2.

You can check your answers on the following pages.

1) How many source tables are available for this topic? **953**

2) From what year is the oldest table that gives population counts by educational attainment?

1934 (The 1880 table that appears for this topic has a universe of “schools” and therefore does not provide “population counts” by educational attainment.)

3) How many time series does this table contain? 7

4) Which 3 source tables are used to create this time series table?

NP57 from 1990 STF3, NP037C from 2000 SF 3a and B15002 from 2012 ACS 5-Year

5) Can you think of an advantage to using this table rather than the table of "Persons 18 Years and Over by Educational Attainment [7]"?

A large portion of people aged 18-24 are still working to complete a degree. The age 25+ table captures the population after most have completed their formal education.

6) What type of "geographic integration" does this table use?

Nominal

7) With this type of integration, what should we keep in mind as we compare data across time?

This table won't tell us how much of a city's population changes were due to boundary changes, such as through annexation. Also, a city that changed its name or merged with another (e.g., Norwood Young America, MN, in 1997) will be missing values for some years.

8) How many places are included in this table? **30,544 (= the number of data rows, not counting the two header rows)**

9) Why do you think some places are missing values for certain years?

Possibilities: They didn't exist yet or ceased to exist at some point. They were unincorporated places that the Census did not identify in some years. The city changed its name or merged with another.

10) How many place records are there for Minnesota? **916**

11) Defining “college graduates” as anyone with a bachelor’s degree or higher, which columns should we highlight?

AG, AI, AK, & AM: “Bachelor’s degree” for both years and the “Graduate or professional degree” for both years

12) How many college graduates were living in White Bear Lake in 1990? **4,445**

13) Which city had the largest increase? How much was it?

Minneapolis: +40,568

14) What was the total population of age 25+ in St. Paul in 2008-2012? **174,459**

15) Which city had the highest percentage of college grads in 2008-2012? **Woodland: 79.8%**

16) Which city had the largest increase in its proportion of college graduates? **Carver: +43.7**