

NHGIS Extraction & Analysis

Exercise 2

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

Research Question

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

Objectives

- Create and download an extract of NHGIS time series data
- Unzip data file and open in Microsoft Excel
- Analyze the data using Microsoft Excel
- Validate data analysis work using answer key

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 Minnesota Population Center 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...

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? _____

Find Data for the Period of Interest

- With the topic already selected, click the **Years** filter button, 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' listed here 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 Table tab, located just 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]

Learn About the Table in the Data Finder

- 3) Click the table name to see additional information. How many time series does this table contain? _____
- 4) Which 3 source tables are used to create this 1 time series table? _____

- 5) What advantage is there in using this table rather than the 'Persons 18 Years and Over by Educational Attainment [7]'? _____

- 6) What type of 'geographic integration' does this table use? _____
- 7) In the Select Data grid, click on 'Nominal' in the Geographic Integration column. With this type of integration, what should we keep in mind as we compare data across time? _____

Create a Data Extract

Creating a data extract requires the user to select the table(s), specify a geographic level, and select the data layout structure...

- Click the plus sign to the left of the table name to add it to your Data Cart.
- Click the green Continue button under your Data Cart.
- On the Data Options screen, select the 'Place' geographic level.
 - (In census terminology, cities, villages, and town centers are all 'places'.)
- Click the green Continue button under your Data Cart.
- On the Review and Submit screen, check the box to 'Include additional descriptive header row (best for spreadsheets)', add an extract description if you wish, and click Submit.

Download the Data Extract

From the Extracts History page, you will be able to download your data extract once it has finished processing, typically within a few minutes. You may leave this page and return once you have received the email alerting you to your finished extract.

If you refresh your browser window (click on the loop icon at top, or press F5), you will see the extract status change from 'queued' to 'in progress' to 'complete', at which time you will be able to click the 'tables' link to download the data.

- Return to the Extracts History page if not currently there.
- Right-click on the 'tables' link for the extract you created.
- Choose 'Save Target As...' (or 'Save link as...').
- Save the zip file into 'Documents'.
- Find the 'Documents' folder (through the Start menu).
- Right-click on the 'nhgis####_csv.zip' file (where '####' is the number of this extract), and select Extract All... Then click the Extract button.
- Open the now-extracted nhgis0002_csv folder and confirm that it contains two files: a comma separated values (.csv) file and a text (.txt) file.

Analyze the Table in Microsoft Excel

- Double-click on the 'nhgis0002_ts_nominal_place.csv' 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? _____

- Copy the Minnesota records along with the 2nd row, containing descriptions, to a new worksheet. Using the Filter tool on the STATE column is a quick way to isolate these.

10) How many place records are there for Minnesota? _____

- To see all the field descriptions, select the top row and 'Wrap Text'.

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

- Change the font color for these columns to highlight them.
- Create 2 new variables called "CollegeGrad90" and "CollegeGrad0812", and sum 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 variable called 'ChangeCollegeGrad', and compute the total change in college grads between 1990 and 2008-2012 for all places.

13) Which city had the highest 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 variables called Total90 and Total0812, 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 25+ of St. Paul in 2008-2012? _____

- Create 2 more new variables called %College90 and %College0812. 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 variables between 1990 and 2008-2012

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

You have finished Exercise 2. Please check your answers on the next two pages.

Answers:

1) How many source tables are available when you filter only on Topic = 'Educational Attainment'? 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 1 time series table?

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

5) What advantage is there in using this table rather than the 'Persons 18 Years and Over by Educational Attainment [7]'? A large portion of people aged 18-24 are still actively working to complete a degree. The 25+ table helpfully 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

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 highest increase? How much was it? Minneapolis: +40,568

14) What was the total population 25+ of 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 highest increase in its proportion of college graduates?
Carver: +43.7