Tutorial for working with SEER, SEER*STAT, and Joinpoint Software

University of Nebraska Medical Center

Tutorial Information

- This tutorial will cover many topics that relate to Surveillance, Epidemiology, and End Results Program (SEER) data and SEER*related software.

- SEER is a program of the National Cancer Institute which primarily provides information on cancer statistics, in an effort to reduce cancer burden on the U.S. population.

- SEER collects and maintains U.S. mortality data for all causes of death, not just cancer deaths.

Tutorial Information

- **SEER data is free to use but those who want to use it must get permission from SEER.**

- **This is accomplished through submitting a request on the SEER website and signing a data-use agreement. There is a detailed section in this tutorial that will walk you through this process step-by-step.**

Tutorial Information (cont.)

- There are many sections to this tutorial, and not all may be relevant to your needs. If you haven't used SEER data or software before, you will want to start with these two:
  - Getting Access to SEER
  - Downloading SEER*Stat Software (the software used to query SEER data)

- Note: gaining initial access to SEER data may take at least a business day or two, so plan accordingly. Also, you may need to contact your IT department to be able to download SEER-related software (if you don't have administrative rights on your computer), so keep that in mind as well.

Tutorial Information (cont.)

- There are different ways to access SEER data. Most people will use their internet connection and password from SEER to access data through SEER*Stat.
- You can download SEER data for off-line use if you may not have internet access and want to query SEER data, or you can download SEER data to use in SAS. If these situations are not relevant to you, feel free to skip those sections.
- There are many different examples regarding how to use SEER data (e.g. frequencies, incidence rates), and each will walk you through an example. Again, jump to the section that is of interest to you.

Tutorial Contents

Getting Access to SEER

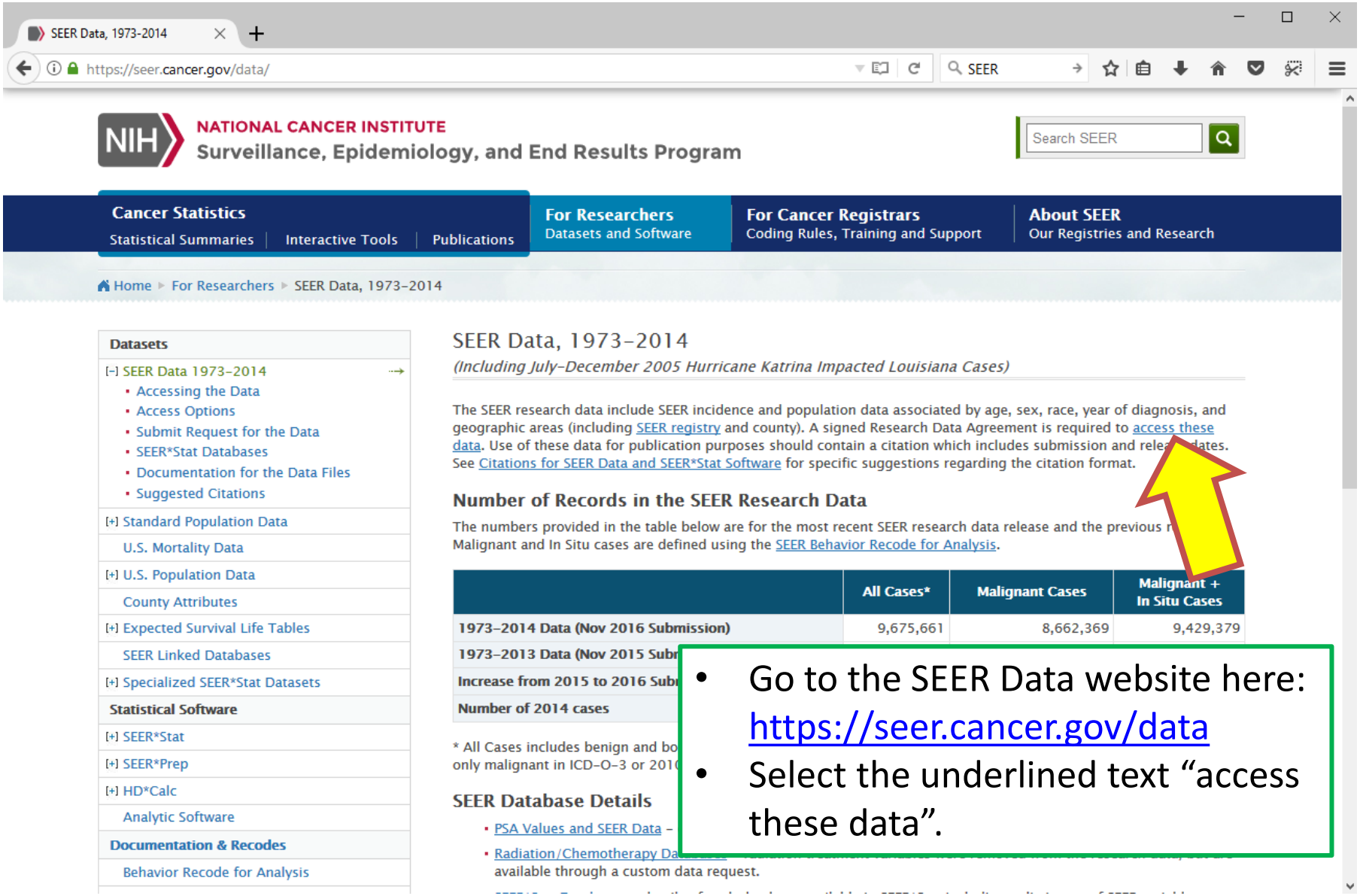
University of Nebraska Medical Center

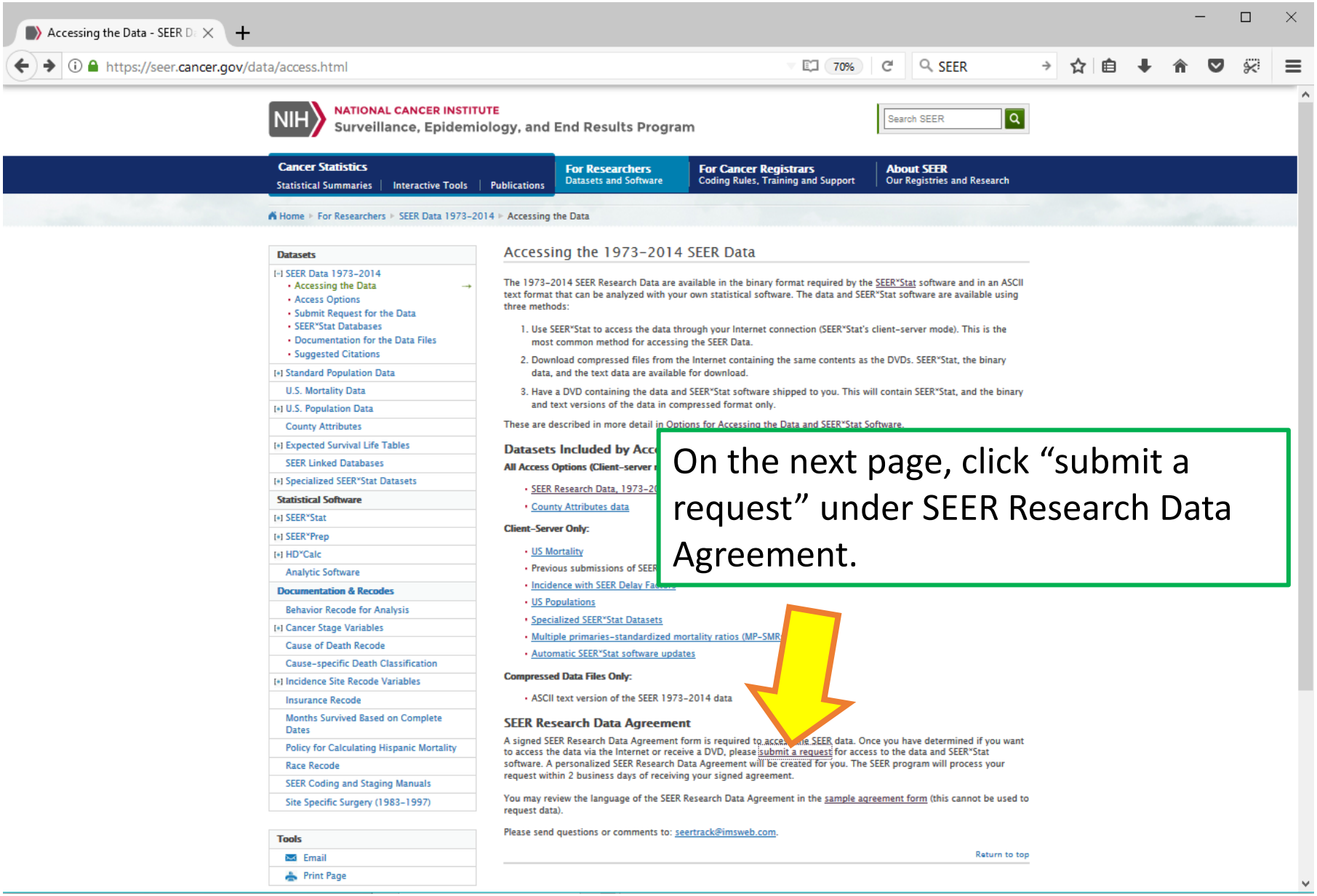
Getting Access to SEER Data

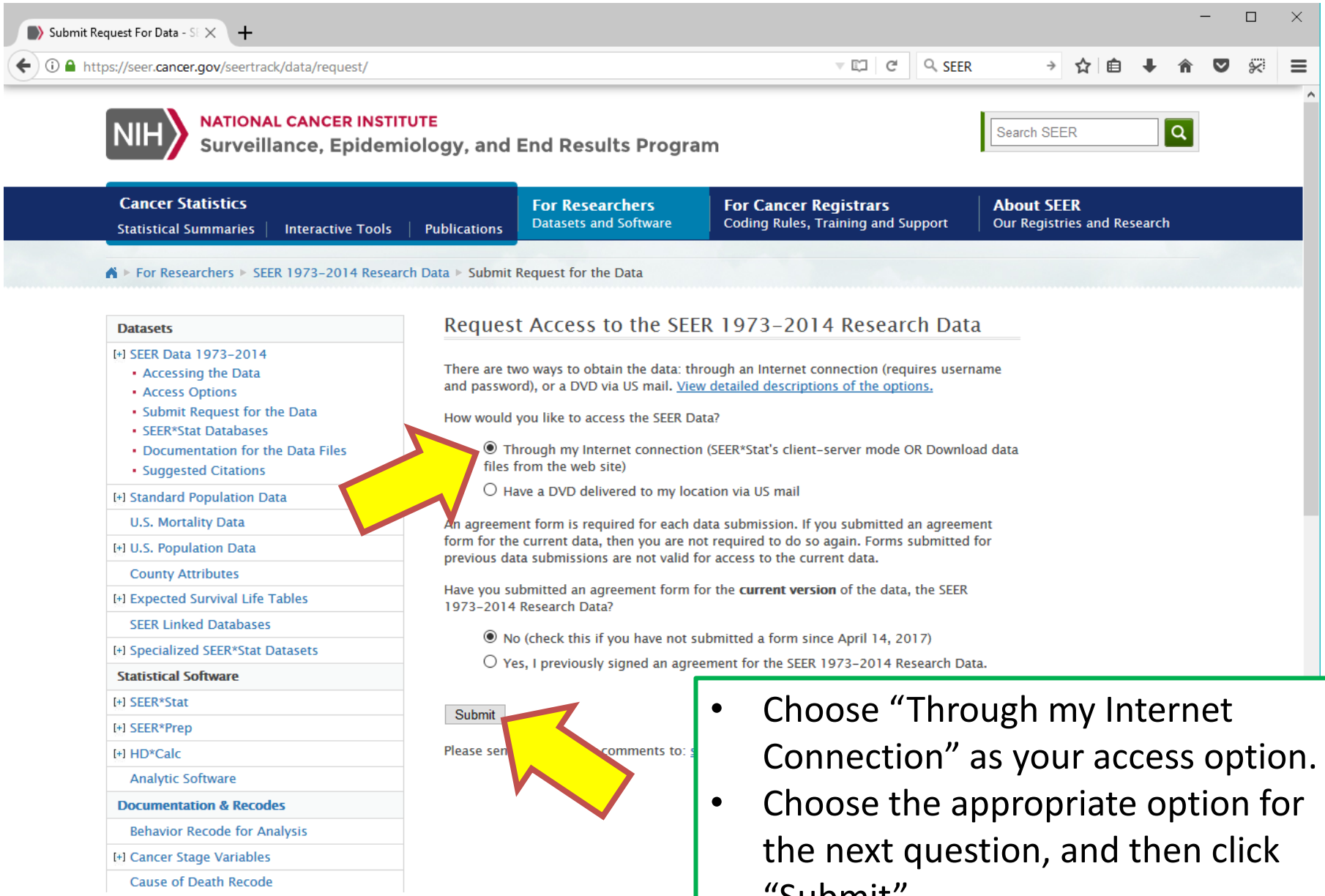
This section will instruct you on how to get access to SEER data, using the following steps:

- **Submitting a data request**
- **Receiving a verification email**
- **Sending a signed data use agreement**
- **Receiving a username and password**

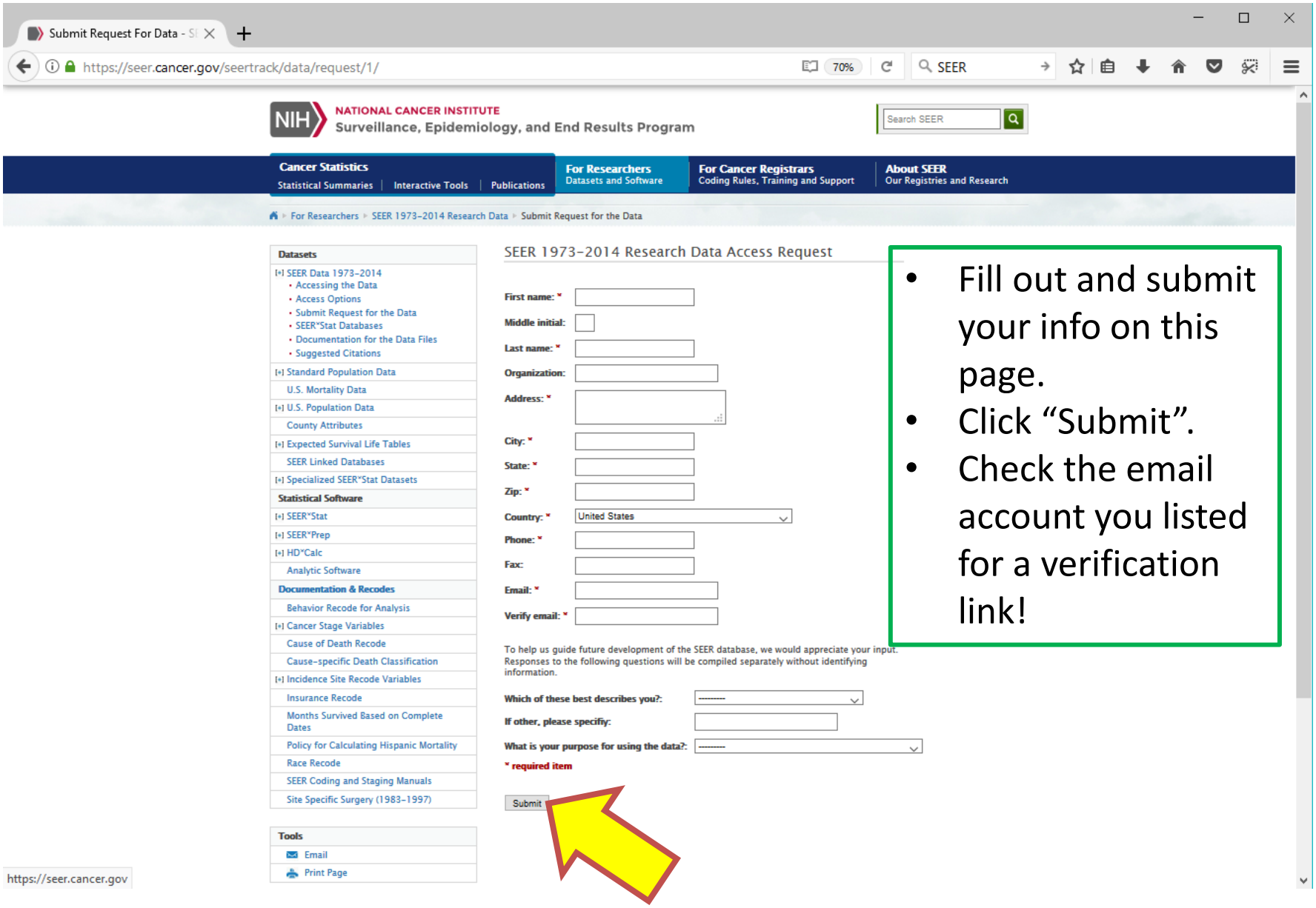
Once you receive your username and password, you will be able to download SEER data and SEER*Stat software.


SEER Data, 1973-2014
https://seer.cancer.gov/data/
SEER
NATIONAL CANCER INSTITUTE
Surveillance, Epidemiology, and End Results Program
Search SEER
Cancer Statistics
Statistical Summaries | Interactive Tools | Publications
For Researchers
Datasets and Software
For Cancer Registrars
Coding Rules, Training and Support
About SEER
Our Registries and Research
Home ▸ For Researchers ▸ SEER Data, 1973–2014
Datasets
[−] SEER Data 1973–2014
Accessing the Data
Access Options
Submit Request for the Data
SEER*Stat Databases
Documentation for the Data Files
Suggested Citations
[+] Standard Population Data
U.S. Mortality Data
[+] U.S. Population Data
County Attributes
[+] Expected Survival Life Tables
SEER Linked Databases
[+] Specialized SEER*Stat Datasets
Statistical Software
[+] SEER*Stat
[+] SEER*Prep
[+] HD*Calc
Analytic Software
Documentation & Recodes
Behavior Recode for Analysis
SEER Data, 1973–2014
(Including July–December 2005 Hurricane Katrina Impacted Louisiana Cases)
The SEER research data include SEER incidence and population data associated by age, sex, race, year of diagnosis, and geographic areas (including SEER registry and county). A signed Research Data Agreement is required to access these data. Use of these data for publication purposes should contain a citation which includes submission and rele dates. See Citations for SEER Data and SEER*Stat Software for specific suggestions regarding the citation format.
Number of Records in the SEER Research Data
The numbers provided in the table below are for the most recent SEER research data release and the previous r
Malignant and In Situ cases are defined using the SEER Behavior Recode for Analysis.
All Cases*
Malignant Cases
Malignant + In Situ Cases
1973–2014 Data (Nov 2016 Submission)
9,675,661
8,662,369
9,429,379
1973–2013 Data (Nov 2015 Subr
Increase from 2015 to 2016 Subr
Number of 2014 cases
* All Cases includes benign and bo
only malignant in ICD-O-3 or 201
SEER Database Details
PSA Values and SEER Data –
Radiation/Chemotherapy Da
available through a custom data request.
• Go to the SEER Data website here: https://seer.cancer.gov/data
• Select the underlined text “access these data”.

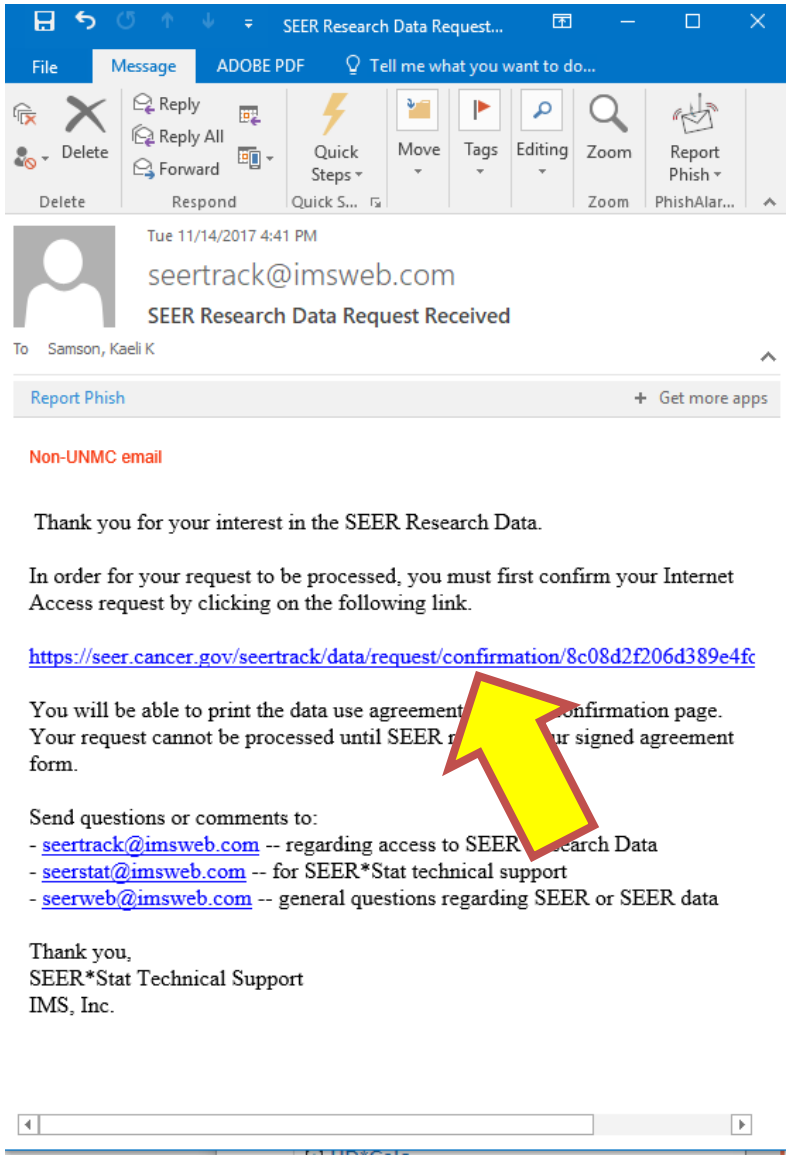

Accessing the Data - SEER D
https://seer.cancer.gov/data/access.html
70%
SEER
NIH
NATIONAL CANCER INSTITUTE
Surveillance, Epidemiology, and End Results Program
Search SEER
Cancer Statistics
Statistical Summaries
Interactive Tools
Publications
For Researchers
Datasets and Software
For Cancer Registrars
Coding Rules, Training and Support
About SEER
Our Registries and Research
Home
For Researchers
SEER Data 1973–2014
Accessing the Data
Datasets
SEER Data 1973–2014
Accessing the Data
Access Options
Submit Request for the Data
SEER*Stat Databases
Documentation for the Data Files
Suggested Citations
Standard Population Data
U.S. Mortality Data
U.S. Population Data
County Attributes
Expected Survival Life Tables
SEER Linked Databases
Specialized SEER*Stat Datasets
Statistical Software
SEER*Stat
SEER*Prep
HD*Calc
Analytic Software
Documentation & Recodes
Behavior Recode for Analysis
Cancer Stage Variables
Cause of Death Recode
Cause-specific Death Classification
Incidence Site Recode Variables
Insurance Recode
Months Survived Based on Complete Dates
Policy for Calculating Hispanic Mortality
Race Recode
SEER Coding and Staging Manuals
Site Specific Surgery (1983–1997)
Tools
Email
Print Page
Accessing the 1973–2014 SEER Data
The 1973–2014 SEER Research Data are available in the binary format required by the SEER*Stat software and in an ASCII text format that can be analyzed with your own statistical software. The data and SEER*Stat software are available using three methods:
1. Use SEER*Stat to access the data through your Internet connection (SEER*Stat's client-server mode). This is the most common method for accessing the SEER Data.
2. Download compressed files from the Internet containing the same contents as the DVDs. SEER*Stat, the binary data, and the text data are available for download.
3. Have a DVD containing the data and SEER*Stat software shipped to you. This will contain SEER*Stat, and the binary and text versions of the data in compressed format only.
These are described in more detail in Options for Accessing the Data and SEER*Stat Software.
Datasets Included by Acc
All Access Options (Client-server
SEER Research Data, 1973–2
County Attributes data
Client-Server Only:
US Mortality
Previous submissions of SEER
Incidence with SEER Delay Fa
US Populations
Specialized SEER*Stat Datasets
Multiple primaries-standardized mortality ratios (MP-SMR
Automatic SEER*Stat software updates
Compressed Data Files Only:
ASCII text version of the SEER 1973–2014 data
SEER Research Data Agreement
A signed SEER Research Data Agreement form is required to acce SEER data. Once you have determined if you want to access the data via the Internet or receive a DVD, please submit a request for access to the data and SEER*Stat software. A personalized SEER Research Data Agreement will be created for you. The SEER program will process your request within 2 business days of receiving your signed agreement.
You may review the language of the SEER Research Data Agreement in the sample agreement form (this cannot be used to request data).
Please send questions or comments to: seertrack@imsweb.com.
Return to top
On the next page, click "submit a request" under SEER Research Data Agreement.



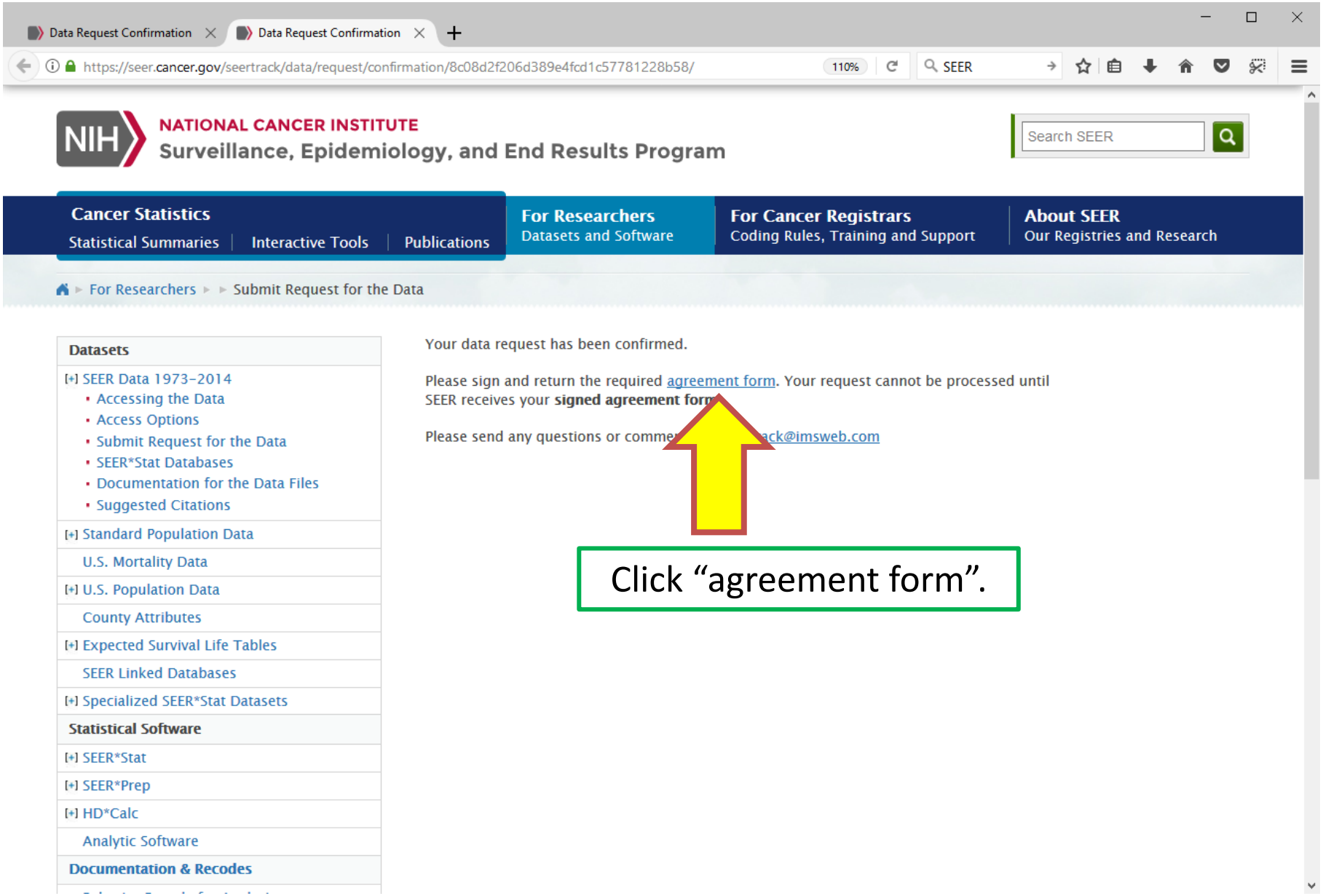
- Choose "Through my Internet Connection" as your access option.
- Choose the appropriate option for the next question, and then click "Submit".

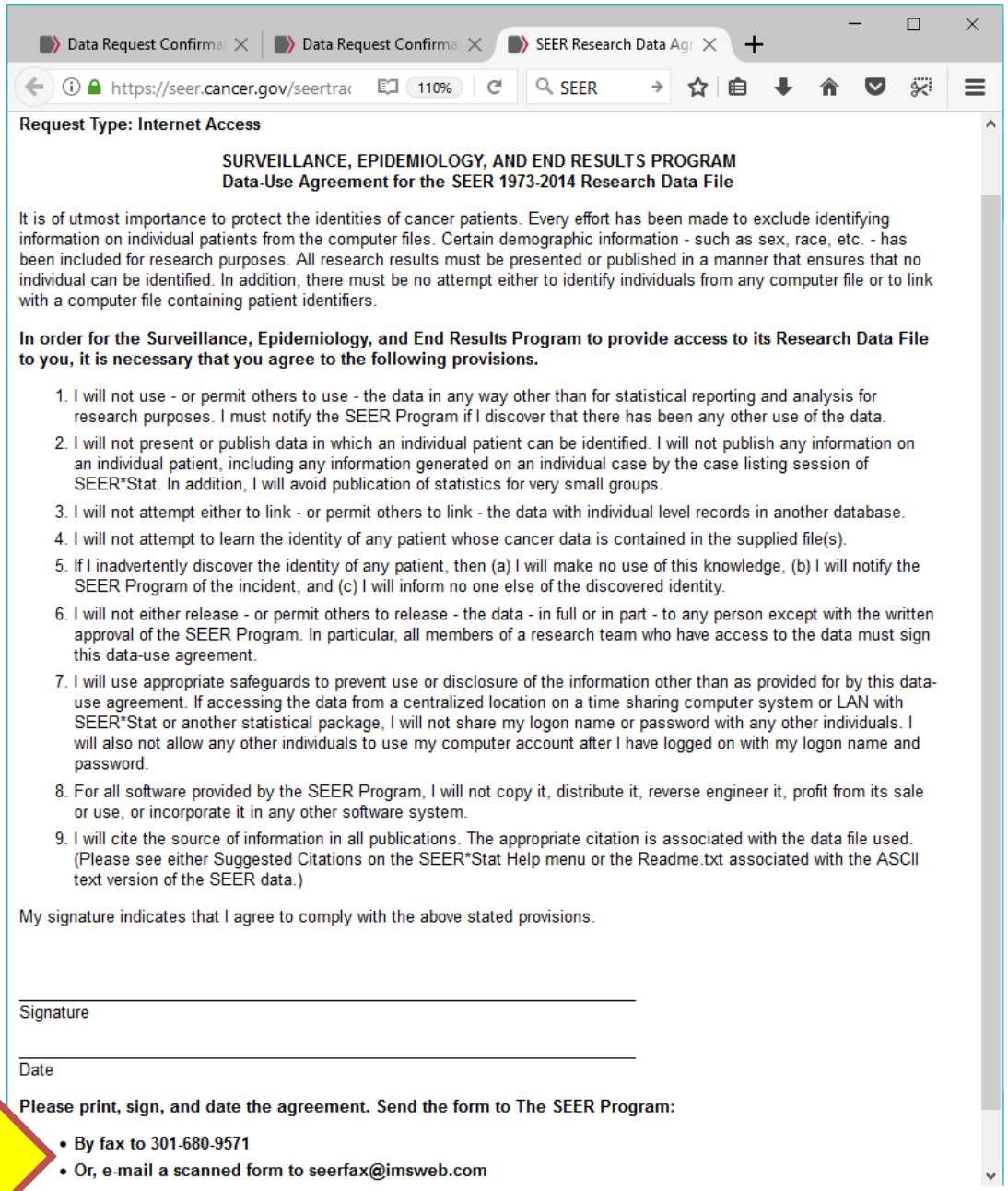


- Fill out and submit your info on this page.
- Click “Submit”.
- Check the email account you listed for a verification link!



- Check your email for a link to verify your access request.
- Click the link to take you to a website with the agreement form.


Data Request Confirmation
Data Request Confirmation
https://seer.cancer.gov/seertrack/data/request/confirmation/8c08d2f206d389e4fcd1c57781228b58/
110%
SEER
NATIONAL CANCER INSTITUTE
Surveillance, Epidemiology, and End Results Program
Search SEER
Cancer Statistics
Statistical Summaries
Interactive Tools
Publications
For Researchers
Datasets and Software
For Cancer Registrars
Coding Rules, Training and Support
About SEER
Our Registries and Research
For Researchers
Submit Request for the Data
Datasets
[+] SEER Data 1973–2014
Accessing the Data
Access Options
Submit Request for the Data
SEER*Stat Databases
Documentation for the Data Files
Suggested Citations
[+] Standard Population Data
U.S. Mortality Data
[+] U.S. Population Data
County Attributes
[+] Expected Survival Life Tables
SEER Linked Databases
[+] Specialized SEER*Stat Datasets
Statistical Software
[+] SEER*Stat
[+] SEER*Prep
[+] HD*Calc
Analytic Software
Documentation & Recodes
Your data request has been confirmed.
Please sign and return the required agreement form. Your request cannot be processed until SEER receives your signed agreement form.
Please send any questions or comments to seertrack@imsweb.com
Click “agreement form”.



Request Type: Internet Access

SURVEILLANCE, EPIDEMIOLOGY, AND END RESULTS PROGRAM
Data-Use Agreement for the SEER 1973-2014 Research Data File

It is of utmost importance to protect the identities of cancer patients. Every effort has been made to exclude identifying information on individual patients from the computer files. Certain demographic information - such as sex, race, etc. - has been included for research purposes. All research results must be presented or published in a manner that ensures that no individual can be identified. In addition, there must be no attempt either to identify individuals from any computer file or to link with a computer file containing patient identifiers.

In order for the Surveillance, Epidemiology, and End Results Program to provide access to its Research Data File to you, it is necessary that you agree to the following provisions.

1. I will not use - or permit others to use - the data in any way other than for statistical reporting and analysis for research purposes. I must notify the SEER Program if I discover that there has been any other use of the data.
2. I will not present or publish data in which an individual patient can be identified. I will not publish any information on an individual patient, including any information generated on an individual case by the case listing session of SEER*Stat. In addition, I will avoid publication of statistics for very small groups.
3. I will not attempt either to link - or permit others to link - the data with individual level records in another database.
4. I will not attempt to learn the identity of any patient whose cancer data is contained in the supplied file(s).
5. If I inadvertently discover the identity of any patient, then (a) I will make no use of this knowledge, (b) I will notify the SEER Program of the incident, and (c) I will inform no one else of the discovered identity.
6. I will not either release - or permit others to release - the data - in full or in part - to any person except with the written approval of the SEER Program. In particular, all members of a research team who have access to the data must sign this data-use agreement.
7. I will use appropriate safeguards to prevent use or disclosure of the information other than as provided for by this data-use agreement. If accessing the data from a centralized location on a time sharing computer system or LAN with SEER*Stat or another statistical package, I will not share my logon name or password with any other individuals. I will also not allow any other individuals to use my computer account after I have logged on with my logon name and password.
8. For all software provided by the SEER Program, I will not copy it, distribute it, reverse engineer it, profit from its sale or use, or incorporate it in any other software system.
9. I will cite the source of information in all publications. The appropriate citation is associated with the data file used. (Please see either Suggested Citations on the SEER*Stat Help menu or the Readme.txt associated with the ASCII text version of the SEER data.)

My signature indicates that I agree to comply with the above stated provisions.

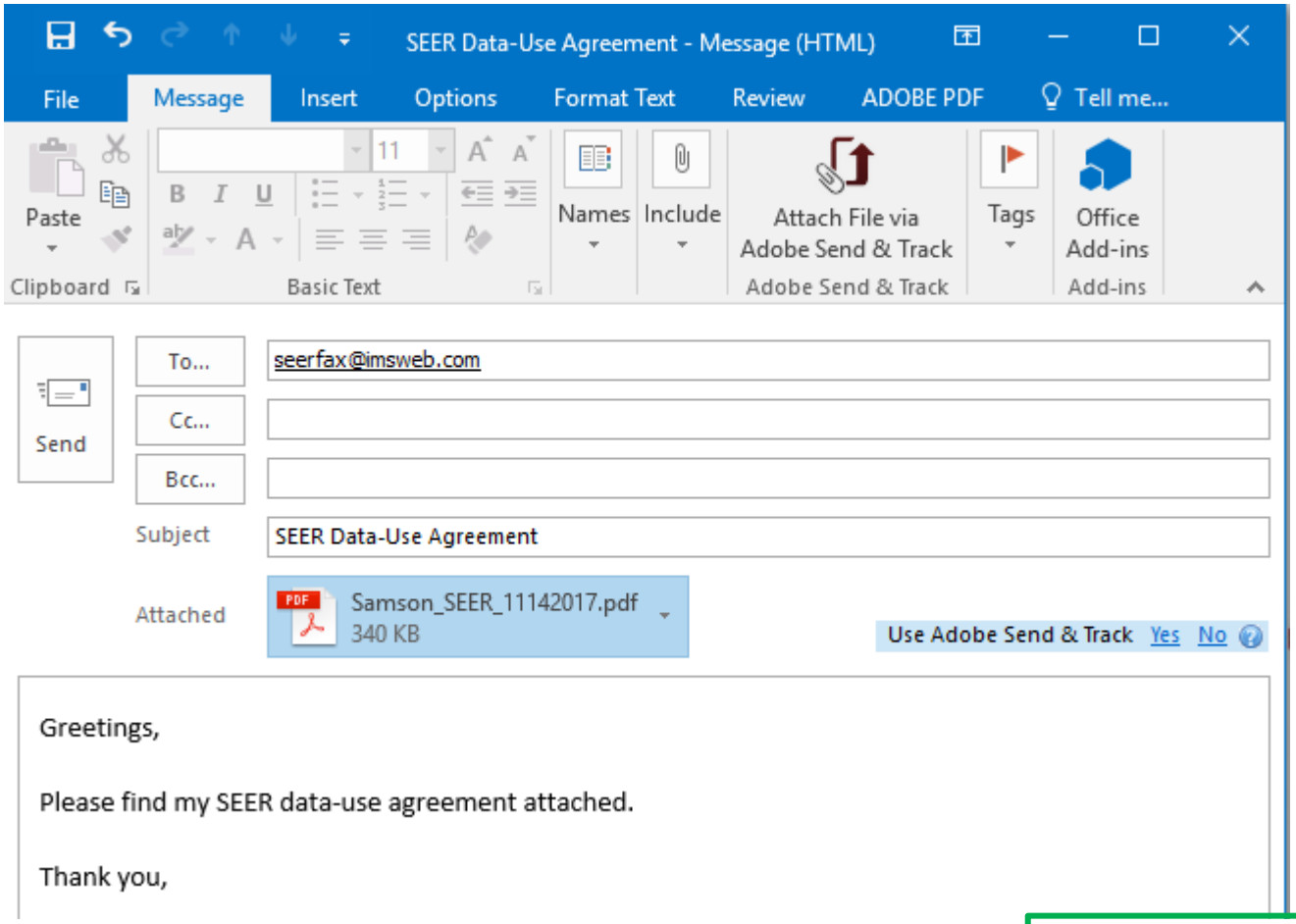
_____
Signature

_____
Date

Please print, sign, and date the agreement. Send the form to The SEER Program:

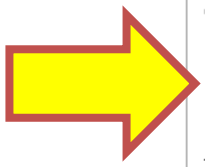
- **By fax to 301-680-9571**
- **Or, e-mail a scanned form to seerfax@imsweb.com**

- Print off this page, then sign and date it.
- Scan and email it to the email address listed at the bottom of the page, or fax it to the number shown at the bottom of the page



Example of scanning and emailing the signed and dated agreement.

- Wait for a response from SEER.
- This will typically take 1-3 business days.
- Their response will include a username and password.



-----Original Message-----
From: seertrack@imsweb.com [mailto:seertrack@imsweb.com]
Subject: SEER Data Request Approved

Thank you for your interest in the SEER Research Data. Your signed Research Data Agreement is on file at SEER. Your username and password have been generated for Internet access and they are shown below. Please note that both the username and password are case sensitive.

Username: XXXXXXX
Password: XXXXXXX

These will allow you to utilize the SEER*Stat client-server system and/or download the files which make up the SEER Research Data DVD. These options are described at the following URL:

http://seer.cancer.gov/data/options.html

You can change your password once you log into SEER*Stat from the "Client Server User Information" option located under the Profile menu.

The review of PSA data has been completed for 2004+ cases, therefore PSA values are available for 2004-2014 prostate cases. For more information, see https://seer.cancer.gov/data/psa-values.html.

Radiation treatment variables have been removed from the public research database starting with the November 2016 data submission. These variables are available through a custom data request process after signing an additional data use agreement that describes the completeness of the radiation treatment variable and the potential biases associated with use of the radiation data. To request access to

Downloading SEER*Stat software to query SEER data

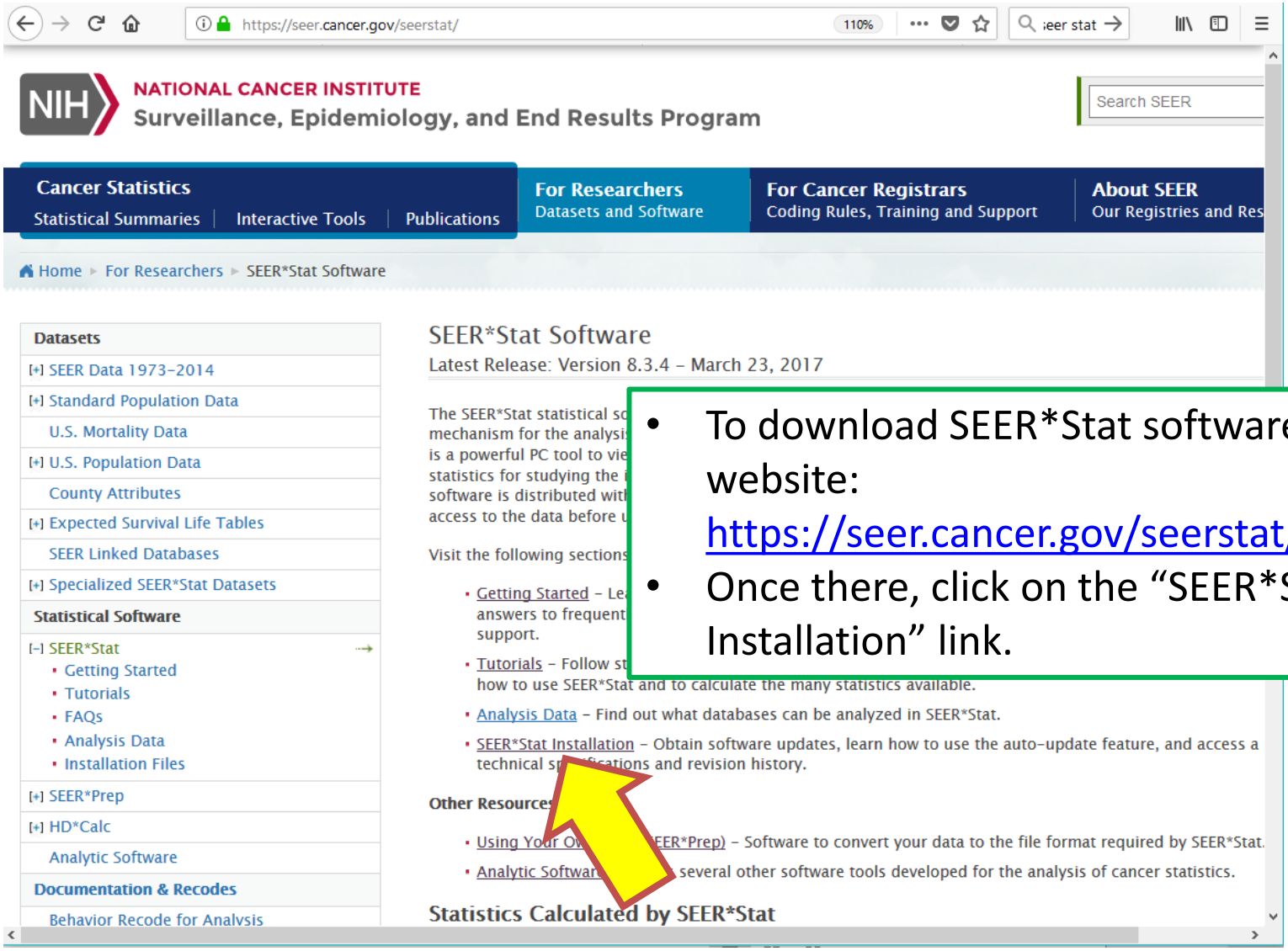
University of Nebraska Medical Center

Downloading SEER*Stat software to query SEER data

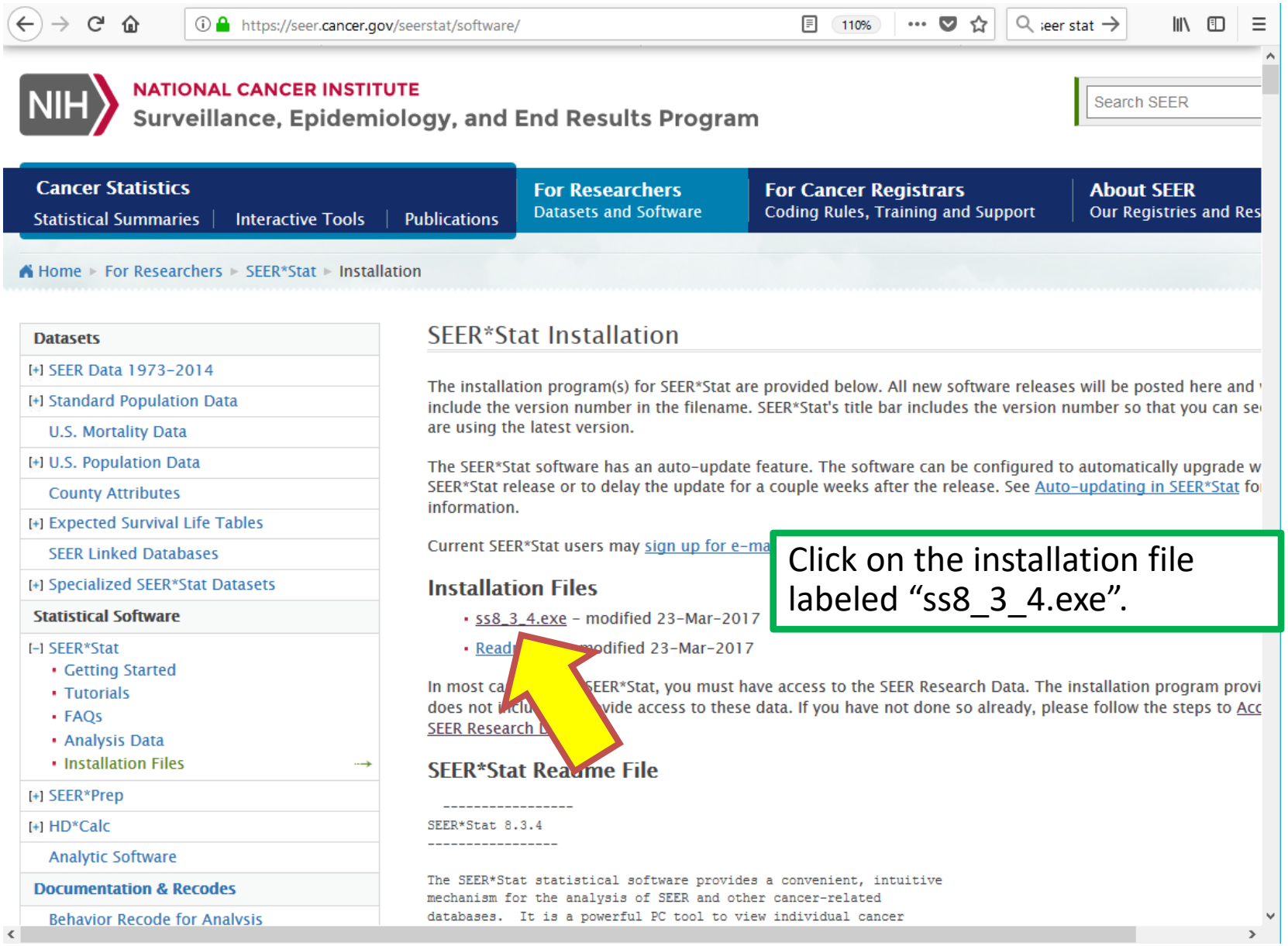
- **SEER*Stat is a versatile tool for making queries of SEER data.**
- **It is particularly useful for easily calculating age-adjusted rates.**
- **Note: SEER*Stat is a Windows application and will not run on Mac OSX!**

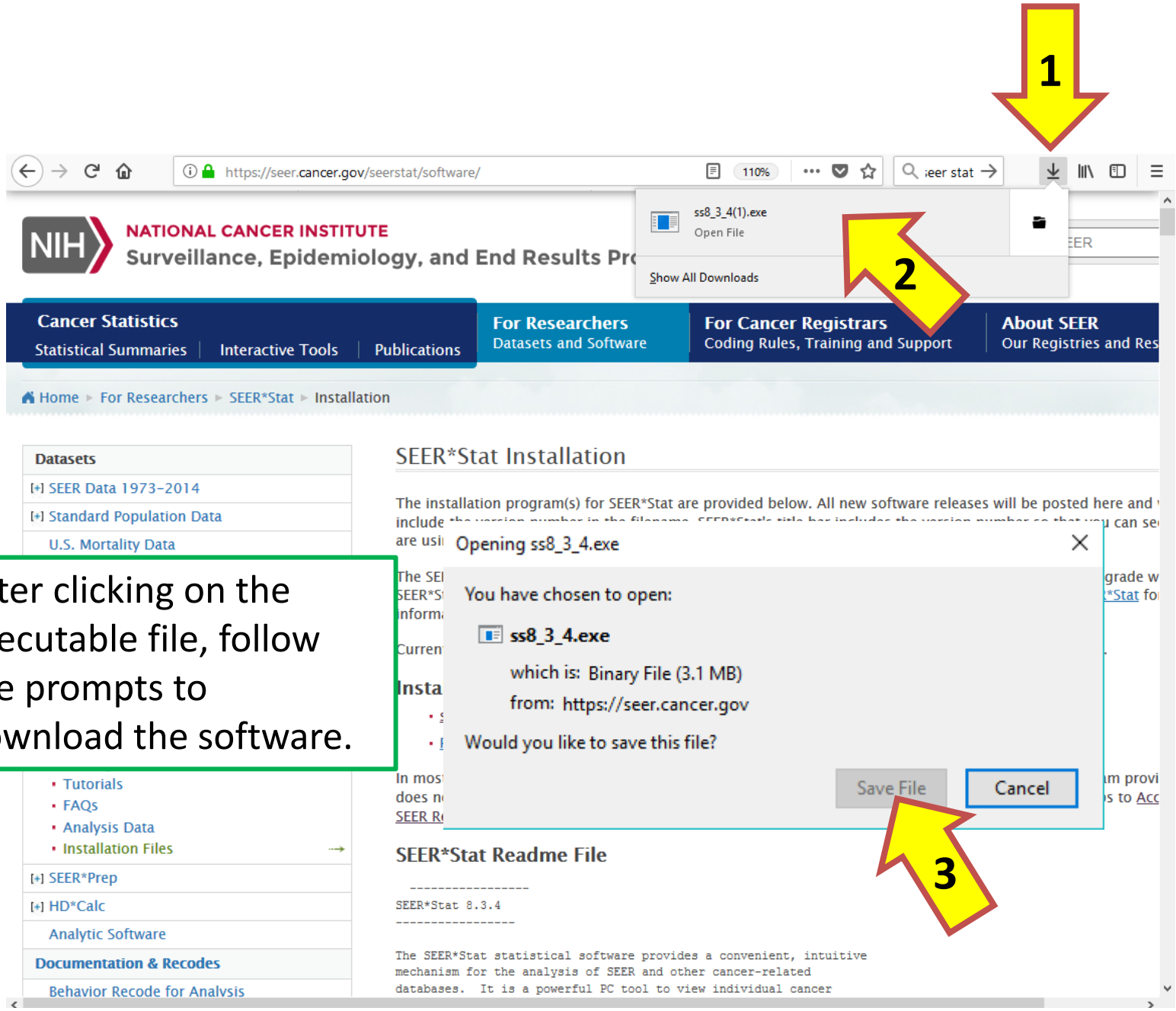
Downloading SEER*Stat software to query SEER data (cont.)

- **After downloading SEER*Stat you can instantly access SEER data by simply entering your username and password, given you have an internet connection. If you may not always have regular access to the internet, you may want to consider downloading the binary SEER data for off-line use in SEER*Stat (there is a section for doing this later in the tutorial, if needed).**

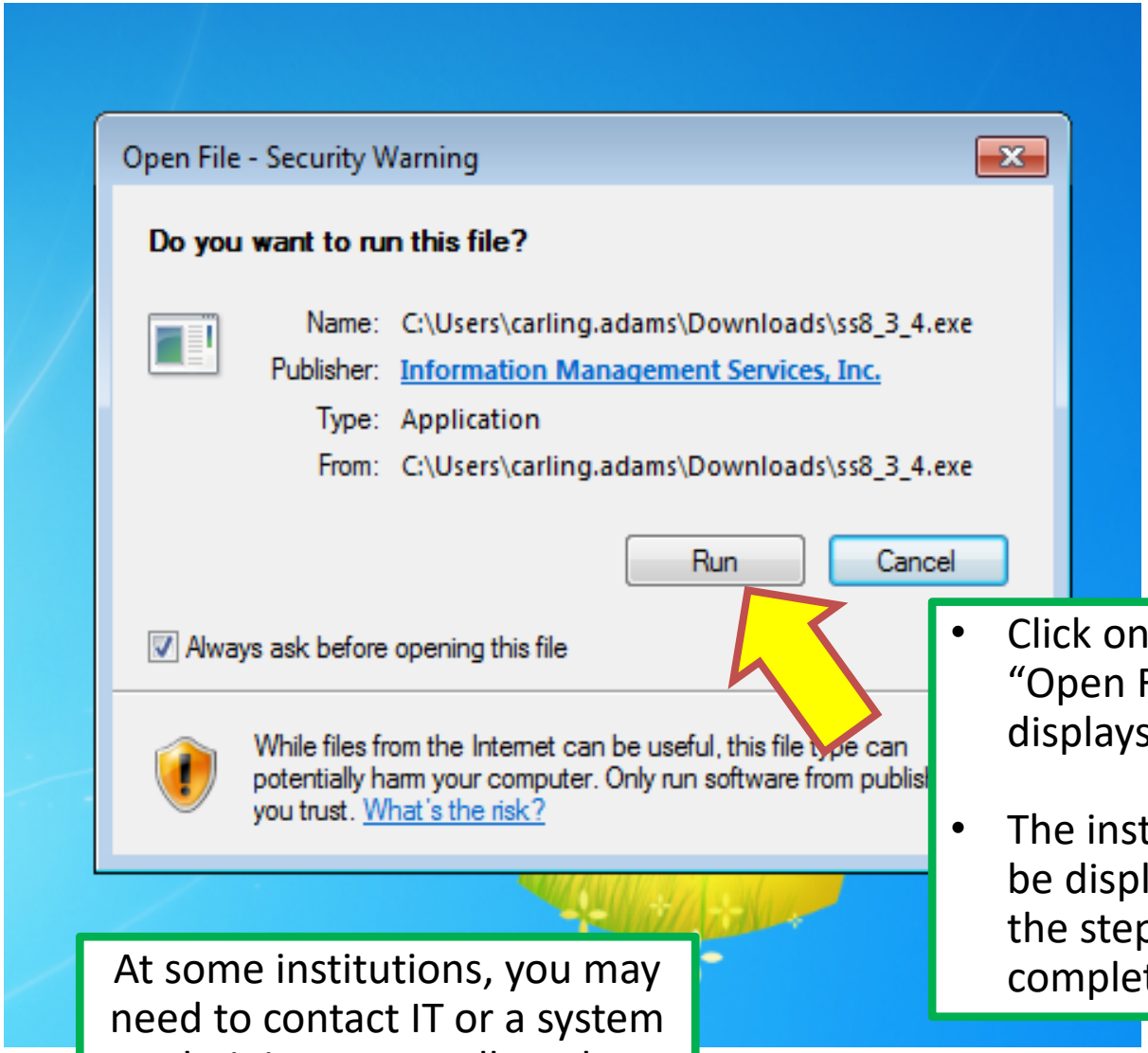


- To download SEER*Stat software, go the website: https://seer.cancer.gov/seerstat/software/
- Once there, click on the “SEER*Stat Installation” link.


https://seer.cancer.gov/seerstat/software/
110%
seer stat
NATIONAL CANCER INSTITUTE
Surveillance, Epidemiology, and End Results Program
Search SEER
Cancer Statistics
Statistical Summaries
Interactive Tools
Publications
For Researchers
Datasets and Software
For Cancer Registrars
Coding Rules, Training and Support
About SEER
Our Registries and Res
Home ▸ For Researchers ▸ SEER*Stat ▸ Installation
Datasets
[+] SEER Data 1973–2014
[+] Standard Population Data
U.S. Mortality Data
[+] U.S. Population Data
County Attributes
[+] Expected Survival Life Tables
SEER Linked Databases
[+] Specialized SEER*Stat Datasets
Statistical Software
[–] SEER*Stat
Getting Started
Tutorials
FAQs
Analysis Data
Installation Files
[+] SEER*Prep
[+] HD*Calc
Analytic Software
Documentation & Recodes
Behavior Recode for Analysis
SEER*Stat Installation
The installation program(s) for SEER*Stat are provided below. All new software releases will be posted here and
include the version number in the filename. SEER*Stat's title bar includes the version number so that you can se
are using the latest version.
The SEER*Stat software has an auto-update feature. The software can be configured to automatically upgrade w
SEER*Stat release or to delay the update for a couple weeks after the release. See Auto-updating in SEER*Stat fo
information.
Current SEER*Stat users may sign up for e-ma
Installation Files
ss8_3_4.exe – modified 23-Mar-2017
Read modified 23-Mar-2017
In most ca EER*Stat, you must have access to the SEER Research Data. The installation program provi
does not inclu vide access to these data. If you have not done so already, please follow the steps to Acc
SEER Research D
SEER*Stat Readme File
SEER*Stat 8.3.4
The SEER*Stat statistical software provides a convenient, intuitive
mechanism for the analysis of SEER and other cancer-related
databases. It is a powerful PC tool to view individual cancer
Click on the installation file labeled “ss8_3_4.exe”.



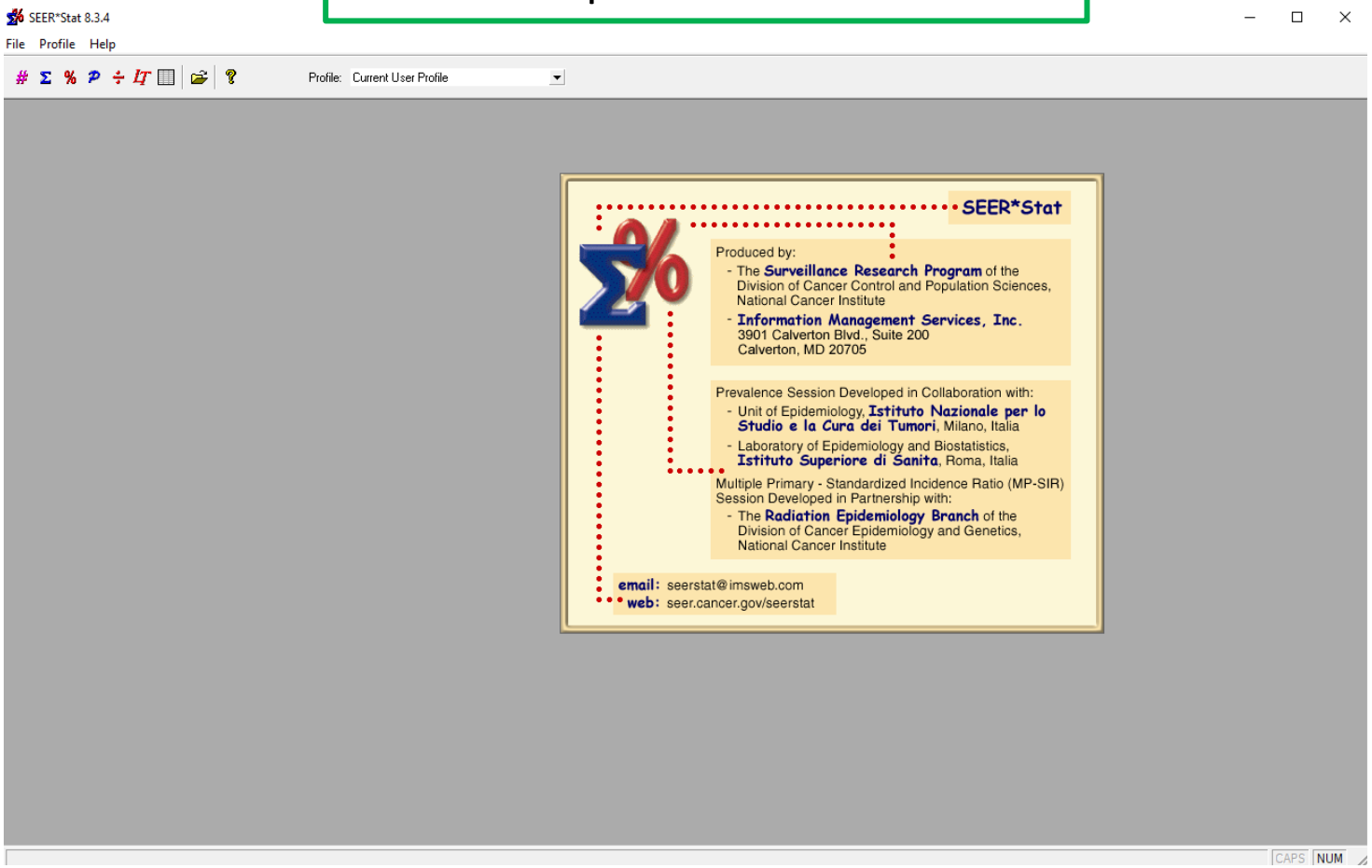
- After clicking on the executable file, follow the prompts to download the software.

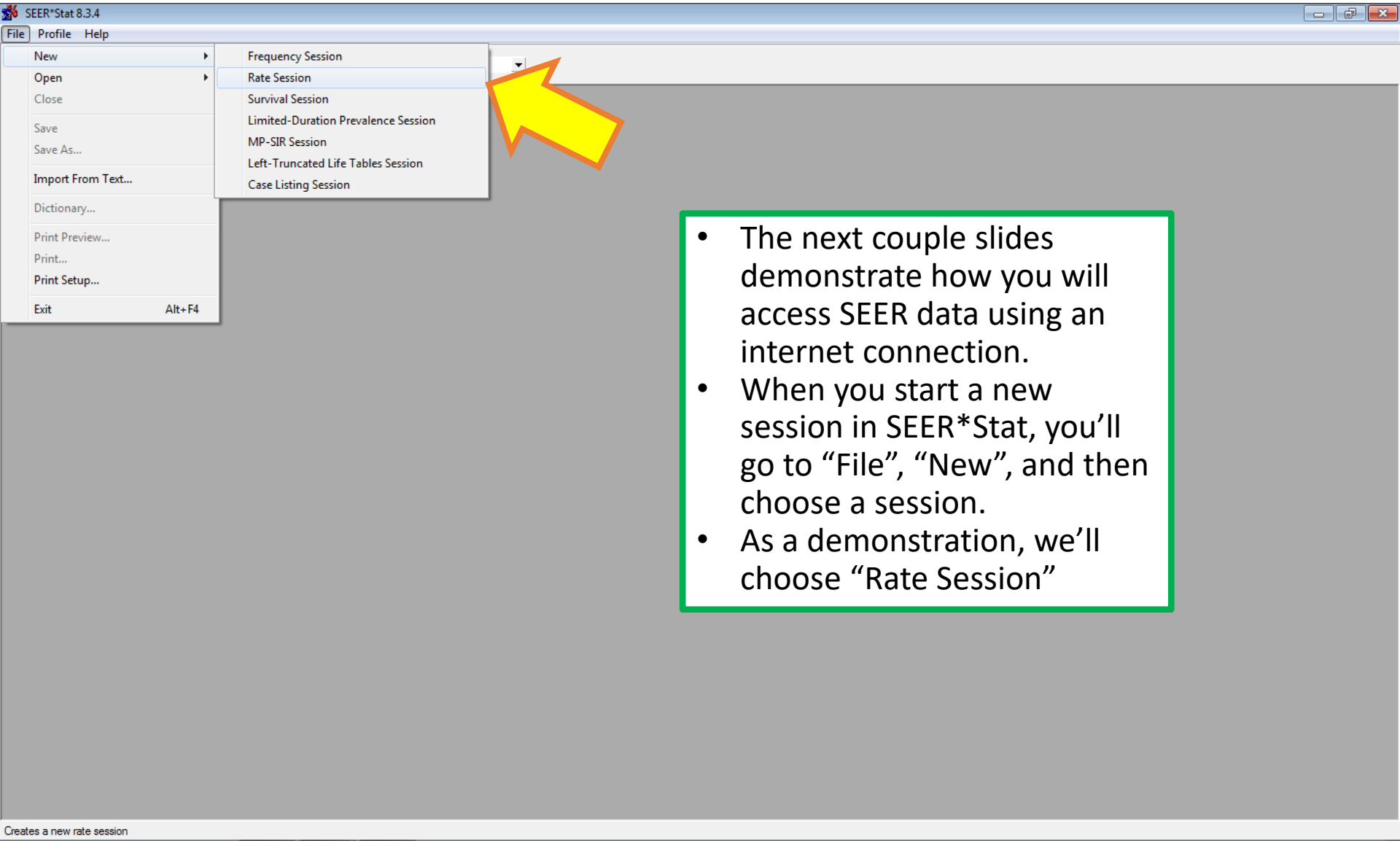


- Click on “Run” when the “Open File” window displays.
- The installation window will be displayed. Click through the steps in this window to complete installation.

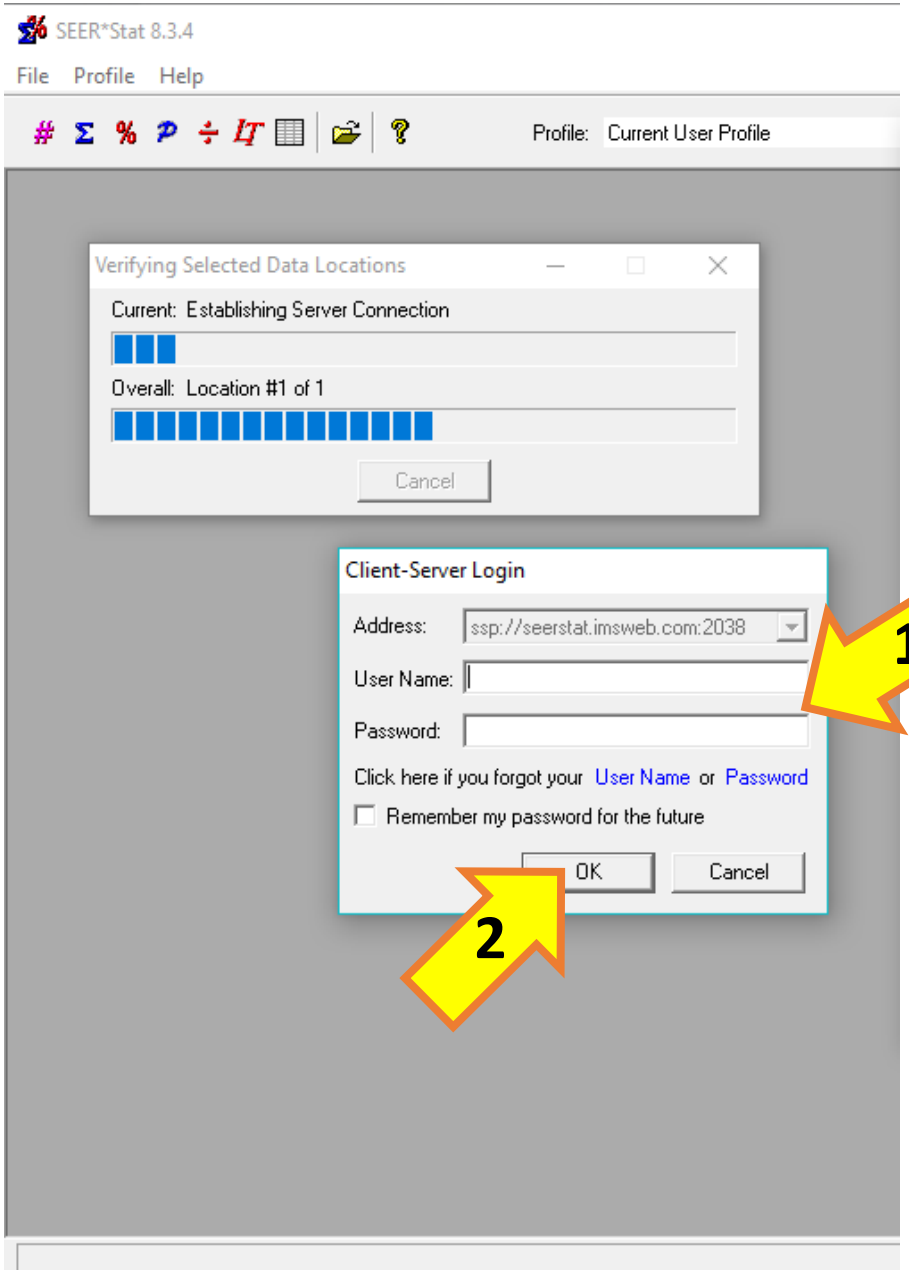
At some institutions, you may need to contact IT or a system administrator to allow the download on your computer.

- After following the prompts to download the software, it should create a shortcut on your desktop.
- Double-click on the desktop icon to open the software.

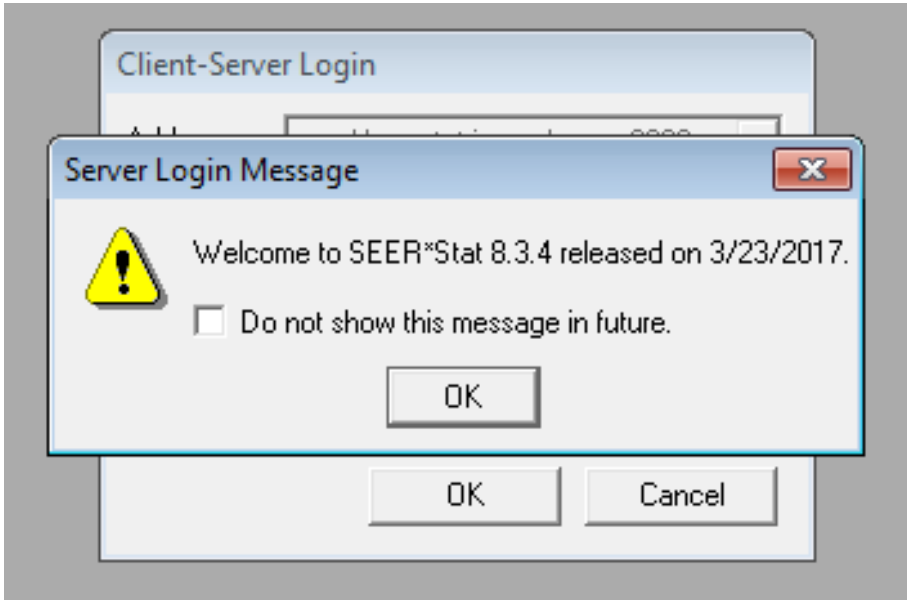




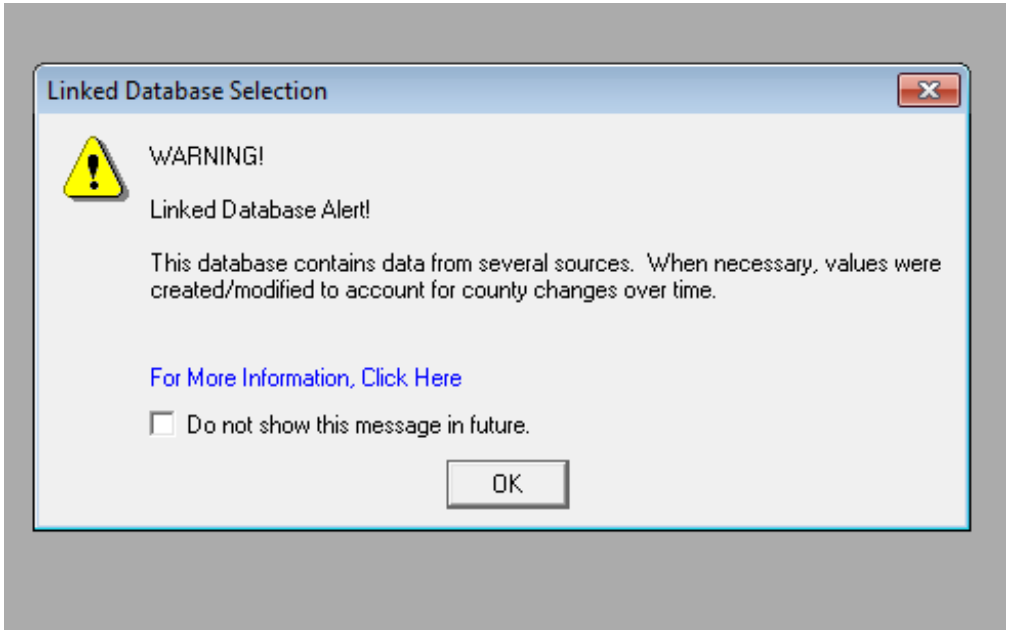
- The next couple slides demonstrate how you will access SEER data using an internet connection.
- When you start a new session in SEER*Stat, you'll go to "File", "New", and then choose a session.
- As a demonstration, we'll choose "Rate Session"



- A login window may appear to connect you to the server.
- Log in using the username and password given to you through an email from SEER.
- Click “OK”.



- Messages will pop up.
- You may click on the link in the "Linked Database Selection" window to read more about that alert.
- Click "OK" for both the welcome and warning messages to continue.
- Now you may proceed with your query.



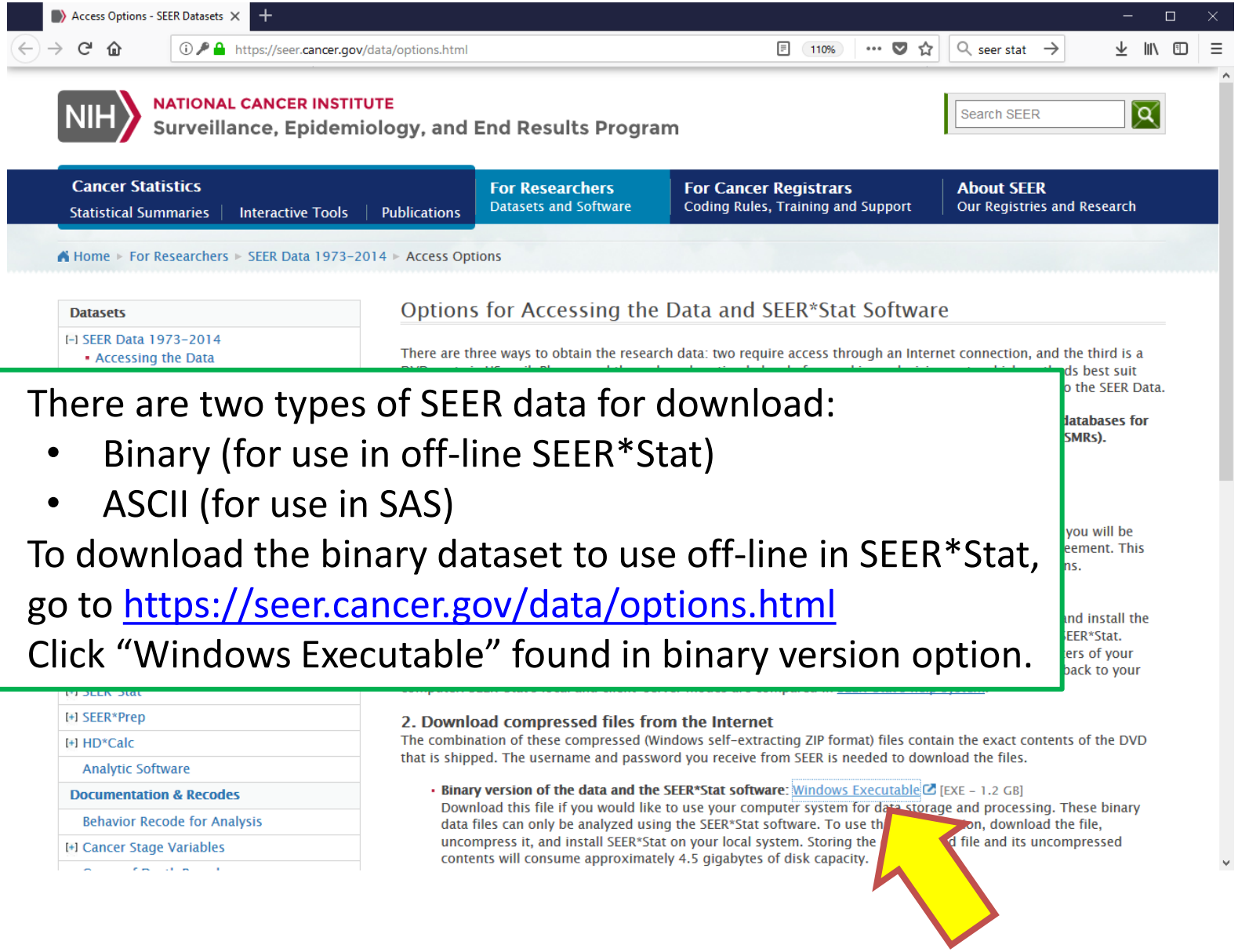
Downloading binary SEER Data to use in SEER*Software while off-line

University of Nebraska Medical Center

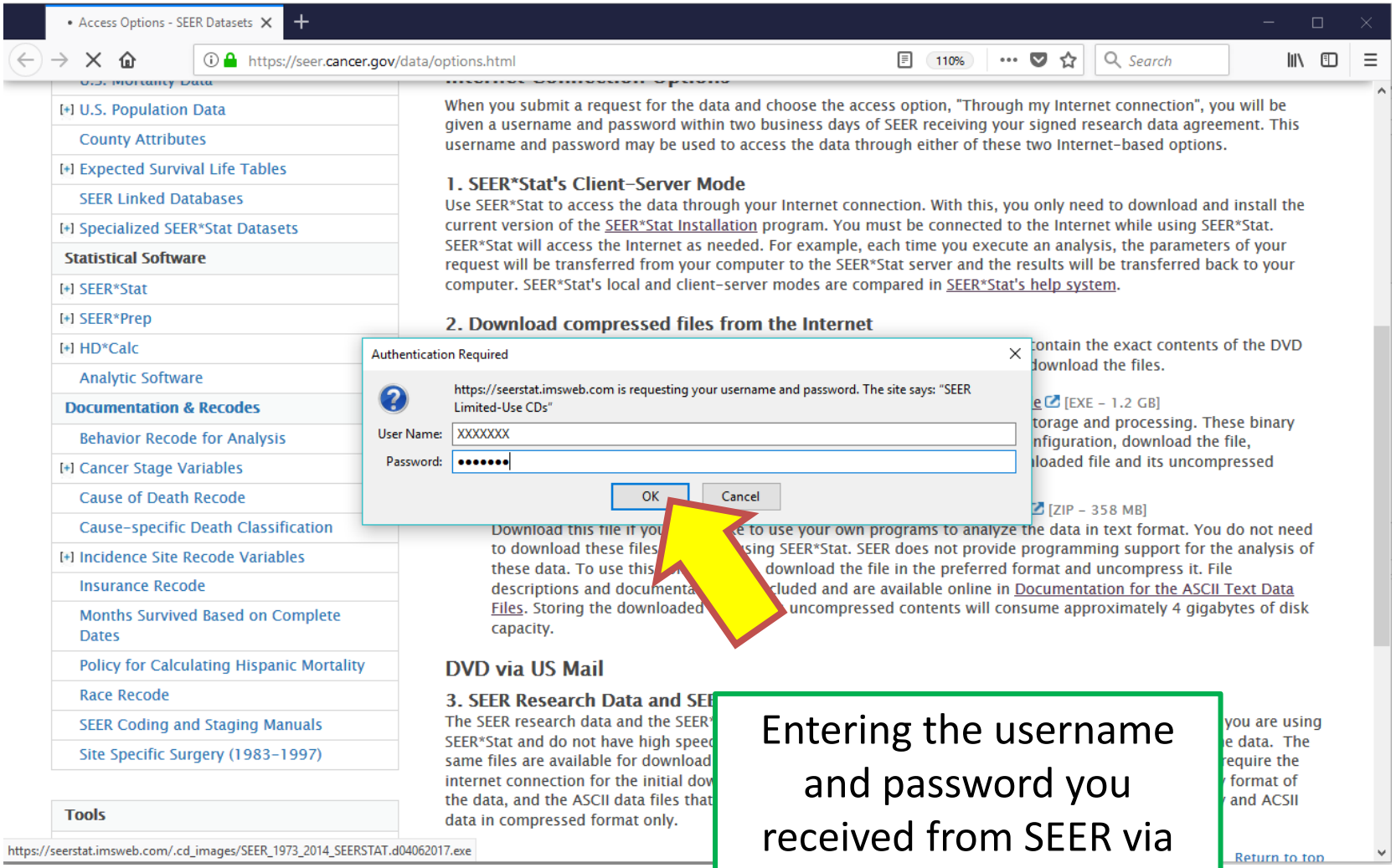
Downloading SEER Data to use in SEER*Software while off-line

- This section will instruct you on how to download binary SEER data to be able to query the data in SEER*Stat while not connected to the internet.

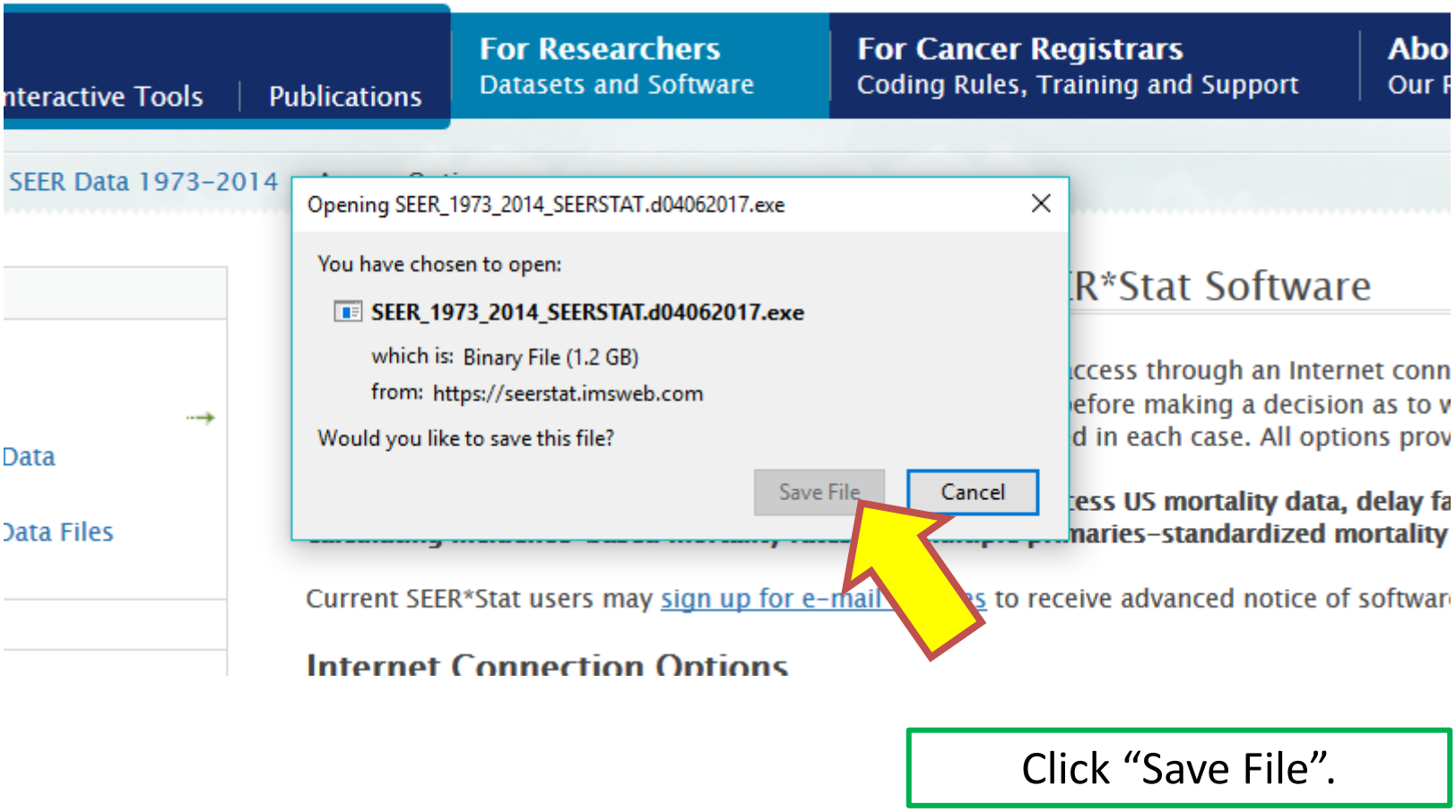
- If you have an internet connection, you can instantly access SEER data by simply entering your username and password when you open SEER*Stat. If you have regular access to the internet, you may want to skip this section.



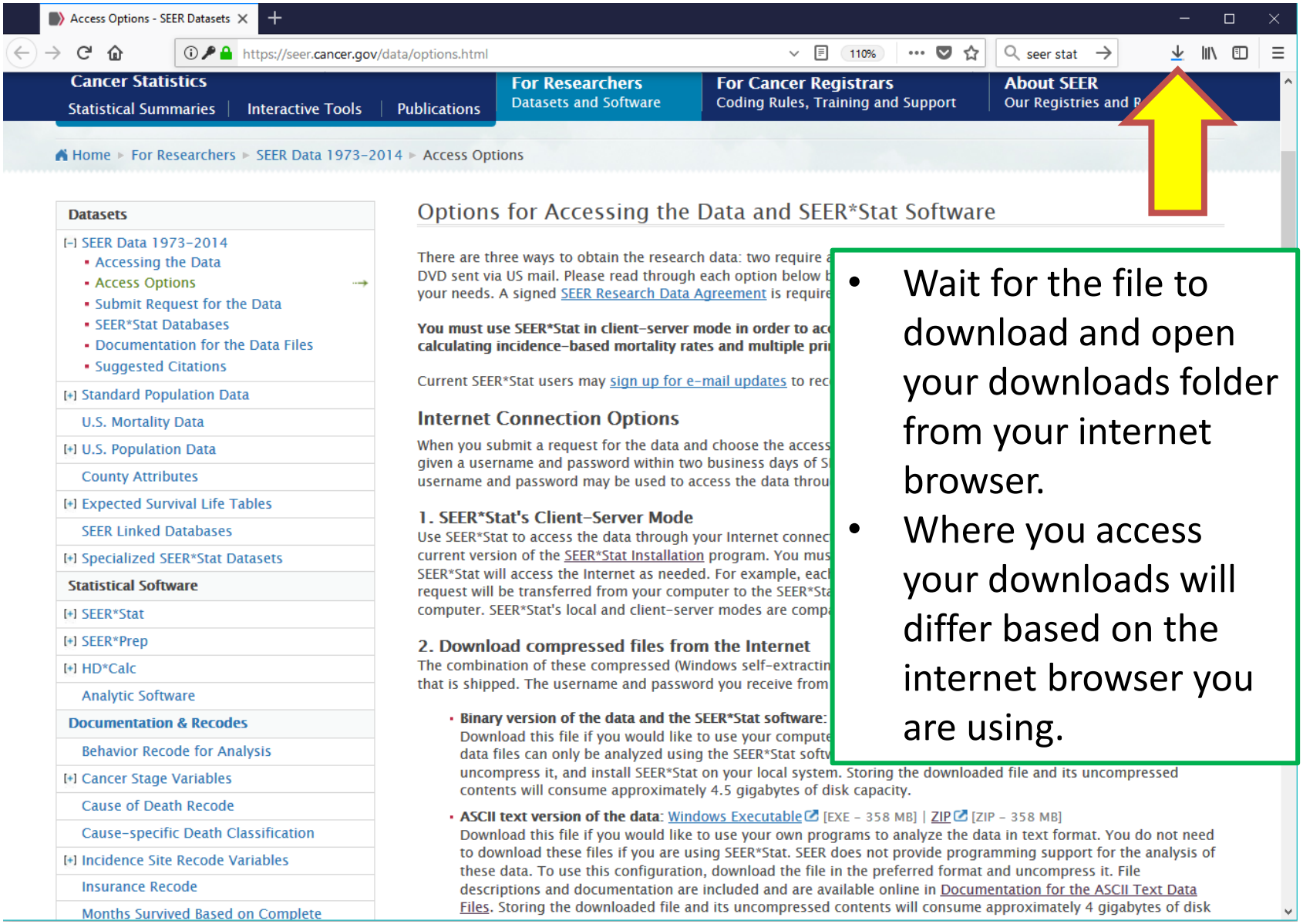
- There are two types of SEER data for download:
  - Binary (for use in off-line SEER*Stat)
  - ASCII (for use in SAS)
- To download the binary dataset to use off-line in SEER*Stat, go to https://seer.cancer.gov/data/options.html
- Click “Windows Executable” found in binary version option.



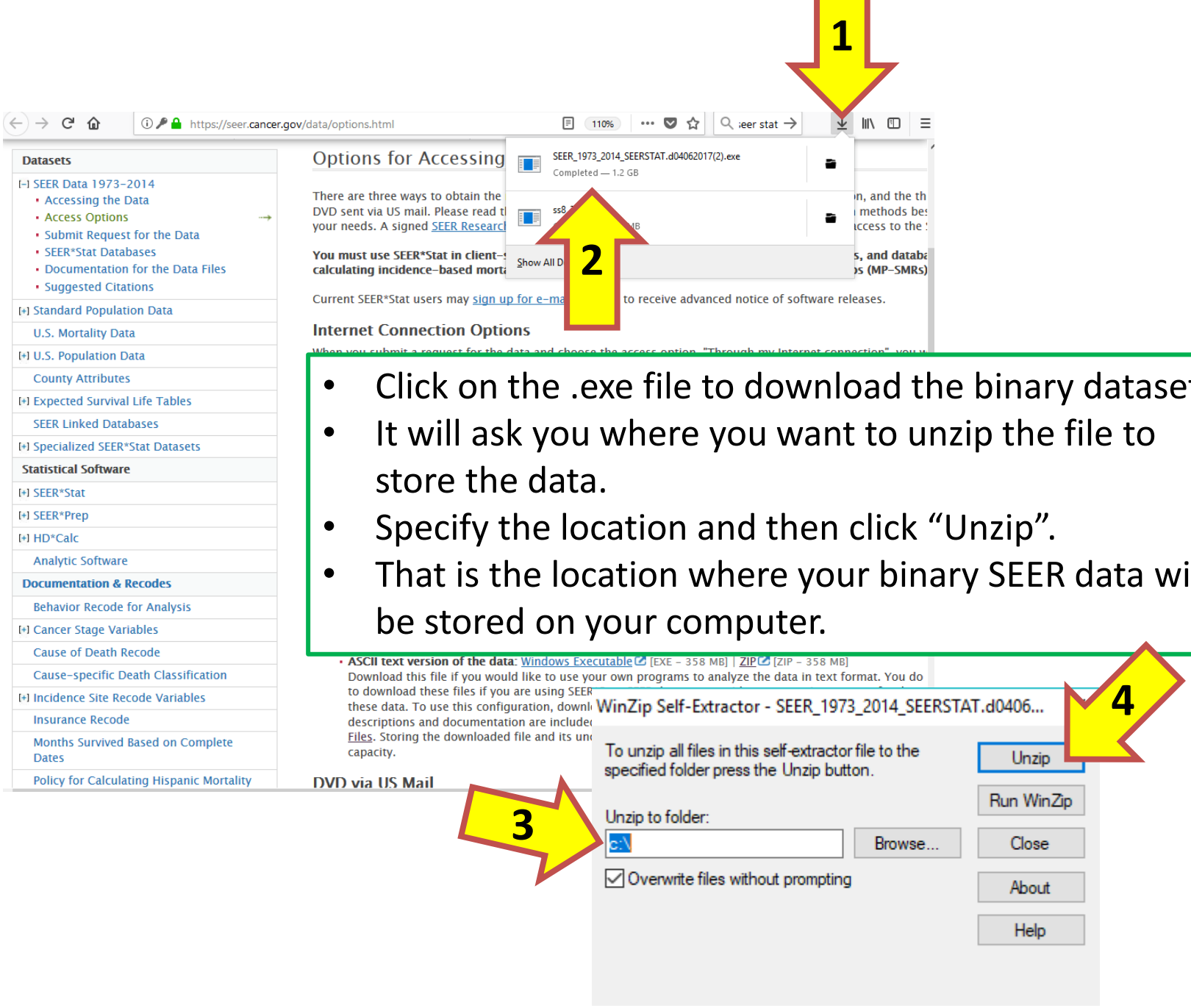
Entering the username and password you received from SEER via email and then click "OK".



Click “Save File”.



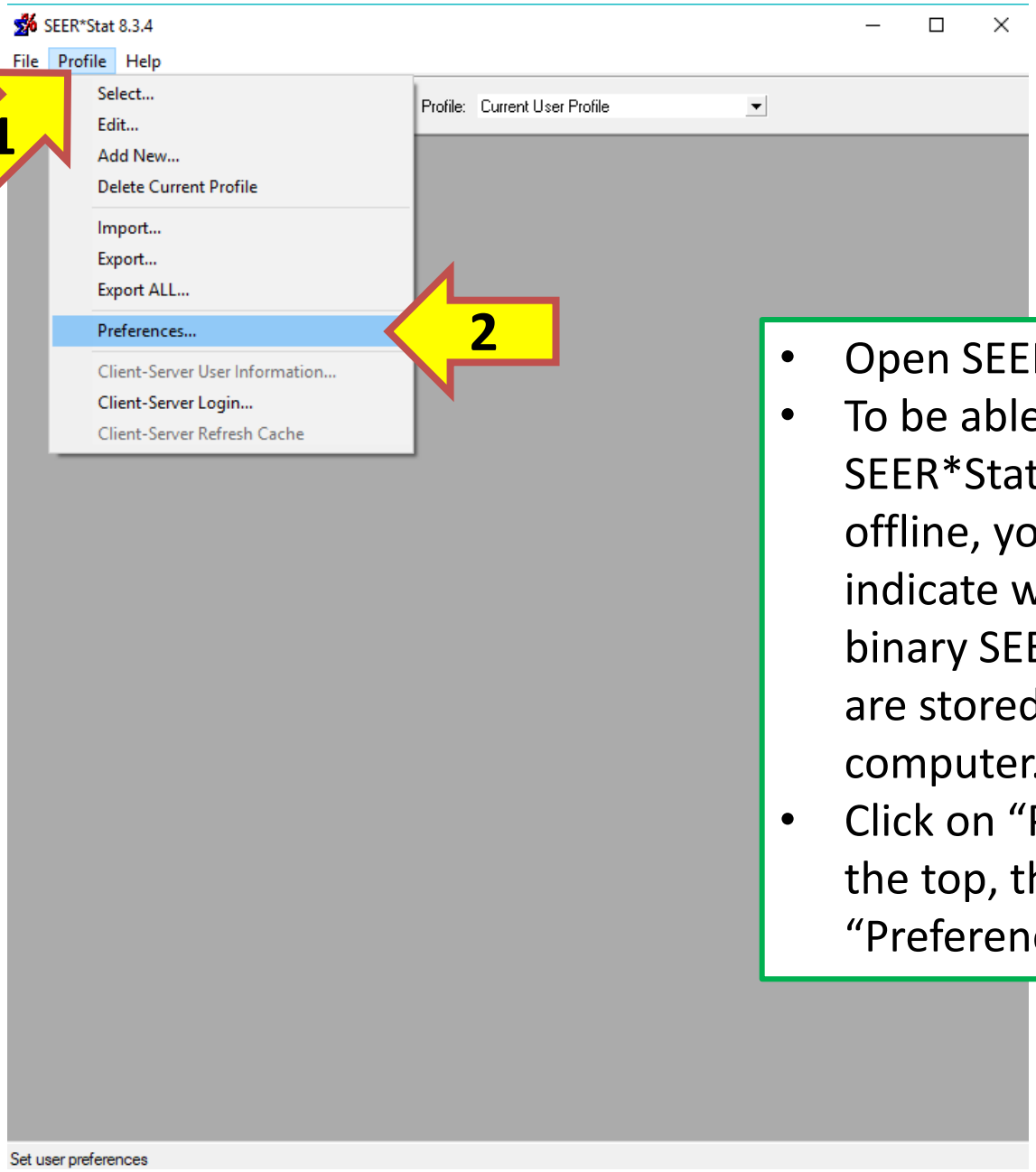
- Wait for the file to download and open your downloads folder from your internet browser.
- Where you access your downloads will differ based on the internet browser you are using.



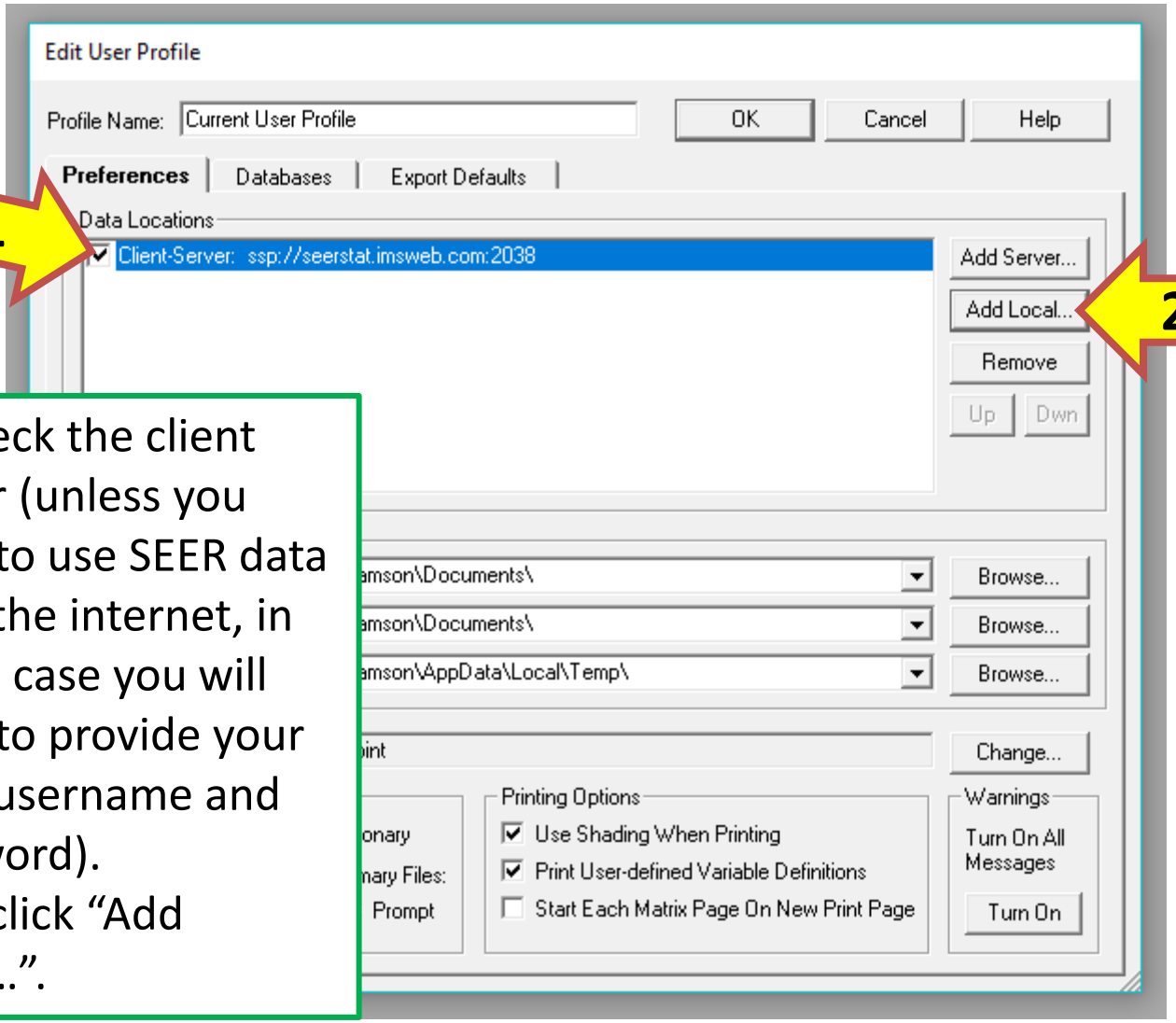
- Click on the .exe file to download the binary dataset.
- It will ask you where you want to unzip the file to store the data.
- Specify the location and then click “Unzip”.
- That is the location where your binary SEER data will be stored on your computer.

Setting up SEER*Stat to use downloaded binary data

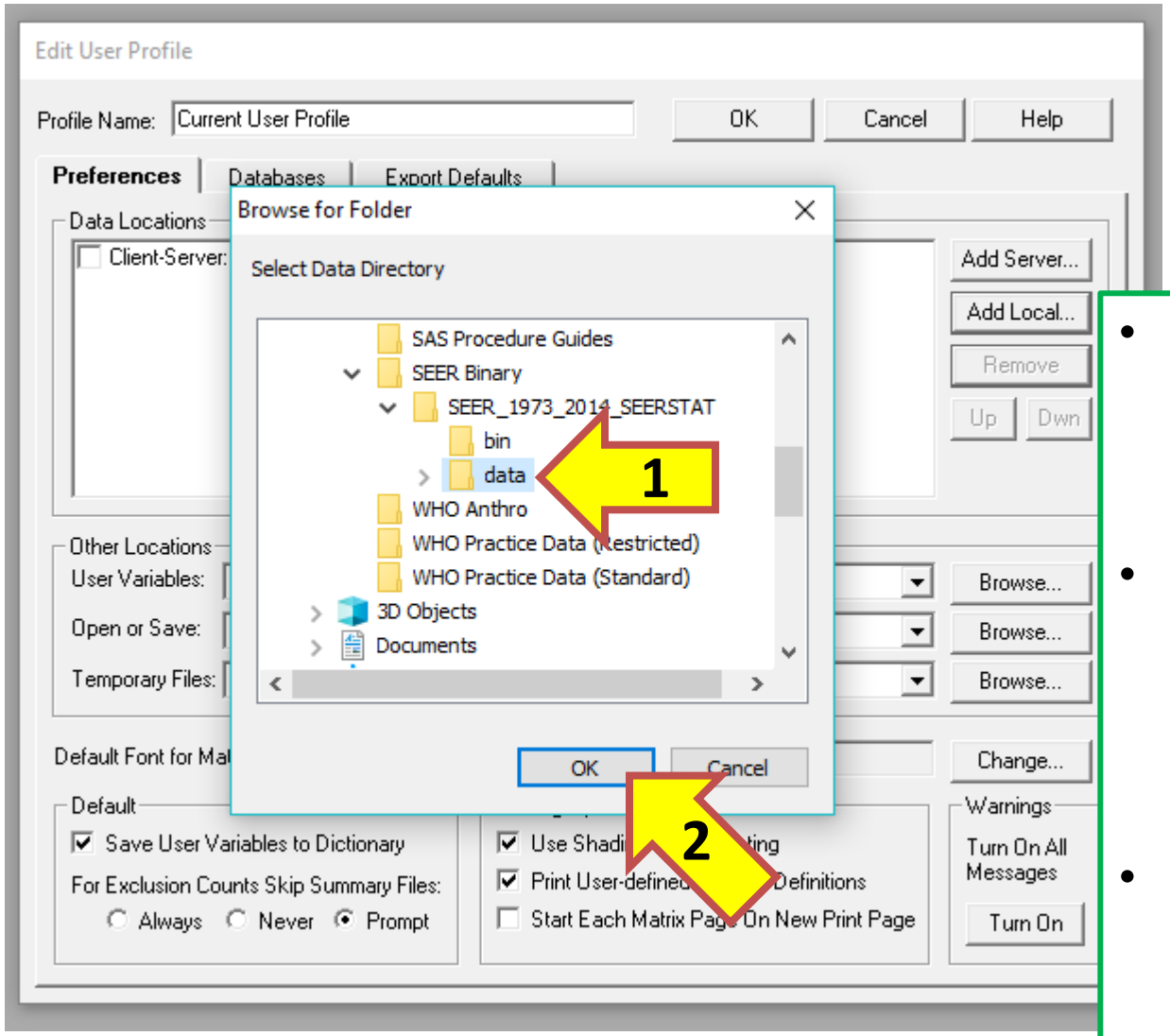
- **Open up your SEER*Stat software by double clicking on the SEER*Stat icon on your desktop.**



- Open SEER*Stat.
- To be able to utilize SEER*Stat software offline, you need to indicate where your binary SEER data files are stored on your computer.
- Click on “Profile” at the top, then “Preferences”.



- Uncheck the client server (unless you want to use SEER data from the internet, in which case you will need to provide your SEER username and password).
- Next click “Add Local…”.



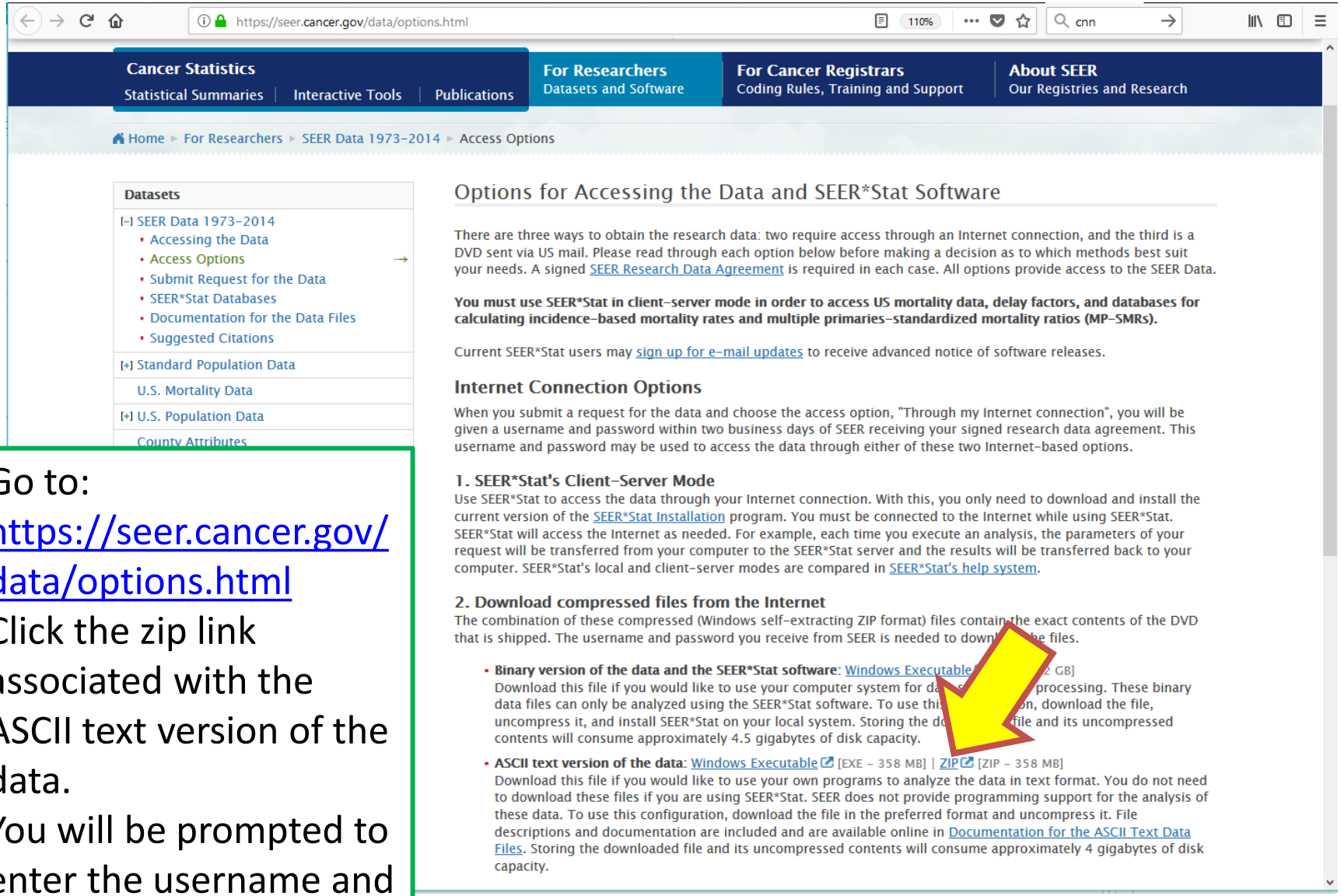
- Locate the SEER Binary data folder you downloaded from the SEER website.
- Select the “data” sub-folder inside the “SEER_1973_2014_SEERSTAT” folder and click “OK”.
- Then click “OK” to close the “Edit User Profile” window.
- Now you may run analyses offline.

Downloading ASCII SEER Data to use in SAS (and reading the file into SAS)

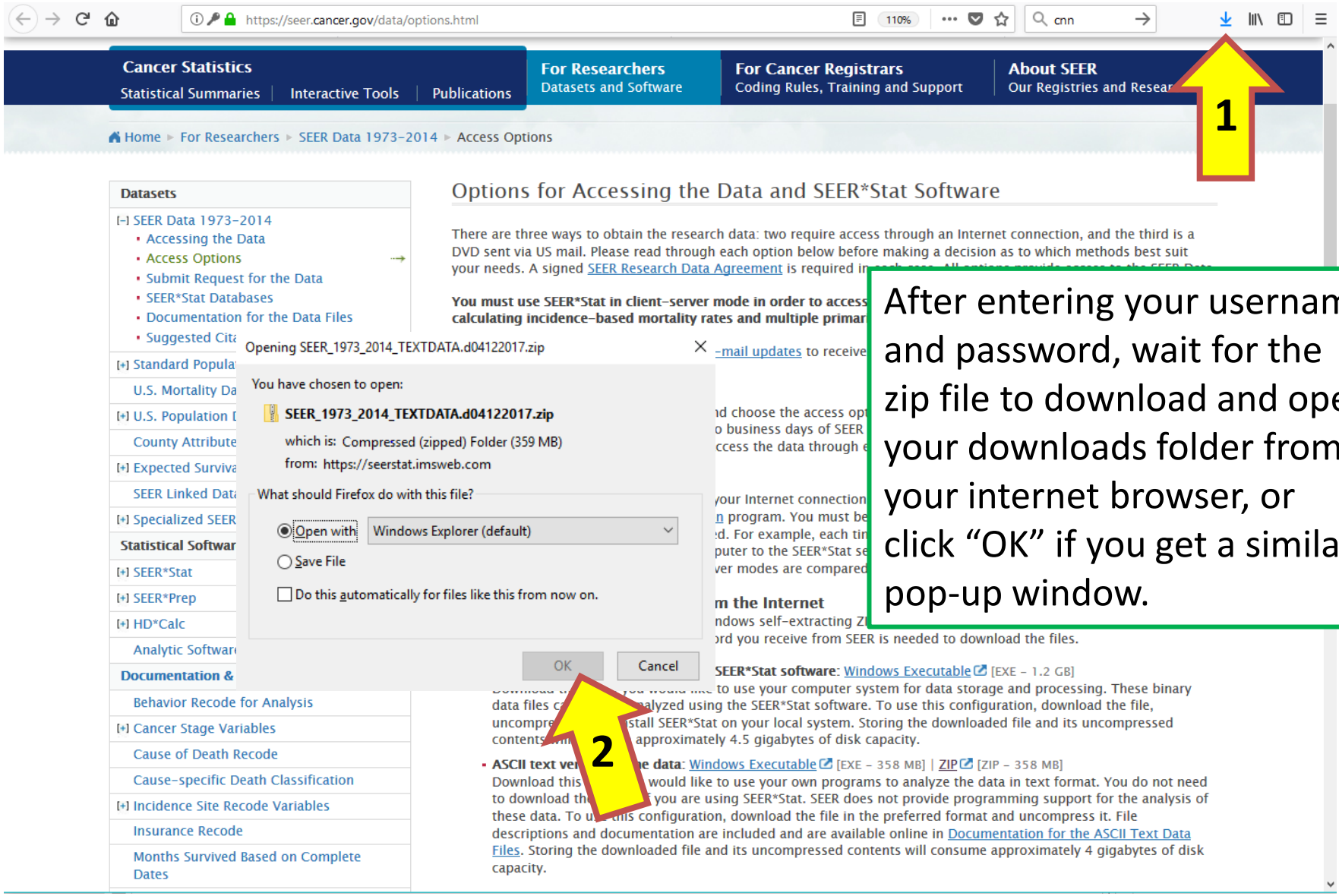
University of Nebraska Medical Center

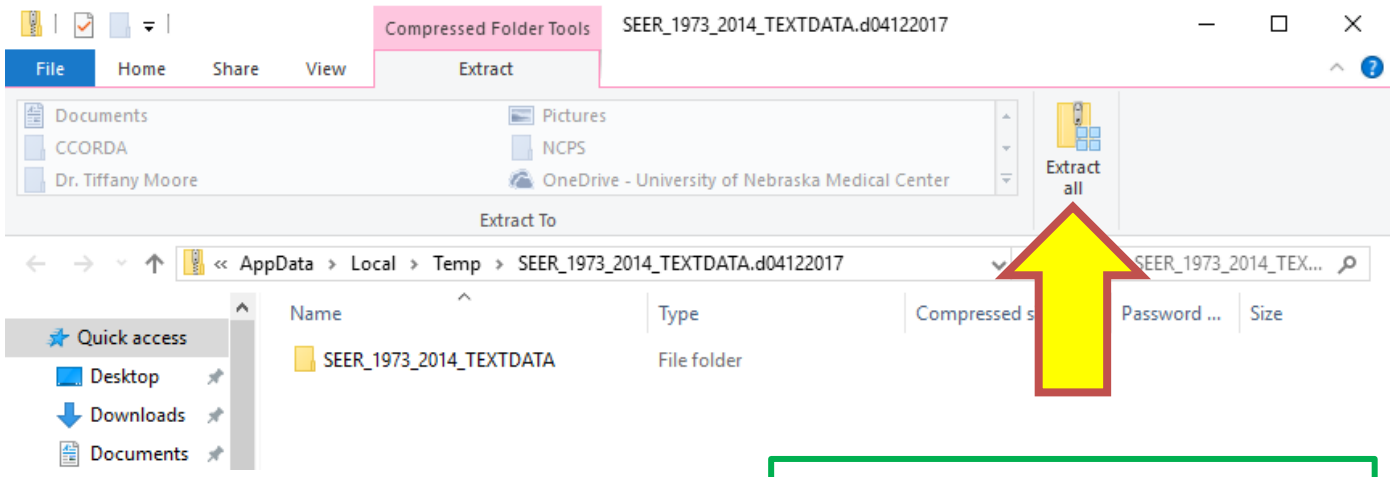
Downloading SEER Data to use in SAS

- This section will instruct you on how to download SEER data to be able to use in SAS.
- Not many people will use this option, as SEER*Stat is the most user-friendly way to access SEER data and calculate age-adjusted rates.
- Note: this ASCII data cannot be used in SEER*Stat; for that, you need to download the binary data (which is covered in the previous section).

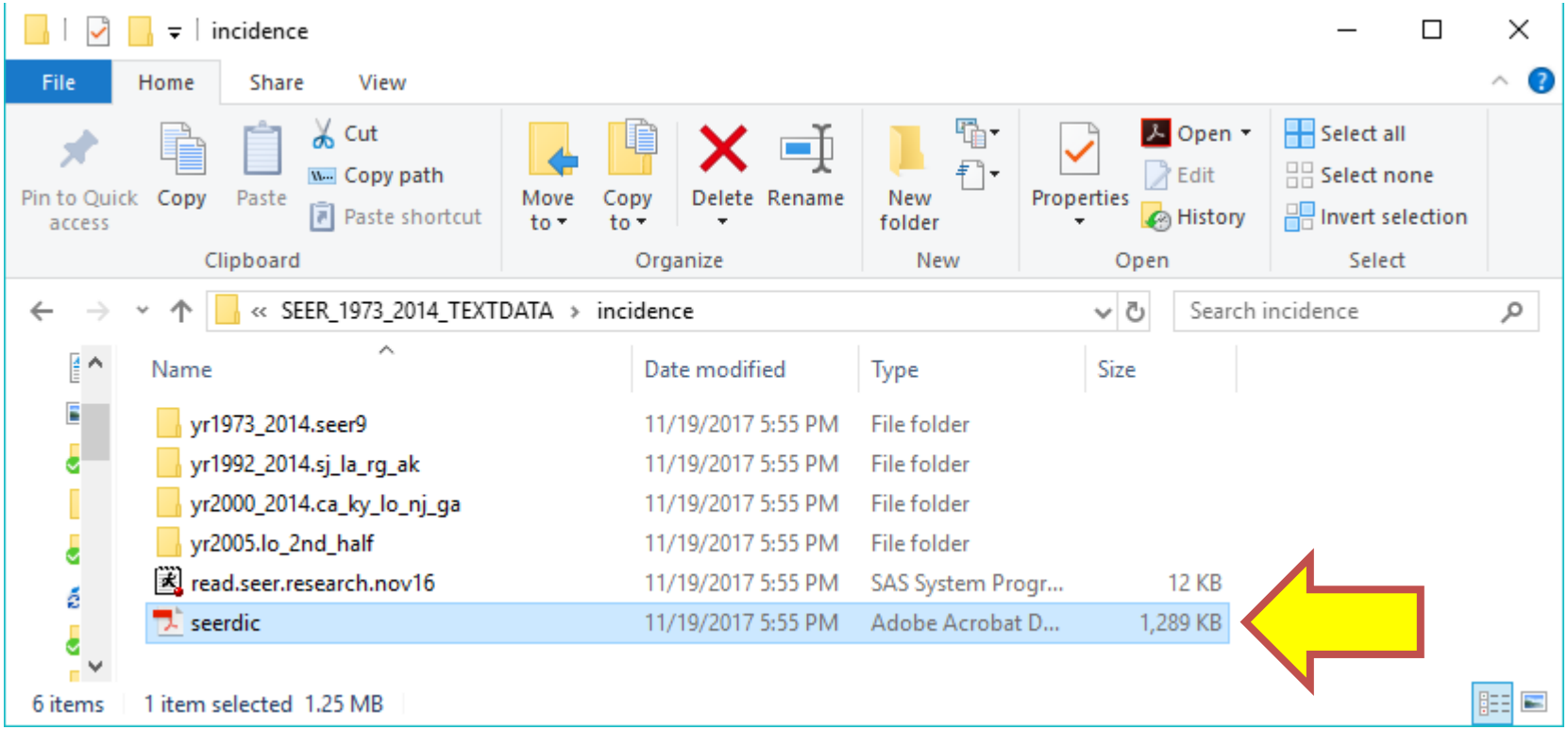


- Go to: https://seer.cancer.gov/data/options.html
- Click the zip link associated with the ASCII text version of the data.
- You will be prompted to enter the username and password SEER sent.


https://seer.cancer.gov/data/options.html
Cancer Statistics
Statistical Summaries | Interactive Tools | Publications
For Researchers
Datasets and Software
For Cancer Registrars
Coding Rules, Training and Support
About SEER
Our Registries and Resear
1
Home ▸ For Researchers ▸ SEER Data 1973–2014 ▸ Access Options
Datasets
[–] SEER Data 1973–2014
Accessing the Data
Access Options
Submit Request for the Data
SEER*Stat Databases
Documentation for the Data Files
Suggested Cita
[+] Standard Popula
U.S. Mortality Da
[+] U.S. Population D
County Attribute
[+] Expected Surviva
SEER Linked Data
[+] Specialized SEER
Statistical Softwar
[+] SEER*Stat
[+] SEER*Prep
[+] HD*Calc
Analytic Softwar
Documentation &
Behavior Recode for Analysis
[+] Cancer Stage Variables
Cause of Death Recode
Cause-specific Death Classification
[+] Incidence Site Recode Variables
Insurance Recode
Months Survived Based on Complete Dates
Options for Accessing the Data and SEER*Stat Software
There are three ways to obtain the research data: two require access through an Internet connection, and the third is a DVD sent via US mail. Please read through each option below before making a decision as to which methods best suit your needs. A signed SEER Research Data Agreement is required in
You must use SEER*Stat in client-server mode in order to access
calculating incidence-based mortality rates and multiple primar
Opening SEER_1973_2014_TEXTDATA.d04122017.zip
You have chosen to open:
SEER_1973_2014_TEXTDATA.d04122017.zip
which is: Compressed (zipped) Folder (359 MB)
from: https://seerstat.imsweb.com
What should Firefox do with this file?
Open with Windows Explorer (default)
Save File
Do this automatically for files like this from now on.
OK
Cancel
2
SEER*Stat software: Windows Executable [EXE – 1.2 GB]
ASCII text ver e data: Windows Executable [EXE – 358 MB] | ZIP [ZIP – 358 MB]
Download this would like to use your own programs to analyze the data in text format. You do not need to download th you are using SEER*Stat. SEER does not provide programming support for the analysis of these data. To u this configuration, download the file in the preferred format and uncompress it. File descriptions and documentation are included and are available online in Documentation for the ASCII Text Data Files. Storing the downloaded file and its uncompressed contents will consume approximately 4 gigabytes of disk capacity.
After entering your username and password, wait for the zip file to download and open your downloads folder from your internet browser, or click "OK" if you get a similar pop-up window.



- Extract the contents of the zipped folder by clicking the “Extract all” button.
- This will prompt you for the location where you would like to save the contents.



- This is what the contents of the incidence extracted folder looks like.
- Double click on the "seerdic" PDF, which is the documentation for this data.

MARITAL STATUS AT DX

NAACCR Item #: 150
SEER*Stat Name: Marital status at diagnosis
Item Length: 1

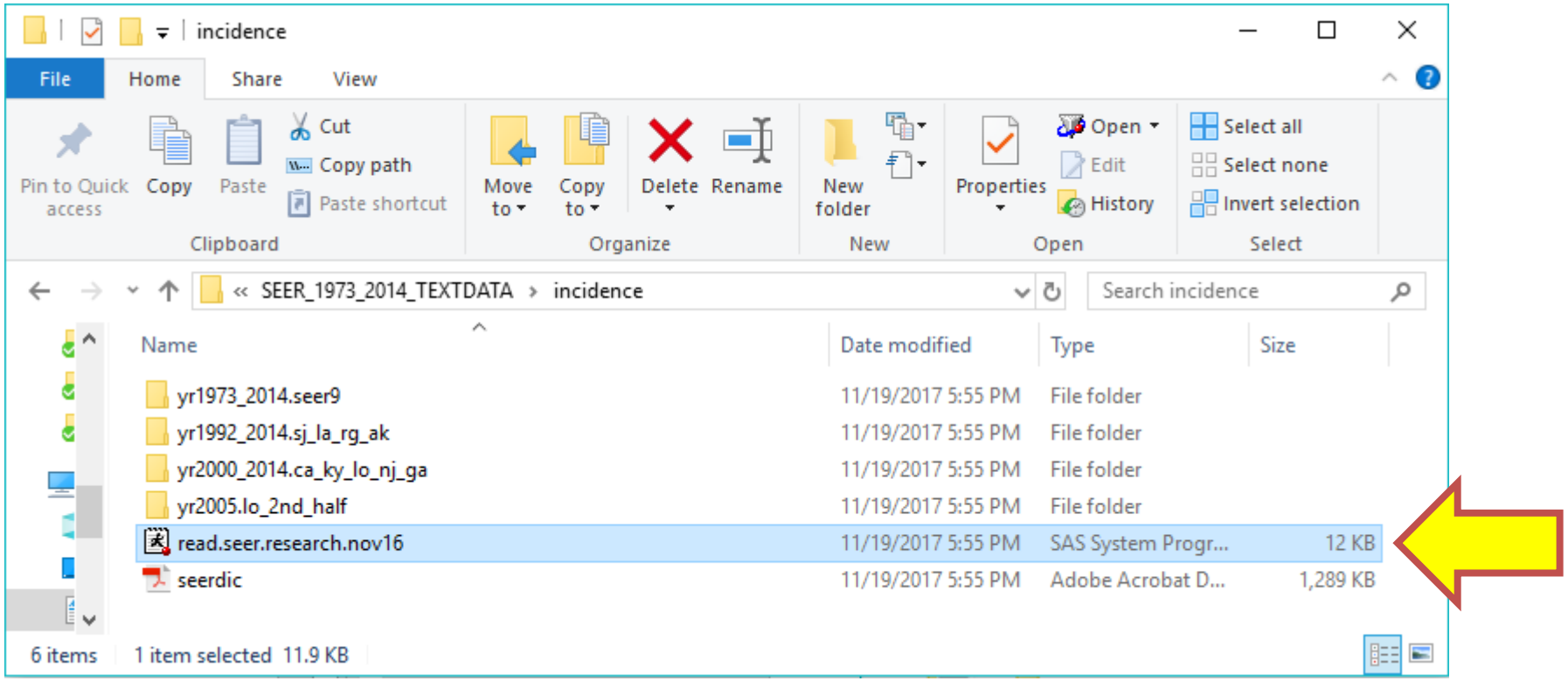
Field Description: This data item identifies the patient's marital status at the time of diagnosis for the reportable tumor.

Code	Description
1	Single (never married)
2	Married (including common law)
3	Separated
4	Divorced
5	Widowed
6	Unmarried or domestic partner (same sex or opposite sex or unregistered)
9	Unknown

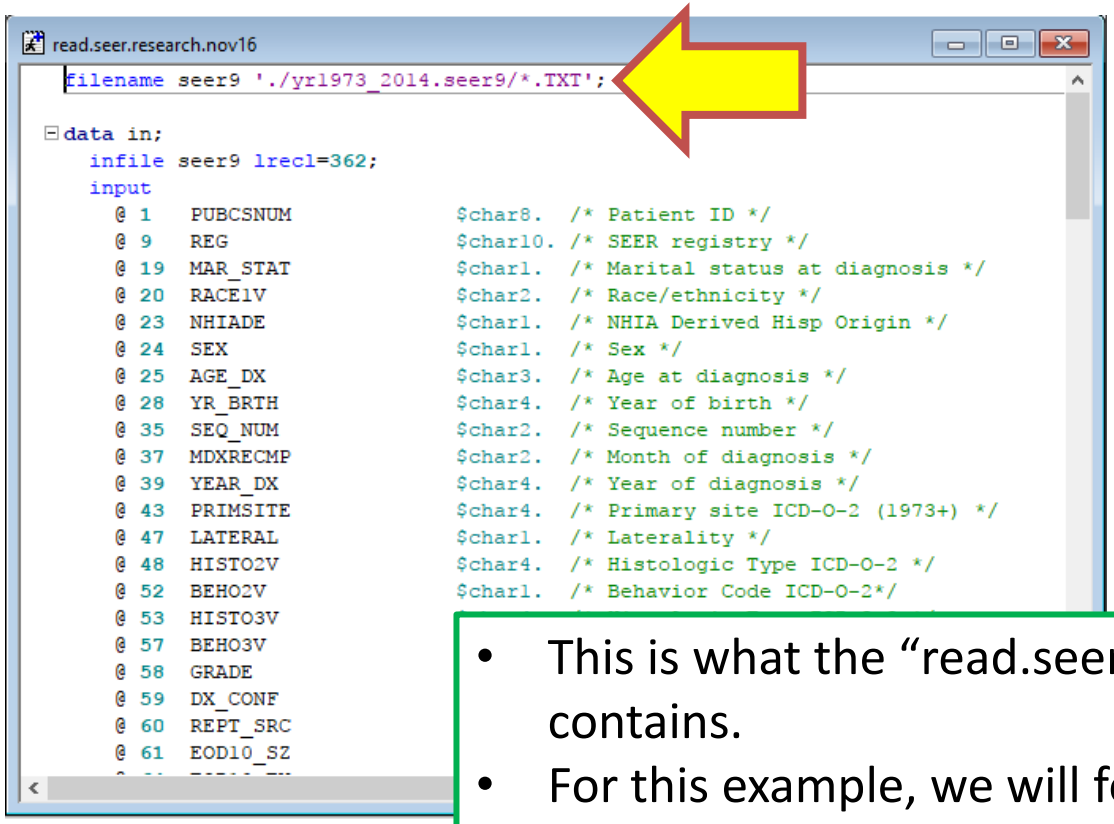
- This is an example of what is included in the documentation.
- Here, it tells you specific information about a specific variable.

Reading in SEER data using SAS

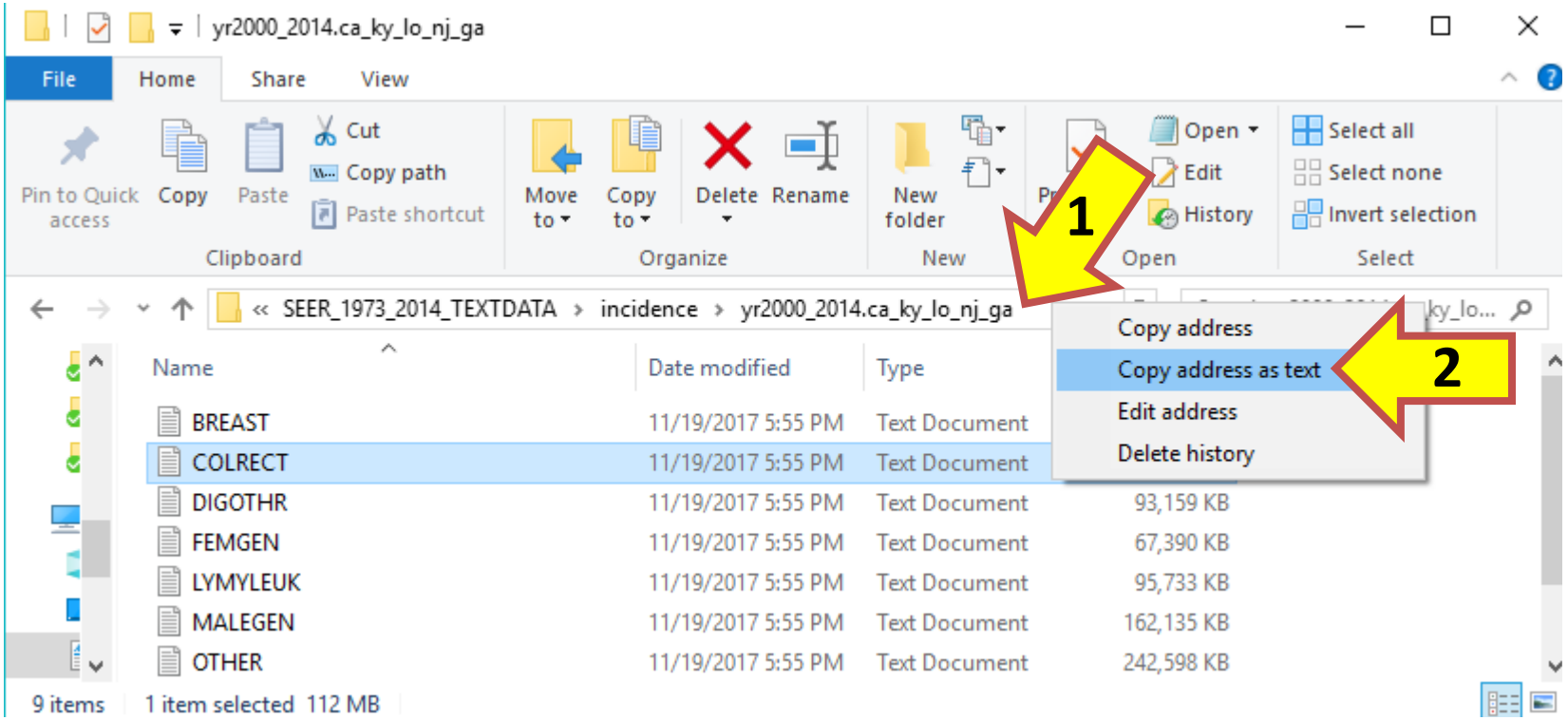
- **Example: Read in colorectal cancer incidence data from the "yr2000_2014.ca_ky_lo_nj_ga" data folder downloaded from the SEER website.**



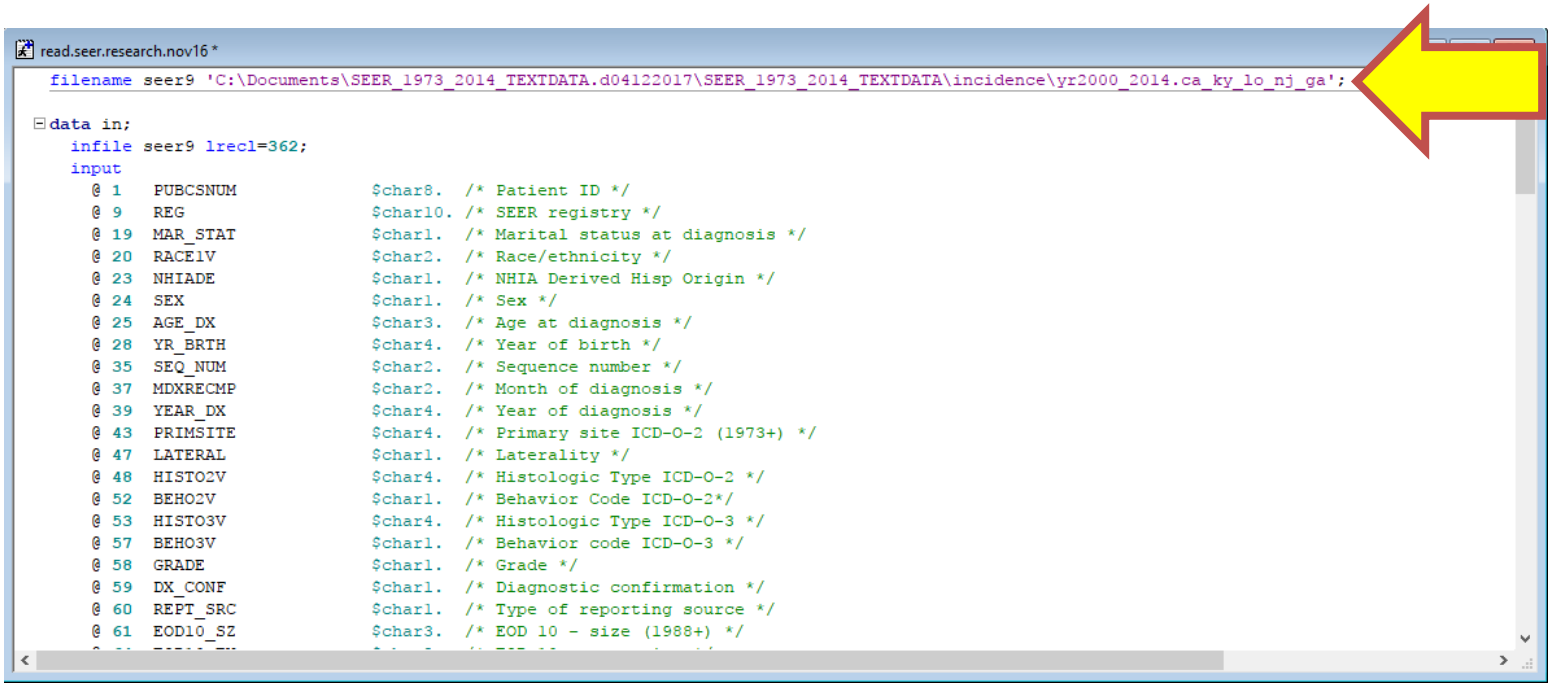
- The “read.seer.research.nov16” in the “incidence” folder is the SAS file you can use to read in the individual .txt data files.
- The .txt files are contained in the four folders listed above.
- Double click on the SAS file.



- This is what the “read.seer.research.nov16” SAS file contains.
- For this example, we will focus on the “yr2000_2014.ca_ky_lo_nj_ga” folder.
- The first thing we need to do is point to the correct data file using the appropriate path.
- We must change the purple text in the top line to indicate where the colorectal cancer text file from the “yr2000_2014.ca_ky_lo_nj_ga” folder is located, which we’ll do on the next slide.



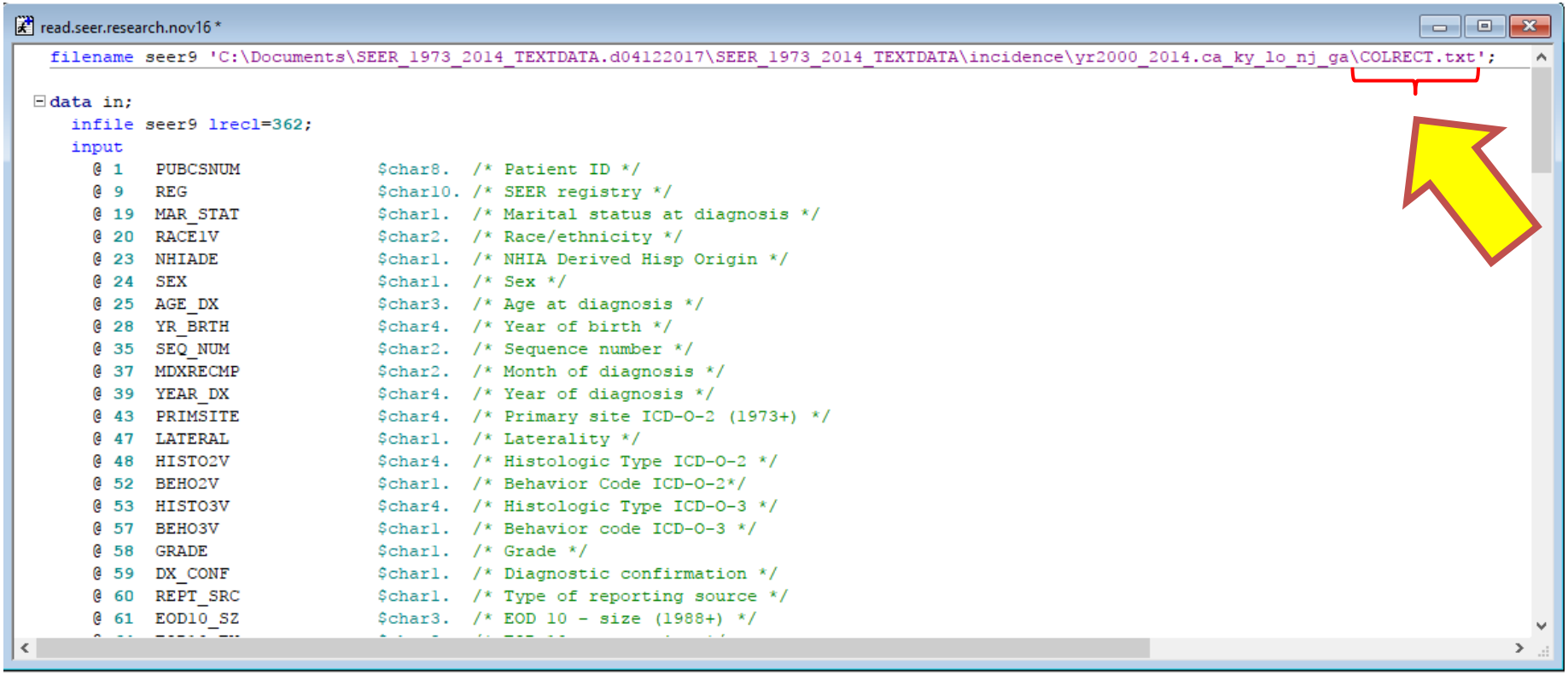
The easiest way to do this is to open the "yr2000_2014.ca_ky_lo_nj_ga" folder (which is contained in the "incidence" folder of the SEER ASCII data downloaded from the website), and then right click the path name at the top of the folder, and click "Copy address as text".



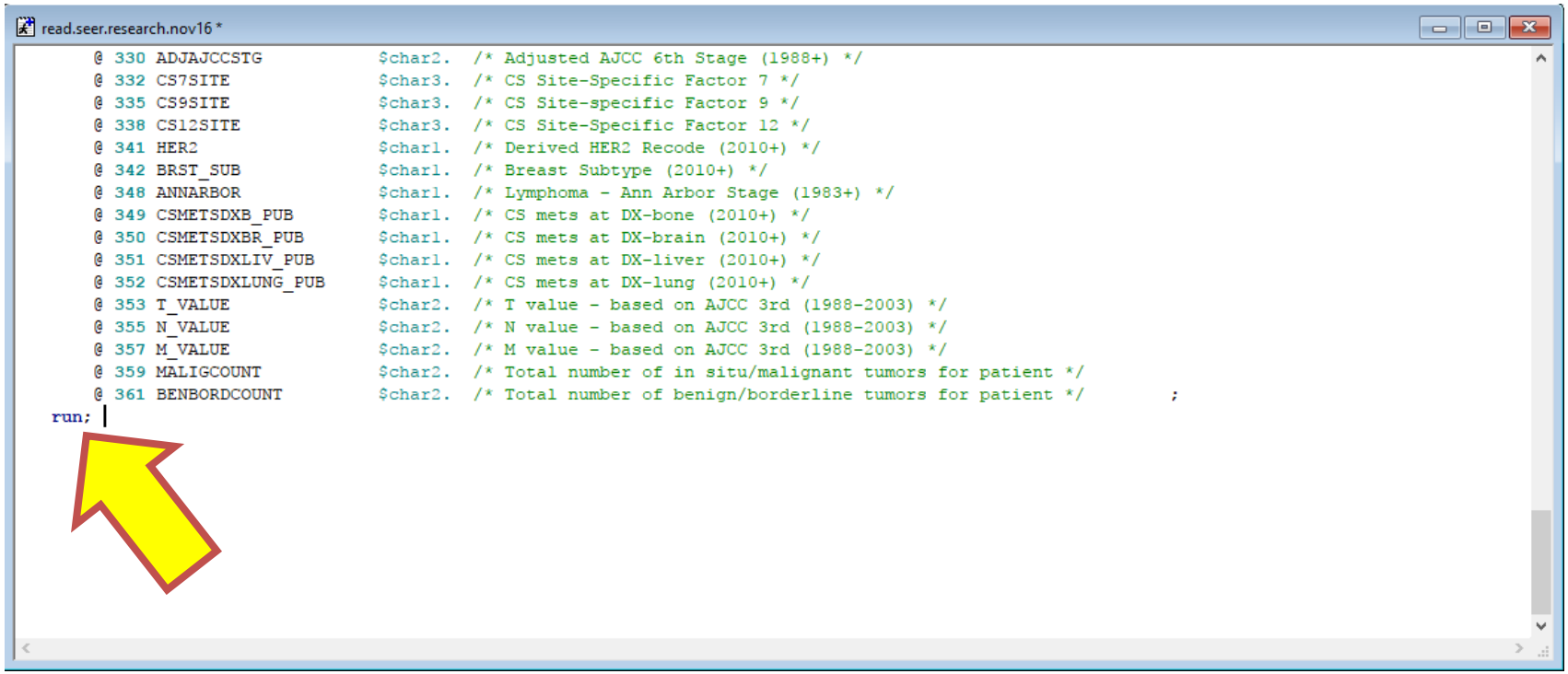
```
read.seer.research.nov16 *
filename seer9 'C:\Documents\SEER_1973_2014_TEXTDATA.d04122017\SEER_1973_2014_TEXTDATA\incidence\yr2000_2014.ca_ky_lo_nj_ga';

data in;
   infile seer9 lrecl=362;
   input
     @ 1   PUBCSNUM            $char8.  /* Patient ID */
     @ 9   REG                 $char10. /* SEER registry */
     @ 19  MAR_STAT            $char1.  /* Marital status at diagnosis */
     @ 20  RACE1V              $char2.  /* Race/ethnicity */
     @ 23  NHIADE              $char1.  /* NHIA Derived Hisp Origin */
     @ 24  SEX                 $char1.  /* Sex */
     @ 25  AGE_DX              $char3.  /* Age at diagnosis */
     @ 28  YR_BRTH             $char4.  /* Year of birth */
     @ 35  SEQ_NUM             $char2.  /* Sequence number */
     @ 37  MDXRECMP            $char2.  /* Month of diagnosis */
     @ 39  YEAR_DX             $char4.  /* Year of diagnosis */
     @ 43  PRIMSITE            $char4.  /* Primary site ICD-O-2 (1973+) */
     @ 47  LATERAL             $char1.  /* Laterality */
     @ 48  HISTO2V             $char4.  /* Histologic Type ICD-O-2 */
     @ 52  BEHO2V              $char1.  /* Behavior Code ICD-O-2*/
     @ 53  HISTO3V             $char4.  /* Histologic Type ICD-O-3 */
     @ 57  BEHO3V              $char1.  /* Behavior code ICD-O-3 */
     @ 58  GRADE               $char1.  /* Grade */
     @ 59  DX_CONF             $char1.  /* Diagnostic confirmation */
     @ 60  REPT_SRC            $char1.  /* Type of reporting source */
     @ 61  EOD10_SZ            $char3.  /* EOD 10 - size (1988+) */
```

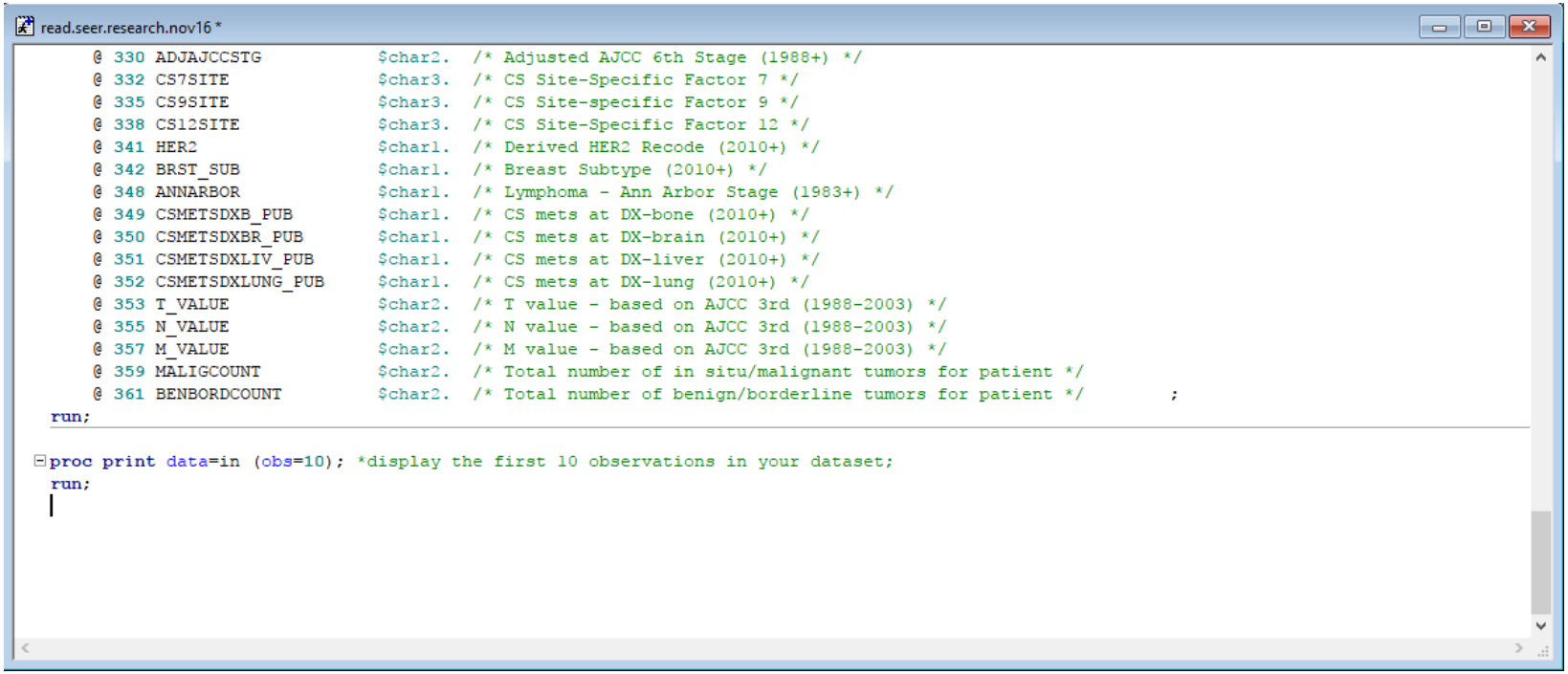
- This is what the path for my “yr2000_2014.ca_ky_lo_nj_ga” folder looks like.
- Yours will likely differ depending on where you saved the SEER data on your computer.
- Next we need to specify the specific .txt file we are interested in as part of the filename.



To specify the .txt file with data for colorectal cancers, I added the name of the .txt file with a backslash and the extension, "\COLRECT.txt", to the end of the address for the "yr2000_2014.ca_ky_lo_nj_ga" folder.

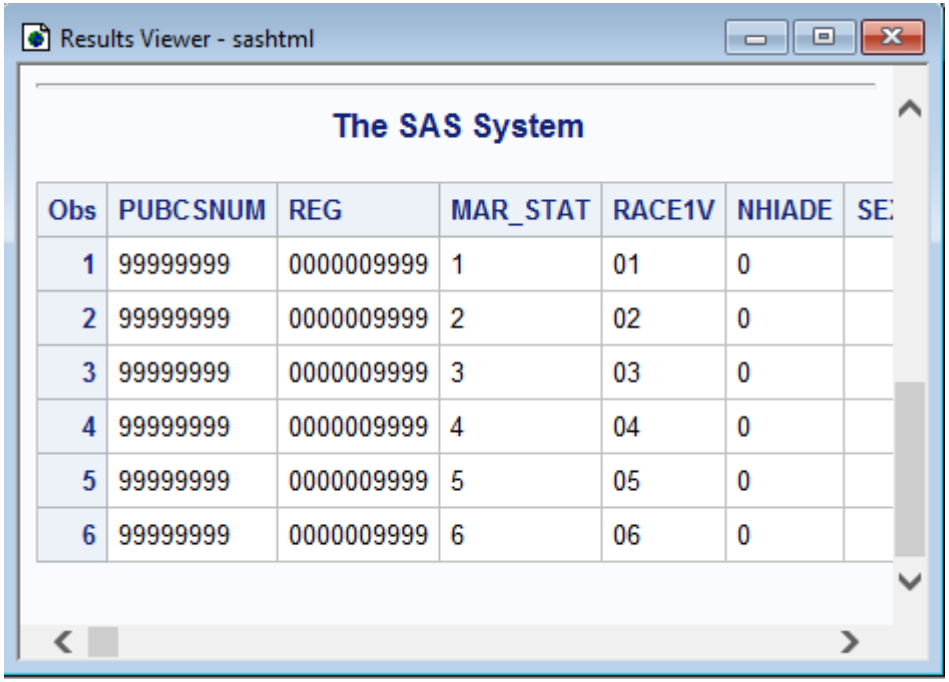


- To create a SAS dataset, you will need to add “run;” to the bottom of the code provided by SEER.
- Now you can highlight all of the SAS code in the file and run the program.
- It will read in the colorectal text file and name the new dataset created from that text file “in”.



To see the first 10 observations in your dataset, type in the following code:

```
proc print data=in (obs=10); *display the first 10
observations in your dataset;
run;
```



Obs	PUBCSNUM	REG	MAR_STAT	RACE1V	NHIADE	SE
1	99999999	0000009999	1	01	0	
2	99999999	0000009999	2	02	0	
3	99999999	0000009999	3	03	0	
4	99999999	0000009999	4	04	0	
5	99999999	0000009999	5	05	0	
6	99999999	0000009999	6	06	0	

- This is an example of what the data will look like.
- The data has been de-identified, but gives you an idea what the data structure looks like.
- From here, you can do many types of analyses using SAS software procedures.

Incidence Counts

Example: Breast Cancer

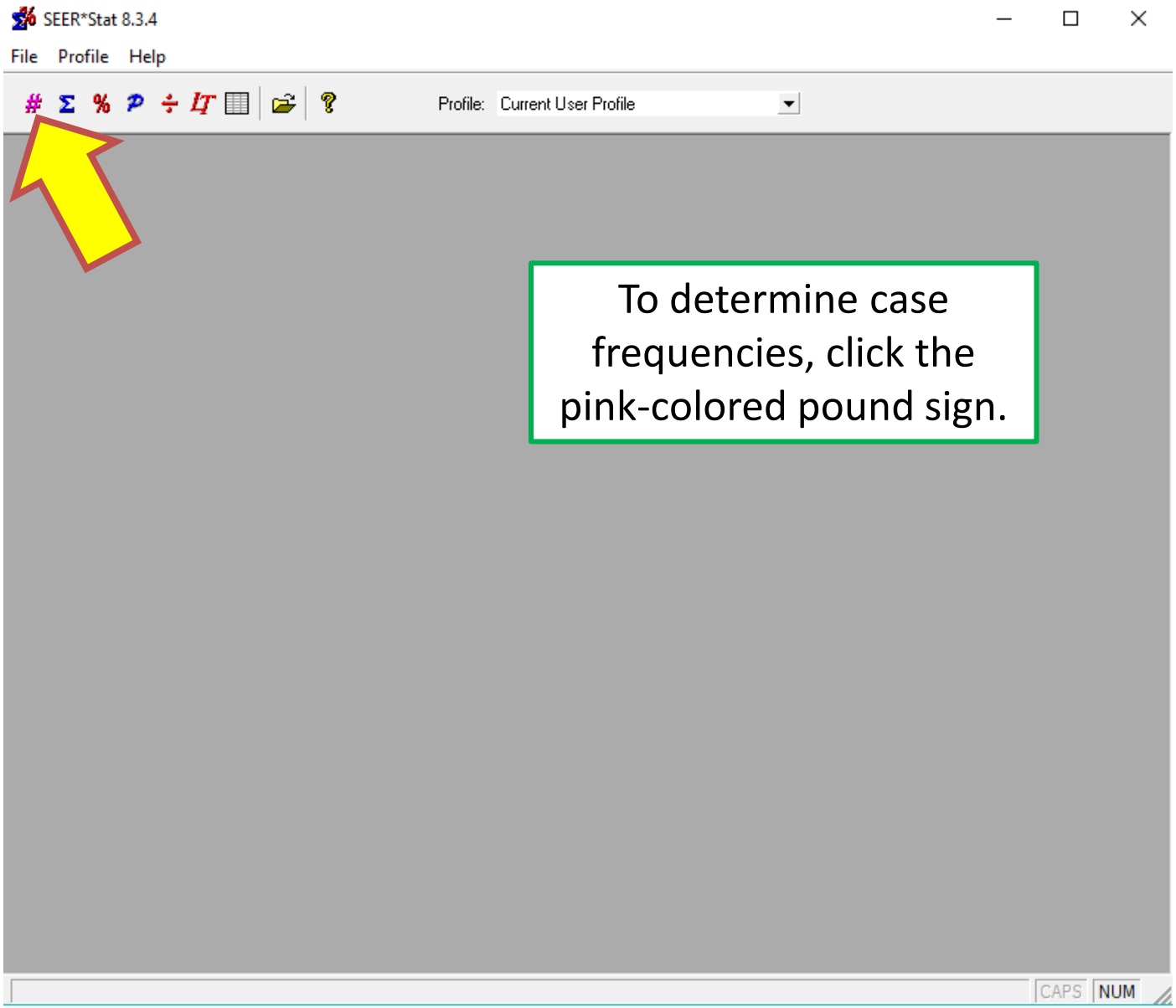
University of Nebraska Medical Center

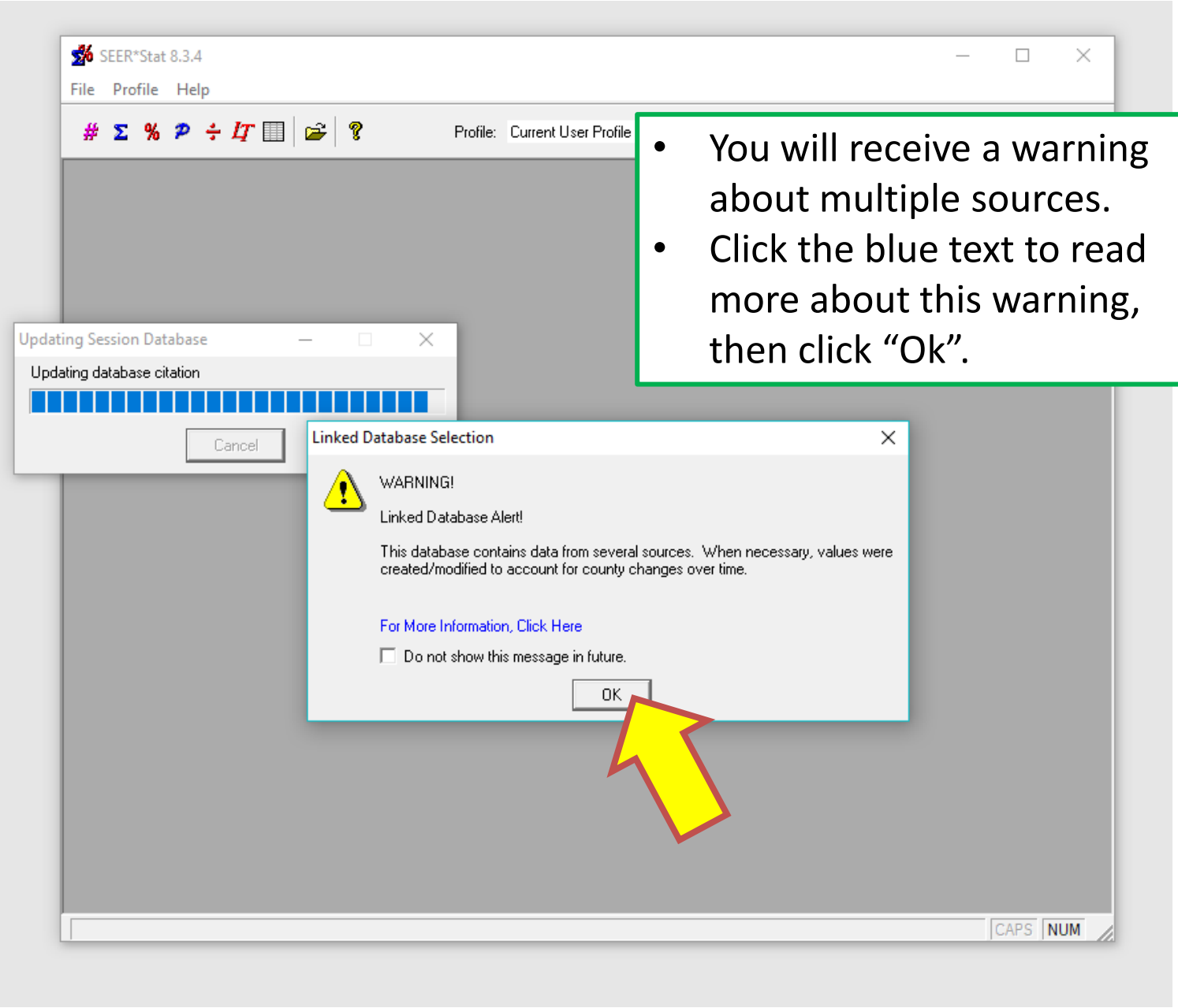
Incidence Counts

- This section will instruct you on how to get incidence counts using SEER data

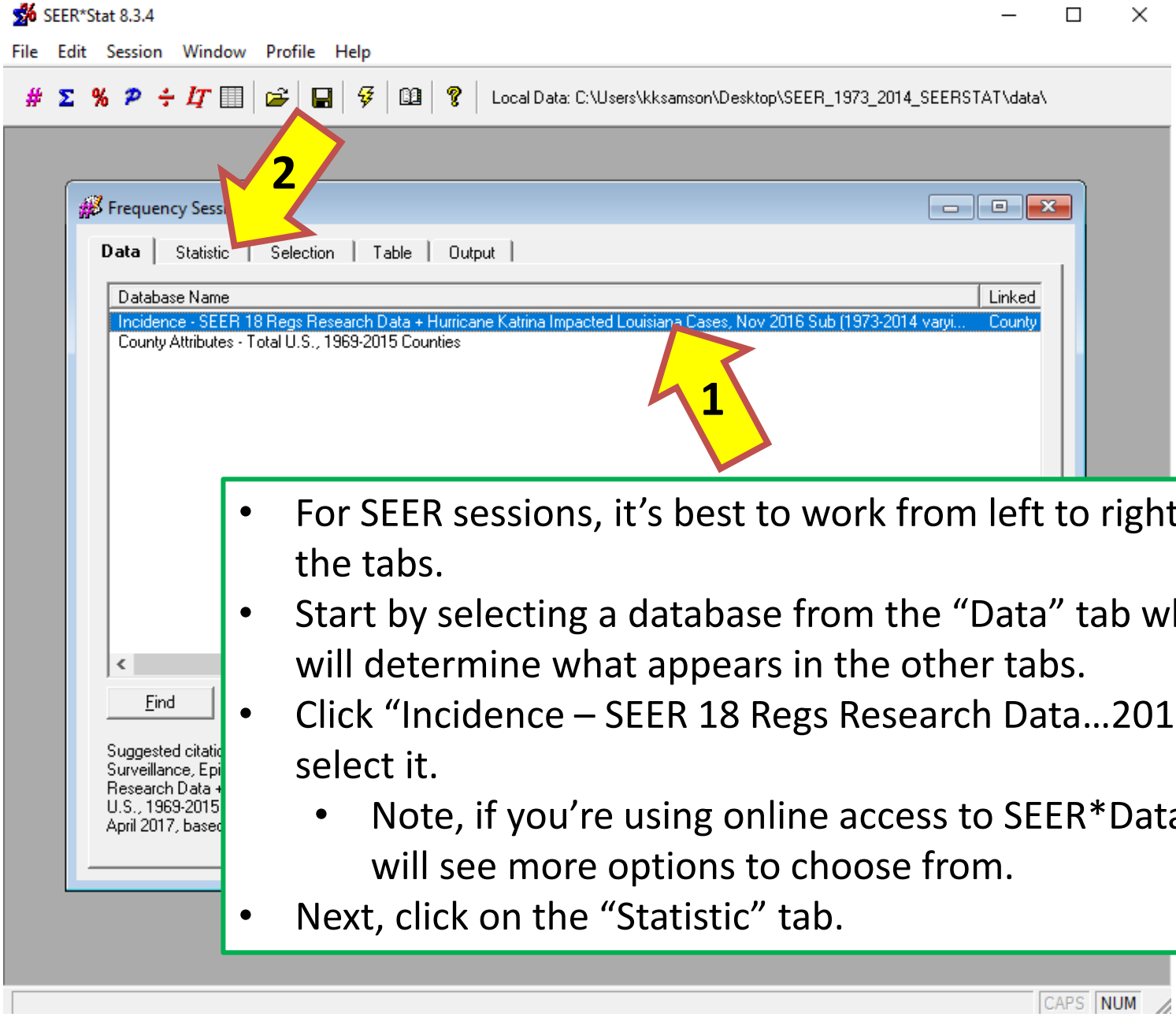
- For this example, we will show you to get access to breast cancer incidence counts using the 2016 SEER 18 registries data, with the following criteria:
  - Exclude anyone under the age of 15 years old
  - Display data by 5-year age categories
  - Display data separately by race

- This example was done using downloaded binary SEER data for use in SEER*Stat. If you are using online data instead (i.e. entering your username and password), a couple of your screens and output may be slightly different from the tutorial.

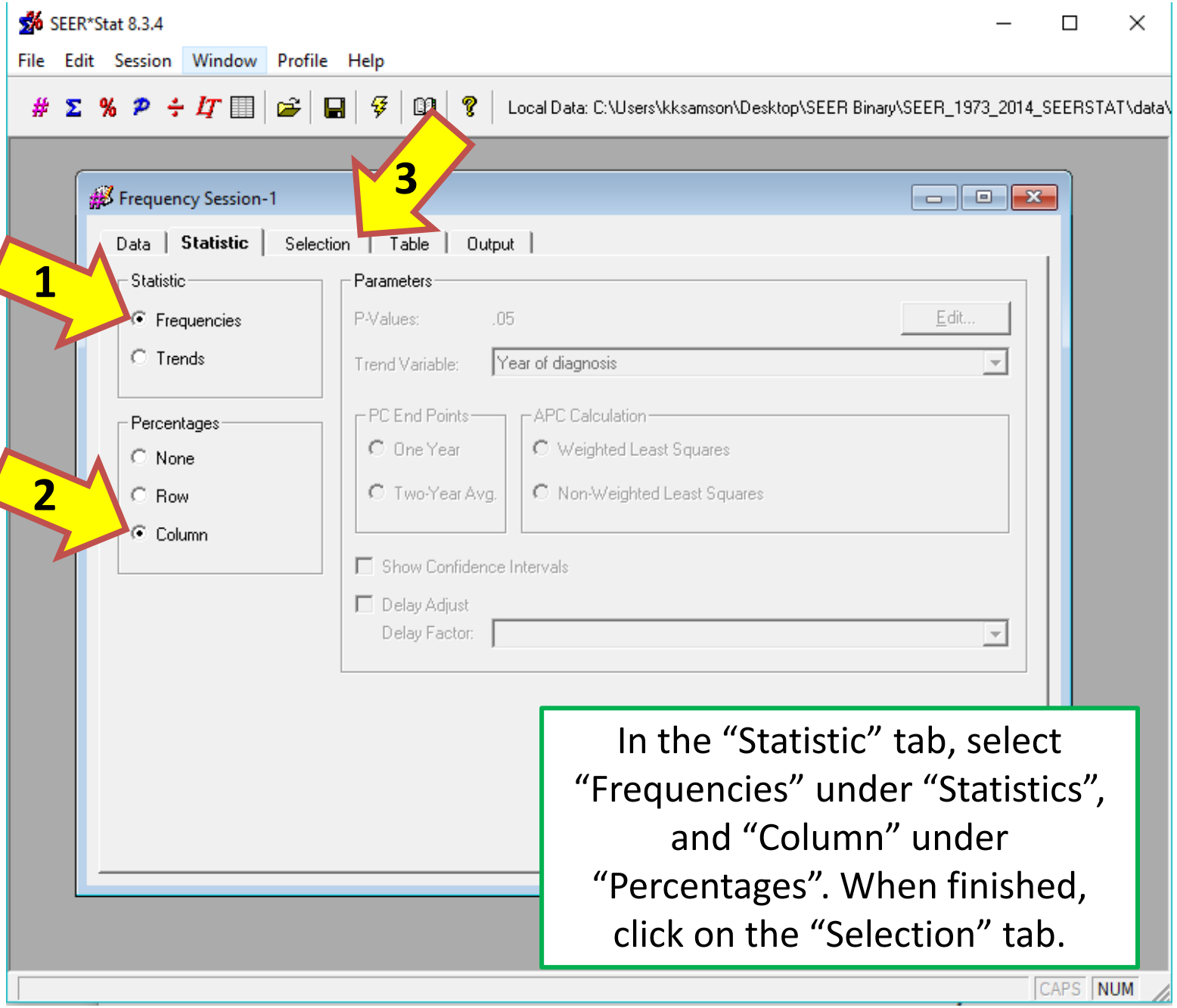

SEER*Stat 8.3.4
File
Profile
Help
Profile:
Current User Profile
To determine case frequencies, click the pink-colored pound sign.
CAPS
NUM

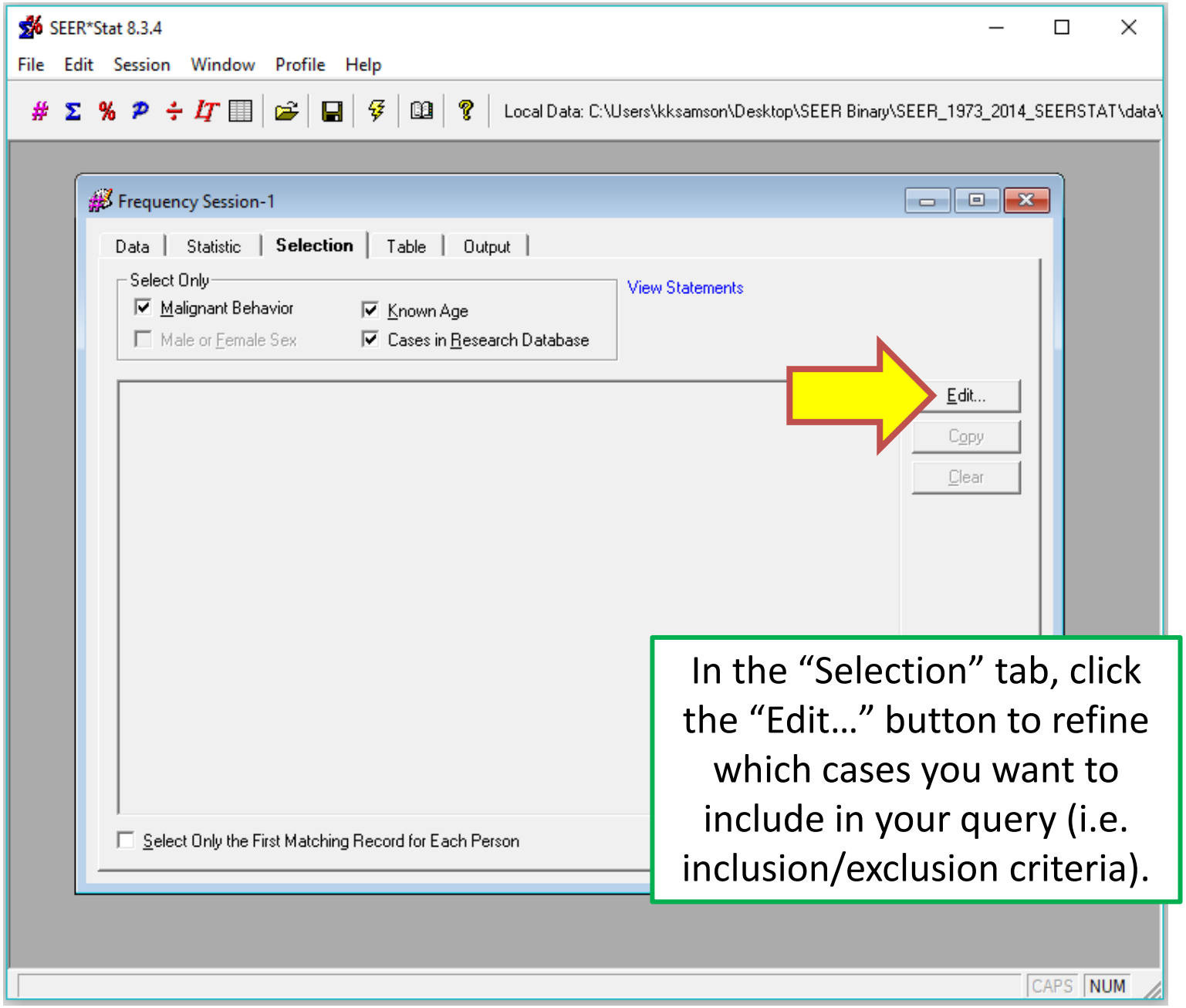


- You will receive a warning about multiple sources.
- Click the blue text to read more about this warning, then click “Ok”.

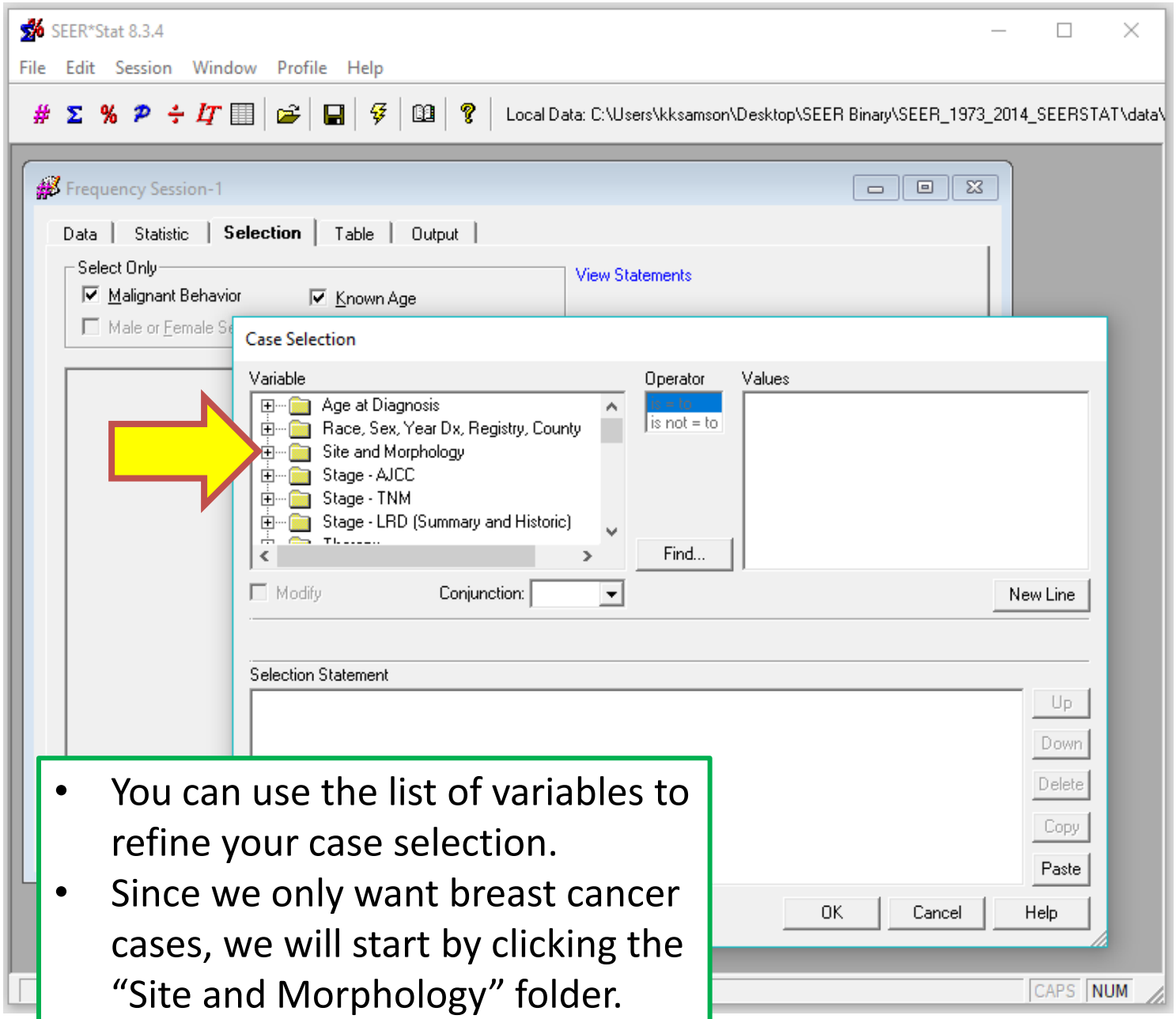


- For SEER sessions, it's best to work from left to right with the tabs.
- Start by selecting a database from the "Data" tab which will determine what appears in the other tabs.
- Click "Incidence – SEER 18 Regs Research Data…2016" to select it.
  - Note, if you're using online access to SEER*Data, you will see more options to choose from.
- Next, click on the "Statistic" tab.

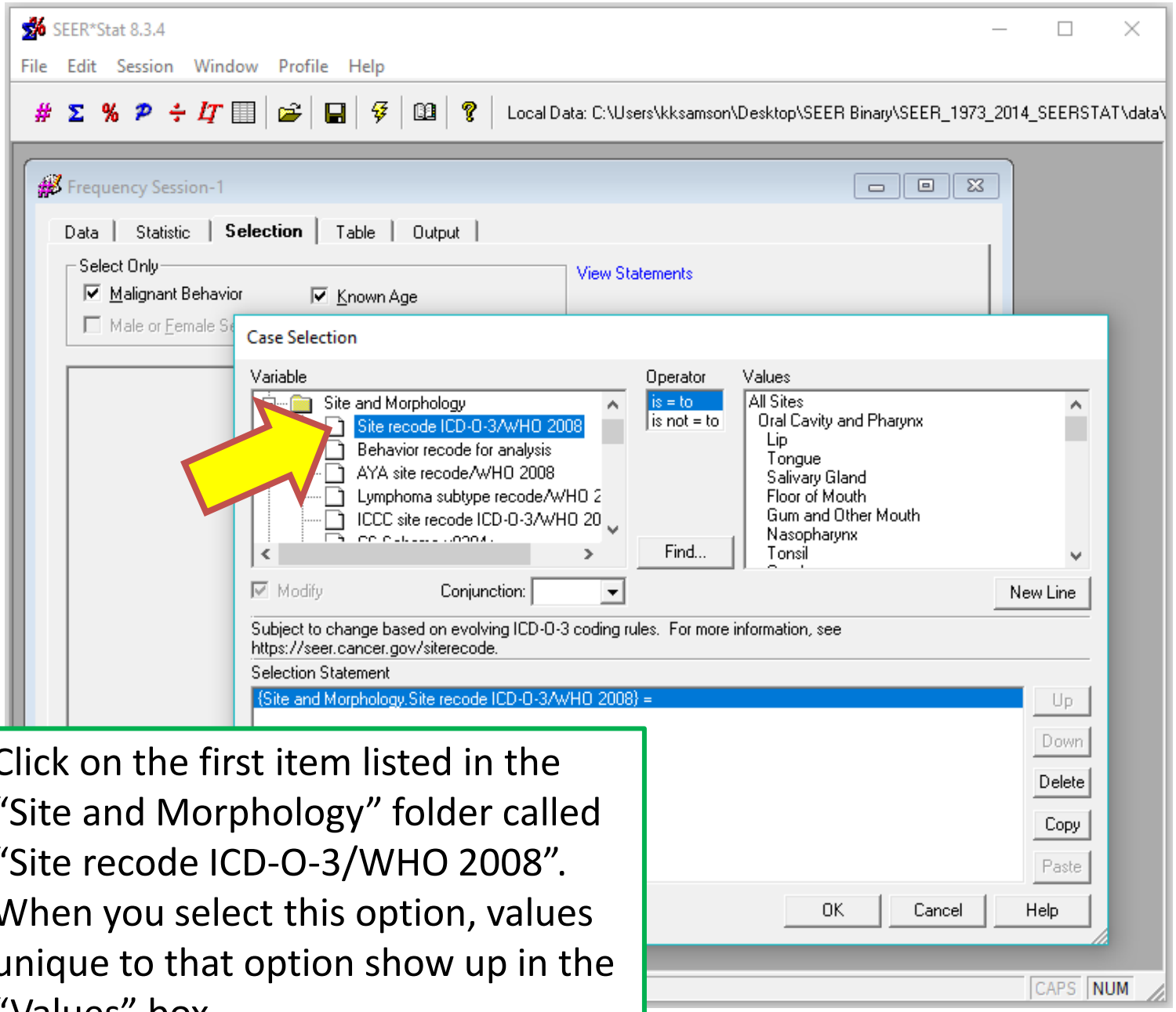

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER Binary\SEER_1973_2014_SEERSTAT\data\
Frequency Session-1
Data Statistic Selection Table Output
1
2
3
Statistic
Frequencies
Trends
Percentages
None
Row
Column
Parameters
P-Values: .05
Edit...
Trend Variable: Year of diagnosis
PC End Points
One Year
Two-Year Avg.
APC Calculation
Weighted Least Squares
Non-Weighted Least Squares
Show Confidence Intervals
Delay Adjust
Delay Factor:
In the “Statistic” tab, select “Frequencies” under “Statistics”, and “Column” under “Percentages”. When finished, click on the “Selection” tab.
CAPS NUM



In the "Selection" tab, click the "Edit…" button to refine which cases you want to include in your query (i.e. inclusion/exclusion criteria).

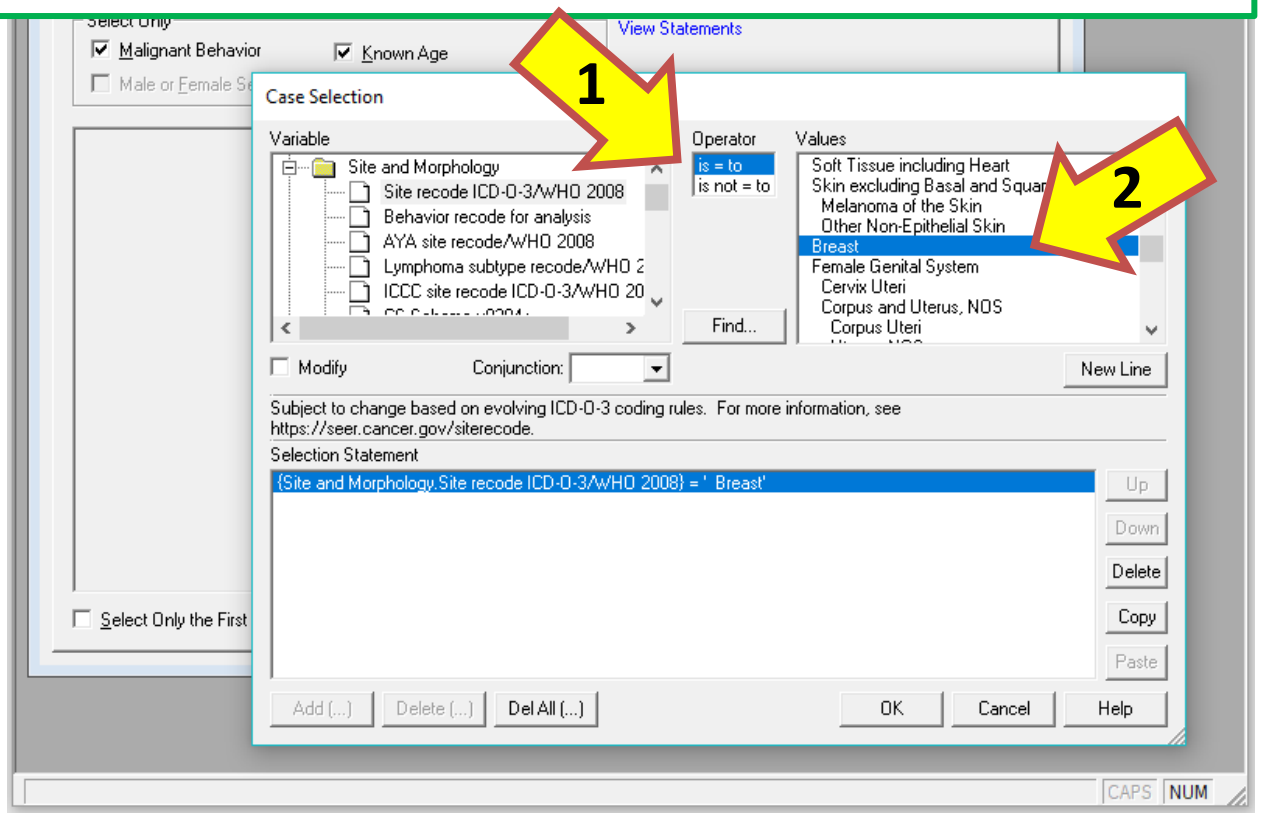


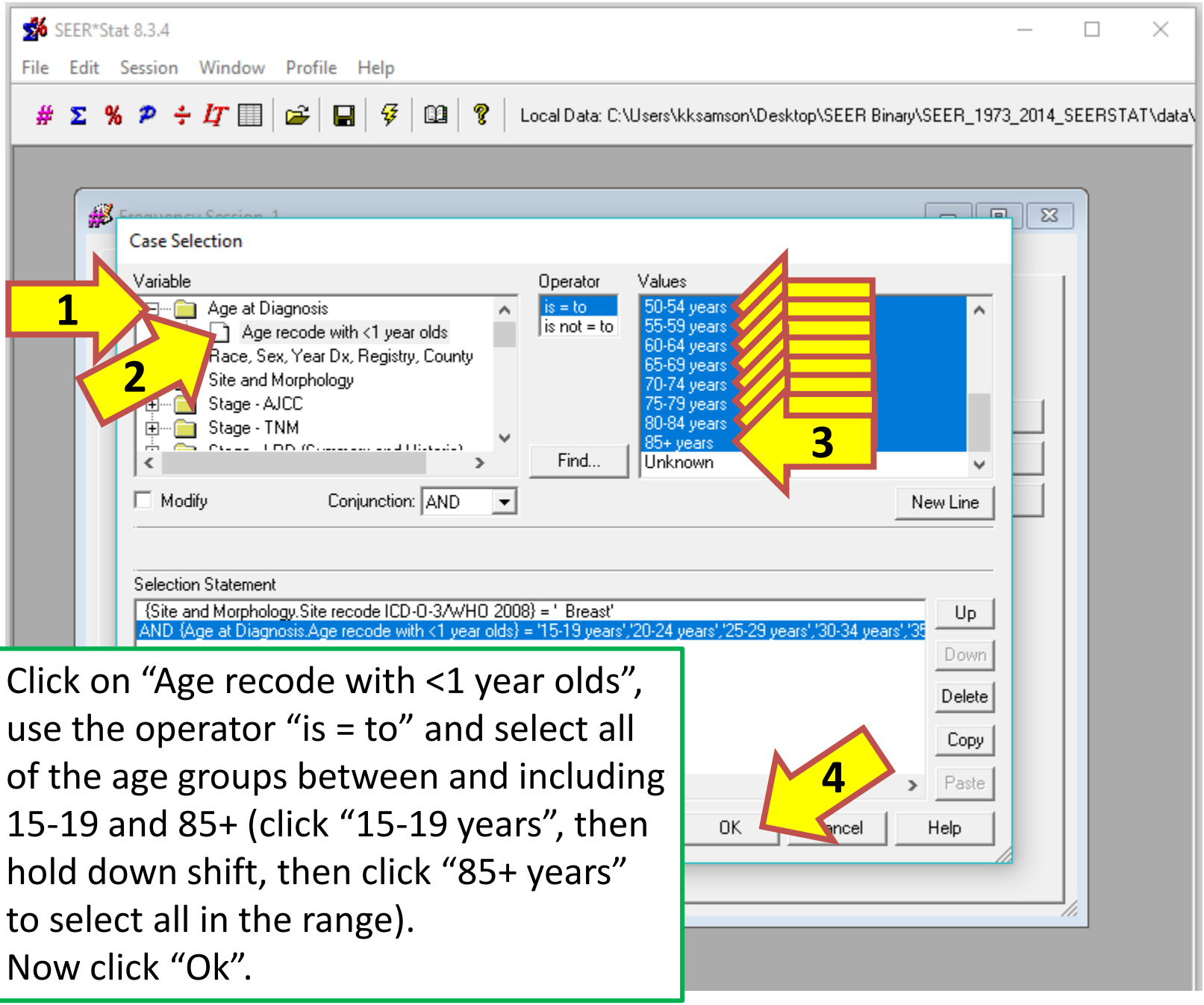
- You can use the list of variables to refine your case selection.
- Since we only want breast cancer cases, we will start by clicking the "Site and Morphology" folder.



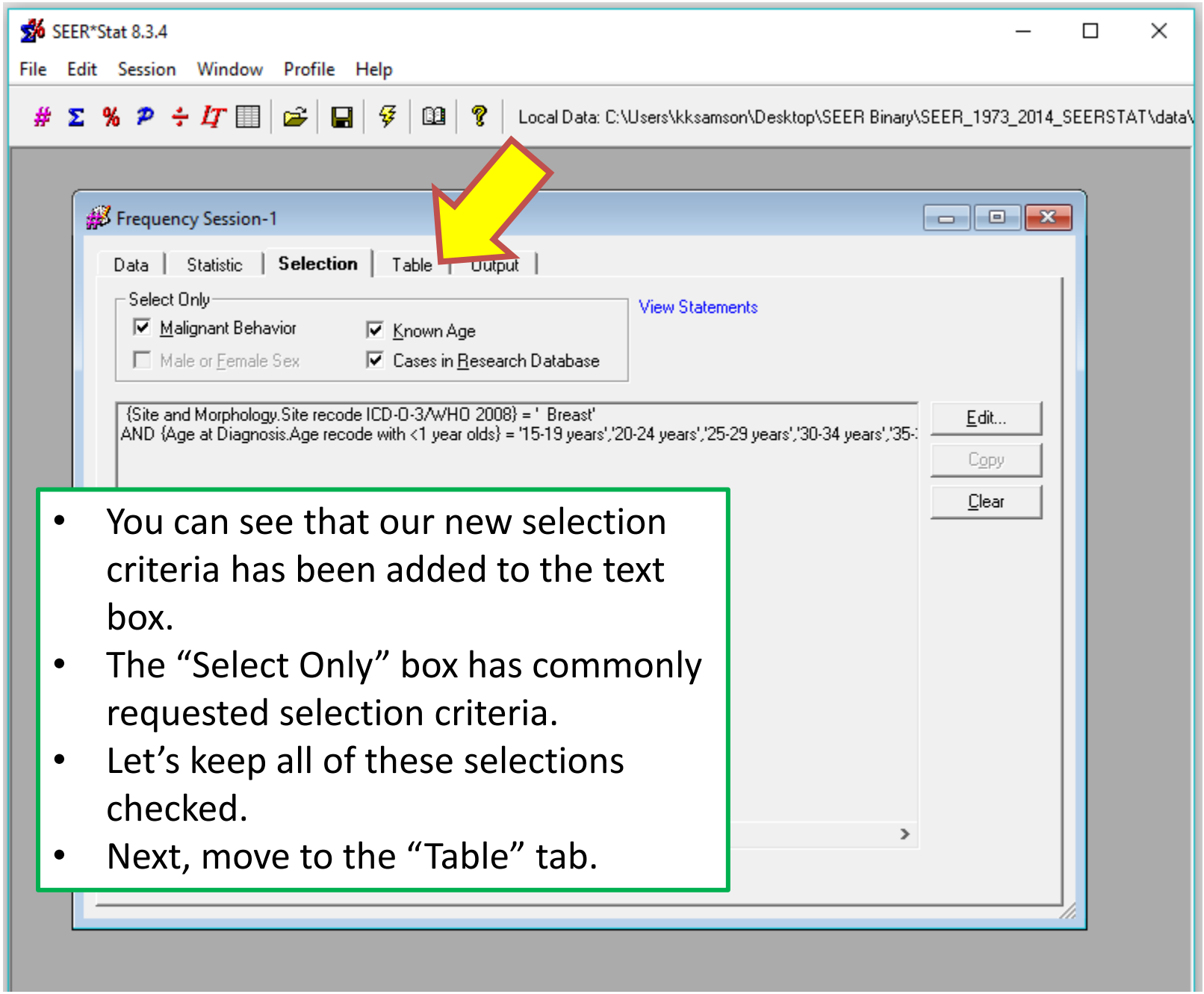
- Click on the first item listed in the "Site and Morphology" folder called "Site recode ICD-O-3/WHO 2008".
- When you select this option, values unique to that option show up in the "Values" box.

- Ensure that the “Operator” value “is = to” is selected.
- In the “Values” box, find and select “Breast”.
- In the “Selection Statement” box below, you will see that code was automatically generated to restrict sites to those with values of “Breast”.
- FYI: if you had chosen “is not = to” in the “Operator” box, then your data would have all cancer cases that are not Breast cancer.
- We will now add additional selection criteria by clicking on the “Age at Diagnosis” folder in the “Variable” box.

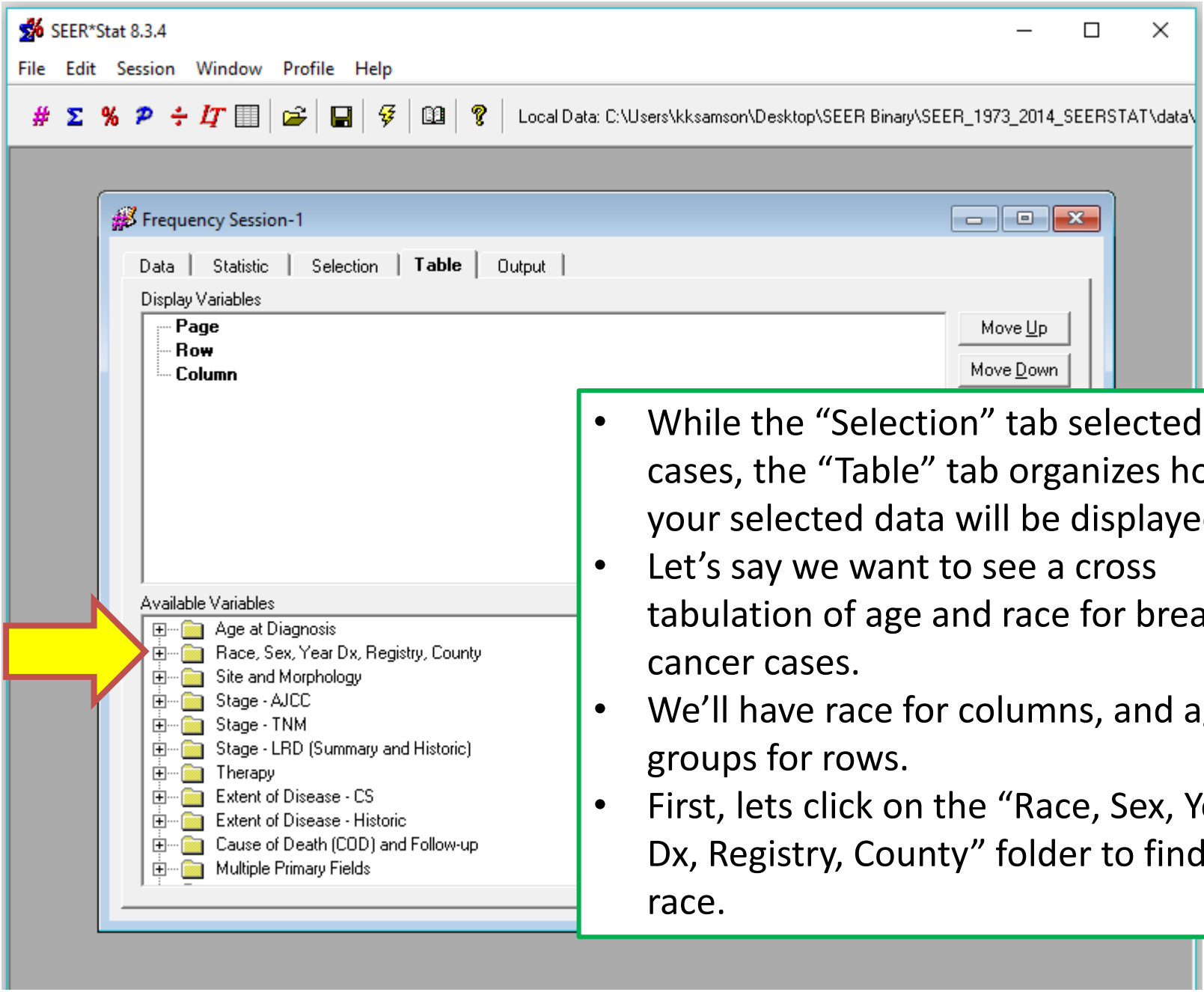




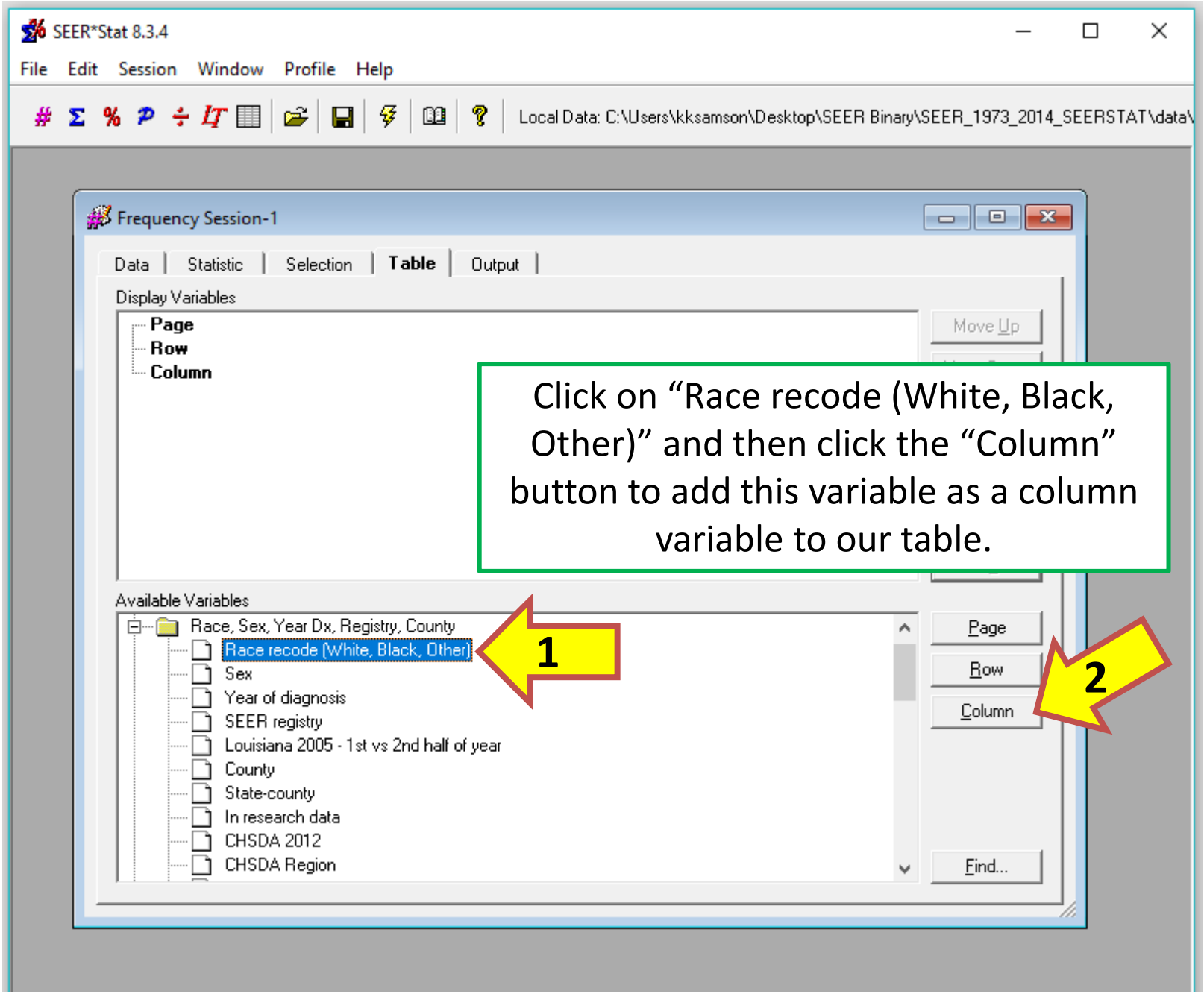
- Click on “Age recode with <1 year olds”, use the operator “is = to” and select all of the age groups between and including 15-19 and 85+ (click “15-19 years”, then hold down shift, then click “85+ years” to select all in the range).
- Now click “Ok”.

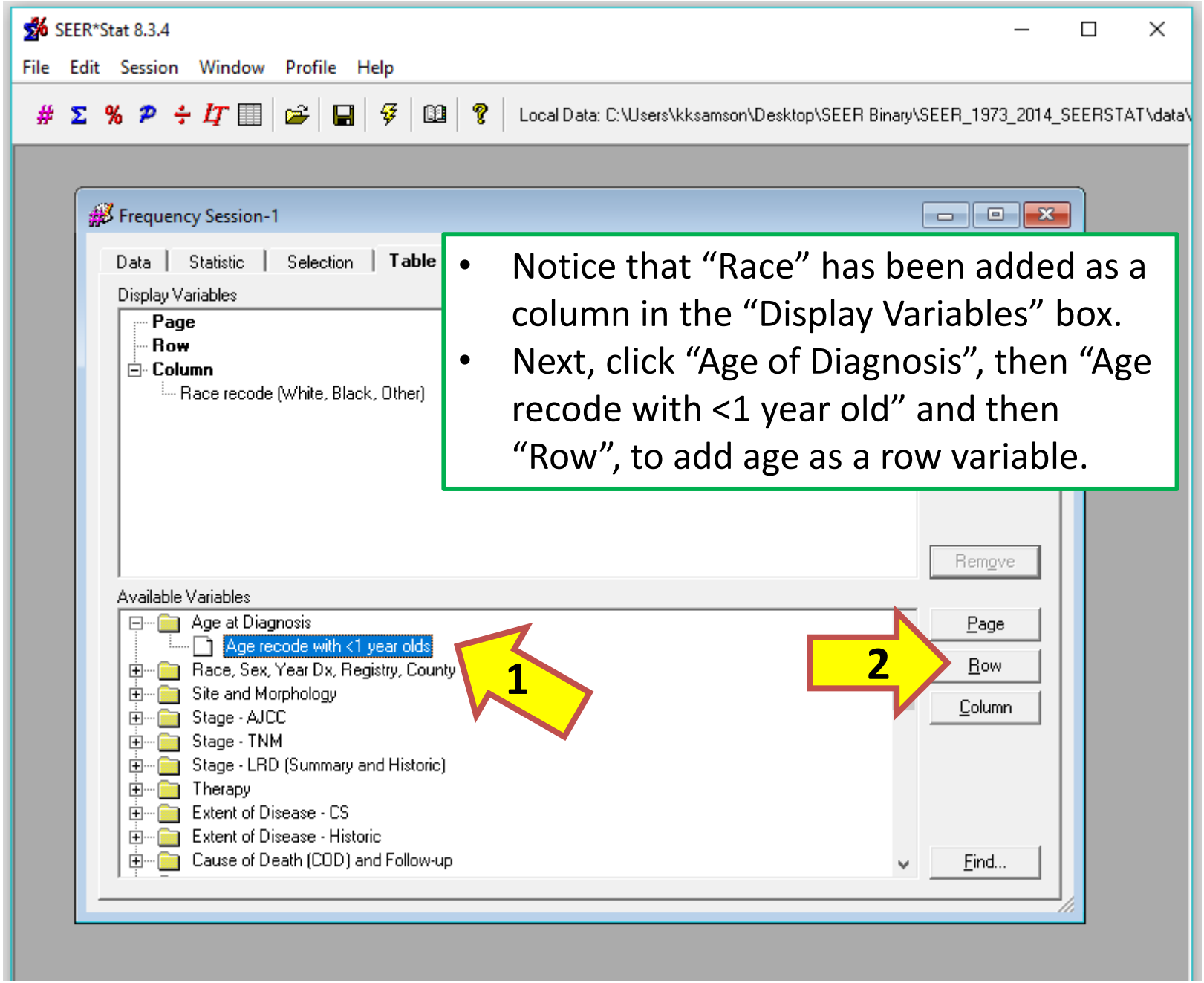


- You can see that our new selection criteria has been added to the text box.
- The “Select Only” box has commonly requested selection criteria.
- Let’s keep all of these selections checked.
- Next, move to the “Table” tab.

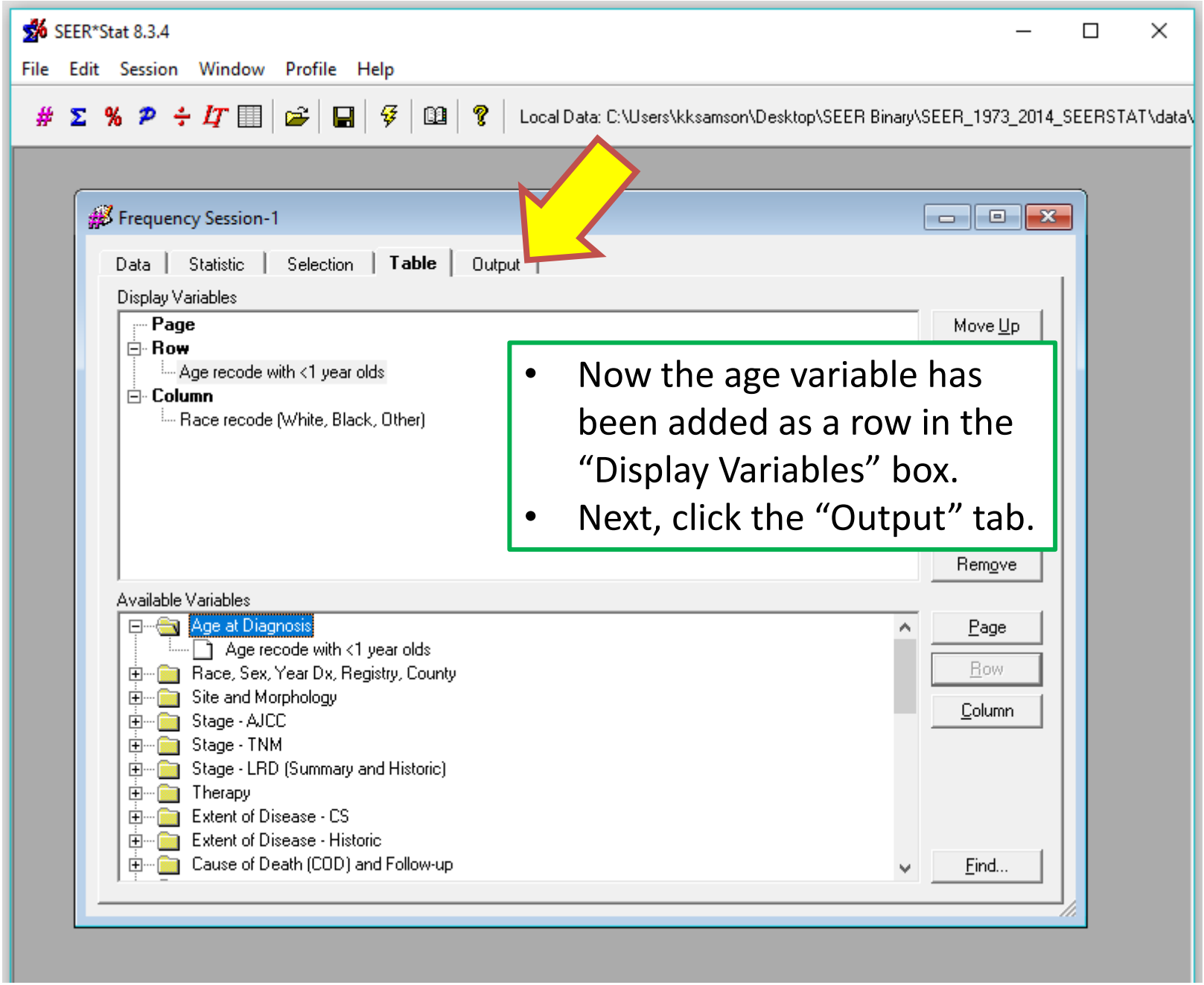


- While the "Selection" tab selected cases, the "Table" tab organizes how your selected data will be displayed.
- Let's say we want to see a cross tabulation of age and race for breast cancer cases.
- We'll have race for columns, and age groups for rows.
- First, lets click on the "Race, Sex, Year Dx, Registry, County" folder to find race.

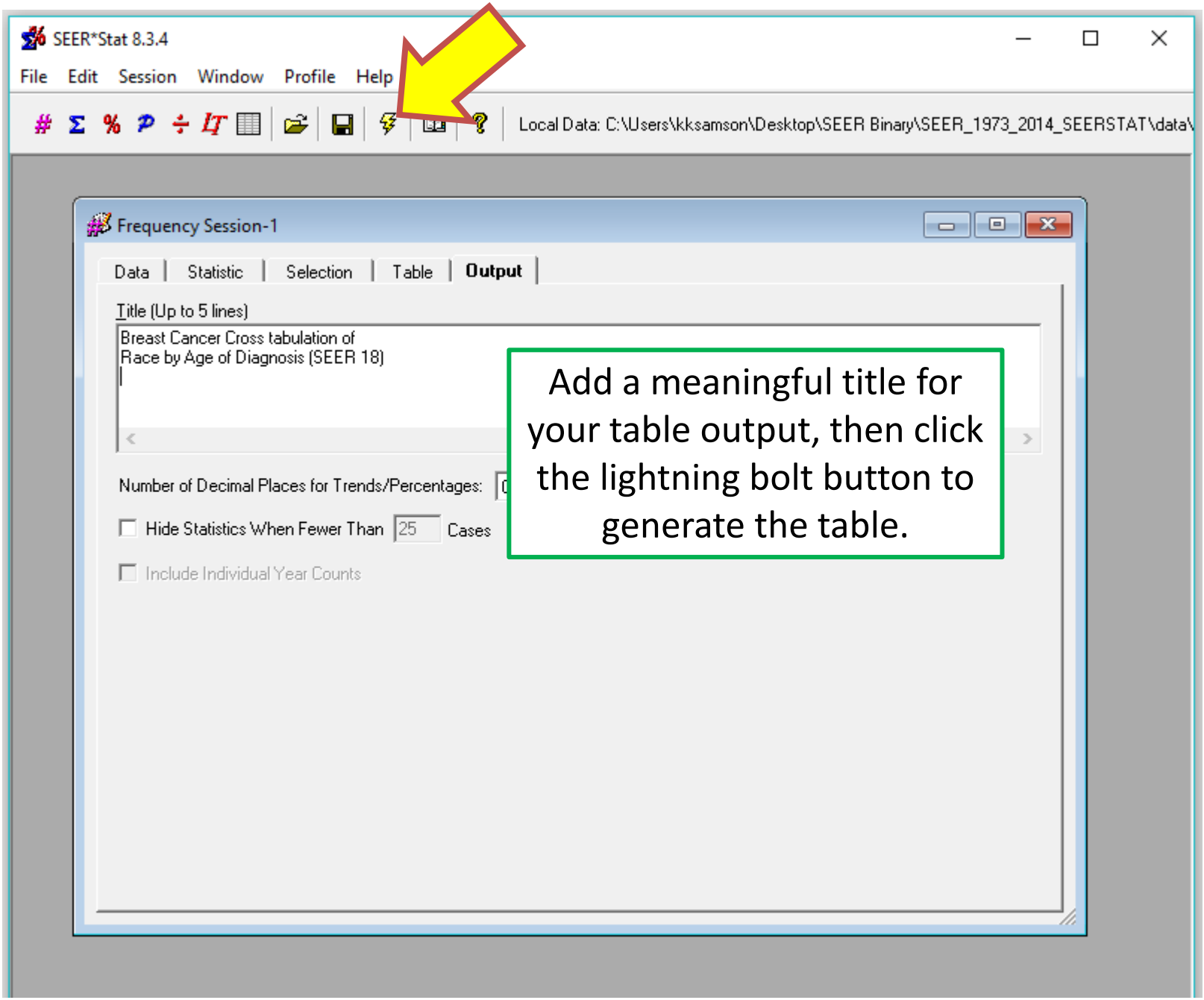

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER Binary\SEER_1973_2014_SEERSTAT\data\
Frequency Session-1
Data Statistic Selection Table Output
Display Variables
Page
Row
Column
Move Up
Click on “Race recode (White, Black, Other)” and then click the “Column” button to add this variable as a column variable to our table.
Available Variables
Race, Sex, Year Dx, Registry, County
Race recode (White, Black, Other)
Sex
Year of diagnosis
SEER registry
Louisiana 2005 - 1st vs 2nd half of year
County
State-county
In research data
CHSDA 2012
CHSDA Region
1
2
Page
Row
Column
Find...

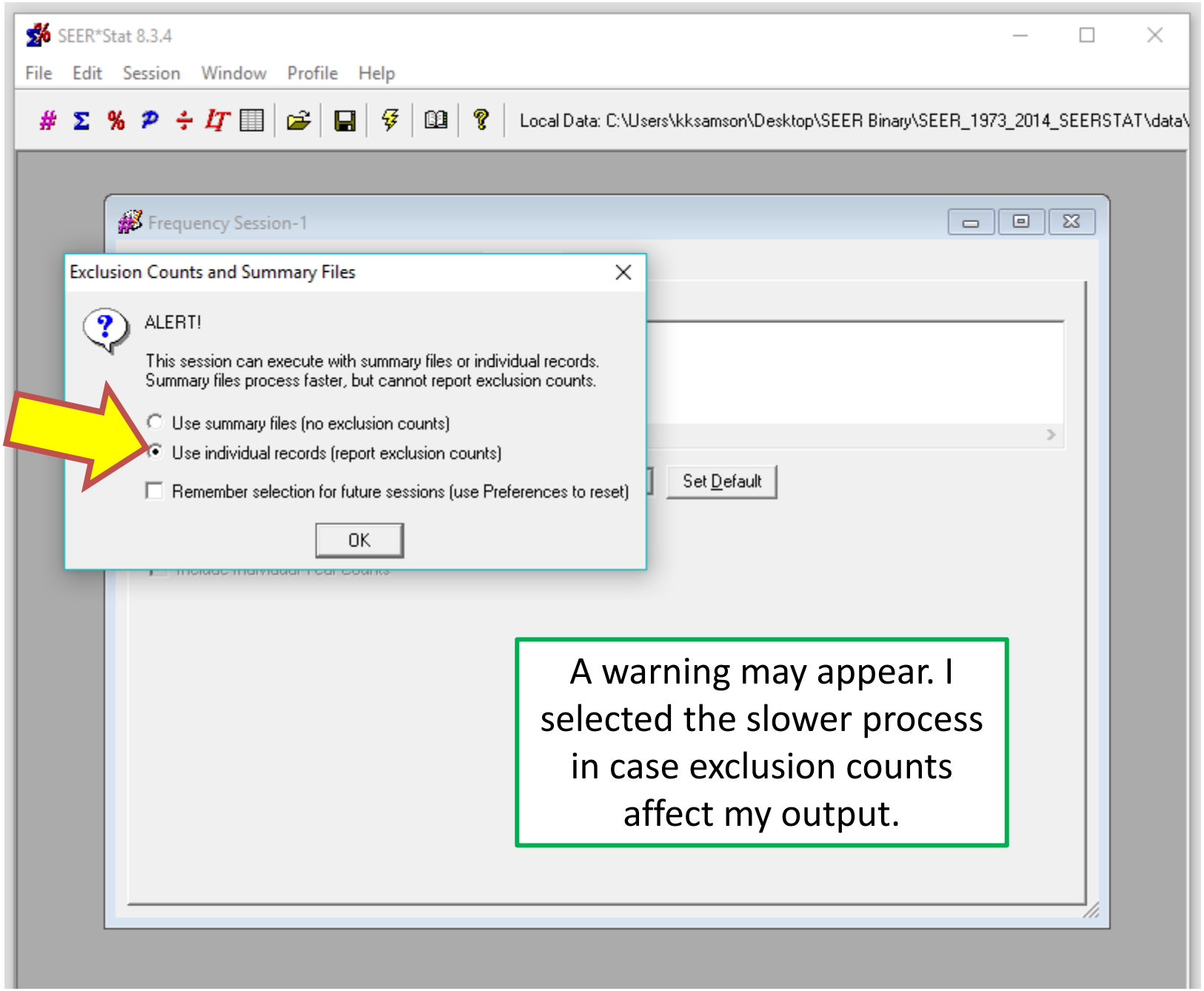


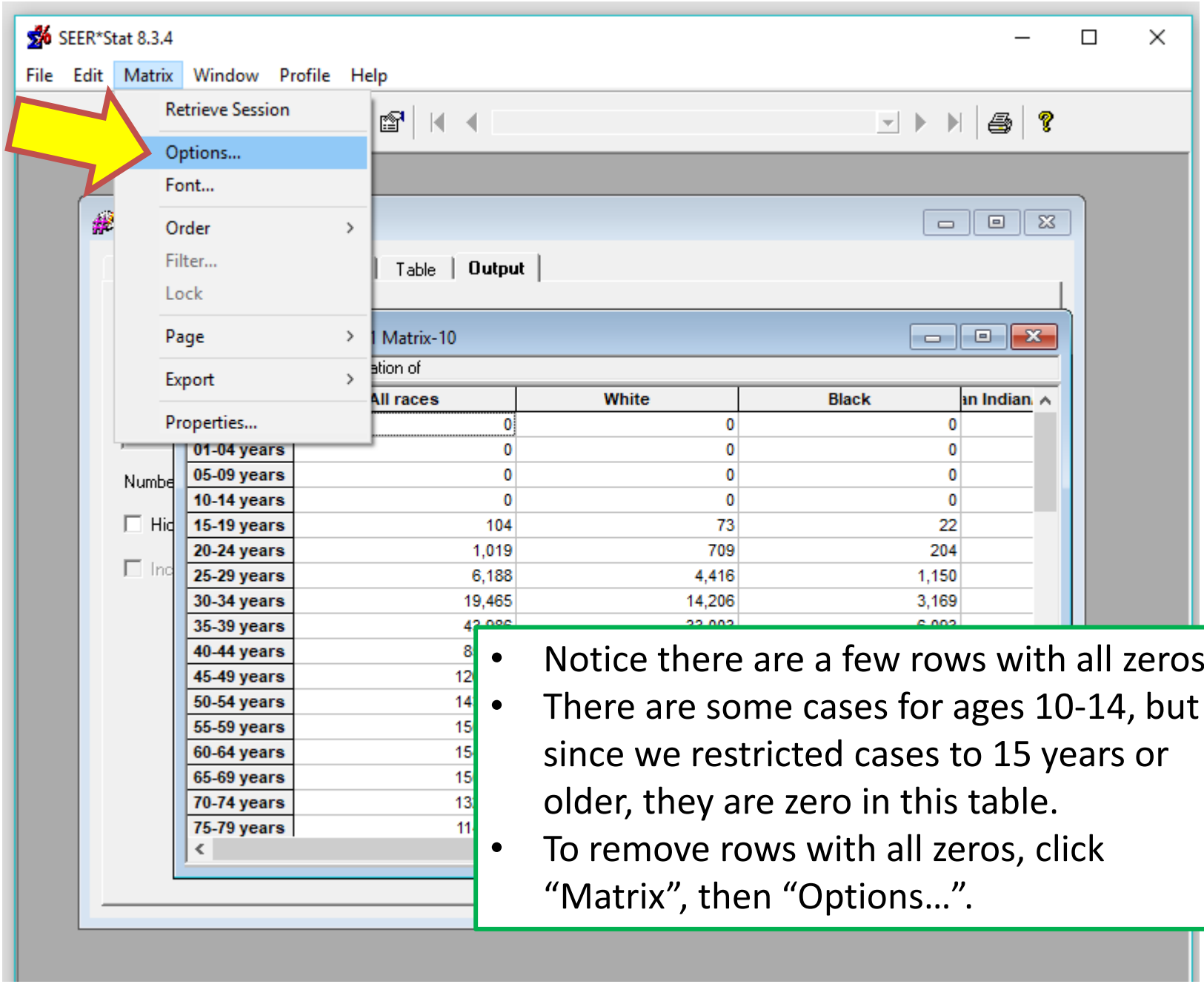
- Notice that "Race" has been added as a column in the "Display Variables" box.
- Next, click "Age of Diagnosis", then "Age recode with <1 year old" and then "Row", to add age as a row variable.



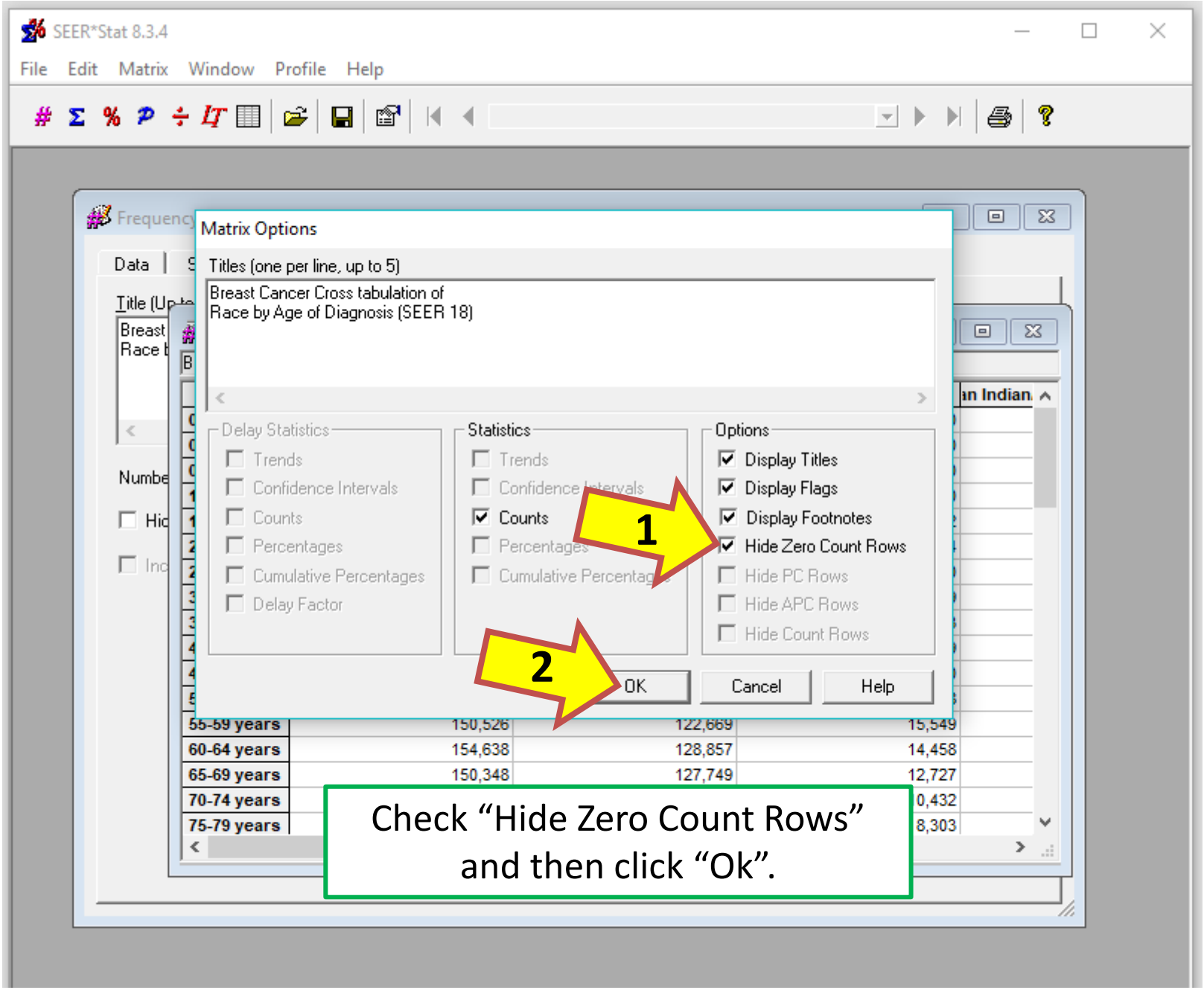
- Now the age variable has been added as a row in the "Display Variables" box.
- Next, click the "Output" tab.

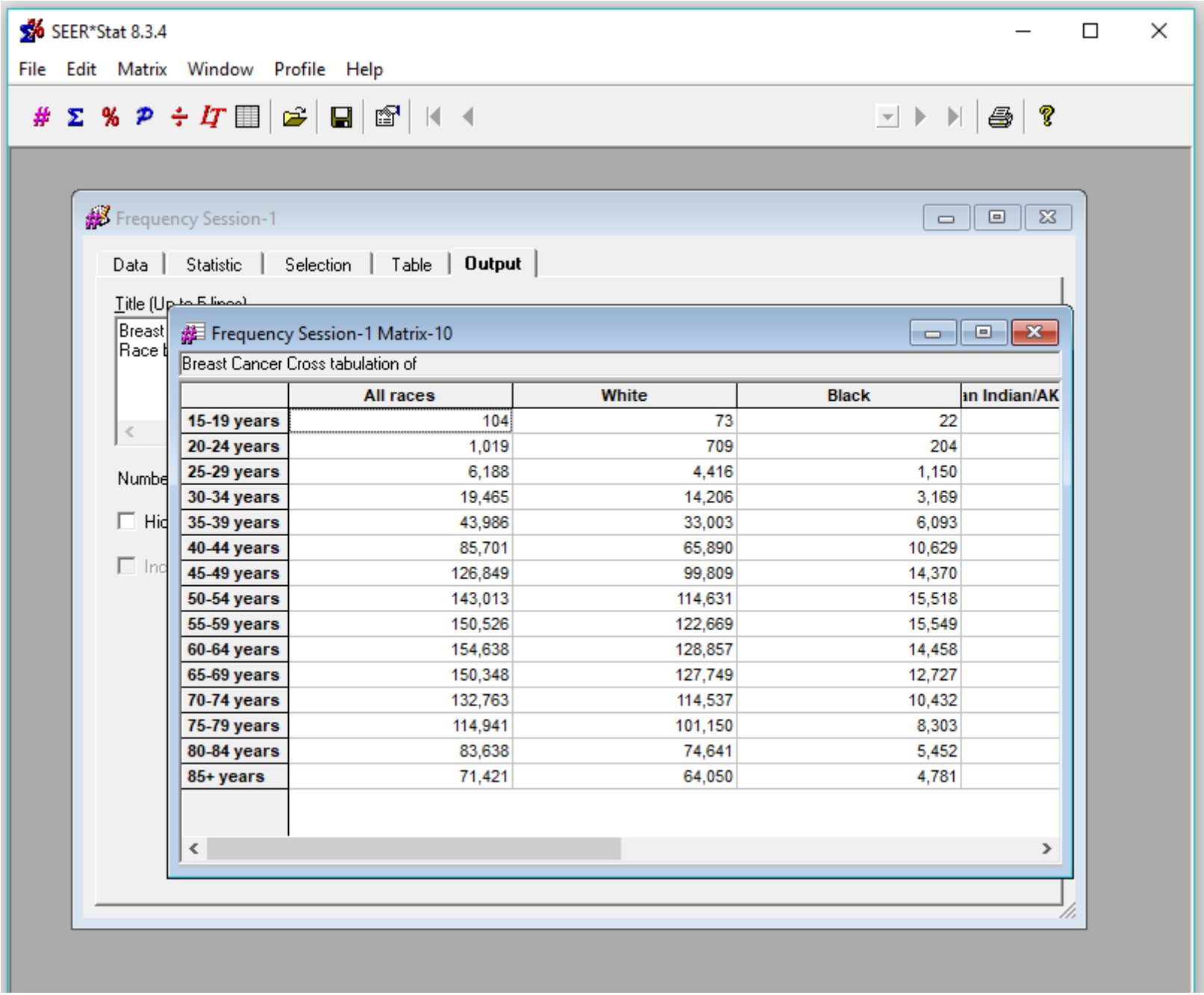

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER Binary\SEER_1973_2014_SEERSTAT\data\
Frequency Session-1
Data Statistic Selection Table Output
Title (Up to 5 lines)
Breast Cancer Cross tabulation of
Race by Age of Diagnosis (SEER 18)
Number of Decimal Places for Trends/Percentages:
Hide Statistics When Fewer Than 25 Cases
Include Individual Year Counts
Add a meaningful title for your table output, then click the lightning bolt button to generate the table.


SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER Binary\SEER_1973_2014_SEERSTAT\data\
Frequency Session-1
Exclusion Counts and Summary Files
ALERT!
This session can execute with summary files or individual records.
Summary files process faster, but cannot report exclusion counts.
Use summary files (no exclusion counts)
Use individual records (report exclusion counts)
Remember selection for future sessions (use Preferences to reset)
OK
Set Default
A warning may appear. I selected the slower process in case exclusion counts affect my output.



- Notice there are a few rows with all zeros.
- There are some cases for ages 10-14, but since we restricted cases to 15 years or older, they are zero in this table.
- To remove rows with all zeros, click “Matrix”, then “Options…”.


SEER*Stat 8.3.4
File Edit Matrix Window Profile Help
Matrix Options
Titles (one per line, up to 5)
Breast Cancer Cross tabulation of
Race by Age of Diagnosis (SEER 18)
Delay Statistics
Trends
Confidence Intervals
Counts
Percentages
Cumulative Percentages
Delay Factor
Statistics
Trends
Confidence Intervals
Counts
Percentages
Cumulative Percentages
Options
Display Titles
Display Flags
Display Footnotes
Hide Zero Count Rows
Hide PC Rows
Hide APC Rows
Hide Count Rows
1
2
OK
Cancel
Help
55-59 years 150,526 122,669 15,549
60-64 years 154,638 128,857 14,458
65-69 years 150,348 127,749 12,727
70-74 years 0,432
75-79 years 8,303
Check “Hide Zero Count Rows” and then click “Ok”.


SEER*Stat 8.3.4
File Edit Matrix Window Profile Help
Frequency Session-1
Data Statistic Selection Table Output
Title (Up to 5 lines)
Breast
Race
Frequency Session-1 Matrix-10
Breast Cancer Cross tabulation of
All races
White
Black
an Indian/AK
15-19 years 104 73 22
20-24 years 1,019 709 204
25-29 years 6,188 4,416 1,150
30-34 years 19,465 14,206 3,169
35-39 years 43,986 33,003 6,093
40-44 years 85,701 65,890 10,629
45-49 years 126,849 99,809 14,370
50-54 years 143,013 114,631 15,518
55-59 years 150,526 122,669 15,549
60-64 years 154,638 128,857 14,458
65-69 years 150,348 127,749 12,727
70-74 years 132,763 114,537 10,432
75-79 years 114,941 101,150 8,303
80-84 years 83,638 74,641 5,452
85+ years 71,421 64,050 4,781

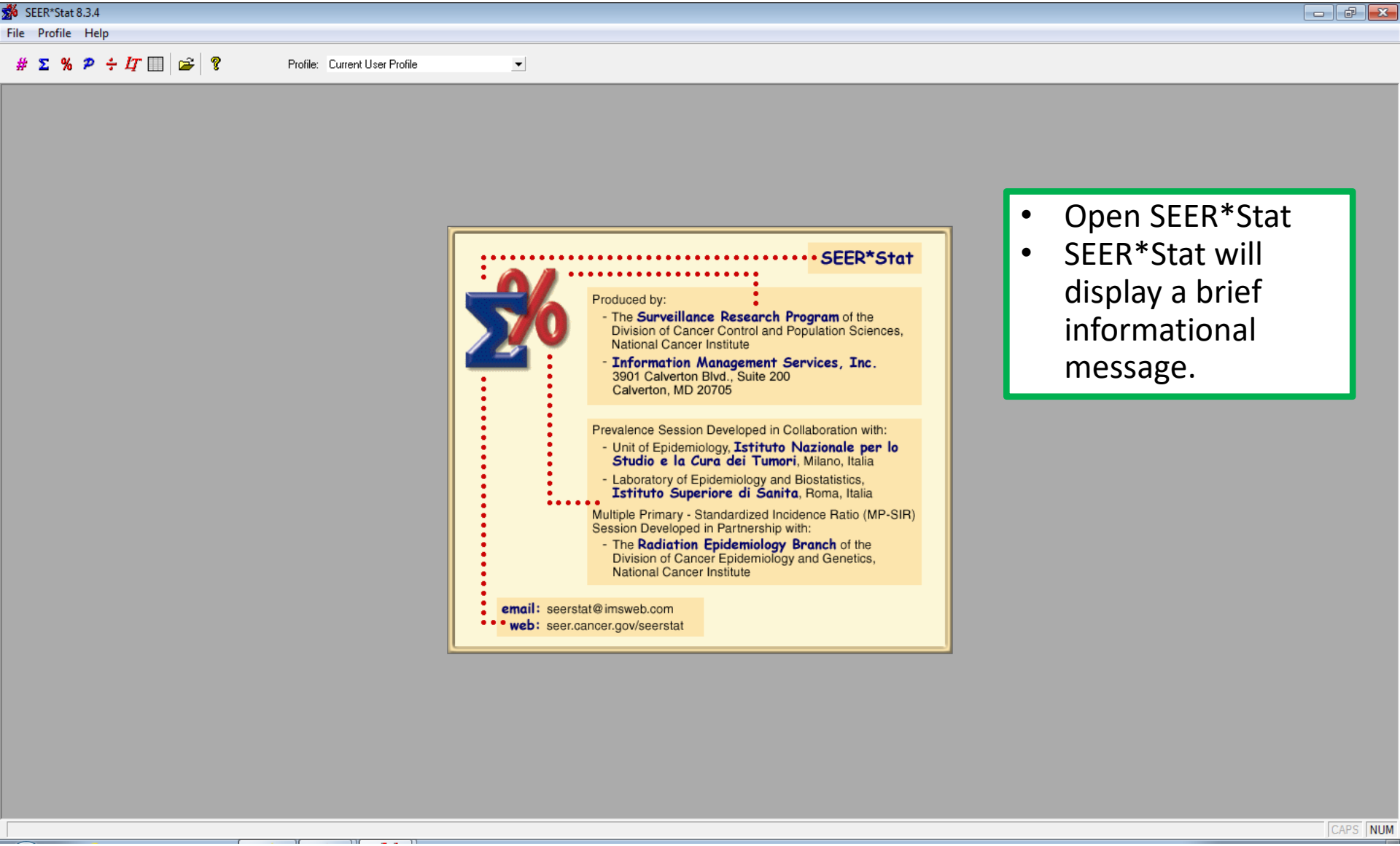
Mortality Counts and Rates

Example: Heart Disease

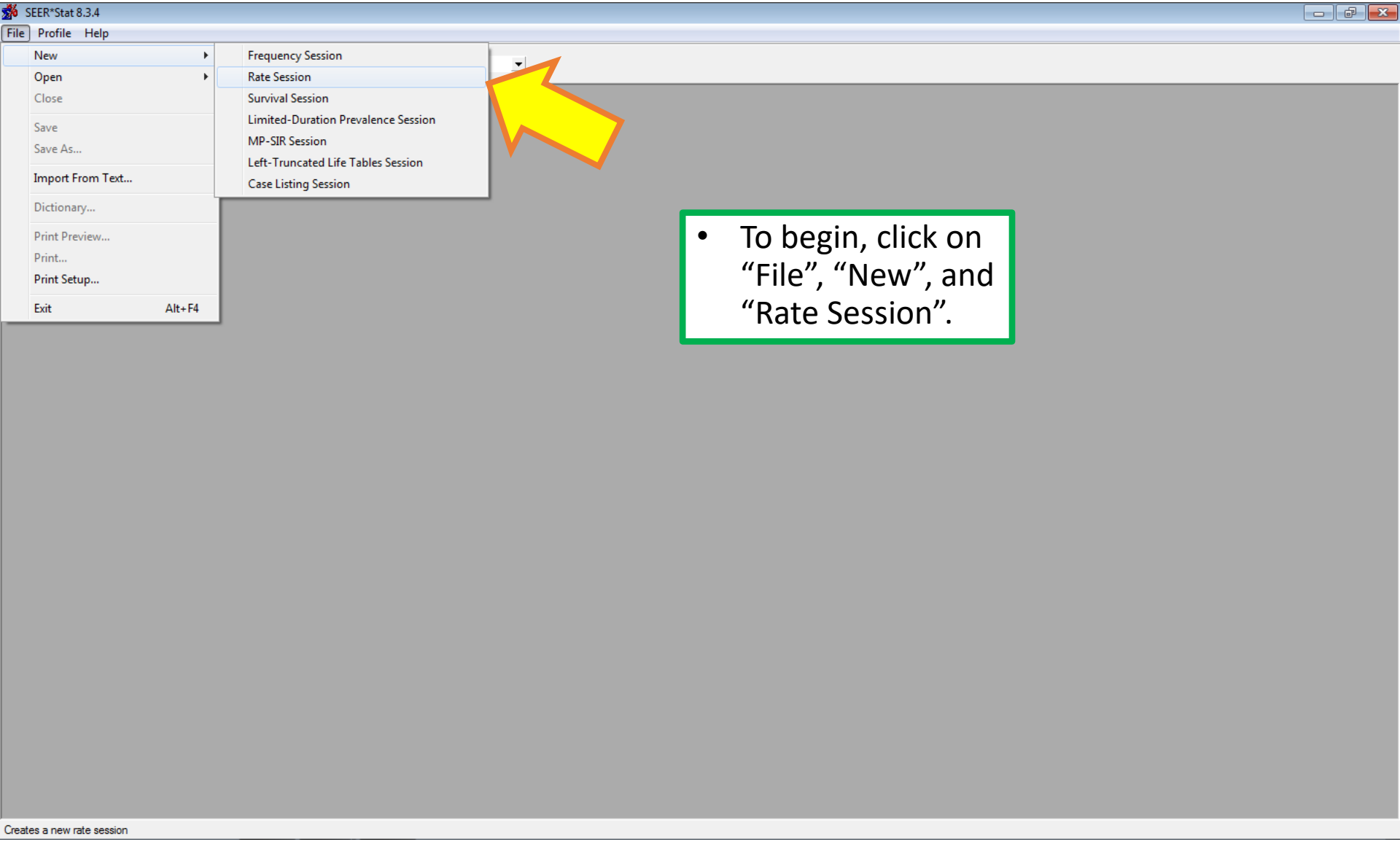
University of Nebraska Medical Center

Mortality Counts and Rates

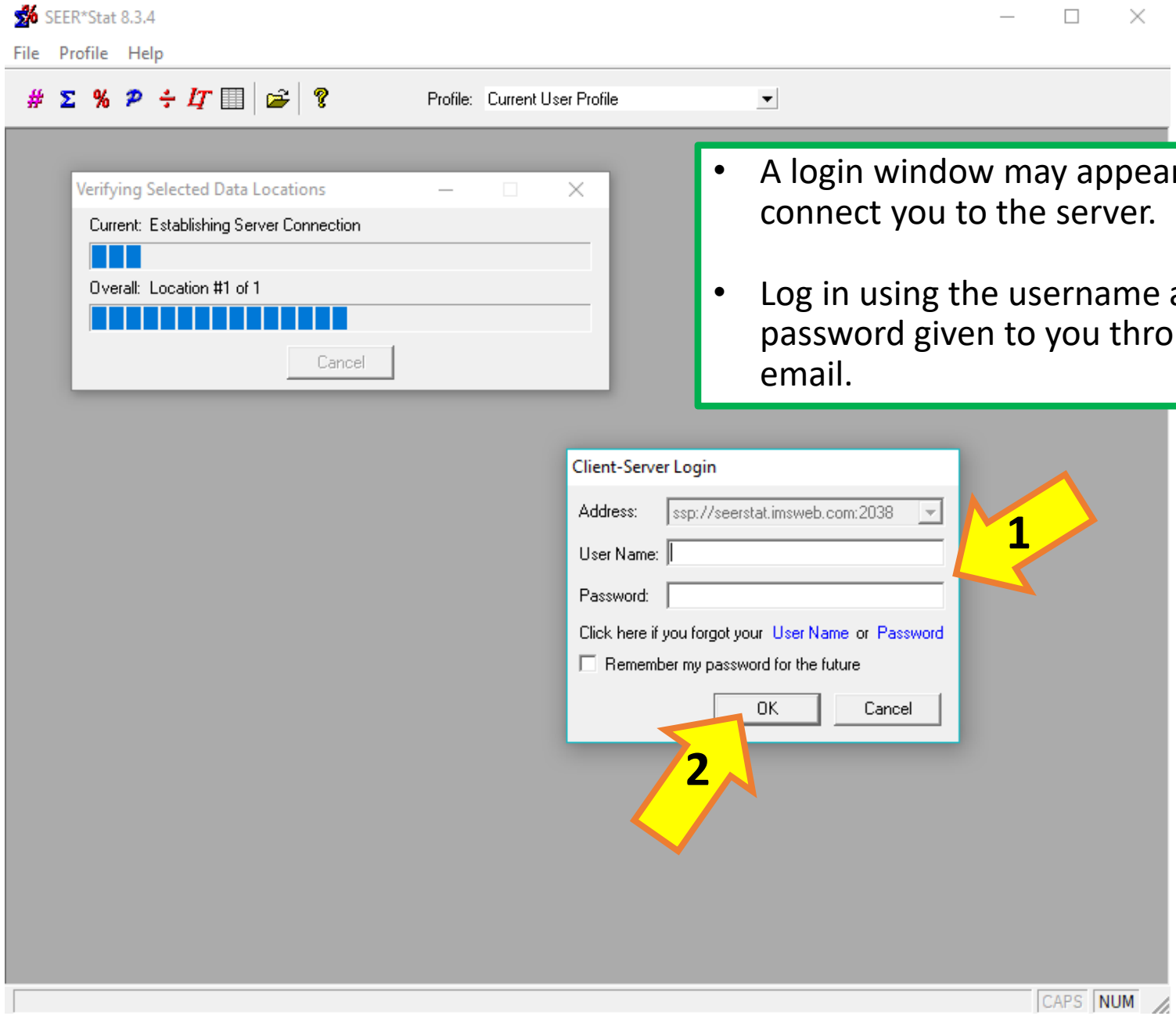
- This section will instruct you on how to get counts and age-adjusted rates for mortality data.
- Mortality data is only available through SEER*Stat.
- For this example, we will look at heart diseases mortality counts and age-adjusted rates in three separate states, to be able to compare them:
  - Nebraska
  - North Dakota
  - South Dakota
- These mortality examples were completed using on-line SEER data, so some screens and output may look different if you're using downloaded binary SEER data.



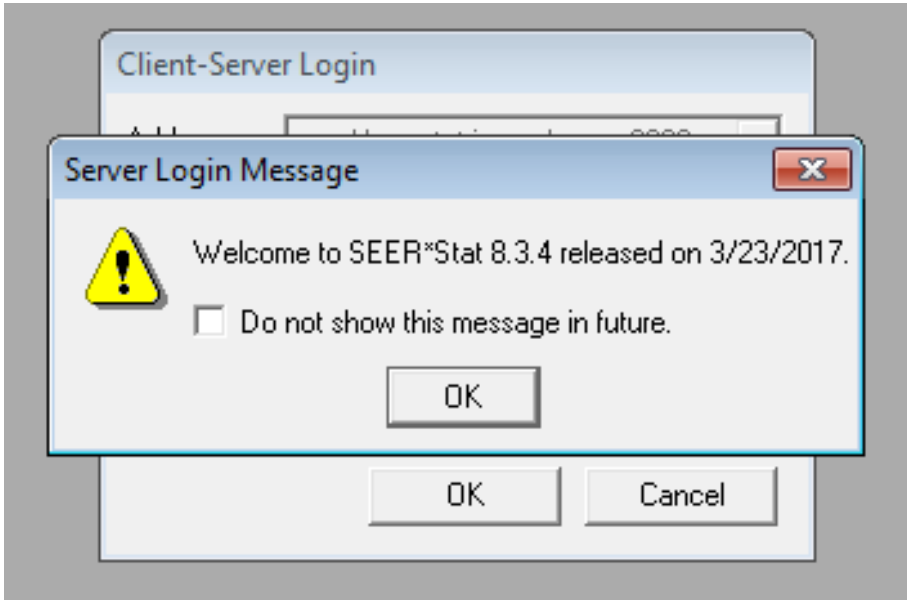
- Open SEER*Stat
- SEER*Stat will display a brief informational message.



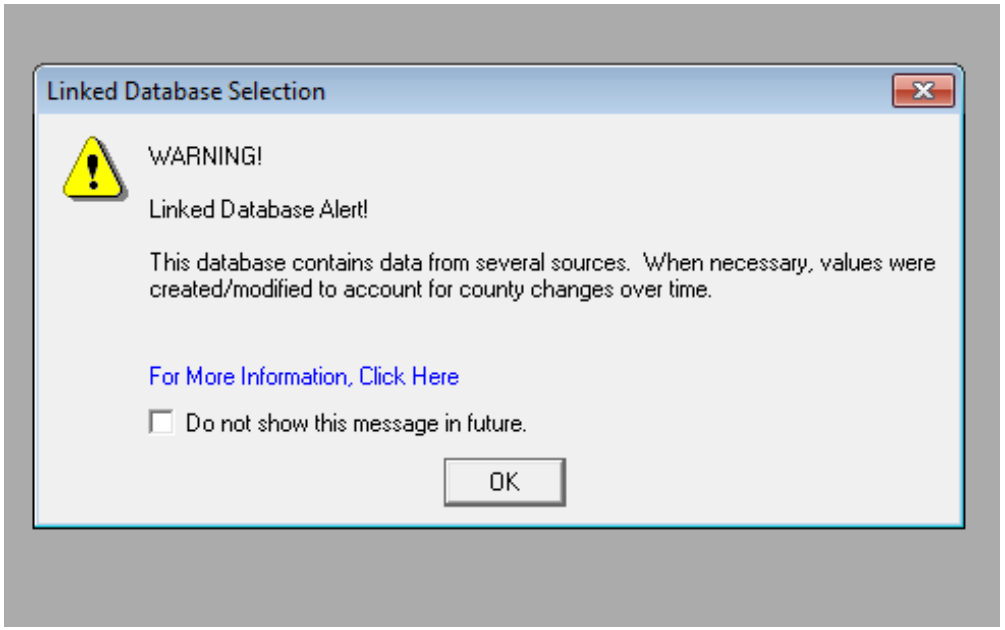
- To begin, click on “File”, “New”, and “Rate Session”.

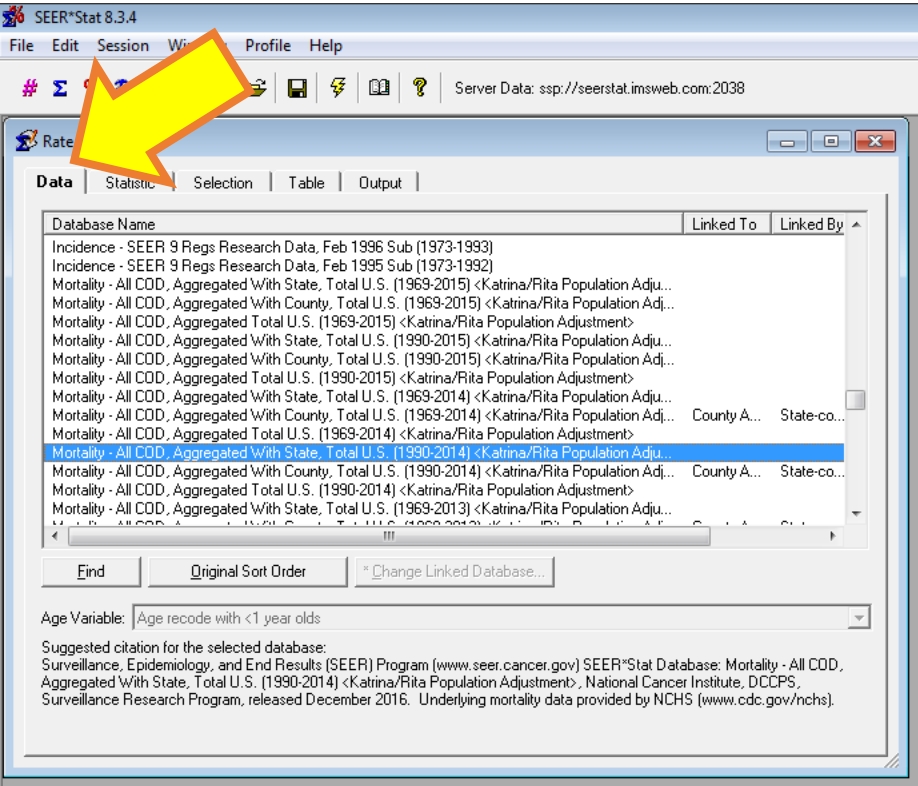


- A login window may appear to connect you to the server.
- Log in using the username and password given to you through email.



Click OK for both the welcome and warning messages to continue.

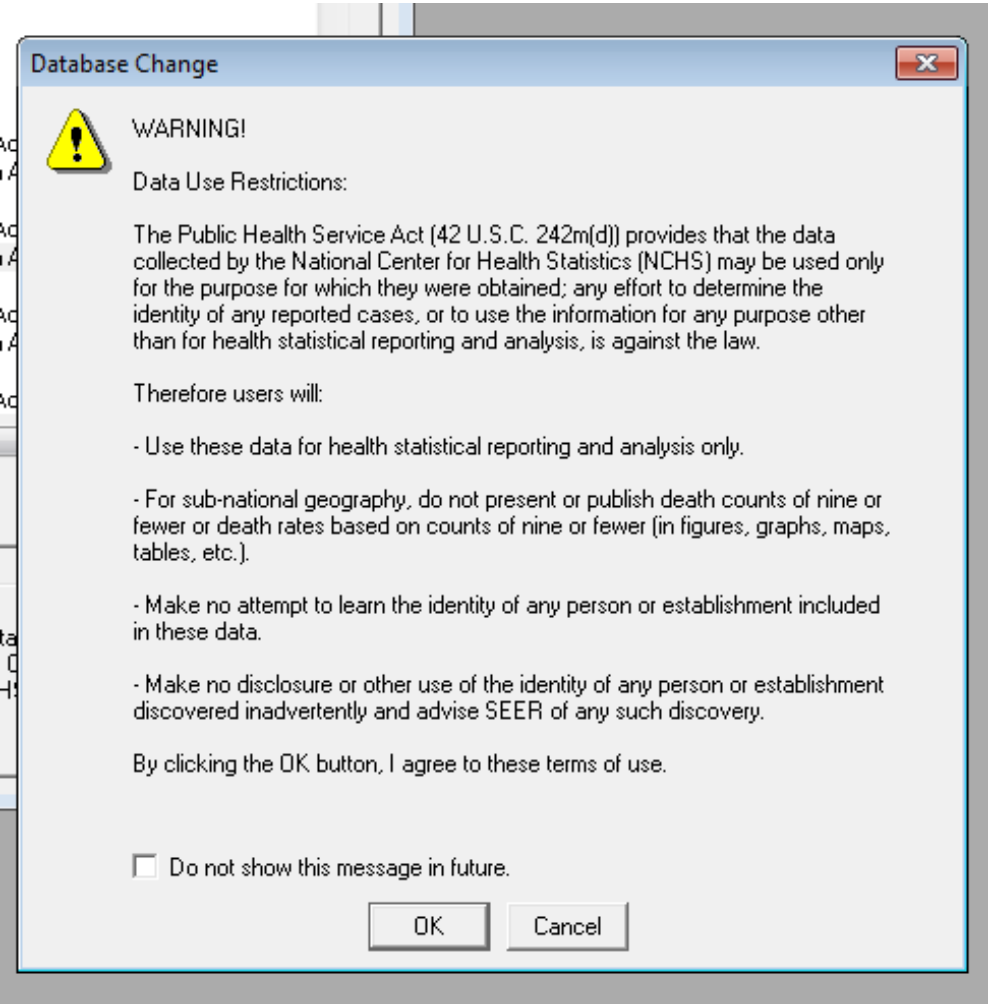




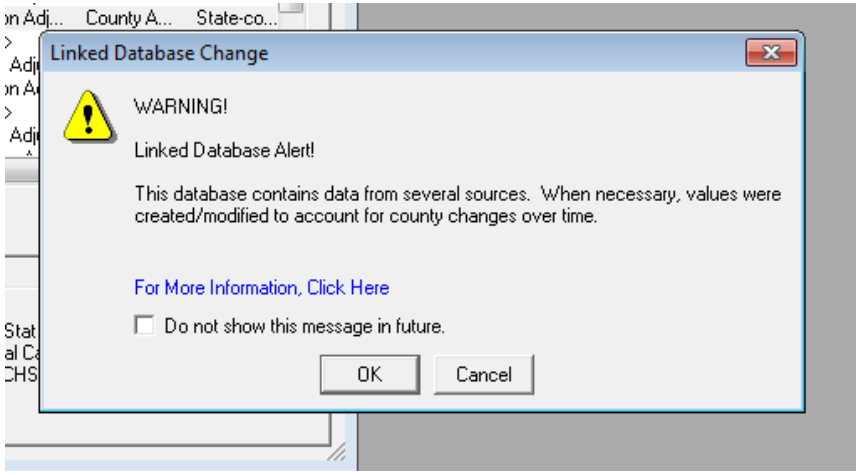
The rate session begins by displaying a window with various tabs.

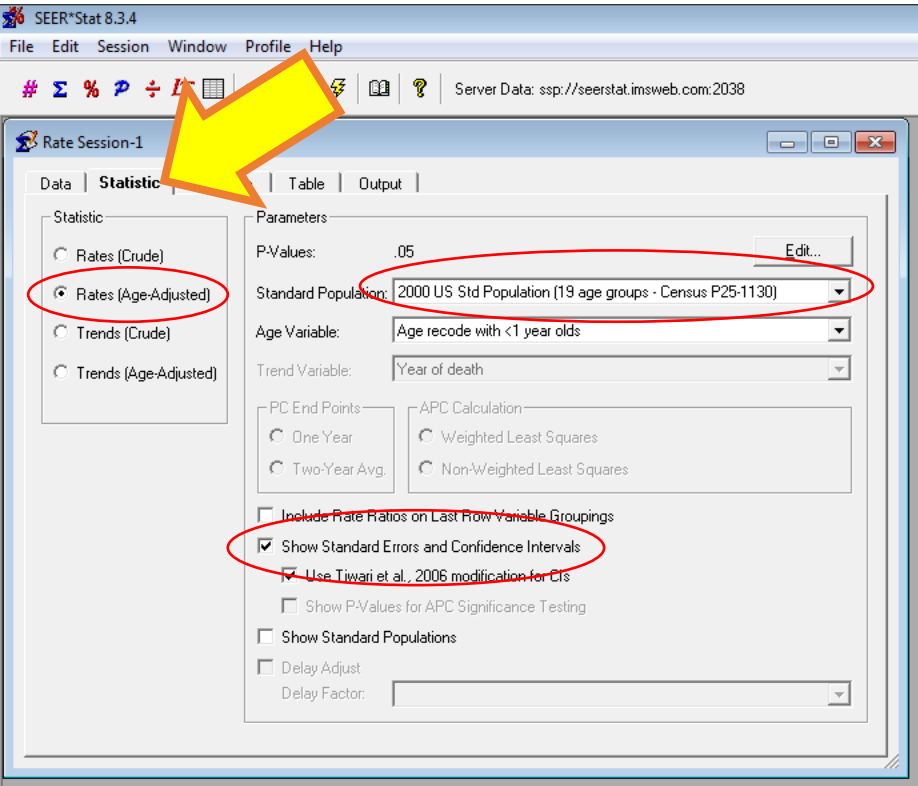
- Go to the “Data” tab.
- Scroll down and select the database titled “**Mortality - All COD, Aggregated With County, Total U.S. (1990-2014) <Katrina/Rita Population Adjustment>**”.

Choosing which database will depend on what data you are looking for. Some of the database options are explained at the website: https://seer.cancer.gov/data/seerstat/



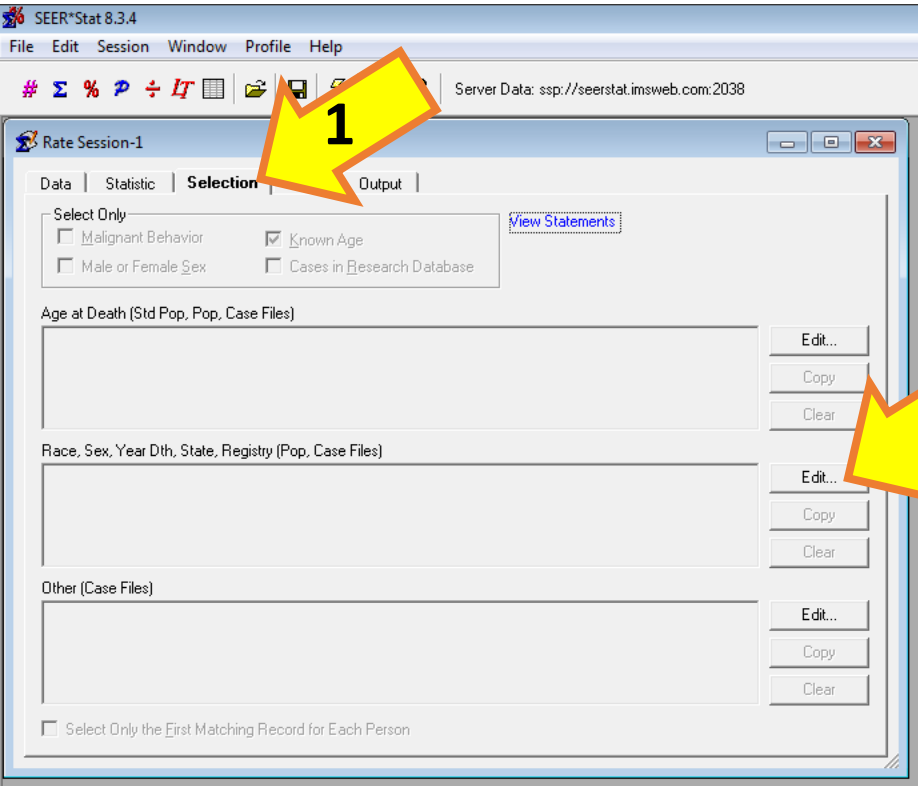
- Clicking on a database will display the warning messages about the data.
- Review and then click “OK” to both of these messages.





- Go to the “Statistic” tab.
- Choose “Rates: Age-Adjusted” under the “Statistic” box.
- Choose “2000 US Std Population” under the “Parameters” box.
- Choose “Show Standard Errors and Confidence Intervals” under the “Parameters” box.

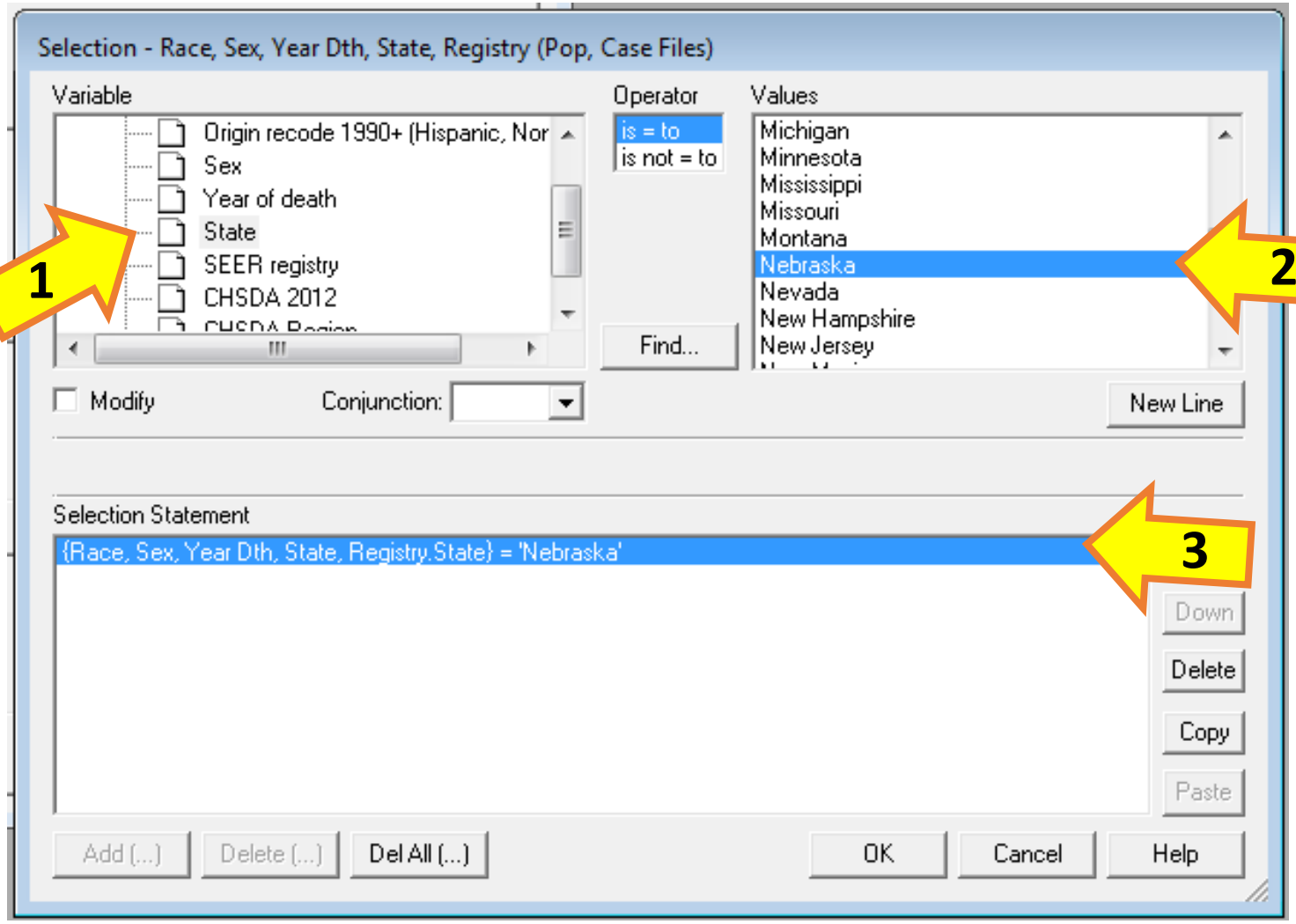
Using an age-adjusted rate adapts the crude rates to standard population age groups. This allows us to accurately compare rates between populations that may not actually have the same age distribution.



- Go to the “Selection” tab.
- Click on “Edit” for the “Race, Sex, Year Dth, State, Registry.State” category.

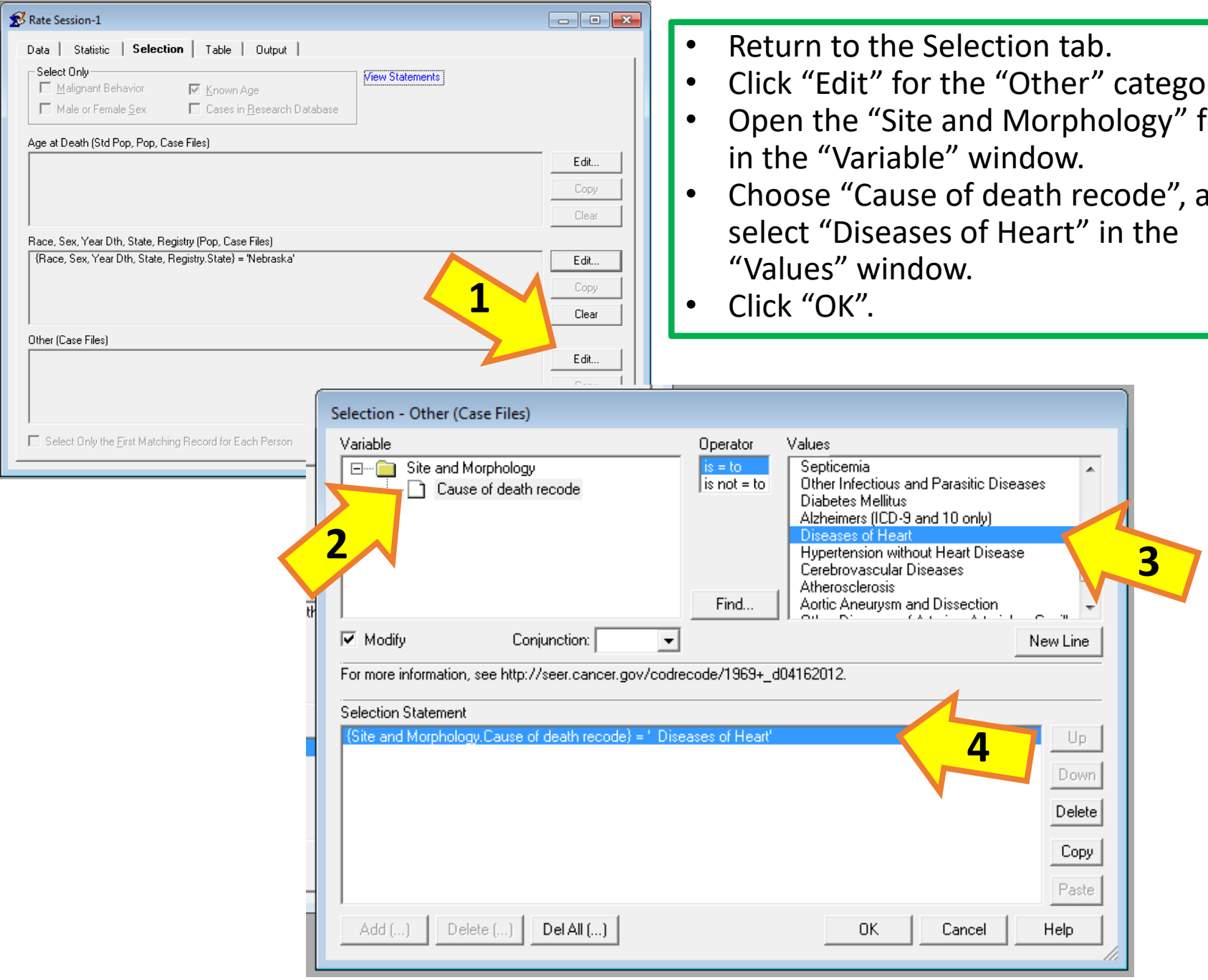
A new window will be displayed.

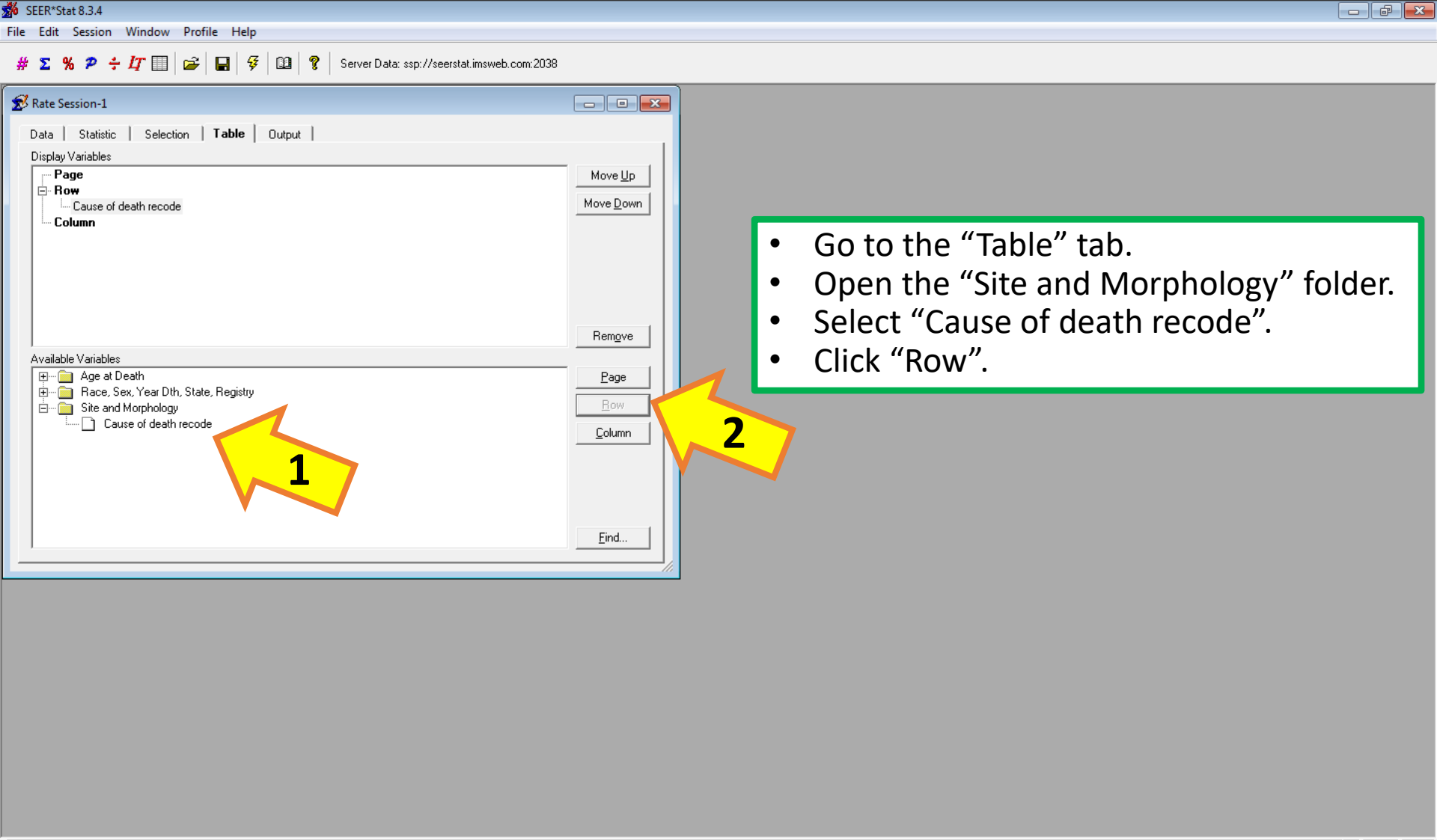
Note: the selection tab determines what data is used for the analysis. This is a good place to specify inclusion/exclusion criteria.



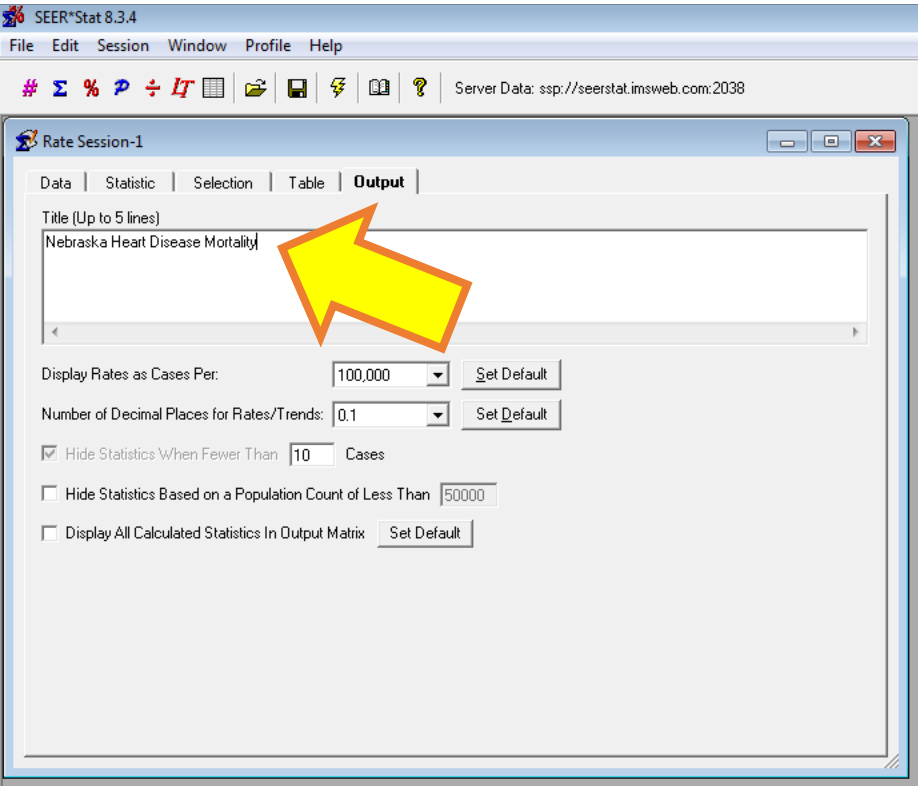
- Click on the “Race, Sex, Year Dth, State, Registry.State” folder in the “Variable” window, and select “State”.
- Select “Nebraska” in the “Values” window. Once selected, the values will display in the “Selection Statement” box.

- Return to the Selection tab.
- Click “Edit” for the “Other” category.
- Open the “Site and Morphology” folder in the “Variable” window.
- Choose “Cause of death recode”, and select “Diseases of Heart” in the “Values” window.
- Click “OK”.

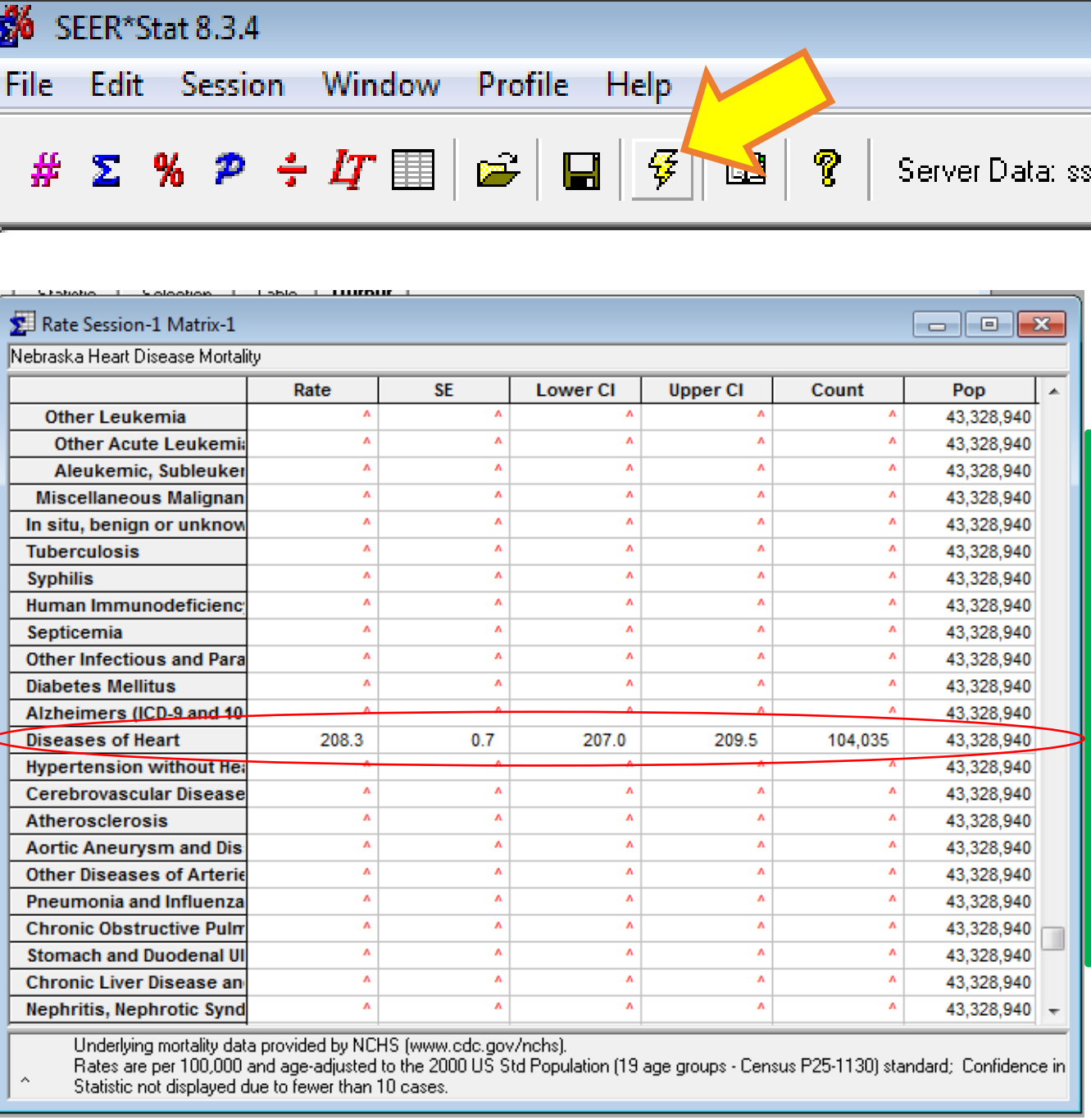




- Go to the “Table” tab.
- Open the “Site and Morphology” folder.
- Select “Cause of death recode”.
- Click “Row”.

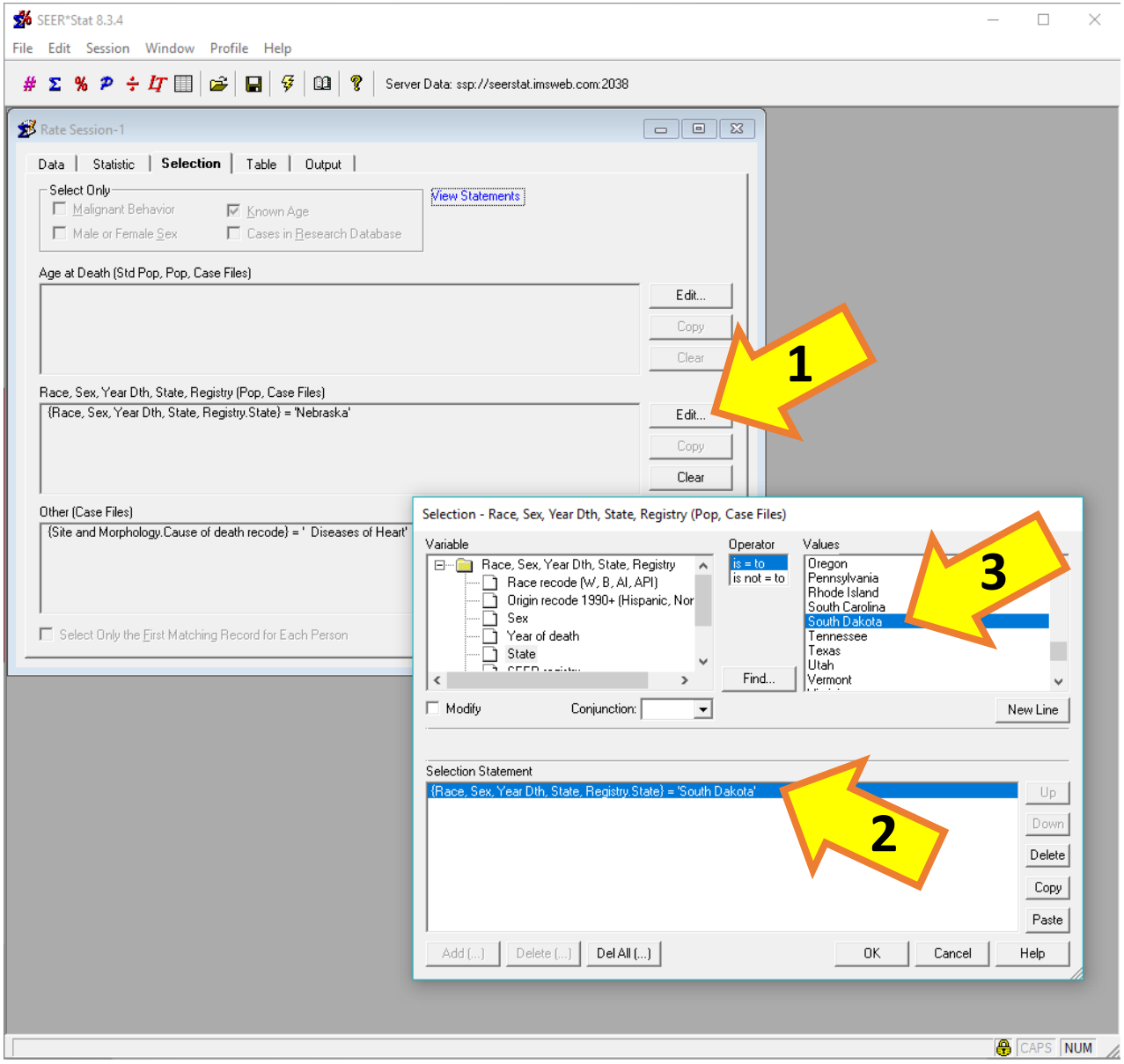


- Go to the “Output” tab.
- Type “Nebraska Heart Disease Mortality” in the “Title” box.



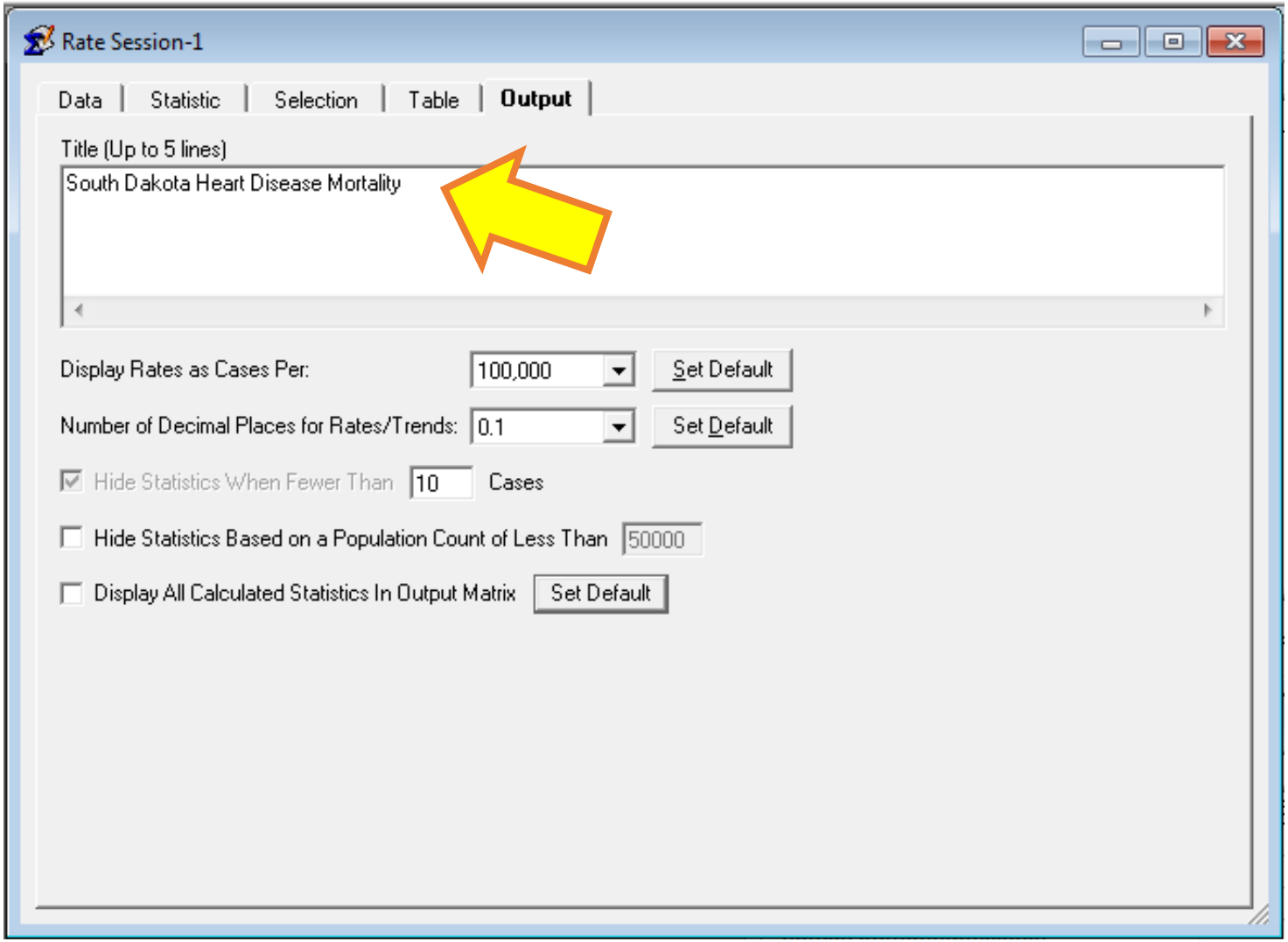
- Click on the lightning bolt icon in the toolbar.

Your table will then be displayed. The table will show all site recode categories for mortality, but will show ^ for all values except those for diseases of heart, since that is the variable we chose in the "Selection" tab.

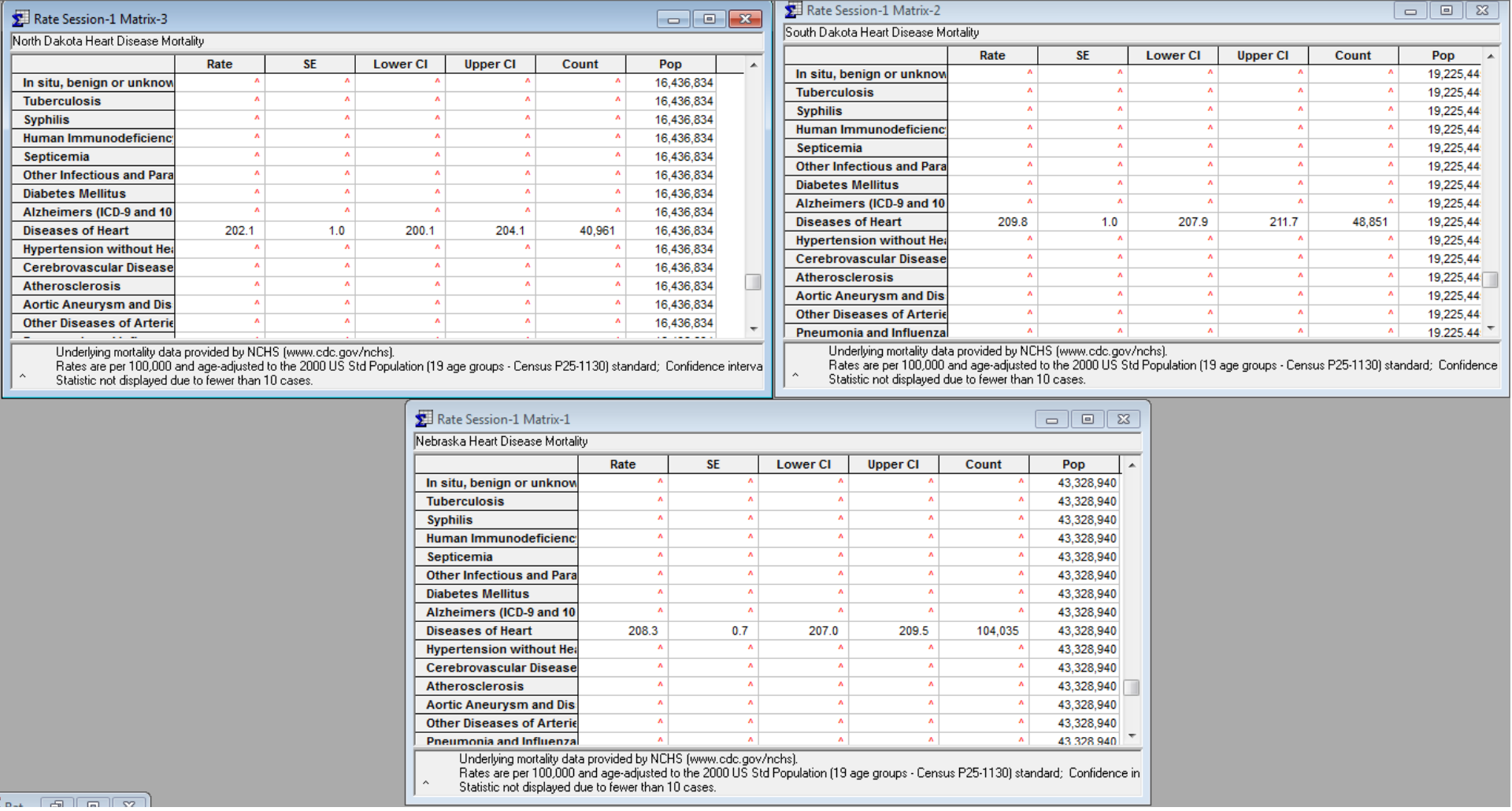


We have produced the table for Nebraska, but we still need tables for South Dakota and North Dakota for comparison.

- Go back to the same rate session window.
- Go to the “Selection” Tab.
- Click “Edit” for the “Race, Sex, Year Dth, State, Registry” category.
- Click on the text in the “Selection Statement” box.
- Choose “South Dakota” in the “Values” window.
- Click “OK”.



- Go to the “Output” tab and make a new title to avoid confusing your tables!



Rate Session-1 Matrix-3

North Dakota Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	16,436,834
Tuberculosis	^	^	^	^	^	16,436,834
Syphilis	^	^	^	^	^	16,436,834
Human Immunodeficienc	^	^	^	^	^	16,436,834
Septicemia	^	^	^	^	^	16,436,834
Other Infectious and Para	^	^	^	^	^	16,436,834
Diabetes Mellitus	^	^	^	^	^	16,436,834
Alzheimers (ICD-9 and 10	^	^	^	^	^	16,436,834
Diseases of Heart	202.1	1.0	200.1	204.1	40,961	16,436,834
Hypertension without He	^	^	^	^	^	16,436,834
Cerebrovascular Disease	^	^	^	^	^	16,436,834
Atherosclerosis	^	^	^	^	^	16,436,834
Aortic Aneurysm and Dis	^	^	^	^	^	16,436,834
Other Diseases of Arterie	^	^	^	^	^	16,436,834

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence interva
^ Statistic not displayed due to fewer than 10 cases.

Rate Session-1 Matrix-2

South Dakota Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	19,225,44
Tuberculosis	^	^	^	^	^	19,225,44
Syphilis	^	^	^	^	^	19,225,44
Human Immunodeficienc	^	^	^	^	^	19,225,44
Septicemia	^	^	^	^	^	19,225,44
Other Infectious and Para	^	^	^	^	^	19,225,44
Diabetes Mellitus	^	^	^	^	^	19,225,44
Alzheimers (ICD-9 and 10	^	^	^	^	^	19,225,44
Diseases of Heart	209.8	1.0	207.9	211.7	48,851	19,225,44
Hypertension without He	^	^	^	^	^	19,225,44
Cerebrovascular Disease	^	^	^	^	^	19,225,44
Atherosclerosis	^	^	^	^	^	19,225,44
Aortic Aneurysm and Dis	^	^	^	^	^	19,225,44
Other Diseases of Arterie	^	^	^	^	^	19,225,44
Pneumonia and Influenza	^	^	^	^	^	19,225,44

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence
^ Statistic not displayed due to fewer than 10 cases.

Rate Session-1 Matrix-1

Nebraska Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	43,328,940
Tuberculosis	^	^	^	^	^	43,328,940
Syphilis	^	^	^	^	^	43,328,940
Human Immunodeficienc	^	^	^	^	^	43,328,940
Septicemia	^	^	^	^	^	43,328,940
Other Infectious and Para	^	^	^	^	^	43,328,940
Diabetes Mellitus	^	^	^	^	^	43,328,940
Alzheimers (ICD-9 and 10	^	^	^	^	^	43,328,940
Diseases of Heart	208.3	0.7	207.0	209.5	104,035	43,328,940
Hypertension without He	^	^	^	^	^	43,328,940
Cerebrovascular Disease	^	^	^	^	^	43,328,940
Atherosclerosis	^	^	^	^	^	43,328,940
Aortic Aneurysm and Dis	^	^	^	^	^	43,328,940
Other Diseases of Arterie	^	^	^	^	^	43,328,940
Pneumonia and Influenza	^	^	^	^	^	43,328,940

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence in
^ Statistic not displayed due to fewer than 10 cases.

- Click the lightning bolt icon.
- Repeat the steps in the last two slides to create your North Dakota table.

Now we have mortality rates for Nebraska, South Dakota, and North Dakota, and can compare the rates between these states.

Rate Session-1 Matrix-1

Nebraska Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	43,328,940
Tuberculosis	^	^	^	^	^	43,328,940
Syphilis	^	^	^	^	^	43,328,940
Human Immunodeficienc	^	^	^	^	^	43,328,940
Septicemia	^	^	^	^	^	43,328,940
Other Infectious and Para	^	^	^	^	^	43,328,940
Diabetes Mellitus	^	^	^	^	^	43,328,940
Alzheimers (ICD-9 and 10	^	^	^	^	^	43,328,940
Diseases of Heart	208.3	0.7	207.0	209.5	104,035	43,328,940
Hypertension without Hea	^	^	^	^	^	43,328,940
Cerebrovascular Disease	^	^	^	^	^	43,328,940
Atherosclerosis	^	^	^	^	^	43,328,940
Aortic Aneurysm and Dis	^	^	^	^	^	43,328,940
Other Diseases of Arterie	^	^	^	^	^	43,328,940
Pneumonia and Influenza	^	^	^	^	^	43,328,940

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence in
^ Statistic not displayed due to fewer than 10 cases.

First, let's compare Nebraska and North Dakota. Nebraska had a count of 104,035 deaths from heart disease between 1990-2014; North Dakota had 40,961 deaths. This seems like a large difference until we notice the discrepancy in population size.

Rate Session-1 Matrix-3

North Dakota Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	16,436,834
Tuberculosis	^	^	^	^	^	16,436,834
Syphilis	^	^	^	^	^	16,436,834
Human Immunodeficienc	^	^	^	^	^	16,436,834
Septicemia	^	^	^	^	^	16,436,834
Other Infectious and Para	^	^	^	^	^	16,436,834
Diabetes Mellitus	^	^	^	^	^	16,436,834
Alzheimers (ICD-9 and 10	^	^	^	^	^	16,436,834
Diseases of Heart	202.1	1.0	200.1	204.1	40,961	16,436,834
Hypertension without Hea	^	^	^	^	^	16,436,834
Cerebrovascular Disease	^	^	^	^	^	16,436,834
Atherosclerosis	^	^	^	^	^	16,436,834
Aortic Aneurysm and Dis	^	^	^	^	^	16,436,834
Other Diseases of Arterie	^	^	^	^	^	16,436,834

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence interva
^ Statistic not displayed due to fewer than 10 cases.

Nebraska had a rate of 208.3 deaths per 100,000. North Dakota had a rate of 202.1 deaths per 100,000. Nebraska had a slightly higher mortality rate due to heart disease between 1990-2014.

Rate Session-1 Matrix-1

Nebraska Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	43,328,940
Tuberculosis	^	^	^	^	^	43,328,940
Syphilis	^	^	^	^	^	43,328,940
Human Immunodeficienc	^	^	^	^	^	43,328,940
Septicemia	^	^	^	^	^	43,328,940
Other Infectious and Para	^	^	^	^	^	43,328,940
Diabetes Mellitus	^	^	^	^	^	43,328,940
Alzheimers (ICD-9 and 10	^	^	^	^	^	43,328,940
Diseases of Heart	208.3	0.7	207.0	209.5	104,035	43,328,940
Hypertension without Hea	^	^	^	^	^	43,328,940
Cerebrovascular Disease	^	^	^	^	^	43,328,940
Atherosclerosis	^	^	^	^	^	43,328,940
Aortic Aneurysm and Dis	^	^	^	^	^	43,328,940
Other Diseases of Arterie	^	^	^	^	^	43,328,940
Pneumonia and Influenza	^	^	^	^	^	43,328,940

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence in
^ Statistic not displayed due to fewer than 10 cases.

Rate Session-1 Matrix-2

South Dakota Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	19,225,44
Tuberculosis	^	^	^	^	^	19,225,44
Syphilis	^	^	^	^	^	19,225,44
Human Immunodeficienc	^	^	^	^	^	19,225,44
Septicemia	^	^	^	^	^	19,225,44
Other Infectious and Para	^	^	^	^	^	19,225,44
Diabetes Mellitus	^	^	^	^	^	19,225,44
Alzheimers (ICD-9 and 10	^	^	^	^	^	19,225,44
Diseases of Heart	209.8	1.0	207.9	211.7	48,851	19,225,44
Hypertension without Hea	^	^	^	^	^	19,225,44
Cerebrovascular Disease	^	^	^	^	^	19,225,44
Atherosclerosis	^	^	^	^	^	19,225,44
Aortic Aneurysm and Dis	^	^	^	^	^	19,225,44
Other Diseases of Arterie	^	^	^	^	^	19,225,44
Pneumonia and Influenza	^	^	^	^	^	19,225,44

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence
^ Statistic not displayed due to fewer than 10 cases.

Nebraska has higher mortality counts than South Dakota. South Dakota has 209.8 deaths per 100,000 and Nebraska has 208.3 deaths per 100,000. South Dakota has a slightly higher mortality rate due to heart disease than Nebraska.

Rate Session-1 Matrix-3

North Dakota Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	16,436,834
Tuberculosis	^	^	^	^	^	16,436,834
Syphilis	^	^	^	^	^	16,436,834
Human Immunodeficienc	^	^	^	^	^	16,436,834
Septicemia	^	^	^	^	^	16,436,834
Other Infectious and Para	^	^	^	^	^	16,436,834
Diabetes Mellitus	^	^	^	^	^	16,436,834
Alzheimers (ICD-9 and 10	^	^	^	^	^	16,436,834
Diseases of Heart	202.1	1.0	200.1	204.1	40,961	16,436,834
Hypertension without Hea	^	^	^	^	^	16,436,834
Cerebrovascular Disease	^	^	^	^	^	16,436,834
Atherosclerosis	^	^	^	^	^	16,436,834
Aortic Aneurysm and Dis	^	^	^	^	^	16,436,834
Other Diseases of Arterie	^	^	^	^	^	16,436,834

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence interva
^ Statistic not displayed due to fewer than 10 cases.

Rate Session-1 Matrix-2

South Dakota Heart Disease Mortality

	Rate	SE	Lower CI	Upper CI	Count	Pop
In situ, benign or unknow	^	^	^	^	^	19,225,44
Tuberculosis	^	^	^	^	^	19,225,44
Syphilis	^	^	^	^	^	19,225,44
Human Immunodeficienc	^	^	^	^	^	19,225,44
Septicemia	^	^	^	^	^	19,225,44
Other Infectious and Para	^	^	^	^	^	19,225,44
Diabetes Mellitus	^	^	^	^	^	19,225,44
Alzheimers (ICD-9 and 10	^	^	^	^	^	19,225,44
Diseases of Heart	209.8	1.0	207.9	211.7	48,851	19,225,44
Hypertension without Hea	^	^	^	^	^	19,225,44
Cerebrovascular Disease	^	^	^	^	^	19,225,44
Atherosclerosis	^	^	^	^	^	19,225,44
Aortic Aneurysm and Dis	^	^	^	^	^	19,225,44
Other Diseases of Arterie	^	^	^	^	^	19,225,44
Pneumonia and Influenza	^	^	^	^	^	19,225,44

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence
^ Statistic not displayed due to fewer than 10 cases.

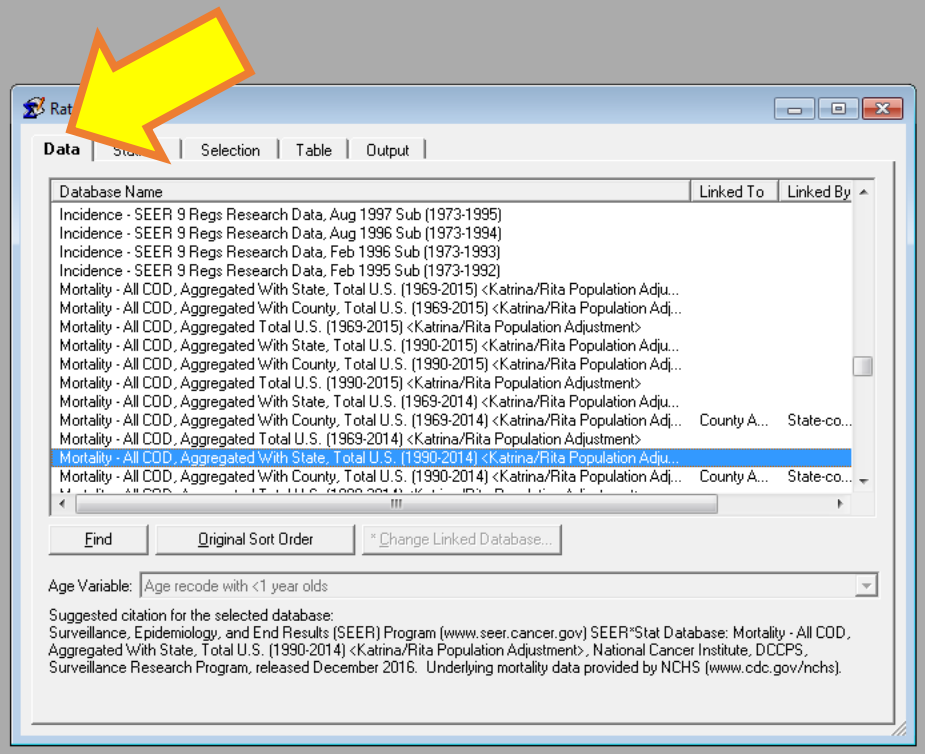
North Dakota has a lower count of heart disease deaths and a lower mortality rate than South Dakota for heart disease mortality.

In our comparisons, South Dakota had the highest mortality rate, followed by Nebraska. North Dakota had the lowest mortality rate of these states from 1990-2014.

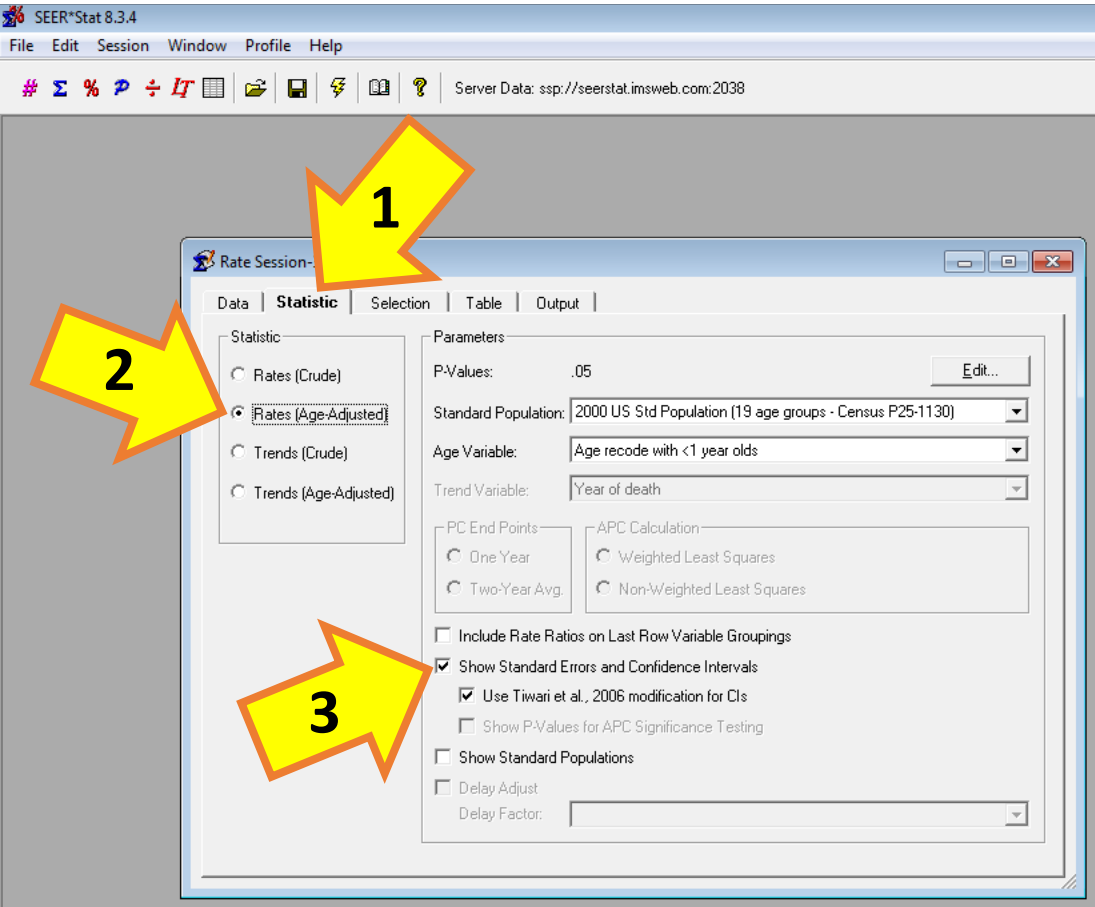
Mortality Counts and Rates (cont.)

- To provide another example, we will look at heart diseases mortality counts and age-adjusted rates in different races within the same state, to be able to compare them:
  - American Indians
  - Whites

- Again, this mortality example was completed using on-line SEER data, so some screens and output may look different if you're using downloaded binary SEER data.



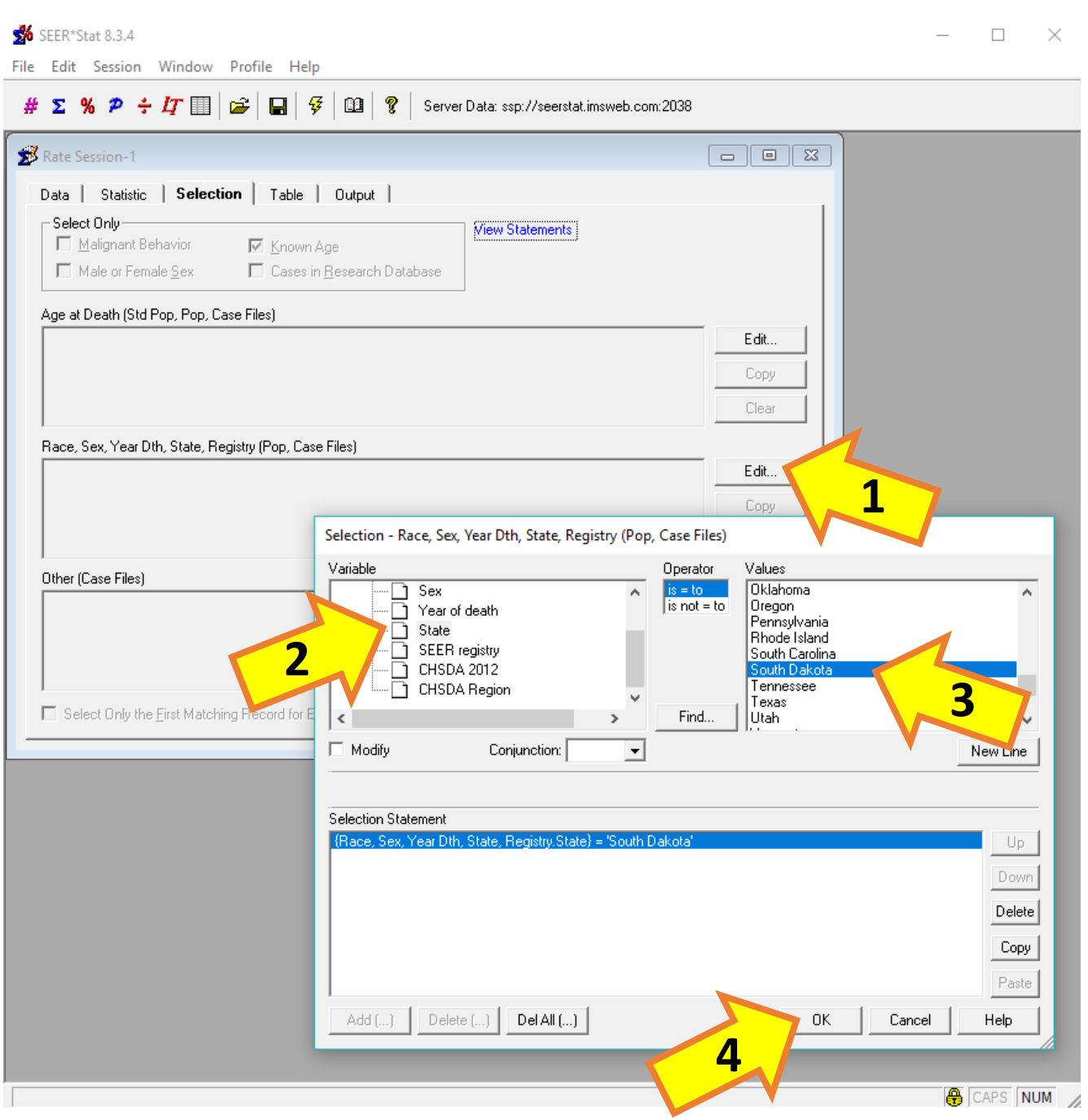
- Start a new rate session.
- Select “Mortality - All COD, Aggregated With County, Total U.S. (1990-2014) <Katrina/Rita Population Adjustment>”.

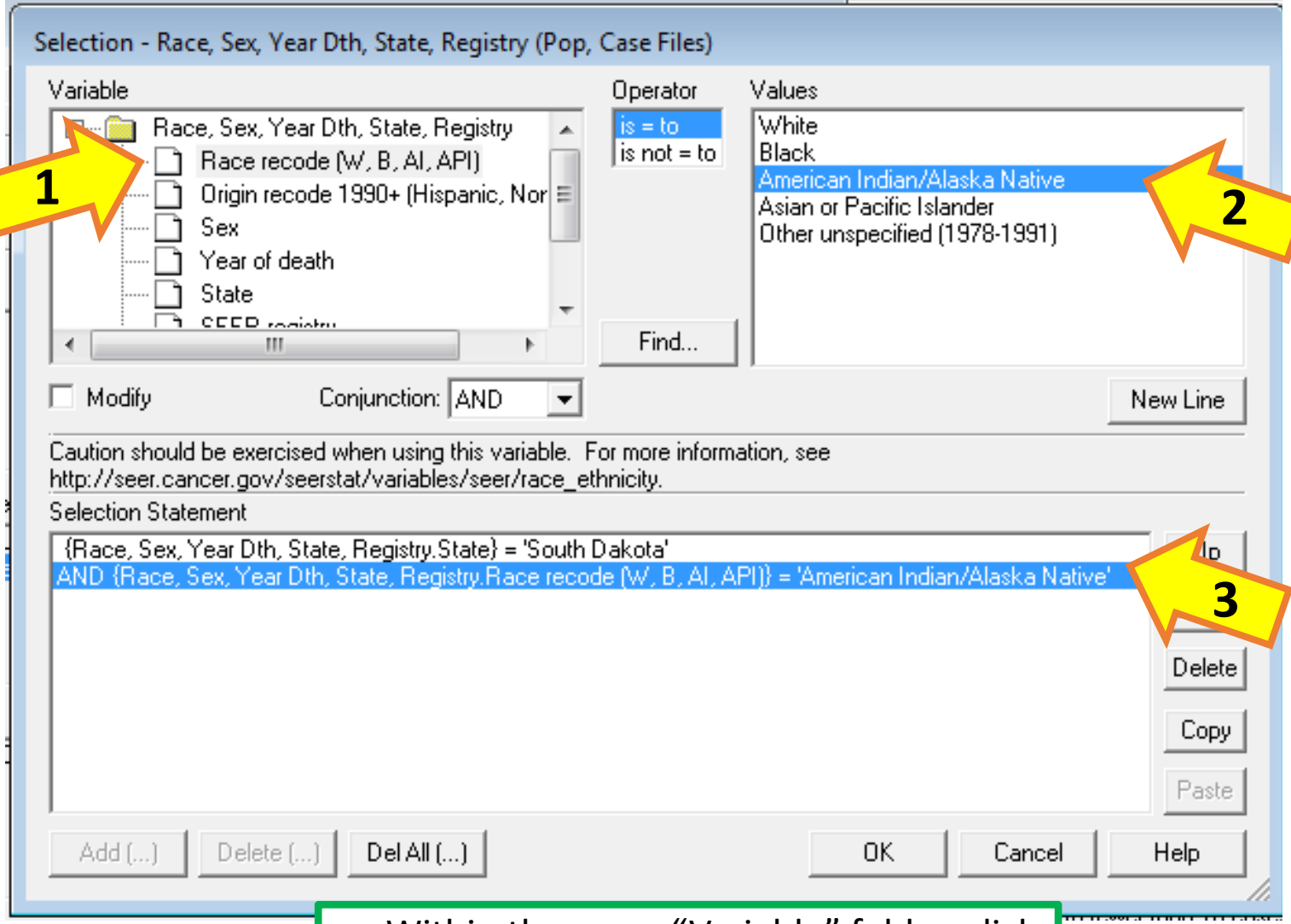


- In the "Statistic" tab, choose an age-adjusted rate, standardized to the 2000 US Population, with standard errors and confidence intervals.

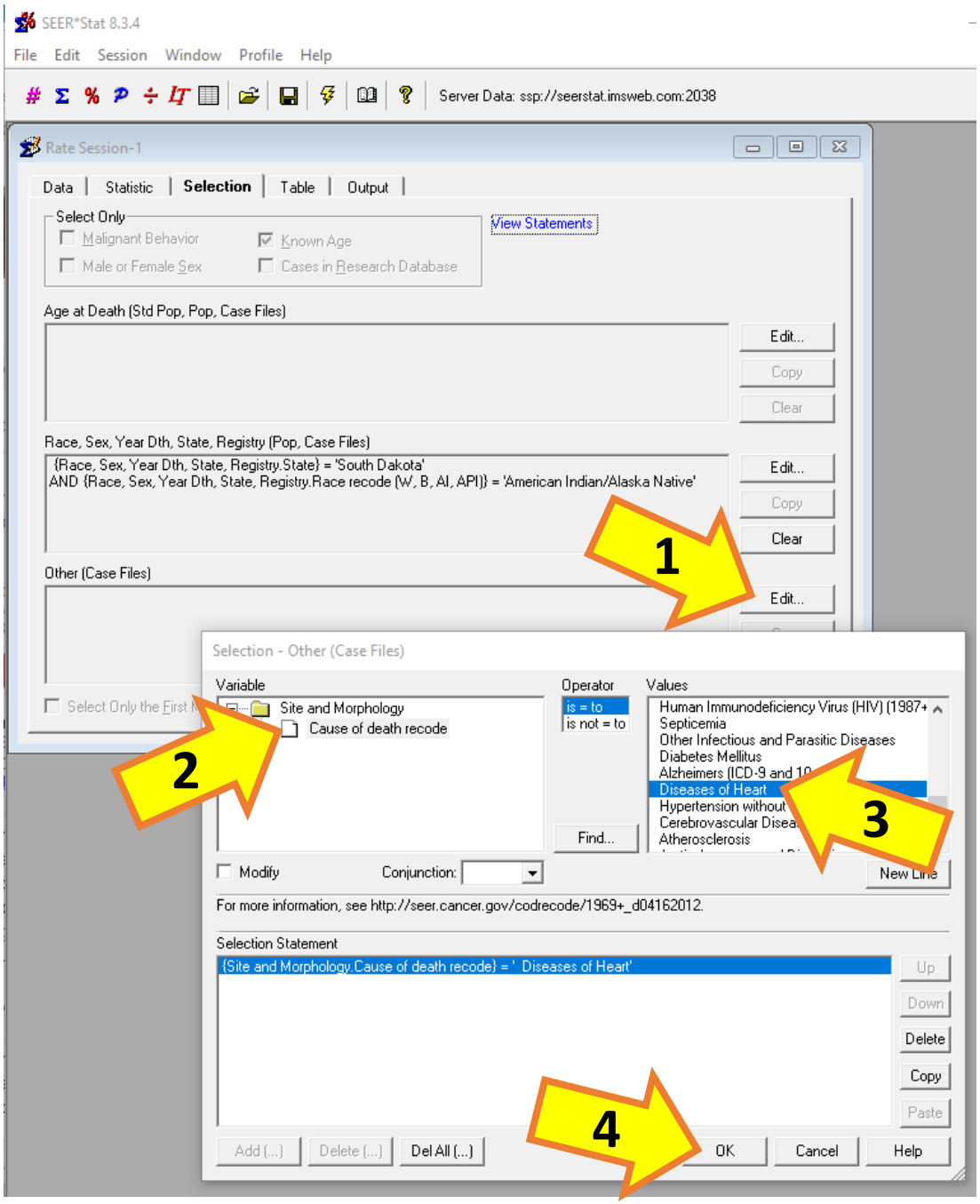
Using an age-adjusted rate adapts the crude rates to standard population age groups. This allows us to accurately compare rates between populations that may not actually have the same age distribution.

- In the “Selection” tab, click “Edit” on the “Race, Sex, Year Dth, State, Registry” category.
- Open the folder titled “Race, Sex, Year Dth, State, Registry” in the “Variables” window.
- Select “State” in this folder.
- Select “South Dakota” in the “Values” window.

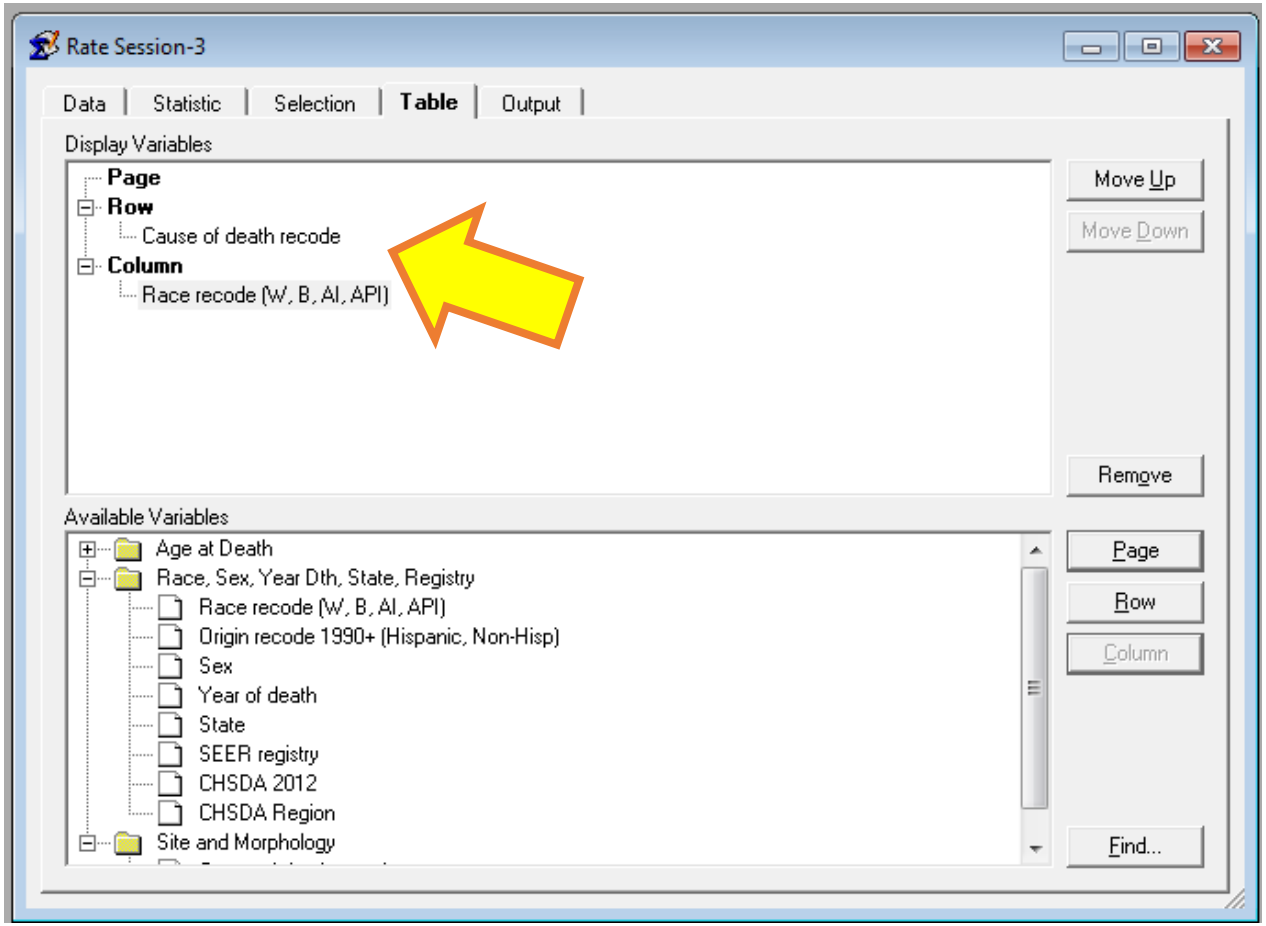




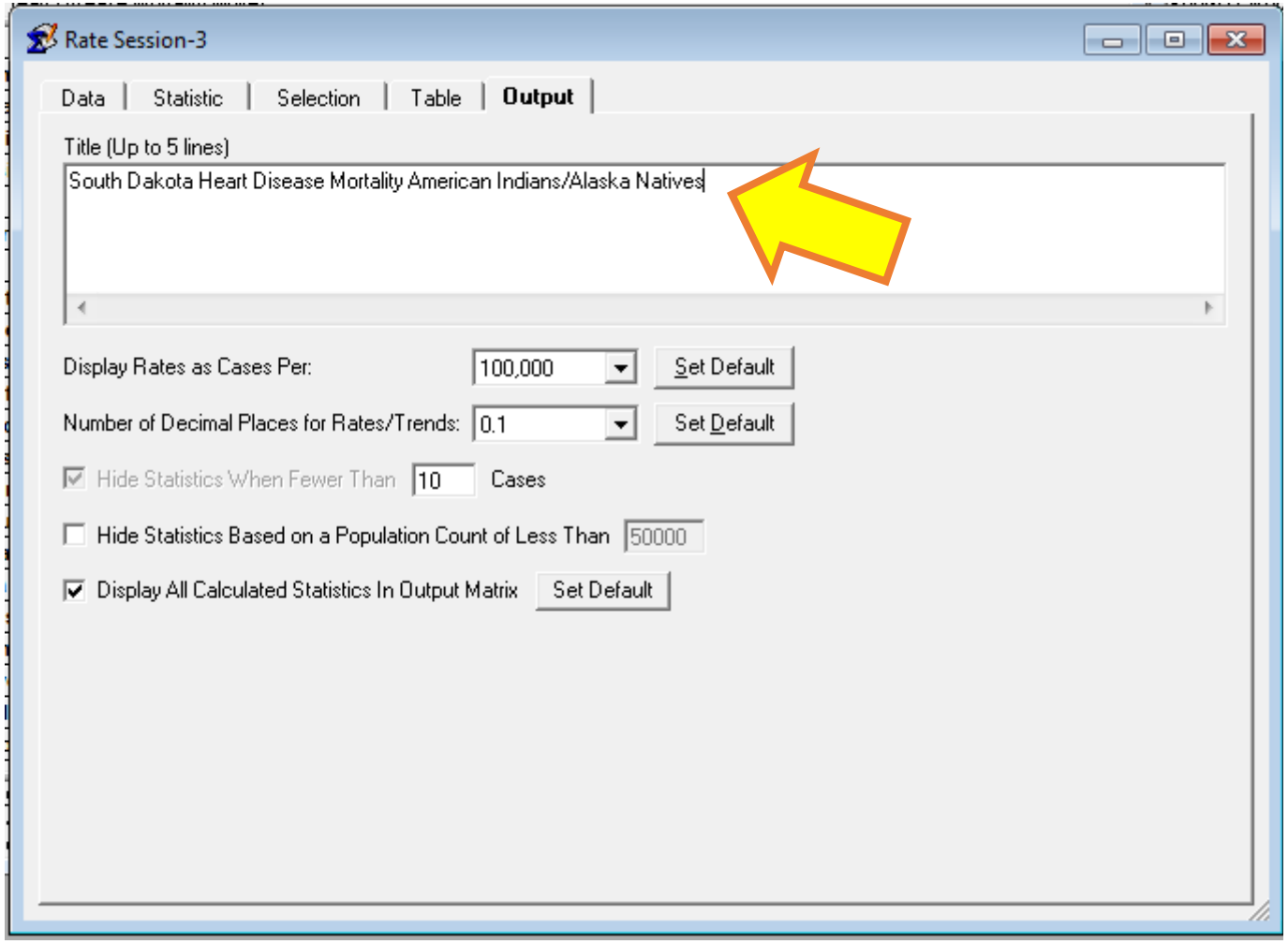
- Within the same “Variable” folder, click on “Race recode” .
- Choose “American Indian/Alaska Native”.
- Click “OK”.



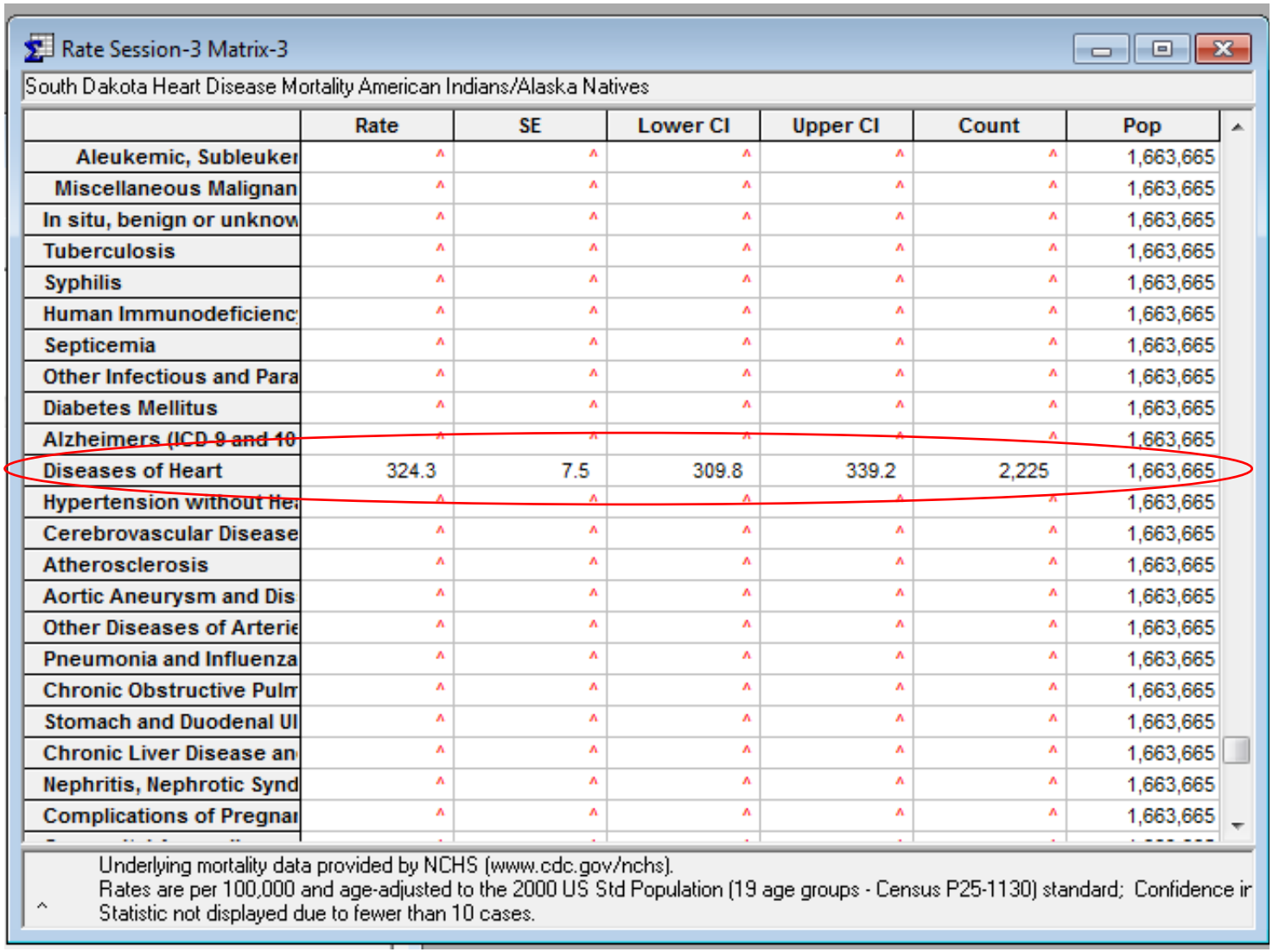
- Click “Edit” in the “Other” category.
- Select “Cause of death recode” in the “Site and Morphology” folder.
- Select “Diseases of Heart” in the “Values” window.
- Click “OK”.



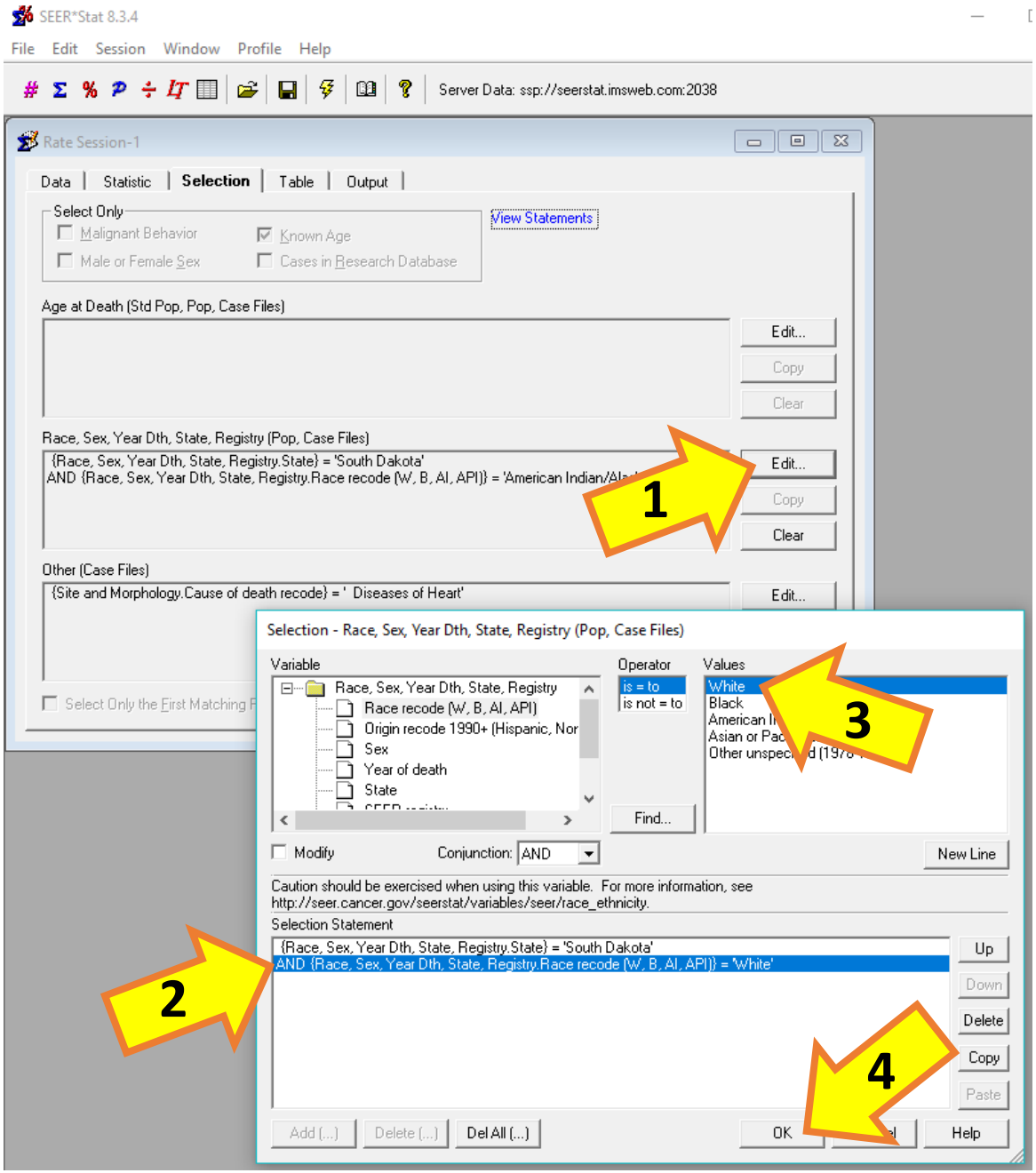
- Go to the Table tab.
- Choose “Cause of death recode” as the row variable.
- Choose “Race recode” as the column variable.



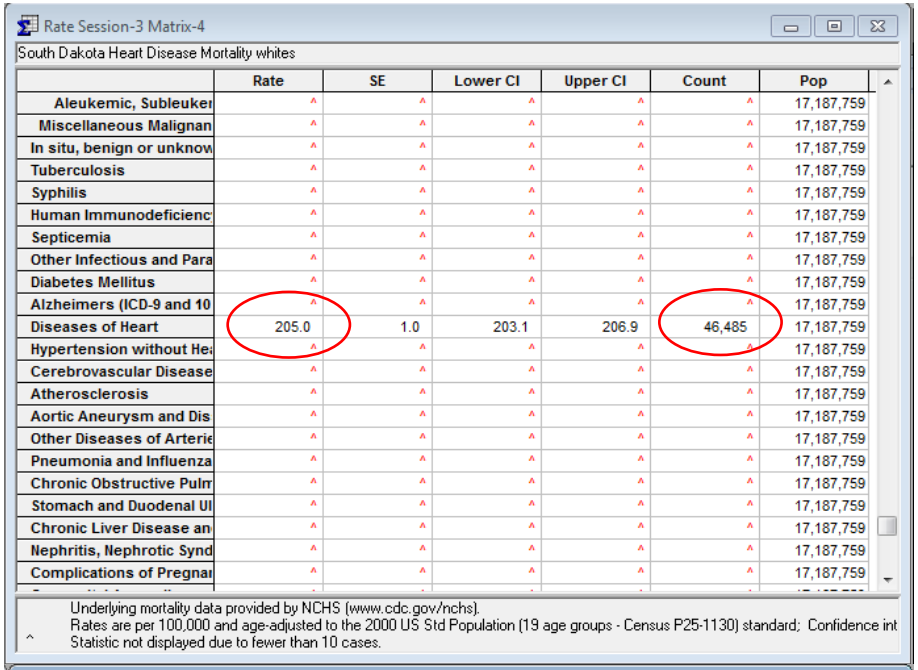
- Go to the “Output” tab.
- Give the table the title “South Dakota Heart Disease Mortality American Indians/Alaska Natives”.



- Click the lightning bolt icon to produce the table.



- Return to the “Selection” tab in the same rate session window.
- Click “Edit” for the “Race, Sex, Year Dth, State, Registry” category.
- Click the “AND {Race, Sex,…” option in the Selection Statement box
- Select “White” in the “Values” window.
- Rename the table in the “Output” tab.
- Click the lightning bolt icon to produce the second table.

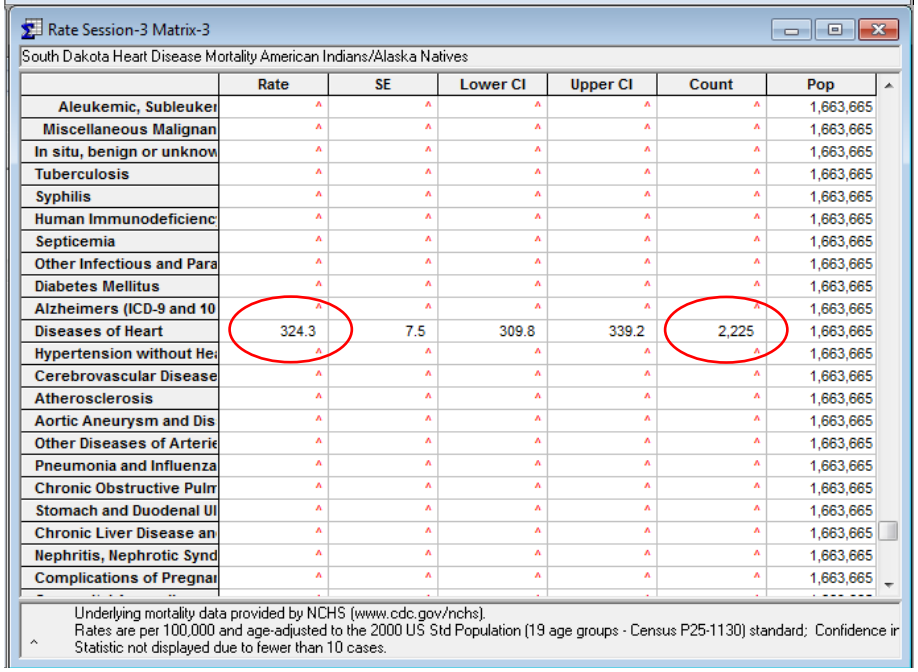


Rate Session-3 Matrix-4

South Dakota Heart Disease Mortality whites

	Rate	SE	Lower CI	Upper CI	Count	Pop
Aleukemic, Subleuker	^	^	^	^	^	17,187,759
Miscellaneous Malignan	^	^	^	^	^	17,187,759
In situ, benign or unknow	^	^	^	^	^	17,187,759
Tuberculosis	^	^	^	^	^	17,187,759
Syphilis	^	^	^	^	^	17,187,759
Human Immunodeficienc	^	^	^	^	^	17,187,759
Septicemia	^	^	^	^	^	17,187,759
Other Infectious and Para	^	^	^	^	^	17,187,759
Diabetes Mellitus	^	^	^	^	^	17,187,759
Alzheimers (ICD-9 and 10	^	^	^	^	^	17,187,759
Diseases of Heart	205.0	1.0	203.1	206.9	46,485	17,187,759
Hypertension without Hea	^	^	^	^	^	17,187,759
Cerebrovascular Disease	^	^	^	^	^	17,187,759
Atherosclerosis	^	^	^	^	^	17,187,759
Aortic Aneurysm and Dis	^	^	^	^	^	17,187,759
Other Diseases of Arterie	^	^	^	^	^	17,187,759
Pneumonia and Influenza	^	^	^	^	^	17,187,759
Chronic Obstructive Pulm	^	^	^	^	^	17,187,759
Stomach and Duodenal Ul	^	^	^	^	^	17,187,759
Chronic Liver Disease an	^	^	^	^	^	17,187,759
Nephritis, Nephrotic Synd	^	^	^	^	^	17,187,759
Complications of Pregnar	^	^	^	^	^	17,187,759

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence int
^ Statistic not displayed due to fewer than 10 cases.



Rate Session-3 Matrix-3

South Dakota Heart Disease Mortality American Indians/Alaska Natives

	Rate	SE	Lower CI	Upper CI	Count	Pop
Aleukemic, Subleuker	^	^	^	^	^	1,663,665
Miscellaneous Malignan	^	^	^	^	^	1,663,665
In situ, benign or unknow	^	^	^	^	^	1,663,665
Tuberculosis	^	^	^	^	^	1,663,665
Syphilis	^	^	^	^	^	1,663,665
Human Immunodeficienc	^	^	^	^	^	1,663,665
Septicemia	^	^	^	^	^	1,663,665
Other Infectious and Para	^	^	^	^	^	1,663,665
Diabetes Mellitus	^	^	^	^	^	1,663,665
Alzheimers (ICD-9 and 10	^	^	^	^	^	1,663,665
Diseases of Heart	324.3	7.5	309.8	339.2	2,225	1,663,665
Hypertension without Hea	^	^	^	^	^	1,663,665
Cerebrovascular Disease	^	^	^	^	^	1,663,665
Atherosclerosis	^	^	^	^	^	1,663,665
Aortic Aneurysm and Dis	^	^	^	^	^	1,663,665
Other Diseases of Arterie	^	^	^	^	^	1,663,665
Pneumonia and Influenza	^	^	^	^	^	1,663,665
Chronic Obstructive Pulm	^	^	^	^	^	1,663,665
Stomach and Duodenal Ul	^	^	^	^	^	1,663,665
Chronic Liver Disease an	^	^	^	^	^	1,663,665
Nephritis, Nephrotic Synd	^	^	^	^	^	1,663,665
Complications of Pregnar	^	^	^	^	^	1,663,665

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence ir
^ Statistic not displayed due to fewer than 10 cases.

With both tables produced, we can now compare heart disease mortality between American Indians/Alaska Natives and whites in South Dakota.

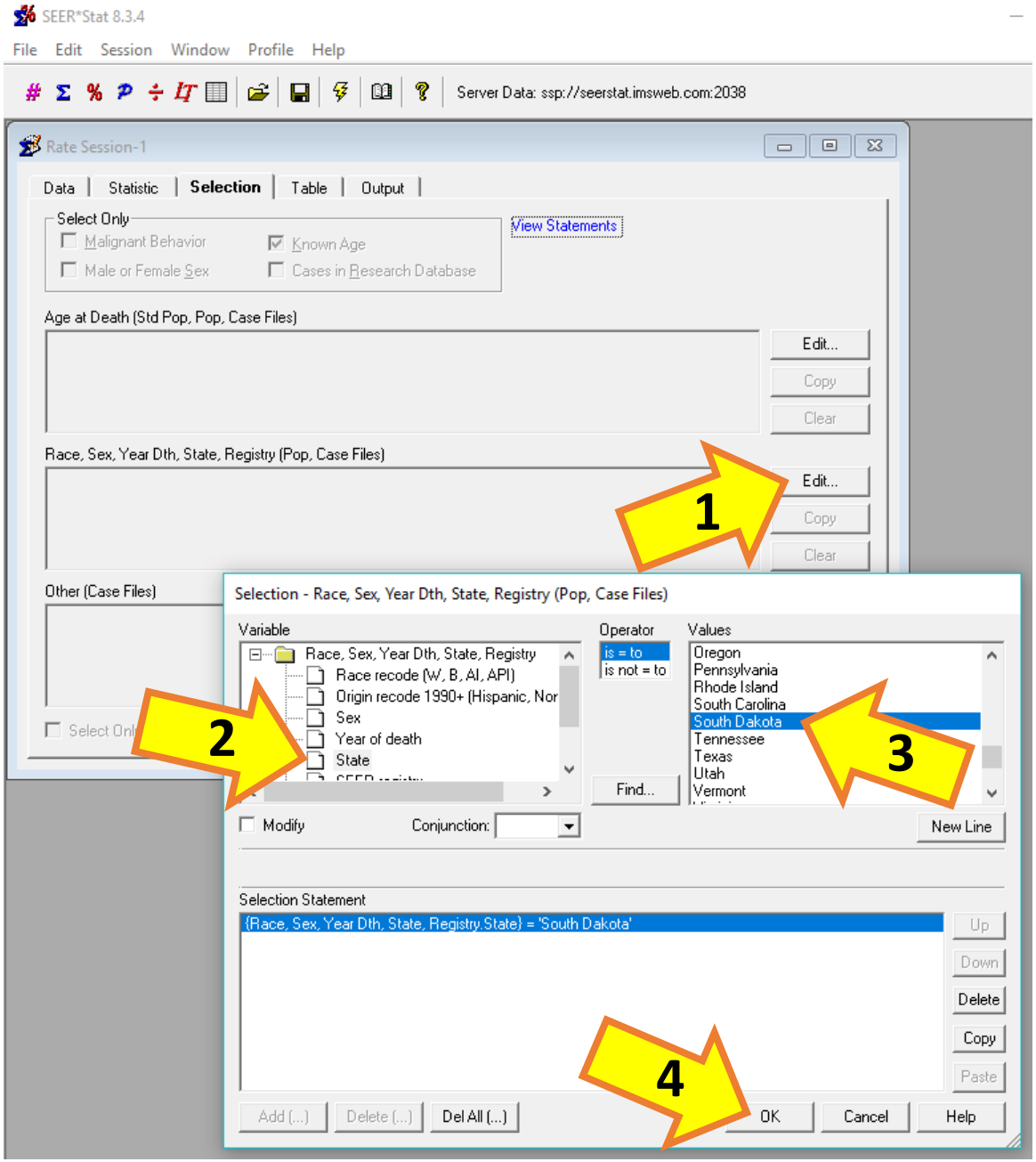
American Indians/Alaska Natives have a lower count of heart disease compared to whites in South Dakota. However, when we look at the mortality rate per 100,000 population, we see that the heart disease mortality rate for American Indians/Alaska Natives is 324.3 deaths per 100,000, compared to the lower 205.0 deaths per 100,000 for whites for the period of 1990-2014.

What if we want to compare age-adjusted mortality for two races in the same table?

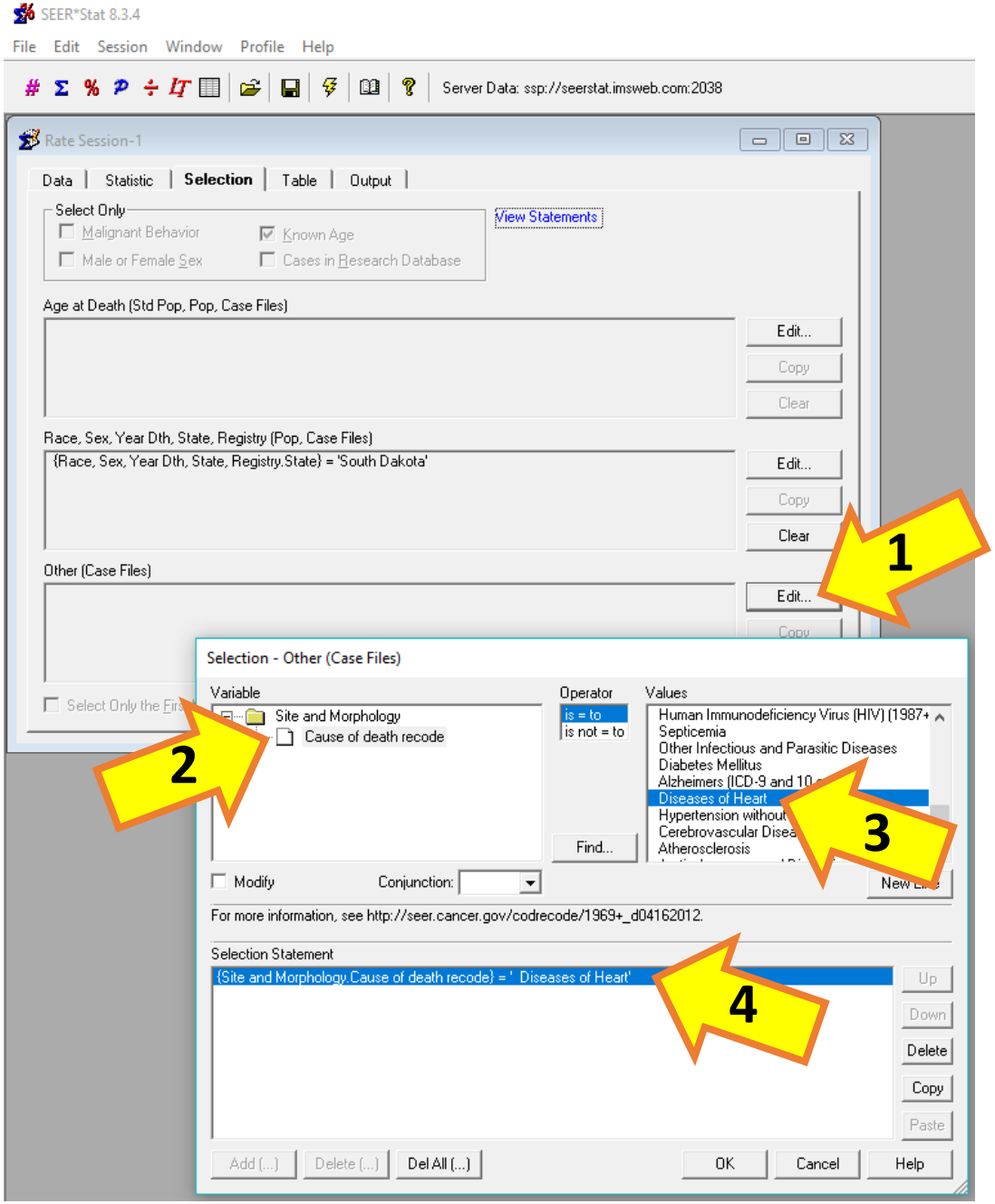
We will start by selecting the same population in the “Data” tab, and following the same steps as before for the “Statistic” tab.

- Start a new rate session.
- Select “Mortality - All COD, Aggregated With County, Total U.S. (1990-2014) <Katrina/Rita Population Adjustment>”.

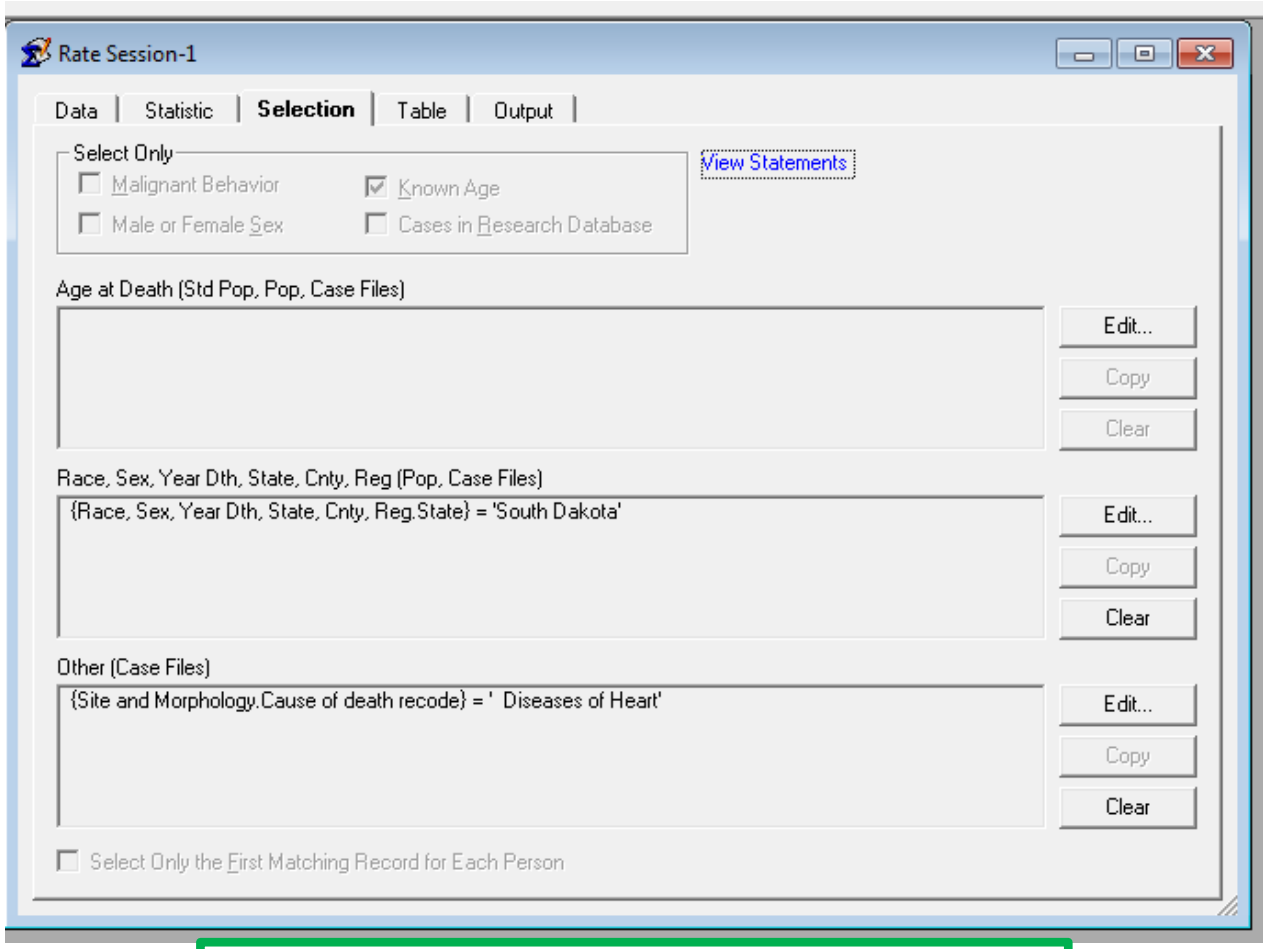
- In the “Statistic” tab, choose an age-adjusted rate, standardized to the 2000 US Population, with standard errors and confidence intervals.



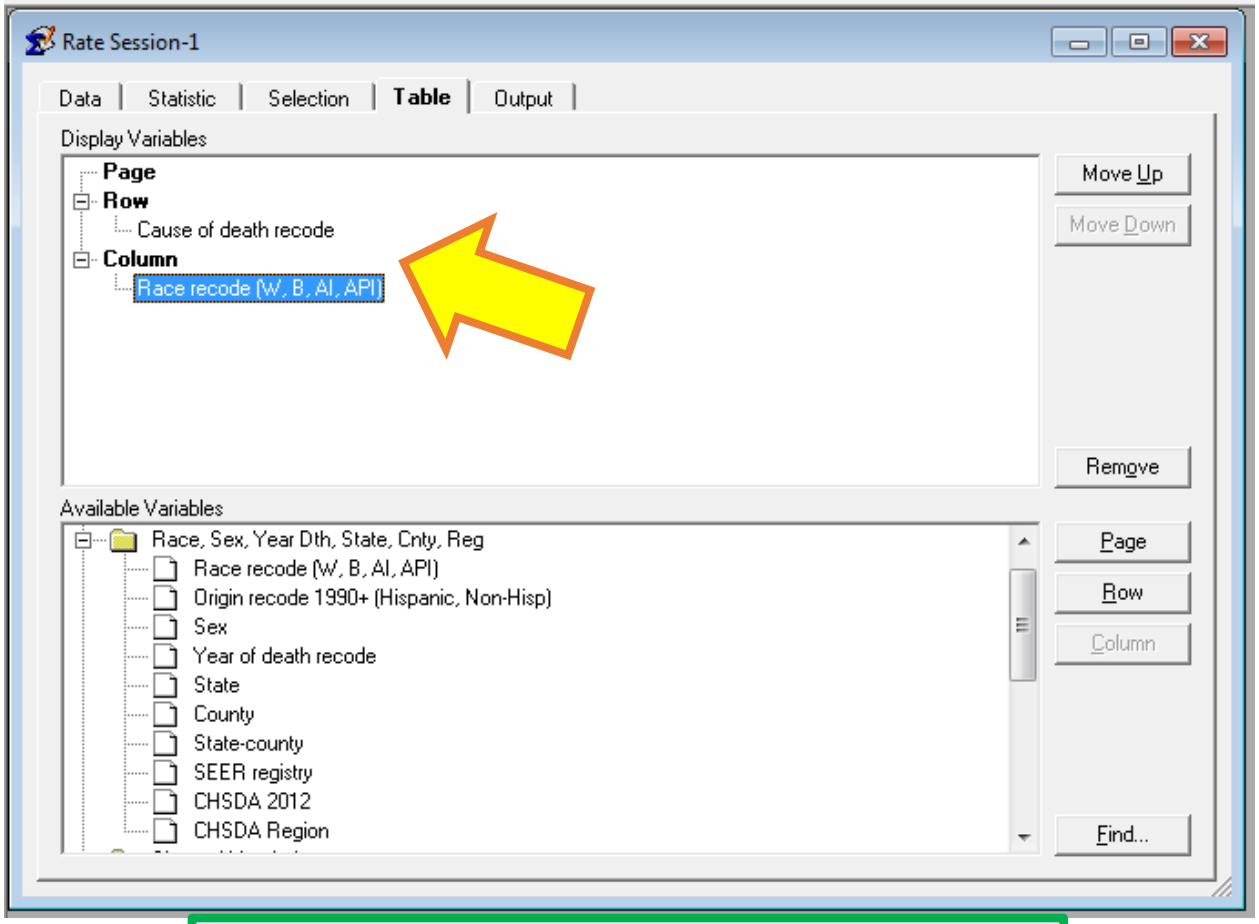
- In the “Selection” tab, click “Edit” on the “Race, Sex, Year Dth, State, Registry” category.
- Open the folder titled “Race, Sex, Year Dth, State, Registry” in the “Variables” window.
- Select “State” in this folder.
- Select “South Dakota” in the “Values” window and click “OK”.



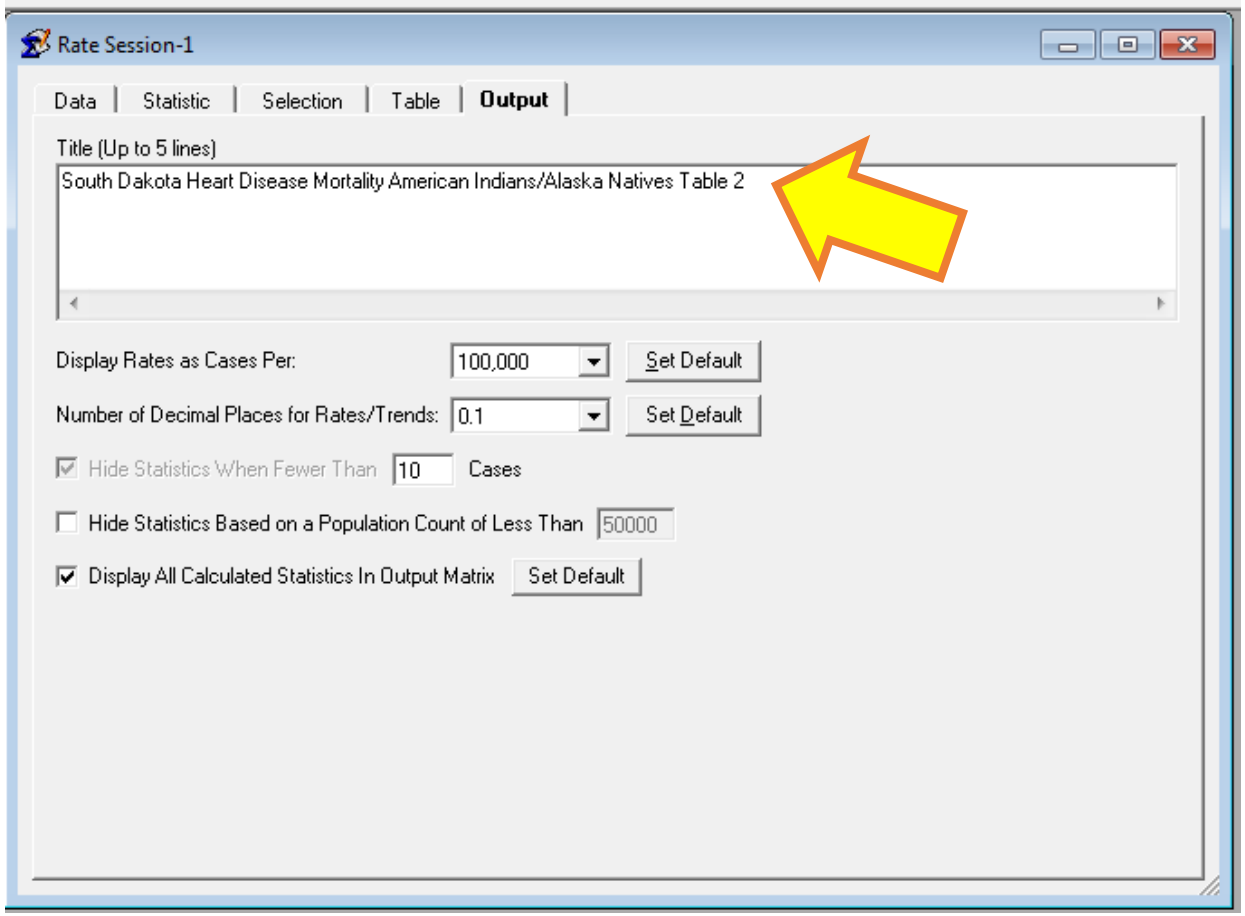
- Click “Edit” in the “Other” category.
- Select “Cause of death recode” in the “Site and Morphology” folder.
- Select “Diseases of Heart” in the “Values” window and click “OK”.



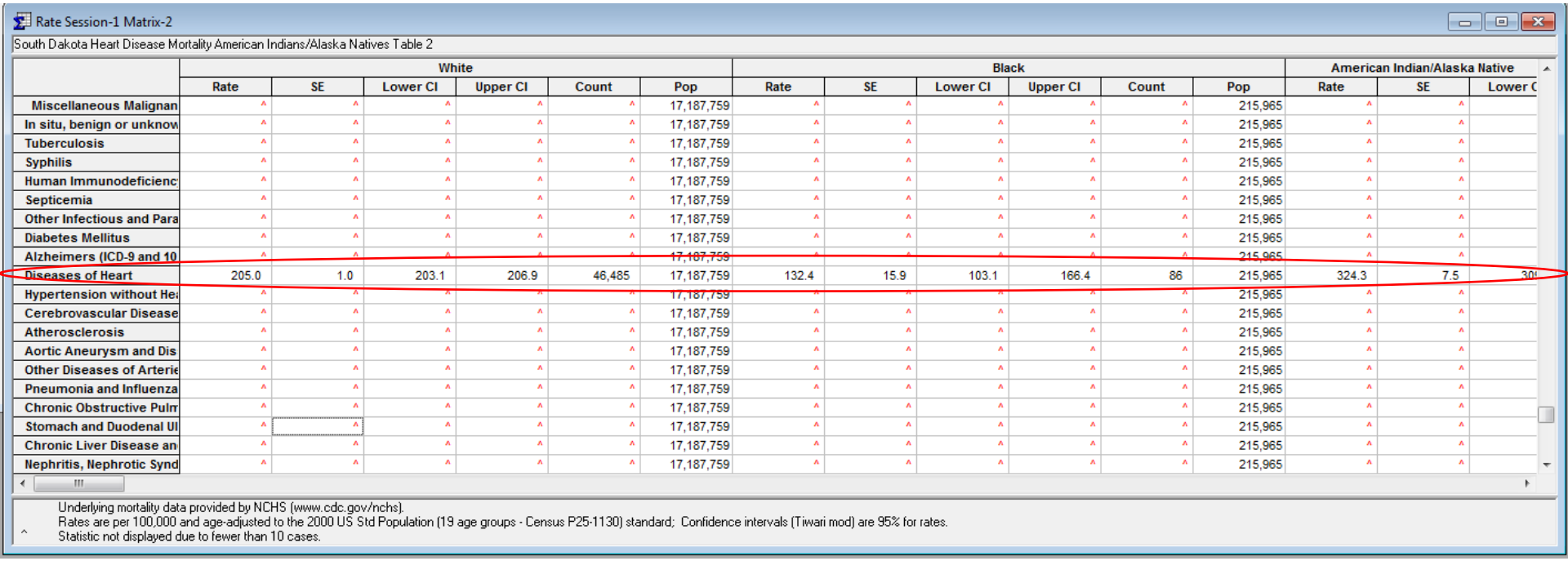
You may notice that we did not select the races that we want to compare to our "Race, Sex, Year Dth, State, Cnty, Reg" category. This will allow us to keep all races in the same table.



- Go to the Table tab.
- Choose “Cause of death recode” as the row variable.
- Choose “Race recode” as the column variable.



- Go to the “Output” tab.
- Give the table the title “South Dakota Heart Disease Mortality American Indians/Alaska Natives Table 2”.



Rate Session-1 Matrix-2

South Dakota Heart Disease Mortality American Indians/Alaska Natives Table 2

	White						Black						American Indian/Alaska Native		
	Rate	SE	Lower CI	Upper CI	Count	Pop	Rate	SE	Lower CI	Upper CI	Count	Pop	Rate	SE	Lower C
Miscellaneous Malignan	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
In situ, benign or unknow	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Tuberculosis	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Syphilis	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Human Immunodeficienc	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Septicemia	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Other Infectious and Para	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Diabetes Mellitus	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Alzheimers (ICD-9 and 10	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Diseases of Heart	205.0	1.0	203.1	206.9	46,485	17,187,759	132.4	15.9	103.1	166.4	86	215,965	324.3	7.5	30
Hypertension without He	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Cerebrovascular Disease	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Atherosclerosis	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Aortic Aneurysm and Dis	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Other Diseases of Arterie	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Pneumonia and Influenza	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Chronic Obstructive Pulm	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Stomach and Duodenal Ul	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Chronic Liver Disease an	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	
Nephritis, Nephrotic Synd	^	^	^	^	^	17,187,759	^	^	^	^	^	215,965	^	^	

Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
Rates are per 100,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard; Confidence intervals (Tiwari mod) are 95% for rates.
^ Statistic not displayed due to fewer than 10 cases.

- Click the lightning bolt icon to produce the table.

All races with data for heart disease mortality in South Dakota are now displayed in the same table.

Finding Age-Adjusted Incidence Rates

Example: Breast Cancer

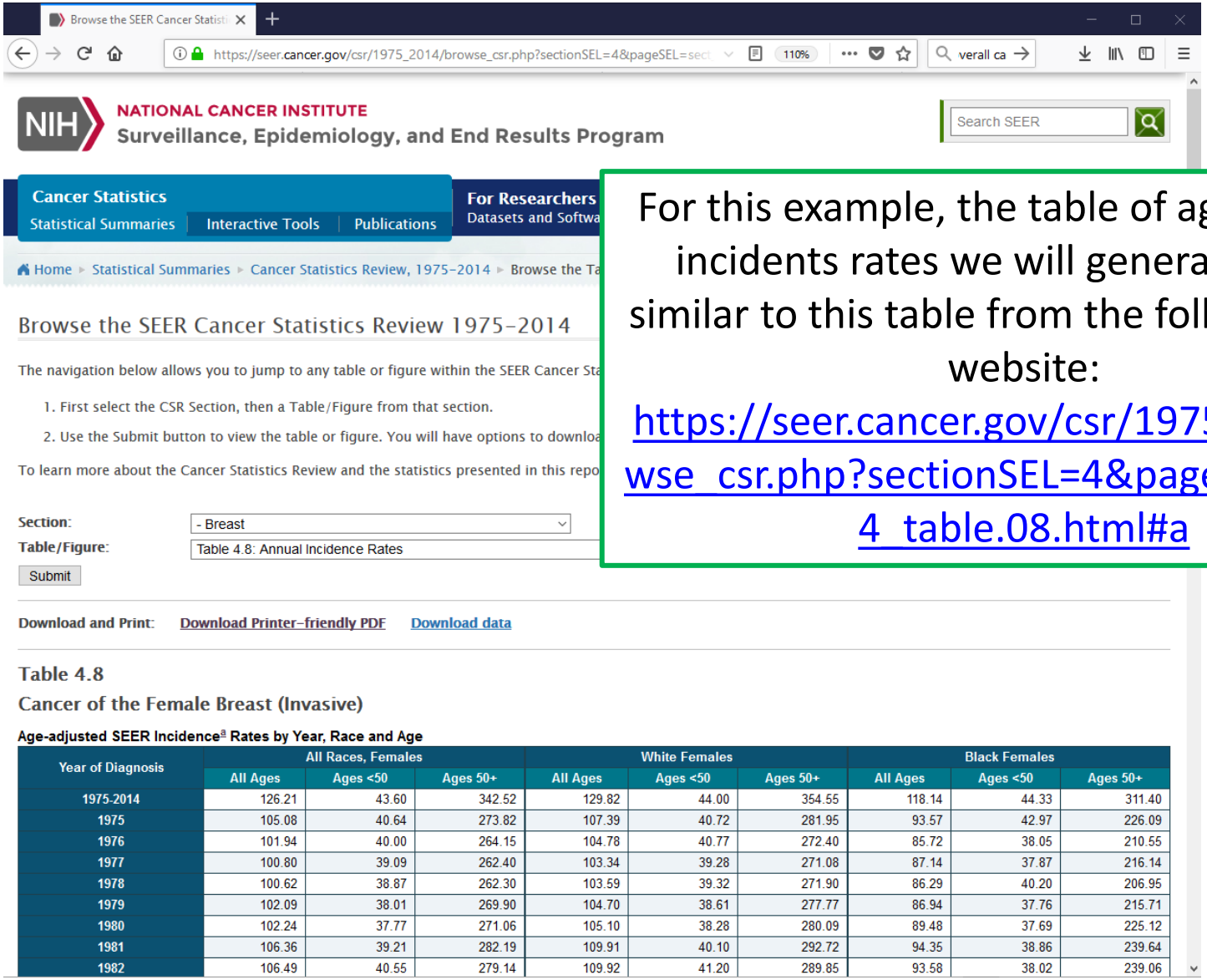
University of Nebraska Medical Center

Age-Adjusted Incidence Rates

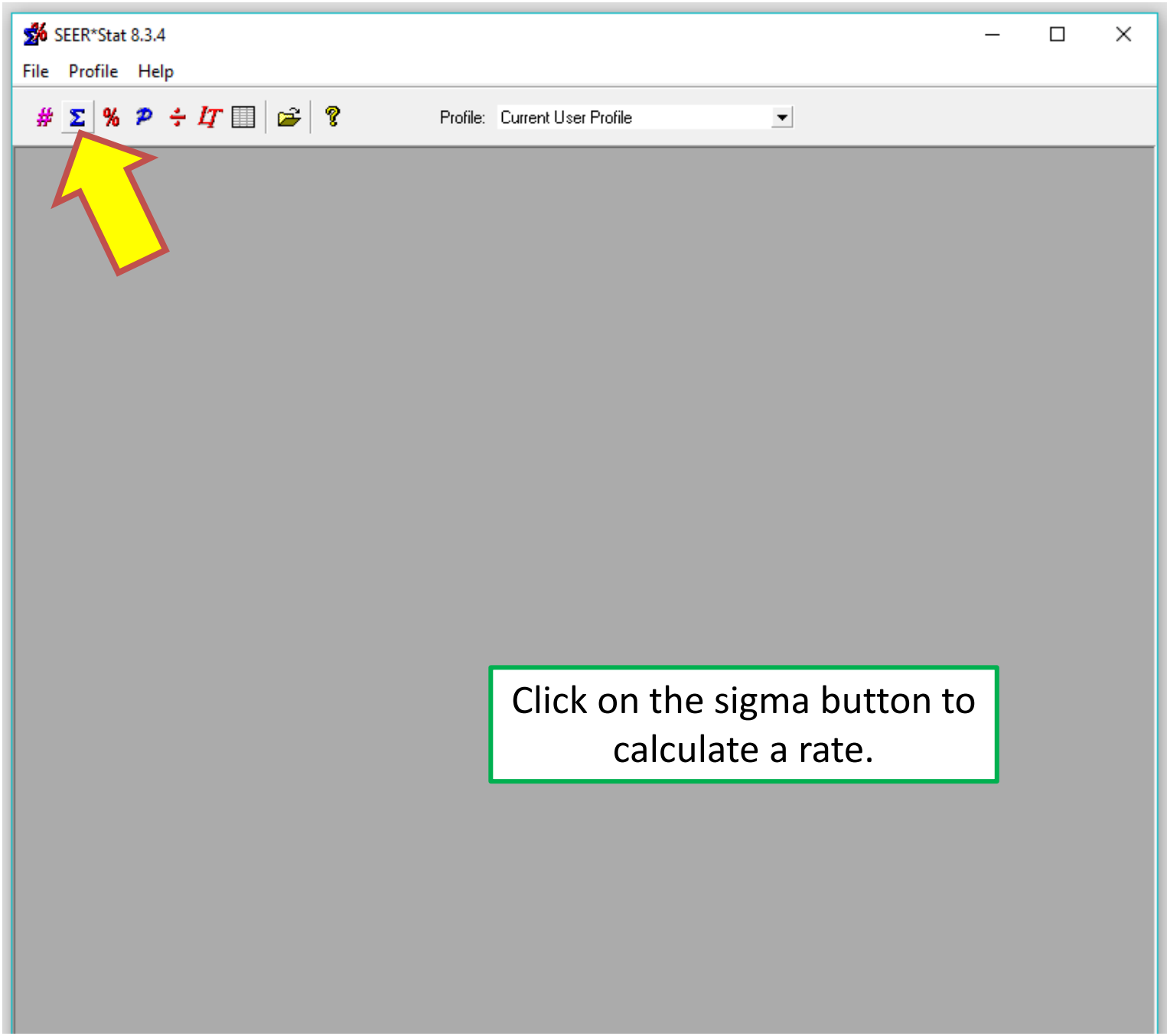
- This section will instruct you on how to get age-adjusted incidence rates using SEER data

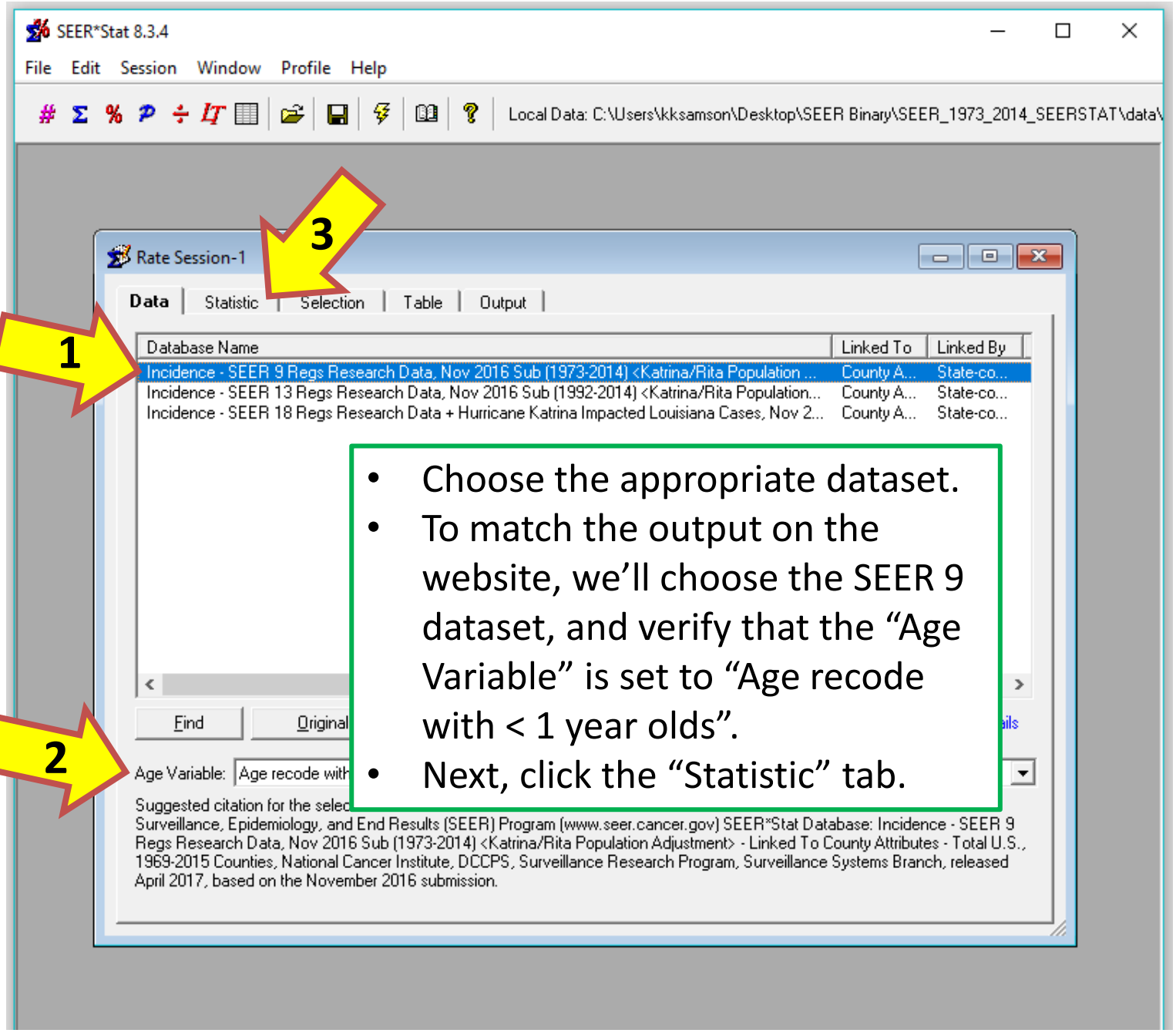
- For this example, we will show you to get access to breast cancer incidence counts using the SEER 9 registries data, with the following criteria:
  - Only include females
  - Exclude 'In Situ' cases
  - Display data by custom age categories
    - < 50 years or 50+ years
  - Display data separately by race
  - Show data by year (2000 – 2014)

- This example was done using downloaded binary SEER data for use in SEER*Stat. If you are using online data instead (i.e. entering your username and password), a couple of your screens and output may be slightly different from the tutorial.

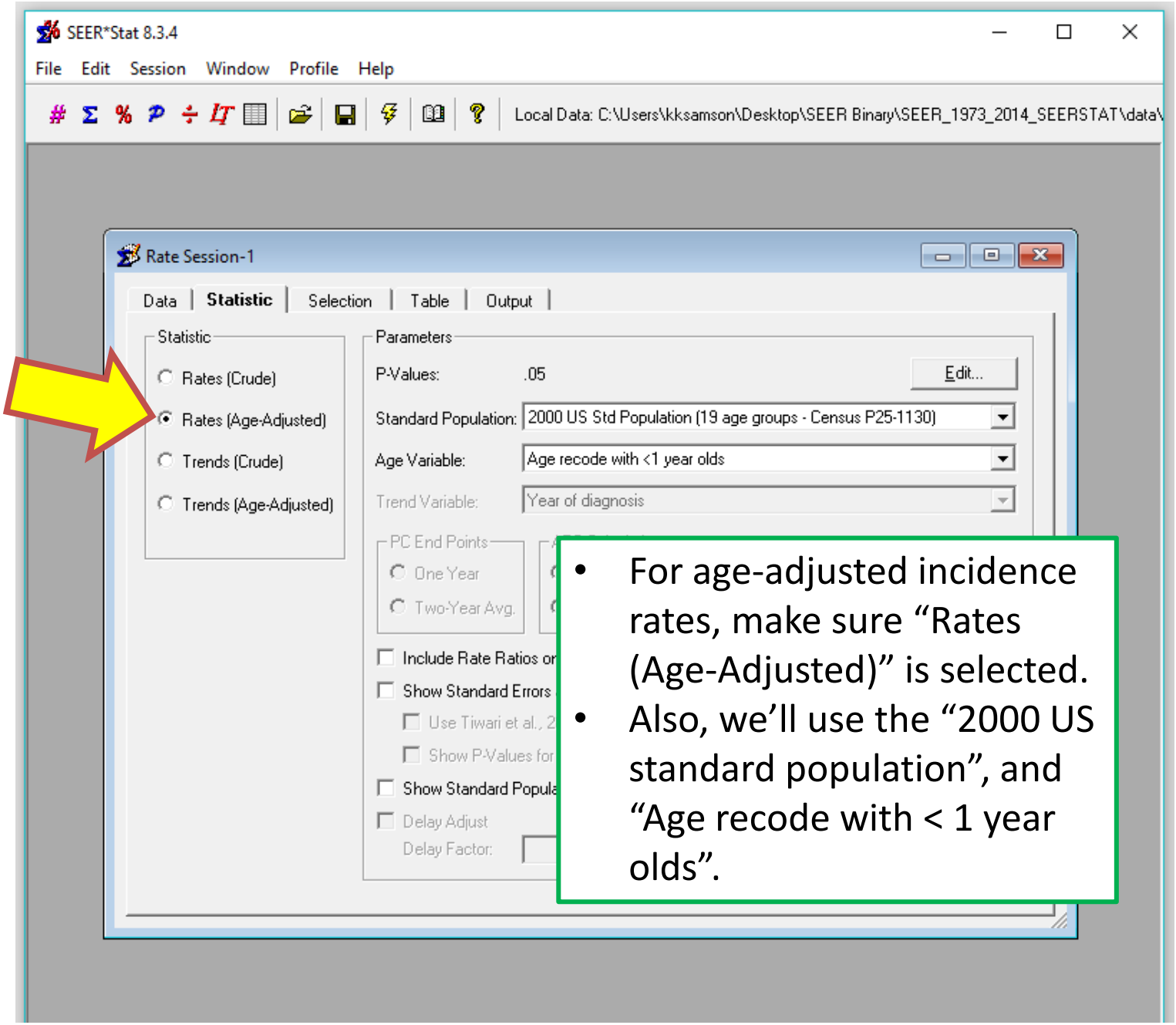


For this example, the table of age-adjusted incidents rates we will generate is very similar to this table from the following SEER website: https://seer.cancer.gov/csr/1975_2014/browse_csr.php?sectionSEL=4&pageSEL=sect_04_table.08.html#a

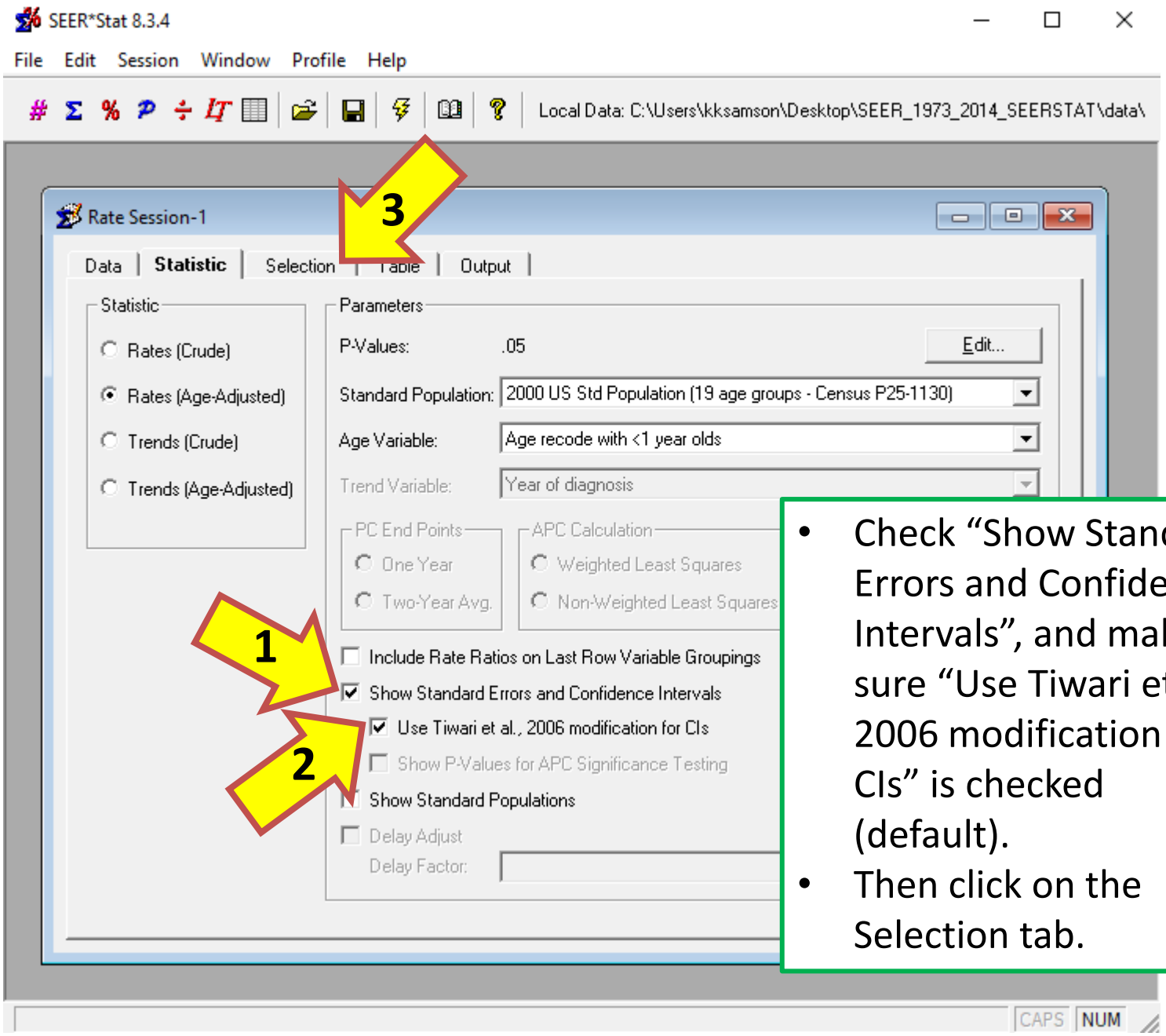

SEER*Stat 8.3.4
File
Profile
Help
Profile:
Current User Profile
Click on the sigma button to calculate a rate.



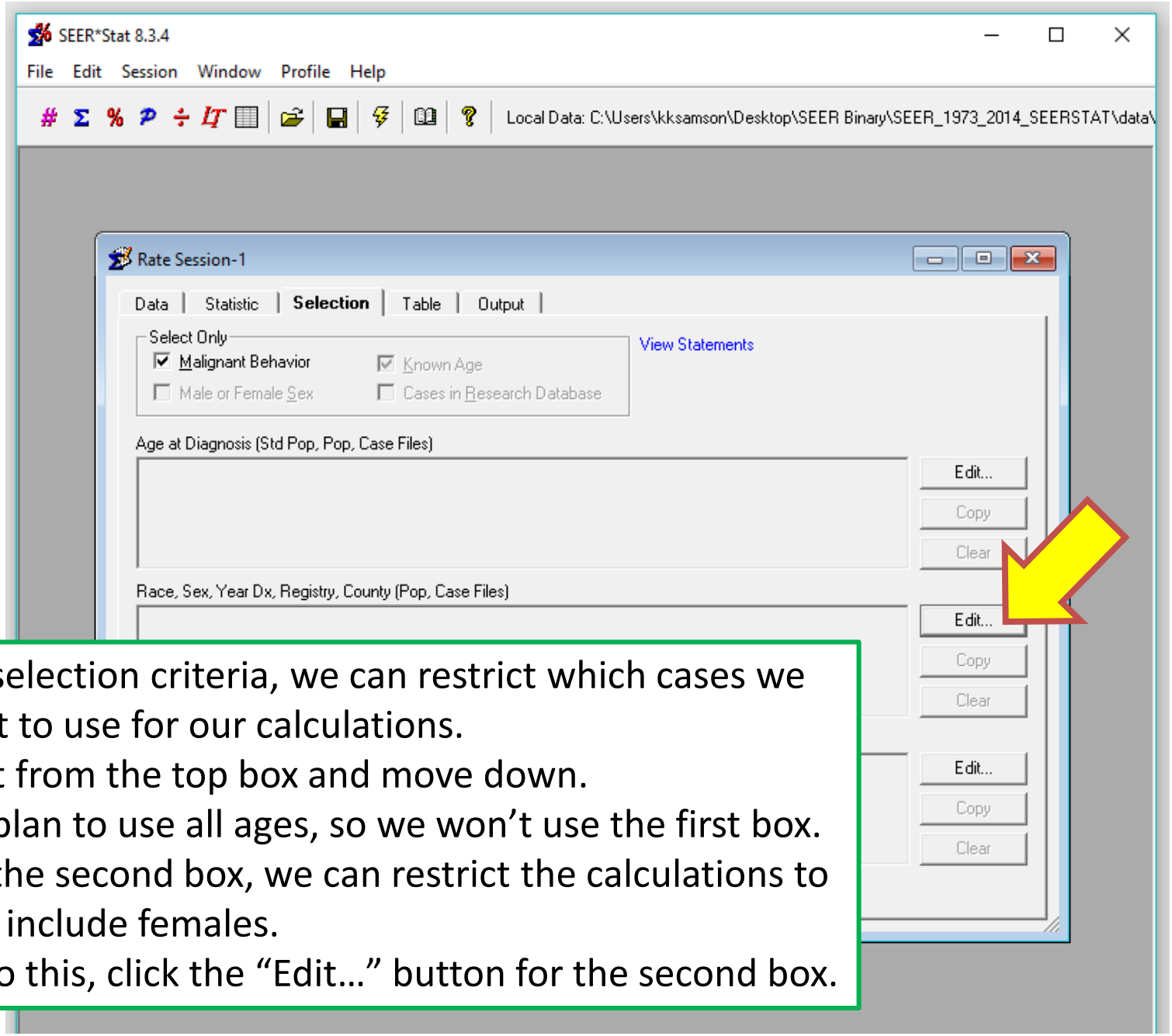
- Choose the appropriate dataset.
- To match the output on the website, we'll choose the SEER 9 dataset, and verify that the "Age Variable" is set to "Age recode with < 1 year olds".
- Next, click the "Statistic" tab.



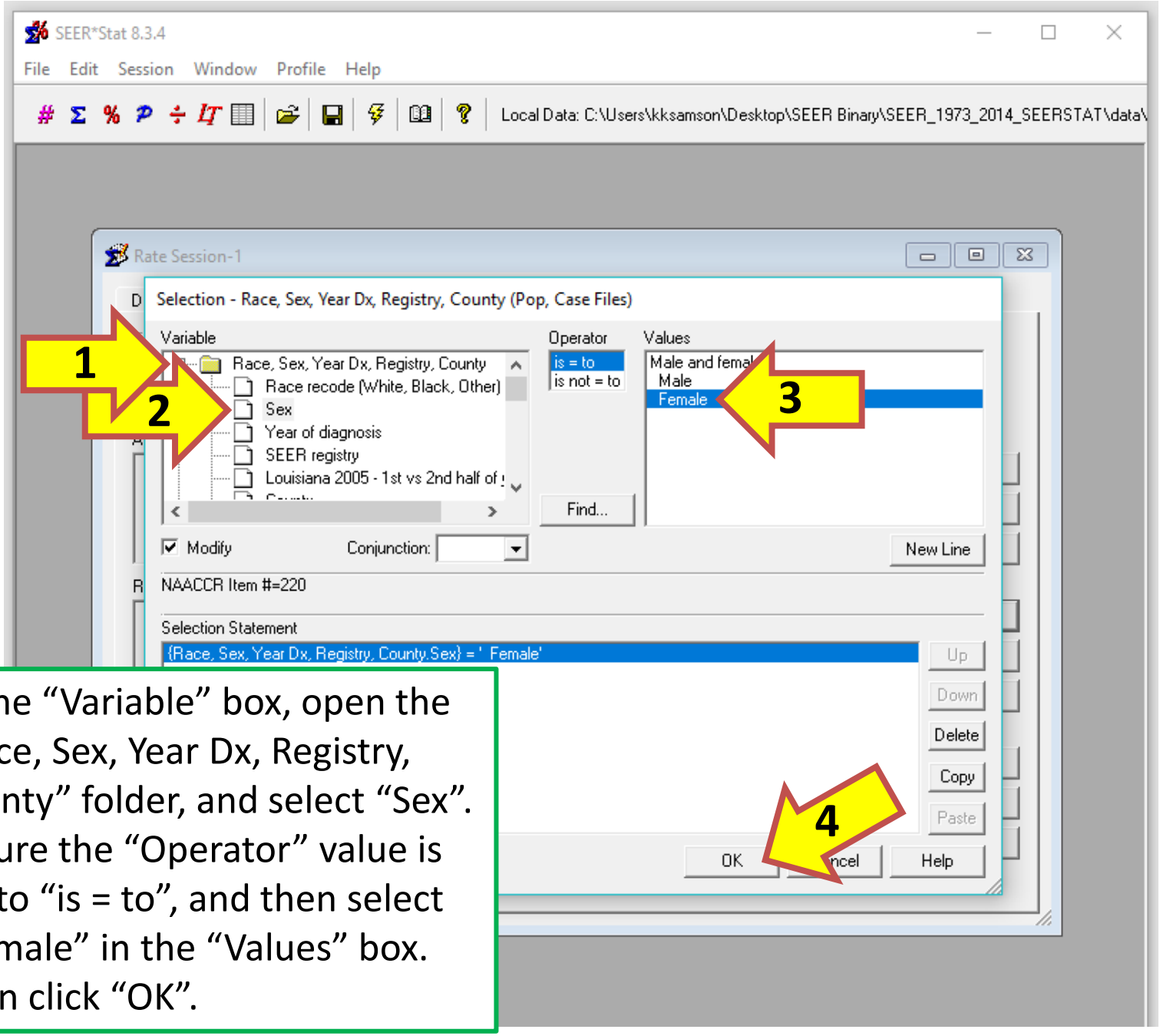
- For age-adjusted incidence rates, make sure “Rates (Age-Adjusted)” is selected.
- Also, we’ll use the “2000 US standard population”, and “Age recode with < 1 year olds”.



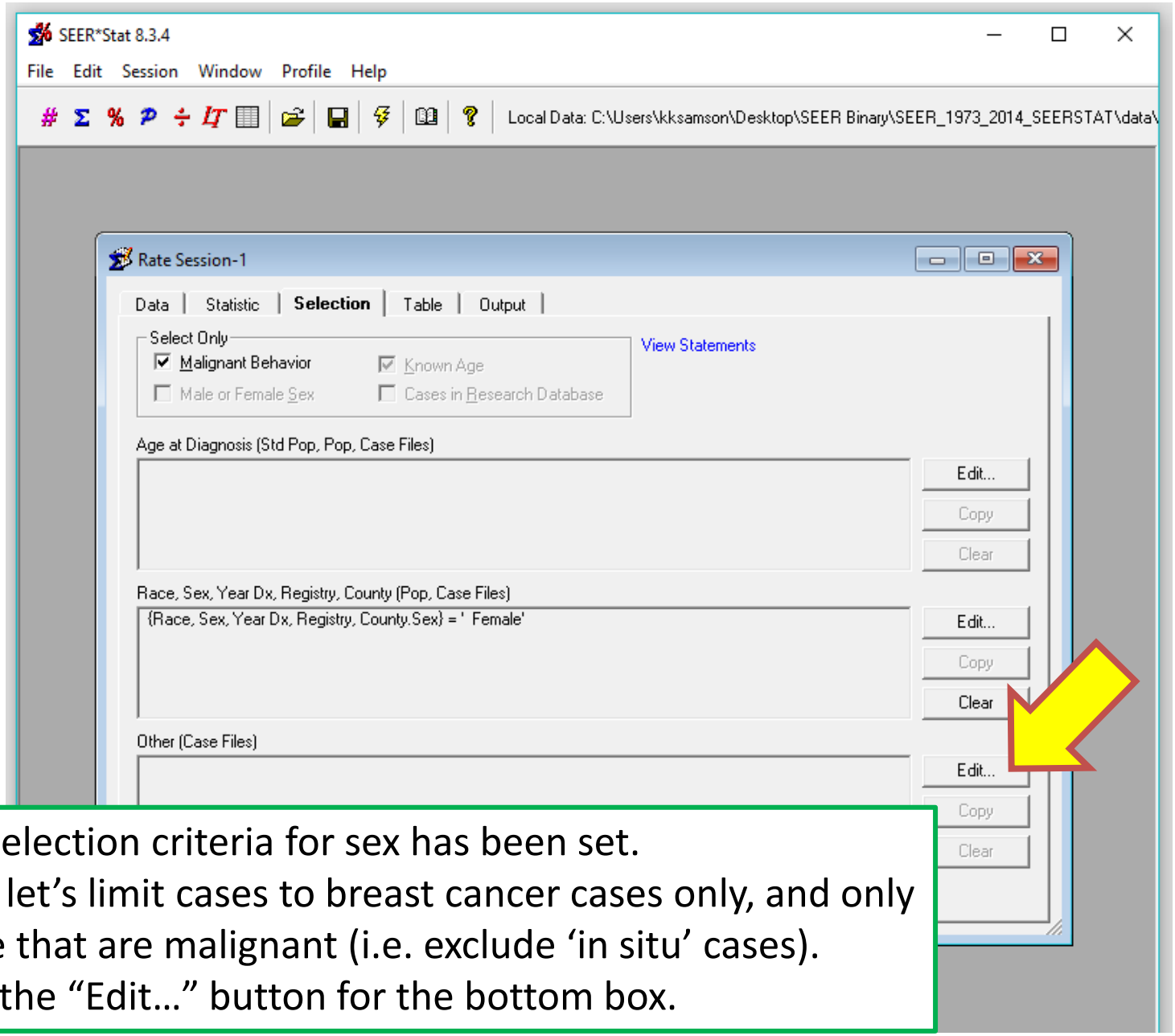
- Check “Show Standard Errors and Confidence Intervals”, and make sure “Use Tiwari et al., 2006 modification for CIs” is checked (default).
- Then click on the Selection tab.



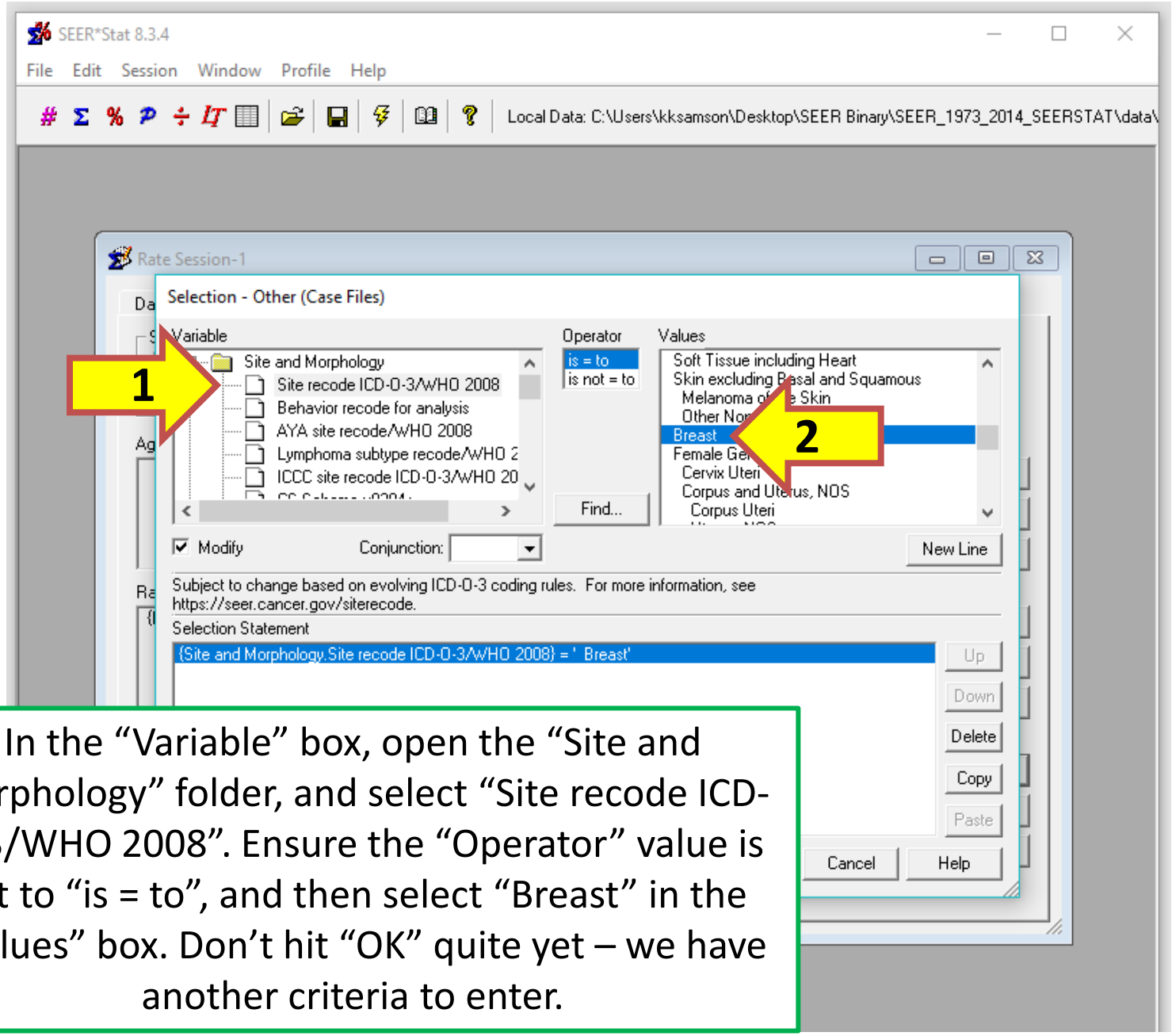
- For selection criteria, we can restrict which cases we want to use for our calculations.
- Start from the top box and move down.
- We plan to use all ages, so we won't use the first box.
- For the second box, we can restrict the calculations to only include females.
- To do this, click the "Edit…" button for the second box.



- In the "Variable" box, open the "Race, Sex, Year Dx, Registry, County" folder, and select "Sex".
- Ensure the "Operator" value is set to "is = to", and then select "Female" in the "Values" box.
- Then click "OK".

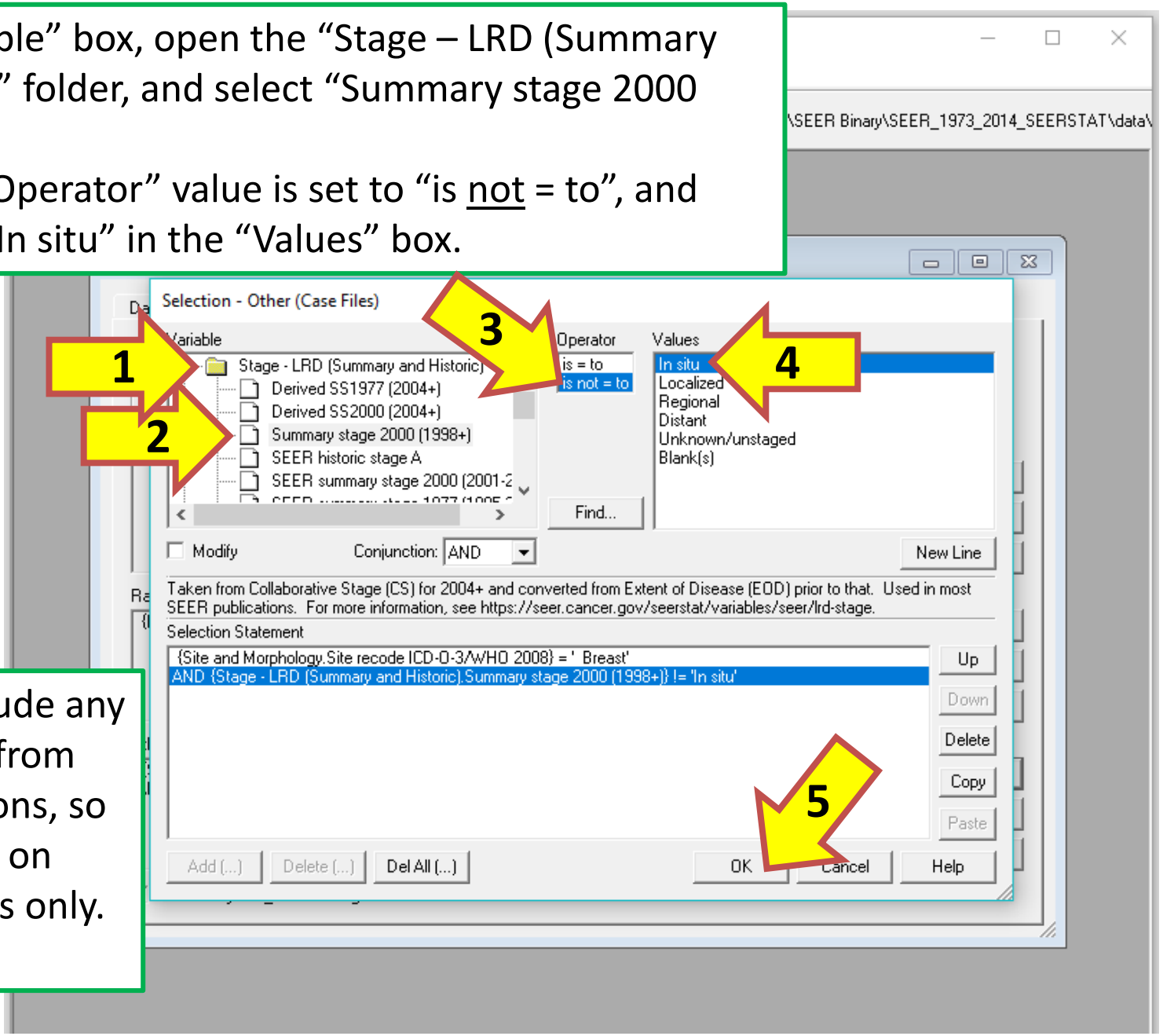


- Our selection criteria for sex has been set.
- Now, let's limit cases to breast cancer cases only, and only those that are malignant (i.e. exclude 'in situ' cases).
- Click the "Edit…" button for the bottom box.

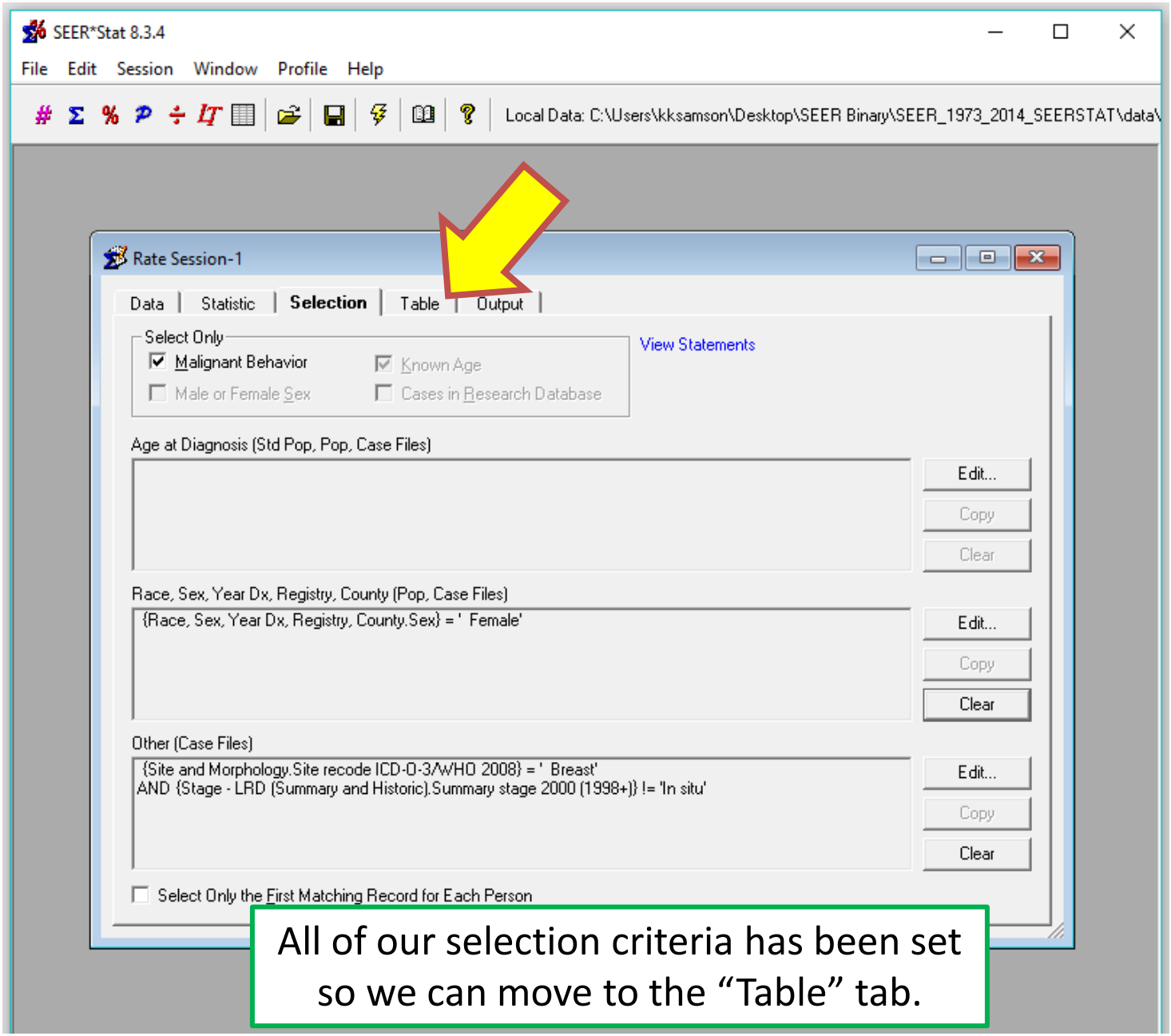


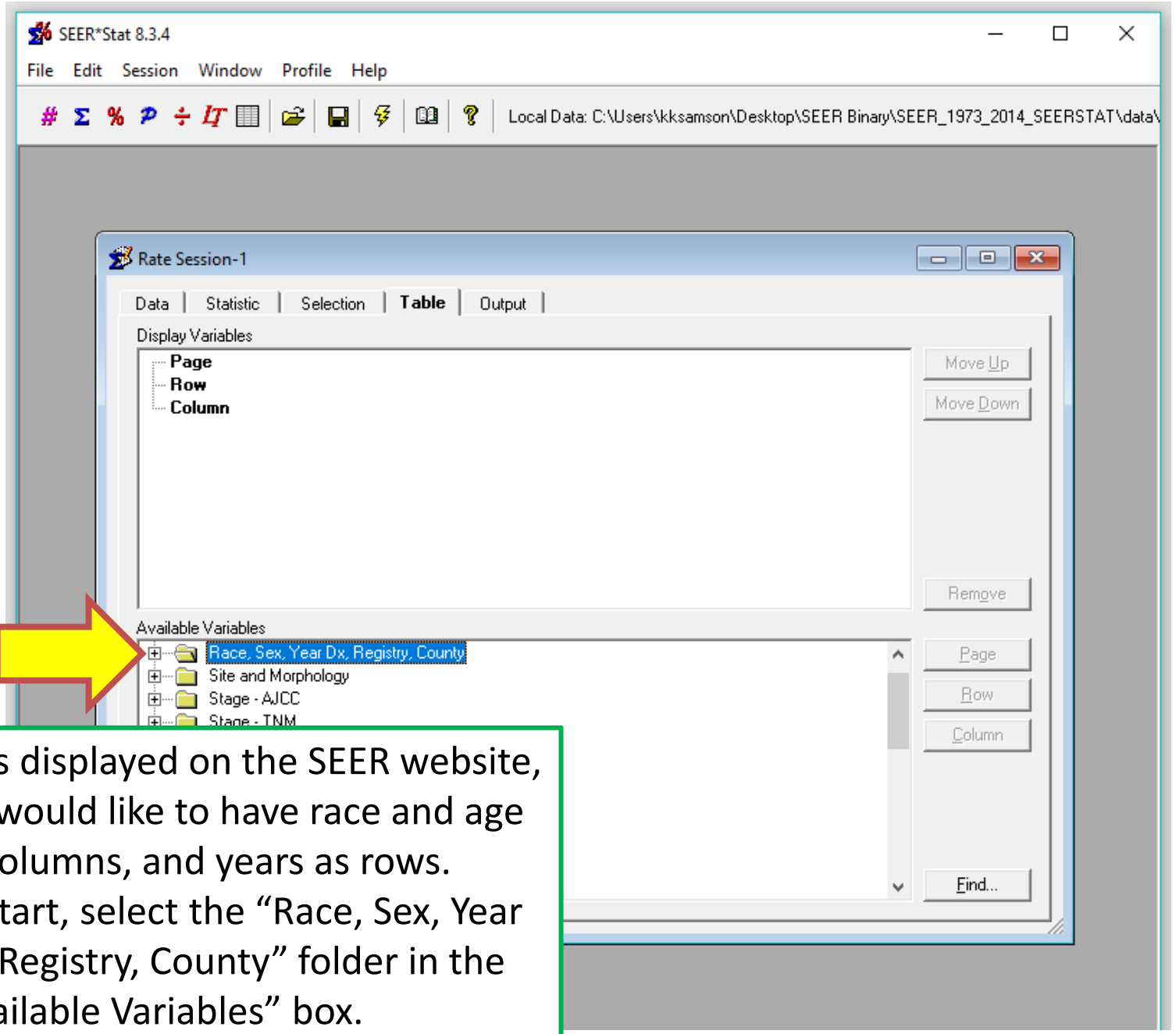
In the “Variable” box, open the “Site and Morphology” folder, and select “Site recode ICD-O-3/WHO 2008”. Ensure the “Operator” value is set to “is = to”, and then select “Breast” in the “Values” box. Don’t hit “OK” quite yet – we have another criteria to enter.

- In the “Variable” box, open the “Stage – LRD (Summary and Historic)” folder, and select “Summary stage 2000 (1998+)”.
- Ensure the “Operator” value is set to “is not = to”, and then select “In situ” in the “Values” box.

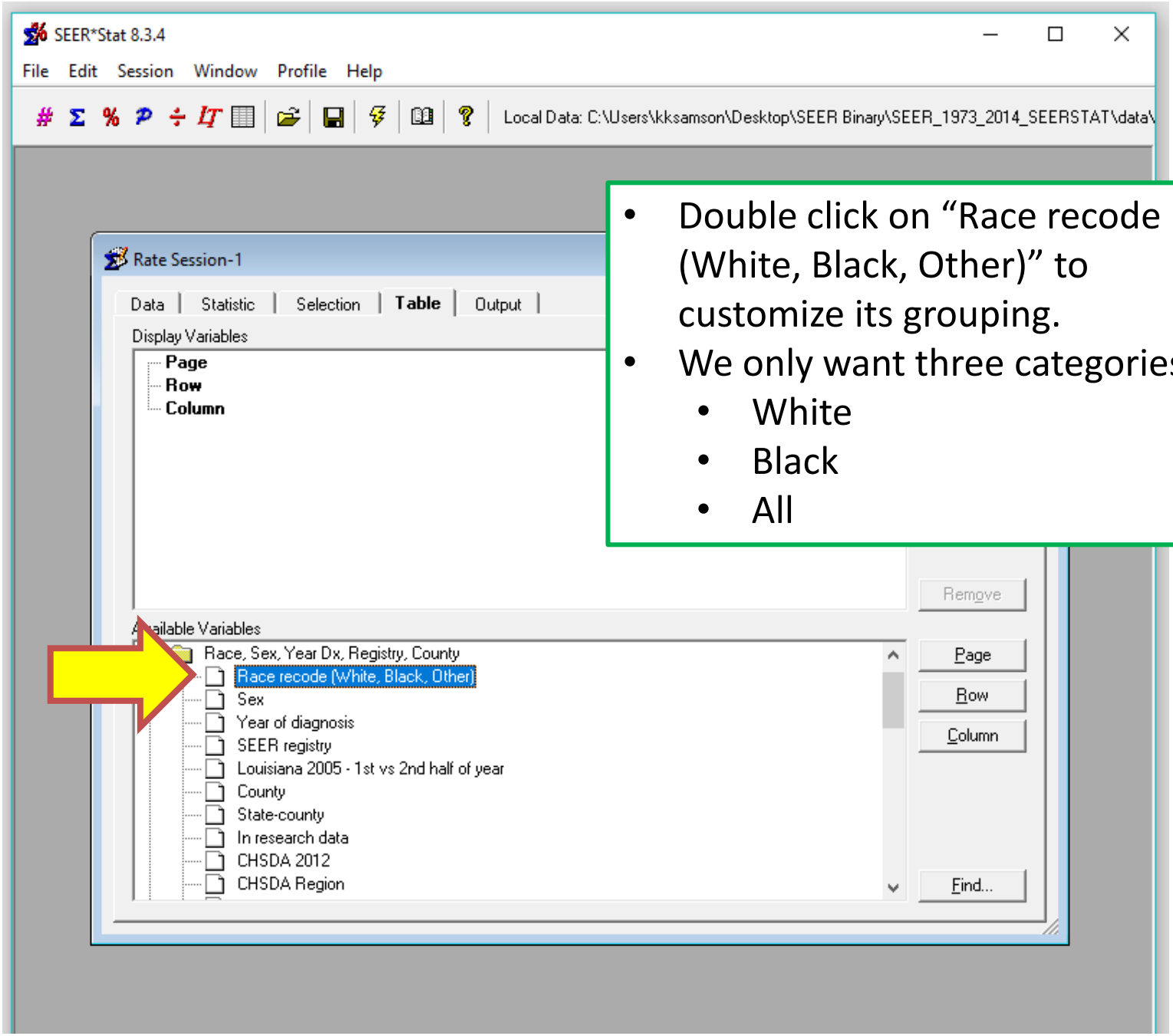


- This will exclude any in situ cases from our calculations, so we can focus on invasive cases only.
- Click “OK”.

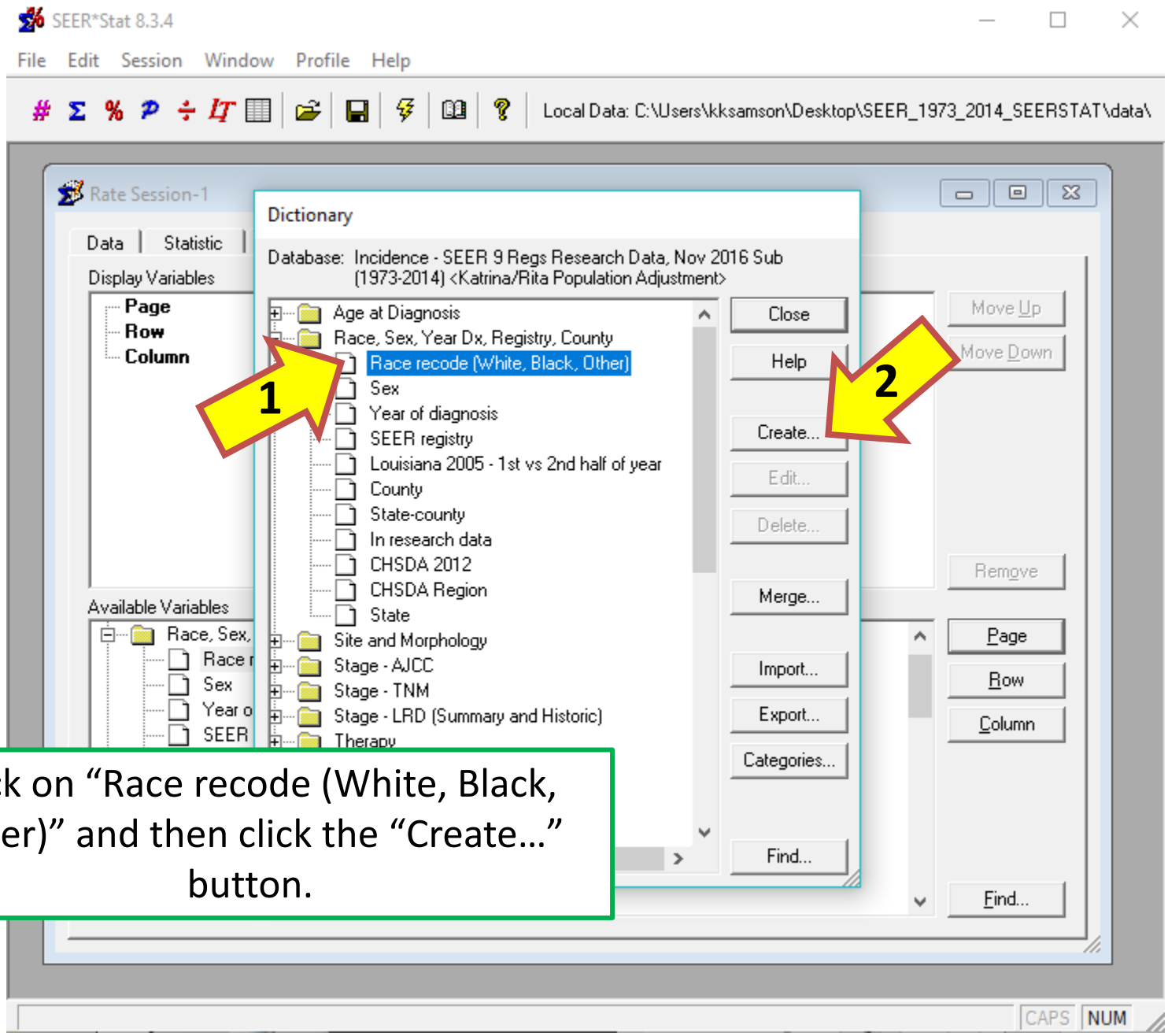

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER Binary\SEER_1973_2014_SEERSTAT\data\
Rate Session-1
Data Statistic Selection Table Output
Select Only
Malignant Behavior
Known Age
Male or Female Sex
Cases in Research Database
View Statements
Age at Diagnosis (Std Pop, Pop, Case Files)
Edit...
Copy
Clear
Race, Sex, Year Dx, Registry, County (Pop, Case Files)
{Race, Sex, Year Dx, Registry, County.Sex} = ' Female'
Edit...
Copy
Clear
Other (Case Files)
{Site and Morphology.Site recode ICD-O-3/WHO 2008} = ' Breast'
AND {Stage - LRD (Summary and Historic).Summary stage 2000 (1998+)} != 'In situ'
Edit...
Copy
Clear
Select Only the First Matching Record for Each Person
All of our selection criteria has been set so we can move to the “Table” tab.



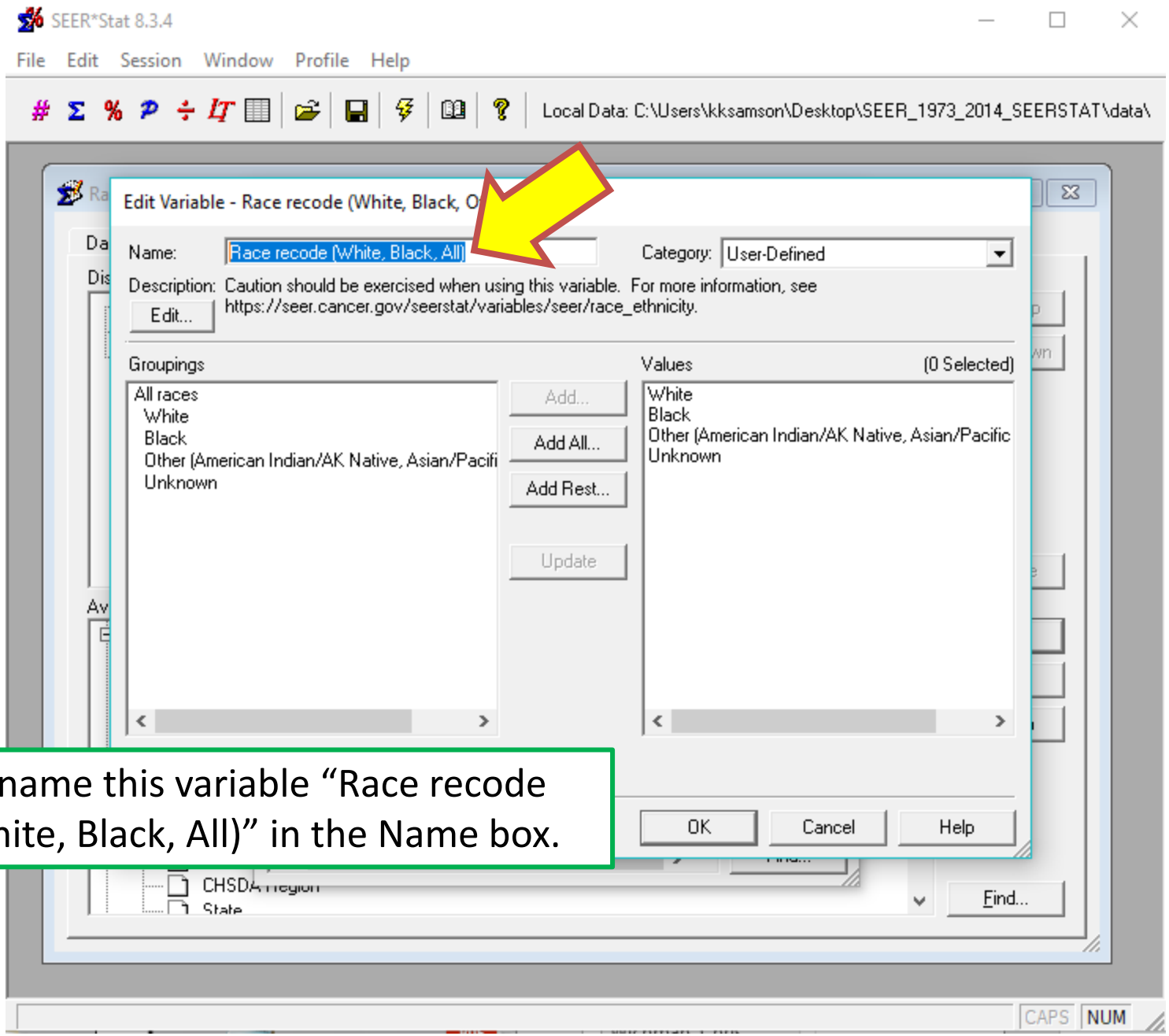
- As is displayed on the SEER website, we would like to have race and age as columns, and years as rows.
- To start, select the “Race, Sex, Year Dx, Registry, County” folder in the “Available Variables” box.



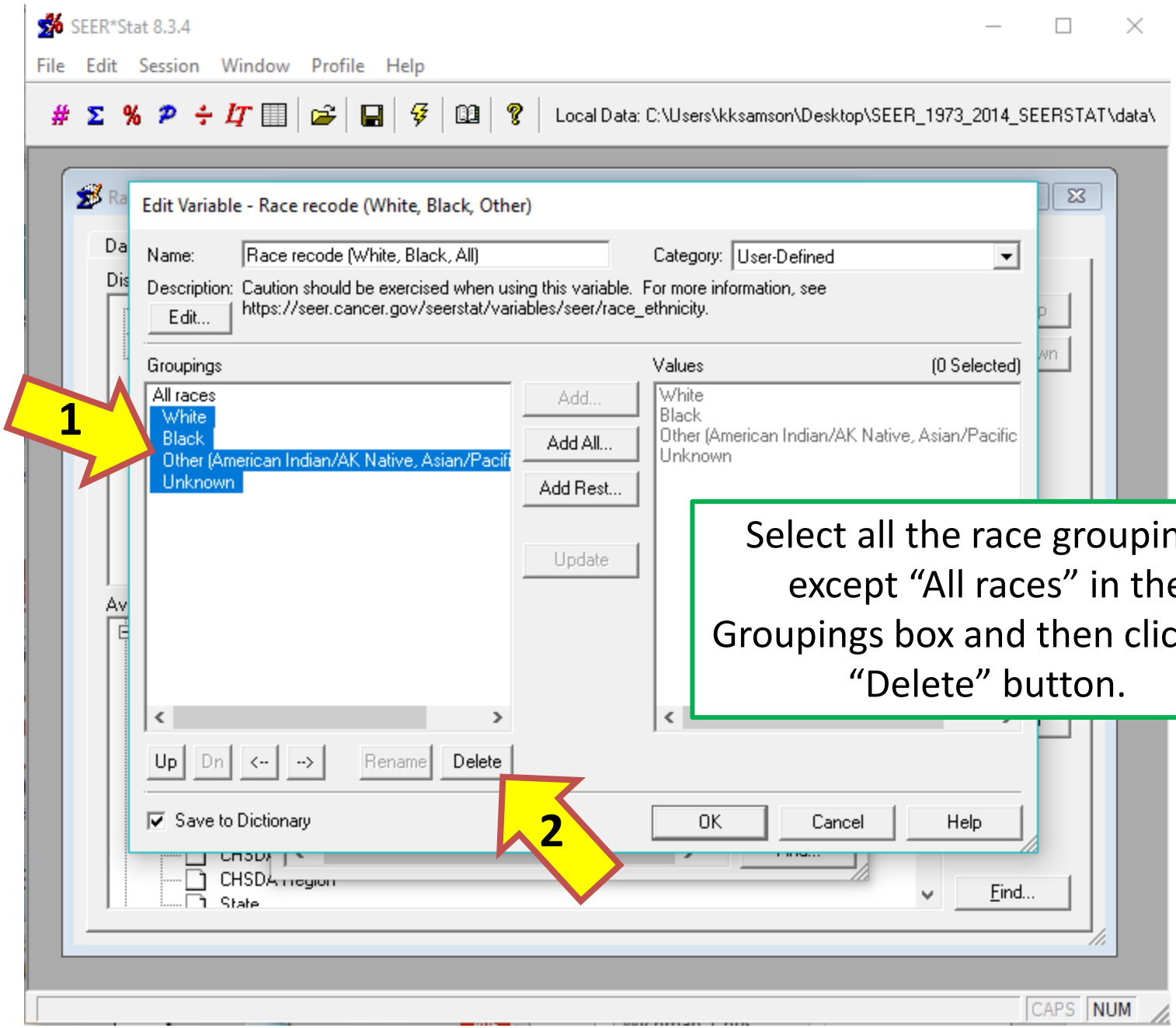
- Double click on “Race recode (White, Black, Other)” to customize its grouping.
- We only want three categories:
  - White
  - Black
  - All



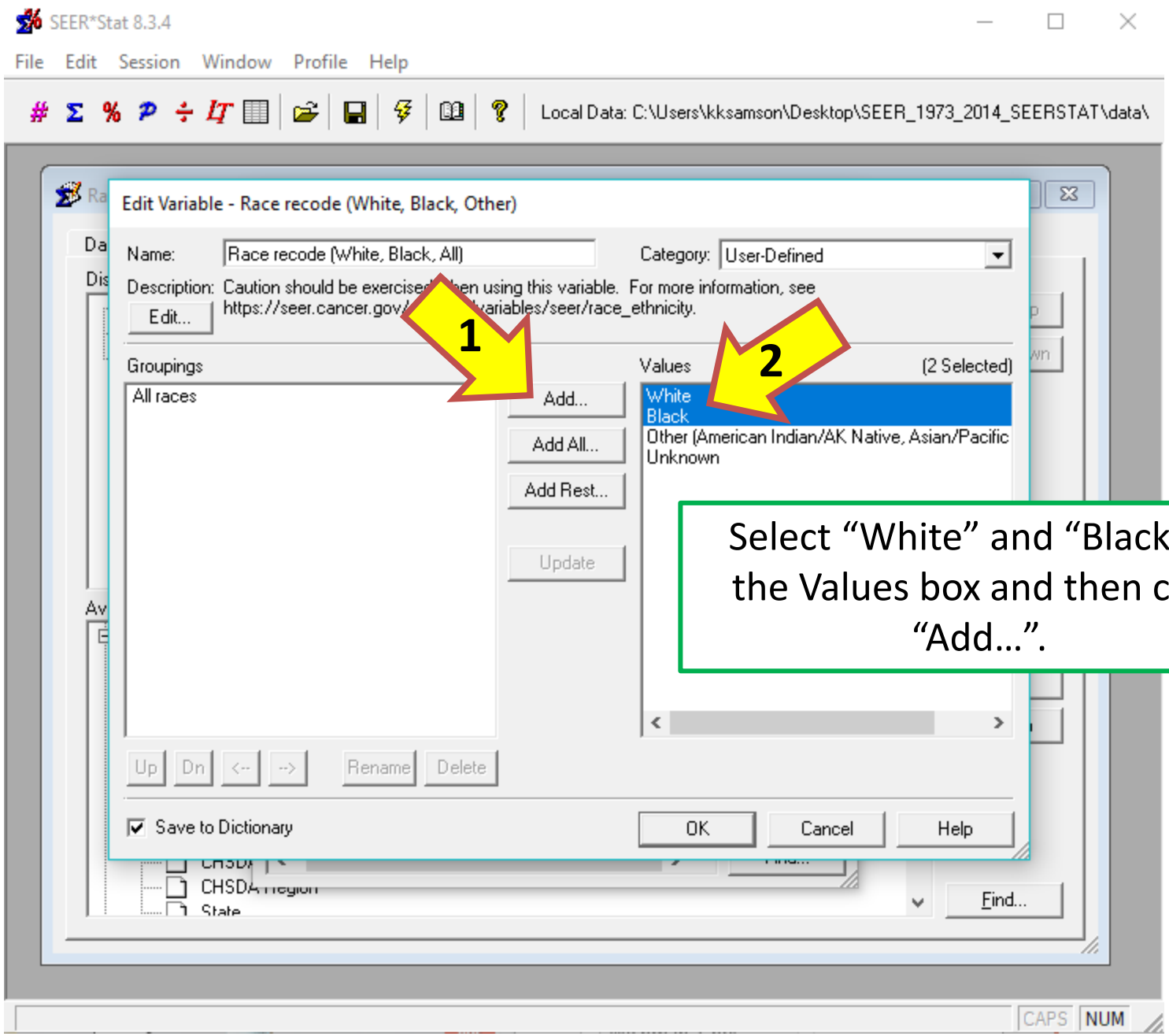
Click on “Race recode (White, Black, Other)” and then click the “Create…” button.



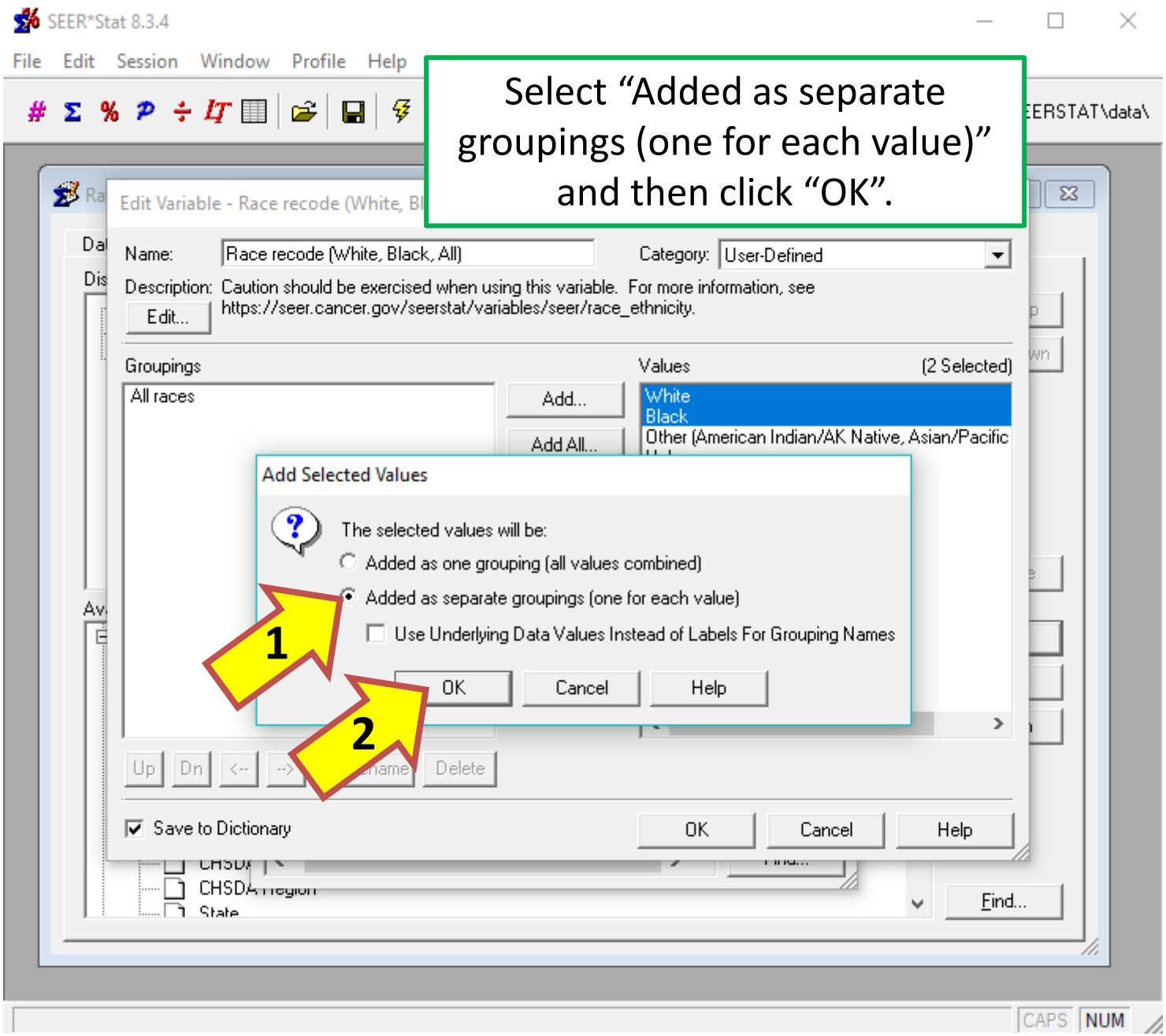
Rename this variable “Race recode (White, Black, All)” in the Name box.

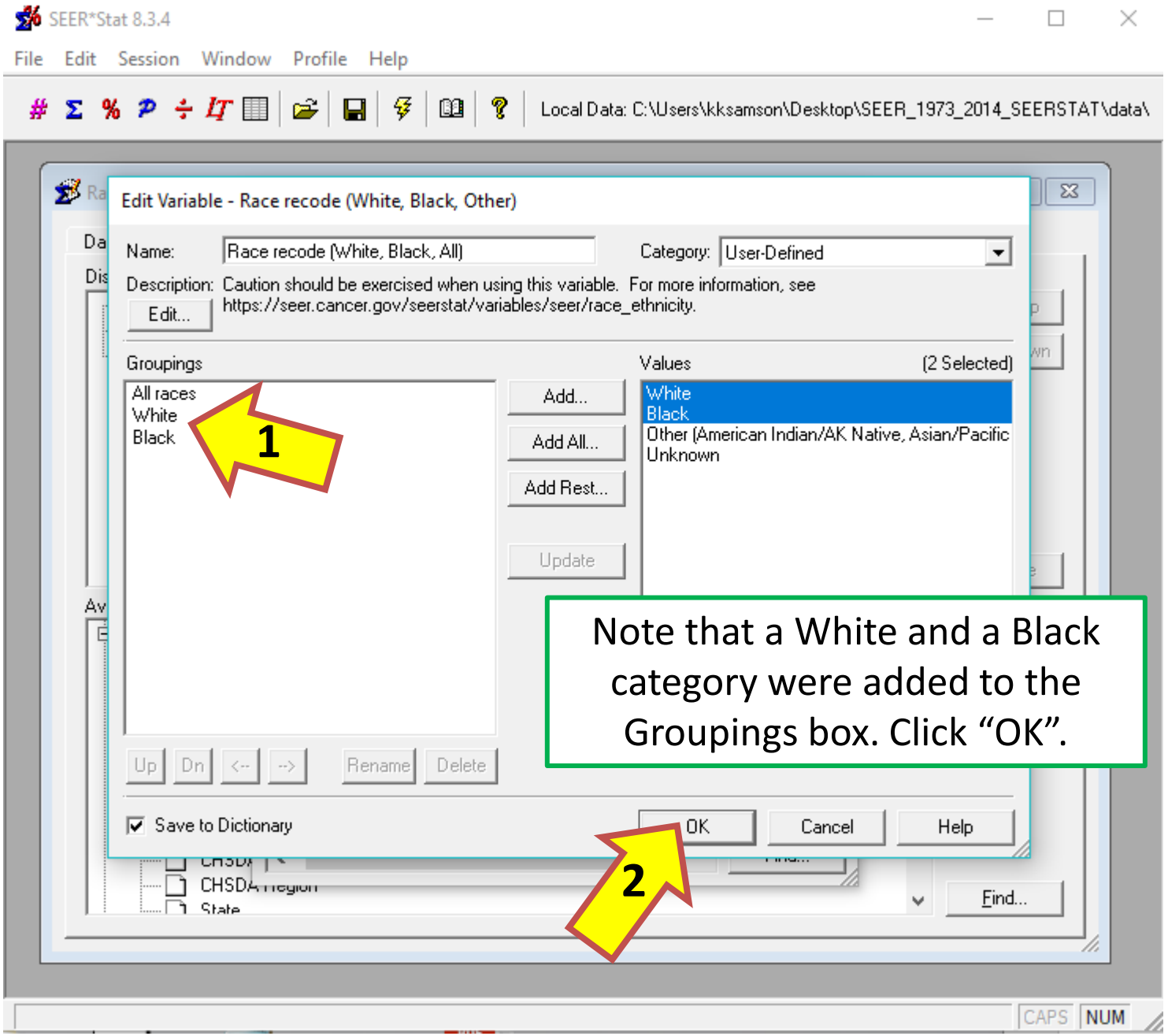


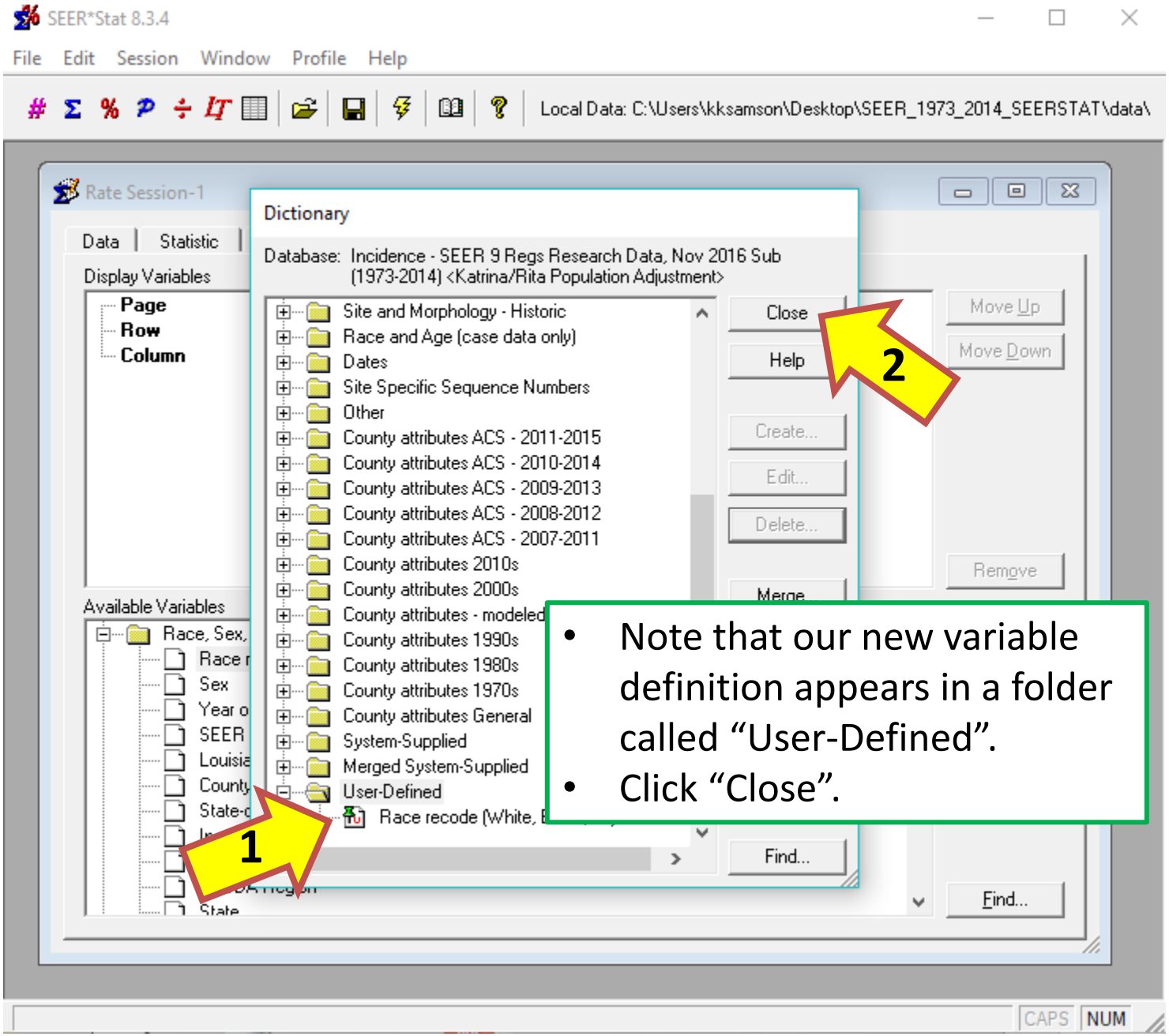
Select all the race groupings, except "All races" in the Groupings box and then click the "Delete" button.

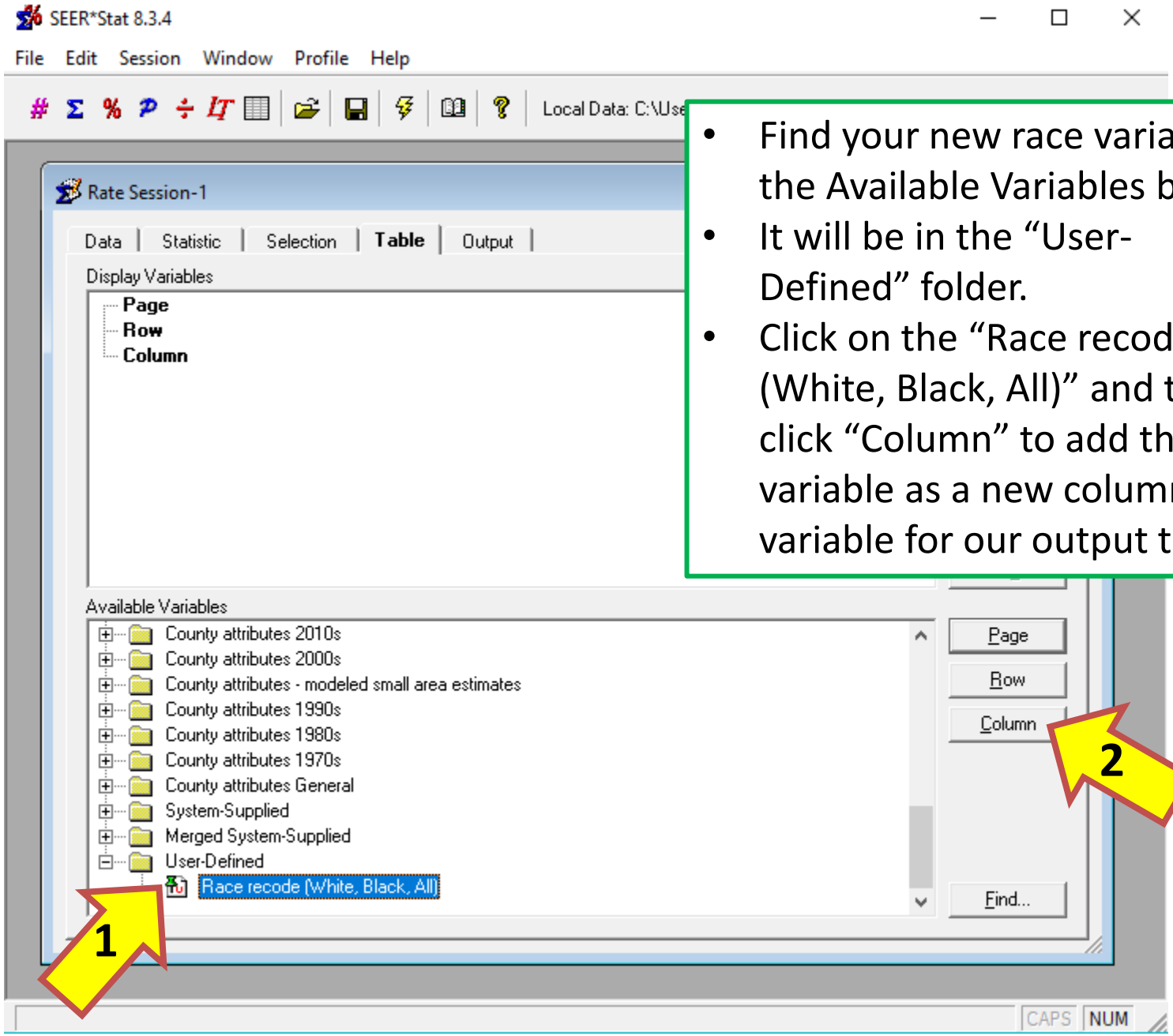


Select “White” and “Black” in the Values box and then click “Add…”.

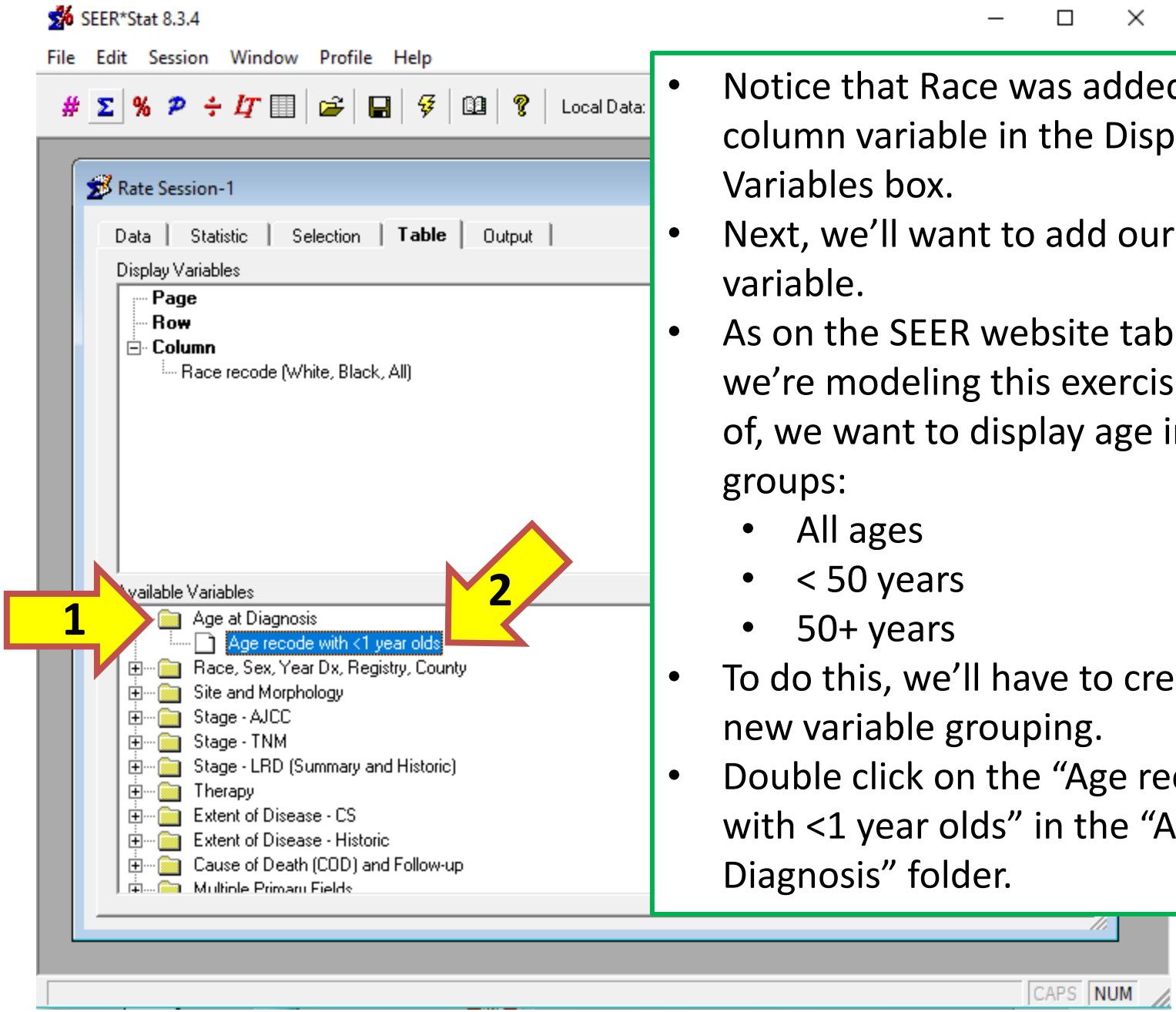

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Select “Added as separate groupings (one for each value)” and then click “OK”.
Edit Variable - Race recode (White, Bl
Name: Race recode (White, Black, All)
Category: User-Defined
Description: Caution should be exercised when using this variable. For more information, see https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity.
Edit...
Groupings
All races
Values (2 Selected)
White
Black
Other (American Indian/AK Native, Asian/Pacific
Add...
Add All...
Add Selected Values
The selected values will be:
Added as one grouping (all values combined)
Added as separate groupings (one for each value)
Use Underlying Data Values Instead of Labels For Grouping Names
OK
Cancel
Help
1
2
Up
Dn
Delete
Save to Dictionary
OK
Cancel
Help
CHSDA Region
State
Find...
CAPS
NUM


SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER_1973_2014_SEERSTAT\data\
Edit Variable - Race recode (White, Black, Other)
Name: Race recode (White, Black, All)
Category: User-Defined
Description: Caution should be exercised when using this variable. For more information, see https://seer.cancer.gov/seerstat/variables/seer/race_ethnicity.
Edit...
Groupings
All races
White
Black
Values (2 Selected)
White
Black
Other (American Indian/AK Native, Asian/Pacific
Unknown
Add...
Add All...
Add Rest...
Update
Up
Dn
<--
-->
Rename
Delete
Save to Dictionary
OK
Cancel
Help
1
2
Note that a White and a Black category were added to the Groupings box. Click "OK".
Find...
CAPS
NUM

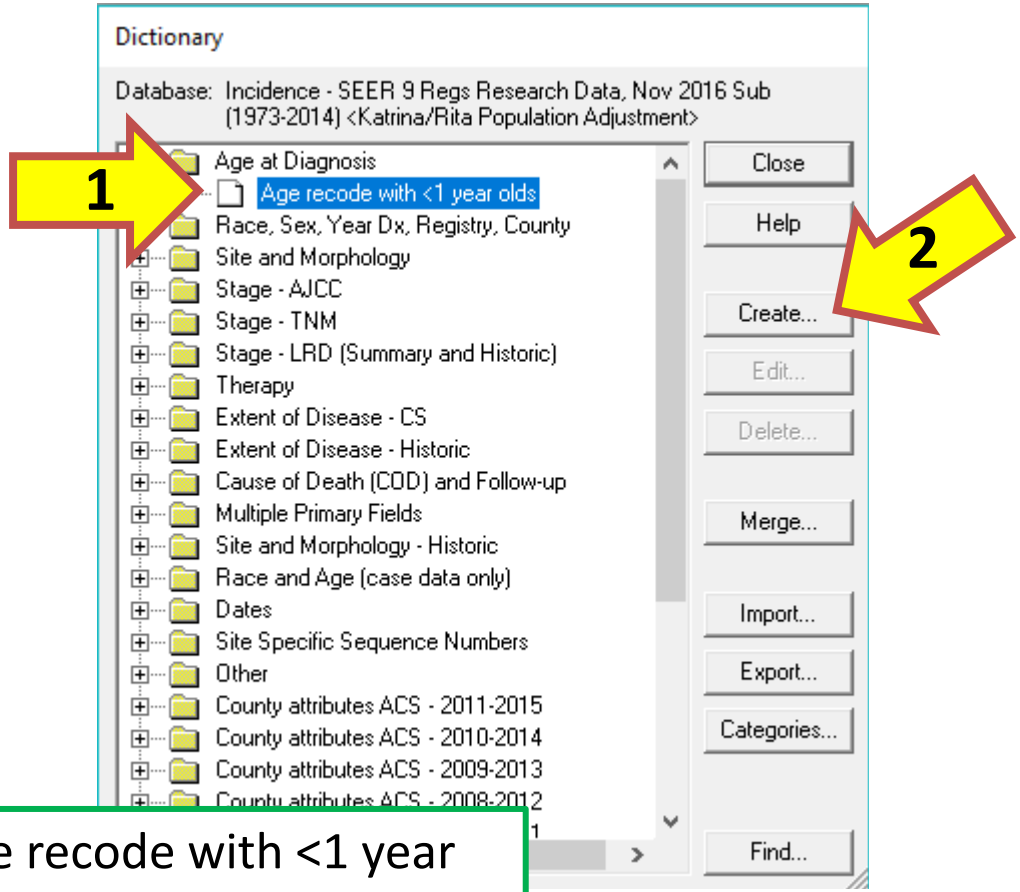

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER_1973_2014_SEERSTAT\data\
Rate Session-1
Data Statistic
Display Variables
Page
Row
Column
Available Variables
Race, Sex,
Sex
SEER
Dictionary
Database: Incidence - SEER 9 Regs Research Data, Nov 2016 Sub (1973-2014) <Katrina/Rita Population Adjustment>
Site and Morphology - Historic
Race and Age (case data only)
Dates
Site Specific Sequence Numbers
Other
County attributes ACS - 2011-2015
County attributes ACS - 2010-2014
County attributes ACS - 2009-2013
County attributes ACS - 2008-2012
County attributes ACS - 2007-2011
County attributes 2010s
County attributes 2000s
County attributes - modeled
County attributes 1990s
County attributes 1980s
County attributes 1970s
County attributes General
System-Supplied
Merged System-Supplied
User-Defined
Race recode (White,
Close
Help
Create...
Edit...
Delete...
Merge
Find...
Move Up
Move Down
Remove
Find...
1
2
• Note that our new variable definition appears in a folder called “User-Defined”.
• Click “Close”.
CAPS NUM



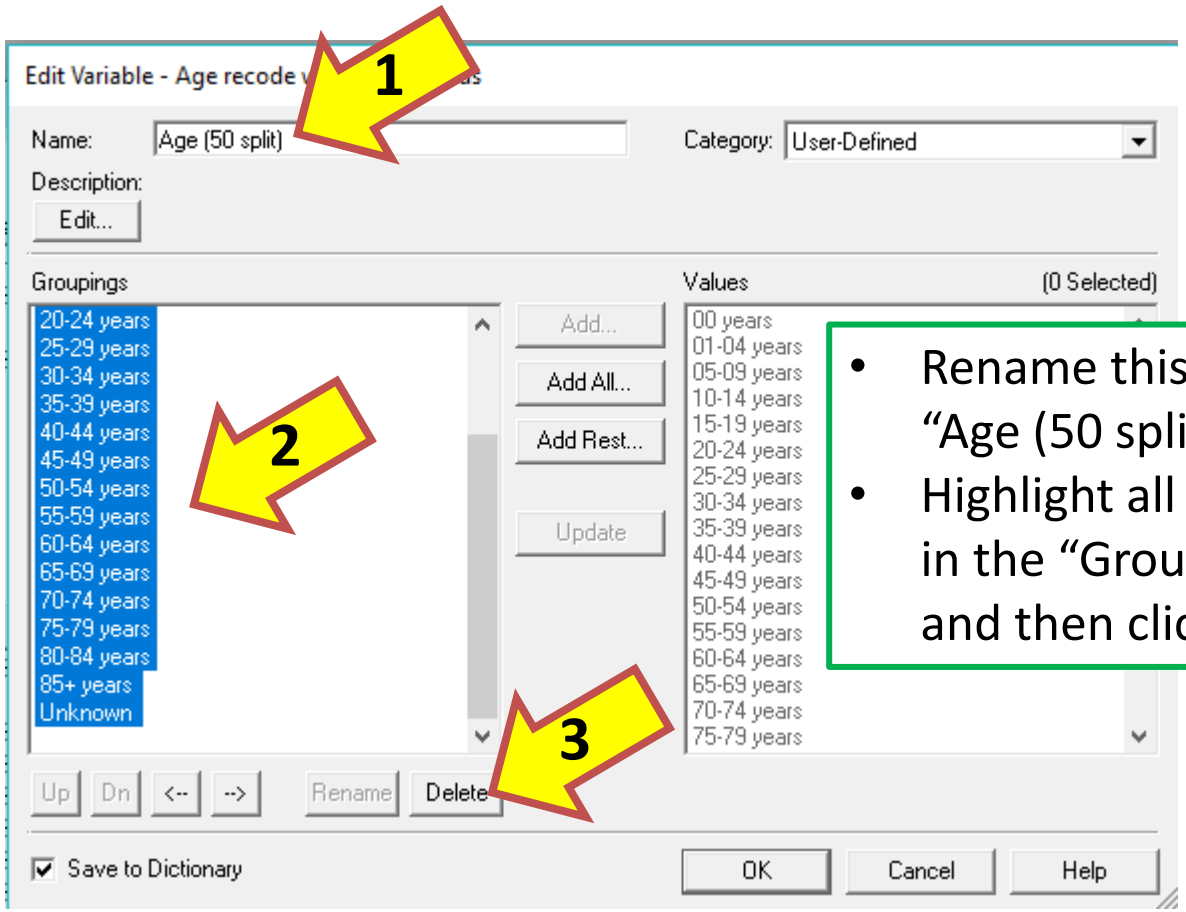
- Find your new race variable in the Available Variables box.
- It will be in the "User-Defined" folder.
- Click on the "Race recode (White, Black, All)" and then click "Column" to add this variable as a new column variable for our output table.



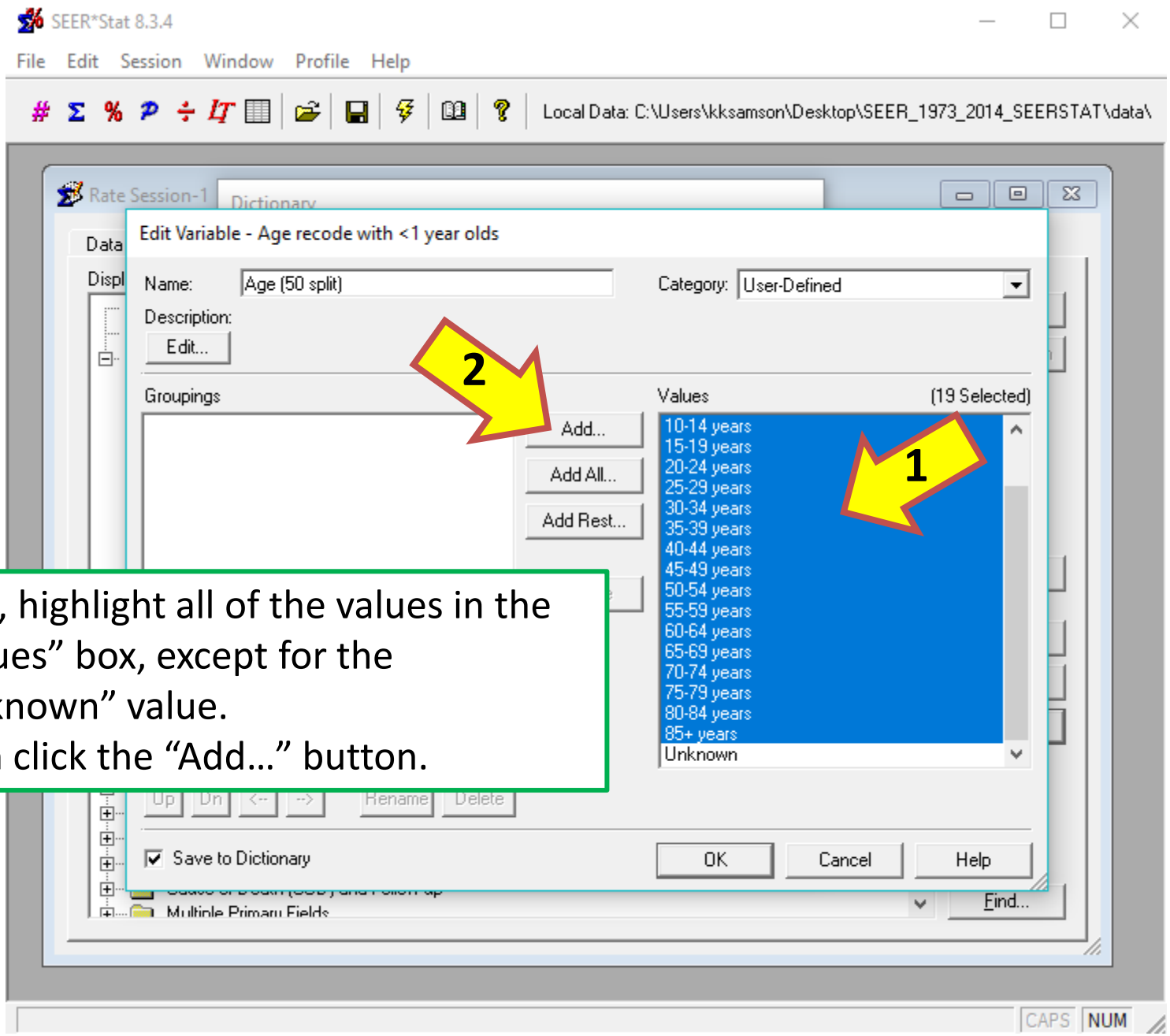
- Notice that Race was added as a column variable in the Display Variables box.
- Next, we'll want to add our age variable.
- As on the SEER website table we're modeling this exercise off of, we want to display age in three groups:
  - All ages
  - < 50 years
  - 50+ years
- To do this, we'll have to create a new variable grouping.
- Double click on the "Age recode with <1 year olds" in the "Age at Diagnosis" folder.



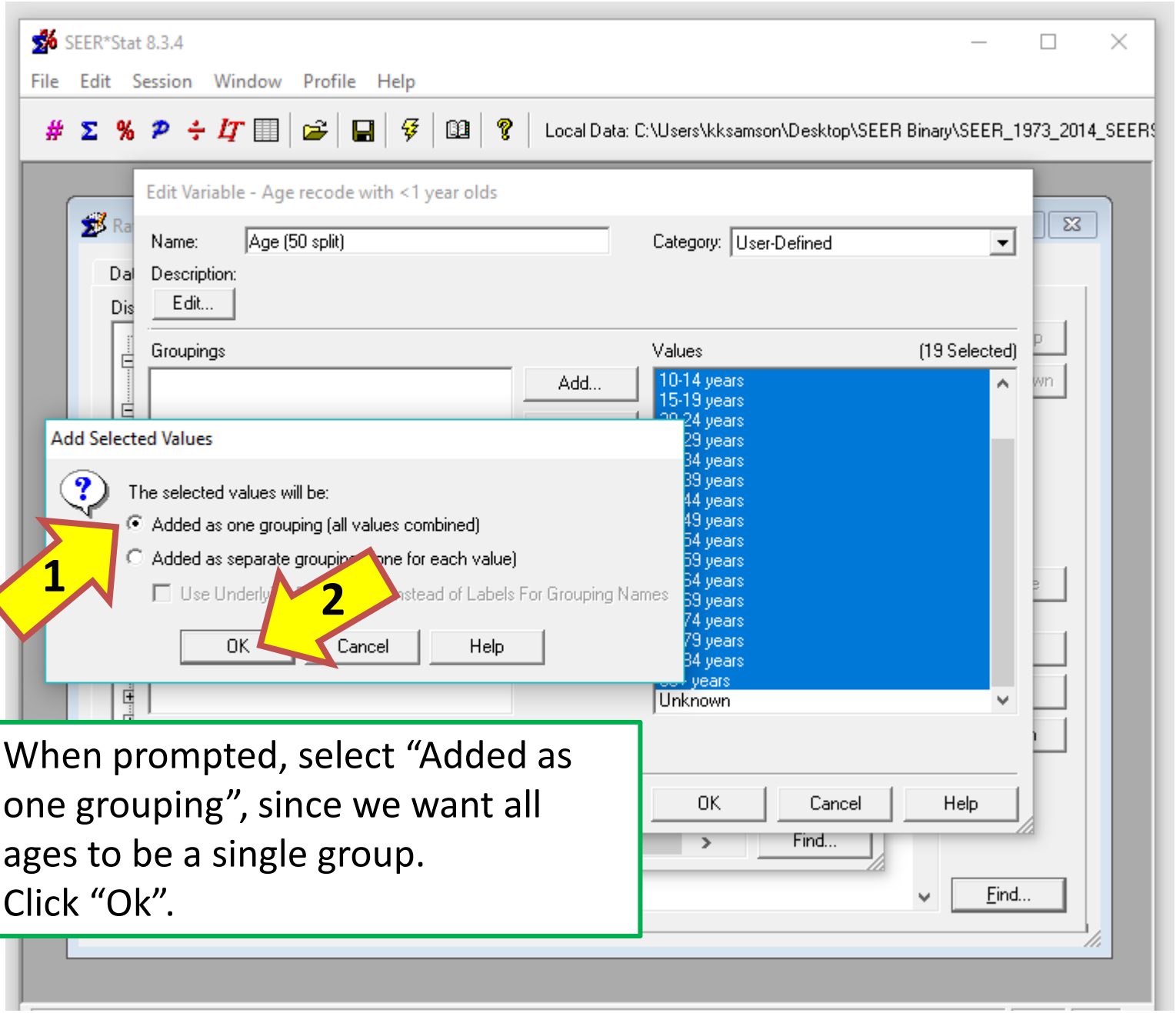
Make sure “Age recode with <1 year olds” is selected and then click “Create…”.



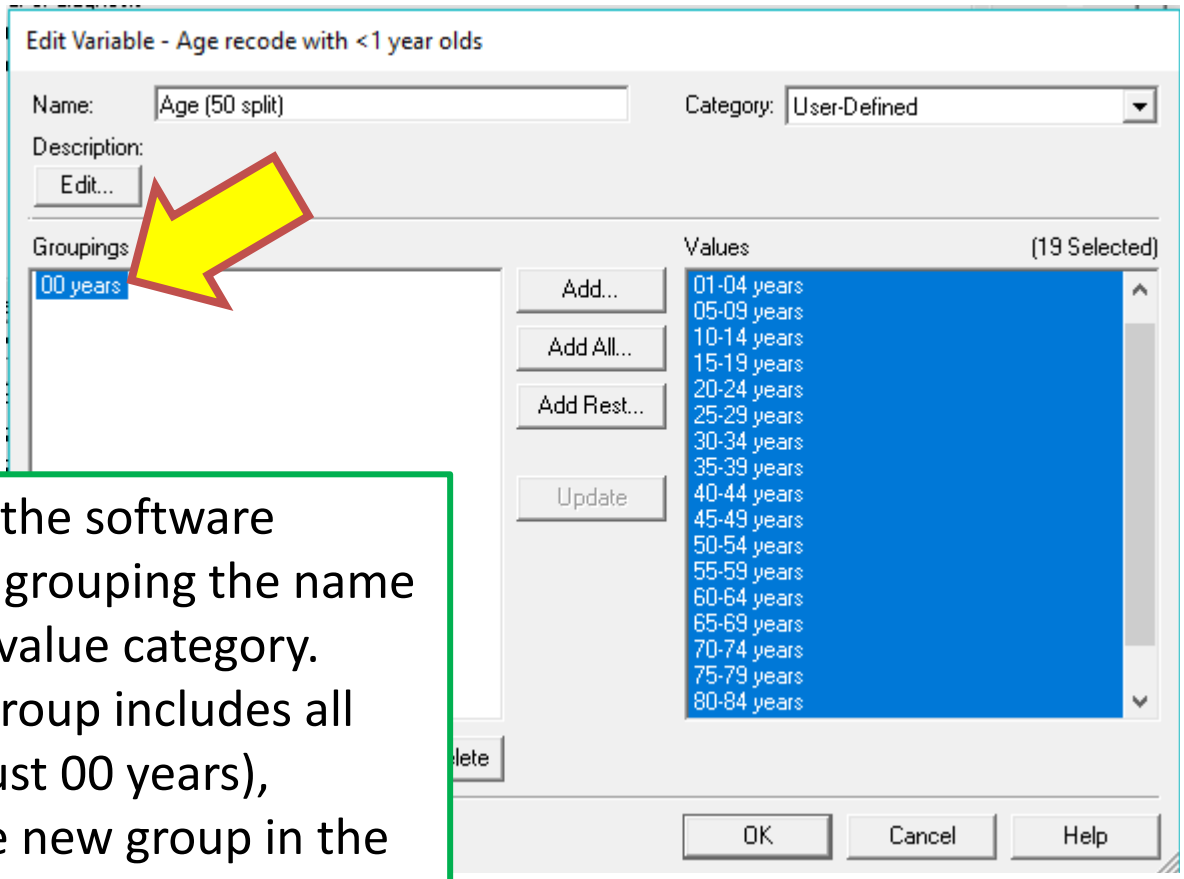
- Rename this new variable “Age (50 split)”.
- Highlight all of the values in the “Groupings” box, and then click “Delete”.



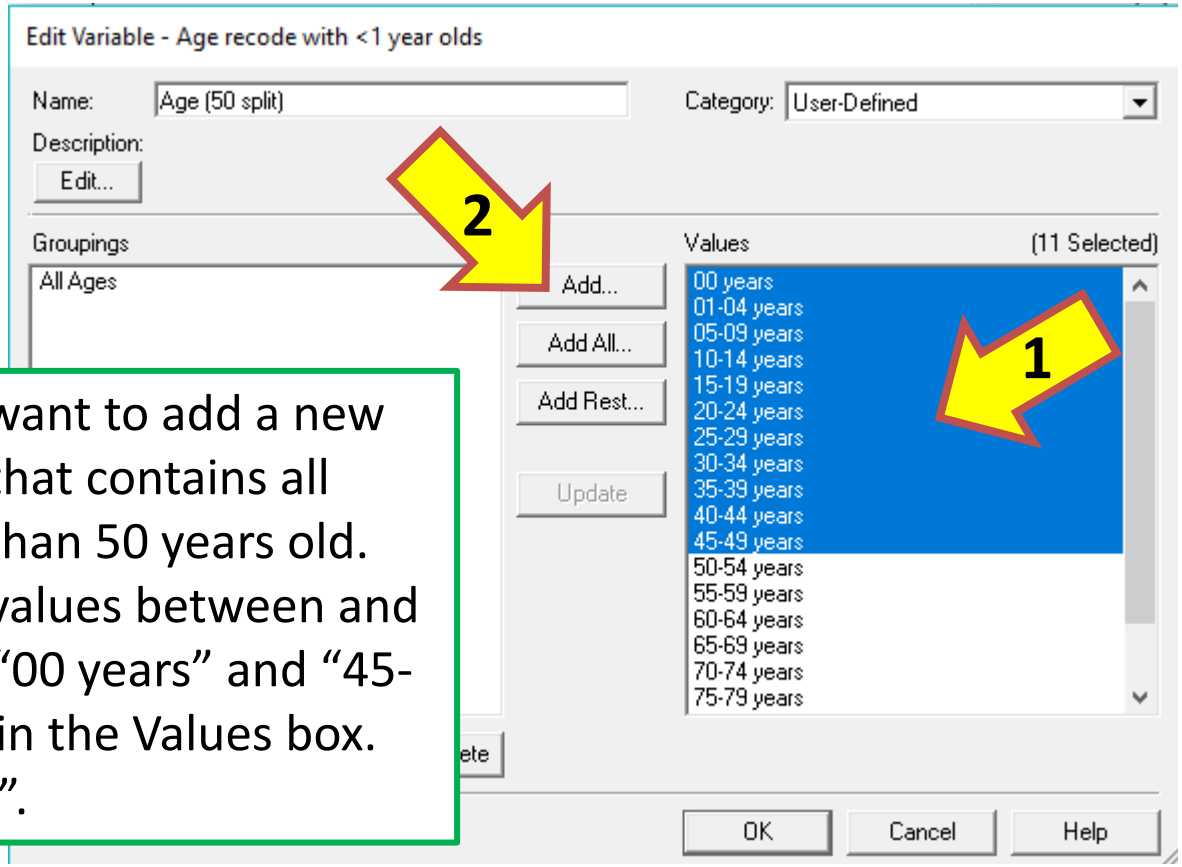
- Next, highlight all of the values in the "Values" box, except for the "Unknown" value.
- Then click the "Add…" button.



- When prompted, select “Added as one grouping”, since we want all ages to be a single group.
- Click “Ok”.

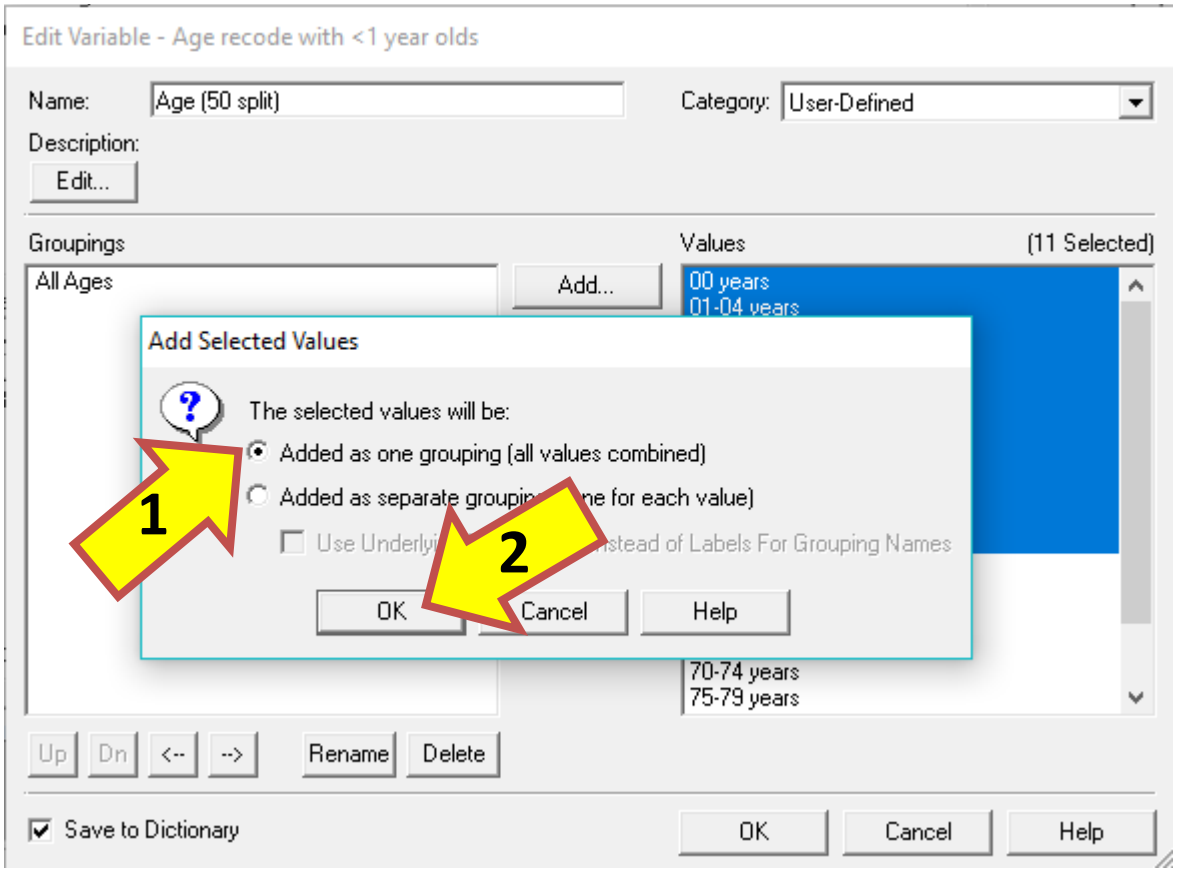


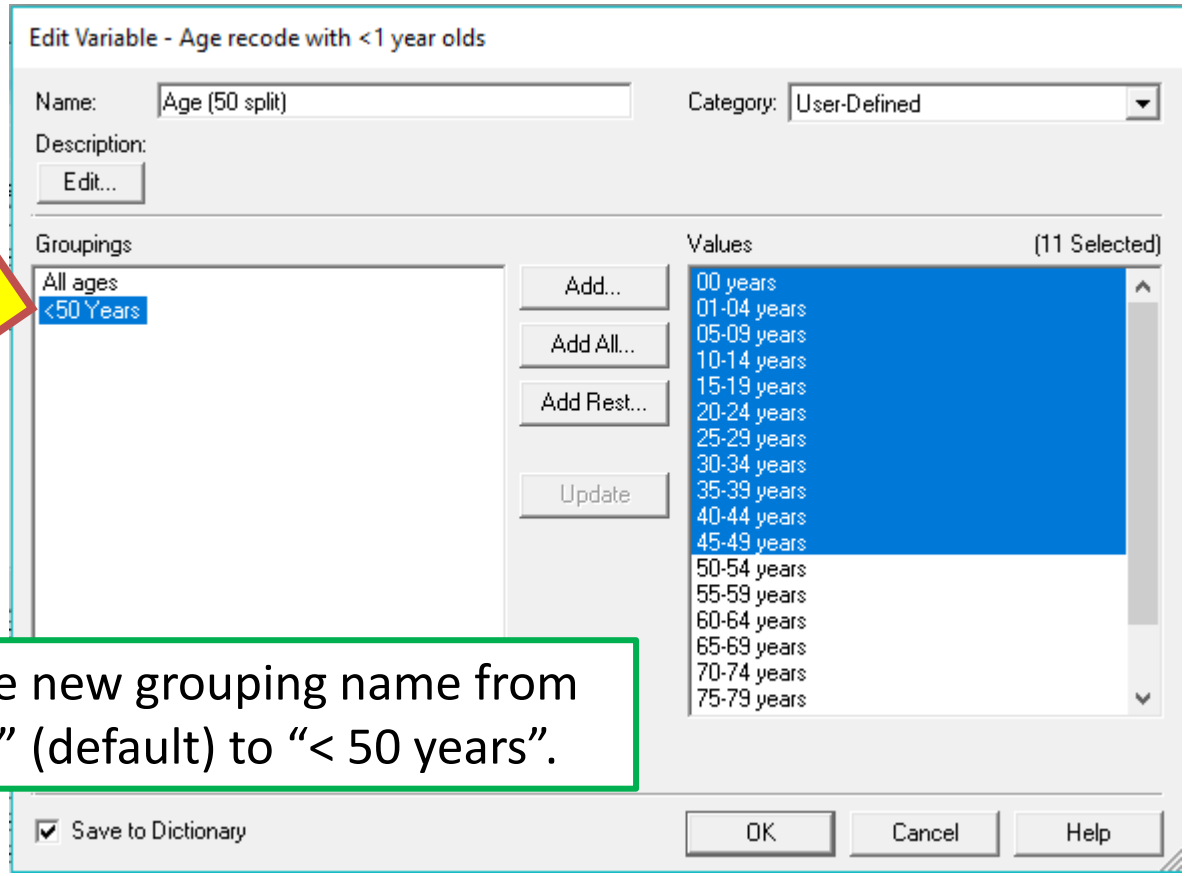
- By default, the software names this grouping the name of the first value category.
- Since this group includes all ages (not just 00 years), rename the new group in the Groupings box “All ages”.
- If the name is not editable, click it once to highlight it, then click again to type.



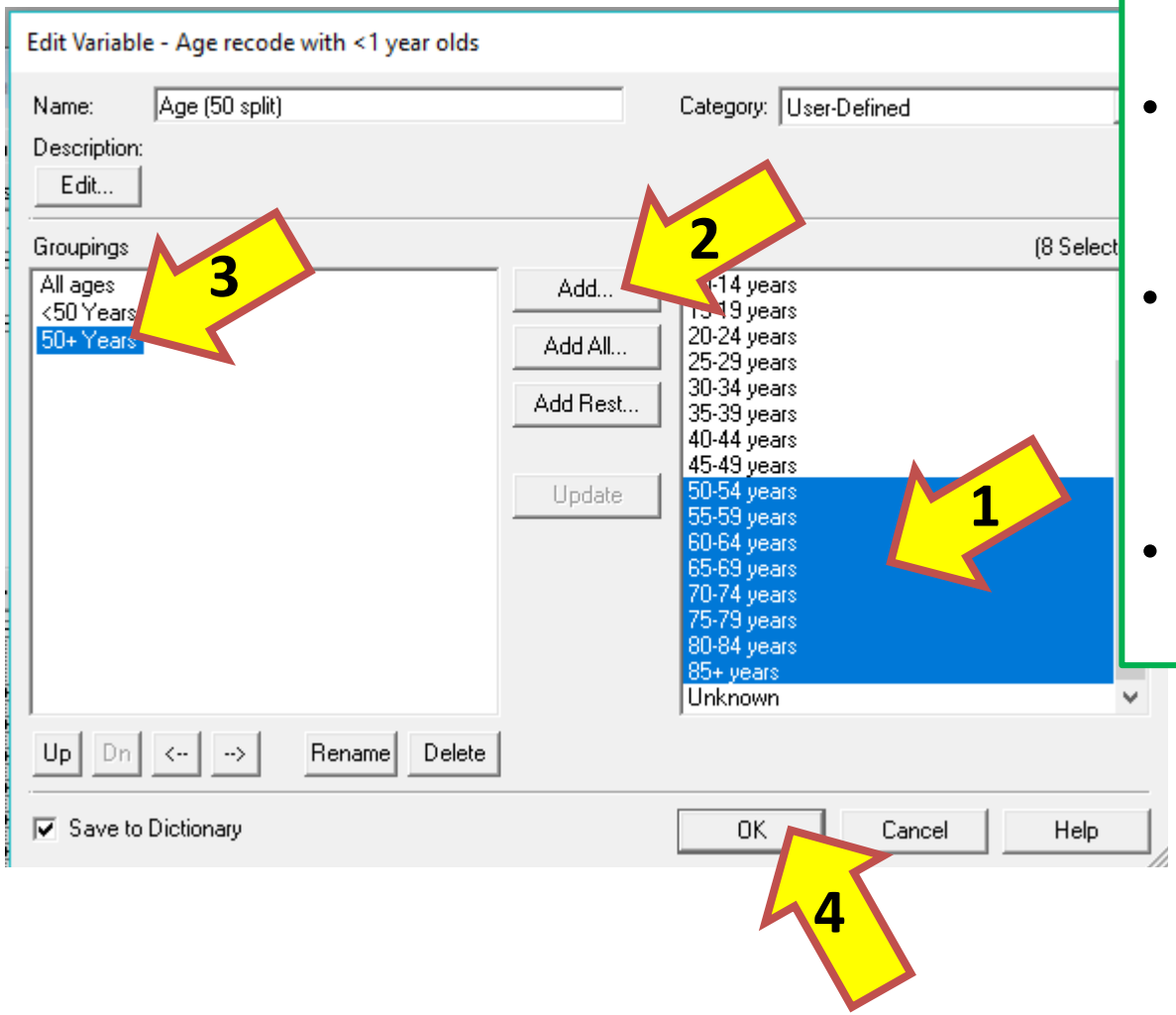
- Next, we want to add a new grouping that contains all ages less than 50 years old.
- Select all values between and including “00 years” and “45-49 years” in the Values box.
- Click “Add”.

As with the previous group creation, select “Added as one grouping” and click “Ok”.

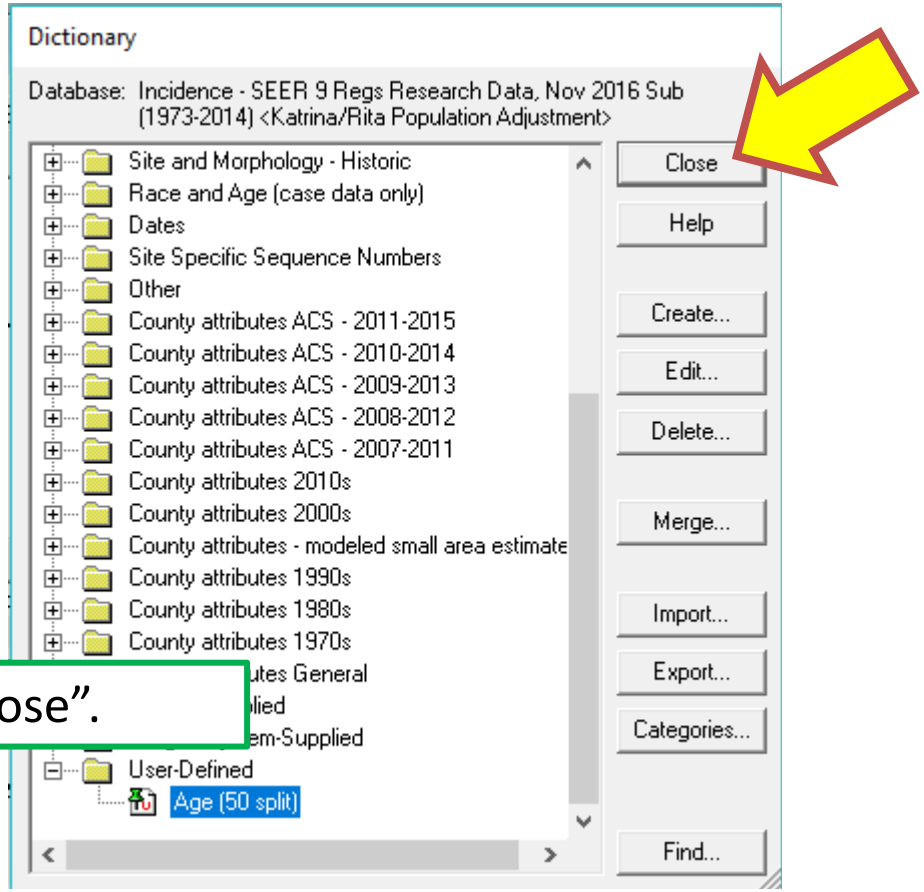




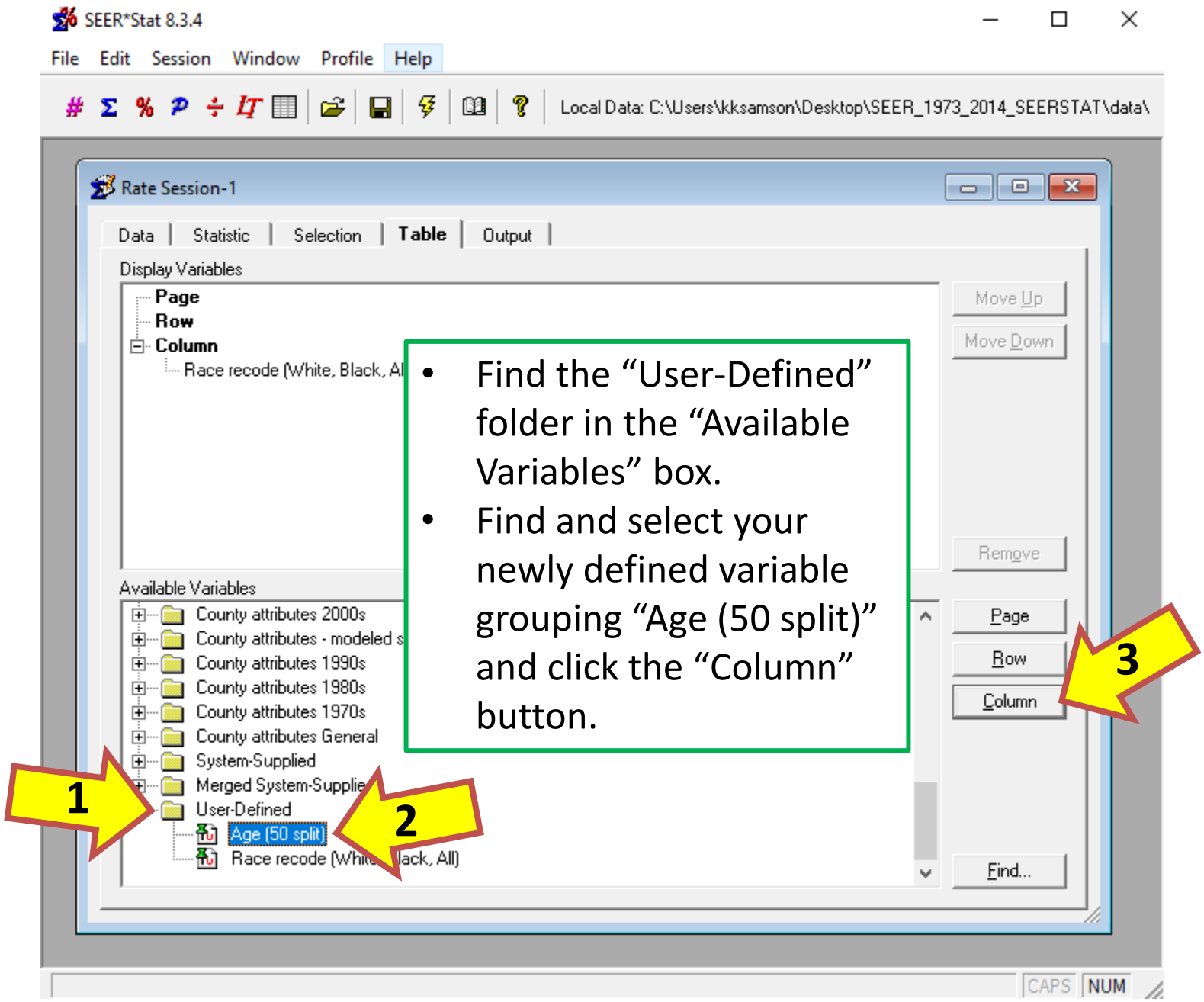
Change the new grouping name from "00 years" (default) to "< 50 years".

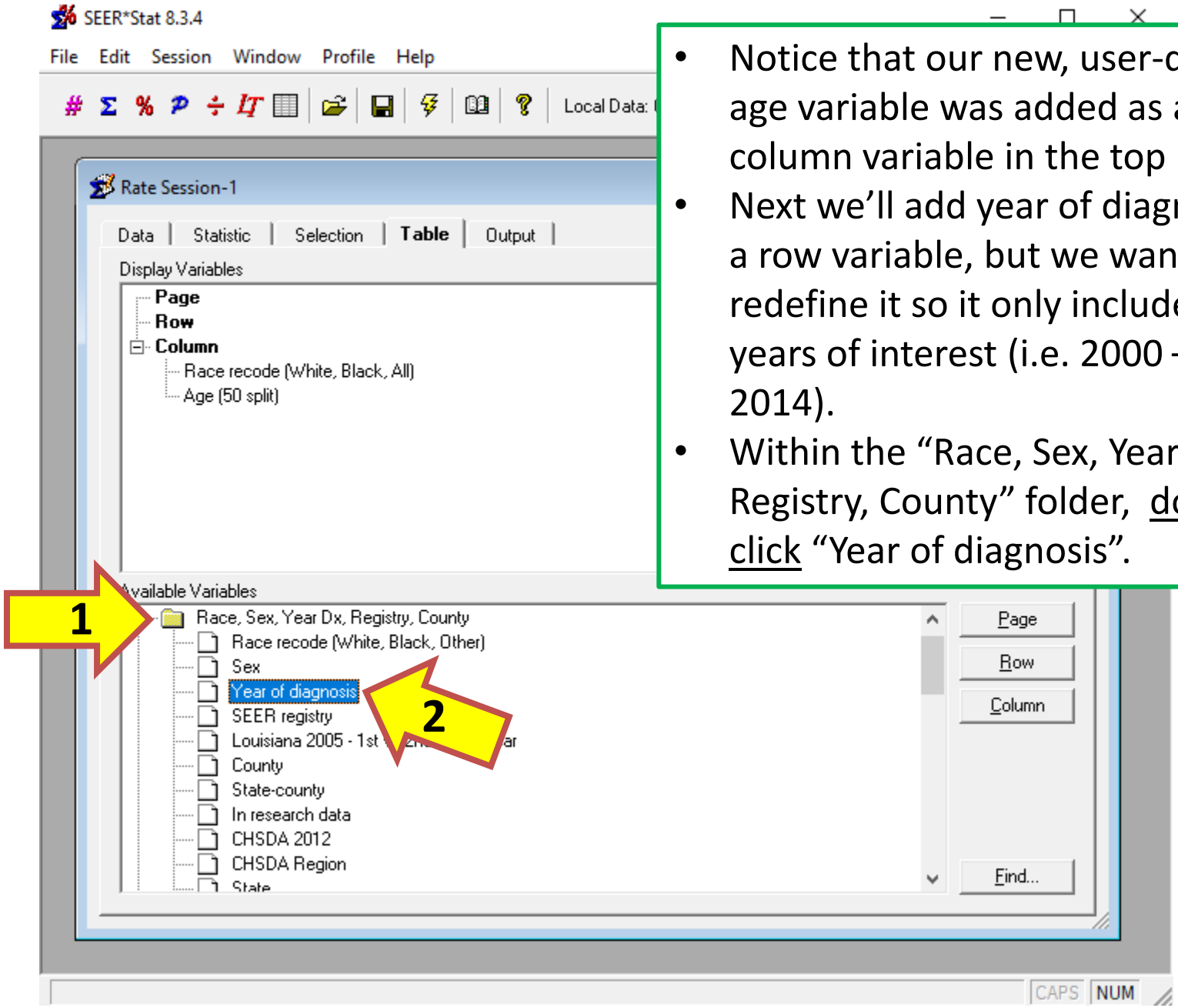


- Follow the previous steps once more, selecting “50-54 years” to “85+ years” (exclude the Unknown group).
- Click “Add…”, select “Added as one grouping”, click “Ok”
- Change the name for this grouping from “50-54 years” (default) to “50+ years”.
- Once you’re finished, click “OK”.

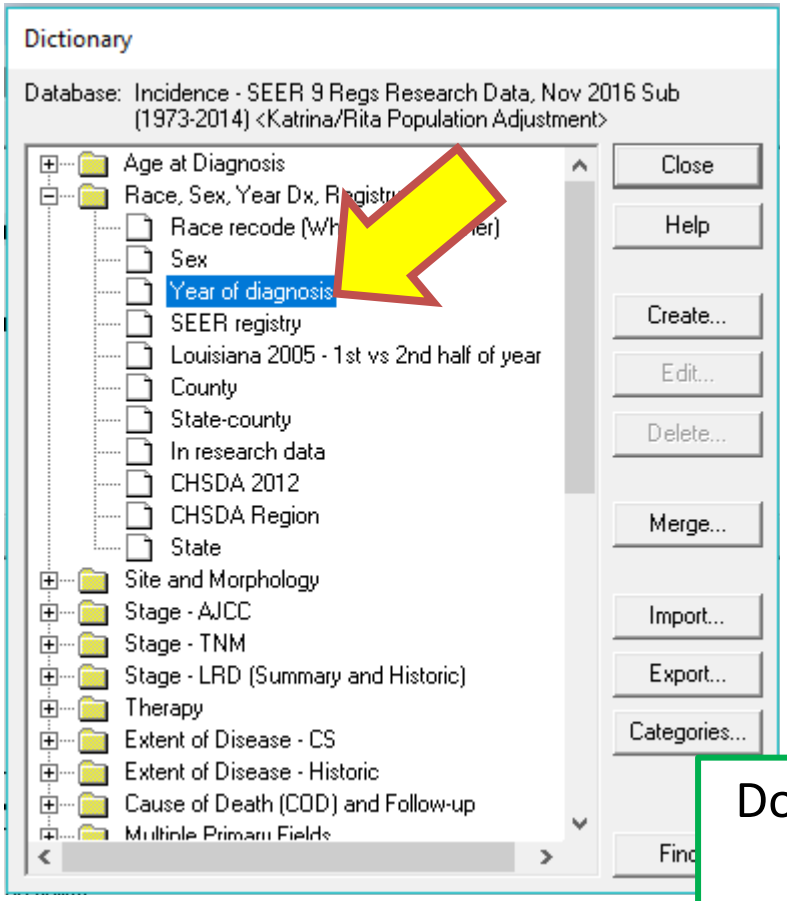

Dictionary
Database: Incidence - SEER 9 Regs Research Data, Nov 2016 Sub (1973-2014) <Katrina/Rita Population Adjustment>
Site and Morphology - Historic
Race and Age (case data only)
Dates
Site Specific Sequence Numbers
Other
County attributes ACS - 2011-2015
County attributes ACS - 2010-2014
County attributes ACS - 2009-2013
County attributes ACS - 2008-2012
County attributes ACS - 2007-2011
County attributes 2010s
County attributes 2000s
County attributes - modeled small area estimate
County attributes 1990s
County attributes 1980s
County attributes 1970s
utes General
lied
em-Supplied
User-Defined
Age (50 split)
Close
Help
Create...
Edit...
Delete...
Merge...
Import...
Export...
Categories...
Find...

Click “Close”.

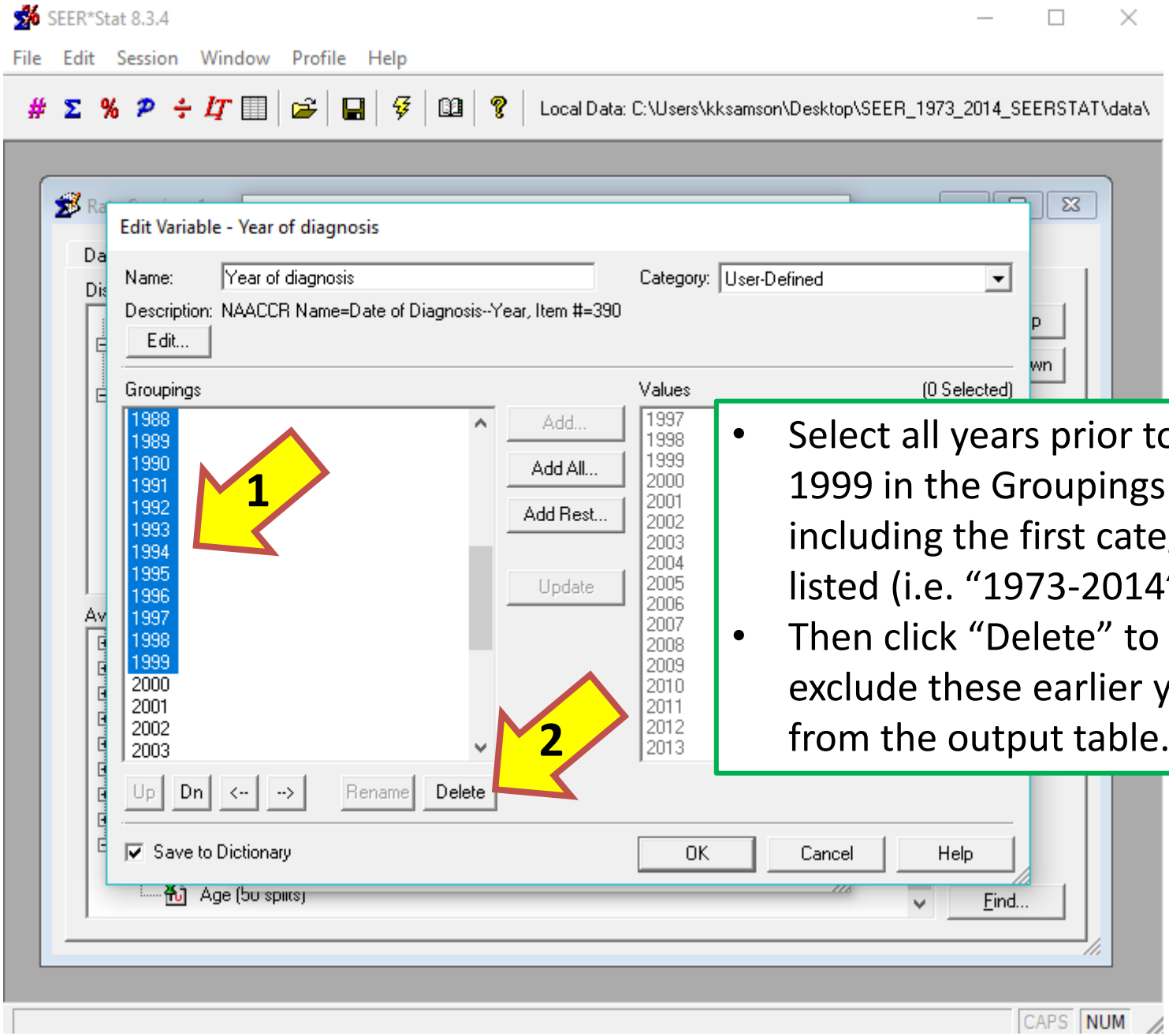

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER_1973_2014_SEERSTAT\data\
Rate Session-1
Data Statistic Selection Table Output
Display Variables
Page
Row
Column
Race recode (White, Black, Al
Move Up
Move Down
Remove
Available Variables
County attributes 2000s
County attributes - modeled s
County attributes 1990s
County attributes 1980s
County attributes 1970s
County attributes General
System-Supplied
Merged System-Supplie
User-Defined
Age (50 split)
Race recode (Whit lack, All)
Page
Row
Column
Find...
1
2
3
• Find the "User-Defined" folder in the "Available Variables" box.
• Find and select your newly defined variable grouping "Age (50 split)" and click the "Column" button.
CAPS
NUM



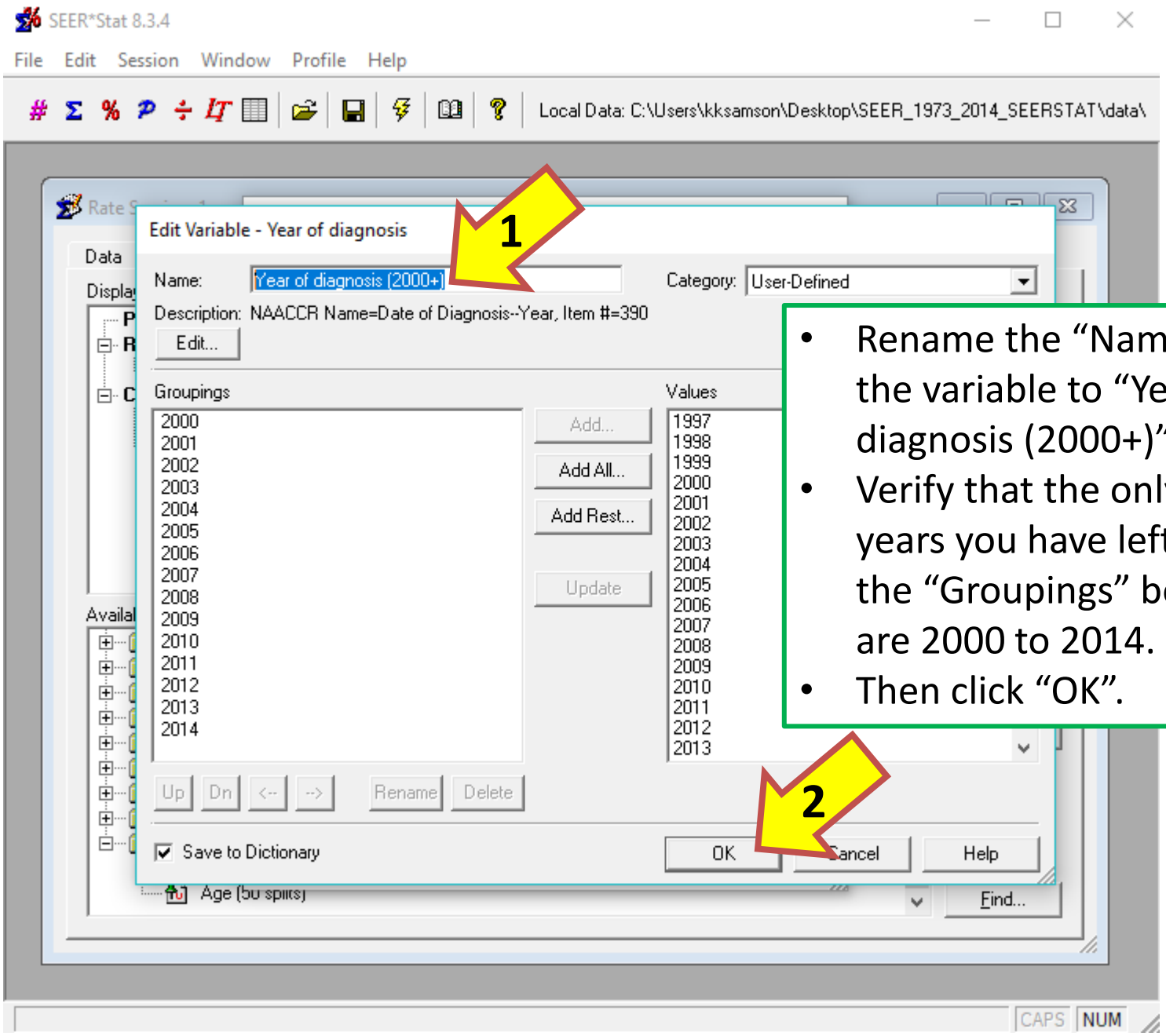
- Notice that our new, user-defined age variable was added as a column variable in the top box.
- Next we'll add year of diagnosis as a row variable, but we want to redefine it so it only includes years of interest (i.e. 2000 – 2014).
- Within the "Race, Sex, Year Dx, Registry, County" folder, <u>double click</u> "Year of diagnosis".



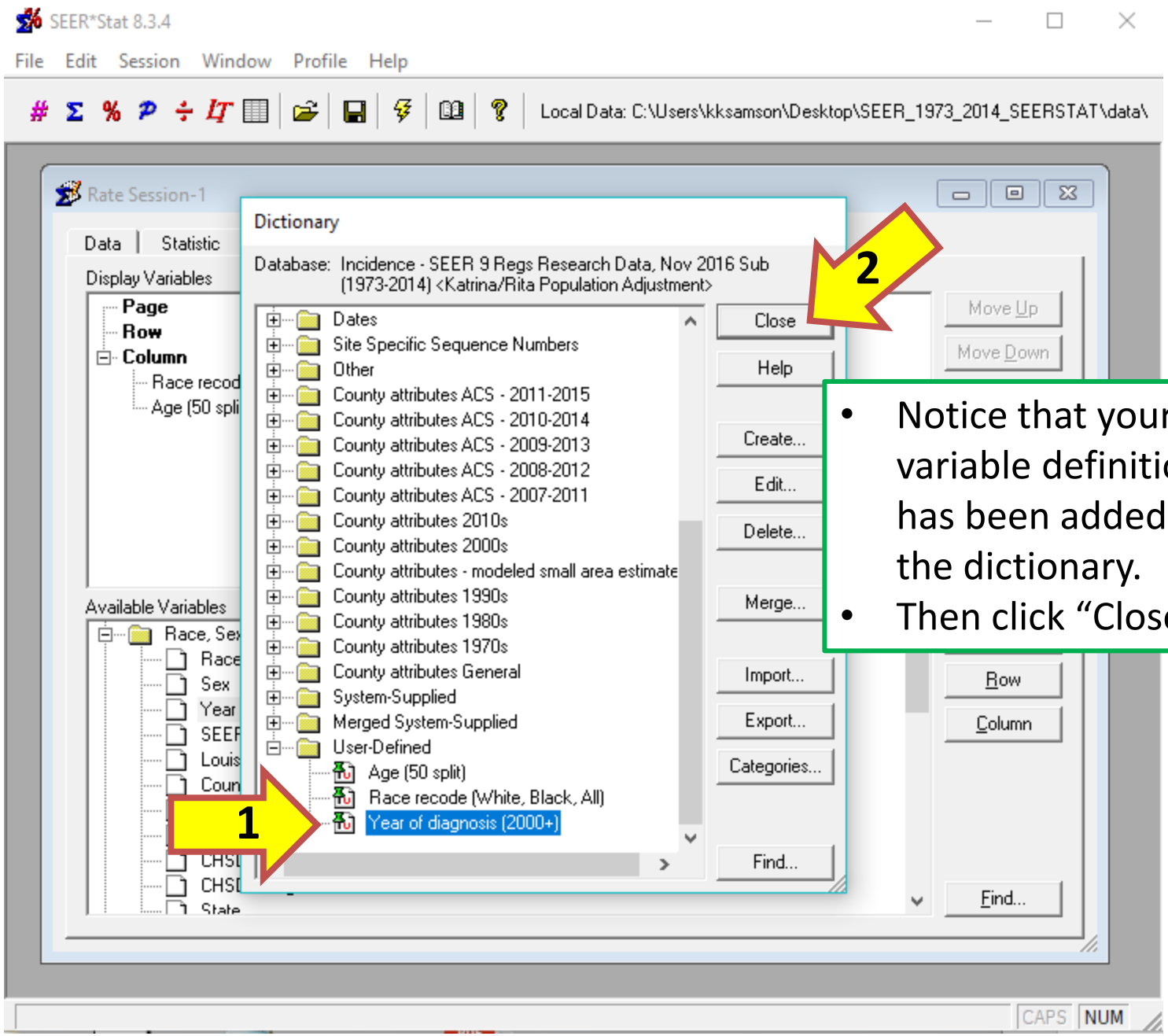
Double click on “Year of diagnosis” in the Dictionary pop-up.



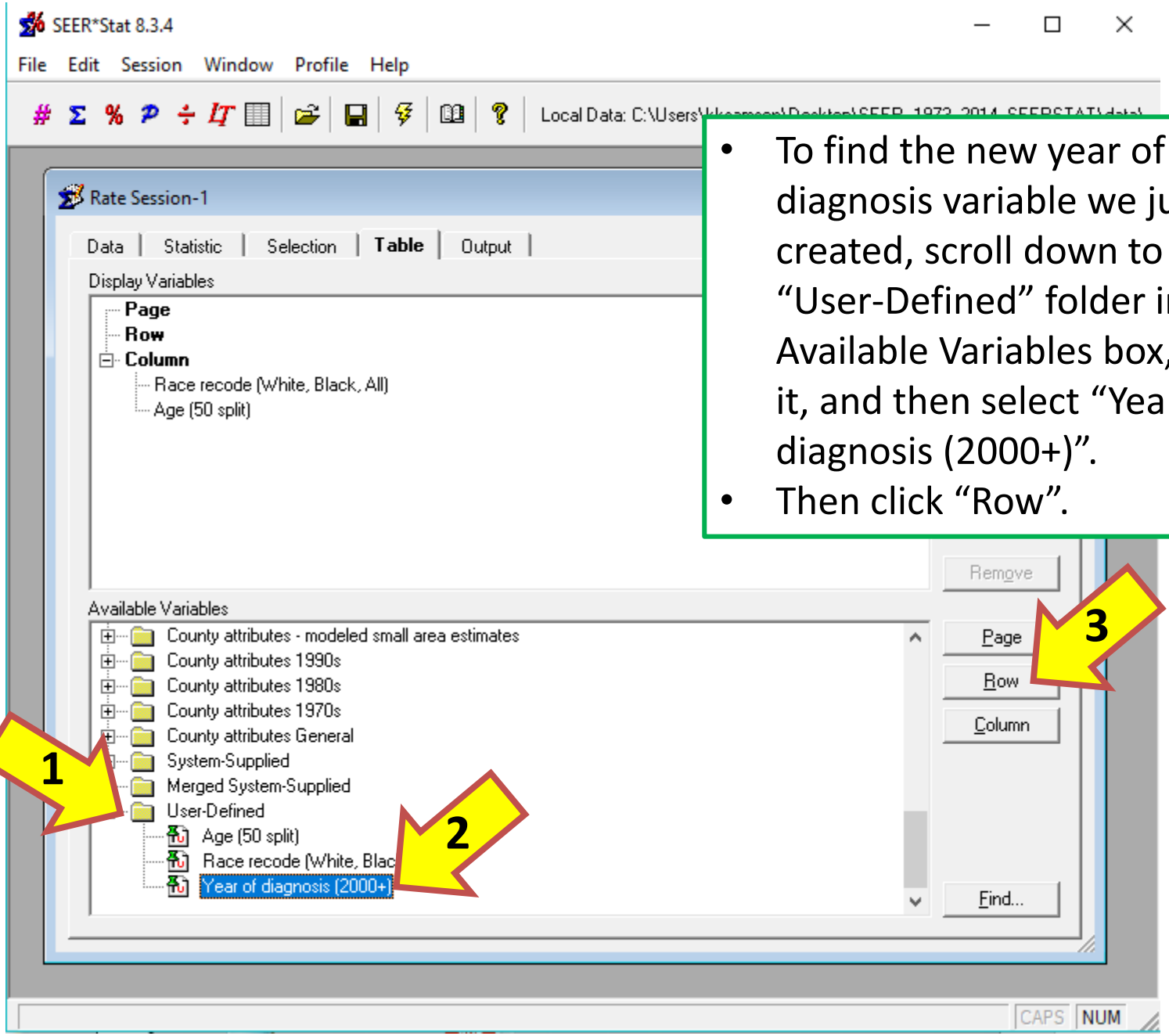
- Select all years prior to 1999 in the Groupings box, including the first category listed (i.e. “1973-2014”).
- Then click “Delete” to exclude these earlier years from the output table.



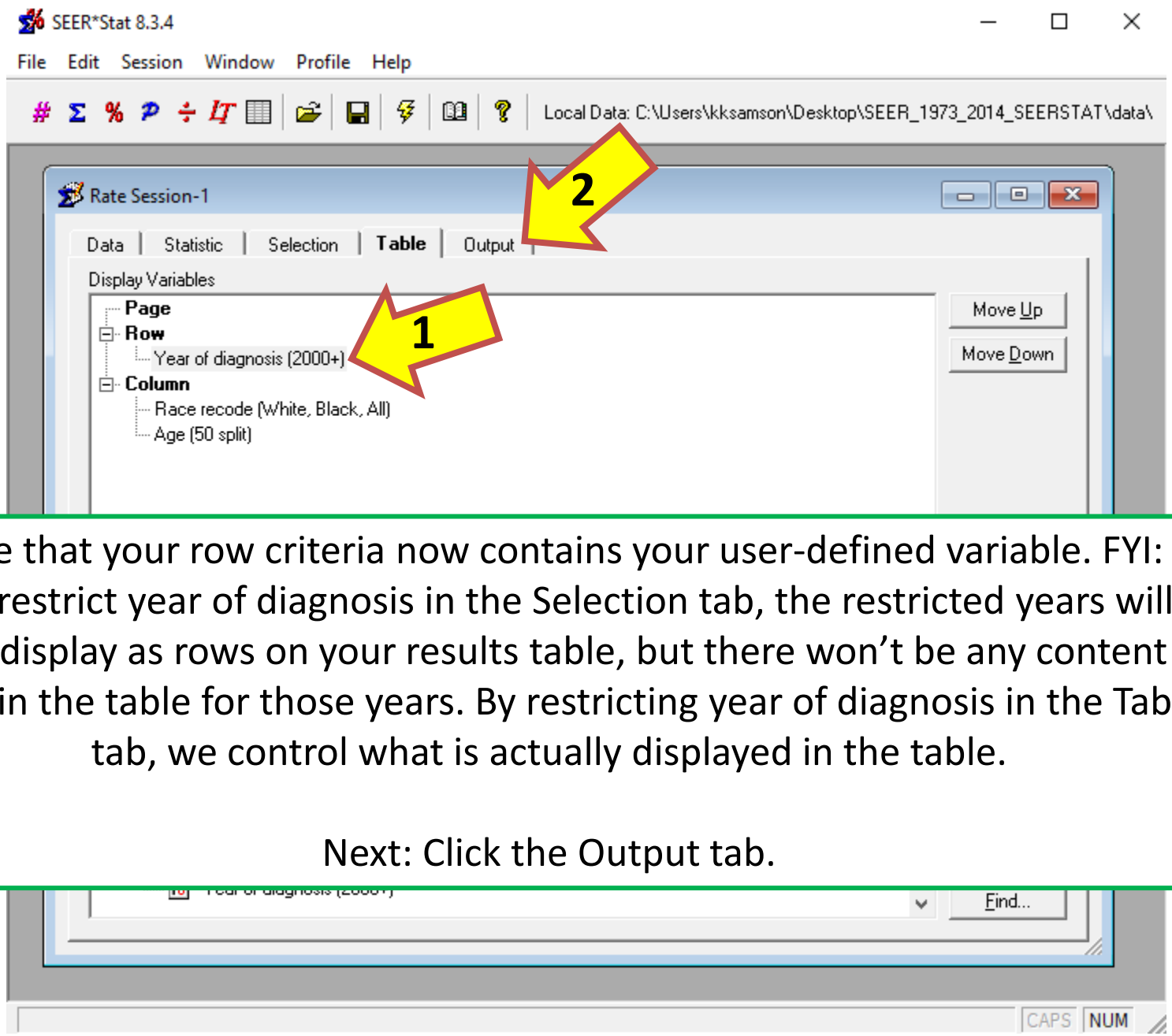
- Rename the "Name" of the variable to "Year of diagnosis (2000+)".
- Verify that the only years you have left in the "Groupings" box are 2000 to 2014.
- Then click "OK".



- Notice that your new variable definition has been added to the dictionary.
- Then click “Close”.

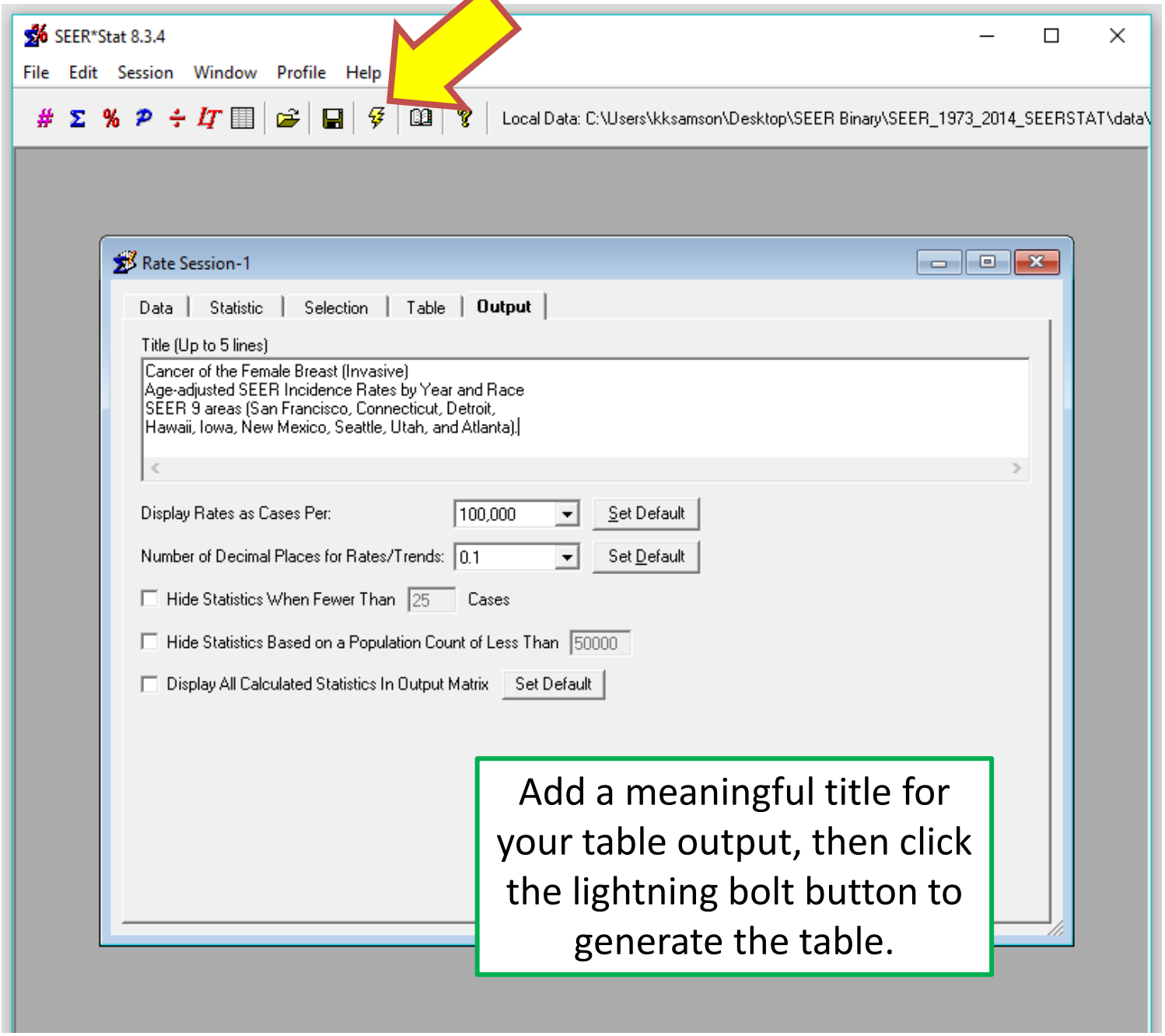


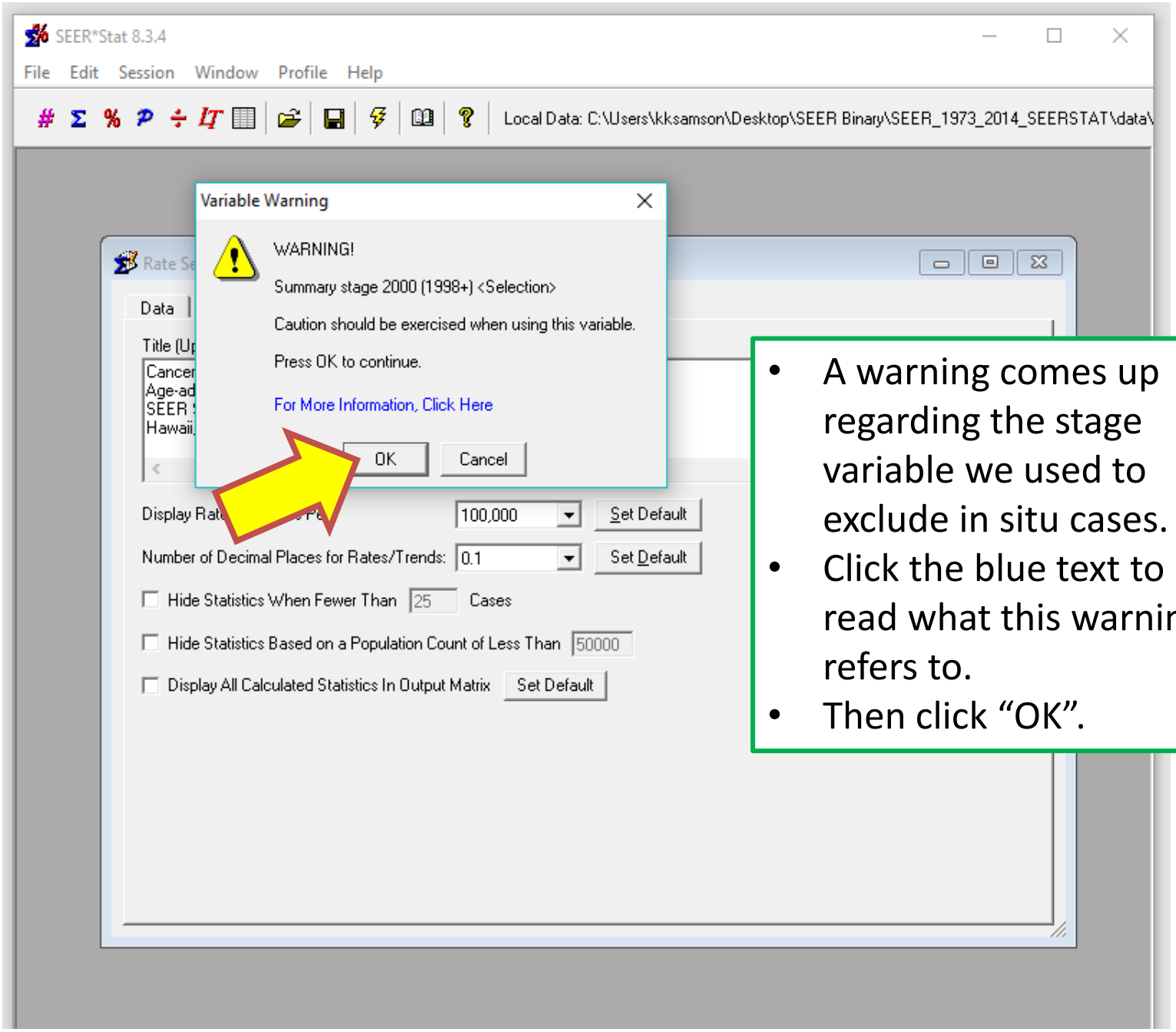
- To find the new year of diagnosis variable we just created, scroll down to the "User-Defined" folder in the Available Variables box, click it, and then select "Year of diagnosis (2000+)".
- Then click "Row".



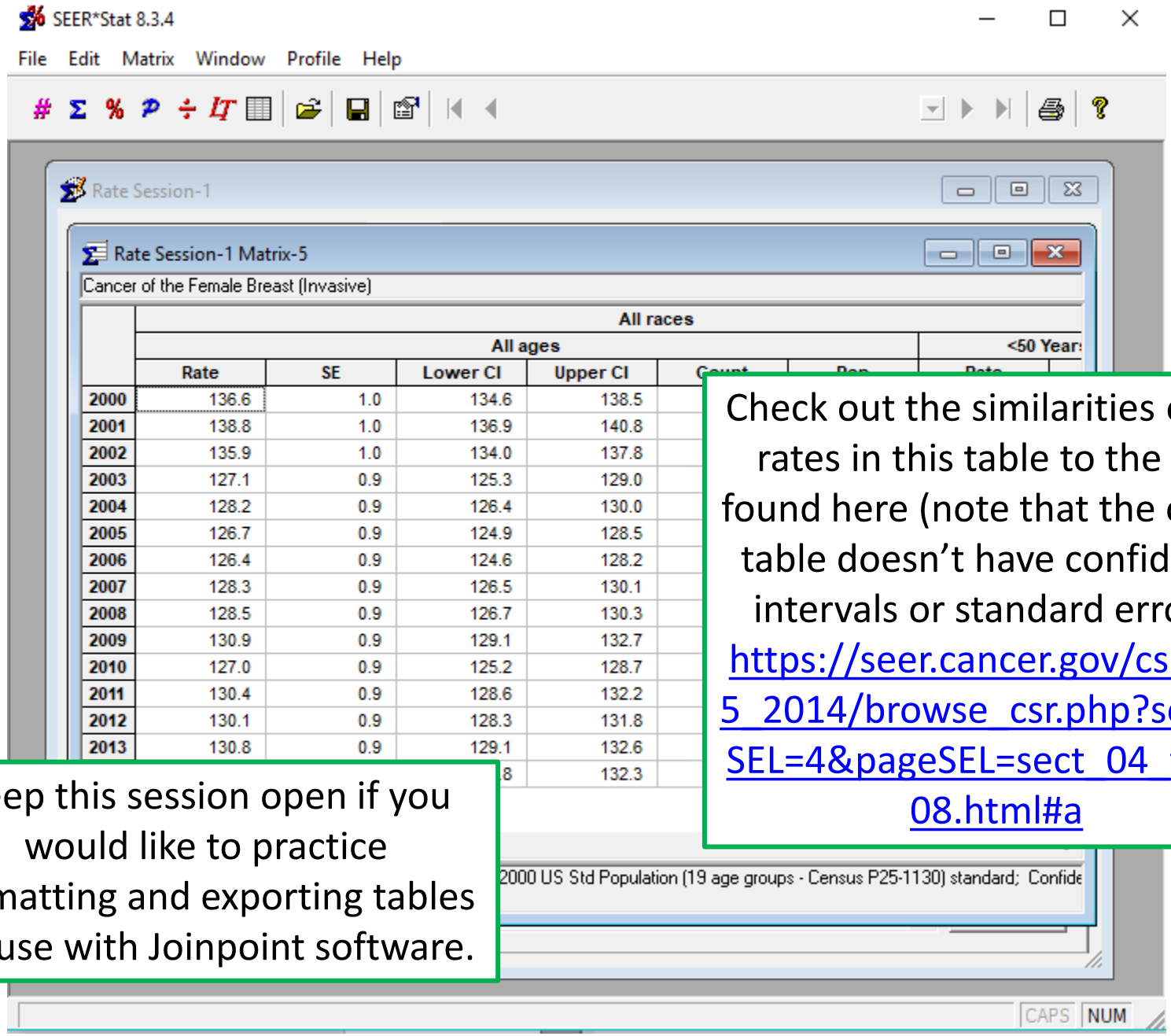
Notice that your row criteria now contains your user-defined variable. FYI: If you restrict year of diagnosis in the Selection tab, the restricted years will still display as rows on your results table, but there won't be any content listed in the table for those years. By restricting year of diagnosis in the Table tab, we control what is actually displayed in the table.

Next: Click the Output tab.


SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER Binary\SEER_1973_2014_SEERSTAT\data\
Rate Session-1
Data Statistic Selection Table Output
Title (Up to 5 lines)
Cancer of the Female Breast (Invasive)
Age-adjusted SEER Incidence Rates by Year and Race
SEER 9 areas (San Francisco, Connecticut, Detroit,
Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta).
Display Rates as Cases Per: 100,000 Set Default
Number of Decimal Places for Rates/Trends: 0.1 Set Default
Hide Statistics When Fewer Than 25 Cases
Hide Statistics Based on a Population Count of Less Than 50000
Display All Calculated Statistics In Output Matrix Set Default
Add a meaningful title for your table output, then click the lightning bolt button to generate the table.



- A warning comes up regarding the stage variable we used to exclude in situ cases.
- Click the blue text to read what this warning refers to.
- Then click “OK”.



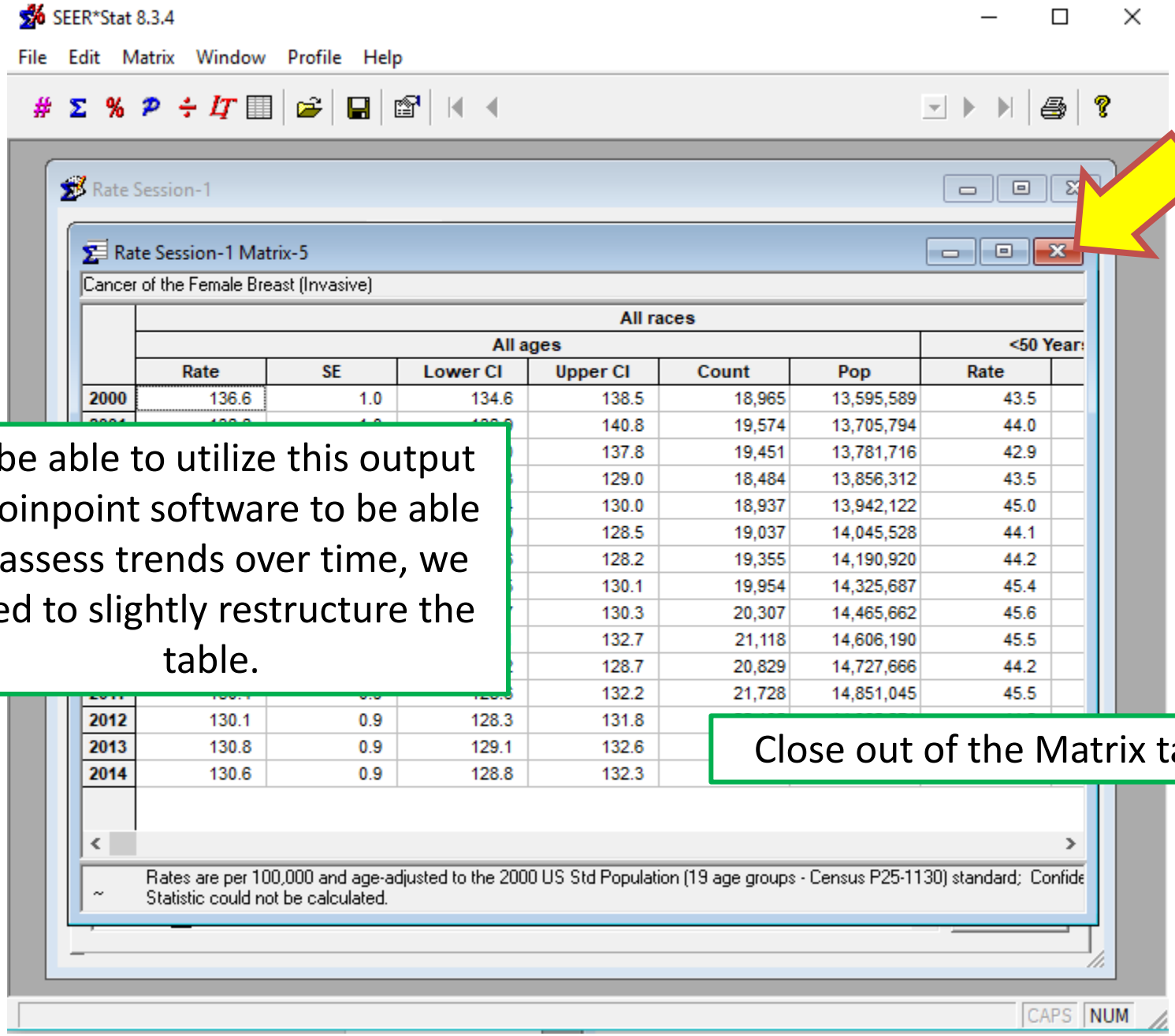
Check out the similarities of the rates in this table to the one found here (note that the online table doesn't have confidence intervals or standard errors): https://seer.cancer.gov/csr/1975_2014/browse_csr.php?sectionSEL=4&pageSEL=sect_04_table.08.html#a

Keep this session open if you would like to practice formatting and exporting tables for use with Joinpoint software.

Reformatting Incidence Output and Exporting Files for Use in Joinpoint

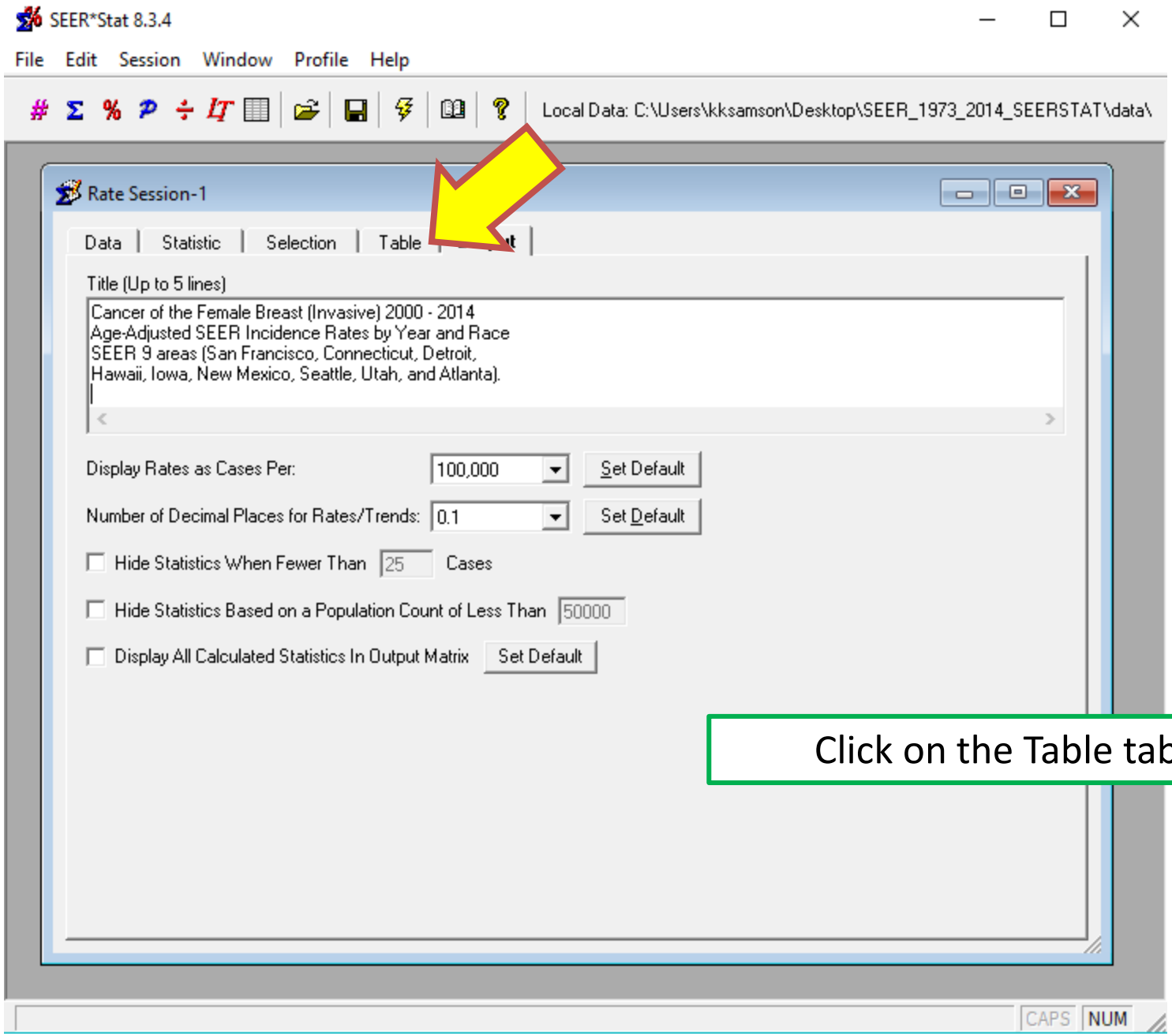
- To run an example in Joinpoint, we will use the data from the previous age-adjusted incidence rate session from this tutorial.

- While the table in the previous example is straightforward and easy to read, it cannot be used in Joinpoint software. Therefore, we will restructure the table so that we can use it.

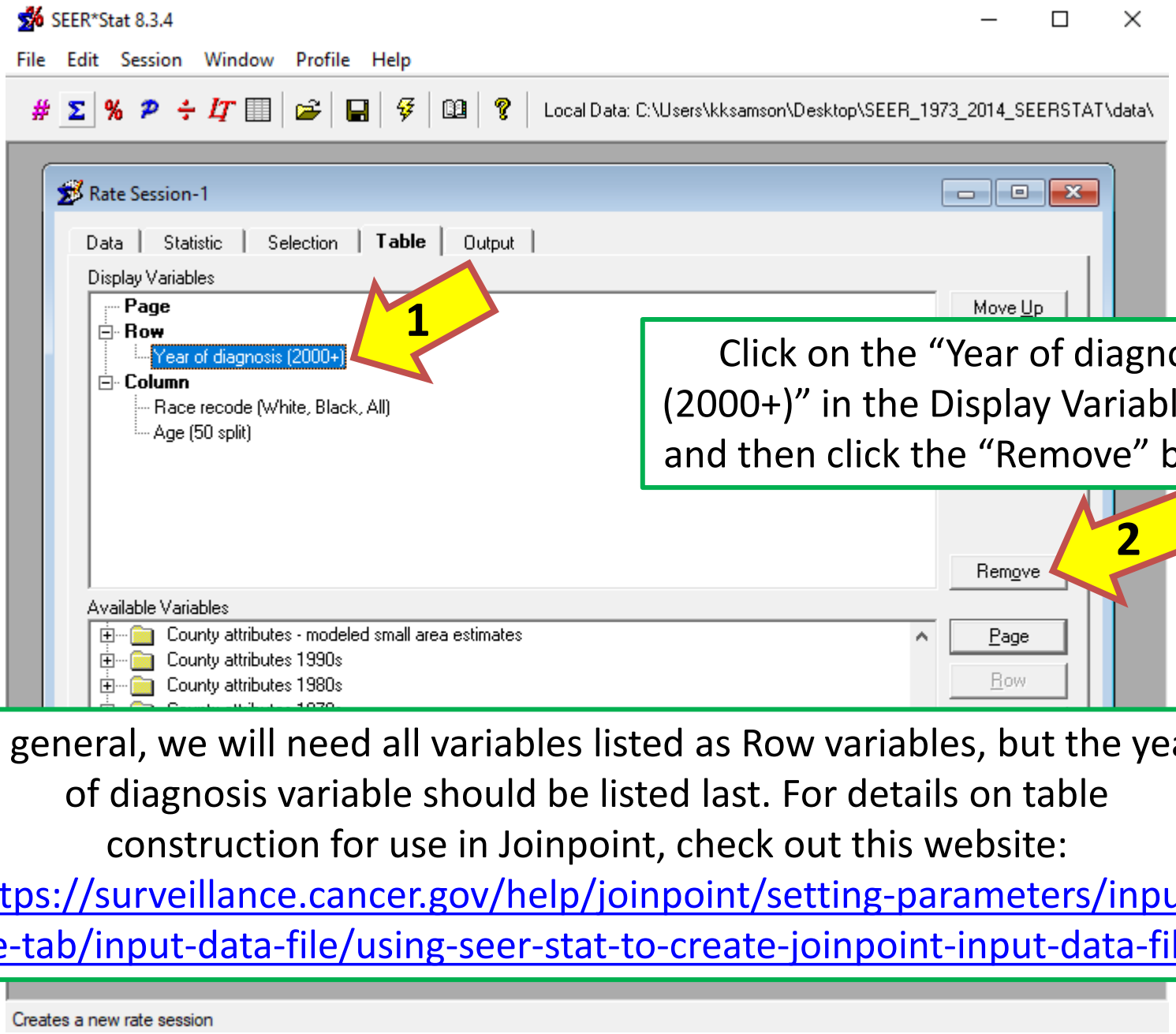


To be able to utilize this output in Joinpoint software to be able to assess trends over time, we need to slightly restructure the table.

Close out of the Matrix table.

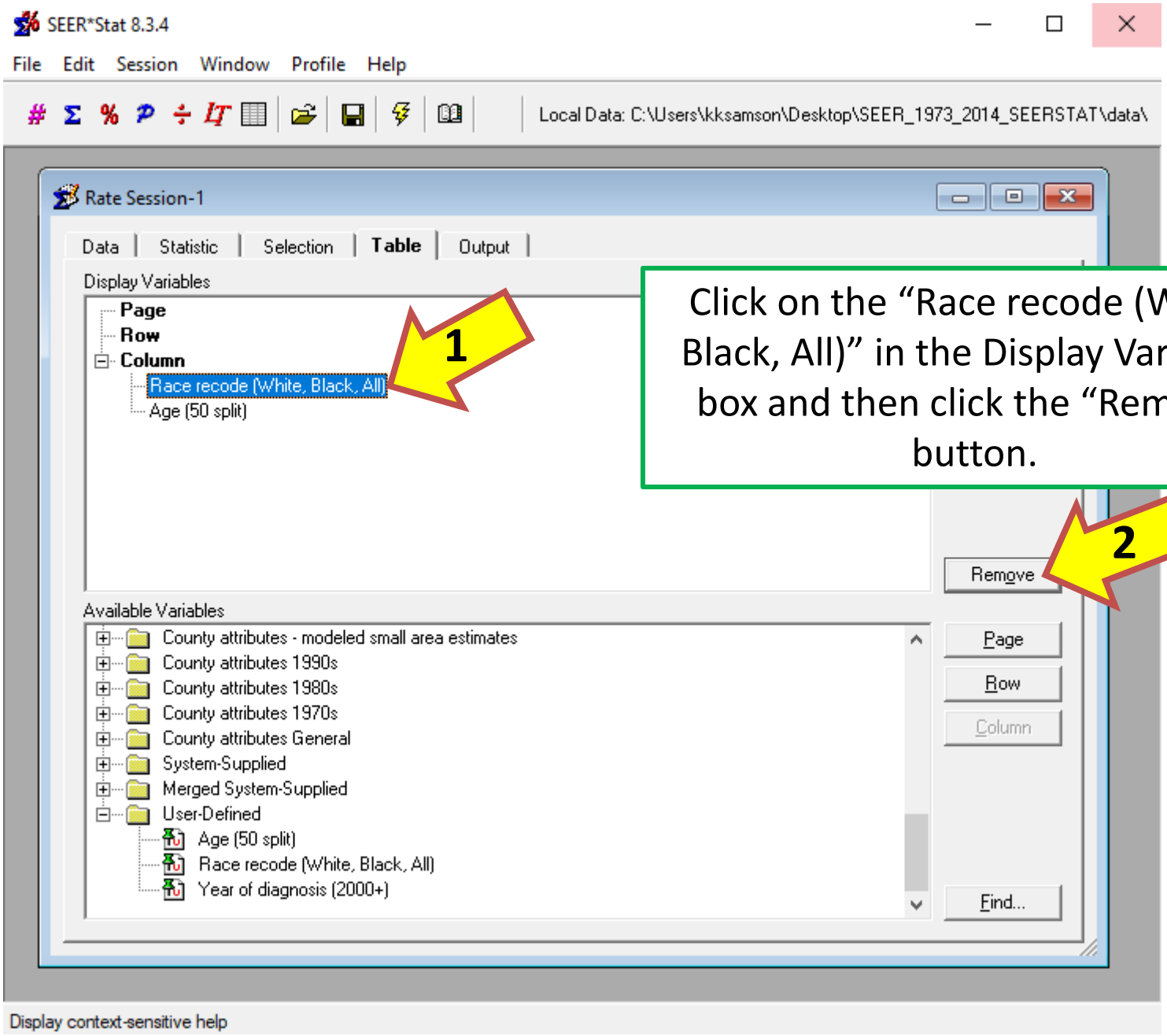


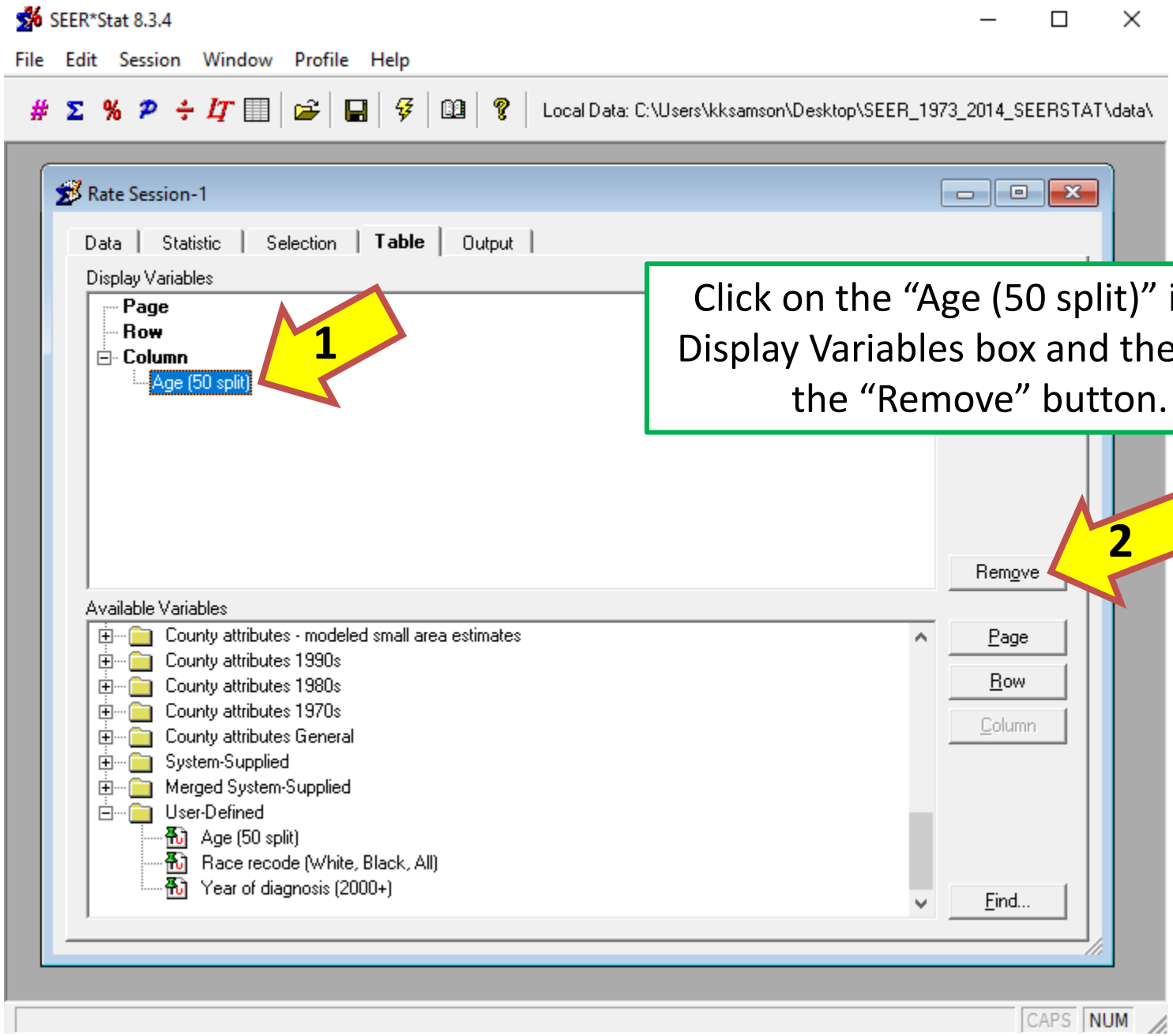
Click on the Table tab.

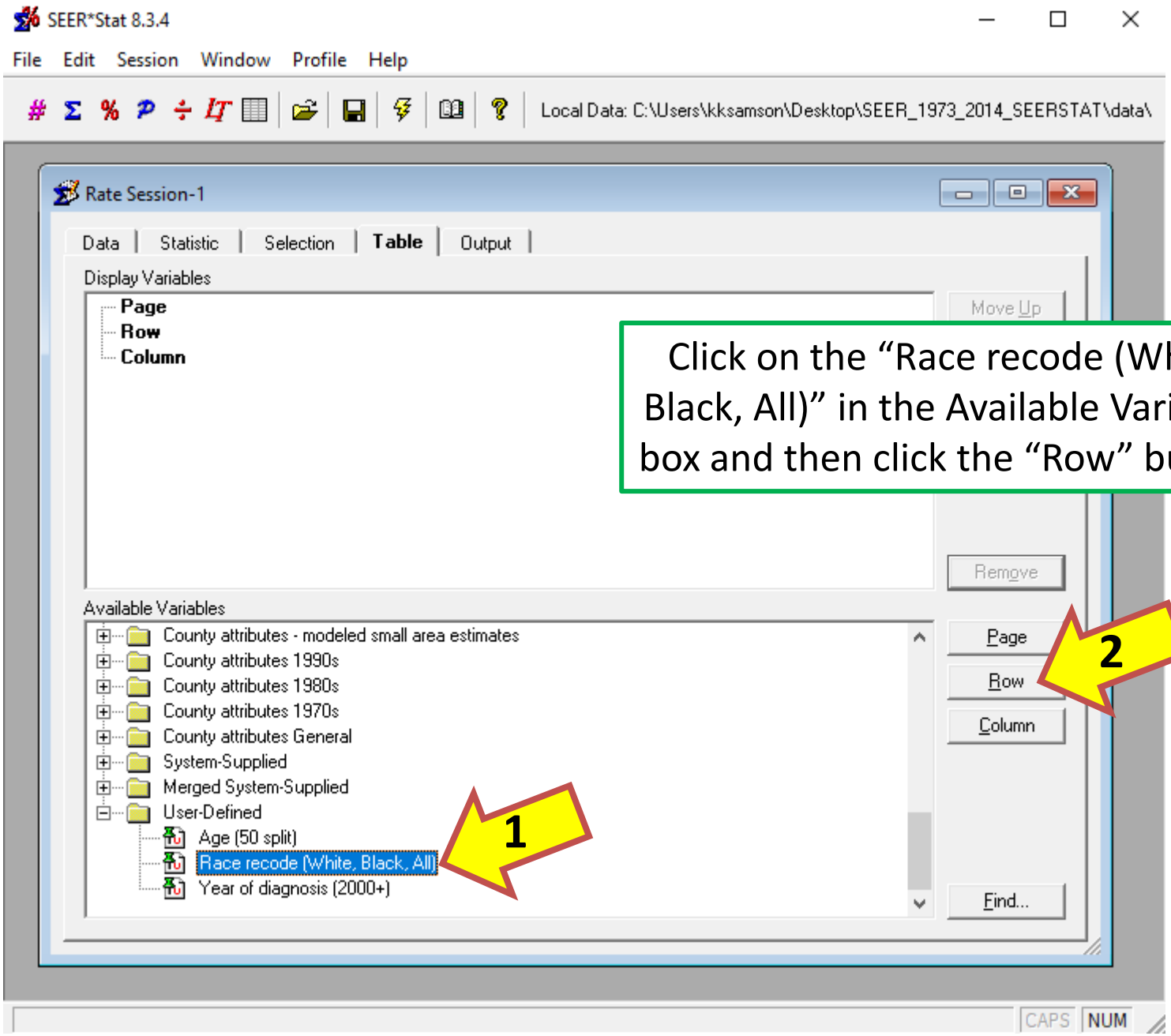


Click on the "Year of diagnosis (2000+)" in the Display Variables box and then click the "Remove" button.

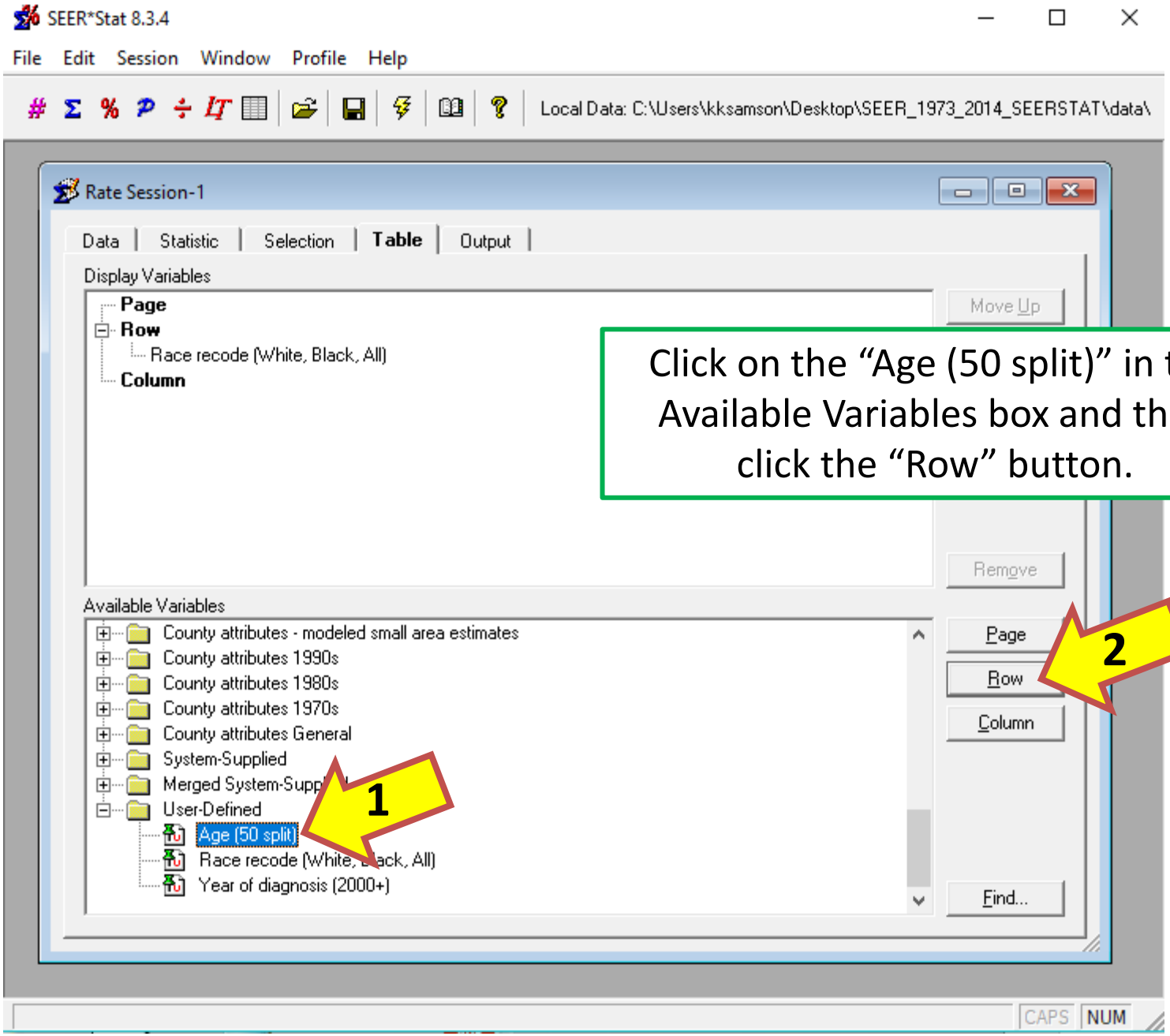
In general, we will need all variables listed as Row variables, but the year of diagnosis variable should be listed last. For details on table construction for use in Joinpoint, check out this website: https://surveillance.cancer.gov/help/joinpoint/setting-parameters/input-file-tab/input-data-file/using-seer-stat-to-create-joinpoint-input-data-files


SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER_1973_2014_SEERSTAT\data\
Rate Session-1
Data Statistic Selection Table Output
Display Variables
Page
Row
Column
Race recode (White, Black, All)
Age (50 split)
1
Click on the "Race recode (White, Black, All)" in the Display Variables box and then click the "Remove" button.
2
Remove
Available Variables
County attributes - modeled small area estimates
County attributes 1990s
County attributes 1980s
County attributes 1970s
County attributes General
System-Supplied
Merged System-Supplied
User-Defined
Age (50 split)
Race recode (White, Black, All)
Year of diagnosis (2000+)
Page
Row
Column
Find...
Display context-sensitive help

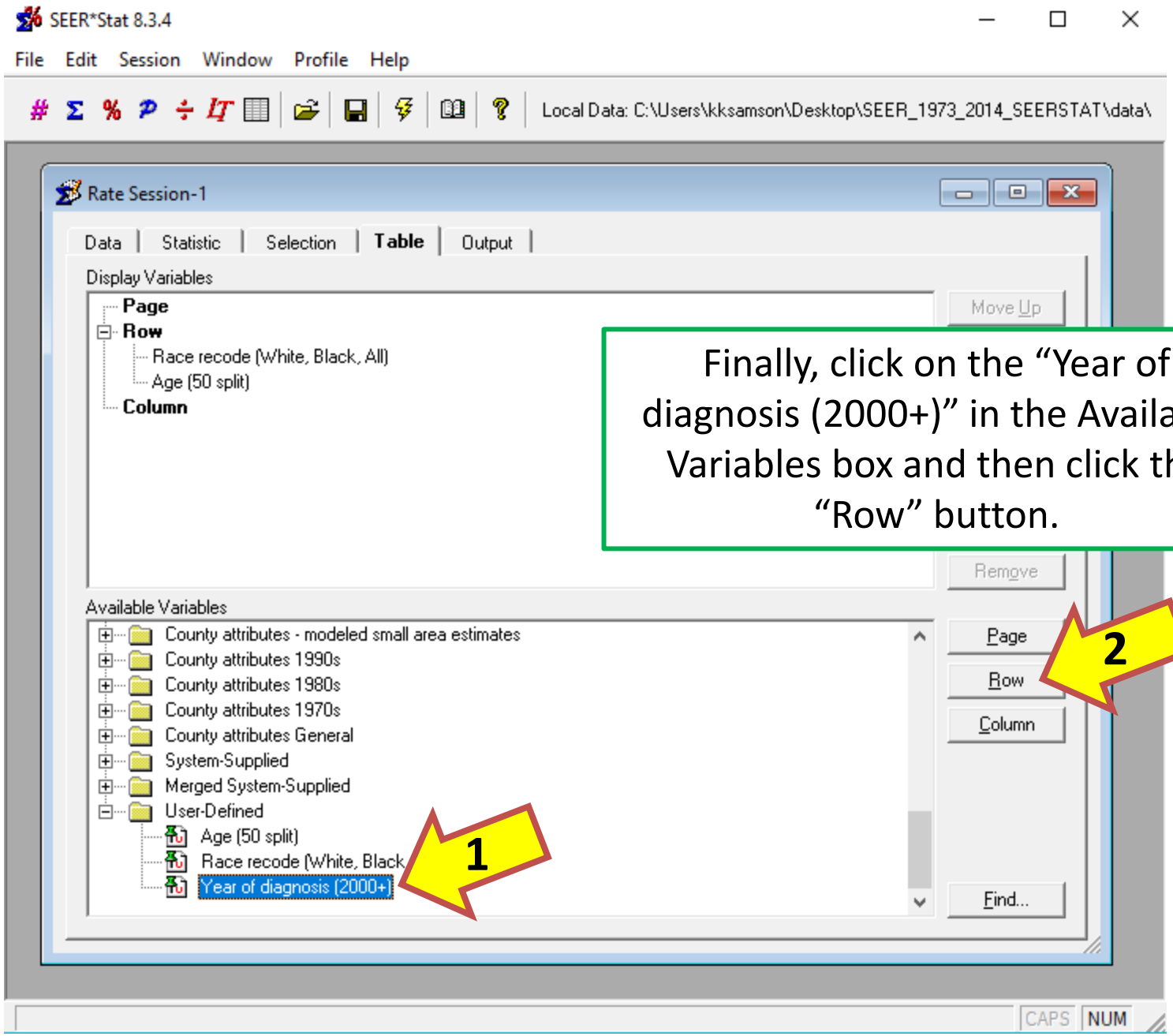

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER_1973_2014_SEERSTAT\data\
Rate Session-1
Data Statistic Selection Table Output
Display Variables
Page
Row
Column
Age (50 split)
1
Click on the "Age (50 split)" in the Display Variables box and then click the "Remove" button.
2
Remove
Available Variables
County attributes - modeled small area estimates
County attributes 1990s
County attributes 1980s
County attributes 1970s
County attributes General
System-Supplied
Merged System-Supplied
User-Defined
Age (50 split)
Race recode (White, Black, All)
Year of diagnosis (2000+)
Page
Row
Column
Find...
CAPS
NUM

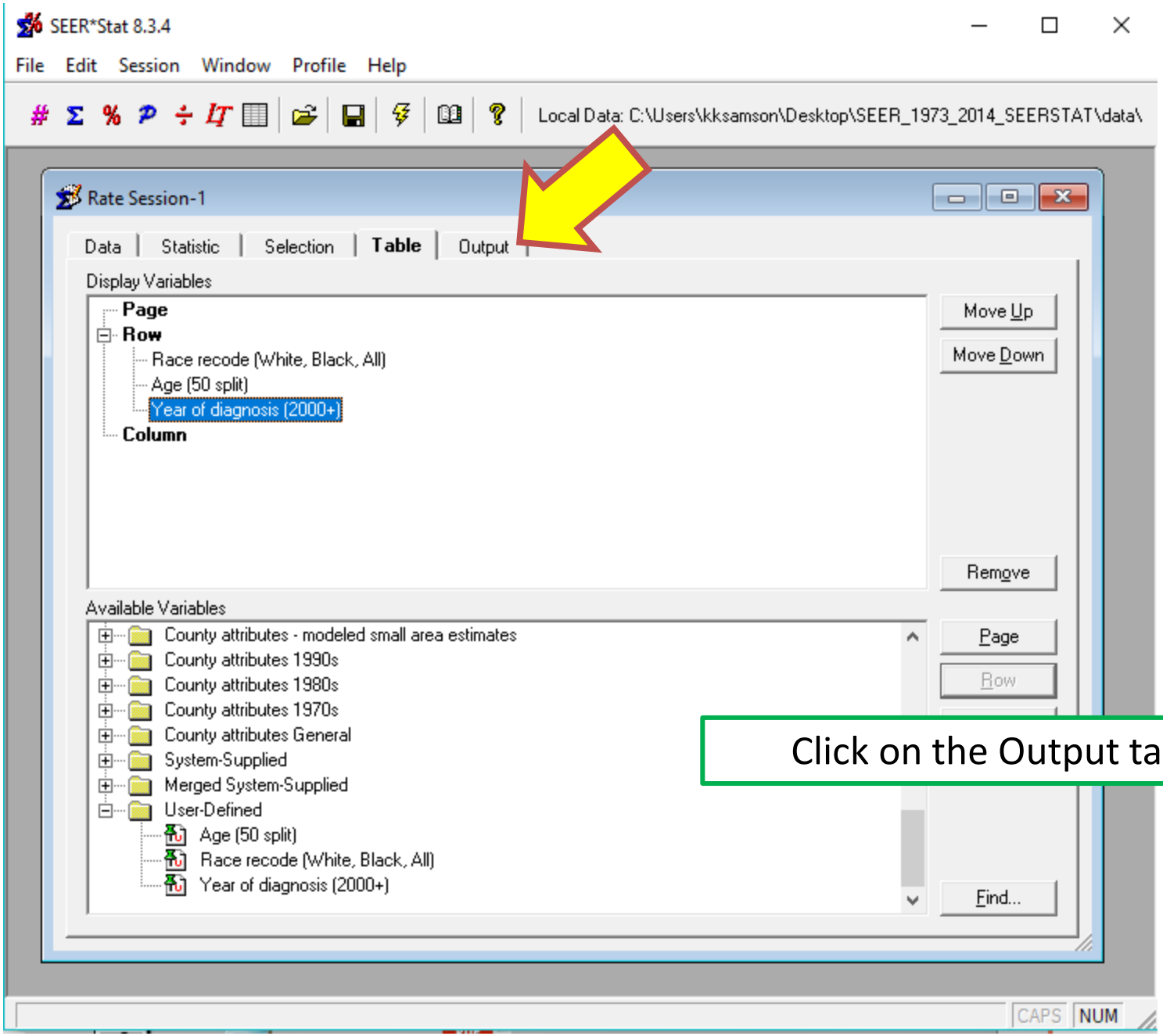


Click on the "Race recode (White, Black, All)" in the Available Variables box and then click the "Row" button.

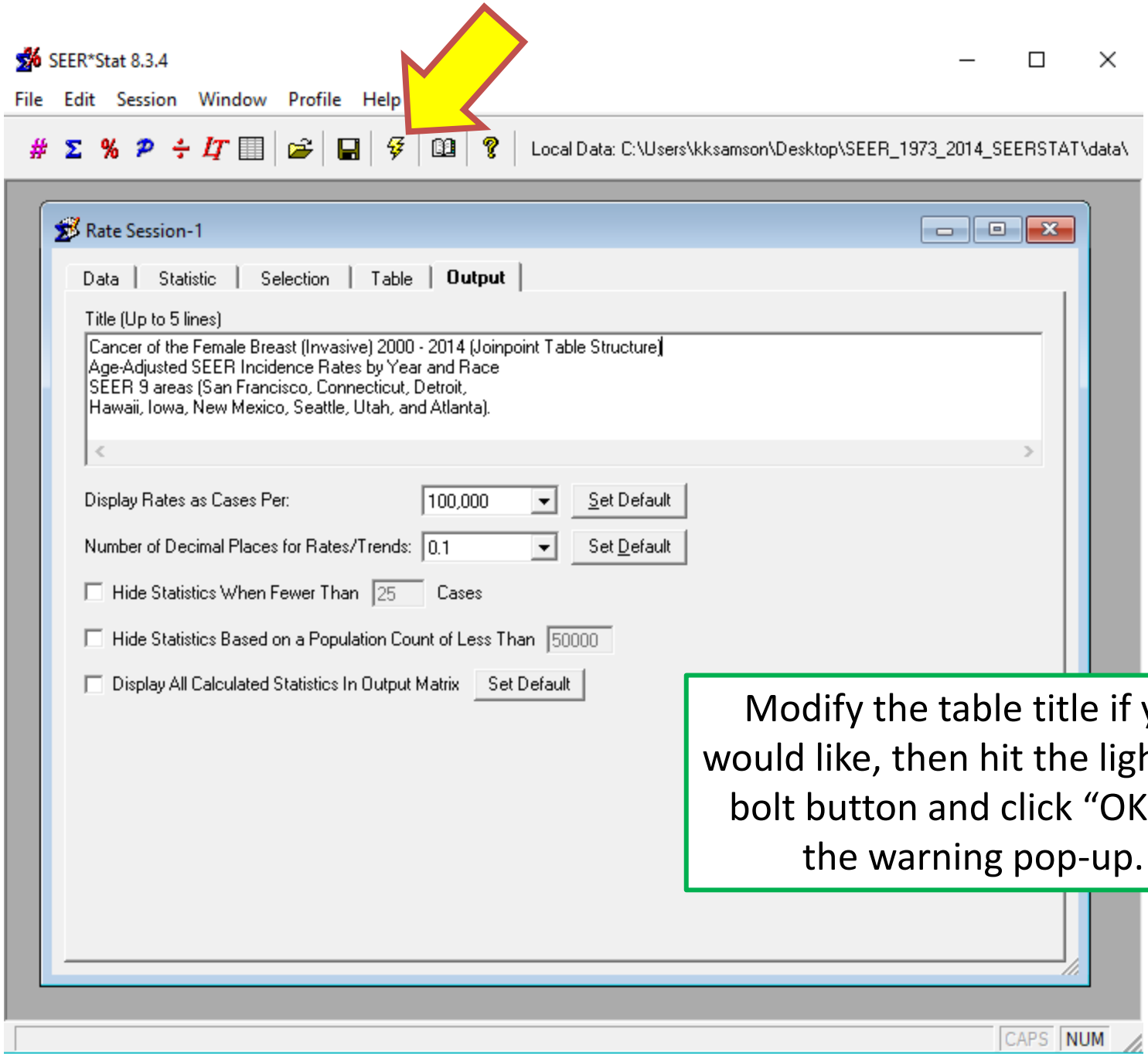


Click on the "Age (50 split)" in the Available Variables box and then click the "Row" button.

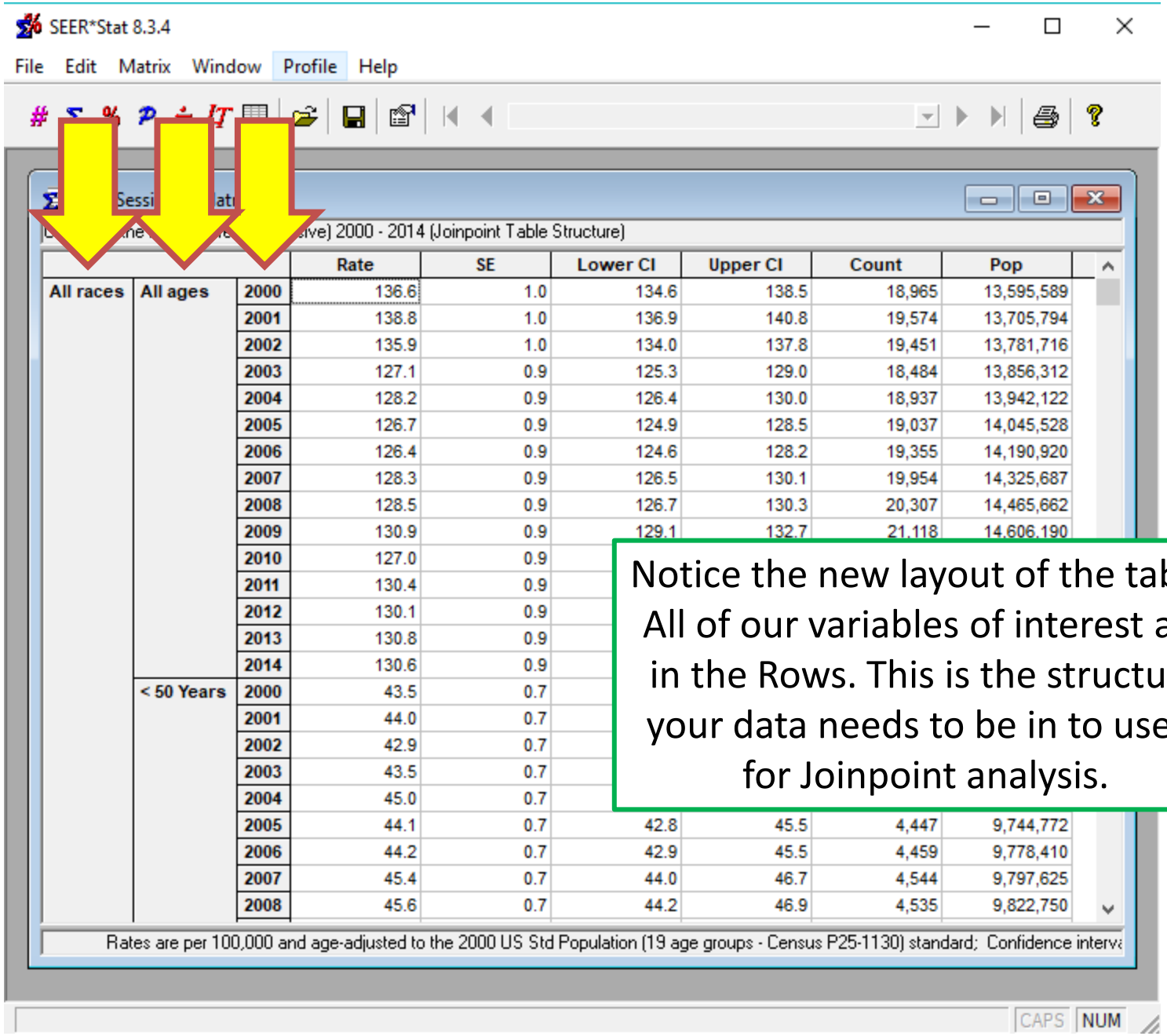

SEER*Stat 8.3.4
File Edit Session Window Profile Help
Local Data: C:\Users\kksamson\Desktop\SEER_1973_2014_SEERSTAT\data\
Rate Session-1
Data Statistic Selection Table Output
Display Variables
Page
Row
Race recode (White, Black, All)
Age (50 split)
Column
Move Up
Remove
Available Variables
County attributes - modeled small area estimates
County attributes 1990s
County attributes 1980s
County attributes 1970s
County attributes General
System-Supplied
Merged System-Supplied
User-Defined
Age (50 split)
Race recode (White, Black
Year of diagnosis (2000+)
Page
Row
Column
Find...
1
2
Finally, click on the "Year of diagnosis (2000+)" in the Available Variables box and then click the "Row" button.
CAPS NUM



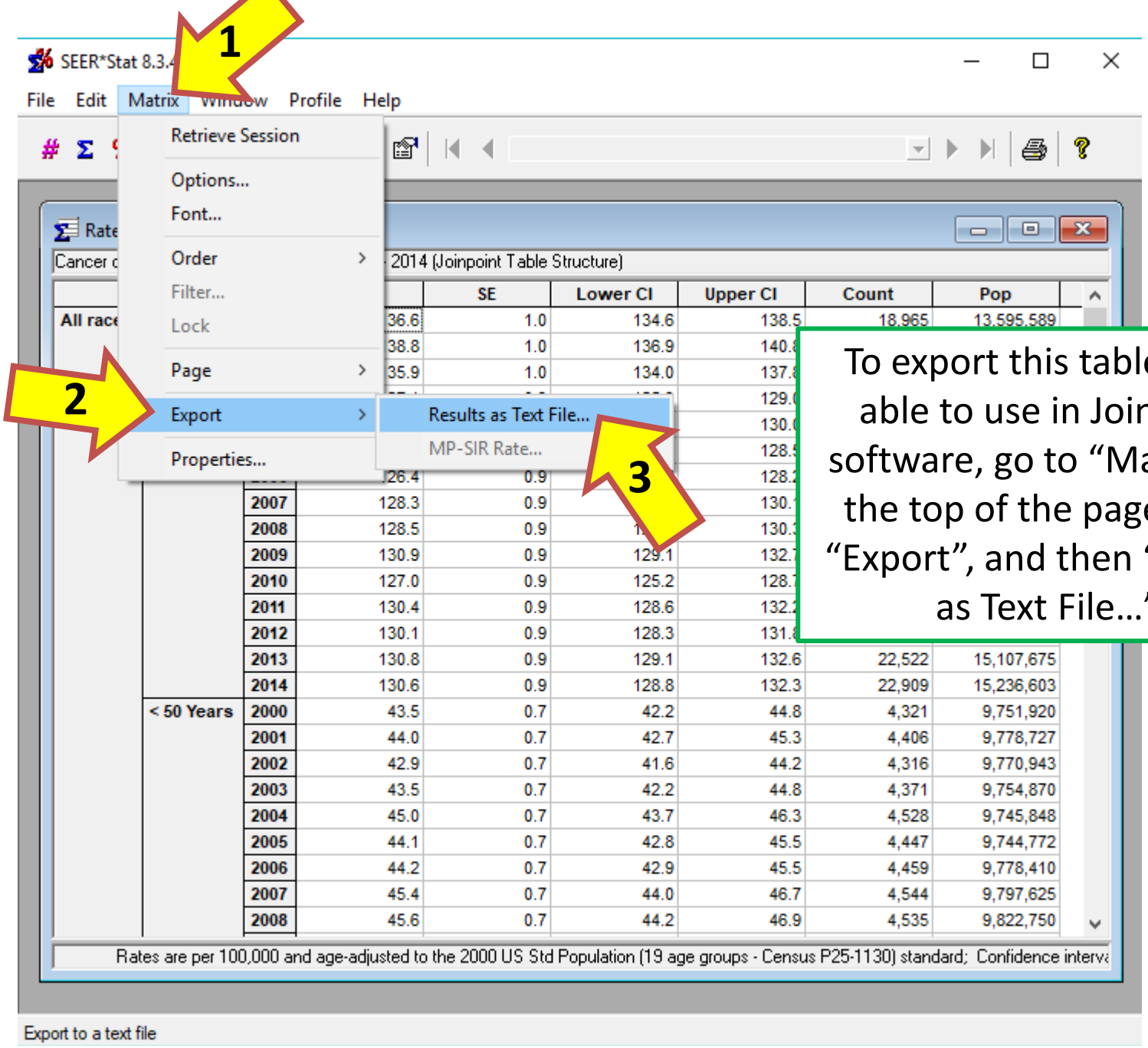
Click on the Output tab.



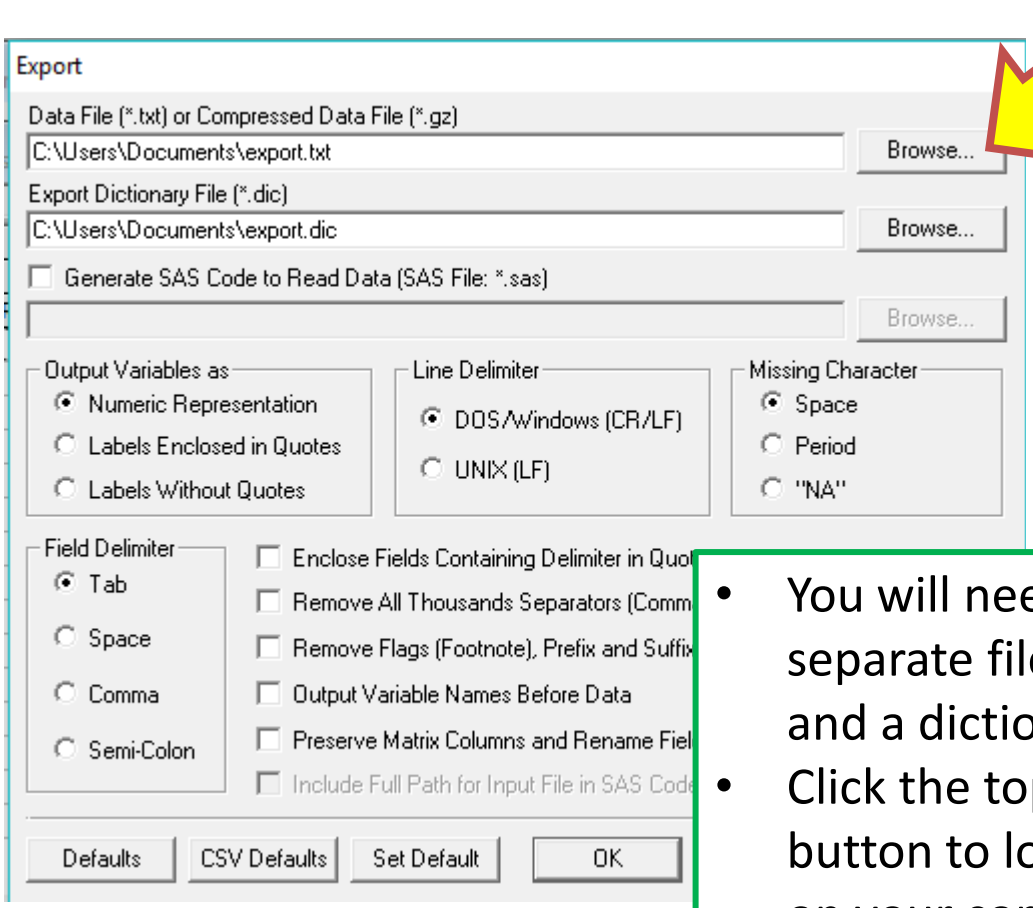
Modify the table title if you would like, then hit the lightning bolt button and click “OK” on the warning pop-up.



Notice the new layout of the table. All of our variables of interest are in the Rows. This is the structure your data needs to be in to use it for Joinpoint analysis.

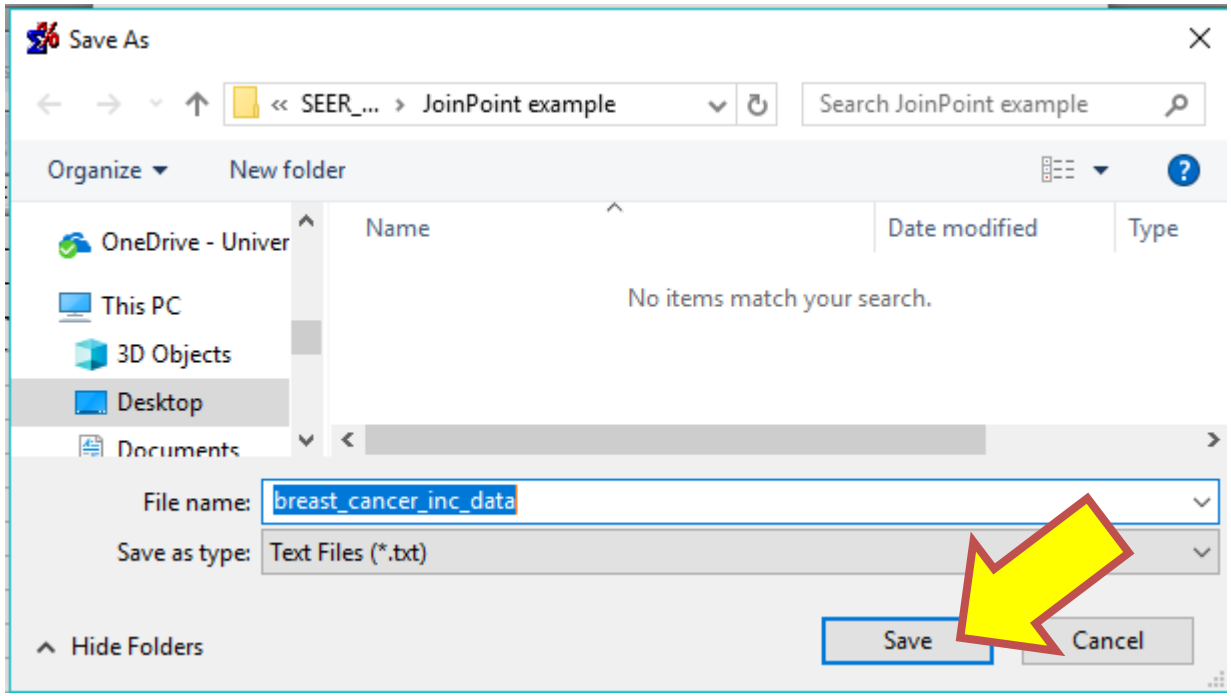


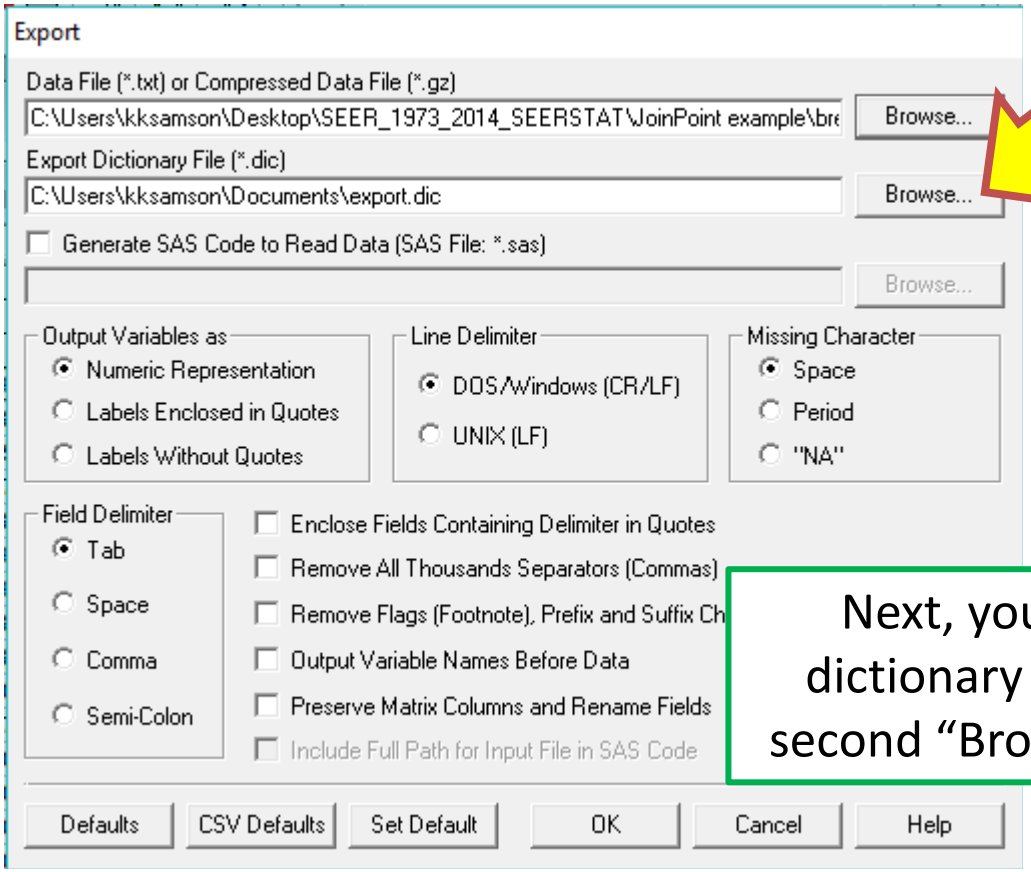
To export this table to be able to use in Joinpoint software, go to “Matrix” at the top of the page, then “Export”, and then “Results as Text File…”.



- You will need to save two separate files: a data file and a dictionary file.
- Click the top “Browse…” button to locate a folder on your computer where you will save the data file.

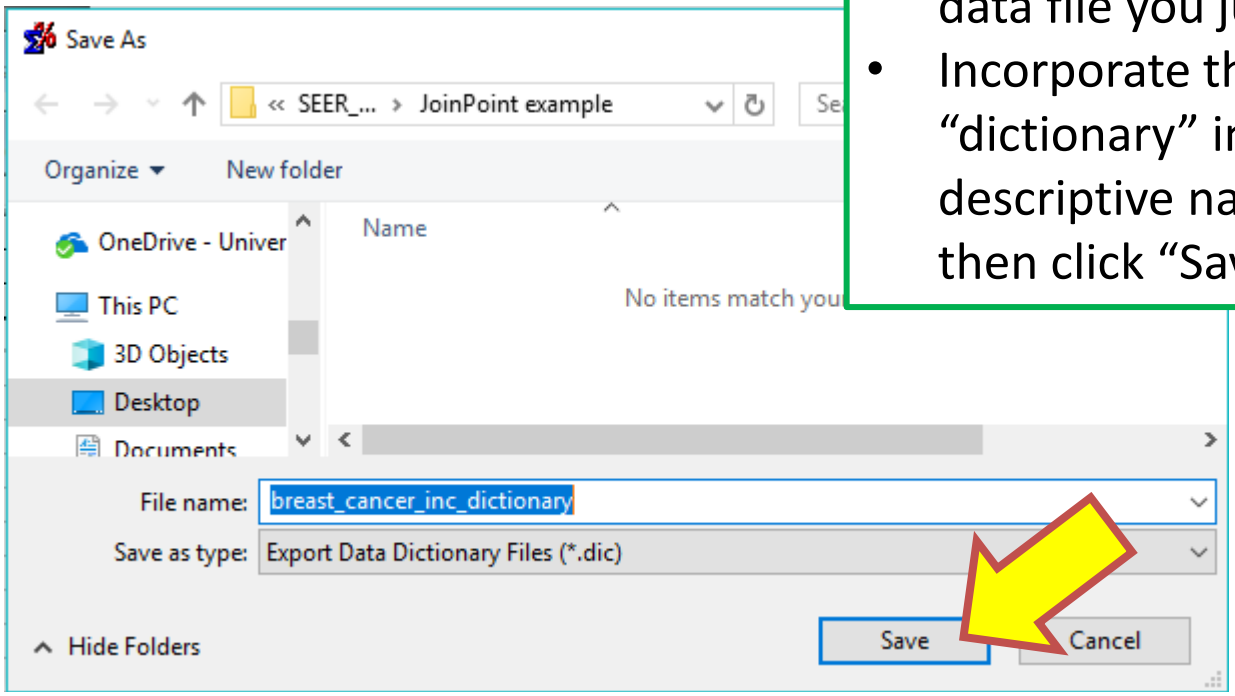
Give the file a descriptive name and then click “Save”.

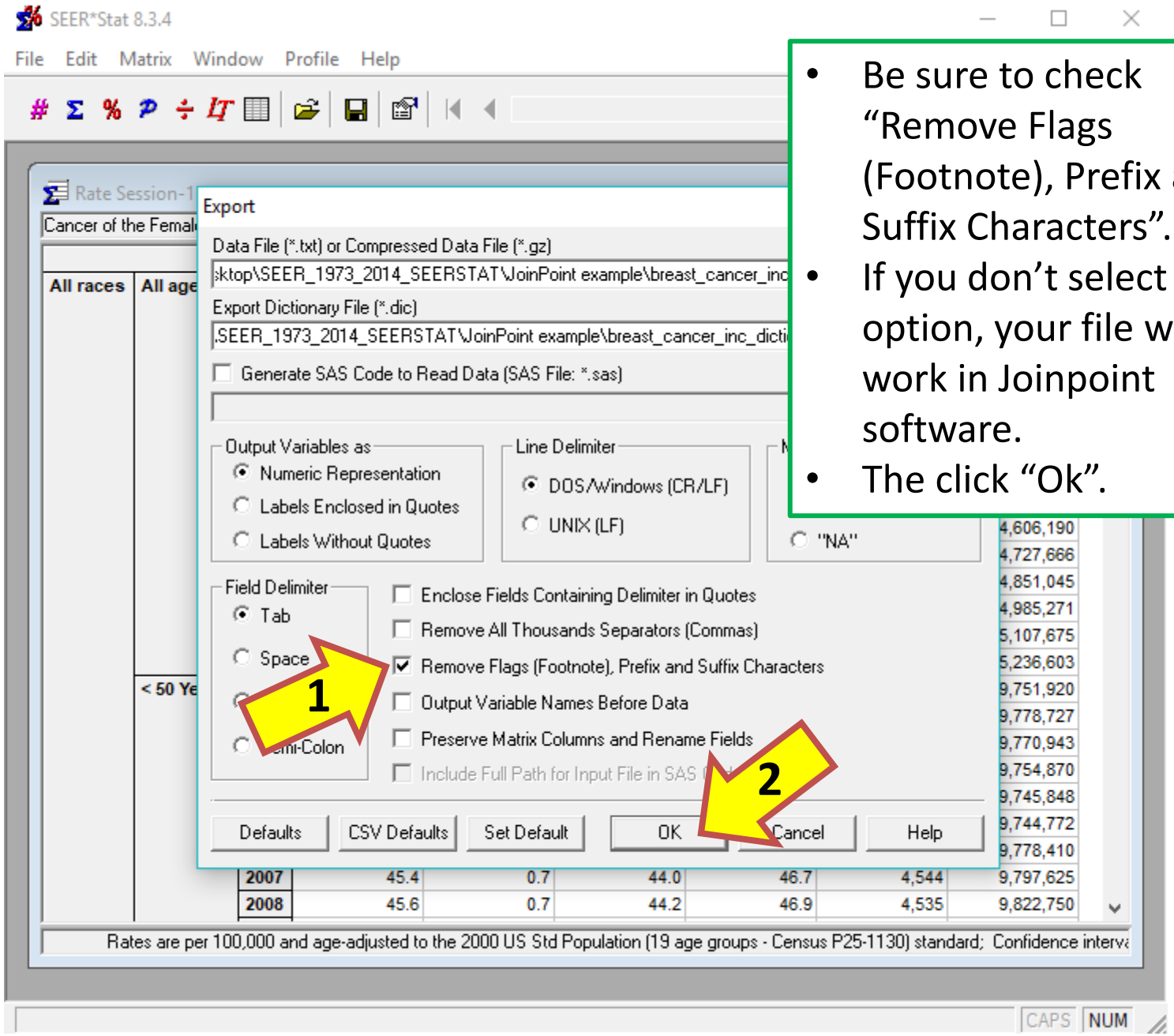




Next, you'll save the dictionary file. Click the second "Browse…" button.

- Save the dictionary file in the same folder as the data file you just saved.
- Incorporate the word “dictionary” in the descriptive name and then click “Save”.

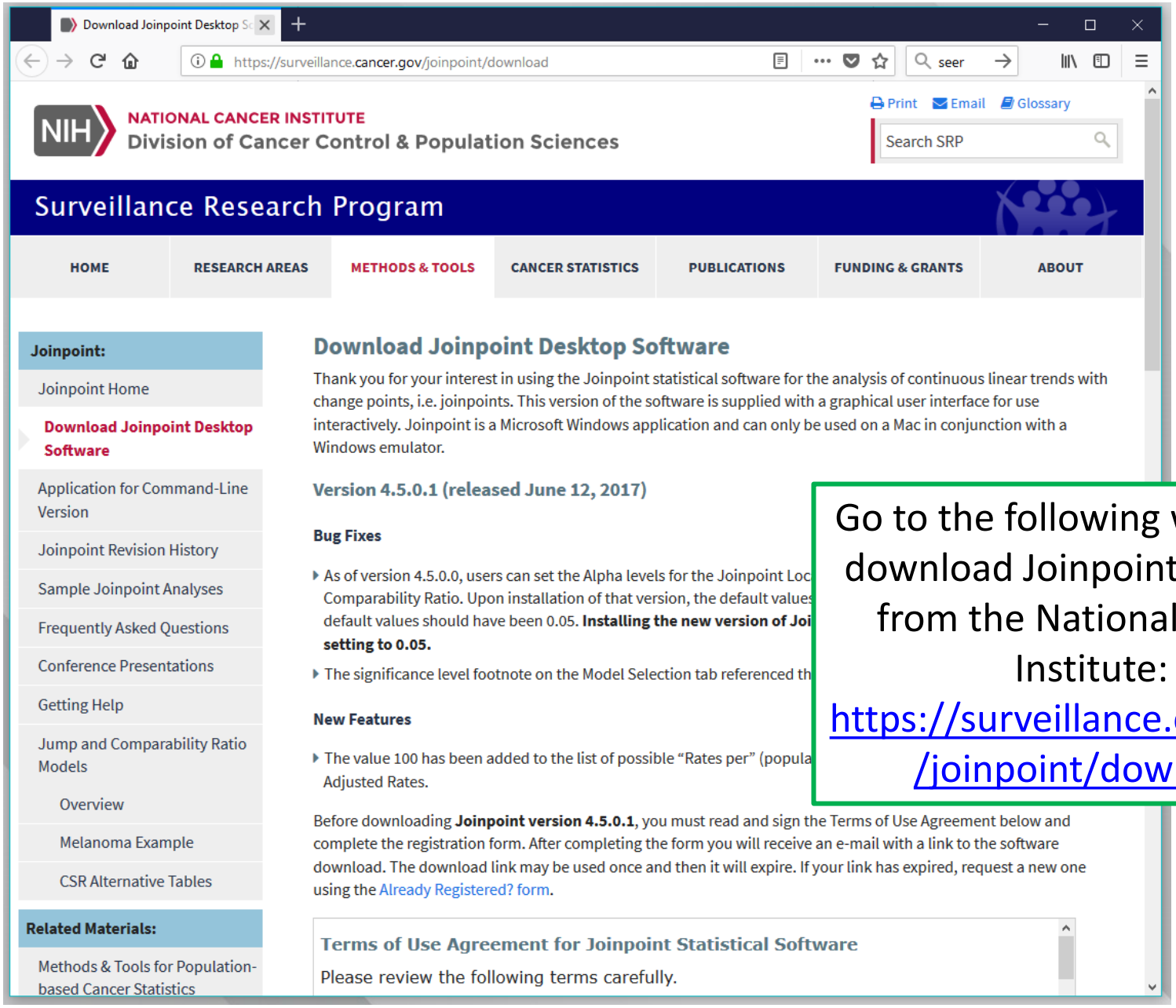




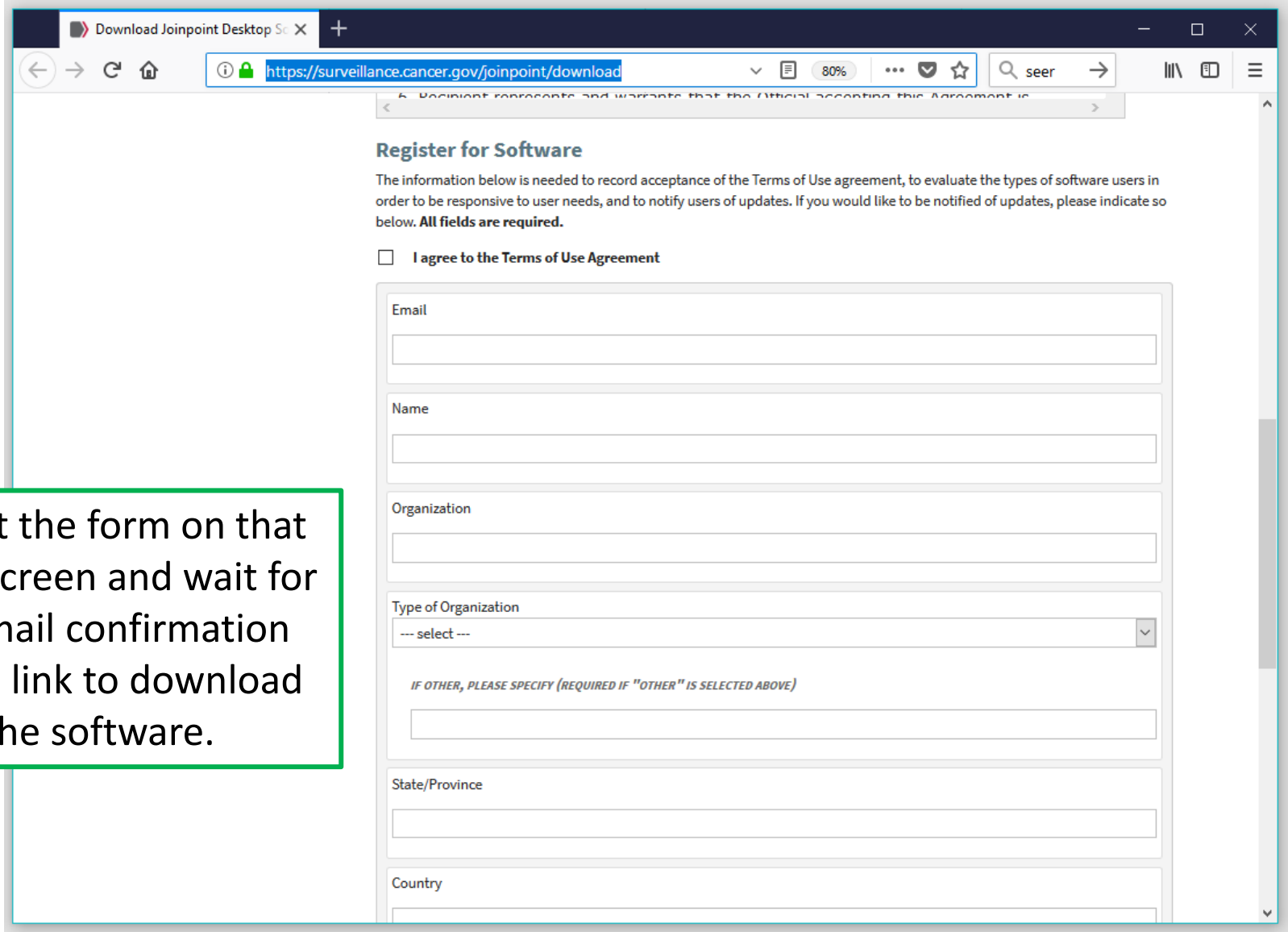
- Be sure to check “Remove Flags (Footnote), Prefix and Suffix Characters”.
- If you don’t select this option, your file won’t work in Joinpoint software.
- The click “Ok”.

Downloading Joinpoint Software

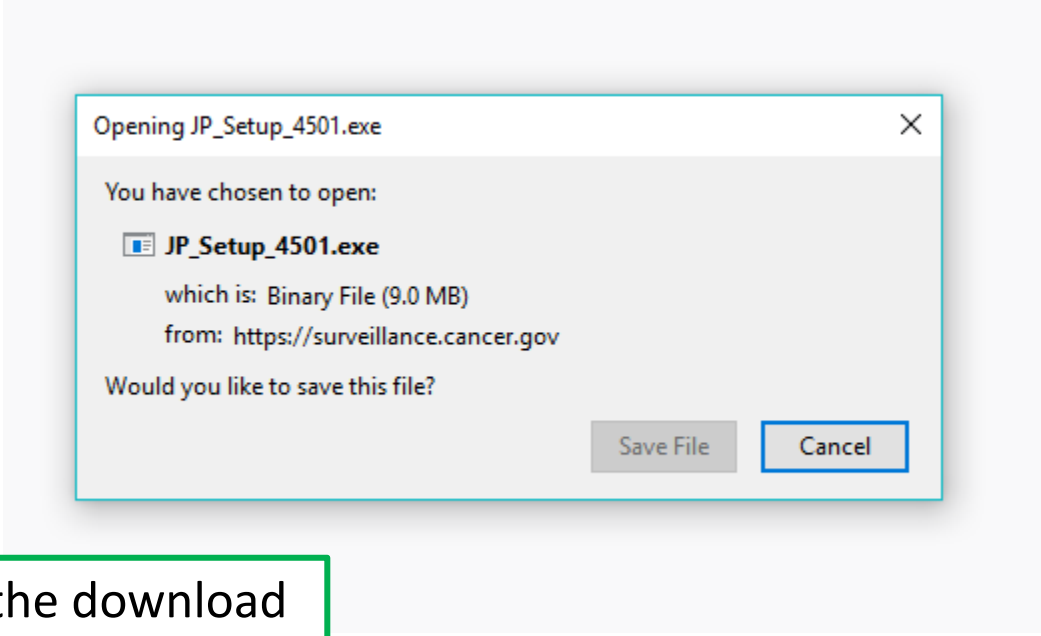
University of Nebraska Medical Center



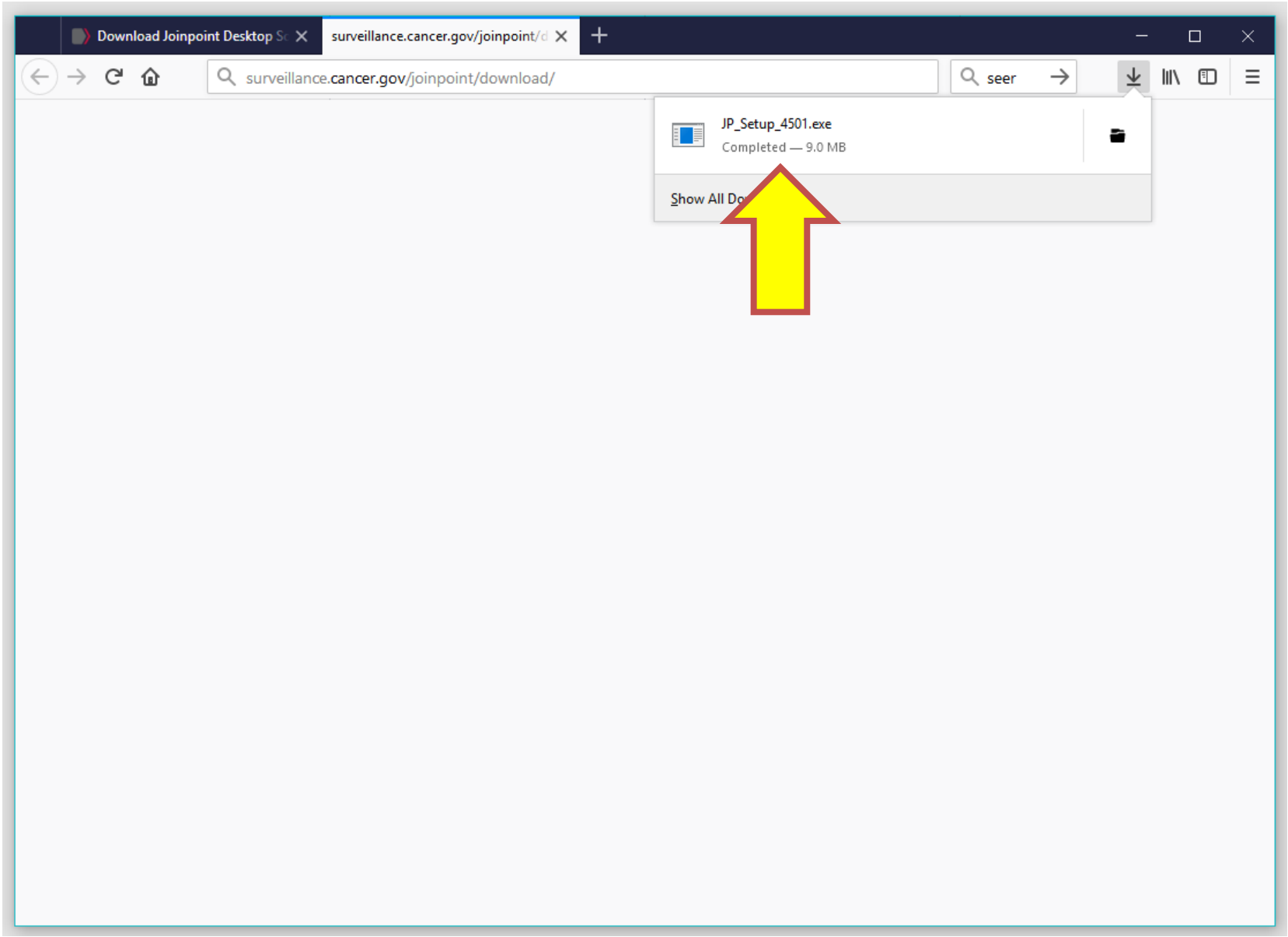
Go to the following website to download Joinpoint software from the National Cancer Institute: https://surveillance.cancer.gov/joinpoint/download

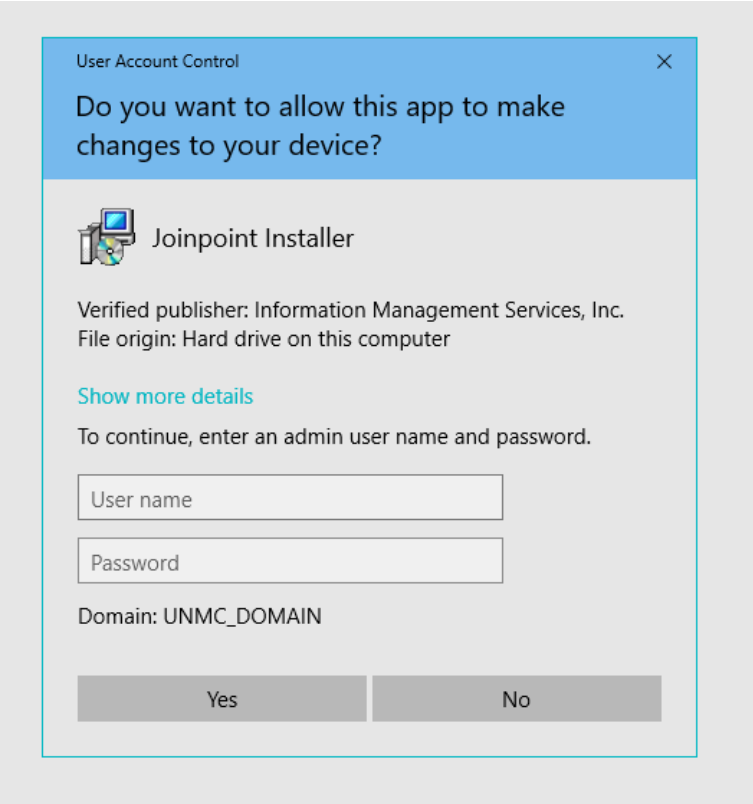


Fill out the form on that same screen and wait for an email confirmation with a link to download the software.



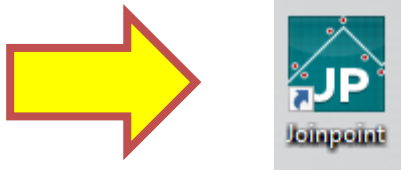
Click on the download link in your email.


Download Joinpoint Desktop
surveillance.cancer.gov/joinpoint/
surveillance.cancer.gov/joinpoint/download/
seer
JP_Setup_4501.exe
Completed — 9.0 MB
Show All



- After clicking on the executable file, follow the prompts to download the software.
- You may need to contact your IT department to get the administrative rights to download executable files.

When the download has completed, you should have a Joinpoint shortcut on your desktop.



Using Joinpoint software to detect significant changes in temporal trends

Example: Breast Cancer Age-Adjusted Incidence Rates

University of Nebraska Medical Center

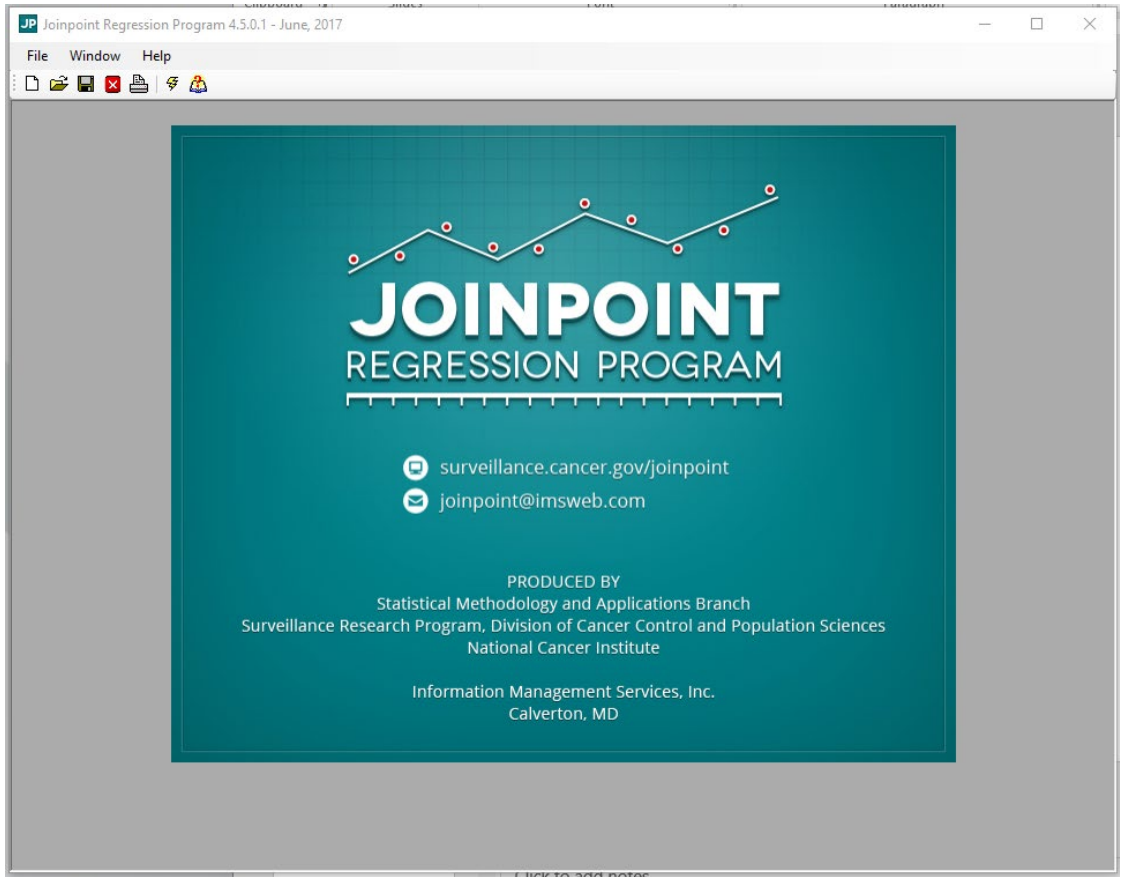
Using Joinpoint

- The primary objective of this tutorial is showing you how to import SEER*Stat results into Joinpoint and then run a session.

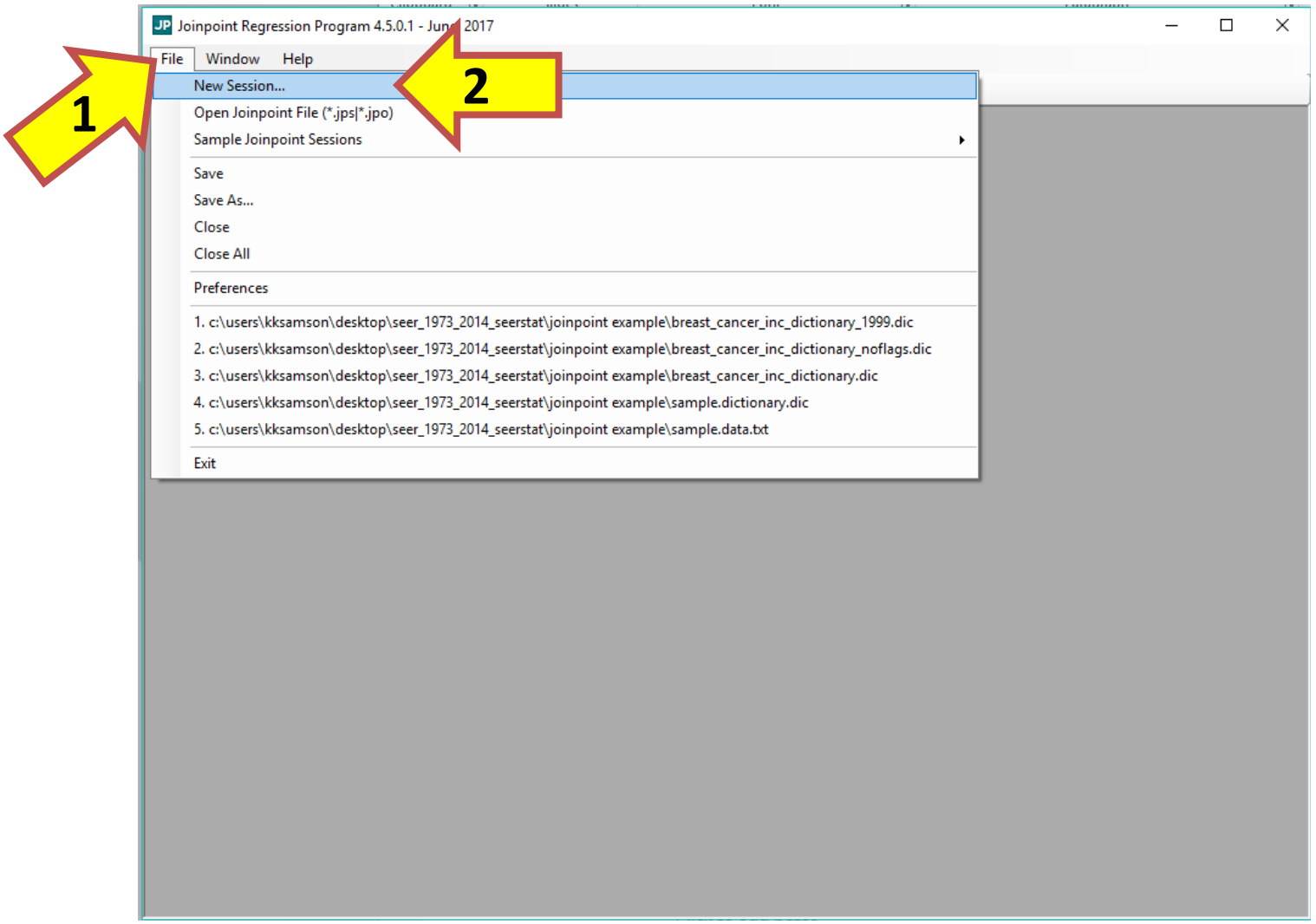
- There are many nuances and options within this software, and we encourage you to check out more information about Joinpoint here:
  - https://surveillance.cancer.gov/joinpoint/example.html
  - https://surveillance.cancer.gov/joinpoint/faq/

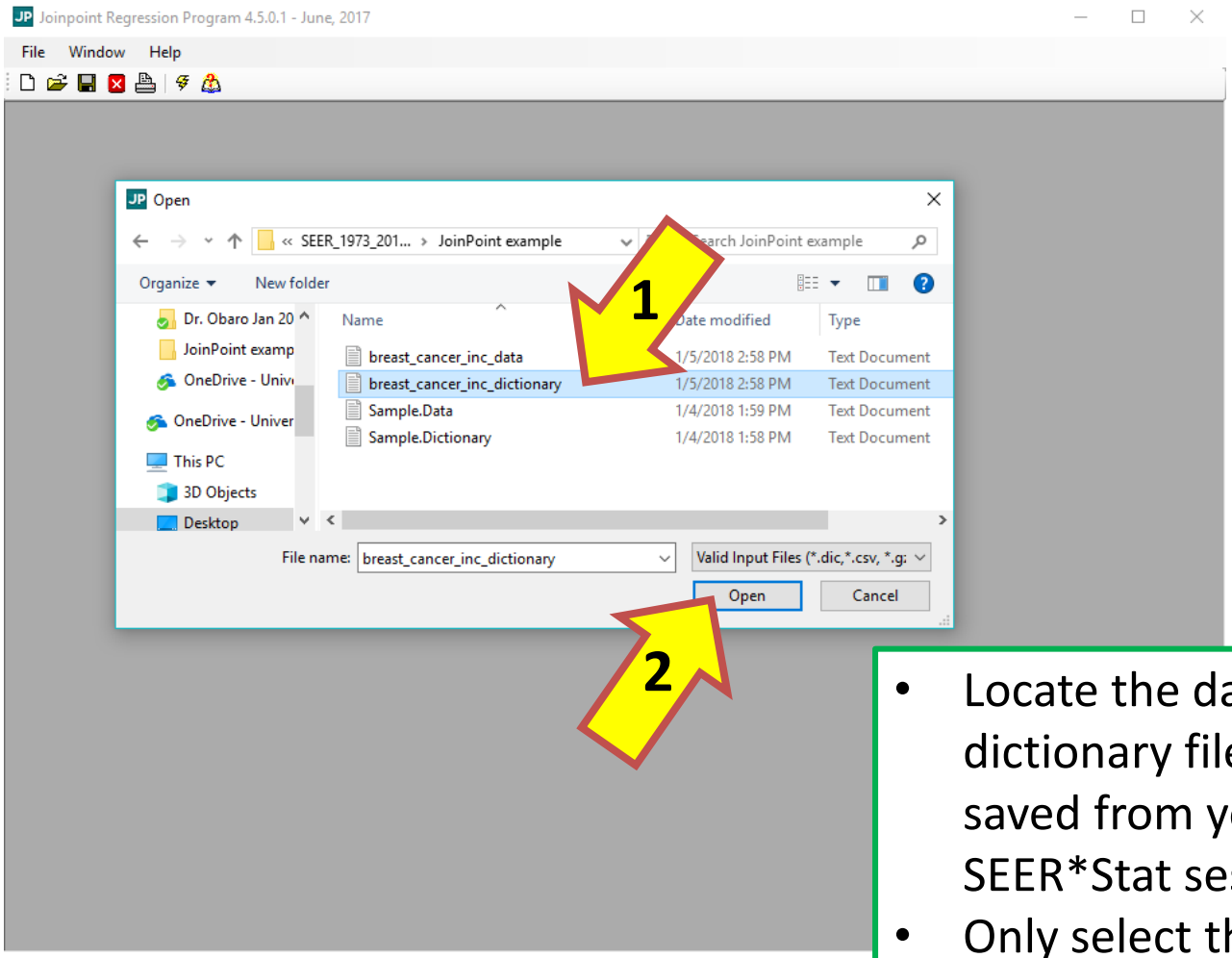
- For this example, we will be importing the breast cancer age-standardized rates from the previous tutorial sections.

Double click the Joinpoint icon on your desktop to open the software.

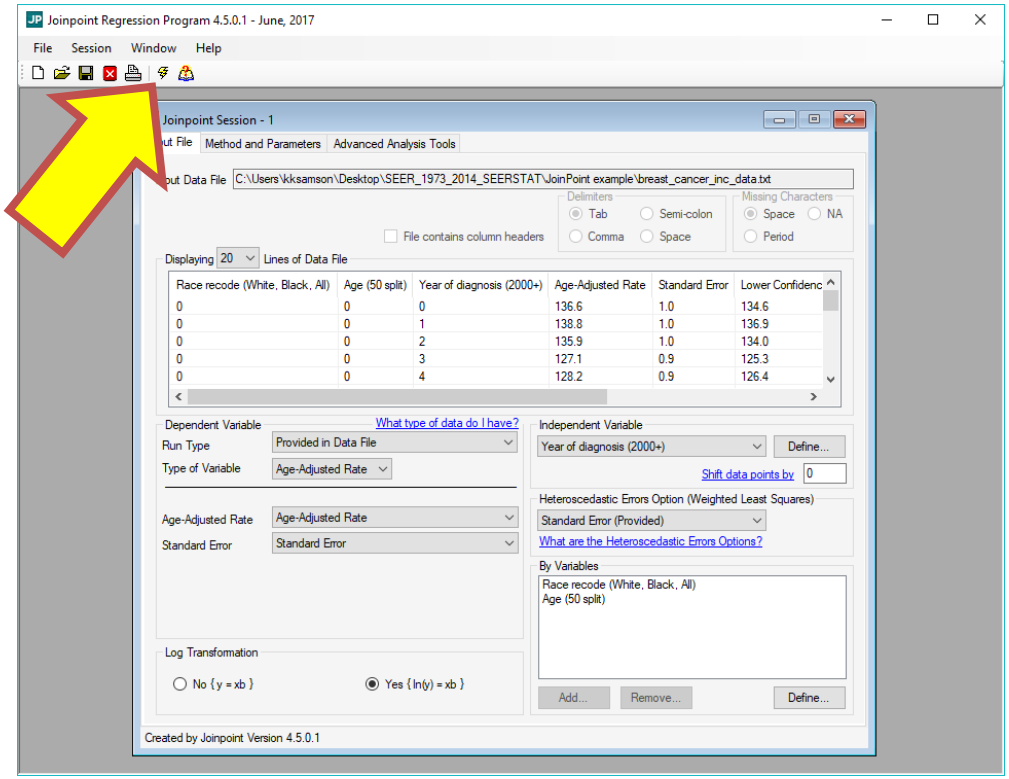


To start a new session, go to File and then “New Session…”





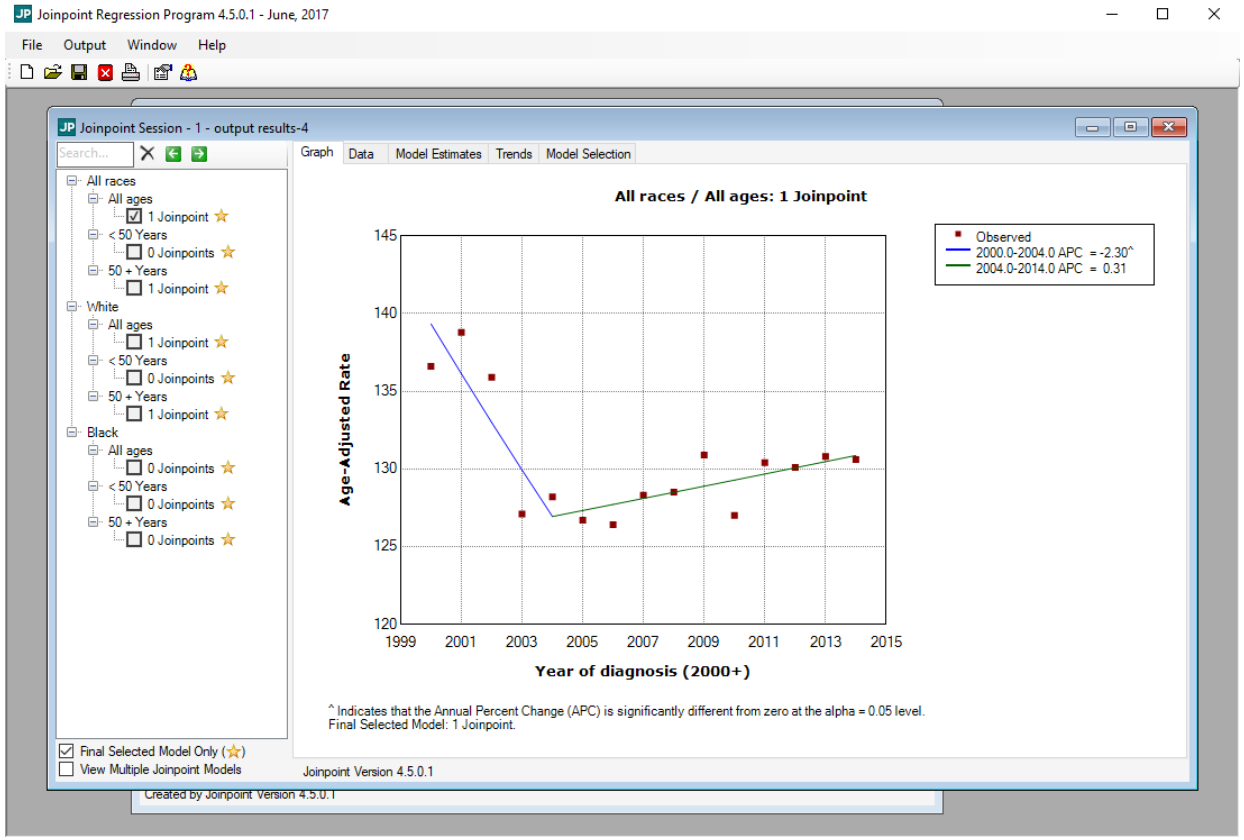
- Locate the data and dictionary files you saved from your SEER*Stat session.
- Only select the **dictionary file** and then click “Open”.



For the purposes of this exercise, we will use the default settings. Note that because we imported the data and dictionary from SEER*Stat, the Joinpoint software automatically detected Run Type, Type of Variable, Age-Adjusted Rate, Standard Error, and Independent/By Variables. If you imported your own data, you would have to manually specify these.

Click the lightning bolt button to run the session.

The software generated a Joinpoint regression for each combination of the By variables. Here, we are seeing the trend for all races and ages. You may click through the other combinations on the side panel to see other Joinpoint regressions. For additional resources on what Joinpoint software does, please see this website: https://surveillance.cancer.gov/help/joinpoint/tech-help/frequently-asked-questions



University of Nebraska Medical Center

BREAKTHROUGHS FOR LIFE.®