GREAT PLAINS
IDeA | Clinical and
Translational Research

Tutorial for working with BRFSS online data tools and downloading data to run analysis in statistical software

April 2020

INTRODUCTION

- This tutorial will cover topics that relate to use of data collected through Behavioral Risk Factor Surveillance system.

- BRFFS is the system of health-related telephone surveys that collect state data about non-institutionalized, adult U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services.

- The states use a standardized core questionnaire, optional modules, and state-added questions administering the surveys continuously throughout the year.

GREAT PLAINS IDeA | Clinical and Translational Research

INTRODUCTION

- The survey is conducted using Random Digit Dialing (RDD) techniques on both landlines and cell phones using two samples.

- Disproportionate stratified sampling (DSS) is used for landline sample i.e. dividing telephone numbers into two density-based strata. The landline sampling ratio of high to medium density is 1:1.5.

- The cellular telephone sample is randomly generated from a sampling frame of confirmed cellular area code and prefix combinations

GREAT PLAINS
IDeA | Clinical and Translational Research

INTRODUCTION

- For landline interviews, individual respondents are randomly selected from all adults, living in a household.
- Cellular interviews are treated as one-person households.
- BRFSS data and comprehensive documentation are freely available to download. There is a detailed section in this tutorial that will walk you through this process step-by-step.
- Detailed methodology of the survey has been explained in the data user guide.

GREAT PLAINS IDeA | Clinical and Translational Research

TUTORIAL CONTENTS

GREAT PLAINS IDeA | Clinical and Translational Research

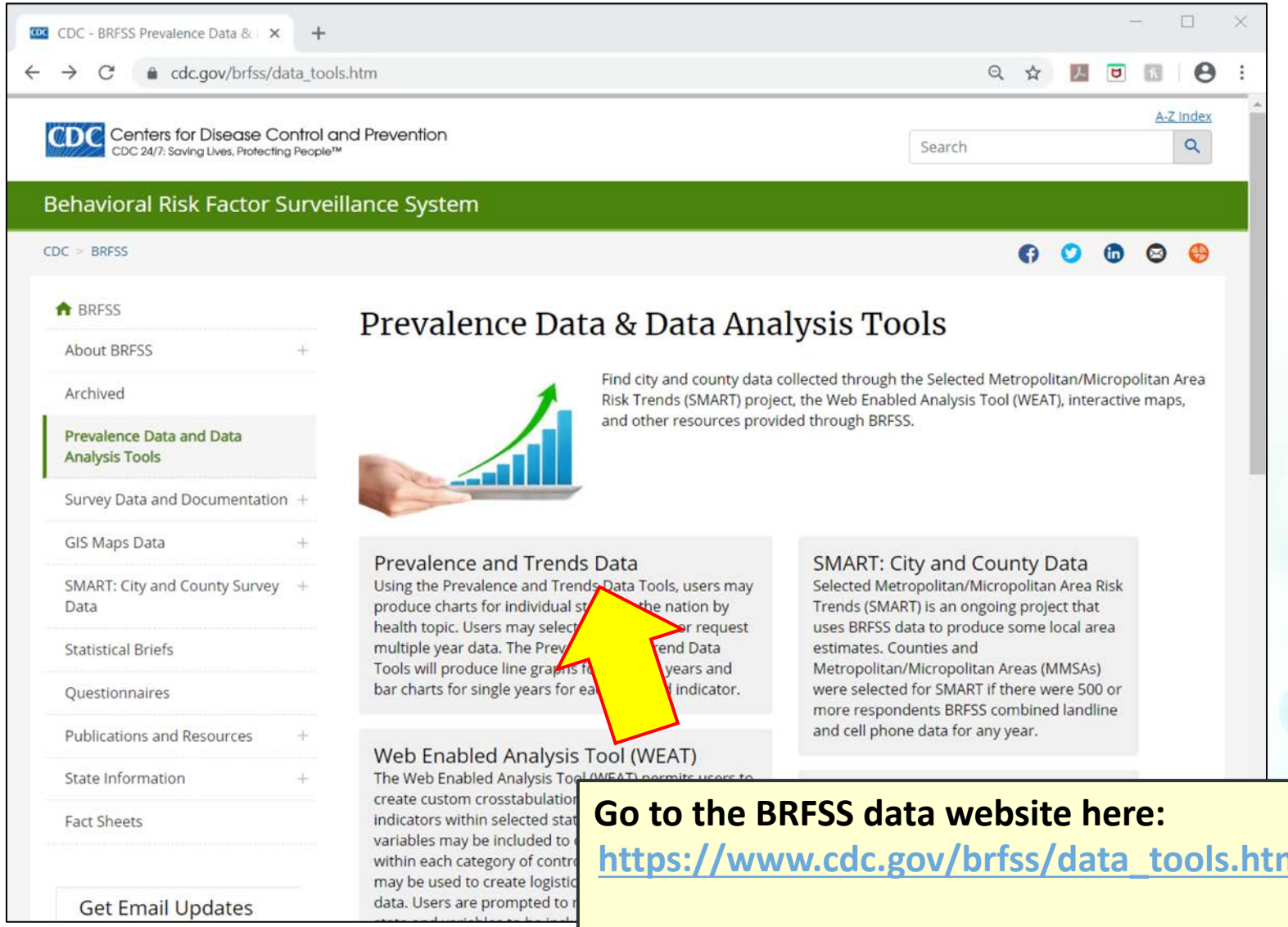
Prevalence And Trends Data Tool

GREAT PLAINS
IDeA | Clinical and Translational Research

PREVALENCE AND TRENDS DATA TOOL

- Quick tool to explore prevalence using a wide range of variables
- Produces charts and maps for individual states or the nation by health topics.
- Particularly useful to produce trends over multiple years.
- Trends are produced as line graphs and single year estimates are produced as maps and bar charts.

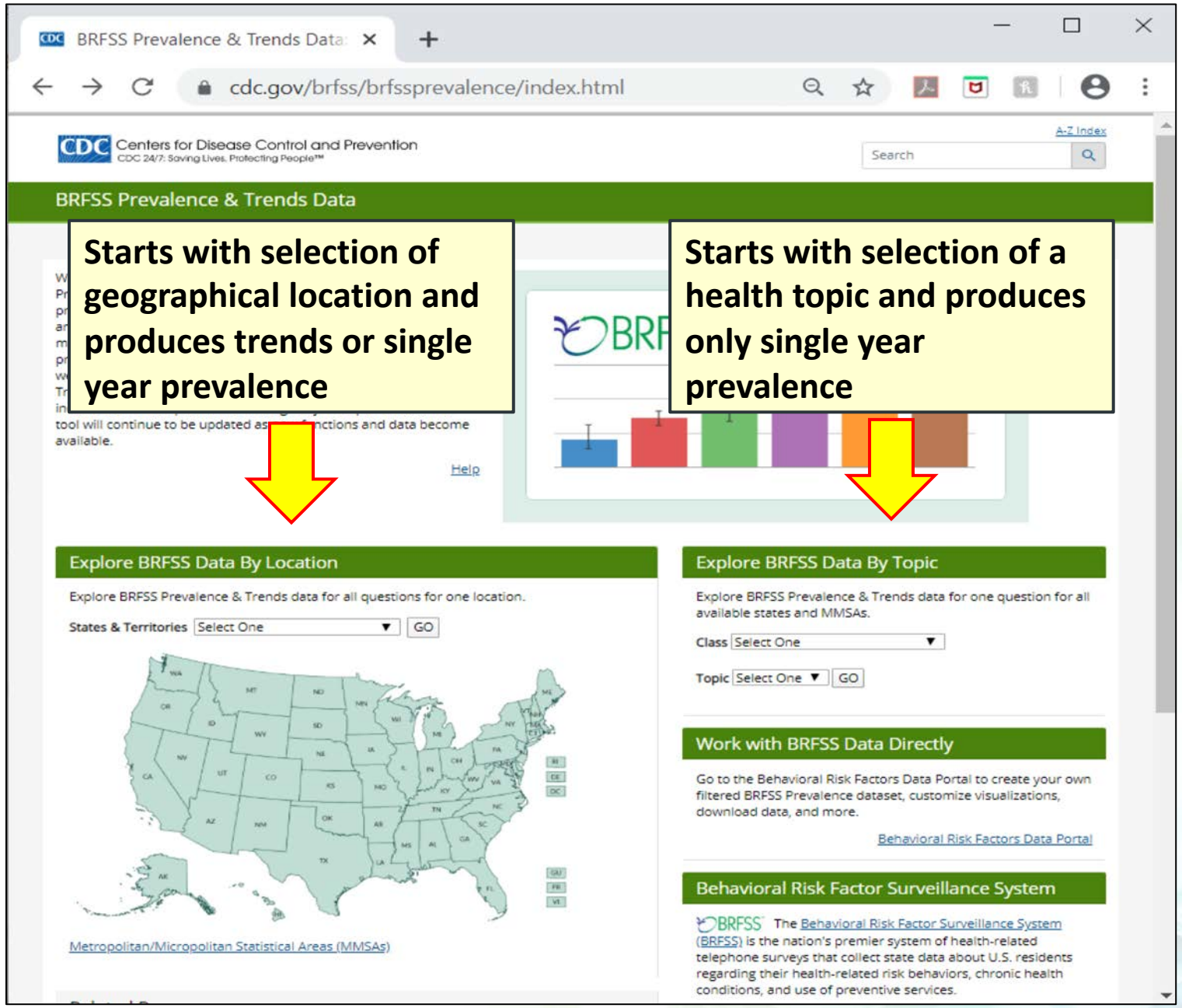
GREAT PLAINS IDeA | Clinical and Translational Research



Go to the BRFSS data website here:
https://www.cdc.gov/brfss/data_tools.htm

Click "Prevalence and Trends Data"

GREAT PLAINS IDeA | Clinical and Translational Research

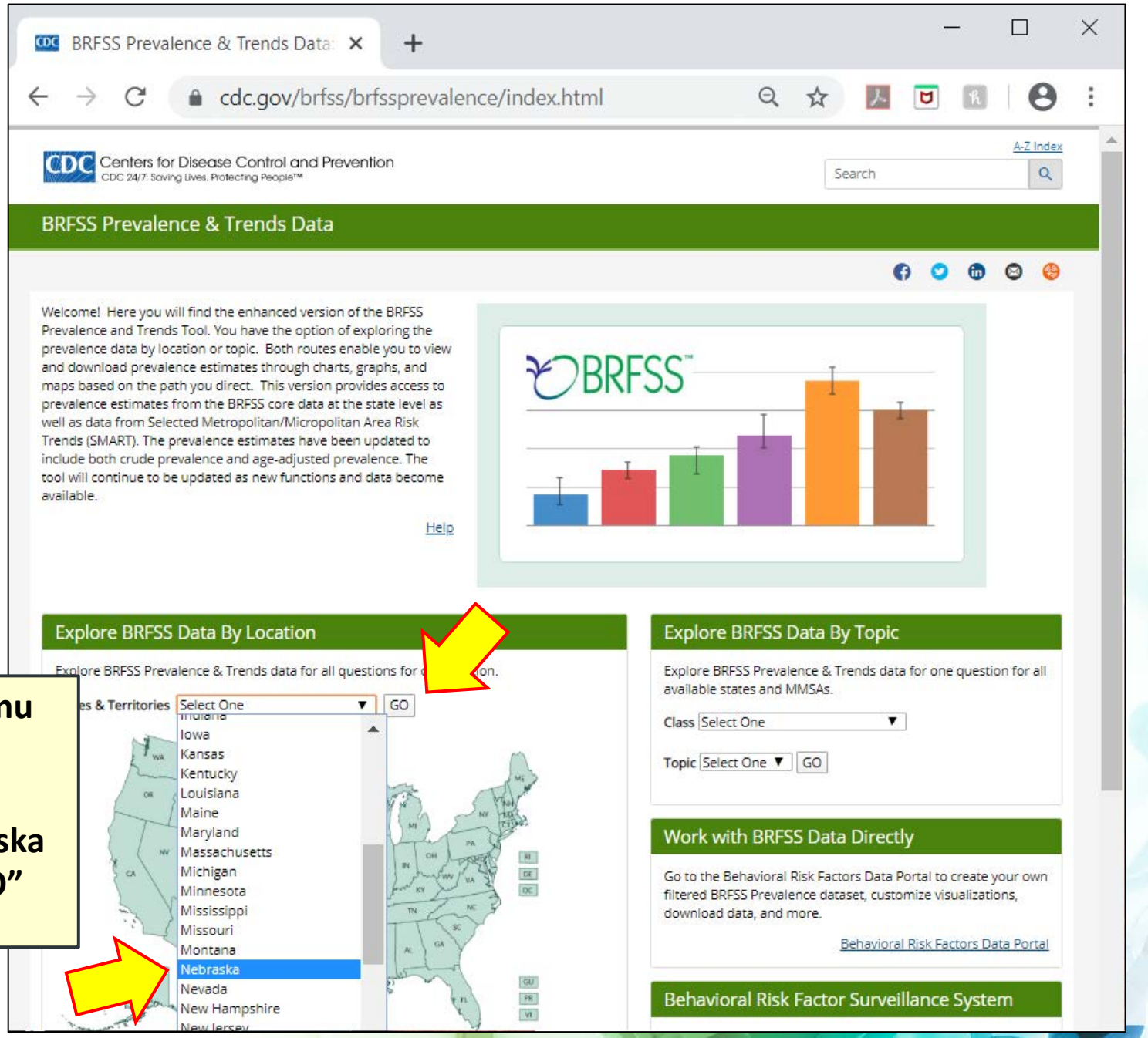


GREAT PLAINS
IDeA | Clinical and Translational Research

PREVALENCE OF BINGE DRINKING

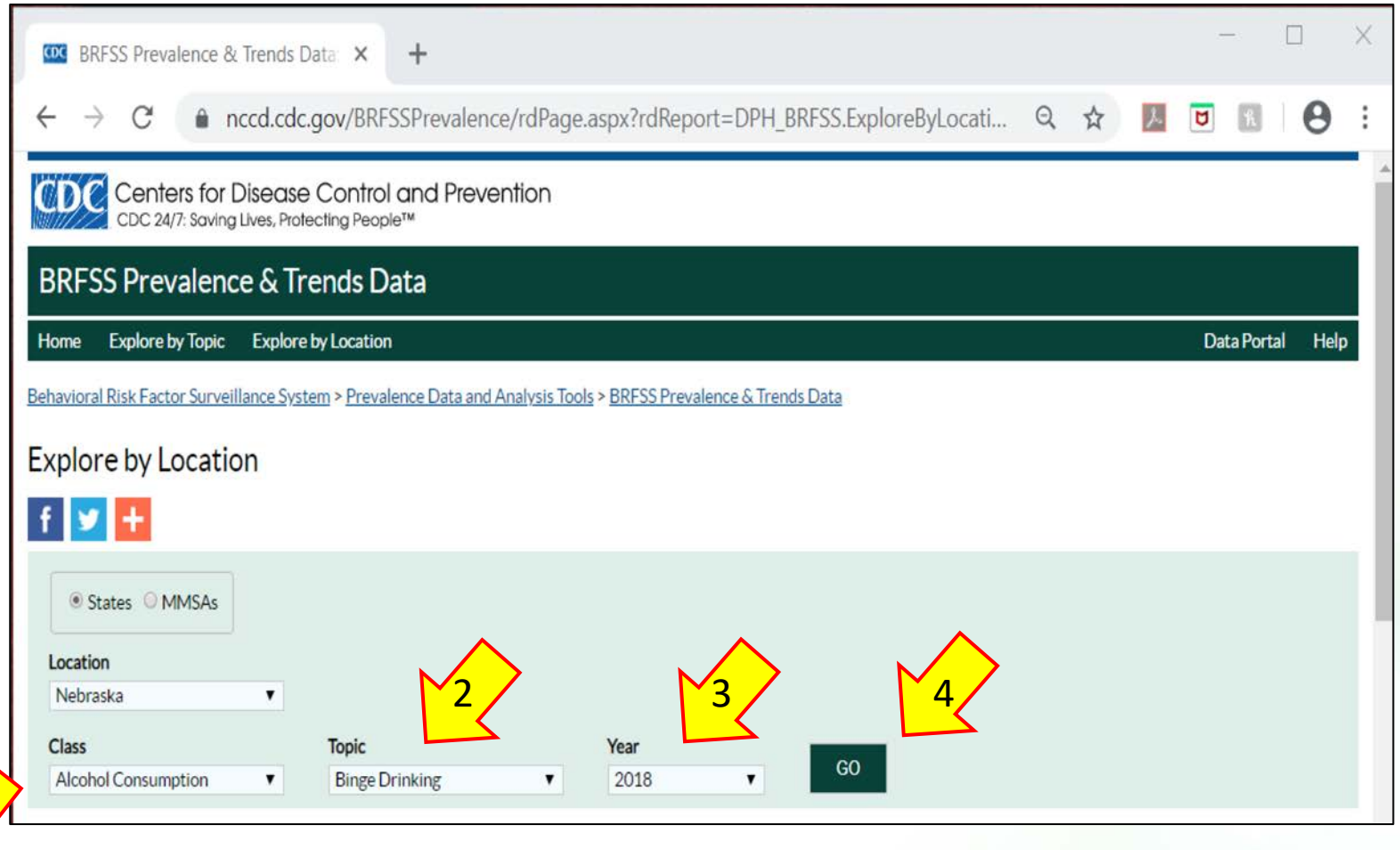
- This section will instruct you on how to get prevalence for a specific topic
- For this example, let us see how to get:
  - Age-adjusted prevalence of binge drinking in Nebraska in 2018
  - Trend of age-adjusted prevalence of binge drinking in Nebraska for all available years (2011-2018)
  - Prevalence of binge drinking in Nebraska by gender in 2018 and trend over years (only gives crude prevalence)

GREAT PLAINS IDeA | Clinical and Translational Research



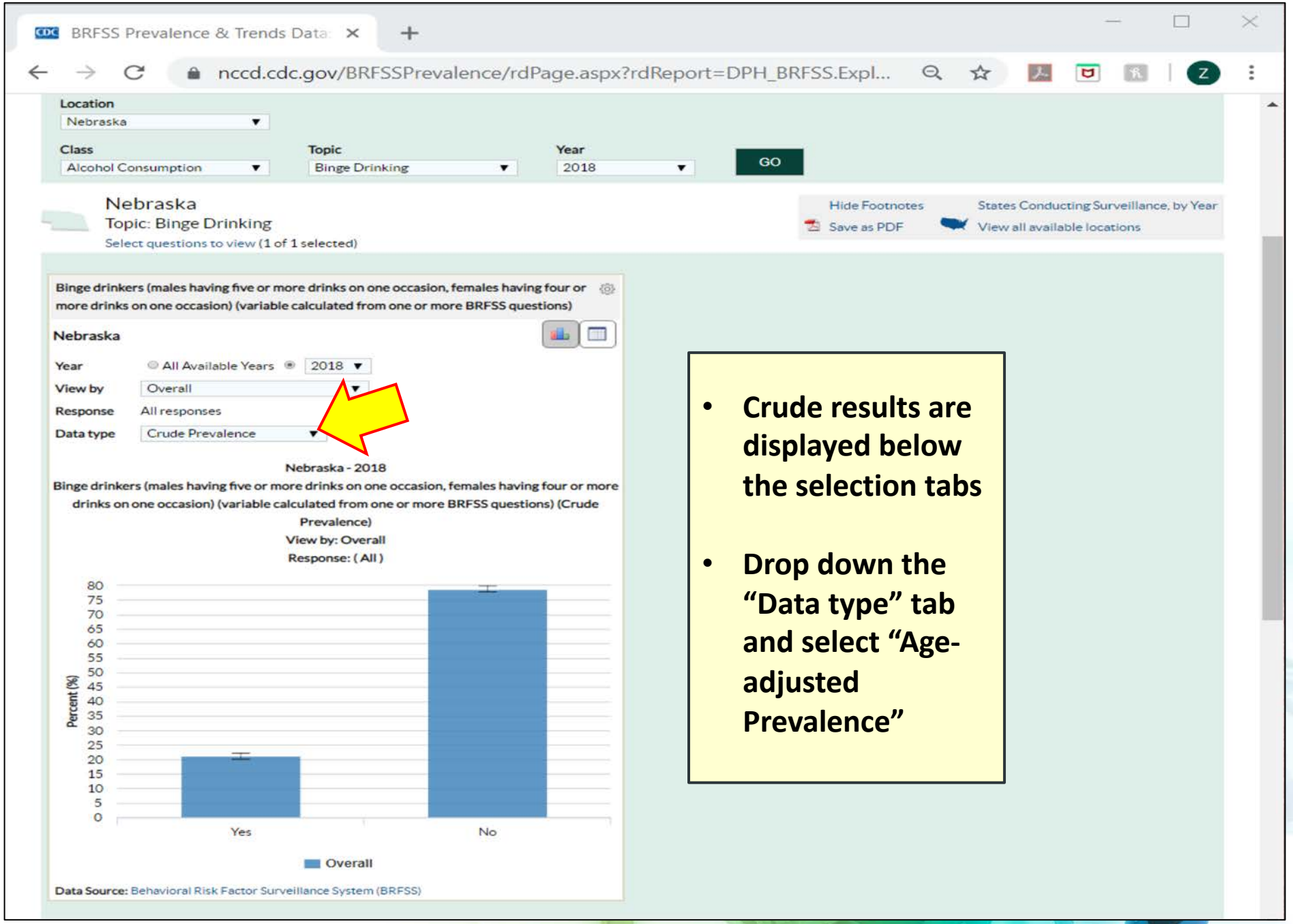
- **Drop the menu down**
- **Select Nebraska and click “GO”**

GREAT PLAINS
IDeA | Clinical and Translational Research



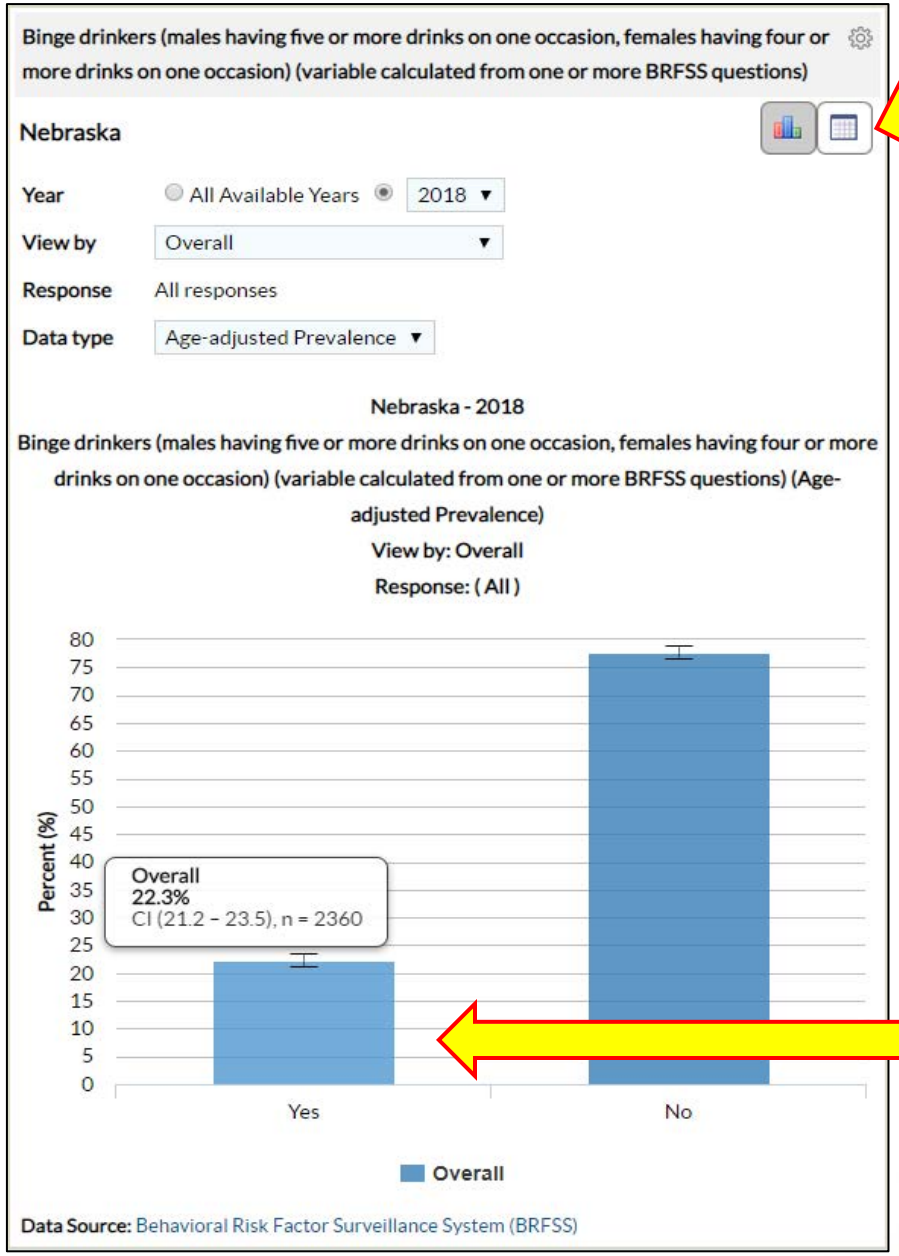
- **Under “class” select “Alcohol Consumption”**
- **Under “topic” select “Binge Drinking”**
- **Under “Year” select “2018”**
- **Click “go”**

GREAT PLAINS IDeA | Clinical and Translational Research

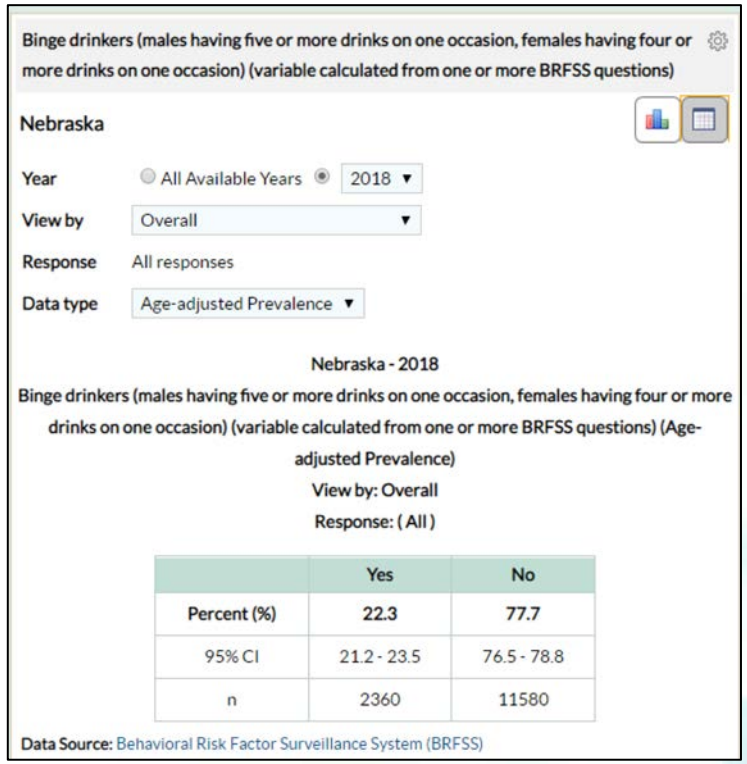


- **Crude results are displayed below the selection tabs**
- **Drop down the "Data type" tab and select "Age-adjusted Prevalence"**

GREAT PLAINS IDeA | Clinical and Translational Research

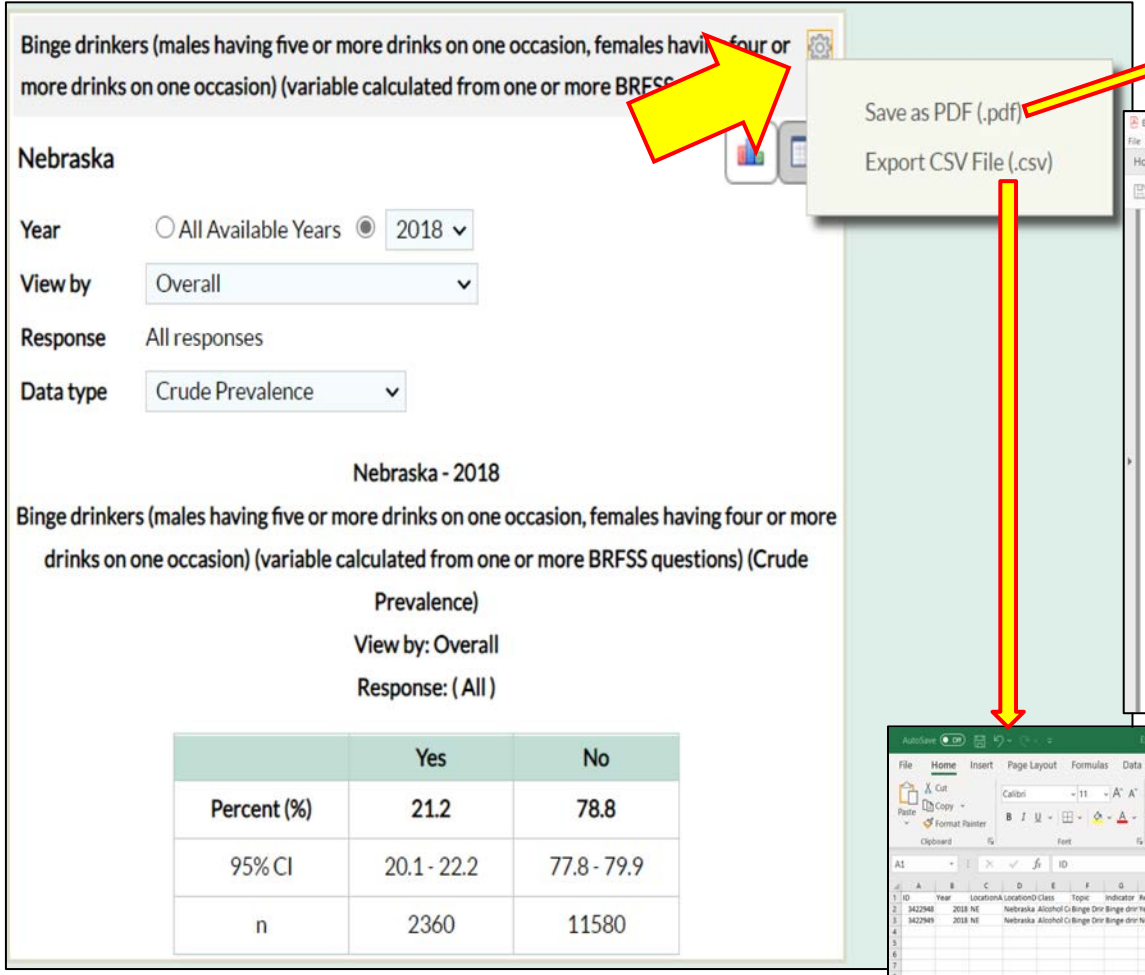


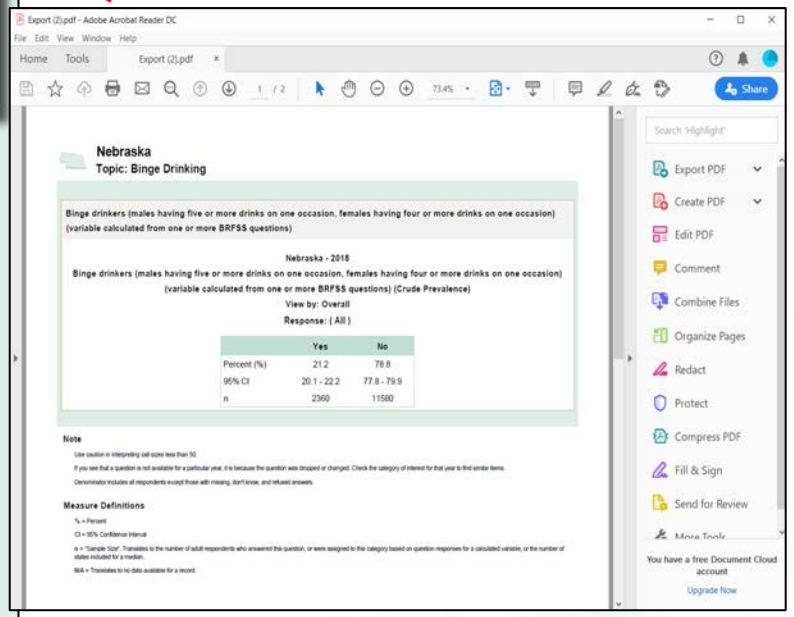
- **Click this button to get the results in tabular form below**

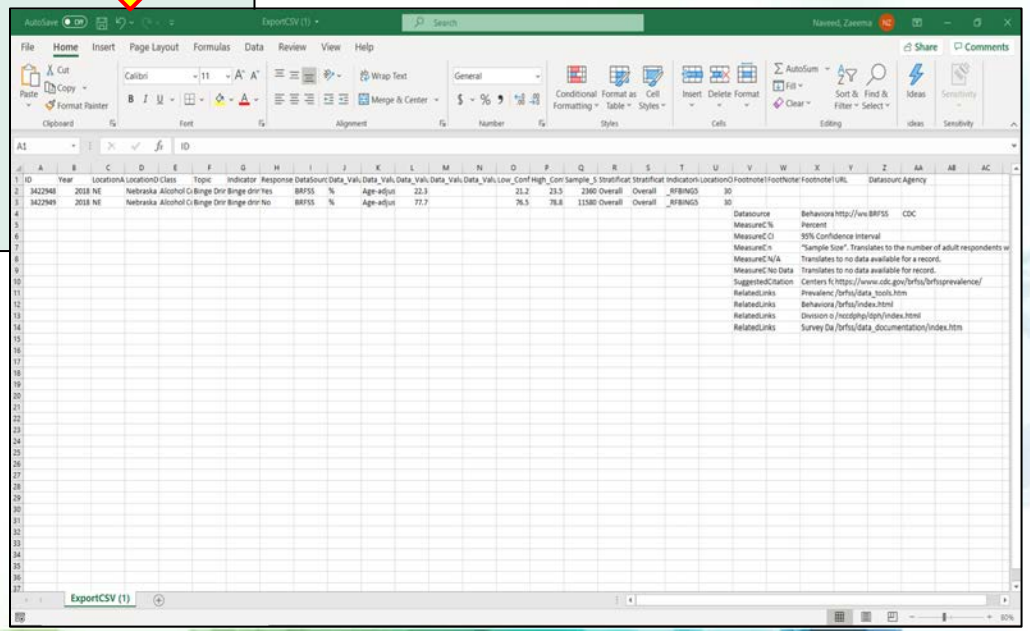


- **Hover the Mouse over the bars to see the numbers**
- **The results show that in 2018, the age-adjusted prevalence of binge drinking in Nebraska was 22.3%**

GREAT PLAINS IDeA | Clinical and Translational Research

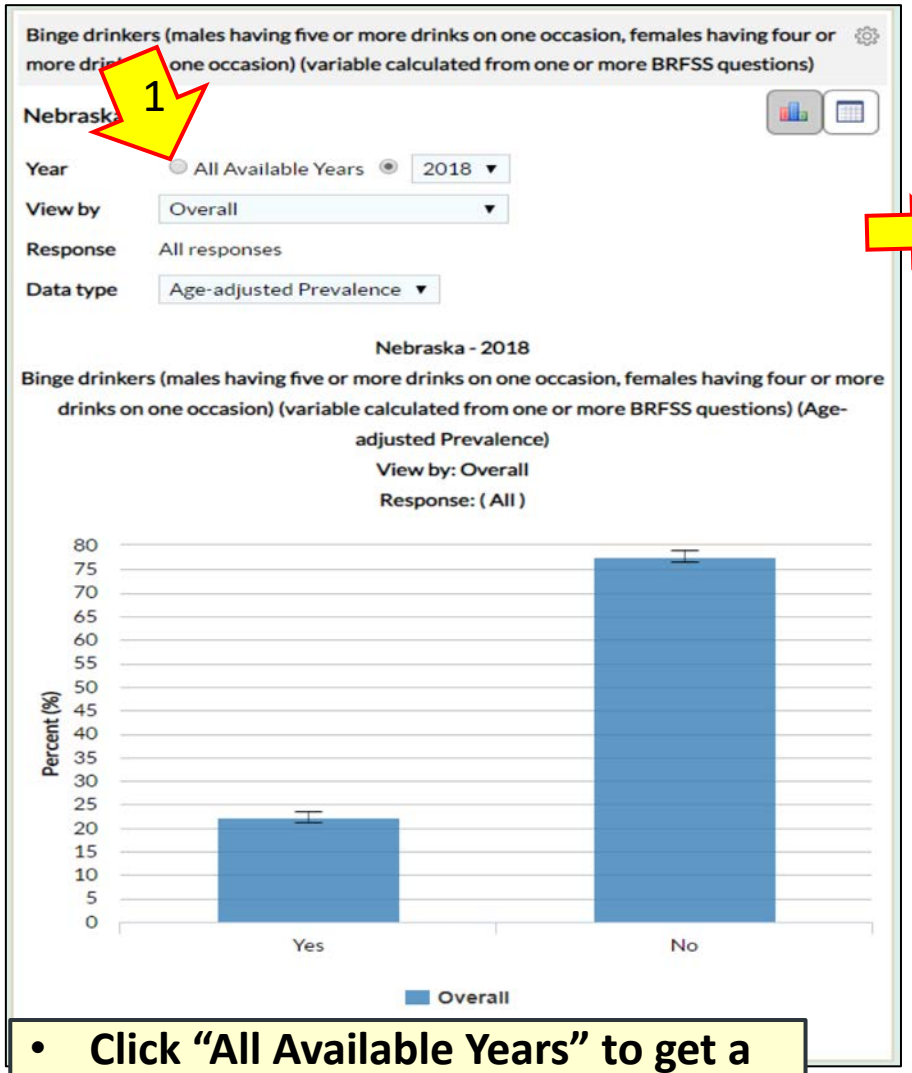


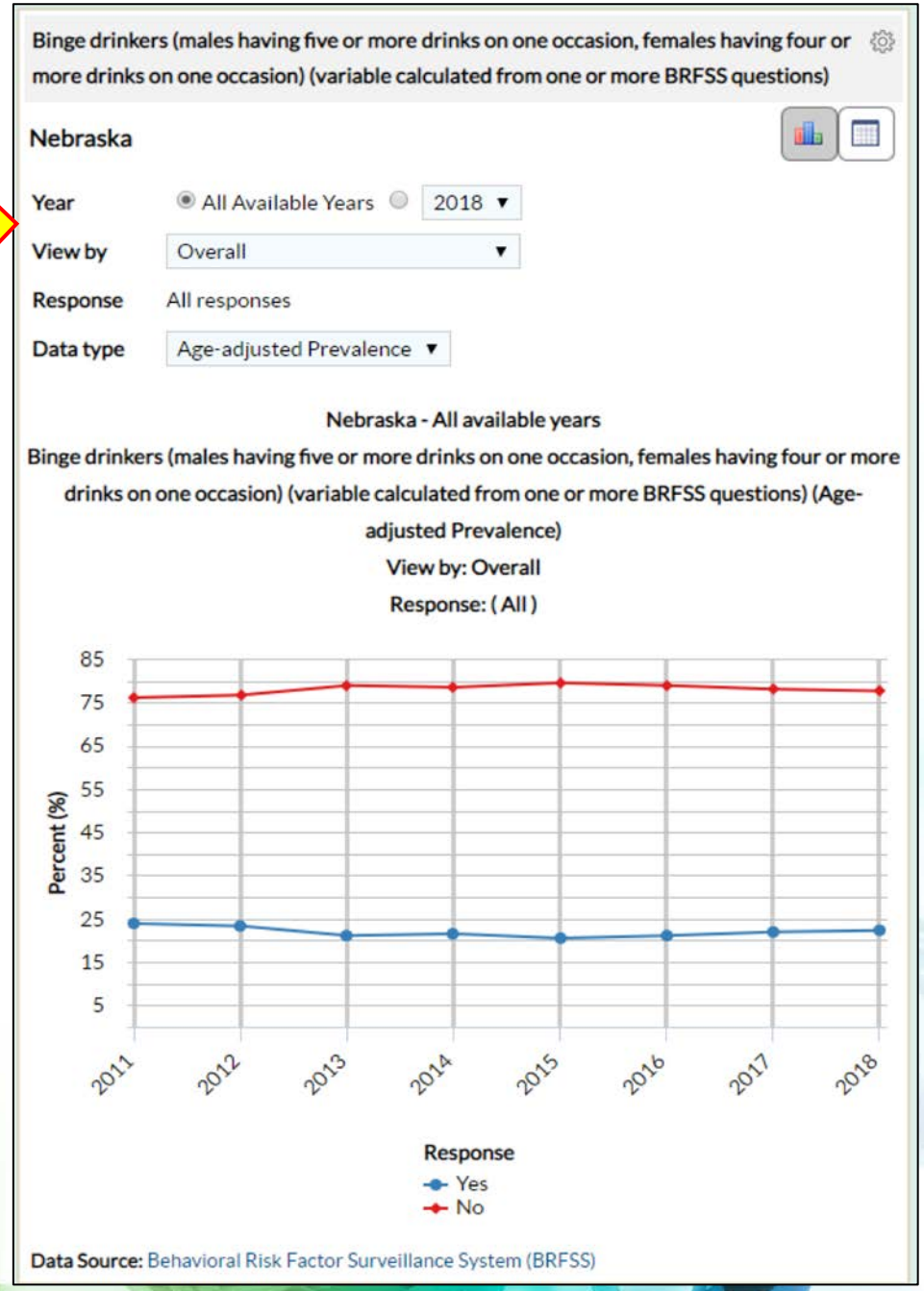




- **Click the gear icon to save the results as PDF or CSV file.**
- **The outputs are shown**

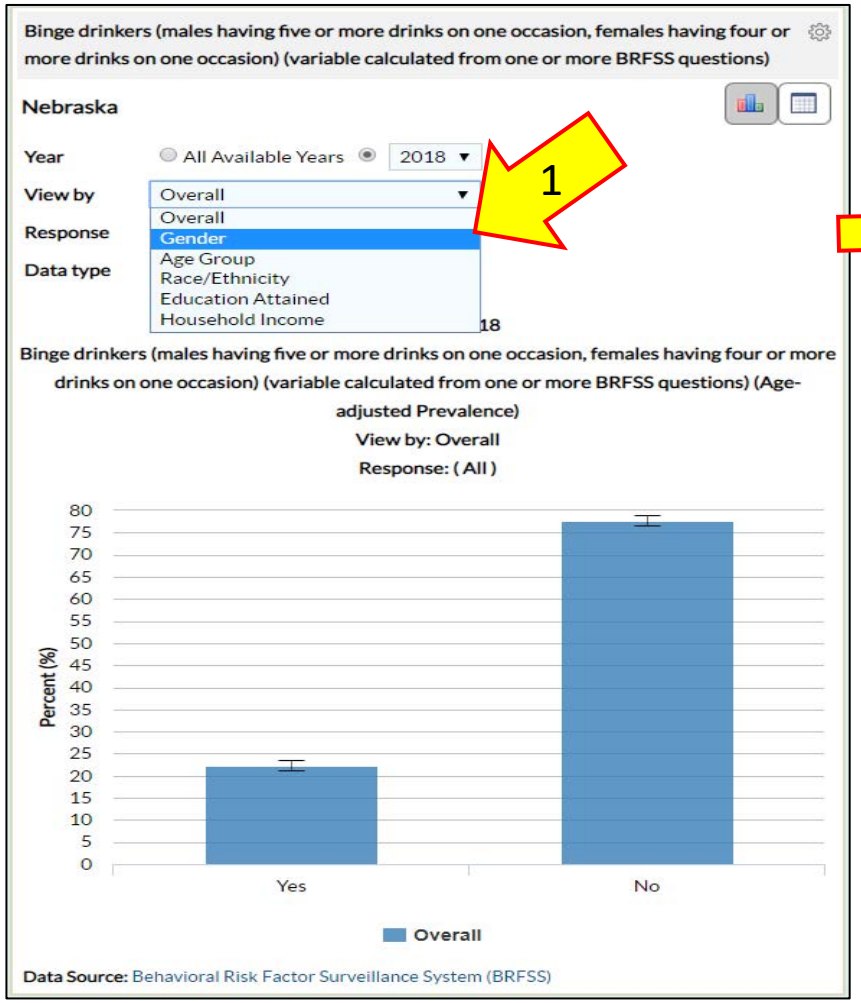
GREAT PLAINS
IDeA | Clinical and Translational Research

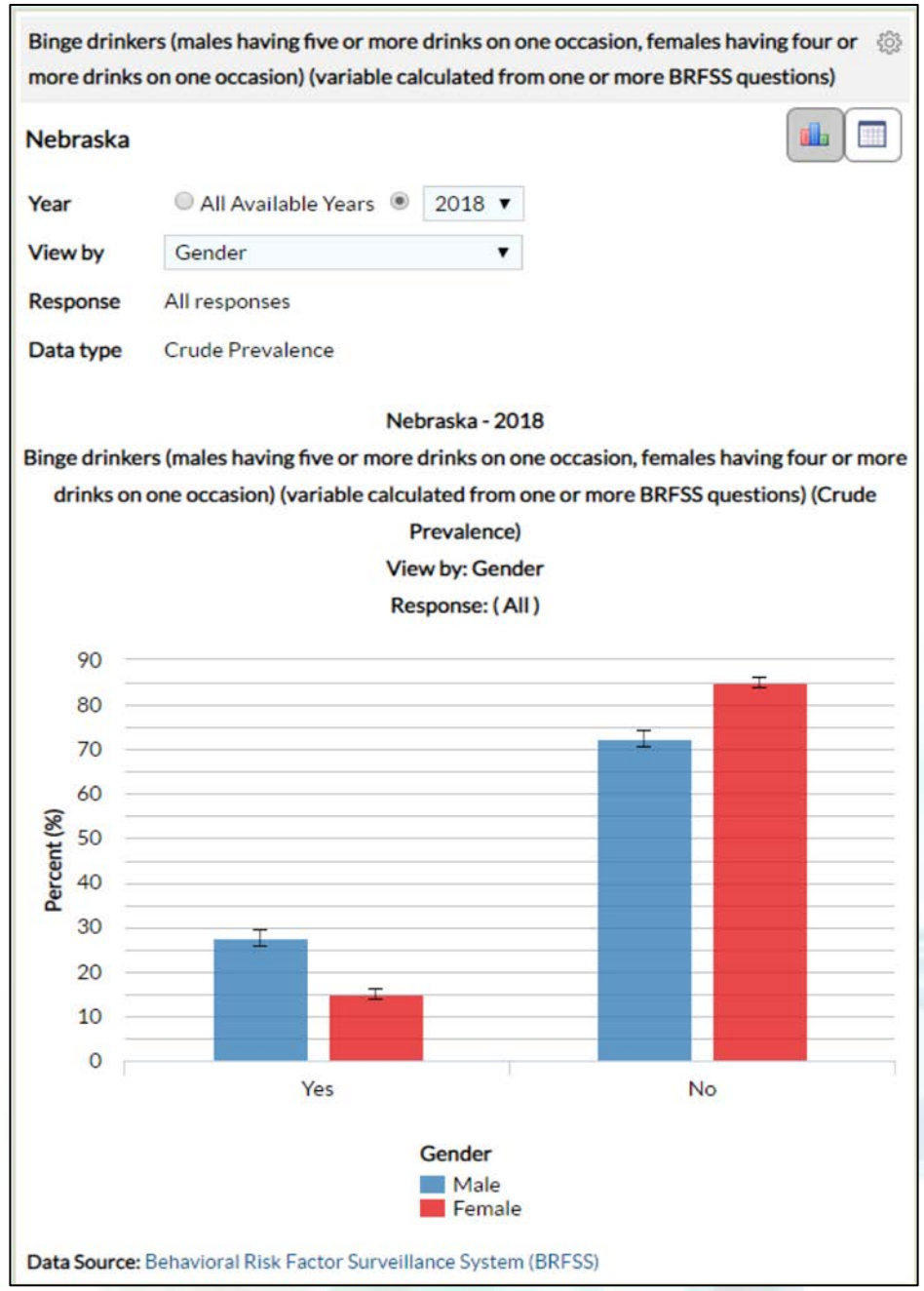




- **Click "All Available Years" to get a trend of binge drinking prevalence as shown on the right**
- **The age-adjusted prevalence of binge drinking in Nebraska has followed nearly a constant trend from 2011 to 2018**

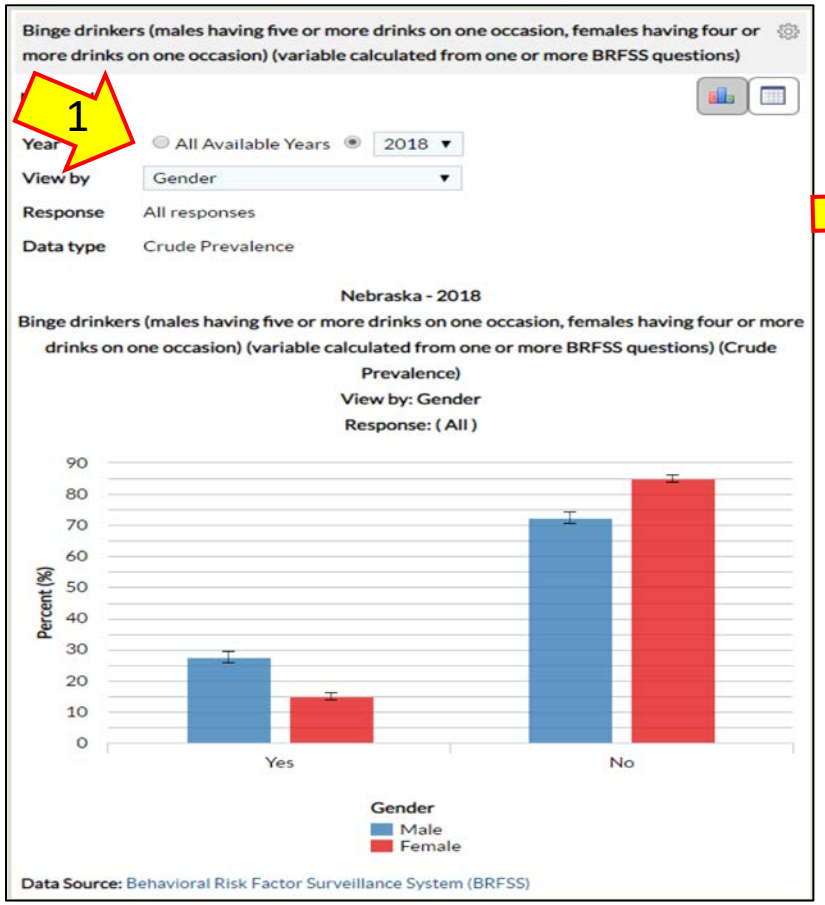
GREAT PLAINS IDeA | Clinical and Translational Research

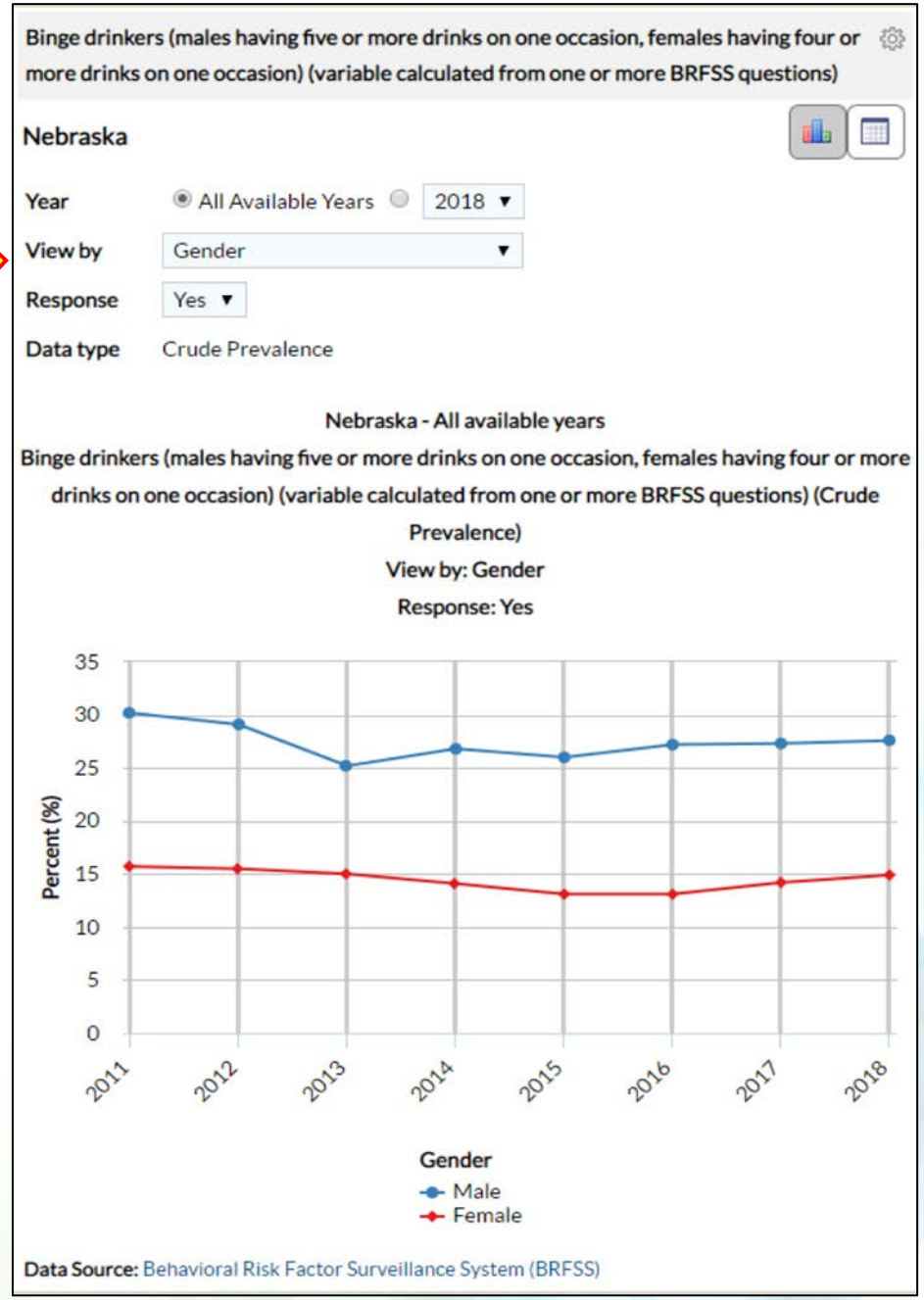




- **Drop down "view by" and select gender to get the results on the right. Only crude prevalence is displayed**
- **In 2018, males had a higher prevalence(27.6%) of binge drinking compared to females (14.9%) in Nebraska (Hover over the bars or convert into tabular form to see the numbers).**

GREAT PLAINS IDeA | Clinical and Translational Research





- **Click "All Available Years" to get trends of binge drinking by gender (crude prevalence)**
- **Among males, the prevalence of binge drinking in Nebraska shows a decrease in 2013 followed by a steady trend**
- **The prevalence of binge drinking among females has followed a steady trend from 2011 to 2018**

GREAT PLAINS IDeA | Clinical and Translational Research

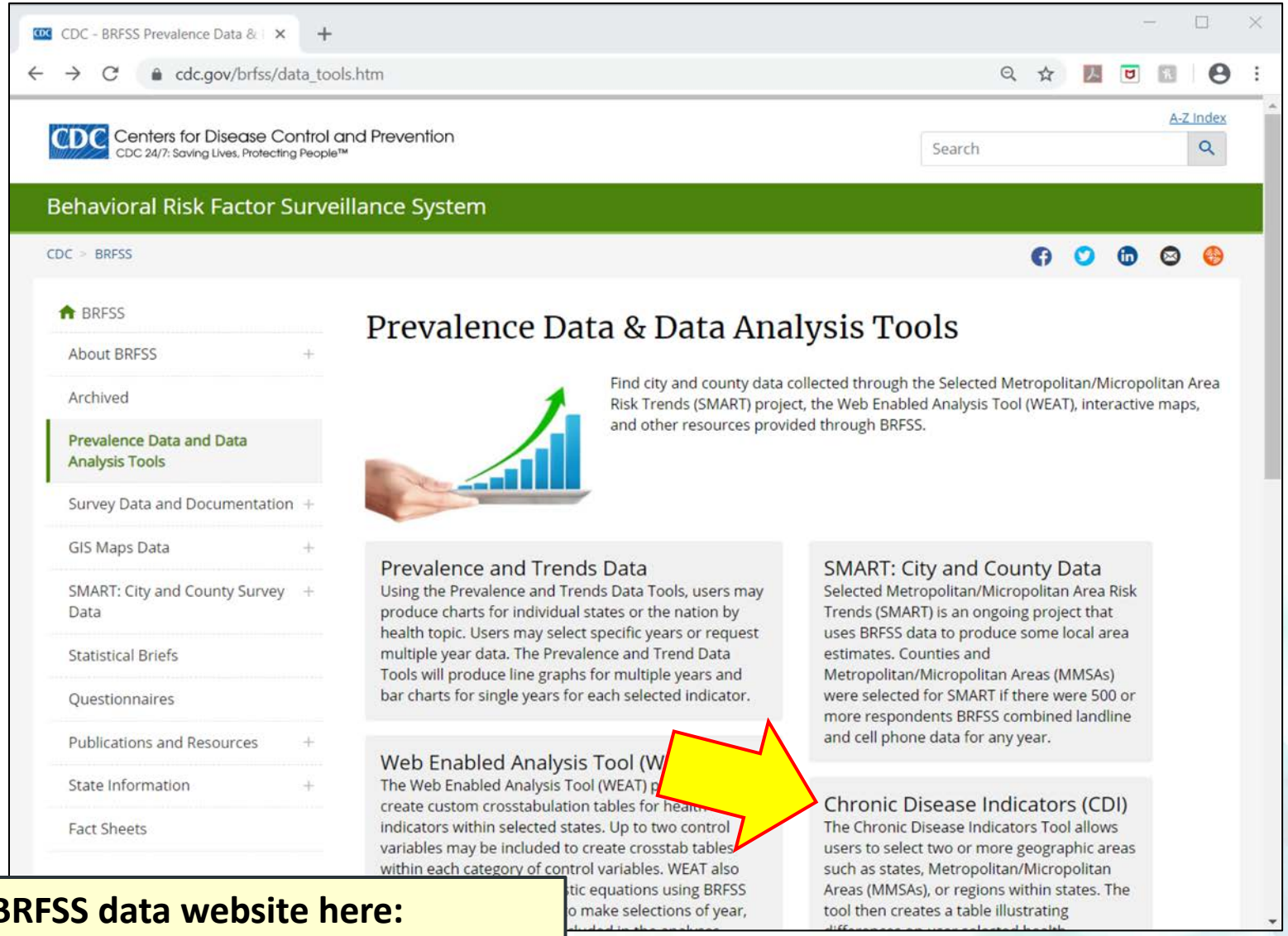
CHRONIC DISEASE INDICATORS TOOL (CDI)

GREAT PLAINS
IDeA | Clinical and Translational Research

PREVALENCE AND TRENDS DATA TOOL

- Quick tool to explore prevalence using a wide range of variables
- Produces charts and maps for individual states or the nation by health topics.
- Particularly useful to produce side by side tables of user selected indicators.
- Has a wider variety of variables compared to the ‘prevalence and trends data’ tool.

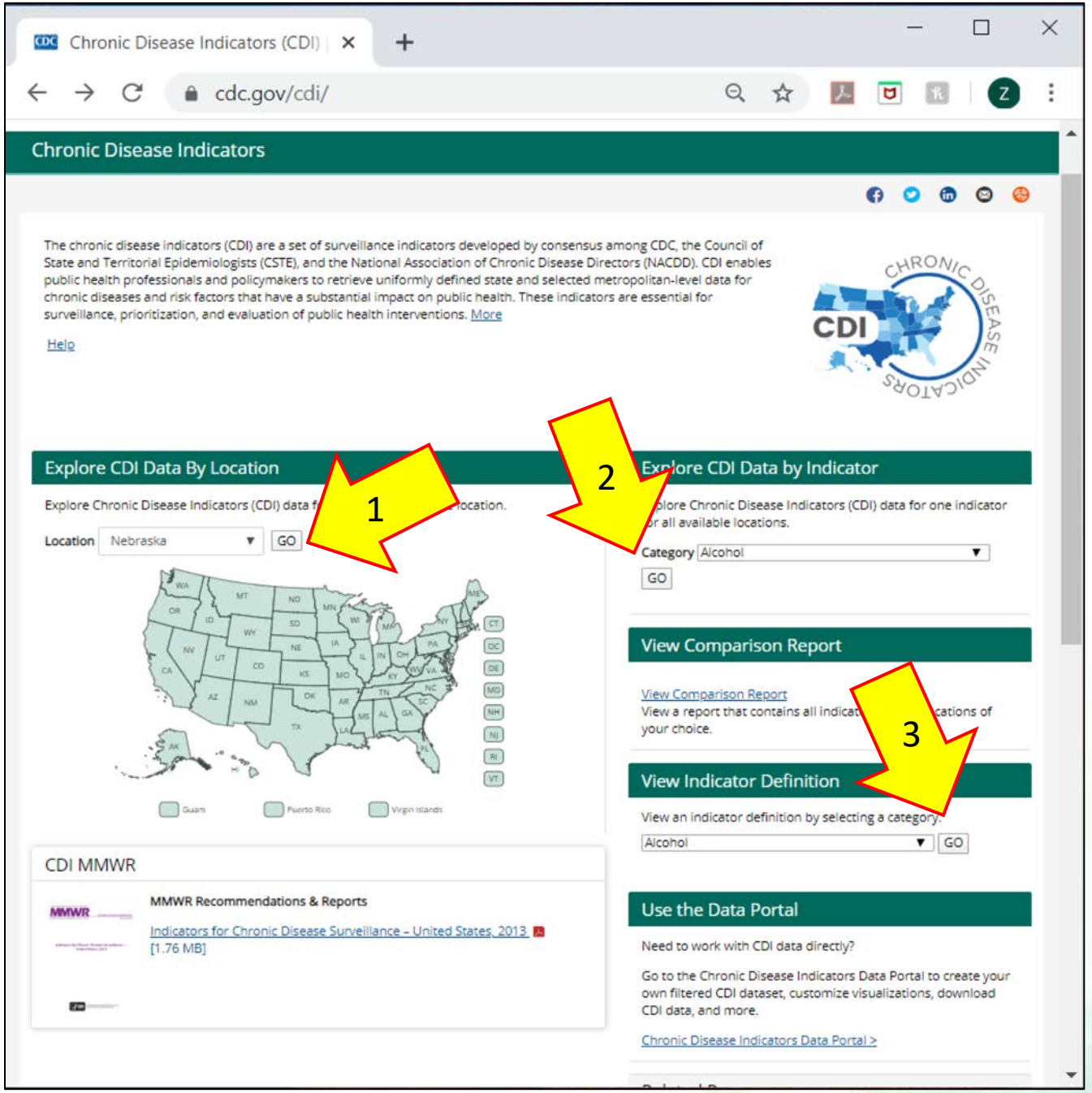
GREAT PLAINS IDeA | Clinical and Translational Research



Go to the BRFSS data website here:
https://www.cdc.gov/brfss/data_tools.htm

Click "Web Enabled Analysis Tool"

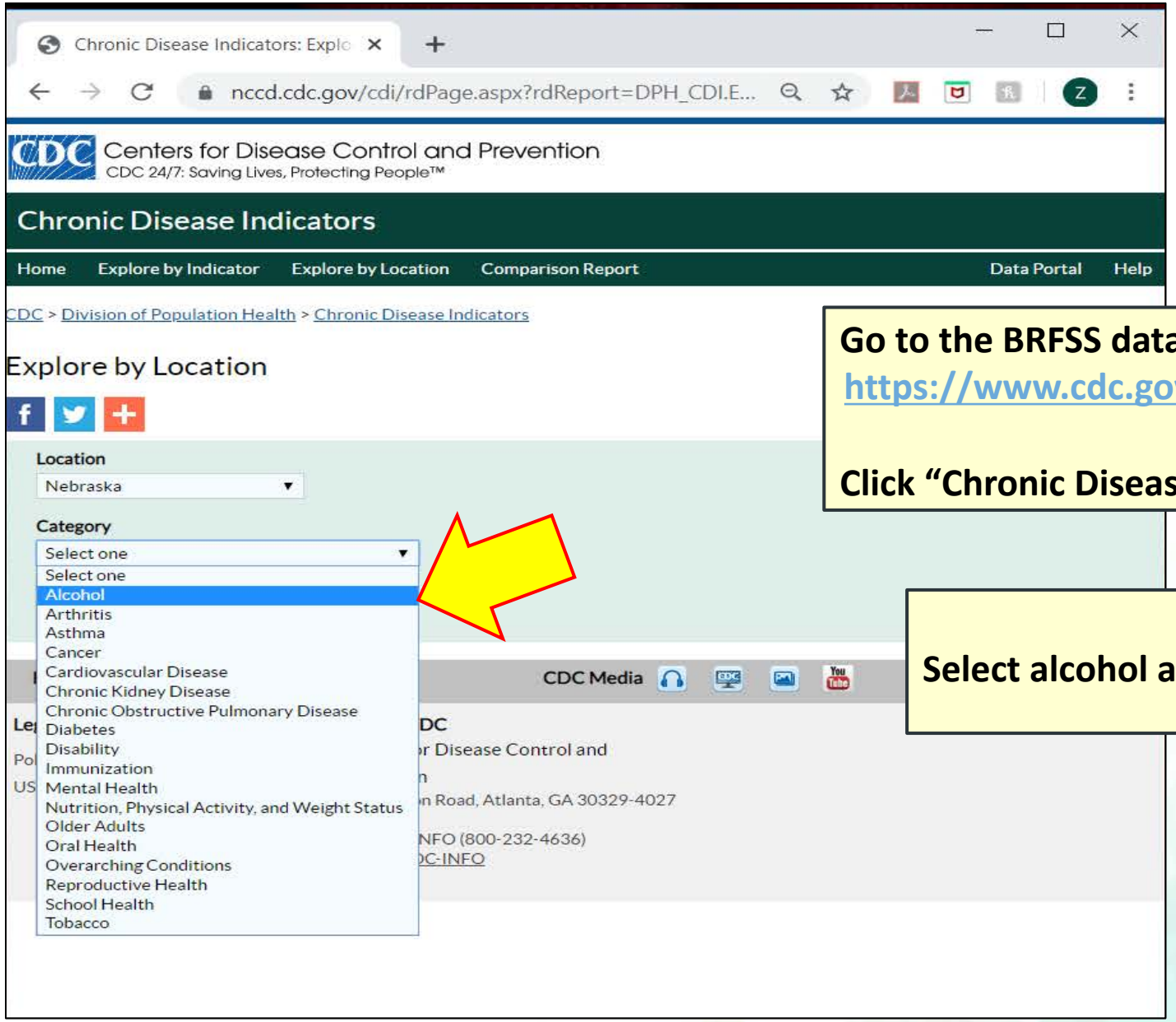
GREAT PLAINS
IDeA | Clinical and Translational Research



1. **To start with a geographic location, drop down ‘Location’, select the state and click ‘Go’.**
2. **To start with a topic drop down ‘Category’, select topic and click ‘Go’.**
3. **Use ‘View Indicator Definition’ to see how an indicator has been defined**

We will start with location. Select Nebraska and click ‘Go

GREAT PLAINS IDeA | Clinical and Translational Research

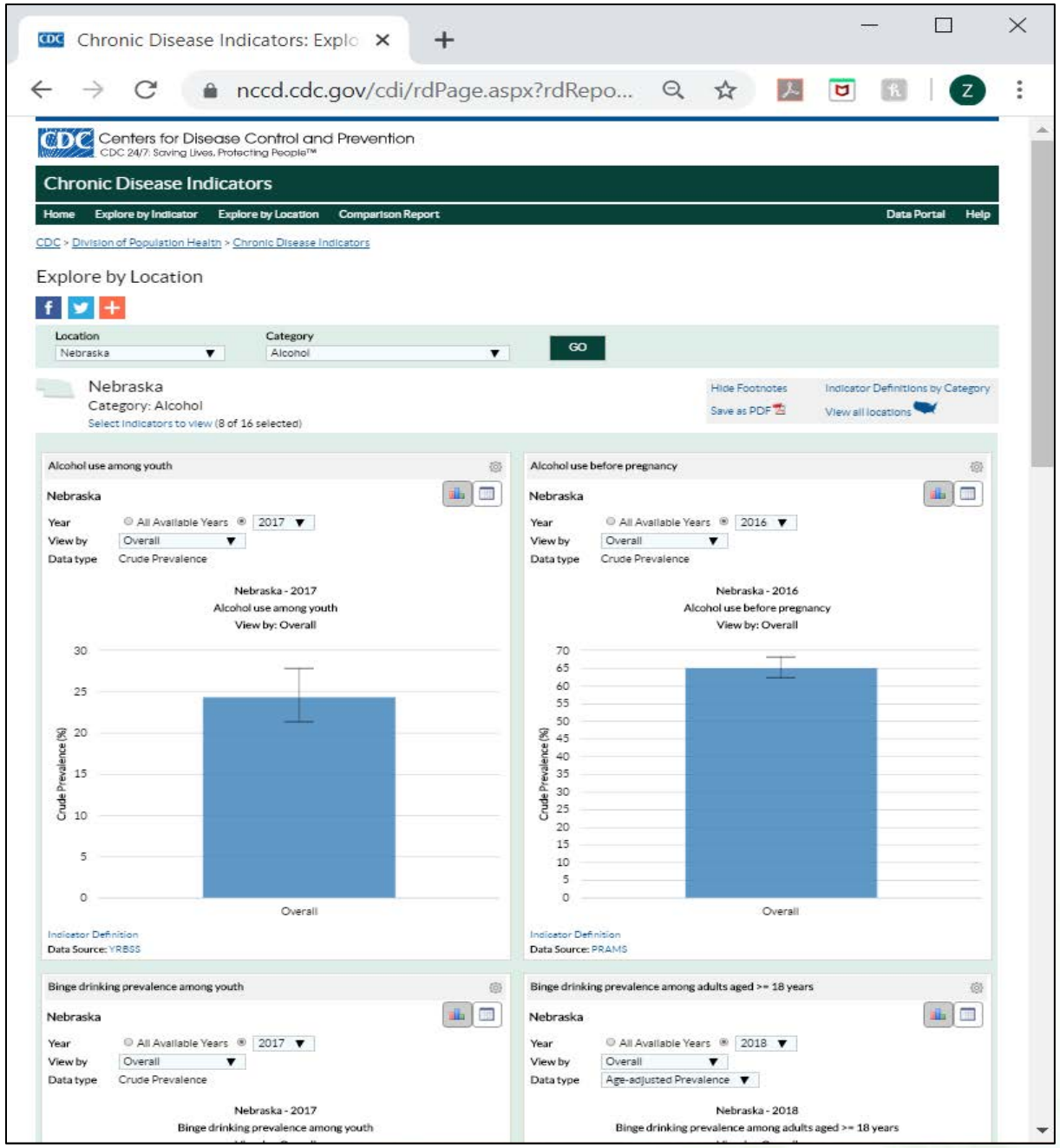


Go to the BRFSS data website here:
https://www.cdc.gov/brfss/data_tools.htm

Click "Chronic Disease Indicators"

Select alcohol and click 'Go'

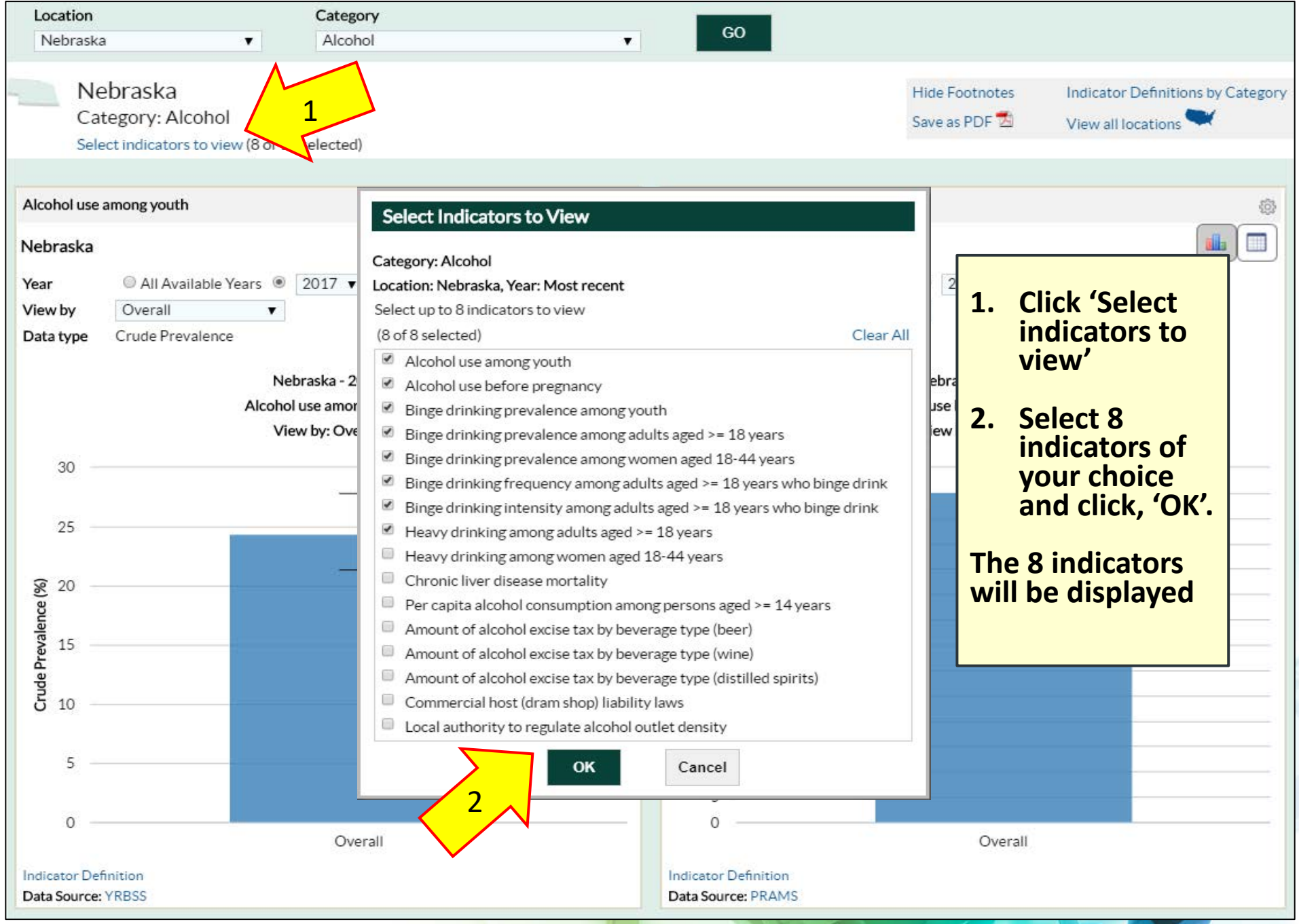
GREAT PLAINS IDeA | Clinical and Translational Research



Shows multiple alcohol indicators side by side.

Scroll down your browser to see a total of eight indicator at a time.

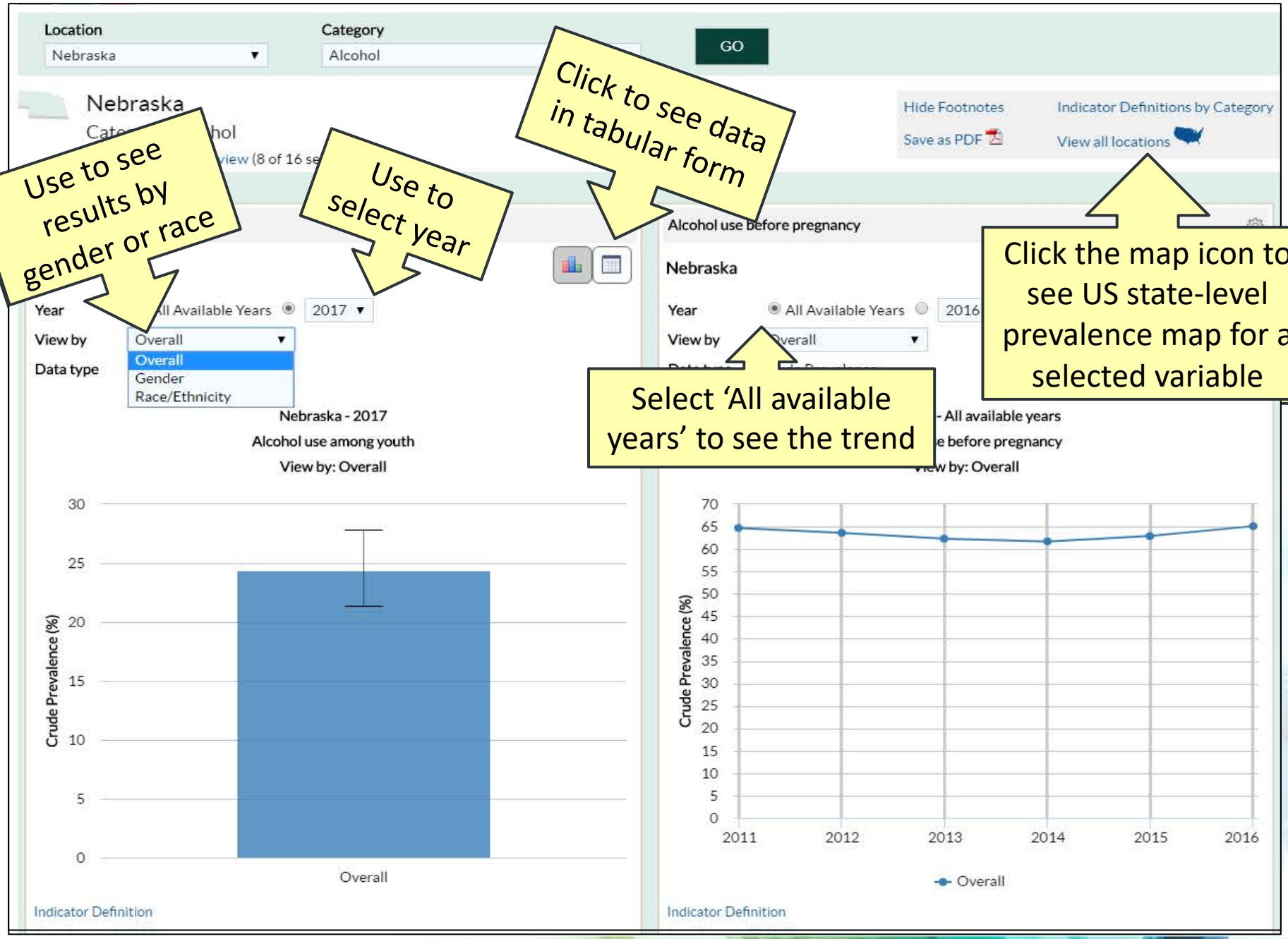
GREAT PLAINS
IDeA | Clinical and Translational Research



1. **Click 'Select indicators to view'**
2. **Select 8 indicators of your choice and click, 'OK'.**

The 8 indicators will be displayed

GREAT PLAINS IDeA | Clinical and Translational Research



GREAT PLAINS IDeA | Clinical and Translational Research

WEB ENABLED ANALYSIS TOOL (WEAT)

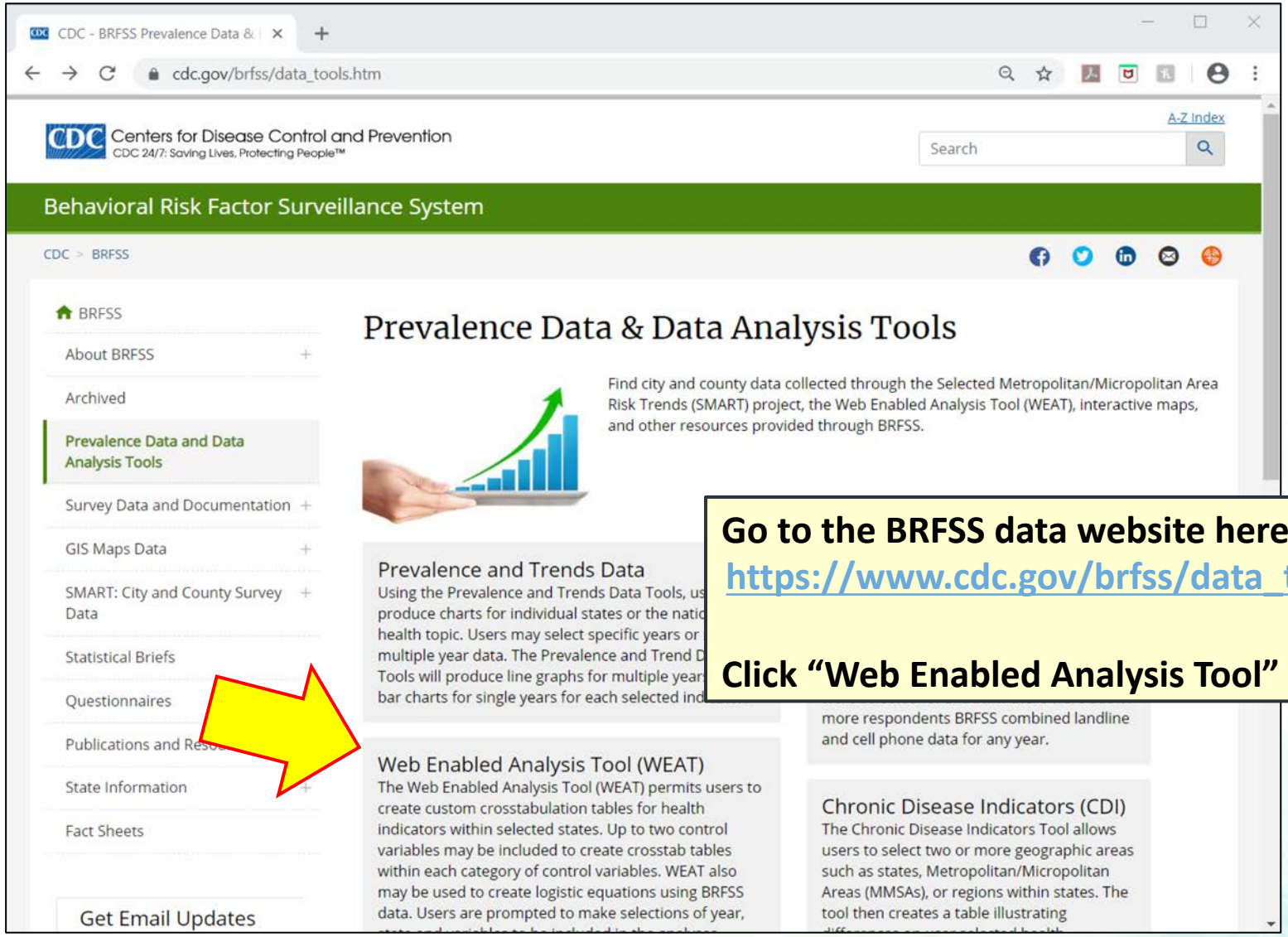
GREAT PLAINS
IDeA | Clinical and Translational Research

WEB ENABLED ANALYSIS TOOL (WEAT)

- **Enables users to create custom crosstabulation tables for variables within selected states**

- **Can be used to run logistic regression**

GREAT PLAINS
IDeA | Clinical and Translational Research



Go to the BRFSS data website here:
https://www.cdc.gov/brfss/data_tools.htm

Click "Web Enabled Analysis Tool"

GREAT PLAINS
IDeA | Clinical and Translational Research

ORAL HEALTH AND TOBACCO USE

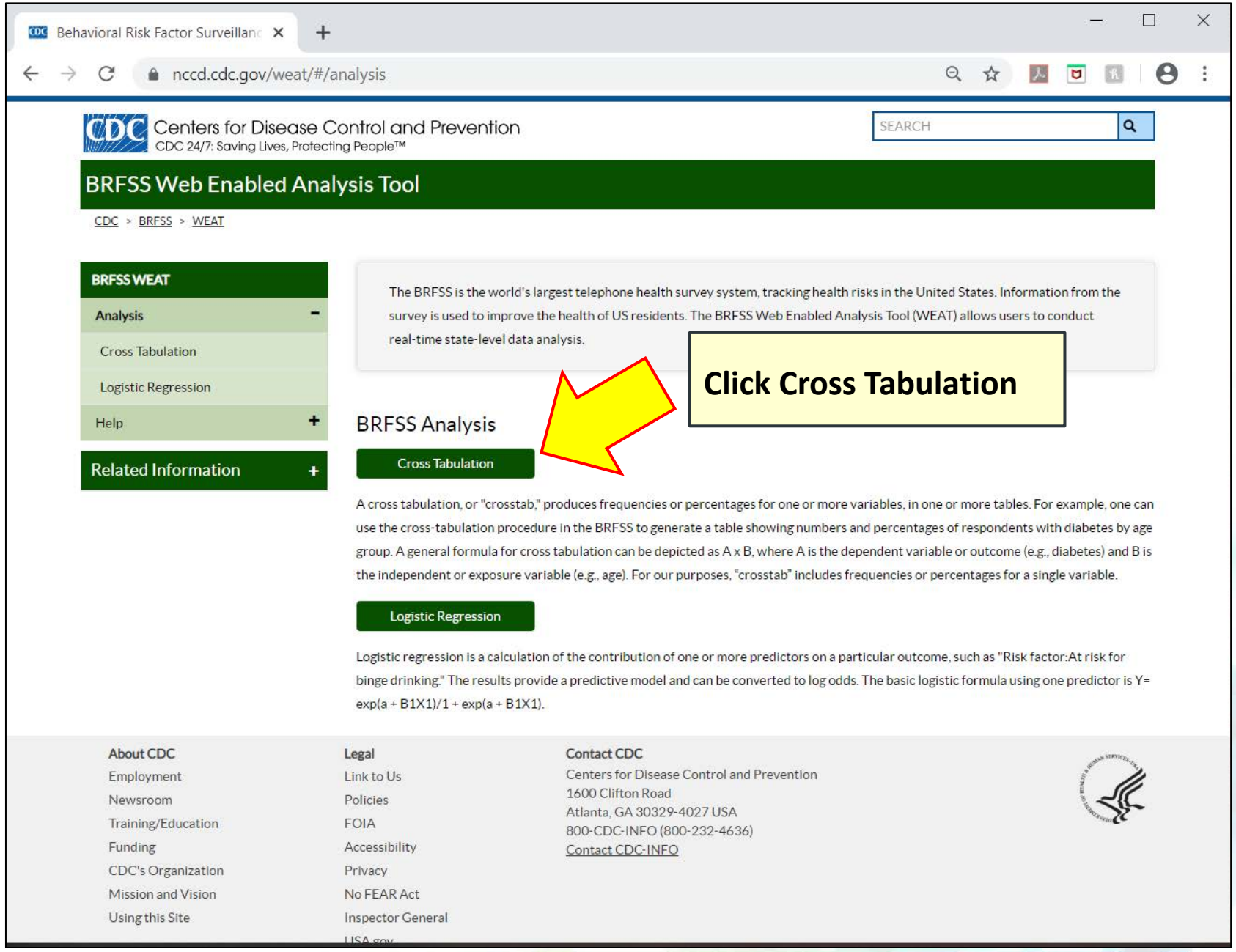
- This section will instruct you on how to use WEAT tool to explore if an association exists between oral health and tobacco use

- For this example, let us see how to get:

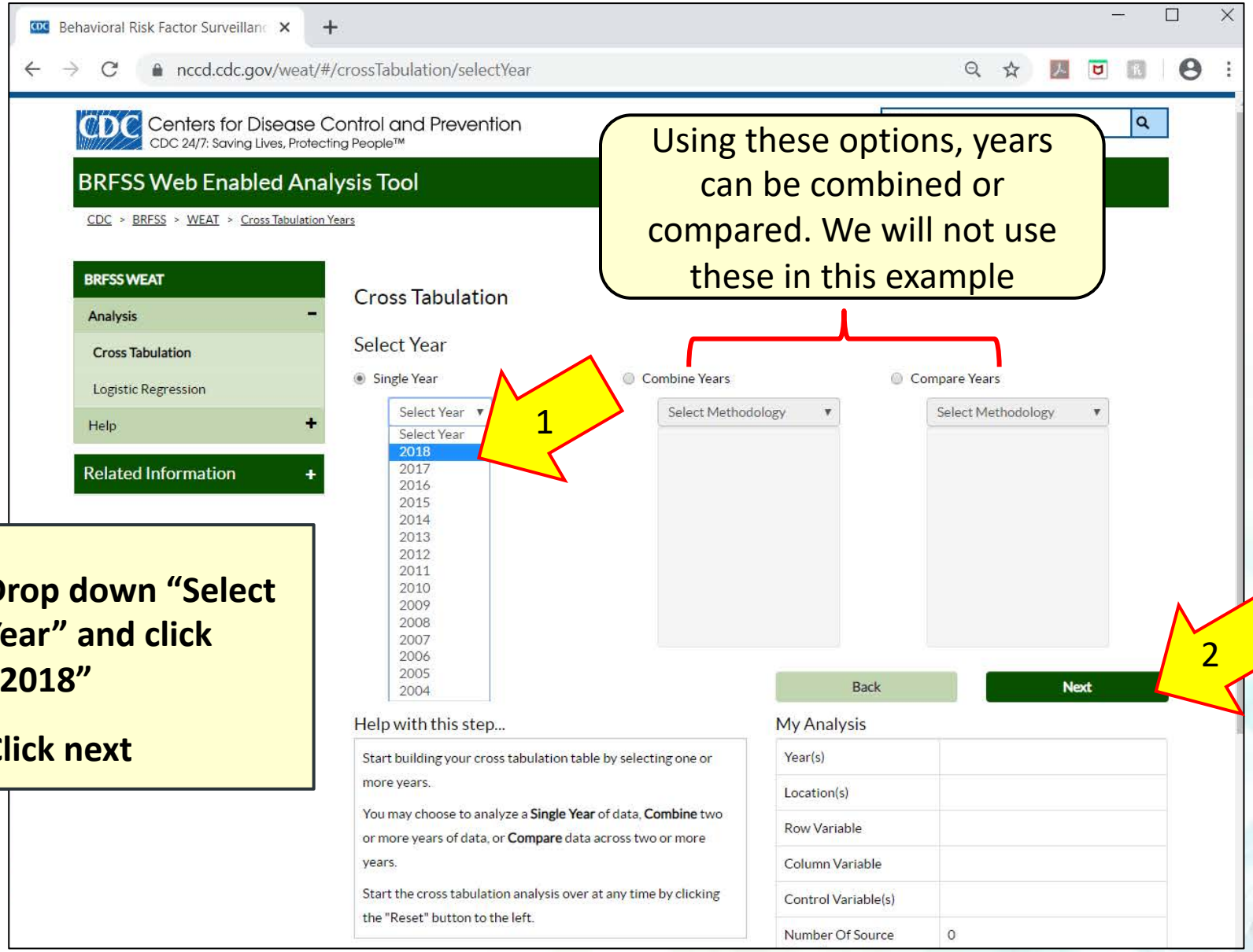
  - A cross-tabulation between current tobacco use and permanent teeth extraction stratified by gender among Nebraska residents in 2018

  - Gender-adjusted measure of association (odds ratio in this case) between current tobacco use and permanent teeth extraction among Nebraska residents in 2018 (as the outcome variable is binary, we will use logistic regression)

GREAT PLAINS IDeA | Clinical and Translational Research

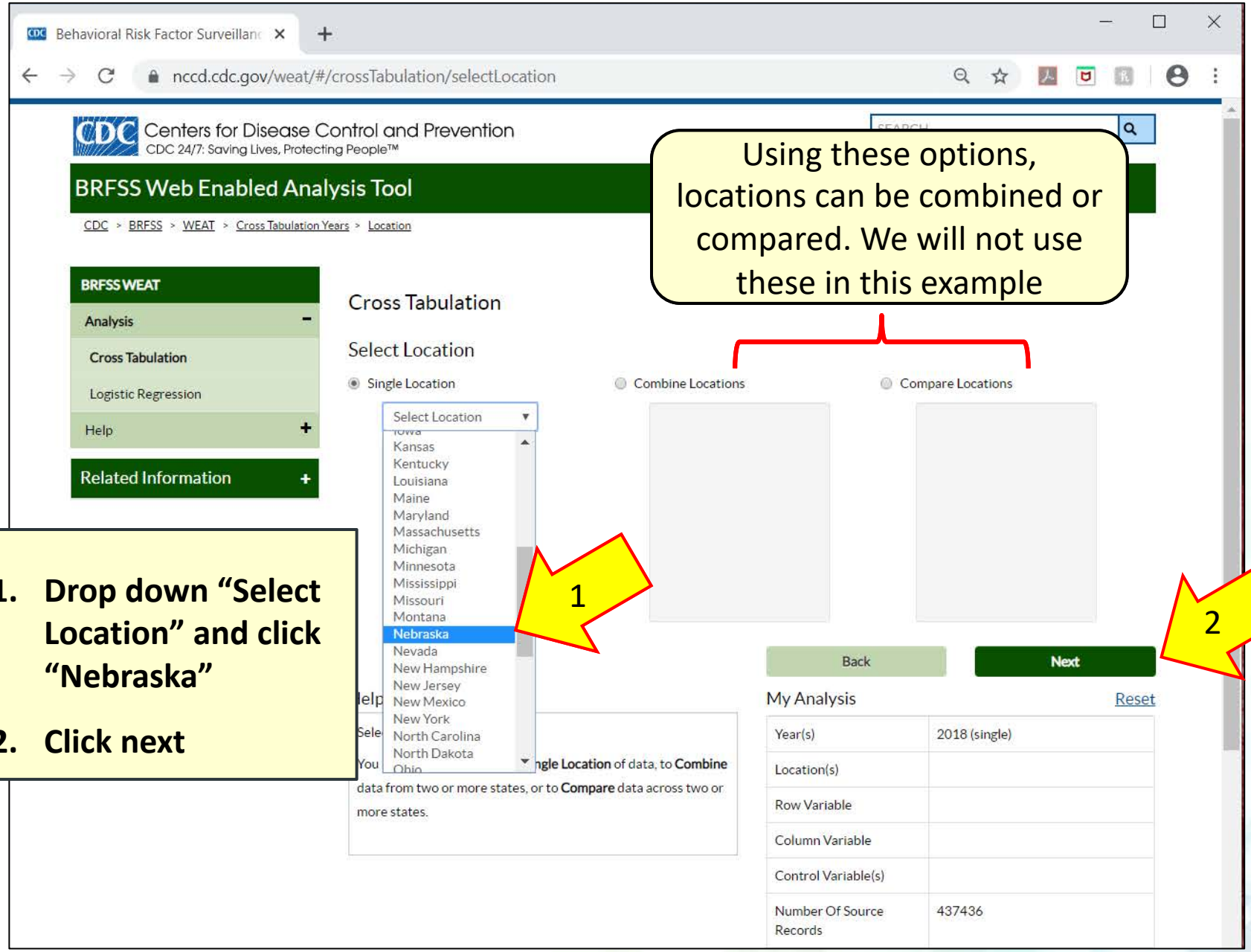


GREAT PLAINS
IDeA | Clinical and Translational Research



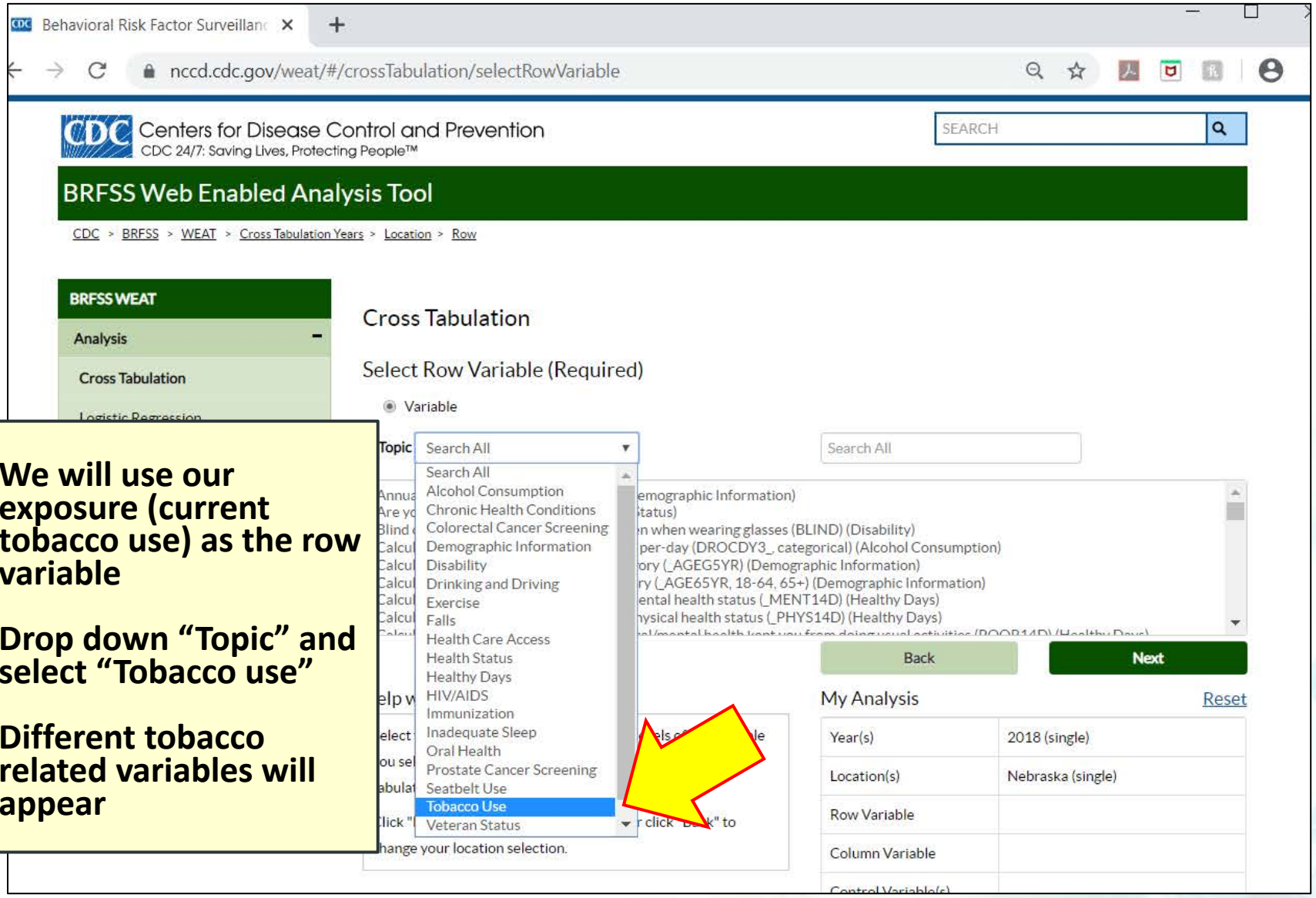
1. **Drop down "Select Year" and click "2018"**
2. **Click next**

GREAT PLAINS IDeA | Clinical and Translational Research



1. **Drop down "Select Location" and click "Nebraska"**
2. **Click next**

GREAT PLAINS IDeA | Clinical and Translational Research

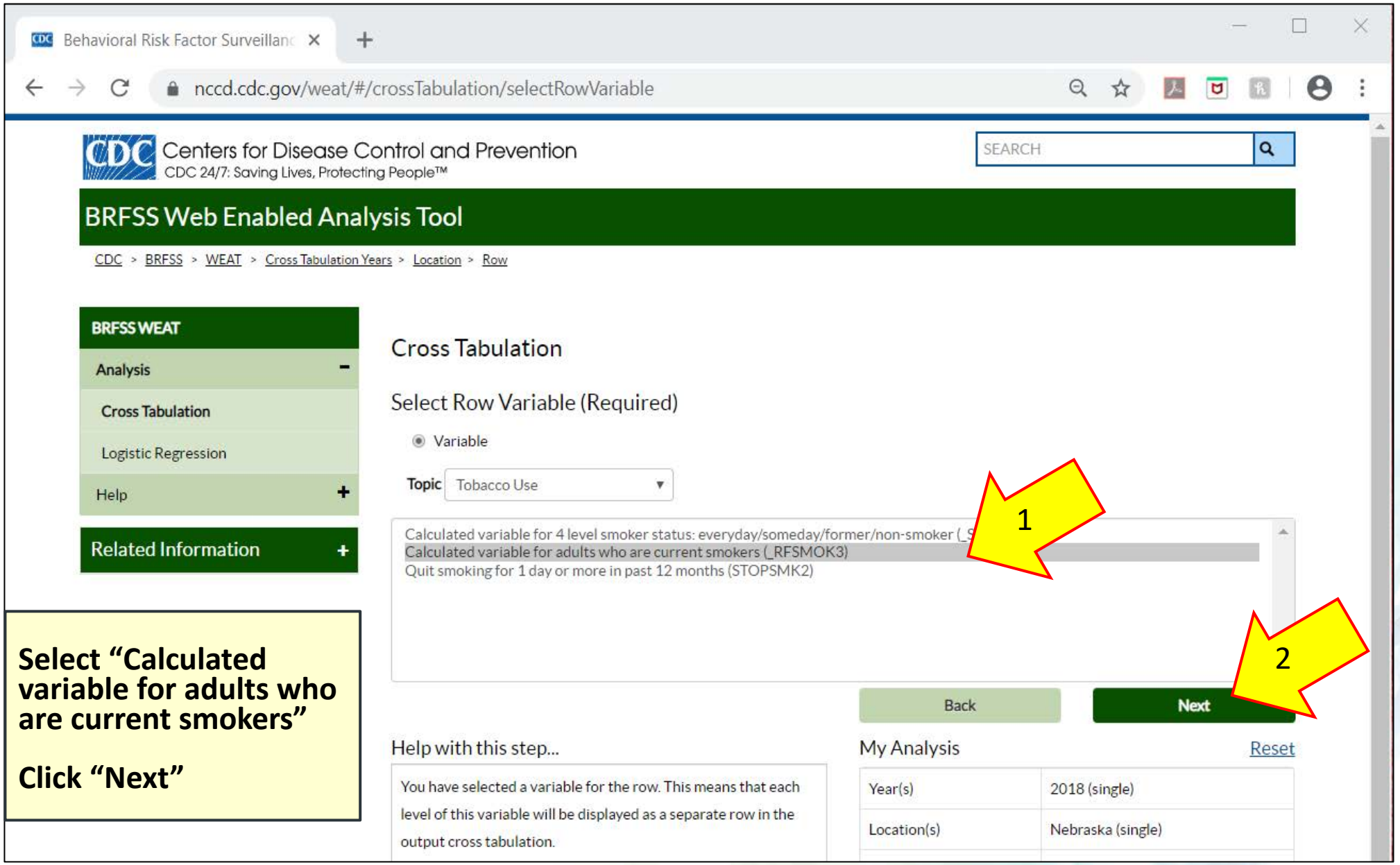


We will use our exposure (current tobacco use) as the row variable

Drop down "Topic" and select "Tobacco use"

Different tobacco related variables will appear

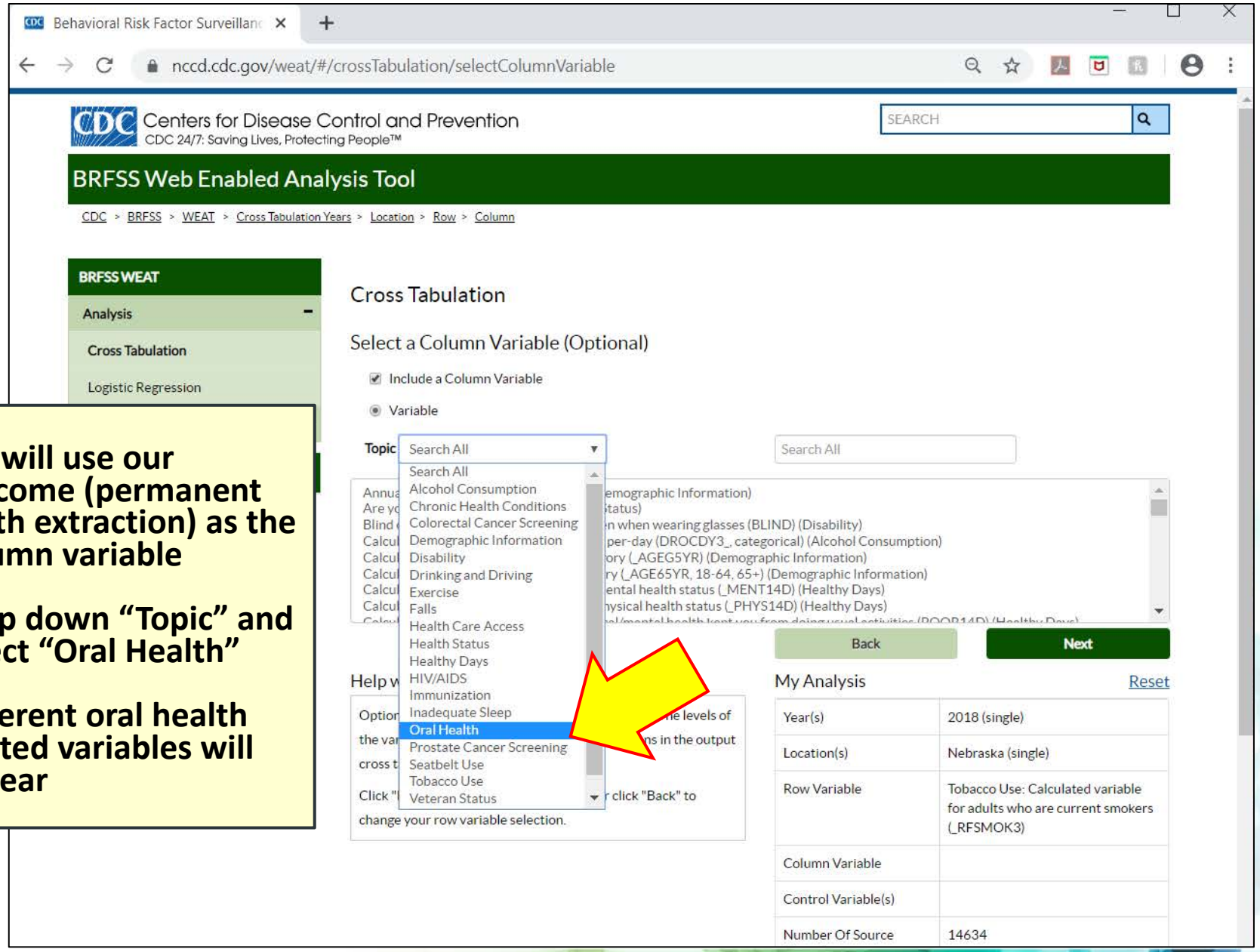
GREAT PLAINS IDeA | Clinical and Translational Research



Select "Calculated variable for adults who are current smokers"

Click "Next"

GREAT PLAINS IDeA | Clinical and Translational Research

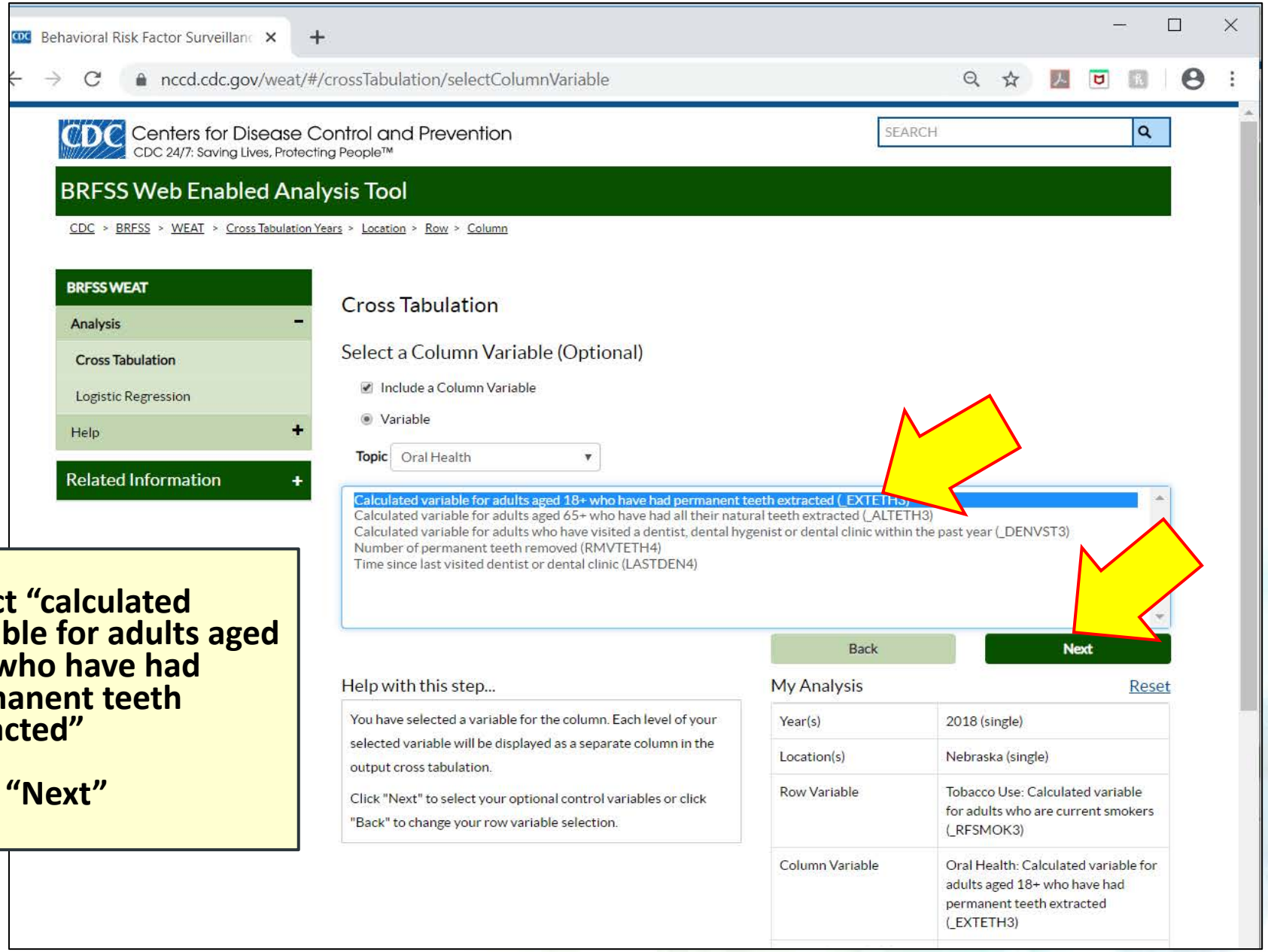


We will use our outcome (permanent tooth extraction) as the column variable

Drop down "Topic" and select "Oral Health"

Different oral health related variables will appear

GREAT PLAINS IDeA | Clinical and Translational Research



Select “calculated variable for adults aged 18+ who have had permanent teeth extracted”

Click “Next”

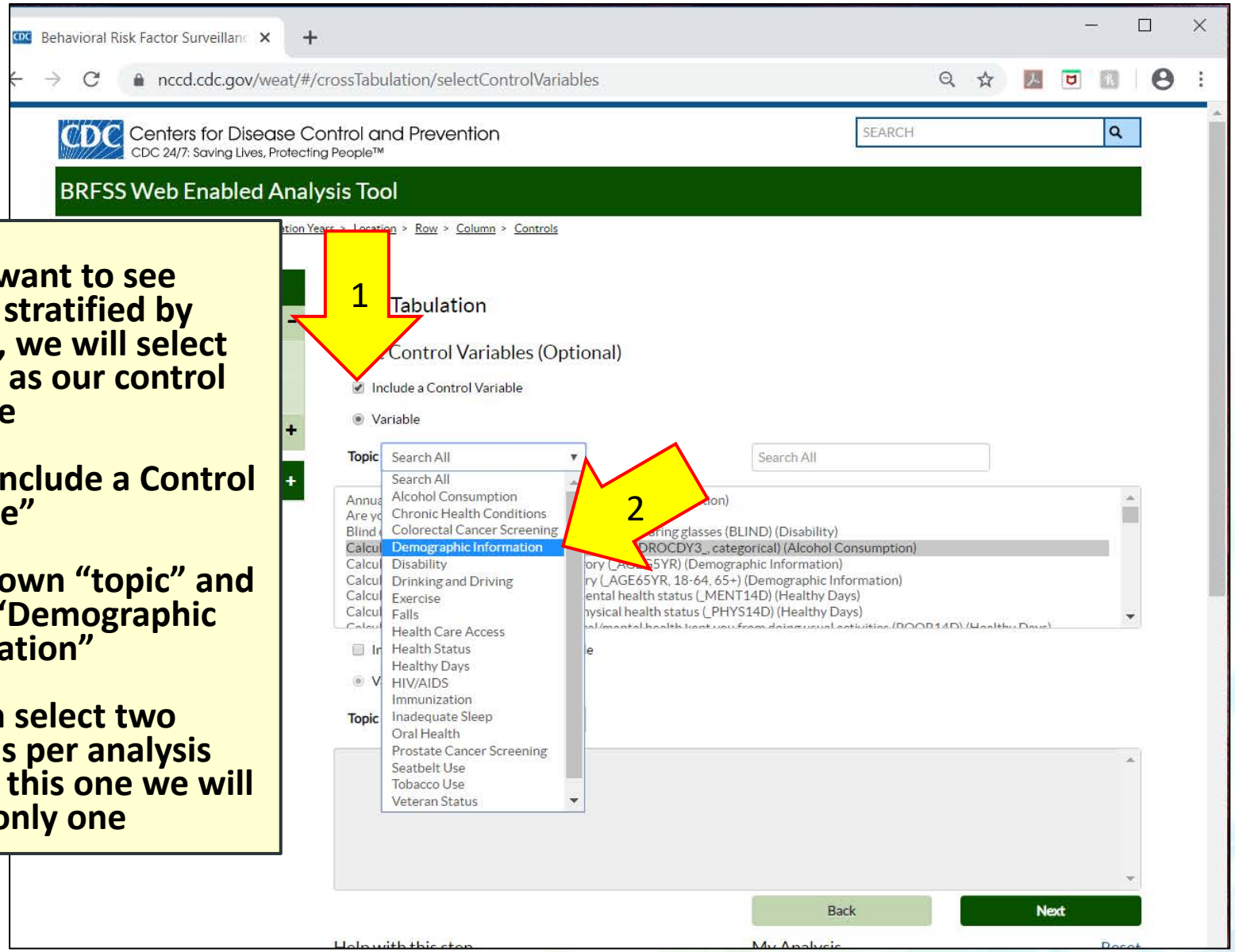
GREAT PLAINS IDeA | Clinical and Translational Research

As we want to see results stratified by gender, we will select gender as our control variable

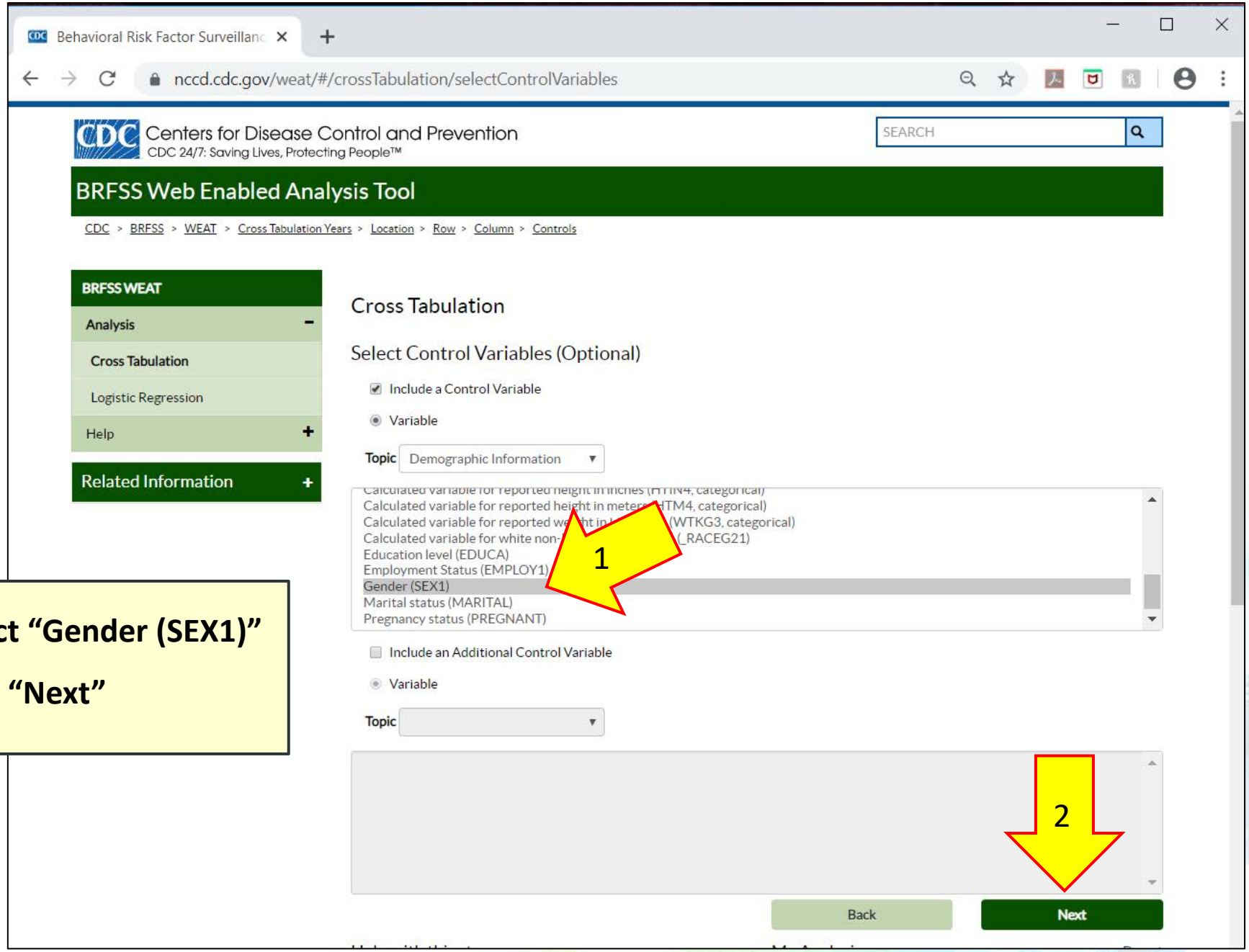
Click “Include a Control Variable”

Drop down “topic” and select “Demographic Information”

We can select two controls per analysis but for this one we will select only one



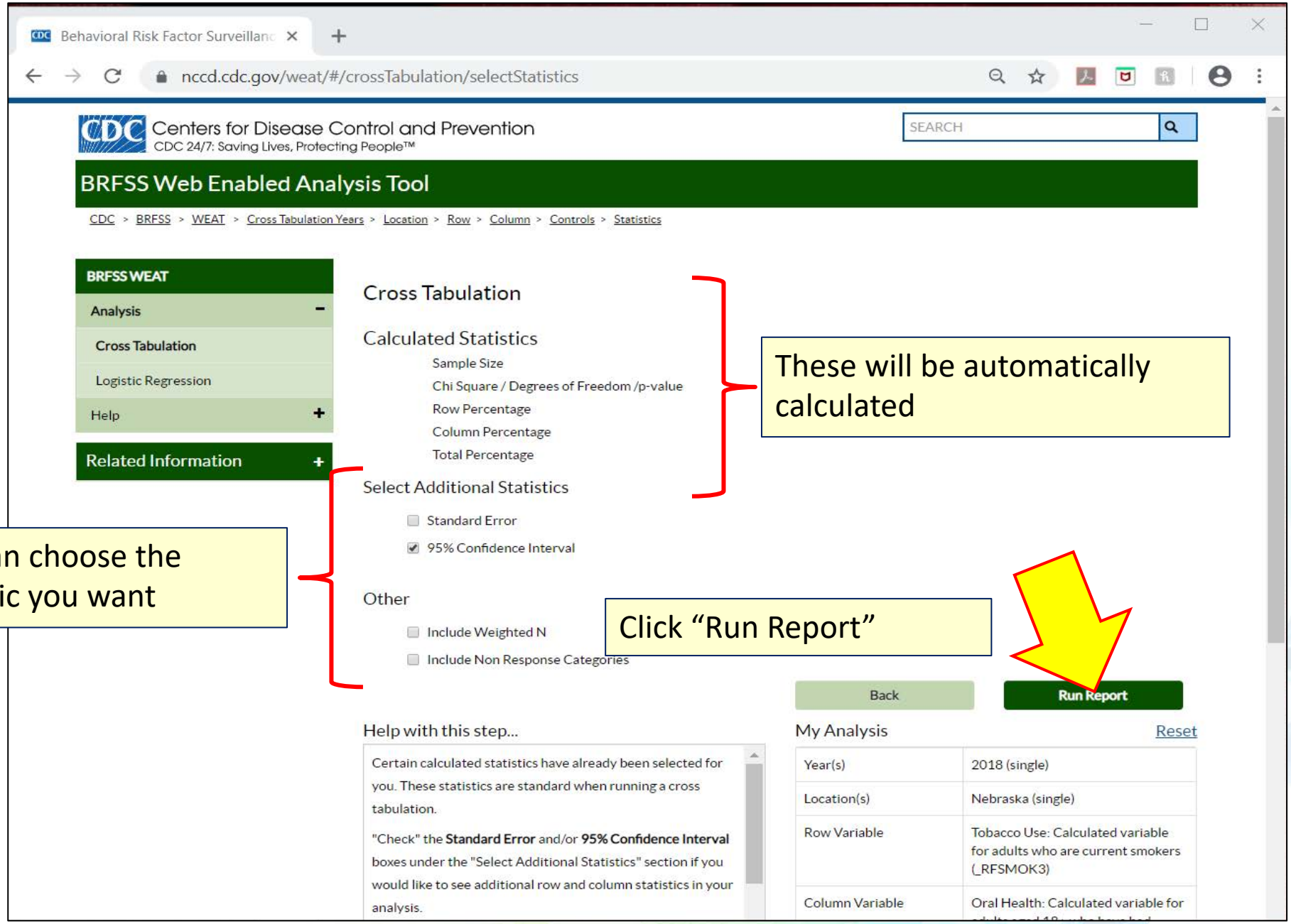
GREAT PLAINS IDeA | Clinical and Translational Research



Select “Gender (SEX1)”

Click “Next”

GREAT PLAINS IDeA | Clinical and Translational Research



GREAT PLAINS IDeA | Clinical and Translational Research

Behavioral Risk Factor Surveillance System
Cross Tabulation, Nebraska, 2018
of Tobacco Use by Oral Health

Sample Size	
Row %	(95% Confidence Interval)
Column %	(95% Confidence Interval)
Total % (Weighted)	(95% Confidence Interval)

		Calculated variable for adults aged 18+ who have had permanent teeth extracted (_EXTETH3)					
		No		Yes		Total	
Calculated variable for adults who are current smokers (_RFSMOK3)							
Former smoker or never smoked	n	6,665		5,396		12,061	
	Row%	64.8%	(63.6 - 66.1)	35.2%	(33.9 - 36.4)	100.0%	(100.0 - 100.0)
	Col%	87.6%	(86.5 - 88.8)	78.2%	(76.6 - 79.8)	84.1%	(83.1 - 85.0)
	%	54.5%	(53.3 - 55.8)	29.6%	(28.5 - 30.6)	84.1%	(83.1 - 85.0)
Current smoker	n	729		1,155		1,884	
	Row%	48.3%	(45.0 - 51.6)	51.7%	(48.4 - 55.0)	100.0%	(100.0 - 100.0)
	Col%	12.4%	(11.2 - 13.5)	21.8%	(20.2 - 23.4)	15.9%	(15.0 - 16.9)
	%	7.7%	(6.9 - 8.4)	8.2%	(7.6 - 8.9)	15.9%	(15.0 - 16.9)
Total	n	7,394		6,551		13,945	
	Row%	62.2%	(61.0 - 63.4)	37.8%	(36.6 - 39.0)		
	Col%	100.0%	(100.0 - 100.0)	100.0%	(100.0 - 100.0)		
	%	62.2%	(61.0 - 63.4)	37.8%	(36.6 - 39.0)		

Wald Chi-Square Value	Degrees of Freedom	p-value
85.39	1	<0.0001

Number of records on the BRFSS dataset for the year and location selected	14,634
Number of records excluded from the analysis	689
Sample Size (Number of records used in the analysis)	13,945
Total Weighted N (Population)	1,395,494

Multiple result tables are produced.

The first table, as shown here, shows overall sample without stratification

In 2018, among former/never smokers living in Nebraska, the prevalence of permanent tooth extraction was 32.5%

In 2018, among current smokers living in Nebraska, the prevalence of permanent tooth extraction was 51.7%

The Chi-Square p-value (<0.0001) suggests a significant association between smoking status and permanent tooth extraction

GREAT PLAINS
IDeA | Clinical and
Translational Research

Controlling for: Gender (SEX1) = Male

Sample Size
Row % (95% Confidence Interval)
Column % (95% Confidence Interval)
Total % (Weighted) (95% Confidence Interval)

Calculated variable for adults who are current smokers (_RFSMOK3)		Calculated variable for adults aged 18+ who have had permanent teeth extracted (_EXTETH3) No		Yes		Total	
Former smoker or never smoked	n	2,978		2,338		5,316	
	Row%	64.7%	(62.9 - 66.6)	35.3%	(33.4 - 37.1)	100.0%	(100.0 - 100.0)
	Col%	86.1%	(84.4 - 87.9)	76.9%	(74.5 - 79.3)	82.6%	(81.2 - 84.1)
	%	53.5%	(51.6 - 55.3)	29.2%	(27.6 - 30.7)	82.6%	(81.2 - 84.1)
Current smoker	n	394		548		942	
	Row%	49.5%	(45.0 - 54.1)	50.5%	(45.9 - 55.0)	100.0%	(100.0 - 100.0)
	Col%	13.9%	(12.1 - 15.6)	23.1%	(20.7 - 25.5)	17.4%	(15.9 - 18.8)
	%	8.6%	(7.5 - 9.7)	8.8%	(7.8 - 9.8)	17.4%	(15.9 - 18.8)
Total	n	3,372		2,886		6,258	
	Row%	62.1%	(60.3 - 63.8)	37.9%	(36.2 - 39.7)		
	Col%	100.0%	(100.0 - 100.0)	100.0%	(100.0 - 100.0)		
	%	62.1%	(60.3 - 63.8)	37.9%	(36.2 - 39.7)		

Wald Chi-Square Value	Degrees of Freedom	p-value
36.83	1	<0.0001

The second table, as shown here, shows results only for males

In 2018, among male former/never smokers living in Nebraska, the prevalence of permanent tooth extraction was 35.3%

In 2018, among male current smokers living in Nebraska, the prevalence of permanent tooth extraction was 50.5%

The Chi-Square p-value (<0.0001) suggests a significant association between smoking status and permanent tooth extraction in males

GREAT PLAINS IDeA | Clinical and Translational Research

Controlling for: Gender (SEX1) = Female

Sample Size
Row % (95% Confidence Interval)
Column % (95% Confidence Interval)
Total % (Weighted) (95% Confidence Interval)

Calculated variable for adults who are current smokers (_RFSMOK3)		Calculated variable for adults aged 18+ who have had permanent teeth extracted (_EXTETH3) No		Yes		Total	
Former smoker or never smoked	n	3,681		3,055		6,736	
	Row%	65.0%	(63.4 - 66.7)	35.0%	(33.3 - 36.6)	100.0%	(100.0 - 100.0)
	Col%	89.1%	(87.5 - 90.6)	79.5%	(77.3 - 81.6)	85.5%	(84.2 - 86.7)
	%	55.6%	(53.9 - 57.3)	29.9%	(28.5 - 31.4)	85.5%	(84.2 - 86.7)
Current smoker	n	335		606		941	
	Row%	46.9%	(42.0 - 51.7)	53.1%	(48.3 - 58.0)	100.0%	(100.0 - 100.0)
	Col%	10.9%	(9.4 - 12.5)	20.5%	(18.4 - 22.7)	14.5%	(13.3 - 15.8)
	%	6.8%	(5.8 - 7.8)	7.7%	(6.8 - 8.6)	14.5%	(13.3 - 15.8)
Total	n	4,016		3,661		7,677	
	Row%	62.4%	(60.8 - 64.0)	37.6%	(36.0 - 39.2)		
	Col%	100.0%	(100.0 - 100.0)	100.0%	(100.0 - 100.0)		
	%	62.4%	(60.8 - 64.0)	37.6%	(36.0 - 39.2)		

Wald Chi-Square Value	Degrees of Freedom	p-value
48.98	1	<0.0001

The third table, as shown here, shows results only for females

In 2018, among female former/never smokers living in Nebraska, the prevalence of permanent tooth extraction was 35.0%

In 2018, among female current smokers living in Nebraska, the prevalence of permanent tooth extraction was 53.1%

The Chi-Square p-value (<0.0001) suggests a significant association between smoking status and permanent tooth extraction in females

GREAT PLAINS IDeA | Clinical and Translational Research

Controlling for: Gender (SEX1) = Don't know/Not Sure

Sample Size
Row % (95% Confidence Interval)
Column % (95% Confidence Interval)
Total % (Weighted) (95% Confidence Interval)

Calculated variable for adults who are current smokers (_RFSMOK3)		Calculated variable for adults aged 18+ who have had permanent teeth extracted (_EXTETH3) No	Yes	Total
Former smoker or never smoked	n	2	2	4
	Row%	*	*	100.0% (100.0 - 100.0)
	Col%	*	*	*
	%	*	*	*
Current smoker	n	0	1	1
	Row%	*	*	100.0% (100.0 - 100.0)
	Col%	*	*	*
	%	*	*	*
Total	n	2	3	5
	Row%	*	*	
	Col%	100.0% (100.0 - 100.0)	100.0% (100.0 - 100.0)	
	%	*	*	

Wald Chi-Square Value	Degrees of Freedom	p-value
0.69	1	0.4056

* Estimate not available if the unweighted sample size for the denominator was < 50 or the Relative Standard Error (RSE) is > 0.3.

Controlling for: Gender (SEX1) = Refused

Sample Size
Row % (95% Confidence Interval)
Column % (95% Confidence Interval)
Total % (Weighted) (95% Confidence Interval)

Calculated variable for adults who are current smokers (_RFSMOK3)		Calculated variable for adults aged 18+ who have had permanent teeth extracted (_EXTETH3) No	Yes	Total
Former smoker or never smoked	n	4	1	5
	Row%	*	*	100.0% (100.0 - 100.0)
	Col%	*	*	*
	%	*	*	*
Current smoker	n	0	0	0
	Row%	*	*	-
	Col%	*	*	*
	%	*	*	*
Total	n	4	1	5
	Row%	*	*	
	Col%	100.0% (100.0 - 100.0)	100.0% (100.0 - 100.0)	
	%	*	*	

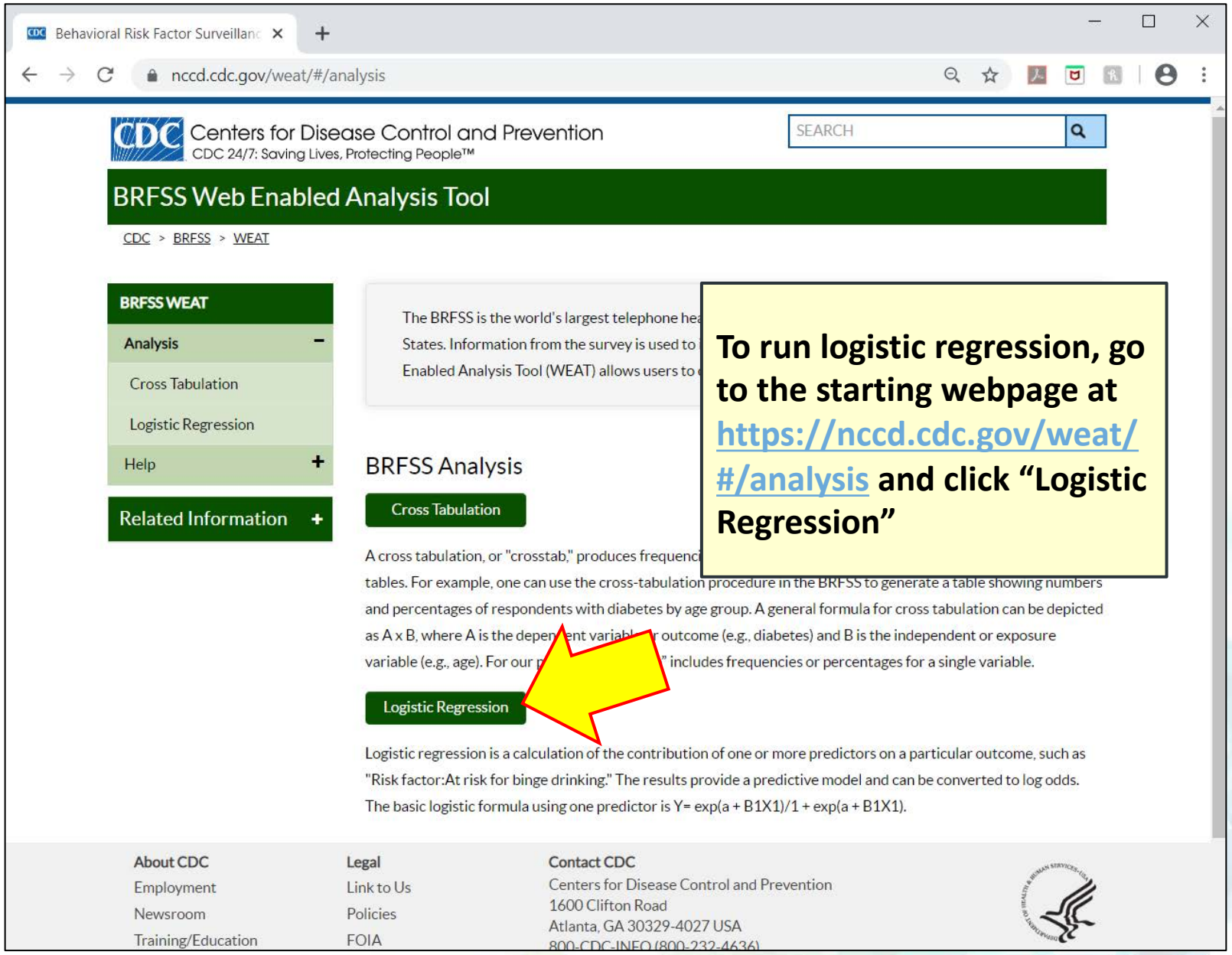
Wald Chi-Square Value	Degrees of Freedom	p-value
.	.	.

- Estimate could not be computed. Some estimates cannot be computed when the denominator sample size is 0, though other causes may exist as well

As the variable "Gender" had 4 categories including male, female, don't know/not sure and refused, we get the cross tabulations for the last two categories.

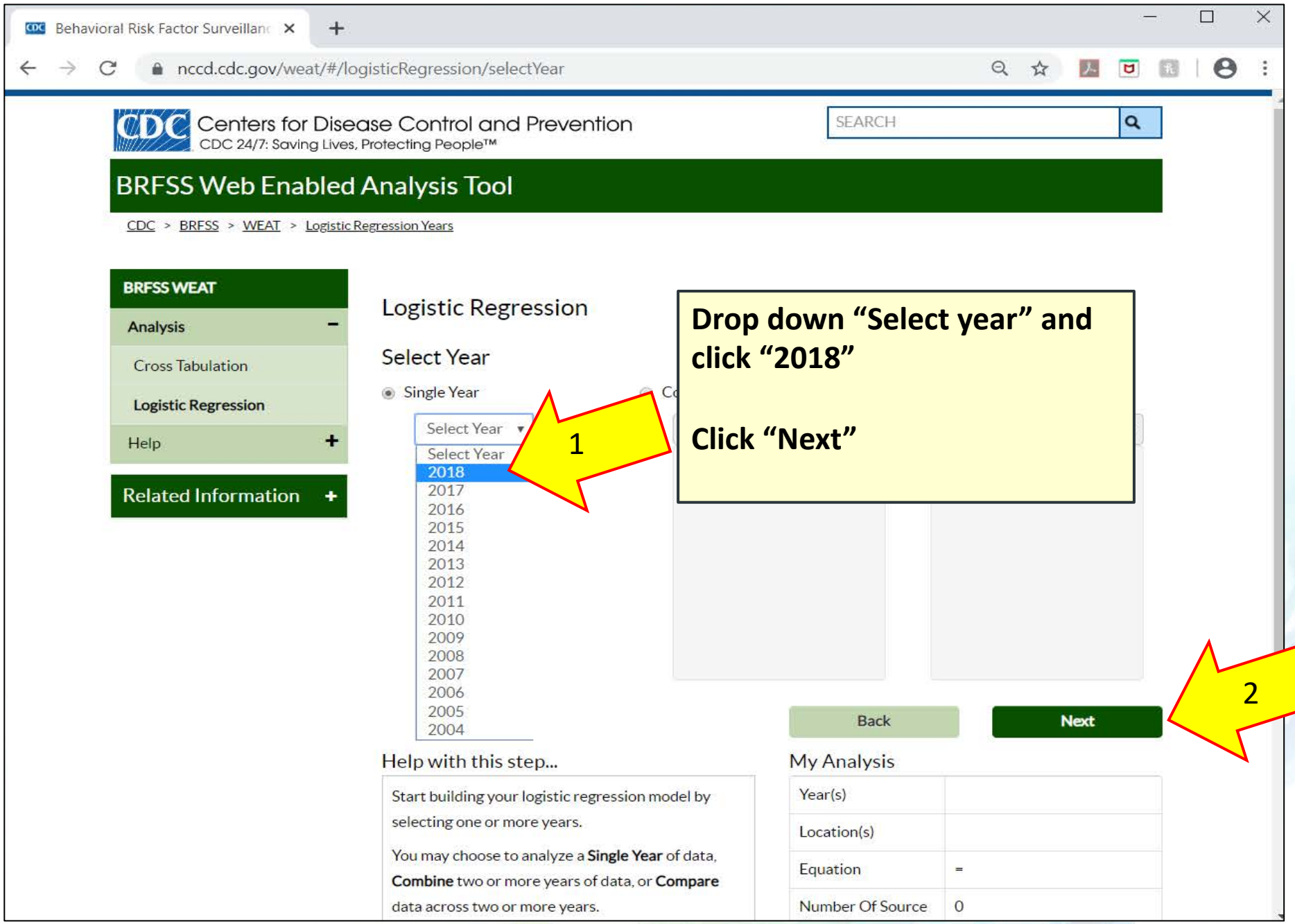
We can ignore these tables as the sample size is too small for any estimate computation. These categories may be converted to missing data in a proper statistical software. This online analysis tool does not provide an option for recoding the variables

GREAT PLAINS IDeA | Clinical and Translational Research

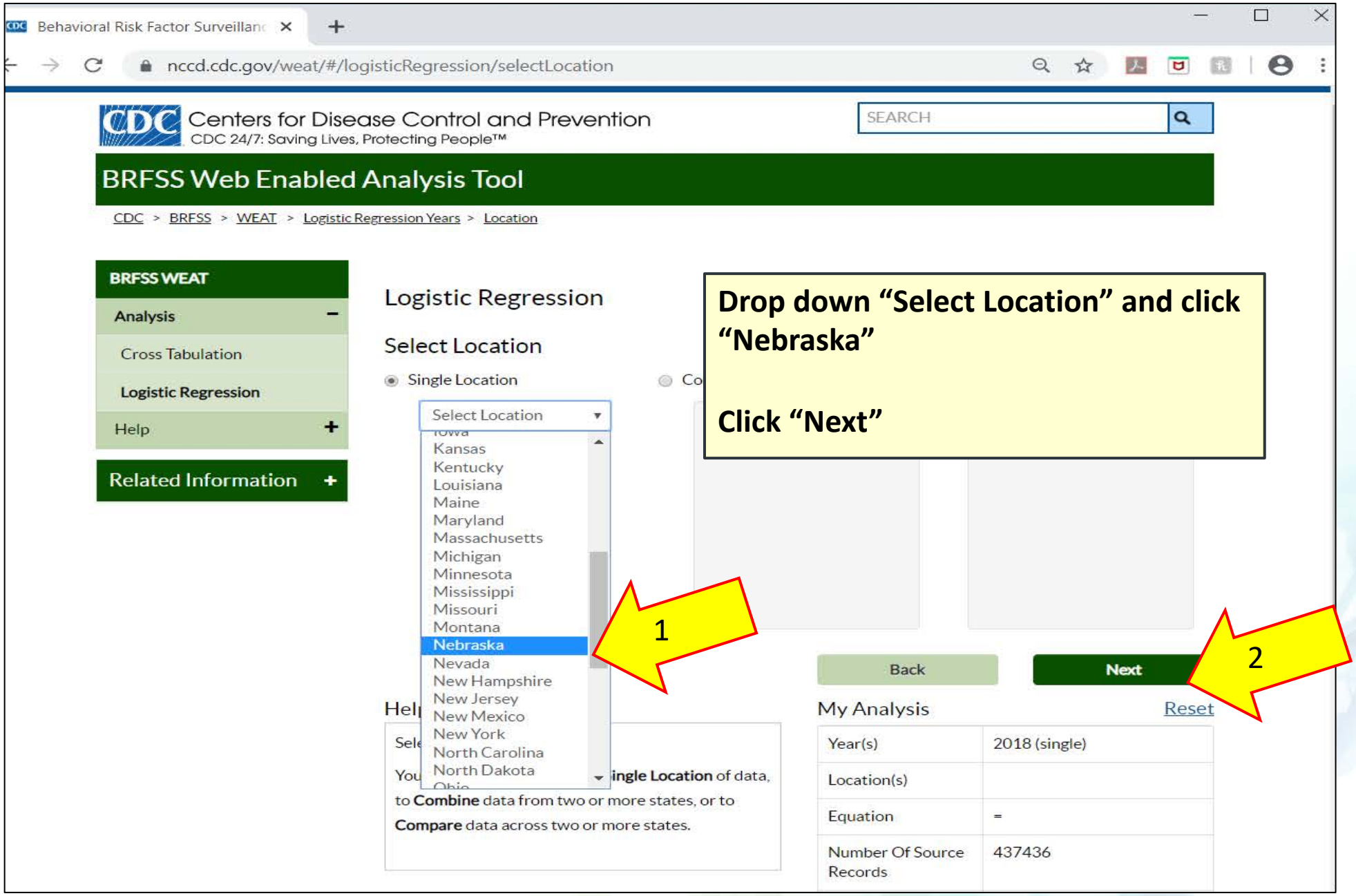


To run logistic regression, go to the starting webpage at https://nccd.cdc.gov/weat/#/analysis and click "Logistic Regression"

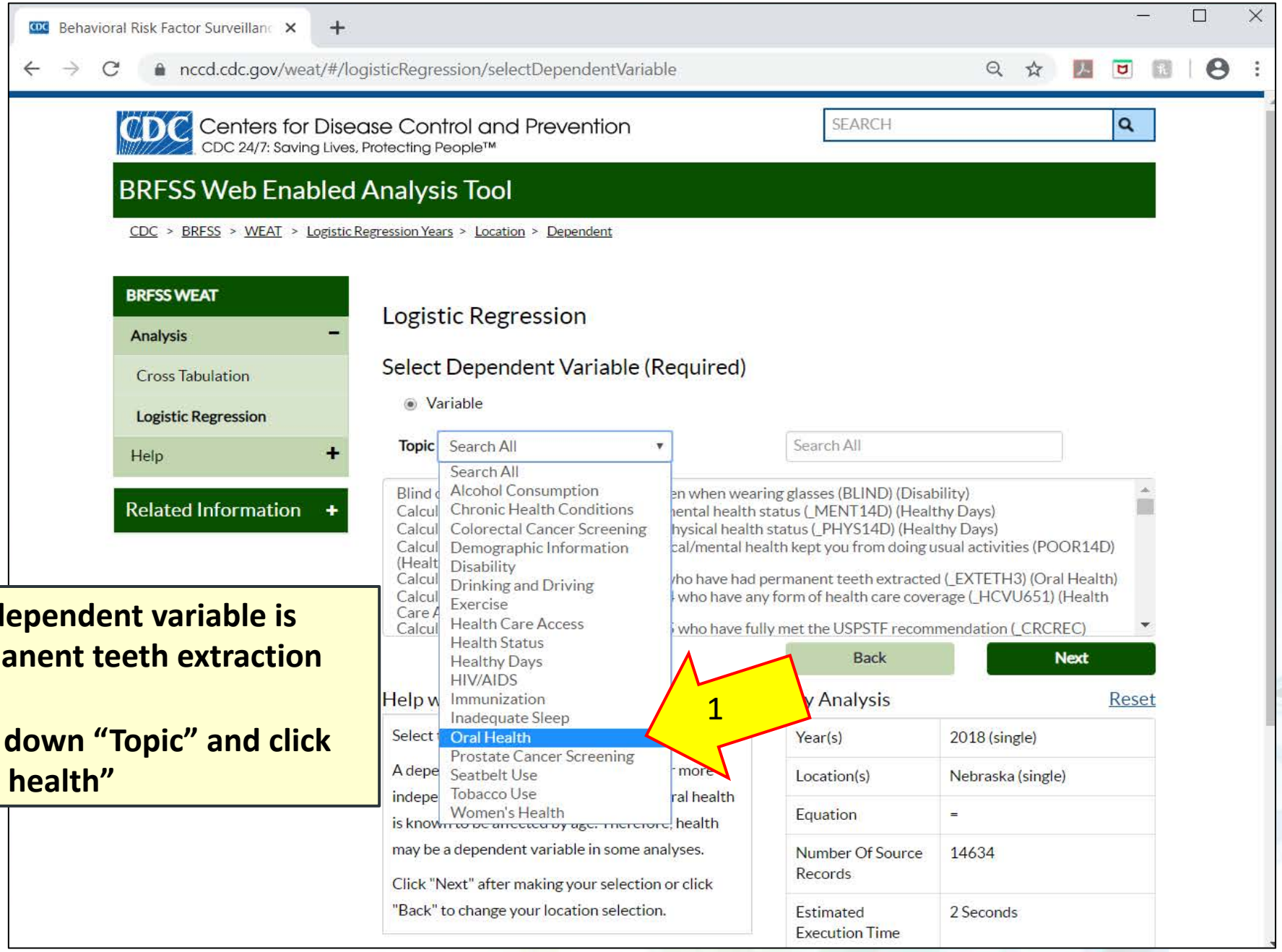
GREAT PLAINS IDeA | Clinical and Translational Research



GREAT PLAINS IDeA | Clinical and Translational Research



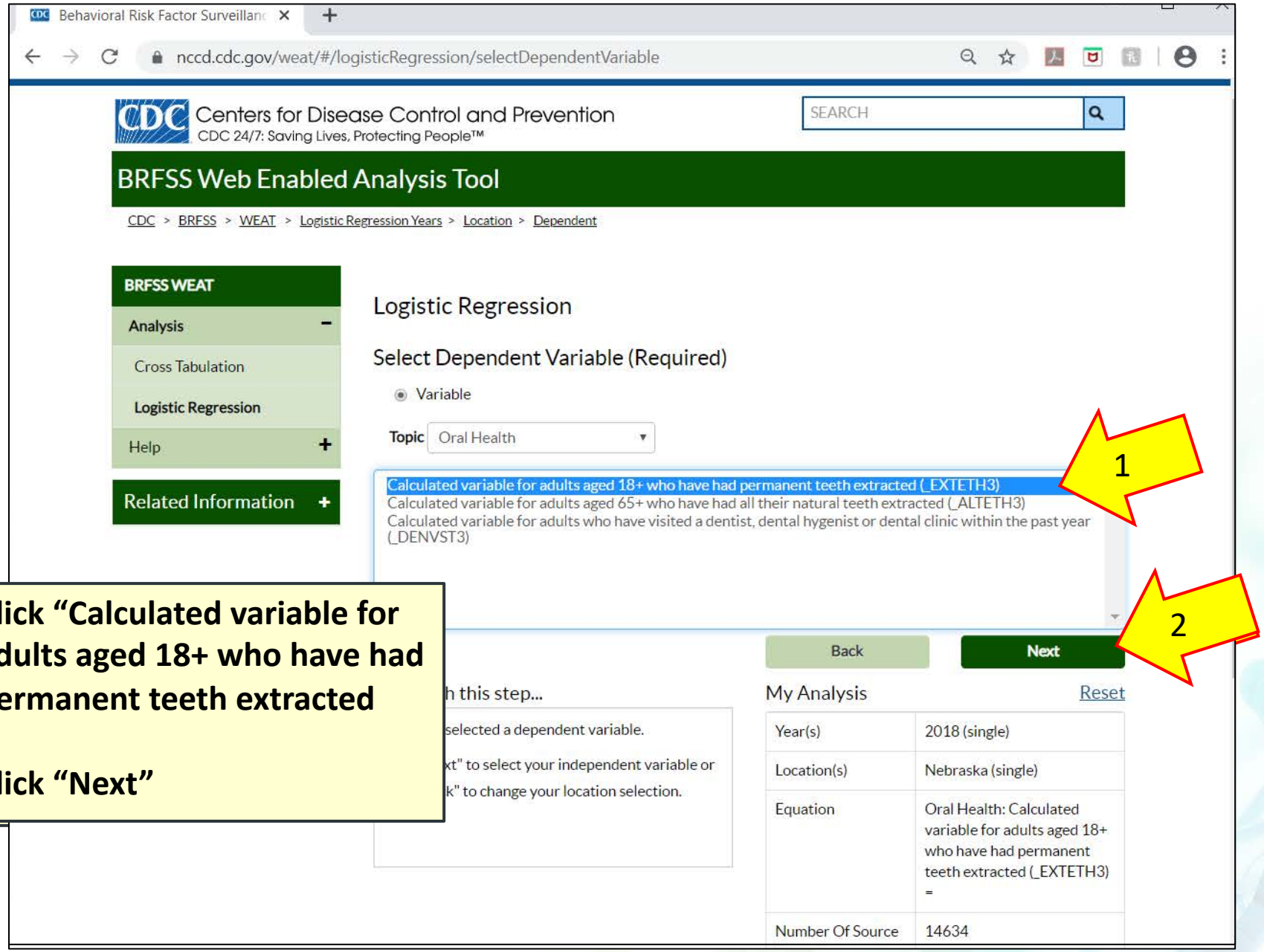
GREAT PLAINS IDeA | Clinical and Translational Research



Our dependent variable is permanent teeth extraction

Drop down “Topic” and click “Oral health”

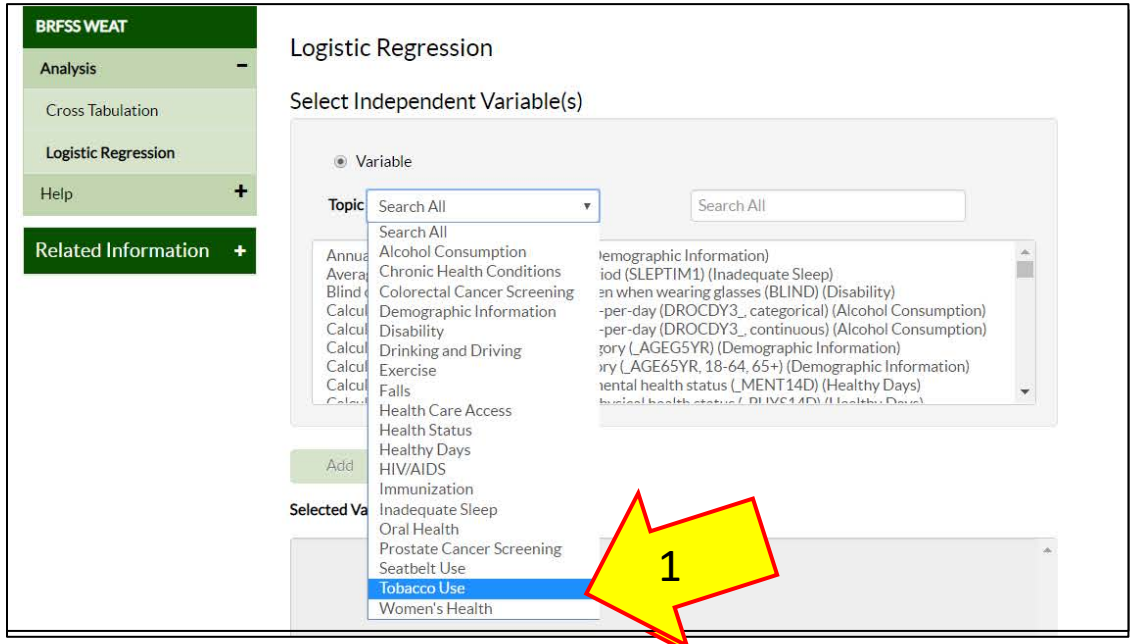
GREAT PLAINS IDeA | Clinical and Translational Research

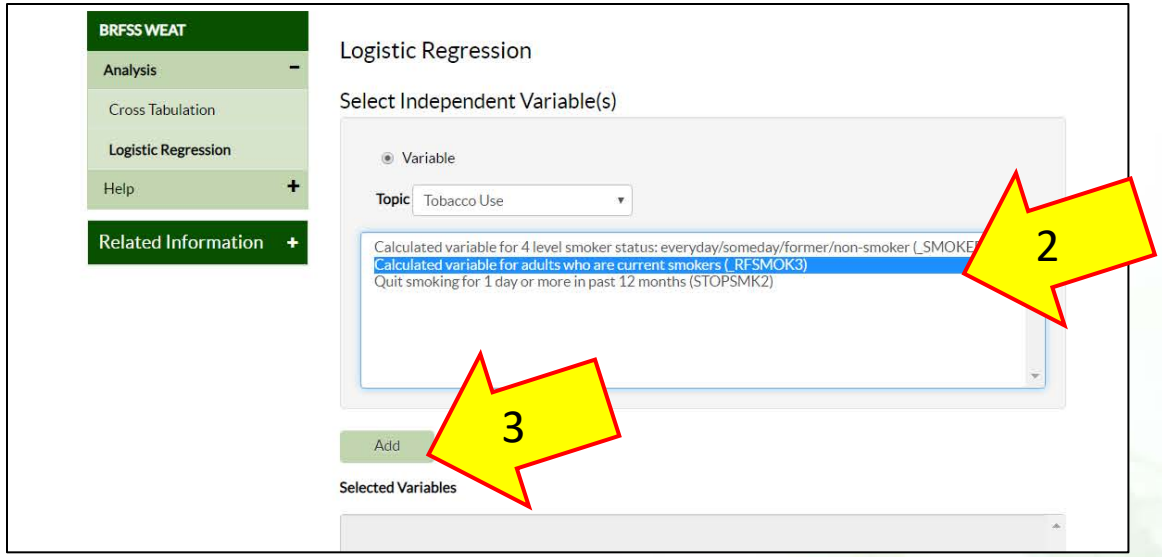


Click "Calculated variable for adults aged 18+ who have had permanent teeth extracted

Click "Next"

GREAT PLAINS IDeA | Clinical and Translational Research





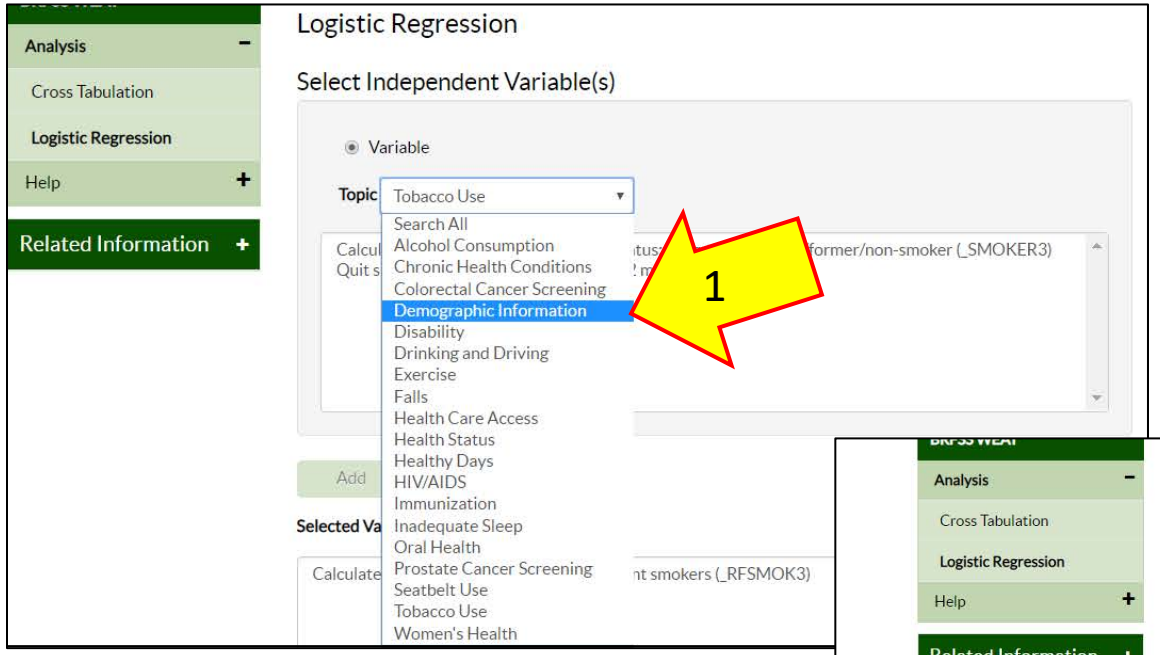
Our primary independent variable (exposure) is current tobacco use and we have age as another independent variable

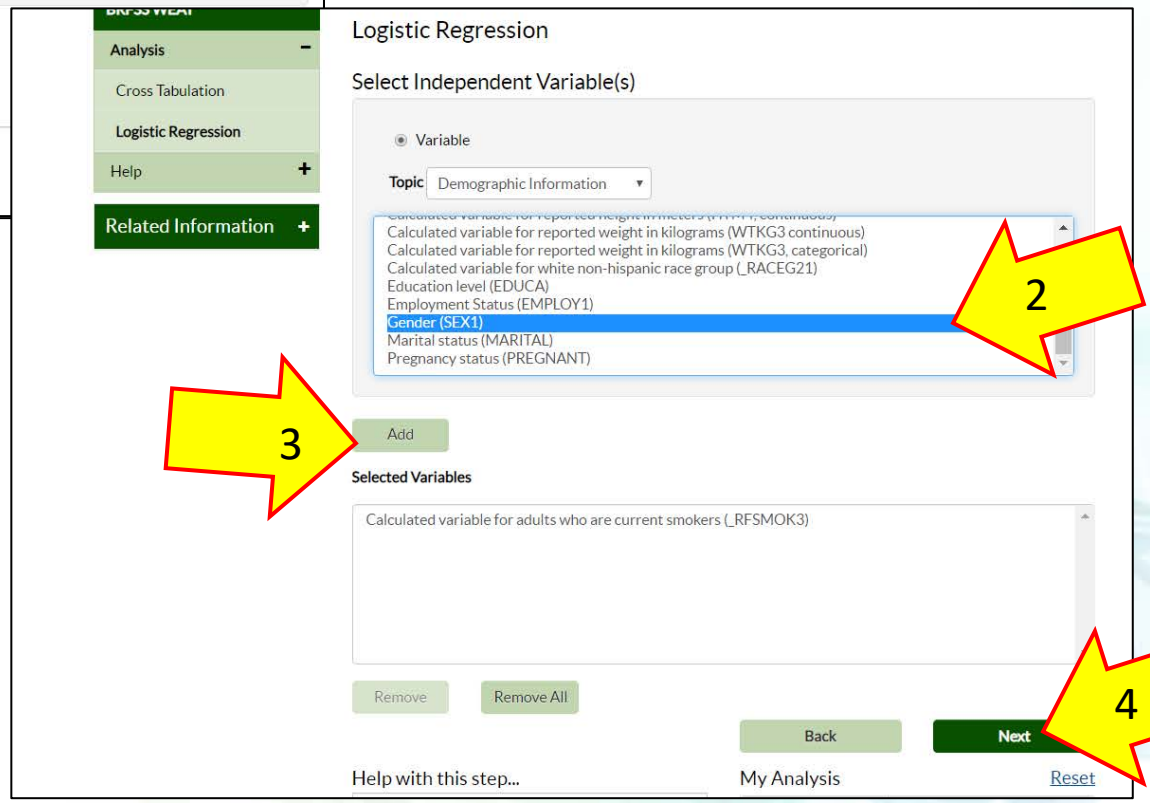
Drop down "Topic" and click "Tobacco Use"

Click "Calculated variable for adults who are current smokers"

Click "Add"

GREAT PLAINS IDeA | Clinical and Translational Research





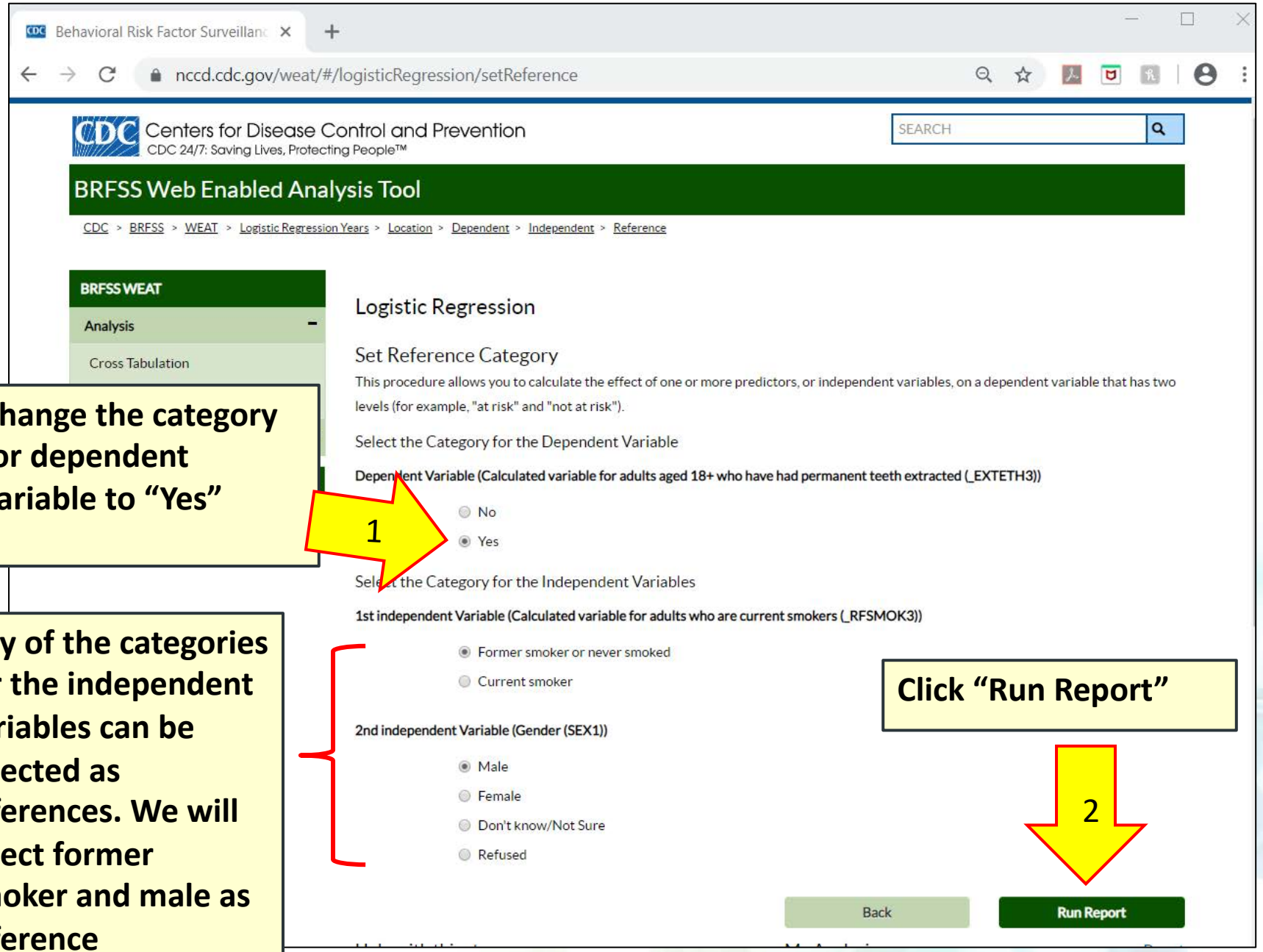
Drop down "Topic" and click "Demographic Information"

Click "Gender"

Click "Add"

Click "Next"

GREAT PLAINS IDeA | Clinical and Translational Research



Change the category for dependent variable to "Yes"

1

Any of the categories for the independent variables can be selected as references. We will select former smoker and male as reference

Click "Run Report"

2

GREAT PLAINS IDeA | Clinical and Translational Research

Download Results Reset

Behavioral Risk Factor Surveillance System

Logistic Regression, Nebraska, 2018

Dependent Variable:	Oral Health: Calculated variable for adults aged 18+ who have had permanent teeth extracted (_EXTETH3)
Independent Variables:	Tobacco Use: Calculated variable for adults who are current smokers (_RFSMOK3)
	Demographic Information: Gender (SEX1)

	Sample Size	Beta	Standard Error	t-statistic	p-value for t-statistic	Odds Ratio	95% CI for Odds Ratio
Calculated variable for adults aged 18+ who have had permanent teeth extracted (_EXTETH3), Modeling category: Yes							
Intercept	13,945	-0.62	0.04	-15.52	<0.0001	0.54	(0.50 - 0.58)
Calculated variable for adults who are current smokers (_RFSMOK3)							
Former smoker or never smoked (REF)	12,061	0.00	0.00	.	.	1.00	(1.00 - 1.00)
Current smoker	1,884	0.68	0.07	9.30	<0.0001	1.98	(1.71 - 2.28)
Gender (SEX1)							
Male (REF)	6,258	0.00	0.00	.	.	1.00	(1.00 - 1.00)
Female	7,677	0.01	0.05	0.14	0.8921	1.01	(0.91 - 1.11)
Don't know/Not Sure	5	2.15	1.08	1.99	0.0465	8.62	(1.03 - 71.92)
Refused	5	0.06	1.25	0.05	0.9608	1.06	(0.09 - 12.25)

Model test statistics

Contrast	Degrees of Freedom	Wald chi-square	p-value for Wald chi-square
Overall Model	5	499.18	<0.0001
Model minus intercept	4	90.59	<0.0001
Intercept	.	.	.
Calculated variable for adults who are current smokers (_RFSMOK3)	1	86.48	<0.0001
Gender (SEX1)	3	3.97	<0.0001

The value of Odds Ratio against "Current smoker" is 1.98 and the p-value is <0.0001.

The odds of permanent teeth extraction is 1.98 times higher for current smokers compared to former/never smokers, after adjusting for gender.

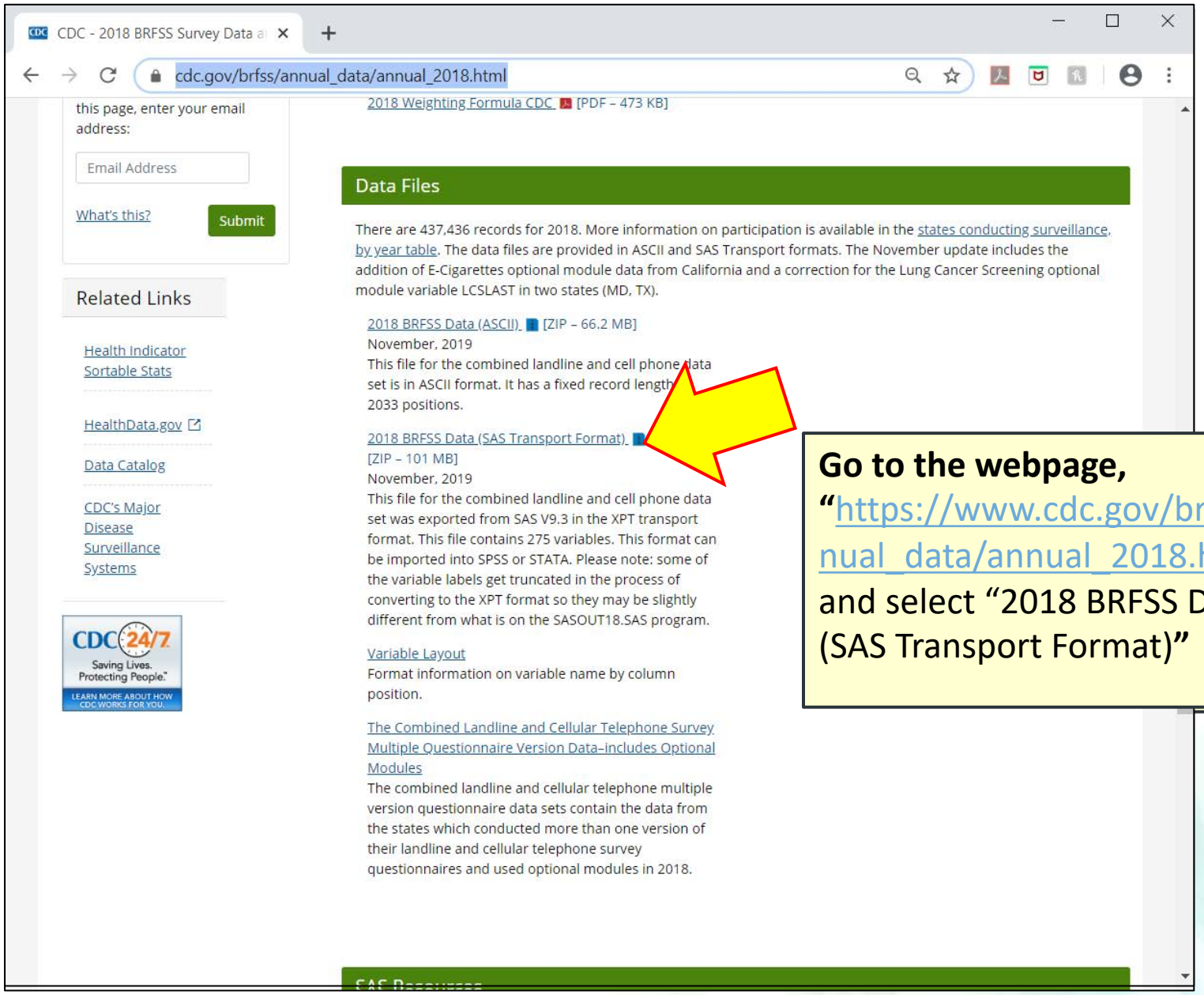
As p-value is less than 0.05, the association is statistically significant.

The odds ratio against "Female" is 1.01 and the p-value is also insignificant (0.8921). So there is no significant difference in permanent tooth extraction for females compared to males after adjusting for smoking status.

As the sample sizes for gender categories of "Don't know/Not sure" and "Refused" are very small, the results associated with them are not precise. This is evident from the very broad confidence intervals

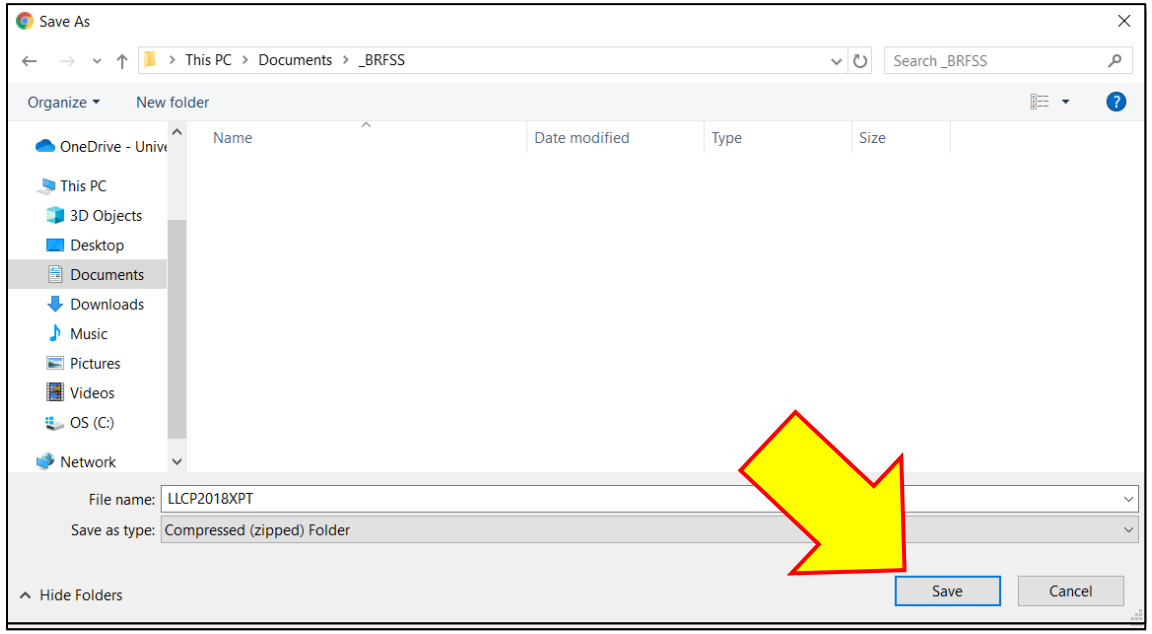
GREAT PLAINS IDeA | Clinical and Translational Research

DOWNLOADING BRFSS DATA

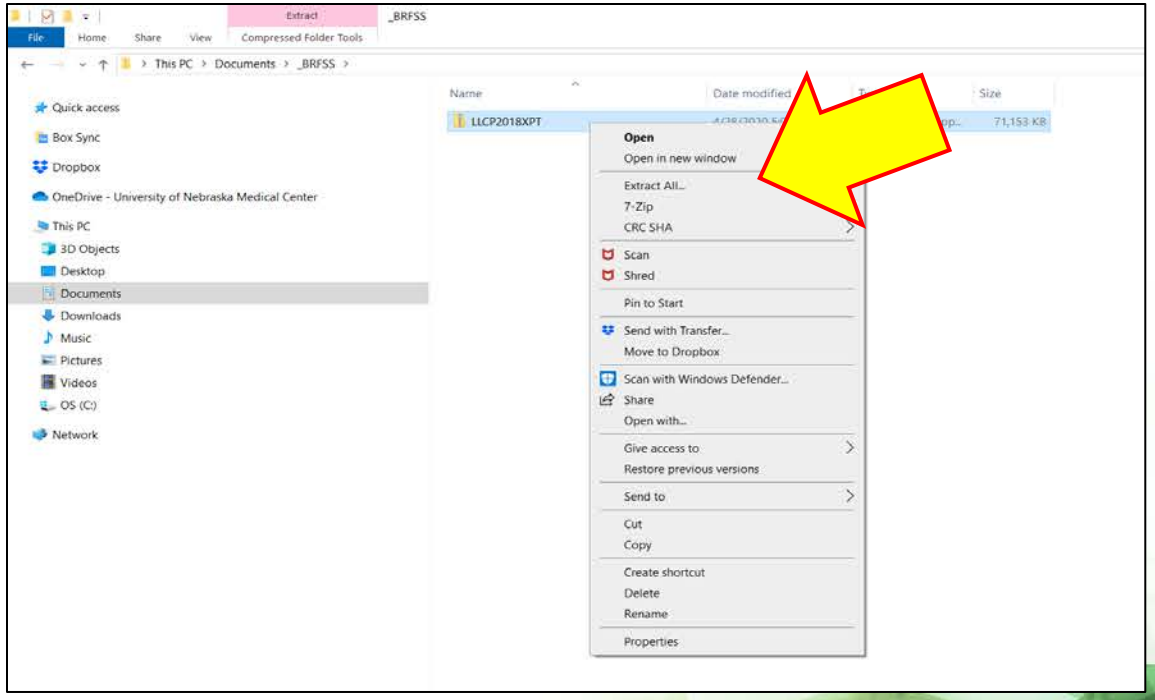


Go to the webpage, "https://www.cdc.gov/brfss/annual_data/annual_2018.html" and select "2018 BRFSS Data (SAS Transport Format)"

GREAT PLAINS IDeA | Clinical and Translational Research

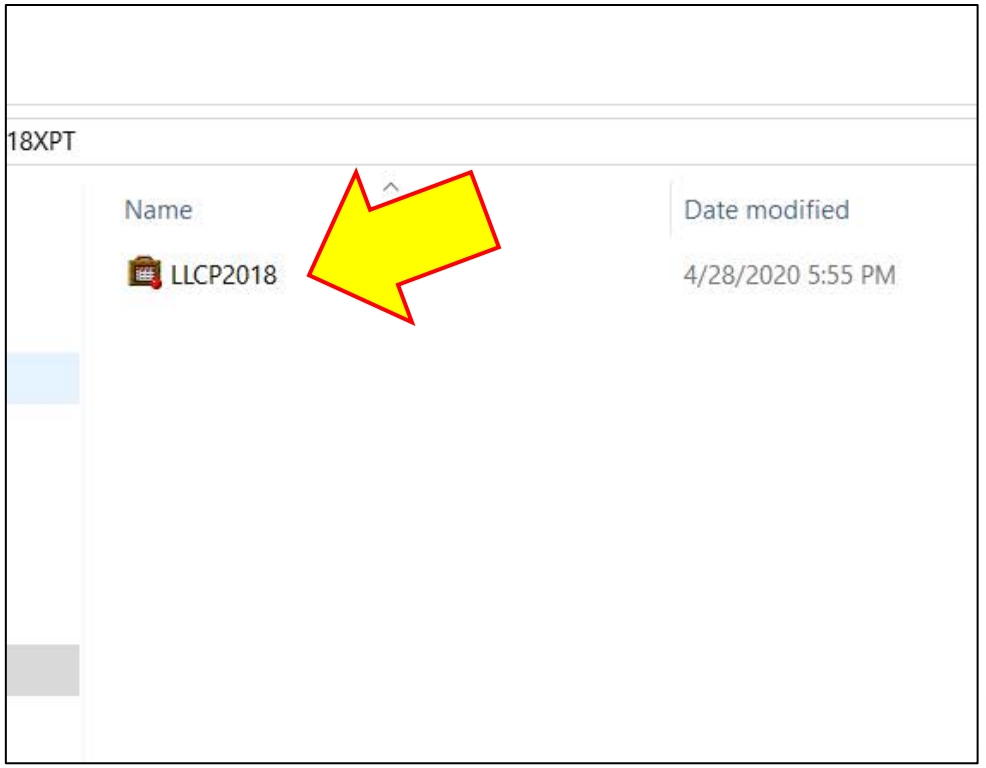


Select destination folder and save the zip file



Right click the zip folder and Extract All

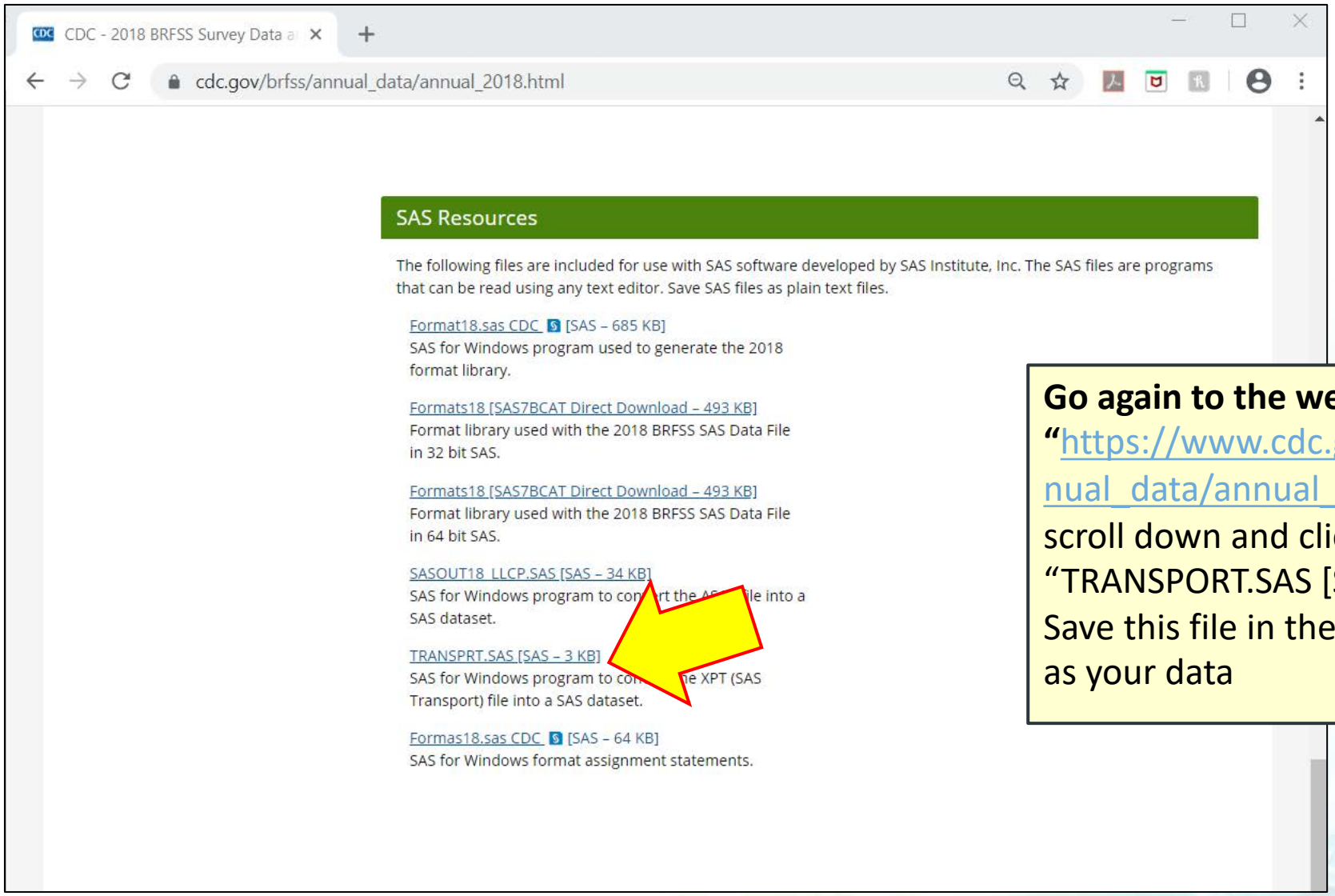
GREAT PLAINS
IDeA | Clinical and Translational Research



Within the extracted folder you will have this SAS export transport file

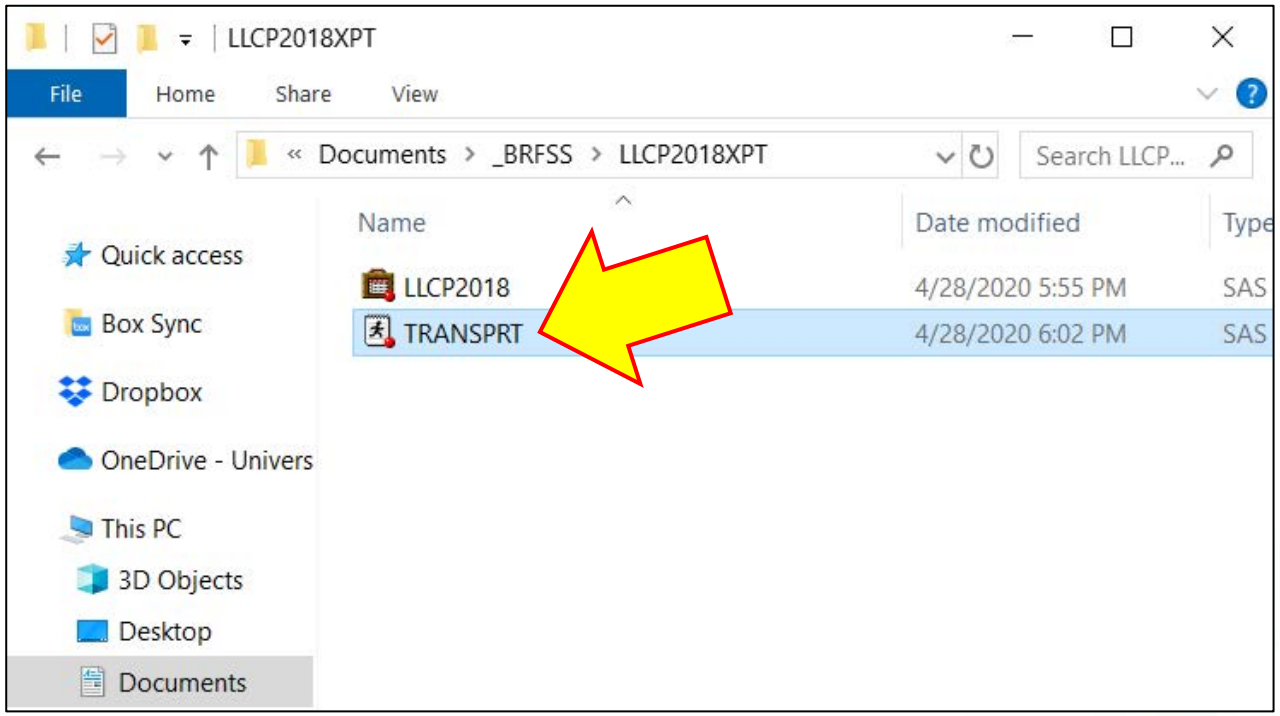
GREAT PLAINS IDeA | Clinical and Translational Research

GETTING STARTED WITH SAS



Go again to the webpage, "https://www.cdc.gov/brfss/annual_data/annual_2018.html**" scroll down and click, "TRANSPORT.SAS [SAS – 3 KB]". Save this file in the same folder as your data**

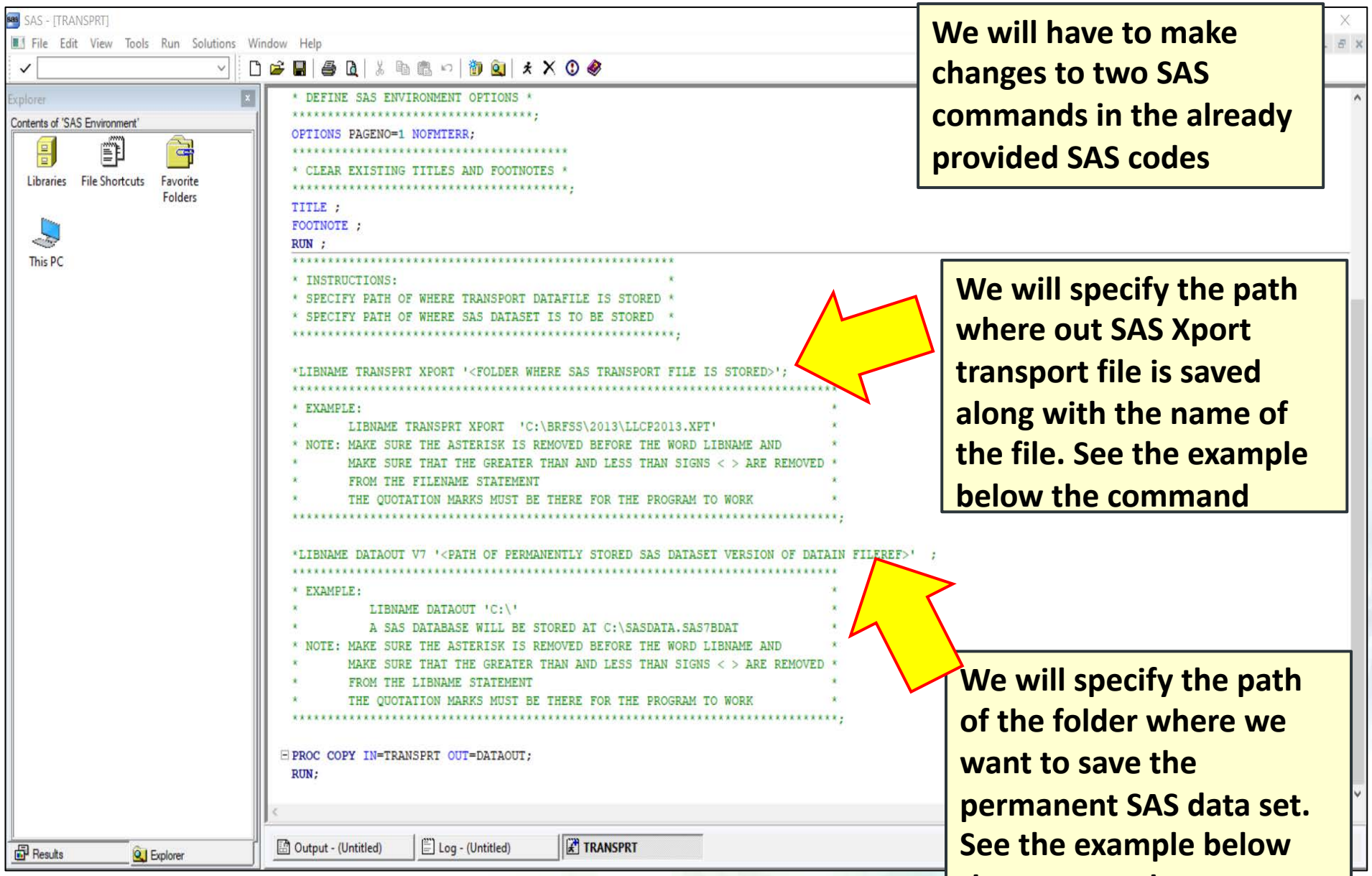
GREAT PLAINS IDeA | Clinical and Translational Research



First, we will convert the SAS Xport transport file to permanent SAS data set

Double click on the SAS system program, "TRANSPRT" that we downloaded from the BRFSS website. This will open the SAS program

GREAT PLAINS IDeA | Clinical and Translational Research

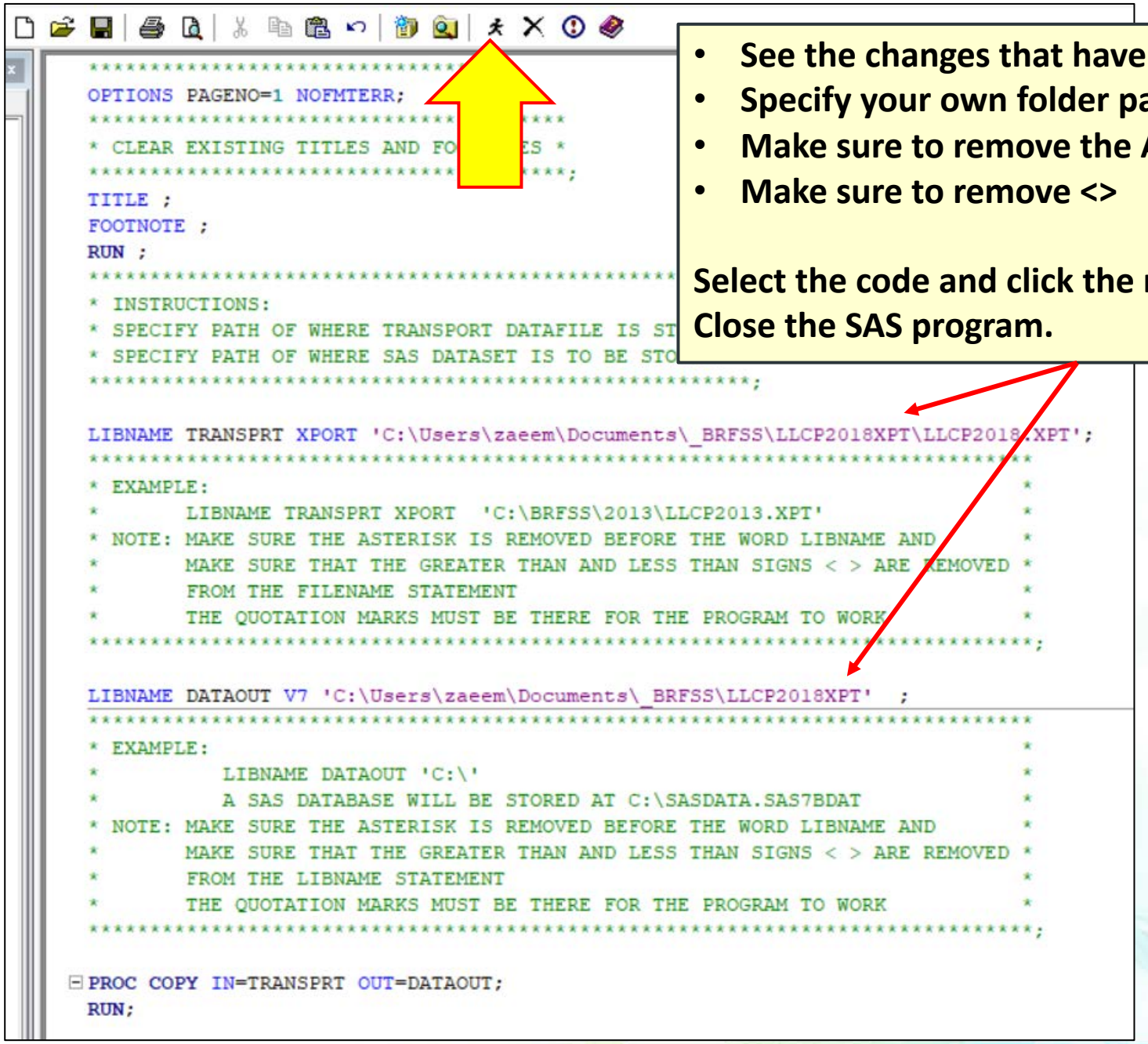


We will have to make changes to two SAS commands in the already provided SAS codes

We will specify the path where out SAS Xport transport file is saved along with the name of the file. See the example below the command

We will specify the path of the folder where we want to save the permanent SAS data set. See the example below the command

GREAT PLAINS
IDeA | Clinical and Translational Research

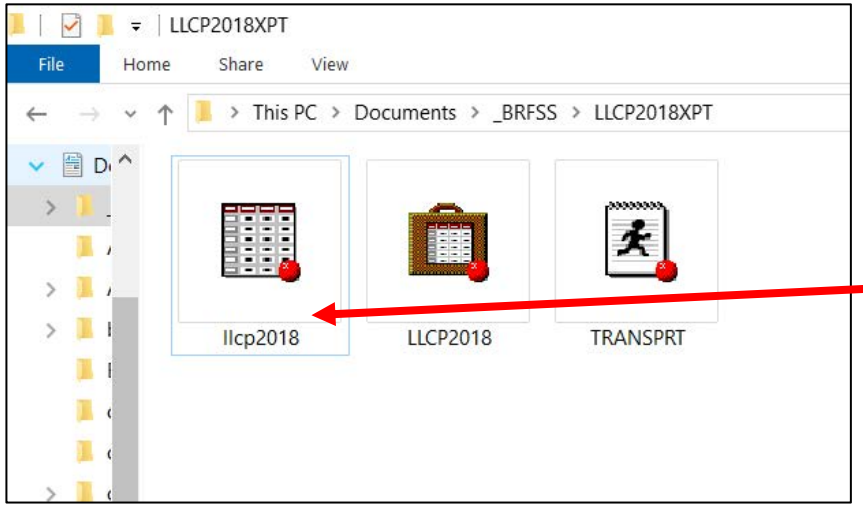


```
OPTIONS PAGENO=1 NOFMTERR;
* CLEAR EXISTING TITLES AND FO    ES *
TITLE ;
FOOTNOTE ;
RUN ;
* INSTRUCTIONS:
* SPECIFY PATH OF WHERE TRANSPORT DATAFILE IS ST
* SPECIFY PATH OF WHERE SAS DATASET IS TO BE STO

LIBNAME TRANSPRT XPORT 'C:\Users\zaeem\Documents\_BRFSS\LLCP2018XPT\LLCP2018.XPT';
* EXAMPLE:                                                                    *
*       LIBNAME TRANSPRT XPORT  'C:\BRFSS\2013\LLCP2013.XPT'                  *
* NOTE: MAKE SURE THE ASTERISK IS REMOVED BEFORE THE WORD LIBNAME AND         *
*       MAKE SURE THAT THE GREATER THAN AND LESS THAN SIGNS < > ARE REMOVED   *
*       FROM THE FILENAME STATEMENT                                           *
*       THE QUOTATION MARKS MUST BE THERE FOR THE PROGRAM TO WORK             *

LIBNAME DATAOUT V7 'C:\Users\zaeem\Documents\_BRFSS\LLCP2018XPT'  ;
* EXAMPLE:                                                                    *
*          LIBNAME DATAOUT 'C:\'                                              *
*          A SAS DATABASE WILL BE STORED AT C:\SASDATA.SAS7BDAT               *
* NOTE: MAKE SURE THE ASTERISK IS REMOVED BEFORE THE WORD LIBNAME AND         *
*       MAKE SURE THAT THE GREATER THAN AND LESS THAN SIGNS < > ARE REMOVED   *
*       FROM THE LIBNAME STATEMENT                                            *
*       THE QUOTATION MARKS MUST BE THERE FOR THE PROGRAM TO WORK             *

PROC COPY IN=TRANSPRT OUT=DATAOUT;
RUN;
```

- See the changes that have been made
- Specify your own folder pathways
- Make sure to remove the Asterix
- Make sure to remove <>

Select the code and click the running man.
Close the SAS program.

GREAT PLAINS
IDeA | Clinical and
Translational Research



Permanent SAS dataset has been saved to the folder

Specify your own path

```
LIBNAME FILE "C:\Users\zaeem\Documents\_BRFSS\LLCP2018XPT";
PROC CONTENTS DATA = FILE.llcp2018 VARNUM;
RUN;
```

Name of the permanent dataset

Open SAS
Use this program to open the permanent data set

GREAT PLAINS IDeA | Clinical and Translational Research

SAS CODE FOR CROSS-TABULATION

```
*READ THE PERMANENT SAS DATASET INTO SAS;

LIBNAME FILE "C:\Users\zaeem\Documents\_BRFSS\LLCP2018XPT";

*PROC CONTENTS SHOWS ALL THE VARIABLES IN THE DATASET;

PROC CONTENTS DATA = FILE.llcp2018 VARNUM;

RUN;

*CREATING A NEW TEMPORARY DATASET FROM PERMANENT DATA SET AND RECODING TO ASSIGN MISSING
VALUES TO MATCH THE RESULTS FROM WEAT TOOL;

DATA BRFSS;

SET FILE.llcp2018;

IF SEX1=9 THEN SEX1=.;

IF _RFSMOK3=9 THEN _RFSMOK3=.;

IF _EXTETH3 = 9 THEN _EXTETH3=.;RUN;

*CROSS TABULATION BETWEEN CURRENT SMOKING AND PERMANENT TEETH EXTRACTION BY GENDER;

PROC SURVEYFREQ DATA = BRFSS;

WHERE _STATE = 31; *SPECIFIES STATE = NEBRASKA (31 IS THE FIPS CODE);

CLUSTER _PSU; *SPECIFIES CLUSTER VARIABLE;

STRATA _ststr; *SPECIFIES THE STRATUM VARIABE;

WEIGHT _llcpwt; *SPECIFIES THE WEIGHT VARIABLE;

TABLE SEX1* _RFSMOK3* _EXTETH3/CHISQ COL ROW;*PRODUCES THE CROSS-TABULATION BY SEX WITH
CHI SQUARE TEST. SEX1 IS SEX, _RFSMOK3 IS CURRENT SMOKING AND _EXTETH3 IS EXTRACTION OF
PERMANENT TEETH;

run;
```

GREAT PLAINS IDeA | Clinical and Translational Research

Table of _RFSMOK3 by _EXTETH3										
Controlling for SEX1=1										
_RFSMOK3	_EXTETH3	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent	Column Percent	Std Err of Col Percent
1	1	2978	368277	8818	53.4669	0.9332	64.7066	0.9432	86.1449	0.8941
	2	2338	200871	5977	29.1628	0.8070	35.2934	0.9432	76.8782	1.2229
	Total	5316	569148	9757	82.6297	0.7288	100.000			
2	1	394	59232	4071	8.5993	0.5706	49.5060	2.3216	13.8551	0.8941
	2	548	60414	3627	8.7710	0.5144	50.4940	2.3216	23.1218	1.2229
	Total	942	119645	5353	17.3703	0.7288	100.000			
Total	1	3372	427508	9350	62.0663	0.8835			100.000	
	2	2886	261285	6778	37.9337	0.8835			100.000	
	Total	6258	688794	10230	100.000					

Rao-Scott Chi-Square Test	
Pearson Chi-Square	88.1499
Design Correction	2.2828
Rao-Scott Chi-Square	38.6144
DF	1
Pr > ChiSq	<.0001
F Value	38.6144
Num DF	1
Den DF	13878
Pr > F	<.0001
Sample Size = 13940	

Here is one of the resulting tables. The estimates are for men (SEX=1)

We can compare these results to the results of the table produced by WEAT Tool as they are similar

We use proc surveyfreq instead of the traditional proc freq because it takes strata and clusters into account

The variable names and details can be looked up in the code book for BRFSS 2018 data

GREAT PLAINS IDeA | Clinical and Translational Research

SAS CODE FOR LOGISTIC REGRESSION

```
*READ THE PERMANENT SAS DATASET INTO SAS;

LIBNAME FILE "C:\Users\zaeem\Documents\_BRFSS\LLCP2018XPT";

*PROC CONTENTS SHOWS ALL THE VARIABLES IN THE DATASET;

PROC CONTENTS DATA = FILE.llcp2018 VARNUM;

RUN;

*CREATING A NEW TEMPORARY DATASET FROM PERMANENT DATA SET AND RECODING TO ASSIGN MISSING
VALUES TO MATCH THE RESULTS FROM WEAT TOOL;

DATA BRFSS;

SET FILE.llcp2018;

IF SEX1=9 THEN SEX1=.;

IF _RFSMOK3=9 THEN _RFSMOK3=.;

IF _EXTETH3 = 9 THEN _EXTETH3=.;RUN;

*LOGISTIC REGRESSION PERMANENT TEETH EXTRACTION AS DEPENDENT VARIABLE AND CURRENT
SMOKING AND GENDER AS INDEPENDENT VARIABLES;

PROC SURVEYLOGISTIC DATA= BRFSS;

WHERE _STATE = 31;

CLUSTER _PSU;

STRATA _STSTR;

WEIGHT _llcpwt;

CLASS _RFSMOK3 (REF="1") SEX1 (REF='1'); *SPECIFYING REFERENCE CATEGORIES. NO CURRENT
SMOKING AND MALES GENDER ARE REFERNCE CATEGORIES;

MODEL _EXTETH3 (EVENT='2')= _RFSMOK3 SEX1; *MODEL WITH TEETH EXTRACTION AS DEPENDENT
VARIABLE AND CURRENT SMOKING AND GENDER AS INDEPENDENT VARIABLES;

RUN;
```

GREAT PLAINS
IDeA | Clinical and Translational Research

Analysis of Maximum Likelihood Estimates					
Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		0.4448	0.3618	1.23	0.2190
_RFSMOK3	2	0.3410	0.0367	9.30	<.0001
SEX1	2	-0.7134	0.3615	-1.97	0.0484
SEX1	7	1.4339	0.7212	1.99	0.0468
NOTE: The degrees of freedom for the t tests is 13878.					

Odds Ratio Estimates			
Effect	Point Estimate	95% Confidence Limits	
_RFSMOK3 2 vs 1	1.978	1.713	2.284
SEX1 2 vs 1	1.007	0.910	1.114
SEX1 7 vs 1	8.622	1.034	71.928
NOTE: The degrees of freedom in computing the confidence limits is 13878.			

The results are shown to the left.

We can compare these results to the results of the logistic regression produced by WEAT Tool as they are similar.

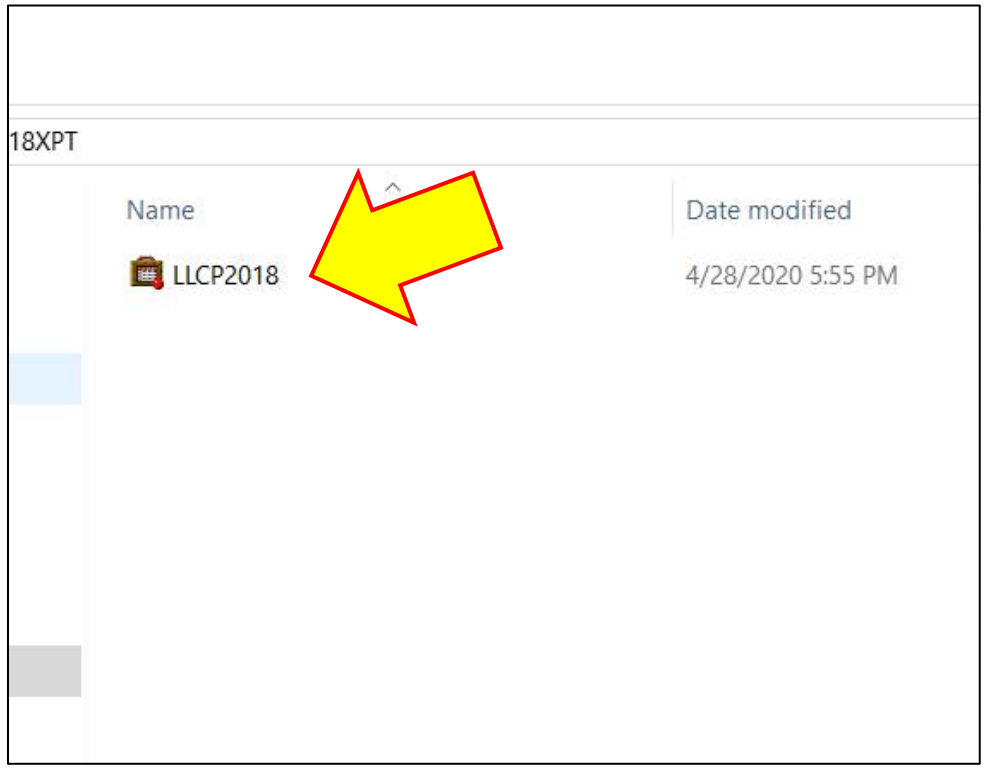
P value for gender is different because we removed one category of gender that was included by the WEAT tool (SEX1=9)

We use proc surveylogistic instead of the traditional proc logistic because it takes strata and clusters into account

Similarly, we use proc surveyreg for linear regression

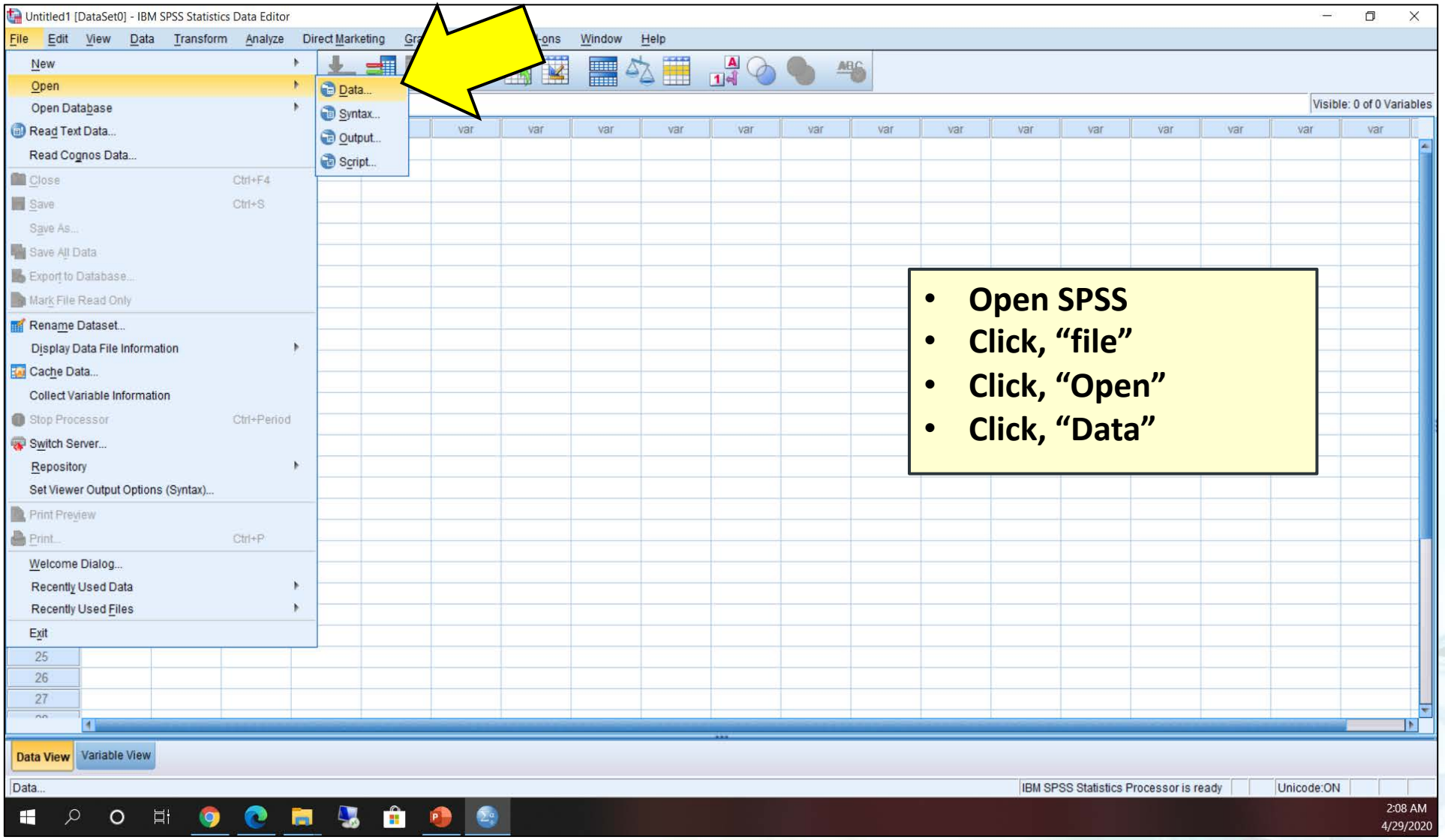
GREAT PLAINS IDeA | Clinical and Translational Research

GETTING STARTED WITH SPSS



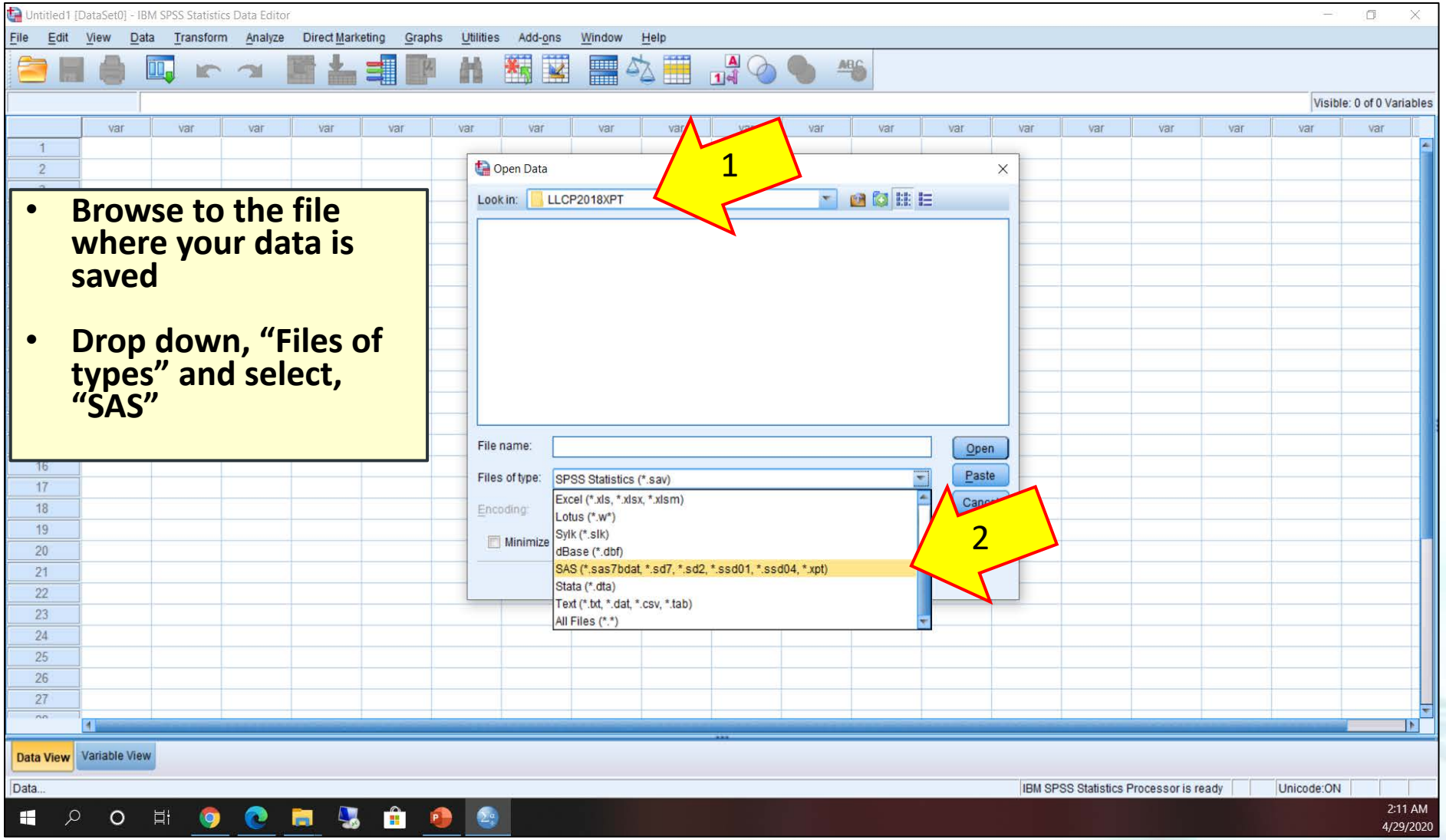
Please review the section, "Downloading BRFFS DATA" starting sliding number 55 as we will start with opening this data set into SPSS

GREAT PLAINS
IDeA | Clinical and Translational Research



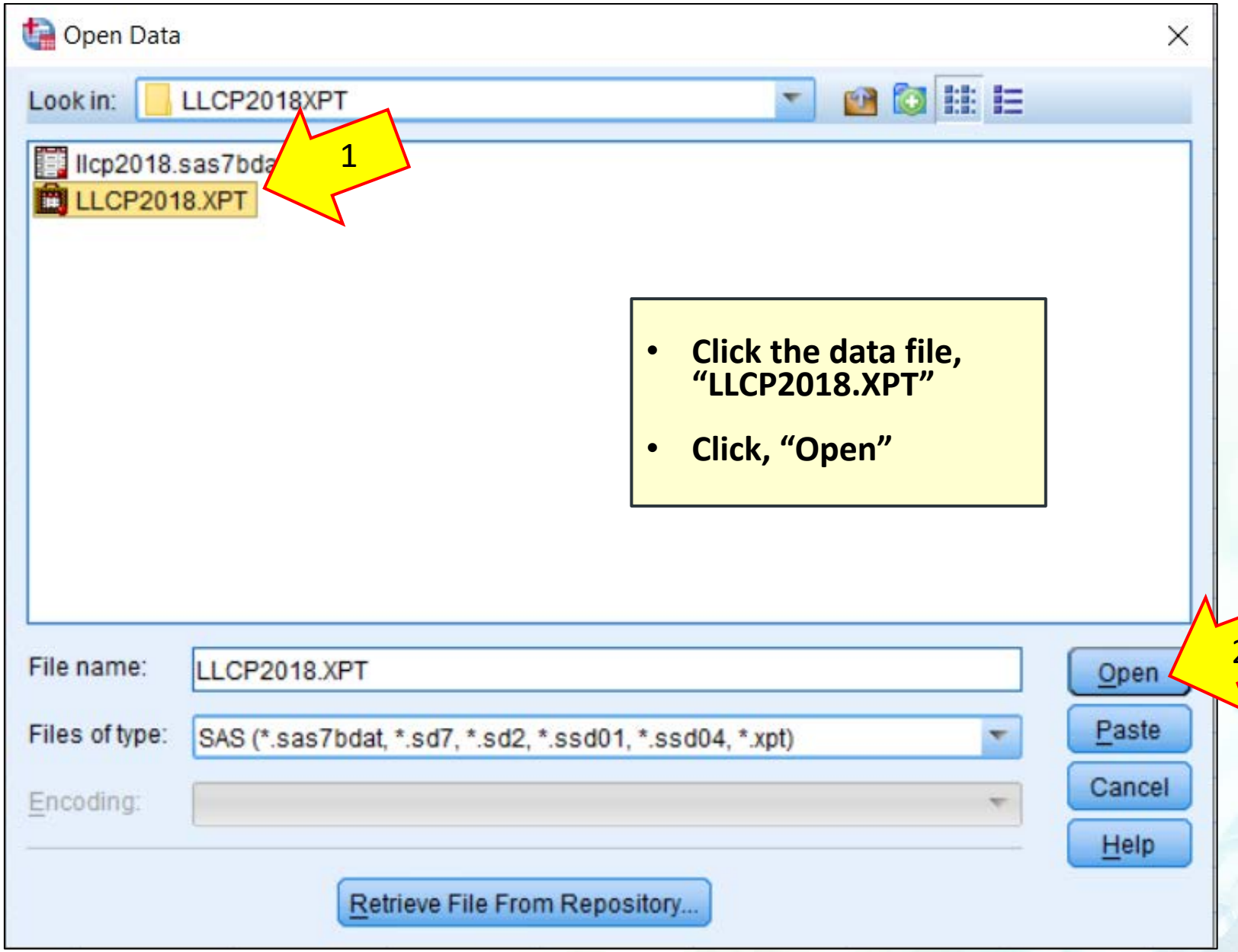
- **Open SPSS**
- **Click, “file”**
- **Click, “Open”**
- **Click, “Data”**

GREAT PLAINS
IDeA | Clinical and Translational Research



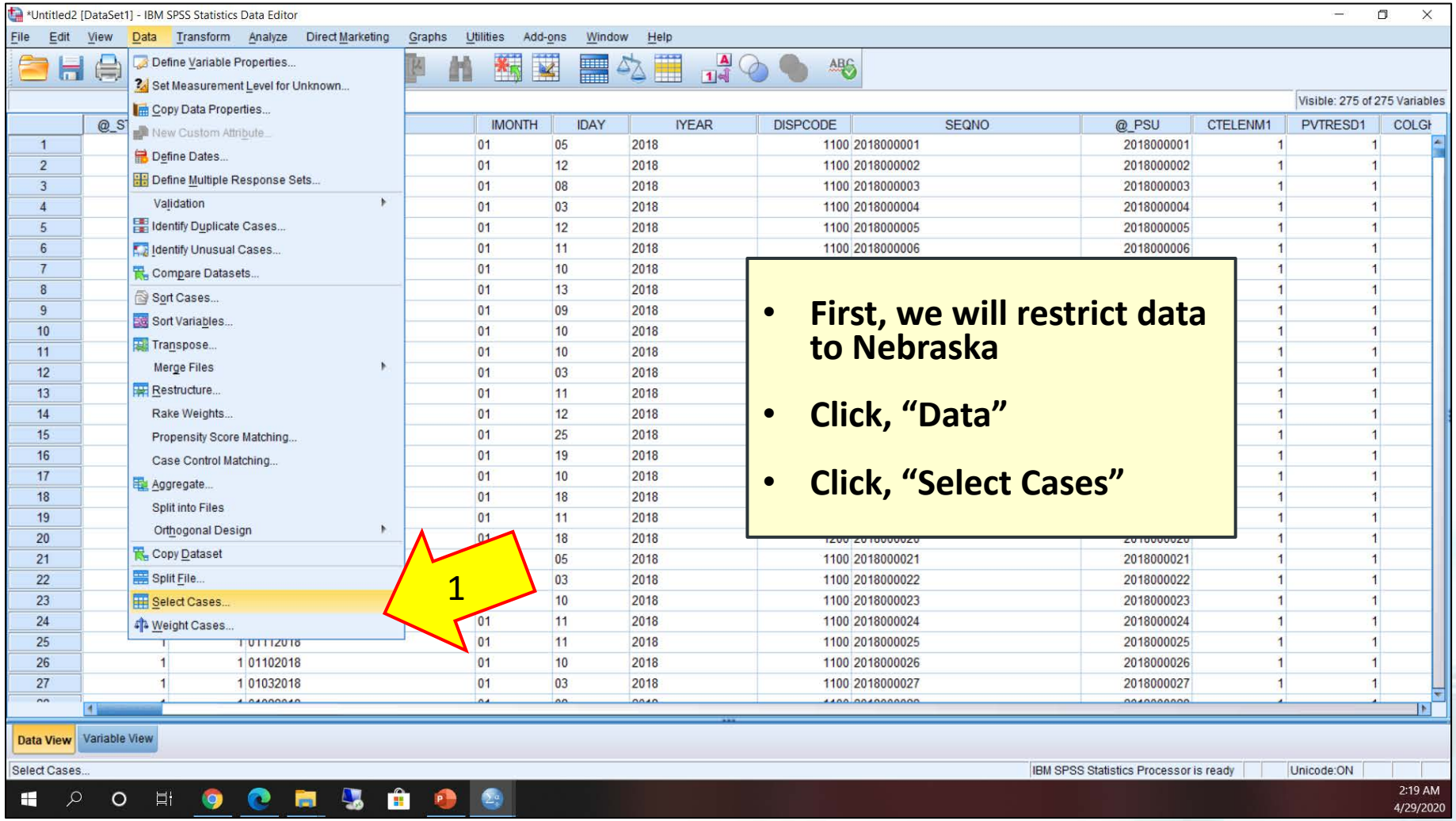
- **Browse to the file where your data is saved**
- **Drop down, "Files of types" and select, "SAS"**

GREAT PLAINS IDeA | Clinical and Translational Research



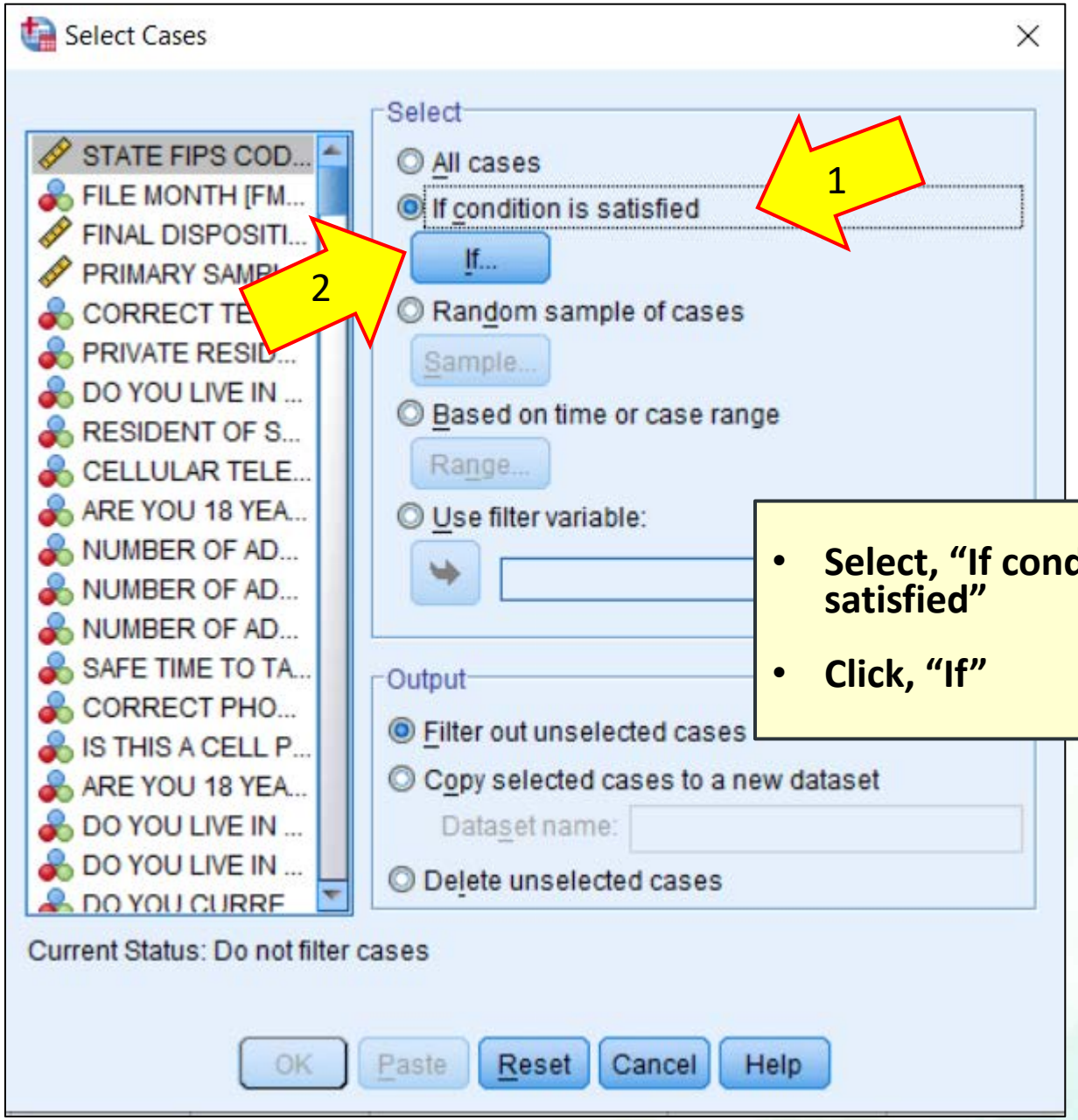
- **Click the data file, "LLCP2018.XPT"**
- **Click, "Open"**

GREAT PLAINS IDeA | Clinical and Translational Research



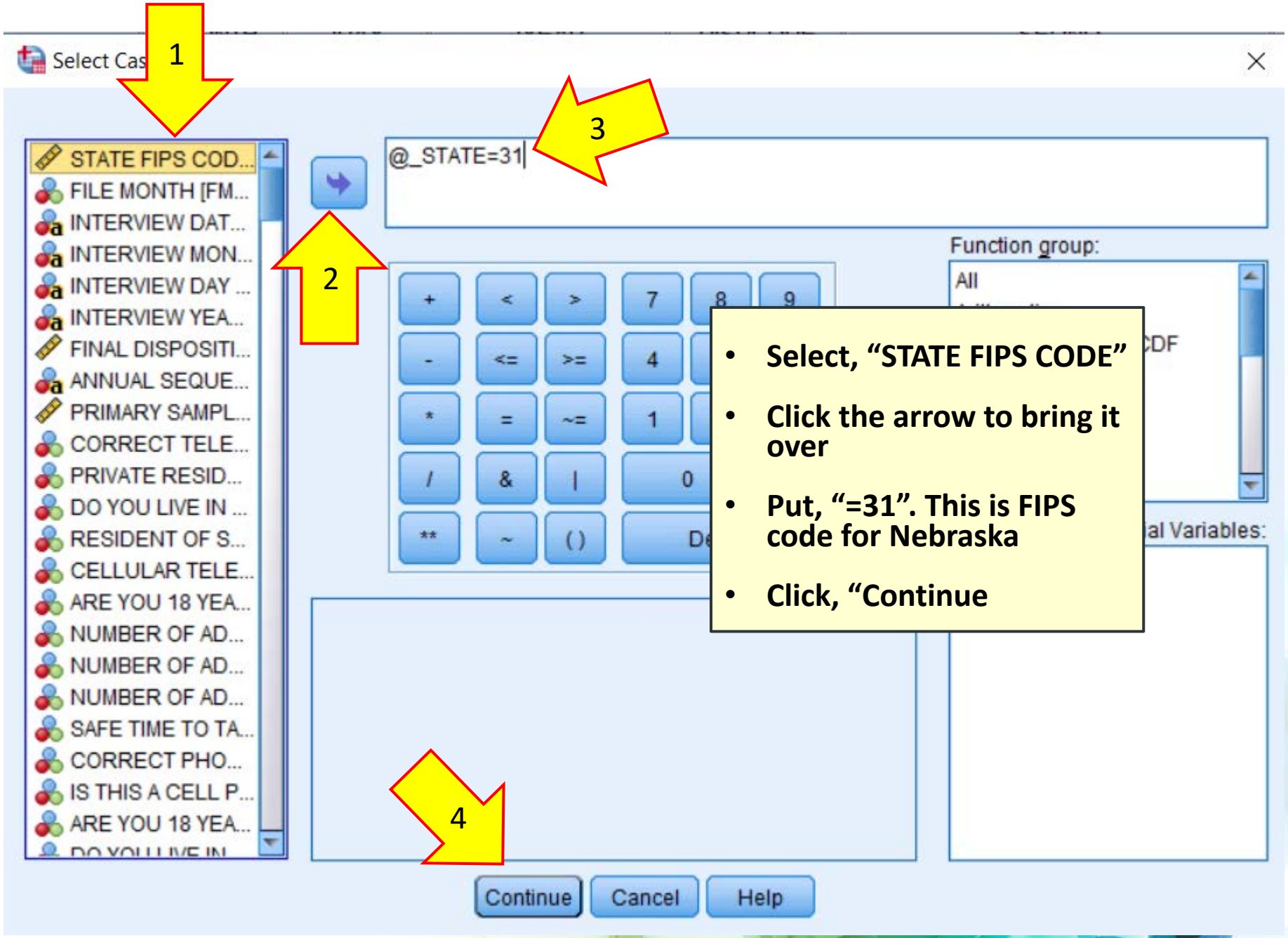
- **First, we will restrict data to Nebraska**
- **Click, “Data”**
- **Click, “Select Cases”**

GREAT PLAINS
IDeA | Clinical and Translational Research



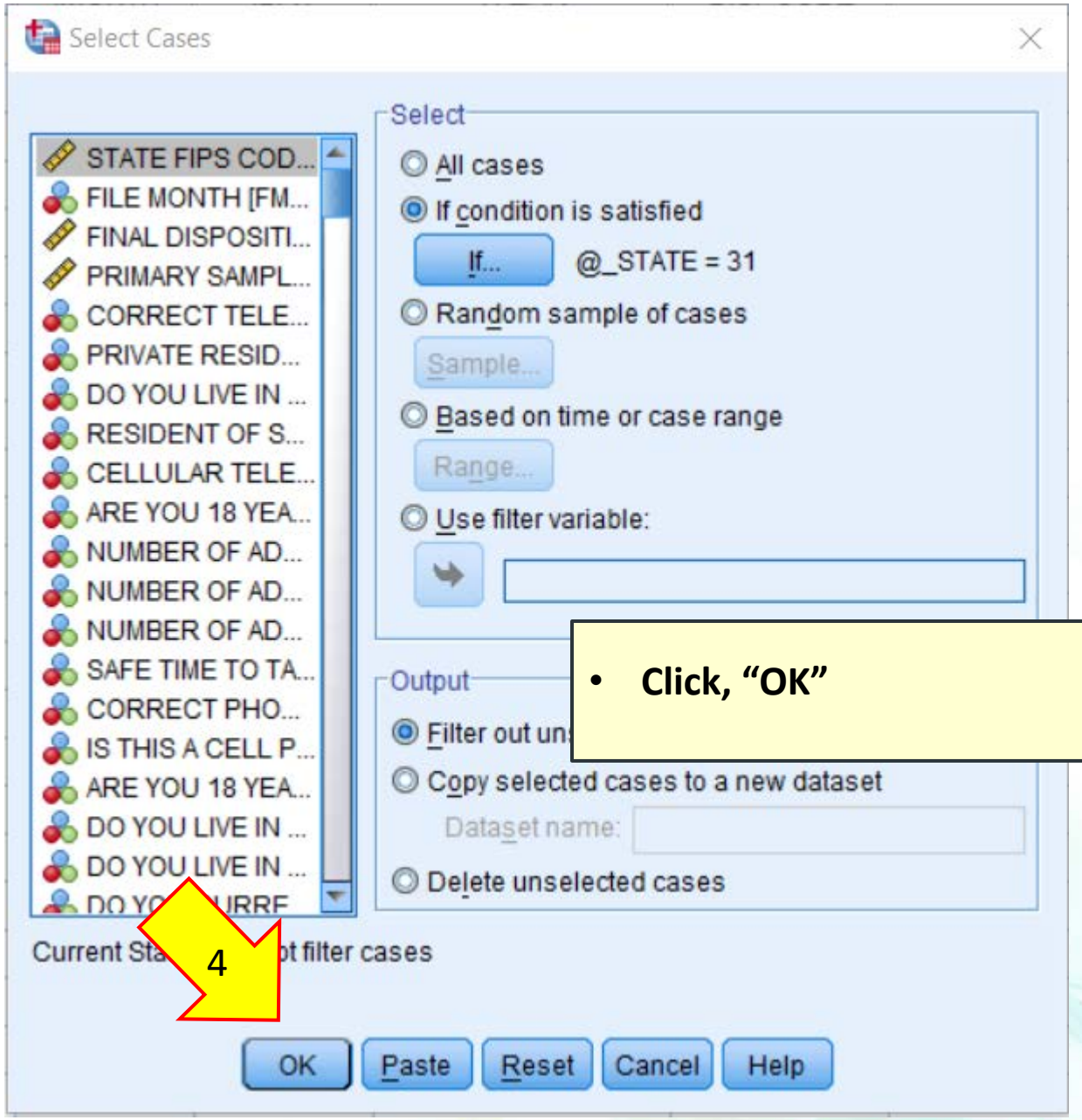
- **Select, “If condition is satisfied”**
- **Click, “If”**

GREAT PLAINS IDeA | Clinical and Translational Research

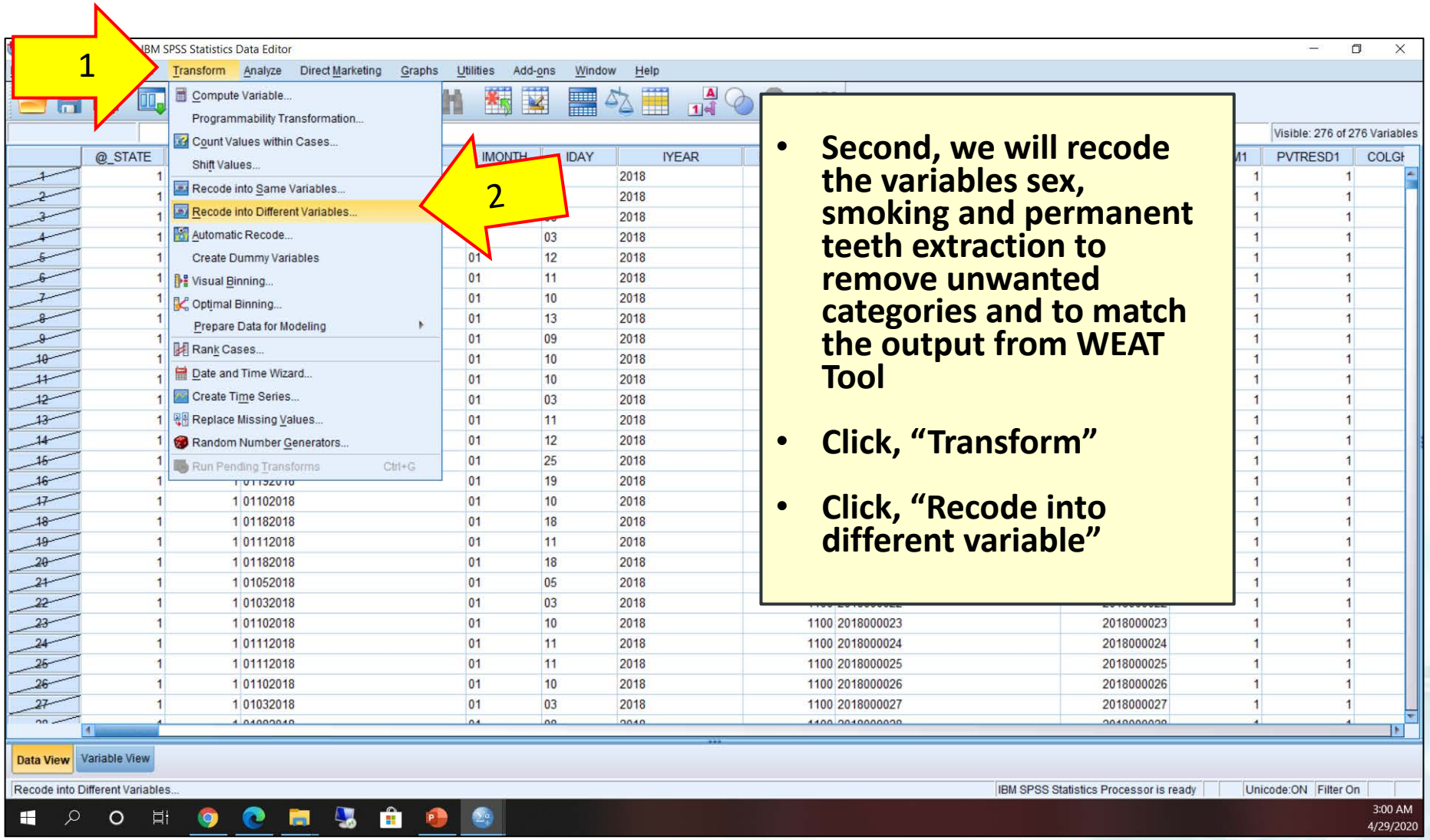


- **Select, "STATE FIPS CODE"**
- **Click the arrow to bring it over**
- **Put, "=31". This is FIPS code for Nebraska**
- **Click, "Continue**

GREAT PLAINS IDeA | Clinical and Translational Research

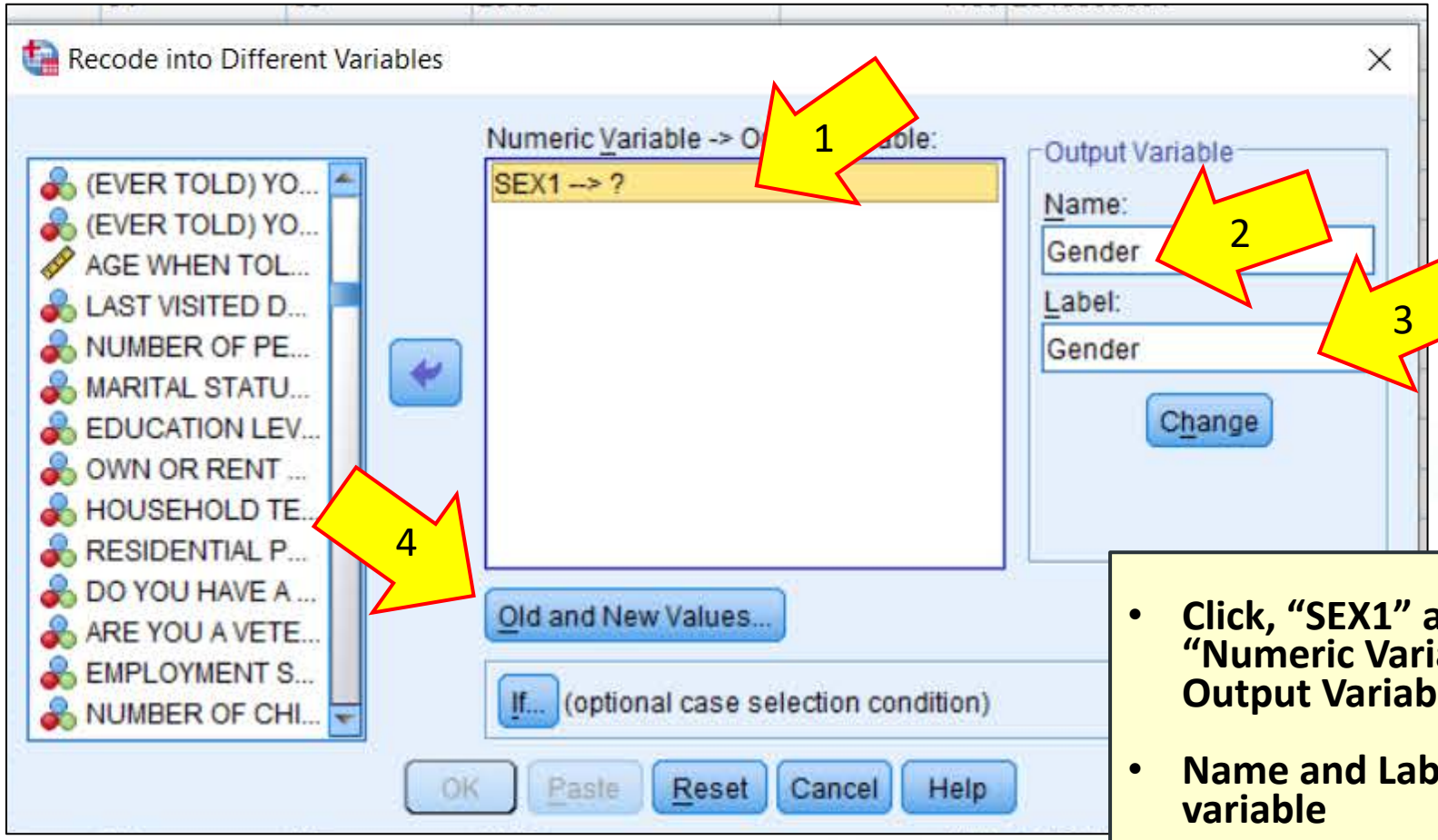

Select Cases
STATE FIPS COD...
FILE MONTH [FM...
FINAL DISPOSITI...
PRIMARY SAMPL...
CORRECT TELE...
PRIVATE RESID...
DO YOU LIVE IN ...
RESIDENT OF S...
CELLULAR TELE...
ARE YOU 18 YEA...
NUMBER OF AD...
NUMBER OF AD...
NUMBER OF AD...
SAFE TIME TO TA...
CORRECT PHO...
IS THIS A CELL P...
ARE YOU 18 YEA...
DO YOU LIVE IN ...
DO YOU LIVE IN ...
Select
All cases
If condition is satisfied
If...
@_STATE = 31
Random sample of cases
Sample...
Based on time or case range
Range...
Use filter variable:
Output
Filter out un
Copy selected cases to a new dataset
Dataset name:
Delete unselected cases
Current Sta
t filter cases
4
OK
Paste
Reset
Cancel
Help

- **Click, "OK"**



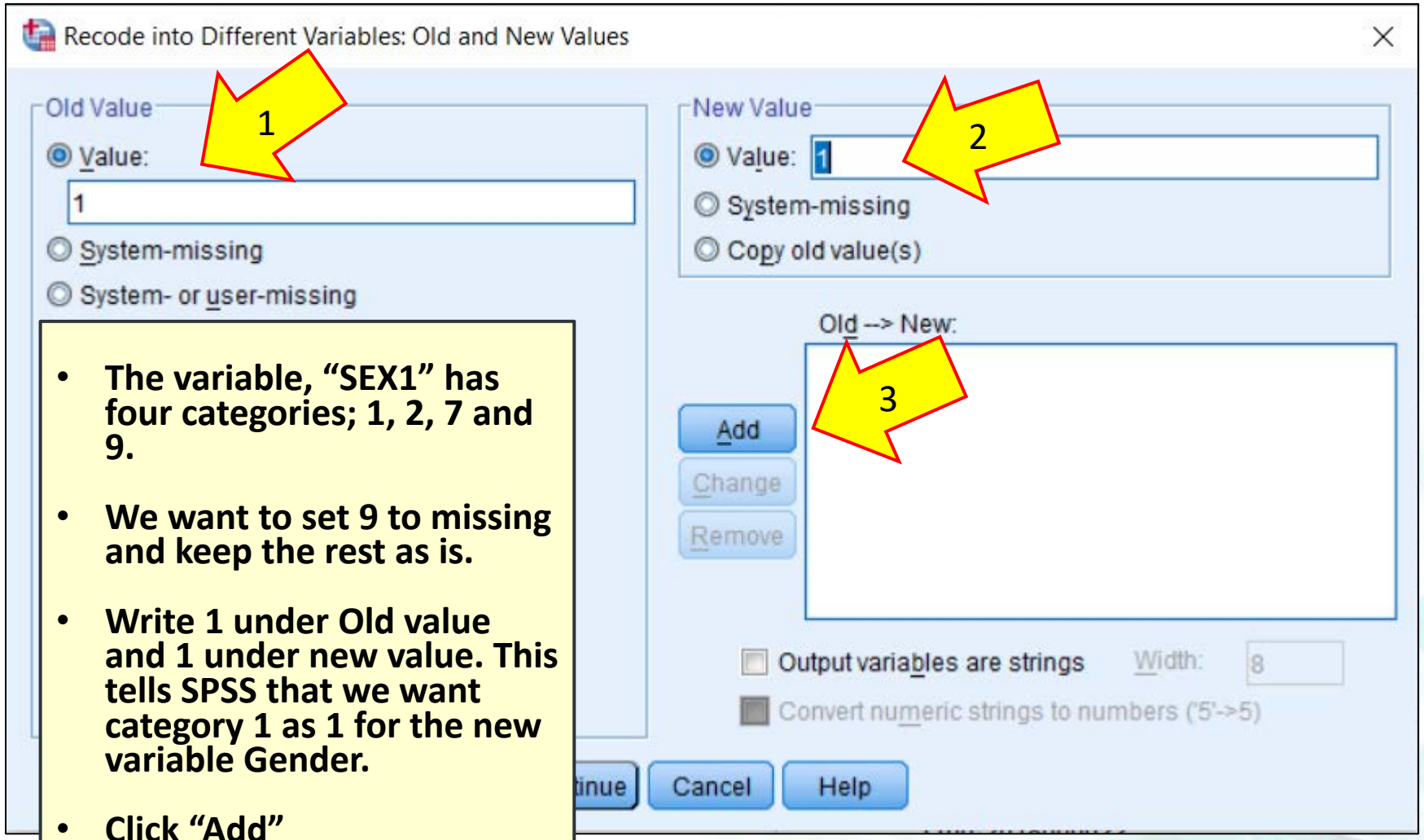
- **Second, we will recode the variables sex, smoking and permanent teeth extraction to remove unwanted categories and to match the output from WEAT Tool**
- **Click, “Transform”**
- **Click, “Recode into different variable”**

GREAT PLAINS IDeA | Clinical and Translational Research



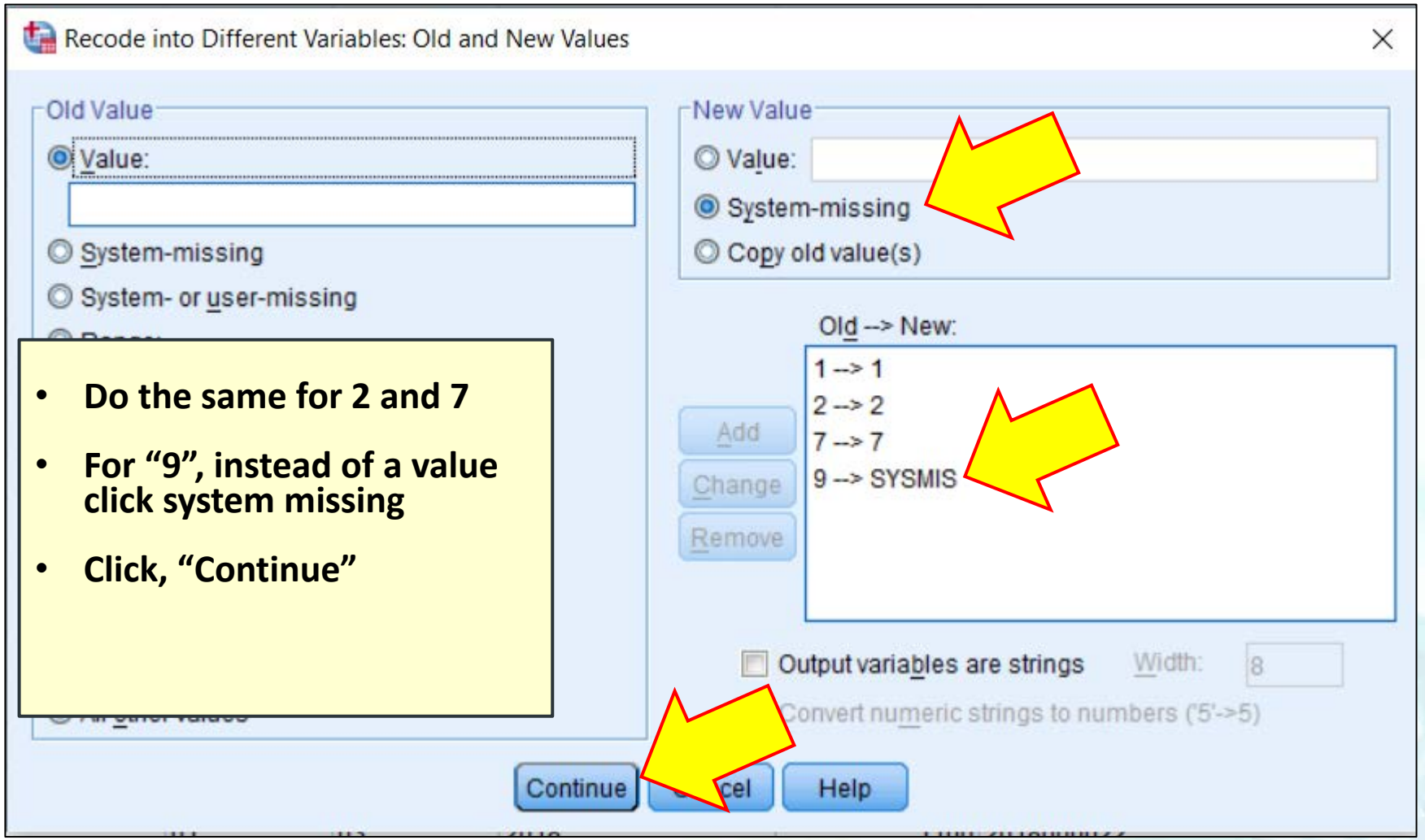
- **Click, "SEX1" and bring it "Numeric Variable→ Output Variable"**
- **Name and Label the variable**
- **Click, "Old and New Values"**

GREAT PLAINS IDeA | Clinical and Translational Research



- **The variable, "SEX1" has four categories; 1, 2, 7 and 9.**
- **We want to set 9 to missing and keep the rest as is.**
- **Write 1 under Old value and 1 under new value. This tells SPSS that we want category 1 as 1 for the new variable Gender.**
- **Click "Add"**

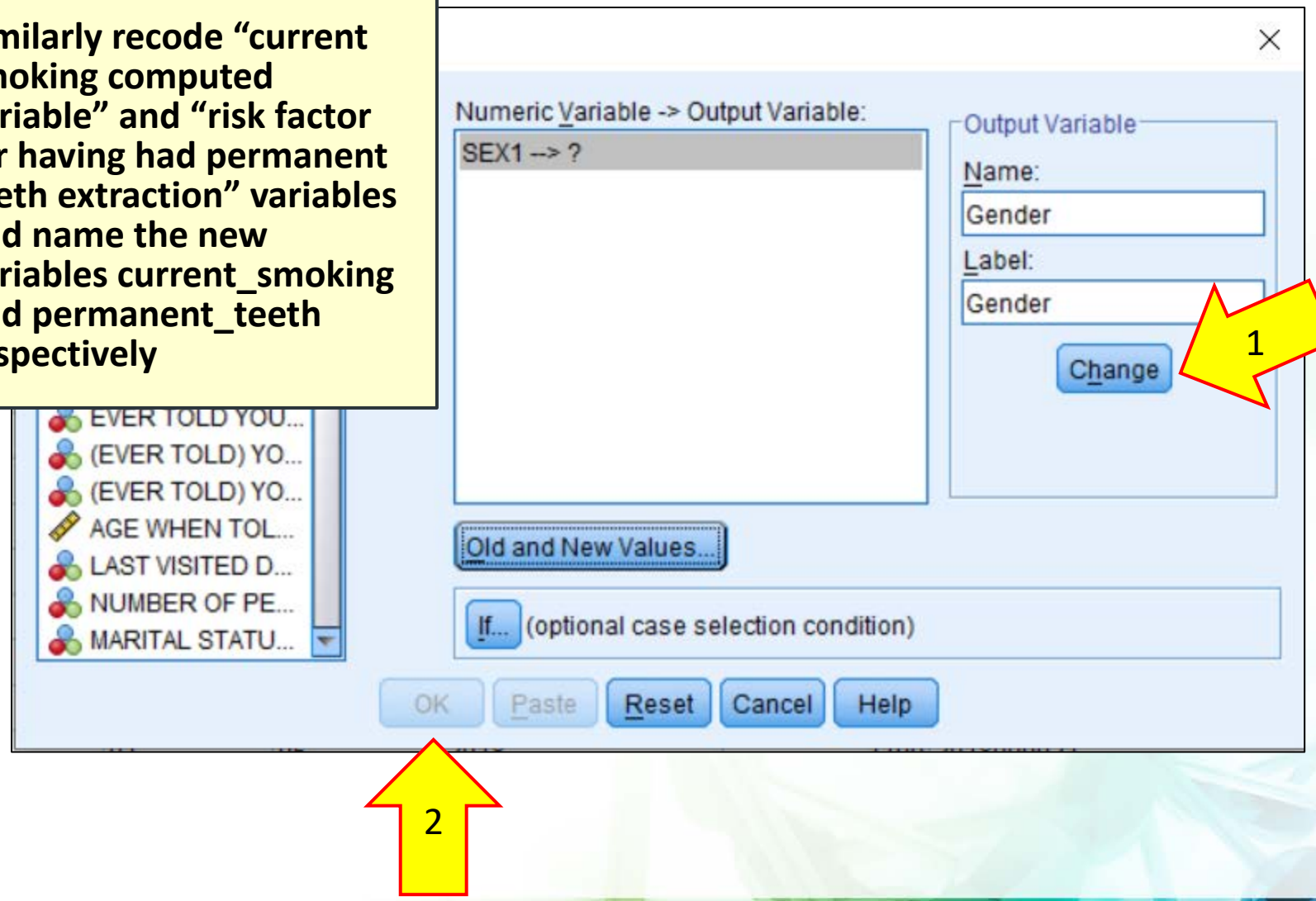
GREAT PLAINS IDeA | Clinical and Translational Research

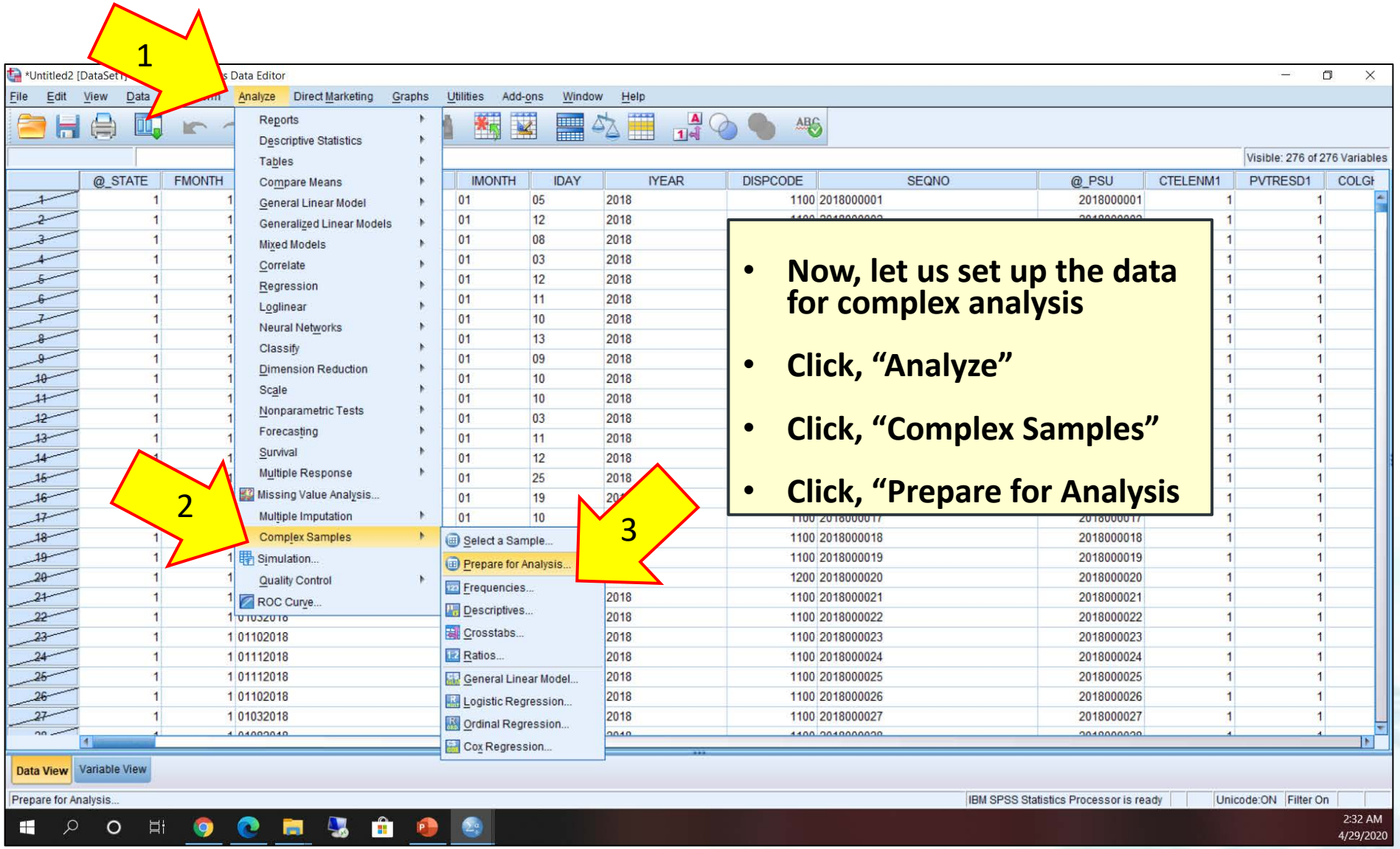


- **Do the same for 2 and 7**
- **For "9", instead of a value click system missing**
- **Click, "Continue"**

GREAT PLAINS IDeA | Clinical and Translational Research

- Click, “Change”
- Click, “OK”
- Similarly recode “current smoking computed variable” and “risk factor for having had permanent teeth extraction” variables and name the new variables current_smoking and permanent_teeth respectively

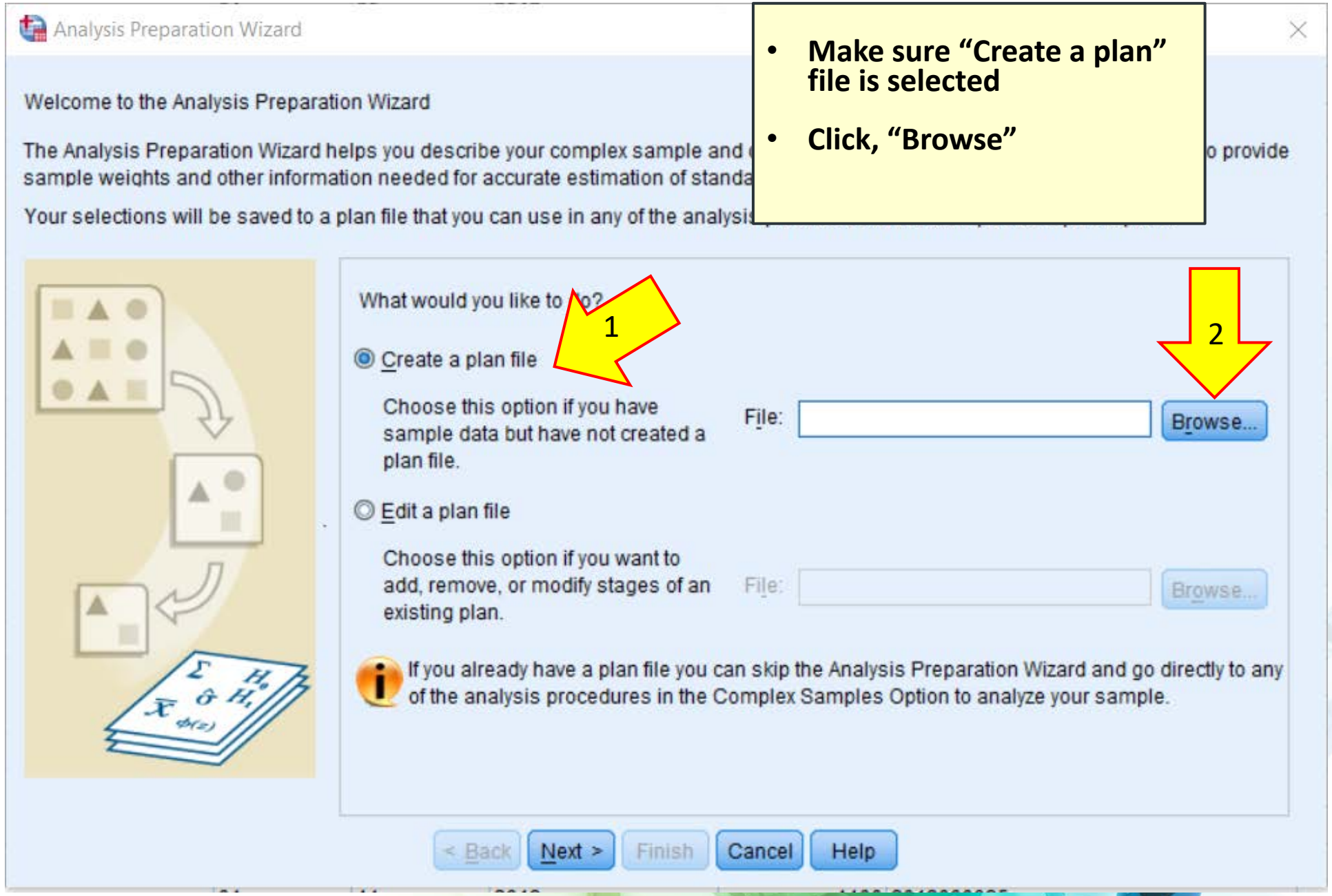




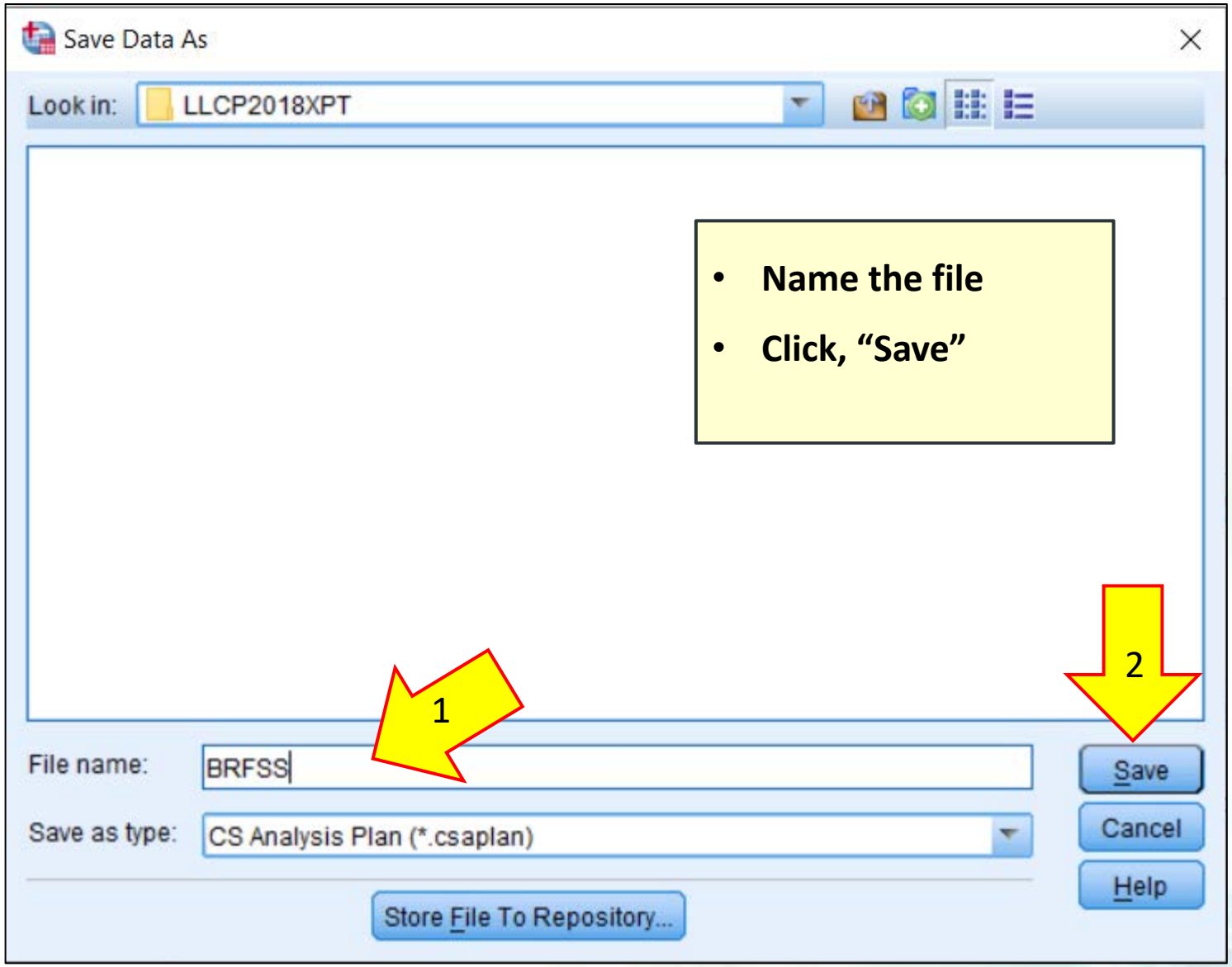
- **Now, let us set up the data for complex analysis**
- **Click, “Analyze”**
- **Click, “Complex Samples”**
- **Click, “Prepare for Analysis**

GREAT PLAINS IDeA | Clinical and Translational Research

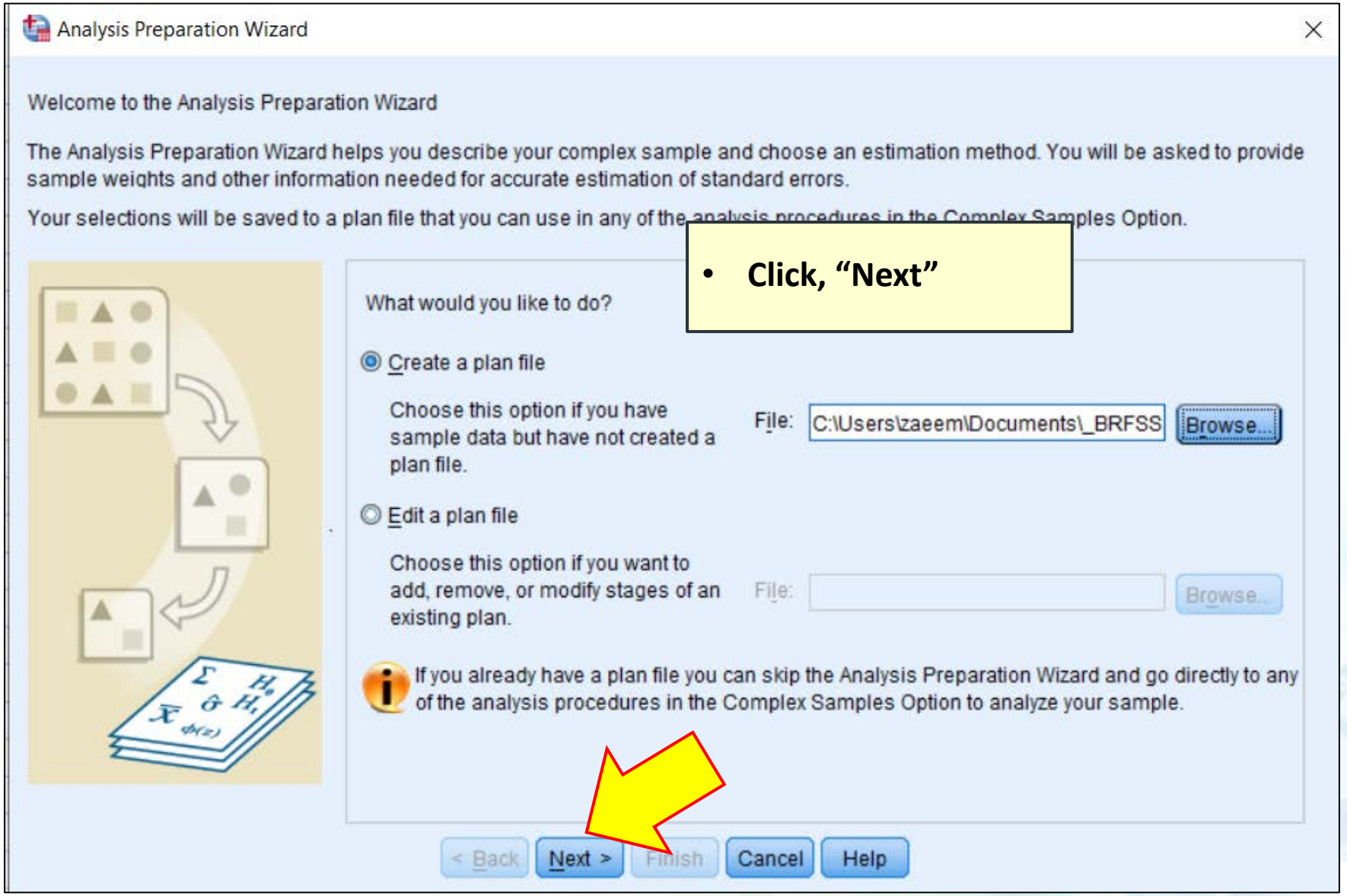
- **Make sure “Create a plan” file is selected**
- **Click, “Browse”**



GREAT PLAINS IDeA | Clinical and Translational Research



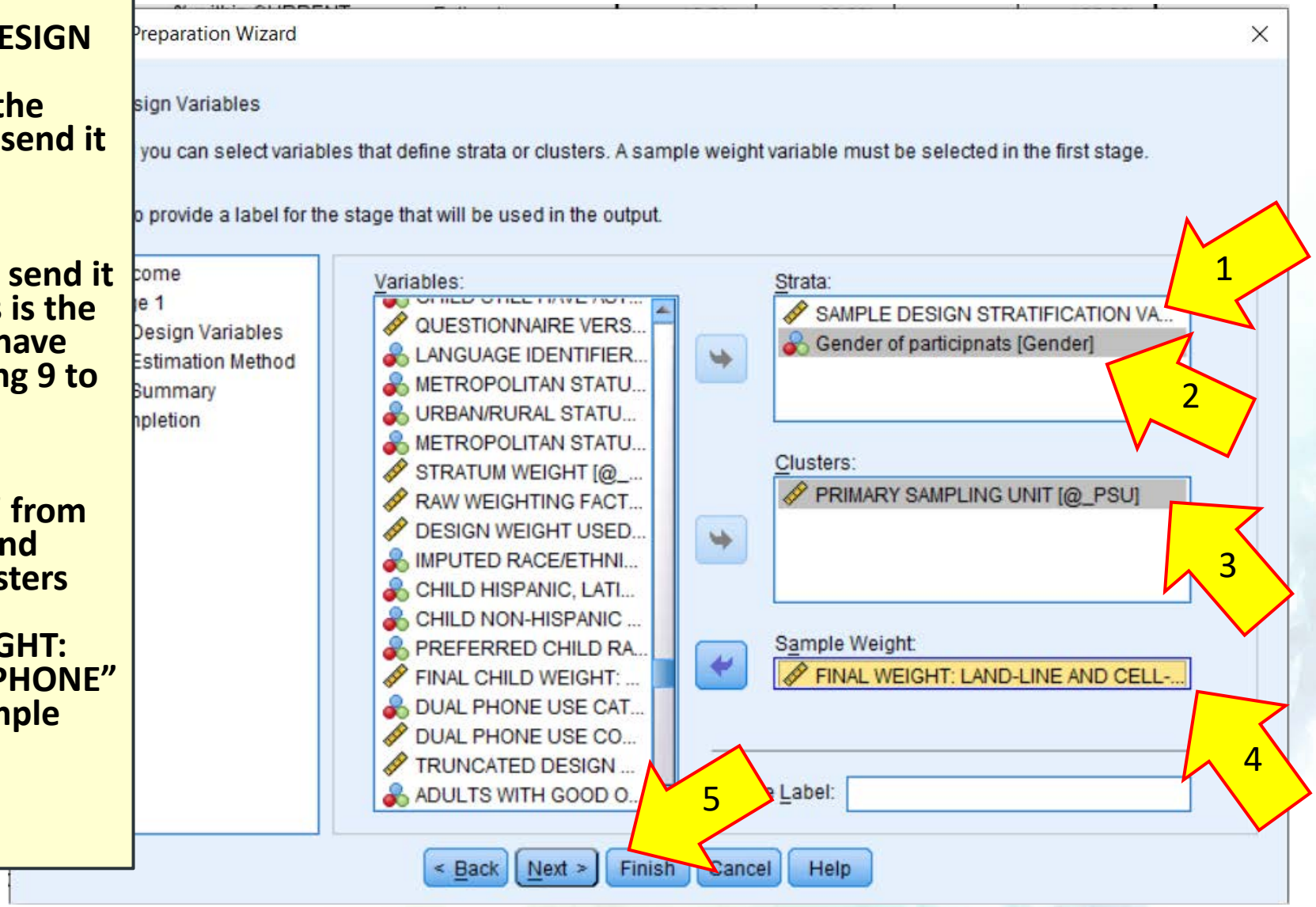
GREAT PLAINS IDeA | Clinical and Translational Research



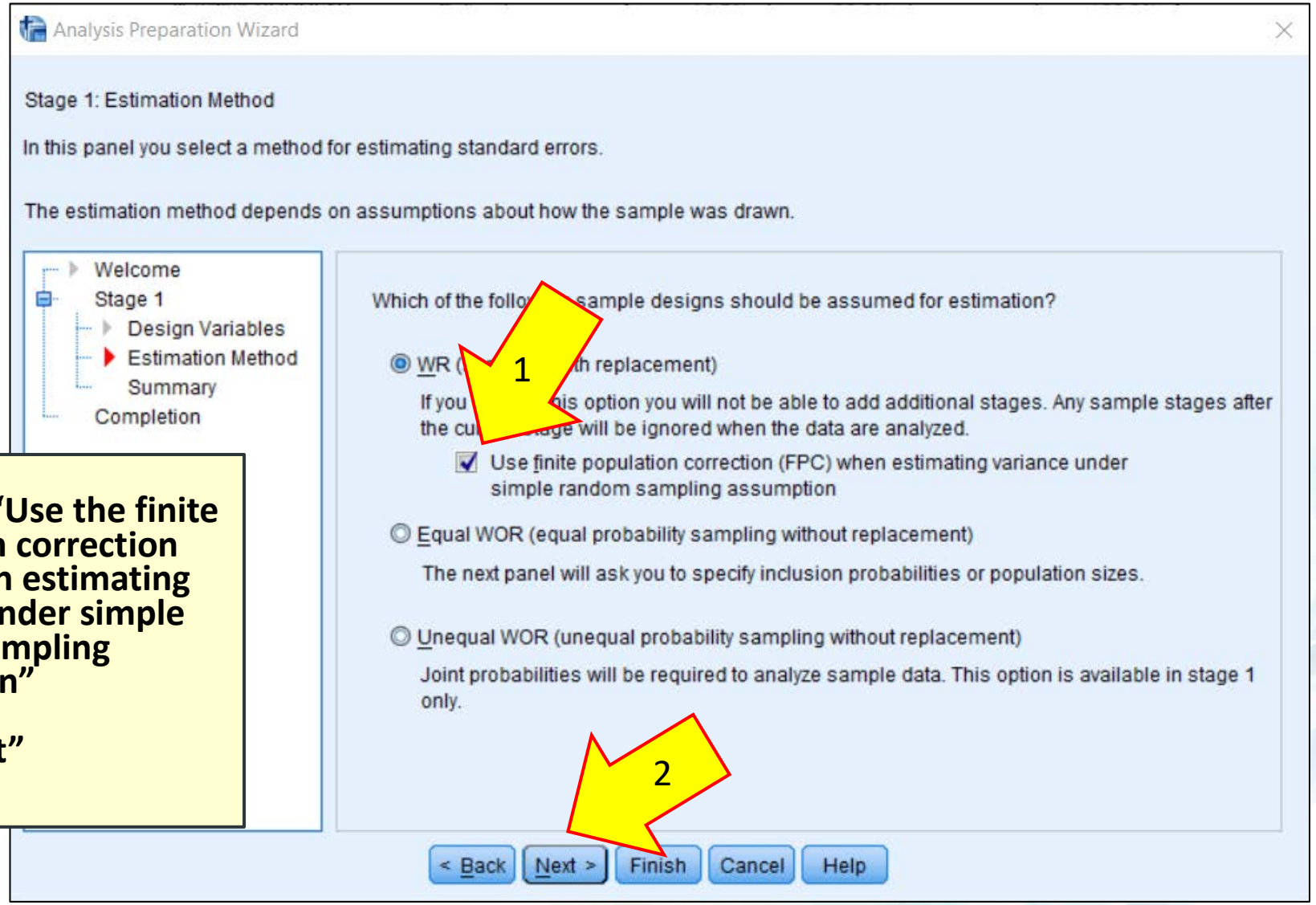
- **Click, “Next”**

GREAT PLAINS IDeA | Clinical and Translational Research

- **Click, “SAMPLE DESIGN STRATIFICATION VARIABLE” from the variables list and send it under Strata**
- **Click “Gender of participants” and send it under strata. This is the variable that we have just created setting 9 to missing**
- **Click, “PRIMARY SAMPLING UNIT” from the variable list and send it under Clusters**
- **Click “FINAL WIRGHT: LAND AND CELL-PHONE” send it under Sample Weight**
- **Click “Next”**



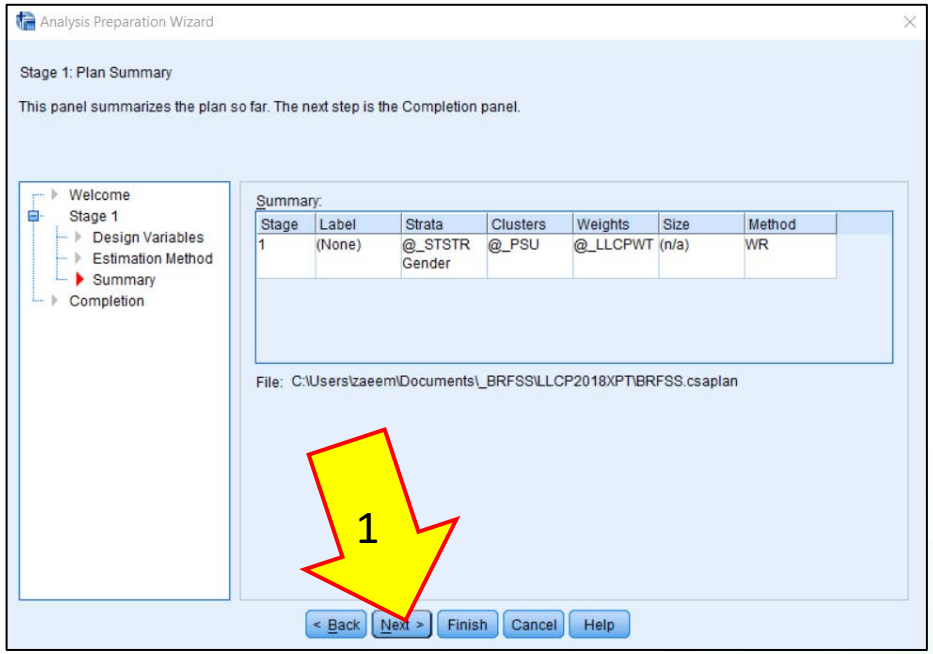
GREAT PLAINS IDeA | Clinical and Translational Research

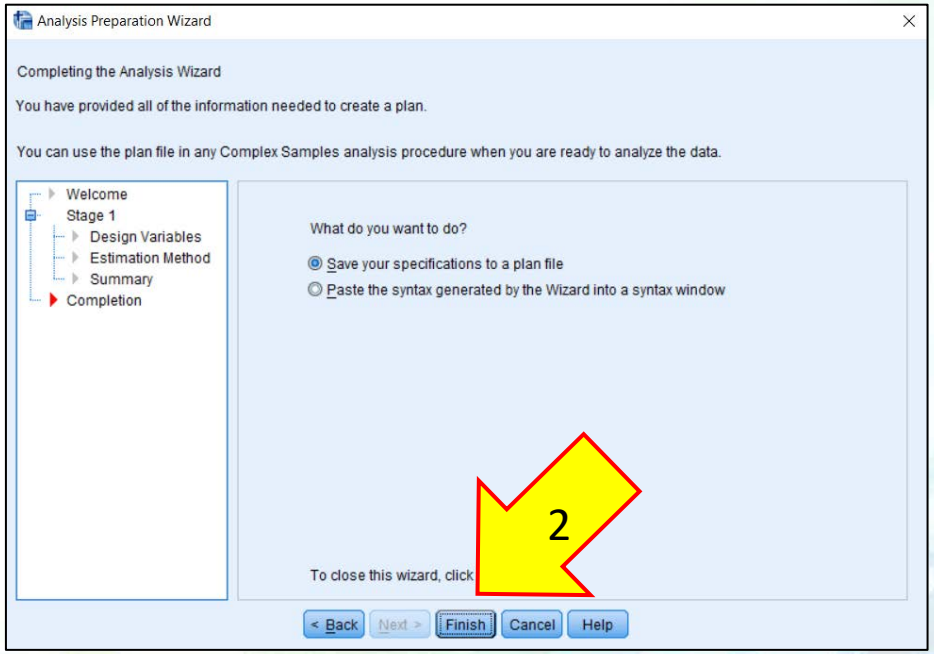


- **Uncheck' "Use the finite population correction (FPC) when estimating variance under simple random sampling assumption"**
- **Click "Next"**

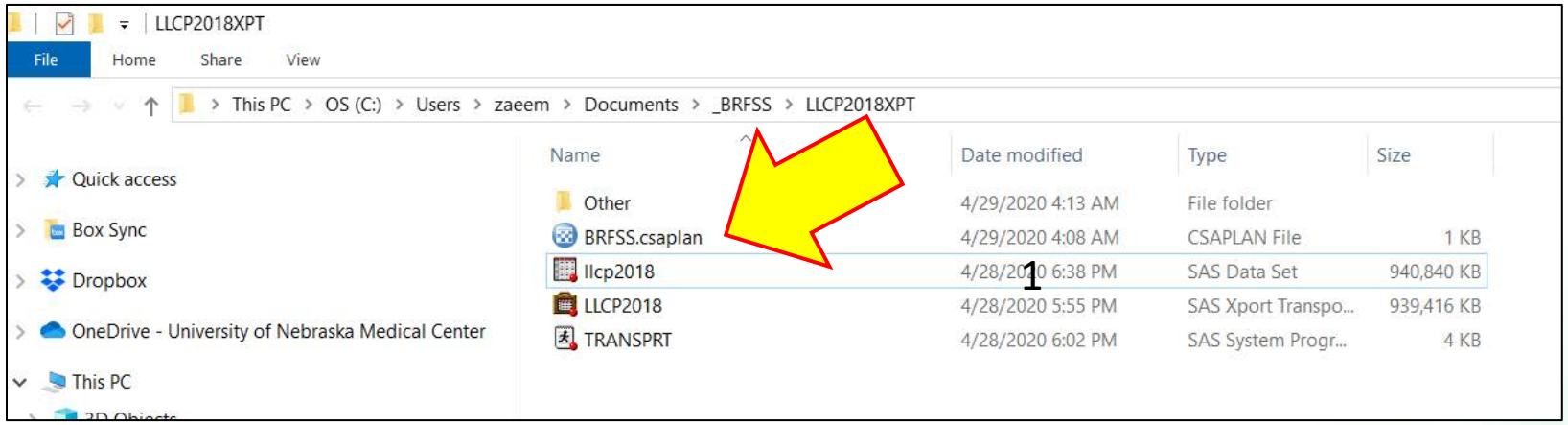
GREAT PLAINS IDeA | Clinical and Translational Research

- **Click, “Next”**
- **Click, “Finish”**



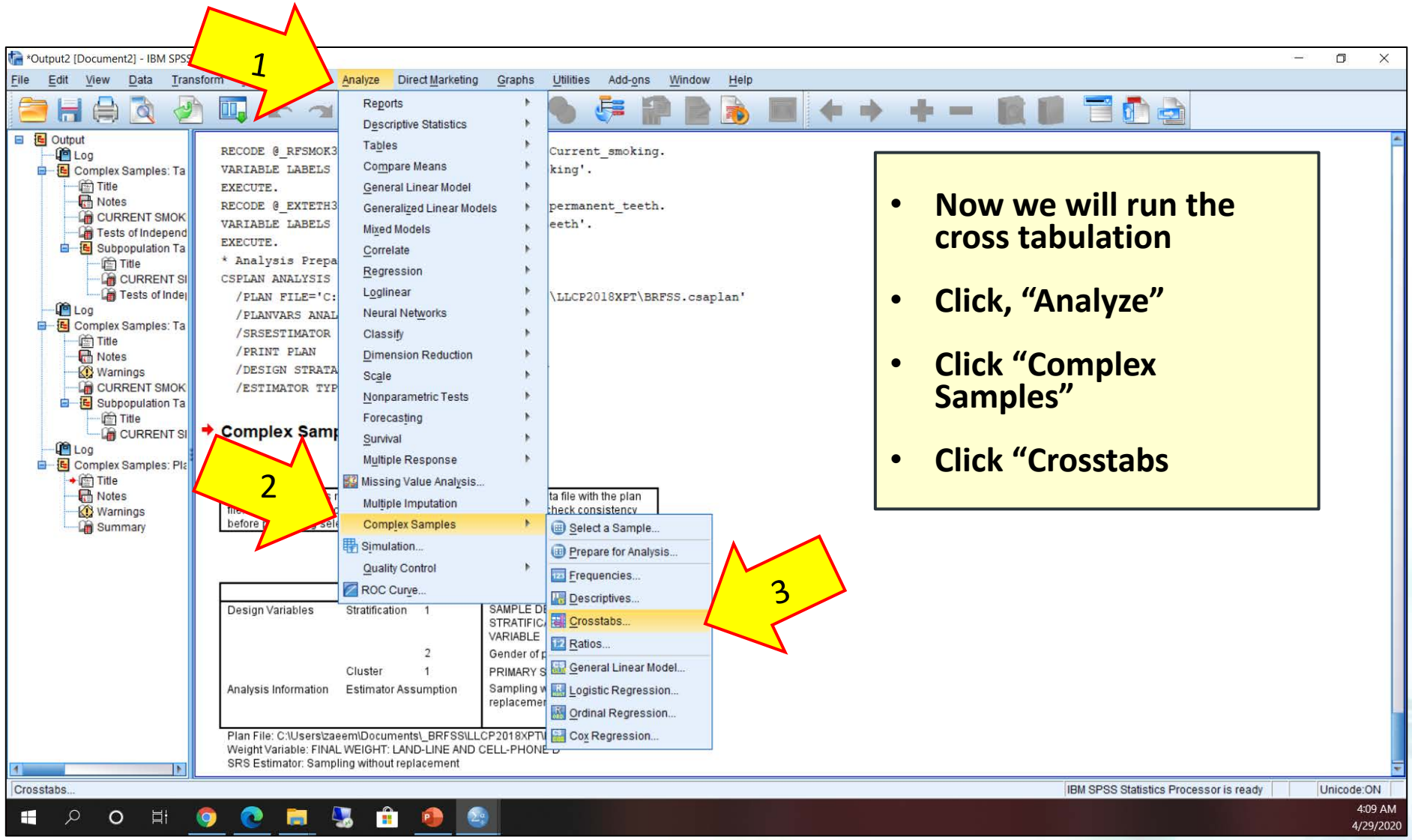


GREAT PLAINS
IDeA | Clinical and Translational Research



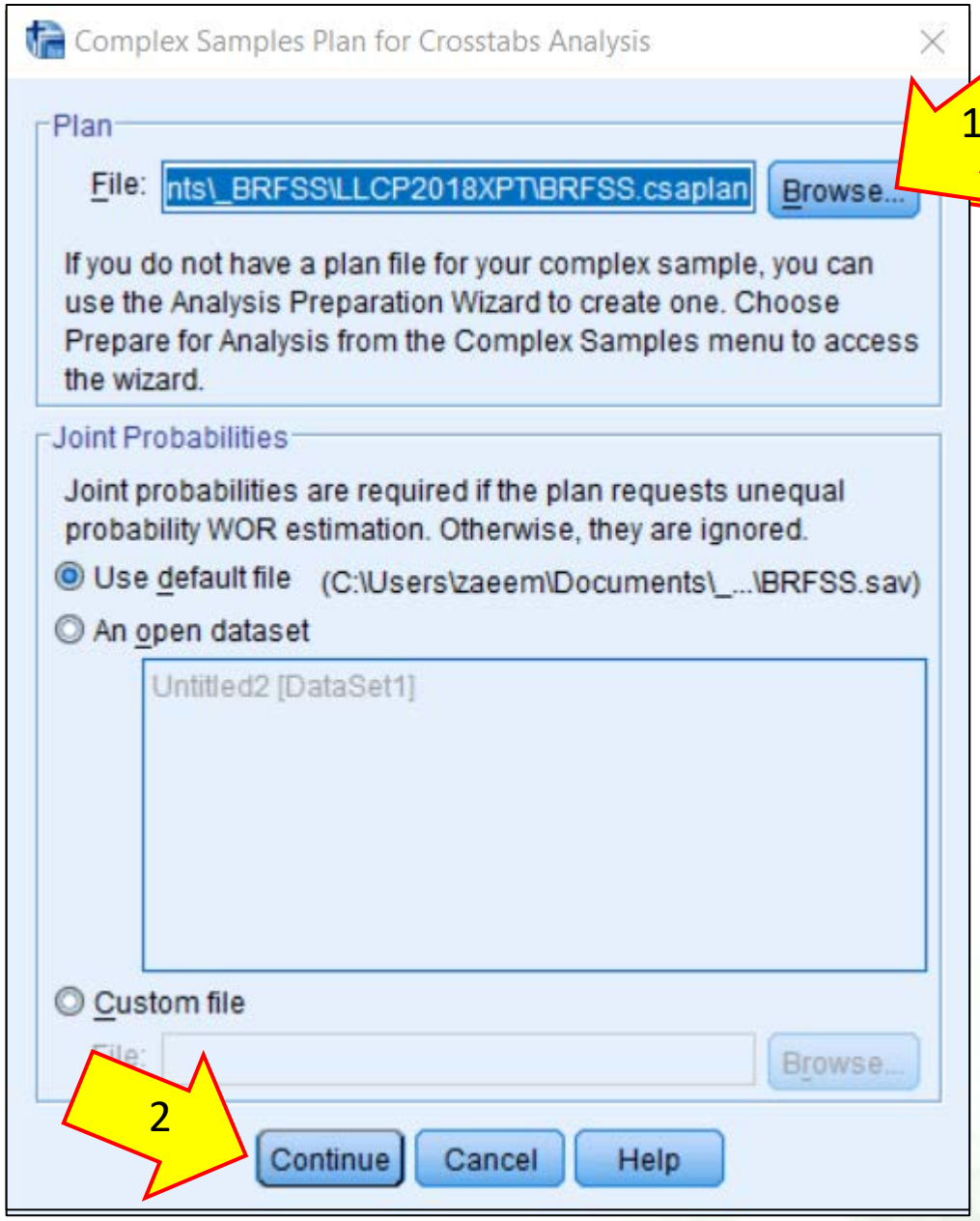
- **The complex sample plan file is saved in your specified folder**

GREAT PLAINS
IDeA | Clinical and Translational Research



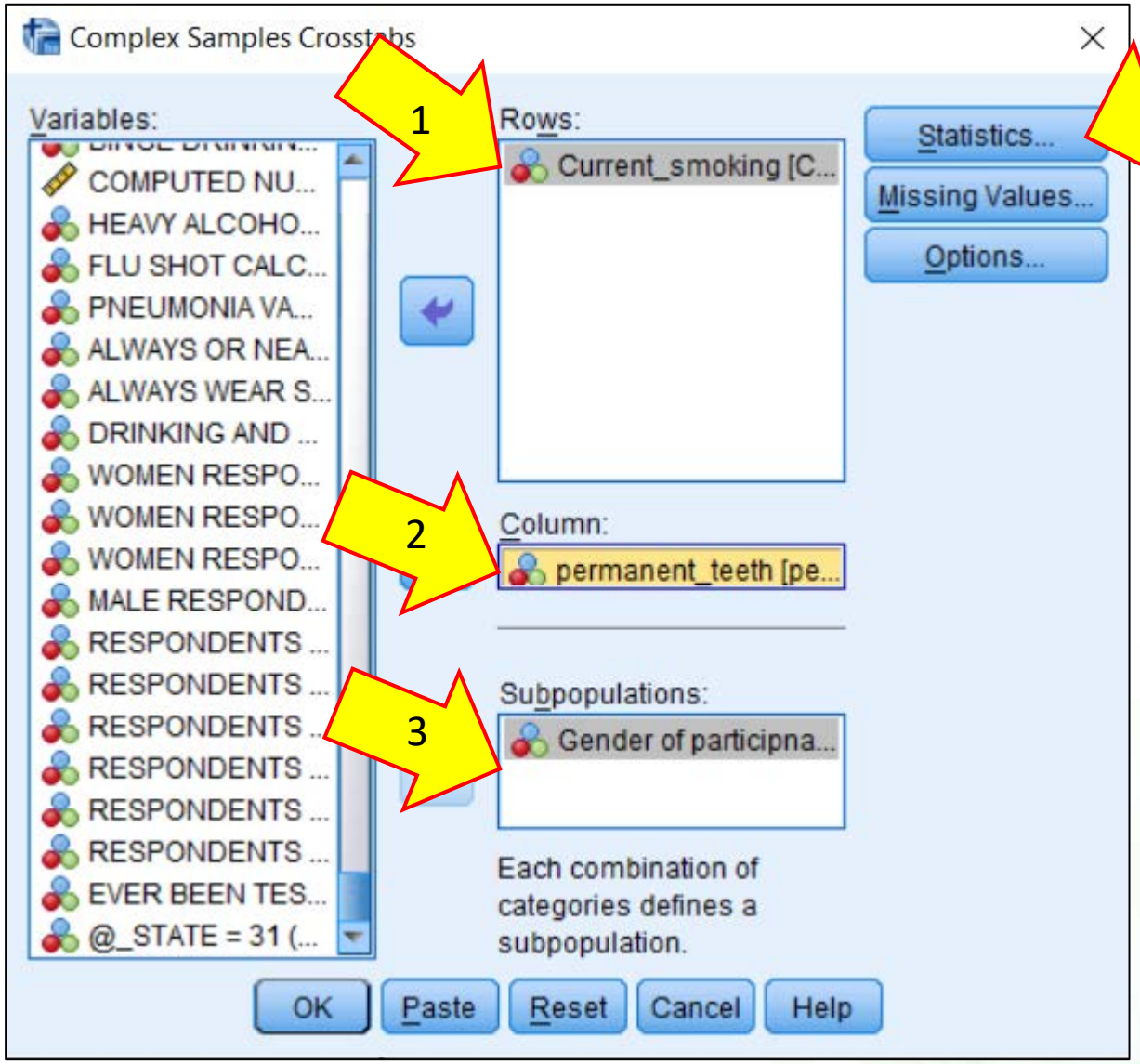
- **Now we will run the cross tabulation**
- **Click, “Analyze”**
- **Click “Complex Samples”**
- **Click “Crosstabs**

GREAT PLAINS
IDeA | Clinical and Translational Research



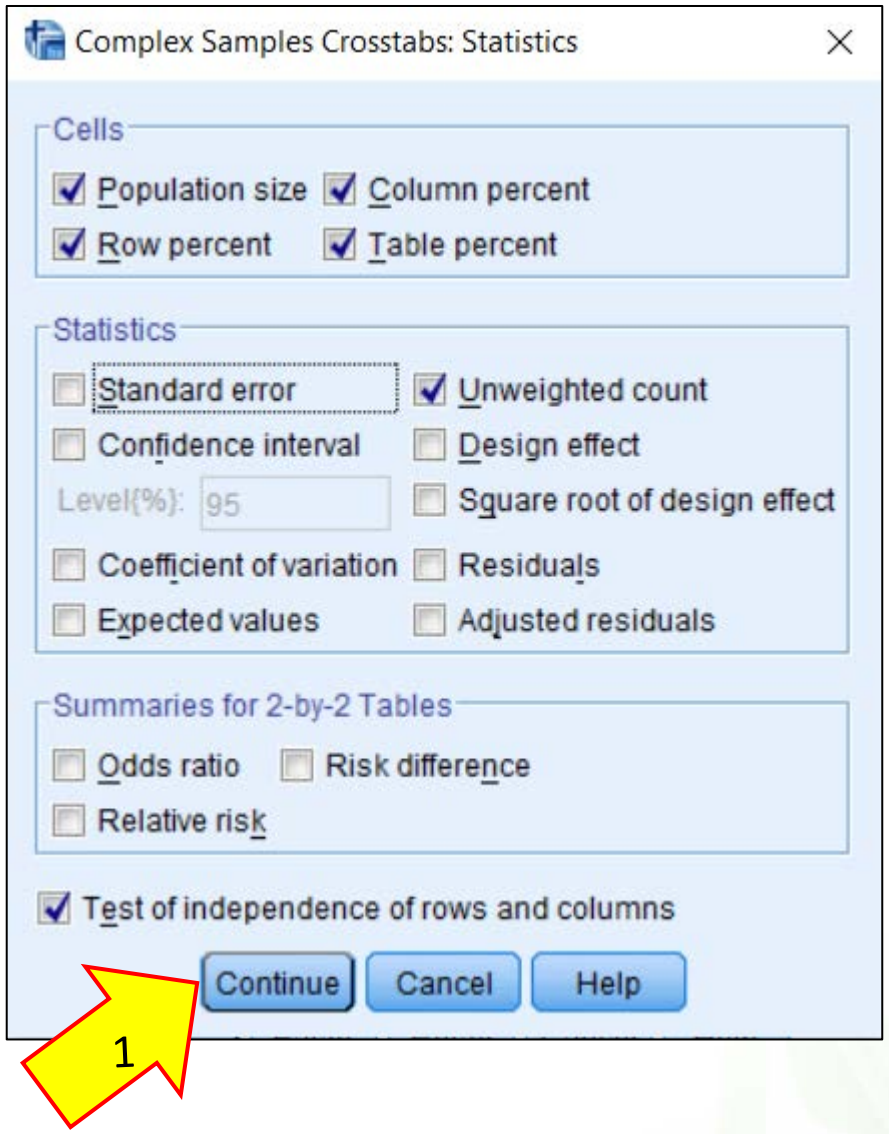
- **Whenever you open SPSS, browse and select the plan file that we just created. This plan file has the information about cluster, strata and weight**
- **Click, "Continue"**

GREAT PLAINS IDeA | Clinical and Translational Research



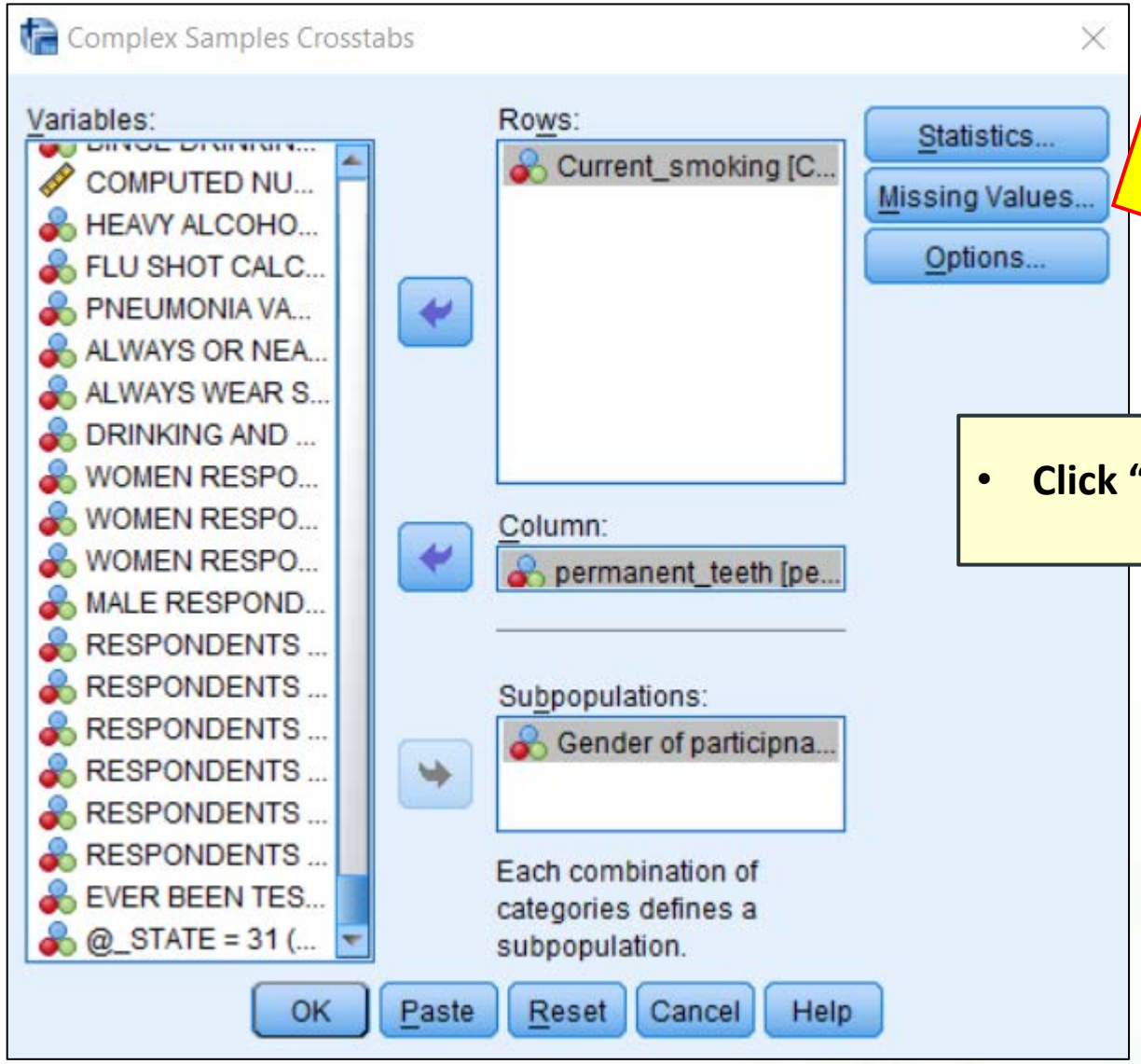
- **Bring "Current_smoking" under Rows**
- **Bring Permanent_teeth under column**
- **Bring "Gender of participants" under Subpopulations**
- **These variables are the ones that we recoded into new variables specifying the missing values**
- **The recoded variables always appear at the bottom of the "Variables"**
- **Click, "Statistics"**

GREAT PLAINS IDeA | Clinical and Translational Research

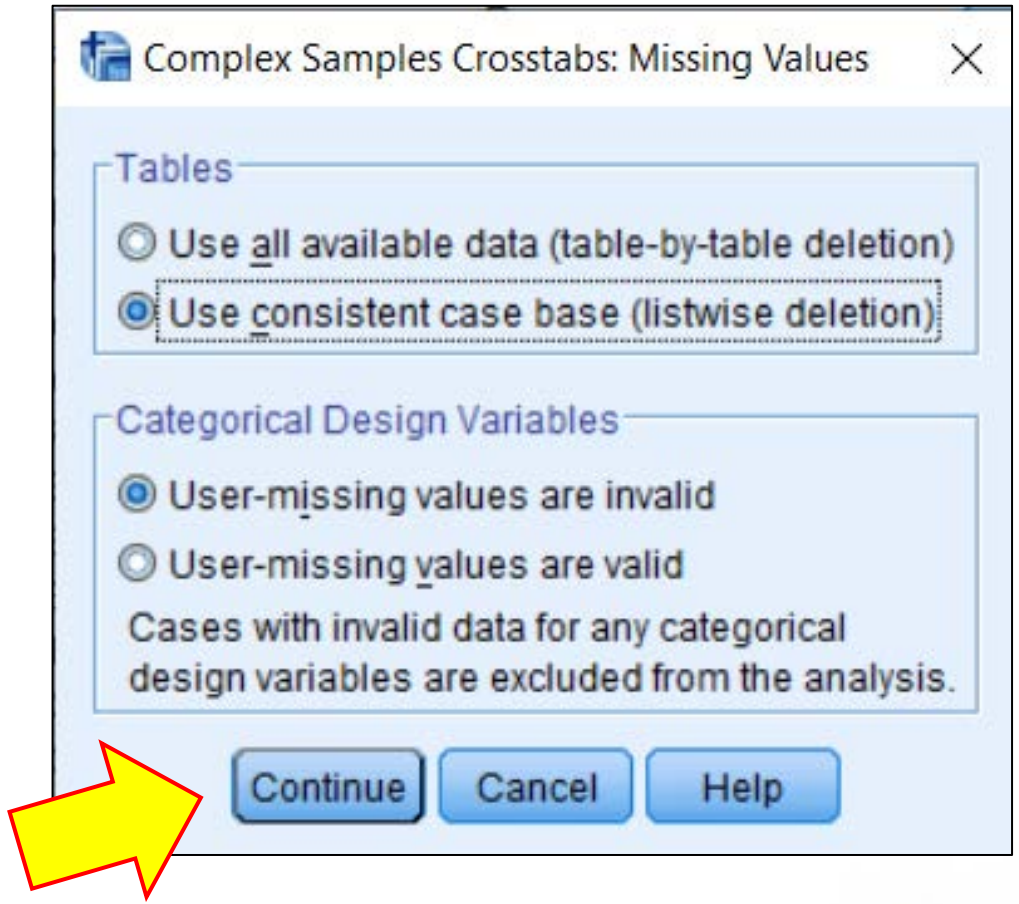


- **You can select the options of your choice or follow what has been shown here**
- **Click, “Continue”**

GREAT PLAINS IDeA | Clinical and Translational Research

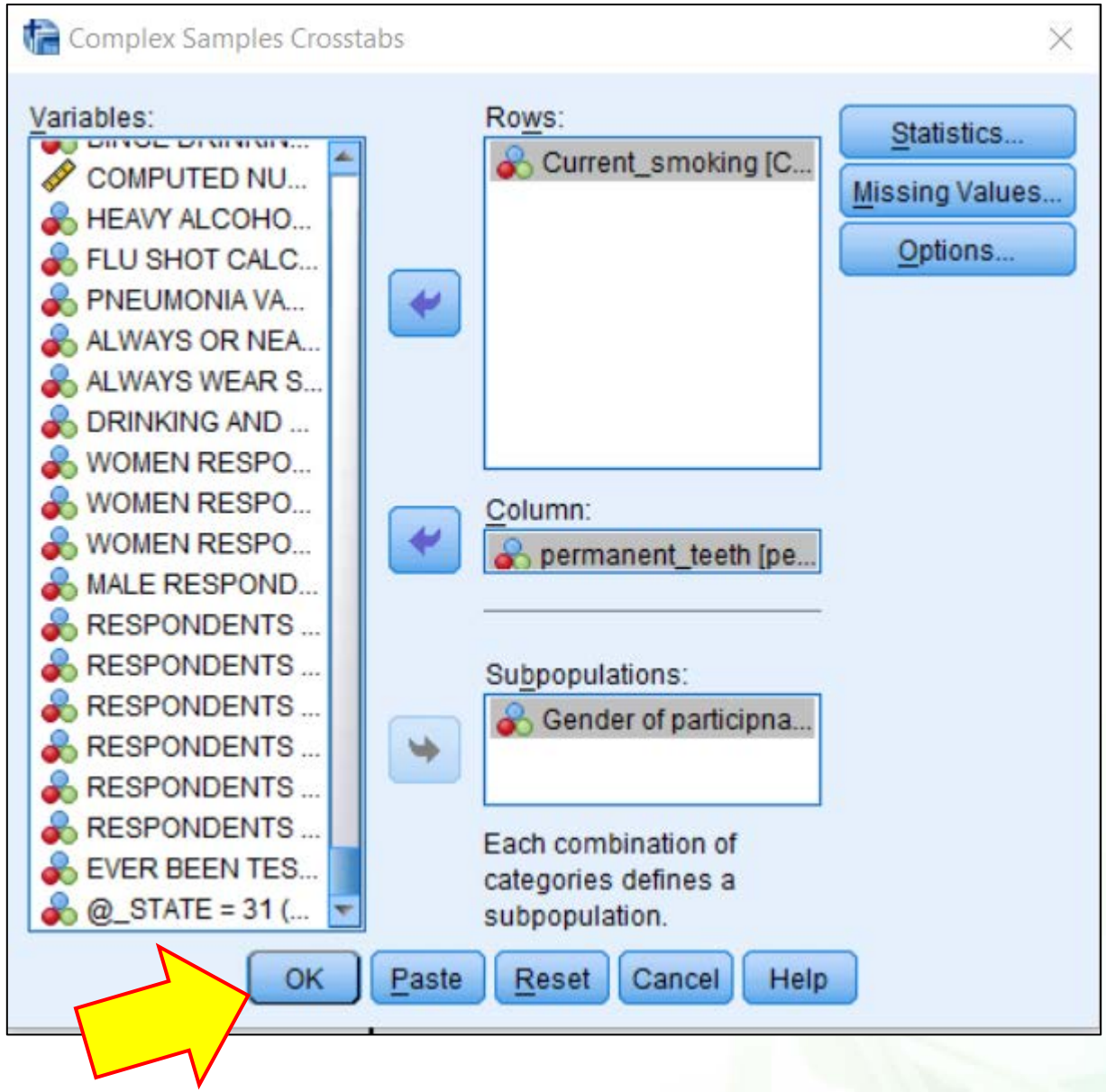


- **Click “Missing Values”**



- **Select the options as shown**
- **Click, “Continue”**

GREAT PLAINS IDeA | Clinical and Translational Research


Complex Samples Crosstabs
Variables:
COMPUTED NU...
HEAVY ALCOHO...
FLU SHOT CALC...
PNEUMONIA VA...
ALWAYS OR NEA...
ALWAYS WEAR S...
DRINKING AND ...
WOMEN RESPO...
WOMEN RESPO...
WOMEN RESPO...
MALE RESPOND...
RESPONDENTS ...
RESPONDENTS ...
RESPONDENTS ...
RESPONDENTS ...
RESPONDENTS ...
RESPONDENTS ...
EVER BEEN TES...
@_STATE = 31 (...
Rows:
Current_smoking [C...
Statistics...
Missing Values...
Options...
Column:
permanent_teeth [pe...
Subpopulations:
Gender of participna...
Each combination of categories defines a subpopulation.
OK
Paste
Reset
Cancel
Help

- **Click, “OK”**

Current_smoking * permanent_teeth

Gender of participnats	Current_smoking			permanent_teeth 1.00	permanent_teeth 2.00	permanent_teeth Total
1.00	1.00	Population Size	Estimate	368276.735	200871.480	569148.216
			Unweighted Count	2978	2338	5316
		% within Current_smoking	Estimate	64.7%	35.3%	100.0%
			Unweighted Count	2978	2338	5316
		% within permanent_teeth	Estimate	86.1%	76.9%	82.6%
			Unweighted Count	2978	2338	5316
		% of Total	Estimate	53.5%	29.2%	82.6%
			Unweighted Count	2978	2338	5316
	2.00	Population Size	Estimate	59231.644	60413.756	119645.401
			Unweighted Count	394	548	942
		% within Current_smoking	Estimate	49.5%	50.5%	100.0%
			Unweighted Count	394	548	942
		% within permanent_teeth	Estimate	13.9%	23.1%	17.4%
			Unweighted Count	394	548	942
		% of Total	Estimate	8.6%	8.8%	17.4%
			Unweighted Count	394	548	942
	Total	Population Size	Estimate	427508.380	261285.236	688793.616
			Unweighted Count	3372	2886	6258
		% within Current_smoking	Estimate	62.1%	37.9%	100.0%
			Unweighted Count	3372	2886	6258
		% within permanent_teeth	Estimate	100.0%	100.0%	100.0%
			Unweighted Count	3372	2886	6258
		% of Total	Estimate	62.1%	37.9%	100.0%
			Unweighted Count	3372	2886	6258

Tests of Independence

Gender of participnats			Chi-Square	Adjusted F	df1	df2	Sig.
1.00	Current_smoking * permanent_teeth	Pearson	88.150	38.523	1	6196	.000
		Likelihood Ratio	86.078	37.618	1	6196	.000
2.00	Current_smoking * permanent_teeth	Pearson	133.683	50.685	1	7615	.000
		Likelihood Ratio	129.795	49.211	1	7615	.000
7.00	Current_smoking * permanent_teeth	Pearson	.203	.	.	.	.
		Likelihood Ratio	.341	.	.	.	.

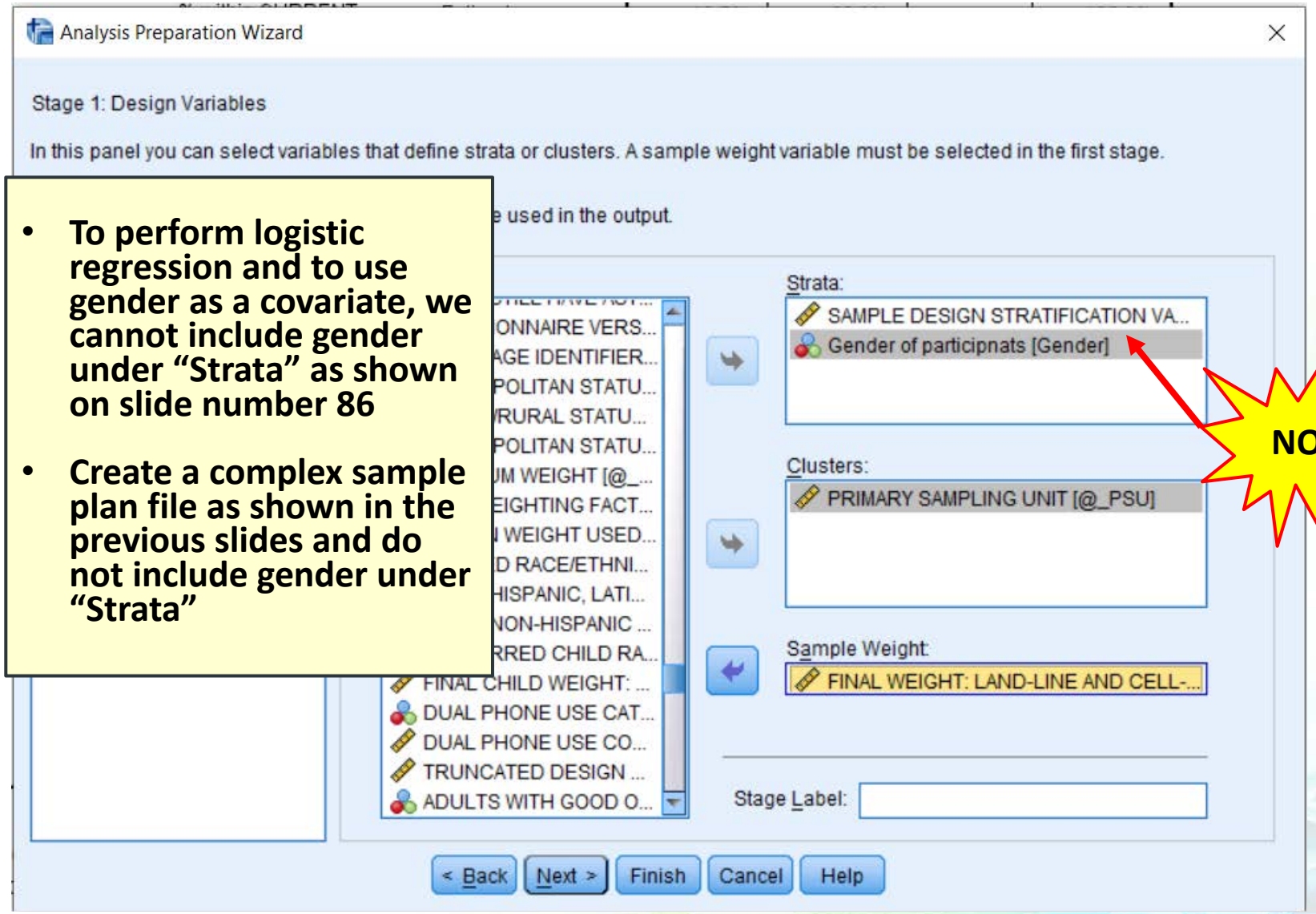
The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic. Significance is based on the adjusted F and its degrees of freedom.

- These are two tables from the results.
- The upper table is the cross tabulation between current smoking and permanent teeth extraction for males.
- As you scroll down in you SPSS, you will see results for other gender categories
- The bottom table shows the results of chi-square test
- Note that we have received the same results as those of the WEAT tool and SASS

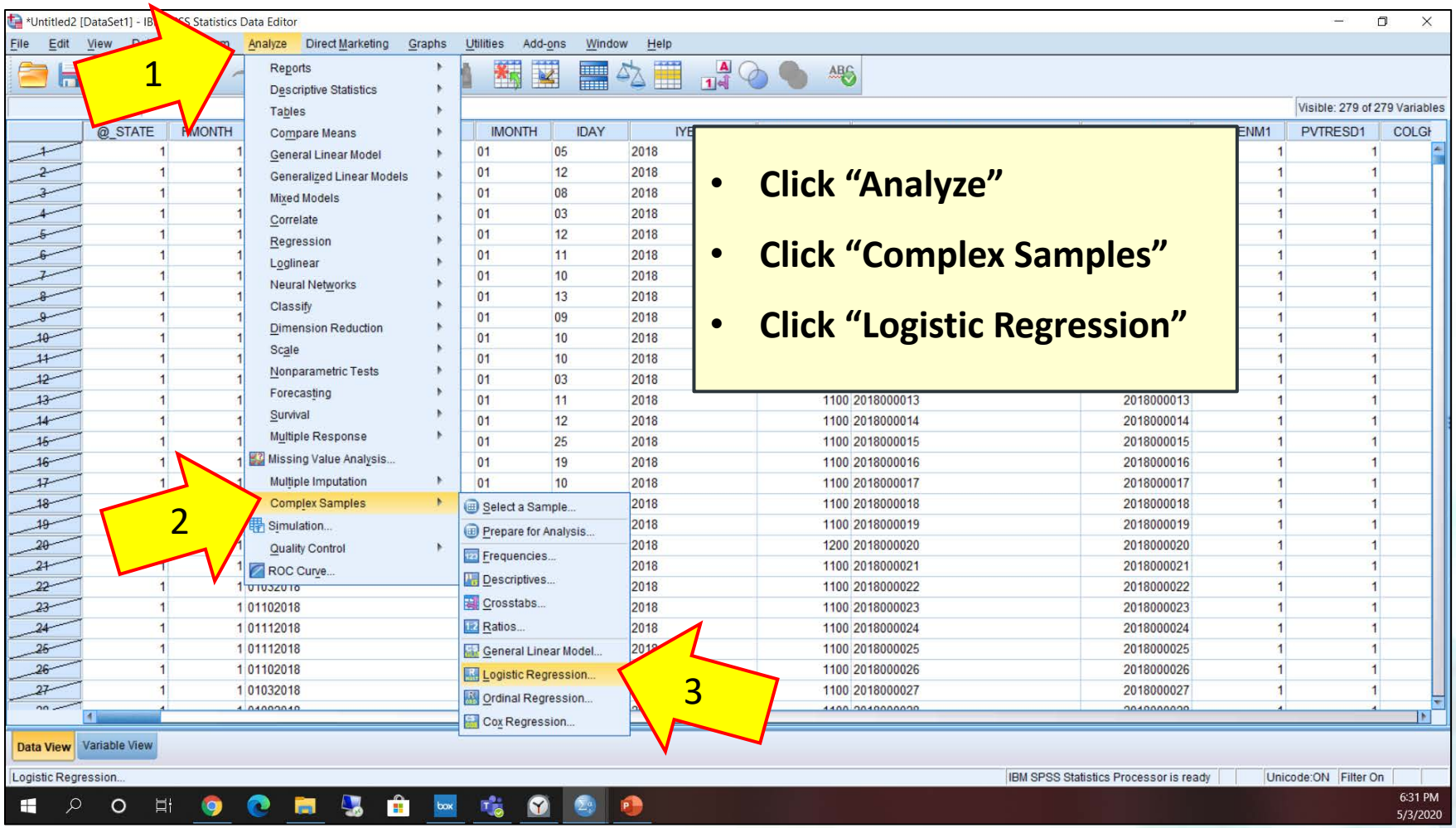
GREAT PLAINS IDeA | Clinical and Translational Research

LOGISTIC REGRESSION

- To perform logistic regression and to use gender as a covariate, we cannot include gender under “Strata” as shown on slide number 86
- Create a complex sample plan file as shown in the previous slides and do not include gender under “Strata”

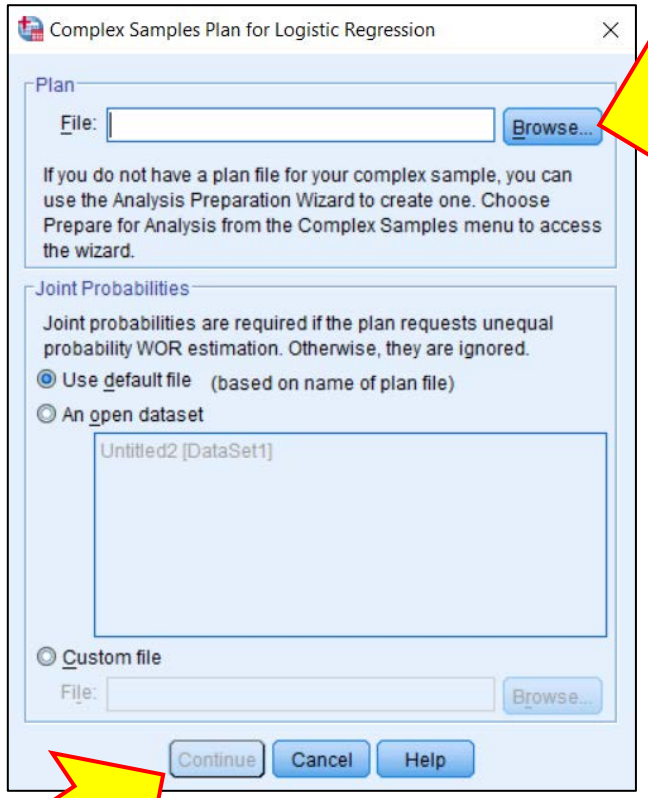


GREAT PLAINS IDeA | Clinical and Translational Research

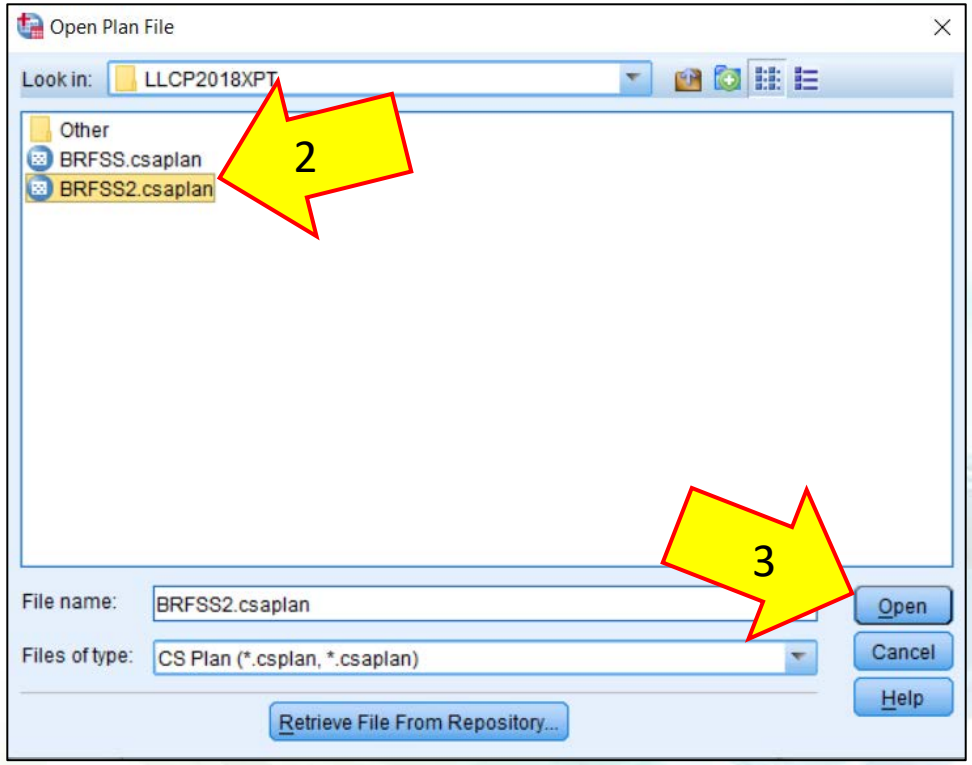


- Click “Analyze”
- Click “Complex Samples”
- Click “Logistic Regression”

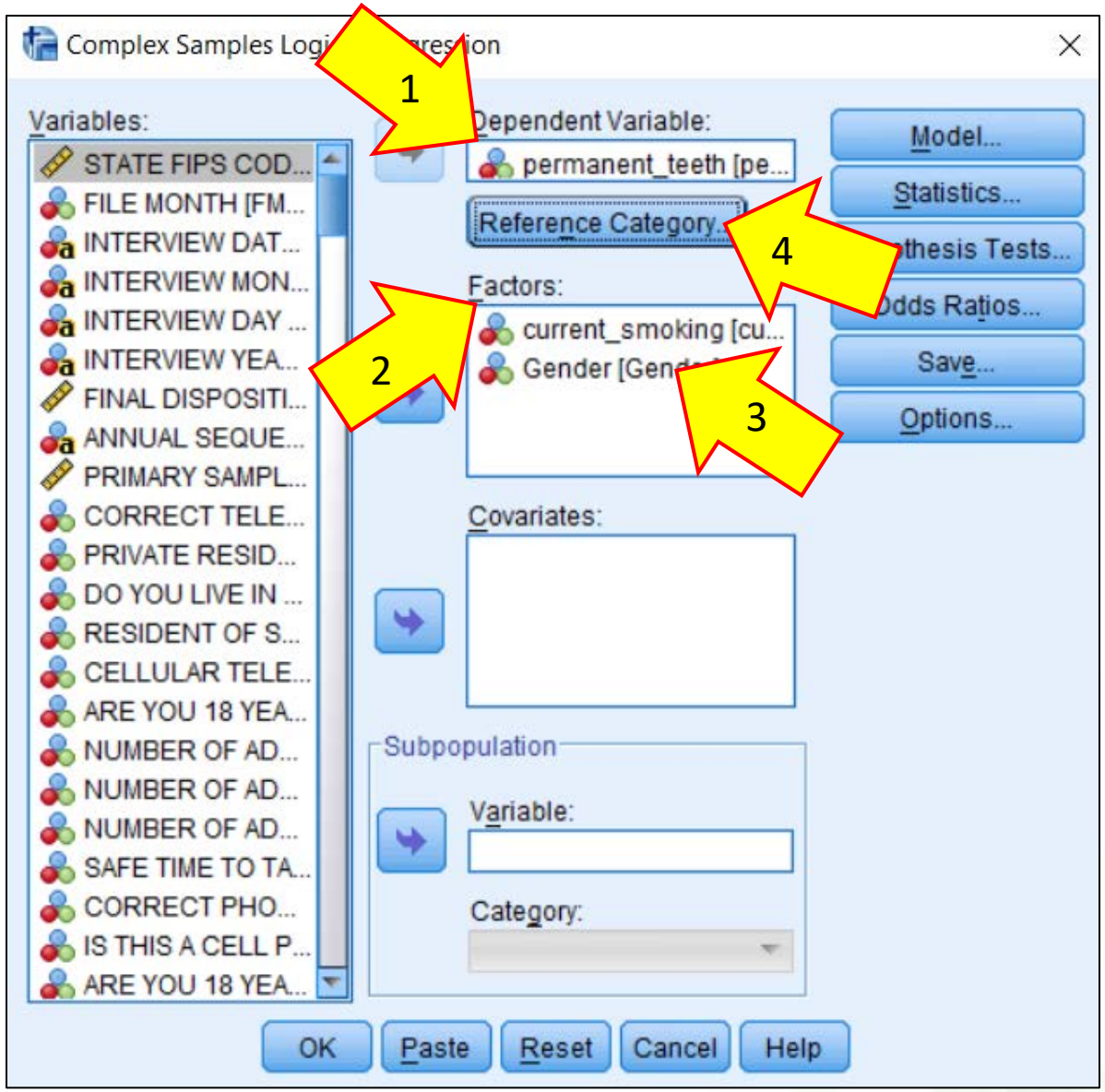
GREAT PLAINS IDeA | Clinical and Translational Research



- **Click "Browse"**
- **Select your saved complex sample plan file.**
- **Click open**
- **Click "Continue"**

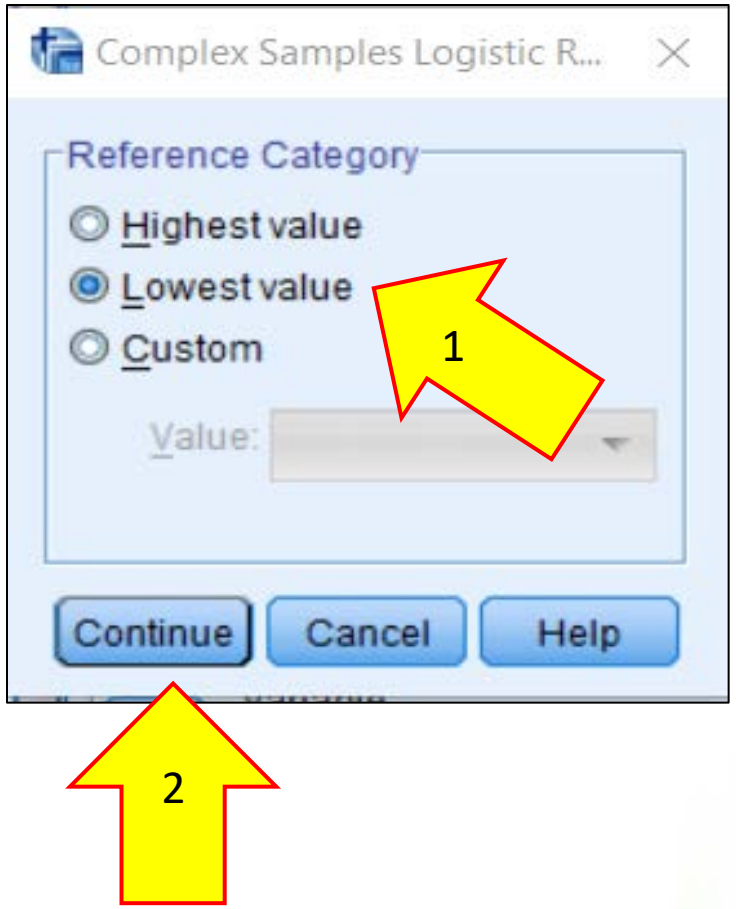


GREAT PLAINS IDeA | Clinical and Translational Research



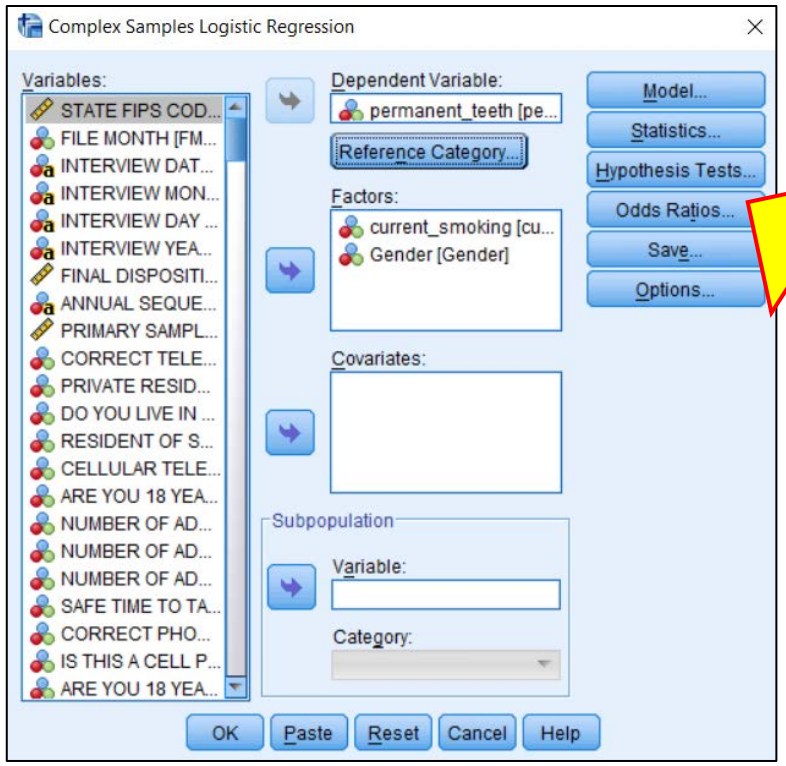
- **Select “permanent_teeth” from variables list and transfer if under “Dependent Variable”**
- **Select “current_smoking” from variables list and transfer it under “Factors”**
- **Select “Gender” from variables list and transfer it under “Factors”**
- **Click “Reference Category”**

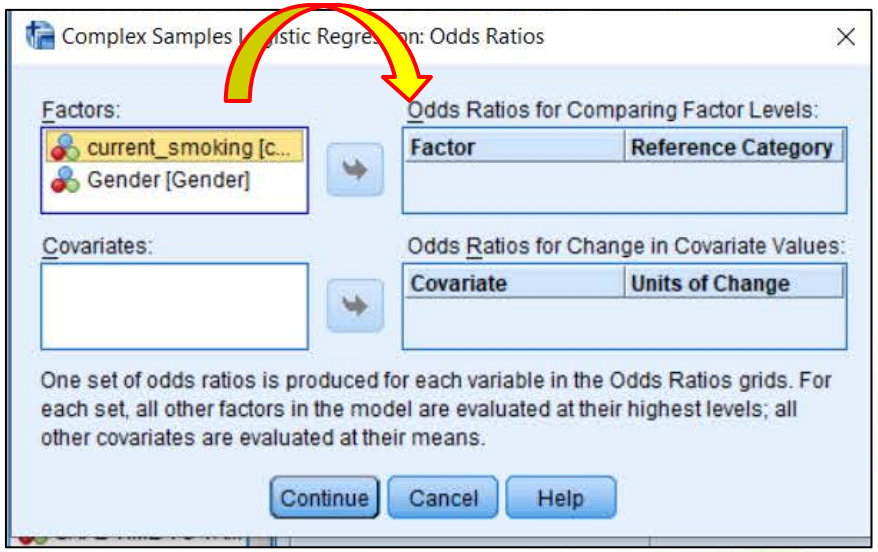
GREAT PLAINS IDeA | Clinical and Translational Research



- Set the reference category to “lowest value”
- The variable “permanent-teeth” has two categories i.e.
  - 1=not at risk
  - 2=at risk
- We want to see the results for those who are at risk, thus we will put 1 as reference (lowest value)
- Click Continue

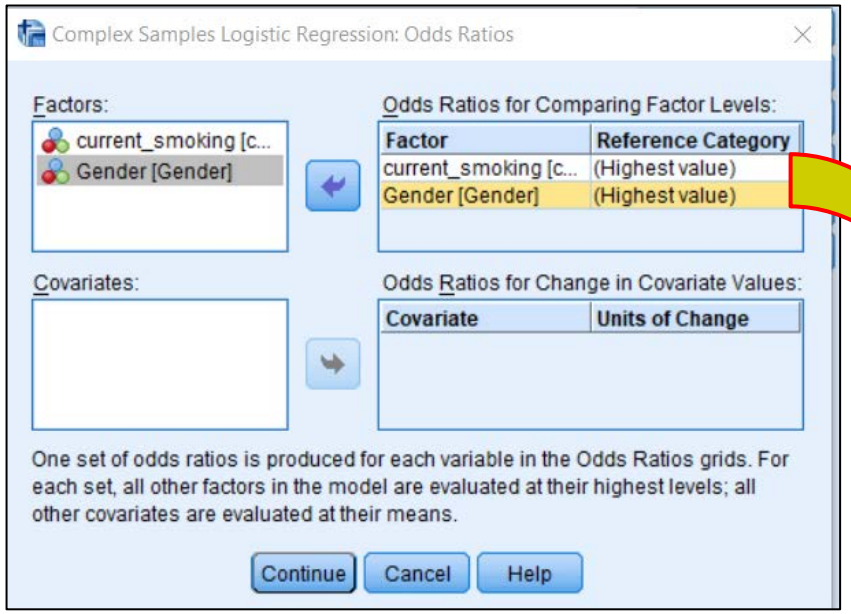
GREAT PLAINS IDeA | Clinical and Translational Research

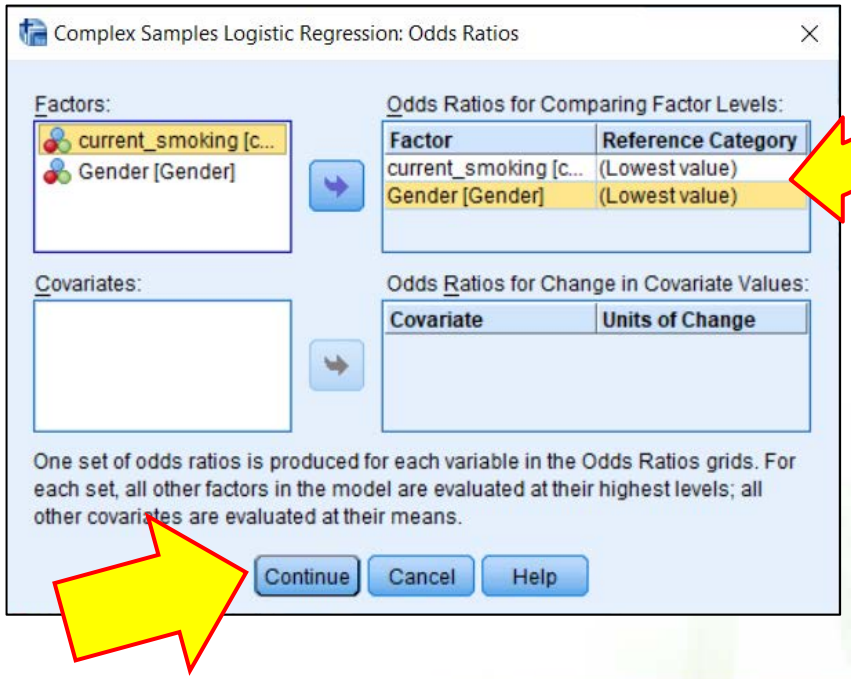




- **Click “Odds Ratios”**
- **On the resulting screen, transfer both, “current_smoking” and “gender” in the factors list to the cell on the right**

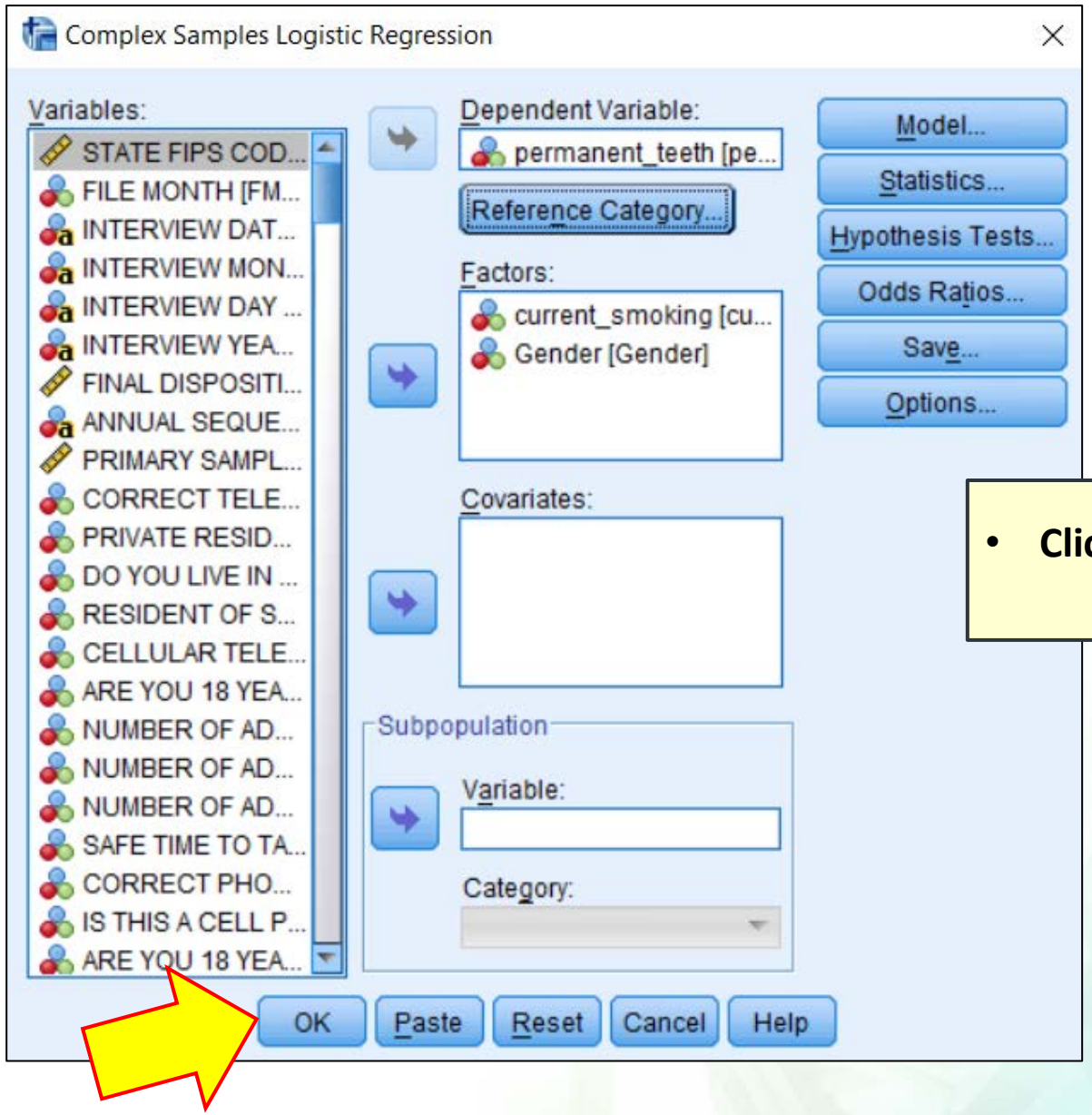
GREAT PLAINS IDeA | Clinical and Translational Research





- **The the options under "Reference Category" one by one and change tem to lowest as shoen in the lower figure**
- **It is very important to set your references correctly to get the desired measurements**
- **Click Continue**

GREAT PLAINS IDeA | Clinical and Translational Research


Complex Samples Logistic Regression
Variables:
STATE FIPS COD...
FILE MONTH [FM...
INTERVIEW DAT...
INTERVIEW MON...
INTERVIEW DAY ...
INTERVIEW YEA...
FINAL DISPOSITI...
ANNUAL SEQUE...
PRIMARY SAMPL...
CORRECT TELE...
PRIVATE RESID...
DO YOU LIVE IN ...
RESIDENT OF S...
CELLULAR TELE...
ARE YOU 18 YEA...
NUMBER OF AD...
NUMBER OF AD...
NUMBER OF AD...
SAFE TIME TO TA...
CORRECT PHO...
IS THIS A CELL P...
ARE YOU 18 YEA...
Dependent Variable:
permanent_teeth [pe...
Reference Category...
Factors:
current_smoking [cu...
Gender [Gender]
Covariates:
Subpopulation
Variable:
Category:
Model...
Statistics...
Hypothesis Tests...
Odds Ratios...
Save...
Options...
OK
Paste
Reset
Cancel
Help

- **Click "OK"**

Odds Ratios 1[a]

			Odds Ratio	95% Confidence Interval	
		permanent_teeth		Lower	Upper
current_smoking	2.00 vs. 1.00	2.00	1.978	1.713	2.284

Dependent Variable: permanent_teeth (reference category = 1.00)
Model: (Intercept), current_smoking, Gender

a. Factors and covariates used in the computation are fixed at the following values: current_smoking=2.00; Gender=7.00

Odds Ratios 2[a]

			Odds Ratio	95% Confidence Interval	
		permanent_teeth		Lower	Upper
Gender	2.00 vs. 1.00	2.00	1.007	.910	1.114
	7.00 vs. 1.00	2.00	8.623	1.034	71.918

Dependent Variable: permanent_teeth (reference category = 1.00)
Model: (Intercept), current_smoking, Gender

a. Factors and covariates used in the computation are fixed at the following values: current_smoking=2.00; Gender=7.00

- **See the SPSS results above**
- **They are similar to what we get from SAS and WEAT tool**

GREAT PLAINS IDeA | Clinical and Translational Research

University of Nebraska Medical Center

BREAKTHROUGHS FOR LIFE.®

GREAT PLAINS
IDeA | Clinical and
Translational Research