/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

STEPS FOR UPDATING CANCER ATLAS

Version: 2023

Editor: Jason Massey

Date Last Edited: 3/23/2023

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Step 1: Replace library with correct pathname to import census data from Social Explorer and cancer data from CDC

Step 2: If needed update code

Step 3: Check new atlas and verify with old

Step 4: Export final file with correct pathname to be merged with shapefile

Step 5: Make a copy of data processing code for next update and replace Old Atlas Dataset with updated atlas dataset

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*

(A) Old Atlas Data

\*/

\*Old Atlas Dataset for commparison;

\*AFTER COPYING CURRENT DATA PROCESSING CODE, VERIFY PATHNAME AND REPLACE WITH MOST RECENT VERSION OF ATLAS DATASET;

**proc** **import**

dbms=dbf

datafile= "C:\Users\JasonMassey\OneDrive - American Cancer Society\Documents\Misc\Cancer\_Atlas\_Updates\2023\_Updates\CountyT1\_SImple\_IncMorPop.dbf"

out=atlas\_old;

**run**;

\*Drop Lables and Formats;

**proc** **datasets** lib=work;

modify atlas\_old;

attrib \_all\_ label='';

**run**;

\*Keep relevant variables from old atlas;

**data** atlas\_old2;

set atlas\_old;

keep ObjectID NAME STATE\_NAME STATE\_FIPS CNTY\_FIPS FIPS MStateID ;

rename FIPS = Geo\_FIPS ;

**run**;

\*Sort old atlas;

**proc** **sort** data = atlas\_old2;

by Geo\_FIPS;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Step 1: Replace library with correct pathname and import cancer data from CDC and census data from Social Explorer

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*Library;

\* REPLACE WITH CORRECT PATHNAME;

libname atlas "C:\Users\JasonMassey\OneDrive - American Cancer Society\Documents\Misc\Cancer\_Atlas\_Updates\2023\_Updates";

**run**;

\*Import Census Data;

**data** ses\_data;

set atlas.SES\_2023; \*<--- REPLACE WITH NEW CENSUS DATA;

**run**;

\*Drop Labels and Formats;

**proc** **datasets** lib=work;

modify ses\_data;

attrib \_all\_ label='';

**run**;

\*Import Cancer Data;

**data** cancer\_data;

set atlas.byarea\_county; \*<--- REPLACE WITH NEW CANCER DATA;

**run**;

\*Drop Labels and Formats;

**proc** **datasets** lib=work;

modify cancer\_data;

attrib \_all\_ label='';

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Step 2: If needed update code

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*

(B) Census Data from Am. Com. Survey

\*/

\*Rename vars;

**data** census\_2;

set ses\_data;

length Geo\_FIPS $ **5**;

Geo\_FIPS = FIPS;

rename

A04001\_001 = TotPop

PCT\_A04001\_003 = PWhitePCT

PCT\_A04001\_004 = PBlackPCT

PCT\_A04001\_005 = PNativePCT

PCT\_A04001\_006 = PAsianPCT

PCT\_A04001\_007 = PPacificPC

PCT\_A17005\_003 = PUnemploye

PCT\_A13003B\_002 = PPovertyPC

A14006\_001 = PMedHHInco;

PNoBachelo = sum(PCT\_A12001\_002 + PCT\_A12001\_003 + PCT\_A12001\_004);

**run**;

\*Keep relevant variables;

**data** census\_3;

set census\_2;

keep

Geo\_FIPS

PWhitePCT

PBlackPCT

PNativePCT

PAsianPCT

PPacificPC

PUnemploye

PPovertyPC

PMedHHInco

PNoBachelo

;

**run**;

/\*

(C) Cancer Data

\*/

\*Updating/renaming cancer sites;

**data** cancer\_data2;

set cancer\_data;

if SITE = 'All Cancer Sites Combined' then cancer2 = 'All Malignant Cancers';

if SITE = 'Cervix' then cancer2 = 'Cervix Uteri';

if SITE = 'Colon and Rectum' then cancer2 = 'Colorectum';

if SITE = 'Corpus and Uterus, NOS' then cancer2 = 'Corpus and Uterus, NOS';

if SITE = 'Esophagus' then cancer2 = 'Esophagus ';

if SITE = 'Hodgkin Lymphoma' then cancer2 = 'Lymphoma';

if SITE = 'Kidney and Renal Pelvis' then cancer2 = 'Kidney and Renal Pelvis';

if SITE = 'Leukemias' then cancer2 = 'Leukemia';

if SITE = 'Liver and Intrahepatic Bile Duct' then cancer2 = 'Liver and Intrahepatic Bile Duct';

if SITE = 'Lung and Bronchus' then cancer2 = 'Lung and Bronchus';

if SITE = 'Male and Female Breast' then cancer2 = 'Breast';

if SITE = 'Ovary' then cancer2 = 'Ovary';

if SITE = 'Pancreas' then cancer2 = 'Pancreas';

if SITE = 'Prostate' then cancer2 = 'Prostate';

if SITE = 'Stomach' then cancer2 = 'Stomach';

if SITE = 'Urinary Bladder' then cancer2 = 'Urinary Bladder';

**run**;

\*Sorting;

**proc** **sort** data = cancer\_data2;

by Sex ;

**run**;

\*Create Fips Code;

**data** cancer\_data3;

set cancer\_data2;

format Crd\_Rate Age\_Rate Best12.;

Crd\_Rate = CRUDE\_RATE;

Age\_Rate = AGE\_ADJUSTED\_RATE;

length Geo\_FIPS $ **5**;

Geo\_FIPS = strip(scan(AREA,**2**,'()'));

**run**;

**proc** **sort** data=cancer\_data3;

by STATE ;

**run**;

/\*

5-Yr Incidence: Non-Sex Specific Sites (all sexes; IE: male and female breast cancer included)

\*/

/\* I need to select incidence only, and flip the data into a wide format\*/

**data** incidence2;

set cancer\_data3;

where EVENT\_TYPE = "Incidence" and sex = "Male and Female" and RACE = "All Races" ;

**run**;

**data** incidence2;

set incidence2;

where SITE = "All Cancer Sites Combined" or SITE = "Cervix" or SITE = "Colon and Rectum" or

SITE = "Corpus and Uterus, NOS" or SITE = "Esophagus" or SITE = "Hodgkin Lymphoma" or

SITE = "Kidney and Renal Pelvis" or SITE = "Leukemias" or SITE = "Liver and Intrahepatic Bile Duct" or

SITE = "Lung and Bronchus" or SITE = "Male and Female Breast" or SITE = "Male and Female Breast, <i>in situ</i>" or

SITE = "Non-Hodgkin Lymphoma" or SITE = "Ovary" or SITE = "Pancreas" or SITE = "Prostate" or SITE = "Stomach" or SITE = "Urinary Bladder";

**run**;

**proc** **sort** data=incidence2;

by Geo\_FIPS ;

**run**;

**PROC** **TRANSPOSE** data= incidence2 out= incidence\_wide ;

BY Geo\_FIPS ;

\*COPY Sex;

\*ID State\_county;

VAR SITE POPULATION COUNT Crd\_Rate Age\_Rate;

**RUN**;

**data** incidence\_wide\_ar;

set incidence\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "Age\_Rate";

**run**;

**data** incidence\_wide\_cnt;

set incidence\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "COUNT";

**run**;

\*Rename Columns tables ;

\*AR;

**data** incidence\_wide\_ar\_2 ;

set incidence\_wide\_ar;

IAll\_AR = COL1\***1** ;

ICRC\_AR = COL3\***1** ;

ILiver\_AR = COL9\***1** ;

ILung\_AR = COL10\***1**;

INHL\_AR = COL13\***1**;

IPancr\_AR = COL15\***1**;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL18;

**run**;

\*Count;

**data** incidence\_wide\_cnt\_2;

set incidence\_wide\_cnt;

IAll\_cnt = COL1\***1** ;

ICRC\_cnt = COL3\***1** ;

ILiver\_cnt = COL9\***1** ;

ILung\_cnt = COL10\***1**;

INHL\_cnt = COL13\***1**;

IPancr\_cnt = COL15\***1**;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL18;

**run**;

/\*

5-y incidence data - Sex Specific Sites (Include Single Sex Values Only. IE: Excluding Male Breast Cancer etc. )

\*/

**data** incidence\_mf;

set cancer\_data3;

where EVENT\_TYPE = "Incidence" and sex ^= "Male and Female" and RACE = "All Races";

**run**;

**data** incidence\_mf2;

set incidence\_mf;

where SITE = "Cervix" or SITE = "Corpus and Uterus, NOS" or SITE = "Female Breast" or SITE = "Female Breast, <i>in situ</i>"

or SITE = "Ovary" or SITE = "Prostate" ;

**run**;

**proc** **sort** data=incidence\_mf2;

by Geo\_FIPS ;

**run**;

**PROC** **TRANSPOSE** data=incidence\_mf2 out=incidence\_mf\_wide ;

BY Geo\_FIPS ;

\*COPY Sex;

\*ID State\_county;

VAR SITE POPULATION COUNT Crd\_Rate Age\_Rate;

**RUN**;

**data** incidence\_mf\_wide\_ar;

set incidence\_mf\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "Age\_Rate";

**run**;

**data** incidence\_mf\_wide\_cnt;

set incidence\_mf\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "COUNT";

**run**;

\*Renaming columns for tables ;

\*AR;

**data** incidence\_mf\_wide\_ar\_2 ;

set incidence\_mf\_wide\_ar;

ICervix\_AR= COL1\***1** ;

ICorpUte\_AR= COL2\***1** ;

IFemBC\_AR = COL3\***1** ;

IFemBCis\_AR= COL4\***1** ;

IOvary\_AR = COL5\***1** ;

Iprost\_AR= COL6\***1** ;

;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL6 ;

**run**;

\*Count;

**data** incidence\_mf\_wide\_cnt\_2;

set incidence\_mf\_wide\_cnt;

ICervix\_cnt= COL1\***1** ;

ICorpUte\_cnt= COL2\***1** ;

IFemBC\_cnt = COL3\***1** ;

IFemBCis\_cnt= COL4\***1** ;

IOvary\_cnt = COL5\***1** ;

Iprost\_cnt= COL6\***1** ;

;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL6 ;

**run**;

/\*

5-Yr mortality: Non-Sex Specific Sites (all sexes; IE: male and female breast cancer included)

\*/

/\* I need to select mortality only, and flip the data into a wide format\*/

**data** mortality2;

set cancer\_data3;

where EVENT\_TYPE = "Mortality" and sex = "Male and Female" and RACE = "All Races" ;

**run**;

**data** mortality2;

set mortality2;

where SITE = "All Cancer Sites Combined" or SITE = "Cervix" or SITE = "Colon and Rectum" or

SITE = "Corpus and Uterus, NOS" or SITE = "Esophagus" or SITE = "Hodgkin Lymphoma" or

SITE = "Kidney and Renal Pelvis" or SITE = "Leukemias" or SITE = "Liver and Intrahepatic Bile Duct" or

SITE = "Lung and Bronchus" or SITE = "Male and Female Breast" or SITE = "Male and Female Breast, <i>in situ</i>" or

SITE = "Non-Hodgkin Lymphoma" or SITE = "Ovary" or SITE = "Pancreas" or SITE = "Prostate" or SITE = "Stomach" or SITE = "Urinary Bladder";

**run**;

**proc** **sort** data=mortality2;

by Geo\_FIPS ;

**run**;

**PROC** **TRANSPOSE** data= mortality2 out= mortality\_wide ;

BY Geo\_FIPS ;

\*COPY Sex;

\*ID State\_county;

VAR SITE POPULATION COUNT Crd\_Rate Age\_Rate;

**RUN**;

**data** mortality\_wide\_ar;

set mortality\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "Age\_Rate";

**run**;

**data** mortality\_wide\_cr;

set mortality\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "Crd\_Rate";

**run**;

**data** mortality\_wide\_pop;

set mortality\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "POPULATION";

**run**;

**data** mortality\_wide\_cnt;

set mortality\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "COUNT";

**run**;

\*Rename Columns for tables ;

\*AR;

**data** mortality\_wide\_ar\_2 ;

set mortality\_wide\_ar;

MAll\_AR = COL1\***1** ;

MCRC\_AR = COL3\***1** ;

MLiver\_AR = COL9\***1** ;

MLung\_AR = COL10\***1**;

MNHL\_AR = COL13\***1**;

MPancr\_AR = COL15\***1**;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL18;

**run**;

\*Count;

**data** mortality\_wide\_cnt\_2;

set mortality\_wide\_cnt;

MAll\_cnt = COL1\***1** ;

MCRC\_cnt = COL3\***1** ;

MLiver\_cnt = COL9\***1** ;

MLung\_cnt = COL10\***1**;

MNHL\_cnt = COL13\***1**;

MPancr\_cnt = COL15\***1**;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL18;

**run**;

/\*

5-y mortality data - Sex Specific Sites (Include Single Sex Values Only. IE: Excluding Male Breast Cancer etc. )

\*/

**data** mortality\_mf;

set cancer\_data3;

where EVENT\_TYPE = "Mortality" and sex ^= "Male and Female" and RACE = "All Races";

**run**;

**data** mortality\_mf2;

set mortality\_mf;

where SITE = "Cervix" or SITE = "Corpus and Uterus, NOS" or SITE = "Female Breast" or SITE = "Female Breast, <i>in situ</i>"

or SITE = "Ovary" or SITE = "Prostate" ;

**run**;

**proc** **sort** data=mortality\_mf2;

by Geo\_FIPS ;

**run**;

**PROC** **TRANSPOSE** data=mortality\_mf2 out=mortality\_mf\_wide ;

BY Geo\_FIPS ;

\*COPY Sex;

\*ID State\_county;

VAR SITE POPULATION COUNT Crd\_Rate Age\_Rate;

**RUN**;

**data** mortality\_mf\_wide\_ar;

set mortality\_mf\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "Age\_Rate";

**run**;

**data** mortality\_mf\_wide\_cr;

set mortality\_mf\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "Crd\_Rate";

**run**;

**data** mortality\_mf\_wide\_pop;

set mortality\_mf\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "POPULATION";

**run**;

**data** mortality\_mf\_wide\_cnt;

set mortality\_mf\_wide;

where \_NAME\_ = "SITE" or \_NAME\_ = "COUNT";

**run**;

\*Renaming columns for tables ;

\*AR;

**data** mortality\_mf\_wide\_ar\_2 ;

set mortality\_mf\_wide\_ar;

MCervix\_AR = COL1\***1**;

MCorpUte\_AR = COL2\***1**;

MFemBC\_AR = COL3\***1**;

MOvary\_AR = COL4\***1**;

Mprost\_AR = COL5\***1**;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL5 ;

**run**;

\*Count;

**data** mortality\_mf\_wide\_cnt\_2;

set mortality\_mf\_wide\_cnt;

MCervix\_cnt =COL1\***1**;

MCorpUte\_cnt =COL2\***1**;

MFemBC\_cnt =COL3\***1**;

MOvary\_cnt =COL4\***1**;

Mprost\_cnt =COL5\***1**;

if \_NAME\_ = "SITE" then delete ;

drop \_NAME\_ COL1-COL5 ;

**run**;

/\*

(D)

- Merge Old Atlas, Census, and Cancer Data -

{ Old Atlas Data,

Census Data,

Incidence Allsexes, Incidence Single Sexes, Mortality Allsexes, Mortality Single Sexes }

\*/

**data** atlas\_new;

merge

atlas\_old2

census\_3

incidence\_wide\_ar\_2 incidence\_wide\_cnt\_2

incidence\_mf\_wide\_ar\_2 incidence\_mf\_wide\_cnt\_2

mortality\_wide\_ar\_2 mortality\_wide\_cnt\_2

mortality\_mf\_wide\_ar\_2 mortality\_mf\_wide\_cnt\_2;

by Geo\_FIPS;

\*Rename Geo\_FIPS to FIPS to match Old Atlas ;

rename Geo\_FIPS = FIPS ;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Step 3: Check table, variables names, types, formats, and frequencies. Make sure matches old atlas structure

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*Copy contents to excel and compare duplicate names etc. ;

**proc** **contents** data = atlas\_old;

**run**;

**proc** **contents** data = atlas\_new;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Step 4: Export final file with correct pathname to be merged with shapefile

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*

PROC EXPORT DATA= atlas\_new

OUTFILE= " [ \*\*\* PATHNAME GOES HERE \*\*\* ] \atlas\_new.dbf"

DBMS=DBF replace;

RUN;

\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Step 5: Make a copy of data processing code for next update and replace Old Atlas Dataset with updated atlas dataset

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/