/\*LOAD THE LIBRARY(S)\*/

libname Raw "Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw";

libname Derived "Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Derived";

options fmtsearch=(raw.iCare4me\_formats raw.iCare4me\_online\_formats raw.tracking\_fmt raw.trt\_fmt);

options nofmterr; **RUN**;

/\*IMPORT THE ZIPCODE DATASET\*/

**PROC** **IMPORT**

DATAFILE="Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\Zipcode Data iCare4Me- Summary-4-6-23 (2).xls"

OUT= Raw.zipcode\_ic4u

DBMS=xls

REPLACE;

GETNAMES=YES;

**RUN**;

**PROC** **IMPORT**

DATAFILE="Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\Zipcode Data iCare4Me- Summary-4-6-23 (2).xls"

OUT= Raw.zipcode\_ic4u\_REG

DBMS=xls

REPLACE;

GETNAMES=YES;

SHEET= "Race,ethnicity&gender";

**RUN**;

/\*PREPARE TO MERGE\*/

/\*FIRST STEP- LOOK AT THE CONTENTS OF THE DATASETS\*/

**PROC** **CONTENTS** data= Raw.zipcode\_ic4u;

**run**;

**PROC** **CONTENTS** data= Raw.zipcode\_ic4u\_REG;

**run**;

**PROC** **SORT** data= Raw.zipcode\_ic4u;

by cg\_zipcode;

**run**;

/\*NEXT STEP- SORT BOTH THE DATASET BY VARIABLE : "cg\_zipcode" \*/

**PROC** **SORT** data= Raw.zipcode\_ic4u\_REG;

by cg\_zipcode;

**run**;

/\*NEXT STEP- MERGE THE TWO SHEETS OF DATASET\*/

**Data** zipcode\_merge;

merge Raw.zipcode\_ic4u\_REG Raw.zipcode\_ic4u;

by cg\_zipcode;

**run**;

**PROC** **CONTENTS** data= zipcode\_merge;

**run**;

/\* CHECKED THE MERGED DATASET- "zipcode\_merge". IT CONTAINS DUPLICATE ZIPCODES AND MISSING DATA. SO, IN THE NEXT STEP WE DROP THE DUPLICATE ZIPCODES\*/

**PROC** **PRINT** data= zipcode\_merge;

**run**;

/\* SORTED THE DATA BY NO DUPLICATE AND DUPLICATE ZIPCODES TO SEPARATE OUT A NON-DUPLICATE DATASET\*/

**proc** **sort** data=zipcode\_merge nodupkey out=nodup\_zipcode dupout= dup\_zipcode;

by cg\_zipcode;

**run**;

/\*CHECKED THE NON-DUPLICATE DATASET\*/

**PROC** **PRINT** data= nodup\_zipcode;

**run**;

/\* QA FOR NON DUPLICATE DATASET \*/

**PROC** **PRINT** data= nodup\_zipcode;

where cg\_zipcode in ("08012" "19143" "19335");

**run**;

/\*CREATED AN OUTPUT OF SUMMARY STATISTICS AND PLOTS FOR THE ZIPCODE LEVEL DATASET\*/

ods rtf file="Q:\Barbara Riegel\Caregiver\_RO1\_2019\documents\output\Summary statistics, plots &SYSDATE..doc";

title "Summary statistics and plots";

/\*CREATAED SUMMARY STATISTICS FOR THE NEW MERGED NO DUPLICATE DATASET\*/

**PROC** **MEANS** data= nodup\_zipcode n nmiss min max mean std maxdec=**2** ;

var Bachelor\_s\_Degree\_or\_Higher

Median\_Household\_Income

Without\_Health\_Care\_coverage

Total\_Households\_in\_ZipCode

Bachelor\_s\_Degree\_or\_Higher

American\_Indian\_\_Alaska\_Native

Asian Black\_AA

Female

Hispanic\_Latino\_\_Any\_Race Male

Native\_Hawiaan\_\_Pacific\_\_\_Island

Not\_Hispanic\_Latino\_\_White\_Alone

Some\_Other\_Race

White

Employer\_Establishments

Employment\_Rate;

**run**;

**PROC** **UNIVARIATE** data= nodup\_zipcode PLOT;

var Bachelor\_s\_Degree\_or\_Higher Median\_Household\_Income Without\_Health\_Care\_coverage Total\_Households\_in\_ZipCode Bachelor\_s\_Degree\_or\_Higher American\_Indian\_\_Alaska\_Native Asian Black\_AA Female Hispanic\_Latino\_\_Any\_Race Male Native\_Hawiaan\_\_Pacific\_\_\_Island Not\_Hispanic\_Latino\_\_White\_Alone

Some\_Other\_Race White Employer\_Establishments Employment\_Rate;

**run**;

ods rtf close;

**Proc** **corr** data= nodup\_zipcode;

var Median\_Household\_Income Without\_Health\_Care\_coverage Total\_Households\_in\_ZipCode Bachelor\_s\_Degree\_or\_Higher American\_Indian\_\_Alaska\_Native Asian Black\_AA Female Hispanic\_Latino\_\_Any\_Race Male Native\_Hawiaan\_\_Pacific\_\_\_Island Not\_Hispanic\_Latino\_\_White\_Alone

Some\_Other\_Race White Employer\_Establishments Employment\_Rate;

**run**;

/\*CHECK THE CONTENTS OF ZIPCODE VARIABLE\*/

**PROC** **CONTENTS** data=derived.scored\_trt ;

**run**;

/\*IMPORTED TWO NEW DATASETS THAT IS THE CONNECTION LINK BETWEEN THE PARTICIPANT LEVEL DATASET TO ZIPCODE LEVEL DATASET\*/

**proc** **import** datafile= "Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\zip codes\ICare4MeOnline-QAZipcodes\_DATA\_2023-04-12\_0915.csv"

out= raw.zipcode\_1

dbms= csv replace;

guessingrows= max;

getnames= yes;

**run**;

**proc** **import** datafile= "Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\zip codes\ICare4Me-QAZipcodes\_DATA\_LABELS\_2023-04-12\_0919 (1).csv"

out= raw.zipcode\_2 (rename= zipcode=cg\_zipcode)

dbms= csv replace;

guessingrows= max;

getnames= yes;

**run**;

/\* STACKEED THE TWO SHEETS OF THE DATASET TOGATHER \*/

**data** zipcodes;

set raw.zipcode\_1 raw.zipcode\_2;

**run**;

**proc** **print** data=zipcodes;

**run**;

/\*PREPARE TO MERGE\*/

/\*FIRST STEP- LOOK AT THE CONTENTS OF THE DATASETS\*/

**proc** **contents** data= zipcodes; **run**;

**proc** **contents** data= raw.zipcode\_1; **run**;

**proc** **contents** data= raw.zipcode\_2; **run**;

/\* SOME OF THE ZIPCODES WERE IDENTIFIED AS 4 NUMBERS INSTEAD OF FIVE. SO, DECIDED TO ADD A ZERO INITIAL

TO THOSE ZIPCODES WITH ONLY 4 NUMBERS AND ELIMINATE THE MISSING DATA\*/

/\*FIRST A NEW DATASET IS CREATED FROM THE CONNECTION LINK- PARTICIPANT LEVEL DATASET TO ZIPCODE LEVEL DATASET\*/

**data** zipcode\_1\_clean;

set zipcodes;

/\*ELIMINATE THE MISSING DATA\*/

if cg\_zipcode = **.** then Delete;

/\*THE VARIABLE "cg\_zipcode" IS STORED AS A NUMERIC VARIABLE IN THE DATASET.

IT NEEDS TO BE CONVERTED INTO A CHARACTER VARIABLE BEFORE PROCEEDING WITH THE NEXT STEP\*/

cgzip\_clean=put(cg\_zipcode,**8.**);

len= countc(cgzip\_clean,,'d');

/\*ADD A ZERO TO THOSE ZIPCODES WITH ONLY 4 NUMBERS\*/

if countc(cgzip\_clean,,'d') = **4** then cgzip\_clean = cats("0",cgzip\_clean);

/\*Remove special characters/ blanks\*/

cgzip\_clean = compress(cgzip\_clean);

**run**;

**proc** **print** data= zipcode\_1\_clean ;

**run**;

/\*QA FOR THE ZIPCODE LEVEL DATA\*/

**proc** **freq** data = zipcode\_1\_clean;

tables cg\_zipcode\*cgzip\_clean\*len/ list missing;

**run**;

**proc** **contents** data= raw.zipcode\_1; **run**;

/\*proc sort data=nodup\_zipcode;\*/

/\*by cg\_zipcode;\*/

/\*run;\*/

/\*CREATED A NEW DATASET FROM THE PREVIOUSLY MERGED ZIPCODE LEVEL DATASET\*/

**data** nodupe\_zipcode\_sorted ;

set nodup\_zipcode;

cgzip\_clean = compress(cg\_zipcode);

**run**;

**proc** **contents** data=nodupe\_zipcode\_sorted;

**run**;

/\* SORTED THE DATA BY "cgzip\_clean" \*/

**proc** **sort** data=nodupe\_zipcode\_sorted;

by cgzip\_clean;

**run**;

/\*SORTED THE CLEANED DATASET BY "cgzip\_clean"\*/

**proc** **sort** data=zipcode\_1\_clean;

by cgzip\_clean;

**run**;

/\*MERGED THE NEW DATASET (nodupe\_zipcode\_sorted) CREATED FROM THE PREVIOUSLY MERGED ZIPCODE LEVEL DATASET (nodup\_zipcode)\*/

**data** merge\_zips dataonly idsonly;

merge nodupe\_zipcode\_sorted (in=dat drop=cg\_zipcode) zipcode\_1\_clean (in=ids);

by cgzip\_clean;

if dat= **1** and ids ne **1** then output dataonly;

if dat ne **1** and ids = **1** then output idsonly;

if dat = **1** and ids = **1** then output merge\_zips;

**run**;

/\*QA CHECK\*/

**proc** **freq** data= merge\_zips;

table cgzip\_clean/list missing;

**run**;

**proc** **freq** data= idsonly;

table cgzip\_clean/list missing;

**run**;

/\*CREATED AN OUTPUT OF SUMMARY STATISTICS AND PLOTS FOR THE DATASETS THAT CONNECTS PARTICIPANT LEVEL DATASET TO ZIPCODE LEVEL DATASET\*/

ods rtf file="Q:\Barbara Riegel\Caregiver\_RO1\_2019\documents\output\Ids no zip &sysdate..doc";

title "Ids with Missing Zipcode Information";

**proc** **print** data= idsonly;

var cgzip\_clean;

**run**;

ods rtf close;

**proc** **print** data=nodupe\_zipcode\_sorted;

**run**;

/\*checking the contents of NEW DATASET CREATED FROM THE CONNECTION LINK- PARTICIPANT LEVEL DATASET TO ZIPCODE LEVEL DATASET\*/

**proc** **contents** data= zipcode\_1\_clean; **run**;

**PROC** **SORT** data= Raw.zipcode\_ic4u\_REG;

by cg\_zipcode;

**run**;

**proc** **contents** data= derived.scored\_bl;**run**;

/\*CHECKED THE CONTENTS FOR THE PARTICIPANT TRACKING DATASET\*/

**proc** **contents** data= raw.participant\_tracking;**run**;

/\*CREATED A NEW DATASET FOR THE PARTICIPANT TRACKING\*/

**data** participant\_track;

set raw.participant\_tracking;

**run**;

/\*CHECK THE CONTENTS OF THE MERGED DATASET\*/

**proc** **contents** data= merge\_zips;**run**;

/\*A NEW DATASET IS CREATED FROM THE PREVIOUSLY MERGED ZIPCODE ONLY LEVEL DATASET\*/

/\*THE CAREGIVER ID VARIABLE- "dyad\_id" WAS FOUND AS A NUMERIC VARIABLE.

AS A NEXT STEP, THE CAREGIVER ID VARIABLE- "dyad\_id" WAS CONVERTED INTO A CHARACTER VARIABLE BEFORE PROCEEDING TO NEXT STEP\*/

**data** merge\_zip\_ch;

set merge\_zips;

chr\_dyad\_id=put(dyad\_id,**8.**);

drop dyad\_id;

**run**;

**proc** **contents** data =merge\_zip\_ch; **run**;

/\*A NEW DATASET IS CREATED AND COMPRESS FUNCTION USED TO REMOVE SPECIAL CHARACTERS FROM THE CAREGIVER ID VARIABLE\*/

**data** merge\_zip\_ch\_2;

set merge\_zip\_ch;

chr\_dyad\_id=compress(chr\_dyad\_id);

**run**;

**proc** **contents** data =merge\_zip\_ch\_2; **run**;

/\*SORTING THE DATASET BY NEW CREATED CAREGIVER ID VARIABLE\*/

**PROC** **SORT** data= merge\_zip\_ch\_2;

by chr\_dyad\_id;

**run**;

/\*proc print data =merge\_zip\_ch;\*/

/\*var chr\_dyad\_id;\*/

/\*run;\*/

/\*\*/

/\*proc print data =participant\_track2;\*/

/\*var chr\_dyad\_id;\*/

/\*run;\*/

/\*A NEW DATASET CREATED FROM PARTICIPANT LEVEL DATASET\*/

**data** participant\_track2 (rename=dyad\_id=chr\_dyad\_id);

set participant\_track;

**run**;

/\*SORTED THE PARTICIPANT LEVEL DATASET BY CAREGIVER ID VARIABLE\*/

**PROC** **SORT** data= participant\_track2;

by chr\_dyad\_id;

**run**;

/\*CHECKED THE CONTENTS OF PARTICIPANT LEVEL DATASET\*/

**proc** **contents** data =participant\_track2; **run**;

/\*MERGED PARTICIPANT LEVEL DATASET WITH ZIPCODE LEVEL DATASET\*/

**data** merged\_baseline ziponly idsonly participant\_tarck3;

merge merge\_zip\_ch\_2 (in=zip) participant\_track2 (in=ids);

by chr\_dyad\_id;

if zip = **1** and ids ne **1** then output ziponly;

if zip ne **1** and ids = **1** then output idsonly;

if zip = **1** and ids = **1** then output merged\_baseline;

if ids = **1** then output participant\_tarck3;

**run**;

**proc** **contents** data =merged\_baseline; **run**;

/\*QA FOR PARTICIPANT LEVEL DATASET\*/

**proc** **freq** data= participant\_tarck3;

tables condition /list missing;

**run**;

/\*CHECK CONTENTS FOR PARTICIPANT LEVEL DATASET\*/

**proc** **contents** data =participant\_tarck3; **run**;

/\*\*\*\*\*SUMMARY OF THE ZIPCODE MEASURES AT THE PARTICIPANT LEVEL\*\*\*\*\*\*\*/

ods rtf file="Q:\Barbara Riegel\Caregiver\_RO1\_2019\documents\output\summary at the participant level &sysdate..doc";

title "summary of the zipcode measures at the participant level";

/\*CREATED A DATASET TO REMOVE FORMATS\*/

**data** no\_format\_merged;

set merged\_baseline;

format \_all\_;

**run**;

/\*QA TO CHECK IF 1 REPRESENTS INTERVENTION AND 0 REPRESENTS HI\*/

**proc** **freq** data= no\_format\_merged;

tables condition /list missing;

**run**;

/\*variables used in proc means are from the dataset- data=nodupe\_zipcode\_sorted\*/

**PROC** **MEANS** data= merged\_baseline n nmiss min max mean std maxdec=**2** ;

var American\_Indian\_\_Alaska\_Native Asian Black\_AA White Native\_Hawiaan\_\_Pacific\_\_\_Island Some\_Other\_Race Total\_Households\_in\_ZipCode

Hispanic\_Latino\_\_Any\_Race Not\_Hispanic\_Latino\_\_White\_Alone Female Male Median\_Household\_Income Bachelor\_s\_Degree\_or\_Higher

Employment\_Rate Employer\_Establishments Without\_Health\_Care\_coverage;

Where condition =**1**;

**run**;

**PROC** **UNIVARIATE** data= merged\_baseline PLOT;

var American\_Indian\_\_Alaska\_Native Asian Black\_AA White Native\_Hawiaan\_\_Pacific\_\_\_Island Some\_Other\_Race Total\_Households\_in\_ZipCode

Hispanic\_Latino\_\_Any\_Race Not\_Hispanic\_Latino\_\_White\_Alone Female Male Median\_Household\_Income Bachelor\_s\_Degree\_or\_Higher

Employment\_Rate Employer\_Establishments Without\_Health\_Care\_coverage;

Where condition =**1**;

**run**;

ods rtf close;

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

/\* NEW \*/

/\* \*/

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

/\*IMPORTED TWO SHEETS OF A NEW ZIPCODE LEVEL DATASET\*/

**PROC** **IMPORT**

DATAFILE="Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\zip codes\Zipcode Data iCare4Me\_Summary\_4.10.23-GT.xls"

OUT= Raw.zipcode\_data

DBMS=xls

REPLACE;

GETNAMES=YES;

**RUN**;

**PROC** **IMPORT**

DATAFILE="Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\zip codes\Zipcode Data iCare4Me\_Summary\_4.10.23-GT.xls"

OUT= Raw.zipcode\_data\_REG

DBMS=xls

REPLACE;

GETNAMES=YES;

SHEET= "Race,ethnicity&gender";

**RUN**;

**data** derived.zipcode\_data;

set Raw.zipcode\_data;

**run**;

**data** derived.zipcode\_data\_REG;

set Raw.zipcode\_data\_REG;

**run**;

**PROC** **CONTENTS** data= derived.zipcode\_data;

**run**;

**PROC** **CONTENTS** data= derived.zipcode\_data\_REG;

**run**;

/\*SORTING THE TWO SHEETS OF THE DATASET\*/

**PROC** **SORT** data= derived.zipcode\_data;

by cg\_zipcode;

**run**;

**PROC** **SORT** data= derived.zipcode\_data\_REG;

by cg\_zipcode;

**run**;

/\*MERGING THE TWO SHEETS OF THE DATASET\*/

**Data** new\_zipcode\_merge;

merge derived.zipcode\_data derived.zipcode\_data\_REG;

by cg\_zipcode;

**run**;

/\*DUPLICATE AND MISSING VARIABLES FOUND\*/

**PROC** **PRINT** data= derived.zipcode\_data;

**run**;

**PROC** **PRINT** data= derived.zipcode\_data\_REG;

**run**;

/\*ELIMINATING THE DUPLICATE AND MISSING VALUES FROM THE DATASET\*/

**proc** **sort** data=new\_zipcode\_merge nodupkey out=new\_nodup\_zipcode dupout= dup\_zipcode\_data;

by cg\_zipcode;

**run**;

**PROC** **PRINT** data= new\_nodup\_zipcode;

**run**;

**PROC** **PRINT** data= new\_nodup\_zipcode;

where cg\_zipcode in ("08012" "19143" "19335");

**run**;

/\*COMPARE BETWEEN TWO DATASETS\*/

**proc** **compare** base=nodup\_zipcode compare=new\_nodup\_zipcode;

id cg\_zipcode;

**run**;

**proc** **contents** data= new\_nodup\_zipcode;

**run**;

/\*EXPORT SUMMARY STATISTICS AND PLOTS FROM THE DERIVED DATASET\*/

ods rtf file="Q:\Barbara Riegel\Caregiver\_RO1\_2019\documents\output\Updated Summary statistics, plots &SYSDATE..doc";

title "An Updated Summary statistics and plots";

**PROC** **MEANS** data= new\_nodup\_zipcode nolabels n nmiss min max mean std median p25 p75 maxdec=**2**;

var Bachelor\_s\_Degree\_or\_Higher Median\_Household\_Income Without\_Health\_Care\_coverage Total\_Households\_in\_ZipCode Bachelor\_s\_Degree\_or\_Higher American\_Indian\_\_Alaska\_Native Asian Black\_AA Female Hispanic\_Latino\_\_Any\_Race Male Native\_Hawiaan\_\_Pacific\_\_\_Island Not\_Hispanic\_Latino\_\_White\_Alone

Some\_Other\_Race White Employer\_Establishments Employment\_Rate Grocery\_Stores hospitals;

where cg\_zipcode not in ("", "cg\_zipcode");

**run**;

**PROC** **UNIVARIATE** data= new\_nodup\_zipcode PLOT;

var Bachelor\_s\_Degree\_or\_Higher Median\_Household\_Income Without\_Health\_Care\_coverage Total\_Households\_in\_ZipCode Bachelor\_s\_Degree\_or\_Higher American\_Indian\_\_Alaska\_Native Asian Black\_AA Female Hispanic\_Latino\_\_Any\_Race Male Native\_Hawiaan\_\_Pacific\_\_\_Island Not\_Hispanic\_Latino\_\_White\_Alone

Some\_Other\_Race White Employer\_Establishments Employment\_Rate Grocery\_Stores hospitals;

where cg\_zipcode not in ("", "cg\_zipcode");

**run**;

ods rtf close;

/\* USING A WHERE STATEMENT IN PROC PRINT TO CHECK THE NAMES OF THE MISSING VARIABLES \*/

**PROC** **PRINT** data= new\_nodup\_zipcode;

where Bachelor\_s\_Degree\_or\_Higher=**.**;

**run**;

**proc** **import** datafile= "Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\zip codes\ICare4MeOnline-QAZipcodes\_DATA\_2023-04-12\_0915.csv"

out= raw.zipcode\_1

dbms= csv replace;

guessingrows= max;

getnames= yes;

**run**;

**proc** **import** datafile= "Q:\Barbara Riegel\Caregiver\_RO1\_2019\data\Raw\zip codes\ICare4Me-QAZipcodes\_DATA\_LABELS\_2023-04-12\_0919 (1).csv"

out= raw.zipcode\_2 (rename= zipcode=cg\_zipcode)

dbms= csv replace;

guessingrows= max;

getnames= yes;

**run**;

/\*STACK THE TWO DATASETS TOGATHER\*/

**data** zipcodes;

set raw.zipcode\_1 raw.zipcode\_2;

**run**;

**proc** **print** data=zipcodes;

**run**;

**proc** **contents** data= zipcodes; **run**;

**proc** **contents** data= raw.zipcode\_1; **run**;

**proc** **contents** data= raw.zipcode\_2; **run**;

/\* adding zero to the zipcode with 4 num\*/

**data** zipcode\_1\_clean;

set zipcodes;

if cg\_zipcode = **.** then Delete;

/\*converting variable to a character variable\*/

cgzip\_clean=put(cg\_zipcode,**8.**);

len= countc(cgzip\_clean,,'d');

if countc(cgzip\_clean,,'d') = **4** then cgzip\_clean = cats("0",cgzip\_clean);

/\*Remove special characters/ blanks\*/

cgzip\_clean = compress(cgzip\_clean);

**run**;

**proc** **print** data= zipcode\_1\_clean ;

**run**;

/\*QA for zipcodes\*/

**proc** **freq** data = zipcode\_1\_clean;

tables cg\_zipcode\*cgzip\_clean\*len/ list missing;

**run**;

**proc** **contents** data= raw.zipcode\_1; **run**;

/\*proc sort data=nodup\_zipcode;\*/

/\*by cg\_zipcode;\*/

/\*run;\*/

**data** new\_nodup\_zipcode\_sorted ;

set new\_nodup\_zipcode;

/\*converting numeric variable to a character variable\*/

cgzip\_clean=put(cg\_zipcode,**8.**);

len= countc(cgzip\_clean,,'d');

if countc(cgzip\_clean,,'d') = **4** then cgzip\_clean = cats("0",cgzip\_clean);

/\*Remove special characters/ blanks\*/

cgzip\_clean = compress(cgzip\_clean);

if cgzip\_clean in (**.**) then Delete;

**run**;

**proc** **freq** data= new\_nodup\_zipcode\_sorted;

table cgzip\_clean/list missing;

**run**;

**proc** **contents** data=new\_nodup\_zipcode\_sorted;

**run**;

**proc** **sort** data=new\_nodup\_zipcode\_sorted;

by cgzip\_clean;

**run**;

**proc** **sort** data=zipcode\_1\_clean;

by cgzip\_clean;

**run**;

**data** new\_merge\_zips dataonly idsonly;

merge new\_nodup\_zipcode\_sorted (in=dat drop=cg\_zipcode) zipcode\_1\_clean (in=ids);

by cgzip\_clean;

if dat= **1** and ids ne **1** then output dataonly;

if dat ne **1** and ids = **1** then output idsonly;

if dat = **1** and ids = **1** then output new\_merge\_zips;

**run**;

/\*CONVERTING NUMERIC VARIABLES TO CHARACTER VARIABLES\*/

**data** new\_merge\_zips2;

set new\_merge\_zips;

dyadid\_clean=put(dyad\_id,**8.**);

drop dyad\_id;

/\*TOTAL\*/

total\_clean=put(total,**8.**);

drop total;

**run**;

**proc** **contents** data= new\_nodup\_zipcode\_sorted; **run**;

**proc** **contents** data=new\_merge\_zips2; **run**;

/\*QA\*/

**proc** **freq** data= new\_nodup\_zipcode\_sorted;

table dyad\_id/list missing;

**run**;

**proc** **freq** data= new\_merge\_zips;

table cgzip\_clean/list missing;

**run**;

**proc** **freq** data= idsonly;

table cgzip\_clean/list missing;

**run**;

**proc** **print** data= idsonly;

var cgzip\_clean;

**run**;

**PROC** **PRINT** DATA=dataonly; **RUN**;

**PROC** **PRINT** DATA= new\_merge\_zips; **RUN**;

**PROC** **SORT** data= Raw.zipcode\_ic4u\_REG;

by cg\_zipcode;

**run**;

**data** session\_analysis;

set derived.session\_analysis;

**run**;

/\*CONVERTING NUMERIC VARIABLES TO CHARACTER VARIABLES\*/

**data** session\_analysis;

set derived.session\_analysis;

if dyad\_idnum = **.** then Delete;

dyadid\_clean=put(dyad\_idnum,**8.**);

drop dyad\_idnum;

drop dyad\_id;

total\_clean=put(total,**8.**);

/\* REDCAP REPEAT INSTANCE = CONVERTING NUMERIC VARIABLES TO CHARACTER VARIABLES \*/

redcap\_rept\_ins\_clean=put(redcap\_repeat\_instance,**8.**);

drop redcap\_repeat\_instance;

**run**;

**proc** **contents** data= session\_analysis; **run**;

/\*SORT THE DATASET BY CAREGIVER ID\*/

**proc** **sort** data=session\_analysis;

by dyadid\_clean;

**run**;

**proc** **sort** data=new\_merge\_zips2;

by dyadid\_clean;

**run**;

**proc** **contents** data= new\_merge\_zips; **run**;

/\*\*\*\*MERGING PARTICIPANT LEVEL DATASET WITH ZIPCODE LEVEL DATASET\*\*\*\*\*/

**Data** new\_session\_merge session\_only zips\_only;

merge session\_analysis(in=session) new\_merge\_zips2 (in=zips);

by dyadid\_clean;

if session=**1** and zips=**0** then output session\_only;

if session = **0** and zips = **1** then output zips\_only;

if session = **1** and zips = **1** then output new\_session\_merge;

**run**;

**proc** **contents** data= new\_session\_merge; **run**;

**proc** **contents** data= zips\_only; **run**;

/\*QA ON NEW DATASET CREATED BY MERGING PARTICIPANT LEVEL DATASET WITH ZIPCODE LEVEL DATASET\*/

**proc** **freq** data=session\_analysis;

table dyadid\_clean/list missing;

**run**;

**proc** **freq** data=new\_merge\_zips2;

table dyadid\_clean/list missing;

**run**;

/\*CREATING A SUMMARY STATISTICS ON THE NEW DATASET CREATED BY MERGING PARTICIPANT LEVEL DATASET WITH ZIPCODE LEVEL DATASET\*/

/\*CREATED A DISTRIBUTION TABLE FOR THE CONTINUOUS VARIABLES OF INTEREST\*/

**PROC** **MEANS** data= new\_session\_merge nolabels n nmiss min max mean std median p25 p75 maxdec=**2**;

var bl\_SDH\_race

bl\_SDH\_income

total

Bachelor\_s\_Degree\_or\_Higher

Median\_Household\_Income

Without\_Health\_Care\_coverage

White

Employer\_Establishments

Employment\_Rate;

**run**;

**PROC** **UNIVARIATE** data= new\_session\_merge PLOT normal;

var bl\_SDH\_race

bl\_SDH\_income

total

Bachelor\_s\_Degree\_or\_Higher

Median\_Household\_Income

Without\_Health\_Care\_coverage

White

Employer\_Establishments

Employment\_Rate;

**run**;

/\*CREATED TABLES FOR THE CATEGORICAL VARIABLES- RACE AND INCOME\*/

**proc** **freq** data= new\_session\_merge;

table bl\_SDH\_race bl\_SDH\_income /list missing;

**run**;

/\*CREATING AN OUTPUT DOCUMENT OF ANALYSIS OF THE VARIABLES WITH ENGAGEMENT AS AN OUTCOME\*/

ods rtf file="Q:\Barbara Riegel\Caregiver\_RO1\_2019\documents\output\Geocoded measures engagement analysis &sysdate..doc";

title "Geocoded measures engagement analysis";

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*ANALYSIS\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **npar1way** data=new\_session\_merge anova;

var white;

class bl\_SDH\_race;

**run**;

ods rtf text="Interpretation: Analysis of variance (ANOVA ) reveals that the mean percentage of white residents within caregiversí zip codes

is significantly higher among white caregivers compared to non-white caregivers. Specifically,

among caregivers who are white, the average percent of white residents with in caregiversí

zip codes is 76 percent. Among caregivers who are non-white, the average percent of non-white residents in caregiversí

zip code is 40 percent.";

**proc** **npar1way** data=new\_session\_merge anova;

var Median\_Household\_Income;

class bl\_SDH\_income ;

**run**;

ods rtf text="Interpretation: Caregiversí zip code level median household income is not significantly associated with care givers self-reported

financial difficulty.";

/\*TO KNOW THE NAMES OF EACH TABLE GENERATED- CHECK LOG TO FIND THE NAME OF THE TABLE LOOKING FOR\*/

ods trace on;

/\*USING A SPEARMAN CORRELATION MATRIX (TO EVALUATE A NON-PARAMETRIC ASSOCIATION BETWEEN TWO CONTINUOUS VARIABLES- "TOTAL" VARIABLE THAT IS AN ORDINAL VARIABLE)\*/

**proc** **corr** data =new\_session\_merge spearman;

ods select SpearmanCorr;

var total;

with bl\_SDH\_race

bl\_SDH\_income

Bachelor\_s\_Degree\_or\_Higher

Median\_Household\_Income

Without\_Health\_Care\_coverage

White

Employer\_Establishments

Employment\_Rate;

**run**;

ods rtf text="Interpretation: From the spearman correlation table we can see that the participant race (r=-0.26), zip code level race (r=0.20), and participant income(r=0.19) exhibit weak,

though statistically significant association with participant engagement (p<0.05).

Participant race is strongly correlated with participant engagement.";

ods trace off;

ods rtf close;

ods rtf file="Q:\Barbara Riegel\Caregiver\_RO1\_2019\documents\output\summary EVALUATED A NON-PARAMETRIC ASSOCIATION &sysdate..doc";

/\*USING A SPEARMAN CORRELATION MATRIX WE EVALUATED A NON-PARAMETRIC ASSOCIATION BETWEEN ZIPCODE LEVEL RACE VARIABLE WITH PARTICIPANT LEVEL RACE VARIABLE

TO ASSESS MULTI-COLLINEARITY\*/

**proc** **corr** data =new\_session\_merge plots= matrix spearman;

var bl\_SDH\_race

bl\_SDH\_income

Median\_Household\_Income

White;

**run**;

/\*MULTIVATIATE MODEL ANALYSIS WITH THE ENGAGEMENT OUTCOME\*/

/\*FORMATED 1 AND \*/

**proc** **format**; value race\_ **1**="non-white" **0**="white"; **run**;

**proc** **glm** data=new\_session\_merge plots=all;

class bl\_SDH\_race bl\_SDH\_income;

model total= bl\_SDH\_race bl\_SDH\_income Median\_Household\_Income white /solution;

lsmeans bl\_SDH\_race;

format bl\_SDH\_race race\_.;

**quit**;

/\*CTEATING A REGRESSION TREE WITH THE 4 PREDICTOR VARIABLES AND ENGAGEMENT AS OUTCOME (A CONTINUOUS VARIABLE)\*/

**proc** **hpsplit** data=new\_session\_merge seed=**123** cvmodelfit;

class bl\_SDH\_race bl\_SDH\_income;

model total= bl\_SDH\_race bl\_SDH\_income Median\_Household\_Income white;

output out=hpsplout;

prune cc (leaves=**3**);

**run**;

/\*RECTEATING THE SAME REGRESSION TREE- EXPERIMENTING WITH THE NUMBER OF LEAVES\*/

**proc** **hpsplit** data=new\_session\_merge seed=**123** cvmodelfit;

class bl\_SDH\_race bl\_SDH\_income;

model total= bl\_SDH\_race bl\_SDH\_income Median\_Household\_Income white;

output out=hpsplout;

prune cc (leaves=**5**);

**run**;

/\*APPLYING THE SAME CODES TO RUN A REGRESSION TREE, RESTRICTING THE MINIMUM LEAF SIZE TO BE NO LESS THAN 11\*/

**proc** **hpsplit** data=new\_session\_merge seed=**123** cvmodelfit minleafsize=**11**;

class bl\_SDH\_race bl\_SDH\_income;

model total= bl\_SDH\_race bl\_SDH\_income Median\_Household\_Income white;

output out=hpsplout;

prune cc (leaves=**3**);

**run**;

**proc** **hpsplit** data=new\_session\_merge seed=**123** cvmodelfit minleafsize=**11**;

class bl\_SDH\_race bl\_SDH\_income;

model total= bl\_SDH\_race bl\_SDH\_income Median\_Household\_Income white;

output out=hpsplout;

prune cc (leaves=**5**);

**run**;

ods rtf close;

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

/\* \*/

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

**proc** **contents** data= new\_session\_merge; **run**;

**proc** **freq** data= new\_session\_merge;

table dyadid\_clean/list missing;

**run**;

**proc** **corr** data =new\_session\_merge plots= matrix spearman;

var bl\_SDH\_race

bl\_SDH\_income

Grocery\_Stores

Hospitals;

**run**;

/\*CREATED A NEW DATASET TO REPLACE MISSING VALUES OF GROCERY STORES AND HOSPITAL VARIABLES WITH ZERO\*/

**data** sessions\_cleaned\_gh;

set new\_session\_merge;

Grocery\_Stores\_new= Grocery\_Stores;

Hospitals\_new= Hospitals;

if Grocery\_Stores =**.** then Grocery\_Stores\_new= **0**;

if Hospitals =**.** then Hospitals\_new= **0**;

**run**;

/\*QA OF THE REPLACED MISSING VALUES OF GROCERY STORES AND HOSPITAL VARIABLES WITH ZERO\*/\*/

proc freq data= sessions\_cleaned\_gh;

table Grocery\_Stores\*Grocery\_Stores\_new/list missing;

table Hospitals\*Hospitals\_new/list missing;

**run**;

ods rtf file="Q:\Barbara Riegel\Caregiver\_RO1\_2019\documents\output\Grocery and hospital analysis &sysdate..doc";

title "GROCERY STORES AND HOSPITAL VARIABLES ANALYSIS";

ods trace on;

**proc** **npar1way** data=sessions\_cleaned\_gh anova wilcoxon;

var Grocery\_Stores\_new Hospitals\_new;

class bl\_SDH\_race ;

ods select ClassMeans WilcoxonBoxPlot;

**run**;

**proc** **npar1way** data=sessions\_cleaned\_gh anova wilcoxon;

var Grocery\_Stores\_new Hospitals\_new;

class bl\_SDH\_income ;

ods select ClassMeans WilcoxonBoxPlot;

**run**;

ods trace off;

ods rtf close;

/\* SUMMARY OF ANALYSIS:ANOVA AND WILCOXEN TEST BOTH REVEALS THAT THE GROCERY STORE VARIABLE IS SIGNIFICANT COMPARED WITH PARTICIPANT LEVEL RACE WHEREAS HOSPITAL VARIABLE IS INSIGNIFICANT WITH PARTICIPANT LEVEL RACE

fROM THE ANOVA TEST WE CAN INTERPRET THAT PARTICIPANT WHO ARE WHITE HAVE AN AVERAGE OF 10 GROCERY STORES IN THEIR ZIPCODES COMPARED TO PARTICIPANTS WHO ARE NONWHITE \*/

**proc** **npar1way** data=new\_session\_merge anova wilcox;

var Grocery\_Stores Hospitals;

class bl\_SDH\_race ;

**run**;

/\*SUMMARY OF ANALYSIS: ANOVA AND WILCOXEN TEST BOTH REVEALS THAT THE GROCERY STORE AND HOSPITAL VARIABLE ARE NOT-SIGNIFICANT WITH PARTICIPANT LEVEL INCOME\*/

**proc** **npar1way** data=new\_session\_merge anova wilcox;

var Grocery\_Stores Hospitals;

class bl\_SDH\_income ;

**run**;

/\*RECTEATING THE SAME REGRESSION TREE- EXPERIMENTING WITH THE NUMBER OF LEAVES\*/

**proc** **hpsplit** data= new\_session\_merge seed=**123** assignmissing = branch;

class bl\_SDH\_race bl\_SDH\_income;

model total= bl\_SDH\_race

bl\_SDH\_income

Median\_Household\_Income

white Grocery\_Stores

Hospitals

Bachelor\_s\_Degree\_or\_Higher

Median\_Household\_Income

Without\_Health\_Care\_coverage

Total\_Households\_in\_ZipCode

Bachelor\_s\_Degree\_or\_Higher

American\_Indian\_\_Alaska\_Native

Asian Black\_AA

Female

Hispanic\_Latino\_\_Any\_Race Male

Native\_Hawiaan\_\_Pacific\_\_\_Island

Not\_Hispanic\_Latino\_\_White\_Alone

Some\_Other\_Race

White

Employer\_Establishments

Employment\_Rate;;

output out=hpsplout;

**run**;

**proc** **hpsplit** data= new\_session\_merge seed=**123** assignmissing = branch;

class bl\_SDH\_race bl\_SDH\_income;

model total= bl\_SDH\_:;

output out=hpsplout;

prune cc;

**run**;