Data Wrangling QBS 181

Midterm

Instructions:

* You may collaborate on the homework but you must write it up yourselves.
* Please save the files in this format:
* Name of the file should be username(database)\_Midterm\_QBS181
* Please use both SQL and R. You can use R markdown to submit the code written in R. For SQL, please follow the same instructions as in the assignments

Visit the website <https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DIQ_I.htm#Data_Processing_and_Editing>

1. The DIQ\_I.xpt(will be uploaded on canvas) file has some problems with its data (e.g., missing values, numeric columns stored as chars, etc.) and need to be cleaned before further use.
2. List the data-related issues you see in this data set

So many NA’s in the data. Some variables should be combine such as DID060 and DIQ060U time which represent length times units to get meaningful value. All the variables stored as continuous variables.

1. How will you address each data-related issue?

In my perspective, I do not change any NA to 0, mean, or median. The reasons that I want to address are some NA's could not be impute such as for DID040, I only have the date which people who develop diabetes and others were NA's. Thus, those NA's were separate from the meaningful data which people who develop diabetes, people who do not develop diabetes were assigned as NA's and there is no real meaning to convert those NA's into other value. Multiply those two related columns and stored the results in the time length variables. Convert those continuous variables to categorical.

1. Give justification for why you chose a particular way to address each issue. For example, if you decide to address missing values by removing rows or filling empty data cells, justify your decision or if you want to create a PHI field like year of Birth
2. Clean the data by addressing each point listed in 1.

Verify that whether the counts of each code or value for various variables are correct as mentioned in the website

select \* from [blu].[data]

--create a new data1 from data

select \* into [blu].[data1] from [blu].[data]

alter table [blu].[data1] alter column [DID040] float

--DIQ010

select COUNT([DIQ010]) from [blu].[data]

where [DIQ010] = 1

select COUNT([DIQ010]) from [blu].[data]

where [DIQ010] = 2

select COUNT([DIQ010]) from [blu].[data]

where [DIQ010] = 3

select COUNT([DIQ010]) from [blu].[data]

where [DIQ010] = 7

select COUNT([DIQ010]) from [blu].[data]

where [DIQ010] = 9

select COUNT(\*)-count([DIQ010]) from [blu].[data]

--DID040

select COUNT([DID040]) from [blu].[data]

where [DID040] between '2' and '78'

select COUNT([DID040]) from [blu].[data]

where [DIQ010] = 80

select COUNT([DID040]) from [blu].[data]

where [DIQ010] = 666

select COUNT([DID040]) from [blu].[data]

where [DIQ010] = 777

select COUNT([DID040]) from [blu].[data]

where [DIQ010] = 999

--change 666 to 0.5 because 666 represents age less than 1, I want to predict mean 0.5 for all ages less than 1

update [blu].[data1] set [DID040] = 0.5

where [DID040] = 666

select AVG([DID040]) from [blu].[data1]

where DID040 != 999

--substitude 48.4 into 999 becuase 999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID040] = 48.4

where [DID040] = 999

--The reason that I keep NULL is that only people had diabetes have age that doctor told them, others don't have.

--Thus, I could not simply impute the mean or median as NULL

select [DID040] from [blu].[data]

--DIQ160

select COUNT([DIQ160]) from [blu].[data]

where [DIQ160] = 1

select COUNT([DIQ160]) from [blu].[data]

where [DIQ160] = 2

select COUNT([DIQ160]) from [blu].[data]

where [DIQ160] = 9

select COUNT(\*)-count([DIQ160]) from [blu].[data]

--DIQ170

select COUNT([DIQ170]) from [blu].[data]

where [DIQ170] = 1

select COUNT([DIQ170]) from [blu].[data]

where [DIQ170] = 2

select COUNT([DIQ170]) from [blu].[data]

where [DIQ170] = 9

select COUNT(\*)-count([DIQ170]) from [blu].[data]

--DIQ172

select COUNT([DIQ172]) from [blu].[data]

where [DIQ172] = 1

select COUNT([DIQ172]) from [blu].[data]

where [DIQ172] = 2

select COUNT([DIQ172]) from [blu].[data]

where [DIQ172] = 7

select COUNT([DIQ172]) from [blu].[data]

where [DIQ172] = 9

select COUNT(\*)-count([DIQ172]) from [blu].[data]

--DIQ175A

select COUNT([DIQ175A]) from [blu].[data]

where [DIQ175A] = 10

select COUNT([DIQ175A]) from [blu].[data]

where [DIQ175A] = 77

select COUNT([DIQ175A]) from [blu].[data]

where [DIQ175A] = 99

select COUNT(\*)-count([DIQ175A]) from [blu].[data]

--DIQ175B

select COUNT([DIQ175B]) from [blu].[data]

where [DIQ175B] = 11

select COUNT(\*)-count([DIQ175B]) from [blu].[data]

--DIQ175C

select COUNT([DIQ175C]) from [blu].[data]

where [DIQ175C] = 12

select COUNT(\*)-count([DIQ175C]) from [blu].[data]

--DIQ175D

select COUNT([DIQ175D]) from [blu].[data]

where [DIQ175D] = 13

select COUNT(\*)-count([DIQ175D]) from [blu].[data]

--DIQ175E

select COUNT([DIQ175E]) from [blu].[data]

where [DIQ175E] = 14

select COUNT(\*)-count([DIQ175E]) from [blu].[data]

--DIQ175F

select COUNT([DIQ175F]) from [blu].[data]

where [DIQ175F] = 15

select COUNT(\*)-count([DIQ175F]) from [blu].[data]

--DIQ175G

select COUNT([DIQ175G]) from [blu].[data]

where [DIQ175G] = 16

select COUNT(\*)-count([DIQ175G]) from [blu].[data]

--DIQ175H

select COUNT([DIQ175H]) from [blu].[data]

where [DIQ175H] = 17

select COUNT(\*)-count([DIQ175H]) from [blu].[data]

--DIQ175I

select COUNT([DIQ175I]) from [blu].[data]

where [DIQ175I] = 18

select COUNT(\*)-count([DIQ175I]) from [blu].[data]

--DIQ175J

select COUNT([DIQ175J]) from [blu].[data]

where [DIQ175J] = 19

select COUNT(\*)-count([DIQ175J]) from [blu].[data]

--DIQ175K

select COUNT([DIQ175K]) from [blu].[data]

where [DIQ175K] = 20

select COUNT(\*)-count([DIQ175K]) from [blu].[data]

--DIQ175L

select COUNT([DIQ175L]) from [blu].[data]

where [DIQ175L] = 21

select COUNT(\*)-count([DIQ175L]) from [blu].[data]

--DIQ175M

select COUNT([DIQ175M]) from [blu].[data]

where [DIQ175M] = 22

select COUNT(\*)-count([DIQ175M]) from [blu].[data]

--DIQ175N

select COUNT([DIQ175N]) from [blu].[data]

where [DIQ175N] = 23

select COUNT(\*)-count([DIQ175N]) from [blu].[data]

--DIQ175O

select COUNT([DIQ175O]) from [blu].[data]

where [DIQ175O] = 24

select COUNT(\*)-count([DIQ175O]) from [blu].[data]

--DIQ175P

select COUNT([DIQ175P]) from [blu].[data]

where [DIQ175P] = 25

select COUNT(\*)-count([DIQ175P]) from [blu].[data]

--DIQ175Q

select COUNT([DIQ175Q]) from [blu].[data]

where [DIQ175Q] = 26

select COUNT(\*)-count([DIQ175Q]) from [blu].[data]

--DIQ175R

select COUNT([DIQ175R]) from [blu].[data]

where [DIQ175R] = 27

select COUNT(\*)-count([DIQ175R]) from [blu].[data]

--DIQ175S

select COUNT([DIQ175S]) from [blu].[data]

where [DIQ175S] = 28

select COUNT(\*)-count([DIQ175S]) from [blu].[data]

--DIQ175T

select COUNT([DIQ175T]) from [blu].[data]

where [DIQ175T] = 29

select COUNT(\*)-count([DIQ175T]) from [blu].[data]

--DIQ175U

select COUNT([DIQ175U]) from [blu].[data]

where [DIQ175U] = 30

select COUNT(\*)-count([DIQ175U]) from [blu].[data]

--DIQ175V

select COUNT([DIQ175V]) from [blu].[data]

where [DIQ175V] = 31

select COUNT(\*)-count([DIQ175V]) from [blu].[data]

--DIQ175W

select COUNT([DIQ175W]) from [blu].[data]

where [DIQ175W] = 32

select COUNT(\*)-count([DIQ175W]) from [blu].[data]

--DIQ175X

update [blu].[data] set [DIQ175X] = NULL

where [DIQ175X] = 'NA'

select COUNT([DIQ175X]) from [blu].[data]

where [DIQ175X] = 33

select COUNT(\*)-count([DIQ175X]) from [blu].[data]

--DIQ180

select COUNT([DIQ180]) from [blu].[data]

where [DIQ180] = 1

select COUNT([DIQ180]) from [blu].[data]

where [DIQ180] = 2

select COUNT([DIQ180]) from [blu].[data]

where [DIQ180] = 7

select COUNT([DIQ180]) from [blu].[data]

where [DIQ180] = 9

select COUNT(\*)-count([DIQ180]) from [blu].[data]

--DIQ050

select COUNT([DIQ050]) from [blu].[data]

where [DIQ050] = 1

select COUNT([DIQ050]) from [blu].[data]

where [DIQ050] = 2

select COUNT([DIQ050]) from [blu].[data]

where [DIQ050] = 7

select COUNT([DIQ050]) from [blu].[data]

where [DIQ050] = 9

select COUNT(\*)-count([DIQ050]) from [blu].[data]

--DIQ060U

select COUNT([DIQ060U]) from [blu].[data]

where [DIQ060U] = 1

select COUNT([DIQ060U]) from [blu].[data]

where [DIQ060U] = 2

select COUNT([DIQ060U]) from [blu].[data]

where [DIQ060U] = 3

select COUNT([DIQ060U]) from [blu].[data]

where [DIQ060U] = 4

select COUNT(\*)-count([DIQ060U]) from [blu].[data]

update [blu].[data] set [DIQ060U] = 12

where [DIQ060U] = 2

--DID060

select COUNT([DID060]) from [blu].[data]

where [DID060] between '1' and '55'

select COUNT([DID060]) from [blu].[data]

where [DID060] = 666

select COUNT([DID060]) from [blu].[data]

where [DID060] = 777

select COUNT([DID060]) from [blu].[data]

where [DID060] = 999

select COUNT(\*)-count([DID060]) from [blu].[data]

alter table [blu].[data] alter column [DID060] float

alter table [blu].[data1] alter column [DID060] float

--change 666 to 0.5 because 666 represents age less than 1, I want to predict mean 0.5 for all ages less than 1

update [blu].[data1] set [DID060] = 0.5

where [DID060] = 666

select AVG([DID060]) from [blu].[data1]

--substitude 102.3 into 999 becuase 999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID060] = 102.3

where [DID060] = 999

--time multiply units and get the results store in DID060

update [blu].[data] set [DID060] = [DID060] \* [DIQ060U]

--DIQ070

select COUNT([DIQ070]) from [blu].[data]

where [DIQ070] = 1

select COUNT([DIQ070]) from [blu].[data]

where [DIQ070] = 2

select COUNT([DIQ070]) from [blu].[data]

where [DIQ070] = 7

select COUNT([DIQ070]) from [blu].[data]

where [DIQ070] = 9

select COUNT(\*) - COUNT([DIQ070]) from [blu].[data]

--DIQ230

select COUNT([DIQ230]) from [blu].[data]

where [DIQ230] = 1

select COUNT([DIQ230]) from [blu].[data]

where [DIQ230] = 2

select COUNT([DIQ230]) from [blu].[data]

where [DIQ230] = 3

select COUNT([DIQ230]) from [blu].[data]

where [DIQ230] = 4

select COUNT([DIQ230]) from [blu].[data]

where [DIQ230] = 5

select COUNT([DIQ230]) from [blu].[data]

where [DIQ230] = 7

select COUNT([DIQ230]) from [blu].[data]

where [DIQ230] = 9

select COUNT(\*) - COUNT([DIQ230]) from [blu].[data]

--[DIQ240]

select COUNT([DIQ240]) from [blu].[data]

where [DIQ240] = 1

select COUNT([DIQ240]) from [blu].[data]

where [DIQ240] = 2

select COUNT([DIQ240]) from [blu].[data]

where [DIQ240] = 7

select COUNT([DIQ240]) from [blu].[data]

where [DIQ240] = 9

select COUNT(\*) - COUNT([DIQ240]) from [blu].[data]

--DID250

select COUNT([DID250]) from [blu].[data]

where [DID250] between '1' and '60'

select COUNT([DID250]) from [blu].[data]

where [DID250] = 0

select COUNT([DID250]) from [blu].[data]

where [DID250] = 7777

select COUNT([DID250]) from [blu].[data]

where [DID250] = 9999

select COUNT(\*)-count([DID250]) from [blu].[data]

--change type to float in data1

alter table [blu].[data1] alter column [DID250] float

--calcute avg without 9999

select AVG([DID250]) from [blu].[data1]

where DID250 != 9999

--substitude 4.5 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID250] = 4.5

where [DID250] = 9999

--DID260

select COUNT([DID260]) from [blu].[data]

where [DID260] between '1' and '15'

select COUNT([DID260]) from [blu].[data]

where [DID260] = 0

select COUNT([DID260]) from [blu].[data]

where [DID260] = 777

select COUNT([DID260]) from [blu].[data]

where [DID260] = 999

select COUNT(\*)-count([DID260]) from [blu].[data]

select [DID260] from [blu].[data]

--DIQ260U

select COUNT([DIQ260U]) from [blu].[data]

where [DIQ260U] = 1

select COUNT([DIQ260U]) from [blu].[data]

where [DIQ260U] = 2

select COUNT([DIQ260U]) from [blu].[data]

where [DIQ260U] = 3

select COUNT([DIQ260U]) from [blu].[data]

where [DIQ260U] = 4

select COUNT(\*)-count([DIQ260U]) from [blu].[data]

alter table [blu].[data] alter column [DID260] float

alter table [blu].[data] alter column [DIQ260U] float

--convert weeks, months, and years to days

update [blu].[data] set [DIQ260U] = 365

where [DIQ260U] = 1

update [blu].[data] set [DIQ260U] = 52.1

where [DIQ260U] = 2

update [blu].[data] set [DIQ260U] = 12

where [DIQ260U] = 3

update [blu].[data] set [DIQ260U] = 1

where [DIQ260U] = 4

--time multiply unit and store in DID260

update [blu].[data] set [DID260] = [DID260] \* [DIQ260U]

--DIQ275

select COUNT([DIQ275]) from [blu].[data]

where [DIQ275] = 1

select COUNT([DIQ275]) from [blu].[data]

where [DIQ275] = 2

select COUNT([DIQ275]) from [blu].[data]

where [DIQ275] = 7

select COUNT([DIQ275]) from [blu].[data]

where [DIQ275] = 9

select COUNT(\*) - COUNT([DIQ275]) from [blu].[data]

--DIQ280

select COUNT([DIQ280]) from [blu].[data]

where [DIQ280] between '2' and '18.5'

select COUNT([DIQ280]) from [blu].[data]

where [DIQ280] = 777

select COUNT([DIQ280]) from [blu].[data]

where [DIQ280] = 999

select COUNT(\*)-count([DIQ280]) from [blu].[data]

--change type to float in data1

alter table [blu].[data1] alter column [DIQ280] float

alter table [blu].[data] alter column [DIQ280] float

--calcute avg without 999

update [blu].[data1] set [DIQ280] = NULL

where [DIQ280] = 777

select AVG([DIQ280]) from [blu].[data1]

where DIQ280 != 999

--substitude 7.3 into 999 becuase 999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DIQ280] = 7.3

where [DIQ280] = 999

--DIQ291

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 1

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 2

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 3

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 4

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 5

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 6

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 77

select COUNT([DIQ291]) from [blu].[data]

where [DIQ291] = 99

select COUNT(\*) - COUNT([DIQ291]) from [blu].[data]

--DIQ300S

select COUNT([DIQ300S]) from [blu].[data]

where [DIQ300S] between '80' and '201'

select COUNT([DIQ300S]) from [blu].[data]

where [DIQ300S] = 7777

select COUNT([DIQ300S]) from [blu].[data]

where [DIQ300S] = 9999

select COUNT(\*)-count([DIQ300S]) from [blu].[data]

--change type to float in data1 adn data

alter table [blu].[data1] alter column [DIQ300S] float

alter table [blu].[data] alter column [DIQ300S] float

--calcute avg without 9999

update [blu].[data1] set [DIQ300S] = NULL

where [DIQ300S] = 7777

select AVG([DIQ300S]) from [blu].[data1]

where DIQ300S != 9999

--substitude 130.5 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DIQ300S] = 130.5

where [DIQ300S] = 9999

--DIQ300D

select COUNT([DIQ300D]) from [blu].[data]

where [DIQ300D] between '17' and '251'

select COUNT([DIQ300D]) from [blu].[data]

where [DIQ300D] = 7777

select COUNT([DIQ300D]) from [blu].[data]

where [DIQ300D] = 9999

select COUNT(\*)-count([DIQ300D]) from [blu].[data]

--change type to float in data1 adn data

alter table [blu].[data1] alter column [DIQ300D] float

--calcute avg without 9999

update [blu].[data1] set [DIQ300D] = NULL

where [DIQ300D] = 7777

select AVG([DIQ300D]) from [blu].[data1]

where DIQ300D != 9999

--substitude 78.2 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DIQ300D] = 78.2

where [DIQ300D] = 9999

--DID310S

select COUNT([DID310S]) from [blu].[data]

where [DID310S] between '80' and '175'

select COUNT([DID310S]) from [blu].[data]

where [DID310S] = 6666

select COUNT([DID310S]) from [blu].[data]

where [DID310S] = 7777

select COUNT([DID310S]) from [blu].[data]

where [DID310S] = 9999

select COUNT(\*)-count([DID310S]) from [blu].[data]

--change type to float in data1 and data

alter table [blu].[data1] alter column [DID310S] float

--calcute avg without 9999

update [blu].[data1] set [DID310S] = NULL

where [DID310S] = 6666

update [blu].[data1] set [DID310S] = NULL

where [DID310S] = 7777

select AVG([DID310S]) from [blu].[data1]

where [DID310S] != 9999

--substitude 123.1 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID310S] = 123.1

where [DID310S] = 9999

--DID310D

select COUNT([DID310D]) from [blu].[data]

where [DID310D] between '18' and '140'

select COUNT([DID310D]) from [blu].[data]

where [DID310D] = 6666

select COUNT([DID310D]) from [blu].[data]

where [DID310D] = 7777

select COUNT([DID310D]) from [blu].[data]

where [DID310D] = 9999

select COUNT(\*)-count([DID310D]) from [blu].[data]

--change type to float in data1 adn data

alter table [blu].[data1] alter column [DID310D] float

--calcute avg without 9999

update [blu].[data1] set [DID310D] = NULL

where [DID310D] = 6666

update [blu].[data1] set [DID310D] = NULL

where [DID310D] = 7777

select AVG([DID310D]) from [blu].[data1]

where [DID310D] != 9999

--substitude 76.8 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID310D] = 76.8

where [DID310D] = 9999

--DID320

select COUNT([DID320]) from [blu].[data]

where [DID320] between '4' and '520'

select COUNT([DID320]) from [blu].[data]

where [DID320] = 6666

select COUNT([DID320]) from [blu].[data]

where [DID320] = 7777

select COUNT([DID320]) from [blu].[data]

where [DID320] = 9999

select COUNT(\*)-count([DID320]) from [blu].[data]

--change type to float in data1 adn data

alter table [blu].[data1] alter column [DID320] float

--calcute avg without 9999

update [blu].[data1] set [DID320] = NULL

where [DID320] = 5555

update [blu].[data1] set [DID320] = NULL

where [DID320] = 6666

update [blu].[data1] set [DID320] = NULL

where [DID320] = 7777

select AVG([DID320]) from [blu].[data1]

where [DID320] != 9999

--substitude 123 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID320] = 123

where [DID320] = 9999

--DID330

select COUNT([DID330]) from [blu].[data]

where [DID330] between '6' and '205'

select COUNT([DID330]) from [blu].[data]

where [DID330] = 6666

select COUNT([DID330]) from [blu].[data]

where [DID330] = 7777

select COUNT([DID330]) from [blu].[data]

where [DID330] = 9999

select COUNT(\*)-count([DID330]) from [blu].[data]

--change type to float in data1 adn data

alter table [blu].[data1] alter column [DID330] float

--calcute avg without 9999

update [blu].[data1] set [DID330] = NULL

where [DID330] = 6666

update [blu].[data1] set [DID330] = NULL

where [DID330] = 7777

select AVG([DID330]) from [blu].[data1]

where [DID330] != 9999

--substitude 116.2 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID330] = 116.2

where [DID330] = 9999

--DID341

select COUNT([DID341]) from [blu].[data]

where [DID341] between '1' and '34'

select COUNT([DID341]) from [blu].[data]

where [DID341] = 0

select COUNT([DID341]) from [blu].[data]

where [DID341] = 7777

select COUNT([DID341]) from [blu].[data]

where [DID341] = 9999

select COUNT(\*)-count([DID341]) from [blu].[data]

--change type to float in data1 adn data

alter table [blu].[data1] alter column [DID341] float

--calcute avg without 9999

update [blu].[data1] set [DID341] = NULL

where [DID341] = 7777

select AVG([DID341]) from [blu].[data1]

where [DID341] != 9999

--substitude 2.4 into 9999 becuase 9999 represent don't know, I make a prediction for don't know

update [blu].[data] set [DID341] = 2.4

where [DID341] = 9999

--DID350

select COUNT([DID350]) from [blu].[data]

where [DID350] between '1' and '20'

select COUNT([DID350]) from [blu].[data]

where [DID350] = 0

select COUNT([DID350]) from [blu].[data]

where [DID350] = 7777

select COUNT([DID350]) from [blu].[data]

where [DID350] = 9999

select COUNT(\*)-count([DID350]) from [blu].[data]

select [DID350] from [blu].[data]

--DIQ350U

select COUNT([DIQ350U]) from [blu].[data]

where [DIQ350U] = 1

select COUNT([DIQ350U]) from [blu].[data]

where [DIQ350U] = 2

select COUNT([DIQ350U]) from [blu].[data]

where [DIQ350U] = 3

select COUNT([DIQ350U]) from [blu].[data]

where [DIQ350U] = 4

select COUNT(\*)-count([DIQ350U]) from [blu].[data]

alter table [blu].[data] alter column [DID350] float

alter table [blu].[data] alter column [DIQ350U] float

--convert weeks, months, and years to days

update [blu].[data] set [DIQ350U] = 365

where [DIQ350U] = 1

update [blu].[data] set [DIQ350U] = 52.1

where [DIQ350U] = 2

update [blu].[data] set [DIQ350U] = 12

where [DIQ350U] = 3

update [blu].[data] set [DIQ350U] = 1

where [DIQ350U] = 4

--time multiply unit and store in DID350

update [blu].[data] set [DID350] = [DID350] \* [DIQ350U]

--DIQ360

select COUNT([DIQ360]) from [blu].[data]

where [DIQ360] = 1

select COUNT([DIQ360]) from [blu].[data]

where [DIQ360] = 2

select COUNT([DIQ360]) from [blu].[data]

where [DIQ360] = 3

select COUNT([DIQ360]) from [blu].[data]

where [DIQ360] = 4

select COUNT([DIQ360]) from [blu].[data]

where [DIQ360] = 5

select COUNT([DIQ360]) from [blu].[data]

where [DIQ360] = 7

select COUNT([DIQ360]) from [blu].[data]

where [DIQ360] = 9

select COUNT(\*) - COUNT([DIQ360]) from [blu].[data]

--DIQ080

select COUNT([DIQ080]) from [blu].[data]

where [DIQ080] = 1

select COUNT([DIQ080]) from [blu].[data]

where [DIQ080] = 2

select COUNT([DIQ080]) from [blu].[data]

where [DIQ080] = 7

select COUNT([DIQ080]) from [blu].[data]

where [DIQ080] = 9

select COUNT(\*) - COUNT([DIQ080]) from [blu].[data]

--rename all the columns

EXEC sp\_RENAME '[blu].[data].SEQN' , 'Repseqnum', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ010' , 'Dtolddiab', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID040' , 'Agediabe', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ160' , 'Prediabetes', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ170' , 'Healthrisk', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ172' , 'Feelatrisk', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175A', 'Familyhis', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175B' , 'Overweight', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175C' , 'Age', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175D' , 'Poordiet', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175E' , 'Race', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175F' , 'Babyover9lbs', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175G' , 'Lackphyact', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175H' , 'HBP', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175I' , 'HBS', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175J' , 'HC', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175K' , 'Hypo', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175L' , 'Exhunger', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175M' , 'Tinglin', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175N' , 'Bluvision', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175O' , 'Infatigue', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175P' , 'Anyoneatrisk', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175Q' , 'Drwarn', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175R' , 'Other', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175S' , 'Gesdiab', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175T' , 'Frequri', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175U' , 'Thirst', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175V' , 'Cravsweet', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175W' , 'Medi', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ175X' , 'POS', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ180' , 'Bloodtest', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ050' , 'Insulin', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID060' , 'Insulintime', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ060U' , 'Unitinsulintime', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ070' , 'Diapill', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ230' , 'Diaspec', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ240' , 'Drsee', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID250' , 'Drtime', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID260' , 'Checkblood', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ260U' , 'Unitcheckblood', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ275' , 'DrcheckA1C', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ280' , 'LastA1C', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ291' , 'DrA1C', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ300S' , 'SBP', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ300D' , 'DBP', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID310S' , 'DrSBP', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID310D' , 'DrDBP', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID320' , 'LDL', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID330' , 'DrLDL', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID341' , 'Drcheckfeet', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DID350' , 'Checkfeet', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ350U' , 'Unitcheckfeet', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ360' , 'Pupils', 'COLUMN'

EXEC sp\_RENAME '[blu].[data].DIQ080' , 'Diabeyes', 'COLUMN'