

Coding Session: Data Wrangling 1

Environmental Data Analytics | Rachel Gonsenhausner

Spring 2020

Setup

Changing column headings for clean dataset join later

```
#Changing column headings
```

```
colnames(Arsenic)
```

```
## [1] "stateFIPS"          "State"
## [3] "countyFIPS"         "County"
## [5] "Year"               "Value"
## [7] "Data.Comment"       "Maximum.Contaminant.Level"
## [9] "PWS.ID"             "CWS.Name"
## [11] "Population.Served"  "Maximum.Contaminant.Level.1"
```

```
head(Arsenic)
```

```
## stateFIPS State countyFIPS County Year Value Data.Comment
## 1 4 Arizona 4001 Apache 2006 6.00 NA
## 2 4 Arizona 4001 Apache 2006 2.00 NA
## 3 4 Arizona 4001 Apache 2007 1.50 NA
## 4 4 Arizona 4001 Apache 2007 4.41 NA
## 5 4 Arizona 4001 Apache 2007 1.50 NA
## 6 4 Arizona 4001 Apache 2007 1.50 NA
## Maximum.Contaminant.Level PWS.ID CWS.Name
## 1 Less than or Equal to MCL (2) AZ0401018 LIVCO WATER COMPANY
## 2 Less than or Equal to MCL (2) AZ0401047 LORD AZ WATER 2 - WINCHESTER TRAILS
## 3 Not Detected (3) AZ0401001 ALPINE DWID
## 4 Less than or Equal to MCL (2) AZ0401004 EAGAR TOWN OF
## 5 Not Detected (3) AZ0401039 OJO BONITO ESTATES DWID
## 6 Not Detected (3) AZ0401010 PINECREST WATER COMPANY
## Population.Served Maximum.Contaminant.Level.1
## 1 745 Maximum Contaminant Level: Less than or Equal to MCL
## 2 366 Maximum Contaminant Level: Less than or Equal to MCL
## 3 582 Maximum Contaminant Level: Not Detected
## 4 5000 Maximum Contaminant Level: Less than or Equal to MCL
## 5 53 Maximum Contaminant Level: Not Detected
## 6 73 Maximum Contaminant Level: Not Detected
```

```
summary(Arsenic)
```

```
## stateFIPS State countyFIPS County
## Min. : 4.0 California :20353 Min. : 4001 Washington : 1983
## 1st Qu.:12.0 New York :13014 1st Qu.:12083 Los Angeles : 1955
## Median :29.0 Pennsylvania:13010 Median :29133 Jefferson : 1807
```

```

## Mean :28.6 Florida :12249 Mean :28658 Marion : 1734
## 3rd Qu.:42.0 Washington :11518 3rd Qu.:42019 Orange : 1716
## Max. :55.0 Missouri : 9180 Max. :55141 Hillsborough: 1679
## (Other) :83935 (Other) :152385
## Year Value Data.Comment
## Min. :1999 0.50 :31557 Mode:logical
## 1st Qu.:2004 1.00 :23465 NA's:163259
## Median :2009 2.50 :10740
## Mean :2009 0.01 : 8238
## 3rd Qu.:2013 1.50 : 7927
## Max. :2018 0.70 : 6825
## (Other):74507
## Maximum.Contaminant.Level PWS.ID
## Greater than MCL (1) : 4914 CA0110001: 20
## Less than or Equal to MCL (2): 49508 CA0110005: 20
## Not Detected (3) :108837 CA0210001: 20
## CA0510004: 20
## CA0510006: 20
## CA0510017: 20
## (Other) :163139
## CWS.Name Population.Served
## GREEN ACRES MHP : 54 Min. : 0
## COUNTRYSIDE MHP : 41 1st Qu.: 109
## MOUNTAIN VIEW MHP : 40 Median : 473
## WHISPERING PINES MHP : 39 Mean : 13635
## AUBURN WATER DISTRICT: 36 3rd Qu.: 3500
## CITY OF WESTMINSTER : 33 Max. :8271000
## (Other) :163016
## Maximum.Contaminant.Level.1
## Maximum Contaminant Level: Greater than MCL : 4914
## Maximum Contaminant Level: Less than or Equal to MCL: 49508
## Maximum Contaminant Level: Not Detected :108837
##
##
##
##

```

```

colnames(Arsenic)[6] <-c("Arsenic_ugL")
colnames(Arsenic)[8] <-c("MCL_Arsenic")
colnames(Arsenic)

```

```

## [1] "stateFIPS" "State"
## [3] "countyFIPS" "County"
## [5] "Year" "Arsenic_ugL"
## [7] "Data.Comment" "MCL_Arsenic"
## [9] "PWS.ID" "CWS.Name"
## [11] "Population.Served" "Maximum.Contaminant.Level.1"

```

```
colnames(Income)
```

```

## [1] "stateFIPS" "State" "countyFIPS" "County" "Year"
## [6] "Value" "Data.Comment" "X"

```

```
head(Income)
```

```
## stateFIPS State countyFIPS County Year Value Data.Comment X
```

```
## 1      1 Alabama      1001 Autauga 2000 42,463      NA NA
## 2      1 Alabama      1001 Autauga 2001 42,183      NA NA
## 3      1 Alabama      1001 Autauga 2002 42,841      NA NA
## 4      1 Alabama      1001 Autauga 2003 44,241      NA NA
## 5      1 Alabama      1001 Autauga 2004 45,379      NA NA
## 6      1 Alabama      1001 Autauga 2005 45,019      NA NA
```

```
summary(Income)
```

```
##      stateFIPS      State      countyFIPS      County
## Min.   : 1.00   Texas   : 4318   Min.    : 1001   Washington: 527
## 1st Qu.:18.00   Georgia: 2703   1st Qu.:18179   Jefferson  : 442
## Median :29.00   Virginia: 2275   Median :29177   Franklin   : 425
## Mean   :30.29   Kentucky: 2040   Mean    :30399   Jackson    : 408
## 3rd Qu.:45.00   Missouri: 1955   3rd Qu.:45083   Lincoln    : 408
## Max.   :56.00   Kansas  : 1785   Max.    :56045   Madison    : 340
##                (Other) :38334                (Other)   :50860
##      Year      Value      Data.Comment      X
## Min.   :2000   No Data: 17   Mode:logical   Mode:logical
## 1st Qu.:2004   38,916 : 10   NA's:53410     NA's:53410
## Median :2008   37,120 : 9
## Mean   :2008   38,407 : 9
## 3rd Qu.:2012   39,219 : 9
## Max.   :2016   31,434 : 8
##                (Other):53348
```

```
colnames(Income)[6] <-c("MHI")
```

```
colnames(PFAS)
```

```
## [1] "stateFIPS"      "State"          "countyFIPS"
## [4] "County"         "Year"           "Value"
## [7] "Data.Comment"   "Status"         "PWS.ID"
## [10] "CWS.Name"       "Population.Served" "Contaminant"
```

```
head(PFAS)
```

```
##      stateFIPS      State      countyFIPS      County      Year      Value      Data.Comment
## 1      1 Alabama      1021 Chilton 2013-2015      20.0      NA
## 2      1 Alabama      1033 Colbert 2013-2015      50.0      NA
## 3      1 Alabama      1033 Colbert 2013-2015      60.0      NA
## 4      1 Alabama      1043 Cullman 2013-2015     180.0      NA
## 5      1 Alabama      1043 Cullman 2013-2015      90.0      NA
## 6      1 Alabama      1049 DeKalb 2013-2015      70.0      NA
##      Status      PWS.ID      CWS.Name      Population.Served
## 1 Detected (2) AL0000213      Clanton Water Department      13500
## 2 Detected (2) AL0000314 Colbert County Rural Water System      10731
## 3 Detected (2) AL0000314 Colbert County Rural Water System      10731
## 4 Detected (2) AL0000413      VAW Water System, Inc.      29985
## 5 Detected (2) AL0000413      VAW Water System, Inc.      29985
## 6 Detected (2) AL0001422      Northeast Alabama Water System      42609
##      Contaminant
## 1      PFOA
## 2      PFOS
## 3      PFOA
## 4      PFOS
```

```
## 5      PFOA
## 6      PFOS
```

```
summary(PFAS)
```

```
##      stateFIPS      State      countyFIPS      County
## Min.   : 1.00    California   :35    Min.   : 1021    Orange      : 13
## 1st Qu.: 6.00    New Jersey  :24    1st Qu.: 6059    Los Angeles: 12
## Median :19.50    Alabama    :21    Median :19600    Bergen      : 7
## Mean   :21.77    Florida    :14    Mean   :21828    Washington : 7
## 3rd Qu.:36.00    North Carolina:11    3rd Qu.:36071    Bucks       : 6
## Max.   :55.00    Pennsylvania:10    Max.   :55131    El Paso     : 6
##              (Other)      :83              (Other)      :147
##      Year      Value      Data.Comment      Status
## 2013-2015:198  40.0 : 10    Mode:logical    Detected (2):198
##              50.0 : 10    NA's:198
##              20.0 : 9
##              43.0 : 6
##              21.0 : 5
##              30.0 : 5
##              (Other):153
##      PWS.ID      CWS.Name      Population.Served      Contaminant
## AL0000314: 2    Artesian Water Company: 2    Min.   : 129    PFOA:107
## AL0000413: 2    ATLANTIC CITY MUA      : 2    1st Qu.: 19426    PFOS: 91
## AL0000577: 2    CALAM - SUBURBAN      : 2    Median : 32573
## AL0000588: 2    Chatsworth            : 2    Mean   : 91716
## AL0000591: 2    CITY OF ANAHEIM       : 2    3rd Qu.: 87450
## AL0000783: 2    CITY OF LATHROP       : 2    Max.   :2100000
## (Other) :186    (Other)              :186
```

```
colnames(PFAS)[6] <-c("PFAS_ppt")
```

```
colnames(Trihalomethane)
```

```
## [1] "stateFIPS"      "State"
## [3] "countyFIPS"     "County"
## [5] "Year"           "Value"
## [7] "Data.Comment"   "Maximum.Contaminant.Level"
## [9] "PWS.ID"         "CWS.Name"
## [11] "Population.Served" "Maximum.Contaminant.Level.1"
```

```
head(Trihalomethane)
```

```
##      stateFIPS      State      countyFIPS      County      Year      Value      Data.Comment
## 1           4    Arizona          4001    Apache    2006    0.25              NA
## 2           4    Arizona          4001    Apache    2006    18.00             NA
## 3           4    Arizona          4001    Apache    2006    2.00              NA
## 4           4    Arizona          4001    Apache    2007    16.50             NA
## 5           4    Arizona          4001    Apache    2007    6.40              NA
## 6           4    Arizona          4001    Apache    2007    3.80              NA
##      Maximum.Contaminant.Level      PWS.ID      CWS.Name      Population.Served
## 1      Not Detected (3)    AZ0401056      ASPC APACHE              540
## 2 Less than or Equal to MCL (2)    AZ0401004      EAGAR TOWN OF              5000
## 3 Less than or Equal to MCL (2)    AZ0401012      ST JOHNS CITY OF              3800
## 4 Less than or Equal to MCL (2)    AZ0401004      EAGAR TOWN OF              5000
## 5 Less than or Equal to MCL (2)    AZ0401018    LIVCO WATER COMPANY              745
```

```
## 6 Less than or Equal to MCL (2) AZ0401012 ST JOHNS CITY OF 3800
## Maximum.Contaminant.Level.1
## 1 Maximum Contaminant Level: Not Detected
## 2 Maximum Contaminant Level: Less than or Equal to MCL
## 3 Maximum Contaminant Level: Less than or Equal to MCL
## 4 Maximum Contaminant Level: Less than or Equal to MCL
## 5 Maximum Contaminant Level: Less than or Equal to MCL
## 6 Maximum Contaminant Level: Less than or Equal to MCL
```

```
summary(Trihalomethane)
```

```
## stateFIPS State countyFIPS County
## Min. : 4.00 Florida : 19237 Min. : 4001 Polk : 2546
## 1st Qu.:20.00 New York : 17283 1st Qu.:20061 Washington : 2345
## Median :29.00 Pennsylvania: 16832 Median :29165 Marion : 2117
## Mean :30.17 Louisiana : 10334 Mean :30235 Hillsborough: 2093
## 3rd Qu.:42.00 Missouri : 9440 3rd Qu.:42003 Jefferson : 1940
## Max. :55.00 Iowa : 8896 Max. :55141 Orange : 1638
## (Other) :105595 (Other) :174938
## Year Value Data.Comment
## Min. :1999 0.25 : 6942 Mode:logical
## 1st Qu.:2007 1.00 : 6002 NA's:187617
## Median :2012 2.00 : 3427
## Mean :2011 0.00 : 2969
## 3rd Qu.:2015 2.50 : 2944
## Max. :2018 0.01 : 2774
## (Other):162559
## Maximum.Contaminant.Level PWS.ID
## Greater than MCL (1) : 5303 C00101140: 20
## Less than or Equal to MCL (2):154064 C00138045: 20
## Not Detected (3) : 28250 FL1030050: 20
## FL1030515: 20
## FL1660596: 20
## FL2100741: 20
## (Other) :187497
## CWS.Name Population.Served
## WHISPERING PINES MHP : 45 Min. : 0
## GREEN ACRES MHP : 42 1st Qu.: 224
## MOUNTAIN VIEW MHP : 39 Median : 990
## SUEZ WATER : 38 Mean : 11522
## FLORIDA GOVERNMENTAL UTILITY AUTHORITY: 35 3rd Qu.: 4805
## RICHLAND, CITY OF : 35 Max. :8271000
## (Other) :187383
## Maximum.Contaminant.Level.1
## Maximum Contaminant Level: Greater than MCL : 5303
## Maximum Contaminant Level: Less than or Equal to MCL:154064
## Maximum Contaminant Level: Not Detected : 28250
##
##
##
##
```

```
colnames(Trihalomethane)[6] <-c("TTHM_ugL")
colnames(Trihalomethane)[8] <-c("MCL_TTHM")
```

```
colnames(Uranium)
```

```
## [1] "stateFIPS"           "State"
## [3] "countyFIPS"          "County"
## [5] "Year"                "Value"
## [7] "Data.Comment"        "Maximum.Contaminant.Level"
## [9] "PWS.ID"              "CWS.Name"
## [11] "Population.Served"   "Maximum.Contaminant.Level.1"
```

```
head(Uranium)
```

```
## stateFIPS State countyFIPS County Year Value Data.Comment
## 1 4 Arizona 4001 Apache 2007 5.50 NA
## 2 4 Arizona 4001 Apache 2008 3.70 NA
## 3 4 Arizona 4001 Apache 2008 31.40 NA
## 4 4 Arizona 4001 Apache 2010 0.50 NA
## 5 4 Arizona 4001 Apache 2011 0.50 NA
## 6 4 Arizona 4001 Apache 2011 7.95 NA
## Maximum.Contaminant.Level PWS.ID CWS.Name
## 1 Less than or Equal to MCL (2) AZ0401013 SPRINGERVILLE TOWN OF
## 2 Less than or Equal to MCL (2) AZ0401004 EAGAR TOWN OF
## 3 Greater than MCL (1) AZ0401022 SANDERS SCHOOL DISTRICT
## 4 Not Detected (3) AZ0401013 SPRINGERVILLE TOWN OF
## 5 Not Detected (3) AZ0401056 ASPC APACHE
## 6 Less than or Equal to MCL (2) AZ0401004 EAGAR TOWN OF
## Population.Served Maximum.Contaminant.Level.1
## 1 1972 Maximum Contaminant Level: Less than or Equal to MCL
## 2 5000 Maximum Contaminant Level: Less than or Equal to MCL
## 3 350 Maximum Contaminant Level: Greater than MCL
## 4 1972 Maximum Contaminant Level: Not Detected
## 5 540 Maximum Contaminant Level: Not Detected
## 6 5000 Maximum Contaminant Level: Less than or Equal to MCL
```

```
summary(Uranium)
```

```
## stateFIPS State countyFIPS County
## Min. : 4.00 California :14019 Min. : 4001 Los Angeles : 1192
## 1st Qu.: 6.00 Pennsylvania : 3645 1st Qu.: 6083 Kern : 1152
## Median :20.00 New Hampshire: 2790 Median :20055 Rockingham : 855
## Mean :22.06 Connecticut : 2461 Mean :22118 San Bernardino: 732
## 3rd Qu.:36.00 Oregon : 2187 3rd Qu.:36071 Riverside : 678
## Max. :55.00 New Mexico : 2150 Max. :55141 Hillsborough : 654
## (Other) :15124 (Other) :37113
## Year Value Data.Comment
## Min. :1999 0.50 : 6009 Mode:logical
## 1st Qu.:2005 0.00 : 2673 NA's:42376
## Median :2009 1.00 : 1796
## Mean :2009 0.75 : 1461
## 3rd Qu.:2013 0.10 : 1087
## Max. :2018 2.00 : 747
## (Other):28603
## Maximum.Contaminant.Level PWS.ID
## Greater than MCL (1) : 1241 CA1910146: 20
## Less than or Equal to MCL (2):24964 CA1910163: 20
## Not Detected (3) :16171 CA2000593: 20
```

```
##                                CA2010007:  20
##                                CA3310031:  20
##                                CA3600270:  20
##                                (Other)   :42256
##                                CWS.Name   Population.Served
##  HILLCREST WATER DISTRICT      :  22   Min.      :    0
##  ALPINE WATER USERS ASSOCIATION :  20   1st Qu.:   95
##  ARROWBEAR PARK CWD            :  20   Median   :   300
##  CANAAN WATER DEPT             :  20   Mean     :  16116
##  GOLDEN STATE WATER MORONGO DELNORTE:  20   3rd Qu.:  2800
##  HILLVIEW WC-OAKHURST/SIERRA LAKES :  20   Max.    :4072307
##  (Other)                       :42254
##                                Maximum.Contaminant.Level.1
##  Maximum Contaminant Level: Greater than MCL      : 1241
##  Maximum Contaminant Level: Less than or Equal to MCL:24964
##  Maximum Contaminant Level: Not Detected         :16171
##
##
##
##
```

```
colnames(Uranium)[6] <-c("Uranium_ugL")
colnames(Uranium)[8] <-c("MCL_Uranium")
```

Changing classes prior to joining datasets

```
# Changing classes of variables before joining
class(Arsenic$Year)
```

```
## [1] "integer"
```

```
Arsenic$Year <- as.numeric(Arsenic$Year)
```

```
class(Income$Year)
```

```
## [1] "integer"
```

```
Income$Year <- as.numeric(Income$Year)
```

```
class(PFAS$Year)
```

```
## [1] "factor"
```

```
# Dates for PFAS given as a range of 2013-2015, so I am changing them all to 2014 so I can integrate in
```

```
PFAS$Year <- ((2013+2015)/2)
```

```
class(PFAS$Year)
```

```
## [1] "numeric"
```

```
class(Trihalomethane$Year)
```

```
## [1] "integer"
```

```
Trihalomethane$Year <- as.numeric(Trihalomethane$Year)
```

```
class(Uranium$Year)
```

```
## [1] "integer"
```

```
Uranium$Year <-as.numeric(Uranium$Year)
```

Joining and filtering data for processed dataset

```
# Joining datasets
```

```
Arsenic.Income <- left_join(Arsenic, Income)
```

```
## Joining, by = c("stateFIPS", "State", "countyFIPS", "County", "Year",  
## "Data.Comment")
```

```
## Warning: Column `State` joining factors with different levels, coercing to  
## character vector
```

```
## Warning: Column `County` joining factors with different levels, coercing to  
## character vector
```

```
Arsenic.Income.PFAS <- left_join(Arsenic.Income, PFAS)
```

```
## Joining, by = c("stateFIPS", "State", "countyFIPS", "County", "Year",  
## "Data.Comment", "PWS.ID", "CWS.Name", "Population.Served")
```

```
## Warning: Column `State` joining character vector and factor, coercing into  
## character vector
```

```
## Warning: Column `County` joining character vector and factor, coercing into  
## character vector
```

```
## Warning: Column `PWS.ID` joining factors with different levels, coercing to  
## character vector
```

```
## Warning: Column `CWS.Name` joining factors with different levels, coercing to  
## character vector
```

```
Arsenic.Income.PFAS.Trihalomethane <- left_join(Arsenic.Income.PFAS, Trihalomethane)
```

```
## Joining, by = c("stateFIPS", "State", "countyFIPS", "County",  
## "Year", "Data.Comment", "PWS.ID", "CWS.Name", "Population.Served",  
## "Maximum.Contaminant.Level.1")
```

```
## Warning: Column `State` joining character vector and factor, coercing into  
## character vector
```

```
## Warning: Column `County` joining character vector and factor, coercing into  
## character vector
```

```
## Warning: Column `PWS.ID` joining character vector and factor, coercing into  
## character vector
```

```
## Warning: Column `CWS.Name` joining character vector and factor, coercing into  
## character vector
```

```
Arsenic.Income.PFAS.Trihalomethane.Uranium <- left_join(Arsenic.Income.PFAS.Trihalomethane, Uranium)
```

```
## Joining, by = c("stateFIPS", "State", "countyFIPS", "County",  
## "Year", "Data.Comment", "PWS.ID", "CWS.Name", "Population.Served",  
## "Maximum.Contaminant.Level.1")
```

```
## Warning: Column `State` joining character vector and factor, coercing into  
## character vector
```

```
## Warning: Column `County` joining character vector and factor, coercing into
```



```
## character vector
## Warning: Column `PWS.ID` joining character vector and factor, coercing into
## character vector
## Warning: Column `CWS.Name` joining character vector and factor, coercing into
## character vector
# Filtering dataset down to just necessary columns
Water.Quality.And.Income.Processed <-
  Arsenic.Income.PFAS.Trihalomethane.Uranium %>%
  select(stateFIPS:Arsonic_ugL, PWS.ID:Population.Served, MHI, PFAS_ppt, TTHM_ugL, Uranium_ugL, MCL_TTHM)
```

Changing MCL column data for clearer visualiation later

```
#Arsenic
levels(Water.Quality.And.Income.Processed$MCL_Arsenic) <- c(levels(Water.Quality.And.Income.Processed$MCL_Arsenic) == 'Greater than or equal to 0.05 mg/L', 'Less than 0.05 mg/L')
Water.Quality.And.Income.Processed$MCL_Arsenic[Water.Quality.And.Income.Processed$MCL_Arsenic == 'Greater than or equal to 0.05 mg/L'] <- 'Greater than or equal to 0.05 mg/L'
Water.Quality.And.Income.Processed$MCL_Arsenic[Water.Quality.And.Income.Processed$MCL_Arsenic == 'Less than 0.05 mg/L'] <- 'Less than 0.05 mg/L'

levels(Water.Quality.And.Income.Processed$MCL_Arsenic) <- c(levels(Water.Quality.And.Income.Processed$MCL_Arsenic) == 'Greater than or equal to 0.05 mg/L', 'Less than 0.05 mg/L')
Water.Quality.And.Income.Processed$MCL_Arsenic[Water.Quality.And.Income.Processed$MCL_Arsenic == 'Greater than or equal to 0.05 mg/L'] <- 'Greater than or equal to 0.05 mg/L'
Water.Quality.And.Income.Processed$MCL_Arsenic[Water.Quality.And.Income.Processed$MCL_Arsenic == 'Less than 0.05 mg/L'] <- 'Less than 0.05 mg/L'

levels(Water.Quality.And.Income.Processed$MCL_Arsenic) <- c(levels(Water.Quality.And.Income.Processed$MCL_Arsenic) == 'Greater than or equal to 0.05 mg/L', 'Less than 0.05 mg/L')
Water.Quality.And.Income.Processed$MCL_Arsenic[Water.Quality.And.Income.Processed$MCL_Arsenic == 'Greater than or equal to 0.05 mg/L'] <- 'Greater than or equal to 0.05 mg/L'
Water.Quality.And.Income.Processed$MCL_Arsenic[Water.Quality.And.Income.Processed$MCL_Arsenic == 'Less than 0.05 mg/L'] <- 'Less than 0.05 mg/L'

#TTHM
levels(Water.Quality.And.Income.Processed$MCL_TTHM) <- c(levels(Water.Quality.And.Income.Processed$MCL_TTHM) == 'Greater than or equal to 0.1 mg/L', 'Less than 0.1 mg/L')
Water.Quality.And.Income.Processed$MCL_TTHM[Water.Quality.And.Income.Processed$MCL_TTHM == 'Greater than or equal to 0.1 mg/L'] <- 'Greater than or equal to 0.1 mg/L'
Water.Quality.And.Income.Processed$MCL_TTHM[Water.Quality.And.Income.Processed$MCL_TTHM == 'Less than 0.1 mg/L'] <- 'Less than 0.1 mg/L'

levels(Water.Quality.And.Income.Processed$MCL_TTHM) <- c(levels(Water.Quality.And.Income.Processed$MCL_TTHM) == 'Greater than or equal to 0.1 mg/L', 'Less than 0.1 mg/L')
Water.Quality.And.Income.Processed$MCL_TTHM[Water.Quality.And.Income.Processed$MCL_TTHM == 'Greater than or equal to 0.1 mg/L'] <- 'Greater than or equal to 0.1 mg/L'
Water.Quality.And.Income.Processed$MCL_TTHM[Water.Quality.And.Income.Processed$MCL_TTHM == 'Less than 0.1 mg/L'] <- 'Less than 0.1 mg/L'

levels(Water.Quality.And.Income.Processed$MCL_TTHM) <- c(levels(Water.Quality.And.Income.Processed$MCL_TTHM) == 'Greater than or equal to 0.1 mg/L', 'Less than 0.1 mg/L')
Water.Quality.And.Income.Processed$MCL_TTHM[Water.Quality.And.Income.Processed$MCL_TTHM == 'Greater than or equal to 0.1 mg/L'] <- 'Greater than or equal to 0.1 mg/L'
Water.Quality.And.Income.Processed$MCL_TTHM[Water.Quality.And.Income.Processed$MCL_TTHM == 'Less than 0.1 mg/L'] <- 'Less than 0.1 mg/L'

#Uranium
levels(Water.Quality.And.Income.Processed$MCL_Uranium) <- c(levels(Water.Quality.And.Income.Processed$MCL_Uranium) == 'Greater than or equal to 0.02 mg/L', 'Less than 0.02 mg/L')
Water.Quality.And.Income.Processed$MCL_Uranium[Water.Quality.And.Income.Processed$MCL_Uranium == 'Greater than or equal to 0.02 mg/L'] <- 'Greater than or equal to 0.02 mg/L'
Water.Quality.And.Income.Processed$MCL_Uranium[Water.Quality.And.Income.Processed$MCL_Uranium == 'Less than 0.02 mg/L'] <- 'Less than 0.02 mg/L'

levels(Water.Quality.And.Income.Processed$MCL_Uranium) <- c(levels(Water.Quality.And.Income.Processed$MCL_Uranium) == 'Greater than or equal to 0.02 mg/L', 'Less than 0.02 mg/L')
Water.Quality.And.Income.Processed$MCL_Uranium[Water.Quality.And.Income.Processed$MCL_Uranium == 'Greater than or equal to 0.02 mg/L'] <- 'Greater than or equal to 0.02 mg/L'
Water.Quality.And.Income.Processed$MCL_Uranium[Water.Quality.And.Income.Processed$MCL_Uranium == 'Less than 0.02 mg/L'] <- 'Less than 0.02 mg/L'

levels(Water.Quality.And.Income.Processed$MCL_Uranium) <- c(levels(Water.Quality.And.Income.Processed$MCL_Uranium) == 'Greater than or equal to 0.02 mg/L', 'Less than 0.02 mg/L')
Water.Quality.And.Income.Processed$MCL_Uranium[Water.Quality.And.Income.Processed$MCL_Uranium == 'Greater than or equal to 0.02 mg/L'] <- 'Greater than or equal to 0.02 mg/L'
Water.Quality.And.Income.Processed$MCL_Uranium[Water.Quality.And.Income.Processed$MCL_Uranium == 'Less than 0.02 mg/L'] <- 'Less than 0.02 mg/L'
```

Saving processed dataset for next session

```
write.csv(Water.Quality.And.Income.Processed, row.names = FALSE, file = "./Data/Processed/CDC_WaterQual.csv")
```