Summary Statistics

Melanie Erkman

3/12/2024

```r
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.1.3
```

```r
#install.packages("psych")
#install.packages("summarytools")
library(psych)
library(summarytools)
```

```
## Warning: package 'summarytools' was built under R version 4.1.3
```

```r
df <- read_excel("kansasv2.xlsx", na="NA")
```

```r
nrow(df)
```

```
## [1] 173140
```

```r
#View(df)
```

```r
colnames(df)
```

```
##  [1] "OrganizationIdentifier"
##  [2] "ActivityIdentifier"
##  [3] "ActivityTypeCode"
##  [4] "ActivityMediaName"
##  [5] "ActivityMediaSubdivisionName"
##  [6] "ActivityStartDate"
##  [7] "MonitoringLocationIdentifier"
##  [8] "MonitoringLocationName"
##  [9] "LocationLatitudeMeasure"
## [10] "LocationLongitudeMeasure"
## [11] "LocationZip"
## [12] "LocationCity"
## [13] "LocationCounty"
## [14] "Location State"
## [15] "HydrologicCondition"
## [16] "HydrologicEvent"
## [17] "ActivityLocation/LatitudeMeasure"
## [18] "ActivityLocation/LongitudeMeasure"
## [19] "SampleCollectionEquipmentName"
## [20] "ResultIdentifier"
## [21] "ResultDetectionConditionText"
## [22] "CharacteristicName"
## [23] "ResultMeasureValue"
## [24] "ResultMeasure/MeasureUnitCode"
## [25] "DataQuality/PrecisionValue"
## [26] "ResultCommentText"
## [27] "USGSPCode"
## [28] "SubjectTaxonomicName"
## [29] "ResultAnalyticalMethod/MethodName"
## [30] "ResultLaboratoryCommentText"
## [31] "DetectionQuantitationLimitTypeName"
## [32] "DetectionQuantitationLimitMeasure/MeasureValue"
## [33] "AvgGrossIncome_County"
```

```
#numeric_summary <- describe(df)
```

```
categorical_summary <- summary(df)
```

```
#print(numeric_summary)
print(categorical_summary)
```

```
##    OrganizationIdentifier ActivityIdentifier ActivityTypeCode
##    Length:173140          Length:173140      Length:173140
##    Class :character        Class :character   Class :character
##    Mode  :character        Mode  :character   Mode  :character
##
##
##
##
##    ActivityMediaName  ActivityMediaSubdivisionName ActivityStartDate
##    Length:173140      Length:173140                Min.   :2018-01-08 00:00:00
##    Class :character   Class :character             1st Qu.:2018-11-28 00:00:00
##    Mode  :character   Mode  :character             Median :2020-07-16 00:00:00
##                                                    Mean   :2020-08-11 03:28:36
##                                                    3rd Qu.:2022-03-08 00:00:00
##                                                    Max.   :2023-12-21 00:00:00
##
##    MonitoringLocationIdentifier MonitoringLocationName LocationLatitudeMeasure
##    Length:173140                Length:173140          Min.   :37.00
##    Class :character             Class :character       1st Qu.:37.88
##    Mode  :character             Mode  :character       Median :38.90
##                                                        Mean   :38.57
##                                                        3rd Qu.:39.20
##                                                        Max.   :40.00
##
##    LocationLongitudeMeasure  LocationZip     LocationCity       LocationCounty
##    Min.   :-102.01          Min.   :64114   Length:173140      Length:173140
##    1st Qu.: -97.42          1st Qu.:66439   Class :character   Class :character
##    Median : -96.91          Median :66869   Mode  :character   Mode  :character
##    Mean   : -96.58          Mean   :66799
##    3rd Qu.: -95.45          3rd Qu.:67147
##    Max.   : -94.61          Max.   :68943
##
##    Location State     HydrologicCondition HydrologicEvent
##    Length:173140      Length:173140       Length:173140
##    Class :character   Class :character    Class :character
##    Mode  :character   Mode  :character    Mode  :character
##
##
##
##
##    ActivityLocation/LatitudeMeasure ActivityLocation/LongitudeMeasure
##    Min.   :37.02                    Min.   :-101.71
##    1st Qu.:39.12                    1st Qu.: -96.59
##    Median :39.38                    Median : -95.78
##    Mean   :39.28                    Mean   : -96.09
##    3rd Qu.:39.68                    3rd Qu.: -95.64
##    Max.   :40.00                    Max.   : -94.61
##    NA's   :140388                   NA's   :140388
##    SampleCollectionEquipmentName ResultIdentifier    ResultDetectionConditionText
##    Length:173140                 Length:173140       Length:173140
##    Class :character              Class :character    Class :character
##    Mode  :character              Mode  :character    Mode  :character
```

```
##
##
##
##
##   CharacteristicName ResultMeasureValue ResultMeasure/MeasureUnitCode
##   Length:173140      Length:173140      Length:173140
##   Class :character    Class :character    Class :character
##   Mode  :character    Mode  :character    Mode  :character
##
##
##
##
##   DataQuality/PrecisionValue ResultCommentText   USGSPCode
##   Mode :logical              Length:173140       Min.   :    4
##   FALSE:242                  Class :character    1st Qu.:  631
##   TRUE :208                  Mode  :character    Median :61601
##   NA's :172690                                   Mean   :40767
##                                                  3rd Qu.:68646
##                                                  Max.   :99997
##                                                  NA's   :32752
##   SubjectTaxonomicName ResultAnalyticalMethod/MethodName
##   Length:173140        Length:173140
##   Class :character     Class :character
##   Mode  :character     Mode  :character
##
##
##
##
##   ResultLaboratoryCommentText DetectionQuantitationLimitTypeName
##   Length:173140               Length:173140
##   Class :character            Class :character
##   Mode  :character            Mode  :character
##
##
##
##
##   DetectionQuantitationLimitMeasure/MeasureValue AvgGrossIncome_County
##   Min.   :     0.00                              Min.   :   950
##   1st Qu.:     0.05                              1st Qu.: 16506
##   Median :     1.00                              Median : 35208
##   Mean   :    47.07                              Mean   : 49284
##   3rd Qu.:     6.00                              3rd Qu.: 75031
##   Max.   :240000.00                              Max.   :218289
##   NA's   :74214                                  NA's   :1680
```

#OrganizationIdentifier

```
summary(df$OrganizationIdentifier)
```

```
##    Length     Class      Mode
##    173140 character character
```

#ActivityIdentifier

```
summary(df$ActivityIdentifier)
```

```
##    Length     Class      Mode
##    173140 character character
```

#ActivityTypeCode

```
summary(df$ActivityTypeCode)
```

```
##    Length     Class      Mode
##    173140 character character
```

#ActivityMediaName

```
summary(df$ActivityMediaName)
```

```
##    Length     Class      Mode
##    173140 character character
```

#ActivityMediaSubdivisionName

```
summary(df$ActivityMediaSubdivisionName)
```

```
##    Length     Class      Mode
##    173140 character character
```

#ActivityStartDate

```
summary(df$ActivityStartDate)
```

```
##                      Min.               1st Qu.                Median
## "2018-01-08 00:00:00" "2018-11-28 00:00:00" "2020-07-16 00:00:00"
##                      Mean               3rd Qu.                  Max.
## "2020-08-11 03:28:36" "2022-03-08 00:00:00" "2023-12-21 00:00:00"
```

#MonitoringLocationIdentifier

```
summary(df$MonitoringLocationIdentifier)
```

```
##    Length     Class      Mode
##    173140 character character
```

#MonitoringLocationName

```
summary(df$MonitoringLocationName)
```

```
##    Length     Class      Mode
##    173140 character character
```

#ResultCommentText

```
summary(df$ResultCommentText)
```

```
##    Length     Class      Mode
##    173140 character character
```

#USGSPCode

```
describe(df$USGSPCode)
```

```
##    vars      n     mean      sd median  trimmed      mad min   max range  skew
## X1    1 140388 40766.83 34620.27  61601 39768.78 42924.24   4 99997 99993 -0.11
##    kurtosis   se
## X1    -1.69 92.4
```

```
summary(df$USGSPCode)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##       4     631   61601   40767   68646   99997   32752
```

#SubjectTaxonomicName

```
summary(df$SubjectTaxonomicName)
```

```
##    Length     Class      Mode
##    173140 character character
```

ResultAnalyticalMethod/MethodName

```
summary(df$`ResultAnalyticalMethod/MethodName`)
```

```
##    Length     Class      Mode
##    173140 character character
```

#ResultLaboratoryCommentText

```
summary(df$ResultLaboratoryCommentText)
```

```
##     Length     Class      Mode
##     173140 character character
```

#DetectionQuantitationLimitTypeName

```
summary(df$DetectionQuantitationLimitTypeName)
```

```
##     Length     Class      Mode
##     173140 character character
```

DetectionQuantitationLimitMeasure/MeasureValue

```
describe(df$`DetectionQuantitationLimitMeasure/MeasureValue`)
```

```
##     vars     n  mean      sd median trimmed  mad min    max   range   skew
## X1     1 98926 47.07 1800.59      1    3.65 1.47   0 240000 240000 123.34
##     kurtosis    se
## X1 16150.18 5.72
```

```
summary(df$`DetectionQuantitationLimitMeasure/MeasureValue`)
```

```
##      Min. 1st Qu.   Median     Mean 3rd Qu.      Max.     NA's
##      0.00    0.05     1.00    47.07    6.00 240000.00    74214
```

#AvgGrossIncome_County

```
describe(df$AvgGrossIncome_County)
```

```
##     vars      n  mean       sd median  trimmed      mad min    max  range skew
## X1     1 171460 49284 46000.78  35208 43601.57 42028.74 950 218289 217339 2.02
##     kurtosis      se
## X1      5.28 111.09
```

```
summary(df$AvgGrossIncome_County)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     950   16506   35208   49284   75031  218289    1680
```