Assignment 1.2

Veera Koppula

12/4/2021

The first assignment is a review of graphical and data anlalysis. This purpose of this assignment is to provide a refresher of R and/or Python. The assignment is divided into three sections. Using R and/or Python, complete the following steps.

1. Import, Plot, Summarize, and Save Data

###Using the US Bureau of Labor Statistics data, choose a dataset that interests you. Then generate summary statistics for 2 variables, plot some of the features (e.g., histograms, box plots, density plots, etc.) of several variables, and save the data locally as CSV files.

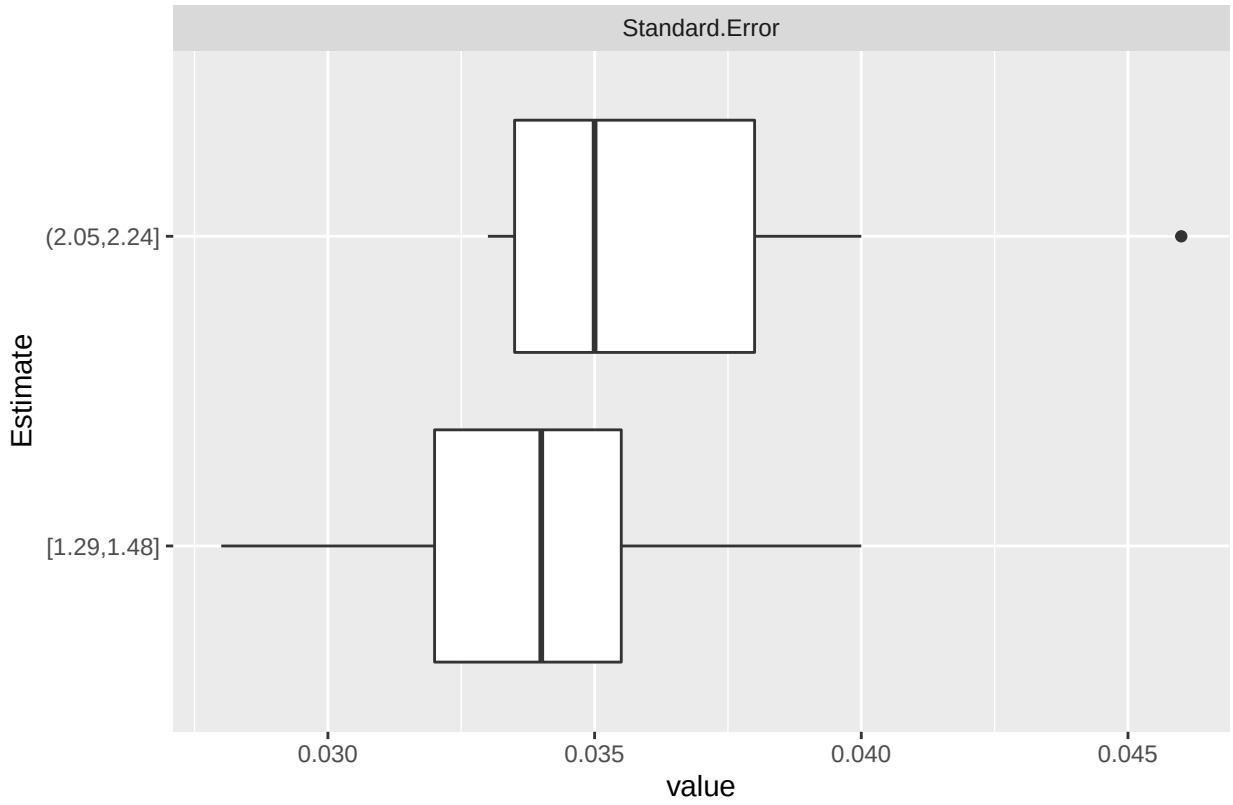
Import the Data saved from BLS data store

merging the two data files to create one file for Household Time Use data as timeuse

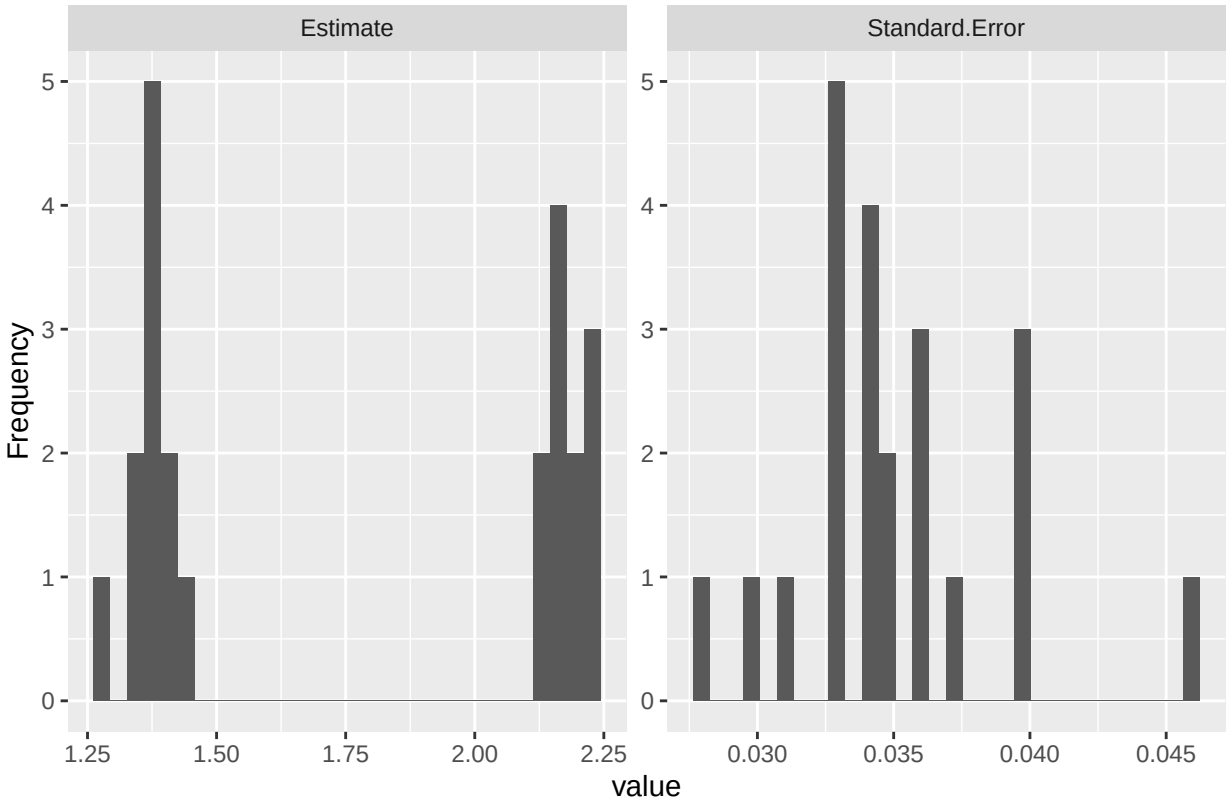
Validating structure of Data

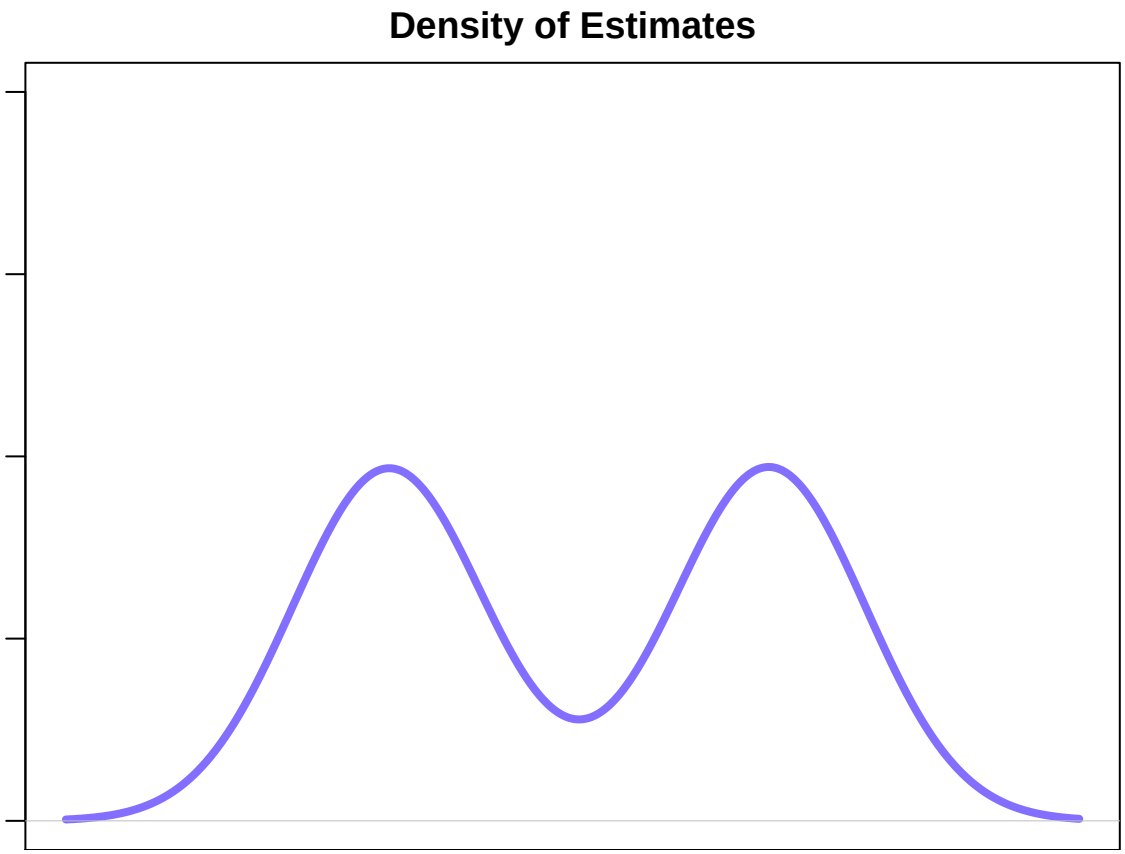
```
## 'data.frame':    22 obs. of  7 variables:
##  $ Year          : chr  "2009" "2010" "2011" "2012" ...
##  $ Period        : chr  "Annual" "Annual" "Annual" "Annual" ...
##  $ Estimate      : num  1.33 1.42 1.37 1.29 1.34 1.38 1.43 1.38 1.41 1.36 ...
##  $ Standard Error: num  0.03 0.037 0.028 0.031 0.033 0.036 0.04 0.034 0.034 0.035 ...
##  $ Gender        : chr  "Men" "Men" "Men" "Men" ...
##  $ Type          : chr  "Average hours per day" "Average hours per day" "Average hours per day" "Ave
##  $ Activity      : chr  "Household activities (includes travel)" "Household activities (includes tra
```

Generating summary statistics for the variables, I will print and add the details at the end of Step3 for the assignemnt Based on the data review there are 2 continuous variables, Estimate and Standard Error. Rest of the 5 variables are discrete. I would like to create some basic plots for the 2 continous vari-


Standard.Error
(2.05,2.24]
[1.29,1.48]
Estimate
0.030
0.035
0.040
0.045
value

ables.


Estimate
Standard.Error
Frequency
value
0
1
2
3
4
5
1.25
1.50
1.75
2.00
2.25
0.030
0.035
0.040
0.045



2. Explore Some Bivariate Relations

###Use the same dataset within the same website to explore some bivariate relations (e.g. bivariate plot, correlation, table cross table etc.)

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |              Expected N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  22
##
##
##                  | timeuse$`Standard Error`
## timeuse$Estimate |     0.028 |      0.03 |     0.031 |     0.033 |     0.034 |     0.035 |     0.036
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             1.29 |         0 |         0 |         1 |         0 |         0 |         0 |         0
```

```
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.045 |     0.045 |    20.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.045 |     0.000 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            1.33 |         0 |         1 |         0 |         0 |         0 |         0 |         0
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.045 |    20.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     0.045 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            1.34 |         0 |         0 |         0 |         1 |         0 |         0 |         0
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.045 |     0.045 |     0.045 |     2.627 |     0.182 |     0.091 |     0.136
##                 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.200 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.045 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            1.36 |         0 |         0 |         0 |         0 |         0 |         1 |         0
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     9.091 |     0.136
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.500 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.045 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            1.37 |         1 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |    20.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                 |     0.045 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            1.38 |         0 |         0 |         0 |         0 |         1 |         0 |         1
##                 |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     0.182 |     0.273
##                 |     0.091 |     0.091 |     0.091 |     0.455 |     1.114 |     0.182 |     1.939
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.500 |     0.000 |     0.500
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.250 |     0.000 |     0.333
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.045 |     0.000 |     0.045
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            1.39 |         0 |         0 |         0 |         1 |         0 |         0 |         0
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.045 |     0.045 |     0.045 |     2.627 |     0.182 |     0.091 |     0.136
##                 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.200 |     0.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.045 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            1.41 |         0 |         0 |         0 |         0 |         1 |         0 |         0
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                 |     0.045 |     0.045 |     0.045 |     0.227 |     3.682 |     0.091 |     0.136
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.250 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.045 |     0.000 |     0.000
```

```
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             1.42 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                  |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                  |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             1.43 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                  |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                  |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             2.14 |         0 |         0 |         0 |         1 |         1 |         0 |         0
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     0.182 |     0.273
##                  |     0.091 |     0.091 |     0.091 |     0.655 |     1.114 |     0.182 |     0.273
##                  |     0.000 |     0.000 |     0.000 |     0.500 |     0.500 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.200 |     0.250 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.045 |     0.045 |     0.000 |     0.000
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             2.16 |         0 |         0 |         0 |         1 |         0 |         0 |         0
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     0.182 |     0.273
##                  |     0.091 |     0.091 |     0.091 |     0.655 |     0.364 |     0.182 |     0.273
##                  |     0.000 |     0.000 |     0.000 |     0.500 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.200 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.045 |     0.000 |     0.000 |     0.000
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             2.17 |         0 |         0 |         0 |         0 |         0 |         1 |         0
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     0.182 |     0.273
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     3.682 |     0.273
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.500 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.500 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.045 |     0.000
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             2.19 |         0 |         0 |         0 |         0 |         0 |         0 |         1
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     0.182 |     0.273
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     0.182 |     1.939
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.500
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.333
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.045
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             2.23 |         0 |         0 |         0 |         1 |         0 |         0 |         0
##                  |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
##                  |     0.045 |     0.045 |     0.045 |     2.627 |     0.182 |     0.091 |     0.136
##                  |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.200 |     0.000 |     0.000 |     0.000
##                  |     0.000 |     0.000 |     0.000 |     0.045 |     0.000 |     0.000 |     0.000
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             2.24 |         0 |         0 |         0 |         0 |         1 |         0 |         1
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     0.364 |     0.182 |     0.273
##                  |     0.091 |     0.091 |     0.091 |     0.455 |     1.114 |     0.182 |     1.939
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.500 |     0.000 |     0.500
```

```
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.250 |     0.000 |     0.333
##                  |     0.000 |     0.000 |     0.000 |     0.000 |     0.045 |     0.000 |     0.045
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##     Column Total |         1 |         1 |         1 |         5 |         4 |         2 |         3
##                  |     0.045 |     0.045 |     0.045 |     0.227 |     0.182 |     0.091 |     0.136
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  150.5167     d.f. =  135     p =  0.1709053
##
##
##
```

150.5167 is a very high Chi value, meaning the data (Estimate and Standard Error) does not fit very well.

The p-value is relatively large which indicates weak evidence against the null hypothesis, so we fail to reject the null hypothesis.

Now let us take a look at the descriptive statistics of the erntire dataset stored earlier.

3. Organize a Data Report

Generate a summary report. Make sure to include: summary for every variable, structure and type of data elements, discuss four results of your data.

```
## timeuse
##
##  7  Variables      22  Observations
## ---------------------------------------------------------------------------
## Year
##        n  missing distinct
##       22        0       11
##
## lowest : 2009 2010 2011 2012 2013, highest: 2015 2016 2017 2018 2019
##
## Value       2009  2010  2011  2012  2013  2014  2015  2016  2017  2018  2019
## Frequency      2     2     2     2     2     2     2     2     2     2     2
## Proportion 0.091 0.091 0.091 0.091 0.091 0.091 0.091 0.091 0.091 0.091 0.091
## ---------------------------------------------------------------------------
## Period
##        n  missing distinct    value
##       22        0        1   Annual
##
## Value      Annual
## Frequency      22
## Proportion      1
## ---------------------------------------------------------------------------
## Estimate
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

```
##       22        0       16    0.997    1.779   0.4471    1.331    1.342
##      .25      .50      .75      .90      .95
##    1.380    1.785    2.170    2.226    2.240
##
## lowest : 1.29 1.33 1.34 1.36 1.37, highest: 2.16 2.17 2.19 2.23 2.24
##
## Value       1.29  1.33  1.34  1.36  1.37  1.38  1.39  1.41  1.42  1.43  2.14
## Frequency      1     1     1     1     1     2     1     1     1     1     2
## Proportion 0.045 0.045 0.045 0.045 0.045 0.091 0.045 0.045 0.045 0.045 0.091
##
## Value       2.16  2.17  2.19  2.23  2.24
## Frequency      2     2     2     1     2
## Proportion 0.091 0.091 0.091 0.045 0.091
## ---------------------------------------------------------------------------
## Standard Error
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##       22        0       10    0.978  0.03505 0.004186  0.03005  0.03120
##      .25      .50      .75      .90      .95
##  0.03300  0.03400  0.03600  0.04000  0.04000
##
## lowest : 0.028 0.030 0.031 0.033 0.034, highest: 0.035 0.036 0.037 0.040 0.046
##
## Value      0.028 0.030 0.031 0.033 0.034 0.035 0.036 0.037 0.040 0.046
## Frequency      1     1     1     5     4     2     3     1     3     1
## Proportion 0.045 0.045 0.045 0.227 0.182 0.091 0.136 0.045 0.136 0.045
## ---------------------------------------------------------------------------
## Gender
##        n  missing distinct
##       22        0        2
##
## Value        Men Women
## Frequency     11    11
## Proportion   0.5   0.5
## ---------------------------------------------------------------------------
## Type
##                      n               missing              distinct
##                     22                     0                     1
##                  value
## Average hours per day
##
## Value      Average hours per day
## Frequency                     22
## Proportion                     1
## ---------------------------------------------------------------------------
## Activity
##                                      n                                  missing
##                                     22                                        0
##                                distinct                                    value
##                                       1 Household activities (includes travel)
##
## Value      Household activities (includes travel)
## Frequency                                      22
## Proportion                                      1
## ---------------------------------------------------------------------------
```

Analysis:

Year - We have data between 2009-2019 and two instances per each year for Men/Women.

Period - The period is a single value within tha data set.

Estimate - The estimate has somewhat even distribution which is not surprising given the data has men and women in equal proportion.

Standard Error - Standard Error picks between 0.033, 0.034,0.035 and 0.040.

Gender - The dataset has equal weightage on gender.

Since the summary descriptive statistics (stats) is generated as a data frame, I would output that in an excel file. But first I want to check what is in it.

```
##                vars  n mean   sd median trimmed  mad  min   max range skew
## Year*             1 22 6.00 3.24   6.00    6.00 4.45 1.00 11.00 10.00 0.00
## Period*           2 22 1.00 0.00   1.00    1.00 0.00 1.00  1.00  0.00  NaN
## Estimate          3 22 1.78 0.42   1.79    1.78 0.60 1.29  2.24  0.95 0.00
## Standard Error    4 22 0.04 0.00   0.03    0.03 0.00 0.03  0.05  0.02 0.85
## Gender*           5 22 1.50 0.51   1.50    1.50 0.74 1.00  2.00  1.00 0.00
## Type*             6 22 1.00 0.00   1.00    1.00 0.00 1.00  1.00  0.00  NaN
## Activity*         7 22 1.00 0.00   1.00    1.00 0.00 1.00  1.00  0.00  NaN
##                kurtosis   se
## Year*             -1.38 0.69
## Period*             NaN 0.00
## Estimate          -2.06 0.09
## Standard Error     0.93 0.00
## Gender*           -2.09 0.11
## Type*               NaN 0.00
## Activity*           NaN 0.00
```

The variable names are converted into row names. I want to assign them to the first column of the dataf frame.

```
##         variable vars  n       mean          sd median    trimmed       mad
## 1          Year*    1 22 6.00000000 3.236694375  6.000 6.00000000 4.4478000
## 2        Period*    2 22 1.00000000 0.000000000  1.000 1.00000000 0.0000000
## 3       Estimate    3 22 1.77863636 0.417239174  1.785 1.77944444 0.6004530
## 4 Standard Error    4 22 0.03504545 0.003884981  0.034 0.03483333 0.0022239
## 5        Gender*    5 22 1.50000000 0.511766316  1.500 1.50000000 0.7413000
## 6          Type*    6 22 1.00000000 0.000000000  1.000 1.00000000 0.0000000
## 7      Activity*    7 22 1.00000000 0.000000000  1.000 1.00000000 0.0000000
##     min    max  range         skew   kurtosis          se
## 1 1.000 11.000 10.000  0.000000000 -1.3781405 0.6900655593
## 2 1.000  1.000  0.000          NaN        NaN 0.0000000000
## 3 1.290  2.240  0.950 -0.002616487 -2.0566792 0.0889556907
## 4 0.028  0.046  0.018  0.850990784  0.9303615 0.0008282806
## 5 1.000  2.000  1.000  0.000000000 -2.0888430 0.1091089451
## 6 1.000  1.000  0.000          NaN        NaN 0.0000000000
## 7 1.000  1.000  0.000          NaN        NaN 0.0000000000
```

Saving the summary output into a XLS