Assignment1.2

Anjani Bonda

9/3/2022

This assignment is a refresher of data analysis and visualization using Python and/or R. Find a data set that interests you and has appropriate data to create some interesting visualizations. A few good sources for finding datasets include Kaggle, UCI ML Repository, and the US Bureau of Labor Statistics. Following steps are performed using R.

Step1: Write a summary of your data and identify at least two questions to explore visually with your data.

The dataset is a series report chosen from BLS (Bureau of Labor Statistics) based on American Time Use Survey. It essentially has data from 2011-2021 with an Estimate of Average hours per day (working and work-related activities including travel). The age group considered here is 15 years and over (both men and women) and for all days.

####Questions I would like to explore: 1. Check if Men and Women have similar proportions/weightage. 2. Check the distribution of 'Estimate' values (avg hrs per day) if even or abnorm

Import the Data saved from BLS data store

merging men and women data to create one file for timeuse analysis

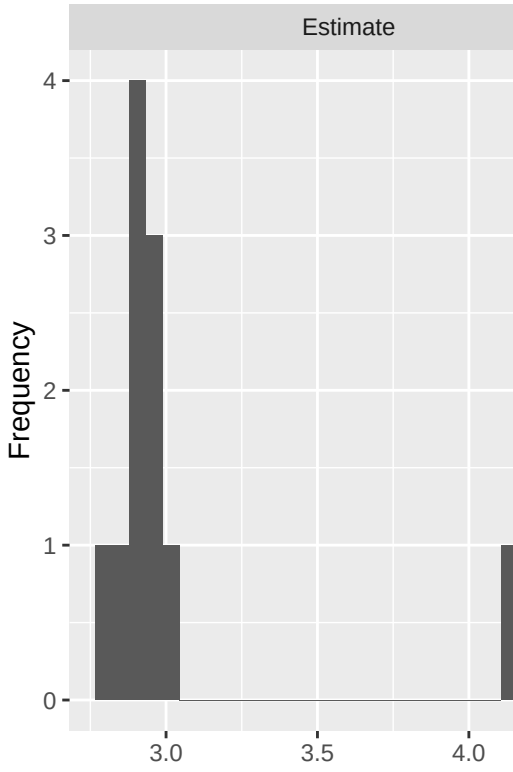
Remove missing values and validate data structure

```
##    Year Period Estimate Standard Error Gender                 Type
## 1  2011 Annual     4.23          0.072    Men Average hours per day
## 2  2012 Annual     4.17          0.073    Men Average hours per day
## 3  2013 Annual     4.20          0.074    Men Average hours per day
## 4  2014 Annual     4.29          0.084    Men Average hours per day
## 5  2015 Annual     4.18          0.075    Men Average hours per day
## 6  2016 Annual     4.39          0.079    Men Average hours per day
## 7  2017 Annual     4.32          0.088    Men Average hours per day
## 8  2018 Annual     4.16          0.082    Men Average hours per day
## 9  2019 Annual     4.36          0.086    Men Average hours per day
## 10 2020 Annual       NA             NA    Men Average hours per day
## 11 2021 Annual     4.18          0.082    Men Average hours per day
## 12 2011 Annual     2.95          0.062  Women Average hours per day
## 13 2012 Annual     2.94          0.071  Women Average hours per day
## 14 2013 Annual     2.77          0.065  Women Average hours per day
## 15 2014 Annual     2.94          0.065  Women Average hours per day
```

```
## 16 2015 Annual     2.92          0.073  Women Average hours per day
## 17 2016 Annual     2.88          0.058  Women Average hours per day
## 18 2017 Annual     2.89          0.075  Women Average hours per day
## 19 2018 Annual     3.02          0.075  Women Average hours per day
## 20 2019 Annual     2.91          0.083  Women Average hours per day
## 21 2020 Annual       NA             NA  Women Average hours per day
## 22 2021 Annual     2.86          0.066  Women Average hours per day
##                                                  Activity
## 1  Working and work-related activities (includes travel)
## 2  Working and work-related activities (includes travel)
## 3  Working and work-related activities (includes travel)
## 4  Working and work-related activities (includes travel)
## 5  Working and work-related activities (includes travel)
## 6  Working and work-related activities (includes travel)
## 7  Working and work-related activities (includes travel)
## 8  Working and work-related activities (includes travel)
## 9  Working and work-related activities (includes travel)
## 10 Working and work-related activities (includes travel)
## 11 Working and work-related activities (includes travel)
## 12 Working and work-related activities (includes travel)
## 13 Working and work-related activities (includes travel)
## 14 Working and work-related activities (includes travel)
## 15 Working and work-related activities (includes travel)
## 16 Working and work-related activities (includes travel)
## 17 Working and work-related activities (includes travel)
## 18 Working and work-related activities (includes travel)
## 19 Working and work-related activities (includes travel)
## 20 Working and work-related activities (includes travel)
## 21 Working and work-related activities (includes travel)
## 22 Working and work-related activities (includes travel)

##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE

##    Year Period Estimate Standard Error Gender                  Type
## 1  2011 Annual     4.23          0.072    Men Average hours per day
## 2  2012 Annual     4.17          0.073    Men Average hours per day
## 3  2013 Annual     4.20          0.074    Men Average hours per day
## 4  2014 Annual     4.29          0.084    Men Average hours per day
## 5  2015 Annual     4.18          0.075    Men Average hours per day
## 6  2016 Annual     4.39          0.079    Men Average hours per day
## 7  2017 Annual     4.32          0.088    Men Average hours per day
## 8  2018 Annual     4.16          0.082    Men Average hours per day
## 9  2019 Annual     4.36          0.086    Men Average hours per day
## 11 2021 Annual     4.18          0.082    Men Average hours per day
## 12 2011 Annual     2.95          0.062  Women Average hours per day
## 13 2012 Annual     2.94          0.071  Women Average hours per day
## 14 2013 Annual     2.77          0.065  Women Average hours per day
## 15 2014 Annual     2.94          0.065  Women Average hours per day
## 16 2015 Annual     2.92          0.073  Women Average hours per day
## 17 2016 Annual     2.88          0.058  Women Average hours per day
## 18 2017 Annual     2.89          0.075  Women Average hours per day
## 19 2018 Annual     3.02          0.075  Women Average hours per day
## 20 2019 Annual     2.91          0.083  Women Average hours per day
```

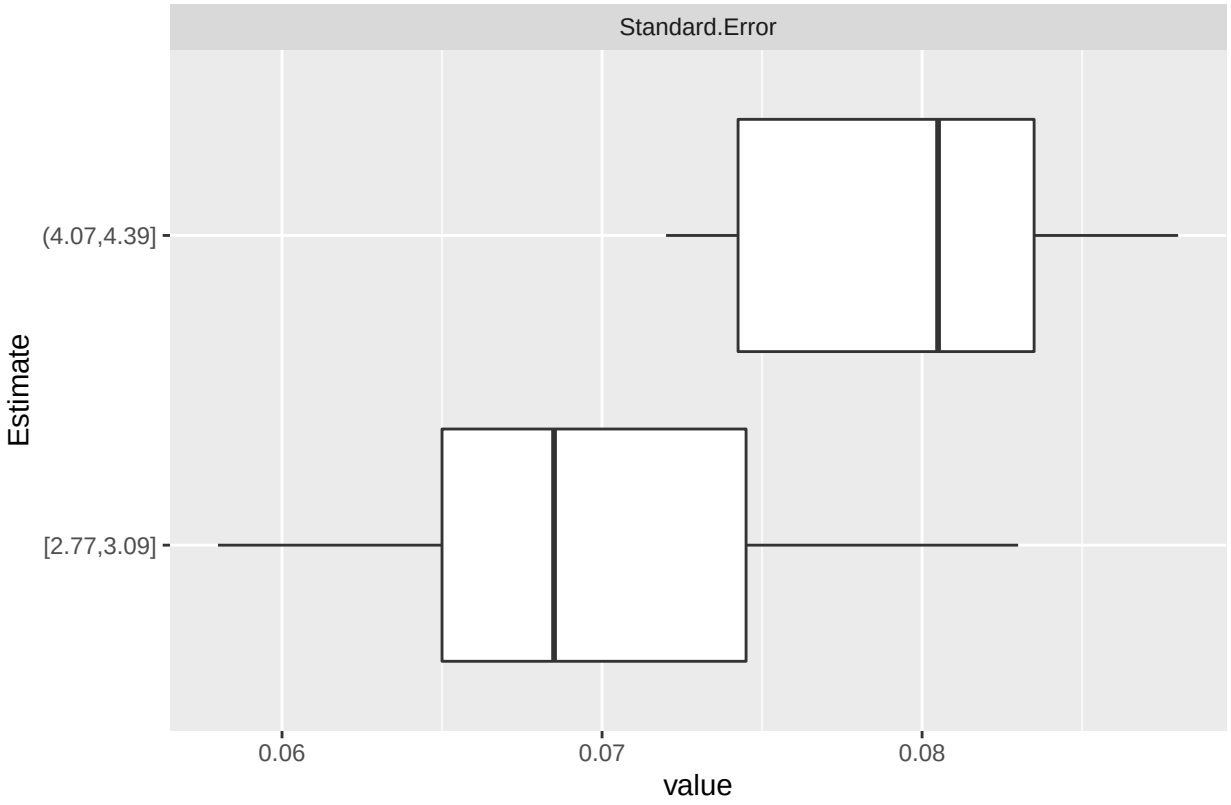
```
## 22 2021 Annual     2.86          0.066  Women Average hours per day
##                                                                Activity
## 1  Working and work-related activities (includes travel)
## 2  Working and work-related activities (includes travel)
## 3  Working and work-related activities (includes travel)
## 4  Working and work-related activities (includes travel)
## 5  Working and work-related activities (includes travel)
## 6  Working and work-related activities (includes travel)
## 7  Working and work-related activities (includes travel)
## 8  Working and work-related activities (includes travel)
## 9  Working and work-related activities (includes travel)
## 11 Working and work-related activities (includes travel)
## 12 Working and work-related activities (includes travel)
## 13 Working and work-related activities (includes travel)
## 14 Working and work-related activities (includes travel)
## 15 Working and work-related activities (includes travel)
## 16 Working and work-related activities (includes travel)
## 17 Working and work-related activities (includes travel)
## 18 Working and work-related activities (includes travel)
## 19 Working and work-related activities (includes travel)
## 20 Working and work-related activities (includes travel)
## 22 Working and work-related activities (includes travel)

## 'data.frame':    20 obs. of  7 variables:
##  $ Year          : chr  "2011" "2012" "2013" "2014" ...
##  $ Period        : chr  "Annual" "Annual" "Annual" "Annual" ...
##  $ Estimate      : num  4.23 4.17 4.2 4.29 4.18 4.39 4.32 4.16 4.36 4.18 ...
##  $ Standard Error: num  0.072 0.073 0.074 0.084 0.075 0.079 0.088 0.082 0.086 0.082 ...
##  $ Gender        : chr  "Men" "Men" "Men" "Men" ...
##  $ Type          : chr  "Average hours per day" "Average hours per day" "Average hours per day" "Ave
##  $ Activity      : chr  "Working and work-related activities (includes travel)" "Working and work-rel
```

Step2: Create a histogram or bar graph from your data

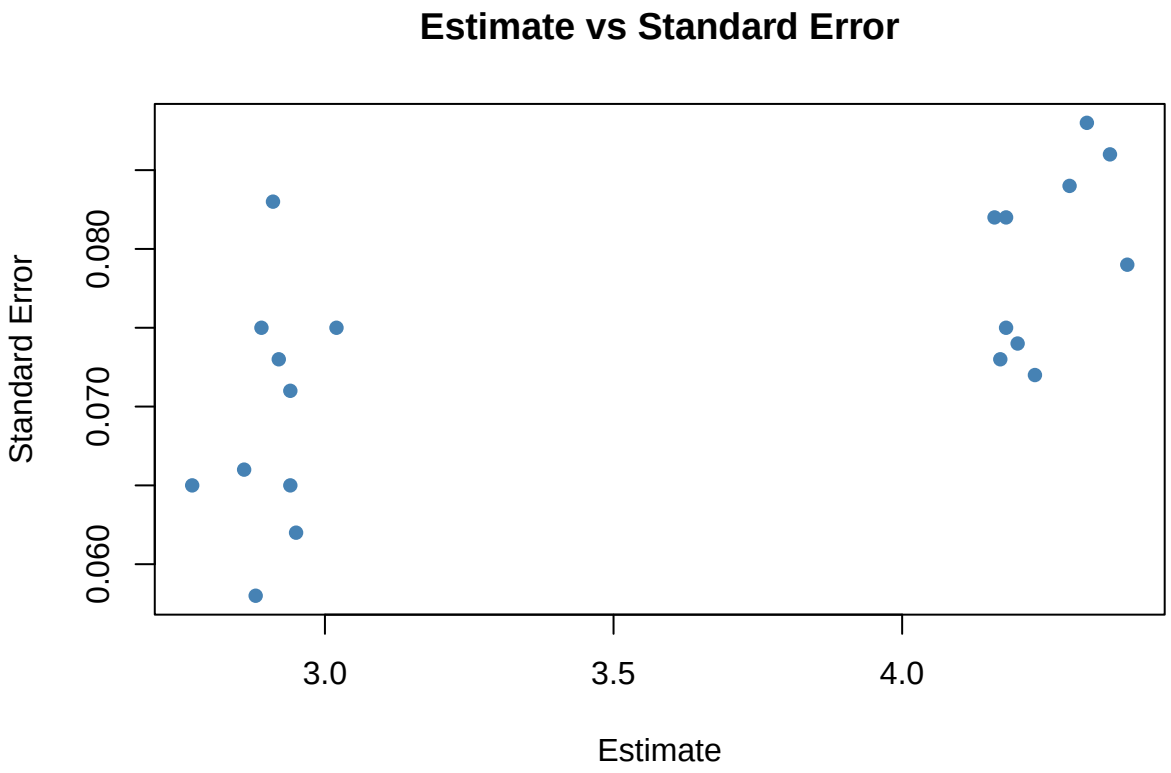


Histogram of the continuous variables (Estimate and Standard Error)

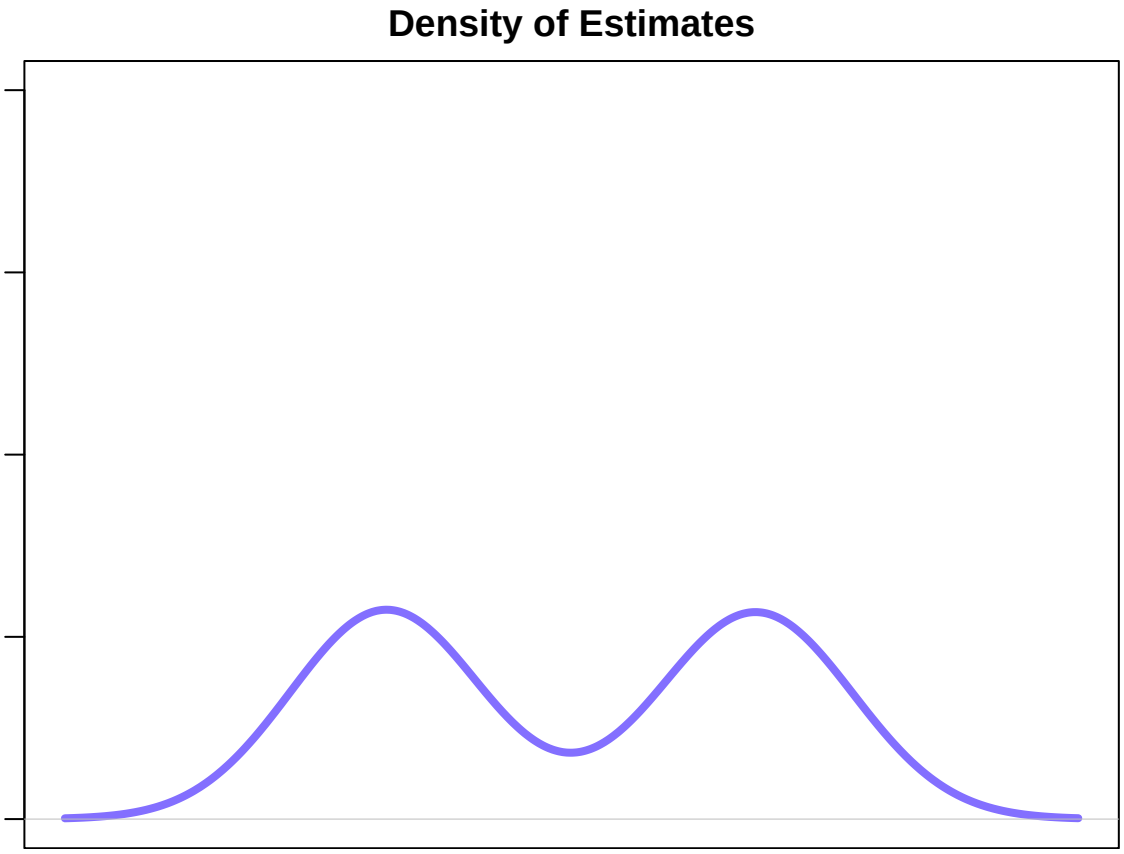
Step3: Create a boxplot from your data.



Step4: Create a bivariate plot from your data.



Step5: Create any additional visualizations that will help to answer the question(s) you want to answer.




```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |              Expected N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  20
##
##
##                  | timeuse$`Standard Error`
## timeuse$Estimate |     0.058 |     0.062 |     0.065 |     0.066 |     0.071 |     0.072 |     0.073
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             2.77 |         0 |         0 |         1 |         0 |         0 |         0 |         0
##                  |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                  |     0.050 |     0.050 |     8.100 |     0.050 |     0.050 |     0.050 |     0.100
##                  |     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 |     0.050 |     0.000 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            2.86 |         0 |         0 |         0 |         1 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |    18.050 |     0.050 |     0.050 |     0.100
##                 |     0.000 |     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 |     0.050 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            2.88 |         1 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |    18.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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
##                 |     0.050 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            2.89 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            2.91 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            2.92 |         0 |         0 |         0 |         0 |         0 |         0 |         1
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     8.100
##                 |     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 |     0.500
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.050
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            2.94 |         0 |         0 |         1 |         0 |         1 |         0 |         0
##                 |     0.100 |     0.100 |     0.200 |     0.100 |     0.100 |     0.100 |     0.200
##                 |     0.100 |     0.100 |     3.200 |     0.100 |     8.100 |     0.100 |     0.200
##                 |     0.000 |     0.000 |     0.500 |     0.000 |     0.500 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.500 |     0.000 |     1.000 |     0.000 |     0.000
##                 |     0.000 |     0.000 |     0.050 |     0.000 |     0.050 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            2.95 |         0 |         1 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |    18.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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 |     0.050 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            3.02 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
```

```
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            4.16 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            4.17 |         0 |         0 |         0 |         0 |         0 |         0 |         1
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     8.100
##                 |     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 |     0.500
##                 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.050
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            4.18 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.100 |     0.100 |     0.200 |     0.100 |     0.100 |     0.100 |     0.200
##                 |     0.100 |     0.100 |     0.200 |     0.100 |     0.100 |     0.100 |     0.200
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             4.2 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            4.23 |         0 |         0 |         0 |         0 |         0 |         1 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |    18.050 |     0.100
##                 |     0.000 |     0.000 |     0.000 |     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 |     0.050 |     0.000
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            4.29 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            4.32 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                 |     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
## ----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##            4.36 |         0 |         0 |         0 |         0 |         0 |         0 |         0
```

```
##                  |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                  |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                  |     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
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##             4.39 |         0 |         0 |         0 |         0 |         0 |         0 |         0
##                  |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                  |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
##                  |     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
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##     Column Total |         1 |         1 |         2 |         1 |         1 |         1 |         2
##                  |     0.050 |     0.050 |     0.100 |     0.050 |     0.050 |     0.050 |     0.100
## -----------------|-----------|-----------|-----------|-----------|-----------|-----------|----------
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  256.6667     d.f. =  238     p =  0.1936414
##
##
##


## timeuse
##
##  7  Variables      20  Observations
## ---------------------------------------------------------------------------
## Year
##        n  missing distinct
##       20        0       10
##
## lowest : 2011 2012 2013 2014 2015, highest: 2016 2017 2018 2019 2021
##
## Value      2011 2012 2013 2014 2015 2016 2017 2018 2019 2021
## Frequency     2    2    2    2    2    2    2    2    2    2
## Proportion  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
## ---------------------------------------------------------------------------
## Period
##        n  missing distinct    value
##       20        0        1   Annual
##
## Value      Annual
## Frequency      20
## Proportion      1
## ---------------------------------------------------------------------------
## Estimate
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##       20        0       18    0.998    3.578   0.7463    2.856    2.878
##      .25      .50      .75      .90      .95
```

```
##    2.917    3.590    4.208    4.324    4.361
##
## lowest : 2.77 2.86 2.88 2.89 2.91, highest: 4.23 4.29 4.32 4.36 4.39
##
## Value      2.77 2.86 2.88 2.89 2.91 2.92 2.94 2.95 3.02 4.16 4.17 4.18 4.20
## Frequency     1    1    1    1    1    1    2    1    1    1    1    2    1
## Proportion 0.05 0.05 0.05 0.05 0.05 0.05 0.10 0.05 0.05 0.05 0.05 0.10 0.05
##
## Value      4.23 4.29 4.32 4.36 4.39
## Frequency     1    1    1    1    1
## Proportion 0.05 0.05 0.05 0.05 0.05
## ---------------------------------------------------------------------------
## Standard Error
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##       20        0       15    0.995   0.0744 0.009674  0.06180  0.06470
##      .25      .50      .75      .90      .95
##  0.06975  0.07450  0.08200  0.08420  0.08610
##
## lowest : 0.058 0.062 0.065 0.066 0.071, highest: 0.082 0.083 0.084 0.086 0.088
##
## Value      0.058 0.062 0.065 0.066 0.071 0.072 0.073 0.074 0.075 0.079 0.082
## Frequency      1     1     2     1     1     1     2     1     3     1     2
## Proportion  0.05  0.05  0.10  0.05  0.05  0.05  0.10  0.05  0.15  0.05  0.10
##
## Value      0.083 0.084 0.086 0.088
## Frequency      1     1     1     1
## Proportion  0.05  0.05  0.05  0.05
## ---------------------------------------------------------------------------
## Gender
##        n  missing distinct
##       20        0        2
##
## Value        Men Women
## Frequency     10    10
## Proportion   0.5   0.5
## ---------------------------------------------------------------------------
## Type
##                     n               missing              distinct
##                    20                     0                     1
##                 value
## Average hours per day
##
## Value      Average hours per day
## Frequency                     20
## Proportion                     1
## ---------------------------------------------------------------------------
## Activity
##                                                  n
##                                                 20
##                                           missing
##                                                 0
##                                          distinct
##                                                 1
##                                             value
```

```
## Working and work-related activities (includes travel)
##
## Value      Working and work-related activities (includes travel)
## Frequency                                                   20
## Proportion                                                   1
## --------------------------------------------------------------------------------
```

Step6: Summarize your results and make a conclusion. Explain how you arrived at this conclusion and how your visualizations support your conclusion.

The high Chi value of 256.67 indicates that the data (Estimate and Standard Error) does not fit very well.

We have data between 2011-2021 and instances per each year for Men/Women.

The estimate is almost an even distribution which is not surprising given the data has men and women in equal proportions.

Standard Error is between 0.058 and 0.088.

The data set has equal weight on gender.