1.3 Assignment: R Refresher

edris safari

9/1/2020

Assignment Description

The first assignment is a review of R.

The purpose of this assignment is to provide a refresher of R and python. The assignment is divided into three sections.

1. Import, Plot, Summarize, and Save Data Using the Bureau of Labor Statistics data, choose a dataset that interest 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.

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.)

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.

Import dataset

```
## [1] "C:/Users/safar/Documents/GitHub/Safarie1103/Bellevue University/Courses/DSC630/Week1"
```

```
## [1] "C:/Users/safar/Documents/GitHub/Safarie1103/Bellevue University/Courses/DSC630/Week1"
```

```
## # A tibble: 21 x 15
##    Year    Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
##    <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 2000   179.  180.  180.  181.  181.  181   181.  182.  182.  183.  183   183.
##  2 2001   184.  184.  185.  186.  186.  186.  186.  187.  187.  188.  188.  188.
##  3 2002   188.  189.  190.  190.  190.  190.  190.  191   191.  192.  192.  191.
##  4 2003   192.  192.  193   193.  193.  193   193.  194.  194.  194.  194.  194.
##  5 2004   194   195.  196.  196.  196.  197.  197.  197.  197.  198.  198.  198.
##  6 2005   198.  200.  201.  201.  201.  201.  201.  201   201.  202.  202.  202.
##  7 2006   203.  204.  205.  206.  206.  206.  206.  207.  207.  208.  208.  207.
##  8 2007   208.  209.  210.  210.  210.  210.  211.  211.  212.  212.  212.  212.
##  9 2008   213.  214.  215.  215.  215.  216.  216.  216.  217.  217.  217.  216.
## 10 2009   217.  218.  219.  219.  219.  219.  219.  220.  220.  221.  220.  220.
## # ... with 11 more rows, and 2 more variables: HALF1 <dbl>, HALF2 <dbl>
```

```
## # A tibble: 21 x 15
##    Year    Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
##    <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 2000   166.  166.  166.  167.  167.  167.  168.  169.  169.  169.  169.  170
##  2 2001   171.  171.  172.  172.  172.  173   174.  174.  174.  175.  175.  175.
##  3 2002   176.  176.  176.  176.  176.  176.  176   176   176.  176.  177.  177.
##  4 2003   178.  178.  179.  178.  179.  180.  180.  180.  181.  182.  182.  184.
##  5 2004   184.  184.  184.  184.  186.  186.  187.  187.  187.  188.  188.  188.
##  6 2005   189.  189.  189.  190.  191.  190.  191.  191.  191.  192.  192.  193.
##  7 2006   194   194   194   194.  194.  194.  195   196.  196.  197.  197.  197
##  8 2007   199.  200   200.  201.  202.  202.  203.  204.  205.  206.  206.  207.
##  9 2008   209.  209.  209.  211.  212.  213.  215.  216.  218.  219.  219.  219.
## 10 2009   220.  219.  219.  218.  218.  218.  217.  217.  217.  218.  217.  218.
## # ... with 11 more rows, and 2 more variables: HALF1 <dbl>, HALF2 <dbl>


## # A tibble: 21 x 15
##    Year    Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
##    <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 2000   112.  117.  122.  121.  121   130.  130.  126.  131.  129.  129   128.
##  2 2001   132.  132   130.  133.  140.  140.  132.  129.  132.  122.  116   111.
##  3 2002   112.  111   116.  122.  123.  125.  126.  126.  126.  126.  125.  123.
##  4 2003   128.  135.  143.  138.  134   136.  137.  141.  145.  137.  133.  132.
##  5 2004   137.  141.  143.  146.  154.  160.  156.  155.  154.  158.  159.  154.
##  6 2005   152.  155.  161.  171.  169.  171.  178.  187.  208   204.  188.  180
##  7 2006   190.  186.  189.  201.  209.  211.  215.  215.  199.  181.  180.  185.
##  8 2007   184.  184.  197.  207.  219.  221.  217.  209.  210.  208.  219.  218.
##  9 2008   219.  219.  231.  240.  257.  276.  281.  266.  258.  232.  190.  171.
## 10 2009   175.  179.  177.  180.  187.  205.  202.  205.  202.  199.  204.  202.
## # ... with 11 more rows, and 2 more variables: HALF1 <dbl>, HALF2 <dbl>
```

```r
summary(all_items_cpi)
```

```
##       Year          Jan             Feb             Mar             Apr
##  2000   : 1   Min.   :178.8   Min.   :179.5   Min.   :180.5   Min.   :180.9
##  2001   : 1   1st Qu.:198.4   1st Qu.:199.5   1st Qu.:200.7   1st Qu.:200.9
##  2002   : 1   Median :220.1   Median :220.6   Median :221.1   Median :221.2
##  2003   : 1   Mean   :220.3   Mean   :221.2   Mean   :222.0   Mean   :222.3
##  2004   : 1   3rd Qu.:239.2   3rd Qu.:240.1   3rd Qu.:241.1   3rd Qu.:241.8
##  2005   : 1   Max.   :266.0   Max.   :267.3   Max.   :267.3   Max.   :266.1
##  (Other):15
##       May             Jun             Jul             Aug
##  Min.   :180.9   Min.   :181.0   Min.   :181.3   Min.   :181.7
##  1st Qu.:199.7   1st Qu.:199.6   1st Qu.:199.8   1st Qu.:199.9
##  Median :220.2   Median :220.3   Median :220.3   Median :220.6
##  Mean   :220.3   Mean   :220.5   Mean   :220.6   Mean   :221.0
##  3rd Qu.:239.1   3rd Qu.:239.2   3rd Qu.:239.2   3rd Qu.:239.4
##  Max.   :262.6   Max.   :263.2   Max.   :263.6   Max.   :264.2
##  NA's   :1       NA's   :1       NA's   :1       NA's   :1
##       Sep             Oct             Nov             Dec
##  Min.   :182.3   Min.   :182.8   Min.   :183.0   Min.   :182.8
##  1st Qu.:200.3   1st Qu.:201.3   1st Qu.:201.2   1st Qu.:201.0
##  Median :221.0   Median :221.4   Median :221.2   Median :220.9
##  Mean   :221.5   Mean   :222.0   Mean   :222.0   Mean   :221.8
```
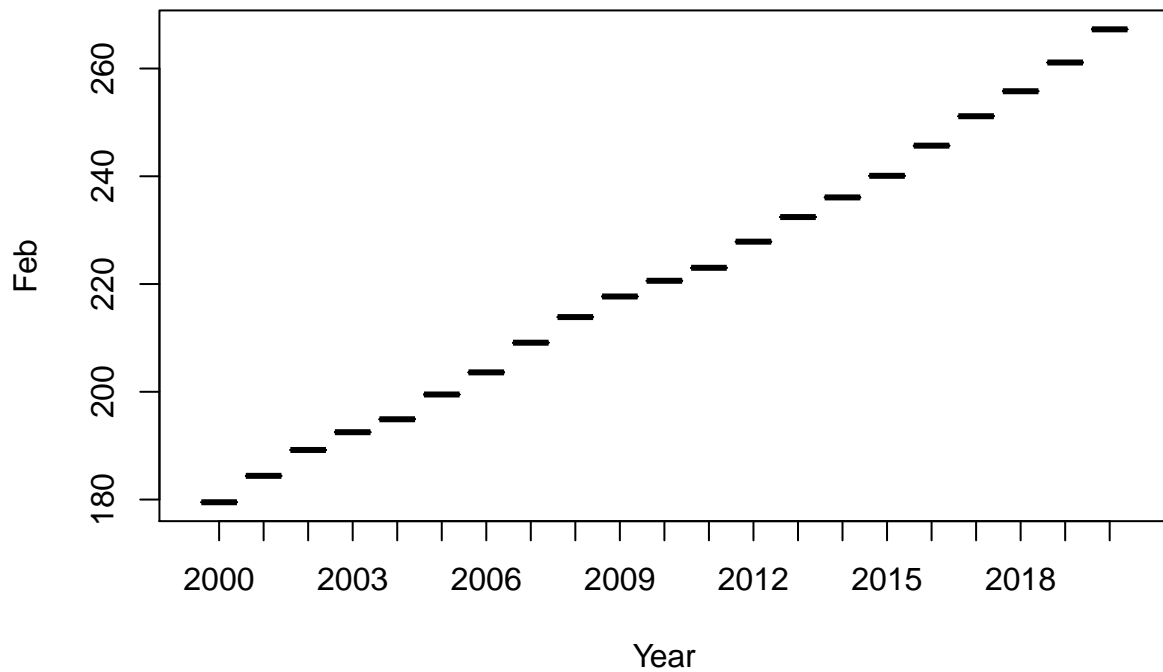
```
##  3rd Qu.:240.0   3rd Qu.:240.6   3rd Qu.:240.5   3rd Qu.:240.0
##  Max.   :264.5   Max.   :265.1   Max.   :265.1   Max.   :264.9
##  NA's   :1       NA's   :1       NA's   :1       NA's   :1
##      HALF1           HALF2
##  Min.   :180.3   Min.   :182.3
##  1st Qu.:199.1   1st Qu.:200.6
##  Median :219.7   Median :220.9
##  Mean   :219.6   Mean   :221.5
##  3rd Qu.:238.0   3rd Qu.:239.9
##  Max.   :261.9   Max.   :264.6
##  NA's   :1       NA's   :1
```

```r
str(all_items_cpi)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    21 obs. of  15 variables:
##  $ Year : Factor w/ 21 levels "2000","2001",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Jan  : num  179 184 188 192 194 ...
##  $ Feb  : num  180 184 189 192 195 ...
##  $ Mar  : num  180 185 190 193 196 ...
##  $ Apr  : num  181 186 190 193 196 ...
##  $ May  : num  181 186 190 193 196 ...
##  $ Jun  : num  181 186 190 193 197 ...
##  $ Jul  : num  181 186 190 193 197 ...
##  $ Aug  : num  182 187 191 194 197 ...
##  $ Sep  : num  182 187 191 194 197 ...
##  $ Oct  : num  183 188 192 194 198 ...
##  $ Nov  : num  183 188 192 194 198 ...
##  $ Dec  : num  183 188 191 194 198 ...
##  $ HALF1: num  180 185 190 193 196 ...
##  $ HALF2: num  182 187 191 194 198 ...
```
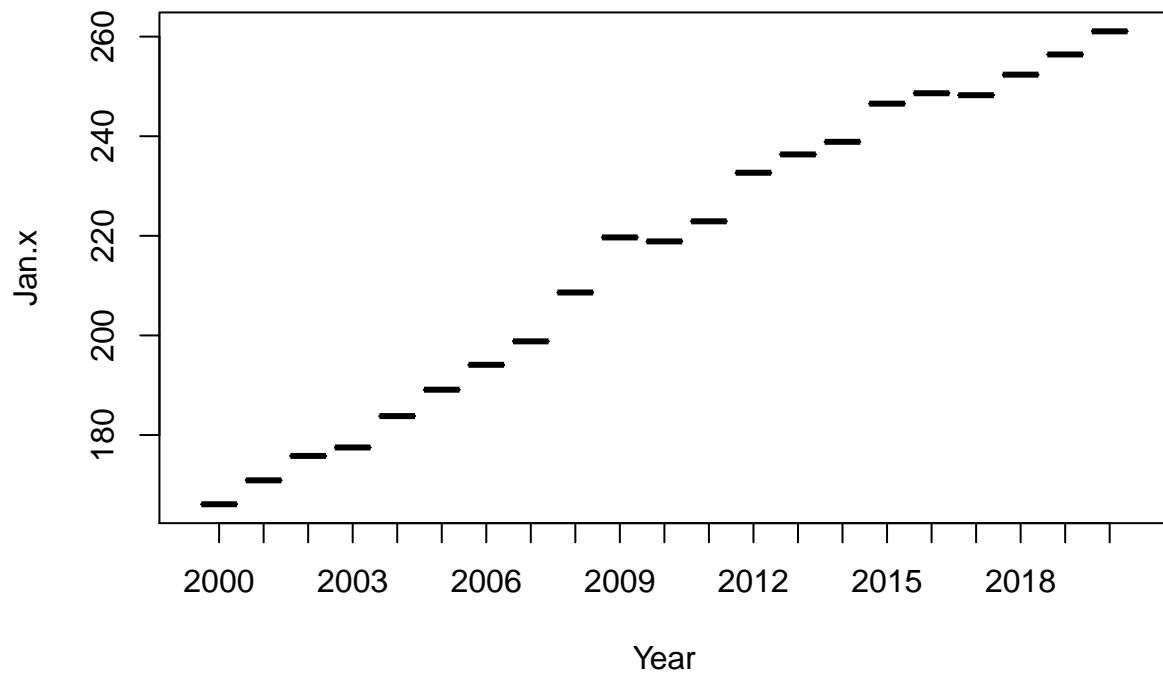
```r
plot(Feb ~ Year, data=all_items_cpi)
```

```
food_and_energy_CPI <- food_cpi %>% left_join(energy_cpi,by="Year")
names(food_and_energy_CPI)
```

```
##  [1] "Year"    "Jan.x"   "Feb.x"   "Mar.x"   "Apr.x"   "May.x"   "Jun.x"
##  [8] "Jul.x"   "Aug.x"   "Sep.x"   "Oct.x"   "Nov.x"   "Dec.x"   "HALF1.x"
## [15] "HALF2.x" "Jan.y"   "Feb.y"   "Mar.y"   "Apr.y"   "May.y"   "Jun.y"
## [22] "Jul.y"   "Aug.y"   "Sep.y"   "Oct.y"   "Nov.y"   "Dec.y"   "HALF1.y"
## [29] "HALF2.y"
```

```
plot(Jan.x~Year,data=food_and_energy_CPI)
```

```r
plot(Jan.y~Year,data=food_and_energy_CPI)
```