ThipRattanavilay_1_2_Assignment_R_Refresher

Thip Rattanavilay

6/12/2021

1. Import, Plot, Summarize, and Save Data

```
##    Year Quarter Number.of.workers..in.thousands....Total
## 1  2009       4                                    98555
## 2  2010       1                                    98143
## 3  2010       2                                    99605
## 4  2010       3                                   100412
## 5  2010       4                                    99958
## 6  2011       1                                    99670
## 7  2011       2                                   100347
## 8  2011       3                                   100495
## 9  2011       4                                   101337
## 10 2012       1                                   102161
##    Number.of.workers..in.thousands....Men
## 1                                   54410
## 2                                   54098
## 3                                   55028
## 4                                   55620
## 5                                   55486
## 6                                   55337
## 7                                   55821
## 8                                   56046
## 9                                   56687
## 10                                  57110
##    Number.of.workers..in.thousands....Women
## 1                                     44145
## 2                                     44045
## 3                                     44577
## 4                                     44792
## 5                                     44472
## 6                                     44333
## 7                                     44526
## 8                                     44449
## 9                                     44650
## 10                                    45051
##    Median.weekly.earnings..in.current.dollars....Total
## 1                                                  747
## 2                                                  748
## 3                                                  742
## 4                                                  746
## 5                                                  750
## 6                                                  750
```

```
## 7                                                    754
## 8                                                    760
## 9                                                    760
## 10                                                   764
##    Median.weekly.earnings..in.current.dollars....Men
## 1                                                  823
## 2                                                  836
## 3                                                  814
## 4                                                  821
## 5                                                  826
## 6                                                  821
## 7                                                  830
## 8                                                  836
## 9                                                  838
## 10                                                 841
##    Median.weekly.earnings..in.current.dollars....Women
## 1                                                  666
## 2                                                  662
## 3                                                  671
## 4                                                  670
## 5                                                  676
## 6                                                  679
## 7                                                  687
## 8                                                  681
## 9                                                  686
## 10                                                 693
##    Median.weekly.earnings..in.constant.dollars....Total
## 1                                                  344
## 2                                                  344
## 3                                                  342
## 4                                                  342
## 5                                                  341
## 6                                                  338
## 7                                                  336
## 8                                                  336
## 9                                                  335
## 10                                                 335
##    Median.weekly.earnings..in.constant.dollars....Men
## 1                                                379
## 2                                                384
## 3                                                374
## 4                                                377
## 5                                                376
## 6                                                370
## 7                                                370
## 8                                                370
## 9                                                369
## 10                                               368
##    Median.weekly.earnings..in.constant.dollars....Women
## 1                                                  307
## 2                                                  304
## 3                                                  309
## 4                                                  308
## 5                                                  308
```

```
## 6                                                    306
## 7                                                    306
## 8                                                    301
## 9                                                    302
## 10                                                   303
```

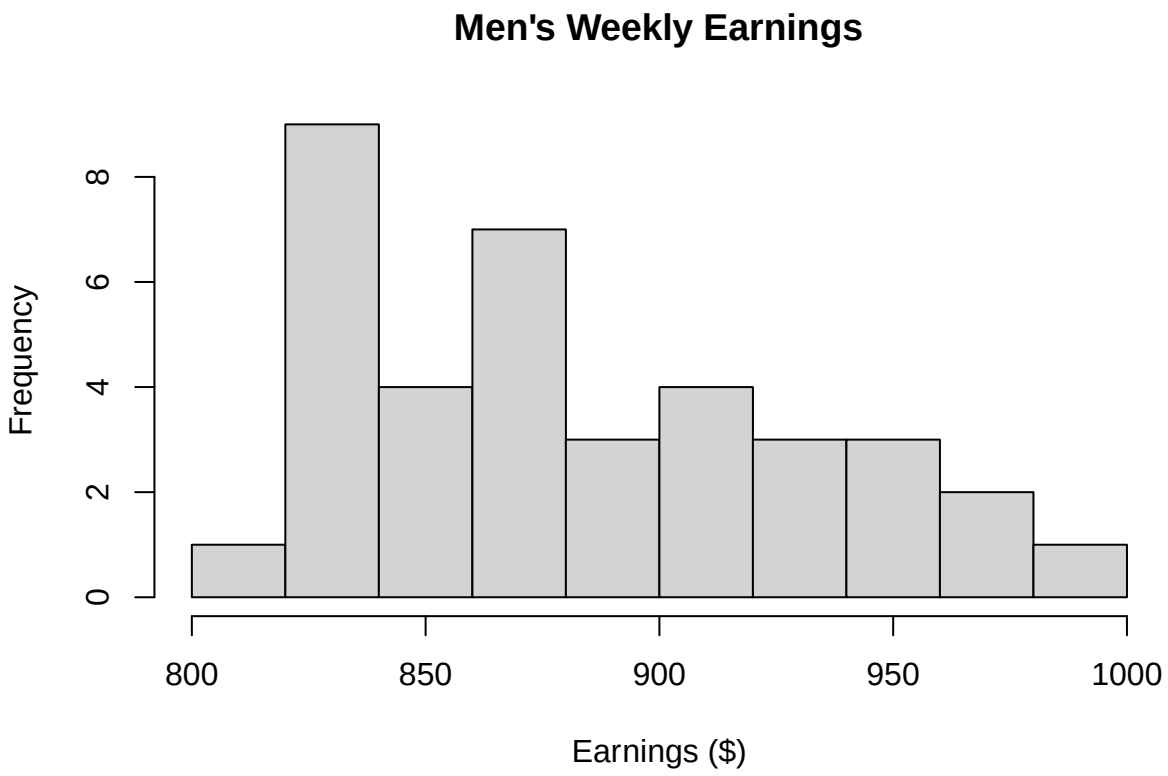
```
# Summary Descriptives
round(stat.desc(earnings$Number.of.workers..in.thousands....Men), 2)
```

```
##      nbr.val     nbr.null       nbr.na          min          max        range
##        37.00         0.00         0.00     54098.00     64237.00     10139.00
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   2196813.00     59486.00     59373.32       514.40      1043.24   9790292.23
##      std.dev     coef.var
##      3128.94         0.05
```

```
round(stat.desc(earnings$Number.of.workers..in.thousands....Women), 2)
```

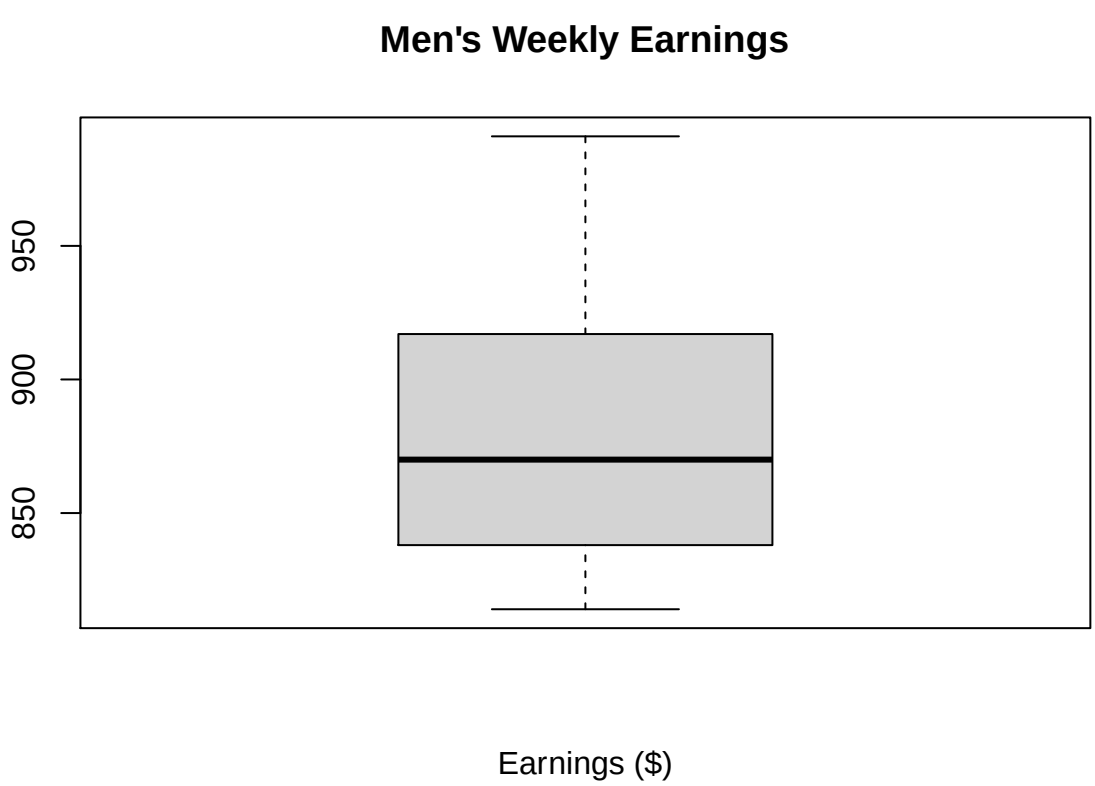
```
##      nbr.val     nbr.null       nbr.na          min          max        range
##        37.00         0.00         0.00     44045.00     51923.00      7878.00
##          sum       median         mean      SE.mean CI.mean.0.95          var
##   1752035.00     46951.00     47352.30       403.58       818.50   6026401.77
##      std.dev     coef.var
##      2454.87         0.05
```

```
# Plot some of the features of several variables
men_weekly_earnings <- earnings$Median.weekly.earnings..in.current.dollars....Men
women_weekly_earnings <- earnings$Median.weekly.earnings..in.current.dollars....Women
hist(men_weekly_earnings, main="Men's Weekly Earnings", xlab = "Earnings ($)")
```




```
hist(women_weekly_earnings, main="Women's Weekly Earnings", xlab = "Earnings ($)")
```


```
boxplot(men_weekly_earnings, main="Men's Weekly Earnings", xlab = "Earnings ($)")
```

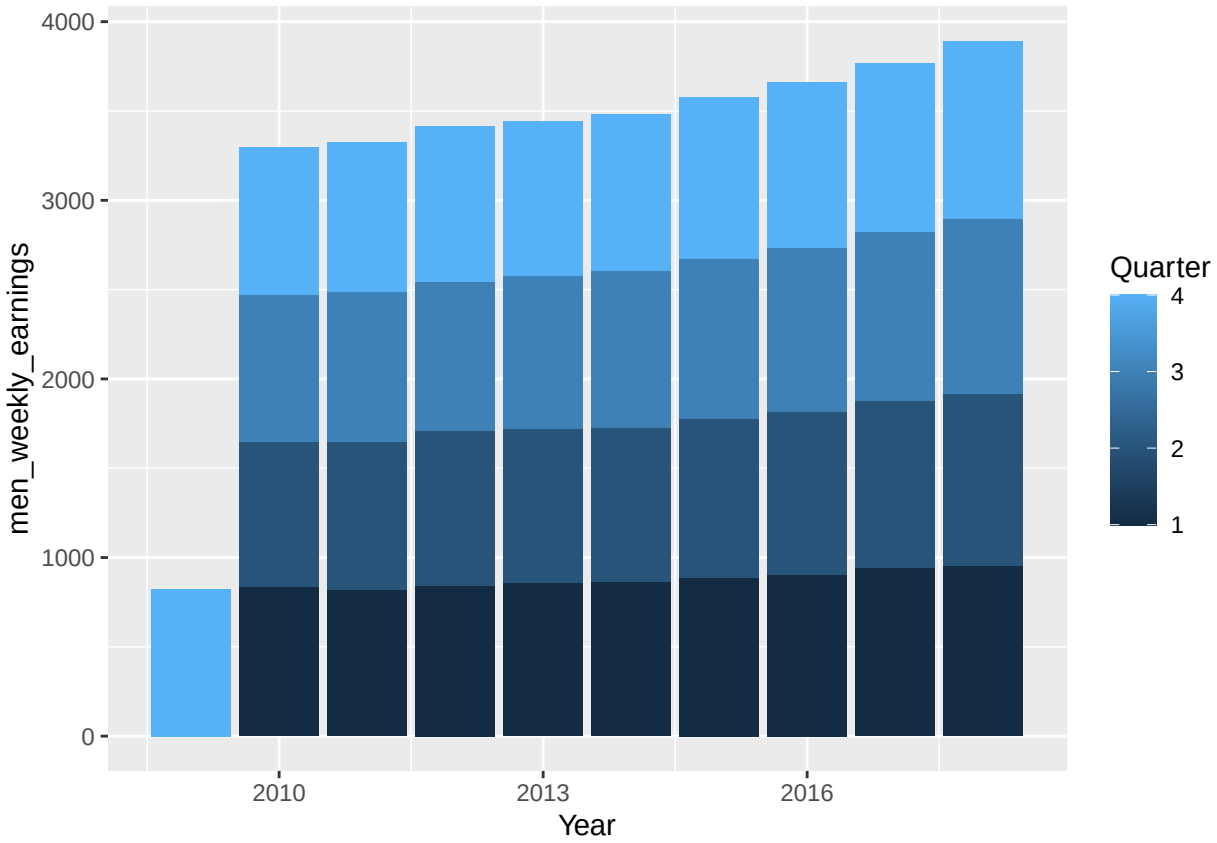



```
boxplot(women_weekly_earnings, main="Women's Weekly Earnings", xlab = "Earnings ($)")
```

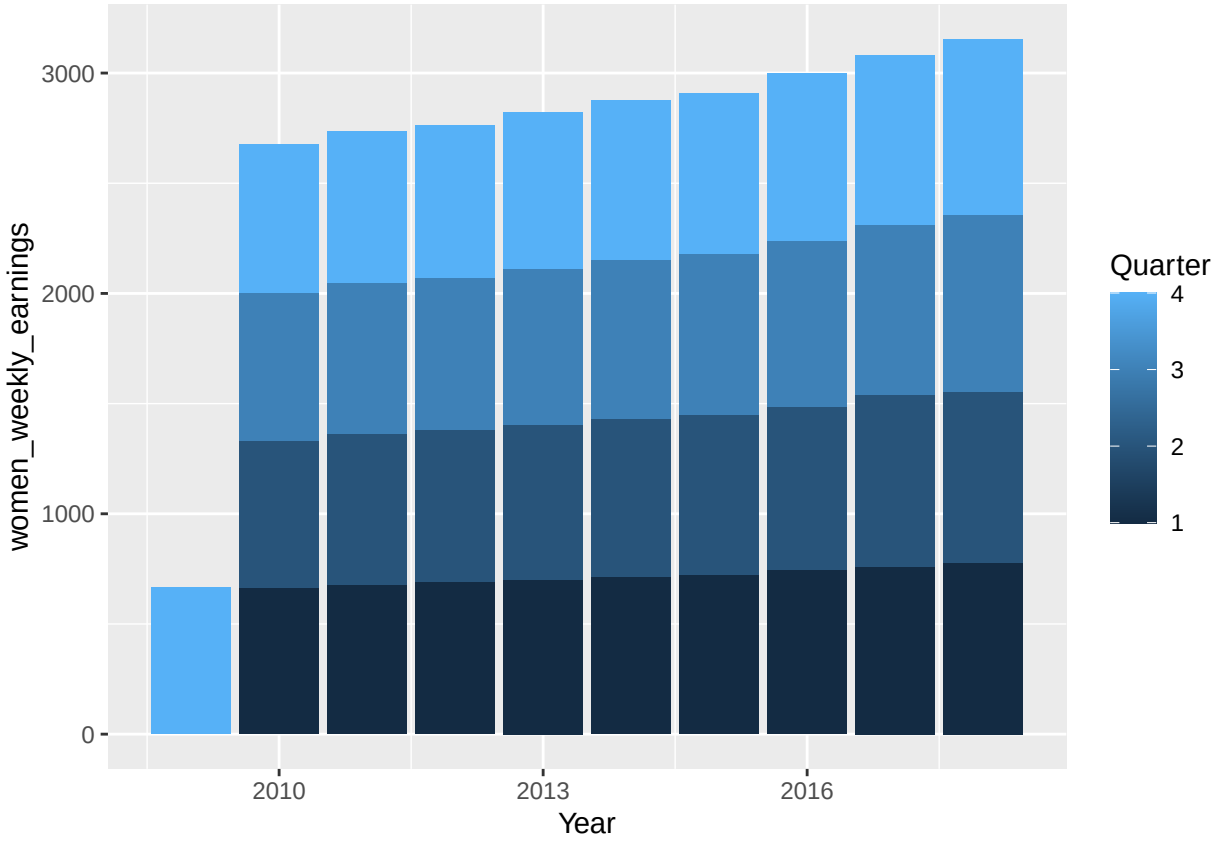

```
# Save data locally
write.csv(earnings, file = 'MenVsWomenMedianEarnings.csv')
```

2. Explore Some Bivariate Relations

```
# Bivariate Relations
ggplot2::ggplot(earnings, ggplot2::aes(x = Year, y = men_weekly_earnings, fill = Quarter)) + ggplot2::g
```




```
ggplot2::ggplot(earnings, ggplot2::aes(x = Year, y = women_weekly_earnings, fill = Quarter)) + ggplot2:
```




```
# Correlation to number of men workers and earnings
men_employees <- earnings$Number.of.workers..in.thousands....Men
cor.test(men_employees, men_weekly_earnings, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  men_employees and men_weekly_earnings
## t = 21.927, df = 35, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9334769 0.9822227
## sample estimates:
##       cor
## 0.9654744
```

```
# Correlation to number of women workers and earnings
women_employees <- earnings$Number.of.workers..in.thousands....Women
cor.test(women_employees, women_weekly_earnings, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  women_employees and women_weekly_earnings
## t = 33.883, df = 35, p-value < 2.2e-16
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9710159 0.9923631
## sample estimates:
##       cor
## 0.9850964
```

3. Organize a Data Report

```
summary(earnings)
```

```
##       Year         Quarter      Number.of.workers..in.thousands....Total
##  Min.   :2009   Min.   :1.000   Min.   : 98143
##  1st Qu.:2012   1st Qu.:2.000   1st Qu.:102161
##  Median :2014   Median :3.000   Median :106342
##  Mean   :2014   Mean   :2.541   Mean   :106726
##  3rd Qu.:2016   3rd Qu.:4.000   3rd Qu.:111463
##  Max.   :2018   Max.   :4.000   Max.   :116160
##  Number.of.workers..in.thousands....Men
##  Min.   :54098
##  1st Qu.:57079
##  Median :59486
##  Mean   :59373
##  3rd Qu.:62082
##  Max.   :64237
##  Number.of.workers..in.thousands....Women
##  Min.   :44045
##  1st Qu.:45051
##  Median :46951
##  Mean   :47352
##  3rd Qu.:49214
##  Max.   :51923
##  Median.weekly.earnings..in.current.dollars....Total
##  Min.   :742.0
##  1st Qu.:764.0
##  Median :790.0
##  Mean   :801.3
##  3rd Qu.:833.0
##  Max.   :897.0
##  Median.weekly.earnings..in.current.dollars....Men
##  Min.   :814.0
##  1st Qu.:838.0
##  Median :870.0
##  Mean   :883.1
##  3rd Qu.:917.0
##  Max.   :991.0
##  Median.weekly.earnings..in.current.dollars....Women
##  Min.   :662
##  1st Qu.:687
##  Median :716
##  Mean   :721
##  3rd Qu.:750
```

```
##  Max.   :801
##  Median.weekly.earnings..in.constant.dollars....Total
##  Min.   :330.0
##  1st Qu.:335.0
##  Median :341.0
##  Mean   :341.4
##  3rd Qu.:346.0
##  Max.   :355.0
##  Median.weekly.earnings..in.constant.dollars....Men
##  Min.   :363.0
##  1st Qu.:370.0
##  Median :377.0
##  Mean   :376.4
##  3rd Qu.:382.0
##  Max.   :392.0
##  Median.weekly.earnings..in.constant.dollars....Women
##  Min.   :298.0
##  1st Qu.:304.0
##  Median :306.0
##  Mean   :307.3
##  3rd Qu.:311.0
##  Max.   :320.0
```

```
str(earnings)
```

```
## 'data.frame':    37 obs. of  11 variables:
##  $ Year                                                : int  2009 2010 2010 2010 2010 2011 2011 201
##  $ Quarter                                             : int  4 1 2 3 4 1 2 3 4 1 ...
##  $ Number.of.workers..in.thousands....Total            : num  98555 98143 99605 100412 99958 ...
##  $ Number.of.workers..in.thousands....Men              : num  54410 54098 55028 55620 55486 ...
##  $ Number.of.workers..in.thousands....Women            : num  44145 44045 44577 44792 44472 ...
##  $ Median.weekly.earnings..in.current.dollars....Total : int  747 748 742 746 750 750 754 760 760 764
##  $ Median.weekly.earnings..in.current.dollars....Men   : int  823 836 814 821 826 821 830 836 838 841
##  $ Median.weekly.earnings..in.current.dollars....Women : int  666 662 671 670 676 679 687 681 686 693
##  $ Median.weekly.earnings..in.constant.dollars....Total: int  344 344 342 342 341 338 336 336 335 335
##  $ Median.weekly.earnings..in.constant.dollars....Men  : int  379 384 374 377 376 370 370 370 369 368
##  $ Median.weekly.earnings..in.constant.dollars....Women: int  307 304 309 308 308 306 306 301 302 303
```

Results

Looking at the earnings of both women and men on a weekly basis, we can easily tell that the men overall are earning more. In fact, the minimum earning for men is larger than that of women.

While exploring the bivariate relations, I witnessed the earnings for men and women increasing each year. However, even though both were increasing, the women were still earning less than men.

In running the correlation of the amount of employees and earnings of men and female workers, there was a strong positive correlation in both areas for men and women. It could be possibly due to with more more workers in an area, the earnings gathered for employees would be hirer since there are more employees being paid increasing the grand total.

With looking at the box plots, I noticed that the higher bound outliers of the women's earnings are about the same of the lower outliers of the men's earnings. With looking at the earnings totals and distributions

between the men's and women's, you can tell that there is a difference in pay between the two groups no matter the industry.