1.3 R Refresher

Gabriel Valenzuela

6/7/2020

## 1. Import, Plot, Summarize, and Save Data

## Warning: package 'ggplot2' was built under R version 3.6.2

## Warning: package 'hexbin' was built under R version 3.6.3

## Year Quarter Number.of.workers..in.thousands....Total  
## 1 2009 4 98,555  
## 2 2010 1 98,143  
## 3 2010 2 99,605  
## 4 2010 3 100,412  
## 5 2010 4 99,958  
## 6 2011 1 99,670  
## 7 2011 2 100,347  
## 8 2011 3 100,495  
## 9 2011 4 101,337  
## 10 2012 1 102,161  
## Number.of.workers..in.thousands....Men  
## 1 54,410  
## 2 54,098  
## 3 55,028  
## 4 55,620  
## 5 55,486  
## 6 55,337  
## 7 55,821  
## 8 56,046  
## 9 56,687  
## 10 57,110  
## Number.of.workers..in.thousands....Women  
## 1 44,145  
## 2 44,045  
## 3 44,577  
## 4 44,792  
## 5 44,472  
## 6 44,333  
## 7 44,526  
## 8 44,449  
## 9 44,650  
## 10 45,051  
## 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

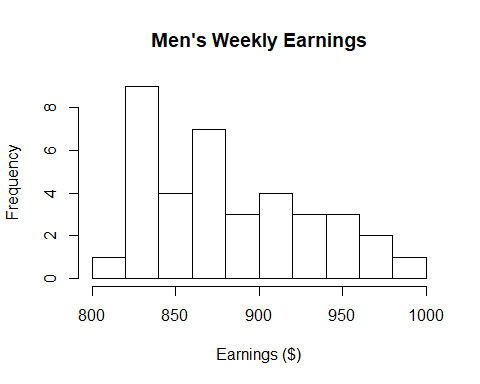
# Summary Descriptives  
round(stat.desc(earnings$Number.of.workers..in.thousands....Men), 2)

## nbr.val nbr.null nbr.na min max range   
## 37.00 0.00 0.00 1.00 37.00 36.00   
## sum median mean SE.mean CI.mean.0.95 var   
## 703.00 19.00 19.00 1.78 3.61 117.17   
## std.dev coef.var   
## 10.82 0.57

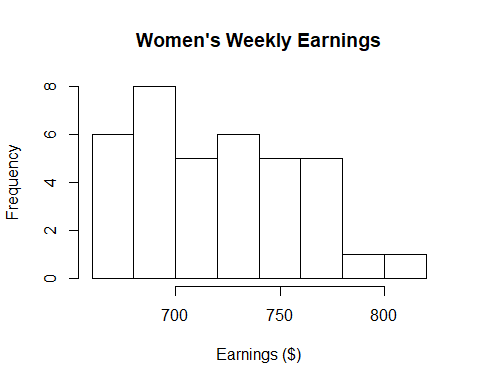
round(stat.desc(earnings$Number.of.workers..in.thousands....Women), 2)

## nbr.val nbr.null nbr.na min max range   
## 37.00 0.00 0.00 1.00 36.00 35.00   
## sum median mean SE.mean CI.mean.0.95 var   
## 680.00 18.00 18.38 1.71 3.47 108.46   
## std.dev coef.var   
## 10.41 0.57

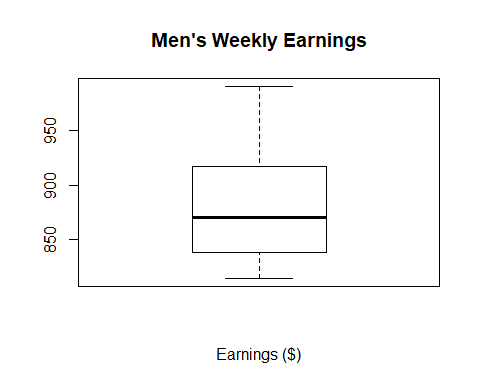
# Plot some of the features of several variables  
men\_weekly\_earnings <- earnings$Median.weekly.earnings..in.current.dollars....Men  
women\_weekly\_earnings <- earnings$Median.weekly.earnings..in.current.dollars....Women  
hist(men\_weekly\_earnings, main="Men's Weekly Earnings", xlab = "Earnings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



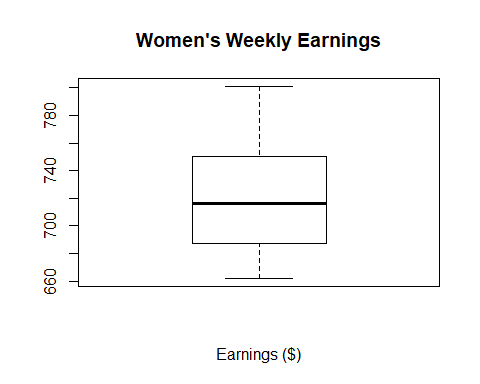
hist(women\_weekly\_earnings, main="Women's Weekly Earnings", xlab = "Earnings ($)")



boxplot(men\_weekly\_earnings, main="Men's Weekly Earnings", xlab = "Earnings ($)")



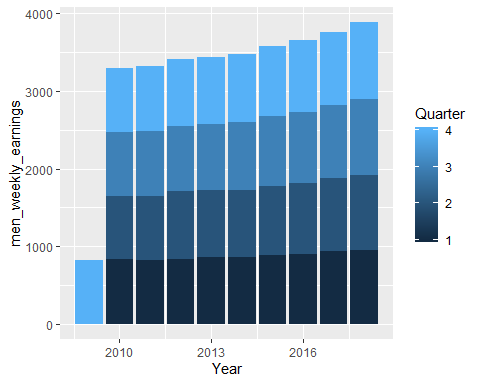
boxplot(women\_weekly\_earnings, main="Women's Weekly Earnings", xlab = "Earnings ($)")



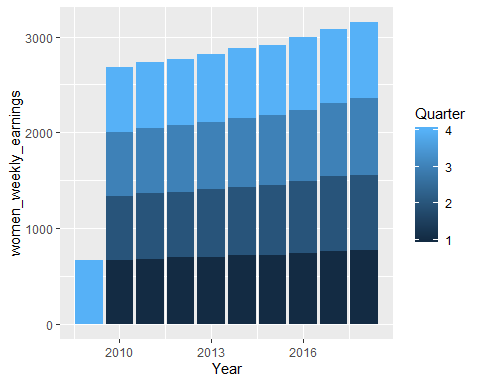
# Save data locally  
write.csv(earnings, file = 'MenVsWomenMedianEarnings.csv')

## 2. Explore Some Bivariate Relations

# Bivariate Relations  
ggplot2::ggplot(earnings, ggplot2::aes(x = Year, y = men\_weekly\_earnings, fill = Quarter)) + ggplot2::geom\_bar(stat = "identity")



ggplot2::ggplot(earnings, ggplot2::aes(x = Year, y = women\_weekly\_earnings, fill = Quarter)) + ggplot2::geom\_bar(stat = "identity")



# 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 = 22.079, df = 35, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9343325 0.9824570  
## sample estimates:  
## cor   
## 0.9659257

# 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 = 27.281, df = 35, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9559872 0.9883377  
## sample estimates:  
## cor   
## 0.9772845

## 3. Organize a Data Report

summary(earnings)

## Year Quarter Number.of.workers..in.thousands....Total  
## Min. :2009 Min. :1.000 Min. : 1   
## 1st Qu.:2012 1st Qu.:2.000 1st Qu.:10   
## Median :2014 Median :3.000 Median :19   
## Mean :2014 Mean :2.541 Mean :19   
## 3rd Qu.:2016 3rd Qu.:4.000 3rd Qu.:28   
## Max. :2018 Max. :4.000 Max. :37   
## Number.of.workers..in.thousands....Men  
## Min. : 1   
## 1st Qu.:10   
## Median :19   
## Mean :19   
## 3rd Qu.:28   
## Max. :37   
## Number.of.workers..in.thousands....Women  
## Min. : 1.00   
## 1st Qu.:10.00   
## Median :18.00   
## Mean :18.38   
## 3rd Qu.:27.00   
## Max. :36.00   
## 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 2011 2011 2012 ...  
## $ Quarter : int 4 1 2 3 4 1 2 3 4 1 ...  
## $ Number.of.workers..in.thousands....Total : num 34 33 35 2 37 36 1 3 4 5 ...  
## $ Number.of.workers..in.thousands....Men : num 2 1 3 6 5 4 7 8 9 11 ...  
## $ Number.of.workers..in.thousands....Women : num 2 1 7 9 5 3 6 4 8 10 ...  
## $ 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.