DaigleInClassLab_Wk4D1.R

2011home

Tue Feb 6 13:08:14 2018

```r
## Chris Daigle
## In class Lab: Wk4D1

# Exercise 1
set.seed(pi)
r <- rnorm(2754, 0, 1)
income <- exp(r)

# find 1/10, 7/20, 15/20, 98/100 quantiles of income, you can use "quantile()"
quant <- quantile(income, c(0.10, 0.35, 0.75, 0.98))

# construct a vector, income.level, as follows:
income.level <- rep(NA, 2754)
income.level[income <= quant[1]] <- "VL"
income.level[income > quant[1] & income <= quant[2]] <- "L"
income.level[income > quant[2] & income <= quant[3]] <- "M"
income.level[income > quant[3] & income <= quant[4]] <- "H"
income.level[income > quant[4]] <- "VH"

# Make an ordered factor from income.level. Specify the levels.
factor_income.level <- factor(income.level, ordered = TRUE, levels <- c("VL", "L", "M
", "H", "VH"))

#rename
levels(factor_income.level) <- c("Very Low", "Low", "Middle", "High", "Very High")

#summary
summary(factor_income.level)
```
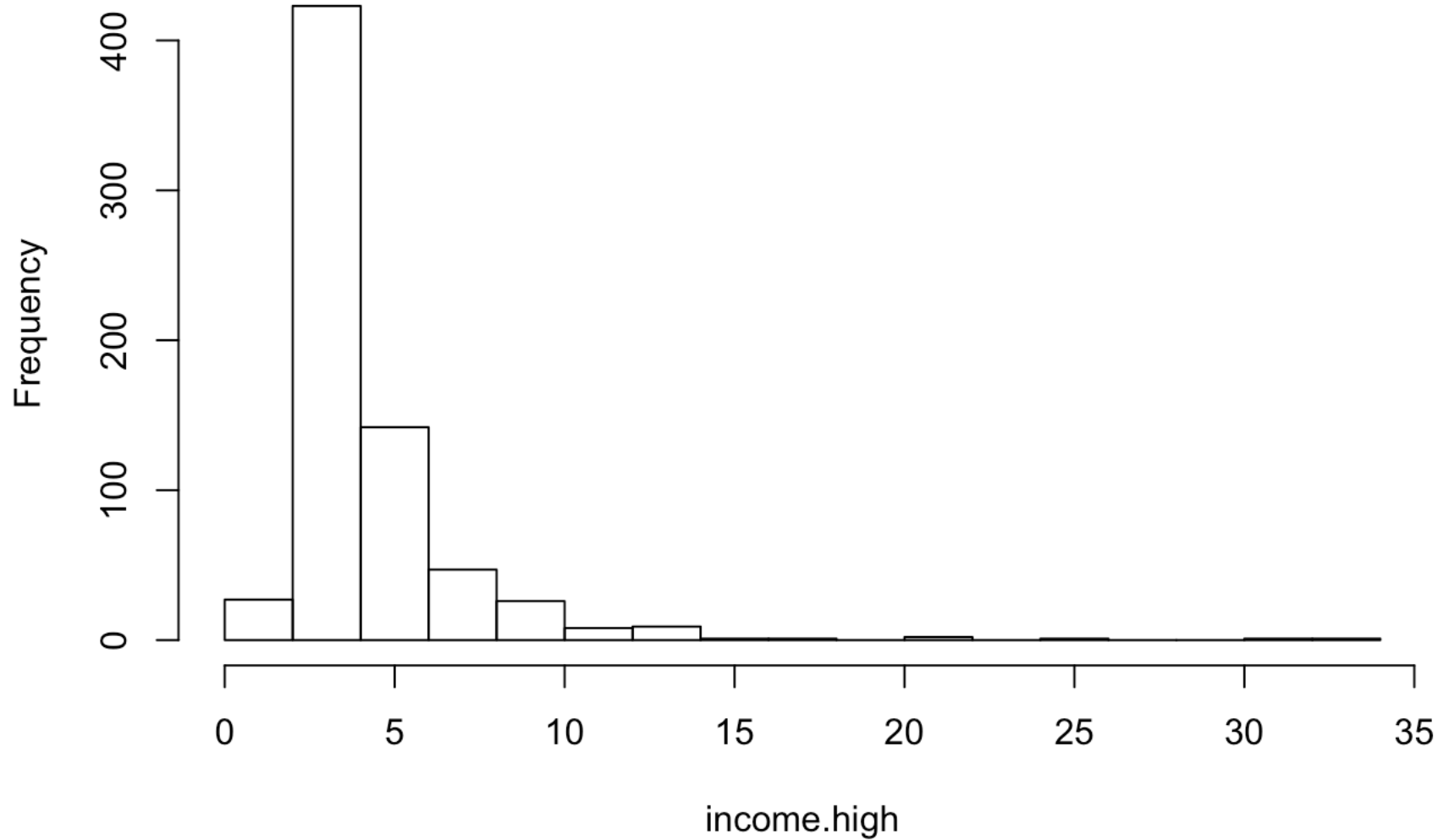
```
##  Very Low       Low    Middle      High Very High
##       276       688      1101       633        56
```

```r
#create a subvector income.high that includes income level is higher than "Middle"
income.high <- income[factor_income.level > "Middle"]
hist(income.high, breaks = 20)
```

Histogram of income.high


```
# calculate average income for people with income higher than middle
mean(income.high)
```

```
## [1] 4.134607
```

```
#what's the gap between average income high and very high
gap = mean(income[factor_income.level == "Very High"]) - mean(income[factor_income.le
vel == "High"])
gap
```

```
## [1] 7.975506
```

```
## Exercise 2
industry <- sample(c("Manufacture", "Service", "IT"), 100, replace = TRUE, prob = c(0
.3, 0.5, 0.2))
stock <- rep(NA,100)
stock[industry == "Manufacture"] <- rnorm(sum(industry == "Manufacture"), 3, 2)
stock[industry == "Service"] <- rnorm(sum(industry == "Service"), 2, 4)
stock[industry == "IT"] <- rnorm(sum(industry == "IT"), 8, 8)

factor.industry <- factor(industry)
summary(factor.industry)
```

```
##          IT Manufacture     Service
##          23          28          49
```

```
stockManufacture <- stock[factor.industry == "Manufacture"]
stockService <- stock[factor.industry == "Service"]
stockIT <- stock[factor.industry == "IT"]

mean <- c(mean(stockManufacture), mean(stockService), mean(stockIT))
sd <- c(sd(stockManufacture), sd(stockService), sd(stockIT))

comp <- cbind(mean, sd)
indNames <-c("Manufacture", "Service", "IT")
rownames(comp) <- indNames
comp
```

```
##                  mean       sd
## Manufacture 2.437437 2.190272
## Service     2.014805 4.034680
## IT          5.477804 7.459002
```