Assignment 8.2

```
library(RcmdrPlugin.IPSUR)

data(RcmdrTestDrive)
```

Perform the below operations:

1. Compute the measures of central tendency for salary and reduction which variable has highest center?

2. Which measure of center is more appropriate for before and after?

The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others

Three measures of central tendency

i. The mean: describes the typical score

ii. The mode: describes the most recurring score 1. Only used with nominal variables

iii. The median: is the 50th Percentile of the distribution 1.

A median is a special case of a percentile, which is the percentage of cases below which a specific percentage of cases fall.

The mean is always the center of any distribution. The mean may often be very misleading because it is sensitive to all observations whereas the median is not.

```
c(mean(before),median(before))

## [1] 73.96607 73.90000

c(mean(after),median(after))

## [1] 73.26726 73.70000

sd(before)

## [1] 1.076412

mad(after)

## [1] 1.55673

IQR(after)/1.349

## [1] 1.556709
```

```
library(e1071)

##
## Attaching package: 'e1071'

## The following object is masked from 'package:Hmisc':
##
##     impute

skewness(before)

## [1] -0.03510369

kurtosis(before)

## [1] -0.7972288

skewness(after)

## [1] -1.164056

kurtosis(after)

## [1] 1.322198
```

The following summary table to know what the best measure of central tendency is with respect to the different types of variable.

Type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

We should take the sample skewness value and compare it to
$2\sqrt{6/n}\approx.$\Sexpr{round(2*sqrt(6/length(before)),3)}

in absolute value to see if it is substantially different from zero.The direction of skewness is decided by the sign (positive or negative)of the skewness value.

We should take the sample kurtosis value and compare it to
$2\cdot\sqrt{24/168}\approx$\Sexpr{round(4*sqrt(6/length(before)),3)}),

in absolute value to see if the excess kurtosis is substantially different from zero. And take a look at the sign to see whether the distribution is platykurtic or leptokurtic.

Calculate and report the skewness and kurtosis for \emph{after}. Based on these values

We should do for this one just like we did previously. We would again compare the sample skewness and kurtosis values (in absolute value) to \Sexpr{round(2*sqrt(6/length(after)),3)} and Sexpr{round(4*sqrt(6/length(after)),3)}, respectively.

The mean is usually the best measure of central tendency to use when your data distribution is continuous and symmetrical, such as when your data is normally distributed. However, it all depends on what you are trying to show from your data.

The mode is the least used of the measures of central tendency and can only be used when dealing with nominal data. For this reason, the mode will be the best measure of central tendency (as it is the only one appropriate to use) when dealing with nominal data. The mean and/or median are usually preferred when dealing with all other types of data, but this does not mean it is never used with these data types

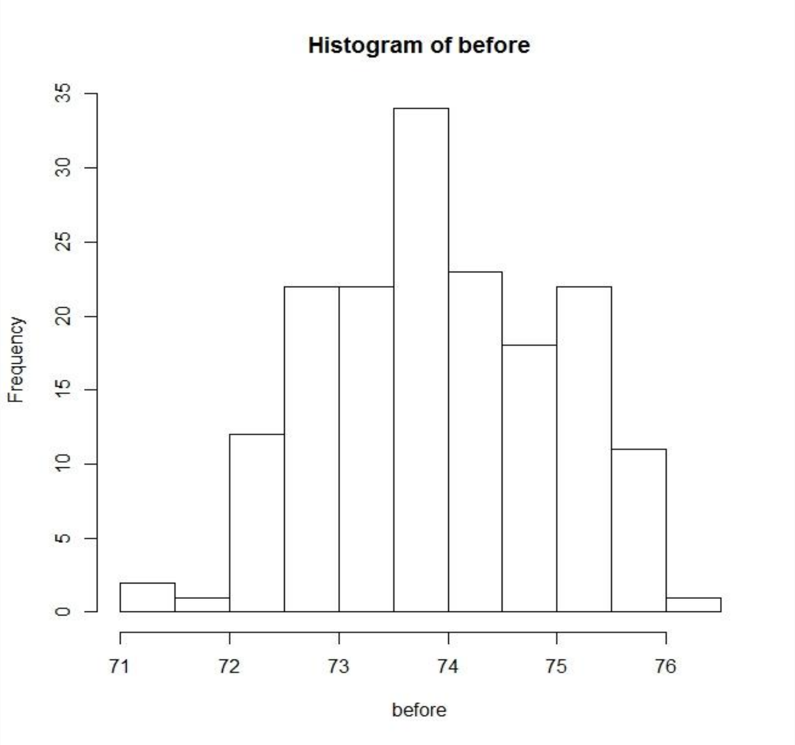
The median is usually preferred to other measures of central tendency when your data set is skewed (i.e., forms a skewed distribution) or you are dealing with ordinal data. However, the mode can also be appropriate in these situations, but is not as commonly used as the median.

Plot histograms of \emph{before} and \emph{after} and compare them.
The graphs are shown below.

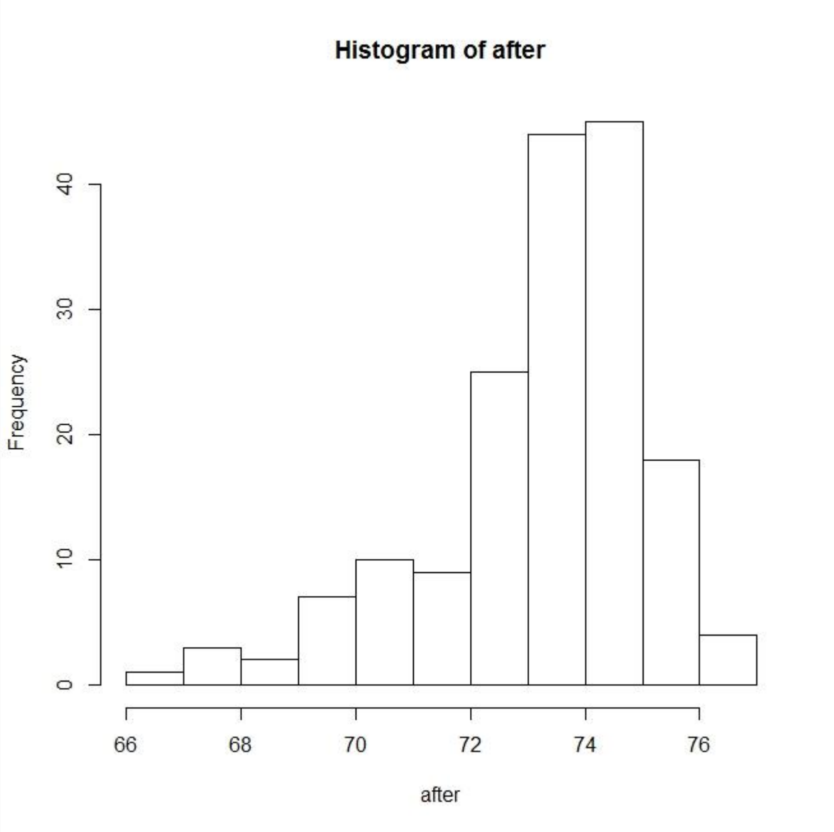
```
\begin{center}
<<echo = FALSE, fig=true, height = 4.5, width = 6>>=
hist(before, xlab="before", data=RcmdrTestDrive)
@
\par\end{center}
```

```
\begin{center}
<<echo = FALSE, fig=true, height = 4.5, width = 6>>=
hist(after, xlab="after", data=RcmdrTestDrive)
@
\par\end{center}
```

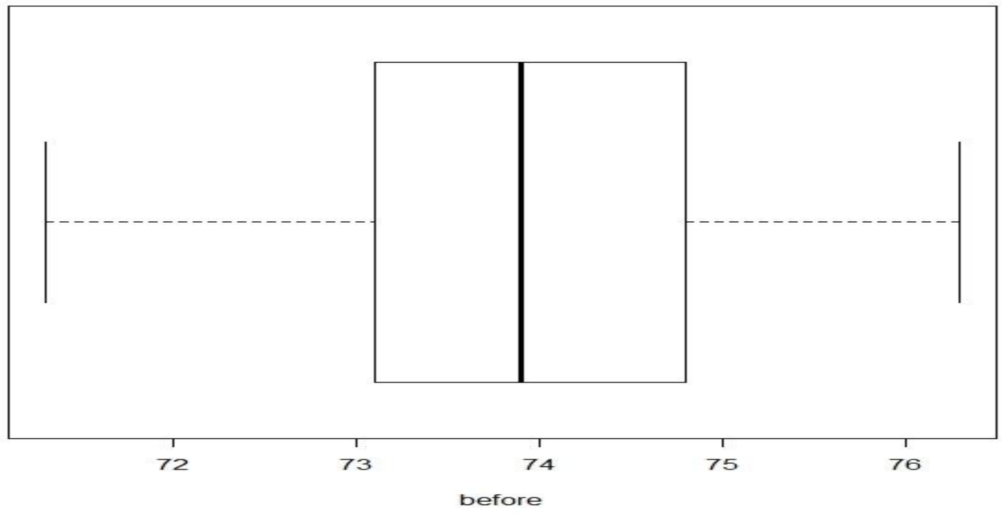
Answers will vary. We are looking for visual consistency in the histograms to our statements.

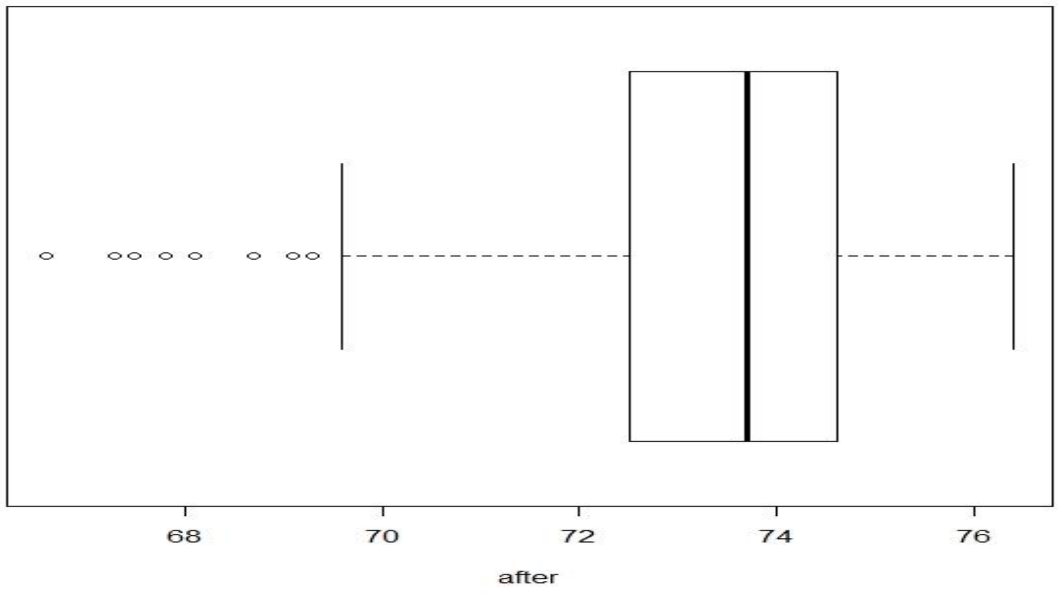


B



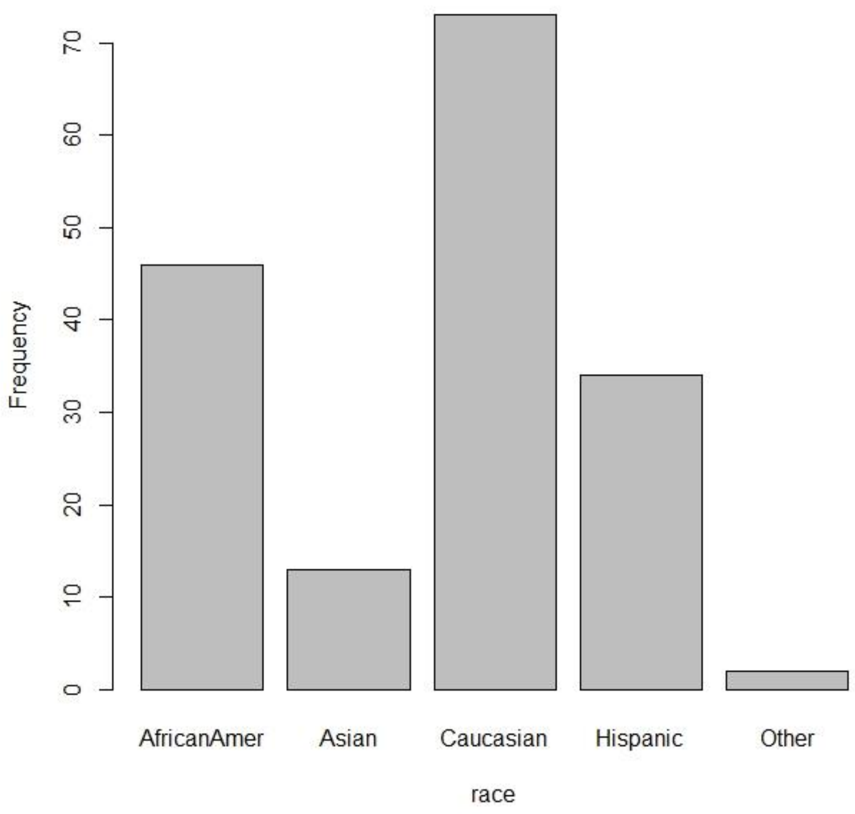
Before has symmetric mount shaped distribution excellent measure of center would be the sample standard deviation. After is left skewed we should use median absolute deviation.It is also acceptable to use IQR but it is necessary to rescale it appropriately by dividing by 1.349.





```
str(RcmdrTestDrive)

## 'data.frame':    168 obs. of  9 variables:
##  $ order    : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ smoking  : Factor w/ 2 levels "Nonsmoker","Smoker": 1 1 1 1 1 1 2 1 1 1
...
##  $ gender   : Factor w/ 2 levels "Female","Male": 1 2 1 1 1 2 2 2 1 1 ...
##  $ race     : Factor w/ 5 levels "AfricanAmer",..: 3 1 3 3 4 3 4 4 3 4 ...
##  $ before   : num  72.6 75.3 75.5 71.3 74.3 73 72.4 73.6 73.7 74.6 ...
##  $ after    : num  75.2 73.2 74.5 74.6 73.8 73.6 70.7 74 75.9 74.8 ...
##  $ salary   : num  619 545 550 616 543 ...
##  $ reduction: int  9 62 19 30 105 43 229 40 101 440 ...
##  $ parking  : int  2 1 4 1 1 1 5 1 2 1 ...
```



R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

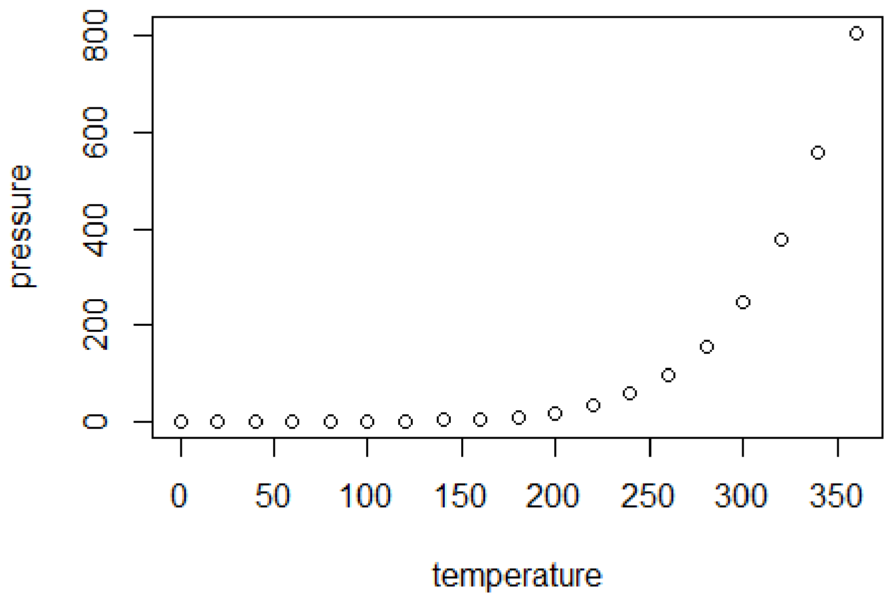
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```r
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.