Assignment - logical operators and dataframes

# Run the following code to generate the data for this assignment. (You don't need to understand this code in order to successfully complete this assignment).

# The questions in this assignment refer to a dataframe that contains information about stocks.
# You can use the following code to generate some random data about ficticious stocks which you can use as
# the data for the assignment. Note that the actual values in the data will should affect the code that you
# write for your answers. However, it does help to have some data on which to test out your answers.

generateRandomStockData <- function (numberOfStocks=2000, seed = 1) {

 set.seed(seed) # this will "prime" the random number generator

 # Generate the tickers

 getTicker <- function(...) {

 # Generate tickers

 numberOfLetters <- sample(2:4, size=1, prob=c(.15, .2, .65) )

 tickerLetters <- sample(letters, size=numberOfLetters, replace=TRUE)

 paste0(tickerLetters, collapse="")

 }

 uniqueTickers <- vector()

 while (length(uniqueTickers) < numberOfStocks){

 #cat("length=",length(uniqueTickers))

 tickers <- c(uniqueTickers, sapply(1:numberOfStocks, getTicker))

 uniqueTickers <- unique(tickers)

 uniqueTickers <- uniqueTickers[1:min(length(uniqueTickers),numberOfStocks)]

 }

 # Generate the exchange

 exchanges <- sample(c("nasdaq","nyse"), size=numberOfStocks, replace=TRUE, prob=c(.6, .4) )

 # Generate the sector

 sectors <- sample(c("tech","energy","pharm","consumer"), size=numberOfStocks, replace=TRUE, prob=c(.3, .3, .2, .2) )

 # Generate the open prices

 numberOfCheapStocks <- floor(.15 \* numberOfStocks)

 numberOf100Stocks <- floor(.40 \* numberOfStocks)

 numberOf1000Stocks <- floor(.18 \* numberOfStocks)

 numberOfRemainingStocks <- numberOfStocks - numberOfCheapStocks -

 numberOf100Stocks - numberOf1000Stocks

 openPrices <- sample(095:105, numberOfCheapStocks, replace=TRUE) / 100

 openPrices <- c(openPrices, sample(9800:10200, numberOf100Stocks, replace=TRUE) / 100)

 openPrices <- c(openPrices, sample(99000:101000, numberOf1000Stocks, replace=TRUE) / 100)

 openPrices <- c(openPrices, sample(1:200000, numberOfRemainingStocks, replace=TRUE) / 100)

 news <- sample(c(TRUE, FALSE), numberOfStocks, replace=TRUE)

 closePrices <- openPrices \* (1 + rnorm(length(openPrices), mean=0, sd=.10))

 closePrices <- round(closePrices, 2)

 df <- data.frame( ticker=uniqueTickers, exchange=exchanges, sector=sectors, open=openPrices, close=closePrices, news=news)

 df <- df[order(df$ticker),]

 rownames(df) <- 1:nrow(df)

 df

}

stocks <- generateRandomStockData(100)

stocks

# The Assignment

Assume that a dataframe named stocks contains information about stocks that traded today. The data is in the format shown below. Note that what is shown below is just some sample data. The dataframe may actually contain thousands of rows. Answer the questions below by writing the R code that will return the correct result.

* The open and close columns represent the opening and closing prices for that day.
* The "news" column contains logical TRUE/FALSE values. The value is TRUE if significant news about that stock came out that day (the news might be good news or bad news - you don't know). If no news came out that day then the value in the "news" column is FALSE.

**Sample stock trading data:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ticker** | **exchange** | **sector** | **open** | **close** | **news** |
| goog | nasdaq | tech | 1050.00 | 1062.50 | FALSE |
| msft | nasdaq | tech | 998.50 | 1001.50 | TRUE |
| mrk | nyse | pharm | 59.50 | 48.00 | TRUE |
| ziop | nasdaq | pharm | 30.20 | 39.20 | FALSE |
| wmt | nyse | consumer | 102.00 | 99.50 | FALSE |
| xon | nyse | energy | 80.00 | 80.05 | TRUE |
| fang | nasdaq | energy | 99.75 | 101.25 | TRUE |
| etc ... | etc ... | etc ... | etc ... | etc ... | etc ... |

You may use multiple R commands to build up to the final answer, however, the final answer must be executed using a single command. For example:

**Sample Question**: Show those stocks that are in the consumer sector and also those stocks that are in the pharm sector

**The following is a CORRECT answer that takes a single command**

> stocks [ stocks$sector == "consumer" | stocks$sector == "pharm" , ]

 ticker exchange sector open close news

3 mrk nyse pharm 59.5 48.0 TRUE

4 ziop nasdaq pharm 30.2 39.2 FALSE

5 wmt nyse consumer 102.0 99.5 FALSE

(CONTINUED ON NEXT PAGE ...)

**The following is also a CORRECT answer but this version uses more than one command**:

> consumerRows <- stocks$sector == "consumer"

> pharmRows <- stocks$sector == "pharm"

> stocks [ consumerRows| pharmRows , ] # this line returns all of the data

 ticker exchange sector open close news

3 mrk nyse pharm 59.5 48.0 TRUE

4 ziop nasdaq pharm 30.2 39.2 FALSE

5 wmt nyse consumer 102.0 99.5 FALSE

**HOWEVER, the following answer is WRONG. It's true that between the two commands all of the data is shown. However, to answer correctly your FINAL COMMAND must SHOW ALL OF THE DATA THAT WAS ASKED FOR**:

> stocks [ stocks$sector == "consumer" , ]

 ticker exchange sector open close news

5 wmt nyse consumer 102 99.5 FALSE

> stocks [ stocks$sector == "pharm" , ]

 ticker exchange sector open close news

3 mrk nyse pharm 59.5 48.0 TRUE

4 ziop nasdaq pharm 30.2 39.2 FALSE

## Questions

Write R code that answers each question. Remember that the order of operations for logical operators are ! is first & is second | is third. For some answers, it might be necessary for you to use (parentheses) to override R's natural order of operations. For example:

 Just as in math order of operations is important
 2+3\*4 is 14 but (2+3)\*4 is 20.

 Similarly in R
 TRUE|FALSE&FALSE is TRUE but (TRUE|FALSE)&FALSE is FALSE

If you are having trouble, it might help for you to review the contents of the Notes files for "dataframes" and for "logical operators". See the "Notes" folder on Canvas for these and other Notes files.

1. Show all nasdaq stocks.
2. Show all nasdaq stocks that went up in price.
3. Write a command that returns the number (i.e. how many) nasdaq tech stocks there are.
4. Show all **tech** and **energy** stocks.
5. Show all tech and energy stocks on the nasdaq that had news today.
6. Show all nasdaq stocks that went up by at least $10.00.
7. Show all nasdaq stocks that went up by at least 10%.
8. Show all nasdaq stocks that went up **or down** by at least **10%**
9. Show all stocks that **had news** and/or fluctuated in price by at least **$10.00**.
10. Show the average opening price of nasdaq stocks that went up
11. Show all **tech** and **energy** stocks that went **up or down** by at least 10%
12. Show all tech and energy stocks on the nasdaq that went up by 10% or that had news today.
13. Show all stocks that were "surprising" in that either they

had **no news** and the price went up or down by at least $10.00

or that **had news** but the price remained stable and didn't fluctuate up or down by more than $1.00

Your answer should include a single command that when executed displays BOTH types of "surprising" stocks but no other stocks.

1. Show the difference between the average price of nasdaq **tech** stocks closing prices and nasdaq **non-tech** stock closing prices
2. Show how many (i.e. a single number) of nasdaq stocks that opened or closed below $1
3. Show all nasdaq stocks that opened or closed below $1 or that opened or closed above $1000
4. Show **tech** stocks on the nasdaq that opened or closed below $1 or that opened or closed above $1000
5. Show **tech** and **fin** stocks on the nasdaq that opened or closed below $1 or that opened or closed above $100
6. Show **tech** stocks on the nasdaq that

had news come out today

 and also
opened or closed below $1 or opened or closed above $1000.

You should provide a single answer that includes a command that returns all the data for stocks that meet the specified conditions but no other stock data .

1. Show **tech** stocks on the nasdaq that opened or closed below $1 or that opened or closed above $1000.
In addition, show **any** stocks that had some news come out today.
2. Answer both parts - provide separate answers for parts (a) and (b)
	1. Did the nasdaq go up or down? Write a command that results in TRUE if the average closing price of nasdaq stocks is above the average opening price of nasdaq stocks.
	2. Write a command that shows those stocks that moved in the opposite direction to the nasdaq as a whole.

	The answer to part (b) should be a single answer that works whether or not the nasdaq went up or down. If the nasdaq went up then show the stocks that went down. If the nasdaq went down, then show the stocks that went up. If you run the same exact command with data from different days, your R code should give the correct answer no matter whether the nasdaq went up or down that day.
3. Show tech stocks whose move was 10% more than the average move for all tech stocks.
4. Show tech stocks whose move was

10% more than the average move for all tech stocks and in the same direction (up or down) of the average move for all tech stocks

or 10% more than the average move for all stocks (regardless of the exchange/sector) and in the same direction as the average move for all stocks

You should provide a single answer that includes a command that returns all the data for stocks that meet the specified conditions but no other stock data .