Module 2 Homework

ADD IN YOUR NAME

February 22, 2018

Module 2 Homework Assignments

This module consists of 4 homework assignments and corresponds to assignments 6, 7, 8, and 9 of the semester.

Remember to set eval to T and echo to T after you have answered a code chunk so that your code and plots will appear when you knit the document.

Module 2 Homework Assignment 1

Question 1

- a) Please review the following functions shown in class today. You should use the help documentation for each function in answering these questions. To access the help use ?function or help(function). Please describe the following for each function:
 - What are the inputs to the function?
 - What are the outputs from the function?
 - When is this function useful?

Example Answer:

- sum()
 - The sum() function takes one or more vectors of numbers as an argument. Additionally, it accepts the na.rm argument which allows sum() to be used even when NA values are present.
 - The output of the sum() function is a single number, the sum of all values contained in the inputs
 - The function is useful when looking to find the sum of multiple numbers. It is also useful with TRUE/FALSE values.

Please discuss the following functions:

- for
- if
- else
- identical()
- str detect()
- b) This is a repeat of an in-class exercise from the Module 2 Day 1 lecture. If you have your code saved from class then you can just place it word for word. Create a loop that will take the square root of a positive number or give us an NA if the number is negative. Save the results in a vector you initialized outside of the loop.

Question 2

For the rest of Assignment 1 we are going to be thinking about investing and retirement.

Let's think about things we could do with money we have left after we've paid our bills. Three main options we will be looking at today:

- Savings account
- Stock market
- Bonds

We are going to look at how the total value of our investment would change over a ten year period, starting with our savings account. Savings rates change, but over the last ten years they were very low. Let's say you have a 2% interest rate on your money. This means that every year you get 2% of the money you have in the bank added to your account.

Let's look at a simple example, where we start with \$100 in our savings account and grow it.

```
# A = P(1+r*t)
# where A = account balance, P = principal, r = interest rate,
# and t = periods of time (in this case, one year)
P = 100
# After one year:
A = P*(1 + 0.02)
print(A)
```

- ## [1] 102
- a) How much money did we make after one year?
- b) Use a for loop to answer this: what would our account balance be after ten years, with an initial investment of \$100 and interest rate of 2% annually?

Hint: you will update the variable account_balance ten times.

You should get the following value:

```
## [1] 121.8994
```

c) How much money would we have in our account after ten years, if each month we add an additional \$100 to our account?

You should get the following value:

```
## [1] 13394.09
```

d) How much of this account balance would be made from interest?

Hint: figure out the amount of total deposits made and subtract that from the account balance.

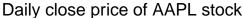
Question 3

Now we are going to read in and clean some stock data. Later, we will build portfolios using this stock data to model your retirement savings!

a) Read in the following files: AAPL.csv, FORD.csv, GE.csv, JPM.csv, EXC.csv and JCP.csv from the Hw_Data folder. List the column names of each dataset.

- b) Two of these data sets have columns that the others don't. Select out the columns that these datasets have, that are unlike the others.
- c) Convert the date colums into a date object for each data frame.
- d) Pick one stock other than AAPL and plot its close price over time. Be sure to include appropriate titles and labels.

Your chart should look like this.





e) What do you notice about the stock's trend over time? (Hint: talk about dips, highs, general trends.)

Module 2 homework assignment 2

Question 1

Please review the following functions shown in class today. You should use the help documentation for each function in answering these questions. To access the help use ?function or help(function). Please describe the following for each function:

- What are the inputs to the function?
- What are the outputs from the function?
- When is this function useful?

Example Answer:

• sum()

- The sum() function takes one or more vectors of numbers as an argument. Additionally, it accepts the na.rm argument which allows sum() to be used even when NA values are present.
- The output of the sum() function is a single number, the sum of all values contained in the inputs
- The function is useful when looking to find the sum of multiple numbers. It is also useful with TRUE/FALSE.

Please discuss the following functions:

```
    %in%
    left_join()
    str_replace()
    str_split_fixed()
    str_trim()
```

Question 2

Rerun your code from last week so that all of our data is clean before we start today.

b) Using your FORD data set and the gather function, transform your close and volume columns so that have a key column called: "measure" and a value column called "amount". My example below uses the AAPL dataset.

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
## # A tibble: 3 x 2
##
     measure amount
##
       <chr>
              <dbl>
## 1
        date
              17539
## 2
        date
              17536
## 3
        date
             17535
```

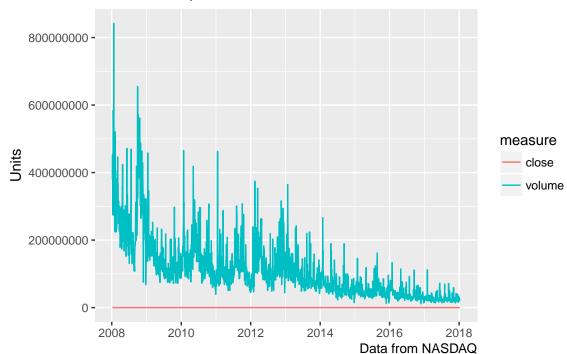
b) Uh oh, our date values were transformed. What we wanted were three columns: date, measure, and value, but we only got two because we forgot to protect our date column. Using your FORD data, create transformed_FORD again but this time make sure that our date column is kept intact.

Your answer should look like this, again, I am using the AAPL data set *Hint: you will need to use the - sign*

```
## # A tibble: 3 x 3
## date measure amount
## <date> <chr> <dbl>
## 1 2018-01-08 close 174.35
## 2 2018-01-05 close 175.00
## 3 2018-01-04 close 173.03
```

c) Part of why gather is so useful is that it makes plotting with ggplot really easy! Now that I have a single column with multiple measure variables, I can use it to control ggplot aesthetics, such as color. Using your new transformed_FORD data, make a plot similar to the one I am making with the AAPL data.

AAPL close price and volume



- d) What units should we use for the y-axis of our plot showing the closing price and volume? Does it make sense to have these two things plotted on the same chart?
- e) Just as before we used gather to transform our data from wide to long, we can also use spread() to transform our data from long to wide. Using spread() turn your long version of FORD, transformed_FORD back into the wide version of the data, your data set should be identical to the FORD data you started with.

Your output should look similar to mine

```
## # A tibble: 3 x 3
## date close volume
## <date> <dbl> <dbl>
## 1 2008-01-08 24.4643 379755714
## 2 2008-01-09 25.6286 453136596
## 3 2008-01-10 25.4314 370297661
```

Question 3

Now we are almost ready to join our data together. But if we do that with the data as it is, look what happens:

```
AAPL_FORD <- left_join(AAPL, FORD, by = c("date"))
names(AAPL_FORD)
```

R saw that all the column names were the same, so it assigned its own! But we won't know which price belongs to which stock by just seeing x and y. So it's better to rename the columns ourselves *before* joining the data.

- a) Rename the close and volume columns to AAPL_close, AAPL_volume, etc. We leave date the same so that we can join on it later.
- b) Join all four data sets into one, and call it stock_data. Hint: you will have to use join multiple times.

Question 4

a) Select all of the close price columns from stock_data and use gather() to create a suitable data frame for plotting, with the columns date, Stock, and Close. Call the data frame close_price_data.

Here is an example of what your data should look like.

b) We could plot our data as is, but the values in the Stock column are not very clean. Rename the values of the Stock column in close_price_data from, for example, AAPL_close to AAPL. Hint: consult your notes on str_split_fixed().

Here is an example of what your data should look like

```
## # A tibble: 6 x 3

## date Stock Close

## < date> <chr> <dbl>
## 1 2018-01-08 AAPL 174.35

## 2 2018-01-05 AAPL 175.00

## 3 2018-01-04 AAPL 173.03

## 4 2018-01-03 AAPL 172.23

## 5 2018-01-02 AAPL 172.26

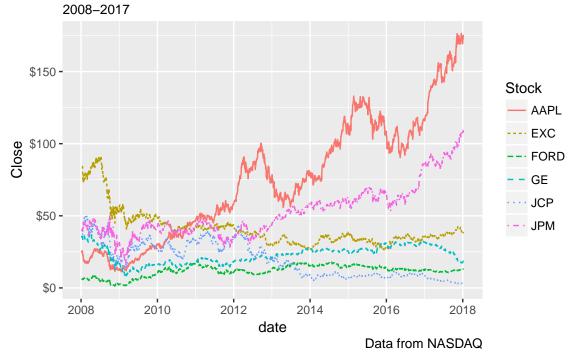
## 6 2017-12-29 AAPL 169.23
```

c) Let's plot our data, now that it's in gathered form. Reproduce the plot below, noting the labels and titles.

Here is what your plot should look like

Warning: Removed 10 rows containing missing values (geom_path).

Daily closing prices for selected stocks



d) What are your major takeaways from this chart? What stocks would you want in your retirement portfolio, if you were retiring in 2018? Did our stocks maintain their relative price order (were the stocks with the highest prices in 2008 still our highest priced stocks in 2018?)

Question 5: Adding bond data

```
bond_data <- read_csv("Hw_Data/bond_data.csv") %>%
  mutate(date = as.Date(date, format = "%Y-%m-%d"))
```

- a) What are the names of the two Bond funds we are looking at?
- b) Join close_price_data with this bond data. (Hint: we are hoping to capture all of the rows in both datasets, so left_join() will not give the desired result.)
- c) Create a dummy variable called Bond that equals TRUE if the Stock name refers to a bond, and FALSE otherwise. What will this variable be useful for?
- d) Plot stock_bond_data, reproducing the chart below.
- ## Warning: Removed 10 rows containing missing values (geom_path).



Module 2 Homework Assignment 3

Question 1

Please review the following functions shown in class today. You should use the help documentation for each function in answering these questions. To access the help use ?function or help(function). Please describe the following for each function:

- What are the inputs to the function?
- What are the outputs from the function?
- When is this function useful?

Example Answer:

- sum()
 - The sum() function takes one or more vectors of numbers as an argument. Additionally, it accepts the na.rm argument which allows sum() to be used even when NA values are present.
 - The output of the sum() function is a single number, the sum of all values contained in the inputs
 - The function is useful when looking to find the sum of multiple numbers. It is also useful with TRUE/FALSE values.

Please discuss the following functions:

- str sub()
- require()
- warning()
- stop()
- range()

Question 2

Now that we have all of data nicely cleaned and put together, it is time for us to use it so that we can analyze it to create our very own retirement strategies.

a)

We will need to look at monthly returns for each of our stocks and bond funds. To do this we will need our stock prices for only the last day of each month for which we have data. Using lubridate: create a year, a month, and a day column in our dataframe for each row which breaks out the components of our date columns.

Here is what your data should look like.

```
## # A tibble: 6 x 7
##
           date Stock Close
                               Bond
                                     year month
                                                    day
##
         <date> <chr> <dbl> <lgl> <dbl> <dbl> <dbl>
## 1 2018-01-08
                   BND 81.27
                               TRUE
                                     2018
                                               1
                                                      8
## 2 2018-01-05
                   BND 81.26
                               TRUE
                                      2018
                                               1
                                                      5
## 3 2018-01-04
                   BND 81.37
                                      2018
                                                      4
                               TRUE
                                               1
                                                      3
## 4 2018-01-03
                   BND 81.41
                               TRUE
                                     2018
                                               1
                                                      2
## 5 2018-01-02
                   BND 81.34
                               TRUE
                                     2018
                                               1
## 6 2017-12-29
                   BND 81.57
                               TRUE
                                     2017
                                              12
                                                     29
```

b) What is the last day of every month in which we have data for each stock or bond? Using our new year, month, and day columns and dplyr, add a new column: max_day which has the number of maximum day for each combination of year, month, and stock

Here is what your data should look like.

```
## # A tibble: 6 x 8
           date Stock Close
##
                               Bond
                                    year month
                                                    day max_day
##
         <date> <chr> <dbl> <lgl> <dbl> <dbl> <dbl>
                                                 <int>
                                                          <dbl>
## 1 2018-01-08
                   BND 81.27
                               TRUE
                                      2018
                                               1
                                                      8
                                                               8
## 2 2018-01-05
                   BND 81.26
                               TRUE
                                     2018
                                               1
                                                      5
                                                               8
## 3 2018-01-04
                   BND 81.37
                               TRUE
                                     2018
                                               1
                                                      4
                                                               8
                                                      3
## 4 2018-01-03
                   BND 81.41
                               TRUE
                                      2018
                                               1
                                                               8
## 5 2018-01-02
                   BND 81.34
                               TRUE
                                     2018
                                               1
                                                      2
                                                               8
## 6 2017-12-29
                   BND 81.57
                               TRUE
                                     2017
                                              12
                                                     29
                                                              29
```

c) Now that we know our last day of the month for which we have data for a given stock, we can use filter to only take our end of month data (rows where day = max_day) and then sort our data to be increasing (so that our last row is in 2017).

Note: We only want data for full months, so make sure not to include January 2018 data

Here is what your data should look like.

```
## # A tibble: 6 x 4
##
           date Stock
                         Close
                                Bond
##
                         <dbl> <lgl>
         <date> <chr>
## 1 2008-01-31
                  BND 78.3500
## 2 2008-01-31 VBLTX 11.7400
## 3 2008-01-31
                 AAPL 19.3371 FALSE
## 4 2008-01-31
                FORD
                       6.6400 FALSE
## 5 2008-01-31
                  JPM 47.4000 FALSE
```

d) Now that we have our month end values for each stock, we can find the growth percentage for each month using dplyr::lag function. Using group_by, mutate, dplyr::lag, and whatever else you need, find the growth in the closing price for each stock in each month.

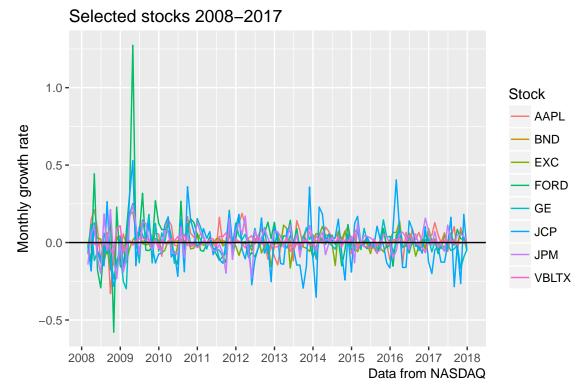
Note that growth is: (final price/initial price) - 1

Here is what the tail of your data should look like.

```
## # A tibble: 6 x 5
  # Groups:
               Stock [6]
##
           date Stock Close Bond
                                          growth
##
         <date> <chr>
                       <dbl> <lgl>
                                           <dbl>
## 1 2017-12-29
                 AAPL 169.23 FALSE -0.015245854
## 2 2017-12-29
                 FORD
                       12.49 FALSE -0.002396166
## 3 2017-12-29
                  JPM 106.94 FALSE 0.023153463
                       17.45 FALSE -0.045926736
## 4 2017-12-29
                   GE
## 5 2017-12-29
                  JCP
                        3.16 FALSE -0.045317221
## 6 2017-12-29
                  EXC
                       39.41 FALSE -0.055142652
```

The first month of data should all have NA values for growth, why?

e) Remove the rows of our dataset with an NA value and then make a line chart showing the monthly growth values for stocks and bonds. Be sure to use appropriate labelling and add a line at 0 using geom_hline().



What are your takeaways from this chart? Do any stocks seem consistently better than any other? Which one seems the most erratic?

e) Why do we care about growth rates and not just the closing price values?

Question 3

It's time to build our retirement portfolio simulator. Let's start by trying to recreate what would happen if we invested in JCPenny in 2008.

```
# Pull out all the dates we will loop through
dates <- unique(month_end$date)
# Invest $100 at the beginning, no monthly contribution
investment_value <- 100

for(d in dates){
    investment_value <- investment_value* (1 + month_end[month_end$date == d & month_end$Stock == "JCP", "growth"])
}
names(investment_value) <- "Value"
print(investment_value)</pre>
## Value
## 1 6.673706
```

Wow, what a terrible investment, our \$100 turned to \$7 in ten years!

a) One easy way we can make our code better is by getting rid of that whole month_end[month_end\$date ... section. We can do this by making a function.

Write a function: growth_finder with the following characteristics

- Inputs: date, stock symbol, reference table (use month end)
- Output: growth value for a stock on a given date

Test your function and make sure you get the same output:

[1] 0.0373881

Let's re-envision our retirement forecasting loop with this new function

```
# Pull out all the dates we will loop through
dates <- unique(month_end$date)
# Invest $100 at the beginning, no monthly contribution
investment_value <- 100

for(d in dates){
  investment_value <- investment_value* (1 + growth_finder(d, "JCP", month_end))
}
print(investment_value)</pre>
```

[1] 6.673706

- b) Now let's replace the right side of our for loop assignment statement: investment_value * (1 + ...) with a new function called: monthly_update
 - Inputs: investment_value at beginning of the month, all inputs to growth_finder
 - Outputs: investment value at end of month

Check that you get the same result as I do:

[1] 103.7388

Let's envision our retirement loop with our new function now:

[1] 6.673706

c) Using the code from above, fill in the below loop so that we can forecast for two stocks instead of just one

Your answer should look like the following:

[1] "our JCPenney value is 3.34 dollars and our Apple value is 437.58 dollars."

Question 4

a) Instead of using a for loop, let's turn the code for question 3d into a function that also allows us to divide up our allocated contributions among our different stocks. Make sure that you get the same results as I do when you run the function:

```
portfolio_finder(month_end, 1000, "JCP", "AAPL", 0.5, 0.5)
```

```
## From our initial allocation of
## 1000
## our investment in
## JCP
## is now worth
## 33
## dollars and our investment in
## AAPL
## is now worth
## 4376
## dollars. In total our portfolio is worth
## 4409
## dollars.
```

So, we've managed to update our portfolio_finder() function to work with 2 stocks, but in our data we have 8 different stocks! Instead of adding in more arguments to our function, we should create a single function to do all of the work for a single stock and then write a wrapper around it so that we can have that function do our work for all 8 stocks separately.

Below I've written a function to do just that and incorporate monthly contributions into our investments

Remember, the initial value is how much we invest on the first day, while the monthly contribution is a fixed sum of money that we add to the investment portfolio each month.

- Inputs: stock symbol, reference table, initial investment, monthly allocation, share of contributions to be invested in that stock
- Output: final value of our investment

b) Using our single_stock_value() function rewrite the portfolio_finder() function.

Inputs: reference table, initial contribution, monthly contribution, vector of stock symbols, vector of allocation shares for initial/monthly contributions

Output: Table showing the stocks invested in and their final values

Test your function to be sure that you get the same results as I do.

```
portfolio_finder(month_end, 100, 50,
                 c("BND", "VBLTX", "AAPL", "FORD", "JPM", "GE", "JCP", "EXC"),
                 c(0.05, 0.05, 0.3, 0.1, 0.1, 0.1, 0.1, 0.2))
##
      Name
               value
## 1
            303.8261
       BND
## 2 VBLTX
           336.3109
## 3
      AAPL 6634.7768
      FORD
           812.7566
## 5
       JPM 1353.9178
## 6
        GE
           499.5667
## 7
       JCP
            160.0110
       EXC 1233.9935
## 8
```

- c) Time to practice good function hygiene and update our portfolio_finder function.
 - what value should all of our contribution shares add up to? Make sure that they do.

Test that your version gives the same results as mine for the following call:

```
##
      Name
               value
## 1
       BND
            3038.261
## 2 VBLTX
            3363.109
      AAPL 66347.768
## 3
## 4
      FORD
           8127.566
## 5
       JPM 13539.178
## 6
           4995.667
        GΕ
## 7
       JCP
           1600.110
```

- d) Now let's try creating some hypothetical portfolios to see how large your investments could have gotten.
 - Create 3 different portfolios, each with a \$100 initial contribution and a \$50 monthly contribution
 - For each portfolio you have to invest in at least 4 of the bonds or stocks and you can only put a maximum of 50% of your funds to any single stock/bond in your portfolio. (No, you can't only buy AAPL)
 - Of your fictional portfolios which one was worth the most in total value at the end of 10 years? Which one was worth the least? What was different about your best and worst portfolios?

Module 2 Homework assignment 4

Question 1

Last week we wrote a function: portfolio_finder which returns the final values of a stock/bond portfolio based on initial and monthly contributions. But this function loses a lot of data since we do not return any of our intermediary monthly values.

Instead, let's write a function that will output a vector showing the value of a single investment over the entire time period

- Inputs: stock symbol, reference table, initial investment, monthly allocation, share of contributions to be invested in that stock
- Output: a vector of the values of the investment over time
- a) Using the single_stock_value function I've given you, write a new function, portfolio_finder2 that returns a data frame with the historical value of multiple stock symbols over time

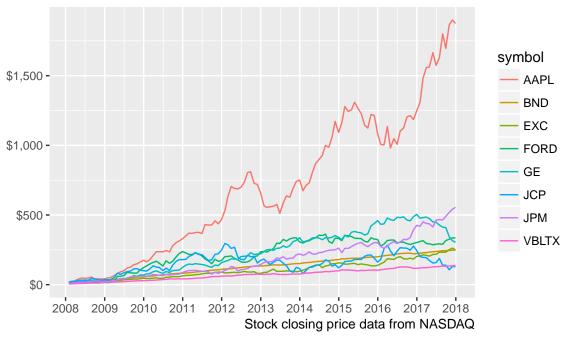
Your output should look like this:

b) Now that we can track our stock performance over time, come up with a fake investment strategy. You must allot at least 5% of your investment and monthly contribution to each stock/bond. Make a line chart showing the performance of each stock in your portfolio. Be sure to correctly label and title your chart.

This is an example, yours will have different lines

Components of hypothetical portfolio

\$100 initial investment and \$20 monthly contribution



Question 2

a) One major thing we are still missing from our portfolio_finder2 is a column which keeps a running total of our portfolio's value. Update the function to include a column with the total value of all stocks/bonds for each month.

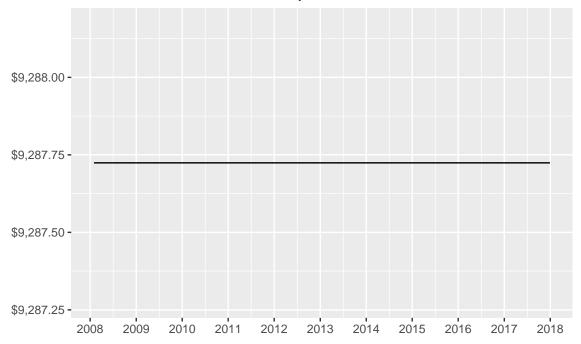
Try using the rowSums function and be sure that your solution works no matter how many stocks you choose

Here is what the output for your new function should look like:

b) Create a hypothetical portfolio and graph its total value over the course of the ten years of investment. Again, you must allocate at least 5% of your initial amount and monthly contributions to each stock/bond. Here is an example plot:

Total value of hypothetical portfolio

\$500 initial investment and \$50 monthly contribution



c) Really what we are interested in though is how much money our portfolio made beyond the money we directly invested (our initial money and our monthly contributions). Write a new function: contribution_adder that uses the output of our portfolio_finder2 function and adds columns which keep a running total of how much contributions we make directly to the portfolio and how much profit our portfolio generates.

Your output should look like this

```
input_data <- portfolio_finder2(month_end, 100, 20,</pre>
                             c("AAPL", "BND"),
                             c(0.5, 0.5))
head(contribution_adder(input_data, 100, 20))
           date
##
                    AAPL
                               BND
                                      Total Contribution
                                                            Profit
## 1 2008-02-01 4685.732 1246.537 5932.269
                                                     100 5832.269
## 2 2008-02-29 4685.732 1246.537 5932.269
                                                     120 5812.269
## 3 2008-03-31 4685.732 1246.537 5932.269
                                                     140 5792.269
## 4 2008-04-30 4685.732 1246.537 5932.269
                                                     160 5772.269
## 5 2008-05-30 4685.732 1246.537 5932.269
                                                     180 5752.269
## 6 2008-06-30 4685.732 1246.537 5932.269
                                                     200 5732.269
```

Question 3

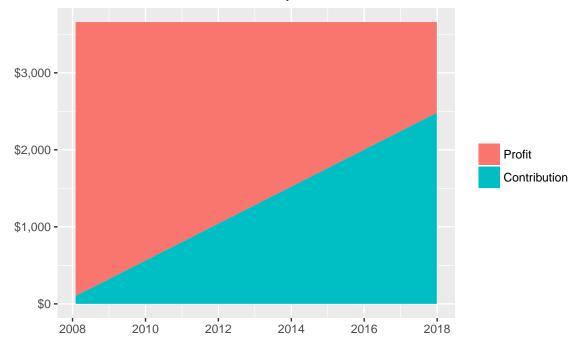
a) Using our new contribution and profit columns create an area chart for a hypothetical portfolio. Again, make sure that each stock/bond receives at least 5% of your contributions. You should not copy my allocation.

Here is an example chart and portfolio

Make sure that your profit is on top of your contribution for the chart, use the factor() function and set the levels argument

Breakout of Hypothetical investment portfolio 2008–2017

\$100 initial investment with \$20 monthly contribution



b) As you have seen previously, some investments are risky but have high risk and high return like stocks. Buying Apple in 2008 would make you a millionaire now while buying JCPenny in 2008 could have wiped out your savings. Bonds, on the other hand are low risk and low return. I want you now to create three types of portfolios: risk averse, risk neutral and risk loving.

For your risk averse portfolio, 50% of your investment allocations should be in bonds (BND and VBLTX)

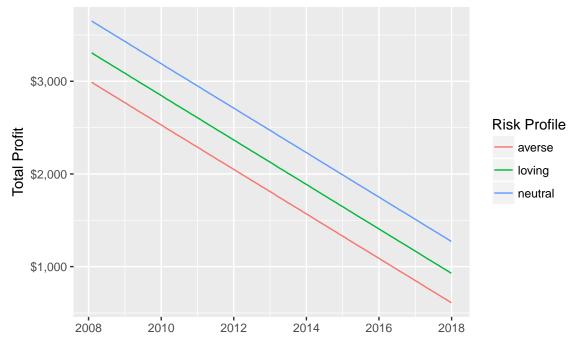
For your risk neutral portfolio, 20% of your investment allocations should be in bonds For your risk loving portfolio, 5% of your investment allocations should be in bonds For all of your portfolios you must have some investment in each stock/bond

- c) Create area plots for our three portfolios and label them appropriately. Be sure to display them in your final submission.
- d) Area plots alone do not really help us compare strategies, create a line plot showing the Total values for our three strategies.

Example below

Total Profit of investment portfolios over ten-year period

\$100 initial contribution with \$20 monthly contributions



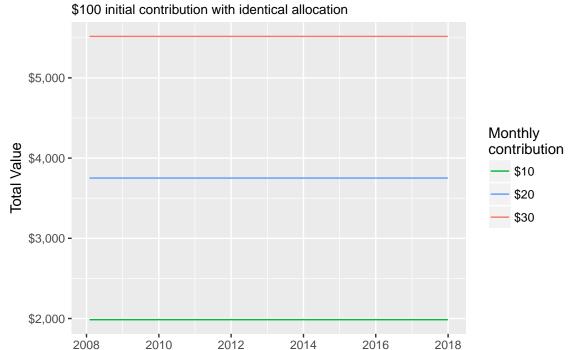
Surprisingly, for my example the risk neutral portfolio outperformed the risk loving portfolio. This is because my risk loving portfolio allocated too much to a few stocks which ultimately underperformed.

e) Based on my example above, why might financial advisors recommend investing in many different stocks/bonds as opposed to just one or two?

Question 4

- a) So far we've looked at how changing the allocation yields different results. One other thing we could change is our initial and monthly contributions. Create three new profiles, low, medium, and high with the same initial value but monthly contributions of \$10, \$20, and \$30 respectively. Each portfolio should have the same allocation of stock/bonds.
- b) Similar to question 3d, create a plot showing the total value of each portfolio over time Example Below

Total Value of investment portfolios over ten-year period



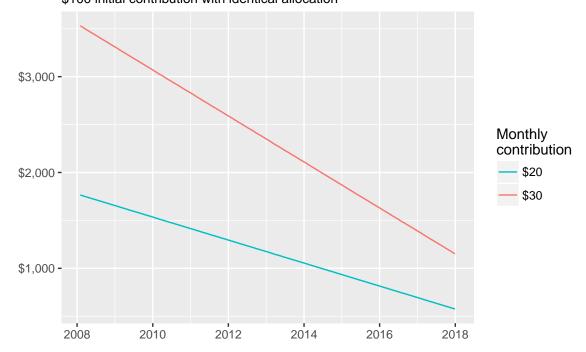
c) How much extra profit did you make for your medium and high contribution portfolios compared to your low contribution portfolios? How much more valuable is your high contribution portfolio than your low contribution portfolio, is the difference greater than the amount of raw contributions you gave to one portfolio over the other?

2018

2008

To find this you can just subtract the profit in your low portfolio from the profit in your high and medium portfolios

Profit of investment portfolios relative to \$10 monthly contribution \$100 initial contribution with identical allocation



For my high contribution portfolio, I invested a total of \$20*\$120 = \$2400 more from my pockets but had a higher profit of almost \$1200. Even though I only spent an extra \$2400 on my high contribution portfolio, after ten years that portfolio is worth \$3600 more than my low-contribution portfolio!

- d) For your own portfolios, how much "extra profit" does your high contribution portfolio have at the end of the ten years than your low contribution portfolio? (for me it is \$1200)
- e) Experts often recommend a mix of stocks and bonds in a retirement portfolio, but recommend having a "growth" portfolio (lots of stocks) when you're young, that you switch to an "income portfolio" (lots of bonds) as you get closer to retirement. Using your insight and the information from this assignment, why might they suggest this strategy?