

# HW #1: Exploring Financial Data

## **Purpose of This Assignment**

In this assignment, you will do some basic exploratory data analysis with a finance twist.

## **Disclaimer**

None of this is to be taken as financial advice. It is not recommended that you put all your money into an asset class based off of a homework assignment. Also, none of the scenarios in this assignment or in this class are based on real people. If you happen to know someone who fits a scenario, then that is just a coincidence.

## Instructions

The submission instructions and grading criteria are on the last page of this pdf. <sup>1</sup>

## Exploring Financial Data

Your friend has just learned that his long-lost relative, who passed away years ago, had named him as a beneficiary in a will that was just found. His Great-Aunt Belinda was an original cypherpunk in the 1980s. Codename 82-ce-31-c5-94-42-2d-c5-aa-ae-cf-7f-a6-45-1b-ff-b7-45-92-e6-93-6a-72-ba-a0-d6-25-9b-77-5c-e2-b6-61-48-68-d6-09-90-c1-0e-12-70-23-9e-44-9b-27-12-01-3a-41-f5-32-8f-8a-0f-00-31-7b-8b-b2-f3-59-0e-20-40, but you can simply call her 03-5b-1a-9f-3b-5f-d4-a2-22-e7-86-11-cb-22-b1-35-ef-68-4c-7c-39-c6-e1-30. Shortly before her passing, she bought \$5000 worth of Bitcoin (BTC) on January 6, 2016. She also created a portfolio with \$2500 in Ripple (XRP) and \$2500 in Litecoin (LTC) She created another portfolio which consisted of \$1250 in Ethereum (ETH), \$1250 in Dash (DASH), \$1250 in Peercoin (PPC), and \$1250 in Stellar (XLM). These were all in the top 10 cryptocurrencies by market capitalization back on January 6, 2016 when she made these purchases. In her will, she says that your friend will have first dibs on one of these portfolios, and his distant cousins who he has never met would have the leftovers. He worries that if he takes too long to decide, then the cousins will try to fight him in court to contest the will. He wants your advice. However, before you make a recommendation, you want to understand these portfolios better by examining the historical price data.

For the remainder of the problem, define the three portfolios as follows: Portfolio 1: portfolio with BTC; Portfolio 2: portfolio with XRP and LTC; Portfolio 3: portfolio with ETH, DASH, PPC, and XLM.

**Note:** As a simplification, **only use closing price data** from January 6, 2016 to January 6, 2023 (**including both of those dates**) for all of sub-questions of Questions 1 and 2. The closing price is in the ‘close’ column for each dataset. Also assume that you can buy fractional amounts (i.e., no need to find the max number of BTC below some fractional amount of BTC). Ignore any *airdrops* or *staking rewards* these cryptocurrencies may have had (if any of these have had any *hard forks* or migrated to a *proof-of-stake consensus algorithm*). Also, every dollar amount (both in the assignment statement and in the dataset) is in terms of USD (US dollars). Your friend is living in the States and cares about the portfolio in USD.

**Tip:** It will be useful for you to think at a high level about out how to make a for-loop for some of the questions (so that if instead of seven, if I had asked you about twenty cryptocurrencies or twenty stocks, you would be able to generalize your approach without much extra effort). However, we may not have learned the functions that would help us write a loop to solve the problem yet, so don’t worry about actually writing one. We will see more loops in the coming weeks and learn more tools/techniques later this semester that would help you simplify your code and make it more efficient. Look back on this assignment in a month to see what you might do differently with your new skills then.

Before you begin, load the data by saving the ‘HW1\_data.rda’ file from the HW 1 D2L folder to your working directory. If you do not know where your working directory folder is, then use the `getwd()` function. If you want to set your working directory to another folder, then you can use `setwd()`:

```
setwd("C:/Documents/Teaching/HW") # Modify this to your folder
```

To load the data, use the `load()` function:

```
load("HW1_data.rData") # Loads data from the working directory into the current session
```

The data was obtained from CoinMarketCap (<https://coinmarketcap.com/>) using their API. We won’t focus on retrieving the data in this assignment; you will learn how to gather data from the web later.

---

<sup>1</sup>Read them very carefully.

Now that we have the data, we can begin to explore the data.

### Starting Simple: Focusing on Individual Cryptocurrencies

1a. (11 points) Of the seven cryptocurrencies, which has had the highest long-run ROI (defined below)?

$$\text{Long-Run ROI} = \frac{\text{value on Jan 6, 2023} - \text{value on Jan 6, 2016}}{\text{value on Jan 6, 2016}}.$$

Write your number to the nearest hundred percent. In other words, if your ROI is 38.12891, then write as 3800. If your ROI is 8.123897, then write as 800. If your ROI is 123.8181, then write as 12400.

1b. (10 points) Which one has the highest mean daily return? Write as a percent, rounded to the nearest hundredth of a percent. Specifically, if your mean return is 0.123841, then write as 12.38. If your mean return is 0.0471841, then write as 4.72. If your mean return is 0.008981, then write as 0.90.

1c. (10 points) Which one has the lowest standard deviation of its daily return (i.e., lowest risk)? Write in the same format as 1b).

### Expanding the Exploration: Examining the Portfolios

2a. (15 points) This is likely the question your friend is most interested in if he plans on cashing out now: How much is each portfolio worth? Find out how much each portfolio was worth on January 6, 2023 (we'll use the same end date so that everyone gets the same answer on the D2L quiz associated to HW 1). Round to the nearest thousand dollars. For example, if the portfolio's worth is 121289, then write as 121000. If the portfolio's worth is 535660, then write your answer as 536000 on the D2L quiz.

2b. (15 points) Plot the value of the three portfolios over time, either as a combined plot or as separate plots. What was the highest value achieved by each of the three portfolios between January 6, 2016 and January 6, 2023? Round to the nearest thousand dollars. Among the three portfolios, which portfolio achieved the highest value?

2c. (6 points) When did Portfolio 1 achieve its all-time highest value within the time window of the dataset? When did Portfolio 2 achieve its all-time highest value within the time window of the dataset? What about Portfolio 3? Write some code so that you can find the month and year for which the all-time high occurred for each of the portfolios. You should not just use the View() function and tell the TA marking to look through the data themselves to confirm the date you found is the all-time high. Imagine that the dataset had tens of thousands of days. Write some code so that someone reading your code can tell that you have found the month and year of the all-time highest portfolio value without looking through all the data themselves.

2d. (15 points) How correlated is Portfolio 2's daily return with Bitcoin's daily return? How correlated is Portfolio 3's daily return with Bitcoin's daily return? Plot two separate scatterplots for the pairwise daily returns, and then find the correlation coefficients using the cor() function. Round to the nearest hundredths place. For example, 0.02113 would be rounded to 0.02 and -0.39912 would be rounded to -0.40.

2e. (18 points) Your friend is excited about his sudden windfall and is considering holding onto the portfolio he chooses in the hopes that it might eventually gain value over time beyond its current value. However, he is also worried that it might continue to drop in value. He is asking you to find the worst drop from all-time high for each of the portfolios. Define a running measure 'drop from all-time high (ATH)':

$$\text{Drop From All-Time High Prior to Today} = \frac{\text{highest value prior to today} - \text{value on today}}{\text{highest value prior to today}}.$$

For each portfolio, create a numeric vector to capture this measure across the days in the dataset. For example, suppose that you had the following numbers for Portfolio 1 (made-up numbers just to illustrate).

date	value
2023-05-01	1
2023-05-02	2
2023-05-03	3
2023-05-04	2
2023-05-05	1
2023-05-06	2
2023-05-07	4
2023-05-08	5
2023-05-09	3

Then the vector capturing the running drop from all-time high would look like this:

drop.from.ATH
0.0000000
0.0000000
0.0000000
0.3333333
0.6666667
0.3333333
0.0000000
0.0000000
0.4000000

The interpretation is the following. The first three zeros in the vector means that each of the first three days were at their all-time high up until those points in time. The fourth number being 0.3333333 means that the fourth day's value is a 33.33333% drop from the ATH up until that point in time (dropping from 3 to 2). Similarly, the fifth number being 0.6666667 means that the fourth day's value is a 66.66667% drop from the ATH up until that point in time (dropping from 3 to 1). Then the next zeros mean that those values are their current all-time highs again. The last entry in the vector being 0.4 means that the last day's value is a 40% drop from the ATH up until that time (a drop from 5 to 3).

For each portfolio, create a vector capturing the running drop from all-time high measure. What is the highest number for Portfolio 1 (how large of a drop was the worst drop from ATH)? What is the highest number for Portfolio 2? What is the highest number for Portfolio 3? Express your numbers as percentages rounded to two decimals (for example, 0.6666667 in the table above would be written as 66.67 on D2L).

## Instructions

First, read all of the instructions and questions carefully for the entire assignment.

Then for each sub-problem, before writing any code, think about the problem logically. Understand what data is available, so that you can think of ways to use that data to create what you want to create. Write down a structured, logical solution approach in words (in plain English, without code). Your approach should be detailed enough so that someone else would be able to implement your approach in code. **Write your approach as R comments before your solution for that problem or sub-problem.** Note that this is an iterative process. As you are solving the problem, if you realize that your approach was fundamentally missing something, then you should revise your solution approach above.

## Grading Criteria

I will mark the assignment in the following way: I will look at the D2L score of your final D2L HW quiz attempt and take off points if your code does not work. **You must submit your answers to the questions in the D2L quiz for the homework to get credit for the assignment.** Your R file must be able to execute properly and replicate your D2L quiz responses to get credit. If your R code has a bug anywhere (if I try to run a line of code and it doesn't work), then you will lose points. If you don't submit an R file, if you don't submit a HW 1 D2L Quiz attempt, or if your R code and plots do not match your responses, then you will get zero credit.

Submit your R code as **a single R file** to the D2L dropbox folder. Do not submit a different R file for each question of the assignment. In the file name, include the homework number, your first name, and your last name. An example would be 'HW1\_firstName\_lastName.R' or 'HW1\_firstName\_lastName.Rmd' (depending on whether or not you used RMarkdown).

**Caution:** Early on in the semester, students do not follow instructions carefully. If you follow the steps below, you can almost guarantee that you will not lose points due to your R code not running properly on my computer. Carefully making sure your code works fine and does not contain irrelevant chunks of code is good practice, and you want to have formed these habits prior to working together in teams on projects.

- 1) Save your current homework R file, and close RStudio.
- 2) Open your homework R file again.
- 3) Start at Line 1 (the first line in your code). Press Ctrl + Enter (Cmd + Enter for Macs).
- 4) Line by line (one line at a time), keep pressing Ctrl + Enter (Cmd + Enter) and check the outputs. Make sure that your final outputs for each question match your D2L quiz responses for the assignment.
- 5) Remove all extraneous bits or chunks of code that were not needed to produce your output. For example, if you had bits of code that were dead ends and the objects you created were not used in your solution, then delete those lines of code. Students have things like 'Attempt 4' or 'Attempt 10' in their code. Before doing this step, you may want to save a separate copy that contains all your dead ends (in case you accidentally delete something you needed). That's great to have a separate copy of your work (saved as a separate R file) containing the dead ends, but in your final submission, only submit code that is part of your working solution. Otherwise, if you include all prior attempts before your working solution, then the TA might accidentally mark the buggy, dead-end code and flag your work for potential academic misconduct if they see you got the right answer on D2L but your code doesn't work (in which case, I would have to spend time to investigate).
- 6) **Repeat Steps 1 - 5 again (to make sure you didn't remove a needed line of code).**
- 7) Save your file with your name as stated in the instructions, and submit your work to D2L. Your work for the assignment should all be in a single R file, not multiple R files and not a zip file.