

## Final project

The syllabus contains the following information about the project:

Students are required to do a group project **using financial data**. The project will require you to do some computer programming. At the end of the semester, each team is required to **submit a report** on your project as well as **a web app** that you will employ on your own server in the Cloud. The precise deadline will be set later in the semester, and late submissions will receive a zero grade. A guideline for the **report length is 4–8 pages** (excluding references, the app, and computer code). You are also required to submit your computer code. All team members are expected to contribute to all aspects of the **project work (statistical analysis, coding, writing)**. Students will be asked to provide confidential feedback about the work contribution of his/her team members through peer evaluations.

See the syllabus for details on grading.

## 1 Introduction and objectives

You just started a new job at an investment company, and your new boss gives you the following assignment on your very first day: You need to **carry out a statistical analysis of stocks** to be sent to clients. The clients are familiar with basic statistical tools, and they are **interested in original questions** and **insightful analysis**.

old wives tale: 指假想的故事

The final project is **open ended** to **simulate a real-world setting**. The goal is to investigate statistically if you can find evidence for an “old wives’ tale” of your choice. For instance, you could investigate whether January stock returns predict annual returns, see **pink** for more information. You don’t have to use stock data, you may also investigate mutual funds, bonds, etc; in this document I call all of this ‘stock’ data. The questions you formulate and what you discuss in your report are totally up to you! **This document describes what your project is required to address at the very least. You will be graded on how well you execute these requirements, but originality is also part of your grade.**

This is a data-related project, so you’ll be using mostly statistical tools from the second part of this class. For **the choice of the data set**, see the instructions **in hw09-R**. You are responsible for cleaning (if needed). You are free to switch to a different data set any time you like.

## 2 Relation with group assignments

All R group assignments this semester are part of the homework and count towards your homework grade unless specifically marked as “ungraded”. They do not count towards your project grade. You may **use code, figures, etc from the group assignments in your final project**, so you may recycle parts of these assignments. You can grow the code/app over time into the code/app for the final project, or you can create different code/apps each time.

### 3 Report

In your report, make sure you: (1) Describe the data set. (2) Formulate a set of questions you want to investigate, i.e., the goals of your project. (3) Analyze the data and draw conclusions.

You do not want to overwhelm clients with unimportant details, so you have to choose what you want to say. Short is often better, think carefully about the points you want to make. For instance, clients will not be interested in an endless number of figures. Make it count what you write!

I've uploaded an example project report from a related class (see Courseworks under "Files/project"). The instructions were different for that class, but it is the best example that I can share. (This example report doesn't have any figures, so please note the above comment that carefully chosen figures can be very helpful.)

### 4 Data

You need to use log-returns for your stock market data, which you may assume to be drawn from a random sample. You can for instance download this data from Yahoo or Google Finance.

Calculate log-returns for your stocks with R. The log-return is defined as

$$\log \left( \frac{S_t^{\text{close}}}{S_t^{\text{open}}} \right)$$

where  $S_t^{\text{close}}$  and  $S_t^{\text{open}}$  are the close and open prices, respectively, on day  $t$ . (Here log is natural log.) Depending on the question you want to explore, you may alternatively measure time in minutes, hours, days, weeks, etc.

### 5 Code

You need to code in R. Write code as clean as you can, in particular avoid repetition. Document your code.

### 6 Web app

You need to create a web app using Shiny. You can take a tutorial here. You're only limited by your own creativity when it comes to app design. For instance, it's up to you whether you want to use drop-down menus, where to place them, and how they look. Your clients want to be able to play with your web app to make investment decisions. It's up to you to impress them with your web app!

Your app needs to have the following capabilities *at a minimum*:

- Given one stock symbol, your code needs to be able to: (1) Display histograms for your data by stock symbol. (2) Display a normal probability plot to see if the data is approximately normal. (3) Create (approximate) confidence intervals for the means and variances given a confidence level. (4) Perform a regression of the log-return on time.
- Given two stock symbols, your code needs to be able to: (1) Test the equality of the two population means. (2) Perform a regression of one log-return on the other.

All regression output needs to include intercept and slope estimates, a diagram of the data with the least-squares line, a graphical depiction of residuals, and  $R^2$ .

## 7 Awards

I will be making two awards: “Best Project” and “Best App”.

## 8 Cloud

Your web app needs to **run in the Google cloud** and needs to be publicly accessible. You will receive cloud credit of \$150 per team. If you make your app available any other way (or not at all) then it will affect your grade. Your server needs to be online when you submit until your course grade is posted on SSOL. You will receive detailed instructions on how to get your app online through Piazza.

## 9 Submission

You need to submit a report and code in one zip file. The file name of your zip file should be `UNI1_UNI2_UNI3.zip`, where `UNI1`, `UNI2`, and `UNI3` are the UNIs of the team members. The file should contain the following:

1. A file called “`projectApp.html`”, which you can download from Courseworks under “`Files/project`”. Change the URL in that file your own URL, and change the team information (UNIs and names) to your own team’s information.
2. Your report in PDF format.
3. All of the code. It needs to be ready to run on my machine. For instance, **all dependent data files must be included (and in the folder that is referenced in your code), and your code needs to check if the required libraries are installed (and install them if needed).**

The deadline for submission is **December 7, 11:59pm**. You need to submit on Courseworks (not Gradescope). Only one person per team has to submit, and this person should make sure all team members are listed as part of the “R project” group in Courseworks.