**Project Plan**

**Description:**

I have decided to base my project on financial data. I will be calculating the risk of various portfolios of stocks in terms of their Beta. The Beta of a portfolio gives a measure of its overall market risk. Two portfolios consisting of the same stocks can have a different beta depending on the proportion of each stock in the portfolio. I will construct multiple portfolios to show the difference in the beta values. I have chosen this project since I am working in the finance industry and some of the measures like the Beta are used very frequently and in different contexts.

**Data Set:**

I plan to use the historic daily or monthly prices for most of the analysis. The price data is available in Yahoo finance and also from other sources like quandl. I will most probably use Yahoo as the data source. I will be sourcing datasets for multiple stocks and also the S&P 500 index. The data transformation and cleanup will be done prior to analysis as necessary.

**Analysis Component:**

For the analysis, I will first create multiple portfolios of stocks. The portfolios will consist of the Stocks, their Prices and their Quantities. I will then find the Beta for one portfolio at a time. To find the Beta, I will have to fetch the historical prices for each of the stocks in the portfolio. I will also need to fetch the historical prices for an index (S&P 500 in this case) that is representative of the broader market. These prices will be fetched dynamically using a URL that is created at runtime using the stock symbols in the portfolio. After getting the historical prices for the each stock and the index, we calculate the Beta for each individual stock by calculating the covariance of the stock with the index and dividing the result with the variance of the index. To find the Beta of a portfolio of stocks, we need to first find the betas for all individual stocks in the portfolio. Each beta will then multiplied by the percentage of the total portfolio that stock represents to get a weighted beta. Adding all the weighted betas will give the Portfolio’s overall beta.

The data will be collected directly from the web using a URL that will be constructed at runtime using the stock symbols from the Portfolio. The source for the data will be Yahoo finance.

The date column in the data will be converted to a Date column in R.

The statistical analysis will consist of all analysis steps that will be needed to create the final Beta for the entire portfolio. In particular, the following analysis will be performed:

* Calculate the monthly returns for each stock in a portfolio.
* Calculate the monthly returns for the S&P 500 Index.
* Calculate the Covariance of each stock with the S&P 500 Index.
* Calculate the Variance of the S&P 500 Index.
* Calculate the Beta of each Stock in the Portfolio.
* Calculate the Beta of the entire Portfolio.
* Repeat the above steps for each portfolio.

We can get an idea of the risk for a portfolio consisting of a particular set of symbols by creating multiple portfolios from that set with different quantities for each stock and then finding the Beta for each Portfolio. In this way we can find a combination that is within our risk tolerance.

**Programming Tools/Technologies:**

The main programming tool will be R. I will be using various packages within R. I will use R markdown for the output. I will be using a graphics package to create the graphical output. I will be mainly using ggplot2, and as an extra challenge, I will use the lattice graphics package.

**Input:**

The input to the program will be through csv files and the web. The URL for getting the datasets will be dynamically generated from the tickers in the portfolio. I might use intermediate temp files to store temporary data if required.

**Visualization:**

I will be generating some bar graphs to show the individual betas of the stocks within a portfolio and of betas across portfolios. I will also be generating graphs to show the returns of stocks versus the S&P 500 index.