MATH 70 MIDTERM

1. (10 points). Random variables U, V, and Z are uncorrelated, and $X = U + aZ$, $Y = V + aZ$, where a is a constant. (a) Prove that the Pearson correlation coefficient between X and Y is $\rho_{XY} = a^2/(1+a^2)$. (b) Using Definition 2 from Week5.pdf, prove that the partial correlation coefficient between X and Y equals zero, $\rho_{XY|Z} = 0$, by the 3×3 matrix inverse.

2. (15 points). (a) Use stock prices from `Most active stocks prices_Feb_01.2023.csv` to compute and display the correlation heatmap using 12 breaks from -0.1 to 1 with the respective colors from `Rcolor.pdf`. Save the graphs in the `pdf` format (ensure all stock symbols and labels are seen by choosing the appropriate graph size, e.g., height=width=50 inches). (b) Report pairs of stocks with r > 0.9. Print out the symbols and the rounded correlation coefficient.

3. (10 points). The same as above, but create a partial correlation matrix heatmap (use the same breaks). Summarize the findings by comparison of ordinary and partial correlations.

4. (15 points). (a) Regress TESLA returns on returns of all other companies and extract only the companies with the p-value regression coefficient < 0.001 to obtain a parsimonious regression (print out `summary`). (b) Look online to understand why these companies have the most significant regression coefficients (write 1-2 sentences as an explanation). (c) Why the coefficient of determination for a parsimonious regression is smaller than for the full model (provide a proof)? (d) Use the 3D plot to display the point returns with TESLA on the Z-axis. Use `par(mfrow=c(1,1),mar=c(1,1,3,1),cex.lab=1.5, cex.main=1.5)` and `ticktype="detailed"` option in `persp`.

5. (10 extra points). Create a running window movie by estimating the correlation matrix on a three-month period/window starting from the first date (use `png` to save the frames). Display the first and the last period of the running window using `mtext(side=3)`. Don't display anything but the image itself. Summarize your finding by running the movie. [Hint: Borrow the animation technique from the `cdf.dyn` code.]

The pre-processing steps for stock prices:
`d=read.csv("Most active stocks prices_Feb_01.2023.csv)`
You must insert the name of the folder where the file is stored before `Most`
This file has 91 columns and 524 rows. The dates can be extracted as `as.character(d[,1])`
To get stock prices for 90 companies, issue `X=as.matrix(d[,2:91])`
Company symbols can be extracted as `nm=names(d)[2:91]`
Full names can be found in the file `mostactive.csv`

Include the graphs, the R codes and their outputs. Don't include the gif file from 5 (will be to big), but provide the R code.

Due to large size submit as a single zip file.