IE 423 Quality Engineering
Project Part 2, due December 4th, 2019

Instructions: Please solve the following exercises using R (http://www.r-project.org/) or Python (https://www.python.org/) as a group of at most 5 members. You are expected to use GitHub Classroom and present your work as an html file (i.e. web page) on your progress journals. There are alternative ways to generate an html page for you work:

- A Jupyter Notebook including your codes and comments. This works for R and Python, to enable using R scripts in notebooks, please check:
  - https://docs.anaconda.com/anaconda/navigator/tutorials/r-lang/
  - https://medium.com/@kyleake/how-to-install-r-in-jupyter-with-irkernel-in-3-steps-917519326e41

  Things are little easier if you install Anaconda (https://www.anaconda.com/). Please export your work to an html file. Please provide your *. ipynb file in your repository and a link to this file in your html report will help us a lot.

- A Markdown html document. This can be created using RMarkdown for R and Python-Markdown for Python

Note that html pages are just to describe how you approach to the exercises in the homework. They should include your codes. You are also required to provide your R/Python codes separately in the repository so that anybody can run it with minimal change in the code. This can be presented as the script file itself or your notebook file (the one with *.ipynb file extension).

The last and the most important thing to mention is that academic integrity is expected! Do not share your code (except the one in your progress journals). You are always free to discuss about tasks but your work must be implemented by yourself. As a fundamental principle for any educational institution, academic integrity is highly valued and seriously regarded at Boğaziçi University.

TASKS
Part 1 *(50 points)*

You will be working on the images uploaded to Moodle for this task. Use the image having the name "group_id.jpg". In other words, each group will have their own image. If you decided not to move on with the same group and formed a new group, please use the image "group_others.jpg". Before working on the tasks, please transform your color image to a greyscale one using an image editor. The other option is to use some package/module to perform the transformation. You will be working on greyscale images for the following tasks:

In the first part of your project, you have performed quality control operations based on the pixel value distribution of the whole image and within the patches of the image. For this task, you are expected to find out out-of-control pixel values considering the each row and each column of the matrix. Implicitly, we are imposing an assumption that the distribution of the pixels should be the same horizontally and vertically.

a) For each row of your image, construct an appropriate chart for monitoring the mean and variance. Find the pixels that are out of control. After finding those pixels, change the value of these pixels to zero (i.e. black color). Display the new image and original image in a plot. What are your observations? Comment on your findings.
b) Perform the same operations on the columns and comment on your findings.

Part 2 ***(50 points)***

Regular control charts built as in Part 1 assume that pixel values are independent of each other (i.e. there is a problem of autocorrelation). This is problematic especially for the type of image you are working on. When we work on an image which has a texture, the pixel values are somehow related (due to the definition of a texture). To avoid from this problem, one can use the following strategy:

- Suppose we extract the patches of images (i.e. windows of certain size). When local structures are important, performing image operations on the patches might be important.
- Assume that your window size is 51x51 and you would like to build a linear regression model to predict the center pixel using the information of other pixels. In other words, the remaining pixels in the patch will be predictors of the center pixel. One can claim that the predictors are also related and using linear regression may not be a good idea because of the multicollinearity. This is a valid claim but we will move on with this strategy to keep things easier for now.
- Using your regression model, you can come up with the predicted values of the center pixels and associated residuals. Check if residuals are looking like Normal distribution.
- Build an appropriate chart on the residuals and find the residuals that are out of control. After finding those residuals (corresponding to certain pixels), change the value of these pixels to zero (i.e. black color). Display the new image and original image in a plot. What are your observations? Comment on your findings.

Can you devise a better way to handle the problem of autocorrelation? Propose alternative approaches to solve this problem. Note that you are not expected to implement your proposals.