Assignment 2

In this assignment, you will analyze real data that I logged on #thedress phenomenon (http://www.slate.com/articles/health_and_science/science/2015/02/what_color_is_this_dress_a_scientist_explains_visual_ambiguity_and_color.html)
and no, studying the perception of the dress is *not* frivolous, far from it:
(http://www.creativitypost.com/science/why_dressgate_matters)
I even got a paper out of it: http://jov.arvojournals.org/article.aspx?articleid=2617976

This data is contained in the workspace "Assignment2Data.mat". All tasks below are to be executed within a single script. This is the .m file that you should upload. You can assume that the person reviewing your code will run the script from the same folder that contains the data file.

Specification of the tasks that the script needs to be able to do (in bold):

BASIC

1) **Load the workspace with the DATA matrix**. Note that the rows represent individual participants. Column 1 represents the self-reported color percept when they first saw the dress. Column 2 represents the self-reported circadian type (an "owl" likes to sleep in and stay up, feels best at night, a "lark" gets up early, goes to bed early and feels best in the morning). Also note that the data is represented in "cells" that contain strings. The content of cells can be accessed by curly braces, so for instance DATA{1,1} will reach into the cell and access the contents of the first row and column.

2) Data parsing is a routine step in almost every data analysis project. It involves the processing of raw data into a format that can be used in further analysis. This assignment is no exception. **Create a new matrix, numericalDATA, of the same size as DATA. It should represent the original data, but as numbers. Go through all data, parse it and assign numerical labels that you store in the numericalDATA matrix**. For color, represent "White/Gold" as 1, "Blue/Black" as 2, "White/Black" as 3 and "Blue/Gold" as 4. Rapid switching should be 5 and anything else or missing data 6. For circadian type, "strong owl" is 1, "owl" is 2, "lark" is 3, "strong lark" is 4 and anything else or missing data is 5. Note that in order to parse the strings, you will need to use the *strcmp* function. The *strcmp* function compares a target string to a test string and returns 1 if the comparison is a match, 0 otherwise. To do this part of the assignment, I recommend to go through all of the data in a loop, and check in each case whether the strings correspond to one of the categories above (e.g. "White/Gold"), then assign the appropriate number (as mentioned above) in the corresponding location of the numericalDATA matrix.

3) **Make a figure that shows a histogram of the self-reported color responses.**

4) **Make a figure that shows a histogram of the self-reported circadian type**

AMBITIOUS

1) Do a chi-square test to see if there is an association between self-reported color responses and self-reported circadian type or whether these variables are independent. What is the p-value? What is the effect size?
2) Column 3 represents the self-reported response of participants as to where they live. Parse those data into meaningful categories and assign numbers to them. Be creative. For instance, we don't have many respondents from African countries, so all of Africa might be a single place category. The same is true for the middle east and Oceania. We have more responses from the US and Europe, so we can be finer grained there. Are there place-level effects on dress perception? Make histograms and do a chi-square analysis to see if there is an association.
3) Column 4 represents the self-reported ethnicity of the participants. Parse and group them into a few ethnic categories that are large enough for a meaningful statistical analysis. Are there effects of ethnicity on dress perception? Make histograms and do a chi-square test.
4) Column 5 represents the self-reported favorite color of the participants. Parse and group them into a few meaningful categories (primary colors?) and determine – with histograms and chi-square tests whether dress perception is related to the favorite color of the participants.
5) Column 6 represents self-reported visual disorders of the participants. Parse and group them into a few meaningful categories, e.g. "Cataracts" or "Myopia" and see whether dress perception is at all related to visual disorders. Make histograms, do chi square tests.

REALLY AMBITIOUS

1) Column 7 represents the time the participants did the study and columns 8 and 9 represent the latitude and longitude of where they logged in from, respectively. Do a manipulation check: Are self-reported larks more likely to do the study in the morning than at night (relative to owls)? Are self-reported owls more likely to do the study at night than in the morning (relative to larks)? Note that you'll need to take latitude and longitude into account to determine local time.
2) Make a heat map of dress perception around the world, using both latitude and longitude. Color the dots appropriately (corresponding on the mean response at that latitude/longitude). Take local response density (high in the US and Europe, low elsewhere) into account. Be creative.
3) Is there an effect of latitude or longitude on dress perception?