Week 14 in CS 234

What is due on Tuesday, 12/05/17?

1. **Project 2 Google Searches - Wrapping up**
    a. Complete your work, especially try to build a simple classifier, so that you see how classification works.
    b. Put all the notebooks and data files you created for the project in the "dav/drop/project02" folder.
    c. Update your blog to contain the links to the notebook HTML files, as well as to the final report of your group.
    d. Make sure your final report addresses all parts of the data science cycle, also showing some of your visualizations.

2. **Notebooks** that were brought to class in Week 13
    a. Complete the Linear Regression notebook, put it on the dav/drop/week13.
    b. Complete the Clustering notebook, put it on the dav/drop/week13

What is due on Friday, 10/27/17?

1. **Project 3: Digital Natives**
    a. Make sure that your HTML public_folder has now a new subfolder for the digital project: `public_html/cs234/digital` and you have in it the `blog.html` file (this should have been created in early November)
    b. If you wrote a paragraph to share with me on Nov 7, please have that as your first entry in the `blog.html` page.
    c. Write **a new entry blog** about our class discussion of Nov 28 about the digital natives reading. Specifically, using what you learned about linear regression in the lecture about regression and the notebook on regression, describe in your own words the meaning of the models that the author created, especially the parameters (coefficients) of the models (what do the negative and positive values mean).
    d. Write **a new entry blog** about your current plan for the final, independent project. Now that you have learned all the steps of the data science cycle and knowing that our topic for the project is "digital natives and ways to characterize them through data", **write a short plan** of what you will do from now until the deadline (**Dec 21st, 4PM**) to complete your project. I will read them over the weekend of Dec 9-10 and give you feedback.

2. **Seaborn Visualization [Improve your viz skills]**
    Seaborn allows you to be faster in making certain statistical visualizations. We saw examples in the notebooks of Week 13, now it's time for you to try to create visualizations of your own.
    a. Create box-plots for your history data [see below if you don't have a history]. Read here how to create box-plots. In the vertical axis there will be the days of the week. The boxplot will show the boxplot for the 8 datapoints you'll have for each day of week (since your

history contains 8 weeks of data). Write a notebook to show how you created the data and then built the visualization. Upload the notebook in the dav/drop/week14 and blog about it in your blog.

b. Create a dot plot for your most visited websites. Read here how to create dot-plots. Your variables will be the top 5 most visited websites in your history, for example: google.com, mail.wellesley.edu, etc. In the y-axis you have the days of the week again. The plot will show how many total visits there are on each website on a particular weekday (all your Mondays in the data, all your Tuesdays, etc.). [Use the same notebook as task a. above].

If you don't have your browser history: work together with one of your project peers who has a history and submit a joint notebook (with both your names on it).

As usually, all the tasks are to advance your learning. If you don't have time to complete them, work on them when you have the time, or you can decide to not complete them at all.