**Schedule/timeline update**

**Following the timeline you proposed in your project proposal, please give us an update regarding finished activities, possible modifications and/or additional activities.**

Our team downloaded all the databases and scrap the list from the web page of Times Higher Education College Rankings 2017; as well the list from QS World University Ranking. We use lapply function combined with a vector list to upload the databases. Currently we have 12 cvs. files containing data and 12 xlsx. the variables descriptions; each file contain different categories of analysis.

In the other hand, as part of the tidying, from all the databases specific columns were eliminated because it did not contain valuable information. Also, using a loop, labels were assigned indicating the description of all the columns in all the databases. For this step it was necessary to use the files containing the data and the files with the description of all the columns. Description labels in columns will allows us to identify in a practical way the data we need.

In addition to the expected work, the data regarding the names of the universities required an extensive cleaning. This procedure was necessary to link the names of the universities on the list of the Times Higher Education College Rankings(THE) and the names that were showed in the databases of the Integrated Postsecondary Education Data System (IPEDS) . For this new task, we applied tools learned doing text mining in our semester project.

Modifications in tidying will be done. For some files the tidying will be done before plotting and for other the data will be transformed in the code of the plot depending of the characteristics that some graphs require.

**Task accomplished:**

Loading and scrapping data

Cleaning variable corresponding to University names to link files

Initial phase of data visualization and analysis

**Preliminary results**

**Briefly summarise your current findings. Is the project going as expected? Will it be possible to achieve the goals you set in the proposal?**

Top universities agree mostly requiring admission test and school records, however it was interesting to find that having a recommendation is as almost as required as the other most usual requirements. This observation is clearer when the color in the graph exposes the difference that the universities have between require and recommended (suggested). Furthermore, the rate of acceptance, even at the different level in the top 50, sports more robustness getting closer to the higher ranks.

As suggested, now the project will focus on answer and analyze specific questions to have a better understanding of these variables. Cleaning the data took more time than expected, nevertheless the time invested in this first part pays off. We expect to achieve the goal we set in our proposal, now with a more specific evaluation.