**Data Acquisition Strategy [finalize data by 4/1]**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Strategy | Responsible | Note |
| 1 | Play with API | Alexis/Ana |  |
| 2 | Contact LinkedIn | Michael |  |
| 3 | Scrape (look at Kimono to start) | Michael | https://www.kimonolabs.com/ |
| 4 | Manual download |  | Last resort: Loss of data depth |

**Timeline**

Apr 1st: Develop API and download data

Apr 4th: Prototype design template

Apr 10th: Prototype due

Apr 10th: Finalize data exploration

Apr 17th: Finalize design template

Apr 24th: First draft

May 1st: Final deadline

**Milestones**

* Thursday, March 13: Project proposal due (part of Homework 3)
* Thursday, April 10: Functional project prototype due
* Week of April 14: Project review with the TFs
* Thursday, May 1: Projects due (including screencast)
* Thursday, May 8: Best project presentations and prizes

**Project Proposal**

You start your project by forming your groups and letting us know what topic you are interested in exploring by submitting [this project data form](https://docs.google.com/forms/d/1L7bb7R9J3yA8NSnz7tpOlOX8TFNAv9AMZCQsBQqTkMk). Each team will only need to submit one form. In addition to the form, you will create a proposal document, addressing the following points. Use these points as headers in your document.

* **Background and Motivation**. Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

Every student who, like us, is in their last year at Harvard is faced with the question of what to do next. One of the benchmarks of the careers that Harvard graduates can embark on can be found in what their peers have actually done before them. We thought about collecting this data online in order to get a better picture of what professional directions Harvard student takes, and how this compares to other schools

* **Project Objectives**. Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

What do Harvard graduates do after they leave Harvard? Where do they work? In What kind of jobs? In Which companies and which geographies?

How does that compare to peer schools such as Stanford and Yale?

* **Data**. From where and how are you collecting your data? If appropriate, provide a link to your data sources.

We will use data aggregated from networking platform LinkedIn, which has in recent years achieved sufficient scale to make the data credibly representative of the Harvard graduate population (in particular the recent one). The data is displayed in a very basic form [here](http://www.linkedin.com/college/alumni?eduSchool=18483&wideRange=true&trk=edu-cp-com-CC-title#pt.psiund=true&pt.pses=%2218483%22&pt.schn=%22Harvard%20University%2C%20John%20F.%20Kennedy%20School%20of%20Government%22&pt.psess=%22%22&pt.psest=1900&pt.psese=%22%22&pt.psee=2021&pt.pseg=2021&pt.yst=%22a%22&).

The data will be collected from the Linkedin API which allows up to 250 calls/day for unregistered applications.

* **Data Processing**. Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

We haven’t yet looked into the details of the API results, but we expect data processing to be manageable. We will report counts and proportions of students at various levels of granularity and sliced according to various dimensions.

* **Visualization**. How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.

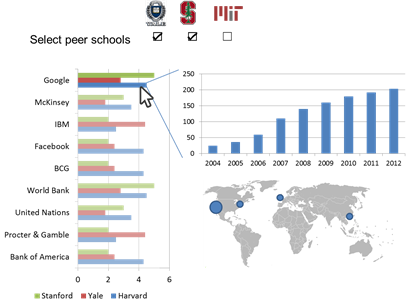
The following ideas will be explored:

- Map of countries and cities where alumni work

- Bar chart of most common employers, with one bar color per school

- Bar chart of most common jobs, with one bar color per school

- Filter interactions between these dimensions (click on google to see what graduates are doing at Google, click on Stanford to only see Stanford students, etc.)



* **Must-Have Features**. These are features without which you would consider your project to be a failure.

Comparison of sector composition between the different schools

Comparison of employer composition between the schools

Peer school selector

* **Optional Features**. Those features which you consider would be nice to have, but not critical.

Peer school selector (allows to choose up to three schools to compare Harvard to)

Use [Glassdoor’s](http://www.glassdoor.com) API to merge information of average company salaries.

* **Project Schedule**. Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

Apr 3rd: Develop API and download data

Apr 10th: Finalize data exploration

Apr 17th: Finalize design template

Apr 24th: First draft

May 1st: Final deadline

You will submit this proposal as part of Homework 3. Based on your proposals we will assign a TF to your team who will guide you through the rest of the project. You will schedule a project review meeting with your TF during regular lecture times of the week marked in the schedule. Make sure all of your team members are present at the meeting. Online students can schedule a Skype or google hangout meeting with their TF.