20 Years of Educated Lives

Anthony Chua, Máté Hekfusz, Toby Le

1. Introduction and Justification

When considering what factors may have influenced an individual's rise to notability, one that is always worth considering is their education, namely their higher education. From the myriad of American presidents who have graduated from the law schools of Harvard and Yale[1], to the scores of musicians, artists, and performers who were trained at Juilliard[2], to the countless renowned individuals who have walked the hallowed halls of Oxbridge[3], there has always been some association between certain universities and the production of individuals, who in this investigation, will be referred to as 'notable'. Moreover, as the literature on the subject has so consistently stressed the importance of education, particularly higher education, thoroughly emphasizing its role in shaping an individual's life trajectory[4], this investigation contends that the full extent of this relationship and its potential connotations would be a worthwhile endeavor to explore. Still, as the university is a broad entity that constitutes many working parts, this investigation posits that one factor is of paramount importance in evaluations of its contribution to individual nobility and that is namely: its prestige. Thus, to explore this idea, this investigation will operate around the central question of: **Does attending a more notable institute of higher education have an effect on an individual's potential to be notable?**

Defined as admiration for something derived from its perceived reputation[5], the prestige of a university has been long associated with the quality of education it offers and resultantly the caliber of students it produces. As the connotation of a prestigious university has become increasingly loaded, with names like Oxbridge and the Ivy League becoming synonymous with the production of elite graduates, employers in all fields[6] have begun to favor candidates who have graduated from institutions with greater levels of prestige, that is more 'notable' ones.

[1] Bolluyt
[2] Wakin (2004)
[3] University of Oxford
[4] Chicago Tribune
[5] Cambridge Dictionary
[6] Mihut (2015)

Consequently, as students have flocked towards notability in order to increase their potential employment in their respective desired fields – even going as far to pay exorbitant fees to guarantee their attendance[7] - these very universities have themselves begun to react by actively skewing their focus towards obtaining greater levels of prestige by insulating themselves from potentially mutually beneficial cooperation[8] rather than strictly pursuing the ideals of academic excellence. Resultantly, whilst the graduates of these institutions may not have the objectively best possible education, they will have a much greater potential to advance in their fields and make important connections that may contribute to their eventual notability, thanks to the external validation received by their alma mater.

It is in consideration of this context, that this investigation seeks to make its independent variable the placement of a notable individual on a hierarchical ranking, namely the Times Higher Education rankings[9]. For a position on this list, is a clear and as close to standardized as possible indicator that the university possesses a significant degree of renown and recognition, something that is more likely to influence its graduates' connections and association, further factors that might propel them into notability. Whilst this investigation recognizes that the rankings are meant to signify the position of the university in 2020, it argues that this is still a valid measure for the prestige of an institution. As reputation is a factor that is advanced and developed over time, any short term or immediate shortcomings or negative incidents are less likely to have a significant impact over the universities' overall placing, regardless of the year – especially when one considers that many of the universities that make this list are several hundred years old. Therefore, it is more than likely that the overwhelming majority of universities present on the current rankings were also present on the rankings during the college years of our specific sample size: notable individuals born between 1940 and 1960.

To fully justify why this investigation has chosen to investigate this relation in its chosen years, it must contextualize the idea that the University in itself is a concept that is both simultaneously younger and older than one might presume. Whilst it has arguably existed since the 11th century when Oxford was the first academic institution to receive the moniker of 'universitas'[10] in 1242, the modern university, especially the research university based on 'liberal ideals', and the university that was not only interested in conveying knowledge but generating new knowledge in all of the academic disciplines, has only emerged in the 19th century[11].

[7] The Atlantic (2018)
[8] Blackmore (2016)
[9] Times Higher Education (2019)
[10] University of Oxford
[11] Psychology Today (2019)

Consequently, whilst it is unquestionable that universities may have contributed to the notability of the countless thousands of students who attended them in the first several centuries of their existence, it would not only be inefficient but methodologically questionable to evaluate the impact of institutions that are entirely different from the ones that we consider to be a part of the contemporary tradition of the university.

Still, whilst the majority of changes in the university as an academic space did occur before the chosen period of this investigation, the study of the attendance at an institute of higher education by a notable individual born in the period of 1940-1960 is particularly significant. This is because they would have likely pursued at least one degree in their higher education in the period from 1960-1980, a time where the university became a much more socially open space that was finally more welcoming and therefore more thoroughly attended by individuals of more diverse backgrounds, national, economic, and ethnic alike[12]. This among other factors, such as increased economic incentives and employer demand, would also lead to a notable increase in university attendance across the world[13], something that provides this investigation with the dual benefits of a tight focus and significant sample size, both of which will contribute to making the results and conclusions more theoretically accurate. Thus, as the time, scope, and subject of this investigation have all been established and justified, it will move on to exploring the methodology used in the investigation and how it will contribute to the analysis of the relationship in question.

[12] Kerr (1991)
[13] Our World in Data

2. Methods

The main dataset, part of a larger effort to document notable people, was generously provided by Professor Etienne Wasmer, one of the database's creators, for research use. It contained 22102 entries of notable people born between 1940-1960 from all countries of the world. This dataset was run through our parser, a custom Python script, which queried Wikidata for information about the alma mater of these twenty-two thousand notable people. After roughly 11 hours of parsing, we acquired a database of 9367 people with one or more higher education institution listed. Though a comprehensive source, Wikidata is not exhaustive, and there are no doubt many notables whose higher education was not listed in their database. To minimize guesswork, we have excluded these people from our analysis.

The other integral source for our research was the freely available *Times Higher Education 2020* university rankings. As we were not allowed to web-scrape the data directly, we used a separate parser to turn it from raw text into a usable database of 1396 universities. Our cross-reference with the main dataset, detailed below, yielded 6021 people who have attended at least one of the universities in the rankings.

One of our key datasets was created by cross-referencing these two sources. We scanned the alma mater database for people who have attended any of the institutions listed in the THE rankings, acquiring various statistics about the universities, including their number of notable graduates in our examined time period and the average notability of these people. There were several difficulties in the creation of this unified database, mainly based on data inconsistency: the university names in the Times Higher Education rankings did not always match the parsed names from Wikidata, for example. Sometimes, the same university was stored in Wikidata, just under a different or an older name, which our initial search did not find. We also had to account for the fact that many universities had several colleges (the University of Oxford being one notable example with its 30-plus college network): to get accurate numbers, we had to manually build lists of colleges and run searches through them. Due to time and manpower constraints, we were only able to do this for the top 30 universities in the THE rankings.

The parsers and miscellaneous scripts (.py files and Jupyter notebooks) and data files (.csv and text format) are all included in our submission. Through them, our results are fully reproducible.

University ranking system

Expanding on the introduction's comments on the Times Higher Education (THE) World University Rankings, this section will provide further details regarding the justification of its use. A publication of worldwide university rankings by THE magazine, it works in cooperation with analytics company Elsevier in order to rank research-intensive universities through both qualitative (e.g.: research survey) and quantitative (e.g.: citation count) measurements across five areas: Teaching, Research, Citations, International Outlook, and Industry Income. (A breakdown for each area is in Figure 1 below.)

This ranking is widely observed and is well-received for relying more heavily on non-subjective criteria. However, it is also criticized for the weight given to citations: Non-English journals and papers are less commonly found because English is the international academic language, so institutions with research published in such journals tend to be less cited and thus gain less research influence; and is emphasis on books as sources for citation, which the social science and humanities are more likely to publish in than the STEM fields, are less covered by citation records, biasing the citation criteria toward some disciplines over others. Since universities ranking beyond 200 are grouped into different ranking blocks rather in individual rankings[14], only universities ranked 1st through 200th were considered for the analysis.

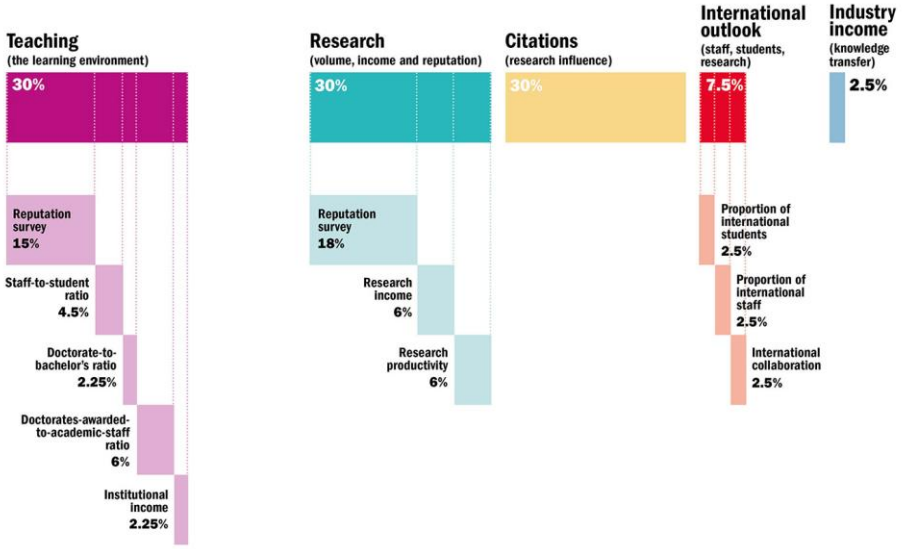


Figure 1. Times Higher Education (THE) World University Rankings 2020 criteria breakdown.

[14] For example, universities that would be ranked 201st to 250th are lumped under the single ranking category 201st-250th, rather than being ranked 201st, 202nd, and so on.

Notability ranking system

A person is considered notable if a record of that person exists on the Wikidata database. The ranking of a person's notability is calculated based on: the number of Wikipedia translations; the sum of the word count of up to seven language editions; number of references in the aforementioned pages; and number of visits to each page by different users.

The reason why Wikidata serves as a more holistic source for notability is because Wikidata entries can exist for figures who are not notable enough to have more than Wikipedia page, or even figures who do not have pages at all. This is particularly conducive to the purposes of this investigation due to its focus on analyzing the most notable individuals from every country. As some countries have significantly low populations, it is highly unlikely that even the most famous figures might be featured on Wikipedia no less the top 100, however, it is far more likely that they will have a Wikidata entry, something that can be considered in the data set and that can be subsequently analyzed.

3. Results and Discussion

Before delving into the relationship between university ranking and graduate notability, let's first tread through basic information about the notable persons datasets. Of the 22102 notables born during 1940-1960, 9361 of them are listed as having attended a higher education institute. Thus, the examination of notables is narrowed down to this educated subset. From this, the breakdowns of education-listed notables from the top 100 notables of each country born during 1940-1960 can be made by gender (Figure 2), primary occupation (Figure 3), and geographic region of attachment (Figure 4).

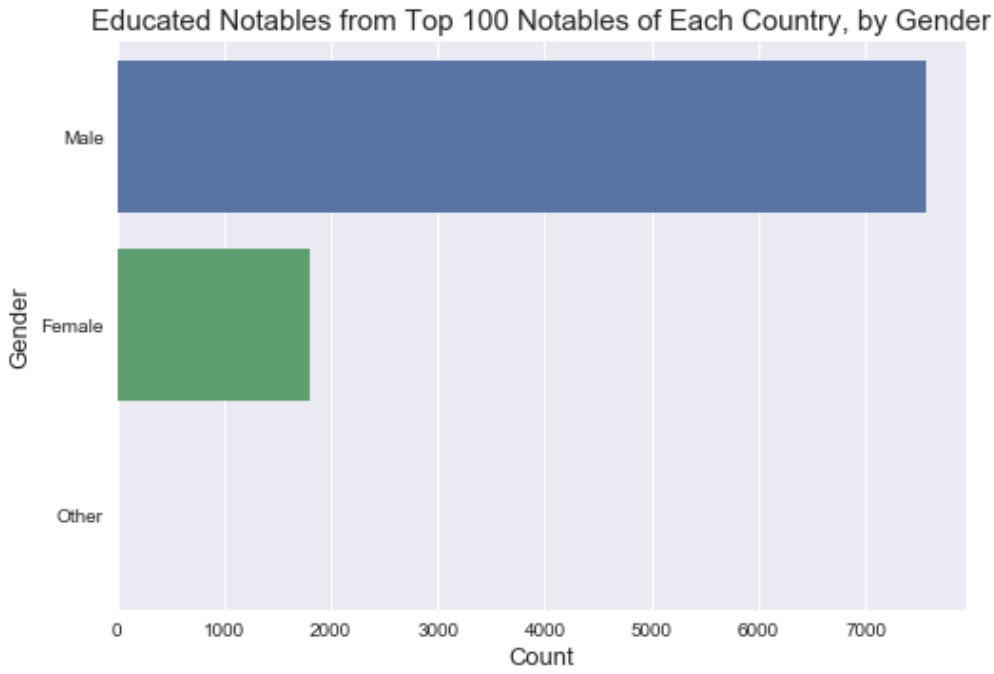


Figure 2. Gender of the educated notables out of the top 100 notables of each country born during 1940-1960. 7561 are listed as Male (81%), 1798 as Female (19%), and 2 as Other (~0%).

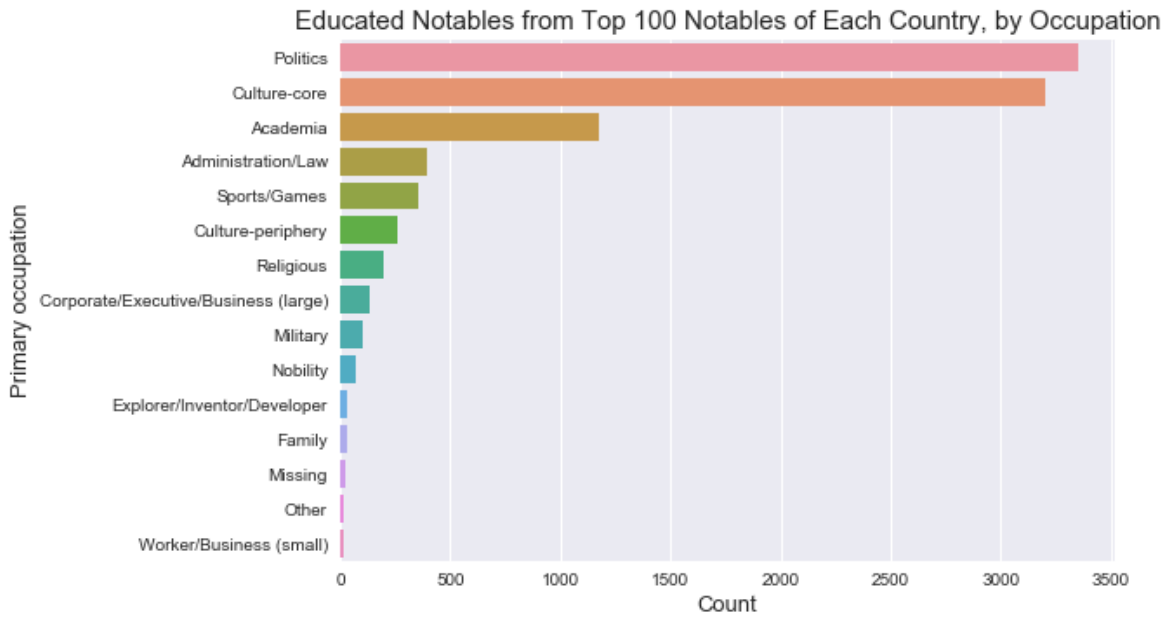


Figure 3. Primary occupation of the educated notables out of the top 100 notables of each country born during 1940-1960.

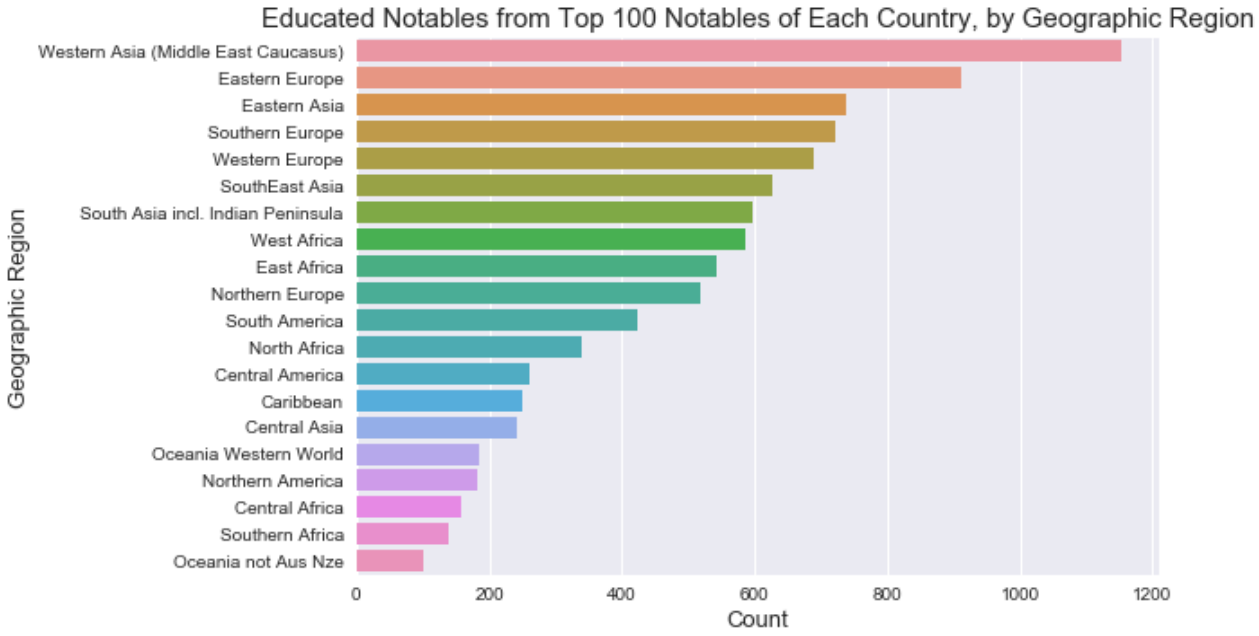


Figure 4. Primary geographic region of attachment of the educated notables out of the top 100 notables of each country born during 1940-1960.

Figures 2 through 4 highlight general trends of highly ranked educated notables born during the 1940s and 1950s. As seen from previous gender trends in class, men outnumber women by 4-to-1. The majority of the notables were known for their roles in Politics, Culture-core, and Academia, trends which when seen through the Cold War context of this period - when the culture war, space race, and political turmoil were all at their peaks - makes sense. Geographically, the popularity of Western and Eastern Asia and much of Europe are indicative of broader trends in both the economic development and political histories of countries in these regions.

Europeans have always had immediate access to universities, thus it is unsurprising that their number of educated notables is one of the highest, as the populations of these countries were on average more capable of obtaining this level of education than any other. As for East Asia and West Asia (specified in this dataset to be the Middle East) it is also expected that the attendance of notables from these countries are high, especially in terms of the culture and political industries. The period from 1960s-1980s in which these individuals would have attended these universities coincided with the oil boom and rapid economic modernization that swept over respectively the Arab World and the East Asian economies. As a result, more individuals, particularly those belonging to more astute families, could afford to pursue an education abroad - one that would also bolster their opportunities for prestige if they chose to return home and seek employment in sectors such as politics. Similarly, as this was a period where the United States essentially held a cultural monopoly, it is only natural that writers, actors, singers, and other artists who are trained in the American tradition abroad were more likely to be well perceived and marked as notable upon returning to their native countries.

Moving on from notable persons for a moment, let’s briefly examine the top 200 universities in the THE2020, with information about their locations (Figure 5) and ages (Figures 6).

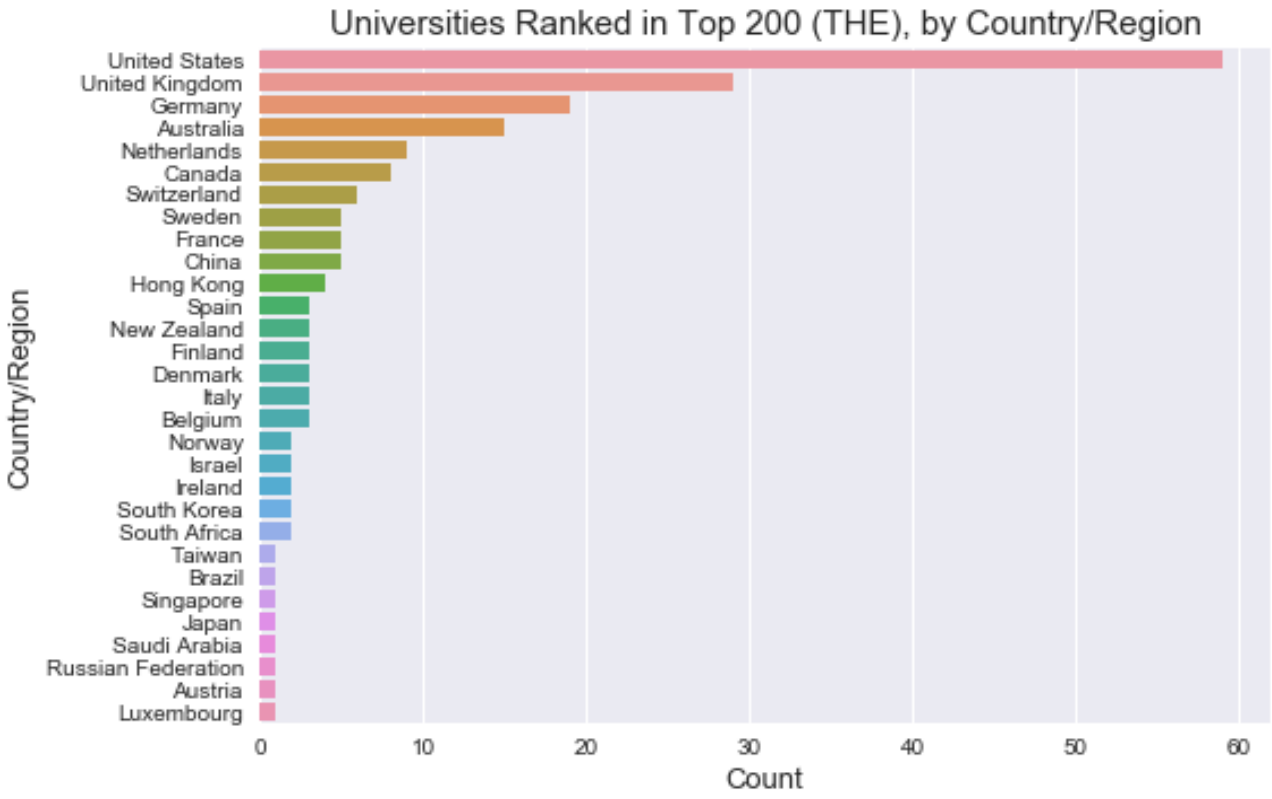


Figure 5. Countries with at least one university in the top 200 of THE2020.

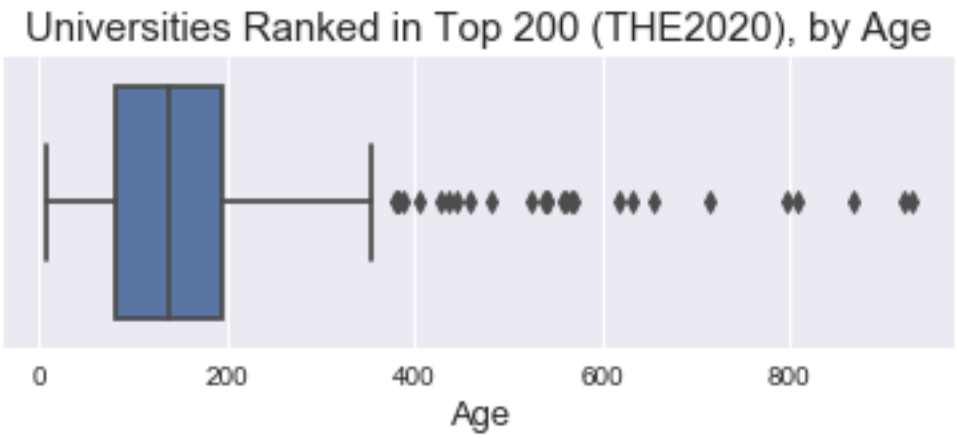


Figure 6. Boxplot of ages of universities in the top 200 of THE2020.

From Figure 5, it seems that the developed world, especially Western nations or those that have adopted Western ideals, has a monopoly on top-ranking universities in both the number of countries that host them, and the number of universities hosted. Figure 6 shows that nearly current universities are no more than 400 years old,

and most are under 200 years old, highlighting that the popularity of the university is a relatively recent trend in the context of human history. Again, this serves to reinforce the research provided in the introduction, as whilst universities have existed from almost a millennium, these were essentially limited to the highest fringes of aristocracy until the 19th century, when European scholars proposed the new model of the university that made it more possible and pragmatic for states to introduce and sponsor a greater number of these institutions.

With this data overview, let's examine whether there is a correlation between the ranking of a university and the notability of its graduates. For a given university, the natural log was taken of the median ranking of the notables who graduated from that university. This value was plotted against the rank of that university, and once repeated for all universities in the THE2020, a linear regression was made (Figure 7a, with detailed regression results in Figure 7b).

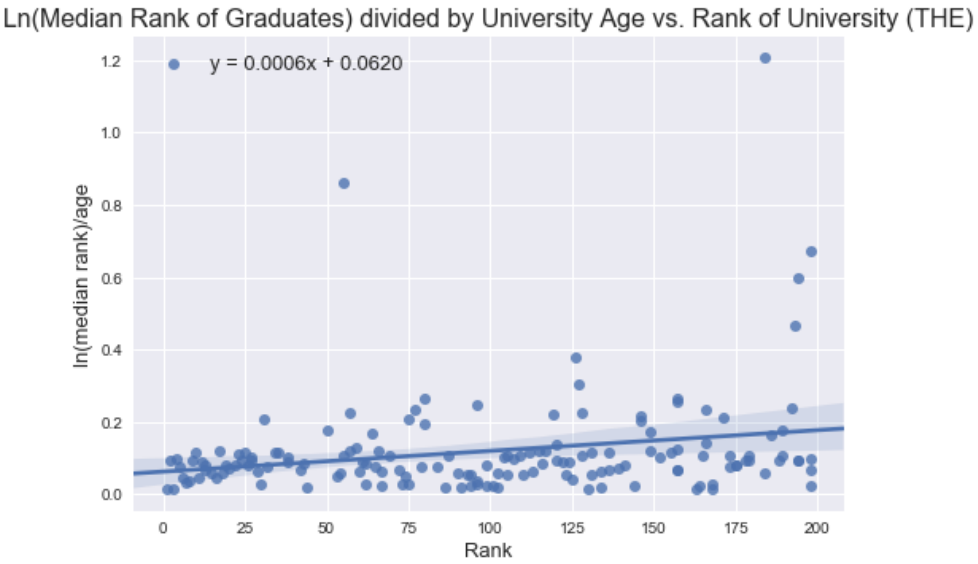
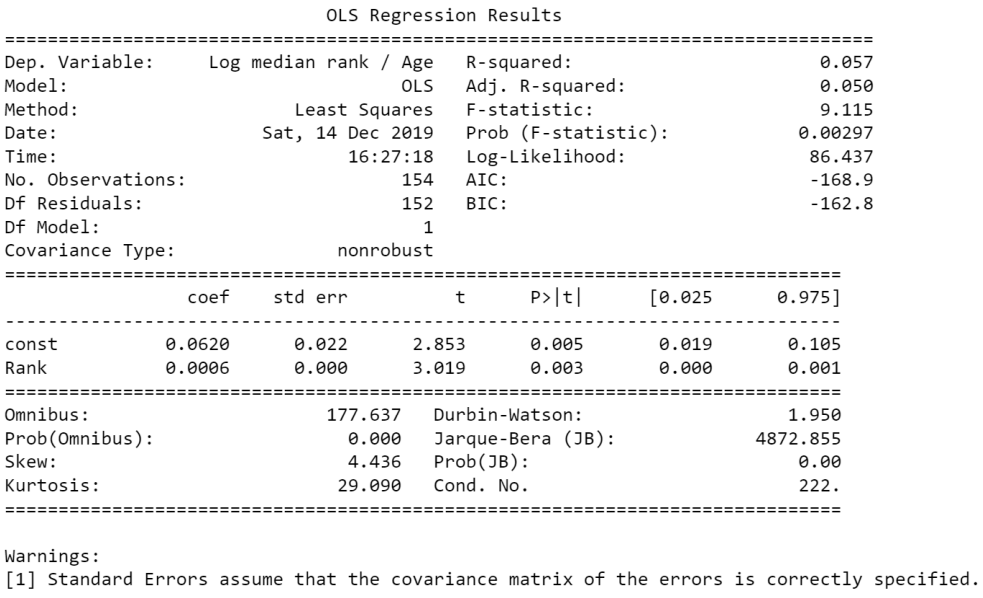


Figure 7a. Linear regression of log(median rank of notables who attended a university) ÷ age of university vs. rank of university. R-squared = 0.057 (see Figure 7b).



OLS Regression Results

Dep. Variable:	Log median rank / Age	R-squared:	0.057
Model:	OLS	Adj. R-squared:	0.050
Method:	Least Squares	F-statistic:	9.115
Date:	Sat, 14 Dec 2019	Prob (F-statistic):	0.00297
Time:	16:27:18	Log-Likelihood:	86.437
No. Observations:	154	AIC:	-168.9
Df Residuals:	152	BIC:	-162.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0620	0.022	2.853	0.005	0.019	0.105
Rank	0.0006	0.000	3.019	0.003	0.000	0.001

Omnibus:	177.637	Durbin-Watson:	1.950
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4872.855
Skew:	4.436	Prob(JB):	0.00
Kurtosis:	29.090	Cond. No.	222.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 7b. Regression results of Figure 7a.

Before interpreting the regression, note that there are a few instances of universities with flawed inputs for age. For those that recently underwent mergers, the Wikidata dataset list the year of merger as the founding year for the universities that merged, rather than the founding year of one of the merging universities. Doing so would cause the y-coordinate for a university to be higher than expected, suggesting that the university is producing more notable graduates over time than it actually is. The coordinate in Figure 7a with a y-value above 1.2 is one such example: it represents Aalto University, which was established in 2010 from a merger of three major Finnish universities with founding dates between 1849 and 1904. However, these cases were in the minority and are not expected to significantly impact the results of the regression.

The results in Figures 7a and 7b suggest a lack of correlation between the university that one attends and the notability that one attains (R-squared = 0.057). Arguably, this is due to the change in the prevalence of careers in the 20th century, since a greater amount of people began to become notable for their work in culture, both core and periphery, as well as for their athletic ability. Whilst attending a notable institution might, as aforementioned increase an individual's credentials and therefore their personal prestige and potential opportunities, higher education is not a necessity to access these fields in the same way that it is for individuals in academia, and to a much lesser extent, politics, which even in this era, was less dependent on education than it was on military training. Considering that this period saw the rise and perpetuation of a great deal of authoritarian regimes, it is unsurprising that fewer of these individuals attended notable institutions, as rather than gain their status through

meritocratic means, they gained their positions and consequent fame/notoriety from a wide range of means ranging from the wide range of military coups which saw relatively uneducated officers become notable, to the promotion of Communist bureaucrats due to their loyalty to the Brezhnev Doctrine rather than any form of educational background.

However, the data does not indicate whether this is because higher education has no effect, or because any form of higher education from a good institution (i.e.: from the top 200 universities) equally affects notability. A future project to determine this would involve comparing, within the same area of attachment, notable persons who have attended institutions from different ranking blocks (e.g.: persons who attended universities ranking 1st-200th versus 501st-1000th). If higher education has no effect, then no statistically significant difference in notability would be found between the notables in different ranking blocks; if the effect of higher education is dependent on university rankings to some degree, then notables who attended universities in worse ranking blocks would be less notable than those from better ranking blocks. It is also possible that the university ranking system used is not a proper reflection of the ability of universities to produce notable graduates; a future study could use other university ranking systems or incorporate them into a university ranking system average, which would then be used to compare against graduate notability.

4. Conclusion: Where to send your kids?

Few would argue that higher education is not beneficial for the vast majority of people when it comes to chances at notability, or at a better life in general. Our findings, however, indicate that for all the effort people and institutions alike put into chasing prestigious higher education, the results are not at all life changing. Indeed, they seem to suggest that attending any of the top 200 higher education institutions will have a similar effect on future notability. As was previously discussed, the reasons are manifold and could range from the numberless avenues to notability to the ever-decreasing threshold to it that comes with our interconnected world.

As many of the world's 'elite' universities remain out of reach for the less financially fortunate, it is a relieving thought that their chances at fame and notability are not overly hampered depending on the institution attended, given that many of the top 200 universities are not 'elite'. This offers a sense of comfort, in that aiming for the best universities is not necessarily as good as a long-term investment than aiming for the last decile. Whether this trend applies to the rest of the world's universities or for specific professions or time periods remains to be seen.

In conclusion, worrying about your kid's future alma mater may not be worthwhile. There is no shame in chasing big names and the 'prestige' they give to their students, but at the same time, one should not despair if these lofty goals are not reached. If one has what it takes, notability will come.

5. Appendix

Below is a list of all additional material included with this study.

1. MAIN FILES
    a. notable-graduates.ipynb – the main Jupyter notebook for data creation and management
    b. notable-graphs.ipynb – the second Jupyter notebook for data analysis and graphs
    c. wikidataParser.py – the main parser that, given a list of Q-codes, grabs the higher education of the people included from Wikidata, if available
    d. THEparser.py – separate parser for the Times Higher Education rankings, converts the plaintext version of the rankings into a usable database
2. DATASETS
    a. MAT_alma_mater_top100_436countries.csv – top 100 most notable people born between 1940-1960 all over the world, from the Brief History of Human Time database
    b. THE2020.csv, THE2020_v2.csv – the formatted THE 2020 rankings in two slightly different variants
    c. merged.csv – the combined dataset of notable people with their education attached
    d. top200unis.csv – the THE top 200 with founding year, number of graduates, notability measurement all included (combination of uniGradsWithStats.csv, unisWithFounding.csv, and top200qcodes.csv files)
    e. uniGrads.csv – dataset of every university from the THE with at least one graduate, number of graduates counted
    f. codeListFull.txt – list of all twenty-two thousand Q-codes that were parsed
    g. unis.txt – plaintext list of THE universities
3. UTILITY FILES
    a. ‘Split code lists’, ‘Processed chunks’ folders – contains the 11 chunks that allowed step-by-step parsing of the main database
    b. THE 2020 Rankings.txt – plaintext of the THE 2020
    c. top200qcodes.csv – Q-code list for the top 200 THE universities
    d. uniGetter.py – script to create unis.txt from the rankings
    e. qCodeGetter.py – web-scraper that gets university Q-codes from Wikipedia with top200links.txt

4. MISCELLANEOUS
    a. ‘Stata’ folder – contains helper files and script to generate Stata commands
    b. missingCodes.txt – list of unreachable Q-codes collected during parsing, mostly pointing to deleted Wikidata entries
    c. Parsing statistics.txt – collected output of wikidataParser.py about running time
    d. wikidataParser2.py – parser built on Wikidata’s SPARQL technology instead of the Qwikidata library, ultimately unused

6. Bibliography

Blackmore, Paul. *Universities Vie For the Metric That Cannot Be Measured: Prestige*. https://www.theguardian.com/higher-education-network/2016/mar/29/universities-vie-for-the-metric-that-cannot-be-measured-prestige.

Bolluyt, Jess. *These Are the Presidents Who Attended Harvard and Other Ivy League Schools*. https://www.cheatsheet.com/culture/presidents-who-attended-ivy-league-schools.html/.

Chicago Tribune. *The importance of a college education*. https://www.chicagotribune.com/suburbs/post-tribune/opinion/ct-ptb-heckler-guest-column-st-0912-story.html.

Gergaud, Olivier, Morgane Laouénan and Etienne Wasmer. "A Brief History of Human Time: Exploring a database of 'notable people'." *Sciences Po Economics Discussion Papers 2016-03* (2016).

Kerr, Clark. *The Great Transformation in Higher Education, 1960-1980*. 1991.

Mihut, Georgiana. "What Germany and Romania have in common: The impact of university prestige on graduate employability." *Working Papers in Higher Education Studies* (2015): 47-67.

Our World in Data. *Tertiary Education*. https://ourworldindata.org/tertiary-education.

Psychology Today. *What Is the Purpose of a "Modern" University?* https://www.psychologytoday.com/gb/blog/sexual-personalities/201901/what-is-the-purpose-modern-university.

The Atlantic. *At Private Colleges, Students Pay for Prestige*. https://www.theatlantic.com/education/archive/2018/09/america-private-college-tuition/569812/.

Times Higher Education. *World University Rankings 2020*. https://www.timeshighereducation.com/world-university-rankings/2020/world-ranking.

University of Oxford. *Famous Oxonians*. http://www.ox.ac.uk/about/oxford-people/famous-oxonians.

—. *Introduction and history*. https://www.ox.ac.uk/about/organisation/history?wssl=1.

Wakin, Daniel. *The Juilliard Effect: Ten Years Later*. https://www.nytimes.com/2004/12/12/arts/music/the-juilliard-effect-ten-years-later.html.