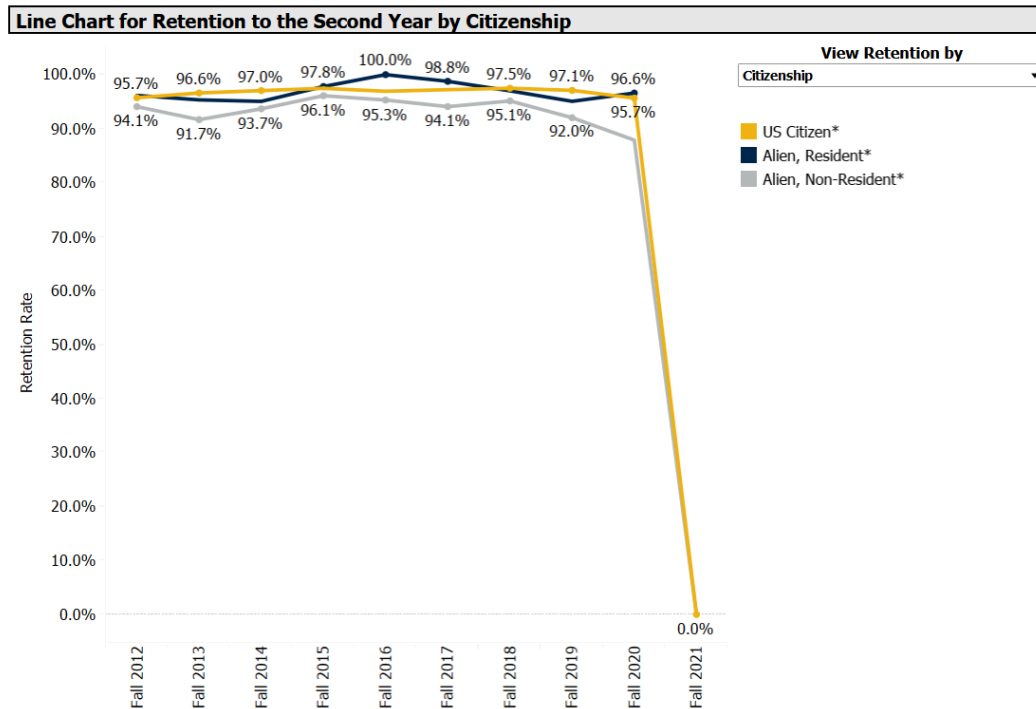


Assignment P5

Jackson McDonald
jmcdonald86@gatech.edu

1 QUESTION 1 - ANALYTICS PROMPT

The visual I decided to use is this pretty basic generated report on retention rates at GTech broken down by citizenship, taken from lite.gatech.edu.



From the graph we can see starting around 2016 the retention rate among resident and non-resident non-citizens (aliens) begins to reverse trend and drops, while US citizen retention rates remain the same. I could draw the conclusion that US policies and political rhetoric starting in 2016 and carrying through the next few years could have impacted the retention rates of non-citizens. The policies and rhetoric coming from that time period were in my opinion xenophobic, and could have created an unwelcome feeling among non-US citizens in the country, as well as GTech, leading to a drop in retention rates after they had been steadily increasing. The lack of any movement or change for US citizens points to some change impacting non-US citizens. The core message from this visual is that the non-US citizens haven't changed in academic ability or passion, external events are impacting their resolve to stay enrolled. For the record, none

of these conclusions can actually be drawn from this simple graph!

If I were to re-design to visualize to show something else, I would add an additional y-axis to show the rates of retention specifically for those who are taking advanced classes (think masters or PHD). What this would show is that the percentage of non-US citizens enrolled in advanced courses in that time shot up as it became harder to generally immigrate into the country and those who had the qualifications to get into advanced degrees were still able to enter the country. This meant that the academic rigour for the average non-US citizen enrolled at GTech from 2016 on-wards was higher than that of US citizens. The xenophobic rhetoric had no impact on non-US citizens desire to stay in the US, but instead a greater number of them were enrolled in challenging advanced courses and thus the retention numbers dipped down naturally as a result, and were over sampled amongst that group of people. The US citizen cohort had no such impact as their undergraduate and graduate degree numbers would have remained the same.

The difference between the two visuals is that the original one placed emphasis on general xenophobic climate as a whole that made non-US citizens feel unwelcome and thus decide to leave the US/GTech and drop retention rates. The second re-designed visual would show that as a result of an increased proportion of non-US citizens enrolled in advanced degrees that their retention rates would naturally drop due to the increased challenge. It shows there is nothing to do with non-US citizens feeling unwelcome, just a change in the proportion of undergrad and graduate non-US citizen representation. The key difference is that the re-designed visual adds context to could be the driving force for the drop in retention rate. While the first visual requires a massive jump in logic, adding another dimension offers a more logical interpretation. It isn't a difference in interpretation, it is just one done with more information.

(If you read this answer I apologize for the terrible visual and my terrible interpretation of it + redesign :))

2 QUESTION 2 - ANALYTICS PROMPT

When it comes to data privacy and deductive disclosures, it is a topic pertinent to my field of education (Surprise I'm sure based on the last answer hahah) where we want to ensure that student data is protected and only accessible by those

who require it (teachers, parents, authorized 3rd parties, researchers). I work on the team that manages the data and I believe it is our responsibility as managers of that data to ensure it is kept private. A mistake in granting someone access who shouldn't get it can be devastating to numerous people. Many times when we present student level information to researchers or 3rd parties, we de-identify data by removing student IDs or randomize them, but as Dr. Joyner's example shows, it can still be determined who is who in certain cases. In an ideal case if a researcher/3rd party comes across a bad job on our part to de-identify the data they should let us know, and also shouldn't purposefully try and determine specific individuals. While data custodians are responsible for safeguarding it, they should not be responsible for actions of nefarious individuals (to the same extent).

If I were to design a dataset or a procedure to reduce the change of deductive disclosure I would first always remove identifiers completely. Even randomizing them can be determined as all random number generators are deterministic. Then I would ask the researchers what form of analysis they are actually trying to accomplish and the relevance of certain variables. If for example they wanted longitudinal data but didn't care about the actual years, just sequence of years, then change the actual years from 2017, 2018, 2019 to Year 1, Year 2, Year 3 etc. This would be the same Year 1- Year 3 format for all even though the actual years could be 3 different ones across all time. Similar things can be done with other variables, like in Dr. Joyner's example, instead of the specific math class, if you could just provide the student was enrolled in a math class, any math class, and suffices for research purposes, then do so. If possible have two separate datasets, one with all the specific information and no one has access to that except the engineering and analysts, and then the dataset with the abbreviated/generic variables is what is also maintained and provided when necessary.

The major drawbacks associated with my design is a lengthy back and forth with researchers and the engineering/analyst team to determine exactly what they want, how they want it, and in what way it can actually be used. The researchers may discuss topics that are too advanced for our team and thus we wouldn't know the best way to go about it. The other issue is that when trying to provide transparency, the researchers may discover that we hold more sensitive data than they anticipated and then request it (de-identified), but it may be so sensitive and so vulnerable to deductive disclosure that it would be impossible

to, and then that research couldn't happen, or the relationship gets strained. This new process would involve a lot of time and energy for an already stretched out team.

3 QUESTION 3 - 2 PAPERS

3.1 Paper 1

Title : Considering Parents in Coding Kit Design: Understanding Parents' Perspectives and Roles

Authors : Junnan Yu, Chenke Bai, Ricarose Roque

The paper focuses on the lack of research around the parents' role when it comes to coding kits for children, and the researchers hope to formalize some concepts for future designers of coding kits to be aware of and incorporate. The researchers were able to gather insights by interviewing 18 parents of children aged 3-9 who had received coding kits. The child age was chosen because of the importance of the parents' role in teaching and development in that time period. The main expectations that parents had from their child using a coding kit was that they could become proficient in coding in the future as they saw it as another language, as well as become comfortable with modern technology. The researchers were able to label the types of roles that parents took on while supporting their child with coding kits, such as teacher, scaffolder, and collaborator. One of the parental concerns highlighted from the interviews was the helplessness some felt when interacting with the coding kits and the inability to fully support their child. Their own lack of coding knowledge or confidence in the technology made them feel inadequate. As a result of this the researchers emphasize designing coding kits to empower parents to adopt scaffolding and teaching roles so that even when they can't do the technical aspects themselves, they can translate the concepts to other areas they are comfortable in.

The reason I chose this paper was the researchers' way of tackling an important aspect of any children's learning tool - the parents' or guardians role. Something we take for granted is that when we build something for kids it can really be for adults to do with their kids, but that might not always be top of mind when designing things. In school we provide reports on kids for the adults to consume, but other times we might make a report for kids to absorb but also needs to be readable and actionable for adults. The researchers look into the primary and

secondary stakeholders, where the line between both types shift depending on the role the parents adopt (such as collaborator and interact with the coding kit as well).

3.2 Paper 2

Title : Digital Liminalities : Understanding Isolated Communities on the Edge

Authors : Rikke Bjerg, Jensen Lizzie Coles-Kemp. Nicola Wendt

The researchers explore the continuity of mental models of members of isolated communities when presented with interfaces that can provide representations outside of their limited physical surroundings. To gather data on this phenomenon, the researchers looked at three distinct groups : container ship employees, communities in Greenland, and welfare claimants in rural England. Liminality refers to the concept of transitioning between boundaries and borders, and in the case of these distinct groups, it is achieved digitally and through technological interfaces. Each group shared the common aspect of extended isolation at certain points of time, even though they may be surrounded by people in close proximity all the time. This feeling of isolation was amplified by the access of technology, the researchers found, as individuals felt a world apart from what they were viewing or experiencing on their phone. In some cases the mere usage of technology made some focus not on the task at hand, but on what the technology could have provided all along, as was the case for the welfare participants. The researchers emphasize that designing with social isolation of users is critical to combat this feeling that can have detrimental physical and mental health effects. Technological advances alone cannot overcome this, such as Facebook's video chat feature, but the design and interaction must accommodate the mental models that isolated people hold and how they can be disrupted less.

The reason I was drawn to this paper was the concept of social isolation in the age of globalization and international connection. While we can feel as connected as ever, for some it can lead to the polar opposite feeling. The researchers were able to find such communities that were already isolated, and technology amplified that feeling. I am also interested in social isolation because of how important it is for humans to function properly. Research shows that the greatest predictor of long lasting health is not working out, or a proper diet, but social connections and small interactions with other people. If technologies can help alleviate social isolation for some, that would be both physically and mentally

beneficial.

4 2 CONFERENCES 1 PAPER EACH

4.1 Creativity and Cognition - Paper 1

Title : CrowdMuse: Supporting Crowd Idea Generation through User Modeling and Adaptation

Authors : Victor GirottoSchool, Erin Walker, Winslow Burleson

The researchers look to take advantage of crowdsourcing when it comes to idea generation and brainstorming. On top of crowdsourcing, certain individuals within groups are provided greater weights in their opinion based on prior contributions to other brainstorming events. The system the researchers create, CrowdMuse, is interactive in nature and allows users to view and manipulate individual ideas, and also see the density of certain idea types created. The researchers also leverage the concept of inspiration when it comes to brainstorming and can present random ideas to participants who would like to see what is currently being offered. The results of experiments determined that the inspiration feature improved the breadth of ideation (number of categories surveyed by one ideator), as opposed to the interactive feature which had a lesser impact. Each feature had no impact on fluency (number of ideas generated) or depth (within category fluency).

I wanted to explore this paper as the concept of crowdsourcing ideas is an interesting one. We do so with concepts such as voting for elected officials or group majority decides what to do, but when it comes to other activities we try to lessen the number of people involved. Sometimes that is for logistical reasons and other times for efficacy and research backed methods. If it were possible to show that ideas could be successfully generated through crowds, that could open up more avenues for more direct input from people. In my own job, it would be great to receive feedback from a wide range of people, but for convenience sake that isn't possible. If there were something like CrowdMuse, that could be the solution!

4.2 International Conference on Educational Data Mining - Paper 2

Title : Increasing Enrollment by Optimizing Scholarship Allocations Using Machine Learning and Genetic Algorithms

Authors : Lovenoor Aulck. Dev Nambi, Jevin West

These researchers wanted to use machine learning to best optimize the resources the university has at their disposal, and one of them is scholarships and funding for prospective students. As universities attempt to target certain groups of students to enroll, they can change the size of the scholarships they offer to increase the odds of a student accepting their offer. To the researchers this is clearly a solvable optimization problem. The optimization problem was to develop a merit-based system that didn't look at ethnicity or demographic details, but on standardized tests and high school GPA's. Tuition due and enrolled status were the two main variables that the researchers wanted to test for correlation, of which they found a negative relationship. They used a gradient boosted tree model to build out the optimization model to determine how much more scholarship should be provided to change an enrollment status from no to yes for a given student that was desired by the university. After this model was deployed in the real world there was a 23% increase in accepted enrollment offers the next year!

I wanted to explore this paper as scholarships are a severely underutilized method of attracting talent from all walks of life (in the US). It isn't for lack of trying by the universities, it is just an intensive and manual task of sending out letters and reviewing applications, which some universities do not have the resources in those departments to deal with. If they had some additional tools which can optimize a process such as awarding scholarship amounts to increase the likelihood of a candidate accepting an offer, then they can shift resources to different departments such as outreach.