PAGE NUMBERS!!!!!

This is fine for a draft title, but fotr the final paper use a title that give reader some idea of what the main question(s) or result(s) were.

Simon Initiative Capstone Project

Jinghao Huang, Jenny Mao, Shihua Pei

Faculty Advisor: Prof. Valerie Ventura

4/21/2021

Abstract

The Open Learning Initiative is trying to evaluate how the COVID pandemic in 2020 has been impacted on user behaviors and students' performances, with an underlying assumption that students performed better after the pandemic. Using data from Georgia State University from Spring 2019 - Fall 2020, we use weekly logins to determine user behavior changes and average assessment correctness for students performance changes. The result shows a statistically significant decrease in logins for both instructors and students after pandemic, as well as higher assessment correctness in 2020 than in 2019. Other factors not listed in the dataset ~~can~~ may also contribute to this effect, including potential changes in course policy and course materials, due to switching to remote learning mode.

Don't center this text. Left-justified with these narrower margins is fine.

1 Introduction

Open Learning Initiative (OLI) is a scientifically-based online learning environment that supports learning and instruction with high-quality, scientifically-based classroom-tested online courses and materials. It allows educational institutions to share courses and materials openly and freely and develop a community of use, research, and development. Research at UCD has shown that students' performances went up during the pandemic when being compared to pre-pandemic performances. This project is interested in exploring whether similar findings in ~~the~~ Georgia State University (GSU) dataset can be found.

citation

redundant

spell out and add citation

a

add citation

The goal of this project is to define any use pattern changes on the OLI platform for instructors and students at GSU statistics course from Spring 2019 - Fall 2020 due to COVID19. This project focuses on comparing OLI use patterns for both instructors and students before and during the pandemic, and investigating how students' performances are changed during the same time period.

Tables should generally have the appearance illustrated in https://texblog.org/2017/02/06/proper-tables-with-latex/ (the examples are in LaTeX, but you can do the same thing in msword).

2 Data

was provided the (OLI)

The data for this paper ~~is given~~ by Open Learning Initiative. We have MySQL files for each semester for the academic year 2019-2020. For each semester we have about 34 csv tables. Listed ~~below~~ in Tables 1 -- 4 are some of the important tables we are using and the variables inside the tables:

(add discussion of variables and meanings)

give the tables captions and table numbers

Students	Instructors	Course	Other
Student	Instructor activities per week	Course	Content
Student activities per week	Instructors	Modules	Activities
Student results per unit		Question summary	Units
Student results per module		Questions by unit	

Table 1: blah blah blah

missing border

Table name	Variable name							
Instructor action per week	user_id	course key	week of year	week logins	dashboard views	gradebook views	gradebook action	activities started
Student action per week	user_id	course key	week of year	week logins	pages first accessed	activities started	learn-by-doing started	did-i-get-this started

Table 2: blah blah blah

Student results per week	user_id	course key	week of year	week logins	pages first accessed	activities started	learn-by-doing started	did-i-get-this started
Student	user_id	first name	last name	course key	total logins	checkpoint mean	quiz mean	checkpoint mean

Table 3: blah blah blah

Semester	2019Spring	2019Fall	2020Spring	2020Fall
Number of students	1,524	1,672	1,686	2,078
Number of instructors (identified by name)	2	3	4	2

Table 4: blah blah blah

3 Methods

The goal of this project is to see whether there were any behavior changes in a GSU Statistics course because of the COVID pandemic starting in the Spring 2020 semester. Primarily we focus on whether there are any use pattern changes in the instructors' and students' behaviors, as well as whether there are any changes in students' performances. To determine the impact of switching to remote teaching mode in GSU starting at the end of March 2020, we compare user behaviors in the Fall and Spring semesters separately, due to different behavior patterns in the Fall and Spring semesters in 2019, before the pandemic.

capitalize

It would be great to illustrate some of this with EDA in the data section, like in your progress report talks.

Besides separating user behaviors by spring and fall semesters, we also focus on the two professors who taught the multiple semesters between Spring 2019 – Fall 2020. Prof. H taught all four semesters where Prof. B taught the first three semesters. The assumption is that each professor has very different use patterns, and students in the same professor's sections would most likely perform similar use patterns. In addition, there are lots of low logins and some double student registration in Spring 2019, we might not consider this semester for student use pattern analysis.

have use patterns more similar to each other than to students in the other professor's sections.

In final paper, tell reader definitively whether you used the Spring 2019 data or not.

The main analysis for the use pattern changes is to perform 95% confidence level and hypothesis testing to determine whether the differences in the two semesters are statistically significant.

The above is fine so far, but in the final paper I want to see a list and description of the analytic methods you actually use.

4 Results

4.1 Instructor Login Changes

We first look at Prof. B and Prof. H's weekly login patterns. Prof. B did not teach in Fall 2020 so we can only compare login patterns for Fall 2019 and Fall 2020. Week 14 was when GSU switched to online teaching mode.

I don't get this. If Prof B didn't teach in Fall 2020, how could you compare fall 2019 with fall 2020?

change to "Prof B"?

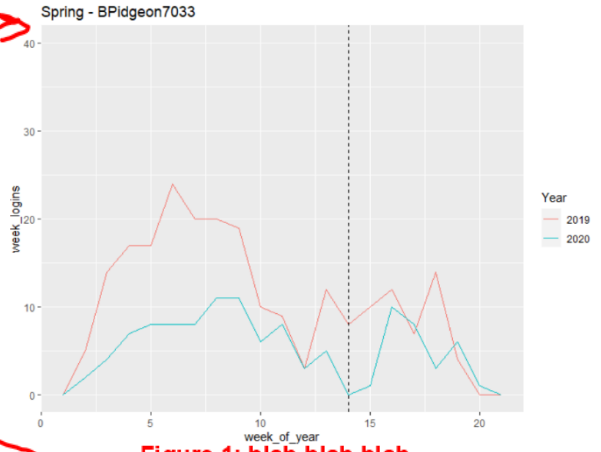


Figure 1: blah blah blah

Give each figure a caption and a figure number, and refer to the figures by number.

Discuss each one, by figure number, and explain how it contributes to understanding of the problem.

Does this need to be "Prof H"?

repace blurry figures with clear ones

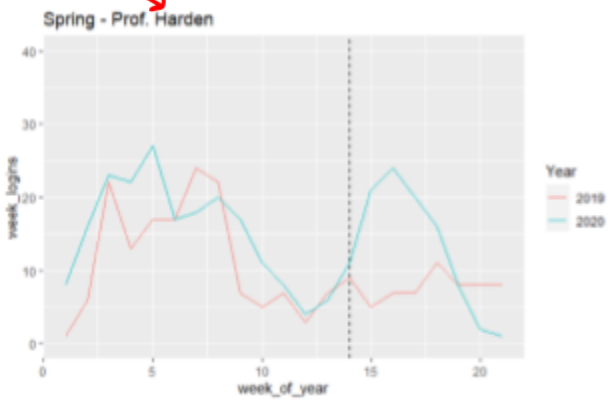


Figure 2: blah blah blah

Figure 3: blah blah blah

It does not look like the login patterns for Prof. B changed much during week 14 in the spring semesters, and the use patterns for Prof. H in the spring semesters does seem to change after week 14. It is much more obvious that the login patterns are different in the fall semesters. Through hypothesis testing we confirm there is a statistically significant decrease in instructor logins in Fall 2020 (p-value = 0.0012). Difference in the spring semesters after week 14 is not statistically significant for neither of the instructors at 5% level. Detailed R codes and plots can be found in Appendix 1.

refer to figures by number so reader knows where to look

The method that you used here should be mentioned in the Methods section (ANOVA? two sample t test after checking for comparable variances? or ... ?)

good - definitely the right thing to do -- but I don't see a tech appx in this draft yet...

4.2 Student Performance Changes

Text referring to figure should precede figure.

Give this figure a caption and a figure number and refer to it by number in the text

Which professor or professors are involved here?

Only one?

Both? If both, does there need to be another factor to account for possible differences between students in different professors' classes?

Etc.

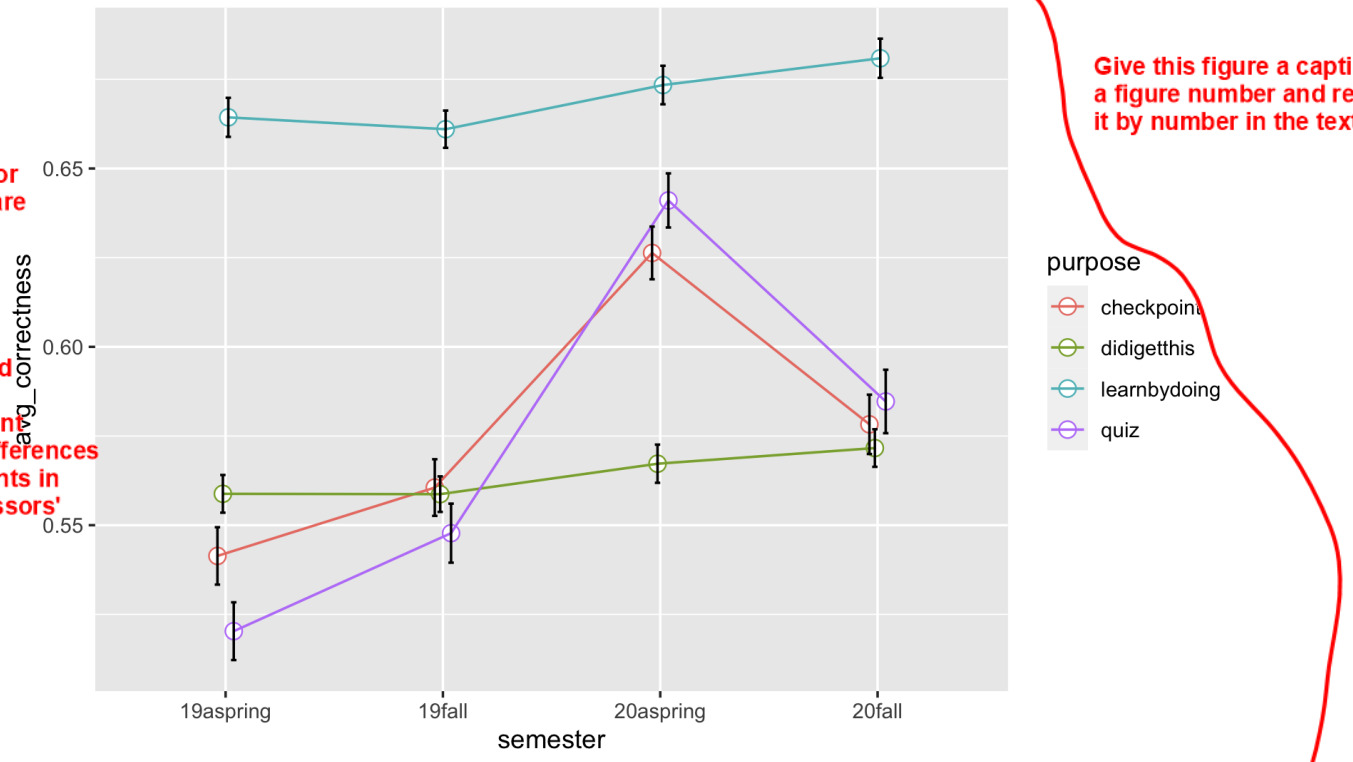


Figure 4: blah blah blah

The plot shows the changes of average correctness for all the sections during the four semesters, grouped by type of assessments. High stake assessments include quiz and checkpoint, while low stake assessments include learn-by-doing and did-i-get-this. From the plot, we can clearly see a huge increase from Fall 2019 to Spring 2020, especially for quiz and checkpoint, followed by a drop in Fall 2020. The average quiz and checkpoint correctness are still higher than those in the 2019 calendar year.

will add in plots/explanations for changes in assessment performances focusing on the 2 professors taught in multiple semesters.

may add in plots/explanations to display changes in correctness by module/unit so we can see how correctness varies as the semester progresses.

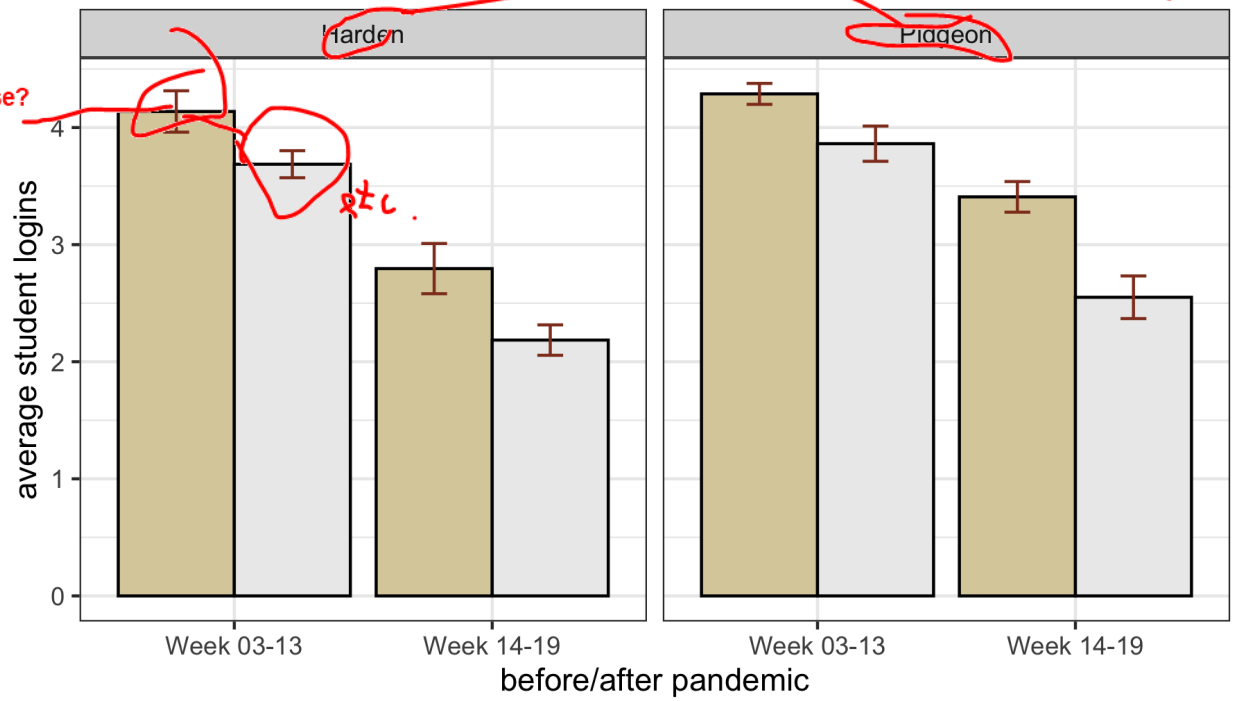
Good, thanks for letting me know

It seems to me a 2-way ANOVA with interactions would be appropriate to supplement Figure 4 above also.

(or a 3-way ANOVA if you also have to account for professor(s) in Fig 4)

Earlier you just said Prof H & Prof B. Choose one way of referring to the professors and stick with it throughout report.

4.3 Student Login Changes



What are these? Why are they in the graph?

What formal statistical analysis can you do to confirm the results that seem to be shown in this graph?

Figure 5: blah blah blah

Text referring to figure should precede figure.

Even before you discuss results, you need to explain to the reader what is in the figure

We observe that the average student logins in Spring 2020 is significantly lower than that of Spring 2019. The average student logins after COVID-19 are lower than before COVID-19. One possible reason for the low logins after COVID19 is that students log in to the system fewer times but have longer study sessions because of remote learning.

Is there evidence for this conjecture in the data you have? If so, explain...

why are these dashed lines here? Do you intend to compare transition at week 14 between 2019 and 2020?

Student weekly logins in Spring 2019

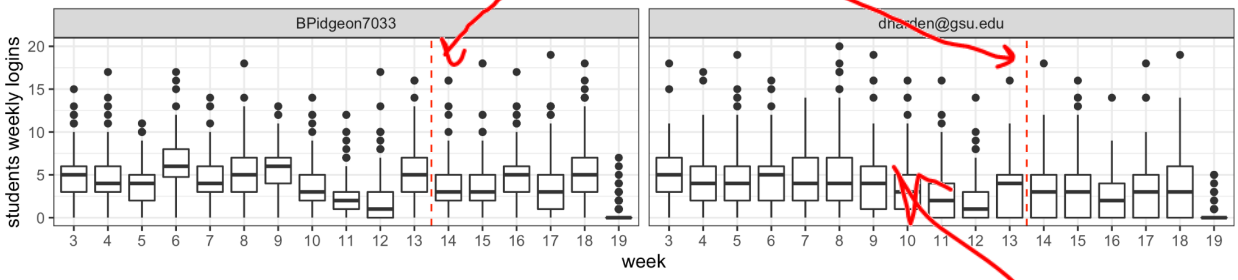


Figure 6: blah blah blah

Student weekly logins in Spring 2020

There is some pretty severe skewing in Figs 6 and 7. What happens if you try a logarithm or similar transformation?

(bad break here. don't allow figure to be separated from title in final report)

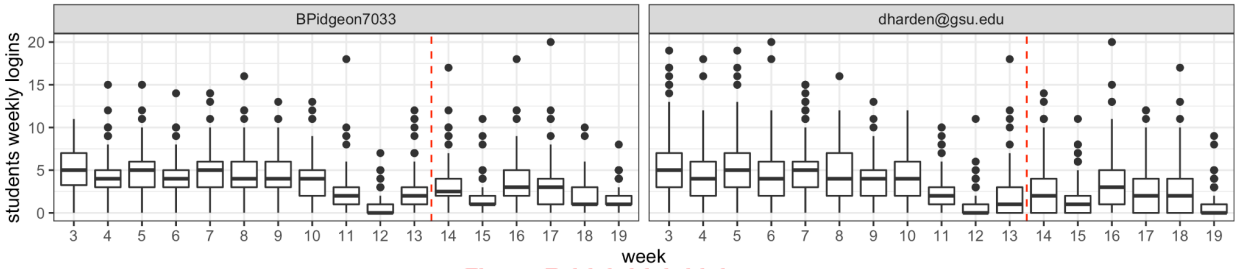


Figure 7: blah blah blah

What kind of analysis can you do to confirm patterns you see in the data?

We observe that student logins per week for different instructors seems to be different. The red lines indicate when the pandemic happened in the spring of 2020. If we compare the students' weekly logins for the same professor in Spring 2019 and Spring 2020, we observe significantly lower students' logins after the pandemic in 2020 than in the same time period of 2019.

Again, text should precede figures it refers to, and it should refe to figures by number.

Interrupted time series analysis? Changepoint model? Or ...?

may add in plots/explanations to display changes in students actions by module/unit (% of opportunities completed, # of hints, etc). May not add this due to different module/unit definitions in 2019 vs. 2020.

What differences between 2019 and 2020 would these models confirm as important?

Thanks. Good to know. In the final report make very clear what you are doing and why.

5 Discussions

focuses on project constraints for now ok

In this project we mainly use student login data and assessment scores to determine student behavior changes, however, there are lots of other effects that are not captured by our dataset. For instance, the change of course policy due to COVID will likely impact student performances. If instructors allow students to drop more high/low assessments than before, students may not perform as well on these assessments as previously. Instructors may also change the type of questions asked in online assessments, which can make the students' performances less comparable in 2019 and 2020 calendar years. Additionally, the pandemic might also lead to the change of material covered in Spring 2020 compared to Spring 2019.

Good start. When you revise this parargraph for the final paper, discuss specific examples of course policy changes, changes of material covered, etc., that support your conjectures.

I look forward to the rest of the discussion in the final paper!

Reference

Good start on references.

Bier, N. (2020), 36-726 Capstone Description. Unpublished assignment sheet, Open Learning Initiative, Carnegie Mellon University, Pittsburgh PA.

RStudio Team (2020). R Studio: Integrated Development Environment for R. RStudio, PBC, Boston MA. URL http://www.rstudio.com/.