Education During Covid-19

Stephanie Gamino

5/6/2022

Introduction

The spread of Covid-19 brought the world to a halt. The virus is easily transmitted by droplets from the infected person's sneeze, cough and even breath. Aside from close contact transmission, corona virus could be transmitted through air and even surfaces which people touch, before touching their noses and mouths. Because it didn't take much for the virus to be transmitted between individuals, the entire world had to quarantine in their homes to control the spread of this infection. At the time, not many knew what was going on, but quarantining and self-isolation could help stop the spread. This was a rare and fast growing situation, but both students and teachers had to make the transition to online learning to continue their education.

Many teachers and professors resorted to using digital learning platforms to ensure students stayed engaged whilst learning online. This paper will focus on which learning platform had the highest engagement throughout 2020 and what subject category the learning platform focused on. I hypothesize that learning platforms that focus on math will have highest engagement. This may be because since math requires some demonstrations and example walk-through, it might cause students to use math digital platforms more and for longer periods of time rater than other subjects.I predict that student engagement will decrease during the months of June through August. Additionally, I presume learning platform engagement will be at it's highest during the last month of their semester which is typically May for the spring semester and December for the fall semester.

Methods

The data used for this analysis was provided by Kaggle which is a machine learning and data science community. The data was taken from an analytic competition called "LearnPlatform COVID-19 Impact on Digital Learning". This data set includes school district information, different learning platforms, and other educational technology that were downloaded in 2020. Within this data set, there was 'product_data' that provided characteristics of different learning platforms. This ranged from each products name, to their URL, to their specific purpose as an electronic tool. For the purposes of this project, I extracted only products specially for digital learning. There were many other products listed such as Google Docs and Google Classroom that fell under school operations purposes. Although those platforms may have been very helpful, they aren't as relevant to learning platforms that interacted with students direct education. There was also 'district_data' which provided characteristics per district. For instance, it provide the district's number, the state it was associated with and location categories of city, rural, and suburb. This information was obtained from the National Center for Education Statistics (NCES), The Federal Communications Commission (FCC), and Edunomics Lab. The last data provided was an 'engagement_data' folder which broke down each district individually and provided downloads there were per day, per digital product, subject category, and the engagement per download. Engagement was measured by one-page loaded event by a given platform. Meaning, if a student used this platform, every time the student would open a new page within the platform it recorded it. An example of this is browser history. Google for example, records all the pages a user loads. A user could load multiple pages from just one website. The categories these products fell under were math, science, and language/reading. The engagement data was taken from the LearnPlatform's Chrome Extension. After combining all this data, there was over 22 million individual downloads recorded. The number of students downloading these products is unknown.

Results

Since there were over 22 million downloads throughout 2020, it is important to visualize that grand number by the category they fall under. Figure 1 shows months on the x axis and the the amount of downloads on the y axis, by the three subject categories.

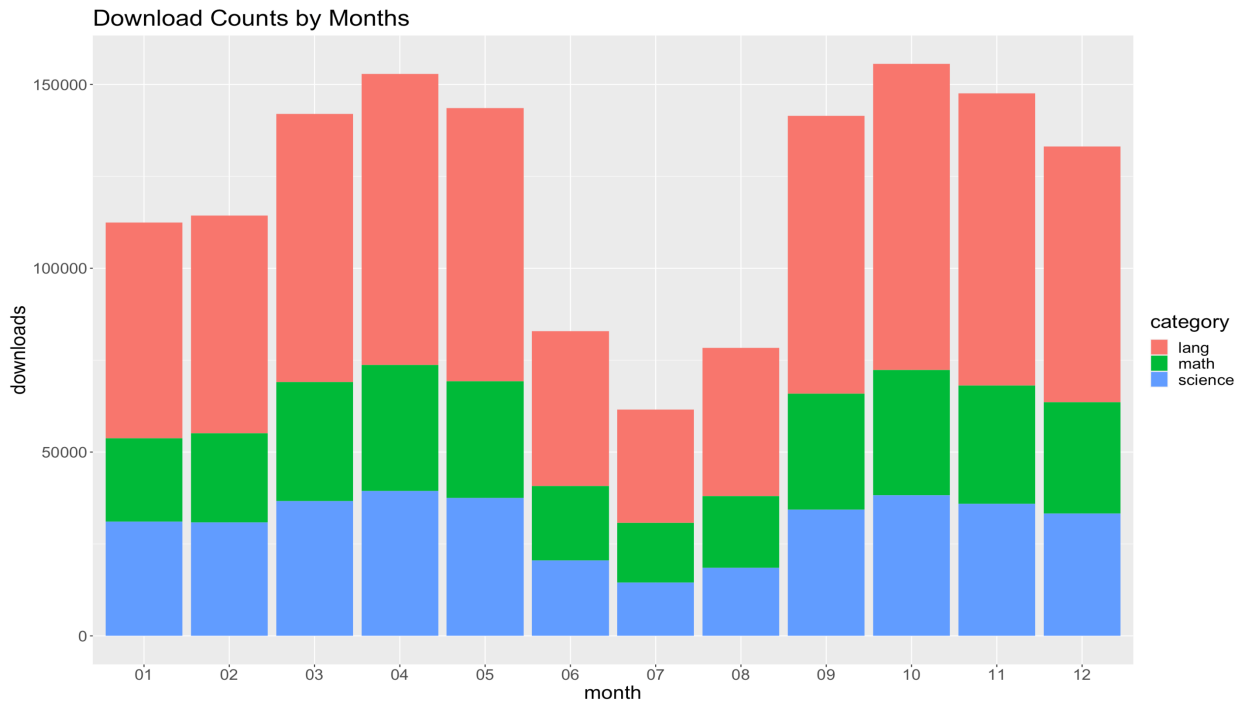


Figure 1: Download Counts by Month: The oranage color represents all language learning platforms, the green is math, and lastly, blue represents science.

Throughout 2020, it seems that digital learning platforms associated with language and reading had higher downloads in comparison to math and science.

Figure 1 shows the raw data of downloads across all the provided states in the US in 2020. The next figure, Figure 2, is a similar visual only, it displaying all digital learning platforms over the overall average engagement.

As shown in Figure 2, across all learning platforms over average engagement, there is one digital learning platform that has the best average engagement. The digital product is called 'i-Ready!' which is a both math and reading oriented program. Although it is both, it is categorized as a math program because a lot of educators use other resources to test students language skills. This product was originally created in 1969 to serve the purpose of publishing student workbooks.

To asses which learning platform had the highest engagement overall and in what category, I used a linear model to show the relationship between engagement and categories of the learning products. The following is the syntax for the linear mixed model I ran: lmer(pct_access ~ category * month + (1 | state) , data = grand_merge, REML= FALSE). In this model, the dependent variable is engagement and how its being effected by its predictor, the learning platform categories, while including an interaction of month and random intercept of state. Figure 3 is an image of the results from the lmer model.

To summarize whats in Figure 3, one must understand how the results are read. It is comparing engagement in the category of language in the first month of January, to the rest of the categories and months by subject. To start, the first two rows in the fixed effects portion, shows categorymath and categoryscience being compared to category of language. To start, we can see that the math category has a positive effect

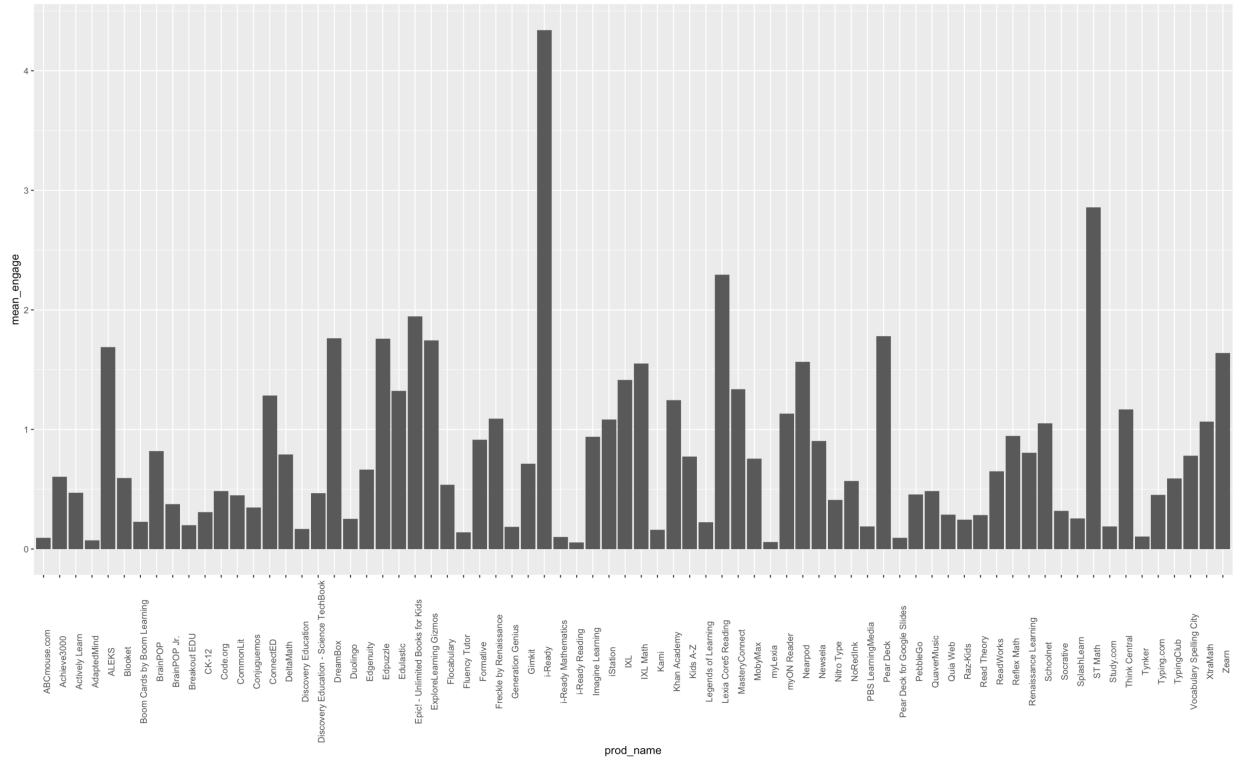


Figure 2: Figure 2: Most Used Learning Platform on Average.

while being compared to language category, while science has a negative effect. Meaning overall usage of science is lower than language and math is being used twice as much as language. Additionally, they provide the comparison of language in January versus the other categories in each month being compared to January. Ultimately, as language usage in January is being compared to math usage in all the rest of the months, one can note that overall there is a decrease or lower effect of engagement in math while being compared to February and the rest of the months. From this we can tell that as time continues within the students semesters, students had the highest engagement in math in January. Similarly, while science was being compared to the language usage in January as well, there was a negative effect. Conveying, the overall engagement for language platforms was higher than science learning platforms. While comparing language to science over the months, it seem there is a negative effect when comparing science usage over the months. Demonstrating, that the engagement on language platforms in January was also greater while being compared to engagement in science in all the other months.

While looking at the random effects, the random intercept of state was added into the model to see if there was any effects of states while comparing the engagement to subject categories. The variance reported was 0.34 which concludes that the random effect of state did not have much effect on subject categories. Overall, the model was able to tell us that language had higher engagement when being compared to the other subject categories throughout the months. Language and math learning platforms had more engagement overall compared to science learning platforms.

Discussion

The intended outcome was finding which learning platform had the highest engagement and in what category. The linear mixed model helped show that language learning platforms had most overall engagement. To further show this, Figure 4 is a visual representation of engagement across the months by subject and by state.

In Figure 4, it is very apparent that language platforms exceed engagement through the states provided.

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: pct_access ~ category * month + (1 | state)
   Data: grand_merge

     AIC      BIC   logLik deviance df.resid
 6782785  6783248 -3391354  6782709  1465697

Scaled residuals:
   Min     1Q Median     3Q    Max
-1.524 -0.348 -0.199 -0.024 32.840

Random effects:
 Groups   Name        Variance Std.Dev.
 state    (Intercept) 0.322    0.5675
 Residual             5.987    2.4468
Number of obs: 1465735, groups:  state, 23

Fixed effects:
                            Estimate Std. Error         df t value Pr(>|t|)
(Intercept)                1.098e+00  1.190e-01  2.257e+01   9.228 4.04e-09 ***
categorymath               1.081e+00  1.911e-02  1.466e+06  56.568  < 2e-16 ***
categoryscience           -3.021e-01  1.718e-02  1.466e+06 -17.586  < 2e-16 ***
month02                   -6.488e-03  1.426e-02  1.466e+06  -0.455 0.649112
month03                   -2.719e-01  1.358e-02  1.466e+06 -20.026  < 2e-16 ***
month04                   -3.221e-01  1.334e-02  1.466e+06 -24.140  < 2e-16 ***
month05                   -4.497e-01  1.352e-02  1.466e+06 -33.250  < 2e-16 ***
month06                   -7.346e-01  1.565e-02  1.466e+06 -46.949  < 2e-16 ***
month07                   -8.959e-01  1.722e-02  1.466e+06 -52.028  < 2e-16 ***
month08                   -5.080e-01  1.585e-02  1.466e+06 -32.050  < 2e-16 ***
month09                    6.300e-02  1.348e-02  1.466e+06   4.674 2.95e-06 ***
month10                    6.128e-02  1.321e-02  1.466e+06   4.639 3.49e-06 ***
month11                   -6.343e-03  1.333e-02  1.466e+06  -0.476 0.634265
month12                   -1.641e-01  1.373e-02  1.466e+06 -11.950  < 2e-16 ***
categorymath:month02      -1.019e-01  2.670e-02  1.466e+06  -3.816 0.000136 ***
categoryscience:month02   -1.556e-02  2.430e-02  1.466e+06  -0.641 0.521784
categorymath:month03      -4.042e-01  2.514e-02  1.466e+06 -16.076  < 2e-16 ***
categoryscience:month03    2.058e-01  2.325e-02  1.466e+06   8.852  < 2e-16 ***
categorymath:month04      -6.725e-01  2.480e-02  1.466e+06 -27.117  < 2e-16 ***
categoryscience:month04    2.822e-01  2.286e-02  1.466e+06  12.343  < 2e-16 ***
categorymath:month05      -7.843e-01  2.518e-02  1.466e+06 -31.150  < 2e-16 ***
categoryscience:month05    2.589e-01  2.314e-02  1.466e+06  11.188  < 2e-16 ***
categorymath:month06      -9.909e-01  2.833e-02  1.466e+06 -34.971  < 2e-16 ***
categoryscience:month06    2.780e-01  2.701e-02  1.466e+06  10.293  < 2e-16 ***
categorymath:month07      -1.047e+00  3.045e-02  1.466e+06 -34.379  < 2e-16 ***
categoryscience:month07    3.179e-01  3.005e-02  1.466e+06  10.580  < 2e-16 ***
categorymath:month08      -9.591e-01  2.863e-02  1.466e+06 -33.499  < 2e-16 ***
categoryscience:month08    9.494e-02  2.771e-02  1.466e+06   3.426 0.000612 ***
categorymath:month09      -6.292e-01  2.518e-02  1.466e+06 -24.982  < 2e-16 ***
categoryscience:month09   -2.322e-01  2.342e-02  1.466e+06  -9.914  < 2e-16 ***
categorymath:month10      -5.250e-01  2.474e-02  1.466e+06 -21.215  < 2e-16 ***
categoryscience:month10   -2.493e-01  2.289e-02  1.466e+06 -10.893  < 2e-16 ***
categorymath:month11      -5.712e-01  2.503e-02  1.466e+06 -22.821  < 2e-16 ***
categoryscience:month11   -2.265e-01  2.318e-02  1.466e+06  -9.772  < 2e-16 ***
categorymath:month12      -5.621e-01  2.548e-02  1.466e+06 -22.063  < 2e-16 ***
categoryscience:month12   -9.777e-02  2.368e-02  1.466e+06  -4.128 3.66e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation matrix not shown by default, as p = 36 > 12.
Use print(x, correlation=TRUE)  or
    vcov(x)        if you need it
```

Figure 3: Figure 3: lmer Model Results

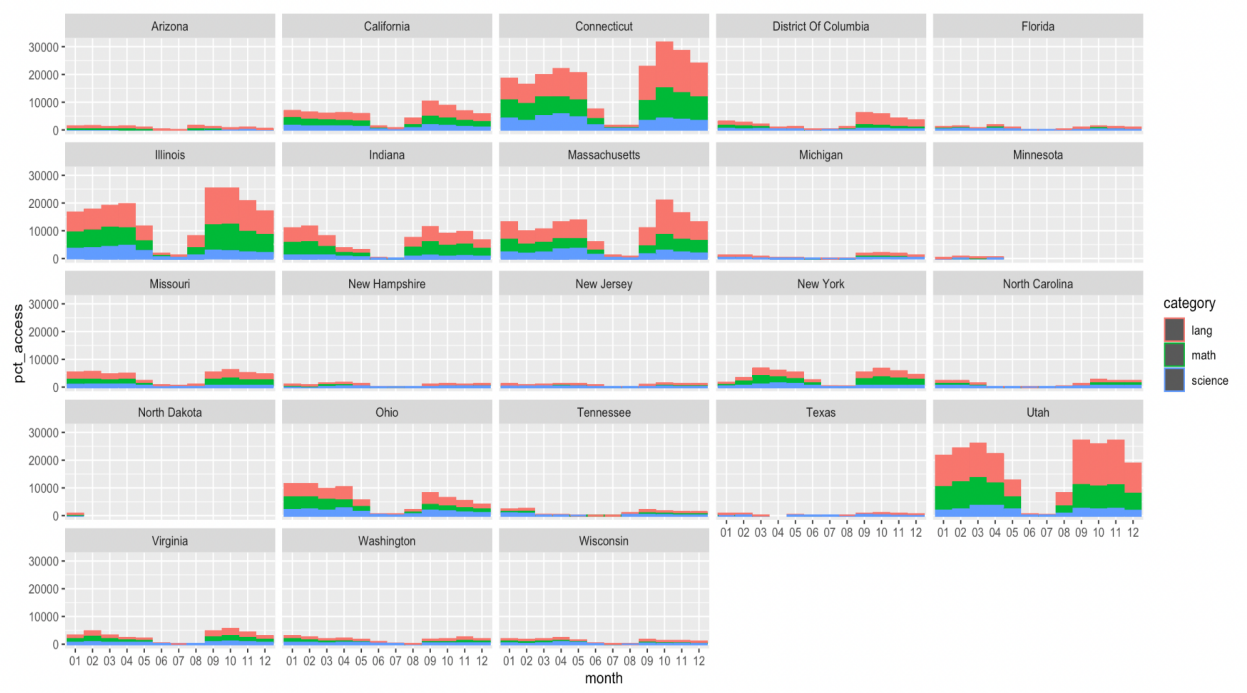


Figure 4: Figure 4: Engagment Across States

Nevertheless, this doesn't fully answer the question. Figure 5, is a visual representation of the all the learning platforms over engagment.

Figure 5 is very similar to Figure 2, but the main difference is that Figure 2 displayed average engagement and Figure 5 shows all engagement records throughout 2020. Similarly like in Figure 2, the digital learning platform 'i-Ready!' has the record for most engagement. This figure shows us the variety of learning platforms and the category they fall under. Although a math learning platform was the most used throughout 2020, there are higher frequencies in using language learning platforms rather than math ones. The second highest used learning platform was a language platform called 'Epic!'.

We can use Figure 4 to visualize my prediction that engagement will be at its highest as the semester comes to an end. May and December are typically the months school sessions end. In Figure 4, one can see that across the states displayed, months 5 and 12 have lower engagement in comparison to the other months. This proves my original prediction incorrect. My other prediction stated that while students were on summer break, months 6-8 would have little to none engagement. Figure 6 is a visual of all the learning platforms engagement across the months.

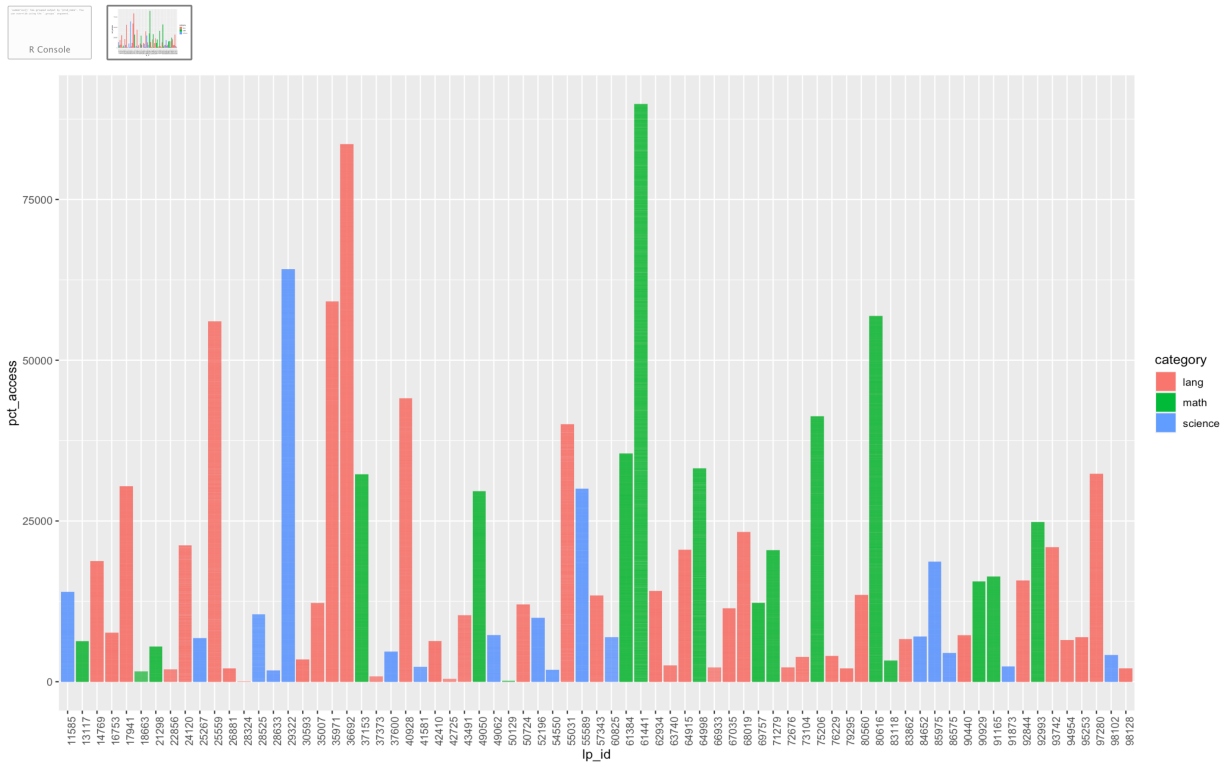
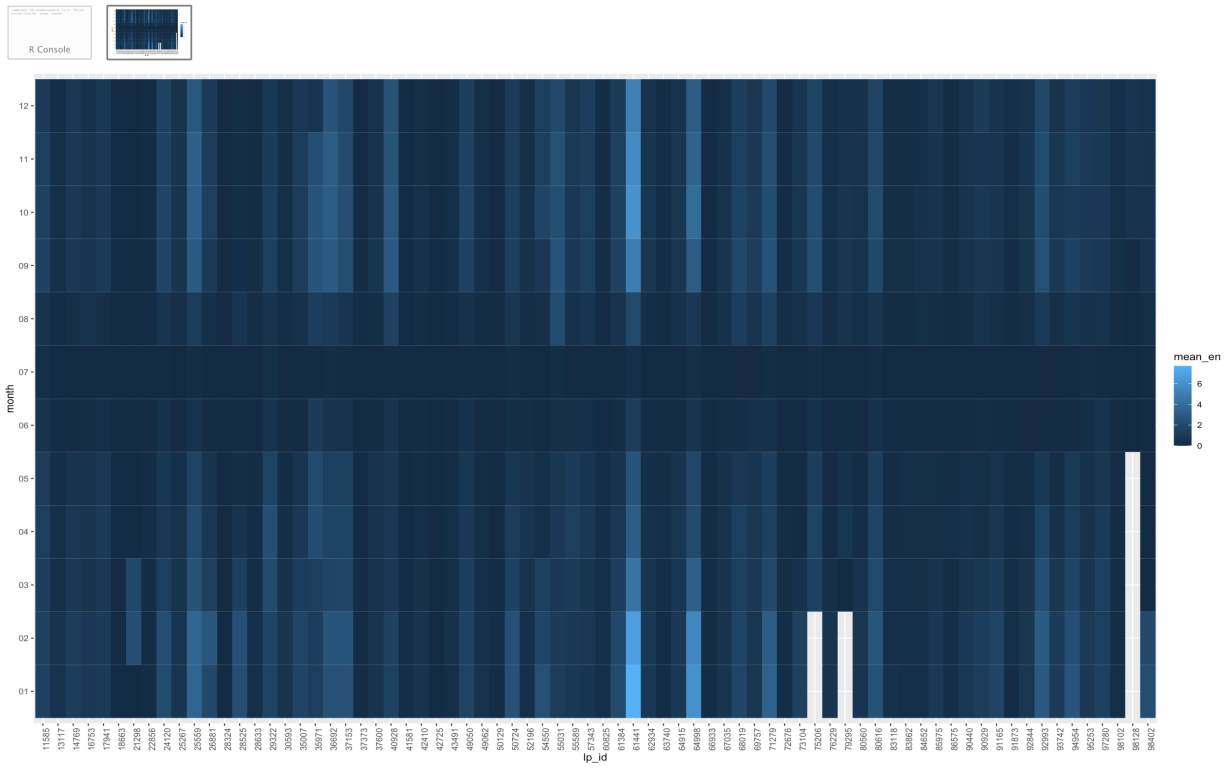


Figure 5: Figure 5: Engagement across Learning Platforms.



This figures is showing all the learning platforms by their id number in the x axis and the months on

the y axis. The color represents engagement across the platforms. One can note that there is very little engagement in months 6 and 8 and close to none engagement in month 7 which is July. This visual helps support my prediction.

Although this data was very detailed in student engagement across month and learning platforms, the data itself was not well distributed across the states provided. This can potentially be because some states have more school districts than others. Additionally, the US has 50 states total and this data only represented 23 states, leaving out information that can possibly change the overall data. Lastly, looking back at at Figure 4, there is hardly any engagement whatsoever in North Dakota, which probably isn't correct. This data was taken from government databases and some states may not have included all their data which can cause ambiguity about engagement levels in those states.

In conclusion, although the data was not well representative across school districts in the US, it provided some interesting findings. During Covid-19 there were more students downloads in learning platforms associated with language and reading. Interestingly, of the 74 learning platforms on average and overall, the math program 'i-Ready!' had the most engagement.

References

Bailey, D. H., Duncan, G. J. Murnane, R. J. & Au Yeung, N. (2021). Achievement gaps in the wake of Covid-19. Educational Researcher, 50, 266-275. https://doi.org/10.3102/0013189X211011237

Kaggle. (2020). LearnPlatform COVID-19 Impact on Digital Learning, Retrieved 04012022.from https://www.kaggle.com/competitions/learnplatform-covid19-impact-on-digital-learning/overview/description.

Kuhfeld, M., Soland, J., Lewis, K., & Morton, E. (2022, March 3). The pandemic has had devastating impacts on learning. what will it take to help students catch up? Brookings. Retrieved May 4, 2022, from https://www.brookings.edu/blog/brown-center-chalkboard/2022/03/03/the-pandemic-has-had-devastating-impacts-on-learning-what-will-it-take-to-help-students-catch-up/

Reilly Ph.D., J. (2018, November 4). Jamie's Linear Mixed Effects Model Notes. Jamie's linear mixed effects model notes. Retrieved May 4, 2022, from https://reillylab.github.io/Jamie_LinearMixedEffects.html

R. (n.d.). A beginner's guide to lmer. A beginner's guide to lmer. Retrieved May 4, 2022, from https://rstudio-pubs-static.s3.amazonaws.com/63556_e35cc7e2dfb54a5bb551f3fa4b3ec4ae.html